A new volcanic stratospheric sulfate aerosol forcing emulator (EVA_H): Comparison with interactive stratospheric aerosol models.

Thomas J. Aubry[1*], Matthew Toohey[2†], Lauren Marshall[3], Anja Schmidt[1,3], A. Mark Jellinek [4]

[1]University of Cambridge, Department of Geography, Cambridge, United Kingdom
[2]GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany
[3]University of Cambridge, Department of Chemistry, Cambridge, United Kingdom
[4]University of British Columbia, Department of Earth, Ocean and Atmospheric Sciences, Vancouver, Canada

Key Points:

- We present EVA_H, an observationally constrained volcanic forcing emulator accounting for the mass, latitude and altitude of erupted SO_2.
- Accounting for the altitude of volcanic injections improves predictions of subsequent perturbations of stratospheric optical properties.
- EVA_H shows enhanced consistency with interactive stratospheric aerosol models, but still lacks sensitivity to eruption latitude.

*Previously at University of British Columbia, Department of Earth, Ocean and Atmospheric Sciences, Vancouver, Canada
†Now at: Institute of Space and Atmospheric Studies, University of Saskatchewan, Saskatoon, Canada.

Corresponding author: Thomas J. Aubry, ta460@cam.ac.uk

Abstract

Idealized models or emulators of volcanic aerosol forcing have been widely used to reconstruct the spatio-temporal evolution of past volcanic forcing. However, existing models, including the most recently developed Easy Volcanic Aerosol (EVA, Toohey et al. (2016)): i) do not account for the height of injection of volcanic SO_2; ii) prescribe a vertical structure for the forcing; and iii) are often calibrated against a single eruption.

We present a new idealized model, EVA_H, that addresses these limitations. Compared to EVA, EVA_H makes predictions of the global mean stratospheric aerosol optical depth that are: i) similar for the 1979-1998 period characterized by the large and high-altitude tropical SO_2 injections of El Chichón (1982) and Mt. Pinatubo (1991); ii) significantly improved for the 1998-2015 period characterized by smaller eruptions with a large variety of injection latitudes and heights. Compared to EVA, the sensitivity of volcanic forcing to injection latitude and height in EVA_H is much more consistent with results from climate models that include interactive aerosol chemistry and microphysics, even though EVA_H remain less sensitive to eruption latitude than the latter models.

We apply EVA_H to investigate potential biases and uncertainties in EVA-based volcanic forcing datasets from phase 6 of the Coupled Model Intercomparison Project (CMIP6). EVA and EVA_H forcing reconstructions do not significantly differ for tropical high-altitude volcanic injections. However, for high-latitude or low altitude injections, our reconstructed forcing is significantly lower. This suggests that volcanic forcing in CMIP6 last millenium experiments may be overestimated for such eruptions.

1 Introduction

Stratospheric volcanic sulfate aerosol radiative forcing (volcanic forcing hereafter) is a major driver of Earth's climate variability. Volcanic eruptions can inject sulfur dioxide (SO_2) into the stratosphere and form long-lived (1-3 years) sulfate aerosol that modify Earth's radiative balance, causing a net cooling at the surface and affecting major modes of climate variability (e.g. Robock (2000); Timmreck (2012); Kremser et al. (2016)). Recently, it has emerged that even relatively small eruptions (injecting less than around 1 teragram (Tg) of SO_2) of the early 21^{st} century exert small but significant radiative forcing (e.g. Schmidt et al. (2018)) and have a statistically discernible cooling effect on sea surface and tropospheric temperatures (Santer et al., 2015).

Models are key tools to reconstruct past volcanic impacts on climate and societies, as well as to predict the impacts of future volcanic eruptions. Interactive stratospheric aerosol models (e.g. Timmreck et al. (2018)) predict the full life cycle of volcanic sulfate aerosol, and the associated radiative and climate response following an injection of volcanic SO_2 into the atmosphere. However, there is a large spread among the forcing predicted by these models for a specified volcanic SO_2 injection (e.g. Zanchettin et al. (2016)). This inter-model uncertainty adds to intra-model uncertainties as well as uncertainties related to constraining eruption source parameters, e.g., the mass of SO_2 and eruption latitude reconstructed from ice cores when investigating the impacts of past eruptions (Toohey & Sigl, 2017; Marshall et al., 2018). Given the computational cost of interactive stratospheric aerosol models, exploring how the propagation of model and source parameter uncertainties affect the assessment of the climate response to a volcanic eruption is challenging and requires significant efforts such as model intercomparison exercises (e.g. Zanchettin et al. (2016); Timmreck et al. (2018)).

Another class of models consist of idealized models or "emulators" of volcanic aerosol evolution which have been developed to reproduce the spatiotemporal evolution of volcanic aerosol and associated perturbations of atmospheric optical properties, e.g. using constraints from ice-cores on the timing and mass of sulfur injected by past eruptions (e.g. Grieser and Schonwiese (1999); Amman et al. (2003); Gao et al. (2008); Crowley

and Unterman (2013); Toohey and Sigl (2017)) or scenarios of future eruptions (Ammann & Naveau, 2010; Bethke et al., 2017). Grieser and Schonwiese (1999), Amman et al. (2003), Gao et al. (2008) and Toohey and Sigl (2017) use emulators based on box models, where each box corresponds to a latitudinal region of the stratosphere. For a prescribed sulfur injection in one of the boxes, the evolution of the mass of sulfate aerosol is governed by timescale(s) for: i) the production of sulfate from SO_2; ii) the mixing between the boxes; and iii) the loss of aerosol to the troposphere. Aerosol properties like stratospheric aerosol optical depth (SAOD) and effective radius are scaled from the mass of sulfate in the boxes. These models generally rely on only a few parameters and are computationally inexpensive so that conducting sensitivity studies to explore uncertainty propagation is straightforward.

The most recently developed idealized model of volcanic forcing is the Easy Volcanic Aerosol model (EVA, Toohey et al. (2016)). Recent reconstructions of volcanic aerosol properties following the Mt. Pinatubo 1991 eruption were used to calibrate the model. EVA also used Gaussian shape functions to produce a realistic latitudinal distribution of extinction whereas most previous models had step-like latitudinal distributions. However, like all idealized models of volcanic forcing currently available, EVA has two important limitations:

1. The vertical structure of the forcing produced by the model does not depend on characteristics of volcanic sulfur injections, in particular plume height.
2. It is calibrated using data from the 1991 Mt. Pinatubo eruption. Given the sensitivity of the relationship between the erupted sulfur mass and the subsequent volcanic forcing on eruption source parameters (such as the latitude or altitude of injection, e.g. Marshall et al. (2019), Toohey et al. (2019)), one should be careful when applying this model to other eruptions. In particular, most eruptions whose plume reaches the stratosphere inject order(s) of magnitude less sulfur than Mt. Pinatubo, with injections between 10 and 20 km altitude (instead of ca. 20-25km for Mt. Pinatubo), and commonly occur in high latitudes instead of the tropics (Carn et al., 2016).

Consequently, the major objective of this study is to extend the EVA methodology to develop EVA_H (with "H" standing for height), an idealized model of volcanic aerosol forcing: i) accounting for plume height to determine the forcing resulting from a sulfur injection; ii) predicting the vertical structure of aerosol extinction; and iii) calibrated against eruptions spanning a large range of mass of erupted sulfur, plume height and latitude. We compare outputs of EVA_H to EVA and to interactive stratospheric aerosol models. We also provide example applications to improve reconstructions of past volcanic forcing and provide fast response to present/future eruptions.

2 Data and model

2.1 Data

Primary datasets used to calibrate the model

Our strategy is to calibrate the model so that its output best reproduces observations of atmospheric optical properties given an input inventory of volcanic sulfur emission estimates. For optical properties, we use the Global Space-based Stratospheric Aerosol Climatology (GloSSAC, version 1.1, Thomason et al. (2018)), which is the National Aeronautics and Space Administration latest reconstruction of extinction from satellite data. It contains latitude and altitude dependent extinction at 525 nm from 1979 to 2016. Typical uncertainties on extinction coefficients are about 10% (Thomason et al., 2018), although uncertainties associated with the processing and combination of the various observational datasets used in GloSSAC remains to be precisely quantified. In addition,

whereas 1984-2005 climatological tropopause height from the Modern-Era Retrospective Analysis for Research and Applications (MERRA, Rienecker et al. (2011)) are provided with the GloSSAC dataset, we use time-varying tropopause height from the NCEP/NCAR reanalysis (Kalnay et al., 1996). This enables us to account for trends related to climate change (Santer et al., 2003) and the large variability of tropopause height at high latitudes when calculating stratospheric aerosol optical depths (see Figure S1 for a comparison of GloSSAC versions and tropopause height treatment).

For the volcanic sulfur emission inventory, we use data reported by Carn et al. (2016) who report the date, location, mass of SO_2 and altitude of volcanic emissions over 1978-2015. Typical uncertainties for the total mass of SO_2 injected by an eruption range from 20% to up to 50-100% (Carn et al., 2016), while typical uncertainties on the injection height are up to 20% (e.g. Carboni et al. (2016); Aubry et al. (2017)).

2.2 Model structure

The new model, EVA_H, maintains the overall approach of EVA (Toohey et al., 2016), i.e.:

- The global mean SAOD at 525 nm and effective radius are scaled from the total mass of SO_4 (Section 2.4 and 3.4).
- Transport equations govern the production, transfer and loss of SO_4 among the model grid boxes (Section 2.3).
- The latitudinal and vertical distribution of extinction is produced using the distribution of SO_4 mass in the model boxes and 2D shape functions (Section 3.3).
- Wavelength-dependent extinction, single scattering albedo, and scattering asymmetry factor are calculated from the effective radius and extinction at 525 nm using Mie theory (Section 3.4).

EVA separates the stratosphere into three latitudinal bands (southern extratropics, tropics and northern extratropics) which is consistent with respect to the structure of the Brewer-Dobson circulation (e.g. Plumb (1996); Neu and Plumb (1999); Butchart (2014)). To add a vertical dimension while maintaining the simplified approach of a box model, we use three vertical bands:

- The lowermost extratropical stratosphere ($\leqslant$ 16 km), where cross-tropopause mixing and transport at mid-latitudes is an important control on the transport of aerosols between the stratosphere and the troposphere.
- The lower stratosphere (16-20 km) where aerosols in the tropics may be transported directly into the lowermost extra-tropical stratosphere due to the latitudinal dependence of isolines of potential temperature.
- The middle stratosphere ($\geq$ 20 km).

The proposed structure including three latitudinal and three vertical bands results in an "8-box" model (Figure 1) if we keep only stratospheric boxes and exclude the uppermost tropical troposphere. To be consistent with the grid of the GloSSAC data, against which the model will be calibrated, the top of the model is at 39.5km altitude, and the tropical boxes comprise latitudes $\leqslant 22.5°$.

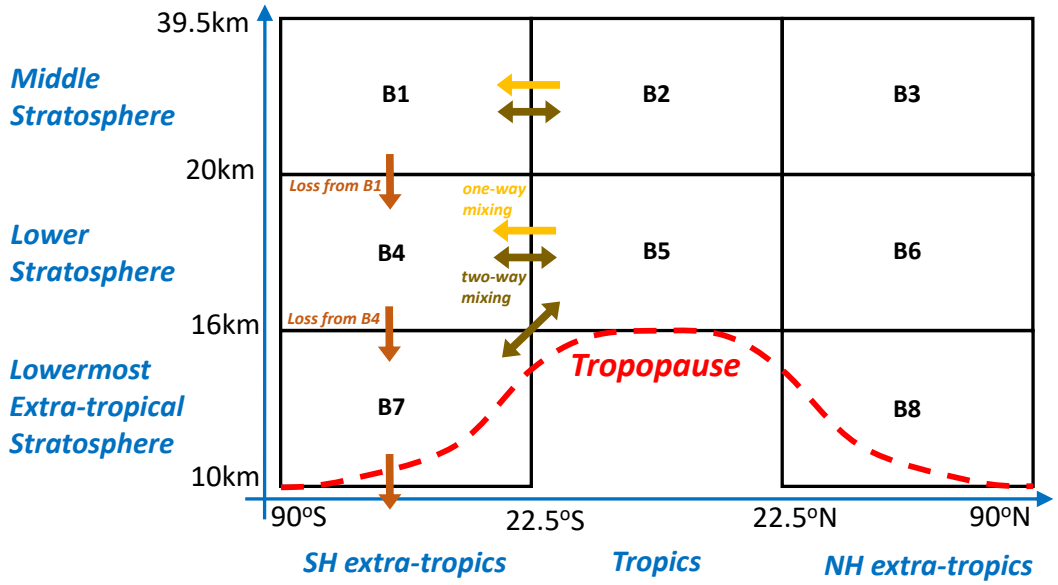

Figure 1. Schematic showing the 8 boxes of EVA_H, and their approximate positions relative to the tropopause sketched by the red dashed line. The boxes are indexed from top to bottom and South to North. Arrows represent examples of SO_4 fluxes from and into the southern hemisphere boxes (boxes 1,4 and 7). The vertical axis is not to scale.

