Spatial interpolation of two-metre temperature over Norway based on the combination of numerical weather prediction ensembles and *in situ* observations

C. Lussana[1] | I. A. Seierstad[1] | T. N. Nipen[1] | L. Cantarello[2]

[1]The Norwegian Meteorological Institute, Oslo, Norway

[2]School of Mathematics, University of Leeds, Leeds, UK

Correspondence
Cristian Lussana, The Norwegian Meteorological Institute, Postboks 43 Blindern, 0313 Oslo, Norway.
Email: cristian.lussana@met.no

Abstract

Accurate hourly two-metre temperature gridded fields available in near real-time are valuable products for numerous applications, such as civil protection and energy production planning. An analysis ensemble of temperature is obtained from the combination of a numerical weather prediction ensemble (background) and *in situ* observations. At the core of the flow-dependent spatial interpolation method lies the analysis step of the local ensemble transform Kalman filter (LETKF). A scaling factor and a localization procedure have been added to correct for deficiencies of the background. Each observation is characterized by its own representativeness, which is allowed to vary in time. We call the method described here an Ensemble-based Statistical Interpolation (EnSI) scheme for spatial analysis and it has been integrated into the operational post-processing systems in use at the Norwegian Meteorological Institute (MET Norway). The benefits of the analysis are assessed over a 1-year time period (July 2017–July 2018) and a case-study is presented for a challenging situation over complex terrain. EnSI gives more accurate results than an interpolation method based exclusively on observations. The analysis ensemble provides a more informative representation of the uncertainty than a spatial analysis based on a single-field background. EnSI reduces the number of large prediction errors in the analysis compared to the background by almost 50%, reduces the ensemble spread and increases its accuracy.

KEYWORDS

Kalman filter, Norway, numerical weather prediction ensembles, spatial interpolation, two-metre temperature, weather stations

1 | INTRODUCTION

Near-surface fields of temperature are among the most widely used products made available by national meteorological services. In this article, we consider the representation of instantaneous two-metre temperature sampled at an hourly rate (or *hourly temperature*) either for the recent past or as soon as the temperature observations from a network of automatic weather stations (i.e. *in situ* measurements) become available. In Norway, near-surface variables available on a

real-time basis are used for instance by national civil protection authorities for the monitoring of floods, avalanches and landslides (Saloranta, 2012; Skaugen and Onof, 2014; Magnusson *et al.*, 2015) and energy companies for the planning of renewable energy production. Real-time gridded temperature fields are also important in public weather forecasting to represent current conditions. For example, the Norwegian Meteorological Institute (MET Norway) uses the gridded datasets of hourly temperatures on the weather forecast portal Yr (https://www.yr.no). Our experience is that the estimated temperature for the current time is an important parameter for the users and its accuracy affects their confidence in the products.

This study deals with spatial interpolation because observations are used to estimate temperature values at unobserved points on a grid. The main idea of this article is to combine two-metre hourly temperature gridded fields derived from numerical model output with the corresponding *in situ* observations to achieve a high-resolution product that (a) can serve real-time applications, (b) is as accurate as observational gridded datasets in data-dense regions, and (c) is as accurate as numerical model output in data-sparse (or data-void) regions. Numerical weather prediction (NWP) local area models are operated by national weather services to serve real-time applications so the use of their output in our work is a natural choice. In addition, NWP models often provide ensemble output. At any observation time, our spatial interpolation scheme considers as its background the two-metre temperature ensemble forecast valid at that observation time (i.e. just 1 h) and returns an analysis ensemble adjusted by taking into account the observations.

The core of our statistical interpolation is based on the analysis step of the filtering problem typical of the state-estimation theory. The prediction step is ignored because it is not applicable to our problem. An excellent overview of linear filters and how they are usually formulated in atmospheric sciences can be found in the recent review paper by Carrassi *et al.* (2018). A brief overview of the relevant literature for our work begins with the original papers on the Kalman filter (Kalman, 1960; Kalman and Bucy, 1961) and the traditional optimal interpolation (OI: Gandin and Hardin, 1965) that was developed to combine a climatological background with meteorological observations in the context of objective analysis. OI has been used for decades as a data assimilation technique in NWP (Lorenc, 1986; Ide *et al.*, 1997) and in oceanography (Kaplan *et al.*, 1997) to combine a model-derived background with the observations. OI replaces the Kalman filter background error covariance matrix with a convenient and constant-in-time approximation. The background error covariance matrix has usually been modelled with analytical functions, such as the ones described by Gaspari and Cohn (1999). An adaptation of OI to statistical interpolation aimed at the production of observational temperature

datasets has been described by Uboldi *et al.* (2008) and it has been subsequently used for daily mean temperature over Norway (Lussana *et al.*, 2018b). Equivalent methodologies based on Kriging have been used by Krähenmann *et al.* (2011), Frei (2014), Hiebl and Frei (2016) and Brinckmann *et al.* (2016) to obtain daily temperature fields across Europe. As pointed out by Amezcua and Leeuwen (2014), the analysis step of the Kalman filter is optimal when "(a) the distribution of the background is Gaussian, (b) state variables and observations are related via a linear operator, and (c) the observational error is of additive nature and has a Gaussian distribution". These conditions are generally considered valid for two-metre temperature, as the widespread application of statistical techniques based on similar assumptions demonstrate (Haylock *et al.*, 2008; Frei, 2014). OI has been used for the interpolation of several other variables, both within the context of meteorology (Lussana *et al.*, 2009) and outside it (e.g. leaf area index: Gu *et al.*, 2006). Examples of OI applications to the combination of numerical model output and independent (i.e. not used in the data assimilation cycle) *in situ* observations have been described for example by: Mahfouf *et al.* (2007) presenting the Canadian Precipitation Analysis (CaPA); Soci *et al.* (2016) combining precipitation from regional reanalyses and *in situ* observations from gauges; and more recently Crespi *et al.* (2019) demonstrating the benefits of the combination of model data and *in situ* observations for the reconstruction of monthly precipitation climatologies in Norway. In the context of the European project Uncertainties in Ensemble of Regional Reanalysis (UERRA, uerra.eu), the work presented by Soci *et al.* (2016) has been applied to two-metre temperature, among the other variables, and the gridded fields have been used for hydrological simulations. The spatial interpolation methodology developed in UERRA is now operated as a service by the Copernicus Climate Change Service (https://climate.copernicus.eu), thus confirming that the combination of model data and *in situ* observations is a widely supported approach in atmospheric sciences.

The original contribution of our study compared to the aforementioned methods is that we consider both the background and the analysis as ensembles. Ensemble datasets provide an ideal way to communicate interpolation uncertainty. NWP output is often available as an ensemble (e.g. Frogner *et al.*, 2019). Observational gridded datasets begin to be available in that form too, for instance the newest E-OBS version is now available as an ensemble dataset (Cornes *et al.*, 2018), as is the high-resolution spatial interpolation of precipitation in the Alps described by Frei and Isotta (2019). We call the method described here an Ensemble-based Statistical Interpolation (EnSI) scheme developed for spatial analysis. The analysis steps of the Ensemble Kalman filter (Evensen, 2003) and of the Local Ensemble Transform Kalman filter (LETKF: Hunt *et al.*, 2007) have been used to transform the background into the analysis ensemble. The spatial interpolation scheme

is flow-dependent as the forecast-error covariances are shaped by the background ensemble. A scaling factor has been introduced to adjust for deficiencies of the background ensemble. The *in situ* observations have been characterized individually by a station-dependent representativeness error that is also varying in time.

Statistical spatial interpolation of near-surface field often borrows methods from data assimilation. Literature on data assimilation is vast, and here we may refer to the classical books by Daley (1993) and Kalnay (2003). As an example of intersection between the two, the calculation of chapter 5 of Kalnay (2003) can be used to derive the OI scheme used by Uboldi *et al.* (2008). In this regard, it is worth spending a few words on the terminology used in spatial interpolation. A reader more familiar with data assimilation might get confused by our use of the word "analysis". In classical linear filter theory (Jazwinski, 2007), the analysis is the best estimate of the unknown true state and it is obtained combining a first-guess, or background field, with the observations. In data assimilation, the word analysis also indicates the process aimed at obtaining the initial conditions to run a numerical model, making the analysis a prerequisite for a weather forecast. In this article, our point of view is closer to that of the classical linear theory. The state vector contains a single meteorological variable (i.e. two-metre temperature) over a two-dimensional grid in terrain-following coordinates. The analysis is meant to be an estimate of the true state of two-metre hourly temperature based on the combination of model data and observations.

The structure of the article is as follows. Section 2 describes our spatial interpolation method. Then, the method is implemented and validated using MET Norway's products. Section 3 presents the NWP model and the observations used. The evaluation is described in section 4.

2 | ENSI, ENSEMBLE-BASED STATISTICAL INTERPOLATION FOR SPATIAL ANALYSIS

The mathematical notation used hereinafter is introduced in appendix A. The spatial interpolation scheme is depicted in Figure 1. The graph illustrates how the spatial interpolation method works for three consecutive hours. The unknown trajectory of the true temperature state is represented by the blue line. The analyses are the estimates of the true state at the observation time steps obtained by combining the background ensemble and the observations with a scheme based on the Kalman filter analysis scheme. In accordance with the linear filter theory (Jazwinski, 2007), the errors are assumed to be Gaussian and unbiased. Therefore, the error PDFs are characterized by the covariance matrices only, because the error mean values are set to zero. The first guesses from the background are the ensemble means $\mathbf{x}^b$ (green triangles). The
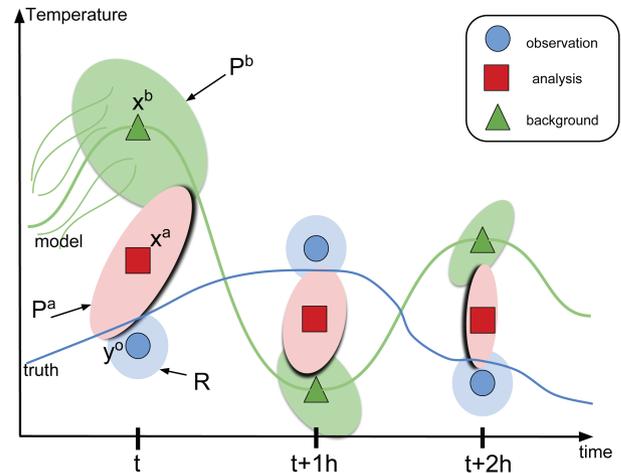
FIGURE 1 Sketch of the spatial interpolation scheme (based on Carrassi *et al.* (2018)). The truth is the blue line. The background is the green line. To avoid confusion, the background ensemble members are only initially indicated with several green lines, then only the ensemble mean is shown. Ellipses represent the error covariance matrices

observed values $\mathbf{y}^o$ (blue circles) are more accurate estimates of the truth than the background, as shown by the smaller areas (i.e. uncertainty) associated to the observation error covariance matrices $\mathbf{R}$ compared to the background error covariance matrices $\mathbf{P}^b$. At a fixed observation time step, the best estimate of the truth is the analysis ensemble mean $\mathbf{x}^a$ (red squares), which is derived from the analysis ensemble $\mathbf{X}^a$. The analysis constitutes a more accurate and precise estimate of the truth than the background because $\mathbf{x}^a$ is closer to the truth than $\mathbf{x}^b$ and its uncertainty, represented by the analysis error covariance matrix $\mathbf{P}^a$, is smaller than the background uncertainty. Figure 1 has been inspired by similar figures reported on the review paper on data assimilation by Carrassi *et al.* (2018). Unlike in data assimilation, in our spatial interpolation method the background is not influenced by the previous analysis steps of the Kalman filter and, as a consequence, the analyses at different hours are independent from each other. Eventually, the output of our scheme is the analysis ensemble $\mathbf{X}^a$, which allows us to compute both $\mathbf{x}^a$ and $\mathbf{P}^a = [1/(k-1)]\mathbf{A}^a\mathbf{A}^{a\mathrm{T}}$.

