DIRECT PREDICTION OF ERROR VARIANCES BY THE TANGENT LINEAR MODEL: A WAY TO FORECAST UNCERTAINTY IN THE SHORT RANGE?

Philippe Veyre
Météo France
Paris, France

Abstract

The direct forecasting of variances of short range errors is studied with the technique of the tangent linear model and its adjoint. In spite of the computational effort, this can be tractable in an operational context for local forecasts: a forecast for a given variable at a given point needs only two integrations: a tangent linear one and an adjoint one. The linear hypothesis is acceptable for synoptic scales up to 72 hours.

Four sets of ten experiments each have been tried for the 24 hour range. Three are made with barotropic models, including one with initial variances from the operational OI, the fourth uses a three level baroclinic model. The forecast variances for the 500hPa geopotential height are compared with actual errors produced by operational models. As reference forecasts, the RMSE and the zonal mean of the RMSE, estimated with the other nine forecast errors, are used to assess the information provided by the method. Quantitative estimation of its skill is given by the continuous Ranked Probability Score.

In the Northern Hemisphere, the forecasts of variances are better than the reference in every case, especially with the initial variances from OI. They are statistically significant for the three barotropic cases. On the Southern Hemisphere, they are conclusive only for the barotropic case which uses the ECMWF analysis and forecasts.

1 INTRODUCTION

Weather forecast is probabilistic by essence: we don't know the initial state precisely, models are only approximations, and the evolution of the atmosphere is characterized by instability and chaos. The usual forecast of the "mean" or the most probable flow gives only a partial answer. We must also give the higher moments of the probability law of the future state. The formulation of final products by forecasters as well as the decisions taken by final users are linked to this information. To get this with numerical tools, three techniques have been suggested: statistics, Monte Carlo and direct forecast of variances.

The statistical method was first tried by *Hoffman and Kalnay* (1983). This is only an extension of the "Model Output Statistics" to standard scores (RMSE, ACC) on large domains. The predictands are differences between successive forecasts or some parameters which characterize the weather type. *Palmer and Tibaldi* (1988) and *Molteni and Palmer* (1988) have studied this scheme extensively. They conclude that this kind of forecast is more a reference for other methods than a usable tool.

Monte Carlo forecasts use a set of numerical integrations from randomly perturbed initial states. Apart from the problem of huge computations, the main question is to construct the initial sample according to the initial uncertainty, and if possible, to choose the most significant elements, in order to decrease the size of the ensemble. This technique is still under development, especially for the medium range.

The third technique is also the oldest, and the most straightforward: the direct numerical forecast of variances. *Epstein* (1967) and *Pitcher* (1977) have tried it with some success, but the computations were far too large, and the second order closure assumption is not very efficient. However, independently, specialists in data assimilation tried to apply the Kalman filter technique to simplified NWP models (*Ghil et al.* , 1981). According to this scheme, it is possible to estimate and forecast the mean state of a stochastic system as well as variances and covariances as long as the error has a linear evolution. *Lacarra and Talagrand* (1988) have shown that this seems to work for the short range.

This technique naturally combines data assimilation and forecast, and model uncertainty can be taken into account. The tangent linear model used to compute variances can be a simplified version of the one used for the direct, nonlinear forecast, in contrast with the Monte

Carlo technique. Above all, the use of the adjoint model allows to calculate variances one by one, and thus to obtain at a low cost the most interesting ones. After the presentation of this method, we shall show different experiments of 24 hours variance forecasts with simple models and simple initial covariances. Results will be compared with actual errors from operational models.

2 DESCRIPTION OF THE METHOD

2.1 Evolution equation for the covariance matrix

The evolution of the atmosphere is given symbolically by the nonlinear equation:

$$X(t) = M_{t,0}(X(0)) \tag{1}$$

$M_{t,0}$ is the resolvant of the dynamic equations and is called 'direct model'. The evolution of a small perturbation $W(t)$ is determined by the linearized counterpart of (1):

$$W(t) = R_{t,0}.W(0) \tag{2}$$

where $R_{t,0}$ is the differential of $M_{t,0}$ at $X(0)$. This is a linear operator depending on the forecast trajectory: $(X(0), \cdots, X(t))$. It is called 'tangent linear model'.

Now, $W(0)$ and thus $W(t)$ are random, unbiased. The covariance matrix at time t is simply given by the following formulas ($W(t)$ is a column vector):

$$P(t) = \left\langle W(t).W(t)^T \right\rangle \tag{3}$$

$$P(t) = R_{t,0}.\left\langle W(0).W(0)^T \right\rangle.R_{t,0}^T \tag{4}$$

$$P(t) = R_{t,0}.P(0).R_{t,0}^T \tag{5}$$

$$P(t) = R_{t,0}.(R_{t,0}.P(0))^T \tag{6}$$

This formulation supposes the model is perfect. If this is not the case, a matrix Q which represents the intrinsic uncertainty of the model must be added to $P(t)$ at each time step. This possibility has been neglected in this study. According to (6), the computation of variances needs the computation of the full matrix, i. e. $2N$ linear integrations, if N is the number of variables. However, the use of (5) allows to get variances one by one, as explained

later. Here, we need the adjoint operator of R for the canonical scalar product (defined by $W(t)^T.W(t)$). This technique of the adjoint model has been described by *Talagrand and Courtier* (1987) and *Courtier and Talagrand* (1987) and is coded in the new ARPEGE/IFS model developed by METEO-FRANCE and ECMWF. If necessary, we must take care of the scalar product used for the coded adjoint. If B is the symmetrical matrix for the scalar product which defines the adjoint R^*, we simply have:

$$R^T = B.R^*.B^{-1} \tag{7}$$

For the case of ARPEGE, B is diagonal and the transformation is very easy. Now, if we want variable number i (a given field at a given point or for a given spectral component), we will apply (5) to a unit vector $U_i = (\delta_{ij})$. The covariance field with the forecast parameter i is obtained through the following operations:

- $R^T.U_i$: gives the "sensitivity" of the forecast at i to initial errors. According to the definition of the adjoint:

$$\left(R^T.U_i\right)^T.U_j = U_i^T.(R.U_j) \tag{8}$$

  the value of this sensitivity field at j gives the forecast response at i to a unit initial perturbation at j.

- $P(0).R^T.U_i$: gives the covariance between the forecast error at i and *initial* errors for any variable.

- $R.P(0).R^T.U_i$: gives the covariance between the forecast error at i and forecast errors for any variable.

Thus, we need only two integrations. This approach is very useful: in an operational context, the variance is needed only over the region of interest, and for the most informative variables.

