Response to Referee #3.

Thank you very much for reviewing our manuscript.

Interactive comment on "Global temperature response to the major volcanic eruptions in multiple reanalysis datasets" by M. Fujiwara et al.
Anonymous Referee #3
Received and published: 23 June 2015

In this manuscript the authors analyze the temperature response to major volcanic eruptions in nine reanalyses datasets. After regressing the reanalysis temperature fields to eliminate the effects of QBO, solar cycle, and ENSO, the authors analyze the time series of global temperature residuals and the zonal mean temperature residuals during the year following the eruptions of Agung, El Chichon, Pinatubo, and Fernandina.

General comments.

- *The idea behind this study is interesting and worth to be explored, but I think that the analyses of the reanalyses datasets should be more detailed. Most of the manuscript is a description of the figure, and does not address the reasons for discrepancies, which makes impossible to assess which reanalyses system is doing a better job during specific time series.*

In general, we found three groups, i.e., (1) newer reanalysis datasets, JRA-55, MERRA, ERA-Interim, and NCEP-CFSR, (2) older reanalysis datasets, JRA-25, ERA-40, NCEP-1, and NCEP-2, and (3) 20CR which is without atmospheric (upper-air) observations assimilated. For (1) and (2), the original observations that have the major impact on the reanalysis temperature are common, which are radiosondes and microwave and infrared sounders on several operational satellites. Therefore, the causes of the differences between (1) and (2), within (1), and within (2) should not be in the original observations assimilated but in the bias correction (i.e., quality control) methods for observational data before the assimilation, in the assimilation scheme, and in the forecast model. The newer reanalysis datasets use newer and thus basically better assimilation scheme and forecast model, with improved data quality control procedures. Even within the newer reanalysis datasets (1), we found some quantitative differences in the volcanic temperature response. At the moment, the exact causes of these differences are unknown, and thus what we can do is to regard these differences as the uncertainty information, i.e., uncertainty of our knowledge on the global temperature response to the major volcanic eruptions.

We will add these points in the revised manuscript.

- *It would be useful to include a figure/table showing the observational systems assimilated by each reanalyses dataset and the period of time in which they were assimilated. Such figure would help interpreting the changes in temperature residuals. Does any of the periods used to analyze the volcanic response include the addition/removal of an observing system? Would this invalidate the analyses for the response to that particular volcano?*

The major observations that are directly relevant to the reanalysis atmospheric (upper-air) temperature data (except for the 20CR) are basically common and summarized below:

- Radiosonde temperature measurements
    - Available throughout the period
    - Spatially much more inhomogeneous than satellite measurements, with far less stations over the oceans and in the Southern Hemisphere (e.g., Seidel et al., 2011)
    - The typical balloon burst altitude of 30 hPa (e.g., Seidel et al., 2011)
- Microwave and infrared sounders on several operational satellites (mostly the "NOAA" satellites)
    - The SSU and MSU instruments in the TOVS[*1)] suite between 1979 and 2005 on several operational satellites
    - The AMSU-A instrument in the ATOVS[*2)] suite from 1998 onward on several operational satellites
    - Spatially much more homogeneous, but with broader vertical weighting functions (e.g., Seidel et al., 2011 for the TOVS suite)
    - There is a technical difference for the satellite data assimilation. The NCEP-1 and NCEP-2 assimilated retrieved temperature profiles, while the other reanalysis systems (except the 20CR) directly assimilated radiance data using a radiative transfer model. The radiance assimilation is considered better than the retrieved data assimilation because the retrieval model can be an additional source of uncertainty.

In addition, there are two other types of temperature measurements as follows.

- Aircraft temperature measurements
    - With impacts only around 200-300 hPa
    - High density of measurement points only over north America, the high-latitude Atlantic Ocean, and the Europe
    - Known warm biases with respect to radiosondes (Ballish and Kumar, 2009; Rienecker et al.,

2011)

- The JRA-25 and JRA-55 only assimilated aircraft horizontal wind measurements, not temperature.

