Resolving the differences in the simulated and reconstructed climate response to volcanism over the last millennium

Feng Zhu[1], Julien Emile-Geay[1], Gregory J. Hakim[2], Jonathan King[3,4], Kevin J. Anchukaitis[3,4,5]

[1]Department of Earth Sciences, University of Southern California, Los Angeles, CA USA
[2]Department of Atmospheric Sciences, University of Washington, Seattle, WA USA
[3]Department of Geosciences, University of Arizona, Tucson AZ USA
[4]Laboratory of Tree-Ring Research, University of Arizona, Tucson AZ USA
[5]School of Geography and Development, University of Arizona, Tucson AZ USA

Key Points:

- We explore model–proxy disagreement on the temperature response to volcanic eruptions over the past millennium.
- Using paleoclimate data assimilation with both real and synthetic data, we show that this discrepancy is due to four main factors.
- Agreement is found for tree-ring density records at the place and season these proxies record.

Plain Language Summary

The response to volcanic eruptions is a critical benchmark of the performance of climate models. Previous studies of the past millennium have identified discrepancies between model simulations and climate reconstructions regarding the temperature response to volcanic eruptions, raising concerns regarding the source of this mismatch and implications for both models and reconstructions. By evaluating the leading sources of differences between simulations and reconstructions, this study shows that accounting for known factors largely bridges the gap.

Corresponding author: Feng Zhu, `fengzhu@usc.edu`

Abstract

Explosive volcanism imposes impulse-like radiative forcing on the climate system, providing a natural experiment to study the climate response to perturbation. Previous studies have identified disagreements between paleoclimate reconstructions and climate model simulations (GCMs) with respect to the magnitude and recovery from volcanic cooling, questioning the fidelity of GCMs, reconstructions, or both. Using the paleoenvironmental data assimilation framework of the Last Millennium Reanalysis, this study investigates the causes of the disagreements, using both real and simulated data. We demonstrate that the disagreement may be resolved by assimilating tree-ring density records only, by targeting growing-season temperature instead of annual temperature, and by performing the comparison at proxy locales. Our work suggests that discrepancies between paleoclimate models and data can be largely resolved by accounting for these features of tree-ring proxy networks.

1 Introduction

Volcanic eruptions substantially affect the climate system through their direct effect on shortwave radiation entering the earth system and the ensuing influences on, and feedback to, major modes of ocean-atmosphere variability (Robock, 2000; Schneider et al., 2009). Eruptions therefore offer unique natural experiments with which to probe the fidelity of climate model simulations, understand the response of the ocean and atmosphere circulation to changes in radiative forcing, assess climate system feedbacks, and evaluate solar radiation management proposals (Soden et al., 2002; Timmreck, 2012). The sporadic and infrequent occurrence of large volcanic eruptions means that developing a deeper understanding of their effect on climate necessarily involves analyzing the response to events prior to the instrumental era. Studies of these historical events using climate models and paleoclimate proxies give conflicting estimates of the average climate response, which we address here using a model–data fusion approach.

Significant disagreements have been identified between paleoclimate reconstructions of the climate system response to volcanic eruptions and climate model simulations (D'Arrigo et al., 2013; Schurer et al., 2013). The IPCC AR5 (Masson-Delmotte et al., 2013) describes a discrepancy in the intensity and duration of the simulated versus proxy-based reconstructed temperature response to explosive volcanism (Fig 1b). CMIP5/PMIP3 model simulations for the last millennium experiment (Schmidt et al., 2012a) show more cooling for a shorter duration than paleoclimate reconstructions. Compounding this uncertainty, the precise timing and location of some volcanic eruptions over the last millennium remain unknown (Sigl et al., 2015; Stevenson et al., 2017) as does the magnitude of the radiative forcing (Timmreck et al., 2009). A critical question is whether this mismatch is an artifact of uncertainties in (1) the paleoclimate proxy observations, (2) the reconstruction process, (3) the forcing estimates, (4) climate model physics, or (5) a combination thereof (Anchukaitis et al., 2012; Timmreck, 2012; D'Arrigo et al., 2013; LeGrande & Anchukaitis, 2015; Stoffel et al., 2015).

Here we explore four major sources of uncertainty in reconstructions of surface air temperature over the past millennium: spatial coverage, seasonality, biological memory, and proxy noise. We do so in the context of a paleoenvironmental data assimilation (PDA) framework, the Last Millennium Reanalysis (LMR) (Hakim et al., 2016; Tardif et al., 2019), which provides an objective basis for optimally combining information from proxies and models. We show here that the discrepancy in Fig. 1b is present in our reconstruction framework (Fig. 1c), but that it can be largely reconciled by accounting for the aforementioned sources of uncertainty.

The remainder of the paper is organized as follows: we first introduce the data and methods (Section 2), then explore the causes of the discrepancy (Section 3), leveraging both real proxy and pseudoproxy experiments (PPEs). A discussion follows in Section 4.

2 Data and methods

2.1 Paleoclimate data assimilation

We apply the paleoenvironmental data assimilation framework of the Last Millennium Reanalysis (LMR) (Hakim et al., 2016; Tardif et al., 2019) to both pseudoproxy and real proxy data networks. LMR uses an offline ensemble data assimilation procedure for multivariate climate field reconstruction (Steiger et al., 2014), where information from a prior expectation of the climate, derived from a climate model, is weighted against information in proxy records. Weights are determined from the relative error in these two estimates of the climate, as defined by the update equation in the Kalman filter, which is optimal if the errors are normally distributed.

The essential components of the procedure are (1) existing climate model data for the prior expectation, which we take from a last millennium simulation from the isotope-enabled Community Earth System Model (iCESM) (Stevenson et al., 2019; Brady et al., 2019); (2) proxy data networks, which we take from the PAGES 2k phase 2 compilation (PAGES 2k Consortium, 2017, Fig. S1), and the Northern Hemisphere Tree-Ring Network Development (NTREND) compilation (Wilson, Anchukaitis, Briffa, Büntgen, et al., 2016; Anchukaitis et al., 2017, Fig. S7); and (3) a "forward operator" or proxy system model (PSM), which predicts the proxies given the climate state. Here the forward operator is a linear regression procedure, univariate on annual temperature for corals and ice cores, and seasonal, univariate on seasonal temperature for maximum latewood density records and seasonal bivariate (temperature, precipitation) for tree-ring width records, as in Tardif et al. (2019). Further details of the LMR data assimilation procedure for paleoclimate reconstruction may be found in Hakim et al. (2016).