We can easily generalize for post-processed variables, such as those given by statistical methods. If $y = V^T.X$ is the linearized version of this post-processing, the variance for y is given simply by:

$$\langle y'^2 \rangle = V^T.R.P(0).R^T.V \tag{9}$$

Here again, we need only two integrations: the vector U_i is simply replaced by the vector V, the "adjoint" of the post-processing. Obviously, this is reduced to one integration if $\langle y'^2 \rangle$ only is needed: $V^T.R$ is obtained directly from $R^T.V$. Of course, all these schemes are linked to $P(0)$. It could be given by a Kalman filter, but this is too expensive. Approximate estimations come from optimal interpolation (OI). Because $P(t)$ depends also on R and thus on the predicted trajectory, we can hope that the lack of accuracy for the initial matrix is not too annoying.

At last, we must notice that we can take a simplified version of R, even if we use a sophisticated model to calculate the trajectory (for instance, the operational forecast).

2.2 Limitations due to the linear hypothesis

The crucial hypothesis is that evolution of errors is governed by the tangent linear model. This problem has been extensively studied by *Lacarra and Talagrand*(1988) with a divergent barotropic model on an f-plane: the linear hypothesis is justified at least up to 48 hours for wavelengths greater than 640 km, and for initial errors of the order of magnitude of real initial errors. We try here to extend the range up to 72 hours, for synoptic and planetary scales (total wavenumber $n < 20$), on the sphere with barotropic and baroclinic models.

The model is a preliminary version of the future operational model ARPEGE/IFS. It is a spectral model with triangular truncation with filtered or divergent barotropic equations or baroclinic primitive equations . Each option has linear tangent and adjoint versions (provided that the trajectory of the basic state is given). All the experiments are adiabatic but with lateral diffusion.

The initial perturbations are differences between two initialized analyses taken randomly, rescaled in order to have $500hPa$ kinetic energy equal to $4m^2s^{-2}$, which corresponds to a geopotential height RMSE of 20m in middle latitudes. This prevents excitation of unrealistic gravity waves . In order to test the influence of the scale of the perturbation, we have defined by truncations large scale ($n < 11$) and small scale ($n > 20$) perturbations.

Filtered barotropic experiments are made with T63 resolution. The correlation coefficients for streamfunction between exact and linear perturbations are respectively 0.98, 0.92 and 0.86 for 24, 48 and 72 hours ranges. Kinetic energy spectra show that at 72 hours the

difference exact minus linear is larger than the exact perturbation itself for $n > 17$. This is also verified with T42 truncation and for large scale or small scale initial perturbation.

Baroclinic cases use a T42, 19 levels model with orography. Due to the lack of boundary layer parametrization, results are not significant for low levels. In spite of the presence of baroclinic instability, the properties are not very different from the barotropic case. Doubling time for kinetic energy are similar (about one day). Correlation coefficients for streamfunction are at 24h more than 0.95 for all levels apart the 32m one, at 48h more than 0.90 for $850hPa$ and above, at 72h more than 0.83 at $800hPa$ and above. The slopes of the regression lines are about one. The limit n where linear error equals exact perturbation energy at $500hPa$ is about 25 for 48h and 20 for 72h (fig.1). Maps (fig.2) show good agreement between linear and exact computations, especially for the positions of extrema. Their amplitudes can be overestimated, of course because of the overestimated energy in small scales. Results are similar with the large scale and small scale perturbations (limit n from 17 to 25 at 72h).

In short, our technique should be usable up to 72h, for scales greater than $n \approx 20$.

2.3 Comparison with the Monte Carlo technique

2.3.1 The experiment

Our aim is to study both methods in a simple, idealized context. We have used the T21 filtered barotropic model. The initial covariance matrix is taken diagonal in spectral space, and corresponds in physical space to a homogeneous variance for streamfunction and an isotropic gaussian correlation function (length scale 800km). The kinetic energy of error is $4m^2s^{-2}$ and the geopotential height RMSE is about 20m in middle latitudes.

According to this matrix, a gaussian sample of 1000 members is generated, in order to compare our method with the Monte Carlo one, with linear and exact integrations. The direct method requires here 968 linear integrations.

2.3.2 Numerical results

The results have been studied for different ranges and parameters (vorticity, streamfunction, geopotential height) (cf. fig.3 for the standard deviation of vorticity at 72h). As it is well known for barotropic instability, the development of errors is linked to the shear of the basic