The key point in using a background ensemble is that a flow-dependent $\mathbf{P}^b$ can be included in our scheme. Because of the limited number of ensemble members, $\mathbf{P}^b$ may contain spurious long-distance correlations and to deal with this issue we have: (a) implemented a gridpoint-by-gridpoint analysis scheme based on Hunt *et al.* (2007), where for each grid point a local domain is defined and only the nearest observations are considered, and (b) introduced an R localization technique (Greybush *et al.*, 2011), where the $\mathbf{R}$ elements are multiplied by a distance-dependent function. The localization used is indicated as "local analysis" by Sakov and Bertino (2011) and they emphasize that for such a method "the impact of observations close to the boundary of the local domain is reduced by artificially increasing its error variance".

Considering the generic ith grid point, the local domain includes only the nearest q observations ($q < p$). As in Sakov and Bertino (2011), upper accents have been used to denote the local version of a variable. The local observation q-vector is $\overset{i}{\mathbf{y}}{}^o$ and its $q \times q$ error covariance matrix is $\overset{i}{\mathbf{R}}$. $\mathbf{P}^b$ is written as:

$$\mathbf{P}^b = \Delta(k-1)^{-1}\mathbf{A}^b\mathbf{A}^{b\mathrm{T}}, \tag{1}$$

where an ensemble scaling factor Δ is used to adjust for under- or over-estimation due to deficiencies in the background ensemble. We opted for a constant Δ that has been manually optimized (see section 4.1.2). A time- and location-dependent ensemble scaling would indeed be desirable and it is considered for future developments. Δ plays a similar role as the covariance inflation factor used in data assimilation (Li *et al.*, 2009; Miyoshi, 2011). Note that in order to get $\mathbf{X}^a$ it is not required to explicitly compute the rather big $m \times m$ matrix $\mathbf{P}^b$. The analysis ensemble mean at the grid point and the (local) error covariance matrix are obtained as per the analysis step of the standard Kalman filter (Kalnay, 2003):

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \overset{i}{\mathbf{K}}_{i,:}(\overset{i}{\mathbf{y}}{}^o - \overset{i}{\mathbf{H}}\mathbf{x}^b); \tag{2}$$

$$\overset{i}{\mathbf{P}}{}^a = (\mathbf{I} - \overset{i}{\mathbf{K}}\overset{i}{\mathbf{H}})\mathbf{P}^b. \tag{3}$$

$\mathbf{I}$ is the $m \times m$ identity matrix. $\overset{i}{\mathbf{H}}$ is the linear observation operator mapping m-vectors into q-vectors (when applied to a matrix it is intended to be applied separately to each of its columns); it consists of a nearest-neighbour interpolation with an adjustment for elevation differences between stations and their nearest grid points using a constant near-surface lapse rate of $-6.5\,°$C/km (i.e. when using the International Standard Atmosphere defined by the International Civil Aviation Organization, the temperature decreases at a constant rate of $6.5\,°$C per km; such an approximation is reasonable also near the Earth's surface (Brunetti *et al.*, 2014)). $\overset{i}{\mathbf{P}}{}^a$ approximates the actual analysis error covariance matrix only in the surroundings of the ith grid point, in the same way as $\mathbf{P}^b$ does for the background error. $\overset{i}{\mathbf{K}}$ is the local gain matrix and by substituting Equation 1 in the expression for the standard Kalman gain matrix $\overset{i}{\mathbf{K}} = \mathbf{P}^b\overset{i}{\mathbf{H}}{}^{\mathrm{T}}\left(\overset{i}{\mathbf{H}}\mathbf{P}^b\overset{i}{\mathbf{H}}{}^{\mathrm{T}} + \overset{i}{\mathbf{R}}\right)^{-1}$, it is possible to write:

$$\overset{i}{\mathbf{K}}_{i,:} = \mathbf{A}^b_{i,:}\overset{i}{\Gamma}{}^a(\overset{i}{\mathbf{H}\mathbf{A}}{}^b)^{\mathrm{T}}\overset{i}{\mathbf{R}}{}^{-1}; \tag{4}$$

$$\overset{i}{\Gamma}{}^a = [\Delta^{-1}(k-1)\mathbf{I} + (\overset{i}{\mathbf{H}\mathbf{A}}{}^b)^{\mathrm{T}}\overset{i}{\mathbf{R}}{}^{-1}(\overset{i}{\mathbf{H}\mathbf{A}}{}^b)]^{-1}. \tag{5}$$

$\mathbf{I}$ is the $k \times k$ identity matrix. $\overset{i}{\mathbf{H}\mathbf{A}}{}^b$ is the $q \times k$ matrix of the background perturbations evaluated at the q station locations.

The analysis ensemble at the ith grid point $\mathbf{X}^a_{i,:}$ is obtained from the equivalence of the following two expressions for $\overset{i}{\mathbf{P}}{}^a$:

$$\overset{i}{\mathbf{P}}{}^a = (k-1)^{-1}\overset{i}{\mathbf{A}}{}^a\overset{i}{\mathbf{A}}{}^{a\mathrm{T}}, \tag{6}$$

$$\overset{i}{\mathbf{P}}{}^a = \mathbf{A}^b\overset{i}{\Gamma}{}^a\mathbf{A}^{b\mathrm{T}}. \tag{7}$$

Equation 7 is derived in Hunt *et al.* (2007) by substituting Equation 1 into Equation 3. Note that $\overset{i}{\Gamma}{}^a$ is the projection of $\overset{i}{\mathbf{P}}{}^a$ onto the ensemble-space of dimension k by the linear transformations $\mathbf{A}^b$ and this enables us to obtain the analysis ensemble members. Equations 6 and 7 lead to:

$$\overset{i}{\mathbf{A}}{}^a_{i,:} = \mathbf{A}^b_{i,:}[(k-1)\overset{i}{\Gamma}{}^a]^{1/2}, \tag{8}$$

such that the analysis ensemble is:

$$\mathbf{X}^a_{i,:} = \mathbf{x}^a_i + \overset{i}{\mathbf{A}}{}^a_{i,:}. \tag{9}$$

As $\mathbf{P}^b$ does not explicitly appear in the equations to obtain the analysis ensemble (i.e. Equations 2, 4, 5, 8), it is natural to apply an R localization technique. In the assumption of independent observation errors, the $q \times q$ matrix $\overset{i}{\mathbf{R}}$ is diagonal and each of its elements has been factorized into the product of two main factors: (a) an observation-dependent error variance, and (b) the localization function. The first factor is defined as the product of σ_o^2, that is the error variance characterizing the observational network as a whole, and an observation-dependent correction factor (e.g. c_j indicates the correction factor for the jth observation). σ_o^2 is constant in time and space and its value is optimized as described in section 4.1.1. The correction factors are introduced so as to give more weight to representative observations, as described in appendix B. The second factor, the localization function ρ, returns a value for two arbitrary points, e.g. the ith grid point and the jth observation location identified by the respective position vectors $\mathbf{r}_i$ and $\mathbf{r}_j$, on the basis of the differences between selected geographical parameters at those two locations. The parameters considered are: the horizontal (radial) distance between the two points $d(\mathbf{r}_i, \mathbf{r}_j)$, their elevation difference $z(\mathbf{r}_i, \mathbf{r}_j)$ and the difference between their land area fractions $w(\mathbf{r}_i, \mathbf{r}_j)$. $\rho(\mathbf{r}_i, \mathbf{r}_j)$ is modelled as:

$$\rho(\mathbf{r}_i, \mathbf{r}_j) = \exp\left\{-\frac{1}{2}\left[\frac{d(\mathbf{r}_i, \mathbf{r}_j)}{D^h}\right]^2 - \frac{1}{2}\left[\frac{z(\mathbf{r}_i, \mathbf{r}_j)}{D^z}\right]^2\right\} \cdot \ldots$$
$$[1 - (1 - w_{\min})|w(\mathbf{r}_i, \mathbf{r}_j)|]. \tag{10}$$

D^h and D^z are reference length-scales used to introduce different covariance suppression rates along the horizontal and

vertical dimensions. 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 1). Contrary to the flow-dependent $\mathbf{P}^b$, the localization function is not related to the atmospheric flow.

Because $\overset{i}{\mathbf{R}}{}^{-1}$ appears in Equation 4 and 5 instead of $\overset{i}{\mathbf{R}}$, it is more convenient to write:

$$\overset{i}{\mathbf{R}}{}^{-1}_{jj} = (\sigma_o^2 c_j)^{-1} \cdot \rho(\mathbf{r}_i, \mathbf{r}_j), \qquad (11)$$

where the two main factors have been separated by the symbol. Note that the localization function ρ in Equation 11 decreases near the boundary of the local domain, therefore the observation error variance is increased.

3 | NWP MODEL AND OBSERVATIONS

The two-metre air temperature forecasts from the AROME-MetCoOp (i.e. Meteorological Cooperation on Operational Numerical Weather Prediction) model for the Nordic countries (Müller *et al.*, 2017) have been used to obtain the EnSI ensemble background. In particular, the Met-CoOp Ensemble Prediction System (MEPS: Frogner *et al.*, 2019) has been considered in our study. MEPS has been running operationally four times a day (0000, 0600, 1200, 1800 UTC) since November 2016 at MET Norway and its ensemble consists of 10 members. The two-metre temperature fields are available over a regular grid of 2.5 km. The finer resolution of topography contributes to the improved two-metre temperature forecast skill of the high-resolution model compared to the global model. For each analysis, we use the available MEPS forecast with the shortest lead time (0–6 h). The forecast fields are downscaled onto a 1 km grid by using the elevation differences between the 2.5 km grid and the 1 km grid. To better account for inversions, the method computes an elevation temperature gradient in a neighbourhood (15×15 km^2) surrounding the point. A smaller neighbourhood yields less reliable estimates of the gradient, while a larger neighbourhood is more likely to incorporate other temperature gradients not due to elevation difference. The air temperature from the lowest model level (approximately 12 m) was used to estimate the gradient as it gave better results compared with using the 2 m temperature (not shown). The downscaling procedure has been implemented in the open-source software *gridpp* (see https://github.com/metno/gridpp). The choice of a 1 km grid has been proven useful for several applications over Norway requiring high-resolution meteorological fields (Saloranta, 2014; Gisnås *et al.*, 2016; Lussana *et al.*, 2018a; 2018b).

