



This is a repository copy of *Unknown eruption source parameters cause large uncertainty in historical volcanic radiative forcing reconstructions.*

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/182661/>

Version: Published Version

---

**Article:**

Marshall, L.R., Schmidt, A., Johnson, J.S. [orcid.org/0000-0002-4587-6722](https://orcid.org/0000-0002-4587-6722) et al. (4 more authors) (2021) Unknown eruption source parameters cause large uncertainty in historical volcanic radiative forcing reconstructions. *Journal of Geophysical Research: Atmospheres*, 126 (13). e2020JD033578. ISSN 2169-897X

<https://doi.org/10.1029/2020jd033578>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>



## RESEARCH ARTICLE

10.1029/2020JD033578

### Key Points:

- We use an aerosol-climate model and statistical emulation to find eruption source parameters from eight bipolar sulfate deposition signals
- Plausible eruptions have a wide range in sulfur dioxide emissions and latitude, and consequently, a large range in radiative forcing
- Resulting uncertainties in global mean surface cooling for the largest unidentified historical eruptions are on the order of 1°C

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

L. R. Marshall,  
[lrn49@cam.ac.uk](mailto:lrn49@cam.ac.uk)

### Citation:

Marshall, L. R., Schmidt, A., Johnson, J. S., Mann, G. W., Lee, L. A., Rigby, R., & Carslaw, K. S. (2021). Unknown eruption source parameters cause large uncertainty in historical volcanic radiative forcing reconstructions. *Journal of Geophysical Research: Atmospheres*, 126, e2020JD033578. <https://doi.org/10.1029/2020JD033578>

Received 22 JUL 2020  
 Accepted 30 MAY 2021







### Author Contributions:

**Conceptualization:** Lauren R. Marshall, Anja Schmidt, Ken S. Carslaw  
**Formal analysis:** Lauren R. Marshall, Jill S. Johnson, Lindsay A. Lee  
**Investigation:** Lauren R. Marshall, Jill S. Johnson  
**Methodology:** Lauren R. Marshall, Anja Schmidt, Jill S. Johnson, Graham W. Mann, Ken S. Carslaw  
**Supervision:** Anja Schmidt, Graham W. Mann, Ken S. Carslaw

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

# Unknown Eruption Source Parameters Cause Large Uncertainty in Historical Volcanic Radiative Forcing Reconstructions

Lauren R. Marshall<sup>1,2</sup> , Anja Schmidt<sup>2,3</sup> , Jill S. Johnson<sup>1</sup>, Graham W. Mann<sup>1,4</sup> , Lindsay A. Lee<sup>1,5</sup> , Richard Rigby<sup>6</sup> , and Ken S. Carslaw<sup>1</sup> 

<sup>1</sup>School of Earth and Environment, University of Leeds, Leeds, UK, <sup>2</sup>Department of Chemistry, University of Cambridge, Cambridge, UK, <sup>3</sup>Department of Geography, University of Cambridge, Cambridge, UK, <sup>4</sup>National Centre for Atmospheric Science, University of Leeds, Leeds, UK, <sup>5</sup>Department of Engineering and Mathematics, Sheffield Hallam University, Sheffield, UK, <sup>6</sup>Centre for Environmental Modelling and Computation, University of Leeds, Leeds, UK

**Abstract** Reconstructions of volcanic aerosol radiative forcing are required to understand past climate variability. Currently, reconstructions of pre-20th century volcanic forcing are derived from sulfate concentrations measured in polar ice cores, mainly using a relationship between the average ice-sheet sulfate deposition and stratospheric sulfate aerosol burden based on a single explosive eruption—the 1991 eruption of Mt. Pinatubo. Here we estimate volcanic radiative forcings and associated uncertainty ranges from ice-core sulfate records of eight of the largest bipolar deposition signals in the last 2,500 years using statistical emulation of a perturbed parameter ensemble of aerosol-climate model simulations of explosive eruptions. Extensive sampling of different combinations of eruption source parameters using the emulators reveals that a very wide range of eruptions in different seasons with different sulfur dioxide emissions, eruption latitudes, and emission altitudes produce ice-sheet sulfate deposition consistent with ice-core records. Consequently, we find a large range in the volcanic forcing that can be directly attributed to the unknown eruption source parameters. We estimate that the uncertainty in volcanic forcing caused by many plausible eruption realizations leads to uncertainties in the global mean surface cooling of around 1°C for the largest unidentified historical eruptions. Our emulators are available online (<https://cemac.github.io/volcanic-forcing-deposition>) where eruption realizations for given ice-sheet sulfate depositions can be explored.

**Plain Language Summary** Explosive volcanic eruptions inject sulfur into the atmosphere, forming sulfate particles that reflect sunlight and cause surface cooling. To understand past climate variability it is important to quantify the climate effects of historical eruptions. However, there are no direct observations of the sulfur emissions and cooling from eruptions that occurred hundreds of years ago, and volcanic climate effects must be inferred from indirect measurements. Volcanic sulfate is deposited onto Earth's surface and can be found in layers of polar ice. Ice-core sulfate records are used to identify the occurrence of past eruptions, and the size of the sulfate signals is used to estimate the amount of sulfate that was present in the atmosphere and consequently the climatic effect of the eruption. We explored the uncertainty in this calculation by using climate model simulations and novel statistics to infer the properties of eruptions that could lead to sulfate deposition that matches that from ice-core records. We found there are many eruption scenarios that could lead to the same polar sulfate deposition but have different resulting magnitudes of surface cooling. Consequently, there is a large uncertainty in the climatic effect of historical eruptions, which affects our understanding of past climate variability.

## 1. Introduction

Explosive volcanic eruptions that inject large amounts of sulfur dioxide (SO<sub>2</sub>) into the stratosphere are major drivers of natural climate variability on multi-annual to decadal timescales (Myhre et al., 2013). The SO<sub>2</sub> is converted to sulfate aerosol, which causes a radiative perturbation, or forcing, by scattering incoming shortwave radiation as well as absorbing near-infrared and infrared longwave radiation and leads to surface and tropospheric cooling. In order to understand and attribute climate variability on millennial timescales,

**Visualization:** Lauren R. Marshall,  
Richard Rigby  
**Writing – original draft:** Lauren R.  
Marshall  
**Writing – review & editing:** Anja  
Schmidt, Jill S. Johnson, Ken S. Carslaw

reconstructions of volcanic aerosol radiative forcing (RF) are therefore required (Crowley et al., 2008; Hegerl et al., 2007; Sigl et al., 2015). The majority of climate models also rely on forcing reconstructions as opposed to volcanic SO<sub>2</sub> emissions (Jungclaus et al., 2017; Schmidt et al., 2011). Accurate reconstructions of all climate forcing agents and an assessment of their uncertainties are vital to understand and evaluate past temperature changes on global and regional scales, to assess ocean heat uptake, which is critical for estimating future sea-level rise (Gregory et al., 2013), and ultimately to compare natural and anthropogenic drivers of climate variability.

Reconstructions of volcanic RF are uncertain because in-situ and remote-sensing measurements of eruptions do not exist before the 1963 eruption of Mt. Agung. When available in the modern era, SO<sub>2</sub> emissions and stratospheric aerosol optical depth (SAOD) are derived from satellite retrievals (since ~1979) (Carn et al., 2016) and ground-based optical measurements (since ~1800s) (Sato et al., 1993; Stothers, 1996, 2001). When direct measurements are not available, volcanic forcing data sets are constructed based on sulfate concentration anomalies measured in polar ice cores. Alternatively, the injected mass of sulfur can be estimated from petrological and geochemical studies of eruption deposits (e.g., Devine et al., 1984; Metzner et al., 2014; Scaillet et al., 2003; Self et al., 2004; Vidal et al., 2016).

Sulfate measured in ice cores provides a record of volcanism with high temporal resolution over thousands of years (Sigl et al., 2014, 2015). However, several assumptions must be made to translate the record of sulfate anomalies into a record of RF. Established methods include using transfer functions to estimate hemispheric stratospheric sulfate aerosol burdens (i.e., the total mass of volcanic sulfate aerosol in the stratosphere of each hemisphere) (Gao et al., 2008; Sigl et al., 2015; Toohey & Sigl, 2017) or SAOD (Crowley & Unterman, 2013) from estimates of the average amount of sulfate deposited on each ice sheet. Transfer functions are derived from the ice-sheet deposition averages and observed stratospheric sulfate burden or SAOD following the 1991 eruption of Mt. Pinatubo, from estimates of radioactive material in the stratosphere and measured in ice cores following nuclear weapons testing in the 1950s and 1960s (Clausen & Hammer, 1988), or from climate model simulations (Gao et al., 2007).

Several factors can affect the relationship between ice-sheet sulfate deposition and stratospheric sulfate burdens and SAOD; therefore it is unlikely that these transfer functions are universally appropriate for eruptions other than 1991 Mt. Pinatubo. For example, sulfate deposition is modulated by the season of eruption (Gao et al., 2008; Toohey et al., 2013; Toohey, Kruger, et al., 2016), atmospheric variability (Robock & Free, 1995; Toohey et al., 2013), and the magnitude of the injection, which can alter atmospheric circulation (Toohey et al., 2013). In addition, the RF efficiency per unit mass of emitted sulfur falls with the magnitude of SO<sub>2</sub> emission due to formation of larger sulfate aerosol particles (Pinto et al., 1989; Timmreck et al., 2010). The sulfate formation (e.g., size and amount) is also impacted by coincident emissions of other species such as water and halogens (e.g., LeGrande et al., 2016; Staunton-Sykes et al., 2021). Further uncertainties are associated with the conversion of estimated sulfate aerosol burdens into SAOD and RF. For example, for the reconstruction of Gao et al. (2008), a linear scaling is applied between the sulfate aerosol burden and SAOD (Stothers, 1984). Other reconstructions scale ice-sheet sulfate deposition to SAOD based on this relationship after 1991 Mt. Pinatubo but attempt to account for changes in particle size for eruptions with larger SO<sub>2</sub> emissions by applying an idealized 2/3 power scaling (Crowley & Unterman, 2013; Toohey & Sigl, 2017). SAOD can then be converted to RF using further conversion factors based on climate model simulations of the 1991 eruption of Mt. Pinatubo (Hansen et al., 2005). However, the relationship between SAOD and RF is uncertain because it depends on several factors that vary from eruption to eruption such as the insolation, cloudiness and surface albedo (Andersson et al., 2015) and consequently the latitude and season of an eruption (Marshall et al., 2020), aerosol-cloud interactions (Gregory et al., 2016; Larson & Portman, 2016; Marshall et al., 2020; Schmidt et al., 2018), the aerosol particle size distribution (Lacis et al., 1992), and the model configuration and radiative transfer schemes used to simulate eruptions (Hansen et al., 2005; Myhre et al., 2013; Wigley et al., 2005).

Simulations of the last millennium as part of the Paleoclimate Modelling Intercomparison Project phase 4 (PMIP4past2k) use the prescribed SAOD timeseries “EVA(2k)” derived by Toohey and Sigl (2017), where the RF is calculated internally by each model (Jungclaus et al., 2017). The spatial and temporal evolution of the prescribed SAOD is based on a simple parameterized transport model, the Easy Volcanic Aerosol (EVA) forcing generator (Toohey, Stevens, et al., 2016), which uses stratospheric SO<sub>2</sub> emissions from the

eVolv2k reconstruction (Toohey & Sigl, 2017). The EVA model is calibrated against the measured evolution of stratospheric aerosol following the 1991 Mt. Pinatubo eruption and does not explicitly account for many microphysical, chemical and dynamical processes that occur following a volcanic eruption. Because many eruptions identified in ice-core sulfate records are unattributed, the eruption season and latitude must be estimated or arbitrarily assigned, which introduces further uncertainty because these factors affect the formation and transport of stratospheric sulfate aerosol and its deposition (Gao et al., 2008; Toohey et al., 2011, 2013). Eruptions are assumed to be tropical if simultaneous sulfate signals occur in both Antarctica and Greenland (i.e., bipolar deposition signals) and are assigned to January if the season is unknown.

Overall, the difficulty with any reconstruction of RF is that it does not scale directly with the ice-sheet deposited sulfate. The magnitude of the forcing (integrated over time) depends on the global spread of the volcanic aerosol in the stratosphere, and the lifetime and microphysical properties of the aerosol (i.e., size, mass, and number of the aerosol particles). All of these depend on the emission strength, the altitude and latitude of emission and the eruption season, so-called “eruption source parameters” (Marshall et al., 2019). Consequently, for any observed ice-sheet volcanic sulfate deposition there is potentially a very wide range of equally plausible “eruption realizations”—that is, eruptions with different combinations of SO<sub>2</sub> emission, eruption latitude, emission altitude, and eruption season that produce ice-sheet sulfate within the uncertainty of the measurements. These multiple eruption realizations will have a wide range of associated forcings.

Current estimates of volcanic forcing uncertainty account for multiple factors, but do not directly account for the effect of multiple plausible eruption realizations. Toohey and Sigl (2017) estimated the uncertainty on SO<sub>2</sub> emissions greater than 20 Tg derived from ice-core sulfate composites to be around 15%–30% (1 standard deviation [SD]) by considering uncertainties in the ice-core composites themselves and in the transfer functions. Hegerl et al. (2006) (see their supplementary material) estimated an overall uncertainty in past RF of 35% (1 SD) and an uncertainty of around 50% for individual volcanic eruptions (1 SD) based on uncertainties in the transfer function between ice-core sulfate and SAOD, in converting SAOD to RF, in the ice-core averaging and the variability of sulfur concentrations between ice cores. However, the possible range in RF has not yet been quantified by considering the range in plausible eruption realizations, rather than estimates of uncertainty derived from the calculation of the reconstructions themselves. Previous sensitivity studies investigating the relationships between eruption source parameters and sulfate deposition have also been based on specific case-studies (Toohey, Kruger, et al., 2016) or at single latitudes (Toohey et al., 2013).

Here, we investigate and quantify comprehensively the uncertainty in volcanic aerosol RF derived from ice-core sulfate records that is introduced by not knowing anything about the eruption other than the sulfate deposition. Using a state-of-the-art aerosol-climate model, we simulate a wide range of large-magnitude explosive eruptions and use the results to build statistical emulators (Marshall et al., 2019) that describe how sulfate deposition and time-integrated RF vary with eruption magnitude, eruption latitude, injection height, and eruption season. The emulators enable us to predict the sulfate deposition and RF for thousands of eruptions that we did not simulate directly. We determine the combinations of eruption source parameters that produce ice-sheet sulfate deposition within the uncertainty range of measurements for eight of the largest deposition signals recorded in the last 2,500 years. We then estimate the range of RFs associated with these parameter combinations. Consequently, we calculate the RF and the associated uncertainty range of eruptions from ice-core sulfate records without making any assumptions about transfer functions and conversions factors, while also accounting for uncertainty in the sulfate deposition.