This study utilizes a lightweight implementation of the LMR framework, `LMRt` (Zhu et al., 2019). As a benchmark, a reconstruction of the spatiotemporal variations of surface temperature over Common Era is conducted, using iCESM as the model prior and the PAGES 2k network as observations. As expected, the DA procedure yields a substantially better estimate of the temporal variability in the temperature field than the prior, as quantified by the pointwise correlation (Fig. S2c, S2d). This reconstruction skill level is comparable to a previous implementation of LMR (Tardif et al., 2019), and supported by the similarity between the reconstructed NHMT using both versions of the code (Fig. S2a). For a more in-depth evaluation of the LMR framework, see Tardif et al. (2019).

2.2 Simulated and instrumental temperature observations

In order to compare paleoclimate reconstructions to climate models, we consider simulations of past millennium climate from the following models: iCESM, as well as the PMIP3 models (Schmidt et al., 2012b; Braconnot et al., 2012), including CESM (Otto-Bliesner et al., 2015), BCC CSM1.1 (Wu et al., 2014), GISS-E2-R (Schmidt et al., 2006), HadCM3 (Gordon et al., 2000), IPSL-CM5A-LR (Dufresne et al., 2013), MIROC-ESM (Watanabe et al., 2011), MPI-ESM-P (Giorgetta et al., 2013), CSIRO (Rotstayn et al., 2012), CCSM4 (Landrum et al., 2012). For more details on each simulation, see Table S1.

We also use two sets of instrumental temperature observations, including the Berkeley Earth instrumental temperature analysis (Rohde et al., 2013) and the Goddard Institute for Space Studies (GISS) Surface Temperature Analysis (GISTEMP) (Hansen et al., 2010). GISTEMP and the gridded precipitation dataset from the Global Precipita-

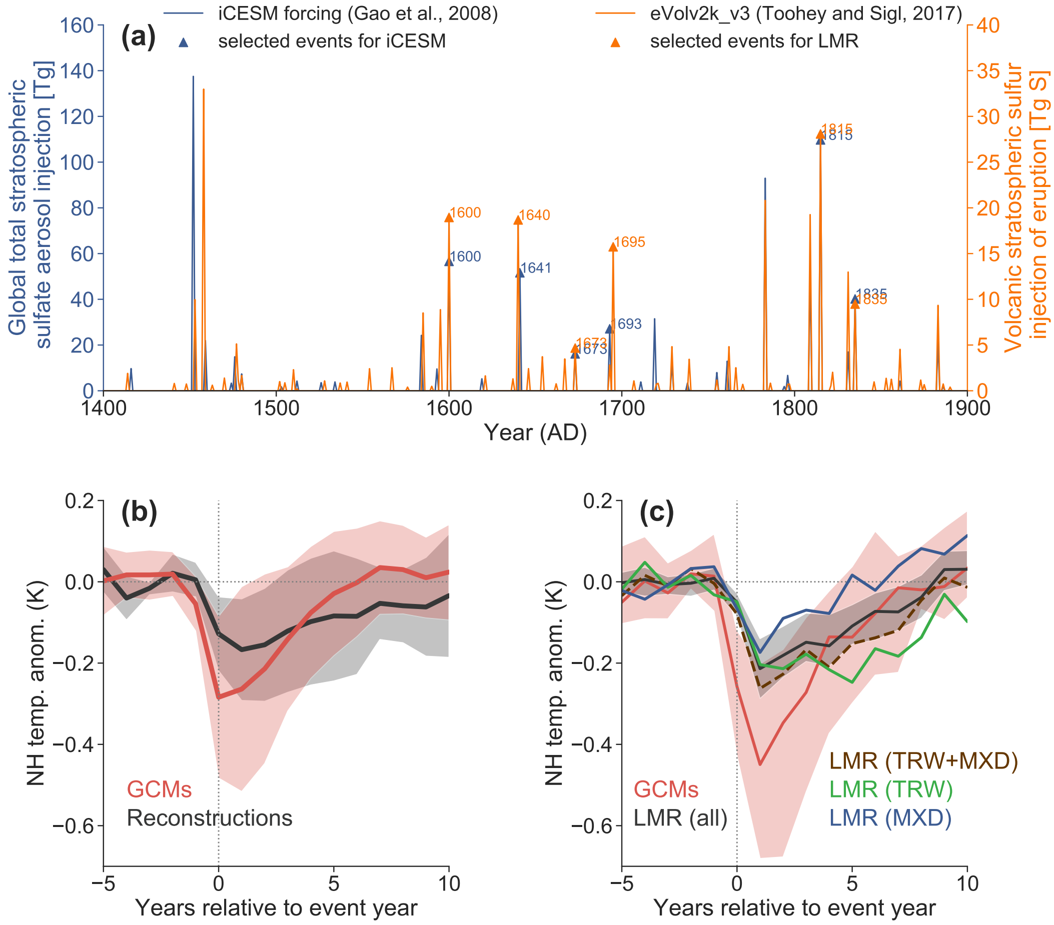

Figure 1. (a) Comparison between the volcanic forcing (Gao et al., 2008) used in the isotope-enabled Community Earth System Model (iCESM) simulation (Stevenson et al., 2019; Brady et al., 2019) and the eVolv2k version 3 Volcanic Stratospheric Sulfur Injection (VSSI) compilation (Toohey & Sigl, 2017). The triangles denote the selected 6 large events between 1400 and 1850 CE. (b) Superposed epoch analysis (SEA) on simulated and reconstructed temperature response to the 12 strongest volcanic eruptions since 1400 AD, reproduced from IPCC AR5 (Masson-Delmotte et al., 2013) Fig. 5.8b. (c) Superposed epoch analysis on annual Northern hemispheric mean temperature (NHMT) simulated by 9 GCMs Section 2.2, Table S1) and LMR reconstructions assimilating the whole network, the tree-ring network, the tree-ring width (TRW) network, and the maximum latewood density (MXD) network, respectively. The shading encompasses the 5% and 95% quantiles of the ensemble.

tion Climatology Centre (GPCC) (Schneider et al., 2014) are also used for PSM calibration in the bivariate framework of Tardif et al. (2019).

2.3 Superposed epoch analysis (SEA)

Superposed epoch analysis (SEA) (Haurwitz & Brier, 1981) is a frequently used measure of temperature response to volcanic eruptions (Adams et al., 2003; Masson-Delmotte et al., 2013; Rao et al., 2019). It consists of aligning temperature anomaly series to the timing of volcanic eruptions within a fixed time window prior to and following the event, and averaging these responses to estimate the typical response to eruptions. The IPCC AR5 (Fig. 1b) considered the reconstructed temperature response to the 12 strongest eruptions since 1400 AD. For our analysis, we selected 6 large and well-dated eruption events over the years 1400-1850 CE that are consistent in timing in both the volcanic forcing used in iCESM (Gao et al., 2008) and the most recent compilation of Volcanic Stratospheric Sulfur Injection (VSSI) (Toohey & Sigl, 2017) (Fig. 1a). For further details, see Text S3. The LMR response to individual events is shown in Fig S10.