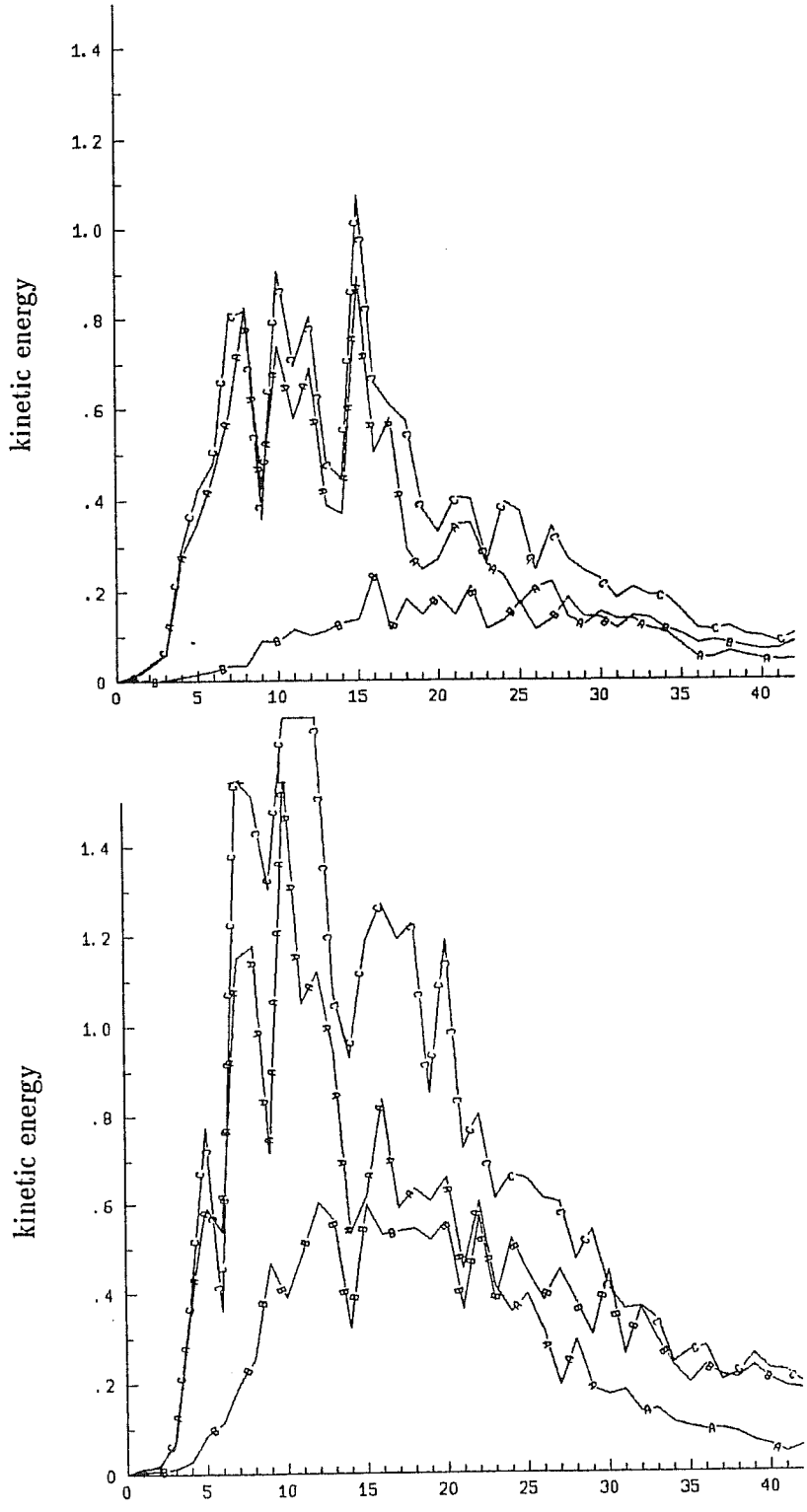
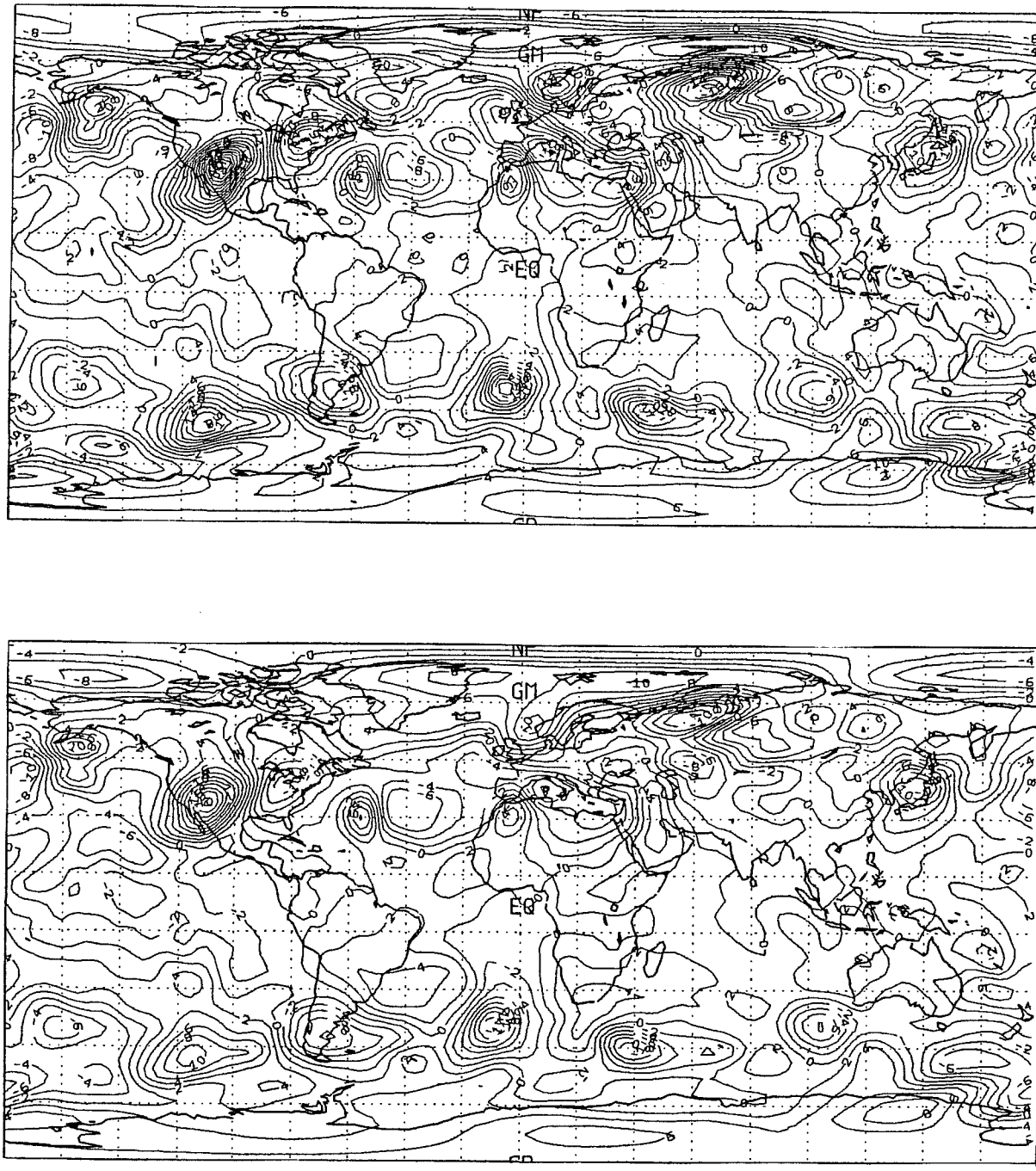


Fig. 1: 500hPa kinetic energy spectra for nonlinear (A) and linear (C) perturbations and for the difference (B). Top: 48h. Bottom: 72h (unit: $m^2 s^{-2}$).



- Fig. 2: 500hPa streamfunction of the perturbation for 72h. Top: linear perturbation. Bottom: nonlinear perturbation (unit: $10^6 m^2 s^{-1}$).

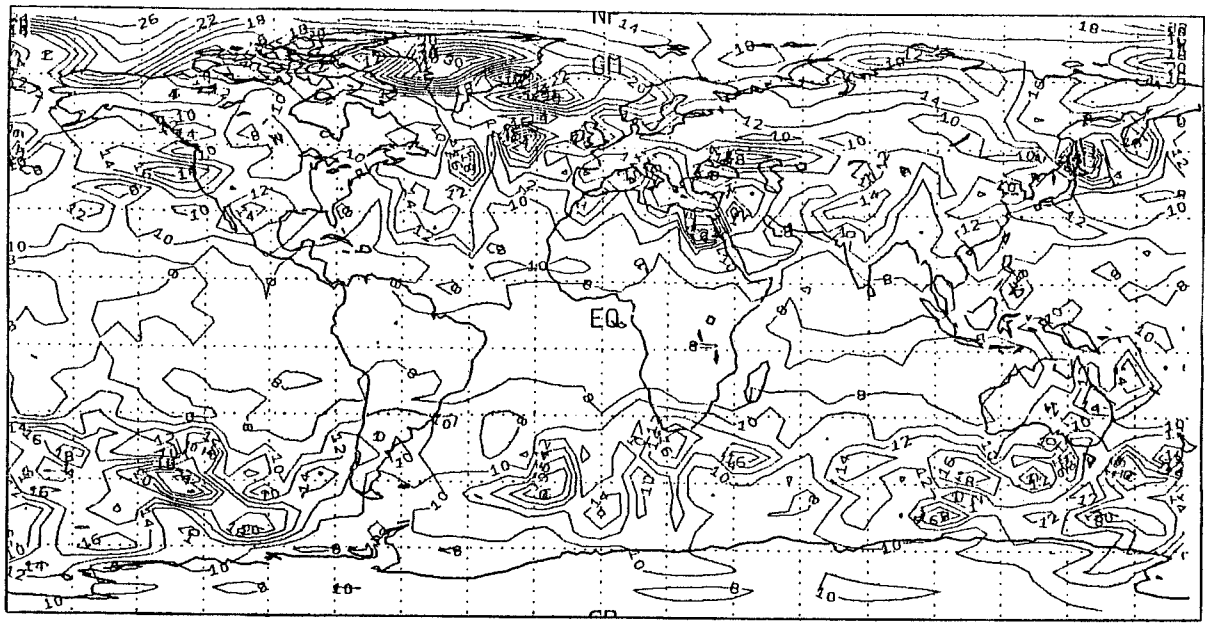
flow. More precisely, the largest variances are localized in regions of strong vorticity gradient, as suggested by *Thompson*(1986).