We describe the model setup and statistical emulation in Section 2 and examine the range in ice-sheet sulfate deposition across our simulations in Section 3.1. In Section 3.2, we constrain the eruption source parameters for eight of the largest bipolar deposition signals recorded in ice cores (Sigl et al., 2015) and examine the volcanic RF and temperature uncertainty in Sections 3.3 and 3.4, respectively. Section 4 contains the discussion, and conclusions are found in Section 5.

## 2. Methods

### 2.1. Model Simulations

Simulations of volcanic eruptions were performed with the UM-UKCA state-of-the-art interactive stratospheric aerosol microphysics model (UK Hadley Centre Unified Model, HadGEM3-GA4 coupled with version 8.4 of the UK Chemistry and Aerosol scheme) as outlined in Marshall et al. (2019). The model has a horizontal resolution of  $1.875^\circ$  longitude by  $1.25^\circ$  latitude with 85 vertical levels up to 85 km and has an internally generated Quasi Biennial Oscillation (QBO) (Osprey et al., 2013). The simulations were atmosphere-only and free running, with year 2000 background conditions that included prescribed climatological sea surface temperatures and sea ice extent (Reynolds et al., 2007). Nonvolcanic  $\text{SO}_2$  emissions were taken from the standard Coupled Model Intercomparison Project phase 5 emissions data set (Lamarque et al., 2010) and  $\text{SO}_2$  emissions from continuously degassing volcanoes were based on Andres and Kasgnoc (1998). UM-UKCA has been used extensively to model volcanic eruptions and has been well evaluated for the 1991 eruption of Mt. Pinatubo (Dhomse et al., 2014). The version of the model used here was also previously used with preindustrial conditions to simulate the 1815 eruption of Mt. Tambora in a multi-model comparison and performed well compared to the other models and to ice-core records of sulfate deposition (Marshall et al., 2018).

Aerosol processes were simulated using the Global Model of Aerosol Processes modal aerosol microphysics scheme (GLOMAP-mode) (Mann et al., 2010, 2012), with aerosol mass and number concentrations simulated using seven log-normal modes: nucleation-soluble (geometric mean radius,  $r < 5$  nm), Aitken-soluble ( $5 < r < 50$  nm), accumulation-soluble ( $50 < r < 250$  nm), coarse-soluble ( $r > 250$  nm), Aitken-insoluble ( $5 < r < 50$  nm), accumulation-insoluble ( $50 < r < 500$  nm), and coarse-insoluble ( $r > 500$  nm). GLOMAP-mode includes primary emissions, new particle formation, condensation, coagulation, cloud processing, sedimentation, and dry and wet deposition. Further details on these processes can be found in Mann et al. (2010). The radiation scheme computes the optical properties accounting for the mean radius of each mode, the refractive index of the mode based on its internally mixed composition and the water content (Bellouin et al., 2013). In the version used here, based on that of Dhomse et al. (2014), UM-UKCA includes stratospheric and tropospheric chemistry (Morgenstern et al., 2009; O'Connor et al., 2014) and aerosols, interactive sulfur chemistry, meteoric sulfuric particles, and aerosol radiative heating, which has been shown to influence volcanic plume dispersion and subsequent radiative effects (Aquila et al., 2012; Mann et al., 2015; Sekiya et al., 2016).

Two ensembles of simulations were conducted, each containing the same 41 eruptions with different values of three eruption source parameters: the mass of  $\text{SO}_2$  emitted, the eruption latitude, and the emission injection height.  $\text{SO}_2$  emissions ranged between 10 and 100 Tg of  $\text{SO}_2$ , eruption latitude ranged between  $80^\circ\text{S}$  and  $80^\circ\text{N}$ , and the bottom injection height varied between 15 and 25 km with a plume depth of 3 km (i.e., injections from 15–18 km to 25–28 km). One ensemble was performed for eruptions on January 1<sup>st</sup> and the other for eruptions on July 1<sup>st</sup> to examine the seasonal dependence of meridional stratospheric aerosol transport and sulfate deposition (the July ensemble is presented in Marshall et al., 2019). The values of the eruption source parameters for each ensemble simulation were selected using a “maximin” Latin hypercube design (Figure S1) to achieve good coverage of the three-dimensional parameter space (Lee et al., 2011; Morris & Mitchell, 1995). Each eruption (simulation) was initialized by injecting the  $\text{SO}_2$  into the grid boxes within the 3 km plume over 24 h on the first day of the month. Both ensembles were initialized during similar easterly QBO phases. A 9-year control simulation was also conducted without any volcanic perturbation. The ensemble simulations were run for three years post eruption, by which time the majority (at least 83%, mean = 92%) of the injected sulfur had been deposited as sulfate. Uncertainties in the deposition are accounted for during statistical emulation, see Section 2.2.

The simulated sulfate deposited in each month ( $\text{kg SO}_4 \text{ km}^{-2}$ ) was calculated by summing the monthly mean dry and wet sulfate deposition flux components for each aerosol mode. The volcanic sulfate deposition was determined by subtracting the climatological monthly mean sulfate deposition derived from the control simulation. These anomalies were integrated over the 3 years of each simulation to produce the total volcanic sulfate deposition. Time-integrated effective RF was similarly calculated by summing the net (shortwave + longwave) top-of-the-atmosphere all-sky global mean radiative flux anomalies over

the 3 years of the simulation. This metric consequently represents the cumulative impact of each eruption and we refer to it as the “time-integrated RF” in units of  $\text{MJ m}^{-2}$ . Additionally, we also examine the 3-year time-integrated volcanic anomalies in global mean SAOD at 550 nm, which we refer to as “time-integrated SAOD.” SAOD is dimensionless, so time-integrated monthly SAOD has units of months. Radiative flux and SAOD anomalies were also derived with respect to climatological values from the control simulation.

## 2.2. Statistical Emulation

Statistical emulators are used as surrogate statistical representations of the UM-UKCA model output, which can be used to predict model output in a fraction of the time (seconds) compared to the simulations themselves (weeks). An emulator, trained on the model data, maps the input parameters (here the  $\text{SO}_2$  emission, eruption latitude, and injection height) to a model output (e.g., total sulfate deposited on Greenland) and can be used to predict that model output for any combination of the input parameters (within the bounds of the simulated ranges) that was not explicitly simulated. By sampling from an emulator thousands of times, a multi-dimensional response surface of the model output can be generated across the input parameter space, based on only a small set of model simulations used to train the emulator.

We generated four Gaussian process emulators (Johnson et al., 2015; Lee et al., 2011; O'Hagan, 2006) of the simulated deposition: total sulfate deposited on Greenland (ice-sheet mean) following eruptions in January and July, and total sulfate deposited on Antarctica (ice-sheet mean) following eruptions in January and July, and four Gaussian process emulators of the time-integrated RF and time-integrated SAOD for each season. Each emulator was constructed using R (RCoreTeam, 2017) and the DiceKriging package (Roustant et al., 2012). Following a Bayesian statistical approach, each model output response is assumed a priori to follow a Gaussian process, which is then updated with the information in the model data from 30 of the 41 ensemble members, known as “training runs”, to generate a posterior Gaussian process that forms our emulator. The emulators are built assuming a linear mean function that includes all parameters and a Matérn covariance structure (Rasmussen & Williams, 2006). In the Bayesian paradigm, these are prior specifications that are updated given the information in the training data. The posterior emulator is not restricted to be linear (i.e., the Gaussian process emulator does not assume linearity in the model output response that it represents). The nonlinearity in volcanic forcing that arises from the growth of aerosol particles that is explicitly simulated in UM-UKCA is therefore represented in the emulator, as shown by Marshall et al. (2019).

The emulator provides a mean prediction of the model output at any parameter combination in the parameter space along with an estimate of the variance in this prediction. These uncertainties increase with distance (in parameter space) from the training runs because there is less information on how the model output varies as a function of the input parameters. The remaining 11 simulations of each ensemble are used to validate the emulators (see below, and Figures S2 and S3) by comparing the emulator prediction (mean prediction and its uncertainty) for the parameter combinations of each simulation to the actual model output of each simulation (Bastos & O'Hagan, 2009).

We account for atmospheric variability within the emulators. This is necessary because the amount of sulfate deposited on the ice sheets is the result of a chain of several processes that includes the large-scale stratospheric transport of sulfate aerosol, stratosphere-troposphere exchange and deposition. The magnitude of sulfate deposition is therefore affected by variability in stratospheric circulation (e.g., because of the QBO and the polar vortices) and tropospheric meteorological variability. Sulfate is deposited under specific (but unknown) atmospheric conditions and is therefore unlikely to be accurately captured by any single climate model simulation. This uncertainty is normally addressed by varying the initial conditions in the model to generate a meteorological ensemble of simulations. Each of our simulations (in different parts of parameter space) are meteorologically different. Therefore, rather than running a meteorological ensemble at each point in parameter space, we account for the internal variability by adding a noise variance term when building the deposition emulators (Roustant et al., 2012). The variance term allows the emulator to vary more smoothly such that the mean emulator prediction does not have to pass exactly through each training point (Andrianakis & Challenor, 2012; Johnson et al., 2011; Salter & Williamson, 2016; Williamson et al., 2015). In this way, we can effectively characterize conventional ensemble variability in the construction of the emulator with the emulator mean prediction reflecting a meteorological ensemble mean. This method negates the need for running conventional meteorological ensemble members at each training

point and therefore reduces the computational cost of our experiment. A figure demonstrating this concept and further text are included in the supporting information (Figure S4, Text S1).

We chose to add a homogeneous noise variance term to each emulator with a magnitude of  $10 \text{ kg SO}_4 \text{ km}^{-2}$  in Greenland and  $2.5 \text{ kg SO}_4 \text{ km}^{-2}$  in Antarctica. These values reflect our estimates of the SD of the deposition over each ice sheet and assume the same variance in deposition for each training run. The values were chosen based on prior knowledge of deposition in Antarctica and Greenland from previous modeling studies and from the deposition variability in previous UM-UKCA simulations. For example, relative standard deviations (RSD) of 16% in Greenland and 9% in Antarctica due to the meteorological state have been estimated (Toohey & Sigl, 2017), based on ensembles of atmosphere-only simulations of eruptions with a range of  $\text{SO}_2$  emission magnitudes from 8.5 to 700 Tg (Toohey et al., 2013). Similarly, Gao et al. (2007) reported  $\sim 10\%$ – $20\%$  differences in sulfate deposition over Greenland and Antarctica among ensemble members following simulations of eruptions with  $\text{SO}_2$  injections ranging from 5 to 122 Tg. We found that across five meteorological ensemble members of atmosphere-only simulations of UM-UKCA of the eruption of Mt. Tambora (Marshall et al., 2018) (simulated during the easterly QBO phase), the SD of the Greenland deposition was  $\sim 3 \text{ kg SO}_4 \text{ km}^{-2}$  (RSD =  $\sim 10\%$ ) and the SD of Antarctica deposition was  $\sim 2 \text{ kg SO}_4 \text{ km}^{-2}$  (RSD =  $\sim 9\%$ ). Given that the simulation of Mt. Tambora was initialized with a 60 Tg  $\text{SO}_2$  injection at the equator, it is reasonable that an average noise variance term for our ensemble with emissions spanning 10 to 100 Tg  $\text{SO}_2$  and across both high and low latitudes, is higher. Regardless of the value of estimated noise, we found that the overall shape and pattern of the emulated surfaces remained the same, but the emulator validation was improved using the given values. We found that higher estimates of this noise variance led to poorer emulator validation and response surfaces with reduced variation in model output versus the eruption source parameters. These values remain a best-estimate and reflect the average variability in deposition across the whole parameter space and across both seasons. Although it is possible to specify a heterogeneous noise term, it is likely this would introduce greater uncertainty given a lack of information as to how internal variability terms may vary across the parameter space. Because our simulations use one set of prescribed SSTs, internal variability associated with ENSO is not included. Similarly, we do not investigate variability associated with different QBO phases.

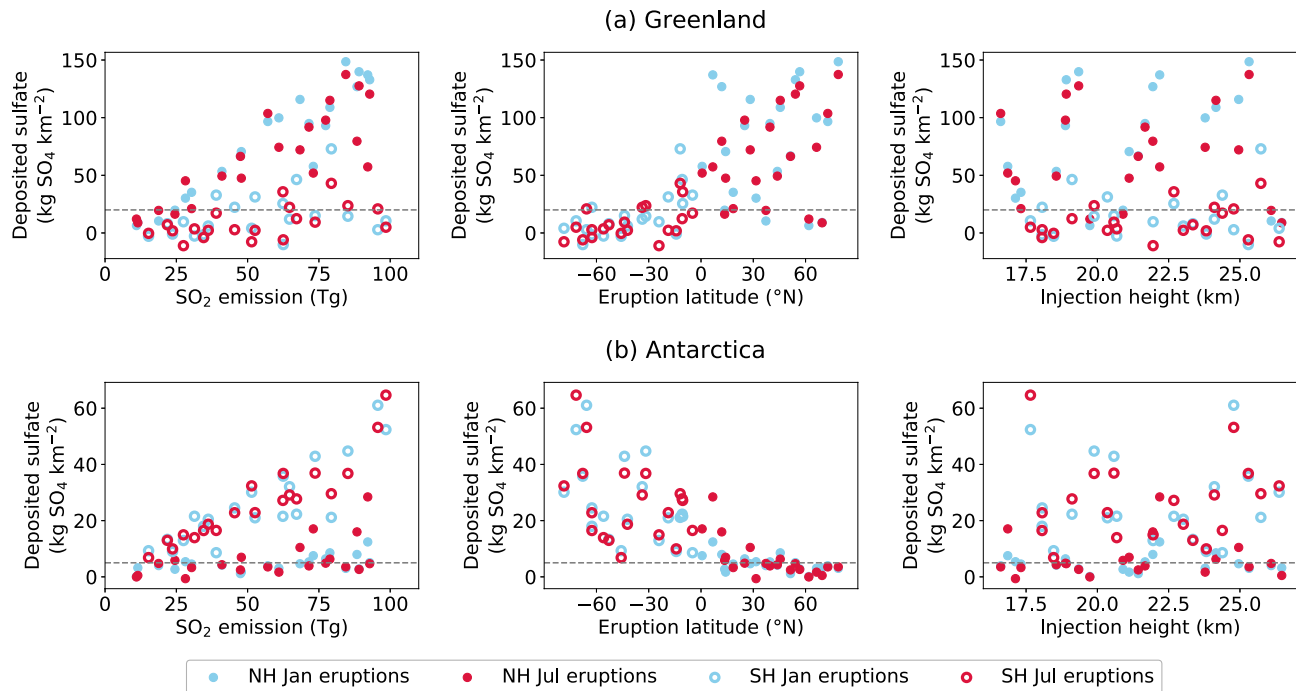
The time-integrated RF and time-integrated SAOD emulators were built without noise because the global mean SAOD and RF signals have relatively low variability compared to the deposition (they are not determined by tropospheric meteorology) and validation of the emulators was reasonable without an additional noise term.

