THE RPN ENSEMBLE PREDICTION SYSTEM

P.L. Houtekamer
Recherche en Prévision Numérique, Atmospheric Environment Service,
Dorval, Canada
phoutekamer@rpn.aes.doe.ca

Louis Lefaivre
Canadian Meteorological Centre, Atmospheric Environment Service,
Dorval, Canada
llefaivre@cmc.aes.doe.CA

Jacques Derome
Department of Atmospheric and Oceanic Sciences,
and Centre for Climate and Global Change Research,
McGill University, Montreal, Canada
derome@zephyr.meteo.McGill.CA

Abstract

For many aspects of numerical weather prediction it is important to have good error statistics. Here one can think of applications as diverse as data assimilation, model improvement and any-range forecasting. In this paper, we propose a method for producing all these statistics from a representative ensemble of forecast states at the appropriate forecast time. To generate the ensemble, an attempt is made to simulate the process of error growth in a forecast model. For different ensemble members the uncertain elements of the forecast (the observations and some components of the model) are perturbed in different ways.

To obtain perturbations to the model we select different options for the parametrization of: horizontal diffusion, vertical diffusion, deep convection, gravity wave drag and orography. As part of the forecast error is due to model error, perturbing the model will lead to an improved ensemble forecast. This also creates the opportunity to use the ensemble forecast for model sensitivity experiments.

For each ensemble member we perform a different six-hour assimilation cycle. For this we randomly perturb the available observations. The perturbed observations are input to the statistical interpolation assimilation scheme, giving a perturbed analysis. This analysis is integrated for six hours with a perturbed version of the T63 forecast model, using perturbed surface fields, to obtain a perturbed first guess for the next assimilation. After cycling for four days we find that the ensemble statistics have become stable. Given a big enough ensemble it is possibly to replace the correlation model that is used by the statistical interpolation procedure, by covariances estimated from the ensemble. The simulated analysis error fields, from the above independently perturbed assimilation cycles, serve as the initial states for an any-range ensemble forecast.

1 INTRODUCTION

At the Canadian Meteorological Centre (CMC) and at Recherche en Prévision Numérique (RPN) it is being attempted to build up an ensemble prediction system starting from first principles. It is also attempted to come up with a product of a rather general utility. Its output should directly be useful for model improvement, any-range ensemble forecasting and improvement of the assimilation system.

It seems evident that in order to predict the forecast error one should have a thorough understanding of its sources. In our opinion the main problem for ensemble forecasting is a lack of such understanding of the origin of the forecast error. An ensemble prediction scheme should therefore come with many tools to diagnose the validity of its components. These tools can then be used to improve not only the ensemble forecast but the entire forecasting system it is using.

As we will explain below all sources of error are combined using a system simulation approach. Here one perturbs the input randomly in order to obtain random output errors. Such an approach might also be referred to as a Monte Carlo method or as a black box approach.

1.1 Why we use a system simulation approach.

The system simulation approach consists of randomly perturbing uncertain aspects (initial coordinates and/or observed values and/or model parameters) of a forecast in agreement with their statistics. If this is done several times a representative ensemble of forecast states is obtained. Mathematicians would consider a system simulation approach to be a simple straightforward Monte Carlo method. In the meteorological literature the term Monte Carlo method has often been used for work where one decides to randomly and independently perturb all coordinates off the initial state. Such perturbations will decay significantly before showing any growth (e.g., Smagorinsky 1969). As this is not what we do, we will mostly refer to our method as being a system simulation approach or system simulation experiment (SSE). This makes it clear that our method is a more general version of an observation system simulation experiment (OSSE). OSSE's have a long history in the meteorological literature (e.g., Charney et al. 1969; Arnold and Dey 1986; Daley and Mayer 1986; Baker et al 1995). OSSE's have also been labeled identical twin experiments. SSE's have been called fraternal twin experiments (Williamson and Kasahara 1971; Kasahara and Williamson 1972).

To our knowledge the Monte Carlo method is the only available method to evaluate an integral in a very high dimensional phase space (e.g., Press et al. 1992). By "very high" we mean dimension D higher than 10. Non Monte-Carlo methods, with order $O(N^2)$ or higher order convergence will typically require of order M^D evaluations of the integrand (with M a constant between 5 and 100 depending on the required accuracy and N the number of evaluations of the integrand). The number M^D will quickly become excessive for any type of integrand. In the current example the evaluation of an integrand will involve an integration with a numerical forecast model. Non Monte-Carlo methods thus seem to be clearly not appropriate for high dimensional spaces. For problems of only moderate dimension, semi-random Monte Carlo methods might

be used. When used for very high dimensional phase spaces, most of these methods will give a performance comparable to pure random Monte Carlo methods. For the meteorological application at hand the dimension would seem to be at least 10000 (which is roughly the number of available observations). This implies that one has to use some type of Monte Carlo method. A Monte Carlo method will not work for an integrand that is strongly peaked in a priori unknown regions. There is some indication that the problem is indeed of this type (Oortwijn and Barkmeijer 1995). If this is indeed the case then one might as well give up as it would seem unlikely that we can ever determine the exact location of these spots. The accuracy of a Monte-Calo method improves with the root of the number of evaluations of the integrand $O(\sqrt{N})$. This rather slow convergence is offset by the rather nice fact that useful accuracy may already be obtained with very small ensembles (N about 10) (Leith 1971; Houtekamer and Derome 1995).

The currently popular optimal perturbation method (e.g., Palmer et al. 1993) may be considered a semi-random Monte Carlo method. It is based on the assumption that the phase space that is relevant for medium-range forecasting is relatively low-dimensional. To make this be the case the relevant space is restricted to the Northern Hemisphere extratropics or some subregion thereof (e.g., Houtekamer 1993). Recent studies (e.g., Palmer et al. 1993) seem to indicate that the typical dimension is of order 100 rather than of order 5. It would seem then that the optimal perturbation method, even if one would get it to work properly, would not be any more efficient than a random Monte Carlo method. As optimal perturbations are less general (being determined for some region and for some forecast time and obeying some constraint on the initial error) the applications of an optimal perturbation ensemble would seem to be limited to medium-range forecasting.

An alternative method is the breeding method (Toth and Kalnay 1993). This method may be considered a simplified implementation of the system simulation method. Historically it is more correct to consider the system simulation method to be a generalization of the breeding method. The generalization considers in particular the evolution of errors during the analysis step and the model errors.

A comparison of the optimal perturbation method, the breeding method and the Monte Carlo method is discussed in Houtekamer and Derome (1995). It is shown by these authors that medium-range ensemble forecasts of rather comparable quality can be obtained with these three methods in the context of a T21 3-level quasi-geostrophic model.

A system simulation method is preferable if one decides to build a forecasting system around it. How we think this should be done is the main subject of this paper.

1.2 The relevance of maximum initial growth rates

It has often been stated in favor of other ensemble prediction methods, such as the breeding method (Toth and Kalnay 1993; Houtekamer and Derome 1994) or the optimal perturbation method (Palmer et al. 1993), that initial error fields should have the fastest possible growth. This would be so because one can not properly sample both the growing and the now-growing perturbations in a high dimensional space. In the authors' opinion this argument is based on a misconception of the problem. In a representative ensemble one has

both the growing and the non-growing perturbations. Both sets of perturbations have proper amplitudes in the ensemble. The growing perturbations suffer in no way from the presence of some additional non-growing components.

To understand why this would be so we now pose an analogue question, of which the reader can easily find the answer himself. Suppose that, for the current state of the atmosphere, the following are available:

1. The estimation of the atmospheric state delivered by a state-of-the-art assimilation cycle.
2. The 8 leading coefficients of an EOF expansion of the above analysis.

