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 #1

This paper examines the dust cycle simulated by 15 models from the Coupled Model Intercomparison Project (CMIP5). Annual mean dust emission, burden, lifetime, deposition, and surface concentration are examined. Large discrepancies are found in global dust emission and burden, while simulated dust deposition and concentration are within a factor of 10 at most stations. Wet deposition is found to contribute about 12-39% of total dust deposition. Overall, the paper is well organized and results are clearly presented. Further improvements are suggested as follows.

Reply: We thank the reviewer for his/her detailed review and helpful comments. The text, tables, and figures are revised as the reviewer suggested.

Major comments:

1. It is not clear why the MERRA2 is included in CMIP5 model comparisons. It's not a fair comparison since meteorological fields and total AOD in the reanalysis are assimilated with observations, but not in CMIP5 models. Although results from the MERRA2 can provide some insights on how well the dust cycle is captured when meteorological fields are constrained with observations, this aspect is not fully discussed in the paper, e.g., how model biases in meteorological fields, such as surface 10 m wind, precipitation, and atmospheric circulation, in CMIP5 models are transformed to biases in dust simulation. I'd suggest either better justifying why the reanalysis is used and the benefits of such a comparison or removing the comparison with MERRA2 results.

Reply: We thank the reviewer for pointing out this. Although there are still gaps in dust fields between MERRA-2 reanalysis and observations, MERRA-2 is a state-of-art aerosol reanalysis and provides a global dust distribution which is better constrained by satellite observations. The comparison of CMIP5 models with MERRA-2 will benefit the identification of model discrepancy.

On the other hand, we also note the dust emission in MERRA-2 is less reliable compared to dust burden and concentration as it is not directly adjusted by the assimilation system. Dust emission in MERRA-2 depends not only on meteorological conditions but also on dust emission parameterizations and thus still of large uncertainty. It is a pity that we can't identify the model biases in dust emission in CMIP5 models and thus we are unable to analyze how model biases in meteorological fields, such as surface 10 m wind, precipitation, and atmospheric circulation, in CMIP5 models are transformed to biases in dust simulation. However, as the development of a referential data is also important for model evaluation, we also mention the limitations when using MERRA-2 data. This will also benefit the development of further dust aerosol reanalysis, for example, by adjoint inversion of dust emission using more specific observations such as lidar observations (Yumimoto et al., 2007).

Because of these benefits, we keep using MERRA-2 to evaluate CMIP5 models. To clarify, in the revised manuscript, we first move the description of MERRA-2 data from Section 2 "Model data" to Section 3.2 "MERRA-2 reanalysis" (under Section 3 "Reference data"). Second, we add more explanations to better justify why MERRA-2 reanalysis is used in Section 3.2: "Because the station observations are limited in space coverage (Figure 1), we also use the aerosol reanalysis from Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) to evaluate the CMIP5 model results." (Lines 213-215). Third, we add the discussions on further development of reanalysis data in Section 6: "It is desirable that future aerosol reanalysis also includes adjoint inversion of dust emissions using more specific observations such as lidar observations as done in Yumimoto et al. (2007)" (Lines 690-692).

2. In section 4, some model discrepancies are attributed to potential causes, such as model biases in vegetation cover (lines 312-314), wind speed and precipitation (lines 343-345). I wonder if it's possible to add analysis to verify these hypotheses by examining a few relevant variables from CMIP5 model output, if available. This will help us better understand the underlying causes of model biases.

Reply: We thank the reviewer for this helpful suggestion. Following the suggestions, we add more analysis to identify the reasons for model discrepancies.

First, we add the analysis on the bare soil fraction, surface wind speed, and soil moisture in the ACCESS1-0, HadGEM2-CC, and HadGEM2-ES, which have similar dust emission parameterizations. The results show HadGEM2-CC/ES simulate much larger bare soil fraction especially in Australia, North America, and South Asia compared to the International Geosphere-Biosphere Programme (IGBP) data used in ACCESS1.0. HadGEM2-CC/ES also simulate significantly larger surface wind speed in Australia and South Asia. In Australia, HadGEM2-CC/ES also simulate slightly smaller soil moisture. These can explain the excessive dust emission in Australia in HadGEM2-CC/ES than ACCESS1-0. Overestimated bare fraction in South Asia and North America can also explain the excessive dust emission simulated by HadGEM2-CC/ES in these regions.

