https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

Geoscientific Model Development
Discussions
Open Access

Effects of forcing differences and initial conditions on inter-model agreement in the VolMIP volc-pinatubo-full experiment

Davide Zanchettin[1], Claudia Timmreck[2], Myriam Khodri[3], Anja Schmidt[4,5], Matthew Toohey[6], Manabu Abe[7], Slimane Bekki[8], Jason Cole[9], Shih-Wei Fang[2], Wuhu Feng[10], Gabriele Hegerl[11], Ben Johnson[12], Nicolas Lebas[3], Allegra N. LeGrande[13,14], Graham W. Mann[10], Lauren Marshall[5], Landon Rieger[6], Alan Robock[15], Sara Rubinetti[1], Kostas Tsigaridis[14,13], Helen Weierbach[16,17]

[1]University Ca' Foscari of Venice, Department of Environmental Sciences, Informatics and Statistics, Via Torino 155, 30172 Mestre, Italy
[2]Max-Planck-Institute for Meteorology, Bundesstr. 53, 20146 Hamburg, Germany
[3]Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques, Institut Pierre-Simon Laplace, Sorbonne Universités /IRD/CNRS/MNHN, Paris, France.
[4]Department of Geography, University of Cambridge,Cambridge, U.K.
[5]Department of Chemistry, University of Cambridge, Cambridge, U.K.
[6]Institute of Space and Atmospheric Studies, University of Saskatchewan, Canada
[7]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, 3173-25 Showa-machi, Kanazawa-ku, Yokohama 236-0001, Japan
[8]Laboratoire Atmosphères, Milieux, Observations Spatiales, Institut Pierre-Simon Laplace, Sorbonne Université/CNRS/UVSQ, Paris, France
[9]Environment and Climate Change Canada, Toronto, ON, Canada
[10] University of Leeds, Leeds, UK
[11]University of Edinburgh,Edinburgh, U.K.
[12]Met Office Fitzroy Road EXETER EX1 3PB, U.K.
[13]NASA Goddard Institute for Space Studies, New York, NY, USA
[14] Center for Climate Systems Research, Columbia University, New York, NY, USA
[15] Department of Environmental Sciences, Rutgers University, New Brunswick, NJ, USA
[16]Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[17]Lamont Doherty Earth Observatory, Columbia University, New York, NY, USA

Correspondence to: Davide Zanchettin (davidoff@unive.it)

Abstract. This paper provides initial results from a multi-model ensemble analysis based on the volc-pinatubo-full experiment performed within the Model Intercomparison Project on the climatic response to volcanic forcing (VolMIP) as part of the sixth phase of the Coupled Model Intercomparison Project (CMIP6). The volc-pinatubo-full experiment is based on ensemble of volcanic forcing-only climate simulations with the same volcanic aerosol dataset across the participating models (the 1991-1993 Pinatubo period from the CMIP6-GloSSAC dataset). The simulations are conducted within an idealized experimental design where initial states are sampled consistently across models from the CMIP6-piControl simulation providing unperturbed pre-industrial background conditions. The multi-model ensemble includes output from an initial set of six participating Earth system models (CanESM5, GISS-E2.1-G, IPSL-CM6A-LR, MIROC-E2SL, MPI-ESM1.2-LR and UKESM1).

The results show overall good agreement between the different models on the global and hemispheric scale concerning the surface climate responses, thus demonstrating the overall effectiveness of VolMIP's experimental design. However, small yet significant inter-model discrepancies are found in radiative fluxes especially in the tropics, that preliminary analyses link with minor differences in forcing implementation, model physics, notably aerosol-radiation interactions, the simulation and sampling of El Niño-Southern Oscillation (ENSO) and, possibly, the simulation of climate feedbacks operating in the tropics. We discuss the volc-pinatubo-full protocol and highlight the advantages of volcanic forcing experiments defined within a carefully designed protocol with respect to emerging modeling approaches based on large ensemble transient simulations. We identify how the VolMIP strategy could be improved in future phases of the initiative to ensure a cleaner sampling protocol with greater focus on the evolving state of ENSO in the pre-eruption period.

Plain text summary

This paper provides metadata and first analyses of the volc-pinatubo-full experiment of CMIP6-VolMIP. Results from six Earth system models reveal significant differences in radiative flux anomalies that trace back to different implementations of volcanic forcing. Surface responses are in contrast overall consistent across models, reflecting the large spread due to internal variability. A second phase of VolMIP shall consider both aspects toward improved protocol for volc-pinatubo-full.

1 Introduction

The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP, Zanchettin et al., 2016) defined a coordinated set of idealized volcanic perturbation experiments to be carried out in alignment with the protocol of the 6th phase of the Coupled Model Intercomparison Project (CMIP6, Eyring et al., 2016). VolMIP aims to assess the diversity of simulated climate responses to large-magnitude explosive volcanic eruptions with global-scale sulfate aerosol and to identify the causes and processes that cause inter-model differences.

Accordingly, VolMIP experiments are based on multi-model ensemble simulations with coupled climate models that are designed upon two pillars. First, the prescribed stratospheric volcanic aerosol optical properties (often referred to as "forcing") used in the simulations must be the same across all participating models. Consensus volcanic forcing data sets were thus defined for each experiment and implemented in terms of zonal and monthly mean optical properties of stratospheric volcanic aerosol. Second, climate conditions defining the initial states of individual members of the ensemble must be selected in a consistent manner across the participating models, so that the diagnosed expected climate response is not biased by potential effects of a preferential phase of ongoing internal variability at the time of the eruption (see, e.g., Zanchettin et al., 2013; Swingedouw et al., 2015; Lehner et al., 2016; Khodri et al., 2017; Coupe and Robock, 2021). This is achieved by defining desired states of climate variability modes to be sampled along the parent CMIP6-DECK piControl simulation representative of unperturbed preindustrial climate conditions.

VolMIP Tier 1 experiments are branched in two main sets, named “volc-pinatubo” and “volc-long”, respectively (Zanchettin et al., 2016). The volc-pinatubo experiments include a main experiment with full forcing (volc-pinatubo-full) and two sensitivity experiments aimed at disentangling the competing effects of the two mechanisms known to determine the seasonal-to-interannual response to volcanic eruptions, i.e., surface cooling and stratospheric warming. The experiments tackle uncertainty and inter-model differences in the climatic response to an idealized 1991 Mt. Pinatubo-like eruption, which is chosen as representative of the largest magnitude of volcanic events that occurred during the instrumental period to date. The volc-pinatubo experiments use a volcanic forcing dataset derived from satellite observations (Thomason et al., 2018) but do neither account for the actual climate conditions at the time of the 1991 Mt. Pinatubo eruption nor for other forcing factors concomitant with the eruption, hence their idealized character. The idealized volc-pinatubo experiments complement full-forcing transient experiments, where historical climate simulations show a general improvement of CMIP6-generation Earth system models in the simulation of the high-latitude climate response to the 1991 Mt. Pinatubo eruption compared to previous CMIP results (Pauling et al., 2021).

This paper provides the first multi-model analysis of the volc-pinatubo-full ensemble. The aim is to provide a first assessment of the experiment and general guidance about emergent gaps of knowledge revealed by major discrepancies across results from different models. An additional aim is to assess the appropriateness of the VolMIP protocol and to propose improvements for a possible second phase of the initiative. Therefore, the paper first includes a short description of the experimental setup (Sect. 2.1) and an overview of the participating models (six at the moment of writing) including technical details of the simulations such as the branching years from the CMIP6-DECK piControl (Sect. 2.2). Statistical techniques used in the multi-model analysis are described in Sect. 3, and results are presented in Sect. 4. Focus is on spatially integrated quantities regarding the energy fluxes at the top of atmosphere and at the surface, basic quantities describing global and hemispheric-scale surface climate parameters (temperature and precipitation). We illustrate how inter-model differences are largely reduced compared to previous analyses thanks to the application of the VolMIP protocol, although inconsistencies remain especially in post-eruption radiative flux anomalies. First insights into dynamical responses are also proposed, where we illustrate how selected modes of large-scale climate variability and climate feedbacks show larger differences across models compared to surface climate responses. The impacts from sampling and ensemble size on post-eruption anomalies and the expected climate response are also illustrated. Sect. 5 discusses the main inconsistencies across models, provides guidance for future studies and suggests revisions to the volc-pinatubo-full experiment in a possible second phase of VolMIP.

2 Characteristics of volc-pinatubo-full

2.1 The experiment

The VolMIP protocol established the CMIP6 stratospheric aerosol dataset to be used for the volc-pinatubo-full experiment. This dataset covers the CMIP6 historical period (1850-2018) and provides zonal and monthly mean stratospheric aerosol extinction, single scattering albedo and asymmetry factor as a function of latitude, height, wavelength and time (Luo, 2018a, b). For the years around the 1991 eruption of Mt. Pinatubo, the CMIP6 stratospheric aerosol data is constructed directly from the Global Space-based Stratospheric Aerosol Climatology (GloSSAC) observational reconstruction (Thomason et al., 2018). GloSSAC is constructed mainly from satellite observations of stratospheric aerosol extinction. For the Pinatubo period, GloSSAC aerosol properties are retrieved from the SAGE II satellite instrument, with some gaps filled via tropical and mid-latitude ground- based lidar measurements, most notably due to the strong extinction by Pinatubo aerosols in the lower tropical stratosphere in the first months after the eruption. Gaps in high latitudes have been filled by regridding the observations onto equivalent latitude instead of geographical latitude. The CMIP6 volcanic aerosol forcing initially made available for CMIP6 historical simulations and the VolMIP volc-pinatubo-full experiment was labeled as version 3, and was based on GloSSAC version 1.0 (Thomason et al., 2018). Subsequently (in August 2018), an updated version (v4) was released, based on GloSSAC v1.1, which corrected erroneous aerosol extinction values which arose through cloud screening of satellite data in the lower stratosphere. Nonetheless, version 3 remains the recommended forcing data set for the VolMIP volc-pinatubo experiments and was used in all simulations included in this study. This choice was deemed to be preferable since inter-model consistency in forcing is one of the pillars of VolMIP and volc-pinatubo-full simulations were already performed by some modeling groups at the time version 4 was released. Therefore, in the case that CMIP6 historical simulations are run with the version 4 data set, inconsistency in the forcing contributes to explaining possible inconsistencies between the volc-pinatubo-full and the historical simulations performed with the same model. Comparison of stratospheric aerosol optical depth (SAOD) for the CMIP6 v3 and v4 data sets confirms that both data sets show a very similar spatiotemporal evolution of the SAOD, with an initial tropical peak, followed by transport of the aerosol to the mid and high latitudes of both hemispheres (Fig. S1). Differences between the two versions are strongest in the tropics in the first few months after the Mt. Pinatubo eruption, and at the highest latitudes of both hemispheres. Rieger et al., (2020) compared CMIP6-historical simulations performed with two models using v3 and v4 of the CMIP6 forcing dataset, and found generally small differences in the simulated post-Pinatubo surface climate response, with the notable exception of temperature response in the tropical stratosphere.

By protocol, a minimum of 25 simulations are branched-off from the parent piControl simulation on June 1st of selected years, sampled in a way that they include equally distributed cold/neutral/warm states of the El Niño-Southern Oscillation (ENSO) and negative/neutral/positive states of the North Atlantic Oscillation (NAO). To this purpose, recommended indices were the Niño3.4 sea-surface temperature index for ENSO, and the two-box index as used by Stephenson et al. (2006) for the NAO, both referring to winter-average data (DJF, with January as reference for the year) for the first post-eruption winter (i.e., including January 1992). Specifically, the NAO index was defined as the difference between spatial averages of 500 hPa geopotential heights over (20–55 N; 90W–60 E) and (55–90 N; 90W–60 E). Sampling of an eastern phase of the Quasi-Biennial Oscillation (QBO), as observed after the 1991 Pinatubo eruption, is preferred for those models that explicitly simulate this mode of stratospheric variability.

2.2 The multi-model ensemble

The volc-pinatubo-full multi-model ensemble includes simulations from six models: the 5th version of the Canadian Earth System Model (CanESM5, Sect. 2.2.1), the NASA Goddard Institute for Space Studies Earth System model (GISS-E2.1-G, Sect. 2.2.2), the CMIP6 version of the Institut Pierre-Simon Laplace coupled atmosphere-ocean general circulation model (IPSL-CM6A-LR, Sect. 2.2.3), the Model for Interdisciplinary Research on Climate, Earth System version2 for Long-term simulations (MIROC-ES2L, Sect. 2.2.4), the Max Planck Institute Earth System Model version 1.2 in its low resolution version (MPI-ESM1.2-LR, Sect. 2.2.5) and the 1st version of the UK Earth System Model (UKESM1, Sect. 2.2.6). The main characteristics of the participating models are reported in Table 1.

2.2.1 CanESM5

The fifth version of the Canadian Earth System Model (CanESM5) couples together models of the atmosphere (CanAM5), land (CLASS), ocean/sea-ice (CanNEMO), atmospheric carbon cycle (CTEM) and ocean biogeochemistry (CMOC) which are briefly described in Swart et al. (2019). The atmosphere is resolved using a T63 horizontal resolution, roughly 2.8° in longitude and latitude, and 49 vertical levels while the ocean is horizontally resolved at roughly 1° and 45 levels. A 1000-year long piControl simulation using CanESM5 is generally stable for heat, water and carbon related quantities, with observed drifts in some variables, e.g., ocean carbon flux, that are much smaller than anthropogenic signals. The climate sensitivity of CanESM5 is greater than CanESM2 (used for CMIP5) with a value of 5.67 K compared with 3.7 K (Zelinka et al., 2020) which is mainly attributed to changes in cloud feedbacks (Virgin et al., 2021).

2.2.2 GISS-E2.1-G

The NASA Goddard Institute for Space Studies Earth System modelE version 2.1, coupled to the GISS ocean (GISS-E2.1-G, Kelley et al., 2020), is one of several versions of the NASA GISS climate model submitted to CMIP6. It uses a regular grid across both the ocean and atmosphere. It has a 1.25° in longitude by 1° in latitude, 40-layer ocean coupled to a 2.5° in longitude by 2° in latitude 40-layer atmosphere with a top at 0.01 mb. The non-interactive version of atmospheric aerosols and chemistry is used in VolMIP (physics version 1). Volcanic eruptions are represented by specifying monthly averaged volcanic sulfate (optical properties described in Lacis et al., 1992) aerosol optical depth in 69 layers (5 km-39.5 km) and 36 latitude bands with no zonal variation (Thomason et al., 2018); the general implementation strategy of the GISS stratospheric volcanic aerosols are described in Sato et al. (1993), which uses a lower resolution version of the GISS model as well as a coarser resolution stratospheric aerosol boundary condition. ENSO variability is relatively large in this version of the model with variance about 50% greater than that observed (Kelley et al., 2020). Equilibrium climate sensitivity is about 3.6 °C for doubling of CO2, with remarkable consistency to previous versions of the model.

