seNorge_2018, daily precipitation and temperature datasets over Norway

Cristian Lussana[1], Ole Einar Tveito[1], Andreas Dobler[1], and Ketil Tunheim[1]

[1]Norwegian Meteorological Institute, Oslo, Norway

Correspondence: Cristian Lussana (critianl@met.no)

Abstract. seNorge_2018 is a collection of observational gridded datasets over Norway for: daily total precipitation; daily mean, maximum and minimum temperatures. The time period covers 1957 to 2017, and the data are presented over a high-resolution terrain-following grid with 1 km spacing in both meridional and zonal directions. The seNorge family of observational gridded datasets developed at the Norwegian Meteorological Institute (MET Norway) has a twenty-year long history and seNorge_2018 is its newest member, the first providing daily minimum and maximum temperatures. seNorge datasets are used for a wide range of applications in climatology, hydrology and meteorology. The observational dataset is based on MET Norway's climate data, which has been integrated by the "European Climate Assessment and Dataset" database. Two distinct statistical interpolation methods have been developed, one for temperature and the other for precipitation. They are both based on a spatial scale-separation approach where, at first, the analysis (i.e., predictions) at larger spatial scales are estimated. Subsequently they are used to infer the small-scale details down to a spatial scale comparable to the local observation density. Mean, maximum and minimum temperatures are interpolated separately, then physical consistency among them is enforced. For precipitation, in addition to observational data, the spatial interpolation makes use of information provided by a climate model. The analysis evaluation is based on cross-validation statistics and comparison with a previous seNorge version. The analysis quality is presented as a function of the local station density. We show that the occurrence of large errors in the analyses decays at an exponential rate with the increase in the station density. Temperature analyses over most of the domain are generally not affected by significant biases. However, during wintertime in data-sparse regions the analyzed minimum temperatures do have a bias between $2°C$ and $3°C$. Minimum temperatures are more challenging to represent and large errors are more frequent than for maximum and mean temperatures. The precipitation analysis quality depends crucially on station density: the frequency of occurrence of large errors for intense precipitation is less than 5% in data-dense regions, while it is approximately 30% in data-sparse regions. The open-access datasets are available for public download at: daily total precipitation (DOI:https://doi.org/10.5281/zenodo.2082320 Lussana, 2018b) ; daily mean (DOI:https://doi.org/10.5281/zenodo.2023997 Lussana, 2018c) , maximum (DOI:https://doi.org/10.5281/zenodo.2559372 Lussana, 2018e) and minimum (DOI:https://doi.org/10.5281/zenodo.2559354 Lussana, 2018d) temperatures

1 Introduction

Long-term observational gridded datasets of near-surface meteorological variables are widely used products. In climatology, they have been used for example to monitor the regional climate (Simmons et al., 2017) and to validate and bias-correct climate simulations (Kotlarski et al., 2017). In meteorology, they are used at national meteorological institutes, such as the Norwegian

5 Meteorological Institute (MET Norway), to monitor and report the weather conditions. In hydrology, they are used as external forcing for hydrological and snow modeling (Saloranta, 2012; Skaugen and Onof, 2014; Magnusson et al., 2015).

seNorge_2018 is a collection of four long-term observational datasets over Norway covering the 61-year time period 1957-2017 for: daily total precipitation (RR), daily mean temperature (TG), daily minimum (TN) and maximum (TX) temperatures. It builds upon the previous work on establishing MET Norway's observational datasets (Tveito and Førland, 1999; Lussana

10 et al., 2018a, b) and the core of its statistical interpolation method is the Optimal Interpolation (OI, Gandin and Hardin, 1965; Kalnay, 2003). A review of the relevant literature for our spatial interpolation applications is given in the paper by Lussana et al. (2018a).

Like the previous versions of seNorge, precipitation and temperature data are provided on a high-resolution grid with 1 km grid spacing in both meridional and zonal directions. seNorge_2018 aims at achieving a higher effective resolution of

15 the analyzed (or predicted) fields than the previous versions. The difference between grid spacing and resolution is described by Grasso (2000). In the context of numerical modeling, Walters (2000) defines the "effective resolution as the minimum wavelength the model can describe with some required level of accuracy (not defined)" and it concludes that as many as 10 gridpoints may be required to properly represent a field. As pointed out by Pielke (2001), there is a subjective component in the number of gridpoints needed to resolve a feature in a field. It is worth spending a few more words on effective resolution

20 in OI. In contrast to in-situ observations which represent point values, our gridded analyses produce areal averages. What this means is that for each grid point, we calculate weighted averages of the nearest observations. The extensions of the spatial supports for those averages depend on the settings of the statistical interpolation, which, in turn, are optimized on the station spatial distribution. The larger the extensions of the spatial supports, the lower the effective resolution of the analysis fields. As demonstrated in the Appendix C of Uboldi et al. (2008), an OI scheme is realizing a low-pass filter whose cut-off wavelength

25 is determined by the OI settings.

The main original aspect of our research is that the spatial interpolation methods automatically adapt OI settings to the local station density, such that in data-dense regions the spatial supports of the areal-averaged analyses are smaller than in data-sparse regions. In other words, the effective resolution of the analysis fields is higher in data-dense than in data-sparse regions. The user of the seNorge_2018 must be aware that (i) the comparison between different sub-regions over the domain

30 is influenced by the respective local station densities, and (ii) variations in the observational network over time will affect temporal trends derived from this dataset. According to Masson and Frei (2016), to overcome those limitations a further post-processing of seNorge_2018 would be required so to create a new dataset, less accurate but suitable for the investigation of long-term variations and trends.

The following definitions of spatial scales are used in the text. Regional scale coincides with the whole domain. Given the importance of the observational network, at an arbitrary point we refer to scales that are defined with respect to the station distribution in its surroundings. Sub-regional scale (or local scale) defines an area -around the point- that includes dozens of observations (10-100). Small-scale defines an area that includes few observations (1-10). Unresolved scale refers to those spatial scales that are smaller than the average distance between a station and its closest neighbours, such that atmospheric fields could not be properly represented by the observational network.

The presented research includes several other original aspects. In the case of precipitation, the measurements have been adjusted for the wind-induced under-catch in a way that is consistent with the method proposed by Wolff et al. (2015). A multi-scale OI scheme has been implemented on precipitation relative anomalies with respect to a reference field that captures the field variability at unresolved spatial scales. Furthermore, a Box-Cox transformation has been used to get the precipitation relative anomalies into a normal shape as required by OI. The use of a reference field as a first-guess for the precipitation patterns has been proven to be a successful approach by Masson and Frei (2014). They found that daily precipitation over the Alpine region is well represented by using the seasonal precipitation mean as a single predictor field in Kriging with external drift. The precipitation observational network is extremely sparse over significant portions of our domain, such as in mountainous regions, where the vast majority of stations are located on the valley floors (Lussana et al., 2018a); and in the Arctic region. As a consequence, we have chosen not to use precipitation climatologies derived by observational gridded datasets. Instead the reference is derived from long-term averages calculated from the output of a high-resolution numerical model. We have used a regional climate simulation with a resolution of 2.5 km, based on the dynamical downscaling of the global reanalysis ERAInterim and available for the time period 2003-2016, to derive the monthly reference fields, as this has been proven skillful for such an application (Crespi et al., 2018).

In the case of temperature, seNorge_2018 is the first seNorge dataset including daily minimum and maximum temperatures. The three temperature variables are treated separately with the same interpolation method. With respect to seNorge2 (Lussana et al., 2018b), the regional spatial trend of temperature is obtained as the blending of a much larger number of sub-regional trends. The analysis method has been implemented on a gridpoint-by-gridpoint basis so to take advantage of a local Kalman gain. As a result, the effective resolution of the analyzed fields is optimized over the domain.

The structure of the paper is as follows. Section 2 presents the observational network and the regional climate simulation used as the precipitation reference. The spatial interpolation methods are described in Section 3. Finally, Section 4 presents the results and the validation of seNorge_2018. With its intricate coastline and complex terrain, Norway is an excellent region for testing spatial interpolation schemes under challenging conditions. In this sense, the evaluation presented can be useful to infer the performances of the presented methods over other regions as well.

2   Data

The in-situ observations are retrieved from MET Norway's climate database and the European Climate Assessment and Dataset (ECA&D, Klein Tank et al., 2002). The spatial domain covers the Norwegian mainland, plus an adjacent strip of land extending

into Sweden, Finland and Russia in order to reduce boundary effects along the Norwegian border. The observations have been quality controlled by experienced staff and with the help of automatic procedures, such as the spatial consistency test described by Lussana et al. (2010). The variables are defined as following: TG is the 24-hour average between 06:00 UTC of the day, reported as time-stamp and 06:00 UTC of the previous day; RR is the accumulated precipitation over the same time interval

5   as TG, moreover RR data has been corrected for the wind-induced under-catch of the gauges; TX and TN are, respectively, the maximum and minimum observed temperatures between 18:00 UTC of the day reported as time-stamp and 18:00 UTC of the previous day. TG and RR share the same day-definition so as to better serve hydrological applications, while for historical reasons TX and TN have a different day definition.

The measured RR value (i.e., RR_{raw}) at an arbitrary location is adjusted for wind-induced under-catch of solid precipitation

10   by means of a procedure similar to the one presented by Wolff et al. (2015):

$$\alpha = \tau_1 + (\tau_2 - \tau_1) \left\{ \exp\left[(TG - T_\tau)/s_\tau \right] / \left(1 + \exp\left[(TG - T_\tau)/s_\tau \right] \right) \right\} \tag{1}$$

$$\gamma = [1 - \alpha] \exp\left[-(W/\theta)^\beta \right] + \alpha \tag{2}$$

$$RR = \gamma^{-1} RR_{raw} \tag{3}$$

where TG is extracted from the analysis field (Sec. 3.1) so to always have a temperature estimate; W is the ten-metre wind

15   speed at the station location extracted from a gridded dataset derived from numerical model output. The (NORA10, Reistad et al., 2011) wind speed dataset, which covers the whole time period 1957-2017, has been downscaled onto the 1 km grid by using a quantile mapping approach (Bremnes, 2004) to match the climatology of the high-resolution numerical weather prediction model (AROME-MEtCoOp, Müller et al., 2017). The wind dataset is available for public download at http://thredds. met.no/thredds/catalog/metusers/klinogrid/KliNoGrid_16.12/FFMRR-Nor/catalog.html. In the original paper by Wolff et al.