One also has the state-of-the-art forecast model used to obtain the above analysis and a tool to reproduce an atmospheric state from the EOF coefficients. Which of the above two states should best be used to obtain a medium-range forecast?

In our opinion an SSE-perturbation corresponds with the state-of-the-art analysis, while the other methods, giving only the most rapidly growing components of the initial error, correspond to the EOF expansion.

The generation of an ensemble of N SSE-perturbations, as proposed in this paper, is about a factor two more expensive as the generation of a similar N-member ensemble of bred perturbations. This is because we do not use opposing initial perturbations as motivated by Houtekamer et al. (1995). The cost of the additional analysis step is rather minor as detailed in section 4.1. The cost of the determination of optimal perturbations would seem higher as the cost of the other methods as a significant number of forward and backward integrations is needed (possibly done at a lower resolution). For the SSE-method, the breeding method and presumably also for the optimal perturbation methods, only a minor fraction of the numerical costs of the ensemble forecast are for the generation of the initial perturbations. It would seem then that the above analogy is justified. We note that a comparison of the cost of the methods is very difficult as the methods are implemented on different sets of platforms.

1.3 What do we need to know?

In order to be able to perturb randomly the uncertain aspects of the forecast model we need first to identify and quantify these aspects. A preliminary selection of these aspects has been the main part of the development work of the RPN-CMC ensemble prediction system (Houtekamer et al. 1995). Fortunately, as we will see, the power of an SSE is that it allows for a validation of its components.

Clearly the observations on which an analysis is (partly) based are not free of error. This is known (e.g., Hollingsworth and Lönnberg 1986; Lönnberg and Hollingsworth 1986) and more or less satisfactorily accounted for in most operational data-assimilation systems. We will simply use the knowledge of observational errors that has been incorporated into the CMC operational data assimilation system (Mitchell et al. 1995).

A forecast also uses some surface fields such as albedo, roughness length and sea surface temperature. Little is known about the accuracy of these fields. We will perturb them in agreement with mostly ill-documented expert opinions of their accuracy. The field of soil-moisture would also seem to be important as the forecast model is sensitive to this field. However, no estimate of the accuracy of this field was available. We decided not to perturb it at all instead of just perturbing it arbitrarily. The accuracy of the surface fields, in particular soil moisture, is currently being studied in more detail at RPN and CMC.

Finally we will consider the error in the forecast model itself. A forecast model consists of a dynamical core and a number of modules to parametrize physical processes. We will assume here that the dynamical core treating the integration of the model equations, discussed in Ritchie and Beaudoin (1994), has achieved a state of perfection. Model errors are then only caused by the physical parametrizations made necessary by the limited resolution of the forecast model. We assume that those parametrizations that were first introduced in forecast models describe the most important processes. We assume also that these most important parametrizations describe the biggest part of the model error. To identify these parametrizations and the problems with them a literature study of model development has been performed. The rather subjective nature of this study should later be corrected by incoming validation results.

It would seem from the above that one hardly knows what uncertainties in the forecasting process are primarily responsible for the forecast error. This uncertainty inevitably slows down the progress in meteorological forecasting. We think it is time for a systematic study. For this study to lead home we will need to have a system that can signal its own short-comings.

1.4 The system simulation approach

We will presently give a general overview of how all knowledge on sources of error can be combined in a System Simulation Experiment (SSE). In Fig. 1, that pictures an SSE, one recognizes the entire forecasting system with the word perturbed added at the appropriate places. Beginning at the top we have the observations. These are randomly perturbed in agreement with their error statistics. The perturbed observations are input to the 6-hour data-assimilation cycle that makes up the top-half of Fig. 1. To start the assimilation cycle we use the operationally available first guess and the perturbed observations, i.e., observations to which we add random errors. As only the observations have been perturbed, the resulting perturbed analysis differs from the non-perturbed analysis only where observations are available. The perturbed analysis is integrated with model version i, where we have in principle as many versions of the model as we have members in the ensemble. The resulting 6-hour forecast serves as a perturbed first guess for the assimilation of the next set of perturbed observations. We note that the first-guess perturbation contains a contribution from model error and from observational error that has evolved during 6 hours with the model dynamics. As the cycling proceeds, the SSE difference patterns reflect the weaknesses of the model and the observing system as well as characteristics of the dynamics. After 4 days, statistics as referred to in section 4.2.1 remain reasonably stable. One thus has to cycle for 4 days before one can confidently use the ensemble statistics. To make the SSE complete, we have also perturbed the surface fields for roughness length, sea-surface temperature and albedo. These fields are perturbed at the beginning of the experiment and then kept constant. In principle,