2.2.3 IPSL-CM6A-LR

The CMIP6 version of the Institut Pierre-Simon Laplace (IPSL) coupled atmosphere-ocean general circulation model is the low-resolution IPSL-CM6A-LR (Boucher et al., 2020), corresponding to a grid resolution of its atmospheric component LMDZ6A-LR of 1.25° in latitude and 2.5° in longitude and 79 vertical levels (Hourdin et al., 2020). LMDZ6A-LR is coupled to the ORCHIDEE (d'Orgeval et al., 2008) land surface component, version 2.0. In IPSL-CM6A-LR, the oceanic component uses the Nucleus for European Models of the Ocean (NEMO), version 3.6 (Madec et al., 2017), which includes other models to represent sea-ice interactions (NEMO-LIM3; Vancoppenolle et al., 2009; Rousset et al., 2015) and biogeochemistry processes (NEMO-PISCES; Aumont et al., 2015). Compared to the 5A-LR model version and other CMIP5-class models, IPSL-CM6A-LR was significantly improved in terms of the climatology, e.g., by reducing overall SST biases and improving the latitudinal position of subtropical jets. The IPSL-CM6A-LR is also more sensitive to CO2 forcing increase (Boucher et al., 2020) and represents a more robust global temperature response than the previous CMIP5 version consistently with current state-of-the-art CMIP6 models (Zelinka et al., 2020).

2.2.4 MIROC-ES2L

The Model for Interdisciplinary Research on Climate, Earth System version2 for Long-term simulations (MIROC-ES2L) consists of the coupled atmosphere-ocean general circulation model called MIROC5.2, the land biogeochemistry component (VISIT), and the ocean biogeochemistry component (OECO2) (Hajima et al., 2020). The horizontal resolution of the atmosphere and the land is set to have T42 spectral truncation, which is approximately 2.8° intervals for latitude and longitude. The atmospheric vertical resolution is 40 layers up to 3 hPa. The horizontal grid of the ocean model is built on a tripolar system that is divided horizontally into 360×256 grid points. (To the south of 63° N, the longitudinal grid spacing is 1° and the meridional spacing becomes fine near the Equator. In the central Arctic Ocean, the grid spacing is finer than 1° because of the tripolar system.) The ocean model has 62 vertical levels. VISIT simulates carbon and nitrogen dynamics on land. The OECO2 is a nutrient–phytoplankton–zooplankton–detritus-type model that is an extension of the previous model, MIROC-ESM (Watanabe et al., 2011). The effective climate sensitivity in MIROC-ES2L is lower than the previous version of MIROC-ESM (4.7 °C for MIROC-ESM, see Andrews et al., 2012, and 2.7 °C for MIROC-ES2L, see Tsutsui, 2020).

2.2.5 MPI-ESM1.2-LR

The Max-Planck-Institute Earth-System-Model (MPI-ESM) is composed of four components: the atmospheric general circulation model ECHAM6 (Stevens et al., 2013), the ocean-sea ice model MPIOM (Jungclaus et al., 2013), the land component JSBACH (Reick et al., 2013), which is directly coupled to ECHAM6, and the ocean biogeochemistry model HAMOCC (Ilyina et al., 2013), which is directly coupled to MPIOM. VolMIP experiments are performed with the MPI-ESM version 1.2 (Mauritsen et al., 2019) in its low resolution (MPI-ESM1.2-LR). In MPI-ESM1.2-LR, ECHAM6.3 is run with a horizontal resolution of T63 (~200 km) and 47 vertical levels, while MPIOM is run with a nominal horizontal resolution of

1.5° and 40 vertical levels. ECHAM6.3 includes modifications of the convective mass flux, convective detrainment and turbulent transfer, the fractional cloud cover and a new representation of radiative transfer with respect to its CMIP5 version (Stevens et al., 2013), while MPIOM remained largely unchanged with respect to the CMIP5 version of MPI-ESM (Jungclaus et al., 2013). A detailed description of all MPI-ESM1.2 updates is given in Mauritsen et al. (2019)., which contains ECHAM6.3. Climate sensitivity in the MPI-ESM1.2 was tuned to match the instrumental record warming by targeting an equilibrium climate sensitivity (ECS) of about 3 K using cloud feedbacks.

2.2.6 UKESM1

The first version of the UK Earth System Model (UKESM1) is fully described by Sellar et al. (2019). UKESM1 differs from its predecessor HadGEM2-ES (Collins et al., 2011) in being developed via a partnership between the UK Met Office and the UK Universities (funded via the UK Natural Environment Research Council). UKESM1 is built around the physical atmosphere-ocean climate model HadGEM3-GC3.1 (Kuhlbrodt et al., 2018; Williams et al., 2018), combining the Global Atmosphere 7.1 (GA7.1) configuration of the UK Met Office Unified Model (Walters et al., 2019; Mulcahy et al., 2018) with the NEMO ocean model (Storkey et al., 2018), the CICE sea-ice model (Ridley et al., 2018) and the JULES land-surface model (Best et al., 2011). Two major developments since HadGEM2-ES include the atmosphere physical model having a well-resolved stratosphere with stratosphere-troposphere chemistry (Archibald et al., 2020) and tropospheric aerosol radiative forcings from the GLOMAP-mode modal aerosol microphysics module (Mann et al., 2010; Bellouin et al., 2013). Another important progression since HadGEM2-ES is that the UKESM1 terrestrial biogeochemistry module within JULES (Clark et al., 2011) has coupled carbon and nitrogen cycles, also with enhanced land management. The main characteristics and behaviour of the UKESM1 deck simulations for CMIP6 (pre-industrial control, abrupt 4xCO2, 1% increasing CO2 and post-industrial historical) are presented in Sellar et al. (2019).

2.2.7 Forcing implementation

The VolMIP protocol recommends volcanic forcing input data below the model tropopause to be replaced by climatological or other values of tropospheric aerosol used by the models (see Zanchettin et al., 2016). This is a potential source of inter-model disagreement already at the forcing level, given differences across models in the vertical structure of the simulated atmosphere and choices made regarding the definition of the tropopause and the climatological reference value of tropospheric aerosol. The presence of forcing differences across participating models is illustrated in Fig. 1 by monthly values of the aerosol optical thickness at 550 nm due to stratospheric volcanic aerosols (variable aod550volso4[1]) averaged over the tropics (30°S-30°N). The amplitude of forcing differences between models vary through time, as can be seen for instance comparing results from IPSL-CM6A-LR and MPI-ESM1.2-LR, which match at the peak of the forcing but differ appreciably during the decaying phase. The largest difference occurs during the initial steep aerosol rise. Differences also concern the stratospheric background

---

[1] Variable names are reported as defined by the CMIP6 "Climate Model Output Rewriter"

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

aerosol, which is present in MPI-ESM1.2-LR, GISS-E2.1-G and UKESM1 but not in other models (note, MPI-ESM1.2-LR volc-pinatubo-full simulations started in January 1991).

2.2.8 Initial conditions

Following the VolMIP protocol, branching from the piControl simulations was designed to sample combined states of NAO and ENSO in such a way that a broad range of internal unperturbed variability is considered at the time of the peak of the applied forcing, i.e., during the first post-eruption winter (DJF 1992, with January setting the reference for the year). The approach therefore aims at selecting climate conditions at the time of the eruption that are preconditioning a broad variety of states of ENSO and NAO in the following winter, to determine whether the climate response is influenced by developing anomalies in such modes.

Figure 2 illustrates the sampled winter average ENSO and NAO states from the piControl simulations for all models. For both modes, indices are standardized where the original DJF time series is modified by removing the long-term average calculated over the whole piControl and then by dividing it by the square root of the variance calculated over the whole piControl. For all models, the sampling complies with the VolMIP protocol as the homogeneous spread of the scatterplots encompassing all quadrants confirms that different unperturbed coupled states of ENSO and NAO are considered corresponding to the first post-eruption winter (DJF 1992, Fig. 2a). Still, some differences across models are apparent, for instance the range of sampled ENSO states, in the standardized DJF Niño3.4 index, is comparatively smaller in CanESM5 than in other models. Also, small biases in the average sampled state of ENSO are appreciable, e.g., the bias to positive ENSO in MIROC-ES2L and to negative ENSO in MPI-ESM1.2-LR and UKESM1. The inter-model differences reflect the application of different sampling algorithms and/or subjective choices. Concerning the first point, initial states of MIROC-ES2L were sampled at 200-year intervals without an explicit consideration of the corresponding ENSO and NAO states. Then, the circular structure emerging for some models in the NAO-ENSO scatterplot corresponding to the last pre-eruption boreal winter (DJF 1991, Fig. 2b) reveals how some modeling groups (CanESM5, IPSL-CM5A-LR and MPI-ESM1.2-LR) targeted the last pre-eruption boreal winter and not the first post-eruption boreal winter, for the selection of initial states. This was done using a sampling algorithm yielding the visible circular structure. The sampling strategy is further discussed in Sect. 4.5.

The simulations are then started on May 31st of the year of the Pinatubo eruption for all models except MPI -ESM1.2-LR, for which the simulations are started on January 1st of the year of the eruption due to technical reasons. This study uses an ensemble of 25 simulations - the minimum ensemble size set by the protocol - for each contributing model, unless otherwise specified. Note that more realizations are available from certain models, for instance 121 realizations are available for GISS-E2.1-G and 40 realizations are available for CanESM5. The realizations considered here are those labeled from r1 to r25 in the metadata.

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

3 Statistical methods and diagnostics

Climate responses to the volcanic forcing are quantified as anomalies with respect to the unperturbed climatology in each model. These are differences between the output of each realization in the volc-pinatubo-full multi-model ensemble and the climatology calculated for the whole length of the piControl, including seasonality. The anomaly method assumes that the unperturbed climate is characterized by uncorrelated white noise, which adds to the forced response in the volc-pinatubo-full simulations. Accordingly, the expected forced response can then be calculated as the ensemble mean anomaly whereas the spread of anomalies reflects noise. The method smears out climate variations on seasonal or longer timescales that emerge in the volc-pinatubo-full simulations and may have already been in progress in the piControl at the time chosen to start the volc-pinatubo-full simulation and might therefore be included in the calculation of the response. For instance, such variations can be due to ocean dynamics or sea-ice changes. This could be relevant, for instance, considering the biases in the averaged sampled states of ENSO in some models (Fig. 2a). Additional approaches to the quantification of the climate responses have been tested, including calculation of paired anomalies, i.e., deviations of the volc-pinatubo-full realizations from the corresponding branch of the piControl. These alternative methods occasionally yield different results compared to the anomaly method shown in the main results concerning the expected climate response, i.e., the ensemble mean, and the evaluation of statistical significance of inter-model differences. Accordingly, different approaches are discussed whenever deemed necessary.

The significance of inter-model differences in the multi-model ensemble is estimated based on the Mann-Whitney U test where the ensemble of each model is tested against the aggregated multi-model ensemble of the other models.

The analysis is performed on monthly values of selected relevant diagnostics spatially averaged over four regions. These are the full globe (hereafter GL), the Northern Hemisphere extratropics (30°-90°N, hereafter NH), the tropics (30°S-30°N, hereafter TR), and the Southern Hemisphere extratropics (30°-90°S, hereafter SH).

A simple assessment of the global upward LW radiation flux across the atmospheric column and of the cloud-albedo feedback in the tropics and their dependence on the mean state of the unperturbed climate is performed (Sect. 4.4). Specifically, the atmospheric LW transmittance is diagnosed through the ratio $LWt/LWs\uparrow$, where LWt is the global average top-of-atmosphere LW radiation (rlut) and $LWs\uparrow$ is the global average upward LW radiation at the surface (rlus) (e.g., Zanchettin et al., 2013). Cloud-albedo interactions in the tropics are diagnosed through the ratio SWt/SWtcs, where SWt (rsut) and SWtcs (rsutcs) are the top-of-atmosphere upward solar radiation under full-sky and clear-sky conditions, respectively, over the TR region. For both diagnostics, anomaly values below one are associated with a strengthening of the underlying processes and feedbacks.

The effect of ensemble size on the uncertainty estimation in the expected climate response (i.e., ensemble mean) is quantified for each model through changes in the standard error of the ensemble mean calculated for different ensemble sizes (Sect. 4.6). In practice, for each ensemble size from 3 to 25, all possible permutations of the full ensemble for the considered size are retrieved. Then, for each size and sub-ensemble obtained from the permutations, the standard error of the ensemble mean is calculated for the anomalies of the variable of interest, i.e., the square root of the variance of the sub-ensemble anomalies divided by the square root of the sub-ensemble size is calculated. Then, means and 5th and 95th percentiles of the so-obtained standard errors for each ensemble size are plotted. The standard error is calculated for anomalies of annual mean values of the year 1992 for two key surface variables: global-mean near-surface air temperature and global-mean precipitation.

4 Results

4.1 Mean state and variability in piControl

The simulated mean state and variability of piControl, i.e., under unperturbed conditions, can affect the post-eruption response through excitation of internal climate modes by the eruption, which in turn also depends on their amplitude and phase at the time of the eruption (see Fig. 2 and Sect. 4.3), and/or through controlling the strength of climate feedbacks that operate through changes in the global-mean surface temperature (see Sect. 4.4). The climate state in piControl is illustrated in the form of Box-Whisker plots of relevant diagnostics calculated for the whole length of the simulations (see Table 1), including global and regional averages of annual-mean near-surface air temperature (Fig. 3a-d), and winter ENSO and NAO indices as defined by the VolMIP protocol (Fig. 3e,f).

There are substantial differences in the simulated near-surface air temperature across models: IPSL-CM6A-LR has an overall cooler climate compared to the other models, of about 1°C in the global-mean surface temperature compared to CanESM5, GISS-E2.1-G and MPI-ESM1.2-LR; in contrast, UKESM1 and MIROC-ES2L have a warmer climate, of more than 1°C compared to other models. The colder global conditions in IPSL-CM6A-LR stem mostly from colder tropics and southern extra-tropics, for the latter, even colder conditions than IPSL-CM6A-LR are found for CanESM5. However, IPSL-CM6A-LR yields substantially warmer and less variable winter sea-surface temperature in the equatorial Pacific compared to other models, possibly reflecting different biases in the models. These differences may affect climate feedbacks as well as dynamical responses. The warmer climate of MIROC-ES2L stems from substantially warmer extra-tropical regions compared to other models, which may affect the response in terms of, among other processes, meridional energy transports and sea ice-albedo feedback. In contrast, the warmer climate of UKESM1 mostly stems from a warmer tropical region compared to other models, which may affect especially cloud-albedo feedbacks operating there.

