Review of 'EMO-5: A high-resolution multi-variable gridded meteorological data set for Europe' by Vera Thiemig et al.

This is a second review of this paper following the open public review period. The authors have addressed some, but not all, comments provided by myself and the other reviewer. The work undertaken to address these review comments has undoubtedly served to improve the paper and increase potential uptake of the product by users. There remain a number of points that I believe require to be addressed before this can be published to enable the analysis to be understandable to the ESSD readership and to enable reproducibility of the results.

Dear reviewer,

thank you for your continuous dedication to improving our manuscript. As in the previous round of review, we have read attentively each and every of your comment and addressed each with care and consideration, so that we can move forward with the full publication of this manuscript.

With best regards,

Vera Thiemig

Major comment 1:

Lines 89-96 have now arguably gone too far the other way. You probably want to say something like: 'Users should be aware that EMO-5 is prepared principally for near real-time rather than climatological applications. While the series are available from 1990 users intending to apply the dataset for climatological applications should take care in its application and consider, in addition, the use of other products to ensure the robustness of their analysis to the choice of dataset. For the station database aspects of the product users may also consider the C3S holdings (Noone et al., 2021) who undertake an expanded suite of delayed mode quality checks and with whom we are actively collaborating regarding the sharing of data sources where the data licensing permits to improve both products. For gridded data products users may consider the E-OBS product and various flavours of global and regional reanalysis served via the C3S climate data store in addition to EMO-5 to assure themselves of the quality of the various products and the robustness of their analyses.' This or similar text would appropriately caveat without ruling out the potential use of EMO-5.

Reply by authors:

We have adopted the first suggested sentence into the paragraph to stress right from the beginning that EMO-5 is not targeted towards climatological applications, and used the rest of the paragraph for what EMO-5 can be considered.

As this is a scientific publication on a data set which is clearly stated to not aim at climatological applications, we do not see reason to refer to other data sets in this paragraph. Other existing datasets, among which E-OBS, are credited in a more prominent location three paragraphs earlier (ll 68-80).

Changes in manuscript:

before:
Based on the information provided in this paper, the scientific community will understand that EMO-5 (version 1 as well as the operational grids) is not a climatological dataset (unlike e.g. the Essential Climate variables produced by C3S, see Noone et al., 2021), but an operational data set based on the maximum amount of quality-controlled information available at any given time. The implications of this are, that we do not advice to use this dataset to determine climate variability and climate change, however any other environmental application, especially those with real-time, high spatial resolution or multi-variable needs is likely to benefit from this data set. Hence, by making the EMO-5 (version 1) data publicly available, we aim to support many other environmental applications and services that would benefit from using those data, such as e.g. hydrological, agricultural or other environmental applications.

now:
Users should be aware that EMO-5 is prepared principally for near real-time rather than climatological applications. EMO-5 (version 1 as well as the operational grids) is an operational data set based on the maximum amount of quality-controlled information available at any given time. Environmental applications, especially those with real-time, high spatial resolution or multi-variable needs are likely to benefit from this data set. Hence, by making the EMO-5 (version 1) data publicly available, we aim to support many other environmental applications and services that would benefit from using those data, such as e.g. hydrological, agricultural or other environmental applications.

Major comment 2:
I leave this to the editor to determine but my view remains that the list of providers table should be in the main text and not the appendix and that the text introducing them in lines 104-106 should be somewhat expanded.

Reply by authors:

Table A1 contains the listing of data providers that have contributed with their data to EMO-5, hence a very relevant information. Due to its substantial length (>1 page) we found it more appropriate to put it in the Appendix, where it is available to each reader as all the other tables of this manuscript, however, without interrupting the core text.

At this point, we agree for the editor to decide.

Major comment 3:
The methodology to my view remains insufficient to enable reproducibility of results. I note that the authors were unclear what I meant here. It is not that the steps aren't present but rather that too many of the steps are described to a perfunctory level and without settings / parameter values given and that this would preclude a reasonable effort at independent replication. For example, the enumerated list line 156-161 solely hints at what was done for some aspects which the table does not cover. E.g. what are the monthly statistics check? Some kind of

climatological check? If so what period of climatology and what are the thresholds? Note that I am using this as an example only. I would urge the authors to carefully reread the methodology and consider whether each and every step is described in sufficient detail that a reader might be able to reasonably approximate their method based upon the description given alone. Too often I am left feeling that there is grossly insufficient detail to enable a reasonable attempt at being able to reconstruct the method. Given that reproducibility is the central tenet of the scientific process I would urge the authors to carefully redraft their methodology providing sufficient details as to approaches and specific parameter choices to enable an independent replication.

Reply by authors:

We have gone through the description of the methodology and added information at numerous places, such as to the availability check, the monthly check and the interpolation methods (including parameterisation).

Changes in manuscript:

before:

A data value is flagged as "missing" if the value is not available, and as "suspect" if the time stamp has been corrected (Rule 1). It is also flagged as "suspect" if it fails the validation against the monthly statistics (rule 2) or the cross-validation (rule 3). A data value is flagged as "rejected" if it falls outside of the defined minimum / maximum range (rule 4) or if the threshold for the maximum rate of change between two values has been exceeded (rule 5).
[…]
The quality of each of the interpolation schemes was derived through a leave-one-out cross-validation. This means that for each iteration of the interpolated field, one station was left out and then later on compared with its interpolated value. This was done for around 4000 randomly chosen stations evenly distributed over high and low density station areas, and those pairs of interpolated and real observations were used to compute the uncertainty estimates. A similar approach was applied by Hofstra et al. (2008) for the E-CA&D data set.
[…] Originally, it was developed for Kriging schemes, but was adapted to the utilized modified SPHEREMAP scheme. Briefly, the method uses the interpolation weights to calculate a weighted variance between the gridded value and the input station data. It is zero if all input data are identical (e.g. areas with zero precipitation) and increases with increasing variance of the input data.

Table 3: Overview of the three spatial interpolation schemes that were evaluated for the purposes of EMO-5.

#	Interpolation scheme	Description
1.	**Inverse Distance Weighting (IDW)**	IDW is a simple and robust scheme with low computational cost. It is a purely geometric scheme based on the assumption that the closer the meteorological station is to the grid cell centre, the more related it is to its actual value. Mathematically, this is expressed by assigning weights to the surrounding stations proportional to their distance d, e.g. $1/d^2$.

2.	**Modified SPHEREMAP**	The original SPHEREMAP (Willmott et al., 1985) is the adaptation to spherical coordinates of Shepard's inverse distance weighting (Shepard, 1968), which is an extension to the IDW scheme described above. Interpolation weights decrease with increasing distance between grid centre and meteorological station, as in IDW, but the equation for the calculation of the interpolation weight depends also on the distance and number of available input stations. Furthermore, the interpolation takes the clustering of stations into account, so the weights of clustered stations were reduced in order not to overweight these data. As the original SPHEREMAP scheme would lead to a neglect of many stations in regions with a high station density we adapted the algorithm. Previously, if at least one station was found within the smallest search radius "epsilon", then this station or the mean of the stations within "epsilon" was utilised and the station outside the "epsilon" neglected. Now, the "epsilon" is set as 1/20 of the initial search radius and the distance from stations within the radius is set to "epsilon" to avoid an overweighting of the nearest station(s). With these modifications the utilisation of at least four stations per grid point is assured. The maximum number of stations used for interpolation is set to 10.
3.	**Ordinary Kriging**	Ordinary Kriging (Krige, 1966) is an advanced geostatistical method based on correlations between observations. The interpolation weights, which are created by means of the variograms, make use of observation data. Briefly, variograms sort the variance between observations by the distance between these observations were taken. Several approaches can be used to compute these variograms, such as calculations of variograms for each station and time separately, or climate zone dependent variograms, but here utilizing one global variogram for all interpolations, as is utilized at the Global Precipitation Climatology Centre / GPCC (Schamm et al., 2014).

now:

A data value is flagged as “missing” if the value is not available, and as “suspect” if the time stamp has been corrected (Rule 1). Observations are expected at time stamps according to WMO regulations (Manual on Codes, 2013), Offsets up to 29 minutes are corrected. Some countries do not report 6hourly, 12hourly or daily precipitations totals at 00UTC, 06UTC, 12UTC and 18UTC, but at 03UTC, 09UTC, 15UTC and 21UTC. Those data are shifted by three hours forward in time to match the expected reporting times. The error due to the time shift is smaller than disaggregate the data to equally distributed hourly totals, which are accumulated to 6hourly and daily totals afterwards. It is also flagged as “suspect” if it fails the validation against the monthly statistics (rule 2) or the cross-validation (rule 3). For the monthly statistics, the mean monthly minimum and maximum for each station is calculated with the available data and updated with an annual cycle. A data value is flagged as “rejected” if it falls outside of the defined minimum / maximum range (rule 4) or if the threshold for the maximum rate of change between two values has been exceeded (rule 5).
[…]
The quality of each of the interpolation schemes was derived through a leave-one-out cross-validation. This means that for each iteration of the interpolated field, one station was left out and then later on compared with its interpolated value. This was done for around 4000 randomly chosen stations evenly distributed over high and low density station areas, and those pairs of interpolated and real observations were used to compute the uncertainty estimates. A similar approach was applied by Hofstra et al. (2008) for the E-CA&D data set. The three tested interpolation schemes used the same setting, as far as possible. These identical settings were: at least four and at maximum ten stations were used to compute the grid cell value and no restriction in the distance to find the nearest four neighbouring stations. The initial search radius to find the nearest stations was calculated as described in Shepard, 1968. Further scheme dependent settings are mentioned in Table 3.
[…] Originally, it was developed for Kriging schemes, but was adapted to the utilized modified SPHEREMAP scheme. Briefly, the method uses the interpolation weights to

calculate a weighted variance between the gridded value and the input station data. It is zero if all input data are identical (e.g. areas with zero precipitation) and increases with increasing variance of the input data. This method do not need any additional information besides grid value, station values and the interpolation weights.

Table 3: Overview of the three spatial interpolation schemes that were evaluated for the purposes of EMO-5.

#	Interpolation scheme	Description	Parameterisation
1.	**Inverse Distance Weighting (IDW)**	IDW is a simple and robust scheme with low computational cost. It is a purely geometric scheme based on the assumption that the closer the meteorological station is to the grid cell centre, the more related it is to its actual value. Mathematically, this is expressed by assigning weights to the surrounding stations proportional to their distance d, e.g. $1/d^2$.	• Minimum Number of stations: 4 • Maximum Number of stations: 10 • No restriction in search radius • Weight function $\sim 1/d^2$ • Initial search radius as in Shepard, 1968
2.	**Modified SPHEREMAP**	The original SPHEREMAP (Willmott et al., 1985) is the adaptation to spherical coordinates of Shepard's inverse distance weighting (Shepard, 1968), which is an extension to the IDW scheme described above. Interpolation weights decrease with increasing distance between grid centre and meteorological station, as in IDW, but the equation for the calculation of the interpolation weight depends also on the distance and number of available input stations. Furthermore, the interpolation takes the clustering of stations into account, so the weights of clustered stations were reduced in order not to overweight these data. As the original SPHEREMAP scheme would lead to a neglect of many stations in regions with a high station density we adapted the algorithm. Previously, if at least one station was found within the smallest search radius "epsilon", then this station or the mean of the stations within "epsilon" was utilised and the station outside the "epsilon" neglected. Now, the "epsilon" is set as 1/20 of the initial search radius and the distance from stations within the radius is set to "epsilon" to avoid an overweighting of the nearest station(s). With these modifications the utilisation of at least four stations per grid point is assured. The maximum number of stations used for interpolation is set to 10.	• Minimum Number of stations: 4 • Maximum Number of stations: 10 • No restriction in search radius • Initial search radius as in Shepard, 1968
3.	**Ordinary Kriging**	Ordinary Kriging (Krige, 1966) is an advanced geostatistical method based on correlations between observations. The interpolation weights, which are created by means of the variograms, make use of observation data. Briefly, variograms sort the variance between observations by the distance between these observations were taken. Several approaches can be used to compute these variograms, such as calculations of variograms for each station and time separately, or climate zone dependent variograms, but here utilizing one global variogram for all interpolations, as is utilized at the Global Precipitation Climatology Centre / GPCC (Schamm et al., 2014).	• Minimum Number of stations: 4 • Maximum Number of stations: 10 • No restriction in search radius • One variogram for whole domain, not season-dependent • Initial search radius as in Shepard, 1968

Major comment 4:

Table 3 could be expanded to provide the precise parameter settings used in your approach. This is presently given solely for Modified SPHEREMAP but presumably both remaining approaches also had to have some of the parameters set to give values. This comment is by way of a further example whereby the method reproducibility is questionable.

Reply by authors:

We have added a column to Table 3, providing the information on the parameterisation. (Please see above our reply to major comment 3)

Major comment 5:
You need to at least briefly describe what the Yamamoto method for uncertainty quantification is and if there were parameters that needed to be given a value the values you chose should be given. Hence there is a need to revisit the paragraph starting line 206, again with a view to method reproducibility.

Reply by authors:

A description of the method can be found in the paragraph starting at line 206. It also contains the reference to the methodology for those with deep reading wishes. This method does not require a parameterisation, which information we have added to the manuscript.

Changes to the manuscript:

There are many methods available to estimate the uncertainty (i.e. reliability) of the gridded values, such as the leave-one-out approach, ensemble creations or the technique developed by Yamamoto (2000). Kriging itself provides an error estimation, but this depends only on the spatial distribution of the applied stations and not on the input data, therefore this error estimation is not applicable here. As the computational time of the grids is highly relevant in order to produce the operational grids as input for emergency management applications, the technique developed by Yamamoto (2000) is used due to its low computational effort. Furthermore, this method takes into account the variability of the surrounding observations, unlike the common Kriging uncertainty that only depends on the variogram and the spatial distribution of stations. This approach was also used, for example, for the E-OBS data set (Haylock et al., 2008). Originally, it was developed for Kriging schemes, but was adapted to the utilized modified SPHEREMAP scheme. Briefly, the method uses the interpolation weights to calculate a weighted variance between the gridded value and the input station data. It is zero if all input data are identical (e.g. areas with zero precipitation) and increases with increasing variance of the input data. This method does not need any additional information besides grid value, station values and the interpolation weights.

Major comment 6:
In the paragraph starting line 251 I think identical stations is perhaps a misleading term. It’s a single station but it has redundant records arising from two or more distinct sources so this isn’t a case of two or more stations but rather two or more copies of the records from a given station. I think it would be better to talk about redundant versions of records from some stations that have been shared multiple times and that you make steps to identify such records and mingle them to produce a single record for any given station. I would suggest a rewrite of this paragraph accordingly so that it is clearer to a reader what is going on here and use redundant records rather than identical stations as the term in particular to be much clearer what is the issue.

Reply by authors:

We agree to this comment and appreciate the suggestion, which we have implemented throughout the manuscript.

Changes in manuscript:

before:
For precipitation, not all stations that fulfil the above criteria are used in the interpolation. This is due to the fact that over time, there was an increasing number of identical stations that were reported by different data providers (e.g. identical stations are often found between the SYNOP as well as national data), albeit sometimes with slightly different values or slightly different coordinates. Not removing those duplicate stations would lead to a multiple counting of the same station during the interpolation, with the result of overweighting of those stations in the grids and less reliable area mean grid-cell values. To correct this, and to assure a gradual change between stations during the interpolation, duplicate stations within a vicinity of 500 metres were identified and merged into one virtual station. The coordinates of the virtual station were taken from the first station of this cluster, while the value was computed as the average of all duplicate stations. This reduced the total number of stations used per grid realisation by an average of 3.4%. Figure 1 shows the number of stations used during the grid creation.

after:
For precipitation, not all station records that fulfil the above criteria are used in the interpolation. This is due to the fact that over time, there was an increasing number of stations that were reported redundantly by different data providers (e.g. SYNOP and national data), albeit sometimes with slightly different values or slightly different coordinates. Hence appearing as multiple station in the database, while in fact they are one station with multiple records. Not removing those duplicate records would lead to a multiple counting of the same station during the interpolation, with the result of overweighting of those stations in the grids and less reliable area mean grid-cell values. To correct this, and to assure a gradual change between stations during the interpolation, redundant records were identified through a vicinity check. Records, i.e. stations within a vicinity of 500 metres to each other were identified and merged into one record, i.e. virtual station. The coordinates of the virtual station were taken from the first station of this cluster, while the value was computed as the average of all duplicate records. This reduced the total number of records used per grid realisation by an average of 3.4%. Figure 1 shows the number of stations used during the grid creation.

