We thank the two reviewers for their valuable comments and constructive suggestions on the manuscript. Below, we explain how the comments and suggestions are addressed and make note of the revision in the revised manuscript.

Reviewer #2

Global dust cycle and uncertainty in CMIP5 models Chenglai Wu, Zhaohui Lin and Xiaohong Liu
Presented in this study is an evaluation of the global dust cycle simulated by 15 models participating in the Coupled Model Intercomparison Project (CMIP5). The models are compared with each other, aerosol reanalysis data and station observations of dust deposition and concentration. Differences between model simulated dust emission, load, deposition and other aspects are discussed. I believe this is a very valuable study which allows us to better understand the state of the art of dust modelling and better understand the areas where research is needed.

It is probably not surprising that very large differences exist between the model simulated features of the dust cycle, as we already know for some time. It remains a challenge for the models to converge to the truth. This study is a valuable reminder of the challenges ahead and contribution to better quantifying the error bars of the aerosol radiative forcing estimated by climate models.

The paper is well written and logically structured, although a more concise description would be my preference.

There are a number of issues, which I suggest the authors to consider:

Reply: We thank Prof. Yaping Shao for his detailed review and encouraging comments. The text, tables, and figures are revised as he suggested.

Abstract appears to be long.

Reply: Thank you for the comment. We have shortened our abstract by about 20% in the revised manuscript.

L11: address their strengths …
Reply: We have changed "address the strengths and weaknesses of these models" to "address their strengths and weaknesses".

L28-29: deposition is a flux, not a sink

Reply: We thank the reviewer for pointing out this. We have changed "wet deposition is a smaller sink than dry deposition" to "wet deposition is smaller than dry deposition".

Model data: a description of the dust schemes examined in this study is given in this section. These schemes differ in a number of aspects. It would be helpful if some statements were given here, how it is ensured that the comparison is fair. For instance, all models have the same spatial resolution? Do they use the same land surface data?

Reply: We thank the reviewer for this good suggestion. The models included here are those models participating in the CMIP5 and used for historical climate change attribution and future climate projection. CMIP5 provides a well-coordinated framework for climate change experiments and the simulations are included in the IPCC AR5. The experiment design in CMIP5 is described in Taylor et al. (2012). Here, we use the historical experiment which cover the period of 1850 to at least 2005. CMIP5 asks the various model groups around the world to run their models with same forcing data including greenhouse and anthropogenic aerosol and precursor emissions, but the groups are allowed to configure the models with their own resolutions and physical parameterizations including dust emission. For dust emission, land surface data is also different as originally set in the models.

In the revised manuscript, we add some statements in Section 2: "Here we use the historical simulations from 15 CMIP5 models (Table 1). CMIP5 provides a well-coordinated framework for climate change experiments (Taylor et al., 2012). The experiment design in CMIP5 is given in Taylor et al. (2009). The models in CMIP5 were run with their own formulations and resolutions and CMIP5 represented a variety of best-effort attempts to simulate the climate system at the time. CMIP5 results have been included in the Fifth Assessment Report of Intergovernmental Panel on Climate Change (Flato et al., 2013). For the historical experiment, the models were run from 1850 to at least 2005 with same forcing data such as greenhouse gas, solar radiation, and anthropogenic aerosol and precursor emissions (Taylor et al., 2009). All the 15 models used here are fully-coupled models." (Lines 96-105) and "Land cover data are crucial for dust modeling and they also varies in different models. Eleven models use prescribed vegetation or roughness and these data are originated from different studies (an example of this can be seen from the difference between MIROC4h and MIROC5, shown in Section 4.2). In other four models (HadGEM2-CC, HadGEM2-ES, MIROC-ESM, MIROC-ESM-CHEM), dust emission scheme is coupled to dynamic vegetation." (Lines 120-125)

In Section 4.1, I suggest to write explicitly the equation for the global dust budget, and state how the individual terms are computed, so that we can easily understand how the quantities examined are related and why they are chosen. For example, while residence time is important for dust deposition, surface shear stress is important for dust emission, so why is residence time compared here, but not surface shear stress?

