AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2015GL067431

Key Points:
- Assessment of the potential gain of skill due to volcanic forcing over the CMIP5 hindcast period
- The system has no prediction skill in global mean TAS if external forcing is neglected
- Significant multiyear seasonal temperature prediction skill due to volcanic aerosols over Europe

Supporting Information:
- Supporting Information S1

Correspondence to:
C. Timmreck,
claudia.timmreck@mpimet.mpg.de

Citation:
Timmreck, C., H. Pohlmann, S. Illing, and C. Kadow (2016), The impact of stratospheric volcanic aerosol on decadal-scale climate predictions, *Geophys. Res. Lett.*, *43*, 834–842, doi:10.1002/2015GL067431.

Received 16 DEC 2015
Accepted 20 DEC 2015
Accepted article online 22 DEC 2015
Published online 19 JAN 2016

The impact of stratospheric volcanic aerosol on decadal-scale climate predictions

Claudia Timmreck[1], Holger Pohlmann[1], Sebastian Illing[2], and Christopher Kadow[2]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]Freie Universität Berlin, Berlin, Germany

Abstract To understand the impact of volcanic aerosol on multiyear seasonal and decadal climate predictions, we performed Coupled Model Intercomparison Project Phase 5-type hindcasts without volcanic aerosol using the German Mittelfristige Klimaprognosen prediction system and compared them to the corresponding simulations including aerosols. Our results show that volcanic aerosol significantly affects the prediction skill for global mean surface air temperature in the first five years after strong volcanic eruptions. Also, on the regional scale a volcanic imprint on decadal-scale variability is detectable. Neglecting volcanic aerosol leads to a reduced prediction skill over the tropical and subtropical Atlantic, Indic, and west Pacific but to an improvement over the tropical east Pacific, where the model has in general no skill. Multiseasonal differences in the skill for seasonal mean temperatures are evident over Continental Europe with significant skill loss due to neglection of volcanic aerosol in boreal winter over central Europe, Scandinavia and over southeastern Europe, and the East Mediterranean in boreal summer.

1. Introduction/Motivation

The need for near-term climate information has fostered decadal climate prediction research [e.g., *Smith et al.*, 2007; *Pohlmann et al.*, 2009; *Meehl et al.*, 2013]. The skill of climate predictions on the time scale up to a decade is in general evaluated using hindcasts, for example, as part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) [*Taylor et al.*, 2012], e.g., *Kim et al.* [2012] and *Doblas-Reyes et al.* [2013]. In these simulations external forcing is assumed to be known over the full simulated decade. For a real decadal forecast this assumption is obviously invalid for unforeseeable abrupt forcing changes as in the case of large volcanic eruptions. These cannot be predicted beforehand, but eventually they will happen, with the potential to affect global and regional climates on short-term and long-term time scales. Hence, they are a potential source of uncertainty in decadal predictions [*Shiogama et al.*, 2010; *Kirtman et al.*, 2013], but once they occur, the forecasts have to be repeated to estimate their impact.

Large tropical volcanic eruptions increase the stratospheric aerosol loading and thereby reduce the global average surface temperature and increase stratospheric temperatures. For instance, as a response to the large tropical Pinatubo eruption in 1991 global surface cooling with a maximum of about 0.4 K [*Thompson et al.*, 2009] and a stratospheric warming of 2–3 K [*Labitzke and McCormick*, 1992] was observed. While stratospheric temperatures recover within 2 years after the eruption, the regeneration of tropospheric temperatures is delayed due to the large thermal inertia of the ocean. Volcanic eruptions also affect tropospheric variability by increasing the likelihood of El Niño and positive Indian Ocean Dipole events [*Maher et al.*, 2015]. Furthermore, after large tropical volcanic eruptions a postvolcanic strengthening of the Northern Hemisphere (NH) polar vortex is observed which leads to a significantly positive North Atlantic Oscillation (NAO) index in the first winter [*Christiansen*, 2008]. Hence, this robust postvolcanic signal might imply an enhanced skill in multiyear seasonal mean prediction for NH winter climate if stratospheric aerosol is taken into account.

After the Pinatubo eruption, the stratosphere has not been affected by a large climate relevant volcanic eruption, but several small volcanic eruptions have affected the lower stratosphere [e.g., *Vernier et al.*, 2011]. Neglecting these stratospheric aerosol changes could possibly have contributed to the overestimation of projected global warming by climate models, compared to the observed global temperature record for the past decade [e.g., *Solomon et al.*, 2011].

The impact of volcanic aerosol on climate and on multiyear seasonal and decadal climate predictability is still largely unexploited [e.g., *Bellucci et al.*, 2015] and has been addressed by only a few studies before. *Collins* [2003] could demonstrate that volcanoes of the size of El Chichón and Pinatubo have the potential to significantly bias the probability density functions of mean sea level pressure and surface air temperature on large

continental-scale seasonal averages. A more robust signal is found for the larger Pinatubo size eruptions but in general care has to be taken as the volcanic disturbance can be strongly influenced or even obscured by internal variability [*Collins*, 2003]. *Shiogama et al.* [2010] studied with the MIROC3 climate model the potential impact of a volcanic eruption of Pinatubo size in 2010 on decadal-scale predictability. They found not only significant surface cooling but also a noticeable increase in global mean surface and upper ocean temperature variability. Furthermore hindcasts with the atmosphere, only Hadley Centre Global Environmental Model, version 1, model [*Marshall et al*., 2009] revealed for the first European postvolcanic winter a strong dependence on the stratospheric conditions in early winter.

