The effect of uncertainties in natural forcing records on simulated temperature during the last Millennium

Lucie J Lücke[1], Andrew P Schurer[1], Matthew Toohey[2], Lauren R Marshall[3, 4], and Gabriele C Hegerl[1]

[1]School of Geosciences, University of Edinburgh, Edinburgh, UK
[2]Institute of Space and Atmospheric Studies, University of Saskatchewan, Saskatoon, SK, Canada
[3] Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK
[4] Now at: Department of Earth Sciences, Durham University, Durham, UK

Correspondence: Lucie Lücke (lucie.luecke@ed.ac.uk)

Abstract. Here we investigate how uncertainties in the solar and volcanic forcing records of the past millennium affect the large-scale temperature response, using a 2-box impulse response model. We use different published solar forcing records and present a new volcanic forcing ensemble which accounts for random uncertainties in eruption dating and sulfur injection amount. The simulations are compared to proxy reconstructions from PAGES 2k and Northern hemispheric tree-ring data. We find that low solar forcing is most consistent with all the proxy reconstructions, even when accounting for volcanic uncertainty and that the residuals are well in line with CMIP6 control variability at centennial timescales. Volcanic forcing uncertainty induces a significant spread in the temperature response, especially at periods of peak forcing. For individual eruptions and superposed epoch analyses volcanic uncertainty can strongly affect the agreement with proxy reconstructions, and may explain previous proxy-model discrepancies.

1 Introduction

To quantify long-term climate variability we rely on simulations from climate models, driven by transient records of past radiative forcing, as well as on climate reconstructions from natural proxy archives. While both sources of information are imperfect in their own ways, they are independent, and thus may complement each other and serve for cross-validation. Proxy reconstructions are subject to a wide range of uncertainties and biases and data availability is sparse in many regions (Wilson et al., 2016; Anchukaitis et al., 2017; Neukom et al., 2018; Lücke et al., 2019, 2021). Models are very useful for detecting underlying mechanisms and drivers of variability, but are only approximations of the real world. Additionally, even a perfect model can only provide data as good as its input parameters: the records of external forcing agents, which are again subject to uncertainties.

Natural climate forcing includes orbital, solar and volcanic forcing. For the last Millennium, explosive volcanism is considered to be the main driver of forced surface temperature variability (Crowley and Lowery, 2000; Hegerl et al., 2003; Schurer et al., 2014; Neukom et al., 2019; Lücke et al., 2019). The influence of solar forcing on large-scale climate variability has been subject to debate, but has probably been minor compared to other forcings (Schurer et al., 2014). Orbital forcing is the only driver that can be precisely calculated for the last few millions of years (Berger, 1978; Berger and Loutre, 1991; Laskar et al., 2004). While it influences seasonal millennial temperature trends (Lücke et al., 2021), it plays only a minor role on shorter timescales.

Volcanic forcing records are most often compiled from sulfate concentrations in ice cores, complemented by documentary evidence to help determine unknown eruption characteristics, including eruption location and precise timing. However, both carry uncertainties and documentary evidence is not always available, especially for eruptions further back in time (Toohey and Sigl, 2017; Guevara-Murua et al., 2014) and thus various knowledge gaps still exist regarding past volcanic forcing. In some cases, large discrepancies are found between the simulated response to volcanic eruptions and the climate response in proxy reconstructions (e.g. Timmreck et al. (2009); Anchukaitis et al. (2012); Wilson et al. (2016); Guillet et al. (2017)), but also between the magnitude of ice core sulfate signals and proxy-based climate response estimates (Esper et al., 2017). Studies have shown that the forcing from eruptions can be overestimated if rapid adjustments are not considered where volcanic forcing is calculated from SAOD rather than forward modelled (Schmidt et al., 2018), and that the response to eruptions is strongly dependent on the eruption season and latitude (e.g. Marshall et al. (2019, 2020); Toohey et al. (2013, 2019)). The latest SAOD reconstruction for the past 2500 years (eVolv2k, Toohey and Sigl (2017)) benefits from improvements in methodologies used to date and synchronise ice core records, resulting in a more accurate estimate of volcanic forcing and resolving some discrepancies in the timing of volcanic events recorded by ice core sulfate records and by temperature sensitive proxy data (Sigl et al., 2015). Based on data availability and current state of knowledge, the most recent Paleoclimate Project Model Intercomparison Project (PMIP4, Jungclaus et al. (2017)) recommends the use of eVolv2k (Toohey and Sigl, 2017) for volcanic forcing in simulations of the 2000 years, replacing the two previous records (Crowley and Unterman, 2013; Gao et al., 2008) from PMIP3 (Schmidt et al., 2011, 2012). Despite these latest advances, substantial uncertainties remain in the reconstruction of volcanic forcing from ice core records regarding e.g. timing, magnitude, injection height and latitude of

eruptions (Stoffel et al., 2015; Schneider et al., 2017; Stevenson et al., 2017; Marshall et al., 2021). To address this, eVolv2k included an estimated uncertainty range in the volcanic sulfur injections for each eruption for the first time (Toohey and Sigl, 2017), providing information that can be used to explore the impact of uncertainties in volcanic forcing.

Solar forcing is primarily driven by photospheric magnetism, leading to varying numbers of sunspots and faculae concentrations on the solar surface, which modulate the total solar irradiance (TSI). Satellite records start in 1978 and have, in combination with various prior instrumental measurements of TSI, produced a well constrained understanding of short-term solar activity, which follows a quasiperiodic 11 year cycle (Fröhlich, 2006). Before the start of the instrumental period, sunspot numbers serve as a proxy for TSI. From AD 1610 onwards, records of sunspot numbers from telescope observations are available. However, prior to the telescopic era, the reconstruction of solar variation is based mainly on cosmogenic isotopes deposited in polar ice cores and tree-rings, of which solar activity can be estimated by applying a chain of physics-based models. Therefore, long-term changes in radiative forcing remain uncertain, and its exact timeline and especially its amplitude is still controversial (Gray et al., 2010). While most studies suggest a small amplitude on centennial timescales (e.g. Vieira et al. (2011); Steinhilber et al. (2009); Wang et al. (2005); Muscheler et al. (2007)), Shapiro et al. (2011) reconstructed the maximum long-term amplitude based on physically possible mechanisms, exceeding other reconstructions by more than 1 Wm^{-2} during the Spörer Minimum. A number of studies have investigated this high versus low amplitude controversy, and indicated that high solar forcing is inconsistent with instrumental data (Feulner, 2011; Judge et al., 2012; Lockwood and Ball, 2020) and proxy temperature reconstructions (Ammann et al., 2007; Hind et al., 2012; Schurer et al., 2014). The uncertainty around the long-term solar cycle is reflected by the inclusion of several solar forcing records in PMIP, with three different reconstructions included in PMIP4 (Jungclaus et al., 2017), an update from the previous seven reconstructions of PMIP3 (Schmidt et al., 2011, 2012).

Given the high degree of uncertainty in the volcanic and solar forcing record, it would be desirable to run a variety of comprehensive climate model simulations using different realisations of transient forcing. However, running such experiments over a whole Millennium requires large computational resources. Additionally, internal variability tends to disguise the effects of forcing, and therefore a large ensemble of runs would be needed for the differences in the forced signal to emerge for each implementation of forcing. Given the computational expense needed, such experiments are not performed by the scientific community.

A computationally very efficient representation of the Earth system is given by simple energy balance models, which simulate the climate system's response to changing radiative forcing (Crowley, 2000; Hegerl et al., 2006; Held et al., 2010; Geoffroy et al., 2013b, a; Millar et al., 2017). Held et al. (2010) introduced an energy balance model with two time scales, in which a fast timescale is associated with the ocean mixed layer, while the slow timescale is associated with the deep ocean response. Such a model can perform well to simulate the large-scale forced response to greenhouse gas and other forcings on hemispheric and global scale, for land and ocean (Haustein et al., 2019).

Here, we use a two-box impulse response model, tuned to climate simulations using volcanic forcing only, to simulate the response to solar and volcanic forcing to study the consequences of uncertainty in the natural forcing history. We perform experiments using a new ensemble of volcanic forcing, as well as different solar forcing reconstructions as input for the response

model. We present the current forcing records, temperature reconstructions and a detailed explanation of the experimental setup in Sect. 2. We compare the response model output to proxy reconstructions of Northern Hemisphere and global temperature, focusing on the analysis of solar and volcanic forcing uncertainty separately in Sect. 3. The residuals between proxy reconstructions and response model simulations reflect a combination of forcing error, structural error in the response model, tuning error, reconstruction uncertainty and internal variability. We put these residuals into the context of CMIP6 pre-industrial control runs and observed unforced variability. Where the residual is similar in magnitude to internal variability, we conclude that the particular forcing estimate used is consistent with the proxy reconstruction, while the spread of the residuals with varying forcing reflects the importance of forcing error to explain discrepancies. Based on this we estimate the role of forcing uncertainty and evaluate if particular forcings are more realistic than others. We discuss our findings and summarise our conclusions in Sect. 4.

2 Data and Methods

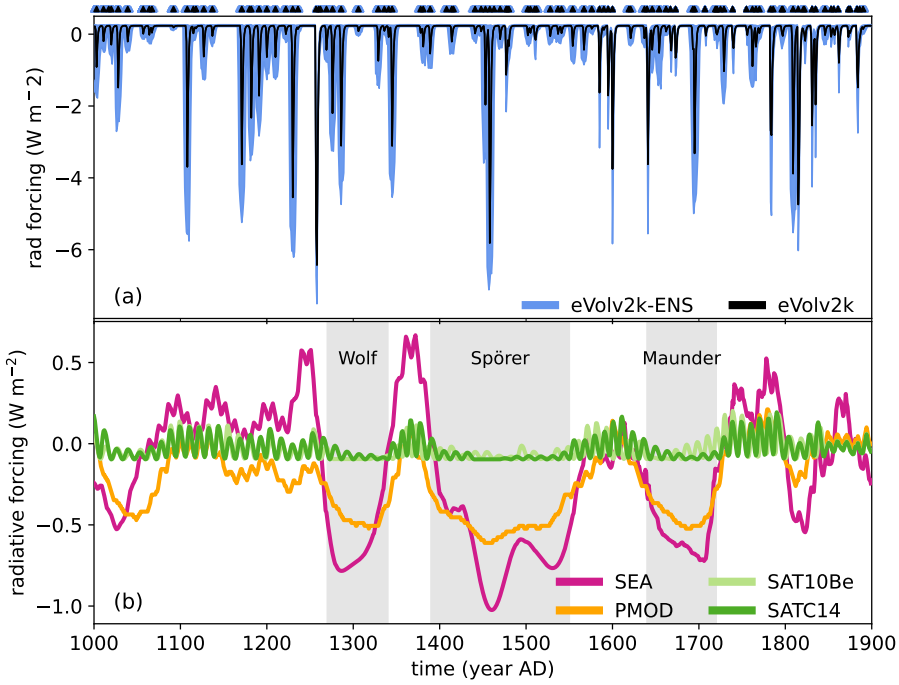

Figure 1. Timeseries of natural forcing records, as anomalies taken over the whole period. (a) global annual mean of the volcanic forcing ensemble (eVolv2k-ENS) and of eVolv2k. Triangles on top mark the associated eruption years. (b) solar forcing reconstructions. The three strongest solar minima during the pre-industrial millennium are shaded in grey.

2.1 Volcanic forcing

The eVolv2k reconstruction (Toohey and Sigl, 2017) provides a time series of volcanic stratospheric sulfur injection (VSSI) and latitudinally resolved stratospheric aerosol optical depth (SAOD). This reconstruction uses ice core records to estimate the timing and amount of VSSI, and the EVA forcing generator (Toohey et al., 2016) to reconstruct the spatiotemporal distribution of SAOD given the estimated eruption latitude, date and VSSI. Estimates of VSSI and SAOD from ice core sulfate measurements have a significant associated uncertainty (Hegerl et al., 2006). The eVolv2k reconstruction provides estimates of the 1-sigma random uncertainty in the VSSI estimates (Toohey and Sigl, 2017). These uncertainties represent uncertainties due to two sources: 1) the uncertainty in the ice sheet (Greenland or Antarctica) mean sulfate flux, based on the variation between the individual ice core records that make up the ice sheet composite, and 2) a representativeness error, corresponding to the error inherent in taking the ice sheet mean as a measure of the full hemispheric sulfate deposition. The latter source of uncertainty is estimated based on the variability in ice sheet average deposition for eruptions of fixed magnitude in aerosol model simulations (Toohey et al., 2013). A total random error is computed by combining the measurement and representativeness error. Average

estimated percent uncertainties in eVolv2k are 34%, with values for eruptions of greater than 10 TgS ranging from 14-40%. There is also uncertainty in the timing of eruptions for all but the relatively few ice core sulfate peaks that can be attributed to historically recorded eruptions. For unidentified eruptions, the year listed in the eVolv2k database represents the year in which excess sulfate is first detected in the ice cores, with an estimated uncertainty of ± 2 years. This uncertainty includes the overall dating uncertainty of the ice core and as well as uncertainty in the time lag between eruption and first deposition of sulfate flux to the ice sheets. To incorporate uncertainties in VSSI amount and eruption timing into the present work, we constructed an ensemble of VSSI time series from the default eVolv2k data. For each member of constructed eVolv2k forcing ensemble (eVolv2k-ENS) the following steps were taken:

- For each eruption individually, we perturbed the VSSI amount by a normally distributed random variable of mean zero and standard deviation of the reported VSSI uncertainty for that eruption. If the resulting value was negative, we set the value to zero.