Due to the small truncation and the horizontal diffusion, the linearity is fully justified up to 72h . The relative differences for streamfunction standard deviations are about 2% for 24 and 48 hours and about 6.5% for 72 hours, i.e. the sampling error expected with 120 trials. We have verified that in Monte-Carlo technique the relative error of the estimation of standard deviation is $\frac{1}{\sqrt{2M}}$ where M is the sample population (i.e. a 10% error needs 50 integrations). Thus, Monte Carlo should be cheaper if we want all the variances. However, the way to build the ensemble is not obvious: the explicit diagonalization of the initial covariance matrix, necessary to generate the sample, can be unfeasible because of its size. And the use of perturbations with optimal growth (*Molteni and Palmer* ,1991) could bias the sample in an unrealistic manner.

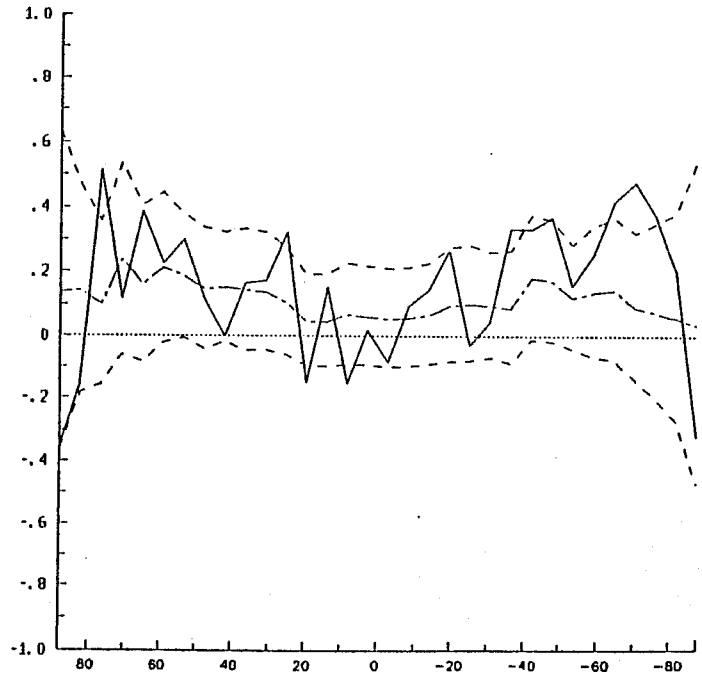
2.3.3 Verification against real data

Verification against real data seems obviously hazardous with such a simple forecast model and such a simple initialization. However, some features are consistent with observation: large errors in middle latitudes, particularly in the vicinity of strong winds, small errors in the intertropical zone: the model represents more or less the natural variability of the atmosphere. Moreover, for the 24h range, the forecast of the basic flow is not too bad. Consequently, we have compared for this range, and for each latitude, the correlation between absolute value of actual error $|e|$ and the forecast standard deviation σ_p with those obtained with the Monte Carlo sample (i.e. longitudinal correlations between the individual perturbations and the standard deviation of the sample). These are taken as references: they tell us what can be hoped even with a perfect knowledge of the probability law. In fig.4, we see that the actual correlation is within the range of one standard deviation of the Monte Carlo correlations. The results seem especially interesting in middle latitudes where the natural variability is not too low, and the length of a latitude circle not too short. Results for other ranges (48h, 72h) are of course less conclusive. This is probably linked to the poor quality of the direct model.

As a result, this first attempt is rather encouraging, but we must keep in mind that



- Fig. 3: 72h forecast of the standard deviation of the vorticity ,88/01/15 case (unit: $10^{-6}s^{-1}$).



- Fig. 4: Correlation between $|e|$ and σ_p (24h forecast of vorticity) for the actual error (solid) and the "errors" simulated by the Monte Carlo sample (mean: dash-dotted, mean $\pm$ standard deviation for the correlation: short dashed) (see text).

the verification is not an obvious question: even with a perfect forecast of the probability law, correlations with actual errors are about 0.2 with large fluctuations. Measuring the skill of variance forecast with this method could lead to misinterpretations: we don't forecast a particular realization of the error, but the probability law of the error. Specific scores must be used.

3 EXPERIMENTS

3.1 Description

In order to test the method with realistic conditions, four sets of ten 24h forecasts of variances each have been implemented and compared with actual errors from operational forecasts (ECMWF and the French model EMERAUDE):

- experiment A :
  - barotropic T21 model
  - initial situations and actual forecast errors from the ECMWF archives (88/12/01 , 88/12/15, 89/01/01 , 89/01/15 , 89/02/01 , 89/02/15 , 89/12/01 , 89/12/15 , 90/01/01 , 90/01/15)
  - initial matrix: as in 2-3 (kinetic energy: $4m^2s^{-2}$)
- experiment B :
  - T21 baroclinic model; three levels hybrid coordinate (200hPa, 500hPa, 830hPa); realistic orography; no physical parametrization.
  - nonlinear normal mode initialization for the basic state and its linearized counterpart for the variances
  - initial situations and actual errors from EMERAUDE archive (every tenth day from 90/12/01 until 91/03/01)
  - initial matrix: derived from the covariance model used for the French operational OI (horizontal and vertical correlation functions, use of hydrostatic and

geostrophic approximations). The standard deviations do not depend on longitude and are about 17m for the 500hPa geopotential height at 45N.

- forecast of variances for the 500hPa vorticity only. We use the adjoint method to get the covariance matrix between 500hPa vorticity and all the variables. Computing the full matrix would be ten times as expensive.

- experiment C : similar to B but with a barotropic model.

- experiment D : similar to C except for the initial matrix. Correlation functions are the same but the variances are taken from the operational OI. They are not homogeneous and depend on the density and the quality of the observations. This estimation is crude, and that is a way to test it.