The DJF Niño3.4 index, which is used to illustrate ENSO, shows substantial differences in the distributions across individual models. IPSL-CM6A-LR yields a warmer mean state of ENSO (above 28 °C) and a smaller variance of ENSO compared to other models. The distribution of ENSO in IPSL-CM6A-LR does not overlap with those of CanESM5, MPI-ESM1.2-LR and UKESM1, whose mean state of ENSO is below 26 °C. ENSO shows a skewed distribution with a long tail toward strong El Niño events in MIROC-ES2L compared to the other models that yield rather symmetric distributions for the Niño3.4 index. There are similar differences across models concerning the NAO in terms of both mean state and variability of the mode. In particular, MIROC-ES2L displays a lower mean value of the non-standardized NAO index, indicating a smaller mean

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

difference between boxes hence a smaller meridional gradient in the 500 hPa geopotential height, and a smaller variance compared to other models.

4.2 TOA and surface radiative fluxes

Left panels in Fig. 4 show the net top-of-atmosphere (TOA) vertical radiative fluxes calculated as anomalies of incoming shortwave (rsdt) minus outgoing shortwave (rsut) minus outgoing longwave radiation (rlut). At the global scale, the models agree on a largest average negative anomaly (i.e., reduced downward flux) of about 2 Wm-2 occurring in the first post-eruption boreal winter and on a persistence of the volcanic perturbation to TOA fluxes until 2.5 years after the eruption, i.e., until the third post-eruption boreal winter. The ensemble spread is also largely consistent across models. There are inter-model differences during the first six post-eruption months, with models clustering into two groups, with slower increase of (hence smaller) radiative anomalies in IPSL-CM6A-LR, CanESM5, UKESM1 and MIROC-ES2L, and a faster increase of (hence larger) radiative anomalies in MPI-ESM1.2-LR and GISS-E2.1-G, with differences between clusters exceeding 0.5 Wm-2. The models agree remarkably on the extra-tropical response, whereas a significantly weaker response is found for IPSL-CM6A-LR in the tropics from the time of the eruption to mid 1992. The smaller TOA net flux anomaly in IPSL-CM6A-LR is produced by a combination of the LW and SW components, especially in the tropics (see supplementary Figs. S2 and S3). All models display weak changes in the outgoing LW radiation in the NH until the second post-eruption boreal summer (1992), when a rather sudden drop takes place consistently in all models (Fig. S3). This may reflect a change in the spatial structure of aerosol forcing, with the aerosol cloud predominating over the NH extratropics in the second post-eruption year. Clear-sky net TOA radiative flux anomalies, calculated as rsdt minus clear-sky outgoing shortwave (rsutcs) minus clear-sky outgoing longwave radiation (rlutcs) showing more significant inter-model differences than full-sky diagnostics for all considered regions (Fig. 4b,d,f,h).

Figure 5 illustrates anomalies of the surface net vertical radiative flux calculated as anomalies of downward shortwave (rsds) plus downward longwave (rlds) minus upward shortwave (rsus) minus upward longwave radiation (rlus). At the global scale, the models agree substantially as shown by the large overlap between ensemble means and envelopes. The largest reduction in the downward net surface fluxes occurs around the first post-eruption boreal autumn and winter, with average anomalies persisting on values below -2 Wm-2 well into the first post-eruption boreal spring. Two models stand out from the ensemble: MIROC-ES2L and GISS-E2.1-G, the former with weaker and the latter with stronger anomalies during the first two post-eruption years. As shown for the TOA fluxes, the models agree well in the extratropics while differences are strongest in the tropics. As also seen for the TOA fluxes, changes in the NH are small until the second post-eruption boreal summer.

Figure 6 illustrates anomalies for the surface upward vertical latent plus sensible heat (LH+SH) flux. All models similarly produce a weak global response, identified by small ensemble mean anomalies and large ensemble spread encompassing positive and negative values. However, the models agree on indicating a tendency toward negative upward heat surface flux anomalies during the first three post-eruption years, arguably linked with reduced surface temperatures and ocean heat losses to the atmosphere, and with a slower development of anomalies compared to the radiative fluxes (peak negative values are observed around the second post-eruption boreal summer). Again, the global response is largely determined by the tropics, with ensemble-mean heat flux anomalies in the SH extratropics remaining always around zero, suggesting a very small sensitivity to the forcing and/or small signal-to-noise.

4.3 Tropospheric and surface climate response

Figure 7 illustrates near-surface air temperature anomalies. At the global scale, there are only sporadic significant differences in the temperature response with maximum expected cooling ranging across models between about -0.27 °C and -0.38 °C and a multi-model mean of about -0.33 °C. During the post-eruption cooling, the difference between expected responses across models can exceed 0.15 °C, linked to the significantly weaker cooling in MIROC-ES2L compared to other models. Using paired anomalies, the consistency across models at the global scale is very strong in the first two post-eruption years, indicating a progressive cooling until late 1992 when a maximum cooling of about 0.3 °C is attained (Fig. S6). There is no evidence of a significantly weaker cooling in MIROC-ES2L compared to other models in the paired anomalies, which reveals that the response identified in Fig. 7 may reflect biased sampled states in this model (Fig. 2), as further discussed below. Thereafter, the ensemble-mean trajectories depart more from each other also in the paired anomalies, with a quicker recovery for MIROC-ES2L and a slower one for CanESM5 compared to other models. Anomalies remain negative to the end of the simulations, with values between around -0.08 and -0.12 °C in year 1995 in models that extended the integration to this time (IPSL-CM6A-LR, MPI-ESM1.2-LR and CanESM5).

The response in the tropics is seen to be the source of the occasional disagreement in the global-mean temperature across models, and reveals model specificities in the cooling phase that do not emerge at the global scale. This is the case for the intermittent significantly colder anomalies seen in MPI-ESM1.2-LR and the warmer conditions of IPSL-CM6A-LR during the second post-eruption boreal summer, which particularly emerge in the paired anomalies. This again suggests a possible effect of biases in the sampled initial conditions in the case of MPI-ESM1.2-LR, considering its slight negative average of sampled ENSO conditions in piControl at the time of peak forcing (Fig. 2) and the fact that paired anomalies do not yield significant differences among models (Fig. S6). Otherwise, this might reflect differences in the applied forcing (see Sect. 4.2) as well as the consequent activation of certain dynamical responses in only some models. In the extra-tropics, surface cooling is consistently stronger in the NH compared to the SH, which can be linked to the larger land cover in the former and to inter-hemispheric asymmetries in the exposure of polar regions to the volcanically induced radiative forcing anomalies. In the NH the response is negligible until the second post-eruption boreal summer, when hemispheric surface temperature anomalies drop until the following boreal winter to reach ensemble-mean values around -0.5 °C in MIROC-ES2L, CanESM5 and MPI-ESM1.2-LR, and around -0.7 °C in GISS-E2.1-G, IPSL-CM6A-LR and UKESM1.

Figure 8 illustrates the precipitation anomalies. The magnitude of the reduction of global-mean precipitation is similar in all models except MIROC-ES2L. Peak ensemble-mean anomalies are smaller than -0.05 mm/day in all models, which is small compared to the ensemble variability of the simulations. The difference between MIROC-ES2L and the other models is reduced if paired anomalies are considered (Fig. S7), which again points to a biased sampling of initial states in this model. The precipitation response is especially small in the extratropics, hence the global reduction of precipitation largely stems from a reduction of the tropical precipitation.

Figure 9 illustrates the response of ENSO and NAO, quantified using indices defined according to the VolMIP protocol, i.e., non-standardized box-based indices (see Sect. 2.2). Niño3.4 anomalies indicate a general tendency of ensemble means toward colder sea-surface temperatures anomalies in the first three post-eruption years, except for MIROC-ES2L with peak warm anomalies of around 1°C in late 1992 and, much less evident, GISS-E2.1-G and MPI-ESM1.2-LR. MIROC-ES2L stands out as significantly different from the other models both concerning the warming in 1992 and the following cooling around 1994. Considering the possible biased sampling of ENSO states, Figs. 9a,c and S9 illustrate paired anomalies of Niño3.4 sea-surface temperatures for the volc-pinatubo-full simulations and corresponding piControl sections, and the anomalies for such piControl sections from the climatology, respectively. Results indicate a weak yet consistent tendency toward cooling (around -0.25°C) until the first post-eruption boreal winter in volc-pinatubo-full and a divergence of ensemble-mean trajectories thereafter, with a clear tendency toward warming in MIROC-ES2L during the second post-eruption winter (Fig. 9c), which peaks at much lower values compared to the anomalies with respect to the climatology (Fig. 9). Anomalies with respect to the climatology for the piControl sections confirm that part of the response detected in MIROC-ES2L is in fact spurious and linked to a biased sampling of warm ENSO states in piControl (Fig. S8). Paired anomalies also show a smaller initial cooling signal in UKESM1 and a stronger initial cooling in CanESM5 compared to the anomalies from the climatology, up to -0.5°C around the third post-eruption summer. Overall, the models thus seem to agree on a weak La Niña-like response in the early phase, whereas the models disagree on the later response of ENSO, with some models suggesting a warm (El Niño-like) response and others suggesting a cold (La Niña-like) response. The different sign and timing of the response highlight the potential influence of the different simulation of ENSO dynamics in the different models. In addition to the sampling bias, two considerations are worthy: first, the large ensemble spread indicates a general low signal-to-noise ratio; then, Niño3.4 SST does not provide a good diagnostic to examine ENSO as dynamical responses may be masked by broad tropical radiative cooling effects.

The NAO response and inter-model agreement are also difficult to interpret based on the chosen diagnostic, due to the apparent low signal-to-noise ratio of the response. There is a weak tendency toward positive NAO anomalies in the first post-eruption winter in GISS-E2.1-G and IPSL-CM6A-LR, and earlier in UKESM1, which contrasts with tendential negative anomalies in CanESM5.

4.4 Feedbacks

Figure 10 illustrates diagnostics that relate to two examples of climate feedbacks. The first describes changes in the atmospheric LW transmittance through the LWt/LWs↑ ratio, defined as difference between the value of the LWt/LWs↑ ratio calculated for the volc-pinatubo-full simulations and for the corresponding piControl sections. This value integrates the effects of diverse processes, including absorption by the volcanic aerosol and high-level clouds, and several feedbacks including Planck, lapse rate and water vapor. The second describes changes in solar radiation linked to the cloud-albedo feedback, estimated through the SWt/SWtcs ratio. There is a tendential lowering of the LWt/LWs↑ ratio in all models, but possibly less strong for IPSL-CM6A-LR, which has a colder mean state of the global surface climate, and strongest for UKESM1, which has a warmer mean state of the global surface climate than all other models except MIROC-ES2L. The inter-model difference between models agrees with the general relation that a warmer climate has a stronger water vapor feedback, which is accounted for in this diagnostic, but other factors influence the thermal radiation response across the atmosphere, possibly including aerosol radiative effects. However, there seems to be no major difference between the first and the second post-eruption year in terms of LWt/LWs↑. Given the substantial difference in aerosol loading between both years, this again suggests that the diagnostics mostly reflect differences in feedbacks operating through changes in the LW radiation.

There is a clear strengthening of the cloud-albedo feedback in all models, as the associated diagnostic is largely below the value of 1 in all models. There seems to be a clustering between models, with MIROC-ES2L and GISS-E2.1-G with a strong response of the feedback during the first post-eruption year and a strong recovery in the second post-eruption year, IPSL-CM6A-LR with a comparatively weak response in the first post-eruption year but a stronger persistence of the signal in the second post-eruption year, and CanESM5, UKESM1 and MPI-ESM1.2-LR with an intermediate behavior. The warmer mean state of tropical temperatures of GISS-E2.1-G and MIROC-ES2L compared to other models (Fig. 3) suggest a dependency on temperature, although the clustering may reflect different choices in the parameterization of clouds in the different models as UKESM1, with the warmest climatological tropical temperatures, and CanESM5, with similar climatological temperatures to MIROC-ES2L, show weaker cloud responses. The diagnostics may also reflect differences in the applied forcing, although the linkage is nontrivial as, for instance, IPSL-CM6A-LR and UKESM1 have similar rsut/rsutcs values but different forcing (compare with Fig. 4).

4.5 Effect of sampling strategy

Figure 11 illustrates the effect of the sampling strategy, i.e., the considered ENSO and NAO conditions to start the volc-pinatubo-full simulations from the piControl following the VolMIP protocol, for the post-eruption climate anomalies. The figure compares empirical distributions for near-surface air temperature anomalies in terms of seasonal-average anomalies for the 1992 boreal summer grouped by different states of ENSO and NAO. For global-average near-surface air temperature, all models show smaller negative temperature anomalies on average in the realizations starting from ENSO+ pre-conditions, and stronger negative temperature anomalies on average for realizations starting from ENSO- pre-conditions. This result is unsurprising and can be explained by the global temperature anomaly resulting from the ENSO state at the time of maximum cooling superimposing on the volcanic cooling. Note, however, -that the anomalies contain the potential effect of sampling biases regarding ENSO, most importantly regarding MIROC-ES2L (Figs. 2 and S9): unbalanced sampling of ENSO can thus lead to biases in the global-average post-eruption anomalies for some models (see also, e.g., Lehner et al., 2016). This hypothesis is supported by analysis of paired anomalies (Fig. S9) showing no substantial change in the response as a function of ENSO pre-conditioning, or even opposite dependencies of the response on ENSO compared to what is seen in the anomaly analysis (compare Fig. 11 and Fig. S9 for, e.g., GISS-E2.1-G). In any case, the distributions of global-mean temperature anomalies for the different ENSO pre-conditions do overlap considerably in most cases also in the anomaly analysis, indicating that the effect of ENSO preconditioning can be overwhelmed by other factors contributing to internal variability. In particular, the distributions for IPSL-CM6A-LR overlap considerably, suggesting that in this model global cooling is weakly sensitive to the ENSO pre-conditioning. The NAO sampling affects the global response with an overall weaker impact compared to ENSO, with only some models showing differences in the ensemble-mean response under different NAO pre-conditioning. This can be understood with the weaker imprint of NAO on the global surface climate compared to ENSO, particularly in summer.

Similar considerations stand for regional cooling over areas deemed most impacted by the two considered modes, i.e., the tropics for ENSO and the Northern Hemisphere for the NAO. Preconditioning of ENSO clearly impacts tropical temperatures in the anomaly analysis, with all models agreeing on a weaker cooling under El-Niño preconditioning compared to neutral or La-Niña preconditioning. Again, paired anomalies weaken the dependency of the response to the state of ENSO, revealing that the evolution of post-eruption tropical temperature anomalies contain the signal of dynamics related to ENSO and unaffected by the forcing. The lack of impact of NAO sampling on NH temperatures can be again explained by the fact that the NAO is predominant in winter whereas direct radiative responses are better identified in the summer season. Further, the NAO hemispheric pattern is strongly heterogeneous and includes both warm and cold regional temperature anomalies within the Northern Hemisphere that tend to compensate for each other leaving a negligible imprint on hemispheric averages.

