# Norwegian Meteorological Institute

# METreport

# *seNorge* observational gridded datasets

## seNorge_2018, version 20.05

Cristian Lussana

The Norwegian Meteorological Institute, Oslo, Norway

Credit: *Stoneman in Norway* by **Thomas Burschil**
(distributed via imaggeo.egu.eu)

# METreport

| Title | Date |
|---|---|
| seNorge observational gridded dataset. seNorge_2018, version 20.05. | August 5, 2020 |
| **Section** | **Report no.** |
| Division for Climate Services | 07/2020 |
| **Author(s)** | **Classification** |
| Cristian Lussana | ● Free ○ Restricted |

**Abstract**

seNorge_2018 is an observational gridded dataset for daily aggregated temperatures and precipitation data over the Norwegian mainland. The time interval covered spans more than 60 years, from the 1st of January 1957 to the present day. Version 20.05 of the historical archive (1957-2019) is described in this document, with special emphasis on the variations in the methods with respect to the previous version of the archive, which is version 18.12.

**Keywords**

observational gridded datasets, temperature, precipitation, climatology, hydrology, statistical interpolation

---

Disciplinary signature

Hans Olav Hygen

Responsible signature

Cecilie Stenersen

# Contents

# 1 Introduction

seNorge indicates a family of observational gridded datasets over Norway. The Norwegian Meteorological Institute (MET Norway) started to produce the seNorge datasets versions 1.0 and 1.1 for the Norwegian mainland in the nineties (*Tveito and Førland*, 1999; *Tveito et al.*, 2000, 2002, 2005; *Mohr*, 2008). The version 2.0 has been produced after 2015 (*Lussana et al.*, 2018b,a). The most recent member of this family is seNorge_2018 and this document describes version 20.05. In particular, we focus on the modifications of methods and data used with respect to version 18.12, which has been documented by *Lussana et al.* (2019).

The main motivation behind the development of seNorge is to provide gridded datasets of a few near-surface atmospheric key-variables for climatological and hydrological applications. Observational gridded datasets exploit the information provided by networks of traditional weather stations over a region and they return estimates of the selected variables potentially anywhere within the region.

The tag "2018" in seNorge_2018 indicates that the statistical models used for spatial analysis and the code have been developed in 2018. An important component of the spatial analysis procedure is Optimal Interpolation (OI *Gandin and Hardin*, 1965), and the reader may refer to *Uboldi et al.* (2008) for the description of the modified OI scheme used. The versioning of seNorge_2018 indicates the production date of the datasets, such that version 20.05 (or ver. 20.05) refers to May 2020. As stated above, the first version of seNorge_2018 was ver. 18.12. A new version may include both updated data sources and further developments of the methods. Note that the new developments are not considered significant enough to change the name of seNorge_2018, because the Bayesian inference behind the statistical methods remains unchanged.

seNorge_2018 includes gridded fields of daily mean, maximum and minimum temperature, and daily total precipitation amounts. The time period covered by the datasets begins the 1st of January 1957 (1957-01-01) and continues to the present day. The dataset is updated on a daily basis. seNorge_2018 datasets can be divided into two archives: an historical archive and an operational archive. The historical archive includes data from 1957-01-01 to a date in the recent past, which is 2019-12-31 for ver. 20.05. The historical archive is re-built periodically. Then, a new version of the archive is created and new Digital Object Identifiers (DOIs) are assigned to the corresponding datasets. The operational archive is updated every day based on the data of the last few days. In addition, each time

a new historical archive is built, the operational archive is also modified, such that there is a seamless transition between the most recent version of the historical archive and the operational archive. In this sense, the operational archive is a provisional archive. In the case of ver. 20.05, the operational archive starts at 2020-01-01.

The motivations that led us to change parts of the methods used for spatial analysis in ver. 20.05 with respect to ver. 18.12 are different for temperature and precipitation. In the case of precipitations, the gridded fields of a regional climate model are used as input data to scale precipitation in the spatial analysis. Since a climate model covering the same region but spanning a longer time period is available, the gridded fields from this model are used to replace those previously used for ver. 18.12. In the case of temperature, the OI in ver. 18.12 was configured to achieve the best agreement -on average- between (independent) observations and cross-validation analyses. However, the optimization procedure favoured some regions and penalized others, depending on the spatial distributions of the observational network. In particular, since most of the stations used are located at low elevations, the optimization procedure was biased towards ensuring the optimal performances for those regions, while the uncertainties in mountainous regions and in data sparse regions was considered only marginally in the optimization. As a consequence, the variability of ver. 18.12 analyses fields in data sparse regions is rather large, larger than in seNorge2 (or seNorge version 2.0) for instance. The ideal situation would be to rely on the OI performances of seNorge_2018 ver. 18.12 in data dense regions and, at the same time, reduce the uncertainty of the temperature fields in data sparse regions down to the level of seNorge2. seNorge_2018 ver. 20.05 has exactly this objective. The variability of the gridded analysis field in data sparse regions is reduced by blending together the results of several different OIs based on slightly different configurations.

The document is organized as follows. Sec. 2 defines seNorge_2018 variables and presents the data used. Sec. 3 describes the variations in the spatial analysis with respect to the methods implemented for ver. 18.12. Sec. 4 compares the gridded fields of ver. 20.05 with those of ver. 18.12 with the aim of studying the effects of variations in the input data and methods over the results. Sec. 5 summarizes the main messages of the document. The data access is then described in the Appendix. After the Appendix, a section with supplementary material is included, in order to make the reading of the main text more fluent by avoiding the interruptions caused by pages and pages of figures.

# 2 Data and Definitions

seNorge covers the Norwegian mainland and it includes all the catchments of rivers flowing in the Norwegian territory, as a consequence seNorge extends also into parts of Sweden and Finland bordering Norway. The domain of seNorge_2018 ver. 20.05 is the same of ver. 18.12. More details and figures on the domain can be found in previous papers, such as *Lussana et al.* (2018b,a, 2019). The gridded data are presented on a high-resolution terrain-following grid with 1 km spacing. The coordinate reference system is Universal Transverse Mercator, zone 33.

The data sources used are: MET Norway's climatological archive (frost.met.no), which includes also observations from stations managed by several public Norwegian institutions, such as The Norwegian Water Resources and Energy Directorate (NVE), the Norwegian Public Roads Administration (Statens vegvesen, SVV) and the Norwegian Institute of Bioeconomy Research (NIBIO); the Swedish Meteorological and Hydrological Institute (SMHI) climatological archive; Finnish Meteorological Institute (FMI) climatological archive. In addition, the daily datasets provided by the European Climate Assessment & Dataset project (ECA&D, https://www.ecad.eu, *Klein Tank et al.*, 2002) are also used.

The definitions of the variables available in seNorge_2018 are:

- RR at a given date is the total amount of precipitation accumulated from 06 UTC of the day previous to that date to 06 UTC of the day in the date. RR is also referred to as daily precipitation. Note that RR at a station location does not correspond to the value observed by the corresponding ombrometer, not only because of the adjustment made by the spatial analysis but also because observations are adjusted for the wind-induced under-catch according to the relationships reported by *Lussana et al.* (2019) and based on *Wolff et al.* (2015).

- TG at a given date is the mean averaged temperature from 06 UTC of the day previous to that date to 06 UTC of the day in the date. TG is also referred to as daily mean temperature. In detail, it is the arithmetic mean of 24 hourly values or a formula based mean value computed from fewer observations (*Førland and Tveito*, 1997).

- TX at a given date is the maximum temperature from 18 UTC of the day previous to

that date to 18 UTC of the day in the date. TX is also referred to as daily maximum temperature.

- TN at a given date is the minimum temperature from 18 UTC of the day previous to that date to 18 UTC of the day in the date. TX is also referred to as daily minimum temperature.

The 63-year time period covered by the historical archive of seNorge_2018 ver. 20.05 ranges from January 1957 to December 2019. In such a long period of time the observational network has continuously changed. The data availability is shown in Fig. 1. There are more observations for precipitation than for temperature. In the case of temperature, there are no significant differences between TG, TX and TN. The area of seNorge_2018 spatial domain measures 606579 km$^2$. In the idealized case of a completely uniform spatial distribution of the observational network, the ratio (area) / (number of observations) corresponds to the area around each observation where that observation is the closest data source. In this sense, that area can be referred to as a sort of "area of influence" of an observation. The radius of the circle equivalent to the area of influence is a measure of the average distance we need to travel in every direction before meeting another observation. This radius is a rough indicator of the observation spatial density. Suppose the number of available observations is 100, then the radius is equal to 44 km. With numbers of available observations equal to 200, 300, 600 and 900, then we get circle radii of, respectively: 31 km, 25 km, 18 km and 15 km. For all variables, it is possible to identify three different regimes in the time series. From 1957 to 1975, there was an increase in the number of available observations. From 1975 to 2000-2005, there was a gradual, slow decrease in the number of observations. The decrease is rather sharp for RR in the period from 2000 to 2005. From 2005 to 2020, the situation is different between the two quantities. In the case of precipitation, there is an oscillatory trend with an average of approximately 700 observations, while for temperature a sharp increase from 300 to 500 observations is shown.

The in-situ data shown in Fig. 1 has been quality controlled according to MET Norway's quality assurance system specifications. In addition, for temperature, a spatial consistency test such as those described by *Lussana et al.* (2010) has been applied. For precipitation, the time series have been visually inspected to identify anomalies with respect to neighbouring stations.

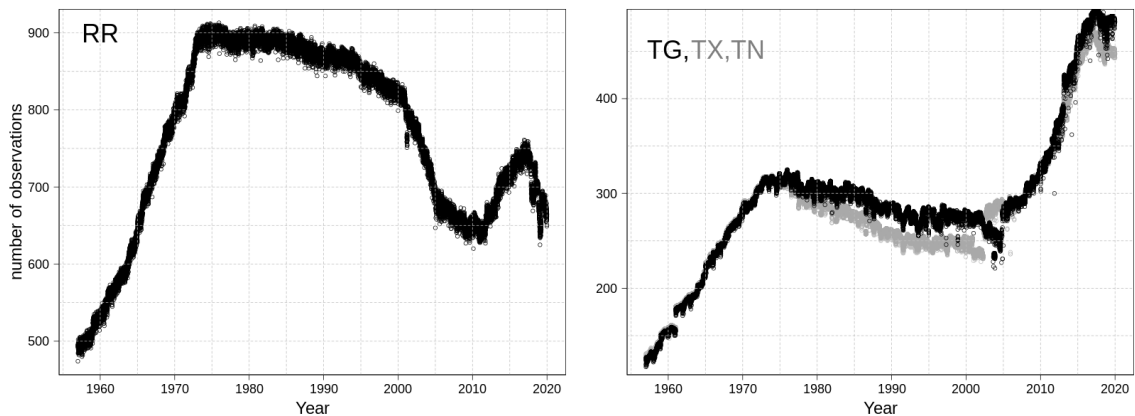By comparing Fig. 1 with Fig. 1 of *Lussana et al.* (2019), it can be noted that on

Figure 1: Time series of number of observations used for spatial analysis that are in the seNorge_2018 domain after the data quality control. The left panel shows RR observations. The right panel shows TG (black dots) and TX, TN (gray dots) observations. Note the different scales on the y-axes between the two panels.

average ver. 20.05 is based on more observations than ver. 18.12. However, similarly to ver. 18.12, the in-situ data used in ver. 20.05 are not evenly distributed across the altitude range. More stations are located in the lowlands, while in the mountains the observational network is quite sparse. This should have an influence on the uncertainties of the spatially interpolated fields of ver. 20.05 similar to what reported for ver. 18.12 and documented by *Lussana et al.* (2019).