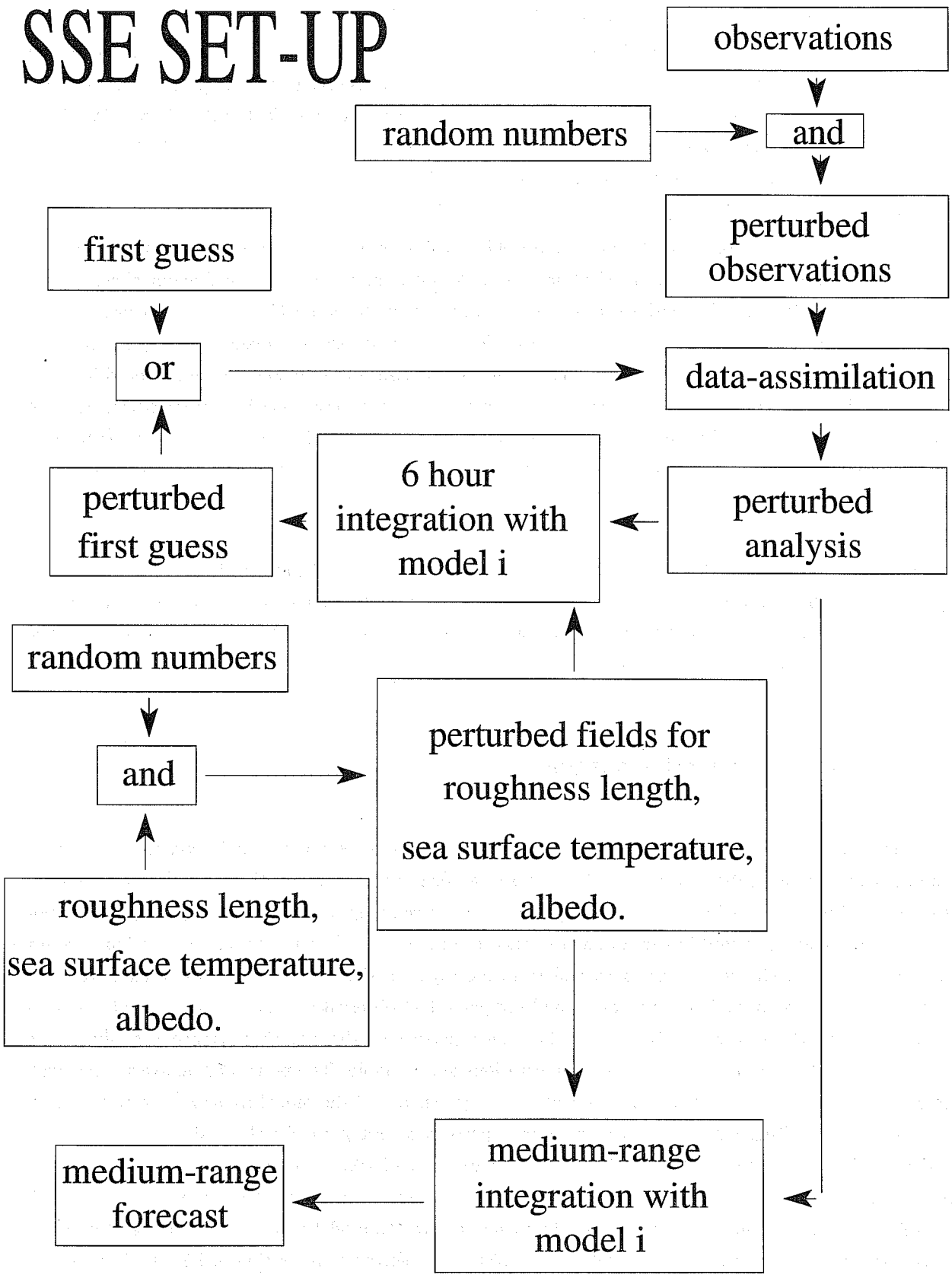

Figure 1: Current design of the system simulation experiment

we should let these perturbations vary slowly (on a time scale of about one month). We note that, except when starting up the perturbed assimilation cycle, the availability of an unperturbed control assimilation cycle is not required.

Going down in Fig. 1 we come to the forecast part. Here the medium-range forecast is started from the perturbed analysis. The forecast is done using the same forecast model i as has been used for the assimilation cycle and using the same perturbed surface fields. Repeating all the above with different random numbers for different ensemble members i, we obtain the ensemble prediction.

In the following sections we explain in detail our strategies for perturbing the observations, the model and the surface fields.

2 PERTURBATION OF THE OBSERVATIONS

In this section we discuss how to generate random perturbations to observations and surface fields. Observations are used by the data-assimilation procedure. This procedure takes a weighted mean of the first guess field and of the observations. To compute the weights the variances and covariances of the observations are used. We can thus assume that proper error statistics for observations are already available from the current data assimilation system. We will consider observations to be either scalars (with an error variance) or vectors (with an error covariance). We will discuss in detail how to randomly perturb observations in agreement with the available statistical knowledge. The perturbation of the surface fields is somewhat more complicated as one has to take horizontal correlations into account.

What follows in this section is a description of the algorithms that we used to generate all kinds of random perturbations. This may allow a more rapid development phase for future SSE's. We do not present any tools to validate the basic statistical knowledge that is used by the algorithms. We speculate that the observational error, in all the forms discussed in this section, is better known than anything else in the SSE.

2.1 Perturbation of scalars

Many observations y_{obs} are simple scalars. Their observational error $y - y_{obs}$ does not depend on the error in any other observation. It is fairly standard to assume that the probability distributions $p(y)$ for y is normal (Gaussian) around the observed value y_{obs}:

$$p(y)dy = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-y_{obs})^2/2\sigma^2} dy = (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp\left[-\frac{1}{2}(y - y_{obs})\sigma^{-2}(y - y_{obs})\right] dy, \quad (1)$$

where σ is the variance of the distribution and the notation of the extreme right-hand side is such that the correspondence with the multi-dimensional case (Eq. 2) will be clear.

To draw a random number from a Gaussian distribution one may start from a random number from a homogeneous distribution and subsequently perform a transformation (Press et al. 1992, p 279) to a standard normal distribution (i.e. $\sigma = 1$). The random number is subsequently multiplied with σ and added to the observed value y_{obs} to serve as a perturbed observation.

The normal distribution has some nice properties. In particular we have the central limit theorem (e.g., Kendall and Stuart 1977) which states that a linear sum of a large number of random arbitrarily distributed numbers will be close to normally distributed. In other words if one starts some computational procedure with non-normally distributed errors then the final errors may well be normally distributed, giving the same result as one would have had if one had started with normally distributed numbers.

One problem with the normal distribution is that out-liers (observations with enormous errors) are very unlikely as can be seen from substitution of a large error in Eq. 1. In practice sometimes something is very wrong with observations, so that the distribution would be non-normal. However it is not clear what this distribution would be and one may also think that quality control checks will remove the gross errors (Lorenc 1984).

2.2 Perturbation of vectors

Some instruments, such as radiosondes, report values at regular positions. The observation forms a vector of p values and the errors in the vector are correlated. It is again standard to assume that the errors have a multinormal distribution around the vector y_{obs} with covariance matrix V. The n-variate probability density p is given by (e.g., Kendall and Stuart 1977):

$$p(y)dy = (2\pi)^{-\frac{1}{2}n} \mid V^{-1} \mid^{\frac{1}{2}} \exp\left[-\frac{1}{2}(y - y_{obs})V^{-1}(y - y_{obs})\right] dy. \tag{2}$$

To draw a random vector from this distribution one will determine the eigenvectors v_i of V and the corresponding eigenvalues λ_i^2. We have the following relation:

$$V = \sum_{i=1}^{n} \lambda_i^2 v_i v_i^T. \tag{3}$$

The eigenvectors are the principal axes of the error ellipsoid that is defined by the covariance matrix. Random projections d_i along the axes will result in a random draw y from the multi-dimensional distribution.

$$y = y_{obs} + \sum_{i=1}^{n} d_i \lambda_i v_i. \tag{4}$$

Here the numbers d_i have been drawn from a standard normal distribution. It is easy to verify that an infinite ensemble of random vectors y has the covariance matrix V as it should.

2.3 Perturbation of fields

We will presently discuss an approximate method to generate random fields having, after averaging over many cases, a certain horizontal correlation length and a prescribed variance field. In principle one might generate a covariance matrix corresponding to this information, perform an eigenvector analysis as described above, generate a random field by randomly projecting on these vectors and sum the result (see Eq. 4). Unfortunately the numerical cost of an eigenvector analysis scales like the third power of the number of horizontal gridpoints. This direct method may thus not be feasible for high resolution grids.

If the horizontal correlations decrease fast enough with distance, then an approximate method to generate random fields can be used. The approximation consists of prescribing the correlation function only if the correlation is above a certain threshold value. The statistical theory (Epstein 1969 and e.g., Kendall and Stuart 1979) is described below.

2.3.1 *Theory of multiple correlations*

Suppose we have random values x_i at the points $i = 2, 3, \cdots, p$. These random numbers are a realization of a random field with variances σ_i and horizontal correlation function ρ. We would like to generate a random number x_1 at point $i = 1$. The unconditional (i.e. when not knowing the values at the other points) variance at this point equals σ_1 and the mean of the unconditional Gaussian distribution equals zero. The horizontal correlation between point $i = 1$ and any of the above points is above the minimum value ρ_{min}. The problem then is to describe the conditional probability distribution for x_1 given the values $x_i, i = 2, 3, \cdots, p$.

