Dear Referee #2,

We would like to thank you for your comments and suggestions to improve the quality of our manuscript "The Beijing Climate Center Climate System Model (BCC-CSM): Main Progress from CMIP5 to CMIP6" by Tongwen Wu et al. Following your suggestions, we added 5 further figures in the revised manuscript. So, all figures in the first manuscript are renumbered. We have also changed the order of presentation for some subsections in Section 4 (Results). The point-to-point responses to your comments are enclosed in the following.

Best regards,

Tongwen Wu and all co-authors

==========

Response to Anonymous Referee #2

This is a description paper of BCC-CSM2-MR. The upgraded schemes are well described, while evaluations for internal variability and long-term trends are insufficient. Authors emphasized improvements in several aspects. Some of them (e.g., representation of QBO) are remarkable improvements but some others are not very convincing. Major revisions and some additional analyses are required as follows:

Major comments:
The paper lacks the information of ensemble size of each historical runs. The following two comments are related to this issue:

1.1 Is the global warming trend of CMIP6 model shown in Fig. 10 the ensemble-mean value? What is the gray shading? Maximums/minimums of CMIP5? If the ensemble size of CMIP6 is only one member, you cannot deny the possibility that the simulated slowdown after 1998 happened by chance due to natural variability. On the other hand, if it is ensemble-mean value, then the slowdown after 1998 implies it is the result of external forcing. In the latter case, you need to discuss whether the difference in the trends between CMIP5 and CMIP6 is due to the difference in external forcing or the model update.

Response：
The mentioned curves (in old Fig. 10 or new Fig. 4) are from a single realization for both BCC models (CMIP5 and CMIP6). After the initial submission of the manuscript, two other members are now available for our CMIP6 exercises, which are added in the new Fig. 4. Gray shaded area shows the spread calculated from 31 CMIP5 models. The caption in Fig. 10 of the initial manuscript was not clear, and is now modified in the new manuscript. Concerning the global warming hiatus or slowdown after 1998 in old Figure 10 (or new Fig.4, it is the overlay

of internal variability and long-term trends. We need to explore it further in the future. Among the three members, two show some features of hiatus.

1.2 Is the improvement of MJO shown in Fig. 7 robust? A correlation coefficient depends on a sample size. Are the sample sizes (the total numbers of years) you used for the analyses of the observation, BCC-CLM1.1m, and BCC-CLM2-MR comparable? If the improvement is true, what do you think is a factor improving the representation of MJO? New convection scheme?

Response: Yes, MJO activities largely depend on period and time length of analysis. But we believe that our analysis in Figure 7 is robust and consistent: all data for observation, BCC-CSM1.1m, and BCC-CSM2-MR are from the same period from 1997 to 2005. We believe that the improvement of MJO simulation in BCC-CSM2-MR is true and can be possibly attributed to deep convection scheme (Table 2). Further investigations are underway.

2 It seems that the results in Fig. 8 and Fig. 11 are inconsistent. Is it due to the difference between the reference datasets (NSIDC, ECMWF forecasts, and Hadley Centre Sea Ice)? Please add some explanation to figure out it and the reason why you used those datasets.

Response:
Figure 11a and 11b (in the old version of manuscript) shows seasonal cycle of sea-ice extent climatology, in which the observation data is derived by National Snow and Ice Data Center (NSIDC; Fetterer et al., 2002) and are directly downloaded from https://svn-ccsm-models.cgd.ucar.edu/tools/proc_ice/trunk/ice_diag/data/SSMI.ice_extent.1981-2005.monthly.regional.txt. Figure 11c and 11d (in the old version of manuscript) shows seasonal cycle of sea-ice thickness averaged for the Northern Hemisphere and the Southern Hemisphere, which is computed based on 1980-2005 global monthly 1/4 °x1/4 °gridded dataset based on European Center for Medium-Range Weather Forecast (Tietsche, et al., 2014).

Figure 8 (in the initial manuscript) presents long term change of sea-ice extent from 1851 to 2012, the observed data are computed based on global monthly 1 °x1 ° gridded dataset of Hadley Centre Sea Ice and Sea Surface Temperature (Rayner et al., 2003).

The observed sea-ice extent in Figs. 11 and 8 (in the initial manuscript) doesn't come from a same source. NSIDC data are mainly derived from satellite observations, so with a higher quality. But they cover a too-short period. This explains why Fig. 8 in the initial manuscript the long-lasting Hadley Centre Sea Ice and Sea Surface Temperature data set. For the sake of consistency, we decided to use only Hadley center data throughout the manuscript.

3 Section 4.2: Please discuss the SST cold bias in the equatorial Pacific and the double ITCZ problem. Evaluations for the subsurface ocean (temperature, zonal currents) are also required.

Response:
We added two more figures to discuss the SST cold bias and the interannual variations of NINO3.4 SST in the equatorial Pacific. As length of the paper is limited and there are already 20 figures, the subsurface ocean (temperature, zonal currents) will be explored in the future.

4 Section 4.6: There is a large seasonal cycle in the East Asian climate. Analyses should be done for each season, especially for JJA and DJF.
Response: We added a further figure to discuss the DJF and JJA precipitation in East Asia.