# 3 Methods

The steps of the procedure used for the production of daily gridded datasets for an arbitrary day are listed in the Algorithm 1. The mathematical notation and the symbols used are defined in Tab. 1. The objective of this Section is to describe the modifications in the methods of seNorge_2018 ver. 20.05 compared to ver. 18.12, while the general description of the methods documented by *Lussana et al.* (2019) is still the reference for seNorge_2018.

In the Algorithm 1, we made it clear that a number of data are required for the production of seNorge_2018. The fixed datasets used are not reported in the algorithm and those are the digital elevation model and the land area fraction over the 1 km regular grid. The gridded fields of wind speed are derived from numerical model output, in this way it is possible to associate a value of wind speed to each observation, independently of

the actual availability of an observed value of wind speed. In fact, observed wind speed values would have been more accurate data but we would have inevitably lost some of the precious observations of temperature and precipitation. We estimated that data loss would have been a worse solution than the less accuracy for the wind speed. The numerical models used are described in the paper *Lussana et al.* (2019) and they are based on MET Norway's products (*Reistad et al.*, 2011; *Müller et al.*, 2017; *Frogner et al.*, 2019).

## 3.1 Precipitation

seNorge_2018 statistical interpolation of precipitation is based on an iterative OI scheme that uses the relative anomalies between observed and reference precipitation values. The references are long-term monthly averages of precipitation obtained through numerical models. In the case of ver. 18.12, the reference datasets used for precipitation are based on hourly precipitation provided by the climate model version of HARMONIE (version cy38h1.2), a seamless NWP model framework developed and used by several national meteorological services. For ver. 20.05, the reference dataset has changed. The long-term monthly averages of precipitation have been derived by the "NorCP" simulation, which provides high-resolution (3 km) climate model data covering the 21-year time period from 1998-2018 by an updated climate model version of HARMONIE (*Lind et al.*, 2020).

In Fig. 2, the typical annual total precipitation fields for the references of ver. 20.05 and ver. 18.12 are compared. First thing to notice is that the domain of the reference for ver. 20.05 is wider than that of ver. 18.12. As a consequence, the Norwegian mainland in ver. 20.05 reference is less likely to be affected by those border effects that may affect ver. 18.12, especially in Southern Norway. Besides, the wider domain of ver. 20.05 should also allow for a better development of the high-resolution dynamics of the dynamical downscaling procedure. As shown in the right panel, the reference of ver. 20.05 has -on average- larger values of precipitation, though the situation can vary significantly between months and regions, as shown in Figs. 11- 22.

## 3.2 Temperatures

seNorge_2018 statistical interpolations of temperature are all based on a two-step spatial analysis scheme. The first step aims at representing the temperature field at a point as the average over a spatial support larger than the local station density, where this is characterized as the average distance between a station and its closest neighbouring stations
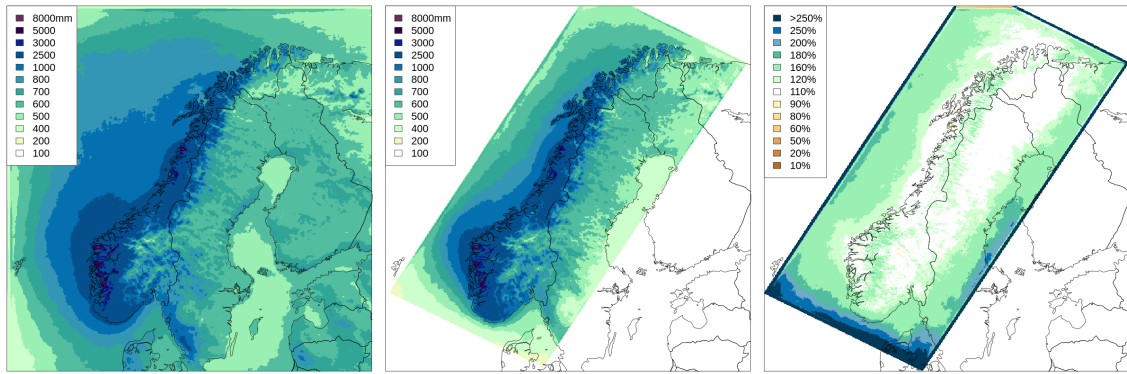
Figure 2: Comparison of typical annual total precipitation fields of the seNorge_2018 reference fields. Left panel: ver. 20.05 reference field. Middle panel: ver. 18.12 reference field. Right panel: relative deviations as percentages of the ver. 18.12 reference field (i.e., ref(ver 20.05) / ref(ver 18.12)).

in the surroundings of that point. The second step refines the temperature representation at a point by including those effects that are observed only by a small number of neighbouring stations. The first step is implemented through the elaboration of sub-regional pseudo-background fields. This part has remained unchanged from ver. 18.12. The second step is implemented by means of an Optimal Interpolation (OI) scheme and this part has been significantly changed, as it is described in the following of this Section.

For each of the three variables TG, TX and TN, the OI runs with more than a single configuration, as it is instead for ver. 18.12. In particular, four different OI are performed. The parameter that is varying within the different OI runs is $D^z$, which is the vertical decorrelation length used in the definition of the background error covariance matrix. For ver. 18.12, $D^z$ was set to 210 m. In the case of ver. 20.05 the four values used are: 210 m; 400 m; 600 m and 800 m. For each combination of the pair variable/OI configuration, a dataset of gridded fields is produced.

There are two more steps that must be performed in order to obtain the final dataset of TG, TX and TN. The first step is the merging of the four datasets into a single dataset for each variable. The second step is the check for consistency between the three variables, specifically to ensure that TX is greater or equal to TG and, at the same time, TN is smaller or equal to TG. Note that such a check is needed, because the spatial analysis is performed for each variable independently from the others, therefore in regions where the analysis uncertainty at a given time is greater than the difference between variables, the consistency between TG, TX and TN may be violated. The check for consistency was

10

implemented already in ver. 18.12 and it first replaces TX with TG for those grid points where TG is greater than TX, then it replaces TN with TG for those grid points where TG is smaller than TX. This simple and practical solution might not always yield satisfactory results, for this reason it is one of the points that should be improved in future versions of seNorge_2018.

The merging of datasets into a single one is a new feature of ver. 20.05 compared to ver. 18.12. Statistical interpolation schemes produce fields where each value is an average over a spatial support and the sizes of those spatial supports vary across the domain. The OI settings influence the sizes of the three-dimensional spatial supports. The vector of OI settings $\lambda_1$ includes $D^z = 210\,\text{m}$ (same as ver. 18.12), then: $\lambda_2$ includes $D^z = 400\,\text{m}$; $\lambda_3$ includes $D^z = 600\,\text{m}$ (same as seNorge2); $\lambda_4$ includes $D^z = 800\,\text{m}$. Apart from $D^z$, all the parameters in the four OI settings do share the same values. The variables considered are TG, TX and TN, they are further abbreviated as $g$, $x$ and $n$, respectively. For example, the notation for the analysis of TG at the $j$th grid point for the OI configuration with $D^z = 800\,\text{m}$ is $\mathbf{x}_j^{g,a}(\lambda_4)$ (see also Tab. 1). In addition to the analysis, the Integral Data Influence (IDI, *Uboldi et al.*, 2008) is available for each point in each field. IDI is an indicator ranging from 0 to 1 that summarizes the amount of information used in the OI that is derived from nearby observations. At the $j$th grid point, if we consider for instance TX, the notation $\mathbf{x}_j^{x,\text{IDI}}(\lambda_2)$ indicates IDI obtained with $D^z = 400\,\text{m}$. Given a particular set of $\lambda$ values, grid points where the IDI values are close to 0 will have the analysis almost identical to the large-scale pseudo-background, since no information from nearby observations has been used in the analysis. Then, the spatial support of the analysis is mostly determined by the pseudo-background characteristics. On the other hand, grid points where the IDI values are close to 1 will use nearby observations to adjust the analysis. In this case, the spatial support of the analysis is mostly determined by the OI settings and the spatial support of the analysis temperature field is smaller for smaller values of $D^z$.

The merging is done such that more weight is given to the field with smaller spatial support (i.e. $\lambda_1$) and weight has been given to the other (coarser) fields only where they provide additional information. In this way, we make use of the observational network in an optimal way by providing a higher resolution field where the network is denser. The relative content of information of a field with respect to another field is evaluated through a function inspired by Shannon's measure of information content, similarly to what reported for the probability density functions in *Tarantola* (2005), Sec. 1.2.5. In

our equations, IDI will replace probability density functions. The fields are ordered from the one with the smallest spatial supports, which is that obtained with $\lambda_1$, to the one with the largest spatial supports, when the OI is performed with $\lambda_4$. Then, relative content of information in one field with respect to its immediate antecedent field with smaller spatial supports is computed. The first field, the one with $\lambda_1$, is assigned the complementary of the relative content of information of the second field. The weights are the normalized relative contents of information. The procedure is described in mathematical terms in the equations that follows.

The TG analysis at the $j$th grid point is obtained as a weighted average of the four OI analysis:

$$\mathbf{x}_j^{g,a} = \sum_{l=1}^{4} \mathbf{w}_j^l \cdot \mathbf{x}_j^{g,a}(\lambda_l) \tag{1}$$

where the weights with $l = 2,\ldots,4$ are obtained as:

$$\mathbf{w}_j^l = \alpha_j^{-1} \cdot \mathbf{x}_j^{g,\mathrm{IDI}}(\lambda_l) \cdot \log\left( \frac{\mathbf{x}_j^{g,\mathrm{IDI}}(\lambda_l)}{\mathbf{x}_j^{g,\mathrm{IDI}}(\lambda_{l-1})} \right) \tag{2}$$

The weight for $l = 1$ is set to:

$$\mathbf{w}^1 = \alpha^{-1} \cdot (1 - \mathbf{w}^2) \tag{3}$$

and $\alpha$ is the vector of normalization factors used to enforce that $\sum_{l=1}^{4} \mathbf{w}_j^l = 1$, $\forall j = 1,\ldots,m$. Note that to avoid problems in Eq. (2), all the IDI values smaller than 0.0000001 have been replaced with 0.0000001.