Second, we also compare the leaf area index, surface wind speed, soil moisture in the four MIROC family models (MIROC4h, MIROC5, MIROC-ESM, MIROC-ESM-CHEM) which adopt similar dust emission parameterizations. Instead of using bare soil fraction directly, MIROC models use leaf area index to determine the vegetation cover empirically for dust emission. The results show compared to MIROC5, MIROC4h simulates significantly smaller surface wind speed and adopts a larger leaf area index in the dust source regions, which leads to much smaller dust emission in MIROC4h. Compared to MIROC5, MIROC-ESM and MIROC-ESM-CHEM simulates larger leaf area index in Australia, South America and southern Africa, which can largely explain the difference of dust emissions in these regions among MIROC5 and MIROC-ESM/MIROC-ESM-CHEM.

In the revised manuscript, we have demonstrated these results by adding Figures 4 and 5 and revising/adding corresponding statements for the reasons of model discrepancies in Section 4.2:

a. "The excessive dust emission in Australia from HadGEM2-CC/ES is mainly ascribed to the excessive bare soil fraction simulated by HadGEM2-CC/ES, as indicated by its comparison with International Geosphere-Biosphere Programme (IGBP) data used in ACCESS1-0 (Figure 4a-4c). The overestimation of bare soil fraction in HadGEM2-ES is also illustrated in Collins et al. (2011). In fact, the ACCESS1-0 model that uses the similar dust emission parameterization but with the prescribed vegetation from IGBP simulates a much lower dust emission than HadGEM2-CC/ES. Compared to ACCESS1.0, HadGEM2-CC/ES simulate larger surface wind speed and slightly smaller soil moisture in Australia (Figures 4d-4i), which can also partly explain the larger dust emission in HadGEM2-CC/ES." (Lines 394-403)
b. "The low dust emission in Australia from MIROC-ESM and MIROC-ESM-CHEM is related to the prognostic vegetation used for dust emission. As shown in Figure 5a-5d, MIROC-ESM and MIROC-ESM-CHEM simulate much larger leaf area index compared to the two other MIROC family models (MIROC4h and MIROC5)." (Lines 408-411)
c. "Small dust emission area in MIROC4h may be mainly due to the weaker surface winds in MIROC4h compared to other three MIROC family models (MIROC5, MIROC-ESM, MIROC-ESM-CHEM) (Figure 5e-5f). In the dust source regions (normalized dust emission flux >0.01), the annual mean surface wind speeds are 3.7, 4.4, 4.1, and 4.1 m s^{-1}, respectively in MIROC4h, MIROC5, MIROC-ESM and MIROC-ESM-CHEM. MIROC4h differs much from other three MIROC models in both dynamic core and physical parameterizations (Watanabe et al., 2010, 2011; Sakamoto et al., 2011), which can explain the weakest surface winds in MIROC4h In North Hemisphere, MIROC4h adopts a larger leaf area index than MIROC5, which can also lead to the smaller dust emission area in MIROC4h (Figure 5a-5b)." (Lines 434-444).

