RESPONSE TO REVIEWER

Review of Manuscript No.: essd-2020-327

Title: STEAD: SLOCLIM: A high-resolution daily gridded precipitation and temperature dataset for Slovenia

Authors: Nina Škrk, Martín de Luis, Katarina Čufar, Zalika Črepinšek, Lučka Kajfež Bogataj, and Roberto Serrano-Notivoli

---

We are grateful to the Reviewers for their thoughtful and constructive feedback, and the Editor for considering a revision. In this Response to the Reviewers' files, we provide complete documentation of the changes that have been made in response to the reviewers' suggestions and comments. The original comments are shown in **bold text** and the author responses are shown indented in plain text. Quotations from the revised manuscript are shown in *italic text*. Line numbers in the author responses refer to locations in the revised manuscript.

 # Martin Hanel (Referee)

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.

> Thank you for your useful comments. Based on the experience of the authors in the creation of previous datasets of precipitation and temperature, we consider that a specific high-resolution dataset for Slovenia will help to tackle very different environmental challenges from an unprecedent climatic perspective in a region with a wide variety of climatic regimes and ecological units.
>
> We have addressed all your comments, trying to be clear in those aspects that can lead to misunderstandings about the dataset creation and its climatic derivatives. Apart from that, although the used methodological process could be discussed in terms of its performance, it was already done in Serrano-Notivoli (2017a, 2017b, 2019), and large and interesting discussions were already resolved. However, we will be happy to answer to all of your doubts and concerns about it. In case of the current open discussion cannot encompass your queries due to out-of-scope issues, we encourage you to read our previous responses to our model-based methodology. In any case, we tried to summarize in this response letter all the methodological questions raised by all the reviewers, complementing them with their corresponding additions of text to the revised manuscript.

> *Serrano-Notivoli, R., Beguería, S., Saz, M. Á., Longares, L. A., and de Luis, M.: SPREAD: a high-resolution daily gridded precipitation dataset for Spain – an extreme events frequency and intensity overview, Earth Syst. Sci. Data, 9, 721–738, https://doi.org/10.5194/essd-9-721-2017, 2017a.*
>
> *Serrano-Notivoli, R., de Luis, M., Saz, M.A. and Beguería, S. Spatially based reconstruction of daily precipitation instrumental data series. Climate Research, 73: 167–186, https://doi.org/10.3354/cr01476, 2017b.*

Serrano-Notivoli, R., Beguería, S., and de Luis, M.: STEAD: a high-resolution daily gridded temperature dataset for Spain, Earth Syst. Sci. Data, 11, 1171–1188, https://doi.org/10.5194/essd-11-1171-2019, 2019.

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.

Thank you for your comment. The choice of 10 nearest stations is not arbitrary. Serrano-Notivoli et al. (2017a and 2017b) (2017b is now cited in the text, line 101) showed that, for precipitation, 10 neighbors are a good compromise between keeping the local variability of precipitation and having enough data to run the model. Despite the number itself does not need to be fixed to 10 (could be, for instance, 11 or 9, or even different depending on the availability of the data), it ensures the model reliability and, based on the number of observations over time, the disposal of data in all days of the period. We want to note that, for temperature, the number of neighboring stations is 15, as stated in Serrano-Notivoli et al. (2019), it is now indicated in the text.

Certainly, the number of stations changes over time, however, as the quality control (QC) process is applied for each day separately, the difference in observations' availability is rather an advantage than a drawback because it allows for using all the data series regardless of their length. Indeed, the models change for each day and location and, for this reason, the QC is iterative, meaning that once all the criteria have been applied, the process starts over again without flagged data, and all the process is repeated until no suspect data is detected. Details can be found in the 3 methodological papers (Serrano-Notivoli et al. 2017a, 2017b and 2019).

As described in the methodology (section 3, first paragraph), the process implies 3 stages: quality control, reconstruction of missing values and gridding. The latter uses the reconstructed data series, which means that the estimates of the final gridded dataset are built from the same 10 neighbors (15 in temperature) in all days of the period. Of course, the same number of stations are used for all the grid points (for all the stations at the reconstruction stage), even for those located in the boundaries of the country. In this particular case, the ideal situation would be having available data series from adjacent countries at daily resolution over the same time period, though we did not have the chance and we only used the stations within the borders of Slovenia.

