CONFESS

Consistent representation of temporal variations of boundary forcings in reanalyses and seasonal forecasts

Evaluation of the impact of improved volcanic forcings on seasonal and near-term predictions.

Roberto Bilbao, Tim Stockdale

www.confess-h2020.eu

Co-ordinated by ECMWF

CONFESS

Consistent representation of temporal variations of boundary forcings in reanalyses and seasonal forecasts

D3.2 Evaluation of impact of improved volcanic forcings on seasonal and near-term predictions, including recommendations for implementation.

Author(s): Roberto Bilbao (BSC), Tim Stockdale (ECMWF),

Dissemination Level: Public

Date:	19/12/2023
Version:	1.0
Contractual Delivery Date:	31/12/2023
Work Package/ Task:	WP3/ T3.2 & T3.3
Document Owner:	BSC
Contributors:	ECMWF
Status:	Final

ECMWF METEO FRANCE cnrm BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación ISAC

CONFESS

Consistent representation of temporal variations of boundary forcings in reanalyses and seasonal forecasts

Research and Innovation Action (RIA)
H2020- LC-SPACE-18-EO-2020 Copernicus evolution: Research activities in support of the evolution of the Copernicus services - Copernicus Climate Change Service (C3S

Project Coordinator: Dr Magdalena Alonso Balmaseda (ECMWF)
Project Start Date: 01/11/2020
Project Duration: 36 months

Published by the CONFESS Consortium

Contact:
ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom
Magdalena.Balmaseda@ecmwf.int

The CONFESS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004156.

Contents

1. Executive Summary

In WP2 several developments were carried out to improve the representation of volcanic aerosols in the IFS and implement the capability of responding to volcanic eruptions in a real-time context in both the ECMWF seasonal forecast system and the BSC decadal forecast system (Deliverable 2.3; Stockdale et al., 2023). In this deliverable sets of seasonal and decadal prediction experiments, integrating these developments, have been evaluated to assess the representation of volcanic aerosols and their impacts. The objective of this deliverable is to advance our understanding on the impact of volcanic eruptions on seasonal and decadal predictions and provide recommendations for the future model developments in the operational prediction systems.

The developments made in CONFESS have allowed for a substantial improvement in the treatment of volcanic aerosol in the ECMWF seasonal prediction system. Seasonal predictions with IFS are now able to properly represent the vertical structure of the stratospheric heating from aerosol. This is done using satellite-derived observational datasets, typically produced some time after a volcanic eruption, and with the EVA_H model in a real-time forecast context. The quality of the aerosol distribution produced by EVA_H seems generally satisfactory, given other uncertainties.

Accurate knowledge of the volcanic aerosol forcing in climate predictions is necessary to account for the climatic impacts following major volcanic eruptions. CMIP6 decadal predictions contributing to the DCPP component C (DCPP-C) show a strong agreement in predicting the radiative response to the volcanic eruptions on decadal timescales, however the atmospheric and oceanic dynamical impacts exhibit greater uncertainty. In the case of a future hypothetical eruption, decadal predictions run with either EVA and EVA_H volcanic forcings will potentially lead to more accurate decadal forecasts than without forcing, however these tools have some limitations in realistically producing the magnitude and latitudinal structure of the forcing.

2. Introduction

2.1. Background

Explosive volcanic eruptions affect climate by injecting large quantities of sulphur dioxide (as well as other gases like water vapour, CO_2 and dust) into the stratosphere, where it oxidises to form sulphate aerosols. The presence of sulphate aerosols in the stratosphere has two main effects: (1) reflects part of the incoming solar radiation, causing a negative radiative forcing that cools the Earth's surface, an effect that may last for several years (until the aerosols return to the surface) and (2) absorb infrared radiation and block the outgoing longwave radiation which may lead to a local warming of the stratosphere (Robock, 2000). These temperature adjustments may subsequently lead to other climate impacts on seasonal-to-decadal timescales (see Marshall et al. (2022), for a review), such as atmospheric and oceanic dynamical changes, which may modulate climate variability and are potentially predictable (Hermanson et al., 2022).

In WP2 several developments were carried out to improve the representation of volcanic aerosols in the IFS and implement the capability of responding to volcanic eruptions in a real-time context in both the ECMWF seasonal forecast system and the BSC decadal forecast system (Deliverable 2.3; Stockdale et al., 2023). For this deliverable sets of seasonal and decadal prediction experiments, integrating these developments, have been run and evaluated to assess the representation of volcanic aerosols and their impacts. The goal of the WP3 is to further our understanding regarding the impact of volcanic eruptions on seasonal and decadal predictions and provide recommendations for the future model developments in the operational prediction systems.

2.1.1. Objectives of this deliverable

Objective of this deliverable:

- To assess the impact of the volcanic aerosol forcing developments in the IFS carried out in WP2 (Deliverable 2.3; Stockdale et al., 2023) in two sets of seasonal prediction experiments.
- Analyse the climatic impacts of the recent large volcanic eruptions on decadal climate predictions with a purposefully designed set of simulations from six CMIP6 decadal prediction systems.
- Evaluate and compare the climate response to the volcanic forcings generated with EVA_H for the recent large volcanic eruptions with the forcings from EVA and CMIP6 using EC-Earth3 decadal hindcasts.
- Guide future developments in volcanic aerosols forcings for the next generation of operational seasonal and decadal prediction systems.

2.1.2. Work performed in this deliverable

For this deliverable sets of seasonal and decadal prediction experiments have been run to assess the representation of volcanic aerosols and their impacts.

ECMWF: A full set of coupled forecasts covering the period 1981-2020 have been run, using various representations of volcanic forcing, including a fixed low-level stratospheric background

corresponding to no volcanic forcing. To assess the impact of volcanic forcing on tropospheric dynamics, two additional case studies were run with the full operational resolution (35km) and a very large ensemble size (101 members).

BSC: We have analysed of the climate response following the eruptions of Mount Agung (1963), El Chichón (1982) and Mount Pinatubo (1991) using a multi-model set of decadal predictions, contributing to the CMIP6 Decadal Climate Prediction Project (DCPP Boer et al., 2016), which follow a purposefully designed experimental protocol (C). We have also ran and analysed sets of decadal predictions, following the DCPP protocol, but with the volcanic aerosol forcings produced with EVA and EVA_H (simple models that predict the stratospheric aerosol forcing evolution), to evaluate the expected uncertainty of using these tools in real time forecasts in the case of a future large eruption.

2.1.3. Deviations and counter-measures

It was hoped to run the ECMWF volcanic integrations for WP3 using Cy49r1 of the IFS, which incorporates the code for the new time-varying tropospheric aerosol climatology developed earlier in CONFESS. Although this cycle was supposed to be formally released at the end of summer 2023, it was delayed at a late stage to the end of the year. The WP3 volcanic aerosol integrations were initially put on hold waiting for the cycle to become available, but once it was clear that this would not happen in time for the Deliverable to be met, the plan was changed to use the existing CY48R1 instead. This means that the performance of the new volcanic aerosol is tested independently of the tropospheric aerosol changes instead of in combination with them. This makes no difference to our ability to give a scientific assessment of the results, and the final combination of new tropospheric and volcanic aerosols for use in ERA6 and SEAS6 will be made after the end of the CONFESS project.

3. Impact of volcanic aerosols on the ECMWF seasonal prediction system.

For the last 30 years, volcanic aerosol loadings in the stratosphere have been relatively low, and therefore have had limited impact on the climate system. However, this state will not persist indefinitely, and operational seasonal forecast systems need to be ready to account for the impact of volcanic eruptions as and when they occur. Based on historical experience, there are three main impacts that might be expected: warming in the stratosphere and a response of stratospheric winds; widespread surface cooling, for both SST and land areas; and in many cases a dynamical response affecting weather patterns in the troposphere, most notably in northern hemisphere winter.

3.1. Re-calculation of optical properties

In the original implementation of GloSSACv2.2 (NASA/LARC/ASDC, 2022) and EVA_H (Aubry et al, 2020) in the IFS, documented in CONFESS Deliverable D2.3, the optical properties for stratospheric sulphate aerosol were left unchanged from their original settings. One of the conclusions of Deliverable 2.3 was that the varying results seen in the IFS and EC-Earth models when using specified aerosol forcings might well be related to differences in optical properties, so it was decided to review the properties used by IFS before proceeding further. The IFS aerosol-radiation scheme relies on each aerosol type being assigned a specific lognormal size distribution and a specific complex refractive index. From this, separate software uses Mie calculations to calculate the optical properties (extinction, single scattering albedo and asymmetry factor) for each of the large number of wavelengths used by the IFS radiation scheme. The resulting optical property data for the stratospheric volcanic aerosol type is then combined with equivalent data for many other aerosol species into a single netcdf file read by the IFS at run-time. This approach allows accurate computation of the radiative effects of each specified aerosol species, but does imply that the particle size distribution is fixed and must be decided in advance.

The existing optical properties used by the IFS were found to have been calculated using assumptions more appropriate for tropospheric sulphate aerosol, namely ammonium sulphate aerosol with a median radius r0 = 0.0355 microns and a geometric standard deviation (sd) of 2.0, implying an effective radius of 0.12 microns. In contrast, stratospheric volcanic aerosol is made of droplets of concentrated sulphuric acid, with a typical concentration of 75%. Further, in-situ measurements suggest a typical sd of 1.20 to 1.25, i.e. a substantially narrower range of sizes for aerosol measured at a particular location. The effective radius of stratospheric aerosol is, however, and inconveniently for the IFS, far from constant. Since we can choose only a single size distribution, we want to choose one that gives the best and most reasonable approximation to the radiative impact across the range of particle sizes which are found at times of significant radiative impact. To do this, we need to understand how the radiative impact of volcanic aerosols depends on the size distribution. The key tools we use to help us do this are the Aerosol Refractive Index Archive from the University of Oxford Earth Observation Data Group (EODG-ARIA, https://eodg.atm.ox.ac.uk/ARIA/), and the ecaeropt code developed at ECMWF.

3.1.1. Complex refractive index of sulphuric acid

The first stage is to establish the complex refractive index of concentrated sulphuric at representative temperatures and concentrations. The concentration is not exactly fixed, in that it will vary according to conditions and be different over time and for different eruptions. We assume, though, that large eruptions are what matter most and that in this scenario water vapour will be limited and the concentration will be high. We choose a value of 75% (see Lacis, 2015). Ideally temperatures should be for lower- to mid-stratospheric temperatures, so 210-230K. Data from EODG-ARIA suggests that there is only weak temperature dependence, while the impact of concentration is somewhat higher. From the available datasets, we have chosen to use Lund Myhre et al (2003), for a concentration of 76% and a temperature of 213K. We note, though, disagreement between different datasets that cannot be explained by differences in concentration and temperature. Our preferred dataset is the most recent, and we hope the most reliable over the infrared. However, the only dataset covering the visible range is by Palmer and Williams (1975). Judged against all the more recent datasets, this appears to be less accurate in the infrared, but importantly has an imaginary refractive index which tends to zero at shorter wavelengths, matching the known fact that pure sulphuric acid is largely transparent at visible wavelengths (see Figure 1).

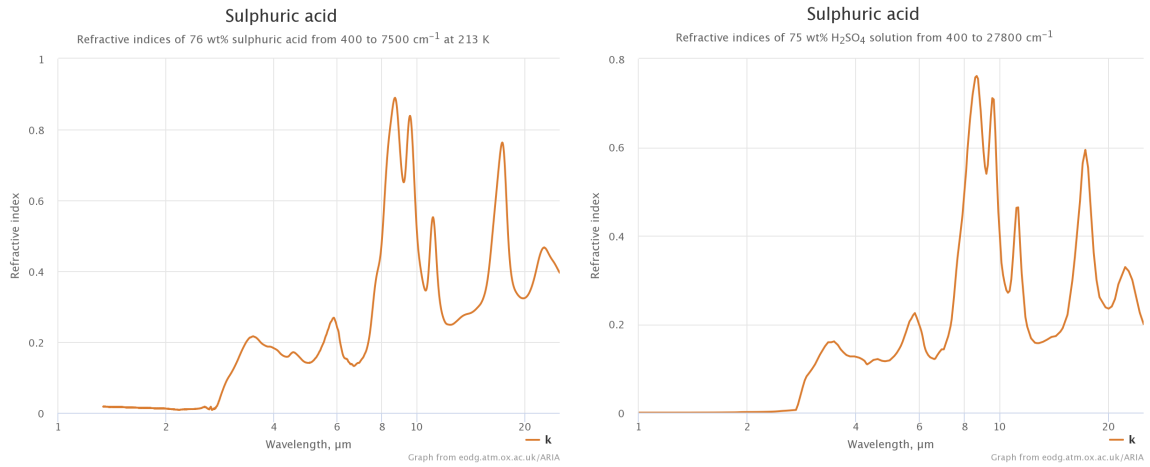

Figure 1: Imaginary part of refractive index of concentrated sulphuric acid as a function of wavelength, according to Lund Myhre et al 2013 (left) and Palmer and Williams 1975 (right). We use a combination of these two datasets, merging at a wavelength of 3 microns. Note different scales. Plots generated at eodg.atm.ox.ac.uk/ARIA.

The Lund Myhre dataset has significantly non-zero values at its smallest wavelengths of around 1.5 microns, and if not changed this would be extrapolated into the visible and lead to large solar absorption, which would be badly wrong. Thus we choose to combine Lund Myhre (above 3 microns) with Palmer and Williams (below 3 microns) to make our preferred dataset - at 3 microns itself, agreement between the datasets is very good. Michael Radke from John Hopkins has been working on new spectroscopic measurements of sulphuric acid, but no data from these experiments has been published. His measurements (personal communication) strongly support Palmer and Williams in the near-infrared and show that Lund Myhre becomes very inaccurate below 2.7 microns. Our combination of the two datasets, with a transition at 3 microns, is thus well supported by the latest data in the near-infrared.

3.1.2. Best fixed choice of volcanic aerosol particle size distribution

The IFS radiation scheme takes as input the mass density of aerosol, expressed as kg/kg which can easily be converted to kg/m3 using the local density of air. Aerosol both scatters and absorbs radiation, and despite the complexity of Mie calculations when the wavelength is comparable to the particle size, absorption is governed by the imaginary component of the refractive index. For sulphuric acid, this is essentially zero below 2 microns. Most absorption will be of terrestrial longwave radiation in the 6-20 micron window, and thus be in the regime where the wavelength is much larger than the particle size. In this case **the absorption depends only on the mass and not the particle size distribution**. However, the mass density is not directly observed, nor provided by EVA_H. Instead we are given the extinction at 525 nm, from which we must estimate the mass density supplied as input to the radiation scheme. Since aerosol particles might be in the range 0.1 to 1 micron, the extinction at this wavelength depends strongly on the size distribution. Thus we need to choose a size distribution which will give realistic estimates of extinction at 525 nm; what happens at other wavelengths will not affect the calculated warming of the stratosphere.

On the other hand, volcanic aerosol also scatters sunlight to space, cooling the surface of the planet. This scattering will depend on the particle size distribution across the visible wavelengths. To estimate it, we can take an assumed size distribution, and calculate the optical properties of the aerosol across all wavelengths in the solar spectrum, roughly 0.25 to 2.5 microns. What matters is not the extinction at these wavelengths, but the backscatter to space, which depends on both single scattering albedo and asymmetry factor. The asymmetry factor is the mean of the cosine of the scattering angle. Isotropic scattering, which will reflect much light back to space, has a zero mean, and is true of particles which are very small compared to the wavelength of light. Particles which are large compared to wavelength scatter forward very strongly, and have g ~ 0.85. The relationship between g and upward scattering is discussed in Moosmuller and Ogden (2017), and a rather rough approximation is given by b=(1-g)/2, given by Sagan and Pollack (1967). Note that the IFS two-stream radiation calculations use different approximations so may not match this exactly.

We calculated the optical properties for many different possible size distributions, to understand the sensitivity of the expected stratospheric heating and surface cooling to the possible choices that could be made. We also considered the observational evidence for particle sizes and the constraint from the ratio of satellite-observed 525/1020 nm extinction. For realistic values (r_eff in the range 0.1 to 0.5 microns), extinction at 525 nm is in the range 4000-5000 per kg/m2 if we assume a relatively narrow width for the size distribution. A wider size distribution is more consistent, if thought of in a vertically integrated sense and accounting for the variation of particle sizes with height. This results in a slightly flatter and lower extinction curve, and reduced sensitivity of the results to r_eff. When it comes to assessing the impact of particle sizes on backscatter and hence surface cooling, there is a partial compensation between smaller particles being less effective scatterers in the infrared, but better at backscattering, such that backscatter has relatively low sensitivity to particle size in the relevant range. This compensation, and a relatively flat extinction profile for our preferred width (1.50) and range of r_eff, mean that the radiative impact of aerosol is quite robust - the values we choose have reasonable validity despite the wide variation of particle sizes observed in reality.

We ran a small number of calibration experiments for the case of Pinatubo. To get the best fit to stratospheric heating, we need 525 nm to be on one side or the other of the peak in extinction. This means an r_eff of 0.15 to 0.2, or 0.45 to 0.50 microns. Given the impact on backward scattering as well as stratospheric heating, and the need to represent modest eruption events as well as large ones, we choose to take the lower value of r_eff=0.15 microns. This gives a robust and reasonably accurate estimation of optical properties in most conditions that matter. The exception is just after a large eruption, when a very large number of small particles first form and particles of ash may also be present.

We checked the impact of the revised optical parameters on the simulation of the SST response to Pinatubo. It is hard to assess what global SST should be, because there was significant ENSO variability in 1991/1992 which was not well forecast by the coupled model, presumably due to the limited quality of the ocean initial conditions (this was before the TAO array was implemented). Instead we prefer to concentrate on NH summer SSTs, which due to the shallow summer mixed layer are quite sensitive to the level of incoming solar radiation, and show marked cooling in the summers 1992 and 1993. The new optical properties result in a reduction in surface cooling (not shown), which is beneficial in 1992, though probably less so in 1993.

Details of the revised optical properties are presented in Table 1.