In Figs. 3- 6 examples of the normalized weights of Eqs. (2)- (3) are shown. The figures show the averaged weights over the days within year 2019 for TG. In the right panels, the fields on the original seNorge_2018 grid are shown. In the left panels, the fields have been aggregated on a coarse resolution grid with square boxes of side length 50 km. In this way, we focus on averages over regions and not on small-scale details. The weight $\mathbf{w}^1$ in Fig. 3 is by far the largest one and for a few boxes on the coarser grid its value is equal to 1. The OI analysis obtained with $\lambda_1$ should have more weight than the others, especially close to an observation. In mountainous regions where the observational network is sparse, $\mathbf{w}^1$ reaches its smallest values. The weight $\mathbf{w}^2$ in Fig. 4 has maximum values in the range 0.25-0.5 exactly in those regions where $\mathbf{w}^1$ has its smallest values. The other two weights, $\mathbf{w}^2$ and $\mathbf{w}^3$, are shown in Figs. 5- 6 and the information of the corresponding OIs are used to integrate the higher-resolutions OI only in mountainous

12

regions where the observational network is sparse. The maximum value of $\mathbf{w}^3$ is 0.25, while the maximum value for $\mathbf{w}^4$ is 0.16.
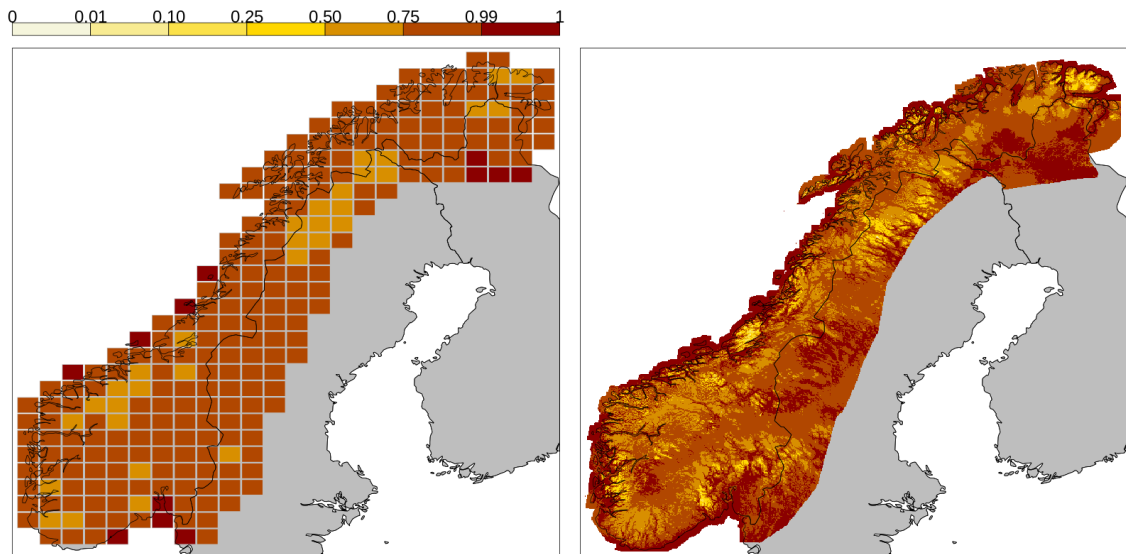


Figure 3: Normalized weights $\mathbf{w}^1$ ($D^z = 210\,\mathrm{m}$) for TG, averaged over 2019. The left panel shows the aggregated field over square boxes with sides of 50 km. The right panel shows the original field on the 1 km grid.
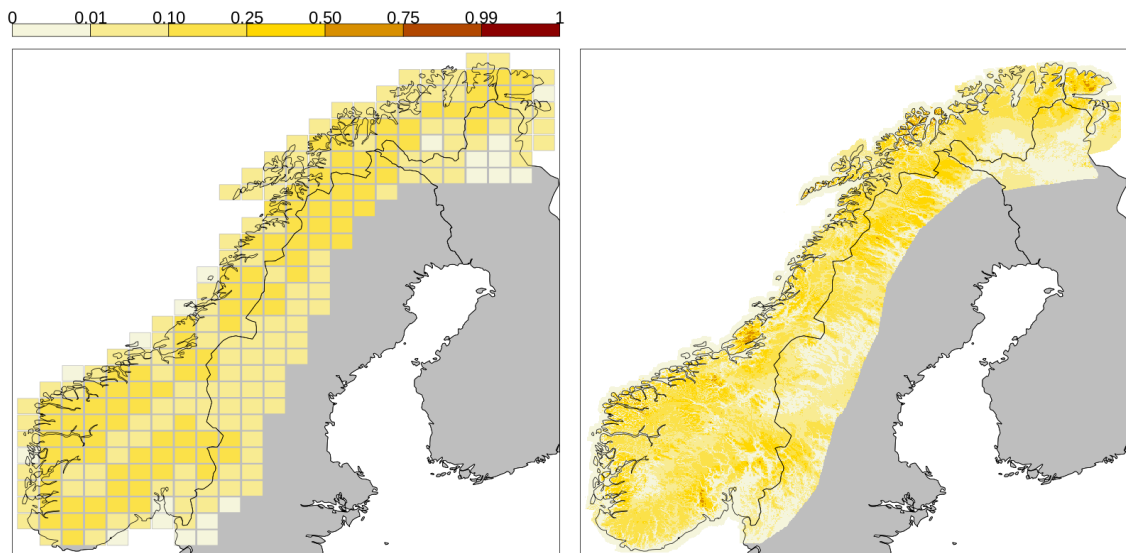


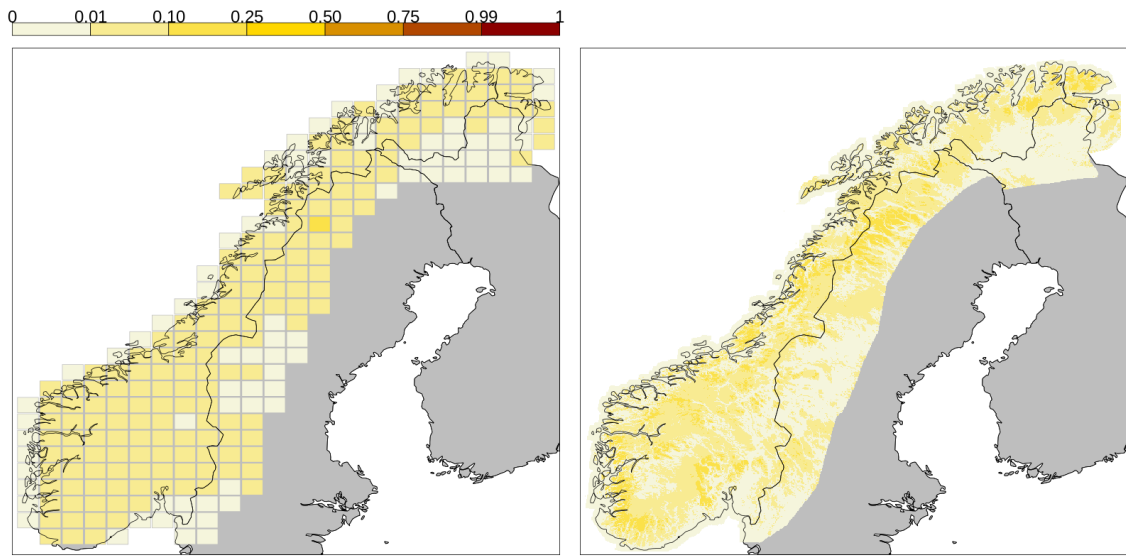Figure 4: Same as Fig. 3 but for $\mathbf{w}^2$ ($D^z = 400\,\mathrm{m}$).

13

Figure 5: Same as Fig. 3 but for $\mathbf{w}^3$ ($D^z = 600\,\text{m}$).
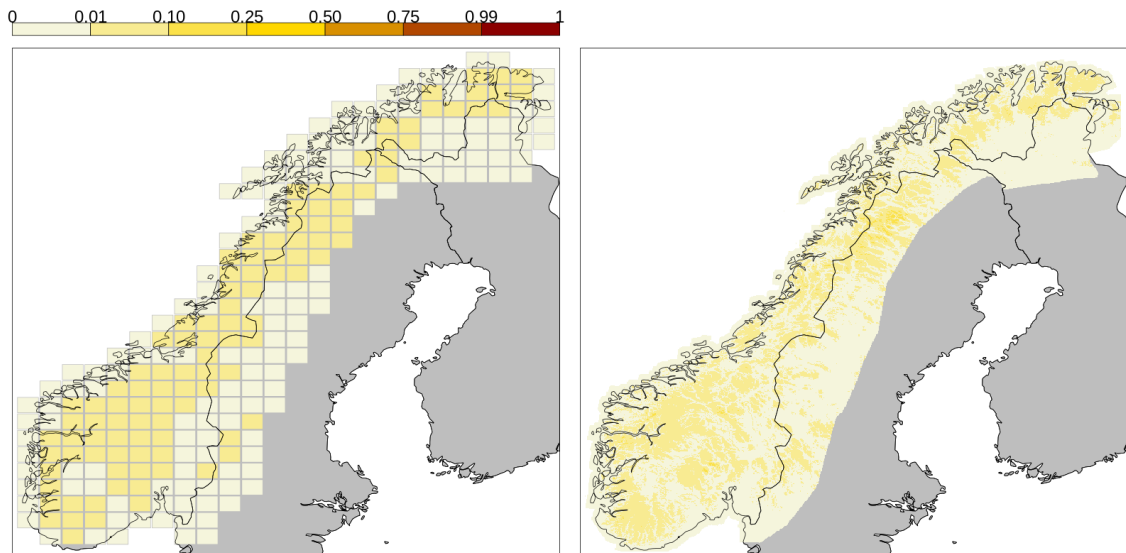


Figure 6: Same as Fig. 3 but for $\mathbf{w}^4$ ($D^z = 800\,\text{m}$).

# 4 Results

A comparison between seNorge_2018 ver. 20.05 and ver. 18.12 is reported in this section for the four variables RR, TG, TX and TN. In particular, the comparison is based on the most recent 30-year period common to both datasets, which is the time period 1988-2017. Note that a complete study to characterize estimation uncertainties based on cross-validation experiments has not been performed yet for ver. 20.05, while it is available in the paper by *Lussana et al.* (2019) for ver. 18.12.

The comparison of total precipitation amounts is shown in Fig. 7 for winter months and in Figs. 23- 25 for the other seasons. The comparison is performed on monthly precipitation totals averaged over the period 1988-2017. The results show the amounts of ver. 20.05 as percentages of the amounts of ver. 18.12. For example, where the monthly total precipitation of ver. 20.05 is approximately equal to the amount of ver. 18-12 ($\pm 5\%$), then the corresponding square box will be filled with the white color. Green/blue colors indicate boxes where ver. 20.05 have larger precipitation amounts than ver. 18.12. Yellow/brown colors indicate boxes where ver. 20.05 have smaller precipitation amounts than ver. 18.12. For the majority of the boxes the two versions are rather similar, that is $\pm 5\%$ or $\pm 10\%$. However, for some boxes in the mountains or along the coast the differences are significant (up to 300% for two boxes in northern Sweden, close to the boundary with Norway). The reasons are the different observational dataset used and the variations in the reference field used for the OI. If compared to the other seasons, the summer months shown in Fig. 24 display a different pattern. Compared to ver. 18.12, ver. 20.05 has less precipitation along the coast and larger precipitation amounts inland.