Each forecast gives the matrix for the spectral components of vorticity (and more for the baroclinic case). It takes 968 integrations of the tangent linear model or its adjoint, i.e. for a CRAY2 computer about 800sCPU for the barotropic cases and 4000sCPU for the baroclinic one. The matrices are converted by the linear balance operator and the spectral-gridpoint operator into covariances in physical space for the 500hPa geopotential height.

3.2 Method of assessment

We have shown earlier that the simple correlation between forecast variances and actual errors is not very convincing. Therefore, we have chosen a score developped for probabilistic forecasts: the Ranked Probability Score (RPS). It was proposed by *Epstein*(1969) and *Murphy*(1971) for discrete, categorical forecasts. The generalization to the continuous case (*Matheson and Winkler*, 1976 and *Unger*, 1985) is straightforward. For a given observation and a given probabilistic forecast, the Continuous RPS is the quadratic distance between the forecast cumulative distribution $F_p(x)$ and the observed one $F_o(x)$(a step function):

$$CRPS = \int_{-\infty}^{+\infty} (F_p(x) - F_o(x))^2 dx \tag{10}$$

with

$$F_o(x) = \begin{cases} 0 & \text{if } x < x_o \\ 1 & \text{if } x \geq x_o \end{cases}$$

This score is defined for each point and each forecast and can be averaged over a domain or for a set of experiments. If the forecast is deterministic, we simply have $CRPS = |x_p - x_o|$ and the mean CRPS gives the Mean Absolute Error (MAE).

If $F_p(x)$ is gaussian centered on x_p , with σ_p as standard deviation, the formula becomes, with $e = x_p - x_o$:

$$CRPS(e, \sigma_p) = 2\sigma_p \left[\frac{e}{\sigma_p} \left(F\left(\frac{e}{\sigma_p}\right) - \frac{1}{2} \right) + G\left(\frac{e}{\sigma_p}\right) - \frac{1}{2\sqrt{\pi}} \right] \tag{11}$$

F is the cumulative density and G the distribution for the normalized gaussian law.

This allows us to compute the CRPS. In this case, the following properties can be shown:

- For a given e, the CRPS is minimum for $\sigma_p = 1.20\,|e|$. This "optimum" CRPS is given by $CRPS_{opt} = 0.595\,|e|$. In this case, the observed error must be known *a priori*.

- If x_p and x_o are two gaussian variables with the same mean (here x_o is not the observation, but the true distribution of the forecast error, that would be obtained by a perfect knowledge of the initial uncertainty and of the matrix Q characterizing the model error), $E[CRPS]$[1] is minimum if and only if $\sigma_p = \sigma_o$, and then:

$$E[CRPS] = \frac{1}{\sqrt{\pi}}\sigma_o = \frac{1}{\sqrt{2}}EAM = 1.1886\,E[CRPS_{opt}] \tag{12}$$

  Even with a perfect forecast, the CRPS is larger than the minimum.

- If the forecast of σ is not perfect, the difference between $E[CRPS]$ and $\frac{1}{\sqrt{2}}EAM$ gives the skill of this forecast. More precisely, if $\sigma_p = \sigma_o(1+\epsilon)$, ϵ gaussian variable with $\sigma_\epsilon^2 \ll 1$:

$$E[CRPS] = \frac{1}{\sqrt{2}}EAM\left(1 + \frac{\sigma_\epsilon^2}{4}\right) \tag{13}$$

- In this case, the standard deviation of this difference is also available. Moreover,if we have *two* independent forecasts of variance, the difference for the CRPS gives the difference of skill, and the statistical significance of the result can be assessed according to the standard deviation of this difference. If $D_{12} = CRPS_1 - CRPS_2$:

$$E[D_{12}] = \frac{1}{\sqrt{2}}EAM\left(\frac{\sigma_{\epsilon_1}^2 - \sigma_{\epsilon_2}^2}{4}\right) \tag{14}$$

[1] $E[.]$ is the expectation

$$\sigma(D_{12}) = 0.278\, EAM \sqrt{\sigma_{\epsilon_1}^2 + \sigma_{\epsilon_2}^2} \tag{15}$$

This formula must be divided by $\sqrt{M}$ if the estimation is made with M independent trials. This gives confidence intervals for the comparison of different methods of forecast.

As for any prediction scheme, a "reference" forecast must be chosen, as "climatology" or "persistence": we must prove that our forecasts are better then trivial forecasts. Because our experiments are for small samples, we have taken the RMSE estimated over the other nine situations, and the zonal mean of this RMSE. The first one is very crude due the small sample. However, we know that, even with large samples (one month or one season), this estimation is fluctuating very much, just because the weather types fluctuate as well. The zonal mean is more reliable in our case, and indeed gives better CRPS.

The second problem is that CRPS is sensitive to biases for the variance forecast, and it is difficult to avoid that because of the choice of the initial covariance matrix. In order to correct the mean level of variances, a kind of "model output statistics" is made for the zonal mean of σ_p:

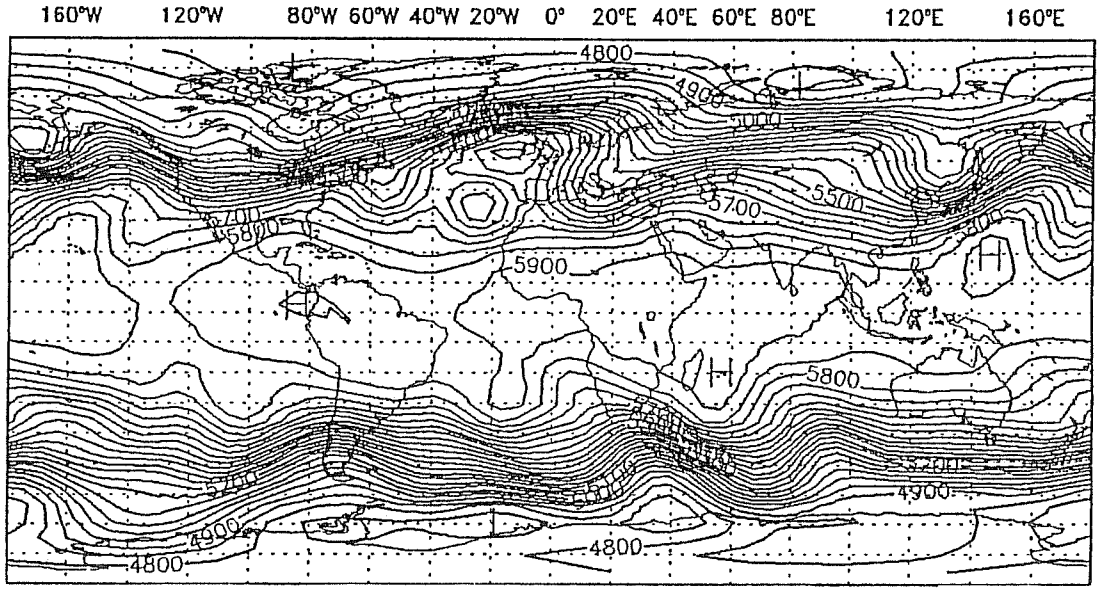
$$\tilde{\sigma}_p = K(lat).\sigma_p \tag{16}$$

where $K(lat)$ is the ratio of the zonal mean of RMSE and the zonal mean of the RMS of σ_p computed from the other nine cases. Therefore, the zonal means of the forecast should have minor biases. This can be considered as an indirect way to correct the initial matrix as well as to take into account the model error. The reference forecast is a "non-forecast": the uncertainty is constant over a latitude circle. Our forecast gives a spatial structure linked to the direct forecast.

