Anonymous Referee #2

This study reports the evaluation against measurements of the output from a dynamical downscaling link between the global Community Earth System Model (CESM) and the WRF-CMAQ modelling system over the East Asia region for a number of meteorological and air quality composition variables. The climatological simulations were for RCP4.5 for 2006-10 and the air quality applications were for winter and summer months in 2013 (principal compositional variables of interest: PM2.5 and O3). The authors report satisfactory prediction of major meteorological variables, although see the first of the general comments below. The paper reports on a major piece of work, with what appear to be generally appropriate methods, and is within the scope for consideration of publication in GMD.

Response: We thank Referee #2 for the constructive comments. Please see below our point-by-point replies to other comments.

General comments

(1) The description of the downscaling (P5-6) indicates that aspects of it involves significant bias corrections, so to what extent is it valid to judge model performance by model-observation statistics? For example, it is stated on P8, lines 1-6, that the improved statistical performance of the modelling approach used in this study may be related to the bias-correction applied. If a bias correction is applied then presumably we expect better model-observation statistics, so have we learned anything fundamental about the model performance by these comparison statistics?

Response: While using bias-corrected ICs/BCs does improve WRF-CMAQ's model performance, it does not make model-observation comparison invalid. While meteorological reanalysis data were used to correct biases in meteorological ICs/BCs based on CESM-NCSU's results and satellite retrievals of O_3 were used to constrain their upper boundary conditions, observational data were used for model performance evaluation. Because GCMs generally suffer from systematic biases to a certain extent, bias correction to the GCM (i.e., CESM) boundary conditions was applied in this study to improve the model performance in simulating regional climate. By comparing to the traditional approach without GCM bias corrections, previous studies (Xu and Yang, 2012; Bruyère et al., 2014; Done et al., 2015) have shown that the improved dynamical downscaling method with GCM bias corrections greatly improves the downscaled climate. The bias-correction technique is also used in the NCAR CESM global bias-corrected CMIP5 output to support WRF/MPAS research (https://rda.ucar.edu/datasets/ds316.1/). Also note that the bias correction is applied to the ICs/BCs, rather than the model results. So, the model-observation comparison will provide insights into the model's capability in capturing observations.

(2) The model-observation statistics should include RMSE instead of, or in place of, the normalized mean error (NME). The former is the statistic usually used alongside the correlation coefficient and mean bias (or normalised mean bias) in the suite of statistics that captures the spectrum of model performance characteristics.

Response: As suggested, we have added the root mean square error (RMSE) in the statistics tables (Table 3, 4 and S3) in place of the normalized mean error (NME), and added the mean absolute gross error (MAGE) in Table 3. The model performed well for T2 and RH2, with MBs of -0.6 ℃ and 0.8%, correlation coefficients of 0.97 and 0.72, MAGEs of 2.4 ℃ and 9.7%, and RMSEs of 3.2 ℃ and 12.6%, respectively. WS10 was moderately overpredicted by 22.2%, with an MB of 0.6 m/s, an MAGE of 1.2 m/s and a RMSE of 1.6 m/s. We have added this in the Section 3.1 of the revised manuscript.

(3) In general, the discussion of model output against meteorological and compositional variations is (i) vague, i.e. non-quantitative (using phrasing like agreed well, satisfactory, etc.), and (ii) lacking explanatory insight, i.e. lists of potential reasons for discrepancy are given which could be written down as potential explanations without needing to do these comparisons. The authors should endeavour to provide more quantitative assessments of model performance, including how their mod-obs statistics compare with expectation and with other studies, and also to provide some more informed analysis of what is the driving explanation for mod-obs discrepancies for particular variables or circumstances.

Response: As suggested, we have provided more quantitative assessments of model performance in terms of MB, NMB, or RMSE in the abstract, result and conclusion sections of the revised manuscript.

The model biases or errors can be attributed to many factors. Pinpointing the exact causes is not a trivial effort, often involving large amounts of sensitivity simulations and in some cases, model further development and improvement that are not permitted with our very limited resources. Nevertheless, we have provided more insights into the model's performance statistics and how they are compared with other studies, wherever possible. For example, we have compared the performance of several meteorological variables with the benchmarks suggested by Emery et al. (2001) in the Section 3.1 of the manuscript. Emery et al. (2001) suggested the benchmarks for satisfactory performance for T2 (MB within ±0.5 ℃, MAGE of ≤ 2.0 ℃) and WS10 (MB within ± 0.5 m/s, MAGE and RMSE of 2.0 m/s). In the climatological application, the MB and MAGE of T2 and the MB of WS10 are close to the benchmark, the MAGE and RMSE of WS10 are within the benchmark, and hence the performance is deemed acceptable.

We have also compared the CMAQ performance of chemical predictions in this study with other studies, as shown in the Section 3.2 of the manuscript. The revised text is as follows:

The CMAQ performance of chemical predictions in this study was comparable to or even better than those of other air quality studies over East Asia (Wang et al., 2009; 2012; Liu et al., 2010; Zheng et al., 2015; Hu et al., 2016; Liu et al., 2016; Zhang et al., 2016a). This study predicted relatively well for most chemical species in most months. Compared with other regional modeling studies, WRF-CMAQv5.0.2 used in this study outperformed MM5/CMAQv4.6, which tend to underpredict the surface concentrations of major species with NMBs generally greater than -40% and overpredict surface O_3 concentrations in most months with NMBs generally higher than 20% over East Asia according to the evaluation results of Zhang et al. (2016a). A relatively good performance of CMAQv5.0.1 was also reported by Hu et al. (2016). Global models such as GEOS-Chem and CESM tend to underpredict $PM_{2.5}$ concentrations (by about -50% as reported by Jiang et al., 2013) and overpredict O_3 concentrations (by about 50% as reported by He and Zhang, 2014; Wang et al., 2013) in China/East Asia because of relatively coarse grid resolution and limitations in some model treatments (e.g., missing emissions of unspeciated primary $PM_{2.5}$, and discrepancies in surface layer height and vertical mixing).

