THE KALMAN FILTER

F Bouttier
Météo-France, Centre national de Recherches Météorologiques[1]
31057 Toulouse cedex, France

Summary : The Extended Kalman Filter is a sequential, optimal algorithm for data assimilation under some linearity hypotheses. Although it is computationally too expensive to implement for operational weather prediction, it is a useful framework to understand similarities and differences betwen various data assimilation and stochastic prediction techniques. Some practical problems and solutions for future applications are reviewed in this paper.

1. INTRODUCTION : KALMAN FILTER THEORY

1.1 Historical framework

The Kalman filter[2] is an algorithm for data assimilation into numerical models. Its main usefulness is that it provides a clean framework for estimating the uncertainty of the analyses and short-range numerical weather predictions. Roughly speaking, it is a tool for estimating the predictability of the analyses : in that sense, it provides the initial conditions that should be used for the other predictability systems described by other papers in this volume.

The Kalman filter is actually just a linearization of the more difficult problem of stochastic-dynamic prediction (*Epstein,* 1963). It was designed by Kalman (1960) in order to apply the well-known technique of least-squares estimation to dynamical systems. Subsequently it has been used with some success in engineering, e.g. for satellite tracking. On the other hand, it has so far only been applied to numerical weather prediction (NWP) in extremely simplified forms. In operational data assimilation for NWP, the main algorithms used have been successive correction schemes, and the so-called Optimal Interpolation (OI) which is an approximation of a part of the Kalman filter (*Lorenc,* 1986). With the advent of adjoint techniques (*Talagrand and Courtier,* 1987), it becomes possible to replace OI by three-dimensional variational data assimilation (3D-VAR) (*Parrish and Derber,* 1992, *Andersson et al.,* 1994), which in turn should soon be superseded by four-dimensional variational data assimilation (4D-VAR) (*Thépaut et al,* 1991). At the same time, some work has been devoted to the development of algorithms for efficient estimation of the short-range evolution of prediction errors, namely the breeding method (Toth and Kalnay 1993) and the singular vector computation (*Buizza et al,* 1992) ;

[1]Current affiliation : European Centre for Medium-Range Weather Forecasts, Reading, UK.
[2]Also known as "the Kalman-Bucy filter".

those methods are aimed at providing the initial condition for realistic ensemble prediction at the medium range.

As we will explain later, all these algorithms can be regarded as different approximations to the extended Kalman filter, and they are to a large extent complementary to each other. Hopefully, in the near future it will be possible to merge advanced data assimilation and ensemble prediction systems into a coherent algorithm for the estimation of analysis and prediction error characteristics. An essential ingredient will be the implementation of an efficient approximation to the extended Kalman filter as a part of the operational assimilation schemes (*Fisher,* 1995). The design of such a high-resolution approximation is a major research topic in data assimilation. It relies on the experience provided by years of experiments with low-resolution versions of the Kalman filter in meteorology and oceanography (*Miller,* 1986 ; *Ghil,* 1989 ; *Parrish and Cohn,* 1985).

1.1 The linear Kalman filter

Note that in this paper we are only going to address the *discrete* algorithm. There is also a more difficult, but theoretically useful, definition which is continuous in time and space (*Jaswinski,* 1970).

We shall use the following standard notations (*Ide et al,* 1995) :

x^t	true fluid state
x^a, x^f	analysis and forecast vectors
$\mathbf{P}^a$, $\mathbf{P}^f$	analysis and forecast estimation error covariances
M	forecast model
η	model error, with covariance $\mathbf{Q}$
y^o	vector of observed values
ε	observation error, with covariance $\mathbf{R}$
H	observation operator
$\mathbf{K}$	analysis gain
x^b	background state used in the analysis

The basic hypotheses made about the assimilation system characterize the operators we are going to use for the evolution of the system state (i.e. the model fields) and for the comparison of that model state to the observations of reality:

(i) $\mathrm{x}^t(t_{i+1}) = M(t_{i+1}, t_i)\mathrm{x}^t(t_i) + \eta(t_i)$ defines M as a **linear** dynamical model of state evolution between times t_i and t_{i+1}, which has random errors η of **known covariances** $\overline{\eta(t_i)\eta^{\mathrm{T}}(t_i)} = \mathbf{Q}(t_i)$

(ii) $\mathrm{y}_i^o = H_i\mathrm{x}^t(t_i) + \varepsilon(t_i)$ defines H as a **linear** model of observation at time t_i, which has random errors ε of **known covariances** $\overline{\varepsilon(t_i)\varepsilon^{\mathrm{T}}(t_i)} = \mathbf{R}(t_i)$

(iii) there are **no error biases** : $\overline{\eta} = 0$, $\overline{\varepsilon} = 0$

(iv) there are **no cross-correlations** between model and observation errors, $\overline{\eta\varepsilon} \equiv 0$, nor between model or observation errors at different times.

Note that the overbar is used to denote an *ensemble mean*, i.e. an average at a given time of all possible values of the argument, according to its probability density function.

The linear Kalman filter equations define the optimal way of using the observations and the model in a sequential way, in order to estimate the true atmospheric state :

$$\text{State forecast} \quad \mathrm{x}^f(t_{i+1}) = M(t_{i+1}, t_i)\mathrm{x}^a(t_i) \tag{1}$$

$$\text{Error covariance forecast} \quad \mathbf{P}^f(t_{i+1}) = M(t_{i+1}, t_i)\mathbf{P}^a(t_i)M^{\mathrm{T}}(t_{i+1}, t_i) + \mathbf{Q}(t_i) \tag{2}$$

$$\text{Kalman gain computation} \quad \mathbf{K}_i = \mathbf{P}^f(t_i)H_i^{\mathrm{T}}[H_i\mathbf{P}^f(t_i)H_i^{\mathrm{T}} + \mathbf{R}_i]^{-1} \tag{3}$$

$$\text{State analysis} \quad \mathrm{x}^a(t_i) = \mathrm{x}^f(t_i) + \mathbf{K}_i[\mathrm{y}_i^o - H_i\mathrm{x}^f(t_i)] \tag{4}$$

$$\text{Error covariance of analysis} \quad \mathbf{P}^a(t_i) = [\mathbf{I} - \mathbf{K}_iH_i]\mathbf{P}^f(t_i) \tag{5}$$

The algorithm consists of performing (1) state and (2) covariance forecasts for each model timestep (t_i, t_{i+1}), inserting the corrections (4) and (5) whenever observations are available. In the latter equations, the matrix $\mathbf{K}_i$ is computed using (3).

The derivation of these equations can be found in the literature (e.g. *Ghil,* 1989). To put it in a nutshell, (1) comes straight from the dynamical model definition, (4) is the general form of a linear analysis, (3) is the gain value which minimizes the r.m.s. estimation error of the analysis under the above-mentioned hypotheses (just like in OI), (2) and (5) are just the ensemble means of the deviations from the truth of (1) and (4), respectively, multiplied by their own transposes.

1.2 The extended Kalman filter

The extended Kalman filter, or EKF, is a generalization of the linear Kalman filter for non-linear systems. It is particularly useful for numerical weather prediction, in which the state evolution is strongly nonlinear, still the estimation error evolution can be assumed to be linear to a large extent (*Lacarra and Talagrand,* 1988). More precisely, the EKF is defined using the following operators :

M tangent linear of the non-linear forecast model M
H tangent linear of the non-linear observation operator H

where the linearization is performed in the vicinity of an estimate of the true state x^t at each timestep. The EKF relies on the following *tangent linear hypothesis* to make on top of the linear Kalman filter hypotheses :

(v) the forecast and observation errors are, to a good approximation, given by the linearized operators $(\mathbf{M}, \mathbf{H})$.

In other words, the dynamical and observation models can both be linearized with respect to $\mathrm{x}(t_i)$.

The EKF equations are simply defined from the linear equations by keeping equations (1) and (4) where the models M and H are now nonlinear, and by substituting their linearized counterparts into the other equations :

$$\begin{aligned}
\text{State forecast} \quad & \mathrm{x}^f(t_{i+1}) = M(t_{i+1}, t_i)\mathrm{x}^a(t_i) && (1)\\
\text{Error covariance forecast} \quad & \mathbf{P}^f(t_{i+1}) = \mathbf{M}(t_{i+1}, t_i)\mathbf{P}^a(t_i)\mathbf{M}^{\mathrm{T}}(t_{i+1}, t_i) + \mathbf{Q}(t_i) && (2\mathrm{E})\\
\text{Kalman gain computation} \quad & \mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^{\mathrm{T}}[\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^{\mathrm{T}} + \mathbf{R}_i]^{-1} && (3\mathrm{E})\\
\text{State analysis} \quad & \mathrm{x}^a(t_i) = \mathrm{x}^f(t_i) + \mathbf{K}_i[\mathrm{y}_i^o - H_i\mathrm{x}^f(t_i)] && (4)\\
\text{Error covariance of analysis} \quad & \mathbf{P}^a(t_i) = [\mathbf{I} - \mathbf{K}_i\mathbf{H}_i]\mathbf{P}^f(t_i) && (5\mathrm{E})
\end{aligned}$$

In other words, the EKF manages the evolution of a multidimensional "error bar", the matrix $\mathbf{P}$, in the vicinity of a non-linear trajectory, x. One should note that, in the NWP literature, the phrase "Kalman filter" is often used to denote an algorithm which is actually an EKF.

