DATA ASSIMILATION WITH AN EXTENDED KALMAN FILTER

Pierre Gauthier
Atmospheric Environment Service
Dorval, Canada

Philippe Courtier
ECMWF
Reading, UK

ABSTRACT

Since it produces a forecast of both the model state and its error covariance, the Kalman filter gives information about the actual accuracy of the analysis and this is what makes it unique. The extension to the nonlinear case has to face the issue of the closure scheme for the different statistical moments needed to describe the error structure. One such method is the *extended Kalman filter* (EKF) and it relies on a low order closure assumption while using the tangent linear model to integrate in time the forecast error covariance. In the present paper, the EKF will be used to assess the impact of simulated radiosonde and Doppler wind Lidar data on the time evolution of the forecast error covariance. Even though no model error has been considered, the results show that, with the radiosonde data alone, there can be a substantial forecast error growth especially in regions where the flow is unstable and no data are available. The error growth is attributed to instability processes that are embedded within the complex flow configuration around which the nonlinear model is linearized to obtain the tangent linear model. Wind data from a proposed satellite-based Doppler Lidar instrument are available in these regions of data voids so that when they are added, the assimilation is now capable to put a stop to the unlimited error growth observed in the previous experiment.

1. INTRODUCTION

The Kalman filter supplies a measure of the accuracy of an analysis in the form of its error covariance; therefore, it allows the impact of different observations sets to be compared. Using simulated observations, this can be used to look at different configurations for a proposed satellite instrument or network of observing stations to find the one that can be expected to be the most beneficial from the point of view of numerical weather prediction. A study of this type has been carried out recently by Cohn and Parrish (1991) who looked at the impact of wind profilers data on the analysis and forecast error by employing a linear Kalman filter.

In the nonlinear case, an extended Kalman filter (EKF) can be used and its basis is similar to the Kalman filter. The EKF differs from the Kalman filter in that it uses a nonlinear model to make a forecast and the forecast error covariances are now integrated with the tangent linear model. Moreover, the EKF is based on a low order closure assumption for the different statistical moments needed to describe the error structure and it may not be appropriate in some situations: this has been the object of some theoretical work by Miller and Ghil (1990). But the EKF brings new features concerning the time evolution of the forecast error covariances because the nonlinear model is linearized around a

time and space dependent flow configuration to yield the tangent linear model. Although linear, this model is complex and it embeds instability processes that are not yet fully understood.

An EKF based on a non-divergent barotropic model has been implemented within the IFS/ARPEGE system that also includes the tangent linear model. The focus will be on the time evolution of the forecast error covariance and not on the analysis itself. When assimilating data, several factors intervene:

- the data themselves with their geographical and time distribution, their nature (wind, geopotential, etc.) and their quality,
- the method used (optimal interpolation, Kalman filtering, variational method, etc.),
- the model used to perform the assimilation.

In order to focus on the impact of the data themselves, identical twin experiments have been conducted to allow for a dynamical consistency between the observations and the assimilating model. This also removes, for the time being, the need to consider model error and none has been included in the experiments to be described below. To assess the usefulness of Doppler wind Lidar data from a satellite based instrument, results are compared between an experiment with simulated radiosonde data alone and others where the simulated Lidar data are added. This is done for different satellite configurations.

2. IMPLEMENTATION OF THE EXTENDED KALMAN FILTER

The derivation of the filter's equations can be found in Jazwinski (1970): it follows the same procedure used in Ghil *et al.* (1981) for the linear Kalman filter. If $\mathbf{X} \in R^n$ stands for the model state of the dynamical system

$$\frac{d\mathbf{X}}{dt} = \mathbf{f}(\mathbf{X}), \tag{2.1}$$

its integration from time t_{k-1} up to t_k, using $\mathbf{X}_{k-1}$ as initial conditions, can be represented symbolically by the relation

$$\mathbf{X}_k = \mathcal{F}_{k-1}(\mathbf{X}_{k-1}) . \tag{2.2}$$

The equations of the EKF are shown to be the following:

$$\begin{aligned}
&\text{a)} \quad \mathbf{K}_{k-1} = \mathbf{P}^{f}_{k-1}\mathbf{H}'^{T}_{k-1}\left(\mathbf{R}_{k-1} + \mathbf{H}'_{k-1}\mathbf{P}^{f}_{k-1}\mathbf{H}'^{T}_{k-1}\right)^{-1}, \\
&\text{b)} \quad \mathbf{X}^{a}_{k-1} = \mathbf{X}^{f}_{k-1} + \mathbf{K}_{k-1}\left(\mathbf{Z}_{k-1} - \mathbf{H}_{k-1}(\mathbf{X}^{f}_{k-1})\right), \\
&\text{c)} \quad \mathbf{P}^{a}_{k-1} = \left(\mathbf{I} - \mathbf{K}_{k-1}\mathbf{H}'_{k-1}\right)\mathbf{P}^{f}_{k-1} \\
&\text{d)} \quad \mathbf{X}^{f}_{k} = \mathcal{F}_{k-1}(\mathbf{X}^{a}_{k-1}) \\
&\text{e)} \quad \mathbf{P}^{f}_{k} = \mathcal{A}_{k-1}\mathbf{P}^{a}_{k-1}\mathcal{A}^{T}_{k-1} + \mathbf{Q}_{k-1},
\end{aligned} \tag{2.3}$$

where the following definitions have been used:

$\mathbf{X}_k^{f,a}$: forecast/analysis at time t_k ;

$\mathbf{P}_k^{f,a}$: forecast/analysis error covariance $\left(= \mathcal{E}\,(\mathbf{X}_k^{f,a} - \mathbf{X}_k^t)\,(\mathbf{X}_k^{f,a} - \mathbf{X}_k^t)^T\right)$;

$\mathbf{H}_k$: forward interpolation of the model state to the observations;

$\mathbf{Z}_k$: vector representing a set of "M" observations available at time t_k;

$\mathbf{R}_k$: observation error covariance matrix
(including measurement and representativeness error);

$\mathbf{K}_k$: gain matrix;

$\mathbf{Q}_k$: model error covariance;

$\mathcal{A}_k$: represents the integration from t_k to t_{k+1} of the *tangent linear model*

$$\frac{d}{dt}\delta\mathbf{X} = \frac{\partial \mathbf{f}}{\partial \mathbf{X}}(\mathbf{X}_s(t))\,\delta\mathbf{X} \tag{2.4}$$

Here, $\mathcal{E}$ stands for the statistical expectancy. Within the EKF framework, the time evolution of the forecast error is governed by the tangent linear model (2.4) that stems from a linearization of (2.1) around a model trajectory $\mathbf{X}_s(t)$ obtained by integrating the nonlinear model, using the current best estimate as initial conditions. Although this model is linear, it is linearized around a complex flow configuration and it embeds instability processes that are usually absent from simpler linear models.

The expression (2.3a) for the gain matrix $\mathbf{K}_k$ has been determined by minimizing a quadratic measure of the analysis error

$$J = \mathcal{E}\,(\mathbf{X}_k^a - \mathbf{X}_k^t)^T\,\mathbf{A}\,(\mathbf{X}_k^a - \mathbf{X}_k^t)$$

where $\mathbf{A}$ can be any arbitrary positive semi-definite matrix. For instance, by taking $\mathbf{A}$ to be identically zero with the exception of one element of its diagonal, this minimizes also the error variance for each component of the model state vector.

Several pieces of information are needed for the integration of (2.3). Initial conditions have to be provided not only for the first-guess $\mathbf{X}_0^f$ but also for its error covariance $\mathbf{P}_0^f$. The *observation error covariance* $\mathbf{R}_k$ must also be specified: it depends on the type of observation and is related to the measurement and forward interpolation errors but also to the natural variability of the atmospheric flow on scales that are not resolved by the assimilating model (representativeness error). Finally, a knowledge of the model error covariance $\mathbf{Q}_k$ is also required. This error can be associated with the discretization error or the absence of some dynamical processes in the numerical model (e.g., imperfect parameterization of the sub-grid scale processes). Its determination could be done adaptively (Dee *et al.*, 1985) by making the *innovation vector* sequence $(\mathbf{Z}_k - \mathbf{H}_k(\mathbf{X}_k^f))$ uncorrelated in time. In most studies however (e.g., Cohn and Parrish, 1991; Dee, 1991), the model error covariance is specified and the form used has to be justified. The work of Phillips (1986) investigates the validity

of such approximations and a discussion of this question in relation to the Kalman filter is presented in Cohn and Parrish (1991). It is important to keep in mind that the results may depend crucially on what is used for the model and observational error (Daley, 1991).

