Discrepancies between the modeled and proxy-reconstructed response to volcanic forcing over the past millennium: Implications and possible mechanisms

Michael E. Mann,[1] Scott Rutherford,[2] Andrew Schurer,[3] Simon F.B. Tett,[3] and Jose D. Fuentes[4]

[1]   We show that a systematic discrepancy between model simulations and proxy reconstructions of hemispheric temperature changes over the past millennium appears to arise from a small number of radiatively large volcanic eruptions. Past work has shown that accounting for this mismatch alone appears to reconcile inconsistencies between the overall amplitude of simulated and proxy-reconstructed temperature changes. We provide empirical support for the previously posited hypothesis that this discrepancy may arise from threshold growth effects in tree line-proximal trees that limit their response to large volcanic cooling events. Such threshold responses could lead to an absence of growth rings (as many as six accumulated years over the past eight centuries) for some fraction of tree line-proximal trees, leading to a potential misalignment of volcanic cooling events in trees from climatically distinct regions, and a further attenuation and smearing of the volcanic cooling signal. Since the high-frequency component of nearly all proxy reconstructions of past hemispheric temperature change is derived from tree ring data, this bias would likely impact nearly all such reconstructions. We show here that the discrepancy may have led to an underestimation bias in past studies attempting to infer equilibrium climate sensitivity from proxy temperature reconstructions of the past millennium.

1.   Introduction

[2]   A body of past work [*Robock*, 2005; *Timmreck et al.*, 2009; *Mann et al.*, 2012a—henceforth MFR12] has noted, and sought to explain, discrepancies between the modeled hemispheric cooling response to the largest few volcanic eruptions over the past millennium and proxy reconstructions that suggest modest or, in some cases, no response to the estimated volcanic forcing. As with all model/data mismatches, the origins of the discrepancy could potentially lie in the reconstructions, the model estimates, or both. Since this mismatch has wider implications for our understanding of how the climate responds to substantial radiative forcing,

including the assessment of equilibrium climate sensitivity (ECS), it is important to investigate and understand both the reason for, and wider implications of the model/ data discrepancy.

[3]   One possibility is that the intrinsic noise of the proxy data might lead to an underestimation of the volcanic cooling signals. MFR12 investigated this hypothesis using networks of synthetic proxy ("pseudoproxy") records with the same hemispheric extent and signal-versus-noise properties as those estimated for the actual proxy data. They found that, while the volcanic cooling signals were often difficult to discern in the individual synthetic proxy records because of the obscuring effects of both proxy noise and regional climatic noise, the tendency for noise cancellation in a large-scale average led to faithful estimates of the volcanic cooling signal in the hemispheric composite.

[4]   Another possible explanation for the mismatch is that existing volcanic radiative forcing estimates (or how they are implemented in climate model simulations) have overestimated the amplitude of forcing and consequent resulting cooling. Sensitivity tests by MFR12 using Energy Balance Model (EBM) simulations, driven by the various published volcanic radiative forcing estimates, and spanning the plausible range of climate sensitivity, show some variation in the predicted cooling, but in all cases, the AD 1258 and AD 1816 volcanic cooling signals are predicted to be

---

substantially larger than inferred from the proxy reconstructions. There is nonetheless the possibility of systematic biases in estimates of volcanic forcing [*Hegerl et al.*, 2006], and complications related to the uncertainty in particle size distributions may play an especially large role for larger eruptions such as the AD 1258 tropical eruption [e.g., *Timmreck et al.*, 2009].

[5] Another, not mutually exclusive possibility is that the reconstructions are biased in a way that underestimates the past responses to the largest volcanic eruptions. *Robock* [2005] proposed that favorable growth conditions related to increased diffuse surface radiation following volcanic eruptions could cause tree ring data, which provide most of high-frequency information in large-scale proxy temperature reconstructions, to underestimate the cooling signals associated with large eruptions. MFR12 explored the *Robock* [2005] mechanism further, finding that its impact is likely modest.

[6] MFR12 argued instead that other possible biological effects might limit the cooling recorded by tree ring temperature reconstructions. They concluded that the use of trees from tree line-proximal locations (thus, at or near the limit of their temperature range) potentially limits the recorded cooling to no more than ~1°C below the late 20th century baseline. For this reason, large volcanic eruptions (for which models predict greater cooling) might place trees below the summer threshold for growth in certain regions. This condition potentially leads to both (a) a loss of temperature sensitivity to further cooling and (b) missing rings and consequent chronological errors, e.g., the ring from a previous growing season (i.e., warmer year) such as AD 1815 might masquerade in some regions for the missing AD 1816 ring. Averaging of chronologies from different regions that accumulate differing chronological errors back in time could thus lead to a smearing, and further loss of amplitude, of the expected cooling signal in hemispheric composites. This hypothesis, as discussed later in more depth, is now being debated vigorously in the peer-reviewed literature [*Anchukaitas et al.*, 2012; *Mann et al.*, 2012b].

[7] Regardless of the precise reason for the model/data mismatch, the impact of the discrepancies, e.g., their implications for past estimates of ECS, may be significant. Here we investigate these issues further. In section 2, we examine more closely the hypothesis that biological effects related to tree growth might be responsible for an underestimation of volcanic cooling. We demonstrate, using the underlying tree ring data, that an apparent misalignment of larger volcanic cooling signals in distinct regions due to hypothesized chronological errors can account for the damped cooling found in hemispheric tree ring composites. We also demonstrate that the signal underestimation problem may extend beyond tree ring growth thickness data, to tree ring density records as well. In section 3, we compare the responses to natural forcing over the past millennium as inferred from reconstructions and as indicated by state-of-the-art (CMIP5) climate model simulations subject to estimated radiative forcing changes. We show that the reconstructed signals are smaller than predicted by the models but, that by simply accounting for the mismatch between the predicted and observed responses to the few largest eruptions of the past millennium through an assumed maximum cooling response threshold—regardless of cause—in the reconstructions, we are able to reconcile

the amplitude of modeled and proxy-reconstructed temperature changes. We furthermore show that the mismatch may have led to downward-biased estimates of ECS in past work. In section 4, we summarize with our key conclusions.

2. Potential Misaligned Volcanic Cooling in Tree Ring Series

[8] MFR12 demonstrated that a numerical tree-growth model, forced by climate model-simulated temperature histories, predicts a substantial underestimation, smearing, and delay of the cooling response to the few largest tropical volcanic eruptions including the AD 1258 eruption (location unknown) and the AD 1815 Tambora eruption. The predicted features closely match those of an actual tree ring-based Northern Hemisphere temperature composite [*D'Arrigo et al.*, 2006; henceforth "D06"]. Since then, a vigorous debate has arisen about the viability of the hypothesis, particularly with regard to the possibility that chronological errors associated with missing rings may degrade the signals recorded in tree ring composites [*Anchukaitas et al.*, 2012; *Mann et al.*, 2012b]. While *Anchukaitas et al.* [2012] criticize certain details of the forward model formulation used by MFR12, *Mann et al.* [2012b] in their response show that the same results are obtained using a simple growing-degree day formulation using a growth threshold temperature ($T_{min} = 10°C$) consistent with (though toward the upper end) of the range reported in the published literature.

