Dear Professor Peter Haynes:

Thank you very much for sending us the comments by the two Referees and the Editor's decision (they are in italic and coloured in black below). Please see below for our answers to all the comments (in blue). The line numbers are those for the further revised (i.e., the latest) version of the manuscript.

Editor Decision: Reconsider after minor revisions (Editor review) (31 Oct 2015) by Peter Haynes
Comments to the Author:
I have received two referees reports on the revised version of the paper.
Referee 1 recommends publish after technical corrections gives
one specific
example of a correction/clarification needed and recommends a thorough rechecking
of the paper. Referee 2 recommends publish after minor corrections.
Therefore please can you provide a further revised version of the paper
making the requested technical/minor corrections (or justify in a response why
they are not necessary). I will then consider the paper further without sending
again to referees.

We provide you a further revised version and a track-changes version.

Two further points: (i) Referee 3 on the first version of the paper did not wish
to provide a report on the revised version. My own judgement is that you have
responded thoroughly to that referee. (ii) Referee 2 requests comment on the
relation to the methods used in the Mitchell et al (2015) paper. Referee 2 did
not raise this point in their report on the first version of the paper, but
nonetheless I think that it would be valuable if you could provide such
comment. (For some reason you have an 'accepted' comment in the Mitchell et
al reference in the revised version but not in the first version. The paper is
now published so the 'accepted' should be removed.)

Please see below for the response to Referee #2's comments.
We provide the complete paper information to Mitchell et al. (2015) and Nash and Saunders (2015) in the revised version. (Note that we also added another paper, Mitchell (2015), whose current status is "accepted.")

*Small suggestion to read revised manuscript once more. New text near line
120 indicates using that 20CR uses monthly volcanic aerosols but elsewhere,
near line 205, text still indicates 20CR uses annual mean volcanic aerosols.*

Corrected.
We have carefully read the manuscript again. Similar minor corrections were also made
(please see the track-changes version).

*The paper focuses on the interesting topic of volcanic impact on the Earth's
atmosphere using nine reanalysis data sets. The study overlaps with that by
Mitchell et al. (2015), but uses different settings for the multiple regression
procedure. I would suggest the authors more thoroughly explain their
statistical approach and more specifically compare it with the Mitchell et al.
(2015).*

The major difference between the method by Mitchell et al. (2015) and the one for the
current study is whether a volcanic index is considered or not in the regression analysis.
One of the important motivations of the current study is to evaluate the response to the 3
major eruptions *individually* and *separately*. The method by Mitchell et al. only gives a
combined or *averaged* response (for the El Chichón and Mount Pinatubo eruptions as
their analysis period is 1979-2009). Therefore, we took the similar approach that was used
by Free and Lanzante (2009) and Randel (2010), where we subtracted the components of
ENSO, QBO, solar cycle, linear trends, and seasonal harmonics first, and then evaluated
the volcanic signals after the subtraction.

We think that the above point is clear in Introduction (in the last 3 paragraphs).

The Mitchell et al. method may give better estimates for the QBO, ENSO, and solar cycle
signals, with statistically more significant ones, in part because the current method gives

greater residuals (as they also include volcanic signals). Also, the Mitchell et al. method can evaluate the statistical significance for the (combined/averaged) volcanic signals in the exact sense (i.e., by logically following the statistics). On the other hand, in the current method we cannot use the statistics to evaluate the statistical significance in a rigorous way. See also our answer to Referee #2's question for "L216" below.

L 154: Why do you have 41 predictor in equation (1)? I thought you use only 40. Also, ao in equation (1) is not zero. How do you define its contribution to different processes, e.g., ENSO effect?

As explained at L139-143, we have 6 seasonal harmonics, and 5 other climatic indices with 6 seasonal harmonics plus a constant further considered (i.e., 7 indices in total for each climatic index). Thus, we have $6 + 5 \times 7 = 41$ predictors. We think that this has been clearly written in the text.

The a_0 is a constant that is not relevant to any predictors/processes. (In practice, a_0 equals to the average of $Y(t)$ because all our indices have been normalized (i.e., the average of each index has been set to be zero) before the regression calculations.)

L 160: Please reiterate why do you think R(t) contains volcanic signal. Formally it is just an error of approximation, a portion of a signal that your predictors can not cover. E.g., if you miss some important predictors, except volcanic index, that residual will go to R(t). Please elaborate on this.

We think that all the important predictors have been covered in our analysis. During the first review process, Referee #2 asked about possible inclusion of NAO index and a Monsoon index. But, as we replied and noted in the text (L155-157), they are not the *forcing* to the *zonal mean* temperature. We have added the following note at L153-155: "These climate indices are those considered by Free and Lanzante (2009), Randel (2010), and Mitchell et al. (2015), though Free and Lanzante did not consider solar cycle and Mitchell et al. considered the AOD as well."

Furthermore, there is a recently accepted paper by Mitchell (2015) who explicitly dealt with this residuals problem.
Mitchell, D. M.: Attributing the forced components of observed stratospheric temperature variability to external drivers, Q. J. Roy. Meteor. Soc., accepted, 2015.

We change L163 " . . . (Randel, 2010)" to:

" . . . (Randel, 2010; Mitchell, 2015). Mitchell (2015) analyzed two reconstructions of the SSU dataset using model predicted responses to external forcings as the climatic indices. After regressing the model-predicted response patterns onto observations, it was shown that the noise residual was very small compared with the forcing signal. If the volcanic predictor had been omitted (as in our study), the residual would essentially be the volcanic pattern."

In addition to this, we revise L199-200 " . . . Mitchell et al. (2015) supports . . . " to:

" . . . Mitchell et al. (2015), in addition to the finding by Mitchell (2015) as described in Sect. 2, supports . . . "

L 190: This has to be proofed more thorough. Please give a quantitative estimate.

We cannot "prove" that R(t) is composed of volcanic signals and random variations (and does not include others). Our logic is as follows.

1) This assumption is the one used by Free and Lanzante (2009) and Randel (2010).
2) All the climatic indices, that have been used in the previous studies (e.g., Free and Lanzante, 2009; Randel, 2010; Mitchell et al., 2015), are considered.
3) As written above, Mitchell (2015) also showed that if the volcanic predictor had been omitted in the regression analysis, the residual would essentially be the volcanic pattern.
4) The signals for ENSO, QBO, and solar cycle are quite similar to those obtained by Mitchell et al. (2015), though there are inevitable, small quantitative differences due to the use of a different method.

We think that discussing the small quantitative differences from the Mitchell et al. results (in particular for the signals other than the volcanic signals) is not very relevant to this paper. The use of a consistent methodology within this paper for comparisons of different datasets, and the investigation of the responses to the 3 major eruptions *individually* and *separately* are the important aspects of this paper.