1.3 Comments on the algorithm

The input to both algorithms is : the definition of the models themselves, the initial condition for $(\mathrm{x}, \mathbf{P})$ (when the filter is started), the sequence of observations y^o, and the sequence of model and observation error covariance matrices $(\mathbf{Q}, \mathbf{R})$. The output is the sequence of estimates $(\mathrm{x}, \mathbf{P})$ of the model state and its estimation error covariance matrix.

A first property of the (linear or extended) Kalman filter is its *sequentiality* : to provide an analysis, the algorithm uses only information from the past, up to the latest observations. That is convenient for real-time applications such as operational NWP ; some other algorithms, like 4D-VAR, are not sequential.

Another property of the Kalman filter is its *optimality* :

- the linear Kalman filter analyses and forecasts are the best linear and unbiased estimates (BLUE) of the true fluid state using all available information from the past. That is true only insofar as the hypotheses for the linear Kalman filter are verified.
- with the same hypotheses and input data, if there is no model error, the optimum of a 4D-VAR analysis provides the same final result as the Kalman filter run over the same time interval. That result shows that the Kalman filter is optimal in the sense of the 4D-VAR cost-function (*Thépaut and Courtier,* 1991 ; *Lorenc,* 1986).
- the EKF algorithm is optimal if the linearization hypotheses are verified exactly. If the models are not linear, the optimality is only approximately true to the extent that the tangent linear hypothesis is verified.

One can see that the EKF is not really optimal. However, experience shows that the tangent linear hypothesis is verified for many interesting NWP problems (*Lacarra and Talagrand,* 1988 ; *Vukicevic,* 1991), which means that the EKF is nearly optimal in those cases. One has just to take steps to ensure that linearity indeed remains a good approximation, which

may not be trivial, as linearity itself tends to depend upon the atmospheric situation and filter characteristics.

Since there are many similarities between 4D-VAR and the EKF, it may be opportune to recall the fundamental differences between them :

- 4D-VAR can be run for assimilation in a realistic NWP framework (*Andersson et al.,* 1994) because it is computationally much cheaper than the EKF.
- 4D-VAR is more optimal than the (linear or extended) KF inside the time interval for optimization because it uses all the observations at once, i.e. it is not sequential, it is a smoother (*Bennett and Budgell,* 1989).
- 4D-VAR relies on the hypothesis that the model is perfect (i.e. $\mathbf{Q} = 0$), whereas (uncorrelated) model errors can be accounted for in the EKF.
- 4D-VAR can only be run for a finite time interval, especially if the dynamical model is non-linear, whereas the EKF can be run forever as soon as the recurrence has been initiated.
- 4D-VAR itself does not provide an estimate of $\mathbf{P}^f$, it has to be extended with specific procedures to replace equations (2E) and (5E) (*Fisher,* 1995).

One should also note that the linearization of a non-linear model, or the use of approximate dynamics (the incremental formulation, see *Courtier et al* (1994)) involve essentially the same hypotheses and algorithmic modifications in 4D-VAR and in the EKF.

1.4 Practical implementation

For realistic implementations, the dynamical and observation models are usually given as operators coded for computers (like their tangent linear and adjoint operators). However, the EKF equations do not allow one to easily avoid building the corresponding explicit matrices to perform the algebraic computations. Even when matrix operators are avoided, a huge computer power is required, as can be seen if one considers a system in which the model has n degrees of freedom (the dimension of x) and p observations (the dimension of y) ; in NWP systems those numbers are quite large[3] :

- $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{K}$ matrices have sizes of order n^2, p^2 or $n \times p$: they cannot be easily stored in computers or data handling systems.
- the first term in the right-hand side of equ. (2E) involves a number of dynamical model integrations which is proportional to n.

[3]In state-of-the-art NWP centres, $n > 10^7$ and $p > 10^5$, and the actual figures are usually 10 to 100 times larger.

- the right-hand side of equ. (3) involves the inversion of a $p \times p$ matrix.
- one should ensure that the $\mathbf{P}$ matrices are always positive definite (this may not always be true due to numerical truncation errors).
- all the coefficients of the $\mathbf{P}$ matrix should remain small enough in absolute value, so that the tangent linear hypothesis remains true when evaluating equ. (2E) which involves all lines and columns of $\mathbf{P}$.

These are the main hurdles which have so far prevented the EKF from being applied to realistic assimilation problems. As of today, the highest-resolution systems for which an EKF has been run used models with $n \simeq 1000$. It is safe to say that no computer improvement will ever allow the use of a genuine Kalman filter algorithm for a state-of-the-art assimilation system, simply because increases in computer power tend to be at least partly used for improvements of the dynamical model itself and to handle more observations. As the computational requirements of Kalman filtering increase like n^2+p^2, the relative discrepancy between the resolution of the operational systems and that of the EKF that can be implemented will only increase with time. The consequence of those numerical cost issues is that there is no point in running an exact implementation of the EKF, except for theoretical research studies. One has to design some efficient approximations to the algorithm.

Another difficulty with the Kalman filter algorithm, which should not be overlooked, is its lack of stability, both for numerical and physical reasons. It is not trivial to keep the positive definiteness of the $\mathbf{P}$ covariance matrix over long periods of time. Depending on the characteristics of the model dynamics and observing network, some error covariances may also grow indefinitely or vanish, which means that the EKF equations have to be somewhat modified in order to enforce some robustness and physical sense, as has been documented in simplified experiments (*Bouttier,* 1994 ; *Evenssen,* 1992).

It is possible to rewrite the EKF equations in order to partly solve those problems. A comprehensive description of the possible formulations and of their implications is out of the scope of this paper. Some useful information can be found in the papers mentioned in the bibliography, and in references therein. The main ideas are the following :

- 4D-VAR assimilation can be used to replace the EKF over a finite time interval, although it assumes $\mathbf{Q} = 0$ and $\mathbf{P}^a$ has to be estimated indirectly as the inverse of the Hessian of the cost-function (*Fisher,* 1995).
- 3D-VAR analysis is an efficient implementation of equations (3E) and (4), although (5E) has to be replaced by an estimate of inverse Hessian (*Fisher,* 1995).
- A symmetric version of (5) is useful to improve the numerical stability :

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}} \qquad (5S)$$

It can also be used to account for suboptimal analyses (*Bouttier,* 1994).

- All occurrences of the $\mathbf{P}$ matrix can be replaced by a symmetric factorization : $\mathbf{P} = S.S^{\mathrm{T}}$ which guarantees its positive definiteness (*Boggs et al.,* 1995).

- The expensive equation (2) can be computed by assuming that $\mathbf{Q}$ is added only at the end of each prediction cycle. One can integrate the lines and columns of the covariance matrix : $\mathbf{MPM}^{\mathrm{T}} = \mathbf{M}(\mathbf{MP})^{\mathrm{T}}$, which suppresses the need for coding the adjoint of the tangent linear model, or for explicitly building the matrix $\mathbf{M}$ (*Gauthier et al,* 1992). Conversely, one can use an adjoint formulation, which allows one to directly compute a subset of the coefficients of $\mathbf{P}^f$ (less the model error), i.e. if u and v are two arbitrary vectors,

$$u^{\mathrm{T}}\mathbf{MP}^a\mathbf{M}^{\mathrm{T}}v = (\mathbf{M}^{\mathrm{T}}u)^{\mathrm{T}}\mathbf{P}^a(\mathbf{M}^{\mathrm{T}}v) \qquad (2\mathrm{A})$$

  which costs only a couple of adjoint integrations, and avoids handling large matrices (*Barkmeijer and Opsteegh,* 1991 ; *Veyre,* 1990 ; *Bouttier,* 1993).

In the sequel we shall see some applications of those variant formulations.

2. EXPERIMENTAL RESULTS

2.1 With simple models

Because of computational constraints, most of the experimental work on Kalman filtering has so far been performed only with very simplified models. One may question the relevance of such results to NWP, because such models do not even include what are considered to be the most fundamental mechanisms of short-range atmospheric dynamics, namely baroclinic instability and the peak of energy at synoptic scales. The expected increases in computer power will hopefully allow some more realistic EKF experiments to take place in the near future. Nevertheless, there are some lessons to be learnt from low-resolution experiments, despite the need to take care when extrapolating them to high-resolution models :

- what kind of new meteorological information is provided by the EKF, like error variance or correlation maps ;

- how do the input parameters (the model dynamics and the observations) work together in the EKF algorithm to provide original information ;

- what kind of practical problems arise when implementing the EKF : non-linearity, specification of $\mathbf{Q}$ and $\mathbf{R}$, numerical problems ;

- what is the impact of a modification to the observing system : the error covariance matrices summarize the estimation errors in a much more comprehensive way (*Cohn and Parrish,* 1991 ; *Gauthier et al,* 1992) than the usual impact experiments which tend to depend on details of the experimental setup ;

- exact implementations of the EKF, even in low-resolution models, are necessary to serve as a reference for the validation of approximate algorithms, which is a necessary step before experimenting with higher-resolution models (*Todling and Cohn,* 1993).