[9] Tree ring researchers [*Anchukaitas et al.*, 2012; and response by *Mann et al.*, 2012b] have criticized MFR12 for failing to provide empirical support for the existence of the hypothesized "missing rings." It is well known and accepted by dendrochronologists that adverse environmental conditions (temperature or drought) can cause at least some trees at some locations to miss an annual growth ring, hence the use of cross dating. There must, however, be some threshold at which all trees in a particular region will not exhibit a growth ring. Consider the extreme case of a year in which the growing season mean temperature at tree line is well below 0°C, e.g., $-10°C$. Even the most conservative growth threshold assumptions [see, e.g., MFR12 for a discussion] would indicate an absence of any growth for tree line-proximal trees that year. Some of the trees may in fact die when subject to these conditions, but if so they are not among those that will in the future be sampled by dendroclimatologists anyway. MFR12 hypothesize that regional cooling from large volcanic eruptions is one possible mechanism that may result in a large-scale pattern of missing rings, but the fundamental hypothesis—that misaligned chronologies will result in a delay, smearing, and underestimation of volcanic cooling—permits any combination of conditions that might cause widespread (i.e., site wide and regional scale) missing rings.

[10] Despite claims to the contrary [e.g., *Anchukaitas et al.*, 2012], it is unclear that any amount of regional cross dating, examination under microscope, or other empirical technique can unequivocally identify a regionally widespread pattern of missing rings resulting from a large-scale climatic anomaly that, as predicted by MFR12, places all trees in the region outside their growth-permitting range. The fundamental challenge is that one cannot identify what simply is not there. One might imagine that a comparison of trees along a transect
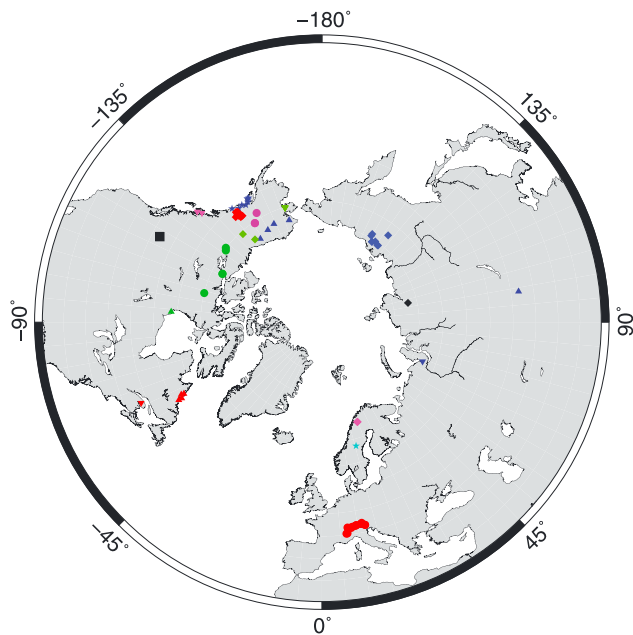
Figure 1. Locations of D06 tree ring records (individual sites and regional composites) used in this study. Each symbol type and color represent the sites included in one of the 19 regional composites.

directed away from the (boreal or alpine) tree line could establish the existence of rings in the tree line-distal locations that are missing at the tree line-proximal locations. The paradox here, however, is that the very feature—sensitivity to a limiting climate conditions—that produces a recognizable, spatially coherent, annual sequence of ring width variation (the reason indeed why tree line-proximal environments are employed by dendroclimatologists) is lost as one moves away from the tree line. Detection, empirically, of a regional-scale pattern of missing rings in tree line-proximal chronologies requires a more nuanced approach.

[11] Here we employ such an approach, using the actual tree ring data used in the D06 dendroclimatic temperature reconstruction. We demonstrate that the apparent temporal misalignment of larger regional volcanic cooling signals in the underlying tree ring data suggests the partial cancellation of what would potentially be considerably larger hemispheric volcanic cooling signals. We provide evidence for a substantially greater cooling in hemispheric composites when the estimated chronological errors are accounted for, and we show that such enhancement of the composite cooling signal is extremely unlikely to have arisen by chance, i.e., by random—rather than systematic—temporal misalignments (all code and data used in this section are available at: http://www. meteo.psu.edu/~mann/supplements/TreeVolcano13/).

2.1. Tree Ring Data

[12] For the purpose of the forgoing analysis, we used the regional tree ring annual growth thickness records of D06, the composite of which forms the hemispheric mean series analyzed by MFR12. These data are archived at the National Geophysical Data Center (http:// www.ncdc.noaa.gov/paleo/pubs/darrigo2006/darrigo2006. html). The regional series are based on a larger number

of 66 individual tree line-proximal sites across North American and Eurasia. In some cases, an individual site represents an entire region (e.g., Polar Urals), while in other cases multiple sites are incorporated into a regional average (e.g., Alps). These data consist of a maximum of 66 distinct site chronologies representing 19 different regions back to 1686, decreasing to eight regions back to AD 1190. The distribution of the tree ring series is shown in Figure 1 [see D06 for further details]. We used the conventionally standardized (STD) series of D06, but qualitatively similar conclusions are obtained using the alternative Regional Curve Standardization ("RCS") versions of the series (see Supplementary Information, henceforth "SI").

2.2. Monte Carlo Procedure

[13] We performed Monte Carlo simulations using estimates of the probabilities for a missing ring in a given year, yielding an ensemble of alternative realizations of the D06 tree ring series consistent with estimated chronological error estimates. Other than the timing of when rings are likely to be missing, which was taken from the MFR12 predictions, the analysis procedure is in all respects entirely independent of the MFR12 forward modeling exercises.

[14] Though local cross dating of trees can be used to identify missing rings in individual cores contributing to local chronologies developed from nearby trees [*Anchukaitis et al*., 2012], it cannot, as discussed above, reliably identify a coherent large-scale pattern of missing rings across an entire climatic region experiencing subgrowth limit summer temperatures, as MFR12 predict to be the case following the largest few tropical volcanic eruptions. We therefore proceeded to generate an ensemble of alternative regional composites (rather than individual cores or chronologies) consistent with estimated chronological errors (e.g., 90% of available series are missing the AD 1258 ring, while 55% of the available series are missing the AD 1816 ring—note that our net estimated age model errors amount to <1% error, i.e., no more than 6 years out of 700+). By fixing the percentage of missing rings rather than using a probability of missing rings, we control for sample size effects and can directly assess the impact of small sample sizes (see, e.g., Figures S1–S3).