20   (2015), they were considering sub-daily precipitation measurements and both temperature and wind were measured at the same location as precipitation. We are operating under different conditions and the requirement of having temperature and wind measurements together with precipitation would reduce the number of suitable observations to a very small subset. As a consequence, in Eqs. (1)- (2) we had to use parameter values which are different from those used by Wolff et al. (2015). We have decided to use seNorge version 1.1 (Mohr, 2008, 2009) as a reference for the extreme values returned by the precipitation

25   adjustment. seNorge version 1.1 includes a precipitation correction based on geographical parameters, summarized in site exposure classes such that a systematic increase of precipitation is carried out. The correction presented in Eqs. (1)- (3) takes advantage of wind and temperature estimates but we do not expect the extreme values of those two corrections to differ significantly. The parameter values used in Eqs. (1)- (3), which have been optimized to better match seNorge version 1.1 extremes, are: $\theta = 4.7449$, $\beta = 0.6667$, $\tau_1 = 0.4930$, $\tau_2 = 0.9134$, $T_\tau = 0.9134$, $s_\tau = 0.7759$.

30   Figure 1 shows the observational network and its evolution in time. The number of available observations was rather stable from 1957 to 2000. In the following decade, the number of RR observations dropped to 500, which was the minimum value, and then it gradually increased again to over 600 in the recent years. The number of temperature observations has been constantly increasing since year 2000, and for 2017 there are about twice as many stations as in 1957. The meteorological stations have been mainly installed to monitor the weather in cities and villages, so the network is denser in urban areas. In the mountainous

regions, the digital elevation model can reach 2000 m but most of the stations are located below the elevation of 1000 m. A difference in the station density between the southern and the northern portion of the domain is also clearly visible, with a higher density in the south of Norway. Ideally, spatial interpolation would require a denser network of observations where the variance of a variable is larger, in order to get a fine-scale representation of the temperature field where it varies the most.

5 However, this is hardly the case in most situations because of the inherent difficulties in station installation and maintenance over complex terrain and in remote areas. As a result, we should expect better performances of the interpolation methods over urban areas and larger analysis uncertainties over data-sparse areas, such as mountainous regions.

As stated in the Introduction, the spatial interpolation depends on the station density. For this reason, the Integral Data Influence (IDI: Uboldi et al., 2008; Lussana et al., 2010) has been introduced as a diagnostic parameter and it is shown in the

10 bottom panels of Fig. 1. IDI is similar to the degrees of freedom introduced by Cardinali et al. (2004) and it has been used also to evaluate the distribution of weather stations (Horel and Dong, 2010). In practice, IDI is obtained as the result of an OI performed by arbitrarily assigning the value of 1 to the observations (i.e., maximum amount of available information) and the reference value of 0 to the background (i.e, basic amount of information available everywhere). The analytical function that usually represents the background error correlation in OI, in the case of IDI is representing the station influence on the analysis

15 according to a predefined metric. This metric is defined as a function of the geographical parameters. For an arbitrary point in space, the geographical parameters are stored in a vector $\mathbf{r}$ having four components: latitude, longitude, altitude and land area fraction (i.e. fraction of land in the 1 km square box centered at the point). The land area fraction is introduced here and used in Sec. 3. Functions are applied to pair of points, such as: $d(\mathbf{r}, \mathbf{s})$ returns the horizontal (radial) distance in km between $\mathbf{r}$ and $\mathbf{s}$; $z(\mathbf{r}, \mathbf{s})$ returns their absolute elevation difference; $w(\mathbf{r}, \mathbf{s})$ returns their absolute land area fraction difference. The correlation

20 function between two points $\mathbf{r}$ and $\mathbf{s}$ is based on Gaussian functions of the form:

$$f_u(\mathbf{r}, \mathbf{s}; D) = \exp\left\{ -\frac{1}{2} \left[\frac{u(\mathbf{r}, \mathbf{s})}{D} \right]^2 \right\} \qquad (4)$$

where: $u()$ is an arbitrary function, such as the ones previously defined, applied to the points; D is a reference length scale governing the decreasing rate. We have chosen to model the station influence using Gaussian functions. For TG, TX and TN, the station influence is factorized into the product of two Gaussian functions: one depending on distances, such that in Eq. (4)

25 $u = d()$ and $D = 50$ km; the other depending on elevation differences, with $u = z()$ and $D = 200$ km. In the case of RR, the station influence depends only on distances, therefore $u = d()$ and $D = 10$ km. The values of the de-correlation length scales used for temperature are consistent with the findings of Sec. 3.1, while for precipitation the value chosen is representative of the smallest spatial scales used in the iterative cycle. Elevation plays always a predominant role for temperature and even only a few stations at higher elevations can provide a reasonable approximation of the sub-regional near-surface temperature lapse

30 rate. Then, it is important to know where this information is not available and the temperature IDI map in Fig. 1 shows those regions. On the other hand, for precipitation we have decided to not consider elevation because we are aware that our network is very sparse at higher elevations and for this reason we have introduced a reference field, as stated in the Introduction. The precipitation IDI map in Fig. 1 highlights the potential of our stations to interact on the horizontal plane. In the following, the cross-validation (CV) approach is used such that the summary statistics derived at station locations can be considered valid for

gridpoints. The CV-IDI at station locations (i.e., IDI at a station location computed without considering the presence of that station) is introduced to link the CV statistics to the IDI of the hypothetical gridpoint represented by a station location. In the two maps of Fig. 1, the IDI is shown over the 1 km grid for TG and RR. The IDIs for TX and TN are very similar to the TG map. In the vicinity of an observation the IDI field tend to stay close to 1 whereas for data sparse areas its value is close to 0. The

5 IDI and CV-IDI values have been divided into classes: values smaller than 0.45 defines observations/gridpoints in data-sparse regions (i.e., where the station influence on the analysis is very limited); values larger than 0.85 defines observations/gridpoints in data-dense regions (i.e., where the station influence on the analysis is substantial), then two intermediate classes have been defined to have an idea of the behaviour of the spatial interpolation scheme in the transition between data-dense and data-sparse regions. Fig. 1a and Fig. 1b show the close relationship between CV-IDI and the station density. In Fig. 1a the distribution of

10 CV-IDI values is shown by boxplots: the horizontal line is the median of the distribution, the box width is the interquartile range. As shown by Fig. 1c, at station locations the IDI values are confined into a smaller range of values than the CV-IDI. Even an isolated station constitutes more information than the background alone, while an isolated gridpoint must have IDI equals to 0 as it is CV-IDI at an isolated stations.

The reference datasets used for precipitation are based on hourly precipitation provided by the climate model version of

15 HARMONIE (version cy38h1.2), a seamless NWP model framework developed and used by several national meteorological services. HARMONIE includes a set of different physics packages adapted for different horizontal resolutions. For the high-resolution, convection permitting simulations in this case, the model has been set-up with AROME physics (Seity et al., 2011) and the SURFEX surface scheme (Masson et al., 2013). The climate runs have been carried out within the HARMONIE script system, covering the period July 2003 to December 2016 on a 2.5 km grid over the Norwegian main land. More details on the

20 climate model can be found in Lind et al. (2016), references therein and on https://www.hirlam.org/trac/wiki/HarmonieClimate. The numerical model does not include measurements from the network of rain-gauges. The mean monthly total precipitation fields have been computed considering the available hourly data and they have been used as reference fields for the spatial interpolation of precipitation as described in Sec. 3.2.

3   Methods

25 The notation used is based on both Ide et al. (1997) and Sakov and Bertino (2011). The number of gridpoints is m. The number of observations is p. Upper-case bold symbols are used for matrices, lower-case bold symbols for vectors and italic symbols for scalars. For an arbitrary matrix $\mathbf{X}$, $\mathbf{X}_i$ means the ith column; $\mathbf{X}_{i,:}$ the ith row; and $\mathbf{X}_{ij}$ the element at the ith row and jth column. For an arbitrary vector $\mathbf{x}$, $\mathbf{x}_i$ denotes the ith element. The superscripts on the upper left hand corner of a symbol identify: analysis a; background b; observation o. Upper accents have been used too. In the case of temperature, where we

30 iterate over the gridpoints, the notation $\overset{i}{\mathbf{X}}$ indicates that matrix $\mathbf{X}$ is valid for the ith gridpoint and in this sense we may refer to it as a local matrix. In the case of precipitation, where we iterate over spatial scales, the same notation $\overset{i}{\mathbf{X}}$ indicates that matrix $\mathbf{X}$ is obtained as a function of the ith spatial scale. Upper accents are not used only for matrices, for instance the in-situ

observations are stored in the p-vector $\mathbf{y}^o$ but in the following we will refer to the $\overset{i}{p}$-vector $\overset{i}{\mathbf{y}}^o$ of the nearest observations to the ith gridpoint.

3.1 Temperature

The same interpolation scheme is used for the mean, the maximum and the minimum daily temperature. The physical consis-
5 tency among them is assured by post-processing the independently analyzed datasets, such that for each gridpoint: minimum temperature is always smaller or equal to the mean, maximum temperature is always bigger or equal to the mean.

The spatial interpolation is implemented on a gridpoint-by-gridpoint basis. It combines a regional pseudo-background field, that is the weighted average of numerous sub-regional fields, with the observations. The temperature analysis at the generic ith gridpoint is written as:

10 $$\mathbf{x}_i^a = \mathbf{x}_i^b + \overset{i}{\mathbf{K}}_{i,:} \left(\overset{i}{\mathbf{y}}{}^o - \overset{i}{\mathbf{y}}{}^b \right) \qquad (5)$$

$\overset{i}{\mathbf{y}}{}^o$ and $\overset{i}{\mathbf{y}}{}^b$ are the $\overset{i}{p}$-vectors of the nearest stations to the ith gridpoint.