Moreover, our quantification of the magnitude and spatial structure of the response to each individual volcanic eruption should be interpreted as the best estimate of a given individual volcanic response after known influences (ENSO, QBO, etc) have been

removed using regression. We cannot exclude the possibility of as-yet unknown influences contaminating the regression residual. However, since volcanic aerosol constitutes a substantial known forcing, and the presence of as-yet unknown influences is speculative, we believe it is reasonable to make use of the regression residual to measure the volcanic response.

L 216: Is the signal statisticaly significant?

We cannot evaluate the statistical significance of our volcanic signals in the exact sense in terms of statistics. This is because of our methodology which is to evaluate the response to the 3 major eruptions *individually* and *separately* as explained above. We can only say, as in the text (i.e., at L226-227), "the volcanic signals discussed above are distinguishable from random variations." In the revised manuscript, we have added the following words to this sentence, "in the sense that these signals are much greater than one SD of the residuals."

L 318-319:
There could be multiple reasons, but heating is controlled by
aerosol absorption (both in IR and nearIR) not just extinction optical depth.
Please clarify.

This is a very good point. We have communicated with the NOAA colleagues to clarify this point. We have added the following sentence in Section 2, right after "(G. Compo and C. Long, private communication, 2015)" in the revised manuscript:
"Furthermore, the stratospheric optical depth data at 550 nm given by Sato et al. (1993) were translated to the optical depth values for ultraviolet, visible, near infrared, and infrared spectral bands (Y.-T. Hou, private communication, 2015)."

Global temperature response to the major volcanic eruptions in multiple reanalysis datasets

M. Fujiwara[1], T. Hibino[1,*], S. K. Mehta[2,**], L. Gray[3,4], D. Mitchell[3], and
J. Anstey[3,***]

[1]Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan
[2]Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Japan
[3]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, UK
[4]NERC National Centre for Atmospheric Science (NCAS), UK
[*]now at: PAP Corporation, Nagoya, Japan
[**]now at: Research Institute, SRM University, Chennai, India
[***] now at: Canadian Centre for Climate Modelling and Analysis, Environment Canada, Canada

Correspondence to: M. Fujiwara (fuji@ees.hokudai.ac.jp)

Abstract. The global temperature responses to the eruptions of Mount Agung in 1963, El Chichón in 1982, and Mount Pinatubo in 1991 are investigated using nine currently available reanalysis datasets (JRA-55, MERRA, ERA-Interim, NCEP-CFSR, JRA-25, ERA-40, NCEP-1, NCEP-2, and 20CR). Multiple linear regression is applied to the zonal and monthly mean time series of temperature for two periods, 1979–2009 (for eight reanalysis datasets) and 1958–2001 (for four reanalysis datasets), by considering explanatory factors of seasonal harmonics, linear trends, Quasi-Biennial Oscillation, solar cycle, and El Niño Southern Oscillation. The residuals are used to define the volcanic signals for the three eruptions separately, and common and different responses among the older and newer reanalysis datasets are highlighted for each eruption. In response to the Mount Pinatubo eruption, most reanalysis datasets show strong warming signals (up to 2–3 K for one-year average) in the tropical lower stratosphere and weak cooling signals (down to -1 K) in the subtropical upper troposphere. For the El Chichón eruption, warming signals in the tropical lower stratosphere are somewhat smaller than those for the Mount Pinatubo eruption. The response to the Mount Agung eruption is asymmetric about the equator with strong warming in the Southern Hemisphere mid-latitude upper troposphere to lower stratosphere. Comparison of the results from several different reanalysis datasets confirms the atmospheric temperature response to these major eruptions qualitatively, but also shows quantitative differences even among the most recent reanalysis datasets. The consistencies and differences among different reanalysis datasets provide a measure of the confidence and uncertainty in our current understanding of the volcanic response. The results of this intercomparison study may be useful for validation of climate model responses to volcanic forc-

ing and for assessing proposed geoengineering by stratospheric aerosol injection, as well as to link studies using only a single reanalysis dataset to other studies using a different reanalysis dataset.

1 Introduction

Explosive volcanic eruptions inject sulphur species to the stratosphere in the form of SO_2 and H_2S which convert to H_2SO_4 aerosols. These aerosols are then transported both vertically and horizontally into the stratosphere by the Brewer–Dobson circulation (Butchart, 2014), stay there to perturb the radiative budget on a timescale of a few years, and thus affect global climate (Robock, 2000). The stratospheric volcanic aerosol layer is heated by absorption of near-infrared solar radiation and upward longwave radiation from the troposphere and surface. In the troposphere, the reduced near-infrared solar radiation is compensated by the additional downward longwave radiation from the aerosol layer. At the surface the large reduction in direct shortwave radiation due to the aerosol layer is the main cause of net cooling there.

Stratospheric aerosol optical depth (AOD) is an indicator of volcanic eruptions that affect global climate and has been estimated from various information (e.g., Sato et al., 1993; Robock, 2000; Vernier et al., 2011). Since 1960 astronomical observations such as solar and stellar extinction and lunar eclipses have become available from both hemispheres, and since 1979 extensive satellite measurements have begun with the Stratospheric Aerosol Monitor (SAM) II on the Nimbus-7 satellite. Extending over a longer period, the global radiosonde network that provides global atmospheric (upper-air) temperature data has been operating since the 1940s, with improved spatial resolution since the late 1950s (Gaffen, 1994). Since 1979, global satellite temperature measurements have begun with the Microwave Sounding Unit (MSU) and Stratospheric Sounding Unit (SSU) instruments on the TIROS-N satellite and on the subsequent several National Oceanic and Atmospheric Administration (NOAA) satellites. Since 1998, the Advanced MSU-A (AMSU-A) instruments on several NOAA satellites have provided global temperature measurements. See, e.g., Cristy et al. (2003), Wang et al. (2012), Wang and Zou (2014), Zou et al. (2014), and Nash and Saunders (2015) for these satellite temperature measurements.

Since the late 1950s, there occurred three major volcanic eruptions that significantly affected global climate, which are Mount Agung (8° S, 116° E), Bali, Indonesia in March 1963, El Chichón (17° N, 93° W), Chiapas, Mexico in April 1982, and Mount Pinatubo (15° N, 120° E), Luzon, Philippines in June 1991. The volcanic explosivity index (VEI) of these eruptions are 6 for Mount Pinatubo, 5 for El Chichón, and 4 for Mount Agung (Robock, 2000). Free and Lanzante (2009) and Randel (2010) used homogenized radiosonde datasets while Santer et al. (2001) and Soden et al. (2002) used MSU satellite data to investigate the tropospheric and stratospheric temperature response to these eruptions. When extracting the volcanic signals, one needs a good evaluation, at the same time, of the components of El Niño Southern Oscillation (ENSO), Quasi-Biennial Oscillation (QBO), and

2

11 year solar cycle as well as seasonal variations and linear trends. Each of the above four studies used a variety of regression analyses.