For the conditional mean $E(x_1|x_2, \cdots, x_p)$ and the conditional variance $\sigma^2_{1,2\cdots p}$ we have (Kendall and Stuart 1979):

$$\frac{E(x_1|x_2, \cdots, x_p)}{\sigma_1} = -\sum_{j=2}^{p} \frac{C^{-1,1j}}{C^{-1,11}} \frac{x_j}{\sigma_j} \tag{5}$$

$$\frac{\sigma^2_{1,2\cdots p}}{\sigma_1^2} = \frac{1}{C^{-1,11}}. \tag{6}$$

Here C is the matrix with the correlations $\rho(i,j)$ between points i and j:

$$C = \begin{pmatrix} \rho(1,1)=1 & \rho(1,2) & \cdots & \rho(1,p) \\ \rho(2,1) & \rho(2,2)=1 & \cdots & \rho(2,p) \\ \vdots & \vdots & & \vdots \\ \rho(p,1) & \rho(p,2) & \cdots & \rho(p,p)=1 \end{pmatrix}. \tag{7}$$

For Eqs. 5 and 6 we need only to have the first row $C^{-1,1*}$ of the inverse C^{-1} of C. To obtain the first row of C^{-1} we first use the fact that the inverse of a matrix times the matrix equals the identity (Eq. 8). We then write out the equation for the first row of C^{-1} in (Eq. 9) and finally we use the fact that C is symmetric to obtain Eq. 10.

$$C^{-1}C = I \Rightarrow \tag{8}$$

$$C^{-1,1*}C = e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Leftrightarrow \tag{9}$$

$$CC^{-1,1*T} = e_1 \tag{10}$$

The column vector $C^{-1,1*T}$ can now be found from the solution of Eq. 10. Note that this solution can be found using a Cholesky decomposition procedure (e.g., Press et al. 1992). Transposing the solution vector gives the first row $C^{-1,1*}$ of the inverse of C. The conditional mean and the conditional variance for x_1 can then be determined from Eqs. 5 and 6. We note that the solution of Eq. 10 is of order p^3 which is acceptable only if p is sufficiently small.

2.3.2 *Generation of an idealized random perturbation field*

Given the above theory for conditional distributions we are ready to generate a random perturbation field. We assume that the variance is constant equal to unity and that the horizontal correlation function has been specified. The gridpoints have been numbered from 1 to N_G.

For the first gridpoint we have no prior information available. We can simply draw a random number from a standard normal distribution. For subsequent gridpoints i we do the following in the order $i = 2, \cdots, N_G$.

1. List the points (with $j < i$) at which, according to the specified correlation function, the correlation with point i is higher than some minimum value ρ_{min}.
2. Determine the conditional distribution at point i given the information at the listed points.

3. Draw a random number from a normal distribution with the above determined conditional mean and variance. This number is the sought after value at point i.

We note here that the numbering of the points of the grid, for instance following meridians, can be done such that the number of required Cholesky decompositions is much smaller than N_G. The computational cost is a function of ρ_{min}. Here one might decide not to prescribe the correlation function at distances where the function itself is no longer certain (in general at large distances correlations are close to zero with a large relative error). One might also accept the function as it is and investigate the convergence in the numerical solution. One might for instance see how $C^{-1,11}$ depends on ρ_{min} (see Eq. 6). For ρ_{min} close to unity this term will be close to unity. With decreasing values of ρ_{min} its value will increase. The amount of increase quantifies the impact of the decreasing ρ_{min} on the conditional variance. After some experimentation one is likely to find, as we did, that the numerical cost of the generation of a random perturbation field is negligible.

Finally we note that an alternative method for the generation of random fields has been described by Evensen (1994). Here the random field is obtained from a random expansion in spherical harmonics which is then transformed to a grid using fast fourier transforms.

2.3.3 *Random geophysical fields*

The above tools have been used to generate random perturbation fields for the sea surface temperature, for the albedo and for the logarithm of the roughness length. This has been described in detail in Houtekamer et al. (1995). Here we will discuss as an example the albedo field.

It has been assumed that the albedo has a horizontal error correlation given by the function:

$$\rho(r) = \frac{1}{(\alpha+1)}((1+cr+\frac{c^2r^2}{3})\exp^{-cr} + \alpha(1+\frac{cr}{N}+\frac{c^2r^2}{3N^2})\exp^{-cr/N}), \tag{11}$$

where α=0.2, N=3 and $c = 0.01km^{-1}$.

It has further been assumed that the rms error in the albedo σ_{AL} is proportional to the albedo AL itself:

$$\sigma_{AL} = 0.116AL + 0.07. \tag{12}$$

The numerical procedure is as follows:

- A random field with unit variance and correlations given by Eq. 11 is determined as explained in section 2.3.2. The generation of many such random fields, using different sequences of random numbers, would show that Eq. 11 is indeed respected but only for $\rho > \rho_{min}$.

- The rms error field for the albedo σ_{AL} is computed from the values of the albedo and Eq. 12.
- The random field determined in step 1 is multiplied with the rms-field determined in step 2. The multiplication would be with the same field of rms-values for a hypothetical set of many random fields. This multiplication has thus no effect on the already imposed correlation function.
- The perturbation field is added to the unperturbed albedo. The resulting perturbed albedo is restricted to be between zero and one.

2.3.4 *Observations with a horizontal correlation*

Observations from satellites have horizontal correlations. This can be accounted for in roughly the same way as in 2.3.2. This time however the list of points is not a list of gridpoints but a list of points where satellite observations are available.

In the case where satellite observations are available at virtually all the gridpoints one may decide to generate global error fields as above and then interpolate the errors to the observation locations. Such an approach will be followed for the new HUMSAT(Humidity from SATellite)-data that are expected to become available to the CMC data-assimilation system. It could also be followed whenever Lidar measured winds become available.

3 PERTURBATION OF THE FORECAST MODEL

3.1 Family relations

In building up a family of forecast models, we have decided upon discrete model perturbations of fixed amplitude. For some physical process A we try to find a plus and minus 1-σ perturbation. For instance for convection we postulate that the Kuo-scheme is a plus 1-σ perturbation of the ideal convection scheme. We also postulate that the Manabe-scheme is a minus 1-σ perturbation to this ideal convection scheme. As we cannot make a smooth transition from the Kuo scheme to the Manabe scheme we are unable to generate a say 0.73 σ perturbation. For convection we use the Kuo scheme for half of the ensemble members and the Manabe scheme for the other half (see Houtekamer et al. (1995) for more details).

One may also have a physical process for which the modeled uncertainty is only in the choice of a parameter. Here one may think of for instance the horizontal diffusion. We assume that its uncertainty is only in the value of the diffusion coefficient. We divided the coefficient by two for half of the cases (Houtekamer et al. 1995). In principle we might have decided to assign random values, drawn from a proper distribution, to this coefficient. In fact it is rather likely that this will be done for some parameters for future versions of the ensemble prediction system.

In a traditional validation of multiple model options one would use $N = 2^m$ forecasts to simultaneously validate m model aspects. One can then directly see which combination of options leads to the best forecast. Operational constraints seem to fix the maximum size of the ensemble to $N = 8$. We could thus validate only $m = 3$ different options for the forecast model. This is unfortunate as we can easily think of more than $m = 3$ options and limiting ourselves to $m = 3$ would thus lead to a perhaps unnecessary underestimation of the model error.

Actually an N-member ensemble spans N-directions in the phase space of atmospheric states. One direction is defined by the ensemble mean. This informs us about the mean quality of the models and can not be used for the validation of an individual model option. However, the remaining $N - 1$ directions might somehow be associated with $N - 1$ model options (to be listed later on). The projection of the forecast error on these $N - 1$ probably non-orthogonal directions gives the "model error of the day". Performing such projections on a daily basis one will most likely come in a position to suggest model improvements.

If one performs half of the model integrations with option A_+ and the other half with A_- then one should, after validation, be able to say which one is the better one. For this to be possible the different model options should somehow be independently distributed over the perturbed models. Otherwise we would not be able to attribute a certain effect to a certain option.

The selection of option i and j, with $i, j = 1, \cdots, N - 1$ is such that looking at the $N/2$ model versions with option i selected we find $N/4$ selections of option j and $N/4$ non-selections of option j, with $i \neq j$. An example of what the resulting model versions may look like has been presented in Table 1. Thus we see for instance that the model to integrate the second ensemble member has option 3, 6 and 7 selected. We do evidently not have all possible 2^7 combinations of options. Thus we have for instance no perturbed model with only the options 1,2,3,4,5 and 6 selected. To have all possible combinations we would need $N = 64 = 2^7$ ensemble members.