2.3 Evolution of sulfate mass in the model boxes

The equations governing the evolution of the mass of sulfur in a model box will follow the approach of EVA, adapted to the new 2-dimensional structure of EVA_H. The calibration of all parameters involved in the equations presented throughout Section 2 is detailed in Section 3. We assume that the evolution of the mass of SO_2 in a box i (see Figure 1 for box indices) $M^i_{SO_2}$ is governed by the equation:

$$\frac{dM^i_{SO_2}}{dt} = S_i - \frac{M^i_{SO_2}}{\tau^i_{prod}} \quad , \tag{1}$$

where S_i is a source term, and τ^i_{prod} is an effective timescale for the conversion of SO_2 into sulfate aerosols. Accordingly, the production of SO_4 in a box i will be of the form:

$$\mathrm{PROD} = \frac{\mathrm{M^i_{SO_2}}}{\tau^{\mathrm{i}}_{\mathrm{prod}}} = \frac{\mathrm{M^i_{vSO_2}}}{\tau^{\mathrm{i}}_{\mathrm{prod}}} + \mathrm{B_i} \quad , \tag{2}$$

where the mass of SO_2 in a box i is decomposed into the mass from volcanic injections $M^i_{vSO_2}$ and a flux B_i, assumed constant, corresponding to background non-volcanic sulfur injections.

We assume that two-way mixing can occur between two adjacent boxes belonging to the same vertical band and/or between the lower tropical stratosphere (box 5) and the lowermost extratropical stratosphere (boxes 7 and 8). The two-way mixing flux from a box i to a box j is proportional to the SO_4 mass difference between the boxes.:

$$\mathrm{MIXING} = \frac{\mathrm{M^i_{SO_4}} - \mathrm{M^j_{SO_4}}}{\tau^{\mathrm{ij}}_{\mathrm{mix}}} \quad , \tag{3}$$

where τ^{ij}_{mix} is a mixing timescale.

As for two-way mixing, we assume that one-way mixing, i.e. residual transport, can happen between two adjacent boxes belonging to the same vertical band and/or between box 5 and boxes 7 and 8. The one-way mixing (OWM) flux from a box i to a box j is proportional to the mass of SO_4 in box i:

$$\mathrm{OWM} = \frac{\mathrm{M^{i}_{SO_4}}}{\tau^{\mathrm{ij}}_{\mathrm{owm}}} \quad , \tag{4}$$

where τ^{ij}_{owm} is a one-way mixing timescale. In EVA, one-way mixing terms are used to represent the residual Brewer-Dobson circulation from the tropics to the extra-tropics not accounted for in the two-way mixing terms.

We assume that the loss of aerosol in box i is proportional to the mass of SO_4 in the same box:

$$\mathrm{LOSS} = -\frac{\mathrm{M^{i}_{SO_4}}}{\tau^{\mathrm{i}}_{\mathrm{loss}}} \quad , \tag{5}$$

where τ^{i}_{loss} is a loss timescale. In EVA_H, we assume that the SO_4 loss flux from a box that is not in contact with the tropopause (i.e., all boxes except boxes 5, 7 and 8) corresponds to a positive flux for the box located directly below. For example, the loss term in box 1, $-\frac{M^{1}_{SO_4}}{\tau^{1}_{loss}}$, corresponds to a flux $+\frac{M^{1}_{SO_4}}{\tau^{1}_{loss}}$ in box 4.

The general equation governing the evolution of aerosol mass $M^{i}_{SO_4}$ in one of the eight boxes i will then be:

$$\frac{dM^{i}_{SO_4}}{dt} = \mathrm{PROD} + \mathrm{MIXING} + \mathrm{OWM} + \mathrm{LOSS} \quad , \tag{6}$$

where the production term PROD is governed by Equation 2, two-way and one-way mixing term(s) MIXING and OWM are governed by Equation 3 and 4 respectively, and the loss term LOSS is governed by equation 5 and may include positive terms related to the loss of aerosols in the box located above box i (e.g. for box 4, cf. fluxes illustrated on Figure 1). Note that timescales τ_{loss}, τ_{mix} and τ_{owm} are not physical timescales and depend on the geometry (e.g. thickness) of the 8 boxes of the model.

The final configuration of the model depends on the following choices:

1. Between which boxes to include two-way and one-way mixing terms
2. The dependence of the timescales τ_{prod}, τ_{loss}, τ_{mix} and τ_{owm} on latitude, altitude, and season.

We further discuss these choices in Section 3.2.

2.4 Scaling for the stratospheric aerosol optical depth

The calibration of the model requires linking the model primary output (i.e. the mass of sulfate in each box) to optical properties that can be directly observed. Following previous studies (e.g. Gao et al. (2008); Crowley and Unterman (2013); Toohey et al. (2016)), we assume that the relationship between the global mean SAOD at 525nm ($SAOD_{525}$) and the total stratospheric SO_4 burden M_{SO4} is adequately represented by a power-law scaling:

$$SAOD_{525} = A \times M^{\alpha}_{SO4} \quad , \tag{7}$$

where α is an exponent and A is a prefactor. In contrast with previous studies (e.g. Gao et al. (2008), Toohey et al. (2016)) , we use observations from a large number of eruptions (19 eruptions with sulfur mass ranging from ca. 10^{-2} to 10^{1} Tg S, latitude from

41°S to 50°N and height from 12 to 25 km) and simulations from interactive stratospheric aerosol models to constrain the exponent (α) of this scaling:

1. Limited direct observational measurements of the stratospheric SO_4 burden exist. Consequently, we identified all SAOD peaks in the 1979-2016 GloSSAC SAOD timeseries, smoothed over 6 months to avoid peaks related to non-volcanic signals. We then defined corresponding SAOD increases by removing the minimum SAOD value between two peaks from the second peak value. We defined the associated SO_2 loading as the mass of sulfur - taken from Carn et al. (2016) - injected by eruptions which occurred no earlier than one month before the minimum SAOD value and no later than one month before the SAOD peak. The chosen one-month lags excludes eruptions for which most SO_2 would likely not been transformed into sulfate aerosols (Toohey et al., 2016). We filtered eruptions for which $H^* = \frac{\text{SO}_2 \text{ inj. height}}{\text{tropopause height}} \leq 1$. Last, we fit SAOD increases as a function of corresponding stratospheric SO_2 injections using a power law (Figure 2.a). We find an exponent of 1 ± 0.2.
2. We use the 1979-2015 experiments run with the Community Earth System Model version 1 with a prognostic aerosol scheme (Whole Atmosphere Community Climate Model, WACCM) using the Neely and Schmidt (2016) volcanic sulfur emission inventory (Mills et al., 2016; Schmidt et al., 2018), with adjusted mass of 10 Tg of SO_2 (instead of 18 Tg in Neely and Schmidt (2016)) and height of 18-20 km (instead of 23-27 km in Neely and Schmidt (2016)) for the 1991 eruption of Mt. Pinatubo. We fit the monthly mean values of global mean SAOD anomaly (i.e. the difference between runs with and without volcanic emissions) at 550 nm to the stratospheric SO_4 burden anomaly using a power-law fit and find an exponent of 1.01 ± 0.01 (Figure 2.b).
3. We use 30 experiments from the MAECHAM5-HAM interactive stratospheric aerosol model, where 8.5 TgS were injected at 6 different set of altitudes and latitudes (Toohey et al., 2019). We fit the monthly mean values of global mean SAOD anomalies at 550 nm to the total stratospheric SO_4 burden anomaly using a power-law fit and find an exponent of 0.84 ± 0.03 (Figure 2.c).
4. We use 41 experiments from the UM-UKCA interactive stratospheric aerosol model, where injection mass, altitude and latitude were varied between 5-50 Tg S, 15-25 km and 80°S-80°N, respectively (Marshall et al., 2019). We fit the monthly mean values of global mean SAOD anomalies at 550 nm to the total stratospheric SO_4 burden anomaly for burden ≤ 10 TgS using a power-law fit and find an exponent of 0.89 ± 0.02 (Figure 2.d).

In agreement with scaling used in previous studies (e.g. Crowley and Unterman (2013); Toohey et al. (2016)), observations and the WACCM run with historical volcanic emission (Figure 2 a-b) suggest that a linear scaling between the stratospheric sulfur burden and the global mean SAOD holds for eruptions of the 1979-2015 period, i.e. for eruptions injecting on the order of or less SO_2 than the 1991 eruption of Mt. Pinatubo ($\simeq$ 9 TgS). However, the observational constraint on α should be considered carefully as it was not derived from an observed relationship between monthly $SAOD_{525}$ and M_{SO4}. It is also very sensitive to the set of eruptions included, with for example a value of $2.3\pm$ 0.8 when excluding the 1991 eruption of Mt. Pinatubo. The two sets of interactive stratospheric aerosol model simulations used here suggest that the value of the exponent to be used in Equation 7 should be close to ca. 0.86 for stratospheric sulfate burdens up to 10 TgS (Figure 2 c-d). Given the proximity of this value to 1 and for simplicity, we will use a linear scaling to calibrate all model parameters in Section 3 - including the prefactor A in Equation 7 - using 1979-2015 observational datasets (Carn et al. (2016) and Thomason et al. (2018)). However, our analysis shows that the assumption of a linear scaling between the mass of sulfate and SAOD should be considered with caution, even for relatively small stratospheric burdens (on the order of those following the Mt. Pinatubo 1991 eruption).

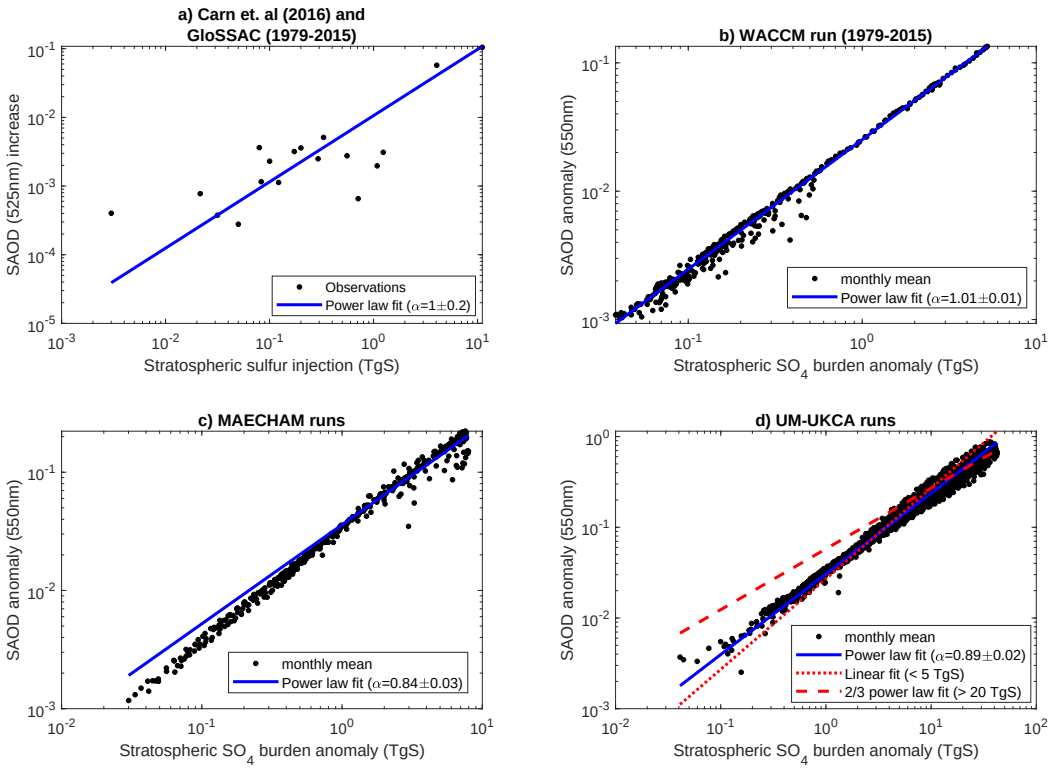

Figure 2. a) Global mean SAOD increase (GloSSAC) as a function of corresponding stratospheric SO_2 loadings (Carn et al. (2016)). b) Global mean SAOD anomaly as a function of the global stratospheric SO_4 burden anomaly in WACCM 1979-2015 run (Schmidt et al., 2018). c) Same as (b) but for MAECHAM's runs (Toohey et al., 2019) and using global mean SAOD anomaly. d) Same as (c) but using UM-UKCA's runs (Marshall et al., 2019). Blue lines show best power law fits for sulfate burden up to 10 TgS, with the exponent α reported in legends. For panel (d), the red dotted line shows a linear fit for burdens smaller than 5 TgS, while the red dashed line shows a 2/3 power law fit for burdens larger than 20 TgS.

For large SO_2 injections, previous studies have suggested that the relationship between the sulfate burden and the SAOD follows a 2/3 power-law (Timmreck et al., 2010; Crowley & Unterman, 2013; Metzner et al., 2014; Toohey et al., 2016), although the critical mass above which a non-linear scaling should apply has been suggested to be as low as 2.5 Tg S (Metzner et al., 2014) and as high as 30 TgS (Toohey et al., 2016). Here we take advantage of the recent simulations of Marshall et al. (2019), with sulfur burdens of up to 50 TgS and a large variety of eruption source parameters, to revisit these results. We perform a linear fit of SAOD vs sulfate burden for burdens $\leq$ 5 TgS, and a 2/3 power-law fit for burdens $\geq$ 20 TgS. These fits are shown on Figure 2.d and intersect for a burden of 10 TgS, which we choose as the threshold sulfate burden M^* above which to apply a 2/3 scaling. This estimate falls in the large range of thresholds previously estimated. Note that when fitting SAOD to sulfate burdens larger than 20 TgS using a power-law fit without a prescribed exponent, we find an exponent of 0.72±0.12 which is compatible although a bit larger than the usually suggested 2/3 power-law. The final scaling we adopt for SAOD at 525 nm in EVA_H is thus

$$SAOD_{525} = \begin{cases} A \times M_{SO4} & \text{if } M_{SO4} \leq M^* \\ A \times (M^*)^{1/3} \times M_{SO4}^{2/3} & \text{if } M_{SO4} > M^* \end{cases}, \quad (8)$$

with $M^* = 10\ TgS$ and where the pre-factor $A \times (M^*)^{1/3}$ for the 2/3 scaling guarantees the continuity at $M_{SO4} = M^*$.

