Anonymous Referee 2:

The present study is dedicated to summer temperature, heat waves and associated implications for human health, agriculture and tourism in an ensemble of convection permitting regional climate model projections. In addition, the added value of the higher model resolution is demonstrated compared with a model version using parameterized convection. The study comprises three novel aspects at once: (1) a relatively large model data base of very high-resolution simulations over a quasi-transient forcing period, (2) the assessment of regional to local climate change patterns based on substantially improved model simulations, and (3) the consideration of derived climate indicators bridging the gap between meteorological heat events and socio-economic implications as well as adaptive requirements.

The paper represents a very valuable contribution to the community – with respect to methodical aspects (new model generation) and practice-relevant research (high-resolution patterns of climate change). Therefore, I recommend this manuscript to be accepted for publication in NHESS with minor revisions.

The minor revisions refer to a list of specific comments (see below) and to two general comments:

A: First, we would like to thank the reviewer for her/his insightful comments, which have greatly contributed to improving the text. In making corrections, we have tried to follow the suggestions as closely as possible.

General comments

General comment 1

(1) The manuscript basically is well presented, but exhibits some linguistic inaccuracies, especially typos. Therefore, I believe the authors themselves can achieve an improvement without explicit language editing by a native speaker. Nonetheless, a careful revision is required since the typos and inaccuracies are quite numerous.

A: We appreciate the feedback and will revise the manuscript.

General comment 2

(2) The GWL 2 and 3 periods seem to be associated with a lower level of temperature increase in central and southern Germany, at least in terms of the mean summer temperature. According to the IPCC and many other studies, I would have expected an above-average warming in Central Europe, given the fact that land masses are warming up stronger than the ocean surface, especially in the Northern Hemisphere extratropical regions (COWL pattern). Is summer less sensitive than the annual mean or is it an issue of the considered GCM-RCM combinations? I suggest that the authors pick up this point in their discussion.

A: Indeed, the warming is slightly stronger integrated over the year. We have rewritten the respective paragraph:

The summer temperature increases with global warming over the whole evaluation area. From the reference period (global warming at 0.46 °C) to GWL2, the increase is on average 1.55 °C (Fig. 6a). From the reference period to GWL3 the average increase is 2.60 °C. When integrated over the year, the

ensemble shows a slightly stronger warming than only over the summer months, indicating that summer temperatures are less sensitive than the annual mean (Fig. 5a). However, the differences are still in the range of 0.11 °C (0.09 °C) above the global warming in GWL2 (GWL3). Therefore, the regional warming in the evaluation area in the considered GCM-RCM combinations is close to the global average and only slightly enhanced. This is less than suggested by the theory of greater warming over land than over the ocean and as generally projected (IPCC 2021). The impact of the bias correction on the climate change signal is considered to be negligible, as the uncorrected data integrated over the year show a nearly identical warming of 0.11 °C (0.07 °C) above the global average in GWL2 (GWL3) in the evaluation area.

Geographical dependence leads to regional variations of warming. Over the evaluation area, warming ranges from 1.45 to 1.64 °C (5th And 95th percentiles) in GWL2 and from 2.44 to 2.76 °C in GWL3. As shown in Fig. 6a and 6c, the strongest increase is observed in the uplands in the north of the domain (GWL2), and in the Black Forest and Swabian Alps in the south (GWL2 and GWL3). Less warming, below the global average, is expected in the Alpine Foreland (GWL2 and GWL3).

To fit in the new structure of the section a following paragraph about the ensemble spread was rearranged as well.

- IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896, 2021.

Specific Comments:

Line 42: CPM stands for ‘convection permitting model’ (not convective).

A: We have corrected the typo.

Line 62: What is meant by quasi-transient? And ‘manor’ is certainly not the right word in this context, I guess it is ‘manner’.

A: We have corrected the typo. We have added the following in the introduction, a detailed description is available in the methods section:

All simulations cover the period from 1971 to 2100 in a quasi-transient manner, where the projection is composed of several time slices.