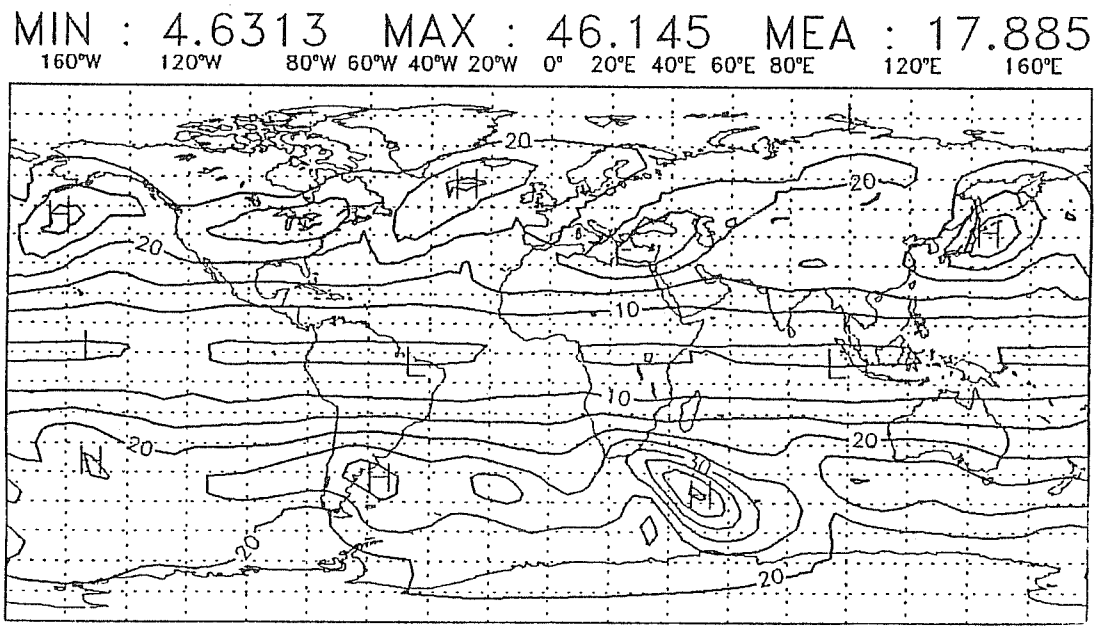
3.3 Results

An example of a forecast is given in fig.5, without any correction: the verifying analysis, the baroclinic forecast, the barotropic forecasts (case C and D) and the initial standard deviation used by case D. The errors grow according to the strength of the vorticity gradient, as mentioned above. The baroclinic and barotropic cases are very similar (we have verified that the direct forecast with the three level baroclinic model is actually better than the barotropic

• (a)

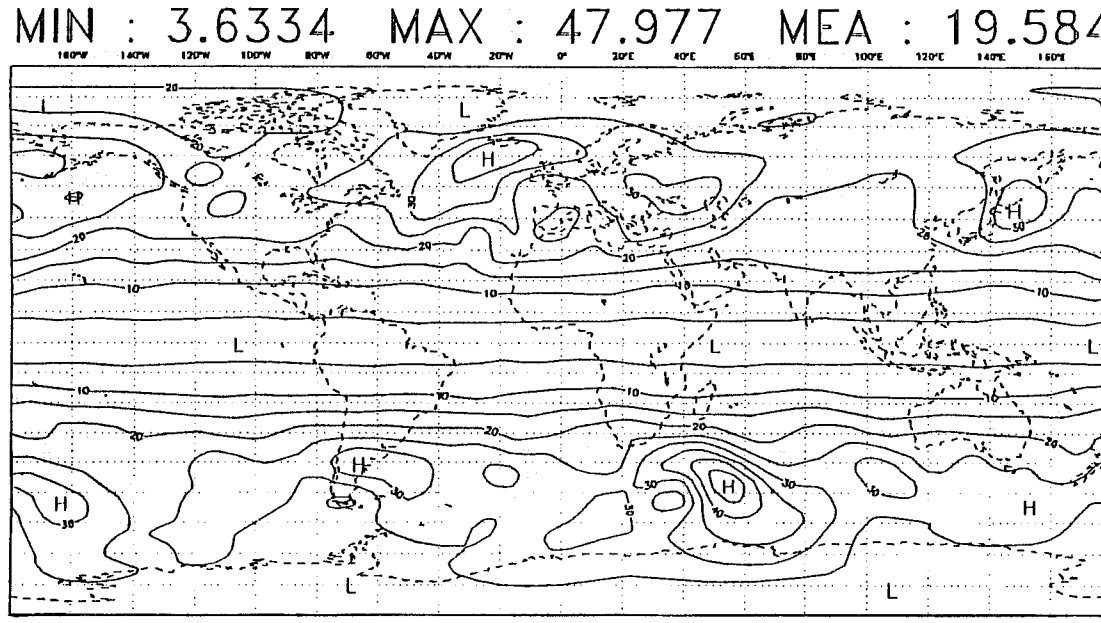


• (b)