The *in situ* observations come from the hourly two-metre air temperature measurements collected by the network of meteorological weather stations directly managed by MET
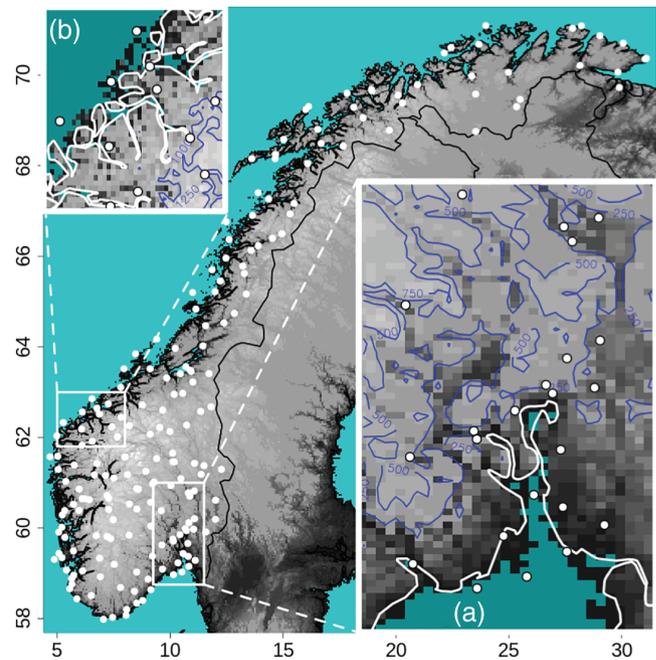
FIGURE 2 Terrain map (grey shades, 1×1 km^2 grid) and station locations over Norway (white dots, stations with more than 90% of valid data in the period from July 2017 to July 2018). Grid points over the sea are shown in cyan. In the boxes, two sub-domains are shown: (a) Oslo Fjord and its surroundings; (b) coastal region in western Norway. The white lines mark the coastlines. The blue lines are the contour lines (m a.m.s.l), sub-domain (a) 250, 500, 750 m; sub-domain (b) 1,000, 1,250 m

Norway. The observations have been quality controlled by experienced staff and with the help of automatic procedures. We have opted for not including a check of the observations against the MEPS temperature fields. In fact, we prefer to allow for large deviations between model data and observations because we know that this might happen during wintertime over complex terrain. The possibility to improve the temperature fields in those situations is highly significant for the purpose of this study. By using professional stations subject to regular maintenance and installed mostly for climatological purposes, we may consider observations as fairly accurate estimates of actual temperatures in the immediate surroundings of station locations. As a consequence of the quality assurance system in place at MET Norway, the rare gross measurements errors, as defined by Gandin (1988), do not have a significant impact on the evaluation.

We focus on the Norwegian mainland, as shown in Figure 2, though MEPS covers a larger domain. The 1 km topography is also shown, together with the land area fraction. The white dots mark the stations used for verification. Panel A shows the Oslo Fjord and its surroundings, an area that is used in several examples in the following. Panel B shows the extremely steep elevation gradients along the western Norwegian coast with high mountains just a few kilometres away from the sea.

Norway with its intricate coastline and complex terrain is an excellent region for testing spatial interpolation schemes under challenging conditions. 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 2,000 m but most of the stations are located below the elevation of 1,000 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. The average distance between a station and its fifth closest station is approximately 50 km in the south and 75 km in the north. On a regional level, the average distance between a station and its twentieth closest station is approximately 125 km in the south and 250 km in the north. Ideally, spatial interpolation would require a denser network of observations where the temperature variance is larger, in order to get a fine-scale representation of the temperature field where it varies the most. 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.

4 | RESULTS AND VALIDATION

The objective of this section is the validation of our method with real data (section 4.2) and, at the same time, we propose two procedures for parameter optimization (section 4.1). The next paragraphs are an introduction to sections 4.2–4.1. First, we outline the evaluation strategy. Then, the diagnostics used in both parameter optimization and evaluation are defined. Next, the OI equations are briefly summarized, since we will make extensive reference to them. Finally, the OI-based Integral Data Influence (IDI: Uboldi *et al.*, 2008; Lussana *et al.*, 2010) for the evaluation of the impact of station density on the analysis is described.

EnSI has been implemented over Norway for the year-long period from July 2017 to July 2018 with the optimal parameter values found in section 4.1 and reported in Table 1. The background and the observations used have been described in section 3. All the available stations have been used for spatial analysis, while for validation only those stations with more than 90% of valid data have been considered. Section 4.2.1 describes the EnSI evaluation over the year-long period and it is mostly based on leave-one-out cross-validation (CV). EnSI has also been compared against two OI-based analysis schemes (section 4.2.2): (a) an observations-only OI which makes no use of NWP model output, and (b) a "classical" OI using a single field as the background, instead of an ensemble. These two methods have been used to create ad hoc gridded

TABLE 1 Optimal values for the parameters of EnSI as it has been implemented over Norway (section 3)

D^h (km)	D^z (m)	w_{min}	σ_o (°C)	Δ
70	200	0.5	1	2

datasets based on the same observational network used for the EnSI implementation. The seNorge2 spatial interpolation method (Lussana *et al.*, 2018a; 2018b) has been used to create the observations-only gridded dataset (henceforth referred to as the seNorge2 dataset). Observational gridded datasets provide accurate data sources in the surroundings of station locations (Isotta *et al.*, 2014) and their performances degrade in data-sparse regions (Hofstra *et al.*, 2008). The seNorge2 method is based on the OI scheme described by Uboldi *et al.* (2008), which is used here as a reference also for the implementation of the classical OI (henceforth referred to simply as OI). The only difference is that seNorge2 estimates the (pseudo)background field from the observations, while OI uses an observation-independent background gridded field. In particular, the EnSI background ensemble mean is used as the OI background, such that OI and EnSI operate under comparable conditions. As a final point in the validation, in section 4.2.3 we have considered a case-study where we investigate the analysis performance in representing temperature inversions, that is one of those cases that may lead to large prediction errors.

Diagnostics can be obtained combining quantities available after the spatial analysis, such as: observed values; background and analysis in the observation space. The innovation (observation-minus-background) statistics are used in section 4.1 for parameter optimization and in section 4.2 as a base reference to assess the analysis improvements over the background. The CV-analysis ensemble at a station location is the analysis ensemble obtained considering all the observations except the one measured at that location. The analysis residuals (analysis-minus-observation) and CV-analysis residuals (CV analysis-minus-observation) statistics are used in section 4.1.2 to set the ensemble scaling factor and in section 4.2 as performance indicators. As a matter of fact, the analysis error (analysis-minus-truth) at grid points is the quantity we would like to study. Since the true temperatures at grid points are unknown, we base our investigation on CV. In the comparison between analysis and background, often the ensemble mean and spread are used instead of the whole ensemble. In those cases, the residuals are given by the ensemble means minus the observations. The statistics we use as verification scores are the mean absolute error (MAE) and the root-mean-squared-error (RMSE). The first quantifies average mean absolute deviations, while the second quantifies the associated spreads. As some users are not interested in small deviations, MAE 1 °C considers only absolute values of CV-analysis residuals or innovations greater than 1 °C. The

3 °C threshold is used at MET Norway to define a significant deviation from the observed temperature that undermines the user confidence in the forecast. For this reason, absolute values of residuals or innovations greater than 3 °C are referred to as "large errors" and their frequencies of occurrence are compared to assess the added value of EnSI in reducing the number of large errors.

OI assumes both the observation error covariance matrix **R** ($p \times p$) and the background error covariance matrices **B** ($m \times m$) as global and static. **H** is the linear observation operator mapping m-vectors into p-vectors (it is implemented as the $\overset{i}{\mathbf{H}}$ operators of section 2), therefore $\mathbf{H}\mathbf{x}^b$ is a p-vector. The **B** matrix does not appear explicitly in the equations; instead, the two matrices used are the error covariance between station locations $\mathbf{S} = \mathbf{H}\mathbf{B}\mathbf{H}^T$ ($p \times p$) and between grid points and station locations $\mathbf{G} = \mathbf{B}\mathbf{H}^T$ ($m \times p$). The OI best linear unbiased estimate can be written as the Kalman filter's analysis step of Equation 2:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{G}(\mathbf{S} + \mathbf{R})^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b). \quad (12)$$

The analysis is shaped by the definitions of **R** and **B**. We assume a diagonal **R** with an observation error variance σ_o^2, that in the following will be related to the error variance characterizing the observational network as a whole introduced in section 2. Analogously, a constant error variance σ_b^2 is introduced for the background. The background error correlations between points are modelled through the function ρ of Equation 10, previously introduced as the localization function. ρ has been used by Lussana *et al.* (2009) as a correlation function for spatial interpolation, with the land use instead of the land area fraction. We can write our assumptions on the error covariance matrices as:

$$\mathbf{R} = \sigma_o^2 \mathbf{I}, \quad (13)$$

$$\mathbf{B}_{ij} = \sigma_b^2 \rho(\mathbf{r}_i, \mathbf{r}_j), \quad (14)$$

such that $\mathbf{x}^a$ of Equation 12 can be rewritten as:

$$\mathbf{x}^a = \mathbf{x}^b + \widetilde{\mathbf{G}}(\widetilde{\mathbf{S}} + \varepsilon^2 \mathbf{I})^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b), \quad (15)$$

where the error variances do not need to be specified. Instead, we need to specify the ratio $\varepsilon^2 \equiv \sigma_o^2/\sigma_b^2$. Because we believe the observations to be, on average, more precise estimates of the truth than the background, we set $\varepsilon^2 = 0.5$. We anticipate here that in section 4.2.2, both seNorge2 and OI implementations use the EnSI optimal values of Table 1.

The station density is one of the most important factors to take into account in evaluating the performances of any spatial interpolation method. We use IDI as a diagnostic to quantify the added value of the observational network over the analysis

at an arbitrary point. Cardinali *et al.* (2004) defined a similar diagnostic and named it "degrees of freedom". IDI has been used also to evaluate the fit-for-purpose of a distribution of weather stations (Horel and Dong, 2010). In practice, IDI is obtained as the analysis in Equation 15 by arbitrarily assigning the value of 1 to the observations (i.e. corresponding to the maximum amount of information available) and the reference value of 0 to the background (i.e. basic amount of information available everywhere). In the vicinity of an observation, the IDI field is close to 1 whereas for data-void regions it is 0. Since we use CV to link statistics at station locations to statistics at grid points, we have introduced the CV-IDI, that is IDI at a station location obtained without considering the presence of that station. Note that when IDI is computed at a station location, its lower bound is always greater than zero because the station presence adds information to the background alone. On the other hand, CV-IDI for a completely isolated station is equal to zero. Consider the observational network used for validation, Figure 3a shows the CV-IDI computed at station locations and Figure 3b shows the IDI field at grid points. Both IDI and CV-IDI are based on Equation 15, modified according to the definition of IDI. The correlations ρ of Equation 14 have been computed with the EnSI optimal values of Table 1. Because of uneven station distribution described in section 3, Figure 3a and especially Figure 3b show that the added value of the observational network over the analysis is limited in complex terrain, where the elevation differences among nearby grid points are significant (compared to D^z in Table 1). The bottom-right panel in Figure 3a shows the close relationship between CV-IDI and station density, which is here defined as the horizontal distance between a station and its nearest five stations. IDI at grid points and CV-IDI at station locations have been divided into four classes: data-sparse regions have values smaller than 0.45, that corresponds to a distance to the nearest five stations of approximately 100 km or more; data-dense regions have IDI greater than 0.85, that corresponds on average to a distance to the nearest five stations of less than 60 km; then, two transitional classes have been defined.