Validation of the emulators is shown in Figures S2 and S3. The emulator predictions follow very close to the 1:1 line (marking a perfect prediction by the emulator) in all cases. The 95% confidence bounds on the emulator predictions are larger for the deposition emulators (Figure S2) compared to the SAOD and RF emulators (Figure S3) because the fit is more uncertain. Overall, the emulators are reasonable surrogates of the UM-UKCA output. Emulated response surfaces of the model outputs were produced by sampling the predicted response of each emulator 1 million times over a three-dimensional grid generated with 100 values of each eruption source parameter (covering the range in values simulated for each parameter).

### 3. Results

#### 3.1. Simulated Ice-Sheet Sulfate Deposition

Figure 1 shows the 82-member UM-UKCA emulator training and validation data set of simulated Greenland and Antarctica sulfate deposition (ice-sheet averages) against  $\text{SO}_2$  emission, eruption latitude, and injection height. The ice-sheet sulfate deposition is greater for eruptions with the largest  $\text{SO}_2$  emissions that are also close to each ice sheet since more aerosol is formed that is more likely to get deposited on the ice sheet due to its proximity. Because the deposition is dependent on both the  $\text{SO}_2$  emission and the eruption latitude, there are eruptions that are close to the ice sheets but have low deposition because the  $\text{SO}_2$  emission was low, and eruptions with high  $\text{SO}_2$  emissions but low deposition because they are far away from the ice sheet. We do not find an obvious relationship between the injection height of the  $\text{SO}_2$  emissions and the magnitude of the ice-sheet sulfate deposition.



**Figure 1.** Time-integrated volcanic sulfate deposited on (a) Greenland (area-mean) and (b) Antarctica (area-mean) in each simulation versus the value of sulfur dioxide emission (left), eruption latitude (middle), and injection height (right) in that simulation. Volcanic deposited sulfate is calculated by subtracting the climatological deposited sulfate in the control from the deposited sulfate in each perturbed simulation. Deposition from the January eruptions are shown in blue and from the July eruptions in red. Northern hemisphere (NH) eruptions are shown by the closed circle markers; southern hemisphere (SH) eruptions are shown by the open markers. There are different scales on the y axes between (a) and (b). Injection height marks the middle of the 3-km deep plume (see Section 2). The horizontal gray dashed lines mark approximate thresholds above which a volcanic signal is deemed significant ( $20 \text{ kg SO}_4 \text{ km}^{-2}$  on Greenland and  $5 \text{ kg SO}_4 \text{ km}^{-2}$  on Antarctica; see text for derivation).

On average, the deposition of sulfate ( $\text{SO}_4$ ) on Greenland is higher than on Antarctica. The maximum deposition on Greenland of  $149 \text{ kg km}^{-2}$  occurs for a January eruption at  $79^\circ\text{N}$ , with a  $\text{SO}_2$  emission of 84 Tg. The maximum simulated Antarctica deposition of  $65 \text{ kg km}^{-2}$  occurs for a July eruption at  $72^\circ\text{S}$  with a  $\text{SO}_2$  emission of 98 Tg. Note that the Latin hypercube design on the training data set means that there are not equivalent eruptions in the northern and southern hemispheres. Lower deposition on Antarctica compared to Greenland was also found in a previous study of tropical eruptions (Toohey et al., 2013), most likely due to stronger meridional transport in the Northern Hemisphere (NH) and increased deposition because the NH is relatively more dynamically active than the Southern Hemisphere (SH). In the SH, the stronger polar vortex inhibits more poleward aerosol transport. The ice-sheet deposition will also vary with  $\text{SO}_2$  emission magnitude as particles grow larger and sediment more quickly out of the atmosphere such that they may be deposited before reaching the ice sheets, and because of stronger polar vortices arising from aerosol-induced stratospheric heating (Toohey et al., 2013, 2014).

Deposition on Greenland is greater for tropical eruptions occurring in January (blue circles in Figure 1a) because more sulfate aerosol is transported to the NH via the Brewer Dobson Circulation (BDC), which is stronger in the winter hemisphere. Similarly, both total SH deposition (Figure S5) and deposition on Antarctica from tropical eruptions is greater if they occur in July (red circles in Figure 1b) following the seasonal cycle of the BDC. For eruptions at latitudes greater than  $\sim 30^\circ\text{N/S}$  the total hemispheric deposition is similar between the seasons (Figure S5), however ice-sheet deposition varies between seasons, but is not consistently larger in either one. Differences in the ice-sheet deposition following eruptions in different seasons for mid-to-high latitude eruptions could be dependent on the  $\text{SO}_2$  emission magnitude, injection height, and seasonal variations in stratosphere-troposphere exchange and sulfate aerosol deposition rates in the mid-latitude storm tracks (Kravitz & Robock, 2011). Seasonal differences may also arise due to internal variability.



There is considerable scatter in the deposition values around zero for eruptions located at high latitudes in the opposite hemisphere to the ice sheet, which has implications for the detection and quantification of volcanic signals. Meteorological variability is the cause of this scatter, which can obscure or enhance the deposition (i.e., the amount of background tropospheric-originating sulfate aerosol deposited on the ice sheets can be very different in the perturbed runs compared to the control climatology). As outlined in Section 2.2, we have estimated that the deposition values have a SD of  $10 \text{ kg SO}_4 \text{ km}^{-2}$  in Greenland and  $2.5 \text{ kg SO}_4 \text{ km}^{-2}$  in Antarctica (i.e., a 2 SD uncertainty of 20 and  $5 \text{ kg SO}_4 \text{ km}^{-2}$  respectively). We have also quantified how high the integrated total could be due to internal variability alone in the absence of a volcanic eruption by examining the variability in 3-year integrated deposition in the control simulation compared to the control climatology and find maximum anomalies also of  $\sim 20 \text{ kg SO}_4 \text{ km}^{-2}$  for Greenland and  $\sim 5 \text{ kg SO}_4 \text{ km}^{-2}$  in Antarctica. Consequently, we define these values as thresholds above which a volcanic signal is deemed significant (gray dashed lines Figure 1). For a few high-latitude eruptions, the deposition is very slightly above these thresholds. We do find that some sulfate aerosol can be transported globally for eruptions at latitudes as high as  $79^\circ$  but given the proximity to the threshold it is also possible that this is noise. There is more anthropogenic sulfate in the NH, which adds to the high meteorological variability simulated in Greenland deposition. Our analysis of a wide range of simulated eruptions highlights the inherent difficulty in detecting and quantifying volcanic deposition anomalies in ice cores (e.g., Cole-Dai et al., 2000; Ferris et al., 2011).

### 3.2. Ice-Sheet Sulfate Deposition Predicted for Thousands of Eruptions

By replacing the UM-UKCA model with statistical emulators that describe how the ice-sheet deposition varies with  $\text{SO}_2$  emission, eruption latitude and injection height, we can predict the sulfate deposition for any eruption that we did not directly simulate within the bounds of our simulated ranges (see Section 2). The emulators, which are built for January and July eruptions separately, enable us to examine the relationship between ice-sheet sulfate deposition and the eruption source parameters in unprecedented detail. This is because the emulated predictions describe how deposition varies continuously for all possible eruptions within the three-dimensional parameter space, effectively interpolating between the model output of the ensemble simulations. An example of the emulator surfaces is shown in Figure S6.

The combinations of eruption source parameters that lead to deposition within a measured range can be found by constraining the emulator predictions for all possible eruptions in the three-dimensional space. To illustrate the constraint procedure we take the emulated deposition and find the source parameters of eruptions that lead to the range in ice-sheet sulfate deposition derived from ice cores (Sigl et al., 2015) for the 1815 eruption of Mt. Tambora (Table 1). The 1815 Mt. Tambora eruption led to the so-called “year without a summer” (e.g., Raible et al., 2016) and remains the best test case since the eruption resulted in a large bipolar deposition signal. The constraint procedure is similar in concept to analysis by Toohey, Kruger, et al. (2016) who constrained the likely parameters for the 536 CE and 540 CE unattributed eruptions by comparing the simulated deposition in potential candidate eruptions with measured ice-sheet deposition. The difference from our approach is that they used the ratio of the ice-sheet sulfate deposition, while we use the absolute magnitude, and we use the emulators to generate a much larger set of candidate eruptions. The three-dimensional constrained parameter spaces for each eruption season (i.e., the eruptions that lead to sulfate deposition of  $39.7 \pm 10.4 \text{ kg km}^{-2}$  in Greenland and  $45.8 \pm 5.3 \text{ kg km}^{-2}$  in Antarctica) are shown in Figure 2a. The uncertainties on the emulator predictions (that the emulator itself derives) are accounted for during the constraint: a combination of parameters (i.e., an eruption realization) is retained if the interval of the emulator mean prediction plus or minus 1 SD overlaps with the uncertainty range of the ice-core-derived sulfate deposition estimate for both ice sheets. The color in Figure 2 shows the emulated mean prediction of the time-integrated RF for each eruption and is explored in Section 3.3.

For an eruption occurring in January, the predicted deposition falls within both the Greenland and Antarctica deposition uncertainty ranges only if the  $\text{SO}_2$  emission is between 75 and 100 Tg (our upper limit) and the latitude of the eruption is between  $22^\circ\text{S}$  and  $56^\circ\text{S}$ . All values of injection height when combined with these ranges of emission and latitude produce plausible deposition; therefore the injection height parameter remains unconstrained. For an eruption in July, only eruptions with  $\text{SO}_2$  emissions greater than 79 Tg and with a latitude between  $6^\circ\text{S}$  and  $62^\circ\text{S}$  can produce deposition that matches both ice-sheet constraints. We assume that since the deposition constraints are above our thresholds derived in Section 3.1 that these

**Table 1**  
Constrained Ranges in Sulfur Dioxide ( $\text{SO}_2$ ) Emission ( $\text{Tg SO}_2$ ), Eruption Latitude ( $^\circ\text{N}$ ), and Time-Integrated RF ( $\text{MJ m}^{-2}$ ) for the 10 Largest Bipolar Ice-Core-Derived Sulfate Deposition Signals in Greenland and Antarctica (Sigl et al., 2015) (in Rank Order of Magnitude)

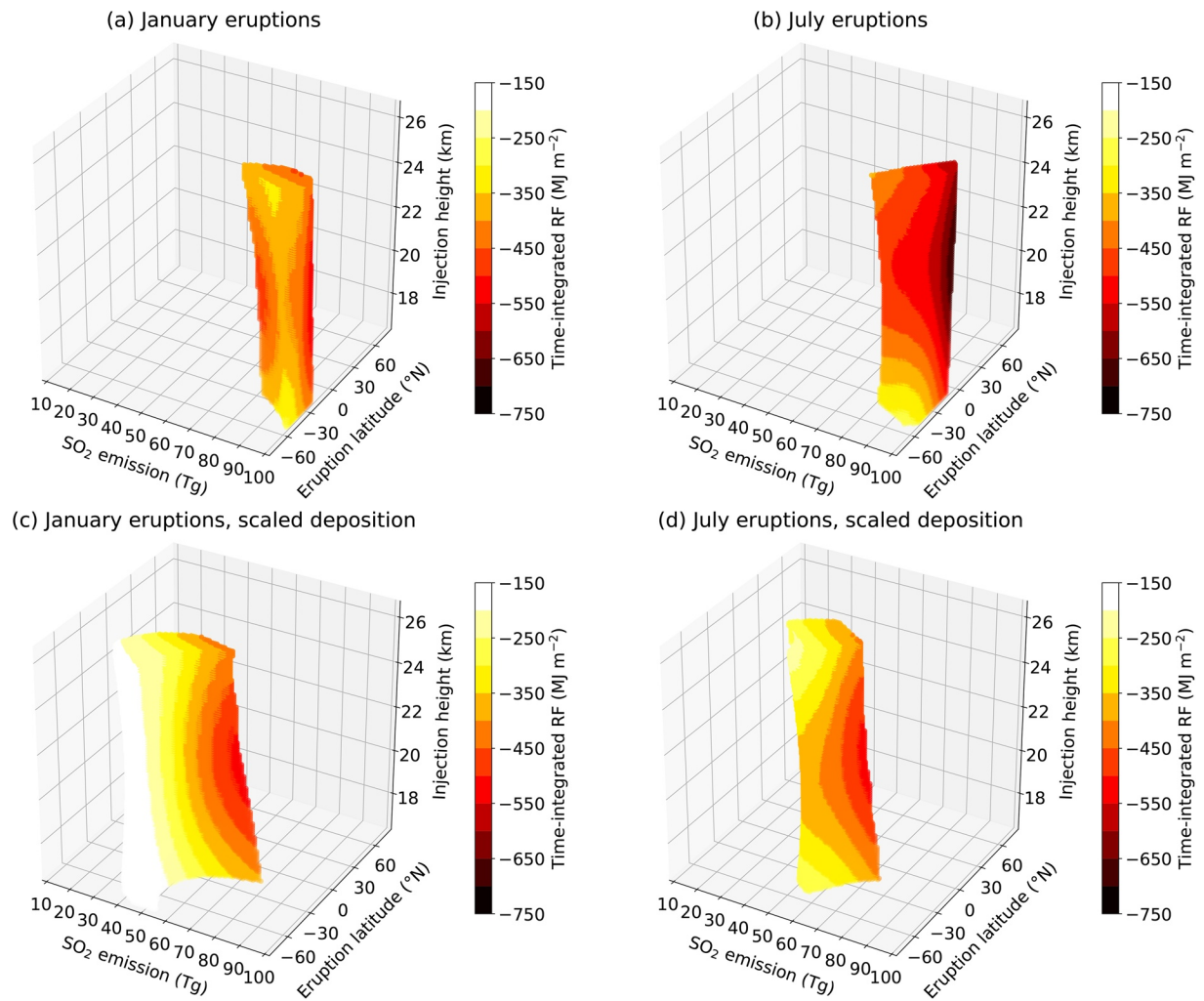
| Year<br>(BCE/CE) | Ice-sheet deposition ( $\text{kg SO}_4 \text{ km}^{-2}$ ) |            | January eruptions |            |                      |                      |           |           | July eruptions |            |                      |                      |             |             | eVolv2k<br>$\text{SO}_2$ |
|------------------|---|------------|-------------------|------------|----------------------|----------------------|-----------|-----------|----------------|------------|----------------------|----------------------|-------------|-------------|--------------------------|
|                  | Greenland   | Antarctica | Lat<br>min        | Lat<br>max | $\text{SO}_2$<br>min | $\text{SO}_2$<br>max | RF<br>min | RF<br>max | Lat<br>min     | Lat<br>max | $\text{SO}_2$<br>min | $\text{SO}_2$<br>max | RF<br>min   | RF<br>max   |                          |
| −426             | 100 ± 26  | 78 ± 21    |                   |            |                      |                      |           |           |                |            |                      |                      |             |             | 119 ± 72                 |
| 1258             | 90 ± 24   | 73 ± 9     |                   |            |                      |                      |           |           |                |            |                      |                      |             |             | 119 ± 43                 |
| −44              | 101 ± 26  | 15 ± 4     | −9                | 40         | 59                   | 100                  | −357      | −648      | 6              | 75         | 62                   | 100                  | −280        | −645        | 77 ± 45                  |
| 1458             | 39 ± 10   | 64 ± 7     | −64               | −40        | 91                   | 100                  | −287      | −432      | −67            | −49        | 96                   | 100                  | −249        | −526        | 66 ± 19                  |
| 540              | 61 ± 16   | 34 ± 4     | −30               | −11        | 73                   | 100                  | −393      | −618      | −28            | 17         | 75                   | 100                  | −400        | −714        | 64 ± 31                  |
| 1815             | 40 ± 10   | 46 ± 5     | −56               | −22        | 75                   | 100                  | −287      | −553      | −62            | −6         | 79 (97)              | 100 (100)            | −270 (−575) | −687 (−685) | 56 ± 18                  |
| 1230             | 56 ± 15   | 23 ± 3     | −20               | −6         | 54                   | 99                   | −362      | −610      | −17            | 19         | 53                   | 100                  | −384        | −702        | 48 ± 21                  |
| 682              | 38 ± 10   | 39 ± 5     | −40               | −19        | 66                   | 100                  | −342      | −565      | −54            | −4         | 70                   | 100                  | −292        | −685        | 54 ± 19                  |
| 574              | 38 ± 10   | 34 ± 4     | −35               | −15        | 59                   | 100                  | −353      | −570      | −49            | −1         | 63                   | 100                  | −306        | −685        | 48 ± 20                  |
| 266              | 61 ± 16   | 11 ± 1     | −11               | 11         | 39                   | 87                   | −242      | −594      | 2              | 36         | 37                   | 88                   | −214        | −568        | 44 ± 23                  |

