AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2016GL070587

Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty

Matthias Bittner[1,2], Hauke Schmidt[1], Claudia Timmreck[1], and Frank Sienz[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]International Max Planck Research School on Earth System Modelling (IMPRS), Hamburg, Germany

Key Points:
- A large ensemble of 15 members is needed to detect a 95% significant impact of the Pinatubo eruption on the NH polar vortex
- If only the two strongest eruptions since 1850 are considered, the CMIP5 models robustly simulate a strengthening of the NH polar vortex
- Including smaller eruptions in the sample reduces the signal-to-noise ratio of the simulated NH polar vortex response

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Bittner,
matthias.bittner@mpimet.mpg.de

Citation:
Bittner, M., H. Schmidt, C. Timmreck, and F. Sienz (2016), Using a large ensemble of simulations to assess the Northern Hemisphere stratospheric dynamical response to tropical volcanic eruptions and its uncertainty, *Geophys. Res. Lett.*, *43*, 9324–9332, doi:10.1002/2016GL070587.

Received 10 MAY 2016
Accepted 23 AUG 2016
Accepted article online 26 AUG 2016
Published online 14 SEP 2016

Abstract The observed strengthening of the Northern Hemisphere (NH) polar vortex after tropical volcanic eruptions appears to be underestimated by coupled climate models. However, there are only a limited number of observed eruptions, which makes the attribution of volcanic signals difficult, because the polar vortex is also influenced by other external forcing factors as well as internal variability. We show with a 100-member ensemble of historical (1850–2005) simulations with the Max Planck Institute Earth System Model that an ensemble larger than what is provided by the Coupled Model Intercomparison Project Phase 5 (CMIP5) models is needed to detect a statistically significant NH polar vortex strengthening. The most robust signal can be found when only the two strongest eruptions (Krakatau and Pinatubo) are considered in contrast to including smaller eruptions to increase the sample size. For these two strongest eruptions, the mean of 15 CMIP5 models shows a statistically significant strengthening of the NH polar vortex as well.

1. Introduction

By injecting large amounts of sulfur dioxide (SO_2) into the stratosphere, explosive tropical volcanic eruptions cannot only impact Earth's radiation budget but also change atmospheric dynamics [e.g., *Robock*, 2000; *Timmreck*, 2012, and references therein]. Observational records indicate a strengthening of the climatological westerlies of the Northern Hemisphere (NH) stratospheric polar vortex in posteruption boreal winters [e.g., *Kodera*, 1995; *Graf et al.*, 2007]. However, the simulated vortex response to volcanic eruptions in the "historical" experiments (1850–2005) of the Coupled Model Intercomparison Project 5 (CMIP5) [*Taylor et al.*, 2011] seems to be small in comparison with observations and to a large extent statistically insignificant [*Driscoll et al.*, 2012; *Charlton-Perez et al.*, 2013].

The observed NH polar vortex response to tropical volcanic eruptions is uncertain due to the limited number of satellite observations, which are only available after the eruptions of El Chichón in 1982 and Pinatubo in 1991. For eruptions prior to the satellite era, the significant shift of the North Atlantic Oscillation in the posteruption winter [*Christiansen*, 2008] is, in general, assumed to indicate a stronger stratospheric polar vortex, but the magnitude of the strengthening is uncertain. This uncertainty is reflected in the large model spread obtained in simulations of the NH polar vortex response to tropical volcanic eruptions [*Charlton-Perez et al.*, 2013]. The uncertainty is due to several sources: forcing uncertainty, spread of the forcings, model uncertainty, and internal variability. The forcing uncertainty is partly due to incomplete knowledge of the aerosol loading and distribution after eruptions before but also during the satellite era [*Stenchikov et al.*, 1998; *Arfeuille et al.*, 2014]. Small differences in the volcanic forcing, and hence in the space-time structure of the temperature anomaly in the lower tropical stratosphere, can have impacts on the response of the NH polar vortex [*Toohey et al.*, 2014]. The spread of the forcings is a source of uncertainty if different eruptions are averaged to increase the sample size. Model uncertainty emerges because different models may simulate different responses to the same volcanic forcing. Internal variability arises in the atmosphere, ocean, or the coupled system in the absence of any external forcing and occurs because of coupled nonlinear dynamical processes [*Madden*, 1976; *Feldstein*, 2000]. The NH polar stratosphere is a region prone to large variability on interannual, interseasonal, and weekly time scales [*Waugh and Polvani*, 2010]. The internal interannual variability in the polar stratosphere arises both from nonlinearities within the polar stratosphere itself [*Holton and Mass*, 1976; *Scott and Polvani*, 2006] as well as from variations remote from the polar stratosphere, such as the quasi-biennial oscillation (QBO) [*Labitzke and Van Loon,* 1988; *Anstey and Shepherd*, 2013] or the El Niño–Southern Oscillation (ENSO) [*van Loon and Labitzke*, 1987; *Sassi et al.*, 2004; *Manzini et al.*, 2006].

An assessment of the response to volcanic eruptions in CMIP5 models has been conducted by *Driscoll et al.* [2012]. The underlying assumption is that the multimodel mean potentially shows a more reliable response to the volcanic forcing compared to a single model because of the reduction of model uncertainty and internal variability (see *Tebaldi and Knutti* [2007] for the climate change signal). Averaging over the nine strongest tropical eruptions, *Driscoll et al.* [2012] conclude that climate models fail to reproduce the observed posteruption dynamical changes, which highlights that even when combining different eruptions to increase the sample size, a substantial uncertainty in the response remains.