- For unidentified eruptions, we perturbed the eruption year by a normally distributed random variable with mean zero and standard deviation 2, rounded to the nearest integer. Eruption months were furthermore randomised based on a uniform distribution centered on the default January eruption month used in eVolv2k.

This procedure was iterated to produce 1000 different timeseries of VSSI, each a possible version of past volcanic activity given the estimated values and uncertainties listed in eVolv2k. For each individual eruption, the eVolv2k-ENS members produce a distribution of potential VSSI amount and timing. The original default eVolv2k values are found at the peak of the distribution, representing the estimated most probable value for each individual eruption. The eVolv2k-ENS VSSI timeseries are then each passed through the EVA v1 forcing generator to create a time series of SAOD. Due to non-linearity in the VSSI to SAOD conversion, the default eVolv2k SAOD is not at the center of the eVolv2k-ENS SAOD distribution, but it occurs at the peak in the VSSI distribution (Fig. S7). The volcanic forcing ensemble therefore represents an estimate of the range of possible volcanic forcing histories given the VSSI values and random uncertainties reported in the eVolv2k data set. It does not however include all possible sources of uncertainty. For example, it does not include potential systematic biases arising from 1) uncertainty in the scaling relations used to covert ice core sulfate into VSSI, or 2) model uncertainty from EVA, including its lack of sensitivity to the height of the sulfur injection (Aubry et al., 2020). Nor does it include potential uncertainty in the attribution of unidentified eruptions to different latitude bands, nor any potential misattribution of sulfate signals to historical eruptions.

To convert the global SAOD timeseries to radiative forcing, we use the average relationship presented for global mean values in Marshall et al. (2020), using exponential scaling. The conversion of NH SAOD to NH radiative forcing is more uncertain and therefore, for simplicity, we use the same relationship, in this way treating all of the SAOD ensembles the same. However it should be noted that this presents a substantial source of uncertainty with respect to the impulse response model output, especially for NH results.

2.2 Solar forcing

Reconstructions of solar irradiance are based on cosmogenic isotope records of ^{14}C (Roth and Joos, 2013; Usoskin et al., 2016) and ^{10}Be (Baroni et al., 2015). ^{14}C is oxidised to CO_2 in the atmosphere and metabolised by trees to form tree-rings (e.g. Suess (1980); Stuiver and Quay (1980)), while ^{10}Be attaches to atmospheric aerosols, which are removed from the atmosphere in different ways and subsequently deposited to ice sheets (e.g. Beer et al. (1983); Usoskin et al. (2009)). They represent independent proxy records (Bard et al., 1997; Steinhilber et al., 2012) and are mostly consistent – especially on long timescales – but deviations during certain periods have been noted due to the different formation rates (Usoskin et al., 2009; Steinhilber et al., 2012).

We include four different reconstructions of solar forcing in our analysis, resulting from either of these isotope records or combined versions, including the PMIP4 records PMOD, SAT10Be and SATC14 (Jungclaus et al., 2017). The latter two are derived from the updated SATIRE-M model (Vieira et al., 2011; Wu, 2017), with SAT10Be using ^{10}Be isotope records as input, and SATC14 using ^{14}C isotope records. PMOD uses only ^{14}C as input (Jungclaus et al., 2017), and represents an update from the older SEA model (Shapiro et al., 2011). PMIP3 includes seven different records (Schmidt et al., 2011, 2012), including, among others, SBF (Steinhilber et al., 2009), which was used as the solar forcing record for runnign HadCM3, and SEA (Shapiro et al., 2011), which represents the maximum amplitude solar forcing record. Apart from SEA and PMOD, all records represent small long-term solar forcing and share a similar amplitude. Due to the similarities in the low solar forcing records, we restrict our analysis to the PMIP4 records, and, for completeness, include SEA, representative of the maximal (although unlikely) physically possible forcing (Fig. 1 (a)). We thus sample two independent isotope records (^{10}Be and ^{14}C), combined with two independent reconstruction methods (SATIRE and SEA/PMOD).

The ensemble variance is dominated by the difference in the long-term amplitude (Fig. 1 (b)) and is especially high during the Spörer minimum (around $1390-1550$), the Wolf minimum (around $1270-1340$) and the Maunder minimum (around $1640-1720$) (Usoskin et al., 2007). Short-term changes are less uncertain and are caused by quasi-periodic oscillations ranging between 9 and 14 years (Nandy et al., 2021), mainly as a result of discrepancies between the different isotope records (Usoskin et al., 2009; Steinhilber et al., 2012). The variance is largest during the mid-15th century, coinciding with the large volcanic eruption in 1458 (Fig. 1 (c)). Baroni et al. (2019) hypothesised that large stratospheric volcanic eruptions may deplete the stratospheric ^{10}Be isotope for around a decade, biasing the isotope records. This could partially explain the large amplitude of solar minima in the SEA record, a record based on ^{10}Be only, which all coincide with periods of major volcanic activity.

2.3 Impulse response model

2.3.1 Model setup and tuning

The response model accounts for fast and slow temperature changes in response to external forcing (Otto et al., 2015; Haustein et al., 2017; Millar et al., 2017), where fast and slow components are associated with the response of the ocean mixed layer and the deep ocean response respectively (Li and Jarvis, 2009). In addition to these two timescale parameters, which need to be determined through model tuning, the equilibrium climate sensitivity (ECS) and the transient climate response (TCR) have to be

set. In order to ultimately compare the output with proxy reconstructions, which typically represent northern hemisphere (NH) summer (May-June-July-August; MJJA) land surface temperature, we tune the response model separately to global annual temperature, and to NH summer land temperature from HadCM3 simulations (Pope et al., 2000; Gordon et al., 2000). We run the response model with TCR=2.0 and ECS=3.3 (Tett et al., 2022) and use ICI5 (Crowley and Unterman, 2013) for volcanic forcing and SBF (Steinhilber et al., 2009) for solar forcing, in agreement with the HadCM3 input. Pre-industrial greenhouse gas concentrations (Meinshausen et al., 2017) were converted into radiative forcing (Etminan et al., 2016) and merged with historical output from FAIR v1.3 (Smith et al., 2018). Aerosol forcing from the Community Emissions Data System (CEDS, Hoesly et al. (2018)) is included from 1820 onwards. We tune the response model by minimising the residuals between the response model output and the all forcings HadCM3 ensemble mean over the period 1400 to 1850. We use the HadCM3 all forced simulations as the target for tuning to ensure an optimal choice of the fast and the slow response time. We test the tuned response model by comparing its forced response separately to the all forcing HadCM3 ensemble mean as well as to individual forcings (Fig. S2 and S3). Note that the response to solar forcing in the HadCM3 solar only simulation is largely masked by residual internal variability in the ensemble mean, which decreases the correlation between HadCM3 and response model. For all, GHG and volcanic forcing we confirm that the best fit corresponds well to the associated HadCM3 ensemble mean.

2.3.2 Forcing uncertainty ensemble

After determining the free model parameters, we run the response model with our forcing uncertainty ensemble as input. However, to ensure an optimal comparison between response model and proxy reconstructions, we first change the climate sensitivity to a more generic setup. A large range of climate sensitivity is found across the CMIP6 models (Fig. S4), reflecting the large uncertainty associated with this parameter (Sherwood et al., 2020). We therefore run the response model with the median climate sensitivity across CMIP6 (ECS=3.3, TCR=2.0) for the main analysis, and conduct additional sensitivity studies with the lower and the upper range of climate sensitivity of CMIP6 models (shown in the Supplementary Information). The lowest climate sensitivity is associated with INM-CM4-8 (ECS=1.84, TCR=1.32) and the highest climate sensitivity is associated with E3SM-1-0 (ECS=5.38, TCR=2.99). This way we can estimate a lower and an upper limit of the temperature response to differences in the natural forcing record.

In order to run the impulse response model with eVolv2k-ENS, we convert the forcing timeseries from SAOD to ERF. For producing global annual data, we use the globally averaged timeseries of SAOD, while for NH summer, we restrict our data to NH SAOD only. For global data, yearly averages are created by using the average from April to March of the following year, following Neukom et al. (2019). For NH summer data, we use a similar strategy, e.g. for a given year we average September of the preceding year to August of the current year. This way, we ensure that a volcanic eruption after the growing season has finished is not taken into account for a given year. The resulting timeseries are converted into ERF using the non-linear scaling (Marshall et al., 2020).

The 1000 timeseries of volcanic radiative forcing, obtained by eVolv2k-ENS are joined with the four solar forcing timeseries, and thus the response model is run 4000 times for each global annual and NH summer land temperature. Greenhouse gas and

aerosol forcing were kept from the previous setup. Additionally, we run the response model with the PMIP4 reference forcing, eVolv2k, and each solar forcing reconstruction.

A summary of the complete process, from model tuning to running the forcing uncertainty ensemble, is given in Fig. S1.

2.4 Evaluation of the forced response

2.4.1 Residual variability of proxy reconstructions

To evaluate the consistency between the output of the impulse response model x, the temperature response to radiative forcing, with the proxy reconstructions y, temperature variability due to forcing and internal climate processes, we calculate the residuals, using the root mean square error (RMSE). We calculate the RMSE over centered,running windows t with length N, using

$$\mathrm{RMSE}_t = \sqrt{\frac{1}{N}\sum_{i=-N/2}^{N/2}((x_{i+t}-\hat{x}_t)-(y_{i+t}-\hat{y}_t))^2}, \qquad t = \frac{N}{2}, \frac{N}{2}+1, \frac{N}{2}+2, ... \tag{1}$$

The mean over the window t is denoted by $\hat{x}_t$ for the response model, and $\hat{y}_t$ for the proxy data. We choose a length $N = 20$ years to analyse the temperature response to volcanic forcing, and a length $N = 200$ years to analyse the response to solar forcing. The former encompasses the recovery time even after a large volcanic eruption such as Tambora generously (Schurer et al., 2013; Otto-Bliesner et al., 2016; Raible et al., 2016). The latter encompasses a full solar cycle, which has an amplitude of around 200 years (Fig. 1 (b)). The residual represents differences between model output and proxies due to internal variability present in the proxy reconstructions, but also includes the discrepancy between radiative forcing in the proxies and the model, proxy reconstruction error and model simulation error.

We can roughly estimate the influence of all of these factors by

- varying the radiative forcing in the response model using an ensemble of different realisations of forcing to estimate forcing discrepancy
- comparing the residuals to control runs from climate models and to estimates of unforced variability in observational data to estimate internal variability
- using an ensemble of different proxy reconstructions to estimate reconstruction error
- comparing the response model simulations to coupled climate model simulations to estimate simulation error.

2.4.2 Proxy reconstructions

We use four tree-ring reconstructions of Northern hemispheric (NH) summer temperature (N-TREND16: Wilson et al. (2016), N-TREND17: Anchukaitis et al. (2017), GCS17: Guillet et al. (2017), SSB15: Schneider et al. (2015) and the multi-proxy reconstruction PAGES 2k (Neukom et al., 2019) for global annual temperature. The N-TREND reconstructions are both compiled

from the northern hemisphere Tree-Ring Network Development dataset (Wilson et al., 2016; Anchukaitis et al., 2017), but using different reconstruction methods. N-TREND16 is a hemispheric land surface temperature reconstruction based on iterative nesting plus scaling to instrumental data. N-TREND17 is a climate field reconstruction using point-by-point regression (Cook et al. (1999)). N-TREND includes measurements of tree-ring width (RW), maximum latewood density (MXD) and blue intensity (BI), and aims to reconstruct midlatitudinal (40–75° N) MJJA temperature over land. SSB15 uses weighted composites based on moving correlations with local temperature and scaling to reconstruct extratropical (30–90° N) June to August (JJA) temperature. They include solely MXD tree-ring chronologies, which are considered to be more sensitive to high frequency variability (Franke et al., 2013; Esper et al., 2015; Lücke et al., 2019). GCS17 used tree-ring data (RW and MXD) to reconstruct extratropical (40–90° N) JJA temperature over land with a bootstrap linear model using principal component analysis. The four reconstructions share several chronologies, however all of them use different methodologies for reconstruction.

PAGES 2k uses a large multi-proxy dataset (PAGES2k Consortium, 2017) and combines seven different reconstruction methods (CPS: Neukom et al. (2014); Mann et al. (2008), PCR: Neukom et al. (2014); Luterbacher et al. (2002), OIE: Shi et al. (2017), M08: Mann et al. (2008), PAI: Hanhijärvi et al. (2013), BHM: Barboza et al. (2014), LMR: Hakim et al. (2016)) to obtain global mean surface temperature (GMST) of the last two Millennia between April and March. For each method Neukom et al. (2019) obtained an ensemble of 1000 reconstructions using various statistical approaches to estimate the reconstruction uncertainty such as bootstrapping or red noise perturbations. For computational reasons we subsample 50 random ensemble members of each method, which we find gives a reasonable estimate of the ensemble spread (compare Fig. S5).

