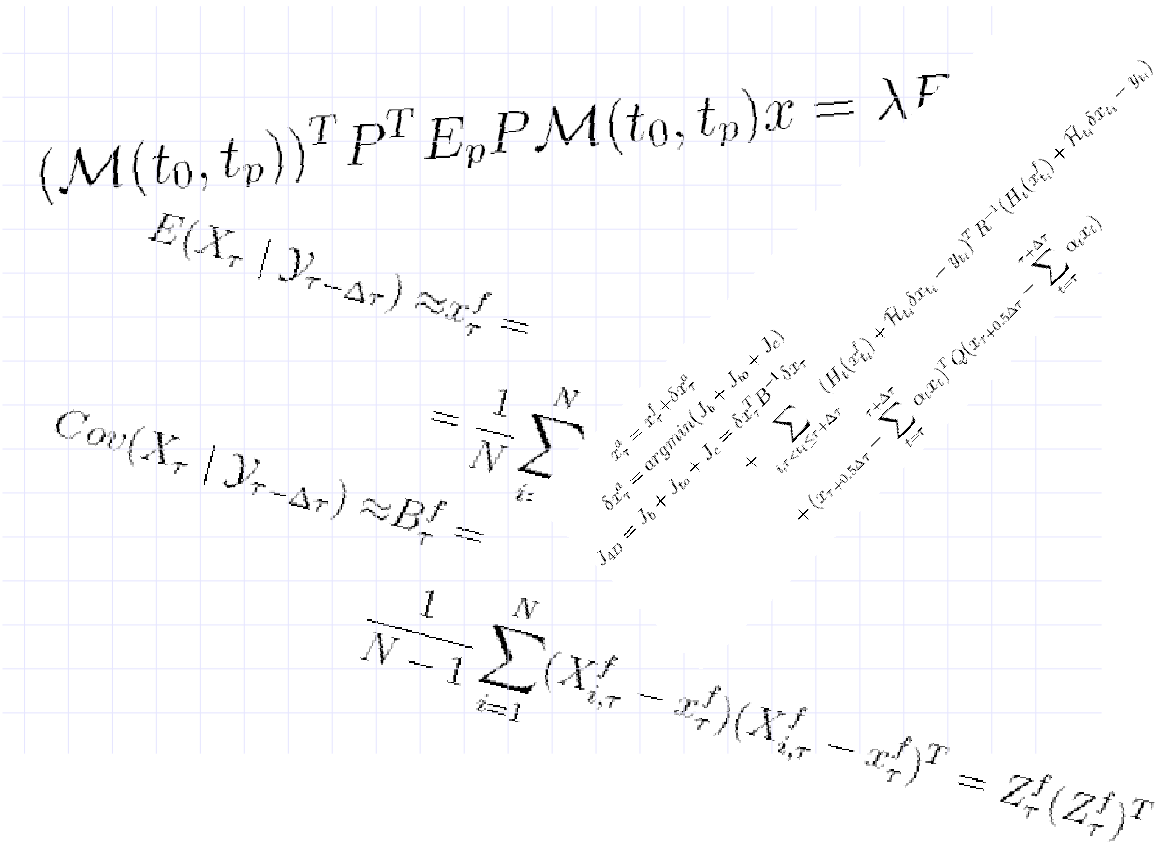
Norwegian Meteorological Institute
met.no

met.no report

no. 2/2010
Remote Sensing

Literature study on the Ensemble Kalman Filters, Variational data assimilation and Ensemble prediction systems:
theoretical relationships and practical considerations

Jelena Bojarova



Norwe
Meteorological Institute
met.no

report

Title	Date
Literature study on the Ensemble Kalman Filters, Variational data assimilation and Ensemble prediction systems: theoretical relationships and practical considerations.	14 January 2010
Section Remote Sensing	**Report no.** no. 2/2010
Author(s) Jelena Bojarova	**Classification** ● Free ○ Restricted
	ISSN 1503-8025
Client(s) RCN (eVITA program)	**Client's reference** eVITA EnKF project, contract 178894

Abstract

We review recent developments in the ensemble methods used in the meteorological data assimilation and prediction systems. The probabilistic formulation of the sequential data assimilation and prediction problem is discussed in the case of the linear state space models. The analytical solution to the problem is possible only under restrictive conditions which are never met in practice. A systematized overview of literature on the computationally feasible approximate solutions to the sequential data assimilation problem, both of the variational (3D-Var, 4D-Var) and the regression (Ensemble Kalman Filter: Stochastic, Square-root and Reduced Rank) type based approach, is presented in this report in connection with non-linear state space models. Some attention is devoted to the non-parametric inference about the meteorological model state based on the Particle Filter approach. A hybrid Ensemble-Variation data assimilation when the rank-deficient flow-dependent ensemble estimate is merged with the full rank constant in time analytically deduced estimate of the forecast error covariance matrix gives promising results. The second topic of this report concerns strategies implemented at different weather services to construct the initial states for ensemble prediction systems with the aim to quantify the uncertainty of the prediction of the events of interest. We discuss both ensemble generation and verification methodologies.

Keywords

Sequential data assimilation, Ensemble Kalmal filter, 3D-Var, 4D-Var, Particle Filter, EPS

Disiplinary signature	Responsible signature
L-A B	Ø H

1. Introduction

The aim of the meteorological data assimilation is to determine an initial data field for numerical weather prediction and to quantify its uncertainty. In general, this task coincides with the problem of a model state observer for the non-linear system, governing the development of the atmosphere. However, several specific features of the meteorological data assimilation makes the task quite challenging.

First of all, the number of observations is of several order of magnitude smaller than the dimension of the model state variable. Therefore, prior information about the model state must be involved in the construction of the model state observer (usually called the analysis state) and must be carefully specified in order to overcome the lack of observations.

The information is usually introduced in the form of the standard first and second order moments: the prior mean of the model state, specified through a short-range forecast and often called a background state or a first-guess field, and the prior variance-covariance matrix of the model state, which contains statistical knowledge about the forecast errors and the physical balances between the different components of the model state variable. Taking the complexity of the physical processes, governing the development of the atmosphere, and the huge dimensionality of the model state $\mathcal{O}(10^6)$ into account, the careful specification of the prior information is a challenging task.

Secondly, the model state variable is just a discrete approximation to the continuous atmospheric state which is observed. The time evolution of the model state is governed by discrete approximations of continuous physical laws. The appropriate discrete approximation of balances as well as the variances of short-range forecast errors strongly depend on the spatial and the temporal scales of motion. So, the observed information must be assimilated taking into account the scale of the phenomena of interest and the model and representativity errors which are caused by the discretisation.

Besides that, an efficient assimilation scheme must be robust to the non-linear dependencies between observed quantities and the model state and must reflect both the spatial and temporal variation in the distribution of observations.

One method to merge a background field and observed quantities in a way consistent with the estimated accuracy of both sources of information was introduced first by Eliassen (1954) and independently by Gandin (1963). Within the meteorological community this method is called the Optimum Interpolation (OI). Under this method the analysed state is constructed as an optimal, in the sense of minimum variance, linear combination of the background state and observed quantities and is a linear regression of the model state on innovations (the deviations of the observed data from the background state projected into observation space). The method was extended to the three-dimensional multivariate analysis (Lorenc, 1982) and was for a long period successfully used for operational weather prediction by many weather services. Being a linear regression technique, the method does not require any strong assumptions on the probabilistic distribution of the model state variable, besides existence of second moments. Main disadvantage of the method is an ability to treat in a proper way only observed quantities linearly related to the model state.

An important step forward in the numerical weather prediction was the development of three-dimensional variational data assimilation scheme (3D-Var)(Parrish and Derber,1992) and its extension to four-dimensional (4D-Var: 3D-Var plus a time-window) variational assimilation scheme (Le Dimet and Talagrand, 1987, Courtier et.al, 1994). Under these methods, the analysed state is determined through a posterior mode of the model state, given the observed quantities and with the background state as a prior. A Gaussian distributional assumption on the prior is essential for the performance of these methods. At the same time these methods provide an optimal solution even in case of a non-linear

observational operator (the operator which transforms the model state variable into observed quantities). The valid physical balance (in essence static) relationship between different model state variable components explicitly enters into the data assimilation procedure through the analytically deduced and statistically derived forecast error covariance matrix. A time-invariant variance-covariance matrix with a simplified structure of the forecast errors (spatial homogeneity and barotropic forecast error structure) reduces efficiency of the data assimilation procedure when three-dimensional variational data assimilation scheme is applied.

In four-dimensional variational data assimilation scheme the observational operator includes a dynamical propagation of the model state over the assimilation window, and in such a weight it provides an implicit evolution of the forecast error variance-covariance matrix during this time window. This implicit propagation of the static variance-covariance matrix by the model dynamics improves the assumed structure of the forecast errors and the sequential observations will be given more proper way in accordance with this. Still the same static variance-covariance matrix in the beginning of each assimilation window, which neglect all information about previously assimilated observations, degrades the assimilation procedure.

As it is well known, having linear dynamics and a linear observational operator under valid Gaussian assumptions on the prior distribution of the model state, the optimal sequential estimation of the model state is Kalman filter (Kalman, 1960, Kalman and Bucy, 1961). If these requirements are fulfiled, the Kalman filter will provide the same solution as the four-dimensional variational assimilation scheme provided that the variance-covariance matrix is properly specified in the beginning of data assimilation window. However, because the numerical equations propagating the development of the atmosphere are non-linear and the dimensionality of the model state variable is unfeasible huge, the Kalman filter recursions cannot be used for the practical implementation of the data assimilation procedure.

A number of generalisations and extensions of the Kalman filter idea has been proposed for the purposes of the meteorological data assimilation. A suboptimal Kalman filter, called the ensemble Kalman filter in a number of versions, is one of the most successful extension of the classical Kalman filter implemented in practice at many weather services. Ensemble Filter, when a number of forecasts are propagated in time and updated sequentially from observations, has been developed in the attempt to produce information about the probability distribution of the atmospheric state (Evensen, 1994; van Leeuwen and Evensen, 1996; Toth and Kalnay, 1993, 1997; Houterkamer and Mitchell, 1998). The fundamental problem of this approach is that the sample size of the practical ensemble is too small in order to directly produce meaningful statistics about the complete distribution of the model state given the observations.

Different assumptions and heuristic methods were tried to tackle this problem. This resulted in a large number of practical implementations of the Ensemble Filter. In common for all implementations is that the dynamical evolution of the probability distribution of the atmospheric state is carried out through propagation of the initial (or analysed) ensemble of the model states forward in time by model dynamics. In such a way a forecast ensemble is obtained. The way in which the forecast ensemble is updated from observations into an analysed (or initial state) ensemble differs for different implementations of the Ensemble filter. Ensemble Filters could be crudely divided into

• the *Resampling approach Ensemble filters* (Kim et.al, 2003; Leeuwen 2003; Anderson and Anderson, 1999), where the ensemble of analysed states is resampled from the ensemble of the forecast states, and

• the *Rescaling approach Ensemble Filters*, where the ensemble of forecast states is transformed into

an ensemble of the analysed states.

The *Ensemble Kalman Filter* (EnKF) belongs to the *Rescaling approach Ensemble Filters* and can be divided into

(1) Perturbed Observations Ensemble Filter (EnKF), where the ensemble of forecast states is stochastically updated during the assimilation step (Evensen, 1994; Houterkamer and Mitchell, 1998),
(2) Square-root Ensemble Filters (ETKF), where the ensemble of forecast state is deterministically updated during the assimilation step (Cohn et al, 1998; Whitaker and Hamil, 2001; Anderson 2001, Bishop et al, 2001; Ott et al. ,2004) and
(3) the Reduced-Rank Kalman Filters where emphasis is put on the computationally feasible propagation in time of the forecast error covariance matrix.

The *Hybrid Ensemble Kalman Filter-Variational assimilation scheme*, which utilizes the advantages from both the Variational (the full-rank forecast error covariance matrix) and Ensemble Assimilation Schemes (the flow-dependent uncertainty about the estimate of the model state), seems to be a promising data assimilation technique. The summary on various practical implementations of the data assimilation schemes, their advantages and simplifications, and theoretical relationships between them are main topics of this report. Data assimilation schemes can provide not only a deterministic estimate of the model state (the analysed state), but also quantify the uncertainty about the analysed state as well. In variational approaches the inverse of the analysed variance-covariance matrix (the Hessian), which determines the curvature of the conditional probability density function, given observations, in a vicinity of its maximum (the analysed state), can theoretically be obtained during the assimilation cycle. In the Ensemble Kalman Filter based approaches the uncertainty of the model state is represented through the ensemble estimate of the analysis error covariance matrix. A number of clever ensemble prediction systems (EPS) were developed and implemented at different weather services in attempt to provide a probabilistic inference about some phenomena of interest.

EPS systems differ by strategies to generate the initial ensemble of perturbations and can crudely be classified into four different classes:

(1) error breeding (Toth and Kalnay, 1993,1997),
(2) singular vectors optimized over a certain forecast length (Buizza et al, 1993, Molteni et. al, 1996),
(3) Kalman filter based rescaling schemes and
(4) system simulation approaches.

A number of studies were performed with the aim to compare different global EPS (Wei and Toth, 2003; Buizza et al., 2005, Wang and Bishop, 2003). These indicated that
• the error breeding scheme may be superior compared with singular vectors at short lead times;
• the ETKF may be superior to the error breeding in a number of aspects;
• the local domain Ensemble Transform Kalman Filter (Ott et al. 2004) may outperform the global domain Ensemble Transform Kalman Filter in resolving medium- and short- range synoptic systems.
Techniques to validate performance of the EPS, different commonly used criterion of verification and relationships between them compose the second topic of this report.

2. The numerical weather prediction as a sequential update problem

2.1. **A general formulation of the sequential update problem.**

Let us denote $\mathcal{X}_\tau$ a m-dimensional state vector describing the atmosphere at time τ, y_τ a p-dimensional vector of observed quantities at time τ and let us denote $\mathcal{Y}_t = \{y_s,\ s \leq t\}$ a collection of all observations available up to time t. The objective of *the weather prediction* is to construct of a conditional density function $p(\mathcal{X}_\tau \mid \mathcal{Y}_t)$ of the state of atmosphere, valid at the time moment τ, given observations $\mathcal{Y}_t$, available up to the time moment $t \leq \tau$. The objective of the *numerical weather prediction* is to construct the conditional density $p(X_\tau \mid \mathcal{Y}_t)$ of X_τ, the discrete approximation of the state of the atmosphere $\mathcal{X}_\tau$, given a set of the observations $\mathcal{Y}_t$. In case if $\tau = t$, the conditional probability density $p(X_\tau \mid \mathcal{Y}_\tau)$ is called the analysis density, and in case if $\tau > t$, the conditional density $p(X_\tau \mid \mathcal{Y}_t)$ is called the forecast density.

Let us denote $t_i,\ i = 1, \ldots, n$ a sequence of time moments when the data assimilation is performed, $\mathcal{M}(t_i, t_{i-1})$ the deterministic dynamical propagator of the model state from one assimilation time to another and $\mathcal{H}_{t_i}$ the deterministic observation operator valid at time t_i. The data assimilation procedure can be described in the following way using the state space model terminology

$$\begin{aligned} y_{t_i} =& \mathcal{H}_{t_i}(X_{t_i}) + \epsilon_{t_i} \\ X_{t_i} =& \mathcal{M}(t_i, t_{i-1})(X_{t_{i-1}}) + T_{i-1}\xi_{t_{i-1}} \end{aligned} \tag{1}$$

where ϵ_{t_i} is a p-dimensional observation error vector, $\xi_{t_{i-1}}$ is a q-dimensional model error vector ($q << m$) and T_{i-1} is a $m \times q$-dimensional projection matrix. Both error terms are stochastically specified.

