Interactive comment on "G2DC-PL+ A gridded 2 km daily climate dataset for the union of the Polish territory and the Vistula and Odra basins" by Mikołaj Piniewski et al.

Joanna Wibig (Referee)

joanna.wibig@geo.uni.lodz.pl

Received and published: 20 December 2020

Interactive comment on G2DC-PL+ A gridded 2 km daily climate dataset for the union of the Polish territory and the Vistula and Odra basins by Piniewski et al.

The manuscript "G2DC-PL+ A gridded 2 km daily climate dataset for the union of the Polish territory and the Vistula and Odra basins" by Piniewski et al , submitted to Earth System Science Data presents a second, updated version of a high-resolution gridded daily precipitation, temperature, relative humidity and wind speed dataset as well as its description.

Overall, I assess very positively the main achievement of the paper – freely available dataset of high resolution daily data of containing key factors for calculating many climate parameters.

The manuscript is well organized and written quite clear. The reference list is complete. The methodology of interpolation method is based on the work by Berezowski et. al and only the differences between both approaches are described here. The description of the evaluation of interpolation errors is somewhat less clear and I would suggest a revision.

The interpolation errors were quantified using two functions: (1) Pearson's correlation coefficient ([-]) and (2) root mean squared error normalized to standard deviation of the observed data [-]. The cross validation was conducted on both a temporal and spatial scale. On the temporal scale the errors were calculated for each day from all stations having data on this day and presented in the form of a descriptive statistics table. On the spatial scale the errors were calculated for each station from all of a station's available daily values. The number of records on the spatial scale calculation was equal to the number of meteorological stations used and results were presented in the form of maps.

Sounds good, but the devil is in the details. In this case it is a standard deviation. For temperature, is the standard deviation calculated for each julian day separately or for the whole record? Nothing in the equation indicates the first possibility. So the range of possibly values of daily minimum and maximum temperature is in the order of 50°C, which means that the standard deviation is very large, as well as the interpolation error of the order of half of the standard deviation. Additionally, the distribution of daily minimum and maximum temperature from periods equal to the multiple of the year is bimodal. What is the statistical meaning of the standard deviation calculated from values from ther bimodal distribution. It is even worse in the case of precipitation. Distribution is far from normal and the standard deviation is a very poor measure of the spread of values.

The values of humidity and wind speed have much smaller dispersion and a less pronounced annual cycle, therefore the standard deviation describe their dispersion better and RMSEsd values are greater.

I have also some other minor comments:

The information on numer of citations (lines 61-63) is unnecassary and I suggest deleting it.

Line 77 should be years in instead of yearsin

Line 80 should be in dataset instead of indataset

Line 81 should be temperature instead of tempearture

Line 82 should be kriging instead of kirigging

In point 2 of subsection 2.4, the ending of the thought seems to be missing

In point 4 of subsection2.5, authors stated that "a map showing values of coefficient b representing the effect of wind exposition of the measurements site is presented in Figure 10", however in this Figure is only the division of stations into stations with low and medium shielding. It would be beneficial to provide some sort of breakdown criterion.

The differentiation of symbols denoting spatial and temporal values of the correlation coefficients and the RMSE would significantly facilitate the tracing of the text.

In the case of precipitation the relationship between the number of available stations and RMSE is intriguing. The increase in the number of stations in the initial period was accompanied by the decline of the RMSE. It is clear. However, in the second part of the analysed record, the decrease in the number of stations was not accompanied by an increase in the RMSE. It is surprising. The authors only stated the fact without comment. And it would be very useful.

C3

When writing about temperature interpolation errors the authors only stated that the conclusion that kriging errors for temperature are not dependent on the density of the observation network seems to hold true. They did not comment this fact. Some discussion on the impact of spatial correlation of variables on the relationship between station density and RMSE would be beneficial.

In my opinion the lack of discussion is the main drawback of the paper.