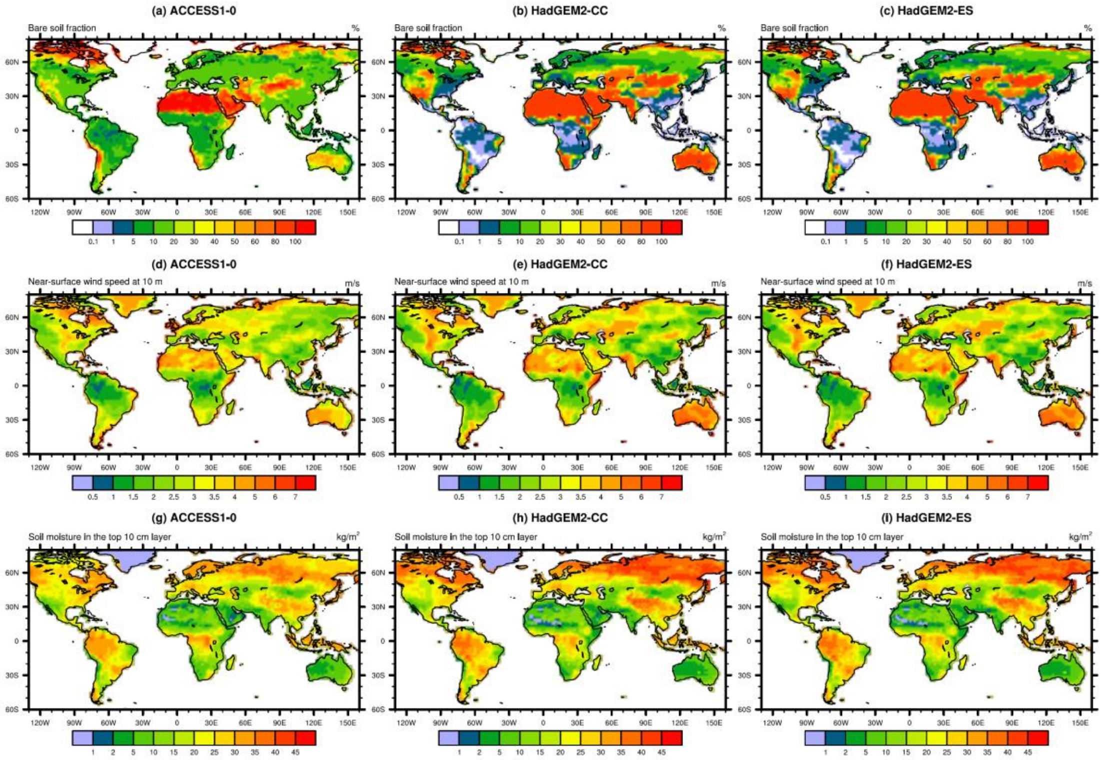

Figure 4. Bare soil fraction (%), near-surface wind speed at 10 m over land (m s^{-1}), soil moisture in the top 10 cm layer (kg m^{-2}) in ACCESS1-0, HadGEM2-CC, and HadGEM2-ES. Note that except bare soil fraction in ACCESS1-0 which is prescribed and set constant for each year, other results are all from model simulations during 1960-2005.

