ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Corso di Laurea Magistrale in Fisica del Sistema Terra

Use of a Kalman filtering technique for near-surface temperature analysis

Relatore:
Dott. Michele Brunetti

Correlatore:
Dott. Cristian Lussana

Presentata da:
Luca Cantarello

Sessione III
Anno Accademico 2015/2016

“There is no such thing as bad weather, only bad clothes”

(norwegian saying)

Abstract

A statistical post-processing of the hourly 2-meter temperature fields from the Nordic convective-scale operational Numerical Weather Prediction model Arome MetCoOp 2.5 Km has been developed and tested at the Norwegian Meteorological Institute (MET Norway). The objective of the work is to improve the representation of the temperature close to the surface combining model data and in-situ observations for climatological and hydrological applications.

In particular, a statistical scheme based on a bias-aware Local Ensemble Transform Kalman Filter has been adapted to the spatial interpolation of surface temperature. This scheme starts from an ensemble of 2-meter temperature fields derived from Arome MetCoOp 2.5 Km and, taking into account the observations provided by the MET Norway network, produces an ensemble of analysis fields characterised by a grid spacing of 1 km. The model best estimate employed in the interpolation procedure is given by the latest avilable forecast, subsequently corrected for the model bias. The scheme has been applied off-line and the final analysis is performed independently at each grid point.

The final analysis ensemble has been evaluated and its mean value has been proved to improve significantly the best estimate of Arome MetCoOp 2.5 km in representing the 2-meter temperature fields, in terms of both accuracy and precision, with a reduction in the root mean squared values as well as in the bias and an improvement in reproducing the cold extremes during wintertime. More generally, the analysis ensemble displays better forecast verification scores, with an overall reduction in the Brier Score and its reliability component and an increase in the resolution term for the zero degrees threshold. However, the final ensemble spread remains too narrow, though not as narrow as the model output.

Riassunto

Una routine di *post-processing* statistico sui campi orari di temperatura a 2 metri prodotti dal modello operativo di previsione numerica Arome MetCoOp 2.5 Km è stata sviluppata e testata presso il Servizio Meteorologico Norvegese (MET Noway). Lo scopo dello studio è quello di migliorare la rappresentazione della temperatura in prossimità della superficie mediante l'utilizzo combinato di dati modellistici e osservazioni in loco, per applicazioni in ambito climatologico e idrologico.

In particolare, lo schema applicato è un adattamento agli scopi dell'interpolazione spaziale della temperatura al suolo di un filtro di Kalman del tipo *Local Ensemble Transform Kalman Filter* (LETKF), implementato in modo tale da valutare e rimuovere il bias del modello. Lo schema prevede l'utilizzo di un ensemble di campi di temperatura al suolo ricavato da Arome MetCoOp 2.5 Km e dei valori di temperatura osservati dalla rete di stazioni gestita da MET Norway; l'analisi finale è costituita da un ensemble di campi di temperatura a 2 metri su una griglia avente passo di 1 km. L'output del modello impiegato come miglior stima nella procedura di interpolazione è dato dal forecast disponibile più recente. Lo schema è stato applicato off-line e l'analisi finale calcolata in modo indipendente per ciascun punto griglia.

Viene mostrato come il valor medio dell'ensemble dell'analisi migliori in modo significativo la miglior stima di Arome MetCoOp 2.5 km nel rappresentare la temperatura a due metri, in termini sia di accuratezza che di precisione, con una riduzione dei valori dell'errore quadratico medio e del bias e un miglioramento nella rappresentazione degli estremi negativi durante le fasi fredde invernali. Più in generale, applicando all'ensemble dell'analisi gli score per la verifica delle previsioni probabilistiche, questa mostra migliori performance rispetto all'ensemble ricavato da Arome MetCoOp 2.5 Km, con una generale riduzione dei valori di Brier Score e della sua componente di affidabilità e un incremento del termine di risoluzione, tutti riferiti alla soglia degli zero gradi Celsius. Tuttavia, lo spread dell'ensemble dell'analisi rimane limitato, sebbene non in maniera così pronunciata quanto l'ensemble ricavato da Arome MetCoOp 2.5 Km.

Contents

Introduction

The temperature of the atmosphere close to the surface (commonly defined as surface temperature or 2-meter temperature) is one of the meteorological variables which has the most relevant impact on human activities. It is vastly employed in meteorology, hydrology and climatology, and its spatial representation is not a trivial issue, especially at a local scale of few kilometers.

The main issue to face in dealing with the surface temperature is the fact, that within the Planetary Boundary Layer (the portion of the atmosphere which is affected by the presence of the surface), the topography and the different surface properties have a relevant impact on the variability of the atmospheric variables, whose spatial scale can become small and local. Therefore, a significative degree of resolution has to be reached in order to obtain a satisfactory representation of the temperature in the proximity of the ground.

Nowadays, the models run to perform the Numerical Weather Predictions (NWP) have reached a high level of accuracy and resolution and they provide useful forecast up to seven days in the future at a synoptic scale and three days at the mesoscale [Bauer et al., 2015]. Their output is usually available on a regular grid, since the equations describing the atmospheric motions are usually discretized in order to be numerically solved. In this respect, the spatial resolution of both Local and Global area models has continuously increased over years and, at the same time, their ability to describe phenomena at finer scales has been improved through better parametrizations or advanced dynamical schemes. As a consequence, on the one hand, numerical models are now able to provide fields of meteorological variables with a high spatial resolution over wide domains, so that surface temperature can be represented in a realistic way also at small scales. On the other hand, numerical models provide only an approximate description of the atmospheric state and their accuracy and precision might not be enough to serve any application. For example, predicted temperature fields may suffer from systematic errors due to surface properties and topography misrepresentations or to even larger errors at specific locations under particular atmospheric conditions because of a too rapid or slow simulated temporal evolution of a forecasted front.

The object of this work is proving that better results can be obtained by using surface observations in support of the model output. The observations are direct measurements of the variable we are interested in, and even if they are sparsely distributed, and vast areas remain commonly uncovered even in the presence of dense measurement networks, they provide information about the state of the atmosphere of their very local surroundings, below the resolution of a typical high-resolution numerical model. Therefore, the observations can be used to add information about those scales which are not resolved by the NWP models today available.

The purpose of this thesis is to build up and test a post-processing scheme based on some statistical hypotheses to better describe the 2-meter temperature over Norway. The post-processing procedures are those techinques which perform some a-posteriori elaborations on the model output in order to improve or inquire the performances of a forecasting system.

Actually, in atmospheric sciences there is a long tradition in the combination of numerical model outcomes and in-situ observations [Daley, 1993], and we are particularly interested in two related and parallel branches of research on this topic.

On the one hand, the objective analysis of a meteorological field given point observations (e.g. the reconstruction of the geopotential at the pressure level of 500hPa by using radiosonde observations) is a classic application of statistical methods to meteorology. More specifically, our interest is in near-surface fields where geostatistical methods have been applied in the context of climatology for variable such as precipitation and temperature at different time aggregation [Brunetti et al., 2012, Brunetti et al., 2014]. On the other hand, initial conditions for numerical weather prediction (NWP) models are obtained by means of Data Assimilation (DA) schemes [Daley, 1993, Kalnay, 2003], which are applications of the Bayesian statistical methods to the problem of the model initiailisation: the model itself constitute the fundamental a-priori guess, and then the observations serve as updated and more accurate information. Several Data Assimilation procedures based on such combination between model and observations have been proposed and developed, with a relevant boost in the context of the geophysical sciences starting from the early 90s. In this regard, two big families of DA schemes can be recognised today: the sequential methods (like the Optimal Interpolation and the Kalman Filter) and the variational approaches (3D and 4D-Var) [Ide et al., 1997, Kalnay, 2003].

In recent years it has become possible to apply Bayesian statistical interpolation techniques (like those employed in the DA) also to the problem of the objective analysis of surface fields aimed at resolving small-scale processes. The three main factors which played a role in this development have been: the constant technological improvements in the measurements of temperature with automatic weather stations, the increase in the density of the station networks and the possibility of state-of-the-art NWP model to properly resolve atmospheric scales that are not too large compared to the usual station density of a meteorological network. Some previous applications of such approach can be found in [Schiemann et al., 2010, Uboldi et al., 2008] and this thesis will give a strong contribution in this direction. In this regard, the interpolation procedure we have adopted is the Local Ensemble Transform Kalman Filter (or LETKF), already proposed as a DA scheme in [Harlim and Hunt, 2005, Hunt et al., 2007]. The fundamental aspect which will be addressed is the characterization and the description of the error affecting both the models and the observation data, which are commonly considered gaussian by hypothesis in a Kalman Filter: a fair assumption for the surface temperature. Moreover, two main implementations will contribute to the originality of the work: the ensemble approach - which exploits the potentialities of the ensemble prediction systems - and the implementation within the scheme of a procedure able to perform the detection and the removal of the model bias, as suggested in [Dee and Da Silva, 1998, Dee and Todling, 2000, Dee, 2003].

The fact that we are not interested in producing any initial condition for a model to run have also some important advantages. As a matter of fact, the DA techinques have to maintain the physical and numerical consistency among the variables and the vertical levels of a model, since the initial condition has to belong to the model space itself. Conversely, we are going to work simply on the 2-meter temperature and we can focus entirely on the spatial resolution issue without any dynamical constraint which a model would pose. Furthermore, we don't either have to replicate any DA cycle, in which the observations are periodically and continuously assimilated in the model resulting in a repetitive application of the scheme: apart from the bias correction routine, in this work we run the scheme off-line and independently at each time step.

The thesis has been carried out within the Division for Climate Service of the Norwegian Meteorological Institute (MET Norway), in Oslo.

The thesis is structured as follows: in the first chapter, the theoretical building of the LETKF scheme adapted to our purposes will be illustrated in details, including the choices regarding the ensemble implementation, the localisation issue, and the bias correction algorithm. In the second chapter we will briefly present the network of observations assimilated in the scheme together with the main properties of the model employed data, originally coming from the 2-meter temperature fields of Arome MetCoOp 2.5 km; there, we will also discuss some of the scheme implementations: the sensitivity of the observation operator, the properties of the background error covariance matrix and the tuning of some important parameters. In the third chapter we will discuss the results obtained running the scheme; the evaluation will start from a case study and then it will be extended to a 6-months period. Lastly, we will sum up the content of the thesis in the conclusions, providing some future perspectives.

Chapter 1

Theoretical aspects

In this chapter, a Kalman Filtering technique aiming at combining numerical model data and in-situ two-meter temperature observations is described in details.

Kalman Filters are often used in Data Assimilation cycles to provide the initial conditions for meteorological, climatological and oceanographic numerical models [Kalnay, 2003]. The implementation of the Kalman Filter considered in this thesis is the Local Ensemble Transform Kalman Filter (LETKF) [Harlim and Hunt, 2005, Hunt et al., 2007]. However, we adapted the LETKF formulation to our application: instead of providing initial condition for a model, we aim at increasing the accuracy and the effective resolution of the model output fields by including the information provided by the station network.

The structure of the chapter is as follows: a first section where we present a summary of the methodology will be followed by an extended and detailed description of the scheme building. Firstly, we will introduce the standard Kalman Filter formulation; then, we discuss the Ensemble Kalman Filter clearifying our implementation choice to recreate the ensemble of fields (i.e. the NMC method); after that, we describe the role of both the observations and the model error in the scheme. At that point, the localization scheme will be presented, followed by the description of the *bias aware* algorithm. Then, the Spatial Consistency Test applied to the observations before entering the scheme will be discussed. The final section will serve as a recap.

1.1 A summary of the methodology

We preliminary sum up here how a Local Ensemble Transform Kalman Filter (LETKF) scheme is built and how it is adapted to our purposes.

At the very base of a LETKF there is the Kalman Filter theory. Kalman Filtering techniques are usually employed to combine two sources of data, that is, in our case, a 2-meter temperature model field defined on a regular grid (the *background*, $\mathbf{x}^f$) and a set of surface temperature *observations*, $\mathbf{y}^o$. Making some fundamental assumptions about the model and the observations (supposed to be unbiased and their respective errors gaussian), the scheme ends up with a final analysis, $\mathbf{x}^a$, still defined over the model grid. Since all the involved variables are conveniently expressed by vectors, their error covariances are in the form of matrices (the background, observations and analysis covariance matrices are usually denoted by $\mathbf{P}^f$, $\mathbf{R}$, and $\mathbf{P}^a$). The Kalman Filter is based on a linear relationship between the analysis increment, $\mathbf{x}^a - \mathbf{x}^f$, and the innovation, $\mathbf{y}^o - \mathbf{y}^f$ (where $\mathbf{y}^f$ is the forecast at the observation locations), where the proportional coefficient shaping the final analysis is the *gain* matrix, usually denoted by $\mathbf{K}$. The Kalman Filter requires also that the background error covariance matrix $\mathbf{P}^f$ is time dependent, computed advacing the previous analysis covariance matrix $\mathbf{P}^a$.

Nonetheless, a LETKF belongs to the family of the Ensemble Kalman Filters (usually shortened as EnKF). An EnKF is based on the idea that both the background and the final analysis are expressed as ensemble fields. This leads to a redefinition of the background covariance matrix $\mathbf{P}^f$ (as well as the analysis covariance matrix $\mathbf{P}^a$), because the error associated to the model (or the analysis) can be defined according to the ensemble spread. A considerable advantage in employing the ensemble approach is the possibility to reduce the dimensionality of the problem (avoiding computations in the model space), simplifying the actual realisation of the algorithm. Since at MET Norway a real Ensemble Prediction System was not available at the time we started our work, we built an ensemble collecting all the past available forecasts for a specific analysis time, retaining the most recent one as the best estimate (this is called NMC method). We also provided a covariance inflation parameter Δ, recursively updated at each time step by means of a scalar Kalman Filter, to compensate for the problem of an often too narrow ensemble spread, which is a common and known issue in the context of the Ensemble Prediction Systems.

However, a difficulty introduced by the reduction in the dimensionality is the fact that, with an ensemble of a number of members of the order of $10^0 \sim 10^1$, it is impossible to reproduce in a reliable way all the possible states in the model space, which in our case are of the order of $10^5 \sim 10^6$, like the number of grid points. That is precisely why a localization procedure is performed in the scheme (which makes the Kalman Filter "Local"). The act of localizing the analysis means performing it independentely at each grid point, selecting only a sub-set of neighbouring stations to be taken into account in the computation of the analysis. This procedure indirectly reduces the area (and thus the dimensionality) over which the model ensemble has to span the model states.

Furthermore, our scheme has been adapted in order to be able to detect and take into account the model bias. As a consequence, we have had to reject the hypothesis about the unbiased condition of the model introducing an unbiased bias prediction variable $\mathbf{b}^p$ having a covariance matrix $\mathbf{P}^b$. The algorithm has been modified accordingly to a *bias aware* scheme, where the model best estimate is recursively corrected for the updated bias, which is estimated at each time step, combining linearly the prediction $\mathbf{b}^p$ with the innovation. To make the bias prediction $\mathbf{b}^p$ explicit, a bias persistence hypothesis has been considered (i.e. the present bias prediction is the previous updated bias, apart from a damping constant μ), whereas the covariance matrix $\mathbf{P}^b$ has been defined on the basis of the background covariance matrix $\mathbf{P}^f$, imposing $\mathbf{P}^b = \gamma\mathbf{P}^f$. In this regard, γ will play the role of controlling the magnitude and the reactivity of the bias correction in the scheme.

Finally, a Spatial Consistency Test to guarantee the observations' quality before they are employed in the scheme is provided, along with a scalar Kalman Filter to update at each time step the observation variance during the scheme run.

1.2 The Local Ensemble Transform Kalman Filter in the presence of model bias

1.2.1 The standard Kalman Filter formulation

Generally speaking, Bayesian statistics provides a theoretical framework to combine multiple sources of information on a specific event. Given a particular realization of an event (i.e. an *observation*), a-priori (or *background*) information for the realization of that event is usually available as well. However, the observations and the background does not provide exactly the same information. Then, Bayesian statistics combines the a-priori information with the observed data to provide a more accurate and precise description on the event realization, which in our work we will consider as the *analysis* state.

The Kalman Filter theory is part of the Bayesian statistics. The model fields constitute the background information and the network of stations provides the two-meter temperature observations. The background information is available on a regular grid, while the point observations are unevenly distributed across the spatial domain. In combing these two different sources of data, the scheme has to deal with their different spatial representations and we need to clarify the notation in advance.

If m is the number of grid points over which the model is run, then we call "model space" the m-dimensional space within which the background field vectors $\mathbf{x}^f$[1] as well as the final analysis $\mathbf{x}^a$ are defined. Moreover, we will often refer to the vector $\mathbf{x}^t$ as the unknown discretised "true" state of the atmosphere that the real and continuos flow assumes in such a space.

Coversely, the observations are distributed in a more irregular and sparse way. If p observations are available at the analysis time, then we call "observations space" the p-dimensional space within which the observations vector $\mathbf{y}^o$ is defined. Furthermore, we define as "observation operator" the function $\boldsymbol{H}(\cdot)$ which maps a state (a vector) of the model space to the corrispondent one in the space of the observations, i.e. $\boldsymbol{H} : \mathbb{R}^m \to \mathbb{R}^p$.

If we treat $\mathbf{x}^a$, $\mathbf{x}^f$ and $\mathbf{y}^o$ as random vector variables, we can define their error vectors as:

---

[1] In the literature the background field is usually referred to with the index "b". Here we refer to it through the letter "f" in order to avoid confusion afterwards when we will deal with the forecast bias.

$$\boldsymbol{\varepsilon}^a = \mathbf{x}^a - \mathbf{x}^t, \tag{1.1}$$

$$\boldsymbol{\varepsilon}^f = \mathbf{x}^f - \mathbf{x}^t, \tag{1.2}$$

and

$$\boldsymbol{\varepsilon}^o = \mathbf{y}^o - \boldsymbol{H}(\mathbf{x}^t), \tag{1.3}$$

where $\boldsymbol{H}(\mathbf{x}^t)$ is the "true" state mapped into the observation space.

Moreover, the error covariance matrices $\mathbf{P}^a$, $\mathbf{P}^f$ and $\mathbf{R}$ are, by definition:

$$\mathbf{P}^a(m \times m) = E(\boldsymbol{\varepsilon}^a(\boldsymbol{\varepsilon}^a)^T), \tag{1.4}$$

$$\mathbf{P}^f(m \times m) = E(\boldsymbol{\varepsilon}^f(\boldsymbol{\varepsilon}^f)^T), \tag{1.5}$$

$$\mathbf{R}(p \times p) = E(\boldsymbol{\varepsilon}^o(\boldsymbol{\varepsilon}^o)^T), \tag{1.6}$$

where $E(\cdot)$ is the expected value (or first moment) operator.

Therefore, the Kalman Filter scheme combines linearly the background $\mathbf{x}^f$ (the a-priori information) together with the observations $\mathbf{y}^o$ to produce an analysis field $\mathbf{x}^a$ which constitutes the a-posteriori estimate of our field, that is:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - \boldsymbol{H}(\mathbf{x}^f)). \tag{1.7}$$

In the equation (1.7) the quantity $\mathbf{y}^o - \boldsymbol{H}(\mathbf{x}^f)$ is the so-called innovation vector, with $\boldsymbol{H}(\mathbf{x}^f)$ representing the background state mapped into the observation space. $\mathbf{K}$ is a matrix of dimension $(m \times p)$ called "gain matrix" which has to be set in order to determine the way the scheme has to weigh the "a-priori" estimate rather than the innovation.

Nonetheless, in order to apply the scheme (1.7), there are some important hypotheses which one should bear in mind. Firstly, it is a requirement that both the observation and the background errors - $\boldsymbol{\varepsilon}^o$ and $\boldsymbol{\varepsilon}^f$ - are unbiased and Gaussianly distributed, that is:

$$E(\boldsymbol{\varepsilon}^o) = \mathbf{0},\ \boldsymbol{\varepsilon}^o \sim N(\mathbf{0}, \mathbf{R}). \tag{1.8}$$

$$E(\boldsymbol{\varepsilon}^f) = \mathbf{0},\ \boldsymbol{\varepsilon}^f \sim N(\mathbf{0}, \mathbf{P}^f). \tag{1.9}$$

Furthermore, we assume that they are statistically uncorrelated:

$$E\left[(\boldsymbol{\varepsilon}^f)(\boldsymbol{\varepsilon}^o)^T\right] = 0. \tag{1.10}$$

Finally, we perform the linearization of the generally non-linear observation operator $\boldsymbol{H}(\cdot)$, and so we can approximate:

$$\boldsymbol{H}(\mathbf{x}) \simeq \mathbf{H}\mathbf{x};$$

where $\mathbf{x}$ is a generic vector of the model space and $\mathbf{H}$ is a matrix of dimension $(p \times m)$.

The gain matrix K

There are obviously infinite possible values for $\mathbf{K}$ in the equation (1.7), and no matter which value one chooses, the linearity of the scheme implies that, if both the background and the observation errors distributions are Gaussian, then also the analysis error $\boldsymbol{\varepsilon}^a$ distribution will be so. Furthermore, if both the background and the observation errors are unbiased (as stated by the hypotheses (1.8) and (1.9)), then the Kalman Filter provides an unbiased analysis (cfr. §1.2.6). In other words:

$$E(\boldsymbol{\varepsilon}^a) = \mathbf{0},\ \boldsymbol{\varepsilon}^a \sim N(\mathbf{0}, \mathbf{P}^a).$$

However, we are interested in finding the "optimal" value of $\mathbf{K}$, which is defined as the value of the gain matrix able to minimize $E((\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a)$, i.e. the mean squared error of the analysis error defined in (1.1).

Starting from (1.7) and substituting both (1.1) and (1.3) inside, we obtain:

$\mathbf{x}^t + \boldsymbol{\varepsilon}^a = \mathbf{x}^f + \mathbf{K}(\boldsymbol{H}(\mathbf{x}^t) + \boldsymbol{\varepsilon}^o - \boldsymbol{H}(\mathbf{x}^f)),$

$\boldsymbol{\varepsilon}^a = -\mathbf{x}^t + \mathbf{x}^f + \mathbf{K}(\boldsymbol{H}(\mathbf{x}^t) + \boldsymbol{\varepsilon}^o - \boldsymbol{H}(\mathbf{x}^f)).$

Now, using (1.2) and computing the square of both members we have:

$\boldsymbol{\varepsilon}^a = \boldsymbol{\varepsilon}^f + \mathbf{K}(\boldsymbol{H}(\mathbf{x}^t) + \boldsymbol{\varepsilon}^o - \boldsymbol{H}(x^f)),$

$(\boldsymbol{\varepsilon}^a)^2 = \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\boldsymbol{H}(\mathbf{x}^t) + \boldsymbol{\varepsilon}^o - \boldsymbol{H}(\mathbf{x}^f))\right)^2,$

$(\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a = \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\boldsymbol{H}(\mathbf{x}^t) + \boldsymbol{\varepsilon}^o - \boldsymbol{H}(\mathbf{x}^f))\right)^T \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\boldsymbol{H}(\mathbf{x}^t) + \boldsymbol{\varepsilon}^o - \boldsymbol{H}(\mathbf{x}^f))\right).$

We have previously seen that - by hypothesis - $\boldsymbol{H}(\cdot)$ can be linearized into a matrix $\mathbf{H}$. This assumption is fairly justifiable in our case, since the observations operator consists only in a regridding process which interpolates the background field over the grid points; however, this is not true in general and we will return on this issue in the section §2.2.

Expliciting $\mathbf{x}^f$ as in (1.2), we have:

$(\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a = \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o - \mathbf{H}\mathbf{x}^f)\right)^T \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o - \mathbf{H}\mathbf{x}^f)\right),$

$(\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a = \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o - \mathbf{H}(\mathbf{x}^t + \boldsymbol{\varepsilon}^f))\right)^T \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o - \mathbf{H}(\mathbf{x}^t + \boldsymbol{\varepsilon}^f))\right),$

$(\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a = \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o - \mathbf{H}\mathbf{x}^t - \mathbf{H}\boldsymbol{\varepsilon}^f)\right)^T \left(\boldsymbol{\varepsilon}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o - \mathbf{H}\mathbf{x}^t - \mathbf{H}\boldsymbol{\varepsilon}^f)\right),$

$(\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a = \left(\boldsymbol{\varepsilon}^f - \mathbf{K}(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)\right)^T \left(\boldsymbol{\varepsilon}^f - \mathbf{K}(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)\right).$

Now we have to minimize the expected (i.e. mean) value of the quantity $(\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a$ with respect to the matrix $\mathbf{K}$, imposing:

$\frac{d}{d\mathbf{K}}E\left[(\boldsymbol{\varepsilon}^a)^T\boldsymbol{\varepsilon}^a\right] = 0,$

that is, expliciting $\boldsymbol{\varepsilon}^a$:

$\frac{d}{d\mathbf{K}}E\left[(\boldsymbol{\varepsilon}^f)^T\boldsymbol{\varepsilon}^f - (\boldsymbol{\varepsilon}^f)^T\mathbf{K}(\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o) - (\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)^T\mathbf{K}^T\boldsymbol{\varepsilon}^f + (\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)^T\mathbf{K}^T\mathbf{K}(\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)\right] = 0.$

Using the rules shown in the appendix to derive an expression with respect to a matrix, we obtain:

$-E(\boldsymbol{\varepsilon}^f(\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)^T) - E(\boldsymbol{\varepsilon}^f(\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)^T) + 2\mathbf{K}E((\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)(\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)^T) = 0,$

$-E(\boldsymbol{\varepsilon}^f(\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)^T) + \mathbf{K}E((\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)(\mathbf{H}\boldsymbol{\varepsilon}^f + \boldsymbol{\varepsilon}^o)^T) = 0.$

Finally we can explicit $\mathbf{K}$ and perform the product:

$\mathbf{K} = E(\boldsymbol{\varepsilon}^f(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)^T)\left(E(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)^T)\right)^{-1},$

$\mathbf{K} = E(\boldsymbol{\varepsilon}^f(\boldsymbol{\varepsilon}^f)^T\mathbf{H}^T - \boldsymbol{\varepsilon}^f(\boldsymbol{\varepsilon}^o)^T)\left(E(\mathbf{H}\boldsymbol{\varepsilon}^f(\boldsymbol{\varepsilon}^f)^T\mathbf{H}^T - \boldsymbol{\varepsilon}^o(\boldsymbol{\varepsilon}^f)^T\mathbf{H}^T - \mathbf{H}\boldsymbol{\varepsilon}^f(\boldsymbol{\varepsilon}^o)^T + \boldsymbol{\varepsilon}^o(\boldsymbol{\varepsilon}^o)^T)\right)^{-1}.$

Exploiting the linearity of the operator $E(\cdot)$ and using both the definitions (1.5) and (1.6) we eventually reach:

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{1.11}$$

The equation (1.11) highlights the importance of the specification of the covariance matrices $\mathbf{P}^f$ and $\mathbf{R}$. This is generally not trivial particularly because these matrices are generally very large to compute. A lot of covariance modeling have been proposed to deal with this problem and a great deal of literature can be found on it (a good review can be found in [Fisher, 2003]).

In this regard, one of the most original feature implemented by the Kalman Filter theory in the context of the Data Assimilation was the fact that it took into account the variation in time of the background error. That is because it evolves the uncertainty through the model at each time step, instead of providing a constant matrix for $\mathbf{P}^f$ as, for example, the Optimal Interpolation scheme does. However, this approach is very computationally expensive since it requires the so-called "tangent linear model" to be defined and applied at each time step and forces the scheme to be splitted in a two-step procedure [Kalnay, 2003]. In the section §1.2.2 we will show how the choice of the ensemble approach will be able to simplify the procedure boosting the computational efficency but retaining the covariance flow-dependence. Later in the chapter we will also discuss our model for the matrix $\mathbf{R}$.

The other important aspect about the matrix $\mathbf{K}$ in (1.11) which is worth pointing out is that its role in the scheme is to weigh the errors of the forecast and the observations. If, for example, the error on the observation is big, then the terms in the matrix $\mathbf{R}$ are so and, as a consequence, the matrix $\mathbf{K}$ will produce small or no adjustments at all to the background, i.e. $\mathbf{x}^a \approx \mathbf{x}^f$. In other words, the scheme will rely more on the model because of the big uncertainty affecting the measurements.

Analysis error covariance matrix $\mathbf{P}^a$ computation

Together with the the analysis field $\mathbf{x}^a$, we aim to describe also how the covariance of the analysis error looks like. Thus, we are interested in computing also the matrix $\mathbf{P}^a$, previously defined in (1.4).

Recovering the definition of $E(\boldsymbol{\varepsilon}^a(\boldsymbol{\varepsilon}^a)^T)$, we have:
$\mathbf{P}^a = E(\boldsymbol{\varepsilon}^f(\boldsymbol{\varepsilon}^f)^T - \mathbf{K}(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)(\boldsymbol{\varepsilon}^f)^T - \boldsymbol{\varepsilon}^f(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)^T\mathbf{K}^T + \mathbf{K}(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)(\mathbf{H}\boldsymbol{\varepsilon}^f - \boldsymbol{\varepsilon}^o)^T\mathbf{K}^T)$;
$\mathbf{P}^a = \mathbf{P}^f - \mathbf{KHP}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{KRK}^T + \mathbf{KHP}^f\mathbf{H}^T\mathbf{K}^T$;
$\mathbf{P}^a = \mathbf{P}^f - \mathbf{KHP}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{K}(\mathbf{R} + \mathbf{HP}^f\mathbf{H}^T)\mathbf{K}^T$.

Now replacing the $\mathbf{K}$ of the fourth term on the right side with (1.11), we obtain:
$\mathbf{P}^a = \mathbf{P}^f - \mathbf{KHP}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}(\mathbf{R} + \mathbf{HP}^f\mathbf{H}^T)\mathbf{K}^T$;
$\mathbf{P}^a = \mathbf{P}^f - \mathbf{KHP}^f - \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T + \mathbf{P}^f\mathbf{H}^T\mathbf{K}^T$;

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f. \tag{1.12}$$

Again, if the matrix $\mathbf{K}$ is small (due to a great deal of uncertainty on observations) the scheme will rely totally on the model and $\mathbf{P}^a \approx \mathbf{P}^f$.

1.2.2 The Ensemble Kalman Filter (EnKF)

The success of the Ensemble Prediction Systems (EPS) in the Numerical Weather Predictions (NWP) has encouraged the development of techniques using analog approaches also in the Data Assimilation context, such as the Ensemble Kalman Filter (or EnKF), in order to provide numerical models with ensembles of initial conditions. The LETKF scheme applied in this study is one of the possible formulation of an EnKF.

In the EnKF framework, the single background field is replaced by a set of k model forecasts $\{\mathbf{x}^{f\,(i)}\ i = 1, ..., k\}$. In a Data Assimilation cycle, each member of the ensemble is independently evolved forward until the analysis time is reached. Then, the observations $\mathbf{y}^o$ are used to adjust directly the background ensemble mean $\overline{\mathbf{x}}^f$[2]:

$$\overline{\mathbf{x}}^f = \frac{1}{k}\sum_{i=1}^{k}\mathbf{x}^{f\,(i)}$$

which is usually assumed as the *a-priori* guess state. The scheme provides an analysis field $\mathbf{x}^a$ which, for convenience, can be interpreted in turn as the mean value $\overline{\mathbf{x}}^a$ of a set of k ensemble members $\{\mathbf{x}^{a\,(i)}\ i = 1, ..., k\}$:

$$\overline{\mathbf{x}}^a = \frac{1}{k}\sum_{i=1}^{k}\mathbf{x}^{a\,(i)}.$$

---

[2] From now on, $(\bar{\cdot})$ will always indicate the ensemble mean.

The need for both a forecast and an analysis ensemble is essential in a Data Assimilation cycle, since in a EPS the model produces an ensemble of outcomes and, at the same time, it has to be initialized with a new one to be run. For our purpose the ensemble approach might seem less essential, since the application of (1.7) is independently performed at each time step. However, an operational EPS will be available at MET Norway in the future and our scheme aims to exploit all its advantages. Moreover, preserving a set of analysis ensemble members as final product, allows us to retain information about the uncertainty affecting the analysis without saving the entire matrix $\mathbf{P}^a$, which would be very computationally expensive.

The background error covariance matrix $\mathbf{P}^f$ which we have previously defined in (1.5), can now be rewirtten in terms of the uncertainty of the background, i.e. the spread of the ensemble around the mean $\overline{\mathbf{x}}^f$, that is:

$$\mathbf{P}^f(m \times m) = \frac{1}{k-1}\mathbf{X}^f(\mathbf{X}^f)^T, \tag{1.13}$$

where $\mathbf{X}^f$ is the background analysis perturbation matrix of dimension $(m \times k)$, whose i-th column is:

$$\mathbf{X}^{f\,(i)} = \mathbf{x}^{f\,(i)} - \overline{\mathbf{x}}^f. \tag{1.14}$$

The equation (1.13) shows how the background covariance matrix usually looks like in an EnKF and it has the advantage of being potentially simple to built, since the only need is the set of forecasts $\{\mathbf{x}^{f\,(i)}\ i = 1, ..., k\}$ with no additional complex covariance model to be formulated. Furthermore, since the ensemble members are individually evolved between one analysis time and the following one, this approach also retains the flow-dependence property of the uncertainty, which is intrinsically expressed by the matrix $\mathbf{P}^f$.