The question remains how large will the uncertainty be for future decadal climate predictions if no volcanic aerosol is taken into account? And how strong has volcanic aerosol affected decadal prediction skill on annual and multiyear seasonal scales over the CMIP5 hindcast period? In this paper we quantify the potential gain of skill due to the consideration of volcanic aerosol between 1962 and 2004 by contrasting hindcasts with the Mittelfristige Klimaprognosen (MiKlip) prediction system baseline 1 [*Pohlmann et al*., 2013] with observed aerosol concentrations as performed in CMIP5 with simulations without volcanic aerosol. The prediction skill for both experiments is compared for surface temperatures for yearly and multiyearly averages. With respect to the pronounced seasonality of the postvolcanic response, we discuss multiyear seasonal means for winter and summer seasons to account for the seasonal dependence of the skill.

2. Experiment

We carried out hindcasts with the baseline 1 (b1) version of the MiKlip prediction system which is based on the coupled Max-Planck-Institut-Earth system model (MPI-ESM) [*Giorgetta et al*., 2013; *Stevens et al*., 2013; *Jungclaus et al*., 2013]. The MPI-ESM is an Earth system model, with atmosphere, ocean, and dynamic vegetation components. As reference simulation the decadal hindcasts carried out with the baseline 1-low resolution (LR) version (b1-LR) [*Pohlmann et al*., 2013] are used. Therefore, the "low resolution" (LR) model configuration of the MPI-ESM is applied (LR: atmosphere: T63L47 and ocean: 1.5°L40). b1-LR has an improved initialization and a higher skill performance over the tropical ocean compared to the previous version used for CMIP5. b1-LR simulations have prediction skill for global surface air temperature with the exception of the eastern Pacific [*Kadow et al*., 2015].

The decadal hindcasts with no volcanic aerosol (b1-NVA) are initialized in the same way as b1-LR for the period 1961–1999. Three ensemble members are calculated for each start year, with restart conditions for the first, the second, and the third of January (lagged-day initialization). The ocean is initialized with salinity and temperature anomalies from the ORAS4 (ocean reanalysis system) [*Balmaseda et al*., 2012]. Atmospheric temperature, divergence, vorticity, and surface pressure 3-D fields are full field initialized with ERA-40 reanalysis data until 1989 and ERA-Interim from 1990 onward [*Uppala et al*., 2005; *Dee et al*., 2011].

The volcanic forcing data set which is used in the b1-LR reference simulations and neglected in b1-NVA is the same as used in the MPI-ESM simulations contributing to CMIP5 and is based on *Sato et al.* [1993] and *Stenchikov et al.* [1998] (Figure S1 in the supporting information). The data set encompasses zonally and monthly means averaged values of single scattering albedo, aerosol extinction, and asymmetry factor as a function of wavelength, pressure, and time [*Schmidt et al*., 2013]. For the years after 1999, the stratospheric aerosol is kept constant at negligible aerosol optical depth (AOD) (1.93e-5) although an increase in stratospheric aerosol has been observed since 2002 [e.g., *Vernier et al*., 2011]. The time series of monthly mean and globally averaged aerosol optical depth (AOD) for the year 1961 to 1999 shows that the last four decades are clearly dominated by large tropical volcanic eruptions (Figure S1). On the multidecadal average the highest AOD values are found in the tropical belt and at high latitudes and minimum values in the subtropics of both hemispheres.

For the hindcast analysis we apply the Murphy-Epstein decomposition and Continuous Ranked Probability Skill Score (MurCSS) tool [*Illing et al*., 2014] software package for standardized evaluation of decadal hindcast-prediction systems in a lead year approach [*Kadow et al*., 2015]. The prediction skill is analyzed in terms of the accuracy of the ensemble mean prediction compared to observations. We calculate the anomaly correlation coefficient (ACC) and the mean squared error skill score (MSESS) [e.g., *Jolliffe and Stephenson*, 2011; *Murphy*, 1988]. Both terms are necessary to quantify the skill of a specific variable in the hindcasts. While the ACC describes the agreement of the phase of the signal (variable) with one of the reference simulations, the MSESS provides information about the signal's amplitude [*Murphy,* 1988]. A positive MSESS indicates a gain in accuracy for the hindcast compared to the reference; zero no gain, negative MSESS loss in accuracy.