An atmospheric reanalysis system provides a best estimate of the past state of the atmosphere using atmospheric observations with a fixed assimilation scheme and a fixed global forecast model (Trenberth and Olson, 1988; Bengtsson and Shukla, 1988). It is an operational analysis system at a particular time (e.g., 1995 for the NCEP-1 system and 2009 for the JRA-55 system), which has been continuously improved with the main motivation being to improve the tropospheric weather prediction. Using a fixed assimilation-forecast model to produce analyses of observational data that were previously analysed in the context of operational forecasting - hence the "re" in "reanalysis" - prevents artificial changes being produced in the analysed fields due to system changes. But, as described above, the observational data inputs still vary over the period of the reanalysis. Currently, there are about 10 global atmospheric reanalysis datasets available worldwide. Table 1 lists the reanalysis datasets considered in this study. It is known that different reanalysis datasets give different results for the same diagnostic. Depending on the diagnostic, the different results may be due to differences either in the observational data assimilated, the assimilation scheme or forecast model, or any combination of these (see, e.g., Fujiwara et al., 2012 for a list of some examples). It is therefore necessary to compare all (or some of the newer) reanalysis datasets for various key diagnostics for understanding of the data quality and for future reanalysis improvements (Fujiwara and Jackson, 2013). To be more specific to the current study, the major observational sources of atmospheric (upper-air) temperature are basically common for all the reanalysis datasets in Table 1 (except for the 20CR which only assimilated surface pressure reports). They are radiosondes and satellite microwave and infrared sounders (i.e., MSU, SSU, and AMSU-A). There are three components that do differ in different reanalysis systems: (1) detailed bias-correction or quality-control methods for the original observations before the assimilation, (2) the assimilation scheme, and (3) the forecast model. Thus, any differences in the analysis results in this study would be due to the differences in these components (except for the 20CR).

Recently, Mitchell et al. (2015) analysed temperature and zonal wind data from nine reanalysis datasets using a linear multiple regression technique during the period from 1979 to 2009 by considering QBO, ENSO, AOD as a volcanic index, and solar cycle, with a focus on the solar cycle response. However, the volcanic response shown by Mitchell et al. is a combined response due to the major eruptions over the period 1979–2009 (i.e., El Chichón in 1982 and Mount Pinatubo in 1991).

Investigation of climatic response to individual volcanic eruptions using multiple reanalysis datasets for the purpose of comparison and evaluation of reanalysis datasets is rather limited. For example, Harris and Highwood (2011) showed global mean surface temperature changes following the Pinatubo eruption using NCEP-1 and ERA-40 reanalysis data for comparison with their model experiments. Analysing all available reanalysis datasets for the 20th-century three major eruptions separately and for the region covering both troposphere and stratosphere will provide valuable infor-

3

mation for model validation as well as on the current reanalysis data quality for capturing volcanic signals. Such an analysis would also be valuable when assessing one of the proposed geoengineering options, i.e., stratospheric aerosol injection to counteract global surface warming (e.g., Crutzen, 2006; Robock et al., 2013).

In the present study, we analyse zonal and monthly mean temperature data from nine reanalysis datasets to investigate the response to the Mount Agung, El Chichón and Mount Pinatubo eruptions separately. The temperature response to the Mount Agung eruption is investigated using four reanalysis datasets (JRA-55, ERA-40, NCEP-1, and 20CR) that cover the period back to the 1960s. A multiple regression technique is used to remove the effects of seasonal variations, linear trends, QBO, solar cycle, and ENSO, and the residual time series is assumed to be composed of volcanic effects and random variations. The remainder of this paper is organized as follows. Section 2 describes the datasets and analysis method. Section 3 provides results and discussion. Finally, Section 4 lists the main conclusions.

2 Data and Method

Monthly mean pressure-level temperature data from the nine reanalysis datasets listed in Table 1 were downloaded from each reanalysis-centre website or the US National Center for Atmospheric Research (NCAR) Research Data Archive (http://rda.ucar.edu/). Zonal means were derived for each dataset before the analysis. All the reanalysis datasets except 20CR assimilated upper-air temperature measurements from radiosondes and from SSU, MSU, and AMSU-A satellite instruments, with varied assimilation techniques. 20CR assimilated only surface pressure reports and used observed monthly sea-surface temperature and sea-ice distributions as boundary conditions for the forecast model. Note also that for the 20CR, monthly latitudinally-varying distributions of volcanic aerosols (averaged for four bands, i.e., 90° N–45° N, 45° N–equator, equator–45° S, and 45° S–90° S) were specified based on data from Sato et al. (1993), and a monthly climatological global distribution of aerosol vertical profiles on a 5° grid was specified based on data from Koepke et al. (1997) (G. Compo and C. Long, private communication, 2015). Furthermore, the stratospheric optical depth data at 550 nm given by Sato et al. (1993) were translated to the optical depth values for ultraviolet, visible, near infrared, and infrared spectral bands (Y.-T. Hou, private communication, 2015). Therefore, 20CR is expected to show volcanic signals even though it did not assimilate upper-air temperature data. The atmospheric forecast model of the 20CR is nearly the same as used in the NCEP-CFSR but with a lower resolution, and thus the NCEP-CFSR also included the same volcanic aerosols. None of the other reanalysis datasets included radiative forcing due to volcanic aerosols in the forecast model. See Mitchell et al. (2015) for further technical comparisons among different reanalysis datasets. For a complete description of each reanalysis, see the reference papers shown in Table 1.

4

Table 1 also shows the period of data availability for each reanalysis dataset. For a direct inter-comparison, we define two analysis periods, namely, between 1979 and 2009 (31 years) for eight reanalysis datasets (all except ERA-40) and between 1958 and 2001 (44 years) for four reanalysis datasets (JRA-55, ERA-40, NCEP-1, and 20CR). The former covers the eruptions of El Chichón in 1982 and Mount Pinatubo in 1991, while the latter also covers the eruption of Mount Agung in 1963. Results from JRA-55, NCEP-1, and 20CR for the El Chichón and Mount Pinatubo eruptions for the two different-period analyses also provide an opportunity to investigate sensitivity to the choice of analysis period.