Line 85: Table 2 is addressed in the text before this is done for Table 1.

A: We have corrected the Table’s location and reference.

Fig. 1: The fine lines in the background of the map seem to be river basin. Maybe a word is useful why these are plotted.

A: The lines in the background represent the German major landscapes. Those regions were added, because often a dependency of the results is visible and they facilitate the interpretation of the results. We have added a description of this background map in the figure description.

Line 96: I wouldn't call it a climatological difference, when two three-year periods are compared with each other. Maybe the authors may want to call it what it is actually: a difference between three-year averages.

A: We have clarified the sentence as suggested.

Subsection 2.1: I suggest to explain in few words the data sources and procedure leading to the HYRAS dataset and to explain what an equilibrium climate sensitivity is (Table 2).

A: We have added a short explanation of equilibrium climate sensitivity.

Regarding HYRAS, we have added the following: "The HYRAS dataset is used as observation, which is based on station data that are aggregated using the REGNIE method of combining a regression model and inverse distance weighting to a gridded dataset (Rauthe et al. 2013, Razimaharo et al. 2020)."

- Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology–Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Meteorol. Z, 22, 235–256, https://doi.org/10.1127/0941-2948/2013/0436, 2013.
- Razafimaharo, C., Krähenmann, S., Höpp, S., Rauthe, M., and Deutschländer, T.: New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS), Theor. Appl. Climatol., 142, 1531–1553, https://doi.org/10.1007/s00704-020-03388-w, 2020.

Line 131: As this study is focussed on heat events, the question arises whether extreme temperature is indeed normal. There are several studies indicating that it is not, suggesting a combined QM approach with different statistical models below and above a temperature threshold. Please add a discussion on this issue.

A: We would rewrite the section on bias correction, also in light of the comments in Review 1 (see response to Review 1). We would add a discussion off the distribution based approach in a revised manuscript.

Furthermore, the use of a parametric approach of fitting an assumed distribution to the data to derive the transfer function is still arbitrarily discussed. Several studies, e.g. Pastén-Zapata et al. (2020), Quian et al. (2021), apply a normal distribution for temperature to get a more robust transfer function. Using a fitted function has the additional advantage that the transfer function is independent of any smoothing interval that may be defined (Kerkhoff et al. 2014). On the other hand parametric approaches introduce additional bias, if the distribution of a variable is not accurately met by the theoretical distribution. Especially for extreme values, a deviating statistic is assumed according to the extreme value distribution. Quantile approaches, allowing different statistical models for extremes,

could potentially reduce uncertainty (e.g. Vrac and Naveau 2007, Berg et al. 1012, Schubert et al. 2017).

- Pastén-Zapata, E., Jones, J.M., Moggridge, H. and Widmann, M., 2020. Evaluation of the performance of Euro-CORDEX Regional Climate Models for assessing hydrological climate change impacts in Great Britain: A comparison of different spatial resolutions and quantile mapping bias correction methods. *Journal of Hydrology*, *584*, p.124653.
- Qian, W. and Chang, H.H., 2021. Projecting health impacts of future temperature: a comparison of quantile-mapping bias-correction methods. *International Journal of Environmental Research and Public Health*, *18*(4), p.1992.
- Kerkhoff, C., Künsch, H.R. and Schär, C., 2014. Assessment of bias assumptions for climate models. *Journal of Climate*, *27*(17), pp.6799-6818.
- Berg, P., Feldmann, H. and Panitz, H.J., 2012. Bias correction of high resolution regional climate model data. *Journal of Hydrology*, *448*, pp.80-92.
- Schubert, D., van der Linden, R., Fink, A.H., Katzfey, J., Phan-Van, T., Maßmeyer, K. and Pinto, J.G., 2017. Klimaprojektionen für die hydrologische Modellierung in Südvietnam. *Hydrologie und Wasserbewirtschaftung*, *61*(6), pp.383-396.

Subsection 2.3.2: The description of UTCI is deficient. I either suggest to refer to the literature, leaving out all equations, or to provide a complete description with all terms figuring in the equations and the full equation for UTCI instead of f().

