Quantifying the contribution of forcing and three prominent modes of variability on historical climate

Andrew P. Schurer[1]*, Gabriele C. Hegerl[1], Hugues Goosse[2], Massimo A. Bollasina[1], Matthew H. England[3], Michael J. Mineter[1], Doug M. Smith[4] and Simon F. B. Tett[1]

[1]School of GeoSciences, University of Edinburgh, Edinburgh, EH9 3JW, United Kingdom
[2]Université Catholique de Louvain, Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Louvain-La-Neuve, B-1348, Belgium
[3]Climate Change Research Centre and ARC Centre for Excellence in Antarctic Science, University of New South Wales, New South Wales 2052, Australia
[4]Met Office Hadley Centre, Exeter, EX1 3PB, United Kingdom
Correspondence to: Andrew Schurer (a.schurer@ed.ac.uk)

Abstract. Climate models can produce accurate representations of the most important modes of climate variability, but they cannot be expected to follow their observed time-evolution. This makes direct comparison of simulated and observed variability difficult, and creates uncertainty in estimates of forced change. We investigate the role of three modes of climate variability, the North Atlantic Oscillation, El-Niño Southern Oscillation and the Southern Annular Mode, as pacemakers of climate variability since 1781, evaluating where their evolution masks or enhances forced climate trends. We use particle filter data assimilation to constrain the observed variability in a global climate model without nudging, producing a near free running model simulation with the time-evolution of these modes similar to those observed. Since the climate model also contains external forcings, these simulations, in combination with model experiments with identical forcing but no assimilation, can be used to compare the forced response to the effect of the three modes assimilated, and evaluate to what extent these are confounded with the forced response. The assimilated model is significantly closer than the "forcing only" simulations to annual temperature and precipitation observations over many regions, in particular the tropics, the north Atlantic and Europe. The results indicate where initialized simulations that track these modes could be expected to show additional skill. Assimilating the three modes cannot explain the large discrepancy previously found between observed and modelled variability in the southern extra-tropics but constraining the El-Niño Southern Oscillation reconciles simulated global cooling with that observed after volcanic eruptions.

1 Introduction

Understanding the causes of observed climate change is crucial not only to gain knowledge on the past but also to improve projections of future change. It is common to split the drivers of climate change into two broad categories, external forcings, which could have natural or anthropogenic origins, and internal variability (see e.g. Hegerl and Zwiers 2011, Eyring et al 2021). Examples of external forcings include changes in greenhouse gases and anthropogenic aerosols as well as volcanic eruptions and changes in solar radiation. Internal variability is caused by chaotic fluctuations generated internally by the

climate system. Separating the external forcing in observed climate change from the internal variability background, is often referred to as detection and attribution (Hegerl and Zwiers 2011). Uncertainty in the model response to external forcing and

35 internal variability are responsible for much of the uncertainty in future projections (see Hawkins and Sutton 2009 and Lehner et al., 2020).

The combined effect of all external forcings or combinations of different forcings can potentially be determined by simulating them in climate models (e.g. Eyring et al. 2016) but these simulations will also include internal variability. For detection and attribution studies, which are primarily focused on the external forced component, the effect of the internal variability on the

40 model signal (often called the fingerprint of change) is reduced by averaging over large ensembles (Gillett et al. 2021). These studies typically treat internal variability as a statistical construct with properties calculated from pre-industrial simulations (piControl - model simulations which do not include any forcings) (see e.g. Allen & Stott, 2003, Ribes et al. 2017). Results can not only be used to understand the past but can also be used to constrain future projections (Kettleborough et al. 2007, Brunner et al. 2020). Other studies use large ensembles (see Maher, Milinski, and Ludwig 2021 for a review) to study the

45 combined role of internal variability and forcing (Deser et al. 2020) and changes in climate variability (Olonscheck et al. 2021). Since each simulation contains one realisation of internal variability the ensemble spread can be used to estimate the possible range of future climates (Deser et al. 2012). While large ensembles can be used to estimate a range of plausible pasts, there is zero probability that any one model simulation will have same detailed evolution as observed, due to chaotic behaviour in the climate system. Therefore determining internal variability and its contribution to past change is challenging and is typically

50 done by removing the forced component from observations either using model results (Hegerl et al. 2018, Friedman et al. 2020) or by de-trending (Knight 2005).

Since climate variability is very complex, acting across multiple time and space scales it is common to only study different aspects of it by isolating distinct modes of variability such as El-Niño Southern Oscillation (ENSO) (Timmermann et al. 2018) or the North Atlantic Oscillation (NAO) (Hurrell et al. 2003). These are typically assumed to be manifestations of internal

55 variability, and are found in piControl simulations, however they can also be influenced by external forcings (e.g Smith et al. 2020; Khodri et al. 2017). By focussing on a particular expression of variability in this way it is possible to study various aspects of its effect on climate. This includes: estimating the past effect of a particular modes of variability (Iles & Hegerl 2017, Hartmann et al 2013); determining whether it's possible to predict how the mode will evolve over the next couple of years (Chen and Cane 2008, Smith et al. 2020); and predicting if this pattern or amplitude is likely to change into the future

60 (Collins et al 2013, Cai et al. 2015). Some studies have attempted to simulate past changes by perturbing climate simulations to mimic the observed evolution of different climate modes. Kosaka and Xie (2013) prescribed observed SSTs over the central Pacific to force a simulation to follow the observed ENSO, while Delworth and Zeng (2016) and Delworth et al. (2016) prescribed heat flux anomalies to mimic the effect of the NAO on the Atlantic Ocean.

In this study, we adopt an existing data assimilation technique (the particle filter, van Leeuwen 2009) that has already been

65 used in a number of studies (for example Goosse et al. 2012) to assimilate three major modes of internal variability in a historical model simulation (starting in 1781) which also accounts for all the most important external forcings. For the start of

the simulations the modes assimilated will mainly rely on proxy reconstructions with instrumental observations used later when it becomes available, In this way the simulations will have both the external forced component of past change, in addition to the correct evolution of three of the most important modes of internal variability, without the need to impose any additional

70 fluxes or SSTs. In combination with simulations without any external forcings (piControl simulations) and simulations with realistic external forcings but non-assimilated internal variability, the roles of the different drivers of past change can be evaluated. In section 2 we will introduce the experimental set-up, while section 3 analyses how well it performs compared to observed climate. The article will then finish with a brief discussion and conclusions section.

2 Experiment Setup

75 We use a set of simulations with the coupled atmosphere–ocean model HadCM3 (Pope et al. 2000, Gordon et al. 2000). The atmosphere has a horizontal resolution of 3.75° × 2.5° in longitude and latitude with 19 vertical levels. The ocean model has a resolution of 1.25° × 1.25° with 20 levels. Despite the relatively low resolution and the no longer state-of-the-art physics modules, we expect the model results to be meaningful, as the model has shown a high level of skill, putting it consistently among the top half of CMIP5 models (see e.g. Flato et al. 2013, Sanderson et al. 2015, Knutti et al. 2013) and performs well

80 compared to CMIP6 models (Tett et al. 2022). The model is sufficiently fast and efficient to run the large quantity of model years required for this study. The forcings used are those described in Schurer et al. (2014), with the exception of the anthropogenic aerosol and ozone emissions which are updated to the CMIP5 protocol. Two sets of experiments have been run covering the period from 1781-2008.

The first set is a 10 member all forcing ensemble from 1781 to 2008 with time varying greenhouse gas and aerosol emissions,

85 land cover changes as well as natural forcings (volcanic, solar, orbital), initialised from existing simulations (the four ALL forced simulations and 4 NoAER simulations presented in Schurer et al., 2014). These simulations are those described and analysed in Brönnimann et al. (2019) and will be referred to as "forcing-only" in the rest of the article.

For the second set of experiments we use the same model set-up with the same forcings as described above, initialised from the same initial conditions, but using a particle filtering method without nudging (van Leeuwen 2009). This technique has

90 already been successfully used in several different studies (e.g. Goosse et al. 2012). It results in a physically consistent near-continuous simulation which tracks a set of chosen indices, which due to the nature of the filter need to have a low dimensionality (usually this would mean only using a small number of indices or strongly filtered spatial fields). In this case the North Atlantic Oscillation (NAO), the El Niño–Southern Oscillation (ENSO), and the Southern Annular Mode (SAM) were chosen as they have been found by numerous studies to have a dominant role on large-scale atmospheric variability (see

95 e.g. Eyring et al 2021). Our simulations cannot be expected to be as close to the "truth" as the Kalman filter or 4D-var approaches commonly employed by reanalyses for example the 20th century reanalysis (Compo et al. 2011), the last millennium reanalysis (Hakim et al., 2016) and the EFK paleo-reanalysis (Franke et al. 2017) which are assimilating much more information such as surface temperature and pressure observations. Instead this set-up will produce an analysis which will

behave like a free running model that happens to share relatively closely (but not perfectly) these three modes of variability with the observations. We can use this technique to see what effect simulating a realistic variability (on a seasonal scale) has on annual and decadal variability, and what aspects of decadal climate evolution are improved compared to fully free running model simulations with the same external forcing (but without the data-assimilation). This analysis will also identify regions where despite the assimilated variability observed changes remain inconsistent with the model simulation which could be interpreted as a possible indicator of model error or data uncertainty.

The particle-filter set-up is based on that described in Dubinkina et al. (2011). In our analysis, 50 model simulations are started in 1781 from initial conditions taken from the same simulations as the ensemble of 10 transient simulations. Every year, the simulations are stopped on the 1st April and the likelihood of each of the 50 simulations (often called "particles") is calculated based on the observations of the chosen three indices over the preceding 12 months. Another set of 50 simulations is then generated with initial conditions taken from the end of these existing 50 simulations, sampled according to their likelihood, with the lowest likelihood particles stopped and multiple new simulations initialised from the higher likelihood particles. A tiny perturbation is made to the atmosphere of each of the initial conditions and 50 new simulations are run for the next year. The likelihood, p, of a particle, ψ, given the observations, d, is based on a Gaussian density,

$$p(d|\psi) = K^{-1}exp\left[-\frac{1}{2}(d - H(\psi))^T C^{-1}(d - H(\psi))\right] \quad (1)$$

where K is a normalization constant. H is an observational operator which maps the model output onto the observational phase space and C is the error covariance matrix which here takes into account the observational uncertainty (see Dubinkina et al. 2011 for more details). The likelihood is thus dependent both on the closeness of each particle to the observed index but also on the relative uncertainty of each index. The number of particles to spawn from each individual simulations is proportional to its likelihood following an iterative process described in Dubinkina et al. (2011) with the only modification that here we only allow each particle to get re-spawned a maximum of 20 times, so maintaining a spread of initial conditions for each new iteration.