The most notable experiments in the meteorological community have involved the following types of models : simple 1-D (*Ghil,* 1989 ; *Dee,* 1991), linear 2-D (*Parrish and Cohn,* 1985), unstable 2-D (*Cohn and Todling,* 1995 ; *Gauthier et al,* 1992 ; *Bouttier,* 1994), 3-D quasi-geostrophic (*Houtekamer,* 1993). Considerable work has also been done in the oceanographical community (*Miller,* 1986 ; *Bennett and Budgell,* 1989 ; *Evenssen,* 1992 ; *Gourdeau and Minster,* 1993) and is continuing steadily.

The most interesting part of the EKF equations is (2) and (5), which feed some information into the error covariance matrix, and (3), which defines how that information feeds back into the model analysis. On the other hand, (1) and (4) are quite trivial in themselves : if an approximation is used to define $\mathbf{P}^f$ in (3), then the set (1), (3) and (4) define the well-known scheme for 3D-VAR or OI assimilation of observations. Thus, a first step in understanding what is the added-value of the EKF is to compute only (2), the forecast equation for the covariance matrix. In *Bouttier* (1993), this has been done using a simple initial condition for $\mathbf{P}^a$ (a matrix with gaussian correlations and longitude-independent variances), assuming $\mathbf{Q} = 0$ and linearizing the model in the vicinity of a realistic non-linear trajectory. One can see in figure 1 how the equation (2) changes the error variances. The error correlations are modified as well, as shown in figure 2. Such experiments show the net effect of the linearized model dynamics on the error covariances. It is not straightforward to explain the observed effects in terms of what we know about the atmospheric dynamics, but some understanding is gained by running similar experiments using academic flows as basic states. It is also useful to perform an adjoint decomposition of equation (2), as explained in *Bouttier* (1993), in order to analyze precisely the forecast error covariances in terms of the dynamics and of the initial covariances. One can show that the variances tend to spread because of wave dispersion, to move because of advection and to increase because of local dynamical instabilities, as one would expect from local sensitivity techniques (*Errico and Vukicevic,* 1992 ; *Rabier et al,* 1994). In unstable situations, the forecast error correlation structure follows closely that of the singular vector which maximizes the local error growth (*Farrel,* 1988 ; *Buizza and Palmer,* 1995) ; in the present experiment it is determined by barotropic instability patterns.

A slightly more complex experiment, performed in (*Bouttier,* 1994), consists of evaluating (2) and (5) cyclically, but without feeding back the covariance information into (3) ; instead, the forecasts (1) and analyses (4) used to define the trajectory for linearization are borrowed from an existing high-resolution assimilation system, in which a static approximation is used for $\mathbf{P}^f$ in (3). Thus, we can see how the information from the dynamics and observations feed into the covariance matrices $\mathbf{P}$ over a long period of time, without allowing the covariances to interfere with the model state analysis. This is a convenient way of experimenting with error covariances, because the model-state analysis will remain reasonably realistic, even if the covariances go wrong. Another feature is that an approximation of the true analysis gain (3) is used, because the covariances are computed only on a low-resolution 2-D grid, with a simplified

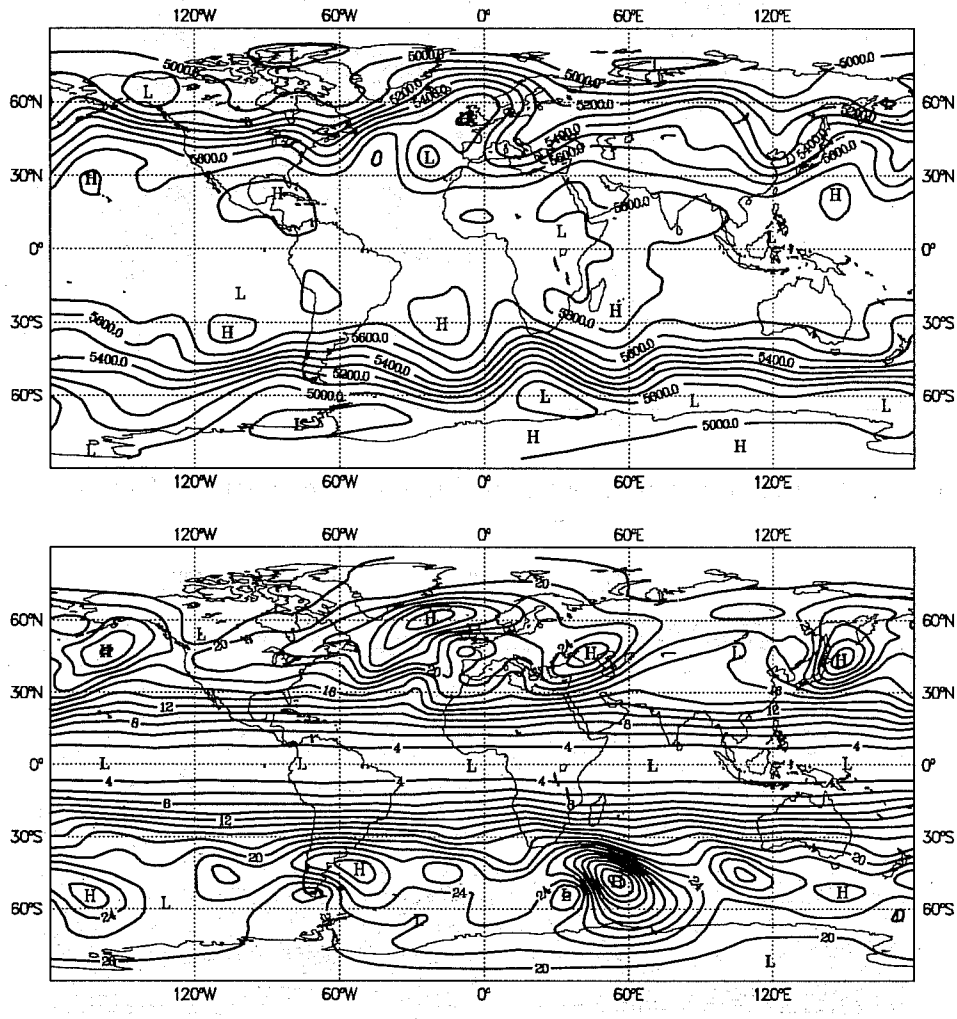

Figure 1: Top panel : 500hPa geopotential height (in m) field used as initial condition for the trajectory. Bottom panel : height standard error field using equation (2E) for 24h with a T21 barotropic model, starting with uniform values in each extratropical hemisphere. (reproduced from *Bouttier*, 1993).

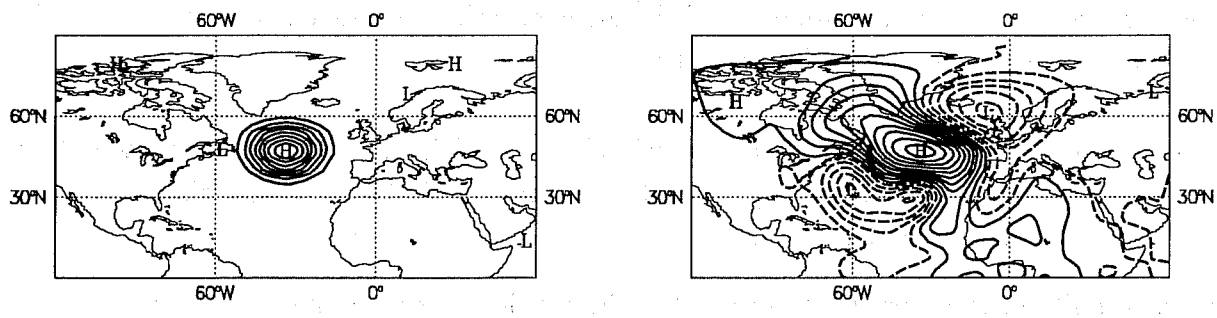

Figure 2: Left panel : initial condition for the height autocorrelation field in the same experiment as in fig.1 . Right panel : same autocorrelation field after the 24h prediction.

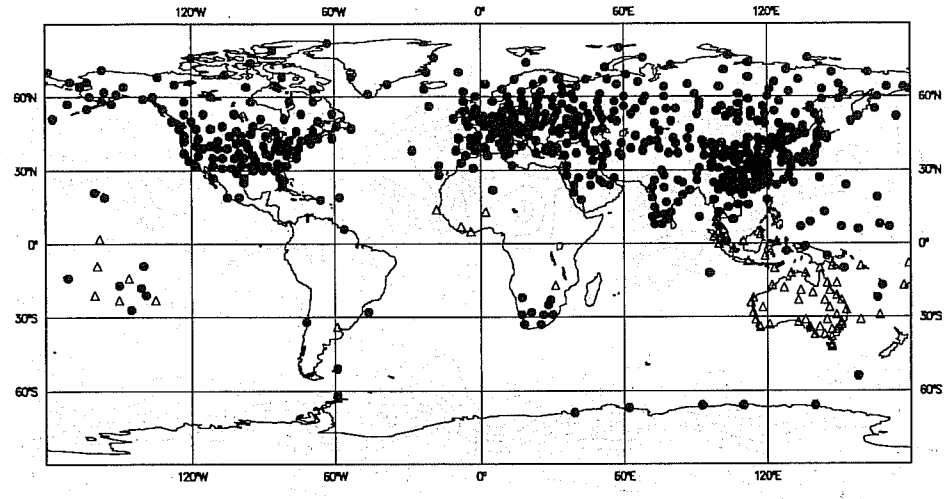

Figure 3: Typical observing network used in the experiments in *Bouttier,* (1994).