2.5 Volcanic SO_2 injection in the model

The Carn et al. (2016) dataset provides the latitude, date, estimated mass of SO_2 and estimated height for each reported volcanic SO_2 injection into the atmosphere. A simple method to include SO_2 in the 8-box model is to inject the entire mass into the box which contains the point defined by the eruption latitude and estimated injection height. However, in the absence of a transport equation for SO_2 in the model, a more realistic approach may be to distribute the SO_2 spatially instead of injecting 100% of the mass in a single box. To determine the spatial distribution of injected SO_2, we investigated patterns of extinction increase in GloSSAC for the first five months following eruptions from the Carn et al. (2016) dataset (see supporting information S1 and Figure S2). We found that the latitudinal and vertical positions of regions of initial extinction increase are in good agreement with the injection latitude and altitude reported in Carn et al. (2016) (Figure S3), and have average extents of 1.2 km and 7^o in height and latitude respectively (Figure S4). Accordingly, in EVA_H, we distribute the SO_2 mass injection among the boxes using Gaussian distributions centered on latitude and altitude estimates from Carn et al. (2016), with widths of 7^o and 1.2 km.

3 Calibration of the model

3.1 Overview of the calibration process

The linear scaling for the global mean SAOD for eruptions injecting less than 10 TgS, in particular all eruptions of the 1979-2015 period, can be written $\sum_{i=1}^{8} w^i AOD^i = A \times \sum_{i=1}^{8} M^i_{SO_4}$ where A is the same prefactor as in Equation 8, AOD^i is the spatially averaged AOD in a box i (i.e., extinction integrated from the lower to the upper vertical boundary of the box), and w^i are weights calculated from the latitudinal extent of each box. For the 1979-2015 calibration period, each box thus follows the scaling $w^i \times AOD^i = A \times M^i_{SO_4}$. To calibrate the box model, we substitute $M^i_{SO_4}$ by $\frac{w^i AOD^i}{A}$ in equation 6. Next, assuming that production timescales τ^i_{prod} are independent of season, the mass of SO_2 $M^i_{vSO_2}$ in a box i of the model at any time t is given by:

$$M^i_{vSO_2}(t) = \sum_{k, t_k \leq t} M^i_k e^{-\frac{t-t_k}{\tau^i_{prod}}} \quad , \tag{9}$$

where k is an index representing eruptions in the Carn et al. (2016) dataset, t_k the date of the k^{th} eruption, and M^i_k the mass of SO_2 injected by eruption k in box i calculated as described in Section 2.5. Consequently, to calibrate the model, we simply calculate model predicted monthly timeseries of weighted AOD ($wAOD^i_{mod}$) in each box over 1979-2015 using the Carn et al. (2016) SO_2 inventory, and find the set of model parameter values minimizing our chosen error metric E:

$$E = \sqrt{\sum_{t=1979}^{2015} \sum_{i=1}^{8} (wAOD^i_{mod} - wAOD^i_{obs})^2} \quad , \tag{10}$$

where $wAOD^i_{obs}(t)$ are the observed timeseries calculated from GloSSAC (Thomason et al., 2018). E is a root mean squared error (RMSE) on AOD calculated over all time steps and all boxes. Figure 3 shows the corresponding SO_2 inputs and $wAOD_{obs}$ timeseries in the 8 model boxes. To calculate E, we run the model without a non-volcanic background injection (terms B_i in Equation 1), and compare its output with $wAOD_{obs}$ time-

series from which we substract a non-volcanic background (black dashed lines on Figure 3). We define this background as the 1999-2002 average because this period has the lowest stratospheric volcanic SO_2 injections in the post-Pinatubo era (e.g. Carn et al. (2016); Schmidt et al. (2018)). We come back to the inclusion and calibration of background injections in the model in Section 3.2.

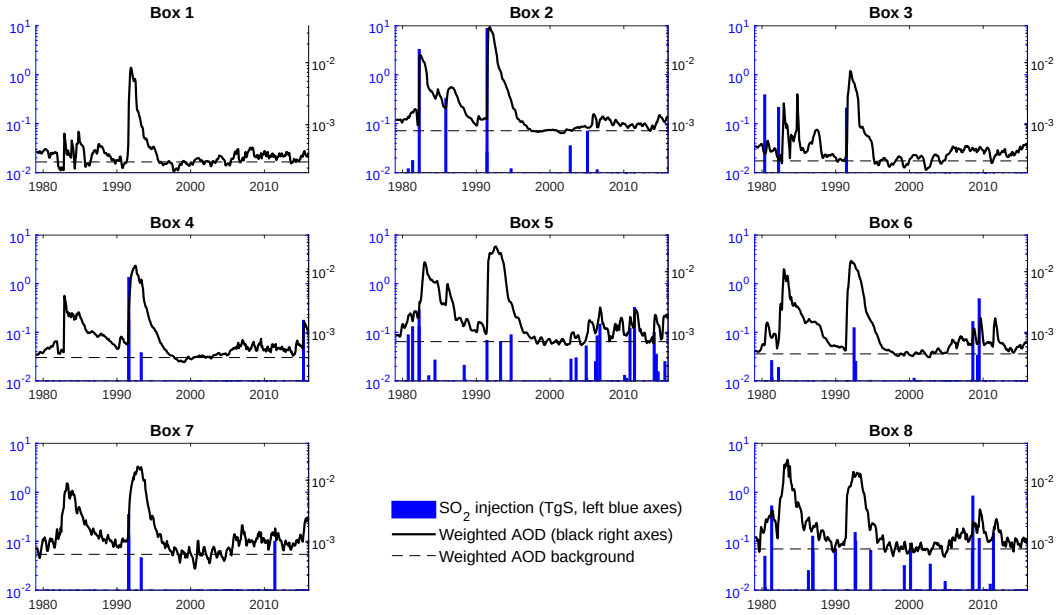

Figure 3. Pre-processed data used to calibrate the model. Each subplot correspond to a box of the model (cf. Figure 1). Blue bars (left y-axes) are SO_2 injections (TgS) in each box calculated using the Carn et al. (2016) SO_2 inventory and distributed among boxes using Gaussian functions (Section 2.5). Black lines (right y-axes) show the AOD of each box (from GloSSAC, obtained by integrating extinction from the lower to upper boundary of a box) weighted by the horizontal spatial extent of boxes. Dashed lines shows the weighted AOD background (1999-2002 average) which was removed from each AOD timeseries before calibration.

Our calibration problem is non-linear and involves between 4 and 54 parameters depending on the choices made for the model configuration, such as the latitudinal and vertical dependence of loss timescales, which will be discussed in Section 3.2. Given a specific model configuration, we use a "genetic algorithm" to find a set of optimal parameter values minimizing the error metric E (Equation 10). Genetic algorithms use strategies inspired from natural selection processes to efficiently solve non-linear optimization problems with a large number of parameters (Goldberg, 1989). Supporting information S4 provides a detailed description of the algorithm used and tests conducted using synthetic AOD datasets.

3.2 Optimal model configuration

Section 3.1 provides an overview of the method employed to calibrate any configuration of the idealized aerosol model described in Section 2. We now have to choose a procedure for deciding which model "configuration" to use, i.e.: i) the dependence of the timescales τ_{prod}, τ_{loss}, τ_{mix} and τ_{owm} on latitude, altitude and season; ii) between which boxes to include two-way and one-way mixing fluxes. Configurations of increasing com-

plexity will include more parameters, and better fit the data. However, we have to decide whether improved fitness is significant given uncertainties in SO_2 and extinction observations.

Figure 4 sketches the methodology used to determine whether a relatively complex "contender" model configuration performs significantly better than a relatively simple "reference" model configuration. Differences between a contender and reference model are kept minimal, e.g. the only difference may be that all boxes have the same loss timescale in the reference model while loss timescale depends on altitude in the contender model, resulting in 3 loss timescales instead of 1. First, we use the Carn et al. (2016) and GloSSAC datasets to calibrate the contender model using a genetic algorithm (supporting information S2). To test whether the calibrated contender model is significantly better than the reference model, we create 100 sets of perturbed model input and output data by randomly perturbating SO_2 injection mass and height (Carn et al., 2016) and weighted AOD timeseries in the 8 boxes (Thomason et al., 2018) by up to 30%, 20% and 10%, respectively. The error E of both the contender and reference model are calculated for each perturbed dataset and we then obtain the probability $p_{cont<ref}$ that the contender model is better than the reference model given uncertainties in observational data used to calibrate the model. We use a significance level of 95%, so that if $p_{cont<ref} \geqslant 0.95$, the contender model becomes the new reference model. If $p_{cont<ref} < 0.95$, we maintain the previous reference model and choose a new contender model by trying a different incremental change in the model configuration. The 95% confidence level chosen is somewhat arbitrary because we would need to better constrain the level of uncertainty in observational data and/or to use uncertainties specific to each eruption to rigorously determine a confidence level. However, it provides us with a threshold to discriminate model configurations that we estimate to be significantly fitter.

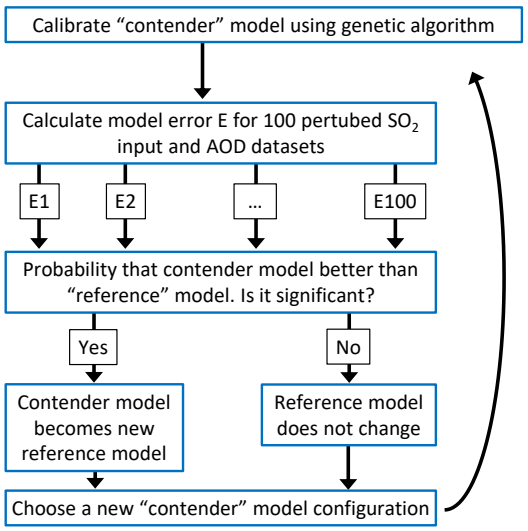

Figure 4. Flowchart of the iterative process employed to determine the optimal model configuration.

In our initial reference model (model "0"), there are no one-way mixing fluxes (Equation 4), two-way mixing fluxes (Equation 3) are horizontal only, and all model parameters are independent of latitude, altitude and season. The resulting model configuration has 4 parameters (A, τ_{prod}, τ_{loss} and τ_{mix}). Table 1 summarizes the result of our iterative process to determine an optimal model configuration (Figure 4). For example, the first row indicates that in the first contender model (model 1), loss timescales τ_{loss}

depend on altitude. Model 1 outperformed model 0 for 63% of the perturbed input/output datasets, which is not significant at the 95% level. The reference model has thus not been changed before testing a new contender model, as reflected in the second row.

The only changes that we retain relative to our initial model configuration are to make loss timescales dependent on latitude and altitude. Making the production timescales dependent on altitude or latitude significantly improved the model error, but the calibration results in ≥ 18 months production timescales in model boxes that do not receive significant injections from the 1982 El Chichòn and 1991 Pinatubo eruptions. When fitting global mean SAOD timeseries following individual eruptions using a 1-box model (not shown), production timescales for the 1982 El Chichòn and 1991 Pinatubo eruptions are 6-9 months whereas production timescales for 6 eruptions injecting smaller SO_2 mass at lower altitude (such as Sarychev Peak 2009 and Nabro 2011) range from 0.5-2 months. Production timescales of 18 months are thus unrealistic, in particular for the lower boxes of the model. In fact, such large production timescales result in an extended aerosol production in other boxes, meaning that a minimum in our error metric is achieved by fitting AOD variability associated with the 21^{st} Century eruptions by a relatively constant term, which is not physically satisfying. Consequently, we maintain a constant production timescale in the model and further discuss this choice in Section 4.

Most other tested changes, such as adding one-way mixing terms or making mixing timescales seasonally-dependent, did not result in significant error improvement. Following our calibration process, we thus do not retain some of the parameterizations implemented in EVA (Toohey et al., 2016) that are physically consistent and result in good predictions of the spatio-temporal evolution of SAOD following the 1991 Pinatubo eruption (e.g. seasonal mixing, one-way mixing). However, the model scripts provided with this paper are not restricted to our choice of configuration but enable the user to choose latitudinal, vertical and seasonal dependence for all model timescales (see supporting information S4).

Table 1. Summary of results of the iterative process used to determine the optimal model configuration (Figure 4). The contender model becomes the new reference model when the probability $p_{cont<ref}$ that the error E of the contender model is smaller than the one of the reference model is larger than 0.95, and that the calibration process leads to physically consistent parameter values (e.g. in terms of range or ranking). Significant probabilities are reported in bold.