A multiple regression technique is applied to extract volcanic signals (e.g., Randel and Cobb, 1994; Randel, 2010; von Storch and Zwiers, 1999, Chapt. 8.4). First, all major variabilities, except for volcanic effects, were evaluated and subtracted from the original zonal and monthly mean temperature data. The major variabilities include seasonal harmonics of the form, $a_1 \sin \omega t + a_2 \cos \omega t + a_3 \sin 2\omega t + a_4 \cos 2\omega t + a_5 \sin 3\omega t + a_6 \cos 3\omega t$, with $\omega = 2\pi/(12\,\mathrm{mon})$, linear trends, two QBO indices, ENSO, and solar cycle. For the latter five climatic indices, the six seasonal harmonics and a constant are further considered to construct seven indices for each of the five indices, as was done by Randel and Cobb (1994). For the two QBO indices, we use 20 and 50 hPa monthly mean zonal wind data taken at equatorial radiosonde stations provided by the Freie Universität Berlin. The cross-correlation coefficient for these two QBO indices is -0.24 for 1979–2009 and -0.21 for 1958–2001. For the ENSO index, we use the Niño 3.4 index, which is a standardized sea surface temperature anomaly in the Niño 3.4 region (5° N–5° S, 170–120° W), provided by the NOAA Climate Prediction Center. As is often done, a time lag for atmospheric response is considered for the ENSO index. We chose 4 months for the lag, following Free and Lanzante (2009). We confirmed that changing the ENSO lag from 0 to 6 months gives somewhat different ENSO signals particularly in the tropical stratosphere but does not alter other signals, including volcanic signals, significantly. For the solar cycle index, we use solar 10.7 cm flux data provided by the NOAA Earth System Research Laboratory. These climate indices are those considered by Free and Lanzante (2009), Randel (2010), and Mitchell et al. (2015), though Free and Lanzante did not consider solar cycle and Mitchell et al. considered the AOD as well. (Note that we do not consider other indices, e.g., the North Atlantic Oscillation index and the Indian Monsoon index because the former is considered to be a response not a forcing and both are considered to be more related to regional response, not zonal mean response.) The multiple regression model that we use in this study is therefore,

$$Y(t) = a_0 + \sum_{l=1}^{41} a_l x_l(t) + R(t), \tag{1}$$

where $Y(t)$ is the zonal and monthly mean temperature time series at a particular latitude and pressure grid point, and a_l is the least squares solution of a parameter for climatic index time series $x_l(t)$. $R(t)$ is the residual of this model which is assumed to be composed of volcanic signals and random variations (Randel, 2010; Mitchell, 2015). Mitchell (2015) analyzed two reconstructions of the SSU

5

dataset using model predicted responses to external forcings as the climatic indices. After regressing the model-predicted response patterns onto observations, it was shown that the noise residual was very small compared with the forcing signal. If the volcanic predictor had been omitted (as in our study), the residual would essentially be the volcanic pattern.

~~Then,~~ Finally, by following Randel (2010), the volcanic signal for each eruption is defined as the difference between the 12 month averaged $R(t)$ after each eruption and the 36 month averaged $R(t)$ before each eruption.

There are several other possible minor variations for the methodological details, i.e., for the multiple regression model, the choice of particular index datasets, and the volcanic signal definition. The use of a consistent methodology is important for comparisons of different datasets. Where possible, however, we will discuss the methodological dependence below.

3   Results and Discussion

3.1   The 1979–2009 Analysis

Figures 1 and 2 show temperature variations in association with the QBO, solar cycle and ENSO from JRA-55 and MERRA, respectively, for the region from 1000 to 1 hPa. The coloured regions are those evaluated as statistically significant at the 95 % confidence level (von Storch and Zwiers, 1999, Chapt. 8.4.6), with an effective degree of freedom where data are assumed to be independent for every three months. Comparing with the results from Mitchell et al. (2015) who used a regression analysis with different details, the setting of this effective degree of freedom may be somewhat too conservative. This is because the regions evaluated as statistically significant are smaller than those in Mitchell et al. (2015) particularly for the solar and ENSO signals in the tropical lower stratosphere, but the general features are quite similar to those shown in Mitchell et al. (2015) although they also considered a volcanic index in the multiple regression analysis. The two QBO variations are displaced vertically by a quarter cycle in the tropics because of their downward phase propagation. The temperature QBO has off-equatorial out-of-phase signals centred around $30°$ N and around $30°$ S because of the associated secondary meridional circulation (Baldwin et al., 2001). The major response to the solar cycle is the tropical lower stratospheric warming. The ENSO response includes the tropical tropospheric warming and a hint of tropical stratospheric cooling, although the statistical significance of this latter signal is weak. The strength of this cooling signal is sensitive to the choice of the time lag for the ENSO index (4 months in this study and 0 month in Mitchell et al., 2015). There also exists midlatitude lower stratospheric warming in both hemispheres for ENSO. The signals of QBO, solar cycle, and ENSO in the other 6 reanalysis datasets (ERA-Interim, NCEP-CFSR, JRA-25, NCEP-1, NCEP-2, and 20CR; not shown) are also similar to those in Mitchell et al. (2015). 20CR shows no QBO signals (and no zonal-wind QBO; not shown) and no tropical stratospheric solar response. NCEP-CFSR shows weaker tropical lower stratospheric solar cycle warming. The

6

overall agreement with the results in Mitchell et al. (2015), in addition to the finding by Mitchell (2015) as described in Sect. 2, supports the assumption that the residual $R(t)$ is composed of volcanic signals and random variations.

Figure 3 shows the residual time series averaged for 30° N–30° S at 50 and at 300 hPa together with the lower-to-middle stratospheric AOD time series averaged for 27.4° N–27.4° S provided by the NASA Goddard Institute for Space Studies (Sato et al., 1993). The AOD time series clearly shows the timing of the El Chichón eruption and Mount Pinatubo eruption and the duration of their impact on the stratospheric aerosol loading. At 50 hPa, all reanalysis datasets show 1–2 K peak warming within one year after the El Chichón eruption, and most (except 20CR and JRA-25) show 2–2.5 K peak warming within one year after the Mount Pinatubo eruption. As described in Sect. 2, 20CR does not assimilate upper-air data, but incorporates annual averages of volcanic aerosols in the forecast model. Thus, 20CR shows a warming signal in association with both eruptions, though the one for Mount Pinatubo is smaller and slower. 20CR also shows warming signals in 1989 and in 1990 though none of the other datasets show the corresponding signals. The warming in JRA-25 is ~ 1 K smaller than other reanalysis datasets except 20CR. This cold bias can be seen at least during the period 1988–1994. This might be in part related to the known stratospheric cold bias in JRA-25 (Onogi et al., 2007). The radiative scheme used in the JRA-25 forecast model has a known cold bias in the stratosphere, and the TOVS SSU/MSU measurements do not have a sufficient number of channels to correct the model's cold bias; after introducing the ATOVS AMSU-A measurements in 1998, such a cold bias disappeared in the JRA-25 data product. It is also possible that the cold bias in JRA-25 during the TOVS era was not constant over time, in particular when unusual, volcanically affected temperature measurements came into the JRA-25 system, which could contribute to the smaller warming signals in our data analysis. As described in Sect. 2, except for 20CR, NCEP-CFSR is the only reanalysis (except 20CR) that included stratospheric volcanic aerosols in the forecast model, but no clear difference is found in comparison with other recent reanalysis datasets. At 300 hPa, all reanalysis datasets show 0.4–0.8 K peak cooling within one year after the Mount Pinatubo eruption. No clear signals are found at 300 hPa for the El Chichón eruption. Note that the standard deviation (SD) of the residual time series is ~ 1 K for tropical 50 hPa and ~ 0.3 K for tropical 300 hPa for all the datasets; thus, the volcanic signals discussed above are distinguishable from random variations in the sense that these signals are much greater than one SD of the residuals.