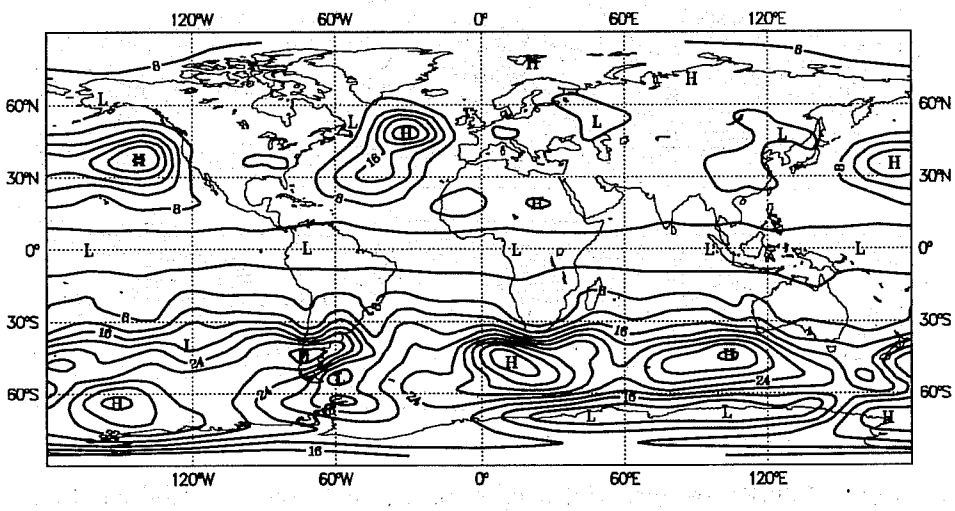

Figure 4: Mean quality of the 500hPa height analysis according to a 16-day integration of equations (2E)-(5E) at T21 resolution on a single level, with a parameterization of non-linear error saturation. Reproduced from *Bouttier,* (1994).

observing network shown in figure 3. Consequently, the symmetric form (5S) for equation (5) has to be used in this experiment, because the analysis gain is no longer truly optimal. Using this experimental setup, an estimation of the error covariances can be performed for several weeks, which yields a lot of useful information :

- the map of the quality of each analysis (fig.4), which depends on the meteorological situation and on the observing network : the quality is lowest along the extratropical jets, except over data-dense areas ;
- the quality of the short-range forecasts, which reflect the advection of the analysis errors by the wind and their amplification in unstable zones ;
- the location of the most useful observations for the quality of each analysis : they are located on the western sides of the continents and on isolated islands (keeping in mind

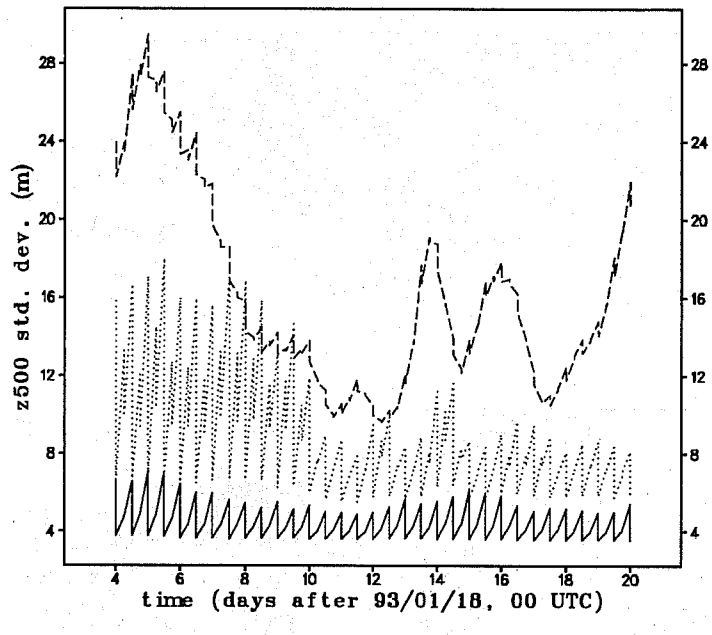

Figure 5: Evolution of the estimation error on 3 selected points during the experiment same experiment as in fig.4 . full line: in the middle of the USA ; dotted line: in the Bay of Biscay ; dashed line: somewhere amid the North Atlantic, far from land.

the low resolution of the experiment) ;

- the time evolution of the quality of the analyses and short-range forecasts, for each model parameter (fig.5) : the behaviour is more sensitive to the atmospheric situation over and near data-sparse areas ;

- the structure functions which would be used by a real EKF in the analysis (fig.6, upper panel) : compared to their equivalents in 3D-VAR or OI, they would lead to a better use of the observations over land for the analysis over the oceans ;

- the structure of the errors in the existing analyses (fig.6, lower panel) : they are almost homogeneous and isotropic over data-dense areas, otherwise they are look like the structure functions themselves.

This information comes on top of the improvement of the analysis itself, which would be expected from a complete implementation of the EKF. Experience in that framework also tells us about some implementation problems :

- the symmetric formulation (5S) is sufficient to force the covariance matrices to remain symmetric and positive definite.

- some kind of parameterization of non-linearities must be introduced into equation (2E), otherwise the tangent linear hypothesis will break down in data-poor areas, which leads to unrealistic large variances.

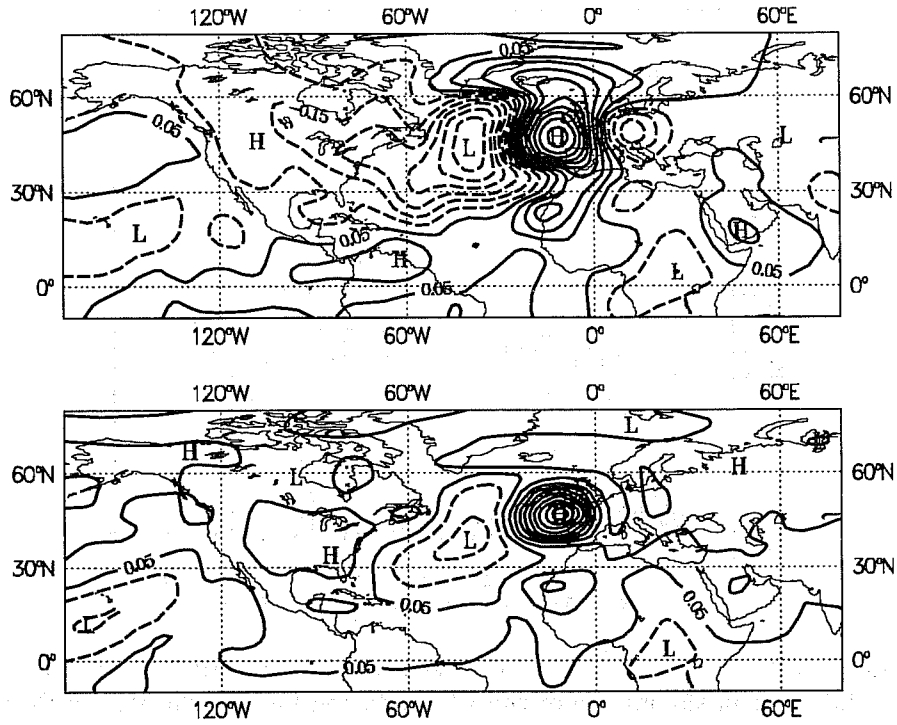

Figure 6: Examples of error structures (height error autocorrelation fields) estimated in the same experiment as fig.4 . Top panel: in a first-guess field ; bottom panel: in an analysis.

- if the model error $\mathbf{Q}$ is ignored, the EKF will run, but the error variances will be too small ; the error correlations will also be too broad, which is unrealistic and may eventually create ill-conditioning problems.

- if the EKF is run with a low resolution, the observation error matrix $\mathbf{R}$ shall be inflated to account for representativeness errors.

- it is possible to estimate some convenient time-averaged values for $\mathbf{Q}$ and $\mathbf{R}$, by computing statistics on the real forecast errors (*Hollingsworth and Lönnberg,* 1986 ; *Lönnberg and Hollingsworth,* 1986). These estimates are themselves a valuable by-product of running the EKF, but if the model used for the error covariance evolution is overly simplified, it will appear that the produced error covariances $\mathbf{P}$ are realistic, but not very informative.

Many other interesting results have been produced by other authors, usually by running a genuine EKF ; some have run into the practical problems explained above, others have gathered some genuinely interesting information, like the impact of various observing systems, of special features of the atmospheric dynamics, or of differences in the numerical formulation of the EKF. The interested reader is invited to look into the references listed at the beginning of this subsection.