A schematic, shown in Fig. 1, illustrates the particle filter technique used here, with just 10 particles and one assimilated index. This toy-example shows the performance for 10 assimilation steps where at each step the initial conditions for the next step are chosen from the previous particles based on the likelihood calculated with equation 1. The assimilated product can be derived at each step for any variable by calculating a likelihood-weighted mean of all simulations using the likelihood in Eq.1. We term this the weighted mean. A sequence of continuous particles can be found by time-reversing the trajectory by linking each particle to its parent (Fig. 1). After only a few assimilation steps back it is inevitable that they will all converge onto one particle and will follow the same sequence thereafter. We term this the continuous particle and for simplicity it will be derived by only following the most likely particle at the final assimilation step backwards. In the particle filter experiment this pseudo-continuous simulation will also exist (with only tiny a perturbation at machine precision being applied to the atmosphere at each restart step). In theory, such a simulation could also exist without any assimilation step at all, but this would require a

prohibitively large number of simulations. Unless stated differently results will be shown for the weighted mean and will be referred to as the "DA" (data assimilation) model simulation.

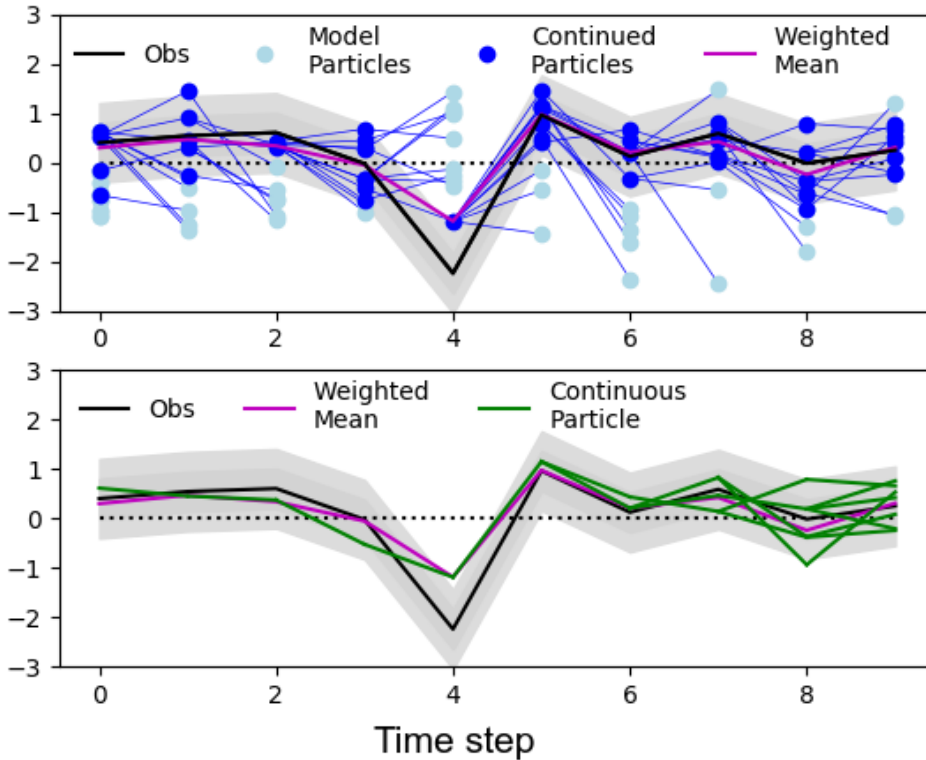



Figure 1 – Schematic showing the performance of the particle filter. *Values are only illustrative and do not represent any particular climate variable. Observations – black with grey uncertainties. Model simulations blue, simulations that are continued use a darker blue. Weighted mean purple, continuous simulations green (note that this quickly converges onto a single simulation as you follow the simulations back in time from the end of the experiment).*

The indices chosen are shown in Fig. 2. The data used to constrain the data assimilation simulation is intended to be the best available for that particular index. Hence, instrumentally observed indices are used when available and proxy reconstructions when they are not. The consequence of this is that the uncertainty of the target index changes through time, as does the relative uncertainty between the indicies and thus the constraint applied to the simulations. The assimilation step was chosen to occur on April 1ˢᵗ, so the simulations are restarted from initial conditions which best agree with the boreal winter which has just occurred (the period in which the most information is assimilated).

For the NAO index the metric used is the mean of the NAO index over the period December to March and is calculated as the standardised difference between the standardised pressure of the Icelandic low and the Azores high, as in Luterbacher et al. (2001). For ENSO the metric used is temperature anomaly over the Nino3.4 region (5N-5S, 170W-120W). For 1781-1881 this is the mean over the months November to February and is obtained by the mean of the proxy reconstructions of Li et al. (2013)

and Emile-Geay et al. (2013). In 1882 the metric changes to take advantage of the more reliable instrumental observations (HadSST3, Kennedy et al. 2011a; 2011b), with two periods used, the mean of the Nino3.4 index from April to September and the mean of the index from October to March, to find particles that follow the ENSO evolution during the whole year. The SAM index is initially the annual reconstruction of Abram et al. (2014). Given that the assimilation occurs on the 1st April every year, it means that the first part of the year used to calculate the SAM index is from the previous assimilation step. In 1957 the filter switches to use the instrumental Marshall index (Marshall 2003). Given that monthly data exists the mean of the period April through to March is used to coincide with the assimilation time-step and the metric for this period is calculated as the mean of the monthly values.

Given the finite number of particles (50), none of the individual simulations is capable of perfectly matching all three indices simultaneously. Which index will be given the most weight will be determined by the uncertainty of that particular index (see eq. 1). The uncertainty associated with each of the indices is shown in the supplement (Fig. S1). The uncertainties in the proxy reconstructions for the ENSO and particularly the SAM are much larger than the NAO instrumental uncertainty. Thus, the analysis initially tracks the NAO better as the filter preferentially tries to fit this index, as shown by correlations between the assimilated and observed index in Fig. 3 (correlations for the "continuous particle" shown in the supplement Fig. S2). Uncertainties in ENSO and then the SAM decrease during the instrumental period (substantially in the case of the SAM) thus the correlation to these indices improves (with the match to the NAO noticeably degrading after 1950). The performance of the filter is also reflected in the number of particles that are re-spawned at each iteration (Fig. S3). This is about 17 till 1881, dropping to 10 when the ENSO instrumental data is used, and below 5 when the SAM instrumental data starts in 1957. As the number of particles which match the observed indices, within their associated uncertainty decreases, so more particles are discontinued and more spawned from the remaining particles. Towards the end of the simulation the filter often only re-spawns from 3 particles, the minimum allowed. It is important to note that improvement in the agreement on an annual scale is not necessarily related to agreement on a decadal scale as modelled values consistently higher or lower than those observed can lead to large decadal disagreement, for example in the decadally-smoothed ENSO index in the mid-20th-century (Fig. 3).

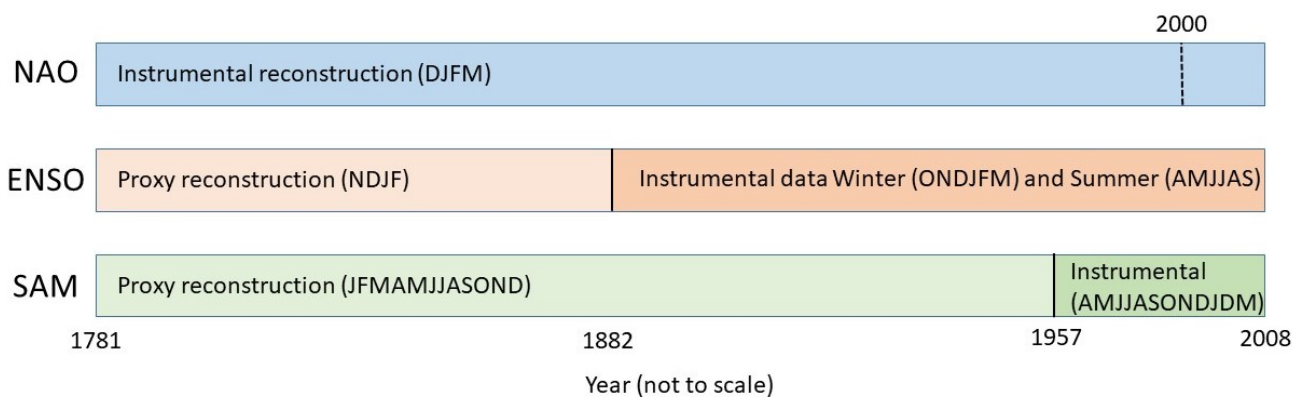



Figure 2 – Schematic diagram of target indices – *NAO instrumental reconstruction uses* (Luterbacher et al. 2001) *up until the year 2000 and then the 20th Century reanalysis* (Compo et al. 2011) *thereafter. ENSO proxy reconstruction uses a*

combination of Emile-Geay et al. (2013) *and* Li et al. (2013) *until 1882 and then HadSST3* (Kennedy et al. 2011a, Kennedy et al. 2011b) *thereafter. SAM proxy reconstruction used is* Abram et al. (2014) *until 1957 and then the Marshall Index* (Marshall 2003) *thereafter.*

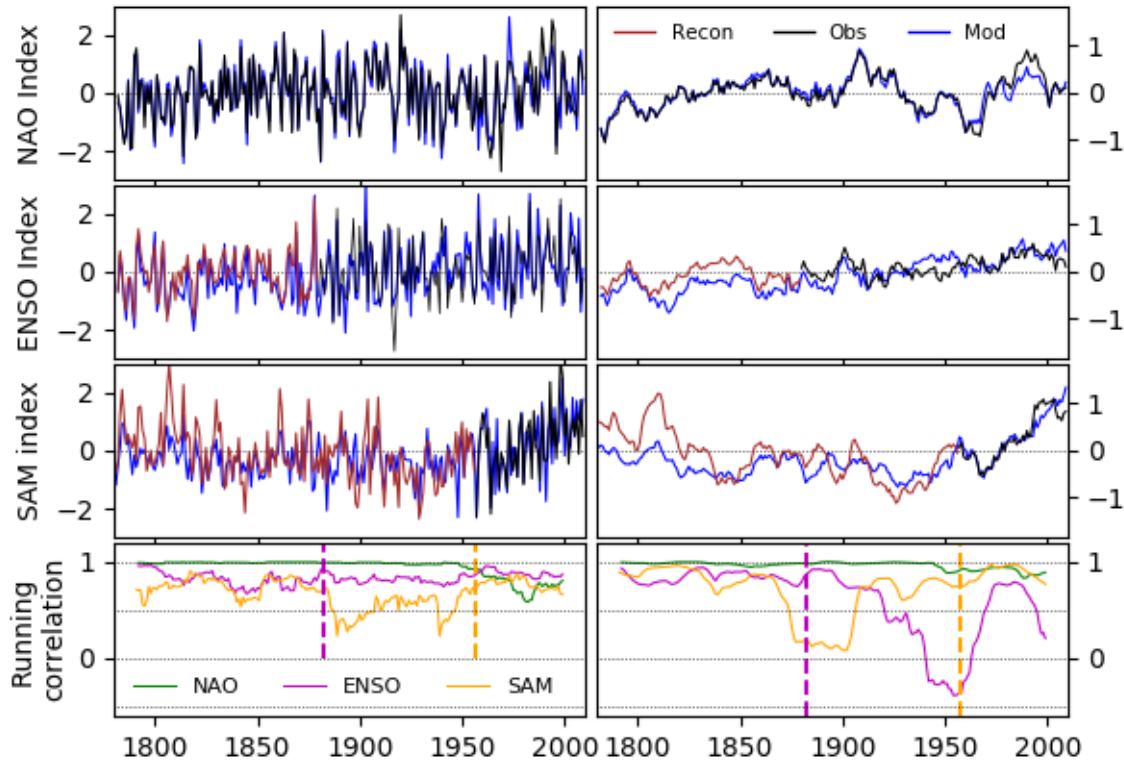



Figure 3 – Performance of filter – Top three rows – time-series of target indices, *NAO (DJFM, climatology period 1782-2008), ENSO (NDJF, climatology period 1882-1992), SAM (Annual mean, climatology period 1957-1995) for the weighted mean analysis (blue), Reconstruction (brown), Instrumental Observations (black). Note that because of changes in the target indices through time (Fig 2) the plotted timeseries does not always exactly correspond to the data assimilated. Left column - annual values, right column – annual values smoothed by an 11 year running mean. Bottom panel shows, left – running 20-year correlation for the annual value, right – 40-year correlations for the decadal values for the NAO (green), ENSO (purple), SAM (orange), vertical dashed line show the switch between proxy and instrumental values for ENSO and SAM.*