Note. Year (BCE/CE) is the onset of the deposition signal. Also included are the eVolv2k (Toohey & Sigl, 2017)  $\text{SO}_2$  emissions (plus or minus 2 SD uncertainty) ( $\text{Tg of SO}_2$ ). For all cases, the minimum and maximum plausible injection plumes were 15–18 km and 25–28 km<sup>a</sup>. The top two signals are unconstrained because they are outside of our parameter space. Although the latitude of 1815 Mt. Tambora is known, we keep the whole range in plausible parameters as an example for if the eruption had not been attributed. Values in parentheses indicate those retained for the latitude of Mt. Tambora only. Values have been rounded to the nearest integer.

<sup>a</sup>For eruption realizations only at the latitude of Mt. Tambora, the injection altitude was constrained to 21.0–26.5 km.

high latitude combinations are plausible; however note that constrained combinations at high latitudes should still be interpreted with caution given the role of internal variability (i.e., nonvolcanic sulfate deposition) and the emulator uncertainty. To match the ice-sheet deposition for an eruption occurring at the latitude of Mt. Tambora (8°S), the eruption must occur in July and the  $\text{SO}_2$  emission must be greater than 97 Tg (the emission altitude is also reduced to between 21.0 and 26.5 km). This emission is higher than the 60 Tg estimate often used to simulate this eruption in climate models (Marshall et al., 2018; Zanchettin et al., 2016), but closer to petrological estimates (73–91 Tg) (Vidal et al., 2016). We have built emulators only for eruptions occurring on January 1<sup>st</sup> and July 1<sup>st</sup>, whereas Mt. Tambora erupted in April 1815. The predicted deposition is generally higher in Greenland for January eruptions, and higher in Antarctica for July eruptions (Figure S6) and deposition following an April eruption would also differ given seasonally varying stratospheric transport of sulfate aerosol and depositional processes (Gao et al., 2008). Furthermore, previous simulations of the 1815 eruption of Mt. Tambora using UM-UKCA with 60 Tg  $\text{SO}_2$  emitted at the equator showed that sulfate deposited on the ice sheets was roughly half that derived from ice-core estimates (Marshall et al., 2018). This indicates that either the  $\text{SO}_2$  emission used in the model was too low or that a structural error exists within the model resulting in a low bias in deposition. Background (nonvolcanic) sulfate deposition simulated in UM-UKCA was found to compare well to ice-core estimates.

Figures 2c and 2d show the constrained parameter space when the ice-core-derived estimates are divided by 2 to account for a potential bias in the model, although the lower scaled estimates are also more likely to be produced by noise (see Figure 1). For eruptions in January the  $\text{SO}_2$  emission now lies between 30 and 78 Tg (a ~20 Tg larger range than for the unscaled deposition where the emissions were capped at 100 Tg) and latitude between 12°S and 80°S. Any higher  $\text{SO}_2$  emission leads to too-high deposition compared to the scaled ice-core estimates. For eruptions in July, the  $\text{SO}_2$  emission ranges between 35 and 75 Tg and latitude between 4°N and 54°S, covering the latitude of Mt. Tambora. This indicates that even if a structural bias does exist in UM-UKCA, there is still a range of eruption source parameter combinations that can lead to the same ice-sheet deposited sulfate. The observationally plausible combinations also include the parameters used to simulate the eruption of Mt. Tambora in preindustrial conditions in Marshall et al. (2018).



**Figure 2.** Constrained parameter space showing the combinations of sulfur dioxide (SO<sub>2</sub>) emission, eruption latitude, and injection height that result in the possible range of ice-core-derived sulfate deposition following 1815 Mt. Tambora from Sigl et al. (2015) for January (a) and July (b) eruptions. (c and d) show the constrained space if the ice-core-derived estimates are divided by 2 (to account for a potential bias in the model) for January and July eruptions, respectively. Parameter combinations are retained if the emulator mean prediction plus or minus 1 SD for both the Antarctica and Greenland deposition overlaps with the ice-core-derived ranges (Table 1). The constrained space is made up of scatter points of the retained parameter combinations and the color of each scatter point shows the corresponding emulator mean prediction of the time-integrated global mean radiative forcing (RF) representing the potential climatic impact for each of these eruptions. Injection height marks the middle of the 3-km-deep plume.

For the 1991 Mt. Pinatubo sulfate deposition (no scaling,  $21.4 \pm 10.7 \text{ kg km}^{-2}$  in Greenland and  $11.1 \pm 1.4 \text{ kg km}^{-2}$  in Antarctica; Sigl et al., 2015) for July eruptions at 15°N (latitude of Mt. Pinatubo and month closest to the real eruption month of June), we return eruption realizations with SO<sub>2</sub> emissions between 18 and 49 Tg and injection heights between 17.5 and 26.5 km (middle of the 3 km plume). The lower SO<sub>2</sub> emission falls within the range of previous estimates (10–20 Tg; Guo et al., 2004; Timmreck et al., 2018), although is higher than the 10 to 14 Tg commonly used in models to simulate this eruption (Dhomse et al., 2014; Feinberg et al., 2019; Kleinschmitt et al., 2017; Mills et al., 2016, 2017; Sukhodolov et al., 2018). For this lower emission, the injection height must be around 26 km, also in line with observations (estimated peak SO<sub>2</sub> was around 25–26 km; Guo et al., 2004; Read et al., 1993). The plausible parameter space for eruptions only at 15°N is shown in Figure S7, alongside the plausible parameter space for the 1815 Tambora deposition for eruptions only at ~8°S.

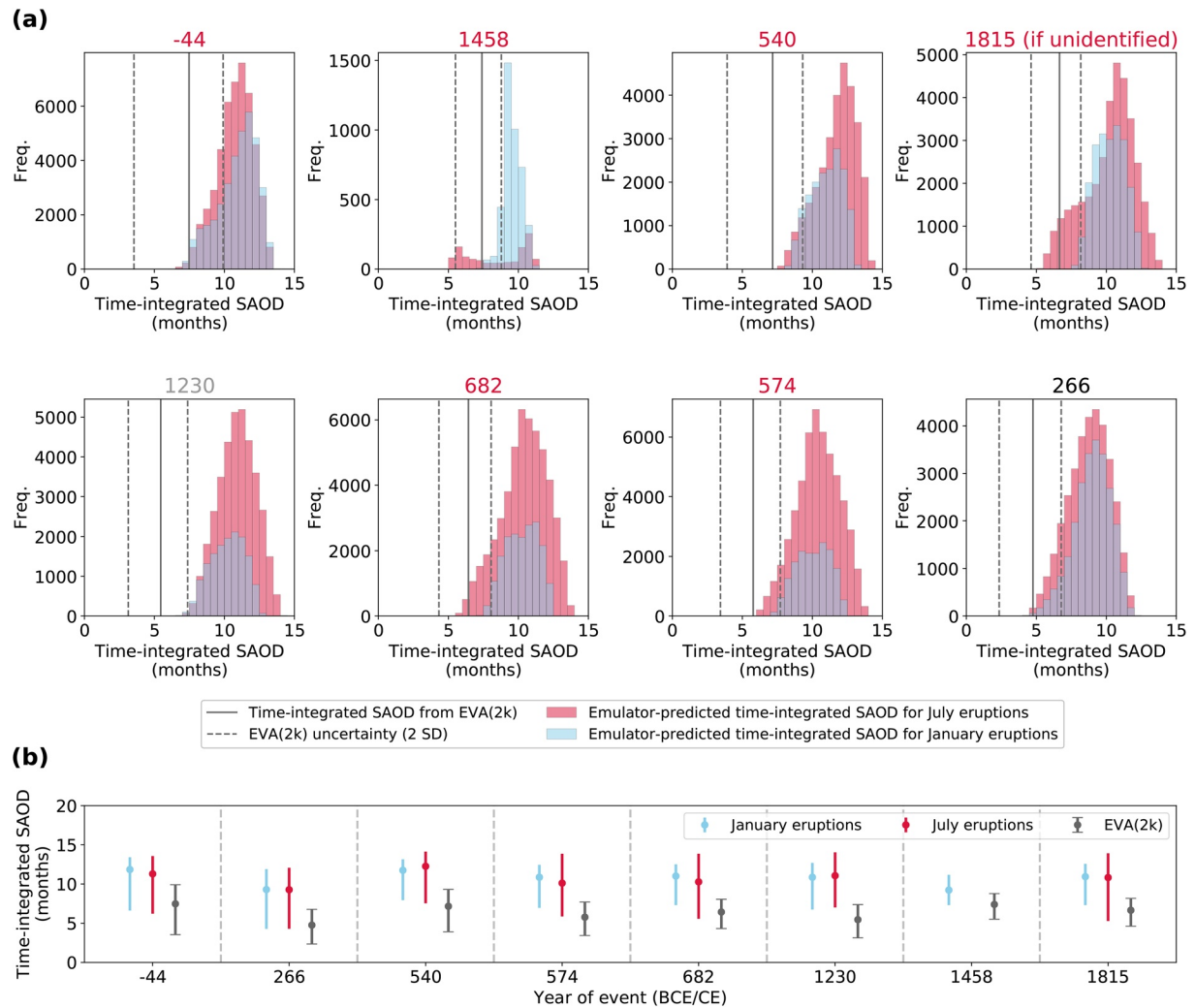
### 3.3. Constraining Volcanic RF for the Largest Bipolar Deposition Signals

We now examine the full range of eruptions that could lead to the ice-core-derived sulfate deposition for eight of the ten largest bipolar signals in the last 2,500 years (Sigl et al., 2015), excluding the top two signals (426 BCE and 1257 CE Samalas), which are too large for the 100 Tg SO<sub>2</sub> upper bound of our parameter space. Lower estimates of the emissions of these two eruptions do span emissions less than 100 Tg (Toohey & Sigl, 2017; Table 1), but we return no plausible eruptions, suggesting that the emissions were larger than 100 Tg, also consistent with previous estimates (Gao et al., 2007; Oppenheimer, 2003; Vidal et al., 2016). Of the eight remaining signals we examine, only one has been unequivocally attributed to a known eruption, which is the 1815 eruption of Mt. Tambora. For each eruption realization retained in the constraint, we examine the time-integrated SAOD and the time-integrated RF predicted by these respective emulators to estimate the range in SAOD and forcing that is consistent with each deposition signal. Because we do not know for certain that a structural bias exists in UM-UKCA, we do not scale the deposition constraints, which may otherwise add an additional bias. By accounting for the emulator uncertainty on the deposition emulators (which also accounts for meteorological ensemble spread) and the uncertainty on the ice-sheet composite observations in the constraint procedure, we provide a conservative estimate of the range in eruption source parameter combinations and subsequently the range in time-integrated RF for the parameter space we have investigated.

Figures 3 and 4 show histograms of the constrained time-integrated SAOD and RF for the eruptions considered in this study. The frequency indicates the number of combinations of eruptions out of the 1 million sample points that are observationally plausible. The three-dimensional constrained parameter spaces as shown for the Tambora example in Figure 2 are shown in Figures S8 (January eruptions) and S9 (July eruptions) for each eruption, and the minimum and maximum values of each parameter are included in Table 1. Based on our model and the parameter ranges we investigate, there are very few combinations for the 1458 CE deposition, suggesting that the SO<sub>2</sub> emission also likely exceeds the 100 Tg bound of our parameter space. For the remaining deposition signals, many eruption realizations are retained with many different combinations of SO<sub>2</sub> emission, eruption latitude and injection height (Table 1, Figures S8 and S9), which lead to a large range in the time-integrated SAOD and RF for both January and July eruptions (Figures 3 and 4).