2.2 With state-of-the-art models

As explained above, it is unthinkable to run the EKF in a high-resolution model. Fortunately there are some techniques which allow one to look at what a high-resolution EKF would produce, at only a fraction of its price. Of course, this involves some approximations, but they seem quite acceptable. The main restriction of such techniques is that one computes the EKF equations only for a very small subspace of the model and for a limited time period, so that it is not possible to apply these technique to data assimilation, which would involve global computations going on over long time periods.

Thanks to its equivalence properties with the Kalman filter, 4D-VAR assimilation performed over a given time interval gives the same result, in terms of analyzed model state, as the Kalman filter for the same model, assuming the model is perfect. To run 4D-VAR, one also has to specify the forecast error covariance matrix $\mathbf{P}^f$ at the beginning of the time interval. Such a matrix is necessarily approximate, as no EKF has been run before launching the 4D-VAR. One has to resort to an empirical model of background error covariance matrix, assuming some simple correlations and variances. It has been shown (*Thépaut and Courtier,* 1991) that both hypotheses (perfect model and approximate $\mathbf{P}^f$) look quite acceptable when the 4D-VAR is run for a time period between 6 and 24 hours. For longer periods some problems are bound to occur due to non-linearity and model errors. At the end of the optimization period, a 24h 4D-VAR produces an analysis which is close to what an EKF would have provided at the same resolution. The increment (the difference between the analysis and the first-guess) is the same for both algorithms : looking at a 4D-VAR increment based on a single observation at the end of the period gives the structure function we would have had in the EKF at the same time, i.e. after the error covariances have been significantly modified by the dynamics (i.e. (2E) for 24h with $\mathbf{Q} = 0$) and does not depend much on the initial covariances. This idea has been applied to a case study in *Thépaut et al* (1993) using a T63L19 primitive-equation model for the evolution of errors. The corresponding increments (fig.7) for a single observation show that the strong baroclinic instability in the past weather evolution leads the EKF (the 4D-VAR) into producing some strongly baroclinic increments. Those increments are simultaneously consistent with the local flow dynamics and with the location of the observation. As the observation has been put in an area of strong instability, the structure of the corresponding increment is very close to that of the singular vector which maximizes the global error growth over the same period. A more comprehensive case study has been performed in (*Thépaut et al,* 1994), using the same methodology, except that a whole set of single-observation experiments has been performed, with simulated observations on the grid of a vertical cross-section of the atmosphere. This provides a vertical cross-section of the 24-h forecast error correlations for a given observation (Fig.8a), as well as a cross-section of the forecast error variances (Fig. 8b). In this particular case the variances are highest in the low-level core of a low-pressure system and along the upper-level jet, and the correlation with an observation situated in the low-level warm sector is tilted in the vertical, and exhibit a clear identification of the air masses near the ground (the correlation is higher with the warm air, lower with the cold air). Such an "intelligent" identification of the meteorological situation is so far completely missing in the current operational analysis systems.

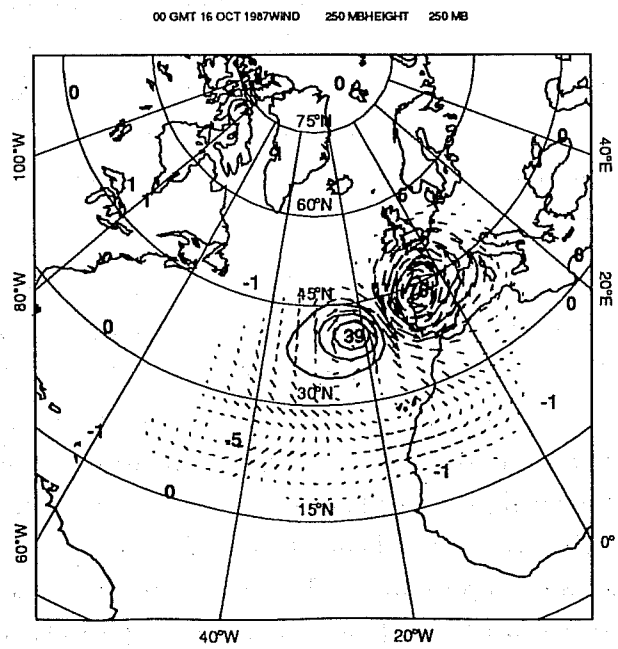

Figure 7: Structure of a 4D-VAR increment (at level 250hPa) produced by a single height observation near an unstable zone. Reproduced from (*Thépaut et al*, 1993).

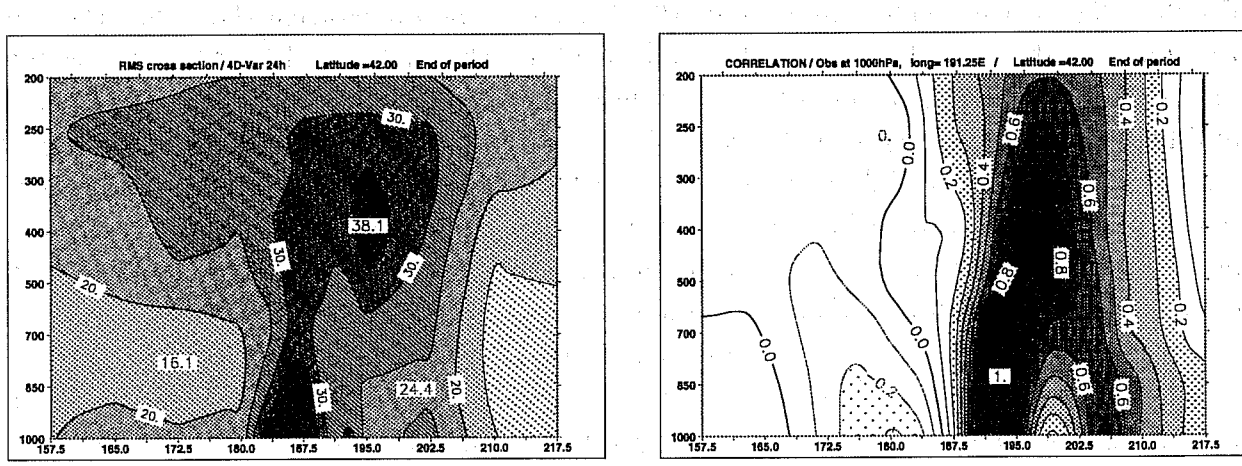

Figure 8: Left panel: vertical cross-section of the first-guess height standard errors assumed by a 4D-VAR analysis in the vicinity of an occluded frontal system. Right panel: vertical cross-section of a first guess height correlation in the same situation. Reproduced from *Thépaut et al*, (1994).

Another technique, the sensitivity study, is less straightforward but even more economical. It consists of studying vectors of the kind $\mathbf{M}^{\mathrm{T}}v$, as defined in (2A), where $\mathbf{M}^{\mathrm{T}}$ is the adjoint operator of the tangent linear in the vicinity of a given short-range forecast (usually a 24 or 48-hour forecast). The vector v can define a particular feature of the prediction (e.g. the value of a particular forecast parameter at a given point, or its average over a given domain, in which case v is the corresponding interpolation/average operator), see *Errico and Vukicevic* (1992) and *Delode et al* (1995), or some part of the prediction error if the forecast is compared to a verifying analysis (then v is the error pattern times the matrix of a user-defined scalar product), see *Rabier et al* (1994). The vector $\mathbf{M}^{\mathrm{T}}v$ is often called the sensitivity pattern of v ; it can be plotted as a meteorological field[4] in order to highlight the areas where the analysis errors have been important for the prediction of v, as well as the structure of those errors. Some examples of sensitivity patterns are shown elsewhere in this volume. Thus, the sensitivity patterns show what kind of analysis increments an optimal assimilation system would have had to generate in order to provide successful predictions. Their study allows for an assessment of the minimal requirements of future EKFs, in terms of resolution of error covariances, so that they will be able to correct the operational forecast failures we experience nowadays. Such EKFs would have the necessary structure functions to produce analyses of good quality, provided they are given observations of good quality.

3. LINKS WITH OTHER TECHNIQUES

Many techniques are being used for data assimilation or predictability estimation, and many more are advocated in the literature. It can be unclear how they compare to each other : some are quite empirical, which does not prevent them from being useful. Because of its clear theoretical foundation, the EKF is a helpful common framework to clarify the ideas behind these techniques.

3.1 Usual data assimilation schemes

A comprehensive comparison of many assimilation techniques is found in *Lorenc* (1986), which demonstrates, among other things, the equivalence between some versions of successive corrections (or nudging), OI, 3D-VAR, 4D-VAR, splines, and the Kalman filter. In operational NWP centres, the most popular algorithms are OI, successive corrections and 3D-VAR (also known as SSI).

Assimilation systems using OI are an approximation of the EKF, because they are sequential, with non-linear prediction cycles according to equation (1), and linear analysis steps according to (4). The analyses are only performed once in a while (every few hours) to save computational time. The main approximation is that (2E) is only computed using an empirical

[4]Once a convenient scalar product has been defined for the analysis perturbations. It should be noted that a natural, and completely objective, choice of scalar product is the one defined by the analysis error covariance matrix. There is no need for the sensitivity pattern scaling to be arbitrarily defined.

evolution law for the error variances. The gain (also called the analysis weight) is approximately computed according to (3), using a kind of banded-matrix approximation in the matrix inversion. The forecast error covariance matrix $\mathbf{P}^f$ is specified in (3) using the approximate error variances, together with a correlation model which is quasi stationary and empirical, according to an analytical, homogeneous and isotropic formulation, and to some geostrophical balance assumption for the mass-wind correlations (*Lorenc,* 1981). The equation (5E) is only computed for the variances. Thus, OI assimilation is an EKF with a static correlation model and linear observation operators.

