Review of essd-2024-433

PL1GD-T - gridded dataset of the mean, minimum and maximum daily air temperature at the level of 2 m for the area of Poland at a resolution of 1 km × 1 km

The authors develop a gridded dataset of daily temperatures for Poland covering the 1951-2020 period, based on observed data from 347 stations. The research is oriented to the validation of the dataset. A very brief example of use is shown at the end of the manuscript.

While the overall objective of constructing a new daily dataset of temperatures for Poland is justified and well contextualized in the introduction, there are several unclear parts regarding the utilized data and the methodological approach.

For instance, regarding the **original data from stations**, a complete characterization of the raw temperature dataset is needed: Where are located the stations? (maybe a map would be useful); Do all of them have daily data or they come from hourly information? Are they automatic or manual stations? How many years of data (and gaps) they have? How is the temporal evolution of data availability? Did you apply any quality control/homogenization/gap filling procedure? This information and a basic statistical characterization are basic to contextualize the starting point of the grid and to understand the results.

In order to be clearer in the **methodological section**, the use of RBF must be further justified with more than a few references (Lines 105-108), mostly not related to temperature. While RBF is not particularly wrong, it is known as a "conservative" interpolation procedure, meaning that it usually reduces the spatial variability of the results which could be a problem in complex orography areas. As an example, the noted underestimation of higher values and overestimation of lower ones, can be attributed, with high confidence, to the interpolation scheme. Same situation can be observed at high elevations. Although higher RMSD values can be related to the scarcity of observations at those altitudes, the RBF is also probably related. None of the cited and well-known datasets used this approximation. Although it can be perfectly valid for this dataset, it must be supported by a justification. In addition, a gridded dataset must include a measure of uncertainty for all estimates to evaluate the reliability of the data at each location and time step.

Lastly, there is not a discussion section in which the new gridded dataset could be compared to others covering the region and cited in the introduction.

Apart from a general recommendation of a professional proofreading of English in the document, due to some confusing expressions, here are the minor comments, line by line:

Introduction:

L31: This statement is not valid for non-European regions.

L36: Which is the rationale to choose this spatial resolution?

L71: Actually, there is an operational product starting in 1960 that is updated regularly: https://www.meteoswiss.admin.ch/dam/jcr:818a4d17-cb0c-4e8b-92c6-1a1bdf5348b7/ProdDoc_TabsD.pdf

Data and methods:

L89: "IMGW-PIB". What is this?

L101-103: This means that all of the stations had hourly data?

L120: How many "m" points you used? Did you set a radius of search? This is important since the availability of stations is not the same throughput the temporal period, and it has an impact on the estimation.

L132-133: If I understood, you're assigning the value of a small 100x100m elevation pixel to the 100 pixels overlapped by a 1km^2, right? The problem with this is that is in those areas with high elevation variability you are assigning a non-representative value to the larger pixel and that can lead to significant biases in temperature estimates. A correct approach would be using the mean or the median elevation of all 100x100 pixels overlapped by one 1x1km pixel.

L139-141: While this is valid to evaluate daily estimates, it cannot be used to evaluate long term trends or even monthly or annual aggregates because, for the comparison between a single pixel and their overlapping observations, you have non-continuous data or even data from different stations. To avoid this issue, it is usual to separate some complete data series (for example 20-30% of the total) and use them to validate the estimates at those locations. In addition, how did you randomly select the 5%? It should be a spatially driven randomization to avoid spatial biases in selecting stations. This procedure is implemented in almost all GIS software.

Results and discussion:

L155-158: This is already stated in previous section

L170 (Figure 1): What is the meaning of coloured lines? Please extend the figure caption.

L193-195: Any interpretation for this? I guess that the RBF is smoothing the extremes.

L196 (Table 3): What is "cRMSD"? About Q95D and Q05D, I guess that they are the difference between observed and estimated Q95 and Q05, respectively. Please extend the table caption to make clear the meaning of all acronyms and the units in which the values are expressed.

L225 (Table 4): At this point, the ratio of means or the ratio of standard deviations would be a better test, since the RMSD is an absolute value and we can't see here if there is a bias related to an over- or under-estimation by altitudes.

Figures 3, 4 and 5: The tonal variation in a single-color scale avoids a correct interpretation. I recommend using a sequential color scale.

L261: Just a conceptual note: you can derive past temperature variability, but this is not (or not necessarily) related to climate change.

L267: These values are surprisingly high. Please clarify if you are showing the average 90th percentile of TX for both periods or the absolute maximum one.

L290-291: This is not results and it should be removed.

L315 (Table 5): Please, state if these values are the average of all the pixels.

Example of application:

L320: Trends significance is not shown. In addition, when you say "selected stations", do you mean the pixels overlapping those cities?

L322-323: This is not a fair comparison since you used the data from stations to build de gridded dataset.