Specific comments

P1, L27: The phrasing "The model showed good ability to predict PM2.5 . . ..and O3. . ." is non-quantitative and vague.

Response: The above sentence has been revised to include more quantitative assessment as follows:
The model showed good ability to predict $PM_{2.5}$ in winter (with a normalized mean bias (NMB) of 6.4% in 2013) and O_3 in summer (with an NMB of 18.2% in 2013) in terms of statistical performance and spatial distributions.
In addition, we have added this in the abstract of the revised manuscript.

P4, L20: Rephrase as "Several modifications in model. . ."

Response: Revised as suggested.

P7, L16: Although the acronym TOR is defined here, there needs to be some further explanation of what it means in practice, particularly in the context of its relevance to model performance evaluation.

Response: TOR represents tropospheric ozone residual or column abundance of O_3. We have clarified this in the Section 2.2 of the revised manuscript.

P12, L18: "were much closer to. . .

Response: Revised as suggested.

Reference

Bruyere, C. L., Done, J. M., Holland, G. J., and Fredrick, S.: Bias corrections of global models for regional climate simulations of high-impact weather, Clim. Dynam., 43, 1847–1856, doi:10.1007/s00382-013-2011-6, 2014.

Done, J. M., Holland, G. J., Bruyere, C. L., Leung, L. R., and Suzuki-Parker, A.: Modeling high-impact weather and climate: lessons from a tropical cyclone perspective, Climatic Change, 129, 381–395, doi:10.1007/s10584-013-0954-6, 2015.

Emery, C., Tai, E., and Yarwood, G.: Enhanced meteorological modeling and performance evaluation for two texas episodes, Report to the Texas Natural Resources Conservation Commission, prepared by ENVIRON, International Corp., Novato, CA, available at: http://www.tceq.state.tx.us/assets/public/implementation/air/am/contracts/reports/mm/EnhancedMetModelingAndPerformanceEvaluation.pdf, 2001.

He, J., and Zhang, Y.: Improvement and further development in CESM/CAM5: gas-phase chemistry and inorganic aerosol treatments, Atmos. Chem. Phys., 14, 9171–9200, doi:10.5194/acp-14-9171-2014, 2014.

Hu, J., Chen, J., Ying, Q., and Zhang, H.: One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system, Atmos. Chem. Phys., 16, 10333–10350, doi:10.5194/acp-16-10333-2016, 2016.

Jiang, H., Liao, H., Pye, H. O. T., Wu, S., Mickley, L. J., Seinfeld, J. H., and Zhang, X. Y.: Projected effect of 2000-2050 changes in climate and emissions on aerosol levels in China and associated transboundary transport, Atmos. Chem. Phys., 13, 7937–7960, doi:10.5194/acp-13-7937-2013, 2013.

Liu, X., Zhang, Y., Cheng, S., Xing, J., Zhang, Q., Streets, D. G., Jang, C., Wang, W., and Hao, J.: Understanding of regional air pollution over China using CMAQ, part I performance evaluation and seasonal variation, Atmos. Environ., 44, 2415–2426, doi:10.1016/j.atmosenv.2010.03.035, 2010.

Liu, X., Zhang, Y., Zhang, Q., and He, M.: Application of online-coupled WRF/Chem-MADRID in East Asia: Model evaluation and climatic effects of anthropogenic aerosols, Atmos. Environ., 124, 321–336, doi:10.1016/j.atmosenv.2015.03.052, 2016.

Wang, K., Zhang, Y., Jang, C., Phillips, S., and Wang, B.: Modeling intercontinental air pollution transport over the trans-Pacific region in 2001 using the Community Multiscale Air Quality modeling system, J. Geophys. Res. Atmos., 114, D04307, doi:10.1029/2008JD010807, 2009.

Wang, K., Zhang, Y., Nenes, A., and Fountoukis, C.: Implementation of dust emission and chemistry into the Community Multiscale Air Quality modeling system and initial application to an Asian dust storm episode, Atmos. Chem. Phys., 12, 10209–10237, doi:10.5194/acp-12-10209-2012, 2012.

Wang, Y., Shen, L., Wu, S., Mickley, L., He, J., and Hao, J.: Sensitivity of surface ozone over China to 2000-2050 global changes of climate and emissions, Atmos. Environ., 75, 374–382, doi:10.1016/j.atmosenv.2013.04.045, 2013.

Xu, Z., and Yang, Z.: An Improved Dynamical Downscaling Method with GCM Bias Corrections and Its Validation with 30 Years of Climate Simulations, J. Climate, 25, 6271–6286, doi:10.1175/JCLI-D-12-00005.1, 2012.

Zhang, Y., Zhang, X., Wang, L., Zhang, Q., Duan, F., and He, K.: Application of WRF/Chem over East Asia: Part I. Model evaluation and intercomparison with MM5/CMAQ, Atmos. Environ., 124, 285–300, doi:10.1016/j.atmosenv.2015.07.022, 2016a.

Zheng, B., Zhang, Q., Zhang, Y., He, K. B., Wang, K., Zheng, G. J., Duan, F. K., Ma, Y. L., and Kimoto, T.: Heterogeneous chemistry: a mechanism missing in current models to explain secondary inorganic aerosol formation during the January 2013 haze episode in North China, Atmos. Chem. Phys., 15, 2031–2049, doi:10.5194/acp-15-2031-2015, 2015.