The local Kalman gain in Eq. (5) is:

$$\overset{i}{\mathbf{K}}_{i,:} = \overset{i}{\mathbf{G}}_{i,:} \left(\overset{i}{\mathbf{S}} + \varepsilon^2 \overset{i}{\mathbf{I}} \right)^{-1} \qquad (6)$$

$\overset{i}{\mathbf{I}}$ is the $\overset{i}{p} \mathrm{x} \overset{i}{p}$ identity matrix and $\varepsilon^2 \equiv \sigma_o^2 / \sigma_b^2$ is the ratio between the constant observed (σ_o^2) and pseudo-background (σ_b^2)
15 error variances that has been set to 0.5, as for seNorge2 (Lussana et al., 2018b). The local pseudo-background error correlation matrices are defined on the basis of the correlation function between pair of points $\rho^T(\mathbf{r}_j, \mathbf{r}_k)$ as:

$$\rho^T(\mathbf{r}_j, \mathbf{r}_k) = f_d(\mathbf{r}_j, \mathbf{r}_k; D_i^h) \ f_z(\mathbf{r}_j, \mathbf{r}_k; D^z) \ [1 - (1 - w_{min})|w(\mathbf{r}_j, \mathbf{r}_k)|] \qquad (7)$$

such that the correlation between the jth gridpoint and the kth station is $\overset{i}{\mathbf{G}}_{jk} = \rho^T(\mathbf{r}_j, \mathbf{r}_k)$. Analogously, the correlation between the jth station and the kth station is $\overset{i}{\mathbf{S}}_{jk} = \rho^T(\mathbf{r}_j, \mathbf{r}_k)$. The Gaussian functions f are defined in Eq. (4). A formulation
20 similar to Eq. (7) has been used in the paper by Lussana et al. (2009), in that case the land area fraction has been replaced by the land use. w_{min} sets the minimum value for the factor related to land area fraction when $w(\mathbf{r}_i, \mathbf{r}_j)$ is maximum (i.e., equals to 1). D^z and w_{min} are fixed over the domain, while D_i^h is allowed to vary between gridpoints, although with some restrictions. In an ideal situation of a very dense observational network, one may consider to rely on adaptive estimates for the three parameters. This is not the case for our station distribution, so we have opted for a "hybrid" configuration (i.e., D^z
25 and w_{min} fixed; D^h adaptive) that would return robust estimates. The impact of large land area fraction differences on ρ^T is less dramatic than those of large horizontal or elevation differences and it also impacts only a limited number of stations along the coast. Eventually, we have manually set $w_{min} = 0.5$ to achieve the desired effect of attenuating the influence of coastal areas over inland areas and vice versa, while at the same time avoiding the introduction of sharp gradients between those two regions. The optimization procedure for D_i^h and D^z is described in the following of this section.

The pseudo-background x_i^b in Eq. (5) is the blending of n sub-regional pseudo-backgrounds and it is in many ways similar to those described by Lussana et al. (2018b). Each sub-regional pseudo-background is defined by a centroid and it includes only the 30 stations closest to this centroid. The pseudo-background field with centroid at $\mathbf{r}_c$ is the m-vector $\overset{c}{\mathbf{x}}^b$ and its value at the ith gridpoint is $\overset{c}{\mathbf{x}}_i^b$. The seNorge_2018 domain has been divided on a 50x50 grid, each cell is a 24 km by 31 km rectangular

5   box and the nodes (i.e., centres of the cells) are the "candidate" centroids. If a node is inside the domain and has at least 30 stations in a neighbourhood of 250 km, then it is a suitable centroid. Those 30 temperature observations are used to estimate a sub-regional pseudo-background field as a function of the elevation only. The analytical function used to model the vertical profile of temperature is the one proposed by Frei (2014) for the alpine region and its parameters have been obtained by fitting the function to the aforementioned 30 observations. We assume that 30 observations can provide a reliable fitting. The generic

10   cth pseudo-background field $\overset{c}{\mathbf{x}}^b$ is derived directly from the digital elevation model by assuming that the cth sub-regional vertical temperature profile is valid for the whole domain. The number of sub-regions n is usually between 500 and 600 and there are significant overlaps between neighbouring sub-regions, such that the continuity of the regional pseudo-background is guaranteed. Finally, $\mathbf{x}_i^b$ is a weighted average of n values:

$$\mathbf{x}_i^b = \frac{\sum_{c=1}^{n} \overset{c}{\mathbf{x}}_i^w \, \overset{c}{\mathbf{x}}_i^b}{\sum_{c=1}^{n} \overset{c}{\mathbf{x}}_i^w} \tag{8}$$

15   where the weights at the ith gridpoint $\overset{c}{\mathbf{x}}_i^w$ are the n IDI values (Sec. 2) and $\overset{c}{\mathbf{x}}_i^w$ is computed considering only those stations included in cth sub-regional pseudo-background field. The settings used in the IDI calculation are similar to those used for precipitation in Fig. 1, in the sense that the station influence decays with horizontal distance only and its de-correlation length scale is set to 27.5 km, that is the average of a box width and height on the 50x50 grid.

    The optimization of D^z and D_i^h of Eq. (7) is based on the statistics of the innovation (i.e. observation minus background) at

20   station locations. As described by Desroziers et al. (2005), the elements of the background error covariance matrix at station locations, which is modeled by us as $\sigma_b^2 \mathbf{S}$, should match the innovation sample covariances. In Tables 1-3, the values of the parameters determining $\sigma_b^2 \mathbf{S}$ are shown for a selection of year (1960, 1970, ...) in the assumption of a constant D^h (i.e., $D_i^h = Dh, \forall i = 1, \ldots, m$). Note that we have also added the average number of stations available for a specific year, the average distances between them and the estimated observation error variance, which is not strictly required to compute $\mathbf{S}$ and it is set to

25   be half of σ_b^2 in our analysis. The TN error variances are significantly higher than those for TG and TX, thus indicating that TN is a more challenging variable to interpolate. D^h and D^z do not differ significantly among TG, TX and TN, probably because the common observational network is the main constrain in determining their value. This justifies our choice to set $D^z = 210$ m for the three variables. The parameter values in the Tables are more influenced by the majority of stations that are located in station-dense areas. Therefore, the value of $D^h = 55$ km can be considered as a suitable reference for the minimum allowed

30   D_i^h value. The procedure used for the D_i^h estimates is similar to the one described for the regional pseudo-background field. D_i^h is a weighted average as the one reported in Eq. (8) where $\overset{c}{\mathbf{x}}_i^b$ is replaced by the cth length scale, which is constant for all gridpoints. For the cth sub-region this length scale is set to the average distance between a station and its nearest 4 stations, provided that this distance is larger than $D^h = 55$ km, otherwise $D^h = 55$ km is used. In this way, the analysis in data-sparse

regions is the result of an interaction between a few (approximately four) stations. At the same time, we take advantage of data-dense areas to increase locally the effective resolution of the analysis without destroying the continuity of the field. Note that the use of extremely different D^h values between data-dense and sparse areas (i.e., with differences around one order of magnitude or more) would result in a rather confusing field to look at. In those cases it would be better to split the domain in

5   sub-domains and operate independently on them.

3.2 Precipitation

The multi-scale OI analyses are the results of successive approximations of the observations over a sequence of decreasing spatial scales that at station locations converge to the observed values.

The interpolation scheme is not applied directly to the RR values (the vector of the raw observed values adjusted for the

10 wind-induced under-catch is indicated as $\mathbf{y}^{\mathrm{rr}}$) but to their anomalies relative to a reference field of monthly precipitation (see Sec. 2, indicated with the abbreviation ref in the following). In addition, a Box-Cox transformation with power parameter set to 0.5 is used and the transformation is indicated with the function $g()$. A similar transformation has been suggested by Erdin et al. (2012), though in the context of combination of radar with gauge data. The ith element of $\mathbf{y}^o$ used in multi-scale OI is:

$$\mathbf{y}_i^o = g\left(\mathbf{y}_i^{\mathrm{rr}}/\mathbf{y}_i^{\mathrm{ref}}\right) \tag{9}$$

15 The analysis procedure can be written as:

$$\mathbf{x}^a = g^{-1}\left[\mathcal{M}_2 \circ \mathcal{M}_3 \circ \ldots \circ \mathcal{M}_l\left(\overset{l}{\mathbf{x}}^a\right)\right] \odot \mathbf{x}^{\mathrm{ref}} \tag{10}$$

where the three fundamental operations are: (1) the composition of several applications of the same statistical interpolation model down a hierarchy of spatial scales of $\{l\,\mathrm{km},\ldots,3\,\mathrm{km},2\,\mathrm{km}\}$, such that the results of a model application are used to initialize the successive one, $\circ$ stands for model composition and $\mathcal{M}_l$ stands for the application of the statistical model to the

20 length scale of l km, besides $\overset{l}{\mathbf{x}}^a$ is the average of the $\mathbf{y}^o$ elements; (2) the Box-Cox inverse-transformation $g^{-1}()$; (3) the elementwise multiplication of vectors $\odot$ so to transform the relative anomalies into RR values and at the same time include the effects of unresolved spatial scales. Ideally, the sequence of spatial scales to be used in Eq. (10) should be bounded between a very large scale (e.g, half the largest domain dimension) and a fine scale corresponding to the average distance between two stations in data-dense areas. The number of scales in between those two extremes is not critical for the final results, provided

25 that they are enough to guarantee a continuous analysis field in all situations. For seNorge_2018, we are using approximately 100 scales with a minimum of 2 km and a maximum of 1400 km. The sequence is unevenly spaced as the difference between two consecutive scales is somewhat proportional to their values.

