Climatic impact of volcanic eruptions over the past millennium

Dissertation

zur Erlangung des Grades

"Doktor der Naturwissenschaften"

im Promotionsfach Geographie

am Fachbereich Chemie, Pharmazie und Geowissenschaften

der Johannes Gutenberg-Universität

in Mainz

Lea Schneider

geb. in Neuwied

Mainz, den 29. Januar 2017

Betreuer der Dissertation

Not available online.

Summary

Volcanic ash clouds, injected into the stratosphere, can alter the Earth's net radiation and initiate widespread surface cooling. Polar ash deposits from the last millennium indicate that past volcanic eruptions impacted global climate much more than any eruption since the onset of modern observations. Paleoclimatic evidence thus contributes to recognize the full range of associated atmospheric processes and potential threats of future eruptions. Additionally, climate reconstructions enable a validation of model simulations and their sensitivity to volcanic forcing. In this dissertation a new annually resolved temperature reconstruction for the Northern Hemisphere is developed to examine volcanic induced cooling events and to gain new insight in the history of volcanism.

High-frequency variability in hemispheric temperature reconstructions is often based on tree-ring width (TRW) measurements, although this parameter was previously accused to be biased by biological memory. The analyzed TRW chronologies respond attenuated and delayed to abrupt temperature changes. Chronologies of maximum latewood density (MXD), in contrast, are more flexible and better suited for examining short-lived cooling events. For Europe the reconstructed temperature drop subsequent to volcanic eruptions is in good agreement with long instrumental records. Estimates of mean volcanic cooling are, however, very sensitive to the selection criteria for climatically relevant eruptions and the varying time lag between ash injection and climatic impact.

Before extending the regional assessment of volcanic signals to a hemispheric scale, the spatial representativeness of the MXD network is assessed. Despite the comparatively small number of multi-centennial MXD chronologies, the ensemble mean can be calibrated successfully against extratropical summer temperatures of the Northern Hemisphere. The resulting reconstruction has an unprecedentedly accentuated high frequency signal over the past 1400 years and comprises a couple of distinct cooling events protruding beyond background year-to-year variability and longer term trends.

Shortcomings in existing archives of volcanism motivate the application of a detection algorithm generating an independent record of climatically relevant eruptions. Designed for the characteristic volcanic cooling pattern and previously approved in a pseudoproxy environment, the algorithm picks up more than a dozen of these events in the MXD derived temperature reconstruction. Their dating generally agrees with historically documented eruptions or ash deposits from polar ice-cores. For some events, however, the exact timing is not in-phase proving that existing archives are either incomplete or imprecise.

Deficiencies in ice-core derived reconstructions of volcanic activity propagate into the output of climate models forced with these datasets and explain some of the observed discrepancies between simulated and reconstructed volcanic cooling estimates. But the detection algorithm presented herein is capable to determine the most significant forcing events in both temperature estimates separately. Assessing the maximum cooling rate based on the detected events reveals a stronger and more persistent volcanic

signal in simulated temperatures. Reconciling the ice-core derived reconstructions with the insights gained from paleoclimatology would improve the next generation of forcing datasets and last millennium model simulations. Besides facilitating data-model comparisons this could also expose more links between climate extremes and human history.

Zusammenfassung

Wird vulkanische Asche in die Stratosphäre eingetragen, kann dies die Strahlungsbilanz der Erde beeinflussen und großräumig oberflächennahe Abkühlung verursachen. Polare Ascheablagerungen aus dem letzten Jahrtausend deuten darauf hin, dass das globale Klima in der Vergangenheit von weit größeren Vulkanausbrüchen beeinflusst wurde als in den letzten Jahrzehnten. Paläoklimatische Untersuchungen sind daher unumgänglich, um alle zugehörigen Prozesse zu verstehen und das Gefahrenpotential zukünftiger Eruptionen abschätzen zu können. Zudem lassen sich mit Klimarekonstruktionen Modellsimulationen und deren Sensitivität gegenüber vulkanischem Strahlungsantrieb validieren. Im Rahmen dieser Dissertation wird eine neue, jährlich aufgelöste Temperaturrekonstruktion für die nördliche Hemisphäre entwickelt, mit deren Hilfe vulkanisch bedingte Abkühlungsereignisse untersucht und neue Erkenntnisse über klimarelevanten Vulkanismus im letzten Jahrtausend gewonnen werden.

Hochfrequente Variabilität in hemisphärischen Temperaturrekonstruktionen basiert häufig auf Jahrringbreite-Messungen (TRW), obwohl dieser Parameter durch den Einfluss biologischer Speicherprozesse in der Kritik steht. Während gezeigt werden kann, dass TRW-Chronologien abrupte Temperaturschwankungen abgeschwächt und zeitlich verzögert wiederspiegeln, stellen sich Chronologien, die auf maximaler Spätholzdichte (MXD) basieren, als deutlich anpassungsfähiger heraus und sind deshalb für die Analyse kurzlebiger Abkühlungsereignisse geeigneter. Für den europäischen Raum stimmt der beobachtete mit dem anhand von Jahrringdichten ermittelten Temperaturrückgang nach Vulkanausbrüchen gut überein. Die durchschnittliche Abkühlungsrate ist jedoch sehr stark von den Auswahlkriterien für klimarelevante Ausbrüche sowie dem zeitlichen Versatz zwischen vulkanischer Aktivität und klimatischer Reaktion abhängig.

Bevor die regionalen Analysen auf hemisphärische Ebene projiziert werden können, wird die Repräsentativität des MXD-Netzwerks ausgewertet. Trotz der vergleichsweise geringen Anzahl an MXD-Chronologien, lässt sich das hemisphärische Mittel erfolgreich gegen extratropische Sommertemperaturen der nördlichen Hemisphäre kalibrieren. Die resultierende Rekonstruktion weist ein besonders akzentuiertes hochfrequentes Signal auf. Über die letzten 1400 Jahre heben sich immer wieder kurzlebige Abkühlungsereignisse von der vorherrschenden Jahr-zu-Jahr Variabilität ab.

Um Schwächen in den bestehenden Archiven vulkanischer Aktivität zu überbrücken, wird ein Detektions-Algorithmus entwickelt, mit dessen Hilfe eine unabhängige Auflistung klimarelevanter Ausbrüche erstellt werden kann. Der auf die spezielle Signatur vulkanisch bedingter Abkühlung getrimmte Algorithmus detektiert über ein Dutzend solcher Ereignisse in der MXD-basierten Temperaturrekonstruktion. Diese decken sich grundsätzlich auch mit historisch dokumentierten Ausbrüchen oder Ascheablagerungen in polaren Eisbohrkernen. Es lassen sich jedoch nicht alle

Abkühlungsereignisse mit einem aus anderen Quellen belegten Ausbruch zeitlich in Einklang bringen. Dies legt nahe, dass die bisherigen Archive für vulkanische Aktivität unvollständig oder ungenau sind.

Wird der vulkanische Strahlungsantrieb fehlerhaft rekonstruiert, pflanzen sich Unstimmigkeiten in Klimamodellen, die auf diese Datensätze zurückgreifen, fort. Mit dem in dieser Arbeit entwickelten Detektionsverfahren lassen sich jedoch sowohl für Temperaturrekonstruktionen als auch –simulationen die bedeutendsten Ereignisse extrahieren und so die jeweils maximalen Abkühlungsraten bestimmen. Dieser neue Ansatz zeigt, dass in den hier untersuchten Modellsimulationen der vulkanische Strahlungsantrieb größere und anhaltendere Abkühlung verursacht als es sich aus rekonstruierten Temperaturreihen ableiten lässt. Durch einen Abgleich des rekonstruierten Strahlungsantriebes mit den paläoklimatischen Informationen können künftige Modellsimulationen der Realität näher kommen und damit neue Zusammenhänge zwischen klimatischen Extremen und Entwicklungen in der Menschheitsgeschichte aufgedeckt werden.

Acknowledgements

Not available online.

Table of contents

1. Introduction

Volcanic eruptions are among the most powerful natural phenomena in the Earth's history (Oppenheimer 2011). Catastrophic environmental scenarios pose a threat to the local and regional population. But even on hemispheric scales societies and economies can be struck by the subtle, yet profound impact of a single eruption: the volcanic cooling (Büntgen et al. 2016, Luterbacher and Pfister 2015). With a temperature decrease of approximately 0.5°C in the global average, the 1991 eruption of Mt. Pinatubo (Philippines) is the largest and climatically most relevant event since the implementation of modern instruments (McCormick et al. 1995). Historic observations, polar ash deposits and other archives, however, suggest that much larger eruptions perturbed global climate during the turn of the last millennium (Lavigne et al. 2013, Sigl et al. 2015, Brazdil et al. 2016). Recognizing the full range of associated atmospheric processes and potential societal aftermaths motivates the enquiry into past volcanism.

The climatic impact of an eruption is closely related to the size and the composition of the volcanic plume. Most relevant is the quantity of gaseous sulfur compounds injected into the stratosphere where transformation and aggregation in sulfate aerosols takes place (Robock 2000). Stratospheric circulation disperses the aerosols and transports the shading haze towards the poles or, if the eruption occurs in the tropics, over the entire globe (Gao et al. 2008). The particles scatter and absorb solar radiation resulting in stratospheric warming and cooling at the Earth's surface. After an abrupt initial drop, surface temperatures recover slowly due to the prolonged stratospheric turnover time of up to three years (Wigley et al. 2005). This characteristic volcanic fingerprint cannot always be deciphered in local temperature records or regional averages, because volcanically induced cooling does not need to be a local extreme (Esper et al. 2013b). In hemispheric or global temperature averages, however, a sequence of cold years protrudes beyond background variability in periods of strong volcanic forcing (Briffa et al. 1998a).

The frequency and magnitude of past volcanic forcing events is reconstructed based on sulfate deposits in polar ice-sheets (Gao et al. 2008, Crowley and Unterman 2013, Sigl et al. 2015). Volcanic deposition signals in Arctic and Antarctic ice-cores yield an estimate for the total stratospheric aerosol injection and models for stratospheric transport simulate the temporal and spatial evolution of the ash cloud. Changes in the interpretation and composition of stacked ice-core records and the associated models resulted in significant discrepancies among published forcing reconstructions. This indicates uncertainties introduced by the underlying proxy data (Sigl et al. 2014) or by their subsequent processing (Plummer et al. 2012). Links to other archives, such as historic documents, are commonly used to confirm the ice-core information, but concurrently manifest the uncertainties in the ice-core derived data (Baillie and McAneney 2015). Evaluating the climatic response by linking reconstructed eruptions with temperature records, however, requires a well-dated and exhaustive list of climatically relevant volcanic eruptions.

Temperature sensitive trees, as an archive for short-lived climatic variations, capture the climatic impact of extreme volcanic events (Anchukaitis et al. 2012, Esper et al. 2013a, Briffa et al. 1998a). The annually growing tree-rings can be dated precisely (Stokes and Smiley 1968) and ample knowledge about their physiologic response to environmental influences has been generated in the past decades (Fritts 1976, Vaganov et al. 2006, Cuny et al. 2015). The abundance of tree-ring chronologies from cold environments, where growth is limited by summer temperatures, offers a wide-spread and dense network of high-resolution temperature proxies for the past millennium (Esper et al. 2016). Combining these chronologies in large-scale averages reveals insight in the history of hemispheric temperatures (Esper et al. 2002, D'Arrigo et al. 2006, Wilson et al. 2016). These long-term temperature reconstructions contribute to disentangle the relationship between internal climate variability and external climate drivers. An assessment of volcanically driven variability throughout the past millennium is, however, still pending.

1.1 The choice of the tree-ring parameter

Annual rings are a typical feature of trees growing in an extra-tropical environment with warm summers, favorable for growth, and cooler winters, characterized by cambial dormancy (Fritts 1976). During the growing season, stem diameter and mass increase through cambial cell division, cell enlargement and cell wall thickening (Cuny et al. 2015). Cell division and cell enlargement initiate before the onset of photosynthetic activity, using assimilates from the previous growing season, and terminate well before cell wall thickening is completed (Fritts 1976). In high altitudes and high latitudes, where temperatures and light availability limit the length of the growing season, the radial growth rate is sensitive to environmental conditions (Rossi et al. 2007). An easy accessible temperature proxy is thus the variability in tree-ring width (TRW) measured along the stem radius.

A physiologically and analytically more complex temperature proxy is the maximum latewood density (MXD). Wood density is controlled by the ratio between cell wall thickness and cell size. A coniferous tree-ring can be separated in light earlywood with large, thin walled cells and dense latewood with small, thick walled cells. The rate of cell wall thickening depends on ambient air temperatures, but the duration of this process compensates much of the environmental influences (Cuny and Rathgeber 2016). Thus, the characteristic transition from earlywood to latewood is unaffected by temperature's year-to-year variability. A particular exception are the small and thick walled cells produced just before the cessation of cambial activity. Their density is distinctly controlled by temperature because the compensatory mechanisms between rate and duration of wall thickening break up for the latest latewood cells (Cuny and Rathgeber 2016). Counterintuitively, the formation and differentiation of these cells can last several weeks and is not finished when cambial activity ceases (Rossi et al. 2007). Hence, MXD data can integrate environmental information of the whole summer with high temperature correlations usually lasting from May/June to August.

The temperature signal in TRW is determined by conditions during the early and previous growing season resulting in differing climatic fingerprints in MXD and TRW. Remobilization processes of stored assimilates and enduring needle generations obscure the immediate environmental influences in TRW data (Pallardy 1997) and conceal abrupt temperature changes as they are observed in the aftermath of volcanic eruptions (Anchukaitis et al. 2012). This is one reason why TRW chronologies correlate weaker with local growing-season temperatures than MXD chronologies. However, the labor intensive production of MXD data limits the number of available chronologies, so that they are absent or underrepresented in large-scale paleoclimatic studies.

1.2 A short review on hemispheric temperature reconstructions

Reconstructing hemispheric or global temperatures for the Common Era has long been in the focus of paleoclimatology and a steadily increasing number of records appear in the assessment reports of the Intergovernmental Panel on Climate Change (IPCC) (Frank et al. 2010, Smerdon and Pollack 2016). Many of these studies combine high-resolution TRW chronologies with other proxy-archives. Although the attention is often primarily on the low frequency component of the resulting reconstructions, an annual resolution facilitates the calibration of the hemispheric ensembles (Wilson et al. 2016).

Tthe latest IPCC-report also addresses the high frequency signal in paleoclimate records (Masson-Delmotte et al. 2013). A couple of annually resolved hemispheric reconstructions and model simulations for the last millennium are investigated with respect to the abrupt temperature changes after volcanic eruptions. Proxy estimates revealed a delayed, less accentuated and weaker response compared to the model simulations. Particularly striking is the model-proxy discrepancy in a study by Mann et al. (2012), that discusses the absence of volcanic cooling in a purely TRW-derived temperature reconstruction. Although the authors omit biologic memory effects in TRW data as the most obvious reason (Anchukaitis et al. 2012, D'Arrigo et al. 2013, Franke et al. 2013, Esper et al. 2013a), the study points unintentionally to the lack of a MXD-derived hemispheric reconstruction. The only density derived record was published by Briffa et al. (1998a) and extends back to 1400CE. Extending this approach further back in time would grant access to short-lived climatic extremes as observed in response to volcanic eruptions over the entire last millennium.

1.3 Research objectives

Together with the network of multi-centennial MXD chronologies the potential for high-frequency analyses in large-scale paleoclimatology is growing. Tapping the potential, this dissertation aims at generating a new millennial hemispheric temperature reconstruction based on MXD data that is able to represent abrupt temperature change. A MXD derived record would reduce the spectral bias introduced in reconstructions by TRW-data (Bunde et al. 2013, Franke et al. 2013) and pave the way for manifold extreme event analyses.

This work accents volcanic induced cooling, because volcanoes are expected to drive the most outstanding negative temperature anomalies (Briffa et al. 1998b, Marotzke and Forster 2015). Beyond characterizing the cooling effect, this study is also targeted at the detection of volcanic events. Deducing the strength and the timing of underlying forcing events from the MXD reconstruction would yield independent access to the history of volcanic activity and allow an evaluation of existing, ice-core derived reconstructions. Complementing eruption dates and forcing records with the climatic impact sheds more light into forcing dynamics and time lags between aerosol emission, stratospheric residence and polar deposition.

Eventually, volcanos are an important agent of external forcing in model simulations. Addressing previously reported mismatches between model and proxy derived estimates of volcanic cooling (Mann et al. 2012, Masson-Delmotte et al. 2013) motivates the investigation of last millennium temperature simulations. A balanced comparison, bridging the peculiarities of model simulations and proxy reconstructions, is the final aim of this dissertation.

1.4 Structure of this dissertation

Because TRW is the most abundant tree-ring parameter, the first aim of this dissertation is justifying the emphasis on MXD chronologies. Section 2 “Signals and memory in tree-ring width and density data” quantifies chronology characteristics associated with extreme events. In addition to the effects of volcanism, the emergence of positive or negative anomalies and the subsequent relaxation in TRW and MXD data is analyzed in comparison to instrumental temperature data.

Before addressing the hemispheric scale, section 3 “European summer temperature response to annually dated volcanic eruptions over the past nine centuries” demonstrates the potential of MXD data to capture the cooling associated with volcanism on a regional scale. The study examines two areas in Europe of particularly good coverage with proxy data and long instrumental temperature observations: the Alps and northern Fennoscandia. Due to potential dating errors in ice-core derived reconstructions, the analyzed volcanic events are drawn from historic observations. Different subsets of volcanic events are compiled to emphasize the importance of using well-dated eruptions. The estimated cooling responses are analyzed with regard to the average magnitude and the average time lag between eruption and temperature drop.

This approach is further pursued in section 4 “Timing and duration of post-volcanic Northern Hemisphere cooling revealed from tree-ring records of maximum latewood density”. Now on hemispheric scale, the study focuses on single sulfur rich volcanic events, dated to the month, revealing the variability of the time lag between eruption and peak cooling. This variability results in discrepancies regarding the averaged cooling response if the volcanic events are either aligned by the eruption date or the cooling peak.

Section 5 “Volcanic induced cooling in instrumental and tree-ring density data” shifts the focus from the temporal evolution of volcanic forcing in section 4 to a spatial dimension. The varying strength of volcanic cooling is assessed in proxy and instrumental temperature data testing the spatial representativeness of the MXD sites. The study also considers the effectiveness of noise cancelation for a hemispheric average based on a relatively sparse proxy network.

The first millennial temperature reconstruction for the Northern Hemisphere solely based on MXD data is presented in section 6 “Revising midlatitude summer temperatures back to AD 600 based on a wood density network”. The new record aims at representing the full spectrum of temperature variability using a network of 15 MXD chronologies longer than 600 years from North America, Europe and Asia.

Significant fingerprints of volcanic cooling in the MXD reconstruction facilitate their independent detection in hemispheric temperature records. An automatic algorithm, developed in section 7 “Detection of designed break functions with an application to volcanic impacts on hemispheric surface temperatures” demonstrates the detection of forcing events in model simulations and pseudo-proxy reconstructions. In the pseudo-proxy environment volcanic forcing is a known parameter and the skill of the detection algorithm can be evaluated based on numerous volcanic events.

Applying the detection algorithm to the MXD-derived temperature reconstruction as demonstrated in section 8 “An alternative record of past volcanism derived from hemispheric summer temperatures over the past millennium” creates a new record of large volcanism during the last millennium. The detected events are used to evaluate different ice-core based reconstructions of volcanic forcing regarding the timing and magnitude of eruptions. In a last step, the maximum cooling response to volcanic events is estimated for tree-ring reconstructions and for model simulations unraveling the previously observed discrepancies.

2. Signals and memory in tree-ring width and density data

Jan Esper[1], Lea Schneider[1], Jason E. Smerdon[2], Bernd R. Schöne[3], Ulf Büntgen[4-6]

[1] *Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany*

[2] *Lamont-Doherty Earth Observatory, Palisades, NY 10964, USA*

[3] *Institute of Geosciences, University of Mainz, 55128 Mainz, Germany*

[4] *Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland*

[5] *Oeschger Centre for Climate Change Research, Bern, Switzerland*

[6] *Global Change Research Centre AS CR, Brno, Czech Republic*

Published in Dendrochronologia 35, 62-70

2.1 Introduction

Volcanic eruptions have been identified as a major natural forcing of the climate system (Oppenheimer, 2011). The aerosols released by large, explosive eruptions tend to cool the earth's surface, but warm the lower stratosphere. Surface cooling results from scattering of incoming solar radiation, i.e. less radiation reaches the ground. Stratospheric warming is triggered by increased absorption of radiation, i.e. more radiation is transferred into sensible heat in 10+ km above ground (Robock, 2000). These processes ought to last for about three years following a large eruption that injects aerosols into the stratosphere (Cole-Dai, 2010, and references therein).

Whereas the climatic effects of volcanic aerosols are assessed based on differing lines of evidence, estimates of the magnitude and persistence of post-eruption cooling are accompanied by large uncertainties. Estimates based on the analysis of satellite and instrumental station data are constrained by the limited number of eruptions during the period of observation (Sear et al., 1987; Self et al., 1981). This limitation is overcome by using annually resolved tree-ring data covering the past centuries to millennia, thereby enabling the assessment of more eruptions and supporting the differentiation from internal climate variability (Briffa et al., 1998; D'Arrigo et al., 2009, 2013; Esper et al., 2013a, 2013b; Fischer et al., 2007; Salzer and Hughes, 2007). However, the noise inherent to proxy data substantially increases the uncertainties of cooling estimates derived from this line of evidence.

Noise in tree-ring data can be differentiated into *(i)* the unexplained, non-climatic (non-temperature) variance typically reducing the magnitude of reconstructed cooling events, and *(ii)* biological memory effects likely biasing the persistence estimates of cooling events. While the unexplained variance and consequences thereof are widely recognized in the paleoclimatic literature (Esper et al., 2005; Lee et al., 2008; von Storch et al., 2004), memory effects are much less acknowledged (Anchukaitis et al., 2012), even if detailed in standard dendrochronological literature (Cook and Kairiukstis, 1990; Fritts, 1976; Matalas, 1962; Schulman, 1956).

Previous work on memory effects in TRW and MXD chronologies is rare, and either focused on just one of the two tree-ring parameters, or in only one region or tree species. Briffa et al. (1998) used a Northern Hemisphere (NH) network of MXD chronologies covering the past 200-600 years to demonstrate the capacity of reconstructing the magnitude and geographic extent of post-volcanic cooling events. Frank et al. (2007a) assessed memory effects in TRW and MXD chronologies from the European Alps subsequent to the 1815 Tambora eruption, revealing a temporally extended response in TRW, by about 5 years, compared to MXD. D'Arrigo et al. (2013) showed a smeared and temporally extended (~10 years) TRW pattern in response to six major eruptions, using hemispheric scale mean TRW and MXD chronologies. Esper (2014) showed a post-eruption delayed response in TRW (~3-4 years), compared to MXD, for a single *Pinus sylvestris* dataset from northern Fennoscandia encompassing 34 volcanic eruptions over the past 900 years. A large-scale assessment of TRW memory effects in response

to the major eruptions of the past millennium, and differentiation from patterns retained in MXD data, has not been completed.

Memory effects in TRW data can arise from physiological processes fundamental to boreal and alpine conifers. Early work by Schulman (1956) and Matalas (1962) already suggested the storage of food products from one year to the next as the main source of non-randomness and autocorrelation in TRW timeseries. These influences appear to have less impact on MXD, representing an estimate of tracheid cell-wall growth that is more directly associated with ambient air temperature changes in cold environments (Schweingruber et al., 1978). For TRW, the storage of starch and sugar in parenchyma ray tissue, the remobilization of carbohydrates from root structures, and the development of needle generations enduring several growing seasons, all likely influence radial increment beyond the instant impact of temperature variability (Pallardy, 1997).

The resulting increased biological memory is imprinted upon the TRW data is reflected by an increased autocorrelation characterizing the chronologies of this parameter. While these general time series characteristics (Esper et al., 2010; Frank and Esper, 2005) and statistical treatment thereof (Cook, 1985) are detailed in the dendrochronological literature, the memory effects in response to extreme events, and underlying processes controlling this behavior, particularly for MXD, are not well understood (Vaganov et al., 2006). The application of methods to remove autocorrelation from tree-ring time series limits the ability to assess the full variance spectrum of past temperature variability (Bunde et al., 2013; Franke et al., 2013), while reconstruction methods may alone alter the scaling of spectral densities in derived dendrochronological and multi-proxy studies (Smerdon et al., 2015). The prewhitening of tree-ring chronologies to reduce biological memory and match the serial correlation of instrumental target data nevertheless has proven useful when assessing long-term hydroclimate variability (Meko, 1981). Changing the spectrum from red to white might be required to avoid low frequency bias in reconstructions of rainfall and associated hydrological phenomena (Cook and Kairiukstis 1990, Meko et al. 2007).

We here assess high frequency signals and memory effects in MXD and TRW chronologies from 11 NH sites covering the past 750+ years. Consideration of these long records enables the evaluation of cooling events in response to a large number (n=29) of major volcanic eruptions. Findings from this long period are compared with TRW and MXD deviations recorded since 1850, a period from which instrumental temperature data are also available. We use these data to additionally address differences between higher and lower latitude tree sites as well as the variability within the NH network, and discuss the results from the two tree-ring parameters with respect to their paleoclimatic significance.

2.2 Material and Methods

Tree-ring data

The tree-ring data used here are a sub-sample of a MXD network compiled by Schneider et al. (2015) for the reconstruction of NH extratropical summer temperatures over the past millennium. Four of the 15 MXD chronologies included in the Schneider et al. network were not considered herin, as the accompanying TRW data from these sites did not contain a significant June-August (JJA) temperature signal. The TRW and MXD chronologies of the remaining 11 sites correlate at $p < 0.05$ with regional summer temperatures, and extend back to AD 1246 or earlier with a minimum replication ≥ 5 series. The network includes data from pine, spruce, and larch (Table 2-1).

Table 2-1 | NH tree-ring sites. * indicates $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ against regional JJA temperatures. Correlations between TRW and MXD are calculated over the 1246-1978 common period using Spline detrended chronologies. Sites ordered by latitude.

Site	Country	Latitude	Longitude	Species	Period	Number	JJA Signal MXD	JJA Signal TRW	Source
Pyrenees (Pyr)	Spain	42.5N	2.5E	Pine	1044–2005	203	***	**	Büntgen et al. 2008
Lötschental (Loe)	Switzerland	47.5N	7.5E	Larch	735–2004	180	***	***	Büntgen et al. 2006
Tyrol (Tyr)	Austria	47.5N	12.5E	Spruce	1047–2003	227	***	***	Esper et al. 2007b
Altai (Alt)	Russia	50.0N	88.0E	Larch	462–2007	59	***	*	Myglan et al. 2012a
Athabasca (Ath)	Canada	52.3N	117.3W	Spruce	1072–1994	102	***	**	Luckman et al., 2005
Jaemtland (Jae)	Sweden	63.5N	15.5E	Pine	1107–1978	158	***	***	Schweingruber et al. 1988
Mangazeja (Man)	Russia	66.7N	82.3E	Larch/Spruce	1246–1990	143	***	**	Schweingruber, ITRDB
Polar Ural (Pol)	Russia	66.9N	65.6E	Larch	778–1990	157	***	***	Briffa et al. 1995
N-Scan (Nsc)	Finland	67.5N	22.5E	Pine	−181–2006	587	***	***	Esper et al. 2012b
Torneträsk (Tor)	Sweden	68.2N	19.7E	Pine	441–2010	124	***	***	Melvin et al. 2013
Camphill (Cam)	Canada	68.3N	133.3W	Spruce	1175–1992	58	***	*	Schweingruber, ITRDB

The spatial distribution of tree sites is uneven, with two sites located in Canada, three in Russia, and six in Europe. Five sites are situated south of 53°N at high-elevation, alpine environments. Six sites are situated north of 63°N in the boreal zone near the northern treeline. Sample replication and chronology length also vary considerably among the sites, with the shortest record originating from Mangazeja (Russia) and the longest from northern Scandinavia (N-Scan, Finland). The number of MXD/TRW measurement series ranges from 58 in Camphill (Canada) to 587 in northern Scandinavia, underlining the variability of the network, and indicating that climate signal strength might considerably change through time (Fig. 2-1).

Detrending and calibration

To assess uncertainty due to detrending methodology, the MXD and TRW data were standardized using different techniques (Fig. 2-2). Regional Curve Standardization (RCS; Esper et al., 2003) and 100-year fixed spline detrending (Spline; Cook and Peters, 1981) were used with both parameters to emphasize low and high frequency variance, respectively. The MXD data were additionally detrended using Hugershoff functions (Hug), whereas for TRW negative exponential curves (NegExp) were chosen, to account for the differently shaped age trends (Fig. 2-1d-e; Cook and Kairiukstis, 1990). All methods (RCS, Spline, Hug/NegExp) were applied using residuals after power transforming the TRW and MXD data (Cook and Peters, 1997). Mean (standard) site chronologies were calculated using the arithmetic

mean, and potential variance changes adjusted to account for temporal replication and inter-series correlation changes (Frank et al., 2007b).

The resulting chronologies were calibrated against instrumental JJA temperatures using the nearest grid point of the 5x5° CRUTEM4v network as detailed in Schneider et al. (2015). We here considered only the tree sites at which both the MXD and TRW chronologies reached $p < 0.05$ against local grid point data, after adjusting the degrees of freedom for autocorrelation in the proxy and instrumental time series (Table 2-1). While this procedure admitted tree-ring chronologies from 11 sites, it should be noted that the summer temperature signal is overall weaker in the TRW chronologies ($\bar{r} = 0.33$) compared to MXD ($\bar{r} = 0.55$). These estimates are conservative, however, as they do not consider any species- or site-specific seasonality in the climate response, and are derived from correlations over the past 150 years, even in areas where early temperature readings had to be infilled using remote instrumental data (details in Schneider et al. 2015). All chronologies were transferred into estimates of past summer temperature variability by centering and scaling them to match the mean and variance of the instrumental target data back to 1850 (Esper et al., 2005).

Autocorrelation and Superposed Epoch Analysis

Basic methods were used to assess memory effects in the TRW and MXD chronologies. We calculated the autocorrelation of the differently detrended (Hug/NegExp, Spline, RCS) chronologies over the past millennium (shorter for the records not covering the full millennium; see Table 2-1) at lags 1 to 20, and

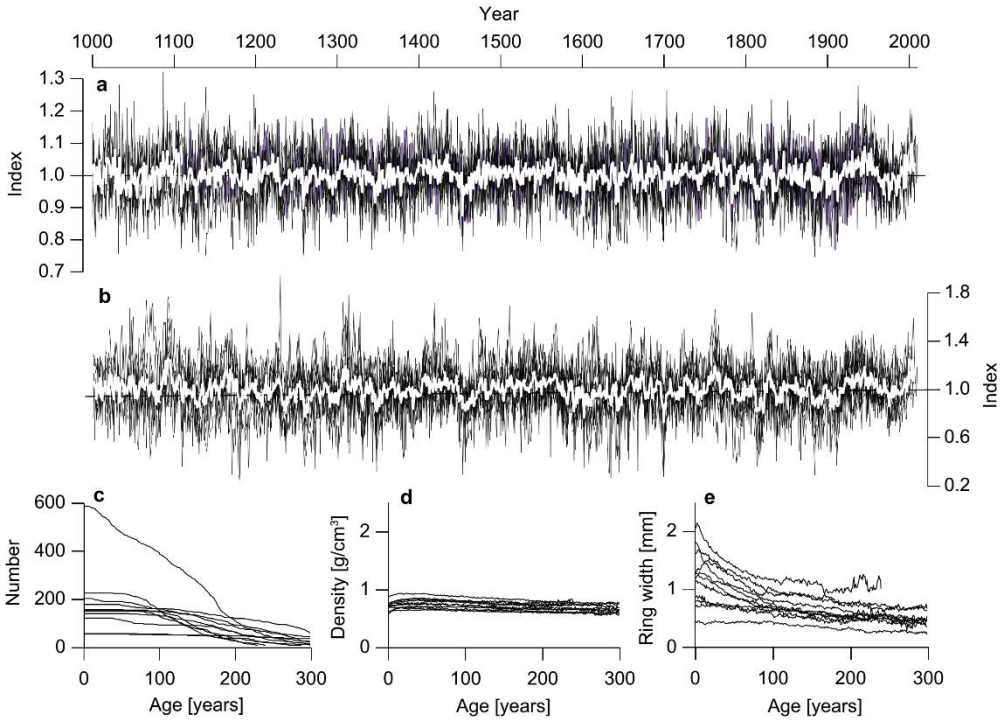


Figure 2-1 | NH MXD and TRW chronologies. (a) The Hug detrended MXD chronologies (black) from 11 locations in the NH shown together with their mean (white) over the past millennium. (b) Same as in (a), but for the NegExp detrended TRW chronologies. (c) Changing sample replication of the 11 chronologies, here shown after aligning the data by cambial age. (d) The age-aligned mean MXD curves ('regional curves') in g/cm^3 over the first 300 years of the tree lifespans. (e) Same as in (d), but for the TRW data in mm.

over 1850-2006 at lags 1 to 10. The latter period was chosen as it additionally permits the calculation of autocorrelations of instrumental temperature data and comparison with tree-ring chronologies.

The MXD and TRW chronologies were decomposed into shorter segments of 15 years and re-aligned by volcanic eruptions using Superposed Epoch Analysis (SEA, Panofsky and Brier, 1958; see Fig. 2-2c for examples in 1257, 1452, 1815). For the past 1000 years, we used 27 annually dated eruptions exceeding a volcanic explosivity index (VEI, Newhall and Self, 1982) of 4 as listed by the Global Volcanism Project (Siebert et al., 2011; for a list of events of the past millennium see Esper et al. 2013b), plus the eruptions in 1257 (Lavigne et al., 2013) and 1452 (Briffa et al., 1998) fingerprinted as aerosol deposition spikes in polar ice core data (Gao et al., 2008; Hammer et al., 1980), and annually dated through tree-ring data (Briffa et al., 1998; LaMarche and Hirschboeck, 1984). We included tropical volcanoes estimated to have released > 1 billion m^3 tephra, and NH extratropical eruptions > 2 billion m^3 tephra to account for the presumed increased significance of lower latitude events (Timmreck et al., 2012; Wegmann et al., 2014).

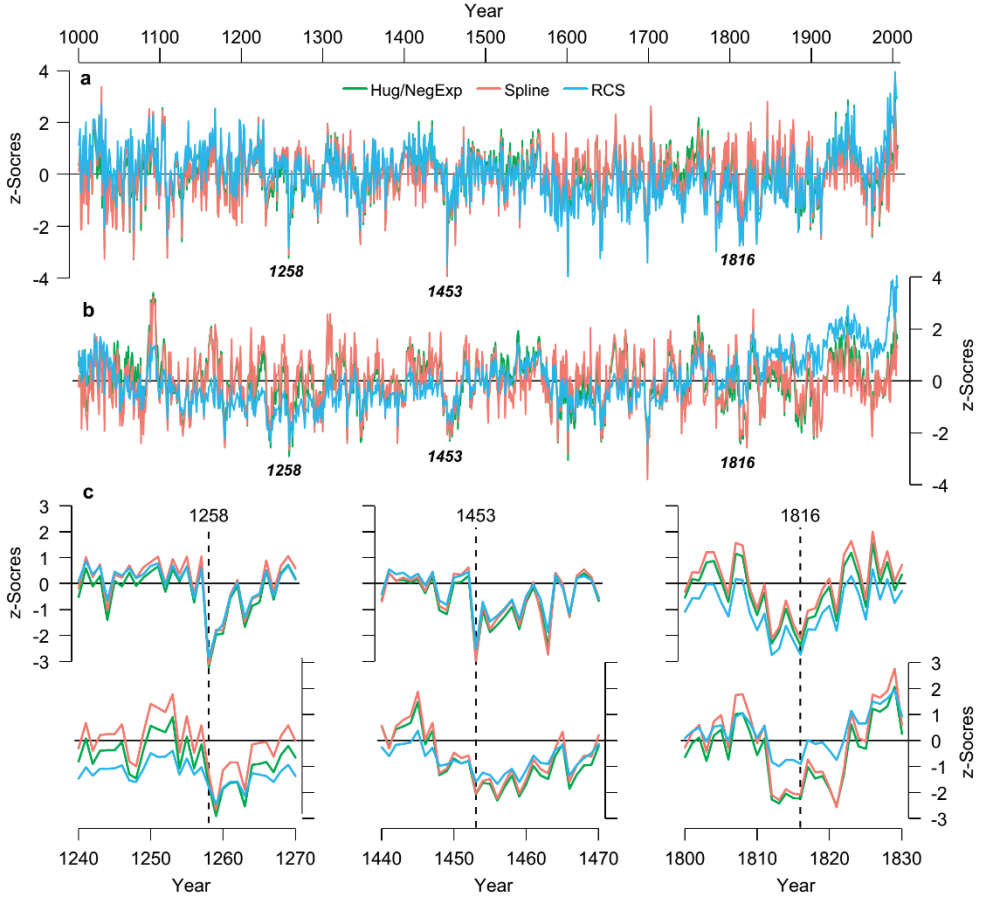


Figure 2-2 | Differently detrended MXD and TRW chronologies. (a) Mean time series of the 11 Hug (green), Spline (red), and RCS (blue) detrended MXD chronologies, normalized over 1000-2006. (b) Same as in (a), but for TRW. (c) MXD (top) and TRW (bottom) patterns around the 1257 (Samalas), 1452 (Kuwae), and 1815 (Tambora) eruptions. Dashed lines indicate the first post-volcanic years. Note there was an additional eruption of unknown origin in 1809.

In addition to these evaluations using data of the past millennium, SEA was applied in the shorter 1850-2006 period, over which instrumental temperature data are available. In this second approach, we re-aligned the proxy and observational data only if local JJA temperatures exceeded -0.5°C in *(i)* the year of a volcanic event, or *(ii)* in one of the two following years. In so doing, we limited this analysis to the locally cold, post-volcanic events in the 11 NH sites considered herein.

We finally used SEA over 1850-2006, considering the 15 coldest and warmest summers recorded in the local temperature data, and aligned the proxy data by these events instead of the volcanic eruptions. In all SEA runs, the displayed MXD, TRW, and JJA data are expressed as anomalies with respect to the five years (-5 to -1) preceding a volcanic (or other) event. Statistical uncertainty was estimated considering the two sigma error range derived from the variance of the mean MXD (TRW) deviations over the five pre-volcanic years (-5 to -1) using the data of the past millennium.

2.3 Results and Discussion

TRW and MXD autocorrelation

The NH TRW chronologies contain substantially more autocorrelation over the past 1000 years than the MXD chronologies, independent of the method used for age-trend removal (Fig. 2-3). All differently detrended chronologies show the typical sequence of largest autocorrelation at lag 1 (= AC1) and subsequently decreasing values. In the RCS detrended data, the mean autocorrelation in both TRW and

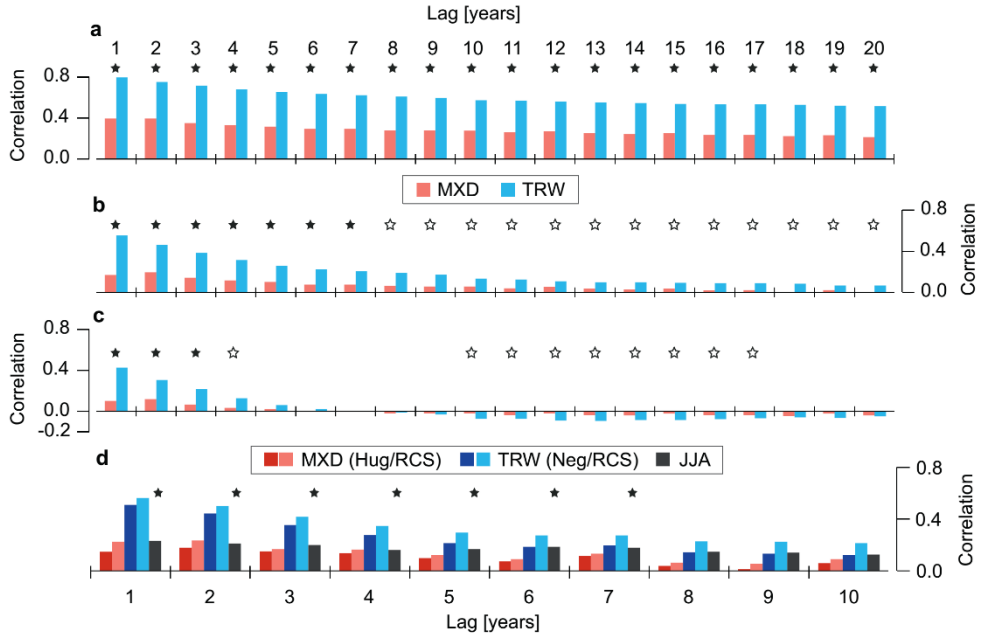


Figure 2-3 | Tree-ring and temperature autocorrelations. (a) Lag 1 to lag 20 autocorrelations of the RCS detrended MXD (red) and TRW (blue) chronologies. Histogram shows the mean values of 11 NH chronologies, calculated over the past millennium. Asterisks indicate statistically significant autocorrelations ($p < 0.05$; empty star if only TRW is significant). (b) Same as in (a), but for the Hug detrended (MXD) and NegExp detrended (TRW) chronologies. (c) Same as in (a), but for the Spline detrended chronologies. (d) Mean lag 1 to lag 10 autocorrelations of 11 tree-ring and JJA temperature time series over 1850-2006. Dark and light red are the Hug and RCS detrended MXD chronologies, dark and light blue the NegExp and RCS detrended chronologies, and black the JJA temperatures. Asterisks indicate significant JJA temperature autocorrelations ($p < 0.05$).

MXD remains positive throughout AC1-AC20, though the TRW chronologies indicate much larger temporal memory ranging from 0.79-0.52 (0.39-0.21 in MXD). In comparison, the Hug/NegExp and Spline detrended chronologies contain much less memory. Autocorrelations are lower at AC1 and approach zero much quicker, turning negative at AC6 (MXD) and AC7 (TRW) in the Spline detrended data, for example. In the individual detrendings (Hug/NegExp, Spline), the initial autocorrelations, AC1-AC5, are notably larger in TRW (0.31 on average) compared to MXD (0.11), revealing increased memory also in these chronology types that are typically used to reconstruct higher frequency climatic changes (Cook and Kairiukstis, 1990).

Comparison of the autocorrelation structure of the RCS and Hug/NegExp chronologies with regional JJA temperatures since 1850 reveals a closer match with MXD (Fig. 2-3d). AC1-AC10 of the local summer temperatures range from 0.23-0.13, whereas particularly the RCS detrended MXD data contain similar memory, ranging from 0.22-0.09. In contrast, the TRW autocorrelations (0.56-0.22 for RCS and 0.51-0.13 for NegExp) exceed the values of the observations, indicating that reconstructions based on this parameter might overestimate the memory retained in local temperature readings.

The TRW/observational mismatch is particularly striking over the first several lags suggesting potential limitations in retaining high frequency climatic variance and recovery from externally forced extreme deviations in the first ~5 years. Testing this assertion, however, requires the application of SEA with such extremes. Statistical treatment to remove autocorrelation and pre-whitening the data is not suitable when assessing temperature, as it disables the reconstruction of any lower frequency variance inherent to this climate element (Franke et al. 2013). Modern tree-ring based climate reconstructions thus consider RCS for tree-ring detrending intending to recover the full frequency spectrum of past temperature variability (e.g., Briffa et al., 1992; Büntgen et al., 2011; Cook et al., 2000, 2002; D'Arrigo et al., 2006; Esper et al., 2002, 2014; Gennaretti et al., 2014; Luckman and Wilson, 2005; Myglan et al., 2012b; Schneider et al., 2015; Wiles et al., 2014).

TRW and MXD signals and memory since AD 1000

The comparison of TRW and MXD variability following 29 major volcanic eruptions of the past millennium reveals substantial differences between the two estimates of response and recovery (Fig. 2-4a). While MXD shows a significant post-volcanic cooling of -0.40°C in year 1 after the eruptions, the mean TRW deviation at 11 NH sites is less distinct. The TRW based estimates are fairly similar throughout years +1 to +5, ranging from -0.13°C to -0.17°C, and characterized by a larger spread among the differently detrended chronologies (the thin grey curves in Fig. 2-4a). This somewhat smeared appearance differs from the MXD pattern, which is dominated by a significantly deviating spike in year +1 and reduced but still noticeable cooling in years +2 and +3. The MXD deviation is quite similar among the different detrendings, and relaxes back to pre-eruption levels in year +4, though does not exceed 0°C before year +7. The MXD signal also remains distinct in year +1 when dividing the NH network into high and mid latitude sites, though the amplitude of temperature deviation differs between

the sub-samples (Fig. 2-4b-c). For TRW, the signal appears to be limited to the high latitude sites, with the trees from lower latitude, alpine environments showing no significant post-volcanic deviation.

The differing patterns in MXD and TRW likely resulted from changing fractions of climatically explained variance, as well as memory effects arising from varying physiological processes controlling radial increment and cell-wall growth. The overall reduced correlation between the TRW chronologies and local JJA temperatures translates into a reduced post-volcanic deviation in the SEA (-0.17°C in year +1, compared to -0.40°C from the MXD data). This initially reduced deviation then remains at a relatively constant level over the next 4-5 years, perhaps due to carbohydrate storage and remobilization effects controlling TRW (Tranquillini, 1964). The MXD volcanic pattern, on the other hand, is overall more transient and appears to be less affected by physiological and structural influences beyond the immediate impact of air temperature variability. It is further noted, however, that the rather large uncertainties displayed in figure 2-4 not only reflect the noise inherent to the proxy data, but also account for the geographically varying temperature response following large volcanic eruptions (D'Arrigo et al., 2013), as has been demonstrated by Briffa et al. (1998) using a dense network of shorter MXD chronologies from the NH extratropics. These interpretations are further examined in the next section in which the proxy deviations are compared with local temperature data over the shorter post-1850 time interval.

Tree-ring and temperature signals since AD 1850

The assessment of recent volcanic eruptions triggering cooling in local temperature data reveals a better fit of MXD compared to TRW (Fig. 2-5). Both the observational and MXD-based summer temperatures

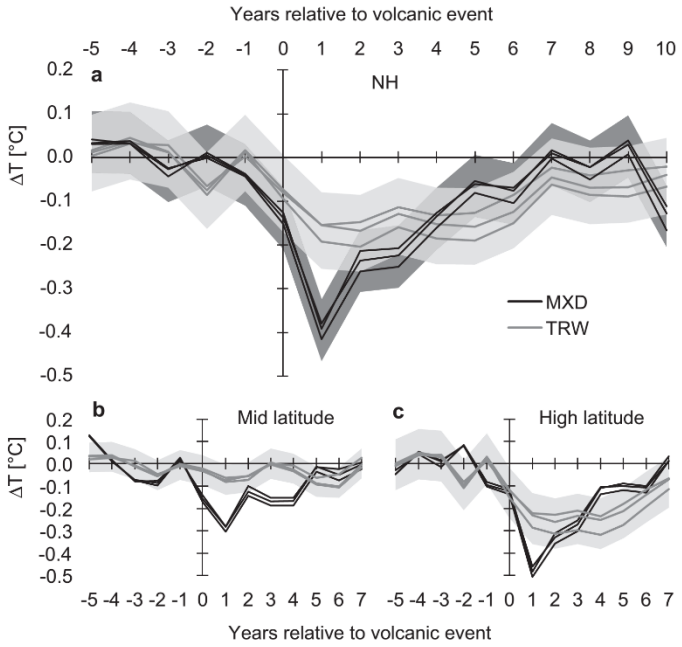


Figure 2-4 | Tree-ring response to volcanic eruptions. (a) SEA of the mean MXD (black) and TRW (grey) data from 11 NH sites aligned by 29 volcanic events over the past 900 years. Single curves represent different detrending methods (Hug/NegExp, Spline, RCS). Shadings indicate the two-sigma uncertainties of the MXD (dark) and TRW data (light). (b) As in (a), but for five sites located south of 53°N (see Table 1). (c), As in (a), but for the six sites located north of 63°N.

reflect cooling in years 0 to +2 ranging from -0.35 to -0.61°C, though maximum deviations are recorded in different years (+2 in MXD, 0 in observations). In contrast, the TRW based estimates do not cohere with the observational data in these years, and do not deviate substantially from climatology (-0.21 to -0.31°C in years 0 and +1), indicating the limited skill of this tree-ring parameter to reflect distinct inter-annual temperature variations. For MXD, it appears somewhat surprising how well the estimates compare with the instrumental cooling magnitude, as the volcanic events aligned in the SEA are limited to yearly anomalies that are colder than -0.5°C in the local observational data (Methods). This selection procedure involved a reduction of the SEA aligned volcanic events by 37% unrelated to local cooling in the gridded temperature data.

The differing responses in years 0 and +2 in the proxy and instrumental data is further explored by dividing the network into sites north of 63°N and south of 53°N (Fig. 2-5b-c). When applying this separation an instantaneous response (year 0) to volcanic activity is recorded at the mid latitude sites, and a delayed response (year +2) at the high latitude sites. While the mid latitude observational temperature response is closely reflected in both the regional TRW and MXD data, the high latitude cooling is only retained in the MXD data. The high latitude TRW chronologies do not indicate any post-volcanic variance differing from other, un-forced years, underscoring the limited skill of this parameter to reconstruct externally forced extremes in boreal environments. This conclusion is, however,

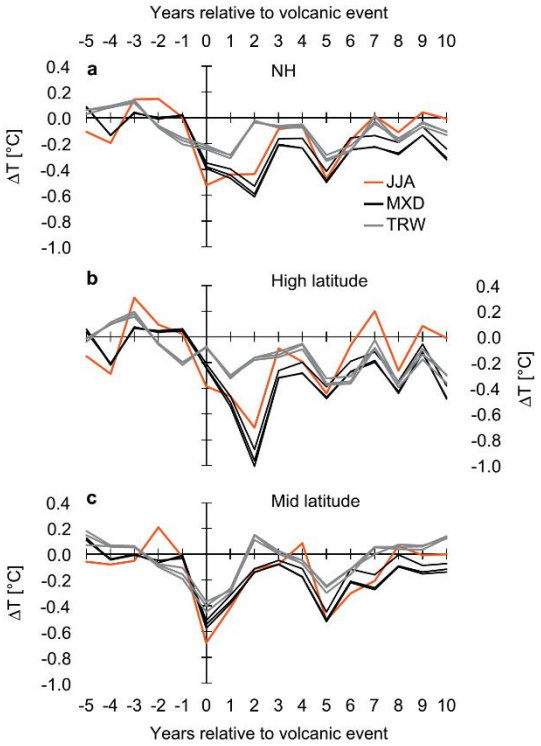


Figure 2-5 | Post-volcanic cooling in proxy and observational data. (a) SEA of the mean MXD (black), TRW (grey), and JJA temperature data (red) from 11 NH sites aligned by 13 volcanic eruptions since 1850. At each site, the events triggering > -0.5°C cooling in years 0, +1, or +2 in the observational data were considered. From a total of 136 potential cases in all 11 NH tree sites, only 85 (63%) reached this criterion, i.e. 51 cases did not exceed -0.5°C at a particular grid point. Single black and grey curves represent different detrendings (Hug/NegExp, Spline, RCS). (b) As in (a), but for the sites located north of 63°N. (c), As in (a), but for the sites locate south of 53°N.

constrained by the limited number of volcanic events leaving a cooling signature in the local instrumental data since 1850. Further explaining the spatially differing post-volcanic patterns in the observational and MXD data requires a spatially resolved analysis of NH temperature fields (Schneider et al., 2015), which is beyond the scope of this current analysis focusing on parameter-specific memory effects.

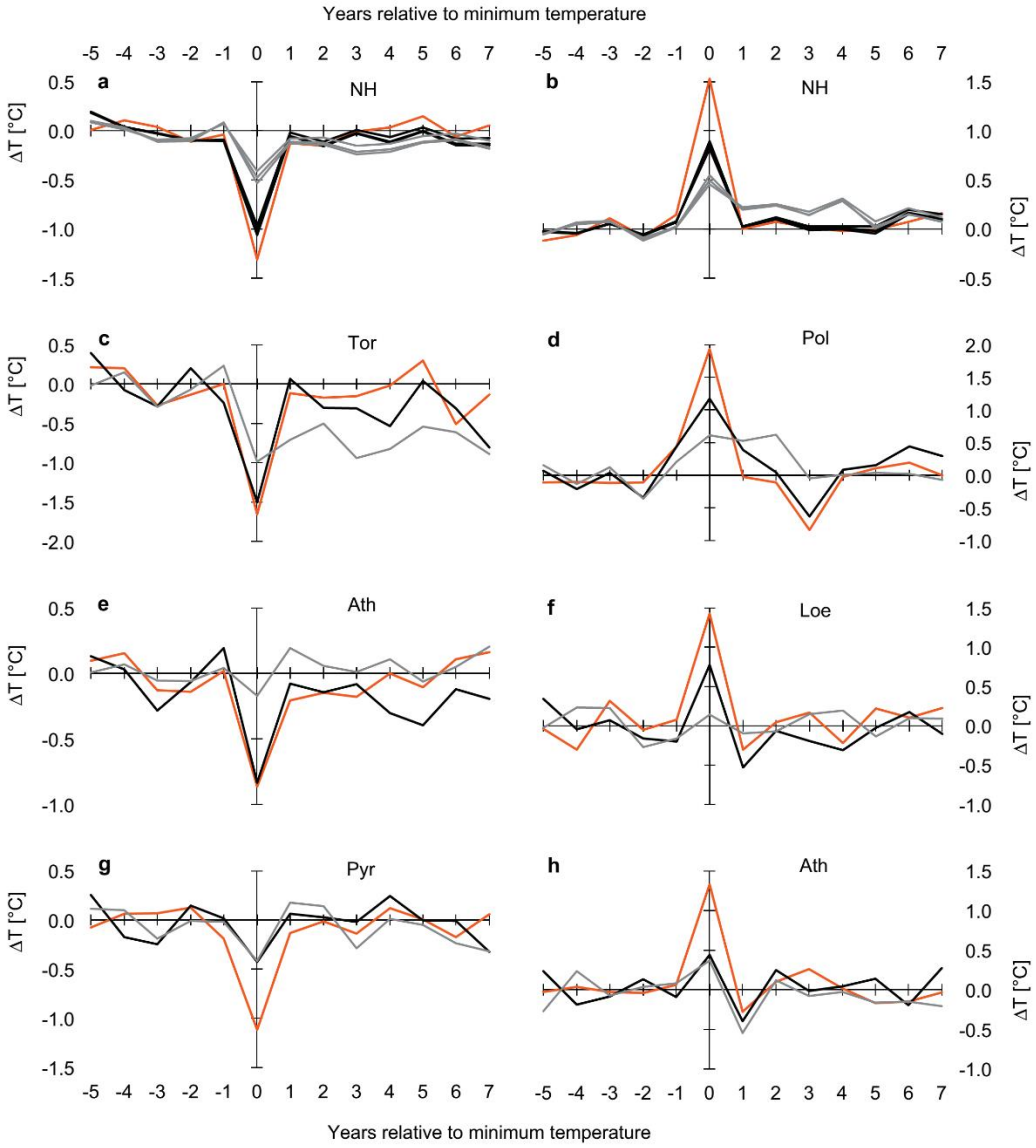


Figure 2-6 | Tree-ring deviations in extremely cold and warm years. (a) SEA of the mean MXD (black), TRW (grey), and JJA temperature (red) data aligned by the 15 coldest summers at the local grid points since 1850. (b) Same as in (a), but for the 15 warmest summers in each region. (c-h) MXD and TRW deviations in selected NH sites.

We finally show the SEA results, based on aligning the tree-ring data by the locally coldest and warmest summers since 1850, to reinforce the *(i)* reduced temperature amplitude retained in the TRW network, and *(ii)* large variability of TRW and MXD patterns among NH sites (Fig. 2-6). While the reduced climatically explained variance in the TRW chronologies ($\bar{r}^2$ = 0.11 with JJA temperatures) easily explains the severe underestimation of extreme JJA temperature deviations (ΔT = 0.84°C in cold and 1.03°C warm years; Fig. 2-6a-b), the spatially variable response patterns cannot be related to a single

controlling factor, such as species composition or site location. If anything, site variability remains as an unexplained characteristic of the proxy network: Tor (Sweden, pine) and Pol (Russia, larch) show good coherence in MXD, but memory effects in TRW (Fig. 2-6c-d); Ath (Canada, spruce) and Loe (Switzerland, larch) show good coherence in MXD, but no signal in TRW (Fig. 2-6e-f); and Pyr (Spain, pine) and Ath (Canada, spruce) show weak signals in both MXD and TRW (Fig. 2-6g-h). It is important to keep these differences in mind when aggregating larger proxy networks and interpreting their common response (Esper et al., 2012a).

2.4 Conclusion

Comparison of long MXD and TRW chronologies from 11 sites in the NH give insight into parameter-specific responses and memory effects. We found a temporally extended response (up to 6 years) in TRW subsequent to 29 volcanic eruptions of the past millennium. In contrast, the MXD response is strongest in the first post-volcanic year (-0.40°C on average) and does not substantially extend beyond the third post-volcanic year. Carbohydrate storage and remobilization likely control the parameter-specific memory effects, increasing the autocorrelation in TRW chronologies beyond the memory inherent to target summer temperature data. While the autocorrelation structure of tree-ring chronologies is highly dependent on the detrending method, memory in TRW is generally larger than in MXD over the first 10 to 20 lags.

Besides the memory effects, the TRW extremes also appear to be reduced in comparison to the MXD deviations over the past 1000 years. TRW-based cooling estimates only reach -0.17°C in the first post-volcanic year, compared to -0.40°C using MXD. Comparisons with instrumental summer temperature data over the post-1850 period support this conclusion, with the MXD estimates being in line with the observations, and the TRW estimates indicating smaller cooling effects. There is no sign that MXD systematically over or underestimates the cooling response to volcanic eruptions (Tingley et al., 2014). The assessments over this much shorter time interval also indicate a better correlation between TRW and observational data in mid latitude locations south of 53°N, but a diminished signal in high latitude locations north of 63°N. This conclusion, however, rests on fewer data points including only 13 volcanic events since 1850, and is not supported by the findings based on data over the past millennium. The latter indicates the high (not mid) latitude TRW data contain a stronger post-volcanic signal, though contaminated by memory effects in subsequent years.

Both characteristics of TRW – the increased memory (~ 6 years) and reduced variability (~ half of the MXD variance) – likely bias post-volcanic cooling estimates, and perhaps other disturbance signals (Esper et al., 2007a). While our study also indicates a large variability among tree sites, it demonstrates the strength of MXD data for assessing pulse-like disturbance events beyond the period covered by observational data. The application of individual detrending methods that emphasize higher frequency variance in resulting chronologies probably does not resolve the limitation of TRW to retain unbiased estimates of extreme events.

2.5 Acknowledgements

Most of the MXD data was measured at WSL in Birmensdorf, Switzerland. We thank all colleagues for developing millennial-length chronologies and making data available. Supported by the JGU Mainz "Inneruniversitäre Forschungsförderung".

2.6 References

Anchukaitis, K., et al., 2012. Tree rings and volcanic cooling. Nature Geoscience 5, 836–837.

Briffa, K.R., Jones, P.D., Bartholin, T.S., Eckstein, D., Schweingruber, F.H., Karlén, W., Zetterberg, P., Eronen, M. (1992) Fennoscandian summers from AD 500: temperature changes on short and long timescales. Climate Dynamics 7, 111–119.

Briffa, K.R., Jones, P.D., Schweingruber, F.H., Shiyatov, S.G., Cook, E.R., 1995. Unusual 20th-Century Summer Warmth in a 1,000-Year Temperature Record from Siberia. Nature 376, 156– 159.