2.4.3 Time-series filtering

We apply a second order Butterworth filter (Mann, 2004, 2008; Neukom et al., 2019) on all temperature timeseries, which we use in two ways: to isolate the response to a specific forcing and to suppress spectral biases/frequency insensitivity of the proxy data. To isolate the response to long-term changes in solar variability we use a bandpass filter between 50 and 300 years. PAGES 2k and N-TREND are both sensitive to temperature variability on this frequency band (Neukom et al., 2019; Lücke et al., 2019). For the volcanic response we focus on annual to multi-decadal frequencies. However, PAGES 2k was shown to be relatively insensitive to high frequency variability due to the large inclusion of low frequency records (Neukom et al., 2019), and thus we set the high frequency threshold at 15 years for global analyses. Tree-ring reconstructions perform well at high frequencies, but are subject to autoregressive biases which affect their high frequency variability (Lücke et al., 2019). We therefore set the high frequency threshold at 5 years. The low frequency threshold remains set at 300 years.

2.4.4 Variability of CMIP6 control runs

We use temperature data from 58 pre-industrial control runs (Eyring et al. (2016), table S1) to estimate unforced temperature variability on centennial timescales and on multi-decadal timescales. Temperature was averaged over both midlatitudinal land summer (MJJA) and the global annual mean. The timeseries were bandpass filtered, and the RMSE was calculated from non-overlapping chunks of $N = 20$ and $N = 170$ years. For $N = 20$ years all individual models are consistent within the lower and upper quartile of the population of all simulations, showing that most of the models roughly agree on the extent of decadal

variability (Fig. S9 (a), (c)). For $N = 170$ years there is less agreement between the individual models, with some models displaying higher variability than others (Fig. S9 (b), (d)). These models are also associated with higher 100 year running GMST trends (Parsons et al., 2020) and we therefore refer to them as high variability types.

2.4.5 Residual variability of observations

We use the infilled HadCRUT5 dataset (Morice et al., 2021), a gridded dataset of global historical surface temperature anomalies from January 1850 to December 2020. The infilled dataset has increased coverage and includes 200 realisations sampling the distribution of various errors and uncertainties. To obtain an estimate of unforced variability, we conduct a total least squares (TLS) regression, using the historical CMIP6 multi-model mean as the fingerprint of forcing (Allen and Stott, 2003; Schurer et al., 2014). The residuals give an estimate of the unforced variability (Schurer et al., 2015; Friedman et al., 2020) and can be compared to the residuals of proxy and model simulations. This procedure was repeated for each of the HadCRUT5 ensemble members and for both global annual and NH summer temperature.

3 Results

3.1 Model performance

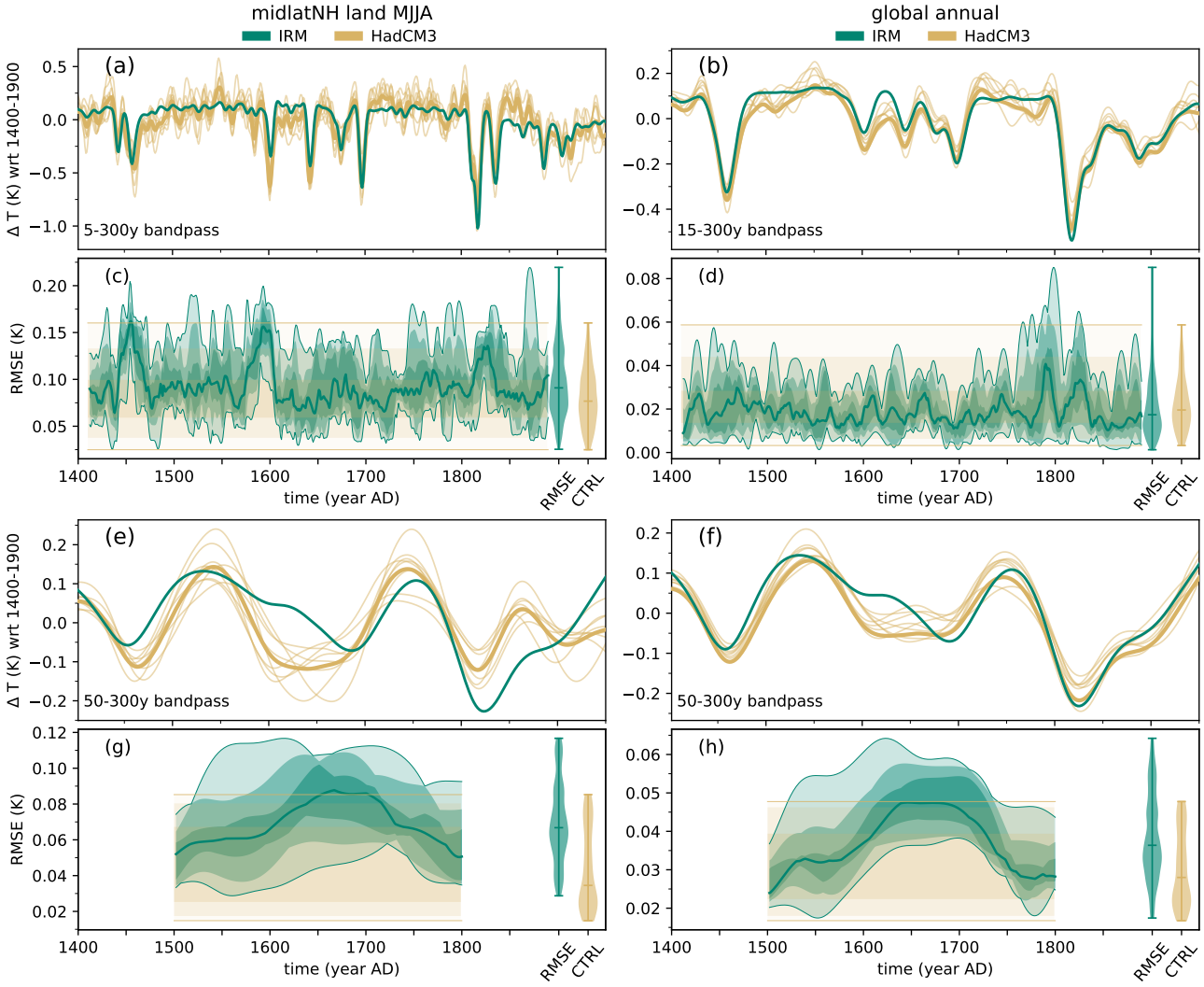

Figure 2. Evaluating the performance of the response model, tuned to HadCM3 and using similar forcing, against simulations from HadCM3. Left: midlatitudinal NH land MJJA temperature. Right: global annual temperature anomalies. (a), (b) Bandpass filtered timeseries of the past 600 years. Thick orange line represents the HadCM3 ensemble mean, thin lines the ensemble members. (c), (d) RMSE between IRM and HadCM3 versus HadCM3 control (CTRL) variability calculated over 20 year running windows. Thick lines represents the median, shading the 75th/90th/100th percentile. The violinplots show the distribution of all slices. (e)–(h) as previous but filtered using a 50-300 year bandpass and residual variability calculated over 200 year windows.

We assess the model performance by comparing the residuals between the response model, tuned to HadCM3 and using HadCM3 forcings, and transient HadCM3 simulations to the standard deviation of HadCM3 control runs over 20 and 200 years sliding windows (Fig. 2). For a perfect response model, which accurately reproduces the climate response to forcing in HadCM3, the residuals would represent the internal variability simulated by HadCM3, and thus match the spread of control variability. However, given that the response model is a simplistic simulation of large-scale climate and does not account for hemispheric transport, seasonal processes or non-linear mechanisms, we expect the residuals to also include model error, and thus to exceed the control variability. Remaining internal variability in the HadCM3 ensemble mean due to the relatively small ensemble size could additionally increase the residuals.

At 20 year timescales the residuals agree at most times well with the spread of control variability, especially for global simulations. However, in the NH, residuals are very large during the mid-15th century as well as around 1800, which are both periods of high volcanic forcing. The larger residuals around 1800 are also present in the global simulation. Both periods are subject to very large volcanic forcing, and thus it is likely that the response model may be too simplistic to accurately model the associated climate processes as simulated by HadCM3. We find a larger discrepancy for both global and NH simulations at the 200 years timescale, although the 90th percentile is within the range of control at most times. The residual clearly exceeds the control variability during the 17th century, which could be explained by a possible non-linearity in the solar and volcanic forcing during that period (Schurer et al., 2014). As observed at the 20 y timescales, residuals are smaller for global simulations, indicating that model error is larger in the NH. This finding is unsurprising, given that regional and seasonal climate processes play a larger role for NH summer temperature as opposed to global annual temperature.

3.2 Solar forcing uncertainty

We compare the proxy reconstructions with response model simulations including the different solar forcing reconstructions. On centennial timescales the difference between simulations including SAT10Be and SATC14 is almost negligible and both follow the proxy reconstructions closely after around AD 1300 (Fig. 3 (a), (b)). In contrast, the temperature response to PMOD and SEA forcing show very large variability, with particularly strong anomalies during the Wolf (around AD 1300) and the Spörer (15th century) minima. Since the availability of proxy records decreases quickly before AD 1300 and therefore reconstructions are much more uncertain (Wilson et al., 2016; Neukom et al., 2019), we use the period from 1300 to 1900 to calculate residuals between proxy reconstructions and response model simulations and compare to estimated observed and modelled internal variability. The residuals between the proxy reconstructions and the full response model ensemble show that even when accounting for volcanic and proxy reconstruction uncertainty the SATIRE forcings consistently generate the lowest residuals, and SEA consistently generates the highest (Fig. 3 (c), (d)), with very little overlap between them. For the northern hemisphere, the former shares a very similar distribution with the root mean square deviation of the CMIP6 control variability, although the distribution shows two peaks, generated by the different proxy reconstructions. Given that the model error ranges around 0.03 K for NH summer simulations and 0.01 K for global annual simulations (Fig. 2), we expect the response model to produce slightly higher residuals than the control variability. This suggests that SATIRE forcing is consistent with the

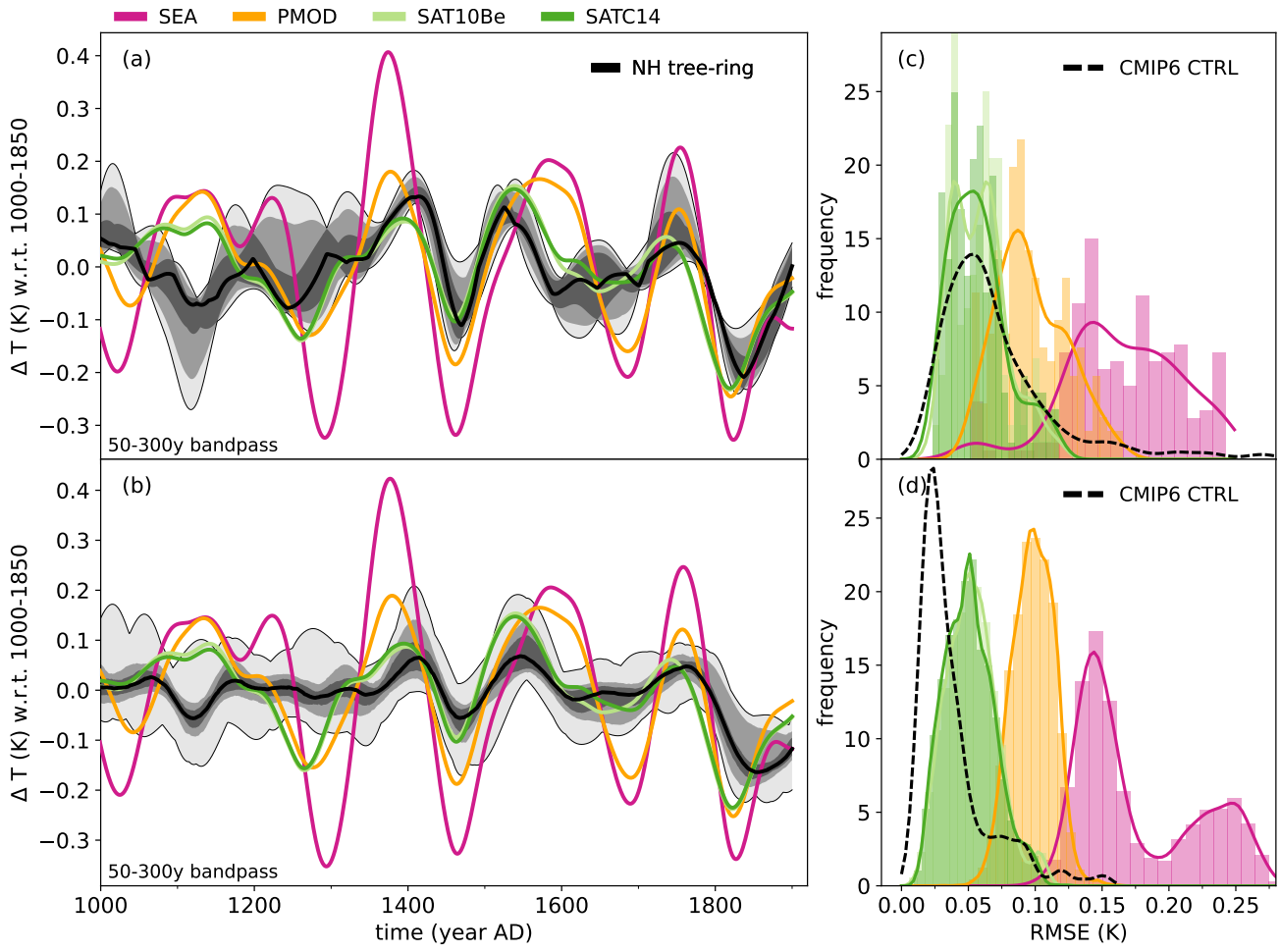

Figure 3. Left: Temperature anomaly of proxy reconstructions and response model ensemble for the different solar forcing records for (a) NH summer land and (b) global mean, filtered by a 50–300 years bandpass. The shading indicates volcanic forcing uncertainty for the response model and reconstruction uncertainty for the proxy data (shading levels at 75th, 90th and 100th percentile, thick line median, thin lines range). Right: Distribution of residuals between response model and proxy reconstructions, calculated over 200 years sliding windows over the period 1300 (dotted line in (a) and (b)) to 1900, for (c) NH summer land and (d) global annual temperature for different solar forcing estimates (colours). The dashed line indicates the variability of CMIP6 control runs.

reconstructions compared to the distribution of control variability. In contrast, both PMOD and SATIRE are inconsistent with the majority of CMIP6 control runs.

