Earth Syst. Sci. Data Discuss., referee comment RC1
https://doi.org/10.5194/essd-2021-401-RC1, 2022

Comment on essd-2021-401

Anonymous Referee #1

Referee comment on "Historical and future weather data for dynamic building simulations in Belgium using the regional climate model MAR: typical and extreme meteorological year and heatwaves" by Sébastien Doutreloup et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-401-RC1, 2022

General comments

- The article is appropriate to support the publication of the data set. It both build on previous work and provides new approaches to producing future weather data files that can be used by a wide range of end users. The data appear to be of quality, complete and usefull.
- As the construction of future weather data files is not a trivial task, most of my comments ask for clarification of certain methodological choices and assumptions. I leave it to you to modify the text accordingly if you deem it necessary to enable end-users to better understand the data they are using.

Specific comments

Abstract

- Maybe add a short sentence explaining that this work is part of the ULiège OCCuPANt project (it might help future readers find the article and the data).

2.1 MAR model and area of interest

- Line 74. “can be considered as a reconstruction simulation”, I understand what you mean but let’s keep in mind that the MAR model is not necessarily perfect and so the MAR-ERA5 can still be biased. I come back to this point in the 2.2.3 Evaluation section.
- Line 77. “represent only the mean evolution”, I understand what you want to say but I would suggest to reformulate the sentence to make it clearer that we are not just talking about the average climate but also its variability (and especially extremes), otherwise you would not be able to use these simulations to produce the XMY files.
- Line 84. What do you mean by “integration domain”? As I understand it is the final domain where data can be extracted and where points from the “relaxation zone” are omitted?
- Line 85. “The spatial resolution of MAR is 5km”. Do you directly downscale ERA5 and GCMs simulations to 5km or do you make intermediate lower resolution simulations to reduce the “resolution jump” between the two?
- Lines 83-84. I would suggest adding the different period you simulated: “[…] over the North Sea from both ERA5 reanalysis for the YYYY-YYYY period and the YYYY-YYYY period for the three ESMs”. I think it would make it easier to understand some parts of the next section.
- Lines 86-87. You indicate that your choice of cities is based on the Urban Heat Island (UHI) effect (“show a temperature increase compared to the neighbouring countryside”) but I do not think the MAR model incorporate any sort of Urban Canopy Models to represent cities. Are you using another approach on urban points (replacing them by rock for example)?

Depending on the approach you are using some points/clarification need to be clear (not necessarily in the article but for the people using the data afterwards):

In the case you were using rocks, depending on the physical properties associated to this cover I think we could expect an underestimation of nocturnal temperatures and of the UHI. Consequently if we were to compare the data from the MAR-ERA5 simulation to observations or even to ERA5 we might find different UHI patterns/intensities.

In the case you were not representing cities at all (which is fine in my opinion), users need to know that the data they are using correspond to the rural conditions surrounding the city.

2.2.1 Choice of representative ESMs

- Line 95. Just by curiosity/for clarification. How is your first criterion defined to preselect the GCMs with the lowest possible bias relative to ERA5? How do you combine the geopotential height and temperature biases? How do you select GCMs based on these biases and how much GCMs do you retain after this first selection?

Once you've made your first selection did you look at the spread of your sub-ensemble compared to the 30 GCMs? On both the historical temperature as well as the expected warming rate? My main concerns is that some GCMs could perform well in reproducing the historical climate (at least in terms of atmospheric circulation) but produce unrealistic warming (some CMIP6 models have been shown to be outside the IPCC's "likely" range of expected warming; Zelinka et al., 2020).

- Line 98. When you talk about "the models spread in 2100", are you comparing end of the century temperatures directly (between models) or warming relative to the historical period? Because later in the paragraph the formulation change to "warming". For example, when you states that "MIROC6 simulates larger warming than average" do you talk about warming relative to the historical period or just that it simulates higher temperatures than the average (in which case it could be either a greater warming rate, a warm bias or a combination of both).