The comparison of daily mean temperatures is shown in Fig. 8 for winter months and in Figs. 26- 28 for the other seasons. The quantity considered is the monthly mean temperature averaged over the period 1988-2017 and the elaboration is based on daily mean temperatures. Then, the differences between ver. 20.05 and ver. 18.12 are shown in the figures. The only significant differences occur during the winter, while for the other season the regional averages are almost everywhere $\pm 0.1°$C. Fig. 8 shows that during winter: ver. 20.05 is colder than ver. 18.12 on the Scandinavian mountains in the Northern part of Sweden; ver. 20.05 is warmer than ver. 18.12 on the northernmost part of Norway, in the region of East-Finnmark.

The comparison of the daily maximum temperature is shown in Fig. 9 for winter months and Figs. 29- 31 for the other months. The quantity considered is the monthly
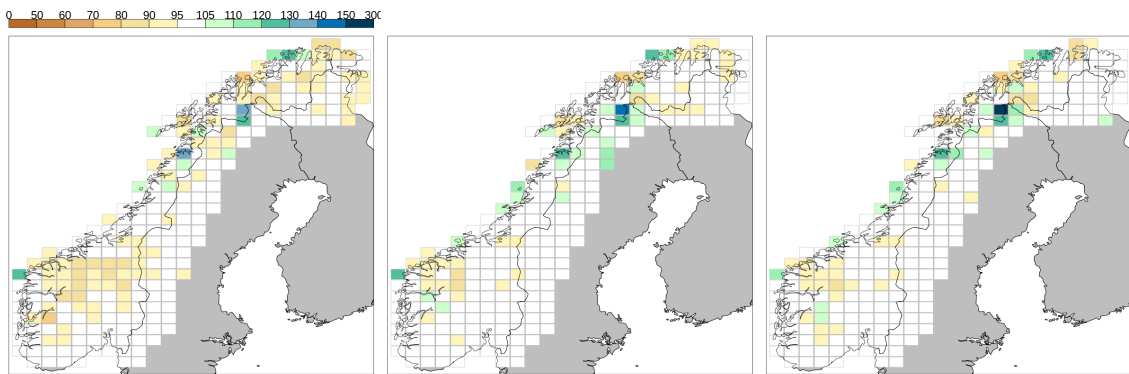
Figure 7: Monthly total precipitation amounts for seNorge_2018 ver. 20.05, as percentage of ver. 18.12, averaged over the 30-year period from 1988 to 2017 for winter months: December, in the left panel; January, in the middle; February, in the right panel. The fields of averaged values over square boxes with sides of 50 km are shown in the panels.
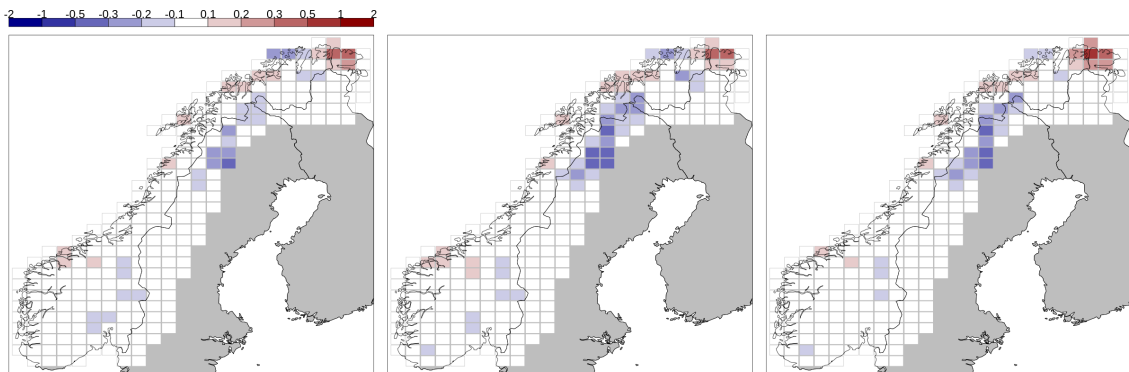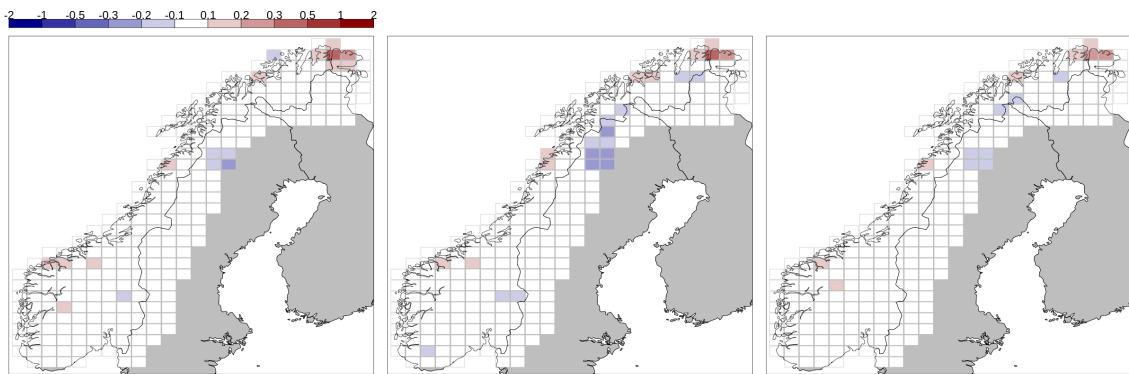


Figure 8: Differences between monthly mean daily temperature of seNorge_2018 ver. 20.05 and ver. 18.12, as °C, averaged over the 30-year period from 1988 to 2017 for winter months: December, in the left panel; January, in the middle; February, in the right panel. The fields of averaged values over square boxes with sides of 50 km are shown in the panels.

mean of daily maximum temperatures. The monthly mean is based on the period 1988-2017. Then, the differences between ver. 20.05 and ver. 18.12 are shown in the figures. In general, the two seNorge_2018 versions are rather similar. The situation is similar to that of daily mean temperature, but the differences between the two versions are even less pronounced.

The comparison of the daily minimum temperature is shown in Fig. 10 for winter months and Figs. 32- 34 for the other months. The quantity considered is the monthly
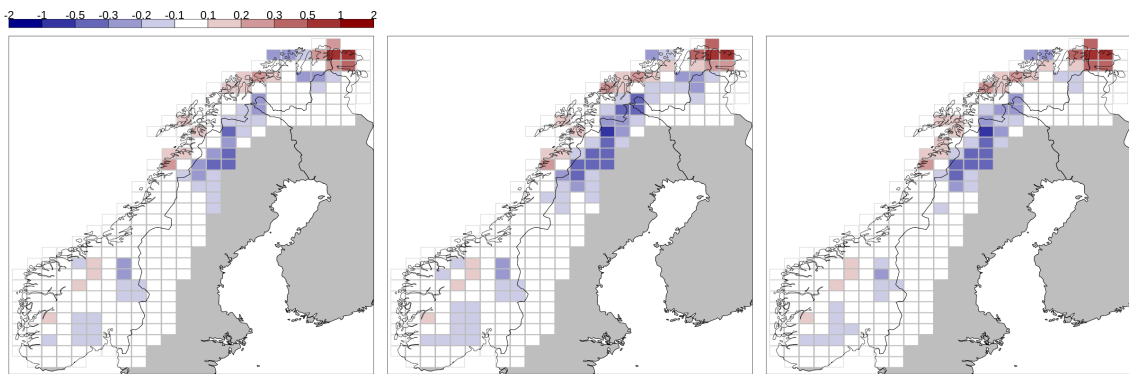
Figure 9: Differences between monthly average of maximum daily temperature of seNorge_2018 ver. 20.05 and ver. 18.12, as °C, averaged over the 30-year period from 1988 to 2017 for winter months: December, in the left panel; January, in the middle; February, in the right panel. The fields of averaged values over square boxes with sides of 50 km are shown in the panels.

mean of daily maximum temperatures. The monthly mean is based on the period 1988-2017. Then, the differences between ver. 20.05 and ver. 18.12 are shown in the figures. As for the other two temperature variables, the largest differences between the two versions are observed for the winter months, even though in the case of minimum temperature there are still significant differences in the northern part of the domain also for September, October and March. The patterns of the deviations resemble also those observed for the daily mean temperature, with a region in northern Norway where ver. 20.05 is warmer (more than 1°C) than ver. 18.12 and a region of the Scandinavian mountains in northern Sweden where ver. 20.05 is colder than ver. 18.12. In addition, the deviations between minimum temperatures over South-Norway in winter are also significant for a few square boxes, up to ±0.3°C.

# 5 Conclusions

seNorge_2018 version 20.05 has been released and it is described in this document. The observational gridded datasets include daily mean, minimum and maximum temperatures and daily total precipitation amounts over the Norwegian mainland for the time period 1957-2019. The gridded fields are made available over a regular grid with 1 km of spacing between grid nodes in both easting and northing directions.

Version 20.05 is an update of ver. 18.12, which covers the same spatial domain but

Figure 10: Differences between monthly average of maximum daily temperature of seNorge_2018 ver. 20.05 and ver. 18.12, as °C, averaged over the 30-year period from 1988 to 2017 for winter months: December, in the left panel; January, in the middle; February, in the right panel. The fields of averaged values over square boxes with sides of 50 km are shown in the panels.

spans a smaller time period, from 1957 to 2017. The data sources for the in-situ observations are the same for the two versions, however the observational datasets for ver. 20.05 have been collected from scratch, that is independently from ver. 18.12. The in-situ observations have been quality controlled based on both visual inspection of the time series and the application of automatic procedures. The set of in-situ observations used for ver. 20.05 is rather similar to the one used for ver. 18.12. A bit more observations have been used in ver. 20.05.

Version 20.05 includes some changes in the methods with respect to ver. 18.12. In the case of precipitation, the reference fields used to scale the daily precipitation amounts in the statistical interpolation have been changed. In the case of temperature, the statistical interpolation has been modified such that the pseudo-background, large-scale, field is adjusted locally on the basis of the surrounding observations over wider regions in ver. 20.05 than in ver. 18.12.

The combined impacts over the results of the updated input observational datasets and the modifications in the methods have been evaluated for the period 1988-2017 by considering monthly averaged deviations between ver. 20.05 and ver. 18.12. Instead of considering grid points, we have aggregated deviations over square boxes with sides of 50 km. In general, the deviations are rather small, especially in data dense regions, and this fact indicates a stability of the methods with respect to variations in some procedures. However, in data sparse regions or where some new observations have been used in ver.

18

20.05, the differences between the two versions are evident even over 30-year averages. For all variables, the largest differences occur during winter and they are more pronounced over few square boxes that are characterized by complex terrain and a sparse observational network. Future plans include a complete evaluation of seNorge_2018 ver. 20.05 based on cross-validation to estimate the improvements in accuracy and precision with respect to ver. 18.12.

# Appendix: Data Access

seNorge_2018 version 20.05 datasets are available for public download under the Norwegian Licence for Open Government Data (NLOD, `https://data.norge.no/nlod/en/`). The data can be accessed from two distinct sources: Zenodo `zenodo.org` and MET's Norway THREDDS data server `thredds.met.no`. A detailed description on how to access the data is available at the seNorge wiki-pages:

- `https://github.com/metno/seNorge_docs/wiki/seNorge_2018`

In the wiki, it is also possible to get the latest news on the datasets and contact the developers by opening issues about specific topics.