4.6 Ensemble size and spread

Figure 12 illustrates the effect of sample size on the uncertainty related to the expected (i.e., ensemble mean) surface temperature and precipitation response, shown in terms of standard error of the mean of post-eruption anomalies (see Sect. 2.2). The standard errors converge toward the value obtained for the 25-member ensemble in all models, starting from the higher and more uncertain estimates obtained for low ensemble sizes. Otherwise, the curves differ across models indicating that they disagree on how the ensemble size affects the standard error of the ensemble mean. Models rank similarly concerning errors in global-mean temperature and global-mean precipitation, reflecting similar relative uncertainty in the response of both variables.

For small ensemble sizes, the amplitude of the 5-95 percentile range of the standard error varies substantially across models, with larger values in IPSL-CM6A-LR, GISS-E2.1-G, MIROC-ES2L and UKESM1 compared to MPI-ESM1.2-LR and CanESM5. This reflects a weak signal-to-noise ratio of the response in the former group of models as seen in their larger full-size standard errors and highlights the exposure of these models to potentially large sampling biases in the expected response when it is estimated from a few events. Overall, uncertainty in the ensemble mean strongly depends on both, ensemble size and model, which, together with the variety of unperturbed climatologies expressed by the models in the piControl, prevents generalization and requires model-specific assessments of the signal-to-noise ratio of post-eruption anomalies.

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

5 Discussion

In the following, we illustrate major gaps of knowledge emerging from our analyses to be addressed in follow up studies (Sect. 5.1) and discuss possible improvements to the experimental design and the protocol of VolMIP in light of a possible second phase of the initiative (Sect. 5.2).

5.1 Gaps of knowledge

Overall, the volc-pinatubo-full results indicate a general agreement in the surface climate response to volcanic forcing among different models compared to previous results. The VolMIP protocol allows models to be compare more powerfully by sampling across different states of dominant climatic modes. This contrasts with the small yet significant inter-model differences in volcanic aerosol optical depth and post-eruption radiative flux anomalies that call for more in-depth analysis of the volc-pinatubo-full simulations and question the efficiency of the VolMIP protocol for constraints on the forcing data across models.

The apparent differences in the aerosol forcing implemented in the different models require further work to be fully assessed and understood. They may reflect differences in model physics, including radiative schemes and parameterizations of aerosol-SW and -LW interactions. We recommend first checking the details of the stratospheric aerosol optical, single scattering albedo and asymmetry parameter depth diagnosed for each model and identifying model specificities regarding the tropopause height, especially over the tropics, and choices in the replacement of aerosol data below the tropopause. Inter-model differences in radiative flux anomalies are seen both at the top-of-atmosphere in full-sky and especially clear-sky diagnostics, and at the surface. They may also indicate inter-model differences in model radiative codes that rely on different spectral band resolution and schemes for aerosol-radiation interactions, or in adjustments/feedbacks, for example cloud adjustments (Schmidt et al., 2018), or the global water balance (Wild, 2020). We recommend analysis of effective radiative forcing or instantaneous radiative forcing calculations (e.g., Smith et al., 2018). In this regard, the VolMIP protocol has defined a group of variables to diagnose volcanic instantaneous radiative forcing (Table 4 in Zanchettin et al., 2016), which were requested to generate volcanic forcing for the volc-pinatubo-surf/strat experiments and can be useful to better constrain the imposed aerosol forcing in the different models.

The idealized nature of the VolMIP experiments does not allow a direct comparison with observations, which must rely on output from full-forcing transient simulations. In this regard, analysis of CMIP6-historical simulations (Pauling et al., 2021) provide a much better agreement with observations compared to CMIP5-historical simulations concerning the global-mean surface temperature response to the 1991 Pinatubo eruption. A comparative assessment of inter-model consistency in the climate response to the 1991 Mt. Pinatubo eruption in CMIP6-historical and in the volc-pinatubo-full experiment could also help to clarify the impact of boundary conditions and choices regarding the correction to the volcanic aerosol input data to confine volcanic aerosol to the stratosphere for volc-pinatubo-full.

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

The tropics emerge as a key region to understand inter-model differences in the volc-pinatubo-full ensemble and assess the realism of the simulated climate response to volcanic forcing. Fiedler et al. (2020) analysed the simulated tropical precipitation across different phases of CMIP and found similar behaviors for CMIP5 and CMIP6 models. In both cases the expected post-eruption reduction in precipitation over land is stronger than what is indicated by observations. This suggests a too-strong response of tropical precipitation to volcanic aerosols persisting across different model generations. However, an explanation based on CMIP5 results indicates that post-eruption precipitation anomalies strongly depend on both the magnitude of applied volcanic forcing and the state of the ocean at the time of eruption (Paik et al., 2020). Our results confirm the strong dependence of the precipitation response - and more generally of the climate response - to both, the mean climate state and the phase of internal climate variability at the time of eruption, beyond the obvious considerations about differences in the magnitude of the applied forcing discussed above. This was highlighted here especially for the case of the biased sampling of ENSO conditions in one of the contributing models (MIROC-ES2L), which reverberated on global-scale responses of temperature and precipitation. In fact, despite our results suggesting that the implementation of the experiment protocol was overall effective for most of the contributing models, room for possible improvements is evident, especially the strictness of the sampling of initial conditions for the volc-pinatubo simulations.

The dependency of post-eruption anomalies on initial conditions emerges as one of the clearest results of our analysis. Based on our analysis, ENSO does not show a robust response neither across models nor within individual models, which implies that its evolution determined by ongoing intrinsic dynamics can significantly affect post-eruption anomalies at the global, hemispheric, and regional scales. Therefore, our results highlight how for an eruption like the 1991 Mt. Pinatubo a biased sampling of internal variability may lead to non-negligible biases in the estimation of the expected climate response and call for caution in the assessment of post-eruption anomalies. The use of paired anomaly calculations mitigates the effect of sampling biases. The role of initial conditions in shaping the climate response to larger magnitude eruptions has been subject of recent studies (e.g., Zanchettin et al., 2019; Pausata et al., 2020). Analysis of the output of the VolMIP volc-long-eq experiment, based on idealized climate simulations of the 1815 Mt. Tambora eruption, will provide context to the general conclusions drawn here for the 1991 Mt. Pinatubo eruption. Also, despite the broad scattering and large overlap of values of ENSO and NAO at the time of peak forcing under unperturbed and perturbed states (Fig. 9e) suggesting a lack of robust response of both modes to volcanic forcing, there are known limitations in the considered indices and further studies shall consider improved diagnostics for both modes of climate variability.

Biases in sampled internal variability may reverberate on misinterpretation of dynamical responses as well. For ENSO at least, conclusions in this regard require additional analyses, which must rely on more reliable diagnostics than the Niño3.4 index employed here, such as relative sea-surface temperatures or sea-surface heights as suggested by previous studies (e.g., Khodri et al., 2017). In addition, the VolMIP Tier 3 volc-pinatubo-slab experiment can provide very useful insights: it uses the same forcing as volc-pinatubo-full but a slab ocean in order to clarify the role of coupled atmosphere–ocean processes for the dynamical response of ENSO (Zanchettin et al., 2016).

Accounting for sampling of initial conditions is relevant when investigating other known dynamical responses, for instance the post-eruption Northern Hemisphere winter warming. Coupe and Robock (2021) found that if the observed sea-surface temperatures are prescribed, the NCAR Community Earth System Model, with the Community Atmospheric Model 5, realistically simulates the observed winter warming after the three largest volcanic eruptions of the late 20th Century, but it fails if the ocean model is coupled to the atmosphere. We foster investigation of the post-eruption winter warming simulated by the volc-pinatubo-full ensemble, and recommend that results are interpreted accounting for the state of ocean variability in each simulation and also for climatological biases/differences in ocean-atmosphere coupled processes. Depending on the scientific question, Atmosphere Model Intercomparison Project (AMIP) style experiments with prescribed sea-surface temperatures might be an alternative approach to coupled climate experiments.

5.2 Implications for VolMIP

Recent advances in the design of climate model experiments makes some afterthoughts necessary regarding the VolMIP protocol, in particular concerning the sampling strategy. The identification of specific conditions of ENSO and NAO (or of any other relevant climatic mode) to start the volc-pinatubo simulations from the piControl might seem not necessary in light of the prospect to increase the integration and assessment of large ensemble experiments within the next phase of CMIP (Deser et al., 2020). If the way forward is toward the so-called “single model initial-condition large ensembles” (SMILEs), discussion about supervised sampling strategies may appear obsolete: SMILEs could provide many realizations of historical eruptions, including the 1991 Mt. Pinatubo, with good sampling of initial conditions as part of the DECK-historical simulations. However, the lack in transient simulations of unperturbed climate evolutions corresponding to periods during and after volcanic forcing would impede fully disentangling forced and intrinsic components of climate evolutions, as evidenced here for some relevant aspects of climate variability including ENSO. In this sense, idealized experiments as those originally proposed for VolMIP remain a valuable contribution to understanding the climate response to volcanic forcing and its simulation.

Another promising approach for the future is also the application and combination of different SMILEs. Maher et al. (2021) demonstrated the utility of combining different types of SMILEs to identify which part of post-eruption climate evolution is a response forced by the volcanic eruption and which one is due to other sources. The combination of different types of SMILEs might be a potential way to move forward to answer open scientific questions, such as the causes of post-eruption winter warming or post-eruption tropical sea-surface temperature variability, by separating and quantifying the forced response from internal variability on a regional scale.

There is a rich debate in the scientific literature about the use of climate anomalies after volcanic eruptions to infer equilibrium climate sensitivity (ECS) or transient climate sensitivity (e.g., Wigley et al., 2005; Boer et al., 2007; Merlis et al., 2014). Pauling et al. (2021) identify no robust connection between ECS and the post-Pinatubo global cooling in an ensemble of CMIP6 historical simulations. Figure 13 illustrates how ECS relates to seasonal-average near-surface air temperature anomalies in the volc-pinatubo-full ensemble. The results overall agree with the conclusion by Pauling et al. (2021) that ECS does not play an important role for the global-mean temperature response to a Pinatubo-like eruption, not in the expected

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

(average) response nor in its uncertainty. Again, MIROC-ES2L stands out from the other models, with smaller post-eruption cooling than observed for MPI-ESM1.2-LR, which has similar ECS. This difference is much reduced for calculations based on paired anomalies, with inter-quartile ranges of both models overlapping in the case of global-mean temperature anomalies for the second post-eruption boreal summer (not shown). It will be important to investigate this relation for the case of a stronger eruption, based on the volc-long experiments.

If activities are continued in a second phase of VolMIP with a Pinatubo-like set of experiments analogous to volc-pinatubo, we make the following considerations and propose the following improvements to the protocol.

Concerning the forcing, the original VolMIP core experiments focused on two historical tropical eruptions (Pinatubo, Tambora) with hemispherically symmetric forcing. However, the transport of aerosol from the tropics to each hemisphere is known to be quite variable for tropical eruptions depending on the eruption latitude and season. While the 1991 Mt. Pinatubo eruption produced a volcanic aerosol cloud that spread relatively evenly in the Northern and Southern Hemispheres, the volcanic aerosol distribution after the 1982 El Chichón eruption and the 1963 Agung eruption were heavily biased to one hemisphere. Previous studies on tropical eruptions already have pointed out the importance of asymmetric volcanic forcing on tropical rain belts or cyclone activities (e.g. Yang et al., 2019; Jacobsen et al., 2020) or for the comparison with proxy data (e.g. Timmreck et al., 2021). Therefore, while we support the vision that VolMIP must remain an idealized volcanic forcing experiment, improvement in the direction of accounting for inter-hemispheric forcing asymmetries should be discussed. In this regard, the original VolMIP protocol included two experiments with strongly asymmetric eruptions, namely volc-long-hlN (high-latitude eruption in the Northern Hemisphere) and volc-long-hlS (high-latitude eruption in the Southern Hemisphere). These experiments, currently set at priority levels Tier 2 and Tier 3, respectively, may provide valuable information as endmembers in an ensemble of idealized volcanic forcing experiments tagging uncertainties due to the spatial structure of the aerosol forcing.

Concerning ensemble size, length and sampling of initial conditions, the current recommended minimum ensemble size (25) seems to be sufficient whereas a longer integration time is proposed (minimum 5 years). Nonetheless, already in the current phase of VolMIP a number of contributing modeling groups generated a much larger ensemble. In the future, a balance must be established between the use of SMILEs and volcanic simulations with controlled selection of initial conditions from a control simulation. For the latter, we foresee a shift of focus from the radiative response to dynamical responses. Accordingly, we recommend shifting the focus only on ENSO for the sampling of initial conditions, since the NAO seems to have an only limited impact on the response and could therefore be neglected. We suggest the protocol be updated so that the ENSO mean state and tendency on the period from the last pre-eruption winter to the onset of the eruption is considered, instead of the state during the first post-eruption winter as in the original VolMIP protocol. Instead of indices based on sea-surface temperatures, we recommend using diagnostics that more closely tie to processes relevant for ENSO dynamics, for instance the equatorial Pacific ocean heat content described by indices such as the Warm Water Volume index (Meinen and McPhaden, 2000). This will allow identification of how ENSO preconditioning affects ENSO's response to the eruption and the role of the state of the equatorial Pacific at the time of eruption for the broad climatic response to be disentangled. Additional modes of climate variability may be considered, which can be identified based on their relevance for the response in follow-up composite analyses. Among potentially relevant modes, the Quasi Biennial Oscillation (QBO) did not have explicit focus in VolMIP, but it is arguable that its representation in climate models will continue improving. However, the prescribed aerosol optical properties at the basis of VolMIP constitute a major limitation to an effective implementation of a sampling strategy for the QBO, since the phase of the QBO affects stratospheric transport including that of volcanic aerosol, hence ultimately the volcanic forcing. The effects of inconsistencies between QBO and prescribed forcing on volc-pinatubo experiments are unknown. Until this gap of knowledge is filled, we recommend continuing to sample an easterly phase of the QBO at the time of the eruption whenever possible.