Successive correction schemes are closely related to OI, except that the observations are used at a more accurate time, and the gain computation (3) is approximated empirically.

The more recent 3D-VAR (*Andersson et al,* 1994) (or SSI, *Parrish and Derber,* 1992) scheme is conceptually very similar to OI. The management of error covariances is performed in much the same way, and the main difference comes from the numerical method for solving (3E), in which the matrix inversion is implicitly replaced by the iterative (thus approximate) solution of a variational version of the same problem. The latter method suppresses some numerical noise found in OI analyses. An important advantage of 3D-VAR is mainly technical ; the structure functions can be made more complex than in OI (*Rabier and Mc Nally,* 1993). The algorithm also makes it possible to account for weakly non-linear observation models $\mathbf{H}$, which has great practical importance for using satellite data (*Andersson et al,* 1994).

3.2 Predictability techniques

The predictability problem is the estimation of the probability density function (pdf) of the predictions. The main components of the pdf are the mean (or the mode) and the error covariance matrix, which is precisely what the EKF estimates in an assimilation system. Equations (1) and (2E) can be used to estimate the pdf in the forecasts as well. Since (2E) is only valid within the tangent linear hypothesis, the EKF is not relevant for medium-range prediction (which pertains to the more general problem of stochastic-dynamic prediction, *Epstein* (1969)). However, the chances are that a good estimate of the medium-range predictability will have to rely on a good estimate of the pdf for shorter ranges, and that is precisely what is reflected in the design of modern ensemble prediction systems (EPSs).

Ensemble prediction itself is a discrete sampling of the forecast pdf, from which in turn some approximation of the state and its error covariance can be computed if the ensemble is large enough ; thus, it replaces equations (1) and (2E) of the EKF. The current EPSs use only a very crude approximation of the analysis error covariances $\mathbf{P}^a$ (*Molteni et al,* 1994), because it is believed that the forecast error will depend more on error amplification during the forecast than on the precise structure of initial analysis error. On the other hand, the evolution of what is sampled from the initial pdf is computed without any linear approximation (the model error is still neglected) : to that extent, an EPS is superior to the EKF. Another feature of current EPSs is the small size of the ensemble (less than 100), which amounts to assuming that the dimensionality or forecast errors is very small ; that hypothesis is supported to some extent

by experiments (*Houtekamer,* 1993 ; *Buizza and Palmer,* 1995) but remains to be discussed. On the other hand, the EKF does not imply such an assumption (although running the EKF at a low resolution amounts to supposing that forecast errors have a low resolution, if model error is not accounted for). Theory dictates that, when the ensemble size is increased, the EPS converges to the EKF (to the complete pdf, actually, including higher-order moments). In conclusion, the EPS is only superior to the EKF pending some hypotheses which remain to be clearly investigated. Those hypotheses are more convincing for long forecast ranges than for the short range. Nevertheless, EPS is considerably easier to implement than an EKF, as it allows one to use directly a realistic forecast model, including physics. On the other hand, EPS is probably not very useful for data assimilation in NWP, as the small ensemble sizes do not allow for a correct estimation of the error correlations (*Fisher,* 1995). Another reason to prefer the EKF is the linearity of the error evolution in atmospheric data assimilation ; in other frameworks, where non-linearity is more important, it is indeed possible to modify the EKF, using an EPS strategy to replace equation (2E), so that non-linearities are accounted for (*Evenssen and van Leeuwen,* 1995).

It is well known that EPS cannot rely solely on statistical sampling of the assumed initial pdf, because most analysis errors will actually be damped in the subsequent prediction : the dimension of the unstable subspace tangent to the prediction is much smaller than the dimension n of the model itself. Modern EPS techniques carry out an estimation of the most unstable part of the daily dynamics. The breeding method (*Toth and Kalnay,* 1993 ; see also in this volume) performs a kind of local Lyapunov vector computation : a set of independent assimilation cycles is emulated by adding some (initially random) perturbations to the reference analysis, using the forecast model to propagate perturbations in time and an empirical rescaling whenever some analyses are performed. In the long run these perturbations (called bred vectors) look like the most likely error patterns in the assimilation system, and they are used to generate the initial perturbations for the subsequent EPS. One can see the breeding method as an approximation for equations (2E) and (5E) in the EKF, with only a small sampling of the error covariance matrix, and a crude approximation for the gain in (5E). There is (so far) no feedback of the bred vectors onto the gain computation. However, the propagation of the errors uses the high-resolution non-linear model with physics, which means their structure is completely coherent with the model dynamics. The subsequent EPS can be seen in the short range as a low-dimensional sampling of (2E) starting with a low-dimensional, high-resolution sampling of the exact analysis error covariance matrix $\mathbf{P}^a$.

At ECMWF, it has been preferred to initialize the EPS using singular vectors (also called optimal modes), which are the most rapidly growing modes in the early stage of the prediction of interest (*Buizza and Palmer,* 1995 ; see also in this volume). The computation of the singular vectors can be regarded as the exact determination of the short-range forecast error matrix $\mathbf{P}^f$ according to equation (2E) of the EKF, in a small subspace of given dimension which is associated to the largest eigenvalues of matrix $\mathbf{P}^f$ (model error is neglected). In other words, the singular vectors generate the subspace of the model state which contains the most uncertain part of the short-range forecast. In that space, equation (2E) is solved exactly ; a very crude approximation is used for the analysis error covariance $\mathbf{P}^a$, as only analysis variances

are accounted for[5]. Actually, experience shows that it is better not to account for analysis error correlations at all, rather than to specify unrealistic correlations, like the barotropic ones assumed in 3D-VAR, which exhibit a spurious orthogonality to the most unstable patterns of the forecast. Singular vectors are convenient for medium-range EPS, as the spread of the ensemble is maximized in the first part of the prediction, so that for a given ensemble size the shape of the medium-range pdf is made as visible as possible. For the short range, singular vectors can be demonstrated to be the forecast error covariance matrix sampling which converges the most rapidly to the exact matrix $\mathbf{P}^f$ computed by the EKF as the ensemble size is increased.

One can see that both methods provide a discrete sampling of the short-range pdf using two different approximations of the EKF :

the breeding method relies on a proper sampling of the analysis error covariances $\mathbf{P}^a$ using an approximation of the complete EKF. The sampling accounts for the instability properties of the past history of the assimilation.

the singular vectors rely on a proper sampling of the short-range low-resolution forecast error covariances $\mathbf{P}^f$ at the beginning of the prediction. The analysis error matrix $\mathbf{P}^a$ is very crudely approximated. Given these approximations, the sampling is mathematically optimal.

Both methods have their drawbacks, and it is probably desirable to try and combine the best features of them in a single algorithm.

Another class of predictability techniques is worth mentioning briefly : those based on the adjoint decomposition (2A) of equation (2E), already described in section 1. Although they can only be used within the validity of the tangent linear hypothesis, they allow for the determination of the error variances (or correlations) of only a few forecast parameters. They can be superior to the EPS because they perform the exact computations of equation (2E) for the parameters of interest (all degrees of freedom of the model dynamics and initial error covariances are taken into account). They are good candidates for the design of specific short-range predictability systems (*Barkmeijer et al,* 1993). However, they depend on the quality of the estimate of the analysis error $\mathbf{P}^a$.

3.3 Advanced assimilation techniques

Some improved algorithms are being developed for data assimilation, based on the breeding method or on 4D-VAR. They can be regarded as versions of existing techniques which have been modified to be closer approximations of the EKF.

It has been proposed to use the breeding method to improve the SSI (3D-VAR) assimilation (*Toth and Kalnay,* 1993). As explained before, 3D-VAR is equivalent to (1), (3) and (4),

[5]The choice of scalar product in the singular vector computation amounts to an assumption about the structure of analysis errors.

whereas the breeding method is equivalent to approximations of (2E) and (5E). To combine both, it is proposed to project (in some sense) each analysis onto the subspace defined by the observations, parallel to the subspace generated by the bred vectors. The bulk of the forecast error covariance matrix $\mathbf{P}^f$ would still be specified in (3) according to empirical correlation models. This procedure is equivalent to setting the variances in $\mathbf{P}^f$ to infinity in the subspace generated by the bred vectors. Since bred vectors are precisely regarded as very uncertain components of the short-range forecasts, the proposed technique could actually be a very convenient low-cost approximation to the full EKF, where the correlations (not the variances) are exactly managed in a small subspace, using a high-resolution, non-linear version of (2E).