Even in the most general cases simplifying assumptions are done in order to justify the application of the state space model theory for data assimilation. For instance, the model error and the observation error are usually assumed to not depend on the state of the atmosphere

$$\xi(X_{t_{i-1}}, t_{i-1}) \equiv \xi(t_{i-1}), \quad \epsilon(X_{t_i}, t_i) \equiv \epsilon(t_i)$$

The sequence of the conditional probability densities $p(X_{t_i} \mid \mathcal{Y}_{t_i})$, $i = 1, \ldots, n$, can be obtained by solving the sequential probability density update problem: construct $p(X_{t_i} \mid \mathcal{Y}_{t_i})$ from $p(X_{t_{i-1}} \mid \mathcal{Y}_{t_{i-1}})$ for $i = 1, \ldots, n$, provided $p(X_{t_0} \mid \mathcal{Y}_{t_0}) \equiv p(X_{t_0})$ is already specified. Applying the definition of the conditional probability,

$$p(X_{t_i} \mid \mathcal{Y}_{t_i}) = \frac{p(X_{t_i}, \mathcal{Y}_{t_i})}{p(\mathcal{Y}_{t_i})} = \frac{p(X_{t_i}, y_{y_i} \mid \mathcal{Y}_{t_{i-1}}) p(\mathcal{Y}_{t_{i-1}})}{p(\mathcal{Y}_{t_{i-1}})}$$

Applying the probability multiplication rule and the factorization of the density over parameter $\mathcal{X}_{t_{i-1}}$ and utilizing the Markovian properties of the state space model, one can obtain

$$p(X_{t_i} \mid \mathcal{Y}_{t_i}) = \frac{1}{c} p(y_{t_i} \mid \mathcal{X}_{t_i}) \int p(X_{t_i} \mid X_{t_{i-1}}) p(X_{t_{i-1}} \mid \mathcal{Y}_{t_{i-1}}) dX_{t_{i-1}} \tag{2}$$

where c is a normalizing constant.

Here the construction of the conditional distribution is based on filtering. For each assimilation cycle i, first the probabilistic knowledge about model state at time t_{i-1}, based on the whole set of history observations $\mathcal{Y}_{t_{i-1}}$, is propagated forward until the next assimilation time t_i. Secondly, the probabilistic knowledge about model state is updated from new observation y_{t_i}.

2.2. **An analytical solution to the sequential update problem.**

Under very restrictive conditions, when

• (K.1) the model dynamical propagator $\mathcal{M}(t_i, t_{i-1})$ and the observation operator $\mathcal{H}_{t_i}$ are linear,

• (K.2) the distributional assumption about the model $\xi(t_{i-1})$ and the observation $\epsilon(t_i)$ errors are Gaussian, $\xi(t_{i-1}) \sim \mathcal{N}(0, Q_{t-1})$, $\epsilon(t_i) \sim \mathcal{N}(0, R_{t_i})$,

• (K.3) the initial model state distribution is Gaussian, $p(X_{t_0} \mid \mathcal{Y}_{t_0}) \equiv p(X_{t_0}) := \mathcal{N}(a_{t_0}, B_{t_0})$

• (K.4) and the model and observations errors are mutually uncorrelated and uncorrelated with the initial model state,

the sequential update of the conditional density function $p(X_{t_i} \mid \mathcal{Y}_{t_i})$ can be expressed analytically. In this case the $p(X_{t_i} \mid \mathcal{Y}_{t_i})$ is Gaussian as well and is completely determined through its two first central moments, $p(X_{t_i} \mid \mathcal{Y}_{t_i}) := \mathcal{N}(a_{t_i}, B_{t_i})$. The parameters a_{t_i}, B_{t_i} can be recursively calculated via the well-known standard Kalman filter equations:

$$
\begin{aligned}
a_{t_i} &= E(X_{t_i} \mid \mathcal{Y}_{t_i}) = a^f_{t_i} + B^f_{t_i}\mathcal{H}^T_{t_i}(R_{t_i} + \mathcal{H}_{t_i}B^f_{t_i}\mathcal{H}^T_{t_i})^{-1}(y_{t_i} - \mathcal{H}_{t_i}a^f_{t_i}) \\
B_{t_i} &= Var(X_{t_i} \mid \mathcal{Y}_{t_i}) = B^f_{t_i} - B^f_{t_i}\mathcal{H}^T_{t_i}(R_{t_i} + \mathcal{H}_{t_i}B^f_{t_i}\mathcal{H}^T_{t_i})^{-1}\mathcal{H}_{t_i}B^f_{t_i}
\end{aligned} \tag{3}
$$

where $a^f_{t_i}$ and $B^f_{t_i}$ are the parameters of the predictive distribution of the model state $p(X_{t_i} \mid \mathcal{Y}_{t_{i-1}})$ given the history of observations $\mathcal{Y}_{t_{i-1}}$. The $p(X_{t_i} \mid \mathcal{Y}_{t_{i-1}})$ is Gaussian as well

$$
\begin{aligned}
a^f_{t_i} &= E(X_{t_i} \mid \mathcal{Y}_{t_{i-1}}) = \mathcal{M}(t_i, t_{i-1})a_{t_{i-1}} \\
B^f_{t_i} &= Var(X_{t_i} \mid \mathcal{Y}_{t_{i-1}}) = \mathcal{M}(t_i, t_{i-1})B_{t_{i-1}}\mathcal{M}^T(t_i, t_{i-1}) + Q_{t_{i-1}}
\end{aligned} \tag{4}
$$

This estimate of the model state (eqn. 3) determines the best linear prediction of the model state on the innovations and the most probable estimate of the model state given observations at the same time. An innovation is a one-step ahead forecast error, $v_{t_i} = y_{t_i} - E(\mathcal{H}_{t_i}X_{t_i} \mid \mathcal{Y}_{t_{i-1}}) = y_{t_i} - \mathcal{H}_{t_i}a^f_{t_i}$. The Kalman filter recursions, which are strictly valid only under these firm conditions (K.1 -K.4), determine not only the time evolution and the update of the conditional mean and the conditional variance of the model state. They describe the development of the whole conditional probability density function given the observations. If the conditions (K.1-K.4) are not valid strictly, the best linear prediction of the model state on the innovations and the mean squared error of the prediction can be constructed. The best linear prediction $x^{ble}_{t_i}$ and the mean squared error $B^{ble}_{t_i}$ are given by

$$
\begin{aligned}
x^{ble}_{t_i} &= E(X_{t_i} \mid \mathcal{Y}_{t_{i-1}}) + cov(X_{t_i}, v_{t_i} \mid \mathcal{Y}_{t_{i-1}})(var(v_{t_i} \mid \mathcal{Y}_{t_{i-1}}))^{-1}v_{t_i} \\
B^{ble}_{t_i} &= Var(X_{t_i} \mid \mathcal{Y}_{t_{i-1}}) - cov(X_{t_i}, v_{t_i} \mid \mathcal{Y}_{t_{i-1}})(var(v_{t_i} \mid \mathcal{Y}_{t_{i-1}}))^{-1}cov(X_{t_i}, v_{t_i} \mid \mathcal{Y}_{t_{i-1}})^{-1}.
\end{aligned} \tag{5}
$$

If the model and observation errors are mutually uncorrelated and uncorrelated with the initial model state (K.4), the equations (5) will have a form notationally similar to the Kalman filter update equations (eqn. 3)

$$
\begin{aligned}
x^{ble}_{t_i} &= a^f_{t_i} + B^f_{t_i}H^T_{t_i}(H_{t_i}B^f_{t_i}H^T_{t_i} + R_{t_i})^{-1}v_{t_i} \\
B^{ble}_{t_i} &= B^f_{t_i} - B^f_{t_i}H^T_{t_i}(H^T_{t_i}B^f_{t_i}H_{t_i} + R_{t_i})^{-1}H_{t_i}B^f_{t_i}
\end{aligned} \tag{6}
$$

where $a^f_{t_i}$ and $B^f_{t_i}$ are the two first conditional moments of X_{t_i}, given the history of observation $\mathcal{Y}_{t_{i-1}}$, $E(X_{t_i} \mid \mathcal{Y}_{t_{i-1}})$ and $Var(X_{t_i} \mid \mathcal{Y}_{t_{i-1}})$, respectively. However, this system (eqn. 6) does not provide the sequential inference about the model state because $x^{ble}_{t_i}$ and $B^{ble}_{t_i}$ do not coinside with the two first

conditional moments $E(X_{t_i} \mid \mathcal{Y}_{t_i})$ and $Var(X_{t_i} \mid \mathcal{Y}_{t_i})$ and do not give any rules on how to sequentially update $a^f_{t_{i+1}}$ and $B^f_{t_{i+1}}$.

2.3. **Smoothing of the unobservable model state.**

Smoothing is an alternative procedure to obtain the conditional probability distribution of the model state based on observations. In this case the conditional distributions of the whole sequence of unobservable model states $(X_{t_0}, X_{t_1}, \ldots, X_{t_n})$ given all available observations $\mathcal{Y}_{t_n}$ is constructed.

Applying the probability density multiplication rule and the Markovian properties of state space model, one can obtain

$$p(X_{t_0}, X_{t_1}, \ldots, X_{t_n} \mid \mathcal{Y}_{t_n}) = \frac{1}{c} p(X_{t_0}) \prod_{i=1}^{n} p(y_{t_i} \mid X_{t_i}) p(X_{t_i} \mid X_{t_{i-1}}). \tag{7}$$

Here c denotes a normalizing constant.

Provided that conditions (K.2-K.4) holds, the posterior distribution $p(X_{t_0}, X_{t_1}, \ldots, X_{t_n} \mid \mathcal{Y}_{t_n})$ is Gaussian and is completely determined through its two first moments,

$$p(X_{t_0}, X_{t_1}, \ldots, X_{t_n} \mid \mathcal{Y}_{t_n} \sim \mathcal{N}(\tilde{a}, \tilde{B}),$$

where $\tilde{a}$ is a $(n+1) \times m$-dimensional vector, $\tilde{a} = (\tilde{a}_{t_0}, \tilde{a}_{t_1}, \tilde{a}_{t_2}, \ldots, \tilde{a}_{t_n})^T = (E(X_{t_0} \mid \mathcal{Y}_{t_n}), E(X_{t_1} \mid \mathcal{Y}_{t_n}), \ldots, E(X_{t_n} \mid \mathcal{Y}_{t_n}))^T$ and $\tilde{B}$ is a $(n+1)m \times (n+1)m$ dimensional matrix, $\tilde{B}_{ij} = cov(X_{t_i}, X_{t_j} \mid \mathcal{Y}_{t_n})$, $0 \leq i, j \leq n$. Numerically, a mean and a variance of the Gaussian distribution can always be obtained by calculating the maximum of the log-density and the curvature of the log-density at the point of the maximum.

$$\begin{aligned} \tilde{a} &= argmin L(X_{t_0}, \ldots, X_{t_n}) \\ &= argmin\{-\log p(X_{t_0}) - \sum_{i=1}^{n} (\log p(y_{t_i} \mid X_{t_i}) + \log p(X_{t_i} \mid X_{t_{i-1}}))\} \\ \tilde{B} &= \left[(\frac{\partial^2 L(X_{t_0}, \ldots, X_{t_n})}{\partial X_{t_i} \partial X_{t_j}})_{0 \leq i,j \leq N} \right]^{-1} \end{aligned} \tag{8}$$

Here the notation *argmin* means "argument that minimizes"

Notice that the joint distributions $p(X_{t_0}, X_{t_1}, \ldots, X_{t_n} \mid \mathcal{Y}_{t_n})$, $i = 0, \ldots, n$ are multiplications of Gaussian ones even if the Gaussian state space model is not linear (condition (K.1) holds). However, the marginal distributions $p(X_{t_i} \mid \mathcal{Y}_{t_n})$, $i = 0, \ldots, n$, are Gaussian only if the Gaussian state space model is linear (condition (K.1) holds). In condition (K.1) holds, the mode and the curvature at the mode, $\tilde{a}$ and $\tilde{B}$, can efficiently be calculated applying the forward Kalman filter and the backward Kalman smoother recursive equations (Durbin and Koopman, 2001).

If the perfect model is assumed ($\xi_i \equiv 0$, $i = 0, \ldots, n-1$), the dimensionality of the minimisation functional is dramatically reduced from $(n+1)m$ to m. In that case the whole uncertainty about the unobservable model state originate from the initial conditions and $p(X_{t_0} \mid \mathcal{Y}_{t_n})$ is the single conditional distribution that should be determined.

$$p(X_{t_0} \mid \mathcal{Y}_{t_n}) = \frac{1}{c} p(X_{t_0}) \prod_{i=1}^{n} p(y_{t_i} \mid X_{t_0}), \tag{9}$$

and $p(X_{t_0} \mid \mathcal{Y}_{t_n}) \sim \mathcal{N}(\tilde{a}, \tilde{B})$,where

$$
\begin{aligned}
\tilde{a} &= argmin L(X_{t_0} \mid \mathcal{Y}_{t_n}) \\
&= argmin\{-\log p(X_{t_0}) - \sum_{i=1}^{n} (\log p(y_{t_i} \mid X_{t_i}))\} \\
\tilde{B} &= \left[\frac{\partial^2 L(X_{t_0})}{\partial X_{t_i}^2}\right]^{-1}
\end{aligned}
\tag{10}
$$

For example, under a Gaussian state space model the minimisation functional $L(X_{t_0} \mid \mathcal{Y}_{t_n})$ is

$$
\begin{aligned}
L(X_{t_0} \mid \mathcal{Y}_{t_n}) =& 0.5(X_{t_0} - a_{t_0})^T B_{t_0}^{-1}(X_{t_0} - a_{t_0}) + \\
& 0.5 \sum_{i=1^n} (y_{t_i} - H_{t_i}\mathcal{M}(t_i, t_0)X_{t_0})^T R_{t_i}^{-1}(y_{t_i} - H_{t_i}\mathcal{M}(t_i, t_0)X_{t_0})
\end{aligned}
\tag{11}
$$

Still for a large-dimensional model state the design of the minimisation procedure (eqn. 10) is a challenging task.

If the state space model is not Gaussian or the conditions (K.1-K.4) do not hold, the mode of the posterior distribution (eqn. 7) still can be obtained by minimising the corresponding functional. It is possible to construct a sequence of Gaussian approximative state space models that in the limit will have a mode/a conditional mean, which will coincide with the mode of the original posterior distribution (Durbin and Koopman, 2001).

3. Approximate solutions to the sequential update problem.

In meteorological data assimilation the strict conditions (K.1-K.4) are never met. Both the dynamical propagator and the observation operator are in principle non-linear, the model and observation errors are correlated with the initial model state and these errors are assumed to obey the Gaussian distribution just for simplicity. The degree of non-linearity and non-Gaussianity differs significantly between different types of observations and between dynamical propagation of different spatial and temporal scale phenomena. Relaxing some of these strict conditions, different approaches for approximate solutions of the problem have been proposed and implemented at different weather services. Essentially, they can roughly be divided into

• the *variational type*, which pretend to estimate mode of the conditional distribution $p(X_{t_0} \mid \mathcal{Y}_{t_n})$ (eqn. 9, 11),

and

• the *regression type*, which pretend to produce the best linear prediction $x_{t_n}^{ble}$ of the X_{t_n} on the whole sequence of available observations (eqn. 5, 6),

approaches.

Let us denote τ a time moment for which the data assimilation should be performed and $\Delta\tau$ to be the length of the assimilation window. Then, under the variational approach, the analysed model state $x_\tau^a \approx E(X_\tau \mid \mathcal{Y}_\tau)$ is estimated by minimizing a certain functional, often called a cost function. The Three-dimensional Variational data assimilation scheme (3D-Var) and the Four-dimensional Variational assimilation scheme (4D-Var) are used worldwide at different weather services with great success. Both schemes are implemented in a so-called incremental formulation. This means the "optimal" increment $\delta x_\tau^a = x_\tau^a - x_\tau^f$ is obtained instead of estimating the whole analysed state. Here x_τ^f is

the best available forecast of the model state at time moment τ based on the history of observations, $x^f_\tau = \mathcal{M}(\tau, \tau - \Delta\tau)x^a_{\tau-\Delta\tau}$, and is often called the background state.