Ref. model #	Cont. model #	Change(s) in cont. model relative to ref. model	$p_{cont<ref}$	Physically consistent?
0	1	Loss timescales depend on altitude	0.63	Y
0	2	Loss timescales depend on latitude	**0.99**	Y
2	3	Loss timescales depend on altitude	**0.98**	Y
3	4	Production timescales depend on altitude	**1**	N: The production timescales of boxes 4-8 are ≃19 months, close to the upper limit fixed (20 months)
3	5	Production timescales depend on latitude	**0.98**	N: The production timescales of extra-tropical boxes are ≃19 months, close to the upper limit fixed (20 months)
3	6	Upwelling term between boxes 2 and 5	0.37	Y
3	7	Mixing between boxes 5 and 7/8	**1**	N: the model becomes insensitive to injection latitude (regardless of injection height)
3	8	Mixing timescales depend on altitude	0.25	Y
3	10	Mixing timescales depend on season	0.44	Y
3	11	Horizontal one-way mixing between the tropics and extra-tropics	0.42	Y
3	12	Horizontal one-way mixing in boxes 1-3	0.27	Y

Table 2 reports the calibrated values of our final choice for the model configuration. We calculate uncertainties on parameter values by calibrating the model against each of the 100 perturbed input/output datasets. The values of the SAOD-sulfate mass scaling factor (A=0.0187), the production timescale (7.8 months) and mixing timescales (10.7 months) are moderately but significantly different from the values used in EVA (0.036, 6 months and 15 months, respectively). The production timescale corresponds to the effective production timescale of SO_2 into radiatively active SO_4 aerosol and should not be interpreted as the decay timescale of SO_2 which is on the order of days to weeks (e.g. Carn et al. (2016)). The loss timescales span an important range (2.3-14.5 months), with most of them being much lower that the value used in EVA (≃ 11 months) which is ex-

pected as EVA_H comprises 3 vertical layers. For boxes 1-3 and 4-6, extratropical loss timescales are significantly smaller than the tropical ones which is consistent with the tropical pipe model (Plumb, 1996; Neu & Plumb, 1999). Most model parameters are well constrained, with relative uncertainties on the order of 25% or less. The one exception is the loss timescale of box 5 (tropical lower stratosphere, the box with the most fluxes in EVA_H), for which uncertainties permit values between ca. 9 and 21 months. Table S1 shows that when calibrating the model using different periods (e.g. 1990-2015 or 1990-1997), obtained parameter values are in close agreement with those obtained in Table 2. Using the full 1979-2015 period results in better constrained parameter values, in particular for the SAOD-sulfate mass scaling factor and the production timescale. We also repeated the calibration process with a mass of 10 Tg of SO_2 for the 1991 Mt. Pinatubo (Table S1) instead of 18 Tg of SO_2 in Carn et al. (2016). Some authors (e.g. Mills et al. (2016)) have argued that a smaller mass is representative of the SO_2 not scavenged by ash and ice on the basis of the reanalysis of Pinatubo SO_2 evolution by Guo et al. (2004). The resulting parameter values are not significantly different from the one shown in Table 2, although values for the SAOD-sulfate mass scaling factor (A, Equation 8) and production timescale (τ_{prod}) respectively lie in the upper and lower range of those exhibited in Table 2.

Table 2. Values of parameters for the final model configuration chosen. The unit of A is TgS^{-1} and all timescales are given in month. The 95% confidence interval is reported in parentheses. We calculate it as the 2.5^{th} and 97.5^{th} quantiles of the parameter value distribution obtained by calibrating the model against each of the perturbed inputs/outputs dataset pair.

Parameter	Value
A (SAOD-M_{SO_4} scaling prefactor)	0.0187 (0.0152 - 0.0231)
τ_{prod} (production timescale)	7.8 (6.3 - 9.2)
$\tau_{loss}^{1,3}$ (loss timescale, extra-tropical middle stratosphere)	2.3 (1.9 - 2.7)
τ_{loss}^{2} (loss timescale, tropical middle stratosphere)	9.5 (7.2 - 16.5)
$\tau_{loss}^{4,6}$ (loss timescale, extra-tropical lower stratosphere)	2.7 (2.3 - 3.1)
τ_{loss}^{5} (loss timescale, tropical lower stratosphere)	14.5 (8.8 - 20.5)
$\tau_{loss}^{7,8}$ (loss timescale, extra-tropical lowermost stratosphere)	3.8 (3.3 - 4.4)
τ_{mix} (mixing timescale, lower and middle stratosphere)	10.7 (9.2 - 12.6)

Last, we find background sulfate injection terms B_i (Equation 2) by fitting a model run without volcanic injections to the background AOD in each box defined as 1999-2002 average. Corresponding background injections and their uncertainties are reported in Table S2. The total injection in the model is 0.28 TgS yr^{-1}, a bit larger but not significantly different from the value of 0.2 TgS yr^{-1} used in EVA.

With all key model parameters calibrated, Figure 5 shows AOD predictions (area-weighted) for each box in comparison to GloSSAC. The northern hemisphere lowermost stratosphere (box 8) accounts for over 25% of the model error E, with an important overestimation of AOD related to post-2005 eruptions and underestimation of AOD related to the 1982 El Chichón eruption. Similar errors are observed for the northern hemisphere lower stratosphere (box 6). In general, the AOD responses associated with the Kasatochi 2008, the Sarychev Peak 2009 and the Nabro 2011 eruptions are slightly overestimated by the model. However, the observed AOD mostly falls within the model prediction confidence interval, whose magnitude is largely driven by uncertainties in injected SO_2 altitude and mass. The model seems to overestimate typical rise and decay timescales of

AOD peaks associated with lower stratospheric injections despite the latitude and altitude dependence of the loss timescales.

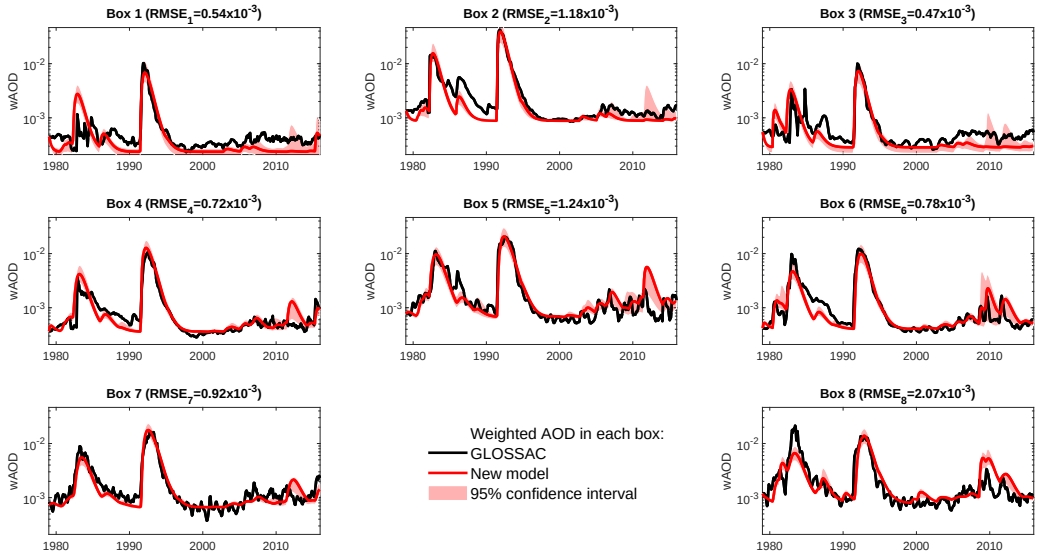

Figure 5. Area-weighted AOD ($wAOD$) in the 8 model boxes. The black line shows observations from GloSSAC and the red line shows EVA_H predictions using the Carn et al. (2016) SO_2 inventory and optimal parameter values. The corresponding RMSE for each box is annotated above each plot. Red shadings show the 95% confidence interval accounting for uncertainties related to SO_2 injections and model parameters.

3.3 Shape functions for prediction of latitudinally and vertically dependent properties

In EVA, Gaussian shape functions (in latitude and height) are used to produce latitude-altitude distribution of extinction given mass of aerosols in the three latitudinal boxes. Here, we use a multilinear regression approach to produce extinction distributions from observations. At each latitude λ and altitude z, we perform a multilinear regression where the extinction timeseries $EXT_{525}(\lambda, z, t)$ from GloSSAC is the dependent variable and the weighted AOD timeseries predicted by the model in the eight boxes $wAOD_i(t)$ (using the Carn et al. (2016) SO_2 inventory) are the independent variables:

$$EXT_{525}(\lambda, z, t) = \sum_{i=1}^{8} c_i(\lambda, z) \times wAOD_i(t) + \epsilon(\lambda, z, t) \tag{11}$$

where $i = 1..8$ is the box index, $\epsilon(\lambda, z, t)$ is the error, and $c_i(\lambda, z)$ are the regression coefficients of box i for latitude λ and altitude z. We impose that coefficients c_i are positive and that their upper limit decay exponentially with distance from the edge of their associated box i. As the global mean SAOD is equal to the sum of $wAOD$ in the 8 boxes as well as to the global mean of the vertical integral of extinction, we also normalize each shape function c_i by its global mean vertical integral. Additional procedures related to smoothing and extension to high-latitudes are described in supporting information S3. The final shape functions of EVA_H are shown in Figure 6. Last, the global mean vertical integral of extinction equals the global mean SAOD and must follow our chosen scaling for SAOD (Equation 8). Consequently, for sulfate burdens larger than M^*, we normalize each shape function by $\left(\frac{M^*}{M_{SO4}}\right)^{1/3}$.

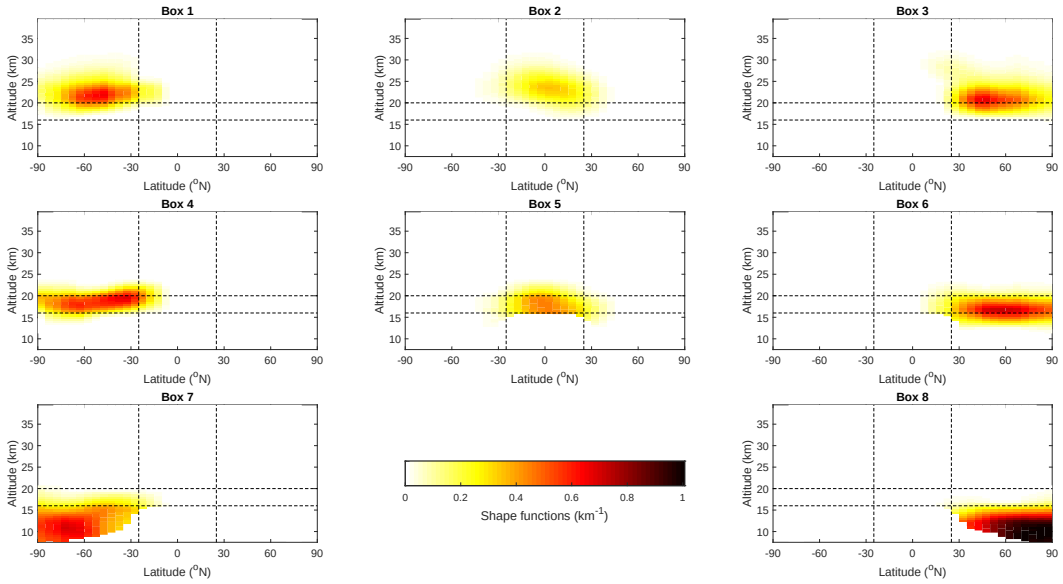

Figure 6. Shape function of EVA_H as a function of latitude and altitude. Shape functions correspond to regression coefficients c_i in Equation 11, modified after extension to high latitude and smoothing (supporting information S3). Dashed lines show latitudinal and vertical boundaries between the model 8 boxes.

3.4 Effective radius and wavelength-dependent optical properties

Climate models without an interactive stratospheric aerosol scheme generally require wavelength-dependent extinction, single scattering albedo and scattering asymmetry factor to simulate the climate response to volcanic eruptions. We adopt the same approach as EVA to produce these parameters (Toohey et al., 2016). We assume that the aerosol size distribution is log-normal with a single mode and a width parameter $\sigma = 1.2$. We then use look-up tables calculated from Mie theory to calculate wavelength-dependent optical properties from the extinction at 525 nm and the effective radius of the aerosol size distribution.

To calculate the global mean effective radius (R_{eff}), Toohey et al. (2016) used the following scaling:

$$R_{eff} = R \times (M_{SO4})^{\beta} \quad , \tag{12}$$

with $\beta = 1/3$, $R = 0.78\ \mu\text{m}(\text{TgS})^{-1/3}$, and setting a minimum effective radius of 0.2 μm. We first test whether an exponent of 1/3 seems appropriate using observations and derived products from GloSSAC and simulations from the three interactive stratospheric aerosol model previously described (Section 2.4). In GloSSAC, extinction at 525nm and 1020nm are the only variables issued from direct observations, while the effective radius is derived from these variables using methods described by Thomason et al. (2008). Consequently, instead of investigating the relationship between the effective radius and the mass of sulfate, we look at the relationship between the SAOD at 525nm and the effective radius (Figure 7), which follows a scaling of the type $R_{eff} = r_1 \times SAOD^{\beta}$ given our assumed linear scaling between SAOD and M_{SO4} for eruptions injecting less than 10 Tg S (Section 2.4). When fitting the global mean effective radius (mass weighted or surface area density weighted) to SAOD using a power-law, both GloSSAC and the simulations from UM-UKCA suggest that a 1/3 scaling is adequate, although simulations from WACCM and MAECHAM suggest values of with $\beta \simeq 0.2$ instead of 1/3. We thus maintain a value of $\beta = 1/3$ as in EVA. We set a minimum value of effective radius of

0.1 μm which seems broadly consistent with GloSSAC and simulations from interactive stratospheric aerosol models (Figure 7). Fitting the effective radius to SAOD using a 1/3 power law, values of r_1 range from 0.47 to 0.93 (for GloSSAC), corresponding to values of R (Equation 12) ranging from 0.17 to 0.26 $\mu m(TgS)^{-1/3}$ using the relationship $R = r_1 \times A^{1/3}$ and our estimate of 0.0187 $(TgS)^{-1}$ for A (Table 1). Such values of R are 3-4 times lower than the value of 0.78 used in EVA. However, Figure S9 shows that EVA_H captures best the evolution of the global mean SAOD at 1020 nm following the 1991 Mt. Pinatubo eruption when using a value of 0.26 (close to the value derived from GloSSAC effective radius and SAOD at 525 nm). We thus use a value of R=0.25$\mu m(TgS)^{-1/3}$ in EVA_H. The local effective radius is then calculated so that: i) its mass-weighted global mean matches Equation 12; and ii) it follows the same spatial distribution as the aerosol mass, raised to the power 1/3.