[15] For each realization, we proceeded as follows. First, for each of 19 regional time series, we randomly inserted "empty" tree rings into a percentage of the available series for a given year as shown in Table 1, assuming all regional time series are independent. Where missing rings were predicted for consecutive years, they were assigned to the same series. For example, a random 50% of the eight series

Table 1. Percentage of Missing Years Projected by MFR12

Year	% Missing
1258	90
1259	50
1452	5
1453	5
1810	5
1811	5
1815	20
1816	55
1817	30
1818	10
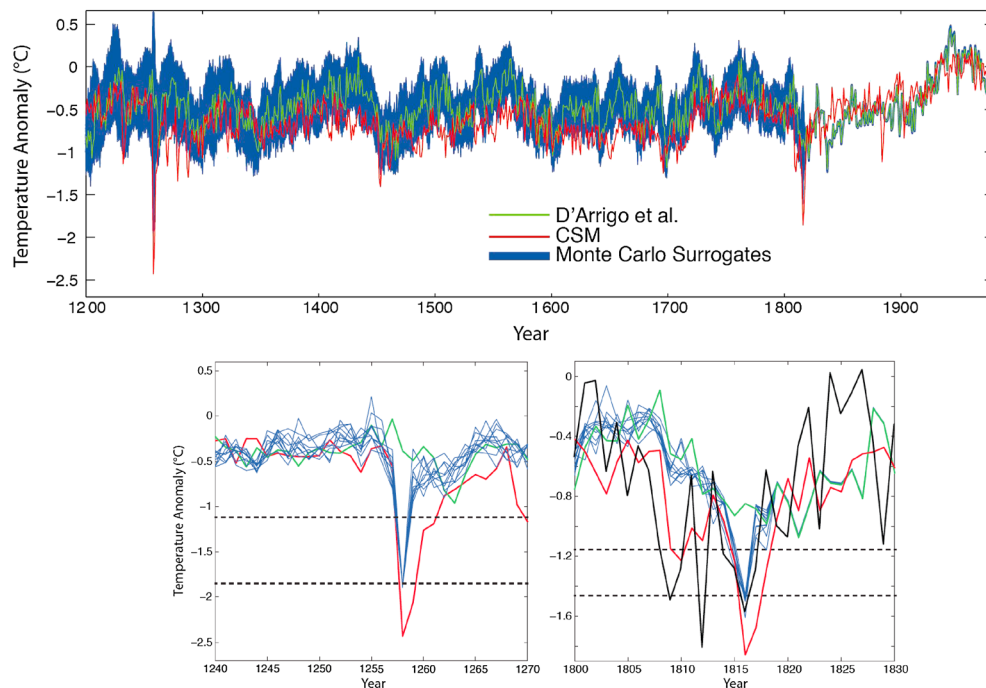
Figure 2. Ensemble of hemispheric tree ring temperature reconstructions derived from available regional tree ring composites resampled to account for predicted age model errors. Shown are the raw composite based on the *D'Arrigo et al.* [2006] tree ring data (green), Monte Carlo surrogate reconstructions (8000 in total—blue curves), and GCM simulation (red). Insets: Expanded views of the response to the AD 1258/1259 and AD 1815 eruptions responses showing the 10 coldest surrogates (blue) for each eruptions and the 2 and 4 sigma significance thresholds for cooling (dashed black). Shown also for AD 1815 eruption is the recently back-extended instrumental Northern Hemisphere land temperature record of *Rohde et al.* [2012] (black). Centering of all series is based on a 1961–1990 modern base period.

available at 1258 missed both AD 1258 and AD 1259 and an additional 40% of the available series missed AD 1258 for a total of 90% of the series missing AD 1258. Second, these regional time series were averaged into continental and then hemispheric means and scaled against the published D06 RCS reconstruction from 1686 to 1978. Where missing rings were inserted, the value for that year and series was set to "missing" and was ignored in the averaging process. As in D06, a nested reconstruction method was used to stabilize the variance with a new nest being created each time the number of available regions declines. As a check on our methodology, we confirmed that we were able to reproduce, with very minor differences, the actual D06 STD and RCS hemispheric reconstructions applying these procedures to the unperturbed regional series (see Figure S4). Northern Hemisphere average series were then computed by processing the regional time series using the same procedure as D06. Each resulting series is then an estimate of the Northern Hemisphere mean temperatures from AD 1190 to AD 1978 (note that the D06 reconstruction extends to AD 747, but our interests only begin with the 1258 eruption). We generated 8000 realizations of these Northern Hemisphere mean temperature series.

[16] The MFR12 prediction of 90% likelihood of missing rings in AD 1258 leads to only one of the eight regional composites contributing to the hemispheric composite in that year. To avoid potential artifacts associated with a unit sample size, we instead imposed a lower, 80% probability of missing rings, which leads to two retained regions (see SI

for further details). Results are nonetheless robust if we restrict the procedure to retain 1, 2, 3, or even 4 (i.e., only 50% missing) regional composites for AD 1258 (see SI).

2.3. Significance Estimation

[17] Randomly removing values for a given year and series will always have an impact on the mean reconstruction for that year and all preceding years (ring-counting proceeding from youngest to oldest). However, if there are no missing rings in reality, then the new alignment would produce surrogate reconstructions that are simply re-aligned noise or re-aligned spatially variable temperatures from adjacent years. Producing a distribution of surrogates through this process thus forms the basis for a null distribution.

[18] If there is, as hypothesized, variable misalignment of volcanic cooling signals among the various regional tree ring series that make up the hemispheric composite, we can think of the actual (D06) hemispheric tree ring series as simply one realization of a large number of potential realizations with random differing misalignments. The larger cooling signal hypothesized will not be evident in the ensemble mean of these realizations, but it should be evident in some subset of realizations, which happen to match the actual dating errors and when accounted for through the resampling process, thus, bringing the misaligned volcanic cooling signals into alignment. The question is whether such signals arise in the Monte Carlo process more often than would be expected from chance alone. This is what we test for in the significance estimation procedure
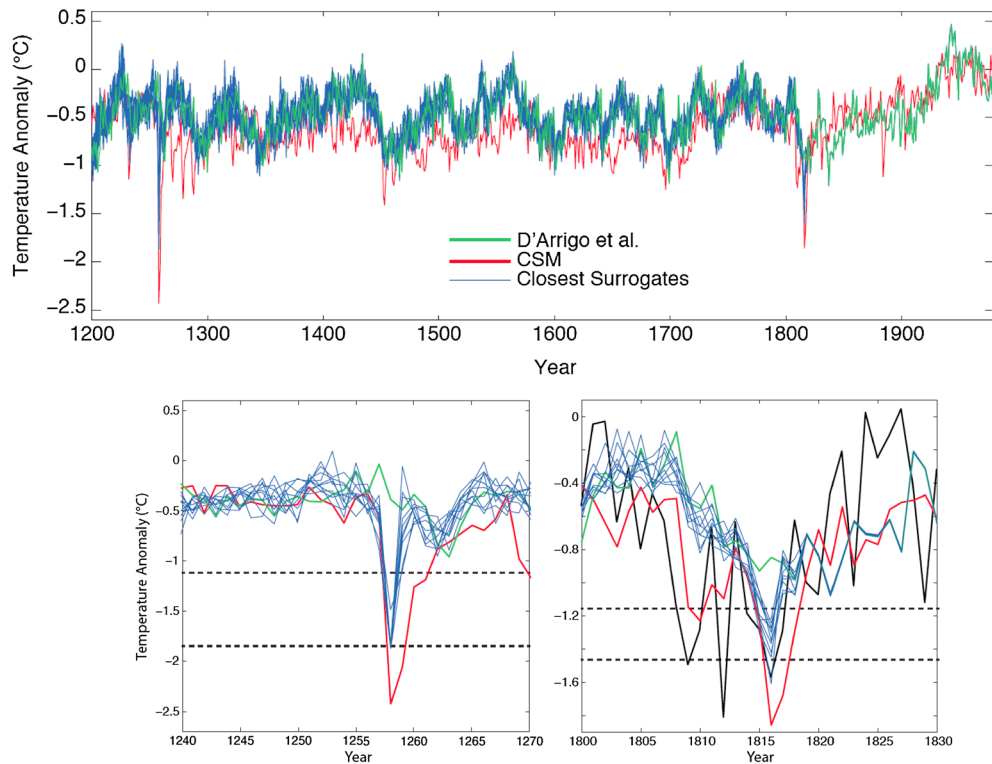
Figure 3. Ensemble of hemispheric tree ring temperature reconstructions derived from available regional tree ring composites resampled to account for predicted age model errors as in Figure 2, but with only the 10 individual surrogates (blue curves) exhibiting the closest overall match to the CSM model simulation shown (both main plot and insets). Other conventions, e.g., color scheme for curves and depiction of significance levels, are as in Figure 2.