3 Results

3.1 Performance Validation

The performance of the particle filter experiment is assessed by analysing how well it reproduces the spatial patterns associated with the assimilated indices. The NAO, ENSO and SAM indicies were calculated in the observational record and in the particle filter simulation and in Fig. 4 the sea level pressure fields are regressed on these indicies. A comparison of the model pattern (right panels) to the observational counterparts (left panels) can be used to assess whether the HadCM3 climate model is capable of accurately simulating the spatial patterns of these modes of variability. In general the model performs well, however there are some noticeable biases in particular in the temperature response to ENSO, which is known to extend too far west in HadCM3 (see also Collins, Tett, and Cooper 2001), and in the temperature response to the SAM. The central panels show the DA simulation fields regressed on the observed index, the difference between these patterns and the observed regression patterns are shown in the supplement (Fig. S5). This represents a test of the assimilation procedure. If the particle filter was performing perfectly the modelled indices would be identical to the observed indices and consequently the central and right panels would be identical. This is almost the case, (although the patterns using the observed index are slightly weaker), which is reflection of the fact that the modelled indices follow the instrumental observations reasonably well although not perfectly (Fig. 3). Regressing the DA simulation fields on the proxy SAM reconstruction results in a much weaker pattern (not shown) reflecting the loss of performance in this earlier period.

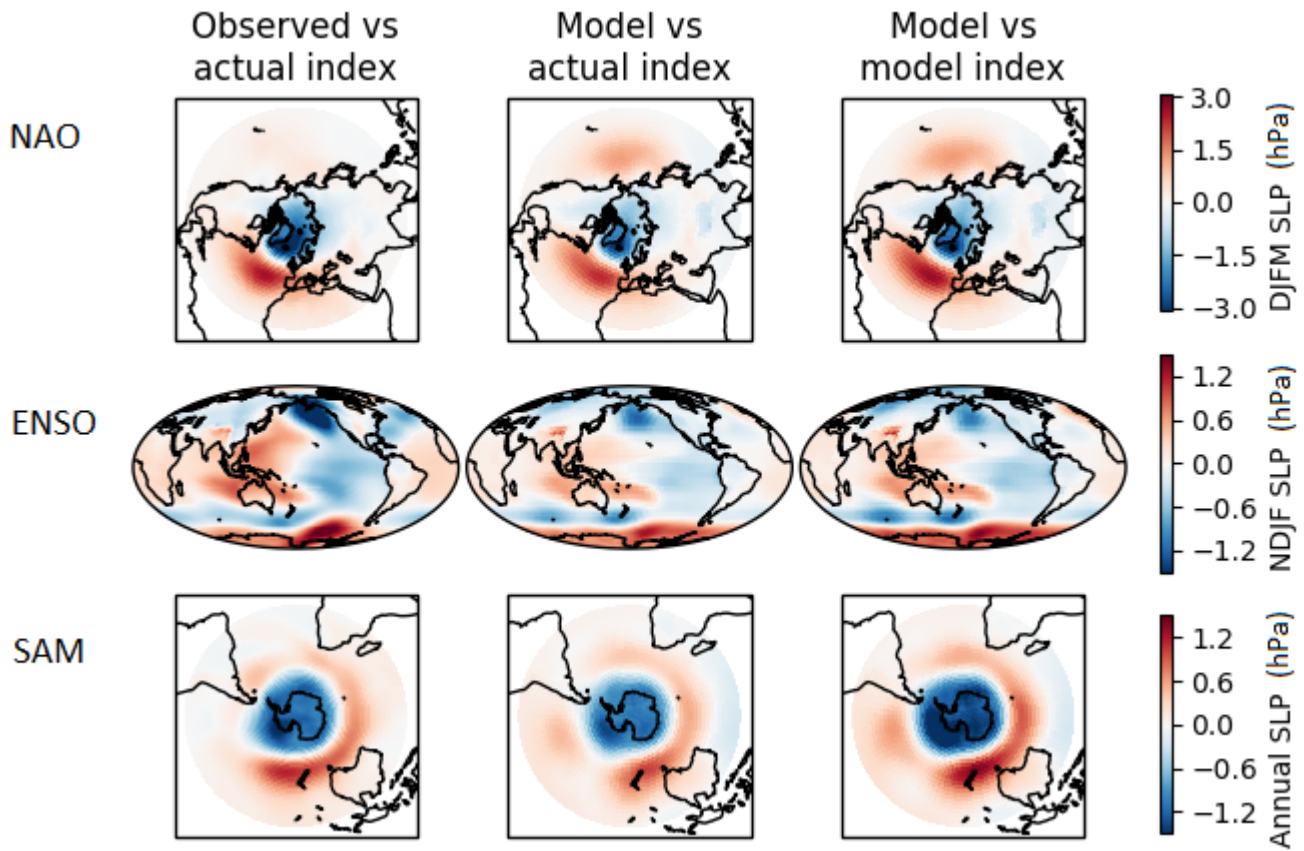



Figure 4 - Regression between index and sea level pressure. From top to bottom: NAO, ENSO, SAM. Left column, regression between the observed index and the observed spatial pattern of sea level pressure (from 20CRv3 Slivinski et al. 2021). Middle column, regression between the observed index and the DA model spatial pattern. Right column regression between the DA model index and its own spatial pattern. All data is first detrended. Regressions calculated over the period 1882-2008 except the SAM regressions which are calculated over the shorter period 1957-2008.

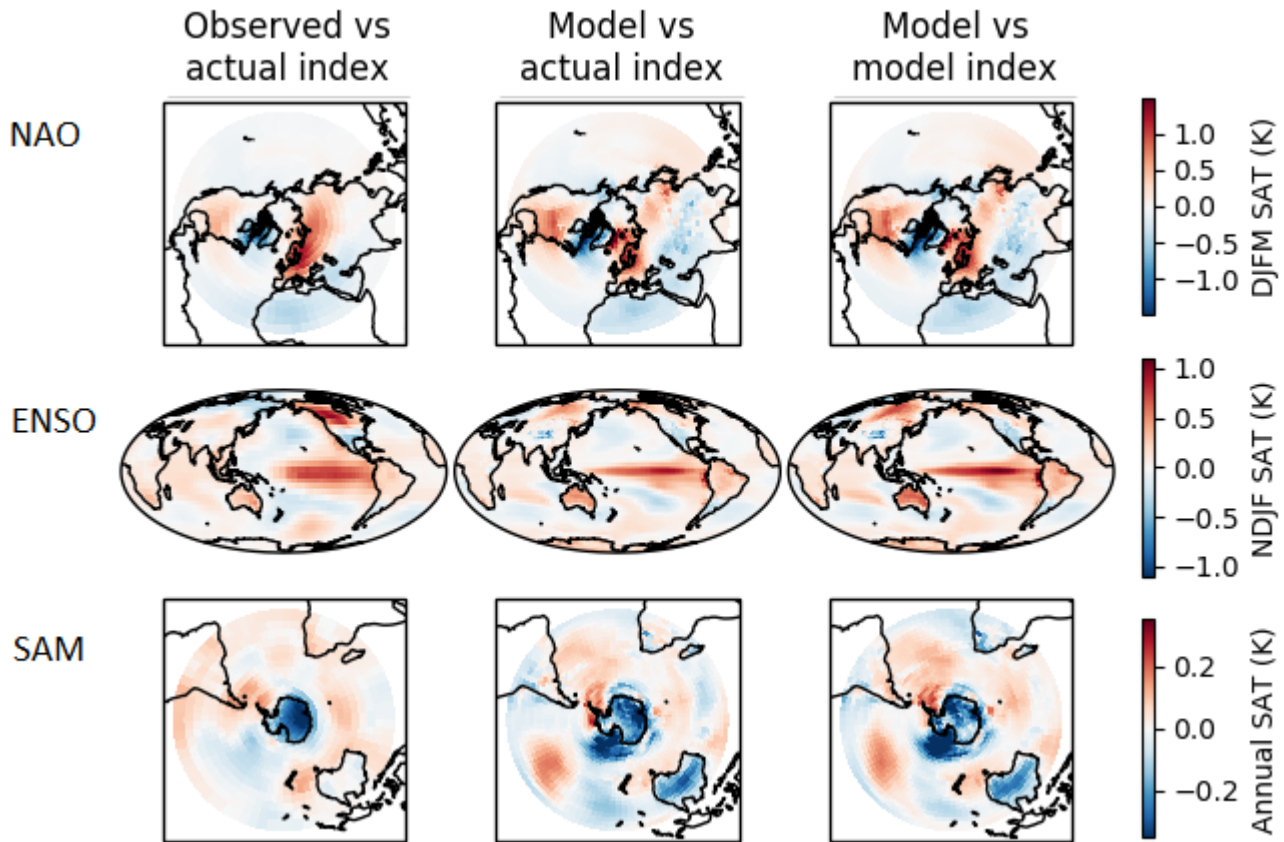



Figure 5 - Regression between index and surface air temperature. From top to bottom: NAO, ENSO, SAM. Left column, regression between the observed index and the observed spatial pattern of SAT (from HadCRUT5 (infilled); Morice et al. 2021). Middle column, regression between the observed index and the DA model spatial pattern and. Right column regression between DA model index and its own spatial pattern. All data is first linearly detrended. Regressions calculated over the period 1882-2008 except the SAM regressions which are calculated over the shorter period 1957-2008 due to data availability. Grey indicates missing data.