Additionally, individual models provide only a small number of ensemble members and due to different forcing sets used by the models, it is difficult to quantify the contribution of internal variability, model, and forcing uncertainty to the total uncertainty. *Pausata et al.* [2015] investigated with a large ensemble the influence of a northern high-latitude eruption on the surface temperature and precipitation. They found that an ensemble of 20–40 members is necessary to discern the volcanic temperature signal from the internal variability on regional scale. However, *Pausata et al.* [2015] focused on a single eruption in an idealized setting. Here we use a 100-member ensemble of transient simulations by the Max Planck Institute Earth System Model in its Low Resolution configuration (MPI-ESM-LR) to provide an estimate of the contribution of internal variability. Similar to *Pausata et al.* [2015] and *Deser et al.* [2010], who investigated the signal-to-noise ratio for the climate change signal, we address the question: "How many ensemble members are needed to identify a significant response to volcanic eruptions?" and focus on the NH polar vortex. Furthermore, we use 15 CMIP5 models to investigate whether or not the practice of averaging over many eruptions is optimal to obtain a robust response of the NH polar vortex.

2. Methods

We use a 100-member ensemble of historical simulations from 1850 to 2005 of the MPI-ESM-LR in an updated version of the same model used in CMIP5 [*Giorgetta et al*., 2013], which is evaluated by numerous studies [e.g., *Reichler et al*., 2012; *Cattiaux and Cassou*, 2013; *Charlton-Perez et al*., 2013]. The MPI-ESM-LR is a fully coupled atmosphere-ocean model [*Jungclaus et al*., 2013; *Stevens et al*., 2013] including land and ocean biogeochemistry processes [*Ilyina et al*., 2013; *Reick et al*., 2013] with a spectral horizontal resolution of the atmosphere of T63 (approximately 1.9° × 1.9° at the equator) and 47 vertical layers. At this vertical resolution, the MPI-ESM-LR does not simulate a QBO [*Schmidt et al*., 2013]. Individual ensemble members only differ in their initial conditions. Anomalies after volcanic eruptions are defined as the difference between the posteruption winter and a pre-eruption reference period. Details about the model version, the simulations, and the eruptions are given in the supporting information.

The minimum number of ensemble members needed to identify a significant response is determined based on a resampling method, explained in detail in the supporting information. In short, we calculate the volcanic anomalies dependent on the ensemble size by randomly drawing 1–100 posteruption winters and reference periods and compute the difference with a paired test design [*von Storch*, 1999; *Wilks*, 2011]. By repeating this procedure 5,000 times, we obtain a mean anomaly and a distribution around the mean dependent on the ensemble size. The minimum number of ensemble members needed to identify a significant volcanic response is defined as the ensemble size at which the averaged anomaly is different from zero at the 95% (or 99%) confidence level. For the multimodel mean response to volcanic eruptions, we use the historical simulations of 15 CMIP5 models [*Sato et al*., 1993; *Ammann*, 2003; *Tanaka et al*., 2003; *Ammann et al*., 2007; *Rotstayn et al*., 2010; *Watanabe et al*., 2011; *Chylek et al*., 2011; *Collins et al*., 2011; *Donner et al*., 2011; *Gent et al*., 2011; *Voldoire et al*., 2012; *Yukimoto et al*., 2012; *Bentsen et al*., 2012; *Wu et al*., 2013; *Hurrell et al*., 2013; *Marsh et al*., 2013; *Schmidt et al*., 2014] (see supporting information).

3. Results

The response of the MPI-ESM-LR to volcanic eruption is analyzed first for the Pinatubo eruption in the first posteruption winter (December-January-February; DJF). The warming of about 4 K in the tropical lower stratosphere exceeds the 95% confidence level with two ensemble members, only (Figure 1, left). This is not surprising, because as MPI-ESM-LR does not simulate a QBO, the internal variability of the tropical stratosphere is small. The observed temperature anomaly is approximately 2–3 K [*Labitzke and McCormick*, 1992], hence smaller than the averaged simulated temperature anomaly. Differences to the observed temperature

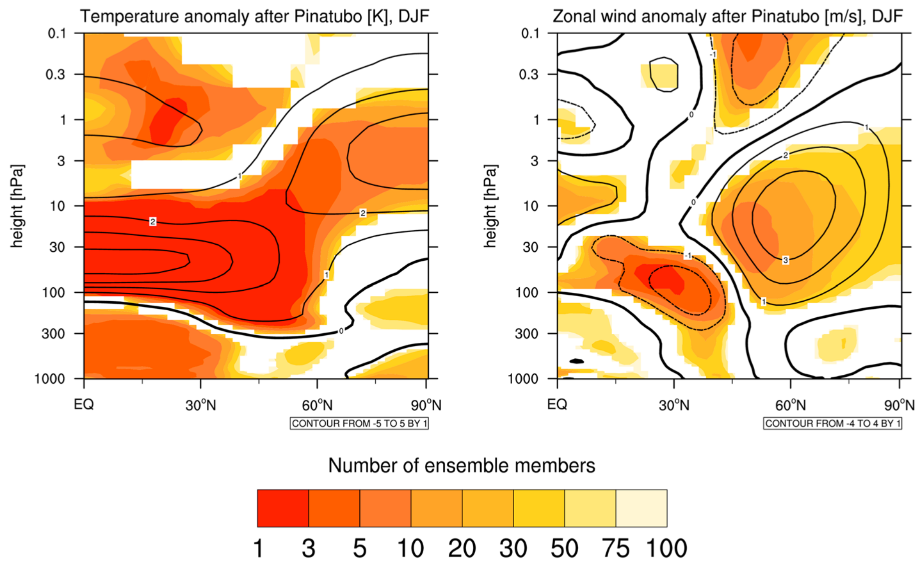

Figure 1. Ensemble average (left) zonal mean temperature and (right) zonal wind anomalies in the first post-Pinatubo DJF (contour lines; from −4.0 K to 4.0 K by 0.5 K and −5.0 m/s to 5.0 m/s by 1 m/s, respectively) , and minimum number of ensemble members needed to detect the anomaly significantly at the 95% confidence level (shading). White color indicates that the anomaly is not statistically significant at the 95% confidence level with 100 ensemble members.

are expected, because the model does not simulate a QBO and does not account for changes in the ozone concentration related to volcanic eruptions [*Cionni et al.*, 2011], which both can substantially influence the temperature perturbation [*Ramachandran et al.*, 2000; *Thomas et al.*, 2009b]. The cooling of the troposphere is significant at low latitudes with already 3–10 ensemble members.