4.1 | Parameter optimization

4.1.1 | Localization

The innovation statistics are used to set the error variance characterizing the observational network σ_o^2 (Equation 11) and the localization parameters D^h and D^z (Equation 10). ρ requires the specification of a third parameter: $w_{\min}$. In an ideal situation of a very dense observational network, one may consider relying on adaptive estimates for all the parameters. This is not the case for our rather sparse network, so we have opted for a simplification that yields robust estimates for the more important parameters. The impact of land area fraction differences on ρ is less dramatic than those of the
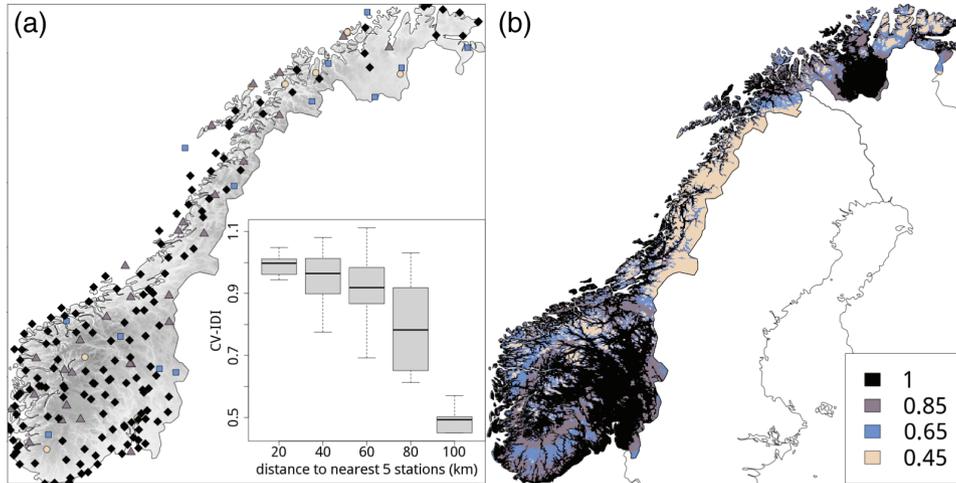
FIGURE 3 Integral Data Influence (IDI) for the observational network used for evaluation. (a) The CV-IDI at station locations. Different symbols have been associated to the colours; the box-plot on the bottom-right corner shows the relationship between CV-IDI and the station density. (b) The IDI at grid points. The colour scales are the same for both panels

other geographical parameters. 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.

The same function ρ is used to define the (static) OI background error correlations in Equation 14 and to shape indirectly the background error correlations in EnSI via the R-localization of Equation 11. **B** can be thought of as a time-averaged version of $\mathbf{P}^b$, such that we can write:

$$\mathbf{S} = \mathbf{HBH}^{\mathrm{T}} = \langle \mathbf{HP}^b \mathbf{H}^{\mathrm{T}} \rangle. \tag{16}$$

The ensemble average $\langle \dots \rangle$ is defined as the average over the hours in the year-long evaluation period. By averaging over thousands of cases, the temporal mean in Equation 16 is damping the spurious long-distance correlations which are present in $\mathbf{P}^b$. That is exactly the task the localization function carries out on an hourly basis.

The innovation sample covariance matrix is related to the sum of observation and background error covariance matrices, $\mathbf{R}$ and $\mathbf{S}$ of section 4, by:

$$\langle (\mathbf{y}^o - \mathbf{Hx}^b)(\mathbf{y}^o - \mathbf{Hx}^b)^{\mathrm{T}} \rangle = \mathbf{R} + \mathbf{S}. \tag{17}$$

This equation follows directly from the definitions of errors (Desroziers *et al.*, 2005). The key point in the optimization of the localization parameters is that $\mathbf{S}$ is present in both Equations 16 and 17. The optimization of the matching between the innovation sample covariance matrix, on the left-hand side of Equation 17, and the sum of $\mathbf{R}$ and $\mathbf{S}$, defined in Equations 13 and 14, allows us to set the values of σ_o^2, D^h and D^z. In particular, with reference to the innovation sample covariance matrix: the diagonal elements are used to set $\sigma_o^2 + \sigma_b^2$, which is the constant value of all the diagonal elements

of $\mathbf{R} + \mathbf{S}$ (it is worth recalling that we impose the constraint $\sigma_o^2/\sigma_b^2 = 0.5$, as in section 4); the off-diagonal elements are used to set D^h and D^z. Then, because of Equation 16 we believe that the same values of σ_o^2, D^h and D^z are also useful for the EnSI localization function.

The optimal values, as returned by a least-mean-squared fitting procedure, are (Table 1): $D^h = 70$ km, $D^z = 200$ m and $\sigma_o = 1\,°\mathrm{C}$. A reference horizontal length scale of $D^h = 70$ km coincides with the meso-β atmospheric scale according to Orlanski (1975) and Thunis and Bornstein (1996). In Figure 4, the decrease of the $\mathbf{S}$ covariances with the increase of the horizontal distance between a pair of observation locations is shown, together with σ_o^2 and σ_b^2 estimates. Only pairs of points whose elevation differences are less than 250 m are shown in the graph. The cloud of empirical points shows a gradual decrease of the covariances, while the theoretical decrease is sharper.

The optimization procedure described is ultimately based on a few subjective decisions. We carried out a sensitivity analysis on the impact of those decisions. A first experiment reveals that the optimal values are robust to deviations of σ_o^2/σ_b^2 from the pre-set value of 0.5: the variations in σ_o^2, D^h and D^z are of the order of a few per cent from their optimal values. A second sensitivity experiment has been made where the stations in the north and in the south of Norway have been considered separately; also in this case the variations of the optimal values were not significant. A third sensitivity experiment has been made by replacing the ensemble mean $\mathbf{x}^b$ in Equation 17 with the individual ensemble members (i.e. the columns of $\mathbf{X}^b$). The results were: D^h between 70 and 80 km; D^z between 200 m and 215 m; σ_o^2 between 1 and 1.4 °C.

Figure 5 shows the values of the localization function $\rho(\mathbf{r}_i, \mathbf{r}_j)$ (Equation 10) for a selection of D^h and D^z values. $\mathbf{r}_i$ corresponds to the city of Oslo (i.e. the blue point in the centre of the domain, see Figure 2 sub-domain A) and $\mathbf{r}_j$ varies
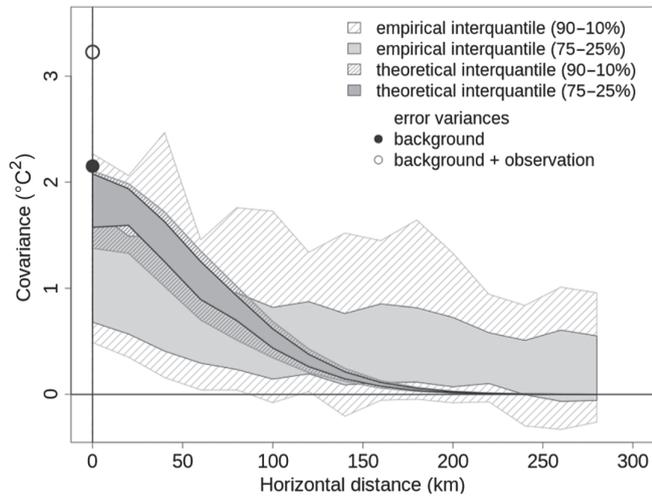
FIGURE 4 Interquantile ranges of: the innovation sample covariances ("empirical" distribution, left-hand side of Equation 17, light grey); $\mathbf{R}+\mathbf{S}$ ("theoretical" distribution, right-hand side of Equation 17, dark grey), when $w_{\min}=0.5$, $D^h=70\,\mathrm{km}$, $D^z=200\,\mathrm{m}$, $\sigma_o=1\,^\circ\mathrm{C}$, assuming $\sigma_o^2/\sigma_b^2=0.5$. σ_o^2 and σ_b^2 are shown as dots on the vertical axis. Horizontal distance bin width is set to 10 km and it has been used to compute the interquantile ranges. Only pairs of stations with absolute elevation differences less than 250 m are shown

FIGURE 5 Localization function $\rho(\mathbf{r}_i,\mathbf{r}_j)$ (Equation 10) for several values of D^h and D^z ($w_{\min}=0.5$). The domain is centred on the blue dot $\mathbf{r}_i$ (Oslo, Figure 2a), $\mathbf{r}_j$ are the grid points within a box of 280 km by 280 km

over the neighbouring grid points. At the bottom-left corner, with $D^h=35\,\mathrm{km}$ and $D^z=100\,\mathrm{m}$, the localization is more extreme and only those locations relatively close to Oslo will have a chance to influence the analysis. On the other hand, when $D^h=140\,\mathrm{km}$ and $D^z=400\,\mathrm{m}$, as it is in the upper-right corner, the analysis in Oslo would include more information, thus making its value representative of a larger region. $\rho(\mathbf{r}_i,\mathbf{r}_j)$ based on the optimal values in Table 1 is shown in the central panel.

4.1.2 | Ensemble scaling factor

The fixed scaling factor Δ in Equation 1 is used to adjust $\mathbf{P}^b$ so as to obtain an analysis that is possibly sharper than the raw ensemble background and more accurate. The optimization aims at reducing the average deviation between CV-analyses and observations.

Figure 6a shows that the Continuous Ranked Probability Score (CRPS: Jolliffe and Stephenson, 2012) depends on the station density even for the background. Since the background ensemble spread does not depend on the station density (not shown), it can be concluded that the station network is denser where the model is performing better (see the black diamonds in Figure 3). It is worth pointing out that since the statistics for data-sparse regions are based on a few stations only, it should be taken more as a rough indication than a robust estimate. The CRPS (Figure 6a) indicates that $\Delta=5$ yields better results for data-sparse areas, while for data-dense areas the CRPS has similar values for all the Δ. According to Figure 6b, in data-dense areas large errors are 30% less likely to occur in

the CV-analysis residual than in the innovation when Δ is set to 1 or 2. On the contrary, with $\Delta=5$ the analysis is less effective in reducing the occurrence of large errors. Note that the statistics shown in Figure 6 are based on the differences at station locations between: (a) CV-analysis residual ensemble mean and observation, and (b) innovation ensemble mean and observation.

The optimal value has been set to $\Delta=2$ (Table 1) as a trade-off between the need to maximize accuracy of the ensemble mean and the reliability and resolution of the analysis ensemble, as summarized in the CRPS score.