3.1. **Variational data assimilation schemes.**

• *The 3-Dimensional Variational data assimilation scheme (3D-Var)*

$$
\begin{aligned}
x^a_\tau &= x^f_\tau + \delta x^a_\tau \\
\delta x^a_\tau &= argmin(J_b + J_o) \\
J_{3D} = J_b + J_o =& 0.5\delta x^T_\tau B^{-1}\delta x_\tau + 0.5(H_\tau(x^f_\tau) + \bar{\mathcal{H}}_\tau\delta x_\tau - y_\tau)^T R^{-1}(H_\tau(x^f_\tau) + \bar{\mathcal{H}}_\tau\delta x_\tau - y_\tau) = \\
& 0.5\eta^T_\tau\eta_\tau + 0.5(H_\tau(x^f_\tau) + \bar{\mathcal{H}}_\tau U^{-1}\eta_\tau - y_\tau)^T R^{-1}(H_\tau(x^f_\tau) + \bar{\mathcal{H}}_\tau U^{-1}\eta_\tau - y_\tau)
\end{aligned}
\tag{12}
$$

Here B is the matrix containing the covariances of the forecast errors of the model state, $\bar{\mathcal{H}}_\tau$ is a tangent-linear observation operator around the background state, U is a square-root of the inverse of the covariance matrix B, $B^{-1} = U^T U$, and $\eta_\tau = U\delta x_\tau$ is the control vector along which the minimization of the cost function J_{3D} is performed. Observations are first collected over the time period $(\tau - 0.5\Delta\tau, \tau + 0.5\Delta\tau)$, then projected to the time moment τ and after that assimilated at the time τ.

• *The 4-Dimensional Variational data assimilation scheme (4D-Var)*

Let t_i denote an actual time when an observation is carried out.

$$
\begin{aligned}
x^a_\tau &= x^f_\tau + \delta x^a_\tau \\
\delta x^a_\tau &= argmin(J_b + J_{to} + J_c) \\
J_{4D} = J_b + J_{to} + J_c =& \delta x^T_\tau B^{-1}\delta x_\tau \\
& + \sum_{i,\tau<t_i\le\tau+\Delta\tau} (H_t(x^f_{t_i}) + \bar{\mathcal{H}}_{t_i}\delta x_{t_i} - y_{t_i})^T R^{-1}(H_t(x^f_{t_i}) + \bar{\mathcal{H}}_{t_i}\delta x_{t_i} - y_{t_i}) \\
& + (x_{\tau+0.5\Delta\tau} - \sum_{t=\tau}^{\tau+\Delta\tau}\alpha_t x_t)^T Q(x_{\tau+0.5\Delta\tau} - \sum_{t=\tau}^{\tau+\Delta\tau}\alpha_t x_t)
\end{aligned}
\tag{13}
$$

The 4-Dimensional variational data assimilation scheme is an extension of the 3D-Var. The increment δx_τ is propagated forward by the tangent-linear dynamical propagator up to the time moment t_i when the actual observations y_{t_i} are carried out, $\delta x_{t_i} = \bar{\mathcal{M}}(t_i, \tau)\delta x_\tau$. In such a way model dynamics are involved as strong constraints in the optimization procedure of the 4D-Var. The optional additional term J_c in the cost-function J_{4D} expresses requirements of the smoothness of the solution in time, one example originating from a low-pass digital filter. Because the dynamical forward integration is involved in the data assimilation procedure, some type of initialisation is necessary. As shown in Lynch and Huang 1992, the digital filter initialisation is very similar to the nonlinear normal mode initialisation provided that there is a clear frequency separation between fastly propagating gravity modes and slowly propagating Rossby modes. The gravity modes are associated with divergent motion and may be created from the unbalanced horizontal pressure gradient or even from the linearly balanced initial state due to non-linear dynamics. The low-pass digital filter prohibits a drift away of the model state due to adjustment of the solution to high frequency oscillations. Here x_t denotes the solution dynamically integrated forward up to the time moment t, $x_t = \mathcal{M}(t, \tau)(x^f_\tau + \delta x^a_\tau)$ and α_t, $t = \tau, \ldots, \tau + \Delta\tau$ are the time filter weights. Formulation and discussion of the time filter weights can be found in Gauthier and Thepaut 2001.

The cost functions J_{3D} and J_{4D} are proportional to the posterior probability density $p(x_\tau \mid \mathcal{Y}_\tau)$ of the model state under the Gaussian state space model (eqn. 11). Thus the $3DVar$ and the $4DVar$ approaches are based on smoothing rather than on filtering, they do not predict the "optimal" model state at the end of the data assimilation window but estimate the most likely model state giving rise to the set of observations. Because the amount of observed quantities is much smaller than the dimensionality of the model state, the prior assumptions on the model state have strong influence on the posterior distribution of the model state. The covariance of the prior distribution B does not change in time, has a very simplified structure and is statistically/analytically deduced.

3.2. **The general ensemble Kalman Filter formulation.**

A Monte Carlo approximation of the conditional probability density function $p(x_\tau \mid \mathcal{Y}_\tau)$ provides one possibility to propagate the forecast error covariance matrix B in time, at least approximately.

The ensemble of the model states, $X^a_{\tau-\Delta\tau} = \left[X^a_{1,\tau-\Delta\tau}, \ldots, X^a_{N,\tau-\Delta\tau}\right]$, where N is the ensemble size, is assumed to represent the conditional probability density function of the model state $p(X_{\tau-\Delta\tau} \mid \mathcal{Y}_{\tau-\Delta\tau})$. Then the conditional predictive distribution of the model state $p(X_\tau \mid \mathcal{Y}_{\tau-\Delta\tau})$ is considered to be represented by the ensemble $X^f_\tau = \left[X^f_{1,\tau}, \ldots, X^f_{N,\tau}\right]$, where each ensemble member is propagated forward in time by model dynamics,

$$X^f_{i,\tau} = \mathcal{M}(\tau, \tau-\Delta\tau) X^a_{i,\tau-\Delta\tau} + \eta_{i,\tau}, \tag{14}$$

where the model error may be eventually sampled from $\eta_{i,\tau} \sim \mathcal{N}(0, Q_\tau)$. If a perfect dynamical model is assumed, the model error term is omitted, namely $Q_\tau \equiv 0$.

For instance, the conditional mean and the conditional covariance of the model state at time τ, given the set of observations assimilated during previous cycles $\mathcal{Y}_{\tau-\Delta\tau}$, are assumed to be estimated by means of the relationship

$$\begin{aligned} E(X_\tau \mid \mathcal{Y}_{\tau-\Delta\tau}) \approx & x^f_\tau = \\ & = \frac{1}{N}\sum_{i=1}^{N} X^f_{i,\tau} \\ Cov(X_\tau \mid \mathcal{Y}_{\tau-\Delta\tau}) \approx & B^f_\tau = \\ & \frac{1}{N-1}\sum_{i=1}^{N}(X^f_{i,\tau} - x^f_\tau)(X^f_{i,\tau} - x^f_\tau)^T = Z^f_\tau (Z^f_\tau)^T \end{aligned} \tag{15}$$

Here $Z^f_\tau = (Z^f_{1,\tau}, \ldots, Z^f_{N,\tau})$ denotes the ensemble of normalized forecast perturbations, $Z^f_{i,\tau} = \frac{1}{\sqrt{N-1}}(X^f_{i,\tau} - x^f_\tau)$.

The *forecast step*, the way how the ensemble of the forecast states at a new assimilation time is constructed from the ensemble of the analysed state obtained from the previous assimilation time, is common for different implementations of the ensemble filter and is given in eqn. (14). Various implementations of the ensemble filter propose different *analysis steps*, the way how the ensemble of the forecast states, representing $p(X_\tau \mid \mathcal{Y}_{\tau-\Delta\tau})$ is transformed into the ensemble of the analysed states to represent $p(X_\tau \mid \mathcal{Y}_\tau)$.

3.3. **Different implementations of the ensemble Kalman filter.**
The *ensemble Kalman filter* EnKF (Evensen, 1994, Houtekamer and Mitchell, 1998) utilises the standard Kalman filter recursions to perform the analysis step. The best linear predictor of the model state x_τ^{ble} on the last innovation $v_\tau = y_\tau - H_\tau x_\tau^f$, given the history of observations $\mathcal{Y}_{\tau-\Delta\tau}$, and its mean squared error B_τ^{ble} are obtained using equation 6. The two first conditional moments are approximated by simply setting $E(X_\tau \mid \mathcal{Y}_\tau) = x_\tau^{ble}$ and $Var(X_\tau \mid \mathcal{Y}_\tau) = B_\tau^{ble}$. In other words, the ensemble Kalman filter retains the "linearity" aspects of Kalman filtering and assumes implicitly Gaussian distributions for uncertainties in the forecast and the observations. The implementations of the EnKF could roughly be divided into 3 different approaches:
• *the ensemble Kalman filter with perturbed observations*, often associated with acronym EnKF (Burgers et al. 1998, Houtekamer and Mitchell 1998),
• *the square-root ensemble Kalman filter*, often associated with the acronym ESRF (Whitaker and Hamill 2002, Tippett et al. 2003, Bishop et al. 2001, Ott and Coauthors. 2004), and
• *the reduced rank square-root Kalman filter* (Heemink, Verlaan and Segers, 2001, Cohn and Todling, 1996, Verlaan and Heemink, 1997).

All filtering algorithms mentioned above are of the rescaling type. The ensemble of forecast states, which is supposed to sample the prediction distribution $p(X_\tau \mid \mathcal{Y}_{\tau-\Delta\tau})$, is transformed, stochastically or deterministically, into the ensemble of analysed states, which is supposed to sample the posterior distribution $p(X_\tau \mid \mathcal{Y}_\tau)$.

• *The ensemble Kalman filter with perturbed observations.*

Under the ensemble Kalman filter with perturbed observations each ensemble member is updated in the following way

$$X_{i,\tau}^a = X_{i,\tau}^f + K_\tau(y_{i,\tau} - \mathcal{H}_\tau(X_{i,\tau}^f)),\ i = 1, \ldots, N \tag{16}$$

where K_τ is a Kalman gain matrix calculated from the ensemble of the forecast states

$$K_\tau = Z_\tau^f(\bar{\mathcal{H}}_\tau Z_\tau^f)^T(\bar{\mathcal{H}}_\tau B_\tau^f \bar{\mathcal{H}}_\tau^T + R_\tau)^{-1} \tag{17}$$

and $y_{i,\tau}$, $i = 1, \ldots, N$ is a simulated ensemble of perturbed observations, where the spread of the ensemble reflects the precision of the observations, namely $y_{i,\tau} \sim \mathcal{N}(y_\tau, R_\tau)$ for each i.

The EnKF analysis update scheme provides a stochastic update of the ensemble of the model state perturbations during the assimilation step. In order to construct the ensemble of analysed states, the ensemble of the forecast states and the ensemble of observations are merged together in observation space taking into account the skill of both ensembles expressed via the respective ensemble spread. While merging, the Gaussian assumptions on the underlying forecast error and observation error distributions are made implicitly.

The perturbed observation approach introduces an additional source of sampling error. Under this scheme, the equations (6) for the sample mean and for the sample variance of analysed state are satisfied only on average, namely

$$E(\bar{X}_\tau^a) = x_\tau^a,$$
$$E(Z_\tau^a(Z_\tau^a)^T) = B_\tau^a,$$

taking into account eqn. (15). At the same time one should stress that adding the noise in perturbation space stabilizes the filter by solving the rank-deficiency problem in an innovative way.

Pham (2001) proposes a similar Ensemble Kalman filter, called the *second-order-exact EnKF*, based on slightly different considerations. During the analysis step the ensemble members are updated as follows

$$X^a_{i,\tau} = X^f_{i,\tau} + K_\tau(y_\tau - H_\tau X^f_{i,\tau}) + \epsilon_{i,\tau} \tag{18}$$

where $\epsilon_{i,\tau},\ i = 1, \ldots, N$, is a second-order-exact sample from the Gaussian distribution,

$$\epsilon_{i,\tau} \sim \mathcal{N}(0, K_\tau R_\tau (K_\tau)^T),$$

with linear constraints

$$\sum_{i=1}^{N} \epsilon_{i,\tau} Z^f_{i,\tau} = 0.$$

In other words the ensemble, of the analysed perturbations $Z^a_\tau = \frac{1}{\sqrt{N-1}}(X^a_{i,\tau} - x^a_\tau)$ has contributions from the space orthogonal to one spanned by the ensemble of forecast perturbations Z^f_τ,

$$Z^a_{i,\tau} = Z^f_{i,\tau} - K_\tau H_\tau Z^f_{i,\tau} + \epsilon_{i,\tau}.$$

• *The square-root ensemble Kalman filter*

Under the ensemble square-root Kalman filter (ESRF) the mean and the spread of the ensemble are adjusted so that they would exactly satisfy the equations (6).

$$\begin{aligned} Z^a_\tau &= Z^f_\tau C \\ \bar{Z}^a_\tau &= 0 \\ X^a_\tau &= x^a_\tau + Z^a_{i,\tau} \end{aligned} \tag{19}$$

where C is an explicitly calculated transformation which preserves the mean of the ensemble and under which the covariance of the the analysis ensemble matches its theoretical value, given by eqn. (6), namely

$$Z^a_\tau (Z^a_\tau)^T = Z^f_\tau C (Z^f_\tau C)^T = (I - K_\tau H_\tau) Z^f_\tau (Z^f_\tau)^T \tag{20}$$

The ESRF provide a deterministic update of the ensemble of model perturbations during the assimilation step. In order to construct the ensemble of analysed states, the ensemble of forecast states is rotated and scaled. The ensemble estimate of the forecast error covariance matrix in observation space is used to determine the scaling and rotation. Because the mean and the variance of the conditional ensemble of the model state, given the whole set of observations up to time τ, satisfy equations 6 by construction, the Gaussian distributions for the forecast error and observation error are made implicitly. The main drawback of the method, in comparison with the perturbed observations approach, is that the ensemble of analysed perturbations is sampled from the space spanned by the ensemble of forecast perturbations only.

There is an infinite amount of square-root transformations C which satisfy requirement (20). Bishop et al. (2001) propose an elegant solution to this equation which allows an explicit look into the

mechanism of Kalman Filtering. Wang et al. (2004) developed the scheme further to construct the transformation preserving the mean of the ensemble.

$$C = G(D+I)^{-1/2}G^T \tag{21}$$

where a diagonal matrix D contains the (N-1) non-zero eigenvalues of the estimated forecast covariance in ensemble space, standardised by the observation error variance, $(\bar{\mathcal{H}}_\tau Z_\tau^f)^T R^{-1}\bar{\mathcal{H}}_\tau Z_\tau^f$, and a $N\times(N-1)$-dimensional matrix G contains the corresponding orthonormal eigenvectors of $(\bar{\mathcal{H}}_\tau Z_\tau^f)^T R_\tau^{-1}\bar{\mathcal{H}}_\tau Z_\tau^f$.

The multiplication from the right by G^T provides a spherical simplex centering of the ensemble after rotation and scaling was performed. Because matrix G is orthogonal, the ensemble of analysed states is centered without destroying its square-root property.

Besides that, Sakov and Oke, 2008, have shown that the symmetric transformation (eqn. 21) provides a unique solution. In the same paper Sakov and Oke generalised the scheme further by noting that the general mean-preserving solution for the ensemble Transform Kalman filters may be written as

$$C = G(D+I)^{-1/2}G^T U^p \tag{22}$$

where U^p is an arbitrary orthonormal mean-preserving matrix

$$U^p 1 = 1, \;\; U^p (U^p)^T = I$$

They provided an efficient algorithm for construction of a random mean-preserving orthogonal matrix U^p. With the transform defined by eqn. (22), a random ensemble of analysed states with the sample mean and sample variance exactly satisfying eqn. (6) can be constructed.

A variety of alternative algorithms to perform deterministic update of the ensemble during analysis exist. The main challenge of the ensemble Kalman filter is the necessity to invert the innovation covariance matrix, which has dimensionality of number of observations. Different algorithms propose different ways to overcome this obstacle. We mention some algorithms which have received large attention in the literature:

(1) a direct approach implemented in the first step of the Physical-space Statistical Analysis System (PSAS) algorithm (Cohn et al. 1998),
(2) a serial assimilation of observations (Houterkamer and Mitchell 2001, Bishop et al. 2001, Whitaker and Hamill 2002) and
(3) the ensemble adjustment Kalman filter (Anderson, 2001).

As it was noticed in Wang and Bishop (2003), the Ensemble Square-root Filters have a very flat eigenvalue spectrum due to the filtering effect of covariance, eqn. (6).