2.2.2 Future socioeconomic scenarios

- Lines 111-112. "scenarios SSP3-7.0 and SSP2-4.5 are reconstructed from the MAR simulations forced by the ESMs with the SSP5-8.5 scenario". Are you aware of previous work that has implemented a similar approach?
- Lines 117-120. Some things are not clear to me on your approach to chose a period representative of others SSPs warming at the end of the century in the SSP5-8.5 GCMs. If I understood correctly you compute the projected average temperature over the whole of Belgium at the end of the century in each ESM and for the SSP3-7.5 and SSP2-4.5 scenarios (on which period?) and then you search for the "closest" decade in the SSP5-8.5 in terms of average temperature and variability (how do you define it?), lastly you extract the period you found in your MAR simulations. On line 117 you say you search for a "decade" (do you compute a moving average or do you just take every 10 years). Also the example you give afterwards show a 20 years period. Lastly, just to be sure, do you apply this approach on the 3 ESMs you have previously selected or on the whole ensemble?

This section, and especially the approach to "reconstruct" different SSPs from a single one is interesting but I would be cautious on multiple points.

Even if the end-of-century warming level of all scenarios is found in the SSP5-8.5, they are probably not perfectly comparable: for example one can expect to see different local conditions between a world that gradually reaches 2°C of warming in 100 years and a world that reaches it in a few decades.

Therefore, when sharing the data, I would make it clear that the scenarios other than the SSP5-8.5 do not really correspond to GCM driven by these scenarios but are reconstructed based on warming level. And for future work, if the same approach is applied, I would suggest to think about presenting the data not in terms of SSPs scenarios but in terms of expected warming level (both global and European).

I come back to another point in the next section 2.2.3. regarding known differences in warming simulated by RCMs and their driving GCMs – which I expect could also be found for MAR simulations.

2.2.3. Evaluation of the MAR simulation

- Line 131. “are very close the average” is debatable with MAR-MIR being more than 2°C hotter on average in summer.
- Line 132. “not statistically significant”. How do you compare both average and how do you test for statistical significance?
- Line 132. “lower than the standard deviation of MAR-ERA5”. Not true for MAR-MIR in summer for temperature.
- Lines 131-134. I would be more cautious about the conclusions.

Since the ESMs driven MAR simulations are compared to another MAR simulation you can not evaluate the intrinsic biases of MAR (which has been shown to have a slight warm bias (between +0,3 and +0,7 °C on average annually with a negative bias in winter and a positive bias in summer) by Wyard et al., 2017 [using an older version]). A stricter evaluation of MAR would have been to compare the MAR-ERA5 simulation to observations.

“expect MAR-MIR which significantly overestimates […]”. Here you correct your prior statements made on lines 131-132.

“Knowing that MAR-MIR simulated the largest warming in 2100 […]”. This point makes me come backs to the questions I asked regarding the way you define warming in the section 2.2.1: is MIROC6 really the ESM with the largest warming relative to the historical period or is it just a more “warmly-biased” ESM than the others?

As I briefly mentioned in the previous section, previous works have shown that some regional climate models are not able to reproduce the warming of their driving GCMs/ESMs over Europe (Schwingshackl et al., 2019; Boé et al., 2020). Different hypothesis have been put forward such as the lack of evolution of aerosols (Boé et al., 2020; Gutiérrez et al., 2020) or the physiological effect of CO2 (Schwingshackl et al., 2019; Boé et al., 2019). See Ribes et al., 2022 for a comparison of CMIP6, CMIP5 and EURO-CORDEX warming over France (which might be comparable to Belgium). I was wondering if you compared the warming simulated by the ESMs to the one simulated by MAR?

Following this point and the one I made previously regarding the use of SSPs names when sharing the data, it might be interesting for future work to build files on projected warming instead of SSPs. I stress these points because each new step in the cascading sources of uncertainties from SSPs, ESMs, RCMs and finally their application (with building energy models for example) makes it extremely difficult in the end to estimate the confidence we can have in the final results.

2.3 Generatin the TMY and XMY files

- Lines 154-159. This section might need to be reformulated. I don’t think I understood the different steps correctly. As an example para is the temperature and the period is 2001-2010:

(1) You take all the hourly data of all the months of each month one by one (January 2001, February 2001, …, November 2010, December 2010). You end up with 120 values of average temperature (12 months times 10 years). Maybe change “daily mean” for “monthly average”?

(2) You compute the median of the distribution of January (then February, March and so on) as the value representing the most typical conditions of this particular month. From your 120 monthly averages you first look at the 10 January averages and chose the closest one to the median. You do this for every month which gives you something of the sort: January 2003, February 2010, March 2006, etc.

(3) You extract the hourly values corresponding of the months you've selected and concatenate them into a single new year of hourly data.

Two questions following these steps:

How do you handle possible jumps in temperature between two months?