Parameter	Old	New
r0	0.0355 microns	0.10 microns
r_eff	0.12 microns	0.15 microns
geometric standard deviation	2.0	1.5
Assumed composition	(NH4)2SO4	H2SO4 (75%)
Optics model	GACP	PW1975_LM2003
Extinction 525 nm	4678 m2/kg	3165 m2/kg

Table 1: A comparison between old and new optical properties used in the IFS for stratospheric volcanic aerosol.

3.2. Volcanic aerosol impact on seasonal forecasts covering 40 years

An extensive set of experiments was run using the low resolution version of the IFS seasonal forecasting configuration at Cycle 48r1. These experiments run at TCo199 resolution (O200 octahedral grid, 55km) with 137 vertical levels, and were coupled with the ORCA1 configuration of NEMO. The forecasts cover the years 1981-2020, and consist of 10 member ensembles starting on the 1st November each year. In order to examine the impact on SST at longer ranges, the forecasts were run for 13 months. We made 4 sets of experiments, all identical apart from the volcanic forcing. There are two control experiments, the first using "SEAS5"-like forcing (damped persistence of initial values in a 3-box model with fixed vertical profiles) representing the approach taken before CONFESS, and the second a constant background stratospheric aerosol with a fixed AOD of 0.0045, representing a situation with no significant volcanic eruptions occurring. Two experiments with time-varying volcanic forcing use either the observed values according to GloSSACv2.2, or the output of EVA_H driven by the list of Carn et al 2016 eruptions. Note that in this latter case, the last eruption is on the 14th August 2015, and after this the amount of volcanic aerosol reduces towards a low background level. In reality there were a number of small eruptions after this date and before the end of 2020, but nothing to cause any major discrepancy. The GloSSACv2.2 dataset extends up to the end of 2021, so covers the full period of these experiments. For visualisations of some of these datasets, see Figures 18 and 25 later in this report.

OLDVOLC	SEAS5-like volcanic forcing (based on simplified CMIP5 GISS data)
CLEAN	Fixed background AOD=0.0045
NEWVOLC	GloSSACv2.2 with background removed
PREDVOLC	EVA_H with Carn et al. 2016 forcing

Table 2: Definition of volcanic forcings used in low and high resolution volcanic experiments.

3.2.1. Impact on stratospheric temperature

We first consider stratospheric temperature over the whole 40 year period. Figure 2 shows the evolution of global mean temperature at 30 hPa and 20 hPa, plotted as an anomaly relative to 1981-2010. Each experiment is bias-corrected using its own forecasts for the reference period 1994-2020, chosen to ensure the bias calibration is made in a period with low volcanic aerosol loading. This ensures that the bias correction is very nearly identical in all experiments, easing interpretation of the results.

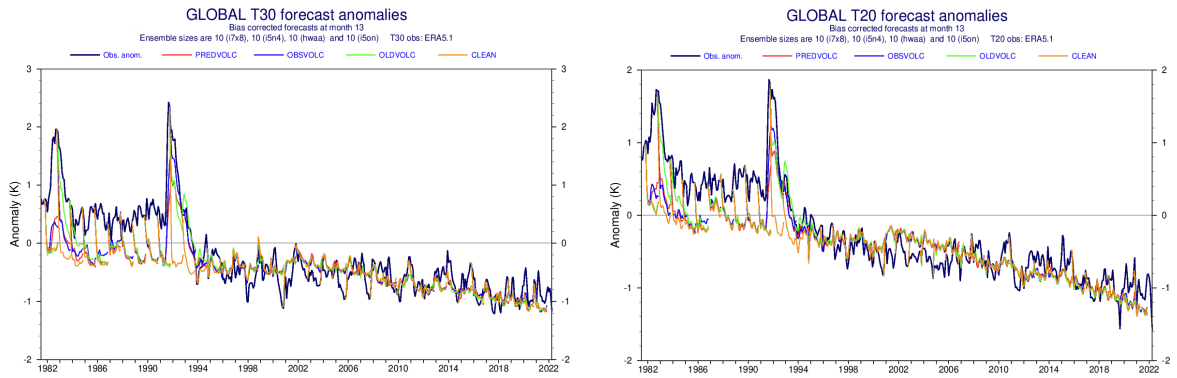

Figure 2: Evolution of global mean temperature at 30 hPa (left) and 20 hPa (right) from late 1981 to end of 2021, comparing ERA5 (blue-black) and the ensemble mean of PREDVOLC (red), NEWVOLC (blue), OLDVOLC (green) and CLEAN (orange) forecasts.

The big picture is the same in all experiments. There is an observed cooling over the whole period, which is stronger at higher altitudes - this is driven by increased CO2 concentrations leading to radiative cooling of the mid-upper stratosphere. In the post-Pinatubo era, there is a hint that the model cooling might be stronger than that in ERA5, and there are also some discontinuities in the downward trend visible in 1986 and 2001. These coincide with large jumps in stratospheric humidity in the ERA5 reanalysis used as initial conditions, in the case of 1986 due to a switch in production streams which gave a huge artificial jump in humidity. ERA5 does not analyse humidity in the stratosphere, and thus values are prone to error. It is also worth noting that the versions of the IFS used both by ERA5 and the Cy48r1 experiments here use a methane oxidation scheme which adds water vapour to the stratosphere in a way that does not change over time. A new version of the scheme will become available in Cy49r1 which accounts for the substantial increase in tropospheric methane concentrations over time, and this will lead to increasing stratospheric moisture over time. Given the sensitivities noticed in these experiments, this might lead to a slightly reduced cooling trend at the 20 hPa level, which would be helpful.

Unlike humidity, variations in O3 initial conditions are relatively modest and are mostly damped away in the forecast, except for the expected response to volcanically-induced temperature perturbations. If ozone was in reality driving large differences in T, its variations must have been much larger than represented in ERA5. This brings us to the second “big picture” result - reanalysis temperatures pre-Pinatubo are above the trend line, while the model values are not. This is a large discrepancy, and is present also in the lower stratosphere (50 and 70hPa, not shown), where the trend is zero but the offset between pre-and post-Pinatubo in ERA5 remains strong. The realism of ERA5 temperature trends becomes dubious at levels of perhaps 10 hPa and certainly above, but for 20-70 hPa the reanalysis temperatures are well constrained by in-situ radiosonde data for this period and are considered reliable. The radiosonde data used to anchor the analysed temperature does have time-varying bias corrections applied, and it is possible that a small part of the relative warming in the early period is due to imperfect bias correction, which is at its largest in these years.

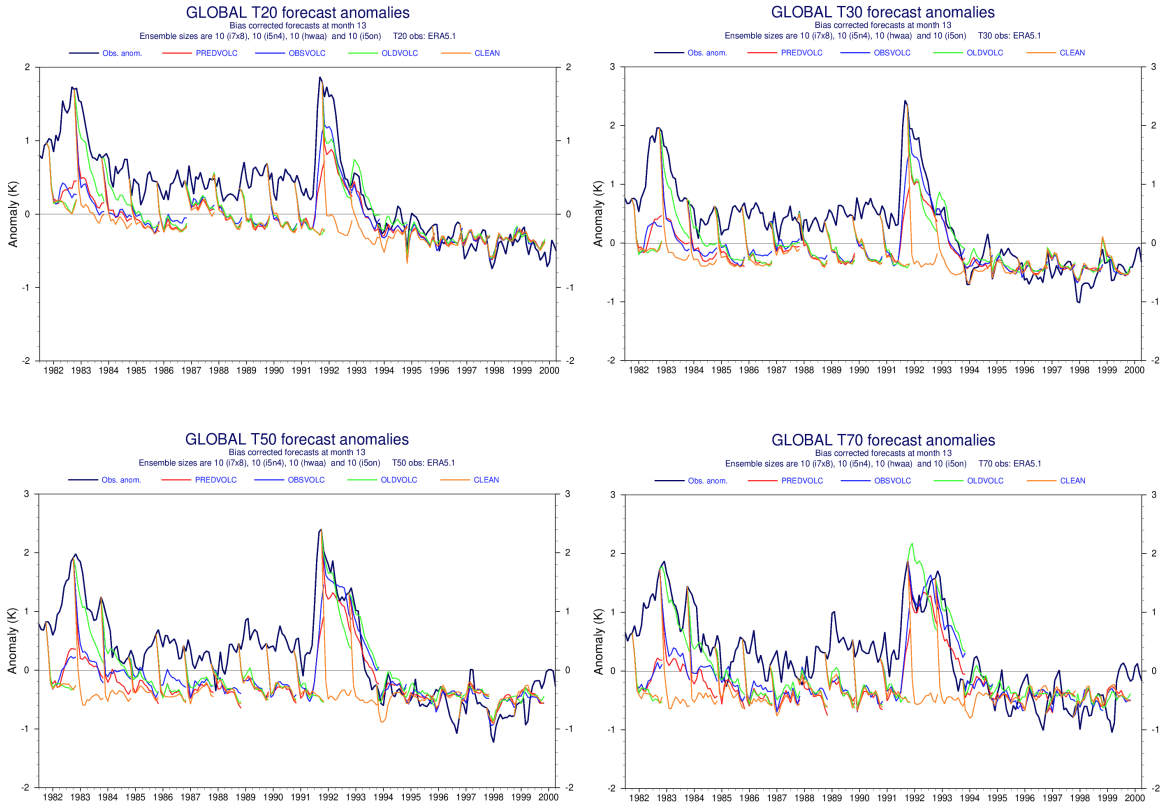

Figure 3: Evolution of global mean temperature at 20 hPa, 30 hPa, 50 hPa and 70 hPa for the period including the El Chichón and Pinatubo eruptions, showing ERA5 (blue-black) and the ensemble mean of PREDVOLC (red), NEWVOLC (blue), OLDVOLC (green) and CLEAN (orange) forecasts.

If we look in more detail at the El Chichón and Pinatubo eruptions (Figure 3), various differences are apparent. The amplitude of response to Pinatubo is reasonable, but runs using the new datasets (PREDVOLC and especially NEWVOLC) have a better vertical structure than OLDVOLC, for reasons discussed in the previous Deliverable D2.3 (Stockdale et al 2023). The success of NEWVOLC is gratifying following the optical property revisions, although there is still insufficient heating at higher altitudes which may be related to an overestimate of the particle sizes in this higher-altitude, lower concentration regime. Considering El Chichón, OLDVOLC has a much bigger signal at all levels, and relative to a post-Pinatubo reference used here this is more realistic. The reason for the stronger signal is simple - the CMIP5 data used (taken from GISS) has a much stronger amplitude of El Chichón relative to Pinatubo than is given by GloSSACv2.2 or implied by Carn et al 2016. The values and ratio of peak tropical extinction from various different datasets are tabulated in Table 3, either in the form of AOD in the visible (525 or 550 nm), or emissions in kT of SO2. We also include values from a more recent emissions dataset MSVOL4 (Carn, 2022), which is considered state-of-the-art and is being proposed for use in CMIP7.

Data source	El Chichón 1982	Pinatubo 1991	Ratio
GISS/SEAS5	0.156	0.215	0.73
GloSSACv2.2	0.065	0.196	0.33
Carn et al	7000	18000	0.39
MSVOL4	7000	15000	0.47

Table 3: The relative strengths of the El Chichón and Pinatubo eruptions from different datasets, in terms of either visible AOD or SO2 emissions in kilotons.

The uncertainty over the correct baseline temperature complicates attempts to estimate what the loading should have been from the El Chichón temperature signal. In the lower stratosphere (70 hPa up to 30 hPa), we would need to add about 0.7-0.8K to the temperatures to match the decade after the eruption. This implies that the SEAS5 forcing from GISS would be too strong at lower levels, most notably 70 hPa, but might not be so much of an overestimate higher up. Conversely, GloSSACv2 and EVA_H forcing, which give a similar magnitude temperature response, seem to be a modest underestimate at lower levels, but a substantial underestimate higher up. We note that GloSSACv2 and EVA_H derive their estimate of El Chichón in very different ways. For GloSSACv2, extinction is in principle estimated from satellite data. However, for the 1982-1984 period, no satellite data is available except at very high latitudes. The data void in the tropics was filled by taking data from a few research flights using LIDAR instruments, and thus a handful of data points had to be extrapolated to produce a multi-year tropics-wide complete dataset. As such, the data values are unusually uncertain for this period, and may well be in error. The issue has been flagged with the producers of the GloSSAC dataset, and they are reviewing all available data to see if better estimates can be produced or not. For Pinatubo, the extinction is estimated from the specified emissions, and then tuned against the relationship between emissions and extinction from the whole period, which in practice means that the data are calibrated against Pinatubo. Thus the extinction relative to Pinatubo is determined by the emissions relative to Pinatubo, which in Carn et al are relatively small (0.39). However, more recent estimates all have a lower value for the Pinatubo eruption than does Carn et al, for example MSVOL4 includes a value of 15000 kT. The fact that both emission estimates have had substantial revisions, and the GloSSAC data set has also been updated (EVA_H was tuned on v1.0, the latest available version is 2.21), means that EVA_H should be retrained using the latest data. This might change the model parameters in various ways, but in particular is likely to lead to a higher level of extinction per unit mass of emissions, so as to preserve the relatively well-observed values of extinction for Pinatubo. As long as the El Chichón estimate does not change (which it has not yet), this would increase the strength of the response to El Chichón. Thus with plausible corrections to the forcing, and allowing for the higher background temperatures, the simulated response of stratospheric temperature to El Chichón might be acceptable.

3.2.2. Impact on SST

The impact on SST and global lower tropospheric temperature can also be examined. As mentioned previously, the results are not easy to interpret because ENSO has a large impact on tropical and global temperatures, and any errors in ENSO evolution in the 13 month forecast will affect the goodness of fit. In mid-latitudes there is some dependence on unpredictable free variability in the observations of what happens. There is a further problem in the northern hemisphere that low frequency trends are not correct due to changes in sulphate aerosol loadings and consequent cloudiness changes not being represented in the model. It is also important to bear in mind that these are initialised forecast runs started at 12 month intervals. The initial conditions for all forecasts (both with and without volcanic forcing) started in the months and years after a large eruption include the cooling effects of volcanic aerosol, and it is thus the divergence between forecasts not the absolute values that is most informative. With these caveats, let us consider Figure 4.

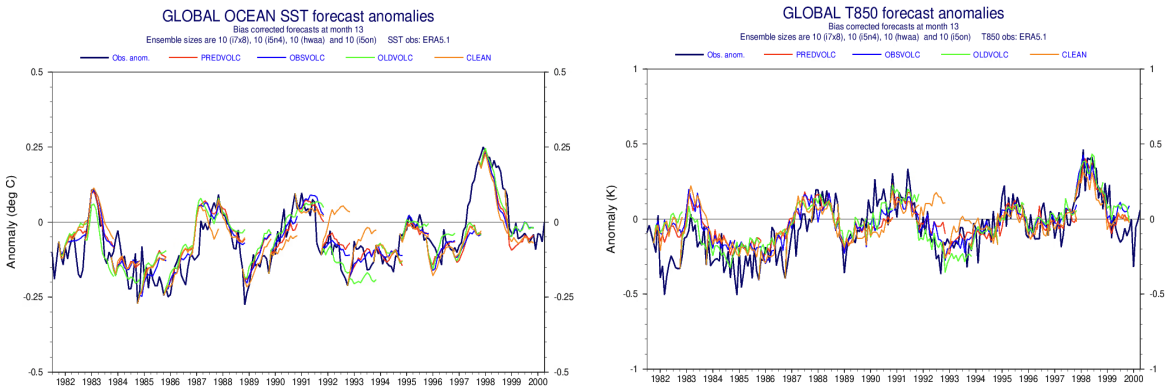

Figure 4: Evolution of global mean SST (left) and 850hPa temperature (right), showing the fit of the various volcanic forcing experiments to ERA5.

Global temperatures bear the clear imprint of ENSO variability, but we can see that the CLEAN experiment (orange) overestimates both SST and T850 in 1993 and especially 1992. The OLDVOLC experiment has the strongest surface cooling after both Pinatubo and El Chichón. For Pinatubo, the cooling is overall too strong, most clearly in 1993 when the NEWVOLC and PREDVOLC experiments have the best consistency with observations. In 1983, the year after El Chichón, OLDVOLC has too cold SST, but does a better job of T850. However, T850 in the models is running warmer than ERA5 throughout most of the 1990s, which might be more to do with other climate forcings, and the apparent agreement of OLDVOLC with ERA5 seems to be for the wrong reason - the shape and amplitude of the peak is better represented by NEWVOLC and PREDVOLC. The peak is of course due to the 1982/83 El Nino, the question is merely as to the extent to which this was offset by volcanic aerosol, and the new simulations suggest very little, compared to CLEAN. Again, this is perhaps more to do with the reduced amplitude of El Chichón relative to Pinatubo in the latest forcing datasets.

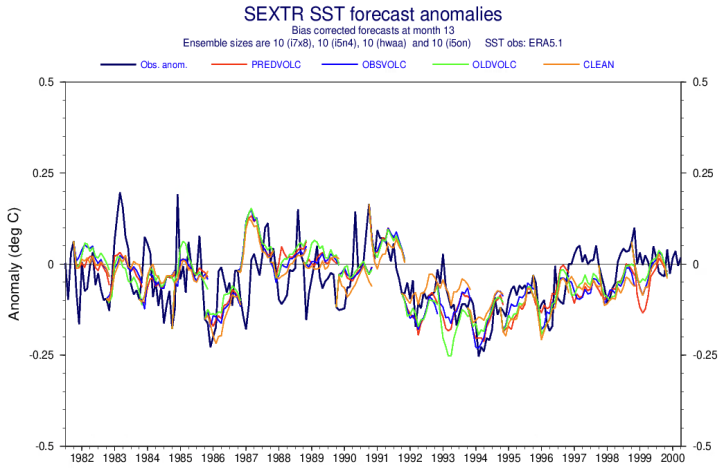

Figure 5: Evolution of southern hemisphere extratropical SSTs, which appear to be more clearly dominated by the Pinatubo eruption.