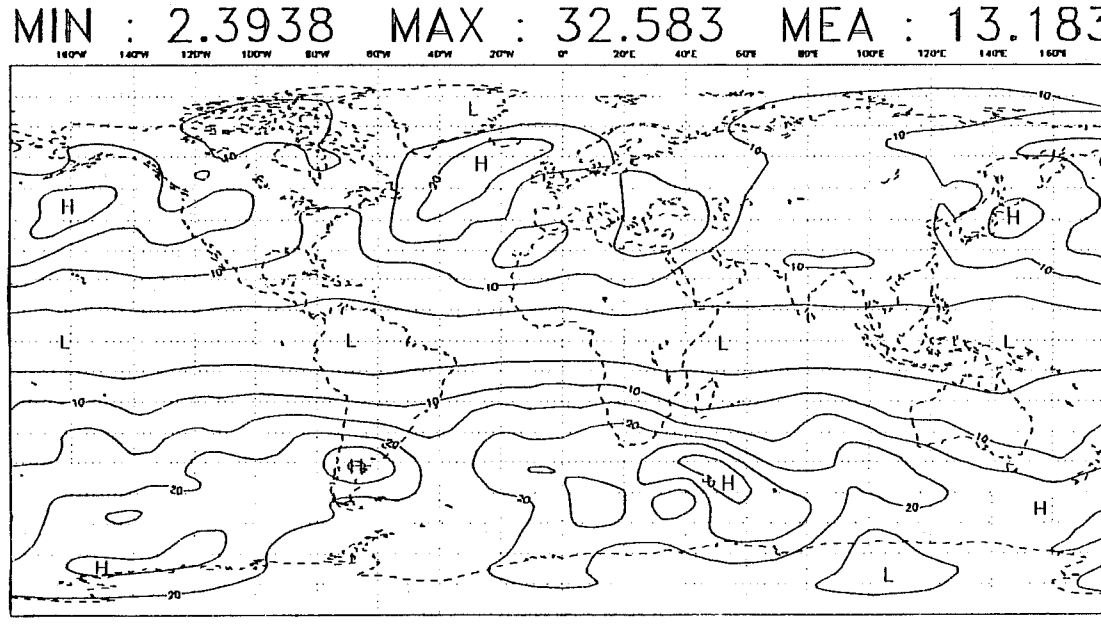


• Fig. 5: 24h forecasts of the standard deviation for the 500hPa geopotential height, 90/12/01 case (unit: m): (a) verifying analysis; (b) baroclinic forecast.

• (c)



• (d)



• Fig. 5 (cont.): (c) barotropic forecast (exp.C); (d) barotropic forecast (exp.D).

forecast, as expected). Our forecasts of variances are not sensitive to the specification of the vertical structures. As for the study of the linear hypothesis, the development of errors in complex flows seems mainly a barotropic phenomenon. In a baroclinic flow, the growth of errors is linked to the strength of the potential vorticity gradient (*Thompson*, 1988). In real flows, the scale of this quantity could be such that it has mainly a vorticity component (scale smaller than the Rossby radius of deformation). The second explanation is of course the choice of the initial matrix: the separability hypothesis for the correlation function favours barotropic structures for initial errors, as is well-known in optimal interpolation.

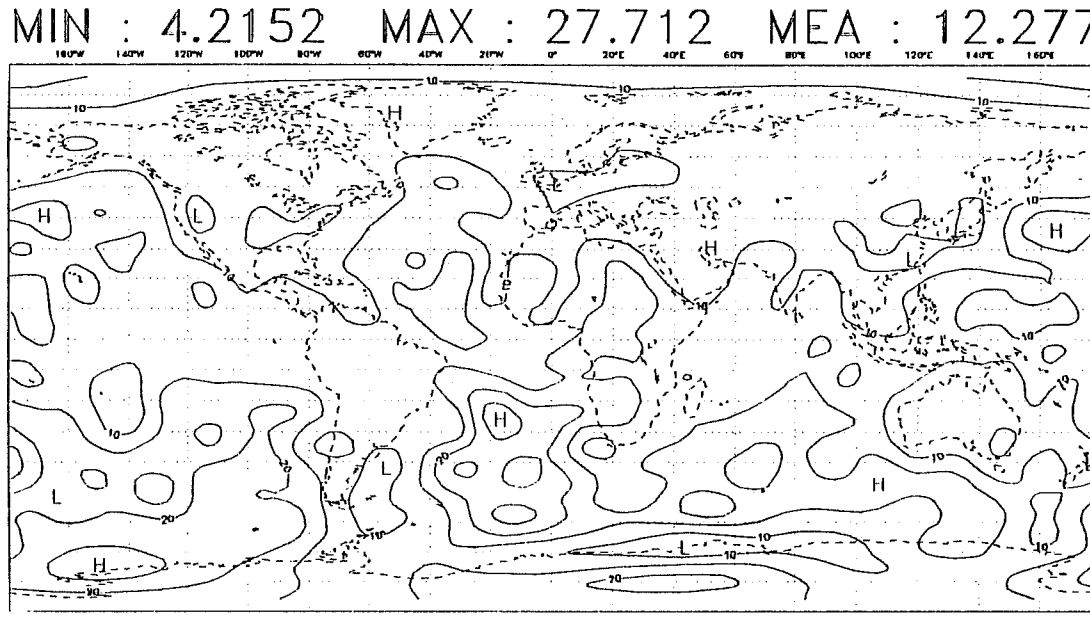
A comparison between experiments B and C is shown in fig.6: we use here the zonal correlations between actual squared errors and forecast variances with the RMSE estimated over the other nine cases as reference. These correlations are better than the reference for middle latitudes. Their order of magnitude (0.2) is similar to those obtained with the perfect Monte Carlo forecasts (fig.4).

In order to give more convincing results, the CRPS has been used over large domains. Here, we shall take the latitudes north of 30N and south of 30S. The reference is the zonal mean RMSE and the forecast standard deviation is corrected according to (16). The confidence intervals come from (14) and (15) with $M = 1200$. M is the equivalent number of independent, individual forecasts for each set of experiments. The number of degrees of freedom can be estimated by the number of spectral components (484) for the whole globe, and therefore about 120 for each domain, and there are ten independent situations. A crude estimate for M is sufficient. We show below the skill for the forecast (S_f) and the reference (S_r), evaluated by (13) and the ratio r which gives the confidence level:

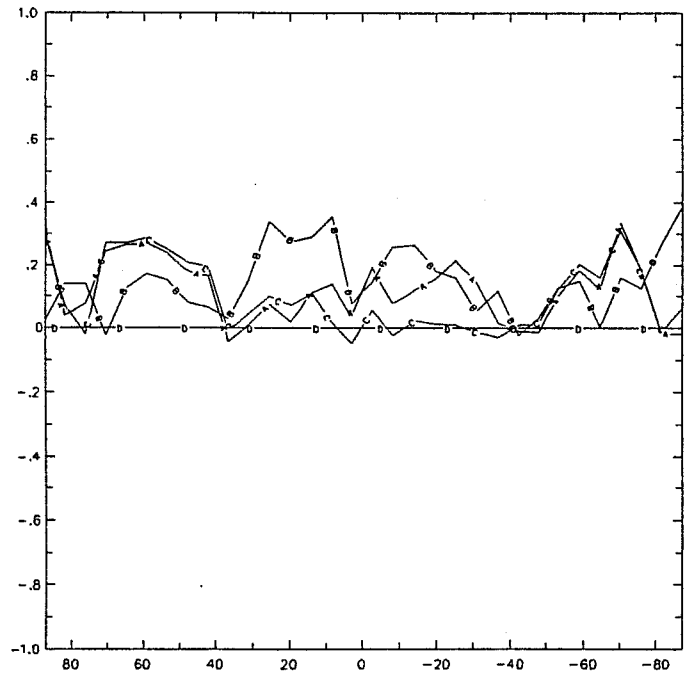
$$r = \frac{\overline{D_{f,r}}}{\sigma\left(\overline{D_{f,r}}\right)}$$

- Exp.A: (barotropic, ECMWF)
  - N30N: $S_f = 18\%$, $S_r = 23\%$, $r = 1.5$
  - S30S: $S_f = 15\%$, $S_r = 19\%$, $r = 1.2$

- (e)



- Fig. 5 (cont.): (e) initial σ used by exp.D.



