

Referee's comments are in **red**, our reply in black, quotes in the revised manuscript in **blue**.

Referee 1's comments

General comments

Wang et al. presents an interesting study about comparison between two downscaling approaches i.e. statistical downscaling (ISIMIP) and dynamical downscaling (WRF) combined with bias correction of the future projection of ESM scenarios (i.e. RCP4.5, RCP8.5, G4) over Beijing provincial region. The idea of comparing statistical with dynamical downscaling is a novel path of research. The authors focus on mean temperature, humidity, and wind speed, which are all relevant for climate impacts. The manuscript is generally clearly written and well structured, and the analysis is done in an organized way. However, some parts need to be clarified to easily understand some contents of the manuscript. I think the manuscript is scientifically sound and merits publication after some minor revisions based on my specific comments below.

Reply: We thank the reviewer for their overall positive and very constructive response, along with their helpful suggestions for improving the manuscript. Our response is given below.

Specific comments:

1. In table 1, assessment on the performance of ISIMIP and WRF with bias correction is done only through RCP4.5 scenario. Is there any reason why RCP8.5 is excluded in this part?

Reply: In the assessment part, we only want to see the performance of QDM methods. So we choose the RCP4.5 scenario during 2008-2017 as our reference to simulate. There is no statistical difference between RCP4.5 and RCP8.5 in the 2007-2017 period so it does not matter whether we use RCP4.5 or RCP8.5.

2. This research used ERA5 reanalysis data as the proxy of observation. The authors need strong justification of choosing ERA5 reanalysis in this research as reanalysis data itself might contain some degree of bias.

Reply: Thanks for your comment. Choosing a reanalysis data is an important thing. ERA5 has a finer temporal resolution (hourly) and spatial resolution (0.25°). We add these sentences in Section 2.1

Here we use ERA5 reanalysis data as our reference. This fifth generation ECMWF atmospheric reanalysis of global climate combining huge amounts of historical observations into global estimates by advanced modelling and data assimilation

systems (Hersbach et al., 2020) has been widely used for meteorological data analysis (Chen et al., 2021; Huo et al., 2021; Lee et al., 2022; Zhang et al., 2022). The performance of ERA5 for temperature (Gong et al., 2020), relative humidity (Zhang et al., 2021) and wind speed (Yu et al., 2019) analyzed over China suggests it well reproduces the observed meteorological data in climatology and interannual variations.

References

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J. et al: The ERA5 global reanalysis, *Q. J. R. Meteorol. Soc.*, 146, 1999-2049, 2020.

Chen, S., Cao, R., Xie, Y., Zhang, Y., Tan, W., Chen, H., Guo, P., and Zhao, P.: Study of the seasonal variation in Aeolus wind product performance over China using ERA5 and radiosonde data, *Atmos. Chem. Phys.*, 21, 11489–11504, <https://doi.org/10.5194/acp-21-11489-2021>, 2021.

Huo, L., Wang, J., Jin, D., Luo, J., Shen, H., Zhang, X., Min, J., and Xiao, Y.: Increased summer electric power demand in Beijing driven by preceding spring tropical North Atlantic warming, *Atmospheric and Oceanic Science Letters*, 15 (1), 100146, 2022.

Lee, S., Lung, S., Chiu, P., Wang, W., Tsai, I., and Lin, T.: Northern Hemisphere Urban Heat Stress and Associated Labor Hour Hazard from ERA5 Reanalysis, *Int. J. Environ. Res. Public Health*, 19, 8163. <https://doi.org/10.3390/ijerph19138163>, 2022.

Zhang, G., Azorin-Molina, C., Wang, X., Chen, D., McVicar, T., Guijarro, J., Chappell, A., Deng, K., Minola, L., Kong, F., Wang, S., and Shi, P.: Rapid urbanization induced daily maximum wind speed decline in metropolitan areas: A case study in the Yangtze River Delta (China), *Urban Climate*, 43, 101147, <https://doi.org/10.1016/j.uclim.2022.101147>, 2022.

Gong, Y., Yang, S., Yin, J., Wang, S., Pan, X., Li, D., and Yi, X.: Validation of the Reproducibility of Warm-Season Northeast China Cold Vortices for ERA5 and MERRA-2 Reanalysis, 61 (9), 1349-1366, <https://doi.org/10.1175/JAMC-D-22-0052.1>, 2022.

Zhang, J., Zhao, T., Li, Z., Li, C., Li, Z., Ying, K., Shi, C., Jiang, L., and Zhang, W.: Evaluation of Surface Relative Humidity in China from the CRA-40 and Current Reanalyses, *Advances in Atmospheric Sciences*, 38, 1958-1976, 2021.