The step-by-step description of the model $\overset{s}{\mathbf{x}}^a = \mathcal{M}_s\left(\overset{s+1}{\mathbf{x}}^a\right)$ for the arbitrarily length scale of s km is:

$$\overset{s}{\mathbf{x}}^b = \overset{s}{\mathbf{\Psi}}\overset{s+1}{\mathbf{x}}^a \tag{11}$$

30 $$\overset{s}{\mathbf{K}} = \overset{s}{\mathbf{G}}\left(\overset{s}{\mathbf{S}} + \varepsilon^2\overset{s}{\mathbf{I}}\right)^{-1} \tag{12}$$

$$\overset{s}{\mathbf{x}}^a = \overset{s}{\mathbf{x}}^b + \overset{s}{\mathbf{K}}\left(\mathbf{y}^o - \overset{s}{\mathbf{H}}\overset{s}{\mathbf{x}}^b\right) \tag{13}$$

In order to reduce the multi-scale OI computational expenses, the original 1 km grid is aggregated onto a new coarser grid, with aggregation factor equal to the integer value nearest to $s/2$ km. The aggregation groups several smaller rectangular boxes into a bigger one (e.g., if $s = 8$ km then 16 of the 1 km by 1 km boxes are aggregated into a single box measuring 4 km by 4 km). The analysis at scale $s + 1$ is used as the background for scale s in an OI scheme. The analysis values are transferred between

5   the two non-matching grid by the operator $\overset{s}{\mathbf{\Psi}}$ of Eq. (11), that is a bilinear interpolation mapping vectors on the $(s+1)$-grid to vectors on the s-grid. The observation operator $\overset{s}{\mathbf{H}}$ of Eq. (13) is also a bilinear interpolation transforming vectors on the s-grid to observation vectors. ε^2 is set 1, which means that observations and background are assumed the have the same error variances. The background error correlation matrices of Eq. (12) are defined on the basis of the correlation function ρ^R:

$$\rho^R (\mathbf{r}_j, \mathbf{r}_k) = f_d (\mathbf{r}_j, \mathbf{r}_k; s) \tag{14}$$

10   The Gaussian function f is defined in Eq. (4). Note that the length scales enter multi-scale OI of Eq. (10) through the correlation function of Eq. (14). An OI scheme such as the one presented in Eq. (11)- (14) is realizing a low-pass filter whose cut-off wavelength is approximately s km (Uboldi et al., 2008) so every iteration over a smaller spatial scale returns a field with more fine-scale details in it. For a given element of the observation vector, there may be a "critical" scale at which the background coincides exactly with the observed value such that its contribution to the innovation in Eq. (13) (i.e., the term in parenthesis)

15   is equal to zero and its analysis value would not change over the subsequent iterations. That critical scale is variable across the domain and depends on the local station density.

4   Results

The evaluation is based mostly on cross-validation exercises and comparison against the seNorge2 datasets of RR and TG. The cross-validation analyses (i.e., CV-analysis) is the analysis value at a station location obtained considering all the available

20   observations except the one measured at that location. The summary statistics of the following variables are used: CV-analysis residuals (i.e, CV-analyses minus observations); analysis residuals (i.e, analyses minus background); and innovations (i.e., background minus observations). At a generic station location, background and CV-analysis are independent from the observation, while the analysis has been computed using the observation. The CV-analysis residual distributions are used in place of the unknown analysis error distributions at gridpoints. For temperature, the comparison between the statistics of CV-analysis

25   residuals and innovations quantifies the improvement of the analysis over the pseudo-background at gridpoints. The statistics of analysis residuals reveal the filtering properties of the statistical interpolation at station locations that are related to the observation representativeness error (Lussana et al., 2010; Lorenc, 1986). The mean absolute error (MAE) and the root mean square error (RMSE) quantify the average mean absolute deviation and the spread, respectively, of the aforementioned variables. The fraction of errors (i.e., absolute deviations) greater than $3°\text{C}$ has been used as a measure of the tails of the distribution of

30   deviation values. Note that the threshold of $3°\text{C}$ is used in MET Norway's verification practice to define a significant deviation that undermine the user confidence in the forecast. In the case of temperature, we have used a leave-one-out cross-validation approach. For precipitation, the algorithm makes the leave-one-out cross-validation approach computationally too expensive.

Thus, for each day a random sample of 10% of the available stations have been reserved for cross-validation and not used in the interpolation.

4.1 Temperature

The main factors determining the quality of the temperature datasets are: the season of the year, the station density and the terrain complexity. We have also investigated the variations of the performances of our interpolation scheme between two different time periods, 1961-1990 (61-90) and 1991-2015 (91-15), and the evaluation scores are similar to the ones presented in the following for the whole 61-year time period.

In Figures 2-3 the TG, TX and TN evaluation results are shown for summer and winter, respectively, as a function of CV-IDI (Sec. 2). The background, analysis and CV-analysis are evaluated through the statistics of innovation, analysis residuals and CV-analysis residuals, respectively. For each of the four CV-ID classes (Sec. 2), the mean value of the score is displayed. The CV-analysis and background always score better in data-dense areas, as expected. On the contrary, the analysis may show the best results in data-sparse regions because an isolated observation constrains the analysis to fit its value almost exactly. For all variables, the spatial interpolation scheme generally performs better during summer than winter when small-scale processes (e.g., strong temperature inversion) are more frequent. The TG analysis error distribution at gridpoints, as estimated by CV-analysis residuals, shows that: during summer, the MAE is between $0.5°C$ and $1°C$ and its RMSE is also around $1°C$, errors larger than $3°C$ are unlikely even in data-sparse regions; during winter, MAE and RMSE double their values and the differences between data-dense and data-sparse regions are larger, the probability of having large errors is approximately 25% in data-sparse regions. The TX analysis error behave similarly to TG, it is worth remarking that the spatial interpolation method during summer performs better on TG than on TX, while in winter the opposite occurs. TN is the most challenging variable to represent, possibly because atmospheric processes at unresolved spatial scales occur more frequently for this variable than for the others. The distribution of the TN analysis error at gridpoints has a much larger spread than those of TG and TX, the TN RMSE is: between $1.5°C$ and $2°C$ in summer; up to approximately $4°C$ during winter in data-sparse regions. The tail of the TN distribution is also longer and large errors are more frequent than for the other daily temperatures. At the same time, the average bias (MAE) is also larger. Our results indicate that a denser station network would be needed to deliver TN products having the same quality as TG and TX.

In Figure 4 the TG dataset is compared with seNorge2 over two multi-year time periods for the winter season. In the other seasons the two datasets are much more similar, the patterns of deviations still follow those of Fig. 4 but the differences are almost always between $-2°C$ and $1°C$. The maps show the average daily difference between the analyses of the two datasets. The lateral and bottom panels are shown so to give a better overview of the extreme values. For most of the Norwegian mainland, seNorge_2018 is colder than seNorge2 and the larger differences are on the mountain tops and in the North, along the coast. The two periods show similar patterns, however for the 91-15 period there are some regions where seNorge_2018 is warmer than seNorge2, such as: the plateau in the North between Finland, Russia and Norway, along the border between Sweden and Norway, and in the mountains of Southern Norway. The seNorge_2018 spatial interpolation procedure builds upon the seNorge2 procedure and a number of modifications have been made, though keeping the scale-separation approach. In

seNorge_2018 a single function has been used to model the sub-regional vertical profile, instead of the three different functions used in seNorge2. At the same time, in seNorge_2018 the blending of sub-regional fields into a regional pseudo-background field is based on a much larger number of sub-regional fields. However, we believe that the variations having the most significant impacts on the differences shwon in Fig. 4 are in the OI settings. seNorge_2018 includes the land area fraction as an additional parameter in Eq. (7) and this improvement causes the differences along the coastline, where stations having significantly different land area fractions are less correlated. The evident difference between the two datasets is in the mountains, with warmer valley floors and colder ridges in seNorge_2018. In particular, the highest peaks of the Scandinavian mountains are on average up to $9°$C colder, while the valley floors are just a few degrees warmer. Those differences can be explained by (1) the reduced D^z value in seNorge_2018 ($D^z = 210$ m instead of $D^z = 600$ m as in seNorge2), which narrows the vertical layer where OI adjustments are effective, therefore temperatures in data-sparse high-altitude regions mostly coincides with background values; (2) the modified procedure for the pseudo-background calculation that realizes a smoother transition between neighbouring sub-regional pseudo-background fields. It should be mentioned also that there had been variations in the observational datasets used for the production of the two gridded datasets. Even though the data sources are the same for both seNorge datasets, seNorge2 is based on an observational dataset that has been produced in 2016, while seNorge_2018 benefits from the latest efforts in data collection and quality control made by MET and the ECA&D team.

In Figure 7, the expected percentage of large errors (i.e., analysis minus true temperature larger than $3°$C) at gridpoints is shown for TG, TX and TN during wintertime. This information completes the evaluation presented in Figures 2-3 because it shows the extensions on the grid of the regions characterized by the different data densities. The scatter-plots in the panels on the bottom left of each maps are different representations of the data in the bottom row of Fig. 3, where the points represents the percentage of large errors err at station locations as a function of CV-IDI. The lines in the scatter-plots are the best fit to the points of the Gaussian function:

$$\text{err}(x) = a \ \exp\left[-0.5\frac{(x-b)^2}{c^2}\right] \tag{15}$$

where x is the CV-IDI for the function fitting performed at station locations, then the IDI fields shown in Fig. 1 are used as x values to estimate err over the grid. The three parameters of the bell curve shape are: the value of the curve's peak a; the position of the center of the peak b, and c which controls the width of the bell. The Gaussian function provides reasonably good fits to the points, the relationship between the station density and the analysis quality is non-linear and the analysis performances decay much faster in data-sparse than in data-dense regions. The parameters values for the three variables are: TG $a = 703.012$, $b = -2.626$, $c = 1.123$; TX $a = 856.420$, $b = -2.924$, $c = 1.114$; TN $a = 448.39$, $b = -2.891$, $c = 1.431$. The probability of having a large analysis error is remarkably small over the domain for TX, while for TN such large errors are quite common. The situation for TG is somewhat in between those two extreme cases and large regions of the domain are unlikely to present large errors. Once again, it is evident that the worst performances occur in those regions characterized by complex terrain and low station density, such as the mountainous region between Norway and Sweden in the northern part of the domain.

4.2 Precipitation

The main factors determining the quality of the precipitation dataset are: the station density and the terrain complexity. The season of the year seems to have a smaller impact on the verification scores.

In Figure 5, the RR dataset is evaluated through the statistics of CV-analysis residuals. In general, the spatial interpolation performs significantly better for station-dense areas, then the performances degrade faster for data-sparse regions as shown by the MAE and RMSE for observations greater than 1 mm/day. The ability in distinguishing between precipitation and no-precipitation is shown by the Equitable Threat Score (ETS) with a threshold of 1 mm/day. In data dense areas the fraction of correct predictions, accounting for hits due to chance, is approximately 70%, while the ETS goes down to 0.4 in data-sparse regions. With respect to intense precipitation (i.e., observed values greater than 10 mm/day), the probability of having large errors (i.e, absolute deviations between CV-analysis and observation greater than 50% of the observed value) is less than 5% in data-dense areas and around 30% in data-sparse areas.