Figure 5 is perhaps the cleanest example we can give of the impact of Pinatubo on surface temperatures. The Southern Ocean SSTs are less influenced by ENSO or aerosol changes in the northern hemisphere, and we see a clear and relatively large drop of 0.2-0.25K in the months following the eruption. The forecasts from 1 November 1991 already include much of this cooling in their initial conditions, although the forecasts for 1992 and 1993 do show a substantial difference between CLEAN (which tends to warm a bit too much), OLDVOLC (which tends to strengthen the cooling too much) and PREDVOLC and NEWVOLC (which are about right).

3.3. Impact of Pinatubo volcanic aerosol in high resolution large ensembles

The impact of volcanic aerosol on the radiative balance of the stratosphere and surface does not greatly depend on model resolution, and can reasonably be assessed with moderate ensemble sizes of lower resolution runs. However, there are long-standing hypotheses of a dynamic impact of volcanic aerosol, in particular on the northern hemisphere winter circulation (Robock and Mao, 1992; Graf et al. 1993; see also Marshall et al 2022). Experience of many modelling groups (e.g. DallaSanta and Polvani, 2022) and at ECMWF has shown that the dynamic pathway (whereby stratospheric heating changes stratospheric winds, leading to a stronger stratospheric polar vortex and then downward propagation to affect the tropospheric circulation) is fickle. We thus choose to examine it using the operational resolution of the forthcoming SEAS6 (TCo319L137), and to use large ensembles (101 members) to attempt to capture even weak signals. This means that the experiments are expensive, so we restrict them to the two winters following the Pinatubo eruption, when the volcanic aerosol is at its strongest out of the last 40 years. We also consider only 5-month integrations from 1st November, covering the full winter season to the end of March but no further.

3.3.1. Dynamical impact on the stratosphere and troposphere

In the first NH winter after the Pinatubo eruption, DJF 1991/92, the volcanic aerosols have a clear dynamical impact on the stratosphere. Figure 6 shows vertical sections of zonal mean differences

between NEWVOLC and CLEAN (top) and PREDVOLC and CLEAN (bottom). The ensemble mean of the seasonal mean wind anomaly at 60N, 10 hPa, often used as a measure of the strength of the stratospheric vortex, is 4 m/s stronger in NEWVOLC. This is modest compared to the observed interannual variability of this quantity, but is a surprisingly good match to other published values such as Figure 1 of Azoulay et al. 2021, which also reports a 4 m/s signal from Pinatubo based on a 100 member ensemble with specified volcanic forcing.

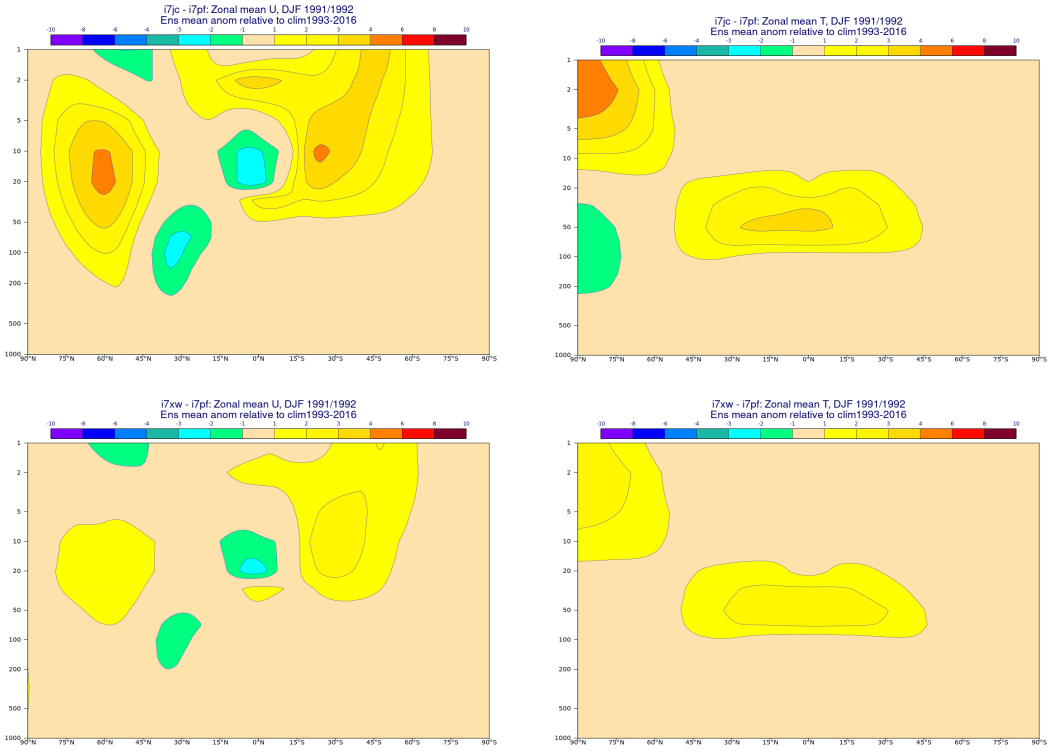

Figure 6: Zonal mean sections of the DJF 1991/92 ensemble mean differences in zonal wind (left) and temperature (right), between NEWVOLC and CLEAN (top) and PREDVOLC and CLEAN (bottom).

However, when we come to look at the results from PREDVOLC (bottom), the zonal wind signal is substantially weaker, not quite reaching 2 m/s. The tropical temperature signal is also weaker in PREDVOLC, but the reduction in the vortex signal seems to be stronger than one might have expected from the relative temperature signals. Threshold effects have been reported in the literature (e.g. Azoulay et al. 2021), but it is not clear whether this is a similar effect or a sensitivity to the structure of the volcanic forcing: the GloSSAC data used in NEWVOLC represent observed latitudinal gradients in aerosol forcing, while EVA_H uses generic functions to represent the distribution in each of its boxes. Signals in the second winter 1992/93 (not shown) are substantially weaker for both experiments.

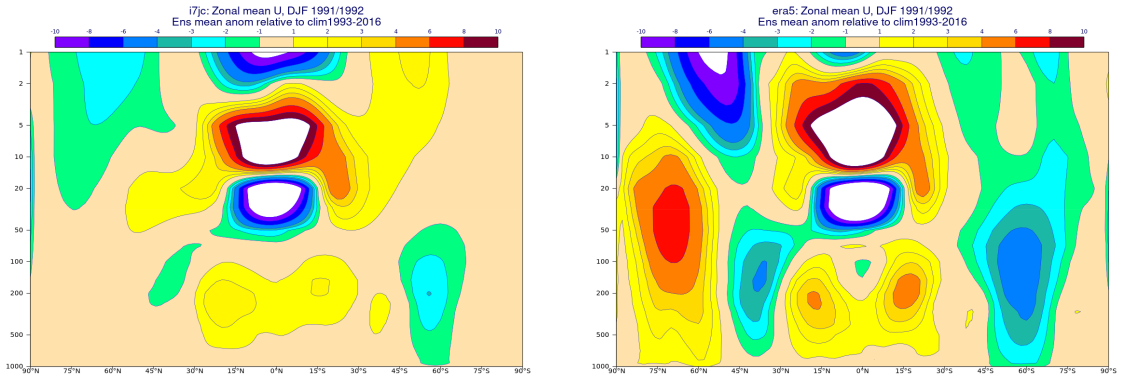

Figure 7: Zonal mean wind anomalies for DJF 1991/1992 relative to 1993-2016 from the ensemble mean of NEWVOLC (left) and ERA5 (right).

To compare with what actually happened in the winter of 1991/1992, we plot the anomalies in NEWVOLC and those from the ERA5 reanalysis (Figure 7). The QBO signal is moderately well reproduced (further improvements are expected in Cy49r2, to be used by SEAS6), and tropospheric anomalies in the tropics and southern hemisphere are also present in the ensemble mean, albeit too weakly. However, although the volcanic aerosol gave a positive wind anomaly at 60N relative to a clean stratosphere (as shown in the previous Figure 6), this was too weak to produce the large positive anomaly observed that year. We could interpret this two ways - either the observed outcome was just chance, and our model should not be expected to match it, or our model signal was essentially correct, but just too weak. The tendency of models to produce a far too weak downward coupling from the NH winter stratospheric vortex to the surface is a well known general property (e.g. Stockdale et al, 2015), that has been established over a larger observed number of cases than is possible with volcanic eruptions, and thus is plausibly part of the story.

We examine this further by looking at the spatial structure of wind anomalies at different levels descending from the stratosphere to the troposphere in Figure 8. For each level, we show on the left the anomaly in the CLEAN experiment, representing all forcings and initial experiments other than the Pinatubo aerosol. In the centre, we show the difference NEWVOLC-CLEAN, representing the impact of Pinatubo. Finally on the right we show the observed anomaly according to ERA5. We invite the reader to consider the extent to which, if we were to hypothetically scale up the signal in the centre column before adding it to the starting signal on the left, we would be able to obtain a predicted signal approaching that observed on the right.

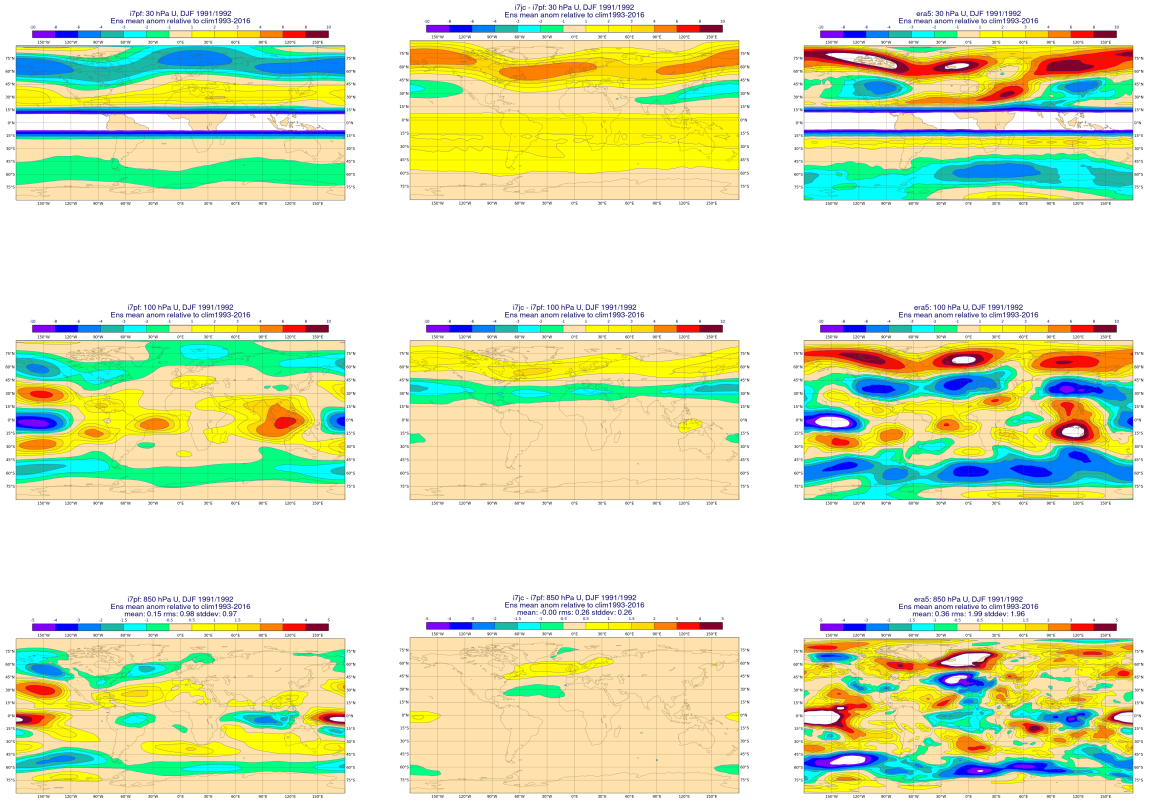

Figure 8: Zonal wind for DJF 1991/92, showing on the left the ensemble mean anomaly from CLEAN, in the centre the ensemble mean difference NEWVOLC-CLEAN, and on the right the observed anomaly from ERA5. Top row: 30 hPa, middle row: 100 hPa, bottom row 850 hPa.

The observed zonal wind anomaly in the Euro-Atlantic region at 850 hPa (bottom right of Figure 8) is in fact part of a barotropic structure that reaches up to 50 hPa - in the stratosphere, it is part of a circumpolar structure (the Northern Annular Mode), but in the troposphere it is only the Euro-Atlantic sector which is active, remaining so down to the surface. Remarkably, this is very similar to the same spatial structure which the model simulates as the response to the volcanic aerosol - a circumpolar signal in the stratosphere, reducing to an NAO-like response in the troposphere. There is admittedly a slight shift in latitude, but the big difference is that the model response is very weak, so that if we add it directly to the non-volcanic signal (left column), which is mostly related to signals originating in the tropical Pacific, the Euro Atlantic signal looks nothing like what was observed. If we were able to create a model with a stronger volcanic response, especially the downward connection to the troposphere, the model prediction might well resemble the observed winter circulation.

3.3.2. Surface temperature response

The volcanic aerosol should produce surface cooling via its direct radiative forcing, together with both warming and cooling from dynamical changes, although as we have seen from the winds, we

expect this latter response to be substantially too weak. Our 101 member ensembles enable us to measure even weak signals in the model response.

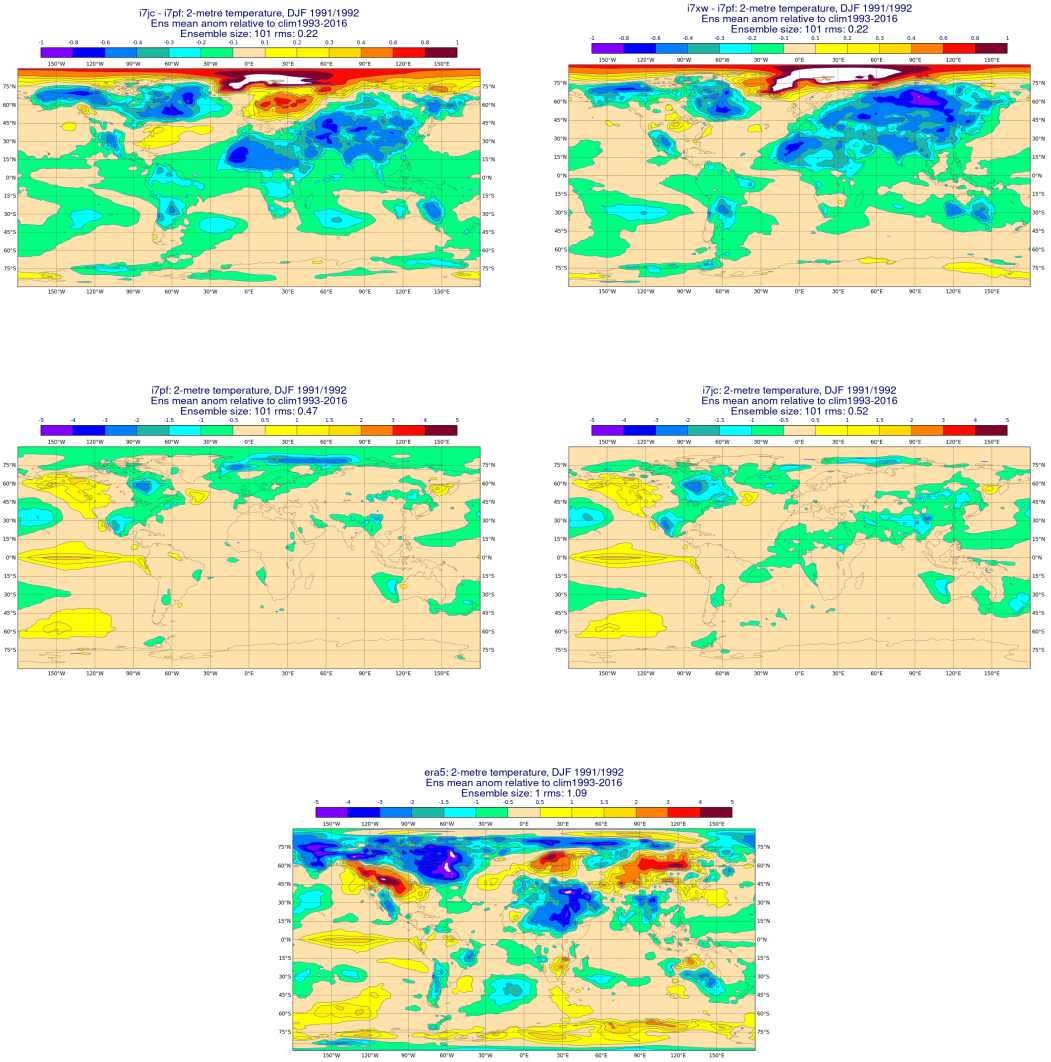

Figure 9: Two-metre temperature signals for DJF 1991/1992. Top row: ensemble mean signal in NEWVOLC (left) and PREDVOLC (right), relative to CLEAN, plotted with a reduced contour interval. Middle row: Ensemble mean anomaly in CLEAN (left) and NEWVOLC (right). Bottom: ERA5 anomaly.

Both NEWVOLC and PREDVOLC show a widespread cooling across the tropics and mid-latitudes relative to CLEAN (Figure 9 top row) as expected from the direct radiative effect of the aerosol in regions of high insolation. However there are also signals in the northern hemisphere which are related to dynamical changes in circulation, with a more definite warming over Scandinavia in PREDVOLC. When we add the signal from PREDVOLC to the anomaly already present in CLEAN (middle left), associated in part with EL Nino, we obtain a temperature which is visibly cooler over tropical areas, in better agreement overall with ERA5 (bottom row). We see that several key dynamic features in the northern hemisphere in the top row (strengthened cooling over California and

Quebec, warming over Scandinavia and cooling over the Middle East) are also present in ERA5, though generally with a much stronger intensity. Not all aspects of ERA5 anomalies are accounted for by the forecasts, notably the warm anomaly in eastern Siberia, and indeed we still expect there to be a role for unforced variability in the observed outcome. Nonetheless the model is giving both dynamic and radiative responses with plausible structures.