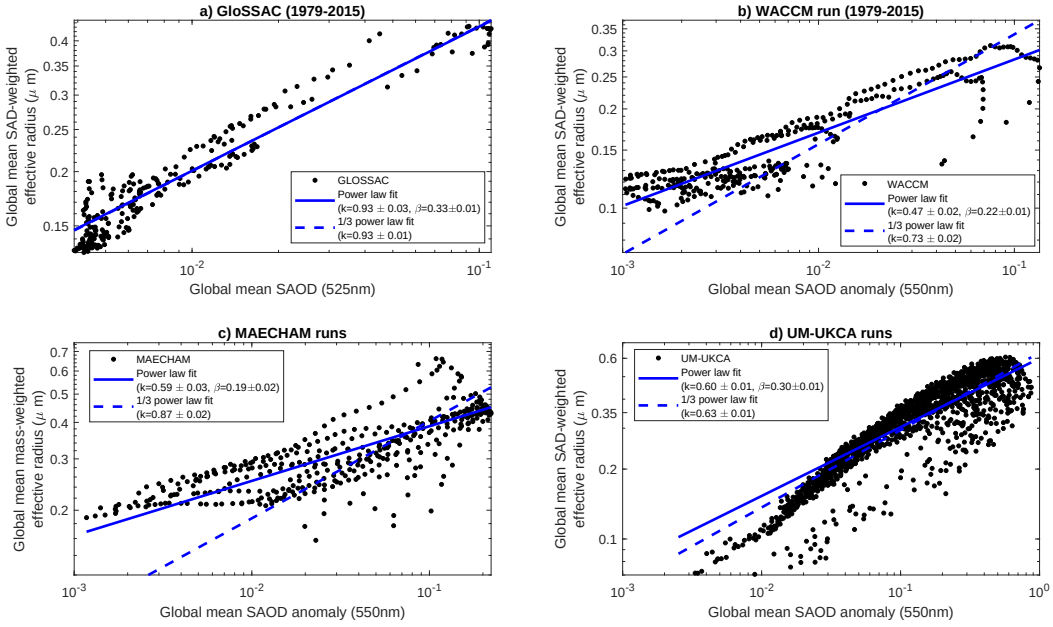

Figure 7. Global mean weighted aerosol effective radius as a function of the global mean SAOD at 525 or 550 nm for GloSSAC (a), WACCM (b), MAECHAM (c) and UM-UKCA (d). Blue lines show power-law fit for each dataset, with fit coefficients value and confidence intervals reported in legend. The effective radius is weighted by the surface area density (SAD) except for MAECHAM for which it was mass-weighted.

4 Comparison of EVA_H with EVA and interactive stratospheric aerosol models

4.1 Comparison with EVA and WACCM for the historical period.

In this subsection, we compare the predictions of EVA_H for the historical period (1979-2015) with those made by:

- EVA, the idealized model on which EVA_H builds, but which does not account for SO_2 injection height, has a prescribed vertical structure, and is calibrated against the 1991 Mt. Pinatubo eruption only.
- WACCM, which includes a prognostic stratospheric aerosol scheme (Mills et al., 2016; Schmidt et al., 2018).

Figure 8 shows the global mean SAOD timeseries for GloSSAC, EVA_H, EVA and WACCM. In panel (a), idealized models are run with the Carn et al. (2016) volcanic SO_2 emissions inventory, against which we calibrated EVA_H. In panel (b), models are run using data from Neely and Schmidt (2016). WACCM uses an adjusted SO_2 mass for the 1991 Pinatubo eruption that has been argued to be representative of the mass of SO_2 not affected by ash and ice scavenging, and results in a good agreement between the model and observations (Mills et al., 2016; Schmidt et al., 2018). For each eruption, we inject exactly the same mass of SO_2 in EVA_H and EVA. Table 3 shows each model's root mean squared error (RMSE) for the two volcanic SO_2 emissions inventories and two different time periods (full 1979-2015 period and post-Pinatubo period).

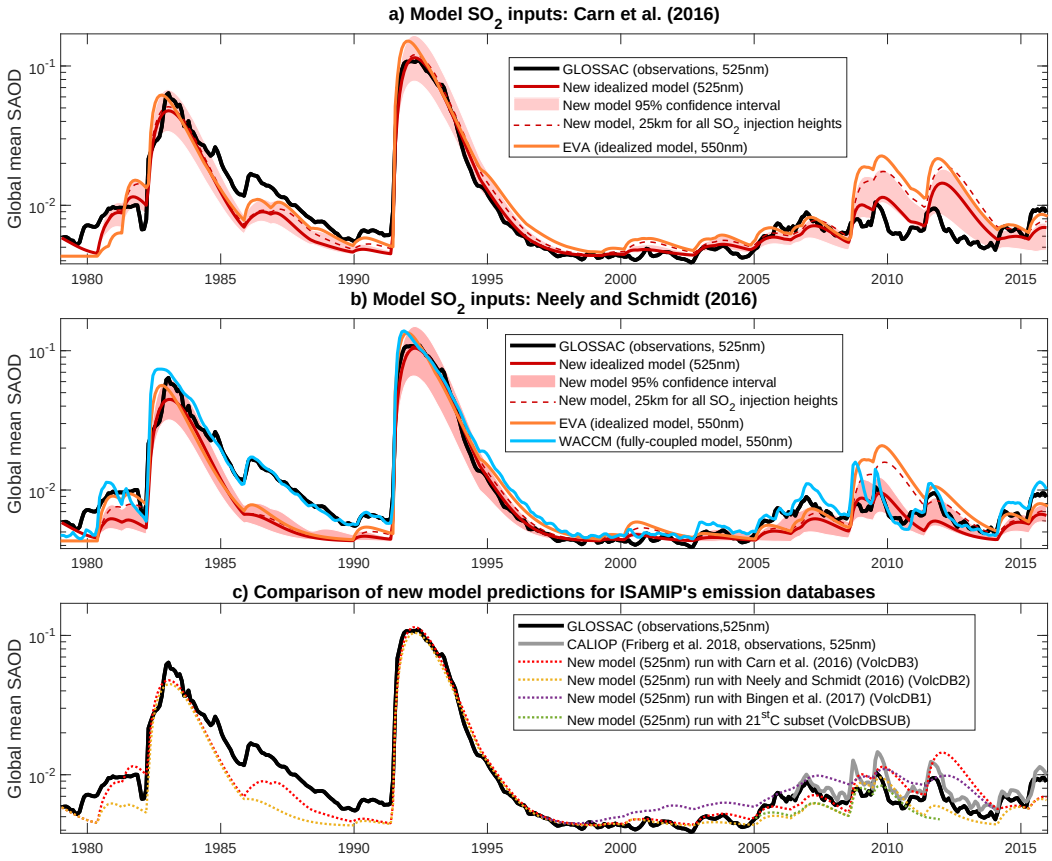

Figure 8. Panels a and b: Global mean SAOD timeseries (525 or 550nm) from observations (GloSSAC) and three different models: EVA (Toohey et al., 2016), EVA_H (this study) and the interactive stratospheric aerosol model WACCM (Mills et al., 2016; Schmidt et al., 2018). Panels a and b show models run with the Carn et al. (2016) SO_2 inventory and the Neely and Schmidt (2016) SO_2 inventory, respectively. Red shadings show the estimated 95% confidence interval related to uncertainties in calibration and SO_2 input parameters. Red dashed line shows predictions from EVA_H with a fixed 25km injection height.
Panel c: Global mean SAOD timeseries (525nm) from observations (GloSSAC and Friberg et al. (2018)) and EVA_H using the volcanic SO_2 emission databases used in the Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP, Timmreck et al. (2018)): Bingen et al. (2017) (VolcDB1, 1997-2012), Neely and Schmidt (2016) (VolcDB2, 1990-2014), Carn et al. (2016) (VolcDB3, 1979-2015) and the subset of the strongest 8 eruptions over 1998-2012 with parameters (SO_2 mass and height) averaged from all other databases used in Timmreck et al. (2018).

Table 3. Root mean squared error (RMSE, $\times 10^{-3}$) of model-predicted global mean SAOD timeseries (Figure 8) relative to the GloSSAC timeseries. We show RMSE calculated with two different SO_2 emission databases (Carn et al. (2016) and Neely and Schmidt (2016)) and over two different time periods. Bold values are outside the RMSE 95% confidence interval of EVA_H. The second row shows RMSE associated with prediction of EVA_H run with a fixed injection height of 25km.

SO_2 database	Carn et al. (2016)		Neely and Schmidt (2016)	
Period	1979-2015	1998-2015	1979-2015	1998-2015
EVA_H	3.8	2.1	4.4	1.2
EVA_H, SO_2 at 25km	4.8	**4.2**	4.3	**2.3**
EVA	7.8	**5.2**	5.9	**3.9**
WACCM	-	-	6.8	1.4

Regardless of the SO_2 emissions inventory used, EVA_H reproduces well the time evolution of the global mean SAOD. Over the 1998-2015 period, it even performs better using the Neely and Schmidt (2016) inventory instead of the Carn et al. (2016) inventory against which it was calibrated. The observed SAOD following the El Chichón 1982 and Mt. Pinatubo 1991 eruptions lies within the estimated 95% confidence interval for model predictions. EVA_H tends to overestimate the global mean SAOD associated with 21^{st} century eruptions when using the Carn et al. (2016) inventory, and to underestimate it when using the Neely and Schmidt (2016) inventory. The main reason is the lower plume height estimates provided in the Neely and Schmidt (2016) inventory that result in less injected SO_2 and shorter-lived SO_4 in our box model. Figure 8 (panel c) gives a more comprehensive overview of the sensitivity of the model predictions to the SO_2 emission inventory using the 4 inventories of the Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP, Timmreck et al. (2018)). In particular, we show that for the 21^{st} century, uncertainties in model prediction related to the different inventories existing are often larger than discrepencies between two SAOD observational datasets (GloSSAC and Friberg et al. (2018)). Regardless of the inventory or SAOD dataset used, the main failure of EVA_H lies in a clear overestimation over the rise and decay time of SAOD associated with 21^{st} century eruptions, despite the latitudinal and vertical dependence of loss timescales in the model.This failure is related to the fact that the production timescale is constant with a value of ca. 7.8 months. Consequently, in addition to overestimating SAOD rise timescales, we also overestimate decay timescales of relatively small eruptions for which the long production timescale compensates the small loss timescales in lower stratospheric boxes. We further discuss this problem and our choice of model configuration for production timescales in the following sections.

Despite imperfections in the prediction and behavior of EVA_H, it represents a clear improvement over EVA. For the 1979-2015 period, EVA_H has a RMSE 30-50% smaller than that of EVA although differences are not significant (Table 3), and for the 1998-2015 period, the RMSE of EVA_H is a factor of ca. 3 lower than EVA, with this difference being significant for both the Carn et al. (2016) and Neely and Schmidt (2016) inventories. In particular, EVA overestimates global mean SAOD over 2008-2014 by almost a factor of 3 using the Carn et al. (2016) inventory. Differences between EVA and EVA_H are not straightforward to interpret as they result from: i) a different model structure; ii) an additional input (injection height) in EVA_H; and iii) different datasets used to calibrate the model. To gain insights on the importance of injection height to accurately predict volcanic forcing, we run EVA_H with all injections height fixed to the Pinatubo 1991 height (25 km in Carn et al. (2016)), which is the only eruption used to calibrate

EVA. In this run, we inject exactly the same mass of SO_2 for each eruption as for the run with observed injection height (only the distribution among boxes changes). The corresponding global mean SAOD prediction is the thin dashed line on Figure 8 a-b with associated RMSE reported in Table 3. It is in close agreement with EVA, demonstrating that accounting for injection height makes a significant difference for accurately capturing volcanic forcing over a large range of volcanic injection parameters (e.g. Pinatubo 1991 vs. Sarychev Peak 2009).

When using the Neely and Schmidt (2016) inventory, EVA_H has slightly lower RMSE on global mean SAOD than WACCM, but with differences between the two models being insignificant (Table 3). In general, WACCM predicts larger SAOD peaks than EVA_H for 21st eruptions, with significant differences for the Kasatochi 2008 eruption. Given the relatively low average injection heights in the Neely and Schmidt (2016) inventory, we suspect that these differences are related to the self-lofting of volcanic gases in WACCM which increases the fraction of sulfur ending in the stratosphere following upper-tropospheric/lower stratospheric injections. This process is absent in EVA_H, and analyses done to determine SO_2 distribution among the box did not reveal any systematic bias between injection heights reported in Carn et al. (2016) and the height at which observed peak extinction enhancements occur following eruptions (supporting information S1, Figure S3). Last, WACCM captures well the short rise and decay timescales of SAOD peaks associated with relatively small volcanic injections in the 21st century in contrast to EVA_H.