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.

In the spirit of focusing on the Slovenian dataset creation and trying to be clear in the taken steps, we avoided most of the details concerning the methodological process, which are already widely described in the aforementioned published papers in this journal.

However, we would like to make clear to the reviewer what involves the wet/dry probability: for each day and location independently, the 10 nearest observations are codified as 0 or 1 depending on the absence or not of precipitation on that day, then a binomial GLM is fitted to these values using the corresponding longitude, latitude and altitude of the same nearest stations as covariates. The model is then applied to obtain a prediction at the target location using the same covariates (at the target location). If the predicted value (predicted probability), ranging from 0 to 1, is higher than 0.5, it is considered a wet day.

In order to determine a day as wet or dry, for each day and location independently, the 10 nearest observations were codified as 0 or 1 depending on the absence or not of precipitation on that day, then a binomial GLM was fitted to these values using their longitude, latitude and altitude as covariates. The model was then applied to obtain a prediction at the target location using its corresponding covariates values. If the predicted value (predicted probability), ranging from 0 to 1, was higher than 0.5, it was considered a wet day.

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

Internal coherence is referred to the verification of the maximum temperature is higher than the minimum temperature. We have stated that in the text now.

l. 110 standard deviation of what?

“Standard deviation of temperature”. Thank you for the notice, we changed it.

l. 113 all suspects were removed?

Modified as suggested.

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.

Thank you for your comment. Since we applied two methodologies (one for precipitation and one for temperature) that share most of the techniques in QC (except for the criteria, obviously) we tried to simplify explanations by: 1) stating first the criteria and 2) then explaining how RVs are created to apply those criteria. Based on your suggestion, we have changed this sentence to make it clear.

In order to perform the QC process by applying the above referenced criteria, we used generalized linear mixed models (GLMMs) and generalized linear models (GLMs) to calculate the reference values (RVs) for each day and location of the original dataset. The information from the 10 nearest stations varied according to the data availability.

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.

Thank you for your timely comment. It has been, indeed, one of the widely discussed issue in the methodological papers due to its great importance on the reliability of the final dataset, as you point out in your comment.

Observed values do not participate in the prediction of the estimates, which allows for making new predictions at any location and day. This means that, regardless of the presence or absence of observation, we are able to make a prediction of the variable at any time of a data series. In this case, the reconstruction process implied the estimation of new data for all days of all locations. Those days and locations with observations were used to check the goodness-of-fit of the models. The models were always different since data involved in each day and location is different. This method represents a great improvement in respect of previous gridding approaches because the estimates always preserve the local variability regardless of farther climatic behaviors in time and space.

Of course, this procedure implies that the uncertainty is different across years (actually, across days) and, for that reason, individual values of uncertainty are given for all precipitation estimates at all grid points and at all days of the period (for each variable, 2 datasets are provided: 1 with the estimates and 1 with their corresponding uncertainty). In the manuscript, we opted by representing the averaged uncertainty (actually, propagated uncertainty) as well as mean climatologies do with average climatic data.

To make clear this issue, and answering to a similar question by Reviewer #3, we added a new figure showing the temporal evolution (annual aggregation) of the uncertainty values in the gridded datasets just to check potential temporal biases due to the different availability of observations over time. We added this figure to the last part of the section 4.3:

The temporal evolution of the annual values of uncertainty (Figure 12) showed a homogeneous behavior in precipitation all over the period except from 2013, when uncertainty rose both un median and highest values. Maximum temperature showed also a relatively static behavior, with higher values of uncertainty in the initial decades (1950-1960s) and at the end of the period (2000s). the lower values were found in the 1980s. With a different pathway, minimum temperature played a different pattern in the first half of the period, with a decrease from 1950 to mid 1960s and then progressively increasing to 2018, being its maximums in 1993 and 2011.

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.