Major comment 7:
The data availability is still to me an issue, although the authors are thanked for making some efforts in this direction. Specifically it helps to have specified the file type. The point about meeting journal minimum requirements is noted. But I assume that the authors wish their data to be used by the broadest possible audience and therefore they should aspire to more than simply treating it as a journal tick box exercise. I am missing here details that may really help a reader to have confidence in the data. So, I would retitle section 4 to be "Data availability, versioning and user support". What is there is good but needs augmenting with, for example:
a. At what delay are various products made available and how are users alerted

to e.g. period of record updates
b. What version control exists, if any?
c. What user support functionality exists?
d. Where and how are data issues and notices handled
This section really should be building the confidence in the user that this is a well documented and well maintained database that they can rely on with confidence. At present its not quite there. The section also may benefit from moving later in the order to come just prior to the conclusions.

Reply by authors:

EMO-5 being a Copernicus product, the general Copernicus user support service will be providing or coordinating user support. In addition, there is a contact person outlined in the JRC Data Catalogue under the link provided in the manuscript.
The version control of EMO-5 will be aligned with the EFAS version control. Documentation on the EFAS version control can be accessed publically on the CEMS wiki: https://confluence.ecmwf.int/display/COPSRV/EFAS+versioning+system
In addition the readme file of EMO-5 will be kept updated.

Major comment 8:
As the authors note in their responses to the initial review in section 5 they are characterising the dataset. I would therefore be more comfortable with section 5 if it were titled 'characterisation' or 'product characterisation'. Evaluation has implications – at least to a native English speaker - that inferences are being made about the correctness or verity of the product. As noted in my initial review the very nature of the problem precludes such an assessment, sadly. Similarly I would change the opening paragraph of this section accordingly.

Reply by authors:

Having consulted multiple native speakers (from UK, Ireland and US) on this issue, we do come to the conclusion that we do not agree on the proposed term "characterisation".
We do evaluate our product from multiple angles through analysis and comparison, and therefore we have been assured by our native speakers that "Evaluation" is indeed the correct term and therefore used as section title.

Major comment 9:
The text on lines 527-529 would need revision to account for the comment above.

Reply by authors:

Please see reply to major comment 8 above.

Major comment 10:
Figures 1-3 remain an issue for me in that too many of the details are simply impossible to discern and the use of multiple different symbols is really hard to untangle. Much of the key text which might help to disentangle and understand the figures is so small as to be indecipherable without zooming in. Considerable efforts

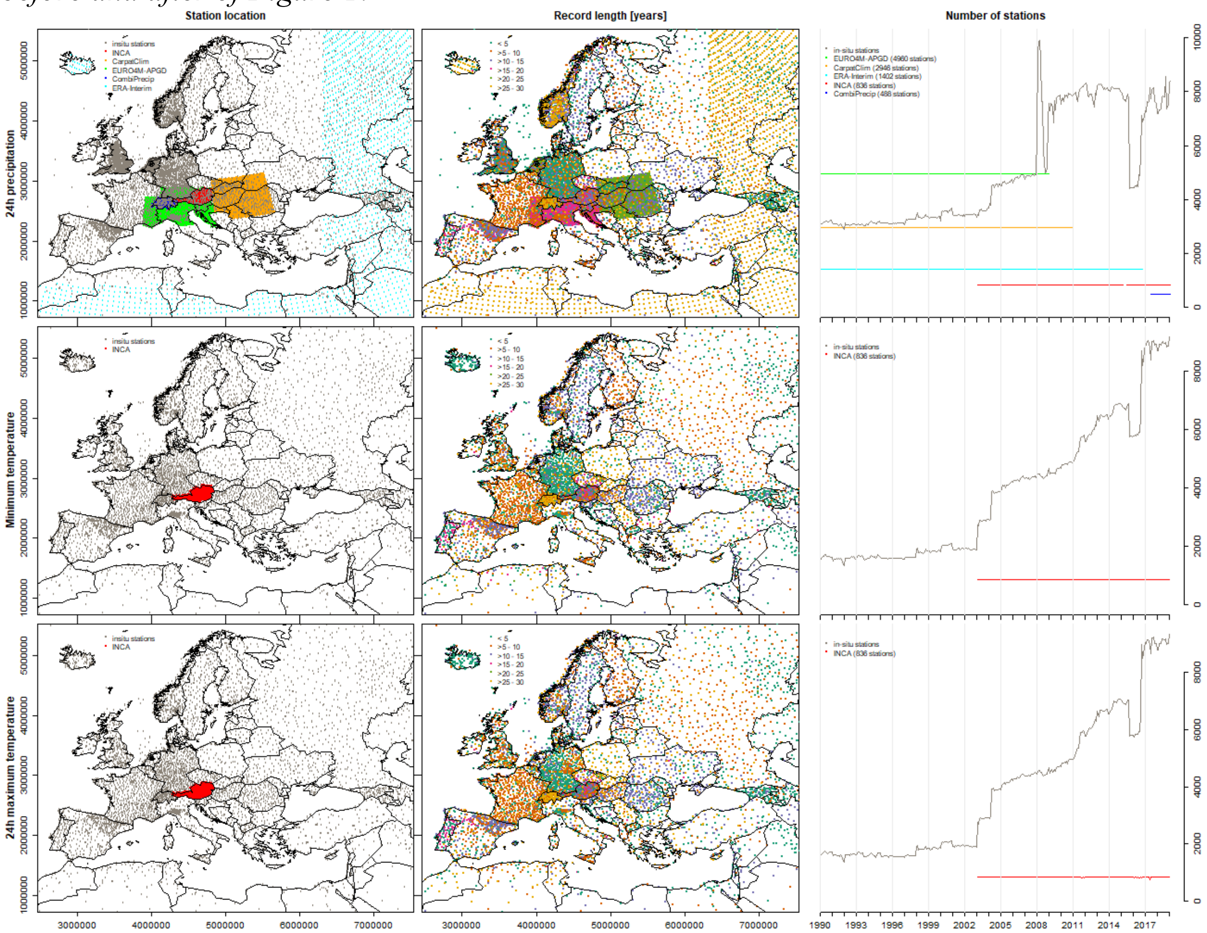
are required to make figures 1-3 more user friendly and, in particular, please ensure all text and numbers are readable at the intended final figure size as readers who print it off shall not have the luxury of being able to zoom.