The ensemble size which is needed to exceed the 95% confidence threshold for the zonal mean zonal wind in the NH polar stratosphere crucially depends on the latitude (Figure 1, right). For the maximum zonal wind anomaly of 3–4 m/s at approximately 60°N and 10 hPa, 15–20 ensemble members are needed to detect a significant response. The simulated maximum anomaly is comparable with reanalysis data in the first post-Pinatubo winter [*Toohey et al.*, 2014], although it is located further northward and at lower altitude in the reanalysis data. Equatorward of the maximum anomaly, the number of ensemble members decreases to 7–15, whereas poleward up to 40 members are necessary for the same anomaly magnitude. This is due to the internal variability which is larger north of 60°N than south of 60°N in reanalysis data as well as in the MPI-ESM-LR [*Schmidt et al.*, 2013]. Still, the model is able to simulate a significantly strengthened polar vortex, albeit with an ensemble larger than what most CMIP5 models provide. The weakening of the subtropical jet due to low-latitude surface cooling and therefore a decreased tropospheric temperature gradient [*Rind et al.*, 1992; *Zanchettin et al.*, 2012] exceeds the 95% confidence level with 3–5 ensemble members.

We now focus on the NH polar vortex, defined as the DJF zonal mean zonal wind at 10 hPa, averaged between 55°N and 65°N [e.g., *Matthes et al.*, 2006; *Butler and Polvani*, 2011]. The mean response of the polar vortex to the Pinatubo eruption is a strengthening of 3–4 m/s independent of the ensemble size (Figure 2a). This is expected because the mean of 5000 resamples should not change significantly if only one ensemble member is drawn or if more ensemble members are drawn and averaged. The confidence intervals, however, change substantially with increasing ensemble size. Only for ensembles larger than 15 members, the positive polar vortex anomaly is significantly different from zero at the 95% confidence level. Almost twice the ensemble size is needed to exceed the 99% confidence level. Most of the CMIP5 models have an ensemble size of 3–5 members, and none has more than 10. Our findings for the MPI-ESM-LR indicate that small ensembles may show no significant response of the polar vortex to the Pinatubo eruption which can be interpreted as a failure, even though the model actually can reproduce a strengthening of the polar vortex but the large internal variability masks the forced response.

To obtain a significant response in observations as well as to circumvent the small ensemble size of simulations, most assessments of volcanic eruptions do not rely on a single or very few cases (like Pinatubo

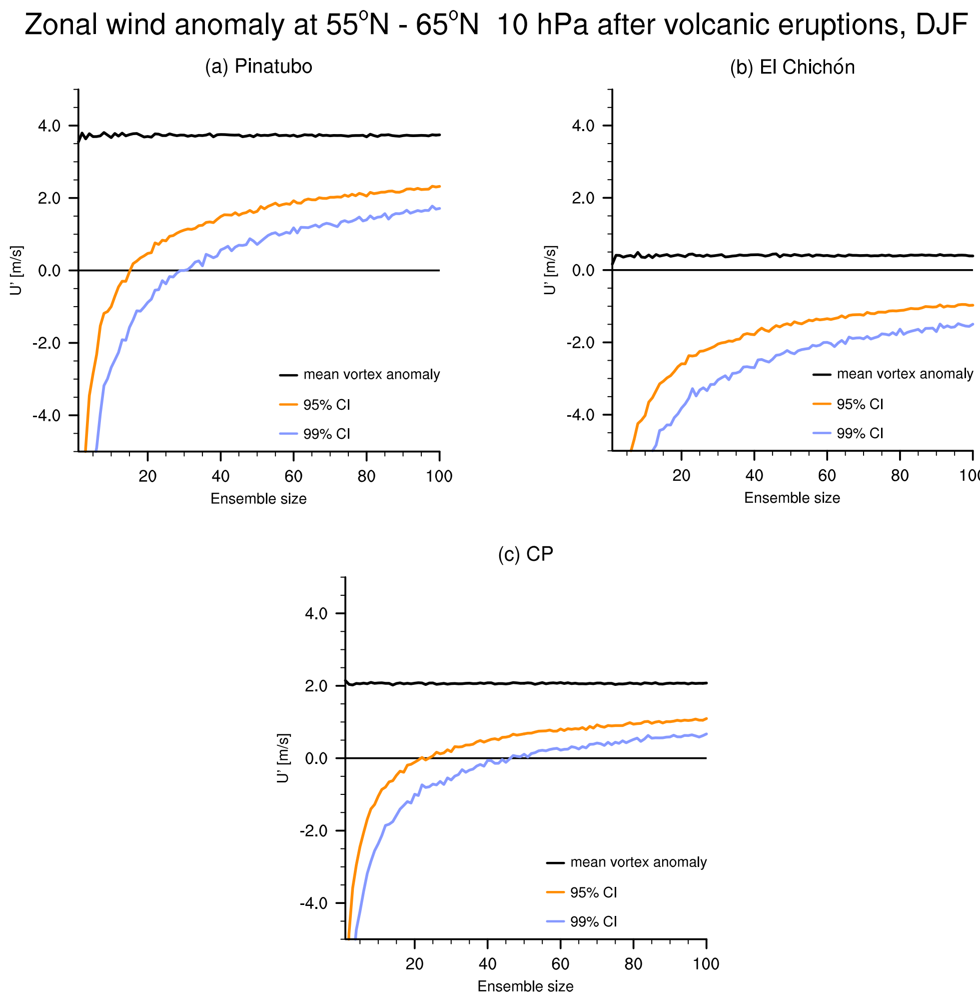

Figure 2. Ensemble average zonal mean zonal wind anomalies between 55°N and 65°N at 10 hPa in the first posteruption winter (DJF) (a) after the Pinatubo eruption, (b) after the El Chichón eruption, and (c) averaged over the Pinatubo and El Chichón eruptions (black), as well as the 95% (orange) and 99% (blue) confidence intervals dependent on the ensemble size.