As stated in previous answers, further details on the methodological approach can be found in the aforementioned references. However, we would like to make clear to the reviewer the operation of the model. The explanatory variables are latitude, longitude, altitude and distance from the coast, and they all always participate in the creation of the estimates. We chose those variables because they are inherent to the climatic data, which means that we will always have available that information about the location. In the earliest stages of method development, we tested several options, including those using different covariates depending on the location and timestep. We observed that this produced a global bias due to the different influences of the variables. The use of all covariates as explanatory variables yielded the best results. Of course, different covariates such as land use / land cover, weather types, or any other factor important to precipitation and temperature spatial distribution would be a great complementary information but, unfortunately, those variables are not always available for all time period and at the required temporal and spatial scale. This has been largely discussed in previous works. Precipitation estimation consists of two stages, predicting first the wet/dry probability and then the magnitude of rainfall (if wet day is predicted through a probability higher than 0.5). This is done through a GLM approach (binomial and quasi-binomial, respectively) (Serrano-Notivoli et al., 2017a, 2017b). Temperature prediction is similar, of course avoiding the wet/dry part, and here we introduce an intermediate stage using GLMMs where the potential seasonal/annual effects are removed from the estimation to avoid biases (Serrano-Notivoli et al., 2019).

We added to the text (L.121-126) the information detailed in a previous question, that works to explain how the model is built from the original data and their corresponding covariates.

In regard of the uncertainty, it comes from the standard error of the individual models applied to each location and day and it follows a normal distribution, meaning that uncertainty informs about a range of probable values being the mean the estimated variable (temperature or preciptiation) and standard deviation the uncertainty. The text from the previous question, included now in the manuscript, helps to understand the different impact of the uncertainty over time. Also, the response of a question of reviewer #3 about uncertainty in L.143-145, is useful to understand how uncertainty was computed for indices:

To calculate the indices, we used a Monte Carlo approach replicating, for all the grid cells, 1000 random realizations based on a normal distribution using the value of the estimate as the mean and the value of the uncertainty in that day as the standard deviation. The indices were computed for all realizations and the final maps represent the median and the percentage of variation based on the standard deviation of all the realizations.

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.

Thank you for your comment. In order to estimate precipitation and temperature variables, the observations acted as dependent variables and the associated geographical information to each station participating in the model, i.e. latitude, longitude, altitude and distance from the coast, acted as independent variables. We have changed this sentence to make it clear.

In order to estimate precipitation and temperature variables, the observations acted as dependent variables and the associated geographical information to each station participating in the model (latitude, longitude, altitude and distance from the coast) were the independent variables.

3. Results

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

Thank you for your suggestion. This article describes a new dataset for Slovenia that was created with an already validated methodology. Previous works using the same methods and published in this journal followed the same structure. The model operation is explained in section 3 and it does not imply the selection of covariates since they are the same for the entire temporal period and spatial domain.

The first two sections, 4.1 and 4.2, actually explain the specific results for Slovenia derived from the application of the model, including its performance by percentiles, elevation ranges, months, by stations and by days. Section 4.3. describes the spatial distribution of the uncertainty for exemplar indices, serving its spatial coherence as evaluation of its performance.

4. Discussion

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

Thank you for your suggestion. The fit with the observations is explained and discussed, providing reasons about potential biases. However, we have included several lines explaining further implications arisen from the dataset estimates and uncertainties. For example, we now state the lower reliability of estimates in high-elevation areas due to the scarcity of observations (L.3555-359):

The reason for this is the complexity of the high geographical variability between stations, which could, to a certain extent, also influence the underestimation of minimum temperatures at higher altitudes. This is shown in the LOO-CV by altitudes, where the bias in the estimates at high elevations is probably due to the lower number of available observations. Grid estimates at these altitudinal ranges (especially >2000 m a.s.l.) may reflect the lower reliability through higher values of uncertainty.

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.

Your comment is, indeed, very appropriate. However, our method is completely different from other interpolation methods in this regard since the estimates are made for specific grid points and not for area-averaged grid boxes. In this particular case of Slovenia, each grid points uses, as covariates in the model, the median longitude, latitude, elevation and distance to the coast of a 1x1 km spatial resolution. Regarding the extreme values smoothing, previous analyses on published works showed agreements between estimates and observations of Pearson>0.93 for daily precipitation and Pearson>0.95 for daily temperature. Of course, it is possible to find several smoothed values over the thousands of considered days, but it was demonstrated that the methodology does not introduce any bias in final results.