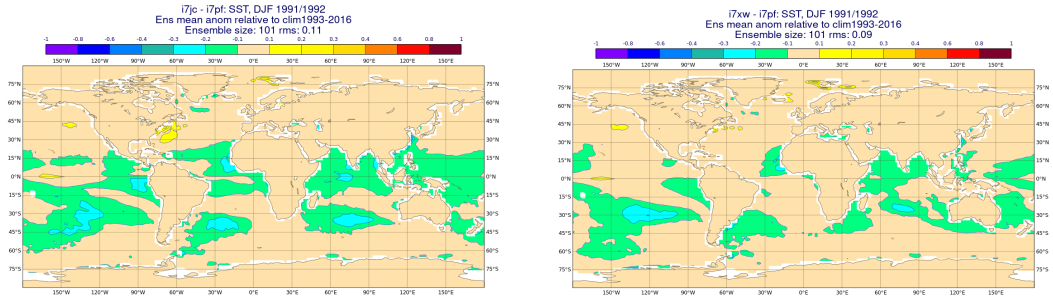

Figure 10: SST signal for DJF 1991/92 in NEWVOLC (left) and PREDVOLC (right), relative to CLEAN. This is not the total impact of Pinatubo, but the difference in SST evolution from the November initial conditions.

We finally include plots of the impact of volcanic aerosol on SST (Figure 10). As expected, this is largely in the tropics and southern mid-latitudes, where insolation is strong in the November-February period. The impact of the EVA_H predicted aerosol (right) is slightly less than that of the GloSSACv2 specified aerosol (left). We cannot relate these differences directly to the observed anomalies, because much of the SST anomaly present in both observations and model runs is implicit in the ocean initial conditions on the 1st November, although we can report that the cooler SSTs when including aerosol are more realistic (not shown). However, this plot does show the expected error in our seasonal forecast of SSTs, even at a short lead time of 2-3 months, if we were to neglect volcanic aerosol in our seasonal forecast systems.

Finally we note the intriguing detail that the amplitude of El Nino is slightly enhanced in both sets of integrations, even though the difference in radiative forcing is only present for a few months. It may possibly be related to reductions in the strength of the tropical hydrological cycle leading to fractionally weaker winds. The effect is small, and is not likely on its own to trigger an El Nino event, but it is still interesting to note that volcanic aerosol has the potential to perturb ENSO, in line with previous studies (e.g. Predybaylo et al., 2017).

4. Impact of the volcanic aerosol forcings on multi-year forecasts

Decadal climate predictions have become a major tool for forecasting the climate of the next few years out to several decades (e.g. Hermanson et al., 2022). On these timescales, part of the predictability arises from internal variability, in particular in the slowly evolving components of the climate system (e.g. the ocean). This predictability can be improved by initialising the model with the observed state to put the model in phase with observed internal variability. The other main source of predictability relates to the changes in external radiative forcings (i.e. changes in the climate system energy balance), which can be of natural (e.g. solar irradiance and volcanic aerosols) or anthropogenic (e.g. greenhouse gas concentrations, land use changes and anthropogenic aerosols) origin.

In recent decades three major tropical volcanic eruptions have occurred: Mount Agung (1963), El Chichón (1982) and Mount Pinatubo (1991). These eruptions of varying intensity (7 Tg, 8 Tg and 18 Tg of SO_2 respectively) had climate impacts on seasonal-to-decadal timescales with high predictive potential (e.g. Timmreck et al., 2016; Ménégoz et al., 2018; Hermanson et al., 2020). Explosive volcanic eruptions affect climate by injecting large quantities of sulphur dioxide (as well as other gases like water vapour, CO 2 and dust) into the stratosphere, where it oxidises to form sulphate aerosols. The presence of sulphate aerosols in the stratosphere has two main effects: (1) reflects part of the incoming solar radiation, causing a negative radiative forcing that cools the Earth's surface, an effect that may last for several years (until the aerosols return to the surface) and (2) absorb infrared radiation and block the outgoing longwave radiation which may lead to a local warming of the stratosphere (Robock, 2000). These temperature adjustments may subsequently lead to other climate impacts on seasonal-to-decadal timescales (see Marshall et al. (2022), for a review), such as atmospheric and oceanic dynamical changes, which may modulate climate variability.

Understanding the sensitivity to the volcanic forcing is particularly relevant in a real-time climate prediction context, since the Volcanic Response Plan (VolRES) following the next major eruption (a Stratosphere-Troposphere Processes and their role in Climate (SPARC) initiative, a core project within the World Climate Research Program (WCRP)) protocol consists in estimating the volcanic forcing of the future eruption using tools such as the Easy Volcanic Aerosol models (e.g. EVA, Toohey et al., 2016; EVA_H, Aubry et al., 2020).

4.1. Impact of volcanic eruptions on CMIP6 decadal predictions: a multimodel analysis.

Large volcanic eruptions can have significant climate impacts on seasonal-to-decadal timescales, some of which occur consistently across eruptions while others depend on aspects such as the magnitude, space–time structure of the volcanic aerosol concentrations, timing during the year and climate background conditions at the time of the eruption. Understanding these commonalities and particularities in the responses, and to what extent they are model-dependent, is essential to make better predictions should a new major volcanic eruption occur. The DCPP jointly with VolMIP (Zanchettin et al., 2016) designed a specific protocol to improve our understanding of the effects of volcanic aerosols upon decadal prediction, which consists in repeating three sets of retrospective

predictions initialised just before the eruptions of Agung (1963), El Chichón (1982) and Pinatubo (1991), but without the associated volcanic forcing (DCPP-C Boer et al., 2016). The impact of the volcanic eruptions is therefore determined by subtracting the hindcasts with and without the volcanic aerosols (DCPP-A - DCPP-C). In this study we have analysed and compared these prediction sets with the baseline predictions including all forcings in six CMIP6 decadal prediction systems (CanESM5, CESM1-1-CAM5-CMIP5, CMCC-CM2-SR5, EC-Earth3, IPSL-CM6A-LR and HadGEM3-GC31-MM). The fact that these simulations are decadal hindcasts which are initialised with the observed state, implies that the climate response might be more realistic (with respect to non-initialised simulations) and directly comparable to observations, as internal variability can modulate the response to the volcanic forcing. To fully exploit the decadal prediction protocol we also compare the predicted surface temperature anomalies with observations to infer the importance of including the volcanic forcing, attribute observed changes and determine to what extent the initial conditions can improve the agreement in the three hindcasts. The results discussed here are a summary of the publication submitted to Earth System Dynamics (Bilbao et al., in rev., https://esd.copernicus.org/preprints/esd-2023-36/).

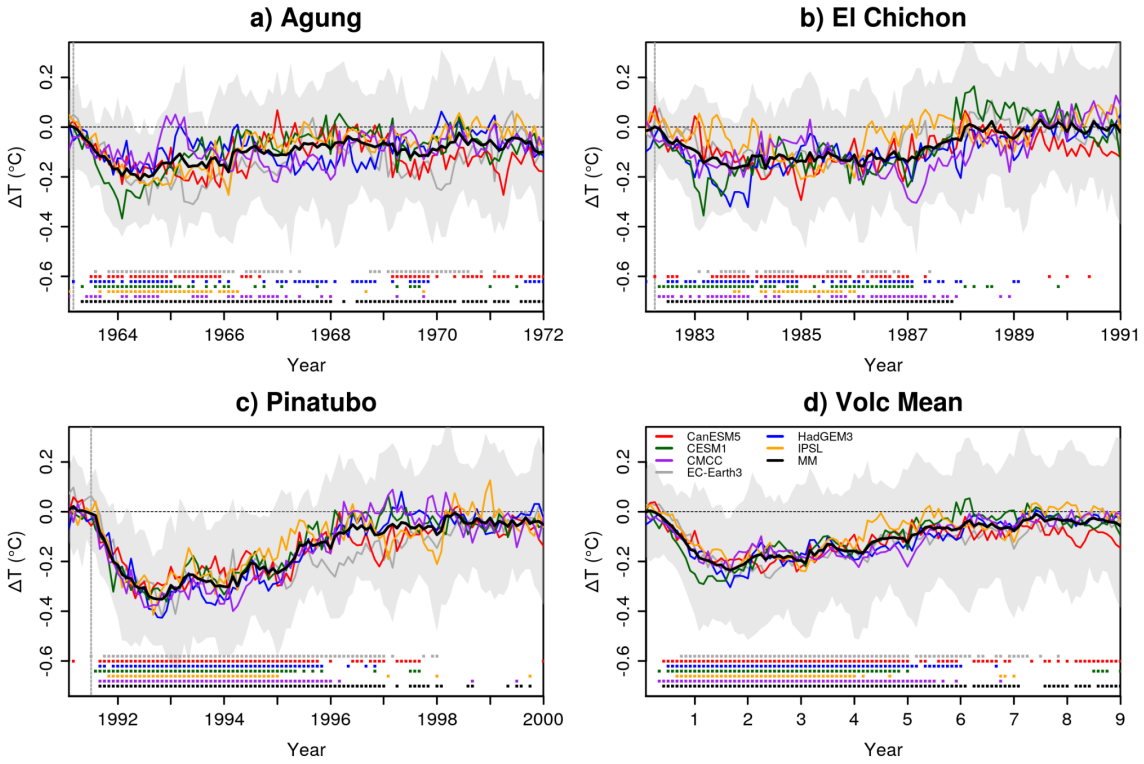

Figure 11. Global mean surface air temperature response (°C) to the volcanic eruptions (DCPP-A minus DCPP-C). The ensemble mean for each model and the multi-model mean are shown. The shading is the multi-model member spread calculated as the 10 th and 90 th percentiles of the entire ensemble. Filled squares on the bottom part of the figure indicate statistically significant differences. The vertical grey dashed lines indicate the approximate time of the eruptions.

All decadal prediction systems simulate a reduction in the global net TOA radiation fluxes, surface temperature (Figure 11) and ocean heat content in response to the volcanic eruptions, with rather small inter-model differences in terms of the ensemble mean response. The magnitude of the eruption does influence the magnitude and persistence of the signals. The geographical pattern of

the surface temperature response is also generally consistent across the models. For example, the first year following the eruptions is characterised by a cooling of the Tropics and subtropics and a warming over the Eurasian Arctic sector, although the warming is not statistically significant for all eruptions (Figure 12). In later years, cooling spreads worldwide, with the strongest anomalies being found over the Arctic, with local cooling anomalies persisting for 5 to 9 years, depending on the eruption magnitude.

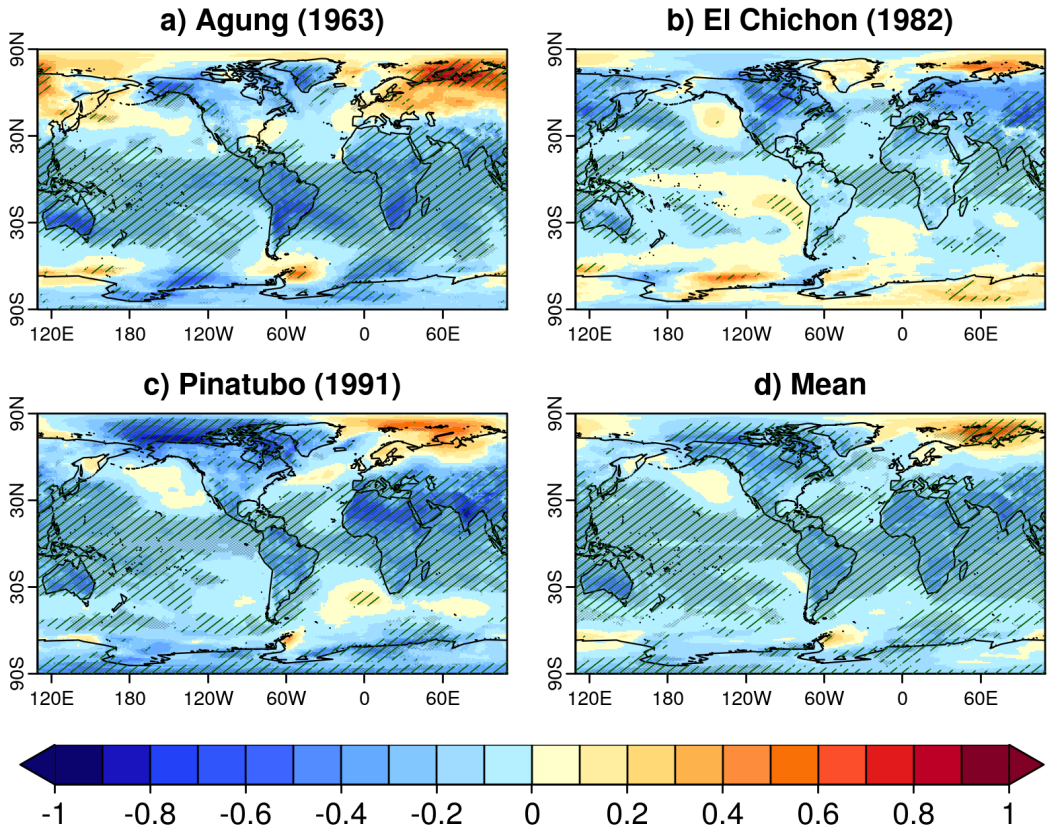

Figure 12. Model mean near-surface air temperature (°C) response (DCPP-A minus DCPP-C) during the first year following the volcanic eruptions (June-May). Hatching indicates statistically significant anomalies, while the shading indicates model agreement.

There are some differences in the predicted radiative response among the three eruptions analysed. The eruption of Pinatubo was the largest, which is reflected by simulating the strongest and most persistent anomalies in TOA radiation fluxes, surface temperature and ocean heat content. The eruptions of Agung and El Chichón are weaker and of comparable intensity, but exhibit evident differences in the geographical distribution and temporal evolution of their forcings. While the eruption of Agung mainly affected the Southern Hemisphere, the eruption of El Chichón affected the Northern Hemisphere, something that is reflected in the TOA radiation and surface temperature anomaly patterns of the response. In contrast, the eruption of Pinatubo had a more meridionally symmetric response.

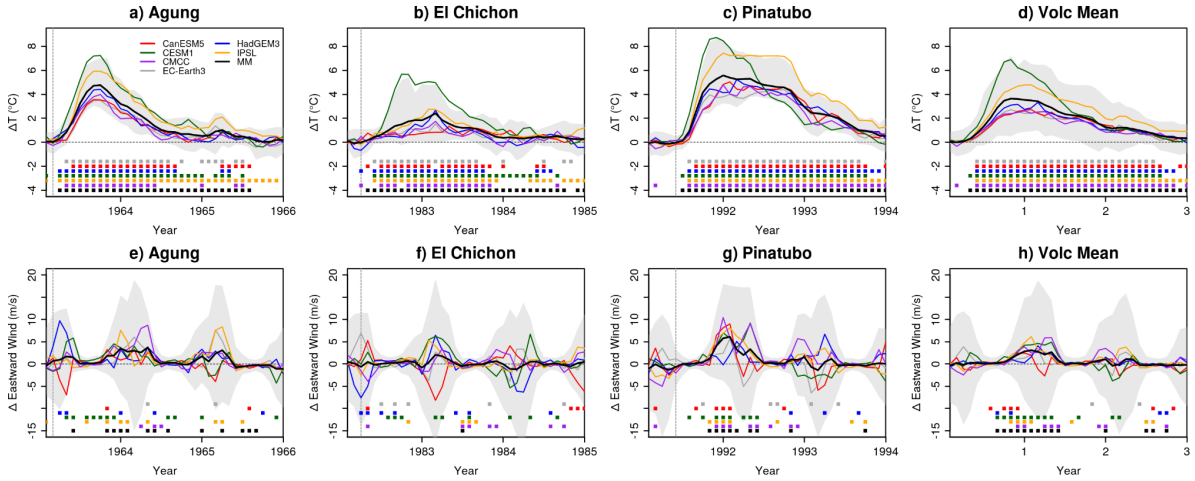

Figure 13. Stratospheric air temperature in the tropics (30 ◦ N - 30 ◦ S at 50 hPa) and polar vortex (average zonal velocity over 55 ◦ N–75 ◦ N at 50 hPa) response (DCPP-A minus DCPP-C) following the volcanic eruptions. The ensemble mean for each model and the multi-model mean are shown. The shading is the multi-model member spread calculated as the 10 th and 90 th percentiles. Filled squares on the bottom part of the figure indicate statistically significant differences.

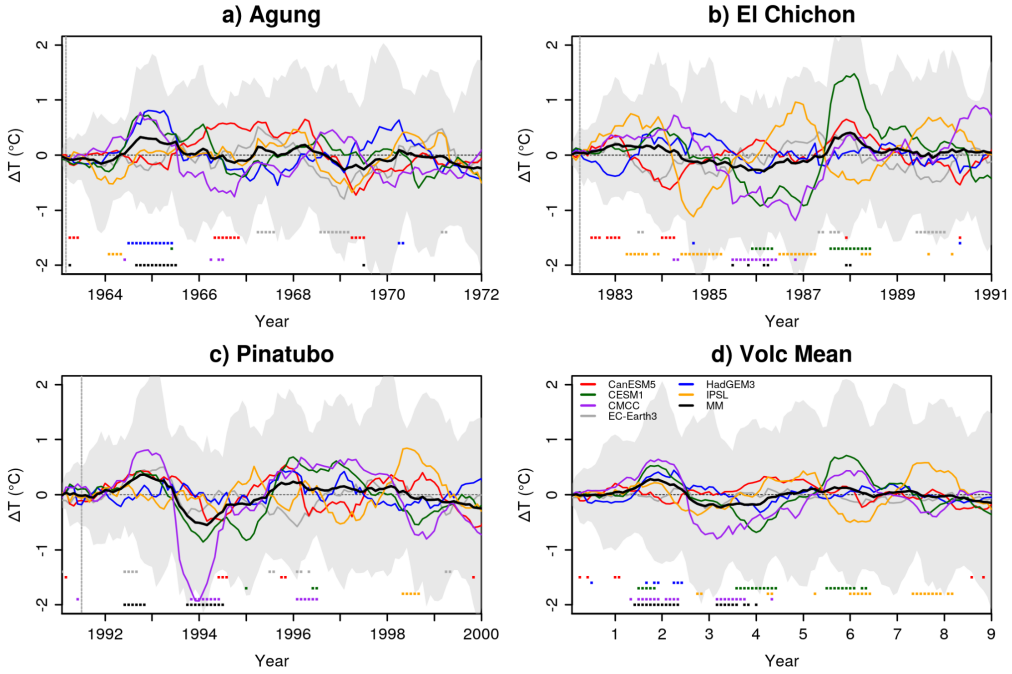

Figure 14. Relative Niño3.4 index response following the eruptions of a) Mount Agung (1963), b) El Chichón (1982), c) Mount Pinatubo (1991) and d) the mean of the three eruptions. Filled squares on the bottom part of the figure indicate statistically significant differences (see methods). The ensemble mean for each model and the multi-model mean are shown. The shading is the multi-model member spread calculated as the 10 th and 90 th percentiles. The vertical dashed lines indicate the approximate time of the eruptions.