Beyond improving predictions for the global mean SAOD, a major motivation for our new idealized model is to better capture the vertical structure of extinction changes associated with volcanic stratospheric sulfur injections. Figure 9 shows the time-altitude evolution of extinction for GloSSAC, EVA_H (run with Carn et al. (2016)), EVA (run with Carn et al. (2016)) and WACCM (run with Neely and Schmidt (2016)) over three latitudinal bands corresponding to extratropical southern latitudes, tropics and extratropical northern latitudes. Two of the major features of extinction time-altitude evolution in GloSSAC are: i) large extinction values extending up to ca. 35 km for the Pinatubo 1991 eruption vs. 20 km for post-2005 eruptions; and ii) a decrease of the altitude of peak extinction values following the 1991 Pinatubo eruption. These features cannot be captured by EVA - which prescribes a Gaussian vertical profile of extinction calibrated against Pinatubo - but are well captured by EVA_H, demonstrating the value of the vertical layers of boxes added (Figure 1) and accounting for plume height. Similarly, WACCM captures these features well. From Figure 9, it is again clear that extinction decay timescales for post-2005 eruptions are overestimated in EVA_H, whereas the fully-coupled aerosol-chemistry-climate model WACCM reproduces well short decay timescales for these eruptions. Last, in GloSSAC, extinction enhancements associated with the El Chichón 1982 eruption occur at lower altitude than those from the 1991 Pinatubo eruption. EVA_H fails to capture this, but the cause is most likely the particularly high injection height reported by Carn et al. (2016) for El Chichón 1982 eruptions (28 km for the phase with the most SO_2 injections). Such height is at the upper end of values found in the literature (e.g. Aubry et al. (2017) and references herein).

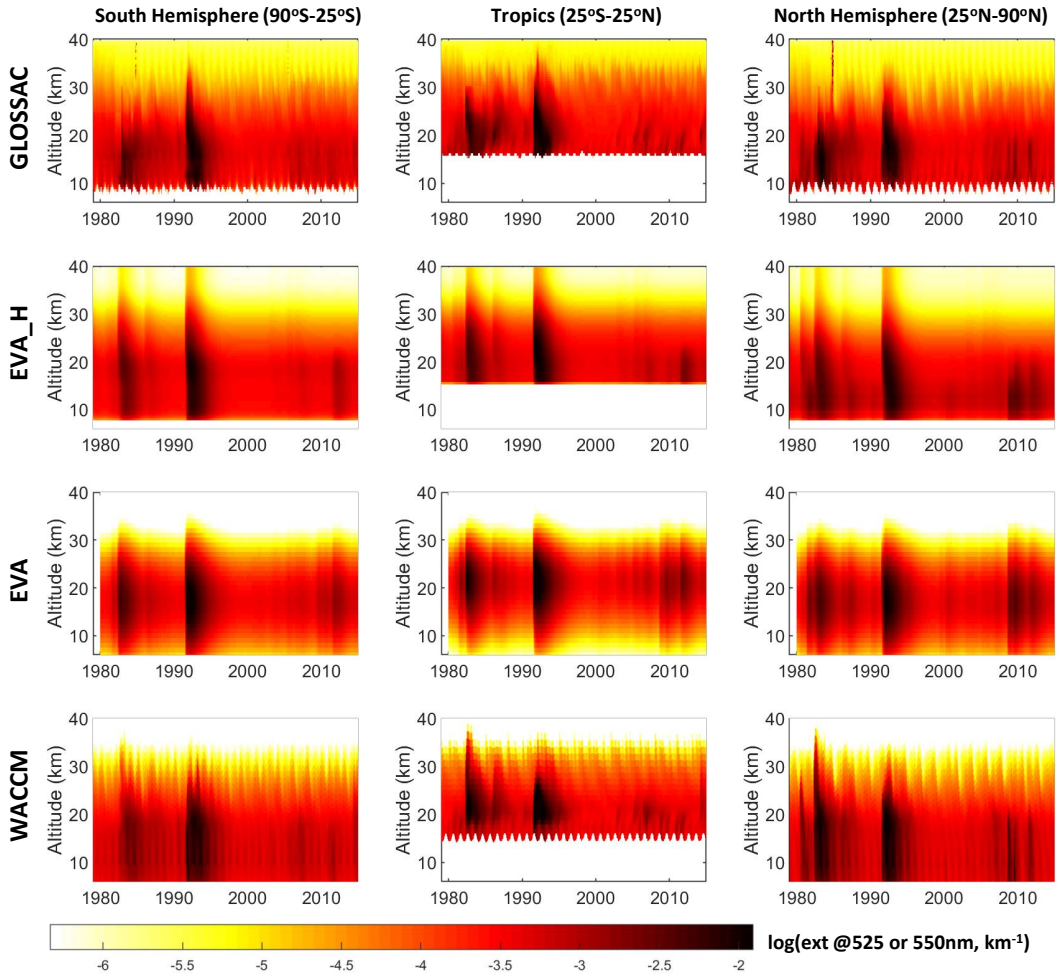

Figure 9. Extinction at ≃525 nm as a function of time and altitude, averaged longitudinally and over three different latitudinal bands: $90^oS - 22.5^oS$, $22.5^oS - 22.5^oN$ and $22.5^oN - 90^oN$ corresponding to the left, center and right columns of plots respectively. The four rows of plots show, from top to bottom, extinction from GloSSAC, EVA_H (run with Carn et al. (2016)), Easy Volcanic Aerosols (run with Carn et al. (2016)) and WACCM (run with Neely and Schmidt (2016)).

4.2 Model sensitivity to injection height and latitude: Comparison with EVA, UM-UKCA and MAECHAM

Figures 8 and 9 show that EVA_H overestimates the decay timescale of SAOD associated with 21st century eruptions, compared to both observations and simulations by WACCM. To further investigate this limitation, we investigate the sensitivity of 2 forcing metrics to injection altitude and latitude:

- The cumulative global mean SAOD at 525 nm, in months, calculated as the time-integrated SAOD between 0 and 38 months following the eruption.
- The e-folding time of the global stratospheric SO_4 burden, in months, calculated using an exponential fit of the SO_4 burden time series between one month after the peak value is reached and the month at which it reaches 10% of its peak value.

We calculate these parameters for a July injection of 8.5 Tg S, and compare the results to simulations conducted with UM-UKCA by Marshall et al. (2019) and with MAECHAM by Toohey et al. (2019). Results for a January eruption are also shown for MAECHAM.

Figure 10 shows cumulative global mean SAOD as a function of injection height and latitude for EVA_H (top left panel) and UM-UKCA (top center panel). Values for UM-UKCA are calculated using a Gaussian process emulator trained with 41 simulations (Marshall et al., 2019). The two models are in broad agreement on the following features: i) cumulative SAOD decreases as the injection latitude increases (in absolute value); ii) cumulative SAOD decreases with decreasing injection height below ca. 20km. However, there are important differences between the two models. First, the cumulative SAOD predicted by UM-UKCA is much larger than that of EVA_H. For example, for tropical injections between 20 and 25km, UM-UKCA has cumulative SAOD of ca. 4.5 months vs. 1.8 months for EVA_H. Second, UM-UKCA is much more sensitive to injection latitude, with the cumulative SAOD of an eruption at $\geq 45^o$ latitude being 30-60% smaller than an eruption with the same injection height in the tropics while this difference is only ca. 20% in EVA_H. Third, the only seasonal effect in EVA_H is related to the tropopause height seasonal cycle which explains the slight differences in cumulative SAOD for injections in the lowermost southern hemisphere stratosphere and lowermost northern hemisphere stratosphere. In contrast, for the July injection shown, UM-UKCA predicts a clearly larger cumulative SAOD and e-folding time for eruptions in the southern hemisphere compared to those in the northern hemisphere for injection heights between 18 and 27 km. This may be related to the more pronounced transport and stratosphere-troposphere exchange in the winter hemisphere (Butchart, 2014) in January-March (i.e., the northern hemisphere), when the aerosol burden of a July eruption peaks in UM-UKCA.

Figure 10 (top right) shows cumulative SAOD for EVA_H, UM-UKCA, MAECHAM and EVA for six sets of injection latitude and height for which simulations were conducted with MAECHAM, for an 8.5 Tg S July injection. Although the cumulative SAOD predicted by MAECHAM and UM-UKCA differ by up to 30%, both interactive stratospheric aerosol models agree remarkably well on the dependence of SAOD to injection latitude for a 24km injection, with a decrease by a factor 2-2.5 between an injection at 4°S and one at 56°N. In comparison, EVA_H produces a weaker dependence with a decrease by ca. 15%. However, for a 56°N injection, EVA_H and MAECHAM are in reasonable agreement for the dependence of cumulative SAOD to injection height. Last, regardless of the set of injection height and latitude used, the cumulative SAOD predicted by EVA is ca. 1.7 months. This constant value is expected as EVA does not account for injection height, and uses injection latitude only to determine the latitudinal distribution of aerosol. The loss timescales are independent of latitude so that the time evolution of the total sulfate burden and global mean SAOD only depend on the injected mass.

Bottom panels of Figure 10 are similar to the top panels, but showing results for the SO_4 e-folding time instead of the cumulative SAOD. EVA_H and UM-UKCA agree well on e-folding time for tropical injections above $\geq 20km$, ca. 12 months, while MAECHAM predicts a smaller value of ca. 8 months. However, for both interactive stratospheric aerosol models, the e-folding time strongly decreases with increasing latitude whereas EVA_H exhibits a weak dependence on latitude. The e-folding time in EVA (12.1 months) is independent of both eruption latitude and height. Overall, the e-folding timescale in EVA_H varies between 9 and 11.5 months for injections heights between 10 and 26km and all latitudes. This range is very small compared to the one of MAECHAM and UM-UKCA, and may appear surprising given that loss timescales τ_{loss} in the model are as small as 2.3 months in extratropical boxes (3.8 months for the lowermost extratropical stratosphere, i.e. boxes 7 and 8). However, the production timescale τ_{prod} is large (7.8 months) and independent of latitude or height. As a result, sulfate is produced long after the peak sulfate burden, and the e-folding timescale largely exceeds the loss timescales for extratropical injections. Last, Figure 10 shows MAECHAM's results for a January eruption

in addition to a July eruption. For an injection height of 24 km and at latitudes spanning 15-56°N, the e-folding timescale and cumulative SAOD tend to be larger for eruptions occuring in winter (January for latitudes considered), which is consistent with the explanation proposed above for the hemispheric asymmetry observed for UM-UKCA e-folding timescale and cumulative SAOD. In contrast, the total stratospheric aerosol burden evolution does not depend on eruption season in EVA and EVA_H.

All in all, comparison with both observations (Figure 8) and interactive stratospheric aerosol models (Figure 8 and 10) suggest that the forcing predicted by EVA_H still lacks sensitivity to eruption latitude. Despite this limitation, it is important to stress the sensitivity of forcing to eruption source parameters is more realistic in EVA_H compared to EVA in which the total sulfate burden and global mean SAOD evolution are independent of both injection altitude and latitude.

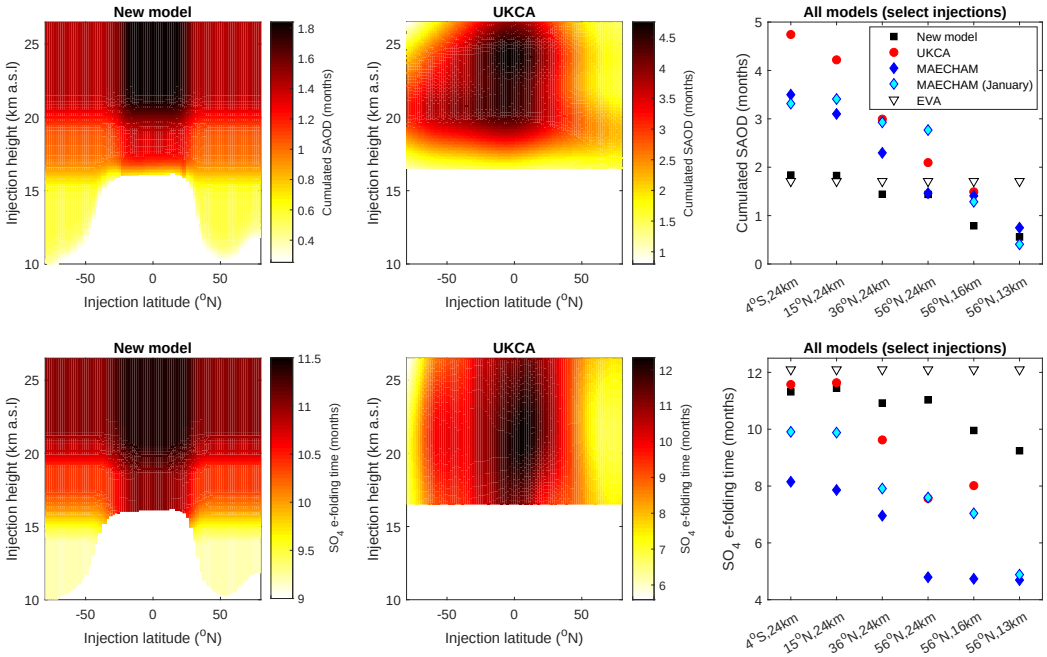

Figure 10. Cumulated global mean SAOD at 525 or 550nm (top, in month) and total stratospheric SO_4 burden e-folding time (bottom, in month) for a July injection of 8.5 Tg S into the stratosphere. The left (EVA_H) and center (UM-UKCA) columns show the sensitivity of these variables as a function of the injection latitude and latitude. The right column show these variables for EVA_H, EVA, UM-UKCA and MAECHAM for six sets of injection latitude and altitude. For MAECHAM, the same variables for a January eruption are shown in cyan.