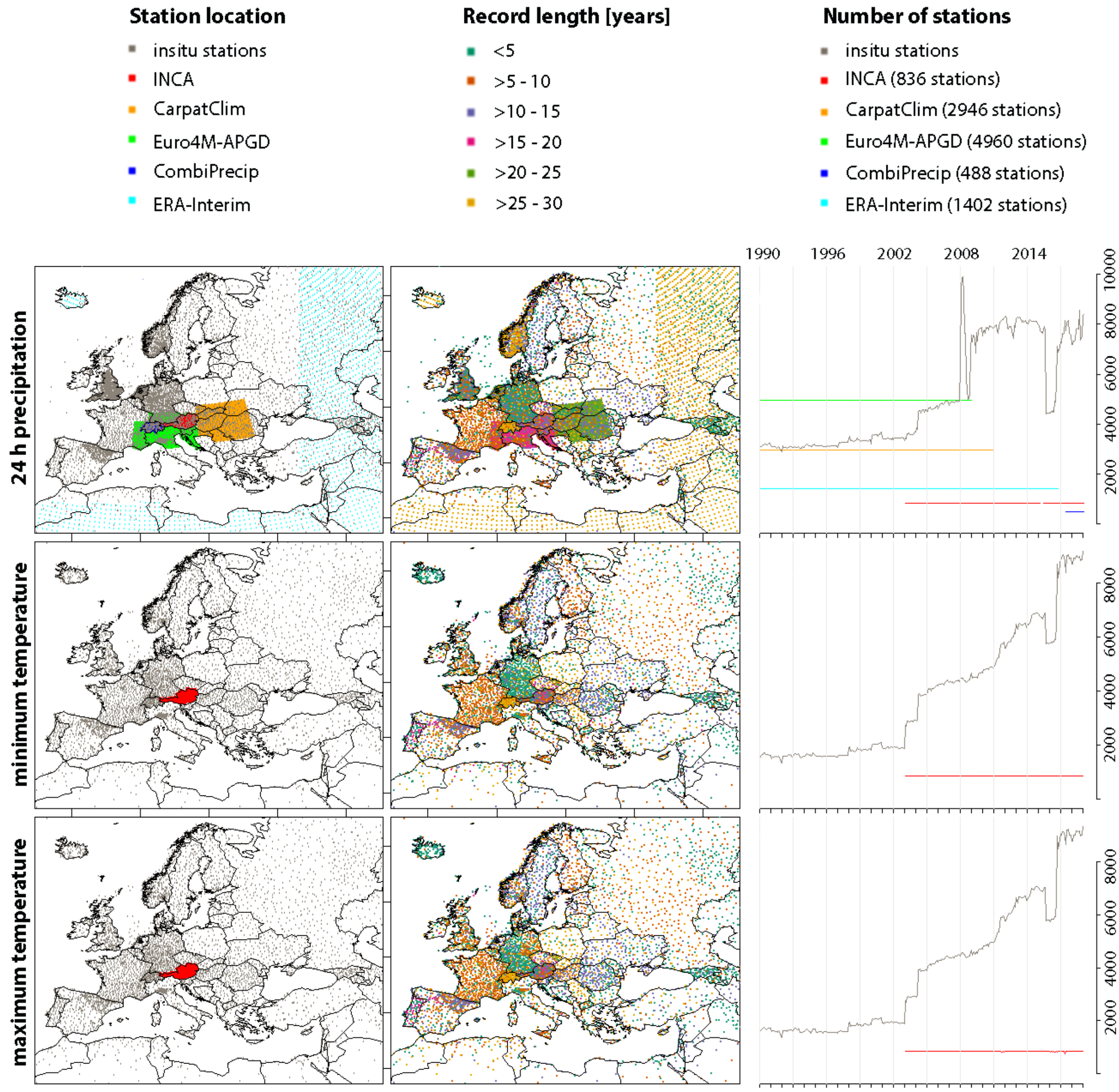
<u>Reply by authors:</u>

Justified comment. The Figures have been updated. Originally the figures have been prepared for online publication with 72 ppt and large zoom potential. We have changed the resolution to 300 ppt, which is the standard for printing and set the image width to a maximum of 16 cm, so that it will not get further compressed once implemented into the manuscript. Also all the labels have been enlarged. The original images will still be provided in the supplement, so that people reading the publication online (we presume that are the majority of readers) will still benefit from the high-res images we originally prepared.

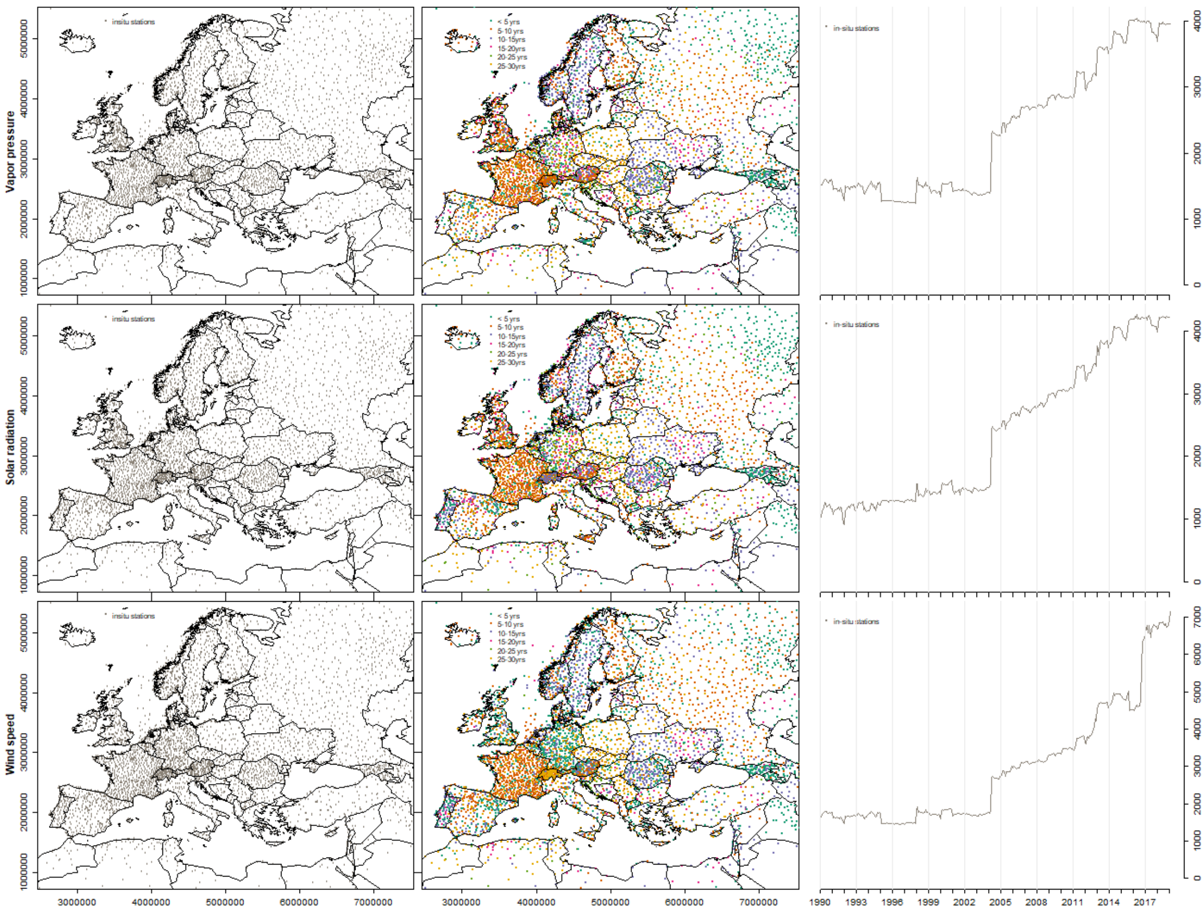
<u>Changes in manuscript:</u>

before and after of Figure 1:

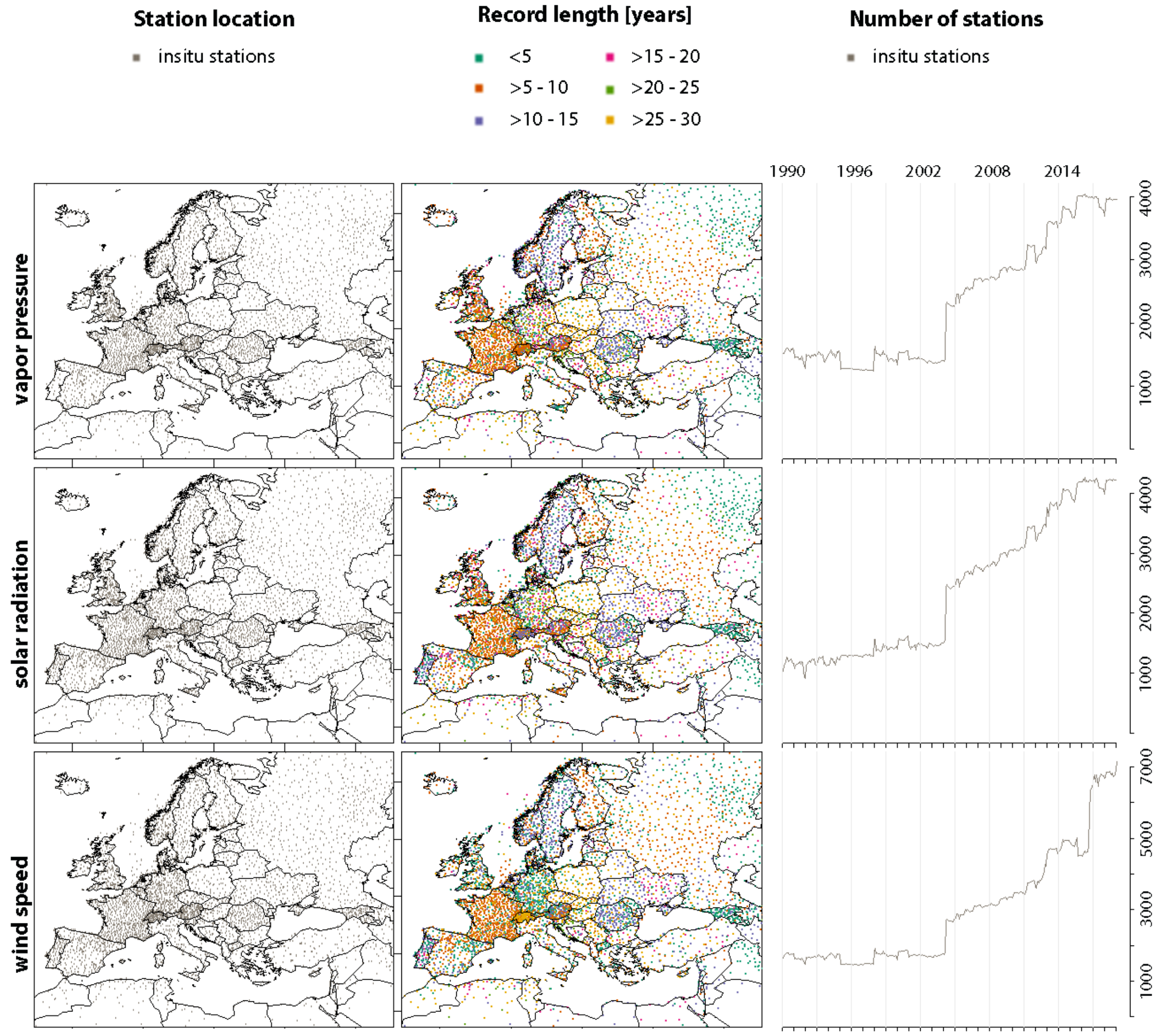




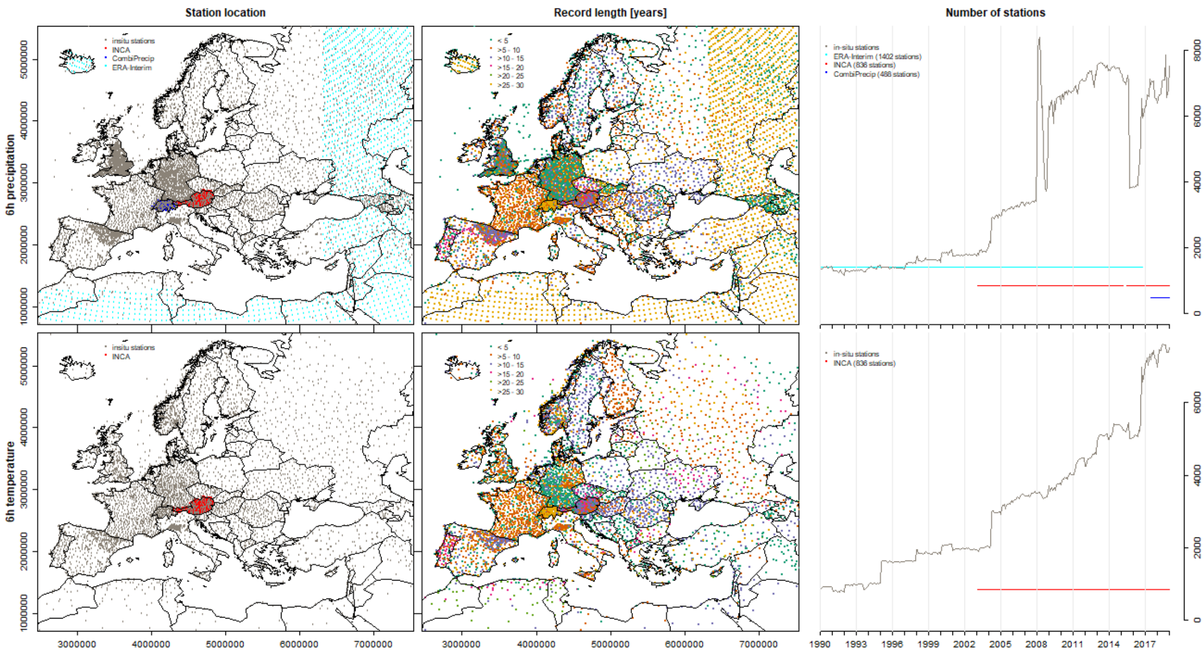
before and after of Figure 2:

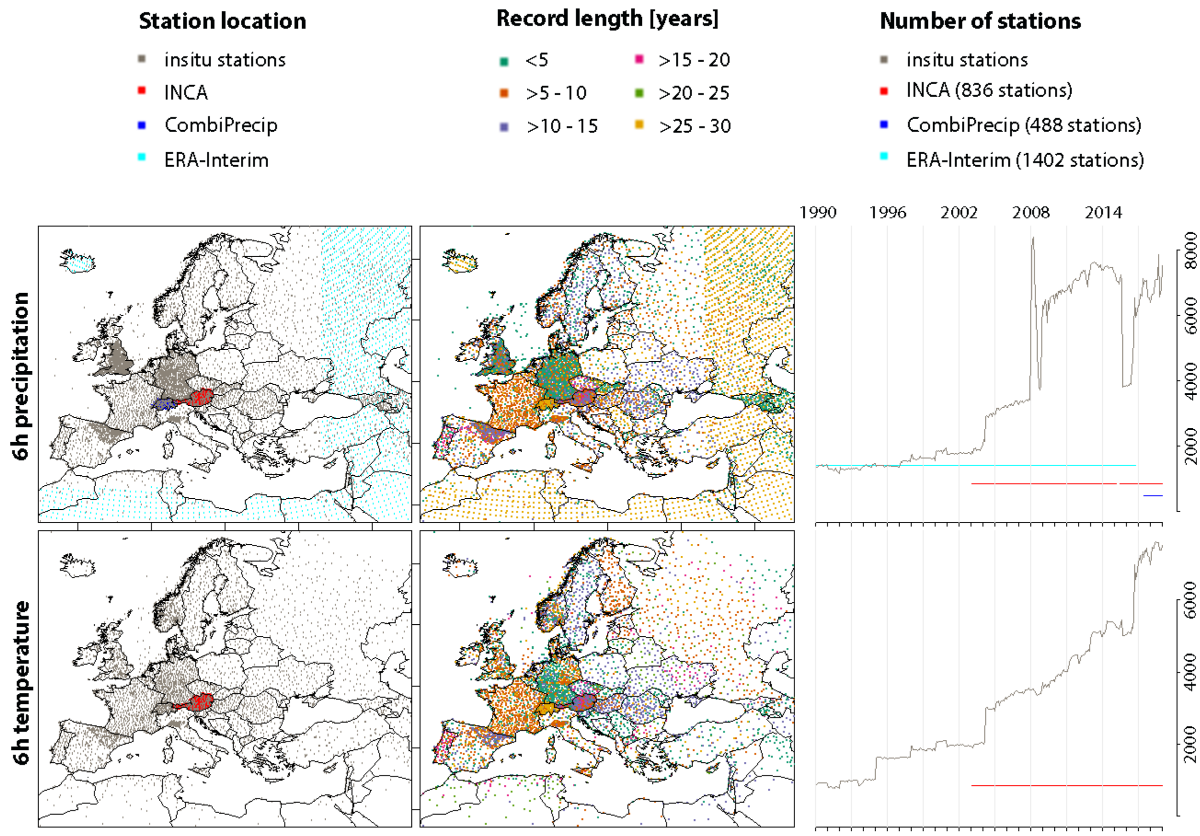


Vapor pressure
Solar radiation
Wind speed
insitu stations
< 5 yrs
5-10 yrs
10-15yrs
15-20yrs
20-25 yrs
25-30yrs
in-situ stations
3000000
4000000
5000000
6000000
7000000
1000000
2000000
3000000
4000000
5000000
1990
1993
1996
1999
2002
2005
2008
2011
2014
2017
0
1000
2000
3000
4000
5000
6000
7000



before and after of Figure 3:





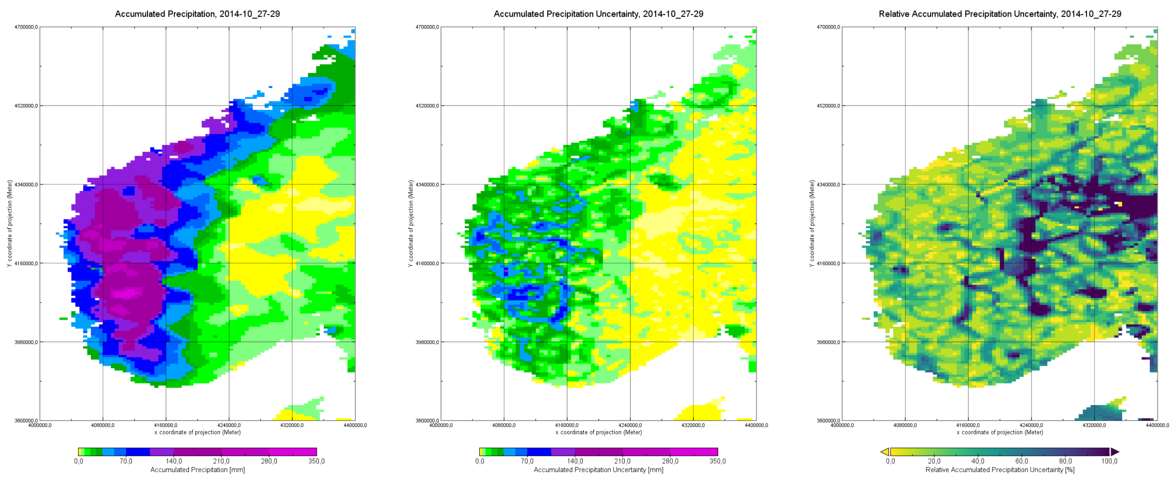
Major comment 11:
Figures 8 and 9 please make the font sizes in these figures larger so they can be read. The keys are impossible to read even scaled to 200% resolution so would be entirely indecipherable for a reader of a printed copy.

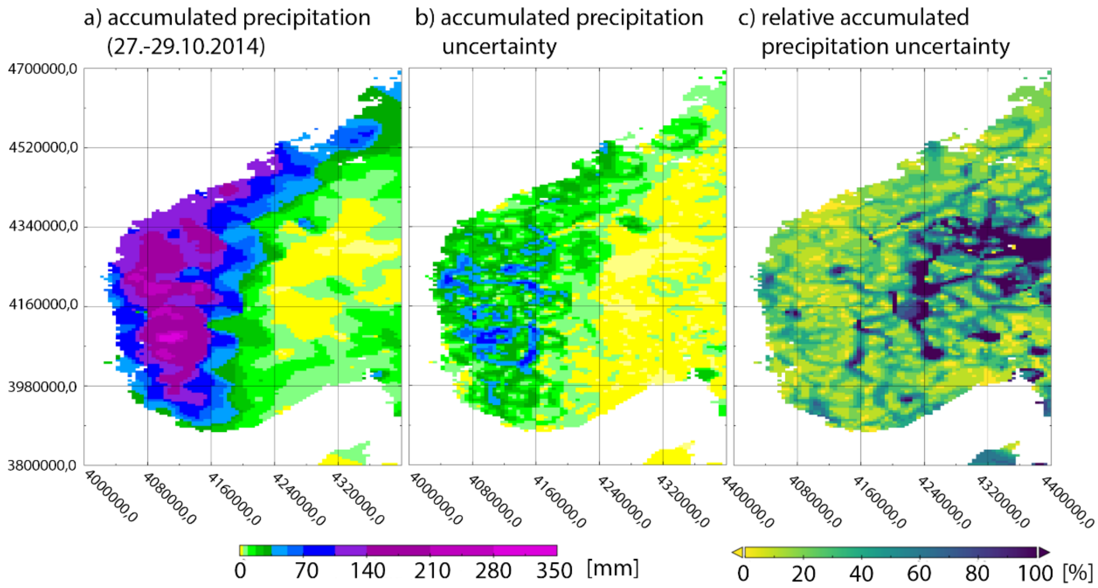
<u>Reply by authors:</u>

We do agree on this point, and have resolved it in the same manner as major comment 10.

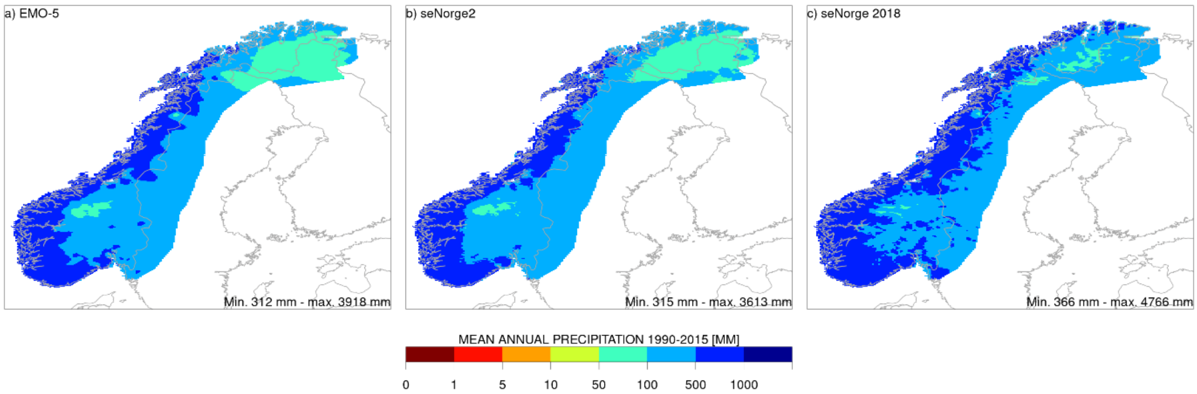
<u>Changes in manuscript:</u>

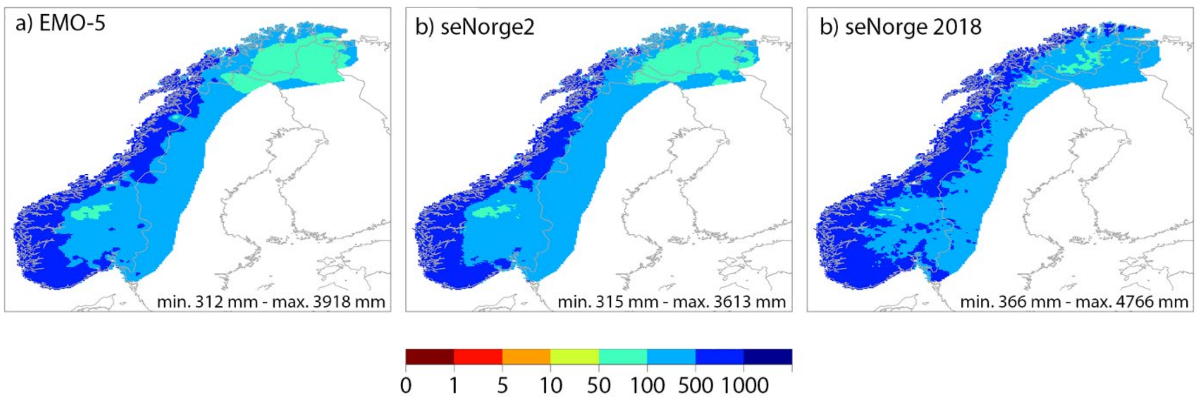
before and after of Figure 8:





before and after of Figure 9:





Major comment 12:
Figure 12 is somewhat improved but still to me very hard to decipher. I find the colour scheme non-intuitive. I'd expect heavier precipitation amounts to be blue not

brown. The colour scheme is also not colour blind friendly

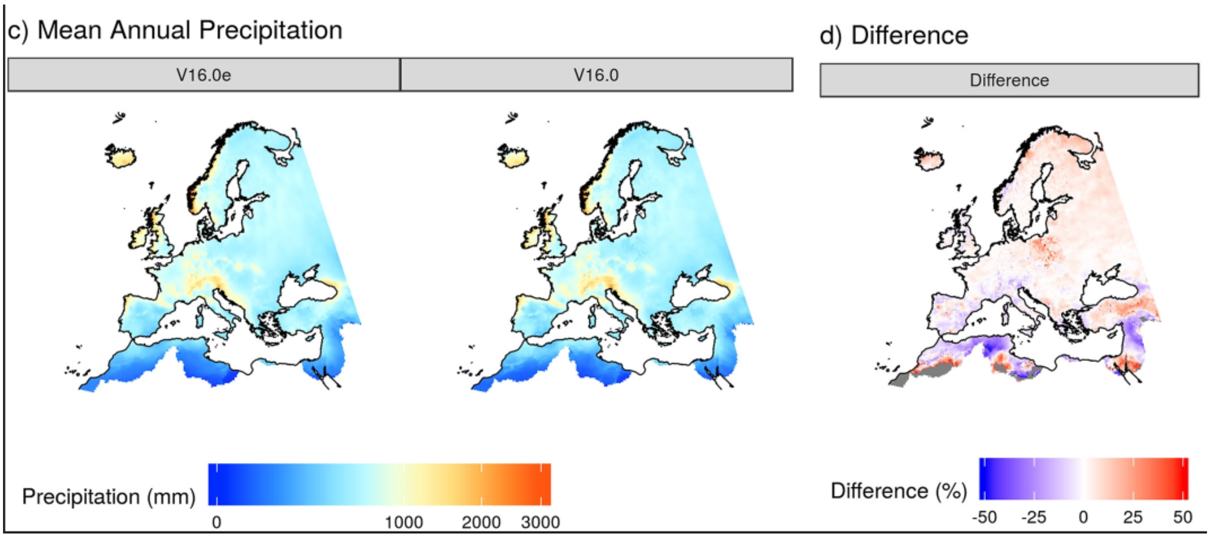
Reply by authors:

The colour scheme used in Figure 12 is a standard colour scheme of ArcGIS, that has also been used within the PESETA projects without any problems during publication.

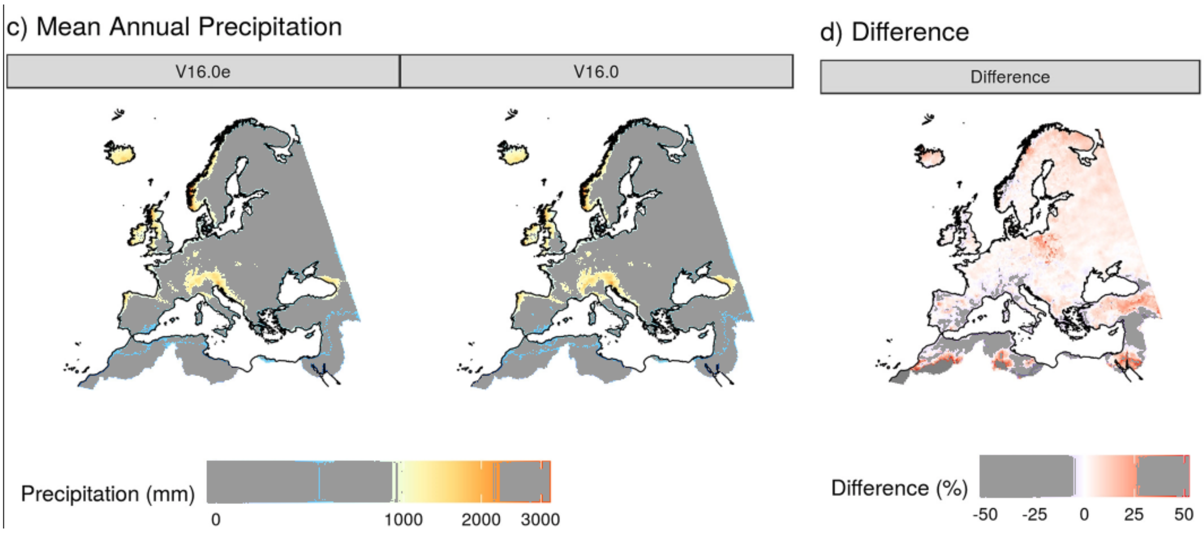
Concerning the comment that the colour scheme used might not be optimal for people with colour vision deficiency seems true (we did a colour blind test). This is however a common problem in almost all publications. Below an example of a similar figure of E-OBS, one can see that tailoring to people with colour vision deficiency is not yet a common practice, as all the areas in grey below would not be differentiable for colour vision deficiency. Nevertheless, they have all been published.

Considering that this is not a requirement of this journal (and neither of others), we have not had it in our mind while preparing the manuscript, and due to the workload it would require at this point we restrain from doing so for this time, but we will have it in mind for future work and publications. At the same time, we believe that this is a problem that needs to be addressed at journal level, as it is not solved by individual authors tailoring to the needs for people with colour vision deficiency.

Colour scheme used by E-OBS (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2017JD028200):



In grey highlighted the problematic areas for people with colour vision deficiency:



Major comment 13:
Following on from point 12 this colour blind issue actually pertains to all figures. None of the colour schema chosen for figures are colour-blind friendly. A substantial proportion of the global population are colour blind. Several colour blind palettes exist see e.g. https://colorbrewer2.org/ including sequential schema that the authors could choose from. For example figure 12 could use https://colorbrewer2.org/#type=sequential&scheme=BuPu&n=9 with the lightest hues pertaining to the lightest precipitation. This schema would be visible and interpretable to all variations of colour-blindness.

Reply by authors:

See reply to major comment 12.

Minor comment 1:
The opening paragraph of the introduction feels like it is missing important context. What aspects of the data quality? Is it their absolute quality? Assurance of their quality? Something else? What do you mean by environmental and risk indicators? Perhaps give an example?

Reply by authors:
We have re-written the opening of the introduction:

Many environmental models rely heavily on the availability of meteorological data. Factors like the accessibility, quality, spatio-temporal coverage as well as spatio-temporal resolution of those meteorological data influence and ultimately determine their modelling capacity. This is further intensified for environmental applications that are running operationally and require quality-controlled, multi-variable meteorological data in near-real time. One prominent example […]

Minor comment 2:
Line 78 I would suggest ‘substantial’ rather than ‘long’. 1990 is not long in the grand

scheme of things from a climatological perspective.

Reply by authors:
Done.

Minor comment 3:
There should be a line break after line 88 assuming this is intended to be a new paragraph?
Reply by authors:
Done.

Minor comment 4:
Line 111 I think you need to say […]others provided data only for […]

Reply by authors:
Done.

Minor comment 5:
Line 184 ECA&D

Reply by authors:
Done.

Minor comment 6:
Line 199 I am unclear what you mean by 'given that the resulting grid quality of all variables allows this' – it makes no sense to me, at least in the context in which it is given. Please clarify.

Reply by authors:
Resolved. We removed ",given that the resulting grid quality of all variables allows this"

Minor comment 7:
Line 246 please be specific which variables or is it all remaining variables in which case say all remaining variables. Also, does this mean that the gridding only considers stations with some minimum set of observed variables and how does this impact station counts etc from sources that have not all variables? Again, this lack of detailed description is precluding reproducibility (see major comments)

Reply by authors:
Yes, this applies to all remaining variables. We changed the sentence to "data coverage for all remaining variables" to clarify this.
It is clearly written that only stations that fulfil this data coverage criterion are used during gridding. Figure 1-3 (right column) shows the correct count of stations used during gridding for each variable.