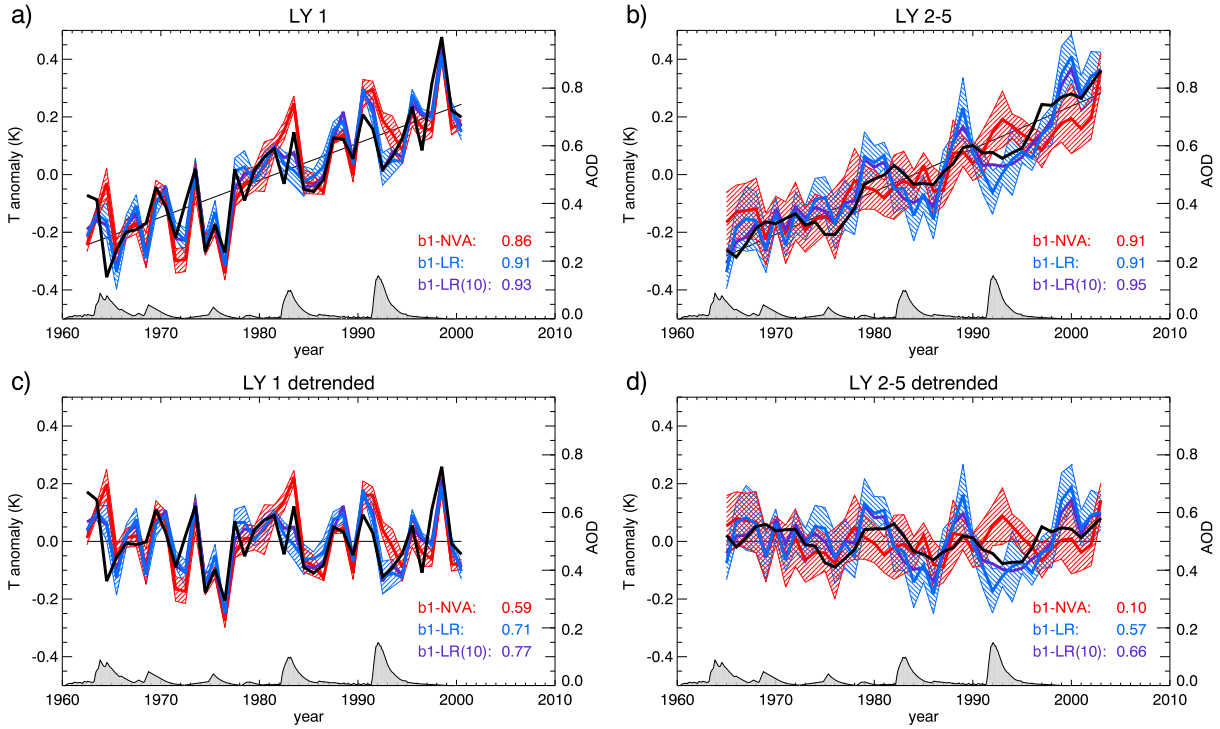

Figure 1. Time series of hindcast and observed anomalies of globally averaged surface temperature from 1962 to 2004 with respect to the mean of 1962–2004 for HadCRUT3v observations [*Brohan et al.*, 2006] (black), b1-LR (blue), and b1-NVA (red). The purple line indicates the 10 ensemble mean of b1-LR. The standard deviation is indicated by the dashed areas. The first prediction year (LY1) (Figures 1a and 1c) and) the second to fifth year ensemble mean of the hindcasts (LY2–5) (Figures 1b and 1d). Detrended time series are shown in Figures 1c and 1d. The numbers indicate the ACC between the hindcasts and the observations over the whole period. The grey shaded region shows the time series of annual-averaged stratospheric aerosol optical depth (AOD) [*Stenchikov et al.*, 1998 and updates].

The reference simulations in our case are the CMIP5 historical simulations. A block bootstrap method [e.g., *Wilks*, 2011], which incorporates autocorrelation functions according to *Goddard et al.* [2013], is applied to estimate significance. For b1-LR 10 ensemble members are available for each run, whereas for b1-NVA only three members exist. Therefore, just the first three ensemble members of b1-LR are considered. Only anomalies of the lead time-dependent time series are considered. Therefore, anomaly calculations result in a lead time-dependent bias adjustment [*International CLIVAR Project Office*, 2011]. Additionally, the anomaly calculation is done in a cross-validated manner, and the grid is time aggregated and spatially averaged on a $5 \times 5°$ grid. We concentrate the analysis on near surface air temperature (TAS) and verify the simulated ensemble mean against Climate Research Unit and Hadley Centre (HadCRUT3v) observations [*Brohan et al.*, 2006]. While we focus on the discussion of the ACC, corresponding maps of the ensemble mean MSESS can be found in the supporting information (Figures S2–S5). We show results of linearly detrended fields exemplary in Figure 1. The corresponding results of detrended fields to Figures 2 and 3 can be found in the supporting information (Figures S3, S5, and S6). Time series are detrended assuming a linear trend using a least squares fit.

3. Results/Discussion

Figure 1 shows a lead time-dependent time series of hindcast and observed anomalies of globally TAS from 1962 to 2005. The volcanic aerosols are cooling the climate on average. Removing the volcanic aerosols results in a warmer global mean temperature of about 0.1 K on average for the hindcast period, which is typical for climate models; see, e.g., *Smith et al.* [2007]. For comparison we also show the ensemble mean time series for the available 10 b1-LR ensemble members which show a slightly higher correlation coefficient. This is in line with *Kadow et al.* [2015] who demonstrate that the prediction skill increases with ensemble size for all lead years. The fact that the 10 member ensemble mean is close to the three member ensemble mean and within its standard deviation reveals that the results discussed here are not dependent on the choice of the