Yu, J., Zhou, T., Jiang, Z., and Zou, L.: Evaluation of Near-surface Wind Speed Changes during 1979 to 2011 over China Based on Five Reanalysis Datasets, *Atmosphere*, 10, 804, <https://doi.org/10.3390/atmos10120804>, 2019.

3. Step 1 of ISIMIP statistical downscaling and bias correction is simplified and make it unclear. Please provide a more detailed description of it.

Reply: Done. We have changed sentences in Section 2.3.

2.3 ISIMIP statistical downscaling and bias correction

This method corrects daily variability on the premise that the monthly trend of the modeled variable is unchanged (Hempel et al., 2013). Here, we take the data from one grid point and for some single month as an example to illustrate the procedure. It includes three steps:

Step 1: We firstly bilinearly interpolate the model data to the same grid points of reanalysis data before bias correction.

Step 2: Monthly bias corrected data are found by multi-year averaged difference between the model output and reanalysis data in our referenced period.

$$M_m^* = \overline{R_m} - \overline{M_m} + M_m \quad (2)$$

The M_m^* is the bias-corrected monthly data, $\overline{R_m}$ and $\overline{M_m}$ are the multi-year averaged values in this month from reanalysis data and model data during the reference period, respectively. M_m is the modeled monthly data. The subscript m represents monthly. In this step, ISIMIP does not correct the daily variability of modeled data.

Step 3: Correct the modeled daily variability to a linear regression residual.

$$\Delta M_d^* = \overline{B} \times (M_d - M_m) \quad (3)$$

The ΔM_d^* is the bias-corrected residual daily data from model. M_d is the modeled daily data. The subscript d represents daily. $(M_d - M_m)$ represents the modeled daily residual values in this month, and residual of reanalysis data can be obtained in the same way. \overline{B} is the linear regression coefficient of daily residual values between reanalysis data and model data during our referenced period. Then, we can get the bias-corrected modeled daily data:

$$M_d^* = M_m^* + \Delta M_d^* \quad (4)$$

The M_d^* is the bias-corrected daily data of model. Therefore ISIMIP corrects the monthly mean and its daily variability. Here, we use the ERA5 reanalysis data as reanalysis data in our study. For convenience we use the term ISIMIP-ESM to denote the output from the ESMs after applying the ISIMIP statistical downscaling and bias correction methodology.

4. The statement “QDM is similar to QM but is non-stationary”. In what sense it is non-stationary?

Reply: Non-stationary means that the mean value or the probability density function (pdf) of a time series changes over time. QM is used to correct the pdf of model data under any scenario to be consistent with the pdf of observed data in the reference period. But the pdf in the future scenario is not the same as in the historical reference period. Therefore, QDM takes the difference between the pdf of model data and observed data in the reference period into account, and applies this difference to the

model data in the future scenario. In this sentence, what we want to express is that QM assumes that the pdf of model data does not change over time (under different scenarios).

5. From figure 3i we see that the WRF simulations have spatial patterns very different from ERA5. Moreover, figure 4 also shows that the pdf of WRF deviates from ERA5 significantly. Are there any reasons behind these findings?

Reply: From the figure 3i and 4, we can see that WRF overestimates the wind speed. In our paper discussion part (line 402-403 and 408-411), we only mentioned the reason for distribution of wind speed from WRF results. A new surface drag parameterization scheme was launched after WRFv3.4, although we chose not to use it here. Zheng et al. (2016) evaluated the effect of two surface drag parameterization schemes on the surface wind speed in Beijing, and found that terrain correction can improve the simulation of wind speed in valley areas. However, WRF still has a systematic overestimation of wind speed. We add the following sentence in the discussion section.

Overestimated surface wind speed in WRF is caused by using smoothed topography in the model (Jimenez and Dudhia, 2012).

References

Jimenez, P., and Dudhia, J.: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model, *J. Appl. Meteorol. Climatol.*, 51 (2), 300-316, 2012.

Zheng, Y., Liu, S., Miao, Y., and Wang, S.: Effects of different topographic correction methods on the simulation of surface wind speed and temperature in parameterization scheme of YSU boundary layer, *Chinese Journal of Geophysics*, 59 (3), 803-815, 2016.

6. Authors use three different terms for ensemble i.e., “ensemble-mean”, “multi-ensemble mean”, “multi-model ensemble mean”. Are they referring to the same metric?

Reply: Yes. We have changed each usage to “ensemble mean”, which means the ensemble mean of four downscaled ESMs results.

7. In Figure 5, the caption states “labelled as ERA5”, but in the figure we cannot find the label “ERA5”. Do you mean the “observed” label? Figure 8 seems correct. Please make it consistent.