3 Causes of the discrepancy

Fig. 1b highlights discrepancies between model simulations and reconstructions in three aspects: (1) the magnitude of the peak cooling (2) the timing of the peak cooling (3) the length of the recovery.

Specifically, model simulations show a stronger peak cooling amplitude, a slightly earlier peak cooling, and a shorter recovery interval than the reconstructions. A similar discrepancy pattern can be seen in the LMR reconstruction assimilating the PAGES 2k network (Fig. 1c). Comparing results for assimilating the PAGES 2k network as a whole (solid dark gray curve) to only its tree-ring records (dashed brown curve), we see that the volcanic signal is indeed captured by the tree-ring network alone, which consists of two main observation types: (1) tree-ring width (TRW) and (2) maximum latewood density (MXD). Assimilating these two proxy types, however, shows different responses to volcanism: TRW yields a lagged peak cooling year compared to a more prolonged recovery for MXD.

We investigate four factors that we hypothesize may account for these differences, motivated by prior studies and existing knowledge of the tree-ring proxy network: (1) spatial coverage (Anchukaitis et al., 2012; D'Arrigo et al., 2013) (2) seasonality (D'Arrigo et al., 2006; Anchukaitis et al., 2017) (3) biological memory (Fritts, 1966; Krakauer & Randerson, 2003; Frank et al., 2007; Esper et al., 2015; Zhang et al., 2015; Lücke et al., 2019), and (4) non-temperature 'noise' (von Storch et al., 2004; Riedwyl et al., 2009; Neukom et al., 2018).

3.1 Spatial coverage of tree-ring proxies

There are 336 TRW records and 59 MXD records over the Northern Hemisphere (NH) in the PAGES 2k network. MXD records in PAGES2k are mainly limited to North America and Scandinavia, while the TRW records cover both North America and Asia. Evaluating the correlation between the LMR reconstruction and the Berkeley Earth instrumental temperature analysis (Rohde et al., 2013) over the instrumental era over 1880–2000, we see that assimilating the TRW network yields a greater improvement over the model prior than assimilating the MXD network (Fig 2a, 2b). Is this difference due to the location or the quantity of each type of proxy record? To investigate this question, we use a pseudoproxy experiment (PPE) (Smerdon, 2011). We set the annual iCESM simulated temperature as our truth, and use it as model prior in the DA framework (a "perfect model" scenario). Pseudoproxies are defined as perfect temperature recorders

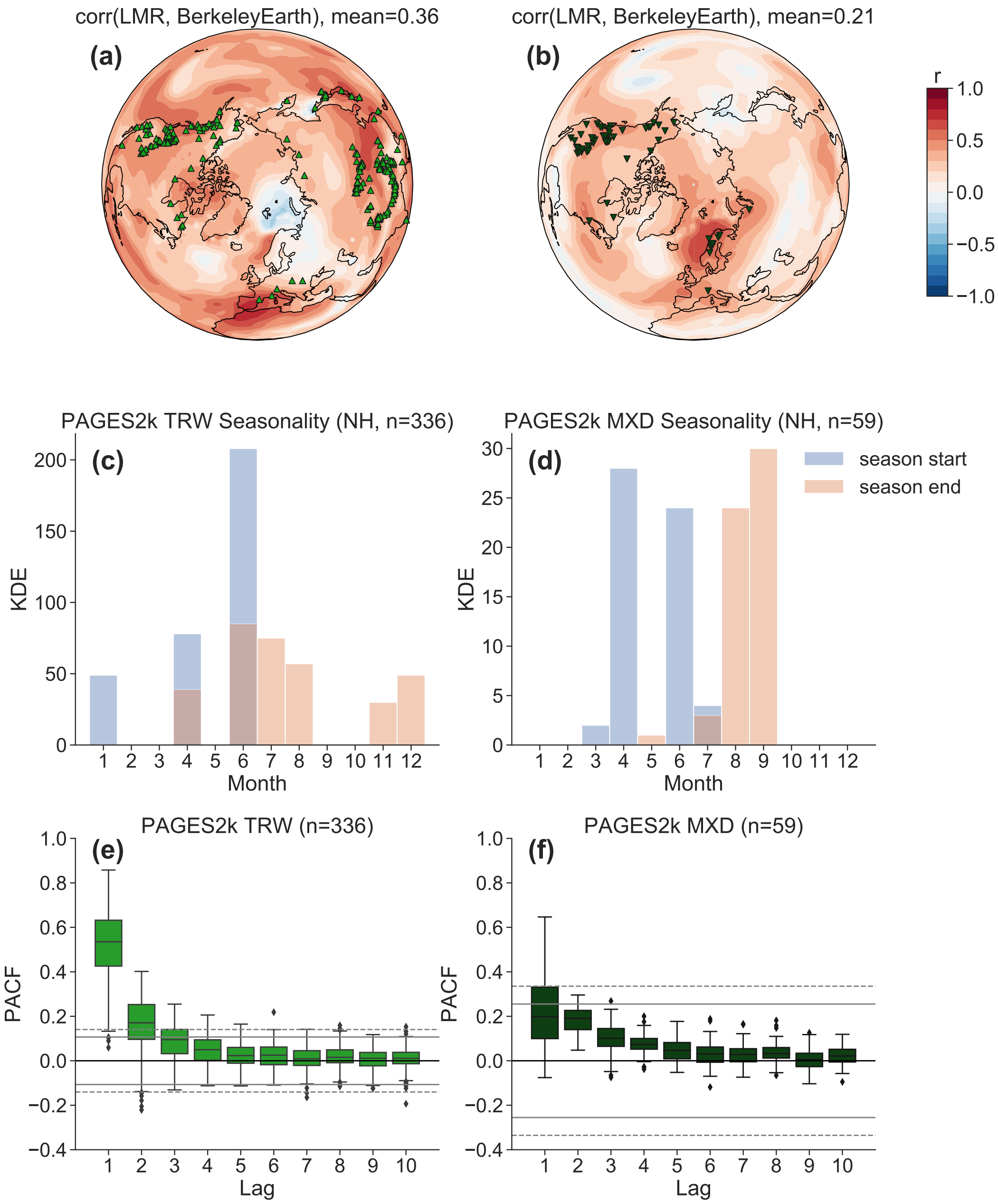

Figure 2. Differences between PAGES 2k TRW and MXD records regarding (a, b) spatial coverage, (c, d) seasonality detected by the algorithm used in Tardif et al. (2019), and (e, f) biological memory quantified by the partial autocorrelation function (PACF). (a) The spatial coverage of TRW network. The color indicates the correlation between LMR reconstruction assimilating the TRW network and the Berkeley Earth instrumental temperature analysis (Rohde et al., 2013). (c) The optimal seasonality of the TRW network. (e) The PACF of the TRW network. (b), (d), and (f) are are as (a), (b), and (e), respectively, but for the MXD network.

at three sets of locations: (1) all the 336 NH PAGES 2k TRW records (2) 50 records over North America and (3) 50 records spread out throughout the NH.