3.2 Correlations of models with observed climate

In order to determine in which regions the particle filter is providing a statistically improved realisation of the observed climate compared to the forcing-only simulations we estimate skill by using correlation coefficients, as suggested in Goddard et al. (2013). Detrended observed anomalies are correlated with those simulated by the model with and without assimilation for each grid cell to determine where the assimilation is significantly improving the agreement. Because the particle filter results are a weighted mean over a number of particles (see supplementary Fig. S3) the internal variabilty not associated with the three assimilated modes will be reduced. Consequently comparing them directly to each of the forcing-only simulations could give

misleading results. To account for this, the average weights which are used in the particle filter experiment for the period analysed are calculated (see supplementary Fig. S4) and then applied at random to the 10 forcing-only simulations to create 100 different equivalently weighted-means. Results are considered significant if correlations calculated between the observations and the particle filter experiment are greater than 95% of the equivalent values calculated between the observations and the weighted means from the 10 forcing-only simulations. In order to focus exclusively on the skill gained from including the assimilation, we calculate the difference between the correlation calculated for the assimilated experiment and the mean correlation calculated for the non-assimilated forcing-only simulations (see Fig. 6). In order to provide a test of bias an equivalent figure displaying mean absolute error (instead of the correlation) is show in the supplement Fig. S10. The correlations between the assimilated and non-assimilated model simulations and the observations are shown in the supplement (Figs. S6, S7).

Spatial patterns of correlation differences are shown in Fig. 6a and b for annual and boreal-winter (DJF) surface air temperature (the focus is on boreal winter as this is when most data is assimilated, results for other seasons are shown in the supplement Fig. S9). In both cases a substatial fraction of the total number of grid squares displays significantly higher correlations than would be expected due to chance. Improved correlations are highest in the central and eastern tropical Pacific due to the effect of ENSO and regions with strong connections to this mode of variability will show similar behaviour in both models and observations, for example the Indian ocean and tropical Atlantic (Klein et al. 1999). Outside of the tropics teleconnections to the ENSO variability are much weaker. Other significant increases in correlation (above that expected by chance) are seen in the northern extratropics, in the north Atlantic, Europe and Siberia, particularly in winter. This is the region most strongly affected by the NAO (see Hurrell et al. 2003 and Figs. 4 and 5) which is assimilated for this period (December – March). To further investigate correlations over this region we have analysed results using a spatial European reconstruction (Luterbacher et al. 2004) over the full period of the experiment; 1782-2000 (Figs. 6c and d), in winter and in the annual mean. Correlations are found to be particularly strong over the northern part of Europe in winter, consistent with the effect of the NAO. The effect of the SAM on temperature is not so clear, partly due to limitations in observational coverage. However, there are high and significant correlations in parts of the Southern Ocean particularly south of Australia and off the southern tip of South America, regions strongly effected by the SAM (see e.g. Fogt and Marshall 2020) and which have good agreement between models and observations in Fig. 4. The results for mean absolute error (Fig. S10) are broadly similar showing reduction in bias in the regions which are found to have higher correlations. Note that Figs. 6a-d use the infilled analysis product of HadCRUT5, but the main findings are insensitive to the infilling process (see Fig. S8).

Differences in how well the assimilated and forced simulations correlate to observed precipitation are shown in Figs. 6g and h. Due to sparcity in observations the anlaysis is restricted to land only regions with a start date of 1950. As for temperature, significant improvements are mainly found in the tropics particularly in the Pacific, but also in Europe and southern Australia consitent with the influence of the NAO and the SAM.

One striking aspect of Fig. 6 is that, even for temperature, the assimilation only significantly improves the correlations for about a third of the globe at interannual timescales, with much of the significantly correlated area in the tropics. This is an

indication of the relative importance of the three modes of variability assimilated here compared to that of other components of internal variability. At higher latitudes the atmosphere's intrinsic variability is much higher, and although important, the relative role of the indices assimilated here can only explain a relatively small component of the total variability. This demonstrates both the limitations and strengths of this approach; while it will not give as accurate a realisation of past climate as reanalyses which assimilate as much observed daily (or sub-daily) data as possible it can instead be used to assess the relative role these three indices play in determining the climate.

One interesting aspect is how much variability these indices can explain on a decadal time-scale (filtering using a running 11-year mean). As Figs. 6c and d show, the most prominent area of improved correlations on these longer time-scales is the North Atlantic, in the sub ploar gyre region, which is an area that has previously been identified as one with a particualarly strong link to the NAO, and which could play a key role in driving ocean circulation (see e.g Zhang et al. 2019 for a review). There are also significant correlation improvements in surface temperature on a decadal scale in the Pacific and Indian Ocean, over northern Eurasia, Africa and in the Southern Ocean. Due to the relatively short period analysed, results for precipitation on a decadal scale are very noisy (Figs. 6i and j). However they do show an increased agreement in the tropical Pacific and in Europe, consitent with the results for surface temperature.

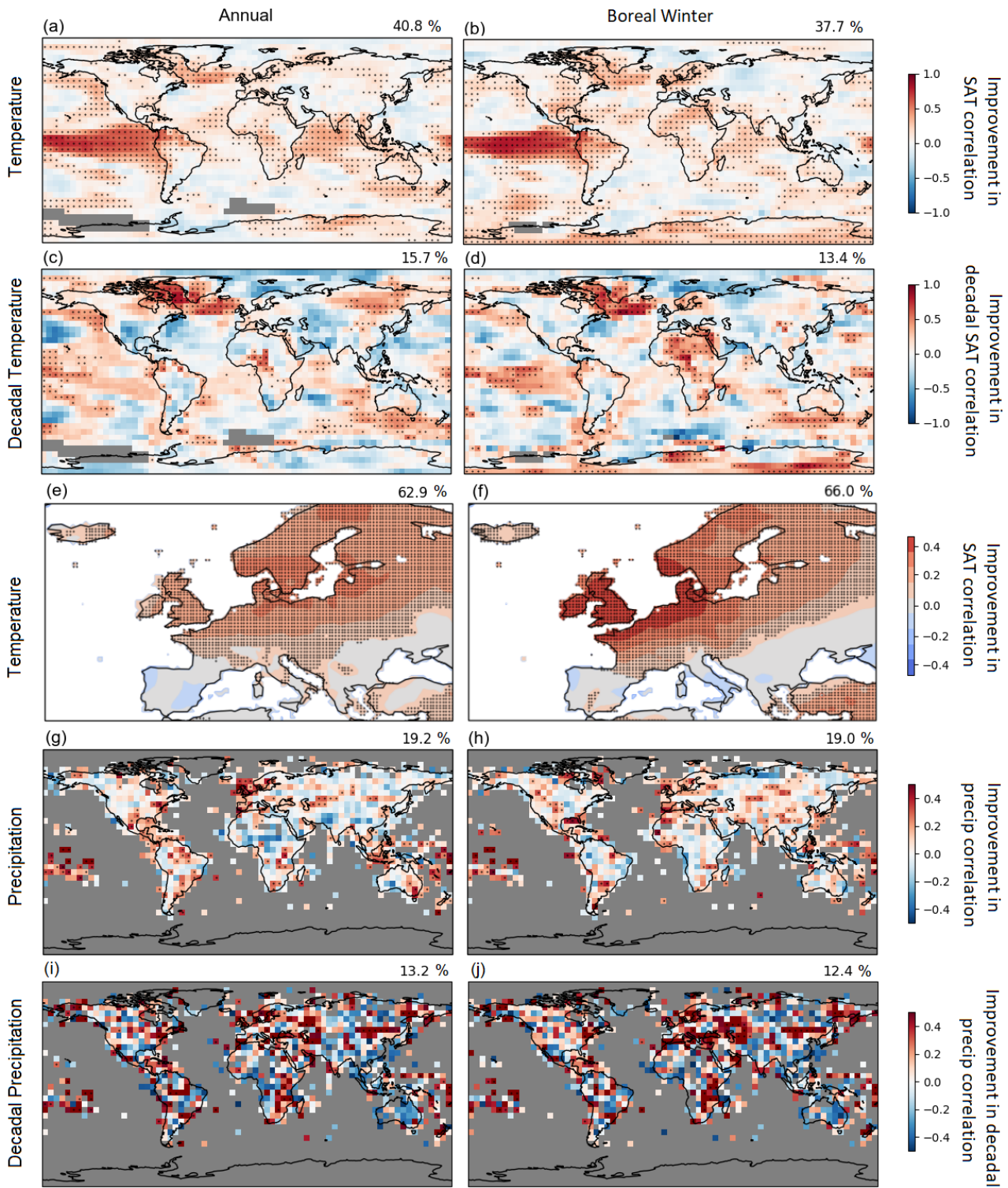



Figure 6 – Improvement in temperature and precipitation annual and NH winter (DJF) correlation between models and observations due to assimilation *(The difference in correlation between observations and the model simulations with and without assimilation. a) and b) Temperature correlations using HadCRUT5 (infilled; Morice et al. 2021) for period 1882-2008. c) and d) decadal (11-year running mean) temperature correlations using HadCRUT5 for period 1882-2008. e) and f) Correlation using a temperature reconstruction (Luterbacher et al. 2004), for the period 1782-2000. g) and h) Correlation using GPCC precipitation dataset (Schneider et al. 2016) for the period 1950-2008.(i) and (f) Decadal (11-year running mean) precipitation correlation using GPCC for the period 1950-2008. Stippling indicates where correlation is bigger than the 95th percentile from the weighted means of the 10 forcing-only model simulations. The number at the top right of every plot indicates the percentage of grid cells which are stippled. If only due to uncorrelated random variability a value of 5% would be expected.*

3.3 Large-scale mean climate

Results for correlation against observed large scale average temperature support the spatial correlations discussed above. The agreement is significantly improved by the assimilation in the global mean and in the tropics and SH extra-tropics (Fig. 7). Temperature in the northern extra-tropics show an improvement in the particle filter during the second half of the century (particularly north of 35N) with larger correlations than in any simulation without data-assimilation, but not during the first half of the century, with values particularly low in the period 1900-1940 and no statistical overall improvement for the period as a whole. A similar result is found in correlations over Europe (Fig. 7h) with very low values at the start of the 20th century. As the NAO has a particularly strong influence on Europe (Hurrell et al. 2003) and the DA model follows the observed NAO closely throughout this period, these weak correlations suggest the effect of the NAO on Europe and northern high latitude regions is not constant through time, although uncertainties in observations could also contribute. Other studies (Weisheimer et al. 2019) have also found non-stationarities in the skill of initialised climate models to predict the NAO with higher values since the 1970s and lower in the mid-20th century. In general, relatively small improvements in correlation of temperature at large scale and over long time periods are expected because at these scales the effect of external forcing will dominate and much of the discrepancy between simulated and observed climate is likely to be due to errors in the simulated forced response. This is particularly true for higher latitudes where previous studies have found that despite a strong regional influence, the NAO only weakly affects large scale temperatures (Iles & Hegerl 2017).