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

Modified as suggested, we added the percentage of original missing values of precipitation and temperature.

The original missing daily data, considering the whole period (1950-2018) for all stations were 60.2% for precipitation, and 59.7% for maximum and minimum temperature. The quality control was applied to the remaining 39.8% and 40.3% of the data, respectively.

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

Yes, it is the total number of removed precipitation and temperature observations per year. We changed the caption of the figure to make it clear.

Figure 3: Total number of removed precipitation (a) and temperature (b) observations from the original dataset by the different criteria per year. The criteria are explained in more detail in Serrano-Notivoli et al. (2017, 2019).

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

Thank you for the suggestion. It is, certainly, not so clear. We meant that, from the removed data, suspect data and suspect zeros were the most frequent situations. Considering that, only a final 1.26% of the total observations were removed. We changed the text to explain that.

From the removed data, the Suspect data and Suspect zeros criteria, which usually represent estimates of low precipitation values based on nearest observations, were the most frequent. While these criteria are commonly frequent in this kind of QC (Serrano-Notivoli et al., 2017, 2019), only 1.26% of the original raw observations were removed.

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

Yes, there is actually a probability-based threshold to define a wet day (probability>0.5) as explained in a previous answer, based on nearest observations. Lines 162-164 explain that, when zeros and very low values prevail in these observations, the probability of dry day is higher.

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

Figure 4 represents the comparison between observations and estimates. Upper row shows this comparison by stations (each dot is a station) and lower row by days (each dot is a day). The columns show the different aggregations, from left to right, mean values, medians in wet days and 95th percentiles. This means that, for example, the first plot (up-left) shows the comparison between the mean observed and estimated daily precipitation, by stations, and the last plot (bottom-right) shows the comparison between the 95th percentile of daily precipitation of all stations, by days (one dot is one day with the 95th percentile of precipitation).

We changed the caption to make clear its interpretation:

Comparison between measurements and estimates by stations (upper row) and by days (lower row). Left column: comparison of mean precipitation; central column: comparison of median precipitation in wet days; right column: comparison of 95th percentiles in wet days. Dashed lines represent ±1 standard deviation of the data.

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

Thank you for your suggestion. We prefer to maintain the point-based representation since this is not a continuous variable to be represented with lines. Each category in Y-axis is independent and the goal of the plot is to show the differences between observations and estimates by temperature ranges. Only when this difference is not too small (which only occurs in one case), both dots (solid and transparent) are connected with a line.

 # Michael Grabner (Referee)

The database seems to be very useful also for dendroecology. Did the authors think to provide a tool or some web application for downloading climatic data just for calculated and selected grid points (maybe in specific selected time periods) in a more widely used format?

Thank you for this comment. The database is already freely available for download from Zenodo: https://doi.org/10.5281/zenodo.410854 in NetCDF format. However, downloading and analysing the data in this format might not be very manageable for a wide range of users. To solve this issue, we have developed a web page with the SLOCLIM dataset with a user-friendly interface: http://www.sloclim.eu. On this web page, users can select any date in the period 1950-2018, and any location on the map of Slovenia to obtain the data (maximum and minimum temperature and amount of precipitation) for each day of the selected period. The data are available for download in .csv format and the system allows for downloading long data series at once. This website provides an easier access to the data for a wider community. Instructions for users and general description are shown on the website. As Dr. Maks Merela contributed in the creation of the web-based access and in the translation of scientific formats to a wider audience, he has been included as co-author of this article.

M. Tomas-Burguera (Referee)

In this paper the authors present a high-resolution (1 x 1 km) dataset of daily precipitation and temperature for Slovenia for the 1950-2018 period. The interest of this dataset is high as it covers a gap in the climatology of Slovenia, a region with both climate and terrain complexity, as mentioned in the manuscript.

The authors used a previously designed methodology by two of them, that has already been tested in Spain and other regions. This methodology is based on estimating daily data of the variable of interest by using GLMs/GLMMs, daily data of the 10 closest weather stations and geographic information, such as latitude, longitude, altitude and distance from the coast. The main advantage of this methodology is that all available data can be used, even if weather stations cover short periods.