[19] Significance estimates were developed based on a null distribution derived from "random eruption" surrogates created as follows: Missing "rings" were applied in the correct percentages (Table 1) to random "eruption" years that are, in fact, known *not to be* eruption years (and do not immediately neighbor known eruption years). The selection of the random "eruption" year was constrained so that the same number of regions was available as with the real eruption year in question (e.g., those random years representing 1258 could only be drawn from 1200 to 1339) to preserve spatial sampling statistics. As with the real eruption case, 8000 Monte Carlo surrogates were produced, and the procedure was repeated with 40 sets of random eruption years. This procedure yields $40 \times 8000 = 320,000$ random hemispheric composites for each event of interest (i.e., AD 1258 and AD 1816). From these random composites, a null distribution is built up of the range of cooling that might be expected for the event (e.g., AD 1258 eruption or AD 1816 eruption) by chance, due simply to the random sampling statistics of the Monte Carlo procedure. From this null distribution, the significance of the actual peak cooling produced from the actual Monte Carlo surrogates can be assessed (Table S1).

2.4. Results

[20] As shown in Figure 2, many of the surrogate reconstructions conform far more closely to the model-predicted temperatures than does the raw unaltered reconstruction. For the AD 1258 eruption, a large number of Monte Carlo

surrogates point toward a distinct ~2°C cooling in AD 1258 (lacking the enigmatic delayed and reduced 1260–62 cooling signal seen in the raw reconstruction). The year AD 1816 now lives up to its billing as the "Year Without A Summer," with surrogates showing cooling of up to ~1.6° C, remarkably consistent with a newly available extension of the instrumental Northern Hemisphere average temperature record [*Rohde et al.*, 2012] and closer to the model-predicted cooling.

[21] If the hypothesis of chronological errors induced by missing rings is correct, then it should be possible to observe enhanced volcanic cooling signals for both major eruptions (AD 1258 and AD 1815) simultaneously in the same individual surrogate series. That is indeed found to be the case. Figure 3 shows the 10 surrogates within the ensemble that most closely match the CSM simulation responses. Each of these surrogates shows cooling for both eruptions simultaneously that is comparable to the largest cooling found among all surrogates for either eruption (i.e., compare Figure 2).

[22] These enhanced cooling responses are highly significant relative to the null hypothesis of chance occurrence due to random sampling variations from the Monte Carlo procedure. The peak AD 1258 cooling from the Monte Carlo surrogates (−1.9°C) breaches the 4.1σ cooling limit derived from the null distribution, corresponding to a 1 in 100,000 event, something highly unlikely to occur by chance alone given a sample of $N = 8000$ surrogates. The peak AD 1816 cooling (−1.6°C) breaches the 4.9σ cooling limit and
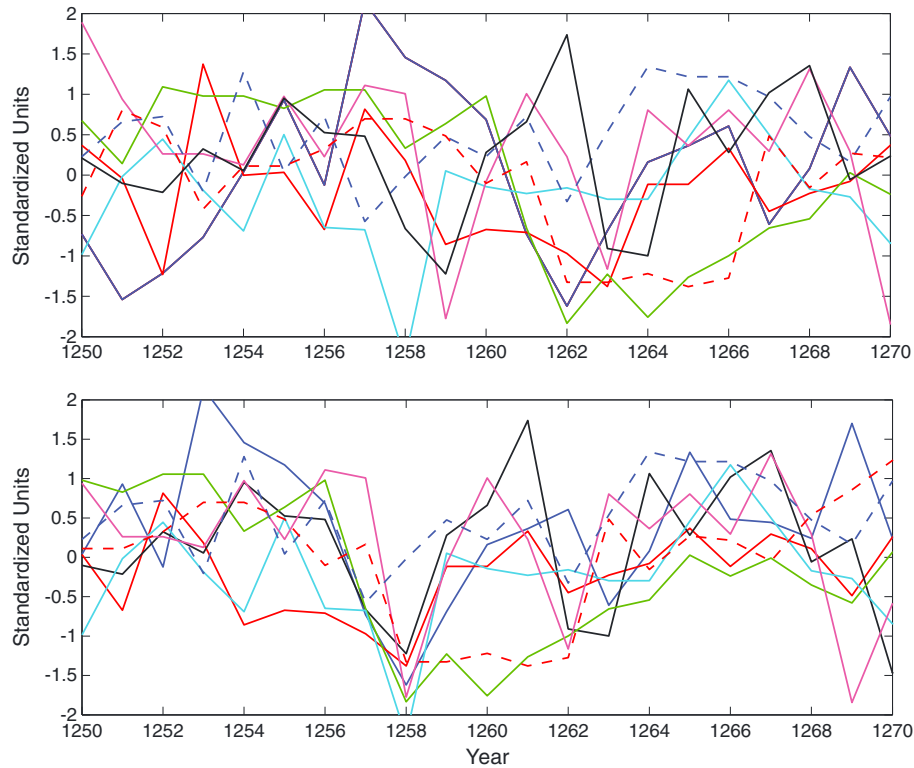
Figure 4. The eight regional series available at 1258 on their original time scales (top) and altered time scales to align a cool year at or later than 1258 (bottom). The series were lagged backward in time (up to a maximum of five years) to align constructively with regard to 1258 (lags are: blue dash 0; cyan 0, magenta 1; black 1; red dash 4; green 4; blue 4; red 5). This example illustrates how the more objective chronological resampling procedure used in our analyses recovers volcanic cooling events that may have been obscured by chronological errors back in time due to missing rings.

is also unlikely to have occurred by chance (see SI for further details). Another measure of the significance of the cooling events is, e.g., the number of surrogates that breach the 99.5[th] percentile. Given $N = 8000$ realizations, we would only expect 40 such cases from chance alone. Yet the actual distribution of Monte Carlo surrogates indicates 336 such cases for AD 1258 and 90 for AD 1816. Furthermore, we observe two surrogates where both the AD 1258 and AD 1816 cooling simultaneously breach the 99.9[th] percentile of the null distribution. As such an event should happen randomly only once in a million realizations, it is obviously highly unlikely to have occurred by chance given an ensemble of $N = 8000$ realizations.