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

Büntgen, U., Frank, D.C., Nievergelt, D., Esper, J., 2006. Summer temperature variations in the European Alps, AD 755–2004. Journal of Climate 19, 5606–5623.

Büntgen, U., Frank, D., Grudd, H., Esper, J., 2008. Long-term summer temperature variations in the Pyrenees. Climate Dynamics 31, 615–631.

Büntgen, U., et al., 2011. European climate variability and human susceptibility over the past 2500 years. Science 331, 578–582.

Bunde, A., Büntgen, U., Ludescher, J., Luterbacher, J., von Storch, H., 2013. Is there memory in precipitation? Nature Clim Change 3, 174–175.

Cole-Dai, J., 2010. Volcanoes and climate. WIREs Clim Change 1, 824–839.

Cook, E.R., 1985. A time series analysis approach to tree-ring standardization. Ph.D. Dissertation, University of Arizona, Tucson, 171pp.

Cook, E.R., Peters, K., 1981. The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies. Tree-Ring Bulletin 41, 45–53.

Cook, E.R., Kairiukstis, L.A. (Eds.), 1990. Methods of Dendrochronology: Applications in Environmental Science. Kluwer Academic Publishers, Dordrecht, 394 pp.

Cook, E.R., Peters, K., 1997. Calculating unbiased tree-ring indices for the study of climatic and environmental change. Holocene 7, 361–370.

Cook, E.R., Buckley B.M., D'Arrigo, R.D., Peterson M.J., 2000. Warm-season temperatures since 1600 BC reconstructed from Tasmanian tree rings and their relationship to large-scale sea surface temperature anomalies. Climate Dynamics 16, 79–91.

Cook, E.R., Palmer, J.G., D'Arrigo, R.D., 2002. Evidence for a 'Medieval Warm Period' in a 1,100 year tree-ring reconstruction of past austral summer temperatures in New Zealand. Geophysical Research Letters 29, doi: 10.1029/2001GL014580.

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

D'Arrigo, R., Wilson, R., Tudhope, A., 2009. Impact of volcanic forcing on tropical temperatures during the last four centuries. Nature GeoScience 2, 51–56.

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 118, doi: 10.1002/jgrd.50692.

Esper J (2014) Memory effects in tree-ring width and maximum latewood density in response to volcanic eruptions: evidence from Northern Fennoscandia. In: Wilson, R.J.S., et al. (Eds.), Tree rings in archaeology, climatology and ecology. TRACE, vol. 13, in press.

Esper, J., Cook, E.R., Schweingruber, F.H. 2002. Low-frequency signals in long tree-ring chronologies and the reconstruction of past temperature variability. Science 295, 2250–2253.

Esper, J., Cook, E.R., Krusic, P.J., Peters, K., Schweingruber, F.H., 2003. Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies. Tree-Ring Research 59, 81–98.

Esper, J., Frank, D.C., Wilson, R.J.S., Briffa, K.R., 2005. Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. Geophysical Research Letters 32, doi: 10.1029/2004GL021236.

Esper, J., Büntgen, U., Frank, D.C., Nievergelt, D., Liebhold, A., 2007a. 1200 years of regular outbreaks in alpine insects. Proceedings of the Royal Society B 274, 671–679.

Esper, J., Büntgen, U., Frank, D., Pichler, T., Nicolussi, K., 2007b. Updating the Tyrol tree- ring dataset. In: Haneca, K., et al. (Eds.), Tree rings in archaeology, climatology and ecology. TRACE, vol. 5, pp. 80–85.

Esper, J., Frank, D.C., Büntgen, U., Verstege, A., Hantemirov, R.M., Kirdyanov, A.V., 2010. Trends and uncertainties in Siberian indicators of 20th century warming. Global Change Biology 16, 386–398.

Esper, J., Büntgen, U., Timonen, M., Frank, D.C., 2012a. Variability and extremes of Northern Scandinavian summer temperatures over the past millennia. Global and Planetary Change 88-89, 1–9.

Esper, J., et al. 2012b. Orbital forcing of tree-ring data. Nature Climate Change 2, 862–866.

Esper, J., Büntgen, U., Luterbacher, J., Krusic, P., 2013a. Testing the hypothesis of post-volcanic missing rings in temperature sensitive dendrochronological data. Dendrochronologia 13, 216–222.

Esper, J., Schneider, L., Krusic, P.J., Luterbacher, J., Büntgen, U., Timonen, M., Sirocko, F., Zorita, E., 2013b. European summer temperature response to annually dated volcanic eruptions over the past nine centuries. Bulletin of Volcanology 75, 736, doi: 10.1007/s00445-013-0736-z.

Esper, J., Düthorn, E., Krusic, P., Timonen, M., Büntgen, U., 2014. Northern European summer temperature variations over the Common Era from integrated tree-ring density records. Journal of

Quaternary Science 29, 487–494.

Fischer, E.M., Luterbacher, J., Zorita, E., Tett, F.B., Casty, C., Wanner, H., 2007. European climate response to tropical volcanic eruptions over the last half millennium. Geophysical Research Letters 34, L05707, doi: 10.1029/2006GL027992.

Frank, D., Esper, J., 2005. Characterization and climate response patterns of a high-elevation, multi-species tree-ring Network in the European Alps. Dendrochronologia 22, 107–121.

Frank, D., Büntgen, U., Böhm, R., Maugeri, M., Esper, J., 2007a. Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quaternary Science Reviews 26, 3298–3310.

Frank, D., Esper, J., Cook, E.R., 2007b. Adjustment for proxy number and coherence in a large-scale temperature reconstruction. Geophysical Research Letters 34, doi: 10.1029/2007GL030571.

Franke, J., Frank, D., Raible, C.C., Esper, J., Bronnimann, S., 2013. Spectral biases in tree-ring climate proxies. Nature Clim Change 3, 360–364.

Fritts, H.C., 1976. Tree Rings and Climate. Academic Press, London, 567 pp.

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 113, D23111, doi: 10.1029/2008JD010239.

Gennaretti, F., Arseneault, D., Nicault, A., Perreault, L., Bégin, Y., 2014. Volcano-induced regime shifts in millennial tree-ring chronologies from northeastern North America. PNAS 111, 10077–10082.

Hammer, C.U., Clausen, H.B., Dansgaard, W., 1980. Greenland ice sheet evidence of post-glacial volcanism and its climatic impact. Nature 288, 230–235.

LaMarche, V.C., Hirschboeck, K.K., 1984. Frost rings in trees as records of major volcanic eruptions. Nature 307, 121–126.

Lavigne, F., et al., 2013. Source of the great A.D. 1257 mystery eruption unveiled, Samalas volcano, Rinjani Volcanic Complex, Indonesia. PNAS 110, 16742–16747.

Lee, T.C., Zwiers, F.W., Tsao, M., 2008. Evaluation of proxy-based millennial reconstruction methods. Climate Dynamics 31, 263–281.

Luckman, B.H., Wilson, R.J.S., 2005. Summer temperatures in the Canadian Rockies during the last millennium: a revised record. Climate Dynamics 24, 131–144.

Matalas, N.C., 1962. Statistical properties of tree ring data. Hydrological Sciences Journal 7, 39–47.

Melvin, T.M., Grudd, H., Briffa, K.R., 2013. Potential bias in updating tree-ring chronologies using regional curve standardisation: re-processing 1500 years of Torneträsk density and ring-width data. Holocene 23, 364–373.

Meko, D.M., 1981. Applications of Box-Jenkins methods of time series analysis to the reconstruction of drought from tree rings. Ph.D. Dissertation, University of Arizona, Tucson, 149pp.

Meko, D.M., et al., 2007. Medieval drought in the upper Colorado River basin. Geophysical Research Letters 34, doi: 10.1029/2007GL029988.

Myglan, V.S., Oidupaa, O.C., Vaganov, E.A., 2012a. A 2367-year tree-ring chronology for the Altai–Sayan region (Mongun-Taiga Mountain Massif). Archaeology, Ethnology and Anthropology of Eurasia 40, 76–83.

Myglan, V.S., Zharnikova, O.A., Malysheva, N.V., Gerasimova, O.V., Vaganov, E.A., Sidorov, O.V., 2012b. Constructing the Tree-Ring Chronology and Reconstructing Summertime Air Temperatures in Southern Altai for the Last 1500 Years. Geography and Natural Resources 33, 22–30.

Newhall, C.G., Self, S., 1982. The volcanic explosivity index (VEI): an estimate of explosive magnitude for historical volcanism. Journal of Geophysical Research 87, 1231–1238.

Oppenheimer, C., 2011. Eruptions that Shook the World. Cambridge University Press, Cambridge, 408 pp.

Pallardy, S.G. (2010) Physiology of Woody Plants. Academic Press, San Diego, 464 pp.

Panofsky, H.A., Brier, G.W., 1958. Some applications of statistics to meteorology. University Park, Pennsylvania, 224 pp.

Plummer, C.T., et al., 2012. An independently dated 2000-yr volcanic record from Law Dome, East Antarctica, including a new perspective on the dating of the c. 1450s eruption of Kuwae, Vanuatu. Climate of the Past 8, 1929–1940.

Robock, A., 2000. Volcanic eruptions and climate. Reviews of Geophysics 38, 191–219.

Salzer, M.W., Hughes, M.K., 2007. Bristlecone pine tree rings and volcanic eruptions over the last 5000 yr. Quaternary Research 67, 57–68.

Schneider, L., Smerdon, J.E., Büntgen, U., Wilson, R.J.S., Myglan, V.S., Kirdyanov, A.V., Esper, J. (2015) Revising northern hemispheric temperature history back to AD 600. Geophysical Research Letters 42, doi: 10.1002/2015GL063956.

Schweingruber, F.H., Fritts, H.C., Bräker, O.U., Drew, L.G., Schaer, E., 1978. The X-ray technique as applied to dendroclimatology. Tree-Ring Bulletin 38, 61–91.

Schweingruber, F.H., Bartholin, T., Schär, E., Briffa, K.R., 1988. Radiodensitometric–dendroclimatological conifer chronologies from Lapland (Scandinavia) and the Alps (Switzerland). Boreas 17, 559–566.

Schulman, E., 1956. Dendroclimatic Changes in Semiarid America. University of Arizona Press, Tucson, 142 pp.

Sear, C.B., Kelly, P.M., Jones, P.D., Goodess, C.M., 1987. Global surface temperatures responses to major volcanic eruptions. Nature 330, 365–367.

Self, S., Rampino, M.R., Barbera, J.J., 1981. The possible effects of large 19th and 20th century volcanic eruptions on zonal and hemispheric surface temperatures. Journal of Volcanology and Geothermal Research 11, 41–60.

Siebert, L., Simkin, T., Kimberly, P., 2011. Volcanoes of the world. University of California Press, London, 551 pp.

Smerdon, J.E., Coats, S., Ault, T.R., 2015. Model-dependent spatial skill in pseudoproxy experiments testing climate field reconstruction methods for the Common Era. Climate Dynamics, in press.

Timmreck, C., 2012. Modeling the climatic effects of large explosive volcanic eruptions. WIREs Climate Change 3, 545–564.

Tingley, M.P., Stine, A.R., Huybers, P., 2014. Temperature reconstructions from tree-ring densities overestimate volcanic cooling. Geophysical Research Letters 41, 7838–7845.

Tranquillini, W., 1964. The physiology of plants at high altitudes. Annual Reviews of Plant Physiology 15, 345–362.

Vaganov, E.A., Hughes, M.K., Shashkin, A.V., 2006. Growth Dynamics of Conifer Tree Rings: Images of Past and Future Environments. Springer, Berlin, 358 pp.

Von Storch, H., Zorita, E., Jones, J.M., Dimitriev, Y., González-Rouco, F., Tett, S.F., 2004. Reconstructing past climate from noisy data. Science 306, 679–682.

Wegmann, M., Brönnimann, S., Bhend, J., Franke, J., Folini, D., Wild, M., Luterbacher, J., 2014. Volcanic influence on European summer precipitation through monsoons: possible cause for “years without summer”. Journal of Climate 27, 3683–3691.

Wiles, G.C., D'Arrigo, R.D., Barclay, D., Wilson, R.S., Jarvis, S.K., Vargo, L., Frank, D., 2014. Surface air temperature variability reconstructed with tree rings for the Gulf of Alaska over the past 1200 years. The Holocene 24, 198–208.

3. European summer temperature response to annually dated volcanic eruptions over the past nine centuries

Jan Esper[1], Lea Schneider[1], Paul J. Krusic[2], Jürg Luterbacher[3], Ulf Büntgen[4], Mauri Timonen[5], Frank Sirocko[6], Eduardo Zorita[7]

[1] *Department of Geography, Johannes Gutenberg University, Mainz, Germany*

[2] *Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden*

[3] *Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus-Liebig University, Giessen, Germany*

[4] *Swiss Federal Research Institute WSL, Birmensdorf, Switzerland*

[5] *Finnish Forest Research Institute, Rovaniemi Research Unit, Rovaniemi, Finland*

[6] *Institute for Geoscience, Johannes Gutenberg University, Mainz, Germany*

[7] *Institute for Coastal Research, HZG Research Centre, Geesthacht, Germany*

Published in Bulletin of Volcanology 75:736

3.1 Introduction

Sulfate aerosols, from volcanic sulfur injected into the stratosphere by explosive eruptions, tend to cool global surface temperatures (Cole-Dai 2010). The aerosols scatter incoming solar radiation and absorb outgoing infrared radiation, thereby warming the lower stratosphere and cooling the earth's surface (Robock 2000). Explosive eruption plumes that pass the tropopause, where the temperature lapse rate reaches an abrupt minimum (~9-17 km asl.), cause large-scale changes in atmospheric optical depth and negative radiative forcing (McCormick et al. 1993). Eruptions of this size are typically classified as having a volcanic explosivity index (VEI) ≥ 5 (Newhall and Self 1982). The tephra volume of such eruptions is estimated to exceed one billion cubic meters.

Estimates of post-volcanic cooling are based on the analysis of surface temperatures following large eruptions (Self et al. 1981; Kelly and Sear 1984; Angell and Korshover 1985; Sear et al. 1987; Robock and Mao 1995). The number of VEI ≥ 5 eruptions captured within the modern instrumental period is small (n = 10, 1901-2012), thus limiting the confidence of estimates based solely on observational data. Estimating the degree of cooling by eruptions prior to the era of instrumental observation necessitates the use of annually resolved temperature proxies that explain a fraction of temperature variance of which only tree-ring, and a few documentary records, have the temporal precision and accuracy to provide adequate information over the past millennium (Frank et al. 2010). The suitability of tree-ring proxy data to detect the thermal signature of explosive eruptions, in space and time, has been successfully demonstrated (Briffa et al. 1998; Hegerl et al. 2003; Anchukaitis et al. 2012; Esper et al. 2013).

The Global Volcanism Program (GVP) has identified 37, annually dated, explosive eruptions in the Northern Hemisphere (NH) and tropics over the past 1000 years that likely injected sulfate aerosols into the stratosphere (Siebert et al. 2010). Though caution is required when working with these data as some of the eruptions have been dated using tree-ring records, which can lead to a circular reasoning when combining tree-ring reconstructed cooling estimates with eruption histories derived from the same proxy data. In addition, the sulfur emission magnitude, as well as the plume altitude, vary among VEI classified eruptions. Alternatively, histories of explosive eruptions derived from sulfate deposition in Greenland and Antarctic ice cores (Crowley 2000; Gao et al. 2008; Crowley and Unterman 2012) can be used to assess post-volcanic cooling (Ammann et al. 2007). However, this approach is constrained by dating uncertainties of the ice core acid layers that increases back in time and limits the temporal precision of inferred post-volcanic cooling estimates (Hammer et al. 1986; Traufetter et al. 2004; Baillie 2010).

The amplitude and duration of post-volcanic surface cooling is not well constrained and recently received critical examination (Anchukaitis et al. 2012; Mann et al. 2012; Esper et al. 2013). Hemispheric scale estimates, derived from observational and annually resolved proxy data (mainly tree-rings), range from ~ 0.0 to -0.4°C (Mass and Portman 1989; Briffa et al. 1998; Jones et al. 2003; D'Arrigo et al. 2009). It has been shown that the cooling signal is stronger during the summer season and in high European latitudes compared to lower latitudes (Fischer et al. 2007; Hegerl et al. 2011). Previous work, utilizing

temperature simulations from Energy Balance and Coupled General Circulation Models (CGCMs) indicate the frequency of stratospheric volcanic clouds may have triggered long-term temperature variations responsible for cold conditions during the Little Ice Age (LIA) in the seventeenth and early nineteenth centuries (Crowley 2000; Wagner and Zorita 2005; Hegerl et al. 2011). Other studies (Robock 2000; Grove 2001; Schneider et al. 2009; Miller et al. 2012) suggest the clustered volcanic eruptions in the thirteenth century, including the 1258/59 unknown event identified in ice core sulfuric acid depositions (Langway et al. 1988), contributed to the transition from the Medieval Warm Period (MWP) to the LIA about 700 years ago (see also Timmreck et al. 2009), a period during which a global reorganization of climate has been suggested (Graham et al. 2007, 2011; Trouet et al. 2009). However, the ability of CGCM's to accurately capture the dynamical response to stratospheric volcanic clouds is not without its own controversy (Stenchikov et al. 2006; Anchukaitis et al. 2010, Zanchettin et al. 2013a,b). An analysis of the dynamic response by twelve Coupled Model Intercomparison Project 5 (CMIP5) simulations to a suite of eruptions from the instrumental period indicated the models consistently overestimate tropical troposphere cooling leading to unstable pressure fields over high latitudes in the NH (Driscoll et al. 2012).

Here we present estimates of post-volcanic cooling over Northern and Central Europe derived from an updated network of tree-ring maximum latewood density (MXD) records covering the past 900 years (Büntgen et al. 2010; Esper et al. 2012a). Tree-ring MXD is a superior parameter for studying the effects of volcanic eruptions – compared to the more commonly used tree-ring width (TRW) measurements – as it is not biased by biological memory effects that tend to smear and lengthen the inferred TRW response to distinct climatic disturbances (Frank et al. 2007). We compare the temperature response to 34 of 37, annually dated and documented, VEI ≥ 5 eruptions, found in the summer temperature sensitive MXD network to the response found in a network of shorter instrumental records back to 1722 C.E. We also perform two sensitivity tests with subsets of volcanic eruptions representing *(i)* different VEI intensities, and *(ii)* the latitude of eruptions. We relate our cooling estimates from annually dated eruptions documented by the GVP with estimates derived from volcanic sulfate peaks identified in ice core records. Finally, our best estimates of post-eruption cooling are related to the annual summer temperature variance from 1111-1976 C.E. to evaluate the statistical significance of volcanic forcing in the context of natural climate variability over the past nine centuries.

3.2 Material and Methods

GVP and ice core data

Thirty-four annually dated large eruptions (VEI ≥ 5) from the NH and (NH and SH) tropics that occurred between 1111-1976 C.E. were used for assessing post-volcanic cooling effects (Table 3-1). Three eruptions (1480, 1482, 1800) that met these criteria were not considered, as these events were dated using dendrochronological methods (Siebert et al. 2010). The 34 eruptions have been precisely dated through documentary evidence and exceed VEI 4 above which stratospheric production of sulfate

Table 3-1

Year	Season	Volcano and Region	VEI[a]	Lat.	Tephra[c]	Long period (SEA1)	Short period (SEA2)	Large events (SEA3)	Small events (SEA4)	Extratropics (SEA5)	Tropics (SEA6)	Gao08 (NH)[d] (SEA7)
1262	[b]	Katla (Iceland)	5	64°N	1.5	✓		✓		✓		
1362	2	Oraefajokull (Iceland)	5	64°N	2.3	✓		✓		✓		
1471	4	Sakura-Jima (Japan)	5 [b]	32°N	1.3	✓			✓	✓		
1477	1 [b]	Bardarbunga (Iceland)	6	65°N	10	✓		✓		✓		
1563	2	Agua de Pau (Azores)	5 [b]	38°N	1	✓			✓	✓		
1586	[b]	Kelut (Indonesia)	5 [b]	8°S	1	✓			✓		✓	
1593	[b]	Raung (Indonesia)	5 [b]	8°S	1	✓			✓		✓	✓
1600	1	Huaynaputina (Peru)	6	17°S	30	✓		✓			✓	
1625	3	Katla (Iceland)	5	64°N	1.5	✓		✓		✓		
1630	3	Furnas (Azores)	5	38°N	2.1	✓		✓		✓		
1631	4	Vesuvius (Italy)	5 [b]	41°N	1.1	✓			✓	✓		
1640	3	Komaga-Take (Japan)	5	42°N	2.9	✓		✓		✓		
1641	1	Parker (Philippines)	5 [b]	6°N	1	✓			✓		✓	✓
1663	3	Usu (Japan)	5	43°N	2.8	✓		✓				
1667	3	Shikotsu (Japan)	5	43°N	3.4	✓		✓				
1673	2	Gamkonora (Indonesia)	5 [b]	1°N	1	✓			✓		✓	✓
1680	[b]	Tongkoko (Indonesia)	5 [b]	2°N	1	✓			✓		✓	
1707	4	Fuji (Japan)	5	35°N	2.1	✓		✓				
1721	2	Katla (Iceland)	5 [b]	64°N	1.2	✓			✓			
1739	3	Shikotsu (Japan)	5	43°N	4	✓	✓	✓				
1755	4	Katla (Iceland)	5 [b]	63°N	1.5	✓	✓	✓				✓
1815	2	Tambora (Indonesia)	7	8°S	160	✓	✓	✓				✓
1822	4	Galunggung (Indonesia)	5	7°S	1	✓	✓		✓		✓	
1835	1	Cosiguina (Nicaragua)	5	13°N	5.7	✓	✓	✓			✓	✓
1854	1	Shiveluch (Russia)	5	56°N	2	✓	✓	✓		✓	✓	
1875	1	Askja (Iceland)	5	65°N	1.8	✓	✓	✓		✓		
1883	3	Krakatau (Indonesia)	6	6°S	20	✓	✓	✓			✓	✓
1902	4	Santa Maria (Guatemala)	6 [b]	15°N	20	✓	✓	✓			✓	
1907	2 [b]	Ksudach (Russia)	5	52°N	2.8	✓	✓	✓		✓		
1912	2	(USA)	6	58°N	28	✓	✓	✓		✓		✓
1913	1	Colima (Mexico)	5	19°N	1.7	✓	✓	✓			✓	
1933	1	Kharimkotan (Kuril Isl.)	5	49°N	1	✓	✓		✓	✓		
1956	1	Bezymianny (Russia)	5	56°N	2.8	✓	✓	✓		✓		
1963	1	Agung (Indonesia)	5	8°S	1	✓	✓		✓		✓	✓

Table 3-1 (continued)

Year	Season	Volcano and Region	VEI[a]	Lat.	Tephra[c]	Long period (SEA1)	Short period (SEA2)	Large events (SEA3)	Small events (SEA4)	Extratropics (SEA5)	Tropics (SEA6)	Gao08 (NH)[d] (SEA7)
Number of events						34	15	22	12	21	13	9
Mean tephra vol. (billion cubic meters)						9.4	16.9	14.0	1.1	3.7	18.8	(23.2)[e]
JJA temperature response		MXD-north (lag +2)				−0.52	−0.65	−0.46	−0.61	−0.41	−0.68	−0.15
MXD-central (lag +1)						−0.18	−0.29	−0.17	−0.19	−0.01	−0.44	−0.18

The 34 annually dated and documented volcanic eruptions (VEI index≥5) in the NH extratropics and tropics from 1111–1976 CE. SEAs 2–7 indicate subsets of events used to estimate JJA temperature responses in European MXD chronologies. Bottom two lines summarize the temperature response in the second post-eruption year in Northern Europe and the first post-eruption year in Central Europe

[a] Volcanic explosivity index

[b] Indicates uncertain assignments.

[c] Estimated volume in billion cubic meters

[d] Ice–core-derived volcanic sulfate deposition signals from Gao et al. (2008). Nine of the total of 40 NH volcanic events reconstructed in Gao et al. (2008) for the 1111–1976 period match the VEI≥4 events listed here. All 40 NH deposition events were used in the SEA

[e] Mean NH stratospheric sulfate aerosol injection (in *Tg*) according to Gao et al. 2008

aerosols is expected based on descriptions of eruption type, duration, and column height (Newhall and Self 1982). The estimated tephra ejecta of these eruptions range from 1 to 160 10^9 m^3 (Table 3-1).

In addition to the 34 VEI ≥ 5 events (SEA 1 in Table 3-1) subsets of eruptions were tested to evaluate the significance of an eruption's size and the volcano's location on observed cooling patterns. These subsets include *(i)* 15 eruptions within the shorter 1722-1976 period, covered by long instrumental temperature data (SEA 2), *(ii)* 22 eruptions ≥ 1.5 10^9 m^3 tephra volume, and 12 eruptions < 1.5 10^9 m^3 tephra volume (SEAs 3 and 4), and *(iii)* volcanoes located in the NH extratropics and tropics (SEAs 5 and 6). Note that the average tephra volume of the tropical volcanoes is much larger (18.8 10^9 m^3) than the extratropical volcanoes (3.7 10^9 m^3).

Finally, we considered a timeseries of sulfate aerosol layers derived from multiple Greenland and Antarctic ice cores (Gao et al. 2008) identifying 40 NH stratospheric events between 1111-1976 C.E. (SEA 7). This record contains a number of major eruptions that are not documented by the GVP, including the 1452/53 Kuwae and 1258/59 unknown events, but are identified in sulfate depositions and tree-ring chronologies (LaMarche and Hirschboeck 1984; Gao et al. 2006; Salzer and Hughes 2007). Though these ice-core derived data are typically used to force CGCMs, the dating and location of a number of eruptions, particularly during the earlier part of the past 900 years, is not certain (Hammer et al. 1986; Baillie 2008, 2010; Plummer et al. 2012; Sigl et al. 2013). This condition might compromise the temporal

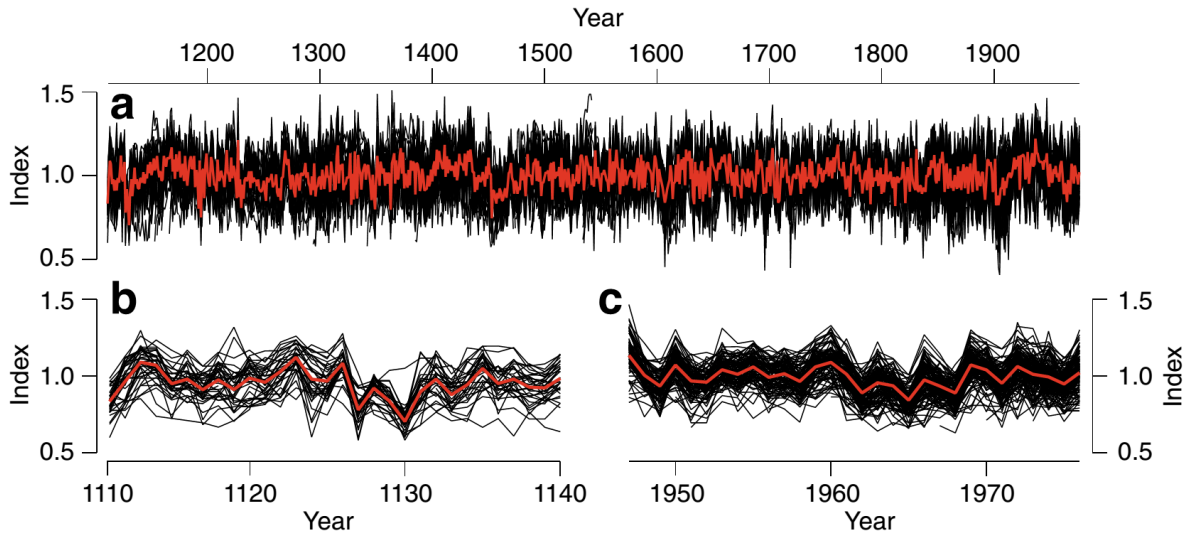


Figure 3-1 | NSC maximum latewood density data. **a** NegExp detrended single MXD measurement series (black) shown together with their bi-weighted robust mean (red) over the 1111-1976 C.E. period. **b** and **c**, Same as in **a**, but shown over the earliest (1111-1140 C.E.) and latest (1947-1976 C.E.) 30-year periods

precision of any post-volcanic, climate assessment using such data. Only nine of the 40 ice core derived eruptions identified in Gao et al. (2008) coincide with a documented VEI ≥ 5 event during the 1111-1976 C.E. period (Table 3-1).

Tree-ring maximum latewood density chronologies

Documentary and ice core derived volcanic events were used to assess pre- and post-eruption June-August (JJA) temperature deviations reconstructed from European MXD chronologies spanning the past 900 years. An MXD chronology is the mean of a collection of MXD measurement series belonging to individual trees growing in an ecologically homogeneous site (Cook and Kairiukstis 1990). Typically, two such measurement series, representing two radii of a stem, are procured from each tree. The raw MXD series (in g/cm^3) need to be detrended/standardized to remove level differences between biologically younger and older tree-rings, which possess slightly denser and lighter latewood, respectively (Schweingruber et al. 1978). This is done by fitting negative exponential curves (NegExp) to the individual measurement series (radii) and calculating ratios between the raw density measurements and the curve values (Cook and Kairiukstis 1990). The procedure removes non-climatic, tree age-related trends and emphasizes common variations.

To produce a millennium-length chronology, MXD radial patterns from living trees, which typically represent the most recent 200-400 years, are crossdated (Douglass 1920) with patterns from relict trees (Büntgen et al. 2011). In the case of the MXD dataset from Northern Scandinavia (Fig. 3-1, NSC), relict material was obtained from trees that fell some hundred years ago into shallow lakes in Finnish Lapland and were preserved (Esper et al. 2012b). In other chronologies used in this study (see below), living trees were combined with historical timbers from old buildings (e.g. the Lötschental, Switzerland; Büntgen et al. 2006) or dry-dead wood in talus (e.g. the Pyrenees, Spain; Büntgen et al. 2008). Latewood

cell-wall growth in these cold environments is controlled by summer temperature (Moser et al. 2010), imprinting a common variance among all single MXD measurement series at a given site (Figs. 3-1b and c). The coherence among individual measurement series is typically higher in MXD compared to TRW data (Esper et al. 2010). The common signal strength of tree-ring chronologies is also controlled by the number of integrated measurement series, which varies among sites and typically decreases back in time (in Figs. 3-1 b and c: 114 series over the recent 1947-1976 C.E. and 34 series over the early 1111-1140 C.E. periods).

Seven NegExp detrended MXD site chronologies, from latitudinal and elevational treeline environments in Northern and Central Europe, were used to assess the spatial and temporal temperature patterns associated with large volcanic eruptions (Table 3-2). The site chronologies are composed of *Pinus sylvestris* from Central (JAE) and Northern Scandinavia (TOR, NSC), *Pinus uncinata* from the Pyrenees (PYR), and *Larix decidua* and *Picea abies* from the Alps (LAU, LOE, TIR). The average number of MXD measurement series, over the common period 1111-1976 C.E., varies considerably among these datasets ranging from 18 in TOR to 49 in NSC. The lag 1 autocorrelation, a measure of the temporal persistence in a timeseries, is < 0.38 in all MXD site chronologies matching the memory inherent to instrumental JJA temperature data from the European stations used in this study (see below).

Table 3-2 | European MXD chronologies. Period refers to the time span during which replication exceeds two MXD measurement series (-181 denotes 181 B.C.). Mean replication is the average number of MXD measurement series over the 1111-1976 C.E. common period. Lag 1 autocorrelation is calculated for the NegExp detrended chronologies over the same period. MXD-north and MXD-central are the mean timeseries of the three MXD site chronologies from Northern Europe (JAE, TOR, NSC) and the four MXD site chronologies in Central Europe (PYR, LAU, LOE, TIR).

MXD chronology	Country	Species	Period	Mean replication (1111–1976)	Lag 1 autocorrelation	Source
Jaemtland (JAE)	Sweden	Pine	1111–1978	29	0.12	Gunnarson et al. 2010
Tornetraesk (TOR)	Sweden	Pine	452–2004	18	0.17	Grudd 2008
N-Scan (NSC)	Finland	Pine	−181–2006	49	0.25	Esper et al. 2012b
MXD-north			1111–1976	96	0.18	
Pyrenees (PYR)	Spain	Pine	1044–2005	46	0.00	Büntgen et al. 2008
Lauenen (LAU)	Switzerland	Spruce	996–1976	22	0.03	Schweingruber et al. 1988
Lötschental (LOE)	Switzerland	Larch	743–2004	46	0.37	
Tirol (TIR)	Austria	Spruce	1047–2003	33	0.07	Esper et al. 2007b
MXD-central			1111–1976	147	0.10	

The regional mean timeseries, MXD-north and MXD-central, were calculated by averaging all the northern and central site chronologies (Fig. 3-2). This consolidation is justified by the significantly high inter-site correlations among the three northern (r_{north} = 0.57) and the four central sites ($r_{central}$ = 0.46) over the common period 1111-1976 C.E. It is important to note that the northern *versus* central site chronologies share no common variance (see the grey curve in Fig. 3-2d centered at r = 0.03) reflecting the distinct climatic dipole structure that exists over Europe as a consequence of internal climate forcings (Barnston and Livezey 1987). Inter-site correlations also decrease back in time – particularly among the

northern sites (see the blue curve in Fig. 3-2d) – likely due to declining sample sizes in the site chronologies. This latter feature points to a weaker climatic signal in the site chronologies and subsequent regional composites during the earliest centuries of the past millennium. Replication of the entire European MXD network declines from 426 measurement series in 1973 to 87 series in 1111 C.E. (Fig. 3-2c).

Instrumental temperature data and calibration of MXD records

The MXD site and regional chronologies were transformed into estimates of average JJA temperature variability by scaling (adjusting the mean and variance; Esper et al. 2005) each chronology against the average JJA temperature of the nearest grid point in the Crutem4 temperature dataset (Jones et al. 2012) over the common period 1901-1976 (Table 3-3). The correlations between MXD chronology and JJA temperatures, at their respective grid points, are lower in Central Europe (ranging from 0.31 to 0.61) than in Northern Europe (0.71 to 0.82), indicating an overall weaker inherent climate signal in the central portion of the network. This tendency is confirmed by the correlations, calculated over a much longer time period (1722-1976), between the mean JJA temperatures recorded at the Stockholm and Uppsala stations and the northern MXD chronologies, to the corresponding correlations computed for Central European chronologies and the long Central England, De Bilt, and Berlin station records (see last column in Table 3-3; Table S3-1).

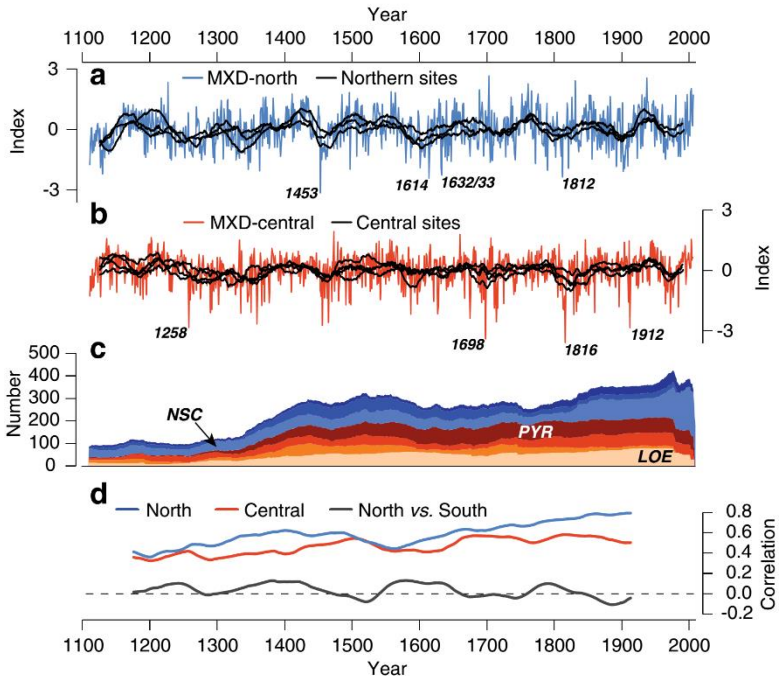


Figure 3-2 | European maximum latewood density records. MXD site chronologies (black) from a Northern (JAE, TOR, NSC) and b Central Europe (PYR, LAU, LOE, TIR) over their common period 1111-1976 C.E. Records were smoothed using a 30-year filter. Blue and red curves are the regional mean timeseries derived from averaging the unsmoothed site records in Northern (blue) and Central Europe (red) respectively. The years of the four most negative deviations are labeled. c Temporal sample depth of all MXD measurement series (stem radii) within each site chronology in Northern (bluish colors) and Central Europe (reddish colors). The well-replicated site chronologies in Northern (NSC) and Central Europe (PYR, LOE) are labeled. d 100-year running inter-site correlations among the three northern (blue) and central site chronologies (red), and between the northern and central regional records (grey).

Comparison of the spatial patterns of MXD summer temperature signals (Fig. 3-3) and the spatial patterns of the long European station record's summer temperature signals (Fig. S3-1) reveals increasing distance between the proxy sites and station locations is an additional source of correlation decay. The significant correlations ($p < 0.05$) between the northern MXD data and the gridded temperature data are spatially more homogeneous, reaching southward to a line across Northern Germany towards Ukraine. The significant portions of the overall weaker and more heterogeneous patterns of the Central European MXD data are centered over the Alps reaching into the central Mediterranean and the Balkans. The spatial overlap between the correlation patterns of the central MXD sites and the long station records (Figs. 3-3 and S3-1) indicates that the distance between proxy and station data affects the correlation results over the long 1722-1976 period in Central Europe, which is particularly obvious for the Mediterranean PYR site (Tab. 3-3: $r_{Crutem4} = 0.40$; $r_{Stations} = 0.17$). A similar feature is seen in Northern Europe, where the JAE site correlates lower than the TOR and NSC sites with the nearest grid points, but correlates better than the far northern MXD sites (TOR, NSC) with the Uppsala and Stockholm stations located in southern Sweden. These spatial associations help explain the overall better fit between the MXD-north mean timeseries and the station derived mean timeseries (JJA-north; see Fig. S3-2), compared to the MXD-central mean timeseries *versus* the Central European station mean (JJA-central). As the distance between proxy and station locations in the central portion of the network is larger, and their association is weaker, somewhat less coherent results should be expected when estimating post-volcanic cooling effects from the MXD-central and JJA-central data over the common 1722-1976 period.

Table 3-3 | JJA temperature signals of European MXD chronologies. Pearson correlation coefficients of the MXD site and regional mean chronologies with JJA temperatures of the nearest grid points from the Crutem4 dataset (Jones et al. 2012) over the 1901-1976 period, together with the correlations with JJA mean temperatures of the long station records in Northern Europe (Uppsala and Stockholm) and Central Europe (Central England, De Bilt, and Berlin) over the 1722-1976 period.

MXD chronology	Crutem4 grid point	Correlation with Crutem4 (1901–1976)	Correlation with long station record (1722–1976)
JAE	67.5°N/12.5°E	0.71	0.59
TOR	67.5°N/22.5°E	0.82	0.56
NSC	67.5°N/22.5°E	0.76	0.52
MXD-north	Mean	0.80	0.61
PYR	42.5°N/2.5°E	0.40	0.17
LAU	47.5°N/7.5°E	0.31	0.29
LOE	47.5°N/7.5°E	0.61	0.42
TIR	47.5°N/12.5°E	0.41	0.27
MXD-central	Mean	0.52	0.36

Coupled general circulation models

In addition to the European MXD and long instrumental station records, we used four millennium-long JJA temperature histories simulated by three CGCMs for the assessment of post-volcanic cooling effects (Supplementary Material). CGCM runs are typically used to attribute the influence of natural and anthropogenic forcings on climate, including the effects of explosive volcanism (Schneider et al. 2009).

The simulations considered here include two millennium-long runs of the ECHO-G model (denoted Erik1 and Erik2; Zorita et al. 2005), as well as combined runs of the Max-Planck-Institute Earth System Model Paleoclimate version (MPI-ESM-P) and the Community Climate System Model version 4 (CCSM4; Gent el al. 2011) downloaded from the CMIP5 archive (Taylor et al. 2012; Fernández-Donado et al. 2013).

We extracted and averaged the simulated temperatures from each model run at five grid points in the vicinity of the northern MXD and station sites to produce a composite, simulated, JJA timeseries (CGCM-north; Fig. S3-3). The same procedure was applied to the seven grid points in vicinity to the central MXD chronologies and their corresponding long central stations (CGCM-central). The model composites are later used for comparison with the proxy derived, volcanic cooling estimates from 1111-1976. Note the simulated temperatures correlate only weakly between the four model runs in Northern Europe ($R_{1111\text{-}1976}$ = 0.12) and Central Europe ($R_{1111\text{-}1976}$ = 0.08), possibly related to the limited geographical region and the varying external forcings used in each model. Whereas the CCSM4 run has been forced with the aerosol deposition data from Gao et al. (2008), Erik1 and Erik2 were forced using eruption estimates from Crowley (2000), and MPI-ESM-P with estimates from Crowley et al. (2008). The simulated summer temperatures also indicate slightly differing long-term trends from 1722-1976, as compared to the JJA-north and JJA-central station means (Fig. S3-3).

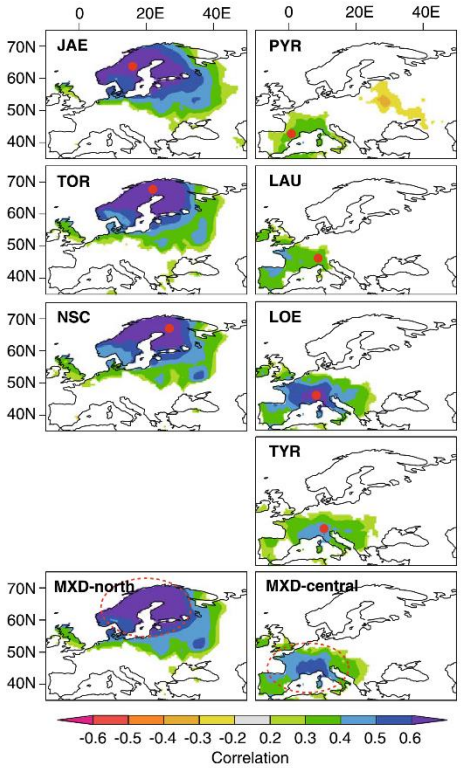


Figure 3-3 | MXD temperature signals. Maps showing the correlation patterns of MXD site chronologies (red dots) with gridded JJA mean temperatures (Mitchell and Jones 2005) over the common 1901-1976 period ($p < 5\%$). Bottom panels indicate the results for the regional mean timeseries, MXD-north and MXD-central.

Superposed epoch analysis (SEA)

To assess post-volcanic cooling, we used SEA (Panofsky and Brier 1958) with *(i)* the temperature-transformed MXD site chronologies and their regional means (MXD-north, MXD-central), *(ii)* the long instrumental station records and their means (JJA-north, JJA-central), and *(iii)* the simulated JJA temperatures of the four CGCM runs and their means (CGCM-north, CGCM-central). In this experiment, the five years before and after a volcanic eruption are analyzed. Instrumental JJA temperature measurements, and their MXD-based and CGCM estimates, are expressed as anomalies with respect to the mean of the five years preceding the eruptions (years -5 to -1). SEA is applied to the 34 annually dated VEI $\geq$ 5 events, documented by the GVP (Siebert et al. 2010), over the 1111-1976 C.E. period, as well as the five additional subsets of those eruptions (SEA2-6 in Table 3-1). We also considered 40 volcanic events derived from sulfate aerosol layers in Greenland and Antarctic ice cores (SEA 7) and those eruptions used in a previous NH, living-tree MXD study of cooling patterns by Briffa et al. (1998) (Supplementary Material).

3.3 Results

Analysis of millennial-length MXD chronologies revealed severe post-volcanic summer cooling in Northern Europe and a reduced, but temporally extended, response in Central Europe associated with 34, precisely located and dated, large eruptions between 1111-1976 C.E. (Fig. 3-4). Northern European JJA temperatures, in year 1 and year 2 after the volcanic events, are -0.28 and -0.52°C. The individual MXD site chronologies from Scandinavia indicate fairly homogeneous patterns in these years (see the thin curves in Fig. 3-4) with a spread about their mean departure as small as ±0.08 °C at lag +1 and ±0.05°C at lag +2. Summer temperatures in Northern Europe rebound to +0.25°C by the fourth post-volcanic year. Cooling in Central Europe lasts until the fourth post-volcanic year (minimum at lag +1 = -0.18°C) and gradually returns to +0.02°C in year five after eruptions. However, relative to the temperature variations prior to stratospheric events (years -5 to -1), only the post-volcanic response in Northern Europe appears exceptional. In Central Europe, the post-volcanic deviations do not differ significantly from the centralized pre-volcanic estimates.

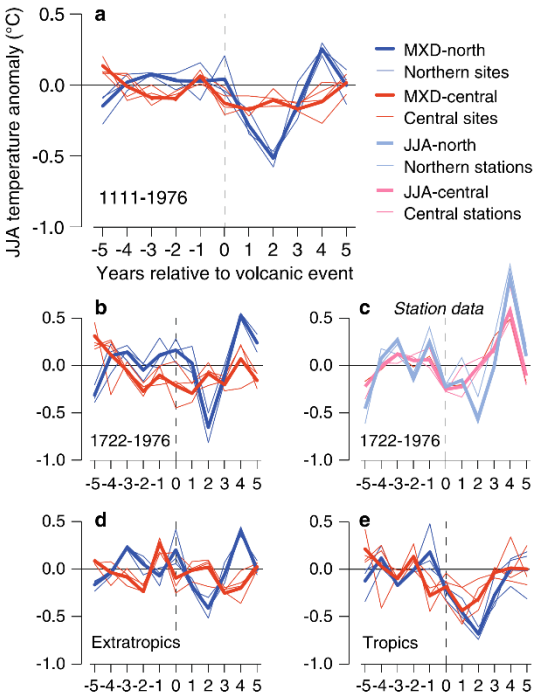


Figure 3-4 | Superposed Epoch Analyses centered on large volcanic eruptions of the past nine centuries. a JJA temperature patterns of MXD-north (blue) and MXD-central (red) five years before and after the 34 large volcanic eruptions (VEI index ≥ 5) within the 1111-1976 C.E. period (SEA1 in Table 1). Thin curves are the SEA timeseries of the individual MXD site records JAE, TOR, and NSC in Northern Europe, and PYR, LAU, LOE, and TIR in Central Europe. b Same as in a, but for the 15 eruptions of the 1722-1976 C.E. period (SEA2). c, Same as in b, but using the JJA instrumental temperatures (instead of the MXD-derived estimates). d and e, Same as in a, but for the 21 eruptions located in the NH extratropics and 13 eruptions in the (NH and SH) tropics, respectively. All SEA timeseries expressed as temperature anomalies with respect to the five years preceding the volcanic events (lags -5 to -1).

The 1111-1976 C.E. MXD derived, SEA temperature estimates (Fig. 3-4a) are strikingly similar to those found in both the 1722-1976 MXD and 1722-1976 instrumental station records (Figs. 3-4b and c), though the spread of SEA temperatures over the shorter period that contains 15 eruptions is larger. In Northern Europe, the dominating feature is the strong cooling in the second post-volcanic year, followed by a dramatic warming (+0.93°C station data) in the fourth. A similar pattern is evident in Central Europe where the JJA-central cooling signal (minimum -0.22°C) in year +1 vanishes among the pre- and post-volcanic temperature variations. Differences between the Central European MXD and instrumental SEA patterns, especially the station's positive anomaly in the fourth post-eruption year, are likely related to *(i)* the varying spatial coverage of the central MXD and observational data (Figs. 3-3 and S3-1), *(ii)* the unexplained temperature variance in the proxy data (larger in Central Europe compared to Northern Europe; Table 3-3), and *(iii)* the reduced number of VEI ≥ 5 volcanic events since 1722 C.E. (n = 15), producing larger uncertainties (e.g., increased variance of the Central European MXD site's response in Fig. 3-4b).

The SEA results reveal stronger post-volcanic responses to tropical eruptions as compared to NH events (Figs. 3-4d and e) over the past 900 years, but relatively minor differences as a consequence of eruption size (i.e., < 1.5 vs. ≥ 1.5 10^9 m^3 tephra volume; Supplementary Material). It remains unclear whether the increased tropical eruption signature is due to the volcano's location – and associated increased

stratospheric transport (Trepte and Hitchman 1992) – or driven by eruption size, as the mean tephra volume of the low latitude events (18.8 10^9 m^3) is much larger than the high latitude events (3.7 10^9 m^3). Also varying sulfur contents might contribute to the differentiation between Tropical and NH eruptions.

The strong temperature cooling found in Northern Europe following VEI ≥ 5 eruptions diminishes if ice-core derived volcanic events (Gao et al. 2008) are considered in the SEA (Fig. 3-5; details in Fig. S3-4), pointing to the importance of utilizing annually dated eruption data when assessing post-volcanic effects. Of course, the simulated Northern European summer temperatures indicate severe post-volcanic cooling in response to the Gao et al. (2008) events, if the CGCM (here CCSM4) has been forced with the same aerosol injection estimates (Fig. S3-5). In this case, post-volcanic cooling is much larger in the regional CCSM4 output (-0.90°C and -0.80°C at lags +1 and +2) than the cooling seen in both the MXD and instrumental data. The overall variance among the four CGCMs considered in the SEAs is significantly high, compared to the proxy and observational data, pointing to the limited validity of simulated temperatures at the scale of continental Europe (Supplementary Material).

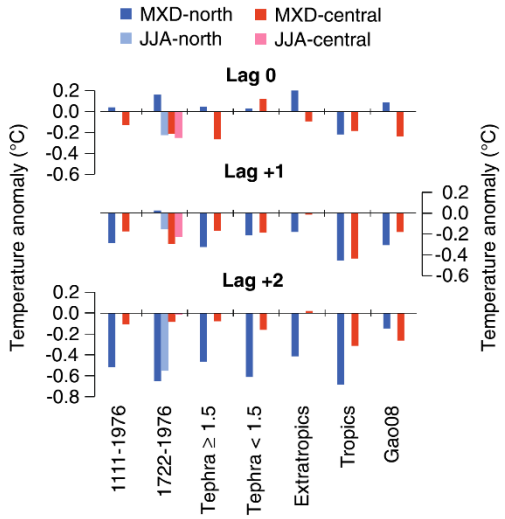


Figure 3-5 | Summarized SEA results for stratospheric volcanic events at lag 0, +1, and +2 in the MXD-north (blue), MXD-central (red), JJA-north (light blue), and JJA-central (light red) datasets.

The assessment of post-volcanic cooling in the context of the full spectrum of summer temperature variance over the 1111-1976 C.E. (n = 866 years) and 1722-1976 C.E. (n = 255 years) periods indicates that only mean deviations at lag +2 in Northern Europe differ significantly ($p < 0.05$; Mann–Whitney–Wilcoxon test) from the mean of all years (Fig. 3-6). Temperature cooling in year 1 after the stratospheric event – the most striking signal found in Central Europe – is not significantly different from the mean of all years, even if the overall variance of summer temperatures is less in Central Europe compared to Northern Europe (see the density functions in Fig. 3-6). Visualization of the annual temperature estimates demonstrates that *(i)* a number of post-volcanic JJA anomalies are actually positive (i.e., on the right side of the centered distributions), and *(ii)* there are frequent cool years that are not associated with stratospheric volcanic events. The latter finding is likely constrained by the incompleteness of the

volcanic record particularly during the earlier centuries of the past 900 years, albeit this argument is not valid for the shorter 1722-1976 C.E. period. The positive deviations point to the importance of "unforced" internal variability of the climate system at the European scale (Jungclaus et al. 2010).

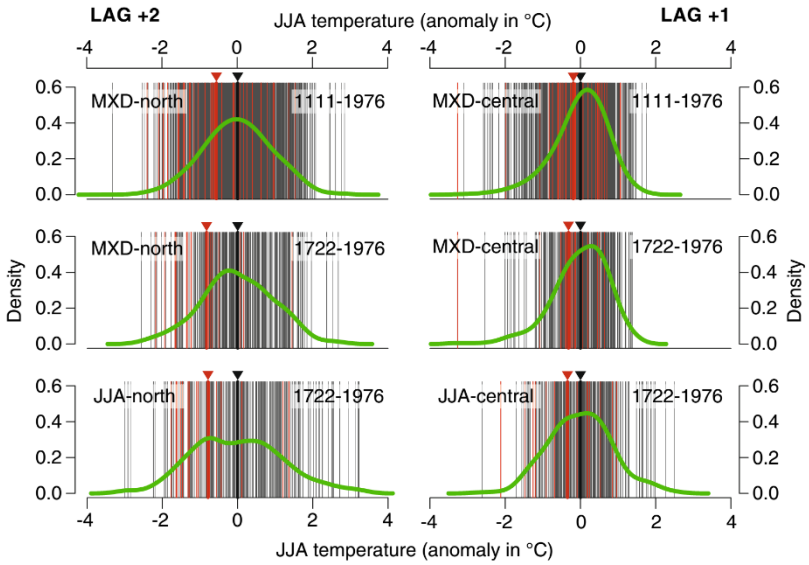


Figure 3-6 | Distributions of reconstructed and recorded JJA temperatures over the 1111-1976 and 1722-1976 C.E. periods. Left column shows temperatures in Northern Europe two years after volcanic eruptions (SEA1: lag +2), right column shows temperatures in Central Europe one year after volcanic eruptions (SEA1: lag +1). Green curves indicate density functions (bandwidth = 0.3) of JJA temperature anomalies with respect to the 1111-1976 and 1722-1976 periods (thin grey and red lines; 866 years in the top panels, 255 years in the middle and bottom panels). Red lines indicate summer temperatures in 34 post-volcanic years (lag +2 in the left, and lag +1 in the right column). Bold red lines and triangles indicate the mean temperature of these lag years. Bold black lines and triangles indicate the mean temperature of all years. Results are for MXD-based (top and middle panels) and observational (bottom panels) JJA temperatures.

3.4 Discussion and Conclusions

The analysis of an MXD network covering the past 900 years, and comparison with long instrumental records since 1722 C.E., revealed severe summer temperature cooling two years after stratospheric volcanic clouds in Northern Europe and a generally weaker response in Central Europe. This spatial pattern supports findings based on a compilation of shorter proxy and instrumental records (including documentary evidence) in response to selected tropical eruptions (Fischer et al. 2007). However, the thermal cooling reported here, based on a complete set of annually dated VEI $\geq$ 5 eruptions from the NH extratropics and tropics, is weaker than that reported in Fischer et al. (2007) and, in Northern Europe, delayed by one year (lag +2 instead of lag +1). Tests with respect to *(i)* eruption size (1–1.5 vs. $\geq$ 1.5 10^9 m^3 tephra volume), *(ii)* volcano location (NH vs. tropics), *(iii)* time period (1111-1976 vs. 1722-1976 C.E.), and *(iv)* volcanic forcing data (documentary vs. ice core reconstructed) demonstrate sensitivity of the cooling estimates to the selection criterion of eruptions. The marginal post-volcanic signals in both the Northern European MXD data and the long instrumental station data, in response to ice core derived sulfate deposition signatures (Gao et al. 2008), suggests caution should be used when considering these volcanic forcing estimates in CGCM studies (Solomon et al. 2007).

Documented versus ice core reconstructed volcanic histories

The documented and annually dated eruption data used here for the assessment of post-volcanic cooling indicates a higher frequency of stratospheric events during the more recent centuries of the past 900 years. There is a noticeable reduction of VEI ≥ 5 eruptions before 1450 C.E. (n = 2 events; Table 3-1), likely caused by incomplete documentary evidence from sparsely populated regions prior to the 16th century. During this early period a number of major volcanic events, including the 1258/59 unknown (Zielinski 1995) and 1452/53 Kuwae eruptions (Hammer et al. 1980; Sigl et al. 2013), have been identified in ice core acid layers from Greenland and Antarctica (Oppenheimer 2003; Kurbatov et al. 2006). These events are represented in the Gao et al. (2008) sulfate aerosol injection estimates used here for comparison of regional scale cooling effects. However, throughout the 1111-1976 C.E. period, only nine of the 40 NH and tropical stratospheric events included in Gao et al. (2008) match the annually dated VEI ≥ 5 events recognized by the GVP (Siebert et al. 2010).

The conclusion from these cooling estimates, based on documented *versus* ice core reconstructed volcanic histories, is somewhat ambivalent. The potentially missing stratospheric events during earlier centuries of the past 900 years suggests cooling estimates from documented eruptions are too small, yet the substantially reduced Northern European cooling obtained from the ice core derived events contradicts this qualification. This conflicting result is likely related to dating uncertainties inherent to the ice core data (Hammer et al.1986; Robock and Free 1995) biasing the SEA-derived cooling estimates towards smaller deviations. Such an interpretation is supported by recent analyses of ice cores from high accumulation sites, questioning the dating of major volcanic events, including the 1452/53 Kuwae eruption (Plummer et al. 2012; Sigl et al. 2013) and challenging the common practice of using particular sulfuric acid layers as markers (e.g. 1258/59; Langway et al. 1988) to align stratigraphy between drill sites (Baillie 2008, 2010). The MXD network analyzed here indicates there were severe cooling events in 1453 C.E. in Northern Europe (coldest year of the past 900 years; labeled in Fig. 3-2) and 1258 C.E. in Central Europe (fourth coldest year; see also Fig. S3-6). The distinct cooling pattern identified at lag +2 in Northern Europe aggregated over 34 annually dated VEI ≥ 5 eruptions, implies the eruption associated with 1453 C.E. cooling even occurred as early as 1451 C.E. Admittedly, this inference is constrained by the limited geographical region (Northern Europe) represented by the MXD network in this study (Fig. 3-3) and the particular response to any single eruption as opposed to the overall mean signal.

Post-volcanic temperature patterns

The northern European cooling pattern reported here appears particularly robust, as the regional MXD data share a high degree of common variance and contain a strong climate signal (64% of MXD-north variance explained by JJA temperatures). The similarity between the SEA results derived from the northern MXD data over the past 900 years, and the northern European instrumental data over the past 260 years, aids the detection of a volcanic signal two years after an eruption. The signal is likely

associated with a positive (negative) sea level pressure and 500 hPa geopotential height anomaly over the central North Atlantic (eastern Scandinavia), connected to anomalous northwesterly and northerly flows towards central Europe (Fischer et al. 2007), suggesting a dynamical response to sub-continental cooling exists.

The high latitude post-volcanic cooling found here is much stronger than the signal reported by Briffa et al. (1998) for the NH extratropics (-0.11°C at lag +1), based on an analysis of a large-scale MXD network in response to 31 selected eruptions over the past 600 years (Fig. S3-4d). The Briffa et al. (1998) experiment, which also included MXD data from low latitude sites, produces an even weaker response (-0.08°C at lag +1 and -0.07 at lag +2) when strictly considering the annually dated NH and tropical VEI ≥ 5 eruptions since 1400 C.E. (SEA 1 in Table 3-1; n = 27 events). The much smaller temperature deviations in the Briffa et al. (1998) NH extratropical MXD network, compared to our findings from Europe, suggest spatially heterogeneous temperature patterns mitigate post-volcanic effects at the hemispheric scale.