I suppose you compute you percentiles on all the hourly data, so you must take a lot of null values of radiation into account (during night-time). Do you know if this can have an effect on your selection, (especially for the percentile 95) compared to a selection with percentiles computed on day-time values?

- Lines 160-173. I would suggest putting these lines in a table if possible. Maybe something like: Variable name | Height above ground level | Units

2.4 Definition of a heatwave and generating the HWE files

- Line 178. Do you know how the 18.2 °C and 29.6 °C thresholds were determined? I did not find the information in Tsachoua & Reynders, 2016.
- Line 184. Since you screen for heatwaves on every pixels the points I mentioned in section 2.1 regarding the possible representation of cities in MAR (either with an Urban Canopy Model or by replacing them by rocks) is of particular importance since heat waves definition are usually made on rural conditions and do not take into account the urban heat island.
- Line 187. Do you compute your statistics on the 1980-2014 period (35 years)? Because Ouzeau et al., 2016 use the 1976-2005 period (30 years).
- Lines 193-194. I think Ouzeau et al., 2016 already had implemented a minimum duration of 5 days for their heatwave detection.

- Lines 202-204. As Nicolas Heijmans noted, building energy models not only need the values during an extreme event but also the ones that lead to it. Maybe add a sentence explaining how the HWE files can be combined with the TMY and XMY files (depending on what has been decided in the ULiège OCCuPANt project).

4. Conclusion

- Line 242. Change "is forced by 4 ESMs" to "is forced by a reanalysis and 3 ESMs".

5.1 Appendix A: Example for Liege city

- Figures A-1 and A-2. You could add axis labels and variable units.

Technical and editorial corrections

- Line 32. If I understood correctly I suggest changing "will warm faster and larger than equatorial regions" to "will warm faster and to higher levels than equatorial regions".
- Line 47. Change "This comfort of life" to "This living comfort".
- Line 61. Remove "Finally".
- Line 70. I would change the order of references so that they are chronologically ascending
- Line 73. I would change "6-hourly assimilated" to "assimilated every 6 hours"
- Line 79. As Nicolas Heijmans suggested, I would uniformise the SSP notations (SSPx-x.x)
- Line 82. Remove "MAR" from "MAR vertical level".
- Line 87. Neighbouring with a "u".
- Line 125. Remove "y" in "in your case".
- Lines 137. Change "the extreme climates" to "the extreme climate".
- Table 1. I would remove "annual" from the "Mean annual temperature […]" cell since

the average is computed for the whole year and the summer.

References

- Boé, J., Somot, S., Corre, L., and Nabat, P. (2020). Large discrepancies in summer climate change over Europe as projected by global and regional climate models: causes and consequences, Climate Dynamics, 54, 2981–3002
- Boé, J. (2021). The physiological effect of CO2 on the hydrological cycle in summer over Europe and land-atmosphere interactions, Climatic Change, 167, 21, https://doi.org/10.1007/s10584-021-03173-2
- Gutiérrez, C., Somot, S., Nabat, P., Mallet, M., Corre, L., Meijgaard, E. v., Perpiñán, O., and Gaertner, M. A. (2020). Future evolution of surface solar radiation and photovoltaic potential in Europe: investigating the role of aerosols, Environmental Research Letters, 15, 034 035, https://doi.org/10.1088/1748-9326/ab6666
- Ribes, A., Boé, J., Qasmi, S., Dubuisson, B., Douville, H., and Terray, L. (2022). An updated assessment of past and future warming over France based on a regional observational constraint, Earth Syst. Dynam. Discuss. [preprint], https://doi.org/10.5194/esd-2022-7, in review
- Schwingshackl, C., Davin, E. L., Hirschi, M., Sørland, S. L., Wartenburger, R., and Seneviratne, S. I. (2019). Regional climate model projections underestimate future warming due to missing plant physiological CO 2 response, Environmental Research Letters, 14, 114 019, https://doi.org/10.1088/1748-9326/ab4949
- Wyard, C., Scholzen, C., Fettweis, X., Van Campenhout, J. and François, L. (2017), Decrease in climatic conditions favouring floods in the south-east of Belgium over 1959–2010 using the regional climate model MAR. Int. J. Climatol., 37: 2782-2796. https://doi.org/10.1002/joc.4879
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. Geophysical Research Letters, 47, e2019GL085782. https://doi.org/10.1029/2019GL085782