Besides the direct radiative cooling, the volcanic eruptions also excited dynamical responses. Since these responses are more sensitive to climatic noise they require larger ensembles to be detected, so we first analysed the multi-model and multi-eruption composite response, formed by 180 members. We note that this approach is useful to increase the ensemble size but can also mask some responses

by including weaker eruptions (c.f. Bittner et al., 2016). The resulting composite response is characterised by a strong tropical warming in the lower stratosphere accompanied with a strengthening of the Northern Hemisphere polar vortex in the first winter, which resembles a positive NAO-like pattern which is, however, not statistically significant (Figure 13). The ENSO response is characterised by the development of weak El Niño-like conditions in the first year after the eruption which then transitions to weak La Niña-like conditions in the second and third years (Figure 14). In the North Atlantic Ocean we have shown that there is a significant enhancement of the mixed layer depth in the Labrador Sea during the three boreal winters following the eruptions, and a weak but significant strengthening of the AMOC during years 2-9 after the eruptions (Figure 15). We have related these responses to a reduction in density stratification in the Labrador Sea.

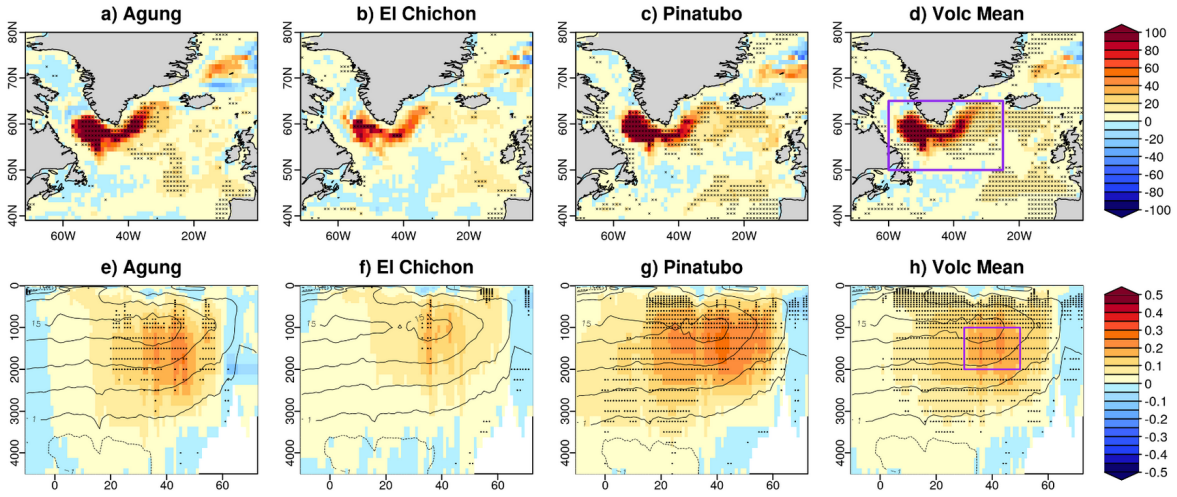

Figure 15. Multi-model and multi-eruption composites response for the mixed layer (February-March-April) depth (m) for years 1-3 and overturning stream function (Sv) years 2-9 to the volcanic eruptions. Stippling indicates statistically significant anomalies.

However, there are important differences in these dynamical responses, both across models and across eruptions. Multi-model composites for individual eruptions show that the acceleration of the Northern Hemisphere polar vortex only occurs in the eruptions of Agung and Pinatubo, while not for El Chichón. The lack of a response for El Chichón is probably related to a combination of factors, from its weak intensity, the geographical pattern of the forcing and the background climate conditions. In the case of the ENSO response, we have shown that the for the eruptions of Agung and Pinatubo, the El Niño-like state develops and peaks in the first year following the eruptions, while for the eruption of El Chichón the El Niño-like state occurs in the same year of the eruption. We have discussed that these differences are probably explained by the geographical pattern of the volcanic forcing (c.f. Pausata et al., 2020), the timing of the eruption and the ocean state (c.f. Predybaylo et al., 2020). We have also shown that there are important inter-model differences in these dynamical responses. For example, not all models simulate an acceleration of the Northern Hemisphere polar vortex. The ENSO response is also model dependent since some models show a strong response and others remain unresponsive. Similarly, for the North Atlantic Ocean we have shown that the multi-model response comes exclusively from two of the models (CMCC-CM2-SR5 and HadGEM3-GC31-MM), which show coherent changes in Labrador Sea stratification, the mixed layer depths, and the AMOCs.

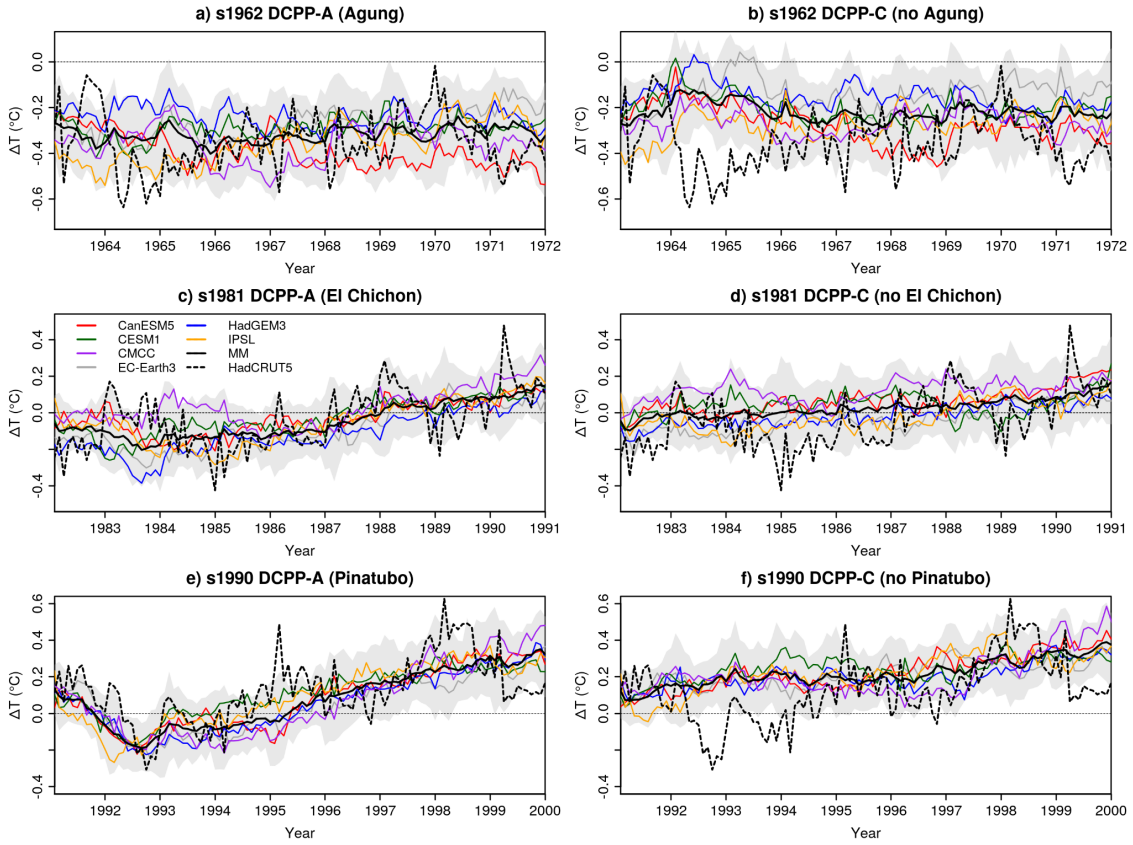

Figure 16. Monthly mean global near-surface temperature anomalies (°C) of the predictions initialised in 1962, 1981 and 1990 for the DCPP-A (with volcanic forcing) and DCPP-C (without volcanic forcing) experiments. HadCRUT5 is used as the observational reference (dashed line). The anomalies have been computed with respect to the period 1970-2005. The shading is the multi-model member spread calculated as the 10 th and 90 th percentiles of the entire ensemble.

To fully exploit these decadal hindcasts and determine whether including the volcanic forcing results in improved predictability in these events, we compare the predicted surface temperature in the three DCPP-A (with volcanic forcing) and the DCPP-C (without volcanic forcing) hindcasts with observations. This protocol also allows us to identify and attribute observed variations to the volcanic forcing. For the global mean surface air temperature, the DCPP-A hindcasts predict the observed anomalies significantly better by reproducing the post volcanic cooling (figure 16). At the local scale, even though the volcanic forcing has a characteristic regional surface air temperature response pattern which evolves with forecast time, an improvement in the DCPP-A hindcasts is only detectable for forecast years 2-5, when the volcanic signal is strongest. For other forecast times considered (year 1 and years 6-9), either the forecast error is greater than the volcanic impacts, the local volcanic signals are overwhelmed by internal variability and/or the regional response to the volcanic forcing is not correctly simulated by the models. In particular we have shown that the volcanic forcing seems to have a weak impact on ENSO, and in the case of Pinatubo degrades the predicted SST anomalies in the tropical Pacific Ocean, as shown in Wu et al. (2023). This is not the case for the other two eruptions, which are no worse in the tropical Pacific with volcanic aerosols included. In contrast, in the North Atlantic Ocean, the volcanic forcing seems to be particularly important for reproducing the observed SST variability in the first few years following the eruptions (figure 17). We also note that

the hindcast corresponding with the eruption of Pinatubo is overall better at predicting the observed anomalies than for the eruptions of Agung and El Chichón (Figures 16 and 17). This could be because the eruption of Pinatubo had a stronger climatic impact and/or because the volcanic forcing is better constrained by the satellite observations available.

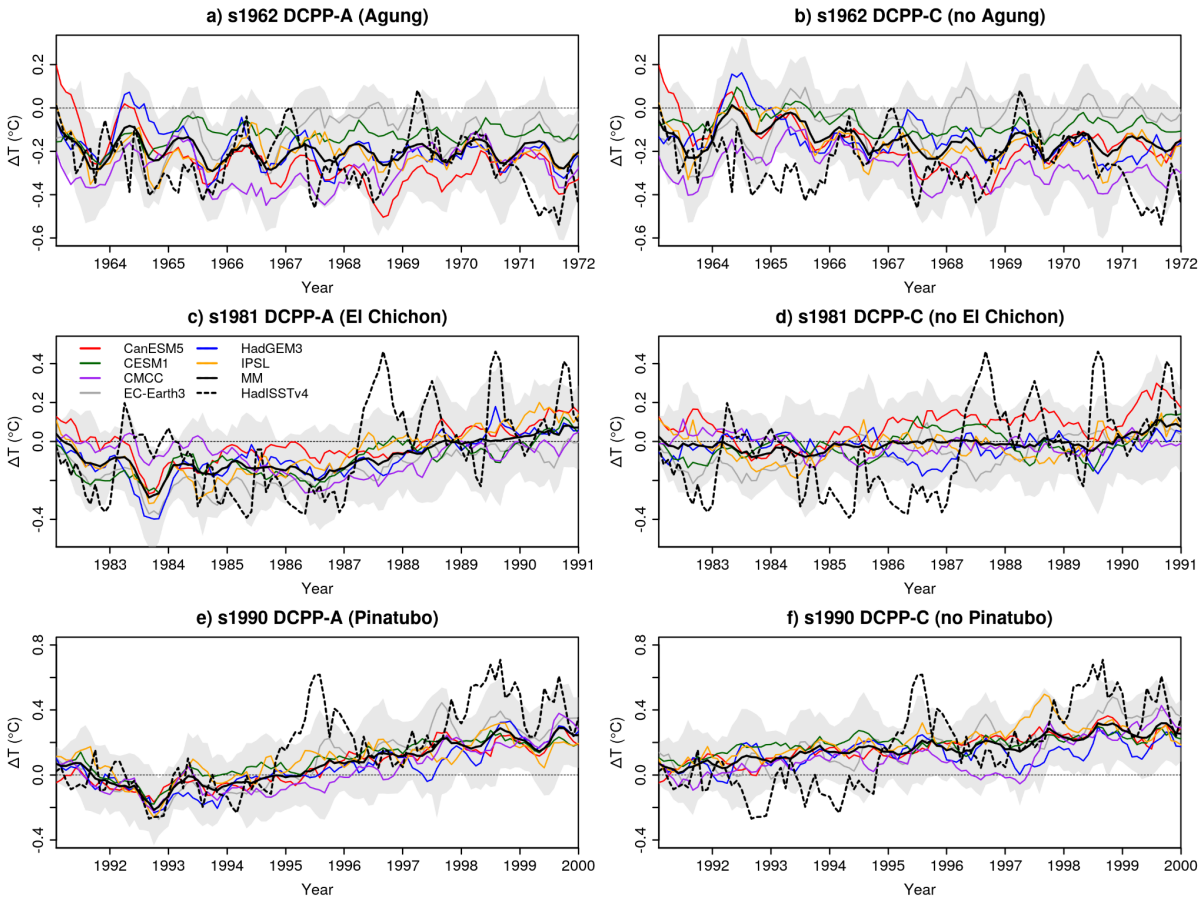

Figure 17. North Atlantic (0°N-60° N, 80° W-0°) SST anomalies (°C) in the predictions initialised in 1962, 1981 and 1990 for the DCPP-A (with volcanic forcing) and DCPP-C (without volcanic forcing) experiments. HadISSTv4 is used as the observational reference (dashed line). The anomalies have been computed with respect to the period 1970-2005. The ensemble mean for each model and the multi-model mean are shown. The shading is the multi-model member spread calculated as the 10 th and 90 th percentiles of all members.

4.2. Evaluating the impact of volcanic forcings generated with EVA and EVA_H in decadal predictions.

In real-time prediction, following a major volcanic eruption, an estimate of the stratospheric sulphate aerosol evolution is needed. For this purpose models such as the Easy Volcanic Aerosol (EVA, Toohey et al., 2016) and its more recent version EVA_H (Aubry et al., 2020), can be used to generate the

stratospheric aerosol forcing due to a volcanic eruption which then can be used as input in climate models. EVA, developed by Toohey et al. (2016), is a simple model of stratospheric aerosol evolution that takes as input the timing and location of an eruption, and the amount of sulphur injected into the stratosphere. It produces forcing files containing aerosol optical properties, including aerosol extinction, single scattering albedo and scattering asymmetry factor as a function of latitude, height, wavelength and time. Recently, Aubry et al. (2020) developed a new version named EVA_H (the H stands for height) which enhanced EVA by accounting for the plume height, predicting the vertical structure of aerosol extinction and calibrating against eruptions spanning a large range of mass of erupted sulphur, plume height, and latitude (EVA was only calibrated against the 1991 Pinatubo eruption).

The objective is to evaluate the climate response to the volcanic forcings produced with EVA and EVA_H using EC-Earth3 decadal hindcasts (Bilbao et al., 2021) for the eruptions of Agung in 1963, El Chichón in 1982 and Pinatubo in 1991, which can inform of the expected uncertainty if these tools were used in a real time forecast. We follow a similar approach to the Decadal Climate Prediction Project (DCPP, Boer et al., 2016). As in DCPP component-C the hindcasts initialised in 1962, 1981 and 1990 (corresponding with the start-dates right before the three major volcanic eruptions) are repeated but with the simulated volcanic forcings from EVA and EVA_H. Comparing these hindcasts with DCPP-A and DCPP-C allows us to determine the expected uncertainty in the climate response when used operationally in the case of a future volcanic eruption.

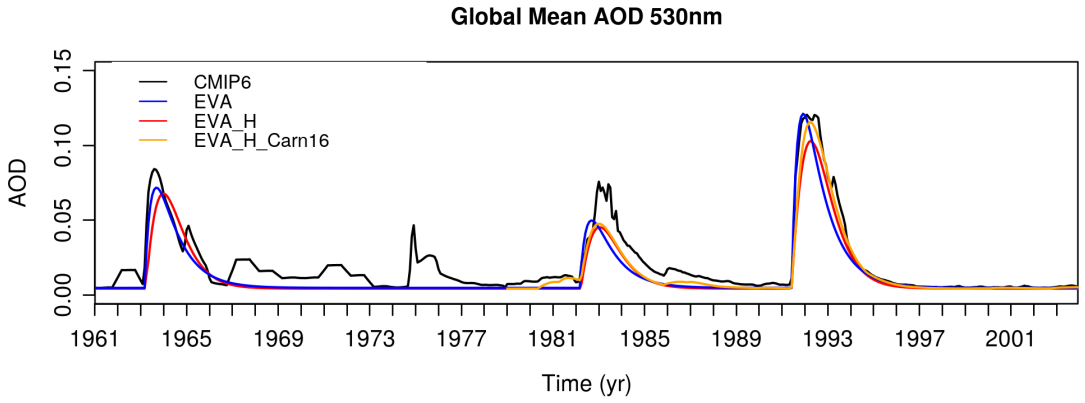

Figure 18: Global mean aerosol optical depth at 530nm. The EVA and EVA_H forcings only have data for the eruptions of Mount Agung in 1963, El Chichón in 1982 and Mount Pinatubo in 1991. The EVA_H_Carn16 forcing was generated with the data from Carn et al. (2016) which includes all volcanic eruptions from 1978-2004.