Because the trace of the ensemble estimate of the model state covariance cannot exceed the ensemble size, this flat eigenvalue spectrum induces a severe underestimation of the analysed error covariance, and as a result a severe underestimation of the forecast error covariance in the beginning of the next assimilation time. This leads to filter divergence. In fact, the whole derivation of the rescaling matrix C is implicitly based on the assumption that the ensemble of forecast perturbations is large enough to represent adequately the forecast error covariance matrix in observation space, $\bar{H}_\tau B_\tau^f \bar{H}_\tau^T$. If the ensemble size is too small, the ensemble estimate of the forecast error covariance will lack contributions from important directions. To increase the spread of the analysis ensemble by multiplying the transformed perturbations by an inflation factor or to sample an additional uncertainty are some

possibilities to overcome the problem. The spread of innovations can be used to design the inflation factor (Wang and Bishop 2003, Dee 1995). The inflation factor Π_τ is defined as

$$\Pi_\tau = \Pi_{\tau-1}\sqrt{\alpha_\tau} \tag{23}$$

where the parameter α_τ is such that

$$\tilde{d}_\tau^T \tilde{d}_\tau \approx trace(\tilde{H}_\tau \alpha_\tau B_\tau^f \tilde{H}_\tau^T + I). \tag{24}$$

Here $\tilde{d}^\tau$ is a vector of the standardised innovations,

$$\tilde{d}_\tau = R_\tau^{-1/2}(y_\tau - H_\tau x_\tau^f),$$

and $\tilde{H}_\tau$ is a standardised observational operator, $\tilde{H}_\tau = R_\tau^{-1/2}\bar{H}_\tau$.

When the inflation factor is implemented the total transformation matrix at time τ becomes

$$C_\tau = \Pi_\tau G_\tau (D_\tau + I)^{-1/2} G_\tau^T \tag{25}$$

Recently, Wang et al. (2007) have shown that such a simplistically designed inflation factor (eq. 24) leads to the overestimation of the true covariance matrix in the subspace spanned by the ensemble members if the dimensionality of the ensemble is much smaller than the dimensionality of the model state in the normalised observation space. In order to improve the filter performance they propose an alternative rescaling matrix

$$C_\tau = \Pi_\tau G_\tau (\rho D_\tau + I)^{-1/2} G_\tau^T \tag{26}$$

where the scalar factor ρ is the fraction of the forecast error variance projected into the ensemble space. It can be estimated by

$$\rho = \frac{\overline{\tilde{d}_\tau^T E_\tau^T E \tilde{d}_\tau} - (N-1)}{\overline{\tilde{d}_\tau^T \tilde{d}_\tau} - p} \tag{27}$$

where p is the number of observations, N is the number of ensemble members and E_τ is a matrix of the eigenvectors of the model error covariance matrix in normalised observation space. As shown in Bishop et al. (2001)

$$E_\tau = \tilde{H}_\tau Z_\tau^f G_\tau D_\tau^{-1/2} / \sqrt{N-1}$$

The averaging is done over a number of independent cases in the expression for ρ.

Noisiness of the ensemble estimate of the covariance matrix is another problem which originates from the small ensemble size. A "covariance localisation" (Gaspari and Cohn 1999) which explicitly damps correlations between model state components at long distances is a practical way to improve the quality of the ensemble estimate of the covariance matrix. Buehner 2005 proposed a covariance localisation scheme applicable in the framework of the ETKF. The mistreat of the statistical balances and long-scale variation are dangers of this approach. Filtering noise directly in ensemble space is another possibility to handle the problem. Ott and coauthors (2004) solved the problem associated with the rank-deficiency of the ensemble square-root filters by implementing a local Ensemble Kalman Filter. In this approach the analysis at each grid-point is performed simultaneously using the model state variable components and the observations in a local region centered at that point. Because the assimilation is performed independently in each local region, the smoothness of the analysed field must be considered. In order to achieve this, the assimilation procedure is explicitly constrained to choose the analysis perturbations which minimize the distance to background state. Notice that the backwards

ensemble rotation, provided by the spherical simplex centering in the ETKF rescaling scheme, serves for the same purpose.

The ensemble of analysed states is strongly correlated with the ensemble of forecast states, such that the ensemble rescaling scheme (eqn. 21) preserves the pattern. There is a hope that this approach will reduce the aliasing of the long scale variations which appears due to local assimilation scheme. Another strong side of this approach is the local Gaussian approximation to the non-Gaussian problem in case of non-linear dynamics. The global ensemble Kalman filter, which assumes implicitly the Gaussian distribution of the uncertainty about the model state, cannot perform data assimilation properly in that case. However, a local Gaussian state space model can provide an efficient approximation of a non-Gaussian model state provided that the local neighbourhood is selected successfully. The local ensemble Kalman filter is supposed to describe well atmospheric processes associated with local energetics, such as the baropropic and baroclinic instabilities, downstream development of the upper-tropospheric wave packets, anticyclinuc wave breaking and other phenomena.

• *Reduced Rank Ensemble Kalman Filter.*

In the Reduced Rank Kalman Filter the full-rank covariance matrix of the model state is approximated by a matrix with a reduced rank. The reduced rank approach can be implemented via an ensemble square-root Kalman filter as well when the N ensemble members are selected in the direction of N leading eigenvectors of the covariance matrix. The reduced rank square-root ensemble Kalman filter algorithm was initially proposed by Verlaan and Heemink (1997, RRSQRT). The analysis step is deterministic and is based on the square-root Kalman filter analysis update

$$\begin{aligned} B_\tau^f =& L_\tau^f (L_\tau^f)^T \\ K_\tau =& B_\tau^f H_\tau^T (H_\tau B_\tau^f H_\tau^T + R_\tau)^{-1} \\ x_\tau^a =& x_\tau^f + K_\tau (y_\tau - H_\tau x_\tau^f) \\ \tilde{L}_\tau^a =& \{(I - K_{\tau H_\tau}) L_\tau^f, K_\tau R_\tau^{1/2}\} \\ L_\tau^a =& \Pi_\tau \tilde{L}_\tau^a \end{aligned} \tag{28}$$

Here L_τ^f is a matrix of perturbations, $L_\tau^f = [I_{1,\tau}, \ldots, I_{q,\tau}]$ and Π_τ is a projection onto the q leading eigenvectors of the matrix ${}_\tau^a(L_\tau^a)^T$. To initialize the filter L_0^a is taken to be the q largest eigenvectors of the static model error covariance matrix B_0

$$L_0^a = [I_{1,0}^a, \ldots, I_{q,0}^a]$$

To propagate the perturbations the dynamical step (eqn. 14) is generalized by implementing a finite differences approach

$$\begin{aligned} x_\tau^f =& \mathcal{M}(\tau, \tau - \Delta\tau)(x_{\tau-\Delta\tau}^a) \\ I_{i,\tau}^f =& \frac{1}{\epsilon}\{\mathcal{M}_{\tau,\tau-\Delta\tau}(x_{\tau-\Delta\tau}^a + \epsilon I_{i,\tau-\Delta\tau}^a) - \mathcal{M}_{\tau,\tau-\Delta\tau}(x_\tau^a - \Delta\tau)\} \end{aligned} \tag{29}$$

Notice that if $\epsilon = 1$ the equation above coincides with (eqn. 14). The model error can be easily accounted for assuming

$$\begin{aligned} \tilde{L}_\tau^f =& [L_\tau^f, Q_\tau^{1/2}] \\ L_\tau^f =& \Pi_\tau^f \tilde{L}_\tau^f. \end{aligned}$$

Here Π_τ^f is again a projection matrix on the q leading eigenvectors of the matrix ${}_\tau^a(L_\tau^a)^T$ and Q_τ is a model error covariance matrix at time τ. Heemink et al. (2001) propose extensions to the reduced rank ensemble Kalman: to nicely combine deterministic and probabilistic updates of the analysis ensemble. Under *the Partially Orthogonal Ensemble Kalman filter* (POEnKF), the matrix of ensemble L_τ^a consists of the q largest eigenvalues of B_τ^a, $I_{1,\tau}, \ldots, I_{q,\tau}$, and N random ensembles $\xi_{1,\tau}, \ldots, \xi_{N,\tau}$ to better sample uncertainty about x_τ^a

$$[L^{a,\tau}, E_{a,\tau}] = [I_{1,\tau}, \ldots, I_{q,\tau}, \xi_{1,\tau}, \ldots, \xi_{N,\tau}].$$

For initialisation of the filter, the random ensemble is sampled from the Gaussian approximation to the model state distribution, $\xi_{i,0} \sim \mathcal{N}(x_0^a, B_0)$. During the analyses step the forecast error covariance matrix is estimated

$$\begin{aligned} E_\tau^{*,f} =& \Pi_\tau^* E_\tau^f \\ B_\tau^f =& L_\tau^f (L_\tau^f)^T + \frac{1}{N-1} E_{\tau+}^{*,f} (E_\tau^{*,f})^T \end{aligned} \tag{30}$$

where Π_τ^* is a projection of the random ensemble to the space orthogonal to the space spanned by L_τ^f. The analysis update of the ensemble is performed in accordance with (eqn. 28) for the deterministic part of the ensemble, L_τ^a and in accordance with (eqn. 16) for the random part of the ensemble, E_τ^a.

In the *Complementary Orthogonal Subspace Filter for Efficient Ensembles* (COFFEE) the random part of the ensemble, E_τ^f is explicitly constrained to sample from the subspace orthogonal to the L_τ^f.

$$E_\tau^f = [\xi_{1,\tau} - x_\tau^f + \eta_{1,\tau}, \ldots, \xi_{N,\tau} - x_\tau^f + \eta_{N,\tau}] \tag{31}$$

where $\eta_{i,\tau} \sim \mathcal{N}(0, (I - \Pi_\tau^f) L_\tau^f (L_\tau^f)^T (I - \Pi_\tau^f)^T)$.

Heemink et al. (2001) shows that adding of structured random noise to the reduced rank Kalman filter significantly improves it behaviour.

The *Singular Evolutive Extended Kalman filter* (SEEK) developed by Pham et al. (1998b) and the *Singular Evolutive Interpolated Kalman filter* (SEIK) developed by Pham 1997 can be considered to be a type of reduced-rank Kalman filter. The idea behide SEEK filter is to make correction only in the direction where error is amplified or is large, keeping at the same time the error small in other directions. Ideally, the reduced-rank covariance matrix should span the attractor which is of low rank. However it seems to be an unrealistic task to approximate the attractor of so complicated non-linear system in such a simple way. In essence, all ensemble Kalman filter implementations severely suffer from the problem of rank deficiency. Whatever rescaling scheme is selected, the rank of the sample covariance of the analysed states will be bounded from above by the amount of ensemble members.

- *The particle type ensemble filters*

In all algrorithms discussed above the implicit resampling of the model state during the assimilation cycle is done mainly in order to compensate for the underestimation of the covariance of the model state errors, which arises from a rank-deficiency of the ensemble Kalman filter. Besides these algorithms, a number of elegant ensemble filter implementations based on the particle filter idea are proposed in the literature (van Leeuwen (2003), Kim et al. (2003), Chin et al. (2007)). The particle filters are non-parametric. The conditional distribution $p(X_{\tau-\Delta\tau} \mid \mathcal{Y}_{\tau-\Delta\tau})$ is approximated by a discrete distribution, located on N analysis states, $x_{i,\tau-\Delta\tau}^a,\ i = 1, \ldots, N$, with equal probability, $p_{1,\tau-\Delta\tau} = p_{2,\tau-\Delta\tau} = \ldots = p_{N,\tau-\Delta\tau} = 1/N$. The conditional distribution $p(X_\tau \mid \mathcal{Y}_{\tau-\Delta\tau})$ is approximated by discrete distribution, located at N forecast states $x_{i,\tau}^f,\ i = 1, \ldots, N$, with the same probability $p_{i,\tau} = 1/N$. The model

dynamics, eqn (14), are used to propagate the particles forward in time. To approximate the conditional distribution $p(X_\tau \mid \mathcal{Y}_\tau)$ at a new assimilation time, the probabilities of particles $p_{i,\tau}$ are changed. They are not anymore equal to each other, while the particle themselves remain untouched. To improve performance of the particle filter, a resampling step is introduced. At the new assimilation time the ensemble of analysed states x^a_τ is resampled from the ensemble of the forecast states x^f_τ in accordance with modified probabilities $p_{i,\tau}$. The conditional distribution $p(X_\tau \mid \mathcal{Y}_\tau)$ is again approximated by a discrete one, with equal probabilities, located on particles $x^a_{i,\tau}, i = 1, \ldots, N$. The resampling prohibits the efficient rank of the ensemble to decrease during the assimilation cycles.

Even though the particle filters theoretically are able to handle the data assimilation properly in case of non-linear and non-Gaussian state space problems, the practical applicability of these filters for the meteorological data assimilation is quite limited. The main problem is a very slow convergence of the non-parametric estimate ($\mathcal{O}(1/N)$), especially in case of a large-dimensional model.

A non-Gaussian extension of the EnKF that uses a mixture of Gaussian probability densities to describe the probability density of the model state is an alternative to handle non-linear and non-Gaussian state space models (Anderson and Anderson 1999, Bengtsson et al. 2003). Under this approach the conditional density $p(X_\tau \mid \mathcal{Y}_{\tau-\Delta\tau})$ is assumed to be

$$p(X_{\tau|\mathcal{Y}_{\tau-\Delta\tau}}) = \sum_{l=1}^{L} \pi^f_{l,\tau} \mathcal{N}(\mu^f_{l,\tau}, B^f_{l,\tau}),$$

where $\mathcal{N}(\mu^f_{l,\tau}, B^f_{l,\tau})$ denotes a Gaussian density with mean $\mu^f_{l,\tau}$ and covariance matrix $B^f_{l,\tau}$. All parameters $\mu^f_{l,\tau}$, $B^f_{l,\tau}$ and mixing probabilities $\pi^f_{l,\tau}$ are estimated from the ensemble of the forecast states $X^f_{i,\tau},\ i = 1, \ldots, N$, assuming that the size N of the ensemble is much larger than the amount of the selected classes L. During the assimilation step, the conditional density $p(X_\tau \mid \mathcal{Y}_\tau)$ is approximated by

$$p(X_\tau \mid \mathcal{Y}_\tau) = \sum_{l=1}^{L} \pi^a_{l,\tau} \mathcal{N}(\mu^a_{l,\tau}, B^a_{l,\tau}),$$

where $\mu^a_{l,\tau}$ and $B^a_{l,\tau}$ are updated in accordance with the Kalman filter recursive formulas, eqn. (6), separately for each class L. The mixing probabilities are updated in such way that the classes closer to observations would have higher weights.

3.4. **The Hybrid Ensemble Kalman Filter - Variational assimilation scheme..**

The Hybrid Ensemble Kalman Filter-Variational Assimilation scheme has received much attention in the literature at present time (Wang et al 2007, Hamill and Snyder 2000). The idea of the hybrid scheme is to combine the best possible achievements of both the variational and the ensemble assimilation schemes. The successful assimilation scheme must point out areas of strongest forecast uncertainty and try extract as much as possible information from the available observations in those areas. When the data assimilation is performed under the Variational or the Ensemble Kalman scheme, in both cases the analysed state in each gridpoint is a weighted sum of the forecast state and the surrounding observations with weights being determined by the relative uncertainty of both sources of information. That is why the realistic estimation of the current forecast uncertainty, which is large in dynamically unstable areas and low in the areas with dense observation network, is important for constructing the analysed state. Even if the number of observations available and utilized by data assimilation scheme

at present is relatively large (from $\mathcal{O}(10^4)$ to $\mathcal{O}(10^6)$), the dimensionality of the model state in high resolutions is still much higher. The prior assumptions on the model state, expressed via the forecast error covariance matrix, influence strongly the construction of the analysed state. In order to perform a proper extrapolation of the observations into the model state, the flow-dependent structure functions of the forecast error covariance are essential.

The forecast error covariance used in variational data assimilation schemes at present does not have this property. It is static and has simplistic structure functions based on spatial homogeneity and isotropy. There is a hope that a forecast error covariance matrix B, which combines together the full rank static forecast error covariance matrix B_{3DVar} and the flow-dependent rank-deficient ensemble forecast error covariance matrix B_e, based on the Kalman filter rescaling ensemble, will improve the variational assimilation scheme.