Figure 4 shows the temperature signals for the El Chichón eruption from the 8 reanalysis datasets. As described in Sect. 2, the volcanic signal is defined as the difference between the 12 month averaged $R(t)$ after each eruption and the 36 month averaged $R(t)$ before each eruption. The coloured regions are also defined by following Randel (2010), i.e., as those regions with positive (negative) values more (less) than twice the SD of annual mean residual $\overline{R(t)}$. The annual mean is taken here because of the use of 12 month average in the volcanic signal definition. For the recent four reanalysis datasets, i.e., JRA-55, MERRA, ERA-Interim, and NCEP-CFSR, the tropical lower stratospheric

warming of 1.2–1.6 K centred around 50–30 hPa is a common signal. There are also Northern Hemisphere high-latitude middle-upper stratospheric warming and tropical upper stratospheric cooling signals, though the latter is comparable to random variations in some of the four datasets and thus its statistical significance is weak. The tropical and midlatitude troposphere is only weakly cooling, with a maximum cooling (0.4–0.8 K) occurring in the upper troposphere at 20–30° N. For JRA-25, the tropical lower stratospheric warming is confined around 100–50 hPa with (statistically insignificant) cooling signals around 50–10 hPa. ~~This may be due to~~ This might be in part related to the cold bias in JRA-25 as described in the previous paragraph. The tropospheric features in JRA-25 are similar to those in the latest four reanalysis datasets. For NCEP-1 and NCEP-2, the tropical stratospheric warming region extends to 10 hPa where it maximises, and the 20–30° N upper tropospheric cooling is largely missing. The major differences of the NCEP-1 and NCEP-2 systems from the recent four reanalysis systems include the lower model top height (3 hPa), older forecast model and assimilation scheme (of the 1990s; see Table 1), and the use of retrieved temperature data for the assimilation of SSU, MSU, and AMSU-A data. It is possible that these factors may be responsible for the different signals of the El Chichón eruption in NCEP-1 and NCEP-2. (See also discussion on the results for the Mount Pinatubo eruption below.) For 20CR, tropical stratospheric warming is present, but again, this is due to the specified volcanic aerosols in the forecast model.

Free and Lanzante (2009) and Randel (2010) analysed the temperature signals for the El Chichón eruption using different homogenized radiosonde datasets globally up to the 30 hPa level. The distribution of the tropical lower stratospheric warming signal is similar, though the peak warming is greater, i.e., 1.6–2 K for Free and Lanzante (2009, their Figure 3) and 2.5–3 K for Randel (2010, his Figure 4). (Note that Free and Lanzante defined the volcanic signals as the difference between the 24 month average after the eruption and the 24 month average before the eruption, but we use the same definition of volcanic signals as Randel (2010) and still obtain roughly a factor of two discrepancy in tropical lower stratospheric warming (1.2–1.6 K from the reanalyses versus 2.5–3 K from the radiosondes).) Free and Lanzante (2009) also show a 20–30° N upper tropospheric cooling of 0.6–0.9 K.

Figure 5 shows the temperature signals for the Mount Pinatubo eruption. For the latest four reanalysis datasets, i.e., JRA-55, MERRA, ERA-Interim, and NCEP-CFSR, the tropical lower stratospheric warming of 2.0–2.8 K (depending on datasets) centred around 50–30 hPa is a common signal. In the upper troposphere, a cooling (0.4–0.8 K) at 20–30° N and at 15–45° S can be seen, with the latter somewhat greater. JRA-25 shows similar upper tropospheric features and relatively similar lower stratospheric features, though for the latter, the warming magnitude is smaller and the "random" variability becomes large above the 50 hPa level because of the reason described above (i.e., the cold bias and its disappearance in 1998). For NCEP-1 and NCEP-2, the tropical tropospheric and stratospheric features are similar to those for the latest four reanalysis datasets, though the lower stratospheric warming magnitude is slightly smaller than in most of the other reanalyses. Compar-

8

ing with the El Chichón case, the NCEP-1 and NCEP-2 systems worked much better to capture the Mount Pinatubo signals for some reasons. For 20CR, the tropical stratospheric warming is not de-
275 tected. This is because of the unknown warming signals in 20CR in 1989 and in 1990 (see Fig. 3) that raised the 36 month averaged base in the volcanic signal definition. As in Fig. 3, there are no relevant signals in AOD around 1989–1990. Thus, the unknown warming signals are likely due to unrealistic (unforced) variations in the 20CR system.

The temperature signals for the Mount Pinatubo eruption shown in Randel (2010) are similar to
280 the present results both in the tropical-midlatitude stratosphere and troposphere, though Randel's stratospheric warming peak value is somewhat greater ($\sim 3\,\mathrm{K}$) and his upper tropospheric cooling is somewhat greater (0.5–1 K) and more uniform in latitude. On the other hand, Free and Lanzante (2009) show that the lower stratospheric warming signal is split near the equator with two maxima (1.6–2 K at $10°$ N and $> 2\,\mathrm{K}$ at $15°$ S, both at 70–50 hPa) and that the upper tropospheric cooling
285 signal has its peak (0.9–1.2 K) around $20°$ S. In summary, the recent four reanalysis datasets (i.e., JRA-55, MERRA, ERA-Interim, and NCEP-CFSR) give more consistent signals for both eruptions compared to the two radiosonde data analyses using different homogenized datasets by Free and Lanzante (2009) and Randel (2010).

3.2   The 1958–2001 Analysis

290 The multiple regression analysis is applied to the four reanalysis datasets, namely, JRA-55, ERA-40, NCEP-1, and 20CR which cover the period of 1958–2001. Figure 6 shows temperature variations associated with the QBO, solar cycle, and ENSO from JRA-55. Comparing with the 1979–2009 analysis results shown in Fig. 1, all variations are quite similar, with the statistically significant regions for the solar cycle variation being much greater both in the tropical stratosphere and in the
295 tropical troposphere. The same is true for NCEP-1 (not shown). 20CR does not have QBO and stratospheric solar-cycle signals, but does show ENSO signals in both 1979–2009 and 1958–2001 analyses; the 20CR ENSO signals are similar to those from all other reanalysis datasets. ERA-40 shows similar results to JRA-55 except for the solar cycle variation. In ERA-40, the tropical lower stratospheric warming signal in association with the solar cycle is very weak and not symmetric
300 about the equator, in contrast to the results by Crooks and Gray (2005) and Mitchell et al. (2015) who both applied a regression analysis during the period 1979–2001.