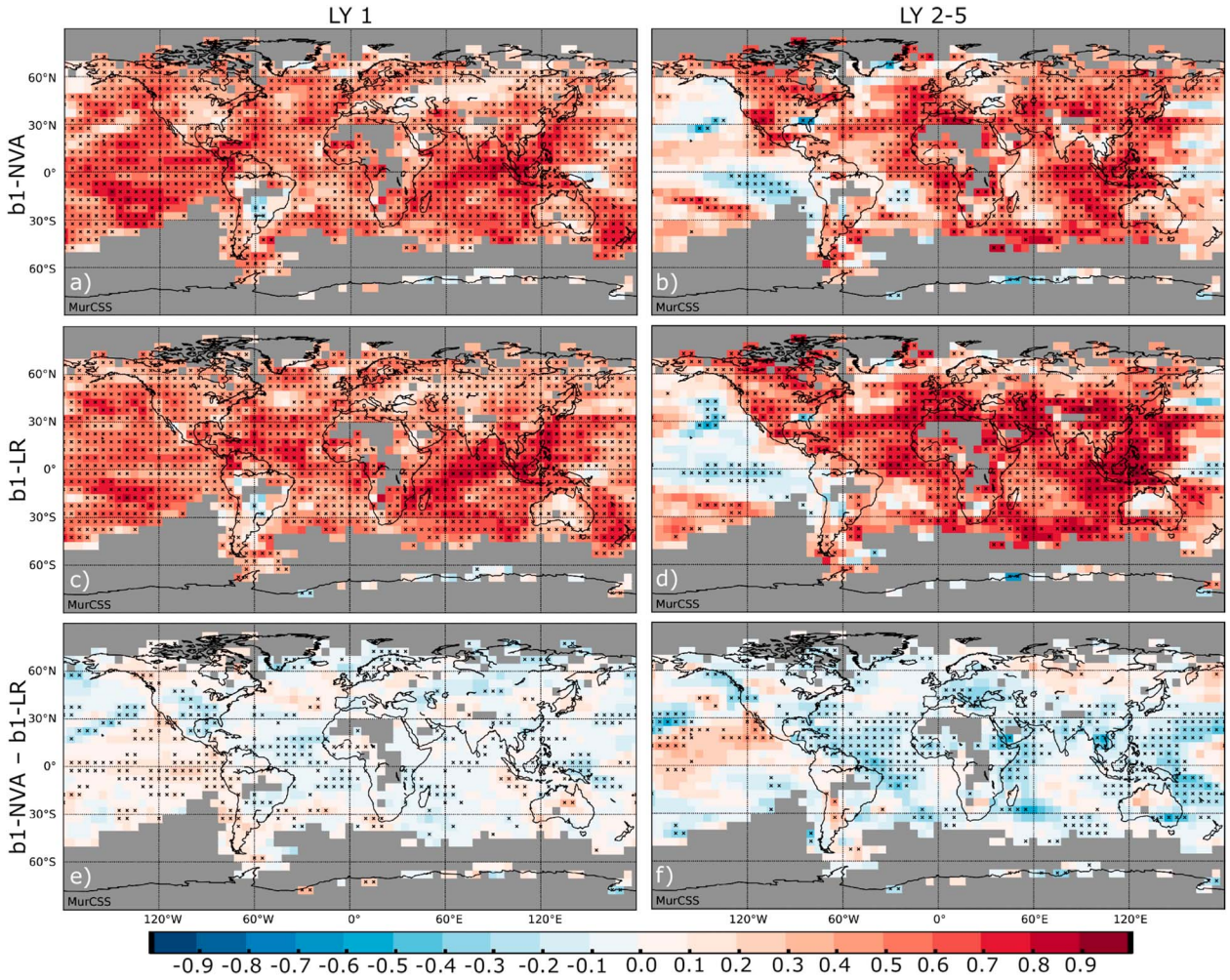

Figure 2. Ensemble mean anomaly correlations of TAS against HadCRUT3v observations [*Brohan et al.*, 2006] over the period 1962–2004. b1-NVA (Figures 2a and 2b), b1-LR (Figures 2c and 2d), and differences between both (Figures 2e and 2f), all averaged over the first prediction year (Figures 2a, 2c, and 2e) and lead years 2–5 (Figures 2b, 2d, and 2f). Crosses denote skill and differences in skill exceeding the 5–95% confidence level. Grid points with missing values in the observations are masked.

selected ensemble members. Both hindcasts with and without volcanic aerosol capture the observed general warming trend (Figures 1a and 1b), but the correlations are higher for b1-LR with a value of 0.91 compared to 0.86 in b1-NVA for the first prediction year (LY1). For the mean of the second to the fifth forecast years (LY2–5) the correlations for b1-LR and b1-NVA are even equal if only three ensemble members are considered for both hindcasts. The relatively high correlation coefficients are clearly determined by the global warming trend on the pentadal time scale considered here. The volcanic impact on global mean TAS is, however, significant for LY1 and LY2–5 for hindcasts started in the first four to six years after the Pinatubo and El Chichón eruptions. These periods are the only ones over the whole time period, where the standard deviations of both hindcast experiments are not overlapping. Over the whole hindcast period, the degradation of simulated TAS with respect to the observations due to the disregard of volcanic aerosol becomes evident only in the detrended time series (Figures 1c and 1d). In LY1 (Figure 1c) the influence of the initialization on the global TAS skill is still visible in the relatively high correlation of 0.71 for b1-LR and 0.59 for b1-NVA. For the LY2–5 (Figure 1d), however, there is almost no correlation (0.1) for b1-NVA with the observations, in contrast to b1-LR (0.57 for 3 and 0.68 for 10 ensembles) which clearly shows the influence of volcanic aerosol on pentadal-scale variability.

Maps of the ensemble mean ACC of TAS (not detrended) in both experiments and the respective differences between b1-NVA and b1-LR are presented for different prediction lead years in Figure 2. Figures 2a–2d show significant positive anomaly correlations in b1-LR and b1-NVA almost everywhere which reflect the observed