5 Comparison of ENSO representation (NINO3.4 time series, amplitude, spatial pattern etc.) is necessary to evaluate the model performance.
Response: We added a figure to show behaviors of Nino3.4 time series. The detail evaluation for ENSO representation will be analyzed in other papers.

Minor issues:
6 L19: models -> model's
7 L27-28 "Compared to BCC CMIP5 models, BCC CMIP6 models show: : :": The expression "models" is inappropriate. This paper compared only one model for each (BCC-CSM1.1m for CMIP5 and BCC-CSM2-MR for CMIP6).
8 L41: More -> more
9 L53: The full name of CMIP5 and Taylor et al. (2012) should appear at L42.
10 L59: Please comment on Section 5 and 6.
11 L64: Please add the full name for NCAR.
12 L64: Coupler -> coupler
Response: 6-12, all done.

13 L65-66: Are tuning parameters also the exactly same between the two models?
Response: Some physical schemes in the two models are not the same. Tuning parameters are not exactly the same, neither.

14 L79: Please add the level of the top layer.
Response: Added as suggested. The tops of atmosphere in BCC-CSM1.1m and BCC-CSM2-MR are at 2.917hPa and 1.459 hPa, repsecctively.

15 L240: What is the CEVSA model?
Response:CEVSA model is the carbon exchange between vegetation, soil and the atmosphere (CEVSA) model

16 L243: Table 2 -> Table 3
Response: Modified.

17 L362: 2014 -> 2005? The end of the CMIP5 historical run is 2005.
Response: Modified. In Table 4, and Fig.1 (renumbered to Fig.3 in the revised manuscript), model results are for the period 1986–2005, while the available CERES-EBAF data are for

2003–2014.

18 L371 "1986-": "1985-" in the caption of Fig. 1
Response: Modified.

19 L374-376: In the mid latitude, discrepancies in the two observations are also large. You cannot discuss the difference between the two models.
Response: Modified as suggested.

20 L396: Precipitation data diagnosed in reanalysis is not necessarily correct. You had better use the observation-based dataset.
Response: Modified as suggested.

21 L406: Figure 4s -> Figure 4
Response: Modified.

22 L409: Were there any reason you used the two different reanalysis datasets (NCEP and ERA-interim) for each analysis?
Response: Modified. All the reanalysis data used to draw figures are now replaced by ERA-interm.

23 L437 "forcing is less adequate": What forcing?
Response: "forcing" denotes gravity wave forcing. Modified to "The amplitudes of the QBO cycles in the simulation are weaker than in the reanalysis, which is possibly due to inadequate gravity wave forcing to drive the QBO."

24 L452: improvementcompared -> improvement compared
Response: Modified.

25 L478-481: Is the lower and deeper NADW better? Please show the observation based values.
Response: Modified. The observation-based value is 25 Sv in Talley et al. (2013).

26 L486 "from 60S to 60N": Why did not you use the average from 90S to 90N?
Response: Only the area from 60°S to 60°N is averaged in Fig. 10 (renumbered to Fig.4 in the revised manuscript). This is mainly motivated by the consideration that HadCRUT4 dataset had very few observations in polar regions in earlier 20th century.

27 L491, L500: HadCRU -> HadCRUT4
Response: Modified.

28 L503: Figure 2c -> Figure 10?
Response: Modified.

29 L529: abruptCO2 -> abrupt CO2
Response: Modified.

30 L535 "the TCR of the new version model BCC-CSM2-MR is lower than BCCCSM1.1m,": lower -> higher? It is inconsistent with L527-528.
Response: The TCR of the new version model BCC-CSM2-MR is lower than BCC-CSM1.1m, but the ECS of BCC-CSM2-MR is slightly higher than BCC-CSM1.1m.

31 L566-567: Are there any impact of the upgraded land surface scheme on the improvement of the rainfall diurnal cycle?
Response: We didn't explore this issue, but we think the upgraded land surface scheme has minor role for the improvement of the rainfall diurnal cycle.

32 L616-620: Please see 31.
Response: Modified.

33 Table 3 "(Guenther et al., 2012)": Please check the font.
Response: Modified.

34 Table 4: What is the value of the net energy at TOA for OBS? 0.81?
Response: The net energy at TOA is 0.81 W m^{-2} for CERES-EBAF and 5.73 W m^{-2} for CERES data.

35 Table 4, Notes, "1981-2014": 1981-2005 for BCC-CSM1.1m? Why are the periods slightly different among the figures or tables (1985-2005 for Fig. 1, 1986-2005 for Fig. 2-3 and 5, 1980-2005 for Fig. 4, 8-9, 13)?
Response: Yes, that's done. We apologize for the inconsistency. However, In Table 4 and Fig. 1 (renumbered to Fig.3 in the revised manuscript), model data are the mean from 1986 to 2005, while observation is only available for 2003–2014.

36 Figure 7 d-f: Please show the range of longitude used for the average to obtain the latitude-time section.
Response: Modified.

37 Figure 14: The result of BCC-CSM1.1m (BCC-CSM2-MR) should be in the bottom (middle) as same as in the other figures.
Response: Yes, that's done. We appologize for the inconsistency.