Dear Benjamin Poschlod,

Thank you for your review of our manuscript and suggestions for improvement. Please find our replies to your comments below in blue.

Best regards,
Emily Collier & Thomas Mölg

General comments

The manuscript by Emily Collier & Thomas Mölg gives a comprehensive overview of a high-resolution 30-year climatological data set over Bavaria. The climate simulations were produced by the WRF model in 1.5 km resolution, nested in a 7.5-km-resolution domain and driven by ERA5 boundary conditions. The authors evaluate the model performance for air temperature, relative humidity, winds, surface pressure, precipitation, and land surface temperature for a 12-month period where they compare simulated values to observational data. Additionally, the effect of the application of nudging is assessed. Generally, the manuscript is well-written, and the figures support the presentation of the data set and its evaluation. In particular, the authors' handling of errors in the data set (e.g. sub-surface temperature in single glacier pixels) and explanation of deviations/biases (e.g. urban heat islands, connection between overestimated air temperature and overestimated radiation) are very valuable features of the data description.

The data are easily accessible and valuable for further application with focus on impact-related studies. Though, the total size of the 30-year daily-resolution data set (~450 GB) may not be easy to handle for users, who are new to the application of high-resolution climate data. On the other hand, users from the field of climate science would be interested in even higher temporal resolution, especially regarding the precipitation data. In sum, I consider the manuscript and the data appropriate for the publication within ESSD, although I recommend minor revisions based on the following remarks.

Thank you for your favorable assessment of our manuscript. Although the total dataset size is ~450 GB, the 3D variables that are most likely to be used for impacts assessments (e.g., near-surface air temperature, humidity and precipitation) amount to a more manageable 57 GB. With regards to the provision of higher temporal resolution precipitation data, please see our response to Michael Warscher for more details.

Specific comments

L1: Title: the data set is described as "convection-resolving". Though, within the whole manuscript, no convective events have been evaluated. Furthermore, the data set is provided in daily resolution, which is why short convective events cannot be investigated properly. Hence, I would suggest replacing "convection-resolving" by "high resolution".

We used the term "convection resolving" to describe the dataset following convention for atmospheric simulations with grid spacings below ~4-km. We did not mean to imply that we analyze convective events, however we agree that the use of this term could be misleading, especially to a wider audience, and therefore changed the title as suggested.

L80 / Table 1: The Kain-Fritsch cumulus scheme is applied for the 7.5 km domain, but not for the 1.5 km domain. According to that, not only deep, but also shallow convection is explicitly resolved in the 1.5 km domain? I would suggest clarifying this in the text.

Yes, as no additional parameterization is employed, deep and shallow convection are explicitly represented in the 1.5 km domain. We added to Sect. 2.1 "As no cumulus parameterization was employed in D2, both deep and shallow convection are assumed to be explicitly resolved.

L143: “For the distributed trend analysis, we did not apply a field significance test (e.g., Wilks, 2016) due to the small sample size.” – Does the “distributed trend analysis” refer to the results in L241 – 246 and Fig. 9? If yes – can you explain why is the sample size too small? If you test the trend at all 351x351 locations, the p-value should be adjusted for statistical tests at many locations (following e.g. Wilks 2016). Moreover, the reference (Wilks, 2016) is missing in your Reference section. Please also clarify, which test or method you used to detect trends.
Here we were referring to the sample size of years. Please note that based on the suggestion of the other reviewer, we removed the trends analysis from the manuscript in favor of expanding the model evaluation. Please see our response to this reviewer for more details.

L180: Has the observational precipitation data from DWD been corrected for undercatch? Especially in (pre-)alpine regions, this plays a major role, in particular for solid precipitation. I would recommend to briefly discuss this source of uncertainty.
We did not correct precipitation for undercatch and have added this information in Section 2.2. We also added to Section 3.1: “The MD is positive at the majority of stations, indicating that WRF generally underestimates observed precipitation. The underestimate is likely greater than reported here, since the observations were not corrected for wind-induced undercatch.”

L385: Figure 4 gives a good overview of the biases averaged for all locations. Though, the spatial distribution of biases would be of high interest as well. As the manuscript is already quite long and contains many figures, I would suggest creating such bias maps and moving these additional figures to a supplementary file.
We added some spatially distributed bias analysis as part of the expanded model evaluation.

Technical corrections
L156: 273.16 unit is missing
We changed to “exceeded the melting point.”

Figures 3,4,5,7,8: Temperature unit is “C” instead of “C”. Figure 6: Here the unit is missing in the figure (and given in the caption instead) Figure 9: Here you use “K” –> Please unify.
We changed the units to degrees Celsius throughout the paper and corrected the figure labels and captions.