

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2020-327-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2020-327**

Martin Hanel (Referee)

Referee comment on "SLOCLIM: a high-resolution daily gridded precipitation and temperature dataset for Slovenia" by Nina Škrk et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-327-RC1>, 2021

The paper documents the development of 1km x 1km gridded dataset of precipitation and maximum and minimum temperature. The authors apply methods of Serrano-Notivoli (2017, 2019) and while there are no methodological innovations, the work itself is not trivial. The dataset is in my opinion important contribution to regional climatology and is potentially useful for following applications from trend assessment to hydrological modeling and as such it is worth publication in ESSD.

I have, however, some concerns about the way the paper presents the application of Serrano-Notivoli (2017, 2019) framework and related uncertainties. To summarize in general, the authors focus more on the presentation of the derived dataset than the derivation itself. Which I believe is opposite the way it is demanded by ESSD. Details are given below. Therefore I suggest to reconsider the paper after major revision of the paper.

Comments:

1. Methods

l. 101 and further - authors use 10 neighboring stations to estimate the central value. It is not clear whether this applies also to the boundary stations or not. In addition, as the data series are not available for the whole 1950-2018 period, it seems that the 10-nearest set changes over time, implying that also the glm or glmm models should change. This should be stated explicitly since it has an impact on uncertainty of the estimated values. It should be further discussed in section 5.

It is also not clear, what was the basis for selecting 10 stations and how this choice impacts the results.

l. 105 the authors are speaking here about wet probability but it is at this point of paper not clear at all, how they get it.

l. 107 what do you mean by "internal coherence"

l. 110 standard deviation of what?

l. 113 all suspects were removed?

l. 115-116 "the RVs ... were then calculated ..." - from the text above it seems that the RVs had to be calculated already before to identify suspects. If this is the case, then I believe it would be better to begin with the glm(m) model to also clear up the QC procedure.

It is also not clearly presented, what have you done to obtain station time-series prior to the start of the measurement and after its end. Does it mean that the model for a location changed when a close station popped-up? This should be clearly described and discussed also later in sect. 5 since it impacts the uncertainty of the estimates, which in principle would be different across years.

2. model description

The key part of the procedure are the GLM and GLMMs models, however, we do not learn much about the choice of the explanatory variables and nor any model assumptions are discussed. I believe it is very important to reveal, how the model was set up, how the variable selection was done and how the uncertainty was estimated.

l. 117-118 authors state that precipitation and temperature were used as dependent variables and lat, lon, alt and distance from the coast as independent - is it really like this? Please, give a precise description of what you have done.

3. Results

- there should be less information on Slovenian climate and more information on model selection, uncertainty assessment etc.

4. Discussion

- limitations and uncertainties of the dataset should be discussed in detail.

Minor comments:

l. 58 Another source of discrepancies is that the grided data represent spatial average rather than point value leading to smoothing of extreme values.

l. 140 Please consider adding information on the overall number of missing values.

Fig. 3 - not entirely clear what is on y-axis? Is it the total number of removed days for a year and all stations?

l. 143 "The majority of the data was removed ..." vs l. 145 "... only 1.26% ... were removed" - please revise the sentence to be clear.

l. 163-164 would it not be better to set up some wet-day threshold?

l. 169 - 173 I was not able to understand what is described here and what precisely is represented by Fig. 4

Fig. 7 perhaps would be better to use lines instead of dots.