At global scale, the CMIP6 control runs show a significantly lower variability than the residuals from all solar forcings. This is roughly consistent with the difference between HadCM3 and the response model run with HadCM3 forcing (Fig. 2) and thus could be caused by model error, or could indicate that global annual variability is slightly underestimated by CMIP6 models.

PMOD is only consistent with the upper tail of the distribution, caused by the high variability type models, and residuals from SEA are pretty much inconsistent with CMIP6 control variability.

It is worth noting that the relative magnitude of residuals associated with the different solar forcings is independent of the temperature target and the reconstruction (Fig. S11 and S12). In addition, it is also independent of response model parameters such as climate sensitivity. While for the low climate sensitivity the differences between the temperature response to SATIRE, PMOD and SEA forcing decrease, SEA still produces consistently larger residuals than SATIRE forcing (Fig. S14). For the high climate sensitivity the differences increase (Fig. S13), but the order remains unchanged: SATIRE produces the smallest residuals, followed by PMOD, and SEA forcing.

3.3 Volcanic forcing uncertainty

To quantify the effect of volcanic forcing uncertainty, we fix the solar forcing to SATC14, the default PMIP4 forcing (Jungclaus et al., 2017). The volcanic ensemble generates a large variety of amplitudes in response to individual eruptions. At times, the spread between the ensemble members exceeds 1 K in the NH, indicating that large uncertainty of the volcanic forcing record progresses into the temperature response (Fig. 4 (a)). The largest spread is found during the 1450s, and after the eruption of Laki (1783). Accordingly, the magnitude of the residuals between response model simulations and proxy reconstructions also vary significantly throughout the Millennium (Fig. 4 (b)). In the NH, the 75th percentile of the residuals obtained from eVolv2k-ENS agrees with the spread of residuals obtained from eVolv2k, which represent the proxy reconstruction uncertainty, at most times. The proxy reconstruction uncertainty tends to be larger during years of low volcanic activity, and decreases around the eruptions, except for the eruption of Samalas in 1257. The residuals of eVolv2k-ENS are largest, and show the largest spread during periods of high volcanic activity. The residuals peak around the eruption of Samalas for both eVolv2k and eVolv2k-ENS, which also features the largest spread of results. This reflects the large uncertainty around this eruption with regard to both forcing and reconstruction. The mismatch between the model response to the eruption of Samalas and the response shown by proxy reconstructions is a well-known discrepancy. It reflects a mismatch between the signals in ice-cores, the source data of the volcanic forcing in model simulations, and tree-rings, the source data of the proxy reconstructions (Guillet et al., 2017; Timmreck et al., 2009). Additionally, the response model and the proxy reconstructions match better going forwards in time, with a particularly poor agreement prior to ca. A.D. 1300. This reflects the decreasing data availability and quality as we go back in time. Prior to A.D. 1300 only few high resolution records from maximum latewood density are available, which reduces the confidence in the proxy reconstructions for the first three centuries (Wilson et al., 2016) and hence in the proxy response to the eruption of Samalas. Nevertheless, given the large ensemble spread between residuals from the eVolv2k-ENS members (around 0.2 K) around this eruption, volcanic forcing uncertainty could additionally explain a part of this discrepancy. A similar observation holds for the mid 1450s eruptions (Esper et al., 2017; Sigl et al., 2015), however in this case the spread of reconstructions is smaller.

While volcanic forcing uncertainty provides a possible contribution to the discrepancy of these specific eruptions, it is unlikely that it can be fully explained by it- at least not with such a simple model. Proxy biases, reconstruction uncertainty and model errors may provide further explanations. Apart from Samalas and the mid-1450s eruptions, the lower range of

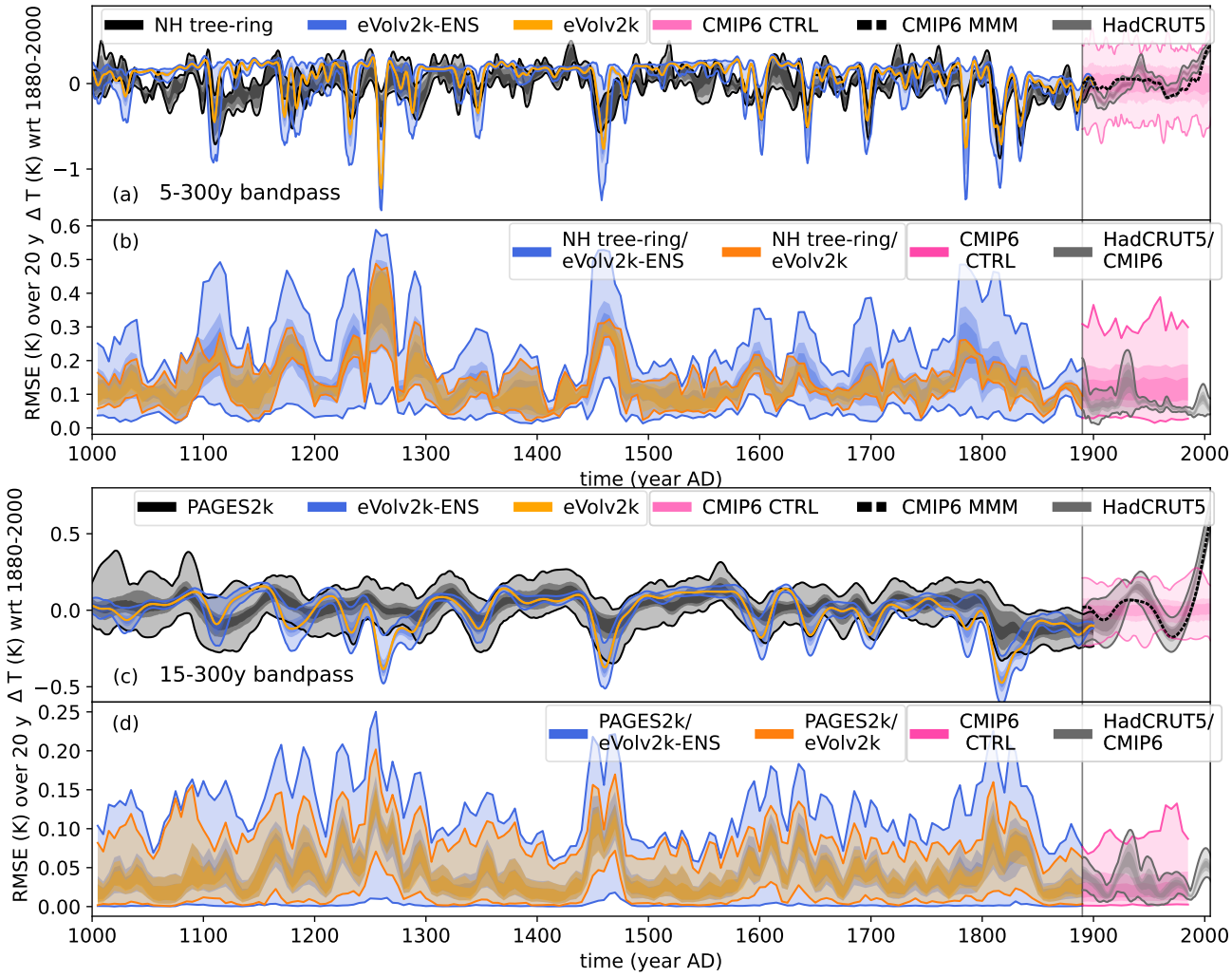

Figure 4. (a) Temperature anomaly from tree-ring reconstructions and response model simulations for NH land summer over the pre-industrial last Millennium (up to AD 1885), filtered by a 5 to 300 years bandpass. Observations, CMIP6 historical multi-model mean (MMM) and CMIP6 control runs are presented from 1885 onwards. Note that the CMIP6 control runs are shown for reference only and are not associated with this particular time period. (b) Residuals between tree-ring reconstructions and response model simulations over 20 y running windows. From 1890 onwards residuals between HadCRUT5 and the CMIP6 multi-model mean are shown and compared to CMIP6 control variability. (c) and (d) as (a) and (b), but for global annual data, and compared to PAGES 2k using a 15 to 300 years filter. The shading in all panels indicates the 75th, 90th and 100th percentile, thin lines the range, thick lines the median (where applicable).

345 residuals is well in line with the observed unforced variability. Thus, for most eruptions we can find an ensemble member that is in line with the uncertainty range of observed internal variability and control variability. It is likely that the lowest range overfits the data to some extent, however it is well in line with observed unforced variability (Fig. 5 (a)). For most of

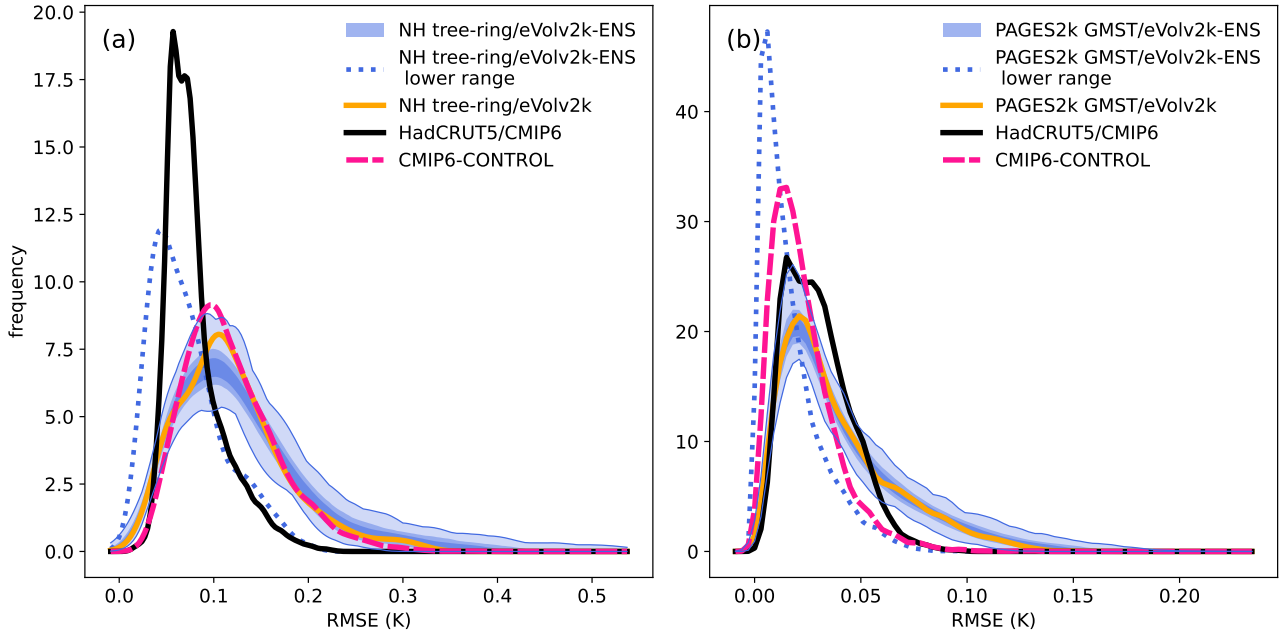

Figure 5. Distribution of residuals between proxy reconstructions and impulse response model simulation from 1300 to 1900, compared to the distribution of CMIP6 control variability and estimated unforced variability in HadCRUT5. (a) shows NH land summer temperature and (b) global annual temperature. The proxy/eVolv2k-ENS lower range represents the lower range of the residuals (thick blue line) shown in Fig. 4 (b) and (d).

the instrumental period– except the 1920s– the 20 y variability of control runs exceeds the estimates of observed unforced variability. Based on these findings, we conclude that for many eruptions, the temperature differences resulting from volcanic forcing uncertainty is in line with the magnitude of internal variability on decadal timescale. This indicates that differences in the response to volcanic eruptions due to uncertainty in the eruption parameters may not emerge from the background of internal variability. This is particularly important for comparing model simulations to proxy reconstructions. For eruptions which exceed the control variability, it is likely that the proxy temperature reconstructions can constrain the volcanic forcing by identifying forcing ensemble members that produce temperature anomalies incompatible with the reconstructions. This includes for example 1345, Samalas, the unknown mid 15th century eruption, the tropical eruptions in the 17th century (1600, 1640, 1695), Laki (1783) and the 19th century eruptions (1809 and Tambora 1815). However, residuals from most of eVolv2k-ENS and residuals from eVolv2k fully agree with the range of control variability after 1300 (Fig. 5 (a)).