4.2.1 EVA and EVA_H Volcanic Forcings

In WP2 task 2.3 (Stockdale et al., 2023) we documented the implementation of the volcanic forcings generated by the EVA_H model (Aubry et al., 2020) and a preliminary evaluation of the forcings for the 1982 eruption of El Chichón and the 1991 eruption of Pinatubo against forcings from EVA and CMIP6 (Thomason et al., 2018). Here we extend the analysis with the 1963 eruption of Mount Agung. We take CMIP6 as the best estimate of the observed volcanic forcing acknowledging that for

the eruptions of Agung and El Chichón there is a larger observational uncertainty with respect to the Pinatubo eruption. Figure 18 shows there are differences among the global mean AOD at 530nm for the CMIP6, EVA and EVA_H forcings for the three eruptions. While the EVA_H volcanic forcing reproduces better the temporal evolution to the CMIP6 forcing, the magnitude is smaller at peak values (~15%). The EVA forcing decays sooner and also underestimates the magnitude of the forcing for the eruptions of Agung and El Chichón with respect to the CMIP6, but not for the eruption of Pinatubo, which is comparable. The EVA_H_Carn16 forcing, which includes information from all volcanic eruptions during this period as the CMIP6, reveals that the apparent underestimation of the Pinatubo forcing is mostly resolved when the eruption of Cerro Hudson (Chile) is also considered. This suggests that EVA may overestimate the magnitude of the forcing for the Pinatubo eruption.

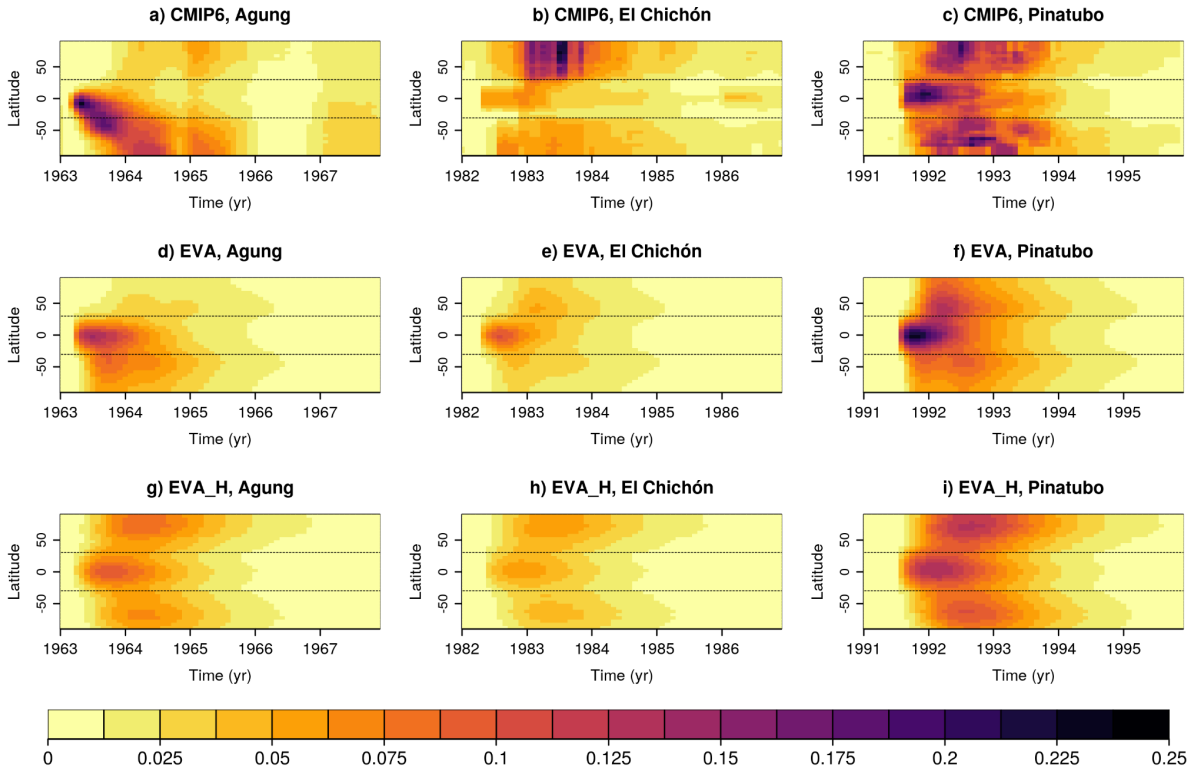

Figure 19: Vertically integrated aerosol optical depth at 530nm as a function of time and latitude for the eruptions of Mount Agung in 1963, El Chichón in 1982 and Mount Pinatubo in 1991 for CMIP6, EVA and EVA_H.

Figure 19 shows that both EVA and EVA_H have deficiencies in simulating the latitudinal structure of the forcing. While the eruption of Pinatubo was mostly hemispherically symmetric, the eruption of Agung mostly affected the Southern Hemisphere and the eruption of El Chichón affected the Northern Hemisphere. For the eruption of Pinatubo, EVA_H simulates reasonably well the three maxima (in the equator and northern and southern hemispheres), while EVA only simulates a strong maxima in the equator, which is overestimated with respect to the CMIP6 forcing. For the other two eruptions we find that the EVA_H simulates overall a similar forcing structure to Pinatubo and cannot account for the latitudinal asymmetry of the forcing. In the case of EVA, the latitudinal asymmetry of

the forcing is better captured for the eruption of Agung, but not for El Chichón, which strongly underestimates the magnitude of the forcing in the Northern Hemisphere.

4.2.2 Global Mean Volcanic Response

To investigate the climate impacts and evaluate the volcanic forcings produced with EVA and EVA_H, we repeat the EC-Earth3 decadal hindcasts initialised in 1962, 1981 and 1990 (described in section 4.1), but with the volcanic forcings estimated with EVA and EVA_H. Following the DCPP-C protocol (described in section 4.1), to determine the volcanic impacts we subtract the DCPP-C (no volcanic forcing) hindcasts from these hindcasts.

We start by analysing the global mean top-of-atmosphere radiation (TOA) flux response, calculated as anomalies of incoming shortwave minus outgoing shortwave and out-going longwave radiation. The predictions with the EVA and EVA_H volcanic forcings show a post-volcanic decrease in global mean TOA comparable with the DCPP hindcasts but with evident differences in the magnitude and the temporal evolution for the three eruptions (Figure 20a-c). For the eruption of Agung, the EVA and EVA_H hindcasts have a weaker TOA response (~40% and ~64% respectively at peak values), consistent with the weaker forcings (Figure 20a). For the eruption of El Chichón, the magnitude of the response is also underestimated by EVA and EVA_H (~50% and ~60% respectively at peak values), and since this eruption is weaker, the TOA response is barely significant (Figure 20b). In the case of the Pinatubo eruption (largest negative anomaly) the impact of the EVA_H forcing on the net incoming energy is approximately 40% weaker than for the CMIP6 forcing, partly due to not including the Cerro Hudson, while EVA forcing is approximately 10% stronger (Figure 20c). The relative magnitude differences among the forcings (Figure 18) and among the TOA responses for the eruptions indicate that the radiative response is not linear and probably associated with the latitudinal differences in the forcing.

In response to the negative TOA anomalies, the global mean surface temperature response cools (figure 20d-f), however there are evident differences with respect to the DCPP-A. For the eruptions of Agung and El Chichón, where both EVA and EVA_H underestimate the magnitude of the forcings, we find that the post-volcanic cooling is greatly underestimated with respect to DCPP-A (figure 20d and e). In the case of Pinatubo a progressive post eruption cooling until approximately 1993, when the cooling reaches its maximum with all three forcings (figure 20f). Consistent with the TOA radiative flux differences, even if the differences in both variables do not relate linearly, the EVA_H forcing yields weaker negative global surface temperature anomalies (~-0.3℃) than CMIP6 (~-0.4℃), while EVA forced anomalies remain closer to the CMIP6 forced ones. Although the CMIP6 and EVA temperature response is similar this is for the wrong reasons we expect both the EVA and EVA_H to be weaker, since the Cerro Hudson eruption is not included. Also, the temporal evolution seems better captured by the EVA_H forcing, since the EVA forcing is initially stronger (consistent with greater negative temperature anomalies early on) and persists for a shorter time.

The global mean temperature in the lower stratosphere (50 hPa) shows strong post eruption warming anomalies, with small ensemble spread in comparison to other variables, and clearly

illustrates fundamental differences induced by the forcings (figure 20g-i). Following the eruptions of Agung and El Chichón, the EVA and EVA_H forcings underestimate the lower stratospheric warming with respect to DCPP-A (by ~24% and ~68% respectively at peak values). Consistent with the results described previously, the EVA_H forcing produces a weak response (~1.5°C) in comparison to CMIP6 forcing (~2.8°C), while with the EVA forcing the response is stronger (~3.6°C). There are evident temporal structural differences in the nature of the response to the idealised forcings, with the EVA_H and CMIP6 signals showing similarities (despite the difference in magnitude), while EVA signal peaks sooner and recovers faster.

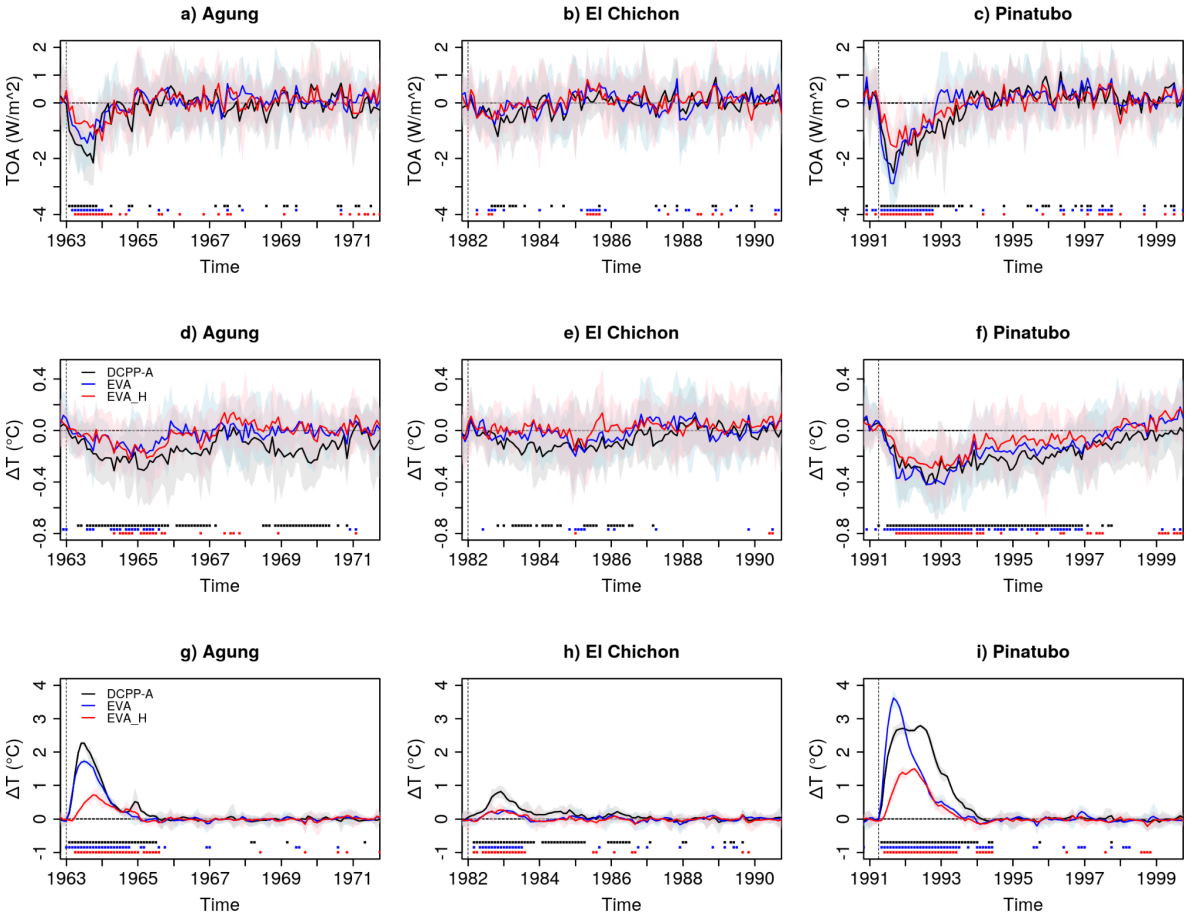

Figure 20: Global mean top-of-atmosphere radiation (W/m²), global mean surface air temperature (°C) and global mean lower stratospheric (50hPa) temperature (°C) response to the volcanic eruptions (volc - no volc), for DCPP-A, EVA and EVA_H. The shading is the multi-model member spread calculated as the 10th and 90th percentiles of the entire ensemble. Filled squares at the bottom part of the figure indicate statistically significant differences according to a bootstrap with resampling with 1000 iterations.

4.2.3 Spatiotemporal Characteristics of the Volcanic Response

Next we explore the impact of the latitudinal variation of the forcing by comparing the TOA radiative flux anomaly maps the first year following the volcanic eruption (figure 21), when the forcing is strongest. As previously shown in section 4.2.1, the volcanic forcings generated with EVA and EVA_H

show some limitations in simulating the latitudinal variation and this is reflected in the TOA response. In the case of Agung, the hindcasts with the EVA and EVA_H forcings do not simulate the decreased TOA in the Southern Hemisphere, as predicted by DCPP-A (figure 21 a,d,g). Similarly, in the case of El Chichón, the hindcasts with the EVA and EVA_H forcings do not simulate the Northern Hemisphere (figure 21 b,e,h) neither, although the anomalies are significantly weaker. For the eruption of Pinatubo, the TOA anomalies in the hindcasts with the EVA and EVA_H forcings are comparable to those from DCPP-A, although the magnitude varies (figure 21 c,f,i). With the EVA forcing the response is slightly stronger than the response to the CMIP6 forcing (~1 W/m^2), while with the EVA_H forcing response is considerably weaker along the equator than (~-2 W/m^2), although this is partly related to the absence of the Cerro Hudson eruption as previously mentioned.

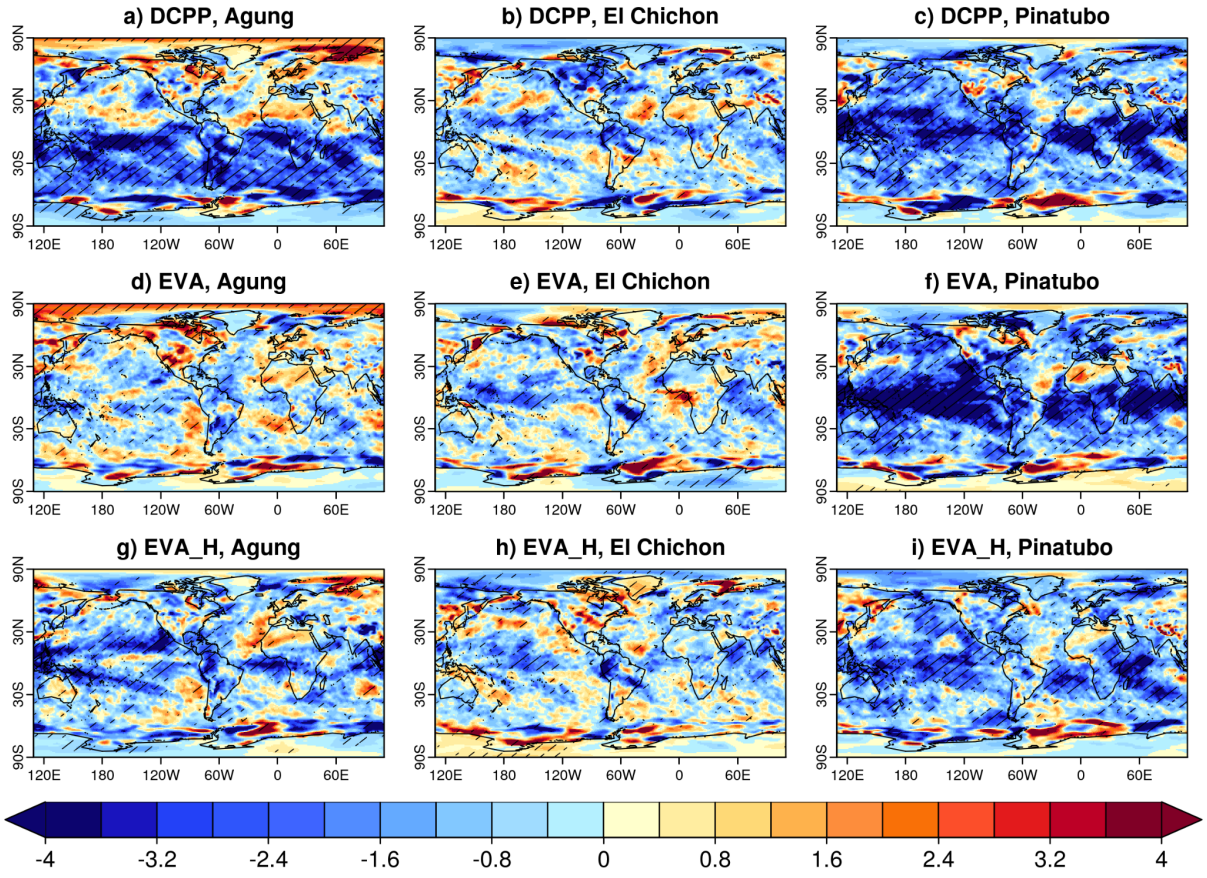

Figure 21: TOA (W/m^2) response (volc - no volc) the first year following the eruptions (June-May) for DCPP, EVA and EVA_H. Hatching indicates statistically significant anomalies according to a bootstrap with resampling with 1000 iterations.

These limitations in simulating the geographical pattern of the radiative response affect the surface temperature response. Figure 22 shows the global mean surface temperature for years 2-5 following the volcanic eruptions. For the eruptions of Agung and El Chichón, we find that the hindcasts with the EVA and EVA_H forcings barely reproduce cool anomalies with respect to DCPP, and probably underestimate the cooling due to the volcanic eruptions. In contrast, for the eruption of Pinatubo, we find that the hindcasts simulate widespread cooling in response to the volcanic forcing, with

maxima over the Arctic. As previously mentioned, the hindcasts with the EVA_H forcing simulate overall weaker cooling.