In the limit where the nonlinearities become negligible, the EKF is optimal and corresponds to the usual Kalman filter. In the nonlinear case, its optimality has to be verified through a Monte-Carlo simulation out of which the forecast covariances can be estimated and compared to what the EKF produces (Veyre, 1991). This is directly related to the validity of the tangent linear approximation that, has been shown to be a reasonably good one for barotropic models over a 24-hour time interval that corresponds to what will be used here (Courtier and Talagrand, 1987; Lacarra and Talagrand, 1988).

The assimilation starts with the analysis cycle and then proceeds with the integration of the resulting analyzed model state and analysis error covariance. The EKF has been implemented within the framework of the ARPEGE/IFS model of ECMWF and Météo-France. The present paper will only be concerned with the specifics of the implementation of the EKF for the barotropic vorticity equation

$$\frac{\partial \zeta}{\partial t} + J(\psi, \zeta + f) = 0,$$

where ψ, ζ and f are respectively the stream function, the relative vorticity and the planetary vorticity while J(a,b) stands for the Jacobian operator. When implementing the EKF, it is important to remember that the linear operators $\mathcal{A}_{k-1}$ and $\mathbf{H}_{k-1}$ appearing in (2.3) are not explicitly known as matrices but only through their action on some vector $\mathbf{X}$. The calculation of

$$\mathbf{M}_k = \mathcal{A}_{k-1}\, \mathbf{P}^a_{k-1}$$

is done by performing N integrations of the tangent linear model from t_{k-1} up to t_k using successively the N columns of $\mathbf{P}^a_{k-1}$ as initial conditions and replacing the corresponding column with the result. Proceeding in this manner, only one N×N matrix is required and the completion of the integation of $\mathbf{P}^f$ is obtained from

$$\mathbf{P}^f_k = \mathcal{A}_{k-1}\, \mathbf{P}^a_{k-1} \mathcal{A}^T_{k-1} = (\mathcal{A}_{k-1}\, \mathbf{M}^T_k) \quad .$$

This requires N more integrations of the tangent linear model using now the rows of $\mathbf{M}_k$ as initial conditions. The cost of integrating the forecast error covariances therefore corresponds to that of 2N integrations of the tangent linear model. For a barotropic vorticity equation model truncated at T21, a 24-hour integration of the EKF with a one-hour timestep was done in 1100 seconds of CPU time on a CRAY 2 computer.

The forward interpolation $\mathbf{H}_k$ summarizes all operations necessary to go from the model state $\mathbf{X}_k$ to the observations. In the present case, the assimilation of an observation of a single wind component

would require that $\mathbf{H}_k$ includes all the operations needed to go from the spectral components of the vorticity to the model wind interpolated at the observation point (see Gauthier *et al.*, 1992). The assimilation of observations of the geopotential requires that the model state be related to it through the nonlinear balance equation

$$\Delta\phi = \nabla\cdot\left((\zeta+f)\nabla\Delta^{-1}\zeta\right) - \frac{1}{2}\nabla\Delta^{-1}\zeta\cdot\nabla\Delta^{-1}\zeta . \tag{2.6}$$

where Δ stands for the Laplacian operator. At mid-latitudes, the planetary vorticity f being usually more important than the relative vorticity ζ, (2.6) can be approximated by linearizing around a state of rest to give the linear balance equation

$$\phi' = \Delta^{-1}\left[\nabla\cdot(f\nabla\Delta^{-1}\zeta)\right]. \tag{2.7}$$

which has been used for sake of simplicity and $\mathbf{H}_k$ is linear.