The 4D-VAR analysis is generally regarded as the basic ingredient of next-generation assimilation systems. As explained before, and despite its theoretical equivalence to the EKF in the linear case, the first 4D-VAR implementations would still differ from the EKF in many aspects : the minimization interval would be limited, and the initial $\mathbf{P}^f$ matrix would be based on simple correlation models, as in 3D-VAR. As a consequence, 4D-VAR would be a nearly exact approximation to the exact EKF inside each optimization time interval, but equations (2E) and (5E) would not be computed between the intervals. This is known as the cycling problem, which is of more concern if the interval is shorter, because then the structure functions depend more on the initial $\mathbf{P}^f$ matrix. Another important approximation coming with 4D-VAR is the perfect model assumption ($\mathbf{Q} = 0$) inside each optimization interval, which is of more concern if the interval is longer, and if an incremental version of 4D-VAR is used to reduce the resolution of the control variable, or to remove the physics from the minimization (*Courtier et al,* 1994). Taking into account the model error in 4D-VAR itself is an area of active research (*Leeuwen,* 1995). Thus, 4D-VAR is a step towards the EKF, but still leaves room for improvement.

3.4 Other techniques

The EKF is by no means the ultimate assimilation algorithm. As explained in the introduction, it relies on many hypotheses which are acceptable only to a limited extent.

An important constraint of the EKF design is the sequentiality in the use of observations, which is not necessary for non-realtime applications such as reanalysis for climate studies or field experiments. The problem of optimal retrospective data analysis is known as smoothing, and there is actually a variant of the Kalman filter, called the Kalman smoother, which can be used whenever data for a given analysis can be taken from the future as well as from the past (*Bennett and Budgell,* 1989). Unfortunately, such an algorithm is even more expensive than Kalman filtering. There are a number of interesting lower-cost alternatives (*Cohn et al,* 1994), among which is 4D-VAR analysis.

An issue of much concern with EKF implementation is the specification of the error covariances for the model $\mathbf{Q}$ and the observation $\mathbf{R}$ errors (although some knowledge about $\mathbf{R}$ has been gathered in the past because it was needed in operational assimilation systems). These matrices contain an amount of information we are simply unable to provide, given the data records available (*Dee,* 1991 ; *Dee,* 1995) : it is necessary to use some indirect techniques to

estimate them. More generally, there are several hypotheses behind the EKF (no biases, no time correlation of model/observation error, linearity, plus additional hypotheses made to simplify the algorithm) that need to be validated in real time using objective data. To that end, a convenient framework is adaptative filtering, which is likely to develop in the years to come.

4. CONCLUSION : FUTURE IMPLEMENTATIONS OF THE EKF

Since the EKF is supposed to provide improved analyses, together with information about their quality which would be very useful for ensemble prediction, it is desirable to implement it in an operational framework. However, such a task is difficult, because the true EKF is too costly, and we do not know much about the model error $\mathbf{Q}$. Anyway, it would make little sense to exactly implement the EKF : it is probably not useful to compute explicitly the huge amount of information contained in the forecast error covariances $\mathbf{P}^f$ at the resolution of an operational model. One shall first think about what we really need from the EKF.

From a physical point of view, the algebraic complexity of the EKF equations stems from the following features of the algorithm :

- any observation can be used to correct any part of the model state,
- there can be some significant estimation errors in all degrees of freedom of the model state,
- any perturbation in the model can interact simultaneously with all model variables,
- there can be some significant correlations between the estimation errors of all couples of model variables.

Physical intuition tells us that what happens in a meteorological data assimilation system is certainly much less complex. Most observations are local in space, and uncorrelated to each other. Like the real atmosphere, the dynamical model phase space is contained in a manifold which is much smaller than the model state dimension. Most meteorological phenomena are known to propagate at a very limited speed, so that a local perturbation at a given timestep will only have a local effect in the next timestep. The combined local character of most observations and dynamics should lead to mostly local error structures, as shown in the mean by the correlation statistics of forecast errors (*Hollingsworth and Lönnberg,* 1986 ; *Lönnberg and Hollingsworth,* 1986). As a consequence, the complexity of the actual forecast error covariances, and hence the actual cost of a well-designed EKF, is probably much smaller than what is apparent from the equations themselves. A few simplification methods are known, which allow one to reduce the cost of the EKF without sacrificing its usefulness ; many more are probably to be discovered and implemented.

The computation of equations (1), (3E) and (4) is to a large extent a solved problem, thanks to the 3D-VAR (or SSI) technique. The main hurdle for the implementation of the EKF is the

evaluation of (2E) and (5E), in which the representation of the error covariances is as much a problem as the computational cost itself. Too many techniques for approximation have been proposed to be cited in this paper ; we are only going to depict the most promising ones.

As already explained, the discrete sampling (or : randomization) of the error pdfs is a way to avoid building explicitly large covariance matrices, while at the same time allowing the computation of the error evolution with the most realistic models, as is done for EPS. It is technically straightforward to implement and can be used to account for non-linearities. However, it is more convenient for estimating the error variances than the correlations (*Fisher,* 1995), so that it is not a very attractive way to provide structure functions for the analysis, unless one is willing to pay the price of a very large ensemble. As a rule, the quality of the error covariances that can be recovered from a discrete sample grows like the square root of the ensemble size, which means that the method is rather inefficient if satisfactory covariances are not already obtained with a few elements.

A promising technique is the eigenvector decomposition of the error covariance matrices. The underlying assumption is that the most important part of the errors, from the point of view of some user-defined scalar product, is contained in a small subset of the model phase space, which is the one generated by the eigenvectors associated to the largest eigenvalues of the error covariance matrices (the smallest eigenvalues may also be of some interest). This subspace is the optimal way of representing as much as possible of the total error covariances while at the same time keeping the problem dimension low. Thanks to the symmetry of the matrices, we obtain by design a diagonal representation of the covariance matrix in the relevant subspace, and we know that the unknown remaining matrix coefficients are all contained in the orthogonal subspace (the cross-correlations are zero). Thus, one can build a good approximation of the complete covariance matrix by using the known eigenvectors and eigenvalues in the most important subspace, and filling the remainder by some convenient empirical models for error covariances, like the ones already used in 3D-VAR or OI. The approximation is numerically efficient if the spectrum of the eigenvalues is steep, so that the unresolved part of the matrix is small ; that property is likely to hold (*Houtekamer,* 1993) as a consequence of the dimensionality of the atmospheric dynamics (only a few areas of instability appear to be really active at a given time, see *Buizza and Palmer* (1995)). One can think of using that approximation to compute only the most interesting part of the matrices on the left-hand side of equations (2E) and (5E) ; there are some efficient iterative algorithms, like the Lanczos method (*Buizza et al,* 1992 ; *Fisher,* 1995), which provide such information directly, without ever having to build explicitly the matrices of the operators used. However, some further studies are required, as one does not know precisely how the "remainder" of the matrix would interact with the subsequent EKF computations : a more fundamental question is, how do the analysis errors and forecast dynamics interact with each other ?

It is also possible to implement the EKF at only a reduced resolution (known as RKF, or reduced Kalman filter, in *Todling and Cohn* (1993)) for the error covariance matrices, as suggested in section 2. Then the matrices have to be blended with an empirical covariance model to provide the high-resolution covariance model necessary for the state analysis in (3). The minimum resolution necessary to obtain realistic results is not well known. Some dramatic

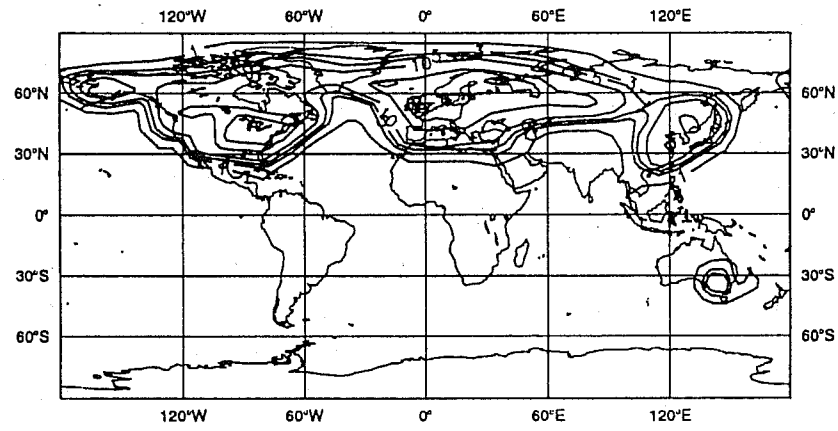

Figure 9: Differences between the background and analysis standard errors for height near level 500hPa, using a Lanczos-based approximation of (5E) where only the most significant subspace of dimension 52 has been resolved. Reproduced from *Fisher,* 1995.

reductions to the EKF cost can be achieved by making other specific hypotheses on the error covariance matrices (some covariances are zero...), or on the model itself (the model is barotropic, or slow, or purely advective...). While the results may be quite enticing (*Cohn,* 1992), the consequences and realism of the hypotheses made are often quite unclear. Ideas for simplifying the EKF should always be validated using a rigorous and meteorologically meaningful framework, following the methodology of *Todling and Cohn* (1993).