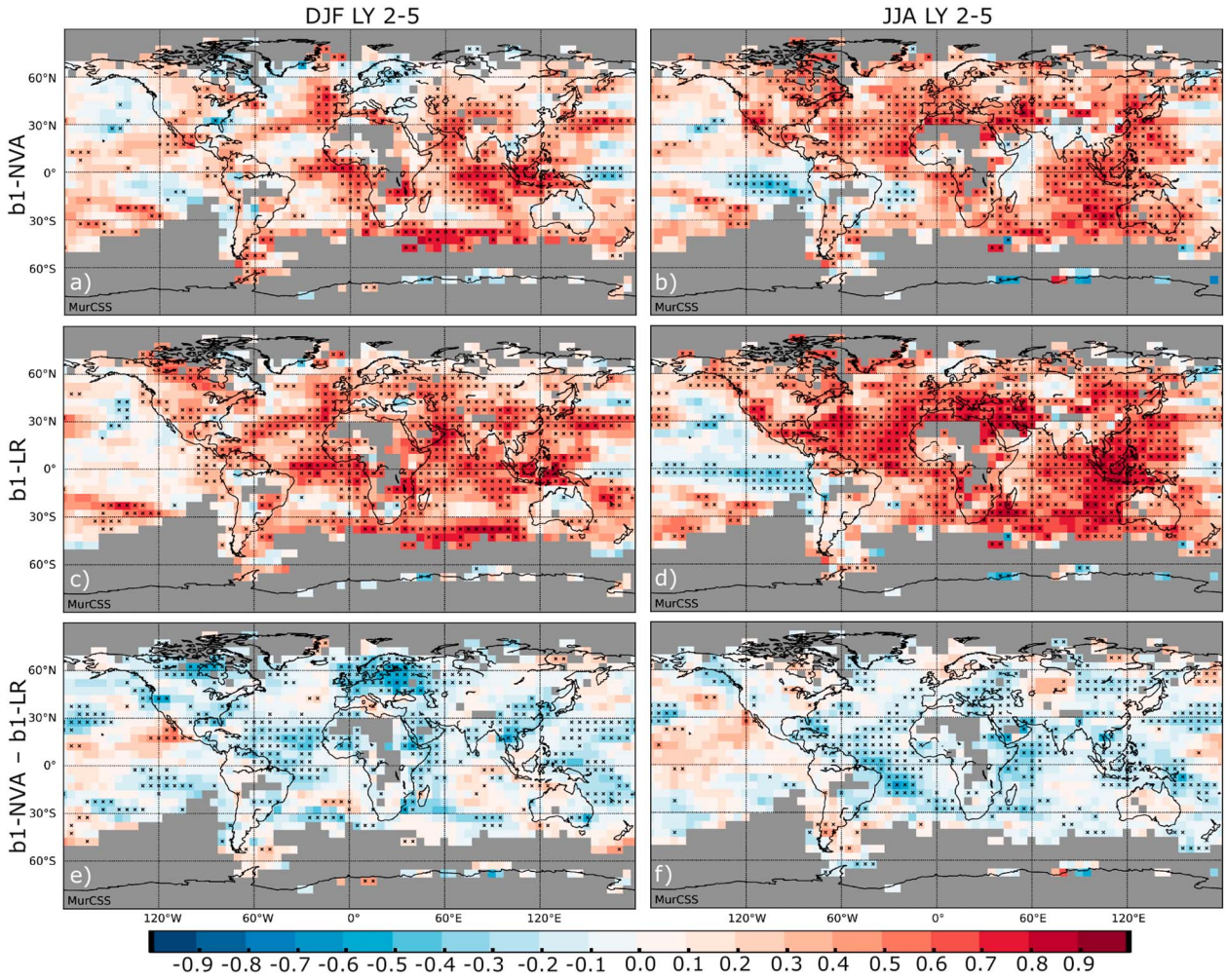

Figure 3. Ensemble multiseasonal mean anomaly correlation of TAS against HadCRUT3v observations [*Brohan et al.*, 2006] over the period 1962–2004 in b1-NVA (Figures 3a and 3b), b1-LR (Figures 3c and 3d), and the differences between both (Figures 3e and 3f) averaged over lead years 2–5 and over NH winter (Figures 3a, 3c, and 3e) and NH summer (Figures 3b, 3d, and 3f). Crosses indicate where the skill and the difference in skill exceed the 5–95% confidence level. Grid points with missing values in the observations are masked.

temperature trend over the last four decades [e.g., *Pohlmann et al.*, 2013; *Kim et al.*, 2012]; see also Figure 1. However, strong regional differences can be found between both simulations which are mostly significant in the tropics and subtropics (Figures 2e and 2f). Negative correlation differences (loss of skill if volcanic aerosol is neglected) can be found over the Atlantic, the Indic, and the western Pacific and positive differences (gain of skill if volcanic aerosol is neglected) over the central and eastern Pacific. Omitting volcanic aerosol hence leads unexpectedly to a higher prediction skill in the eastern tropical Pacific, which is a poorly simulated region in the MPI-ESM [*Jungclaus et al.*, 2013], and in the MiKlip prediction system [*Kadow et al.*, 2015]. Dynamic incompatibility of the oceanic and atmospheric initial states in the equatorial Pacific leads to negative skill [*Thoma et al.*, 2015]. The additional cooling due to the volcanic aerosol enhances this discrepancy. For the hindcasts averaged over LY2–5 the positive and the negative differences in ACC anomalies between b1-NVA and b1-LR become more distinct over the tropical oceans with respect to LY1. While, for example, for LY1 the differences in the ACC between both hindcasts are only weak over the western Pacific and the maritime continent, they are significant for LY2–5 and show an ACC decrease of 0.3–0.5 if volcanic aerosol is neglected. In the North Atlantic region the differences between b1-NVA and b1-LR are insignificant. Therefore, the additional external volcanic forcing is of minor importance there. On average, the anomaly correlation is reduced for LY2–5 in b1-NVA if volcanic aerosol is neglected (Figure 2f). This clearly shows that on the regional scale the volcanic influence is even evident in the nondetrended data.