The uncertainty affecting the analysis can be similarly expressed in terms of a new background error covariance matrix $\mathbf{P}^a$, that is:

$$\mathbf{P}^a(m \times m) = \frac{1}{k-1}\mathbf{X}^a(\mathbf{X}^a)^T, \tag{1.15}$$

where $\mathbf{X}^a$ is again the analysis ensemble perturbation matrix of dimension $(m \times k)$, whose i-th columns are defined as in (1.16):

$$\mathbf{X}^{a\,(i)} = \mathbf{x}^{a\,(i)} - \overline{\mathbf{x}}^a. \tag{1.16}$$

According to (1.7), the analysis can then be expressed as:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{K}(\mathbf{y}^o - \boldsymbol{H}(\overline{\mathbf{x}}^f)). \tag{1.17}$$

At this point we can also highlight the fact that the availability of an ensemble allows us to treat our problem in a dimensionally reduced context. As a matter of fact, the number of the members of a typical EPS, which defines the dimension of the ensemble space, is far smaller than the degrees of freedom of a NWP model. This means that, in our particular case, the number of the ensemble members k is far smaller than the numbers of the grip points m (and also than the number of the observations p). In equation (1.17), we are primarily interested in the matrix $\mathbf{K}$, which has been previously defined in (1.11) and normally requires operations on matrices of dimension $(m \times m)$, $(p \times m)$ and $(p \times p)$. However, since by definition the maximum allowed rank for both the matrices $\mathbf{X}^f$ and $\mathbf{X}^a$ in eqs. (1.13) and (1.15) is $k-1$, then also the rank of both $\mathbf{P}^f$ and $\mathbf{P}^a$ is so, therefore we expect a dimensionally-reduced path to compute $\mathbf{K}$, avoiding at all operations in the model space (i.e. the entire matrix $\mathbf{P}^f$).

If we interpret $\mathbf{X}^f$ as a linear operator which maps a vector of the ensemble space into the model one, we can entirely perform the analysis in the former (as well as in the observation one). To this aim, we replicate here the procedure contained in [Harlim and Hunt, 2005], starting by expliciting the definition of the optimal gain matrix $\mathbf{K}$ according to the definition of $\mathbf{P}^f$ in (1.13). If we substitute it in (1.11) we have:

$\mathbf{K} = \frac{1}{k-1}\mathbf{X}^f(\mathbf{X}^f)^T\mathbf{H}^T\left[\frac{1}{k-1}\mathbf{H}\mathbf{X}^f(\mathbf{X}^f)^T\mathbf{H}^T + \mathbf{R}\right]^{-1}$.

In the above expression we can avoid to operate in the model space if we define the background ensemble observation perturbations as the matrix $\mathbf{Y}^f = \mathbf{H}\mathbf{X}^f$. The i-th column of $\mathbf{H}\mathbf{X}^f$ can be expressed as $\mathbf{H}(\mathbf{x}^{f\,(i)} - \overline{\mathbf{x}}^f)$, and we can interpret it as the first order approximation of a Taylor series for the function $\boldsymbol{H}(\mathbf{x}^{f\,(i)})$ evaluated in $\overline{\mathbf{x}}^f$, i.e. $\boldsymbol{H}(\mathbf{x}^{f\,(i)}) - \boldsymbol{H}(\overline{\mathbf{x}}^f) = \mathbf{H}(\mathbf{x}^{f\,(i)} - \overline{\mathbf{x}}^f)$. Therefore, we can linearly approximate $\mathbf{H}\mathbf{X}^f$ with $\mathbf{Y}^f$, whose i-th column will be $\boldsymbol{H}(\mathbf{x}^{f\,(i)}) - \overline{\mathbf{y}}^f$ and where $\overline{\mathbf{y}}^f = \frac{1}{k}\sum_{i=1}^{k}\boldsymbol{H}(\mathbf{x}^{f\,(i)})$. Now, if we define the variable $\mathbf{A}$:

$\mathbf{A} = \frac{1}{(k-1)^{\frac{1}{2}}}\mathbf{Y}^f$,

then we can rewrite the right member of the above equation according to the matrix identity:

$\mathbf{A}^T(\mathbf{A}\mathbf{A}^{\mathbf{T}} + \mathbf{R})^{-1} = (\mathbf{I} + \mathbf{A}^T\mathbf{R}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{R}^{-1}$,

obtaining:

$$\mathbf{K} = \frac{1}{k-1}\mathbf{X}^f\left[\mathbf{I} + \frac{1}{k-1}(\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}(\mathbf{Y}^f)^T\mathbf{R}^{-1} \tag{1.18}$$

Now, in order to get the expression for $\mathbf{P}^a$ valid in an EnKF scheme, we have to substitute (1.13) in (1.12). Therefore we have:

$\mathbf{P}^a = \frac{1}{k-1}(\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{X}^f(\mathbf{X}^f)^T$;

then, using the previously obtained expression (1.18) for $\mathbf{K}$, we obtain:

$\mathbf{P}^a = \frac{1}{k-1}\mathbf{X}^f(\mathbf{I} - \frac{1}{k-1}\left[\mathbf{I} + \frac{1}{k-1}(\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}(\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f)(\mathbf{X}^f)^T$.

If we now define $\mathbf{B} = \frac{1}{k-1}(\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f$, we can exploit the matrix identity $\mathbf{I} - (\mathbf{I} + \mathbf{B})^{-1}\mathbf{B} = (\mathbf{I} + \mathbf{B})^{-1}$ to simplify the above expression. Then it becomes:

$\mathbf{P}^a = \frac{1}{k-1}\mathbf{X}^f\left[\mathbf{I} + \frac{1}{k-1}(\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}(\mathbf{X}^f)^T$;

$\mathbf{P}^a = \mathbf{X}^f\left[(k-1)\mathbf{I} + (\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}(\mathbf{X}^f)^T$.

Given the definition of the operator $\mathbf{X}^f$, in the above equation we can identify as:

$\tilde{\mathbf{P}}^a(k \times k) = \left[(k-1)\mathbf{I} + (\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}$

the analysis error covariance matrix in the ensemble space, and then we can write:

$$\mathbf{P}^a = \mathbf{X}^f\tilde{\mathbf{P}}^a(\mathbf{X}^f)^T. \tag{1.19}$$

If we compare this result with the expression (1.18), we can note that $\mathbf{K}$ can be rewritten in terms of $\tilde{\mathbf{P}}^a$:

$\mathbf{K} = \mathbf{X}^f\tilde{\mathbf{P}}^a(\mathbf{Y}^f)^T\mathbf{R}^{-1}$.

Therefore, we can transfer the new expression for $\mathbf{K}$ into (1.17):

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}^f\tilde{\mathbf{P}}^a(\mathbf{Y}^f)^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{y}^f), \tag{1.20}$$

where $\mathbf{y}^f = \boldsymbol{H}(\overline{\mathbf{x}}^f)$.

The equation (1.20) allows us to avoid operations in the model space. We can rewrite its right member defining the weight vector $\mathbf{w}^a$:

$\mathbf{w}^a = \tilde{\mathbf{P}}^a(\mathbf{Y}^f)^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{y}^f)$,

which can be interpreted as the analysis increment in the ensemble space.

The equation (1.20) then becomes:

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}^f\mathbf{w}^a. \tag{1.21}$$

At this point we have only to build the analysis ensemble members $\{\mathbf{x}^{a\,(i)},\ i = 1, ..., k\}$ starting from the analysis mean $\overline{\mathbf{x}}^a$. Therefore, we need to find the matrix $\mathbf{X}^a$ as defined by its columns in (1.16).

In order to compute the ensemble members, we start defining the matrix $\mathbf{X}^a$ as:

$$\mathbf{X}^a = \mathbf{X}^f\mathbf{W}^a. \tag{1.22}$$

If we substitute (1.22) in (1.15), we have:

$\mathbf{P}^a(m \times m) = \frac{1}{k-1}\mathbf{X}^f\mathbf{W}^a\left(\mathbf{X}^f\mathbf{W}^a\right)^T = \frac{1}{k-1}\mathbf{X}^f\mathbf{W}^a(\mathbf{W}^a)^T(\mathbf{X}^f)^T$.

Then, comparing the above equation with (1.19) we can define:

$\mathbf{W}^a(\mathbf{W}^a)^T = (k-1)\tilde{\mathbf{P}}^a$,

which is the same as:

$$\mathbf{W}^a(k \times k) = ((k-1)\tilde{\mathbf{P}}^a)^{\frac{1}{2}}. \tag{1.23}$$

There are two main reasons proving the fact that (1.23) is a good definition for $\mathbf{W}^a$. The first one is that the sum of the column of the so-defined matrix $\mathbf{X}^a$ is zero[3] - granting the right sample mean for the analysis ensemble. Secondly, the fact that in (1.23) $\mathbf{W}^a$ depends continuosly

---

[3] $\mathbf{X}^a\mathbf{v} = \mathbf{X}^f\mathbf{W}^a\mathbf{v} = \mathbf{X}^f\mathbf{v} = \mathbf{0}$ because $\mathbf{v} = (1, 1, ...1)^T$ is eigenvector of $\mathbf{W}^a$ as defined in (1.23) [Wang et al. 2004].

on $\tilde{\mathbf{P}}^a$ has a positive effect on the smoothness of the analysis ensemble members in (1.22), since the matrix $\tilde{\mathbf{P}}^a$ will be computed independentely at each grid point as the localization technique will be introduced (cfr. §1.2.5).

Once the matrix $\mathbf{W}^a$ is specified and computed, the ensemble members are then obtained simply performing:

$\mathbf{x}^{a\,(i)} = \mathbf{X}^{a\,(i)} - \overline{\mathbf{x}}^a$.

The NMC method and the redefinition of the best estimate

As we have discussed so far, the Local Ensemble Transform Kalman Filter requires the building of an ensemble of k background fields $\{\mathbf{x}^{f\,(i)}\ i = 1, ..., k\}$ in order to be applied.

Since at MET Norway a real EPS was not operational when this study started[4], we have built the ensemble of background fields adopting the same principles as those of the so-called NMC method, which was introduced in 1992 by [Parrish and Derber, 1992] and that has been employed also at the European Centre for Medium-range Weather Forecast until September 1999 [Fisher, 2003]. The NMC method exploit operational forecasts valid at the same time but initialised in different moments.

Thus, in order to build the matrix $\mathbf{X}^f$ (and then the matrix $\mathbf{P}^f$), for each hour of the day we have gathered all the available predicted surface temperature fields (run with Arome MetCoop 2.5 km) initialiased at different times in the past. This set of forecasts constitute our ensemble of background fields. Since Arome is run up to 66 hours onward at each synoptic hour (00, 06, 12 ,18 UTC) every day, we get a maximum of 11 background fields available for each hour of the day (12 at synoptic hours).

As regards the *a priori* estimate, we have previously seen that, in the context of the EnKF, the typical choice for the first guess in eq. (1.17) is the ensemble mean. Since our ensemble is composed by forecasts realised at different initialisation times, we cannot assume that they represent the surface temperature all with the same accuracy and reliability. As a consequence, we employ the latest forecast (that one launched at the closest synoptic hour in that day) as our best estimate:

$$\mathbf{x}^f_{BE} = \mathbf{x}^{f\,(1)}. \tag{1.24}$$

Then, we shift the uncertainty computed with the ensemble members around the latest forecast.

According to the equation (1.7), the analysis field can be then expressed as:

$$\overline{\mathbf{x}}^a = \mathbf{x}^f_{BE} + \mathbf{K}(\mathbf{y}^o - \boldsymbol{H}(\mathbf{x}^f_{BE})). \tag{1.25}$$

---

[4] The MetCoOp Ensemble Prediction System has become operational during Novembe 2016.

1.2.3 The representativeness error and the observations error covariance matrix R

Since the observation error covariance matrix $\mathbf{R}$ plays a fundamental role in our scheme, some clarifications about the meaning of the observational error should be pointed out. In order to better interpret $\boldsymbol{\varepsilon}^o$ - as defined in the eq. (1.3) - we can rewrite $\mathbf{y}^o$ as:

$$\mathbf{y}^o = \boldsymbol{H}(\mathbf{x}^t) + \boldsymbol{\varepsilon}^o_{ins} + \boldsymbol{\varepsilon}^o_{rep}.$$

That is, $\boldsymbol{\varepsilon}^o$ can be splitted into an instrument component and a representativeness one. The former, $\boldsymbol{\varepsilon}^o_{ins}$, is the uncertainty which in physics is typically associated to a measurement; the latter, $\boldsymbol{\varepsilon}^o_{rep}$, comes from the application of the observation operator $\boldsymbol{H}(\cdot)$ to the vector $\mathbf{x}^t$ [Lorenc, 1986]. As a matter of fact, the "true" atmospheric state $\mathbf{x}^t$ is the discretisation of the real continuous temperature field over a grid of specified step, and its ability to resolve certain spatial scales rather then others is inherentily due to the density of the grid points (which in a numerical model depends in turn on the physics it has been built upon). The act of mapping the true atmospheric state from the model space into the observations one inevitably introduces an error because of the impossibility of reproducing at the station locations features which belong to spatial scales not represented in the starting field $\mathbf{x}^t$. In this sense, $\boldsymbol{H}(\mathbf{x}^t)$ can only approximate the value of $\mathbf{y}^o$ at the station location, and a representativeness error has to be taken into consideration. Usually, this error is far bigger than the measurement one, so $\boldsymbol{\varepsilon}^o$ can be interpreted as mostly a representativity issue [Lussana et al., 2010].

With the reasonable assumption that these errors are independent (i.e. the mistake one makes in applying $\boldsymbol{H}(\cdot)$ at a specific station location is not related to that affecting another station nearby or far away), we expect the matrix $\mathbf{R}$ to be diagonal. In particular, we expect that it reads as:

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & & \sigma_p^2 \end{bmatrix},$$

where σ_i^2 is the variance of the i-th observation.

Since the number of available observations at each time step is relatively small, the istantaneous σ_i^2 value could be affected by sampling error. Then, the variance σ_i^2 is computed as in [Li et al., 2009], that is introducing a temporal smoothing technique by means of a scalar version of the Kalman Filter which updates the observation error with time. Then, unlike the analysis, the observation error in our scheme is computed at each time step and updated recursively.

The observation error $(\sigma_o^2)_{t_i}$ at each analysis time step t_i can be computed as (see the consistency diagnostic on observations error in [Desroziers et al., 2005]):

$$(\sigma_o^2)_{t_i} = \frac{(\mathbf{y}^o - \mathbf{y}^a)^T(\mathbf{y}^o - \mathbf{y}^f)}{s_{t_i}}, \tag{1.26}$$

where $\mathbf{y}^f = H(\mathbf{x}^f)$, $\mathbf{y}^a = H(\mathbf{x}^a)$, and s_{t_i}indicates the number of available observations at the instant t_i.

However, in order to better characterize the variance affecting the single i-th station at a generic t_i istant, we can multiply (1.26) for a station-dependent coefficient, that is: $(\sigma_i^2)_{t_i} = (c_i)_{t_i} \cdot (\sigma_o^2)_{t_i}$.

The initial value $(\sigma_i^2)_{t_0}$ as well as the coefficients $(\overline{c_i})_{t_i}$ are estimated from the seNorge2 database (see Appendix 2).

The $\mathbf{R}(t_i)$ matrix valid at each time step t_i can then be explicited as:

$$\mathbf{R}(t_i) = (\overline{\sigma^2})_{t_{i-1}} \begin{bmatrix} (\overline{c_1})_{t_{i-1}} & 0 & \cdots & \cdots & 0 \\ 0 & (\overline{c_2})_{t_{i-1}} & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & & (\overline{c_s})_{t_{i-1}} \end{bmatrix}.$$

The fact that the $\mathbf{R}(t_i)$ matrix depends on quantities defined at step t_{i-1} is due to the the fact that the equation (1.26) can be applied only after the analysis computation.

The temporal smoothing consists on the update at each analysis time of the “observed” value of the variance $(\sigma_i^2)^{\mathrm{o}}_{t_{i-1}}$ employing a scalar version of the Kalman Filter:

$$(\sigma_i^2)^{\mathrm{a}}_{t_i} = \frac{\upsilon^{\mathrm{o}}(\sigma_i^2)^{\mathrm{p}}_{t_i} + \upsilon^{\mathrm{f}}(\sigma_i^2)^{\mathrm{o}}_{t_i}}{\upsilon^{\mathrm{o}} + \upsilon^{\mathrm{f}}}; \tag{1.27}$$

where $(\sigma_i^2)^{\mathrm{a}}_{t_i}$ indicates the updated “analysis” value of the variance starting from a persistence model which provides the predicted value:

$$(\sigma_i^2)^{\mathrm{p}}_{t_i} = (\sigma_i^2)^{\mathrm{a}}_{t_{i-1}}.$$

The variance υ^{f} in (1.27) is defined as:

$$\upsilon^{\mathrm{f}}_{t_{i+1}} = \kappa \upsilon^{\mathrm{a}}_{t_i},$$

where

$$\upsilon^{\mathrm{a}} = \left(1 - \frac{\upsilon^{\mathrm{f}}}{\upsilon^{\mathrm{f}} + \upsilon^{\mathrm{o}}}\right) \upsilon^{\mathrm{f}}.$$

The values of κ and υ^{o} have been set to $\kappa = 1.03$ and $\upsilon^{\mathrm{o}} = 1$ as in [Li et al., 2009].

This method allows the scheme to adjust the observation error on itself, starting from a predetermined value. The scalar Kalman Filter aims to reduce the impact of possible wrong estimations of the observation variance due to little sample, and to retain possible fluctuation in time.

1.2.4 The background covariance inflation

We have seen in the section §1.2.2 that in the EnKF context the background error covariance matrix $\mathbf{P}^f$ is essentially defined by the ensemble spread. However, it is a well-known issue in the Data Assimilation research the fact that the ensemble spread produced by an EPS system is usually too small to be employed to produce reliable initial conditions, with a potential negative impact on the quality of the forecasts. This underestimation of the forecast error is then usually addressed inflating the covariance matrix $\mathbf{P}^f$. Plenty of possible solutions have been developed, but for our scheme we choose to stick with the solutions proposed in [Li et al., 2009], where the temporal smoothing of the observational error (see previous section) is performed simultaneously together with the covariance inflation for the matrix $\mathbf{P}^f$.

Inflating the matrix $\mathbf{P}^f$ requires an inflation factor Δ - whose value needs to be tuned - to be introduced, such that:

$$\mathbf{P}^f = \Delta \cdot \mathbf{P}^f_e. \tag{1.28}$$

In order to make the tuning of Δ less arbitrary as possible, we link it to a quantity ϵ defined as the ratio between the observation variance and that of the forecast:

$$\epsilon = \frac{\sigma_o^2}{\sigma_f^2},$$

forcing ϵ to assume only the values in $\{0.1,\ 0.2,\ 0.3,\ 0.4\}$.

These values are all smaller than $\epsilon = 1$, in order to impose the observation variance to be smaller than that of the model. As a matter of fact, we aim our scheme to add information effectively with respect to the model output, and this is possible only if it relies more on the observations (i.e. their variance is smaller than the model's one), as we have seen when we comment the role of the gain matrix $\mathbf{K}$ in the section §1.2.1.

If we define σ_f^2 as the mean value of the model variance mapped in the observation space, we can write:

$$\sigma_f^2 = \frac{1}{s} \cdot Tr(\mathbf{H}\mathbf{P}^f\mathbf{H}^T) = \Delta \cdot \frac{1}{s} \cdot Tr(\mathbf{H}\mathbf{P}^f_e\mathbf{H}^T).$$

Therefore, Δ can be written as (using the definition of ϵ):

$$\Delta = \frac{\sigma_f^2}{\frac{1}{s} \cdot Tr(\mathbf{H}\mathbf{P}_e^f\mathbf{H}^T)} = \frac{\sigma_o^2}{\epsilon \cdot \frac{1}{s} \cdot Tr(\mathbf{H}\mathbf{P}_e^f\mathbf{H}^T)}.$$

Since the variance of both the observations and the model can vary a lot from one hour to the following one, we perform the same temporal smoothing (i.e. a scalar Kalman Filter) we introduced in the previous section also for Δ.

Then we compute online:

$$(\Delta)_t^{\mathrm{a}} = \frac{\upsilon^{\mathrm{o}}(\Delta)_t^{\mathrm{p}} + \upsilon^{\mathrm{f}}(\Delta)_t^{\mathrm{o}}}{\upsilon^{\mathrm{o}} + \upsilon^{\mathrm{f}}},$$

where:

$$(\Delta)_t^{\mathrm{o}} = \Delta(\epsilon(min(CV_{score})),$$

and holding the persistency hypothesis:

$$(\Delta)_t^{\mathrm{p}} = (\Delta)_{t-1}^{\mathrm{a}}.$$

The Cross Validated score (CV_{score}) is defined as:

$$CV_{score} = \sqrt{\frac{1}{s} \cdot \sum_{i=1}^{s} (y_i^o - \check{y}_i^a)^2},$$

where the Cross Validated analysis $\check{y}_i^a$ is the value of the analysis computed at the i-th station location without taking into account the observed value in that point. The CV residual $y_i^o - \check{y}_i^a$ is a way to evaluate the quality of the analysis itself, because it provides an estimate of how much the scheme is able to reproduce the observed value even in the absence of the local constraint given by that observation.

Therefore, this algorithm tunes the inflation parameter Δ selecting the value of ϵ which minimizes the CV_{score} (pursuing the best possible analysis value) and then it is averaged with the previous value.

υ^{o}, υ^{f} and υ^{a} are the same as defined in the section §1.2.3.

1.2.5 The localisation technique

As we have previously noted, since the number of the ensemble members k is usually far smaller than that of the grid points m ($k \ll m$), the adoption of the ensemble approach in a Kalman Filter scheme (cfr. §1.2.2) is able to simplify the problem. On the other hand, it also limits the possibility of representing globally in a reliable way all the possible system states. As a matter of fact, the background error covariance matrix $\mathbf{P}^f$ as redefined in (1.13) has rank at most $k-1$.

This means that it can describe the uncertainty in a m-dimensional grid only in a space spanned by at most $(k-1)$ vectors. Since $\mathbf{K}$ depends on $\mathbf{P}^f$, this has obvious repercussions also on the quality of the final analysis $\overline{\mathbf{x}}^a$.

A possible way to deal with this problem is to make the analysis local instead of global. That is, the analysis is computed independentely at each grid point and only a reduced number of observations close to the grid point itself are considered. In this way, the range of the possible system states to be represented is reduced by the restricted domain over which the analysis is computed.

Many localisation techniques can be found in the literature. In this study, we adopt the approach suggested by [Greybush et al., 2011] and apply the localisation directly to the observation error covariance matrix $\mathbf{R}$. As a matter of fact, looking at the gain matrix $\mathbf{K}$ in (1.18), only $\mathbf{R}$ plays a role in the model space. Then, a localisation function f_{Rloc} such as:

$$f_{Rloc} = \exp\left[-\frac{1}{2}\left(\frac{d_h(i,j)^2}{(D_h)^2} + \frac{d_z(i,j)^2}{(D_z)^2}\right)\right] \tag{1.29}$$

is applied to the elements of the error covariance matrix $\mathbf{R}$ in order to gradually reduce the influence of the observations j far from the i-th grid point. In (1.29) the values of the parameters D_h and D_z have to be set in order to define a de-localisation distance (both horizontal and vertical) beyond which the variance associated to the j-th observation on the analysis at the i-th grid point is reduced by a factor $\frac{1}{e}$.

Therefore, at each grid point, a new matrix $\mathbf{R}_{loc}$ containing the diagonal terms shaped by the localisation function f_{Rloc} will apply. That is:

$$\mathbf{R}_{loc} = f_{Rloc} \cdot \mathbf{R}.$$

1.2.6 Estimation and correction of the model bias

One of the most original aspect of this study is the implementation of the model bias correction within the LETFK scheme.

So far we have described a *bias-blind* scheme, in which both the forecast and observations errors were considered unbiased. Unfortunately, this is not a realistic hypothesis. In particular, as a great deal of literature on Data Assimilation has shown ([Dee and Da Silva, 1998, Dee and Todling, 2000, Dee, 2003]), the presence of the bias in the model forecasts does affect the quality of the analysis, and so it is good practice trying to remove it.

Unlike the procedure used to compute the analysis, the algorithm which will be described here provides a running estimate for the bias, updating its value at each time step t_i using information about the bias at time t_{i-1}. In introducing the time dependence for the vectors, matrices and operators, we will refer to the notation as defined in [Ide et al., 1997].

Nonetheless, we will retain the assumption that the observations are unbiased, or at least that their systematic error is small compared to the model one and so it can be neglected.

In taking into account the forecast bias, we are basically removing the hypothesis (1.9). We can conveniently redefine the background field $x^f(t_i)$ splitting its uncertainty $\boldsymbol{\varepsilon}^f(t_i)$ into a zero mean component error $\tilde{\boldsymbol{\varepsilon}}^f(t_i)$ and the bias $\mathbf{b}^f$:

$$\mathbf{x}^f(t_i) = \mathbf{x}^t(t_i) + \boldsymbol{\varepsilon}^f(t_i) = \mathbf{x}^t(t_i) + \tilde{\boldsymbol{\varepsilon}}^f(t_i) + \mathbf{b}^f. \tag{1.30}$$

The forecast bias is then defined as the non-zero mean forecast error $\mathbf{b}^f$:

$$\mathbf{b}^f = E(\boldsymbol{\varepsilon}^f(t_i)) = E(\mathbf{x}^f(t_i) - \mathbf{x}^t(t_i)) \neq 0. \tag{1.31}$$

The background bias (1.31) is usually due to the presence of systematic errors in the model and it tends to show persistence in time and space. That is, it is usually a slowly varying quantity with a well defined spatial pattern which is usually inferred by comparing the observed climatology with that predicted by the model.

Similarly to (1.31), the observations and the analysis bias can be defined as well:

$$\mathbf{b}^o = E(\boldsymbol{\varepsilon}^o); \tag{1.32}$$

$$\mathbf{b}^a = E(\boldsymbol{\varepsilon}^a), \tag{1.33}$$

and we assume $\mathbf{b}^o = \mathbf{0}$ for what it has been said before.

The presence of the bias forces us to modify the *bias-blind* interpolation scheme. As a matter of fact, as we have previously seen, if both the observations and the background fields in (1.7) are unbiased, then the analysis $\mathbf{x}^a$ is an unbiased estimate of the true atmospheric state $\mathbf{x}^t$ without regarding the specification of $\mathbf{K}$. Conversely, if the observations or (as in our case) the forecast is biased, then the analysis field as computed in (1.7) will be biased no matter which value of the gain matrix $\mathbf{K}$ we choose. We can prove this fact as follows, starting from (1.7) and neglecting the time-dependences for simplicity:

$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f)$;

now applying the expectation value operator $E(\cdot)$ to both the members of the equation we find:

$E(\mathbf{x}^a) = E(\mathbf{x}^f) + \mathbf{K}(E(\mathbf{x}^o) - \mathbf{H}E(\mathbf{x}^f))$,

$E(\mathbf{x}^t + \boldsymbol{\varepsilon}^a) = E(\mathbf{x}^t + \boldsymbol{\varepsilon}^f) + \mathbf{K}(E(\mathbf{H}\mathbf{x}^t + \boldsymbol{\varepsilon}^o) - \mathbf{H}E(\mathbf{x}^t + \boldsymbol{\varepsilon}^f))$.

Then using the definitions (1.31), (1.32) and (1.33) we obtain:

$\mathbf{b}^a = \mathbf{b}^f + \mathbf{K}(\mathbf{b}^o - \mathbf{H}\mathbf{b}^f)$.

As we can see in the equation above, the only way to have an unbiased analysis (i.e. $\mathbf{b}^a = 0$) is assuming $\mathbf{b}^o = \mathbf{b}^f = 0$.

In order to modify the scheme in a way so that it is able to compute and correct the bias resulting in an unbiased analysis, we follow what [Dee and Da Silva, 1998, Dee and Todling, 2000, Dee, 2003] proposed. The fundamental idea behind their works is that we can infer information about the systematic error introduced by the model comparing directly the forecast with the observations at each analysis time, that is the innovation (or the *observed-minus-forecast residuals*) vector $\mathbf{v}(t_i)$:

$\mathbf{v}(t_i) = \mathbf{y}^o(t_i) - \boldsymbol{H}_i(\mathbf{x}^f(t_i))$.

Linearizing $\boldsymbol{H}(\cdot)$ and expliciting both $\mathbf{x}^o$ and $\mathbf{x}^f$ we can rewrite $\mathbf{v}$ as follows:

$\mathbf{v}(t_i) = \boldsymbol{H}_i(\mathbf{x}^t(t_i)) + \boldsymbol{\varepsilon}^o(t_i) - \boldsymbol{H}(\mathbf{x}^f(t_i)) = \mathbf{H}_i\mathbf{x}^t(t_i) + \boldsymbol{\varepsilon}^o(t_i) - \mathbf{H}_i\mathbf{x}^f(t_i) = \mathbf{H}_i\mathbf{x}^t(t_i) + \boldsymbol{\varepsilon}^o(t_i) - \mathbf{H}_i\mathbf{x}^t(t_i) - \mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i) = \boldsymbol{\varepsilon}^o(t_i) - \mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i)$.

We can now compute the mean and the error covariance of $\mathbf{v}(t_i)$:

$E(\mathbf{v}(t_i)) = \mathbf{b}^o - \mathbf{H}\mathbf{b}^f$;

$E((\mathbf{v}(t_i) - E(\mathbf{v}(t_i)))(\mathbf{v}(t_i) - E(\mathbf{v}(t_i)))^T) = E((\boldsymbol{\varepsilon}^o(t_i) - \mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i) - \mathbf{b}^o + \mathbf{H}_i\mathbf{b}^f)(\boldsymbol{\varepsilon}^o(t_i) - \mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i) - \mathbf{b}^o + \mathbf{H}_i\mathbf{b}^f)^T) = \mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}(t_i)$.

The last expression is true if we assume to know the unbiased value of $\mathbf{b}^f$, i.e. $E(\mathbf{b}^f) = \mathbf{0}$, and if the hypothesis in (1.10) still holds.

As [Dee and Da Silva, 1998] argue, tha quantity $\mathbf{v}(t_i)$ can be seen as a *measurement model* for the background bias at time t_i. However, since the bias is by definition a slowly varying quantity, we expect it to be better identified over a time average. Therefore, we build a secondary Kalman Filter scheme which updates the bias $\mathbf{b}^f$ at every time step combining linearly an *a-priori* estimate (also called “predicted bias” or “bias prediction”), $\mathbf{b}^p$, with the observed quantity $\mathbf{v}$ corrected for the bias prediction $\mathbf{b}^p$:

$$\mathbf{b}^f(t_i) = \mathbf{b}^p(t_i) - \mathbf{L}_i[\mathbf{y}^o(t_i) - \mathbf{H}_i(\mathbf{x}^f(t_i) - \mathbf{b}^p(t_i))], \tag{1.34}$$

where $\mathbf{L}_i$ is the $(m \times p)$ “bias gain” matrix valid at time t_i:

$$\mathbf{L} = \mathbf{P}^b(t_i)\mathbf{H}_i^T[\mathbf{H}_i\mathbf{P}^b(t_i)\mathbf{H}_i^T + \mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}(t_i)]^{-1}, \tag{1.35}$$

and where $\mathbf{P}^b$ is the error covariance matrix for the bias prediction $\mathbf{b}^p(t_i)$ (and for the bias estimate $\mathbf{b}^f(t_{i-1})$ as well if the persistence model in (1.36) holds):

$$\mathbf{P}^b(t_i) = E(\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^b(t_i))^T).$$

The bias prediction $\mathbf{b}^p$ at the time t_i can be explicited as:

$$\mathbf{b}^p(t_i) = \mathbf{b}^t(t_i) + \boldsymbol{\varepsilon}^b,$$

where $\boldsymbol{\varepsilon}^b$ is the zero-mean error component of the bias prediction: $E(\boldsymbol{\varepsilon}^b) = \mathbf{0}$. In other words, we are assuming the bias prediction to be unbiased too.

Furthermore, in this study we adopt the persistence hypothesis: the bias prediction $\mathbf{b}^p$ to be employed at analysis time t_i is that resulting by the previous update performed as in (1.34) at time t_{i-1}, reduced by a damp coefficient μ:

$$\mathbf{b}^p(t_i) = \mu \cdot \mathbf{b}^f(t_{i-1}). \tag{1.36}$$

In order to avoid confusion and not to complicate notation, we consider eq. (1.36) as if $\mu = 1$ in the remainder of this chapter.

The condition (1.36) seems reasonable since the bias is a slowly varying quantity. Then we can write:

$$\mathbf{b}^f(t_i) = \mathbf{b}^f(t_{i-1}) - \mathbf{L}_i \left[\mathbf{y}^o(t_i) - \mathbf{H}_i(\mathbf{x}^f(t_i) - \mathbf{b}^f(t_{i-1}))\right]. \tag{1.37}$$

Now we are interested in computing the optimal value for the matrix $\mathbf{L}_i$, as we previously did for $\mathbf{K}$ in the section Â§1.2.1.