Reply: Yes. We have changed the “observed” label to the “ERA5” in the revised figure 5.

8.It is stated that “wind speed of all four ESMs outputs have correlation coefficients <0.1 with ERA5”. I would suggest that authors describe in the discussion section the consequence of this low value to the results of the analysis.

Reply: Done. We add some sentences after line 404.

Correlation coefficients of wind speed between WRF downscaling results and ERA5 for all four ESM raw results are all < 0.1. Zha et al. (2020) and Jiang et al. (2017) also found similar low correlations. Zha et al. (2020) projects the near-surface wind speed over eastern China based on a CMIP5 dataset, and found that 18 of the 24 ESMs analyzed show negative correlations with observed wind speed during 1979-2005. The low correlations are to be expected when considering variability in simulated weather at high temporal resolution. Jiang et al. (2017) found that differences in CMIP5 model wind responses to the East Asian monsoon in China are related to model parameterization and horizontal resolution.

References

Zha, J., Wu, J., Zhao, D., and Fan, W.: Future projections of the near-surface wind speed over eastern China based on CMIP5 datasets, *Climate Dynamics*, 54, 2361-2385, 2020.

Jiang, Y., Xu, X., Liu, H., Dong, X., Wang, W., and Jia, G.: The underestimated magnitude and decline trend in near-surface wind over China, *Atmospheric Science Letters*, 18 (12), 475-483, 2017.

9.The statement “Humidity and windspeed anomalies from ISIMIP appear somewhat spatially anti-correlated, while for WRF there are no particular patterns” need to be clarified. What does the term “anti-correlated” mean? Is it a negative correlation?

Reply: This sentence is based on the G4-RCP8.5. This sentence has been rewritten.

Humidity anomalies from ISIMIP have a difference under G4 relative to RCP8.5 in the southwest of the domain, where windspeed anomalies show an obvious positive change, while for WRF there are no particular patterns.

10.In Table 3, I would suggest using Asterisk sign to indicate that the differences are significant, instead of writing the number in bold.

Reply: Sorry we prefer to use bold – as is the case on many articles. But perhaps it can be left to the journal style and editor decision on what approach to take.

Referee 2's comments

General Comments:

This paper looks at statistical and dynamical downscaling for the Beijing region under three future climate scenarios. It also appears to be the first time that dynamical downscaling has been used with a geoengineering scenario. The paper is fairly straightforward but seems to be a good starting point for future work on geoengineering modeling. Probably the biggest question I have is regarding the choice of analyzed variables. The authors compare temperature, humidity, and wind speeds under the different scenarios; temperature is of course important, but, at least to me, the latter two are not so much. I wonder if the authors thought about analyzing precipitation instead.

Reply: As you mentioned, the effects of SAI on precipitation is also very important. Precipitation has often been studied and in general changes have been widely reported, not least in our own work e.g. in a $\frac{1}{4}^\circ$ global food and streamflow analysis (Wei et al 2018 in *ACP* doi: 10.5194/acp-18-16033-2018). But in this analysis, we wanted to focus on the meteorological elements that impact human health. Temperature, humidity and wind speed are factors that affect apparent temperature, especially in extreme events for big cities. Humidity and windspeed have been far less studied in geoengineering scenarios than precipitation, and so does merit some analysis.

Another general question I have is: how is the ISIMIP spatial downscaling done? The two steps in Section 2.3 address bias correction and temporal downscaling, but how do you go from the coarse ESM grids to the ERA5 grid?

Reply: We have rewritten the Section 2.3 in more detail.

Step 1: We firstly bilinearly interpolate the model data to the same grid points of reanalysis data before bias correction.

Step 2: Monthly bias corrected data are found by multi-year averaged difference between the model output and reanalysis data in our referenced period.

$$M_m^* = \overline{R_m} - \overline{M_m} + M_m \quad (2)$$

The M_m^* is the bias-corrected monthly data, $\overline{R_m}$ and $\overline{M_m}$ are the multi-year averaged values in this month from reanalysis data and model data during the reference period, respectively. M_m is the modeled monthly data. The subscript m represents monthly. In this step, ISIMIP does not correct the daily variability of modeled data.

Step 3: Correct the modeled daily variability to a linear regression residual.

$$\Delta M_d^* = \overline{B} \times (M_d - M_m) \quad (3)$$

The ΔM_d^* is the bias-corrected residual daily data from model. M_d is the modeled daily data. The subscript d represents daily. $(M_d - M_m)$ represents the modeled daily residual values in this month, and residual of reanalysis data can be obtained in the

same way. \bar{B} is the linear regression coefficient of daily residual values between reanalysis data and model data during our referenced period. Then, we can get the bias-corrected modeled daily data:

$$M_d^* = M_m^* + \Delta M_d^* \quad (4)$$

The M_d^* is the bias-corrected daily data of model. Therefore ISIMIP corrects the monthly mean and its daily variability. Here, we use the ERA5 reanalysis data as reanalysis data in our study. For convenience we use the term ISIMIP-ESM to denote the output from the ESMs after applying the ISIMIP statistical downscaling and bias correction methodology.