- GNSS[*3)]/GPS[*4)] Radio Occultation temperature measurements
  - From 2001 onward (CHAMP[*5)]: 2001-2008; FORMOSAT-3/COSMIC[*6)]: from 2006 onward; and MetOp-A[*7)]: from 2008 onward)
  - Assimilated only in the ERA-Interim, NCEP-CFSR, and JRA-55
  - Not covering the periods of the volcanic eruptions considered in this study; thus, their impacts on our results are only indirect through the evaluation of other forced variabilities

[*1)]TOVS: Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder
[*2)]ATOVS: Advanced TIROS Operational Vertical Sounder
[*3)]GNSS: Global Navigation Satellite System
[*4)]GPS: Global Positioning System
[*5)]CHAMP: CHAllenging Minisatellite Payload
[*6)]FORMOSAT-3/COSMIC: Constellation Observing System for Meteorology, Ionosphere, and Climate on the Republic of China Satellite (ROCSat) renamed to FORMOSAT
[*7)]MetOp-A: MetOp is a series of three polar orbiting meteorological satellites operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)

All these observations are assimilated in all the reanalysis systems except 20CR and except noted. In practice, radiosondes and microwave and infrared sounders are the main sources of reanalysis temperature. Therefore, it is difficult to attribute the differences among different reanalysis datasets to original observations assimilated. Rather, we can see that there are two key years from the observations viewpoint, i.e., the year 1979 when data from operational (TOVS) satellites appeared and the year 1998 an advanced (ATOVS) satellite instruments appeared. For our current study, the eruptions of Mount Pinatubo and El Chichon occurred during the TOVS period, while the eruptions of Mount Agung (and other three volcanos) occurred during the period when only radiosondes were available for upper-air temperature measurements. Thus, the uncertainty for the global temperature response to the Mount Agung eruption is considered greater than that to the Mount Pinatubo and El Chichon eruptions.

We will add the major points from the above discussion in Introduction of the revised manuscript.

Ballish, B. A., and Kumar, V. K.: Systematic differences in aircraft and radiosonde temperatures: implications for NWP and climate studies, Bull. Amer. Meteorol. Soc., 89, 1689-1708,

doi:10.1175/2008BAMS2332.1, 2008.
Seidel, D. J., Gillett, N. P., Lanzante, J. R., Shine, K. P., and Thorne, P. W.: Stratospheric temperature trends: our evolving understanding, WIREs Clim. Change, 2, 592-616, doi:10.1002/wcc.125, 2011.

- *Given the change in temperatures simply due to the inclusion of additional datasets, would it be more appropriate to divide the data record in periods with a specific set of instruments (i.e. no instrument is added/dropped) and perform separate regression analyses for each period?*

As explained above, except for 20CR, there is basically no difference in terms of the original observations assimilated. The year 1979 is the key year, and that is the reason why many reanalysis datasets start from 1979. Considering this fact, we made two separate data analyses, one for the period 1979-2009 and the other for the period 1958-2001. It is technically possible to make another test analysis for the period 1958-1978 by using the four reanalysis datasets. But in this case, we are afraid that what we will see would be the impact of a shorter time period of the regression analysis, rather than the impact of the difference in the types of observations.

Specific comments.

- *Fig 4: the high top models and low top models differ quite a bit from each other in terms, for instance, of altitude of the maximum. Is there a specific reason behind that distinguish the behavior of high- and low-top models?*

Thank you for pointing this out.
The stratospheric warming for the El Chichon eruption in the NCEP-1 and NCEP-2 is located at 10 hPa, the top boundary for these reanalysis systems, while that for the other reanalysis systems (including the 20CR) is located around 50 hPa. The major differences of the NCEP-1 and NCEP-2 from the other reanalysis systems include the lower model top height (3 hPa), older forecast model and assimilation scheme (of the 1990s; see Table 1, the fourth column, of Mitchell et al. (2015)), 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 Chichon eruption in NCEP-1 and NCEP-2.

However, this is not true for the Mount Pinatubo eruption: All the reanalysis systems except the 20CR show a lower stratospheric warming signal centered around 50 hPa. The NCEP-1 and NCEP-2 systems worked much better to capture the Mount Pinatubo signals for some reasons.

The 20CR did not assimilate any upper-air observations but took into account the volcanic aerosols in the forecast model, and these facts should be responsible for the different response.

We will add these discussions in the revised manuscript.

- *page 13325 L 11: 20CR shows "unknown warming signals" in 1989/1990. There is no hypothesis about the origin of these signals?*

As written above, the 20CR did not assimilate any upper-air observations but took into account the volcanic aerosols in the forecast model. In practice, the 20CR uses the same monthly-mean aerosol index data shown in Figure 3 (i.e., taken from Sato et al. (1993)) which were, for the case of 20CR, averaged into 4 evenly spaced latitude bins (i.e., 90S-45S, 45S-equator, equator-45N, and 45N-90N). Figure 3 does not show any relevant AOD signals in 1989/1990. Thus, the unknown warming signals are probably due to unrealistic (unforced) variations in the 20CR system.

This discussion will be added in the revised manuscript.