We start expliciting $\mathbf{v}(t_i)$ and $\mathbf{b}^p(t_i)$ in equation (1.34):

$$\mathbf{b}^f(t_i) = \mathbf{b}^t(t_i) + \boldsymbol{\varepsilon}^b(t_i) - \mathbf{L}_i \left[\boldsymbol{\varepsilon}^o(t_i) - \mathbf{H}_i\mathbf{x}^t(t_i) - \mathbf{H}_i\mathbf{x}^t(t_i) - \mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i) - \mathbf{H}_i\mathbf{b}^t(t_i) + \mathbf{H}_i\mathbf{b}^t(t_i) + \mathbf{H}_i\boldsymbol{\varepsilon}^b(t_i)\right],$$

which can be rewritten as:

$$\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i) = [\mathbf{I} - \mathbf{L}_i\mathbf{H}_i]\boldsymbol{\varepsilon}^b(t_i) - \mathbf{L}_i \left[\boldsymbol{\varepsilon}^o(t_i) - \mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i)\right].$$

Now computing the mean squared error of the left member of the above equation, we have:

$$\begin{aligned}
E\left[(\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i))^T(\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i))\right] = {} & E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\boldsymbol{\varepsilon}^b(t_i)\right] - E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\boldsymbol{\varepsilon}^b(t_i)\right] \\
& - E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^b(t_i)\right] + E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^b(t_i)\right] \\
& - E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\mathbf{L}_i\boldsymbol{\varepsilon}^o(t_i)\right] + E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\mathbf{L}_i\boldsymbol{\varepsilon}^o(t_i)\right] \\
& + E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i)\right] - E\left[(\boldsymbol{\varepsilon}^b(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i)\right] \\
& - E\left[(\boldsymbol{\varepsilon}^o(t_i))^T\mathbf{L}_i^T\boldsymbol{\varepsilon}^b(t_i)\right] + E\left[(\boldsymbol{\varepsilon}^f(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\boldsymbol{\varepsilon}^b(t_i)\right] \\
& + E\left[(\boldsymbol{\varepsilon}^o(t_i))^T\mathbf{L}_i^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^b(t_i)\right] - E\left[(\boldsymbol{\varepsilon}^f(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^b(t_i)\right] \\
& + E\left[(\boldsymbol{\varepsilon}^o(t_i))^T\mathbf{L}_i^T\mathbf{L}_i\boldsymbol{\varepsilon}^o(t_i)\right] - E\left[(\boldsymbol{\varepsilon}^f(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\mathbf{L}_i\boldsymbol{\varepsilon}^o(t_i)\right] \\
& - E\left[(\boldsymbol{\varepsilon}^o(t_i))^T\mathbf{L}_i^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i)\right] + E\left[(\boldsymbol{\varepsilon}^f(t_i))^T\mathbf{H}_i^T\mathbf{L}_i^T\mathbf{L}_i\mathbf{H}_i\boldsymbol{\varepsilon}^f(t_i)\right].
\end{aligned}$$

We want to minimize this quantity, so we perform:

$$\frac{d}{d\mathbf{L}}E\left[(\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i))^T(\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i))\right] = 0,$$

which is (according to the rules reported in the appendix):

$$\begin{gathered}
- E\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^b(t_i))^T\right]\mathbf{H}_i^T - E\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^b(t_i))^T\right]\mathbf{H}_i^T + 2\mathbf{L}_i\mathbf{H}_iE\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^b(t_i))^T\right]\mathbf{H}_i^T + \\
- E\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^o(t_i))^T\right] + 2\mathbf{L}_i\mathbf{H}_iE\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^o(t_i))^T\right] + E\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^f(t_i))^T\right]\mathbf{H}_i^T + \\
- 2\mathbf{L}_i\mathbf{H}_iE\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^f(t_i))^T\right]\mathbf{H}_i^T - E\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^o(t_i))^T\right] + E\left[\boldsymbol{\varepsilon}^b(t_i)(\boldsymbol{\varepsilon}^f(t_i))^T\right]\mathbf{H}_i^T + \\
+ 2\mathbf{L}_iE\left[\boldsymbol{\varepsilon}^o(t_i)(\boldsymbol{\varepsilon}^b(t_i))^T\right]\mathbf{H}_i^T - 2\mathbf{L}_i\mathbf{H}_iE\left[\boldsymbol{\varepsilon}^f(t_i)(\boldsymbol{\varepsilon}^b(t_i))^T\right]\mathbf{H}_i^T + 2\mathbf{L}_iE\left[\boldsymbol{\varepsilon}^o(t_i)(\boldsymbol{\varepsilon}^o(t_i))^T\right] + \\
- 2\mathbf{L}_i\mathbf{H}_iE\left[\boldsymbol{\varepsilon}^f(t_i)(\boldsymbol{\varepsilon}^o(t_i))^T\right] - 2\mathbf{L}_iE\left[\boldsymbol{\varepsilon}^o(t_i)(\boldsymbol{\varepsilon}^f(t_i))^T\right]\mathbf{H}_i^T + 2\mathbf{L}_i\mathbf{H}_iE\left[\boldsymbol{\varepsilon}^f(t_i)(\boldsymbol{\varepsilon}^f(t_i))^T\right]\mathbf{H}_i^T = 0
\end{gathered}$$

Employing the definitions of the matrices $\mathbf{P}^b(t_i)$ and $\mathbf{R}(t_i)$ and neglecting all the cross error covariances, we have:

$$-2\mathbf{P}^b(t_i)\mathbf{H}_i^T + 2\mathbf{L}_i\mathbf{H}_i\mathbf{P}^b(t_i)\mathbf{H}_i^T + 2\mathbf{L}_i\mathbf{R}(t_i) + 2\mathbf{L}_i\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T = 0.$$

We can finally explicit $\mathbf{L}_i$:

$$\mathbf{L}_i = \mathbf{P}^b(t_i)\mathbf{H}_i^T\left[\mathbf{H}_i\mathbf{P}^b(t_i)\mathbf{H}_i^T + \mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}(t_i)\right]^{-1}.$$

We can now show that the update bias estimate $\mathbf{b}^f(t_i)$ computed as in equation (1.37) is unbiased if both $\mathbf{b}^p(t_i)$ and $\mathbf{y}^0(t_i)$ are so. Subtracting $\mathbf{b}^t(t_i)$ to both members of (1.37) we have:

$$\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i) = (\mathbf{b}^p(t_i) - \mathbf{b}^t(t_i)) - \mathbf{L}_i\left[\mathbf{y}^o(t_i) - \mathbf{H}_i(\mathbf{x}^f(t_i) - \mathbf{b}^p(t_i))\right],$$

then averaging and expliciting $\mathbf{y}^o(t_i)$ and $\mathbf{x}^f(t_i)$:

$$E(\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i)) = E(\mathbf{b}^p(t_i) - \mathbf{b}^t(t_i)) - \mathbf{L}_i\left[E(\mathbf{H}_i\mathbf{x}^t(t_i) + \boldsymbol{\varepsilon}^o(t_i)) - \mathbf{H}_i(E(\mathbf{x}^t(t_i) + \boldsymbol{\varepsilon}^f(t_i)) - E(\mathbf{b}^p(t_i)))\right],$$

$$E(\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i)) = E(\mathbf{b}^p(t_i) - \mathbf{b}^t(t_i)) - \mathbf{L}_i\left[E(\boldsymbol{\varepsilon}^o(t_i)) - \mathbf{H}_i(\mathbf{b}^f(t_i) - E(\mathbf{b}^p(t_i))))\right]$$

If the bias prediction $\mathbf{b}^p(t_i)$ is unbiased and the observations are so, then $E(\mathbf{b}^p(t_i) - \mathbf{b}^t(t_i)) = 0$ and $E(\boldsymbol{\varepsilon}^o(t_i)) = 0$. So we obtain:

$$E(\mathbf{b}^f(t_i) - \mathbf{b}^t(t_i)) = \mathbf{0},$$

and so the bias estimate given by the equation (1.34) is unbiased if the above mentioned hypotheses hold.

The bias algorithm we have discussed so far provides a running estimate of the forecast bias which is employed to adjust the background field in the algorithm for the analysis field at time t_i, that is:

$$\mathbf{x}^a(t_i) = (\mathbf{x}^f(t_i) - \mathbf{b}^f(t_i)) + \mathbf{K}_i \left[\mathbf{y}^o(t_i) - \mathbf{H}_i(\mathbf{x}^f(t_i) - \mathbf{b}^f(t_i))\right]. \tag{1.38}$$

The equations (1.34) and (1.38) together constitute the so-called *bias-aware* scheme.

An interesting point in (1.38) is whether the gain matrix $\mathbf{K}_i$ is still the same as the equation (1.11), i.e. the matrix valid in a *bias-blind* scheme. We can show that this is true with a little bit of algebra, following what is showed in the Appendix of [Dee and Todling, 2000].

First of all, we rewrite our algorithm with a more convenient notation, defining the unbiased forecast $\tilde{\mathbf{x}}^f(t_i)$ as:

$$\tilde{\mathbf{x}}^f(t_i) = \mathbf{x}^f(t_i) - \mathbf{b}^f(t_i).$$

Then the equation (1.38) can be rewritten as:

$\mathbf{x}^a(t_i) = \tilde{\mathbf{x}}^f(t_i) + \mathbf{K}_i\left[\mathbf{y}^o(t_i) - \mathbf{H}_i(\tilde{\mathbf{x}}^f(t_i))\right].$

If we repeat the procedure of the section §1.2.1, we find:

$$\mathbf{K}_i = \left[\tilde{\mathbf{P}}^f(t_i)\mathbf{H}_i^T - \mathbf{X}^T(t_i)\right]\left[\left(\mathbf{H}_i\tilde{\mathbf{P}}^f(t_i)\mathbf{H}_i^T + \mathbf{R}(t_i) - \mathbf{H}_i\mathbf{X}^T(t_i) - \mathbf{X}(t_i)\mathbf{H}_i^T\right)\right]^{-1} \tag{1.39}$$

where $\tilde{\mathbf{P}}^f(t_i)$ and $\mathbf{X}(t_i)$ are the unbiased background error covariance matrix and the cross covariance matrix, respectively:

$$\tilde{\mathbf{P}}^f(t_i) = E(\tilde{\boldsymbol{\varepsilon}}^f(t_i)(\tilde{\boldsymbol{\varepsilon}}^f(t_i))^T), \tag{1.40}$$

$$\mathbf{X}(t_i) = E(\boldsymbol{\varepsilon}^o(t_i)(\tilde{\boldsymbol{\varepsilon}}^f(t_i))^T). \tag{1.41}$$

In this case $\mathbf{X}(t_i)$ cannot be neglected (as in the case of the hypothesis (1.10)) because $\boldsymbol{\varepsilon}^o(t_i)$ and $\tilde{\boldsymbol{\varepsilon}}^f(t_i)$ are not independent:

$\tilde{\boldsymbol{\varepsilon}}^f(t_i) = \tilde{\mathbf{x}}^f(t_i) - \mathbf{x}^t(t_i) = \mathbf{x}^f(t_i) - \mathbf{b}^f(t_i) - \mathbf{x}^t(t_i) = \boldsymbol{\varepsilon}^f(t_i) + \mathbf{b}^t(t_i) - \mathbf{b}^f(t_i) = [\mathbf{I} - \mathbf{L}_i\mathbf{H}_i](\boldsymbol{\varepsilon}^f(t_i) - \boldsymbol{\varepsilon}^b(t_i)) + \mathbf{L}_i\boldsymbol{\varepsilon}^o(t_i).$

Then, if we define the matrices $\mathbf{P}$ and $\mathbf{S}$ as (for simplicity, from now on we drop the time-dependence):

$\mathbf{P} = \mathbf{P}^b + \mathbf{P}^f;$

$\mathbf{S} = \mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R};$

we can explicit $\tilde{\mathbf{P}}^f$in (1.40) substituting $\tilde{\boldsymbol{\varepsilon}}^f$ inside. Therefore, we have:

$$\begin{aligned}\tilde{\mathbf{P}}^f &= [\mathbf{I}-\mathbf{LH}]\,\mathbf{P}\,[\mathbf{I}-\mathbf{LH}]^T+\mathbf{LRL}^T\\ &= \mathbf{P}-\mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^f-\mathbf{PH}^T\mathbf{S}^{-1}\mathbf{HP}^b.\end{aligned}$$

Moreover, $\mathbf{X}$ in eq. (1.41) can be explicited as:

$$\mathbf{X}=\mathbf{RL}^T=\mathbf{RS}^{-1}\mathbf{HP}^b.$$

At this point we have all we need to show that $\mathbf{K}$ in eq. (1.39) is the same as that in eq. (1.11). The first member of the right term of the equation (1.39) can be rewritten as:

$$\begin{aligned}\tilde{\mathbf{P}}^f\mathbf{H}^T-\mathbf{X}^T &= \mathbf{PH}^T-\mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^f\mathbf{H}^T-\mathbf{PH}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T-\mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{R}\\ &= \mathbf{PH}^T-\mathbf{P}^b\mathbf{H}^T\mathbf{S}^{-1}\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{HP}^b\mathbf{H}^T+\mathbf{R}\right]-\mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T\\ &= \mathbf{PH}^T-\mathbf{P}^b\mathbf{H}^T-\mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T\\ &= \mathbf{P}^f\mathbf{H}^T\left[\mathbf{I}-\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T\right]\\ &= \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right];\end{aligned}$$

instead, the second term reads as:

$$\begin{aligned}\mathbf{H}\tilde{\mathbf{P}}^f\mathbf{H}^T+\mathbf{R}-\mathbf{HX}^T-\mathbf{XH}^T &= \mathbf{HPH}^T-\mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^f\mathbf{H}^T-\mathbf{HPH}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T+\mathbf{R}-\mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\mathbf{R}-\mathbf{RS}^{-1}\mathbf{HP}^b\mathbf{H}^T\\ &= \mathbf{S}-\mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{HP}^b\mathbf{H}^T+\mathbf{R}\right]-\mathbf{HP}^f\mathbf{H}^T\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T-\mathbf{RS}^{-1}\mathbf{HP}^b\mathbf{H}^T\\ &= \mathbf{S}-\mathbf{HP}^b\mathbf{H}^T-\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right]\mathbf{S}^{-1}\mathbf{HP}^b\mathbf{H}^T\\ &= \left[\mathbf{I}-\mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\right]\left[\mathbf{S}-\mathbf{HP}^b\mathbf{H}^T\right]\\ &= \left[\mathbf{I}-\mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\right]\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right].\end{aligned}$$

Therefore, substituting the two previous results in (1.39), we finally obtain:

$$\begin{aligned}\mathbf{K} &= \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right]\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right]^{-1}\left[\mathbf{I}-\mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\right]^{-1}\\ &= \mathbf{P}^f\mathbf{H}^T\mathbf{S}^{-1}\left[\mathbf{I}-\mathbf{HP}^b\mathbf{H}^T\mathbf{S}^{-1}\right]^{-1}\\ &= \mathbf{P}^f\mathbf{H}^T\left[\mathbf{S}-\mathbf{HP}^b\mathbf{H}^T\right]^{-1}\\ &= \mathbf{P}^f\mathbf{H}^T\left[\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right]^{-1}.\end{aligned}$$

So we have proved the fact that the matrix $\mathbf{K}$ for the bias-aware scheme is the same as that valid in the bias-blind one (cfr. eq. (1.11)).

The bias covariance model

As we can see in (1.35), the bias prediction error covariance matrix $\mathbf{P}^b$ plays a particularly important role in shaping the matrix $\mathbf{L}$, and it has to be specified in order to apply the bias correction algorithm. To this aim, we adopt a model for $\mathbf{P}^b$ previously proposed by [Dee and Da Silva, 1998] and employed with good results in [Dee and Todling, 2000], which states:

$$\mathbf{P}^b = \gamma \mathbf{P}^f, \tag{1.42}$$

where γ is a parameter to be set.

The hypothesis in (1.42) is based on the assumption that the spatial correlation of the uncertainty $\boldsymbol{\varepsilon}^b$ associated to the bias estimation $\mathbf{b}^p$ has the same dynamics as that of the random error $\boldsymbol{\varepsilon}^f$ associated to the model forecast. That is, we don't formulate a separate physical model for the evolution of $\mathbf{b}^p$, but we rather make it dependent on the forecast. If we put (1.42) in (1.35) we obtain:

$$\mathbf{L} = \gamma \mathbf{P}^f \mathbf{H}^T [(\gamma + 1)\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}]^{-1},$$

which is equivalent to:

$$\mathbf{L} = \frac{\gamma}{\gamma+1} \mathbf{P}^f \mathbf{H}^T \left[\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \frac{1}{\gamma+1}\mathbf{R} \right]^{-1}. \tag{1.43}$$

It is clear from (1.43) that the scalar γ in (1.42) shapes the matrix $\mathbf{L}$ and so in turn the rapidity with which the predicted bias is adjusted toward the bias corrected residuals. In particular, smaller values of γ tend to preserve the bias prediction, giving little weight to the innovation. Bigger values of γ enhance the importance of the bias corrected residuals resulting in a very noisy rapid response for the bias correction.

The consequence of the hypothesis given by the eq. (1.42) is not limited to the bias update algorithm, since the latter is used to correct the background in the bias-aware analysis $\overline{\mathbf{x}}^a$. We can then rewrite the eq. (1.38) as:

$$\begin{aligned}
\overline{\mathbf{x}}^a &= \left(\overline{\mathbf{x}}^f - \mathbf{b}^f\right) + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\left(\overline{\mathbf{x}}^f - \mathbf{b}^f\right)\right) \\
&= \overline{\mathbf{x}}^f - \mathbf{b}^p + \mathbf{L}\left(\mathbf{y}^o - \mathbf{H}\left(\overline{\mathbf{x}}^f - \mathbf{b}^p\right)\right) + \mathbf{K}\left(\mathbf{y}^o - \mathbf{H}\left(\overline{\mathbf{x}}^f - \mathbf{b}^p\right)\right) - \mathbf{KHL}\left(\mathbf{y}^o - \mathbf{H}\left(\overline{\mathbf{x}}^f - \mathbf{b}^p\right)\right) \\
&= \overline{\mathbf{x}}^f - \mathbf{b}^p + (\mathbf{L} + \mathbf{K} - \mathbf{KHL})\left(\mathbf{y}^o - \mathbf{H}\left(\overline{\mathbf{x}}^f - \mathbf{b}^p\right)\right).
\end{aligned}$$

The role played by $\mathbf{K}$ in (1.38) is now played by the $\mathbf{L} + \mathbf{K} - \mathbf{KHL}$ matrix, which can be rewritten as:

$$\begin{aligned}
\mathbf{L}+\mathbf{K}-\mathbf{KHL} &= \mathbf{L}+\mathbf{K}\left(\mathbf{I}-\mathbf{HL}\right)\\
&= \mathbf{L}+\mathbf{K}\left[\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}-\frac{\gamma}{1+\gamma}\mathbf{HP}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}\right]\\
&= \mathbf{L}+\mathbf{K}\left[\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}-\frac{\gamma}{1+\gamma}\mathbf{HP}^f\mathbf{H}^T\right)\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}\right]\\
&= \mathbf{L}+\mathbf{K}\left[(1+\gamma)^{-1}\left(\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right)\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}\right]\\
&= \mathbf{L}+(1+\gamma)^{-1}\mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right)^{-1}\left(\mathbf{HP}^f\mathbf{H}^T+\mathbf{R}\right)\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}\\
&= \frac{\gamma}{1+\gamma}\mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}+(1+\gamma)^{-1}\mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}\\
&= \mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}.
\end{aligned}$$

Therefore, the bias-aware scheme for $\overline{\mathbf{x}}^a$ which includes the hypothesis in (1.42) reads as:

$$\overline{\mathbf{x}}^a=\overline{\mathbf{x}}^f-\mathbf{b}^p+\mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}\left(\mathbf{y}^o-\mathbf{H}\left(\overline{\mathbf{x}}^f-\mathbf{b}^p\right)\right);$$

with $\mathbf{K}$ redefined as:

$$\mathbf{K}=\mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}.$$

In applying the bias-aware scheme in a ensemble context, it is essential to redefine also the $\tilde{\mathbf{P}}^a$ matrix, i.e. the anaylsis error covariance matrix in the dimensionally-reduced ensemble space. In order to do that, we have to recompute the matrix $\mathbf{P}^a$, starting from $\boldsymbol{\varepsilon}^a$:

$$\begin{aligned}
\boldsymbol{\varepsilon}^a &= \overline{\mathbf{x}}^a-\mathbf{x}^t\\
&= \boldsymbol{\varepsilon}^f-\boldsymbol{\varepsilon}^b+\mathbf{K}\left(\boldsymbol{\varepsilon}^o-\mathbf{H}\left(\boldsymbol{\varepsilon}^f-\boldsymbol{\varepsilon}^b\right)\right)\\
&= \boldsymbol{\varepsilon}^f-\boldsymbol{\varepsilon}^b+\mathbf{K}\boldsymbol{\varepsilon}^o-\mathbf{KH}\left(\boldsymbol{\varepsilon}^f-\boldsymbol{\varepsilon}^b\right)\\
&= \mathbf{K}\boldsymbol{\varepsilon}^o-\left(\mathbf{I}-\mathbf{KH}\right)\left(\boldsymbol{\varepsilon}^f-\boldsymbol{\varepsilon}^b\right).
\end{aligned}$$

Now, applying the usual definition of $\mathbf{P}^a$ (cfr. eq. (1.4)) we obtain:

$$\begin{aligned}
\mathbf{P}^a &= E\left[\boldsymbol{\varepsilon}^a\left(\boldsymbol{\varepsilon}^a\right)^T\right]\\
&= \mathbf{KRK}^T-\left(\mathbf{I}-\mathbf{KH}\right)\left(\mathbf{P}^f+\mathbf{P}^b\right)\left(\mathbf{I}-\mathbf{KH}\right)^T\\
&= \mathbf{KRK}^T+(1+\gamma)\mathbf{P}^f-(1+\gamma)\mathbf{P}^f\mathbf{H}^T\mathbf{K}^T-(1+\gamma)\mathbf{KHP}^f+(1+\gamma)\mathbf{KHP}^f\mathbf{H}^T\mathbf{K}^T\\
&= (1+\gamma)\mathbf{P}^f-(1+\gamma)\mathbf{P}^f\mathbf{H}^T\mathbf{K}^T-(1+\gamma)\mathbf{KHP}^f+(1+\gamma)\mathbf{K}\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)\mathbf{K}^T\\
&= (1+\gamma)\mathbf{P}^f-(1+\gamma)\mathbf{P}^f\mathbf{H}^T\mathbf{K}^T-(1+\gamma)\mathbf{KHP}^f+(1+\gamma)\mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)^{-1}\left(\mathbf{HP}^f\mathbf{H}^T+(1+\gamma)^{-1}\mathbf{R}\right)\mathbf{K}^T\\
&= (1+\gamma)\mathbf{P}^f-(1+\gamma)\mathbf{P}^f\mathbf{H}^T\mathbf{K}^T-(1+\gamma)\mathbf{KHP}^f+(1+\gamma)\mathbf{P}^f\mathbf{H}^T\mathbf{K}^T=(1+\gamma)\mathbf{P}^f-(1+\gamma)\mathbf{KHP}^f\\
&= (1+\gamma)\left(\mathbf{I}-\mathbf{KH}\right)\mathbf{P}^f.
\end{aligned}$$

Therefore, we can replicate what we did in the paragraph § in order to obtain a value of $\mathbf{K}$ valid in an EnKF, taking into account the covariance inflation (cfr. §) and applying the before-mentioned identity to $\mathbf{A} = \left(\triangle \cdot (1+\gamma) \cdot \frac{1}{k-1}\right)^{\frac{1}{2}} \mathbf{Y}^f$:

$$\begin{aligned}
\mathbf{K} &= \triangle \cdot \frac{1}{k-1}\mathbf{X}^f(\mathbf{X}^f)^T\mathbf{H}^T \left[\triangle \cdot \frac{1}{k-1}\mathbf{H}\mathbf{X}^f(\mathbf{X}^f)^T\mathbf{H}^T + \frac{1}{1+\gamma}\mathbf{R}\right]^{-1} \\
&= \triangle \cdot (1+\gamma) \cdot \frac{1}{k-1}\mathbf{X}^f(\mathbf{Y}^f)^T \left[(1+\gamma) \cdot \frac{1}{k-1}\mathbf{Y}^f(\mathbf{Y}^f)^T + \mathbf{R}\right]^{-1} \\
&= \mathbf{X}^f \left[\frac{1}{\triangle} \cdot \frac{1}{1+\gamma} \cdot (k-1)\mathbf{I} + (\mathbf{Y}^f)^T\mathbf{R}^{-1}\mathbf{Y}^f\right] (\mathbf{Y}^f)^T\mathbf{R}^{-1}.
\end{aligned}$$

Then, putting the above definition in $\mathbf{P}^a$:

$$\begin{aligned}
\mathbf{P}^a &= (1+\gamma)\left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{P}^f \\
&= \Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{X}^f\left(\mathbf{X}^f\right)^T \\
&= \Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\left[\mathbf{X}^f\left(\mathbf{X}^f\right)^T - \mathbf{K}\mathbf{H}\mathbf{X}^f\right] \\
&= \Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\left[\mathbf{X}^f\left(\mathbf{X}^f\right)^T - \mathbf{X}^f\left[\Delta^{-1}\left(1+\gamma\right)^{-1}\left(k-1\right)\mathbf{I} + \left(\mathbf{Y}^f\right)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}\left(\mathbf{Y}^f\right)^T\mathbf{R}^{-1}\mathbf{H}\mathbf{X}^f\left(\mathbf{X}^f\right)^T\right] \\
&= \Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\mathbf{X}^f\left[\mathbf{I} - \left[\mathbf{I} + \Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\left(\mathbf{Y}^f\right)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}\Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\left(\mathbf{Y}^f\right)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]\left(\mathbf{X}^f\right)^T \\
&= \Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\mathbf{X}^f\left[\mathbf{I} + \Delta\left(1+\gamma\right)\left(k-1\right)^{-1}\left(\mathbf{Y}^f\right)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}\left(\mathbf{X}^f\right)^T \\
&= \mathbf{X}^f\left[\Delta^{-1}\left(1+\gamma\right)^{-1}\left(k-1\right)\mathbf{I} + \left(\mathbf{Y}^f\right)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}\left(\mathbf{X}^f\right)^T \\
&= \mathbf{X}^f\tilde{\mathbf{P}}^a\left(\mathbf{X}^f\right)^T.
\end{aligned}$$

Then, provided the hypothesis in (1.42), the anaylsis error covariance matrix in the ensemble space $\tilde{\mathbf{P}}^a$ reads as:

$$\tilde{\mathbf{P}}^a = \left[\Delta^{-1}\left(1+\gamma\right)^{-1}\left(k-1\right)\mathbf{I} + \left(\mathbf{Y}^f\right)^T\mathbf{R}^{-1}\mathbf{Y}^f\right]^{-1}.$$

1.2.7 The Spatial Consistency Test

Our scheme is based on the idea that the observations coming from the MET Norway network of stations could correct and adjust the model forecast, providing information. An important requirement is that the observations used in the scheme are as reliable as possible.

In section §1.2.3 we discussed the predominance of the representativity error in $\boldsymbol{\varepsilon}^o$, reducing the instrument error to a minor role. Even though, several kinds of malfunctioning in the observation process can result in a large instrument error and a totally unrealistic observed value. In this case, a simple comparison with the climatological mean can easily lead to exclude that measurement. However, there are more subtle cases in which an observation is still affected by a large error, but not enough to be totally unrealistic from a climatological point of view. We refer to them as "gross errors".

To this aim, a Spatial Consistency Test to check for the the observations quality is performed in the scheme. For every observation entering the scheme the gross probability error (GEs) is computed and discarted if the value is greater than a certain thresold.

In order to detect the presence of possible gross errors in the observations employed in our scheme, we adopt the Spatial Consistency Test already proposed in [Lussana et al., 2010]. In particular, for each observation, the Cross Validated analysis residuals for each i-th observation $y_i^o - \check{y}_i^a$ is evaluated and the measurement discarded if:

$$(y_i^o - \check{y}_i^a)^2 > T^2(\sigma_{oi}^2 + \check{\sigma}_{ai}^2). \tag{1.44}$$

In equation (1.44), the Cross Validated analysis $\check{y}_i^a$ is the value of the analysis computed at the i-th station location without taking into account the observed value in that point. The CV residual $y_i^o - \check{y}_i^a$ is one of the possible way to evaluate the quality of the analysis itself, because it provides an estimate of how much the scheme is able to reproduce the observed value even in the absence of the local constraint given by that observation. σ_{oi}^2 and $\check{\sigma}_{ai}^2$ are the i-th observation and CV analysis variances, whereas T^2 is a constant parameter (here set to $T^2 = 40$).

In the appendix A of [Lussana et al. 2010] it is shown that:

$$y_i^o - \check{y}_i^a = \frac{1}{z_i}(\mathbf{S} + \mathbf{R})^{-1}(\mathbf{y}^o - \mathbf{y}^f), \tag{1.45}$$

where $\mathbf{S}$ is the background error covariance matrix computed at station locations:

$$\mathbf{S} = E\left[(\mathbf{y}^f - \mathbf{y}^t)(\mathbf{y}^f - \mathbf{y}^t)^T\right],$$

with $\mathbf{y}^f = \boldsymbol{H}(x^f)$ and $\mathbf{y}^t = \boldsymbol{H}(x^t)$. In the EnKF context and taking into account the inflation parameter $\triangle$, we can define $\mathbf{S}$ as:

$$\mathbf{S} = \mathbf{H}\mathbf{P}^f\mathbf{H}^T = \triangle \cdot \mathbf{H}\mathbf{P}_e^f\mathbf{H}^T = \triangle \cdot \frac{1}{k-1}\mathbf{H}\mathbf{X}(\mathbf{X}^f)^T\mathbf{H}^T = \triangle \cdot \frac{1}{k-1}\mathbf{Y}^f(\mathbf{Y}^f)^T.$$

In the equation (1.45), z_i is defined as the i-th diagonal element of the matrix $(\mathbf{S} + \mathbf{R})^{-1}$, that is:

$$z_i = \left[(\mathbf{S} + \mathbf{R})^{-1}\right]_{i,i}.$$

In our scheme we have decided to implement a localization procedure for the matrix $\mathbf{S}$, multiplying it at each station location i for the scalar function:

$$S_{loc} = \exp\left[-\frac{1}{2}\left(\frac{d_H(i,j)^2}{(D_H)^2} + \frac{d_V(i,j)^2}{(D_v)^2}\right)\right],$$

where $d_H(i,j)$ and $d_V(i,j)$ indicate the horizontal and vertical distances between each pair of station (i,j) and D_H and D_v are two constant de-correlation paramters. Thus we perform:

$$\mathbf{S}_{loc} = S_{loc} \cdot \mathbf{S},$$

and we can substitute the new covariance in the equation (1.45), that is:

$$y_i^o - \check{y}_i^a = \frac{1}{z_i}(\mathbf{S}_{loc} + \mathbf{R})^{-1}(\mathbf{y}^o - \mathbf{y}^b).$$

In the appendix B of [Lussana et al. 2010] it is also shown that the sum of the variances in (1.44) can also be computed as:

$$\sigma_{oi}^2 + \check{\sigma}_{ai}^2 = \frac{1}{z_s}.$$

The SCT is performed for each value of the parameter ϵ defined in the section §1.2.4 and the respective observations excluded.

1.2.8 Summary of the complete scheme

We can finally summarize the set of essential equations which constitutes the complete scheme adopted in our work, including the localization, the bias correction and the NMC method.

The analysis mean at each grid point is computed as:

$$\overline{\mathbf{x}}^a = \hat{\mathbf{x}}^f + \mathbf{K}(\mathbf{y}^o - \hat{\mathbf{y}}^f),$$

where:

$$\begin{cases} \hat{\mathbf{y}}^f = \boldsymbol{H}(\hat{\mathbf{x}}^f) = \mathbf{H}\hat{\mathbf{x}}^f \\ \hat{\mathbf{x}}^f = \mathbf{x}_{BE}^f - \mathbf{b}^f \\ \mathbf{b}^f = \mathbf{b}^p - \mathbf{L}(\mathbf{y}^o - \hat{\mathbf{y}}^f) \\ \mathbf{L} = \gamma \cdot (\gamma + 1)^{-1} \cdot \mathbf{K} \\ \mathbf{K} = \mathbf{X}^f \tilde{\mathbf{P}}^a (\mathbf{Y}^f)^T \mathbf{R}_{loc}^{-1} \\ \tilde{\mathbf{P}}^a = \left[\Delta^{-1} \cdot (1+\gamma)^{-1} \cdot (k-1) \cdot \mathbf{I} + \left(\mathbf{Y}^f\right)^T \mathbf{R}_{loc}^{-1} \mathbf{Y}^f\right]^{-1} \\ \mathbf{Y}^f = \mathbf{H}\mathbf{X}^f \end{cases} .$$

$\mathbf{x}_{BE}^f$ has been previously defined in eq. (1.24), $\mathbf{R}_{loc}$ in section §1.2.3, Δ in §1.2.4 and γ in §1.2.6. The analysis ensemble members $\{\mathbf{x}^{a\,(i)}\ i = 1, ..., k\}$ are defined as:

$$\mathbf{x}^{a\,(i)} = \mathbf{X}^{a\,(i)} - \overline{\mathbf{x}}^a,$$

where:

$$\begin{cases} \mathbf{X}^a = \mathbf{X}^f \mathbf{W}^a \\ \mathbf{W}^a = ((k-1)\tilde{\mathbf{P}}^a \end{cases} \qquad .$$

Chapter 2

The implementation of the LETKF-based scheme at MET Norway

In this chapter, we illustrate how we implemented the LETKF-based scheme described in §1 at MET Norway from a practical point of view. We will present the main characteristics of the data sources, as well as the diagnostics of the scheme, including some considerations about the observation operator sensitivity and discussing the behaviour of some of the scheme implementations and the parameters setting.