Figure 7 shows the time series of residual $R(t)$ and stratospheric AOD averaged over the trop-ics for the period between 1958 and 2001. The AOD time series shows the timing of the Mount Agung eruption in March 1963 as well as the El Chichón and Mount Pinatubo eruptions. The fea-
305 tures at both 50 and 300 hPa for the El Chichón and Mount Pinatubo eruptions are quite similar to the 1979–2009 analysis results shown in Fig. 3, including the 20CR's smaller and slower Mount Pinatubo signal at 50 hPa. For the Mount Agung eruption, $\sim 2.5\,\mathrm{K}$ peak warming is seen within one year after the eruption except for 20CR. At 300 hPa, a sudden cooling occurred about one year later,

9

i.e., in mid-1964 for all the datasets, which is probably related to the Mount Agung eruption. The
cooling might have continued for more than one year. ERA-40 shows anomalous $\sim 1\,\mathrm{K}$ warming
in the mid-1970s at both levels, which are not present in other reanalysis datasets (see also Fig. 14
of Kobayashi et al., 2015). The AOD time series in Fig. 7 shows a small increase in the mid-1970s
which is probably due to the eruption of Mount Fuego ($14°\,\mathrm{N}$, $91°\,\mathrm{W}$), Guatemala, in October–
December 1974 (VEI 4, Smithsonian Institution National Museum of Natural History Global Vol-
canism Program, http://www.volcano.si.edu/, last accessed August 2015). But, the magnitude and
the sign (i.e., warming) at 300 hPa seems unrealistic. Before the introduction of horizontally dense
satellite measurements in 1979, the upper-air temperature is constrained basically only by horizon-
tally inhomogeneous, relatively sparse radiosonde data (see, e.g., Fig. 2 of Uppala et al., 2005). Also,
the ERA-40 system is a relatively old system (the 2001 version of the ECMWF analysis system).
These two facts are possible reasons for the ERA-40's anomalous warming in the mid-1970s. A
stream change of the reanalysis execution could also be a potential reason. For the ERA-40, there
were three execution streams, that is, 1989–2002, 1957–1972, and 1972–1988 (Uppala et al., 2005).
But the stream change point of 1972 is unlikely to explain the anomalous warming starting around
the end of 1974.

Figure 8 shows the temperature signals for the Mount Agung eruption from 4 different reanalysis
datasets. All except 20CR show Southern Hemisphere lower stratospheric warming centred at 40–
$30°\,\mathrm{S}$ and 100–50 hPa, with an extension to equatorial latitudes at 50 hPa. The maximum warming
value varies with dataset, that is, 1.6–2 K for NCEP-1, 2–2.4 K for JRA-55, and 2.4–2.8 K for ERA-
40. The reason for the weak signal in 20CR is in the fact that 20CR does not assimilate upper-air
temperature observations but does consider volcanic aerosol loading in the forecast model. The mod-
elled aerosol loading was probably too weak to simulate the lower stratospheric warming signals.
For all the four reanalysis datasets, the 300 hPa cooling shown in Fig. 7 is not captured with the
current volcanic-signal definition (i.e., 12 month average after the eruption started).

Free and Lanzante (2009) showed a very similar Southern Hemisphere midlatitude lower strato-
spheric warming signal ($> 2\,\mathrm{K}$) in association with the Mount Agung eruption using a homogenized
radiosonde dataset. Sato et al. (1993) showed that the aerosols emitted from the Mount Agung erup-
tion were transported primarily to the Southern Hemisphere. The uncertainty of the Mount Agung
signal is considered to be much greater than that of the El Chichón and Mount Pinatubo signals be-
cause of the unavailability of satellite temperature data during the 1960s and because of the limited
number of available reanalysis datasets. A tentative conclusion is that the JRA-55 dataset is the most
reliable for studies of the Mount Agung eruption, since it is currently the only available dataset that
employs the most up-to-date reanalysis system.

The El Chichón signal from the 1958–2001 analysis (not shown) is very similar to the one from
the 1979–2009 analysis for JRA-55 and 20CR shown in Fig. 4. For NCEP-1, the warming signal
in the tropical 30–10 hPa region shown in Fig. 4 becomes weaker, thus showing better agreement

with the results from the modern reanalysis datasets (e.g., JRA-55). ERA-40 shows similar signal to JRA-55 at least up to the 10 hPa level globally. The Mount Pinatubo signal from the 1958–2001 analysis (not shown) is very similar to the one from the 1979–2009 analysis for JRA-55, NCEP-1, and 20CR. ERA-40 shows similar signal to JRA-55 at least up to the 20 hPa level globally.

350 Figure 9 provides a useful summary plot for the volcanic effects on the temperature at 50 hPa and at 300 hPa using JRA-55 from the 1958–2001 analysis together with the AOD latitudinal time series. The aerosol loading due to the Mount Agung eruption in March 1963 extended primarily to the Southern Hemisphere, that due to the El Chichón eruption in April 1982 was very large in the tropics and extended primarily to the Northern Hemisphere, and that due to the Mount Pinatubo eruption

355 in June 1991 was very large in the tropics and extended to both hemispheres. The tropical lower stratosphere warmed after these three major volcanic eruptions with a time scale of 1–2 years. The warming after the Mount Agung eruption is not equatorially symmetric and is shifted to the Southern Hemisphere and to somewhat lower levels, in association with the distribution of aerosol loading. The tropical troposphere became cooler after the Mount Pinatubo eruption but the tropospheric re-

360 sponse is not as clear for the other two eruptions. The high latitude response is also unclear both in the troposphere and stratosphere due to high random variations that mask any volcanic signals, if they exist.

4   Conclusions

Monthly and zonal mean temperature data from nine reanalysis datasets were analysed to charac-

365 terize the response to the three major volcanic eruptions during the 1960s to 1990s. Multiple linear regression analysis was applied to evaluate seasonal variations, trends, QBO, solar cycle and ENSO components, and the residual time series $R(t)$ was assumed to be composed of volcanic signals and random variations. The volcanic signals were defined as the difference between the 12 month averaged $R(t)$ after each eruption and the 36 month averaged $R(t)$ before each eruption. Two separate

370 analyses were performed, that is, one for the period 1979–2009 (31 years) using eight reanalysis datasets and the other for 1958–2001 (44 years) using four reanalysis datasets. The former covered the eruptions of El Chichón (April 1982) and Mount Pinatubo (June 1991), while the latter also covered the eruption of Mount Agung (March 1963).

The general features of the response to QBO, solar cycle, and ENSO were found to be quite similar

375 to those shown in Mitchell et al. (2015) who also used a multiple linear regression with different methodological details, in particular, considering a volcanic index as well. Also, these signals were at least qualitatively similar among reanalysis datasets, with a notable exception that 20CR shows no QBO signals and no tropical stratospheric solar response.

The latitude-pressure distribution of El Chichón and Mount Pinatubo temperature response was

380 quite similar at least among the recent four reanalysis datasets (JRA-55, MERRA, ERA-Interim, and

NCEP-CFSR) and between the 1979–2009 and 1958–2001 analyses. For the Mount Pinatubo eruption, tropical lower stratospheric warming and tropical upper tropospheric cooling were observed. For the El Chichón eruption, tropical lower stratospheric warming was observed, but tropospheric cooling was much weaker than the Mount Pinatubo case. For the Mount Agung eruption, JRA-55, ERA-40, and NCEP-1 showed Southern Hemisphere lower stratospheric warming centred at 40–30° S and 100–50 hPa, with an equatorial extension to 50 hPa. Thus, the Agung signal was asymmetric about the equator and very different from the El Chichón and Pinatubo signals. We suggest that this may be due to differences in the transport of volcanic aerosols (Sato et al., 1993).