4.3 EVA_H limitations and future developments

In light of Sections 4.1 and 4.2, the most important future improvement to EVA_H is to implement a dependence of the production timescale τ_{prod} on the injection parameters. The currently constant timescale results in a lack of sensitivity of the model-predicted forcing to the eruption latitude. The calibration methodology and/or datasets used in our study did not enable us to constrain such dependence, with unrealistically high values of τ_{prod} obtained when implementing a height or latitude dependence (Section 3.2). If we calibrate a model with height-dependent production timescales bounded to a maximum of 2.5 months for boxes 4-8, it is significantly outperformed by the model config-

uration retained with constant production timescales (using the same performance criteria as in Section 3). The primary reason is that with all other parameters being kept constant, a smaller production timescale results in larger SAOD peaks. Consequently, when enabling smaller production timescales in boxes 4-8, the overestimation of SAOD over the 21st century is worsened although the predicted rise and decay timescales compare better with observations (Figure S10). A solution and potential future development is to make the scaling factor A (Equation 7) dependent on height as well, so that SAOD signals associated with both the 1991 Pinatubo and the 21st century eruptions can be reproduced, despite the tendency of smaller production timescales to produce larger SAOD peaks. However, constraining the sulfate mass-SAOD scaling with available observations and interactive stratospheric aerosol models is already challenging, even at global scale (see Section 2.4), and such solution would largely increase the complexity of both the calibration process and the box model. In addition, we cannot exclude that the apparent overestimation of SAOD peak and rise timescale for 21st century eruptions is a consequence of errors in the observational datasets chosen to calibrate the model (Carn et al. (2016) and Thomason et al. (2018)). For example, Figure 8.c shows that for two SO_2 emission inventories, EVA_H tends to underestimate post-2000 SAOD which would facilitate the implementation of short production timescales in boxes 4-8 while maintaining good predictions for the Pinatubo eruption. Altogether, given the significant improvements of EVA_H over EVA, we choose to maintain the model configuration resulting from the calibration process described in Section 3. The scripts provided make it trivial for users of EVA_H to implement different values of production timescales in each box, in which case we recommend values of 0.5-2.5 months in boxes 4-8 (see Section 3.2 for justification of these values and Figure S10 for the corresponding model run).

Given the empirical nature of EVA_H, its calibration and predictions are limited by the parameter space covered by the set of eruptions used for calibration. In particular, the calibration of parameters of boxes 1-3 ($\geq 20km$) is constrained mostly by two large tropical eruptions (El Chichón 1982 and Pinatubo 1991). Furthermore, whereas the ice-core and geological records suggests that some of the most important volcanic events of the Common Era injected material well above 30 km in the atmosphere (e.g. Samalas 1257, Vidal et al. (2015)), no eruptions used to calibrate EVA_H injected sulfur above ca. 25 km. Until future eruptions contribute to fill this gap, interactive stratospheric aerosol model experiments could be valuable to help inform idealized models outside the parameter space in which they are calibrated.

Following our calibration procedure, seasonal mixing was not included in our chosen model configuration, in contrast to EVA, because it did not significantly improve the model performance as defined by our error metric (Equation 10). However, the seasonality of stratospheric mixing is apparent both in observations and models (e.g. Butchart (2014)) and is implemented as an option in EVA_H (see Supporting Information S4). Last, whereas interactive aerosol size evolution is key to accurately predict volcanic forcing (e.g. Mann et al. (2015)), the parameterization we use for aerosol effective radius is simplistic (Section 3.4) and effective radius does not affect, e.g., the model sulfate loss timescales. Improving the representation of aerosol size distribution in the box model is thus an important area of future development.

5 Examples of application of EVA_H: Reconstruction of past volcanic forcing and fast response during volcanic eruptions.

A major application of EVA (Toohey et al., 2016; Toohey & Sigl, 2017) is to produce forcing datasets for the experiments of the Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP, Zanchettin et al. (2016)) and the Paleoclimate Modeling Intercomparison Project (Jungclaus et al., 2017; Kageyama et al., 2018). For VolMIP, the large spread among predictions from state-of-the-art aerosol-chemistry-climate models indeed prevented the identification of consensual forcing datasets derived

from these models, motivating the use of an idealized model. Consequently, an important question is whether using EVA_H would significantly affect forcing datasets used in VolMIP or PMIP. We test this hypothesis using:

- A Tambora (1815)-like eruption with the same injections conditions as those used in Zanchettin et al. (2016) (Figure 3), i.e. 60 Tg of SO_2 at 0°N and 24 km altitude in April.
- An Eldgjá (939)-like eruption with 32 Tg of SO_2 (Toohey & Sigl, 2017) at 63.6°N and 12.5 km altitude (Moreland (2017), 17.5 km for plume top which corresponds to ca. 12.5 km for the umbrella cloud) in April.

The resulting global mean SAOD timeseries for EVA_H and EVA are shown in Figure 11, along with VolMIP runs from 4 interactive stratospheric aerosol models for the Tambora case.

For the Mt. Tambora case (Figure 11, left), the peak SAOD predicted by EVA_H is 20% smaller than the one predicted by EVA, which is largely due to our lower value of the threshold sulfate burden above which we apply a 2/3 scaling for SAOD (Equation 8). However, differences between EVA_H and EVA are not statistically significant. This result is not surprising given the similarity of injections parameters (tropical injection at $\simeq 25km$) for Mt. Tambora 1815 and Mt. Pinatubo 1991, against which EVA is calibrated. We thus expect a reasonable agreement between EVA_H and EVA for high-altitude tropical injections, and in particular for most experiments of VolMIP. Figure 11 also shows for EVA_H the uncertainty related to model parameter values and injection parameters, with uncertainty on the erupted mass of SO_2 taken from Toohey and Sigl (2017) and a 20% uncertainty on injection height. Although the predicted SAOD is uncertain by a factor of 2, the spread among predictions of interactive stratospheric aerosol models remains much larger. The predictions of two models (WACCM and UM-UKCA) are also clearly incompatible with the predictions of EVA_H. Although no conclusion can be made on which models are more realistic given the absence of SAOD observations and large uncertainties on the SO_2 mass and injection altitude for the 1815 Tambora eruption, these results stress again the large magnitude of inter-model spread, even in the face of the important uncertainties related to constraining sulfate injections from ice-cores or model calibration against recent eruptions.

For the Eldgjá case (Figure 11, right), there are significant differences between the SAOD predicted by EVA and EVA_H. If we use a latitude of 63.6° but a height of 25 km in EVA_H (similar to that of the Pinatubo 1991 eruption), the peak SAOD is 40% smaller than the one predicted by EVA. This difference is solely due to differences in model structure (including sensitivity to eruption latitude) and calibration processes. When we use the estimated injection height of 12.5 km for this eruption (Moreland, 2017), the resulting SAOD is significantly lower than the one predicted by EVA_H with a 25 km injection height or the one predicted by EVA. In particular, the predicted SAOD is 50-90% smaller than the one predicted by EVA. As a consequence, we conclude that: i) using EVA_H instead of EVA would significantly affect the forcing reconstruction for extra-tropical eruptions; and ii) injection height is an important parameter that should be accounted for - when constrained - in past volcanic forcing reconstruction. A comprehensive reconstruction of volcanic forcing associated with all eruptions for which the mass, latitude and altitude of injection are constrained is beyond the scope of this paper but is the subject of ongoing workwhich will greatly benefit from recent efforts to better constrain eruption source parameters(Burke et al., 2019; Gautier et al., 2019; Hartman et al., 2019). However, the preliminary results shown in Figure 11 reinforce the discussion of uncertainties given by Toohey and Sigl (2017), and help to quantify the degree to which the recommended PMIP4 forcing represents an upper estimate for extra-tropical eruptions.

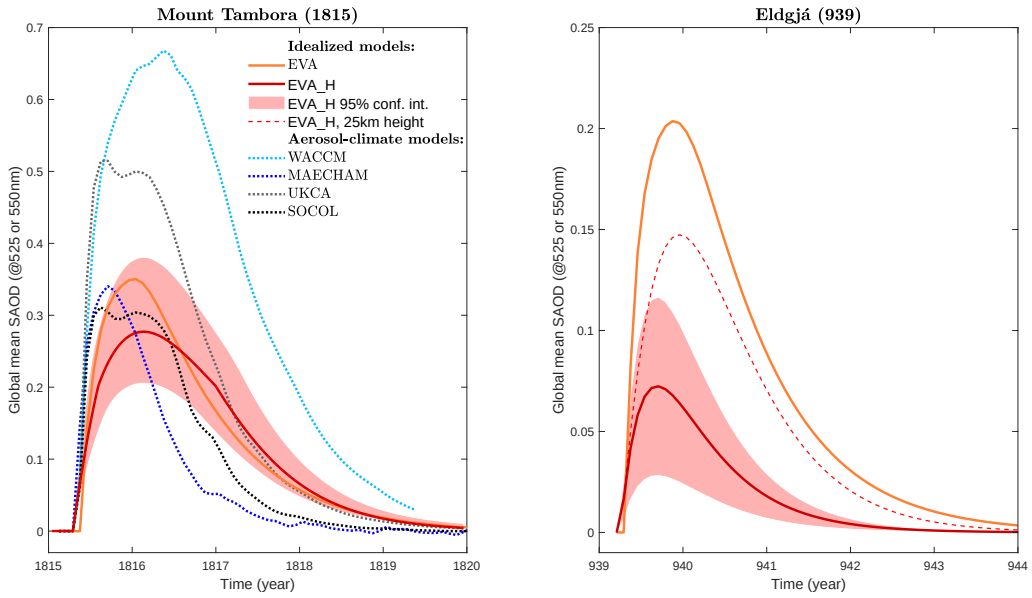

Figure 11. Global mean SAOD anomalies following a volcanic SO_2 injection with source parameters similar to those estimated for: i) Left: the 1815 Mt. Tambora eruption (60 Tg of SO_2, 0°N, 24km a.s.l., April); and ii) Right: the 939 eruption of Eldgjá (32 Tg of SO_2, 63.6°N, 12.5 km a.s.l, April). The orange and red continuous lines respectively show predictions from EVA and EVA_H, with shadings showing the 95% confidence interval for EVA_H. The red thin dashed line show results from EVA_H ran with a 25 km injection height. On the left plot, dotted lines show interactive stratospheric aerosol model runs from the VolMIP Tambora experiment (Zanchettin et al., 2016; Marshall et al., 2018). These models are WACCM (Mills et al., 2016), UM-UKCA (Dhomse et al., 2014), SOCOL (Sheng et al., 2015) and MAECHAM (Niemeier et al., 2009). We use the latest runs available after some modeling groups updated their contributions, and always use runs with point injection (as opposed to band injection) for modeling groups that tested both types of injection of volcanic SO_2.

As a final comment to this section, one of the main advantages of EVA_H over interactive stratospheric aerosol models is that it is computationally inexpensive. Consequently, it can be used to produce rapid estimates of future SAOD perturbations immediately following volcanic eruptions. A recent example of such application of the model is the June 2019 eruption of Raikoke (Kurile Islands). Shortly after first estimates of SO_2 loading and injection height were available, we ran EVA_H and provided global mean SAOD predictions to members of the "Volcano Response" (VolRes) initiative (https://wiki.earthdata.nasa.gov/dis
The model was run 1000 times to span the large range of SO_2 mass and height estimates available during the first few days after the eruption. The figures provided to the community are shown on Figure S11, and were shared with the VolRes community less than 30 minutes after deciding to apply EVA_H to the Raikoke 2019 eruption. EVA_H predicts relatively small perturbations of SAOD confined to the Northern Hemisphere, with a peak value of 9×10^{-3} at most for global mean SAOD. This upper estimate was later refined to 6.5×10^{-3} after a more detailed SO_2 injection profile was provided. Following our discussion of EVA_H limitations (Section 4), we expect that the rise and decay timescales of SAOD shown on Figure S11 are overestimated. It will be an interesting test for the model to compare Figure S11 with SAOD observations over the next year.

6 Conclusions

We take advantage of recently developed datasets of volcanic SO_2 injections (Carn et al., 2016) and atmospheric optical properties (GloSSAC, Thomason et al. (2018)) to develop EVA_H, a new idealized model of volcanic aerosol forcing that accounts for the mass, latitude and height of the sulfur injected by a volcanic eruption. Compared to the most recently developed idealized model (EVA, Toohey et al. (2016)) that did not account for injection altitude and was calibrated only against the 1991 Mt. Pinatubo eruption, we show that EVA_H:

- Captures significantly better the global mean stratospheric aerosol optical depth variations during the 21st century.
- Captures well the vertical evolution of extinction following eruptions of the 1979-2015 period.
- Exhibits a forcing sensitivity to the eruption latitude and injection height that is in better agreement with observations and interactive stratospheric aerosol model results.

Despite this latter improvement, an extensive comparison of EVA_H with interactive stratospheric aerosol model simulations shows that the latter remain more sensitive to the eruption latitude.

We apply EVA_H to discuss potential biases and uncertainties in EVA-based volcanic forcing datasets recommended for use in VolMIP (Zanchettin et al., 2016) and PMIP (Jungclaus et al., 2017), components of phase 6 of the Coupled Model Intercomparison Project. While the volcanic forcing constructed from EVA_H does not significantly differ for high-altitude tropical volcanic injections, it is significantly lower for high-latitude or low altitude emissions. As a consequence, we expect that the forcing produced by EVA_H would be similar for most experiments of VolMIP (Zanchettin et al., 2016) but may have significant differences with EVA(eVolv2k) (Toohey & Sigl, 2017), the reference volcanic forcing dataset used in PMIP (Jungclaus et al., 2017; Kageyama et al., 2018).

In contrast to interactive stratospheric aerosol models, idealized models like EVA and EVA_H are computationally inexpensive and can be used to extensively explore eruption source parameter space, which is for example required to rigorously quantify uncertainties associated with reconstructed forcing of past eruptions. We provide Matlab ® scripts that enable to run EVA_H in the configuration selected in our study (Section 3.2), but also in different configurations, e.g. with additional dependence of mixing timescales on season or production timescales on height and latitude. All scripts are available in Supporting Information S4 and EVA_H.zip, or via T.J.A.'s website (https://sites.google.com/view/thomasjaubry/
, GitHub (https://github.com/thomasaubry/EVA_H), and Code Ocean [link to be added at next revision or proofing stage after review of the Code Ocean team] where users without a Matlab ® license can run the EVA_H model.