Reply: We thank the reviewer for a good suggestion. We have added several equations to explicitly explain the global dust budget and dust residence time in Section 4.1 (Lines 248-270):

First, we present the global dust budgets in CMIP5 models. The key global budget terms include global dust emission (E; kg s^{-1}), dust deposition (D; kg s^{-1}), and dust burden (B; kg), defined respectively as

$$E = \int F_e dS \qquad (1)$$

$$D = \int F_d dS \qquad (2)$$

$$B = \int m_b dS \qquad (3)$$

where F_e is emission flux (kg m^{-2} s^{-1}); F_d is deposition flux (kg m^{-2} s^{-1}); m_b is column dust concentration (kg m^{-2}); S is surface area (m^2). m_b is an integration of dust concentration (C; kg m^{-3}) over the entire column:

$$m_b = \int C dz \qquad (4)$$

The mass equation for dust aerosols around the globe is:

$$\int E dt = \int D dt + \Delta B \qquad (5)$$

Or

$$\bar{E}\Delta t = \bar{D}\Delta t + \Delta B \qquad (6)$$

where ΔB is the change of dust burden between the start time and the end time; $\bar{E}$ is mean global dust emission; $\bar{D}$ is mean global dust deposition; and Δt is the cumulative time. For a long-term period, ΔB is relatively small (i.e., $\Delta B \approx 0$), then

$$\bar{E} = \bar{D} \qquad (7)$$

Dust deposition can be separated into two terms: dry deposition and wet deposition. According to Eq. (6), the mean dust lifetime (also called residence time; $\bar{T}$) can be defined by assuming $\bar{E} = 0$ as:

$$\bar{T} = \frac{\bar{B}}{\bar{D}} \qquad (8)$$

where $\bar{B}$ is mean global dust burden.

L222: may be useful to state, whether we are talking about the same size range. If it is not the same size range, then it is not meaningful to emphasis the range of 735-8196 Tg /a, and a size range correction is necessary. I am not sure whether I missed something, but it is not clear to me whether this is the total emission for the particle size range 0 – 20 microns for all models, or the emission for some models using size range 0 – 20 microns and some 0 – 63 microns.

Reply: We thank the reviewer for pointing out this. The results in this study are based on all the dust particles included in each model. We have this clarified in Section 2: "as only the total dust emission, deposition, and concentration for the whole size range are provided, we are unable to investigate the difference in the mass partitioning among different dust sizes and its evolution, which will be left for future studies" (Lines 138-140)

We agree with the reviewer that comparison of dust emission results should take into account the different size range. Although we are unable to make all the model results comparable, we classify the models into three groups and the dust size range in each group is identical or similar. Therefore, the global dust emissions in each group are comparable (Table 3). Other results are compared by re-ordering the models (Tables 4-5, Figures 3, 6, 8-9). The difference in the dust size range can be recognized if we compare the results from different groups. Accordingly, the statements in the main text have been revised:

a. "The results show that the global dust emission in these models ranges by a factor of 4-5 for the same size range" (Abstract, Lines 17-18)
b. "The dust size ranges considered in the models are not exactly the same. Three models (ACCESS1-0, HadGEM2-CC/ES) consider dust particles with diameter from 0.06 to 63 μm, and estimated global dust emissions range from 2218 to 8186 Tg yr^{-1}. Seven models (GFDL-CM3, four MIROC models and two MRI models) consider dust particles in the diameter of 0.2-20 μm, and they estimate global dust emission in the range of 735-3598 Tg yr^{-1}. The remaining five models consider dust particles in diameter below 10-16 μm and they estimate global dust emission of 1677-3698 Tg. If ACCESS1-0 and HadGEM2-CC/ES are excluded, these estimation here are similar to those of AeroCom models in the similar size range, which gave dust emissions in the range of 514-4313 Tg yr^{-1} (Huneeus et al., 2011)" (Section 4.1, Lines 274-283)
c. "Overall, the models with largest dust size ranges (ACCESS1-0, HadGEM2-CC/ES) simulate smaller fraction of wet deposition (12-19 %) than other models (16-39 %)." (Section 4.1, Lines 311-313)
d. "It is interesting to mention that if ACCESS1-0 with largest dust particle size range (0.06-63 μm in diameter) and largest fraction (91%) for continental deposition is excluded, other six models simulate quite similar fraction of continental deposition (78-83%)." (Section 4.3, Lines 526-529)
e. "The results show that the global dust emission in these models differs much: from 2218 to 8186 Tg yr^{-1} (size range of 0.06-63 μm in diameter), from 735 to 3598 Tg yr^{-1} (size range of 0.06-20 μm in diameter), and from 1677 to 3698 Tg yr^{-1} (size <16 μm in diameter). The global dust emission ranges by a factor of 4-5 for dust particles in the same size range." (Section 5, Lines 599-604)

L245: I recall that in earlier studies dry and wet depositions are about the same order of magnitude, the finding that wet deposition makes only 12-39% of the total deposition is somewhat surprising.

