RESPONSE LETTER

Review of Manuscript No.: essd-2019-52
Title: STEAD: A high-resolution daily gridded temperature dataset for Spain
Author(s): Serrano-Notivoli R, Beguería S, de Luis M

GENERAL COMMENTS

Overall this is a thorough piece of research, and the content is entirely suitable for this journal. The gridding method, including the quality control procedure, are well documented, but some clarifications are necessary prior to publication.

> Thank you for your comments. We tried to be clear in the description of the methodological process and the resulting products. We hope that the changes made in the manuscript are fair enough for a straightforward understanding.

SPECIFIC COMMENTS

A common problem with datasets that grid tmax and tmin separately is that there is no guarantee that tmax will be greater than tmin in the final dataset. I have checked the dataset and there are several days where tmax<tmin values occur in certain grid cells. These mostly occur across the edges of the gridding domain. The highest frequency (11) of tmax<tmin is during the year 1996, although such occurrences are apparent for most years. I do not advocate changing the methodology to account for this, but this limitation needs to be highlighted in the paper.

> Thank you for checking carefully the dataset, we appreciate the concern. We realized that the available dataset at the *digitalCSIC* repository is a previous version rather than the final one. We have uploaded the correct version, which was used to make the analysis in the manuscript. Sorry for the mistake. The methodological process creates new estimates for maximum and minimum separately. However, all the stages, from the quality control to the gridding, consider the differences between them to regularly check the internal consistency, so it is impossible to get situations in which tmax<tmin.

The methodology is broadly similar to the method used in the SPREAD precipitation dataset, produced by the same authors. I refer to the use of Reference Values and Generalized Linear Models, as for precipitation the skewed nature of the data and zero-cutoff were also taken into account. Nonetheless, given that there is a connection between STEAD and SPREAD I think that the precipitation dataset needs to be mentioned earlier in the introduction, and in the Methods Section the differences in the method used here for the temperature variables should be indicated.

> Based on your suggestion, we added an extended explanation of SPREAD-STEAD relationship in the introduction:
>
> *"The experience acquired in the SPREAD dataset (Serrano-Notivoli et al., 2017a) development set the basis for a solid and reliable daily gridded precipitation datasets creation. Using the same framework with a complete renewal of the core calculations, we developed a new methodology for daily temperature datasets reconstruction and grids creation."*
>
> and also at the beginning of the methods section:
>
> *"The key stages of the methodological process (calculation of RV, quality control, gap filling and gridding) are the same to that used to create the SPREAD dataset (Serrano-Notivoli et al., 2017a). However, the method basics are completely different since the RV creation has been refined, the quality control has been adapted to temperature data, and*

the gap filling and gridding processes include now an improved standardization procedure."

Section 2: Information needs to be provided about the time schedule over which the daily maximum and minimum values were recorded. This may not be available for all stations but where available it should be described briefly in this section, e.g. are tmax/tmin calculated over the full 24-hour period and does this change over time.

Since we used the daily products provided by AEMET and MAGRAMA, we don't have the information about the moment of the day in which maximum and minimum temperatures were recorded. Anyway, this issue doesn't affect to the temperature estimates because the method creates a prediction for each original observation, independently of when it was recorded. The final reconstructed series faithfully represents the temporal structure of the original ones, regardless of the moment of recording.

We added a sentence in section 2 explaining that we used the daily maximum and minimum values of temperature observations:

"Daily maximum and minimum values of temperatures series were used from all the observatories."

Section 2: As pointed out by reviewer #2, the changing number of input stations over time can have a profound influence on the gridded data, and is important for users who want to calculate long-term trends from the data to be aware of this. This limitation of the dataset needs to be highlighted.

We contributed to the minimization of the impact of the data availability over the time introducing a standardization procedure. The methodological approach creates spatial references that are standardized with the temporal structure of the series to avoid biases or incoherencies. Furthermore, the provided uncertainty values for each of the estimates inform about the reliability of the data. This was already mentioned in the discussion section (6th paragraph). However, we emphasized this interesting subject in section 2:

"[…] Despite the differences in the data availability through time, the methodological process creates spatial references that are standardized with the temporal structure of the series to avoid biases or incoherencies. In this regard, the chosen spatial resolution accurately reflects the local characteristics of daily temperature in most of the temporal period, while the provided uncertainty values help to understand the reliability of the estimates when the original data have higher variability."

Section 4.2: The verb "depurate" appears to me to be wrongly used for this procedure. Suggest changing to simply "Quality-controlled dataset".

Modified as suggested.

Section 5 (Discussion) lines 20-23: The key point about producing these gridded datasets is that the final values should reflect grid-box average values that are based on limited spatial sampling (unless the method produces values representative of point- values, which I understand that it does not). This relates to the comments by reviewer #2 about the choice of gridding resolution. This general aim of gridding is not articulated well in this discussions section and needs revision.

Whereas your comment is very interesting, maybe there is a misunderstanding at this point. The temperature estimates (as well as their corresponding uncertainty values) are created for specific individual locations represented by 4 parameters (i.e. latitude, longitude, altitude and distance to the coast). The representativeness of the grid-box in each case falls on that each of those parameters are the median value of all covering that area. For instance, in the STEAD dataset, each gridpoint is the centroid of a squared area of 5 x 5 km with the median of all the possible values of the parameters covering that area.

We added an explanation in the description of the grid (section 2) to avoid misunderstandings:

"[…] The predictor parameters (i.e. latitude, longitude, altitude and distance to the coast) for each grid point were computed as the median of all the possible values of those parameters, covering an area of 5 x 5 squared km in which the grid point is the centroid. […]"

Abstract and Conclusions: One of the key aspects of this paper is the use of many more stations than used in other datasets for the region. This needs to be stated more clearly in the abstract and conclusions, as the phrase "full total of available 5520 observatories" does not convey to the reader (especially those not familiar with the station density across the region) that this is an important feature of this dataset.

Despite the full total available number of stations is 5,520, this figure is for the whole period, so the station density has greatly varied though time. However, we included in the abstract and conclusion, as suggested, the theoretical density:

"[…] (about 1 station per 90 km2 considering the whole period) […]"