In the experiments that will be presented, $\mathbf{P}_0^f$ has been initialized by assuming the first-guess error correlation on the stream function to be homogeneous and isotropic. This point is discussed at length in Appendix B of Gauthier *et al.* (1992) where it is also shown that under these conditions, $\mathbf{P}_0^f$ is diagonal. The aim here was to choose something reasonable although the correct form could be obtained by integrating the EKF over a sufficiently long time so that the forecast error covariance converges to the asymptotic form that it would reach for a stationary observation network (Daley, 1991). But this would require a good knowledge of the model error which was not considered in our experiments.

3. ASSIMILATION OF A SIMULATED RADIOSONDE NETWORK

All the experiments will use synthetic data and are of the identical twin type (Daley, 1991). Consequently, no model error is taken into account and, as pointed out in Cohn and Parrish (1991), this can imply that there is no global error growth and it usually produces overly optimistic results. The observations have been simulated according to the radiosonde network for measurements of wind and geopotential at 500 mb associated with the 633 TEMP stations and wind only for the 135 PILOT stations: all reporting stations on 14 July 1990, 0 h UTC have been considered. As can be seen from Fig.1, this observation network covers the continents of the Northern hemisphere rather well but gives only a weak coverage of the oceanic regions with important data voids especially in the Southern hemisphere. Synthetic observations were generated from a 24-hour integration of the barotropic vorticity equation spectrally truncated at T21: the initial conditions $\mathbf{X}_0^t$ used were those of 14 July 1990 at 0h UTC. At the location of each station, values of u, v and ϕ have been generated every 6 hours for 24 hours according to what would be provided by the station; the network has been assumed to be the same for each synoptic time. In reality, the radiosonde network has a periodicity of 12 hours and not 6 hours, but the choice made here can be justified by the fact that various other sources of