Reply: We thank the reviewer for the comment. We explore the earlier studies on global dust budget. There are several studies which did show wet deposition was about the same order of magnitude of dry deposition (e.g., Luo et al., 2003). However, there are some studies which showed dry deposition is significantly larger than wet deposition. For example, Ginoux et al. (2004) estimated wet deposition accounts for 10 % of total deposition. In addition, the fourteen AeroCom models estimated the fraction of wet deposition ranges from 16 to 66 %. Therefore, the results of CMIP5

(12-39 %) should lie in the meddle to low end of previous estimates. To clarify, we add more discussions on this in the revised manuscript: "Early model studies estimated the fraction of global wet deposition ranges from 10 % (Ginoux et al., 2004) to 49 % (Luo et al., 2003). The 14 AeroCom models estimated the fraction of global wet deposition in the range of 16-66 %. Therefore, this result of 12-39 % lies at the middle to low end of previous estimates." (Lines 305-308)

Section 4.2: some dust emission schemes are already adjusted to satellite observed dust load (so much emission is allowed such that the dust load matches the satellite observed global dust load). I think it would be useful to point out which these models are.

Reply: We thank the reviewer for the suggestion. Six models adopt source erodibility to make the simulated dust patterns close to the observations. We have explicitly mentioned these models in Section 2 (Model data): "In addition, to make the simulated dust patterns close to the observations, the dust schemes in six models (ACCESS1-0, CESM, CSIRO-Mk3-6-0, GFDL-CM3, HaGEM2-CC/ES) further adopt a source erodibility (also called source function) on dust emission. CESM adopts a source erodibility from Zender et al. (2003), and other five models use that of Ginoux et al. (2001)."(Lines 116-120). We also add more discussions in Section 4: "Note CSIRO-Mk3-6.0 and GFDL-CM3, which adopt the same dust emission scheme and source erodibility (Section 2), show similar dust emission regions." (Lines 447-449).

Section 5, Discussion and Conclusion: Experience shows that differences in land surface schemes can have a major effect on dust emission estimates, in particular the simulation of soil moisture. It may be useful to say something about it.

Reply: We thank the reviewer for the comment. Land surface state especially soil moisture is vital for dust emission. The difference in land surface state is involved in the difference in dust emission as well. Following the suggestion, we have added "soil moisture" when discussing "the uncertainty in many aspects of the model" (Lines 667-668). We also add more discussions: "In addition, it is also helpful to setup more constrained experiments to separate the sensitivity of model estimates to individual factors, by varying one single factor such as dust emission scheme (e.g., Wu and Lin, 2013) and land surface scheme (e.g., Lin et al., 2012), or using identical emissions (e.g., Textor et al., 2007)" (Lines 671-675)

L587-589: Again, is the size issue considered in the comparison? Because mass is proportional to size cubed, a small difference in size range can result in huge differences in the dust budget terms. If size correction is not done, then what we can learn from such an assessment study is limited.

Reply: We thank the reviewer for the comment. The size issue is not considered fully in the original manuscript, although we have mentioned its impacts. Now in the revised manuscript, we consider carefully the size difference and compared the model results by classifying the models into three groups according to the size ranges (Tables 3). The description in the main text is revised accordingly (see my reply to the comment on L222 above). Here in the Discussions Section, we clarify: "The results show that the global dust emission in these models differs much: from 2218 to 8186 Tg yr^{-1} (size range of 0.06-63 μm in diameter), from 735 to 3598 Tg yr^{-1} (size range of 0.06-20 μm in diameter), and from 1677 to 3698 Tg yr^{-1} (size <16 μm in diameter). The global dust emission ranges by a factor of 4-5 for dust particles in the same size range." (Lines 599-604) and "We have compared the global dust emission and burden among the models with the same dust size range considered." (Lines 693-694)

I suggest, separate the discussion with conclusion. As it is very a long section.

Reply: we thank the reviewer for the suggestion. We have separated the previous "5. Discussion and Conclusions" section into two sections: "5. Conclusions" and "6. Future work"

Uno et al. (2006 JGR), Textor et al (2006; 2007 ACP) have done model comparisons. These papers may be interesting to this study.

Reply: We thank the reviewer for pointing out these relevant studies to us. These studies have done great job in quantifying the uncertainties in regional or global dust modeling. Much have learned from these studies. In the revised manuscript, we have cited these references with some discussions (Line 54, Lines 669-670, Line 675).