2.1 The input data

2.1.1 The surface temperature observations

The Norwegian Meteorological Institute (MET Norway) together with other public institutions[1] manage a vast network of manual and automatic weather stations distributed over the Norwegian mainland and part of the neighboring areas of Sweden and Finland. Several different kind of observations are performed by these stations, and all of them are stored in the climate database of MET Norway (*KlimaDataVareHuset*, or KDVH). In our scheme we employ the hourly sampled air temperature observations performed by the network. These data will constitute the set of hourly 2-meter temperature observations we provide our scheme with, at each time step.

The spatial distribution of the stations in the network performing such observations during February 2016 is shown in fig. 2.1. The network is particularly dense in the south and around Oslo. The number of stations available is not constant in time and the spatial distribution might be affected by minor variations from one hour to the following one. In figure (2.2) the number of hourly available observations during a six months period (from February to July 2016) is shown:

---

[1] Among others: the Norwegian Water Resources and Energy Directorate (NVE), The Norwegian Public Roads Administration (Statens Vegvesen), the Norwegian Institute of Bioeconomy research (NIBIO), the Swedish Meteorological and Hydrological Institute (SMHI), the Finnish Meteorological Institute (FMI).

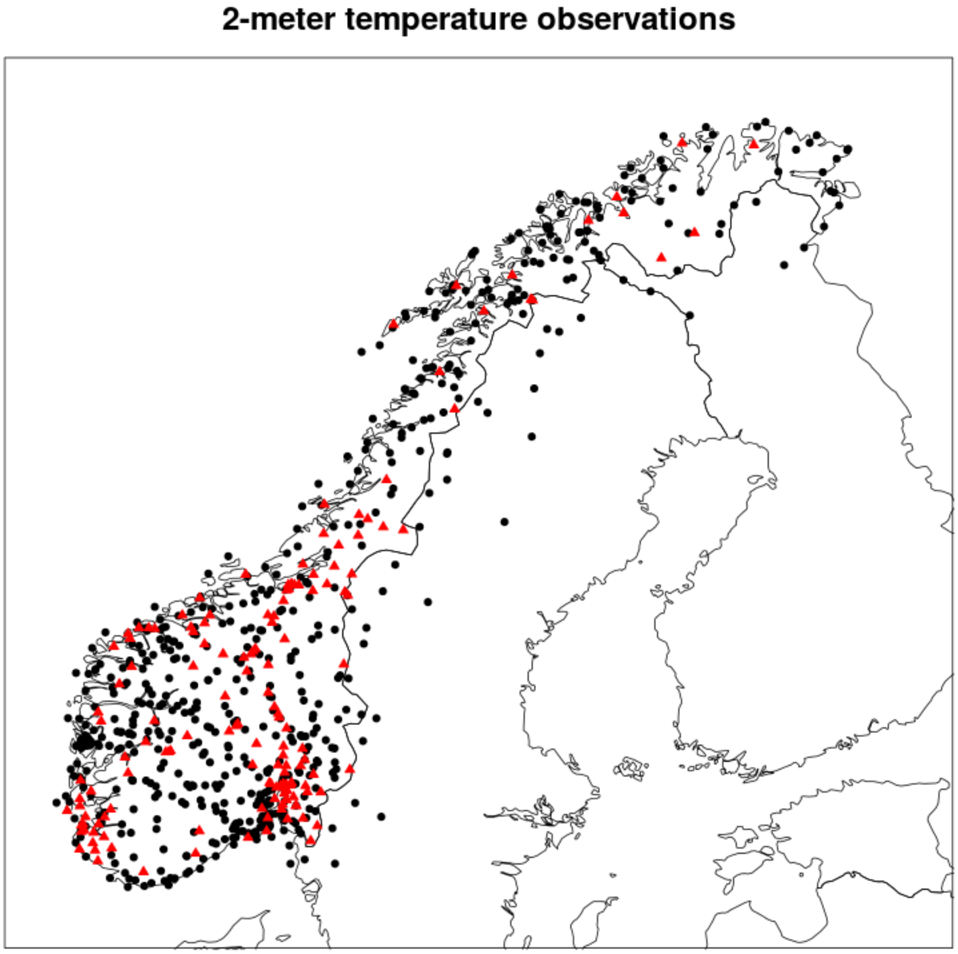


Figure 2.1: Spatial distribution of the stations performing measurements of the 2-meter temperature during February 2016. The red triangles indicate those stations which report less than 90% of the hourly measurements during the period.

the value fluctuates slighlty above 700, with some occasional dips to around 400 and rare complete lacks.

The horizontal (and vertical) density of the stations in the network will be a useful criterion to set some reasonable values for the de-correlation distance D_h (D_z) (cfr. §2.4). The histrograms in fig. (2.3) show the distribution of the mean horizontal and vertical displacement between each station in the network and its first three nearest neighbors[2].

---

[2] Here we compute the "first three nearest neighbors" by means of the function: $f = \exp\left[-\frac{1}{2}\left(\frac{\triangle h}{40}\right)^2 - \frac{1}{2}\left(\frac{\triangle z}{500}\right)^2\right]$, where $\triangle h$ and $\triangle z$ are the horizontal and the vertical distances between each station and all the other ones. We retain, for each station, the neighbor stations associated to the three biggest values of the function f.

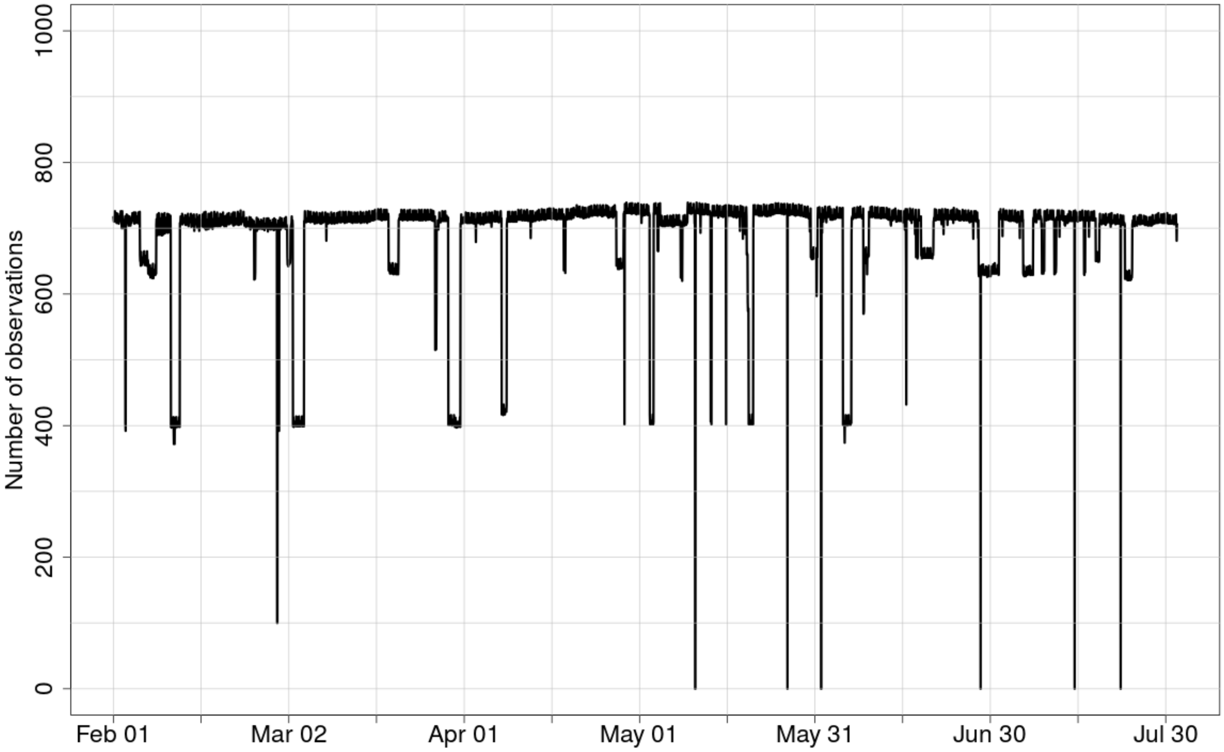


Figure 2.2: Number of available hourly observations from 1st February to 31st July 2016.

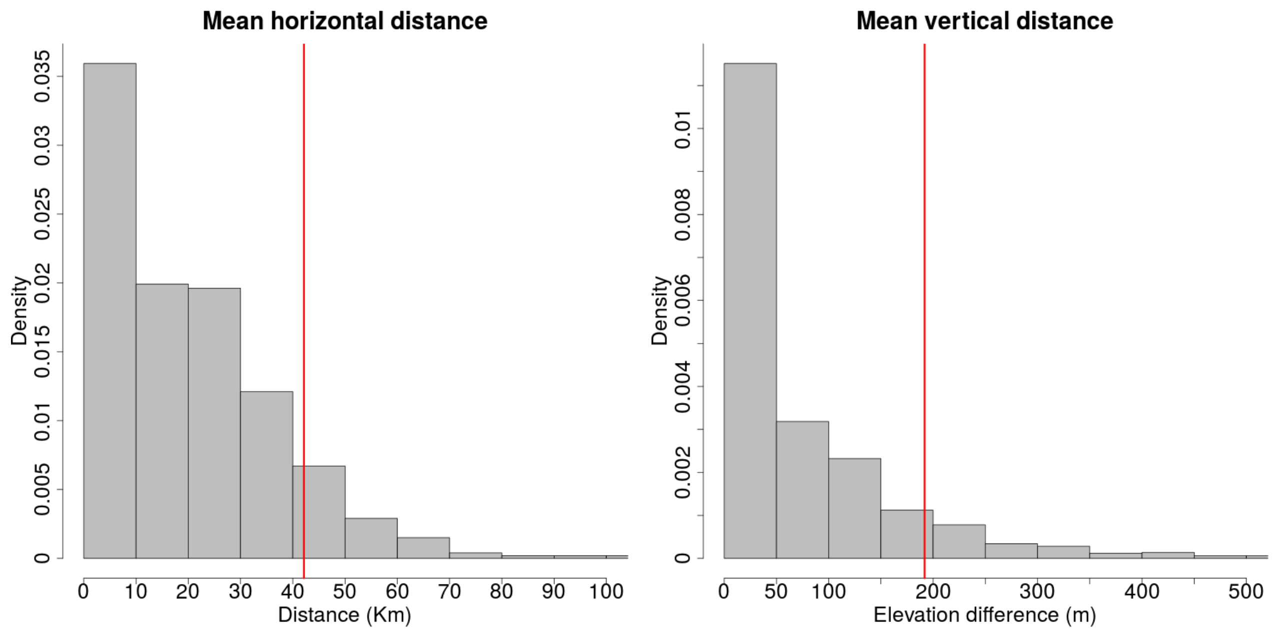


Figure 2.3: Distribution of the mean horizontal distance (left) and of the altitude difference (right) between each station of the network and the first three nearest neighbors (see the footnote 2). The red line indicates the position of the 90th percentile of the distribution.

2.1.2 The Arome MetCoOp 2.5 km model and the predicted 2-meter temperature fields

In our Kalman Filter scheme, the background (the first guess of the final analysis) is given by the predicted values of the hourly 2-meter temperature coming from the high-resolution operational numerical model run at MET Norway: Arome MetCoOp. Such a model has been developed by MET Norway in a joint collaboration with the Swedish Meteorological and Hydrological Institute (SMHI). It is based on the convection-permitting model AROME run by Météo France, with several modifications in the microphysics, the assimilation cycle and in the surface description, which better conform to the environmental properties and the atmospheric conditions of the Nordic region. The model is forced at boundaries (both upper and lateral) by the global model ECMWF Integrated Forecasting System. It is run four times per day (0000, 0600, 1200, 1800 UTC) over a grid of 2.5 km of horizontal resolution and on 65 vertical layers. Each run produces a 66 hours forecast. More details about the dynamical set up and the physical parametrizations of the model can be found in [Müller et al., 2017]. An example of the 2-meter temperature field produced by the model (and its domain) is shown in fig. (2.4).

The data employed in the scheme as background are not the raw output of the 2-meter temperature provided by Arome MetCoOp 2.5 km. They undergo some preliminary post-processing procedures at MET Norway before being employed in the scheme.

Firstly, a correction based on the difference between the predicted and the observed values

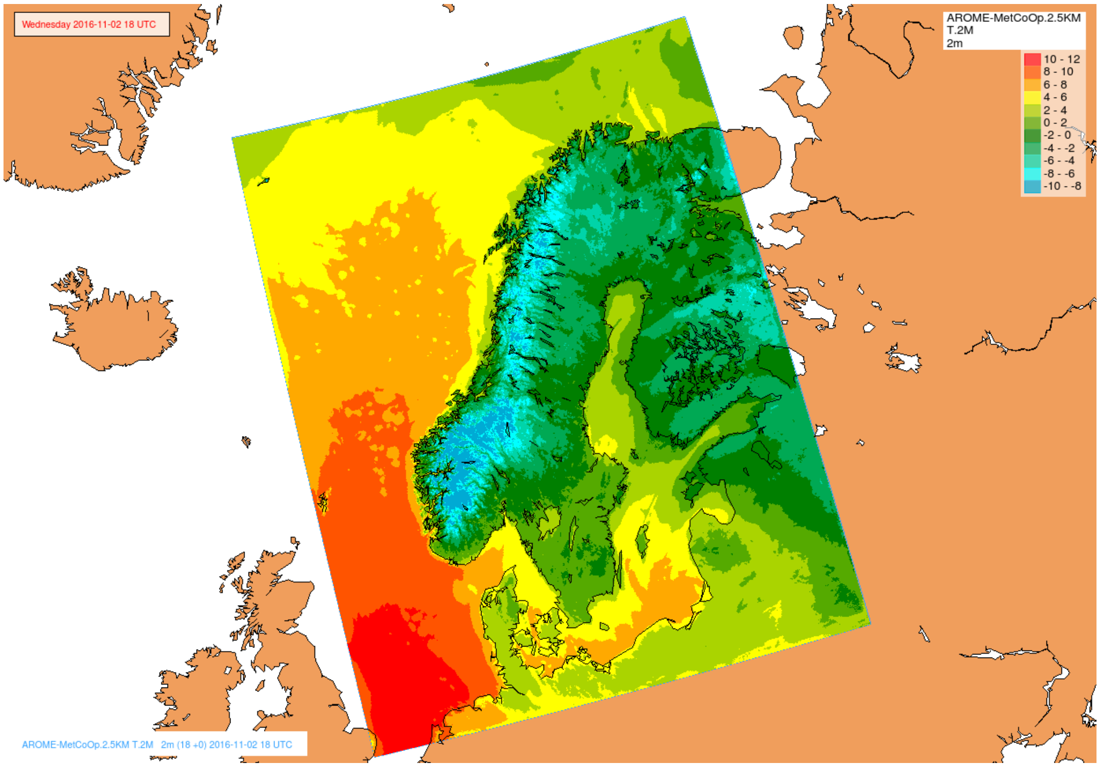


Figure 2.4: An exemple of a 2-meter temperature field produced by AROME-MetCoOp 2.5 km.

recorded during the previous days (with more importance given to the previous 24 hours) at the station location is performed by means of a kriging technique over the neighboring grid points within an area of 25 km and 200 meters of vertical elevation difference. Moreover, the temperature values are corrected according to the elevation difference between the model and the real altitude, employing the mean lapse-rate, i.e. $-6^{\circ C}/_{Km}$.

Lastly, the data are regridded onto a grid of 1 km of spacing (see fig. (2.5)), adopting a nearest neighbor method and correcting the values for the difference between the previous and the new grid point's elevation by using a near-surface vertical lapse rate inferred by the temperature values in a surrounding squared area with a 15 km side. A mean lapse-rate of $-6^{\circ C}/_{Km}$ is employed if the difference in elevation between the highest and the lowest grid point within the area is less than 30 meters.

This latter regridding procedure is performed in order to properly represent the Norwegian drainage network, since the final analysis could be used in the future for climatological and hydrological applications. This grid depicts the norwegian orography in details, with a range in the elevation between $z = -3\,m\ AMSL$ and $z = 2306\,m\ AMSL$.

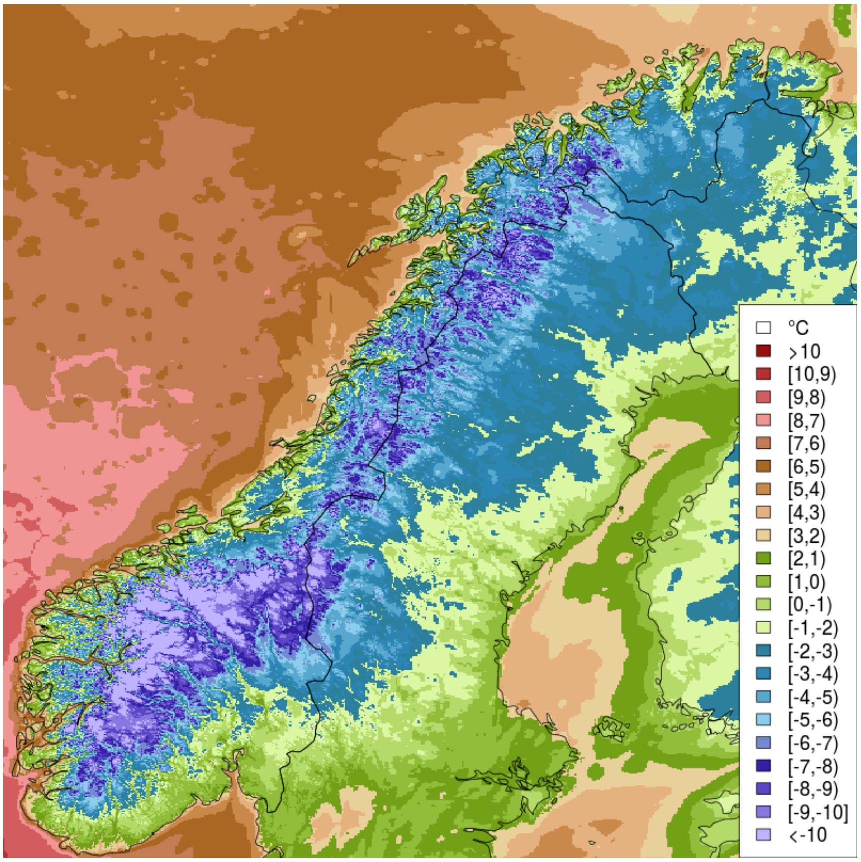


Figure 2.5: The same 2-meter temperature field as in fig. (2.4), after the regridding procedure onto a grid of 1 km of spatial resolution.

2.2 Observation operator sensitivity

Before applying the whole scheme, we evaluated the sensitivity of the observation operator.

The observation operator is meant to map a model state into the observation space. In our scheme we consider it as a linear operator (i.e. a matrix of dimension $(p \times n)$), which interpolates the model data defined over the grid onto the station locations.

In order to do that, for each i-th available observation, a set of grid points $j = 1, ..., N$ within an area of radius R is considered (see the example in figure (2.6)) and the near-surface temperature lapse rate Γ_i is deduced through a linear regression, fitting their temperature values $x^f_{j=1,...,N}$ against their altitude $z^f_{j=1,...,N}$. Then, the value of the background over the station location y^f_i is computed according to the station elevation z^o_i. So, for each station i:

$$y^f_i = \overline{x} + \Gamma_i \cdot (z^o_i - \overline{z}),$$

where $\overline{x}$ is the mean temperature value computed over the grid points $j = 1, ..., N$:

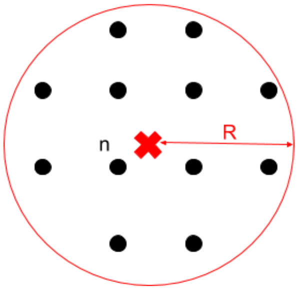


Figure 2.6: An example of how $\boldsymbol{H}(\cdot)$ works. The red cross is the station location at which the observation operator interpolates the background spreading over the surrounding grid points (black circles) within an area of radius R.

$$\overline{x} = \frac{1}{N}\sum_{j=1}^{N} x_j^f;$$

and $\overline{z}$ is the mean elevation value computed over the grid points $j = 1, ..., N$:

$$\overline{z} = \frac{1}{N}\sum_{j=1}^{N} z_j^f.$$

Hourly observations are representative of the meteorological variability belonging to the meso-gamma atmospheric scale [Orlanski, 1975]. In this regard, the observation operator should provide background values representative of a comparable atmospheric scale, using only the closest grid points to the i-th station to perform the interpolation. On the other hand, for the linear operator to be robust from a statistical point of view respect to the outliers in the background field (for example, because of differences in elevation between the digital elevation model and the station location) it may be better to consider a larger neighbourhood of grid points than the few grid points closest to the station location. As a consequence, a compromise should be found.

We evaluated two quantities to infer the best neighborhood dimension: the value of the near-surface temperature lapse rate and the absolute value of the innovation $|y^f - y^o|$ (i.e. the difference between the observed and the predicted value at station location). In particular, we study their statistical distribution (median and interquartile range) as the dimension of the area over which the lapse rate is estimated varies. We expect that the value of the lapse rate shows some important discontinuities when we move from smaller areas to bigger ones, as a result of the gradually inclusion of information even far from the station location. Instead, the variations in the innovation statistical properties as a function of the radius reveal information about the observation representativeness error (cfr. §1.2.3). The smaller the innovation values, the more comparable are the atmospheric scales described by both the observations and the background field (i.e. the smaller the observation representativeness error).

We considered the period between January 2015 and August 2016 and neighborhoods having an area of radius 2, 5, 10, 20, 25, 50, 75 and 100 km. Moreover, we filtered out all the values associated with a positive near-surface lapse rate, since we are mainly interested in the behaviour of the observation operator in normal conditions (no inversions taken into account). Results for each season and for both daytime (14 UTC) and nighttime (4 UTC) are shown in figs. (2.7)-(2.10).

The lapse-rate figures (2.7)-(2.8) show remarkable day-night differences, especially during JJA (bottom-left panels), MAM (top-right panels) and (less clearly) during SON (bottom-right panels). During nighttime in the warmer seasons (JJA, MAM: bottom-left and top-right panels in fig. (2.7)), the lapse rate remain fairly constant across the radius interval, with values around $\Gamma \simeq -7°C/Km$ in JJA and $\Gamma \simeq -8°C/Km$ in MAM; conversely, during the day (same panels, fig. (2.8)), the values become gradually bigger as the radius grows, with a transition from a lapse-rate of $\Gamma \simeq -8°C/Km$ to $\Gamma \simeq -6°C/Km$ in JJA and from $\Gamma \simeq -9°C/Km$ to $\Gamma \simeq -8°C/Km$ in MAM. During DJF differences between day and night are almost absent (compare top-left panels in figs. (2.7) and (2.8)), with steadily decreasing values from $\Gamma \simeq -7 \div -8°C/Km$ to $\Gamma \simeq -9°C/Km$. Furthermore, a sudden discontinuity between 2 and 5 km in all seasons and all hours of the day is clearly evident. In almost all cases (apart from DJF), the interquartile range of the lapse-rate distributions narrows as the radius increases.

The innovation values are generally smaller in the interval between 2 and 25 km, and then they tend to increase as the radius gets bigger (see figs. (2.9)-(2.10)). Generally speaking, they seem to be optimized by a radius of 5-10 km during nighttime (see fig. (2.9)) and 20-25 km during daytime (see fig. (2.10)). More specifically, the values are bigger during DJF (see top-left panels, figs. (2.9)-(2.10)), with a transition from $|y^f - y^o| \simeq 1.0°C$ to $|y^f - y^o| \simeq 1.3°C$ during nighttime and from $|y^f - y^o| \simeq 0.9°C$ to $|y^f - y^o| \simeq 1.1°C$ during daytime, as the radius increases. Values are smaller in the warmer seasons: during nighttime generally within $0.7°C < |y^f - y^o| < 0.9°C$ in JJA and within $0.9°C < |y^f - y^o| < 1.1°C$ during MAM (top-right and bottom-left panels in fig. (2.9)); during daytime within $0.9°C < |y^f - y^o| < 1.1°C$ in JJA and within $0.8°C < |y^f - y^o| < 1.0°C$ during MAM (top-right and bottom-left panels in fig. (2.10)). As regards the interquartile range of the innovation values, it becomes larger as the radius increases, especially during the colder seasong (DJF, SON: see top-left and bottom-right panels in figs. (2.9)-(2.10)).

All the previous considerations led us to choose $R = 10\,km$ as the optimal radius. This value represents a compromise between the minimization of the innovation values (a middle way between the daytime and the nighttime best radius) and the stabilization of the lapse rate as a function of R, which is affected by a strong discontinuity in the range between 2 and 5 km.

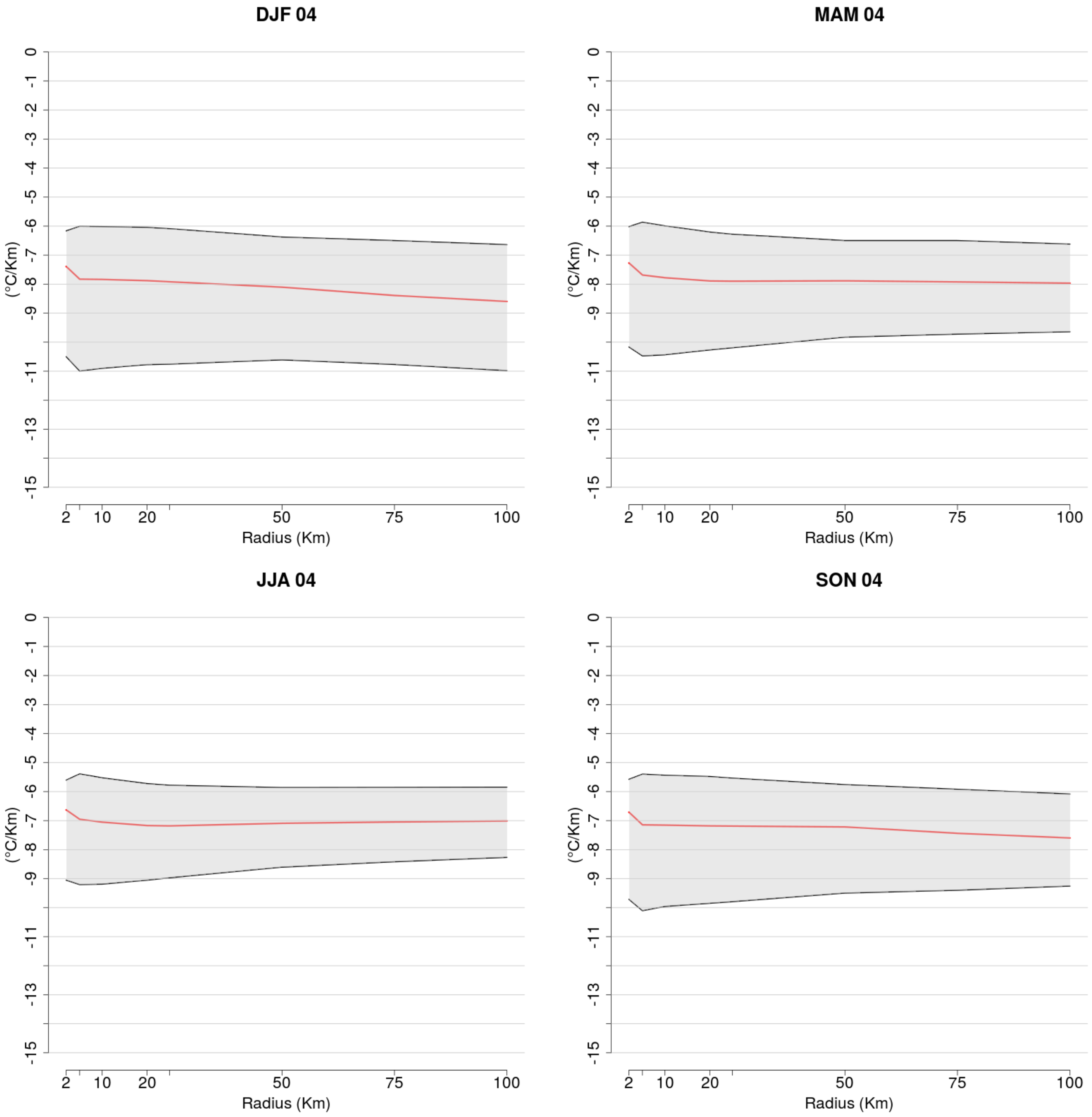


Figure 2.7: Nighttime near-surface temperature lapse rate as a function of the radius of the neighborhood for each season. Red lines indicates the median, the gray shaded area the interquartile range.

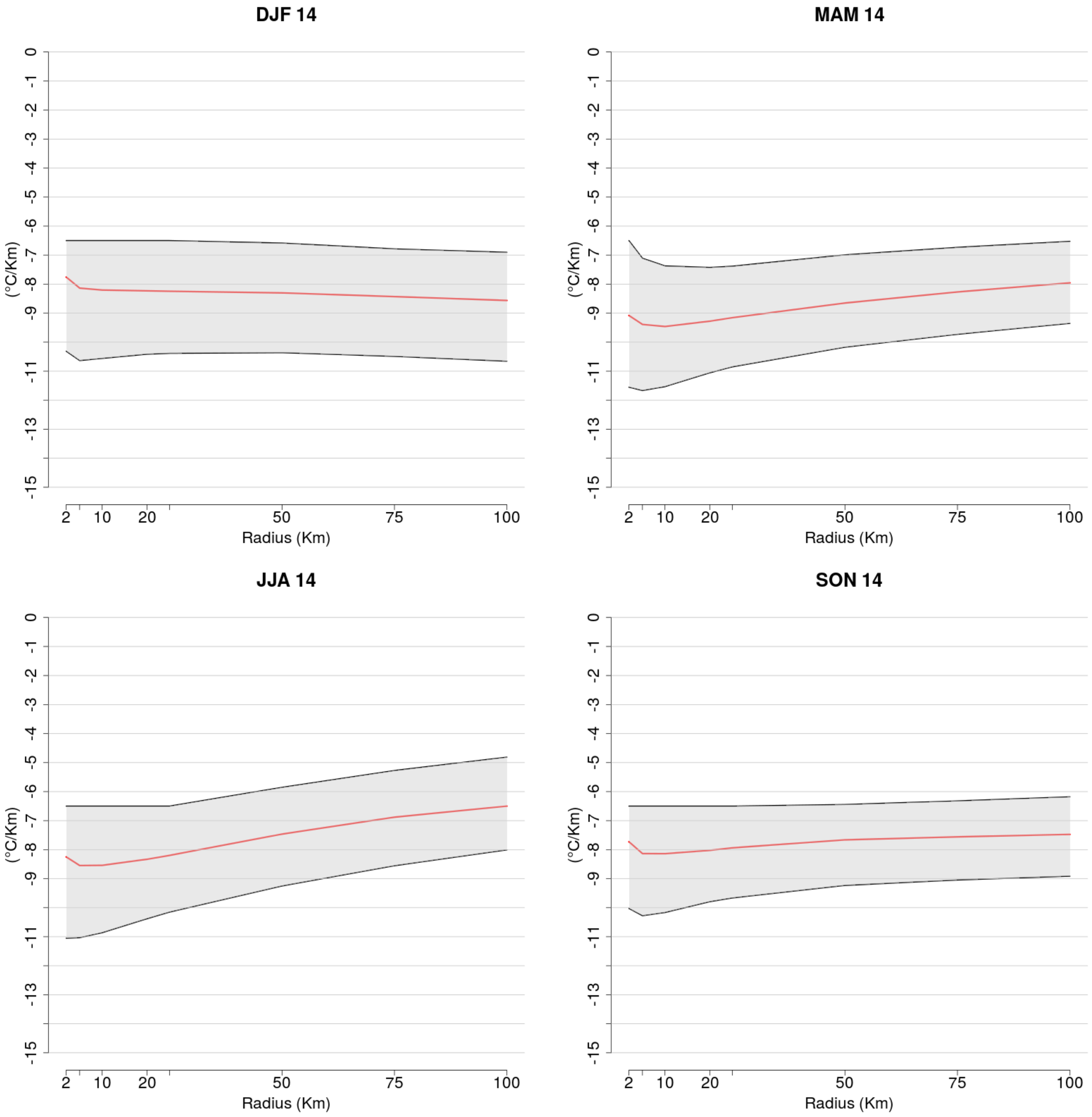


Figure 2.8: Daytime near-surface temperature lapse rate as a function of the radius of the neighborhood for each season. Red lines indicates the median, the gray shaded area the interquartile range.