The close relationship between the terrain and the annual total precipitation is shown in Fig. 6, where values up to 4700 mm/year are reconstructed along the Norwegian coast. Figure 6 shows also the differences between seNorge_2018 and seNorge2 mean annual total precipitation in the period 1957-2015. On average, seNorge_2018 has significantly higher precipitation values than seNorge2, especially in data-sparse mountainous regions, and this is due to: i) the correction of rain-gauge data for the wind-induced under-catch, that increase the observed precipitation mostly during winter, and ii) the modified statistical interpolation scheme that uses more information in data-sparse regions, which often tend to increase the precipitation analysis when compared to seNorge2. As described in the seNorge2 paper Lussana et al. (2018a), this dataset is likely to underestimate precipitation so we have designed seNorge_2018 to returns higher precipitation values because this would better agree with the Norwegian water balance and eventually improve the results of hydrological simulations based on seNorge_2018 than for seNorge2. This last point deserves to be verified in the near future.

As for temperature in Sec. 4.1, also for precipitation the expected percentage of large errors over the grid is shown in Fig. 7. The three parameters of Eq. (15) for RR are: $a = 50.438$, $b = -0.676$, $c = 0.846$. The elevation is not considered in the RR IDI definition, so the map look rather different from the temperature maps. For data-dense regions, the expected percentage of large analysis errors for intense precipitation is less than 10%. With respect to the IDI calculation (Sec. 2), by setting $D = 10$ km in Eq. (4) (instead of $D = 50$ km as for temperature) we have imposed a fast transition between data-dense and data-sparse regions. For data-sparse regions the percentage of large errors is approximately 40%.

The evaluation carried out in Fig. 7 for both temperature and precipitation quantifies impacts on the analysis quality due to variations in the station distribution, thus providing a useful tool in the strategic planning of future observational networks.

5 Conclusions

seNorge_2018 provides 61-year (1957-2017) datasets of daily mean, maximum, minimum temperatures, and daily total precipitation over Norway and parts of Finland, Sweden and Russia. The plan at MET Norway is to update the historical dataset once a year, while at the same time provisional daily estimates for the current year are computed every day. MET Norway has

an open data policy and all the datasets, as well as most of the observations used in the calculations, are available for public download via its web services.

The observational datasets have been obtained through statistical methods that build upon our previous works. The interpolation schemes automatically adapts their settings to the local station density and this allows for an higher effective resolution in data-dense areas, while in data-sparse regions the analysis is always the estimate of at least a few stations.

The main factor determining the quality of the temperature analysis are: the season of the year, the station density and the terrain complexity. In the case of precipitation, those factors are: the station density and the terrain complexity. Because of the importance of the combination of station density and terrain, we have widely used the IDI concept in our evaluation.

The new seNorge_2018 shows significant differences when compared to its predecessor seNorge2, both for TG and especially for RR. While first qualitative evaluations indicate that this is an improvement, an indirect evaluation where seNorge_2018 would be used as the forcing data for snow- and hydrological modeling is needed to confirm this.

seNorge_2018 is the first MET Norway's observational dataset providing TX and TN from 1957. The temperature analysis has the largest errors during winter and the TN is the most challenging variable to represent. For TG and TX, large analysis errors are expected only in winter and limited to almost data-void areas, such as the mountain tops. TN may present large analysis errors more often than TG and TX and for larger portions of the domain, especially in mountainous regions.

The long-term average of the RR dataset is shaped by the monthly fields of a high-resolution numerical model, so to fill the gaps in data-sparse regions that are common, given the high variability of daily precipitation. The ability of the method to correctly distinguish between precipitation and no-precipitation depends critically on the station density. In the North, the sparser observational network is associated with a high occurrence of large analysis errors. The evaluation shows that large analysis errors are unlikely in the data-dense regions of Southern Norway, even for intense precipitation.

6   Code and data availability

The open-access datasets are available for public download at: daily total precipitation (DOI:https://doi.org/10.5281/zenodo.2082320 Lussana, 2018b) ; daily mean (DOI:https://doi.org/10.5281/zenodo.2023997 Lussana, 2018c) , maximum (DOI:https://doi.org/10.5281/zenodo.2559372
Lussana, 2018e) and minimum (DOI:https://doi.org/10.5281/zenodo.2559354 Lussana, 2018d) temperatures. seNorge_2018 is daily updated by MET Norway and the most recent data are available at http://thredds.met.no/thredds/catalog/senorge/ seNorge_2018/catalog.html. The spatial interpolation software is avalalble at (DOI:https://doi.org/10.5281/zenodo.2022479 Lussana, 2018a).

Acknowledgements.   This research has been partially funded by the Norwegian project "Felles aktiviteter NVE-MET tilknyttet nasjonal flom- og skredvarslingstjeneste".

References

Bremnes, J. B.: Probabilistic wind power forecasts using local quantile regression, Wind Energy, 7, 47–54, https://doi.org/10.1002/we.107, https://onlinelibrary.wiley.com/doi/abs/10.1002/we.107, 2004.

Cardinali, C., Pezzulli, S., and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, Quarterly Journal of the Royal Meteorological Society, 130, 2767–2786, https://doi.org/10.1256/qj.03.205, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.03.205, 2004.

Crespi, A., Lussana, C., Brunetti, M., Dobler, A., Maugeri, M., and Tveito, O. E.: High-resolution monthly precipitation climatologies over Norway (1981–2010): Joining numerical model data sets and in situ observations, International Journal of Climatology, 0, https://doi.org/10.1002/joc.5933, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5933, 2018.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/10.1256/qj.05.108, http://dx.doi.org/10.1256/qj.05.108, 2005.

Erdin, R., Frei, C., and Künsch, H. R.: Data transformation and uncertainty in geostatistical combination of radar and rain gauges, Journal of Hydrometeorology, 13, 1332–1346, 2012.

Frei, C.: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances, International Journal of Climatology, 34, 1585–1605, 2014.

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

Grasso, L. D.: The differentiation between grid spacing and resolution and their application to numerical modeling, Bulletin of the American Meteorological Society, 81, 579, 2000.

Horel, J. D. and Dong, X.: An evaluation of the distribution of Remote Automated Weather Stations (RAWS), Journal of Applied Meteorology and Climatology, 49, 1563–1578, 2010.

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.: Unified notation for data assimilation: operational, sequential and variational, Practice, 75, 181–189, 1997.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003.

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

Kotlarski, S., Szabó, P., Herrera, S., Räty, O., Keuler, K., Soares, P. M., Cardoso, R. M., Bosshard, T., Pagé, C., Boberg, F., Gutiérrez, J. M., Isotta, F. A., Jaczewski, A., Kreienkamp, F., Liniger, M. A., Lussana, C., and Pianko-Kluczyńska, K.: Observational uncertainty and regional climate model evaluation: A pan-European perspective, International Journal of Climatology, 0, https://doi.org/10.1002/joc.5249, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5249, 2017.

Lind, P., Lindstedt, D., Kjellström, E., and Jones, C.: Spatial and Temporal Characteristics of Summer Precipitation over Central Europe in a Suite of High-Resolution Climate Models, Journal of Climate, 29, 3501–3518, https://doi.org/10.1175/JCLI-D-15-0463.1, 2016.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 112, 1177–1194, https://doi.org/10.1002/qj.49711247414, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711247414, 1986.

Lussana, C.: BLISS - Bayesian statistical interpolation for spatial analysis, https://doi.org/10.5281/zenodo.2022479, the Norwegian Meteorological Institute, 2018a.

Lussana, C.: seNorge_2018 daily total precipitation amount 1957-2017, https://doi.org/10.5281/zenodo.2082320, the Norwegian Meteorological Institute, 2018b.

Lussana, C.: seNorge_2018 daily mean temperature 1957-2017, https://doi.org/10.5281/zenodo.2023997, the Norwegian Meteorological Institute, 2018c.

Lussana, C.: seNorge_2018 daily minimum temperature 1957-2017, https://doi.org/10.5281/zenodo.2559354, the Norwegian Meteorological Institute, 2018d.

Lussana, C.: seNorge_2018 daily maximum temperature 1957-2017, https://doi.org/10.5281/zenodo.2559372, the Norwegian Meteorological Institute, 2018e.

Lussana, C., Salvati, M., Pellegrini, U., and Uboldi, F.: Efficient high-resolution 3-D interpolation of meteorological variables for operational use, Advances in Science and Research, 3, 105–112, 2009.

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

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, Earth System Science Data, 10, 235–249, https://doi.org/10.5194/essd-10-235-2018, https://www.earth-syst-sci-data.net/10/235/2018/, 2018a.

Lussana, C., Tveito, O. E., and Uboldi, F.: Three-dimensional spatial interpolation of 2m temperature over Norway, Quarterly Journal of the Royal Meteorological Society, 144, 344–364, https://doi.org/10.1002/qj.3208, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3208, 2018b.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resources Research, 51, 2707–2723, 2015.

Masson, D. and Frei, C.: Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types, Hydrology and Earth System Sciences, 18, 4543, 2014.

Masson, D. and Frei, C.: Long-term variations and trends of mesoscale precipitation in the Alps: recalculation and update for 1901–2008, International Journal of Climatology, 36, 492–500, 2016.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, 2013.

Mohr, M.: New routines for gridding of temperature and precipitation observations for "seNorge. no", Met. no Report, 8, 2008, 2008.

Mohr, M.: Comparison of versions 1.1 and 1.0 of gridded temperature and precipitation data for Norway, Norwegian Meteorological Institute, met no note, 19, 2009.

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

Pielke, R. A.: Further Comments on "The Differentiation between Grid Spacing and Resolution and Their Application to Numerical Modeling", Bulletin of the American Meteorological Society, 82, 699–700, https://doi.org/10.1175/1520-0477(2001)082<0699:FCOTDB>2.3.CO;2, https://doi.org/10.1175/1520-0477(2001)082<0699:FCOTDB>2.3.CO;2, 2001.

Reistad, M., Breivik, Ø., Haakenstad, H., Aarnes, O. J., Furevik, B. R., and Bidlot, J.-R.: 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, 2011.

Sakov, P. and Bertino, L.: Relation between two common localisation methods for the EnKF, Computational Geosciences, 15, 225–237, 2011.