The ensemble spread for global temperature is less variable, and the biggest spread just about exceeds 0.3 K (Fig. 4 (c)). This is partly caused by the smaller variability of global annual temperature in contrast to NH land summer temperature, and thus could be an effect of the tuning. It could also reflect a possibility that random errors average out in the global reconstruction.

However, note that high frequency variability is strongly suppressed due to the bandpass filter that we applied to global annual temperature (compare Fig. S15 (a), (c)).

At global scale the 90th percentile of the uncertainty ensemble agrees at all times with the range of eVolv2k residuals. This implies that after removing the higher frequencies with the 15 to 300 years bandpass filter, spread due to volcanic forcing uncertainty is small compared to the proxy uncertainty range. We therefore concentrate on the analysis of the residuals from NH reconstructions and simulations. However, we note that the lower range of the eVolv2k-ENS spread perfectly matches the lower range of the control variability at all times for the global residuals, and exceeds the upper range only during periods of very large volcanic forcing. eVolv2k lies well within the range of control variability, and at times of low volcanic forcing the residuals of PAGES 2k/eVolv2k, CMIP6 CTRL and HadCRUT5/CMIP6 match almost perfectly. However, it is likely that the lowest range is an overfit of the data (see Fig. 5 (b)), due to too many degrees of freedom.

Our results suggest that internal variability and proxy reconstruction uncertainty may at times disguise the effects of volcanic forcing uncertainty, which range in the same order of magnitude. However, overall, they confirm the large-scale quality of proxy reconstructions, where the distribution of residuals between eVolv2k and proxy reconstructions overlaps well with the distribution of control variability and observed unforced variability (Fig. 5).

Note that due to the way eVolv2k-ENS was compiled, it does not represent systematic uncertainty, but randomly combines the uncertainties associated with individual eruptions. Thus, the distribution of the residuals over the whole Millennium do not differ very much between the different ensemble members (Fig. 5). Compared to eVolv2k, the main difference can be found at the upper tail of the distribution, which, for most ensemble members, is shifted towards larger residuals. In Fig. 4 (b) and (d), the lower range represents the minimum residual which can be generated by one ensemble member over 20 years. Due to the limited ensemble size of 1000 ensemble members, randomly combining 105 eruptions, the lower range does not represent one individual ensemble member but a combination of them over the whole millennium for each 20 year slice. While this may strongly overfit the data due to the large number of degrees of freedom, it can give information about the very lowest boundary of residuals that can be achieved by tweaking the volcanic response. Remarkably, the lower range of NH residuals fits the estimate of observed variability relatively well (Fig. 5 (a)), although the left tail of the distribution indicates that a significant number of residuals approach zero. This is not consistent with any other data source, and thus clearly indicates overfitting. For global data, the same observation can be made, and the distribution of the lower range is considerably smaller than control and estimated observed variability. However, the large discrepancy between the distribution of eVolv2k-ENS and the distribution of the lower range of residuals indicates that choice of the volcanic forcing record can influence the variability of last Millennium temperature significantly, especially at, but not limited to, higher frequencies.

As a sensitivity study, we repeat our analysis with the highest and the lowest climate sensitivity presented by a CMIP6 model. We find that for very high climate sensitivity, the differences between the ensemble members are much elevated (Fig. S16), making the residuals overall less consistent with control variability (Fig. S18 (a), (b)). This concerns both residuals obtained from eVolv2k and eVolv2k-ENS, although the best fit considerately overlaps with control and observed variability. In contrast, very low climate sensitivity suppresses the response to forcing and thus decreases the spread of eVolv2k-ENS (Fig. S17) and the distribution of residuals is shifted towards zero (Fig. S18 (c), (d)). Consequentially, if climate sensitivity was very low, the

effects of volcanic forcing uncertainty may be to a very large extent disguised by internal variability.

The exact effect of timing and magnitude uncertainty on the temperature response to individual eruptions can be quantified by focusing on 11 eruptions with the largest eruption magnitudes (Fig. 6 (a)–(k)). Due to their better sensitivity at higher frequencies, we limit our analysis to NH temperature. Furthermore, we consider only eruptions past AD 1300 due to the increased quality of proxy reconstructions (Wilson et al., 2016). We compare the eVolv2k-ENS ensemble spread to the proxy reconstructions, and find the forcing ensemble members that minimise the residuals between response model and each proxy reconstruction. For most eruptions we find at least one ensemble member that matches the proxy reconstruction very closely in the immediate aftermath of an eruption, in particular 1345, 1640, 1695, 1783 and all 19th century eruptions. This has a significant effect on the superposed epoch analysis which combines these eruptions (Fig. 6 (l)), and brings the simulated volcanic response in line with the proxy reconstructions. Eruptions which still show a significant discrepancy include 1453 and 1458 and 1600.

Comparing the eruption parameters of eVolv2k-ENS best fit and eVolv2k, we find that for most eruptions the sulfate injection magnitude favoured by the proxy reconstructions is significantly reduced (table S2 and S22). Only for 1453 all four reconstructions suggest a larger sulfate injection magnitude than eVolv2k, while for 1835 three reconstructions agree on a slightly larger one. For all other eruptions considered, the majority of proxy reconstructions suggest smaller amplitudes ranging from a reconstruction average of 32% for Laki (1783) to 88% for Tambora (1815). From the four reconstructions considered, GCS17 stands out regarding the magnitude, suggesting a higher magnitude for about half the eruptions considered (1453, 1695, 1809, 1815 and 1835).

For undated eruptions, eVolv2k-ENS accounts not just for the uncertainty of sulfate injection, but also for uncertainty in eruption month and year. Out of the eleven eruptions shown in Fig. 6, this concerns 1345, 1453, 1458, 1694 and 1809. While the eruption year plays a role regarding the spread of the temperature response, the eruption month does not seem to account for much variance (Fig. S24). It is likely, that this is caused by the simple nature of our temperature response model, which does not take seasonal processes into account, and thus we cannot draw any significant conclusions regarding the role of the eruption season. The relative eruption year varies much between the different undated eruptions, as well as the different reconstructions, with no visible trend (S23). We find that only for 1695 all reconstructions agree on a slightly later eruption date, ranging between $+1$ to $+3$ years. For the other eruptions results range between -2 and $+3$ years.

The wide range of volcanic forcing uncertainty found in the temperature response to the 11 individual eruptions leads to a similarly large range in the epoch analysis combining these eruptions (Fig. 6 (l)). At the time of the maximum amplitude, two years after an eruption, the difference between the different ensemble members ranges around 0.8 K. The ensemble range encompasses the range of proxy reconstructions at all times, while the simulation using eVolv2k shows a larger response until around four years after the eruption. At this point we find a quick recovery from the cooling, while the proxy reconstructions show more persistent cooling, leading to an overlap between proxy reconstructions and eVolv2k response. In contrast, the weaker cooling and slower recovery is replicated by the best fitting eVolv2k-ENS members. Here, we find an almost perfect overlap within the proxy reconstruction range, although the epoch analysis still shows a slight discrepancy in the timing and

magnitude of the peak cooling, and a more prolonged response in the proxy reconstruction. The latter could be explained by biological memory processes in the tree-ring data (Lücke et al., 2019). Note that the response model does not account for internal variability, but the volcanic response in the proxy reconstruction is affected by it. Therefore we do not necessarily expect the best fitting ensemble members to be the true implementation of volcanic forcing over these eruptions. However, it should be noted that the overall response– including magnitude, peak cooling and recovery time– shown by the epoch analysis is heavily affected by volcanic forcing, and can differ by up to 0.8 K during peak cooling. This exceeds the magnitude of the difference between PMIP3 models and proxy reconstructions (Masson-Delmotte et al., 2013).

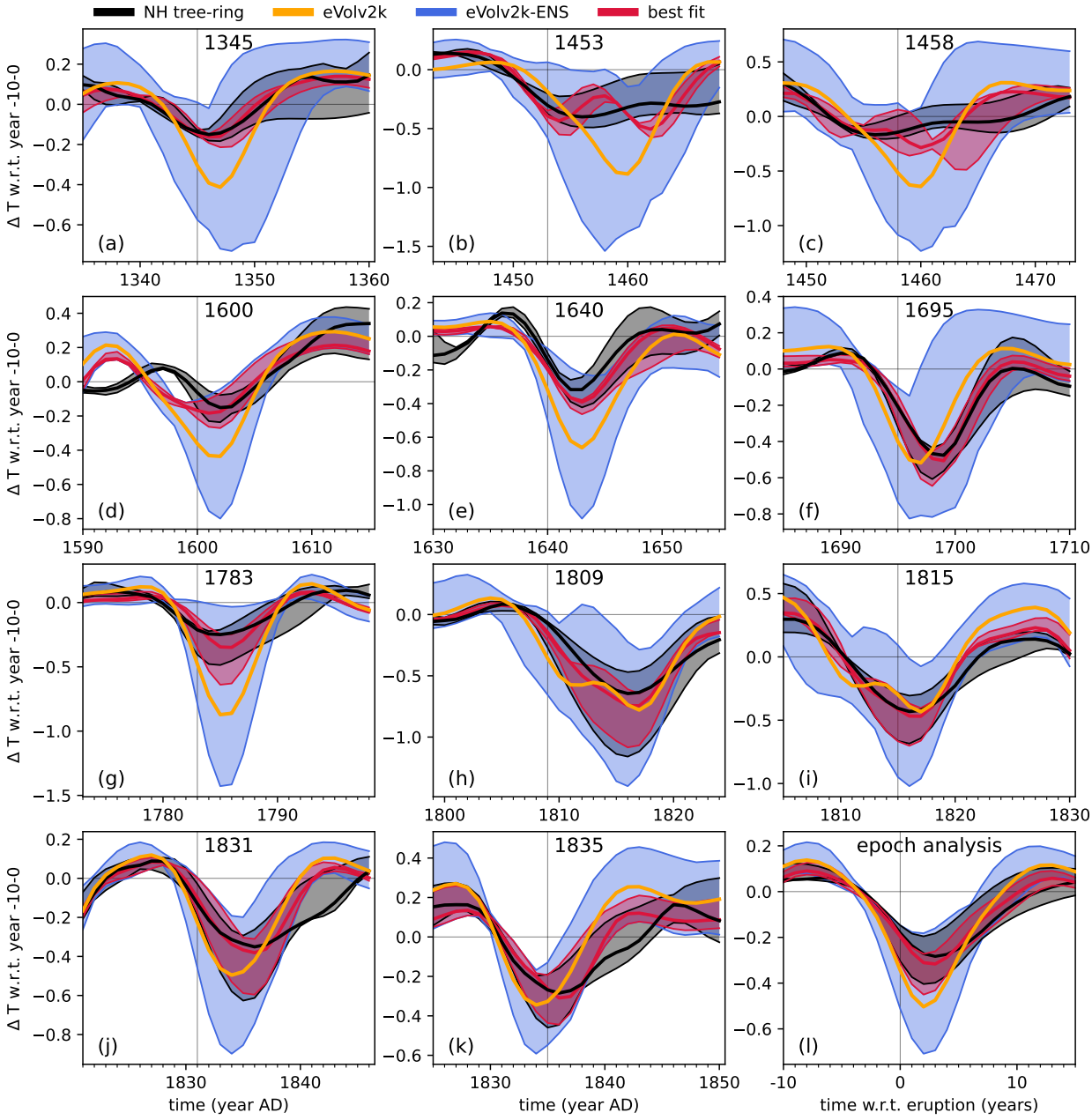

Figure 6. (a)–(k): Temperature response to the 11 biggest eruptions post AD 1300, filtered by a 5 to 300 years bandpass. Title indicates the eruption year in eVolv2k. Anomalies are taken w.r.t. ten years prior to the eruption in eVolv2k, up to the eruption year. (l) Epoch analysis over all eruptions shown in panels (a)–(k). In all panels the shading indicates the range, thick line the median (where applicable).