The digital object identifiers (DOIs) of the datasets are:

- TG, daily mean temperature (*Lussana*, 2020b). DOI is 10.5281/zenodo.3923706.

- RR, daily total precipitation amount (*Lussana*, 2020a). DOI is 10.5281/zenodo.3923703. In addition to RR, the archive also provides RRa (RR alternative), which is the field of daily total precipitation without the correction for the wind-induced undercatch.

- TX, daily maximum temperature (*Lussana*, 2020d). DOI is 10.5281/zenodo.3923700. In addition to TX, the archive also provides TXa (TX alternative), which is the field of daily maximum temperature without the check for consistency against TG and TN.

- TN, daily minimum temperature (*Lussana*, 2020c). DOI is 10.5281/zenodo.3923697. In addition to TN, the archive also provides TNa (TN alternative), which is the field of daily minimum temperature without the check for consistency against TG and TX.

# Supplementary Material
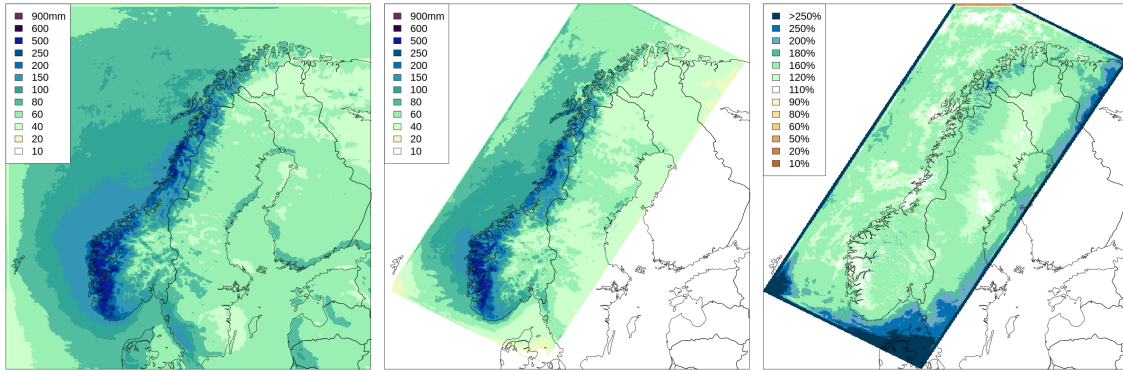


Figure 11: same as Fig. 2 but for the typical January.
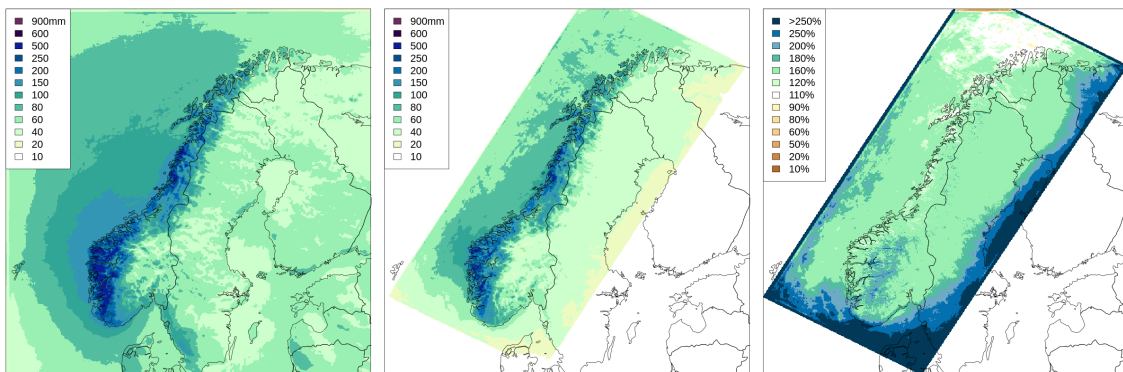


Figure 12: same as Fig. 2 but for the typical February.
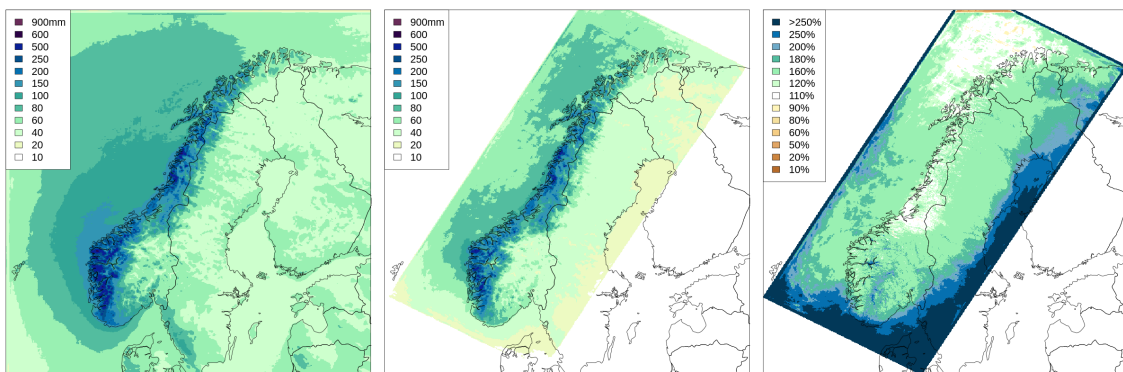


Figure 13: same as Fig. 2 but for the typical March.
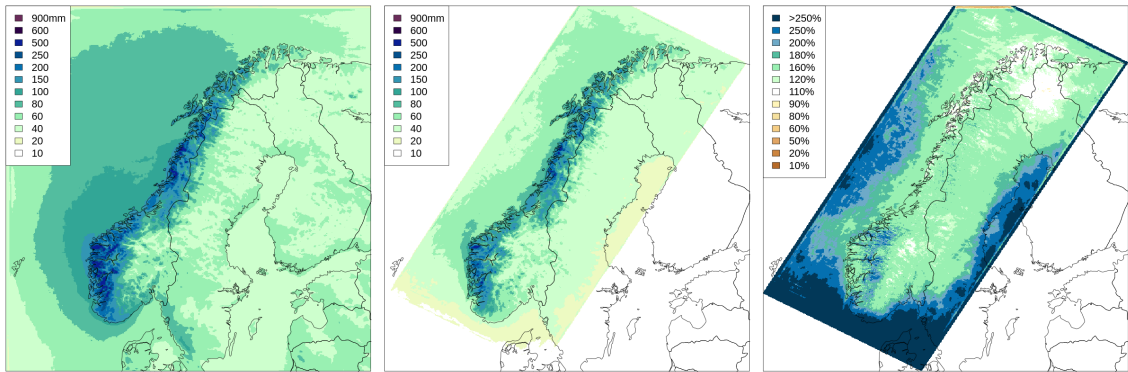
Figure 14: same as Fig. 2 but for the typical April.
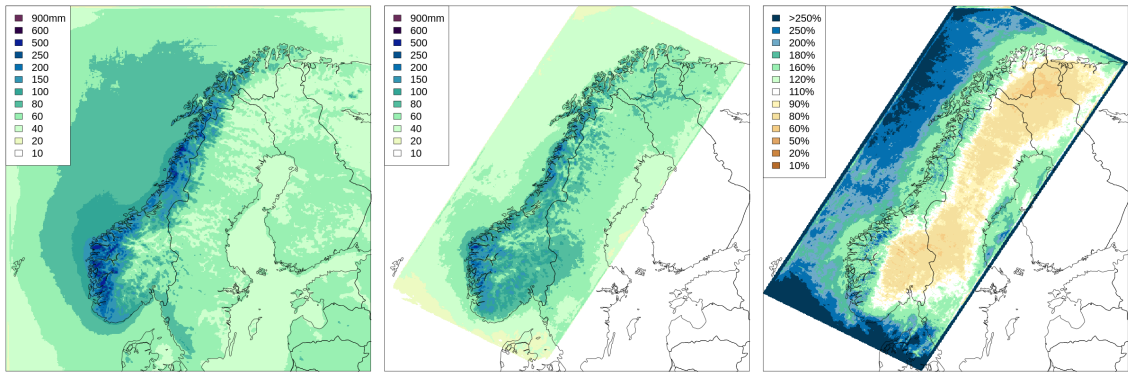
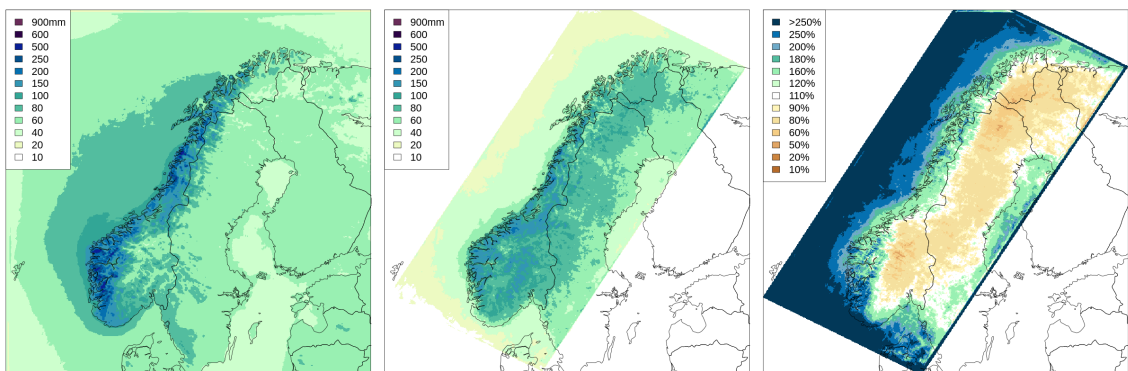

Figure 15: same as Fig. 2 but for the typical May.
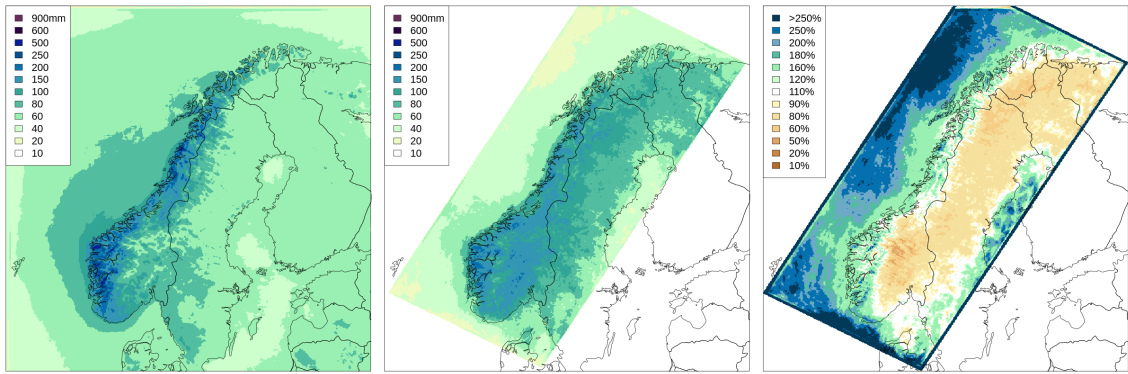


Figure 16: same as Fig. 2 but for the typical June.