5 Saloranta, T.: Simulating snow maps for Norway: description and statistical evaluation of the seNorge snow model, The Cryosphere, 6, 1323–1337, 2012.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France Convective-Scale Operational Model, Monthly Weather Review, 139, 976–991, https://doi.org/10.1175/2010MWR3425.1, 2011.

Simmons, A. J., Berrisford, P., Dee, D. P., Hersbach, H., Hirahara, S., and Thépaut, J.-N.: A reassessment of temperature variations and 10 trends from global reanalyses and monthly surface climatological datasets, Quarterly Journal of the Royal Meteorological Society, 143, 101–119, https://doi.org/10.1002/qj.2949, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2949, 2017.

Skaugen, T. and Onof, C.: A rainfall-runoff model parameterized from GIS and runoff data, Hydrological Processes, 28, 4529–4542, 2014.

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

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

Walters, M. K.: commentary and analysis, Bulletin of the American Meteorological Society, 81, 2475–2479, https://doi.org/10.1175/1520-0477(2000)081<2475:CAACOT>2.3.CO;2, https://doi.org/10.1175/1520-0477(2000)081<2475:CAACOT>2.3.CO;2, 2000.

Wolff, M. A., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T., and Brækkan, R.: Derivation of a new continuous adjustment 20 function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study, Hydrology and Earth System Sciences, 19, 951–967, https://doi.org/10.5194/hess-19-951-2015, https://www.hydrol-earth-syst-sci.net/19/951/2015/, 2015.
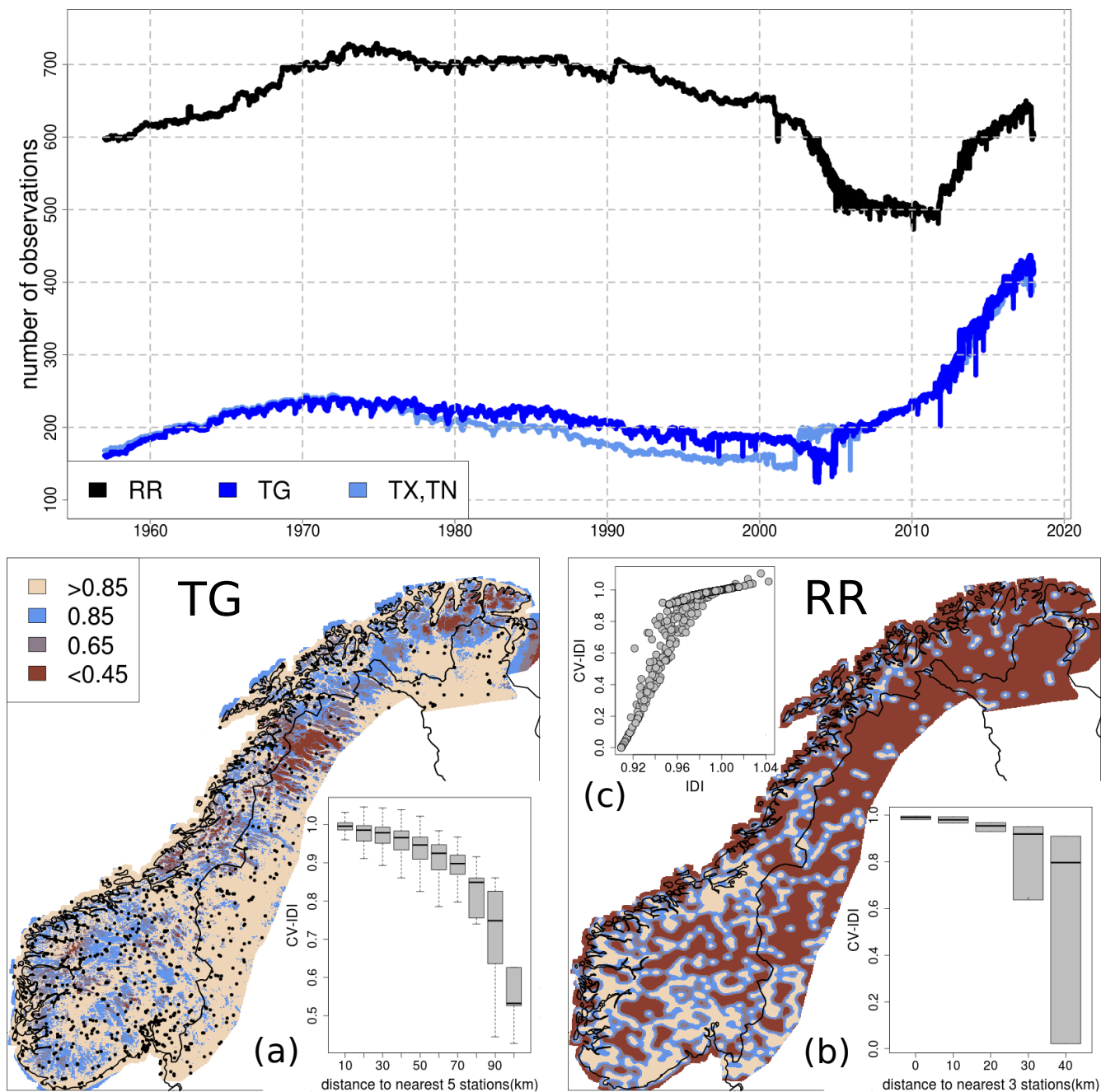
Figure 1. The observational network for the four variables: RR, TG, TX and TN. A detailed description is given in Sec. 2. The top panel shows the time series for the number of available observations over the Norwegian mainland. On the bottom panels, TG (left) and RR (right) observational networks are shown (TX and TN are similar to TG). The displayed fields are the IDI at gridpoints. Panels *a* (TG) and *b* (RR) show the CV-IDI as a function of the distance to the nearest stations. In panel *c*, the two parameters measuring the local station influence on the analysis, CV-IDI and IDI, are compared.
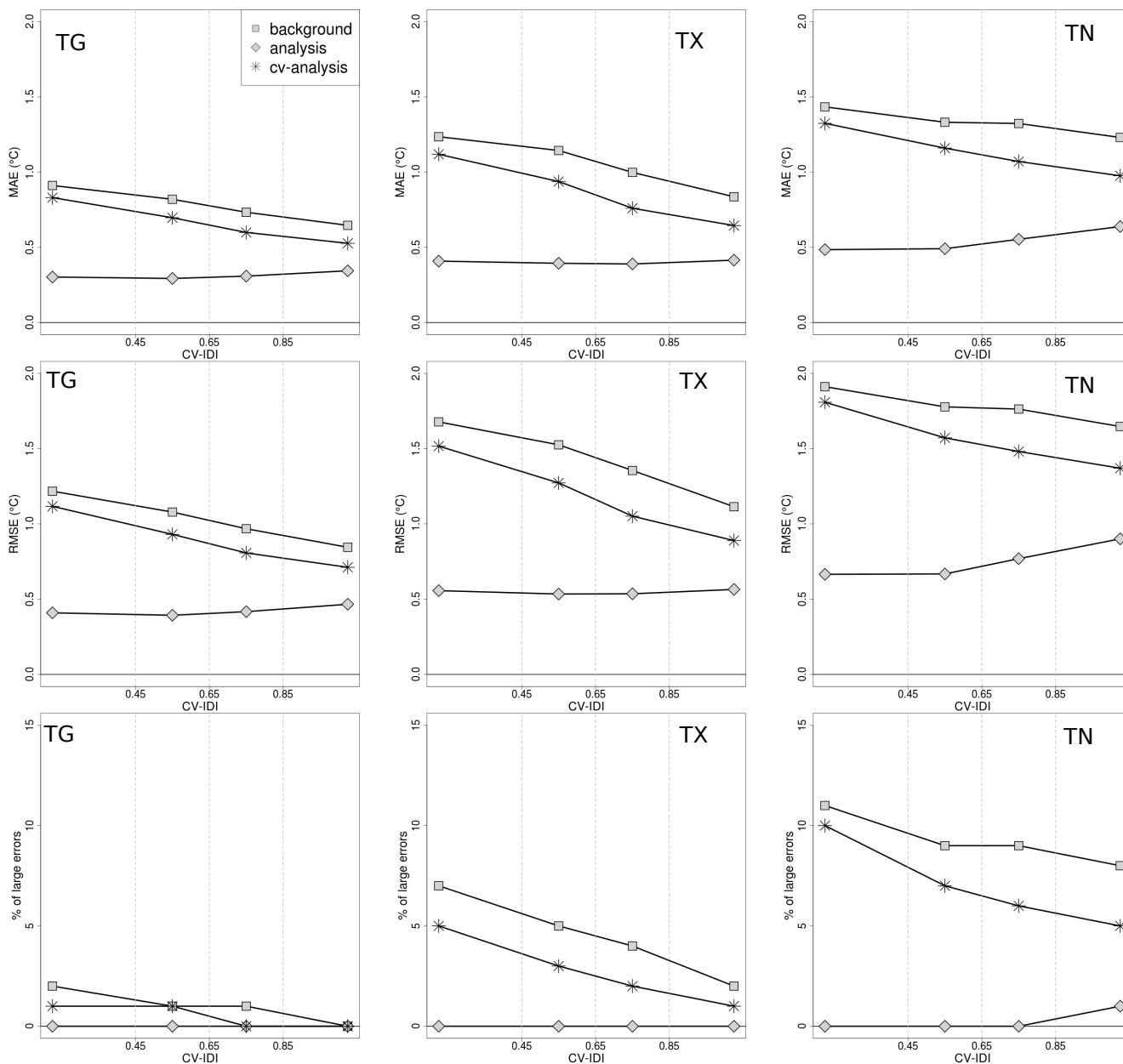
Figure 2. TG, TX and TN, verification scores as a function of CV-IDI based on the summer seasons (June-July-August) within the 61-year time period 1957-2017. The verification scores are based: for the analysis on the analysis residuals; for the CV-analysis on the CV-analysis residuals; for the background on the innovation. On the top row, the mean absolute error (MAE). In the middle, the root mean square error (RMSE). On the bottom row, the percentage of large errors. A large error is defined as the absolute value of innovation or residual larger than 3°C.