The phase of modes of variability with longer characteristic timescales may be important as well. For instance, Illing et al. (2018) identify significant regional differences in near-surface air temperature over the North Atlantic, sea-ice area fraction, frost days, and precipitation between two Pinatubo-like experiments, which were initialized in years with different phases of the Pacific Decadal Oscillation. Then, the state of the North Atlantic ocean circulation as described by the Atlantic Multidecadal Variability may affect atmospheric responses to the volcanic eruption as well (e.g., Omrani et al., 2016; Coupe and Robock, 2021). Therefore, their consideration in the sampling protocol should be considered, either with a strict explicit definition of their phase or with its a posteriori assessment in case of response biases across models.

The usefulness of the volc-pinatubo-full experiment cannot be fully understood unless in connection with the companion volc-pinatubo-strat/surf experiments, which are also Tier 1 VolMIP experiments and were designed to disentangle dynamical responses to the two primary thermodynamic consequences of aerosol forcing, i.e., surface cooling and stratospheric heating. Analysis of these experiments will allow us to clarify the main pathways through which volcanic aerosols affect atmospheric circulation and surface climates. This type of mechanistic experiments might be useful also for new questions to be addressed in a potential second phase of VolMIP that focus on the impact of volcanic aerosol on stratospheric/atmospheric dynamics and chemistry.

As a final consideration, a general critical aspect about VolMIP is the long turn over time between the experiment design, the integration of the simulations, and the analysis of the output. When most participating modeling groups performed the volc-pinatubo-full simulations, the experiment protocol was over five years old. Hence, some of the questions raised in the VolMIP overview paper in 2016 that steered the set up of VolMIP experiments have been answered in the meantime, while new questions have arisen. Still, we have outlined the potential for VolMIP to contribute to answering these emergent new questions, thanks to its well designed experimental protocol and, especially to the international community that has been built around the initiative.

6 Conclusions

First results from the VolMIP volc-pinatubo-full experiment reveal a dichotomy in the simulated climate response to a Pinatubo-like eruption, which is seen as broad inter-model consistency of post-eruption surface climate and hydroclimate anomalies contrasting with small yet significant differences in post-eruption radiative flux anomalies. Despite further analysis of the output of volc-pinatubo-full needed to explain such inter-model differences, the preliminary results shown here indicate that they reflect differences in the applied forcing. As well-constrained volcanic forcing is a pillar of VolMIP, any ambiguity in the protocol - possibly the treatment of volcanic aerosol input data at and below the tropopause - shall be amended in a possible second phase of the initiative. Then, the statistical consistency diagnosed in the near-surface air temperature and precipitation response may simply reflect the large intrinsic variability of the associated processes compensating for forcing uncertainties, which is also seen in single-model analyses as dependency of the response on the climate state at the time of eruption. Improved assessment of initial condition influences on direct radiative and dynamical responses is therefore also recommended towards a refinement of the volc-pinatubo sampling protocol.

Code and data availability

The CMIP6 data are available at https://esgf-data.dkrz.de/projects/esgf-dkrz/. The time series used in the analysis will be made available via permanent public repository with doi upon final publication. In the meanwhile, they are available at: https://vesg.ipsl.upmc.fr/thredds/catalog/VOLMIP/volc-pinatubo-full/catalog.html.

Competing interests

The authors declare no competing interest.

Acknowledgments

Climate modeling at GISS is supported by the NASA Modeling, Analysis and Prediction program, and resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. The simulations by MIROC-ES2L used the Earth Simulator and JAMSTEC Super Computing System, supported by TOUGOU/SOUSEI, the Integrated Research Program for Advancing Climate Models (grant number JPMXD0717935715)/Program for Risk Information on Climate Change, through the Ministry of Education, Culture, Sports, Science, and Technology of Japan. MPI-ESM1.2-LR simulations were performed at the Deutsches Klimarechenzentrum (DKRZ). This work used DKRZ resources granted by its Scientific Steering Committee (WLA) under project ID bm0963. This work used the French HPC resources of TGCC under the allocations 2016-A0030107732, 2017-R0040110492, and 2018-R0040110492 (project gencmip6) provided by GENCI (Grand Équipement National de Calcul Intensif). As such it benefited from the French state aid managed by the ANR under the “Investissements d'avenir” program with the reference ANR-11-IDEX-0004-17-EURE-0006. This study benefited from the ESPRI (Ensemble de Services Pour la

Recherche l'IPSL) computing and data center (https://mesocentre.ipsl.fr) which is supported by CNRS, Sorbonne Université, École Polytechnique, and CNES and through national and international grants.

Financial support

This research has been supported by the Deutsche Forschungsgemeinschaft Research Unit VolImpact (FOR2820, grant no. 398006378, CT) within the project VolClim, and the German Federal Ministry of Education and Research (BMBF) within the research programme “ROMIC-II, ISOVIC” (FKZ: 01LG1909B, SWF). The simulation by MIROC-ES2L has been supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (Integrated Research Program for Advancing Climate Models, grant no. JPMXD0717935715). This work was undertaken in the framework of the French L-IPSL LABEX and the IPSL Climate Graduate School EUR and benefited from the FNS "SYNERGIA "EffeCts of lArge voLcanic eruptions on climate and societies: UnDerstanding impacts of past Events and related subsidence cRises to evAluate potential risks in the future"(CALDERA) project under French CNRS Grant Agreement number CRSII5_183571 - CALDERA. AR is supported by US National Science Foundation grant AGS-2017113.

Author contributions

DZ, CT, and MK designed the study. DZ performed the statistical analyses, drafted the initial manuscript and coordinated the writing process. CT, MK, AS and MT contributed to writing the paper. CT performed the experiments with the MPI-ESM1.2-LR model with the help of SWF; MK and NL the IPSL-CM6A-LR simulations with the help of SB; KT performed the GISS-E2.1-G simulations with the help of HW and ANL; MA performed the MIROC-ES2L simulations. All authors contributed to discussion and finalization of the manuscript.

References

Andrews, T., Gregory, J. M., Webb, M. J., and Taylor, K. E.: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models, Geophys Res Lett 39:1–7, https://doi.org/10.1029/2012GL051607, 2012.

Archibald, A. T., O’Connor, F. M., Abraham, N. L., Archer-Nicholls, S., Chipperfield, M. P., Dalvi, M., Folberth, G. A., Dennison, F., Dhomse, S. S., Griffiths, P. T., Hardacre, C., Hewitt, A. J., Hill, R. S., Johnson, C. E., Keeble, J., Köhler, M. O., Morgenstern, O., Mulcahy, J. P., Ordóñez, C., Pope, R. J., Rumbold, S. T., Russo, M. R., Savage, N. H., Sellar, A., Stringer, M., Turnock, S. T., Wild, O., and Zeng, G.: Description and evaluation of the UKCA stratosphere–troposphere chemistry scheme (StratTrop vn 1.0) implemented in UKESM1, Geosci. Model Dev., 13, 1223–1266, https://doi.org/10.5194/gmd-13-1223-2020, 2020.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. Geophysics Model Development, 8, 2465– 2513. https://doi.org/10.5194/gmd-8-2465-2015, 2015.

Bellouin, N., Mann, G. W., Woodhouse, M. T., Johnson, C., Carslaw, K. S., and Dalvi, M.: Impact of the modal aerosol scheme GLOMAP-mode on aerosol forcing in the Hadley Centre Global Environmental Model, Atmos. Chem. Phys., 13, 3027–3044, https://doi.org/10.5194/acp-13-3027-2013, 2013.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description ? Part 1: Energy and water fluxes, Geosci. Model Dev., 4(3), 677–699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

Bittner, M., Schmidt, H., Timmreck, C., Sienz, F.: Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty. Geophys Res Lett., 43(17):9324–32., 2016.

Boer, G., Stowasser, M., and Hamilton, K.: Inferring climate sensitivity from volcanic events, Climate dynamics, 28, 481–502, https://doi.org/10.1007/s00382-006-0193-x, 2007.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, E., Lionel, Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, Journal of Advances in Modeling Earth Systems, 12, e2019MS002010, https://doi.org/10.1029/2019MS002010, 2020

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description ? Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4(3), 701–722, https://doi.org/10.5194/gmd-4-701-2011, 2011.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S.: Development and evaluation of an Earth-System model ? HadGEM2, Geosci. Mod. Dev., 4(4), 1051–1075, https://doi.org/10.5194/gmd-4-1051-2011, 2011.

Coupe, J. and Robock, A.: The Influence of Stratospheric Soot and Sulfate Aerosols on the Northern Hemisphere Wintertime Atmospheric Circulation, J. Geophys. Res. Atmos., 126, e2020JD034513, https://doi.org/10.1029/2020JD034513, 2021.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., and others: Insights from Earth system model initial-condition large ensembles and future prospects, Nat. Clim. Chang., 10, 277–286, https://doi.org/10.1038/s41558-020-0731-2, 2020.

d'Orgeval T., Polcher, J., Rosnay, P. d.: Sensitivity of the west African hydrological cycle in orchidee to infiltration processes. Hydrol. Earth Sys. Sci., 12(6):1387–1401, 2008.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fiedler, S., Crueger, T., D?Agostino, R., Peters, K., Becker, T., Leutwyler, D., Paccini, L., Burdanowitz, J., Buehler, S. A., Cortes, A. U., Dauhut, T., Dommenget, D., Fraedrich, K., Jungandreas, L., Maher, N., Naumann, A. K., Rugenstein, M., Sakradzija, M., Schmidt, H., Sielmann, F., Stephan, C., Timmreck, C., Zhu, X., and Stevens, B.: Simulated Tropical Precipitation Assessed across Three Major Phases of the Coupled Model Intercomparison Project (CMIP), Mon. Wea. Rev., 148, 3653–3680, https://doi.org/10.1175/MWR-D-19-0404.1, 2020.

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S., and Kawamiya, M.: Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks, Geosci. Model Dev., 13, 2197–2244, https://doi.org/10.5194/gmd-13-2197-2020, 2020.

Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., Musat, I., Idelkadi, A., Fairhead, L., Foujols, M.-A., Mellul, L., Traore, A.-K., Dufresne, J.-L., Boucher, O., Lefebvre, M.-P., Millour, E., Vignon, E., Jouaud, J., Diallo, F. B., Lott, F., Caubel, A., Meurdesoif, Y., and Ghattas, J.:. LMDZ6: Improved atmospheric component of the IPSL coupled model. Journal of Advances in Modeling Earth Systems, 12, e2019MS001892, https://doi.org/10.1029/2019MS001892, 2020.

Illing, S., Kadow, C., Pohlmann, H., Timmreck, C.: Assessing the impact of a future volcanic eruption on decadal predictions. Earth Syst. Dynam., 9, 701–715, https://doi.org/10.5194/esd-9-701-2018, 2018.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., and N??ez-Riboni, I.: Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, J. Adv. Model. Earth Sy., 5, 287–315, https://doi.org/10.1029/2012MS000178, 2013.

Jacobson, T, W, P., Yang, W., Vecchi, G. A:,Horowitz, L.W.: Impact of volcanic aerosol hemispheric symmetry on Sahel rainfall, Climate Dynamics, 55, 1733 – 1758, https://doi.org/10.1007/s00382-020-05347-7, 2020.

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., Mikolajewicz, U., Notz, D., and von Storch, J. S.: Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, J. Adv. Model. Earth Sy., 5, 422–446, https://doi.org/10.1002/jame.20023, 2013.

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T. L., Cook, B. I., Cruz, C. A., Del Genio, A. D., Elsaesser, G. S., Faluvegi, G., Kiang, N. Y., Kim, D., Lacis, A. A., Leboissetier, A., LeGrande, A. N., Lo, K. K., Marshall, J., Matthews, E. E., McDermid, S., Mezuman, K., Miller, R. L., Murray, L. T., Oinas, V., Orbe, C., Garc?a-Pando, C. P., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Shindell, D. T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G., Weng, E., Wu, J., and Yao, M.-S.: GISS-E2.1: Configurations and Climatology, J. Adv. Model. Earth Sy., 12, e2019MS002025, https://doi.org/10.1029/2019MS002025, 2020.

Khodri, M., Izumo, T., Vialard, J., Janicot, S., Cassou, C., Lengaigne, M., Mignot, J., Gastineau, G., Guilyardi, E., Lebas, N., Robock, A., McPhaden, M. J.: Tropical explosive volcanic eruptions can trigger El Niño by cooling tropical Africa. Nat. Commun. 8, 778, 2017.

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., Hill, R., Graham, T., Ridley, J., Blaker, A., Calvert, D., Copsey, D., Ellis, R., Hewitt, H., Hyder, P., Ineson, S., Mulcahy, J., Siahaan, A., and Walton, J.: The Low-Resolution Version of HadGEM3 GC3.1: Development and Evaluation for Global Climate, J. Adv. Model. Earth Sy., 10, 2865–2888, https://doi.org/10.1029/2018MS001370, 2018.

Lacis, A., Hansen, J., and Sato, M.: Climate forcing by stratospheric aerosols, Geophys. Res. Lett., 19, 1607–1610, https://doi.org/10.1029/92GL01620, 1992.

Lehner, F., Schurer, A. P., Hegerl, G. C., Deser, C., and Fr?licher, T. L.: The importance of ENSO phase during volcanic eruptions for detection and attribution, Geophys. Res. Lett., 43, 2851–2858, https://doi.org/10.1002/2016GL067935, 2016.

Luo, B.: Stratospheric aerosol data for use in CMIP6 models, ftp://iacftp.ethz.ch/pub_read/luo/CMIP6/Readme_Data_Description.pdf (last access: 12 August 2019), 2018a.

Luo, B.: Release Notes Stratospheric Aerosol Radiative Forcing and SAD version v4.0.01850 - 2016, ftp://iacftp.ethz.ch/pub_read/luo/CMIP6_SAD_radForcing_v4.0.0/Release_note_v4.0.0.pdf (last access: 12 August 2019), 2018b.

Madec, G., Bourdallé-Badie, R., Bouttier, P. A., Bricaud, C., Bruciaferri, D., Calvert, D., and Vancoppenolle, M.: NEMO ocean engine (Version v3.6). Notes du Pôle de modélisation de l'Institut Pierre-simon Laplace (IPSL), 27. https://doi.org/10.5281/zenodo.1472492, 2017.

Maher, N., Milinski, S., and Ludwig, R.: Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble, Earth Syst. Dynam., 12, 401–418, https://doi.org/10.5194/esd-12-401-2021, 2021.

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., Pickering, S. J., and Johnson, C. E.: Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model, Geosci. Model Dev., 3, 519–551, https://doi.org/10.5194/gmd-3-519-2010, 2010.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fl?schner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimen?z-de-la-Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., M?bis, B., M?ller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2, J. Adv. Model. Earth Sy., 11, 998–1038, https://doi.org/10.1029/2018MS001400, 2019.