Significance of cooling estimates

In Central Europe the lower coherence among MXD sites, as well as the weaker climate signal of the central portion of the network (27% of MXD-central variance explained by JJA temperatures), may bias the post-volcanic estimates towards reduced deviations, thereby affecting the statistical evaluation of significant cooling events with respect to the full spectrum of reconstructed summer temperature variability (Fig. 3-6). On the other hand, the weak post-volcanic signal seen in the central MXD data over the past 900 years is also found in the long instrumental station records over the past 260 years suggesting a lower summer temperature sensitivity to stratospheric volcanic clouds in Central Europe. The similar temperature patterns found in the MXD based and instrumentally based SEAs indicates that those factors which could potentially bias the MXD network response, including enhanced tree growth due to increased diffuse light in post-volcanic years (Farquhar and Roderick 2003), are negligible (see also Krakauer and Randerson 2003). Our findings indicate that the prominent cooling following Tambora in 1816 C.E. (the "year without a summer"; Stothers 1984), as well as in 1912 C.E. (perhaps Novarupta), resulted from stratospheric volcanic clouds that caused atypical summer cooling over Central Europe.

It is important to note that the post-volcanic cooling estimates presented here are spatially restricted to Europe and cannot be transferred to global or even hemispheric dimensions. At this limited continental scale the density, length and quality of both the MXD network and long instrumental station data is unique, enabling assessments of cooling effects based on an exceptionally large number of stratospheric events (34 over the 1111-1976 C.E., and 15 over the 1722-1976 C.E. period). The key finding derived from this condition suggests the relaxation time of eruption induced climate anomalies to be on the order of at most a few years. This finding questions how large volcanic eruptions might initiate decadal, or even centennial scale, temperature changes through feedback mechanisms in the climate system

(Crowley 2000; Robock 2000; Grove 2001; Schneider et al. 2009). While the temporally limited climate response, together with the reduced sensitivity found in response to ice-core derived forcing timeseries, belies the ability of large volcanic eruptions to initiate long-term temperature changes through feedback mechanisms in the climate system, there may be longer relaxation times in other systems – e.g. sea ice (Miller et al. 2012) and ocean temperatures (Church et al. 2005, Gleckler et al. 2006).

This conclusion is supported by the significance of observed, post-volcanic cooling with respect to the full spectrum of summer temperature variability found over the past 900 and 260 years. Figure 3-6 shows only the lag +2 cooling events in Northern Europe deviate at the 95% level from the mean summer temperature of all years over these periods. In Central Europe, the maximum likelihood of post-volcanic temperature cooling reaches approximately 80-85% (JJA-central over the 1722-1976 C.E. period at lag +1). Further research on *(i)* the dating uncertainty of eruptions, particularly during the MWP-LIA transition period (Esper et al. 2002) during which a global reorganization of climate has been suggested (Graham et al. 2007), as well as *(ii)* the development of millennial scale MXD records, that are less biased by biological memory effects than TRW records (Frank et al. 2007, Esper et al. 2007a), is needed to assess the ability of stratospheric volcanic clouds to trigger long-term temperature changes.

Eruption selection schemes

Besides the length of skillful temperature reconstructions, the identification and selection of eruption years appears relevant when assessing post-volcanic cooling effects. Consideration of invariable selection criteria (e.g. tephra volume > 1.0 10^9 m^3) seems advisable, particularly if the period covered by the temperature reconstructions, and thereby the number of volcanic events, is limited. SEA results based on just a dozen eruptions will be sensitive to the inclusion or exclusion of single events, e.g. inclusion of a certain VEI 5 (or even VEI 4) event but exclusion of another VEI 5 event, for example. Similarly, inclusion of selected dendro-dated or ice-core derived events – or temporal shifting of the ice core data to match the temperature proxies – is not recommended as such procedures would likely advance inflated post-volcanic cooling estimates. The approach used here, considering only the annually dated events exceeding a pre-defined VEI threshold, is again constrained by differing sulfur emission magnitudes and eruption plume altitudes. These climatically important measures vary considerably among the VEI = 5 eruptions, for example.

The results shown here using state-of-the-art CGCMs suggest consideration of simulated post-volcanic cooling estimates, as a guideline for empirically based estimates, is not advisable at the sub-continental scale. The simulated summer temperatures over Central and Northern Europe do not cohere among the models, a finding that is largely controlled by the differing volcanic histories used to force the models. As expected, the simulated post-volcanic cooling effects appear much larger if the CGCM runs are aligned by the exact same volcanic events used to force the models. In addition, differences in the models innate climate dynamics, as well as the limited geographical region (grid points in Central and Northern Europe) likely contribute to the inconsistency among the simulations.

3.5 Acknowledgements

Supported by the Mainz Geocycles Research Centre. J.L. acknowledges support from the EU/FP7 project ACQWA (NO212250), the DFG Projects PRIME 2 (`PRecipitation In past Millennia in Europe-extension back to Roman times') within the Priority Program `INTERDYNAMIK' and `Historical climatology of the Middle East based on Arabic sources back to ad 800'.

3.6 References

Ammann CM, Joos F, Schimel DS, Otto-Bliesner BL, Tomas RA (2007) Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model. Proc Nat Acad Sci 104:3713–3718

Anchukaitis KJ, Buckley BM, Cook ER, Cook BI, D'Arrigo RD, Ammann CM (2010) The influence of volcanic eruptions on the climate of the Asian monsoon region. Geophys Res Lett 37. doi:10.1029/2010GL044843

Anchukaitis KJ, et al. (2012) Tree rings and volcanic cooling. Nature Geosc 5:836–837

Angell JK, Korshover J (1985) Surface temperature changes following the six major volcanic episodes between 1780 and 1980. J Clim Appl Meteorol 24:937–951

Baillie MGL (2010) Volcanoes, ice-cores and tree-rings: one story or two? Antiquity 84:202–215

Baillie MGL (2008) Proposed re-dating of the European ice core chronology by seven years prior to the 7th century AD. Geophys Res Lett 35. doi:10.1029/2008GL034755

Barnston AG, Livezey RE (1987) Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. Mon Wea Rev 115:1083–1126

Briffa KR, Jones PD, Schweingruber FH, Osborn TJ (1998) Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. Nature 393:450–455

Büntgen U, Frank DC, Nievergelt D, Esper J (2006) Summer temperature variations in the European Alps, AD 755–2004. J Clim 19:5606–5623

Büntgen U, Frank DC, Grudd H, Esper J (2008) Long-term summer temperature variations in the Pyrenees. Clim Dyn 31:615–631

Büntgen U, Frank D, Trouet V, Esper J (2010) Diverse climate sensitivity of Mediterranean tree-ring width and density. Trees 24:261–273

Büntgen U, et al. (2011) European climate variability and human susceptibility over the past 2500 years. Science 331:578–582

Church J, White N, Arblaster JM (2005) Significant decadal-scale impact of volcanic eruptions on sea level and ocean heart content. Nature 438:74–77

Cole-Dai J (2010) Volcanoes and climate. WIREs Clim Change 1:824–839

Cook ER, Kairiukstis, LA (eds) (1990) Methods of dendrochronology: applications in environmental science. Kluwer, Dordrecht

Crowley TJ (2000) Causes of climate change over the past 1000 years. Science 289:270–277

Crowley T, Zielinski G, Vinther B, Udisti R, Kreutz K, Cole-Dai J, Castellano E (2008) Volcanism and the Little Ice Age. PAGES Newsletter 16:22-23

Crowley TJ, Unterman MB (2012) Technical details concerning development of a 1200-yr proxy index for global volcanism. Earth Syst Sci Data Discuss 5:1–28

D'Arrigo R, Wilson R, Tudhope A (2009) Impact of volcanic forcing on tropical temperatures during the last four centuries. Nature GeoSci 2:51–56

Douglass AE (1920) Evidence of climate effects in the annual rings of trees. Ecology 1:24–32

Driscoll S, Bozzo A, Gray LJ, Robock A, Stenchikov G (2012) Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions. J Geophys Res 117. doi:10.1029/2012JD017607

Esper J, Cook ER, Schweingruber FH (2002) Low-frequency signals in long tree-ring chronologies and the reconstruction of past temperature variability. Science 295:2250–2253

Esper J, Frank DC, Wilson RJS, Briffa KR (2005) Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. Geophys Res Lett 32. doi:10.1029/2004GL021236

Esper J, Büntgen U, Frank DC, Nievergelt D, Liebhold A (2007a) 1200 years of regular outbreaks in alpine insects. Proceed Royal Soc B 274:671–679

Esper J, Büntgen U, Frank D, Pichler T, Nicolussi K (2007b) Updating the Tyrol tree- ring dataset. In: Haneca K, et al. (eds), Tree rings in archaeology, climatology and ecology. TRACE 5:80–85

Esper J, Frank DC, Büntgen U, Verstege A, Hantemirov RM, Kirdyanov AV (2010) Trends and uncertainties in Siberian indicators of 20th century warming. Glob Change Biol 16:386–398

Esper J, Büntgen U, Timonen M, Frank DC (2012a) Variability and extremes of Northern Scandinavian summer temperatures over the past millennia. Glob Plan Change 88-89:1–9

Esper J, Frank DC, Timonen M, Zorita E, Wilson RJS, Luterbacher J, Holzkämper S, Fischer N, Wagner S, Nievergelt D, Verstege A, Büntgen U (2012b) Orbital forcing of tree-ring data. Nature Clim Change 2:862–866

Esper J, Büntgen U, Luterbacher J, Krusic P (2013) Testing the hypothesis of post-volcanic missing rings in temperature sensitive dendrochronological data. Dendrochronologia, http://dx.doi.org/10.1016/j.dendro.2012.11.002

Farquhar, GD, Roderick ML (2003) Pinatubo, diffuse light, and the carbon cycle, Science 299:1997–1998

Fernández-Donado L, et al. (2013) Large-scale temperature response to external forcing in simulations and 50 reconstructions of the last millennium. Clim Past 9:393–421

Fischer EM, Luterbacher J, Zorita E, Tett FB, Casty C, Wanner H (2007) European climate response to tropical volcanic eruptions over the last half millennium. Geophys Res Lett 34, L05707. doi:10.1029/2006GL027992

Frank D, Büntgen U, Böhm R, Maugeri M, Esper J (2007) Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. Quat Sc Rev 26:3298-3310

Frank D, Esper J, Zorita E, Wilson RJS (2010) A noodle, hockey stick, and spaghetti plate: a perspective on high-resolution paleoclimatology. WIREs Clim Change 1:507–516

Gao C, et al. (2006) The 1452 or 1453 A.D. Kuwae eruption signal derived from multiple ice core records: Greatest volcanic sulfate event of the past 700 years. J Geophys Res 111. doi:10.1029/2005JD006710

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. J Geophys Res 113 D23111. doi:10.1029/2008JD010239

Gent PR., et al. (2011) The Community Climate System Model Version 4. J Climate 24:4973–4991

Gleckler PJ, AchutaRao K, Gregory JM, Santer BD, Taylor KE, Wigley TML (2006) Krakatoa lives: the effect of volcanic eruptions on ocean heat content and thermal expansion. Geophys Res Lett 33. doi:10.1029/2006GL026771

Graham NE et al. (2007) tropical Pacific – mid-latitude teleconnections in medieval times. Clim Change 83:241–285

Graham NE, Ammann CM, Fleitmann D, Cobb KM, Luterbacher J (2011) Support for global climate reorganization during the "Medieval Climate Anomaly. Clim Dyn 37:1217–1245

Grove JM (2001) The initiation of the ''Little Ice Age'' in regions round the North Atlantic. Clim Change 48:53–82

Grudd H (2008) Tornetraesk tree-ring width and density AD 500–2004: a test of climatic sensitivity and a new 1500-year reconstruction of north Fennoscandian summers. Clim Dyn 31:843–857

Gunnarson BE, Linderholm HW, Moberg A (2010) Improving a tree-ring reconstruction from west-central Scandinavia – 900 years of warm-season temperatures. Clim Dyn. doi:10.1007/s00382-010-0783-5

Hammer CU, Clausen HB, Dansgaard W (1980) Greenland ice sheet evidence of post-glacial volcanism and its climatic impact. Nature 288:230–235

Hammer CU, Clausen HB, Tauber H (1986) Ice-cored dating of the Pleistocene/Holocene boundary applied to a calibration of the ^{14}C timescale. Radiocarbon 28: 284-291

Hegerl GC, Crowley TS, Baum SK, Kim K-Y, Hyde WT (2003) Detection of volcanic, solar and greenhouse gas signals in paleo-reconstructions of Northern Hemispheric temperature. Geophys Res Lett 30. doi:10.1029/2002GL016635

Hegerl G, Luterbacher J, González-Rouco F, Tett SFB, Crowley TJ, Xoplaki E (2011) Influence of human and natural forcing on European seasonal temperatures. Nat Geosci 4:99–103

Jones PD, Moberg A, Osborn TJ, Briffa KR (2003) Surface climate responses to explosive volcanic eruptions seen in long European temperature records and mid-to-high latitude tree-ring density around the Northern Hemisphere. Geophys Monogr Ser 139:239–254

Jones PD, Lister DH, Osborn TJ, Harpham C, Salmon M, Morice CP (2012) Hemispheric and large-scale land surface air temperature variations: an extensive revision and an update to 2010. J Geophys Res 117 D05127. doi:10.1029/2011JD017139

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

Kelly PM, Sear CB (1984) Climatic impact of explosive volcanic eruptions. Nature 311:740–743

Krakauer NY, Randerson JT (2003) Do volcanic eruptions enhance or diminish net primary production? Evidence from tree rings. Glob Biogeochem Cycl 17. doi:10.1029/2003GB002076

Kurbatov AV, Zielinski GA, Dunbar NW, Mayewski PA, Meyerson EA, Sneed SB, Taylor KC (2006) A 12,000 year record of explosive volcanism in the Siple Dome Ice Core, West Antarctica. J Geophys Res 111 D12307. doi:10.1029/2005JD006072

LaMarche VC, Hirschboeck KK (1984) Frost rings in trees as records of major volcanic eruptions. Nature 307:121–126

Langway CC, Clausen HB, Hammer CU (1988) An inter-hemispheric volcanic time-marker in cores from Greenland and Antarctica. Ann Glaciol 10:102–108

Mann ME, Fuentes JD, Rutherford S (2012) Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. Nature Geosc 5:202–205

Mass CF, Portman DA (1989) Major volcanic eruptions and climate: a critical evaluation. J Clim 2:566–593

McCormick PM, Wang PH, Poole LR (1993) Stratospheric aerosols and clouds. In: Hobbs PV (ed) Aerosol-cloud-climate interactions. Academic Press, San Diego, pp 205–222

Miller, GH et al. (2012) Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks. Geophys Res Lett 39 L02708. doi:10.1029/ 2011GL050168

Moser L, Fonti P, Büntgen U, Esper J, Luterbacher J, Franzen J, Frank D (2010) Timing and duration of European larch growing season along an altitudinal gradient in the Swiss Alps. Tree Physiol 30:225–233

Mitchell TM, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol 25:693–712

Newhall CG, Self S (1982) The volcanic explosivity index (VEI): an estimate of explosive magnitude for historical volcanism. J Geophys Res 87:1231–1238

Oppenheimer C (2003) Ice core and palaeoclimatic evidence for the timing and nature of the great mid-13th century volcanic eruptions. Int J Climatol 23:417–426

Panofsky HA, Brier GW (1958) Some applications of statistics to meteorology. Univ. Park, Pennsylvania.

Plummer CT, et al. (2012) An independently dated 2000-yr volcanic record from Law Dome, East Antarctica, including a new perspective on the dating of the c. 1450s eruption of Kuwae, Vanuatu. Clim Past Discuss 8:1567–1590

Robock A, Free MP (1995) Ice cores as an index of global volcanism from 1850 to the present. J Geophys Res 100:11549-11567

Robock A, Mao J (1995) The volcanic signal in surface temperature observations. J Clim 8:1086–1103

Robock A (2000) Volcanic eruptions and climate. Rev Geophys 38:191–219

Salzer MW, Hughes MK (2007) Bristlecone pine tree rings and volcanic eruptions over the last 5000 yr. Quat Res 2007, 67:57–68

Schneider DP, Ammann CM, Otto-Bliesner BL, Kaufman DS (2009) Climate response to large, high-latitude and low-latitude volcanic eruptions in the Community Climate System Model. J Geophys Res 114 D15101. doi:10.1029/2008JD011222

Schweingruber FH, Fritts HC, Bräker OU, Drew, LG, Schaer E (1978) The X-ray technique as applied to dendroclimatology. Tree-Ring Bull 38:61–91

Schweingruber FH, Bartholin T, Schär E, Briffa KR (1988) Radiodensitometric– dendroclimatological conifer chronologies from Lapland (Scandinavia) and the Alps (Switzerland). Boreas 17:559–566

Sear CB, Kelly PM, Jones PD, Goodess CM (1987), Global surface temperatures responses to major volcanic eruptions, Nature 330:365–367

Self S, Rampino MR, Barbera JJ (1981) The possible effects of large 19th and 20th century volcanic eruptions on zonal and hemispheric surface temperatures. J Vol Geothermal Res 11:41–60

Sigl M, et al. (2012) A new bipolar ice core record of volcanism from WAIS Divide and NEEM and implications for climate forcing of the last 2000 years J Geophys Res doi:10.1029/2012JD018603, in press

Plummer CT, et al. (2012) An independently dated 2000-yr volcanic record from Law Dome, East Antarctica, including a new perspective on the dating of the c. 1450s eruption of Kuwae, Vanuatu. Clim Past Discuss 8:1567–1590

Siebert L, Simkin T, Kimberly P (2010) Volcanoes of the world. Univ. California Press, London

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) (2007) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge Univ Press, Cambridge

Stenchikov G, Hamilton K, Stouffer RJ, Robock A, Ramaswamy V, Santer B, Graf HF (2006) Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models. J Geophys Res 111. doi:10.1029/2005JD006286

Stothers RB (1984) The great Tambora eruption in 1815 and its aftermath. Science 224:1191–1198

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experimental design. Bull Amer Meteor Soc 93:485-497

Timmreck C, Lorenz SJ, Crowley TJ, Kinne S, Raddatz TJ, Thomas MA, Jungclaus JH (2009) Limited temperature response to the very large AD 1258 volcanic eruption. Geophys Res Lett 36. doi:10.1029/2009GL040083

Traufetter F, Oerter H, Fischer H, Weller R, Miller H (2004) Spatio-temporal variability in volcanic sulphate deposition over the past 2 kyrs in snow pits and firn cores from Amundsenisen, Antarctica. Claciol 50:137–146

Trepte CR, Hitchman MH (1992) tropical stratospheric circulation deduced from satellite aerosol data. Nature 355:626–628

Trouet V, Esper J, Graham NE, Baker A, Scourse JD, Frank DC (2009) Persistent positive North Atlantic Oscillation mode dominated the Medieval Climate Anomaly. Science 324:78–80

Wagner S, Zorita E (2005) The influence of volcanic, solar and CO2 forcing on the temperatures in the Dalton Minimum (1790–1830): a model study. Clim Dyn 25:205-218

Zanchettin D, Timmreck C, Bothe O, Lorenz S, Hegerl G, Graf H-F, Luterbacher J, Jungclaus, JH (2013a) Delayed winter warming: a decadal dynamical response to strong tropical volcanic eruptions. Geophys Res Lett 40:204–209

Zanchettin D, Bothe O, Graf H-F, Luterbacher J, Jungclaus JH, Timmreck C (2013b) Background conditions influence decadal climate response to strong volcanic eruptions. J Geophys Res. doi:10.1002/jgrd.50229

Zielinski GA (1995) Stratospheric loading and optical depth estimates of explosive volcanism over the last 2100 years derived from the Greenland Ice Sheet Project 2 ice core. J Geophys Res 100:20937–20955

Zorita E, Gonzlez-Rouco F, von Storch H, Montavez JP, Valero F (2005) Natural and anthropogenic modes of surface temperature variations in the last thousand years. Geophys Res Lett 32 L08707. doi:10.1029/2004GL021563

4. Timing and duration of post-volcanic Northern Hemisphere cooling revealed from tree-ring records of maximum latewood density

L. Schneider[1] , J. Smerdon[2], U. Büntgen[3], V. S. Myglan[4] & J. Esper[1]

[1] *Department of Geography, University of Mainz, Germany*

[2] *Lamont-Doherty Earth Observatory, Palisades, New York, USA*

[3] *Swiss Federal Research Institute WSL, Birmensdorf, Switzerland*

[4] *Siberian Federal University, Krasnoyarsk, Russia*

Published in Di Filippo, A., Piovesan, G., Romagnoli, M., Helle, G., Gärtner, H. (Eds.), 2014

4.1 Introduction

Assessments of large-scale temperature responses to volcanic eruptions are important for improving our understanding of a key natural climate forcing, and more generally for understanding how changes in radiative forcings impact climate feedbacks and modes of variability. Stratospheric aerosols, the main cause of climatic perturbations associated with volcanoes, can vary in origin, size and composition. Their climatic effects may be altered by seasonality and source. Each of these factors determines the degree to which volcanic eruptions have widespread (i.e. hemispheric) impacts on climate (Robock 2000).

Reliable information about the characteristics of volcanic eruptions becomes more limited with increasing time before present. The use of proxy-records allows estimates of volcanic activity and climatic impact before the onset of meteorological observations. The sulphur concentration in ice cores is a common measure to deduce volcanic forcing by translating it into atmospheric sulphate loadings (Gao et al. 2008) or radiation equivalents (Crowley 2000).

The timing and scale of a putative climatic fingerprint is, however, difficult to estimate without using a more direct measure for temperature. Past studies (Briffa et al. 1998a, D'Arrigo et al. 2009) showed that maximum latewood-density (MXD) measurements from tree-rings indeed offer good approximations of the short-term summer cooling associated with volcanic events if the volcanic signal is extracted by averaging over several event years.

This method is nevertheless restricted to the relatively small number of well-dated volcanoes because dating errors reduce magnitude and introduce lag effects in the estimated climatic response (Esper et al. 2013). In fact, monthly dating is required if taking into account the example of a volcanic eruption around the turn of the year. Such a volcano could be assigned to two different calendar years while atmospheric dispersion and timing of the climatic response might be very similar. Ice core dates may also be affected by misdating, particularly those associated with deep-core estimates (Baillie 2008), which is an additional drawback if the most significant events are to be derived from sulphur-records.

To evaluate possible biases that may arise from such dating uncertainties, we here use a compilation of six sulphur-rich monthly dated eruptions that took place during the past millennium for comparison with MXD-based northern hemispheric (NH) summer temperature variability. Events are first analysed on a continental to hemispheric level to evaluate the lag between eruption and maximum cooling. These estimates are then used to adjust the process of averaging over multiple events. The new shape of a typical volcanic cooling pattern is tested for significance using a bootstrap approach.

4.2 Data and methods

Northern-hemispheric temperature-reconstructions

A total of 15 MXD site chronologies spanning at least 600 years were used to reconstruct summer-temperature over larger continental regions (Tab. 4-1). We processed each site separately and applied a power transformation (Cook & Peters 1997) and a Hugershoff detrending (Briffa et al. 1998b) to account for spread-vs.-level relationships and biological age trends. After calculating residuals from the estimated growth curves we built a robust mean for all years with a replication ≥ three series. To remove variance-changes arising from varying replication and signal strength an empirical variance-stabilization (Frank et al. 2007) was applied: we smoothed the absolute deviations from the total mean with a 200-year cubic spline and used the spline function to adjust variance. The chronologies were calibrated against gridded temperature data (Jones et al. 2012) for June-August (JJA) using the closest grid-point (Tab. 4-1). The common period for the calibration was set to 1926-1987, although two MXD-chronologies drop out in the late 1970s. To retain the full temperature variance a scaling was performed in order to translate tree-ring indices to temperature means (Esper et al. 2005). Since all proxy-records explained a significant amount of local temperature variation (Tab. 4-1), all site-records were averaged in order to estimate large-scale anomalies. Continental averages were calculated for North America, Europe and Asia using 4, 7 and 4 site-chronologies, respectively. Additionally a simple northern hemispheric average was built as the arithmetic mean of the 3 continental records. Here, we exclusively focus on close-ups for six volcanic events between 1600 and 1883.

Table 4-1 | Tree-ring sites and measures of replication for the 1580-1903 period. The corresponding grid-points refer to the CRUTEM4-dataset, which was used for calibration. The Pearson-correlation between tree-ring data and summer temperature is given as r_{MXD_JJA}.

Site name	*Continent*	*Location*	*Replication*		*Corresponding grid-point*	r_{MXD_JJA}	*Reference*
			Mean	*Minimum*			
Alaska	N-America	68.8N/-142.4E	75	47	67.5N/-132.5E	0.53	Anchukaitis et al. 2013
Altai	Asia	50.2N/90.0E	17	15	52.5N/87.5E	0.62	Myglan et al. 2012
Athabasca	N-America	52.3N/117.3E	28	8	52.5N/-117.5E	0.47	Luckman et al. 1997
Campbell	N-America	68.3N/-133.3E	19	14	67.5N/-132.5E	0.32	Schweingruber, ITRDB
Jaemtland	Europe	63.5N/15.5E	31	14	62.5n/17.5E	0.64	Schweingruber et al. 1988
Lauenen	Europe	46.4N/7.3E	23	11	47.5N/7.5E	0.39	Schweingruber et al. 1988
Lötschental	Europe	47.5N/7.5E	58	44	47.5N/7.5E	0.60	Büntgen et al. 2006
Mangazeja	Asia	66.7N/82.3E	36	3	67.5N/82.5E	0.60	Briffa et al. 2001
Nscan	Europe	67.5N/22.5E	50	29	67.5N/22.5E	0.80	Esper et al. 2012
Polarural	Asia	66.9N/65.6E	30	10	67.5N/67.5E	0.81	Briffa et al. 1995
Pyrenees	Europe	42.5N/2.5E	70	62	42.5N/2.5E	0.40	Büntgen et al. 2008
Quebec	N-America	57.5N/-76.0E	15	7	57.5N/-77.5E	0.69	Schweingruber 2007
Torneträsk	Europe	68.2N/19.7E	26	11	67.5N/17.5E	0.83	Melvin et al. 2013
Tyrol	Europe	47.5N/12.5E	35	16	47.5N/12.5E	0.38	Esper et al. 2007
Zhaschiviersk	Asia	67.5N/142.6E	21	14	66.3N/143.8E	0.27	Briffa et al. 2001

Superposed epoch analysis

A common tool to assess the quantity of volcanic cooling is the superposed epoch analysis (SEA; Panofsky & Brier 1958) that averages temperature patterns after a certain number of events. Including at least 10 years prior and 20 years after the volcanic events allows comparing pre- and post-volcanic time-series characteristics and ensures full coverage of the retention period. Herein all SEAs are displayed as anomalies to the ten pre-volcanic years (lag-10 to lag-1). The year of eruption is referred to as the reference point and located at year zero in the SEA. Since this study focuses on timing and duration of the climatic response, we filtered out the (1) climatically most relevant and (2) well dated volcanic events during the last 800 years - reflecting the original length of the temperature reconstruction.

(1) In a first step an ice-core sulphur record (Gao et al. 2008) was used to identify the most relevant events. The sulphur-record features a global, northern and southern hemispheric sulphate-loading time-series. The 20 highest peaks were chosen not only from the northern hemispheric record but also from the global one (Tab. 4-2). Proceeding only with those events that were recorded in both records ensures that high-latitudinal events close to the arctic ice sheet depositions are not misinterpreted. (2) The remaining peaks were compared with a list of monthly dated volcanoes (Siebert et al. 2010). If no major eruption (volcanic explosivity index ≥ 4) was documented in the same year or one year prior, the event was rejected. The others were assigned to a volcano name and region (Tab. 4-2).

Table 4-2 | Selection of volcanic events over the last 800 years. Only six out of sixteen globally relevant eruptions could be identified as a monthly dated event.

Gao2008		*Month of*	*Volcano name*	*Volcano*
Glo	*NH*	*eruption begin*		*region*
1227	1227	-		
1258	1258	-		
1275	-			
1284	1284	-		
1328	1328	-		
1341	-			
1452	1452	-		
1459	1459	-		
-	1476			
1584	1584	-		
1600	1600	Feb	HUAYNAPUTINA	Perú
1641	1641	Jan	PARKER	Philippines
1693	-			
1719	1719	-		
-	1729			
1783	1783	Jun	GRIMSVÖTN	Iceland
1809	1809	-		
1815	1815	Apr	TAMBORA	Indonesia
1831	1831	-		
1835	1835	Jan	COSIGUINA	Nicaragua
1883	1883	Aug	KRAKATAU	Indonesia
-	1912			
-	1925			
1963	-			

To assess the significance of post-volcanic temperature patterns a bootstrapping technique was employed (Zanchettin et al. 2013). The actual set of six event-years was replaced 1000 times by a set of six randomly chosen years between 1580 and 1903. By calculating SEAs with these random years instead of those containing volcanic events the probability distribution of temperature anomalies can be

quantified for each time step. Deviations from the 95 and 99 percentiles of this distributions indicate a significant temperature anomaly.

4.3 Results

The two sets of 20 maximum sulphate loadings – global and northern hemisphere – show an overlap of 80% (Tab. 4-2). Among all 16 common years only six could be assigned to a monthly dated volcano from the record in Siebert et al. (2010). Although the 20 biggest spikes in the sulphur records are evenly distributed over the 800 years of record, all of these six dated events took place in the second half of the millennium. This indicates either incompleteness in the documentary volcano record or dating uncertainties in the ice core records. However, it limits our analyses to the recent half of the record. Despite the 1783 Grimsvötn eruption, the remaining events are of tropical origin, which is fundamental for a widespread and even distribution of volcanic ash over both hemispheres (Budner & Cole-Dai 2003).

The spatial distribution of long-density records is somewhat weighted toward Europe regarding the total number of sites and their individual replication (Tab. 4-1). On average, European sites also correlate strongest with summer-temperature (r_{ave} = 0.58) for the 62 years of calibration period. However, sites in North America and Asia also contribute strongly to the amount of explained temperature variance (r_{ave} = 0.51 and 0.57, respectively). While local records are strongly driven by internal climate variability and regional characteristics, these fractions tend to level out if the scale is increased. In hemispheric or continental averages volcanic forcing is unmasked and often leaves a distinct cooling peak (Jones et al. 2003). This is also observed for the climatic effect of the six volcanoes studied herein. While the most negative NH temperature anomalies in the six epochs analysed can be attributed to the volcanic events in five of six cases (either at lag0 or at lag1) the continental records show a number of similarly cold years before or after the actual cooling period (Fig. 4-1).

Although volcanic cooling is an outstanding feature in these epochs, it varies strongly in timing, magnitude and shape. Temperature decreases mostly abruptly, but is either followed by immediate retention (e.g. after Huaynaputina and Grimsvötn) or it takes two to three years to recover to the original level (e.g. after Mt. Parker, Tambora or Krakatau). The lowest values for hemispheric temperature occur in the year of the sulphur peak or one year later. However, there is no consistent pattern that links this lag to the month of eruption. Eruptions during winter, for instance, can result in a different timing for minimum summer-temperature (Mt. Parker vs. Huaynaputina and Cosiguina). The Grimsvötn eruption in contrast shows maximal summer cooling in 1783 although the ejection of ash does not start before June of the same year. If a mean magnitude is to be derived from these events, such lag-effects result in blurring of the cooling pattern.

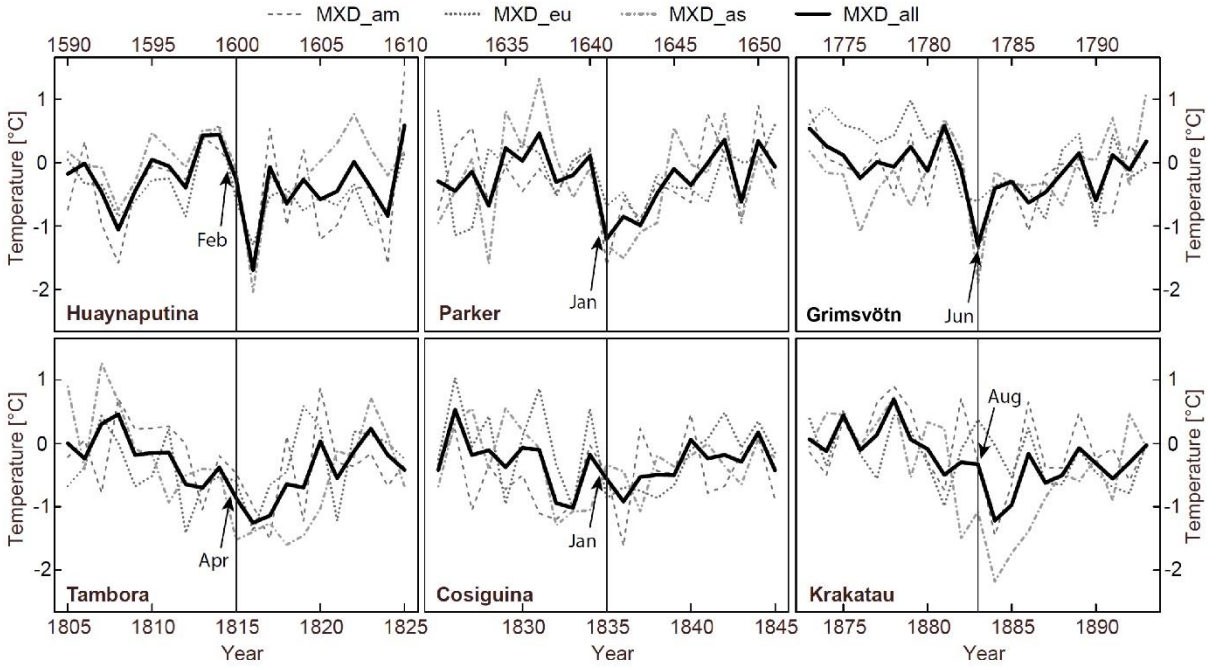


Figure 4-1 | Epochs of strong volcanic events as revealed by continental and hemispheric summer-temperature reconstructions. Temperatures represent anomalies with respect to the 10 years prior to the event. Vertical lines in the middle of each epoch indicate the year of the sulphur-peak in the ice-core records. Arrows give the eruption month as derived from documentary evidence.

To estimate the effect of seasonal transportation patterns and resulting lag-structures, SEAs were calculated in two different ways: (1) For SEA1 the documented years of eruption (which match the years of sulphur peaks) were used as the reference point in the centre of the SEA. (2) In a second approach (SEA2) the reference points were synchronized with the years of minimum post-volcanic summer-temperature in order to estimate the maximum cooling effect. According to Figure 4-1 these are the years 1601, 1641, 1783, 1816, 1836 and 1884. The resulting SEAs for continental and hemispheric averages show distinct differences in magnitude and timing of post-volcanic cooling (Fig. 4-2). After adjusting the reference points the temperature drop is now observed at lag 0 with a rather sharp drop. Only in the European SEAs a clear pattern is missing and cooling is only weakly significant against random fluctuation. The bootstrap-bands are a robust estimate of variance for the SEAs and all values

before and after the volcanic spikes stay within the limits of the 99 percentile. The positive effect of averaging regarding the signal-to-noise ratio becomes apparent in a significantly reduced bandwidth for the hemispheric SEAs. The reference period for the calculation of anomalies is also accompanied by a slightly smaller bandwidth, especially for the Asian SEAs.

With respect to the standard error of the mean the most significant cooling is found for the northern hemispheric SEA2 with a temperature difference of 0.93°C between lag -1 and lag 0. This value is smallest for Europe with only 0.44°C.

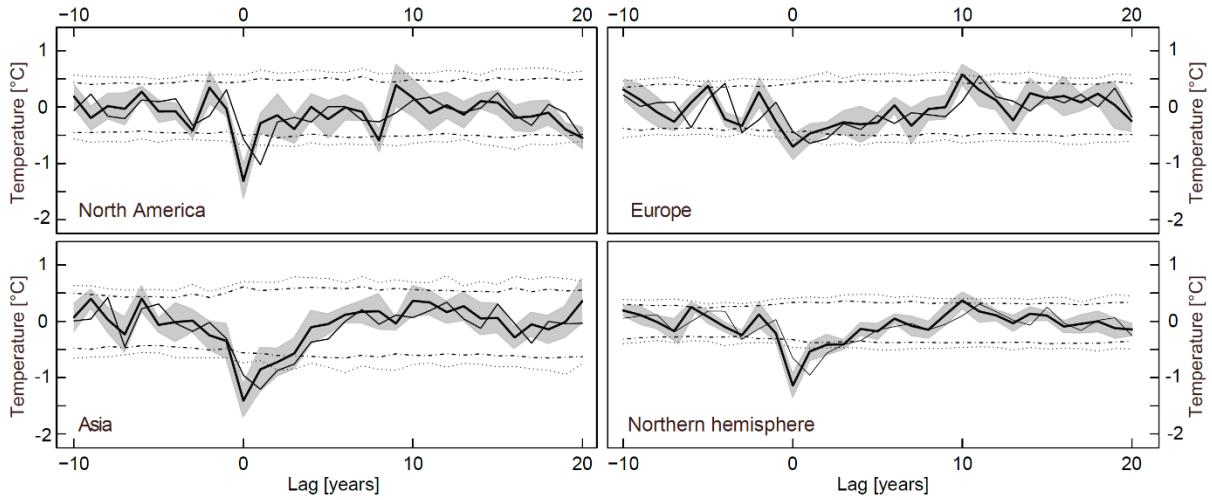


Figure 4-2 | SEAs for different continental averages and the Northern hemisphere. The bold lines represent SEA2 with epochs centred around 1601, 1641, 1783, 1816, 1836 and 1884. SEA1 (thin lines) used the documented/sulphur record years 1600, 1641, 1783, 1815, 1835 and 1883. Temperatures represent anomalies with respect to the 10 years prior to the event. Dashed lines indicate the 95 and 99 percentiles of a bootstrap approach to assess significance. The grey bands show the standard error of the mean (standard deviation divided by the square root of the sample size) for each time-step.

4.4 Discussion

This study documents the potential of MXD-data in storing the climatic response to volcanic eruptions, which is an important information regarding the calibration of sulphur-records and climate models. Although the number of events studied herein had to be limited due to methodological reasons we found distinct patterns of post-volcanic cooling with the help of averaged SEAs. At the same time the basic need for a detailed analysis of single events was revealed. Despite the influence of other climate forcing parameters and internal variability the volcanic impact resulted in a cooling in all 6 epochs and allowed for an improved alignment according to the cooling pattern. This alignment had to follow a visual inspection of the single temperature time-series because no robust relationship between the seasonality of the eruption and the offset between eruption and cooling could be established. Climatic responses are observed in the next growing season or in that of the following year. Only the high-latitudinal eruption of Grimsvötn in summer 1783 translates in immediate cooling whereas magnitude and duration of the cooling associated with this event are in line with eruptions of tropical source.

Compared to other studies (Briffa et al. 1998a, D'Arrigo et al. 2009), the observed cooling-response in SEA2 for North America, Asia and the hemisphere is fairly strong in magnitude. To some extent this is caused by the new approach of aligning by the climatic impact, but it is also a corollary of using only the 6 biggest events. While these SEA2s show temperature anomalies of more than -1°C, the European SEA2 is only weakly significant. This can be attributed to the large portion of alpine records, which respond less to volcanic forcing (Esper et al. 2013).

The two- to three-years temperature depressions in the SEA1s suggest a slow emergence of volcanic cooling with the minimum temperature observed at lag1 after the eruption. In fact, this pattern is an artefact of shifting seasonality, which should be considered in the interpretation. In a larger setup with more volcanoes this approach could, however, help to quantify the number of volcanoes with response at lag0, lag1 or lag2.

According to theoretical sulphur transport (Gao et al. 2008) and observational data of atmospheric optical depth (Crowley et al. 2013) it is more likely to find a sharp temperature decline with a subsequent - possibly fast - recovery. This concept would be better in line with the pattern that was found for the SEA2s and justifies the adjustment of the temperature-minima. Irregular lag-structures, however, limit the potential to upscale this approach by incorporating more monthly dated volcanic events. But using a simple temperature-minimum between lag0 and lag2 as reference point – as in this study – solves the problem although it might introduce a bias in the adjustment procedure if noise or internal variability alter the volcanic signal.

4.5 Conclusion

This study presents a data-adaptive way of analysing the volcanic impact on temperature records across the NH. The method suggests results that are considerably different from previous studies concerning both duration and magnitude of post-volcanic cooling. Adding more volcanoes of smaller size will likely mitigate the increased magnitude. The shape of the climatic response shows an abrupt temperature drop and a short recovery period. Since the alignment of volcanic impacts followed a visual inspection, it would be desirable to objectify this step. The high quality and spatial coverage of the underlying database indicate that the detection of cooling patterns can potentially be automated by applying objective detection algorithms (Hendry & Pretis 2013). This would help to expand this method to more advanced climate time-series and an extended set of volcanoes, which would likely result in more significant results.

4.6 References

Anchukaitis, K. J. et al. (2013): Tree-Ring-Reconstructed Summer Temperatures from Northwestern North America during the Last Nine Centuries. *Journal of Climate* 26(10): 3001-3012.

Baillie, M. G. L. (2008): Proposed re-dating of the European ice core chronology by seven years prior to the 7th century AD. *Geophysical Research Letters* 35(15).

Briffa, K. R., Jones, P. D., Schweingruber, F. H., Shiyatov, S. G., Cook, E. R. (1995): Unusual 20th-Century Summer Warmth in a 1,000-Year Temperature Record from Siberia. *Nature* 376(6536): 156-159.

Briffa, K. R., Jones, P. D., Schweingruber, F. H., Osborn, T. J. (1998a): Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. *Nature* 393(6684): 450-455.

Briffa, K. R., Schweingruber, F. H., Jones, P. D., Osborn, T. J., Shiyatov, S. G., & Vaganov, E. A. (1998b): Reduced sensitivity of recent tree-growth to temperature at high northern latitudes. *Nature* 391(6668): 678-682.

Briffa, K. R., Osborn, T. J., Schweingruber, F. H., Harris, I. C., Jones, P. D., Shiyatov, S. G., Vaganov, E. A. (2001): Low-frequency temperature variations from a northern tree ring density network. *Journal of Geophysical Research-Atmospheres* 106(D3): 2929-2941.

Budner, D. M., Cole-Dai, J. (2003): The number and magnitude of explosive volcanic eruptions between 904 and 1865 A.D.: Quantitative evidence from a new South Pole ice core. In: Robock, A., Oppenheimer, C. (ed.): Volcanism and the Earth's Atmosphere. American Geophysical Union. 165-176.

Büntgen, U., Frank, D. C., Nievergelt, D., Esper, J. (2006): Summer temperature variations in the European Alps, AD 755-2004. *Journal of Climate* 19(21): 5606-5623.

Büntgen, U., Frank, D., Grudd, H., Esper, J. (2008): Long-term summer temperature variations in the Pyrenees. *Climate Dynamics* 31(6): 615-631.

Cook, E. R., Peters, K. (1997): Calculating unbiased tree-ring indices for the study of climatic and environmental change. *Holocene* 7(3): 361-370.

Crowley, T. J. (2000): Causes of climate change over the past 1000 years. *Science* 289(5477): 270-277.

D'Arrigo, R., Wilson, R., Tudhope, A. (2009): The impact of volcanic forcing on tropical temperatures during the past four centuries. *Nature Geoscience* 2(1): 51-56.

Esper, J., Frank, D. C., Wilson, R. J. S., Briffa, K. R. (2005): Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. *Geophysical Research Letters* 32(7).

Esper, J., Büntgen, U., Frank, D. C., Nievergelt, D., Liebhold, A. (2007): 1200 years of regular outbreaks in alpine insects. *Proceedings of the Royal Society B-Biological Sciences,* 274(1610): 671-679.

Esper, J. et al. (2012): Orbital forcing of tree-ring data. *Nature Climate Change* 2(12): 862-866.

Esper, J. et al. (2013): European summer temperature response to annually dated volcanic eruptions over the past nine centuries. *Bulletin of Volcanology* 75(7).

Frank, D., Esper, J., Cook, E. R. (2007): Adjustment for proxy number and coherence in a large-scale temperature reconstruction. *Geophysical Research Letters* 34(16).

Gao, C. 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).

Hendry, D. F., Pretis F. (2013): Anthropogenic Influences on Atmospheric CO2. In: Fouquet, R. (ed.): Handbook on Energy and Climate Change. Edward Elgar. 287-326.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., Morice, C. P. (2012): Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research-Atmospheres* 117.

Jones, P. D., Moberg, A., Osborn, T. J., Briffa, K. R. (2003): Surface climate responses to explosive volcanic eruptions seen in long European temperature records and mid-to-high latitude tree-ring density around the Northern Hemisphere. In: Robock, A., Oppenheimer, C. (ed.): Volcanism and the Earth's Atmosphere. American Geophysical Union. 239-254.

Luckman, B. H., Briffa, K. R., Jones, P. D., Schweingruber, F. H. (1997): Tree-ring based reconstruction of summer temperatures at the Columbia Icefield, Alberta, Canada, AD 1073-1983. *Holocene* 7(4): 375-389.

Melvin, T. M., Grudd, H., Briffa, K. R. (2013): Potential bias in 'updating' tree-ring chronologies using regional curve standardisation: Re-processing 1500 years of Tornetrask density and ring-width data. *Holocene* 23(3): 364-373.

Myglan, V. S., Oidupaa, O. C., Vaganov, E. A. (2012): A 2367-year tree-ring chronology for the Altai-Sayan region. Archeaology, Ethnology & Anthropology of Eurasia 40(3): 76-83.

Panofsky, H. A., Brier, G. W. (1958): Some applications of statistics to meteorology. University Park, Pa. 224.

Robock, A. (2000): Volcanic eruptions and climate. *Reviews of Geophysics* 38(2): 191-219.

Schweingruber, F. H., Bartholin, T., Schar, E., Briffa, K. R. (1988): Radiodensitometric-Dendroclimatological Conifer Chronologies from Lapland (Scandinavia) and the Alps (Switzerland). *Boreas* 17(4): 559-566.

Schweingruber, F. H. (2007): Tree-ring measurements of Picea mariana (Black spruce) from sample QUEBEC-382. Retrieved from doi.pangaea.de/10.1594/PANGAEA.597213 website.

Siebert, L., Simkin, T., Kimberly, P. (2010): Volcanoes of the world. Univ. California Press. 551.

Zanchettin, D. et al. (2013): Delayed winter warming: A robust decadal response to strong tropical volcanic eruptions? *Geophysical Research Letters* 40(1): 204-209.

5. Volcanic induced cooling in instrumental and tree-ring density data

L. Schneider[1], J.E. Smerdon[2], F. Pretis[3] & J. Esper[1]

[1] *Department of Geography, University of Mainz, Germany*

[2] *Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA*

[3] *Programme for Economic Modelling, Oxford Martin School, University of Oxford, Oxford, UK*

Published in Hevia, A., Sánchez-Salguero, R., Linares, J.C., Olano, J.M., Camarero, J.J., Gutiérrez, E., Helle, G., Gärtner, H. (Eds.), 2016

5.1 Introduction

Deciphering the impact of large volcanic eruptions on large-scale climate can yield valuable information on climate sensitivity to radiative perturbations at short timescales (Masson-Delmotte et al. 2013). However, during the era of instrumental climate observations spanning roughly the past 150 years, the number of volcanic events is relatively small, and polar ice cores indicate that the amount of radiation-absorbing sulphate injected into the stratosphere was much larger for several eruptions occurring earlier in the last millennium (Gao et al. 2008; Crowley & Unterman 2013). Analysis of climate archives with high temporal resolution, such as tree-rings, can increase the number of detected events and elucidate the full range of possible volcanic impact. Additionally, the large spatial distribution of tree-ring data provides a wide-angle perspective on climate variability that can dampen local anomalies and amplify externally driven climate variability. Thus, a large-scale tree-ring composite can be an appropriate tool for assessing volcanic feedbacks in the climate system.

A good understanding of the relevant proxy/climate-relationship is a prerequisite for analyzing volcanic-induced cooling using proxy reconstructions. Although temperature sensitivity is well established for tree-rings from high latitudes and altitudes (Fritts 1976), and although it is known that especially tree-ring density data are suitable for studying abrupt temperature changes (Esper et al. 2013, 2015), there are a few pitfalls associated with calculating cooling estimates for volcanic events from tree-ring records:

Usually a linear relationship between temperature and tree-growth is assumed. This can be altered, especially in the case of volcanic events, by the influence of light availability (Robock 2005; Tingley et al. 2014).

The network of proxy sites can be biased towards regions with weaker or stronger influence of volcanic activity or response to volcanic forcing. While this is a general problem of relatively sparse proxy networks, it is of particular importance for the evaluation of a climate forcing using point-source data.

Aggregating or averaging spatial data might reduce the observed amplitude of volcanic cooling.

The hypotheses on light availability referenced in *(i)* are based on large-scale experiments. On local scales, proxy-derived temperatures were found to be in good agreement with long instrumental records (Esper et al. 2013), so that it seems likely that the integration of data over larger regions causes the offset observed in large-scale studies (Tingley et al. 2014). Here we address such potential effects by analyzing a hemispheric composite of maximum latewood density (MXD) chronologies. A summer-temperature reconstruction based on these data reflects distinct cooling in response to the largest eruptions of the last millennium (Schneider et al. 2015). By comparing this dataset with observational data from the 19th and 20th centuries, we intend to test its susceptibility to *(ii)* and *(iii)*, and to verify the cooling estimates derived from this record. We find that spatial aggregation yields systematic underestimation of volcanic induced cooling despite a reasonable hemispheric coverage of the MXD sites.

5.2 Data and methods

The proxy network represents all available MXD chronologies longer than 600 years from the Northern Hemisphere (NH). Data were processed using Regional Curve Standardization (Esper et al. 2003) and scaled (Esper et al. 2005) to local grid-point temperatures in order to derive local temperature reconstructions. The NH average is a 'composite-plus-scaling' (CPS) reconstruction (Von Storch et al. 2006) of extratropical (30-90°N) land-temperatures during the summer months June-August. For details see Schneider et al. (2015).

The instrumental target is the CRUTEM4v-dataset (Jones et al. 2012) with the same spatial (30-90°N, landmass) and seasonal (June-August) coverage. The gridded data reach back to 1850 for many parts of Europe and central North America, but temperature readings do not start before the late 19th or early 20th century for most grid points. In order to provide homogeneous spatial coverage, the NH-mean was calculated after applying a gap filling procedure via the regularized expectation maximization algorithm using ridge regression (Schneider 2001). At the local scale, this method can introduce significant variance changes at grid points with large uncertainty ranges for the infilled data portion. This bias applied to six sites within the network of 15 sites (see Fig. 5-1c and f). Abruptly declining variance in the early portion of these records was adjusted to the level of the late 20th century in order to allow reasonable comparisons on a site-by-site level.

Past climate forcing of volcanic eruptions is usually based on sulphate deposition from multiple ice cores (Masson-Delmotte et al. 2013). Here, proxy and observational data were analyzed for volcanic signals considering the latest global ice core record (Crowley & Unterman 2013). It comprises a reconstruction of stratospheric sulphate expressed in aerosol optical depth (AOD) estimates. We include the volcanic events exceeding an AOD of 0.03 since 1874. Prior to that year the network of observational data is very sparse and the amount of gridpoints with data available is below 25%. The analysis period ends in 1976 representing the last year of the oldest (i.e. first developed) MXD chronology. Accordingly we included 1883 (Krakatau, Indonesia), 1902 (Santa Maria, Guatemala), 1912 (Novarupta, Alaska) and 1963 (Agung, Indonesia) with peaking AOD values in 1884, 1903, 1912 and 1964. The temporal lag for tropical eruptions is caused by the delayed dispersion of the ash-column towards higher latitudes. Since AOD values remain at an elevated level for at least one more year, we also consider this subsequent year. Temperature anomalies in response to volcanic activity were calculated with respect to the 5 pre-eruption years and averaged over the four eruptions.

In order to illustrate how the volcanic signal in proxy reconstructions can be affected by data processing, NH temperatures were reconstructed using the observations from the 15 gridboxes closest to the proxy sites. This pseudo-reconstruction, free of proxy induced noise, was rebuilt a 1000 times using alternative proxy networks, each consisting of 5 randomly chosen input records per continent (North America, Europe and Asia).

5.3 Results

Averaging summer temperatures during years of peaking stratospheric sulphate injection yields widespread cooling in the NH with a mean of 0.35°C below the 5 pre-eruption years (Figs. 5-1a and e). Central North America, southern Europe, western and eastern Asia are key cooling regions, whereas northwestern North America, eastern Europe and central Asia either show no significant cooling or they warm slightly. This pattern is replicated by the proxy records with an outstanding cooling response in northwestern Asia. Local gridpoint temperatures suggest a very similar cooling magnitude in line with the proxy data, and although the response in observational data is on average slightly lower, there is no clear evidence for a general over- or underestimation (Figs. 5-1c and e).

The second year after the sulphate spike is still dominated by cool conditions (-0.25°C on average), but with a clear shift of the key cooling regions (Fig. 5-1b). The most obvious change is found over Europe, where significant cooling affects the central and northern regions, while southern Europe and western Asia already display warming anomalies. The proxy records again agree with this pattern, and in keeping with the reduced overall cooling, there is more heterogeneity in the local responses: While some sites show even stronger cooling compared to the first post-volcanic year, others exhibit warming relative to the 5 pre-eruption years (Figs. 5-1d and f). Averaging the 15 local estimates results in a somewhat stronger cooling than the average of the spatial field over the whole hemisphere. In the first post-volcanic year, reconstructed and observed temperatures are well below the overall average, whereas in the second year only the reconstructed anomalies suggest a slight overestimation (Fig. 5-1e and f).

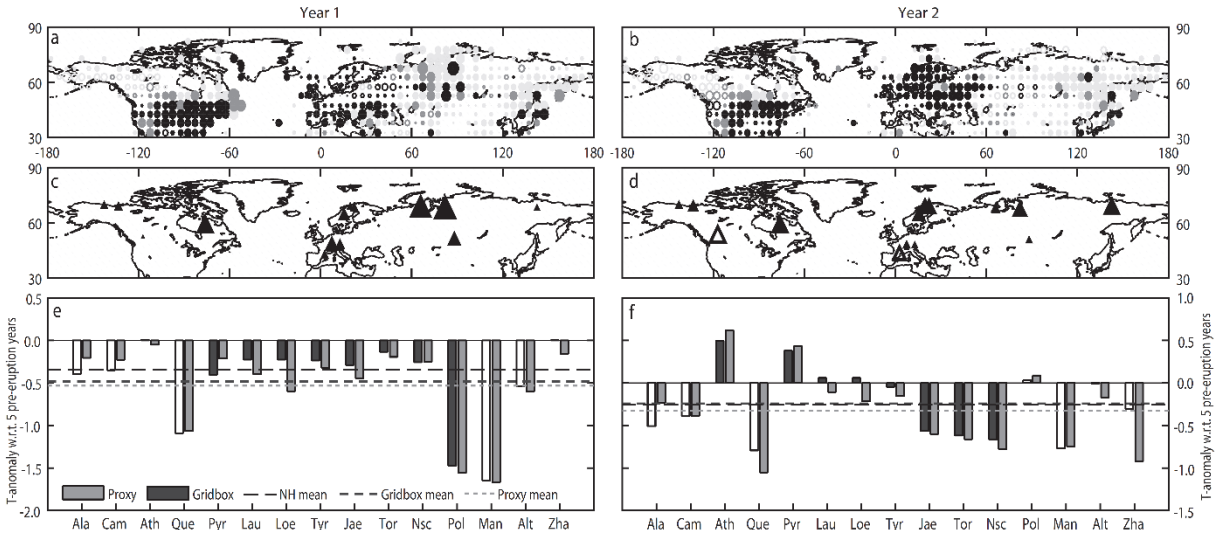


Figure 5-1 | Summer temperature cooling in response to volcanic eruptions in 1884, 1903, 1912 and 1964. (a) Anomalies of gridded summer temperature in years with peaking AOD with respect to the 5 pre-eruption years. Strongest cooling (warming) is indicated with the biggest filled (unfilled) dots. The greyscale represents the number of events covered by a gridbox before gap filling (light grey: 1964, medium grey: 1903, 1912, 1964, dark grey: 1884, 1903, 1912, 1964). (b) As in (a), but for the subsequent year. (c) MXD-sites used for the NH-reconstruction. Lowest (highest) MXD-values in the years with peaking AOD are indicated with the biggest filled (unfilled) triangles. (d) As in (c), but for the subsequent year. (e) Summer temperature anomalies for the 15 MXD sites in years with peaking AOD derived from proxy reconstructions and the gridded temperature field. Unfilled bars indicate gridboxes with short temperature records. A gap filling and variance stabilization were applied. (f) As in (e), but for the subsequent year.

For large-scale examinations, proxy chronologies are compiled into one record, which is then scaled to the NH mean temperature (CPS reconstruction). The NH reconstruction based on the 15 MXD chronologies, representing sites with relatively strong cooling, reveals a considerably reduced response to volcanic forcing compared to the fully-sampled NH mean temperature. Using the reconstructed temperatures as response estimates, the summers were only 0.28°C (0.16°C) cooler in the first (second) year following an eruption. If local grid point temperatures for the 15 proxy sites are used as inputs in a pseudo-reconstruction, the observed cooling is similarly weakened: -0.28°C and -0.14°C, respectively for the first and second years.

As these findings contradict the overestimation of cooling found on local scale (Fig. 5-1e), we recomputed the pseudo reconstruction based on observational data using a randomized site selection scheme. This approach should result in a balanced mix of reconstructions that over- as well as underestimate volcanic cooling with a relatively even distribution around the actual cooling anomaly. The majority of pseudo reconstructions, however, exhibited less volcanic cooling than the overall average (Fig. 5-2). Even doubling the number of input records to a total of 30 sites does not change this result.

5.4 Discussion

Investigating the spatial pattern of summer temperature in response to volcanic eruptions revealed that some regions are not cooling despite massive releases of radiation absorbing sulphate into the stratosphere. The warming in central Asia and western North America is somewhat questionable, however, as data coverage is particularly sparse in these regions and the signal mainly reflects conditions in response to a single eruption (the 1960's eruption of Agung), which caused no clear cooling spike in the NH mean. Some of the observed temperature changes are certainly not significant considering the small number of studied events and uncertainties in the temperature field, i.e. cooling patterns might change if a larger number of events were included (Esper et al. 2013). Nevertheless, the displacement of cooling in the second year is of particular relevance for a potential bias induced by an uneven spatial distribution of proxy records: While in the current network, average cooling at the 15 proxy-sites was much stronger than the NH average in the first year, the values agreed much better in the second year. This indicates that a proxy network that appropriately represents cooling in the second post-eruption year is not necessarily accurate in the first year.

Spatial assembling and processing of the proxy data is an additional source of uncertainty. For the events in 1883, 1902, and 1912, pseudo reconstructions suggest a systematic underestimation of cooling, which is rarely found in years without volcanic forcing. Even the cold anomaly in 1907 (Fig. 5-2) can be ascribed to a volcanic eruption (Ksudach, Russia). There are also short periods of warmth that are difficult to reproduce with any proxy network. This phenomenon, however, cannot be associated with a common driver.

A reason for the underestimation of post-volcanic cooling in sparse proxy networks can be a more spatially homogeneous temperature field when external forcing is active. A reduced percentage of internal variability in such years results in less noise cancellation when calculating large-scale averages and thereby an enhanced temperature peak with respect to long-term variance. For a proxy network of limited spatial coverage, noise cancellation is less effective and, thus, the cooling peak less pronounced. This effect necessarily yields an underestimation of forced temperature changes when interpreting reconstructions based on a proxy network with limited spatial coverage using CPS or linear regression.

