

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

Anonymous Referee #2

Received and published: 4 May 2019

General Comments

This manuscript describes a gridded temperature data set at daily and 5 km x 5 km spatial resolution for Spain. Compared to existing similar data sets it represents a valuable contribution considering the data set temporal and spatial resolution, the period covered (1901 to 2014 for continental Spain) and the powerful methodology employed, which favours the use of a high number of stations and the computation of uncertainty of the estimated variables. Overall I think it is a well written manuscript and the data set may indeed be useful for future studies, therefore it is a perfectly suited article and data set for this journal.

However, I think it requires some clarifications and corrections before being acceptable for publication. For example it is not clear to this reviewer why the spatial resolution

C1

used was chosen and what is the impact of the variable station number (and therefore density) upon the quality of the estimates during the different years considered in the dataset. These details may influence further studies (such as in the interpretation of temporal trends) so I think they should be commented here, in the description of the data set proposed. As explained in more detail below, I propose to modify or add some of the existing figures related to this aspect and also to the global uncertainty and mean error time series.

Another aspect that should be clarified are the data sources. The manuscript describes two institutions (AEMET and MAGRAMA) but in the data set description link provided by the authors in the manuscript, if I understood correctly, 12 providers are listed (which, according to descriptor "dc.relation.isbasedon" are: Spanish Meteorological Agency (AEMET); Ministry of Agriculture and Environment; Servei Meteorològic de Catalunya (METEOCAT); Navarra Government; SAIH Cantábrico; SAIH Duero; SAIH Ebro; SAIH Guadalquivir; SAIH Hidrosur; SAIH Júcar; SAIH Miño-Sil; SAIH Segura; SAIH Tajo), so that implies 10 sources more than those mentioned originally in the manuscript. Obviously one of the two descriptions is not correct so this should be clarified and corrected.

Finally, there are a number of formal aspects that should be amended such as the use of correct symbol units, references quoted in the text but not listed in the references section, or minor problems with English language (I do not intend to be exhaustive in this aspect but I list some issues below). For all the above, despite this is a very interesting contribution, I do not think the manuscript can be accepted in its current form and I recommend major revision.

Specific Comments

1. Page 2, line 20. Please add here the horizontal resolution of the E-OBS dataset, i.e. "with a horizontal resolution of .." or something similar.
2. Page 2, lines 23. The first part and last parts of the sentence "The Spanish territory

C2

perfectly captures... this variability" is unclear and potentially misleading... what do you mean by "the great climatic variability"? The climatic variability of the world? I suggest to rewrite it, for example "The Spanish territory exhibits a great climatic variability with very different regimes in a relatively small area that leads to high risks such as.." (please complete with what you think are those risks) or something similar.

3. Page 3, line 13. Please complete the sentence with all data sources (as mentioned earlier) if there are more than two - or correct the 12 data sources given in the data set description repository.

4. Page 3, figure 1. Map bottom panel of Canary Islands: please enlarge axis font sizes they can hardly be read

5. Page 4, line 2. Please give a range of mean distances between stations, as it is done later for mean station elevations.

6. Page 4, Figure 2. This is a very important part of the data set description, as it gives an idea of temporal changes in mean altitude, minimum distance and number of stations. The methodology used by the authors allows the use of a variable number of stations, not constant in time, which is good because it maximizes the information introduced in the estimation. However, this introduces variability due to the changes in the station data set so this should be carefully described. Authors chose to show mean values only, which I think it is rather limited, and combined in the bottom panel two variables. I suggest to consider a three panel figure, one for each of the three variables considered, and to plot for each one the median and percentiles 25 and 75 - alternatively, if variables examined follow a Gaussian distribution mean values plus standard deviation could be also a possibility - however I would favour the first option. Then the new figure should be briefly commented, and in particular, justify properly (here or in section 4.1) the selection of the 5 km x 5 km horizontal resolution, which in the current version seems arbitrary in this section (in the discussion the reader finds that it is consistent with an already existing precipitation data set for the same region,

C3

but this is not either mentioned in section 2 and could be also considered).

7. Page 5, section 3.1. Please introduce properly the meaning of TMAX and TMIN.

8. Page 10, line 25. How many iterations of the procedure described are typically required?

9. Page 11, line 17. Pearson -> Pearson correlation coefficient ? [at least the first time you mention it]

10. Page 14, section 4.1. Please justify or just comment briefly (if justified previously) the selection of the horizontal resolution.

11. Page 14, line 7. Suggest: case -> case using a leave-one out cross validation (LOO-CV) [this is mentioned later but not in the text]

12. Page 14, line 9. I think there is a rounding problem here and 0.95 should be 0.96 according to figure 7c - please check.

13. Page 14, figure 7. It is good the authors chose to show the four panels with the same x-axis and y-axis ranges to compare them properly. However, the panels are shown as rectangles and not as squares thus using a different scale for x and y axis. Could you please fix this?