[23] The increased AD 1258 cooling and disappearance of (likely spurious) AD 1260–62 cooling arise from a realignment of much larger cooling signals that are present in individual tree ring series but interfere destructively before they are brought into alignment (Figure 4). Of course, there is still the potential issue of a threshold cooling limit as argued by MFR12 for tree line-located trees, so even accounting for the smearing and damping effect of chronological errors, it is not possible to fully recover the true cooling. An analysis of simulation results from MFR12 supports this conclusion, demonstrating that the procedure used, while recovering some of the cooling, systematically underestimates its magnitude (see SI).

[24] Finally, while MFR12 focused entirely on tree ring width data, it is worth noting that a similar underestimation

of the largest volcanic forcing events might arise in tree ring density data from tree line locations, owing to the possibility of missing rings which would, in principle, serve again to smear and degrade the signals of interest. *Briffa et al.* [1998] specifically sought to detect short-term volcanic cooling signals back to AD 1400 using a reconstruction of Northern Hemisphere mean temperatures based on a composite of maximum latewood tree ring density data, largely from tree line-proximal (boreal forest) environments. We have swapped the short-term cooling signals from *Briffa et al.* [1998] for those of D06, based on a composite combining the low-frequency component ($f < 0.05$ cycle/yr) of D06 with the high-frequency ($f > 0.05$ cycle/yr) component of *Briffa et al.* [1998]. As shown in Figure 5, this alternative reconstruction provides a similar picture to the D06 reconstruction. The response to the AD 1815 Tambora eruption, for example, is similarly muted in the alternative reconstruction. The only notable difference is the curiously large cooling signal for the AD 1601 Huaynaputina eruption in the alternative reconstruction. In most volcanic forcing estimates [see, e.g., *Jones and Mann*, 2004; *Jansen et al.*, 2007], this eruption does not rank, in terms of radiative forcing, among even the top five eruptions during the post AD 1400 period. Yet it shows the largest signal in this case. In this sense, the alternative reconstruction, using tree ring density data for the high-frequency signal, is even more inconsistent with model-estimated responses. We conclude that the general underestimation of tree ring-based estimates of volcanic
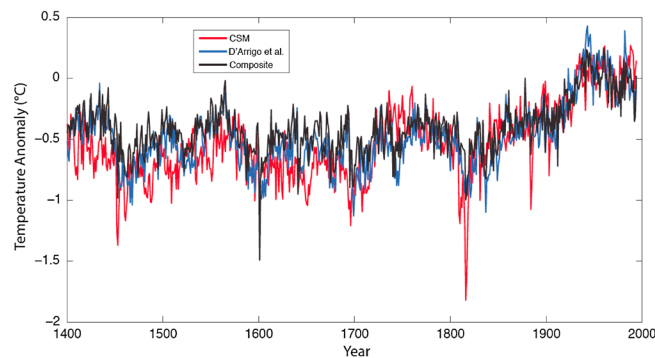
Figure 5. Comparison of CSM model-predicted NH mean series (red) with both D06 (blue) and splice (black) of D06 low-frequency and *Briffa et al*. [1998] high-frequency components.

cooling relative to model-simulated (and, in the case of AD 1815, instrumentally recorded) temperatures may apply to reconstructions based on tree ring density as well as those based on ring widths.

3. Modeled vs. Reconstructed Volcanic Cooling

[25] Given that there are discrepancies between the modeled and proxy-reconstructed response to large volcanic forcing events, it is worth trying to pinpoint the source of this discrepancy and to understand its consequences. Is the mismatch specific to volcanic forcing or is it found in the diagnosed responses to other natural radiative forcings? If the forcing is specific to the response to volcanic forcing, how might it be accounted for in comparisons of reconstructed and modeled climate responses? We investigate such questions through (1) a comparison of proxy-reconstructed responses with those diagnosed from the recent CMIP5 past millennium intercomparison project (section 3.1) and (2) an investigation of the impact that threshold responses might have on estimates of ECS using climate models of known sensitivity and looking at the potential impact that biased responses would have on the diagnosis of ECS (section 3.2).

3.1. Reconciling Forced Temperature Responses in CMIP5 Past Millennium Simulations vs. Paleoclimate Reconstructions.

[26] One approach to comparing reconstructed and modeled responses to past volcanic forcing involves the use of a fingerprint approach, where the contribution from the fingerprint of external forcings (as estimated from a model or ensemble of models) to reconstructed temperatures can be estimated using a total least squares (TLS) detection and attribution technique [*Allen and Stott*, 2003]. TLS performs a multiple linear regression where noise (in this case due to internal variability) is present on both the regressor and target of regression. *Schurer et al*. [2013—henceforth S13] recently used such an approach to estimate forced signals in various proxy reconstructions of Northern Hemisphere mean temperature using an ensemble of models of the past millennium including several CMIP5 simulations [see *Schmidt et al*., 2011]. The TLS detection approach yields a scaling factor β that describes the ratio of the amplitude of the signal detected in a given reconstruction to the amplitude of the multimodel mean predicted signal.

[27] Analyzing the reconstructions for the past millennium, S13 find that the scaling factors β on average are significantly less than one, i.e., the reconstructions appear to have too weak a signal when compared with the multimodel fingerprint. One possible explanation for this discrepancy is that the climate sensitivity of the models is too high (the average CMIP5 ECS value is slightly greater than 3°C—see *Andrews et al*., 2012), in which case, the simulated temperature response to external forcing would indeed be greater than the observed response. If the climate sensitivity were systematically too high, one would expect the models to overestimate the observed response to external forcing regardless of the time scale (e.g., regardless of the smoothing length applied to the series prior to comparison). As it turns out, however, this is not the case (see S13).

[28] S13 show that the mismatch between modeled and estimated responses can be explained almost entirely from discrepancies between the predicted and proxy-reconstructed cooling response to the few largest volcanic eruptions. They performed an analysis where intervals surrounding very large eruptions were masked from the fingerprint procedure (a criterion of optical depths in excess of 0.25 was employed, leading to the elimination of both the AD 1258 and AD 1815 eruptions, and the mid-15th century Kuwae eruption; for each of these eruptions, intervals spanning 5 years on either side were masked out). The majority of the scaling factors were then found to lie around unity, indicating a model response that is consistent with the reconstructions. The uncertainty intervals for β also increased as expected, since the largest few volcanic eruptions represent the strongest nonanthropogenic signals in the climate record of the past millennium. Hence, by masking them, one is removing a substantial potential constraint on the scaling factors.