Acknowledgments

We sincerely thank three anonymous reviewers, the associate editor, and the editor Lynn Russel for their thorough and constructive comments on the manuscript. We thank Alan Robock for his comments on an early version of this manuscript in T.J.A.'s doctoral dissertation. T.J.A. acknowledges funding from the Royal Society through a Newton International Fellowship (grant number NIF\R1\180809), from the Sidney Sussex college through a Junior Research Fellowship and from the University of British Columbia through a Four Year Fellowship. T.J.A. and A.M.J acknowledge support from a discovery grant of the Natural Sciences and Engineering Research Council of Canada (NSERC) and from a NSERC Accelerator grant. M.T. acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the research unit FOR 2820 "VolImpact: Re-

visiting the volcanic impact on atmosphere and climate – preparations for the next big volcanic eruption" (project number 398006378). L.M. and A.S. are funded by the U.K. Natural Environment Research Council (NERC) via the "Vol-Clim" grant (NE/S000887/1). L.M. was funded by NERC through the Leeds-York NERC Doctoral Training Program (NE/ L002574/1). We thank Margot Clyne and VolMIP participants for providing the latest version of SAOD timeseries of models who participated to the VolMIP Tambora experiment. We thank Johan Friberg for sharing SAOD derived from CALIOP. We thank Simon Carn and the NASA for making the MSVOLSO2L4 volcanic SO_2 emission inventory available on https://catalog.data.gov/dataset/multi-satellite-volcanic-sulfur-dioxide-l4-long-term-global-database-v3-msvolso2l4-at-ges-. We thank Larry Thomason and the NASA for making GloSSAC available. These data were obtained from the NASA Langley Research Center Atmospheric Science Data Center. Reanalysis data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/ We thank members of VolRes, and in particular Fred Prata, Simon Carn, Lieven Clarisse, Nicolas Theys and Jean-Paul Vernier, for sharing estimates of the vertical distribution of SO_2 associated with the June 2019 Raikoke eruption on the VolRes mailing list.

References

Amman, C. M., Meehl, G. A., Warren, W. M., & Zender, C. S. (2003). A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophysical Research Letters*, *30*, 1657-1660.

Ammann, C. M., & Naveau, P. (2010). A statistical volcanic forcing scenario generator for climate simulations. *Journal of Geophysical Research: Atmospheres*, *115*(D5). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD012550` doi: 10.1029/2009JD012550

Aubry, T. J., Jellinek, A. M., Carazzo, G., Gallo, R., Hatcher, K., & Dunning, J. (2017). A new analytical scaling for turbulent wind-bent plumes: Comparison of scaling laws with analog experiments and a new database of eruptive conditions for predicting the height of volcanic plumes. *Journal of Volcanology and Geothermal Research*. doi: 10.1016/j.jvolgeores.2017.07.006

Bethke, I., Outten, S., Otterå, O. H., Hawkins, E., Wagner, S., Sigl, M., & Thorne, P. (2017). Potential volcanic impacts on future climate variability. *Nature Climate Change*, *7*(11), 799. doi: 10.1038/nclimate3394

Bingen, C., Robert, C. E., Stebel, K., Brühl, C., Schallock, J., Vanhellemont, F., ... Pinnock, S. (2017). Stratospheric aerosol data records for the climate change initiative: Development, validation and application to chemistry-climate modelling. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2017.06.002

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. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0012821X19303395` doi: https://doi.org/10.1016/j.epsl.2019.06.006

Butchart, N. (2014). The Brewer-Dobson circulation. *Reviews of Geophysics*, *52*(2), 157–184. Retrieved from `http://dx.doi.org/10.1002/2013RG000448` doi: 10.1002/2013RG000448

Carboni, E., Grainger, R. G., Mather, T. A., Pyle, D. M., Thomas, G. E., Siddans, R., ... Balis, D. (2016). The vertical distribution of volcanic SO_2 plumes measured by IASI. *Atmospheric Chemistry and Physics*, *16*(7), 4343–4367. Retrieved from `https://www.atmos-chem-phys.net/16/4343/2016/` doi: 10.5194/acp-16-4343-2016

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

Crowley, T., & Unterman, M. (2013). Technical details concerning development of a 1200 yr proxy index for global volcanism. *Earth System Science Data*, *5*(1), 187–197. doi: 10.5194/essd-5-187-2013

Dhomse, S. S., Emmerson, K. M., Mann, G. W., Bellouin, N., Carslaw, K. S., Chipperfield, M. P., ... Thomason, L. W. (2014). Aerosol microphysics simulations of the Mt. Pinatubo eruption with the UM-UKCA composition-climate model. *Atmospheric Chemistry and Physics*, *14*(20), 11221–11246. doi: 10.5194/acp-14-11221-2014

Friberg, J., Martinsson, B. G., Andersson, S. M., & Sandvik, O. S. (2018). Volcanic impact on the climate – the stratospheric aerosol load in the period 2006–2015. *Atmospheric Chemistry and Physics*, *18*(15), 11149–11169. doi: 10.5194/acp-18-11149-2018

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*. doi: 10.1029/2008JD010239,

Gautier, E., Savarino, J., Hoek, J., Erbland, J., Caillon, N., Hattori, S., ... Farquhar, J. (2019). 2600-years of stratospheric volcanism through sulfate isotopes. *Nature Communications*, *10*, 1 - 7. doi: https://doi.org/10.1038/s41467-019-08357-0

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning* (1st ed.). Addison-Wesley Longman Publishing Co., Inc.

Grieser, J., & Schonwiese, C.-D. (1999). Parameterization of spatio-temporal patterns of volcanic aerosol induced stratospheric optical depth and its climate radiative forcing. *Atmosfera*, *12*(2).

Guo, S., Bluth, G. J. S., Rose, W. I., Watson, I. M., & Prata, A. J. (2004). Re-evaluation of SO2 release of the 15 June 1991 Pinatubo eruption using ultraviolet and infrared satellite sensors. *Geochemistry, Geophysics, Geosystems*, *5*(4). doi: 10.1029/2003GC000654

Hartman, L., Kurbatov, A., Winski, D., Cruz-Uribe, A., Davies, S., Dunbar, N., ... Fudge, T. (2019). Volcanic glass properties from 1459 C.E. volcanic event in South Pole ice core dismiss Kuwae caldera as a potential source. *Scientific Reports*, *9*. doi: 10.1038/s41598-019-50939-x

Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., ... Zorita, E. (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. Retrieved from `https://www.geosci-model-dev.net/10/4005/2017/` doi: 10.5194/gmd-10-4005-2017

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

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., ... others (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. doi: 10.1175/1520-0477(1996)077

Kremser, S., Thomason, L. W., von Hobe, M., Hermann, M., Deshler, T., Timmreck, C., ... Meland, B. (2016). Stratospheric aerosol—observations, processes, and impact on climate. *Reviews of Geophysics*, *54*(2), 278-335. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015RG000511` doi: 10.1002/2015RG000511

Mann, G., 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*, 52–53. doi: https://doi.org/10.22498/pages.23.2.52

Marshall, L., Johnson, J. S., Mann, G. W., Lee, L., Dhomse, S. S., Regayre, L., ... Schmidt, A. (2019). Exploring how eruption source parameters affect volcanic radiative forcing using statistical emulation. *Journal of Geophysical Research: Atmospheres*, *124*(2), 964-985. doi: 10.1029/2018JD028675

Marshall, L., Schmidt, A., Toohey, M., Carslaw, K. S., Mann, G. W., Sigl, M., ... Tummon, F. (2018). Multi-model comparison of the volcanic sulfate deposition from the 1815 eruption of Mt. Tambora. *Atmospheric Chemistry and Physics*, *18*(3), 2307–2328. doi: 10.5194/acp-18-2307-2018

Metzner, D., Kutterolf, S., Toohey, M., Timmreck, C., Niemeier, U., Freundt, A., & Krüger, K. (2014, Oct 01). Radiative forcing and climate impact resulting from SO_2 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. Retrieved from `https://doi.org/10.1007/s00531-012-0814-z` doi: 10.1007/s00531-012-0814-z

Mills, M. J., Schmidt, A., Easter, R., Solomon, S., Kinnison, D. E., Ghan, S. J., ... others (2016). Global volcanic aerosol properties derived from emissions, 1990–2014, using CESM1 (WACCM). *Journal of Geophysical Research: Atmospheres*, *121*(5), 2332–2348. doi: 10.1002/2015JD024290

Moreland, W. M. (2017). *Explosive activity in flood lava eruptions: A case study of the 10th century eldgjá eruption, Iceland* (Unpublished doctoral dissertation). University of Iceland.

Neely, R., & Schmidt, A. (2016). VolcanEESM: Global volcanic sulphur dioxide (SO2) emissions database from 1850 to present -Version 1.0. *Cent. Environ. Data Anal.*, *839*. doi: 10.5285/76ebdc0b-0eed-4f70-b89e-55e606bcd568

Neu, J. L., & Plumb, R. A. (1999). Age of air in a "leaky pipe" model of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, *104*(D16), 19243–19255. Retrieved from `http://dx.doi.org/10.1029/1999JD900251` doi: 10.1029/1999JD900251

Niemeier, U., Timmreck, C., Graf, H.-F., Kinne, S., Rast, S., & Self, S. (2009). Initial fate of fine ash and sulfur from large volcanic eruptions. *Atmospheric Chemistry and Physics*, *9*(22), 9043–9057. doi: 10.5194/acp-9-9043-2009

Plumb, R. A. (1996). A "tropical pipe" model of stratospheric transport. *Journal of Geophysical Research: Atmospheres*, *101*(D2), 3957–3972. Retrieved from `http://dx.doi.org/10.1029/95JD03002` doi: 10.1029/95JD03002

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., ... Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, *24*(14), 3624-3648. doi: 10.1175/JCLI-D-11-00015.1

Robock, A. (2000). Volcanic eruptions and climate. *Reviews of Geophysics*, *38*, 191-219. doi: 10.1029/1998RG000054

Santer, B. D., Solomon, S., Bonfils, C., Zelinka, M. D., Painter, J. F., Beltran, F., ... Wentz, F. J. (2015). Observed multivariable signals of late 20th and early 21st century volcanic activity. *Geophysical Research Letters*, *42*(2), 500–509. doi: 10.1002/2014GL062366

Santer, B. D., Wehner, M. F., Wigley, T., Sausen, R., Meehl, G., Taylor, K., ... others (2003). Contributions of anthropogenic and natural forcing to recent tropopause height changes. *Science*, *301*(5632), 479–483. doi: 10.1126/science.1084123

Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., ... Toon, O. B. (2018). Volcanic radiative forcing from 1979 to 2015. *Journal of Geophysical Research: Atmospheres*, *123*(22), 12,491-12,508. doi: 10.1029/2018JD028776

Sheng, J.-X., Weisenstein, D. K., Luo, B.-P., Rozanov, E., Stenke, A., Anet, J., ... Peter, T. (2015). Global atmospheric sulfur budget under volcanically quiescent conditions: Aerosol-chemistry-climate model predictions and valida-

tion. *Journal of Geophysical Research: Atmospheres*, *120*(1), 256–276. doi: 10.1002/2014JD021985

Thomason, L. W., Burton, S. P., Luo, B.-P., & Peter, T. (2008). Sage II measurements of stratospheric aerosol properties at non-volcanic levels. *Atmospheric Chemistry and Physics*, *8*(4), 983–995. Retrieved from `https://www.atmos-chem-phys.net/8/983/2008/` doi: 10.5194/acp-8-983-2008

Thomason, L. W., Ernest, N., Millán, L., Rieger, L., Bourassa, A., Vernier, J.-P., ... Peter, T. (2018). A global space-based stratospheric aerosol climatology: 1979–2016. *Earth System Science Data*, *10*(1), 469–492. doi: 10.5194/essd-10-469-2018

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

Timmreck, C., Graf, H.-F., Lorenz, S. J., Niemeier, U., Zanchettin, D., Matei, D., ... Crowley, T. J. (2010). Aerosol size confines climate response to volcanic super-eruptions. *Geophysical Research Letters*, *37*(24).

Timmreck, C., Mann, G. W., Aquila, V., Hommel, R., Lee, L. A., Schmidt, A., ... Weisenstein, D. (2018). The Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP): motivation and experimental design. *Geoscientific Model Development*, *11*(7), 2581–2608. Retrieved from `https://www.geosci-model-dev.net/11/2581/2018/` doi: 10.5194/gmd-11-2581-2018

Toohey, M., Krüger, K., Schmidt, H., Timmreck, C., Sigl, M., Stoffel, M., & Wilson, R. (2019). Disproportionately strong climate forcing from extratropical explosive volcanic eruptions. *Nature Geoscience*, *12*(2), 100. doi: 10.1038/s41561-018-0286-2DO

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. doi: 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*, *2016*, 1–40. doi: 10.5194/gmd-2016-83

Vidal, C. M., Komorowski, J.-C., Métrich, N., Pratomo, I., Kartadinata, N., Prambada, O., ... Surono (2015, Aug 08). Dynamics of the major plinian eruption of samalas in 1257 A.D. (Lombok, Indonesia). *Bulletin of Volcanology*, *77*(9), 73. Retrieved from `https://doi.org/10.1007/s00445-015-0960-9` doi: 10.1007/s00445-015-0960-9

Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., ... Tummon, F. (2016). The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): Experimental design and forcing input data. *Geoscientific Model Development*, *2016*, 1–33. doi: 10.5194/gmd-2016-68