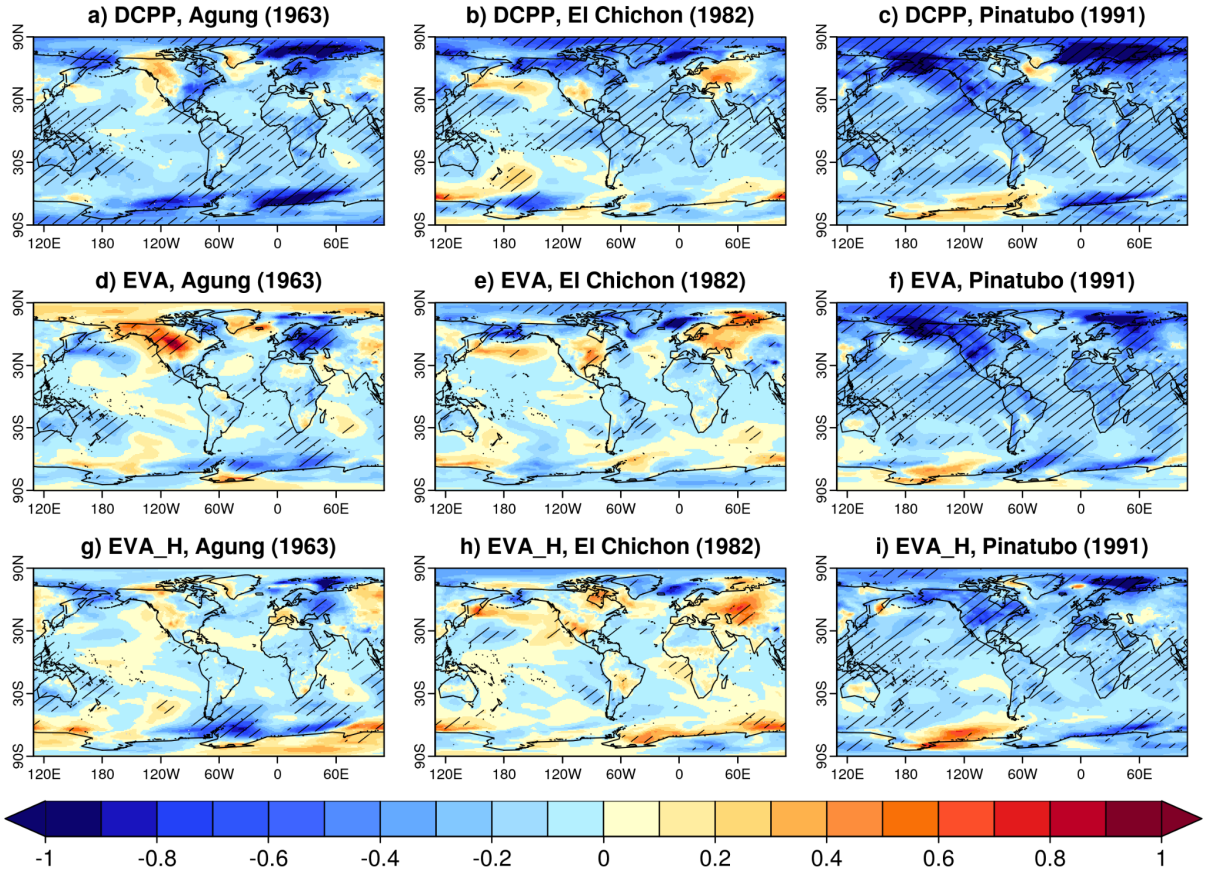

Figure 22: Surface air temperature (°C) response (volc - no volc) for years 2-5 following the eruptions for DCPP, EVA and EVA_H. Hatching indicates statistically significant anomalies according to a bootstrap with resampling with 1000 iterations.

4.2.4 Impact of volcanic eruptions on climate variability

The volcanic radiative response may lead to other climate impacts on seasonal-to-decadal timescales (e.g. Swingedouw et al., 2017; Marshall et al., 2022), such as atmospheric and oceanic dynamical changes, which may modulate climate variability. For example, previous studies have shown that the post eruption stratospheric warming impacts the atmospheric circulation by increasing in the polar vortex strength and the surface winds, resulting in a warming of the North Eurasian continent the first winter after the eruption (e.g. Hermanson et al., 2020). Given the differences found in the stratospheric temperature response (Figure 20g-i) it might be expected to find a response in the Northern Hemisphere polar vortex, however no signal was found probably due to the small ensemble size (10 members per eruption).

Volcanic eruptions also impact El Niño Southern Oscillation (ENSO), increasing the likelihood of El Niño-like response in the first year following an eruption (e.g. McGregor et al., 2020). However in

these simulations we do not find a response. This is again probably partly due to the small ensemble size and also because the response seems to be model dependent and EC-Earth3 is one of the models with barely any response in this respect, as shown in Bilbao et al. (in review).

The Atlantic Ocean is another region of relevance following volcanic eruptions. Studies have shown that on multiannual to decadal timescales, the strength of the AMOC increases in response to large volcanic eruptions (e.g. Stenchikov et al., 2009; Hermanson et al., 2020) which may impact the Atlantic Multidecadal Variability (AMV). Again, no signal was found in these hindcasts. As shown in Bilbao et al. (in review), the AMOC response (and the mixed layer depth response in the Labrador Sea) is model dependent and EC-Earth3 is one of the models that does not show a response.

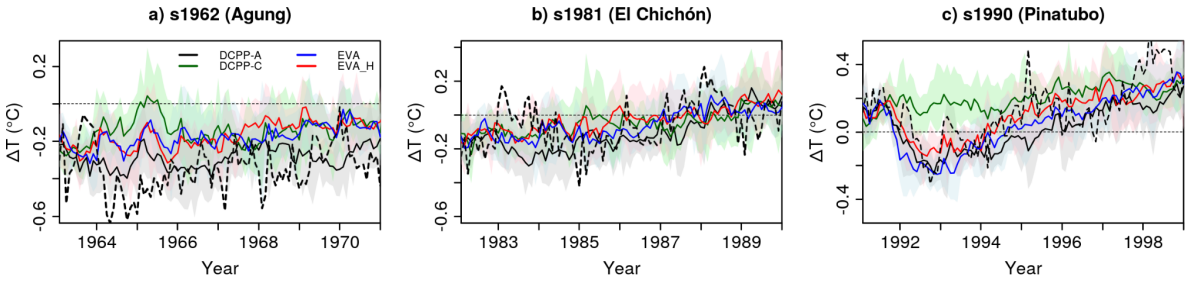

Figure 23: Global mean surface air temperature anomalies (°C) in the predictions initialised in 1962, 1981 and 1990 for the DCPP-A, DCPP-C, EVA and EVA_H hindcasts. HadCRUT5 is used as the observational reference (dashed line). The anomalies have been computed with respect to the period 1970-2005 (see methods for further information). The ensemble mean for each hindcast is shown. The shading is the multi-model member spread calculated as the 10th and 90th percentiles of the ensemble.

4.2.5 Comparison of the predicted surface temperature with observations

In section 4.1 we showed the importance (and limitations) of including the volcanic forcing in CMIP6 decadal hindcasts to reproduce the observed surface temperature variations, which result in overall better predictions. The focus of this section is to determine whether using the volcanic forcings simulated by EVA and EVA_H result in improved predictability for these three events as with the CMIP6 forcing. Figure 23 shows that overall the EC-Earth3 hindcasts which include the volcanic forcing (DCPP-A, EVA and EVA_H) tend to reproduce the HadCRUT5 global mean temperature anomalies more closely than when the volcanic forcing is omitted (DCPP-C). This is particularly relevant for the 1991 eruption of Pinatubo (Figure 23c), in which including the volcanic forcing is important to simulate the observed global mean surface temperature variability in the early 90s. For this eruption, despite the differences among the hindcasts with the CMIP6, EVA and EVA_H forcings discussed earlier, the observed anomalies are generally within the uncertainty of the hindcast ensembles (within the 10th and 90th percentiles). Note however that the EVA_H forcing for the Pinatubo produces slightly weaker surface temperature anomalies and this could be due to the absence of the Cerro Hudson eruptions, which also had a significant climate impact.

For the eruptions of Agung and Pinatubo it is not possible to make further conclusions in this respect since the s1962 and s1981 hindcasts struggle to reproduce the observed variability. Although the EVA and EVA_H have limitations in reproducing the magnitude and latitudinal structure of the forcing, probably underestimating the global mean surface temperature response, the DCPP-A hindcast also struggles to predict the observed variability. This could be partly associated with the uncertainties in both the surface temperature observations and in reconstructing the volcanic forcings in the pre-satellite era. These two eruptions are also weaker in comparison to Pinatubo, which implies that internal variability may dominate the response and therefore it is harder to accurately predict the temperature variations on these timescales.

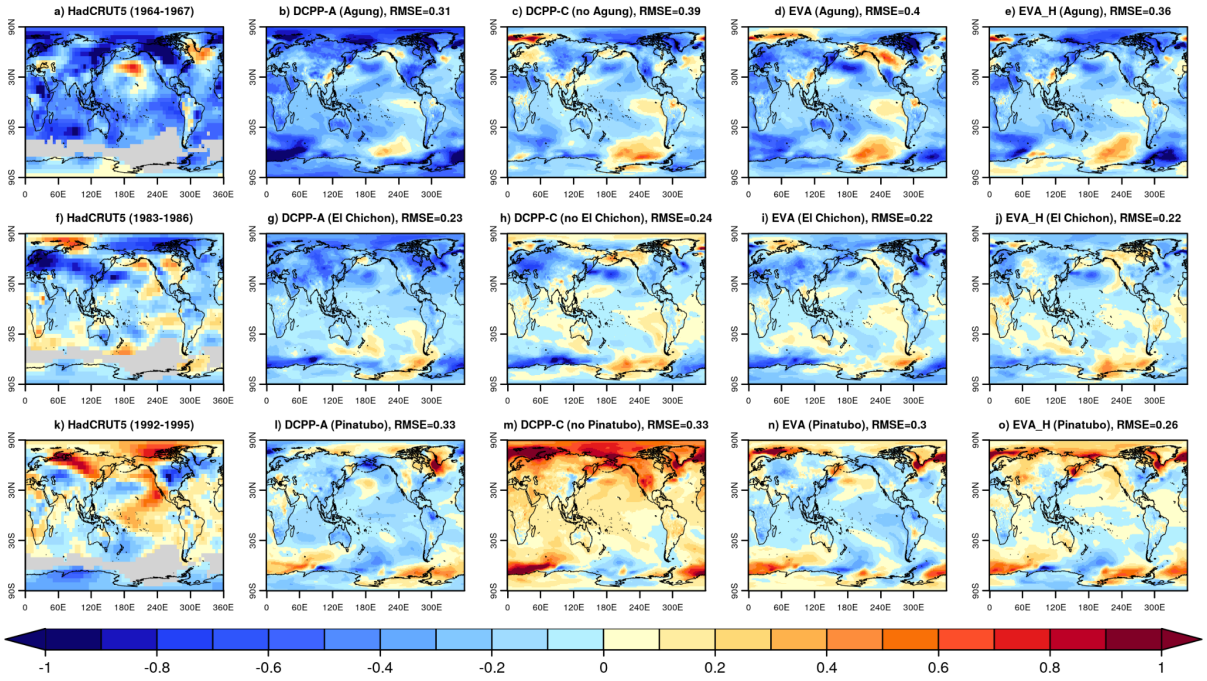

Figure 24: Surface air temperature anomalies (°C) for forecast years 2-5 in the predictions initialised in 1962, 1981 and 1990 for the DCPP-A, DCPP-C, EVA and EVA_H hindcasts. HadCRUT5 is used as the observational reference (dashed line). The anomalies have been computed with respect to the period 1970-2005. The ensemble mean for each hindcast is shown.

The regional surface temperature anomalies predicted by the EC-Earth3 hindcast sets were also compared with the HadCRUT5 observations, focusing on forecast years 1, 2-5 (Figure 24) and 6-9. Overall the surface temperature anomaly patterns are largely consistent among the four hindcast experiments, with magnitude variations associated with the post volcanic cooling pattern (shown in the previous section). Figure 24 shows the surface temperature anomalies for HadCRUT5 and the four hindcast experiments for forecast years 2-5, when the impact is strongest. It is evident that overall the DCPP-C hindcasts simulate warmer conditions since the post volcanic cooling is not included, particularly for the eruption of Pinatubo. The rest of hindcasts experiments (DCPP-A, EVA and EVA_H) which include the volcanic forcing predict anomalies closer to observations. However, the hindcasts do not accurately reproduce the regional variability, particularly in the tropical Pacific.

As in Bilbao et al. (in review) we compute the area weighted RMSE for the different forecast times to determine the forecast error and detect improvements in the regional pattern associated with the volcanic forcing. While Bilbao et al. (in review) show that a detectable improvement in the regional pattern associated with the volcanic forcing (in the CMIP6 multi-model mean) in forecast years 2-5, when limiting the comparison to EC-Earth3 alone, an improvement at the regional scale is not always evident. This is shown by the RMSE values in the titles of the panels of Figure 24. This might be because (1) the forecast error at the regional level is greater than the volcanic impact, (2) the local volcanic response is overwhelmed by internally generated variability and/or (3) the regional response to the volcanic forcing is not correctly simulated by the models.

5. Conclusions and recommendations for implementation.

Volcanic aerosol can have a substantial impact on the earth system. The importance of correctly modelling its details, however, depends on the questions being asked and the timescales being considered. For decadal forecast systems, the questions focus on understanding the overall impact of past and hypothetical future eruptions, with a focus on the overall level of radiatively-driven cooling and hemispheric distributions. At the seasonal timescale, the emphasis is on understanding the detailed change in the forecast for the coming months caused by volcanic aerosol on top of all the other specific initial conditions (including the accumulated impact of the volcanic aerosol by the start of the forecast) and other forcing factors for a specific forecast; dynamical impacts on tropospheric circulation and the vertical structure of heating profiles are also important. Because of these different emphases, we summarise our conclusions for the two timescale separately.

5.1 Seasonal forecasting systems

As shown in our previous Deliverable D2.3 (Stockdale et al, 2023), and considering additionally the work reported here on IFS optical properties of stratospheric sulphate aerosol, the developments made in CONFESS have allowed for a substantial improvement in the treatment of volcanic aerosol in seasonal prediction systems.

For seasonal predictions with IFS, we are now able to properly represent the vertical structure of the stratospheric heating from aerosol. We can do this using satellite-derived observational datasets, typically produced some time after a volcanic eruption, and importantly we can also do it using the EVA_H model in a real-time forecast context, once an eruption has occurred and we have estimates of both the mass of SO_2 released and its injection height. Thanks to the VolRes community, this information is proving to be readily available for significant eruptions, not quite instantly but certainly on a timescale suitable to be used for real-time predictions. The quality of the aerosol distribution produced by EVA_H seems generally satisfactory, given other uncertainties. We note the suggestion from Figure 19 that the hemispheric distributions may not be as good as we like, and for the case of Agung this is on the face of it true. For the case of El Chichón, as discussed, the observed tropical values in GloSSAC may not be accurate enough to be a useful constraint. Physical modelling of the transport, dispersion and conversion to SO4 for a given injection of SO2 might prove to be a better constraint on the quality of EVA_H than the poorly observed aftermath of older eruptions.

Another way to assess and constrain the quality of the EVA_H model or future upgrades of it might be to focus on the impact of the more minor eruptions that have occurred over the last 20 years or so, where the stratospheric aerosol has been better observed. Figure 25 shows a time-latitude plot of extinction at the equator, from 1979 to 2021, with GloSSACv2.2 at the top and the output of a modified version of EVA_H at the bottom. This modified version is driven by the MSVOL4 dataset (Carn et al 2022) which presently extends to 2022. Since this has a different estimate of the strength of the Pinatubo eruption, the coefficient controlling the amplitude has been rescaled to maintain the fit to the GloSSAC data, which has the side effect (beneficial for El Chichon at least) of slightly increasing the amplitude of other volcanic events.

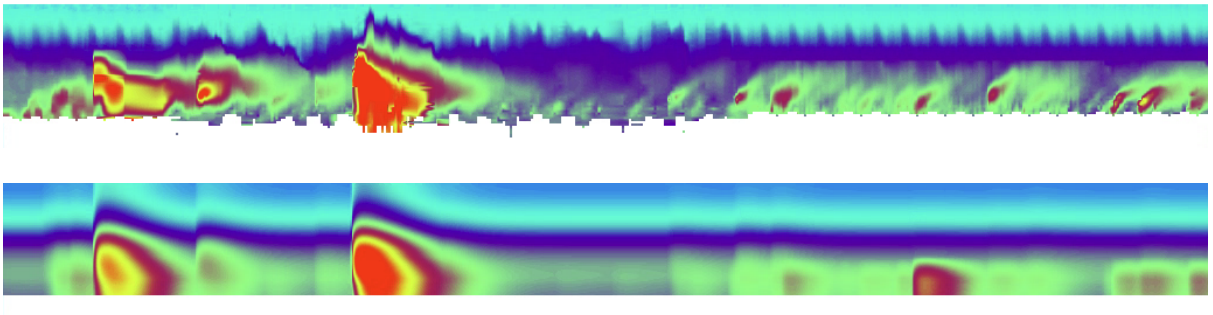

Figure 25: TIme-height plot of extinction at 2.5N from 1979 to end of 2021, according to GloSSACV2.2 (top) and EVA_H modified to use MSVOL4 inputs (bottom).

Although there is a reasonable fit between model and data for Pinatubo and (to a lesser extent) for El Chichon, the agreement in the 2000s is more hit and miss. Each of the modelled eruptions has a corresponding signal in observations, but the relative strength of the eruptions varies greatly between the datasets. Note also the exquisite level of detail in the observed data for this period, showing the upward transport of aerosol at the equator as part of the "stratospheric tape recorder", familiar from stratospheric humidity transport. Unlike the earlier part of the record, there is no good reason to doubt the essence of what the observations are showing. Equally, the MSVOL4 data for the amount of SO2 injected by each eruption is based on satellite data and one might think this should also be reasonably reliable. The biggest uncertainty by far, however, is in how much SO2 is injected at which height. EVA_H reads a "plume height" and assumes that all of the SO2 is detrained in a narrow height window around this. Errors in this assumption, which are very likely especially for lower-altitude plumes, will have a large impact on the estimate of how much SO2 actually reaches the stratosphere and at which altitude. EVA_H also clearly lacks the ability to produce spatial detail in the transports, but arguably the uncertainty in the injection data is the bigger problem.