A: The complexity of the overall calculation is beyond the scope of this paper, so we have decided to refer to the literature.

Line 207: It should be mentioned that this statement refers to the reanalysis-driven experiment. The enhanced spread is probably related to the fact that the model has a higher genuine resolution than the validation data, implying higher temperature differences in mountainous areas.

A: We have added that information.

Fig. 4: Panel c is unclear to me: is it a comparison of the bias (then the caption is wrong saying that it is the 2.8 km minus 7 km scale) or does it indicate that the negative bias of the 7 km run is more or less compensated by the 2.8 km run. I would prefer seeing a bias reduction in panel c because it is more intuitive for the reader.

A: The mean square error skill score (MSESS) is displayed (Eq. 3) in c. We have corrected an error in describing the labeling in the caption. We hope that this resolves the reviewers concern.

Beginning of section 4: I miss a statement about what model resolution is used for the subsequent analyses. I guess it is the 2.8 km scale since the bias could be reduced noticeably.

A: This is correct, the 2.8km resolution was used in the following. We have added a statement regarding the used resolution in Section 4 to 6.

Line 247: Have the authors tested whether the density is indeed skewed left. At first sight, it looks quite normal.

A: That is a valid point. We have recalculated the skewness. In the reference period (1971-2000) it is between -0.24 and -0.18 in the ensemble, while in the observation it is -0.17. Since these deviations from 0 are relatively small, we have removed this aspect from the manuscript.

Line 252: Please explain what FWHM actually tells us.

A: We have added the definition and meaning of the full width at half maximum in relation to the changes in temperature distribution shown:

A parametrization of the spread of the distribution is made in terms of the Full Width of Half Maximum (FWHM), which is defined as the width of the distribution at the level of the half peak value. [...] Regarding the temperature distribution, an increasing FWHM indicates a more variable daily temperature, leading to higher amplitudes and to a stronger increase in the frequency of warm extremes on the right side of the curve compared to the shift of the curve median.

Line 277: At the end of this sentence the authors may include a '(not shown)'.

A: Line 277 comprises following statements "*Overall, the mean temperature over Germany rises in a warmer climate predominantly in late summer as well as in the winter half-year, with the smallest increase in spring. This leads to a general shift of the summer maximum temperatures to later summer*" The paragraph serves as a summary of the analysis above. The statement in line 277 was discussed in line 234 to 237 based on the evaluation presented in Fig. 5a. We have clarified that the paragraph is intended as a summary and hope that this addresses the reviewer's concern.

Fig. 9: It should be mentioned that the thick solid line refers to the ensemble mean. To be clear please add 50% 'of the study region'.

A: We have changed the description as suggested.

Line 357: Why is the British model now claimed an outlier whereas previously it was not because blocking situations may be better represented in this model?

A: In section 5, we found good agreement of the analysis of the bias corrected data of all 4 simulations, including HadGEM2-ES, with the present day conditions derived from the observation. In the following, we found particularly long heat waves in a future climate in the simulation driven by HadGEM2-ES. Those long, and therefore persistent, warm spells could originate from a different representation of large scale circulation patterns in the driving GCM, namely blocking situations.

In Section 6, we proceed analogously and first evaluate the present day conditions in all simulations. However, compared to the other 3 simulation chains, we found a significantly higher UTCI in the simulation driven by HadGEM2-ES. We therefore consider these results to be an outlier. We attribute this difference primarily to the fact that unlike the analysis in Section 5, no bias correction was applied to the hourly data used in Section 6.

We would add the following explanation to the respective paragraph:

There is good agreement between three of the four ensemble members, showing a similar range of UTCI over the reference period 1971-2000. The simulation driven by HadGEM2-ES results in a significantly higher number of days with UTCI > 32°C. We attribute this difference mainly to higher summer temperatures in this simulation, which unlike the previous analysis of daily data was not subject to bias correction. To minimize the influence of possible outliers, we consider the ensemble median in the following analysis.