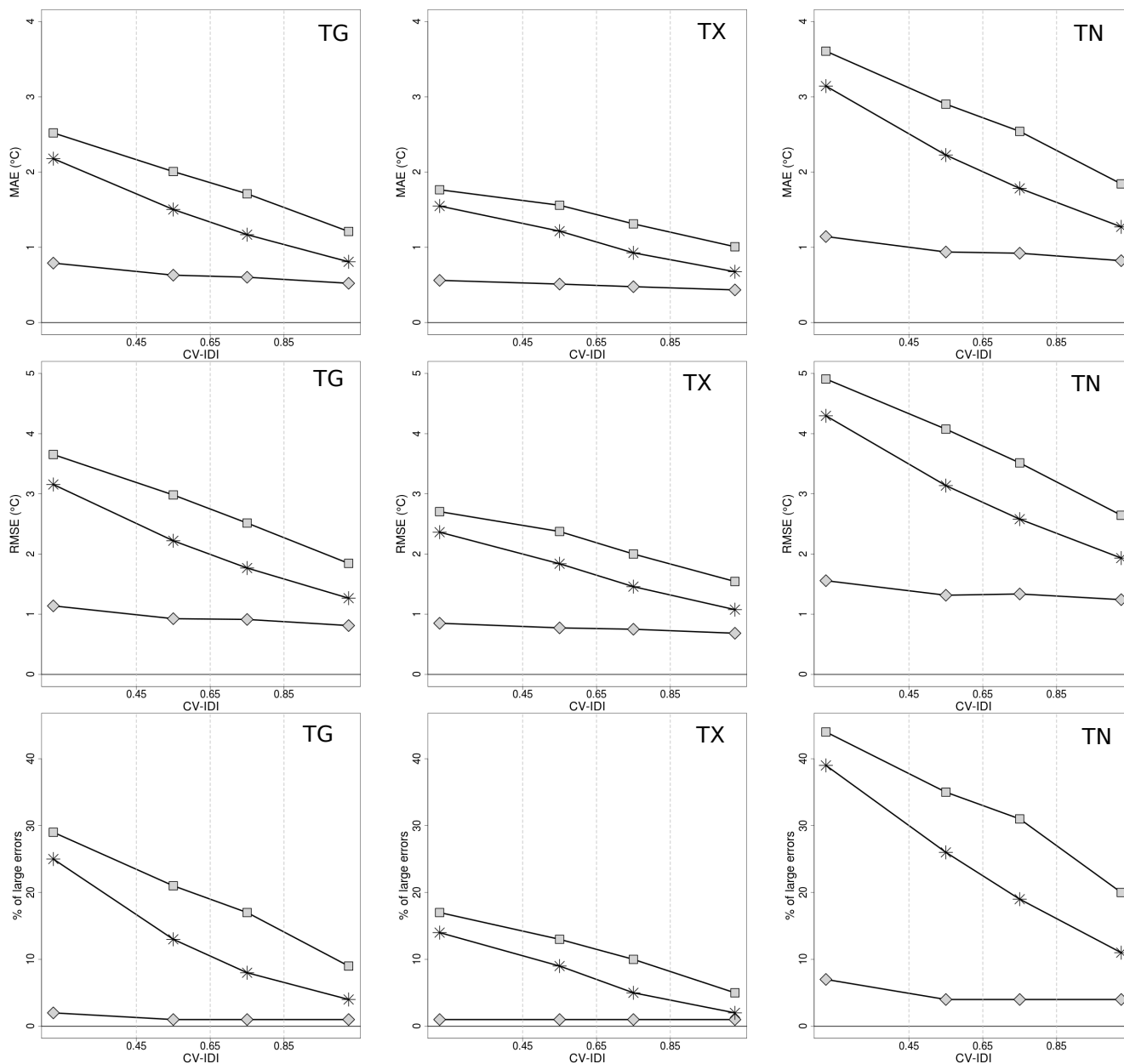
Figure 3. TG, TX and TN, verification scores as a function of CV-IDI based on winter seasons (December-January-February) within the 61-year time period 1957-2017. See Fig. 2.
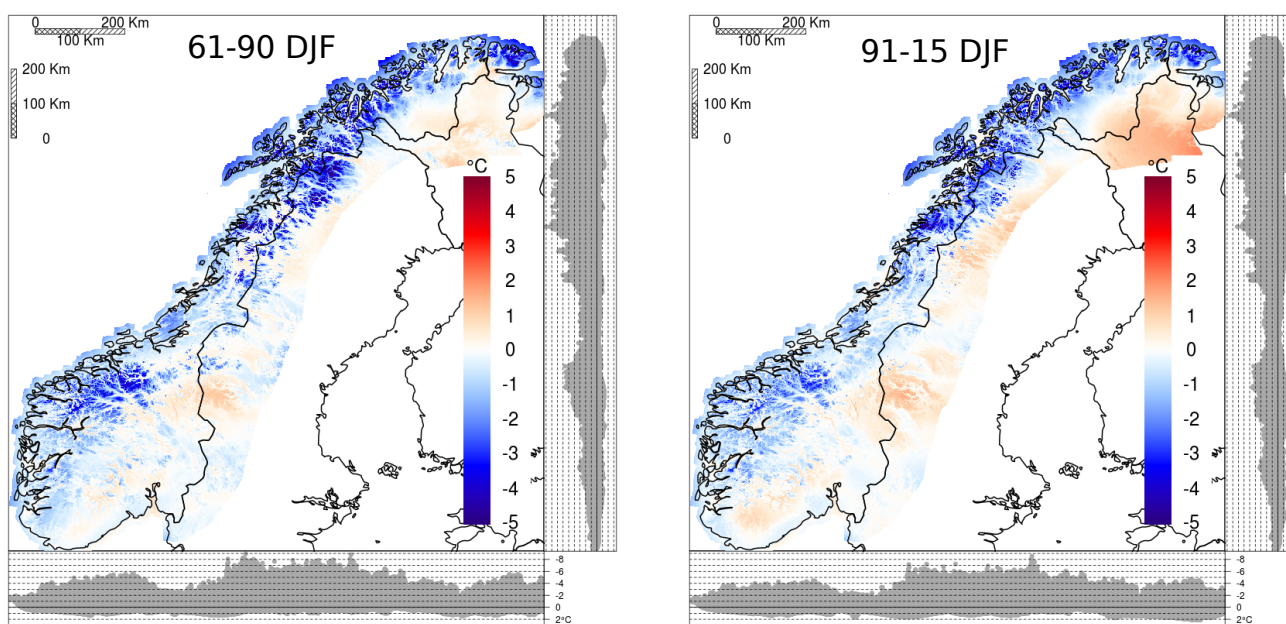
Figure 4. mean TG difference between seNorge_2018 and seNorge2 based on daily analysis in December, January and February. On the left panel, the 30-year period 1961-1990 is considered. On the right, the mean is based on the 25-year period 1991-2015. The lateral and bottom panels in both graphs show the projection of the differences on the y- and the x- axis, respectively.
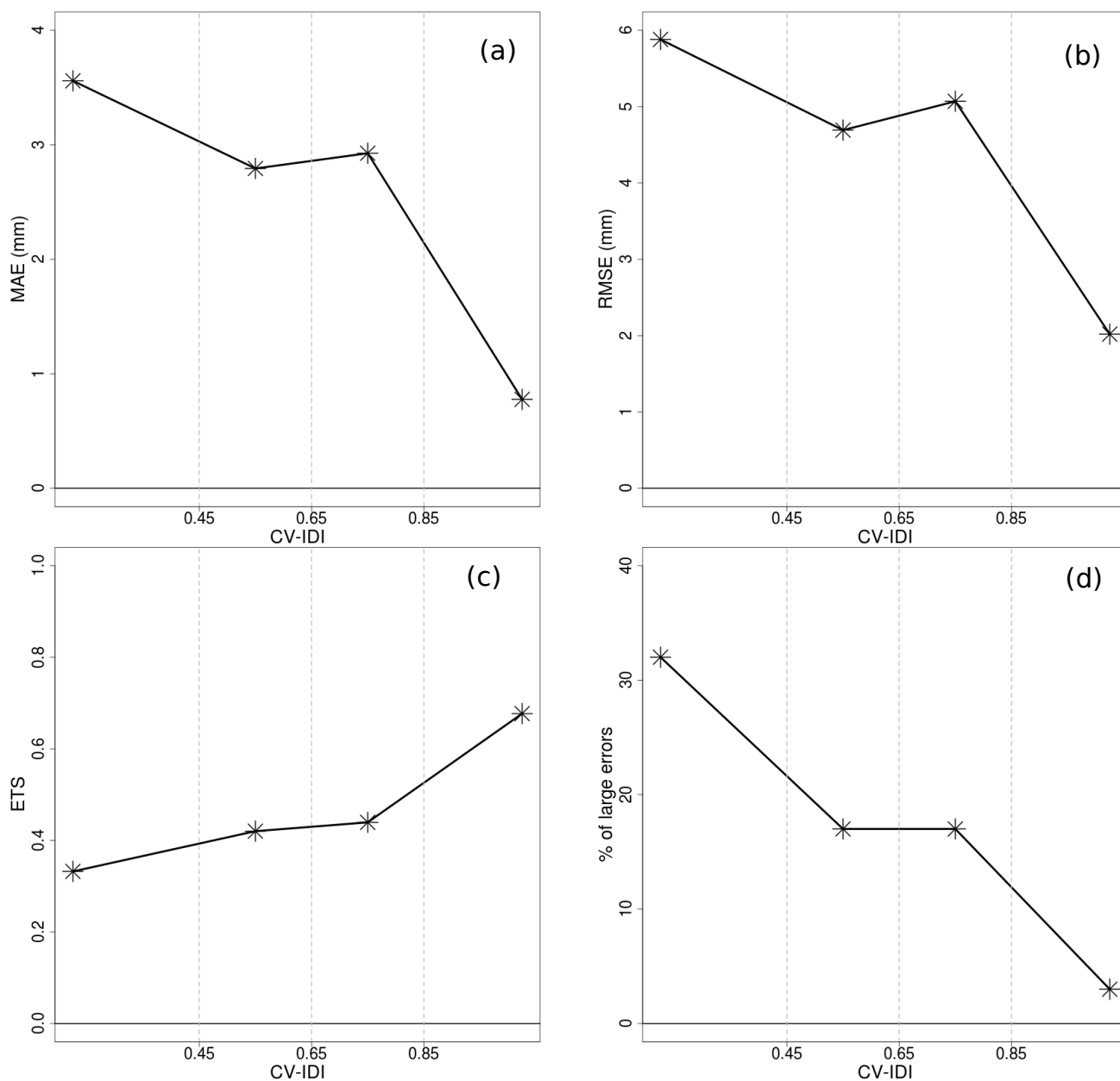
Figure 5. RR, CV-analysis scores as a function of CV-IDI based on CV-analysis residuals in the 61-year time period 1957-2017. Panel a, Mean Absolute Error considering only observations greater than 1 mm/day. Panel b, Root-Mean-Squared Error considering only observations greater than 1 mm/day. Panel c, Equitable Threat Score with threshold equals to 1 mm/day. Panel d, Percentage of large errors in case of intense precipitation (i.e., greater than 10 mm/day). A large error is defined as the absolute value of a CV-analysis residual larger than 50% of the observed value.
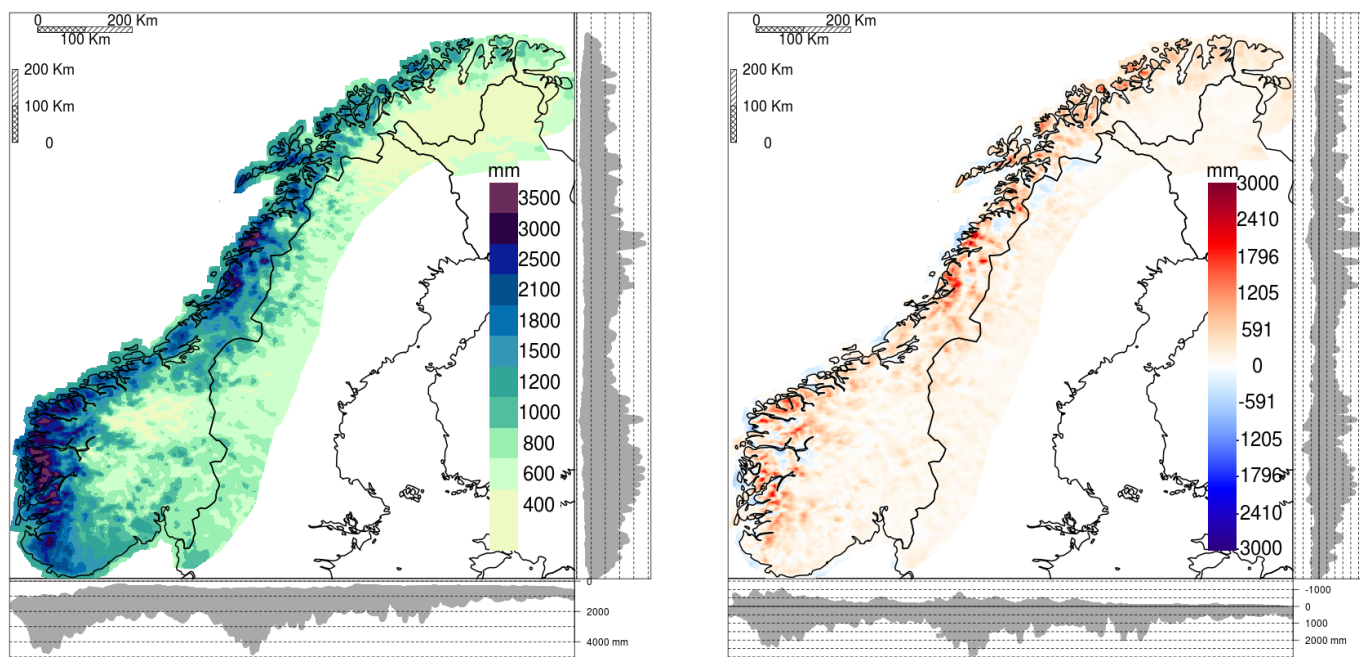
Figure 6. RR annual total precipitation. On the left, the mean annual total precipitation based on the 61-year period 1957-2017: the lower precipitation class includes values smaller than 400 mm; the upper precipitation class values between 3500 mm and 4700 mm. On the right, mean annual total precipitation difference between seNorge_2018 and seNorge2 based on the 51-year period 1957-2015. On each graph, the lateral and bottom panels show the projection of the differences on the y- and the x- axis, respectively.
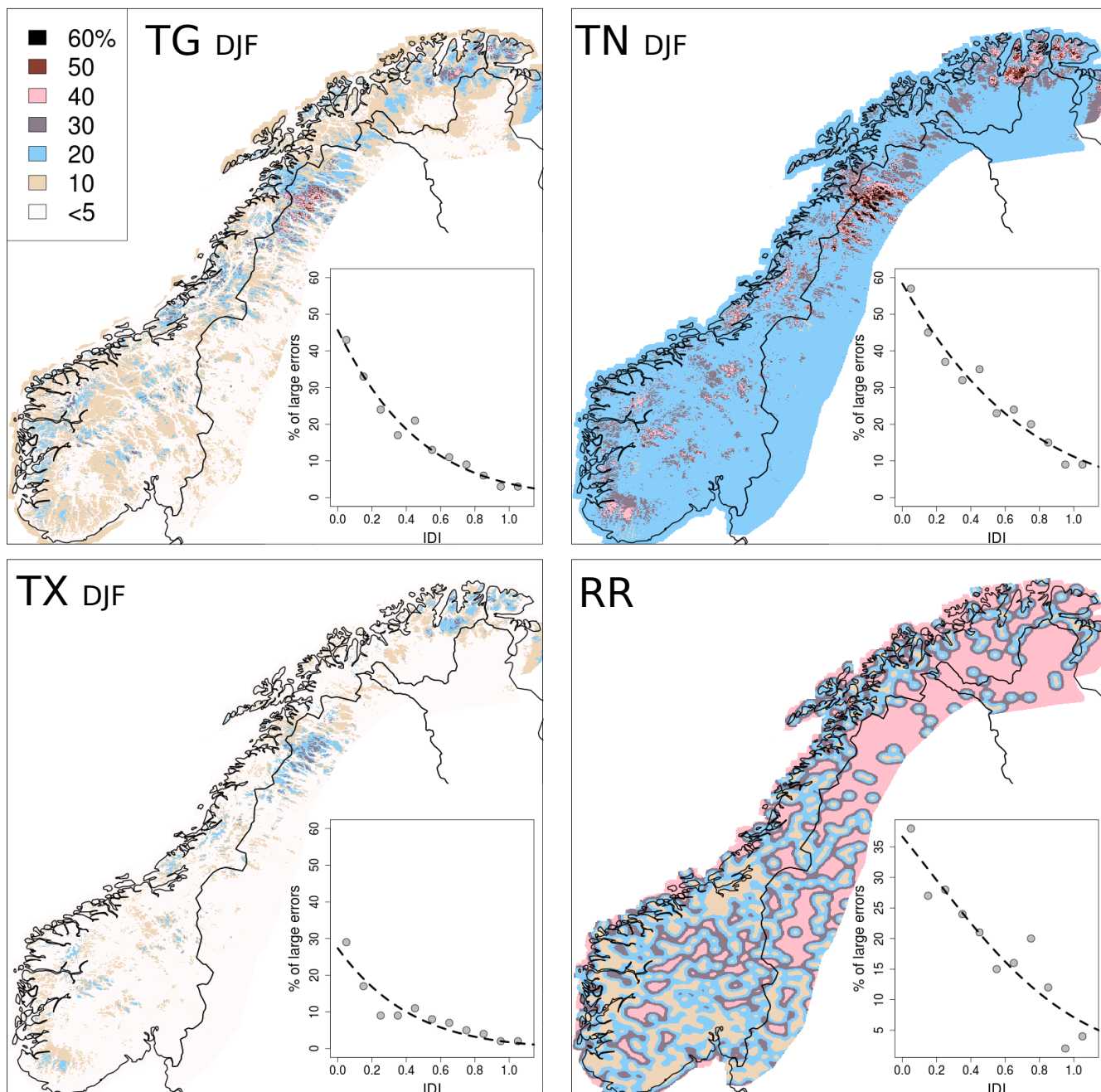
Figure 7. Expected percentage of large errors on the grid. The colour scale is the same for all the maps. Wintertime (DJF) temperatures are considered, large errors are defined as deviations between analysis and unknown truth larger than $3°$C. All precipitation data has been considered, large errors are deviations between analysis and unknown truth larger than 50% when the analysis value is greater than 10 mm/day. The insets show the relation between IDI and: (dots) the percentage of large errors observed at station locations (on the x-axis, CV-IDI instead of IDI); (line) the best-fit function used to infer the expected percentage of large errors at gridpoints.