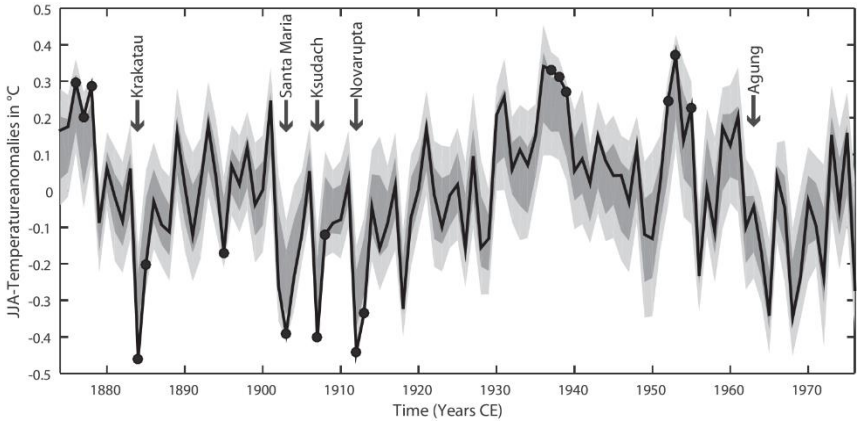


Figure 5-2 | Summer temperature derived from the average of all grid points (black line) and from 1000 pseudo reconstructions (dark grey: 25th-75th percentile, light grey 10th and 90th percentile). Black dots indicate years in which the NH average exceeds the 25-75 percentile range. Arrows indicate selected volcanic eruptions.

5.5 Conclusion

Our analysis focused on potential biases in assessing the strength of volcanic forcing using proxy-based temperature reconstructions. At the local scale there is no systematic deviation from the instrumental record in MXD-based temperature estimates, which is in line with the findings in Esper et al. (2015) who used a different set of volcanic events and similar proxy data. The hemispheric integration, as analyzed herein, revealed that there is no ideal spatial proxy distribution to prevent over- or underestimation of post-volcanic climatic cooling because the response pattern changes over time. The proxy network used in this study includes a bias towards regions with stronger cooling in the first year that abates in the second year. A way to overcome this bias would be to drastically increase the number of predictor chronologies.

By merging data in large-scale reconstructions, it is likely that volcanic cooling is underestimated even if the proxy chronologies are a perfect representation of local temperature and of the average NH climatic response. During years of volcanic activity, climate variability is additionally altered by external forcing, probably changing the spatial patterns in the temperature field resulting in different spatial characteristics of temperature anomalies that eventually suppress the volcanic signal in the proxy reconstruction. In contrast to the proxy distribution, this problem is not implicitly resolved using a denser proxy network, but it is possible to estimate the size of the effect by investigating the ratio between peak amplitude and the long-term variance.

References

Crowley, T. J., & Unterman, M. B. (2013): Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth System Science Data, 5*(1), 187-197.

Esper, J., Cook, E. R., Krusic, P. J., Peters, K., & Schweingruber, F. H. (2003): Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies. *Tree-Ring Research, 59*(2), 81-98.

Esper, J., Frank, D.C., Wilson, R. J. S., & Briffa, K.R. (2005): Effect of scaling and regression on reconstructed temperature amplitude for the past millennium. *Geophysical Research Letters* 32, doi: 10.1029/2004GL021236.

Esper, J., Schneider, L., Krusic, P. J., Luterbacher, J., Büntgen, U., Timonen, M., . . . Zorita, E. (2013): European summer temperature response to annually dated volcanic eruptions over the past nine centuries. *Bulletin of Volcanology, 75*(7).

Esper, J., Schneider, L., Smerdon, J., Schöne, B., & Büntgen, U. (2015): Signals and memory in tree-ring width and density data. *Dendrochronologia, 35*, 62-70.

Fritts, H. C. (1976). Tree rings and climate / H. C. Fritts: Academic Press.

Gao, C. 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).

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012): Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research-Atmospheres, 117*.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González Rouco, J. F., . . . Timmermann, A. (2013). Information from Paleoclimate Archives. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. M. Midgley (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.

Robock, A. (2005): Cooling following large volcanic eruptions corrected for the effect of diffuse radiation on tree rings. *Geophysical Research Letters, 32*(6).

Schneider, L., Smerdon, J. E., Buntgen, U., Wilson, R. J. S., Myglan, V. S., Kirdyanov, A. V., & Esper, J. (2015): Revising midlatitude summer temperatures back to AD600 based on a wood density network. *Geophysical Research Letters, 42*(11), 4556-4562.

Schneider, T. (2001): Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate, 14*(5), 853-871.

Tingley, M. P., Stine, A. R., & Huybers, P. (2014): Temperature reconstructions from tree-ring densities overestimate volcanic cooling. *Geophysical Research Letters, 41*(22), 7838-7845.

Von Storch, H., Zorita, E., Jones, J., Gonzalez-Rouco, F., & Tett, S. (2006): Testing climate reconstructions - Response. *Science, 312*(5782), 1872-1873.

6. Revising mid-latitude summer-temperatures back to AD 600 based on a wood density network

Lea Schneider[1], Jason E. Smerdon[2], Ulf Büntgen[3], Rob J. S. Wilson[2,4], Vladimir S. Myglan[5], Alexander V. Kirdyanov[6] and Jan Esper[1]

[1] *Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany*

[2] *Lamont-Doherty Earth Observatory, Palisades, NY 10964, USA*

[3] *Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland*

[4] *School of Geography and Geosciences, University of St Andrews, St Andrews KY16 9AL, Scotland, UK*

[5] *Institute for the Humanities, Siberian Federal University, Krasnoyarsk, 660041, Russia*

[6] *V.N. Sukachev Institute of Forest SB RAS, Akademgorodok, Krasnoyarsk, 660036, Russia*

Published in Geophysical Research Letters 42, doi: 10.1002/2015GL063956

6.1 Introduction

Integration of climate proxies and development of continental-scale temperature reconstructions of the past millennium has greatly improved our understanding of the spatial patterns of long-term climate variability (*PAGES 2k Consortium*, 2013). It has been shown (*Bunde et al.*, 2013; *Franke et al.*, 2013), however, that the spectral characteristics of temperature fluctuations in proxy-based reconstructions differ systematically from observations, with the proxies overestimating the ratio of low- to high-frequency variability. This "red bias" (*Franke et al.*, 2013) inherent to reconstruction spectra is a fundamental constraint limiting the assessment of pre-instrumental climatic extremes (*Anchukaitis et al.*, 2012; *D'Arrigo et al.*, 2013; *Esper et al.*, 2013), which in turn has implications for the estimation of the Earth´s climate sensitivity through comparisons of climate model simulations with proxy-based reconstructions of preindustrial temperature variability (*Frank et al.*, 2010).

Almost all millennial-scale temperature reconstructions assessed in the IPCC AR5 (*Masson-Delmotte et al.*, 2013) are either solely based on tree-ring data or a combination of these data with multiple additional proxies - most of which have lower temporal resolution. The inter-annual variance in these records is nevertheless dominated by tree-ring width (TRW) measurements, a parameter known to contain substantial biological memory (*Frank and Esper*, 2005) due to the persistence of needle generations and storage of reserves throughout dormancy (*Fritts*, 1976). These effects translate into an overestimation of the persistence of post-volcanic temperature cooling (*Frank et al.*, 2007a; *Masson-Delmotte et al.*, 2013) and biased spectral characterization of tree-ring based temperature reconstructions (*Bunde et al.*, 2013; *Franke et al.*, 2013). Maximum-latewood-density (MXD), in contrast, has the capacity to more directly respond to and recover from pulse-like cooling (*Briffa et al.*, 1998) and disturbance (*Esper et al.*, 2007a) events. Less physiological noise results in a stronger common signal among trees and stronger coherence between observational target and proxy data (*Esper et al.*, 2013). Moreover, compared to TRW, the relative magnitude of biologic age trends in MXD is smaller with respect to the standard deviation of the samples, therefore reducing the distortion effects that arise during the necessary process of detrending, which again supports the fidelity of millennial-scale temperature variance estimated from this parameter (*Esper et al.*, 2012). Nevertheless, a large-scale compilation of the longest MXD chronologies has not yet been used to estimate hemispheric-scale high-to-low frequency variability in temperature over the last millennium.

We introduce herein a network of the 15 longest available MXD records (Tab. S1) from Asia, Europe, and North America (4/7/4), assess the climate signal inherent in these data, and develop an alternative NH reconstruction. The proxy network comprises in total less site-chronologies than the most recent TRW-based reconstruction (*D'Arrigo et al.*, 2006), but is similarly replicated at the onset of the last millennium. Excluding TRW from our evaluations allows us to assess the potential benefits of a parameter free of biological memory effects and its ability to theoretically overcome the spectral limitation associated with TRW-derived estimates.

6.2 Data and Methods

Proxy Network

Proxy data for this study are publicly available at the International Tree-Ring Data-Base or contributed by the original authors (*Anchukaitis et al.*, 2013; *Briffa et al.*, 2013; *Büntgen et al.*, 2006; *Büntgen et al.*, 2008; *Esper et al.*, 2007b; *Esper et al.*, 2012; *Luckman and Wilson*, 2005; *Melvin et al.*, 2013; *Myglan et al.*, 2012; *Schweingruber et al.*, 1988). All datasets were corrected separately for biological and site-related noise using Regional Curve Standardization (RCS) (*Briffa et al.*, 1992; *Esper et al.*, 2003). Residuals between the power-transformed (*Cook and Peters*, 1997) MXD series and an expected age-related growth curve were calculated and averaged using the robust bi-weight mean to form site chronologies. Tree-ring data from Athabasca and Mangazia comprised site- or species-specific offsets in the mean growth level. These sites were divided into subgroups, according to species composition and growth rate, detrended separately using RCS, and re-combined to form mean site chronologies (Text S1) (*Luckman and Wilson*, 2005).

The varying sample replication of the single site chronologies necessitates variance stabilization, here achieved by fitting a 100-year spline through the absolute departure values and dividing the chronologies by this spline curve. Other plausible tree-ring standardization methods (e.g. signal-free detrending or replication-based variance-correction) were also tested, yielding similar results (Fig. S6-1, Tab. S6-2). Fifteen detrended site chronologies longer than 600 years and replicated with at least 3 samples served as the final predictor network (Fig. 6-1, Tab. S6-1). The shortest record starts in 1363 and the oldest terminates in 1976, the last year of the calibration period. All records passed a benchmarking experiment evaluating the strength of local JJA-temperature responses via a random process. A 95% confidence level is derived from the probability distribution of correlations between local temperature and 1000 red noise time-series with first-order autocorrelation matching each MXD record. Local correlations between each MXD chronology and temperature served as the weights in the hemispheric composites

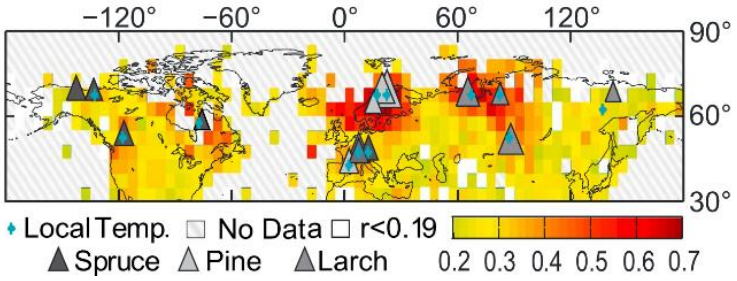


Figure 6-1 | June-July-August temperature correlation field represented by NH extratropical tree-ring density records. The MXD data collection (triangles) comprises temperature sensitive pine, larch, and spruce site chronologies. The size of the triangles represents the length of the chronologies (637-2187 years). Colors indicate the average Pearson's correlation coefficient of all statistically significant ($p < 0.05$, one sided) correlations between the MXD chronologies and gridded observational temperatures from 1901-1976. White cells indicate areas not (significantly) represented by any MXD chronology, and blue stars mark the grid cells used for local weighting and pseudo-proxy experiments.

used in the reconstruction procedure. Records from the Alps and Scandinavia were averaged to regional composites to mitigate the European sampling bias.

Instrumental Target

The instrumental target is monthly mean summer temperature (June-August) derived from the 5x5° CRUTEM4v network (30°-90°) (*Jones et al.*, 2012). The field was truncated in 1901, due to an increasing number of empty grid-cells (53% at the onset of the 20th century, 86% in 1850) and subsequent variance reduction in the NH mean due to the gapfilling via the regularized expectation maximization algorithm using ridge regression (*Schneider*, 2001). Regional and hemispheric temperature averages were weighted by area, using the cosine-latitude of local grid cell centers. For analyses of growth responses, we chose CRUTEM4v temperatures from the grid cell containing the corresponding tree site. Cells with very short records were replaced by records from neighboring cells (e.g. in eastern Siberia; Fig. 6-1). Monthly averaged diurnal temperature maxima for JJA, tested as an alternative instrumental target (*Wilson et al.*, 2007), yielded neither consistent nor significant improvements in the calibration results.

Calibration/Validation

The regional MXD chronologies were combined into an ensemble estimate of hemispheric temperature variability by calculating weighted composites based on moving correlations with local temperature. The composite averages were scaled against NH extra-tropical temperatures over the 1901-1976 period (*von Storch et al.*, 2009). Calibration and validation statistics (R^2 and RE metrics (*Briffa et al.*, 1988)) were estimated using a 38-year holdout window incremented by one year between 1901 and 1976 to derive an ensemble of 76 plausible reconstruction members (Fig. S6-2). The procedure enabled the estimation of a calibration error derived from minimum and maximum reconstructed temperatures for each year. The decreasing number of predictors back in time, typical for dendroclimatological studies, results in increasing uncertainty in earlier periods, which was accounted for by recalculating the reconstruction and associated errors each time a site-record dropped out (Fig. S6-3) (*Cook et al.*, 2002). Such a nested approach also allowed the variance to be adjusted to account for the diminishing predictor-network, as well as validation metrics and sampling error estimates to vary in time. Artificial reduction of site-specific replication by randomly drawing from all MXD series in the 20th century revealed only negligible differences in estimated errors (Fig. S6-4). Moreover, the sensitivity of the reconstruction against a changing predictor network and different calibration choices was assessed (Figs. S6-5,S6-6). The recent end of the proxy network is characterized by a quickly declining number of sites. The 1976-2002 period was thus reconstructed with only one nest containing 7 records extending into the 21st century.

6.3 Results

The MXD network has reasonable coverage over the last millennium with all datasets reaching back to AD 1363. Each chronology explains a significant fraction of regional summer-temperature variability, verified in a benchmarking experiment, and the combined correlation fields ($p < 0.05$) from all records represent 90% of the NH extra-tropical landmass (Fig. S6-1). In addition to validating the MXD data against local observations and regional temperature fields, the combined site chronologies display strong coherence at continental (N-America, Europe, Asia) scales (Fig. S6-2). The explained variance among 30-year low-pass filtered continental means over the common AD 1363-1976 period is $\overline{R^2}_{1363\text{-}1976} = 0.45$ ($R^2_{America/Asia} = 0.47$, $R^2_{America/Europe} = 0.40$, $R^2_{Asia/Europe} = 0.49$) indicating sensitivity to common forcing and/or modes of variability at multi-decadal to centennial time-scales. The correlation coefficients decline during the early, less-replicated portion of the records. At inter-annual time-scales there is no coherence between the continental composite ($\overline{R^2}_{1363\text{-}1976} = 0.03$), corresponding to the reduced spatial coherency of surface temperature fields in the higher frequency domain (*Jones et al.*, 2012).

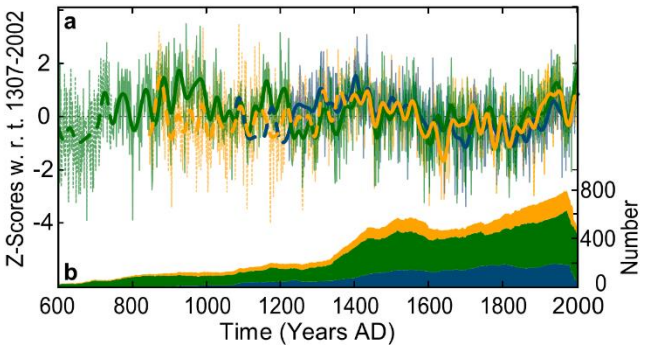


Figure 6-2 | Continental MXD chronologies. a, Continental-scale MXD compilations derived from 4 site chronologies in North America (blue), 7 chronologies in Europe (green), and 4 chronologies in Asia (yellow). Thick curves are filtered reconstructions using a 30-year low-pass butterworth filter. Continental averages with less than three sites are dashed. b, The overall number of samples by continent.

The large-scale reconstruction correlates (Pearson's r) at $r_{1901\text{-}1976} = 0.60$ with modern JJA temperatures, and estimates pre-instrumental temperatures to have spanned 0.54±0.34°C between the coldest 30-year periods during the LIA (1627-1656 and 1809-1838) and the warmest such period during late medieval times (1396-1425) (Fig. 6-3). The same temperature range is reconstructed from the LIA to the warmest most recent 30-year period (1927-1956). The reconstruction indicates moderate summer temperatures from ~ AD 900-1600 at the hemispheric scale, with maxima occurring in the late 9th and mid-15th centuries, followed by cooler summers from ~ 1600-1900, and subsequent 20th-century warming.

The new summer temperature reconstruction was validated using stepwise ensemble techniques, a sliding calibration/verification approach, and pseudoproxy experiments (Text S6-1; Figs. S6-2, S6-3 and S6-7; see *Smerdon* 2012 for a review). Despite considerable empirical uncertainty, the reconstruction proved robust to a number of methodological choices during calibration (Figs. S6-6a-h). Potential divergence issues (*Esper et al.*, 2010; *Wilson et al.*, 2007), which become apparent in some of the local

records in the late 20th century, were evaluated by compiling two independent records derived from 8 "non-diverging" and 5 "diverging" MXD site chronologies (Figs. S6-8 – S6-10). Despite low site replication, both subsamples share a considerable amount of low-frequency variability in their common period (1363-1976) followed by a more significant offset in the late 20th century. In the complete composite, decoupling between the instrumental and proxy-derived temperatures is noted in the most recent decades. The new JJA temperature reconstruction also coheres closely with a shorter reconstruction based on 300+ MXD sites distributed over the NH extra-tropics, the so-called "Schweingruber MXD network" (*Briffa et al.*, 2001) (Fig. 6-3b). These two almost independent datasets (estimated data overlap < 1%) correlate at r = 0.70 over the past ~ 300 years. An offset of ~0.2°C evolving before 1450 is likely related to the changing spatial coverage in the Schweingruber network, while sample and site replication remain stable during this period in our new network of long MXD records.

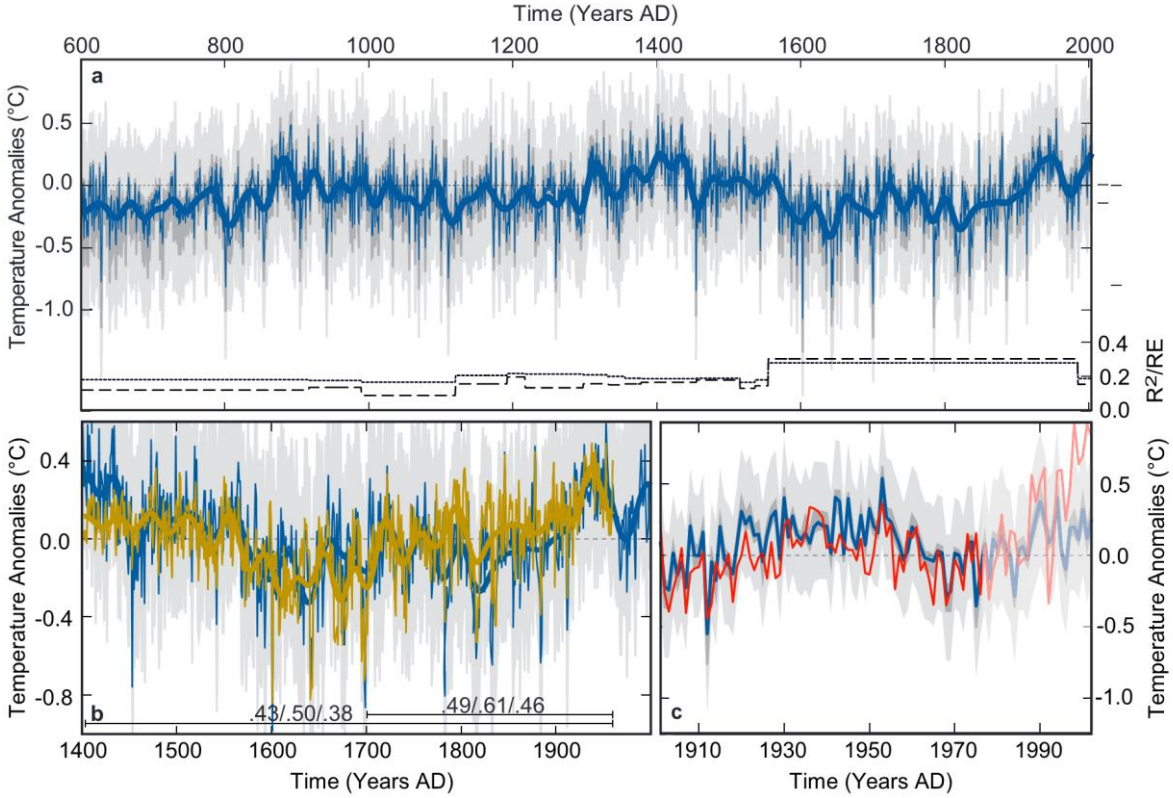


Figure 6-3 | NH extra-tropical temperature reconstruction. a, Nested June-July-August temperature reconstruction (blue) with combined uncertainty estimates derived from the calibration (dark grey) and sampling (light grey) errors. Anomalies w.r.t. 1961-1990. Bottom panel shows the time-varying explained variance (R2; dotted) and reduction of error statistic (RE; dashed) of differently replicated nests over the past 1400 years. All nests passed the 99% threshold of a red-noise benchmarking exercise (see Supplementary Methods). b, MXD-based temperature-reconstructions from this study (blue) and ref. [Briffa et al., 2001] (brown) as anomalies w.r.t 1402-1960. Values at the bottom are explained variances (R2) between the actual, low-pass, and high-pass filtered records (the latter is not shown graphically). c, The MXD-based reconstruction as in a (blue) together with instrumental JJA temperatures averaged over the NH extra-tropics (red) during the 1901-1976 period common to all predictors. The less replicated most recent decades are shown in faded colours. Anomalies w.r.t. 1961-1990.

6.4 Discussion and Conclusion

Comparison of the new MXD-based reconstruction with the multi-proxy and TRW-based records assessed by the IPCC (*Masson-Delmotte et al.*, 2013) reveal substantial differences at low and high frequencies (Fig. 6-4). Modern warming is poorly represented in the new record questioning the ability of MXD to capture very warm temperatures. The last several decades of the reconstruction are, however, represented by a smaller number of chronologies. Among these, only a couple appear to be impacted by divergence, suggesting deficiencies regarding spatial representation or chronology quality, rather than a whole-scale limitation of the proxy. While the overall sequence of warmer conditions during medieval times, subsequent LIA cooling, and 20th-century warming is in accordance with the IPCC collection of reconstructions, the new record deviates substantially regarding the initiation of the LIA and the millennial-scale temperature amplitude. The Medieval Warm Period is less pronounced in our reconstruction, and summer temperatures remain at a relatively warm level throughout the ~1300-1450 period, when the existing reconstructions indicate a prolonged cooling trend. Residuals are largest in the late 15th and 16th centuries during which the MXD-based estimates point to warmer conditions, thereby delaying the onset of the LIA until the early 17th century – a substantial change important to the discussion about potential drivers of NH cooling (*Schurer et al.*, 2014) and the interpretation of potential societal responses (*Büntgen et al.*, 2011).

Some of these differences possibly arise as a result of varying seasonal or spatial coverage. Although most of the reconstructions in the IPCC-ensemble are likely biased toward a summer-signal, many of them target annual temperatures. For TRW-based reconstructions, this assumption is ascribed to the longer season portrayed by ring-width due to biological driven persistence (*D'Arrigo et al*., 2006; *Frank et al.*, 2007). Thus, the more confined seasonal fingerprint of MXD could explain deviations to some extent. Other studies, in contrast, state that multi-decadal variability should be largely independent of the season chosen (*Cook et al.*, 2004) and we find some evidence for that at least during the observational period (Figs. S6-6i-j). By focusing strictly on long MXD-records this reconstruction has a reduced spatial coverage during the calibration period compared to the most recent tree-ring based reconstruction (*D'Arrigo et al.*, 2006). Coherency with observational data is nevertheless good in the mid- and low-frequency domains and does not imply discrimination based on the network-extent (Fig. S6-6 – S6-11). Despite these findings, it is also clear that spatially expanding the extent of the NH MXD-network would greatly benefit future paleoclimatological research

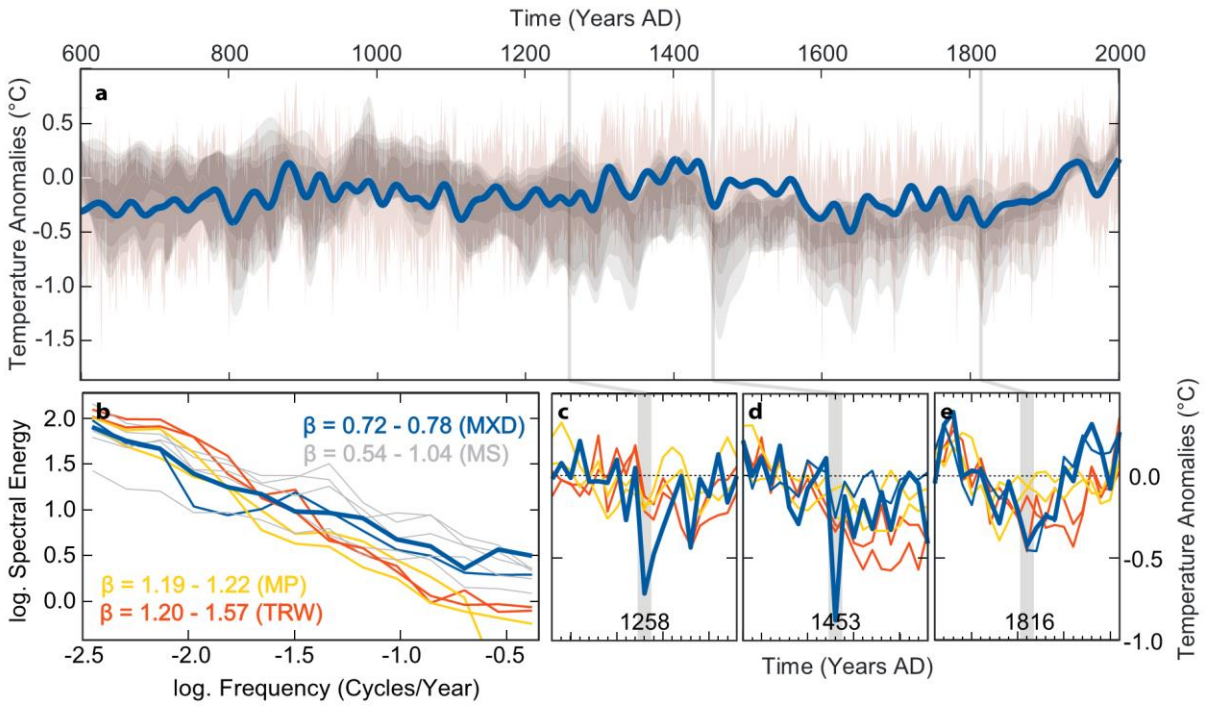


Figure 6-4 | Comparison of Northern Hemispheric temperature reconstructions. a, This study (blue) shown together with the distribution quantiles (grey shading) derived from 15 reconstructions assessed in the IPCC AR5 [Masson-Delmotte et al., 2013] after 30-year low-pass filtering. Note that most reconstructions are scaled to annual mean temperature. Anomalies w.r.t. 1901-1976. Grey bars indicate major volcanic eruptions. b, Spectral energy and β-values for this study (bold blue), the MXD-based reconstruction from Briffa et al. [2001] (thin blue), two TRW-based reconstructions (red) [D'Arrigo et al., 2006; Frank et al. 2007b], two multi-proxy-based reconstructions (MP) (yellow) [Moberg et al., 2005; Juckes et al. 2007] and 5 model simulations (grey, MS). Mean and variance are adjusted over the 1923-1960 period in the time-domain, representing the common period during the calibration interval. c-e, Temperature anomalies following the volcanic eruptions in 1257, 1452, and 1815 w.r.t. 10 preceding years. Colors as in b.

A remaining rationale for deviating long-term trends and the delayed LIA onset likely originates from the different response of MXD data, compared to TRW, to a cluster of major volcanic eruptions in the 13th and subsequent centuries (compared to the relative lack of such events during medieval times) as revealed in Greenland and Antarctic ice core data (*Gao et al.*, 2008). For the TRW and multiproxy reconstructions, spectral power is shifted from annual and multi-decadal timescales to lower frequencies, compared to the model simulations and MXD reconstruction, which indicates prolonged responses to climate-variations in the traditional data (Fig. 6-4b; Text S6-1). At the large spatial scale addressed here, these differences become visible during the periods subsequent to the three largest volcanic eruptions of the past millennium in 1257, 1452, and 1815 (Figs. 6-4c-e). Whereas the TRW-dominated reconstructions show a weak and temporally extended post-volcanic response, the MXD-based reconstructions reveal distinct post-volcanic cooling and recover to mean climatology much faster. Of particular interest is the temperature drop after the 1257 event, recorded in ice cores, providing further evidence for the precise dating of tree-ring records over the entire millennium (*Anchukaitis et al.*, 2012). The MXD-based reconstruction from *Briffa et al.* (2001) indicates less cooling after the 1452 event, most likely related to decreased spatial replication in the earliest part of this record. Cooling in response

to Tambora (1815) is less distinct as it coincides with the generally cold Dalton Minimum (~1790-1830) (*Wagner and Zorita*, 2005).

Model-simulated temperatures support the fast recovery from volcanic-induced cooling (*Wigley et al.*, 2005) as recorded in the MXD-based reconstruction, although some models reveal a much stronger – and potentially too strong – temperature response (*Landrum et al.*, 2013). The temperature trend from the Medieval Warm Period into the LIA also appears reduced in some model simulations (*Fernández-Donado et al.*, 2013). It remains unclear, however, whether this is due to less persistent volcanic cooling, an underestimation of internal variability, or the influence of solar variability. Considering the differing spectral characteristics identified herein (Fig. 6-4b), and confirmed by *Bunde et al.* (2013) and *Franke et al.* (2013), it seems important that our new MXD-based reconstruction has similar spectral properties to the simulated temperatures at frequencies ranging from 0.5-0.003 cycles per year. Although millennium-long oscillations are poorly sampled, our calculated beta-values indicate that the new temperature history is not red-biased, as is typical for the TRW and multi-proxy reconstructions. Consideration of the new reconstruction might thus strengthen efforts to estimate the climate sensitivity to volcanic eruptions and the persistence of subsequent lower tropospheric cooling. The late initiation of the LIA around 1580 revealed in our reconstruction might additionally help to further elucidate ecosystem responses and historical linkages throughout the past millennium.

6.5 Acknowledgements

We thank all of our colleagues for developing millennial-length MXD chronologies and making these data available. Supported by the Mainz Geocycles Research Centre. Data are available via the NOAA online server. Lamont contribution #7902.

6.6 References

Anchukaitis, K. J., R. D'Arrigo, L. Andreu-Hayles, D. Frank, A. Verstege, A. Curtis, B. M. Buckley, G. C. Jacoby, and E. R. Cook (2013), Tree-ring-reconstructed summer temperatures from northwestern North America during the last nine centuries, *Journal of Climate, 26*(10), 3001-3012.

Anchukaitis, K. J., et al. (2012), Tree rings and volcanic cooling, *Nat. Geosci., 5*(12), 836-837.

Briffa, K. R., P. D. Jones, J. R. Pilcher, and M. K. Hughes (1988), Reconstructing summer temperatures in northern Fennoscandinavia back to Ad 1700 using tree-ring data from Scots Pine, *Arctic Alpine Res., 20*(4), 385-394.

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*(6684), 450-455.

Briffa, K. R., T. M. Melvin, T. J. Osborn, R. M. Hantemirov, A. V. Kirdyanov, V. S. Mazepa, S. G. Shiyatov and J. Esper (2013), Reassessing the evidence for tree-growth and inferred temperature change during the Common Era in Yamalia, northwest Siberia, *Quat. Sci. Rev. 72*, 83-107.

Briffa, K. R., T. J. Osborn, F. H. Schweingruber, I. C. Harris, P. D. Jones, S. G. Shiyatov, and E. A. Vaganov (2001), Low-frequency temperature variations from a northern tree ring density network, *J. Geophys. Res.-Atmos., 106*, 2929-2941.

Briffa, K. R., P. D. Jones, T. S. Bartholin, D. Eckstein, F. H. Schweingruber, W. Karlén, P. Zetterberg, and M. Eronen (1992), Fennoscandian summers from Ad-500 - temperature-changes on short and long timescales, *Clim. Dynam., 7*(3), 111-119.

Bunde, A., U. Büntgen, J. Ludescher, J. Luterbacher, and H. von Storch (2013), Is there memory in precipitation?, *Nat. Clim. Change, 3*(3), 174-175.

Büntgen, U., D. Frank, D. Nievergelt, and J. Esper (2006), Summer temperature variations in the European Alps, AD 755-2004, *Journal of Climate, 19*(21), 5606-5623.

Büntgen, U., D. Frank, H. Grudd, and J. Esper (2008), Long-term summer temperature variations in the Pyrenees, *Clim. Dynam., 31*(6), 615-631.

Büntgen, U., et al. (2011), 2500 years of European climate variability and human susceptibility, *Science, 331*(6017), 578-582.

Cook, E. R., and K. Peters (1997), Calculating unbiased tree-ring indices for the study of climatic and environmental change, *Holocene, 7*(3), 361-370.

Cook, E. R., R. D'Arrigo, and M. E. Mann (2002), A well-verified, multiproxy reconstruction of the winter North Atlantic Oscillation index since AD 1400, *Journal of Climate, 15*(13), 1754-1764.

Cook, E. R., J. Esper, R. D'Arrigo (2004), Extra-tropical Northern Hemisphere temperature variability over the past 1000 years, *Quat. Sci. Rev., 23*, 2063-2074.

D'Arrigo, R., R. Wilson, and G. Jacoby (2006), On the long-term context for late twentieth century warming, *J. Geophys. Res.-Atmos., 111*(D3).

D'Arrigo, R., R. Wilson, and K. J. Anchukaitis (2013), Volcanic cooling signal in tree ring temperature records for the past millennium, *J. Geophys. Res.-Atmos., 118*(16), 9000-9010.

Esper, J., E. R. Cook, P. J. Krusic, K. Peters, and F. H. Schweingruber (2003), Tests of the RCS method for preserving low-frequency variability in long tree-ring chronologies, *Tree-Ring Res., 59*(2), 81-98.

Esper, J., U. Büntgen, D. Frank, D. Nievergelt, and A. Liebhold (2007a), 1200 years of regular outbreaks in alpine insects, *P. R. Soc. B., 274*(1610), 671-679.

Esper, J., U. Büntgen, D. Frank, T. Pichler, and K. Nicolussi (2007b), Updating the Tyrol tree-ring dataset, in *Tree Rings in Archaeology, Climatology and Ecology, TRACE*, edited by K. Haneca, pp. 80-85.

Esper, J., D. Frank, U. Büntgen, A. Verstege, R. M. Hantemirov, and A. V. Kirdyanov (2010), Trends and uncertainties in Siberian indicators of 20th century warming, *Global Change Biol., 16*(1), 386-398.

Esper, J., et al. (2012), Orbital forcing of tree-ring data, *Nat. Clim. Change, 2*(12), 862-866.

Esper, J., L. Schneider, P. J. Krusic, J. Luterbacher, U. Büntgen, M. Timonen, F. Sirocko, and E. Zorita (2013), European summer temperature response to annually dated volcanic eruptions over the past nine centuries, *B. Volcanol., 75*(7), 736-736.

Fernández-Donado, L. et al. (2013), Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, *Clim. Past, 9*, 393-421.

Frank, D., and J. Esper (2005), Characterization and climate response patterns of a high-elevation, multi-species tree-ring network in the European Alps, *Dendrochronologia, 22*(2), 107-121.

Frank, D., U. Büntgen, R. Böhm, M. Maugeri, and J. Esper (2007), Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target, *Quat. Sci. Rev., 26*(25-28), 3298-3310.

Frank, D., J. Esper, and E. R. Cook (2007), Adjustment for proxy number and coherence in a large-scale temperature reconstruction, *Geophysical Research Letters, 34*, doi: 10.1029/2007GL030571.

Frank, D., J. Esper, C. C. Raible, U. Büntgen, V. Trouet, B. Stocker and F. Joos (2010), Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate, *Nature, 463*, 527-530.

Franke, J., D. Frank, C. C. Raible, J. Esper, and S. Brönnimann (2013), Spectral biases in tree-ring climate proxies, *Nat. Clim. Change, 3*(4), 360-364.

Fritts, H. C. (1976), *Tree rings and climate / H. C. Fritts*, Academic Press, London ; New York.

Gao, C., A. Robock, and C. Ammann (2008), Volcanic forcing of climate over the past 1500 years: An improved ice core-based index for climate models, *J. Geophys. Res., 113*.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice (2012), Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010, *J. Geophys. Res.-Atmos.*, *117*.

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*(4), 591-609.

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, *Journal of Climate*, *26*(4), 1085-1111.

Luckman, B. H., and R. J. S. Wilson (2005), Summer temperatures in the Canadian Rockies during the last millennium: a revised record, *Clim. Dynam.*, *24*(2-3), 131-144.

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

Masson-Delmotte, V., et al. (2013), Information from paleoclimate archives, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley, pp. 383–464, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Melvin, T. M., H. Grudd, and K. R. Briffa (2013), Potential bias in 'updating' tree-ring chronologies using regional curve standardisation: Re-processing 1500 years of Torneträsk density and ring-width data, *Holocene*, *23*(3), 364-373.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlen (2005), Highly variable Northern Hemisphere temperatures reconstructed from low- and high-resolution proxy data, *Nature*, *433*(7026), 613-617.

Myglan, V. S., O. C. Oidupaa, and E. A. Vaganov (2012), A 2367-year tree-ring chronology for the Altai–Sayan region (Mongun-Taiga Mountain Massif), *Archaeology, Ethnology and Anthropology of Eurasia*, *40*(3), 76-83.

PAGES 2k Consortium (2013), Continental-scale temperature variability during the past two millennia, *Nat. Geosci.*, *6*(5), 339-346.

Schneider, T. (2001), Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, *Journal of Climate*, *14*(5), 853-871.

Schurer, A. P., S. F. B. Tett, and G. C. Hegerl (2014), Small influence of solar variability on climate over the past millennium, *Nat. Geosci.*, *7*(2), 104-108.

Schweingruber, F. H., T. Bartholin, E. Schar, and K. R. Briffa (1988), Radiodensitometric-dendroclimatological conifer chronologies from Lapland (Scandinavia) and the Alps (Switzerland), *Boreas*, *17*(4), 559-566.

Smerdon, J.E. (2012), Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, *WIREs Climate Change*, *3*, 63-77, doi:10.1002/wcc.149.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of Cmip5 and the experiment design, *B. Am. Meteorol. Soc.*, *93*(4), 485-498.

Thomson, D. J. (1982), Spectrum estimation and harmonic-analysis, *P. Ieee.*, *70*(9), 1055-1096.

Wagner, S., and E. Zorita (2005), The influence of volcanic, solar and the Dalton Minimum (1790-1830): CO2 forcing on the temperatures in a model study, *Clim. Dynam.*, *25*(2-3), 205-218.

Wigley, T. M. L., C. M. Ammann, B. D. Santer, and S. C. B. Raper (2005), Effect of climate sensitivity on the response to volcanic forcing, *J. Geophys. Res.-Atmos.*, *110*(D9).

Wilson, R., R. D'Arrigo, B. Buckley, U. Büntgen, J. Esper, D. Frank, B. Luckman, S. Payette, R. Vose, and D. Youngblut (2007), A matter of divergence: Tracking recent warming at hemispheric scales using tree ring data, *J. Geophys. Res.-Atmos.*, *112*(D17).

7. Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation

F. Pretis[1], L. Schneider[2], J.E. Smerdon[3] & D.F. Hendry[1]

[1] *Programme for Economic Modelling, Oxford Martin School, University of Oxford, Oxford, UK*

[2] *Department of Geography, University of Mainz, Germany*

[3] *Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA*

Published in Journal of Economic Surveys 30, doi: 10.1111/joes.12148

7.1 Introduction

Breaks in time series come in many shapes and may occur at any point in time – distorting inference in-sample and leading to forecast failure out-of-sample if not appropriately modelled. Often an approximate shape of a break can be postulated *a-priori*, either from previous observations or theory. For example, smooth transitions are common in economic time series following recessions or policy interventions, while sudden drops followed by smooth reversions to the mean are typical in climate time series such as temperature records after a large volcanic eruption (e.g., Kelly & Sear 1984). While the approximate form of a break may be known, the timings and magnitudes of breaks are often unknown. Here we propose an econometric approach for detecting breaks of any specified shape in regression models using an indicator saturation procedure. Our approach is based on recent developments in variable selection within regression models that involve more variables than observations (Castle et al. 2011). By selecting over a complete set of designed break indicators, our approach produces estimates of the break magnitude and timing without imposing limits on the number of breaks that may occur, even at the start or end of a sample.

A structural break is defined as a time-dependent change in a model parameter resulting from a change in the underlying data generating process. For example, a volcanic eruption leading to a rapid climatic cooling corresponds to a temporary shift in the mean of the surface temperature process. The detection of structural breaks in time series has received significant attention in the recent literature – with a growing interest in econometric models of climate change (e.g., Estrada et al. 2013, Pretis, Mann, & Kaufmann 2015). The focus has primarily remained on breaks in the mean through the form of step functions (Step-Indicator Saturation – SIS, Castle, Doornik, Hendry, & Pretis 2015, Pretis 2015b), smooth transition functions (González & Teräsvirta 2008), breaks in regression coefficients (see e.g. Bai & Perron 1998, 2003, Perron & Zhu 2005, Perron & Yabu 2009), or individual outliers or groups of outliers that can be indicative of different forms of breaks (Impulse-Indicator Saturation – IIS, see Hendry et al. 2008).

Broadly grouped into 'specific-to-general' and 'general-to-specific', there exist a plethora of approaches for the detection of structural breaks. Perron (2006) provides a broad overview of specific-to-general methods, some of which are subject to an upper limit on the number of breaks, a minimum break length, co-breaking restrictions, as well as ruling out breaks at the beginning or end of the sample, though Strikholm (2006) proposes a specific-to-general algorithm that allows for breaks at the start or end of a sample and relaxes the break-length assumption.

Indicator saturation (IIS, SIS) provides an alternative approach using an extended general-to-specific methodology based on model selection. By starting with a full set of step indicators in SIS and removing all but significant ones, structural breaks can be detected without having to specify a minimum break length, maximum break number or imposed co-breaking. Crucially this also allows model-selection to be conducted jointly with break-detection as non-linearities, dynamics, theory-motivated variables, and break functions are selected over simultaneously.

Step-functions and impulses are nevertheless only the simplest of many potential break specifications and may not provide the closest approximation to the underlying break. Ericsson (2012) proposes a wide range of extensions to impulse and step-shifts. Here we show that the principle of step-indicator saturation can be generalized to any form of deterministic break function. An advantage over existing methods is an expected higher frequency of detection when a break function approximates the true

break,[1] high flexibility as multiple types of break functions can be selected over, and improvements in forecasting where designed functions act as continuous intercept corrections. Moreover, by being a structured search, the retention of irrelevant effects can be controlled.

The method is illustrated using an econometric model of climate variation – detecting volcanic eruptions in a time series of Northern Hemisphere mean temperature spanning roughly 1200 years, derived from a fully-coupled global climate model simulation. Our technique demonstrates that eruptions can be statistically detected without prior knowledge of their occurrence or magnitude- and hence may prove useful for estimating the past impact of volcanic events using proxy-reconstructions of hemispheric or global mean temperatures. Specifically, this can lead to an improved understanding of the effect of stratospheric aerosols on temperatures (with relevance to geo-engineering and pollution control), and more generally, the break detection procedure can be applied to evaluate policy interventions (e.g., the Montreal Protocol: see Estrada et al. 2013, and Pretis & Allen 2013), correct measurement changes by detecting and subsequently removing shifts, and function as a robust forecasting device.

Section 7.2 introduces the methodology and investigates the properties of break detection in the presence of breaks and under the null of no breaks. Section 7.3 applies the method to detect volcanic eruptions in simulated climate data, and considers designed indicator functions as a robust forecasting device. The conclusions of our work are discussed in section 7.4.

7.2 Break Detection using Designed Indicator Functions

Breaks are intrinsically stochastic without prior knowledge of their timings and magnitudes. Using a full set of break functions allows us to model the responses deterministically. The detection of structural breaks in regression models can be formulated as a model selection problem where we select over a full set of break functions, a subset of which accurately describes the underlying 'true' break. Consider a simple model as:

$$\mathbf{y} = \mathbf{Z}\beta + \varepsilon \tag{7.1}$$

where $\mathbf{y}$ and ε are $(T \times 1)$ vectors, β is a $(k \times 1)$ vector, and $\mathbf{Z}$ is a $(T \times k)$ matrix $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_k)$ of rank k. We investigate the presence of structural breaks in any of the β where $\mathbf{z}$ may be a constant, trend or random variable. For each break type at any point in time for each variable whose coefficient is allowed to break, we augment the above model by a $(T \times T)$ break function matrix $\mathbf{D}$:[2]

$$\mathbf{y} = \mathbf{Z}\beta + \mathbf{D}\gamma + \varepsilon \tag{7.2}$$

where γ is a $(T \times 1)$ vector. The specification of $\mathbf{D}$ is such that the first column $\mathbf{d}_1$ $(T \times 1)$ is set to denote some specified break function $d(t)$ of length L, where $d_{1,t} = d(t)$ for $t \leq L$ and 0 otherwise, $d_{1,t} = 0$ for $t > L$. All further columns $\mathbf{d}_j$ (for $j = 2, ..., T$) in $\mathbf{D}$ are set such that $d_{j,t} = d_{j-1,t-1}$ for $t \geq j$ and 0 otherwise. The break matrix $\mathbf{D}$ is then defined as $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T)$, where $\mathbf{d}_j$ denotes a vector

[1] For example step-indicator saturation exhibits higher power in detecting step shifts than using impulses alone – see Castle, Doornik, Hendry, & Pretis (2015).

[2] For k breaking variables this implies augmenting the full-sample model by k $(T \times T)$ matrices.

with break at time $t = j$:

$$\begin{aligned}
\mathbf{D} &= (\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_T) \\
\mathbf{d}_1 &= (d_1, d_2, ..., d_{L-1}, d_L, 0, ..., 0)' \\
\mathbf{d}_2 &= (0, d_1, d_2, ..., d_{L-1}, d_L, 0, ..., 0)' \\
\mathbf{d}_3 &= (0, 0, d_1, d_2, ..., d_{L-1}, d_L, 0, ..., 0)' \\
&\vdots
\end{aligned} \tag{7.3}$$

This specification provides a general framework within which multiple break types can be analysed – Table 7.1 provides a non-exhaustive overview.[3] The form of the break function $d(t)$ has to be designed

Table 7.1: Break Function Specifications

	Break value: $d(t)$	**Length:**
Deterministic Breaks		
General Case	$d(t)$	L
Impulses (IIS)	1	1
Step Shifts (SIS)	1	$T-t+1$
Broken Trends	t	$T-t+1$
Volcanic Functions	see eq. (7.31)	3
Random Variables		
Coeff. on z_t (MIS)	$z_t \cdot d_{t,SIS}$	$T-t+1$

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \dots & \dots & \dots & 0 \\ d_2 & d_1 & 0 & \dots & \dots & \vdots \\ \vdots & d_2 & d_1 & 0 & \dots & \vdots \\ d_L & \vdots & d_2 & d_1 & 0 & \vdots \\ \vdots & d_L & \vdots & d_2 & d_1 & 0 \\ 0 & 0 & d_L & d_3 & d_2 & d_1 \end{pmatrix}$$

a-priori, but this is implicitly done in most structural break detection methods. For example, outlier detection through finding impulses (IIS, in Hendry et al. 2008) sets the break vector in $\mathbf{d}_1$ such that $d(t) = 1$ and $L = 1$, while a search for step-shifts (SIS) sets $d(t) = 1$ and the length to $T-t+1$, i.e. the break function continues until the end of the sample. Breaks in linear trends (see e.g., Perron & Yabu 2009, Perron & Zhu 2005, Estrada et al. 2013) can be constructed by setting $d(t) = t$ and the length to $T-t+1$. Pretis, Mann, & Kaufmann (2015) apply indicator saturation using broken trends and step shifts to evaluate climate models. Breaks in coefficients on random variables z_t (see e.g. Bai & Perron 2003, Ericsson 2012, Kitov & Tabor 2015) can be constructed by interacting z_t with a full set of step shifts. Sudden declines followed by a smooth recovery to the mean in hemispheric temperature responses are introduced here as volcanic functions and considered in sections 7.2.1 and 7.3. Linear combinations of multiple break functions can allow for varying lengths of breaks without pre-specification.

Searching for breaks in k variables implies that the complete break matrix across all k variables $\mathbf{D}$ is of dimension $(T \times kT)$. The inclusion of kT additional variables leads to the total number of variables N exceeding the number of observations, $N > T$, even for $k = 1$. Thus, a methodology allowing for more variables than observations is required.

Selection of models with more variables than observations has primarily relied on either shrinkage-based penalised likelihood methods (Tibshirani 1996, Zou & Hastie 2005, Tibshirani 2011) or general-

[3] While the framework presented here provides an encompassing specification for many break types, the construction of $\mathbf{D}$ is not limited to this particular case. Additional sets of specifications for step shifts are considered in Castle, Doornik, Hendry, & Pretis (2015). The appeal of the specification here is that the definition of $\mathbf{D}$ allows for a general framework under which properties can be analyzed where many of the previously proposed cases are a special case of $\mathbf{D}$.

to-specific methodology in the econometrics literature (see e.g. Castle et al. 2011). Kock & Teräsvirta (2015) compare the general-to-specific model selection algorithms *Autometrics*, to *QuickNet* – an artificial neural-network method proposed by H. White (2006), and to a shrinkage-based bridge estimator from Huang et al. (2008) in the context of forecasting.

Here we rely on general-to-specific model selection due to methods based on forward-step-wise searches not performing as well in break detection contexts (see section 7.2.1 for a simple comparison, or Epprecht et al. n.d., and Hendry & Doornik 2014 for comparisons on general variable selection.). Cox and Snell (1974) discuss some of the challenges of the general variable selection problem and Hoover & Perez (1999) show the feasibility of general-to-specific model selection for $N << T$. When facing more variables than observations, the general-to-specific approach is closely linked to robust statistics. Saturating a model with a full set of $0/1$ indicator functions from which selections are made is equivalent to a robust one-step M-estimator using Huber's skip function (see Johansen & Nielsen 2009, Johansen & Nielsen 2013 for the iterated case, and Johansen & Nielsen 2016 for an overview). Here we generalize this allowing for any form of designed break function in place of impulses, and formulate break detection as a model selection problem.

To estimate model (7.2) saturated with a full set of break functions $\mathbf{D}$ (so $N > T$), we rely on a block-partitioning estimation procedure (Doornik 2010, Hendry & Johansen 2015). For this, we partition $\mathbf{D}$ into b blocks of n_i variables such that $n_i << T$ and $\sum_{i=1}^{b} n_i = N$. In the simplest case of testing for a break in a single variable (e.g. the intercept), a split-half approach (see Figure 7.1 and *Algorithm 1* in supplementary material 7.5.4) is feasible: initially we include the first half of $\mathbf{D}_1 = (\mathbf{d}_1, ..., \mathbf{d}_{T/2})$ and retain only significant break indicators. We repeat the step for the second half of break functions d_j (for $j = T/2+1, ...T$) and finally combine the retained sets and only keep significant indicators. This split-half approach is considered here for analytical tractability in section 7.2.1.

In practice, however, we rely on a multi-split and multi-path search to lower the variance of the estimators, allow for any number of variables for a given set of observations, and to avoid a breakdown of the procedure if the breaks cannot be adequately modelled through split-half indicators.[4] This can be implemented through the general-to-specific model selection algorithm *Autometrics* (*Algorithm 2* in supplementary material 7.5.4), described in Doornik 2009a, or the *gets* package in the statistical software environment *R* (Pretis, Sucarrat, & Reade 2015). The algorithm (referred to as multi-path throughout the paper) avoids path dependence through a tree-structure and uses a parallel step-wise backwards search, while testing for encompassing and congruence (Hendry 1995). See Hendry & Pretis 2013 for an application of the algorithm to econometric modelling of climate change. A simulation-based comparison to shrinkage methods is provided in section 7.2.1.

7.2.1 Properties of Designed Break Functions in the Presence of Breaks

To assess the theoretical power of the proposed methodology, we first investigate the properties in the benchmark case of a single break matched by a correctly timed break indicator in section 7.2.1. Section 7.2.1 then assesses the properties of break indicator saturation when the break date and magnitude are unknown. Section 7.2.1 investigates uncertainty around the break date and 7.2.2 describes the properties in the presence of no breaks. Theory results are derived for general designed functions, simulation examples are based on a volcanic break as characterised by equation (7.31) below.

[4] In a simple split-half analysis there may be an identification problem if the sample-split coincides perfectly with a structural break. This is overcome by varying the block partitioning as is done in the software implementations of the algorithm.

Power for Known Break Date

We investigate the theoretical power of detecting a break in a time series given a known break date. Consider a data generating process (DGP) coinciding with the model for a single known break in an intercept:

$$y_t = \mu + \lambda d_t + \varepsilon_t \tag{7.4}$$

where $\varepsilon_t \sim \mathrm{IN}(0, \sigma_\varepsilon^2)$. The break shifts μ to $\mu + \lambda d_t$ where d_t is a break function of length L beginning at time $t = T_1$ where $(T_1 + L) \leq T$ such that $d_t \neq 0$ for $T_1 \leq t < (T_1 + L)$ and 0 otherwise. The estimators $\widehat{\mu}$ and $\widehat{\gamma}$ (where $\widehat{\gamma}$ is the estimator for λ) in a correctly specified model for a known break are given by:

$$\begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma} - \lambda \end{pmatrix} = \begin{pmatrix} T_d^{-1} \left(\sum_{t=T_1}^{T_1+L-1} d_t^2 \sum_{t=1}^{T} \varepsilon_t - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=T_1}^{T_1+L-1} d_t \varepsilon_t \right) \\ T_d^{-1} \left(\sum_{t=T1}^{T_1+L-1} d_t \varepsilon_t - \sum_{t=T1}^{T_1+L-1} d_t \sum_{t=0}^{T} \varepsilon_t \right) \end{pmatrix} \tag{7.5}$$

where $T_d = T \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \frac{1}{T} \left(\sum_{t=T_1}^{T_1+L-1} d_t \right)^2 \right]$. The estimators are unbiased for the break and intercept: $E[\widehat{\mu} - \mu] = 0$ and $E[\widehat{\gamma} - \lambda] = 0$. The variance of the estimators is given by:

$$V \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma} - \lambda \end{pmatrix} = \sigma_\varepsilon^2 T_d^{-1} \begin{pmatrix} \sum_{t=T_1}^{T_1+L-1} d_t^2 & -\sum_{t=T_1}^{T_1+L-1} d_t \\ -\sum_{t=T_1}^{T_1+L-1} d_t & T \end{pmatrix} \tag{7.6}$$

The distribution of the break estimator is then:

$$(\widehat{\gamma} - \lambda) \sim N \left(0, \sigma_\varepsilon^2 \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \sum_{t=T_1}^{T_1+L-1} d_t \bar{d} \right]^{-1} \right) \tag{7.7}$$

where $\bar{d} = 1/T \sum_{t=1}^{T} d_t$. For the special case when step-indicators are chosen as the functional form of d_t and the single break lasts from $t = 0$ to $t = T_1 < T$, equation (7.5) simplifies to $\widehat{\mu} - \mu = \bar{\varepsilon}_2$ and $\widehat{\gamma} - \lambda = \bar{\varepsilon}_1 - \bar{\varepsilon}_2$, where $\bar{\varepsilon}_1 = 1/T_1 \sum_{t=1}^{T_1} \varepsilon_t$ and $\bar{\varepsilon}_2 = 1/(T - T_1) \sum_{t=T_1+1}^{T} \varepsilon_t$.[5]

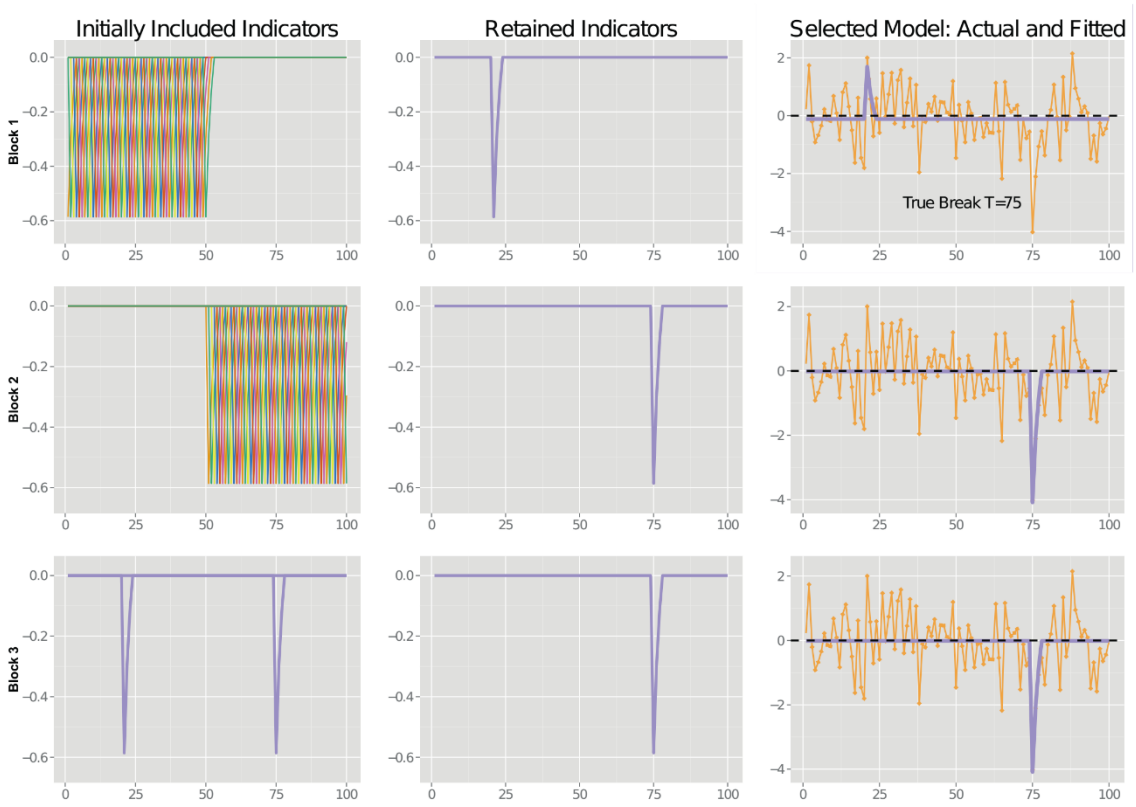
Potency for an Unknown Break Date

When the break date is unknown, we propose to saturate the regression model using a full set of specified break indicators and select significant breaks through an extended general-to-specific algorithm (Castle et al. 2011). We assess the methodology in the selection context using two main concepts: the null retention frequency of indicators is called the gauge, comparable to the size of a test denoting its (false) null rejection frequency, but taking into account that indicators that are insignificant on a pre-assigned criterion may nevertheless be retained to offset what would otherwise be a significant misspecification test (see Johansen & Nielsen 2016 for distributional results on the gauge). The non-null retention frequency when selecting indicators is called its potency, comparable to a similar tests power for rejecting a false null hypothesis.