As a final illustration, here are the ideas for a future operational system at ECMWF (*Courtier,* 1993 ; *Fisher,* 1995) :

- the basic analysis algorithm will be an incremental 4D-VAR run sequentially at a high resolution, with physics, for intervals between 6 and 24 hours,
- the 4D-VAR initial analysis error covariance matrix will be computed exactly in a low-dimensional subspace, based on an eigenvector identification using a Lanczos-like algorithm,
- the short-range forecast error covariance matrix will be computed for the next 4D-VAR analysis using a low-resolution version of equation (2E) of the EKF,
- in the subspace which is not resolved, the analysis and forecast error covariances will be complemented in the whole model space with a complex empirical correlation and variance model, based on the existing experience with the 3D-VAR analysis.

The validation of that scheme will rely on the use of 4D-VAR as a reference, and on a careful monitoring of the effective error statistics. A part of that work has already been implemented for 3D-VAR : the result of a low-dimensional evaluation of equation (5E) at high resolution is shown in fig.9 .

The development of such approximate extended Kalman filters for operational assimilation systems will probably be a major area of interest for data assimilation and short-range predictability in the years to come.

Acknowledgements

Some important results presented in this seminar paper have been kindly provided by collaborators at ECMWF, mainly F. Rabier, E. Klinker, M. Fisher, J.-N. Thépaut and P. Courtier. Thanks are also due to C. Edis and to many developers in the joint IFS/Arpège project between Météo-France and ECMWF.

References

Andersson, E., J. Pailleux, J.-N. Thépaut, J. Eyre, A. McNally, G. Kelly and P. Courtier, 1994: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Quart. Jour. Roy. Met. Soc.,* **120,** 627-653.

Barkmeijer, J., Houtekamer, P. and Wang, X., 1993: Validation of a skill prediction method. *Tellus,* **45A,** 424-434.

Barkmeijer, J., and J. Opsteegh, 1991: Local skill prediction with a simple model. *ECMWF workshop proceedings, New developments in predictability, 13-15 November 1991,* available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Bennett A. and W. Budgell, 1989: The Kalman Smoother for a Linear Quasi-Geostrophic Model of Ocean Circulation. *Dyn. Atmos. Oceans,* **13,** 219-267.

Boggs, D., M. Ghil and C. Keppene, 1995: A stabilized sparse-matrix U-D square-root implementation of a large-state extended Kalman filter. *Proceedings of the 2nd WMO symposium on data assimilation,* Tokyo, 219-224.

Bouttier, F., 1993: The dynamics of error covariances in a barotropic model. *Tellus,* **45A,** 408-423.

Bouttier, F., 1994: A dynamical estimation of error covariances in an assimilation system. *Mon. Wea. Rev.,* **122,** 2376-2390.

Buizza, R., J. Tribbia, F. Molteni and T. Palmer, 1992: Computation of Optimal Unstable Structures for a Numerical Weather Prediction Model. *Tellus,* **45A,** 388-407. Buizza, R. and T. Palmer, 1995: the singular-vector structure of the atmospheric general circulation. *J. Atm. Sci.,* **52,** 1434-1456.

Cohn, S., 1992: Dynamics of Short-Term Univariate Forecast Error Covariances. Accepted for publication in *Mon. Wea. Rev.*

Cohn, S., and D. Parrish, 1991: The Behavior of Forecast Error Covariances for a Kalman Filter in Two Dimensions. *Mon. Wea. Rev.,* **119,** 1757-1785.

Cohn, S., N. Sivakumaran and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.,* in press.

Courtier, P., 1993: Introduction to numerical weather prediction data assimilation methods. *ECMWF seminar proceedings,* 6-10 Sept 1993, pp.189-207, available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-VAR, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,**

1367-1387.

Dee, D., 1991: Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.*, **117**, 365-384.

Dee, D., 1995: Testing the perfect-model assumption in variational data assimilation. *Proceedings of the 2nd WMO seminar on data assimilation*, Tokyo, pp.49-54.

Delode, F., J. Pailleux and F. Bouttier, 1995: Etude de la sensibilite de la prevision aux conditions initiales a l'aide du modele Arpège et de son adjoint. (In French) Note de centre GMAP no.4, Available from CNRM/Météo-France, Toulouse, France.

Epstein, E., 1969: Stochastic Dynamic Prediction. *Tellus*, **21**, 739-759.

Errico, R. and T. Vukicevic, 1992: Sensitivity Analysis using an Adjoint of the PSU-NCAR Mesoscale Model. *Mon. Wea. Rev.*, **120,** 1644-1660.

Evenssen, G., 1992 : Using the Extended Kalman Filter With a Multilayer Quasi-Geostrophic Ocean Model. *Jour. Geophys. Res.*, **99, C11,** 19995-19924.

Evenssen, G. and P. J. van Leeuwen, 1995: Advanced Data Assimilation Based on Ensemble Statistics. Proceedings of the 2nd WMO symposium on data assimilation, Tokyo, 153-163.

Farrel, B., 1988: Optimal Excitation of Neutral Rossby Waves. *J. Atmos. Sci.*, **45**, 163-172.

Fisher, M., 1995: Estimating the covariance matrix of analysis and forecast errors in variational data assimilation. *ECMWF Res. Dept. Tech. Memo. no.220,* available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Gauthier, P., P. Courtier and P. Moll, 1992: Assimilation of Simulated Wind Lidar Data with a Kalman Filter. *Mon. Wea. Rev;,* **121,** 1803-1820.

Ghil, M., 1989: Meteorological Data Assimilation for Oceanographers. Part I : Description and Theoretical Framework. *Dyn. of Atm. and Oceans*, **13**, 171-218.

Gourdeau, L. and J.-F. Minster, 1993: Two Years of Geosat Sea Level Anomalies Assimilated in a Linear Model of the Tropical Atlantic Ocean. Preprint, available at UMR39/GRGS, 18 Av. Belin, 31055 Toulouse, France.

Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111-136.

Houtekamer, P., 1993: Global and local skill forecasts. *Mon. Wea. Rev.,* **121,** 1834-1846.

Ide, K., A. Bennett, P. Courtier, M. Ghil and A. Lorenc, 1995: Unified notations for data assimilation: operational, sequential and variational. *J. Met. Soc. Japan,* submitted.

Jaswinski, A., 1970: *Stochastic processes and filtering theory.* Academic Press, 376pp.

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

Leeuwen, P., 1995: The generalized inverse of a non-linear open quasi-geostrophic ocean model with the representer method. *Proceedings of the 2nd WMO symposium on data assimilation,* Tokyo, 69-74.

Lönnberg, P., and A. Hollingsworth, 1986: The Statistical Structure of Short-Range Forecast Errors as determined from Radiosonde Data. Part II: The Covariance of Height and Wind

Errors. *Tellus*, **38A**, 137-161.

Lacarra, J.-F., and O. Talagrand, 1988: Short-Range Evolution of Small Perturbations in a Barotropic Model. *Tellus,* **40A**, 81-95.

Lorenc, A., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177-1194.

Lorenc, A., 1981: A Global Three-Dimensional Multivariate Statistical Interpolation Scheme. *Mon. Wea. Rev.,* **109,** 701-721.

Miller, R., 1986: Toward the Application of the Kalman Filter to Regional Open Ocean Modeling. *J. Phys. Oceanogr.,* **16,** 72-86.

Molteni, F., R. Buizza, T. Palmer and T. Petroliagis : The ECMWF ensemble prediction system : methodology and validation. Aug 1994, 55 pp. *ECMWF Res. Dept. Tech. Memo. no.202,* available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Parrish, D., and S. Cohn, 1985: *A Kalman Filter for a Two-dimensional Shallow-water Model: Formulation and Preliminary Experiments.* NMC Office note no.304 .

Parrish, D. and J. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Mon. Wea. Rev.*, **120,** 1747-1763.

Rabier, F., E. Klinker, P. Courtier, A. Hollingsworth, 1994: Sensitivity of 2-day forecast errors over the Northern Hemisphere to initial conditions. *ECMWF Res. Dept. Tech. Memo. no.203,* available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Rabier, F., and T. Mc Nally, 1993 : Evaluation of Forecast Error Covariance Matrix. *ECMWF Res. Dept. Tech. Memo. no.195,* available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Talagrand, O. and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113,** 1311-1328.

Thépaut, J.-N. and P. Courtier, 1991: Four-dimensional data assimilation using the adjoint of a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, **117,** 1225-1254.

Thépaut, J.-N., P. Courtier, G. Belaud and G. Lemaître, 1994 : Dynamical structure functions in 4DVAR: a case study. Proceedings of the 1995 WMO conference on data assimilation, Tokyo, pp. 129-134.

Thépaut, J.-N., R. Hoffman and P. Courtier, 1993 : Interactions of Dynamics and Observations in a Four-Dimensional Variational Assimilation. *Mon. Wea. Rev.,* **121,** 3393-3414.

Todling, R. and S. Cohn, 1993: Suboptimal Schemes for Atmospheric Data Assimilation Based on the Kalman Filter. Sumitted to *Mon. Wea. Rev.*

Toth, Z. and E. Kalnay, 1993: Ensemble Forecasting at NMC : The Generation of Perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317-2330.

Veyre, P., 1990: Study of the Forecast of the Forecast Skill with the Kalman Filter Technique. *1st WMO International Symposium on Assimilation of Observations* , Clermont-Ferrand, July 1990.

Vukicevic, T., 1991: Nonlinear and Linear Evolution of Initial Forecast Errors. *Mon. Wea. Rev.*, **119**, 1602-1611.