The result of assimilating these three pseudoproxy networks is shown in Fig. S3 (a, b, and c), indicating that better spatial coverage yields a more accurate reconstruction in the PDA framework, all other things being equal. This is reflected in SEA as well: Fig. 3a shows that assimilating 50 records spread throughout the NH yields a stronger and more accurate peak cooling amplitude than assimilating 50 records over North America, suggesting that a broad spatial coverage is more important than the sheer number of records for resolving peak cooling amplitude. Location does matter to some degree with regard to the large-scale teleconnection patterns, and optimal placement could be determined with the approach of Comboul et al. (2015), but this is beyond the scope of this investigation.

3.2 Seasonality

An implicit assumption in reconstructing annual temperature with tree-ring proxies is that growing season temperature is representative of annual temperature (PAGES 2k Consortium, 2017). However, the correlation between summer and annual temperatures in the Northern Hemisphere is high for the oceans but relatively low over continents (Fig. S3f), where the tree-ring records are located. Trees register climate primarily during their growing season, which varies as a function geography, species, and climate (Fritts, 1966; St. George, 2014; St George & Ault, 2014; Wilson, Anchukaitis, Briffa, Büntgen, et al., 2016). Though the PAGES 2k metadata contain some information about the seasonal sensitivity of all proxies, we follow Tardif et al. (2019) and identify optimal (in a least square sense) seasonal windows for each proxy record. The start and end month of the growing season thus identified are shown in Fig. 2c, 2d. While in the Northern Hemisphere both TRW and MXD proxies record largely boreal summer conditions, the optimal seasonality for TRW is often broader but typically less consistent than that for MXD.

As before, we use a PPE to investigate the impact of growth seasonality on the temperature reconstruction. We generate pseudoproxies as perfect recorders of local summer (JJA) temperature and we perform experiments targeting both JJA temperature and annual temperature. As expected, a much better reconstruction is obtained for the boreal summer temperature field than annual temperature (Fig. S3d, S3e). This is also evident in reconstructions using real proxies and instrumental temperature (Fig. S4). In summary, summer-sensitive trees can only reconstruct annual temperature to the extent that the summer and annual mean are correlated. While such seasonal effects result in quite different assessments of reconstruction fidelity, this difference is hardly noticeable in SEA (Fig. 3b).

3.3 Biological memory

Another important difference between TRW and MXD is biological memory, whereby tree growth reflects the influence of climate in previous years (Fritts, 1966; Krakauer & Randerson, 2003; Frank et al., 2007; Esper et al., 2015; Zhang et al., 2015). We quantify the persistence in TRW and MXD in the PAGES2k using the partial autocorrelation function (PACF) (Fig. 2e, 2f). As expected (Breitenmoser et al., 2012; Esper et al., 2015; Lücke et al., 2019), we find that biological memory in TRW across the PAGES2k network is large and significant for lag-1 and lag-2, while for MXD it is limited.

Comparing the proxy composites and the corresponding average instrumental temperature at proxy locales, we see that the MXD composite captures contemporaneous temperature variations, including the accurate timing of cooling events, while the TRW composite appears to smooth interannual variability and integrate temperatures over 2 to 5 years (Fig. S5a, S5b), leading to lagged and persistent cooling events (Frank et al., 2007).

To investigate the impact of such biological memory on the magnitude of reconstructed volcanic cooling, we again turn to PPEs. We simulate a short-term memory effect in ring width by now designing TRW pseudoproxies as a 5-yr moving average of the annual temperature simulated by iCESM, as shown in Fig. S5c. Assimilating these idealized smoothed pseudoproxies yields a prolonged temperature recovery and a peak cooling that is both damped and lagged (Fig. 3c, the solid light green curve). We find that this overall result is not sensitive to the precise design of the filter used to construct the smoothed pseudoproxies, so long as it captures this multiple year climate integration in some way.

3.4 Proxy system noise

So far, our PPEs have assumed nearly noiseless temperature recorders for simplicity (a signal-to-noise ratio (SNR) of infinity, wherein SNR is defined as the ratio of the standard deviation of signal and that of noise, following existing practice (Smerdon, 2011)). In reality, of course, proxies are imperfect recorders of climate conditions. To make the PPE more realistic, we now add uncorrelated Gaussian white noise into the previously described pseudo-TRW records. Using a linear regression procedure (Fig. S6), we estimate a SNR of 0.3, comparable to the estimate of Wang et al. (2014). Since we have already emulated the biological memory utilizing the moving average filter, we consider white noise instead of red noise to avoid adding more memory into the pseudoproxies. The addition of noise to the previous case yields a more similar SEA pattern to the real-world case (Fig. 3c, solid dark green curve): a more damped and prolonged recovery compared to the noiseless case.

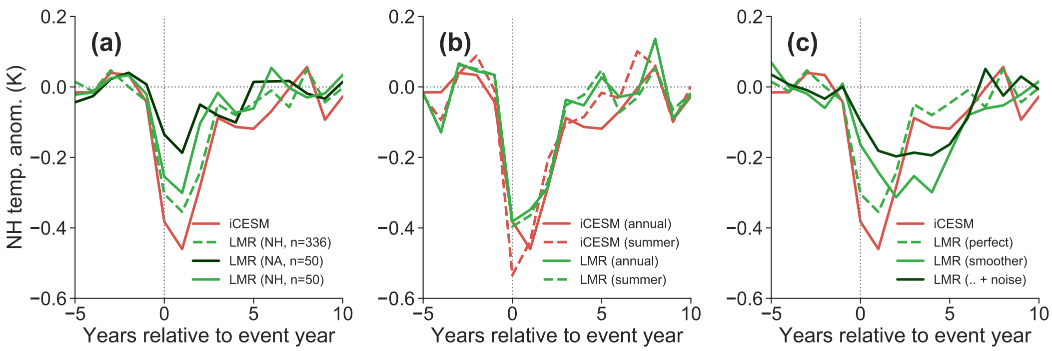

Figure 3. SEA in pseudoproxy experiments, evaluating the impact of (a) spatial coverage, (b) seasonality, and (c) biological memory and noise. (a) the red curve indicates the target, and the dashed light green curve, the solid dark green curve, and the solid light green curve indicate the LMR reconstruction assimilating 336 pseudo-TRW records over the NH, 50 records over North America, and 50 records over the NH, respectively. (b) The solid red curve denotes the annual target, the dashed red curve indicates the boreal summer target, and the green curves indicate the LMR annual and summer reconstructions, respectively. (c) The solid red curve indicates the annual target, and the green curves indicate the LMR reconstruction assimilating pseudo-TRW as perfect temperature recorders (dashed), as temperature smoothers (solid), and the temperature smoother with Gaussian noise superposed with a signal-to-noise ratio as 0.3 (dark solid).