The paper is well written and the use of climate terminology is adequate.

I also consider the paper is correctly distributed in their sectioning, and the authors provided different verification tests of the variables, both in time and space.

I would like to make some general comments to this paper.

- While the authors used a method based on the spatial structure of climate data, other methods relies on the temporal structure of climate data to estimate missing data.

My first comment is related with this point, as I would like to know the opinion of the authors regarding strengths and drawbacks of the methodology they used compared with methods based on the temporal structure of the climate data.

Thank you for your constructive comments about the extensive work that we performed over Slovenia by considering all the available climatic information. The potential of using all the data is the most important advantage regarding those methods that only consider the temporal structure (TS) of the data series.

TS methods require neighboring series having a minimum length and overlapping in significant periods, leading to the rejection of short data series and removing precious information for the reconstruction. Also, two premises are assumed in TS methods: 1) data series and their relationships are stationary in the overlapping period; and 2) they have a similar temporal structure. This situation is aggravated with daily data due to the great spatial variability of the variables.

Slovenia is a highly diverse region in regard of climatic characteristics, with large and abrupt variations of environments in small areas. As a consequence, the proper estimation of climate at high temporal and spatial resolutions require the finest methods with complete and well spatially distributed data, avoiding the assumptions of *a priori* climatic behaviors.

- Also regarding the methodology. If I have understood correctly, the authors first estimated missing data of weather stations and in another iteration of the algorithm they estimated climate data at each grid cell of 1 x 1 km. If this is correct, it turns out the authors used estimated data as if they were observed data. Is this correct? If so, what is the possible impact of this on the obtained results in the opinion of the authors?

Thank you for the comment, which is correct. As indicated in section 3, first paragraph, after the quality control of the original raw data, new estimates are created for all the missing values of the original data series. Then, the same statistical procedure is applied to each of the grid cells. This approach ensures that the grid cells will have the same neighboring data series all days of the reconstruction period, avoiding potential inhomogeneities resulting from the selection of series with different structure to estimate data at the same grid cell.

However, the estimated data is not equated to observations since we provide an individual uncertainty for all the estimates. This uncertainty, arising from the error of the individual models, helps to understand the reliability of the data. Since all the gridded data are estimates (unlike the reconstructed data series, which also contain observations), all the grid cells have associated 2 values: 1 estimate (precipitation or temperature) and 1 uncertainty value (in the same units). Like the rest of gridded datasets, SLOCLIM is a representation of the reality, except that we provide a measure of the reliability of each portion of space and time.

This issue was largely discussed in Serrano-Notivoli et al., 2017 and 2019, both works published in this journal. You can find the details at the links provided as a response to the first reviewer.

- Slovenia is a complex territory, with high elevation areas. Unfortunately, only a few number of stations are available above 1000 m. While some of the validation tests are shown in terms of elevation, some discussion regarding its possible impact on the obtained dataset is missing.

Thank you for the suggestion. Indeed, the scarcity of observations at high elevations is a common issue in grids creation all over the world, for this reason we opted to show a specific validation based on altitudinal ranges. However, the uncertainty values of grid cells at high elevations already show larger errors (lower reliability) than those at lowlands. To state the potential impact on the estimates, we added a few lines to the discussion section:

The reason for this is the complexity of the high geographical variability between stations, which could, to a certain extent, also influence the underestimation of minimum temperatures at higher altitudes. This is shown in the LOO-CV by altitudes, where the bias in the estimates at high elevations is probably due to the lower number of available observations. Grid estimates at these altitudinal ranges (especially >2000 m a.s.l.) may reflect the lower reliability through higher values of uncertainty.

- Station network changes all over the period, with an initial increase in the number of stations followed by a constant decrease after the 80's. This constant change on the number of true observations can have a great impact on the quality of obtained results. While validation of the results is presented by altitudes and by months, why the authors did not provide a validation based on decades (for example)? I think it could be really interesting to assess the impact of changes in the station network in the obtained results.

Thank you for the suggestion. It's true that the availability of the observations decreases from the end of the 70s until 2018. However, this is apparently not related to the quality of the data since the quality control did not detected a higher number of incoherent values from that date (Figure 3). Actually, the number of flagged data dramatically decreases from the 80s, showing a more spatially consistent information. This is probably due to an effort of homogenizing the meteorological network, despite the lower number of stations.

