

## ***Interactive comment on “STEAD: A high-resolution daily gridded temperature dataset for Spain” by Roberto Serrano-Notivoli et al.***

**Anonymous Referee #3**

Received and published: 15 May 2019

### **\* 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.

### **\* 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

C1

most years. I do not advocate changing the methodology to account for this, but this limitation needs to be highlighted in the paper.

- 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.

- 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.

- 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.

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

- 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.

- Abstract and Conclusions: One of the key aspects of this paper is the use of many

C2

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.

---

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2019-52>, 2019.