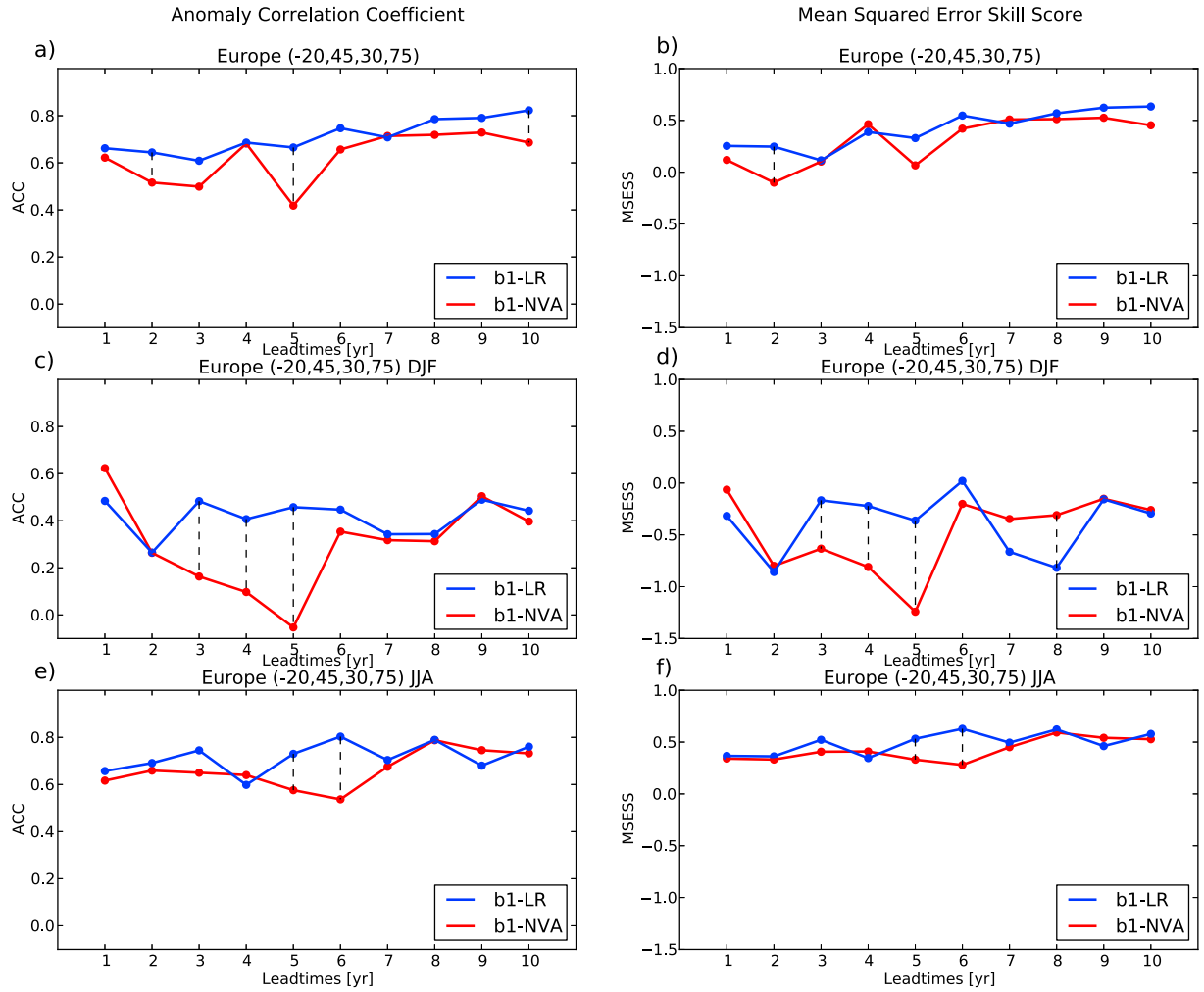

Figure 4. ACC (Figures 4a, 4c, and 4e) and) MSESS (Figures 4b, 4d, and 4f) of European-averaged ensemble mean TAS as a function of lead years verified against HadCRUT3v data [*Brohan et al.*, 2006]. b1-LR simulations are in blue and b1-NVA simulations in red. Multiannual averages are in Figures 4a and 4b, NH winter means in Figures 4c and 4d, and NH summer means in Figures 4e and 4f. Statistical significant differences between the hindcast ensembles are indicated with dashed lines.

Figure 3 shows ACC maps of TAS (not detrended) in both hindcasts and the differences between them for multiyear seasonal means (winter and summer, LY2–5). A pronounced seasonal dependence of the skill has already been found for European temperatures in previous decadal simulations with the MPI-ESM [*Müller et al.*, 2012] with positive skill over northern Europe during NH winter and over central and south-eastern Europe during summer and autumn. Our simulations reveal that these regional increases in skill in NH winter are due to volcanic aerosol because there is no correlation between the b1-NVA hindcasts and HadCRUT temperature visible over central Europe and Scandinavia. Similar features are also found for the Canadian Arctic. The high prediction skill in b1-LR during NH summer over southeastern Europe and the East Mediterranean is also related to volcanic aerosol, while the prediction skill for the Iberian Peninsula and the western Mediterranean is independent from the volcanic forcing. This is also true for the North Atlantic in the region of the Azores high. As already seen in the multiannual averages (Figure 2) the skill is reduced over the tropical and subtropical Atlantic if volcanic aerosol is not considered. The reduction is, however, modulated by the seasonal cycle with a maximum in the subtropics in the winter hemisphere. In NH summer there are also clear ACC differences visible over the Arabian Peninsula and the Arabian Sea as well as over the maritime continent with a significant loss of skill if volcanic aerosol is not considered. Part of these multiyear seasonal mean differences in the tropical region can be explained by the position of the Intertropical Convergence Zone and are related to seasonal changes in cloud cover which modulates the stratospheric aerosol radiative forcing. As for the multiannual mean the average ACC

for the multiseasonal mean is over the CMIP5 period substantially decreased when volcanic aerosol is omitted (Figures 3e and 3f).

The neglection of volcanic aerosol leads to a degraded skill in the hindcasts over north and central Europe in particular in the winter season. For the European-averaged (30°N–75°N, 20°W–45°E) ensemble mean TAS, the anomaly correlation of b1-LR is clearly higher compared to b1-NVA for all lead times in the annual mean as well as in the seasonal (winter and summer) ones (Figure 4). The differences in the detrended values (Figure S6) show a similar behavior. The effect is strongest in the fifth year in NH winter. This is in agreement with *Zanchettin et al.* [2012] who analyzed the volcanic impact on NH winter surface temperature in the last millennium simulations. They found noticeable differences in the latitudinal cooling responses in particular over the Arctic region which experiences a delayed and much stronger posteruption cooling than the tropical latitudes. In NH summer the differences between the hindcasts are smaller but still significant in the fifth and sixth hindcast years.