14. Page 14, figure 7. Please indicate (in the figure caption is fine) the meaning of the dashed lines - confidence intervals perhaps? In that case specify the level.

15. Table 2 (and 3). Please use correct units where needed.

16. Table 2 title. When enumerated the variables displayed on the table, the last one is "Range: minimum and maximum..." but on the last row, first column, it appears "Pearson". Please correct. Note that including the units, as requested above, may help detect these problems.

17. Table 2 (and 3). When you say decimal places you mean decimal digits? Please

C4

check.

18. Table 3. Suggest adding in the first row, first column, the label "Altitude (m)" , referred to the values listed on the first row, next columns.

19. Page 18, Figure 11. This is a very important part of the study and I think it deserves more attention. The uncertainty (upper panel) initially increases, ca. from 1900 to 1905, and then decreases from 1975. However, if I understood correctly, authors mention only a decreasing trend (page 17, line 10). Could you please clarify this? I suggest adding a background grid to the figure to allow an easier visual analysis.

20. Page 18, Figure 11. I think that, given the variability of terrain heights in Spanish territory and different station densities at different altitudes, it is necessary to stratify Figure 11 into an additional 6 panel figure, considering to split the aggregated uncertainty and mean error values into different station altitudes. Looking at Table 3 altitude classes, probably 3 station groups, for example those with altitudes contained in the following intervals [0, 500), [500, 1500), [1500,) m above sea level, would be enough. Future studies examining aspects at different terrain heights may largely benefit from this additional figure to better interpret subsequent results.

21. Results given in sections 4.3 and 4.4 refer both to Figure 12. I suggest to split that figure into two figures (first and last four panels respectively) so that it is easier to read the comments referred to each part of the figure.

22. Regarding Figure 12f and 12h I noted that the uncertainty values (expressed in days) over the islands (both Balearic and Canary Islands) are either very low (Figure 12f) -except for the highest terrains in Canary islands - or very high (Figure 12h). Could you please comment this result?

Technical Comments & Minor Details

23. Page 1, line 24. Typo?: team -> teams? Please check meaning and correct if necessary.

C5

24. Page 1, line 26. Jones et al 2010: reference quoted but not listed in references section. Is it perhaps Jones et al 2012?

25. Page 1, line 26. Please check citation journal style: Willmott and Matsura (1900-2014) (2001) -> Willmott and Matsura (2001) ?

26. Page 1, line 29. Check citation journal style: 2015 and 2018 -> 2015; 2018?

27. Page 2, line 13 and line 15. "e.g.:" -> "e.g." as in line 9, same page?

28. Page 2, line 27. English: did not considered -> did not consider

29. Page 3, line 12. check: down -> bottom map panel

30. Page 3, line 12. Aemet -> AEMET

31. Page 3, line 18. English, suggest: this moment -> then [or "that moment"]

32. Page 5, Figure 3 caption. Suggest: RV -> Reference Values (RV) [I know it is already defined in the text, but this change improves the readability of the figure]

33. Page 5, section 3.1. Please use correct Celsius degree symbols as you have done elsewhere in the manuscript.

34. Page 5, line 7 (and elsewhere in the manuscript). English: please check the meaning of suspected, suspect and suspicious and use properly.

35. Page 8, line 8. English: supplemental -> supplemental material

36. Page 9, line 1. English: finish -> finishes

37. Page 10. For some reason, in this page, text before equations end with a "." and not ":" as it is done elsewhere in the manuscript - please check.

38. Page 11, line 2. English, suggest check: obtain -> obtaining

39. Page 12, line 11. English: "fulfil" is correctly spelled, but it is in the British form - being the American form "fulfill". As you use previously in the manuscript "neighbor"

C6

(American form) this is not consistent: you should chose either American or British forms, but not a mixture.

40. Page 13, line 5. English: what -> which? Please check and correct if necessary.
41. Page 13, Figure caption 6. upper (bottom) line -> upper (bottom) row
42. Page 16, line 14 (and elsewhere in text): a X% -> X% (remove "a" if only values are given)
43. Page 16, line 15. English: slightly -> slight [as in line 17 same page]
44. Page 17, line 2. Please expand "approx."
45. Page 18, line 13. Sentence "The northern half..", please check English - by "than" you mean "to"?, otherwise compared to what?
46. Page 18, last line. higher -> highest?
47. Page 21, line 14. Hofstra et al 2008: reference quoted but not listed in the references section. Idem in same page references Jarvis et al 2001, Jones et al 2010 (is it 2012?), Kurtzmann and Kadmon 1999 (is it 2009?) (next page) Hubbard et al You et al ... please check carefully and make necessary corrections.
48. Page 21, line 11. English: lies > lied?
49. Page 21, line 14 English: comes -> come
50. References: please check alphabetical order - the last one (Van Den ...) should not be there.

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