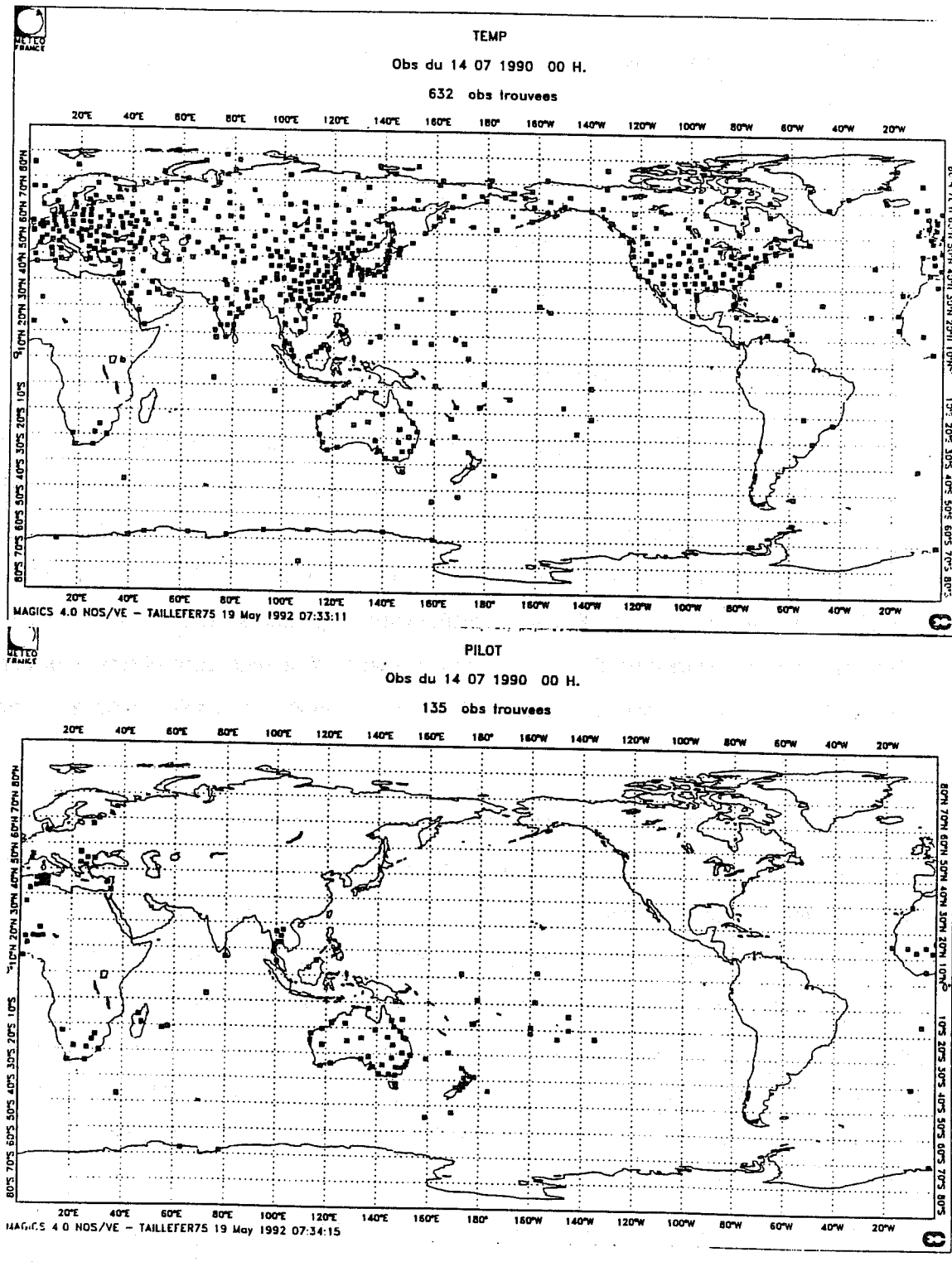

Figure 1. Typical radiosonde network at 00Z for TEMP and PILOT stations. This network is assumed to be the same every 6 hours.

information such as the new emerging ACAR have not been considered here. No noise was added to the interpolated model state but observational error is introduced at the assimilation stage: it has been assumed to be uncorrelated with standard deviations of 2.24 ms^{-1} and 11.2 m respectively on wind and geopotential height observations: these are operational estimates of the error made by radiosondes.

Assimilation with the EKF was then accomplished by initializing the first-guess error covariance $\mathbf{P}_0^f$ to be homogeneous and isotropic and the correlation length was set to 1200 km while the uniform variance corresponds to that of a geopotential error level of 21 m (standard deviation) at mid-latitudes. The focus will not be on the analysis itself but on the time evolution of the covariances that can be influenced by the nature and distribution of observations (through the interpolation $\mathbf{H}_k'$), their error structure (through $\mathbf{R}_k$) and the initial assumptions made on $\mathbf{P}_0^f$. Moreover, the first-guess is taken to correspond to $\mathbf{X}_0^t$ thereby implying that $\mathbf{X}_k^a = \mathbf{X}_k^f$ at all times because the observations correspond to an exact model state. By proceeding in this manner, only (2.3-a,c,e) are active in the EKF equations, the tangent linear model that is therefore a linearization around the true trajectory.

By choosing the initial estimate of $\mathbf{P}_0^f$ to correspond to what is found for the operational analysis error level, the time evolution of the forecast error covariance should not be very different from what would be obtained by linearizing around the current best estimate instead of the true solution because the tangent linear model has been verified to give an accurate representation of the forecast error over a 24-hour time interval (Courtier and Talagrand, 1987; Lacarra and Talagrand, 1988): this holds for the large scales resolved by the T21 barotropic model used here. This would cease to be true however if the assimilation was made over periods of time exceeding the limit of validity of the tangent linear model. As a result, the assimilation deviates from a strict implementation of the extended Kalman filter but the time evolution of the forecast error covariance should not be significantly different.

Fig.2 shows the analysis error variance at t = 0, 12 and 24 hours. At t = 0, it is as it would be if optimal interpolation was used but when the time dimension comes into play, the variance is now governed by the tangent linear model that captures the dynamics of error growth and propagation of information. This is clearly illustrated by Fig.2-b,c showing that the region west of South America experiences a significant increase in the error level which can be attributed to three factors. First of all, a more important error growth rate is hinted at by the local vorticity gradient that points towards local barotropic instability in the area: this can be seen on Fig.3a where the vorticity and the error variance over that region have been plotted at the end of the assimilation time interval. The maximum variance in that area is about 4 times the initial constant level and the error has therefore doubled in 24 hours. Secondly, this area being totally devoid of any observations, nothing acts to bring the error level down.