An example of how EnSI modifies the patterns of uncertainty in temperature estimation is shown in Figure 7. For both EnSI and OI the model parameters are set to the values in Table 1. Even though the example refers to just 1 h of the case-study presented in section 4.2, it helps us to describe some general properties of EnSI. Figure 7a shows the background ensemble standard deviation for each grid point. The background uncertainty is of course independent of the observational network; the largest uncertainty is for temperatures on the valley floors. The EnSI analysis ensemble standard deviations are shown in Figure 7b. The analysis reduces the ensemble spread in densely observed areas, without destroying the spatial pattern of the uncertainty given by the numerical model. The largest analysis uncertainties are still on the valley floors, and data-void areas maintain the same ensemble spread as in the background. Without an ensemble background but with a single-field background, it is possible to apply methods such as the OI presented in section 4. Figure 7c shows the analysis standard deviation at
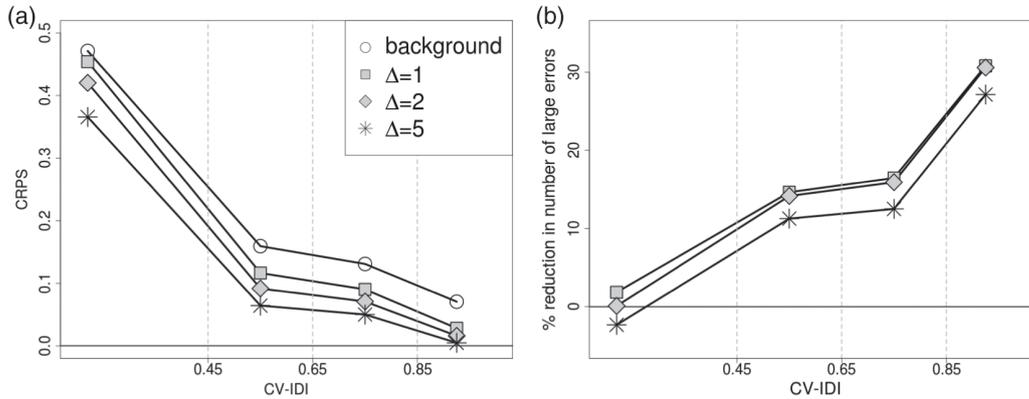
FIGURE 6 CV-analysis residual and innovation ensemble mean statistics as a function of: (a) the scaling factor Δ (Equation 1), and (b) CV-IDI; based on data from July 2017 to July 2018. (a) shows the averaged Continuous Ranked Probability Score (CRPS) for classes of CV-IDI (Figure 3). (b) shows the averaged percentage of reduction in the number of large errors when comparing the CV-analysis residual against the innovation for classes of CV-IDI. A negative reduction indicates an increase in the number of large errors

grid points, which can be obtained as per the analysis step of the Kalman filter in Equation 3. σ_o^2 is obtained as in Eqn (32) of Lussana *et al.* (2010), which gives $\sigma_o^2 = 0.798(°C)^2$. The analysis uncertainty displayed in Figure 7c depends totally on the observational network: it is smaller close to station locations, while it increases up to the maximum value of $\sigma_b^2 = 1/\varepsilon^2\sigma_o^2$ in data-void areas. According to OI, the temperature estimates on the valley floors are less uncertain than those on the mountain ridges because valley floors are better covered by observations. The uncertainty patterns in Figure 7c are very different from those in Figure 7b, even though the standard deviations at station locations are rather small in both figures. The added value of the ensemble is that the analysis uncertainty is shaped not only by the observational network but also by the weather, as simulated by the NWP model.

4.2 | Validation

4.2.1 | Evaluation over one year of hourly data

The benefits of the analysis over the background depend on the observation density (Figure 8), especially for the reduction of large errors. At grid points where the nearest five stations are less than 30 km away, the fraction of large errors is reduced by almost 50%, while the improvements for MAE and RMSE are between 20 and 25%. MAE 1 °C behaves like the reduction of large errors, and in data-dense regions the improvement is over 40%. The analysis improvement over the background gradually decreases with decreasing station density; nonetheless the analysis scores better than background even at points where the distance to the five nearest stations is between 100 and 140 km.

The expected reduction in the number of large errors over grid points is shown in Figure 9. A relationship between IDI at the ith grid point (x_i^{IDI}) and the percentage of reduction in

the number of large errors (x_i^r) can be assumed, such as:

$$x_i^r = 39.757 \exp\left\{-\frac{1}{2}\left[\frac{x_i^{IDI} - 1.171}{0.367}\right]^2\right\}, \quad (18)$$

where the three coefficients (39.757, 1.171, 0.367) have been optimized so to get the best match between the Gaussian function and the empirical data used in Figure 6b. We have tested other functions (e.g. polynomial) to model this relationship, but eventually the Gaussian function provided the best-fitting results. According to Equation 18, in the case of an ideal station density (IDI = 1) the percentage of reduction in the number of large errors is approximately 35%. In Figure 9, boxes A and B show data-dense regions. Because of the extremely complex terrain in B (Figure 2), IDI (Figure 3) indicates that the observations' influence onto the analysis is limited to a small number of grid points. For this reason, the analysis is more effective in reducing large errors in A than in B. Note that Equation 18 can also be used for planning variations in the observational network by assessing the potential of new stations in reducing large prediction errors.

4.2.2 | Comparison against OI-based spatial analyses

The EnSI analyses are compared against seNorge2. The CV-analysis residual statistics have been compared in Figure 10. For EnSI, the residuals are the CV-analysis ensemble means minus the observations. The stations have been classified according to the station density as represented by the CV-IDI (Figure 3a). The class labelled with CV-IDI = 0.5 includes all those stations with CV-IDI < 0.55. Then, CV-IDI has been divided in bins of 0.1, up to CV-IDI = 1.1; within each bin the distributions of CV-analysis residuals are shown through the median (dots) and the interquartile range (IQR,

(a)   (b)   (c)
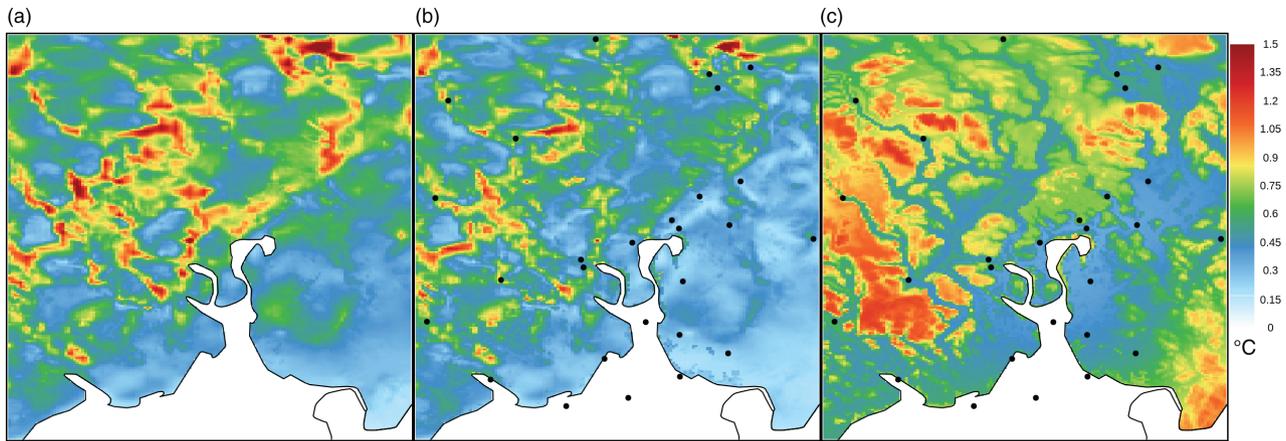
FIGURE 7 25 January 2018, 0900 UTC, uncertainty of the temperature estimates at grid points over sub-domain A in Figure 2: (a) model-derived background, ensemble standard deviation; (b) EnSI analysis, ensemble standard deviation; (c) OI analysis, standard deviation as derived from the analysis error covariance matrix, Equation 3. Black dots in (b,c) mark the station locations
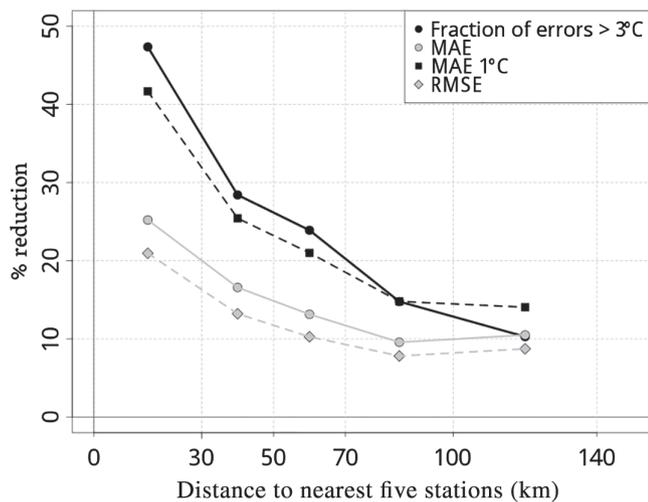
FIGURE 8 CV-analysis residual and innovation ensemble mean statistics as a function of the station density (distance to the nearest five stations) based on data from July 2017 to July 2018. Analysis improvement over the background is shown as % reduction for four different verification metrics (section 4). MAE 1 °C is the mean absolute error (MAE) beyond 1 °C, that is, errors less than 1 °C are not counted

envelope). The regression lines for the median are also shown. EnSI performs constantly better than seNorge2, as highlighted by the regression lines. In data-dense areas, the percentage of large errors (Figure 10c) is around 10% for both datasets. In data-sparse regions, EnSI performs significantly better than seNorge2 for all the verification scores. The spreads of the distributions of RMSE (Figure 10a) and MAE (Figure 10b) are always smaller for EnSI, especially in data-sparse areas, thus indicating that the combined analysis is more stable with respect to variations in the station density than seNorge2. Furthermore, from Figure 10 it can be inferred that EnSI ensemble mean at grid points has: MAE between 0.8 and 1.8 °C, RMSE between 1 and 2.5 °C and an occurrence of large errors between 8 and 18%.

EnSI has been compared also against the OI of Equation 15, with MEPS ensemble mean as the background. The OI background could also have been the output of a deterministic NWP model, if this is available. If we consider MAE, RMSE and percentage of large errors over the year-long period under examination, the verification scores for OI are almost identical to the verification scores of EnSI presented in Figure 10. This is not surprising, given that the EnSI localization function has been optimized so as to match the OI settings, as described in section 4.1.1.

4.2.3 | A wintertime case-study

During 25 January 2018 a low-pressure system pushing westwards in northern Norway brought unsettled conditions to parts of southern and middle Norway (Figure 11). The advection of moist air from over the Atlantic Ocean towards the mountains in the central part of the region caused intense precipitation on the west coast. East of the mountains, no precipitation has been registered. The satellite images show the presence of clouds for most of the day over the whole region. In the first part of the day, southerly winds advected warm air over Oslo Fjord. At the same time, localized föhn episodes associated with winds from west-northwest are likely to have occurred in the inland valleys and in the lateral valleys overlooking Oslo Fjord, as suggested by localized temperature rises during night-time. Then, the wind gradually turned everywhere from west-northwest and temperatures dropped.