Meinen, C. S. and McPhaden, M. J.: Observations of Warm Water Volume Changes in the Equatorial Pacific and Their Relationship to El Niño and La Niña, Journal of Climate, 13, 3551–3559, https://doi.org/10.1175/1520-0442(2000)013<3551:OOWWVC>2.0.CO;2, 2000.

Merlis, T. M., Held, I. M., Stenchikov, G. L., Zeng, F., and Horowitz, L. W.: Constraining Transient Climate Sensitivity Using Coupled Climate Model Simulations of Volcanic Eruptions, J. Climate, 27, 7781–7795, https://doi.org/10.1175/JCLI-D-14-00214.1, 2014.

Mulcahy, J. P., Jones, C., Sellar, A., Johnson, B., Boutle, I. A., Jones, A., Andrews, T., Rumbold, S. T., Mollard, J., Bellouin, N., Johnson, C. E., Williams, K. D., Grosvenor, D. P., and McCoy, D. T.: Improved Aerosol Processes and Effective Radiative Forcing in HadGEM3 and UKESM1, J. Adv. Model. Earth Sy., 10, 2786–2805, https://doi.org/10.1029/2018MS001464, 2018.

Omrani, N.-E., Bader, J., Keenlyside, N. S., and Manzini, E.: Troposphere–stratosphere response to large-scale North Atlantic Ocean variability in an atmosphere/ocean coupled model, Clim Dyn , 46, 1397–1415, https://doi.org/10.1007/s00382-015-2654-6, 2016.

Paik, S., Min, S.-K., Iles, C. E., Fischer, E. M., Schurer, A. P.: Volcanic-induced global monsoon drying modulated by diverse El Niño responses. Science Advances, 6, 21, doi:10.1126/sciadv.aba1212, 2020.

Pauling, A. G., Bushuk, M., and Bitz, C. M.: Robust Inter-Hemispheric Asymmetry in the Response to Symmetric Volcanic Forcing in Model Large Ensembles, Geoph. Res. Lett., 48, e2021GL092558, https://doi.org/10.1029/2021GL092558, 2021.

Pausata, F. S. R., Zanchettin, D., Karamperidou, C., Caballero, R., and Battisti, D. S.: ITCZ shift and extratropical teleconnections drive ENSO response to volcanic eruptions, Science Advances, 6, eaaz5006, https://doi.org/10.1126/sciadv.aaz5006, 2020.

Reick, C. H., Raddatz, T., Brovkin, V., and Gayler, V.: Representation of natural and anthropogenic land cover change in MPI-ESM, J. Adv. Model. Earth Sy., 5, 459–482, https://doi.org/10.1002/jame.20022, 2013.

Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., and Schroeder, D.: The sea ice model component of HadGEM3-GC3.1, Geosci. Model Dev., 11, 713–723, https://doi.org/10.5194/gmd-11-713-2018, 2018.

Rieger, L. A., Cole, J. N. S., Fyfe, J. C., Po-Chedley, S., Cameron-Smith, P. J., Durack, P. J., Gillett, N. P., and Tang, Q.: Quantifying CanESM5 and EAMv1 sensitivities to Mt. Pinatubo volcanic forcing for the CMIP6 historical experiment, Geosci. Model Dev., 13, 4831–4843, https://doi.org/10.5194/gmd-13-4831-2020, 2020.

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., Benshila, R., Chanut, J., Levy, C., Masson, S., and Vivier, F.: The Louvain-La-Neuve sea ice model LIM3.6: Global and regional capabilities. Geophysics Model Development, 28, 2991– 3005. https://doi.org/10.5194/gmd-8-2991-2015, 2015.

Sato, M., Hansen, J. E., McCormick, M. P., and Pollack, J. B.: Stratospheric aerosol optical depths, 1850-1990, J. Geophys. Res., 98, 22987–22994, https://doi.org/10.1029/93JD02553, 1993.

Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., Bardeen, C. G., Conley, A., Forster, P. M., Gettelman, A., Portmann, R. W., Solomon, S., and Toon, O. B.: Volcanic radiative forcing from 1979 to 2015, J. Geophys. Res.-Atmos., 123, 12–491, 2018.

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S. T., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G., and Zerroukat, M.: UKESM1: Description and Evaluation of the U.K. Earth System Model, Journal of Advances in Modelling Earth Systems, 11, 4513–4558, https://doi.org/10.1029/2019MS001739, 2019.

Smith, C. J., Kramer, R. J., Myhre, G., Forster, P. M., Soden, B. J., Andrews, T., et al.: Understanding rapid adjustments to diverse forcing agents. Geophysical Research Letters, 45,12,023–12,031, https://doi.org/10.1029/2018GL079826, 2018.

Stephenson, D. B., Pavan, V., Collins, M., Junge, M. M., and Quadrelli, R.: North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment, Clim. Dyn., 27, 401–420, doi:10.1007/s00382-006-0140-x, 2006.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, J. Adv. Model. Earth Sy., 5, 146–172, https://doi.org/10.1002/jame.20015, 2013.

Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., Calvert, D., Graham, T., Hewitt, H. T., Hyder, P., Kuhlbrodt, T., Rae, J. G. L., and Sinha, B.: UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions, Geosci. Model Dev., 11, 3187–3213, https://doi.org/10.5194/gmd-11-3187-2018, 2018.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019.

Swingedouw, D., Ortega, P., Mignot, J., Guilyardi, E., Masson-Delmotte, V., Butler, P. G., Khodri, M., and Séférian, R.: Bidecadal North Atlantic ocean circulation variability controlled by timing of volcanic eruptions, Nature Comm., 6, 1–12, https://doi.org/10.1038/ncomms7545, 2015.

Thomason, L. W., Ernest, N., Millán, L., Rieger, L., Bourassa, A., Vernier, J.-P., Manney, G., Luo, B., Arfeuille, F., and Peter, T.: A global space-based stratospheric aerosol \hack\breakclimatology: 1979–2016, Earth Syst. Sci. Data, 10, 469–492, https://doi.org/10.5194/essd-10-469-2018, 2018.

Timmreck, C., Toohey, M., Zanchettin, D., Brönnimann, S., Lundstad, E., and Wilson, R.: The unidentified eruption of 1809: a climatic cold case, Clim. Past, 17, 1455–1482, https://doi.org/10.5194/cp-17-1455-2021, 2021.

Tsutsui, J.: Diagnosing Transient Response to CO2 Forcing in Coupled Atmosphere-Ocean Model Experiments Using a Climate Model Emulator, Geophys. Res. Lett., 47, e2019GL085844, https://doi.org/10.1029/2019GL085844, 2020.

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G., & Morales Maqueda, M. A.: Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation. Ocean Modelling, 27, 54– 69. https://doi.org/10.1016/j.ocemod.2008.11.003, 2009.

Virgin, J. G., Fletcher, C. G., Cole, J. N. S., von Salzen, K., and Mitovski, T.: Cloud Feedbacks from CanESM2 to CanESM5.0 and their influence on climate sensitivity, Geosci. Model Dev. Discuss. [preprint], 14, 5355–5372, https://doi.org/10.5194/gmd-14-5355-2021, 2021.

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geosci. Model Dev., 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, 2019.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., and Kawamiya, M.: MIROC-ESM 2010: model description and basic results of CMIP5-20c3m experiments, Geosci. Model Dev., 4, 845–872, https://doi.org/10.5194/gmd-4-845-2011, 2011.

Wigley, T. M. L., Ammann, C. M., Santer, B. D., and Raper, S. C. B.: Effect of climate sensitivity on the response to volcanic forcing, J. Geophys. Res., 110, https://doi.org/10.1029/2004JD005557, 2005.

Wild, M.: The global energy balance as represented in CMIP6 climate models, Clim. Dyn., 55, 553–577, https://doi.org/10.1007/s00382-020-05282-7, 2020.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M.

J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations, Journal of Advances in Modeling Earth Systems, 10, 357–380, https://doi.org/10.1002/2017MS001115, 2018.

Yang, W., Vecchi, G. A., Fueglistaler, S., Horowitz, L. W., Luet, D. J., Muñoz, A. G., Paynter, D., and Underwood, S.: Climate Impacts From Large Volcanic Eruptions in a High-Resolution Climate Model: The Importance of Forcing Structure, Geophysical Research Letters, 46, 7690–7699, https://doi.org/10.1029/2019GL082367, 2019.

Zanchettin, D., Bothe, O., Graf, H. F., Lorenz, S. J., Luterbacher, J., Timmreck, C., and Jungclaus, J. H.: Background conditions influence the decadal climate response to strong volcanic eruptions, J Geophys Res Atm., 118, 4090–4106, https://doi.org/10.1002/jgrd.50229, 2013.

Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., Hegerl, G., Robock, A., Pausata, F. S. R., Ball, W. T., Bauer, S. E., Bekki, S., Dhomse, S. S., LeGrande, A. N., Mann, G. W., Marshall, L., Mills, M., Marchand, M., Niemeier, U., Poulain, V., Rozanov, E., Rubino, A., Stenke, A., Tsigaridis, K., and Tummon, F.: The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): experimental design and forcing \hack\newline input data for CMIP6, Geosci. Model Dev., 9, 2701–2719, https://doi.org/10.5194/gmd-9-2701-2016, 2016.

Zanchettin, D., Timmreck, C., Toohey, M., Jungclaus, J. H., Bittner, M., Lorenz, S. J., and Rubino, A.: Clarifying the Relative Role of Forcing Uncertainties and Initial-Condition Unknowns in Spreading the Climate Response to Volcanic Eruptions, Geophys. Res. Lett., 46, 1602–1611, https://doi.org/10.1029/2018GL081018, 2019.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., and Taylor, K. E.: Causes of Higher Climate Sensitivity in CMIP6 Models, Geophysical Research Letters, 47, e2019GL085782, https://doi.org/10.1029/2019GL085782, 2020.

Model	**Institute**	**Atm. Res.**	**Ocean Res.**	**QBO**	**Ozone**	**ECS (K)**	**Reference**
CanESM5	Canadian Centre for Climate Modelling and Analysis	~2.8°x2.8° 49 levels up to 1 hPa	1°x1°; 45 levels		Prescribed	5.6	(Swart et al., 2019)
GISS-E2.1-G	NASA Goddard	2°x2.5°; 40 levels up to 0.01 hPa	1°x1.25 °; 40 levels	No, easterly phase	Prescribed, based on prognostic	3.59	(Kelley et al., 2020)

	Institute for Space Studies			dominates	O3 simulations with the same model version (CMIP6 physics version 3)		
IPSL-CM6A-LR	Institut Pierre-Simon Laplace	1.25°x2.5°; 79 vertical levels up to 80km	1° nominal resolution with refinement of 1/3° in the equatorial region; 75 vertical levels	yes	Prescribed	5.01	Boucher et al. (2020)
MIROC-ES2L	Japan Agency for Marine-Earth Science and Technology/ Atmosphere and Ocean Research	~2.8°x2.8°; 40 levels up to 3 hPa	>63°N: ~60 km (zonal) x ~33 km (meridional), <63°N:	no	Prescribed	2.7	(Hajima et al., 2020)

	Institute, The University of Tokyo/ National Institute for environmental Studies		1° (long.) x 0.5-1° (lat.), 62 vertical levels				
MPI-ESM1.2-LR	MPI für Meteorologie	1.9°x1.9°, 47 lev, up to 0.01 hPa, 13 lev. above 100 hPa	GR15, 40 levels	no	prescribed	2.83	(Mauritsen et al., 2019)
UKESM1	UK Met Office and UK Natural Environment Research Council (NERC).	1.875° longitude x 1.25° latitude (~135 km horiz) with 85 vertical levels (hybrid-height) to 85km (50 of 85 levels below 18 km).	1°x1° horizontal resolution with 75 vertical levels	yes	Interactive via stratosphere - troposphere chemistry (Archibald et al., 2020)	4.7	(Sellar et al., 2019)

Table 1: Characteristics of the models participating in the volc-pinatubo-full experiment. ECS stands for Equilibrium Climate Sensitivity.

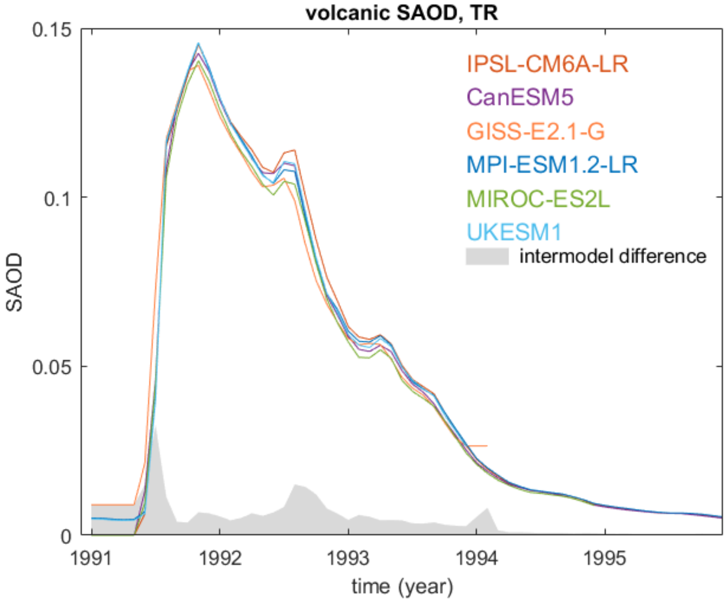


Figure 1: Monthly mean aerosol optical depth at 550 nm due to stratospheric volcanic aerosols (variable aod550volso4) averaged over the tropics (30°S-30°N) used in the volc-pinatubo-full experiments.

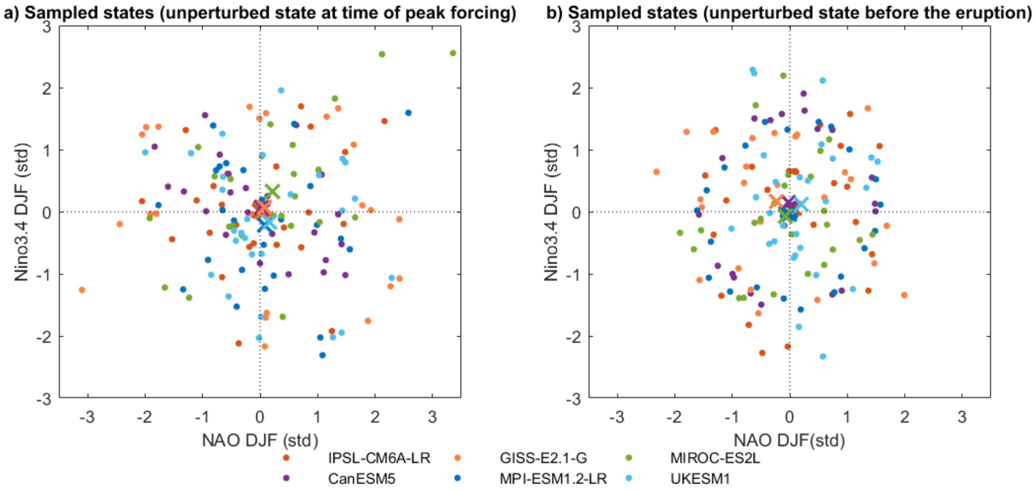


Figure 2: Scatterplots of the sampled states of winter NAO and Niño3.4 indices across the participating models. According to the VolMIP protocol, the sampled states in the piControl correspond to the first post-eruption winter in the volc-pinatubo-full simulations (DJF 1992) when there is a peak in the prescribed forcing dataset (panel a). Panel b shows sampled states during the winter preceding the eruption (DJF 1991). Each colored dot corresponds to one out of the 25 realizations for each model, crosses correspond to ensemble means. The color corresponding to each model is shown in the bottom of the figure. For both variability

modes, the shown values are from indices obtained by standardization of the original December-January-February average time series over the whole piControl simulation.