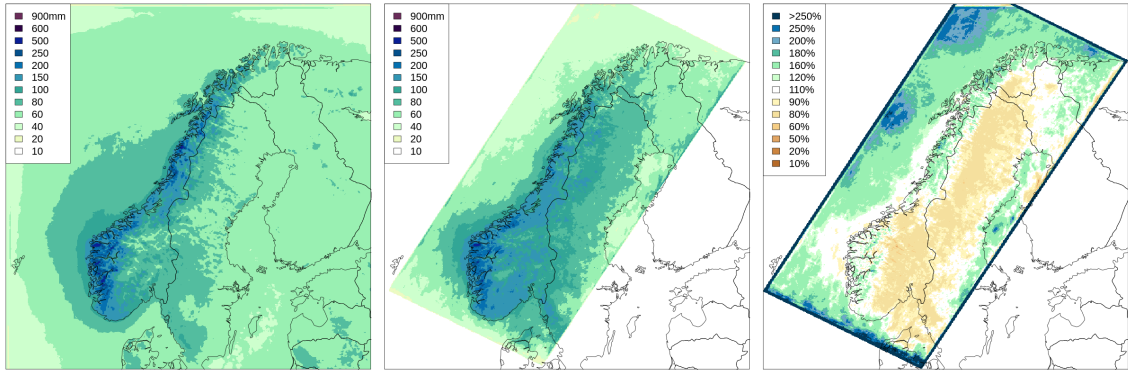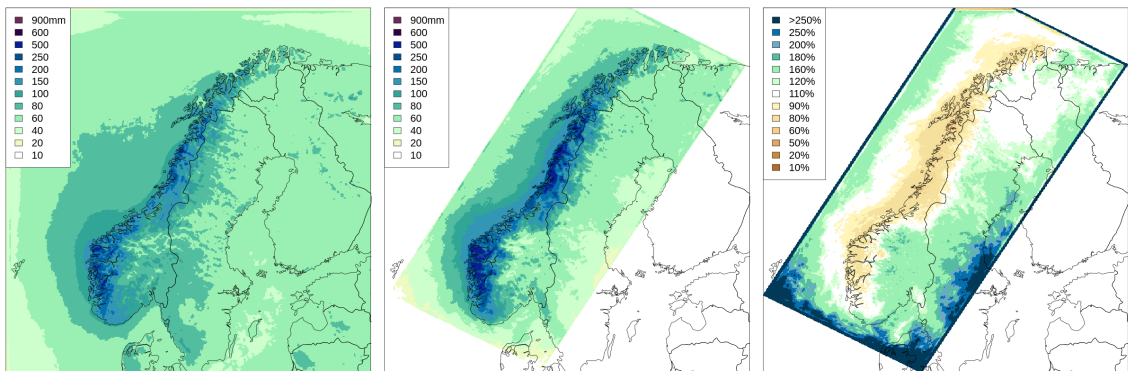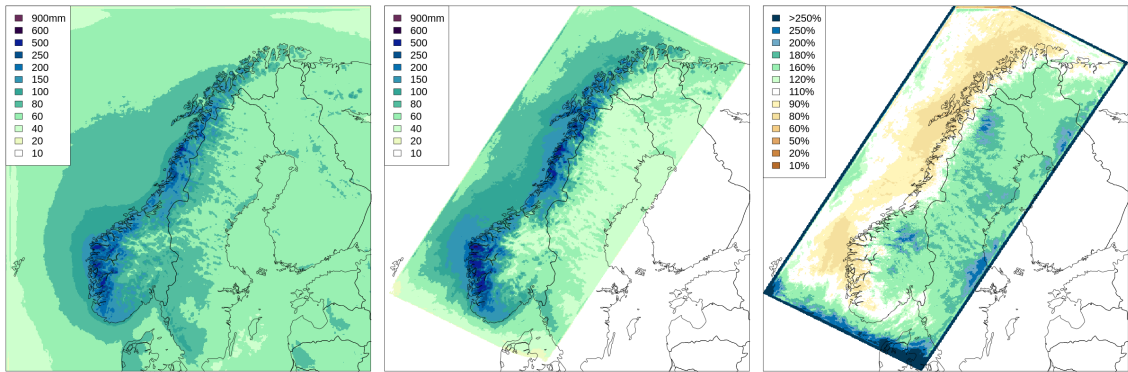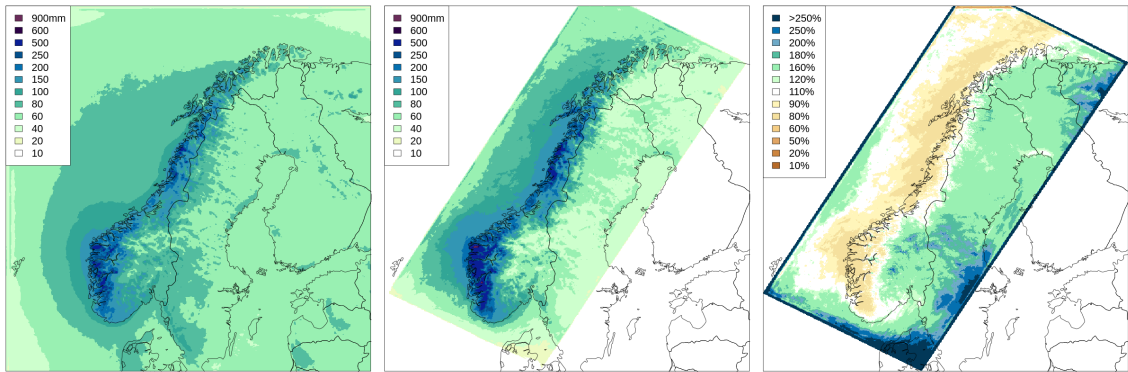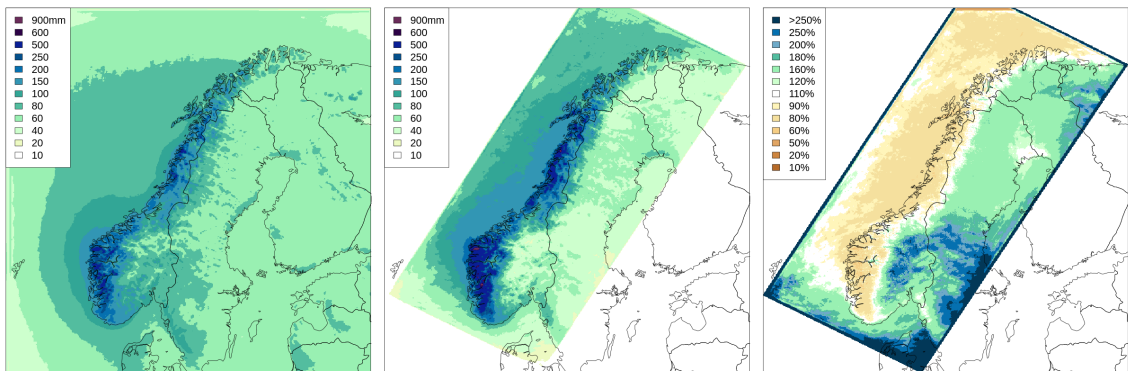Figure 17: same as Fig. 2 but for the typical July.



Figure 18: same as Fig. 2 but for the typical August.



Figure 19: same as Fig. 2 but for the typical September.

Figure 20: same as Fig. 2 but for the typical October.



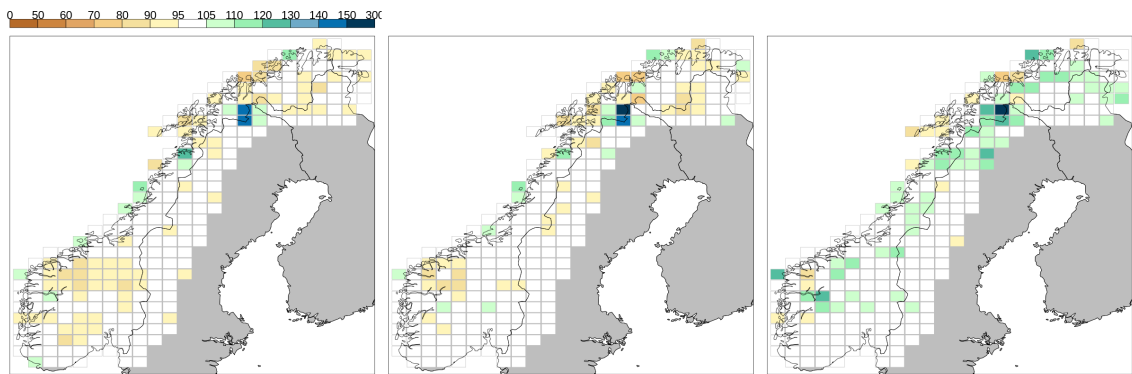Figure 21: same as Fig. 2 but for the typical November.



Figure 22: same as Fig. 2 but for the typical December.

Figure 23: Same as Fig. 7 but for spring months: March, in the left panel; April, in the middle; May, in the right panel.
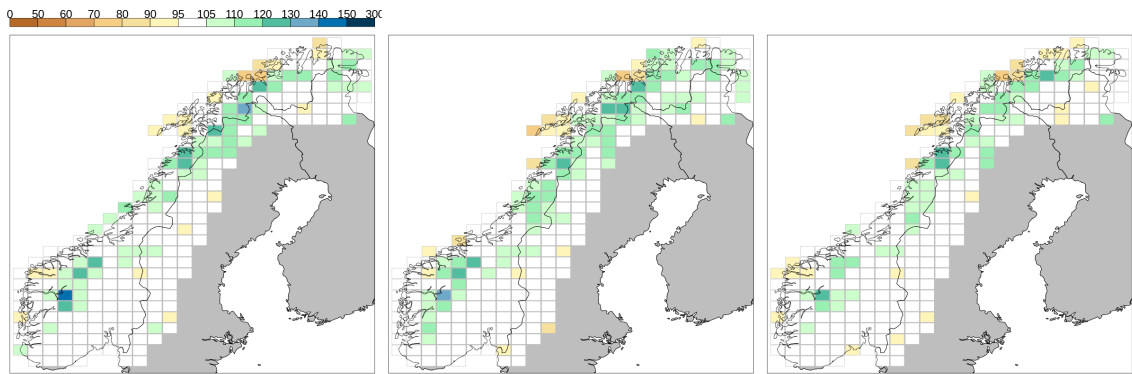