At 0900 UTC some valleys in both Oslo Fjord (O, Figure 11) and inland (I, Figure 11) were experiencing warming episodes. According to Figure 12, the decrease in temperature started at 1100 UTC in domain I and after 1400 UTC over domain O. By 2300 UTC, the temperature was dropping fast all over the domain. At 0900 UTC (Figure 11, left-column), the EnSI analysis increment shows that the background has been adjusted so to better represent the
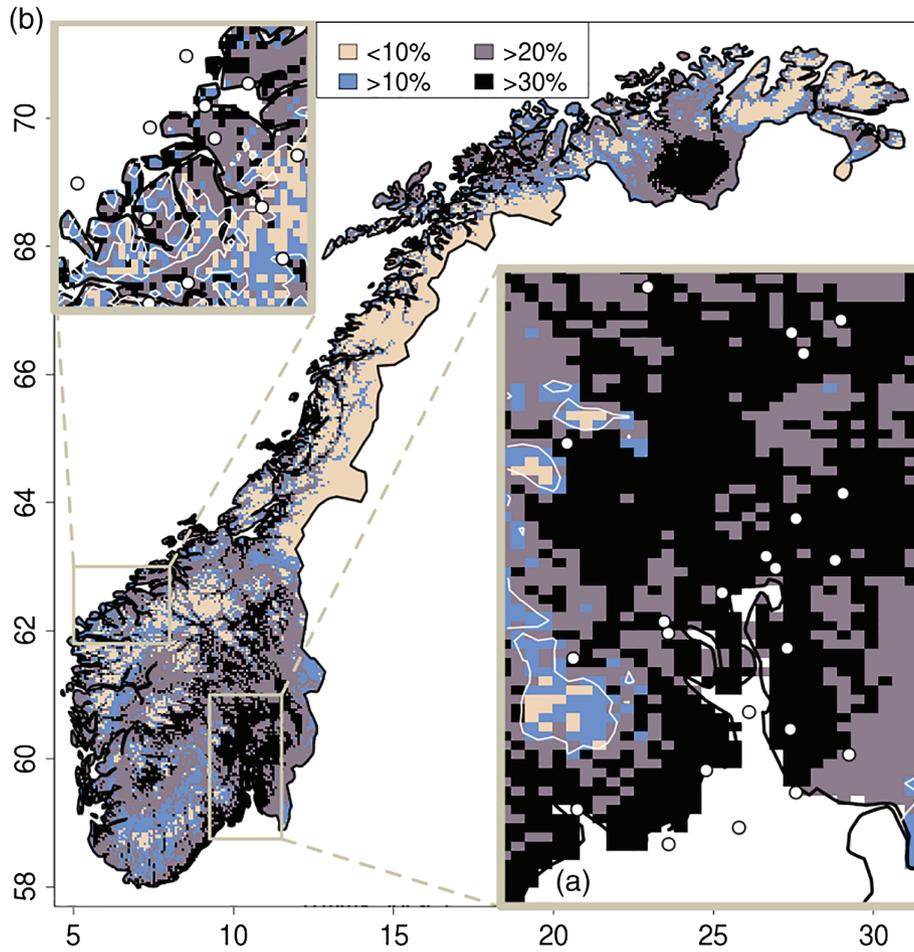
FIGURE 9 Estimates at grid points of the percentage of reduction in the number of large errors in the analysis compared to the background (section 4.2). Boxes A and B are the same as in Figure 2
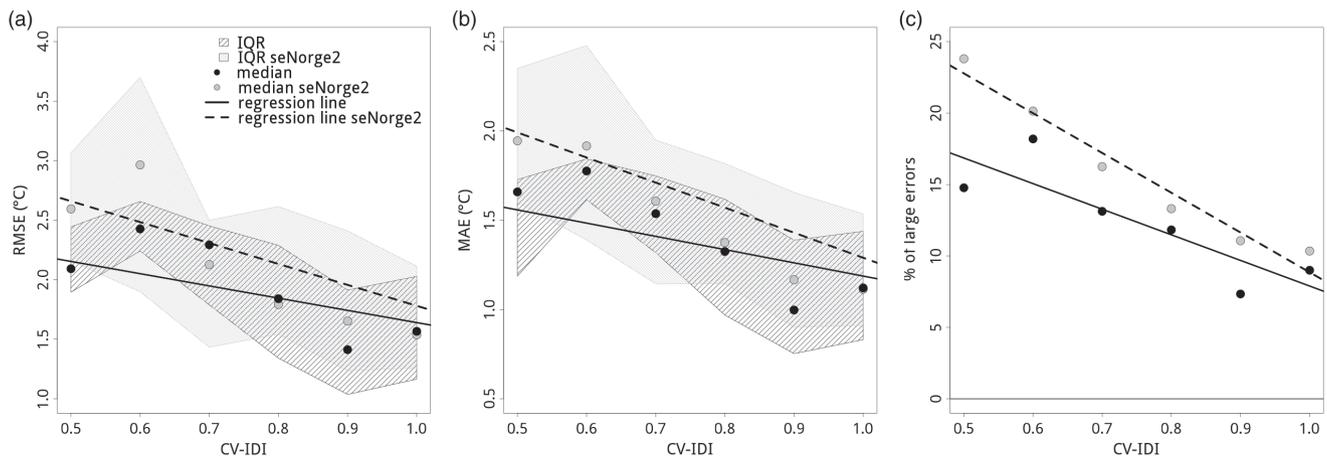
FIGURE 10 CV-analysis residual ensemble mean and seNorge2 CV-analysis residuals as a function of CV-IDI based on data from July 2017 to July 2018: (a) RMSE; (b) MAE; (c) percentage of large errors

warming, especially over the southeastern part of the domain where increments of $5\,°C$ can be found in lateral valleys. At 2300 UTC (Figure 11, right-column), the analysis increment shows a composite pattern with positive values over the O domain and negative values over the I domain. As a matter of fact, the situation is more complex at 2300 UTC and the

interpolation uncertainty is larger, especially over domain I, as is shown by the fields of the ensemble spread (Figure 11, bottom-row).

The EnSI analysis increments over sub-domain A of Figure 2 are shown in Figure 13 and they are compared with the OI analysis increments. Since the observational network
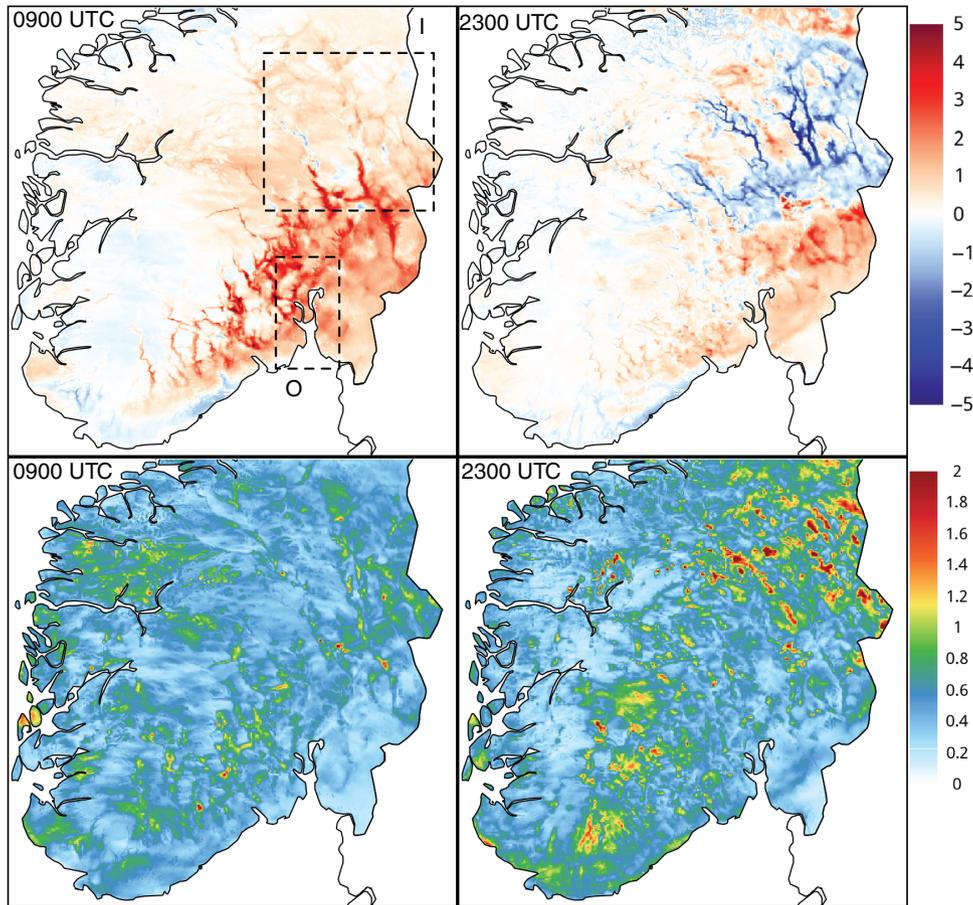
FIGURE 11 25 January 2018. Analysis ensemble mean increment (top row) and analysis ensemble standard deviation (bottom row) at 0900 and 2300 UTC over southern and middle Norway. Units are °C. The two rectangles in the top-left panel display the two sub-domains (O, Oslo Fjord; I, inland) used in Figure 12
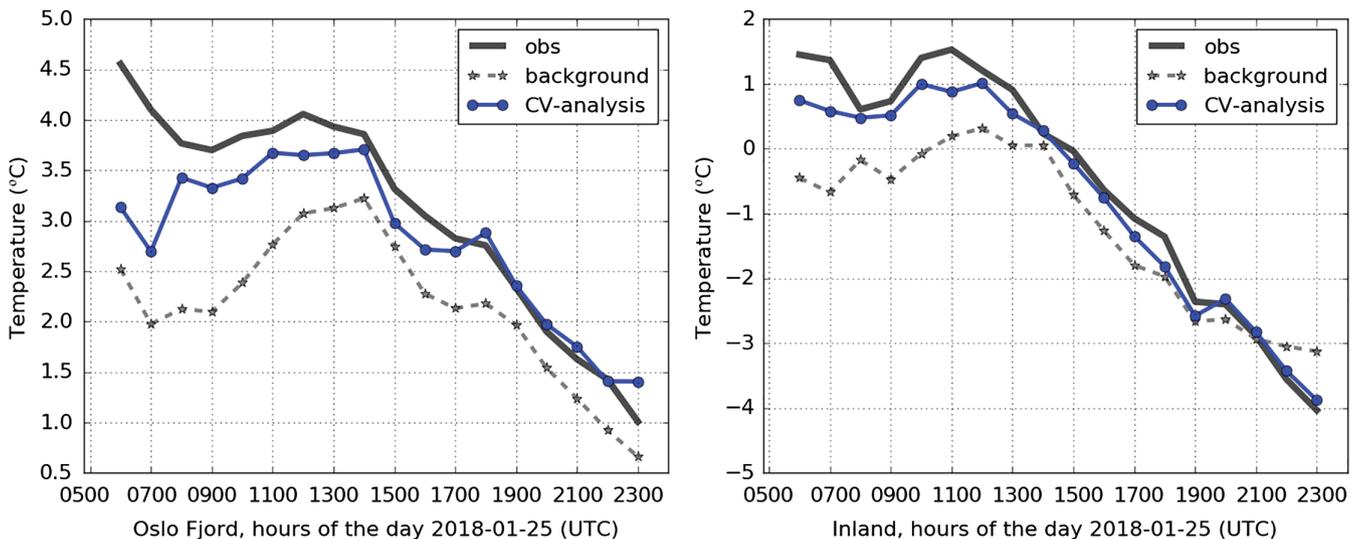
FIGURE 12 25 January 2018. Time series of spatially averaged observations, background ensemble mean, and CV-analysis ensemble mean for Oslo Fjord and the inland areas (Figure 11)

used is the same, the two analysis increments for the same hour are not too different. However, the topography influence is much more evident in the OI because of the predominant role of D^z in suppressing the correlations in ρ (Equation 10).