Table 1: selection of model options for an 8-member ensemble

	module number						
model no	1	2	3	4	5	6	7
1	+	+	+	+	+	+	+
2	-	-	+	-	-	+	+
3	+	-	-	+	-	-	+
4	-	+	-	-	+	-	+
5	+	+	+	-	-	-	-
6	-	-	+	+	+	-	-
7	+	-	-	-	+	+	-
8	-	+	-	+	-	+	-

A table like Table 1 can be constructed recursively for ensemble size $N = 2^m, m = 1, 2, \cdots$. We start with a 2-member ensemble (Table 2). Option 1 is selected for the first member and not for the second.

Table 2: selection of model options for a 2-member ensemble

	module number
model no	1
1	+
2	-

To construct a table for a 4-member ensemble we refer to the contents of the second column of Table 2 as a block. This block thus contains a plus above a minus. We take this block and move it sideways to form the upper half of the third column for the table (Table 3) for a 4-member ensemble. We move the block downwards to form the last half of the new second column, which completes the second column. The complete the third column we mirror the contents of the block (plus becoming minus and minus becoming plus) and put it in the lower part of the third column. Finally we make a fourth column by putting plusses in the upper part and minuses in the lower part.

Table 3: selection of model options for a 4-member ensemble

	module number		
model no	1	2	3
1	+	+	+
2	-	-	+
3	+	-	-
4	-	+	-

Defining the second to fourth columns of Table 3 to be a block and performing the same operations we finally obtain Table 1.

We note that the above way of creating family relations is well suited to our current situation where the number of available options is roughly equal to the number of ensemble members that can be integrated. It is expected that the number of available options will increase with time. It is hoped that the same will happen with the number of ensemble members that can be run.

3.2 Choice of options

Evidently we still have to decide which model perturbations will be associated with the above mentioned 7 options. For preliminary testing we decided on the following set. This set is a modification and extension of the set discussed in Houtekamer et al. (1995).

1. Vertical Diffusion. If the option is selected the model always has a significant vertical diffusion. If the option is not selected the vertical diffusion is a hundred times weaker in stable conditions above the

planetary boundary layer.

2. Not used. This option has as yet not been used. This will allow us to calibrate the effect of an (absent) model change by looking exclusively at the effect of random observational errors. These errors will cause some uncertainty (noise) in the evaluation of a module. This noise-level can hopefully be estimated from using a dummy-module as proposed here.

3. Intensity of Gravity Wave Drag. If the option is selected the gravity wave drag is 3 times more intense (everywhere) than when it is not selected.

4. Convection. If this option is selected the Kuo scheme for convection and a relatively advanced radiation scheme are used. If it is not selected the Manabe scheme is used for convection and we use a simpler radiation scheme.

5. Version of Gravity Wave Drag code. If this option is not selected we use a version of the gravity wave drag code where wave drag begins at relatively low Froude numbers (McLandress and McFarlane 1993). If it is selected the wave drag starts higher up in the atmosphere (McFarlane 1987).

6. Orography. If the option is selected we use a 1.4-σ envelope orography, otherwise we use a mean orography.

7. Horizontal Diffusion. If this option is selected the e-folding time for horizontal diffusion at the smallest scale (wavenumber 63) is equal to 29 hours. If the option is not selected this e-folding time equals 58 hours.

The above list is clearly subject to change. It will be modified when validation results and/or new model versions become available.

3.3 Validation

So far the process of model improvement has often been a 1-dimensional process. One suspects that some aspect of the model may be improved upon and thus writes an improved subroutine for this aspect. One will then organize some tests where the performance of the old model is compared with the new one. If these tests show that the change leads to an improvement, one accepts the change. Following this process forecast models are gradually improved.

The above process of model improvement is perfect if all subroutines of the model have independent effects on the forecast error. This case has been illustrated in Fig. 2a, which shows the case where in a two-dimensional space we are confronted with a model error vector (1,1). Proposed change A will cause a vector in the direction (1,0). Proposed change B will cause a vector in the direction (0,1). If change A is tuned first then we apply the vector with coefficient 1, leading us to (1,0) and a remaining error (0,1). Applying change B will take care of the remaining model error, leading to a perfect model. If we had tuned process B first we would also have come up with a perfect model.

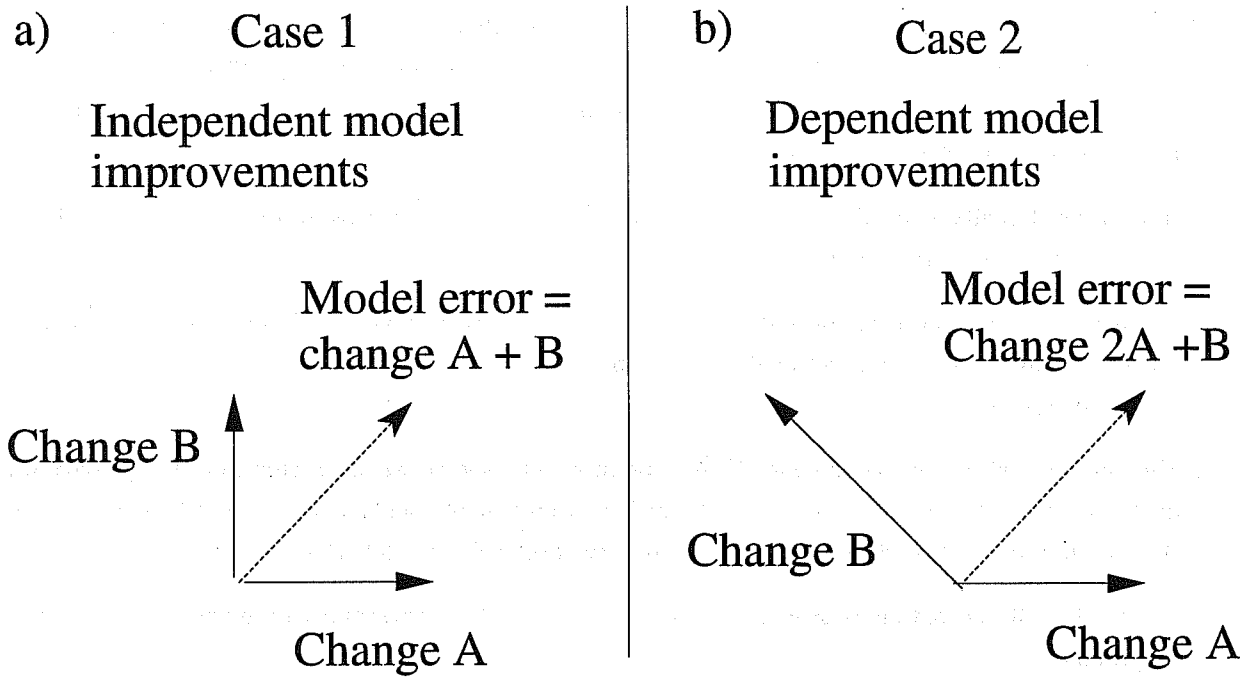

Figure 2: The case of dependent and independent model changes

If, instead, the effects of different subroutines are correlated, then improving the model in the above way becomes a very difficult process. This can be seen from the example shown in Fig. 2b. Here we have model error (1,1). Proposed change A will cause a vector in the direction (1,0). Proposed change B will cause a vector in the direction (-1,1). Applying change A first will bring us as before to (1,0). Proposed change B will then be implemented with coefficient 0.5 bringing us to (0.5,0.5) with a remaining model error of (0.5,0.5). If we first apply change B then we see that it will receive coefficient zero. Change B is thus rejected. Then applying change A brings us to (1,0) with remaining model error (0,1). In this case of a simultaneous consideration of process A and B we would have given amplitude 2 to change A and amplitude 1 to the apparently perfect model change B. The remaining model error would have been zero.