and El Chichón) but average over several tropical volcanic eruptions [*Stenchikov et al.*, 2006; *Driscoll et al.*, 2012]. The benefit of a larger sample size comes with the caveat of introducing uncertainty due to the spread of the forcings. Moreover, the historical eruptions which are included are mostly of smaller magnitude than the Pinatubo eruption, and given the large internal variability of the polar vortex, a smaller forcing will be even more difficult to detect. Figure 2b shows that the mean response in the MPI-ESM-LR of the polar vortex after the 1982 El Chichón eruption, which is about one third the magnitude of Pinatubo in terms of SO_2 emission [*Bluth et al.*, 1997], is not significantly different from zero even with an ensemble of 100 members. Averaging El Chichón and Pinatubo posteruption values yields a mean polar vortex anomaly of approximately 2 m/s but only significantly different from zero at the 95% confidence level with more than 20 ensemble members (Figure 2c). Thus, doubling the sample size by adding a smaller eruption to the one of Pinatubo reduces the mean response as expected but also increases the ensemble size needed to identify a significant polar vortex response.

However, a strengthened polar vortex was observed after the El Chichón eruption, where the MPI-ESM-LR does not simulate a significant positive anomaly (Figure 2b). Missing and/or incorrect processes in the model, for example, interactive stratospheric chemistry, the QBO, or uncertainties in the prescribed volcanic forcing, could contribute to a too weak response of the polar vortex, which is an important question for future research. But given the large internal variability of the NH polar stratosphere, it is impossible to attribute the observed polar vortex anomaly after El Chichón to the volcanic forcing alone. The eruptions of El Chichón and Pinatubo both

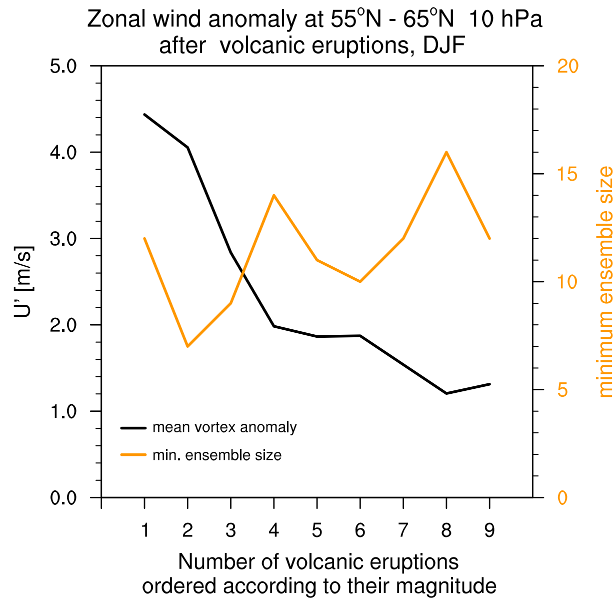

Figure 3. Ensemble average zonal mean zonal wind anomalies between 55°N and 65°N at 10 hPa in the first posteruption DJF averaged over successively smaller eruptions as ranked in Table S1 (black) and minimum ensemble size necessary to detect a zonal mean zonal wind anomalies significantly different from zero at the 95% confidence interval (orange). The *x* axis shows the numbers of volcanic eruptions which are averaged to obtain the polar vortex anomaly.

occurred during an El Niño event, which has been shown to weaken the NH polar vortex [e.g., *van Loon and Labitzke*, 1987; *Manzini et al.*, 2006]. However, the combination of multiple forcings is not a linear combination of the responses to the single forcings [*Calvo et al.*, 2009], which makes it difficult to quantify the influence of the volcanic eruptions. Additionally, it has been shown in modeling studies that the initial climate state affects the response to volcanic eruption [*Thomas et al.*, 2009a; *Zanchettin et al.*, 2013]. *Pausata et al.* [2016] shows, for a high-latitude eruption, that the impact on ENSO strongly depends on the initial state of the tropical Pacific itself, which in turn can affect the polar vortex response. However, due to our large ensemble, we cover a large spectrum of possible states of the central Pacific, which is on average in a neutral state. Satellite observations of the polar vortex after volcanic eruptions are available only for El Chichón and Pinatubo. If one includes only two winters after these eruptions, the sample size of four posteruption winters is very likely too small to allow conclusive statements about the magnitude of the expected posteruption polar vortex anomaly.

It is possible that the polar vortex in the real atmosphere is more susceptible to volcanic forcing than the climate models. Testing this hypothesis is difficult, because of the limited number of volcanic eruptions with well observed stratospheric responses, and because all eruptions will differ among each other not only in their magnitude, but also in their exact location, their eruption season, and their climatic initial conditions, which all might influence the polar vortex response [*Toohey et al.*, 2011; *Timmreck*, 2012]. With the limited number of observations, it is impossible to conclude that the model underestimates the dynamical response of the NH polar vortex. The polar vortex response in the first two winters following the El Chichón and Pinatubo eruptions in ERA-Interim reanalysis data [*Dee et al.*, 2011] is well within the ±1 standard deviation of the simulated response averaged over 100 ensemble members (supporting information).

Increasing the sample size reduces the uncertainty of the mean polar vortex signal, but the mean signal itself is reduced as smaller eruptions are included in the sample. This poses the question whether there is an "ideal" sample of eruptions for which the volcanic signal can be detected most robustly. When only the strongest eruption over the "historical" period (Krakatau) is considered, 12 ensemble members are necessary to identify the mean polar vortex anomaly of approximately 4.5 m/s (Figure 3). Adding the Pinatubo eruption to the sample only slightly reduces the mean polar vortex anomaly, but because the sample size is doubled the minimum ensemble size decreases to 7 members. Adding smaller eruptions, in general, further weakens the polar vortex anomaly, but because the eruptions have a much weaker impact on the polar vortex, the minimum ensemble size is always larger compared to the average of Krakatau and Pinatubo.