Considering the four factors above, we are thus able to simulate the observed discrepancy between simulated and reconstructed NH temperature response to volcanic eruptions. Can this knowledge be used to minimize this discrepancy?

3.5 Reconciling the model–proxy discrepancy

In the present context, noise in proxies reflects non-temperature influence on their formation, including biophysical processes and other climate influences, which we are not able to model properly due to lack of scientific understanding, lack of site-level calibration data, or other practical concerns (e.g., computational efficiency). The other factors can, however, be corrected for: to account for the limited spatial coverage, we perform SEA at proxy locations instead of the whole NH; to minimize the seasonal bias, we target boreal summer temperature instead of annual temperature; and to mitigate memory effects, we assimilate MXD records only, leaving out TRW and mixed chronologies.

As a result, we are able to almost entirely account for the proxy–model discrepancy in Fig. 1 with the PAGES 2k network (Fig. 4a, Fig S11). The same strategy can be used for other proxy networks. For comparison, applying it to the NTREND network (Wilson, Anchukaitis, Briffa, Büntgen, et al., 2016; Anchukaitis et al., 2017) (Fig. S7) yields similar agreement between simulated and reconstructed temperature (Fig. 4b, Fig S12). These results stand in sharp contrast to results where spatial coverage, seasonality, and biological memory are not taken into account (Fig. S8).

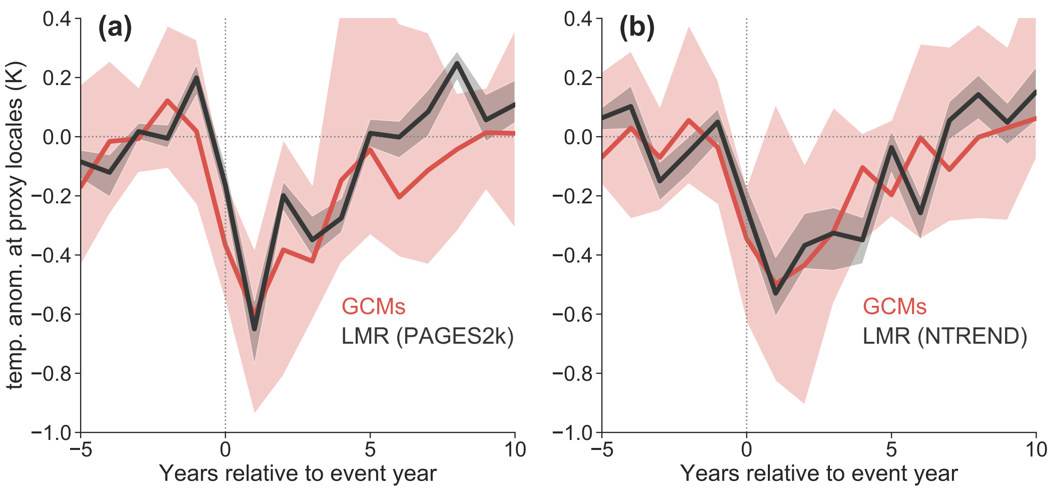

Figure 4. (a) Same as Fig. 1c, after resolving differences in the model and proxy domains associated with seasonality, spatial distribution, and biological memory. (b) Same as (a) but using the NTREND MXD network. A version of this figure showing each model simulation is available in Fig. S9, and one using more eruption events is available in Fig. S13

That the discrepancy of Fig. 1b can be largely reconciled by accounting for known characteristics of the proxy data is reassuring, and bodes well for using volcanic eruptions of the past millennium as a test bed for climate models. We now discuss the broader implications of this result.

4 Discussion

Using recent proxy compilations and climate field reconstruction techniques, we have showed that it is possible to largely resolve the discrepancy between the simulated and reconstructed temperature response to explosive volcanism since 1600 CE. We find that this gap was the result of four main factors: spatial coverage, proxy seasonality, biological memory, and proxy noise. While proxy noise is difficult to account for in model-data intercomparisons, the first three factors can be, if care is taken in evaluating comparable quantities. In particular, since our reconstructions are more reliable at locations where proxy are available than at distal locations (Anchukaitis et al., 2017), carrying out the comparison at proxy sites is a simple and effective way to reduce the mismatch. That this is true even in the data assimilation framework (Steiger et al., 2014) suggests that expanding the spatial extent of proxy network is necessary to resolve global-scale patterns.

Previous studies have argued that the simplified representation of the radiative effects of stratospheric aerosols common to CMIP5-era GCMs can produce overly strong responses to volcanic forcing (Timmreck et al., 2009; Timmreck, 2012; Stoffel et al., 2015; LeGrande et al., 2016). This is due in part to the modeled stratospheric aerosol microphysics not including the self-limiting processes known to operate at large sulfate concentrations (Timmreck et al., 2009). Uncertainties in the timing, location, and magnitude of volcanic forcing itself can also greatly affect the simulated response, independently of simplifications to the model physics. Progress in representing this forcing (Toohey & Sigl, 2017; Aubry et al., 2019), as well as improvements in model resolution and processes (e.g. active stratospheric chemistry) in PMIP4 (Kageyama et al., 2018), may lead to closer model-data matches in future work. Regardless of these factors, our analysis suggests that a critical ingredient of minimizing the model-reconstruction mismatch is to evaluate simulated temperature at the times and places where it is recorded by the proxy sensors.

Naturally, reconstructions may be improved as well. While this study has focused on the uncertainties in proxy measurements in the context of paleoenvironmental data assimilation, more work should be done to reduce sources of uncertainty within the data assimilation method itself, such as the forward operator error, the model prior, and the localization scheme, as the coupling of all these uncertainty sources can potentially affect the SEA comparison. In particular, forward operators that allow for non-contemporaneous influences of the state on the proxies (e.g. time-integration, as is believed to be the case for TRW (Vaganov et al., 2006; Anchukaitis et al., 2006)) would enable us to make better use of the information contained in TRW records. While such process-oriented models have been proposed (Tolwinski-Ward et al., 2011), their application to the DA context is contingent upon accurate specification of observation error variance and correcting for biases in the model prior. Both tasks remain active research areas (Dee et al., 2016).