4. Conclusions/Summary

Our hindcasts with (b1-LR) and without volcanic aerosol (b1-NVA) show (despite a warmer average climate in the latter case) that volcanic aerosol significantly affects the decadal prediction skill of global mean TAS in the first five years after strong volcanic eruptions but not on average over the CMIP5 hindcast period. The predication skill is dominated by the global warming trend. Therefore, the differences between both hindcasts in the anomaly correlation are small for the whole prediction period. The situation, however, is completely different if we remove the global warming trend from the data. A fivefold higher ACC for LY2–5 is found for global mean detrended TAS if volcanic aerosol is taken into account. The low ACC of 0.1 between b1-NVA and HadCRUT surface temperatures shows that the MPI-ESM has almost no prediction skill in global mean surface air temperature for LY2–5 if external forcing (CO_2, volcanic) is neglected. While the volcanic influence on global mean TAS is relatively small in the nondetrended time series, its disregard in the hindcasts leads to significant regional degradations in the prediction skill of TAS over the last 40 years. Omitting volcanic forcing over the CMIP5 hindcast period leads to a loss of skill over the Atlantic, the Indic, and the western Pacific. On average the ACC is for LY2–5 smaller even for the nondetrended data. Seasonal differences between both hindcasts appear very prominently over Continental Europe with significant losses in NH winter over central Europe, Scandinavia and over southeastern Europe, and the East Mediterranean in NH summer. Neglecting volcanic aerosol leads on the other hand to a higher TAS prediction skill over the central and eastern Pacific (at least in the MiKlip b1-LR system). The b1-LR system has no skill in the eastern Pacific. Hence, it might be that this unexpected feature will weaken or disappear in different versions of the MiKlip prediction system where the skill over the east Pacific is improved as in the case of the higher resolved vertical baseline 1 system using the MPI-ESM-MR model. [*Pohlmann et al*., 2013] or in the version with full full-field ocean initialization [*Kruschke et al*., 2015].

Our results are strongly influenced by the strength and number of volcanic eruptions having occurred in the simulated time period. The gain in skill due to volcanic aerosol might therefore be larger or smaller over other hindcast periods dependent on the frequency, the strength, geographical location, and eruption season of the considered volcanic eruptions. This might be relevant also for the most recent decades and the role of small volcanic eruptions in the explanation of the recent warming trend. We could not tackle this question with our experimental setup as the applied volcanic data set ends in 1999 and was set constant afterward. This will be the area of future research. Furthermore, it is important to note that the degradation of prediction skill due to a disregard of volcanic aerosol could only be considered as an upper bound for future forecasts. In an "operational" approach, volcanic aerosol information might be absent for forecasts for which the eruption occurred after the start date, but for successive forecasts information about the volcanic forcing is most likely available so presumably near-real time forecast with operational systems will be repeated if a large volcanic eruption occurs. There exists an apparent mismatch between the observed and simulated seasonal to interannual dynamical responses to volcanic eruptions during the instrumental period. Observations suggest that volcanic eruptions are followed by an anomalously strong NH winter polar vortex, and significant positive anomalies in the NAO and Northern Annular Mode, but CMIP5 models do not robustly reproduce this behavior [e.g., *Driscoll et al*., 2012; *Charlton-Perez et al*., 2013]. There is also a large spread in the response of the oceanic thermohaline circulation [*Ding et al*., 2014] to volcanic forcing in the CMIP5 models. Thus, our results might be biased due to the capabilities of the applied model version. Preferably similar hindcast experiments as presented here should be done in a multimodel framework as the ensemble approach can overcome the systematic model biases from individual models and therefore enhance the decadal prediction skill [*Kim et al*., 2012].

Acknowledgments
We thank Hauke Schmidt for valuable comments on an earlier version of this manuscript. Comments from two anonymous reviewers helped to improve the clarity of the paper. The research was supported by the German Federal Ministry for Education and Research through the "MiKlip" program (FKZ: 01LP1130A (CT), 01LP1144A (HP), and 01LP1160A (CK, SI). We acknowledge the use of observational surface air temperature Hadley Centre data (http://www.metoffice.gov.uk/hadobs/hadcrut3/) and of the European Centre for Medium-Range Weather Forecasts reanalyses data for the initialization (ORAS4, ERA-40, and ERA-Interim). Computations were carried out at the German Climate Computing Centre (DKRZ). Processing was done with the MurCSS decadal evaluation tool [*Illing et al*., 2011]. 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

Balmaseda, M. A., K. Mogensen, and A. T. Weaver (2012), Evaluation of the ECMWF ocean reanalysis system ORAS4, *Q. J. R. Meteorol. Soc.*, *139*, 1132–1161, doi:10.1002/qj.2063.

Bellucci, A., et al. (2015), Advancements in decadal climate predictability: The role of nonoceanic drivers, *Rev. Geophys.*, *53*, 165–202, doi:10.1002/2014RG000473.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

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.

Collins, M. (2003), Predictions of climate following volcanic eruptions, in *Volcanism and the Earth's Atmosphere*, edited by A. Robock and C. Oppenheimer, pp. 283–300, AGU, Washington, D. C.

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

Ding, Y., J. A. Carton, G. A. Chepurin, G. Stenchikov, A. Robock, L. T. Sentman, and J. P. Krasting (2014), Ocean response to volcanic eruptions in Coupled Model Intercomparison Project 5 (CMIP5) simulations, *J. Geophys. Res. Oceans*, *119*, 5622–5637, doi:10.1002/2013JC009780.

Doblas-Reyes, F. J., I. Andreu-Burillo, Y. Chikamoto, J. García-Serrano, V. Guemas, M. Kimoto, T. Mochizuki, L. R. L. Rodrigues, and G. J. van Oldenborgh (2013), Initialized near-term regional climate change prediction, *Nat. Commun.*, *4*, 1715, doi:10.1038/ncomms2704.