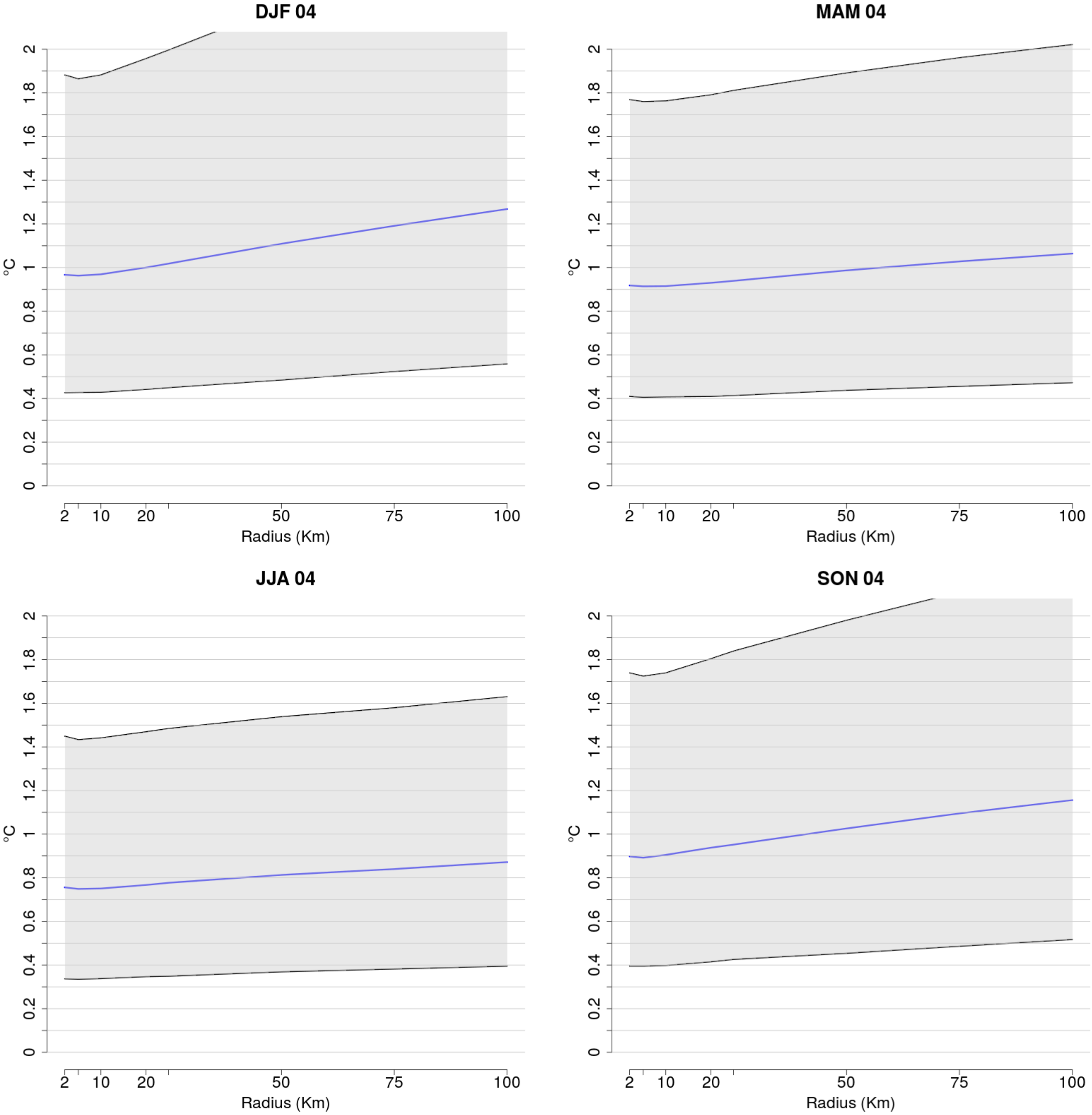


Figure 2.9: Nighttime innovation value as a function of the radius of the neighborhood for each season for the originale H operator. Blue lines indicates the median, the gray shaded area the interquartile range.

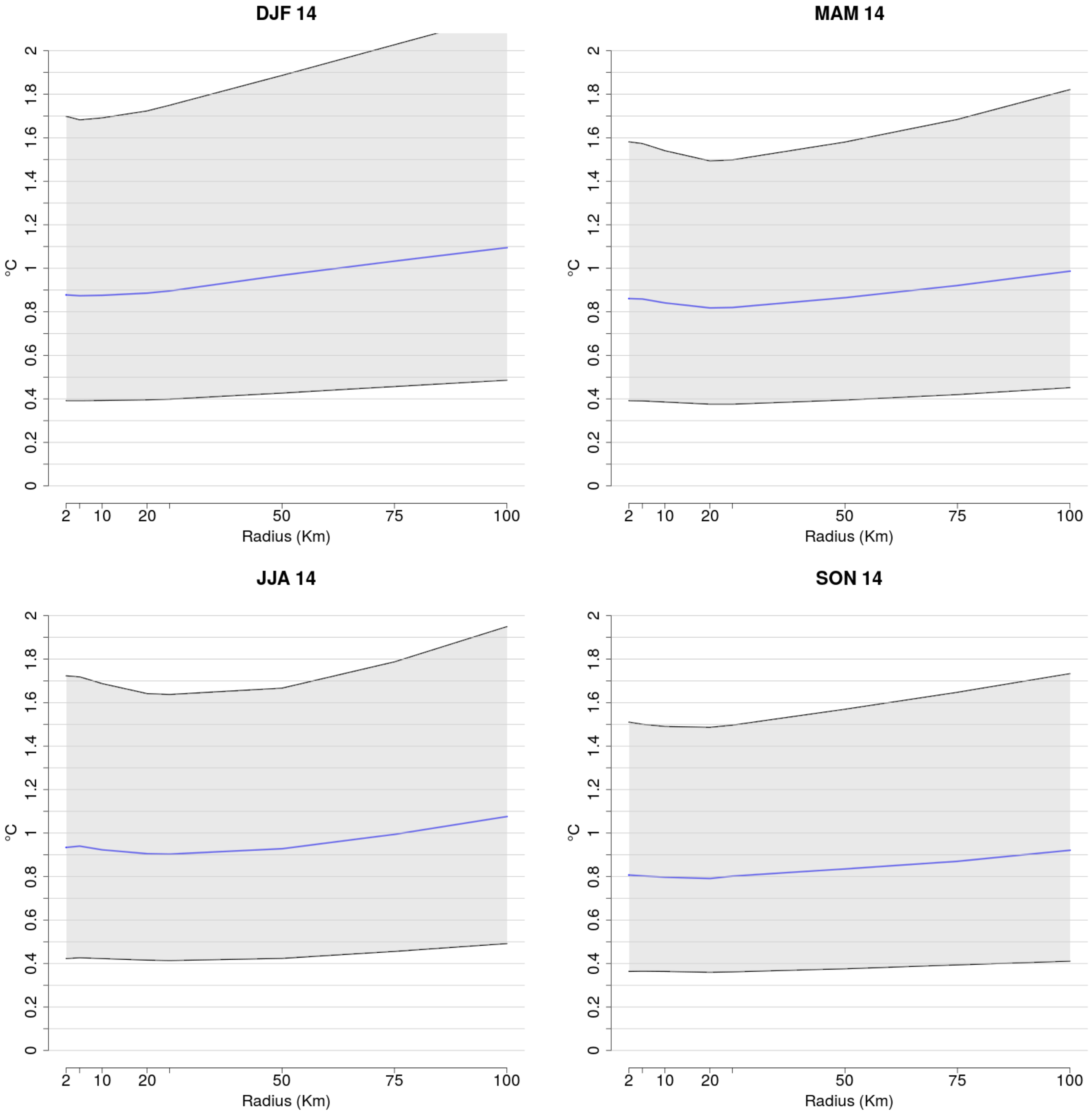


Figure 2.10: Daytime innovation value as a function of the radius of the neighborhood for each season for the original H operator. Blue lines indicates the median, the gray shaded area the interquartile range.

2.3 The spatial correlation and the variance of the background

The background error covariance matrix $\mathbf{P}^f$ is a good indicator of the spatial variability of the surface temperature as predicted by the model ensemble (cfr. eq. 1.13). Its role in the scheme is well depicted by the equation (1.11), where it clearly shapes the gain matrix $\mathbf{K}$. In other words, the matrix $\mathbf{P}^f$ defines the way in which the information is transported across the model space once the contribution of the observations and their error is known.

In order to visualize it better in our scheme, we compute the spatial correlation associated to some specific grid points (in the area around the three main norwegian cities: Oslo, Trondheim and Bergen) and the background variance over the whole model domain. We employ the spatial correlation instead of the spatial covariance because of its more straightforward interpretation as normalized quantity. We considered a period of one month, February 2016, and averaged data in order to obtain the hourly values of both correlation and variance.

Results valid for four different hours of the day (00, 06, 12, 18) are shown in figs. (2.11)-(2.12).

The spatial correlation referred to the selected subset of grid points is meant to show the strong impact of the norwegian territory on the spatial patterns of temperature. As a matter of fact, Norway has a very complex orography, with a variety of mountain ridges, valleys, fiords, shores and plains. This diversity in the landscape has a relevant impact on the surface temperature variability and this is clearly evident in fig. (2.11): without regarding the hour of the day, the correlation pattern along the western coast in the south seems to be confined by the presence of the inland mountains, whereas around Oslo, where the territory is more gentle, the temperatures are much more coherent. Moreover, values show a daily cycle: correlation patterns at 12 (bottom-left panel, fig. (2.11)) have a smaller spatial extension than they have during nightime (top panels, same figure).

The background variance (fig. (2.12)) shows a strong spatial variability as well. It is generally smaller in the southern Norway and along the coasts (with values $\sigma_o^2 < 1.5\ (^{\circ}C)^2$), whereas larger values are present in the hinterland, especially in the northern part of the country, where peaks such that $\sigma_o^2 > 5\ (^{\circ}C)^2$ are present. Moreover, a daily cycle is noticeable: values are larger during nighttime (top panels in fig. (2.12)) than daytime (bottom panels in fig. (2.12)).

These results gives us an indication of how the scheme does not spread the observations impact in the same way everywhere over the model domain and also how it is constrained and shaped by the orographic pattern.

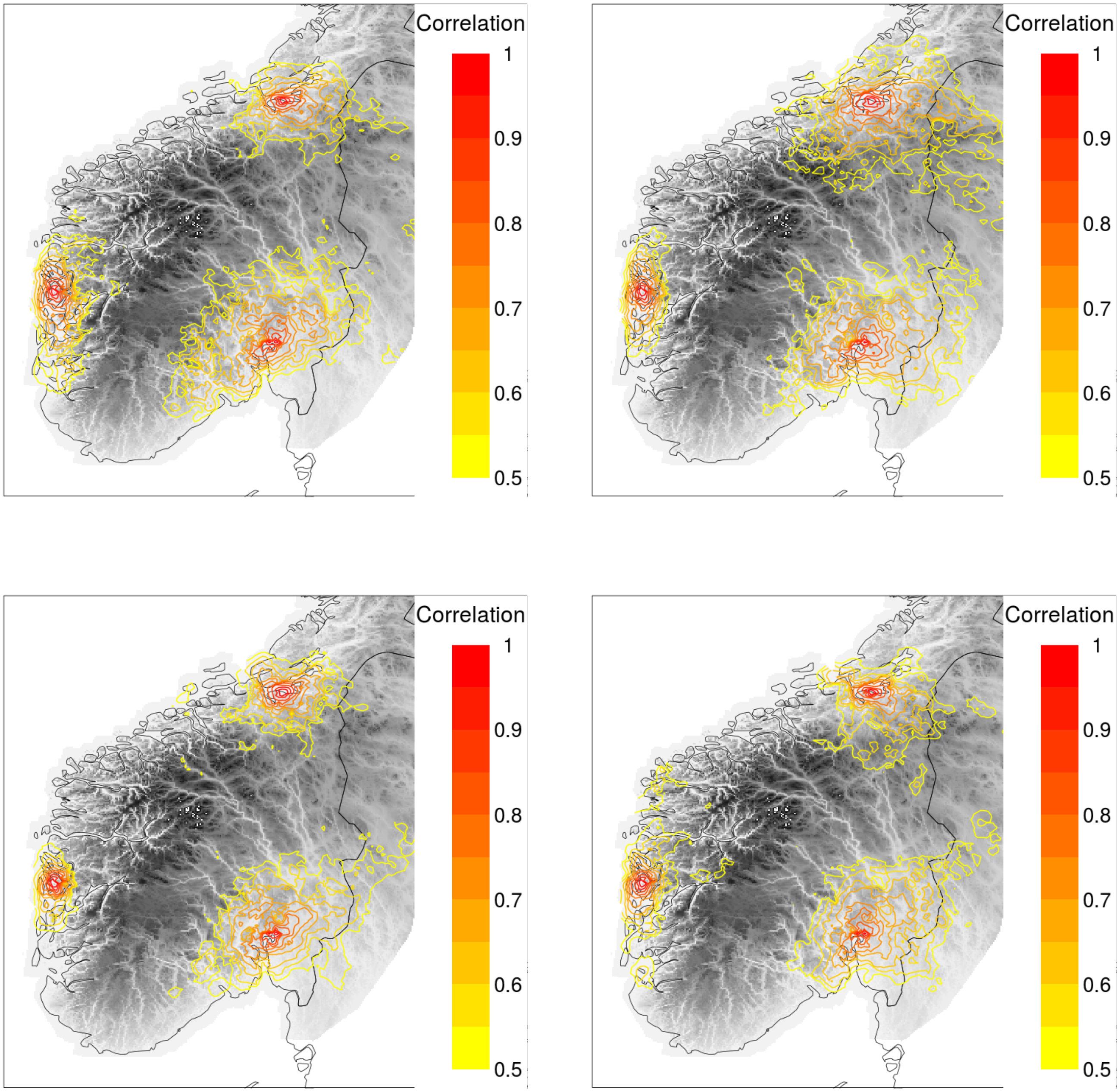


Figure 2.11: Spatial correlation of the predicted surface temperature. The monthly mean for February 2016 for four different hours (top left: 00, top right: 06, bottom left: 12, bottom right: 18) referred to the three grid points closest to stations Oslo Blindern (southern Norway, on the east coast), Bergen Florida (southern Norway, along the west coast) and Trondheim Lade (central Norway, along the coast) is shown. Only contours between 0.5 (yellow) and 1 (red) are drawn.

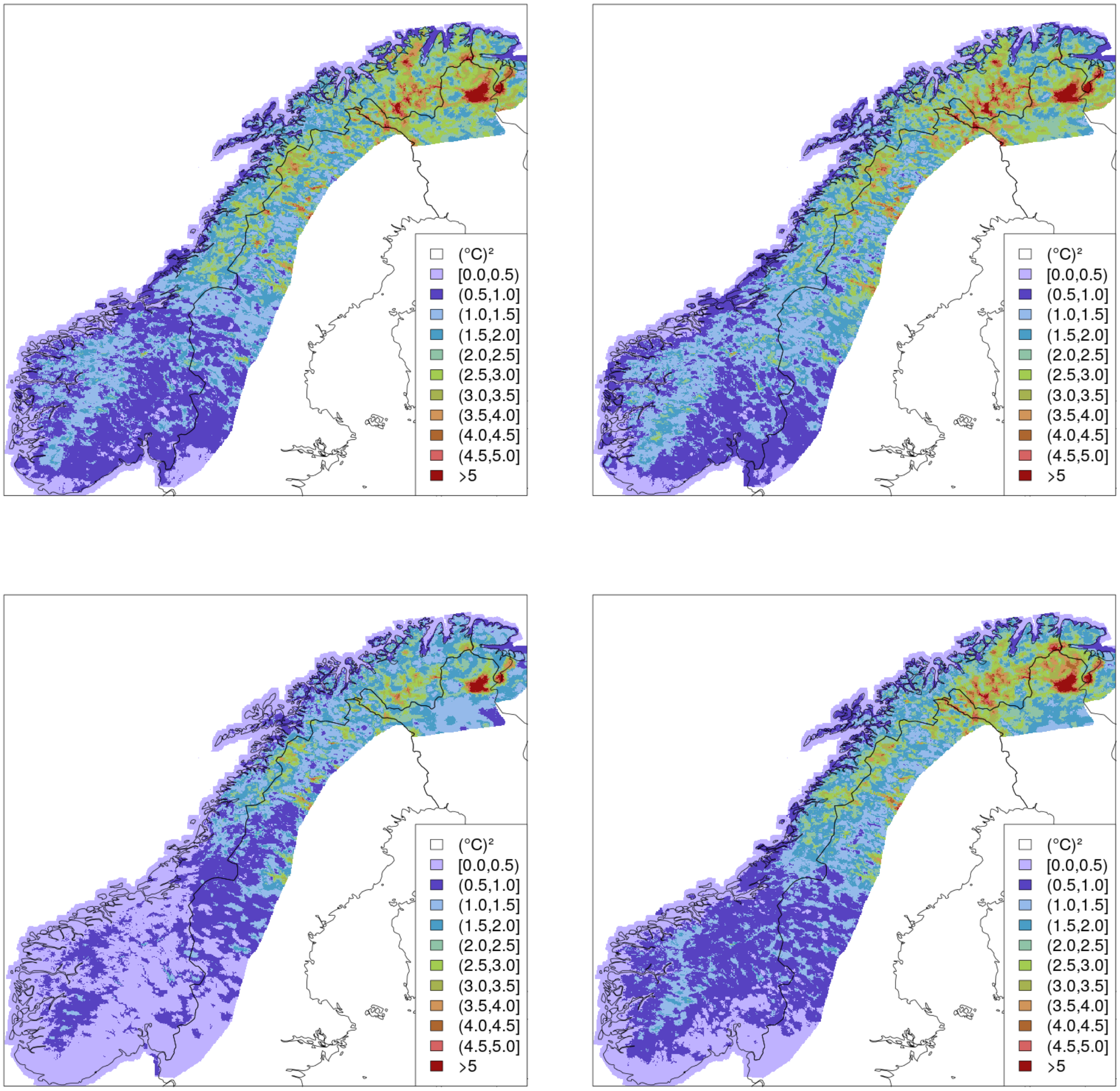


Figure 2.12: Variance of the predicted surface temperature. The monthly mean for February 2016 for four different hours (top left: 00, top right: 06, bottom left: 12, bottom right: 18) is shown.

2.4 Optimisation of parameter values

Besides the background and the observation error (described by the the matrices $\mathbf{P}^f$ and $\mathbf{R}$), the quality of the final surface temperaure analysis depends also on the calibration of a set of parameters controlling the localization distance (D_h and D_z, cfr. §1.2.5) and the bias correction (γ, cfr. §1.2.6). The values tested during the tuning phase have been listed in the table 2.1.

In order to understand which set of parameters produces the best analysis, we run the scheme several times performing a comparison among the final results in terms of the forecast skills and the statistical properties. We searched for those parameters that optimize the Cross Validation analysis and the scores described in Appendix 3. The final choice is highlighted in table 2.1, where the selected parameters are bold. We have not varied the bias-damping parameter μ during our tests since we did not want to affect the bias persistence hypothesis.

γ	0.01	0.2	**0.25**	0.3	0.5	0.9
$D_h[km]$	25	**50**	60	75	150	
$D_z[m]$	**200**	500				
μ	**0.9**					

Table 2.1: Lists of parameters to set and values assigned to them. The bold ones are those evaluated as the best ones.

2.4.1 The setting of the bias-correction parameter

γ controls the magnitude and the reactivity of the bias correction to the difference between the forecast and the observation. As a matter of fact, the bigger the value of γ, the greater the weight given to the innovation in the equation (1.37) through the matrix $\mathbf{L}$. Accordingly, the updated bias $\mathbf{b}^f$ will be adjusted fastly to the istantaneous value of the innovation or it will rely more on the bias prediction (i.e. the bias updated coming from the previous time instant damped by the scalar μ, cfr. eq. (1.36)).

The impact of γ on the scheme is quite evident in fig. (2.13), where a time series of the mean value of the updated bias $\mathbf{b}^f$ over the 80 grid points closest to the cross-validating stations is shown. Very small values of γ (e.g. $\gamma = 0.01$, the blue line in fig. (2.13)) lead the scheme to be nearly bias-blind, making the bias correction totally ineffective and superfluos. Coversely, high values of γ (e.g. $\gamma = 0.9$, the green line in fig. (2.13)) make the scheme very reactive, and its behavour is almost the same as the istantaneous innovation (black line, same figure). However, this reactivity undermines the stability of the scheme itself: in the right panel of fig. (2.14), a plot of the ensemble mean surface temperature analysis for $\gamma = 0.9$ highlihghts an area where the temperature tends to diverge after some days of running ($-25^\circ\,C < T < -22^\circ\,C$) with respect to the observed value ($T \simeq -15^\circ\,C$). Then, as a conservative solution, we chose the value $\gamma = 0.25$ as the best compromise, also after having verified that small variations around this value ($\gamma = \{0.2, 0.3\}$) do not affect in a significant way the performances of the final analysis. This solution also avoids the stability problem (as it is evident from the left panel of fig. (2.14)). Moreover, the fig. (2.13) shows that this value is big enough to make the scheme reactive to the most significant bias deviations.

Figure 2.13: Daily timeseries of the quantity $\mathbf{b}^f$ for three different values of γ, averaged over the cross-validating stations. The black line indicates the innovation value $y^o - y^b$.

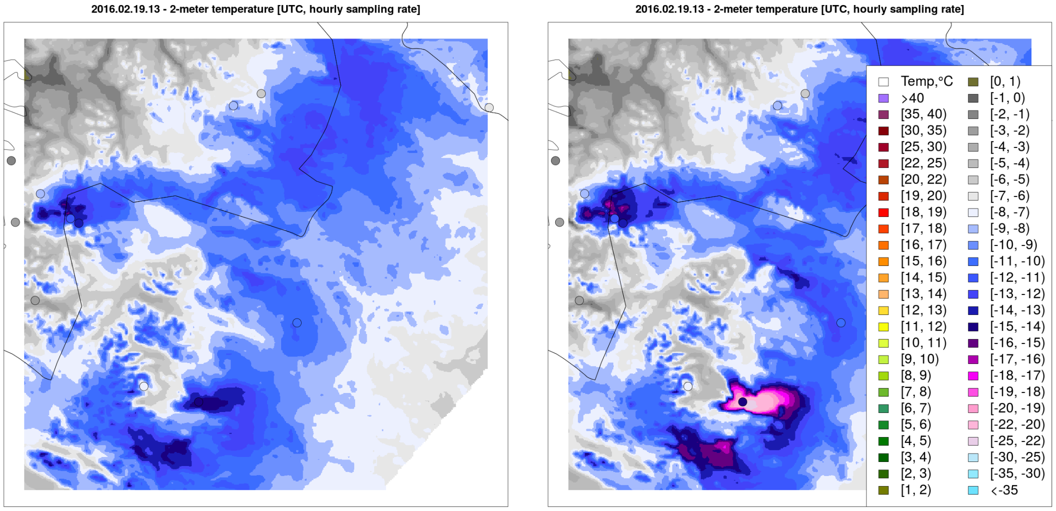


Figure 2.14: Plot of the ensemble mean of the final analysis over a restricted area in the northern Norway for $\gamma = 0.25$ (left panel) and $\gamma = 0.9$ (right panel).

2.4.2 The setting of the vertical localisation distance

D_z is the localisation parameter in eq. (1.29) which selects a decaying value for the elevation difference between each grid point and the surrounding observations.

In order to set this parameter, we had to take into account the station density in the network, in order to guarantee the presence of enough stations in the neighbor of the grid point without including stations located at a totally different elevation; as a matter of fact, mixing information at different altitudes could alter the spatial coherence of the surface temperature pattern in presence of certain orographic features like valleys and mountain ridges. According to the fig. (2.3) in chapter §2, we decided to set $D_z = 200\ m$. This value includes the 90th percentile of the distribution (indicated by a red line in fig. (2.3)) and so ensures that the analysis at the most of the grid points can involve at least three stations in such an interval. Nonetheless, we tested also the value $D_z = 500\ m$ to confirm our hypothesis. Even though the statistical and the probabilist scores do not seem heavily affected, in fig. (2.15) a plot of the analysis ensemble mean shows how bigger values of D_z tend to destroy the coherence of the 2-meter temperature field (look at the differences between the right panel, where$D_z = 500\ m$, and the left one, where $D_z = 200\ m$) which in turn results in a worse definition of some important orographic features, like narrow valleys and ridges.

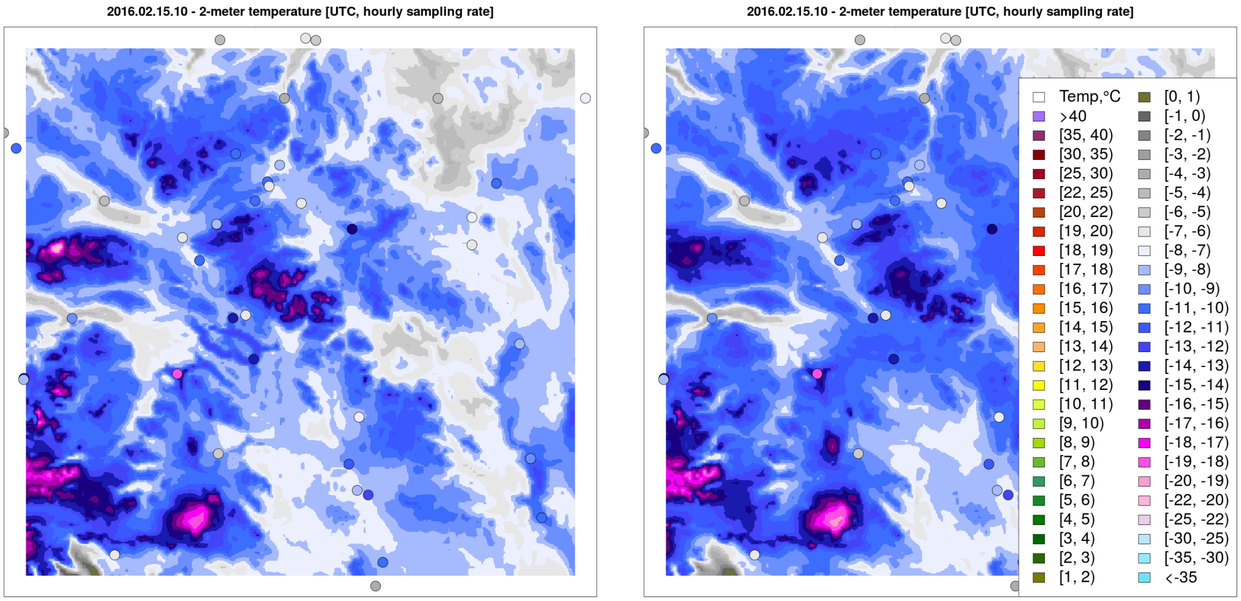


Figure 2.15: Plot of the ensemble mean of the final analysis over a restricted area in the southern Norway for $D_z = 200\,m$ (left panel) and $D_z = 500\,m$ (right panel).

2.4.3 The setting of the horizontal localisation distance

D_h is the localisation parameter in eq. (1.29) which selects a decaying value for the horizontal distance between each grid point and the surrounding observations.

As in the previous case, the station density in the network has played a fundamental role to set this parameter. According to the fig. (2.3) in chapter §2, the value of D_h which includes the 90th percentile of the distribution is $D_h = 50\ km$. This threshold ensures that the analysis at the most of the grid points can involve at least three stations within such a distance. Nonetheless, we tried to vary the value of D_h in order to study its impact on the scheme performances. Figs. (2.16)-(2.17) show the root mean squared error and the brier score for a subset of the D_h parameter (that is $D_h = \{25, 50, 150\}\ km$). Even though only minor differences are noticeable, both $D_h = 25\ km$ and $D_h = 150\ km$ tends to increase the root mean squared error - with a decrease in the precision. Moreover, the choice of $D_h = 150\ km$ tends to increase the value of the brier score for the 0 celsius threshold - with a deterioration of the overall performances of the analysis ensemble - whereas the $D_h = 25\ km$ value could mine the analysis quality in those grid points with few stations around. Lastly, the other values taken into consideration (listed in the table (2.1), all bigger then $D_h = 50\ km$ but smaller than $D_h = 150\ km$) do not seem to affect in a considerable way the performances of the scheme. For all these reasons, we opt for $D_h = 50\ km$ as decorrelation value.

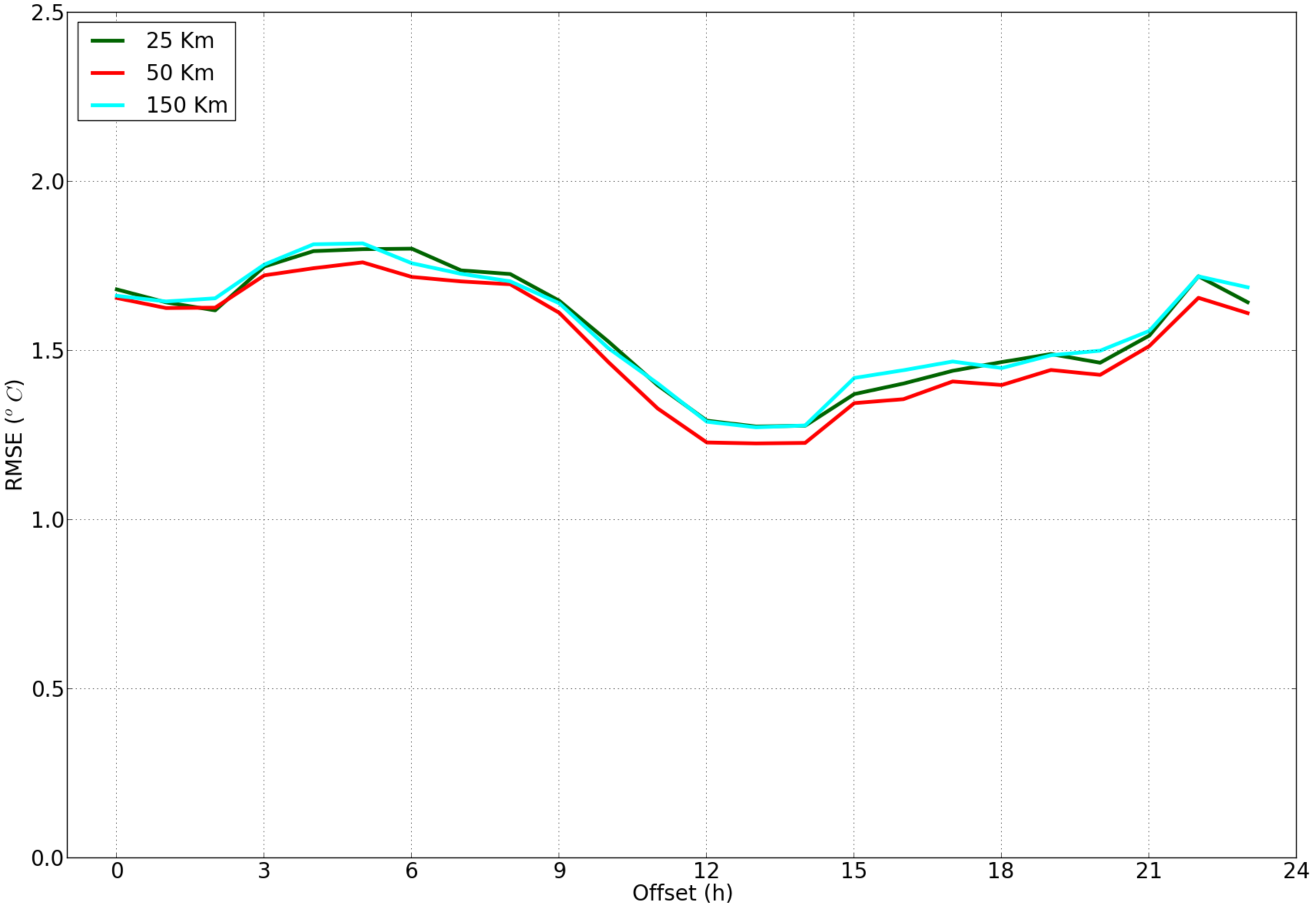


Figure 2.16: Mean hourly values (February 2016) of the root mean squared error computed for three different values of D_h, averaged over the cross-validating stations.