With regard to proxies, we have shown that the lagged cooling exhibited in previous reconstructions can be explained as the consequence of using TRW records. Other proxies that integrate climate information over multiple years likely have a similar impact in multiproxy reconstructions. Since MXD records are more faithful paleo-temperature sensors than TRW records (Esper et al., 2015, 2018), we call here for increased collection and development of MXD records (St. George & Esper, 2019), particularly at locations where they are presently absent or cover only part of the last millennium, e.g. the North American treeline and at high elevations in Asia (Anchukaitis et al., 2017; Esper et al., 2018).

That the LMR-reconstructed boreal summer temperature at proxy locations lies within the range of simulated responses (Fig. 4) suggests that, based on this metric alone, there is no strong evidence that GCMs from the PMIP3 and CESM Last Millennium Ensemble taken as a whole systematically overestimate the climate response to explosive volcanism since 1600 CE. Nonetheless, we also find evidence that the model response is larger than the reconstructed response (Fig S11, S12). In the presence of noise, statistical reconstructions are inevitably damped because of regression dilution (Frost & Thompson, 2000; Wang et al., 2015; Neukom et al., 2018), a caveat that applies to LMR as well, to the extent that it uses regression-based forward operators to translate climate states to proxy values. Here we have mitigated this problem by focusing on the recent period with relatively high proxy coverage, but it is undoubtedly an ingredient in the mismatch observed for earlier eruptions. wi Indeed some of the largest documented discrepancies in the magnitude of peak cooling are for earlier eruptions like 1257 (Samalas) (Lavigne et al., 2013; Wilson, Anchukaitis, Briffa, Büntgen, et al., 2016), with its extremely large inferred forcing (Timmreck et al., 2009) and proportionally large response, which far exceeds what is seen in existing reconstructions (Fig. S10-S12,(Stoffel et al., 2015)). There is also lingering uncertainty as to the magnitude, timing, and location of two major events during the 1450s (Sigl et al., 2015; Toohey & Sigl, 2017; Hartman et al., 2019). These earlier and quite large eruptions register as the largest sulfate depositions of the past millennium (59.42 ± 10.86 Tg in 1258 CE and 32.98 ± 4.8 Tg in 1458 CE, according to Toohey and Sigl (2017)), but are not clearly expressed in our reconstructions (Figs. S11–12). Nonetheless, the composite changes very little when these and weaker events are included (Fig S13).

The key contribution of this work is to demonstrate that it is possible to largely resolve most of the discrepancies between simulated and reconstructed responses to volcanic cooling by explicitly accounting for the characteristics of the proxy network in this comparison. The agreement over the past 400 years, while comforting, is a rather mild test of model performance. Indeed, these eruptions, while larger than the 1991 Pinatubo eruption used to calibrate radiative forcing (Gao et al., 2008; LeGrande et al., 2016), are relatively small compared to the recorded range of the past 2,500 years (Sigl et al., 2015). Thus, parameterizations based on Pinatubo scaling between sulfate aerosols and radiative forcing would be expected to be relatively accurate for these events, though they might not be for larger eruptions. In addition, the last 400 years are more data-rich than earlier periods (Figures S1, S7), leading to relatively well-constrained reconstructions. As we explore earlier, larger eruptions, proxy attrition and its attendant bias in the spatial sampling of volcanic cooling episodes leads to wider differences between simulations and reconstructions. Only with denser observational coverage will this comparison become more informative of model performance.

Acknowledgments

The authors acknowledge support from the Climate Program Office of the National Oceanographic and Atmospheric Administration (grants NA18OAR4310426 to USC, NA18OAR4310422 to UW, and NA18OAR4310420 to UA) GJH also acknowledges support from the NSF through grant AGS0–1702423. Code and data are available at `https://github.com/fzhu2e/lmrvolc` (a placeholder for now, to be filled upon paper acceptance).

References

Adams, J., Mann, M. E., & Ammann, C. M. (2003, November). Proxy evidence for an El Niño-like response to volcanic forcing. *Nature*, *426*(6964), 274–278. doi: 10.1038/nature02101

Anchukaitis, K. J., Breitenmoser, P., Briffa, K. R., Buchwal, A., Buntgen, U., Cook, E. R., … Wilson, R. J. S. (2012, 12). Tree rings and volcanic cooling. *Nature Geosci*, *5*(12), 836–837.

Anchukaitis, K. J., Evans, M. N., Kaplan, A., Vaganov, E. A., Hughes, M. K., Grissino-Mayer, H. D., & Cane, M. A. (2006, February). Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought. *Geophys. Res. Lett.*, *33*, L04705. doi: 10.1029/2005GL025050

Anchukaitis, K. J., Wilson, R., Briffa, K. R., Büntgen, U., Cook, E. R., D'Arrigo, R., … Zorita, E. (2017, 5 1). Last millennium Northern Hemisphere summer temperatures from tree rings: Part II, spatially resolved reconstructions. *Quaternary Science Reviews*, *163*, 1–22. doi: 10.1016/j.quascirev.2017.02.020

Aubry, T. J., Toohey, M., Marshall, L., Schmidt, A., & Jellinek, A. M. (2019). A new volcanic stratospheric sulfate aerosol forcing emulator (eva_h): Comparison with interactive stratospheric aerosol models. *Journal of Geophysical Research: Atmospheres*, *n/a*(n/a). doi: 10.1029/2019JD031303

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., … Zhao, Y. (2012, June). Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*, *2*(6), 417–424. doi: 10.1038/nclimate1456

Brady, E., Stevenson, S., Bailey, D., Liu, Z., Noone, D., Nusbaumer, J., … Zhu, J. (2019). The Connected Isotopic Water Cycle in the Community Earth System Model Version 1. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2547–2566. doi: 10.1029/2019MS001663

Breitenmoser, P., Beer, J., Brönnimann, S., Frank, D., Steinhilber, F., & Wanner, H. (2012, January). Solar and volcanic fingerprints in tree-ring chronologies over the past 2000years. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *313-314*, 127–139. doi: 10.1016/j.palaeo.2011.10.014

Comboul, M., Emile-Geay, J., Hakim, G. J., & Evans, M. N. (2015). Paleoclimate sampling as a sensor placement problem. *Journal of Climate*, *28*, 7717–7740. doi: 10.1175/JCLI-D-14-00802.1

D'Arrigo, R., Wilson, R., & Anchukaitis, K. J. (2013). Volcanic cooling signal in tree ring temperature records for the past millennium. *Journal of Geophysical Research-Atmospheres*, *118*(16), 9000-9010. doi: {10.1002/jgrd.50692}

D'Arrigo, R., Wilson, R., & Jacoby, G. (2006). On the long-term context for late twentieth century warming. *Journal of Geophysical Research: Atmospheres*, *111*(D3). doi: 10.1029/2005JD006352

Dee, S. G., Steiger, N. J., Emile-Geay, J., & Hakim, G. J. (2016). On the utility of proxy system models for estimating climate states over the common era. *Journal of Advances in Modeling Earth Systems*, *8*. doi: 10.1002/2016MS000677

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., … Vuichard, N. (2013, May). Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics*, *40*(9-10), 2123–2165. doi: 10.1007/s00382-012-1636-1

Esper, J., George, S. S., Anchukaitis, K., D'Arrigo, R., Ljungqvist, F. C., Luterbacher, J., … Büntgen, U. (2018). Large-scale, millennial-length temperature reconstructions from tree-rings. *Dendrochronologia*, *50*, 81–90.