Figure 24: Same as Fig. 7 but for summer months: June, in the left panel; July, in the middle; August, in the right panel.
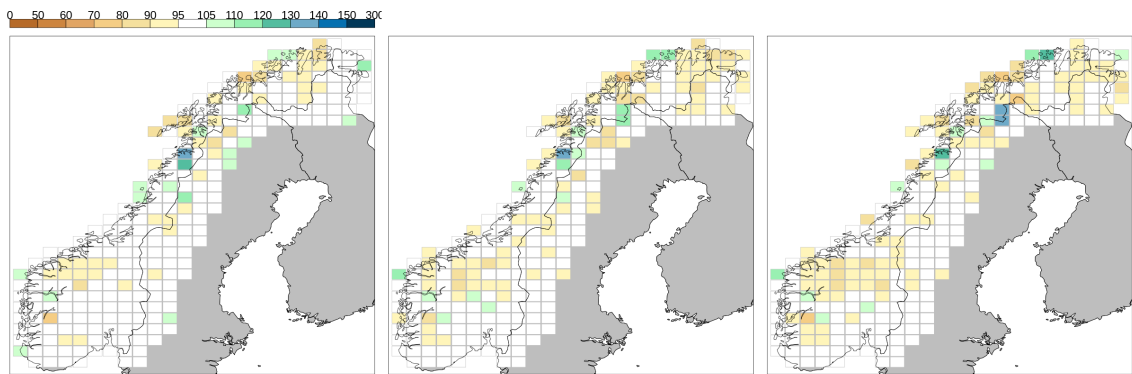


Figure 25: Same as Fig. 7 but for autumn months: September, in the left panel; October, in the middle; November, in the right panel.
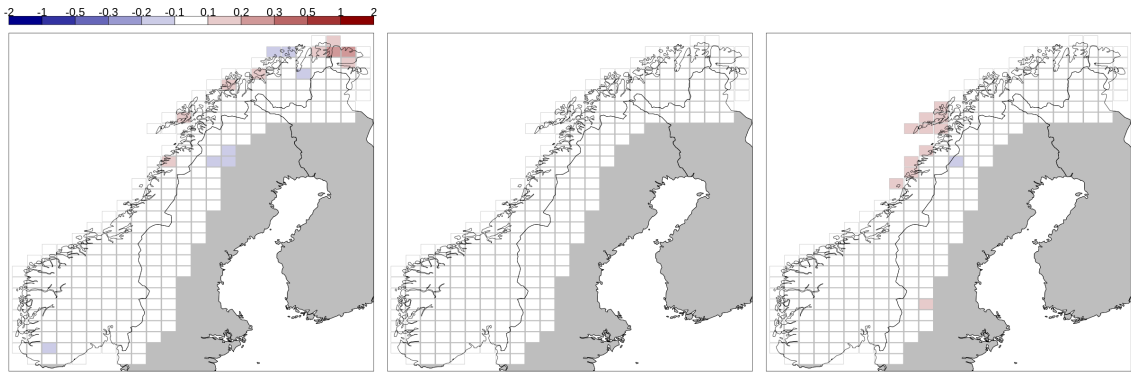
Figure 26: Same as Fig. 8 but for spring months: March, in the left panel; April, in the middle; May, in the right panel.
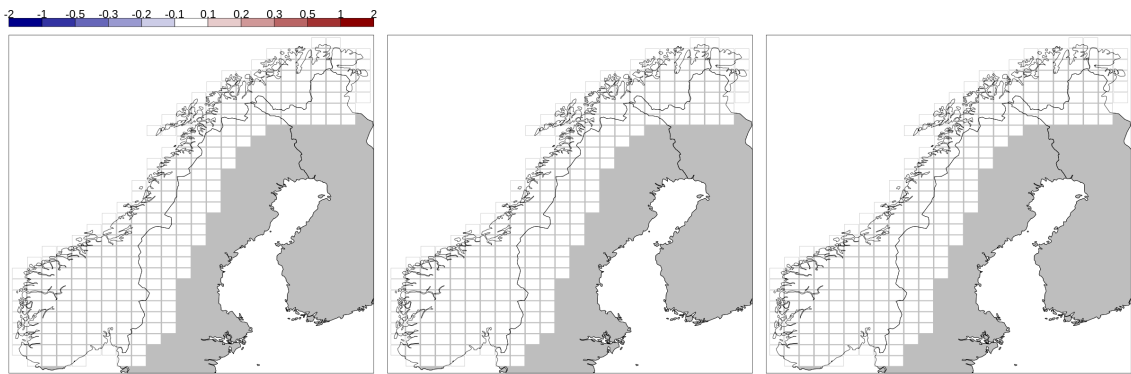


Figure 27: Same as Fig. 8 but for summer months: June, in the left panel; July, in the middle; August, in the right panel.
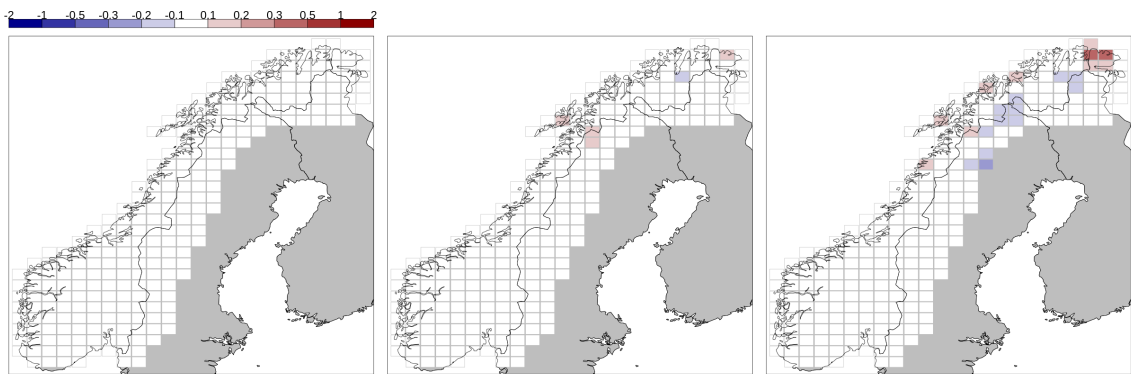


Figure 28: Same as Fig. 8 but for autumn months: September, in the left panel; October, in the middle; November, in the right panel.

Figure 29: Same as Fig. 9 but for spring months: March, in the left panel; April, in the middle; May, in the right panel.



Figure 30: Same as Fig. 9 but for summer months: June, in the left panel; July, in the middle; August, in the right panel.



Figure 31: Same as Fig. 9 but for autumn months: September, in the left panel; October, in the middle; November, in the right panel.
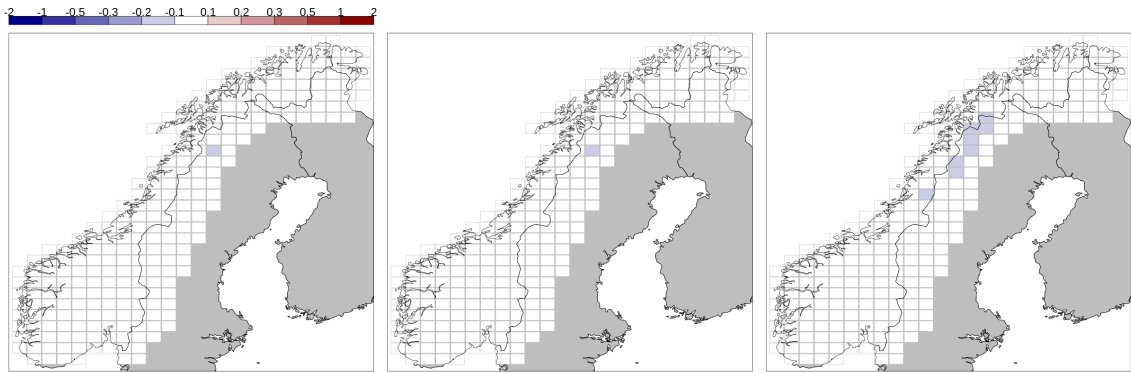
Figure 32: Same as Fig. 10 but for spring months: March, in the left panel; April, in the middle; May, in the right panel.
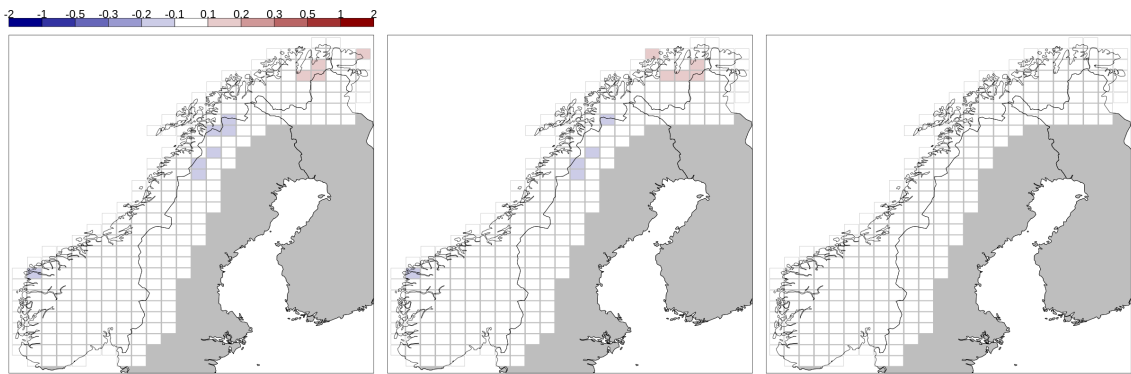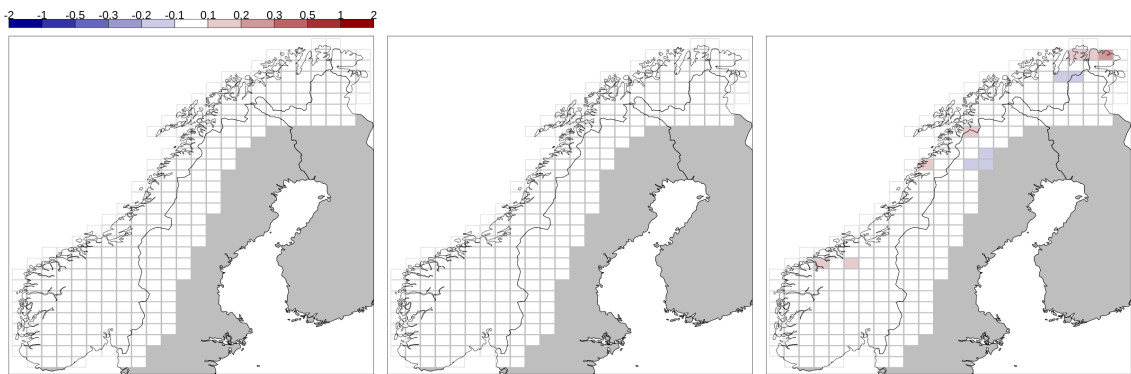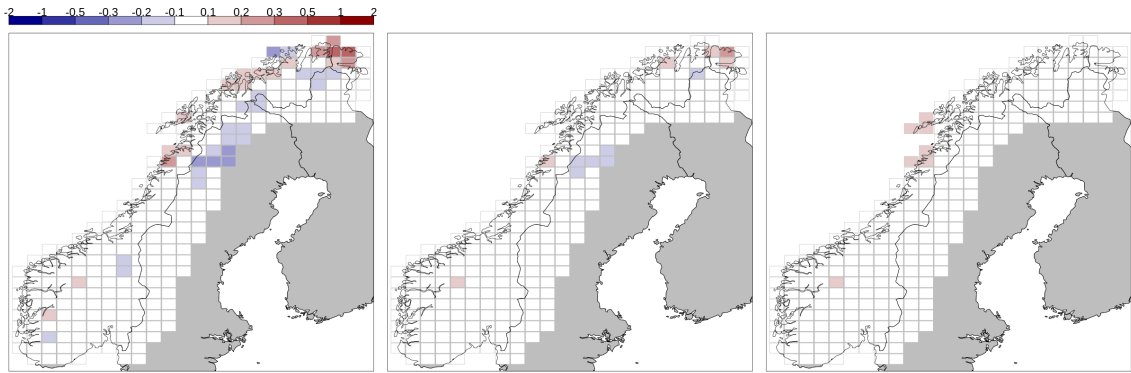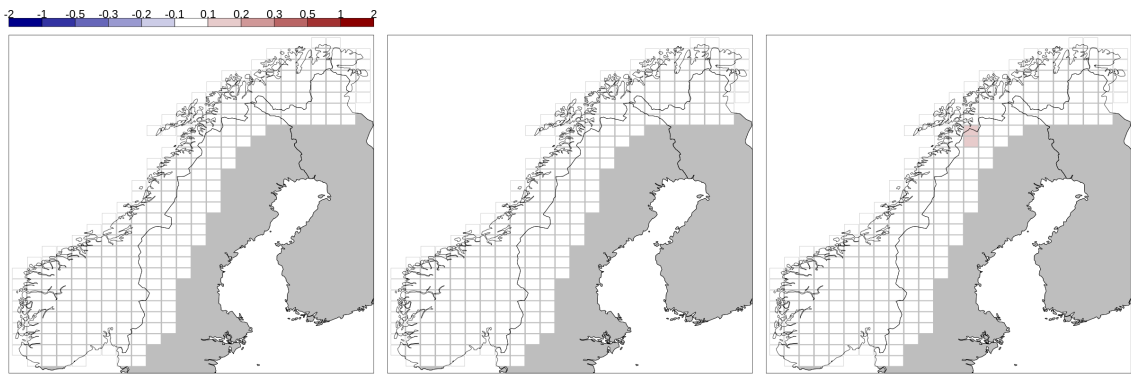