- *page 13326 L18-20: As for the previous comment, why would ERA40 show a 1K warming not present in the other reanalyses? What causes that warming? Is it overestimation of the volcanic signal, wrong dynamics? No hypothesis?*

Before the introduction of horizontally dense satellite measurements in 1979, the upper-air temperature is constrained basically only by horizontally 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 why there occurred some unrealistic meandering in the upper tropospheric and lower stratospheric temperature during this period in the ERA-40 system. A stream change of the reanalysis execution could also be a potential reason. For the ERA-40, there were three streams, i.e., 1989-2002, 1957-1972, and 1972-1988 (Uppala et al., 2005). But, the stream change point of 1972 probably cannot explain the anomalous warming starting around the end of 1974.

This discussion will be added in the revised manuscript.

- *page 13327 L 21: "the former MAY correspond: : :" Why MAY? It should be possible to check in the lat-lon data, correct?*

We will completely remove the discussion on the temperature response to the three smaller-scale eruptions. Please see below.

- *page 13328 L2: Could the opposite response in the case of Fernandina be due to lingering effects of Agung in the three years before the Fernandina eruption?*

As you pointed out, one possibility is the lingering effects of the Mount Agung eruption. For the three smaller-scale eruptions, we may need different definitions for each (e.g, different base period). However, doing this would take time and make this paper complicated. Therefore, we decided that we will completely remove the discussion on the temperature response to the three smaller-scale eruptions from this paper.

- *page 13328 L5: are aerosol heating rates included in the reanalyses output? If so, the cause of the warming could be checked.*

The 20CR and the NCEP-CFSR are the only reanalysis systems that considered volcanic aerosols in the forecast model. Therefore, there is no volcanic signal in the heating rate data for the other reanalysis datasets. Any temperature changes in association with the volcanic eruptions came from the temperature observations in the reanalysis systems except for the 20CR and NCEP-CFSR.

- *page 13328 L 9: the structure in the residuals similar to the QBO response could be due to aerosol-induced effects in dynamics (e.g. Aquila et al. (2014) in the case of a tropical geoengineering aerosol injection). However, why would it be present only in the case of Fernandina? Any hypothesis?*

Again, we decided that we will completely remove the discussion on the temperature response to the three smaller-scale eruptions from this paper. (The paper by Aquila et al. is very interesting. In particular, comparing their Fig. SM4 (a weaker case) with their Fig. 3, the weaker the aerosol loading becomes, the lower the tropical temperature pattern becomes, being more similar to our Figure 10. However, the large difference between Aquila et al.'s Fig. SM4 and our Figure 10 is that the former still has a tropical lower stratospheric warming signal which is essential to explain the circulation and further temperature changes by Aquila et al., but the latter does not have. Thus, for our data analysis, the lingering effects of the Mount Agung eruption cannot be excluded.)

- *page 13329 L20: 20CR shows no QBO signals in the temperature fields or has no QBO at all? If 20CR assimilated only surface pressures, either the underlying model has a way of generating* *the QBO or there is no QBO at all in the model.*

The 20CR does not have the QBO in zonal wind and in temperature. This means that the forecast model of the 20CR does not have spontaneous QBO-like oscillations. This is also true at least for the NCEP-CFSR (Saha et al., 2010, pages 1026-1027), JRA-55 (Kobayashi et al., 2014), and MERRA (Coy, 2014). For your information, for a new reanalysis dataset, MERRA-2 (the data will be publicly available later in 2015 (Steven Pawson, private communication, December 2014)), the forecast model has spontaneous QBO-like oscillations by increasing the parametrized non-orographic (convective) gravity wave forcing in the tropics compared to the MERRA (Coy, 2014).

Coy, L.: Effects of new data types and data assimilation system upgrades on middle atmosphere dynamics, presented at the SPARC Data Assimilation workshop, at the NOAA Center for Weather and Climate Prediction (NCWCP), 8 September 2014.

Kobayashi, C., Endo, H., Ota, Y., Kobayashi, S., Onoda, H., Harada, Y., Onogi, K., and Kamahori H.: Preliminary results of the JRA-55C, an atmospheric reanalysis assimilating conventional observations only, Sci. Online Lett. Atmos., 10, 78–82, doi:10.2151/sola.2014-016, 2014.

Aquila, V., Garfinkel, C. I., Newman, P. A., Oman, L. D., Waugh, D. W. (2014). Modifications of the quasibiennial oscillation by a geoengineering perturbation of the stratospheric aerosol layer. Geophysical Research Letters, 41. http://doi.org/10.1002/(ISSN)1944-8007.

Thank you very much for pointing us to this very interesting paper.