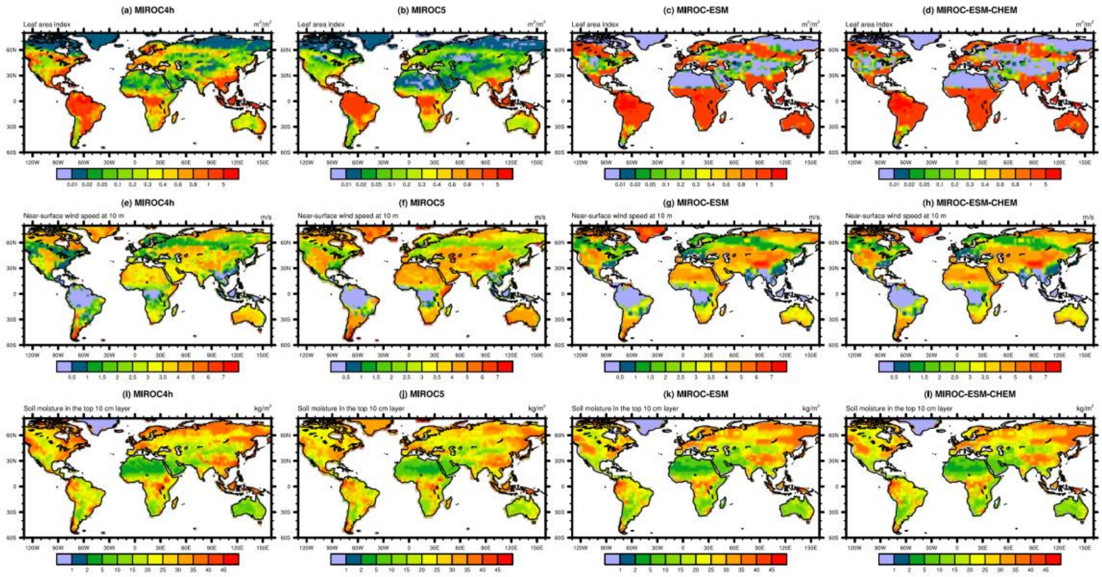

Figure 5. Minimum leaf area index of a calendar year (m^2 m^{-2}), annual mean surface wind speed at 10m (m s^{-1}), and mean soil moisture in the top 10 cm layer (kg m^{-2}) during 1960-2005 in four MIROC family models. For each grid box, monthly mean leaf area index for each month of a calendar year is first derived based on the average of 1960-2005, and then the minimum of leaf area index among these months (i.e., January to December) is plotted.

3. Previous studies of dust simulation in CMIP5 models are thoroughly reviewed in the introduction but not in the result section. Please consider adding discussion and comparisons with current findings in the analysis.

Reply: We thank the reviewer for this good suggestion. The studies of Evan et al. (2014) and Wu et al. (2018b) investigate the dust cycle in specific regions, which can complement our study focusing on dust cycle at global scale. The study of Pu and Ginoux (2018) investigated the dust optical depth in seven CMIP5 models and several of our findings are consistent with theirs. In the revised manuscript, we add discussions on previous studies and comparisons with our study in the result section:

a. “This result is consistent with Pu and Ginoux (2018) that investigated the global distribution of dust optical depth in seven CMIP5 models.” (Lines 333-335)
b. “The extent of “dust belt” can be more clearly seen when we zoom in specific regions such as North Africa (Evan et al., 2014) and East Asia (Wu et al., 2018b). For example, in East Asia, although the CMIP5 models can reproduce the dust emissions in the deserts of northern China and southern Mongolia, they differ greatly in the edges of these deserts, with three models (MIROC5, CanESM2, and CSIRO-MK3-6-0) simulating dust emission over Tibetan Plateau and seven models (e.g., ACCESS1-0) simulating dust emission in the southern part of North China (Wu et al., 2018b).” (Lines 348-354)
c. “The total amount of dust emission in North Africa and East Asia have been presented in Evan et al. (2014) and Wu et al. (2018b), respectively. Here we show the results for all the nine regions in the globe and their comparison.” (Lines 377-

379).

d. "The large scatter of CMIP5 results in North America and Australia is also indicated by dust optical depth, as shown in Pu and Ginoux (2018)" (Lines 388-390)
e. "This is also consistent with the overestimation of dust optical depth in Australia by HadGEM2-CC/ES compared to satellite observations (Pu and Ginoux, 2018)." (Lines 551-553)

Minor comments:

1. Section 2, CMIP5 models have different horizontal resolutions. Did you interpolate model results to the same grid for comparison?

Reply: We interpolate model results to the coarsest resolution among all the models when generating multi-model statistics. To clarify, we add in the revised manuscript: "These models have different horizontal resolutions (Table 1). To generate multi-model statistics of dust emission intensity (Section 4.2), individual model results are interpolated to the coarsest resolution among these models (i.e., 2.8° ×2.8°) using area conserve remapping (http://www.ncl.ucar.edu/Document/Functions/Contributed/area_conserve_remap_Wrap.shtml, accessed on 6 June 2020)." (Lines 143-148)

2. Line 118, I don't think GFDL-CM3 model uses dynamic vegetation to update dust source map. Please double check.

Reply: We thank the reviewer for pointing out this. We have checked GFDL-CM3 doesn't use dynamic vegetation to update dust source map. We have corrected this in the text and Table 1.