Additionally, following your recommendation we made a new figure showing the temporal evolution (annual aggregation) of the uncertainty values in the gridded datasets just to check potential temporal biases due to the different availability of observations over time. We added this figure to the last part of the section 4.3:

The temporal evolution of the annual values of uncertainty (Figure 12) showed a homogeneous behavior in precipitation all over the period except from 2013, when uncertainty rose both un median and highest values. Maximum temperature showed also a relatively static behavior, with higher values of uncertainty in the initial decades (1950-1960s) and at the end of the period (2000s). the lower values were found in the 1980s. With a different pathway, minimum temperature played a different pattern in the first half of the period, with a decrease from 1950 to mid 1960s and then progressively increasing to 2018, being its maximums in 1993 and 2011.

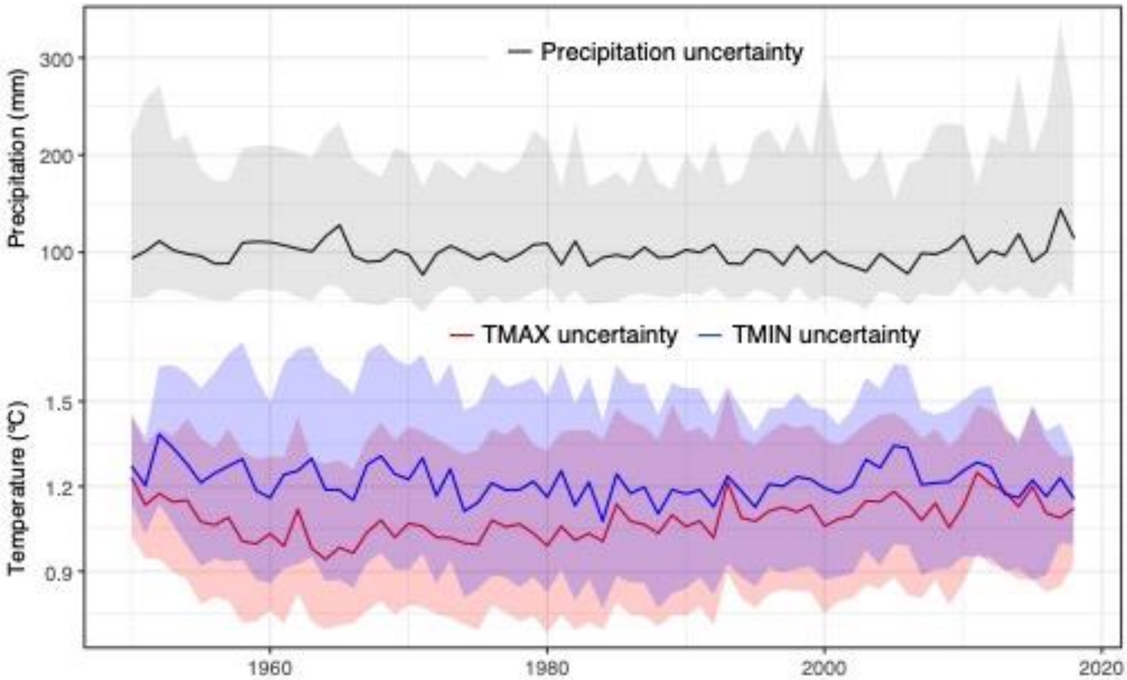

Figure 12. Temporal evolution of annual median values (lines) in precipitation and temperature gridded datasets. Shaded areas show the range between 5th and 95th percentile of uncertainty values.

- One of the key points of the used methodology is the estimation of the uncertainty of each estimated data. The authors should provide more information regarding the calculation of the uncertainty and some kind of validation of the estimated uncertainty could also be provided. Maybe a comparison between uncertainty of estimated data and MAE? A temporal perspective of the estimated uncertainty would also be of high interest to evaluate the impact of changes in station network on the uncertainty of obtained data.