4 Discussion and Conclusion

In this study, we have, for the first time, estimated the effects of both volcanic and solar forcing uncertainty on simulated temperature during the last millennium. This includes the uncertainty of magnitude and timing of volcanic eruptions and the uncertainty of the long-term variation of the solar cycle. Our main findings include:

Solar forcing:

- We confirm previous findings that the SEA solar reconstruction is not consistent with global and hemispheric proxy reconstructions using an independent approach. The same applies for the new PMOD record, producing consistently higher residuals between model simulations and proxy reconstructions than SATIRE forced simulations. For the first time we have shown that this finding is not affected by volcanic forcing uncertainty. We therefore advise that the PMOD and the SEA forcing records should not be used in future analyses.

- We find that low amplitude solar forcing estimates keep the residuals between model simulations and proxy reconstructions consistent with estimates of internal variability from control runs on centennial timescale. Residual variability between the large-scale temperature response to SATIRE solar forcing records and proxy reconstructions agrees exceptionally well with variability from control simulations, even when accounting for uncertainty in the volcanic record and proxy reconstruction uncertainty. Given that temperature reconstructions from proxy records and climate model simulations represent two completely independent sources regarding data and methodology, this agreement strongly supports the quality of proxy reconstructions as well as the SATIRE solar forcing records at these timescales.

Volcanic forcing:

- Our study also shows that volcanic forcing uncertainty has the potential to resolve several discrepancies between model simulations and proxy reconstructions for some individual eruptions, such as 1640, 1695, Laki or Tambora. The best fit between proxy reconstructions and temperature response to eVolv2k-ENS tends to favour weaker sulfate injection. This is a well-known discrepancy, which could result from 1) a weaker response of biological proxies due to memory processes (Lücke et al., 2019; Zhu et al., 2020), 2) reflect a systematic error in the transfer functions used to translate ice core sulfate flux to VSSI, or 3) imply that climate sensitivity of the response model is too high, resulting in an overestimation of the forced response. While the exact cause for this discrepancy remains open, our results show that volcanic uncertainty cannot be neglected for the evaluation of the response to volcanic forcing and has the potential to increase the agreement between proxy reconstructions and climate model simulations remarkably.

- These uncertainties add up considerably in superposed epoch analyses, where different ensemble members can induce a large spread in the amplitude. Thus, when including poorly constrained eruptions the epoch analysis does not seem to be suitable to entirely isolate the volcanic signal. We recommend to focus on comparing better constrained individual eruptions, and/or to include an estimate of forcing uncertainty.

Unforced variability:

- Comparison of simulated control variability with unforced variability estimated from the observations (see Sect. 2.4.5) suggests that on decadal timescales control variability may be overestimated for NH land summer and underestimated for global annual data. The latter confirms previous findings by Neukom et al. (2019). Incompletely removed external forcing in residuals between proxy reconstructions and model simulations, arising from discrepancies in their forcing history, induce a large spread of results. This supports findings by Mann et al. (2022) who suggested that such residual external forcing explains much of the multidecadal variability previously attributed to the Atlantic Multidecadal Oscillation.

- We also find that within the range of volcanic forcing uncertainty, residual variability between model simulations and proxy reconstructions is consistent with both control and observed unforced variability on decadal timescales. All three data sources agree well within their own range of uncertainties. This agreement between two independent records, i.e. the forcing and the climate response, emphasises the quality of both proxy reconstructions of the last Millennium within their band of frequency sensitivity and volcanic forcing reconstructions within their uncertainty range.

These conclusions are, to a greater or lesser extent, all sensitive to our assumptions, input data and methods. Here, we discuss these sensitivities.

Primarily, the results rely on the performance of the impulse response model and the quality of the proxy reconstructions within the chosen frequency bands both of which are subject to a range of uncertainties and biases. The impulse response model is a simple large-scale simulation of energy balance, and does not account for complex processes within the climate system, seasonal or hemispheric phenomena, or abrupt changes on sub-annual timescale. Such models are therefore better suited for the simulation of global annual temperature. In contrast, large-scale proxy reconstructions are impacted by the lack of high quality high resolution proxy records covering the whole globe and reconstructions of extra-tropical northern hemisphere climate tend to be more reliable than global reconstructions. Therefore a balance has been struck between these two contrasting issues. We use the impulse response model here to simulate regional and seasonal temperature, such as midlatitudinal NH summer, to allow for a better comparison to tree-ring datasets. Therefore we carefully evaluated the model performance which agrees sufficiently well with that of HadCM3.

Impulse response models were initially developed for modeling the response to slow-acting forcing, in particular greenhouse gases (Held et al., 2010; Rypdal, 2012; Geoffroy et al., 2013b, a) and have only recently been used for simulating the response to volcanic forcing (Haustein et al., 2019; Marshall et al., 2020). The latest studies have found that in order to capture the full range of CMIP6 model behaviour, impulse response models require at least 3 boxes, and the 2-box model used here will not capture the full response of a more complex model (Tsutsui, 2016, 2020; Cummins et al., 2020). We have carefully tested our model output by comparing the forced response and the residual variability to HadCM3 simulations– which is also subject to model uncertainty. The response model produced systematically larger residuals compared to HadCM3, and was not able to reproduce certain non-linearities in the forced response. Nevertheless it provides a good approximation of forced pre-industrial temperature variability and is able to run a large ensemble of possible realisations of forcing combinations. The conversion between SAOD and radiative forcing is a further source of uncertainty, but likely less important than the overall uncertainty on

the SAOD timeseries across the EVA ensemble. For example if the NH conversion was reduced, the NH radiative forcing input into the impulse response model would be lower, which could result in better agreement between the proxy reconstructions and model output. These issues all contribute to our expectation of a significant model error, which has been estimated by comparing the residuals between HadCM3 and response model simulations using HadCM3 forcing. We found that the model error is small compared to the uncertainty associated with climate sensitivity, at least for the specific filtering and target frequencies. Climate sensitivity induces a large systematic uncertainty in the amplitude of simulated variability in the response model. However, this uncertainty affects the amplitude equally for each simulation, and thus may, in case of higher climate sensitivity, increase or, in case of lower climate sensitivity, decrease the relative differences between the temperature response, but does not change the conclusions of this article.

Despite these simplifications and assumptions our results are robust, as they are based on relative comparisons of different datasets under the same set of assumptions, such as the comparison of the temperature response to the different forcing records to the proxy data. Our results are also consistent with those obtained from current generation (CMIP6) global climate simulations for the historical period and the observational record. While we do not expect the response model simulations to be a perfect implementation of the forced response throughout the last millennium, they give a good approximation within the range of the various uncertainties associated with the datasets. A more sophisticated model may be able to produce more pronounced differences in the forced response, for example regarding the influence of the seasonality of volcanic eruptions on the temperature response. However, the high agreement between the four independent data sources– proxy reconstructions, the simulated response to solar and volcanic forcing, CMIP6 control variability and direct observations– strongly supports our conclusions.

Code and data availability. The data and code generated during this study will be available on the University of Edinburgh DataShare server (upload pending). eVolv2k is available from the World Data Center for Climate: http://doi.org/10.1594/WDCC/eVolv2k_v1 and eVolv2k–ENS will be hosted on the same server (upload pending). The HadCM3 simulations are available from the Center for Environmental Data Analysis: https://catalogue.ceda.ac.uk/uuid/9b51eb5d44524e4c9aebcbbc53d79a27. The HadCRUT5 dataset is available from the Climatic Research Unit: https://crudata.uea.ac.uk/cru/data/temperature/#datdow. CMIP6 simulations are available via the World Climate Research Programme: https://esgf-node.llnl.gov/projects/cmip6/. The solar reconstructions are available from the Paleoclimate Modelling Intercomparison Project - Phase 4: https://pmip4.lsce.ipsl.fr/doku.php/data:solar and the Paleoclimate Modelling Intercomparison Project Phase III: https://wiki.lsce.ipsl.fr/pmip3/doku.php/index.

All Paleo datasets are publicly available from the NOAA National Centers for Environmental Information (NCEI), under the World Data Service (WDS) for Paleoclimatology. Both N-TREND reconstructions are available from: https://doi.org/10.25921/kztr-jd59. The PAGES 2k GMST reconstruction is available from: https://doi.org/10.25921/tkxp-vn12. The reconstruction generated by Guillet et al. (2017) is available from: https://doi.org/10.25921/42gh-z167. The reconstruction generated by Schneider et al. (2015) is available from: https://doi.org/10.25921/6mdt-5246.

Author contributions. L.L. performed the computations, analysed the data, designed the figures and drafted the manuscript. A.S. provided data and performed the computations to estimate observed unforced variability. A.S. and G.H. conceived the idea and supervised the project. M.T. developed eVolv2k–ENS and wrote section 2.1. All authors discussed the results and contributed to the final manuscript.

Competing interests. The authors declare that they have no conflict of interests.

Acknowledgements. LL, AS, LM and GH were funded by the Natural Environment Research Council (NERC) project Vol-Clim (NE/S000887/1). LL was further supported by a studentship from the NERC E3 Doctoral training partnership [grant number NE/L002558/1]. AS and GH were also funded by the NERC project GloSAT (NE/S015698/1), and by NERC under the Belmont forum, Grant PacMedy (NE/P006752/1), and AS received funding from a Chancellors fellowship at the University of Edinburgh. This paper benefitted from discussion at events of the Past Global Changes (PAGES) working group "Volcanic Impacts on Climate and Society" (VICS). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and thank all the climate modelling groups for producing and making available their model output. We acknowledge the Northern Hemisphere Tree-Ring Network Development (N-TREND), Past Global Changes (PAGES) project and the other authors of proxy reconstructions for providing publicly available data. We thank Karsten Haustein for developing and sharing the code for his impulse response model.

References

Allen, M. R. and Stott, P. A.: Estimating signal amplitudes in optimal fingerprinting, part I: theory, Climate Dynamics, 21, 477–491, https://doi.org/10.1007/s00382-003-0313-9, 2003.

Ammann, C. M., Joos, F., Schimel, D. S., Otto-Bliesner, B. L., and Tomas, R. A.: Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model, Proceedings of the National Academy of Sciences, 104, 3713–3718, https://doi.org/10.1073/pnas.0605064103, 2007.

Anchukaitis, K. J., Breitenmoser, P., Briffa, K. R., Buchwal, A., Büntgen, U., Cook, E. R., D'Arrigo, R. D., Esper, J., Evans, M. N., Frank, D., Grudd, H., Gunnarson, B. E., Hughes, M. K., Kirdyanov, A. V., Krner, C., Krusic, P. J., Luckman, B., Melvin, T. M., Salzer, M. W., Shashkin, A. V., Timmreck, C., Vaganov, E. A., and Wilson, R. J. S.: Tree rings and volcanic cooling, Nature Geoscience, 5, 836–837, https://doi.org/10.1038/ngeo1645, 2012.

Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D'Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B. E., Hegerl, G., Helama, S., Klesse, S., Krusic, P. J., Linderholm, H. W., Myglan, V., Osborn, T. J., Zhang, P., Rydval, M., Schneider, L., Schurer, A., Wiles, G., and Zorita, E.: Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions, Quaternary Science Reviews, 163, 1–22, https://doi.org/10.1016/j.quascirev.2017.02.020, 2017.

Aubry, T. J., Toohey, M., Marshall, L., Schmidt, A., and Jellinek, A. M.: A New Volcanic Stratospheric Sulfate Aerosol Forcing Emulator (EVA_H): Comparison With Interactive Stratospheric Aerosol Models, Journal of Geophysical Research: Atmospheres, 125, https://doi.org/10.1029/2019jd031303, 2020.

Barboza, L., Li, B., Tingley, M. P., and Viens, F. G.: Reconstructing past temperatures from natural proxies and estimated climate forcings using short- and long-memory models, The Annals of Applied Statistics, 8, https://doi.org/10.1214/14-aoas785, 2014.

Bard, E., Raisbeck, G. M., Yiou, F., and Jouzel, J.: Solar modulation of cosmogenic nuclide production over the last millennium: comparison between $^{1}4$C and $^{1}0$Be records, Earth and Planetary Science Letters, 150, 453–462, https://doi.org/10.1016/s0012-821x(97)00082-4, 1997.

Baroni, M., Bard, E., and Team, A.: A new 10Be record recovered from an Antarctic ice core: validity and limitations to record the solar activity, in: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, p. 6357, 2015.

Baroni, M., Bard, E., Petit, J.-R., and and, S. V.: Persistent Draining of the Stratospheric 10 Be Reservoir After the Samalas Volcanic Eruption (1257 CE), Journal of Geophysical Research: Atmospheres, https://doi.org/10.1029/2018jd029823, 2019.

Beer, J., Andree, M., Oeschger, H., Stauffer, B., Balzer, R., Bonani, G., Stoller, C., Suter, M., Woelfli, W., and Finkel, R. C.: Temporal 10Be Variations in Ice, Radiocarbon, 25, 269–278, https://doi.org/10.1017/s0033822200005579, 1983.