An ensemble size of 12 members, which is needed in the MPI-ESM-LR when all nine tropical eruptions are considered, is larger than what is available for any of the CMIP5 models. Of course, the exact ensemble size given here is valid only for the MPI-ESM-LR. Other climate models with a different representation of the polar vortex variability and strength as well as a potentially different sensitivity to volcanic forcing may yield a different number of ensemble members. However, the findings presented for the MPI-ESM-LR display that

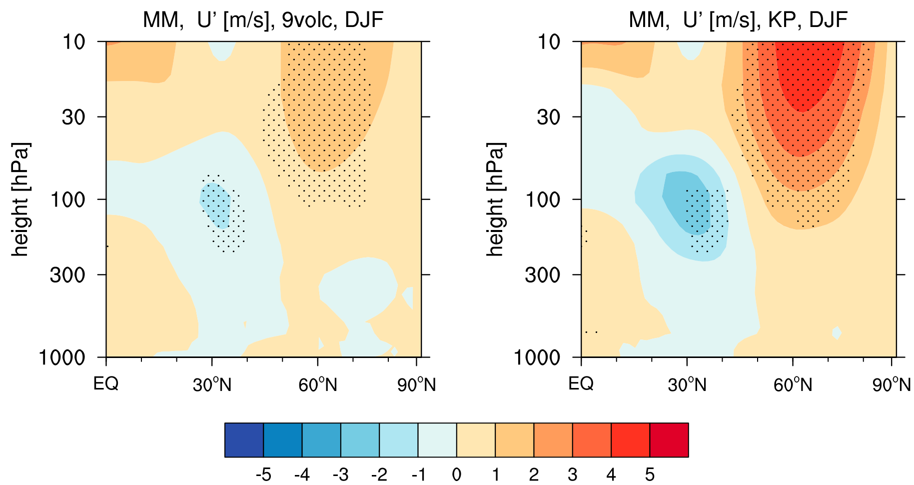

Figure 4. Multimodel mean of the zonal mean zonal wind anomaly in the first winter (left) after the nine strongest tropical eruptions since 1880 and (right) after the two strongest eruptions since 1880—Krakatau and Pinatubo. Stippling indicates where at least 14 of 15 models agree on the sign of the anomaly.

results gained from small ensembles may lead to false conclusions about the inability of coupled climate models to represent the dynamical response to volcanic eruptions.

To underpin this statement, we analyzed the dynamical response of the zonal mean zonal wind to tropical volcanic eruptions of 15 CMIP5 models. Based on the finding that the most robust response of the NH polar vortex to volcanic eruptions is achieved when only the strongest eruptions, Krakatau and Pinatubo, are considered, we show the zonal mean zonal wind anomaly averaged over these eruptions as well as for the nine strongest eruption in the historical simulations as used by *Driscoll et al.* [2012] (Figure 4). A positive wind anomaly in the NH polar vortex region, indicating a strengthening of the climatological westerlies, can be seen even when the nine strongest eruptions are considered. However, the magnitude of the anomaly is of about 1–2 m/s, only, which is much smaller compared to the observed strengthening of the polar vortex after the eruptions of El Chichón and Pinatubo. Still, the models show a consistent behavior with 14 of the 15 models agreeing on the sign of the anomaly. When only the two strongest eruptions are considered, the magnitude of the positive polar vortex anomaly is considerably larger with 4–5 m/s and all models agree on the sign. Hence, a larger and more robust response can be obtained when smaller eruptions are neglected and potentially if all models use identical forcing fields as proposed for the volcanic forcing model intercomparison project (VolMIP) [*Zanchettin et al.*, 2016]. This questions the conclusions of the apparent failure of coupled climate models to reproduce the stratospheric dynamical response to volcanic eruptions drawn from composites which include smaller eruptions.

4. Summary and Conclusion

We used a 100-member ensemble of historical simulation with the MPI-ESM-LR to quantify the response to tropical volcanic eruptions and its uncertainty associated with internal interannual variability. The focus is on zonal wind changes particularly with regard to a NH polar vortex strengthening. It is shown that state-of-the-art climate models are able to simulate such a strengthening. This evidence, however, depends crucially on the number of ensemble members involved.

In contrast to, e.g., the tropical stratospheric temperature anomalies, a large number of ensemble members are required to detect NH polar vortex strengthening in the first post-Pinatubo winter. This number depends on the latitude considered and ranges from 7 at the southward flank of the maximum positive wind anomaly to more than 40 members at high latitudes. The maximum ensemble mean NH polar vortex anomaly of about 3–4 m/s is significantly different from zero for an ensemble larger than 15 members. Although it is likely that the exact number of necessary ensemble members is model dependent, a too small ensemble can potentially show no significant response to a volcanic eruption due to the large internal variability of the polar stratosphere. Adding smaller eruptions reduces the response of the polar vortex, so that a larger number of ensemble members are required for a statistically significant signal. Thus, increasing the sample size for the sake of better statistics does not necessarily improve the detectability compared to a single, strong eruption.

We cannot rule out the possibility of a too weak model response. The failure to simulate a vortex strengthening after the El Chichón eruption may hint at that, but the observed anomaly might also have occurred by chance. The necessary ensemble size to simulate a significant strengthening of the polar vortex in the MPI-ESM-LR will be influenced not only by the eruption magnitude but also by the interannual variability of the NH polar stratosphere and by the representation of relevant processes. In particular, the lack of the QBO in the MPI-ESM-LR might influence the results, because the QBO impacts the NH polar vortex and its variability [*Baldwin et al*., 2001; *Anstey and Shepherd*, 2013]. In the MPI-ESM-LR the interannual variability in DJF is reduced compared to a MPI-ESM model version which simulates a QBO [*Schmidt et al*., 2013]. Including the QBO can therefore potentially increase the necessary ensemble size. Moreover, the interannual variability of the NH polar stratosphere might change due to increased CO_2 levels and/or changing ozone concentrations. Hence, two eruptions with comparable magnitude might yield a different minimum ensemble size in different climatic background states.