Here we investigate the feasibility of the proposed method by deriving the analytical properties of the split-half approach for an unknown break. Figure 7.1 illustrates the split-half method for a single unknown break. In practice we rely on a multi-path, multi-block search algorithm (such as *Autometrics*, see *Algorithm 2* in section 7.5.4) to reduce the variance of the estimators.

[5] See supplementary material 7.5.3 for proof.

Figure 7.1: Split-half approach for a single unknown break of the shape of a volcanic function at T=75. Left column shows included indicators in each step, middle column shows the retained indicators, right column graphs the selected model with actual and fitted data. Block 1 (top panel) includes the first half of break functions and retains a single one as the mean is lowered in the second half due to the presence of a break at t=75. Block 2 (middle panel) then includes the second half retaining the correct break function. Block 3 uses the union of retained indicators from blocks 1 and 2 in which now the first indicator is rendered insignificant by the mean being correctly estimated due to the second indicator capturing the break. Using a saturating set of break functions at 1% the break at T=75 is detected without prior knowledge and is the only break function retained.



Consider a single break falling into the first half of the sample beginning at time T_1 for L periods such that $0 < T_1 < T_1 + L < T/2$. In matrix form the DGP is given as:

$$\mathbf{y} = \lambda \mathbf{d}_{T_1} + \boldsymbol{\varepsilon} \tag{7.8}$$

where $\boldsymbol{\varepsilon} \sim \mathrm{IN}(\mathbf{0}, \sigma^2)$ for simplicity and the $(T \times 1)$ vector $\mathbf{d}_{T_1}$ denotes a break at $t = T_1$ for L periods. Using a split-half approach, we assess the properties of detecting the single break when the break date is unknown. The split-half model for the first half of break functions is:

$$\mathbf{y} = \mathbf{D}_1 \gamma_{(1)} + \mathbf{v} \tag{7.9}$$

where $\gamma_{(1)} = (\gamma_1, \gamma_2, ...\gamma_{T/2})'$ and $\mathbf{D}_1 = (\mathbf{d}_1, ..., \mathbf{d}_{T/2})$. The estimator $\widehat{\gamma}_{(1)}$ equals:[6]

$$\begin{aligned} \widehat{\gamma}_{(1)} &= \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\mathbf{y} = \lambda \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\mathbf{d}_{T_1} + \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\boldsymbol{\varepsilon} \\ &= \lambda \mathbf{r} + \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\boldsymbol{\varepsilon} \end{aligned} \tag{7.10}$$

where the $(T/2 \times 1)$ vector $\mathbf{r}$ is equal to one at $t = T_1$ and zero otherwise, $r_t = 1_{\{t=T_1\}}$. It follows that $E[\widehat{\gamma}_{(1)}] = \lambda \mathbf{r}$ and $V[\widehat{\gamma}_{(1)}] = \sigma_\varepsilon^2 \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}$. We find for the first half, for normal error terms:

$$\left(\widehat{\gamma}_{(1)} - \lambda \mathbf{r}\right) \sim N\left(\mathbf{0}, \sigma_\varepsilon^2 \left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\right) \tag{7.11}$$

Therefore conventional t-tests can be used to assess the significance of individual indicators. The estimator $\widehat{\gamma}_{(2)}$ on the second half of indicators, $\mathbf{D}_2 = (\mathbf{d}_{T/2+1}, ..., \mathbf{d}_T)$, will miss the break in the DGP in the first half described by $\mathbf{d}_{T_1}$ and equals:

$$\widehat{\gamma}_{(2)} = \lambda \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1}\mathbf{D}_2'\mathbf{d}_{T_1} + \left(\mathbf{D}_2'\mathbf{D}_2\right)^{-1}\mathbf{D}_2'\boldsymbol{\varepsilon} \tag{7.12}$$

For step-shifts, Castle, Doornik, Hendry, & Pretis (2015) show that the indicator in $\mathbf{D}_2$ closest to the sample split will be retained in the second set of indicators. For the general form of break functions, retention in $\mathbf{D}_2$, when there is a break in the first half, will depend on the specific functional form. However, conditional on the break indicator being correctly retained in the first set $\mathbf{D}_1$, retention of irrelevant indicators in $\mathbf{D}_2$ does not affect the correct identification of the break overall: let $\mathbf{D}_{1*}$ and $\mathbf{D}_{2*}$ denote the set of retained break functions in the first and second set respectively, where retention is based on a retention rule such as $\mathbf{d}_j$ is retained if $|t_{\widehat{\gamma}_j}| \geq c_\alpha$. The final step in the split-half procedure is then to combine the retained indicators using $\mathbf{D}_U = [\mathbf{D}_{1*}\mathbf{D}_{2*}]$ and estimate the model:

$$\mathbf{y} = \mathbf{D}_U \gamma_{(U)} + \mathbf{v} \tag{7.13}$$

This yields the estimator $\widehat{\gamma}_{(U)}$ unbiased for the true break:[7]

$$\widehat{\gamma}_{(U)} = \lambda \mathbf{r} + \left(\mathbf{D}_U'\mathbf{D}_U\right)^{-1}\mathbf{D}_U'\boldsymbol{\varepsilon} \tag{7.14}$$

[6] Proof given in supplementary material 7.5.2.

[7] In practice selection bias can be controlled using bias-correction after orthogonalization of the selected regressors – see Hendry & Krolzig (2005) for the orthogonal case, Pretis (2015b) for bias-correction of step-functions, and Castle, Doornik, & Hendry (2015) for bias-correction with correlated variables.

The carried-forward break function in $\mathbf{D}_{1*}$ correctly identifies the true break, and coefficients on all other break functions will thus be zero in expectation. The proof is identical to that given for the first half of indicators in the supplementary material 7.5.2. This shows that, conditional on retaining the correct break indicator in $\mathbf{D}_1$, the retention of indicators in $\mathbf{D}_2$ does not affect the correct identification of the break, when the first and second set are combined and re-selected over. The distribution of the final split-half estimator is then given by:

$$\left(\widehat{\gamma}_{(U)} - \lambda \mathbf{r}\right) \sim N\left(\mathbf{0}, \sigma_\varepsilon^2 \left(\mathbf{D}_U' \mathbf{D}_U\right)^{-1}\right) \tag{7.15}$$

Re-selection then results in only the true break indicator being retained in expectation.[8]

This result generalizes the specific case of step indicators presented in Castle, Doornik, Hendry, & Pretis (2015). Even though the break date and magnitude are unknown, the use of a fully saturated set of break indicators allows us to obtain an unbiased estimate of the break magnitude and timing. The estimator then follows a normal distribution subject to correct specification of the break function. Thus the estimated coefficient at the break time, $\widehat{\gamma}_{T_1}$, is in expectation equal to the break magnitude, while all other estimated coefficients are mean-zero in expectation. This result generalizes to multiple breaks falling in a single split. As in the case of the known break timing, the variance of the estimator depends on the specified break function. Let $\delta_{k,j}$ denote the (k,j) element of the matrix $\left(\mathbf{D}_1' \mathbf{D}_1\right)^{-1}$. The variance of the coefficient at the breakpoint in the first half is therefore:

$$V[\widehat{\gamma}_{T_1}] = \sigma_\varepsilon^2 \delta_{T_1,T_1} \tag{7.16}$$

For iid error terms ε, and $\mathbf{D}$ specified as a full set of step functions, the split-half model (without selection) yields $\delta_{j,j} = 2$, so the break coefficient has twice the error variance. For the proposed volcanic function (derived and assessed in detail in section 7.3) modelling a single drop followed by a reversion to the mean, we find that $\delta_{j,j} = 3.7$, thus $V[\widehat{\gamma}_{T_1}] = 3.7\sigma_\varepsilon^2$. This can be compared to the known-break/single-indicator case where the variance is given by equation (7.6) and for the volcanic function equals $2.3\sigma_\varepsilon^2$ (for T=100). Due to collinearity of break functions, the variance of the estimator is higher in a fully saturated model. In the more general case, δ_{T_1,T_1} depends on the specification of the break function but can be computed *a-priori*. The t-statistic is then given as:

$$t_{\widehat{\gamma}_{T_1}} = \frac{\widehat{\gamma}_{T_1}}{\widehat{\sigma}_\varepsilon \sqrt{\delta_{T_1,T_1}}} \approx \frac{(\widehat{\gamma}_{T_1} - \lambda)}{\sigma_\varepsilon \sqrt{\delta_{T_1,T_1}}} + \frac{\lambda}{\sigma_\varepsilon \sqrt{\delta_{T_1,T_1}}} \sim N\left(\frac{\lambda}{\sigma_\varepsilon \sqrt{\delta_{T_1,T_1}}}, 1\right) \tag{7.17}$$

In practice we use sequential elimination of the break indicators or a multi-path search to eliminate insignificant indicators reducing the variance of the estimators from a saturated model (7.16) closer to the single break (7.6) and increasing the power of detection.

For dynamic time series models, the above approach can be extended by including time-dependent co-variates. Valid conditioning (e.g. through the inclusion of auto-regressive terms in the case of non-iid errors) can be ensured by always including the covariates in each block estimation step and only selecting over the break functions. Johansen & Nielsen (2009) provide the asymptotics under the null of no break

[8] The split-half approach is not the only way of analysing the theory of indicator saturation: rather than splitting the functions into a first and second half, alternatively one could consider including every other break function in two sets such that $\mathbf{D}_1$ covers breaks at $t = 1,3,5...$ and $\mathbf{D}_2$ covers breaks at $t = 2,4,6....$ Retention frequencies in this setup can be derived using the results in section 7.2.1.

for the special case of impulses for stationary and unit-root non-stationary autoregressive processes (see Johansen & Nielsen 2013, for the iterated version). The case for general break functions is discussed in section 7.2.2, and the supplementary material 7.5.5 provides simulation results for an AR(1) model and DGP.[9]

Simulation Performance based on Volcanic break functions: Table 7.2 reports simulation results (T=100) for a DGP with a single unknown volcanic break at $t = T_1 = 25$ of magnitude λ followed by a smooth reversion to the mean.[10] Equation (7.31) provides the exact functional form. Simulations are assessed by the retention/detection frequency (potency) for a single break and average retention of spurious breaks (gauge).[11]

The trade-off between potency and level of significance of selection α is shown in Figure 7.2 for a single volcanic break. A multi-path search generally increases the power of detection relative to the split-half approach. Figure 7.3 shows the results for split-half (dashed) and multi-path (solid) selection when using volcanic functions for a break of $\lambda = 6$. Consistent with derived theory (7.16), the estimator has 3.7 times the variance of the error term when using split-half estimation for the given function. Using a multi-path search reduces the variance drastically. Any selection bias of the multi-path search estimates can be controlled through bias correction after selection (see Castle et al. 2011 and Pretis 2015b). The supplementary material 7.5.5 provides simulation results for a simple autoregressive DGP and model.

Table 7.2: Potency of detecting an unknown break when using split-half and multi-path searches. Statistics were generated from 1000 simulations and detection significance was set to $\alpha = 0.01$, with a length of $L = 3$. Break magnitude λ corresponds to the full response in standard deviations of the error term ($\sigma_\varepsilon = 1$) over the entire break, the trough is 0.58λ.

	Split-Half		Multi-Path	
	Potency	Gauge $\mathbf{D}_1$	Potency	Gauge
$\lambda = 6$, trough=3.48	0.69	0.013	0.88	0.015
$\lambda = 4$, trough=2.23	0.30	0.013	0.50	0.014
$\lambda = 2$, trough=1.16	0.06	0.013	0.11	0.015

Comparison to Shrinkage-based Methods Shrinkage-based methods using penalised likelihood estimation (Tibshirani 2011, Zou & Hastie 2005) provide an alternative to the general-to-specific algorithm used here in selecting models with more variables than observations. Figure 7.4 shows the simulation outcomes comparing multi-path indicator saturation (for $\alpha = 0.01$), the Lasso (Tibshirani 1996, estimated using LARS, see Efron et al. 2004) where cross-validation is used to determine the penalty, and

[9] While our analysis concentrates on small-sample properties, the asymptotic rates of convergence will generally depend on the specification of the break function – varying scaling to obtain non-degenerate limit distributions may therefore be required. In the case of step functions ($d_t = 1, L = T$) and the simple no-intercept case, pre-multiplying the estimator by $\sqrt{T}$ yields asymptotic normality for the break estimator when $T^{-1}\sum_{t=T_1}^{T_1+L-1} d_t^2 = T^{-1}L \to \tau$ as $T \to \infty$. In other words, the ratio of break length to the sample size remains constant as the sample size increases – this can be interpreted as obtaining more information on the break period or sampling at higher frequencies as $T \to \infty$. A similar analysis can be applied to the volcanic functions considered here, where either the break length scales with the sample size, or alternatively the magnitude increases similar to the asymptotic analysis for a single impulse in Doornik et al. (1998).

[10] For a volcanic break, λ denotes the entire temperature response over the specified length L, thus the trough will be less than λ. For the present specification of $L = 3$, the initial trough of the function equals 0.58λ.

[11] All simulations and applications using the multi-path search *Autometrics* are coded using the *Ox* programming language (Doornik 2009b). Simulations using the Lasso are coded using the package `glmnet` (Friedman et al. 2010) in *R*.

Figure 7.2: Left: Potency of detecting a volcanic break of magnitude λ for level of significance α using split-half and multi-path selection. Right: Proportion of spuriously retained break indicators (gauge). Break magnitude λ corresponds to the full response in standard deviations of the error term ($\sigma_\varepsilon = 1$) over the entire break, the peak is 0.58λ.

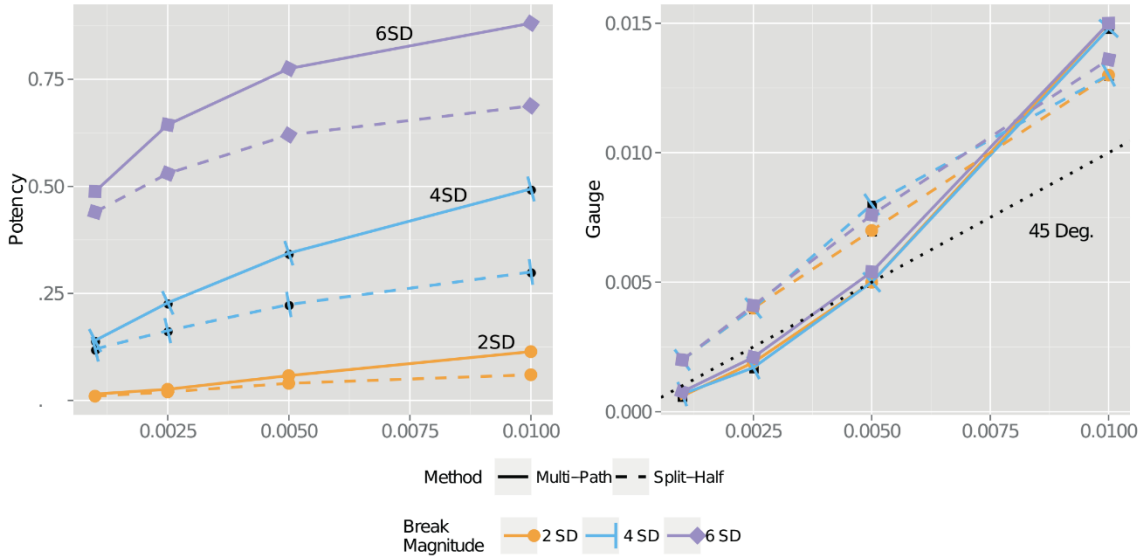


Figure 7.3: Estimated break indicator and variance for unknown break using split-half (blue) and multi-path (yellow) selection. The left panel shows a simulated time series with the true break shown in blue and the fit in green. The middle panel shows the distribution of the estimated coefficient, the right panel shows the variance of the coefficient. Dashed lines show the true break magnitude and analytical variance of the split-half coefficient.

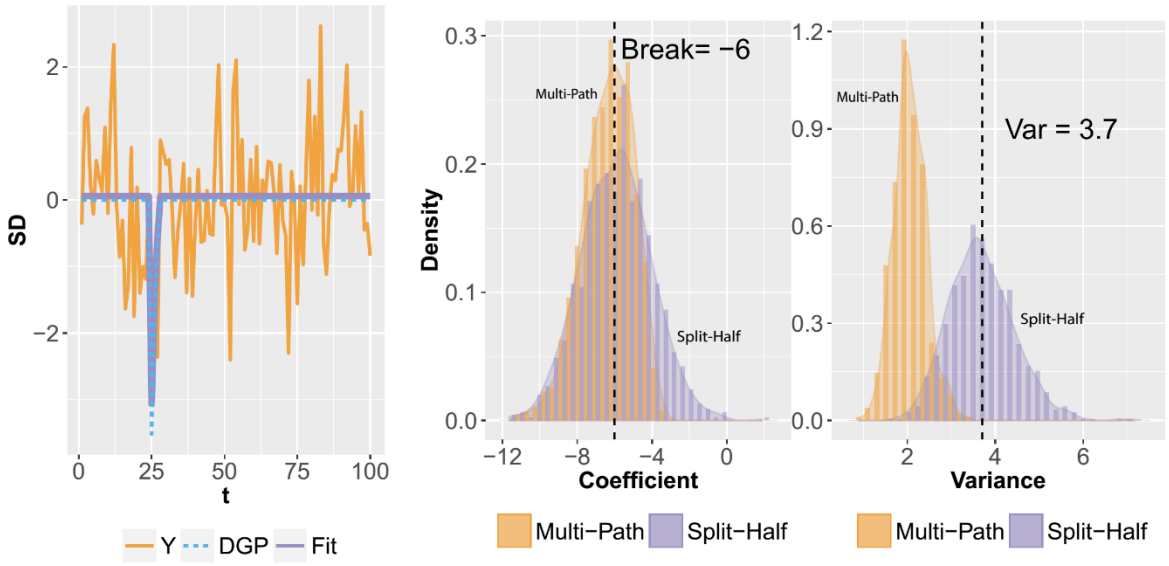
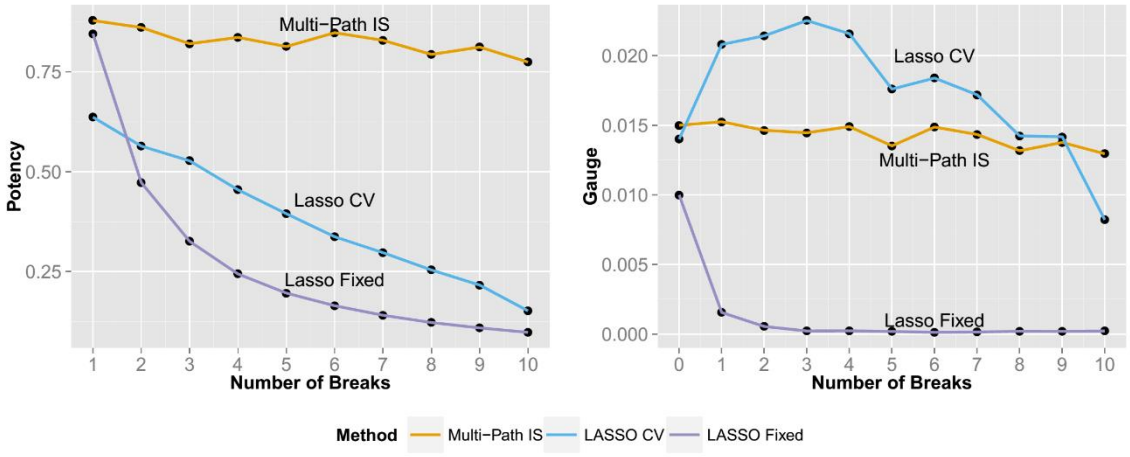


Figure 7.4: Left: Average potency of detecting increasing numbers of volcanic breaks using multi-path indicator saturation at $\alpha = 0.01$ (IS, orange), cross-validated Lasso (blue), and Lasso with fixed penalisation parameter (green) where the penalisation parameter is set such that the false-positive rate approximates that of the indicator saturation procedure under the null of no break. Right: Corresponding false-positive rate (gauge). $M = 1000$ replications.



the Lasso where the penalty is set such to approximate the false-positive rate of the IS procedure under the null of no breaks (≈ 0.01). The simulation uses a total break magnitude of 6 standard deviations (implying a trough of $3.48\sigma_\varepsilon$) for an increasing number of evenly spaced breaks from 0 up to 10 in a sample of $T = 100$. The general-to-specific multi-path algorithm exhibits stable power exceeding that of the penalised likelihood methods across any number of breaks. The false-positive rate remains stable and close to the theory level of 0.01. The shrinkage-based procedures, due to their similarity to forward-selection, show decreasing potency as the number of breaks increases, and the false-positive rate is difficult to control.

Uncertainty on the break date

An estimated uncertainty on the break magnitude and coefficient path (the time-varying intercept in the regression) can be computed given the distribution of the break estimator (see Pretis 2015b). While of considerable interest, it is non-trivial, however, to quantify the uncertainty around the timing of the break (see Elliott & Müller 2007). This is particularly true for the literature focusing on break detection using general to specific methodology. Here we investigate the uncertainty around the timing of estimated break points when using break indicator saturation by computing the analytical power of a single break indicator when the break function is correctly specified but the break time is not. This is a simplification as it only considers a single mis-timed indicator, while the indicator saturation approach includes a saturating set.

Consider a DGP with just a single break in the mean:

$$y_t = \lambda d_{T_1,t} + \varepsilon_t \tag{7.18}$$

The break shifts $E[y_t]$ from 0 to $\lambda d_{T,1}$ at $t = T_1$ where d_{T_1} is a break function of length L beginning at

time $t = T_1$ such that $T_1 + L < T$ and $d_{T_1} = (0, ..., d_1, d_2, ..., d_L, 0, ..., 0)$. The corresponding model is then:

$$y_t = \gamma d_{j,t} + v_t \tag{7.19}$$

When the break date is correctly specified $d_{j,t} = d_{T_1,t}$, so the estimator for λ is given by:

$$\widehat{\gamma}_{t=T_1} - \lambda = \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2 \right)^{-1} \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t} \varepsilon_t \right) \tag{7.20}$$

Similarly for a test of the hypothesis: $\lambda = 0$, the t-statistic has a non centrality of $E\left[t_{\widehat{\gamma},t=T_1}\right] = \psi = \frac{\lambda \sqrt{\left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)}}{\sigma_\varepsilon}$ and the normal distribution:

$$t_{\widehat{\gamma}} \approx \frac{\widehat{\gamma}_{t=T_1} \sqrt{\left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)}}{\sigma_\varepsilon} \sim N(\psi, 1) \tag{7.21}$$

The non-centrality ψ increases in the break magnitude λ, varies with the break length L, and will depend on the underlying break function given by d_t.

Now consider the model being incorrectly specified for the break date, such that $d_{j,t} \neq d_{T_1,t}$ but is shifted by K periods $d_{j,t} = d_{T_1 \pm K,t}$. The estimator for λ is then:

$$\widehat{\gamma}_{=T_1 \pm K} - \lambda = \lambda \left[\left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2 \right)^{-1} \left(\sum_{t=T_1}^{T_1+L} d_{j,t} d_{T_1,t} \right) - 1 \right] + \left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2 \right)^{-1} \left(\sum_{t=T_1}^{T_1+L} d_{j,t} \varepsilon_t \right) \tag{7.22}$$

For a fixed length L and a forced mis-timing, it follows that $\widehat{\gamma}_{t \neq T_1}$ is not an unbiased estimator for λ. Note that if d_j is functionally specified correctly such that the only difference to the true break function is through K lags, $\mathbf{d}_j = \mathbf{d}_{T_1 \pm K}$, then it holds that $\left(\sum_{t=T_1}^{T_1+L} d_{j,t}^2\right) = \left(\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2\right)$. Equally $\left(\sum_{t=T_1}^{T_1+L} d_{j,t} d_{T_1,t}\right) = \left(\sum_{t=T_1}^{T_1+L} d_{T_1 \pm K,t} d_{T_1,t}\right)$ for $K \leq L$ and 0 for $K > L$. Using this, we derive an expression for the approximate t-statistic associated with the estimator given a break function time mis-specified by K lags:

$$E\left[t_{\widehat{\gamma},t=T_1 \pm K}\right] \approx \frac{E\left[\widehat{\gamma}_{=T_1 \pm K}\right]}{\sigma_\varepsilon \left(\sum_{t=T_1}^{T_1+L} d_{T_1}^2\right)^{-1/2}} = \frac{\lambda_1 \left(\sum_{t=T_1}^{T_1+L} d_{T_1 \pm K} d_{T_1}\right)}{\sigma_\varepsilon \left(\sum_{t=T_1}^{T_1+L} d_{T_1}^2\right)^{1/2}} \tag{7.23}$$

This is equal to the non-centrality of the correct break date ψ scaled by a factor less than one, decreasing with the distance K from the correct date:

$$E\left[t_{\widehat{\gamma},t=T_1 \pm K}\right] \approx \psi \left(\frac{\sum_{t=T_1}^{T_1+L} d_{T_1 \pm K,t} d_{T_1,t}}{\sum_{t=T_1}^{T_1+L} d_{T_1,t}^2} \right) \leq \psi \tag{7.24}$$

For a given break specification d_t and break length L, the corresponding power function can be computed to provide an approximate measure of power for detection of a break at $t = T_1$ in the neighbourhood of T_1. Note that $E\left[t_{\widehat{\gamma},t=T_1 \pm K}\right]$ is zero outside a neighbourhood of L. The associated t-statistic of a break indicator further away from the true break date T_1 than the break length L is zero in expectation, since $\left(\sum_{t=T_1}^{T_1+L} d_{j,t} d_{T_1,t}\right) = 0$ for $K > L$. Intuitively, longer breaks increase the likelihood that a break indicator that is not perfectly coincident with the break date will appear significant, and we can expect the retention

to be equal to the nominal significance level outside a $t = T_1 \pm L$ interval.

As before we consider the special case of volcanic functions and also provide results from step shifts for comparison. Figure 7.5 shows the analytical as well as simulated non-centrality and power around a true break date at t=26 of length $L = 3$ for $\alpha = 0.05$. The Monte Carlo simulations match the theoretical powers and non-centralities closely.

For no break, the analytical power is uniform and equal to the nominal significance level. When there is a break outside of the interval $T_1 \pm L$ the expected retention of the break indicator equals the nominal significance level. For a step shift of a forced length, given (7.24), the non-centrality decreases linearly as the numerator falls by $1/L$ per shifted period relative to the correct break date. For longer breaks this implies that the power around the true break date is close to uniform. In the case of volcanic functions, due to the particular functional form, the power and retention probability drop more rapidly and peak clearly around the true break date. The special case presented here only considers the properties of a single time-mis-specified indicator of a fixed length in the model. However, model selection in the indicator saturation approach alleviates many of these concerns in practice. When selecting from a full set of break functions (see section 7.2.1) it is less likely that a break function at $T_{1\pm K}$ appears significant because the correct T_1 indicator is included in the same model, a mis-timed indicator in a fully saturated model would likely appear significant only if a chance draw of the error offsets the shift.

7.2.2 Properties under the Null of No Break

Under the null hypothesis when there are no breaks in the DGP, there are two primary concerns regarding the inclusion of a full set of break functions in the statistical model. First, when including a full set of break functions, break indicators may be retained spuriously, and second, there may be concerns about the effect on the distributions of coefficients on variables that are known to be relevant – in other words, does saturating a model with irrelevant variables affects relevant ones?

First, we consider the spurious retention of break indicators. Under the null of no breaks, $\lambda = 0$, the DGP from (7.8) is given by:

$$\mathbf{y} = \boldsymbol{\varepsilon} \tag{7.25}$$

Based on the above results, when using a split-half approach with a full set of break indicators, the expectation of the estimated coefficients in the first half is given by:

$$E[\widehat{\gamma}_{(1)}] = E\left[\left(\mathbf{D}_1'\mathbf{D}_1\right)^{-1}\mathbf{D}_1'\boldsymbol{\varepsilon}\right] = \mathbf{0} \tag{7.26}$$

The same result generalizes to the union of retained indicators $\mathbf{D}_U$. Thus, the t-statistics of the included break functions will be centred around zero in expectation when there is no break. Using the selection rule that retains the break function $\mathbf{d}_j$ if $|t_{d_j}| > c_\alpha$, then $\alpha T/2$ indicators will be retained on average in each half. Combining the retained indicators in the final set, αT indicators are retained in expectation. The proportion of spurious indicators can thus be controlled through the nominal significance level of selection. The properties under the null are confirmed below using Monte-Carlo simulations.

Table 7.3 and Figure 7.6 report the simulation results when there are no breaks in the DGP but a full set of break functions (of the form of volcanic functions) is included. When using a split-half approach with a 1-cut variable selection decision based on the absolute t-statistic, the proportion of irrelevant retained indicators is close to the nominal significance level. In practice, when using a multi-path, multi-split pro-

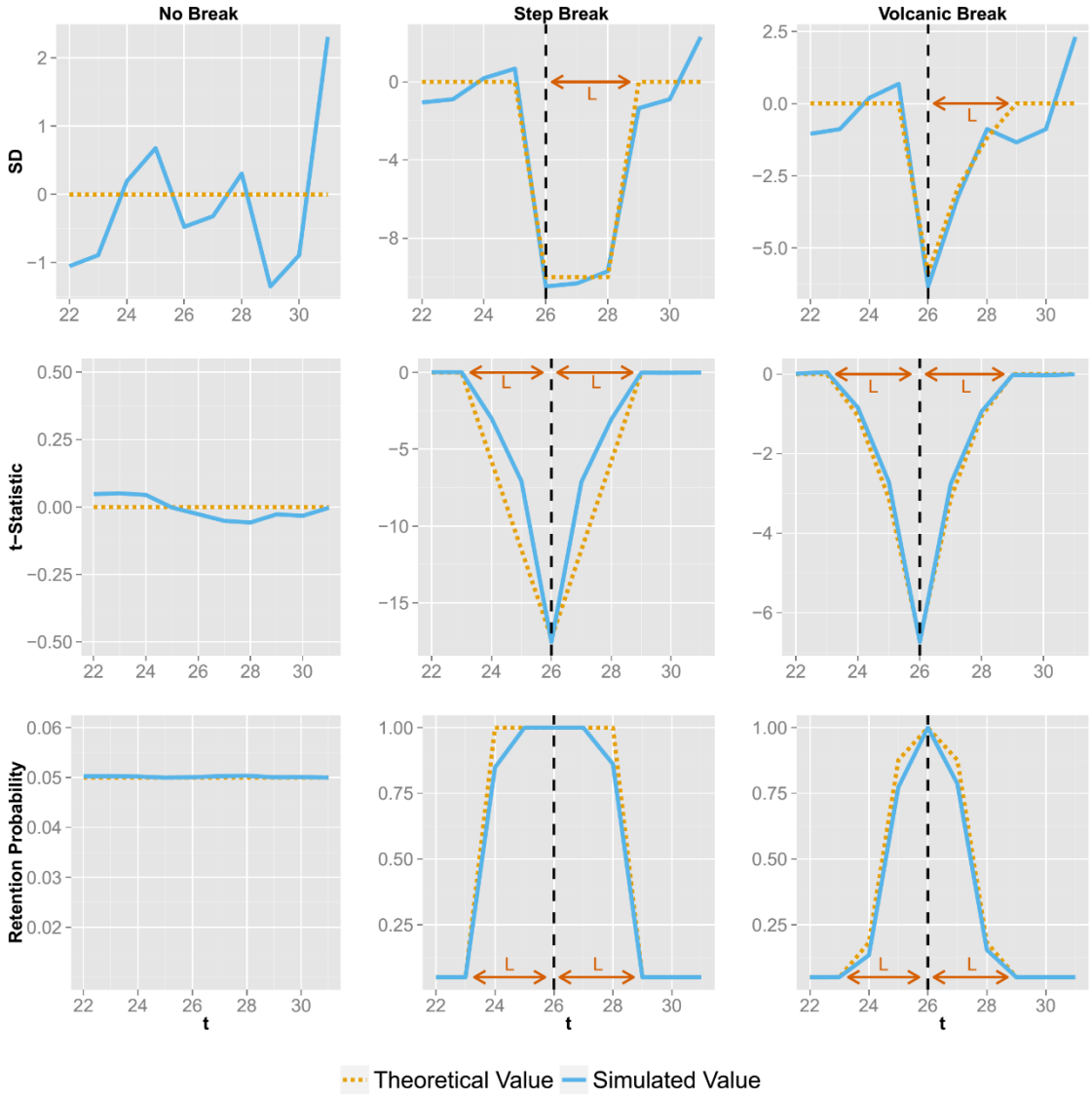
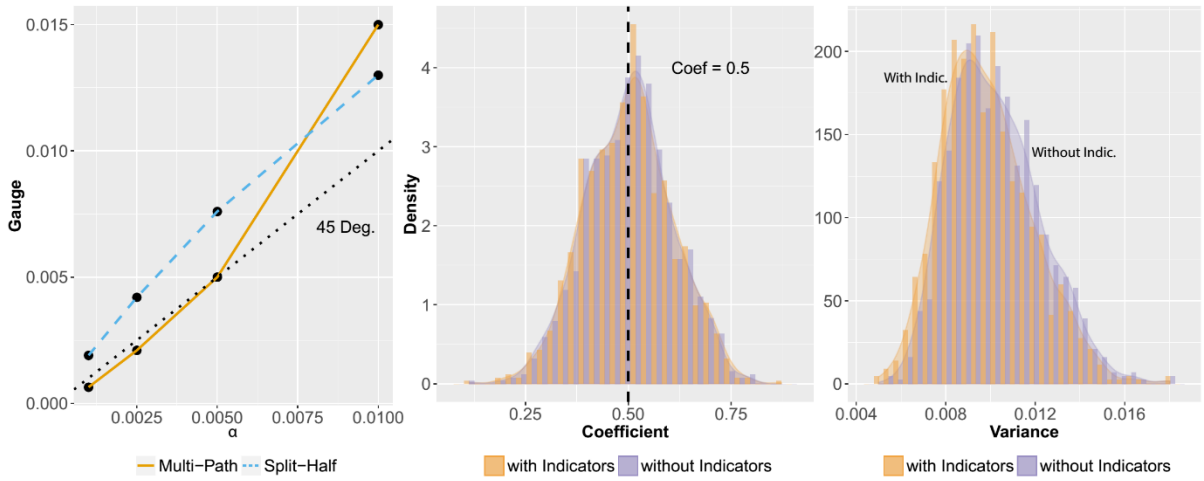


Figure 7.5: Power and retention frequency around the break date where the timing of the break functions is imposed without selection: Simulated data with and without shifts (top), associated non-centrality and simulated t-statistics (middle), analytical and simulated power (bottom) around break $\lambda = -10$ at $T = 26$ of length $L = 3$ and interval $T \pm K$ for $\alpha = 0.05$. Left shows no break, middle a step-break, and right panel a volcanic function break. Analytical non-centralities and power are shown in yellow, simulated t-statistics and retention are shown in blue. Dashed lines mark the break occurrence. Outside of an interval $T = 26 \pm L$ the retention probability and analytical power are equal to the nominal significance level $= \alpha = 0.05$.

Figure 7.6: Simulation results under the null of no break. Left: Proportion of irrelevant retained break functions (gauge) using split-half and multi-path selection for varying α when there is no break. Middle and Right: Simulated distributions and densities of coefficient $\widehat{\beta}$ (true $\beta = 0.5$) on forced parameter x_t: with (yellow) – and without (blue) – a full set of break functions.



cedure (here implemented through *Autometrics*) the gauge is close to the nominal significance level for low levels of α. A conservative approach (low $\alpha \leq 1\%$) is recommended in practice.[12] When compared to results in Castle, Doornik, Hendry, & Pretis 2015, there is little notable difference between different specifications of break functions, consistent with the analytical properties of irrelevant indicators.

Table 7.3: Retention of spurious volcanic break functions when there is no break

Significance Level	Split-Half One-Cut		Multi-Path Search
	Gauge $\mathbf{D}_1$	Gauge $\mathbf{D}_2$	Gauge $\mathbf{D}$
$\alpha = 0.05$	0.056	0.054	0.30
$\alpha = 0.01$	0.013	0.012	0.015
$\alpha = 0.005$	0.007	0.007	0.005
$\alpha = 0.0025$	0.004	0.004	0.002
$\alpha = 0.001$	0.002	0.002	0.001

We now assess the second consideration, which is the effect of including a full set of break indicators when theory variables $\mathbf{X}$ are included in the model but are not selected over ('forced'). These could include contemporaneous covariates or autoregressive dynamic variables. For the specific case when the elements of $\mathbf{D}$ are specified to be impulse indicators, Johansen & Nielsen (2009) derive the asymptotic distribution of β in the full split-half approach in stationary and unit-root non-stationary regressions using the equivalence of IIS and one step Huber-skip M estimators. For an iterated procedure (e.g. resembling the multi-block approach in *Autometrics*) the distributional results under the null for IIS are derived in Johansen & Nielsen (2013). For the general form of designed indicator functions, we follow theory for the sub-steps of split-half estimation where N$\ll$T in each step, and appeal to simulation results for the

[12] Results of high gauge for high significance levels (e.g. $\alpha \geq 0.05$) are consistent with previous results found by Bergamelli & Urga (2013) for step-functions. Once a large number of spurious breaks is retained, it becomes more likely to keep additional spurious breaks. The results for the gauge in Table 7.3 are consistent with the distributional theory for the gauge in Johansen & Nielsen 2016.

overall algorithm. Consider a simple DGP:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{7.27}$$

where $\varepsilon \sim \text{iid}(0, \sigma_\varepsilon^2)$ and the elements of $\mathbf{X}$ (dynamic or static) are assumed to be relevant and not selected over. The model relying on the split-half approach saturated with the first half of the break functions is then:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{D}_1\gamma_{(1)} + \mathbf{v} \tag{7.28}$$

where the true $\gamma_{(1)} = \mathbf{0}$. Following Hendry & Johansen (2015), given that there is no break in the DGP, the inclusion of a full set of irrelevant additional variables $\mathbf{D}_1$ need not affect the distribution of the included relevant parameters β. Orthogonalizing $\mathbf{X}$ and $\mathbf{D_1}$ by regressing each column of $\mathbf{D}_1$ on $\mathbf{X}$ yields the estimator $\widehat{\beta}^*$ with asymptotic distribution:[13]

$$\sqrt{T}\begin{pmatrix}\widehat{\beta^*} - \beta \\ \widehat{\gamma}_{(1)} - \mathbf{0}\end{pmatrix} \xrightarrow{D} N\left[\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}, \sigma_\varepsilon^2 \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{XX}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathbf{D}_1\mathbf{D}_1|\mathbf{X}}^{-1}\end{pmatrix}\right] \tag{7.29}$$

The distribution of the parameters $\widehat{\beta^*}$ on the correct variables $\mathbf{X}$ is unaffected by the inclusion of the orthogonalized break indicators $\mathbf{D}_1$ when there is no break. The equivalent result holds when the second half of break indicators $\mathbf{D}_2$ is included and the resulting union of retained indicators from $\mathbf{D}_1$ and $\mathbf{D}_2$ given that N<T. Orthogonalization relative to shifts, however, is not necessary for estimation in practice. Figure 7.6 shows the simulated distribution of $\widehat{\beta}$ for a single x_t when a full set of break functions is included and selected at $\alpha = 0.005$ (yellow) and when break functions are not included (blue). The distribution of $\widehat{\beta}$ is unaffected by the saturation of a full set of break functions. In practice the main risk is the spurious retention of break indicators, but this can be controlled through a conservative selection mechanism (low α).

7.3 Empirical Illustration for Climate Time-Series: Detection of Volcanic Eruptions from Simulated Model Surface Air Temperature Data

Large volcanic eruptions that inject significant amounts of sulphate aerosols into the stratosphere cause short-lived (multi-year) radiative imbalances that induce surface cooling. Over the course of the last several millennia there have been numerous eruptions that have had impacts on global mean temperatures. Identifying their climatic fingerprint is an important scientific endeavour that relies critically on the robust characterization of the timing and magnitude of past volcanism. An accurate understanding of the impact of past eruptions can lead to more accurate estimates of the effect of stratospheric aerosols – to guide policy from geo-engineering to pollution controls. Records of climatically relevant events primarily rely on sulphur deposits in ice cores (see e.g., Gao et al. 2008, and Crowley & Unterman 2012). However, there remains uncertainty in the precise timing, magnitude, and climatic impact of past volcanic activity (Brohan et al. 2012, Mann et al. 2012, Anchukaitis et al. 2012, G. Schmidt et al. 2011,

[13] Where $T^{-1}\mathbf{X}'\mathbf{X} \xrightarrow{P} \Sigma_{XX}$ for stochastic $\mathbf{X}$, and $\mathbf{D}_1$ is scaled such that either the break length scales with the sample size, or alternatively the break magnitude increases such that $T^{-1}\mathbf{D}_1'\mathbf{D}_1 \to \Sigma_{\mathbf{D}_1}$ is constant, and for stochastic $\mathbf{X}$ it holds that $T^{-1}(\hat{\mathbf{u}}'\hat{\mathbf{u}}) \xrightarrow{P} \mathbf{\Sigma}_{\mathbf{D}_1\mathbf{D}_1|\mathbf{X}}$ where $\hat{\mathbf{u}} = \mathbf{D}_1 - \mathbf{X}\hat{\Gamma}$ from orthogonalization regressions. See supplementary material 7.5.6 for proof based on Hendry & Johansen (2015).

Baillie & McAneney 2015). Statistical methods such as the break detection methodology presented herein can therefore augment previous volcanic reconstruction estimates by providing additional characterizations of the timing and magnitude of temperature responses to volcanic eruptions when coupled with large-scale proxy estimates of past temperature variability e.g., from tree-rings. As a synthetic evaluation of the performance of the break-indicator saturation method, we search for volcanic eruptions in surface air temperature output from model simulations. While there is some disagreement on the timing, magnitude, and climatic impact of real eruptions over the past several millennia, the present simulation is forced with deterministic (known, imposed) eruptions. It therefore can function as a useful tool for assessing the detection efficacy of the proposed statistical methodology in real-world scenarios when the timing and exact DGP of volcanic eruptions are uncertain.

For our empirical illustration we use the Northern Hemisphere (NH) mean surface air temperature from the historical simulation of the National Center for Atmospheric Research (NCAR) Community Climate System Model 4 (CCSM4) and the Last Millennium (LM) simulation (Landrum et al. 2013). These simulations were made available as part of the Coupled and Paleoclimate Model Intercomparison Projects Phases 5 and 3 (CMIP5/PMIP3), respectively (Taylor et al. 2012). Collectively, the two simulations span the period 850-2005 C.E. To imitate potential proxy reconstructions (e.g., tree-ring based), temperatures for extratropical land areas (30 deg −90 deg N) were extracted from the model and only summer months (June-August) were used to build annual averages. This time period is expected to show the strongest cooling in response to an eruption (e.g., Zanchettin et al. 2013 argue for a winter-warming effect) and is associated with the seasonal sampling window of many proxies such as dendroclimatic records. Temperatures are reported as anomalies relative to the 1850-1999 mean. The model is forced with the volcanic reconstruction by Gao et al. (2008) that reports volcanic activity as stratospheric sulphate loadings in teragrams (Tg). While the model is forced with multiple radiative forcing conditions (e.g., solar irriadiance, greenhouse gases, volcanoes, land cover changes, and anthropogenic aerosol changes), for the present experiments we treat these as unknown and work with the univariate NH mean temperature series, although multivariate models with more forcing variables could improve the detection algorithm. For a real-world scenario, however, estimates of climate-forcing and -sensitivity are uncertain (IPCC 2013) and may prove to be of limited use in explaining non-volcanic temperature variation in proxy reconstructions, particularly in the presence of changes in measurement (see e.g. Pretis & Hendry 2013).

7.3.1 Simulation Setup

We design a break function to capture the temperature response to a large-scale volcanic eruption using a simple zero-dimensional energy balance model (EBM) that equates incoming to outgoing energy derived from simple physics-based models of climate (see e.g., section 1 in Rypdal 2012, section 1 in Schwartz 2012, or Pretis 2015a for linking system energy balance models to econometric system models):

$$C\frac{dT'}{dt} = F - \theta T' \tag{7.30}$$

where θ is the climate feedback, C is the heat capacity, T' the temperature deviation from steady state (similar to the measured temperature anomaly as a departure from a long-term average) and F denotes radiative forcing (the variable that in our system describes the volcanic shock). The feedback response time of the model is given by $\tau = \frac{C}{\theta}$. Assuming a volcanic forcing effect of an impulse injection of stratospheric aerosols of F decaying exponentially at rate $-1/\gamma$ yields the following functional form of

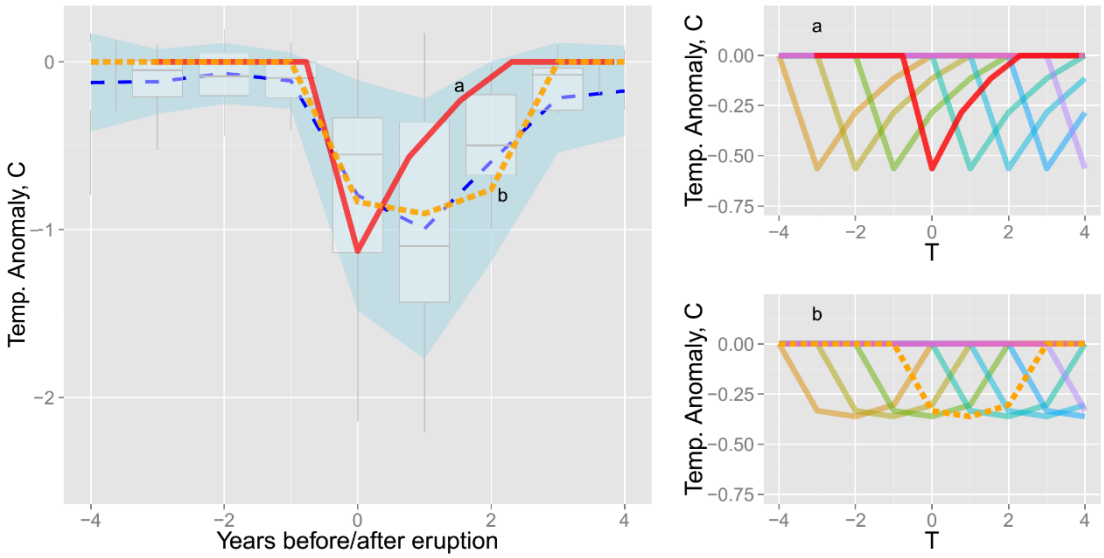


Figure 7.7: Superposed Epoch Analysis of the Model temperature response to simulated volcanic eruptions and sets of volcanic functions. Left: Superposed epoch analysis (Mass and Portman, 1989) of NH mean model temperature response to volcanoes with sulphate emissions $> 20Tg$ (42 events, dashed blue) with 1 sample standard deviation bands (shaded light blue) and distribution over volcanoes (box-plots). Approximate temperature response using a zero-dimensional energy balance model (EBM) used as volcanic function (a) is given in red and function (b) in orange. Right: Sets of EBM-based volcanic break functions for the two different specifications (a) (top) and (b) (bottom) to approximate the temperature response in years T relative to an eruption at $t = 0$.

a volcanic function for the associated temperature response:[14]

$$T_t' = d_t = \begin{cases} \frac{1}{C} e^{\frac{-\theta}{C}t} F \left(\frac{\theta}{C} - \frac{1}{\gamma} \right)^{-1} \left[e^{t\left(\frac{\theta}{C} - \frac{1}{\gamma}\right)} - 1 \right] & t \leq L \\ 0 & t < T_1, t > L \end{cases} \tag{7.31}$$

Intuitively, equation (7.31) states that a volcanic eruption through F leads to a sudden drop in temperatures, followed by a smooth reversion back to the original equilibrium. Different parameter calibrations are explored in the simulation section below. The main results are reported for a normalized temperature response where the feedback response time is set to 1, and the length of the volcanic impact is set to $L = 3$ to approximate the theory. The decay of stratospheric aerosols is modelled as $\gamma = 0.5$ (function a) and $\gamma = 3$ (function b) to capture one-period and two-period cooling respectively. On visual inspection (see Figure 7.7) these calibrations closely match the average-model response based on a superposed epoch analysis of all large-scale volcanic eruptions in the climate model (Mass & Portman 1989). The average model response in temperature is a drop by approximately 1–1.5 degrees C, followed by a smooth reversion to the previous mean over a 3-4 year period. While Gao et al. (2008) estimate the retention time for sulphate aerosols to be 2-3 years, a climatic perturbation of 4 years is in line with findings by Landrum et al. (2013). It is important to emphasize that the in-sample response to a volcanic eruption is not used to design the break function – the method is not trained and evaluated on the same set of observations.

In a more theoretical approach, which avoids particular shape parameters, a single peak (impulse) could be followed by auto-regressive reversion to the mean where we search over a full set of impulses and full set of breaking autoregressive coefficients. The DGP for the response variable NH temperature (T_t) is:

[14] See supplementary material 7.5.7 for a derivation. For break detection, the function is normalized to sum to 1 over L.

$$T_t = f(X_t, V_t) + \varepsilon_t \tag{7.32}$$

To simulate sampling uncertainty of a proxy-based reconstruction we generate 100 replications of the outcome by adding $\varepsilon_t \sim \mathrm{IN}(0, \sigma_\varepsilon^2)$ to the NH mean temperature. The main results here are presented for simulations setting $\sigma_\varepsilon = 0.2$ which is half the sample standard deviation of the NH time-series of 0.4: the effect of the magnitude of noise is explored in Figure 7.9. The function $f(X_t, V_t)$ mapping volcanic, V_t, and other forcing, X_t, on to temperature is unknown, and the observed forcing variables V_t and X_t are equally treated as unknown. As a proof of concept we consider two models (intercept-only, and AR(1) with intercept)[15] to detect eruptions:

$$y_t = \mu + \boldsymbol{\gamma}' \mathbf{d}_t + v_t \tag{7.33}$$

$$y_t = \rho y_{t-1} + \mu + \boldsymbol{\gamma}' \mathbf{d}_t + v_t \tag{7.34}$$

where $\mathbf{d}_t$ is a full set of volcanic break functions (7.31) to be selected over.[16] To reduce computational requirements due to the varying simulation setup, the full-sample is split into 10 sub-samples of $T = 115$ observations each.[17] There is little difference between full-sample and sub-sampling performance aside from computational speed (supplementary material 7.5.8 provides the results for a full-sample simulation). Selection is conducted at $\alpha = 0.01$ implying an expected gauge of 1% (approximately 1 break function spuriously retained per subsample). Higher retention of break functions can be an indicator of model misspecification. Simulations are evaluated based on the retention frequency of known individual volcanic events (potency), the average potency over all volcanoes and the proportion of spurious eruptions detected (gauge).

7.3.2 Illustration Results

Figure 7.8 and Tables 7.4 and 7.5 show the results of detected volcanic events in 100 replications of the modelled NH mean temperature[18] using the model (a) volcanic function. The retained volcanic breaks coincide predominantly with the simulated volcanic eruptions. Few spurious volcanoes are detected, and those that are spurious exhibit retention frequencies drastically lower than those of volcanoes used to force the model.

Most large-scale simulated volcanic eruptions are detected consistently: 74% of all larger ($> 20Tg$) NH eruptions are detected on average within an interval of ± 1 year (57% of all global eruptions, many of which appear to have had little impact on NH temperatures). Consistent with the basic analytical

[15] Unless otherwise stated, results refer to the intercept-only case. The intercept term and the auto-regressive terms are not selected over.

[16] Given the specification of the volcanic break function and if σ_ε was the only noise added to the DGP, then the approximate expected non-centrality for a single unknown break using (7.17) is $\lambda \left(0.2\sqrt{3.7}\right)^{-1} \approx 0.4^{-1}\lambda$ where λ is the full temperature response following a volcanic eruption. Since the specified volcanic function has an approximate trough of 0.58 λ, a temperature drop of 1 degree after a volcanic eruption implies that overall $\lambda \approx 1.7$. Thus in absence of additional noise and for a single volcanic break with an immediate temperature response of 1 degree, the expected t-statistic is approximately ≈ 4.3. The analytical probability of detecting this eruption is roughly: $P(|t| > c_\alpha) \approx 0.96$ for $\alpha = 0.01$. Large eruptions should be consistently detected if the break function is correctly specified and if σ_ε was the only source of noise.

[17] The total sample size is $T = 1155$, resulting in 9 sub-samples of $T = 115$ observations and one sub-sample of $T = 120$ observations. Significance levels are scaled accordingly. Using a 3Ghz processor the sub-sample approach requires ≈ 5 seconds to cover the entire sample for one replication (across 10 subsamples), compared to ≈ 5 minutes for one replication using a full-sample approach.

[18] Retained volcanic functions with positive coefficients are dropped since these likely constitute positive outliers. The focus here lies on the detection of volcanic events which have a negative temperature response.

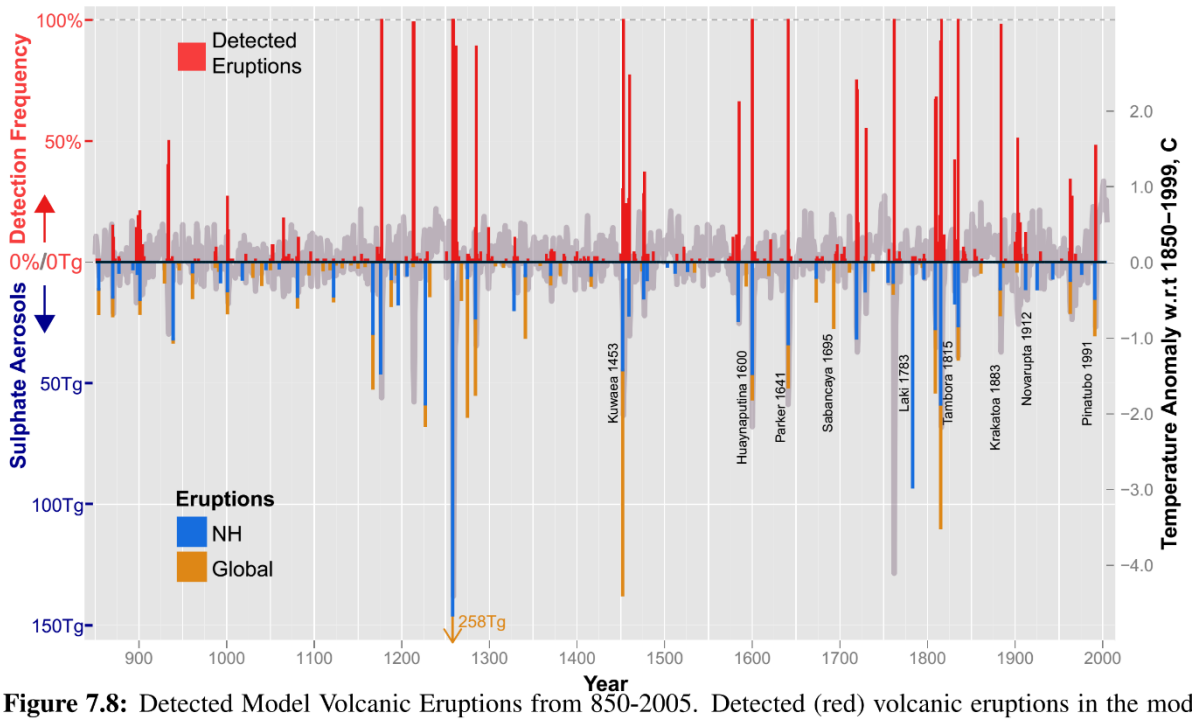


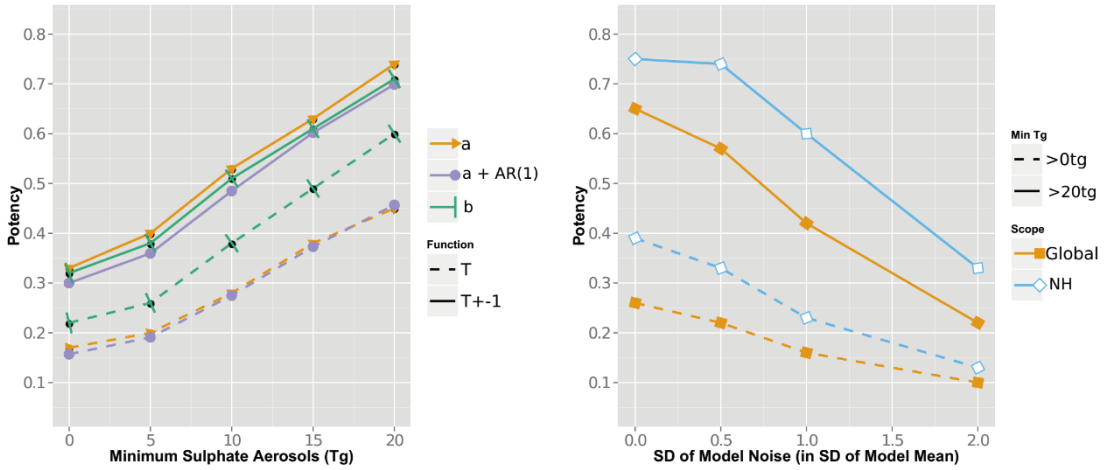
Figure 7.8: Detected Model Volcanic Eruptions from 850-2005. Detected (red) volcanic eruptions in the model temperature series from 850-2005 using function (a) modelling a single-period drop followed by a reversion to the mean together with an intercept. Bar height indicates detection frequency $[0, 100\%]$ across 100 simulations. Stacked sulphure deposition record used to force model temperatures are shown for Northern hemisphere (blue) and global measurements (orange) in Tg. Simulated model mean temperature anomalies used to detect the above volcanic eruptions are shown in grey. Mean NH surface temperature data is taken from the Last Millenium and historical simulation of the NCAR CCSM4 model as part of the CMIP5/PMIP3 data archive.

results presented in the previous section, the intervals of selection around the correct break dates are small. While increasing the band from 0 to 1 generally yields an increase in potency, outside of ± 1 year there is little difference (see Table 7.4). An uncertainty in break dates of $+/-$ 1 year can be the result of a monthly dated volcanic forcing record coupled with an annually dated temperature record, e.g. a December eruption will mainly affect the following year. The season of sulphure injection – before or after summer – can cause offsets in the timing of the temperature response. Equally there may be regional sampling biases based on the construction of the NH mean surface air temperature.

Augmenting the designed break functions (a) by an autoregressive model results in nearly similar potency and gauge relative to the baseline model using just a constant (see Table 7.4 and Figure 7.9).

The retention frequency of volcanic functions increases with the magnitude of sulphate emissions of the volcanic eruption (Figure 7.9). While the overall potency for all volcanoes in the Northern Hemisphere within a 1-year interval is 33%, this increases to 74% when larger volcanic eruptions over 20Tg are considered. Given that potency covers all of the volcanic forcing, much of which is small in magnitude, the result is unsurprising. In particular, the lower potency for small eruptions is not driven by an inconsistency in selection of the same volcano over multiple experiments, but rather in the variation in temperature response between volcanoes. Eruptions in 1641 (Parker) and 1600 (Huaynaputina) are detected 100% of the time while the eruption of 1783 (Laki) is not detected in any of the outcomes. In contrast to most of the other volcanoes, Laki is a high-latitude volcano. Because the CCSM4 model uses spatially resolved sulphate estimates, this eruption only affects the northernmost areas and causes only a minor hemispheric cooling of -0.15 degrees, which is much lower in magnitude than that of any of the

Figure 7.9: Left: detection potency of NH eruptions for given minimum sulphate emissions and timing for functions (a), (a) + AR(1), and (b) at the precise timing T (dashed) and in the interval of $T \pm 1$ (solid). Right: Detection for varying levels of noise added in the simulation for function (a) for all eruptions (dashed) and large eruptions over 20Tg (solid).



other major volcanic events (see Figure 7.7).[19]

Equally, the potency is affected by the chosen standard deviation of the noise process added to the model mean. The main results here are reported for added noise with a standard deviation of half the sample standard deviation. Figure 7.9 shows the potency for varying levels of noise.