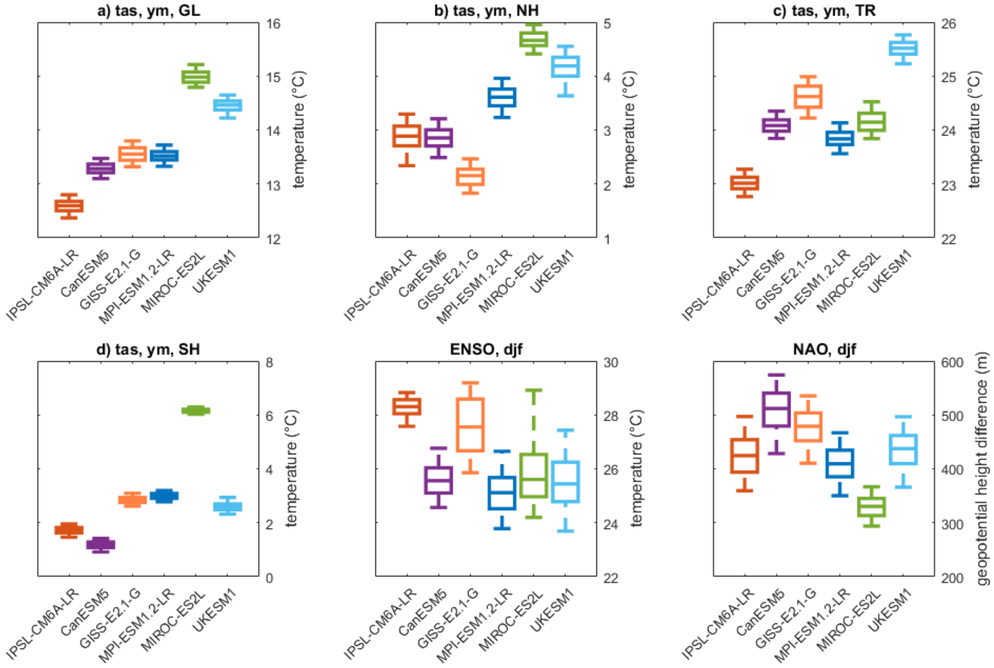


Figure 3: Mean state and variability of unperturbed climates of the participating models. Distributions of selected climatic parameters are shown as Box-Whisker plots (median, 25th-75th and 5th-95th percentile ranges) for the piControl. Panels a-d refers to near-surface annual-mean (ym) near-surface air temperature (tas) spatially averaged over the whole globe (GL), the Northern Hemisphere extratropics (NH) , the tropics (TR) and the Southern Hemisphere extratropics (SH). Following the VolMIP protocol, NAO is shown as the difference between the 500 hPa geopotential height in the two boxes that are used for the calculation of the index (see methods); ENSO is the average sea-surface temperature in the Niño3.4 region. Both indices are winter-average (December to February) time series.

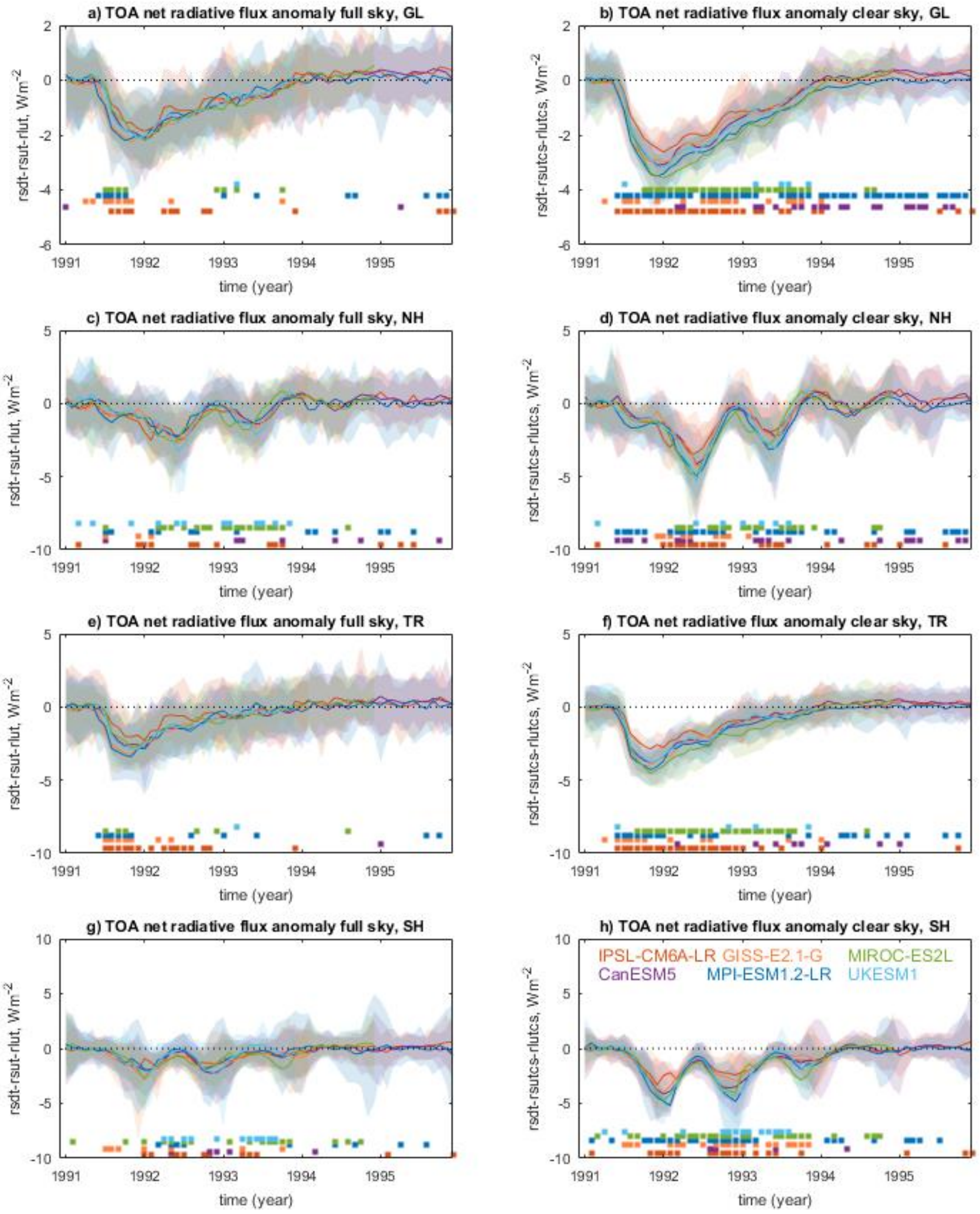


Figure 4: Top-of-atmosphere (TOA) net vertical radiative flux anomalies under full sky (left panels) and clear sky (right panels) conditions in the volc-pinatubo-full multi-model ensemble for a,b) global (GL), c,d) Northern Hemisphere extratropical (NH) , e,f) tropical (TR) and g,h) Southern Hemisphere extratropical (SH) mean. Full-sky anomalies are calculated as incoming shortwave (rsdt) minus outgoing shortwave (rsut) minus outgoing longwave radiation (rlut); clear-sky anomalies are calculated as rsdt minus

clear-sky outgoing shortwave (rsutcs) minus clear-sky outgoing longwave radiation (rlutcs). For each model the shading illustrates the ensemble envelope and the line the ensemble mean. Positive anomalies indicate increased downward flux. Squares at the bottom indicate when one model output is significantly different ($p<0.05$) from the ensemble members of all the other models according to the Mann-Whitney U test.

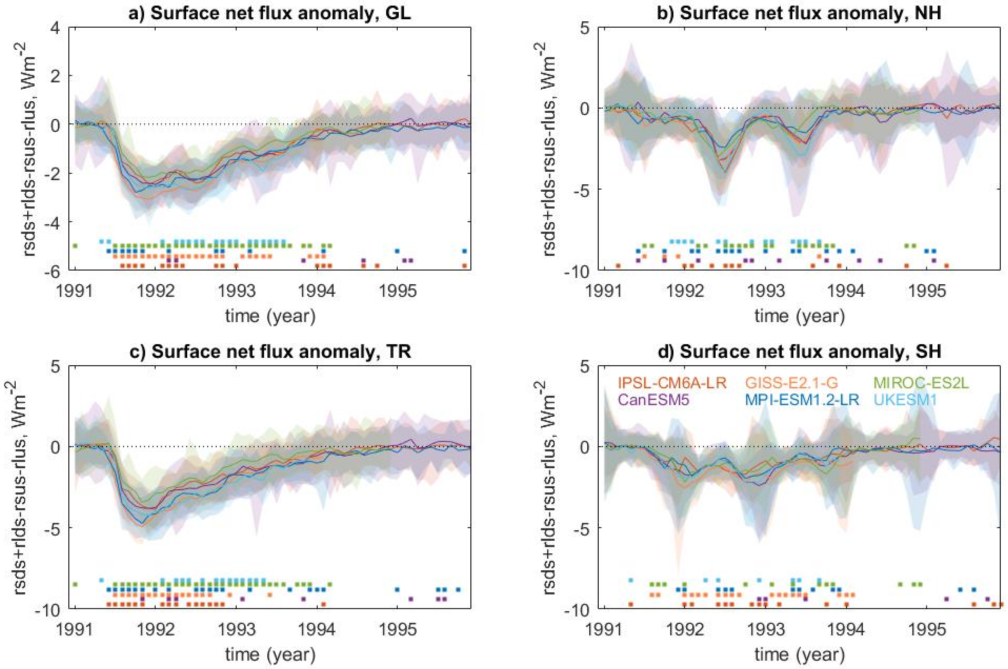
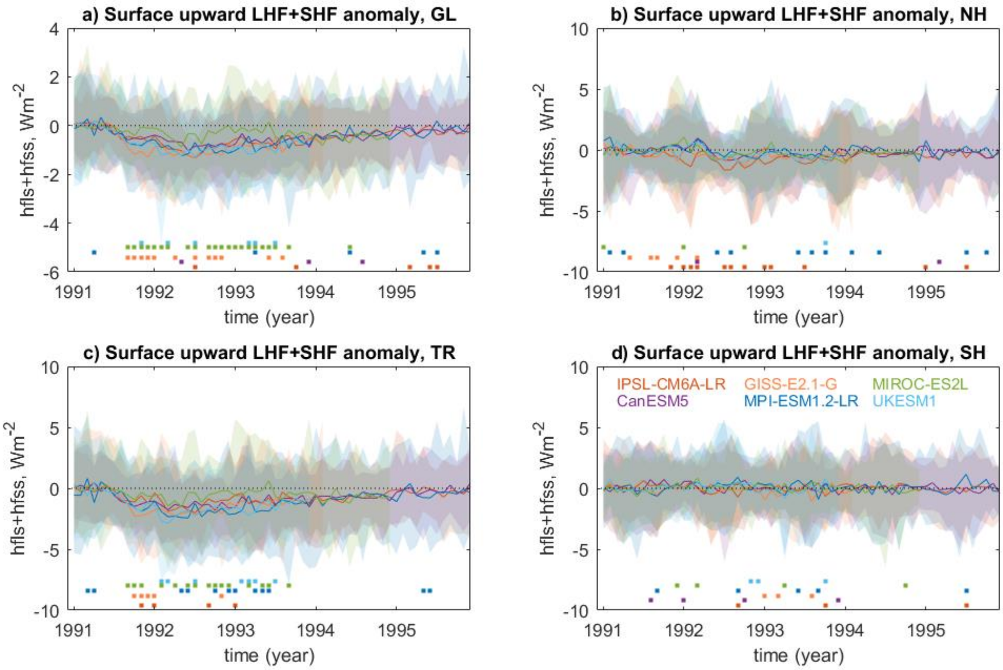


Figure 5: Surface net vertical radiative flux anomalies in the volc-pinatubo-full multi-model ensemble for a) global (GL), b) Northern Hemisphere extratropical (NH) , c) tropical (TR) and d) Southern Hemisphere extratropical (SH) mean. The net flux anomalies are calculated as downward shortwave (rsds) plus downward longwave (rlds) minus upward shortwave (rsus) minus upward longwave radiation(rlus). For each model the shading illustrates the ensemble envelope and the line the ensemble mean. Positive anomalies indicate increased downward flux. Squares at the bottom indicate when one model output is significantly different ($p<0.05$) from the ensemble members of all the other models according to the Mann-Whitney U test.

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021



955

Figure 6: As in Fig. 5 but for surface upward vertical latent heat flux (LHF) plus sensible heat flux (SHF) (or hfls+hfss).