[29] One possible explanation of this result mentioned earlier is that the volcanic radiative forcing itself might have been systematically overestimated in the multimodel mean. *Hegerl et al*. [2006] estimated a total uncertainty in the magnitude of the overall volcanic forcing time series of ~35%, which could accommodate scaling factors as low as estimated by S13 (~0.75). Alternatively, the problem may lie with the threshold biological growth effects explored by MFR12 and, further, in section 2 above. Since tree ring data provide all or much of the high-frequency information for the majority of proxy reconstructions investigated by S13, potential underestimation of the short-term cooling response to
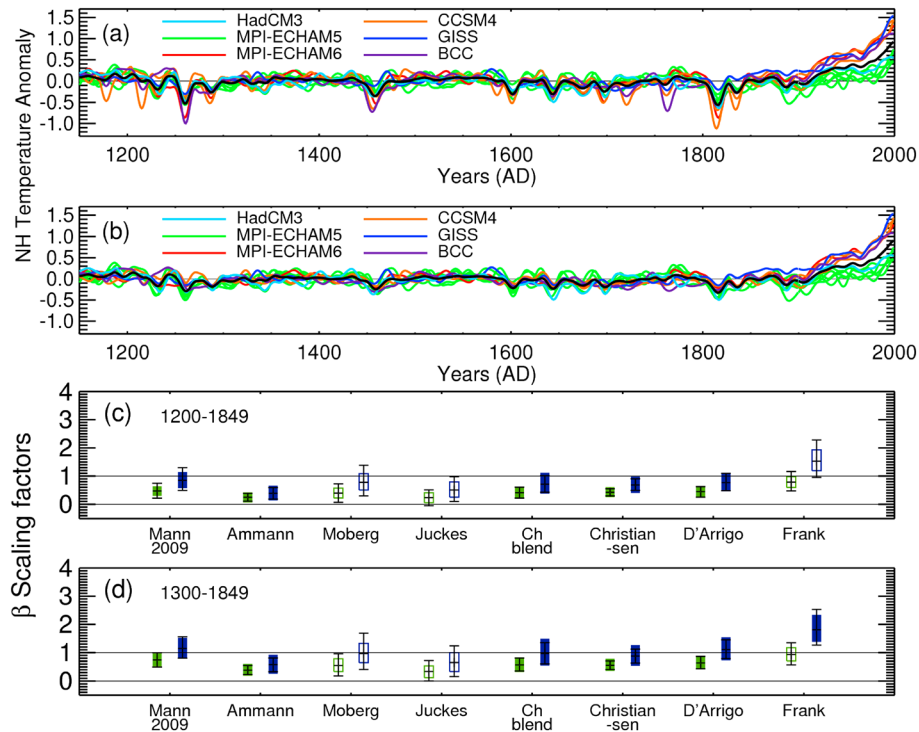
Figure 6. Effect of truncated cooling on TLS fingerprint estimates for multimodel last millennium simulations. (a) Decadally smoothed Northern Hemisphere mean temperatures for the region 0–90N land and ocean for the various simulations, the multimodel mean is shown in black. (b) As in Figure 6a but with cooling truncated at −1°C relative to 1961–1990. TLS estimates of scaling factors β for various published hemispheric temperature reconstructions (see S13 for details) are shown based on both (c) AD 1200–1849 and (d) AD 1300–1849 interval both with (blue) and without (green) the −1°C truncation. Best-fit scaling factors (crosses) are shown with 5–95% ranges (vertical rectangles—where a filled rectangle indicates that the fit passes a residual consistency check; see S13 for details); results from an analysis with noise variance scaled to the residual variance are shown by a vertical line. Fingerprints are detectable if scaling factors are significantly above zero and consistent with the reconstruction if not significantly different from unity. (Note the Christiansen and Ljungqvist reconstruction has been rescaled to fit the instrumental record following S13 prior to the analysis).

large eruptions due to the use of tree ring data would explain why the reconstructions appear to match well the low-frequency forced temperature response predicted by climate models, but show too small a short-term cooling response to volcanic eruptions in comparison with the model simulations. While the precise extent and larger implications of this effect are still a matter of debate, it is certainly worth asking what implications such threshold response limitations could have on derived climate sensitivity estimates.

[30] To test the threshold effect, whether due to the hypothesized tree growth limitations of MFR12 or other processes, we applied (see Figure 6) a temperature threshold of −1.0°C maximum cooling approximating the apparent cooling threshold inferred by MFR12, to the annual hemispheric temperature series simulated by the same models as analysed in S13 [CCSM4—*Landrum et al.*, 2013; MPI-ECHAM5—*Jungclaus et al.*, 2010; MPI-ESM-P—*Giorgetta et al.*, 2012; HadCM3—*Pope et al.*, 2000; *Gordon et al.*, 2000; GISS-E2-R; Bcc-csm-1-1—*Wu*, 2012] (see Figure 1a). This results in simulated pseudo temperatures with substantially reduced volcanic cooling (compare Figures 6a and 6b). We then employ the same TLS analysis technique as used in S13 to test whether these truncated model results are

more consistent with the NH reconstructions analysed in S13 [*Ammann et al.*, 2007; *Juckes et al.*, 2007; *Mann et al.*, 2009; *Moberg et al.*, 2005; *D'Arrigo et al.*, 2006; *Frank et al.*, 2007; *Christiansen and Ljungqvist*, 2011; *Hegerl et al.*, 2007 (CH-Blend reconstruction)]] (where both models and reconstructions are decadally smoothed prior to the analysis), and whether this effect can explain the discrepancy in scaling values discussed above. As Figures 6c and 6d show, fingerprints based on these pseudo temperatures are consistent with a far greater proportion of the actual temperature reconstructions than are the raw simulated temperatures. While scaling factors are significantly below unity using the raw simulated temperatures for nearly all reconstructions, they are consistent with a value of unity for nearly all reconstructions using the pseudo temperatures. This conclusion applies to intervals that both include (AD 1200–1849) and do not include (AD 1300–1849) the very large AD 1258 eruption (Figures 6c and 6d, respectively).

[31] In summary, the low (less than unity) scaling factors are very likely linked to the response to the few largest volcanic eruptions. Errors in the ability of proxy data to resolve the full magnitude of short-term volcanic cooling can account for the model-data discrepancies (i.e., accounting
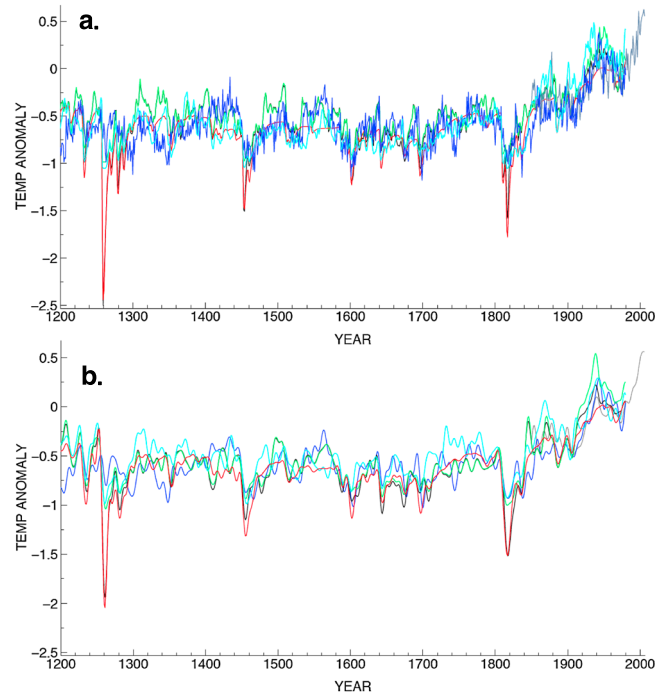
Figure 7. Comparison of EBM-simulated NH series with (black) and without (red) stochastic noise forcing, against actual tree ring (blue) temperature reconstruction and simulated tree-growth temperature series with two different realizations of noise forcing (green; cyan). Shown are both the (a) annual and (b) decadally smoothed series.

for potential such errors leads to scaling factors that are consistent with unity).