Minor comment 8:

Line 281 – please specify which land sea mask is used to enable reproducibility

Reply by authors:
Our own in-house land-sea mask has been used. With regards to the reproducibility, the method can be reproduced with any land-sea mask. It is clear that not a 100% copy of EMO-5 can be reproduced as this would mean that the raw source data would need to be shared, which the data licence agreement does not permit.

Minor comment 9:
Line 290 – which DEM is used? Again, you need to specify to enable reproducibility

Reply by authors:
Reference was added.

Arnal, L., Asp, S.-S., Baugh, C., de Roo, A., Disperati, J., Dottori, F., Garcia, R., Garcia-Padilla, M., Gelati, E., Gomes, G., Kalas, M., Krzeminski, B., Latini, M., Lorini, V., Mazzetti, C., Mikulickova, M., Muraro, D., Prudhomme, C., Rauthe-Schöch, A., Rehfeldt, K., Salamon, P., Schweim, C., Skoien, J.O., Smith, P., Sprokkereef, E., Thiemig, V., Wetterhall, F., Ziese, M.: EFAS upgrade for the extended model domain – technical documentation, EUR 29323 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-79-92881-9, doi: 10.2760/806324, JRC111610, 2019.

Minor comment 10:
Line 302 – add 'as follows' to the end of this sentence to be clear that the modifications are then described in the next paragraph. Either that or describe what those modifications were here.

Reply by authors:
Implemented as suggested.

Minor comment 11:
Line 315 – presumably the 5.5 has units. What are these? K^2?

Reply by authors:
Implemented as suggested.

Minor comment 12:
Line 359 if a new paragraph should have a line break. Same at line 365

Reply by authors:
Implemented as suggested.

Minor comment 13:
Lines 378-380 the enumeration should match the section ordering that follows

Reply by authors:

Implemented as suggested.

Minor comment 14:
On line 430 'two blue patches' is very colloquial. Can more scientifically robust language be used in redrafting please?

Reply by authors:
Sentence was re-written by our native English speaking colleague to: "However, the presence of two conspicuous blue-coloured areas shows that EMO-5 is characterized by higher precipitation than seNorge2018."

Minor comment 15:
Line 454-455 was there really somewhere in Norway with no precipitation for 1400 consecutive days or is this some aggregate of this statistic over some region? As written this is really unclear and a redraft is required for clarity here.

Reply by authors:
We put with this statement a pointer to a weakness of the data set we promote with this paper. We believe that it is very honest to mention this and we should not hide it. For clarification, we have added that these values are unrealistic.

Change to the manuscript:
[…] This concerns for example the high values of the Consecutive Dry Days (CDD) - up to 1400 days without precipitation - in the EMO-5 data set. These unrealistic values can be limited to a small region in the northeast. […]

Minor comment 16:
Line 463 if a new paragraph should have a line break added

Reply by authors:
Implemented as suggested.

Minor comment 17:
Line 496 substantial rather than larger (I think)

Reply by authors:
Sentence has been modified by our native English-speaking colleague to "For 13 out of the 15 selected events, EMO-5 shows greater precipitation amounts over the event duration […]".

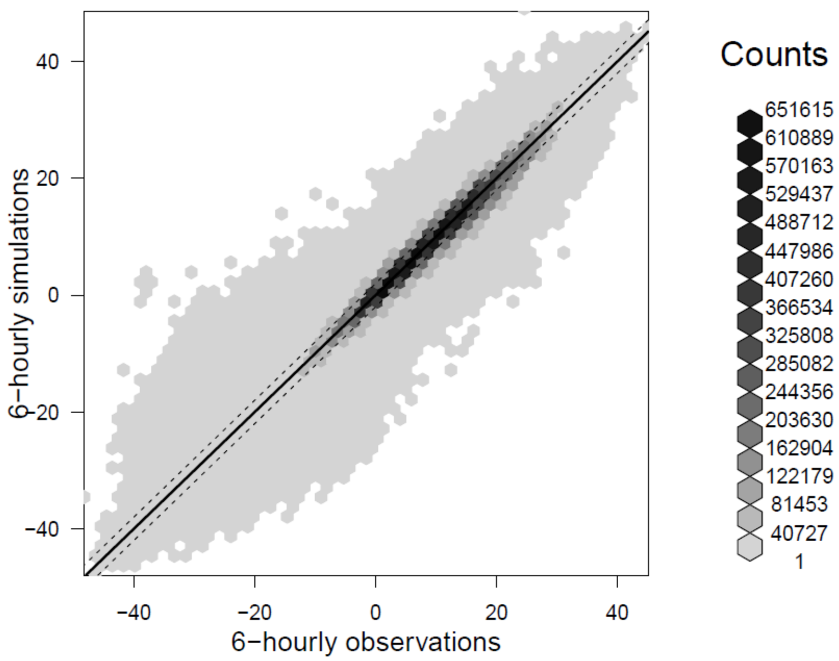
Minor comment 18:
Line 507 station at end of sentence should be stations?

Reply by authors:
Implemented as suggested.

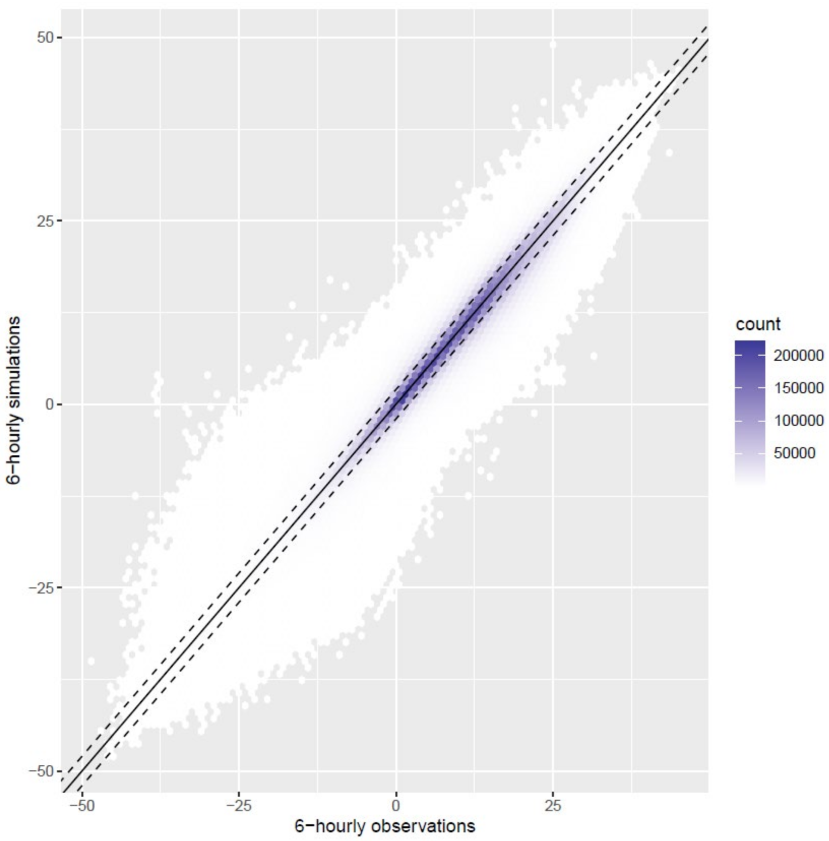
Minor comment 19:
In figure 7 could more sensible bin boundaries be used? It feels really odd to use counts ending in random numbers rather than 0, 40,000, 80,000 etc.

Reply by authors:
It was not possible to define the values of the buckets with the R package we had originally used. However we have investigated on how to improve it to satisfy the review, please see below for the revised visualisation. Please note that we have added the entire aspect only on the wish of the reviewer, we are still not entirely convinced that this aspect shall even be in the manuscript as it is a bit beyond the scope and blurs the paper in our view, but as we are trying to compromise with the reviewer we have added it to the manuscript.

previously:



now:



50
25
0
−25
−50
6−hourly simulations
−50
−25
0
25
6−hourly observations
count
200000
150000
100000
50000