The CMIP5 multimodel mean shows a robust strengthening of the NH polar vortex of about 4–5 m/s in the first winter after the Krakatau and Pinatubo eruptions. Even for the averaged response of the nine strongest eruptions, a strengthening of the polar vortex is simulated, but the magnitude is small. Hence, the coupled climate models do not completely fail to reproduce the observed dynamical response to tropical volcanic eruption, as previously stated [*Driscoll et al*., 2012]. However, the large internal variability of the NH polar stratosphere hampers the detection of a significant strengthening of the polar vortex when smaller tropical eruptions are included in the sample.

As nature offers only "one realization," the detectability of the volcanic signal in the real atmosphere is problematic, especially in regions with large internal variability such as the NH polar vortex. Attributing the polar vortex anomalies after the two available observed events of El Chichón and Pinatubo to the volcanic forcing is difficult, because the polar vortex is also influenced by other factors such as ENSO, the QBO, and solar irradiance. As demonstrated by the relatively large ensemble needed to identify the response to strong eruptions such as Pinatubo, one must be careful in evaluating the model response using the very limited observational record in the stratosphere after volcanic eruptions.

Acknowledgments
The research was supported by the German Ministry for Education and Research through the "MiKlip" program (FKZ: 01LP1130A (MB), 01LP1517B (CT), and 01LP1145A (FS)). Computations were carried out at the German Climate Computing Centre (DKRZ). Primary data and scripts used in the analysis and other supporting information that may be useful in reproducing the author's work are archived by the Max Planck Institute for Meteorology and can be obtained by contacting publications@mpimet.mpg.de

References

Ammann, C. M. (2003), A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate, *Geophys. Res. Lett.*, *30*(12), 1657, doi:10.1029/2003GL016875.

Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner, and R. A. Tomas (2007), Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model, *Proc. Natl. Acad. Sci. U.S.A.*, *104*, 3713–3718, doi:10.1073/pnas.0605064103.

Anstey, J. A., and T. G. Shepherd (2013), High-latitude influence of the quasi-biennial oscillation, *Q. J. R. Meteorol. Soc.*, *140*, 1–21, doi:10.1002/qj.2132.

Arfeuille, F., D. Weisenstein, H. MacK, E. Rozanov, T. Peter, and S. Brönnimann (2014), Volcanic forcing for climate modeling: A new microphysics-based data set covering years 1600–present, *Clim. Past*, *10*, 359–375, doi:10.5194/cp-10-359-2014.

Baldwin, M. P., et al. (2001), The quasi-biennial oscillation, *Rev. Geophys.*, *39*, 179–229, doi:10.1029/1999RG000073.

Bentsen, M., et al. (2012), The Norwegian Earth System Model, NorESM1-M—Part 1: Description and basic evaluation, *Geosci. Model Dev. Discuss.*, *5*(3), 2843–2931, doi:10.5194/gmdd-5-2843-2012.

Bluth, G. J. S., W. I. Rose, I. E. Sprod, and A. J. Krueger (1997), Stratospheric loading of sulfur from explosive volcanic eruptions, *J. Geol.*, *105*, 671–684, doi:10.1086/515972.

Butler, A. H., and L. M. Polvani (2011), El Niño, La Niña, and stratospheric sudden warmings: A reevaluation in light of the observational record, *Geophys. Res. Lett.*, *38*, L13807, doi:10.1029/2011GL048084.

Calvo, N., M. A. Giorgetta, R. Garcia-Herrera, and E. Manzini (2009), Nonlinearity of the combined warm ENSO and QBO effects on the Northern Hemisphere polar vortex in MAECHAM5 simulations, *J. Geophys. Res.*, *114*, D13109, doi:10.1029/2008JD011445.

Cattiaux, J., and C. Cassou (2013), Opposite CMIP3/CMIP5 trends in the wintertime Northern Annular Mode explained by combined local sea ice and remote tropical influences, *Geophys. Res. Lett.*, *40*, 3682–3687, doi:10.1002/grl.50643.

Charlton-Perez, A. J., et al. (2013), On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models, *J. Geophys. Res. Atmos.*, *118*, 2494–2505, doi:10.1002/jgrd.50125.

Christiansen, B. (2008), Volcanic eruptions, large-scale modes in the Northern Hemisphere, and the El Niño-Southern Oscillation, *J. Clim.*, *21*(5), 910–922, doi:10.1175/2007JCLI1657.1.

Chylek, P., J. Li, M. K. Dubey, M. Wang, and G. Lesins (2011), Observed and model simulated 20th century Arctic temperature variability: Canadian Earth System Model CanESM2, *Atmos. Chem. Phys. Discuss.*, *11*(8), 22,893–22,907, doi:10.5194/acpd-11-22893-2011.

Cionni, I., V. Eyring, J. F. Lamarque, W. J. Randel, D. S. Stevenson, F. Wu, G. E. Bodeker, T. G. Shepherd, D. T. Shindell, and D. W. Waugh (2011), Ozone database in support of CMIP5 simulations: Results and corresponding radiative forcing, *Atmos. Chem. Phys.*, *11*(21), 11,267–11,292, doi:10.5194/acp-11-11267-2011.

Collins, W., et al. (2011), Development and evaluation of an Earth-System model—HadGEM2, *Geosci. Model Dev.*, *4*, 1051–1075, doi:10.5194/gmd-4-1051-2011.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597, doi:10.1002/qj.828.

Deser, C., A. Phillips, V. Bourdette, and H. Teng (2010), Uncertainty in climate change projections: The role of internal variability, *Clim. Dyn.*, *38*(3–4), 527–546, doi:10.1007/s00382-010-0977-x.

Donner, L. J., et al. (2011), The Dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3, *J. Clim.*, *24*(13), 3484–3519, doi:10.1175/2011JCLI3955.1.

Driscoll, S., A. Bozzo, L. J. Gray, A. Robock, and G. Stenchikov (2012), Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions, *J. Geophys. Res.*, *117*, D17105, doi:10.1029/2012JD017607.

Feldstein, S. B. (2000), The timescale, power spectra, and climate noise properties of teleconnection patterns, *J. Clim.*, *13*(24), 4430–4440, doi:10.1175/1520-0442(2000)013<4430:TTPSAC>2.0.CO;2.