Evidently the temperature responses were different for different volcanic eruptions. In particular, wide-spread upper tropospheric cooling was observed only for the Mount Pinatubo case, and the Mount Agung lower stratospheric response was found to be asymmetric about the equator. The characteristics in the temperature response are related to the transport of stratospheric aerosols together with the amount of sulphur species emitted into the stratosphere. Depending on the location, season, and magnitude of the eruption, the climatic response can be very different (e.g., Trepte and Hitchman, 1992). This needs to be taken into account when evaluating the stratospheric sulphur injection as a geo-engineering option, and thus accurate estimations of stratospheric circulation and transport are essential for assessing the climate impacts. Also, it should be noted that accurate evaluation of naturally induced variability such as QBO, solar cycle, and ENSO is necessary to detect the effects of artificial injection.

Finally, we conclude that the four most recently developed reanalysis datasets, i.e., JRA-55, MERRA, ERA-Interim, and NCEP-CFSR are equally good for studies on the response to the El Chichón and Mount Pinatubo eruptions. The NCEP-1, NCEP-2, and JRA-25 showed different tropical stratospheric signals particularly for the El Chichón eruption, though the original upper-air temperature observations assimilated are basically common, and this is most probably in association with the use of older analysis systems. The 20CR did not assimilate upper-air observations and gives very different volcanic signals, despite including volcanic aerosols in the forecast model. Of the currently available datasets that extend back far enough (JRA-55, ERA-40, NCEP-1, and 20CR) the JRA-55 dataset is probably the most ideally suited for studies of the response to the Mount Agung eruption because it is the only dataset that employs the most recent reanalysis system.

420

References

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The quasi-biennial oscillation, Rev. Geophys., 39, 179–229, doi:10.1029/1999RG000073, 2001.

Bengtsson, L. and Shukla, J.: Integration of space and in situ observations to study global climate change, B. Am. Meteorol. Soc., 69, 1130–1143, 1988.

Butchart, N.: The Brewer–Dobson Circulation, Rev. Geophys., 52, 157–184, doi:10.1002/2013RG000448, 2014.

Christy, J. R., Spencer, R. W., Norris, W. B., and Braswell, W. D.: Error estimates of version 5.0 of MSU-AMSU bulk atmospheric temperature, J. Atmos. Ocean. Tech., 20, 613–629, 2003.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The twentieth century reanalysis project, Q. J. Roy. Meteor. Soc., 137, 1–28, doi:10.1002/qj.776, 2011.

Crooks, S. A. and Gray, L. J.: Characterization of the 11-year solar signal using a multiple regression analysis of the ERA-40 dataset, J. Climate, 18, 996–1015, 2005.

Crutzen, P. J.: Albedo enhancement by stratospheric sulfur injections: a contribution to resolve a policy dilemma?, Climatic Change, 77, 211–220, doi:10.1007/s10584-006-9101-y, 2006.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Free, M. and Lanzante, J.: Effect of volcanic eruptions on the vertical temperature profile in radiosonde data and climate models, J. Climate, 22, 2925–2939, doi:10.1175/2008JCLI2562.1, 2009.

Fujiwara, M. and Jackson, D.: SPARC Reanalysis Intercomparison Project (S-RIP) planning meeting, 29 April–1 May 2013, Exeter, UK, SPARC Newsletter, 41, 52–55, 2013.

Fujiwara, M., Polavarapu, S., and Jackson, D.: A proposal of the SPARC Reanalysis/Analysis Intercomparison Project, SPARC Newsletter, 38, 14–17, 2012.

Gaffen, D. J.: Temporal inhomogeneities in radiosonde temperature records, J. Geophys. Res., 99, 3667–3676, doi:10.1029/93JD03179, 1994.

Harris, B. M. and Highwood, E. J.: A simple relationship between volcanic sulfate aerosol optical depth and surface temperature change simulated in an atmosphere–ocean general circulation model, J. Geophys. Res., 116, D05109, doi:10.1029/2010JD014581, 2011.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo,

K. C., Ropelewski, C., Wang, J., Jenne R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, 1996.

465 Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., and Potter, G. L.: NCEP–DOE AMIP-II reanalysis (R-2), B. Am. Meteorol. Soc., 83, 1631–1643, 2002.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP–NCAR 50-year reanalysis: monthly means CD-ROM and documentation, B. Am. Meteorol. Soc., 82, 247–267, 2001.

470 Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 reanalysis: general specifications and basic characteristics, J. Meteorol. Soc. Jpn., 93, 5–48, doi:10.2151/jmsj.2015-001, 2015.

Koepke, P., Hess, M., Schult, I., and Shettle, E. P.: Global aerosol data set, Report No. 243, Max-Planck-Institut für Meteorologie, Hamburg, ISSN 0937-1060, 44 pp., 1997.

475 Mitchell, D. M., Gray, L. J., Fujiwara, M., Hibino, T., Anstey, J. A., Ebisuzaki, W., Harada, Y., Long, C., Misios, S., Stott, P. A., and Tan, D.: Signatures of naturally induced variability in the atmosphere using multiple reanalysis datasets, Q. J. Roy. Meteor. Soc., ~~accepted,~~ 141, 2011–2031, doi:10.1002/qj.2492, 2015.

Mitchell, D. M.: Attributing the forced components of observed stratospheric temperature variability to external drivers, Q. J. Roy. Meteor. Soc., accepted, 2015.

480 Nash, J. and Saunders, R.: A review of Stratospheric Sounding Unit radiance observations for climate trends and reanalyses, Q. J. Roy. Meteor. Soc., ~~accepted,~~ 141, 2103–2113, doi:10.1002/qj.2505, 2015.

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R.: The JRA-25 reanalysis, J. Meteorol. Soc. Jpn., 85, 369–432, doi:10.2151/jmsj.85.369, 2007.

485 Randel, W. J.: Variability and trends in stratospheric temperature and water vapor, in: The Stratosphere: Dynamics, Transport and Chemistry, Geophys. Monogr. Ser. 190, edited by: Polvani, L. M., Sobel, A. H., and Waugh, D. W., American Geophysical Union, Washington, D.C., USA, 123–135, 2010.

Randel, W. J. and Cobb, J. B.: Coherent variations of monthly mean total ozone and lower stratospheric temperature, J. Geophys. Res., 99, 5433–5447, 1994.

490 Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications, J. Climate, 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

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

Robock, A., MacMartin, D. G., Duren, R., and Christensen, M. W.: Studying geoengineering with natural and anthropogenic analogs, Climatic Change, 121, 445–458, doi:10.1007/s10584-013-0777-5, 2013.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H.,

500 Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen,

Y., Han, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP climate forecast system reanalysis, B. Am. Meteorol. Soc., 91, 1015–1057, doi:10.1175/2010BAMS3001.1, 2010.