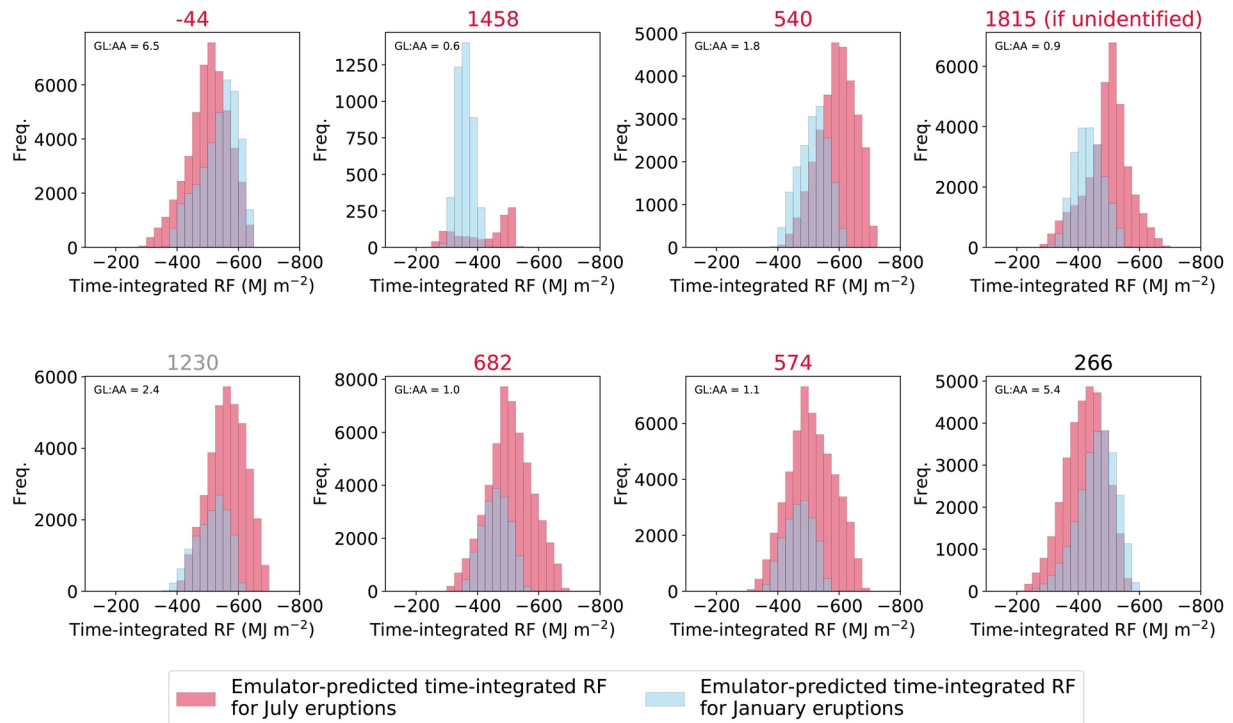
The largest range in time-integrated SAOD constrained for a single season is 8.6 months for a July 1815 Mt. Tambora-like eruption. Although we know the latitude of Mt. Tambora, we retain the whole range in the constrained set of parameter combinations to illustrate a case where the eruption had not been attributed. This range in SAOD translates into a time-integrated RF between  $-270$  and  $-687$  MJ m<sup>-2</sup> (i.e., a range of 417 MJ m<sup>-2</sup>). The largest range in time-integrated SAOD across the minimum and maximum constrained values from both seasons is also 8.6 months for 1815 Mt. Tambora due to the range across the July eruptions (the minimum and maximum SAOD values for January eruptions remain within the July values; Figure 3). The largest range across both seasons in time-integrated RF also occurs for the Mt. Tambora deposition due to the range across July eruptions (Figure 4). The RF distributions do not always match the SAOD distributions because the RF is also dependent on the insolation, surface albedo and cloud cover. This difference highlights the uncertainty associated with using a single factor to convert global mean SAOD to RF in previous reconstructions (e.g., Sigl et al., 2015). The smallest range across both seasons in time-integrated SAOD is 6.2 months and in time-integrated RF is  $-276$  MJ m<sup>-2</sup>, both for the 1458 CE deposition. However, this small range only occurs because the eruption realizations were capped at 100 Tg of SO<sub>2</sub>.

The constrained time-integrated RF values are different for eruptions in different seasons because different combinations of parameters are plausible for each season and because, even for the same combinations of parameters, the RF can differ between seasons. Emulator-predicted time-integrated RF for eruptions across the three-dimensional parameter space and in both seasons are shown in Figure 5. The highest time-integrated RF occurs for eruptions at the equator if they occur in July (Figure 5b), but for eruptions in January the highest forcing occurs for eruptions slightly south of the equator (Figure 5a). This is likely because of seasonal variations in the position and strength of the tropical pipe that controls the tropical confinement of aerosol and its hemispheric transport (Holton et al., 1995; Seviour et al., 2012) and subsequently affects the time-integrated RF. Mid-latitude eruptions also lead to stronger time-integrated RF if they occur in winter because it takes up to 8 months for the peak aerosol burden to be reached (in the UM-UKCA simulations)



**Figure 3.** (a) Range and distributions of the emulator-predicted 3-year time-integrated global mean stratospheric aerosol optical depth (SAOD) (emulator mean) of observationally plausible eruption realizations for the eight bipolar deposition signals investigated (in rank order of magnitude). The red distribution is the time-integrated global mean SAOD for July eruptions and the blue distribution is for January eruptions. The frequency (Freq.) indicates the number of combinations of eruptions out of the 1 million sample points that are plausible. The vertical gray lines mark the time-integrated SAOD derived for each of these eruptions from the Paleoclimate Modelling Intercomparison Project phase 4 (PMIP4past2k) EVA(2k) data set using the medium and upper and lower estimates (2 SD) of the stratospheric sulfur dioxide (SO<sub>2</sub>) emissions. The year (BCE/CE) of the onset of each deposition signal is shown at the top of each panel and color-coded depending on whether the constrained space is capped at 100 Tg. The title is gray if only the July eruptions are capped at 100 Tg, red if both January and July eruptions are capped at 100 Tg and black if neither of the sets of eruptions are capped at 100 Tg (there are no cases where the January eruptions are capped at 100 Tg but the July eruptions are not). (b) Range in the time-integrated global mean SAOD for all deposition signals from the January and July emulators (red and blue lines) and from EVA(2k) using the 2 SD estimates (gray lines). The circles mark the modal SAOD value of each of the constrained distributions taken from a kernel density estimate and the SAOD value from EVA(2k) using the median estimate of the stratospheric SO<sub>2</sub> emission. The error bars remain uncapped for the emulator-predictions because these uncertainties will be larger given the 100 Tg bound of our parameter space and the emulator uncertainties. Because so few combinations of parameters are retained for July eruptions matching the 1458 deposition signals, this data is not shown in panel (b).

and the higher aerosol burden subsequently coincides with peak summer insolation. More combinations are retained in total for the July eruptions (discounting 1458) and except for the 44 BCE and 266 CE deposition, the time-integrated RF of the plausible July eruptions reach stronger values compared to the plausible January eruptions. Compared to the January eruptions, the plausible July eruptions are either shifted toward or expanded into the NH or expanded further into both hemispheres, reaching stronger values of time-integrated RF near the equator (Figures S8 and S9). Tropical NH eruptions occurring in January would lead to higher Greenland deposition and lower Antarctic deposition because more aerosol is transported to the NH and deposition would fall outside of the ice-sheet constraints. The time-integrated RF also depends on seasonal differences in large-scale transport and insolation. The time-integrated RF emulated surfaces



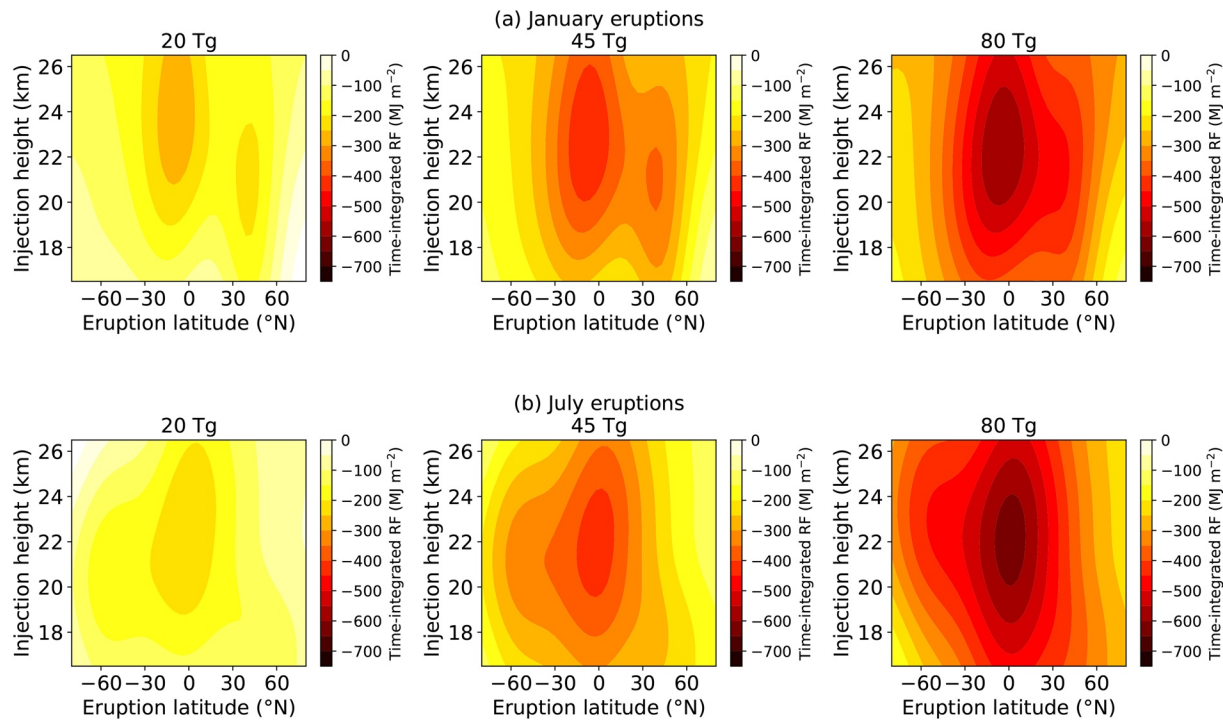
**Figure 4.** Range and distributions of the emulator-predicted 3-year time-integrated global mean radiative forcing (RF) (emulator mean) of observationally plausible eruption realizations for the eight deposition signals constrained in rank order of magnitude. The red distribution is the time-integrated global mean RF for July eruptions and the blue distribution is for January eruptions. The frequency (Freq.) indicates the number of combinations of eruptions out of the 1 million sample points that are plausible. The year (BCE/CE) of the onset of each deposition signal is shown at the top of each panel and color-coded depending on whether the constrained space is capped at 100 Tg. The title is gray if only the July eruptions are capped at 100 Tg, red if both January and July eruptions are capped at 100 Tg and black if neither of the sets of eruptions are capped at 100 Tg. The Greenland (GL) to Antarctica (AA) deposition ratio is shown in the upper left corner. The x-axis is reversed to show stronger forcing toward the right.

are complex (see also discussion in Marshall et al., 2019), for example for emissions of 80 Tg SO<sub>2</sub> the time-integrated forcing for eruptions near the equator is stronger for a July eruption compared to a January eruption, but this is opposite for an SO<sub>2</sub> emission of 20 Tg (Figure 5). Several of the plausible January eruptions also have combinations with slightly lower SO<sub>2</sub> emissions than the July combinations (Table 1).

Following the 44 BCE and 266 CE eruptions, much more sulfate is deposited on Greenland than on Antarctica (Greenland to Antarctica deposition ratios are 6.5 and 5.4, respectively, Figure 4) than the other eruptions (ratios are less than 2.4). Consequently, the plausible January eruptions are in the tropics and the plausible July eruptions are shifted toward the NH mid-latitudes (especially for 44 BCE) with the closer proximity of these eruptions to Greenland balancing the reduction in poleward transport due to being in the summer hemisphere. The range in plausible SO<sub>2</sub> emissions remains similar between seasons for both cases and therefore the more similar time-integrated RF distributions for both seasons (compared to the other six eruptions), despite differences in eruption latitude, may also be because of differences in time-integrated RF related to large-scale transport, the position and strength of the tropical pipe, and insolation (Figure 5).

For all eight constrained deposition signals, the injection height across all plausible parameter combinations is not constrained as all values of injection height are plausible when combined with different eruption latitudes and SO<sub>2</sub> emissions. However, for some combinations of SO<sub>2</sub> emission and eruption latitude a narrower range of injection heights are plausible. Figure S8 shows that the January constrained parameter space is often sloped, with lower emissions that have the highest injection heights, and higher emissions with lower injection heights. The lack of constraint across all plausible combinations is important since the time-integrated RF is dependent on the injection height (Figure 5).

Our estimated midpoint values of plausible SO<sub>2</sub> emissions for each eruption are generally higher than the eVolv2k stratospheric SO<sub>2</sub> estimates (by ~25 Tg on average, Table 1) used to derive the PMIP4past2k



**Figure 5.** Emulated response surfaces of the time-integrated global mean radiative forcing (RF) at fixed sulfur dioxide emissions of 20, 45, and 80 Tg for January eruptions (a) and July eruptions (b) (as Figure 7 in Marshall et al., 2019 except RF is slightly smaller due to a difference in anomaly method: anomalies here are 36-month integrals with respect to a climatological background as opposed to a single paired control simulation). The contour plots show the emulator mean prediction of time-integrated RF against the latitude and injection height of the emissions for each of these emission magnitudes. The surface was built by sampling the emulator 1,600 times over a grid of 40 equally spaced values of the eruption latitude and injection height. The injection height values are the middle of the 3 km plume.

prescribed SAOD time series, EVA(2k) (Toohey & Sigl, 2017), although our lower estimates overlap the eVolv2k estimates (except for the 1458 CE deposition). The eVolv2k  $\text{SO}_2$  emissions were calculated from the ice-sheet sulfate deposition using transfer functions (see Section 1) and the uncertainty (2 SD) was estimated by considering random errors in the ice-core composites and in the transfer functions. For the 540 CE eruption, our plausible eruptions include the eruption latitude suggested by Toohey, Kruger, et al. (2016) ( $15^\circ\text{N}$ ) only if the eruption occurred in July, but do not include their estimated emission (50 Tg). For the 44 BCE signal, our plausible eruptions include the latitude of the Okmok volcano in Alaska ( $53^\circ\text{N}$ ), suggested by McConnell et al. (2020) as the source, only if the eruption occurred in July.