With the above we have demonstrated that one may easily get stuck when applying the traditional method of incremental model improvement. We argue that it is best to test as many simultaneous model improvements as possible. Mathematically the description is as follows:

The impact of model option $i, i = 1, \cdots, N-1$ at forecast time t can be estimated from:

$$\overline{f}(t) = \frac{1}{N}\sum_{k=1}^{N} f_k(t)$$

$$u_i(t) = \frac{2}{N}\sum_{k=1}^{N} f_k(t)\delta_{k,i} - \overline{f}(t) \qquad \text{(average influence of option } i\text{)},$$

with $\delta_{k,i} = 1$ if option i has been selected for use in model k and $\delta_{k,i} = 0$ otherwise. Here the individual forecasts $f_k(t), k = 1, \cdots, N$ have first been averaged to obtain an ensemble mean forecast $\overline{f}(t)$. This is then subtracted from the mean over the individual forecast done with option $i, i = 1, \cdots, N-1$ selected. Note that since by design any given option i appears (is selected) for half of the forecast, that average is computed over $N/2$ forecasts. This gives the $N-1$ difference fields $u_i(t), i = 1, \cdots, N-1$ that are indicative of the impact of using a certain option. The impact of not using option i is given by $-u_i(t), i = 1, \cdots, N-1$.

Let us write $a(t)$ for the verifying analysis of the forecast at time t. We would like to decompose the forecast error $\overline{f}(t) - a(t)$ of the ensemble mean into the non-orthogonal components $u_i(t), i = 1, \cdots, N-1$ in order to see which model component was responsible for the forecast error.

In fact we would like to have an improved forecast $\tilde{f}(t)$:

$$\tilde{f}(t) = \overline{f}(t) + \sum_{j=1}^{N-1} c_j(t) u_j(t), \tag{13}$$

where the new forecast error $\tilde{f}(t) - a(t)$ no longer has a projection on $u_i(t), i = 1, \cdots, N-1$. This gives the system of $N-1$ linear equations:

$$(a(t) - \overline{f}(t), u_i(t)) = \sum_{j=1}^{N-1} c_j(t)(u_j(t), u_i(t)), i = 1, \cdots, N-1. \tag{14}$$

The projection coefficients $c_j(t), j = 1, \cdots, N-1$ can be solved from the system of $N-1$ equations. If $c_j(t) = 1$ then apparently the selection of option j has a beneficial effect on the forecast. If on the other hand $c_j(t) = -1$ then option j had better not be selected.

In designing the ensemble forecasting system we had an option which consisted of having a gravity wave drag scheme, as in the operational model, and not selecting the option consisted of switching the gravity wave drag off. To compute the coefficient c we used an inner product $(,)$ for the 500 mb geopotential height error north of $30°N$. We averaged the coefficients for day 1,2,3,4 and 5 for three ensemble forecasts, one in April 1994, one in July 1994 and one in December 1994. We found a mean coefficient of $c = 0.8$. Thus it seemed from this preliminary evaluation that switching off the gravity wave drag had a rather negative impact on the ensemble forecast and that the optimal coefficient for the gravity wave drag would be close to the one used in the current operational model. As we do not think the gravity wave drag is free of error we now have as option $i = 3$ a coefficient with 1.5 times the current operational value. When the option is not selected we use 0.5 times the current operational value.

For the other model changes the initial evaluation was not so clear. For the horizontal diffusion (currently option 7) we seem to find that the optimal diffusion coefficient should increase with forecast time, as already suggested by Jakimov et al. (1992). The optimal value would be between the extreme values chosen here. For convection (currently option 4) we find an optimal coefficient of 0.18 with considerable spread around this

value. For the orography we found a coefficient of -0.02. This seems to indicate that an ideal model should have a 0.7 σ envelope orography. Option 5 and 1 were not yet available for these preliminary tests.

The above describes a rather optimistic view of the use of the ensemble forecast for model improvement. So far we did not consider the effect of the initial errors and errors in the surface fields on the forecast error. We assume that after averaging over many cases these effects will cancel out. This may indeed be so for the determination of optimal coefficients. It is much less obvious that we will be able to cleanly estimate the remaining uncertainty in a parameter. On the other hand if the ensemble forecast is used operationally for, say, one year then one has an enormous data base to evaluate the $N-2$ model changes. We propose to maintain one dummy model change (currently option 2) to be able to study the estimation problems caused by the random error components. Ideally we would have for c_2 a narrow distribution around zero but this may not be so due to the estimation problems.

4 THE DATA-ASSIMILATION SYSTEM

The global data assimilation system is as described by Mitchell et al (1995). It operates in a 6-hour cycle with a data cutoff time of plus or minus 3 hours. One cycle consists of performing (a) a 6-h forecast to produce a first guess, (b) taking a weighted mean of the first guess and the observations using a statistical interpolation scheme (often referred to as optimum interpolation or OI), and (c) a nonlinear normal mode initialization to eliminate spurious large-amplitude gravity waves in the subsequent 6-hour forecast.

4.1 Parallel assimilation cycles

Using a unique formulation of the OI procedure in step (b) allows for an important computational economy. For the 8 parallel assimilation cycles we have to evaluate the linear equation:

$$A_i = F_i + W(O_i - HF_i), i = 1, \cdots, 8, \tag{15}$$

where F_i is the perturbed first guess i, O_i is the i^{th} set of perturbed observations, A_i is the perturbed analysis number i, H is an operator to interpolate the first guess to the position of the observations and W determines the weight given to the observations. In practice the calculation of the weights W is the computationally most demanding part of the assimilation. However the weights W do not necessarily depend on the index i. This has been exploited to obtain a reduction of the numerical cost of step (b) with a factor of (currently) 5.

We can have the same weights, for different ensemble members, only if the data-selection and the quality control deliver exactly the same sets of observations (with differences only in the reported values). To force the analyses to use exactly the same sets of observations we introduced some changes in step (b). We now

perform part of the quality control for aircraft observations just before these observations are perturbed. This permits the removal of a quality control algorithm in the superobing procedure, which combines some very nearby observations (Lorenc 1981). As it happens, many more super observations are formed due to this modification. We also decided not to perturb a surface pressure if this will cause the observation to be assigned to different analysis levels for different members of the ensemble. This would otherwise happen about 20 times every 6 hours. Finally we use the quality control decisions made for the first set of observations O_1 and the first guess F_1 for the 7 other sets as well. We did not have the impression that quality control decisions would otherwise be very different between the different cycles.

A similar reduction cannot be obtained with a 3D-VAR or 4D-VAR scheme because these schemes are iterative. The iterations start at rather different first guesses and they lead to rather different analyses.

4.2 Feedback towards the OI-scheme

From the SSE one has every 6 hours a set of simulated 6 hour forecast errors available. One may think then that some ensemble statistics may be used to improve the operational data assimilation system. Here one would use the variances as supplied by the SSE together with an analytical correlation model (Evensen 1995). One may be even more ambitious and replace the correlation model by correlations obtained from the ensemble (Evensen 1994).