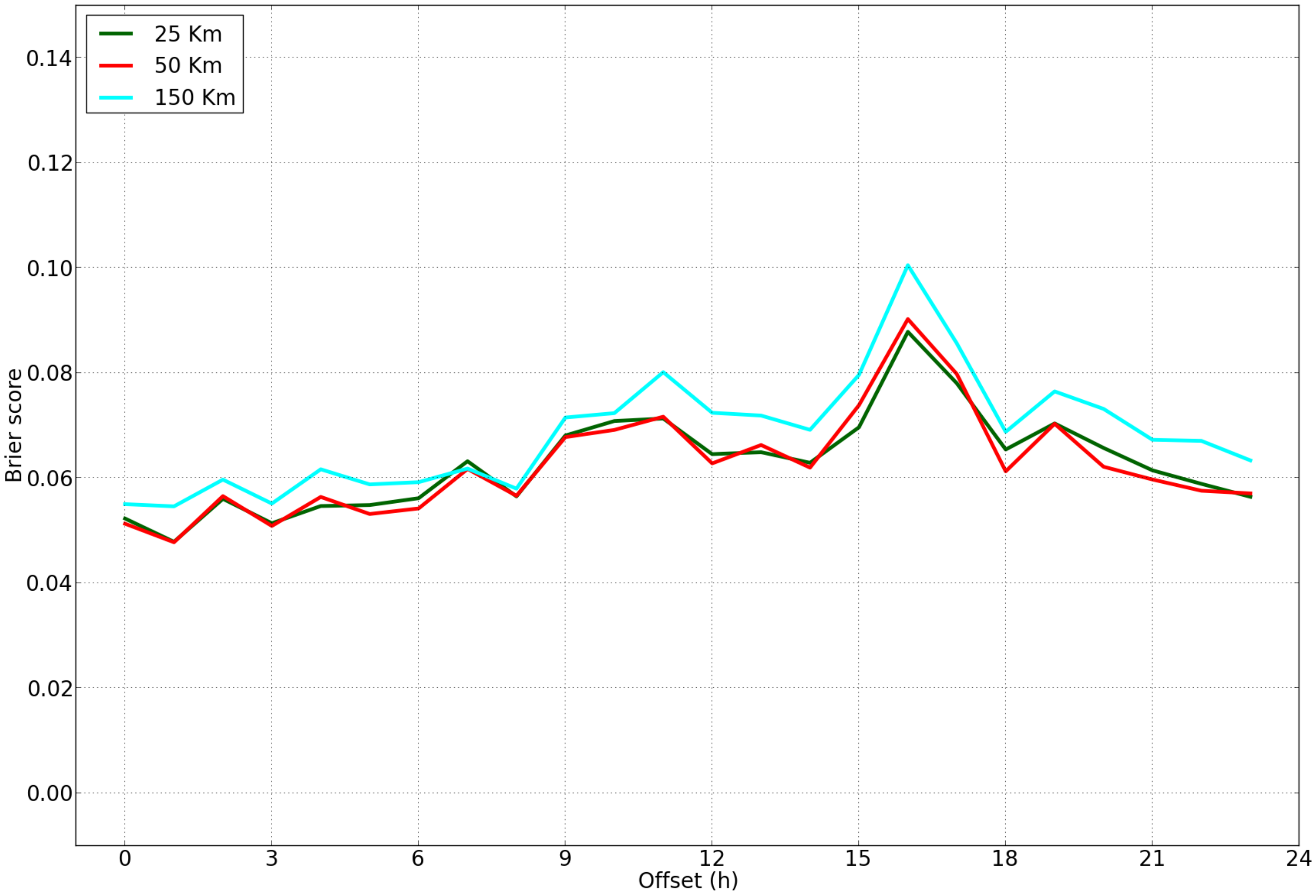


Figure 2.17: Mean hourly values (February 2016) of the Brier Score computed for three different values of D_h, averaged over the cross-validating stations.

Chapter 3

Results

In this chapter we present the main outcomes resulting from the scheme application. We performed an evaluation of the analysis quality by means of both statistical and probabilistic indices, in order to address the behaviour of both the mean value and the ensemble as a whole. We will compare the final analysis with the Arome MetCoOp 2.5 km fields, performing the assessment for both a case study of seven days in February 2016 and an extended period of 6 months from February to July 2016. We will try to sum up and draw the big picture of the results in the last part of the chapter.

3.1 Scheme evaluation

The added value of the scheme with respect to the model only is the contribution given by the observations, which coveys information about the surface temperature of their surrounding area, taking implicitly into account features and properties of the surface at scales lying below the resolution of Arome MetCoOp 2.5 km. The final result is a 2-meter temperature field, still in the model space, which is better (as we are going to prove) at describing the local behaviour of the temperature close to the surface than the model only.

In order to quantify such ability, we needed to look at the statistical and the probabilistic skills of both our final analysis and the model prediction. We used the observations at the station locations as a common reference, since we were interested in detecting the ability of the analysis and the model in reproducing the temperature field locally. Yet, we could not use the observations used in the scheme to verify the scheme itself: in those places we forced the analysis precisely towards those values. Therefore, we had to perform the evaluation using a restricted set of stations which was not employed in the scheme, in order to achieve an independent evaluation of the quality of the analysis. The value of the analysis computed over these locations through the observation operator is a Cross-Validated (CV) analysis.

In figure 3.1 the sample of 80 stations used for our CV analysis is shown. They were randomly selected among the stations performing more than 90% of valid observations during February 2016 (cfr. black circles in fig. 2.1) and removed from the list of stations which the scheme assimilates.

An one month period of evaluation (February 2016) has been employed in order to perform the tuning of the parameters (cfr. §2.4). Once the best set of parameters have been selected, a case study and an extended period of six months (from February to July 2016) have been used for a deeper evaluation.

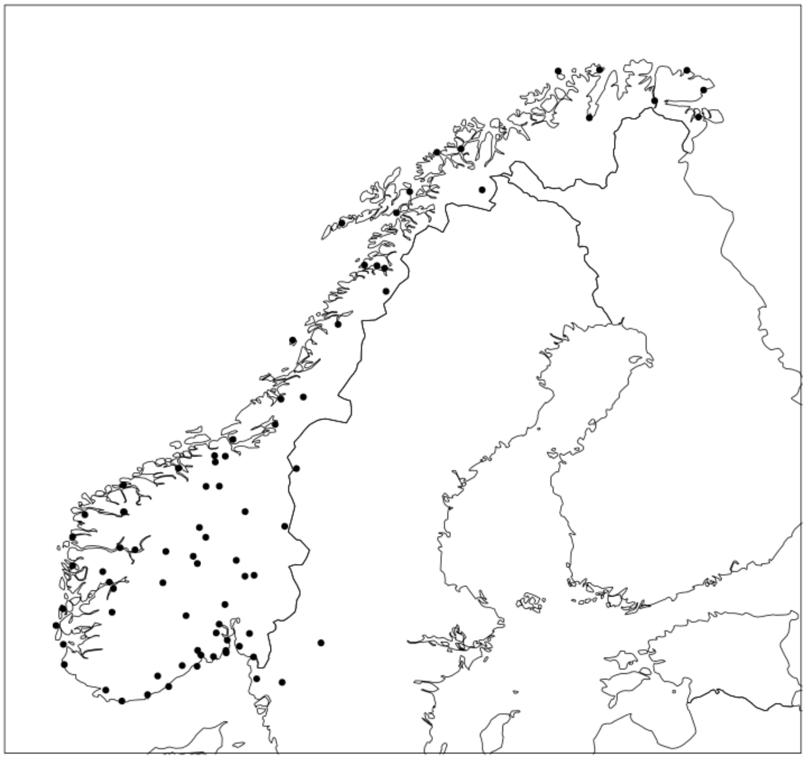


Figure 3.1: Sub-sample of stations used for the cross-validated analysis.

As we pointed out in §1.2.2, an ensemble of fields is basically describable by means of its mean value and its uncertainty (the latter conveyed by the spread of its members). Therefore, the evaluation has been carried out addressing the behaviour of both the mean value and the ensemble spread of the final analysis.

The mean value can be considered as a deterministic (or nonprobabilistic) prediction. The most useful tools to evaluate its skills against the observation are the bias and the root mean square error (RMSE). The former is a good indicator of the accuracy of a prediction, the latter is commonly employed to evaluate the precision.

The uncertainty affecting the mean value is conveyed by the ensemble spread. Thus, in order to evaluate its plausibility we have to treat the ensemble as a probabilistic prediction, using some skill scores like the Brier Score, the reliability diagram and the PIT histogram to infer information about its performances.

In order to compare the analysis with Arome, we have employed the best estimate as the deterministic prediction (cfr. §1.2.2), rescaling all the model ensemble members around it to recreate a coherent ensemble to be used as probabilistic prediction in the evaluation process.

The definitions of all the employed scores are listed in the Appendix 3.

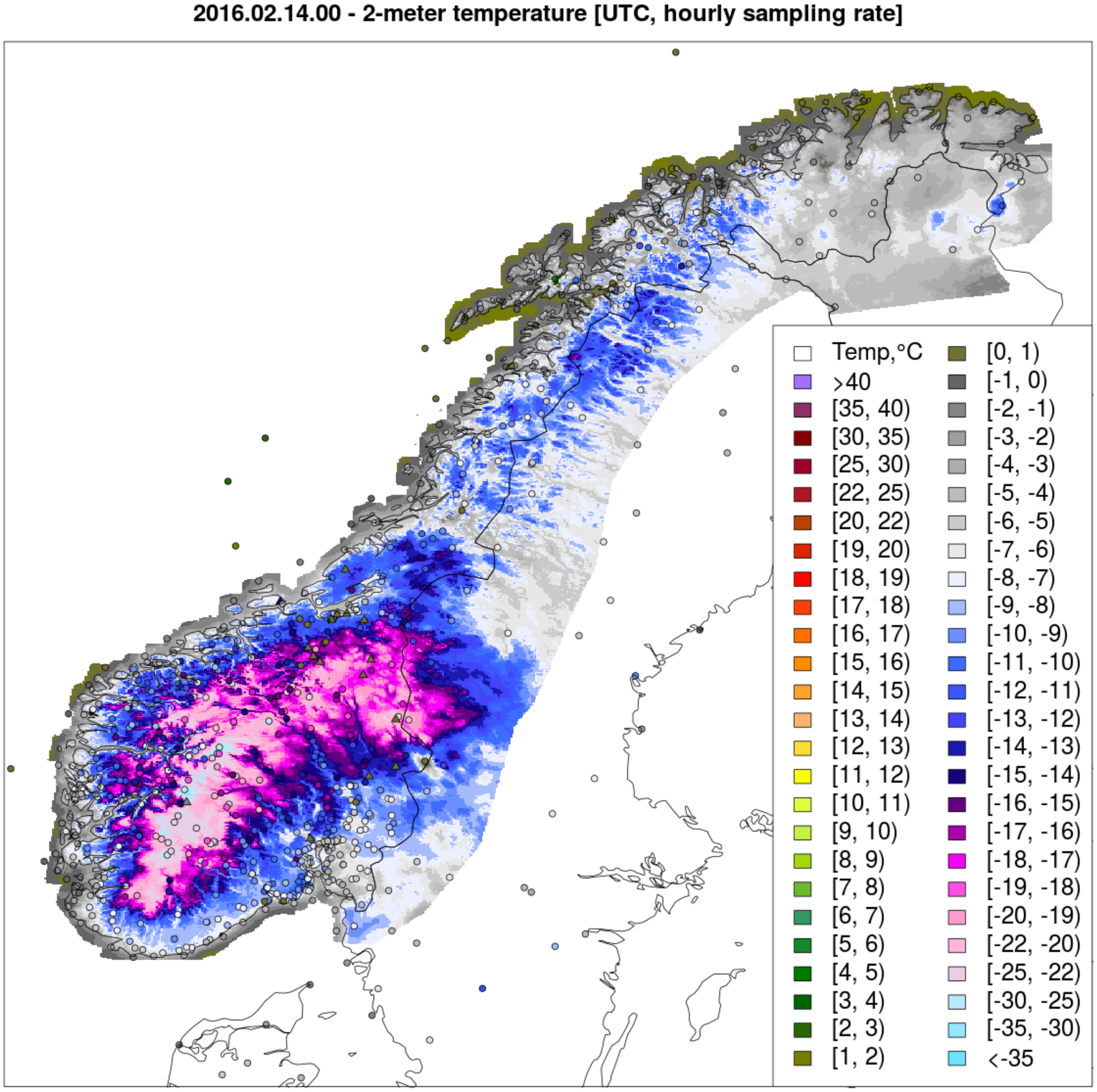


Figure 3.2: An example of the scheme application: analysis ensemble mean, 2-meter temperature, at 00 UTC on 14th February 2016. Circles indicate the valid observed values. Triangles represent the discarded observations.

3.2 The evaluation of the scheme for the selected parameters

Once we select the best set of parameters (see §2.4, values bold in Table 2.1), we evaluated the performances of the scheme for both a case study and a 6-months period (from February to July 2016).

3.2.1 February 2016: a case study

As we have previously noted, Norway has a very complex orography which can lead to a strong surface temperature variability. Such a variability is enhanced under certain conditions. For istance, stable periods after cold air advections during wintertime lead to widespread temperature inversions. Since our scheme is aimed at improving the spatial description of the 2-meter temperature field, we tried to test its reliability in such particularly challenging conditions and selected the week between the 12^{th} and th 19^{th} February 2016. In figure (3.2), an example of the hourly final 2-meter temperature analysis field (the ensemble mean) during the mentioned period is shown.

As regards the statistical indices, figure (3.3) shows the hourly timeseries of the bias (upper panel) and the root mean squared error (bottom panel) affecting both the analysis (red line) and the model (black line) during the considered period. Their evolution in time shows a superposition of both a daily and a synoptic cycle, with a clear alternation of maxima and minima (expecially between the 12th and the 16th) and a more long-term fluctuation resulting in an overall decrease of the RMSE values and an overall increase of the bias across the period. In both cases the analysis mean performs better than the model best estimate (smaller values of both bias and rmse), with a significant improvement in both accuracy and precision. The model bias fluctuates within $-1.2^\circ\,C \div 1.2^\circ\,C$ and it is almost the double of the analysis bias, confined in the interval $-0.5^\circ\,C \div 0.5^\circ\,C$. The interquantile range plotted for the analysis bias (the red shadow which takes into account also the other analysis members) includes almost constantly the zero line and its width varies between a minimum of $0.1^\circ\,C \div 0.2^\circ\,C$ (around the 14th Februry) and a maximum of $0.5^\circ\,C \div 0.6^\circ\,C$ (around the 18th). The RMSE values of the model best estimate are generally larger than the analysis mean by $0^\circ\,C \div 0.5^\circ\,C$, and both fluctuate between $1.0^\circ\,C \div 3.5^\circ\,C$.

The improvement in precision from a spatial point of view is also evident in figure (3.4), where a map shows a comparison of the mean values of the RMSE at the CV station locations in the southern Norway between the model (right panel) and the analysis ensemble mean (left panel). The main reductions in the RMSE values affects the stations along the south-eastern coast, with values decreasing from $1^\circ\,C \div 2^\circ\,C$ (green and yellow circles) in the model panel towards $0.5^\circ\,C \div 1.0^\circ\,C$ (blue circles) in the analysis one.

In figure (3.5), a Quantile-Quantile plot coupling a random sub-set of the predicted values against the corresponding observed ones is shown. In such an image it is possible to appreciate the increased ability of the analysis in reproducing the most extreme and coldest values (i.e.

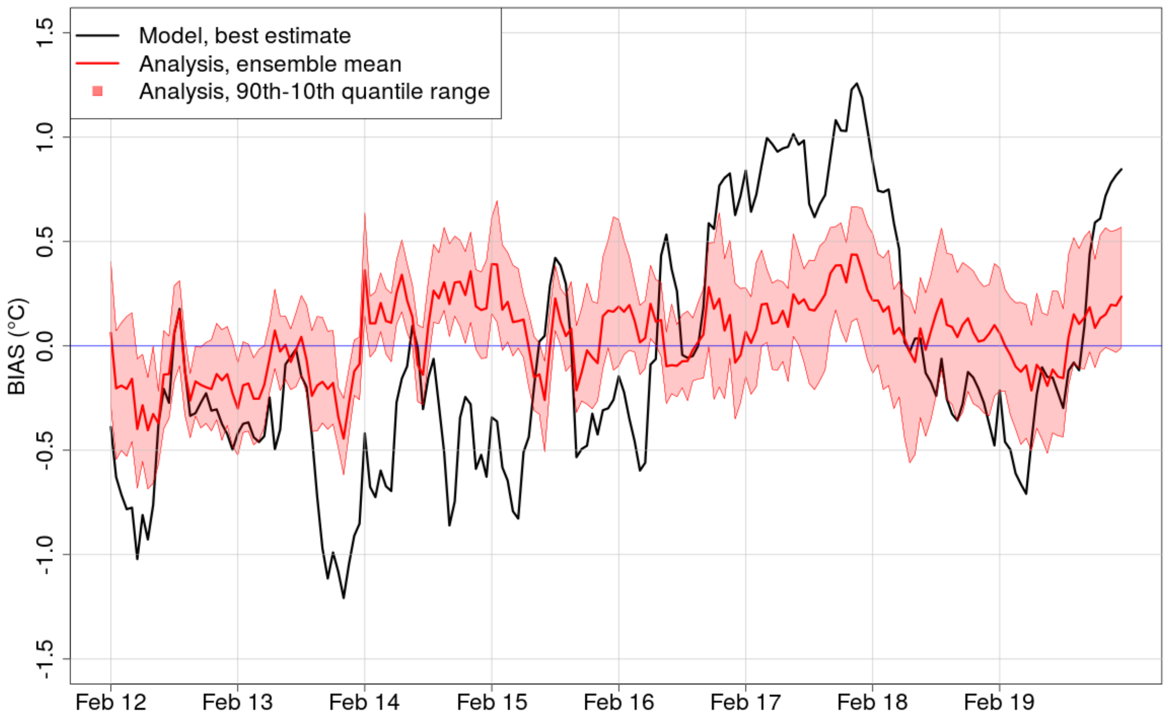
$T_{2m} < -20^{\circ}\,C$) with respect to the model only. As a matter of fact, even though both the model and the analysis seem to overestimate the observed values, the analyis sample (red circles) tends to remain closer to the gray diagonal (representing the perfect score) in the bottom-left corner of the figure than the model only (black circles), resulting in a better correspondence between the ideal prediction and the actual one. A slight improvement is also evident for the values in the top-right corner (i.e. $T_{2m} > 5^{\circ}\,C$).

Figures from (3.6) to (3.8) show the Brier Score computed for the threshold $T_{2m} > 0^{\circ}C$ together with its components (reliability and resolution). An overall improvement of the analysis ensemble is clearly evident in all the line charts, with a significative reduction in both the Brier Score and its reliability component and an equivalent increase in the resolution term.

Figure (3.6) shows both the daily (top panel) and the hourly mean (bottom panel) of the Brier Score during the considered period. The daily values fluctuate below $BS = 0.1$ and show a rapid transition within the period, with a net increase in the values for both analysis (red line) and model (black line): after a minimum on the 14th, a maximum is reached around the 17th February; the analysis (red line) shows smaller values (by $0.01 \div 0.04$) especially in the second half of the period. The hourly timeseries (bottom panel) shows a sistematic reduction in the Brier Score in favor of the analysis by $0.01 \div 0.02$ across all the hours of the day; the values of the hourly Brier Score for the threshold $T_{2m} > 0^{\circ}\,C$ tend to be larger during daytime and fluctuate within $0.04 < BS < 0.09$ for the model and within $0.03 < BS < 0.07$ for the analysis.

Figure (3.7) displays the daily timeseries of both the reliability (top panel) and the resolution term (bottom panel) of the Brier Score. Apart from the 13th February, the analysis (red line) shows a significant reduction in the reliability component with respect to the model (black line), which shows two peaks above $BS_{REL} > 0.01$, whereas the analysis values remains below $BS_{REL} < 0.007$. In the resolution chart (bottom panel), the analysis (red line) displays values constantly larger than the model (black line), but the difference is more marked in the second half of the period, when the values are subjected to an overall increase from $BS_{RES} < 0.05$ to $BS_{RES} > 0.15$.

Lastly, the hourly timeseries of both the components of the Brier Score are shown in figure (3.8). The top panel shows the reliability term: both the analysis (red line) and the model (black line) fluctuate around $BS_{REL} = 0.05$, even if the model displays higher values especially during the nightime. The bottom panel shows the resolution term: again, the analysis (red line) displays values slightly and constantly larger (by $\Delta BS_{RES} \simeq 0.01$) than the model (black line), with an overall increase in the resolution term during the daytime.



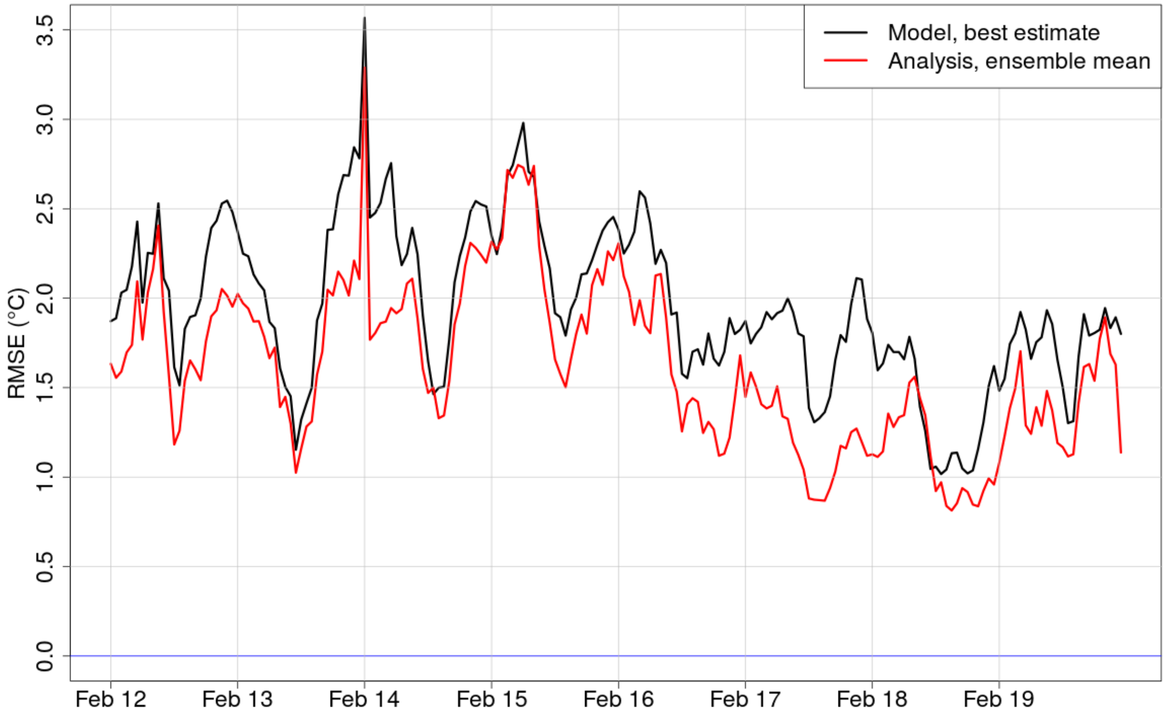


Figure 3.3: Bias (top figure) and root mean squared error (bottom panel) affecting the model (black line) and the analysis (red line) during the case-study period. The blue line marks the perfect score.

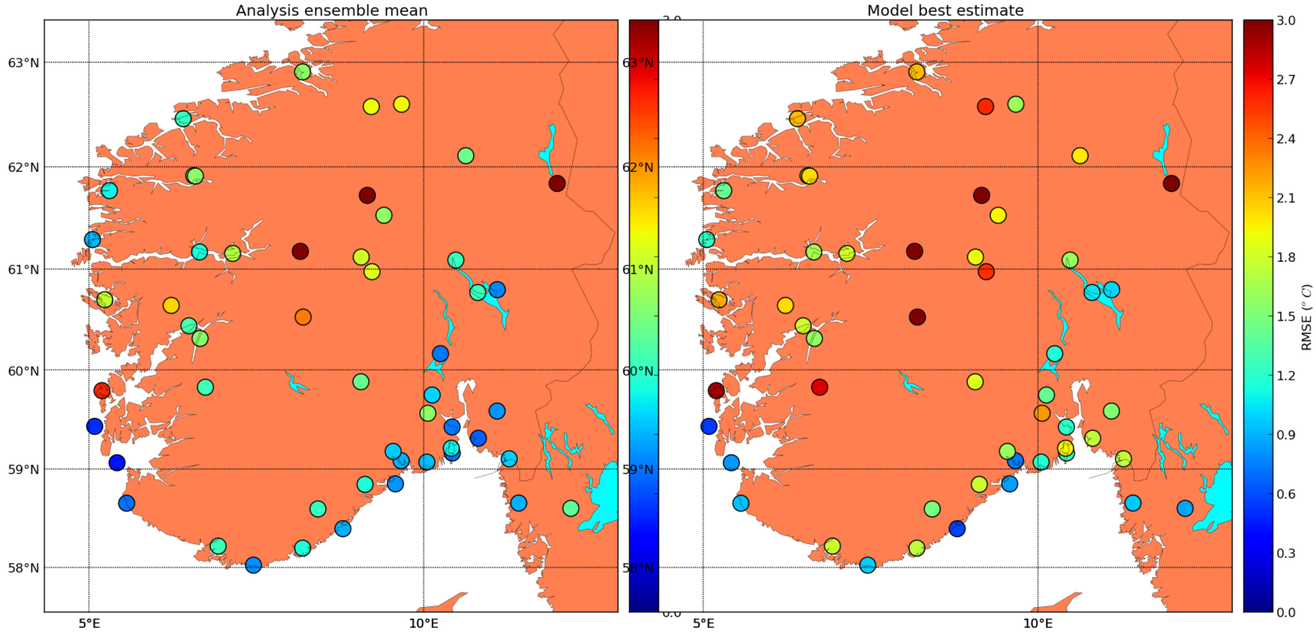


Figure 3.4: Analysis (left panel) and model (right panel) root mean squared error averaged over the case study period for each of the cross-validated stations in the southern Norway.

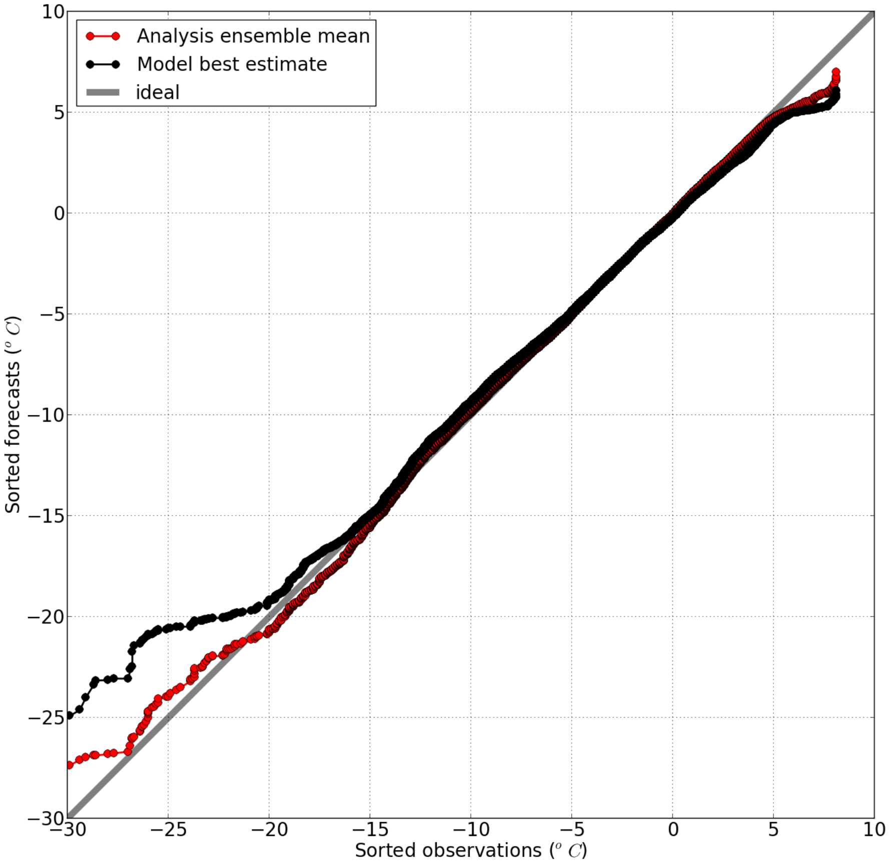


Figure 3.5: Quantile-Quantile plot of a random sub-sample of the predicted values against their observed ones

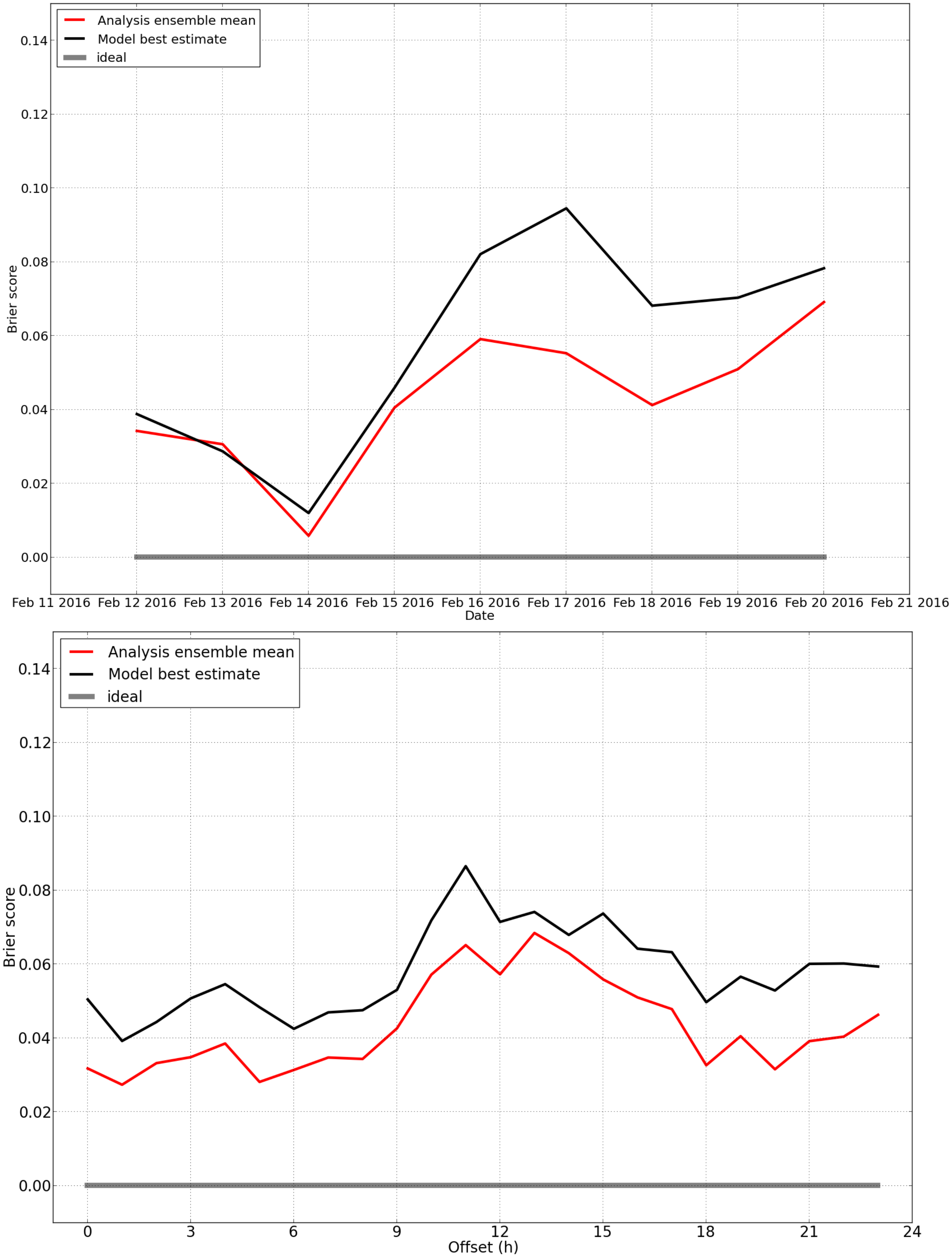


Figure 3.6: Brier Score computed for the threshold $T_{2m} > 0^{\circ}C$ during the case study period. Daily timeseries (top panel) and mean hourly values (bottom panel).