Finally, why is quantile delta mapping not done for the historical WRF simulations (in which case I note that it would simplify to the regular quantile mapping method)? Wouldn't it be fairer to compare ERA5 with bias-corrected ISIMIP/statistical downscaling and bias-corrected WRF/dynamical downscaling?

Reply: In our historical simulation, we only want to see the performance of ESM raw data, ISIMIP results and WRF results – without bias correction. The performance of WRF on the temperature and humidity is good in our historical simulation, but for the wind speed WRF shows an overestimation. If WRF showed good performance on all three variables, we would not need to do bias correction for WRF results, but the analysis shows that we do in fact need to bias correct the WRF output. We think this is a worthwhile point to make, and we do compare the bias corrected WRF with ERA5 and uncorrected WRF in fig. 6.

Specific Comments:

Abstract, line 20: It is not clear to me what “larger spatial ranges” means. I also do not see further reference to this in the main text. While I can see from Figure 9 that the range (maximum minus minimum) of temperature and of humidity is larger in WRF than ISIMIP, it's not obvious that this is the case for wind since although WRF has lower minima, ISIMIP has higher maxima.

Reply: This sentence refers to temperature ranges across the Beijing province. We have rewritten this sentence.

In the 2060s WRF produces consistently larger spatial ranges of surface temperatures, humidities and windspeeds than ISIMIP downscaling across the Beijing province for all three future scenarios.

pg 3, line 60: Give the full name of ISIMIP at least once (perhaps in the abstract as well).

Reply: Done. We have rewritten this sentence.

The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP, <https://www.isimip.org/>) consortium has produced methods (Hempel et al., 2013) widely used to correct the bias from CMIP5 (Climate Model Intercomparison Project phase 5) and GeoMIP outputs (McSweeney et al., 2016; Moore et al., 2019; Kuswanto et al., 2021).

pg 12 / Figure 4: Are the authors concerned about RH values exceeding 100%?

Reply: In Fig. 4, RH from ISIMIP-HadGEM and ISIMIP-BNU do have some values exceeding 100%. Hempel et al. (2013) which defines the ISIMIP method, avoids negative values for radiation and wind speeds but it does not mention that the humidity may exceed 100%. So, the method has a systematic problem. But we note that the fraction of data concerned is small. We have add this to the fig. 4 caption:

Values of humidity exceeding 100% can occur with ISIMIP downscaling.

pg 14, line 277: ISIMIP seems to increase error and variance in wind speed in the case of HadGEM2-ES.

Reply: Yes, we have rewritten this sentence

ISIMIP greatly reduces errors and variance (except for HadGEM2-ES) but does not improve correlation.

pg 16, line 309-310: Assuming I'm looking correctly, it doesn't look like WRF-QDM-MIROC and WRF-QDM-MIROC-CHEM are wetter than ERA5 from June to October. If anything, they look drier, especially the former.

Reply: Yes, we have deleted that statement and rewritten this sentence.

For humidity, the overall performances of the WRF-QDM-ESM are not good and they all show a dry bias relative to ERA5 from July to the following May.

pg 18, line 335-336: Wind speeds in ISIMIP look highest (as opposed to lowest) in the southeast of the domain.

Reply: Yes. We have deleted the southeast.

The windspeed in the southwest of the domain from ISIMIP-ESM is low, while WRF-QDM-ESM winds are lowest in the city center.

pg 20, line 360: To me, only the G4-RCP8.5 anomalies for ISIMIP look anti-correlated. I might suggest removing this sentence.

Reply: We have rewritten this sentence.

Humidity anomalies from ISIMIP have a difference under G4 relative to RCP8.5 in the southwest of the domain, where windspeed anomalies show an obvious positive change, while for WRF there are no particular patterns.

pg 24, line 417: I think the authors meant to say that WRF relative humidity exhibits a dry bias in winter.

Reply: Yes, we have rewritten this sentence.

WRF relative humidity, however, is always drier than that of ERA5 in winter whether revised or not.

pg 25, line 429-430: This sentence is confusing as written, and the first part is not obvious since WRF-QDM-ESM shows opposite trends in roughly equal parts of the domain when comparing G4 and the 2010s. I might suggest removing this sentence.

Reply: Done.