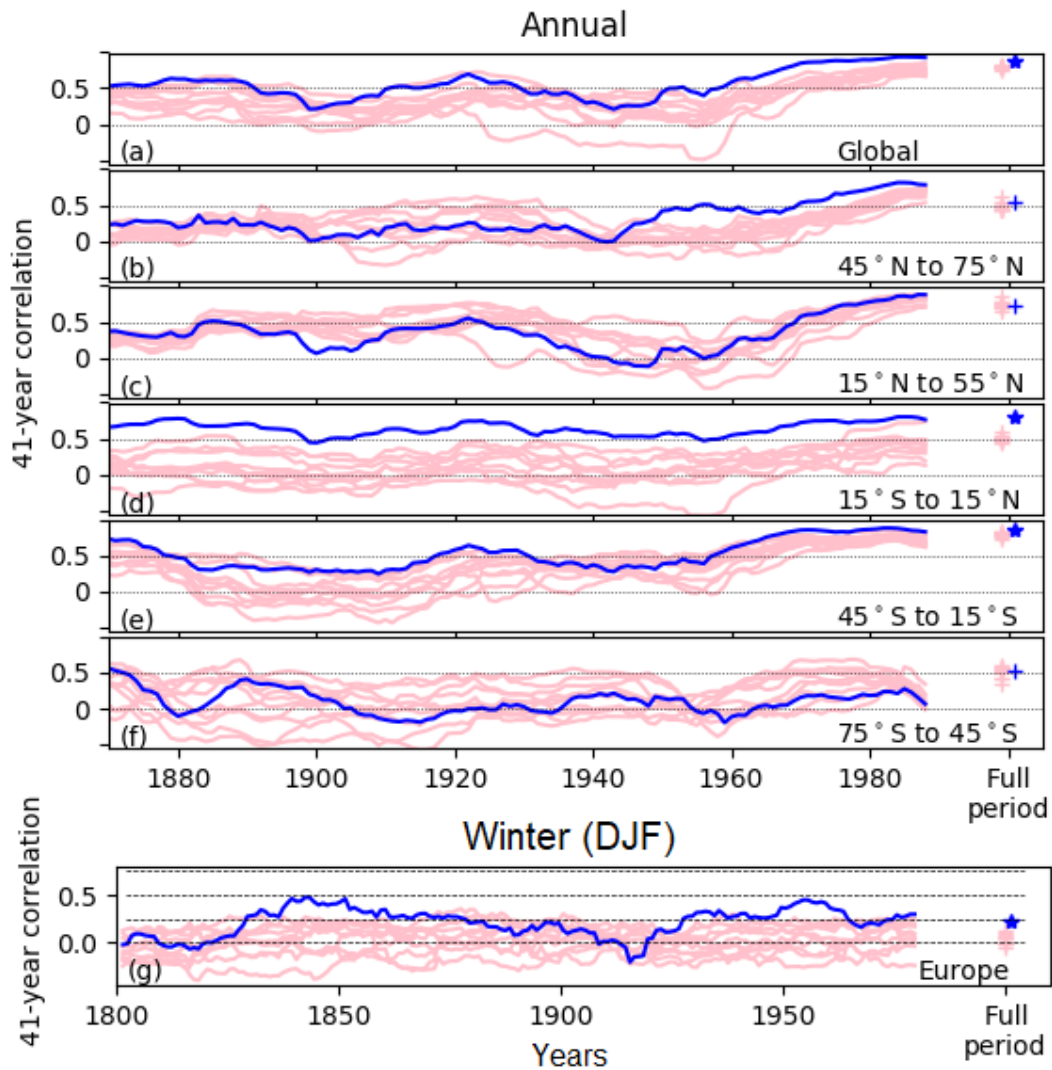



Figure 7 41-year sliding correlations for annual zonal-mean and winter European air temperature. (a-f) *Correlations are calculated between observed and simulated air temperature. (a-f) Correlations between zonal mean temperatures for latitudes given in the bottom right of each panel between HadCRUT5 (non-infilled) and* forcing-only *simulation (pink) and data-assimilation simulation (blue). All simulations masked to available observational data coverage. (g) Correlations between a reconstruction of European Winter (Luterbacher et al 2004) and simulations. Correlations over the full period on right of plot, if correlation greater than all 10* forcing-only *simulations the blue symbol is plotted as a star (instead of a plus).*

305   Previous studies, for example Friedman et al. (2020) and Hegerl et al. (2018) have found that observed decadal temperature variability in the southern hemisphere, particularly the extra-tropics, is far greater than that simulated by the latest climate models, and is outside the CMIP5 ensemble spread. As we are assimilating the key modes of variability for this region (the ENSO and SAM), it is an interesting question to investigate whether our DA model simulation can offer insights into this discrepancy. Time series of surface air temperatures averaged over zonal bands, including the SH extra-tropics are shown in

310   Fig. 8. As expected, the observed variability over 65 to 30S is outside the range of the forcing-only simulations. Assimilation does little to improve this situation, suggesting that it is neither the SAM nor the ENSO variability which is responsible for the discrepancy. However, this could partially be due to possible deficiencies in the model's pattern of the SAM (see Figs. 4 and 5). Other intriguing possibilities are limitations in model variability in the Southern Ocean (Hyder et al. 2018), potentially due to the relatively coarse resolution (Beadling et al. 2020), as well as possible biases in SST measurements such as those found

315   in the NH Atlantic and Pacific (Chan et al. 2019), and using coastal weather stations (Cowtan, et al. 2018).

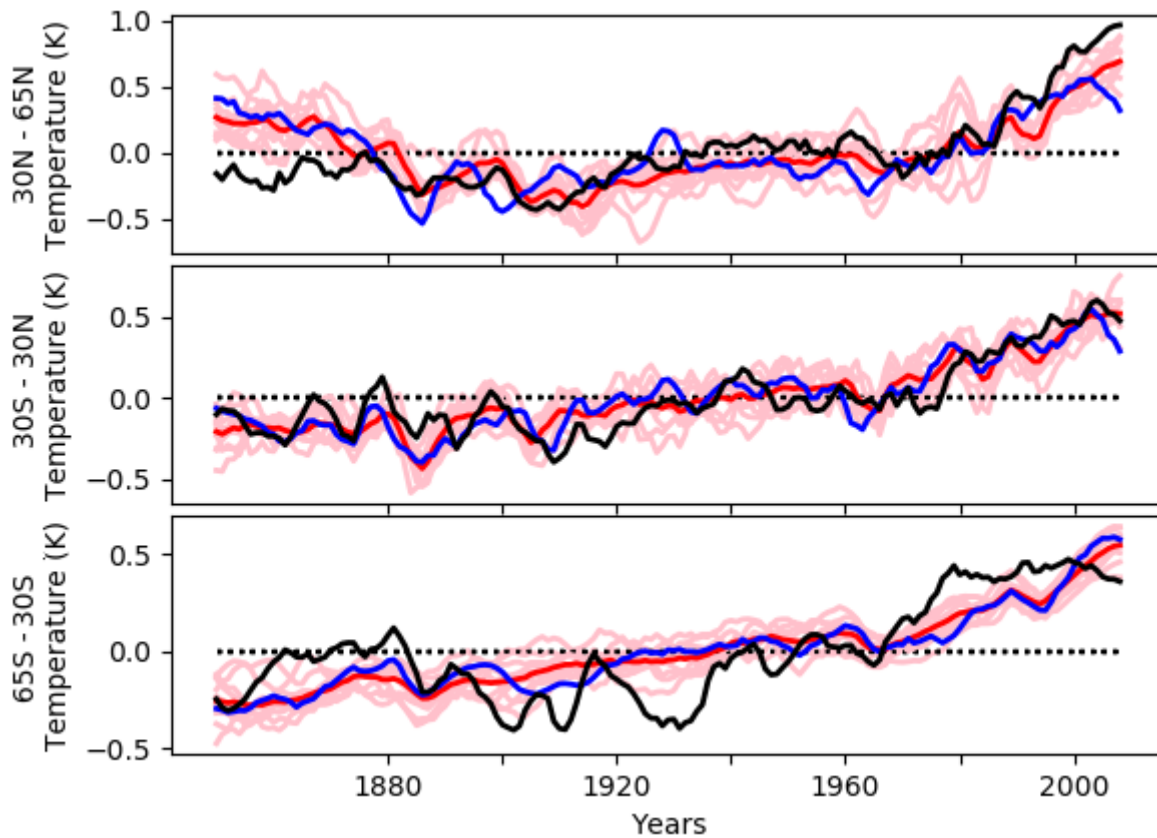
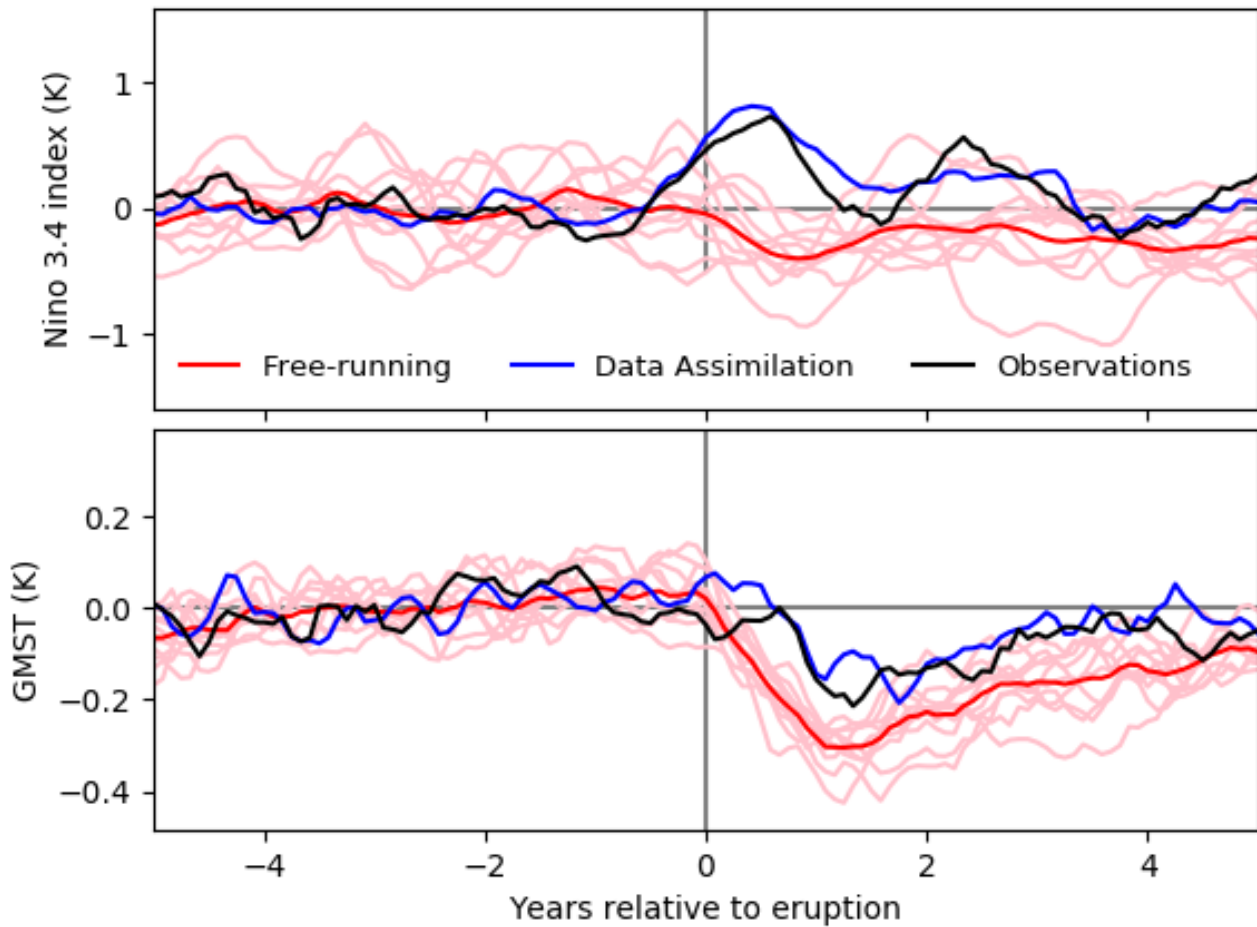