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.

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.

Goddard, L., et al. (2013), A verification framework for interannual-to-decadal predictions experiments, *Clim. Dyn.*, *40*(1–2), 245–272, doi:10.1007/s00382-012-1481-2.

International CLIVAR Project Office (2011), Data and bias correction for decadal climate predictions. Southampton, GB, International CLIVAR Project Office, 5 pp. (ICPO Publication Series, 150). [Available from http://eprints.soton.ac.uk/171975/1/ICPO150_Bias.pdf.]

Illing, S., C. Kadow, O. Kunst, and U. Cubasch (2014), MurCSS: A tool for standardized evaluation of decadal hindcast systems, *J. Open Res. Software*, 2:e24, doi:10.5334/jors.bf.

Jolliffe, I. T., and Stephenson, D. B. (2011) Introduction, in *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed., edited by I. T. Jolliffe and D. B. Stephenson, John Wiley, Chichester, U. K., doi:10.1002/9781119960003.ch1

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

Kadow, C., S. Illing, O. Kunst, H. W. Rust, H. Pohlmann, W. A. Müller, and U. Cubasch (2015), Evaluation of forecasts by accuracy and spread in the MiKlip decadal climate prediction system, *Meteorol. Z.*, Special Issue on "Verification and process oriented validation of the MiKlip decadal prediction system" PrePub doi:10.1127/metz/2015/0639.

Kim, H.-M., P. J. Webster, and J. A. Curry (2012), Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts, *Geophys. Res. Lett.*, *39*, L10701, doi:10.1029/2012GL051644.

Kirtman, B., et al. (2013), Near-term climate change: Projections and predictability, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., Cambridge Univ. Press, Cambridge, U. K., and New York.

Kruschke, T., H. W. Rust, C. Kadow, W. A. Müller, H. Pohlmann, G. C. Leckebusch, and U. Ulbrich (2015), Probabilistic evaluation of decadal predictions of Northern Hemisphere winter storms, *Meteorol. Z.*, doi:10.1127/metz/2015/0641.

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

Maher, N., S. McGregor, M. H. England, and A. S. Gupta (2015), Effects of volcanism on tropical variability, *Geophys. Res. Lett.*, *42*, 6024–6033, doi:10.1002/2015GL064751.

Marshall, A. G., A. A. Scaife, and S. Ineson (2009), Enhanced seasonal prediction of European winter warming following volcanic eruptions, *J. Clim.*, *22*, 6168–6180.

Meehl, G. A., et al. (2013), Decadal climate prediction: An update from the trenches, *Bull. Am. Meteorol. Soc.*, *95*, 243–267, doi:10.1175/BAMS-D-12-00241.1.

Müller, W. A., J. Baehr, H. Haak, J. H. Jungclaus, J. Kröger, D. Matei, D. Notz, H. Pohlmann, J. S. von Storch, and J. Marotzke (2012), Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology, *Geophys. Res. Lett.*, *39*, L22707, doi:10.1029/2012GL053326.

Murphy, A. H. (1988), Skill scores based on the mean square error and their relationships to the correlation coefficient, *Mon. Weather Rev.*, *116*, 2417–2424.

Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke (2009), Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic, *J. Clim.*, *22*, 3926–3938.

Pohlmann, H., W. A. Müller, K. Kulkarni, M. Kameswarrao, D. Matei, F. S. E. Vamborg, C. Kadow, S. Illing, and J. Marotzke (2013), Improved forecast skill in the tropics in the new MiKlip decadal climate predictions, *Geophys. Res. Lett.*, *40*, 5798–5802, doi:10.1002/2013GL058051.

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

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*, 98–116, doi:10.1002/jame.20014.

Shiogama, H., S. Emori, T. Mochizuki, S. Yasunaka, T. Yokohata, M. Ishii, T. Nozawa, and M. Kimoto (2010), Possible influence of volcanic activity on the decadal potential predictability of the natural variability in near-term climate predictions, *Adv. Meteorol.*, 657318, doi:10.1155/2010/657318.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model, *Science*, *317*, 796–799, doi:10.1126/science.1139540.

Solomon, S., J. S. Daniel, R. R. Neely, J.-P. Vernier, E. G. Dutton, and L. W. Thomason (2011), The persistently variable "background" stratospheric aerosol layer and global climate change, *Science*, *333*(6044), 866–870, doi:10.1126/science.1206027.

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*, 13,837–13,858, 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*, 146–172, doi:10.1002/jame.20015.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Thoma, M., R. J. Greatbatch, C. Kadow, and R. Gerdes (2015), Decadal hindcasts initialised using observed surface wind stress: Evaluation and prediction out to 2024, *Geophys. Res. Lett.*, *42*, 6454–6461, doi:10.1002/2015GL064833.

Thompson, D. W. J., J. M. Wallace, P. D. Jones, and J. J. Kennedy (2009), Identifying signatures of natural climate variability in time series of global-mean surface temperature: Methodology and insights, *J. Clim.*, *22*, 6120–6141, doi:10.1175/2009JCLI3089.1.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Quart. J. R. Meteorol. Soc.*, *131*, 2961–3012, doi:10.1256/qj.04.176.

Vernier, J.-P., et al. (2011), Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, *Geophys. Res. Lett.*, *38*, L12807, doi:10.1029/2011GL047563.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, *Int. Geophys. Ser.*, 3rd ed., vol. 100, Academic Press, Elsevier, Amsterdam, Netherlands.

Zanchettin, D., C. Timmreck, H.-F. Graf, A. Rubino, S. J. 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.*, doi:10.1007/s00382-011-1167-1.