The equivalent time-integrated SAOD for each of the eight eruptions from the EVA(2k) reconstruction are included in Figure 3 (vertical lines). We used three runs of EVA (with no background aerosol included so that the SAOD represents the volcanic SAOD only) that were run using the lower-end, middle, and upper-end eVolv2k  $\text{SO}_2$  estimates and isolated our eruptions (upper and lower injections were calculated by summing or subtracting the two SD uncertainties from the middle best-estimate predictions). For each eruption, we summed the SAOD over 3 years to compare directly to the time-integrated SAOD derived from the UM-UKCA simulations. The SAOD time series for each of the eight eruptions from the EVA(2k) reconstruction are shown in Figure S10. Since our constrained  $\text{SO}_2$  emissions reach higher values than the eVolv2k estimates, it is not surprising that the time-integrated SAOD from EVA(2k) for each of the eruptions is toward the lower end, or in the case of the 1230 CE eruption, almost outside of our SAOD ranges (Figure 3). Assuming our model captures well the ice-sheet deposition and the aerosol optical properties, our results suggest that the EVA(2k) SAOD may be an underestimate.

### 3.4. Estimating the Uncertainty in the Global Mean Surface Temperature Response

The time-integrated RF has an uncertainty range (using the minimum and maximum values) of on average  $359 \text{ MJ m}^{-2}$  for the eight historical eruptions and  $371 \text{ MJ m}^{-2}$  if excluding 1458. To put this uncertainty range in context, the time-integrated RF for the 1991 eruption of Mt. Pinatubo, which led to peak global cooling of  $\sim 0.5^\circ\text{C}$  (Dutton & Christy, 1992) has been estimated to be between  $-133$  and  $-229 \text{ MJ m}^{-2}$  (Marshall et al., 2019) and is  $-203 \text{ MJ m}^{-2}$  in the IPCC AR5 (Myhre et al., 2013) volcanic forcing series (integrated over 1991 to 1996). Our uncertainty range on the time-integrated RF of these eight eruptions therefore equates to approximately 2–3 times the total RF of 1991 Mt. Pinatubo. To estimate the range in global mean annual surface temperature response, we used the Finite Amplitude Impulse-Response simple climate-carbon-cycle model, FaIR (version 1.4, Smith et al., 2018). FaIR was run using the Representative Concentration Pathway 4.5 scenario (Meinshausen et al., 2011) and forced with the global post-eruption annual mean RF from each of the 82 perturbed parameter ensemble members starting in the year 2020 (we found that results were insensitive to the year of eruption emplacement). An example of the forcing and temperature anomaly simulated for one ensemble member is shown in Figure S11. The peak global mean annual surface cooling across the 82 simulations ranged between  $-0.1^\circ\text{C}$  (for a January eruption at  $62^\circ\text{N}$  with an  $\text{SO}_2$  emission of 11 Tg) and  $-1.6^\circ\text{C}$  (July eruption at  $7^\circ\text{N}$  and 92 Tg). We built two emulators of the peak surface cooling, which validated well (not shown). Following the same approach to constraining the RF (Section 3.3), we calculated the range in peak cooling for each of the deposition signals. Across all deposition signals the peak surface cooling for the plausible eruptions is between  $-0.5^\circ\text{C}$  and  $-1.7^\circ\text{C}$ . These anomalies are consistent with reconstructed summer temperature anomalies following large millennium eruptions from tree-ring reconstructions (Sigl et al., 2015; Stoffel et al., 2015). The range in peak cooling averaged across the seven eruptions constrained (excluding 1458) was  $0.9^\circ\text{C}$  (Figure S12).

## 4. Discussion

Using a wide range of aerosol climate model simulations of explosive eruptions, we have built statistical emulators that describe how polar sulfate deposition, SAOD and RF depend on the  $\text{SO}_2$  injection magnitude, eruption latitude, injection altitude and eruption season. Using measured deposition constraints from ice cores we have explored the range of eruption realizations that could match the deposition constraints and examined the SAOD and RF of the plausible eruptions. These emulators can be explored by the interested reader using our online tool available at <https://cemac.github.io/volcanic-forcing-deposition>.

### 4.1. Comparison of Volcanic SAOD and Forcing Uncertainty With Previous Studies

Our results demonstrate that there is a large range in time-integrated volcanic SAOD and time-integrated RF from plausible eruptions that match ice-sheet sulfate deposition. Uncertainty associated with the forcing of past eruptions is not a new result, but our study is novel in the methodology used to quantify it, and our uncertainty can be explicitly attributed to unknown eruption source parameters. Compared to the PMIP-4past2k EVA SAOD reconstruction our uncertainty ranges in time-integrated SAOD are higher (Figure 3b). For example, we derive an SAOD range of 8.3 months for the 682 CE eruption (across both January and July eruptions) whereas for EVA the range is 3.7 months. For unidentified eruptions, our estimates account for uncertainties in eruption latitude and injection altitude, as well as partially accounting for uncertainties in the season, that are not included in the EVA reconstructions.

For reconstructed  $\text{SO}_2$  emissions, we find that our estimates of the uncertainty ranges are fairly consistent with the ranges from the eVolv2k estimates (that are used by EVA to produce the SAOD) for the eruptions that are within our upper limit of 100 Tg (266 January and July and 1230 January), although the magnitude of our  $\text{SO}_2$  emissions are generally higher (Section 3.3); for 1230 January we derive a range in  $\text{SO}_2$  of 45 Tg versus 42 Tg from eVolv2k, and for 266, a range of 48 Tg for January and 51 Tg for July, versus 46 Tg from eVolv2k (Table 1). For the Mt. Pinatubo 1991 deposition signals, we also find that our constrained  $\text{SO}_2$  emissions are at the higher end of the 10–20 Tg range derived from observations, which may suggest that our  $\text{SO}_2$  emissions and consequently SAOD magnitude for the eight deposition signals are upper estimates. On the other hand, assuming our model is correct because we do still return plausible  $\text{SO}_2$  emissions in line with observations for Pinatubo, the  $\text{SO}_2$  and SAOD from eVolv2k and EVA could be underestimated. At present,



these differences remain hard to reconcile. We have also only explored the time-integrated SAOD and have not considered the temporal evolution such as the peak SAOD and decay timescale. The uncertainty ranges in SO<sub>2</sub> emissions that we derive are also consistent with modeling results by Toohey et al. (2013) who estimated an uncertainty of at least 25% in reconstructed sulfate aerosol burdens depending on whether a tropical eruption occurred in January or July. They demonstrated that ice-sheet deposition could be similar for eruptions of very different magnitudes, although higher than the emissions perturbed in this study, with equivalent Antarctica deposition simulated for a 170 Tg SO<sub>2</sub> eruption in July and a 300 Tg SO<sub>2</sub> eruption in January. Although Toohey et al. (2013) only considered eruptions at 15°N, they reported a plausible range of about 70 Tg in the SO<sub>2</sub> emission of a 1257 Samalas-like eruption. Our constrained SO<sub>2</sub> emissions would also be larger if we had simulated eruptions with emissions greater than 100 Tg.

The corresponding percentage uncertainties in time-integrated RF based on the minimum and maximum forcing for the three eruptions we have been able to constrain fully are ±26% (1230 January), ±42% (266 January) and ±45% (266 July). Hegerl et al. (2006) suggested that RF for individual volcanic eruptions was uncertain by 50% (1 SD) based on implicitly accounting for uncertainties in the transfer function between ice-core sulfate and SAOD, in converting SAOD to RF, in the ice-core averaging and the variability of sulfur concentrations between ice cores. Although a direct comparison to this study is difficult since we have examined time-integrated forcing and the total range in forcing by explicitly accounting for unknown eruption source parameters, it appears our uncertainty is less for these three eruptions.

Our uncertainty on the estimated surface temperature response to these large eruptions (on the order of 1°C) is also consistent with simulations by Zanchettin et al. (2019) who also found peak cooling following 1815 Tambora to differ by around 1°C when using the upper and lower estimates of the EVA SAOD.

#### 4.2. Implications

Our constrained SO<sub>2</sub> emissions are at the higher end of the eVolv2k estimates used to derive the PMIP4past2k EVA SAOD reconstruction, suggesting that the transfer functions used to link ice-sheet deposition fluxes and stratospheric sulfate burdens do not hold for eruptions with SO<sub>2</sub> emissions larger than 1991 Mt. Pinatubo (in agreement with Toohey et al., 2013). Alternatively, there may be a structural bias in UM-UKCA that leads to too-low polar ice-sheet deposition. Regardless of the absolute magnitude, we show here that a large range of volcanic forcings for each deposition signal must be considered plausible given the wide range of eruption source parameter combinations that reproduce measured deposition anomalies.

A large range in volcanic forcing for past eruptions has important implications for understanding and attributing climate variability on millennial timescales because different volcanic forcings will lead to different climate responses. Only one realization of EVA(2k) with the medium SO<sub>2</sub> predictions is used in PMIP4past2k simulations, thus none of this uncertainty is accounted for. Our results also suggest there are many latitudes other than 0°N (to which unidentified eruptions leading to bipolar deposition signals are assigned) and outside of the tropics that could lead to the deposition signals. Furthermore, if the eruption occurred in January there is a higher likelihood that the eruption was south of the equator in the southern tropics to extratropics.

#### 4.3. Limitations

An inherent difficulty in quantifying volcanic deposition anomalies in a model simulation is the effect of natural variability. It is possible that some of the observationally plausible parameter combinations that we identified, especially at high latitudes, could be false positives that arise due to variations in the nonvolcanic sulfate deposition. However, this ambiguity also applies to the real world (e.g., Ferris et al., 2011; Trautetter et al., 2004). Most (except 44 BCE and 266 CE) of the ice-sheet deposition constraints we consider are large enough in magnitude that it is unlikely that these signals could be produced by nonvolcanic sulfate anomalies in our simulations (i.e., they are much greater than our thresholds of 20 kg SO<sub>4</sub> km<sup>-2</sup> on Greenland and 5 kg SO<sub>4</sub> km<sup>-2</sup> on Antarctica [Figure 1]). Our results have revealed that many eruptions could be missed in the ice-core records, or the volcanic sulfate deposition could be overestimated because of internal variability and false attribution. Accounting for the emulator uncertainty due to using the emulator in place of the model itself also increases the amount of parameter combinations that are plausible. However, by

design, we have not been able to separate this uncertainty from natural variability. This separation could be achieved in a follow-up study by conducting a small number of meteorological ensembles (i.e., with slightly perturbed starting conditions and different ENSO phases) for a set of eruptions at different points in our parameter space.

Our estimated RF uncertainty ranges are likely to be lower bounds for several reasons. Our SO<sub>2</sub> emissions were capped at 100 Tg and we considered eruptions only in January and July, although it is possible that the simulated deposition in these seasons are end members such that the spread may be well estimated. We considered only 1 SD uncertainty on the emulator-predicted deposition during constraint (not more), and we did not include the emulator uncertainty on the SAOD and RF predictions, although this was found during validation to be very small. We also only considered eruptions during the easterly QBO phase. The QBO has an important role for the meridional transport out of the tropics (Barnes & Hofmann, 2001; Choi et al., 1998; Hitchman et al., 1994; Hommel et al., 2015; Punge et al., 2009) and its variability will therefore affect the volcanic sulfate deposition and that from nonvolcanic sources. Accounting for the variability in the QBO could lead to a larger number of plausible eruptions being retained or may help to reconcile the phases that may have occurred during these eruptions based on the match to ice-sheet deposition. The possible low bias in deposition following 1815 Tambora as simulated by our model could also be related to an incorrect QBO phase.

We have constrained eruptions from preindustrial sulfate deposition signals using simulations conducted with a present-day atmosphere, where the background nonvolcanic sulfur emissions and consequently sulfate deposition is higher and large-scale atmospheric circulation is faster. However, we find that our present-day July emulators predict deposition following a 1815 Mt. Tambora-like eruption that matches that from UM-UKCA simulations of Mt. Tambora in preindustrial conditions where the eruption was simulated in April (Marshall et al., 2018) and therefore this difference is unlikely to significantly impact our results.

## 5. Conclusions

We have explored the plausible eruption source parameters and the 3-year time-integrated global mean SAOD and 3-year time-integrated global mean RF of eruptions corresponding to eight of the largest bipolar ice-sheet sulfate deposition signals over the past 2,500 years, without relying on transfer functions and scaling factors derived from a single eruption. Instead, we have used a state-of-the-art aerosol-climate model and statistical emulation to predict the sulfate deposition for thousands of eruptions that we did not simulate directly, with any SO<sub>2</sub> emission between 10 and 100 Tg, eruption latitude between 80°S and 80°N, and emission altitude between 15–18 km and 25–28 km.

Our results show that many quite different eruption realizations (i.e., eruptions with different combinations of SO<sub>2</sub> emission, eruption latitude, emission altitude, and eruption season) are consistent with ice-core-derived sulfate deposition records, leading to uncertainties in the time-integrated RF of  $\sim 360 \text{ MJ m}^{-2}$  and uncertainties in the peak global mean annual surface temperature anomaly on the order of 1°C for the largest eruptions in the past 2,500 years. These estimated uncertainties are in agreement with previous statements on the uncertainty of past volcanic forcing (Hegerl et al., 2006; Toohey & Sigl, 2017) but are derived explicitly since we account for the range in plausible eruption realizations, rather than estimates of uncertainty derived from the calculation of the reconstructions themselves. We also find that extratropical eruptions could be responsible for many of these large bipolar deposition signals. Future work should continue to explore how high the eruption latitude could be and still result in a detectable (above natural variability) bipolar sulfate deposition signal.

For the two largest signals in the last 2,500 years (426 BCE Unidentified and 1258 CE Samalas), we do not return any plausible eruption realizations, which suggests that the SO<sub>2</sub> emissions of these eruptions were greater than the 100 Tg bound of our parameter space. For the remaining eight deposition signals investigated, the plausible eruption realizations have a large range in RF because there are many combinations of eruption latitude, SO<sub>2</sub> emission, emission height, and season that produce ice-sheet sulfate deposition values that are within the uncertainty ranges of measured anomalies. These variations in eruption source parameters affect the amount of aerosol that is formed, the particle size distribution, the horizontal and vertical distribution of the aerosol, its lifetime and hence its radiative effect (Marshall et al., 2019).