Figure 8 – Mean annual surface air temperature averaged over three latitudinal bands *Assimilated model (blue), 10 forcing-only simulations (light pink),forcing-only ensemble mean (red), observations (HadCRUT5; non-infilled; black). All timeseries filtered using a 5-year running mean and simulations are masked to available observational data coverage.*

3.3 Temperature response to large volcanic eruptions

Volcanic eruptions have an important role on annual as well as decadal climate (Robock 2000). However, the global mean temperature response has been found to be stronger in climate models than in observations (Chylek et al. 2020, Bindhoff et al 2013). It has been suggested that this discrepancy could be explained, at least partially, by ENSO events which coincide with all major volcanic eruptions during the instrumental era (Lehner et al. 2016), although the role of the volcanic eruption in triggering the occurrence of an El-Nino is still disputed, see . Studies of the last millennium which rely on proxy evidence offer mixed results from a longer term perspective with some recent studies, typically based on tree-ring evidence supporting an El-Nino like response in the year following an eruption (e.g. McGregor et al. 2010, Li et al. 2013), while studies based on coral data from the ENSO region itself do not find any significant relationship (Tierney et al. 2015, Dee et al. 2020). Models of different complexities also simulate an El-Nino response, although they disagree on the underlying mechanisms (Ohba et al. 2013, Khodri et al. 2017, Predybaylo et al. 2020, Maher et al. 2015, Hermanson et al. 2020), while the majority of studies concentrate on tropical eruptions, modelling studies have also found an important role for high latitude eruptions, also the mechanism maybe different (Pausata et al. 2015). Given that the data-assimilation simulation has been designed to follow the observed ENSO index it provides an ideal test bed for investigating the effect of the ENSO on the climate after volcanic eruptions. Although it cannot provide any evidence to address the question of whether a forced link exists between the eruption and the ENSO state.

To investigate the temperature response following volcanic eruptions we employ a commonly used technique called an epoch analysis (e.g. Hegerl et al. 2003). This involves averaging the response across multiple eruptions in order to reduce internal variability to focus on the forced response. Here we choose the 5 strongest tropical eruptions which coincide with the period of instrumental observations: Krakatoa (August 1883), Santa Maria (October 1902), Agung (March 1963), El Chichon (April 1982), Pinatubo (June 1991). The global mean surface air temperature and Nino3.4 index are calculated for the simulated and observed climate for 60 months after each eruption occurred relative to a pre-eruption mean (calculated as the mean over the 60-months directly before the eruption) and the results for all 5 are averaged together. Fig. 9 shows the results for the epoch analysis for the last 5 major tropical eruptions. The ENSO index concurrent with the eruptions is on average slightly positive when the volcanic eruptions occur and becomes increasing positive for the year after the eruptions. This behaviour is captured remarkably well in the DA simulation but not by any of the forcing-only simulations. Based on the results from the forcing-only simulations one would conclude that the model overestimates the response to volcanic eruptions, since all model simulations have larger cooling following the eruption than that observed. Assimilating ENSO, however resolves this clear mismatch, showing that, at least for this model, if a simulation has the correct ENSO evolution the temperature response is very close to that observed.

355   **Figure 9 – Response to 5 tropical volcanic eruptions.** *Epoch analysis for Top -observed Nino3.4 SST anomaly index. Bottom - global mean surface air temperature. Observations are HadSST4 (Kennedy et al. 2019), and HadCRUT5 (non-infilled), all simulations are masked to the observations. Both variables are plotted as anomalies from the mean of the 5-years before the eruption date.*

4 Discussion and Conclusions

360   In this study, we have introduced a new experimental set-up, based on existing data-assimilation techniques, to produce a near-continuous and near free-running model simulation with a realisation of the ENSO, NAO and SAM similar to that which actually occurred. We have demonstrated that this method adequately captures the assimilated modes and offers an improved realisation of the climate above that of free-running simulations without assimilation in several key regions in both temperature and rainfall. This includes: annual and winter mean surface temperature over much of the northern extratropics, large parts of

365   tropical surface temperature, and tropcial and European rainfall. On a decadal scale improvements in skill are found over much of the tropical Pacific, over some parts of the Southern Ocean and in the sub-polar gyre region of the northern Atlantic. Decadal forcast experiments have shown that the NAO (Smith et al. 2020) and ENSO (Barnston et al. 2019; Dunstone et al. 2020) are predictable on seasonal to annual timescales. This analysis highlights where a prediction system has the potential to better reproduce observed trends, if the prediction systems were capable of following the correct realisation of these modes.

370   Interestingly our results also display some evidence of non-stationarity, with the assimilation improving agreement to observations more in some periods than others, suggesting that the skill in decadal forecasts could vary through time.

Past modelling studies have shown that the observations in the southern extratropics are far more variable on the decadal scale than model simulations. By assimilating the ENSO and the SAM we can show that getting the evolution of these modes of variability closer to reality does not help to explain this discrepency. This suggests that the difference must have an alternative

375   origin, such as observational errors or other model deficiencies. Outside this region though, there is no major discrepancy in Surface air temperature, providing support for the ability of climate model's to reproduce observed climate evolution.

Consistent with most models, the HadCM3 simulations analysed here show too much cooling following volcanic eruptions over the last 150 years. Correctly assimilating ENSO results in simulated cooling that is consistent with the observations. This demonstrates that the model response is not necessarily too strong, rather that the forcing-only simulations do not capture the

380   correct ENSO response. This conclusion highlights that relying on model simulations which do not capture the observed ENSO variability could lead to misleading conclusions.

We therefore consider this use of a particle filter to be a promising experimental design which could be developed further in the future to investigate the causes of past variability or the effect of potential future changes in key circulation modes using a storyline approach. We note that while we have chosen to assimilate three modes of variability in this study, it would be

385   possible to assimilate any number of them. Assimilating fewer modes would reduce the degrees of freedom for the filter and hence allow an improved performance for that mode or alternatively the ability to use less particles, while assimilating more has the potential to reduce the performance of the filter unless the number of particles are also increased.

Data availability

Model simulation data is all publicly available. The particle filter simulation data is available at https://doi.org/10.7488/ds/3829

390   The all-forced HadCM3 simulations (without assimilation) is available at at https://doi.org/10.7488/ds/3827.

Competing interests.

The contact author has declared that none of the authors has any competing interests.

Author contributions.

395   AS, GH and HG designed the study. AS and MM setup the climate model experimental design with support from HG and ST. AS ran the model simulations and conducted all analysis. All authors contributed to the manuscript writing.

Acknowledgements

The model simulations and AS, GH, MM and ST were funded by the ERC funded project TITAN (EC-320691) and made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). AS and GH were

400   further funded by NERC under the Belmont forum, Grant PacMedy (NE/P006752/1), the NERC projects GloSAT (NE/S015698/1) and Vol-Clim (NE/S000887/1) and AS received funding from a Chancellors fellowship at the University of Edinburgh. DMS was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. HG is research director at the Fonds de la Recherche Scientifique (F.R.S.-FNRS).   MHE was supported by the Australian Research Council (SR200100008). We are grateful to the reviewers for insightful comments that improved the manuscript and its

405   presentation and to the editor for helpful direction

References

Abram, Nerilie J., Robert Mulvaney, Françoise Vimeux, Steven J. Phipps, John Turner, and Matthew H. England. 2014. "Evolution of the Southern Annular Mode during the Past Millennium." *Nature Climate Change* 4 (7): 564–69.

410   https://doi.org/10.1038/nclimate2235.

Allen, M. R., and P. a. Stott. 2003. "Estimating Signal Amplitudes in Optimal Fingerprinting, Part I: Theory." *Climate Dynamics* 21 (5–6): 477–91. https://doi.org/10.1007/s00382-003-0313-9.

Barnston, Anthony G., Michael K. Tippett, Meghana Ranganathan, and Michelle L. L'Heureux. 2019. "Deterministic Skill of ENSO Predictions from the North American Multimodel Ensemble." *Climate Dynamics* 53 (12): 7215–

415   34. https://doi.org/10.1007/S00382-017-3603-3.

Beadling, R. L., J. L. Russell, R. J. Stouffer, M. Mazloff, L. D. Talley, P. J. Goodman, J. B. Sallée, H. T. Hewitt, P. Hyder, and Amarjiit Pandde. 2020. "Representation of Southern Ocean Properties across Coupled Model Intercomparison Project Generations: CMIP3 to CMIP6." *Journal of Climate* 33 (15): 6555–81. https://doi.org/10.1175/JCLI-D-19-0970.1.

420   Brönnimann, Stefan, Jörg Franke, Samuel U. Nussbaumer, Heinz J. Zumbühl, Daniel Steiner, Mathias Trachsel, Gabriele C. Hegerl, et al. 2019. "Last Phase of the Little Ice Age Forced by Volcanic Eruptions." *Nature Geoscience* 12 (8): 650–56. https://doi.org/10.1038/s41561-019-0402-y.

Brunner, Lukas, Carol McSweeney, Andrew P. Ballinger, Daniel J. Befort, Marianna Benassi, Ben Booth, Erika Coppola, et al. 2020. "Comparing Methods to Constrain Future European Climate Projections Using a Consistent Framework." *Journal of Climate* 33 (20): 8671–92. https://doi.org/10.1175/JCLI-D-19-0953.1.

Cai, Wenju, Guojian Wang, Agus Santoso, Michael J. McPhaden, Lixin Wu, Fei-Fei Jin, Axel Timmermann, et al. 2015. "Increased Frequency of Extreme La Niña Events under Greenhouse Warming." *Nature Climate Change 2014 5:2* 5 (2): 132–37. https://doi.org/10.1038/nclimate2492.