The hybrid assimilation schemes are proposed in two different formulations, which are theoretically equivalent (Wang et al. (2007)). Hamill and Snyder (2000) use the same cost function as in 3D-Variational data assimilation (eqn. 12) but with the forecast error covariance matrix B being equal

$$B := B_\tau^f = \alpha B_{3DVar} + (1-\alpha)(Z_\tau^f (Z_\tau^f)^T \cdot L) \tag{32}$$

where Z_τ^f is the ensemble of the normalised forecast perturbations, L is the prescribed correlation matrix used for the covariance localisation and $A \cdot B$ denotes the Schur product of the A and B matrices, $(A \cdot B)_{i,j} = A_{i,j} B_{i,j}$.

Lorenc (2003) and Buehner (2005) propose a modification to the cost function by augmenting the set of control variables

$$\begin{aligned} J =& \frac{1}{2}\eta_\tau^T \eta_\tau + \frac{1}{2} a_\tau^T A^{-1} a_\tau \\ &+ \frac{1}{2}(H_\tau(x_\tau^f + \Delta x_\tau) - y_\tau)^T R^{-1} (H_\tau(x_\tau^f + \Delta x_\tau)^T - y_\tau) \end{aligned} \tag{33}$$

which is to be minimised with respect to the "optimal" increment Δx_τ. The analysis increment Δx_τ is expressed via two sets of control variables, namely the usual variational ones, which are of the size of the model state, and a set of new control variables associated with the flow-dependent structures resolvable by the ensemble

$$\begin{aligned} \Delta x_\tau &= \beta_1 \Delta x_{1,\tau} + \beta_2 \Delta x_{2,\tau} \\ &= \beta_1 (B_{3DVar})^{1/2} \eta_\tau + \beta_2 (B_\tau^e)^{1/2} a_\tau \\ &= \beta_1 (B_{3DVar})^{1/2} \eta_\tau + \beta_2 Z_\tau^f a_\tau \end{aligned} \tag{34}$$

Here β_1 and β_2 are empirically estimated weights and matrix A is a filter stabiliser.

Buehner (2005) proposes to augment the control variable set further in order to implement the covariance localisation in the context of the ETKF. He points out that the square-root of the localised ensemble covariance can be expressed as

$$Z_{loc,\tau}^f = [diag(Z_{1,\tau}^f) L^{1/2}, diag(Zf_{2,\tau}) L^{1/2}, \ldots, diag(Z_{N,\tau}^f) L^{1/2}], \tag{35}$$

where, again, L is the prescribed correlation matrix for the localisation, $Z_{i,\tau}^f$ is the i-th perturbation and $diag(Z_{i,\tau}^f)$ is a $m \times m$ diagonal matrix with the $Z_{i,\tau}^f$ on the diagonal. Equation (35) explicitly proves that the covariance localisation reduces the effect of sampling error. The square-root $Z_{loc,\tau}^f$ of the sample estimate of the forecast error covariance matrix has larger dimensionality, which is at most

Nr, in comparison with the dimensionality of the square-root matrix Z^f_τ, which is N. Here r is the dimensionality of L. In this case the set of variational control variables should be augmented by Nr new control variables associated with the flow-dependent structures.

One important remark is that it is not straight-forward to apply the covariance localisation. Essentially, the covariance localisation lacks physical background and destroys important balances between the model state components. For example, the geostrophic component wind components should be in an approximate linear balance with the spatial geopotential gradient. One possibility is to apply the localisation to streamfunction only and let other variable components to adjust to each other after the localisation.

$$
\begin{aligned}
J =& \frac{W_b}{2}(\delta x_\tau)^T B^{-1}_{3DVar}\delta x_\tau + \frac{W_a}{2}a^T A^{-1} a \\
&+ \frac{1}{2}(H_\tau(x^f_\tau + \Delta x_\tau) - y_\tau)^T R^{-1}(H_\tau(x^f_\tau + \Delta x_\tau)^T - y_\tau)
\end{aligned}
\tag{36}
$$

The analysis increment is related to the control vector as follows

$$
\begin{aligned}
\Delta x_\tau &= \Delta x_{1,\tau} + \Delta x_{2,\tau} \\
&= U\eta_\tau + U^e \eta^e_\tau a
\end{aligned}
\tag{37}
$$

where η_τ is a set of control variables associated with the structure of B_{3DVar}, U is a transformation from the control vector space into the model space, η^e_τ is a set of variables estimated from the ensemble of the forecast states Z^f_τ and U^e is a transformation back to the model space that is supposed to preserve some important balances explicitly. Matrix A is an empirical correlation matrix (to be specified) which describes the structure of a_i-2D field. The 2D-fields a_i can be of a much coarser resolution than the model state.

Wang et al. (2007) has shown that the Hybrid Ensemble Transform- Optimal Interpolation scheme can work even without localisation in the case of a simplified dynamical model. The estimate of the forecast error covariance matrix is stabilised by merging the rank-deficient flow-dependent covariance matrix with the full rank static covariance matrix.

4. Ensemble prediction systems

The methods discussed in the previous section are all concerned with data assimilation. The issue of data assimilation is to merge the uncertainty in the forecast model and in the observation in an optimal way in order to construct the initial model state for the weather prediction. The ensemble methods used in data assimilation allow to construct not only a deterministic initial state for the forecasting but to quantify the uncertainty about the initial model state as well. The representation of uncertainty is done essentially through estimation/modelling of the analysis/forecast error covariance. One may say that for these models the Gaussian assumptions about the distribution of the model state are made implicitly. Even variational data assimilation schemes allow theoretically the quantification of the uncertainty about the initial model state. The inverse of the Hessian, which determines the curvature of the conditional probability density function, given observations, in a vicinity of its maximum, could be used as a measure of the uncertainty. However, the huge dimensionality of the model state does not allow to perform the inverse.

The ensemble prediction systems (EPS) were proposed by Leith (excellent review is given by Ehrendorfer (1997)) and have different aim. A successful EPS should sample the uncertainty about the initial

model state (the initial PDF) in such a way that it could describe a relevant part of the PDF (usually non-Gaussian) during and after the integration period of interest for the phenomena of interest. The phenomena of interest can often be related to the model state variable via non-linear small-dimensional transform. The EPS systems can roughly be divided into three groups dependent on how the initial ensemble of perturbations is created:

(1) sampling of dynamically unstable directions : the singular vectors optimized over certain forecast length and error breeding;
(2) sampling of the analyses errors :Kalman filter based rescaling;
(3) comprehensive sampling of different sources of uncertainty about the forecasting system: a system simulation approach.

Combinations of these three approaches also exist.

4.1. **Singular Vectors.**

Singular vectors represent those directions in the model space at initial time that give the maximum linear growth for a specific forecast period and over prespecified area. Maximisation is performed with respect to a certain norm. Typically singular vectors are maximised using the energy norm both at the initial and at the final time and are abbreviated TE SV (total energy singular vectors). TE SV are solutions of the following generalised eigenvalue problem

$$(\mathcal{M}(t_0,t_p))^T P^T E_p P \mathcal{M}(t_0,t_p) x = \lambda E_0 x \tag{38}$$

where λ is an eigenvalue corresponding to x, E_0 and E_p is energy norms at the initial and at the final time, $\mathcal{M}(t_0,t_p)$ is a tangent-linear dynamical propagator over the period $t_0 < t \leq t_p$ and P is a projection operator to the prespecified area.

The total energy norm is calculated via the total dry energy transformation, which was first proposed by Talagrand (1981) and was studied in details by Errico (2000)

$$< x, Ex >= \sum_l (u_l^T D_l u_l + v_l^T D_l v_l + \frac{c_p}{T_r} t_l^T D_l t_l) + R_d T_r (ln(P_s))^2 \tag{39}$$

where l is a vertical model level, D_l is the vertical transformation, $D_l = P_{l+1/2} - P_{l-1/2}$, u_l, v_l, t_l and P_l are wind components, temperature and pressure at the model level l, P_s is the surface pressure, $c_p = 1004$ J K^{-1} kg^{-1} is the specific heat for dry air at constant pressure, $T_r = 273$ K is the reference temperature and $R_d = 287$ J mol^{-1} K^{-1} is the gas constant for dry air. This quantity was used to generate initial perturbations in Buizza et al. (1993). The theoretical expression for the total energy norm can be found in Barkmeijer et al. (1999), for instance.

In the case of linear dynamics the set of singular vectors $X_{SVk}(t_0)$ corresponding to different λ_k, $\lambda_1 > \lambda_2 > \ldots \lambda_N$, would span the most rapidly growing direction at the optimization time t_p ($t_p = 48$ hours is typically assumed). The evolved singular vectors $\mathcal{M}(t_p,t_0)X_{SVk}(t_0)$ will form the E-orthogonal set at the optimisation time t_p.

The derivations of singular vectors is based on the model dynamics. These vectors are believed to sample the unstable linear subspace as efficiently as possible. To generate the initial perturbations for the EPS the singular vectors are rescaled to represent the system with a realistic initial spread. Both the optimisation time and the phenomena of interest influence the construction of the singular

vectors. For example, by maximizing the total energy norm in the small target area the system can be designed for a specific region and process of interest.

Such singular vectors are called Targeted Singular vectors. Frogner and Iversen(2001) and Hersbach et al. (2000) generated the targeted ensemble prediction systems (TEPS) for parts of Europe. TEPS provides the initial and the boundary fields for LAMEPS, a high resolution limited area ensemble prediction model. The model seems to be successful in forecasting extreme weather events and even large scale precipitation which involves strong mesoscale variability and is strongly influenced by orography. A detailed description and verification of LAMEPS can be found in Frogner and Iversen (2002).

To make EPS based on Singular Vectors more appropriate for the short range forecasting, the optimized (at some future time, usually 24 or 48 hours) singular vectors may be combined with the evolved singular vectors from the previous optimisation time. Hamill et al 2003 propose an alternative approach. They generate the ensemble of Analysis Error Covariance Singular Vectors (AEC SV) by solving the generalised eigenvalue problem (eqn 38) but with the initial energy norm E_0 in the equation above being replaced with the inverse of the analyses error covariance norm P_a^{-1}. So the idea is to sample the quickly growing directions, which initially have structures consistent with the analysis error covariance. Hamill et al. (2003) have shown that evolved and appropriately rescaled AEC SV v_{t_p} satisfy the following generalized eigenvalue equation

$$E_p^{1/2} X_{t_p}^f (E_p^{1/2} X_{t_p}^f)^T v_{t_p} = \lambda v_{t_p} \tag{40}$$

provided that the ensemble size is large and that the model dynamics is nearly linear. The forecast ensemble $X_{t_p}^f$ is obtained by dynamical forward integration of the analysis ensemble $X_{t_0}^a$, which is generated using the ensemble square-root Kalman filter algorithm (eqn. 20). Thus, the evolved singular vectors v_{t_p} are expressed as a linear combination of the forecast ensemble, $v_{t_p} = E_p^{1/2} X_{t_p}^f a$, where the vector a is obtained solving an equivalent (eqn. 40) but smaller, with the dimensionality of the ensemble, eigenvalue problem

$$(E_p^{1/2} X_{t_p}^f)^T (E_p^{1/2} X_{t_p}^f) a = \lambda a \tag{41}$$

Then the initial AEC singular vectors u_{t_0}, which are considered to give rise to v_{t_p}, are estimated as the same linear combination a but of the analyses ensemble $X_{t_0}^a$, namely

$$u_{t_0} = X_{t_0}^a a.$$

The typical structures of the initial-time AEC SV were significantly different from the typical structures of the total energy singular vectors (eqn. 38) and were similar to the subsequent forecast error structures, but smaller in amplitudes. However, Buehner and Zadra (2006) show that the shape of the evolved singular vectors is almost independent of the initial norm.

4.2. **Breeding vectors.**

The analyses error, or the initial-time error, consist of the random errors introduced by inaccuracies of the assimilated observations and the growing errors associated with the instabilities of the evolving flow, which are dynamically generated from the errors introduced at the previous assimilation times. Even if the growing part of the error is only a portion of the total analysis error, their impact on the forecast error is large. Therefore, creating an ensemble of initial states with a limited ensemble size seems to be appropriate to focus on the efficient sampling in the direction of growing errors (Ehrendorfer (1997), Toth and Kalnay (1997)). To sample the growing error direction Toth and Kalnay proposed a method called *breeding* of the growing vectors. The idea is to add an arbitrary perturbation to the initial state

at time t_0, to let it grow for a short time period (t_0, t_p), while the error growth is approximately linear, and to downscale the evolved perturbation so that it has the amplitude of the initial perturbation. The obtained perturbation is added to the analysis state at time t_p and the process is continued resetting. The down scaling of the evolved perturbations helps to eliminate decaying directions.

Theoretically the breeding perturbations are related to the local Lyapunov vectors of the atmosphere (Trevisan and Legnani (1995)).

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log_2 \left[\frac{p_i(t)}{p_i(0)}\right] \tag{42}$$

where p is a linear perturbation spanning the phase space of a system with orthogonal vectors. When the Lyapunov exponents are interpreted locally, each of them can be associated with a perturbation. The breeding technique is based on the fact that any random perturbation introduced an infinitely long time earlier develops into the leading local Lyapunov vector, the perturbation p with the largest exponent λ (eqn. 42).

Singular vectors (eqn. 38) provide another possibility to approximate the Lyapunov vectors. At the same time, as we have mentioned above, both optimisation time t_p and the optimisation area influence strongly singular vectors too.

In order to allow the initial perturbations constructed via the breeding error technique resemble the analysis error in a better way, a regional rescaling is introduced. The idea is to have larger perturbation amplitude in the regions sparsely observed. The scaling factor is a smooth function of horizontal location. A perturbation traveling into a poorly observed oceanic area is allowed to grow freely, while those reaching a well-observed area are scaled down to the size of the estimated analysis error (Augustine et al. (1992)).

4.3. **The perturbed observations approach.**

Houterkamer et al. (1996a) have developed an approach, alternative to the selective sampling, to generate the initial perturbation. This approach is in operational use at MSC since 1996. The initial conditions are generated by assimilating randomly perturbed observations, using different model versions in a number of independent data assimilation cycles. This is a type of system simulation experiment, when all uncertain parts of the forecasting system are subject to perturbations. In other words, the idea is to sample carefully all sources of uncertainty which determine the forecast error. Sources of uncertainty that are considered to have a significant impact on the forecast errors are observation errors (both measurement and representativity), model errors (the effect of unresolved scales and parametrisation of the physical processes), data assimilation processing errors (the unrealistic structure-functions of the covariance matrix) and the erroneous boundary fields (including imperfect estimation of the surface fields). To account for all these sources of errors the initial perturbation are generated using different sets of perturbed observations, different dynamical models and different sets of perturbed surface fields. At present the MSC scheme to sample different sources of uncertainty has significantly developed further by allowing elaborated perturbations of essential parameters of physical parametrisation.

4.4. **The Kalman Filter based rescaling.**

The Kalman Filter based rescaling approach (ETKF) to generate ensemble prediction system can be called a generalized breeding. The ensemble of the forecast perturbations is rescaled into the ensemble of the analysis perturbations. Under the error breeding methodology the forecast perturbations are

downscaled by a scalar or an ad-hoc matrix, in a case of a masked breeding, into the analysis perturbations. In contrast to this methodology, the forecast perturbations under the ETKF rescaling scheme are downscaled into the analysis perturbations by a carefully designed rescaling matrix which reflects both the relative forecast/observations uncertainty and the spatial observation distribution (eqn. 21). Theoretically, in case of a full-rank model, the analysis perturbations generated by the ETKF form a square-root of the analysis error covariance matrix. One drawback of the ETKF rescaling scheme is that if the number of the ensemble perturbations is much smaller than the number of directions to which the forecast error variance projects, the transformation (eqn. 20) heavily underestimates the analysis error covariance matrix (Wang and Bishop 2003). A simplistic way to overcome the problem is to multiply each ensemble member by a scalar to force the spread of the analysis perturbations to be consistent with the analysis error variance on a spatially averaged basis. It is still questionable if the perturbations modified in such a heuristic manner are able to describe the time development of the PDF of the model state. Wang and Bishop (2003) have shown that the fastest growth in the ensemble perturbation subspace generated by the ETKF is larger in the total energy norm than the fastest growth in the ensemble constructed via breeding.