### 3.2. Impact of Threshold Responses on Estimated ECS

[32] *Hegerl et al.* [2006] used various proxy reconstructions of Northern Hemisphere mean temperature back to AD 1300 based either partly or entirely on tree ring data, to constrain "Charney" ECS estimates and found that the use of this additional constraint tended to point toward lower-end climate sensitivities, e.g., a median ECS of $\Delta T_{2xCO2} = 2.6°C$, somewhat toward the low end of the typically cited [e.g., *Meehl et al.*, 2007] 2.0°–4.5° range. The mean ECS diagnosed from the climate models used in the current CMIP5 multimodel intercomparison [*Andrews et al.*, 2012], for comparison, is slightly greater than 3.0°C (3.37°C).

[33] Here we examine the potential impact of threshold-like proxy temperature responses to abrupt cooling from volcanic eruptions on past ECS estimates derived from reconstructions of northern hemisphere temperatures. We employed EBM simulations as in MFR12 using the same radiative forcing estimates and midrange value of ECS of $\Delta T_{2xCO2} = 3°C$ as in MFR12. Separate analyses were done using starting date AD 1200 (which contains the AD 1258 eruption) and AD 1300 (which was employed by *Hegerl et al.* [2006] to eliminate the impact of the very large data/model misfit for the AD 1258 response). The EBM was driven additionally with Gaussian white noise weather forcing, yielding a red noise natural variability component, the amplitude of which was chosen to give the approximate variance breakdown of *Crowley* [2000] of 65% forced vs. 35% internally generated variability.

[34] For each modeled temperature series, a synthetic "tree ring" temperature series was calculated using the forward tree growth model protocol of MFR12 which accounts for threshold growth responses, i.e., yields a truncated cooling response for very large eruptions with a maximum cooling relative to the 20th century baseline of slightly greater than ~1°C (note that the additional effect of potential missing rings and resulting chronological errors is considered later). An ensemble of 1000 independent realizations of this process was produced. Figure 7 shows a sample from the ensemble of 1000 Northern Hemisphere mean temperature series and its associated synthetic "tree ring" temperature series.

[35] For each realization, an ECS estimate is obtained by minimizing the mean squared error between the target temperature series and a distribution of purely forced EBM responses, where the EBM sensitivity parameter is varied over a broad range of values. Consistent with *Hegerl et al.* [2006], we smoothed the time series to emphasize decadal and longer time scales prior to the analysis (parallel analyses using annual resolution data yield similar conclusions—see SI). Application of this analysis procedure to the ensemble of series yields a distribution of ECS values consistent with the simulated temperatures. The procedure was applied to both the simulated temperature series themselves and the synthetic "tree ring" temperature series derived from the simulated temperatures as described above. The resulting ECS distributions (Figure 8) are shown for both pre-instrumental intervals AD 1300–1849 (Figure 8a) and AD 1200–1849 (Figure 8b).

[36] The analysis procedure yields the correct results when supplied with the ensemble of simulated temperature series, with the distribution centered on the true ECS value
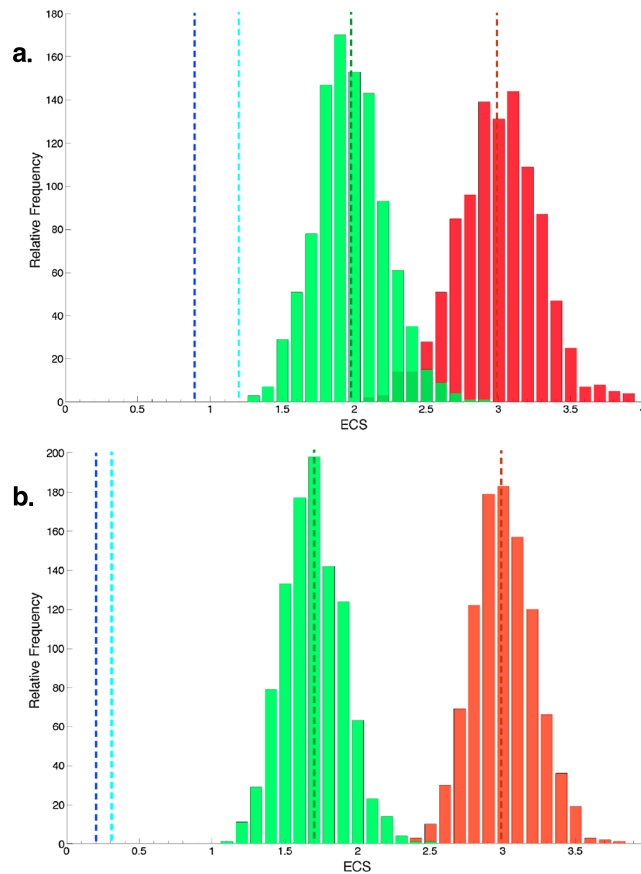
Figure 8. PDF of ECS using decadally smoothed data between (a) AD 1300–1849 and (b) AD 1200–1849 (red = simulated actual temperature series; green = synthetic tree ring temperature series). Shown by dashed vertical lines are mean of the ESC distribution for simulated temperature series (red), mean of ECS distribution for synthetic tree ring temperature series (green), ECS estimate using MFR12 simulated tree ring temperature series where chronological error accumulation due to inferred missing rings is taken into account (cyan), and sensitivity estimate for D06 tree ring temperature reconstruction (blue). True value of ESC is 3.0 in both cases.

($\Delta T_{2xCO2} = 3°C$) ranging from roughly 2.5 to 3.5°C depending on the particular noise realization. The conclusion is insensitive to whether the full AD 1200–1849 or truncated AD 1300–1849 interval is used. By contrast, use of the synthetic "tree ring" temperature series leads to a significant underestimation bias in the ECS distribution, a direct result of the systematic underestimation of the largest volcanic cooling signals (which are indeed the largest forced climate signals during the pre-industrial time periods analyzed). The underestimation bias is at least −1°C, with the synthetic tree ring series yielding ECS values centered on $\Delta T_{2xCO2} \sim 2°C$ if the AD 1300–1849 interval is used. An even lower value $\Delta T_{2xCO2} \sim 1.7°C$ is obtained if the full AD 1200–1849 interval is used, due to the impact of the greatly underestimated AD 1258/1259 cooling.

[37] The sensitivities derived from the actual D06 tree ring temperature reconstruction are even lower than those estimated for the synthetic tree ring series ($\Delta T_{2xCO2} \sim 1°C$ if the AD 1300–1849 interval is used, and $\Delta T_{2xCO2} < 1°C$ if the AD 1200–1849 interval is used). They are, however, comparable to the ECS value derived (Figure 8) from the MFR12 simulation where the additional impact of chronological errors accumulated back in time due to "missing rings" is taken

into account (the MFR12 simulation in question itself involves a Monte Carlo calculation. It would thus be prohibitive to perform an ensemble of a large number of realizations in this case, so we simply show results based on the representative realization featured in MFR12). These chronological errors lead to a misalignment between forcing and estimated response, and a further reduction in inferred ECS. Of course, errors in the estimated radiative forcing can also lead to a misalignment between forcing and response, and a reduction in estimated ECS. The key point here, however, is that even in the absence of chronological errors from hypothesized "missing rings"—a matter about which there is currently a vigorous debate in the peer-reviewed literature [e.g., *Mann et al.*, 2012b; *Anchukaitis et al.*, 2012] —the existence of a threshold beyond which tree ring reconstructions cannot record further cooling, alone, will lead to a systematic underestimate of ECS.