Chan, Duo, Elizabeth C. Kent, David I. Berry, and Peter Huybers. 2019. "Correcting Datasets Leads to More Homogeneous Early-Twentieth-Century Sea Surface Warming." *Nature*. Nature Publishing Group. https://doi.org/10.1038/s41586-019-1349-2.

Chen, Dake, and Mark A. Cane. 2008. "El Niño Prediction and Predictability." *Journal of Computational Physics* 227 (7): 3625–40. https://doi.org/10.1016/J.JCP.2007.05.014.

Chylek, Petr, Chris Folland, James D. Klett, and Manvendra K. Dubey. 2020. "CMIP5 Climate Models Overestimate Cooling by Volcanic Aerosols." *Geophysical Research Letters* 47 (3). https://doi.org/10.1029/2020GL087047.

Collins, M., S. F.B. Tett, and C. Cooper. 2001. "The Internal Climate Variability of HadCM3, a Version of the Hadley Centre Coupled Model without Flux Adjustments." *Climate Dynamics 2001 17:1* 17 (1): 61–81. https://doi.org/10.1007/S003820000094.

Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, et al. 2011. "The Twentieth Century Reanalysis Project." *Quarterly Journal of the Royal Meteorological Society* 137 (654): 1–28. https://doi.org/10.1002/QJ.776.

Cowtan, Kevin, Robert Rohde, and Zeke Hausfather. 2018. "Evaluating Biases in Sea Surface Temperature Records Using Coastal Weather Stations." *Quarterly Journal of the Royal Meteorological Society* 144 (712): 670–81. https://doi.org/10.1002/qj.3235.

Dee, Sylvia G., Kim M. Cobb, Julien Emile-Geay, Toby R. Ault, R. Lawrence Edwards, Hai Cheng, and Christopher D. Charles. 2020. "No Consistent ENSO Response to Volcanic Forcing over the Last Millennium." *Science* 367 (6485): 1477–81. https://doi.org/10.1126/SCIENCE.AAX2000.

Delworth, Thomas L., and Fanrong Zeng. 2016. "The Impact of the North Atlantic Oscillation on Climate through Its Influence on the Atlantic Meridional Overturning Circulation." *Journal of Climate* 29 (3): 941–62. https://doi.org/10.1175/JCLI-D-15-0396.1.

Delworth, Thomas L., Fanrong Zeng, Gabriel A. Vecchi, Xiaosong Yang, Liping Zhang, and Rong Zhang. 2016. "The North Atlantic Oscillation as a Driver of Rapid Climate Change in the Northern Hemisphere." *Nature Geoscience* 9 (7): 509–12. https://doi.org/10.1038/ngeo2738.

Deser, Clara, Adam Phillips, Vincent Bourdette, and Haiyan Teng. 2012. "Uncertainty in Climate Change Projections: The Role of Internal Variability." *Climate Dynamics* 38 (3–4): 527–46. https://doi.org/10.1007/S00382-010-0977-X/FIGURES/17.

Deser, C., F. Lehner, K. B. Rodgers, T. Ault, T. L. Delworth, P. N. DiNezio, A. Fiore, et al. 2020. "Insights from Earth System Model Initial-Condition Large Ensembles and Future Prospects." *Nature Climate Change 2020 10:4* 10 (4): 277–86. https://doi.org/10.1038/s41558-020-0731-2.

Dubinkina, Svetlana, Hugues Goosse, Yoann Sallaz-Damaz, Elisabeth Crespin, and Michel Crucifix. 2011. "Testing a Particle Filter to Reconstruct Climate Changes over the Past Centuries." In *International Journal of Bifurcation and Chaos*, 21:3611–18. World Scientific Publishing Co. Pte Ltd. https://doi.org/10.1142/S0218127411030763.

Dunstone, Nick, Doug Smith, Steve Yeager, Gokhan Danabasoglu, Paul Arthur Monerie, Leon Hermanson, Rosie Eade, et al. 2020. "Skilful Interannual Climate Prediction from Two Large Initialised Model Ensembles." *Environmental Research Letters* 15 (9): 094083. https://doi.org/10.1088/1748-9326/AB9F7D.

Emile-Geay, Julien, Kimberly M. Cobb, Michael E. Mann, and Andrew T. Wittenberg. 2013. "Estimating Central Equatorial Pacific SST Variability over the Past Millennium. Part II: Reconstructions and Implications." *Journal of Climate* 26 (7): 2329–52. https://doi.org/10.1175/JCLI-D-11-00511.1.

Eyring, Veronika, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, Ronald J. Stouffer, and Karl E. Taylor. 2016. "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization." *Geoscientific Model Development* 9 (5): 1937–58. https://doi.org/10.5194/GMD-9-1937-2016.

Fogt, Ryan L., and Gareth J. Marshall. 2020. "The Southern Annular Mode: Variability, Trends, and Climate Impacts across the Southern Hemisphere." *Wiley Interdisciplinary Reviews: Climate Change* 11 (4): e652. https://doi.org/10.1002/WCC.652.

Franke, Jörg, Stefan Brönnimann, Jonas Bhend, and Yuri Brugnara. 2017. "A Monthly Global Paleo-Reanalysis of the Atmosphere from 1600 to 2005 for Studying Past Climatic Variations." *Scientific Data* 4 (June): 170076. https://doi.org/10.1038/sdata.2017.76.

Friedman, Andrew R., Gabriele C. Hegerl, Andrew P. Schurer, Shih-Yu Lee, Wenwen Kong, Wei Cheng, and John C. H. Chiang. 2020. "Forced and Unforced Decadal Behavior of the Interhemispheric SST Contrast during the Instrumental Period (1881–2012): Contextualizing the Late 1960s–Early 1970s Shift." *Journal of Climate* 33 (9): 3487–3509. https://doi.org/10.1175/JCLI-D-19-0102.1.

Gillett, Nathan P., Megan Kirchmeier-Young, Aurélien Ribes, Hideo Shiogama, Gabriele C. Hegerl, Reto Knutti, Guillaume Gastineau, et al. 2021. "Constraining Human Contributions to Observed Warming since the Pre-Industrial Period." *Nature Climate Change 2021 11:3* 11 (3): 207–12. https://doi.org/10.1038/s41558-020-00965-9.

Goddard, L., A. Kumar, A. Solomon, D. Smith, G. Boer, P. Gonzalez, V. Kharin, et al. 2013. "A Verification Framework for Interannual-to-Decadal Predictions Experiments." *Climate Dynamics* 40 (1–2): 245–72. https://doi.org/10.1007/S00382-012-1481-2/FIGURES/14.

Goosse, Hugues, Elisabeth Crespin, Svetlana Dubinkina, Marie France Loutre, Michael E. Mann, Hans Renssen, Yoann Sallaz-Damaz, and Drew Shindell. 2012. "The Role of Forcing and Internal Dynamics in Explaining the 'Medieval Climate Anomaly.'" *Climate Dynamics* 39 (12): 2847–66. https://doi.org/10.1007/s00382-012-1297-0.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood. 2000. "The Simulation of SST, Sea Ice Extents and Ocean Heat Transports in a Version of the Hadley Centre Coupled Model without Flux Adjustments." *Climate Dynamics* 16 (2–3): 147–68. https://doi.org/10.1007/s003820050010.

Hakim, Gregory J., Julien Emile-Geay, Eric J. Steig, David Noone, David M. Anderson, Robert Tardif, Nathan Steiger, and Walter A. Perkins. 2016. "The Last Millennium Climate Reanalysis Project: Framework and First Results." *Journal of Geophysical Research: Atmospheres* 121 (12): 6745–64. https://doi.org/10.1002/2016JD024751.

Hawkins, Ed, and Rowan Sutton. 2009. "The Potential to Narrow Uncertainty in Regional Climate Predictions." *Bulletin of the American Meteorological Society* 90 (8): 1095–1108. https://doi.org/10.1175/2009BAMS2607.1.

Hegerl, Gabriele C., Stefan Brönnimann, Andrew Schurer, and Tim Cowan. 2018. "The Early 20th Century Warming: Anomalies, Causes, and Consequences." *Wiley Interdisciplinary Reviews: Climate Change*. Wiley-Blackwell. https://doi.org/10.1002/wcc.522.

Hegerl, Gabriele C., Thomas J. Crowley, Steven K. Baum, Kwang Yul Kim, and William T. Hyde. 2003. "Detection of Volcanic, Solar and Greenhouse Gas Signals in Paleo-Reconstructions of Northern Hemispheric Temperature." *Geophysical Research Letters* 30 (5). https://doi.org/10.1029/2002GL016635.

Hegerl, G, and F Zwiers. 2011. "Use of Models in Detection and Attribution of Climate Change - Hegerl - 2011 - Wiley Interdisciplinary Reviews: Climate Change - Wiley Online Library." *… Interdisciplinary Reviews: Climate Change*.

Hermanson, Leon, Roberto Bilbao, Nick Dunstone, Martin Ménégoz, Pablo Ortega, Holger Pohlmann, Jon I. Robson, et al. 2020. "Robust Multiyear Climate Impacts of Volcanic Eruptions in Decadal Prediction Systems." *Journal of Geophysical Research: Atmospheres* 125 (9): e2019JD031739. https://doi.org/10.1029/2019JD031739.

Hurrell, James W., Yochanan Kushnir, Geir Ottersen, and Martin Visbeck. 2003. "An Overview of the North Atlantic Oscillation." *Geophysical Monograph Series* 134: 1–35. https://doi.org/10.1029/134GM01.

Hyder, Patrick, John M. Edwards, Richard P. Allan, Helene T. Hewitt, Thomas J. Bracegirdle, Jonathan M. Gregory, Richard A. Wood, et al. 2018. "Critical Southern Ocean Climate Model Biases Traced to Atmospheric Model Cloud Errors." *Nature Communications 2018 9:1* 9 (1): 1–17. https://doi.org/10.1038/s41467-018-05634-2.

Iles, Carley; Hegerl, Gabriele. 2017. "Role of the North Atlantic Oscillation in Decadal Temperature Trends." *Environmental Research Letters* 12. https://doi.org/10.1088/1748-9326/aa9152.

Kennedy, J J, N A Rayner, R O Smith, D E Parker, and M Saunby. 2011a. "Reassessing Biases and Other Uncertainties in Sea-Surface Temperature Observations Measured in Situ since 1850, Part 1: Measurement and Sampling Uncertainties." J. Geophys. Res., 116, D14103. http://www.bom.gov.au/jcomm/vos/.

———. 2011b. "Reassessing Biases and Other Uncertainties in Sea-Surface Temperature Observations Measured in Situ since 1850, Part 2: Biases and Homogenisation."J. Geophys. Res., 116, D14104

Kennedy, J. J., N. A. Rayner, C. P. Atkinson, and R. E. Killick, (2019). An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 data set. Journal of Geophysical Research: Atmospheres, 124, 7719– 7763.