As stated in section 4.1.2, the background ensemble allows EnSI to use more information because the analysis increments are shaped not only on the basis of geographical parameters characterizing the surface but also on similarities between
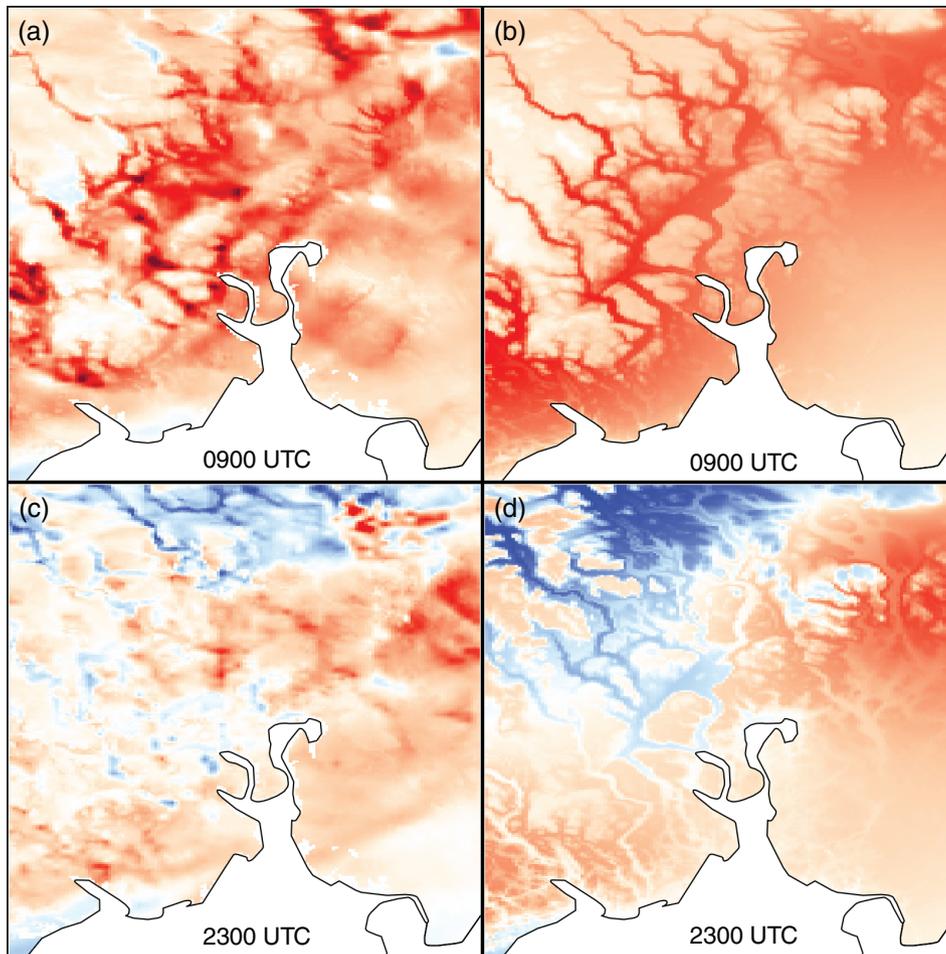
FIGURE 13 25 January 2018. Analysis increments over sub-domain A in Figure 2: (a) 0900 UTC, EnSI ensemble mean increment; (b) 0900 UTC, OI increment; (c) 2300 UTC, EnSI ensemble mean increment; (d) 2300 UTC, OI increment. The same colour scale as for the analysis increments of Figure 11 is used

ensemble members. For instance, consider two nearby valleys where only one of them has observations on the valley floor. OI will give significant weights to those observations in the analyses in both valleys. EnSI will do the same only if the weather is similar in both valleys, otherwise the analysis increments will be different.

Finally, in this paragraph we consider only the Oslo Fjord region (Figure 2a) and we focus on the vertical profile of near-surface temperature at 0900 UTC. Figure 14 shows the profiles as they can be reconstructed by values at station locations or at grid points. The observed profile (dots) is rather complex and it shows a "thermal inversion" in between the elevations of 100 and 200 m, where the two-metre temperature is increasing with elevation. This situation is quite typical in winter and local thermal inversions not resolved by the NWP model may cause large errors. The background ensemble (light grey envelope) follows the observed temperature profile, although it appears to be systematically colder and the narrow temperature inversion is not represented. The analysis ensemble (black) matches the observed temperature profile better and most of the observations fall within its envelope,

which is also characterized by a smaller spread than the background. The observation dots are coloured on the basis of the correction factors (appendix B). Note that those observations with small values of the correction factors – between 0.25 and 0.5 – are constraining the analysis even more towards them.

5 | SUMMARY AND CONCLUSIONS

The EnSI scheme developed at MET Norway provides two-metre temperature ensembles, on a real-time basis, that constitute a convenient alternative to the direct use of raw NWP ensembles. The flow-dependent interpolation method is derived from the analysis step of the LETKF and combines a background ensemble with *in situ* observations. We have introduced an ensemble scaling factor, a localization procedure, and station-dependent correction factors into the definition of observation error variances. The observation errors are characterized such that the most representative stations have a higher influence on the analysis.

EnSI has been validated over Norway by using the operational MEPS and the MET Norway observational network.
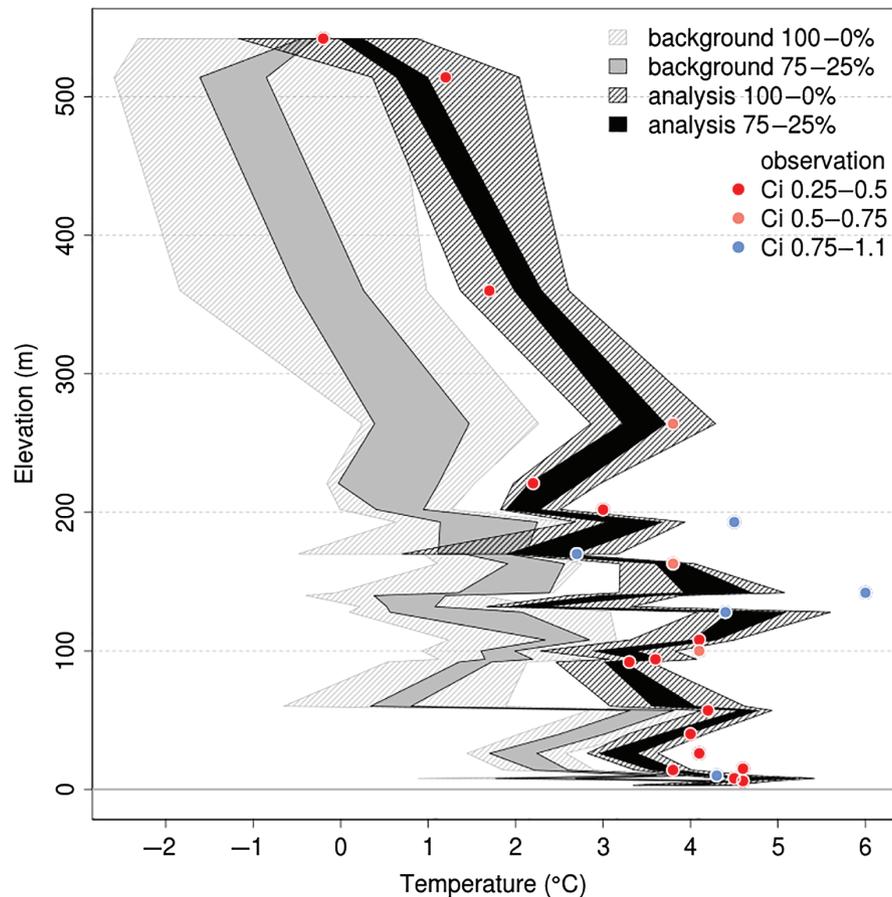
FIGURE 14   25 January 2018, 0900 UTC. Vertical temperature profiles in Oslo Fjord (Figure 2a): observations (dots, colours refer to the correction factors (appendix B)); background ensemble (grey); analysis ensemble (black)

The procedure presented for parameter optimization may constitute a guide for the application of the scheme to other regions. The evaluation has been carried out using 1 year of hourly data. In data-dense areas, the analysis reduces the number of predictions that deviate more than 3 °C from the actual temperature by 35% compared to the background. This number is closer to 50% for grid points having the nearest five stations within a radius of 30 km. Even in data-sparse areas, such as regions where the average distance between a point and its five nearest stations is larger than 100 km, the combined analysis reduces the number of large errors. EnSI analysis can capture thermal inversions in near-surface temperature better than the background, as we have shown for a typical wintertime situation. That is particularly important as thermal inversions are the source of large prediction errors in Norway.

The temperature analysis has also been compared to an ad hoc version of the seNorge2 observational gridded dataset, which is quite popular in Norway for applications requiring high-resolution fields. In data-dense areas, EnSI performs similarly or slightly better than the seNorge2 interpolation method. In data-sparse regions, not surprisingly large errors are more likely to occur in seNorge2 than in EnSI analyses. The fact that EnSI can provide more precise and accurate analyses than observational gridded datasets is an important result.

If compared to a classical OI with a (single) background field, the EnSI analysis ensemble mean verification scores are almost identical to those of the OI analysis. However, the EnSI analysis ensemble includes weather-related uncertainty that the OI analysis error covariance matrix is unable to represent. The flow-dependent temperature patterns represented by the NWP ensemble are kept in the EnSI analysis ensemble.

The method can be improved by taking into account more sophisticated optimization schemes for the scaling factor and the localization function, which can also be allowed to vary in space and time. The use of a denser network of observations may also lead to a better representation of the temperature field because of the reduced de-correlation lengths that could be used in the localization procedure. In this sense, the use of third-party data (e.g. crowd-sourcing and amateur station data) constitutes an interesting option to gain access to massive amounts of weather data.

ACKNOWLEDGEMENTS

ORCID

C. Lussana 🄳 https://orcid.org/0000-0003-3159-4895
L. Cantarello 🄳 https://orcid.org/0000-0002-1005-8463

REFERENCES

Amezcua J, Van Leeuwen PJ. (2014) Gaussian anamorphosis in the analysis step of the EnKF: a joint state-variable/observation approach. *Tellus A: Dynamic Meteorology and Oceanography*, 66(1), 1–18, 23493. https://doi.org/10.3402/tellusa.v66.23493.

Brinckmann, S., Krähenmann, S. and Bissolli, P. (2016) High-resolution daily gridded data sets of air temperature and wind speed for Europe. *Earth System Science Data*, 8(2), 491–516. https://doi.org/10.5194/essd-8-491-2016

Brunetti, M., Maugeri, M., Nanni, T., Simolo, C. and Spinoni, J. (2014) High-resolution temperature climatology for Italy: interpolation method intercomparison. *International Journal of Climatology*, 34(4), 1278–1296. https://doi.org/10.1002/joc.3764.

Cardinali, C., Pezzulli, S. and Andersson, E. (2004) Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 130(603), 2767–2786. https://doi.org/10.1256/qj.03.205.

Carrassi, A., Bocquet, M., Bertino, L. and Evensen, G. (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5), e535. https://doi.org/10.1002/wcc.535.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M. and Jones, P.D. (2018) An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391–9409. https://doi.org/10.1029/2017JD028200.