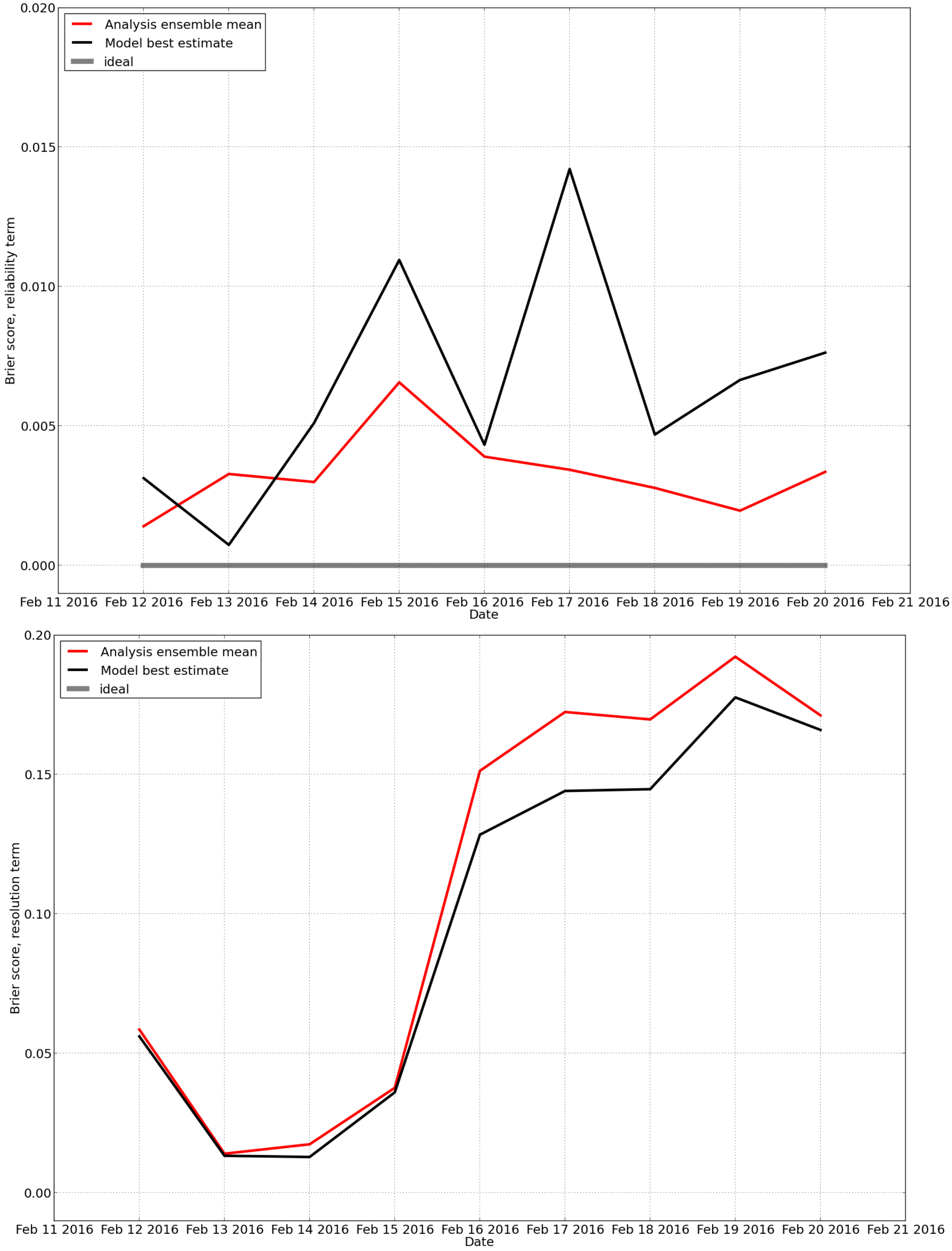


Figure 3.7: Daily timeseries of both the reliability (top panel) and the resolution (bottom panel) component of the Brier Score computed for the threshold $T_{2m} > 0^{\circ} C$ during the case study period. Ideal value not present in the resolution term.

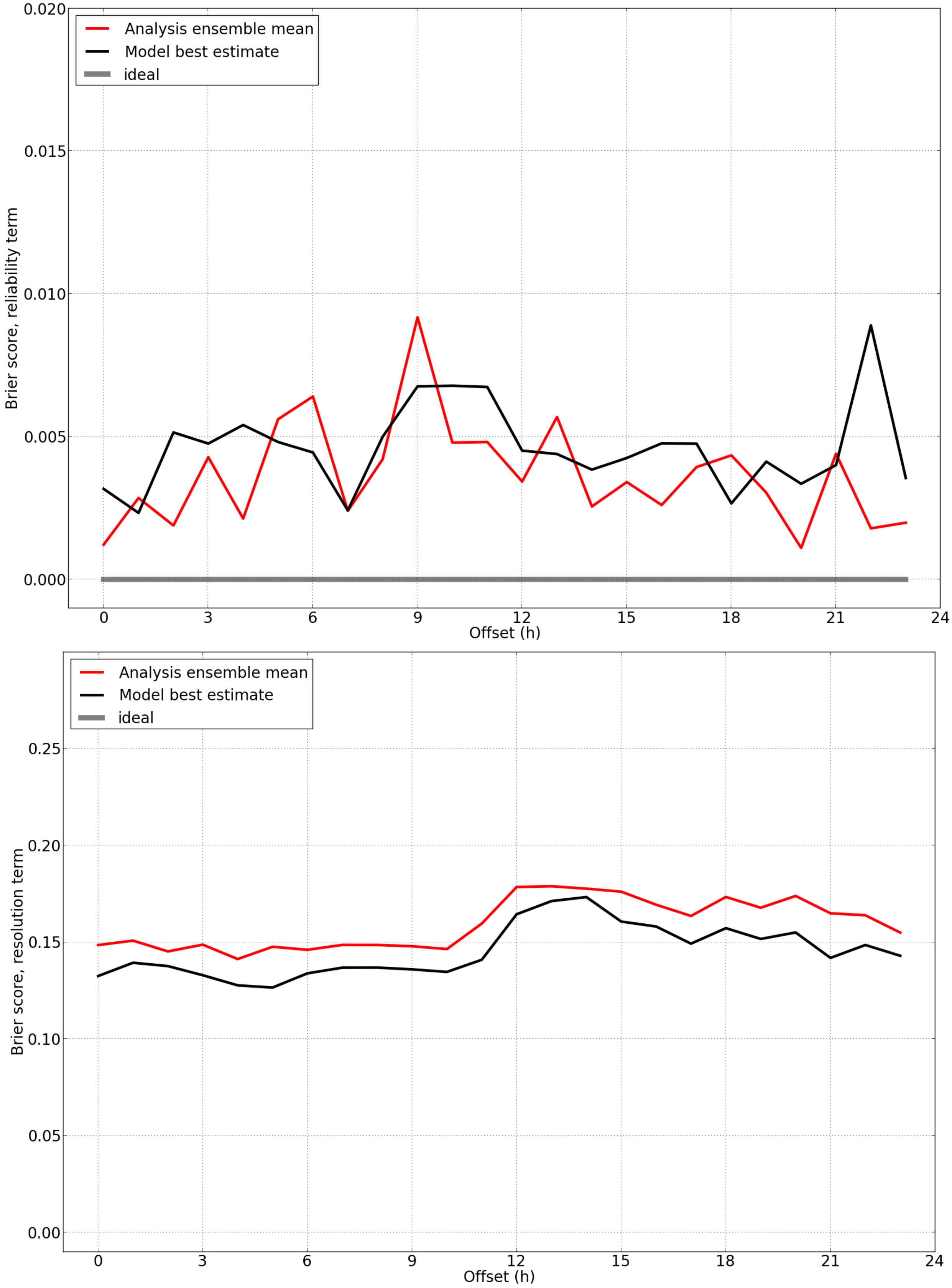


Figure 3.8: Mean hourly values of both the reliability (top panel) and the resolution (bottom panel) component of the Brier Score computed for the threshold $T_{2m} > 0^{\circ}\,C$ during the case study period. Ideal value not present in the resolution term.

As an example of the potential of the scheme in such extreme cold phases during wintertime, in figure (3.9) a map of the probability of $T_{2m} < -20^\circ C$ for the same istant as that of fig. (3.2) is shown. This kind of elaboraton is possible in virtue of the fact that the final outcome is an ensemble of analysis. As it appears evident from the figure, only the mountain regions in the south display middle to high probabilities of suffering these extreme values of temperature.

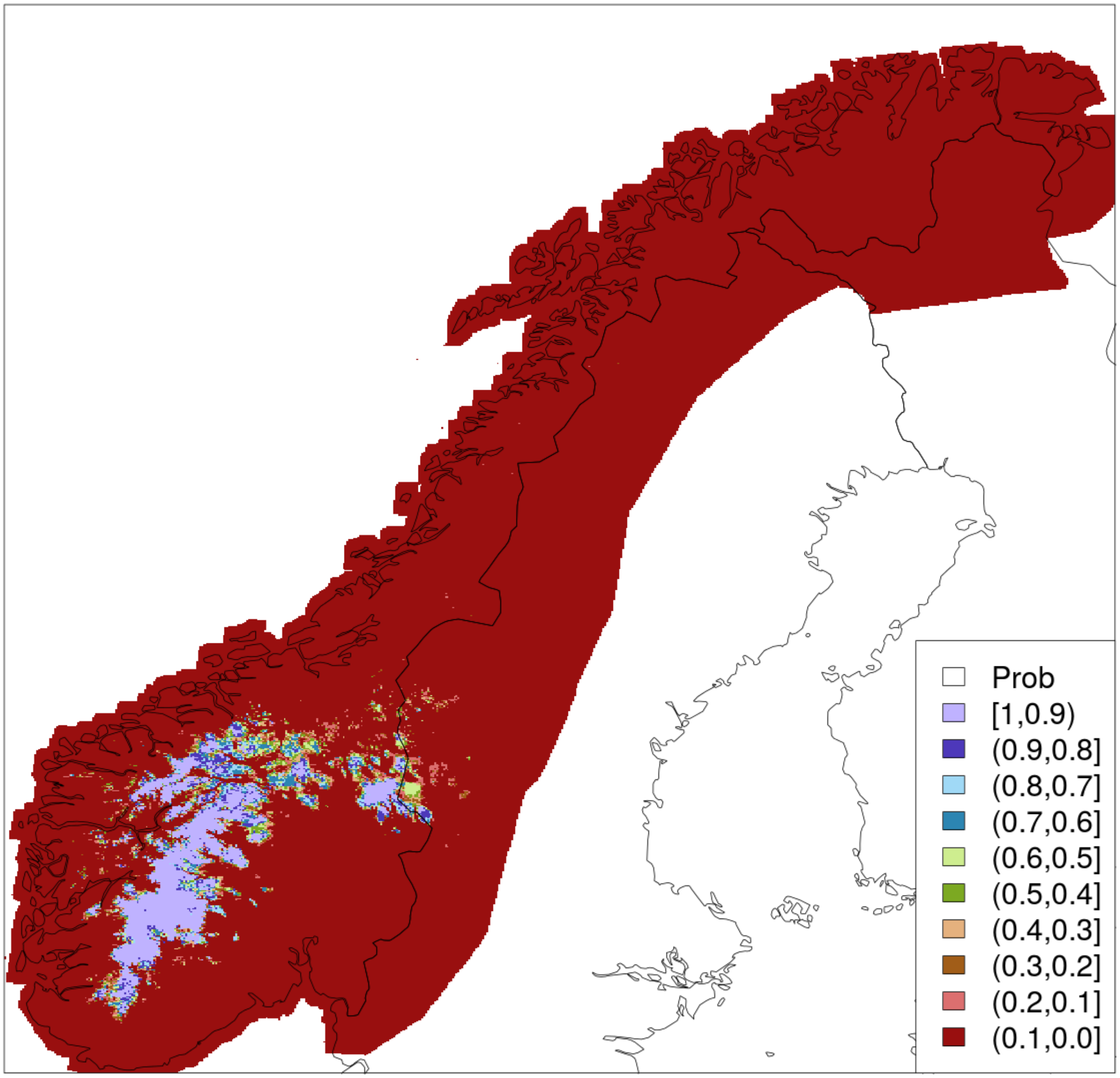


Figure 3.9: Probability of $T_{2m} < -20^\circ C$ at 00 UTC on 14th February 2016.

3.2.2 A 6-months period evaluation

Here we present an evaluation of the scheme performances over an extended period of 6 months, from February to July 2016.

The daily timeseries of the bias (top panel, fig. (3.10)) and the RMSE (bottom panel, same figure) continue to display a significant gain in the analysis performances with respect to the model only. Both the bias and the RMSE show a strong synoptic-scale weather dependence, with broad and periodic fluctuations which highlight the transitions of the synoptic systems. The model bias oscillates between $-0.6^\circ\, C < BIAS < +1.0^\circ\, C$, even if the range becames more narrow in the second half of the period, with values within $-0.3^\circ\, C < BIAS < +0.6^\circ\, C$. The analysis bias is confined between $-0.2^\circ\, C < BIAS < +0.5^\circ\, C$ across the entire period. The model RMSE lies in the interval $1.0^\circ\, C < RMSE < 2.8^\circ\, C$, whereas the analysis shows a general reduction in values up to $0.2^\circ\, C \div 0.3^\circ\, C$, fluctuating within $0.8^\circ\, C < RMSE < 2.6^\circ\, C$. Missing data in fig. (3.10) are due to the lacks in the observations already reported in chapter §2 (cfr. fig. (2.2)).

Moreover, a comparison over the cross-validated stations between the mean value of the RMSE affecting the model and that affecting the analysis mean is shown in Fig. (3.11). The decrease in the values of the RMSE is particularly evident in the eastern Norway, with a transition from $1.2^\circ\, C < RMSE < 1.7^\circ\, C$ (yellow and green circles) to $RMSE < 1.2^\circ\, C$ (green and blue circles).

The reliability diagram in fig. (3.12, always for the threshold $T_{2m} > 0^\circ\, C$) proves that the improvements of the analysis (red line) affect almost all the predicted probability intervals with respect to the model only (black line). Both the analysis and the model tend to underestimate the occurrence of the event for small predicted probabilities (both the red and the black line lies over the gray line of the perfect prediction), whereas they tend to overestimate the event for predicted probabilities close to 1. However, the analysis performs better than the model lying closer to the gray perfect score line. Furthermore, both the model and the analysis perform better than the climaotlogy (the horizontal dashed line in the upper part of the diagram).

In figs. (3.13)-(3.12) the probabilistic scores are shown. Fig. (3.13) displays both the daily (top panel) and the hourly mean (bottom panel) timeseries of the brier score for the threshold $T_{2m} > 0^\circ C$. Both suggest a relevant increase of the overall predicting skills in favour of the analysis: the daily analysis brier score (red line, top panel) lies below the model one (black line). Values drop down to zero in the second half of the period (i.e. from May), when the occurrences (and the predictions) of the temperature below zero become rare. The mean hourly timeseries (bottom panel) shows again a decrease in the analysis brier score with respect to the model: the reduction is more evident during nighttime ($\Delta BS \simeq 0.01$) and less during daytime ($\Delta BS \simeq 0.002$).

Both the reliability and the resolution components of the brier score for the same threshold are displayed in the following figs. (3.14) and (3.15)). The reliability component (both figures, top panels) shows a significant reduction in favor of the analysis (red line), with daily values constantly below $BS_{REL} < 0.01$ during the whole period, whereas the model (black line) shows several peaks up to values bigger than $BS_{REL} > 0.02$. In the hourly mean timeseries (fig. (3.14)), the analysis (red line) is almost half of the model (black line) and both lies below $BS_{REL} < 0.02$. The resolution component (both figures, bottom panels) shows a small increase in favor of the analysis (red line), with values larger than those of the model (black line) by $\Delta BS_{RES} \simeq 0.02 \div 0.03$ across almost the entire period. On average, in the mean hourly timeseries, this is more evident during nighttime (by $\Delta BS_{RES} \simeq 0.01$).

Lastly, the pit histrograms in figs. (3.16) and (3.17) indicate an improvement of the quality in the analysis ensemble with respect to the model one, in terms of both ensemble spread and members equiprobability. Even if in fig. (3.16) both the model and the analysis ensemble seem to show a strong underdispersion (a lot of observations fall outside the ensemble members), this condition seems less pronounced for the analysis ensemble (right panel). However, since we are validating a set of observations against a probability density distribution, we re-computed the pit histrogram also after having perturbed the ensemble with a random perturbation, in order to take into account the observation uncertainty, as suggested in [Hamill, 2001]. The two new histrograms in figure (3.17) show less underdispersion and still a better analysis ensemble with respect to the model one. Another possible way to evaluate the quality of the analysis ensemble is comparing the RMSE of the analysis mean with the ensemble spread. In fig (3.18) the daily timeseries of both is plotted. The analysis ensemble spread (green line) lies below the RMSE (yellow line); the former remains between $0.3^{\circ}\, C < ENS_{spread} < 0.8^{\circ}\, C$, whereas the latter oscillates within $0.8^{\circ}\, C < RMSE < 2.6^{\circ}\, C$. This could mean that the spread of the final analysis ensemble is still too narrow to be able to include the observed value among the possible outcomes expressed by the ensemble members. However, it has to be pointed out that, in this case, the correction proposed by [Hamill, 2001] has not been applied.

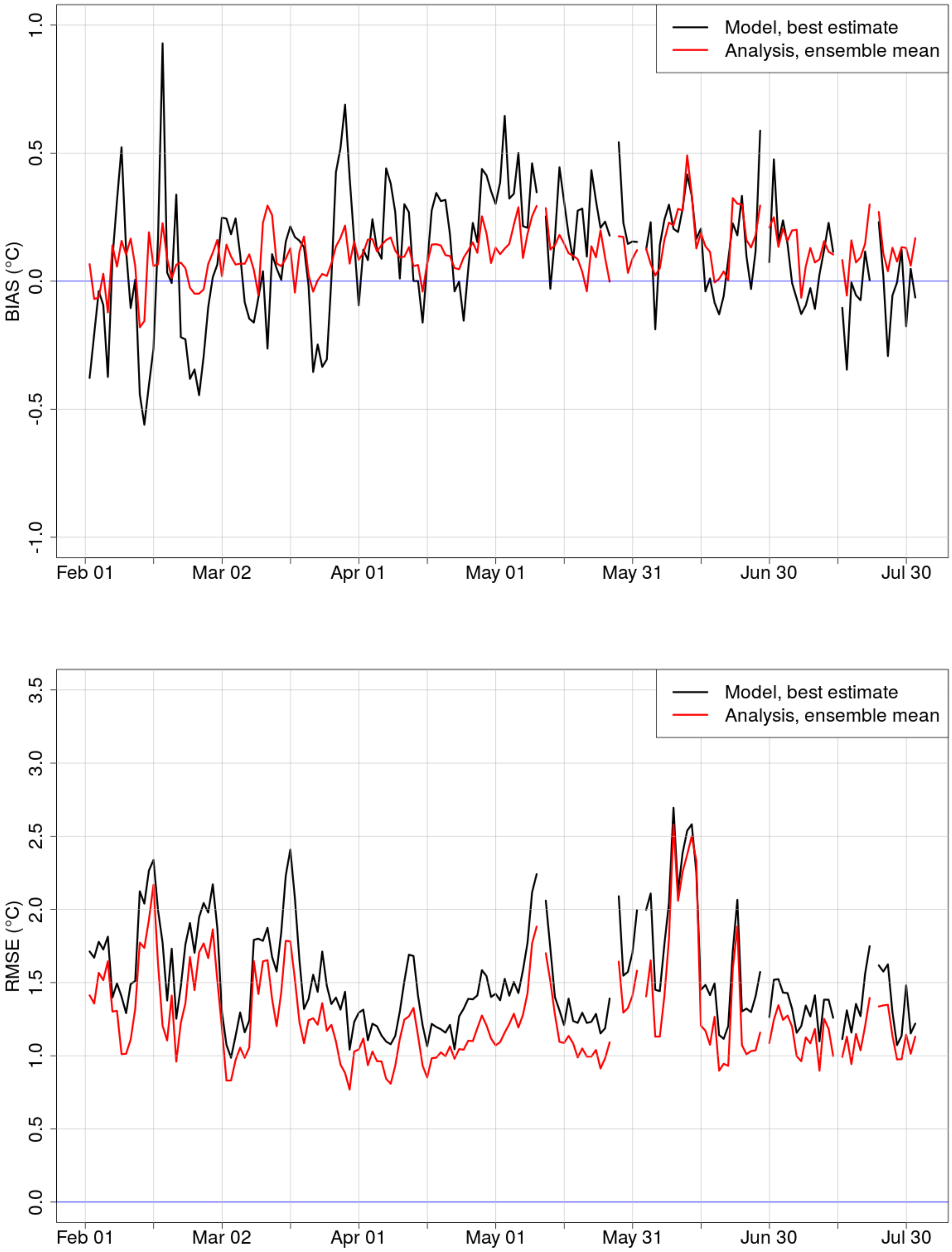


Figure 3.10: Bias (top panel) and root mean squared error (bottom panel) affecting the model (black line) and the analysis (red line) during the 6-months period. The blue line marks the perfect score.

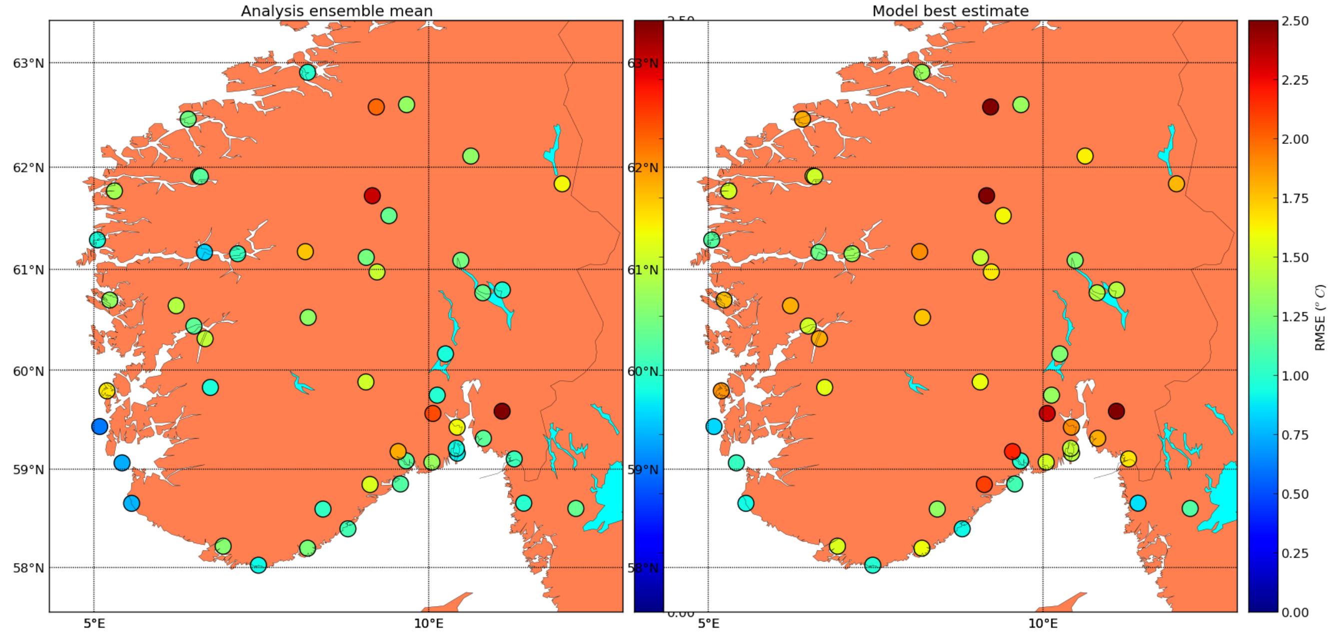


Figure 3.11: Analysis (left panel) and model (right panel) root mean squared error averaged over the 6-months period for each of the cross-validated stations in the southern Norway.

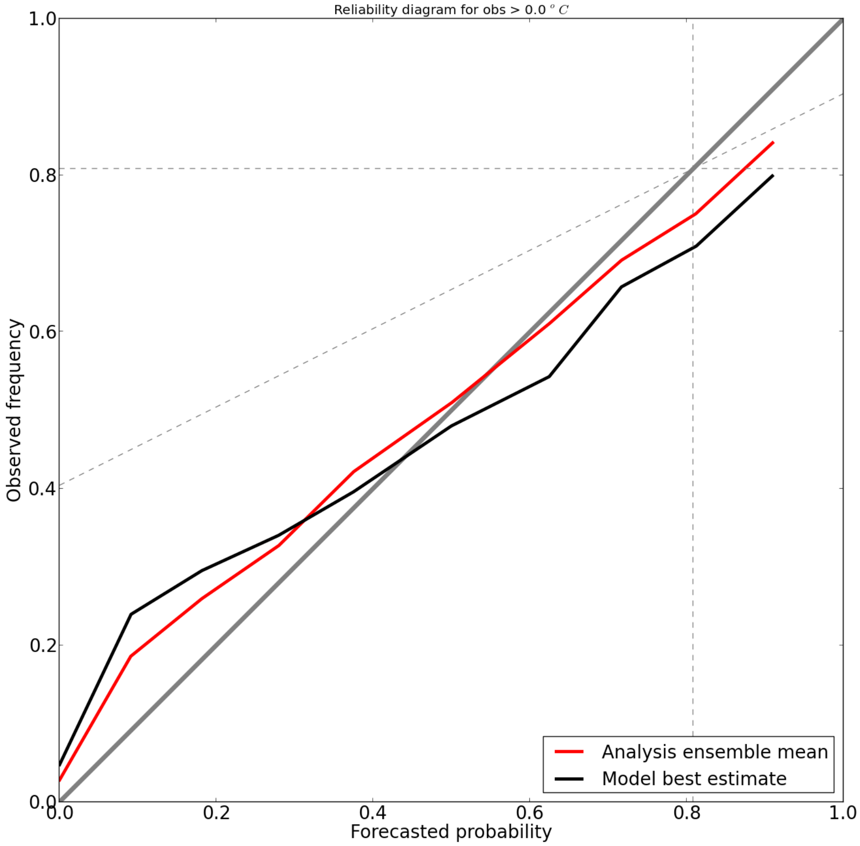


Figure 3.12: Reliability diagram comparing the analysis and the model ensemble computed for the threshold $T_{2m} > 0° \ C$. The solid gray line indicates the ideal prediction. Dashed horizontal line represents the climatology.

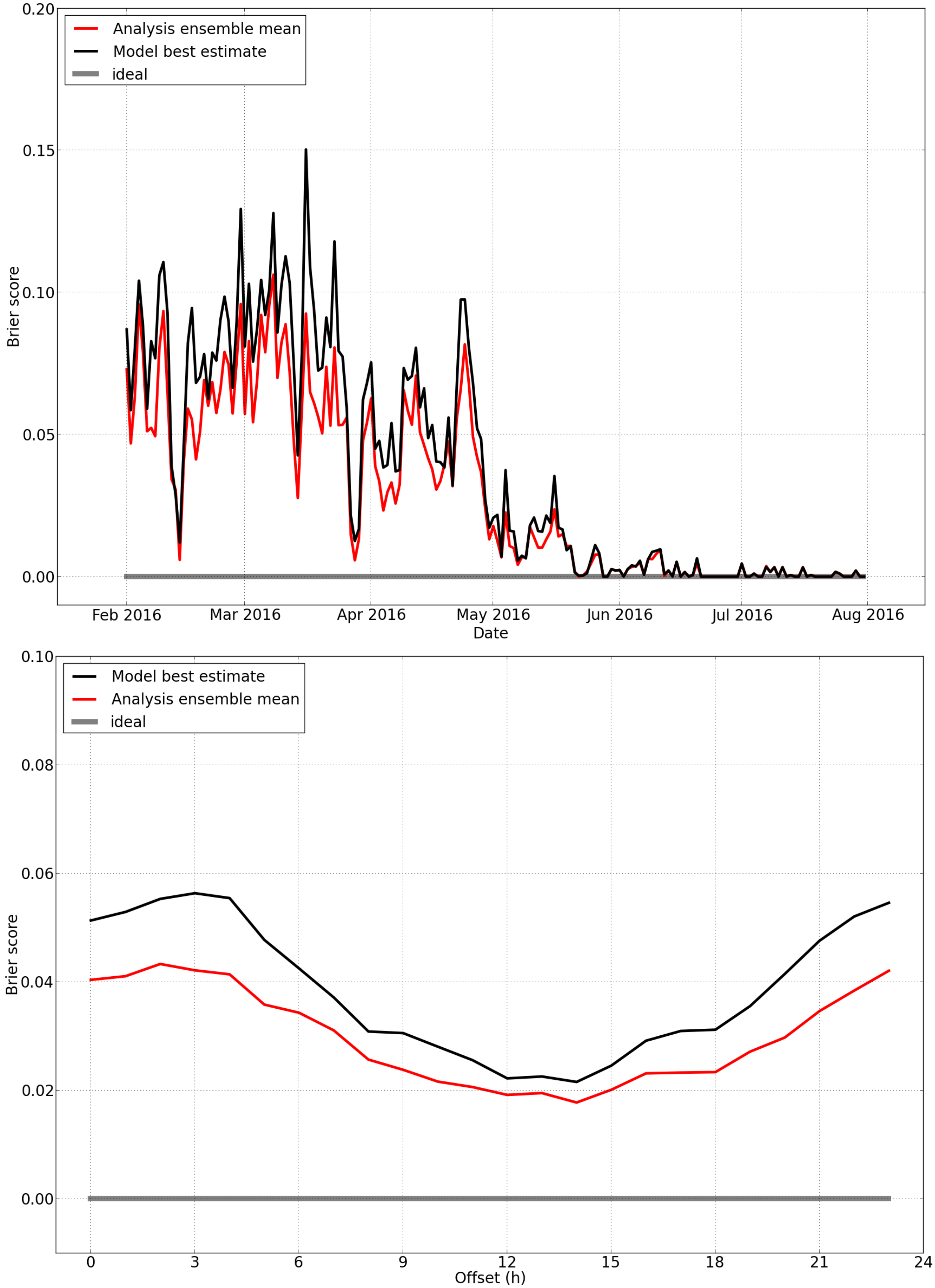


Figure 3.13: Brier score computed for the threshold $T_{2m} > 0^{\circ} C$ during the 6-months period. Daily timeseries (top panel) and mean hourly values (bottom panel).

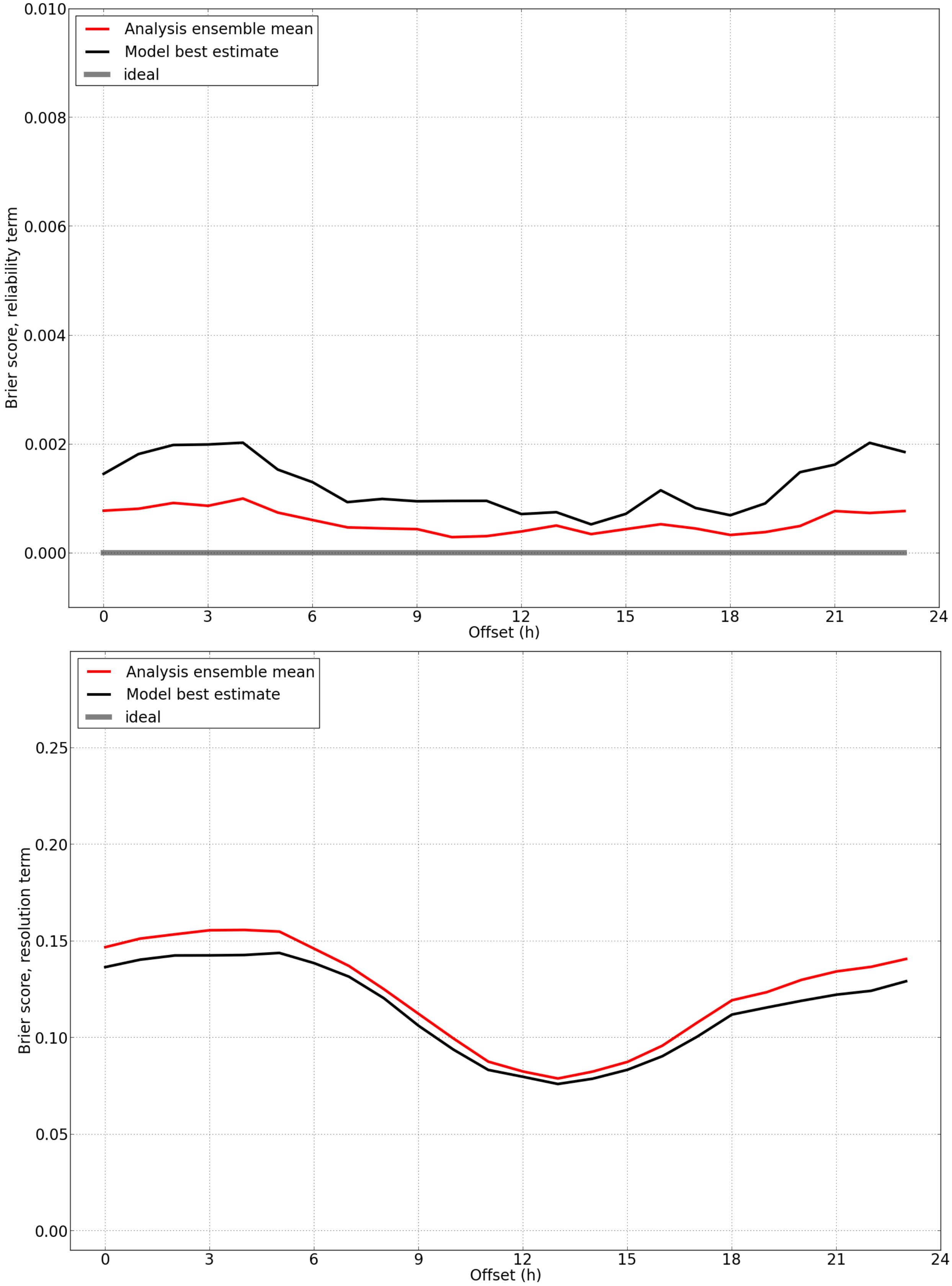


Figure 3.14: Mean hourly values of both the reliability (top panel) and the resolution (bottom panel) component of the Brier Score computed for the threshold $T_{2m} > 0^\circ C$ during the 6-months period. Ideal value not present in the resolution term.