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

Frank, D., Büntgen, U., Böhm, R., Maugeri, M., & Esper, J. (2007, December). Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. *Quaternary Science Reviews*, *26*(25), 3298–3310. doi: 10.1016/j.quascirev.2007.08.002

Fritts, H. C. (1966). Growth-rings of trees: their correlation with climate. *Science*, *154*(3752), 973–979.

Frost, C., & Thompson, S. G. (2000). Correcting for regression dilution bias: comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *163*(2), 173–189. doi: 10.1111/1467-985X.00164

Gao, C., Robock, A., & Ammann, C. (2008). Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models. *Journal of Geophysical Research: Atmospheres*, *113*(D23). doi: 10.1029/2008JD010239

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., … Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597. doi: 10.1002/jame.20038

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., … Wood, R. A. (2000, February). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, *16*(2-3), 147–168. doi: 10.1007/s003820050010

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., … Perkins, W. A. (2016). The last millennium climate reanalysis project: Framework and first results. *Journal of Geophysical Research: Atmospheres*, *121*, 6745 – 6764. doi: 10.1002/2016JD024751

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Rev. Geophys.*, *48*, RG4004. doi: 10.1029/2010RG000345

Hartman, L. H., Kurbatov, A. V., Winski, D. A., Cruz-Uribe, A. M., Davies, S. M., Dunbar, N. W., … Yates, M. G. (2019). Volcanic glass properties from 1459 c.e. volcanic event in south pole ice core dismiss kuwae caldera as a potential source. *Scientific Reports*, *9*(1), 14437. doi: 10.1038/s41598-019-50939-x

Haurwitz, M. W., & Brier, G. W. (1981, October). A Critique of the Superposed Epoch Analysis Method: Its Application to Solar–Weather Relations. *Monthly Weather Review*, *109*(10), 2074–2079. doi: 10.1175/1520-0493(1981)109<2074:ACOTSE>2.0.CO;2

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., … Zhou, T. (2018, March). The PMIP4 contribution to CMIP6 – Part 1: Overview and over-arching analysis plan. *Geoscientific Model Development*, *11*(3), 1033–1057. doi: https://doi.org/10.5194/gmd-11-1033-2018

Krakauer, N. Y., & Randerson, J. T. (2003). Do volcanic eruptions enhance or diminish net primary production? Evidence from tree rings. *Global Biogeochemical Cycles*, *17*(4). doi: 10.1029/2003GB002076

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., & Teng, H. (2012, 2014/05/05). Last millennium climate and its variability in ccsm4. *Journal of Climate*, *26*(4), 1085–1111. doi: 10.1175/JCLI-D-11-00326.1

Lavigne, F., Degeai, J.-P., Komorowski, J.-C., Guillet, S., Robert, V., Lahitte, P., … de Belizal, E. (2013). Source of the great a.d. 1257 mystery eruption unveiled, samalas volcano, rinjani volcanic complex, indonesia. *Proceedings of the National Academy of Sciences*, *110*(42), 16742-16747. doi: 10.1073/pnas.1307520110

LeGrande, A. N., & Anchukaitis, K. J. (2015). Volcanic eruptions and climate. *PAGES Magazine*, *23*(2), 46–47.

LeGrande, A. N., Tsigaridis, K., & Bauer, S. E. (2016, September). Role of atmospheric chemistry in the climate impacts of stratospheric volcanic injections. *Nature Geoscience*, *9*(9), 652–655. doi: 10.1038/ngeo2771

Lücke, L., Hegerl, G., Schurer, A., & Wilson, R. (2019, September). Effects of memory biases on variability of temperature reconstructions. *Journal of Climate*. doi: 10.1175/JCLI-D-19-0184.1

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Rouco, J. G., … Timmermann, A. (2013). Information from Paleoclimate Archives. In T. F. Stocker et al. (Eds.), *Climate Change 2013: The Physical Science*

Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 383–464). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/CBO9781107415324.013

Neukom, R., Schurer, A. P., Steiger, N. J., & Hegerl, G. C. (2018, May). Possible causes of data model discrepancy in the temperature history of the last Millennium. *Scientific Reports*, *8*(1), 1–15. doi: 10.1038/s41598-018-25862-2

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., … Strand, G. (2015). Climate variability and change since 850 CE: An ensemble approach with the community earth system model. *Bull. Amer. Meteor. Soc.*, *97*(5), 735–754. doi: 10.1175/BAMS-D-14-00233.1

PAGES 2k Consortium. (2017, 07). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*, 170088 EP. doi: 10.1038/sdata.2017.88

Rao, M. P., Cook, E. R., Cook, B. I., Anchukaitis, K. J., D'Arrigo, R. D., Krusic, P. J., & LeGrande, A. N. (2019). A double bootstrap approach to Superposed Epoch Analysis to evaluate response uncertainty. *Dendrochronologia*, *55*, 119–124.

Riedwyl, N., Küttel, M., Luterbacher, J., & Wanner, H. (2009). Comparison of climate field reconstruction techniques: Application to Europe. *Climate Dynamics*, *32*(2-3), 381–395.