Berger, A. and Loutre, M. F.: Insolation values for the climate of the last 10 million years, Quaternary Science Reviews, 10, 297–317, https://doi.org/10.1016/0277-3791(91)90033-q, 1991.

Berger, A. L.: Long-Term Variations of Daily Insolation and Quaternary Climatic Changes, Journal of the Atmospheric Sciences, 35, 2362–2367, https://doi.org/10.1175/1520-0469(1978)035<2362:ltvodi>2.0.co;2, 1978.

Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K.: Drought Reconstructions for the Continental United States, Journal of Climate, 12, 1145–1162, 1999.

Crowley, T. J.: Causes of Climate Change Over the Past 1000 Years, Science, 289, 270–277, https://doi.org/10.1126/science.289.5477.270, 2000.

Crowley, T. J. and Lowery, T. S.: How Warm Was the Medieval Warm Period?, AMBIO: A Journal of the Human Environment, 29, 51–54, https://doi.org/10.1579/0044-7447-29.1.51, 2000.

Crowley, T. J. and Unterman, M. B.: Technical details concerning development of a 1200 yr proxy index for global volcanism, Earth System Science Data, 5, 187–197, https://doi.org/10.5194/essd-5-187-2013, 2013.

Cummins, D. P., Stephenson, D. B., and Stott, P. A.: Optimal Estimation of Stochastic Energy Balance Model Parameters, Journal of Climate, 33, 7909–7926, https://doi.org/10.1175/jcli-d-19-0589.1, 2020.

Esper, J., Schneider, L., Smerdon, J. E., Schöne, B. R., and Büntgen, U.: Signals and memory in tree-ring width and density data, Dendrochronologia, 35, 62–70, https://doi.org/10.1016/j.dendro.2015.07.001, 2015.

Esper, J., Büntgen, U., Hartl-Meier, C., Oppenheimer, C., and Schneider, L.: Northern Hemisphere temperature anomalies during the 1450s period of ambiguous volcanic forcing, Bulletin of Volcanology, 79, https://doi.org/10.1007/s00445-017-1125-9, 2017.

Etminan, M., Myhre, G., Highwood, E. J., and Shine, K. P.: Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing, Geophysical Research Letters, 43, https://doi.org/10.1002/2016gl071930, 2016.

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

Feulner, G.: Are the most recent estimates for Maunder Minimum solar irradiance in agreement with temperature reconstructions?, Geophysical Research Letters, 38, n/a–n/a, https://doi.org/10.1029/2011gl048529, 2011.

Franke, J., Frank, D., Raible, C. C., Esper, J., and Brönnimann, S.: Spectral biases in tree-ring climate proxies, Nature Climate Change, 3, 360–364, https://doi.org/10.1038/nclimate1816, 2013.

Friedman, A. R., Hegerl, G. C., Schurer, A. P., Lee, S.-Y., Kong, W., Cheng, W., and Chiang, J. C. H.: 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, 3487–3509, https://doi.org/10.1175/jcli-d-19-0102.1, 2020.

Fröhlich, C.: Solar Irradiance Variability Since 1978, in: Solar Variability and Planetary Climates, pp. 53–65, Springer New York, https://doi.org/10.1007/978-0-387-48341-2_5, 2006.

Gao, C., Robock, A., and Ammann, C.: Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models, Journal of Geophysical Research, 113, https://doi.org/10.1029/2008jd010239, 2008.

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J. L., and Tytéca, S.: Transient Climate Response in a Two-Layer Energy-Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs, Journal of Climate, 26, 1859–1876, https://doi.org/10.1175/jcli-d-12-00196.1, 2013a.

Geoffroy, O., Saint-Martin, D., Olivié, D. J. L., Voldoire, A., Bellon, G., and Tytéca, S.: Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments, Journal of Climate, 26, 1841–1857, https://doi.org/10.1175/jcli-d-12-00195.1, 2013b.

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., and Wood, R. A.: 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, 147–168, https://doi.org/10.1007/s003820050010, 2000.

Gray, L. J., Beer, J., Geller, M., Haigh, J. D., Lockwood, M., Matthes, K., Cubasch, U., Fleitmann, D., Harrison, G., Hood, L., Luterbacher, J., Meehl, G. A., Shindell, D., van Geel, B., and White, W.: SOLAR INFLUENCES ON CLIMATE, Reviews of Geophysics, 48, https://doi.org/10.1029/2009rg000282, 2010.

Guevara-Murua, A., Williams, C. A., Hendy, E. J., Rust, A. C., and Cashman, K. V.: Observations of a stratospheric aerosol veil-
625 from a tropical volcanic eruption in December 1808: is this the *Unknown* ∼1809 eruption?, Climate of the Past, 10, 1707–1722, https://doi.org/10.5194/cp-10-1707-2014, 2014.

Guillet, S., Corona, C., Stoffel, M., Khodri, M., Lavigne, F., Ortega, P., Eckert, N., Sielenou, P. D., Daux, V., (Sidorova), O. V. C., Davi, N., Edouard, J.-L., Zhang, Y., Luckman, B. H., Myglan, V. S., Guiot, J., Beniston, M., Masson-Delmotte, V., and Oppenheimer, C.: Climate response to the Samalas volcanic eruption in 1257 revealed by proxy records, Nature Geoscience, 10, 123–128,
630 https://doi.org/10.1038/ngeo2875, 2017.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, https://doi.org/10.1002/2016jd024751, 2016.

Hanhijärvi, S., Tingley, M. P., and Korhola, A.: Pairwise comparisons to reconstruct mean temperature in the Arctic Atlantic Region over the
635 last 2,000 years, Climate Dynamics, 41, 2039–2060, https://doi.org/10.1007/s00382-013-1701-4, 2013.

Haustein, K., Allen, M. R., Forster, P. M., Otto, F. E. L., Mitchell, D. M., Matthews, H. D., and Frame, D. J.: A real-time Global Warming Index, Scientific Reports, 7, https://doi.org/10.1038/s41598-017-14828-5, 2017.

Haustein, K., Otto, F. E. L., Venema, V., Jacobs, P., Cowtan, K., Hausfather, Z., Way, R. G., White, B., Subramanian, A., and Schurer, A. P.: A Limited Role for Unforced Internal Variability in Twentieth-Century Warming, Journal of Climate, 32, 4893–4917,
640 https://doi.org/10.1175/jcli-d-18-0555.1, 2019.

Hegerl, G. C., Crowley, T. J., Baum, S. K., Kim, K.-Y., and Hyde, W. T.: Detection of volcanic, solar and greenhouse gas signals in paleoreconstructions of Northern Hemispheric temperature, Geophysical Research Letters, 30, n/a–n/a, https://doi.org/10.1029/2002gl016635, 2003.

Hegerl, G. C., Crowley, T. J., Hyde, W. T., and Frame, D. J.: Climate sensitivity constrained by temperature reconstructions over the past
645 seven centuries, Nature, 440, 1029–1032, https://doi.org/10.1038/nature04679, 2006.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., and Vallis, G. K.: Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing, Journal of Climate, 23, 2418–2427, https://doi.org/10.1175/2009jcli3466.1, 2010.

Hind, A., Moberg, A., and Sundberg, R.: Statistical framework for evaluation of climate model simulations by use of climate proxy data from the last millennium - Part 2: A pseudo-proxy study addressing the amplitude of solar forcing, Climate of the Past, 8, 1355–1365,
650 https://doi.org/10.5194/cp-8-1355-2012, 2012.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., ichi Kurokawa, J., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geoscientific Model Development, 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

655 Judge, P. G., Lockwood, G. W., Radick, R. R., Henry, G. W., Shapiro, A. I., Schmutz, W., and Lindsey, C.: Confronting a solar irradiance reconstruction with solar and stellar data, Astronomy & Astrophysics, 544, A88, https://doi.org/10.1051/0004-6361/201218903, 2012.

Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., Egorova, T., Evans, M., González-Rouco, J. F., Goosse, H., Hurtt, G. C., Joos, F., Kaplan, J. O., Khodri, M., Goldewijk, K. K., Krivova, N., LeGrande, A. N., Lorenz, S. J., Luterbacher, J., Man, W., Maycock, A. C., Meinshausen, M., Moberg, A., Muscheler, R., Nehrbass-Ahles, C., Otto-Bliesner, B. I., Phipps, S. J., Pongratz, J.,
660 Rozanov, E., Schmidt, G. A., Schmidt, H., Schmutz, W., Schurer, A., Shapiro, A. I., Sigl, M., Smerdon, J. E., Solanki, S. K., Timmreck, C., Toohey, M., Usoskin, I. G., Wagner, S., Wu, C.-J., Yeo, K. L., Zanchettin, D., Zhang, Q., and Zorita, E.: The PMIP4 contribution to

CMIP6 - Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations, Geoscientific Model Development, 10, 4005–4033, https://doi.org/10.5194/gmd-10-4005-2017, 2017.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A. C. M., and Levrard, B.: A long-term numerical solution for the insolation quantities of the Earth, Astronomy & Astrophysics, 428, 261–285, https://doi.org/10.1051/0004-6361:20041335, 2004.

Li, S. and Jarvis, A.: Long run surface temperature dynamics of an A-OGCM: the HadCM3 4 × CO2 forcing experiment revisited, Climate Dynamics, 33, 817–825, https://doi.org/10.1007/s00382-009-0581-0, 2009.

Lockwood, M. and Ball, W. T.: Placing limits on long-term variations in quiet-Sun irradiance and their contribution to total solar irradiance and solar radiative forcing of climate, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 476, 20200 077, https://doi.org/10.1098/rspa.2020.0077, 2020.

Lücke, L. J., Hegerl, G. C., Schurer, A. P., and Wilson, R.: Effects of Memory Biases on Variability of Temperature Reconstructions, Journal of Climate, 32, 8713–8731, https://doi.org/10.1175/jcli-d-19-0184.1, 2019.

Lücke, L. J., Schurer, A. P., Wilson, R., and Hegerl, G. C.: Orbital Forcing Strongly Influences Seasonal Temperature Trends During the Last Millennium, Geophysical Research Letters, 48, https://doi.org/10.1029/2020gl088776, 2021.

Luterbacher, J., Xoplaki, E., Dietrich, D., Rickli, R., Jacobeit, J., Beck, C., Gyalistras, D., Schmutz, C., and Wanner, H.: Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500, Climate Dynamics, 18, 545–561, https://doi.org/10.1007/s00382-001-0196-6, 2002.

Mann, M. E.: On smoothing potentially non-stationary climate time series, Geophysical Research Letters, 31, n/a–n/a, https://doi.org/10.1029/2004gl019569, 2004.

Mann, M. E.: Smoothing of climate time series revisited, Geophysical Research Letters, 35, https://doi.org/10.1029/2008gl034716, 2008.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, Proceedings of the National Academy of Sciences, 105, 13 252–13 257, https://doi.org/10.1073/pnas.0805721105, 2008.

Mann, M. E., Steinman, B. A., Brouillette, D. J., Fernandez, A., and Miller, S. K.: On the Estimation of Internal Climate Variability During the Preindustrial Past Millennium, Geophysical Research Letters, 49, https://doi.org/10.1029/2021gl096596, 2022.

Marshall, L. R., Johnson, J. S., Mann, G. W., Lee, L., Dhomse, S. S., Regayre, L., Yoshioka, M., Carslaw, K. S., and Schmidt, A.: Exploring How Eruption Source Parameters Affect Volcanic Radiative Forcing Using Statistical Emulation, Journal of Geophysical Research: Atmospheres, 124, 964–985, https://doi.org/10.1029/2018jd028675, 2019.

Marshall, L. R., Smith, C. J., Forster, P. M., Aubry, T. J., Andrews, T., and Schmidt, A.: Large Variations in Volcanic Aerosol Forcing Efficiency Due to Eruption Source Parameters and Rapid Adjustments, Geophysical Research Letters, 47, https://doi.org/10.1029/2020gl090241, 2020.

Marshall, L. R., Schmidt, A., Johnson, J. S., Mann, G. W., Lee, L. A., Rigby, R., and Carslaw, K. S.: Unknown Eruption Source Parameters Cause Large Uncertainty in Historical Volcanic Radiative Forcing Reconstructions, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2020jd033578, 2021.

Masson-Delmotte, V., Schulz, M., Abe-Ouichi, A., Beer, J., Ganopolski, A., Rouco, J. G., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A.: Information from Paleoclimate Archives, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., Fraser, P. J., Montzka, S. A., Rayner, P. J., Trudinger, C. M., Krummel, P. B., Beyerle, U., Canadell, J. G., Daniel, J. S., Enting, I. G., Law, R. M., Lunder, C. R., O'Doherty, S., Prinn, R. G., Reimann, S., Rubino, M., Velders, G. J. M., Vollmer, M. K., Wang, R. H. J., and Weiss, R.: Historical greenhouse gas concentrations for climate modelling (CMIP6), Geoscientific Model Development, 10, 2057–2116, https://doi.org/10.5194/gmd-10-2057-2017, 2017.