4.2.1 *Estimating parameters for a traditional scheme*

From the difference between the N first guess fields F_i and the ensemble mean $\overline{F}$ first guess field we can estimate the statistical properties of the first guess error. From the ensemble one could estimate the error variances and the correlation lengths. This application has been extensively discussed by Houtekamer et al. (1995). In general they found for the Northern Hemisphere extra tropics that errors at the lower levels were at larger scales as compared to the OI-functions. At the higher levels the errors were found to be at smaller scales as compared to the OI-functions. It would seem that in this case the OI used the better functions. For the tropics the height errors were found to be at larger scales. This ensemble result is supported by a theoretical study (Daley 1993). For the Southern Hemisphere extra tropics virtually the same correlation lengths were found. In general Houtekamer et al. (1995) found that error amplitudes were underestimated, from which they concluded that some error sources had been underestimated in the SSE. In particular they suggested increasing the differences between the different models.

Following this conclusion and considering the general bad performance of the models without gravity wave drag the perturbation to the gravity wave drag module (to switch it off) has been considerably revised (a more advanced gravity wave drag routine is now used as a perturbation). The uncertainty in magnitude is now considered to be around the currently operational values. In addition a perturbation to the vertical diffusion scheme has been introduced (module 1). As the experiment has not yet been redone it is not clear to what

extent we have resolved our problems.

The information present in the ensemble could also be used to estimate variances for a 3D-VAR system. This has not yet been done.

4.2.2 *Towards dynamical forecast error covariances*

A generally recognized problem of the OI-scheme is its inability to account for the dynamics of the day. It is possible to introduce this dynamical information using statistics from the ensemble. With this information the scheme becomes four-dimensional and we will thus refer to it as a 4D-OI scheme.

In a traditional OI-application the weights W for Eq. 15 are calculated from the equation:

$$W = F_M H^T (H F_M H^T + O_M)^{-1}. \tag{16}$$

Here the matrix O_M describes the error covariances of the observation vector O and the matrix F_M describes the forecast error covariances as obtained from a predetermined covariance model. The dimension of the matrix that has to be inverted equals the length N_O of the observation vector. If the matrix inversion is performed to compute W we have a process of order N_O^3. To make this feasible it is standard to cut the observation vector into small pieces and to consider a number of local problems. The currently operational 3D-OI scheme at CMC uses at most $N_O = 24$ observations at a time and performs a separate analysis at every gridpoint. This procedure allows for a correlation model that is very dependent on geographical location and height.

One may wonder if the term HF_MH^T might be directly obtained from the ensemble of N perturbed first guess fields. Here we have to consider 3 cases.

1. $N < N_O$. The number of ensemble members is smaller than the number of observations. It is impossible to obtain a covariance matrix of rank N_O from the first guess fields. According to HF_MH^T we have $N_O - N$ directions with zero first guess error. This implies that the first guess can be corrected only in N directions by the N_O observations. Consequently, the first guess will diverge from the actual state as the analysis cycle proceeds (filter divergence).
2. $N = N_O$. The number of ensemble members equals the number of observations. An analysis can be performed. However the analysis will be optimal for the errors as simulated by the ensemble. It will not be optimal for the true forecast error. Thus the properties of the ensemble will start to diverge from the properties of the actual errors (filter divergence) and the analysis will not be as successful as suspected from the small differences between the N analyses.
3. $N > N_O$. As the ensemble size increases the problem slowly becomes better posed. We presume that at $N = 2N_O$ the problem will be well posed. However, this is something that should be studied further.

From the above it is clear that the use of the ensemble statistics for dynamically relevant weights has to be done with great care. One might decide on a rather conservative use of the ensemble as follows:

$$F_\Sigma = \alpha F_{ENS} + (1-\alpha)F_M + \beta F_M, \tag{17}$$

$$\alpha = min(N/2N_O, 1), \tag{18}$$

$$W = F_\Sigma H^T (H F_\Sigma H^T + O_M)^{-1}. \tag{19}$$

Here F_Σ is the covariance matrix for the total forecast error, F_{ENS} is covariance matrix obtained from the ensemble, β is a tuneable parameter and W is the new matrix with weights. With the parameters α and β we protect the system (at least partially) against filter divergence. We suspect that we can confidently obtain statistics from the ensemble only when the number of ensemble members is twice the number of observations.

It may well be that the number of ensemble members N is large enough to maintain a representative ensemble, but that the system simulation suffers from some other problem. In particular it may happen that some sources of error have been underestimated (see section 4.2.1). This will lead to a spread in the ensemble that is too small. Consequently the 4D-OI will give too much confidence to the first guess and not enough to the observations. The 4D-OI will then be sub-optimal. For this reason one might introduce some feedback as described below.

In general the variance of the difference between the observations and the first guess should equal the sum of the error variances of the observations and of the first guess. We should thus have, if everything works properly:

$$\sum_{j=1}^{N_O} ((H\overline{F})_j - O_j)^2 = \text{Trace}(O_M + H F_\Sigma H^T), \tag{20}$$

with $\overline{F}$ being the ensemble mean first guess, O_j unperturbed observation number j and F_Σ determined with $\beta = 0$. It may also be that we have to choose a parameter $\beta > 0$ to make Eq. 20 hold. This would imply that to the extent that apparently we do not understand the forecast error we give confidence to the traditional correlation model.

As we mentioned, the use of the ensemble statistics for the analysis of the same ensemble has the risk of filter divergence. In other words, as the system responds to the ensemble statistics the ensemble is no longer representative. We would reduce this problem if we used two ensembles and we analyzed the members of one ensemble with the statistics obtained from the other.

It would seem that the above scheme would allow the use of fully dynamically determined forecast error covariances in the analysis scheme as soon as we have the ability to perform about 48 ensemble integrations. The above scheme would appear to demand only minor modifications in the current formulation of the 3D-OI. As such it is in sharp contrast with the 4D-VAR method. We also note that the presented scheme, in a natural

way, takes account of the model error. This is not possible with the 4D-VAR method. Another advantage of the presented scheme is that it does not require a separate tool to transport covariance information from one analysis step to the next.

A somewhat different scheme, based on roughly the same ideas has been presented by Kalnay and Toth (1994). They propose to perform the analysis in two steps. First a number of regional analyses is performed to regionally correct the first guess in the directions present in the ensemble with amplitudes as suggested by the observations. Then in a subsequent step a global 3D-VAR algorithm is applied. In our scheme a regional correction is ensured by the selection of only 24 observations to apply a local correction.

5 SPREAD IN THE ENSEMBLE

The traditional use of ensembles is for medium-range forecasting. In general one has observed that the spread in the ensemble is smaller than the difference between the forecast and the verifying analysis. The conclusion has often been that one should find more rapidly growing perturbations or that one should give them bigger initial amplitudes. As there is little information on the actual projections of the initial unstable structures on the analysis error one is in fact free to do so.

With an SSE one does not have such freedom. We claim to perturb all the uncertain aspects of the forecasting process. If nevertheless the spread in the ensemble is too small then it must necessarily be because we did not properly model some sources of forecast error.

Table 4: RMS-spread and forecast error

	error in meters															
day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
8 models	6	8	11	14	18	24	30	36	42	46	49	52	55	58	59	60
1 model	6	6	9	11	14	18	24	30	34	40	45	49	52	54	55	56
error	9	15	23	34	42	43	48	59	69	76	80	81	77	71	70	71

In Table 4 we show the rms-difference between the individual ensemble members and the ensemble mean. These differences have been computed for the height of the 500 mb level at the northern hemisphere. Results are from forecast day 0 to 15 for the forecast starting on 0 UT 18 April 1994. The first line gives the rms-spread for the experiment where we have 8 different models. These 8 models differ in option 4,6 and 7 as well as in a gravity wave drag which is switched on or off. The second line describes a similar experiment but with a unique version of the model. Finally the last line gives the differences between a T63 control forecast and a high resolution T119 analysis.