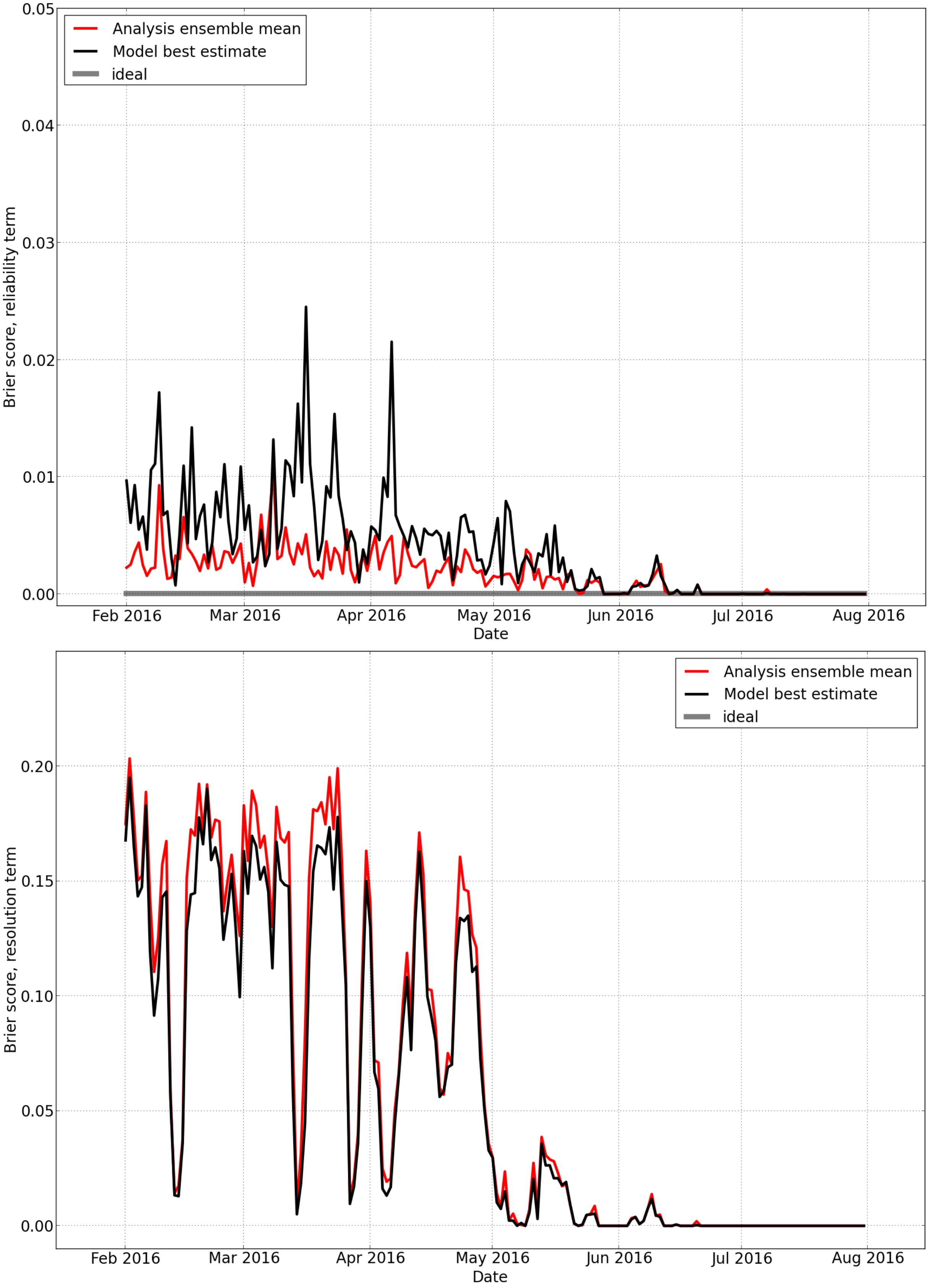


Figure 3.15: Daily timeseries of both the reliability (top panel) and the resolution (bottom panel) component of the Brier Score computed for the threshold $T_{2m} > 0^{\circ}\,C$ during the 6-months period. Ideal value not present in the resolution term.

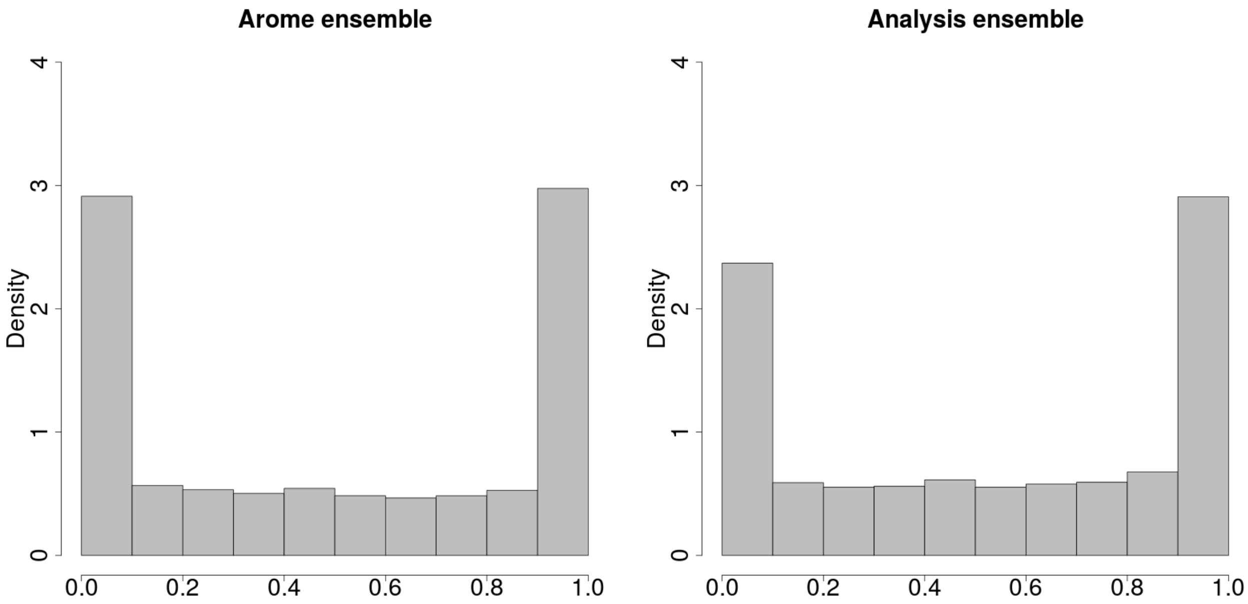


Figure 3.16: Pit histrograms resulting by the validation of the observations against the model (left) and the analysis (right) ensemble.

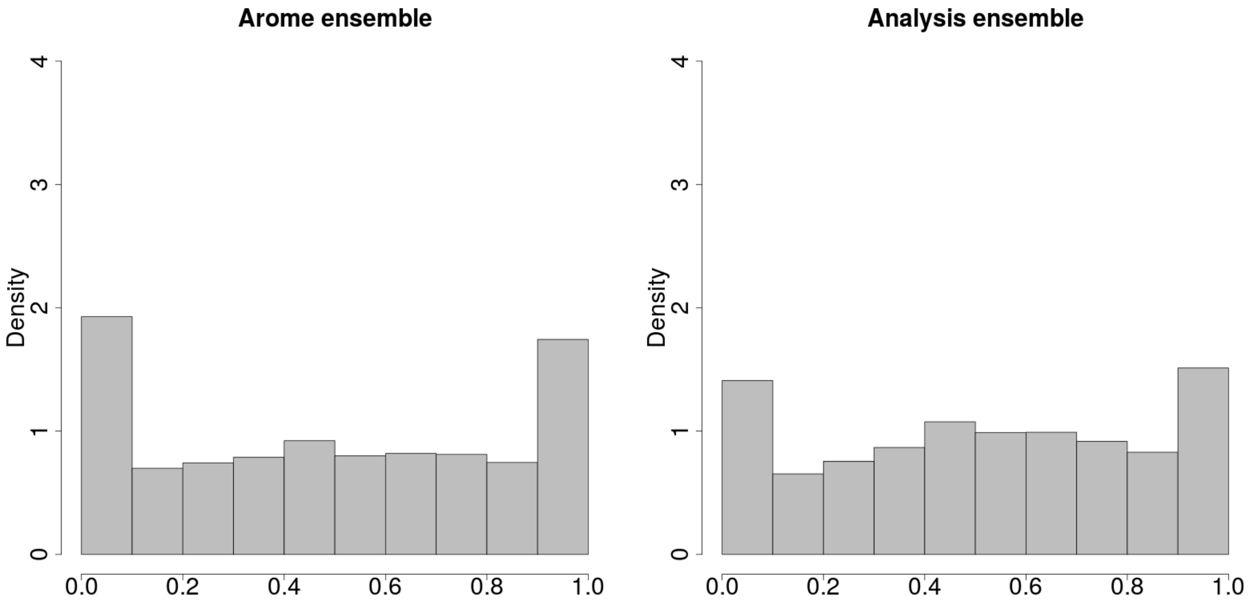


Figure 3.17: Pit histrograms resulting by the validation of the observations against the model (left) and the analysis (right) ensemble. The observation uncertainty is taken into account as in [Hamill, 2001].

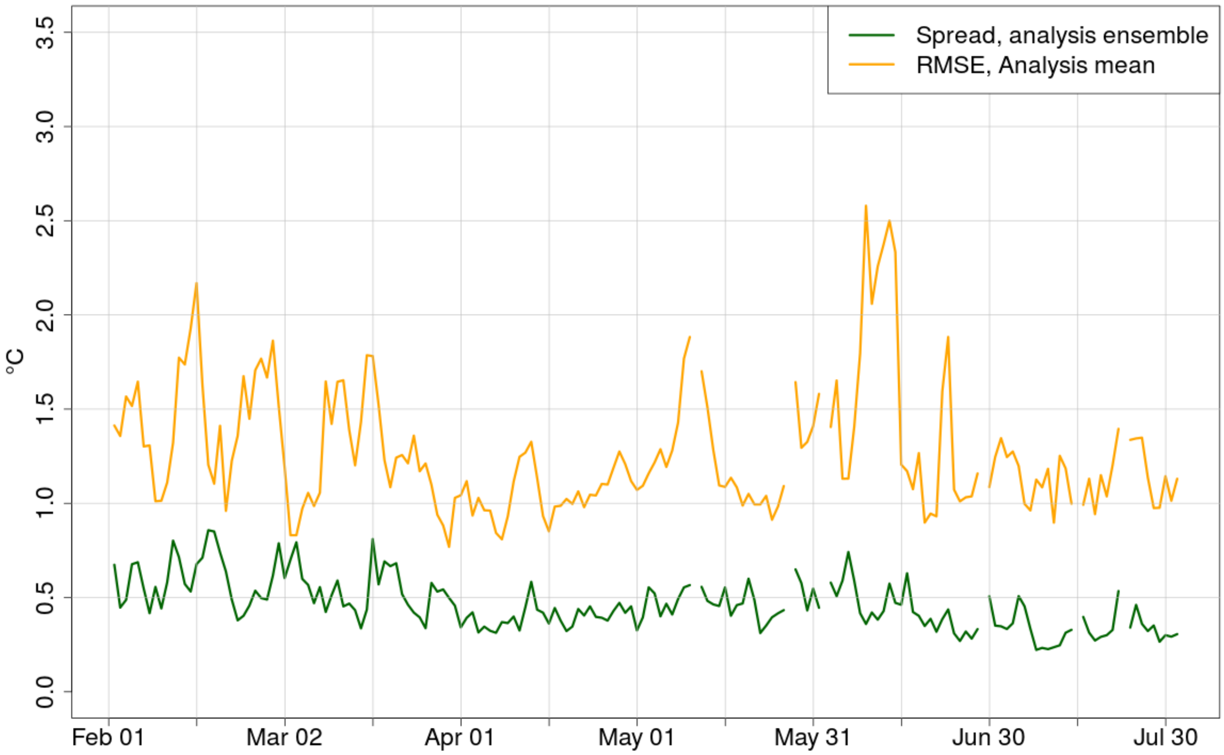


Figure 3.18: Comparison between the analysis ensemble spread (green line) and the analysis RMSE (yellow line) during the 6-months period.

The observation variance and the covariance inflation

The observation variance and the covariance inflation are updated recursively at each time step within the scheme, through a scalar Kalman Filter and are strongly dependent on the parameters choice (cfr. §1.2.3 and §1.2.4). In figure (3.19) a hourly timeseries shows how these two quantities evolve during the considered period, given the parameters set in section §2.4 and highlighted in Table 2.1. The covariance inflation parameter (purple line) shows the biggest fluctuations, with values whithin $2 < \Delta < 25$; several peaks and dips are present probably due to the transitions of the synoptic weather patterns, which in turns affects the flow-dependant predictability and then the ensemble spread (i.e. the matrix $\mathbf{P}^f$ which is employed to compute Δ). The observation variance (orange line) shows several synoptic-scale variations as well, but it fluctuates witihn a more narrow margin, i.e. $1 < \sigma_o^2 < 4$. Phases with higher values of Δ seem to be correlated with higher values of σ_o^2: this makes sense in vitue of the definition of Δ (cfr. eq. (1.28)).

Missing data are due to the before-mentioned lacks of observations; however, the scheme rapidly recovers towards the previous values.

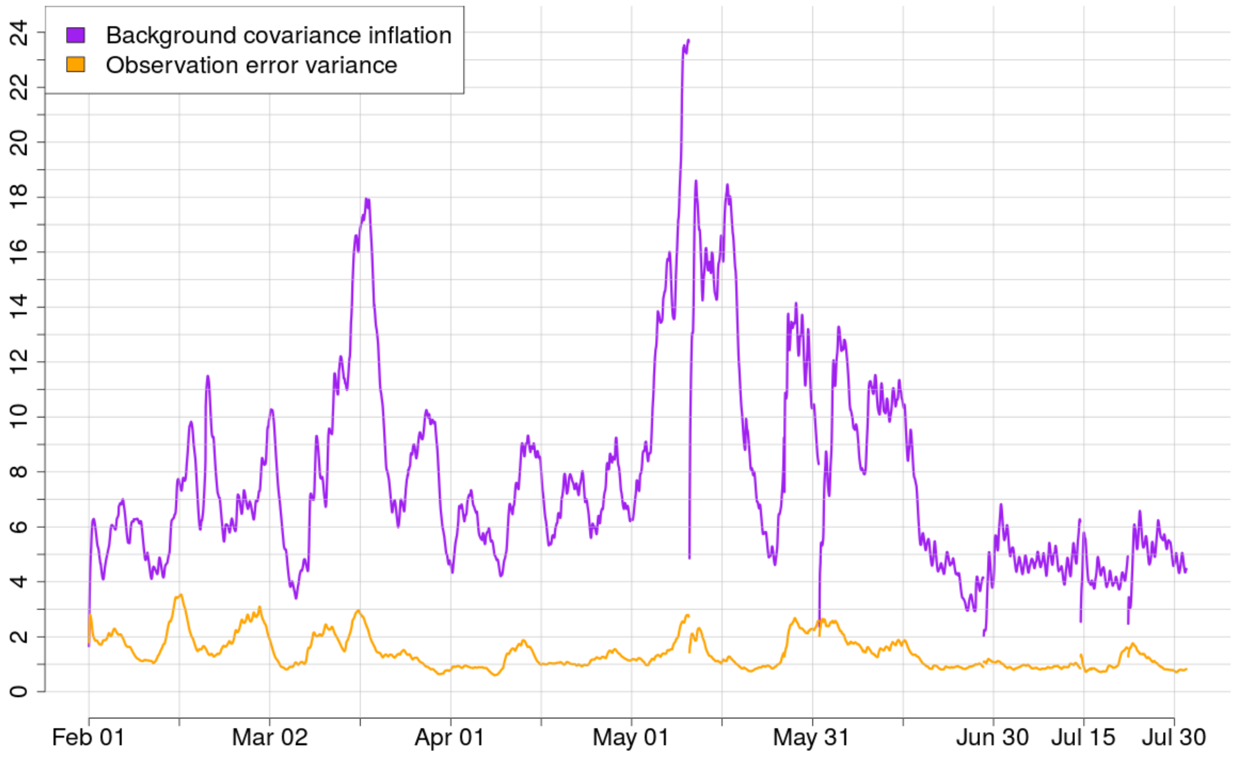


Figure 3.19: Daily timeseries of both the observation error variance (orange line) and the inflation coefficient (purple line) during the 6-months evaluation period.

3.3 Summary and discussion

In this chapter we dropped the main outcomes resulting from the scheme application, starting from a specific case study and then extending our evaluation to a long-term period. All the displayed results have confirmed the added value of our scheme in representing the 2-meter temperature with respect to the Arome model only. The results can be summarized as follows:

1. The analysis ensemble mean shows better accuracy and precision than the model best estimate, with smaller values of both bias and RMSE, during the case study as well as the 6-months period. The decrease in the RMSE values is more evident in the eastern Norway.

2. The analysis ensemble mean represents better the coldest extreme of temperature (i.e. $T_{2m} < -20^{\circ}\,C$) during the case study with respect to the model best estimate.

3. The analysis ensemble shows smaller values of Brier Score and of its reliability component with respect to the Arome ensemble for the threshold of zero celsius; it also displays an increase in the resolution term. Improvements affect daytime in particular.

4. The analysis ensemble shows better predicting skills than the model also in the reliability diagram, with a reduced tendency to underestimate observations at small predicted probabilities and to overestimate observations at larger ones.
5. The analysis ensemble spread is still too narrow and underdispersed, even if it is wider than the model one.
6. The inflation parameter Δ and the observation variance σ_o reacts promptly to the changes of the synoptic weather patterns, with fluctuations over time-scales of the order of some days or weeks.

All these results indicate that the network of observations actually contributes to add information, and helps with taking into account those properties of the surface which are not resolved by the model but still impact on the 2-meter temperature field over Norway. The showed improvement in reproducing the cold extremes is particularly important and promising, especially because such events can damage and put at risks human activities and infrastructures (e.g. agriculture, pipes, roads, etc...); the possibility of converting the final analysis ensemble in a probabilistic prediction (as we did in fig. (3.9)) can represent a very useful and concrete application.

The narrowness of the final analysis ensemble spread is still a weak spot for the scheme, but a real model ensemble should be tested in order to understand if part of the problem could lie in the choice we made in section §1.2.2.

The promptness in the change of the inflation parameter Δ and the observation variance σ_o justifies the choice of providing a recursive (or flow dependent) adjustment for them.

Conclusions and outlook

In this thesis, a statistical post-processing of the Arome MetCoOp 2.5 km 2-meter temperature fields based on a Local Ensemble Transform Kalman Filter scheme has been developed and tested. The research has been carried out at the Norwegian Meteorological Institute (MET Norway).

After a detailed description of the scheme in the first chapter (§1), its implementation at MET Norway and the diagnostics have been discussed in the second one (§2). Firstly, a description of the two data sources employed in the scheme was reported (§2.1). Secondly, a study of the sensitivity of the observation operator $\boldsymbol{H}(\cdot)$ led us to choose $R = 10\,km$ as radius of the area employed to compute the near-surface temperature gradient to adjust the value for the elevation difference between the station and the closest grid point (§2.2). Then, we showed the form of the background error covariance matrix $\mathbf{P}^f$, which influences how the information is transfered across the grid once the observations have been assimilated; as we saw there, its form is strongly dependent on the orography and it is affected by a day-night cycle (§2.3). After that, we illustrated how we tuned the horizontal (D_h) and the vertical (D_z) localisation distances as well as the bias magnitude and reactivity controller γ, all of them to be set before the scheme application (§2.4). The scheme proved to be very robust with respect to the parameters changes, showing a very mild response as γ, D_h and D_z were varied. Nevertheless, we reached a satisfactory level of stability for $\gamma = 0.25$, $D_h = 50\,km$ and $D_z = 200\,m$.

In the third chapter we displayed and commented the results of the evaluation performed during both a case study in February 2016 and a long-term period of 6 months. During the case study (§3.2.1), results showed an overall improvement in the representation of the 2-meter temperature in favor of the analysis, which proved to be more accurate and more precise than the model alone, and better at reproducing the extremes during the cold winter phase; we also proved the usefulness of an analysis ensemble in computing probabilistic forecast. The improved accuracy and precision of the scheme during the case study are confirmed in the 6-months assessment (§3.2.2) together with the improvements in the probabilistic scores like the Brier Score and its components (reliability and resolution) as well as the reliability diagram. Rather, the evaluation of the ensemble quality, performed through the PIT histrogram and the comparison between the ensemble spread and the RMSE, showed some weak spots in the methodology: the analysis ensemble displayed a certain degree of overconfidence and its spread was too narrow. The 6-months timeseries of the covariance inflation parameter Δ and the observation variance σ_o - allowed to be time-depending by means of a scalar Kalman Filter - showed a good reactivity in adapting to the synoptic weather transitions.

The method developed in this work is robust and very promising, and several next steps are worth being considered for future implementations. Firstly, a real ensemble of model field should be used, instead of the one recreated in this work through the NMC method (cfr. §1.2.2): in this regard, an operational ensemble model have been employing at MET Norway since November 2016. Furthermore, the spatial representation of other meteorological variables could benefit from this type of post-processing procedure, and precipitations (for which also radar data are available) might be the next one. If the progresses continue in the future, MET Norway will adapt this scheme for the operational implementation.

Appendix

A.1 Matrix derivative

In this appendix we fix some important rules to perform the derivatives with respect to a matrix.

If $\mathbf{K}$ is a matrix of dimension $m \times n$, $\mathbf{x}$, $\mathbf{x}_1$ and $\mathbf{x}_2$ are vectors of length m and $\mathbf{y}$, $\mathbf{y}_1$ and $\mathbf{y}_2$ are vectors of length n, the derivative of a linear scalar with respect to $\mathbf{K}$ is a constant matrix of dimesion $m \times n$:

$$\frac{\partial}{\partial \mathbf{K}} \left(\mathbf{x}^T \mathbf{K} \mathbf{y} \right) = \mathbf{x} \mathbf{y}^T,$$

or:

$$\frac{\partial}{\partial \mathbf{K}} \left(\mathbf{y}^T \mathbf{K}^T \mathbf{x} \right) = \mathbf{x} \mathbf{y}^T$$

because $\mathbf{x}^T \mathbf{K} \mathbf{y} = \mathbf{y}^T \mathbf{K}^T \mathbf{x}$, being $\mathbf{x}^T \mathbf{K} \mathbf{y}$ a scalar.

Furthermore, the derivative of a quadratic scalar in $\mathbf{K}$ is a matrix of dimesion $m \times n$ linearly dependent on $\mathbf{K}$:

$$\frac{\partial}{\partial \mathbf{K}} \left(\mathbf{y}_1^T \mathbf{K}^T \mathbf{K} \mathbf{y}_2 \right) = 2 \mathbf{K} \mathbf{y}_1 \mathbf{y}_2^T,$$

and:

$$\frac{\partial}{\partial \mathbf{K}} \left(\mathbf{x}_1^T \mathbf{K} \mathbf{K}^T \mathbf{x}_2 \right) = 2 \mathbf{x}_1 \mathbf{x}_2^T \mathbf{K}.$$

We can demostrate at least the first one:

$$
\begin{aligned}
\left[\frac{\partial}{\partial \mathbf{K}}\left(\mathbf{x}^T\mathbf{K}\mathbf{y}\right)\right]_{i,m} &= \frac{\partial}{\partial k_{i,m}}\sum_{i',m'} x_{i'}k_{i',m'}y_{m'} \\
&= \sum_{i',m'}\frac{\partial}{\partial k_{i,m}}\left(x_{i'}k_{i',m'}y_{m'}\right) \\
&= \sum_{i',m'} x_{i'}\frac{\partial k_{i',m'}}{\partial k_{i,m}}y_{m'} \\
&= \sum_{i',m'} x_{i'}\delta_{i,i'}\delta_{m,m'}y_{m'} \\
&= x_i y_m \\
&= \left[\mathbf{x}\mathbf{y}^T\right]_{i,m}
\end{aligned}
$$

A.2 The $\mathbf{c_i}$ coefficients

In this appendix we illustrate how we compute the initial value of the observation variance $(\sigma^2)_{t_i}$ and the c_i coefficients employed in the matrix $\mathbf{R}$ (§1.2.3). Both of them are based on the 2-meter temperature data in the seNorge2 database [Lussana et al., 2016], taking into consideration all the measurements in a period of time T formed by the three hours centered on the analysis time t_i in the 45 days before and after with respect to that instant

The initial value $(\sigma_o^2)_{t_0}$ was then computed as:

$$
(\sigma_o^2)_{t_0} = \frac{1}{T}\sum_{t=1}^{T}\frac{(y^o - y^a)_t^T(y^o - y^f)_t}{s_{t_0}} = \frac{1}{T}\sum_{t=1}^{T}(\sigma_o^2)_t.
$$

The observation variance σ_o^2 computed in eq. (1.26) is a scalar averaged over all the observations available at that time istant t_i. However, the representativeness error can differ a lot from one station to another, so we need a coefficient to link the overall variance to that of a specified station. Then the station-dependent coefficients $(c_i)_{t_i}$are computed as:

$$
(c_i)_{t_i} = \left\{\tfrac{1}{T}\textstyle\sum_T\left[(y_i^o - y_i^a)_t(y_i^o - y_i^f)_t\right]\right\}/(\sigma^2)_{t_0}.
$$

A.3 Forecast verification scores

In this appendix the definitions of the scores and of the other forecast verification tools employed in the chapter §3 are listed. They are all defined for a generic set of $k = 1, ..., n$ couples of predicted (y_k^f) and observed values (y_k^o). o_k indicates a binary variable taking the value 1 or 0 according to the occurence of an event or not, whereas p_k is its predicted probability of occurring. A more extended disquisition can be found in [Wilks, 2011].

Bias The bias is defined as the mean sistematic error affecting a forecast, i.e. its deviation from the observed value:

$$BIAS = \sum_{k=1}^{n} \left(y_k^f - y_k^o \right).$$

The bias is a good indicator of the accuracy of a forecasting system. It is usually applied to evaluate deterministic (nonprobabilistc) forecasts.

Root mean squared error (RMSE) The root mean squared error is defined as the squared root of the mean squared error, i.e. the mean of the squared deviation of the predicted value from the observed one:

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(y_k^f - y_k^o \right)^2}.$$

The root mean squared error is a measure of the precision of a forecasting system. It is usually applied to evaluate deterministic (nonprobabilistc) forecasts.

Brier Score The Brier Score provides an overall indication of the quality of a probabilistic forecast:

$$BS = \frac{1}{n} \sum_{k=1}^{n} \left(p_k - o_k \right)^2.$$

The Brier Score is employed to evaluate, on average, at what extent the forecasted probability of occurrence of an event actually matches the frequency of occurrence of that same event. It can take any value between 0 and 1: a perfect forecasting system will have $BS = 0$, since the predicted probability and the occurrence have the same value.

The Brier Score can be algebraically decomposed in three terms, splitting the forecasted probabilities in $i = 1, ..., I$ sub-intervals (e.g. $p_i = 0$, $p_i = 0.1$, $p_i = 0.2$, and so on), each having N_i couples of p_i and o:

$$BS = \underbrace{\frac{1}{n} \sum_{i=1}^{I} N_i \left(p_i - \bar{o}_i \right)^2}_{Reliability} - \underbrace{\frac{1}{n} \sum_{i=1}^{I} \left(\bar{o}_i - \bar{o} \right)^2}_{Resolution} + \underbrace{\bar{o} \left(1 - \bar{o} \right)}_{Uncertainty},$$

where $\bar{o}_i$ is the mean value of o in each of the $i = 1, ..., I$ intervals and $\bar{o}$ the mean value of o over all the n occurrences.

The reliability term provides an indication of how much, on average, the forecasted proability is calibrated, that is close to the actual frequency of occurrence, in each of the I sub-intervals.

In particular, the term is weighted by N_i, the number of cases in each sub-interval i. The more precise is a forecast, the smaller this term should be.

The resolution term evaluates at what extent, on average, the forecasting system is able to provide forecasts which takes into account the peculiarity of a weather situation, instead of following the climatology. The predicted probability p_i is only implicitly included in this term, through the variable $\bar{o}_i$. The better is a forecast in discerning among different forecasts for different situations, the larger the difference between the frequency of occurrence in each of the sub-interval $\bar{o}_i$ and the overall mean $\bar{o}$ and the bigger this term.

The uncertainty term does not depend on the forecasts but only on the observations. It quantifies the variability in the climatology, i.e. it is maximum when $\bar{o} = \frac{1}{2}$ and the event has an overall equal probability of occurring or not in the considered set of data.

Reliability diagram The reliability diagram provides a graphical way to represent the quality of a forecast. Performing the same split of the the forecasted probabilities as in the Brier Score, the values of p_i and $\bar{o}_i$ are, in these case, plotted in a diagram with the probability lying along the horizontal axis and the observed frequencies along the vertical one. The reliability diagram shows Clearly, a perfect forecasting system will show a line exactly along the diagonal. Sometimes, the climatology is drawn as a horizontal line.

PIT histogram The Probability Integral Transfor histogram (or PIT histogram) is employed to evaluate the quality of an ensemble forecast, especially the spread and the equiprobability of its members. To build the histogram, for each couple of ensemble forecast and observation, a value between 0 and 1 is attributed to the observation according to its relative position within the ensemble members (for example, if the observation lies beyond the largest member, it will be given a 1; conversely, if it lies below the lowest member, it will be given a 0). The form of the PIT histrogram provides information about the quality of the ensemble forecast. A flat histogram represents the condition of “perfect forecast”.

Acknowledgements

I really want to thank Cristian Lussana whose help, advice and support have been very precious to me in the last seven months. At the same time, the suggestions and the tips of Michele Brunetti from Italy have played the role of a a very useful external eye on my ongoing work in Oslo. A special thank goes also to Thomas Nippen, whose forecast verification software VERIF has been useful and very time-saving during the analysis of the results. More generally, I have found a very welcoming and nice environment at the Meteorologisk Institutt and I am very happy and satisfied with the time spent there. I have really appreciated the company of Oskar and Øyvind who shared the office with me, and the friendly support of Ole Einar Tveito and of all the people working at the Climate Division.

During my staying in Norway, I have also had some good time with people outside work, starting from my flatmates came in succession: Fernando, Dennis, Neringa and Jayson. Moreover, if my norwegian is now (still bad but) not as bad as the day I arrived there is thank to Maren and all the people met at the Language Café, with Rålf (who knows if the spelling is right...) at the top of the list.

Since this is also the last act of an entire cycle of study, I want to address a big final thank to all the people who supported me in this five-and-something years, whoever they are: family, relatives, friends.

Bibliography

[Bauer et al., 2015] Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55.

[Brunetti et al., 2012] Brunetti, M., Lentini, G., Maugeri, M., Nanni, T., Simolo, C., and Spinoni, J. (2012). Projecting north eastern italy temperature and precipitation secular records onto a high-resolution grid. *Physics and Chemistry of the Earth, Parts A/B/C*, 40:9–22.

[Brunetti et al., 2014] 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.

[Daley, 1993] Daley, R. (1993). *Atmospheric data analysis*. Number 2. Cambridge university press.

[Dee, 2003] Dee, D. P. (2003). Detection and correction of model bias during data assimilation. *Meteorological Training Course Lecture Series (ECMWF)*.

[Dee and Da Silva, 1998] Dee, D. P. and Da Silva, A. M. (1998). Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124(545):269–295.

[Dee and Todling, 2000] Dee, D. P. and Todling, R. (2000). Data assimilation in the presence of forecast bias: The geos moisture analysis. *Monthly Weather Review*, 128(9):3268–3282.

[Desroziers et al., 2005] 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.

[Fisher, 2003] Fisher, M. (2003). Background error covariance modelling. In *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, pages 45–63.

[Greybush et al., 2011] 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.

[Hamill, 2001] Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129(3):550–560.

[Harlim and Hunt, 2005] Harlim, J. and Hunt, B. R. (2005). Local ensemble transform kalman filter: An efficient scheme for assimilating atmospheric data. *preprint.*

[Hunt et al., 2007] 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 et al., 1997] Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C. (1997). Unified notation for data assimilation: Operational, sequential and variational (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):181–189.

[Kalnay, 2003] Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability.* Cambridge university press.

[Li et al., 2009] 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, 1986] Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194.

[Lussana et al., 2016] Lussana, C., Tveito, O. E., and Uboldi, F. (2016). An observational gridded dataset of temperature for norway.

[Lussana et al., 2010] 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.

[Müller et al., 2017] 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. (2017). Arome-metcoop: A nordic convective scale operational weather prediction model. *Weather and Forecasting*, (2017).

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

[Parrish and Derber, 1992] Parrish, D. F. and Derber, J. C. (1992). The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747–1763.

[Schiemann et al., 2010] Schiemann, R., Liniger, M., and Frei, C. (2010). Reduced space optimal interpolation of daily rain gauge precipitation in switzerland. *Journal of Geophysical Research: Atmospheres*, 115(D14).

[Uboldi et al., 2008] 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.

[Wilks, 2011] Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*, volume 100. Academic press.