Gent, P. R., et al. (2011), The Community Climate System Model Version 4, *J. Clim.*, *24*(19), 4973–4991, doi:10.1175/2011JCLI4083.1.

Giorgetta, M. A., et al. (2013), Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5, *J. Adv. Model. Earth Syst.*, *5*, 572–597, doi:10.1002/jame.20038.

Graf, H.-F., Q. Li, and M. A. Giorgetta (2007), Volcanic effects on climate: Revisiting the mechanisms, *Atmos. Chem. Phys.*, *7*(17), 4503–4511, doi:10.5194/acp-7-4503-2007.

Holton, J. R., and C. Mass (1976), Stratospheric vacillation cycles, *J. Atmos. Sci.*, *33*(11), 2218–2225, doi:10.1175/1520-0469(1976)033<2218:SVC>2.0.CO;2.

Hurrell, J. W., et al. (2013), The Community Earth System Model: A framework for collaborative research, *Bull. Am. Meteorol. Soc.*, *94*(9), 1339–1360, doi:10.1175/BAMS-D-12-00121.1.

Ilyina, T., K. D. Six, J. Segschneider, E. Maier-Reimer, H. Li, and I. Núñez-Riboni (2013), Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, *J. Adv. Model. Earth Syst.*, *5*(2), 287–315, doi:10.1029/2012MS000178.

Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. von Storch (2013), Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, *J. Adv. Model. Earth Syst.*, *5*(2), 422–446, doi:10.1002/jame.20023.

Kodera, K. (1995), On the origin and nature of the interannual variability of the winter stratospheric circulation in the northern hemisphere, *J. Geophys. Res.*, *100*, 14,077–14,087, doi:10.1029/95JD01172.

Labitzke, K., and M. P. McCormick (1992), Stratospheric temperature increases due to Pinatubo aerosols, *Geophys. Res. Lett.*, *19*, 207–210, doi:10.1029/91GL02940.

Labitzke, K., and H. Van Loon (1988), Associations between the 11-year solar cycle, the QBO and the atmosphere. Part I: The troposphere and stratosphere in the northern hemisphere in winter, *J. Atmos. Sol. Terr. Phys.*, *50*(3), 197–206, doi:10.1016/0021-9169(88)90068-2.

Madden, R. A. (1976), Estimates of the natural variability of time-averaged sea-level pressure, *Mon. Weather Rev.*, *104*(7), 942–952, doi:10.1175/1520-0493(1976)104<0942:EOTNVO>2.0.CO;2.

Manzini, E., M. A. Giorgetta, M. Esch, L. Kornblueh, and E. Roeckner (2006), The influence of sea surface temperatures on the Northern Winter Stratosphere: Ensemble simulations with the MAECHAM5 Model, *J. Clim.*, *19*(16), 3863–3881, doi:10.1175/JCLI3826.1.

Marsh, D. R., M. J. Mills, D. E. Kinnison, J.-F. Lamarque, N. Calvo, and L. M. Polvani (2013), Climate change from 1850 to 2005 simulated in CESM1(WACCM), *J. Clim.*, *26*(19), 7372–7391, doi:10.1175/JCLI-D-12-00558.1.

Matthes, K., Y. Kuroda, K. Kodera, and U. Langematz (2006), Transfer of the solar signal from the stratosphere to the troposphere: Northern winter, *J. Geophys. Res.*, *111*, D06108, doi:10.1029/2005JD006283.

Pausata, F. S. R., A. Grini, R. Caballero, A. Hannachi, and Ø. Seland (2015), High-latitude volcanic eruptions in the Norwegian Earth System Model: The effect of different initial conditions and of the ensemble size, *Tellus B*, *67*, doi:10.3402/tellusb.v67.26728.

Pausata, F. S. R., C. Karamperidou, R. Caballero, and D. S. Battisti (2016), ENSO response to high-latitude volcanic eruptions in the Northern Hemisphere: The role of the initial conditions, *Geophys. Res. Lett.*, *43*, doi:10.1002/2016GL069575.

Ramachandran, S., V. Ramaswamy, G. L. Stenchikov, and A. Robock (2000), Radiative impact of the Mount Pinatubo volcanic eruption: Lower stratospheric response, *J. Geophys. Res.*, *105*, 24,409–24,429, doi:10.1029/2000JD900355.

Reichler, T., J. Kim, E. Manzini, and J. Kroger (2012), A stratospheric connection to Atlantic climate variability, *Nat. Geosci.*, *5*(11), 783–787, doi:10.1038/ngeo1586.

Reick, C. H., T. Raddatz, V. Brovkin, and V. Gayler (2013), Representation of natural and anthropogenic land cover change in MPI-ESM, *J. Adv. Model. Earth Syst.*, *5*(3), 459–482, doi:10.1002/jame.20022.

Rind, D., N. K. Balachandran, and R. Suozzo (1992), Climate change and the middle atmosphere. Part II: The impact of volcanic aerosols, *J. Clim.*, *5*, 189–208, doi:10.1175/1520-0442(1992)005<0189:CCATMA>2.0.CO;2.

Robock, A. (2000), Volcanic eruptions and climate, *Rev. Geophys.*, *38*, 191–219, doi:10.1029/1998RG000054.

Rotstayn, L. D., M. A. Collier, M. R. Dix, Y. Feng, H. B. Gordon, S. P. O'Farrell, I. N. Smith, and J. Syktus (2010), Improved simulation of Australian climate and ENSO-related rainfall variability in a global climate model with an interactive aerosol treatment, *Int. J. Climatol.*, *30*(7), 1067–1088, doi:10.1002/joc.1952.

Sassi, F., D. Kinnison, B. A. Boville, R. R. Garcia, and R. Roble (2004), Effect of El Niño–Southern Oscillation on the dynamical, thermal, and chemical structure of the middle atmosphere, *J. Geophys. Res.*, *109*, D17108, doi:10.1029/2003JD004434.

Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack (1993), Stratospheric aerosol optical depths, 1850–1990, *J. Geophys. Res.*, *98*, 22,987–22,994, doi:10.1029/93JD02553.

