Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-43-RC1, 2019

Open Access Earth System Science Data Discussions

Interactive comment on "seNorge_2018, daily precipitation and temperature datasets over Norway" *by* Cristian Lussana et al.

Anonymous Referee #1

Received and published: 14 May 2019

The authors present a new version of the seNorge dataset which is a gridded dataset for precip and temperature based on observational data. The resolution of the dataset is high with 1 km - especially considering the challenges the complex terrain poses to this effort. A state-of-the-art scale-separation approach is used for gridding which adds detail with subsequent additions of maps with even finer scales. Uncertainty estimates are made and uncertainty is related to station density.

The study is interesting and relevant and the manuscript fits within the scope of the journal. The group of authors are leading in this field and approach in the paper can serve as a guide for other NMHSs to provide gridded dataset. However, the main concerns relate to the readability of the manuscript. There will be specialists among the readers of ESSD who might disagree with this concern, but to have an impact which

Printer-friendly version

Discussion paper

is beyond the small group of specialists in advanced geostatistical techniques, a little more background and explanation is required. In addition, there are some issues on the approach that need further clarification.

Main concerns

*) The story line is sometimes hard to follow and the readability would improve with some more illustration. For instance: a brief introduction to optimal interpolation could be helpful on page 5. Another example are the scales you introduce on page 9 (line 25-27). A visualisation of some of these scales would help in guiding the reader to understand the approach. Similarly on page 10, line 15-16: can you visualize this scale somehow and show how this scale changes over time?

Figure 1: I understand that the explanation of the colour coding in fig 1a and 1b is complex, but now the reader has to read a substantial part of the article first before he/she understands what you are plotting here. A intuitive explanation for IDI might help for the reader who has a look at the figures first before deciding to read the paper. Explanation of abbreviations IDI and CV-IDI also helps. In the precipitation plot I'm missing the station locations.

*) A smaller issue is the structure of the text, a critical look would help here. For instance, on page 3, line 10 you start to claim that your approach will capture field variability at unresolved spatial scales. The next line is not an explanation how this is achieved, but deals with something quite different. It is untill line 17 that the reader is informed how you take-in the information on the unresolved spatial scale. Another example is on page 5, line 30. You write '...and Fig 1 shows those regions'. It would help if you guide the reader more explicitly where to look in Fig 1 (which regions/colours, which subfigure).

*) Relating to the interpretation of the results: Figure 2 shows that the analysis of TN is increasing with increasing CV-IDI Both the background and CV-analysis are decreasing with CV-IDI, but for summer this is not seen in the analysis. In winter this effect is absent

as well. My first guess was that this might be the influence of the urban stations, picking up the urban heat island effect. Can you comment on this?

*) In addition: In the introduction there is a paragraph concerning the effective resolution of grids. This makes me curious about the effective grid resolution of seNorge_2018 compared to seNorge2. Is there a way to quantify this? This is an interesting aspect, since the number of station observations (∼density of network) is similar in both datasets.

Other points the authors may want to look into

*) On page 4 (bottom) you describe the increase in station density and that many of these stations are installed in cities and villages. I was wondering if this aspect would give you a possibility to assess the Urban Heat island effect in Norway's larger cities? A comment on this would be interesting.

*) Page 5, line 25. Wouldn't the complexity of the topography be a relevant function here, and if so, have you looked into topography complexity? (slope, aspect, elevation)

*) Page 7, line 4-5. It is a good thing that physical consistency is enforced. A brief explanation how this is done is helpful, i.e. are you simply setting tn to tg where you find that tn is larger than tg, or is there a slightly more sophisticated approach?

*) On page 8, line 4: what is the motivation to choose a 50x50 grid? Where there any sensitivity analysis to support the choice?

*) The usage of 100 scales with a minimum of 2 and a maximum of 1400 km is unclear to the reviewers. Can you visualize some of these scales? Is there a more graphical way to explain how you use these successive scales for downscaling. Related to this, can you visualize the critical scale mentioned on page 10, line 15? Such a graphical representation convinces readers about the innovation of your method.

*) page 11, line 24: can you comment if you think there are other ways that might alleviate this problem with TN without having to install new stations?

*) page 12, line 5: here you claim that the addition of the land area fraction in equation 7 improves the temperature fields. What you are showing is the difference. Intuitively I see where you are going, but showing an improvement requires the cross validation, and the reviewer has not seen evidence that the new dataset improves considerably along the coastline.

*) page 12, line 31: I agree with your statement, but the reverse does not seem to be true. In the Oslo fjord the station density is very high but the TN quality is as low as in less dense regions.

*) The results section starts with the description of the CV. There are some concerns about the validation methods. A LOOCV or random sampling with k=folds does assume data points are spatially independent. This does not hold for data dense regions, moreover these data points will be predicted accurately due to their spatial dependence (especially using LOOCV). It is expected that the current approach will result in an underestimation of predictions errors. A way of making the validation procedure less spatially dependent (and less computationally expensive) could be to splits the datapoint into k-equal area folds.

*) page 6, line 5,8: What is the motivation to choose the numerical boundaries for CV-IDI to indicate data dense and sparse regions?

*) page 10, line 19: This is not a general description of CV but LOOCV.

*) General remark on figures: in most of the figures I'm missing the subfigure annotations (a,b,c). Please include a raster grid with lat/lon for the spatial plots. For regularly gridded raster plots one scale bar is sufficient.

*) Figure 4: Good to include lat/lon averaged differences, am I assuming correctly that the grey area are the min/max values (between -4 and +1.5)? The lateral and bottom panels y-axis temperatures are hard to read. This does not support the text on page 11, line 28, which suggests that almost all differences are between -2 and +1.

*) page 7, line 18: what does the i mean in the definition of G and S as these latter quantities appear not to be related to grid point i?

*) Table 1: The caption states that the third station is used while in the text (page 8, line 33) the average distance to the nearest four stations is used.

Very minor issues

*) page 3, line 27: “Finally, Section 4...” → This suggests there is no chapter 5.

*) page 7, line 18: typo in gridpoint

*) page 10, line 7: I assume “to have the same error”

*) page 13, line 8: Shouldn’t 0.4 be 40%?

*) page 13, line 18: I guess this should be “paper of Lussana”

*) page 17, missing pages in the Reistad citation

*) Figure 7: Using black in the color scale is inconvenient since country borders and coastlines are also black.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-43, 2019.

Printer-friendly version

Discussion paper