

Dear Reviewer,

We would like to thank the reviewer for the precious comments that will help us in improving the quality and readability of our manuscript. We are deeply grateful for that.

In general, We will revise the whole manuscript so to improve its readability. In particular, the issues raised by the reviewer will be addressed. The additional work we planned onto the manuscript will include:

- The quantification of the effective resolution for the precipitation field and its comparison with seNorge2
- A step-by-step graphical representation of the precipitation spatial interpolation algorithm
- The quantification of the improvements due to the introduction of land area fraction in the analysis, as compared to seNorge2 (which does not use land area fraction)

The answers to the comments follow. The reviewer's comments are reported in blue. For brevity's sake, missing answers mean that we will adjust the text as suggested by the reviewer.

Main concerns

*) The story line is sometimes hard to follow and the readability would improve with some more illustrations. For instance: a brief introduction to optimal interpolation could be helpful on page 5. Another example are the scales you introduce on page 9 (line25-27). 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?

We will extend the description of OI. As for the scales, we will include a dedicated Figure.

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. An 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.

In the precipitation plot, stations are located in the middle of beige areas. We will draw the points marking 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 of how this is achieved, but deals with something quite different. It is until 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?

Preliminary notes:

1. At an arbitrary location, CV-analysis and background are -by construction- independent of the observed value. On the other hand, the observed value is used in the calculation of the analysis.
2. OI provides the best (i.e. minimum analysis error variance) linear unbiased estimate of the unknown true value of a quantity. The analysis is the expected value of such an estimate. In practice:
 - a. In data-dense areas: the analysis at an arbitrary location is influenced by several nearby observations, such that the analysis is always a bit different from the observation measured at that location.
 - b. In data-sparse areas: the analysis at an arbitrary location is much more influenced by the associated observation than by other observations, as a consequence the analysis tends to stay closer to that observation.

Points 1-2 explain why the CV-analysis and background performances increase with the increasing of the station density, while the analysis performance decreases (or stay constant) with the increase in station density.

As the reviewer pointed out, the TN analysis behaves quite differently from TG and TX analyses. Our explanation is that the procedure used for the construction of the pseudo-background provides less satisfactory results for TN than for TG and TX. For TN, on average the background differs from the observations more than for TG and TX. The OI realizes a trade-off between observations and background and a worse background causes inevitably a more uncertain OI estimate. However, in data-dense

regions the observations “cooperate” to pull the analyses towards the observations. As a consequence, TN analyses score better for data-dense than for data-sparse regions. In conclusion, we think that the current pseudo-background is more suited for TG and TX than for TN. We will emphasize this conclusion in the manuscript. It is worth remarking that despite the increased uncertainty, the TN fields do provide valuable information for the computation of e.g. climate indices.

*) 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.

The quantification of the effective resolution of meteorological gridded fields is a fascinating and challenging problem. We will apply a scale-decomposition approach (based on 2D wavelet) to the daily precipitation fields of seNorge_2018 and seNorge2. Then, we will study the differences in terms of aggregated statistics for each scale. At the moment, we don't plan to analyze the effective resolution of temperature fields.

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.

The norwegian cities, even the largest ones, are surrounded by forests, they include large green areas, are usually orographically complex and they are invariably close to the sea. Such a heterogeneity of land uses and the alternation of land and water possibly makes the accurate assessment of UHI effects more difficult than elsewhere. Our guess is that a much denser station network would be needed to properly represent the UHI over Norwegian cities. Ideally, at least one observation point for each significant change in land use should be present and this is not always the case for our network. However, our network is dense enough to assess the impact of cities and study the differences between urban areas and forests at a regional level.

*) 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)

No, we haven't looked into that. In the first place, the successful application of a statistical model based on several geographical parameter for precipitation (on a daily time scale) implicitly assumes the local availability of several stations (e.g, one for each slope/aspect/elevation classes in a valley) and this is often not the case for the Norwegian network, as for many other networks in mountainous regions. An interesting analysis on this point has been made by Masson and Frei (2014), where they pointed out that "Our results confirm that the consideration of topography effects is important for spatial interpolation of precipitation in high-mountain regions. But a single predictor may be sufficient and taking appropriate account of the spatial autocorrelation (by kriging) can be more effective than the development of elaborate predictor sets within a regression model."

Masson, D. and Frei, C.: Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types, *Hydrol. Earth Syst. Sci.*, 18, 4543-4563, <https://doi.org/10.5194/hess-18-4543-2014>, 2014.

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

We have implemented a brute force approach that will be described in the text. It is not that different from the simple method described by the reviewer.

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

We will include the motivations in the text. Ideally, one may estimate a pseudo-background field centered on each gridpoint of the 1km grid. To speed up the elaboration the background is computed on a 50x50 grid, then "downscaled" onto the 1km grid. In our case, the average distance between nearby points used as centroids for the background calculation (i.e., centers of each grid cell, 50x50 grid) is 27.5 km (averaged width and height, see page 8 line 18), which is less than the average distance between a station and its 3-4 closest ones. This way, we are sure to represent in the final background field all the relevant features observed by the network.

*) 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.

We will introduce a graphical step-by-step description of the interpolation method for precipitation so to make the whole discussion about the spatial scale less abstract.

*) 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?

As an alternative to the installation of new stations, it would be possible to modify the procedure used to compute the background field. The problem with TN seems to be present both on the regional scale and on the local scale. A better background will improve at least the representation at the regional scale. The use of numerical model output fields (e.g., reanalysis) could also allow us to improve the large scale TN.

*) 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.

The reviewer remark is correct. We need to show that improvement. We will include in the text the verification based on cross-validation and considering coastal stations only.

*) 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.

Fig. 7, TN DJF. The graph in the box (% of large errors vs IDI) shows that TN in data dense region is less likely to provide large errors than in data sparse regions. The color scale in the TN map is the same as for TG and TN so to show the larger uncertainties associated with TN.

*) 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 prediction errors. A way of making the validation

procedure less spatially dependent (and less computationally expensive) could be to split the data point into k-equal area folds.

Our description of the CV procedure used for precipitation is incomplete, since for precipitation we designed the procedure used for the random selection of stations such that the chosen stations are not too close to each other (i.e., approximately equidistant). We will explain it better in the text.

In general, it is exactly to avoid underestimating the prediction errors that we have introduced the discussion on station density and analysis errors, that bring us to Figs. 2-3-5-7. Prediction errors (or analysis uncertainties) do vary over the domain, mostly depending on station density, and this aspect is taken into account in our study.

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

At a specific location (and with the intention to use it as a surrogate for a gridpoint): a value of CV-IDI=0 means that the observations do not have any influence at all on the CVanalysis; CV-IDI=1 means that the observations will strongly condition the CVanalysis, no matter the background value. The numerical boundaries are somewhat arbitrarily chosen: with CV-IDI values > 0.85 we are sure that the observations will strongly influence the CVanalysis; CV-IDI <0.45 means that the observations will play a minor role in the CVanalysis, if compared to the background. The two intermediate classes have been chosen so to ensure having some observations in each class.

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

We will modify the text.

*) 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 page11, line 28, which suggests that almost all differences are between -2 and +1.

Yes, correct. We will modify the text accordingly.

*) 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 ?

G and S are related to the i -th gridpoint because for each gridpoint a different horizontal decorrelation length is used.

*) 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%?

ETS is normally indicated as a number between $-\frac{1}{3}$ and 1 (<https://www.wmo.int/pages/prog/arep/wwrp/new/jwgfvr.html>).

*) 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.