- Fig. 6: Mean correlation between actual squared errors and forecast variances (10 situations) for the baroclinic forecast (A), the barotropic forecast (C), and the reference (B) (see text).

- Exp.B: (baroclinic, EMERAUDE)
  - N30N: $S_f = 23\%$, $S_r = 25\%$, $r = 0.4$
  - S30S: $S_f = 18\%$, $S_r = 18\%$, $r = 0.1$
- Exp.C: (barotropic, EMERAUDE)
  - N30N: $S_f = 21\%$, $S_r = 25\%$, $r = 1.2$
  - S30S: $S_f = 17\%$, $S_r = 18\%$, $r = 0.3$
- Exp.D: (barotropic, EMERAUDE, initial σ from OI)
  - N30N: $S_f = 19\%$, $S_r = 25\%$, $r = 1.7$
  - S30S: $S_f = 24\%$, $S_r = 18\%$, $r = -1.7$

The score of the reference forecast gives an estimation of the natural variability of the skill of the direct prediction, as it forecasts an almost constant value for the variance. This level of "skill variability" can be compared with the geographical variability of common scores: for the 24h range, the RMSE, for instance, is rather uniform. Consequently, for this range, it is hard to beat the reference.

For the extratropical Northern Hemisphere, the results are significant for the three barotropic experiments: the longitudinal structures forecast by the scheme actually give information. This is especially true when initial variances from OI (exp.D) are used. However, the difference for the CRPS with case C is too small and is not really significant. The baroclinic case is also better than the reference, but now the ratio r is not large enough to conclude. This can be due to the sampling: the difference between B and C is small, as previously mentioned, and is not statistically significant.

For the Southern Hemisphere, the results are better than the reference for the first three experiments, and significant only for the first one. With the initial variances from OI, the method is significantly wrong. We know that the initial variances are very unrealistic in data sparse regions. The method used by OI to compute the variance for the guess tends to diverge: it does not advect this quantity. With simpler initial variances, the forecast is not too bad. In addition, we use an analysis to verify the direct forecast and to compute the

"actual" error. This can be hazardous here: in the absence of data, the 24h forecast should not be very different from the subsequent analysis. But it would be absurd to conclude that the forecast is excellent. Therefore, the estimation of the real 24h forecast error should be more rigorous: we should take into account the uncertainty of the verifying analysis itself. Of course, this problem is less important over the Northern Hemisphere, or for longer ranges.

4 CONCLUSIONS

We have shown that a direct forecast of variances could be an alternative to the Monte Carlo method for the short range. It uses in a simple way the information given by the data assimilation system. It is very versatile: we can compute the variances one by one and thus choose the most interesting ones. Each column of the covariance matrix needs one tangent linear and one adjoint integration. Even if the direct forecast is made with a sophisticated model, approximations can be made for the linear model: we are sure that the "perturbed" forecasts would be unbiased if this is the case for the direct one: the basic "trajectory" is computed once. This property is not verified by the Monte Carlo method.

There are of course intrinsic limitations: the linearity and the cost. The linear hypothesis seems to work for the synoptic case (total wavenumber less than 20) up to 72h in the free atmosphere. It should be confirmed by models with physical parameterizations. The cost depends on the required information. For the full field of variances, and even with a simple baroclinic model as used here, it is not feasible. Here, the method is the Monte Carlo one, provided that the sampling is correct. But nobody needs the variance anywhere and for any quantity.

The different experiments use very simple models and initial covariance matrices. In spite of that, they prove that at least in the Northern Hemisphere, the forecast variances give information about the 24h forecast errors of operational models. A specific score for the forecast of the probability density, the continuous RPS, seems to be a very convenient tool to estimate the skill of the variance forecast. Probably, more work is needed to enlighten this problem of skill estimation.

The baroclinic experiment does not give any improvement and is even slightly worse. For

our cases, we have found that the growth of error is mainly a barotropic process. However, this result could be linked to the structure of the initial matrix: the separability assumption for the correlation function does not favour baroclinic structures.

The use of the initial variances as given by an operational OI seems to give positive results in the Northern Hemisphere. But clearly these quantities could be used more efficiently with some corrections.

Of course, we need more cases, different ranges, more sophisticated initial matrices, more precise forecast trajectory. But the result is there and it is statistically significant in spite of the simplicity: this method should be robust, and not too sensitive to the choice of the initial covariances.

Acknowledgements

The author wishes to thank Olivier Talagrand, Michel Déqué and the ARPEGE team in Paris for stimulating discussions and practical help with the ARPEGE/IFS code. This research is partly supported by EDF/DER.

References

Courtier, P. and Talagrand, O., 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. Q.J.R. Meteor. Soc.,113, 1329-1347.

Epstein, E.S., 1967: Stochastic dynamic prediction. Tellus, 21, 739-759.

Epstein, E.S., 1969: A scoring system for probability forecasts of ranked categories. J. Appl. Meteor.,8, 985-987.

Ghil, M., Cohn, S.E., Tavantzis, J., Bube, K. and Isaacson, E., 1981: Applications of estimation theory to numerical weather prediction. In: L. Bengtsson, M. Ghil and E. Källén (Editors), Dynamic meteorology: data assimilation methods. Springer Verlag, New York, pp. 139-224.

Hoffman, R.N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. Tellus, 35A, 100-118.

Lacarra, J. and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. Tellus, 40A, 81-95.

Matheson, J.E., and R.L. Winkler, 1976: Scoring rules for continuous probability distributions. Management Science, 22, 1087-1096.

Molteni, F., and T. Palmer, 1988: An experimental scheme for the prediction of forecast skill at ECMWF. Proceedings of the ECMWF Workshop on predictability in the medium and extended range, 16-18 May 1988, pp.367-402.

Molteni, F., Mureau, R. and T.N. Palmer, 1991: Atmospheric instability and ensemble weather prediction. ECMWF Research Dept. Tech. Memo. No. 177.

Murphy, A.H., 1971: A note on the ranked probability score. J. Appl. Meteor.,10, 155-156.

Palmer, T.N., and S. Tibaldi, 1988: On the prediction of forecast skill. Mon. Wea. Rev., 116, 2453-2472.

Pitcher, E.J., 1977: Application of stochastic dynamic prediction to real data. Mon. Wea. Rev., 109, 1367-1379.

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

Thompson, P.D., 1986: A simple approximate model of stochastic-dynamic prediction for small initial errors and short range. Mon. Wea. Rev., 114, 1709-1715.

Thompson, P.D., 1988: Stochastic-dynamic prediction of three dimensional quasi-geostrophic flow. J. Atmos. Sci., 45, 2669-2679.

Unger, D.A., 1985: A method to estimate the continuous Ranked Probability Score. Proceedings of the 9th Conference on Probability and Statistics in Atmospheric Sciences, Virginia Beach, 9-11 October 1985, pp.206-213.