Figure 33: Same as Fig. 10 but for summer months: June, in the left panel; July, in the middle; August, in the right panel.
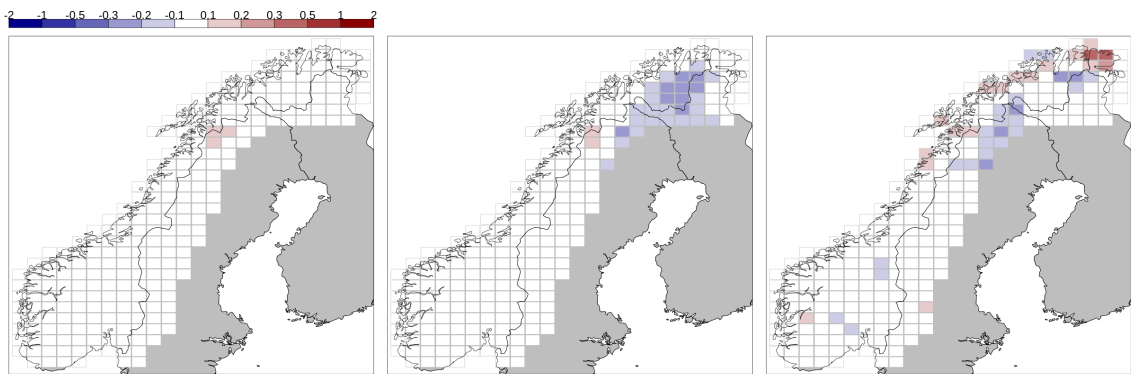


Figure 34: Same as Fig. 10 but for autumn months: September, in the left panel; October, in the middle; November, in the right panel.

Table 1: Overview of variables and mathematical notation. All the vectors are column vectors if not otherwise specified. If $\mathbf{X}$ is a matrix, $\mathbf{X}_i$ is its $i$th column (column vector) and $\mathbf{X}_{i,:}$ is its $i$th row (row vector). If $\mathbf{x}$ is a column vector, $\mathbf{x}_j$ is its $j$th element.

| symbol | description | dimension |
|---|---|---|
| $m$ | number of grid points | - |
| $p$ | number of observations | - |
| $l$ | index of OI configuration | $l = 1, \dots, 4$ |
| $\lambda$ | OI parameters | not specified |
| $\mathbf{x}^{g,a}(\lambda)$ | TG analysis over the grid, obtained with OI parameters $\lambda$ | m |
| $\mathbf{x}^{x,a}(\lambda)$ | TX analysis over the grid, obtained with OI parameters $\lambda$ | m |
| $\mathbf{x}^{n,a}(\lambda)$ | TN analysis over the grid, obtained with OI parameters $\lambda$ | m |
| $\mathbf{x}^{g,\text{IDI}}(\lambda)$ | TG IDI over the grid, obtained with OI parameters $\lambda$ | m |
| $\mathbf{x}^{x,\text{IDI}}(\lambda)$ | TX IDI over the grid, obtained with OI parameters $\lambda$ | m |
| $\mathbf{x}^{n,\text{IDI}}(\lambda)$ | TN IDI over the grid, obtained with OI parameters $\lambda$ | m |
| $\mathbf{w}^l$ | weight for OI with $\lambda_l$, used in the merging | m |
| $\alpha$ | normalization factor for the merging | m |

**Algorithm 1** seNorge_2018 procedure for the production of daily gridded datasets for an arbitrary day.

---

**Require:** daily observations of temperature and precipitation collected from the data sources

**Require:** daily averaged wind speed gridded field from numerical model output

**Require:** Gridded fields of long-term monthly averages of precipitation from numerical model output

Production of 4 provisional versions of TG fields, OI with $\lambda_l$, $l = 1, \ldots, 4$

Merging of the TG versions into the definitive TG field

Adjust RR observations for wind-induced undercatch based on TG and the wind fields

Production of RR gridded fields

Production of 4 provisional versions of TX fields, OI with $\lambda_l$, $l = 1, \ldots, 4$

Merging of the TX versions into the definitive TX field

Production of 4 provisional versions of TN fields, OI with $\lambda_l$, $l = 1, \ldots, 4$

Merging of the TN versions into the definitive TN field

Check consistency between TX and TG. Modify TX where the consistency is violated

Check consistency between TN and TG. Modify TN where the consistency is violated

---

# References

Førland, E., and O. Tveito (1997), Temperatur og snødata for flomberegning, *DNMI Report*, *28*(97), 51.

Frogner, I.-L., A. T. Singleton, M. Ø. Køltzow, and U. Andrae (2019), Convection-permitting ensembles: challenges related to their design and use, *Quarterly Journal of the Royal Meteorological Society*, *0*(0), doi:10.1002/qj.3525.

Gandin, L. S., and R. Hardin (1965), *Objective analysis of meteorological fields*, vol. 242, Israel program for scientific translations Jerusalem.

Klein Tank, A., J. Wijngaard, G. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, et al. (2002), Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment, *International journal of climatology*, *22*(12), 1441–1453.

Lind, P., D. Belušić, O. B. Christensen, A. Dobler, E. Kjellström, O. Landgren, D. Lindstedt, D. Matte, R. A. Pedersen, E. Toivonen, and F. Wang (2020), Benefits and added value of convection-permitting climate modeling over fenno-scandinavia, *Climate Dynamics*, pp. 1–20, doi:10.1007/s00382-020-05359-3.

Lussana, C. (2020a), senorge_2018 daily total precipitation amount 1957-2019, doi:10.5281/zenodo.3923703, observational gridded dataset over the Norwegian mainland.

Lussana, C. (2020b), senorge_2018 daily mean temperature 1957-2019, doi:10.5281/zenodo.3923706, observational gridded dataset over the Norwegian mainland.

Lussana, C. (2020c), senorge_2018 daily minimum temperature 1957-2019, doi:10.5281/zenodo.3923697, observational gridded dataset over the Norwegian mainland.

Lussana, C. (2020d), senorge_2018 daily maximum temperature 1957-2019, doi:10.5281/zenodo.3923700, observational gridded dataset over the Norwegian mainland.

Lussana, C., F. Uboldi, and M. R. Salvati (2010), A spatial consistency test for surface observations from mesoscale meteorological networks, *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 1075–1088.

Lussana, C., T. Saloranta, T. Skaugen, J. Magnusson, O. E. Tveito, and J. Andersen (2018a), seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, *Earth System Science Data*, *10*(1), 235–249, doi: 10.5194/essd-10-235-2018.

Lussana, C., O. E. Tveito, and F. Uboldi (2018b), Three-dimensional spatial interpolation of 2m temperature over Norway, *Quarterly Journal of the Royal Meteorological Society*, *144*(711), 344–364, doi:10.1002/qj.3208.

Lussana, C., O. E. Tveito, A. Dobler, and K. Tunheim (2019), senorge_2018, daily precipitation, and temperature datasets over norway, *Earth System Science Data*, *11*(4), 1531–1551, doi:10.5194/essd-11-1531-2019.

Mohr, M. (2008), New routines for gridding of temperature and precipitation observations for "senorge. no", *Met. no Report*, *8*, 2008.

Müller, M., M. Homleid, K.-I. Ivarsson, M. A. Køltzow, M. Lindskog, K. H. Midtbø, U. Andrae, T. Aspelien, L. Berggren, D. Bjørge, et al. (2017), Arome-metcoop: a nordic convective-scale operational weather prediction model, *Weather and Forecasting*, *32*(2), 609–627.

Reistad, M., Ø. Breivik, H. Haakenstad, O. J. Aarnes, B. R. Furevik, and J.-R. Bidlot (2011), A high-resolution hindcast of wind and waves for the north sea, the norwegian sea, and the barents sea, *Journal of Geophysical Research: Oceans*, *116*(C5).

Tarantola, A. (2005), *Inverse Problem Theory and methods for model parameter estimation*, SIAM, Philadelphia.

Tveito, O., E. Førland, R. Heino, I. Hanssen-Bauer, H. Alexandersson, B. Dahlström, A. Drebs, C. Kern-Hansen, T. Jónsson, E. Vaarby Laursen, et al. (2000), Nordic temperature maps, *DNMI report*, *9*(00).

Tveito, O. E., and E. J. Førland (1999), Mapping temperatures in Norway applying terrain information, geostatistics and GIS, *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography*, *53*(4), 202–212.

Tveito, O. E., H.-C. Udnæs, Z. Mengistu, R. Engeset, and E. J. Førland (2002), New snow maps for Norway, in *Proceedings XXII Nordic Hydrological Conference 2002*, pp. 527–532.

Tveito, O. E., I. Bjørdal, A. O. Skjelvåg, and B. Aune (2005), A gis-based agro-ecological decision system based on gridded climatology, *Meteorological Applications*, *12*(1), 57–68.

Uboldi, F., C. Lussana, and M. Salvati (2008), Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks, *Meteorological Applications*, *15*(3), 331–345.

Wolff, M. A., K. Isaksen, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan (2015), Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a norwegian field study, *Hydrology and Earth System Sciences*, *19*(2), 951–967, doi:10.5194/hess-19-951-2015.