Robock, A. (2000). Volcanic eruptions and climate. *Rev. Geophys.*, *38*, 191-220. doi: 10.1029/1998RG000054

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., … Mosher, S. (2013). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, *2013*. doi: 10.4172/2327-4581.1000103

Rotstayn, L. D., Jeffrey, S. J., Collier, M. A., Dravitzki, S. M., Hirst, A. C., Syktus, J. I., & Wong, K. K. (2012, July). Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: a study using single-forcing climate simulations. *Atmos. Chem. Phys.*, *12*(14), 6377–6404. doi: 10.5194/acp-12-6377-2012

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., … Vieira, L. E. A. (2012a). Climate forcing reconstructions for use in pmip simulations of the last millennium (v1.1). *Geoscientific Model Development*, *5*(1), 185–191. doi: 10.5194/gmd-5-185-2012

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., … Vieira, L. E. A. (2012b, January). Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1). *Geosci. Model Dev.*, *5*(1), 185–191. doi: 10.5194/gmd-5-185-2012

Schmidt, G. A., Ruedy, R., Hansen, J. E., Aleinov, I., Bell, N., Bauer, M., … Yao, M.-S. (2006, January). Present-Day Atmospheric Simulations Using GISS ModelE: Comparison to In Situ, Satellite, and Reanalysis Data. *Journal of Climate*, *19*(2), 153–192. doi: 10.1175/JCLI3612.1

Schneider, D. P., Ammann, C. M., Otto-Bliesner, B. L., & Kaufman, D. S. (2009). Climate response to large, high-latitude and low-latitude volcanic eruptions in the Community Climate System Model. *Journal of Geophysical Research: Atmospheres*, *114*(D15). doi: 10.1029/2008JD011222

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., & Rudolf, B. (2014, January). GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, *115*(1), 15–40. doi: 10.1007/s00704-013-0860-x

Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F., & Phipps, S. J. (2013). Separating forced from chaotic climate variability over the past millennium. *Journal of Climate*, *26*(18), 6954–6973.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., … Woodruff, T. E. (2015, 07 30). Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*, *523*(7562), 543–549.

Smerdon, J. E. (2011). Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. *WIREs Clim Change.* doi: 10.1002/wcc.149

Soden, B. J., Wetherald, R. T., Stenchikov, G. L., & Robock, A. (2002). Global Cooling After the Eruption of Mount Pinatubo: A Test of Climate Feedback by Water Vapor. *Science*, *296*(5568), 727-730.

St. George, S. (2014). An overview of tree-ring width records across the Northern Hemisphere. *Quaternary Science Reviews*, *95*, 132–150. doi: 10.1016/j.quascirev.2014.04.029

St George, S., & Ault, T. R. (2014). The imprint of climate within Northern Hemisphere trees. *Quaternary Science Reviews*, *89*, 1–4.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., & Roe, G. H. (2014, 2014/04/08). Assimilation of time-averaged pseudoproxies for climate reconstruction. *Journal of Climate*, *27*(1), 426–441. doi: 10.1175/JCLI-D-12-00693.1

Stevenson, S., Fasullo, J. T., Otto-Bliesner, B. L., Tomas, R. A., & Gao, C. (2017). Role of eruption season in reconciling model and proxy responses to tropical volcanism. *Proceedings of the National Academy of Sciences*, *114*(8), 1822–1826. doi: 10.1073/pnas.1612505114

Stevenson, S., Otto-Bliesner, B. L., Brady, E. C., Nusbaumer, J., Tabor, C., Tomas, R., … Liu, Z. (2019). Volcanic Eruption Signatures in the Isotope-Enabled Last Millennium Ensemble. *Paleoceanography and Paleoclimatology*, *0*(0). doi: 10.1029/2019PA003625

St. George, S., & Esper, J. (2019, January). Concord and discord among Northern Hemisphere paleotemperature reconstructions from tree rings. *Quaternary Science Reviews*, *203*, 278–281. doi: 10.1016/j.quascirev.2018.11.013

Stoffel, M., Khodri, M., Corona, C., Guillet, S., Poulain, V., Bekki, S., … Masson-Delmotte, V. (2015, October). Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1,500 years. *Nature Geoscience*, *8*(10), 784–788. doi: 10.1038/ngeo2526

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., … Noone, D. (2019, July). Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Climate of the Past*, *15*(4), 1251–1273. doi: https://doi.org/10.5194/cp-15-1251-2019

Timmreck, C. (2012). Modeling the climatic effects of large explosive volcanic eruptions. *Wiley Interdisciplinary Reviews: Climate Change*, *3*(6), 545–564. doi: 10.1002/wcc.192

Timmreck, C., Lorenz, S. J., Crowley, T. J., Kinne, S., Raddatz, T. J., Thomas, M. A., & Jungclaus, J. H. (2009). Limited temperature response to the very large AD 1258 volcanic eruption. *Geophysical Research Letters*, *36*(21). doi: 10.1029/2009GL040083

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., & Anchukaitis, K. J. (2011). An efficient forward model of the climate controls on interannual variation in tree-ring width. *Climate Dynamics*, *36*(11-12), 2419-2439. doi: 10.1007/s00382-010-0945-5

Toohey, M., & Sigl, M. (2017, November). Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE. *Earth System Science Data*, *9*(2), 809–831. doi: https://doi.org/10.5194/essd-9-809-2017

Vaganov, E. A., Hughes, M. K., & Shashkin, A. V. (2006). *Growth dynamics of conifer tree rings* (Vol. 183). New York, NY: Springer-Verlag.

von Storch, H., Zorita, E., Jones, J. M., Dimitriev, Y., González-Rouco, F., & Tett, S. F. B. (2004, October). Reconstructing Past Climate from Noisy Data.

Science, *306*, 679-682. doi: 10.1126/science.1096109

Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., & Rajaratnam, B. (2015). Fragility of reconstructed temperature patterns over the common era: Implications for model evaluation. *Geophysical Research Letters*, *42*, 7162–7170. doi: 10.1002/2015GL065265

Wang, J., Emile-Geay, J., Guillot, D., Smerdon, J. E., & Rajaratnam, B. (2014). Evaluating climate field reconstruction techniques using improved emulations of real-world conditions. *Climate of the Past*, *10*(1), 1–19. doi: 10.5194/cp-10-1-2014

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., … Kawamiya, M. (2011, January). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*(4), 845–872. doi: 10.5194/gmd-4-845-2011

Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'Arrigo, R., … Zorita, E. (2016, February). Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context. *Quaternary Science Reviews*, *134*, 1–18. doi: 10.1016/j.quascirev.2015.12.005

Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'Arrigo, R., … Zorita, E. (2016, 2 15). Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context. *Quaternary Science Reviews*, *134*, 1–18. doi: 10.1016/j.quascirev.2015.12.005

Wu, T., Song, L., Li, W., Wang, Z., Zhang, H., Xin, X., … Zhou, M. (2014, February). An overview of BCC climate system model development and application for climate change studies. *Journal of Meteorological Research*, *28*(1), 34–56. doi: 10.1007/s13351-014-3041-7

Zhang, H., Yuan, N., Esper, J., Werner, J. P., Xoplaki, E., Büntgen, U., … Luterbacher, J. (2015, August). Modified climate with long term memory in tree ring proxies. *Environmental Research Letters*, *10*(8), 084020. doi: 10.1088/1748-9326/10/8/084020

Zhu, F., Emile-Geay, J., Hakim, G. J., Tardif, R., & Perkins, A. (2019, December). *LMR Turbo (LMRt): a lightweight implementation of the LMR framework.* Zenodo. doi: 10.5281/zenodo.3590258