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

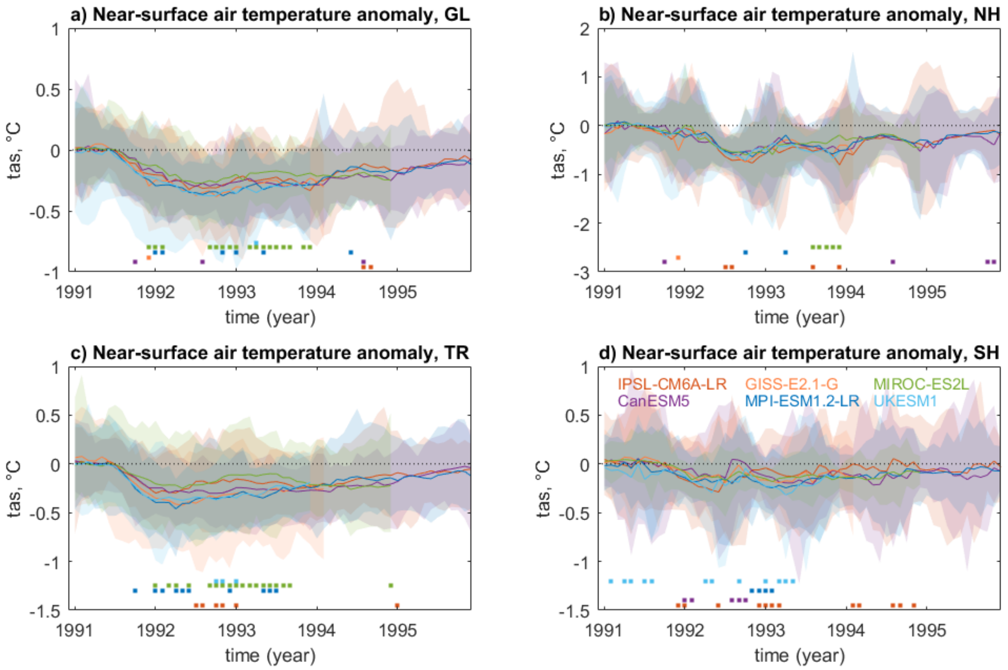
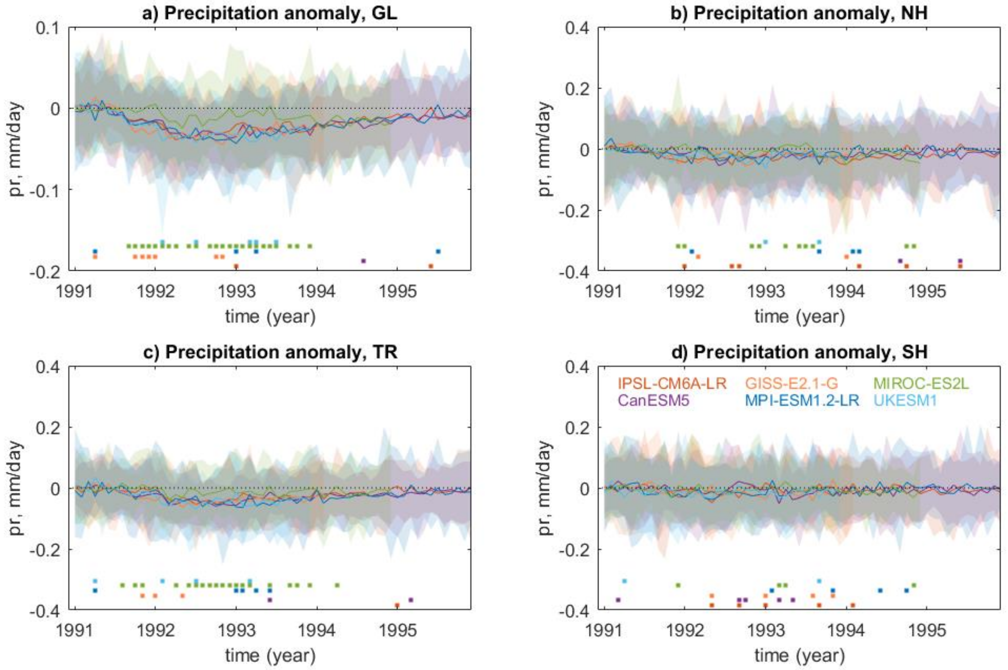


Figure 7: As in Fig. 5 but for near-surface air temperature (tas).



960 **Figure 8: As in Fig. 5 but for anomalies of total precipitation (pr).**

https://doi.org/10.5194/gmd-2021-372
Preprint. Discussion started: 12 November 2021

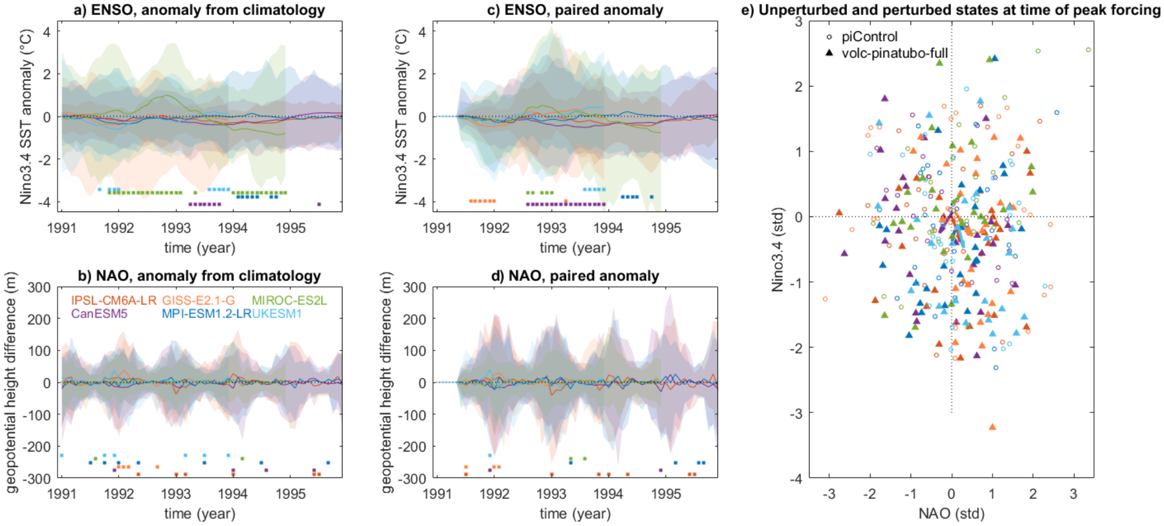


Figure 9: ENSO and NAO anomalies in the volc-pinatubo-full multi-model ensemble. (a-d)Anomalies for the indices calculated according to the VolMIP protocol as deviations from the unperturbed climatology (a,b) and as paired anomalies (c,d). For each model the shading illustrates the ensemble envelope and the line indicates the ensemble mean. Squares at the bottom indicate when one model output is significantly different ($p<0.05$) from the others according to the Mann-Whitney U test. (e) scatterplots of standardized indices for the winter (DJF) season at the time of peak forcing in the volc-pinatubo-full. Shown are also corresponding values under unperturbed conditions following the VolMIP sampling protocol (see methods and Figure 2). Arrows indicate ensemble mean changes between unperturbed and volcanically-perturbed anomalies.

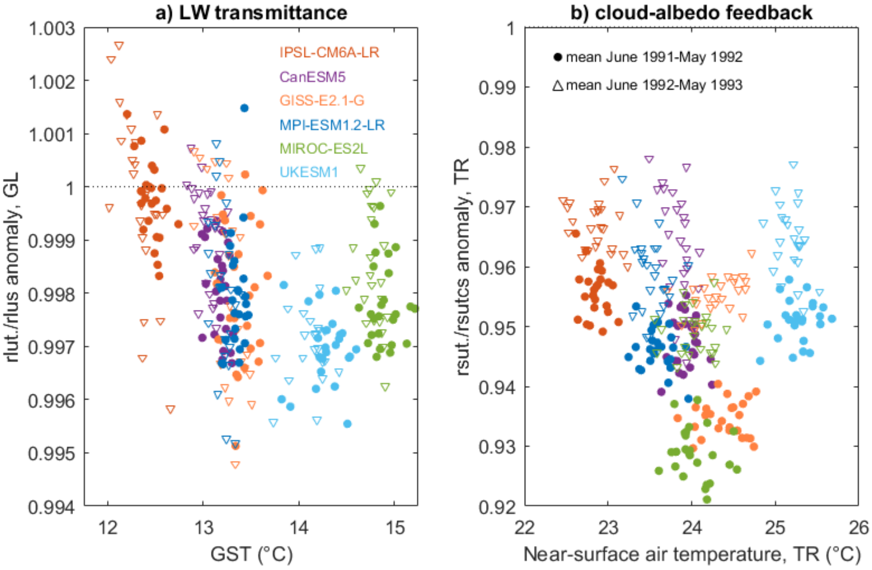


Figure 10: Response of the atmospheric LW transmittance and of the cloud-albedo feedback. a) scatterplot of the ratio of the global-average LWt/LWs↑ (or rlut/rlus) versus the post-eruption global-mean surface temperature (GST). b) scatterplot of the ratio of SWt/SWtcs for the tropical region (rsut/rsutcs) versus the tropical-mean surface temperature. Analysis is based on averages for the periods from June 1991 to May 1992 (filled circles) and from June 1992 to May 1993 (empty triangles).

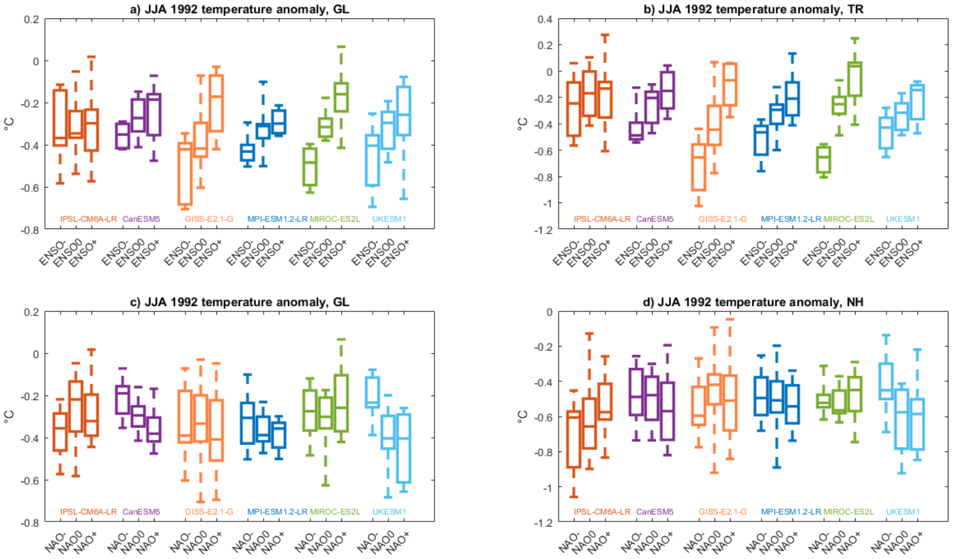


Figure 11: Box-whisker plots of near-surface air temperature anomalies of selected areas for the second post-eruption boreal summer (JJA 1992) in the volc-pinatubo-full multi-model ensemble against the sampling conditions identified by the VolMIP protocol. Sampling conditions are identified by standardized winter-average (DJF) ENSO and NAO states (positive if >0.5, negative if <-0.5, neutral/zero if in between) during the first post-eruption winter under unperturbed conditions (see Sect. 2.2). These states are accordingly used to cluster the temperature anomalies.

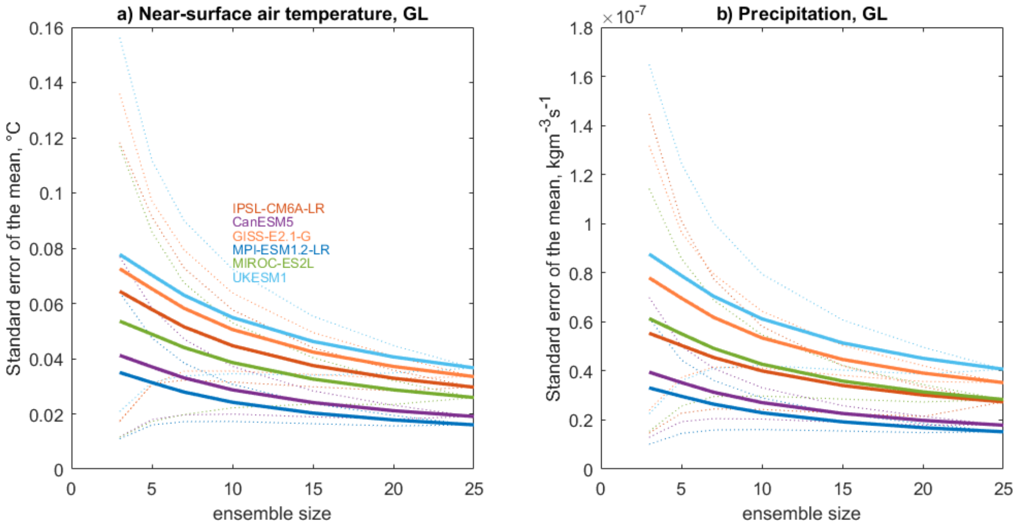


Figure 12: Ensemble size and spread. Standard error of the mean of anomalies calculated for different ensemble sizes across the different models for two key surface variables: global-mean near-surface air temperature (a) and global-mean precipitation (b). Shown are means (thick line) and 5-95th percentiles ranges (see Sect. 2.2).

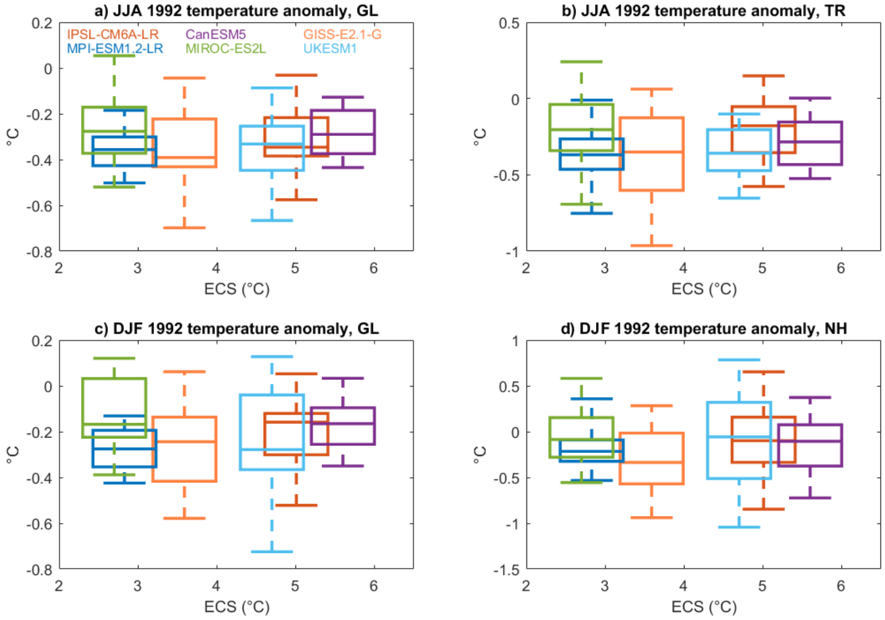


Figure 13: Box-whisker plots of simulated post-eruption near-surface air temperature anomalies in the volc-piantubo-full simulations as function of equilibrium climate sensitivity (ECS). a,b) global-mean (GL) and tropical-mean (TR) anomalies for the second post-eruption boreal summer; c,d) GL and Northern Hemisphere-mean (NH) anomalies for the second post-eruption boreal winter.