3. Line 130, please add "in diameter" after "have the larger size range of 0.0632-63.2 μm"

Reply: Done.

4. Lines 166-167, it seems that dust burden and deposition are not affected by the assimilation of total AOD, right? Please clarify.

Reply: We thank the reviewer for the comment. The MERRA-2 aerosol reanalysis is generated using the increment analysis update procedure. The procedure first derives the AOD increment and then derives 3-dimentional analysis increment for aerosol mixing ratio. This affects the aerosol burden and thus aerosol deposition, but it doesn't affect dust emission. To clarify, in the revised manuscript, we add a description about the aerosol assimilation procedure: "The MERRA-2 aerosol reanalysis uses increment analysis update procedure, which derives 3-dimensional analysis increment for aerosol mixing ratio based on the aerosol optical depth (AOD) analysis increment (Randles et al., 2017). The procedure further affects the aerosol deposition flux." (Lines 223-226)

5. Lines 191-196, can you provide how many years of data are available for dust deposition and surface concentration and add the info to Table 2? Deposition data cover "several to hundreds of years", while CMIP5 data are averaged over 1960-2005. Can you add a short discussion on how the inconsistency of data may affect the comparison?

Reply: We thank the reviewer for a good suggestion. We have examined carefully the periods for the observations used in the study. For surface dust concentration, in the revised manuscript, we have added the information in Table 2.

For the fraction of wet fraction shown in Table 2 in the original manuscript, we find the observation periods were mostly less than 2 years and the observations may be less representative of a climatology. Therefore, in the revised manuscript we don't use the observations for fraction of wet deposition and deleted the comparison results of fraction of wet deposition.

For the deposition flux at 84 stations, the observation periods varied depending on the different observation type. This dataset is directly from AeroCom archive. As some of observation periods were already given in previous studies and the exact periods for ice core data at 5 stations are not available, we prefer to point out these studies to the readers and provide an informative description in the revised manuscript: "The observation periods varied for different stations. Dust deposition from DIRTMAP is from sediment traps and following Tegen et al. (2002), we only use those 41 stations with deployment period larger than 50 days. Original data of Ginoux et al. (2001) contains both measurements and model estimates. We only use the measurements from Ginoux et al. (2001) which consists of 10 stations and the observation periods varied from 1 to 20 years (see sites # 2, 3, 4, 5, 6, 7, 8, 14, 15, 16 in Table 6 of Ginoux et al. (2001)). Data of Mahowald et al. (1999) was derived from ice core data and consists of 6 stations. Except at one of station (i.e., Renland) where the period was 5 years (i.e., 1813-1819 excluding 1816-1817), the exact observation periods at other 5 stations were not provided and generally covered a time slice of tens of years or more for current climate. In addition, Mahowald et al. (2009) further compiled 27 stations from several campaigns and the observation periods mostly covered one to four years." (Lines 178-189)

Although there is mismatch in the temporal coverage between observations and simulation, we mainly focus on the global dust cycle based on multi-year means of both observations and simulations and the impacts on our conclusion due to the mismatch should be not significant. We also add a discussion on the impacts of the inconsistency of data: "We consider the dataset above as a climatology although some of them did not cover a long enough period such as tens of years. Therefore, for the stations with shorter period of observations but large dust variability at interannual to decadal timescales, some model discrepancies may be induced due to the inconsistence between these observations and the model results that are averaged over a period of 45 years. We will discuss this in next sections." (Lines 202-207) and "The biases may also be partly explained by the consistency between the observations and simulations, especially for those observation which were made at a relatively short-term period (one to several years), as mentioned in Section 3.1." (Lines 516-518).

6. Line 220, can you please clarify how dust lifetime is calculated?

Reply: Dust lifetime is defined as the division of global dust burden (Tg) by total deposition (Tg yr^{-1}) and its unit is changed from years to days. In the revised manuscript, we add an equation (Eq. 8) for the definitions of dust lifetime.