This study is a step forward in quantifying some of the uncertainty inherent in calculating the RF of past eruptions, as well as demonstrating the applicability and benefit of statistical emulators. Our approach has potential for further constraining volcanic forcing uncertainty of past eruptions. For example, future work should explore using larger ranges of eruption source parameters and multiple constraints such as the temporal evolution of the ice-sheet sulfate deposition, proxy and instrumental records of surface temperature changes and ice-core sulfur isotope measurements that can be used to broadly constrain injection altitude (e.g., Burke et al., 2019). Extension of our approach to other aerosol-climate models would be vital in assessing model uncertainties such as in the formation and transport of sulfate aerosol and depending on the results could enable some of the source parameters of these eruptions to be further constrained. However, given uncertainties in many of these records, it is likely that many eruption realizations would still be plausible. The uncertainty in volcanic RF quantified here stems ultimately from the range of ice-core-derived sulfate deposition estimates. The larger this range, the higher the number of eruption realizations that produce plausible deposition. Therefore, future work to narrow the measurement uncertainty range will feed through to a reduced number of plausible eruption realizations and a narrower range of associated forcing estimates. Our approach could also be applied to more spatially resolved model outputs such as hemispheric or regional forcing on which the climate response depends, and which are likely to have greater uncertainties than estimated for the global mean.

Instead of using just one realization of the potential volcanic aerosol forcing, the uncertainty on these estimates should be accounted for in future model simulations by running several ensembles with different possible volcanic forcings to facilitate a more complete and robust understanding of millennial climate change.

**Acknowledgments**

Lauren R. Marshall was funded by the UK Natural Environment Research Council (NERC) through the Leeds-York NERC Doctoral Training Program (NE/L002574/1). Anja Schmidt, Lauren R. Marshall, and Ken S. Carslaw acknowledge funding from NERC grant NE/S000887/1 (VOL-CLIM). Anja Schmidt, Ken S. Carslaw, and Graham W. Mann received funding from NERC grant NE/N006038/1 (SMURPHS). Jill S. Johnson and Ken S. Carslaw acknowledge funding from NERC grant NE/J024252/1 (GASSP). Ken S. Carslaw acknowledges funding from NERC grant NE/P013406/1 (A-CURE). Jill S. Johnson and Ken S. Carslaw were supported by the UK China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership China as part of the Newton Fund. Graham W. Mann received funding from the National Centre for Atmospheric Science, one of the NERC research centers via the AC-SIS long-term science program on the Atlantic climate system. Ken S. Carslaw is currently a Royal Society Wolfson Merit Award holder. The authors thank Matthew Toohey for providing the EVA(2k) simulations, and Hans de Leeuw and Thomas Aubry for helpful discussions. The authors also thank Editor Ping Yang and an Associate Editor and three anonymous reviewers for their useful and constructive comments that helped to improve this paper. This work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>) and JASMIN super-data-cluster (<https://doi.org/10.1109/Big-Data.2013.6691556>), via the Centre for Environmental Data Analysis.

**Data Availability Statement**

Summary model output of the time-integrated global mean SAOD, time-integrated RF, and Antarctica and Greenland sulfate deposition for each of the 82 model simulations are included in Table S1 and available at: <https://catalogue.ceda.ac.uk/uuid/89cb57d9191b4b81965acb7f154c66df>.

**References**

Andersson, S. M., Martinsson, B. G., Vernier, J. P., Friberg, J., Brenninkmeijer, C. A. M., Hermann, M., et al. (2015). Significant radiative impact of volcanic aerosol in the lowermost stratosphere. *Nature Communications*, 6, 7692. <https://doi.org/10.1038/ncomms8692>

Andres, R. J., & Kasgnoc, A. D. (1998). A time-averaged inventory of subaerial volcanic sulfur emissions. *Journal of Geophysical Research*, 103(D19), 25251–25261. <https://doi.org/10.1029/98JD02091>

Andrianakis, I., & Challenor, P. G. (2012). The effect of the nugget on Gaussian process emulators of computer models. *Computational Statistics & Data Analysis*, 56(12), 4215–4228. <https://doi.org/10.1016/j.csda.2012.04.020>

Aquila, V., Oman, L. D., Stolarski, R. S., Colarco, P. R., & Newman, P. A. (2012). Dispersion of the volcanic sulfate cloud from a Mount Pinatubo-like eruption. *Journal of Geophysical Research*, 117, D06216. <https://doi.org/10.1029/2011jd016968>

Barnes, J. E., & Hofmann, D. J. (2001). Variability in the stratospheric background aerosol over Mauna Loa Observatory. *Geophysical Research Letters*, 28(15), 2895–2898. <https://doi.org/10.1029/2001gl013127>

Bastos, L. S., & O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics*, 51(4), 425–438. <https://doi.org/10.1198/TECH.2009.08019>

Bellouin, N., Mann, G. W., Woodhouse, M. T., Johnson, C., Carslaw, K. S., & Dalvi, M. (2013). Impact of the modal aerosol scheme GLOMAP-mode on aerosol forcing in the Hadley Centre Global Environmental Model. *Atmospheric Chemistry and Physics*, 13(6), 3027–3044. <https://doi.org/10.5194/acp-13-3027-2013>

Burke, A., Moore, K. A., Sigl, M., Nita, D. C., McConnell, J. R., & Adkins, J. F. (2019). Stratospheric eruptions from tropical and extra-tropical volcanoes constrained using high-resolution sulfur isotopes in ice cores. *Earth and Planetary Science Letters*, 521, 113–119. <https://doi.org/10.1016/j.epsl.2019.06.006>

Carn, S. A., Clarisse, L., & Prata, A. J. (2016). Multi-decadal satellite measurements of global volcanic degassing. *Journal of Volcanology and Geothermal Research*, 311, 99–134. <https://doi.org/10.1016/j.jvolgeores.2016.01.002>

Choi, W., Grant, W. B., Park, J. H., Lee, K.-M., Lee, H., & Russell, III, J. M. (1998). Role of the quasi-biennial oscillation in the transport of aerosols from the tropical stratospheric reservoir to midlatitudes. *Journal of Geophysical Research*, 103(D6), 6033–6042. <https://doi.org/10.1029/97jd03118>

Clausen, H. B., & Hammer, C. U. (1988). The Laki and Tambora eruptions as revealed in Greenland ice cores from 11 locations. *Annals of Glaciology*, 10, 16–22. <https://doi.org/10.1017/s0260305500004092>

Cole-Dai, J. H., Mosley-Thompson, E., Wight, S. P., & Thompson, L. G. (2000). A 4100-year record of explosive volcanism from an East Antarctica ice core. *Journal of Geophysical Research*, 105(D19), 24431–24441. <https://doi.org/10.1029/2000jd900254>

Crowley, T., & 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. <https://doi.org/10.5194/essd-5-187-2013>

Crowley, T., Zielinski, G., Vinther, B. M., Udisti, R., Kreutz, K., Cole-Dai, J., & Castellano, E. (2008). Volcanism and the little ice age. *PAGES Newsletter*, 16, 22–23. <https://doi.org/10.22498/pages.16.2.22>

- Devine, J. D., Sigurdsson, H., Davis, A. N., & Self, S. (1984). Estimates of sulfur and chlorine yield to the atmosphere from volcanic eruptions and potential climatic effects. *Journal of Geophysical Research*, *89*(B7), 6309–6325. <https://doi.org/10.1029/JB089iB07p06309>
- Dhomse, S. S., Emmerson, K. M., Mann, G. W., Bellouin, N., Carslaw, K. S., Chipperfield, M., et al. (2014). Aerosol microphysics simulations of the Mt. Pinatubo eruption with the UM-UKCA composition-climate model. *Atmospheric Chemistry and Physics*, *14*, 11221–11246. <https://doi.org/10.5194/acp-14-11221-2014>
- Dutton, E. G., & Christy, J. R. (1992). Solar radiative forcing at selected locations and evidence for global lower tropospheric cooling following the eruptions of El Chichón and Pinatubo. *Geophysical Research Letters*, *19*(23), 2313–2316. <https://doi.org/10.1029/92gl02495>
- Feinberg, A., Sukhodolov, T., Luo, B. P., Rozanov, E., Winkel, L. H. E., Peter, T., & Stenke, A. (2019). Improved tropospheric and stratospheric sulfur cycle in the aerosol–chemistry–climate model SOCOL-AERv2. *Geoscientific Model Development*, *12*(9), 3863–3887. <https://doi.org/10.5194/gmd-12-3863-2019>
- Ferris, D. G., Cole-Dai, J., Reyes, A. R., & Budner, D. M. (2011). South Pole ice core record of explosive volcanic eruptions in the first and second millennia A.D. and evidence of a large eruption in the tropics around 535 A.D. *Journal of Geophysical Research*, *116*, D17308. <https://doi.org/10.1029/2011JD015916>
- Gao, C., Oman, L., Robock, A., & Stenchikov, G. L. (2007). Atmospheric volcanic loading derived from bipolar ice cores: Accounting for the spatial distribution of volcanic deposition. *Journal of Geophysical Research*, *112*, D09109. <https://doi.org/10.1029/2006jd007461>
- 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. <https://doi.org/10.1029/2008jd010239>
- Gregory, J. M., Andrews, T., Good, P., Mauritsen, T., & Forster, P. M. (2016). Small global-mean cooling due to volcanic radiative forcing. *Climate Dynamics*, *47*(12), 3979–3991. <https://doi.org/10.1007/s00382-016-3055-1>
- Gregory, J. M., Bi, D., Collier, M. A., Dix, M. R., Hirst, A. C., Hu, A., et al. (2013). Climate models without preindustrial volcanic forcing underestimate historical ocean thermal expansion. *Geophysical Research Letters*, *40*(8), 1600–1604. <https://doi.org/10.1002/grl.50339>
- Guo, S., Rose, W. I., Bluth, G. J. S., & Watson, I. M. (2004). Particles in the great Pinatubo volcanic cloud of June 1991: The role of ice. *Geochemistry, Geophysics, Geosystems*, *5*, Q05003. <https://doi.org/10.1029/2003gc000655>
- Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., et al. (2005). Efficacy of climate forcings. *Journal of Geophysical Research*, *110*, D18104. <https://doi.org/10.1029/2005jd005776>
- Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., & Zorita, E. (2007). Detection of human influence on a new, validated 1500-year temperature reconstruction. *Journal of Climate*, *20*(4), 650–666. <https://doi.org/10.1175/jcli4011.1>
- Hegerl, G. C., Crowley, T. J., Hyde, W. T., & Frame, D. J. (2006). Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature*, *440*(7087), 1029–1032. <https://doi.org/10.1038/nature04679>
- Hitchman, M. H., McKay, M., & Trepte, C. R. (1994). A Climatology of Stratospheric Aerosol. *Journal of Geophysical Research*, *99*(D10), 20689–20700. <https://doi.org/10.1029/94jd01525>
- Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., & Pfister, L. (1995). Stratosphere-troposphere exchange. *Reviews of Geophysics*, *33*(4), 403–439. <https://doi.org/10.1029/95rg02097>
- Hommel, R., Timmreck, C., Giorgetta, M. A., & Graf, H. F. (2015). Quasi-biennial oscillation of the tropical stratospheric aerosol layer. *Atmospheric Chemistry and Physics*, *15*(10), 5557–5584. <https://doi.org/10.5194/acp-15-5557-2015>
- Johnson, J. S., Cui, Z., Lee, L. A., Gosling, J. P., Blyth, A. M., & Carslaw, K. S. (2015). Evaluating uncertainty in convective cloud microphysics using statistical emulation. *Journal of Advances in Modeling Earth Systems*, *7*(1), 162–187. <https://doi.org/10.1002/2014ms000383>
- Johnson, J. S., Gosling, J. P., & Kennedy, M. C. (2011). Gaussian process emulation for second-order Monte Carlo simulations. *Journal of Statistical Planning and Inference*, *141*(5), 1838–1848. <https://doi.org/10.1016/j.jspi.2010.11.034>
- Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., et al. (2017). The PMIP4 contribution to CMIP6-Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations. *Geoscientific Model Development*, *10*(11), 4005–4033. <https://doi.org/10.5194/gmd-10-4005-2017>
- Kleinschmitt, C., Boucher, O., Bekki, S., Lott, F., & Platt, U. (2017). The Sectional Stratospheric Sulfate Aerosol module (S3A-v1) within the LMDZ general circulation model: Description and evaluation against stratospheric aerosol observations. *Geoscientific Model Development*, *10*(9), 3359–3378. <https://doi.org/10.5194/gmd-10-3359-2017>
- Kravitz, B., & Robock, A. (2011). Climate effects of high-latitude volcanic eruptions: Role of the time of year. *Journal of Geophysical Research*, *116*, D01105. <https://doi.org/10.1029/2010jd014448>
- Lacis, A., Hansen, J., & Sato, M. (1992). Climate forcing by stratospheric aerosols. *Geophysical Research Letters*, *19*(15), 1607–1610. <https://doi.org/10.1029/92gl01620>
- Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., et al. (2010). Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmospheric Chemistry and Physics*, *10*, 7017–7039. <https://doi.org/10.5194/acp-10-7017-2010>
- Larson, E. J. L., & Portmann, R. W. (2016). A temporal Kernel method to compute effective radiative forcing in CMIP5 transient simulations. *Journal of Climate*, *29*(4), 1497–1509. <https://doi.org/10.1175/JCLI-D-15-0577.1>
- Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., & Spracklen, D. V. (2011). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, *11*(23), 12253–12273. <https://doi.org/10.5194/acp-11-12253-2011>
- LeGrande, A., Tsigaridis, K., & Bauer, S. (2016). Role of atmospheric chemistry in the climate impacts of stratospheric volcanic injections. *Nature Geoscience*, *9*, 652–655. <https://doi.org/10.1038/ngeo2771>
- Mann, G. W., Carslaw, K. S., Ridley, D. A., Spracklen, D. V., Pringle, K. J., Merikanto, J., et al. (2012). Intercomparison of modal and sectional aerosol microphysics representations within the same 3-D global chemical transport model. *Atmospheric Chemistry and Physics*, *12*(10), 4449–4476. <https://doi.org/10.5194/acp-12-4449-2012>
- Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., et al. (2010). Description and evaluation of GLOMAP-mode: A modal global aerosol microphysics model for the UKCA composition-climate model. *Geoscientific Model Development*, *3*(2), 519–551. <https://doi.org/10.5194/gmd-3-519-2010>
- Mann, G. W., Dhomse, S., Deshler, T., Timmreck, C., Schmidt, A., Neely, R., & Thomason, L. (2015). Evolving particle size is the key to improved volcanic forcings. *Past Global Changes Magazine*, *23*(2), 52–53. <https://doi.org/10.22498/pages.23.2.52>
- Marshall, L., Johnson, J. S., Mann, G. W., Lee, L., Dhomse, S. S., Regayre, L., et al. (2019). Exploring how eruption source parameters affect volcanic radiative forcing using statistical emulation. *Journal of Geophysical Research: Atmospheres*, *124*, 964–985. <https://doi.org/10.1029/2018JD028675>
- Marshall, L. R., Schmidt, A., Toohey, M., Carslaw, K. S., Mann, G. W., Sigl, M., et al. (2018). Multi-model comparison of the volcanic sulfate deposition from the 1815 eruption of Mt. Tambora. *Atmospheric Chemistry and Physics*, *18*(3), 2307–2328. <https://doi.org/10.5194/acp-18-2307-2018>