The fastest growth γ is defined via a linear combination b of the perturbations to obtain the direction of the fastest growth of the total energy over the prespecified time period,

$$max \frac{b^T (Z^f_{t_p})^T S Z^f_{t_p} b}{b^T (Z^a_{t_0})^T S Z^a_{t_0} b}, \tag{43}$$

where S is an appropriate norm, the total energy norm (eqn. 39) in this case. In fact, the linear combination b is a leading eigenvector of $A^{-1/2} B A^{-1/2}$, where $A = (Z^a_{t_0})^T Z^a_{t_0}$, $B = (Z^f_{t_p})^T Z^f_{t_p}$ and $t_0 < t < t_p$ is the optimisation period (Wang and Bishop 2003). Thus the fastest error growth is the leading eigenvalue of $A^{-1/2} B A^{-1/2}$.

The analysis perturbations constructed via the ETKF have a very flat spectrum, especially in comparison with the perturbation constructed via error breeding.

Besides that the ETKF perturbation seems to be able to resolve a wider range of innovation variance than the breeding perturbations.

At the same time we would like to stress that the Kalman Filter based rescalig schemes allow to sample the uncertainty valid at the analysed time t_0, measuring the uncertainty via the variance-covariance of the model state. This means that the Gaussian assumptions about the PDF of the model state valid at t_0 are done implicitly. It is questionable wether the ensemble of limited size of the analyses states constructed in such way and dynamically propagated forward to the time interest is able describe adequately the relevant part of the PDF for the phenomena of interest.

5. Verification of the Ensemble Systems

The verification of ensemble prediction system for NWP involves some specific problems. The quality of the prediction system can be evaluated only based on observations, because only the observations reflect the true state of the atmosphere. At the same time the assessment of the quality can be done only statistically, based on a large number of realizations of both observations and ensemble predictions. Scores which are commonly used for evaluating the ensemble prediction system are extensively discussed in Toth et al. (2003) or Stanski et al. (1989).

A verification methodology for the ensemble prediction system concerns three different subjects:

• prediction of the occurrence of a particular binary event,

• prediction of a probability distribution of a random (one-dimensional) variable
• and representation of the uncertainty about the estimate of the (full-dimensional) model state.

To validate the prediction of an occurrence of the particular binary event $\mathcal{E}$, as for example "the surface pressure is 3hPa smaller than its climatological value",
• the Brier score (the Probability score) and its decomposition,
• the Relative Operating Characteristics (ROC) and the area under ROC and
• the Relative Economic Value are usually used.
They all are based on the *Reliability Diagram* (a relation between forecasted and observed frequencies of the binary event $\mathcal{E}$) and try to produce quantitative summaries from it.

A validation of the prediction of the probability distribution of a random one dimensional variable is closely related to the validation of an occurrence of a particular binary event. The random one-dimensional variable X is described via an enclosing set of binary events $\mathcal{E}_l = \{X \leq x_l\}$, $l = 1, \ldots, L$, with increasing sequence of thresholds $x_0 < x_1 < \ldots x_l < \ldots < x_L$.
bullet Discrete and Continuous Rank Probability Scores, which are generalizations of the Brier Score, are common measures for the validation.
Specific measures based on the ensemble estimate of the probability distribution are reported often in parallel.
• The Rank histogram (Candille and Talagrand, 2005) the utilizes indistinguishability hypothesis (verifying observations should be free from observational error) and
• the Skill Score (Sc) concerns the concept of a prior-posterior probability density in a Bayesian framework (a parametric estimate of the probability density) (Wilson et al., 1999). • the Skill Score (Sc) concerns the concept of a prior-posterior probability density in a Bayesian framework (a parametric estimate of the probability density) (Wilson et al., 1999).

A validation of the representation of the uncertainty about the estimate of the model state is of different nature. In applications concerning "Gaussian" type data assimilation (Hybrid Variational or Ensemble Kalman Filter data assimilation) the uncertainty about the point estimate of the model state is represented through the variance-covariance matrix of the forecast error. In that case the validation should primarily reflect how well the ensemble is able to span a subspace essential for the dynamical development of the variance-covariance matrix. The diagnostics should reflect

(1) the dynamical consistency and dominant scales of the variability
(spectral density
and horizontal and vertical cross-correlations within
and between model state components),
(2) the spread-skill relationship
(the "spread-skill" plot,
the resolved range of innovations variance,
Perturbation versus Error Correlation Analysis (PECA)) and
(3) the span of the dynamically unstable directions
(fastest growth of perturbation energy in the ensemble space,
perturbation correlation with the Eady index,
E-dimension).

In the following we will concentrate on a description of some of the validation techniques mentioned above and outline some basic relationships between them.

5.1. **Prediction of the occurrence of a particular binary event.**
Let M denote the total amount of realizations of a prediction of a particular event of interest, $\mathcal{E}$, over which the verification is performed, let $r_{i,j} = \{0, 1\}$, $1 \le i \le N$, $j = 1, \dots, M$, denote the prediction of the event $\mathcal{E}$ by the ensemble member i during the realization j and let $o_j = \{0, 1\}$, $1 \le j \le M$, denote the observation of the event $\mathcal{E}$ during the realization j. The random quantities $r_{i,j}$ and o_j are equal 1 if $\mathcal{E}$ occurs and equal 0, otherwise.

The predicted probability of the occurrence of $\mathcal{E}$ in the realization j, p_j, is

$$p_j = \frac{1}{N} \sum_{i=1}^{N} r_{i,j}.$$

For each realization p_j is a discrete random variable with $N + 1$ possible outcomes, i.e. $p_j = \pi_k$, $\pi_k := \{0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N}{N}\}$ for each $j = 1, \dots, M$. The distribution of p_j will depend on the ensemble size and the predictability of the event. Let $g_k = g(\pi_k) := \frac{1}{M} \sum_{j=1}^{M} \mathcal{I}_{p_j = \pi_k}$, $0 \le k \le N$, denote the frequency of the occurrence of the outcome π_k. Here $\mathcal{I}_{\mathcal{A}}$ is the index of a random event A: $\mathcal{I}_{\mathcal{A}}$ is 1 if $\mathcal{A}$ is true, and $\mathcal{I}_{\mathcal{A}}$ is 0 if $\mathcal{A}$ is false. In the similar way let us denote o_k, $0 \le k \le N$, the frequency with which the event $\mathcal{E}$ indeed occurs in different realizations of the prediction system when it is predicted by the system with probability of the occurrence π_k, $o_k = o(\pi_k) := \frac{1}{M} \sum_{j=1}^{M} o_j \mathcal{I}_{p_j = \pi_k}$.

• The *Reliability Diagram*
is a plot of o_k against π_k. The histogram of the probability realizations $p_j, j = 1, \dots, M$ or the frequency g_k, $k = 0, \dots, N$ is reported as well. This information is the complete representation of the performance of the ensemble prediction system in predicting the event $\mathcal{E}$.

Several quantitative measures for the performance of the Reliability Diagram are proposed in the literature.

• *Brier score* (Brier, 1950)
is defined as

$$\mathcal{B} = \frac{1}{M} \sum_{j=1}^{M} (p_j - o_j)^2 = E(p - o)^2 = E(E(p - o)^2 \mid p) \tag{44}$$

where p and o are random variables describing the predicted probability of occurrence and the occurrence itself of the event $\mathcal{E}$. The better is the prediction system the lower is Brier score.

Murphy 1973 proposed a decomposition of the Brier score into three informative components. Let $g(p)$ and $o(p)$ denote density functions which describe corresponding relationship between the frequencies of occurrence g_k or o_k and the outcomes of the predicted probability π_k, $0 \le k \le N$. Let o_c denote the climatological frequency of the occurrence of the $\mathcal{E}$, $o_c = \sum_{k=0}^{N} o_k g_k$. Then Murphy decomposition reads

$$\begin{aligned} \mathcal{B} =& E_p((p - o(p))^2) + E_p(o(p)(1 - o(p))) \\ =& \int_0^1 g(p)(p - o(p))^2 dp - \int_0^1 g(p)(o(p) - o_c)^2 dp + o_c(1 - o_c) \\ =& \sum_{k=0}^{N} g_k(\pi_k - o_k)^2 - \sum_{k=0}^{M} g_k(o_k - o_c)^2 + o_c(1 - o_c) \end{aligned} \tag{45}$$

The first two terms in the Murphy decomposition of Brier score $\mathcal{B}$ characterize the prediction system. The first term in the decomposition is called a *reliability* (BS_{rel}) and it measures the statistical consistency between the predicted and the observed frequencies of occurrence of $\mathcal{E}$. The second term (BS_{res}) is called a *resolution* and it implies that the predicted probability should be case dependent (different from the climatological one). The third term is called *uncertainty* (BS_s) and it depends on the nature of the event $\mathcal{E}$ but not on the prediction system.

Sometimes *Relative Brier Score* (BSS) and its decomposition are reported instead.

$$BSS = \frac{\mathcal{B}}{BS_s} = BS_{rel} + BS_{res} = \frac{E_p(p - o(p))^2}{o_c(1-o_c)} + 1 - \frac{E_p(o(p) - o_c)^2}{o_c(1-o_c)} \tag{46}$$

Increasing the forecast lead time BSS_{rel} grows and BSS_{res} decreases. The development with time of the total BSS will depend on the relative impact of both components.

• *Relative Operating Characteristics* or Receiver Operating Characteristics (ROC)
is an alternative qualitative measure of the performance of the EPS and it is closely related with the Reliability Diagram (Mason and Graham, 2002). The ROC curve is a plot of Hit rate against False-alarm rate varying the strategy of prediction of the event $\mathcal{E}$ by the ensemble system. For the ensemble system of size N there are $N+2$ different prediction strategies of the event $\mathcal{E}$: predict if at least l, $1 \le l \le N$, ensemble members predict the event ($\mathcal{D}_l$) and two degenerate prediction strategies, namely never predict ($\mathcal{D}_0$) and always predict ($\mathcal{D}_{N+1}$). For the selected informative prediction strategy $\mathcal{D}_l$, $l = 1, \ldots, N$, the jth realization of the EPS will predict event $\mathcal{E}$ if and only if $p_j \ge \pi_l$. Let $m_o = \frac{1}{M}\sum_{j=1}^{M} o_j$ denote the total amount realizations when the event $\mathcal{E}$ has occurred, $m_1^l = \frac{1}{M}\sum_{j=1}^{M} o_j \mathcal{I}_{p_j \ge \pi_l}$ denote the total amount of realizations when the event $\mathcal{E}$ was predicted when it has indeed occurred, $m_l^0 = \frac{1}{M}\sum_{j=1}^{M}(1-o_j)\mathcal{I}_{p_j \ge \pi_l}$ denote the total amount of realizations when the random event was predicted but has not occurred and $m_l^{-1} = \frac{1}{M}\sum_{j=1}^{M} o_j \mathcal{I}_{p_j \le \pi_l}$ denote the total amount of realizations when the random event was not predicted but has indeed occurred. For the degenerate prediction strategies we have $(m_0^1, m_0^0, m_0^{-1}) = (0, 0, m_o)$ and $(m_{N+1}^1, m_{N+1}^0, m_{N+1}^{-1}) = (m_o, M - m_o, 0)$.

Then the *Hit rate* H_l and the *False-alarm rate* F_l corresponding to the prediction strategy $\mathcal{D}_l$, $l = 0, \ldots, N+1$ are defined as follows

$$H_l = \frac{m_l^1}{m_o}, \quad F_l = \frac{m_l^0}{M - m_o}.$$

Plotting H_l against F_l for all $l = 0, \ldots, N+1$ we obtain *ROC curve*. The closer comes the curve to the left upper corner ($(H, F) = (1, 0)$ - "the most hit the least false alarms") the better is the EPS system predicting $\mathcal{E}$.

• *The area under ROC*
is often is reported as another quantitative measure of the probabilistic skill of the EPS. Under certain conditions, the area under ROC can be used to measure skillfulness of the EPS statistically. Area under ROC, A, can easily be calculated as

$$A = 1 - \frac{m_o(M - m_o)}{\mathcal{F}} \tag{47}$$

where $\mathcal{F}$ is a total number of "inversions" among realizations of the EPS. An "inversion" we call a situation when the predicted probability of any hit $p_{j_{hit}} \geq 0, j_{hit} : o_{j_{hit}} = 1$, is smaller than the predicted probability of any false-alarm $p_{j_{false-alarm}}, j_{false-alarm} : o_{j_{false-alarm}} = 0$. The total amount of all hits with the predicted probabilities lower than each false-alarm can be obtained easily from the ranks of the realizations corresponding to each hit, $r_i,\ i = 1, \ldots, m_o$. The realizations are ordered by decreasing the predicted probability of the random event $\mathcal{E}$.

If all predicted probabilities are different for different realizations, the total number of "inversions" is given by the following formula

$$\mathcal{F} = \sum_{i=1}^{m_o} r_i - \frac{m_o(m_o+1)}{2} \tag{48}$$

If ties are present among p_j, as it is often the case for the ensemble prediction system, a correction for the ties should be done (DeLong *et al.* 1988). In that case the area under ROC is obtained as

$$A_{ties} = 1 - \frac{m_o(M-m_o)}{\mathcal{F}_{ties}} = 1 - \frac{1}{m_o(M-m_o)} \sum_{p=1}^{P} h_p(f_p + \tilde{f}_p) \tag{49}$$

where P is a number of distinct segments among p_j, h_p is a number of hit in segment p, f_p is a number of false-alarms with the predicted probability of the event higher than that associated with segment p and $\tilde{f}_p$ is a number of false-alarms with the predicted probability of the event higher or equal than that associated with segment p.

For large amount of realization the distribution of $\mathcal{F}$ can be well approximated by the Gaussian one:

$$\mathcal{F} \sim \mathcal{N}(a_{\mathcal{F}}, b_{\mathcal{F}}). \tag{50}$$

where $a_{\mathcal{F}} = \frac{m_o(M-m_o)}{2}, b_{\mathcal{F}} = \frac{m_o(M-m_o)(M+1)}{12}$.

The exact distribution of the $\mathcal{F}$ is known as well. It is has the same distribution as the Mann-Whitney U-statistics (Bamber 1975), which is symmetric and is defined via a recurrence formula (Conover, 1973, 1999).

The same distributional theory holds even for $\mathcal{F}_{ties}$ with a certain adjustment for ties (Conover, 1999). For large number of realizations,

$$\mathcal{F}_{ties} \sim \mathcal{N}(a_{\mathcal{F}}^{ties}, b_{\mathcal{F}}^{ties}) \tag{51}$$

where

$$\begin{aligned} a_{\mathcal{F}}^{ties} =& \mathcal{N}(\frac{m_o(M-m_o)}{2}, \\ b_{\mathcal{F}}^{ties} =& \frac{m_o(M-m_o)(M+1)}{12} - \frac{m_o(M-m_o)}{12M(M-1)} \sum_{p=1}^{P} (\tau_p(\tau_p+1))(\tau_p-1)) \end{aligned}$$

The Mann-Whitley U-statistics test differences in central tendencies of two independent samples. High observed value of $\mathcal{F}$ ($\mathcal{F}_{ties}$) will indicate that there is a statistically significant difference between the predicted probabilities of the hit events and the false-alarm events. However, it is important that

realizations of p_j indeed form an independent sample in order to use the distribution of the statistics to quantify the significance.

• *Relative Economic value*
of the EPS is one more way to quantify the performance of the EPS. Relative Economic Value relates Brie Score and Relative Operating Characteristics between themselves. Relative Economic value clearly demonstrates the advantage of EPS systems in comparison with deterministic forecasts (Richardson, 2001).

We say that if the verification event $\mathcal{E}$ occurs, it will cost the user L_u if he have already taken a preventive action which cost C or the user will pay the total cost $L_u + L_a$ if he have not taken the preventive action. The user must select a strategy to take the preventive action or not by minimizing the cost over a large number of case, i.e by minimizing the expected cost. If the user decides to protect himself by taking a preventive action, he will do that always as soon as the event $\mathcal{E}$ is predicted. Accepting the certain prediction strategy $\mathcal{D}_l = (m_l^1, m_l^0, m_l^{-1})$, $0 \leq l \leq N+1$, the expected expense will be

$$\begin{aligned} ee_l &= \frac{m_l^{-1}}{M}(L_a + L_u) + \frac{m_l^0}{M}C + \frac{m_l^1}{M}(C + L_u) \\ &= o_c(L_u + L_a) + L_a(F_l(1 - o_c)\alpha - H_l o_c(1 - \alpha)) \end{aligned} \tag{52}$$

where $\alpha = C/La$ is a quantity which will characterize the user, namely which fraction of the potential avoidable loss L_a the user is prepared to spend on the preventive action, and $o_c = m_o/M$ is a climatological frequency of the event $\mathcal{E}$.