7. Lines 226-228, only one model year (2000) is used in AeroCom model intercomparisons, while 46-year averages (1960-2005) are used here. This may contribute to the discrepancy as well.

Reply: We thank the reviewer for the comment. To identify the impacts from different model years, we also compare the CMIP5 model results on year 2000 as AeroCom project. The results show dust emission in year 2000 from CMIP5 models ranges from 773 to 8183 Tg yr^{-1}, and dust burden in year 2000 ranges from 2.7 to 42 Tg. These ranges are similar to those based on 46-year averages (1960-2005), which are 735-8186 Tg yr^{-1} and 2.5-41.9 Tg, respectively. Therefore, the difference in model years selected for comparison could only result in slight difference of comparison results and thus can't change our statements and conclusions.

8. Lines 280-281, "... (Somalia, Ethiopia, and Kenya), East India, and northern part of Indo China Peninsula, which are rarely regarded as potential dust sources". Nogal Valley of Somalia and the Chalbi desert in Kenya are dust sources (Ginoux et al. 2012).

Reply: We thank the reviewer for pointing out this. We are sorry our previous statements were not correct. Now in the revised manuscript, we modified our statements by deleting "the Eastern Africa (Somalia, Ethiopia, and Kenya)" and adding the reference of Ginoux et al. (2012).

9. Lines 291-292, previous studies in addition to "Wu et al. 2018" also identified dust sources in North America, such as Prospero et al. (2002) and Ginoux et al. (2012). Please add more references here.

Reply: We have added more references as suggested by the reviewer.

10. Lines 301-302, "The models consistently simulate the largest dust emission in

North Africa...", is this consistent with AeroCom results?

Reply: Yes. This is consistent with AeroCom results. To clarify, in the revised manuscript, we add a sentence: "This is consistent with previous model intercomparison of AeroCom (Huneeus et al., 2011)" (Lines 381-382)

11. Line 340, 0.1 of erodibility?

Reply: Yes. 0.1 of erodibility is set as a threshold for dust emission occurrence. We have clarified this by changing "a geomorphic source erodibility with a threshold value of 0.1" to "a geomorphic source erodibility with its threshold of 0.1" in the revised manuscript.

12. Line 408, does AOD assimilation affect dust deposition in MERRA2?

Reply: Yes. AOD assimilation affect dust deposition through impacting dust concentrations in MERRA-2. We have clarified this in Section 3.2: "The MERRA-2 aerosol reanalysis uses increment analysis update procedure, which derive 3-dimensional analysis increment for aerosol mixing ratio based on the aerosol optical depth (AOD) analysis increment (Randles et al., 2017). The procedure further affects the aerosol deposition flux." (Lines 223-226)

13. Line 418, "classified into two groups", based on what criteria?

Reply: We classified the stations into the two groups based on their distance from the dust source regions. The stations in the second group are farther from the dust source regions than the first group. In the revised manuscript, the analysis of fraction of wet deposition is removed due to the relatively short period in observation (please see our reply to minor moment #5).

14. Line 457, please add "surface" before "dust concentration"

Reply: Done.

15. Lines 476-477, the vertical distribution of dust could be another reason. The model may simulate higher dust concentration above the surface.

Reply: We thank the reviewer for the comment. We agree with the reviewer that the vertical diffusion of dust may be another reason as MIROC5 and MIROC4h differ much in dynamics and physical parameterizations. Therefore, in the revised manuscript, we add a sentence to clarify this: "Another reason may lie in the vertical diffusion of dust, which also determines the distance of its horizontal transport." (Lines 561-562)

16. Line 496, please add a statistical significance level to correlation coefficients.

Reply: We add the significant test and the results shows the correlation coefficients are all statistically significant at the 0.005 level. In the revised manuscript, we add a sentence to mention this result: "All the correlation coefficients are statistically significant at the 0.005 level" (Lines 581-582)

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).