Crespi, A., Lussana, C., Brunetti, M., Dobler, A., Maugeri, M. and Tveito, O.E. (2019) High-resolution monthly precipitation climatologies over Norway (1981–2010): joining numerical model data sets and *in situ* observations. *International Journal of Climatology*, 39(4), 2057–2070. https://doi.org/10.1002/joc.5933.

Daley, R. (1993) *Atmospheric Data Analysis*, Vol. 2. Cambridge, UK: Cambridge University Press.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005) Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3385–3396. https://doi.org/10.1256/qj.05.108.

Evensen, G. (2003) The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9.

Frei, C. (2014) Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. *International Journal of Climatology*, 34(5), 1585–1605. https://doi.org/10.1002/joc.3786.

Frei C, Isotta FA. (2019) Ensemble spatial precipitation analysis from rain gauge data: methodology and application in the European Alps. *Journal of Geophysical Research: Atmospheres*, 124(11), 5757–5778. doi:https://doi.org/10.1029/2018JD030004.

Frogner IL, Singleton AT, Køltzow MØ, Andrae U. (2019) Convection-permitting ensembles: challenges related to their design and use. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/https://doi.org/10.1002/qj.3525.

Gandin, L.S. (1988) Complex quality control of meteorological observations. *Monthly Weather Review*, 116(5), 1137–1156. https://doi.org/10.1175/1520-0493(1988)116⟨;1137:CQCOMO⟩2.0.CO;2.

Gandin, L.S. and Hardin, R. (1965) *Objective analysis of meteorological fields*, Vol. 242, Israel Program for Scientific Translations, Jerusalem.

Gaspari, G. and Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554), 723–757.

Gisnås, K., Etzelmüller, B., Lussana, C., Hjort, J., Sannel, A.B.K., Isaksen, K., Westermann, S., Kuhry, P., Christiansen, H.H., Frampton, A. and Åkeman, J. (2016) Permafrost map for Norway, Sweden and Finland. *Permafrost and Periglacial Processes*, 28(2), 359–378.

Greybush, S.J., Kalnay, E., Miyoshi, T., Ide, K. and Hunt, B.R. (2011) Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, 139(2), 511–522.

Gu, Y.X., Bélair, S., Mahfouf, J.F. and Deblonde, G. (2006) Optimal interpolation analysis of leaf area index using MODIS data. *Remote Sensing of Environment*, 104(3), 283–296. https://doi.org/10.1016/j.rse.2006.04.021.

Haylock, M., Hofstra, N., Klein Tank, A., Klok, E., Jones, P. and New, M. (2008) A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research*, 113(D20), 1–12. https://doi.org/10.1029/2008JD010201.

Hiebl, J. and Frei, C. (2016) Daily temperature grids for Austria since 1961: concept, creation and applicability. *Theoretical and Applied Climatology*, 124(1), 161–178. https://doi.org/10.1007/s00704-015-1411-4.

Hofstra, N., Haylock, M., New, M., Jones, P. and Frei, C. (2008) Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research*, 113(D21). https://doi.org/10.1029/2008JD010100.

Horel, J.D. and Dong, X. (2010) An evaluation of the distribution of Remote Automated Weather Stations (RAWS). *Journal of Applied Meteorology and Climatology*, 49(7), 1563–1578.

Hunt, B.R., Kostelich, E.J. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1), 112–126.

Ide, K., Courtier, P., Ghil, M. and Lorenc, A.C. (1997) Unified notation for data assimilation: operational, sequential and variational. *Journal of the Meteorological Society of Japan*, 75(1B), 181–189.

Isotta, F.A., Frei, C., Weilguni, V., Tadić, M.P., Lassegues, P., Rudolf, B., Pavan, V., Cacciamani, C., Antolini, G., Ratto, S.M., Munari, M., Micheletti, S., Bonati, V., Lussana, C., Ronchi, C., Panettieri, E., Marigo, G. and Vertačnik, G. (2014) The climate of daily precipitation in the Alps: development and analysis of a high-resolution grid dataset from pan-Alpine rain-gauge data. *International Journal of Climatology*, 34(5), 1657–1675.

Jazwinski, A.H. (2007) *Stochastic Processes and Filtering Theory*. Mineola, NY: Courier Dover Publications.

Jolliffe, I.T. and Stephenson, D.B. (2012) *Forecast Verification*. Oxford: Wiley.

Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45.

Kalman, R.E. and Bucy, R.S. (1961) New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1), 95–108.

Kalnay, E. (2003) *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge, UK: Cambridge University Press.

Kaplan, A., Kushnir, Y., Cane, M.A. and Blumenthal, M.B. (1997) Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysical Research*, 102(C13), 835–860. https://doi.org/10.1029/97JC01734.

Krähenmann, S., Bissolli, P., Rapp, J. and Ahrens, B. (2011) Spatial gridding of daily maximum and minimum temperatures in Europe. *Meteorology and Atmospheric Physics*, 114(3), 151. https://doi.org/10.1007/s00703-011-0160-x.

Li, H., Kalnay, E. and Miyoshi, T. (2009) Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135(639), 523–533.

Lorenc, A.C. (1986) Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474), 1177–1194. https://doi.org/10.1002/qj.49711247414.

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

Lussana, C., Salvati, M., Pellegrini, U. and Uboldi, F. (2009) Efficient high-resolution 3-D interpolation of meteorological variables for operational use. *Advances in Science and Research*, 3(1), 105–112.

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

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

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A. and Jonas, T. (2015) Evaluating snow models with varying process representations for hydrological applications. *Water Resources Research*, 51(4), 2707–2723.

Mahfouf, J.F., Brasnett, B. and Gagnon, S. (2007) A Canadian precipitation analysis (CaPA) project: description and preliminary results. *Atmosphere–Ocean*, 45(1), 1–17.

Miyoshi, T. (2011) The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139(5), 1519–1535. https://doi.org/10.1175/2010MWR3570.1.

Müller, M., Homleid, M., Ivarsson, K.I., Køltzow, M.A., Lindskog, M., Midtbø, K.H., Andrae, U., Aspelien, T., Berggren, L., Bjørge, D., Dahlgren, P., Kristiansen, J., Randriamampianina, R., Ridal, M. and Vignes, O. (2017) AROME-MetCoOp: a Nordic convective-scale operational weather prediction model. *Weather and Forecasting*, 32(2), 609–627.

Orlanski, I. (1975) A rational subdivision of scales for atmospheric processes. *Bulletin of the American Meteorological Society*, 56, 527–530.

Sakov, P. and Bertino, L. (2011) Relation between two common localisation methods for the EnKF. *Computational Geosciences*, 15(2), 225–237.

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

Saloranta, T.M. (2014) Simulating more accurate snow maps for Norway with MCMC parameter estimation method. *The Cryosphere Discussions*, 8(2), 1973–2003.

Skaugen, T. and Onof, C. (2014) A rainfall-runoff model parameterized from GIS and runoff data. *Hydrological Processes*, 28(15), 4529–4542.

Soci, C., Bazile, E., Besson, F. and Landelius, T. (2016) High-resolution precipitation re-analysis system for climatological purposes. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1), 29879. https://doi.org/10.3402/tellusa.v68.29879.

Thunis, P. and Bornstein, R. (1996) Hierarchy of mesoscale flow assumptions and equations. *Journal of the Atmospheric Sciences*, 53(3), 380–397.

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

APPENDIX A: MATHEMATICAL NOTATION

The notation used is based on both Ide *et al.* (1997) and Sakov and Bertino (2011). The number of grid points is m. The number of ensemble members is k. 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 background ensemble members are combined by columns into the $m \times k$ matrix $\mathbf{X}^b$. The ensemble mean is the m-vector $\mathbf{x}^b = (1/k)\,\mathbf{X}^b \mathbf{1}$, where $\mathbf{1}$ is the m-vector with all elements equal to 1. The $m \times k$ perturbation matrix $\mathbf{A}^b$ has columns $\mathbf{A}^b_i = \mathbf{X}^b_i - \mathbf{x}^b$. A similar notation applies to the analysis, only with the superscript a instead of b. The *in situ* observations are stored in the p-vector $\mathbf{y}^o$.

APPENDIX B: CORRECTION FACTORS

The correction factors aim at improving the analysis quality by giving extra weight to representative observations. The observation representativeness is assessed with respect to the

other observations in the network, without considering the background. In fact, the correction factors are used to include into the analysis those small-scale features that may not be properly resolved by the background but that are resolved by the observational network. For this reason, the background is not mentioned in this appendix.

The correction factors are computed at each analysis step, independently from the results of the other steps. This choice allows the correction factors to adapt instantaneously to weather variations. On the other hand, robust quantitative estimates require the simultaneous consideration of several consecutive time steps. To overcome this last limitation, the correction factors are allowed to take values only between the range from 0.25 to 1. We impose the most representative observations to be weighted four times as much as the less representative ones. In the following, first we will introduce γ: the relative representativeness of an observation with respect to the average observation representativeness of the observational network. Then, we will use the distribution of γ values to re-scale the correction factors within the predefined range (0.25 to 1).

To assess the observation representativeness, the definitions of large- and local-scale temperature are introduced. Consider the jth station, the large-scale temperature (y_j^L) is the temperature reconstructed at that station location using dozens of the surrounding stations. The local-scale temperature (y_j^l) is the adjusted large-scale temperature using the few closest stations only. In practice, y_j^L is obtained by fitting the temperature vertical profile proposed by Frei (2014) to the 20 nearest stations (up to 50 in data-dense areas). Depending on the observation density in the neighbourhood of the jth station, the region considered has a radius ranging from approximately 125 to 250 km. The OI Equation 15 is used to obtain y_j^l as the analyses, with y^L as the pseudo-background values. The OI parameters are set to the same values of Table 1, except for D^h that is location-dependent as it is set to the average distance between a station and its nearest five stations. For the observational network described in section 3, D^h assumes in most cases values between 50 and 70 km.

The relative representativeness of the jth observation with respect to the average observation representativeness γ_j is determined considering the deviations of the observations from the large- and local-scale temperatures. If the observation y_j^o is close either to y_j^L or y_j^l, then it is reasonable to assume that this observation represents a feature resolved by the observational network. This level of confidence will be even higher if y_j^o is close to both y^L and y^l. On the other hand, non-representative y_j^o deviates from y_j^L and y_j^l, and it may introduce unrealistic features in the analysis (e.g. bull's-eye effects). γ_j is defined as:

$$\gamma_j = \frac{(y_j^o - y_j^l)(y_j^o - y_j^L)}{\overline{\gamma}},\tag{A1}$$

with the normalization coefficient $\overline{\gamma} = \left[\sum_{n=1}^{p}(y_n^o - y_n^l)(y_n^o - y_n^L)\right]/p$.

Given the distribution of values obtained by Equation 19 for a single hour, the correction factor c_j for the jth observation is determined as ($p_{x\%}$ indicates the xth-percentile):

$$c_j = \begin{cases} 0.25, & \text{if } \gamma_j < p_{25\%} \\ 0.25 + 0.75\frac{\gamma_j - p_{25\%}}{p_{75\%} - p_{25\%}}, & \text{if } p_{25\%} \leq \gamma_j < p_{75\%} \\ 1, & \text{if } \gamma_j \geq p_{75\%} \end{cases}\tag{A2}$$