The proportion of spuriously detected volcanoes (gauge) at around 0.02 is close to the nominal significance level ($1/T \approx 0.01$). The fact that it is slightly higher is likely due to the misspecification of the model, which is only run on a constant (including an auto-regressive term in the alternate specification) and set of break functions. Any variability in temperature other than volcanic eruptions may be spuriously attributed to the shape of the volcanic functions. This could be controlled by augmenting the model with additional dynamics (e.g. further autoregressive terms, long-term fluctuations through sine-cosine processes) or known forcing series.

Results for volcanic functions (b) are reported in the supplementary material 7.5.9 (see Tables 7.9 and 7.8). Volcanic functions (b) that capture the slower initial decline in temperature yield a slightly higher potency when measured at the precise timing (see Figure 7.9). Potency for $t = T_i$ for all i NH volcanoes using (b) is 0.32 vs 0.17 for (a) (0.23 vs 0.11 for Global). This result stems from the single drop in function (a) often being most significant in the second period after an eruption if the cooling lasts for two periods. Once we consider the interval of $T_i \pm 1$ years or volcanoes of larger scale the results are nearly identical for functions (a) and (b). Differentiation between one or two-period cooling following an eruption, and thereby further improvements in detection, could be implemented by searching over functions of type (a) and (b) simultaneously controlling the gauge appropriately.

In summary, large-scale volcanic eruptions can consistently be detected within a ± 1 year interval. Even though the model is likely mis-specified when using only a constant, few spurious volcanic eruptions are retained. The signal to noise ratio remains, however, crucial in detection. When the method is

[19] There is considerable uncertainty on the impact of the Laki eruption, for example A. Schmidt et al. (2012) find the observed NH peak temperature response to Laki to be around -1 degree, suggesting that the last-millennium simulation used here may not reflect the entire impact of the eruption. Notably, the eruption's noxious fumes at the time were discussed in White's (1789) treatment of phenology.

applied to real world proxy reconstructions where lower temperature spikes and higher noise levels can be expected, a well specified baseline model for the temperature process will be required against which volcanic events can be detected to ensure a high power of detection.

Forecasting during Breaks

While breaks (such as volcanic eruptions) are by their nature stochastic, using a deterministic approach through a full set of break functions allows us to account for the underlying breaks and model the responses deterministically. This can improve forecasts during breaks if the break function is well specified. Once the break is observed (in this case a volcanic eruption), a forecasting model can be augmented with a break indicator where the magnitude is determined through estimation in the first break period. This indicator then acts as a continuous intercept correction, thereby improving the forecast performance during the break. To illustrate this concept, Figure 7.10 shows a 1-step forecast for NH model mean temperatures following the simulated 1641 eruption, together with the root-mean-squared (RMSE) forecast errors for all NH ($> 20Tg$) model eruptions based on volcanic function (a). Using volcanic indicators to forecast through the breaks yields on average a lower forecast RMSE (RMSE = 0.51) when compared to a simple AR(1) model (RMSE = 0.71) or even a robust forecasting device (RMSE = 0.66) (Clements & Hendry 1999).[20] Crucially this depends on the correct specification of the break function – for volcanic eruptions further improvements could be achieved by switching to volcanic function (b) if the initial cooling lasts for two periods. Detection of breaks based on theory-informed break functions can therefore act as a robust forecasting device through a continuous intercept correction from climate to economic time series.

7.4 Conclusion

Saturating a regression model with a full set of designed break functions, and removing all but significant ones through a general-to-specific algorithm yields unbiased estimates of the break magnitude and time. By initializing the model with a full set of break functions many of the shortcomings associated with a forward selection or specific-to-general approach in break detection can be avoided. Analytical properties and non-centralities can be derived for any deterministic break function and can be extended to breaks in random variables when interacted with the deterministic break specifications. The break detection procedure exhibits desirable properties both in the presence of breaks (stable potency across multiple breaks) and under the null hypothesis of no breaks where the spurious retention of break functions can be controlled through a chosen significance level of selection. The multi-path algorithm (*Autometrics*) outperforms shrinkage-based estimators, especially when facing multiple breaks. We provide some initial insight into uncertainty on the break date by assessing the retention probability of mis-timed break estimators. Break indicator saturation appears to be effective for detecting large-scale temperature responses to volcanic eruptions. This was shown using surface air temperature output from a combined last millennium and historical climate simulation. Statistically searching over a set of break functions consistently detects large eruptions from the simulated surface air temperatures without prior knowledge of their occurrence. This holds promise for future volcanic detection efforts using real-world proxy re-

[20]The robust forecasting device is based on first differences using the forecasting model for $T+1|T$ given by: $y_{T+1|T} = y_T + \widehat{\rho}\Delta y_T$ where ρ is estimated using an AR(1) model. No error bars are shown on the green robust forecast in Figure 7.10 due to the non-standard distribution of the forecast.

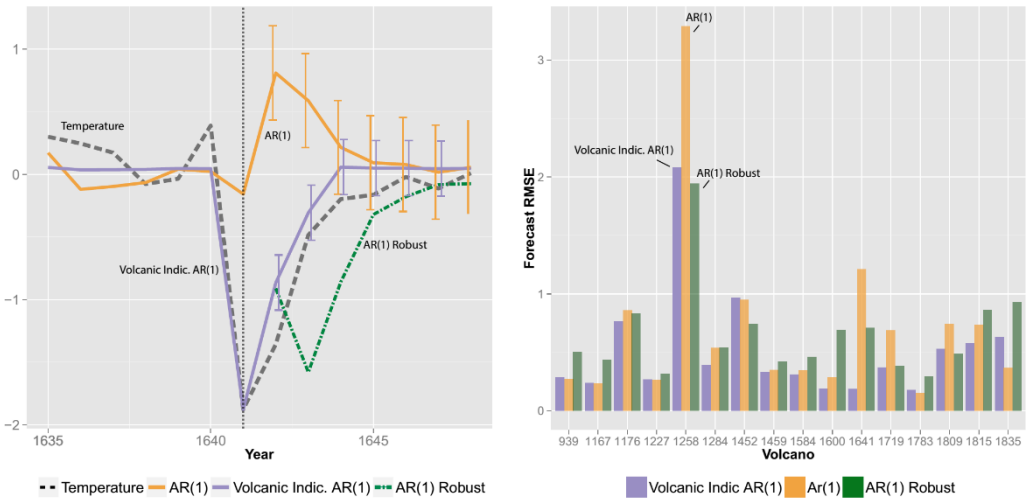


Figure 7.10: 1-Step forecasts through volcanic eruptions using break indicators. Left: Forecast performance across different methods: model mean temperature during the simulated 1641 eruption (yellow), 1-step forecasts from 1641 onwards are shown for using an AR(1) model with volcanic indicator (red), an AR(1) model without a volcanic indicator (blue) and a robust AR(1) forecast (green) (Clements & Hendry 1999). Models are estimated from 1605 until 1641. Right: 1-step forecast root-mean-squared-error over all NH model volcanic eruptions ($> 20Tg$) for an AR(1) model with volcanic indicator (red), without (blue), and robust AR(1) forecast (green). Using volcanic indicators, on average, improves the forecast performance during the break period. However, when no break occurs (little to no temperature response), using a break indicator can result in higher RMSE as seen for example for the 1783 model Laki eruption.

constructions of temperature variability over the last several millennia. More broadly, break detection using designed functions and indicator saturation provide a framework to analyse the detection of breaks of any designed shape at any point in time, with applications ranging from the detection of previously unknown events (such as shifts in time series due to measurement changes or policy impacts), to acting as a robust forecasting device during breaks – from economic recessions to volcanic eruptions.

Table 7.4: Potency and Gauge for Volcanic Functions (a)

Function (a)	T	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
Potency NH Tg$>$ 20	0.45	0.74	0.74	0.74
Potency NH Tg$>$ 0	0.17	0.33	0.34	0.35
Potency Global Tg$>$ 20	0.32	0.57	0.59	0.59
Potency Global Tg$>$ 0	0.11	0.22	0.25	0.26
Gauge NH	0.02			
Gauge Global	0.02			
Function (a) + AR(1)				
Potency NH Tg$>$ 20	0.46	0.70	0.70	0.70
Potency NH Tg$>$ 0	0.16	0.30	0.31	0.31
Potency Global Tg$>$ 20	0.31	0.52	0.54	0.54
Potency Global Tg$>$ 0	0.11	0.20	0.22	0.23
Gauge NH	0.02			
Gauge Global	0.02			

Table 7.5: Potency of detection of Volcanic Eruptions $> 20Tg$ using volcanic functions (a) for intervals $t = T \pm 1,2,3$

NH Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
939	31.83	0	0.02	0.02	0.03
1167	29.535	0	0	0	0
1176	45.761	0.06	1	1	1
1227	58.644	0.01	0.02	0.06	0.06
1258	145.8	1	1	1	1
1284	23.053	0.14	0.97	0.97	0.97
1452	44.6	0.3	1	1	1
1459	21.925	0.26	0.98	0.98	0.98
1584	24.228	0.11	0.77	0.8	0.8
1600	46.077	1	1	1	1
1641	33.805	1	1	1	1
1719	31.483	0.75	1	1	1
1783	92.964	0.02	0.02	0.03	0.05
1809	27.558	0.67	0.99	0.99	0.99
1815	58.694	0.91	1	1	1
1835	26.356	1	1	1	1
Global Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
854	21.387	0	0.02	0.03	0.03
870	22.276	0	0.25	0.25	0.25
901	21.283	0	0.34	0.5	0.54
939	33.128	0	0.02	0.02	0.03
1001	21.011	0	0.4	0.4	0.4
1167	52.114	0	0	0	0
1176	45.761	0.06	1	1	1
1227	67.522	0.01	0.02	0.06	0.06
1258	257.91	1	1	1	1
1275	63.723	0	0.06	0.08	0.08
1284	54.698	0.14	0.97	0.97	0.97
1341	31.136	0	0	0	0.01
1452	137.5	0.3	1	1	1
1459	21.925	0.26	0.98	0.98	0.98
1584	24.228	0.11	0.77	0.8	0.8
1600	56.591	1	1	1	1
1641	51.594	1	1	1	1
1693	27.098	0	0	0.03	0.07
1719	31.483	0.75	1	1	1
1783	92.964	0.02	0.02	0.03	0.05
1809	53.74	0.67	0.99	0.99	0.99
1815	109.72	0.91	1	1	1
1835	40.16	1	1	1	1
1883	21.864	0	0.98	0.98	0.98
1963	20.87	0	0.43	0.63	0.63
1991	30.094	0	0.48	0.48	0.48

References

Anchukaitis, K. J., Breitenmoser, P., Briffa, K. R., Buchwal, A., Büntgen, U., Cook, E. R., ... others (2012). Tree rings and volcanic cooling. *Nature Geoscience*, *5*(12), 836–837.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*, 47–78.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*, 1–22.

Baillie, M., & McAneney, J. (2015). Tree ring effects and ice core acidities clarify the volcanic record of the first millennium. *Climate of the Past*, *11*(1), 105–114.

Bergamelli, M., & Urga, G. (2013). Detecting multiple structural breaks: a Monte Carlo study and application to the Fisher equation for US. *Discussion Paper: Cass Business School, London*.

Brohan, P., Allan, R., Freeman, E., Wheeler, D., Wilkinson, C., & Williamson, F. (2012). Constraining the temperature history of the past millennium using early instrumental observations. *Climate of the Past Discussions*, *8*(3), 1653–1685.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, *3 (1)*, DOI: 10.2202/1941-1928.1097.

Castle, J. L., Doornik, J. A., & Hendry, D. F. (2015). Bias correction after selection with correlated variables. *University of Oxford Economics Discussion Paper*.

Castle, J. L., Doornik, J. A., Hendry, D. F., & Pretis, F. (2015). Detecting location shifts by step-indicator saturation during model selection. *Econometrics*, *3*, 240-264.

Castle, J. L., & Shephard, N. (Eds.). (2009). *The Methodology and Practice of Econometrics*. Oxford: Oxford University Press.

Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*. Cambridge, Mass.: MIT Press.

Cox, D. R., & Snell, E. J. (1974). The choice of variables in observational studies. *Applied Statistics*, 51–59.

Crowley, T. J., & Unterman, M. B. (2012). Technical details concerning development of a 1200-yr proxy index for global volcanism. *Earth System Science Data Discussions*, *5*(1), 1–28.

Doornik, J. A. (2009a). Autometrics. In J. L. Castle & N. Shephard (Eds.), (pp. 88–121). Oxford: Oxford University Press.

Doornik, J. A. (2009b). An object-oriented matrix programming language Ox 6.

Doornik, J. A. (2010). *Econometric model selection with more variables than observations* (Working paper). University of Oxford: Economics Department.

Doornik, J. A., & Hendry, D. F. (2015). Statistical model selection with big data. *Cogent Economics and Finance*, DOI:10.1080/23322039.2015.1045216.

Doornik, J. A., Hendry, D. F., & Nielsen, B. (1998). Inference in cointegrated models: UK M1 revisited. *Journal of Economic Surveys*, *12*, 533–572.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407–499.

Elliott, G., & Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series

regressions. *Journal of Econometrics*, *141*(2), 1196–1218.

Epprecht, C., Guegan, D., & Veiga, Á. (n.d.). Comparing variable selection techniques for linear regression: Lasso and autometrics. *Documents de travail du Centre d'Economie de la Sorbonne 2013.80*.

Ericsson, N. R. (2012). Detecting crises, jumps, and changes in regime. *Working Paper, Board of Governors of the Federal Reserve System, Washington, DC*.

Estrada, F., Perron, P., & Martínez-López, B. (2013). Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geoscience*, *6*, 1050-1055.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. Retrieved from `http://www.jstatsoft.org/v33/i01/`

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 (1984–2012)*, *113*(D23).

González, A., & Teräsvirta, T. (2008). Modelling autoregressive processes with a shifting mean. *Studies in Nonlinear Dynamics & Econometrics*, *12*(1).

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F., & Doornik, J. A. (2014). *Empirical Model Discovery and Theory Evaluation.* Cambridge MA: MIT Press.

Hendry, D. F., & Johansen, S. (2015). Model discovery and Trygve Haavelmo's legacy. *Econometric Theory*, *31*, 93-114.

Hendry, D. F., Johansen, S., & Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*, *23*, 337-339.

Hendry, D. F., & Krolzig, H.-M. (2005). The properties of automatic Gets modelling. *Economic Journal*, *115*, C32–C61.

Hendry, D. F., & Pretis, F. (2013). Anthropogenic Influences on Atmospheric CO2. In R. Fouquet (Ed.), *Handbook on energy and climate change* (pp. 287-326–121). Cheltenham: Edward Elgar.

Hoover, K. D., & Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, *2*, 167–191.

Huang, J., Horowitz, J. L., & Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics*, 587–613.

IPCC. (2013). *Fith assessment report: Climate change 2013: Working Group I Report: The Physical Science Basis*. Geneva: IPCC. Retrieved from `https://www.ipcc.ch/report/ar5/wg1/`

Johansen, S., & Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In J. L. Castle & N. Shephard (Eds.), (pp. 1–36). Oxford: Oxford University Press.

Johansen, S., & Nielsen, B. (2013). Outlier detection in regression using an iterated one-step approximation to the huber-skip estimator. *Econometrics*, *1*(1), 53–70.

Johansen, S., & Nielsen, B. (2016). Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics*.

Kelly, P. M., & Sear, C. B. (1984). Climatic impact of explosive volcanic eruptions. *Nature*, *311*(5988), 740–743.

Kitov, O., & Tabor, M. N. (2015). Detecting structural breaks in linear models: A variable selection approach using multiplicative indicator saturation. *University of Oxford Economics Discussion Paper*.

Kock, A. B., & Teräsvirta, T. (2015). Forecasting macroeconomic variables using neural network models and three automated model selection techniques. *Econometric Reviews*, *forthcoming*, DOI:10.1080/07474938.2015.1035163.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., & Teng, H. (2013). Last millennium climate and its variability in ccsm4. *Journal of Climate*, *26*(4), 1085–1111.

Mann, M. E., Fuentes, J. D., & Rutherford, S. (2012). Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures. *Nature Geoscience*, *5*, 202205.

Mass, F., Clifford, & Portman, A., David. (1989). Major volcanic eruptions and climate- a critical evaluation. *Journal of Climate*, *2*(6), 566–593.

Perron, P. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics*, *1*, 278–352.

Perron, P., & Yabu, T. (2009). Testing for shifts in trend with an integrated or stationary noise component. *Journal of Business & Economic Statistics*, *27*(3).

Perron, P., & Zhu, X. (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics*, *129*(1), 65–119.

Pretis, F. (2015a). Econometric models of climate systems: The equivalence of two-component energy balance models and cointegrated vars. *University of Oxford Economics Discussion Paper 750*.

Pretis, F. (2015b). Testing for time-varying predictive accuracy: a bias-corrected dynamic indicator saturation approach. *University of Oxford Economics Discussion Paper*.

Pretis, F., & Allen, M. (2013). Climate science: Breaks in trends. *Nature Geoscience*, *6*, 992-993.

Pretis, F., & Hendry, D. (2013). Some hazards in econometric modelling of climate change. *Earth System Dynamics*, *4*(2), 375–384.

Pretis, F., Mann, M. L., & Kaufmann, R. K. (2015). Testing competing models of the temperature hiatus: assessing the effects of conditioning variables and temporal uncertainties through sample-wide break detection. *Climatic Change*, *131*(4), 705–718.

Pretis, F., Sucarrat, G., & Reade, J. (2015). gets: general-to-specific model selection and indicator saturation. *R package version 0.4*.

Rypdal, K. (2012). Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *117*(D6).

Schmidt, A., Thordarson, T., Oman, L. D., Robock, A., & Self, S. (2012). Climatic impact of the long-lasting 1783 Laki eruption: Inapplicability of mass-independent sulfur isotopic composition measurements. *Journal of Geophysical Research: Atmospheres*, *117*(D23), n/a–n/a. Retrieved from `http://dx.doi.org/10.1029/2012JD018414` doi: 10.1029/2012JD018414

Schmidt, G., Jungclaus, J., Ammann, C., Bard, E., Braconnot, P., Crowley, T., ... others (2011). Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1. 0). *Geoscientific Model Development*, *4*(1).

Schwartz, S. E. (2012). Determination of earth's transient and equilibrium climate sensitivities from observations over the twentieth century: strong dependence on assumed forcing. *Surveys in Geophysics*, *33*(3-4), 745–777.

Strikholm, B. (2006). *Determining the number of breaks in a piecewise linear regression model* (Tech. Rep.). SSE/EFI Working Paper Series in Economics and Finance.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58:1*, 267–288.

Tibshirani, R. (2011). Regression shrinkage and selection via the Lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273–282.

White, G. (1789). *The Natural History of Selborne*. Oxford: Reprint by Anne Secord, Oxford University Press, 2013.

White, H. (2006). Approximate nonlinear forecasting methods. In G. Elliot, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp. 459–512). Elsevier.

Zanchettin, D., Timmreck, C., Bothe, O., Lorenz, S. J., Hegerl, G., Graf, H.-F., ... Jungclaus, J. H. (2013). Delayed winter warming: A robust decadal response to strong tropical volcanic eruptions? *Geophysical Research Letters*, *40*(1), 204–209.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.

8. An independent record of large volcanic events over the past millennium from reconstructed summer temperatures

L. Schneider[1], J.E. Smerdon[2], F. Pretis[3,4], C. Hartl-Meier[1] and J. Esper[1]

[1]*Department of Geography, Johannes Gutenberg University, Mainz, Germany*

[2]*Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, USA*

[3]*Programme for Economic Modelling, INET at the Oxford Martin School, University of Oxford, Oxford, UK*

[4]*Department of Economics, University of Oxford, Oxford, UK*

Submitted to Environmental Research Letters

8.1 Introduction

Knowledge about past volcanism can improve our understanding of the sensitivity of the climate system to exogenous radiative forcings. Associating large volcanic eruptions with climate anomalies and estimating their magnitude of impact, however, requires a comprehensive and well-dated record of past volcanism (Esper et al. 2013a). The vast majority of known eruptions in the past millennium have been dated by historical observations (Globale Volcanism Program 2013). In a climatological context, however, the amount and character of volcanic aerosols dispersed into the atmosphere are of greatest interest and not well characterized by historical observations. In particular, radiation absorbing sulfate aerosols injected into the stratosphere undergo multiyear transport and distribution resulting in significant alterations of the earth's energy budget (Robock 2000). In contrast to historical observations, volcanic particles deposited in polar ice sheets have been successfully exploited to estimate the amount and composition of sulfate aerosols from volcanic eruptions before the onset of modern observations. Combined with a model for atmospheric dispersion, ice-core deposition records are used to derive radiative forcing reconstructions to evaluate the impacts of volcanic eruptions in climate model simulations (Schmidt et al. 2011).

A concern for ice-core derived estimates of volcanic events is their potential for dating errors due to uncertainties in the age-depth models or false assignment of reference horizons to certain eruptions (Baillie and McAneney 2015, Sigl et al. 2015). Even for a correctly dated ash layer, the time lag between ash injection, atmospheric perturbations and polar deposition can be up to 2.5 years and therefore results in misinterpretations of environmental effects (Plummer et al. 2012). Irregular snow accumulation causes high spatial variability in the amount of deposited volcanic material and, hence, additional large uncertainties in the magnitude of sulfate records (Hegerl et al. 2006, Sigl et al. 2014). Finally, translating the sulfurous ash deposits into radiative forcing estimates requires several physical assumptions about the character of aerosols, their visual properties and atmospheric transport, and thus again adds further uncertainty to the quantification of the volcanic history. These shortcomings motivate the need to develop alternative and independent approaches to reconstruct the timing and climatic impact of volcanic eruptions directly from estimates of their large-scale temperature responses.

Previous studies of annually resolved climate proxies (Briffa et al. 1998a), model simulations (Atwood et al. 2016) and even early observational data (Jones et al. 2003) have revealed volcanic eruptions to cause severe surface temperature cooling at hemispheric scales. The pulse-like forcing of these events triggers structural breaks (shifts in the mean) in timeseries of hemispheric or global mean temperatures, the magnitude of which exceed the natural range of climate variability (Naveau et al. 2003). Our study is therefore motivated by the assumption that detecting and separating such breaks without prior knowledge of their occurrence will generate independent evidence of past volcanism constraining the timing and climatic effects of large eruptions.

We use a new Northern Hemispheric reconstruction of mean extratropical summer temperatures derived from maximum latewood density (MXD) records (Schneider et al. 2015) to test this hypothesis. Although some recent work has suggested that tree-ring records smooth and underestimate the cooling associated with volcanic eruptions (Mann et al. 2012), numerous studies have demonstrated a distinct cooling response in MXD data (Anchukaitis et al. 2012, D'Arrigo et al. 2013, Esper et al. 2013a). In contrast, temperature reconstructions based on tree-ring width, the far more abundant dendrochronological parameter, were shown to underestimate abrupt cooling events due to biological memory effects (Esper et al. 2015). The Schneider et al. (2015) reconstruction covers temperatures over more than a thousand summers and thus represents an ideal record for large eruptions. In this paper, we detect volcanic-induced breaks in this reconstruction independent of ice-core estimates using an indicator saturation method. This technique has been previously validated in a pseudoproxy context (Pretis et al. 2016) and is used to construct an independent chronology of volcanic eruptions that is then compared with existing ice-core-derived forcing records. We conclude by discussing the implications of our alternative record of past volcanism and evaluate our findings in the context of additional proxy reconstructions and climate model simulations.

8.2 Data and Methods

Tree-ring reconstructions

The employed Northern Hemisphere (NH) summer temperature reconstruction is based solely on MXD data (Schneider et al. 2015 (SCH15)). Although the MXD network is not as dense as for tree-ring width, four North American, seven European and four Asian MXD chronologies are available, each longer than 600 years (Fig. 8-1a). Together these chronologies enable a skillful calibration and transfer into summer temperature estimates of the extratropical NH. In contrast to the original SCH15 reconstruction, data processing was performed with a focus on high frequency variability: we removed age related noise from the density chronologies by calculating residuals from Hugershoff functions fit to the individual power-transformed MXD measurement series (Cook and Kairiukstis 1990, Cook and Peters 1997). This classical method (Briffa et al. 1998b) allows for a better approximation of the age trend and produces index chronologies that most accurately represent high-frequency variance, but at the cost of losing multidecadal to multicentennial trends (Esper et al. 2012).

We additionally use two recently developed summer temperature reconstructions derived from MXD plus tree-ring width networks (Stoffel et al. 2015 (STO15), Wilson et al. 2016 (WIL16)). STO15 chose site-specific detrending methods, combined local chronologies using principal components, and applied a high-pass filter to the final reconstruction for the analysis of volcanic signals. Wilson et al. (2016) aggregated published reconstructions, which were mostly detrended with a focus on preserving low-frequency variability.

Volcanic forcing records and model simulations

The third phase of the Paleoclimate Modelling Intercomparison Project (PMIP3) uses either the forcing timeseries of Gao et al. (2008, G08) or Crowley and Unterman (2013, C13) as the volcanic component of the radiative boundary conditions in the last millennium model simulations (Schmidt et al. 2011). Both volcanic reconstructions combine sulfate concentration records from partly overlapping sets of polar ice cores (Fig. 8-1a) to model atmospheric sulfate aerosol loadings and optical depth changes. The recently published record by Sigl et al. (2015, S15) uses new but fewer ice core records in their radiative forcing estimates, which are not accompanied by an atmospheric transport model, but will be implemented in phase four of PMIP (Jungclaus et al. 2016, Kageyama et al. 2016). Despite their common purpose, the three volcanic reconstructions differ regarding the targeted volcanic eruption estimate, their spatial and temporal resolution and timing of events. For comparison with the dendrochronological records, we converted G08 and C13 into annual, global averages of radiative forcing (see SI1) using linear scalings (Stothers 1984, Wigley et al. 2005). To further homogenize these estimates with the format of S15, we calculated time-integrated sums for each volcanic event from both the G08 and C13 records. The integrated value was assigned to the year of peak forcing, because this is most relevant for the anticipated temperature response. In S15, atmospheric transport is not yet modeled and therefore peak forcing years cannot be determined. In this record, the timing of the integrated forcing events corresponds to the year in which the sulfate concentrations initially exceeded a pre-defined threshold (Sigl et al. 2013; M. Sigl 2016, personal communication). Peak forcing can be expected in the same or subsequent year (Gao et al. 2008, Crowley and Unterman 2013).

For assessing the effect of volcanic forcing in climate models, we also investigated monthly surface temperature patterns in five PMIP3 last-millennium simulations after interpolating all model fields to 5x5° latitude-longitude grids. For comparison to the proxy reconstructions, the simulated summer (June-August) temperatures were extracted between 30 and 90° North and an area-weighted mean was calculated. Two of the simulations (BCC_CSM1.1 and CCSM4) used the G08 volcanic forcing estimate and two used C13 (GISS-E2-R and MPI-ESM-P) (Masson-Delmotte et al. 2013). IPSL-CM5A-LR was forced with an unpublished record from Ammann et al. (2007). Note that problems with the implementation of the forcing timeseries were reported for IPSL-CM5A-LR (Atwood et al. 2016; see SI2). Hereinafter the last-millennium simulations will be referred to by the following abbreviations: BCC, CCSM, GISS, MPI and IPSL.

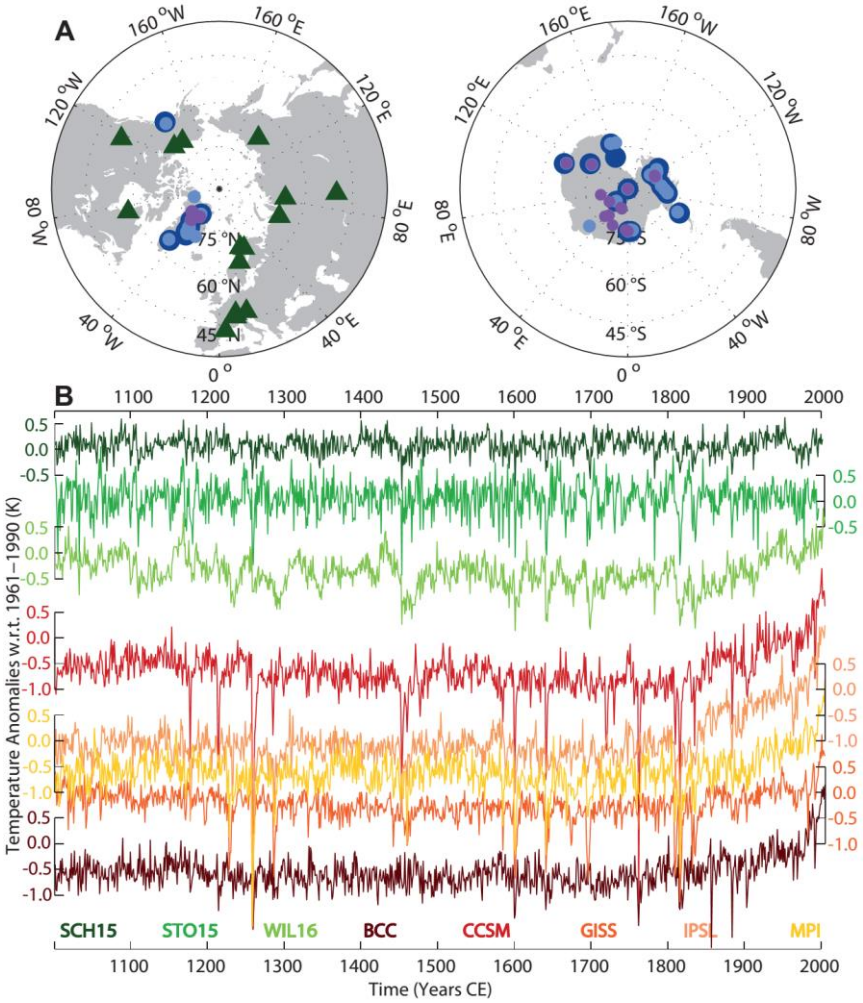


Figure 8-1 | Natural archives for past volcanism and last millennium climate model simulations. (a) Maps of the Northern and Southern Hemisphere in stereographic projection showing the origin of MXD chronologies used in SCH15 (green triangles) and the sets of ice-core records used for 3 different reconstructions of volcanic forcing (G08 light blue dots; C13 dark blue dots; S15 purple dots). (b) In greenish colors three recent NH summer-temperature reconstructions based on solely (SCH15) or partially (STO15; WIL16) MXD-data. Records are presented as in their original publications, i.e. with slightly different seasonal and spatial coverages. In reddish colors JJA mean temperatures for 30-90°N derived from climate model simulations for the last millennium as part of the PMIP3 experiments.

Detection Algorithm

For the detection of volcanic impacts in temperature reconstructions, we applied a method adopted from econometric timeseries modeling for the detection of outliers and breaks, which has been successfully demonstrated in the context of volcanic cooling (Pretis et al. 2016). The detection of breaks (and outliers) is formulated as a problem of model selection, in which a regression model for summer temperature is saturated with a full set of break indicators, representing volcanic eruptions at each step in time, of which all but significant indicators are removed (Doornik 2009, implemented in Doornik and Hendry 2013, see also Hendry et al. 2008, Castle et al. 2015, and Johansen and Nielsen 2016 for a general overview of indicator saturation). The novelty of this modeling approach applied to volcanic eruptions is in the design of the break function: It adapts the shape of the temperature response with an abrupt cooling

followed by a smooth recovery to the mean. A geometric decay by ~50% for two more years after the initial temperature drop (corresponding to Wigley et al. (2005) and Robock (2000)) is the simplified approximation to the modeled fading of stratospheric aerosol concentration (e.g. in G08 and C13).

The magnitude of the response is estimated jointly with selection over break functions. Its coefficient is the time-integrated temperature response over the three-year period of the volcanic function. The timing of the break can be interpreted as the year of peak forcing, equivalent to the categorization of the G08 and C13 forcing records.

The results of the break detection can be used to estimate the climatic impact of paleo-volcanism. Multiple break events from the temperature record are therefore averaged in a superposed epoch analyses (SEA, Panofsky and Brier 1958). To contextualize the estimated cooling for the different temperature records, the range of high-frequency temperature variability is calculated as the average standard deviation in moving 11-year windows over the past millennium.

8.3 Results and Discussion

Detection effectiveness

The algorithm automatically searches for both positive and negative breaks, where the detection of positive events likely captures outlying observations not well-approximated by the statistical model of temperatures (here modelled as autoregressive of order three). Abrupt warming should not exceed natural climate variability because there is no apparent external driver for such an event. Thus, negative spikes of the same magnitude are likely within this range of natural variability and not necessarily caused by a volcanic eruption. In order to avoid the detection of positively-signed (and spurious negative) breaks, the significance level is set conservatively (α=0.0065) so that only large events are picked up by the algorithm. This reduces the likelihood of a falsely identified break to less than 1% (Pretis et al. 2016) at the expense of missing breaks caused by smaller eruptions. Using SCH15 and the prescribed significance level, we detect 14 negative breaks in the temperature record during the past millennium and zero positive breaks (Fig. 8-2a, Tab. S8-1a and c). Common to all detected breaks in SCH15 is a sudden temperature drop, but the subsequent recovery-pattern varies strongly among the events (Fig. 8-2a). Immediate reversion to average temperatures is recorded after 1601, for instance, while temperature further decreases after the initial drop in 1698. Despite these differing response patterns, specifying a multi-year cooling in the detection model yields more rigorous results than an assumed single-year deviation: running the algorithm with a cooling model using a single-year impulse more often returns positive breaks, evidence for a less selective volcanic detection.

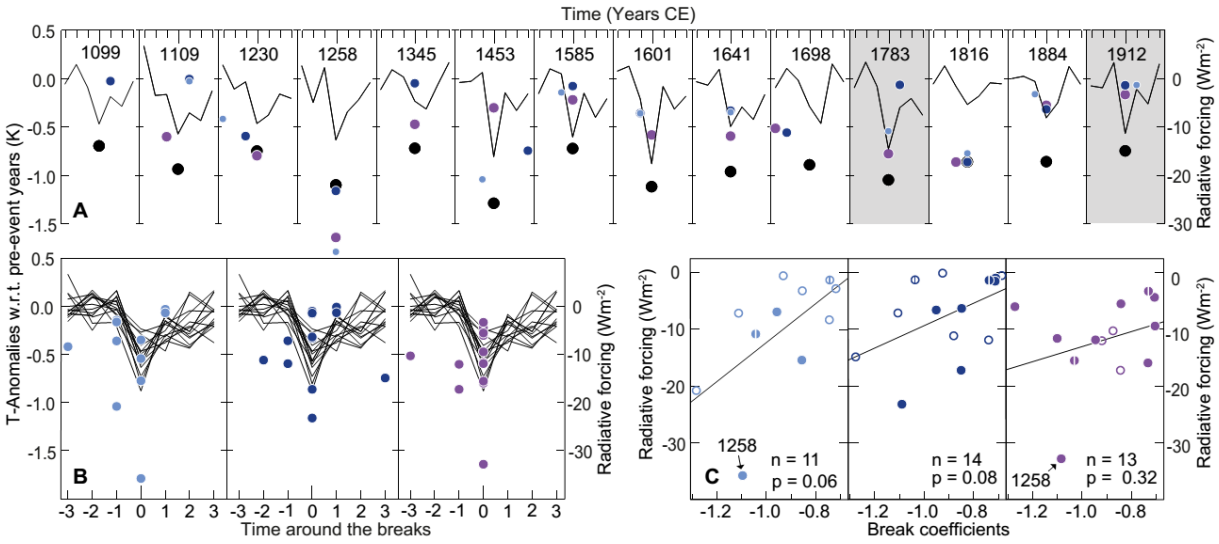


Figure 8-2 | Detected breaks in the temperature reconstruction SCH15. (a) Time frames for the 14 events ±3 years showing the temperature record and the coefficient for the strength of the detected break (black dot). Light blue (G08), dark blue (C13) and purple (S15) dots represent the strongest time-integrated forcing event in the corresponding period. (b) Temperature-anomalies and forcing events superposed for all breaks. Temperatures refer to the left axis and volcanic forcings to the right. Colors as in (a). (c) Scatterplots for the break coefficients and the equivalent events from the three forcing records. Filled dots indicate that break and forcing occurred in the same year and circles a temporal offset of up to three years. Dots with vertical bars refer to high latitude eruptions. In the lower right corner the number of data pairs (n) and the p-value of the regression slope is presented. Colors as in (a).

The number of detected events represents a conservative minimum of volcanic eruptions with hemispheric impact on summer temperatures during the last millennium. Because the detection model identifies events outside the range of internal climate variability we assume many smaller events to be masked. The known very large eruptions of the past millennium, including Samalas (1257), Tambora (1815) and Huaynaputina (1600), are all detected. Each of the 14 events can also be associated with a radiative forcing peak from C13, if we allow for a dating mismatch of up to ±3 years (Fig. 8-2a). Only 11 events are linked to G08 forcing events. Likewise, 13 of the 14 events are in line with peaks in S15 including nine cooling breaks ranking among the 13 strongest forcing events. The remaining four breaks correspond to global or NH forcing peaks ranking among the top 33 events in S15. These numbers are slightly weaker than suggested by previous pseudo-proxy experiments in which it was possible to detect all of the 15 largest events using the G08-forced CCSM simulation (Pretis et al. 2016)[1]. The high success rate is, however, fostered by a very pronounced volcanic forcing in CCSM. Additionally, an inaccurate or incomplete forcing reconstruction could compromise the successful verification of events detected in proxy-derived temperatures. Other reasons for a break detection in a year that is not associated with a

1 In Pretis et al. (2016) the reported proportion of correct detection for large eruptions is only 74%. Four events of significant sulfur injection in the G08-dataset (939, 1167, 1227 and 1783) did not reliably translate in breaks in the temperature record. The reason is that the model simulation behind the pseudoproxy-reconstruction was forced with an older version of G08 that was revised and corrected later. Considering the original version of the forcing record reveals that in all of the four cases breaks were detected in the years that were erroneously equipped with large sulfur peaks, moving the correctly detected proportion of large eruptions from 74% to 100%.

peak in the forcing record could be a mis-specified model in the detection algorithm or a deficient assumption about forced and internal climate variability.

Regarding the total number of events since 1001 CE documented in S15, we detect ~31% of the global eruptions (n=36) and ~3% of the NH eruptions (n=58) with the latter being usually much weaker and thus unable to cause cooling significantly outside the range of natural climate variability. These numbers are, however, dependent on the threshold chosen to separate background sulfate variability in the ice-core record from an actual volcanic source. With 118 events during the past millennium, Sigl et al. (2015) chose an intermediate threshold compared to the earlier forcing reconstructions G08 (67 events) and C13 (193 events).

Performance on other reconstructions and model simulations

If the STO15 and WIL16 reconstructions are used as the basis of the detection experiment, 20 and 15 volcanic events are detected, respectively. There are eight events detected in all tree-ring based reconstructions: these include 1258, 1453, 1601, 1641, 1698, 1783, 1816 and 1912 (Tab. S8-1a and Fig. S8-1). Five additional events are common among two records. For SCH15, only the smallest breaks, 1345 and 1099, are neither supported by STO15 nor WIL16. Using the reconstructions that additionally include ring-width data (STO15 and WIL16) yields three new strong cooling events that are likely volcanically driven: 1169/70, 1832 and 1835/6 (top 20 in S15). At the same time, the multiple positive breaks in STO15 and WIL16 indicate that these temperature records reconstruct abrupt warming events that are similar in absolute magnitude to volcanic-forced cooling. The setup for the detection algorithm was designed to impede the detection of positive breaks in SCH15. In STO15 and WIL16, however, four and five additional positive events are detected, respectively. Positive breaks occur largely independent of their negative counterparts, but appear to be more frequent in periods with reduced spatial coverage in the tree-ring networks (Tab. 8-1c). This could be related to model misspecification or increased noise during these periods. The magnitudes of positive breaks rank among the lower half of all events indicating that these are expected false-positives or small outliers with t-values close to the selection-significance threshold and thus likely retained by chance. With a lower significance level it is likewise possible to prevent the detection of positive breaks in STO15 and WIL16, but the number of remaining negative breaks would then shrink as well and be smaller than for SCH15. A different autoregressive structure in ring-width chronologies, potentially caused by biologic memory effects (Esper et al. 2015), make these records less applicable for the detection of volcanic events.

For comparison to the results from the temperature reconstructions, the detection algorithm is applied to forced-transient model simulations of the last millennium which yields between eight (BCC) and 37 (GISS) negative breaks. The wide range in the frequency of breaks is symptomatic of the internal variability associated with each model simulation and, most importantly, the very different model sensitivities to the ice-core derived estimates of volcanic forcing with some of them likely overestimating the climatic impact (Marotzke and Forster 2015, Stoffel et al. 2015). Major breaks in

years of strong forcing indicate that CCSM, GISS and MPI translate their volcanic forcing effectively into significant temperature anomalies, which the detection algorithm, in turn, successfully detects. The top ten negative breaks in these simulations comprise five to six events that are among the ten strongest forcing events of the respective ice-core forcing record. This result is similar to the agreement between the proxy-reconstructions and S15 (Tab. S8-1a), although forcing and climatic response are estimated from independent sources in the real-world comparison. Among the top 20 events in CCSM, GISS and MPI, at most one event is detected in periods of zero volcanic forcing indicating a high selectivity of the algorithm, but some of the weaker breaks occur in the year neighboring peak forcing. In BCC, CCSM, and GISS very strong and persistent forcing events like Samalas in 1258 and Tambora in 1815 result in a second detected break in the previous or subsequent year. BCC breaks only in eight instances and agrees least with the G08 volcanic forcing record used in the simulation. Of particular concern is the strongest break in BCC (1857), which is not associated with any forcing in G08, and the absence of a break in BCC for 1452/3 (2^{nd} strongest forcing in G08). Together with the generally small number of negative breaks in this simulation, there is some evidence for an underestimation of the volcanic effect in this model. For IPSL, these considerations cannot be evaluated because the original forcing data were not published.

The ranking of the simulated break magnitudes is often inconsistent with the size of the respective forcing peaks. Potential reasons are (1) forcing events in the Southern Hemisphere, which do not result in NH cooling (e.g. 1275 in G08) or extreme NH-cooling with a less outstanding global sulfate flux (e.g. 1903 in BCC). The latter scenario applies if the volcanically driven cooling coincides with a cold phase of internal climate variability and can result in an overestimation of volcanic cooling.

With 13 (GISS) to 21 (CCSM) cases, positive breaks are more common in the model simulations than in the proxy reconstructions. In contrast to STO15 and WIL16, almost all positive breaks in the simulations overlap with negative breaks or occur during the 20^{th} century warming. In 1000 observations and with a 1% likelihood of falsely identified breaks, some of the positive events are probably random outliers that might not be detected with a reduced significance threshold. Most of the positive breaks, however, are associated with considerably large coefficients and would not drop out before many of the negative breaks. The more regular occurrence of positive breaks (next to negative ones and during the 20^{th} century) indicates differences in the timeseries dynamics of the simulated records compared to the reconstructions: positive breaks following negative breaks could be indicative of a more abrupt termination of the cooling signal relative to the slower volcanic recovery based on geometric decay. In CCSM and IPSL, the recovery pattern for more than half of the negative breaks is altered through this effect. For the events in the 20^{th} century, greenhouse gas forcing is likely the driver of the anomalous positive deviations. Their detection is nonetheless surprising because the homogenous increase in greenhouse gas concentrations should be emulated by the autoregressive terms in the model. Although beyond the scope of this study these findings deserve more attention in subsequent analyses.

New insights on the timing of volcanic eruptions

Dendrochronological records are characterized by a precise annual age control. Reviewing the dating accuracy of the ice core records with respect to the breaks in the MXD reconstruction reveals improvement in the more recent S15 reconstruction (Fig. 8-2b). In G08, the oldest volcanic record, peak forcing matches only four of the detected breaks in the SCH15 reconstruction, and three breaks are not reproduced by any event even when considering the ±3 year uncertainty range. In the C13 record, containing most volcanic events, the eruption year is exactly matched in 7 out of 14 years, and all detected breaks are matched by a sulfate spike within the ±3 year uncertainty window. In S15, 10 events are synchronous among the 14 detected peaks, indicating an improved age control in this most recent ice-core chronology. However, S15 records no acidity spike in 1099, which is detected in the MXD reconstruction. As for the mismatch of the peaks in 1108 and 1815 (1109 and 1816 in SCH15), these differences likely reflect the ambiguity in the allocation of the forcing peak to a specific year mentioned in the data description. While S15 is so far not accompanied by a model for atmospheric aerosol dispersion, the application of the "Easy Volcanic Aerosol" forcing generator (Toohey et al. 2016) is in progress and will enable a more straightforward assignment of peak forcing years as calculated herein for G08 and C13.

The most ambiguous break in the tree-ring record is 1698. Although temperatures drop in two subsequent years during this seventh strongest break, the event does not accompany a forcing peak. The ice core records C13 and S15, in contrast, allocate large bipolar spikes in 1696 and 1695, respectively (Fig. S8-2b), which are similar in magnitude to the major eruptions in 1600 (Huaynaputina, Peru) and 1641 (Parker, Philippines). WIL16 and STO15 both feature breaks in 1695 as well as in 1698. STO15 and WIL16 rank these events oppositely (i.e. WIL16 ranks 1698 higher), but the temporally extended cooling in 1698 supports a more vigorous forcing for this later event. Three historically observed eruptions occurred in 1693 and 1694, but none of these explains the significant sulfate influx in both polar regions or the substantial cooling in the proxy records (Global Volcanism Program 2013): the volcanoes are either too far north to inject aerosols into the Southern Hemisphere or too small to have a global impact (Tab. S8-2). From the analysis of temperature reconstructions we conclude that there was a major forcing event in 1698 potentially accompanied by another, probably smaller, eruption in 1695. Aligning the major S15 peak with the 1698 break event shifts a smaller forcing peak from 1693 to 1696. Although asynchronous with the break in 1695, this provides additional evidence for a second eruption in the 1690s.

The MXD-based temperature reconstruction SCH15 suggests the cooling in 1453 to be the most pronounced during the past millennium. A significant volcanic impact in 1453 is supported by breaks in STO15 and WIL16 (rank 5 and 3, respectively). Although G08, C13 and S15 consistently report a major event in the 1450s, the character and timing of the eruption remains ambiguous among the forcing records (Fig. S8-2a). G08 peaks in 1452 and 1459 with the second spike being very small. C13 suggests a small eruption first in 1451 and then the major event in 1456. S15 agrees with C13 on the relative

magnitude, finds both eruptions slightly stronger and re-dates them to 1453 (-6 W/m^2) and 1458 (-20 W/m^2). In 1458, a forcing centered over the Southern Hemisphere, as suggested by Plummer et al. (2012), would explain the missing evidence in NH temperature reconstructions. In a Greenland ice-core Cole-Dai et al. (2013) found the signals in 1453 and 1458 to be of similar absolute magnitude but smaller than the Antarctic counterpart in 1458. None of the temperature reconstructions, however, breaks a second time in the 1450s (Fig. S8-2a and Tab. S8-1). With its major event in 1452, G08 is closest to our interpretation of the three proxy records and C13 and S15 are in noteworthy disagreement. More research is therefore needed to evaluate this obvious misfit including severe reconstructed cooling in 1453 and the absence of cooling in the late 1450s.

Besides 1458, there is only one eruption among the ten largest forcing events from S15 that does not appear as a break in one of the three temperature records. The 1809 forcing peak from a large tropical eruption of unknown source (Cole-Dai et al. 2009) did not result in a break in the temperature reconstructions (Fig. S8-2c). In line with STO15 and WIL16, SCH15 reports a cooling in 1809. In contrast to other events, temperatures do not return to the climatological mean after the initial drop, but remain low until the Tambora eruption in 1815. If the forcing records are correct in assigning a major sulfate loading to the 1809 eruption, then this prolonged cooling, potentially amplified by the Dalton Minimum (Raible et al. 2016), obscured the detection of the volcanic eruption.

The relationship between forcing magnitude and temperature response

Volcanic cooling is sensitive to the altitude, latitude, and character of the volcanic eruption (Hansen et al. 1997). The relationship between cooling patterns and forcing estimates is thus expecetd to be variable. The ice-core sulfate records agree, for example, on an enormous peak in 1258, but the reconstructed cooling is less extreme (Fig. 8-2c). Measuring and calibration uncertainties of single events in ice-core and tree-ring derived reconstructions further complicate such comparisons and impede the verification of the forcing magnitude based on the climatic impact. This is assessed by fitting linear regressions between the detected breaks and sulfate peaks from G08, C13 and S15 (Fig. 8-2b). G08 and C13 cohere relatively well with the break coefficients, but their regression models are based on large intercepts (22.1 and 10.4 W/m^2) which were physically expected to be zero. The S15 volcanic forcing is not significantly correlated with the break coefficients, mainly due to the differences in 1258 and 1453. The more minor breaks are associated with very small forcing events in G08 and C13 (smallest forcing = -0.5 W/m^2 and -0.1 W/m^2, respectively) and more substantial events in S15 (smallest forcing = -3.3 W/m^2). The latter better explains a temperature response outside the range of internal variability and results in a linear regression with an intercept close to the origin for S15. The relatively weak forcing associated with the maximum cooling in 1453 points to a major inconsistency in the 1450s forcing and temperature patterns. Although now off by 5 years, a forcing of -20 W/m^2 (in 1458) would better explain the strong 1453 cooling observed in the tree-ring reconstruction and would be well in line with the linear regression model. The pronounced radiative forcing in 1258 has been discussed previously (Timmreck

et al. 2009) and is likely too large due to nonlinear aerosol microphysics in the volcanic plume of that eruption.

Volcanic cooling in reconstructed and simulated temperatures

To isolate the influence of non-volcanic climate deviations, we average the cooling in response to strong volcanism over multiple events during the last millennium before comparing the proxy- and model-derived cooling estimates (Fischer et al. 2007). Previous studies with the same intension are challenged by the selection of key volcanic eruptions (Wilson et al. 2016). If these are drawn from ice-core forcing records, proxy reconstructions will necessarily underestimate the cooling compared to the simulated temperatures (Masson-Delmotte et al. 2013). The cooling signal in model simulations is stronger and more accentuated because the SEA events can explicitly be determined from the model forcing. For temperature reconstructions, in contrast, the true forcing is unknown and the signal becomes indistinct due to misdated or overestimated ice-core sulfate peaks. Drawing the events from precisely dated eruptions, on the other hand, yields a weaker cooling estimate in model simulations for the same reasons (Esper et al. 2013b). To avoid such discriminations we select the SEA events based on the break detection results for the respective temperature record. The cooling estimate is thereby maximised due to the intenional focus on the strongest deviations.

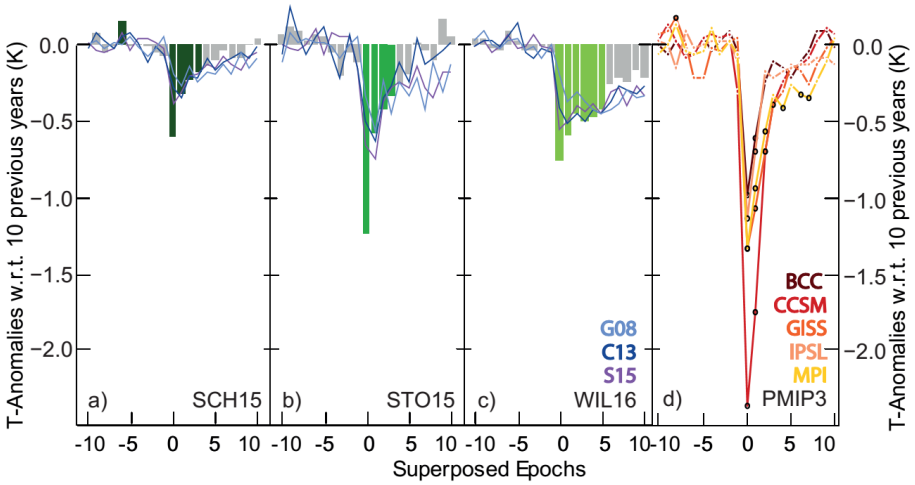


Figure 8-3 | Superposed Epoch Analyses for the 10 strongest negative breaks in the temperature records. (a) Temperature anomalies w.r.t. 10 pre-event years in the SCH15 reconstruction for the 10 strongest breaks. Colored bars indicate a significant ($p \leq 0.01$) cooling/warming. Purple and blue lines represent the cooling response in the SCH15 reconstruction based on the 10 strongest global forcing events derived from three different volcanic forcing timeseries (color codes as in Fig. 8-1). (b) and (c) as in (a), but for the reconstructions STO15 and WIL16. (d) Temperature anomalies w.r.t. 10 pre-event years in the PMIP3 model simulations for the 10 strongest breaks in the respective simulations. Black circles indicate significant ($p \leq 0.01$) cooling/warming.

This averaged climatic response to major volcanic events varies between the three reconstructions. STO15 shows a much stronger cooling signal, and WIL16 a longer cold phase compared to SCH15. Application of SEA to SCH15, STO15 and WIL16 reveals maximum post-volcanic cooling that ranges

from -0.60±0.13°C (SCH15)[2] to -1.23±0.21°C (STO15)[2] for the ten largest breaks (Fig. 8-3a-c). Using instead the top ten events from S15 in the SEA, the temperature peaks are 28% (WIL16) to 46% (STO15) smaller compared to the detection-derived SEA (Fig. 8-3a-c, Tab. S8-3). Our results do not confine the absolute magnitude of the expected temperature drop in response to large volcanic eruptions, because of the remarkable spread across the reconstructions. The discrepancy in absolute magnitude by a factor of two corresponds to the differences in variance among the NH-mean reconstructions (Fig. 8-1b): the standard deviation in the high-frequency domain is 0.16°C in SCH15 and 0.30°C in STO15 (Tab. S8-3). It reflects that the reconstructions vary with respect to their instrumental targets, spatial domains, methods used for detrending and network compilation. However, in all reconstructions, the ten major volcanic events cool the earth surface by four times this standard deviation. For comparison, the modeled temperature response to volcanic forcing exceeds the standard deviation by five (BCC, IPSL, MPI; six for GISS) to nine (CCSM) times, and is thus much stronger than the reconstructed values. The absolute values range from -0.98°C (BCC) to -2.37°C (CCSM).

Biological memory effects do not reduce the average cooling magnitude in the large-scale compilations including ring-width chronologies (STO15 and WIL16). The ratio between high-frequency variability and volcanic cooling is very similar to the purely MXD-based record SCH15 (Tab. S8-3), although the detection algorithm was more effective in the latter. The temporal persistence of low temperature estimates after the initial drop in WIL16 is, however, characteristic for ring-width data (Fig. 8-3c). The much faster recovery of the signal in STO15 suggests a small influence of ring-width chronologies in this reconstruction. Potentially the generally much noisier ring-width chronologies are suppressed in the principal component analysis applied in STO15. Similar to SCH15 the cooling in STO15 is reduced by half in the year subsequent to the initial drop (Fig. 8-3a-b). This fast relaxation exceeds the duration observed in simulated temperatures in which the post-event year is on average only 30% less cool than the event year itself. This result contradicts previous studies in which reconstructed temperatures were largely based on ring-width (e.g. Masson-Delmotte et al. 2013).

8.4 Conclusion

Large-scale temperature reconstructions can serve as an archive for climatically relevant volcanism. In a successful application of a new algorithm we detected all of the most relevant eruptions during the last millennium. The exact timing and, to a lesser degree, the magnitude of these cooling events can add useful information to existing forcing reconstructions. Using an independently derived list of volcanic eruptions to estimate the climatic response reveals an average cooling signal more accentuated and stronger in contrast to the signal in response to ice-core derived forcing events. Discrepancies in magnitude between tree-ring reconstructions are ascribed to differences in the year-to-year variability of temperature. Considering the range of internal, high-frequency temperature variability at hemispheric

[2] The error range refers to the standard deviation of the temperature in the central year of the SEA.

scale also reveals that most model simulations feature a more pronounced and more persistent volcanic signal than proxy reconstructions.

8.5 Acknowledgement

We thank Matthew Toohey for help with the G08-dataset and Sloan Coats for providing PMIP3 model fields. Supported by the German Science Foundation, Grant # 161/9-1 “Development of density chronologies for eastern and southern Europe”. LDEO contribution #XXXX. Supported by the Robertson Foundation and British Academy. The authors declare no competing financial interests.

8.6 References

Ammann, C. M., Joos, F., Schimel, D. S., Otto-Bliesner, B. L. and Tomas, R. A. (2007) 'Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model', *Proceedings of the National Academy of Sciences of the United States of America,* 104(10), 3713-3718.

Anchukaitis, K. J. et al. (2012) 'Tree rings and volcanic cooling', *Nature Geoscience,* 5(12), 836-837.

Atwood, A. R., Wu, E., Frierson, D. M. W., Battisti, D. S. and Sachs, J. P. (2016) 'Quantifying Climate Forcings and Feedbacks over the Last Millennium in the CMIP5-PMIP3 Models', *Journal of Climate,* 29(3), 1161-1178.

Baillie, M. G. L. and McAneney, J. (2015) 'Tree ring effects and ice core acidities clarify the volcanic record of the first millennium', *Climate of the Past,* 11(1), 105-114.

Briffa, K. R., Jones, P. D., Schweingruber, F. H. and Osborn, T. J. (1998a) 'Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years', *Nature,* 393(6684), 450-455.

Briffa, K. R., Schweingruber, F. H., Jones, P. D., Osborn, T. J., Shiyatov, S. G. and Vaganov, E. A. (1998b) 'Reduced sensitivity of recent tree-growth to temperature at high northern latitudes', *Nature,* 391(6668), 678-682.

Castle, J. L., Doornik, J. A., Hendry, D. F. and Pretis, F. (2015) 'Detecting location shifts by step-indicator saturation during model selection', *Econometrics,* 3(2), 240-264.

Cole-Dai, J., Ferris, D., Lanciki, A., Savarino, J., Baroni, M. and Thiemens, M. H. (2009) 'Cold decade (AD 1810-1819) caused by Tambora (1815) and another (1809) stratospheric volcanic eruption', *Geophysical Research Letters,* 36.

Cole-Dai, J., Ferris, D. G., Lanciki, A. L., Savarino, J., Thiemens, M. H. and McConnell, J. R. (2013) 'Two likely stratospheric volcanic eruptions in the 1450s CE found in a bipolar, subannually dated 800 year ice core record', *Journal of Geophysical Research-Atmospheres,* 118(14), 7459-7466.

Cook, E. and Kairiukstis, L. (1990) *Methods of dendrochronology : applications in the environmental sciences,* Dordrecht, Netherlands: Kluwer Academic Publishers.

Cook, E. R. and Peters, K. (1997) 'Calculating unbiased tree-ring indices for the study of climatic and environmental change', *Holocene,* 7(3), 361-370.

Crowley, T. J. and Unterman, M. B. (2013) 'Technical details concerning development of a 1200 yr proxy index for global volcanism', *Earth System Science Data,* 5(1), 187-197.

D'Arrigo, R., Wilson, R. and 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.

Doornik, J. A. (2009) 'Autometrics' in Castle, J. and Shephard, N., eds., *The methodology and practice of econometrics: a festschrift in honour of David F. Hendry,* Oxford: Oxford University Press, 88-121.

Doornik, J. A. and Hendry, D. F. (2013) *PcGive 14,* London: Timberlake Consultants.

Esper, J., Büntgen, U., Luterbacher, J. and Krusic, P. J. (2013a) 'Testing the hypothesis of post-volcanic missing rings in temperature sensitive dendrochronological data', *Dendrochronologia,* 31(3), 216-222.