The revised optical properties in the IFS allow us to represent the aerosol optics reasonably well over the range of aerosol size distributions likely to be encountered after volcanic eruptions of any significance. The optical properties are not perfect, however, and will mis-represent what happens when there are only low levels of aerosol present, which admittedly matters very little as regards radiative impact.

The use of EVA_H to represent the time variation of volcanic aerosol across multiple decades allows us to run the ECMWF seasonal forecasting system many decades into the past, using existing datasets of volcanic eruptions. This is relevant and valuable for the forthcoming SEAS6 operational implementation at ECMWF, which is planning to include reference re-forecasts from 1961 through to the present. The biggest challenge in reproducing stratospheric temperatures in the decades before the standard seasonal calibration period (starting in 1993) is not the volcanic aerosols but other forcings affecting stratospheric temperature trends, most likely ozone. The biggest challenge in a real-time forecast situation will be the reliability of the SO2 injection estimates.

For immediate practical purposes, the most disappointing aspect of the results is the very small dynamical impact seen in the troposphere and at the surface in boreal winter, as assessed for the two winters following the Pinatubo eruption. A moderate dynamical signal is present in the polar stratospheric vortex, but penetrates only very weakly down to the surface. The fact that the lower

tropospheric signal matches the (much larger) anomaly observed supports the hypothesis that a real signal is present but that the model does not capture its amplitude properly, over against a null hypothesis that the observed variability is just chance. The failure of models to propagate stratospheric signals down to the surface has been documented (eg. Stockdale et al, 2015), and is seen across all years with variability of the polar vortex, not just the small sample of volcanic years. An alternative hypothesis is that we are not correctly representing the volcanic forcing near the tropopause. Compared to earlier model representations, which had more weight close to the tropopause, it is notable that both GloSSACv2 and the EVA_H which is trained on it have the volcanic aerosol diminishing as it approaches the level of the mean climatological tropopause. This may be due to sampling the alternating periods of tropospheric and stratospheric air which exist at a given height because of the dynamical variability of the tropopause. The producers of the GloSSAC data set have solicited views on possible future improvements, and the proper distinction between stratospheric and tropospheric air masses was raised by more than one group (including us). Using observational data to disentangle different aerosol types above and below a dynamically varying tropopause may not be easy, and an alternative method to estimate how volcanic aerosol varies in the region of the tropopause might be to explicitly model the transport.

Given the present state of the art, our recommendation is to go ahead and use GloSSACV2.2 and EVA_H, together with the revised optical properties for volcanic stratospheric aerosol in the IFS. If EVA_H can be retuned using the latest GloSSAC and MSVOL emission datasets, so much the better. Nonetheless it is important to be aware of the limitations of this approach, chief of which are uncertainty in the actual stratospheric injection details following eruptions, and uncertainty with what is happening to aerosol concentrations around the tropopause. It is important to continue to liaise with the volcanic community to improve observational data estimates in the tropopause region, and to work towards an “operationalisation” of volcanic emission datasets. In the longer term, explicitly modelling the volcanic aerosol within the forecast models, or possibly training machine-learning systems on such integrations, will allow yet further improvements in our representation of volcanic aerosol. All of these improvements need to be in the context of trying to improve the general ability of models to propagate signals from the winter stratosphere down to the surface.

5.1 Decadal prediction systems

In this deliverable we have shown the importance of including the impact of large volcanic eruptions on decadal climate predictions to produce accurate forecasts. We have presented the results from two analyses: (1) the impact of volcanic eruptions on decadal predictions based on the recent eruptions of Mount Agung (1963), El Chichón (1982) and Mount Pinatubo (1991) using a multi-model set of decadal predictions which contribute to the CMIP6 Decadal Climate Prediction Project (DCPP Boer et al., 2016), and (2) we have analysed decadal predictions for the three eruptions but with volcanic aerosol forcings produced with EVA and EVA_H (simple models that predict the stratospheric aerosol forcing evolution) to evaluate the expected uncertainty of using these tools in real time forecasts in the case of a future large eruption.

We have shown that the CMIP6 decadal prediction contributing to the DCPP component C (DCPP-C) exhibit a strong agreement in predicting the radiative response to the volcanic eruptions, simulating a reduction in global mean top-of-atmosphere radiation fluxes, surface temperature and ocean heat content. The characteristic geographical patterns of the response are consistent across the models and share similarities across the volcanic eruptions, however some differences across models and eruptions arise due to the varying magnitude and spatiotemporal structure of the volcanic forcing. Taking advantage of the large multi-model ensemble we have analysed the dynamical responses in the Northern Hemisphere atmospheric circulation, in the tropical Pacific Ocean and the North Atlantic Ocean, focusing on the multi-model response but also highlighting that there are important differences both across models and across eruptions. Comparing the predicted surface temperature anomalies in the two sets of hindcasts (with and without volcanic forcing) with observations we show that including the volcanic forcing results in overall better predictions. The volcanic forcing is found to be particularly relevant for reproducing the observed SST variability in the North Atlantic Ocean following the 1991 eruption of Pinatubo, however in the tropical Pacific Ocean the predicted SST anomalies are degraded. These results are a summary of the publication submitted to Earth System Dynamics (Bilbao et al., in rev., https://esd.copernicus.org/preprints/esd-2023-36/).

Comparing the volcanic forcings for the eruptions of Agung, El Chichón and Pinatubo, generated with EVA and EVA_H with the CMIP6 forcing, we have shown that these tools have some limitations in reproducing the magnitude and latitudinal structure of the forcing, despite the inherent observational uncertainties. Although EVA_H exhibits a more similar temporal evolution of the forcing compared to EVA, it fails to capture the latitudinal variations crucial for hemispherically asymmetrical eruptions like Agung and El Chichón. The differences in the volcanic forcings lead to variations in the radiative responses simulated by EC-Earth3 decadal predictions, as evidenced by the differences in TOA, surface temperature and lower stratospheric temperatures. For the eruptions of Agung and El Chichón, both hindcasts with EVA and EVA_H forcings underestimate the radiative response with respect to DCPP-A. For the eruption of Pinatubo the hindcasts with the EVA forcing exhibit a stronger radiative response than DCPP-A, while in the hindcasts with the EVA_H forcing it is weaker. Using EVA_H has allowed us to reveal that the weaker response is partly because the eruption of Cerro Hudson was not taken into consideration, which made a substantial contribution to the forcing, and when this eruption is included the forcing is much closer to the CMIP6 forcing. This shows the potential of EVA_H not only to predict the forcing for a future volcanic eruption, but also to understand the past.

In summary, the results show that including the volcanic forcing is necessary to make skillful climate predictions following major volcanic eruptions, particularly in capturing the direct radiative effects, which are consistently reproduced by climate models (at least those analysed). In contrast, the dynamical impacts exhibit greater uncertainty, as they require large ensembles to detect responses, the responses are model dependent, they can be affected by the background climate conditions and there is evidence that models might be deficient in simulating some of the impacts (i.e. on the tropical Pacific variability). Regarding real-time predictions, our study indicates that, for Pinatubo-like eruptions, both EVA and EVA_H forcings can be reasonable choices, at least for reproducing the

surface temperature response, given the inherent forecast uncertainty. However, for eruptions akin to Agung and El Chichón, EVA and EVA_H exhibit limitations, despite the greater observational uncertainty than in the case of Pinatubo, and likely result in an underestimation of the radiative response. This underscores the immense value of tools like EVA and EVA_H, emphasising the need for ongoing development to contribute to minimising uncertainties in decadal forecasts.

6. References

CONFESS Deliverables:

Stockdale, T., Senan, R. and Bilbao, R. (2022). CONFESS D2.2: Harmonized CAMS and CMIP6 datasets for aerosols.

Stockdale, T., Bilbao, R. and Ortega, P. (2023). CONFESS D2.3: Simplified volcanic aerosol prediction module validated and interfaced to IFS and EC-Earth.

Peer reviewed articles:

Aubry, T. J., Toohey, M., Marshall, L., Schmidt, A., & Jellinek, A. M. (2020). A new volcanic stratospheric sulfate aerosol forcing emulator (EVA_H): Comparison with interactive stratospheric aerosol models. Journal of Geophysical Research: Atmospheres, 125, e2019JD031303.

Azoulay, A., Schmidt, H., & Timmreck, C. (2021). The Arctic polar vortex response to volcanic forcing of different strengths. *J. Geophys. Res.: Atmospheres*, **126**, e2020JD034450. https://doi.org/10.1029/2020JD034450

Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze, T., Bretonnière, P.-A., Caron, L.-P., Castrillo, M., Cruz-García, R., Cvijanovic, I., Doblas-Reyes, F. J., Donat, M., Dutra, E., Echevarría, P., Ho, A.-C., Loosveldt-Tomas, S., Moreno-Chamarro, E., Pérez-Zanon, N., Ramos, A., Ruprich-Robert, Y., Sicardi, V., Tourigny, E., and Vegas-Regidor, J.: Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth, Earth System Dynamics, 12, 173–196, https://doi.org/10.5194/esd-12-173-2021, 2021.

Bilbao, R., Ortega, P., Swingedouw, D., Hermanson, L., Athanasiadis, P., Eade, R., Devilliers, M., Doblas-Reyes, F., Dunstone, N., Ho, A.-C., Merryfield, W., Mignot, J., Nicolì, D., Samsó, M., Sospedra-Alfonso, R., Wu, X., and Yeager, S.: Impact of volcanic eruptions on CMIP6 decadal predictions: A multi-model analysis, Earth Syst. Dynam. Discuss. [preprint], https://doi.org/10.5194/esd-2023-36, in review, 2023.

Bittner, M., Schmidt, H., Timmreck, C., and Sienz, F.: Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty, Geophysical Research Letters, 43, 9324–9332, https://doi.org/https://doi.org/10.1002/2016GL070587, 2016.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geoscientific Model Development, 9, 3751–3777, 2016.

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

Carn, S., 2022: Multi-Satellite Volcanic Sulfur Dioxide L4 Long-Term Global Database V4, Greenbelt, MD, USA, Goddard Earth Science Data and Information Services Center (GES DISC), Accessed December 2023. 10.5067/MEASURES/SO2/DATA405

DallaSanta K, Polvani LM (2022) Volcanic stratospheric injections up to 160 tg(s) yield a eurasian winter warming indistinguishable from internal variability. Atmospheric Chemistry and Physics Discussions 2022:1–31

Driscoll, S., Bozzo, A., Gray, L. J., Robock, A., and Stenchikov, G. (2012), Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions, J. Geophys. Res., 117, D17105,

Graf, H., Kirchner, I., Robock, A., and Schult, I., 1993: Pinatubo eruption winter climate effects: Model versus observations, *Clim. Dynam.*, 9, 81–93. https://doi.org/10.1007/BF00210011

Hermanson, L., Bilbao, R., Dunstone, N., Ménégoz, M., Ortega, P., Pohlmann, H., Robson, J. I., Smith, D. M., Strand,G., Timmreck, C., Yeager, S., and Danabasoglu, G.: Robust Multiyear Climate Impacts of Volcanic Eruptions in Decadal Prediction Systems, Journal of Geophysical Research: Atmospheres, 125, 2020.

Hermanson, L., Smith, D., Seabrook, M., Bilbao, R., Doblas-Reyes, F., Tourigny, E., Lapin, V., Kharin, V. V., Merryfield, W. J., Sospedra-Alfonso, R., Athanasiadis, P., Nicoli, D., Gualdi, S., Dunstone, N., Eade, R., Scaife, A., Collier, M., O'Kane, T., Kitsios, V., Sandery, P., Pankatz, K., Früh, B., Pohlmann, H., Müller, W., Kataoka, T., Tatebe, H., Ishii, M., Imada, Y., Kruschke, T., Koenigk, T., Karami, M. P., Yang, S., Tian, T., Zhang, L., Delworth, T., Yang, X., Zeng, F., Wang, Y., Counillon, F., Keenlyside, N., Bethke, I., Lean, J., Luterbacher, J., Kolli, R. K., and Kumar, A.: WMO Global Annual to Decadal Climate Update: A Prediction for 2021–25, Bulletin of the American Meteorological Society, 103, E1117 – E1129, https://doi.org/https://doi.org/10.1175/BAMS-D-20-0311.1, 2022.

Lacis., A., 2015: Volcanic aerosol radiative properties. *Past Global Changes Magazine* 23(2) 50-51 https://doi.org/10.22498/pages.23.2.50

Lund Myhre, C. E., Christensen, D. H., Nicolaisen, F. M. and Nielsen, C.J., 2003: Spectroscopic Study of Aqueous H2SO4 at Different Temperatures and Compositions: Variations in Dissociation and Optical Properties. *The Journal of Physical Chemistry A* 107 (12), 1979-1991. DOI: 10.1021/jp026576n

Marshall L.R., Maters, E.C., Schmidt, A., Timmreck, C., Robock, A., & Toohey, M., (2022) Volcanic effects on climate: recent advances and future avenues. Bulletin of Volcanology 84:131.

McGregor, S., Khodri, M., Maher, N., Ohba, M., Pausata, F. S. R., and Stevenson, S.: The Effect of Strong Volcanic Eruptions on ENSO, chap. 12, pp. 267–287, American Geophysical Union (AGU), ISBN 9781119548164, https://doi.org/https://doi.org/10.1002/9781119548164.ch12, 2020.

Ménégoz, M., Bilbao, R., Bellprat, O., Guemas, V., & Doblas-Reyes, F., (2018) Forecasting the climate response to volcanic eruptions: prediction skill related to stratospheric aerosol forcing. Environmental Research Letters 13(6):064,022

Moosmüller, H.; Ogren, J.A., 2017: Parameterization of the Aerosol Upscatter Fraction as Function of the Backscatter Fraction and Their Relationships to the Asymmetry Parameter for Radiative Transfer Calculations. *Atmosphere*, 8, 133. https://doi.org/10.3390/atmos8080133

NASA/LARC/SD/ASDC, 2022. Global Space-based Stratospheric Aerosol Climatology Version 2.2. Available at: https://doi.org/10.5067/GLOSSAC-L3-V2.2

Palmer, K.F., & Williams, D., 1975: Optical constants of sulfuric acid; application to the clouds of Venus? *Applied Optics*, 14 1, 208-19 .

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.

Predybaylo, E., G. L. Stenchikov, A. T. Wittenberg, and F. Zeng (2017), Impacts of a Pinatubo-Size Volcanic Eruption on ENSO, *J. Geophys. Res. Atmos.*, 122, 925–947, doi:10.1002/2016JD025796.

Predybaylo, E., Stenchikov, G., Wittenberg, A. T., and Osipov, S.: El Niño/Southern Oscillation response to low-latitude volcanic eruptions depends on ocean pre-conditions and eruption timing, Communications Earth Environment, 12, 2662–4435, https://doi.org/10.1038/s43247-020-0013-y, 2020.

Robock, A. and Mao, J., 1992: Winter Warming from Large Volcanic Eruptions, *Geophys. Res. Lett.*, 12, 2405–2408. https://doi.org/10.1029/92GL02627

Robock, A.: Volcanic eruptions and climate, Reviews of Geophysics, 38, 191–219, https://doi.org/https://doi.org/10.1029/1998RG000054, 2000.

Sagan, C.; Pollack, J.B., 1967: Anisotropic nonconservative scattering and the clouds of Venus. *J. Geophys. Res.*, 72, 469–477.

Stenchikov, G., Delworth, T. L., Ramaswamy, V., Stouffer, R. J., Wittenberg, A., and Zeng, F.: Volcanic signals in oceans, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/https://doi.org/10.1029/2008JD011673, 2009.

Stockdale, T. N., Molteni, F. and Ferranti, L., 2015: Atmospheric initial conditions and the predictability of the Arctic Oscillation. *Geophys. Res. Lett.*, 42: 1173–1179. doi: 10.1002/2014GL062681.

Swingedouw, D., Mignot, J., Ortega, P., Khodri, M., Menegoz, M., Cassou, C., & Hanquiez, V.: Impact of explosive volcanic eruptions on the main climate variability modes. Global and Planetary Change, 150, 24– 45, 2017.

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 climatology: 1979–2016, Earth System Science Data, 10, 469–492, https://doi.org/10.5194/essd-10-469-2018, 2018.

Timmreck, C., Pohlmann, H., Illing, S., and Kadow, C.: The impact of stratospheric volcanic aerosol on decadal-scale climate predictions, Geophysical Research Letters, 43, 834–842, https://doi.org/https://doi.org/10.1002/2015GL067431, 2016.

Toohey, M., Stevens, B., Schmidt, H., & Timmreck, C. (2016). Easy Volcanic Aerosol (EVA v1.0): An idealized forcing generator for climate simulations. Geoscientific Model Development, 2016, 1– 40.

Wu, X., Yeager, S. G., Deser, C., Rosenbloom, N., and Meehl, G. A.: Volcanic forcing degrades multiyear-to-decadal prediction skill in the tropical Pacific, Science Advances, 9, eadd9364, https://doi.org/10.1126/sciadv.add9364, 2023.

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., (2016). The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): experimental design and forcing input data for CMIP6, Geoscientific Model Development, 9, 2701–2719, https://doi.org/10.5194/gmd-9-2701-2016.

Document History

Version	Author(s)	Date	Changes
1	Roberto Bilbao (BSC), Tim Stockdale (ECMWF)	dd/mm/yyyy	

Internal Review History

Internal Reviewers	Date	Comments
Stephanie Johnson (ECMWF)	19/12/2023	
Constantin Ardilouze (Meteo France)	21/12/2023	Good for submission

Estimated Effort Contribution per Partner

Partner	Effort
BSC	14
ECMWF	3.0 months
Total	17