Kettleborough, J. A., B. B. B. Booth, P. A. Stott, M. R. Allen, J. A. Kettleborough, B. B. B. Booth, P. A. Stott, and M. R. Allen. 2007. "Estimates of Uncertainty in Predictions of Global Mean Surface Temperature." *Journal of Climate* 20 (5): 843–55. https://doi.org/10.1175/JCLI4012.1.

Khodri, Myriam, Takeshi Izumo, Jérôme Vialard, Serge Janicot, Christophe Cassou, Matthieu Lengaigne, Juliette Mignot, et al. 2017. "Tropical Explosive Volcanic Eruptions Can Trigger El Niño by Cooling Tropical Africa." *Nature Communications 2017 8:1* 8 (1): 1–13. https://doi.org/10.1038/s41467-017-00755-6.

Klein, Stephen A., Brian J. Soden, Ngar-Cheung Lau, Stephen A. Klein, Brian J. Soden, and Ngar-Cheung Lau. 1999. "Remote Sea Surface Temperature Variations during ENSO: Evidence for a Tropical Atmospheric Bridge." *Http://Dx.Doi.Org/10.1175/1520-0442(1999)012<0917:RSSTVD>2.0.CO;2*, April. https://doi.org/10.1175/1520-0442(1999)012<0917:RSSTVD>2.0.CO;2.

Knight, Jeff R. 2005. "A Signature of Persistent Natural Thermohaline Circulation Cycles in Observed Climate." *Geophysical Research Letters* 32 (20): L20708. https://doi.org/10.1029/2005GL024233.

Knutti, Reto, David Masson, and Andrew Gettelman. 2013. "Climate Model Genealogy: Generation CMIP5 and How We Got There." *Geophysical Research Letters*. https://doi.org/10.1002/grl.50256.

Kosaka, Yu, and Shang-Ping Xie. 2013. "Recent Global-Warming Hiatus Tied to Equatorial Pacific Surface Cooling." *Nature* 501 (7467): 403–7. https://doi.org/10.1038/nature12534.

Leeuwen, Peter Jan van. 2009. "Particle Filtering in Geophysical Systems." *Monthly Weather Review* 137 (12): 4089–4114. https://doi.org/10.1175/2009MWR2835.1.

Lehner, Flavio, Clara Deser, Nicola Maher, Jochem Marotzke, Erich M. Fischer, Lukas Brunner, Reto Knutti, and Ed Hawkins. 2020. "Partitioning Climate Projection Uncertainty with Multiple Large Ensembles and CMIP5/6." *Earth System Dynamics* 11 (2): 491–508. https://doi.org/10.5194/ESD-11-491-2020.

Lehner, F., A.P. Schurer, G.C. Hegerl, C. Deser, and T.L. Frölicher. 2016. "The Importance of ENSO Phase during Volcanic Eruptions for Detection and Attribution." *Geophysical Research Letters* 43 (6): 2851–58. https://doi.org/10.1002/2016GL067935.

Li, Jinbao, Shang Ping Xie, Edward R. Cook, Mariano S. Morales, Duncan A. Christie, Nathaniel C. Johnson, Fahu Chen, et al. 2013. "El Niño Modulations over the Past Seven Centuries." *Nature Climate Change* 3 (9): 822–26. https://doi.org/10.1038/nclimate1936.

555 Luterbacher, Jürg, Daniel Dietrich, Elena Xoplaki, Martin Grosjean, and Heinz Wanner. 2004. "European Seasonal and Annual Temperature Variability, Trends, and Extremes since 1500." *Science (New York, N.Y.)* 303 (5663): 1499–1503. https://doi.org/10.1126/science.1093877.

Luterbacher, J., E. Xoplaki, D. Dietrich, P. D. Jones, T. D. Davies, D. Portis, J. F. Gonzalez-Rouco, et al. 2001. "Extending North Atlantic Oscillation Reconstructions Back to 1500." *Atmospheric Science Letters* 2 (1–4): 114–24. https://doi.org/10.1006/asle.2001.0044.

560 Maher, Nicola, Shayne McGregor, Matthew H. England, and Alexander sen Gupta. 2015. "Effects of Volcanism on Tropical Variability." *Geophysical Research Letters* 42 (14): 6024–33. https://doi.org/10.1002/2015GL064751.

Maher, Nicola, Sebastian Milinski, and Ralf Ludwig. 2021. "Large Ensemble Climate Model Simulations: Introduction, Overview, and Future Prospects for Utilising Multiple Types of Large Ensemble." *Earth System Dynamics* 12 (2): 401–18. https://doi.org/10.5194/ESD-12-401-2021.

565 Marshall, GJ. 2003. "Trends in the Southern Annular Mode from Observations and Reanalyses." *J. Clim.* 16: 4134–43. https://doi.org/10.1175/1520-0442(2003)016<4134:titsam>2.0.co.

McGregor, S., A. Timmermann, and O. Timm. 2010. "A Unified Proxy for ENSO and PDO Variability since 1650." *Climate of the Past* 6 (1): 1–17. https://doi.org/10.5194/CP-6-1-2010.

Morice, C. P., J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J.H. Dunn, T. J. Osborn, P. D. Jones, 570 and I. R. Simpson. 2021. "An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set." *Journal of Geophysical Research: Atmospheres* 126 (3): e2019JD032361. https://doi.org/10.1029/2019JD032361.

Ohba, Masamichi, Hideo Shiogama, Tokuta Yokohata, and Masahiro Watanabe. 2013. "Impact of Strong Tropical Volcanic Eruptions on ENSO Simulated in a Coupled GCM." *Journal of Climate* 26 (14): 5169–82. 575 https://doi.org/10.1175/JCLI-D-12-00471.1.

Olonscheck, Dirk, Andrew P. Schurer, Lucie Lücke, and Gabriele C. Hegerl. 2021. "Large-Scale Emergence of Regional Changes in Year-to-Year Temperature Variability by the End of the 21st Century." *Nature Communications 2021 12:1* 12 (1): 1–10. https://doi.org/10.1038/s41467-021-27515-x.

Pausata, Francesco S.R., Leon Chafik, Rodrigo Caballero, and David S. Battisti. 2015. "Impacts of High-Latitude 580 Volcanic Eruptions on ENSO and AMOC." *Proceedings of the National Academy of Sciences of the United States of America* 112 (45): 13784–88. https://doi.org/10.1073/PNAS.1509153112/-/DCSUPPLEMENTAL.

Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton. 2000. "The Impact of New Physical Parametrizations in the Hadley Centre Climate Model: HadAM3." *Climate Dynamics* 16 (2–3): 123–46. https://doi.org/10.1007/s003820050009.

585 Predybaylo, Evgeniya, Georgiy Stenchikov, Andrew T. Wittenberg, and Sergey Osipov. 2020. "El Niño/Southern Oscillation Response to Low-Latitude Volcanic Eruptions Depends on Ocean Pre-Conditions and Eruption

Timing." *Communications Earth & Environment 2020 1:1* 1 (1): 1–13. https://doi.org/10.1038/s43247-020-0013-y.

Ribes, Aurélien, Francis W. Zwiers, Jean-Marc Azaïs, and Philippe Naveau. 2017. "A New Statistical Approach to Climate Change Detection and Attribution." *Climate Dynamics* 48 (1–2): 367–86. https://doi.org/10.1007/s00382-016-3079-6.

Robock, Alan. 2000. "Volcanic Eruptions and Climate." *Reviews of Geophysics* 38 (2): 191–219. https://doi.org/10.1029/1998RG000054.

Sanderson, Benjamin M., Reto Knutti, and Peter Caldwell. 2015. "Addressing Interdependency in a Multimodel Ensemble by Interpolation of Model Properties." *Journal of Climate* 28 (13): 5150–70. https://doi.org/10.1175/JCLI-D-14-00361.1.

Schneider, Udo, Markus Ziese, Anja Meyer-Christoffer, Peter Finger, Elke Rustemeier, and Andreas Becker. 2016. "The New Portfolio of Global Precipitation Data Products of the Global Precipitation Climatology Centre Suitable to Assess and Quantify the Global Water Cycle and Resources." *Proceedings of the International Association of Hydrological Sciences* 374 (October): 29–34. https://doi.org/10.5194/PIAHS-374-29-2016.

Schurer, A.P., S.F.B. Tett, and G.C. Hegerl. 2014. "Small Influence of Solar Variability on Climate over the Past Millennium." *Nature Geoscience* 7 (2): 104–8. https://doi.org/10.1038/ngeo2040.

Slivinski, L. C., G. P. Compo, P. D. Sardeshmukh, J. S. Whitaker, C. McColl, R. J. Allan, P. Brohan, et al. 2021. "An Evaluation of the Performance of the Twentieth Century Reanalysis Version 3." *Journal of Climate* 34 (4): 1417–38. https://doi.org/10.1175/JCLI-D-20-0505.1.

Smith, D. M., A. A. Scaife, R. Eade, P. Athanasiadis, A. Bellucci, I. Bethke, R. Bilbao, et al. 2020. "North Atlantic Climate Far More Predictable than Models Imply." *Nature 2020 583:7818* 583 (7818): 796–800. https://doi.org/10.1038/s41586-020-2525-0.

Tett, Simon F. B., Jonathan M. Gregory, Nicolas Freychet, Coralia Cartis, Michael J. Mineter, and Lindon Roberts. 2022. "Does Model Calibration Reduce Uncertainty in Climate Projections?" *Journal of Climate* 35 (8): 2585–2602. https://doi.org/10.1175/JCLI-D-21-0434.1.

Tierney, Jessica E., Nerilie J. Abram, Kevin J. Anchukaitis, Michael N. Evans, Cyril Giry, K. Halimeda Kilbourne, Casey P. Saenger, Henry C. Wu, and Jens Zinke. 2015. "Tropical Sea Surface Temperatures for the Past Four Centuries Reconstructed from Coral Archives." *Paleoceanography* 30 (3): 226–52. https://doi.org/10.1002/2014PA002717.

Timmermann, Axel, Soon-Il An, Jong-Seong Kug, Fei-Fei Jin, Wenju Cai, Antonietta Capotondi, Kim M. Cobb, et al. 2018. "El Niño–Southern Oscillation Complexity." *Nature 2018 559:7715* 559 (7715): 535–45. https://doi.org/10.1038/s41586-018-0252-6.

Weisheimer, Antje, Damien Decremer, David MacLeod, Christopher O'Reilly, Tim N. Stockdale, Stephanie Johnson, and Tim N. Palmer. 2019. "How Confident Are Predictability Estimates of the Winter North Atlantic Oscillation?" *Quarterly Journal of the Royal Meteorological Society* 145 (S1): 140–59. https://doi.org/10.1002/qj.3446.

Zhang, Rong, Rowan Sutton, Gokhan Danabasoglu, Young Oh Kwon, Robert Marsh, Stephen G. Yeager, Daniel E. Amrhein, and Christopher M. Little. 2019. "A Review of the Role of the Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated Climate Impacts." *Reviews of Geophysics* 57 (2): 316–75. https://doi.org/10.1029/2019RG000644.