Esper, J., Frank, D. C., Timonen, M., Zorita, E., Wilson, R. J. S., Luterbacher, J., Holzkämper, S., Fischer, N., Wagner, S., Nievergelt, D., Verstege, A. and Büntgen, U. (2012) 'Orbital forcing of tree-ring data', *Nature Climate Change,* 2(12), 862-866.

Esper, J., Schneider, L., Krusic, P. J., Luterbacher, J., Büntgen, U., Timonen, M., Sirocko, F. and Zorita, E. (2013b) 'European summer temperature response to annually dated volcanic eruptions over the past nine centuries', *Bulletin of Volcanology,* 75(7).

Esper, J., Schneider, L., Smerdon, J. E., Schone, B. R. and Büntgen, U. (2015) 'Signals and memory in tree-ring width and density data', *Dendrochronologia,* 35, 62-70.

Fischer, E. M., Luterbacher, J., Zorita, E., Tett, S. F. B., Casty, C. and Wanner, H. (2007) 'European climate response to tropical volcanic eruptions over the last half millennium', *Geophysical Research Letters*, 34(5).

Gao, C., Robock, A. and 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,* 113(D23), doi:10.1029/2008JD010239.

Global Volcanism Program (2013) 'Volcanoes of the World, v. 4.5.2', in Venzke, E., eds., *Smithsonian Institution,* downloaded 27 Jul 2016. http://volcano.si.edu/list_volcano_holocene.cfm.

Hansen, J., Sato, M. and Ruedy, R. (1997) 'Radiative forcing and climate response', *Journal of Geophysical Research-Atmospheres,* 102(D6), 6831-6864.

Hegerl, G. C., Crowley, T. J., Hyde, W. T. and Frame, D. J. (2006) 'Climate sensitivity constrained by temperature reconstructions over the past seven centuries', *Nature,* 440(7087), 1029-1032.

Hendry, D. F., Johansen, S. and Santos, C. (2008) 'Automatic selection of indicators in a fully saturated regression', *Computational Statistics,* 23(2), 317-335.

Johansen, S. and Nielsen, B. (2016) 'Asymptotic Theory of Outlier Detection Algorithms for Linear Time Series Regression Models', *Scandinavian Journal of Statistics,* 43(2), 321-348.

Jones, P. D., Moberg, A., Osborn, T. J. and Briffa, K. R. (2003) 'Surface Climate Responses to Explosive Volcanic Eruptions Seen in Long European Temperature Records and Mid-to-High Latitude Tree-Ring Density Around the Northern Hemisphere' in Robock, A. and Oppenheimer, C., eds., *Volcanism and the Earth's atmosphere,* Washington, DC: American Geophysical Union, 239–254.

Jungclaus, J. H. et al. (2016) 'The PMIP4 contribution to CMIP6 - Part 3: the Last Millennium, Scientific Objective and Experimental Design for the PMIP4 past1000 simulations', *Geosci. Model Dev. Discuss.*, 2016, 1-34.

Kageyama, M., Braconnot, P., Harrison, S. and Zhou, T. (2016) 'PMIP4-CMIP6: the contribution of the Paleoclimate Modelling Intercomparison Project to CMIP6', *Geoscientific Model Development Discussions*.

Mann, M. E., Fuentes, J. D. and Rutherford, S. (2012) 'Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures', *Nature Geoscience,* 5(3), 202-205.

Marotzke, J. and Forster, P. M. (2015) 'Forcing, feedback and internal variability in global temperature trends', *Nature,* 517(7536), 565-U291.

Masson-Delmotte, V. et al. (2013) 'Information from Paleoclimate Archives' in Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M., 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*, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 383–464.

Naveau, P., Ammann, C. M., Oh, H.-S. and Guo, W. (2003) 'An Automatic Statistical Methodology to Extract Pulse-Like Forcing Factors in Climatic Time Series: Application to Volcanic Events' in Robock, A. and Oppenheimer, C., eds., *Volcanism and the Earth's atmosphere,* Washington, DC: American Geophysical Union, 177-186.

Panofsky, H. A. and Brier, G. W. (1958) *Some applications of statistics to meteorology,* University Park, Pa.,: Mineral Industries Extension Services, College of Mineral Industries, Pennsylvania State University.

Plummer, C. T., Curran, M. A. J., van Ommen, T. D., Rasmussen, S. O., Moy, A. D., Vance, T. R., Clausen, H. B., Vinther, B. M. and Mayewski, P. A. (2012) 'An independently dated 2000-yr volcanic record from Law Dome, East Antarctica, including a new perspective on the dating of the 1450s CE eruption of Kuwae, Vanuatu', *Climate of the Past,* 8(6), 1929-1940.

Pretis, F., Schneider, L., Smerdon, J. E. and Hendry, D. F. (2016) 'Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-Indicator Saturation', *Journal of Economic Surveys,* 30(3), 403-429.

Raible, C. C. et al. (2016) 'Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects', *Wiley Interdisciplinary Reviews-Climate Change,* 7(4), 569-589.

Robock, A. (2000) 'Volcanic eruptions and climate', *Reviews of Geophysics,* 38(2), 191-219.

Schmidt, G. A. et al. (2011) 'Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0)', *Geoscientific Model Development,* 4(1), 33-45.

Schneider, L., Smerdon, J. E., Büntgen, U., Wilson, R. J. S., Myglan, V. S., Kirdyanov, A. V. and Esper, J. (2015) 'Revising midlatitude summer temperatures back to AD600 based on a wood density network', *Geophysical Research Letters,* 42(11), 4556-4562.

Sigl, M. et al. (2013) 'A new bipolar ice core record of volcanism from WAIS Divide and NEEM and implications for climate forcing of the last 2000 years', *Journal of Geophysical Research-Atmospheres,* 118(3), 1151-1169.

Sigl, M. et a. (2014) 'Insights from Antarctica on volcanic forcing during the Common Era', *Nature Climate Change,* 4(8), 693-697.

Sigl, M. et al. (2015) 'Timing and climate forcing of volcanic eruptions for the past 2,500 years', *Nature,* 523(7562), 543-+.

Stoffel, M. et al. (2015) 'Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1,500 years', *Nature Geoscience,* 8(10), 784-+.

Stothers, R. B. (1984) 'The Great Tambora Eruption in 1815 and Its Aftermath', *Science,* 224(4654), 1191-1198.

Timmreck, C., Lorenz, S. J., Crowley, T. J., Kinne, S., Raddatz, T. J., Thomas, M. A. and Jungclaus, J. H. (2009) 'Limited temperature response to the very large AD 1258 volcanic eruption', *Geophysical Research Letters,* 36.

Toohey, M., Stevens, B., Schmidt, H. and Timmreck, C. (2016) 'Easy Volcanic Aerosol (EVA v1.0): An idealized forcing generator for climate simulations', *Geoscientific Model Development Discussions*.

Wigley, T. M. L., Ammann, C. M., Santer, B. D. and Raper, S. C. B. (2005) 'Effect of climate sensitivity on the response to volcanic forcing', *Journal of Geophysical Research-Atmospheres,* 110(D9), doi: 10.1029/2004JD005557.

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

9. Conclusion and perspectives

This dissertation examines the climate sensitivity in response to volcanic perturbations from a paleoclimate perspective. Compiling a MXD-derived millennial temperature reconstruction for the Northern Hemisphere offers new insights in the magnitude, timing and frequency of volcanic induced cooling.

In contrast to previous large-scale reconstructions, this study is solely based on MXD-data, the less abundant but more temperature sensitive dendrochronological parameter compared to TRW (chapter 2). TRW responds with a temporally extended depression of up to six years to the short-lived volcanic cooling. MXD, in contrast, mirrors the abrupt temperature drop with subsequent recovery of up to three years in close agreement with the observed temperature response. Additionally, biologic memory effects and associated higher autocorrelation in the TRW chronologies results in reduced variability and suppressed extreme events. There is no evidence for an over- or underestimation of volcanic cooling in MXD-chronologies on the local scale (Tingley et al. 2014, Mann et al. 2012).

Regionally averaged MXD-records from the Alps and northern Fennoscandia also reproduce the cooling in response to volcanism (chapter 3). In these regions, the significance of the MXD-estimates is enhanced thanks to a particularly strong correlation with summer temperatures during more than 200 years of instrumental records. The volcanic cooling estimates are, however, very sensitive to the collection of volcanic events selected based on different criteria. Although the record of historically documented eruptions seems to miss some of the largest events, the average cooling in response to these events is stronger than in response to ice-core derived volcanic forcing events. This controversy exhibits the risk of dating errors in forcing reconstructions or, more generally, to ambiguities regarding the history of climatically relevant volcanism. Another source of uncertainty is the time lag between the volcanic eruption, peak forcing and maximum cooling. The average maximum temperature drop in northern Europe is as much as two years delayed, suggesting additional dynamic processes to augment the cooling beyond the effect of volcanic aerosols on net radiation. The more immediate but much weaker response detected in the Alps indicates a (i) temporally and (ii) spatially very heterogeneous impact of volcanic forcing.

(i) Averaging over a larger spatial domain mitigates some of these regional extremes and on a hemispheric scale the maximum cooling is not delayed by two years. However, the analysis of single events, dated to the month, revealed that the time lag between eruption and peak cooling can vary between a few months and 1.5 years (chapter 4). Superimposing and averaging multiple volcanic events thus results in a dampened and temporally extended mean volcanic cooling signal.

(ii) Temperature field observations for the extratropical Northern Hemisphere confirm the spatial heterogeneity of volcanic cooling observed in Europe (chapter 5). Whereas almost

all regions respond with negative anomalies to volcanism, the strength and timing of the temperature drop varies significantly. Although the MXD-sites represent regions of strong and weak volcanic cooling, the integrated temperature response seems slightly biased towards regions of enhanced impact. However, incomplete noise cancelation in the sparse proxy network probably results in a reduced peak size for short-lived, externally forced climate anomalies opposing a potential overestimation.

The aggregation of this MXD network yields a successfully calibrated summer temperature reconstruction for the extratropic Northern Hemisphere (chapter 6). Although volcanic events are clearly superimposed on background variability in this record, previous considerations hamper the selection of key event years. Shortcomings of existing volcanic archives additionally motivate the independent detection of volcanic cooling events in temperature records. Hence, a detection algorithm designed for the characteristic fingerprint of volcanic cooling is developed and successfully applied in a pseudoproxy environment (section 7).

In the MXD-derived temperature reconstruction the algorithm detects 14 cooling events over the last millennium comprising a new complementary archive of climatically relevant volcanism. It provides evidence that ice-core derived forcing reconstructions improved with respect to magnitude and timing of volcanic events, but still include periods of ambiguous volcanic activity like the 1450s and the 1690s. Averaging over ten detected events yields a maximum volcanic signal of -0.60±0.13°C and a more accentuated peak than previous estimates. Although comparisons among other temperature reconstructions and climate model simulations are impeded due to very different amplitudes of background variability, most of the analyzed model simulations seem to overestimate the effect of volcanism. These findings are supported by the increased persistence of volcanic cooling in simulated temperatures.

This dissertation overcomes limitations of the high-frequency signal in large-scale paleoclimatology and quantifies the climatic impact of volcanism during the past millennium. The analyses highlight the importance of harmonizing different environmental proxy archives and give new impetus for the improvement of existing records of volcanism, especially for reconstructions of volcanic forcing. The detected cooling events fill the gap between historically observed eruptions and the subsequent polar ash deposition. They can act as additional reference horizons in future attempts of dating and aligning ice-core chronologies. For the atmospheric transport models used in forcing reconstructions a refined intra-annual signal would be beneficial, because all eruptions with unknown seasonality are assumed to begin in April (Stevenson et al. 2016). Insights in the time lag between eruption and summer cooling can be used to better represent the pace of ash cloud propagation, distribution and deposition. In a spatial context, volcanic cooling can be analyzed with respect to regional peculiarities. Are there reoccurring patterns of particularly intense or weak volcanic cooling indicating a regionally varying sensitivity for volcanic forcing? Or does the cooling pattern differ for each eruption suggesting a dependency on the

location and type of eruption? Any inference about stratospheric transport and climatic impact could have implications for the atmospheric models in forcing reconstructions.

While advances in proxy-model comparisons contribute to a better understanding of the climate system and future global change, this dissertation also provides new benchmarks for probing into human history. Abrupt changes in the environment are often considered as societal turning points revealing the vulnerability of past cultures (Oppenheimer 2011). More detailed knowledge about strength, frequency and timing of volcanic induced cooling might construct – or deconstruct – links between human behavior and climate.

10. References

Anchukaitis, K. J., Breitenmoser, P., Briffa, K. R., Buchwal, A., Buntgen, U., Cook, E. R., D'Arrigo, R. D., Esper, J., Evans, M. N., Frank, D., Grudd, H., Gunnarson, B. E., Hughes, M. K., Kirdyanov, A. V., Korner, C., Krusic, P. J., Luckman, B., Melvin, T. M., Salzer, M. W., Shashkin, A. V., Timmreck, C., Vaganov, E. A. and Wilson, R. J. S. (2012) 'Tree rings and volcanic cooling', *Nature Geoscience*, 5(12), 836-837.

Baillie, M. G. L. and McAneney, J. (2015) 'Tree ring effects and ice core acidities clarify the volcanic record of the first millennium', *Climate of the Past*, 11(1), 105-114.

Brazdil, R., Reznickova, L., Valasek, H., Dolak, L. and Kotyza, O. (2016) 'Climatic effects and impacts of the 1815 eruption of Mount Tambora in the Czech Lands', *Climate of the Past*, 12(6), 1361-1374.

Briffa, K. R., Jones, P. D., Schweingruber, F. H. and Osborn, T. J. (1998a) 'Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years', *Nature*, 393(6684), 450-455.

Briffa, K. R., Schweingruber, F. H., Jones, P. D., Osborn, T. J., Shiyatov, S. G. and Vaganov, E. A. (1998b) 'Reduced sensitivity of recent tree-growth to temperature at high northern latitudes', *Nature*, 391(6668), 678-682.

Bunde, A., Buntgen, U., Ludescher, J., Luterbacher, J. and von Storch, H. (2013) 'Is there memory in precipitation?', *Nature Climate Change*, 3(3), 174-175.

Büntgen, U., Myglan, V. S., Ljungqvist, F. C., McCormick, M., Di Cosmo, N., Sigl, M., Jungclaus, J., Wagner, S., Krusic, P. J., Esper, J., Kaplan, J. O., de Vaan, M. A. C., Luterbacher, J., Wacker, L., Tegel, W. and Kirdyanov, A. V. (2016) 'Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD', *Nature Geoscience*, 9(3), 231-U163.

Crowley, T. J. and Unterman, M. B. (2013) 'Technical details concerning development of a 1200 yr proxy index for global volcanism', *Earth System Science Data*, 5(1), 187-197.

Cuny, H. E. and Rathgeber, C. B. K. (2016) 'Xylogenesis: Coniferous Trees of Temperate Forests Are Listening to the Climate Tale during the Growing Season But Only Remember the Last Words!', *Plant Physiology*, 171(1), 306-317.

Cuny, H. E., Rathgeber, C. B. K., Frank, D., Fonti, P., Makinen, H., Prislan, P., Rossi, S., del Castillo, E. M., Campelo, F., Vavrcik, H., Camarero, J. J., Bryukhanova, M. V., Jyske, T., Gricar, J., Gryc, V., De Luis, M., Vieira, J., Cufar, K., Kirdyanov, A. V., Oberhuber, W., Treml, V., Huang, J. G., Li, X. X., Swidrak, I., Deslauriers, A., Liang, E., Nojd, P., Gruber, A., Nabais, C., Morin, H., Krause, C., King, G. and Fournier, M. (2015) 'Woody biomass production lags stem-girth increase by over one month in coniferous forests', *Nature Plants*, 1(11).

D'Arrigo, R., Wilson, R. and 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.

D'Arrigo, R., Wilson, R. and Jacoby, G. (2006) 'On the long-term context for late twentieth century warming', *Journal of Geophysical Research-Atmospheres,* 111(D3).

Esper, J., Buntgen, U., Luterbacher, J. and Krusic, P. J. (2013a) 'Testing the hypothesis of post-volcanic missing rings in temperature sensitive dendrochronological data', *Dendrochronologia,* 31(3), 216-222.

Esper, J., Cook, E. R. and Schweingruber, F. H. (2002) 'Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability', *Science,* 295(5563), 2250-2253.

Esper, J., Schneider, L., Krusic, P. J., Luterbacher, J., Buntgen, U., Timonen, M., Sirocko, F. and Zorita, E. (2013b) 'European summer temperature response to annually dated volcanic eruptions over the past nine centuries', *Bulletin of Volcanology,* 75(7).

Esper, J., Krusic, P.J., Ljungqvist, F.C., Luterbacher, J., Carrer, M., Cook, E., Davi, N.K., Hartl-Meier, C., Kirdyanov, A., Konter, O., Myglan, V., Timonen, M., Treydte, K., Trouet, V., Villalba, R., Yang, B., Büntgen, U. (2016) 'Review of tree-ring based temperature reconstructions of the past Millennium', *Quaternary Science Reviews,* 145, 134-151.

Frank, D., Esper, J., Zorita, E. and Wilson, R. (2010) 'A noodle, hockey stick, and spaghetti plate: a perspective on high-resolution paleoclimatology', *Wiley Interdisciplinary Reviews-Climate Change,* 1(4), 507-516.

Franke, J., Frank, D., Raible, C. C., Esper, J. and Bronnimann, S. (2013) 'Spectral biases in tree-ring climate proxies', *Nature Climate Change,* 3(4), 360-364.

Fritts, H. C. (1976) *Tree rings and climate / H. C. Fritts,* London ; New York: Academic Press.

Gao, C., Robock, A. and 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,* 113(D23).

Lavigne, F., Degeai, J. P., Komorowski, J. C., Guillet, S., Robert, V., Lahitte, P., Oppenheimer, C., Stoffel, M., Vidal, C. M., Surono, Pratomo, I., Wassmer, P., Hajdas, I., Hadmoko, D. S. and 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 of the United States of America,* 110(42), 16742-16747.

Luterbacher, J. and Pfister, C. (2015) 'The year without a summer', *Nature Geoscience,* 8(4), 246-248.

Mann, M. E., Fuentes, J. D. and Rutherford, S. (2012) 'Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures', *Nature Geoscience,* 5(3), 202-205.

Marotzke, J. and Forster, P. M. (2015) 'Forcing, feedback and internal variability in global temperature trends', *Nature,* 517(7536), 565-U291.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González Rouco, J. F., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X. and Timmermann, A. (2013) 'Information from Paleoclimate Archives' in Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M., 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*, Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 383–464.

McCormick, M. P., Thomason, L. W. and Trepte, C. R. (1995) 'Atmospheric Effects of the Mt-Pinatubo Eruption', *Nature*, 373(6513), 399-404.

Oppenheimer, C. (2011) *Eruptions that shook the world*, Cambridge, UK: Cambridge University Press.

Pallardy, S. G. (1997) *Physiology of woody plants*, San Diego: Academic press.

Plummer, C. T., Curran, M. A. J., van Ommen, T. D., Rasmussen, S. O., Moy, A. D., Vance, T. R., Clausen, H. B., Vinther, B. M. and Mayewski, P. A. (2012) 'An independently dated 2000-yr volcanic record from Law Dome, East Antarctica, including a new perspective on the dating of the 1450s CE eruption of Kuwae, Vanuatu', *Climate of the Past*, 8(6), 1929-1940.

Robock, A. (2000) 'Volcanic eruptions and climate', *Reviews of Geophysics*, 38(2), 191-219.

Rossi, S., Deslauriers, A., Anfodillo, T. and Carraro, V. (2007) 'Evidence of threshold temperatures for xylogenesis in conifers at high altitudes', *Oecologia*, 152(1), 1-12.

Sigl, M., McConnell, J. R., Toohey, M., Curran, M., Das, S. B., Edwards, R., Isaksson, E., Kawamura, K., Kipfstuhl, S., Kruger, K., Layman, L., Maselli, O. J., Motizuki, Y., Motoyama, H., Pasteris, D. R. and Severi, M. (2014) 'Insights from Antarctica on volcanic forcing during the Common Era', *Nature Climate Change*, 4(8), 693-697.

Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., Buntgen, U., Caffee, M., Chellman, N., Dahl-Jensen, D., Fischer, H., Kipfstuhl, S., Kostick, C., Maselli, O. J., Mekhaldi, F., Mulvaney, R., Muscheler, R., Pasteris, D. R., Pilcher, J. R., Salzer, M., Schupbach, S., Steffensen, J. P., Vinther, B. M. and Woodruff, T. E. (2015) 'Timing and climate forcing of volcanic eruptions for the past 2,500 years', *Nature*, 523(7562), 543-+.

Smerdon, J. E. and Pollack, H. N. (2016) 'Reconstructing Earth's surface temperature over the past 2000 years: the science behind the headlines', *Wiley Interdisciplinary Reviews-Climate Change*, 7(5), 746-771.

Stevenson, S., Otto-Bliesner, B., Fasullo, J. and Brady, E. (2016) '"El Nino Like" Hydroclimate Responses to Last Millennium Volcanic Eruptions', *Journal of Climate*, 29(8), 2907-2921.

Stokes, M. and Smiley, T. (1968) *An introduction to tree-ring dating*, Chicago, Illinois, USA: University of Chicago Press.

Tingley, M. P., Stine, A. R. and Huybers, P. (2014) 'Temperature reconstructions from tree-ring densities overestimate volcanic cooling', *Geophysical Research Letters,* 41(22), 7838-7845.

Vaganov, E., Hughes, M. and Shashkin, A. (2006) *Growth Dynamics of Conifer Tree Rings. Images of Past and Future Environments.*, Berlin, New York: Springer.

Wigley, T. M. L., Ammann, C. M., Santer, B. D. and Raper, S. C. B. (2005) 'Effect of climate sensitivity on the response to volcanic forcing', *Journal of Geophysical Research-Atmospheres*, 110(D9).

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

List of Figures and Tables

Appendix I: Supplementary material for chapter 3

SUPPLEMENTARY MATERIAL

INCLUDING ONE TABLE AND SIX FIGURES

European summer temperature response to annually dated volcanic eruptions over the past nine centuries

Jan Esper[1,*], Lea Schneider[1], Paul J. Krusic[2], Jürg Luterbacher[3], Ulf Büntgen[4], Mauri Timonen[5], Frank Sirocko[6], Eduardo Zorita[7]

[1]*Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany*

[2]*Department of Physical Geography and Quaternary Geology, Stockholm University, 10691 Stockholm, Sweden*

[3]*Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus-Liebig University, 35390 Giessen, Germany*

[4] *Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland*

[5]*Finnish Forest Research Institute, Rovaniemi Research Unit, 96301 Rovaniemi, Finland*

[6]*Institute for Geoscience, Johannes Gutenberg University, 55099 Mainz, Germany*

[7]*Institute for Coastal Research, HZG Research Centre, 21494 Geesthacht, Germany*

Coupled General Circulation Models (CGCMs)

Four millennium-long simulations from three CGCMs were used for assessing post-volcanic cooling effects in Europe. These include two runs of the ECHO-G model (Erik1 and Erik2; Zorita et al. 2005), and two simulations downloaded from the CMIP5 dataset (Taylor et al. 2012). The latter two are the Max-Planck-Institute Earth System Model Paleoclimate version (MPI-ESM-P; Giorgetta et al. 2013), and the Community Climate System Model version 4 (CCSM4; Gent el al. 2011). The simulations retrieved from the CMIP archive cover the periods 850-1850 and 1850-2005 C.E., and were combined here to produce a single continuous temperature timeseries spanning the past millennium. The two ECHO-G simulations provide continuous coverage from 1000-1990 C.E. (see Fernández-Donado et al. 2013 for an overview on paleo model simulations).

All models are coupled atmosphere-ocean models. The ECHO-G model was developed at the Max-Planck-Institute for Meteorology in Hamburg and consists of the atmospheric model ECHAM4 and the ocean model HOPE. ECHO-G was used, with 21 others, in the Fourth Assessment Report issued by Intergovernmental Panel on Climate Change (Solomon et al. 2007) to produce future climate projections. The horizontal resolution of the ECHO-G model atmosphere (3.75 x 3.75 degrees) is also coarser than in the CMIP models.

The CMIP models used here can be considered as sophisticated climate models of the latest generation and are being used for climate projections included in next IPCC Assessment Report due in 2013. The MPI-ESM-P model consists of the spectral atmospheric model ECHAM6 with a horizontal resolution of approximately 1.87 x 1.87 degrees. The atmosphere model is coupled to the ocean model OM. The CCSM4 model was developed at the National Center for Atmospheric Research. The atmospheric model is a finite-difference model with a considerably finer horizontal resolution of about 1 x 1.25 degrees (latitude x longitude).

The external climate forcings used to drive the ECHO-G and CMIP5 models differ. All simulations include a representation of past solar energy output (total solar irradiance; TSI) assuming solar energy varies equally at all wavelengths of the electromagnetic spectrum. All simulations essentially assume the same shape of past TSI variability, derived from concentrations of the cosmogenic isotope ^{10}Be in polar ice cores, but the amplitude of the TSI variability may vary among simulations. The ECHO-G simulations assume a change in TSI of 0.3% between present (1961-1990) and the Late Maunder Minimum (1675-1715), whereas the CMIP models assume smaller changes of only 0.15%. Volcanic forcing in the ECHO-G simulations was implemented as a simple reduction of TSI, in an attempt to parameterize the effects of volcanic eruptions in a climate model with a coarsely resolved stratosphere. The estimations of volcanic eruptions affect on short wave radiative forcing are based on Crowley (2000). The CMIP models implement a more sophisticated volcanic forcing scheme with spatially and seasonally resolved variations of atmospheric optical depth. Estimates of optical depth changes are based

on different datasets of acidity measures in polar ice cores. The MPI-ESM-P model uses the optical depth reconstruction from Crowley (2008), the CCSM4 model used estimates from Gao et al (2008).

The ECHO-G and CCSM4 models consider past variations of carbon dioxide and methane estimated from trapped air bubbles in polar ice cores. In the MPI-ESM-P model, the concentration of carbon dioxide and methane are interactively simulated by a carbon-cycle module that incorporates the geochemistry of the ocean, the land biosphere and their interaction with climate.

Supplementary Table

Table S1. Long instrumental station records from Northern and Central Europe. "Period" specifies the time span common to all MXD chronologies (the Lau chronology ends in 1976) and the shorter (Stockholm, Berlin) and longer (Uppsala, Central England, De Bilt) station records. For station locations see Fig. S1

	Station	Country	Period	Source
N-Eur	Uppsala	Sweden	1722-1976	Moberg and Bergström (1997)
	Stockholm	Sweden	1756-1976	Moberg and Bergström (1997)
C-Eur	Centr. England	England	1722-1976	Manley (1974)
	De Bilt	Netherlands	1722-1976	van den Dool et al. (1978)
	Berlin	Germany	1756-1976	Vose et al. (1992)

Supplementary Figures

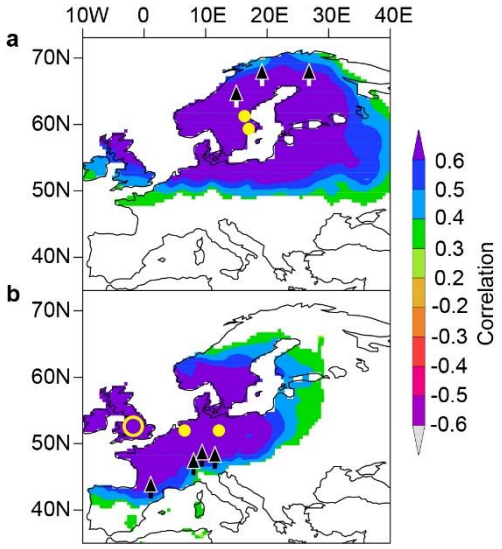


Fig. S1 Spatial domains of the long instrumental station records from Northern and Central Europe. **a** Correlation patterns of the mean JJA temperature of the Uppsala and Stockholm stations (yellow circles) with gridded summer temperatures over the 1901-1976 period ($p < 0.01$). Black triangles are the JAE, TOR, and NSC MXD sites (from west to east). **b** Same as in **a**, but for the mean timeseries of the Central England, De Bilt, and Berlin stations (from west to east). Triangles are the PYR, LOE, LAU, and TIR MXD sites (from west to east)

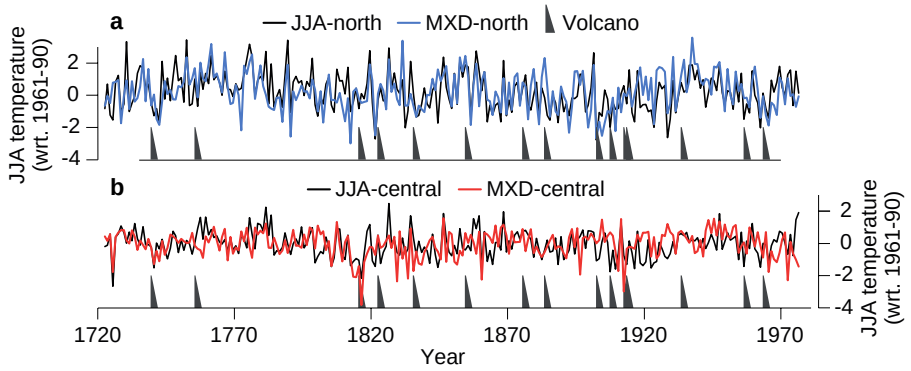


Fig. S2 Summer temperature from long instrumental station and MXD records in Northern and Central Europe. **a** JJA mean temperature from the Stockholm and Uppsala stations (black; JJA-north) together with the mean MXD record of JAE, TOR, and NSC (blue; MXD-north) over their 1722-1976 common period. Grey triangles indicate the 15 annually dated volcanic eruptions, VEI ≥ 5. The MXD timeseries was scaled to the instrumental data over the 1722-1976 period. **b** Same as in **a**, but for JJA mean temperatures from Central England, De Bilt, and Berlin (JJA-central) together with the mean MXD timeseries from PYR, LAU, LOE, and TIR (MXD-central). All timeseries expressed as temperature deviations from the 1961-1990 mean

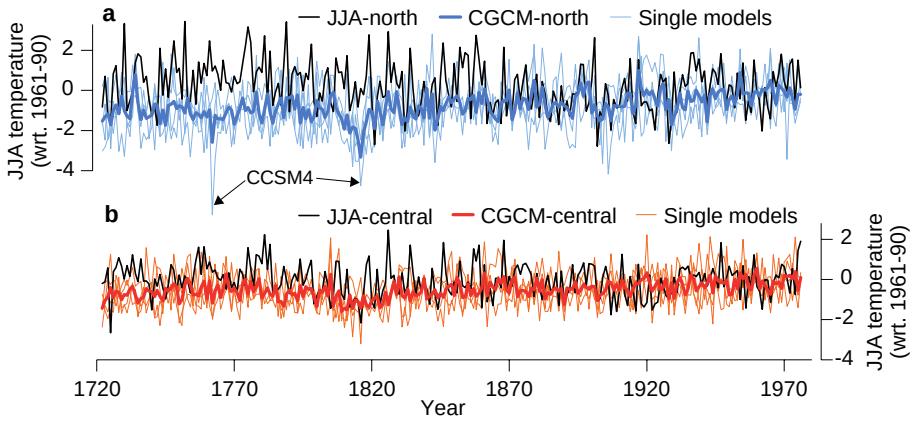


Fig. S3 Summer temperature from long instrumental stations and CGCM simulations. **a** JJA mean temperature of the Stockholm and Uppsala stations (JJA-north, black) together with the arithmetic mean of four CGCM simulations (CGCM-north, blue) over their 1722-1976 common period. Single model runs (CCSM4, Erik1, Erik2, MPI-ESM-P; thin curves) represent JJA temperatures of five grid points in proximity to the meteorological stations and MXD sites in Northern Europe. Severe negative deviations of the CCSM model in 1762 and 1816 following ice core derived volcanic sulfate depositions of 8.4 Tg and 59.7 Tg (Gao et al. 2008), respectively, are labeled. **b** Same as in **a**, but for JJA mean temperatures from the Central England, De Bilt, and Berlin stations (JJA-Central) together with the CGCM timeseries integrating summer temperatures at seven grid points in proximity to the station and MXD sites in Central Europe. All temperatures expressed as anomalies with respect to the 1961-1990 mean

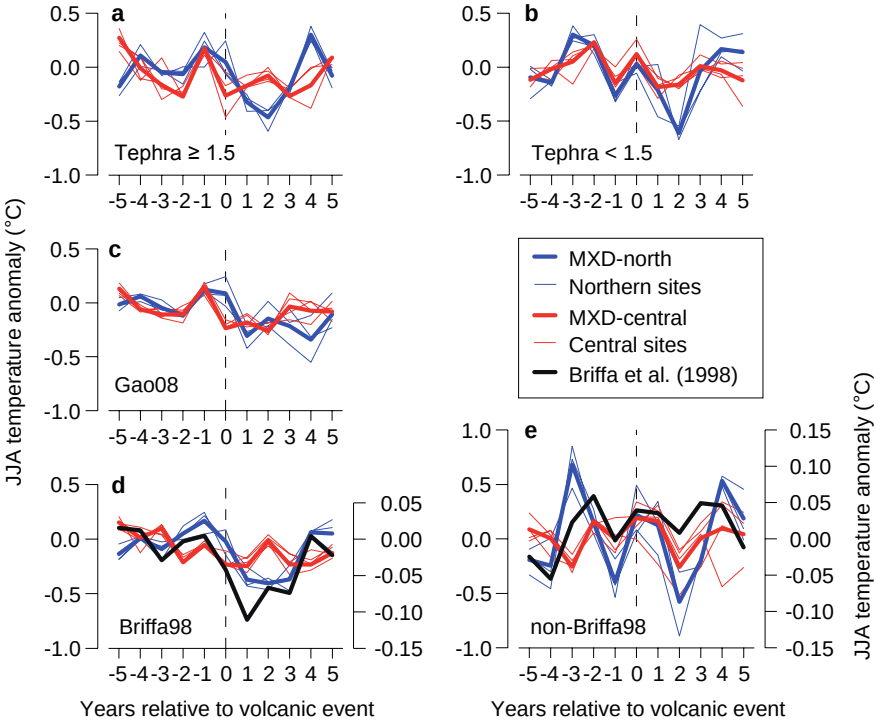


Fig. S4 Superposed Epoch Analyses (SEA) centered on documented and ice core reconstructed volcanic events. **a** JJA temperature patterns of MXD-north (blue) and MXD-central (red) five years before and after 22, very large volcanic eruptions (tephra volume ≥ 1.5 10^9 m^3) between 1111-1976 C.E. (SEA3 in Table 1). Thin curves are the SEA timeseries of the individual MXD site records JAE, TOR, and NSC in Northern Europe, and PYR, LAU, LOE, and TIR in Central Europe. **b** Same as in **a**, but for the 12 eruptions with tephra volumes 1–1.5 10^9 m^3 over the same period (SEA4). **c**, Same as in **a**, but for the 40 Northern Hemispheric volcanic events derived from sulfate depositions in ice cores over the 1111-1976 period (Gao et al., 2008; SEA7). **d** Same as in **a**, but the 31 volcanic events used in Briffa et al. (1998) over the 1400-1976 period. Black curve is the MXD-based JJA temperature reconstruction from Briffa et al. (1998) representing a large fraction of the Northern Hemisphere (scale on the right axis). **e** Same as in **d**, but for the 11 volcanic events between 1400-1976 that meet the criteria of **a** (VEI index ≥ 5, annually dated) and not used in Briffa et al. (1998). All SEA timeseries expressed as temperature anomalies with respect to the five years preceding volcanic events (lags -5 to -1). The temperature scale of the Briffa98 and non-Briffa98 SEAs is on the right-hand axes of panels **d** and **e**

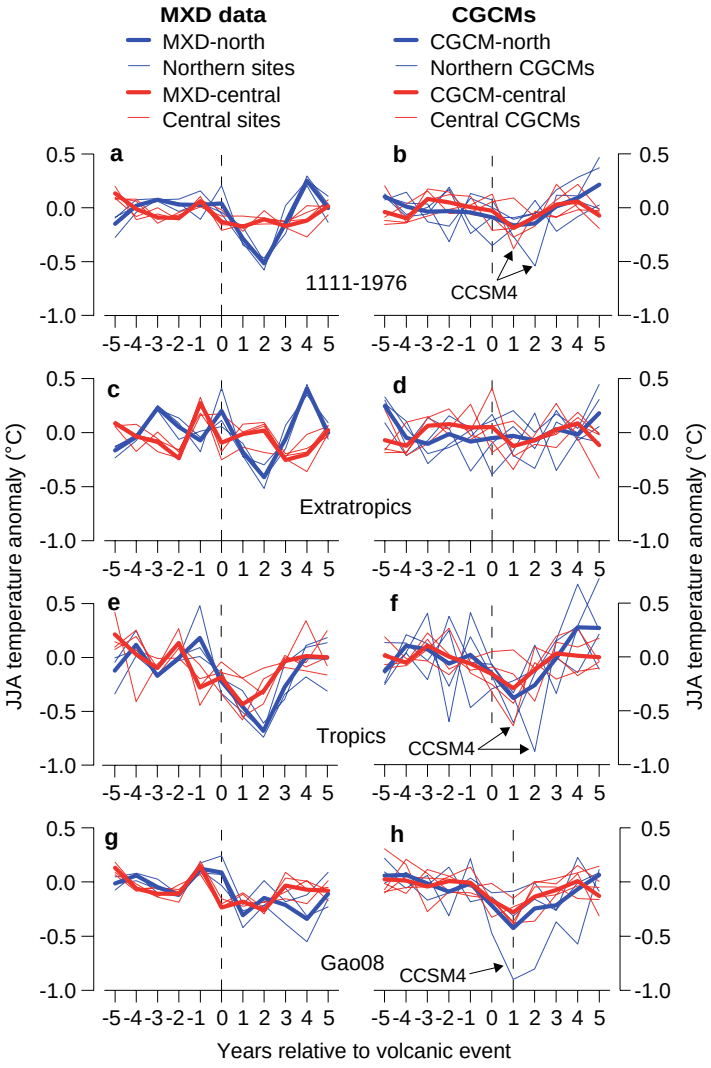


Fig. S5 SEA centered on documented and ice core derived volcanic eruptions. Left column shows results for MXD reconstructed JJA temperatures, right column shows results for CGCM JJA temperatures. **a** and **b** Temperature patterns in Northern (blue) and Central Europe (red) five years before and after 34 annually dated volcanic eruptions (VEI index ≥ 1.5) over the 1111-1976 C.E. period (SEA1 in Table 2). Thin curves are the SEA timeseries of the individual MXD site records (JAE, TOR, NSC in the North; PYR, LAU, LOE, TIR in Central Europe) and the long CGCM runs (ERIC1, ERIC2, MPI-ESM-P, and CCSM4). The simulations average temperature patterns of five grid points in proximity to the MXD and instrumental sites in Northern Europe, and seven grid points in proximity to the MXD and instrumental sites in Central Europe (see Table 3). The bold (red and blue) curves are the arithmetic means of the four CGCMs. **c-f** Same as in **a** and **b**, but for the 21 Extratropics and 13 Tropical stratospheric eruptions over the 1111-1976 C.E. period (SEA5 and SEA6 in Table 1). **g-h** Same as in **a** and **b**, but for the 40 NH eruptions identified by Gao et al. (2008) based on ice core data. All SEA timeseries expressed as temperature anomalies with respect to the five years preceding volcanic events (lags -5 to -1)

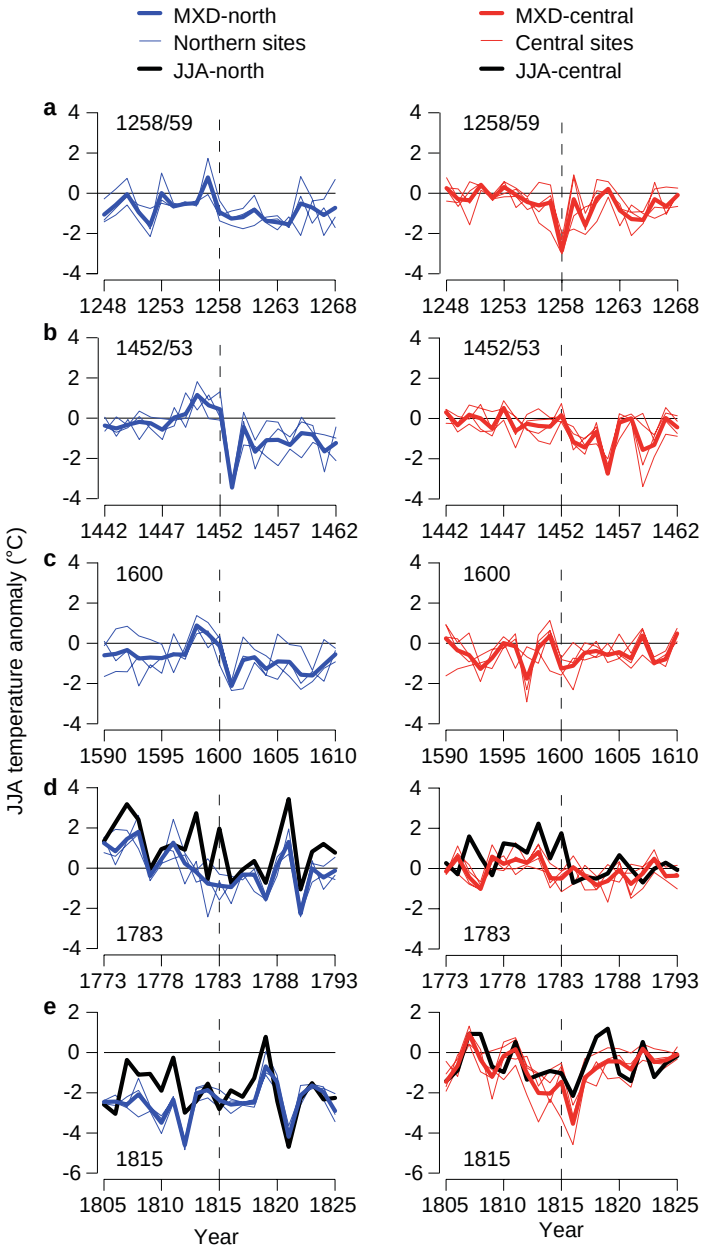


Fig. S6 JJA temperature anomalies (with respect to the 1961-1990 period) recorded in European MXD and long station records for the **a** 1258/59 (unknown), **b** 1452/53 (Kuwae), **c** 1600 (Huaynaputina), **d** 1783 (Laki), and **e** 1815 (Tambora) events. The zoom-in plots show the MXD site (thin blue and red curves) and regional chronologies (thick blue and red curves) together with the JJA-north and JJA-central temperature timeseries (black, panels **d** and **e**) ten years before and after the volcanic events.

References

Briffa KR, Jones PD, Schweingruber FH, Osborn TJ (1998) Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. Nature 393:450–455

Crowley TJ (2000) Causes of climate change over the past 1000 years. Science 289:270–277

Crowley T, Zielinski G, Vinther B, Udisti R, Kreutz K, Cole-Dai J, Castellano E (2008) Volcanism and the Little Ice Age. PAGES Newsletter 16:22–23

Crowley TJ, Unterman MB (2012) Technical details concerning development of a 1200-yr proxy index for global volcanism. Earth Syst Sci Data Discuss 5:1–28

Fernández-Donado L, et al. (2013) Large-scale temperature response to external forcing in simulations and 50 reconstructions of the last millennium. Clim Past 9:393–421

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. J Geophys Res 113 D23111. doi:10.1029/2008JD010239

Gent PR., et al. (2011) The Community Climate System Model Version 4. J Climate 24:4973–4991

Giorgetta MA, et al. (2013) Climate change from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project 5. J Advances Modell Earth Syst, in review

Manley G (1974) Central England temperatures: monthly means 1659 to 1973. Quart J Royal Met Soc 100:389–405

Moberg A, Bergström H (1997) Homogenization of Swedish temperature data. Part III: The long temperature records from Stockholm and Uppsala. Int J Climatol 17:667–699

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) (2007) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge Univ Press, Cambridge

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experimental design. Bull Amer Meteor Soc 93:485-497

van den Dool HM, Krijnen HJ, Schuurmans CJE (1978) Average winter temperatures at De Bilt (the Netherlands) 1634-1977. Clim Change 1:319–330

Vose RS, et al. (1992) The Global Historical Climatology Network (GHCN): long term monthly temperature, precipitation, sea level pressure, and station pressure data. Report ORNL/CDIAC-53, NDP-041 (available from Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1992)

Zorita E, Gonzlez-Rouco F, von Storch H, Montavez JP, Valero F (2005) Natural and anthropogenic modes of surface temperature variations in the last thousand years. Geophys Res Lett 32 L08707. doi:10.1029/2004GL021563

Appendix II: Supplementary material for chapter 6

Geophysical Research Letters

Supporting Information for

Revising mid-latitude summer-temperatures back to AD 600 based on a wood density network

Lea Schneider[1], Jason E. Smerdon[2], Ulf Büntgen[3], Rob J. S. Wilson[4], Vladimir S. Myglan[5], Alexander V. Kirdyanov[6] and Jan Esper[1]

[1]Department of Geography, Johannes Gutenberg University, 55099 Mainz, Germany

[2]Lamont-Doherty Earth Observatory, Palisades, NY 10964, USA

[3]Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland

[4]School of Geography and Geosciences, University of St Andrews, St Andrews KY16 9AL, Scotland, UK

[5]Institute for the Humanities, Siberian Federal University, Krasnoyarsk, 660041, Russia

[6]V.N. Sukachev Institute of Forest SB RAS, Akademgorodok, Krasnoyarsk, 660036, Russia

Contents of this file

Text S6-1. Supplementary Methods

Proxy Selection and Chronology Building For this study we compiled the most complete set of MXD-chronologies longer than 600 years. All datasets consist of relict and recent wood samples providing a relatively homogeneous age structure throughout the millennium. While this is a preferable feature for a RCS-detrending, not all sites revealed sufficient homogeneity to perform RCS with only one Regional Curve. Distinct inconsistencies regarding the tree species, the exact sampling site or the measurement technics made a split RCS approach mandatory for Athabasca, Jämtland and Mangazeja. This approach comes at the expense of low-frequency variability and was only applied when the smaller data portion significantly increased chronology length or sample replication. For Jämtland, neither was the case and hence, the update (Gunnarson et al. 2011) could not be integrated.
An advantage of MXD is the negligible level-to-spread relationship in contrast to ring width (Cook and Peters 1997). Thus a power transformation is not essential, but was employed in a few cases (10-20%) for individual measurements.

Benchmarking and Pseudoproxy Experiments The reconstruction performance was benchmarked using red noise time-series of adjusted first order autocorrelation as an alternative proxy network. Based on 100 different noise realizations, each nest of the true reconstruction exceeded the 99% quantile of the red-noise benchmarking experiments using the R^2 and RE metrics ($R^2_{min/max}$= .08/.10 and $RE_{min/max}$ = -.37/-.24).
Pseudoproxy experiments (see *Smerdon* 2012 for a review) were performed using the surface temperature field from the CCSM4 last-millennium simulation completed for the PMIP3 project [*Landrum et al.*, 2013; *Taylor et al.*, 2012]. The pseudoproxy network was derived from a 5x5 interpolated version of the surface temperature field and was sampled based on the location of the original site records. Signal-to-noise ratios of each pseudoproxy chronology reflected the correlation between the actual proxy record and their corresponding locally observed temperatures. Land-only NH extratropical JJA temperatures served as the reconstruction target in the model, with all grid cells from 30°-90°N being weighted for area.

Frequency Spectra Frequency spectra were calculated for 5 PMIP3 last millennium simulations (BCC, CCSM, GISS, IPSL, MPI) [*Taylor et al.*, 2012] and six annually resolved NH temperature reconstructions [*Masson-Delmotte et al.*, 2013] (both linearly detrended) over their common period 1402-1960. The extraction of model data followed the procedure detailed above for the CRUTEM4v temperature data. All records were scaled with respect to 1923-1960 representing the most recent common period in the length of a calibration interval. In the multi-taper spectra [Thomson, 1982] frequencies below 0.0025 and above 0.5 cycles per year were eliminated to allow for spectral characterizations from 2-300 years. The remaining part was regularly binned and spectra expressed as average bin values to guarantee equal weighting of all frequencies. β-values in Fig. 4b were derived from the power law [Franke et al., 2013] $P(f) \propto f^{-\beta}$, where P is the spectral energy and *f* the frequency. $-\beta$ equals the linear slope of spectral energy plotted against frequency on a logarithmic scale.

Site name	*Country*	*Location*	*Species*	*Period*	*n*	*r*	*Reference*
Alaska	US	68.8N/142.4W	Picea glauca	1073-2002	246	0.31**	Anchukaitis et al. 2013
Altai	Russia	50.0N/88.0E	Larix sibirica	462-2007	59	0.60**	Myglan et al. 2012
Athabasca	Canada	52.3N/117.3W	Picea engelmannii	1072-1994	102	0.56**	Luckman et al. 2005
Camphill	Canada	68.3N/133.3W	Picea glauca	1175-1992	58	0.47**	Schweingruber, ITRDB
Jämtland	Sweden	63.5N/15.5E	Pinus silvestris, Picea abies	1107-1978	158	0.63**	Schweingruber et al. 1988
Lauenen	Switzerland	46.4N/7.3E	Picea abies	982-1976	206	0.33**	Schweingruber et al. 1988
Lötschental	Switzerland	47.5N/7.5E	Larix decidua	735-2004	180	0.62**	Büntgen et al. 2006
Mangazeja	Russia	66.7N/82.3E	Larix sibirica, Picea obovata	1246-1990	143	0.70**	Schweingruber, ITRDB
Nscan	Finland	67.5N/22.5E	Pinus silvestris	-181-2006	587	0.82**	Esper et al. 2012
Polarural	Russia	66.9N/65.6E	Larix sibirica, Picea obovata	778-2006	141	0.76**	Briffa et al. 2013
Pyrenees	Spain	42.5N/2.5E	Pinus uncinata	1044-2005	203	0.40**	Büntgen et al. 2008
Quebec	Canada	57.5N/76.0W	Picea mariana	1352-1989	83	0.59**	Schweingruber, ITRDB
Torneträsk	Sweden	68.2N/19.7E	Pinus silvestris	441-2010	124	0.79**	Melvin et al. 2013
Tyrol	Austria	47.5N/12.5E	Picea abies	1047-2003	227	0.42**	Esper et al. 2007
Zhaschiviersk	Russia	67.5N/142.6E	Larix sibirica	1311-1991	55	0.30*	Schweingruber, ITRDB

Table S6-1. Site information of the 15 MXD-datasets. 'Period' covered by MXD samples is given in years AD, 'n' denotes the number of samples and 'r' the Pearson correlation coefficient against a nearby CRUTEM4v grid cell over the 1901-1976 period. ** indicates 95% significance, * 90% significance.

	RCS 1	RCS 2	RCS SF
Detrending	RCS	RCS	RCS + Signal Free
RC-smoothing	100 yrs spline	Age-dependent spline	Age-dependent spline
Variance Stabilization	100 yrs spline through absolute departure	Replication-dependent (51 yrs running window)	Replication-dependent (51 yrs running window)
Software	ARSTAN (Cook and Krusic)	RCSigfree (Cook, Krusic and Melvin)	RCSigfree (Cook, Krusic and Melvin)

Table S6-2. Detrending approaches. RCS1 is the detrending setup finally chosen. RCS2 and RCS SF are alternative detrendings assessed in this study. Besides the basic detrending choice (1st row), differing Regional Curve smoothings (2nd row), and variance-stabilization techniques (3rd row) were tested. Additionally, two software packages were used.

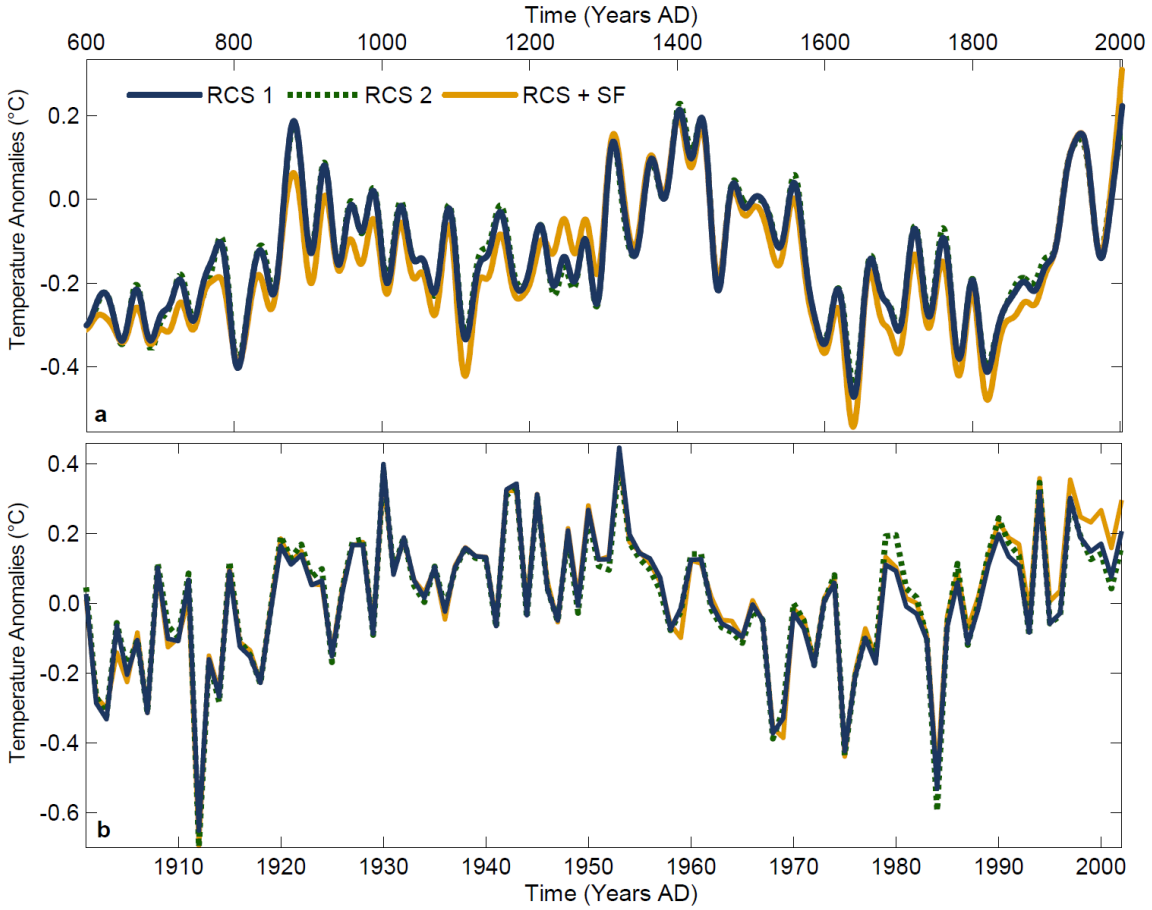


Figure S6-1. Reconstructions based on different detrending methodologies (see Tab. S2). a, Three low-pass filtered (30-years) reconstructions over the past 1400 years. b, The unsmoothed records over the 20th century. Temperature anomalies refer to the 1901-1976 period.

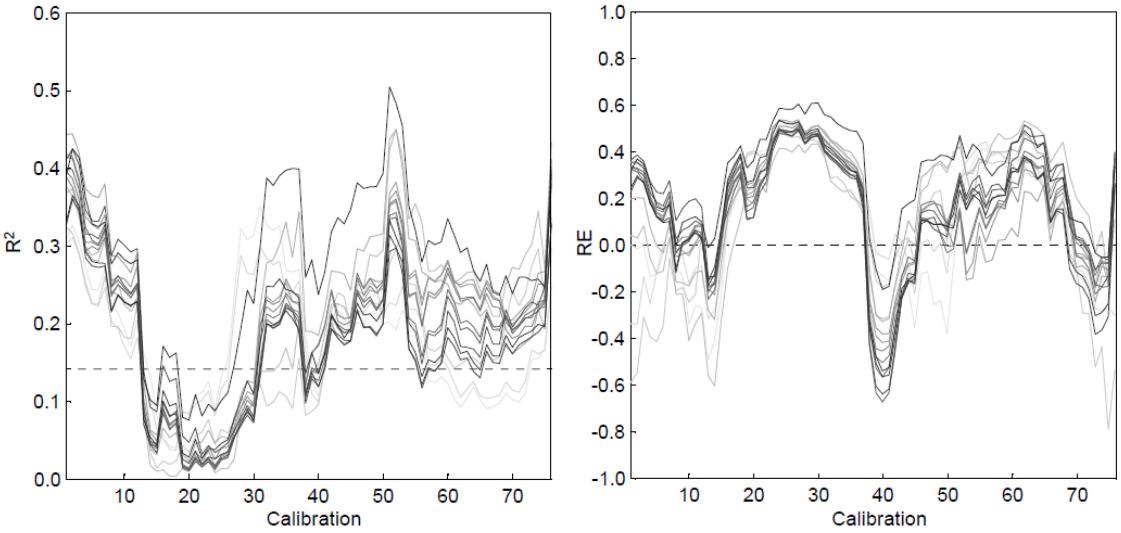


Figure S6-2. Validation metrics of 76 different calibration/validation periods and 15 nests (grey lines). The first calibration (validation) uses the 1901-1938 (1939-1976) period. This period is shifted by one year with each calibration so that the second setup uses 1902-1939 (1940-1976 and 1901) for calibration (validation), resulting in 76 values for each nest (plotted on the y-axis). Colors brighten with thinning replication. Dashed lines indicate $p < 0.01$ (one sided) for R^2 and zero for RE.

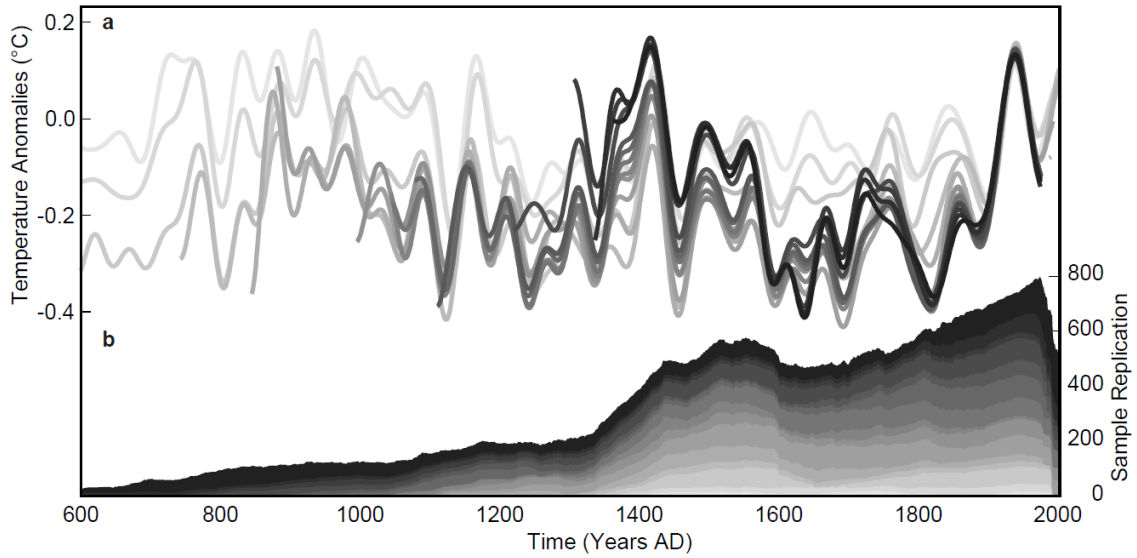


Figure S6-3. Reconstruction nests and replication of the single sites. a, Reconstructions calculated using changing site numbers, with the brightest (darkest) and longest (shortest) reconstruction based on 1 (15) record(s). All time-series are smoothed using a 30-year low-pass filter and expressed as anomalies w.r.t 1901-1976. The final reconstruction integrates the best-replicated reconstruction segments available at each period in time. b, Number of samples per site.