Neglecting information given by the EPS the user would base the selection of strategy on the "climatological frequency" only. The optimal (expected) climatological expense ee_c is

$$ee_c = min\{C + o_c L_u, o_c(L_a + L_u)\}$$

This expression is based on the deterministic strategy to prevent always if the preventive action gives positive gain an average ($C < o_c L_a$), corresponding to the degenerate prediction strategy $\mathcal{D}_{N+1}$ ($H_{N+1} = 1, F_{l+1} = 1$), and to prevent never if the preventive action costs on average too much ($C \geq o_c L_a$), corresponding to the degenerate prediction strategy $\mathcal{D}_0$ ($H_0 = 0, F_0 = 0$).

The Relative economic value V_l, dependent on the prediction strategy $\mathcal{D}_l$, is the reduction in the expected expense due to the EPS in proportion to the reduction in the expected expense due a the perfect forecast,

$$V_l = \frac{ee_c - ee_l}{ee_c - ee_{perf}} \tag{53}$$

where $ee_{perf} = o_c(C + L_u)$ is the expected expense based on the perfect forecast ($H_{perf} = 1, F_{perf} = 0$).

It is possible to show that V_l achieves the maximum value always for $\alpha = o_c$ and that the maximal value depends on the selected prediction strategy and that it is equal to the Kiupers score KS_l of the forecast

$$V_{max,l} = H_l - F_l = KS_l$$

The Relative Economic value is positive only for a range of users, namely

$$\frac{m_l^{-1}}{M - m_l^1 - m_l^0} < \alpha < \frac{m_l^1}{m_l^1 + m_l^0} \tag{54}$$

Evaluating the EPS system from the perspective of the user (α), the user must select his optimal prediction strategy which will give the largest Relative Economic value $V^{opt}(\alpha) = V_{l^{opt}(\alpha)}(\alpha)$. How close $V^{opt}(\alpha)$ is to its the maximal value $KS_{l^{opt}(\alpha)}$ depends on how close α is to the climatological frequency o_c.

• The *Overall Economic value*
is the expected expense over all users. Let the population of users be described by a density function $u(\alpha)$, $0 < \alpha < 1$. All users which are predicted to gain (in the long run) from the preventive action $((\alpha < p) \equiv (C < pL_a))$ will take the preventive action and all user which are predicted to lose in long run on preventing $(\alpha \geq p) \equiv (C \geq p_{L_a})$ will take all cost if the event will happen indeed. The overall economic expense for a deterministic forecast p will be

$$ee_F(p) = \int_0^p \alpha u(\alpha) d\alpha + o(p) \int_p^1 u(\alpha) d\alpha.$$

Here we assume that $L_u = 0$ for simplicity. The Overall Economic value for the EPS, $ee_F = \int_0^1 ee_F(p) g(p) dp$, can be decomposed into following components

$$\begin{aligned} ee_F &= \int_0^1 g(p) \int_0^p (\alpha - o(p)) u(\alpha) d\alpha dp + o_c \\ &= ee_c + \int_0^1 g(p) \int_{o(p)}^p (\alpha - o(p)) u(\alpha) d\alpha dp - \int_0^1 g(p) \int_{o_c}^{o(p)} (o(p) - \alpha) u(\alpha) d\alpha dp. \end{aligned} \tag{55}$$

For a finite size ensemble,

$$ee_F = \sum_{k=0}^N g_k \int_{o_k}^{\pi_k} (\alpha - o_k) u(\alpha) d\alpha - \sum_{k=0}^N g_k \int_{o_c}^{o_k} (o_k - \alpha) u(\alpha) d\alpha + ee_c$$

For the uniform distribution of users, $u(\alpha) \sim \mathcal{U}(0,1)$, the overall expected Economic value ee_F becomes directly related to the Brier score.

$$\begin{aligned} ee_F &= \frac{1}{2} \sum_{k=0}^N g_k (\pi_k - o_k)^2 - \frac{1}{2} \sum_{k=0}^N g_k (o_c - o_k)^2 + \frac{1}{2} o_c (1 - o_c) + \frac{1}{2} o_c \\ &= \frac{1}{2} (BS_{rel} - BS_{res} + BS_s) + \frac{1}{2} o_c \end{aligned} \tag{56}$$

In other words, the Brier score is essentially the Overall Economical Value for users without preference.

5.2. **Prediction of the probability distribution of a scalar random variable.**

• *The Ranked Probability Score*

(RPS) is a generalization of the Brier score and is used to evaluate the probabilistic ensemble prediction of a scalar variable, x. The Ranked Probability Score can be applied in Discretized (DRPS) or in Continuous (CRPS) form dependent on the support of the scalar variable (a discrete or a continuous one). For evaluation of an ensemble prediction system from the perspective of a scalar variable the Relative Ranked Probability Score (RPS_s) and it decomposition into normalized reliability (RPS_{rel}) and normalized resolution score (RPS_{res}) can be applied. An extensive discussion on these scores is given in Candille and Talagrand (2005). For example, in discrete form the $DRPS_s$, $DRPS_{rel}$ and $DRPS_{res}$ are defined as follows

$$
\begin{aligned}
DRPS =& \frac{1}{L}\sum_{l=1}^{L}\mathcal{B}(E_l) = DPRS_{rel} + DPRS_{res} \\
unc =& \frac{1}{L}\sum_{l=1}^{L} p_{lc}(1-p_{lc}) \\
DRPS_s =& 1 - \frac{DRPS}{unc} \\
DRPS_{rel} =& \frac{\frac{1}{L}\sum_{l=1}^{L} E(p_l - o(p_l))^2}{unc} \\
DRPS_{res} =& 1 - \frac{\frac{1}{L}\sum_{l=1}^{L} E(o(p_l) - p_{cl})^2}{unc}
\end{aligned}
\tag{57}
$$

where $\mathcal{E}_l = \{x \leq \xi_l\}$, $\xi_1 < \xi_2 < \ldots \xi_L$, $l = 1, \ldots, L$ is a sequence of events, $\mathcal{B}(\mathcal{E})$ is the Brier score score for the probabilistic prediction of event $\mathcal{E}$ and p_{lc} is a climatological frequency for the occurrence of event $\mathcal{E}_l$.

• *The Continuous Ranked Probability Score* (CRPS)

is obtained by transforming a finite sum over thresholds in DRPS (eqn. 57) into an integral over x.

$$
CRPS = \frac{1}{M}\sum_{j=1}^{M}\int (F_j(\xi) - H(\xi - x_{obs,j}))^2 d\mu(\xi) = \int \mathcal{B}(\mathcal{E}_\xi) d\mu(\xi) \tag{58}
$$

where $\mathcal{B}(\mathcal{E}_\xi)$ is the Brier score for the event $\mathcal{E}_\xi = \{x \leq \xi\}$, $F_j(\xi)$ is the jth realization of the probability distribution of the scalar variable x, $F_j(\xi) = P(\mathcal{E}_\xi) = \frac{1}{N}\sum_{k=1}^{N}\mathcal{H}(\xi - x_{kj})$, x_{kj} is the prediction of the random variable x by ensemble member k in the realization j of the EPS, $\mathcal{H}(y)$ is the Heaoyiside function ($\mathcal{H}(y) = 0$ if $y < 0$, $\mathcal{H}_y = 1$ if $y > 0$) and $d\mu(\xi)$ is a measure with which the integration is performed.

The Continuous Ranked Probability Score can be decomposed into similar components as the Discrete Ranked Probability Score

$$
CRPS = Reli - Resol + U \tag{59}
$$

The uncertainty component $U = \int F_c(\xi)(1 - F_c(\xi))d\mu(\xi)$ is corresponding to unc in (eqn. 57). Here $F_c(\xi)$ defines a climatological probability function of the random variable x. Components $Reli$ and $Resol$ correspond to the reliability and resolution components in the Brier Score decomposition.

Candille-Talagrand (Candille and Talagrand (2005)) and Hersbach-Lalaurette (Hersbach (2000)) are two different decompositions of the $CRPS$ into *Reli* and *Resol* components.

• *The Rank Histogram*

is another measure of the probabilistic ensemble prediction of a scalar variable x. It measures whether the verifying observation $x_{obs,j}$ (neglecting the observation error) is statistically indistinguishable from the N ensemble members $x_{i,j}$, $i = 1, \ldots, N$, $j = 1, \ldots, M$. The Rank Histogram is defined as follows. It consists of $N + 1$ bins s_k:

$$s_k = \sum_{j=1}^{M} \mathcal{I}_{\{x_{[k-1],j} \leq x_{obs,j} < x_{[k],j}}, \quad k = 1, \ldots, N+1 \tag{60}$$

where $-\infty = x_{[0]} < x_{[1]} < \ldots < x_{[k]} < \ldots < x_{[N]} < x_{[N+1]} = +\infty$ are order statistics of the j-th realization of the ensemble prediction system. The flatter is the Histogram, the more reliable is ensemble prediction system.

The quantity

$$\delta = \frac{N+1}{MN} \sum_{k=1}^{N+1} (s_k - \frac{M}{N+1})^2 \tag{61}$$

measures the deviation of the histogram from a flat one. A value of δ which is much larger than 1 means that the ensemble prediction system is unreliable. Very small values of δ indicate that observations are not random or not independent.

• *The Skill Score*

is one more measure of the EPS based on the ensemble estimate of the probability distribution of a random scalar variable. Let $\mathcal{E}_{\Delta X}$ denote a verifying event

$$\mathcal{E}_{\Delta X} = \{ \mid X - x_{obs} \mid \leq \Delta X \}.$$

Let $P_c(\mathcal{E}_{\Delta X})$ denote the climatological probability of the event $\mathcal{E}_{\Delta X}$ and $P_{ens}(\mathcal{E}_{\Delta X})$ denote the one estimated from the ensemble. From the Bayesian perspective they will correspond to the prior and the posterior distribution of the event respectively.

The Skill Score gives a measure of $P_{ens}(\mathcal{E}_{\Delta X})$ relative to $P_c(\mathcal{E}_{\Delta X})$:

$$Sc = \frac{P_{ens}(\mathcal{E}_{\Delta X}) - P_c(\mathcal{E}_{\Delta X})}{1 - P_c(\mathcal{E}_{\Delta X})} \tag{62}$$

Under this scoring measure the EPS is considered to be skillful in predicting event $\mathcal{E}_{\Delta X}$ if the posterior probability of the event is larger than the prior one, in other words Sc is positive for the skillful EPS. The Skill score is positively oriented and it is sensitive to the location and sharpness of the ensemble estimated distribution with respect to the verifying observation. The Skill Score takes into account differences in the predictability of the event too, because it is sensitive to the sharpness (and location) of the climatological probability as well.

Because the ensemble size in EPS is usually small, the $P_{ens}(\mathcal{E}_{\Delta X})$ should be estimated parametrically. The prediction of geopotential height and surface and upper-air temperatures is found to obey a normal distribution, the prediction of precipitation obeys a Gamma or a Kappa distribution, the prediction of wind obeys a Weibull distribution, the prediction of cloud cover obeys a Beta distribution and the prediction of the visibility is assumed to obey a lognormal distribution.

5.3. **Representation of the uncertainty about the model state estimate.**
When an ensemble system is used for "Gaussian" type data assimilation purposes, the main aim of the ensemble is to adequately represent the initial uncertainty about the model state and its development. For the "Gaussian" type data assimilation systems the covariance matrix of the forecast error is considered to capture the uncertainty about the estimate of the model state. A successful ensemble should not only capture the time-and-space dependent variation of the spread of the distribution of different model state components, but reflect also the flow-dependent cross-correlations between model state components. The best verification tool is to run a period of data assimilation cycles: successful ensemble should improve the efficiency in assimilating observations. However, it is a real challenge to construct the proper estimate of the forecast error covariance matrix when the dimensionality of the model state is so high and the size of the ensemble is so small. A number of diagnostics and verification tools can highlight specific features of estimates of the forecast error covariance matrix.

• *Diagnostic plots*
of the spectral variance of different model state components as well as plots of the cross-correlations between model state components are very useful, investigating the dynamical consistency of the forecast error covariance estimate.

• The *spread-skill relationship*
is considered to be an important characteristic of the ensemble system. There is not any unique quantity which would summarize the spread of the ensemble.

The *spread-skill relationship plot* is a plot root-mean-square error of the ensemble mean $E_m = (P_m^T P_m)_S^{1/2}$ and the estimate of the ensemble spread $E_{sp} = (\frac{1}{N}\sum_{i=1}^{N}(P_i^T P_i)_S)^{1/2}$. Here $P_m = \bar{x} - x_a$ and $P_i = X_i - x_m$, where X_i denote the ensemble members i, $1 \leq i \leq N$, $\bar{x}$ denotes the ensemble mean, x_a denotes the verifying analysis and subscript S denotes the norm. The total energy norm is often used (eqn. 39). The correlation between spread and skill is related with the magnitude of the spread variability, namely the more the spread departs from its climatological mean value, the more useful is the spread as a predictor of skill (Whitaker and Loughe (1998)). *The resolved range of innovation variance* is one more way to measure spread-skill relationship of the ensemble (Wang and Bishop (2003)). First, a scatterplot of squared innovations against ensemble estimate of the variance in observation space is constructed using all observation quantities. After that a relationship (regression) of the innovation variance on the ensemble variance is tried. For the perfect ensemble, when the observation error, model error and forecast error are mutually uncorrelated, the relationship should look like as a stright line with 45 degrees slope, because the innovation variance is a sum of the forecast and observation variance (taken in observation space). Although in reality the representativity error and the model error may be correlated with the forecast error, still observing the resolved range of the innovation variance can tell us something about how well the spread of ensemble represent the skill of the ensemble can be made.

Wei and Toth (2003) propose another measure to quantify skill-spread relationship called *Perturbations versus Error Correlation Analysis* (PECA). First the optimal linear combination of ensemble perturbations P_{opt}, which gives the best prediction of the forecast error P_m in L_2 norm, is obtained:

$$P_{opt} = \sum_{i=1}^{N} \alpha_i P_i,$$

where the weights α_i, $1 \leq i \leq N$, are obtained by solving the least-square problem

$$min \mid P_m - P_{opt} \mid_{L_2} \tag{63}$$

The PECA is defined as a pattern anomaly correlation between the forecast error P_m and the optimal perturbation P_{opt}

$$PAC(P_m, P_{opt}) = \frac{cov(P_m, P_{opt})}{var(P_m)^{1/2} var(P_{opt})^{1/2}} \tag{64}$$

The successful ensemble should explain the largest part of the forecast error variance via the optimal perturbation. • *Sampling of dynamically unstable directions*
is one more important characteristics of the ensemble in representing the forecast error variance-covariance. The correlation of the optimal perturbation with the Eady index (Hoskin and Valdes (1990)), the fastest growth of the energy of perturbations in the space spanned by the ensemble in the total energy norm and the E-dimension (Oszkowski et.al, 2005) can be used as measures of the ensemble performance.

The estimation of *the fastest energy growth* is given in the section on Singular vectors (eqn. 43).

The *Eady index* expresses the maximum normal mode error growth rate in baroclinic developments and it is defined as

$$\sigma_{Eady} = 0.3125 \frac{f}{N_b} \frac{du}{dz} \tag{65}$$

where f is the Coriolis parameter, N_b is the buoyancy frequency and $\frac{du}{dz}$ is the vertical wind shear. Interpreting the correlation one should remember that the Eady index corresponds to instabilities caused by baroclinic development only.