Millar, R. J., Nicholls, Z. R., Friedlingstein, P., and Allen, M. R.: A modified impulse-response representation of the global near-surface air temperature and atmospheric concentration response to carbon dioxide emissions, Atmospheric Chemistry and Physics, 17, 7213–7228, https://doi.org/10.5194/acp-17-7213-2017, 2017.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2019jd032361, 2021.

Muscheler, R., Joos, F., Beer, J., Müller, S. A., Vonmoos, M., and Snowball, I.: Solar activity during the last 1000 years inferred from radionuclide records, Quaternary Science Reviews, 26, 82–97, https://doi.org/10.1016/j.quascirev.2006.07.012, 2007.

Nandy, D., Martens, P. C. H., Obridko, V., Dash, S., and Georgieva, K.: Solar evolution and extrema: current state of understanding of long-term solar variability and its planetary impacts, Progress in Earth and Planetary Science, 8, https://doi.org/10.1186/s40645-021-00430-x, 2021.

Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., González-Rouco, F., Linsley, B. K., Moy, A. D., Mundo, I., Raible, C. C., Steig, E. J., van Ommen, T., Vance, T., Villalba, R., Zinke, J., and Frank, D.: Inter-hemispheric temperature variability over the past millennium, Nature Climate Change, 4, 362–367, https://doi.org/10.1038/nclimate2174, 2014.

Neukom, R., Schurer, A. P., Steiger, N. J., and Hegerl, G. C.: Possible causes of data model discrepancy in the temperature history of the last Millennium, Scientific Reports, 8, https://doi.org/10.1038/s41598-018-25862-2, 2018.

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufmann, D. S., Lücke, L., Rehfeld, K., Schurer, A., Zhu, F., Brönniman, S., Hakim, G. J., Henley, B. J., Ljungqvist, F. C., McKay, N., Valler, V., and von Gunten, L.: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, Nature Geoscience, https://doi.org/10.1038/s41561-019-0400-0, 2019.

Otto, F. E. L., Frame, D. J., Otto, A., and Allen, M. R.: Embracing uncertainty in climate change policy, Nature Climate Change, 5, 917–920, https://doi.org/10.1038/nclimate2716, 2015.

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G.: Climate Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model, Bulletin of the American Meteorological Society, 97, 735–754, https://doi.org/10.1175/bams-d-14-00233.1, 2016.

PAGES2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, Scientific Data, 4, https://doi.org/10.1038/sdata.2017.88, 2017.

Parsons, L. A., Brennan, M. K., Wills, R. C., and Proistosescu, C.: Magnitudes and Spatial Patterns of Interdecadal Temperature Variability in CMIP6, Geophysical Research Letters, 47, https://doi.org/10.1029/2019gl086588, 2020.

Pope, V. D., Gallani, M. L., Rowntree, P. R., and Stratton, R. A.: The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3, Climate Dynamics, 16, 123–146, https://doi.org/10.1007/s003820050009, 2000.

Raible, C. C., Brönnimann, S., Auchmann, R., Brohan, P., Frölicher, T. L., Graf, H.-F., Jones, P., Luterbacher, J., Muthers, S., Neukom, R., Robock, A., Self, S., Sudrajat, A., Timmreck, C., and Wegmann, M.: Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects, WIREs Climate Change, 7, 569–589, https://doi.org/10.1002/wcc.407, 2016.

Roth, R. and Joos, F.: A reconstruction of radiocarbon production and total solar irradiance from the Holocene 14C and CO2 records: implications of data and model uncertainties, Climate of the Past, 9, 1879–1909, https://doi.org/10.5194/cp-9-1879-2013, 2013.

Rypdal, K.: Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions, Journal of Geophysical Research: Atmospheres, 117, n/a–n/a, https://doi.org/10.1029/2011jd017283, 2012.

Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., Bardeen, C. G., Conley, A., Forster, P. M., Gettelman, A., Portmann, R. W., Solomon, S., and Toon, O. B.: Volcanic Radiative Forcing From 1979 to 2015, Journal of Geophysical Research: Atmospheres, 123, 12 491–12 508, https://doi.org/10.1029/2018jd028776, 2018.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), Geoscientific Model Development, 4, 33–45, https://doi.org/10.5194/gmd-4-33-2011, 2011.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), Geoscientific Model Development, 5, 185–191, https://doi.org/10.5194/gmd-5-185-2012, 2012.

Schneider, L., Smerdon, J. E., Büntgen, U., Wilson, R. J. S., Myglan, V. S., Kirdyanov, A. V., and Esper, J.: Revising midlatitude summer temperatures back to A.D. 600 based on a wood density network, Geophysical Research Letters, 42, 4556–4562, https://doi.org/10.1002/2015gl063956, 2015.

Schneider, L., Smerdon, J. E., Pretis, F., Hartl-Meier, C., and Esper, J.: A new archive of large volcanic events over the past millennium derived from reconstructed summer temperatures, Environmental Research Letters, 12, 094 005, https://doi.org/10.1088/1748-9326/aa7a1b, 2017.

Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F. B., and Phipps, S. J.: Separating Forced from Chaotic Climate Variability over the Past Millennium, Journal of Climate, 26, 6954–6973, https://doi.org/10.1175/jcli-d-12-00826.1, 2013.

Schurer, A. P., Tett, S. F. B., and Hegerl, G. C.: Small influence of solar variability on climate over the past millennium, Nature Geoscience, 7, 104–108, https://doi.org/10.1038/ngeo2040, 2014.

Schurer, A. P., Hegerl, G. C., and Obrochta, S. P.: Determining the likelihood of pauses and surges in global warming, Geophysical Research Letters, 42, 5974–5982, https://doi.org/10.1002/2015gl064458, 2015.

Shapiro, A. I., Schmutz, W., Rozanov, E., Schoell, M., Haberreiter, M., Shapiro, A. V., and Nyeki, S.: A new approach to the long-term reconstruction of the solar irradiance leads to large historical solar forcing, Astronomy & Astrophysics, 529, A67, https://doi.org/10.1051/0004-6361/201016173, 2011.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., Heydt, A. S., Knutti, R., Mauritsen, T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B., and Zelinka, M. D.: An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence, Reviews of Geophysics, 58, https://doi.org/10.1029/2019rg000678, 2020.

Shi, F., Zhao, S., Guo, Z., Goosse, H., and Yin, Q.: Multi-proxy reconstructions of May-September precipitation field in China over the past 500 years, Climate of the Past, 13, 1919–1938, https://doi.org/10.5194/cp-13-1919-2017, 2017.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Büntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fischer, H., Kipfstuhl, S., Kostick, C., Maselli, O. J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D. R., Pilcher, J. R., Salzer, M., Schüpbach, S., Steffensen, J. P., Vinther, B. M., and Woodruff, T. E.: Timing and climate forcing of volcanic eruptions for the past 2,500 years, Nature, 523, 543–549, https://doi.org/10.1038/nature14565, 2015.

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., and Regayre, L. A.: FAIR v1.3: a simple emissions-based impulse response and carbon cycle model, Geoscientific Model Development, 11, 2273–2297, https://doi.org/10.5194/gmd-11-2273-2018, 2018.

Steinhilber, F., Beer, J., and Fröhlich, C.: Total solar irradiance during the Holocene, Geophysical Research Letters, 36, https://doi.org/10.1029/2009gl040142, 2009.

Steinhilber, F., Abreu, J. A., Beer, J., Brunner, I., Christl, M., Fischer, H., Heikkila, U., Kubik, P. W., Mann, M., McCracken, K. G., Miller, H., Miyahara, H., Oerter, H., and Wilhelms, F.: 9,400 years of cosmic radiation and solar activity from ice cores and tree rings, Proceedings of the National Academy of Sciences, 109, 5967–5971, https://doi.org/10.1073/pnas.1118965109, 2012.

Stevenson, S., Fasullo, J. T., Otto-Bliesner, B. L., Tomas, R. A., and Gao, C.: Role of eruption season in reconciling model and proxy responses to tropical volcanism, Proceedings of the National Academy of Sciences, 114, 1822–1826, https://doi.org/10.1073/pnas.1612505114, 2017.

Stoffel, M., Khodri, M., Corona, C., Guillet, S., Poulain, V., Bekki, S., Guiot, J., Luckman, B. H., Oppenheimer, C., Lebas, N., Beniston, M., and Masson-Delmotte, V.: Estimates of volcanic-induces cooling in the Nothern Hemisphere over the past 1500 years, Nature Geoscience, 8, 784–788, https://doi.org/10.1038/ngeo2526, 2015.

Stuiver, M. and Quay, P. D.: Changes in Atmospheric Carbon-14 Attributed to a Variable Sun, Science, 207, 11–19, https://doi.org/10.1126/science.207.4426.11, 1980.

Suess, H. E.: The Radiocarbon Record in Tree Rings of the Last 8000 Years, Radiocarbon, 22, 200–209, https://doi.org/10.1017/s0033822200009462, 1980.

Tett, S. F. B., Gregory, J. M., Freychet, N., Cartis, C., Mineter, M. J., and Roberts, L.: Does Model Calibration Reduce Uncertainty in Climate Projections?, Journal of Climate, 35, 2585–2602, https://doi.org/10.1175/jcli-d-21-0434.1, 2022.

Timmreck, C., Lorenz, S. J., Crowley, T. J., Kinne, S., Raddatz, T. J., Thomas, M. A., and Jungclaus, J. H.: Limited temperature response to the very large AD 1258 volcanic eruption, Geophysical Research Letters, 36, https://doi.org/10.1029/2009gl040083, 2009.

Toohey, M. and Sigl, M.: Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE, Earth System Science Data, 9, 809–831, https://doi.org/10.5194/essd-9-809-2017, 2017.

Toohey, M., Krüger, K., and Timmreck, C.: Volcanic sulfate deposition to Greenland and Antarctica: A modeling sensitivity study, Journal of Geophysical Research: Atmospheres, 118, 4788–4800, https://doi.org/10.1002/jgrd.50428, 2013.

Toohey, M., Stevens, B., Schmidt, H., and Timmreck, C.: Easy Volcanic Aerosol (EVA v1.0): an idealized forcing generator for climate simulations, Geoscientific Model Development, 9, 4049–4070, https://doi.org/10.5194/gmd-9-4049-2016, 2016.

Toohey, M., Krüger, K., Schmidt, H., Timmreck, C., Sigl, M., Stoffel, M., and Wilson, R.: Disproportionately strong climate forcing from extratropical explosive volcanic eruptions, Nature Geoscience, 12, 100–107, https://doi.org/10.1038/s41561-018-0286-2, 2019.

Tsutsui, J.: Quantification of temperature response to CO2 forcing in atmosphere–ocean general circulation models, Climatic Change, 140, 287–305, https://doi.org/10.1007/s10584-016-1832-9, 2016.

Tsutsui, J.: Diagnosing Transient Response to CO_2 Forcing in Coupled Atmosphere-Ocean Model Experiments Using a Climate Model Emulator, Geophysical Research Letters, 47, https://doi.org/10.1029/2019gl085844, 2020.

Usoskin, I. G., Solanki, S. K., and Kovaltsov, G. A.: Grand minima and maxima of solar activity: new observational constraints, Astronomy & Astrophysics, 471, 301–309, https://doi.org/10.1051/0004-6361:20077704, 2007.

Usoskin, I. G., Horiuchi, K., Solanki, S., Kovaltsov, G. A., and Bard, E.: On the common solar signal in different cosmogenic isotope data sets, Journal of Geophysical Research: Space Physics, 114, n/a–n/a, https://doi.org/10.1029/2008ja013888, 2009.

Usoskin, I. G., Gallet, Y., Lopes, F., Kovaltsov, G. A., and Hulot, G.: Solar activity during the Holocene: the Hallstatt cycle and its consequence for grand minima and maxima, Astronomy and Astrophysics, 587, A150, https://doi.org/10.1051/0004-6361/201527295, 2016.

Vieira, L. E. A., Solanki, S. K., Krivova, N. A., and Usoskin, I.: Evolution of the solar irradiance during the Holocene, Astronomy & Astrophysics, 531, https://doi.org/10.1051/0004-6361/201015843, 2011.

Wang, Y.-M., Lean, J. L., and N. R. Sheeley, J.: Modeling the Sun's Magnetic Field and Irradiance since 1713, The Astrophysical Journal, 625, 522–538, https://doi.org/10.1086/429689, 2005.

Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'Arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., Hegerl, G., Helama, S., Klesse, S., Krusic, P. J., Linderholm, H. W., Myglan, V., Osborn, T. J., Rydval, M., Schneider, L., Schurer, A., Wiles, G., Zhang, P., and Zorita, E.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, Quaternary Science Reviews, 134, 1–18, https://doi.org/10.1016/j.quascirev.2015.12.005, 2016.

Wu, C. J.: SATIRE-M reconstruction of spectral solar irradiance over the Holocene., https://doi.org/10.17617/3.11, 2017.

Zhu, F., Emile-Geay, J., Hakim, G. J., King, J., and Anchukaitis, K. J.: Resolving the Differences in the Simulated and Reconstructed Temperature Response to Volcanism, Geophysical Research Letters, 47, https://doi.org/10.1029/2019gl086908, 2020.