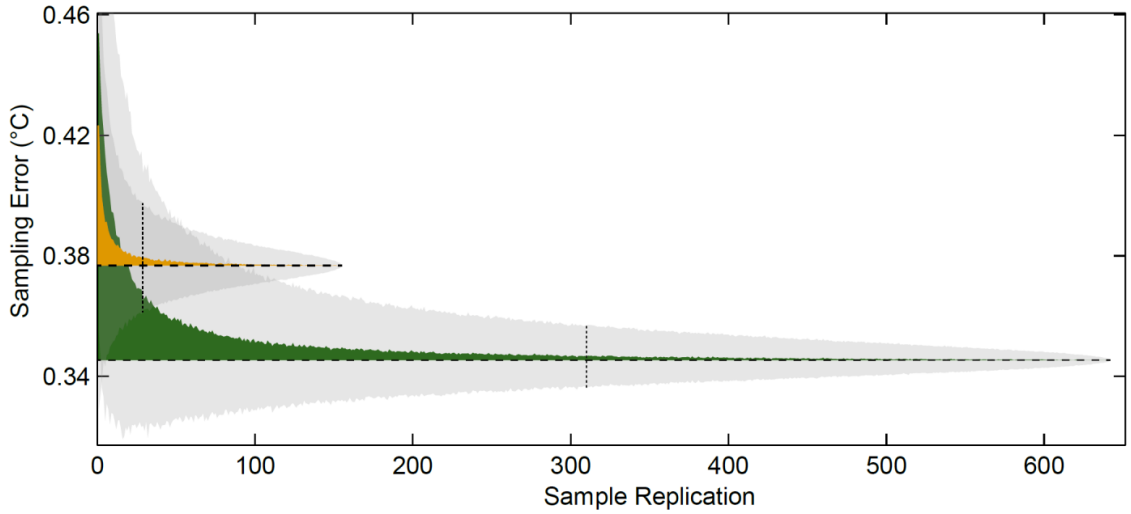


Figure S6-4. Replication error estimates. The green and yellow areas show the additional error introduced by decreasing sample replications for the nests including maximum and minimum numbers of sites, respectively. The dashed, horizontal lines represent the sampling error of the nests, here combined with the replication error to form the total reconstruction error. The latter error changes over time due to a declining number of old or relict tree material. From a total number of 642 (155) samples covering the 1921-1976 period, random draws produced stepwise changes in the total number of samples, from which the reconstruction error was recalculated. Monte Carlo simulations provide the mean standard error with respect to replication (upper bound of the green area) and the 95% confidence intervals (grey shades) derived from 1000 different draws. Dotted, vertical lines indicate the minimum replication actually found in the respective nests.

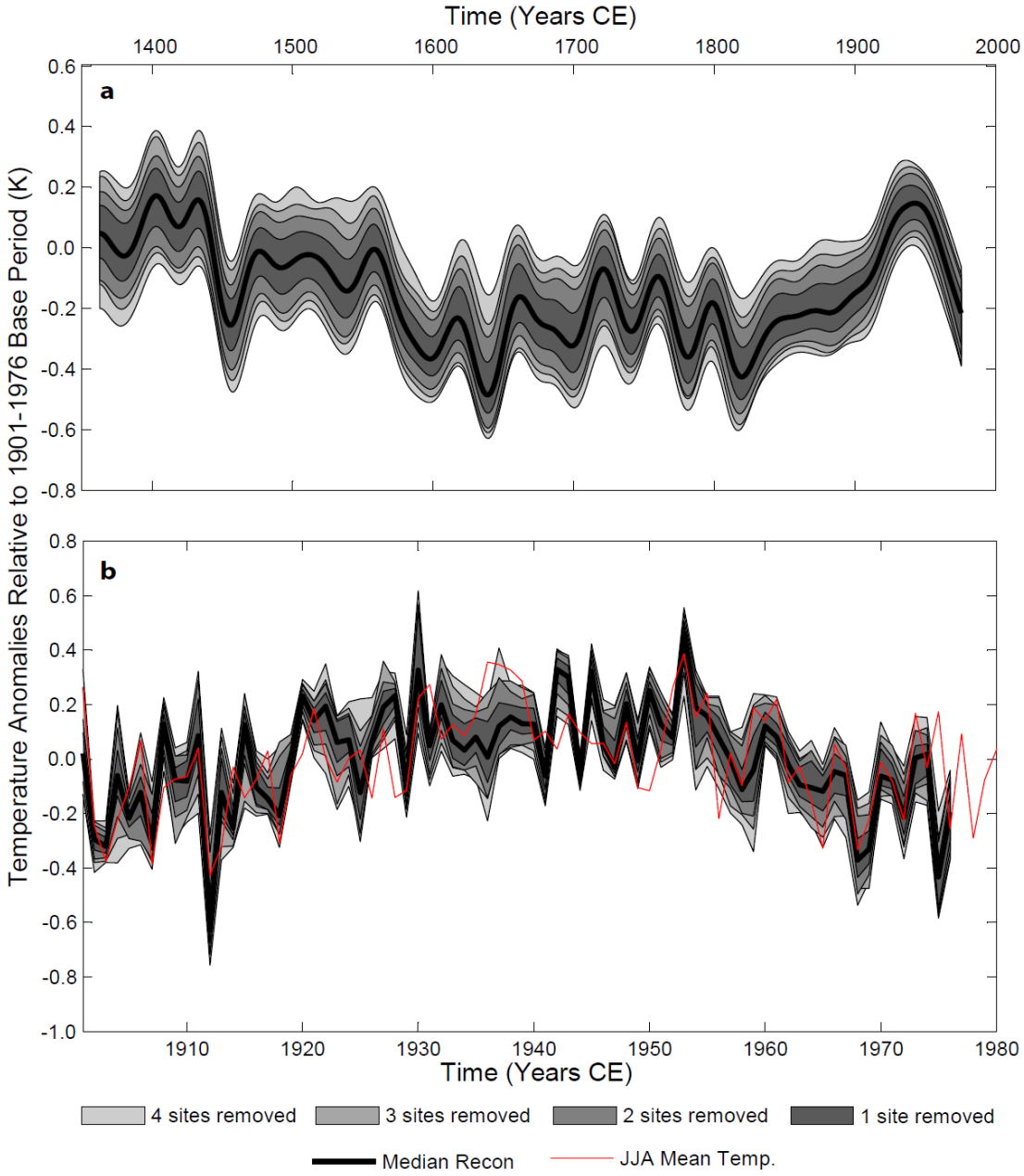


Figure S6-5. Jackknife tests. Uncertainty estimates for a truncated predictor network in the low (a) and high (b) frequency domain. In the upper panel records were smoothed with a 30-years spline. The upper and lower confidence intervals represent the maximum and minimum extent of alternative reconstructions based on a reduced number of site chronologies. For the removal of 2-4 chronologies the sites were randomly drawn and 100 realizations calculated.

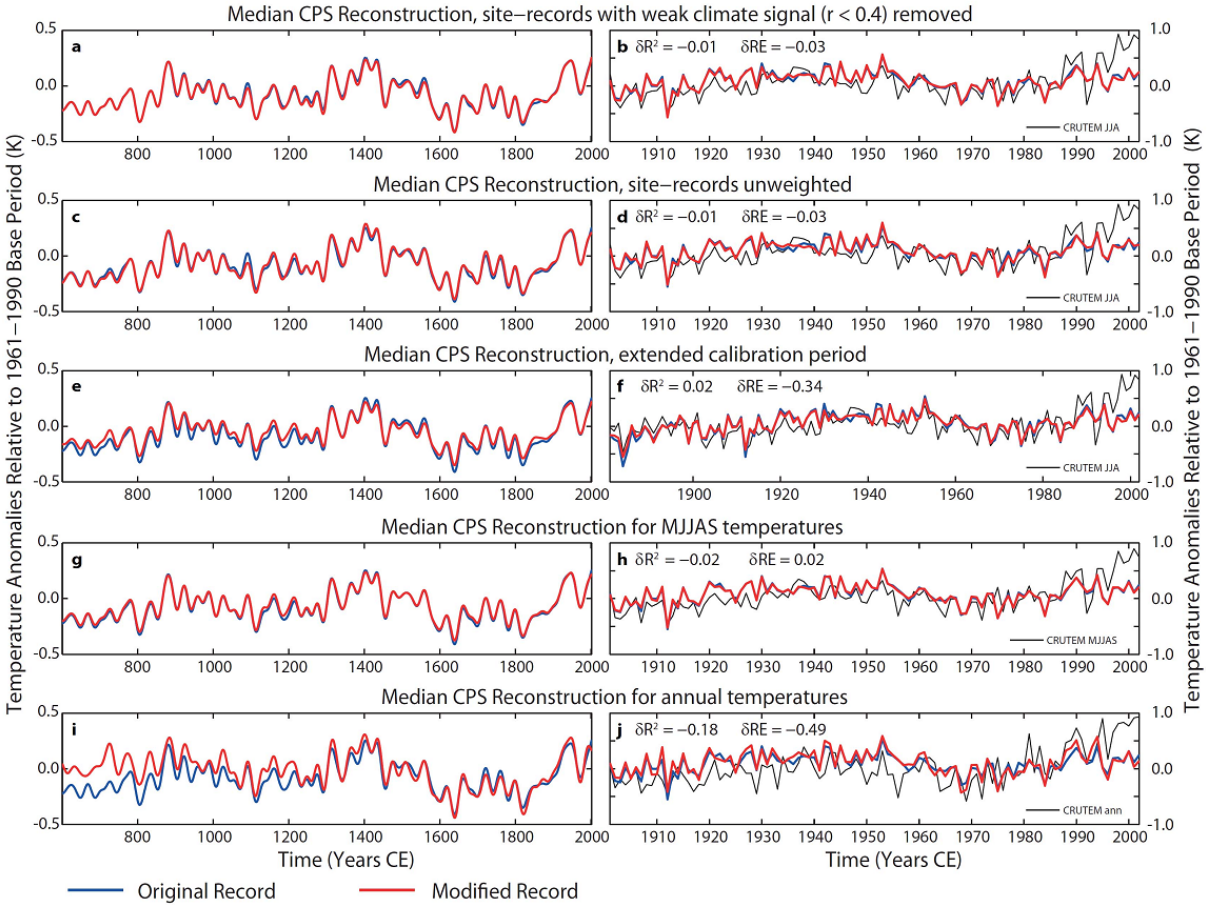


Figure S6-6. Alternative Calibrations. The original reconstruction was modified in multiple ways regarding the proxy network and the instrumental target. a and b, 3 site chronologies with local correlations of r < 0.4 were removed from the network. c and d, To warranty homogeneous spatial coverage proxy records were incorporated without weighting. e and f, The calibration period was extended back to 1881. g and h, The seasonal window was extended from May to September. i and j, Annual mean temperatures were used as a target. All data on the left hand side are smoothed with a 30-years spline, the right hand side shows the respective calibration period.

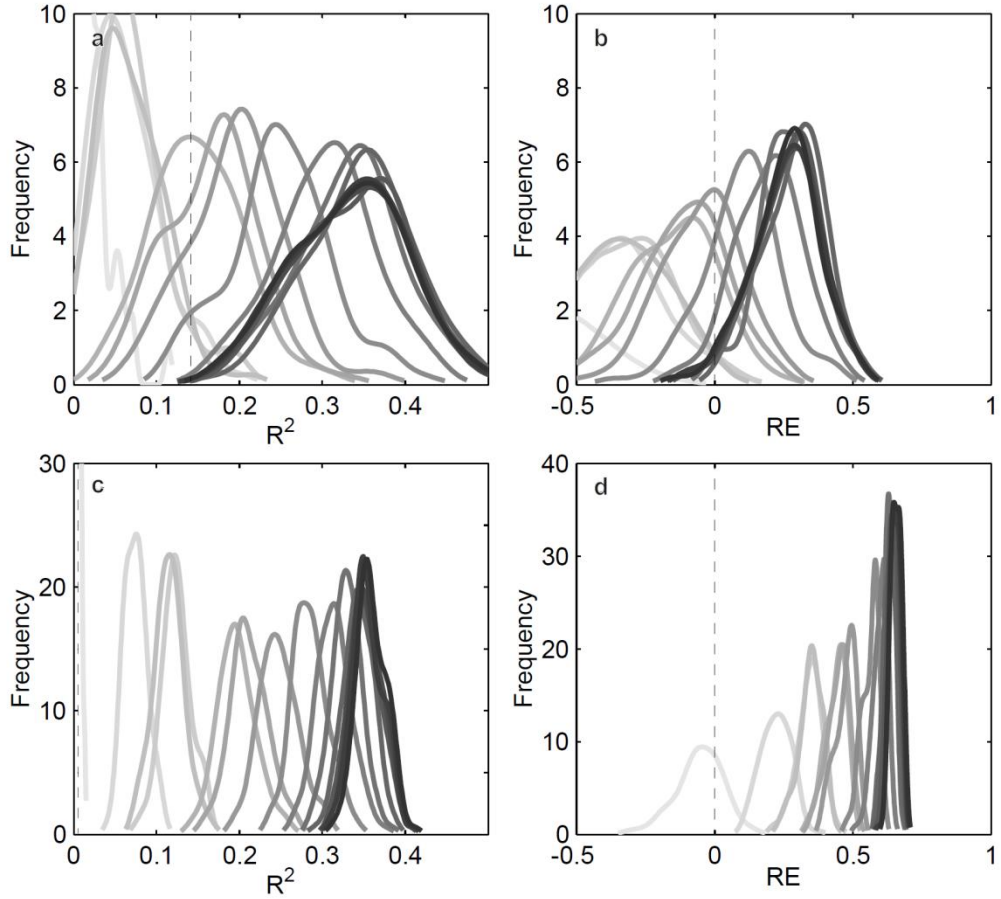


Figure S6-7. Validation metrics for pseudoproxy experiments. a and b, Nonparametric kernel-smoothing distributions of R^2 and RE of 100 different noise realizations using the sliding calibration and validation approach over the 1901-1976 period. Grey scales represent the 15 nests as in Fig. S2. c and d, Same as in a, but calculated over fixed and extended calibration and validation periods (1901-1976 and 850-1900). R^2 of nest #1 is ill-fitted in a and c due to its asymptotic approximation of zero. Dashed lines as in Fig. S2.

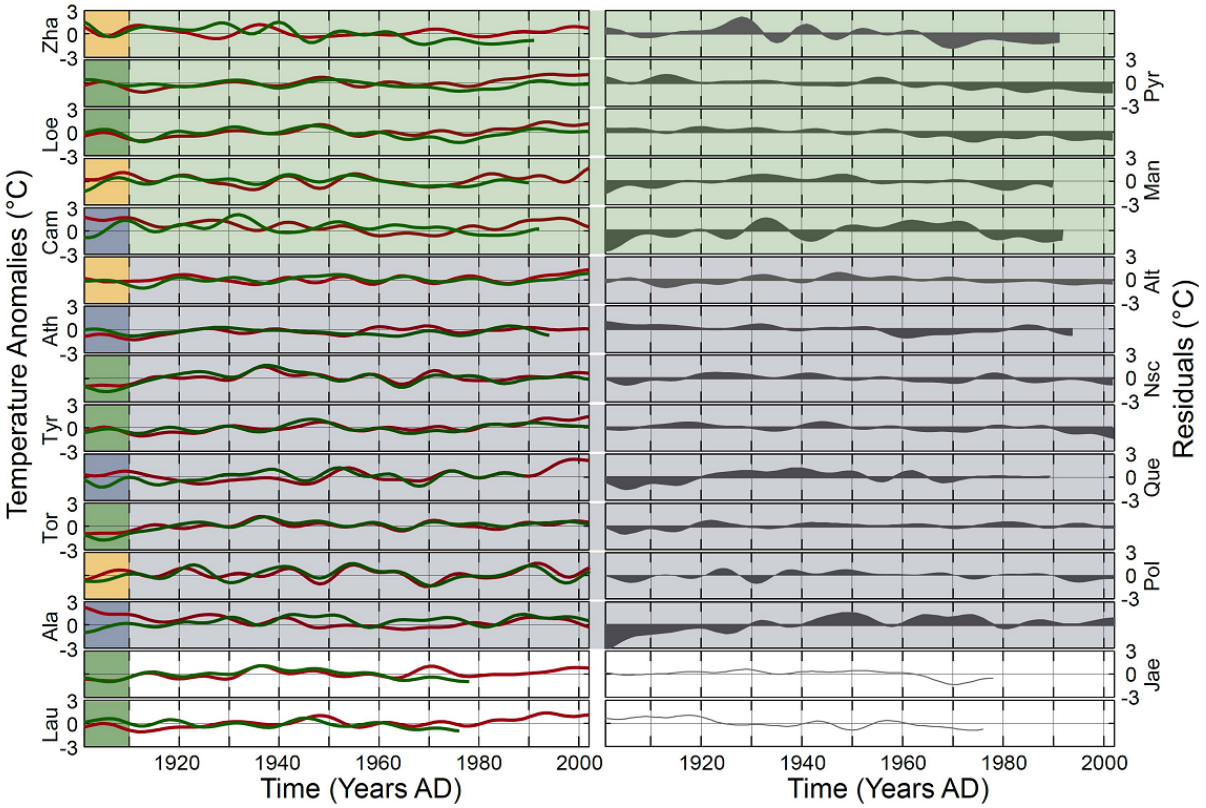


Figure S6-8. Local observational and modeled temperatures. Left panel shows the proxy-derived (green) and instrumental (red; CRUTEM4v) JJA temperatures at the 15 NH tree sites, all smoothed using a 10-year low-pass filter. Right panel shows the corresponding residuals emphasizing potential, late 20th century divergence issues. Sums of the residuals during the 1971-1990 period are decreasing from top to bottom, starting with largest late 20th century residuals in Zha. Referring to these sums a grouping in "diverging" (greenish background) and "non-diverging" (bluish background) proxy-records was established. The site-records from "Jae" and "Lau" are too short to be assigned to any of the groups. Green boxes on the left indicate sites from Europe, yellow from Asia, and blue from North America.

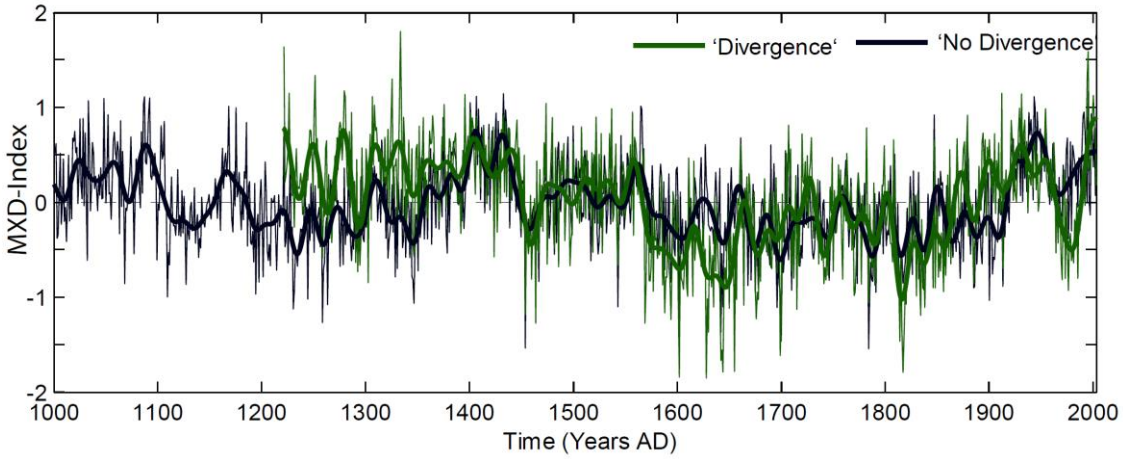


Figure S6-9. Mean records for potentially "diverging" and "non-diverging" MXD-chronologies (as classified in Fig. S6) smoothed with a 30-year low-pass filter. Records are truncated when based on less than 3 site-chronologies.

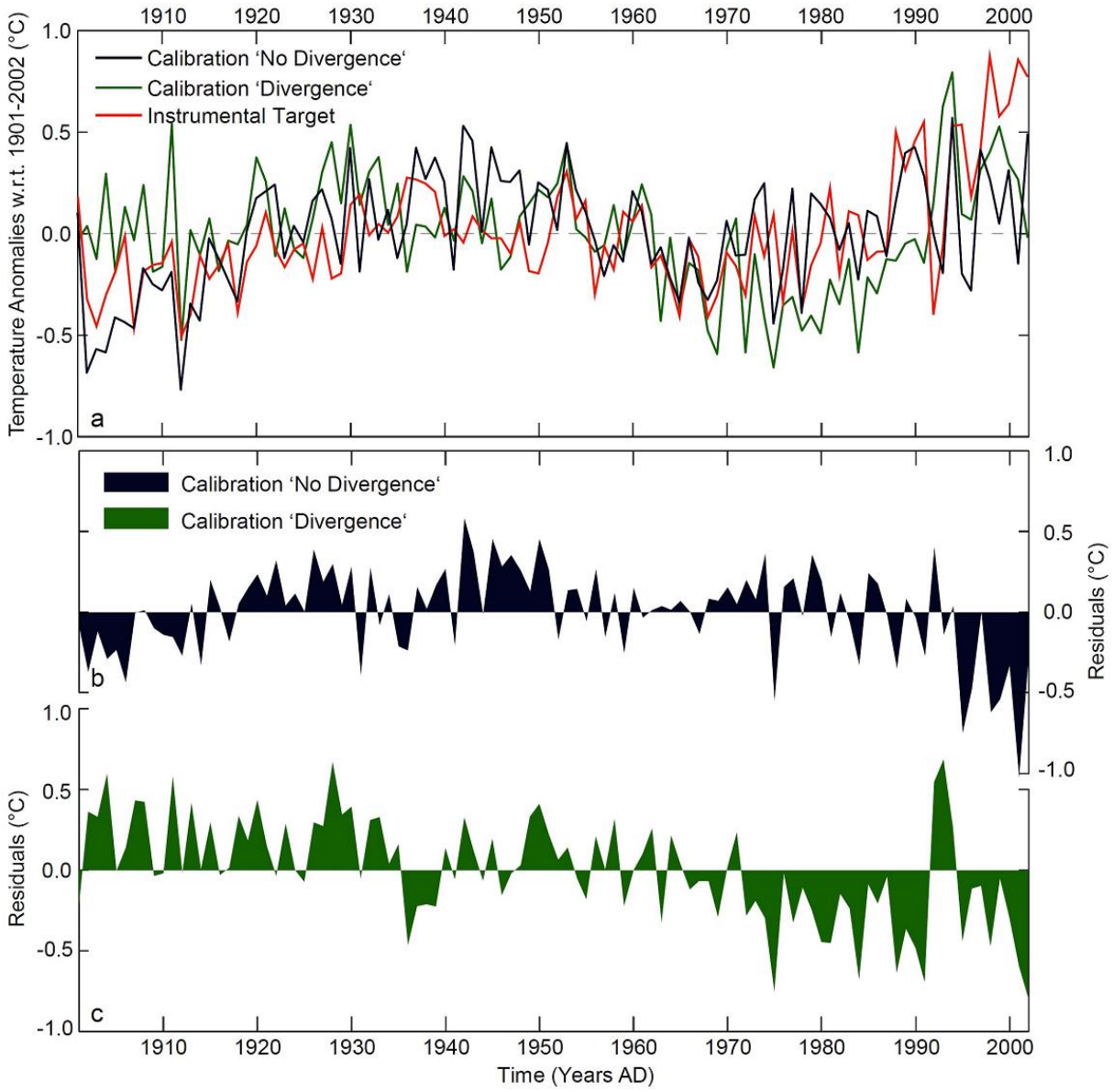


Figure S6-10. Reconstructions using mean curves of 5 "diverging" and 8 "non-diverging" records. a, Mean curves shown together with instrumental JJA temperatures from 1901-2002. b and c, Residuals from the observational temperature record are stronger in the "diverging" group, not only in the post 1970 period, but also during earlier periods. Warming temperatures at the turn of the millennium are not captured by the "diverging" or by the "non-diverging" data.

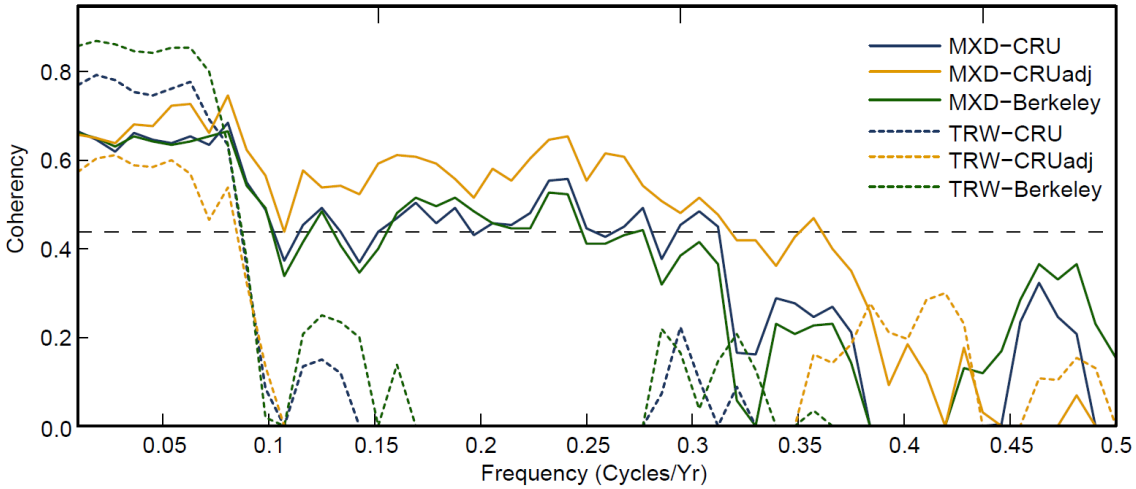


Figure S6-11. Coherency spectra for the 1881-1995 period. The multi-taper spectra are calculated using this reconstruction ("MXD") and the most recent tree-ring width-based reconstruction ("TRW", D'Arrigo et al. 2006) vs. the CRUTEM- and Berkeley-datasets as observational records [*Jones et al.*, 2012; *Rohde et al.*, 2013]. "CRU" and "Berkeley" refers to a simple JJA average of all grid-points between 30 and 90°N, whereas CRUadj includes an areal weighting and gapfilling as applied in this study.

References

Cook, E. R., and K. Peters (1997), Calculating unbiased tree-ring indices for the study of climatic and environmental change, *Holocene*, 7(3), 361-370.

Franke, J., D. Frank, C. C. Raible, J. Esper, and S. Brönnimann (2013), Spectral biases in tree-ring climate proxies, *Nat. Clim. Change*, *3*(4), 360-364.

Gunnarson, B. E., H. W. Linderholm, and A. Moberg (2011), Improving a tree-ring reconstruction from west-central Scandinavia: 900 years of warm-season temperatures, *Clim. Dyn.*, 36, 97-108.

Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice (2012), Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010, J. Geophys. Res.-Atmos., 117.

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, *Journal of Climate*, *26*(4), 1085-1111.

Masson-Delmotte, V., et al. (2013), Information from paleoclimate archives, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley, pp. 383–464, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Smerdon, J.E. (2012), Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments, WIREs Climate Change, 3:63-77, doi:10.1002/wcc.149.

Rohde et al. (2013), A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview 1:1*, http://dx.doi.org/10.4172/2327-4581.1000101.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of Cmip5 and the experiment design, *B. Am. Meteorol. Soc.*, *93*(4), 485-498.

Thomson, D. J. (1982), Spectrum estimation and harmonic-analysis, *P. Ieee.*, *70*(9), 1055-1096.

Appendix III: Supplementary material for chapter 7

7.5 Supplementary Material

7.5.1 Power for a known break

The DGP is:

$$y_t = \mu + \lambda d_t + \varepsilon_t \tag{7.35}$$

The break shifts μ to $\mu + \lambda d_t$ where d_t is a break function of length L beginning at time $t = T_1$ where $T_1 + L \leq T$ such that $d_t \neq 0$ for $T_1 \leq t < T_1 + L$ and 0 otherwise. The correctly specified model is:

$$y_t = \mu + \gamma d_t + \varepsilon_t \tag{7.36}$$

Expressions for the estimators $\widehat{\mu}, \widehat{\gamma}$ are given by:

$$\begin{aligned}
\begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\gamma} - \lambda \end{pmatrix} &= \begin{pmatrix} T & \sum_{t=T_1}^{T_1+L-1} d_t \\ \sum_{t=T_1}^{T_1+L-1} d_t & \sum_{t=T_1}^{T_1+L-1} d_t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} d_t \varepsilon_t \end{pmatrix} \\
&= T^{-1} \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \left(\sum_{t=T_1}^{T_1+L-1} d_t \right)^2 \right]^{-1} \times \\
&\quad \begin{pmatrix} \sum_{t=T_1}^{T_1+L-1} d_t^2 \sum_{t=1}^{T} \varepsilon_t - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=T_1}^{T_1+L-1} d_t \varepsilon_t \\ -\sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=t=1}^{T} \varepsilon_t + T \sum_{t=T_1}^{T_1+L-1} d_t \varepsilon_t \end{pmatrix} \\
&= \begin{pmatrix} T_d^{-1} \left(\sum_{t=T_1}^{T_1+L-1} d_t^2 \sum_{t=1}^{T} \varepsilon_t - \sum_{t=T_1}^{T_1+L-1} d_t \sum_{t=T_1}^{T_1+L-1} d_t \varepsilon_t \right) \\ T_d^{-1} \left(\sum_{t=T1}^{T_1+L-1} d_t \varepsilon_t - \sum_{t=T1}^{T_1+L-1} d_t \sum_{t=0}^{T} \varepsilon_t \right) \end{pmatrix}
\end{aligned} \tag{7.37}$$

where $T_d = T \left[\sum_{t=T_1}^{T_1+L-1} d_t^2 - \frac{1}{T} \left(\sum_{t=T_1}^{T_1+L-1} d_t \right)^2 \right]$.

7.5.2 Proof of equation (7.10)

Proof that $\widehat{\gamma}_{(1)} = \lambda \mathbf{r} + (\mathbf{D}_1' \mathbf{D}_1)^{-1} \mathbf{D}_1' \varepsilon$.

$$\begin{aligned}
\widehat{\gamma}_{(1)} &= \lambda \left(\mathbf{D}_1' \mathbf{D}_1 \right)^{-1} \mathbf{D}_1' \mathbf{d}_{T_1} + \left(\mathbf{D}_1' \mathbf{D}_1 \right)^{-1} \mathbf{D}_1' \varepsilon \\
&= \lambda \left(\mathbf{D}_1' \mathbf{D}_1 \right)^{-1} \left(\mathbf{D}_1' \mathbf{D}_1 \right) \mathbf{r} + \left(\mathbf{D}_1' \mathbf{D}_1 \right)^{-1} \mathbf{D}_1' \varepsilon \\
&= \lambda \mathbf{r} + \left(\mathbf{D}_1' \mathbf{D}_1 \right)^{-1} \mathbf{D}_1' \varepsilon
\end{aligned} \tag{7.38}$$

using:

$$\left(\mathbf{D}_1' \mathbf{D}_1 \right) \mathbf{r} = \mathbf{D}_1' \begin{pmatrix} \mathbf{d}_1 & \cdots & \mathbf{d}_{T_1} & \cdots & \mathbf{d}_{T/2} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1_{T_1} \\ \vdots \\ 0 \end{pmatrix} = \mathbf{D}_1' \mathbf{d}_{T_1}$$

Assuming that the retained indicators of the first half $\mathbf{D}_{1*}$ include the true break indicator, the same result applies to the estimator $\widehat{\gamma}_{(U)}$ over the retained indicators $D_U = [\mathbf{D}_{1*} \mathbf{D}_{2*}]$ where $\widehat{\gamma}_{(U)} = \lambda \mathbf{r} + (\mathbf{D}_U' \mathbf{D}_U)^{-1} \mathbf{D}_U' \varepsilon$.

7.5.3 Proof of generalization of step-functions for known break

Proof that $\widehat{\mu}-\mu=\bar{\varepsilon}_2$ and $\widehat{\gamma}-\lambda=\bar{\varepsilon}_1-\bar{\varepsilon}_2$ when d_t take the form of SIS-type step indicators and a single break from $t=0$ to T_1 is considered. Note, under these specifications $\sum_{t=T_1}^{T1+L} d_t^2=\sum_{t=T_1}^{T1+L} d_t=T_1$. Then

$$\widehat{\mu}-\mu=\frac{T_1\left(\sum_{t=1}^{T}\varepsilon_t-\sum_{t=1}^{T_1}\varepsilon_t\right)}{T\left(T_1-T_1^2/T\right)}=\frac{\sum_{t=T_1+1}^{T}\varepsilon_t}{T-T_1}=\bar{\varepsilon}_2 \tag{7.39}$$

and

$$\begin{aligned}
\widehat{\gamma}-\lambda &= \frac{T\sum_{t=1}^{T_1}\varepsilon_t-T_1\sum_{t=1}^{T}\varepsilon_t}{T_1(T-T_1)} && (7.40)\\
&= \frac{T\sum_{t=1}^{T_1}\varepsilon_t-T_1\left(\sum_{t=1}^{T_1}\varepsilon_t+\sum_{t=T_1+1}^{T}\varepsilon_t\right)}{T_1(T-T_1)} && (7.41)\\
&= \frac{T_1(T-T_1)\bar{\varepsilon}_1-T_1(T-T_1)\bar{\varepsilon}_2}{T_1(T-T_1)} && (7.42)\\
&= \bar{\varepsilon}_1-\bar{\varepsilon}_2 && (7.43)
\end{aligned}$$

using $\bar{\varepsilon}_1=1/T_1\sum_{t=1}^{T_1}\varepsilon_t$ and $\bar{\varepsilon}_2=1/(T-T_1)\sum_{t=T_1+1}^{T}\varepsilon_t$.

7.5.4 Algorithm Specification

Here we outline the split-half algorithm (*Algorithm 1*, see Figure 7.1) used for theory-derivations and simulations, and sketch the multi-path algorithm (*Algorithm 2*) used in simulations and in practice:

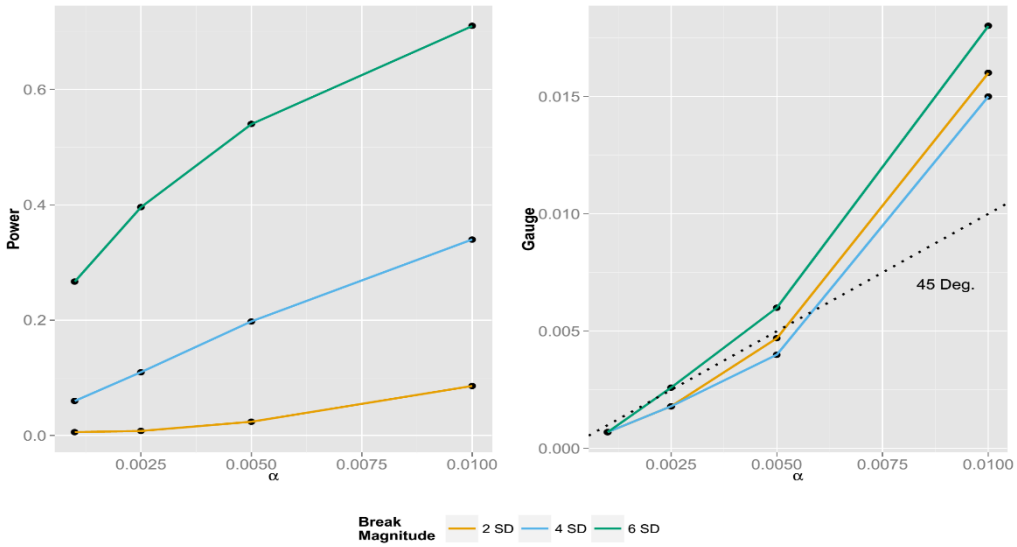
***Algorithm 1*: Split-Half Indicator Saturation for Designed Break Functions.** Choose a target level of significance α.

1. Design a break function and construct the $(T\times T)$ break matrix $\mathbf{D}$ using equation (7.4).
2. Split the break matrix $\mathbf{D}$ into two sets: $\mathbf{D}_1=(\mathbf{d}_1,...,\mathbf{d}_{T/2})$ and $\mathbf{D}_2=(\mathbf{d}_{T/2+1},...,\mathbf{d}_T)$.
3. Compute the least squares estimators $\widehat{\gamma}_{(1)}$ and $\widehat{\gamma}_{(2)}$ using (7.10) and (7.12) fixing an intercept (and additional covariates if required) in each regression.
4. Construct the union set of k retained break functions as $\mathbf{D}_U=(\mathbf{d}_j,...,\mathbf{d}_{j+k})$ where $\mathbf{d}_j$ is included if $|t_{\widehat{\gamma}_{i,j}}|\geq c_\alpha$ in step 3, where $i=1,2$ for each set.
5. Compute the least squares estimator $\widehat{\gamma}_{(U)}$ using (7.14), fixing an intercept (and additional covariates if required) in each regression. Retain final break functions j if $|t_{\widehat{\gamma}_{U,j}}|\geq c_\alpha$.

***Algorithm 2*: Multi-Path Indicator Saturation for Designed Break Functions.**

1. Design a break function and construct the $(T\times T)$ break matrix $\mathbf{D}$ using equation (7.4).
2. Use a mulit-path and multi-block algorithm: for the present simulations and application we rely on the parallel step-wise backwards tree-search algorithm *Autometrics* (Doornik 2009a) fixing an intercept (and additional covariates if required) in the general unrestricted model (7.2).
3. Retain the break functions from the terminal model chosen by the specified tiebreaker criterion.

Figure 7.11: Left: Potency of detecting a volcanic break of magnitude λ for level of significance α using multi-path selection for an autoregressive DGP and model. Right: Proportion of spuriously retained break indicators (gauge). Break magnitude λ corresponds to the full response over the entire break, the peak is 0.58λ.



7.5.5 Simulation results for a simple dynamic DGP

Here we present the results on power and false-positive rate when using a multi-path search and the DGP includes an autoregressive term of order one. The DGP is given by:

$$y_t = \rho y_{t-1} + \lambda d_t + \varepsilon_t \tag{7.44}$$

where $\varepsilon_t \sim \mathrm{IN}(0, \sigma_\varepsilon^2)$, $\sigma_\varepsilon^2 = 1$ and $\rho = 0.3$. The model includes an AR(1) term and a full set of break functions. Figure 7.6 and Table 7.6 provide the results. The null-retention (gauge) remains well calibrated to the nominal size, the power is slightly lower compared to the static case, likely due to the shape of the break mimicking an auto-regressive process and thereby being detected less frequently.

Table 7.6: Power of detecting an unknown break when multi-path searches in an auto-regressive DGP and model. Statistics were generated from 1000 simulations and detection significance was set to $\alpha = 0.01$, with a length of $L = 3$. Break magnitude λ corresponds to the full response over the entire break, the peak is 0.58λ, T=100.

	Multi-Path	
	Potency	Gauge
$\lambda = 6$, peak=3.48	0.71	0.018
$\lambda = 4$, peak=2.23	0.34	0.015
$\lambda = 2$, peak=1.16	0.09	0.016

7.5.6 Proof of null result when forced variables are included

We investigate the approximate distributional consequences of including a full set of irrelevant break variables into a model with assumed K relevant variables $\mathbf{X}$ with associated coefficients β. This is based on the results of Hendry & Johansen (2015). The DGP is:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{7.45}$$

where $\varepsilon \sim \mathrm{iid}(0, \sigma_\varepsilon^2 \mathbf{I})$. The first-half model is given as:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{D}_1 \gamma_{(1)} + \mathbf{v} \tag{7.46}$$

where the true $\gamma_{(1)} = \mathbf{0}$. Consider first orthogonalizing $\mathbf{X}$ and $\mathbf{D_1}$ by regressing each column of $\mathbf{D}_1$ on $\mathbf{X}$, this yields the $(K \times T/2)$ matrix of coefficients $\widehat{\Gamma}$:

$$\widehat{\Gamma} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}_1 \tag{7.47}$$

The $(T \times T/2)$ residuals $\widehat{\mathbf{u}}$ are defined through:

$$\mathbf{D}_1 = \mathbf{X}\widehat{\Gamma} + \widehat{\mathbf{u}} \tag{7.48}$$

where $\widehat{\mathbf{u}}'\mathbf{X} = \mathbf{0}$. Substituting for $\mathbf{D}_1$ in the model equation yields:

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\beta + \left(\mathbf{X}\widehat{\Gamma} + \widehat{\mathbf{u}}\right)\gamma_{(1)} + \mathbf{v} && (7.49)\\
&= \mathbf{X}\left(\beta + \widehat{\Gamma}\gamma_{(1)}\right) + \widehat{\mathbf{u}}\gamma_{(1)} + \mathbf{v} && (7.50)\\
&= \mathbf{X}\beta^* + \widehat{\mathbf{u}}\gamma_{(1)} + \mathbf{v} && (7.51)
\end{aligned}$$

where $\beta^* = \left(\beta + \widehat{\Gamma}\gamma_{(1)}\right) = \beta$ as $\gamma_{(1)} = 0$. Then:

$$\begin{aligned}
\begin{pmatrix} \widehat{\beta^*} - \beta \\ \widehat{\gamma}_{(1)} - \mathbf{0} \end{pmatrix} &= \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\widehat{\mathbf{u}} \\ \widehat{\mathbf{u}}'\mathbf{X} & \widehat{\mathbf{u}}'\widehat{\mathbf{u}} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\varepsilon \\ \widehat{\mathbf{u}}'\varepsilon \end{pmatrix} && (7.52)\\
&= \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \\ (\widehat{\mathbf{u}}'\widehat{\mathbf{u}})^{-1}\widehat{\mathbf{u}}'\varepsilon \end{pmatrix} && (7.53)
\end{aligned}$$

since $\widehat{\mathbf{u}}'\mathbf{X} = \mathbf{0}$. Under the assumption that $T^{-1}\mathbf{X}'\mathbf{X} \xrightarrow{P} \blacksquare_{XX}$, the asymptotic distribution is then:

$$\sqrt{T}\begin{pmatrix} \widehat{\beta^*} - \beta \\ \widehat{\gamma}_{(1)} - \mathbf{0} \end{pmatrix} \xrightarrow{D} N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma_\varepsilon^2 \begin{pmatrix} \blacksquare_{\mathbf{XX}}^{-1} & \mathbf{0} \\ \mathbf{0} & \blacksquare_{\mathbf{D}_1\mathbf{D}_1|\mathbf{X}}^{-1} \end{pmatrix}\right] \tag{7.54}$$

The distribution of the parameters β on the correct variables $\mathbf{X}$ is unaffected by the inclusion of the break indicators $\mathbf{D}_1$ when there is no break. The equivalent result holds when the second half of break indicators $\mathbf{D}_2$ is included. In practice the main risk is the spurious retention of break indicators, but this can be controlled through a conservative selection rule.

7.5.7 Model Response To Large Volcanic Eruption

The basic energy balance model in terms of deviations from the steady state is given by:

$$C\frac{dT'}{dt} = F(t) - \theta T' \tag{7.55}$$

The differential equation can be solved using an integrating factor. Re-arranging the above model yields:

$$\frac{dT'}{dt} + \frac{\theta}{C}T' = \frac{1}{C}F(t) \tag{7.56}$$

$$\frac{dT'}{dt} + \frac{1}{\tau}T' = \frac{1}{C}F(t) \tag{7.57}$$

Using the integrating factor $I(t) = e^{\frac{t}{\tau}}$ yields

$$\frac{d}{dt}T'e^{\frac{t}{\tau}} = \frac{F(t)}{C}e^{\frac{t}{\tau}} \tag{7.58}$$

$$T'e^{\frac{t}{\tau}} = \int_0^t \frac{F(u)}{C}e^{\frac{u}{\tau}}du \tag{7.59}$$

$$T' = \frac{1}{C}e^{\frac{-t}{\tau}}\int_0^t F(u)e^{\frac{u}{\tau}}du \tag{7.60}$$

This expression can be used to solve for the temperature response following a large volcanic eruption under the assumption that the volcanic eruption is approximated by a pulse forcing followed by exponential decay at rate $-1/\gamma$. Let $F(t)$ denote the volcanic forcing:

$$F(t) = 0 \text{ for } t \leq 0 \text{ and } F(t) = Fe^{\frac{-t}{\gamma}} \text{ for } t > 0 \tag{7.61}$$

Using this expression yields:

$$T' = \frac{1}{C}e^{\frac{-t}{\tau}}\int_0^t F(u)e^{\frac{u}{\tau}}du \tag{7.62}$$

$$T' = \frac{1}{C}e^{\frac{-\theta}{C}t}F\left(\frac{\theta}{C}-\frac{1}{\gamma}\right)^{-1}\left[e^{t\left(\frac{\theta}{C}-\frac{1}{\gamma}\right)}-1\right] \tag{7.63}$$

7.5.8 Simulation Results using a Full Sample Search

Table 7.7 provides simulation results when a full-sample rather than a sub-sample approach is used for function (a) with an intercept-only model. There is little difference relative to the sub-sampling method at the cost of increased computational time. Using a 3Ghz processor the sub-sample approach requires ≈ 5 seconds to cover the entire sample for one replication (across 10 subsamples), compared to ≈ 5 minutes for one replication using a full-sample approach, similar to the analysis in Doornik & Hendry (2015).

Table 7.7: Potency and Gauge for Volcanic Functions (a) using a full sample search

	T	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
Potency NH Tg$>$ 20	0.32	0.67	0.67	0.67
Potency NH Tg$>$ 0	0.22	0.53	0.57	0.58
Potency Global Tg$>$ 20	0.23	0.52	0.57	0.58
Potency Global Tg$>$ 0	0.22	0.53	0.57	0.58
Gauge NH	0.02			
Gauge Global	0.02			

7.5.9 Results for Volcanic Function(b)

Table 7.8: Potency and Gauge for Volcanic Functions (b)

	T	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
Potency NH Tg$>$ 20	0.60	0.71	0.72	0.72
Potency NH Tg$>$ 0	0.22	0.32	0.33	0.34
Potency Global Tg$>$ 20	0.40	0.55	0.56	0.57
Potency Global Tg$>$ 0	0.15	0.22	0.25	0.26
Gauge NH	0.02			
Gauge Global	0.02			

Table 7.9: Detection of Volcanic Eruptions $> 20Tg$ using volcanic functions (b)

NH Volcano	Tg	Potency $t = T$	$t = T \pm 1$	$t = T \pm 2$	$t = T \pm 3$
939	31.83	0	0	0.02	0.02
1167	29.535	0.01	0.02	0.04	0.04
1176	45.761	0.6	1	1	1
1227	58.644	0	0.01	0.03	0.03
1258	145.8	1	1	1	1
1284	23.053	0.76	0.82	0.82	0.82
1452	44.6	0.7	1	1	1
1459	21.925	0.81	1	1	1
1584	24.228	0.51	0.66	0.69	0.69
1600	46.077	0.99	1	1	1
1641	33.805	0.92	1	1	1
1719	31.483	0.94	0.98	0.98	0.98
1783	92.964	0	0.01	0.01	0.01
1809	27.558	0.9	0.98	0.98	0.98
1815	58.694	0.65	1	1	1
1835	26.356	0.77	0.9	0.98	0.98
Global Volcano	**Tg**	**Potency $t = T$**	**$t = T \pm 1$**	**$t = T \pm 2$**	**$t = T \pm 3$**
854	21.387	0	0.03	0.04	0.04
870	22.276	0	0.24	0.24	0.24
901	21.283	0	0.5	0.63	0.65
939	33.128	0	0	0.02	0.02
1001	21.011	0	0.34	0.34	0.34
1167	52.114	0.01	0.02	0.04	0.04
1176	45.761	0.6	1	1	1
1227	67.522	0	0.01	0.03	0.03
1258	257.91	1	1	1	1
1275	63.723	0	0.01	0.01	0.01
1284	54.698	0.76	0.82	0.82	0.82
1341	31.136	0	0.01	0.02	0.03
1452	137.5	0.7	1	1	1
1459	21.925	0.81	1	1	1
1584	24.228	0.51	0.66	0.69	0.69
1600	56.591	0.99	1	1	1
1641	51.594	0.92	1	1	1
1693	27.098	0	0.01	0.03	0.11
1719	31.483	0.94	0.98	0.98	0.98
1783	92.964	0	0.01	0.01	0.01
1809	53.74	0.9	0.98	0.98	0.98
1815	109.72	0.65	1	1	1
1835	40.16	0.77	0.9	0.98	0.98
1883	21.864	0	0.95	0.95	0.95
1963	20.87	0	0.47	0.48	0.48
1991	30.094	0	0.38	0.38	0.38

Appendix IV: Supplementary material for chapter 8

Supplementary Information (SI)

1) Computation of radiative forcing estimates

G08 is accompanied by a dataset with "monthly and spatially dependent stratospheric loading from volcanic eruptions". C13 is available as annually averaged global aerosol optical depth estimates. To conform to the format of S15 all records had to be converted into radiative forcing estimates. First, G08 was transferred in aerosol optical depth. Summing up the altitudinal layers yields column mass densities of sulfur in kg/m^2. A global area-weighted average mass density is then multiplied by the surface of the Earth. Using the scaling suggested by Stother et al. (1984), we divided the total sulfur mass by 150Tg to retrieve global aerosol optical depth estimates. C13 and G08 were then multiplied with -20 W/m^2 which yields the radiative forcing in W/m^2.

2) Reported problems with the volcanic radiative forcing in climate models

For GISS-E2-R and IPSL-CM5A-LR problems were reported with the implementation of the volcanic forcing. In IPSL-CM5A-LR a daily aerosol drift likely reduced the strength of the forcing (Atwood et al. 2016). The misspecification of GISS-E2-R in combination with G08 (not used here, Schmidt et al. 2014) resulted in a G08 forcing roughly twice as strong as the input quantity

Supplementary Table S8-1a | Detection results for negative breaks in various records. Detected breaks sorted by the estimated break coefficient (highest on top) and scaled by the value of the break function*. Truncated for CCSM and GISS. The first 10 events (8 for BCC) of each record were used for the superposed epoch analyses. Colors indicate that these events are also among the top 10 in one of the forcing records. The three reconstructions were compared with S15, BCC and CCSM with G08, GISS and MPI with C12.

	SCH15	STO15	WIL16	BCC	CCSM	GISS	IPSL	MPI
1	1453	1258	1641	1857	1762	1258	1258	1258
2	1601	1816	1601	1884	1259	1809	1816	1600
3	1258	1641	1453	1600	1258	1600	1177	1816
4	1783	1453	1698	1815	1600	1286	1309	1641
5	1641	1601	1258	1903	1641	1641	1287	1460
6	1109	1783	1783	1762	1177	1832	1232	1286
7	1698	1695	1816	1258	1213	1228	1452	1228
8	1816	1032	1832	1259	1453	1816	1601	1668
9	1884	1836	1230		1992	1815	1102	1809
10	1230	1109	1108		1816	1585	1884	1695
11	1912	1170	1169		1719	1835	1168	1806
12	1585	1698	1695		1835	1983	1810	1832
13	1345	1585	1835		1815	1694	1748	1584
14	1099	1740	1812		1809	1018	1629	1652
15		1884	1912		1884	1696	1690	1835
16		1667			1900	1041	1963	1759
17		1466			1214	1442	1145	1225
18		1912			1929	1674		1442
19		1179			1452	1460		
20		1004*			1460…	1331…		

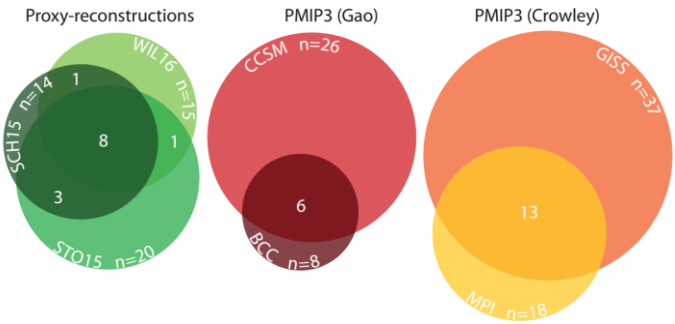
* The lowest break in STO15 (1004) has a high coefficient of -8.5°C, but occurs in the third year of a volcanic break detected in 1002. For a more accurate comparison with other events, the coefficients are scaled by the volcanic function, which is relatively low in the third year. The high coefficient is a remnant of a technical insufficiency. Due to the applied autoregressive model of third order, the temperature cannot be modelled before 1004. Due to this lag a volcanic break function in 1002 acts as a single-year impulse in 1004 (third year of the volcanic function). This seems to be the better approximation for this break than the volcanic function in 1004. Running the algorithm with impulse indicators instead of volcanic functions yields a strong (rank 8) break in 1004.

Supplementary Table S8-1b | Forcing events from ice-core sulfur reconstructions. Timing of the ten strongest time-integrated events of annually averaged global radiative forcing with volcanic source. Although there exists an update for the G08-record, these years refer to the original version which was used to force the model simulations.

	1	2	3	4	5	6	7	8	9	10
G08	1258	1452	1815	1762	1275	1213	1284	1809	1600	1641
C13	1258	1816	1456	1229	1696	1809	1286	1600	1641	1884
S15	1258	1458	1815	1230	1783	1809	1108	1641	1601	1171

Supplementary Table S8-1c | Detection results for positive breaks in various records. Detected breaks sorted by the estimated break coefficient (highest on top). Truncated for CCSM.

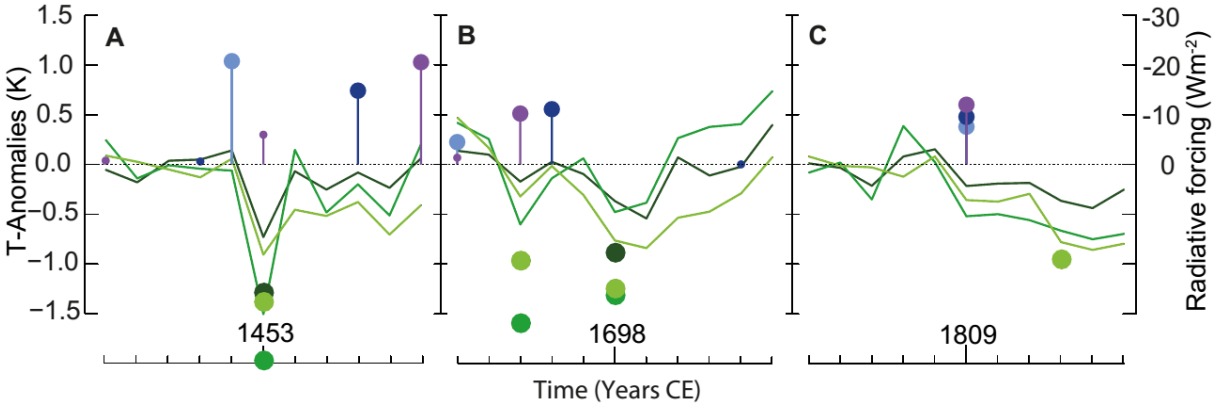
	SCH15	STO15	WIL16	BCC	CCSM	GISS	IPSL	MPI
1		1656	1033	1801	1972	2003	1892	2001
2		1167	1425	1869	1980	1039	1307	1999
3		1478	1499	1998	1602	1586	1818	1978
4		1058	1168	1996	1920	1019	1986	1980
5			2003	1936	1917	1173	1454	1916
6				1933	2003	1999	1859	1179
7				1999	1984	1833	1165	1987
8				1943	1991	1949	1998	1805
9				1971	1993	1989	1949	1158
10				2005	1899	1997	1970	2003
11				1980	1886	2001	1996	1940
12				2003	1997	1984	1310	1651
13				1994	1836	1260	1288	1461
14				1885	1215		1179	2005
15				1859	1817		1749	1052
16				2001	1988		1885	1996
17				1989	1928		1817	
18				1985	1178		1234	
19					2000		1260	
20					1994…			



Supplementary Figure S8-1 | Number of break functions detected in last millennium temperature records. The size of the circles represents the total number of detected breaks and the area of overlap the number of overlapping break dates. Among the breaks in the proxy-reconstructions 8 events are detected in all of the three records. Overlap between the three groups is not shown.

Supplementary Table S8-2 | Late 17th century volcanism as presented by the Global Volcanism Program. All volcanic eruptions between 1690 and 1707 with VEI > 4, or VEI=4 for northern hemispheric and tropical events from "Volcanoes of the World 4.5.0." (downloaded on 27 Jul 2016).

Volcano Name	Latitude	Longitude	Start Year	Start Month	VEI	Evidence Method (dating)
Chikurachki	50.324	155.461	1690+-10	0	4	Tephrochronology
Serua	-6.312	130.017	1693	6	4?	Historical Observations
Hekla	63.983	-19.666	1693	2	4	Historical Observations
Hokkaido-Komagatake	42.063	140.677	1694	7	4	Historical Observations
Fujisan	35.361	138.728	1707	12	5	Historical Observations



Supplementary Figure S8-2 | Examples for discrepancies between ice-core and tree-ring records. (a) and (b) 11-years windows for periods with a distinct mismatch between forcing records and detected breaks. (c) 11-years window around the strongest forcing event without a break detected in the neighboring 3 years. Temperature reconstructions (SCH15 dark green; STO15 green; WIL16 light green) are presented as anomalies w.r.t 5 pre-event years. Dots indicate the break coefficients of the respective breaks in the underlying temperature records. Forcing events in G08 (light blue), C13 (dark blue) and S15 (purple) are plotted inverse with stems, whereas the center of the dots represents the forcing estimate. Bigger dots are assigned to the most prominent forcing event in the respective period.

Supplementary Table S8-3 | Results for superposed epoch analyses with different event-years. The first row gives the average standard deviation of running 11-years windows over the past millennium, multiplied with -4. The results for the SEAs are based on the ten strongest events either chosen from the detected breaks (break SEA) or from the forcing records (forcing SEA). For the forcing SEAs the following records were chosen: S15 for the temperature reconstructions (SCH15, STO15, WIL16), G08 for the simulations BCC and CCSM, and C13 for the simulations GISS and MPI. The 10 events for each SEA are listed in Table S1a and b.

	SCH15	STO15	WIL16	BCC	CCSM	GISS	IPSL	MPI
Mean SD of running 11-yrs blocks * - 4	-0.64	-1.19	-0.77	-0.86	-1.12	-0.83	-0.86	-0.99
Cooling in year 0 (break SEA) [°C]	-0.60	-1.23	-0.76	-0.98	-2.37	-1.34	-1.12	-1.34
Cooling in year 1 (break SEA) [°C]	-0.32	-0.58	-0.59	-0.61	-1.75	-1.06	-0.70	-0.93
Cooling in year 0 (forcing SEA) [°C]	-0.38	-0.66	-0.55	-0.38	-1.50	-1.23		-1.08

References

Atwood, A. R., Wu, E., Frierson, D. M. W., Battisti, D. S. and Sachs, J. P. (2016) 'Quantifying Climate Forcings and Feedbacks over the Last Millennium in the CMIP5-PMIP3 Models', *Journal of Climate*, 29(3), 1161-1178.

Schmidt, G. A., Annan, J. D., Bartlein, P. J., Cook, B. I., Guilyardi, E., Hargreaves, J. C., Harrison, S. P., Kageyama, M., LeGrande, A. N., Konecky, B., Lovejoy, S., Mann, M. E., Masson-Delmotte, V., Risi, C., Thompson, D., Timmermann, A., Tremblay, L. B. and Yiou, P. (2014) 'Using palaeo-climate comparisons to constrain future projections in CMIP5', *Climate of the Past*, 10(1), 221-250.

CURRICULUM VITAE

Not available online.