[38] The above results thus suggest that the truncated response to large volcanic cooling events hypothesized by MFR12 to result from temperature thresholds in tree growth responses, and additionally the potential chronological errors due to missing growth rings, if present will both lead to an underestimate of ECS when attempting to constrain this

quantity through a comparison of climate model simulated temperatures and tree ring-based temperature reconstructions from past centuries. Since virtually all decadally resolved hemispheric proxy temperature reconstructions of the past millennium rely at least partly on tree ring data, this bias is likely to afflict nearly all studies using paleoreconstructions of the past millennium to constrain ECS.

4.   Summary and Conclusions

[39]   First, we showed that the presence of missing rings in regional tree ring temperature composites as hypothesized in MFR12 is not only plausible from a theoretical perspective, but appears to have circumstantial support in the behavior of the actual underlying regional tree ring data and resulting hemispheric temperature composites. When corrected for through a Monte Carlo temporal resampling procedure that accounts for estimated chronological errors, realizations with considerably larger volcanic cooling signals emerge in the hemispheric tree ring temperature composite for the AD 1258 and AD 1815 eruptions. These larger cooling signals are considerably more consistent with climate model simulated temperatures responses than the uncorrected tree ring temperature composite and, in the case of the AD 1815 eruption, with a newly published Northern Hemisphere instrumental temperature series that indicates a substantial (~1.6°C relative to 20th century base period) hemispheric cooling response to this eruption.

[40]   We have investigated the potential role that threshold tree growth responses might play in explaining discrepancies between the amplitude of estimated and modeled forced responses of Northern Hemisphere mean temperature over the past millennium. We find that accounting for potential threshold limits of cooling in proxy reconstructions can reconcile proxy reconstructions with the results of the CMIP5 multimodel simulations of the past millennium. We furthermore find that the hypothesized underestimation of volcanic cooling will likely have led to an underestimation of ECS in past such studies, though the precise magnitude of the underestimate will depend on the relative mix of tree ring and nontree ring proxy data used.

References

Allen, M. R., and P. A. Stott (2003), Estimating signal amplitudes in optimal fingerprinting, Part I: Theory, *Clim. Dyn.*, *21*, 477–491.

Ammann, C. M., and E. Wahl (2007), The importance of the geophysical context in statistical evaluations of climate reconstruction procedures, *Clim. Change*, *85*, 71–88.

Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas (2007), Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model, *Proc. Natl. Acad. Sci.*, *104*, 3713–3718.

Anchukaitis, K., et al. (2012), Tree rings and volcanic cooling, *Nat. Geosci.*, *5*, 836–837.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor (2012), Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere–ocean climate models, *Geophys. Res. Lett.*, *39*, L09712, doi:10.1029/2012GL051607.

Briffa, K. R., P. D. Jones, F. H. Schweingruber, and T. J. Osborn (1998), Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years, *Nature*, *393*, 450–455.

Christiansen, B., and F. C. Ljungqvist (2011), Reconstruction of the extratropical NH mean temperature over the last millennium with a method that preserves low-frequency variability, *J. Climate*, *24*, 6,013–6,034.

Crowley, T. J. (2000), Causes of climate change over the past 1000 years, *Science*, *289*, 270–277.

D'Arrigo, R., R. Wilson, and G. Jacoby (2006), On the long-term context for late twentieth century warming, *J. Geophys. Res.*, *111*, D03103, doi:10.1029/2005JD006352.

Frank, D., J. Esper, and E. R. Cook (2007), Adjustment for proxy number and coherence in a large-scale temperature reconstruction, *Geophys. Res. Lett.*, *34*, L16709, doi:10.1029/2007gl030571.

Giorgetta, M. A., et al., Climate change from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project 5, *Journal of Advances in Modeling Earth Systems* (submitted).

Gordon, C., et al. (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dyn.*, *16*, 147–168.

Hegerl, G. C., T. J. Crowley, W. T. Hyde, and D. J. Frame (2006), Climate sensitivity constrained by temperature reconstructions over the past seven centuries, *Nature*, *800*(440), 1,029–1,032.

Hegerl, G. C., T. J. Crowley, M. Allen, W. T. Hyde, and H. N. Pollack (2007), Detection of human influence on a new, validated 1500-year temperature reconstruction, *J. Climate*, *20*, 650–666.

Jansen, E., et al. (2007), Palaeoclimate, in *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller, pp. 433–497, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jones, P. D., and M. E. Mann (2004), Climate over past millennia, *Rev. Geophys.*, *42*, RG2002, doi:10.1029/2003RG000143.

Juckes, M. N., M. R. Allen, K. R. Briffa, J. Esper, G. C. Hegerl, A. Moberg, T. J. Osborn, and S. L. Weber (2007), Millennial temperature reconstruction intercomparison and evaluation, *Clim. Past*, *3*, 591–609.

Jungclaus, J. H., et al. (2010), Climate and carbon-cycle variability over the last millennium, *Clim. Past*, *6*, 723–737.

Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng (2013), Last millennium climate and its variability in CCSM4, *J. Climate*, *26*, 1,085–1,111.

Mann, M. E., et al. (2009), Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, *326*, 1,256–1,260.

Mann, M. E., J. D. Fuentes, and S. Rutherford (2012a), Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures, *Nat. Geosci.*, *5*, 202–205.

Mann, M. E., J. D. Fuentes, and S. Rutherford (2012b), Reply to "Tree rings and volcanic cooling", *Nat. Geosci.*, *5*, 837–838.

Meehl, G. A., et al. (2007), Global Climate Projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller, 747–845, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlén (2005), Highly variable Northern Hemisphere temperatures reconstructed from low- and high resolution proxy data, *Nature*, *433*, 613–617; Corrigendum, 439,1014.

Pope, V. D., M. L. Galliani, P. R. Rowntree, and R. A. Stratton (2000), The impact of new physical parameterizations in the Hadley Centre climate model – HadAM3, *Clim. Dyn.*, *16*, 123–146.

Robock, A. (2005), Cooling following large volcanic eruptions corrected for the effect of diffuse radiation on tree rings, *Geophys. Res. Lett.*, *32*, L06702, doi:10.1029/2004GL022116.

Rohde, R., et al. (2012), A new estimate of the average Earth surface land temperature spanning 1753 to 2011, *Geoinfor Geostat: An Overview*, 1:1, doi:10.4172/gigs.1000101.

Schmidt, G. A., et al. (2011), Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), *Geosci. Model Dev.*, *4*, 33–45.

Schurer, A., G. C. Hegerl, M. E. Mann, S. F. B. Tett, and S. J. Phipps (2013), Separating forced from chaotic climate variability over the past millennium, *J. Climate* (in press).

Timmreck, C., et al. (2009), Limited temperature response to the very large AD 1258 volcanic eruption, *Geophys. Res. Lett.*, *36*, L21708, doi:10.1029/2009GL040083.

Wu, T. W. (2012), A mass-flux cumulus parameterization scheme for large-scale models: Description and test with observations, *Clim. Dyn.*, *38*, 725–744.