We observe that the differences between the ensemble members grow almost exponentially until forecast day 6. After that the growth rate is falling off until saturation occurs at about forecast day 15. When we perturb

the model the rms-spread is about 25 % larger than with a unique version of the model. The differences between the actual forecast error and the rms-spread in the ensemble is roughly a factor two between forecast day one and six. Apparently we have modeled only half of the present day medium-range forecast error. With the incorporation of the new options 1 and 5 we will obtain a larger spread but we will most likely have to continue looking for additional not yet modeled sources of error.

6 DISCUSSION

We have presented a scheme for the generation of an ensemble of initial perturbations using what we call a system simulation experiment. As explained in the introduction, such experiments have a long history in meteorology. However, to the best of the authors' knowledge, this is the first time that such as scheme is being proposed to simultaneously improve the analysis, to perform an ensemble forecast and to arrive at a number of model validations.

The scheme, which mathematically would be referred to as a Monte Carlo method, attempts to produce a set of representative error fields at the initial time of a forecast. For this purpose a number of parallel assimilation cycles is run using representative perturbations to the observations, the surface fields and the forecast model.

The scheme can be used for the simultaneous validation of a number of model options. A validation of different options for horizontal diffusion, deep convection, gravity wave drag and orography has already been performed while developing the SSE. It was then decided to modify the options for gravity wave drag and to add a new option for vertical diffusion. A thorough validation of the different options is planned in the near future. It will then also be estimated to what extent maintaining these options is of importance for the ensemble forecast. Future developments of the scheme will be guided by the outcome of these experiments.

As a by-product, the scheme delivers a set of supposedly representative first-guess error fields. As the error fields should be representative, their statistical properties can be, and have been, validated against statistical information used by the statistical interpolation (3D-OI) data-assimilation system. Such a validation cannot meaningfully be done for other ensemble prediction methods that only attempt to obtain the growing part of the initial error. Ultimately we expect that all the analysis statistics can be obtained from a big enough ensemble. We expect that the scheme will replace 4D-VAR applications if these become available.

When used for ensemble prediction, we observe that the simulated forecast errors are deficient in amplitude by about a factor two. This happens probably because the different forecast models are still too similar. This situation should improve as new versions of the RPN-CMC forecast model become available.

In conclusion we would like to state that, although many details need to be sorted out, we have a most promising method for ensemble prediction. Unfortunately we have been able to perform only a small number of ensemble forecasts so that we can, as yet, not substantiate our claims.

Acknowledgements

In developing the ensemble prediction method the authors profited from the knowledge and help of Peter Bartello, Christiane Beaudoin, Bruce Brasnett, Gilbert Brunet, Cécilien Charette, Clément Chouinard, Yves Delage, Bernard Dugas, Luc Fillion, Louis Garand, Claude Girard, Eugenia Kalnay, Norman McFarlane, Herschell Mitchell and Zoltan Toth.

REFERENCES

Arnold, C.P. Jr. and C.H. Dey, 1986: Observing-Systems Simulation Experiments: past, present and future. *Bull. Amer. Meteor. Soc.*, **67**, 687–695.

Baker, W.E., G.D. Emmitt, F. Robertson, R.M. Atlas, J.E. Molinari, D.A. Bowdle, J. Paegle, R.M. Hardesty, R.T. Menzies, T.N. Krishnamurti, R.A. Brown, M.J. Post, J.R. Anderson, A.C. Lorenc and J. Elroy, 1995: Lidar-measured winds from space: a key component for weather and climate prediction. *Bull. Amer. Meteor. Soc.*, **76**, 869–888.

Charney, J, M. Halem and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci*, **26**, 1160–1163.

Daley, R. and T. Mayer, 1986: Estimates of global analysis error from the global weather experiment observational network. *Mon. Wea. Rev.*, **114**, 1642–1653.

Daley, R., 1993: Atmospheric Data Assimilation on the Equatorial Beta Plane. *Atmos.-Ocean*,**31**, 421–450.

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

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Calo methods to forecast error statistics. *J. Geophys. Res.*, **99C5**, 10143–10162.

Evensen, G., 1995: Advanced sequential methods with Nonlinear dynamics. *Second international symposium on assimilation of observations in meteorology and oceanography, Vol I, 13-17 March 1995, Tokyo, Japan*, 147–152.

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

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

Houtekamer, P.L., and J. Derome, 1994: Prediction experiments with two-member ensembles. *Mon. Wea. Rev.*, **122**, 2179–2191.

Houtekamer, P.L., and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, **123**, 2181–2196.

Houtekamer, P.L., L. Lefaivre, J. Derome, H. Ritchie and H.L. Mitchell, 1995: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, accepted.

Jakimov, G, E. Yakimiw, and A. Robert, 1992: An implicit formulation for horizontal diffusion in gridpoint models. *Mon. Wea. Rev.*, **120**, 124–130.

Kalnay, E. and Z. Toth, 1994: Removing growing errors in the analysis cycle. *Proceedings of the 10th conference on numerical weather prediction, 18-22 july, 1994*, Portland, OR, USA, 212–215.

Kasahara, A. and D. Williamson, 1972: Evaluation of tropical wind and reference pressure experiments for observing systems. *Tellus*, **24**, 100–115.

Kendall, Sir., M., and A. Stuart, 1977: *Kendall's advanced theory of statistics, 4 th ed, Vol 1: Distribution Theory*, Mac Millan Publishing Co., Inc., New York. 473pp.

Kendall, Sir., M., and A. Stuart, 1979: *Kendall's advanced theory of statistics, 4 th ed, Vol 2: Inference and relationship*, Mac Millan Publishing Co., Inc., New York. 749pp.

Leith, C.E., 1971: Atmospheric predictability and two-dimensional turbulence, *J. Atmos. Sci.*, **28**, 145–161.

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.

Lorenc, A.C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 701-721.

Lorenc, A.C., 1984: Analysis methods for the quality control of observations, in *Proceedings of the ECMWF workshop on the use and quality control of meteorological observations for numerical weather prediction, 6-9 Nov. 1984*, pp. 397–428.

McFarlane, N.A., 1987: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere, *J. Atmos. Sci.*, **44**, 1775–1800.

McLandress, C. and N.A. McFarlane, 1993: Interactions between orographic gravity wave drag and forced stationary planetary waves in the winter northern hemisphere middle atmosphere, *J. Atmos. Sci.*, **50**, 1966–1990.

Mitchell, H.L., C. Chouinard, C. Charette, R. Hogue and S.J. Lambert, 1995: Impact of a revised analysis algorithm on an operational data assimilation system, *Mon. Wea. Rev.*, submitted.

Oortwijn, J, and J. Barkmeijer, 1995: Perturbations which optimally trigger weather regimes. *J. Atmos. Sci.*, submitted.

Palmer, T., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *ECMWF Seminar Proc.*, Validation of models over Europe, Volume I, 7-11 Sept 1992, 21-66.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, 1992: Numerical recipes in fortran. The art of scientific computing. Second Ed. Cambridge University Press.

Ritchie, H. and C. Beaudoin, 1994: Approximations and sensitivity experiments with a baroclinic semi-lagrangian spectral model. *Mon. Wea. Rev.*, **122**, 2391-2399.

Smagorinsky, J. 1969: Problems and promises of deterministic extended range forecasting. *Bull. Amer. Meteorol. Soc.*, **50**, 286-311.

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteorol. Soc.*, **74**, 2317-2330.

Williamson, D.L., and A. Kasahara, 1971: adaptation of meteorological variables forced by updating. *J. Atmos. Sci.*, **28**, 1313–1324.