Table 1. TG annual statistics: "n" is the average number of stations; "d" (km) is the average distance between a station and its nearest third station; D^h (km), D^z (m), σ_b^2 $((^\circ C)^2)$, σ_o^2 $((^\circ C)^2)$ are the optimal values given the 1-year innovation statistics and the constraint $\sigma_o^2/\sigma_b^2 = 0.5$.

year	n	d	D^h	D^z	σ_b^2	σ_o^2
1960	398	55	60	206	2.24	1.12
1970	669	42	45	217	1.86	0.93
1980	639	42	58	201	2.45	1.22
1990	600	44	57	202	1.33	0.66
2000	627	44	55	206	1.28	0.64
2010	639	45	52	206	2.45	1.23

Table 2. TX annual statistics (see Table. 1).

year	n	d	D^h	D^z	σ_b^2	σ_o^2
1960	395	55	56	207	2.09	1.05
1970	669	42	57	201	1.67	0.84
1980	616	45	37	216	2.09	1.05
1990	563	47	55	206	1.40	0.70
2000	596	46	56	210	1.32	0.66
2010	638	45	57	206	2.09	1.04

Table 3. TN annual statistics (see Table. 1).

year	n	d	D^h	D^z	σ_b^2	σ_o^2
1960	396	55	50	217	4.42	2.21
1970	670	42	53	222	3.80	1.90
1980	615	45	62	211	4.58	2.29
1990	560	47	52	210	2.88	1.44
2000	596	46	51	210	2.99	1.49
2010	637	45	64	212	4.70	2.35