505 Santer, B. D., Wigley, T. M. L., Doutriaux, C., Boyle, J. S., Hansen, J. E., Jones, P. D., Meehl, G. A., Roeckner, E., Sengupta, S., and Taylor, K. E.: Accounting for the effects of volcanoes and ENSO in comparisons of modeled and observed temperature trends, J. Geophys. Res., 106, 28033–28059, doi:10.1029/2000JD000189, 2001.

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

Soden, B. J., Wetherald, R. T., Stenchikov, G. L., and Robock, A.: Global cooling after the eruption of Mount Pinatubo: a test of climate feedback by water vapor, Science, 296, 727–730, doi:10.1126/science.296.5568.727, 2002.

Trenberth, K. E. and Olson, J. G.: An evaluation and intercomparison of global analyses from the National
515 Meteorological Center and the European Centre for Medium-Range Weather Forecasts, B. Am. Meteorol. Soc., 69, 1047–1057, 1988.

Trepte, C. R. and Hitchman, M. H.: Tropical stratospheric circulation deduced from satellite aerosol data, Nature, 355, 626–628, doi:10.1038/355626a0, 1992.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K.,
520 Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen,
525 J.: The ERA-40 reanalysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

Vernier, J.-P., Thomason, L. W., Pommereau, J.-P., Bourassa, A., Pelon, J., Garnier, A., Hauchecorne, A., Blanot, L., Trepte, C., Degenstein, D., and Vargas, F.: Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade, Geophys. Res. Lett., 38, L12807, doi:10.1029/2011GL047563, 2011.

530 von Storch, H. and Zwiers, F. W.: Statistical Analysis in Climatic Research, Cambridge Univ. Press, Cambridge, UK, 484 pp., 1999.

Wang, L., Zou, C.-Z., and Qian, H.: Construction of stratospheric temperature data records from Stratospheric Sounding Units, J. Climate, 25, 2931–2946, doi:10.1175/JCLI-D-11-00350.1, 2012.

Wang, W. and Zou, C.-Z.: AMSU-A-only atmospheric temperature data records from the lower troposphere to
535 the top of the stratosphere, J. Atmos. Ocean. Tech., 31, 808–825, doi:10.1175/JTECH-D-13-00134.1, 2014.

Zou, C.-Z., Qian, H., Wang, W., Wang, L., and Long, C.: Recalibration and merging of SSU observations for stratospheric temperature trend studies, J. Geophys. Res., 119, 13180–13205, doi:10.1002/2014JD021603, 2014.

16

Figure 1. Latitude–pressure distribution of the temperature variations in association with (top left) QBO 20 hPa zonal wind index, (top right) QBO 50 hPa zonal wind index, (bottom left) solar cycle index, and (bottom right) ENSO index from JRA-55 reanalysis data for the period 1979–2009. The units are in Kelvin per standard deviation (SD) of each index (note that each index time series was standardized before the regression analysis). Solid and dashed lines denote positive and negative values, respectively. The contour interval is 0.2 K for QBO, and 0.1 K for solar cycle and ENSO. Coloured regions denote those greater (orange) and smaller (blue) than random variations with the 95 % confidence interval at each location.

Figure 2. As in Fig. 1 but for MERRA reanalysis data.

Figure 3. Time series of temperature residual $R(t)$ (including volcanic signals and random variations) averaged for $30°$ N–$30°$ S for the 1979–2009 regression analysis from eight reanalysis datasets at **(a)** 50 hPa and **(b)** 300 hPa. **(c)** Time series of aerosol optical depth at 550 nm averaged for $27.4°$ N–$27.4°$ S and integrated for the region 15–35 km. Vertical dotted lines indicate the starting date of the two volcanic eruptions.

Figure 4. Latitude-pressure distribution of the temperature response to the El Chichón eruption in April 1982 for the 1979–2009 analysis from eight reanalysis datasets. Solid and dashed lines denote positive and negative values, respectively. The contour interval is 0.4 K. Coloured regions denote those with positive and greater (orange) and negative and smaller (blue) than twice the SD of annual mean residual $\overline{R(t)}$ at each location.

Figure 5. As in Fig. 4 but for the Mount Pinatubo eruption in June 1991.

Figure 6. As in Fig. 1 but for the period 1958–2001.

Figure 7. As in Fig. 3 but for the 1958–2001 regression analysis from four reanalysis datasets. Vertical dotted lines indicate the starting date of the three volcanic eruptions.

Figure 8. As in Fig. 4 but for the Mount Agung eruption in March 1963 for the 1958–2001 analysis from four reanalysis datasets.

Figure 9. Time-latitude distribution of temperature residual $R(t)$ (including volcanic signals and random variations) for the 1958–2001 regression analysis from JRA-55 reanalysis data at **(a)** 50 hPa and **(b)** 300 hPa. Thirteen-month running average has been taken for $R(t)$. The contour interval is 1.0 K for **(a)** and 0.25 K for **(b)**. The regions with 0–1 K (> 1 K) are coloured in orange (red) in **(a)**. The regions with 0 to -0.25 K (< -0.25 K) are coloured in light (dark) blue. **(c)** Time-latitude distribution of aerosol optical depth at 550 nm integrated for the region 15–35 km. The contour interval is 0.04. The regions with 0.04–0.12 (> 0.12) are coloured in orange (red) in **(c)**.

Table 1. List of global atmospheric reanalysis datasets considered in this study .

Dataset	Centre	Year[1]	Period	Reference
ERA-Interim	ECMWF	2007	1979–present	Dee et al. (2011)
ERA-40	ECMWF	2001	Sep 1957–Aug 2002	Uppala et al. (2005)
JRA-55	JMA	2009	1958–present	Kobayashi et al. (2015)
JRA-25 / JCDAS	JMA and CRIEPI	2004	Jan 1979–Jan 2014	Onogi et al. (2007)
MERRA	NASA	2008	1979–present	Rienecker et al. (2011)
NCEP-CFSR	NOAA/NCEP	2007	1979–~~2009~~March 2011, ~~2010~~April 2011[2]–present	Saha et al. (2010)
NCEP-DOE AMIP-II R-2 (NCEP-2)	NOAA/NCEP and DOE AMIP-II	1998	1979–present	Kanamitsu et al. (2002)
NCEP-NCAR R-1 (NCEP-1)	NOAA/NCEP and NCAR	1995	1948–present	Kalnay et al. (1996); Kistler et al. (2001)
NOAA-CIRES 20CR v2 (20CR)	NOAA and CIRES/Univ. Colorado	2008	Nov 1869–Dec 2012	Compo et al. (2011)

[1] For the version of the operational analysis system that was used for the reanalysis.

[2] The model horizontal resolution has increased in ~~2010~~April 2011 in the NCEP-CFSR.