The Ensemble dimension, also known as *The Bred Vector dimension*, measures the effective dimension spanned by a N-dimensional ensemble in a local geographical region at a particular time. The E-dimension characterizes the effective number of dominant directions in the vector space spanned by the ensemble perturbations and is defined as

$$E_{dim}(\lambda_1, \ldots, \lambda_N) = \frac{(\sum_{i=1}^{N} \sqrt{\lambda_i})^2}{\sum_{i=1}^{N} \lambda_i} \tag{66}$$

where λ_i, $i = 1, \ldots, N$ are eigenvalues of the local forecast error covariance matrix $B_L = (P_L^T P_L)_S$. Here subscript S denotes the total energy norm and P_L denotes the ensemble of local perturbations $P_L = (P_{L,1}, P_{L,2}, \ldots, P_{L,N})$. The local perturbation $P_{L,i}$ contains perturbations of all dynamical variables $X_{j,i} - \bar{x}_j$ of the global perturbation P_i, belonging to the local area L. Small values of E_{dim}, $1 \leq E_{dim} << N$, reflect presence of a few leading directions of variability (the remaining ones are small compared to the leading ones) in the local area and large values of the E_{dim}. The large values of E_{dim}, $E_{dim} \approx N$, reflect the nearly equal spread of the variability among all direction, what can correspond to noise. Oszkowski et al. (2005) point out that a number of atmospheric scenarios, such as pure baroclinic instabilities, complex processes involving baroclinic and barotropic instabilities, the divergence of ageostrophic geopotential fluxes etc. result in a low E-dimension.

6. Conclusions

The basic purposes of the current literature study could be summarized as follows:

- to investigate the theoretical relationships between the sequential and variational data assimilation schemes especially in a perspective of the practical implementation;
- to investigate the relationships between the ensemble data assimilations and the ensemble prediction systems.
- in addition we have provided a summary of widely used verification methods to qualify the performance of the ensemble prediction and ensemble data assimilation systems.

The variational data assimilation methods, such as the Three-Dimensional and the Four-Dimensional Variational Data Assimilation Schemes, are successfully implemented worldwide by the weather services for the operational weather prediction. At the same time, there is a lot of space for further improvements.

The Variational Data Assimilation schemes compute a posterior mode of the probability density function, valid at the beginning of the data assimilation window, maximising the correspondent density function numerically. The full rank background error covariance matrix is necessary for the procedure of the numerical maximisation. The number of assimilated observation is of several magnitudes smaller that the dimensionality of the model state. Thus the prior assumptions about the probability density function at the beginning of the data assimilation window, expressed via the background forecast and the background forecast error covariance matrix, will have strong impact on the posterior mode. The Variational Data Assimilation Schemes lack affordable procedure for the explicit update of the evolution of the forecast/analysis/forecast-at-the-begging-of-the-next-assimilation-window error covariances. The static constant covariance is used at the beginning of each assimilation window, what degrade the performance of the variational assimilation schemes.

The sequential data assimilations methods based on the Kalman Filter recursions would provide the Gaussian approximation to the posterior probability density function valid at the end of the data assimilation window. However, the Kalman Filter recursions require an explicit forward propagation of the forecast error covariance. The various implementations of the Ensemble Kalman Filter were proposed in order to afford the practical implementation of the sequential methods, where the forecast error covariance matrix in propagated forward approximately. First, a number of model states, an ensemble, is selected so that they together would represent covariance matrix at the initial time; each model state is propagated forward in time using the forward model propagator; the forecast error covariance at the time of interest is estimated from this propagated ensemble; the propagated forecast ensemble is rescaled/resampled into the analyses ensemble that should represent the update of the uncertainty about the model state after new observations are assimilated. However, the rank deficiency of the model state covariance estimate from the ensemble with a limited size creates serious problems implementing Ensemble Kalman Filter data assimilation schemes.

Theoretically, the Hybrid Ensemble-Variational data assimilation scheme, where the the background covariance matrix is modelled via contributions from both the full-rank static constant covariance matrix and the rank-deficient flow-dependent covariance matrix estimated from the ensemble of the forecast state, could be an alternative method. However, it still should be proven that involving the rank-deficient estimate of the forecast error covariance matrix would not degrade the performance of the variational assimilation scheme. The Schur-product of matrices is traditionally used to increase the rank of the ensemble estimate of the forecast error covariance matrix. It is important to notice that a Schur-product of matrices is not a linear transformation, therefore an extreme care should

be taken on investigating impact of the Schur-product on the physical balances between the model state components. In the long term perspective, the flow-dependent forecast error covariance matrix should be modelled using local structures, such as wavelets or Intrinsic Mode Functions, with only the parameters estimated from the ensemble.

The Data Assimilation Schemes provides a point estimate of the model state with a possibility to quantify uncertainty about the estimate. The Ensemble Prediction System are essentially different from the Ensembles Methods for Data assimilation and they are constructed with the aim to provide a probabilistic inference about some phenomena of interest (a low-dimensional transform of the model state variable) during and after a certain integration period. For the long range forecasts, the Lyapunov exponent based techniques, such as singular vectors and breeding vectors, are traditionally used. For the shorter forecast range the initial model state will have an impact on the probabilistic inference during and after integration period. Thus Ensemble Prediction Systems, which adequately represent the probability distribution of the initial model state, are required. Sampling initial uncertainty, the ETKF based rescaling scheme outperforms Breeding and Singular Vectors techniques. The Hybrid ETKF-PF (Particle Filters) Scheme has a potential to become an efficient ensemble prediction system for the short-moderate range forecasts.

The choice of the verification methods is important quantifying the performance of the Ensemble Prediction Systems. The majority of the commonly used verification tools such as the Brier Score, Relative Operating Characteristics and area under ROC as well as the Relative Economic Value represent different quantitative measures of the performance of the Reliability diagram for the particular event of interest. It should be taken care when the conclusions are extrapolated on subspace of the model state support not covered by the event. The Overall Economic value can validate the performance of the Ensemble Prediction system for the users with particular preference expressed through a cost-loss ratio.

Acknowledgements. This research is funded by the eVITA program of the Research Council of Norway, under contract 178894. The author is grateful to Nils Gustafsson and Ole Vignes for a thoughtful reading and fruitful comments on this manuscript.

7. References

Andersson, J.L., and S.L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review* **127**, 2741-2758.

Anderson, J.L., 2001: An ensemble adjustment filter for data assimilation. *Monthly Weather review 129*, 2884-2903.

Augustine, S.J., S.L. Mullen, and D.P. Baumhefner, 1992: Examination of actual analysis differences for use in Monte Carlo forecasting. *Proc. 16th Climate Diagnostics Workshop, Los Angeles*, CA, NOAA/NWS/NMC/CPC, 375-378.

Bamber, D., 1975: The area above the ordinal dominance graph and the area below the relative operating characteristics graph. *J. Math. Psychol.*, **12**, 387-415.

Barkmejer, J., R. Buizza and T,N. Palmer, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.* **125**, 2333-2351.

Beck, A., and M. Ehrendorfer, 2005: Singular-Vector-Based covariance Propagation in a Quasi-geosprophyc Assimilation System. *Mon. Wea. Rev.,* **133**, 1295-1310.

Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *J. Atmospheric Research* **108**(D24), 8775, doi:10.1029/2002JD002900.

Bishop, C.H., B.J. Etherton and S.J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, **129**, 420-436.

Buehner, M., 2005. Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Q.J.R. Meteorol. Soc.*, **131**, 1013-1043, 1013-1043.

Buehner, M., and A. Zadra, 2006: Impact of flow-dependent analysis error covariance norm on extra-tropical singular vectors. *Quart. J. Roy. meteor. Soc.*, **132**, 625-646.

Buizza R., Tribba J., Monteni F. and Palmer T., 1993: Computation of optimal unstable structures for a numerical weather prediction. *Tellus*, **45**, 388-407

Burgers, G, P.J. Van Leeuwen, and G. Evensen, 1998 : Analysis scheme in the ensemble Kalman filtering. *Monthly Weather Review* **126**, 1719 -1724.

Candille, G. and O. Talagrand, 2005: Evaluation of Probabilistic Prediction System for a Scalar Variable. *Quart. J. Roy. Meteor. Soc.*, , 1-20.

Chin T.M., M.J. Turmor, and J,B. Jewell, M. Ghil, 2007: an Ensemble-Based Smoother with Retrospectively Updated Weights for Highly Nonlinear Systems. *Monthly Weather Review* **135**, 186-202.

Cohn S.E., and R. Todling, 1996: Approximate data assimilation scheme for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74**, 63-75.

Cohn S.E, A.M. da Silva, J. Guo, M. Sienkiewich, and D. Lamich, 1998: Assesing the effect of data selection with the DAO Physical-space Statistical Analysis System (PSAS). *Monthly Weather review* **126**, 2913-2926.

Conover, W.J., 1973: Rank tests for one sample, two samples and k-samples without assumptions on a continuous distribution function. *Ann. Stat.*, **1**, 1105-1125.

Conover, W.J., 1999: *Practical nonparametric statistics*, Wiley, Chichester, UK.

Courtier P., J.-N. Thepaut and A. Hollngswirth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Q.J. of Royal Meteorolo. Society* **120**, 1367-1387.

Dee, D.P., 1995: On line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.*, **123**, 1128-1196.

Ehrendorfer, M., 1997: Predicting the uncertainty of the numerical weather forecasts. A review. *Meteorol. Z.*, **6**, 147-183.

Errico, R. M., 2000: Interpretation of the total energy and rotational energy norms to determination of singular vectors. *Quart. J. Roy. Meteor. Soc.*, **126**, 1581-1600.

Evensen, E., 1994: Sequential data assimilation with a nonlinear QG model using Monte Carlo methods to forecast error statistics. *J. Geophys Res.* **99**, 10143-10162.

Evensen, G., and P.J. Van Leeuwen, 1996: Assimilation of geosat altimiter data for the Agulhas current using the Ensemble Kalman filter with a quasi-geostropic model. *Monthly Weather review* **124**, 85-96

Frogner, I.-L., and T. Iversen, 2002: High-resolution limited area ensemble prediction based on low-resolution targeted singular vectors. *Quat. J. Roy. Meteor. Soc.*, **128**, 1231-1341.

Gaspari, G., and S. Cohn, 1999: Construction of the correlation function in two and three dimensions. *Q.J.R Meteorolog. Soc.*, **125**

Gauthier, P. and J.-N. Thepaut, 2001: Impact of the Digital filter as a week constraint in the preoperational 4DVAR Assimilation System of Meteo-France. *Mon. Wea. Rev*, **129**, 2089-2102.

Hamill, T.M. and C.Snyder, 2000: A hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme. *Monthly weather Rev.*, **128**, 2905-2919.

Hamill, T.M., C. Snyder and J.S. Whitaker, 2003: Ensemble Forecast and the Properties of Flow-Dependent Analysis-Error Covariance Singular Vectors. *Monthly Weather Review*, **131**, 1741-1758.

Heemink A.W., M. Verlaan, and A.J. Segers, 2001: Variance reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718-1728.

Hersbach, H, 2000: Decomposition of the Continuous Ranked Probability Score fro the Ensemble Prediction System. *Weather and Forecasting*, **15**, 559-570.

Hoskins, B.J., and P.J. Valdes, 1990: On the existence of storm tracks. *J. Atmospheric Science*, **42**, 1854-1864.

Houtekamer , P.L. and H.L.Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review* **126**, 796-811.

Houtekamer, P.L. and H.L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review* **129**, 123-137.

Kalman, R. 1960: A new approach to linear filter and prediction theory. *J. Basic. Engr.* **82D**, 35-45

Kalman, R and R. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic Engr.* **83D**, 95-108.

Kim, S., G.L. Eyink, J.M. Restrepo, F.J. Alexander, and G. Johnson, 2003: Ensemble filtering for non-linear dynamics. *Monthly Weather review* **131**, 2586-2594.

Le Dimet, F.X. and O. Talagrand, 1986: Variational Algoritm for Analysis and Assimilation of Meteorological Observations. Theoretical aspects. *Tellus* **38A**, 97-110.

van Leeuwen, P.J., 2003: A variance-minimizing filter for large-scale applications. *Monthly Weather Review* **131**, 2071-2084.

Lorenc A.C., 2003: The potential of the ensemble Kalman Filter for NWP- A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183-3203.

Lynch, P. and X.Y.Huang, 1992: Initialisation of the HIRLAM model using digital filter. *Mon. Wea. Rev*, **120**, 1019-1034.

Mason, S.J. and N.E. Graham, 2002: Areas beneath the Relative Operating Characteristics (ROC) and Relative Operating Levels (ROL) curves: Statistical Significance and Interpretation. *Q.J. Roy. Meteor. Soc.*, **128**, 2145-2166.

Molteni, F, Buizza R., Palmer T.N. and Petroliagis T. , 1996: The ECMWF ensemble prediction systems: methodology and validation. *Q. J. of the Roual Meteorol. Soc.*, **122**, 73-119.

Murphy, A.H., 1973: A New Vector partition of the Probability Score. *J. of Applied Meteorology*, **12**, 595-600.

Oczkowski, M., I. Szunyong and D.J. Patil, 2005: Mechanizm for the development of Locally Low-Demensional Atmospheric Dynamics. *J. Atmospheric Sciences* **62**, 1135-1156.

Ott, E. and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415-428.

Parrish, D. F. and J.C. Derber, 1992: The National Meteorological Center's statistical-interpolation analysis system. *Monthly Weather Review* **120**, 1747-1763.

Pham, D.T., 1997: Dimensions, predictability and reduced-rank Kalman filtering in data assimilation. *Proc. Third Bilateral French-Russian Conf.: Predictabilty of Atmospheric and Oceanic Circulations, Nancy, France, (INRA-Miscow State University)*

Pham, D.T. , J. Verron and M.C. Roubaud, 1998b: A singular evolutive extended Kalman filter for data assimilation in oceanography. . *Mar.Syst.*, **16**, 323-340.

Pham, D.T., 2001: Stochastic methods for sequential data assimilation in strongly non-linear systems. *Monthly Weather Review* **129**, 1194-1207.

Richardson, D.S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Q. J. Roy. Meteor. Soc.*, **127**, 2473-2489.

Sakov, P., and P.R. Oke, 2008: Implification of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, **136**, 1042-1053.

Talagrand, 0., 1981: A study of the dynamics of four-dimensional data assimilation. *Tellus*, **33**, 43-60.

Toth Z. and Kalnay E., 1993 : Ensemble forecasting at NMC: The generation of Perturbations. *Bulletin of the American Meteorol. Soc.*, **74**, 2317-2330.

Toth Z. and Kalmay E., 1997: Ensemble Forecasting at NCEP and the Breeding Method. *Monthly Weather review*, **125**, 3297-3319.

Toth, Z., O. Talagrand, G. Candille and Y. Zhu, 2003 : Probability and Ensemble Forecasts. /textit-Forecast Verification. A practitioner's Guide in Atmospheric Science, I. Jollifee and D.B. Stephenson, editors, John Wiley & Son, Ltd, Chichester, United Kingdom, ISBN:0-471-49759-2, 137-163.

Trevisan, A. and R. Legnani, 1995: Transient error growth and local predictability: A study in the Lorenz system, *Tellus*, **47A**, 103-117.

Verlaan M., and A.W. Heemink, 1997: Tidal flow forecasting using reduced-rank square root filters. *Stochastic Hydrological Hydraulics,* **11**, 349-368.

Wang, X., C.H. Bishop, 2003: A comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecasting Schemes. *J. Atmospheric Sciences* **60**, 1140-1158.

Wang, X., C.H Bishop and S.J. Julier, 2004: Which is Better, an Ensemble of Positive- negative Paairs or a Centered Spherical Simplex Ensemble ? *Monthly Weather Review*, **132**, 1590-1605.

Wang, X., T.M. Hamill, J.S. Whitaker and C.H. Bishop, 2007: A comparison of Hybrid Ensemble transform Kalman Filter -Optimum Interpolation and ensemble Square Root Filter Analysis Schemes. *Month. Weather Rev.*, **135**, 1055-1076.

Wang, X., C. Snyder and T.M. Hamill, 2007: On the theoretical equivalence of Different Proposed Ensemble-3DVar Hybrid Analysis Schemes. *Monthly Weather Rev.*, **135**, 222-227.

Wei, M., and Z. Toth, 2003: A New Measure of Ensemble Performance: Perturbation versus Error Correlation analysis (PECA). *Monthly Weather Review*, **131**, 1549-1565.

Wilson, L.J., W.R. Burrows and A. Lanzinger, 1999: A strategy for Verification of Weather Element Forecast from an Ensemble Prediction System. *Monthly Weather Rev.*, **127**, 956-970.

Whitaker, J., and T.M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Monthly Weather review* **130**, 1913-1924.

Whitaker, J., and A.E. Loughe, 1998: The relationship between Ensemble Spread and Ensemble Mean Skill. *Monthly Weather Review*, **126**, 3292-3302.