- Marshall, L. R., Smith, C. J., Forster, P. M., Aubry, T. J., Andrews, T., & Schmidt, A. (2020). Large variations in volcanic aerosol forcing efficiency due to eruption source parameters and rapid adjustments. *Geophysical Research Letters*, *47*, e2020GL090241. <https://doi.org/10.1029/2020GL090241>
- McConnell, J. R., Sigl, M., Plunkett, G., Burke, A., Kim, W. M., Raible, C. C., et al. (2020). Extreme climate after massive eruption of Alaska's Okmok volcano in 43 BCE and effects on the late Roman Republic and Ptolemaic Kingdom. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27), 15443–15449. <https://doi.org/10.1073/pnas.2002722117>
- Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1), 213–241. <https://doi.org/10.1007/s10584-011-0156-z>
- Metzner, D., Kutterolf, S., Toohey, M., Timmreck, C., Niemeier, U., Freundt, A., & Kruger, K. (2014). Radiative forcing and climate impact resulting from SO<sub>2</sub> injections based on a 200,000-year record of Plinian eruptions along the Central American Volcanic Arc. *International Journal of Earth Sciences*, *103*(7), 2063–2079. <https://doi.org/10.1007/s00531-012-0814-z>
- Mills, M. J., Richter, J. H., Tilmes, S., Kravitz, B., MacMartin, D. G., Glanville, A. A., et al. (2017). Radiative and chemical response to interactive stratospheric sulfate aerosols in fully coupled CESM1 (WACCM). *Journal of Geophysical Research: Atmospheres*, *122*, 13061–13078. <https://doi.org/10.1002/2017JD027006>
- Mills, M. J., Schmidt, A., Easter, R., Solomon, S., Kinnison, D. E., Ghan, S. J., et al. (2016). Global volcanic aerosol properties derived from emissions, 1990–2014, using CESM1(WACCM). *Journal of Geophysical Research: Atmospheres*, *121*, 2332–2348. <https://doi.org/10.1002/2015JD024290>
- Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, A. C., Johnson, C. E., Osprey, S. M., & Pyle, J. A. (2009). Evaluation of the new UKCA climate-composition model - Part 1: The stratosphere. *Geoscientific Model Development*, *2*(1), 43–57. <https://doi.org/10.5194/gmd-2-43-2009>
- Morris, M. D., & Mitchell, T. J. (1995). Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, *43*(3), 381–402. [https://doi.org/10.1016/0378-3758\(94\)00035-T](https://doi.org/10.1016/0378-3758(94)00035-T)
- Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., et al. (2013). Anthropogenic and natural radiative forcing. In T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., et al. (2014). Evaluation of the new UKCA climate-composition model—Part 2: The troposphere. *Geoscientific Model Development*, *7*(1), 41–91. <https://doi.org/10.5194/gmd-7-41-2014>
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, *91*(10–11), 1290–1300. <https://doi.org/10.1016/j.ress.2005.11.025>
- Oppenheimer, C. (2003). Ice core and palaeoclimatic evidence for the timing and nature of the great mid-13th century volcanic eruption. *International Journal of Climatology*, *23*(4), 417–426. <https://doi.org/10.1002/joc.891>
- Osprey, S. M., Gray, L. J., Hardiman, S. C., Butchart, N., & Hinton, T. J. (2013). Stratospheric variability in twentieth-century CMIP5 simulations of the met office climate model: High top versus low top. *Journal of Climate*, *26*(5), 1595–1606. <https://doi.org/10.1175/jcli-d-12-00147.1>
- Pinto, J. P., Turco, R. P., & Toon, O. B. (1989). Self-limiting physical and chemical effects in volcanic eruption clouds. *Journal of Geophysical Research*, *94*(D8), 11165–11174. <https://doi.org/10.1029/JD094iD08p11165>
- Punge, H. J., Konopka, P., Giorgetta, M. A., & Müller, R. (2009). Effects of the quasi-biennial oscillation on low-latitude transport in the stratosphere derived from trajectory calculations. *Journal of Geophysical Research*, *114*, D03102. <https://doi.org/10.1029/2008jd010518>
- Raible, C. C., Brönnimann, S., Auchmann, R., Brohan, P., Frölicher, T. L., Graf, H.-F., et al. (2016). Tambora 1815 as a test case for high impact volcanic eruptions: Earth system effects. *WIREs Climate Change*, *7*, 569–589. <https://doi.org/10.1002/wcc.407>
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.
- RCoreTeam. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Read, W. G., Froidevaux, L., & Waters, J. W. (1993). Microwave limb sounder measurement of stratospheric SO<sub>2</sub> from the Mt. Pinatubo Volcano. *Geophysical Research Letters*, *20*(12), 1299–1302. <https://doi.org/10.1029/93gl00831>
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. <https://doi.org/10.1175/2007jcli1824.1>
- Robock, A., & Free, M. P. (1995). Ice cores as an index of global volcanism from 1850 to the present. *Journal of Geophysical Research*, *100*(D6), 11549–11567. <https://doi.org/10.1029/95jd00825>
- Roustant, O., Ginsbourger, D., & Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, *51*(1), 1–55. <https://doi.org/10.18637/jss.v051.i01>
- Salter, J. M., & Williamson, D. (2016). A comparison of statistical emulation methodologies for multi-wave calibration of environmental models. *Environmetrics*, *27*(8), 507–523. <https://doi.org/10.1002/env.2405>
- Sato, M., Hansen, J. E., McCormick, M. P., & Pollack, J. B. (1993). Stratospheric aerosol optical depths, 1850–1990. *Journal of Geophysical Research*, *98*(D12), 22987–22994. <https://doi.org/10.1029/93jd02553>
- Scaillet, B. J., Luhr, F., & Carroll, M. R. (2003). Petrological and volcanological constraints on volcanic sulfur emissions to the atmosphere. In A. Robock, & C. Oppenheimer (Eds.), *Volcanism and the Earth's Atmosphere* (pp. 11–40). AGU. <https://doi.org/10.1029/139gm02>
- Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., et al. (2018). Volcanic radiative forcing from 1979 to 2015. *Journal of Geophysical Research: Atmospheres*, *123*(22), 12,491–12,508. <https://doi.org/10.1029/2018jd028776>
- Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., et al. (2011). Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geoscientific Model Development*, *4*(1), 33–45. <https://doi.org/10.5194/gmd-4-33-2011>
- Sekiya, T., Sudo, K., & Nagai, T. (2016). Evolution of stratospheric sulfate aerosol from the 1991 Pinatubo eruption: Roles of aerosol micro-physical processes. *Journal of Geophysical Research: Atmospheres*, *121*(6), 2911–2938. <https://doi.org/10.1002/2015JD024313>
- Self, S., Gertisser, R., Thordarson, T., Rampino, M. R., & Wolff, J. A. (2004). Magma volume, volatile emissions, and stratospheric aerosols from the 1815 eruption of Tambora. *Geophysical Research Letters*, *31*, L20608. <https://doi.org/10.1029/2004gl020925>
- Seviour, W. J. M., Butchart, N., & Hardiman, S. C. (2012). The Brewer-Dobson circulation inferred from ERA-Interim. *Quarterly Journal of the Royal Meteorological Society*, *138*(665), 878–888. <https://doi.org/10.1002/qj.966>
- Sigl, M., McConnell, J. R., Toohey, M., Curran, M., Das, S. B., Edwards, R., et al. (2014). Insights from Antarctica on volcanic forcing during the Common Era. *Nature Climate Change*, *4*(8), 693–697. <https://doi.org/10.1038/nclimate2293>
- Sigl, M., Winstrup, M., McConnell, J. R., Welten, K. C., Plunkett, G., Ludlow, F., et al. (2015). Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature*, *523*(7562), 543–549. <https://doi.org/10.1038/nature14565>

- Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., & Regayre, L. A. (2018). FAIR v1.3: A simple emissions-based impulse response and carbon cycle model. *Geoscientific Model Development*, *11*(6), 2273–2297. <https://doi.org/10.5194/gmd-11-2273-2018>
- Staunton-Sykes, J., Aubry, T. J., Shin, Y. M., Weber, J., Marshall, L. R., Abraham, N. L., et al. (2021). Co-emission of volcanic sulfur and halogens amplifies volcanic effective radiative forcing. *Atmospheric Chemistry and Physics*, *21*, 9009–9029. <https://doi.org/10.5194/acp-21-9009-2021>
- Stoffel, M., Khodri, M., Corona, C., Guillet, S., Poulain, V., Bekki, S., et al. (2015). Estimates of volcanic-induced 640 cooling in the Northern Hemisphere over the past 1,500 years. *Nature Geoscience*, *8*(10), 784–788. <https://doi.org/10.1038/ngeo2526>
- Stothers, R. B. (1984). The great Tambora eruption in 1815 and its aftermath. *Science*, *224*(4654), 1191–1198. <https://doi.org/10.1126/science.224.4654.1191>
- Stothers, R. B. (1996). Major optical depth perturbations to the stratosphere from volcanic eruptions: Pyrheliometric period, 1881–1960. *Journal of Geophysical Research*, *101*(D2), 3901–3920. <https://doi.org/10.1029/95jd03237>
- Stothers, R. B. (2001). Major optical depth perturbations to the stratosphere from volcanic eruptions: Stellar extinction period, 1961–1978. *Journal of Geophysical Research*, *106*(D3), 2993–3003. <https://doi.org/10.1029/2000jd900652>
- Sukhodolov, T., Sheng, J. X., Feinberg, A., Luo, B. P., Peter, T., Revell, L., et al. (2018). Stratospheric aerosol evolution after Pinatubo simulated with a coupled size-resolved aerosol–chemistry–climate model, SOCOL-AERv1.0. *Geoscientific Model Development*, *11*(7), 2633–2647. <https://doi.org/10.5194/gmd-11-2633-2018>
- Timmreck, C., Graf, H. F., Lorenz, S. J., Niemeier, U., Zanchettin, D., Matei, D., et al. (2010). Aerosol size confines climate response to volcanic super-eruptions. *Geophysical Research Letters*, *37*, L24705. <https://doi.org/10.1029/2010gl045464>
- Timmreck, C., Mann, G. W., Aquila, V., Hommel, R., Lee, L. A., Schmidt, A., et al. (2018). The Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP): Motivation and experimental design. *Geoscientific Model Development*, *11*(7), 2581–2608. <https://doi.org/10.5194/gmd-11-2581-2018>
- Toohey, M., Kruger, K., Bittner, M., Timmreck, C., & Schmidt, H. (2014). The impact of volcanic aerosol on the Northern Hemisphere stratospheric polar vortex: Mechanisms and sensitivity to forcing structure. *Atmospheric Chemistry and Physics*, *14*(23), 13063–13079. <https://doi.org/10.5194/acp-14-13063-2014>
- Toohey, M., Kruger, K., Niemeier, U., & Timmreck, C. (2011). The influence of eruption season on the global aerosol evolution and radiative impact of tropical volcanic eruptions. *Atmospheric Chemistry and Physics*, *11*(23), 12351–12367. <https://doi.org/10.5194/acp-11-12351-2011>
- Toohey, M., Kruger, K., Sigl, M., Stordal, F., & Svensen, H. (2016). Climatic and societal impacts of a volcanic double event at the dawn of the Middle Ages. *Climatic Change*, *136*(3–4), 401–412. <https://doi.org/10.1007/s10584-016-1648-7>
- Toohey, M., Kruger, K., & Timmreck, C. (2013). Volcanic sulfate deposition to Greenland and Antarctica: A modeling sensitivity study. *Journal of Geophysical Research: Atmospheres*, *118*(10), 4788–4800. <https://doi.org/10.1002/jgrd.50428>
- Toohey, M., & Sigl, M. (2017). Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE. *Earth System Science Data*, *9*(2), 809–831. <https://doi.org/10.5194/essd-9-809-2017>
- Toohey, M., Stevens, B., Schmidt, H., & Timmreck, C. (2016). Easy Volcanic Aerosol (EVA v1.0): An idealized forcing generator for climate simulations. *Geoscientific Model Development*, *9*(11), 4049–4070. <https://doi.org/10.5194/gmd-9-4049-2016>
- Traufetter, F., Oerter, H., Fischer, H., Weller, R., & Miller, H. (2004). Spatio-temporal variability in volcanic sulphate deposition over the past 2 kyr in snow pits and firn cores from Amundsenisen, Antarctica. *Journal of Glaciology*, *50*(168), 137–146. <https://doi.org/10.3189/172756504781830222>
- Vidal, C. M., Metrich, N., Komorowski, J. C., Pratomo, I., Michel, A., Kartadinata, N., et al. (2016). The 1257 Samalas eruption (Lombok, Indonesia): The single greatest stratospheric gas release of the Common Era. *Scientific Reports*, *6*, 34868. <https://doi.org/10.1038/srep34868>
- Wigley, T. M. L., Ammann, C. M., Santer, B. D., & Raper, S. C. B. (2005). Effect of climate sensitivity on the response to volcanic forcing. *Journal of Geophysical Research*, *110*, D09107. <https://doi.org/10.1029/2004jd005557>
- Williamson, D., Blaker, A. T., Hampton, C., & Salter, J. (2015). Identifying and removing structural biases in climate models with history matching. *Climate Dynamics*, *45*(5–6), 1299–1324. <https://doi.org/10.1007/s00382-014-2378-z>
- Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., et al. (2016). The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): Experimental design and forcing input data for CMIP6. *Geoscientific Model Development*, *9*(8), 2701–2719. <https://doi.org/10.5194/gmd-9-2701-2016>
- Zanchettin, D., Timmreck, C., Toohey, M., Jungclaus, J. H., Bittner, M., Lorenz, S. J., & Rubino, A. (2019). Clarifying the relative role of forcing uncertainties and initial-condition unknowns in spreading the climate response to volcanic eruptions. *Geophysical Research Letters*, *46*(3), 1602–1611. <https://doi.org/10.1029/2018gl081018>