Thank you for your comment. Methodological details are discussed in previous publications in this journal (see second response and provided links to the first reviewer). We preferred to focus the paper on the specific data implications for Slovenia. While a comparison between uncertainty and MAE may help to evaluate the uncertainty of reconstructed data series it is not useful for gridded datasets since they don't have observations to compare. To respond your request of looking at a temporal perspective of uncertainty we made a new figure (Figure 12), already explained in previous comment, and added to the text.

I also have some specific comments

Lines 81-82. "Although some data series begin before 1950, we decided to limit the research to the years 1950 to 2018, when the station network remained stable over time and space". I don't fully understand this sentence. When the authors said that the station network remained stable over time and space, what exactly do they refer to? In figure 2 it is clear that the number of stations reduced constantly from the maximum number slightly higher than 100 to a number of only 20 in 2018.

The availability of daily observations before 1950 is really scarce and spatially biased. We chose to start in that year because, despite the large variation over time, it represents a constant availability of more than 20 stations covering the whole country. We consider that as the minimum number of representative observations of the great diversity of climates in Slovenia.

- In Figure 9 and figure 11 the authors provided the estimated uncertainty for some derived indices of temperature and rainfall. While in the methodology section the authors mentioned how they estimated the uncertainty of each estimated data of temperature and rainfall, how the authors derived the uncertainty of the indices?

Thank you for your comment. We added to the methodological section the explanation about the computation of the indices and their uncertainty:

To calculate the indices, we used a Monte Carlo approach replicating, for all the grid cells, 1000 random realizations based on a normal distribution using the value of the estimate as the mean and the value of the uncertainty in that day as the standard deviation. The indices were computed for all realizations and the final maps represent the median and the percentage of variation based on the standard deviation of all the realizations.

I would like to make some comments/suggestions regarding some of the thresholds used in temperature quality control.

- "For temperature we used also five criteria: (1) internal coherence;" I assume the authors refer to the coherence between Tmax and Tmin when they write "internal coherence". This should be clarified in the manuscript. Just for curiosity, is this coherence test based on Tmax > Tmin or Tmax ≥ Tmin?

Internal coherence is based on the assumption of Tmax > Tmin.

- The following two comments are more a suggestion for successive works or some update of the database than comments to modify this manuscript.

I think the thresholds used to classify a daily data of temperature as an "out of range" data could be enhanced, especially those used for minimum temperature.

While in the introduction, the authors says. "Moreover, temperature ranges from -35 to +40 °C (BertalaniÄ□ et al., 2006) show the very extreme character of the seasons". Then, in the explanation of the quality control they say: "(3) removal of those days out of range considering maximum temperature (TMAX) ≥ 50 °C or TMAX ≤ –30 °C and minimum temperature (TMIN ≥ 40 °C or TMIN ≤ –35 °C)"

If temperature ranges from -35 to +40ºC, the consideration of Tmin= -35ºC as an out of range value could imply deleting real extreme values. On the other hand, I consider as too high the Tmin>= 40ºC threshold.

Thank you for your suggestion. We appreciate the recommendations, which we will consider in further research. We based the threshold of minimum temperature in the minimum absolute recorded temperature in the original database (-34.5 ºC). The upper threshold of 40ºC is probably too high for minimum temperature. However, this part is only the initial removing of non-possible data, and the quality control method already have, in further steps, procedures to compare data between stations and find spatial incoherences.

- ";(4) removal of all days in a month with a standard deviation equal to zero (suspect repeated values in the series);" With this criterion, the authors only considered repeated values when all the month has the same value. Well, this criterion is as valid as others, but I think this could be easily modified to be more restrictive. Many different thresholds are used in the literature, but when temperature data with decimal precision is provided, 10 consecutive days showing the same temperature are very unlikely. I would suggest the authors to use 7-10 consecutive days as a new threshold.

Thank you for the suggestion, we are working on improving the detection of repeated values at a daily scale and your recommendation will be considered.

Table 5.- December is missing.

Thank you. Modified as suggested.

Figure 1.- As the authors mentioned the complexity of terrain in Slovenia in the manuscript, they could represent a digital elevation model in this figure to help the readers of the manuscript to better understand this complexity. In this figure, the duration of data of each meteorological station could be also represented.

Thank you. Modified as suggested.