Schmidt, G. A., et al. (2014), Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, *J. Adv. Model. Earth Syst.*, *6*(1), 141–184, doi:10.1002/2013MS000265.

Schmidt, H., et al. (2013), Response of the middle atmosphere to anthropogenic and natural forcings in the CMIP5 simulations with the Max Planck Institute Earth system model, *J. Adv. Model. Earth Syst.*, *5*(1), 98–116, doi:10.1002/jame.20014.

Scott, R. K., and L. M. Polvani (2006), Internal variability of the winter stratosphere. Part I: Time-independent forcing, *J. Atmos. Sci.*, *63*(11), 2758–2776, doi:10.1175/JAS3797.1.

Stenchikov, G., K. Hamilton, R. J. Stouffer, A. Robock, V. Ramaswamy, B. Santer, and H.-F. Graf (2006), Arctic Oscillation response to volcanic eruptions in the IPCC AR4 climate models, *J. Geophys. Res.*, *111*, D07107, doi:10.1029/2005JD006286.

Stenchikov, G. L., I. Kirchner, A. Robock, H.-F. Graf, J. C. Antuña, R. G. Grainger, A. Lambert, and L. Thomason (1998), Radiative forcing from the 1991 Mount Pinatubo volcanic eruption, *J. Geophys. Res.*, *103*, 13,837–13,857, doi:10.1029/98JD00693.

Stevens, B., et al. (2013), Atmospheric component of the MPI-M Earth System Model: ECHAM6, *J. Adv. Model. Earth Syst.*, *5*(2), 146–172, doi:10.1002/jame.20015.

Tanaka, T. Y., K. Orito, T. T. Sekiyama, K. Shibata, M. Chiba, and H. Tanaka (2003), MASINGAR, a global tropospheric aerosol chemical transport model coupled with MRI/JMA98 GCM, *Pap. Meteorol. Geophys.*, *53*(4), 119–138, doi:10.2467/mripapers.53.119.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2011), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., and R. Knutti (2007), The use of the multi-model ensemble in probabilistic climate projections, *Philos. Trans. Math. Phys. Eng. Sci.*, *365*(1857), 2053–2075, doi:10.1098/rsta.2007.2076.

Thomas, M. A., C. Timmreck, M. A. Giorgetta, H.-F. Graf, and G. Stenchikov (2009a), Simulation of the climate impact of Mt. Pinatubo eruption using ECHAM5—Part 1: Sensitivity to the modes of atmospheric circulation and boundary conditions, *Atmos. Chem. Phys.*, *9*(2), 757–769, doi:10.5194/acp-9-757-2009.

Thomas, M. A., M. A. Giorgetta, C. Timmreck, H.-F. Graf, and G. Stenchikov (2009b), Simulation of the climate impact of Mt. Pinatubo eruption using ECHAM5—Part 2: Sensitivity to the phase of the QBO and ENSO, *Atmos. Chem. Phys.*, *9*(9), 3001–3009, doi:10.5194/acp-9-3001-2009.

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

Toohey, M., K. Krüger, U. Niemeier, and C. Timmreck (2011), The influence of eruption season on the global aerosol evolution and radiative impact of tropical volcanic eruptions, *Atmos. Chem. Phys.*, *11*, 12,352–12,367, doi:10.5194/acp-11-12351-2011.

Toohey, M., K. Krüger, M. Bittner, C. Timmreck, and H. Schmidt (2014), The impact of volcanic aerosol on the Northern Hemisphere stratospheric polar vortex: Mechanisms and sensitivity to forcing structure, *Atmos. Chem. Phys.*, *14*(11), 13,063–13,079, doi:10.5194/acp-14-13063-2014.

van Loon, H., and K. Labitzke (1987), The Southern Oscillation. Part V: The anomalies in the lower stratosphere of the northern hemisphere in winter and a comparison with the quasi-biennial oscillation, *Mon. Weather Rev.*, *115*(2), 357–369, doi:10.1175/1520-0493(1987)115%3C0357:TSOPVT%3E2.0.CO;2.

Voldoire, A., et al. (2012), The CNRM-CM5.1 global climate model: Description and basic evaluation, *Clim. Dyn.*, *40*(9–10), 2091–2121, doi:10.1007/s00382-011-1259-y.

von Storch, H. (1999), Misuses of statistical analysis in climate research, *Anal. Clim. Var. Appl. Stat. Tech.*, 11–26, doi:10.1007/978-3-662-03744-7_2.

Watanabe, S., et al. (2011), MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments, *Geosci. Model Dev.*, *4*(4), 845–872, doi:10.5194/gmd-4-845-2011.

Waugh, D. W., and L. M. Polvani (2010), *The Stratosphere: Dynamics, Transport, and Chemistry*, *Stratos. Dyn. Transp. Chem.*, edited by L. M. Polvani, A. H. Sobel, and D. W. Waugh, 190, 43–57, doi:10.1029/GM190.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, *Int. Geophys*., 3rd ed., Elsevier, Amsterdam, Neth.

Wu, T., et al. (2013), Global carbon budgets simulated by the Beijing climate center climate system model for the last century, *J. Geophys. Res. Atmos.*, *118*, 4326–4347, doi:10.1002/jgrd.50320.

Yukimoto, S., et al. (2012), A new global climate model of the Meteorological Research Institute: MRI-CGCM3 model description and basic performance, *J. Meteorol. Soc. Japan*, *90A*, 23–64, doi:10.2151/jmsj.2012-A02.

Zanchettin, D., C. Timmreck, H.-F. Graf, A. Rubino, S. Lorenz, K. Lohmann, K. Krüger, and J. H. Jungclaus (2012), Bi-decadal variability excited in the coupled ocean-atmosphere system by strong tropical volcanic eruptions, *Clim. Dyn.*, *39*(1–2), 419–444, doi:10.1007/s00382-011-1167-1.

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

Zanchettin, D., et al. (2016), The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): Experimental design and forcing input data for CMIP6, *Geosci. Model Dev.*, *9*, 2701–2719, doi:10.5194/gmd-9-2701-2016.