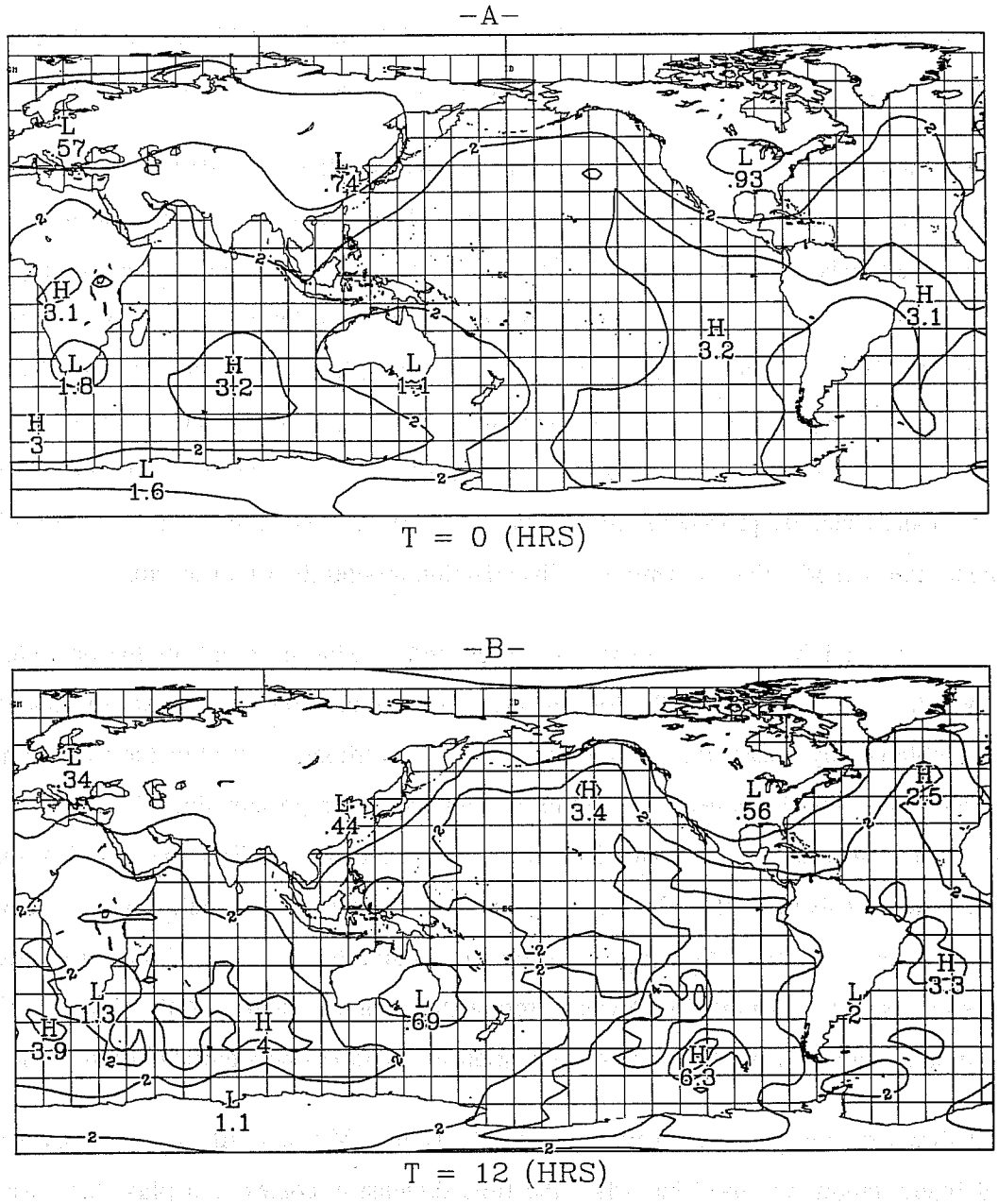

Figure 2. Analysis error variance resulting from an assimilation with the EKF. a) t = 0 hrs, b) 12 hrs, c) 24 hrs. Units used are 1×10^{-12} s^{-2} with contour levels of 1 unit and the characteristic length of the first-guess error is 1200 km.

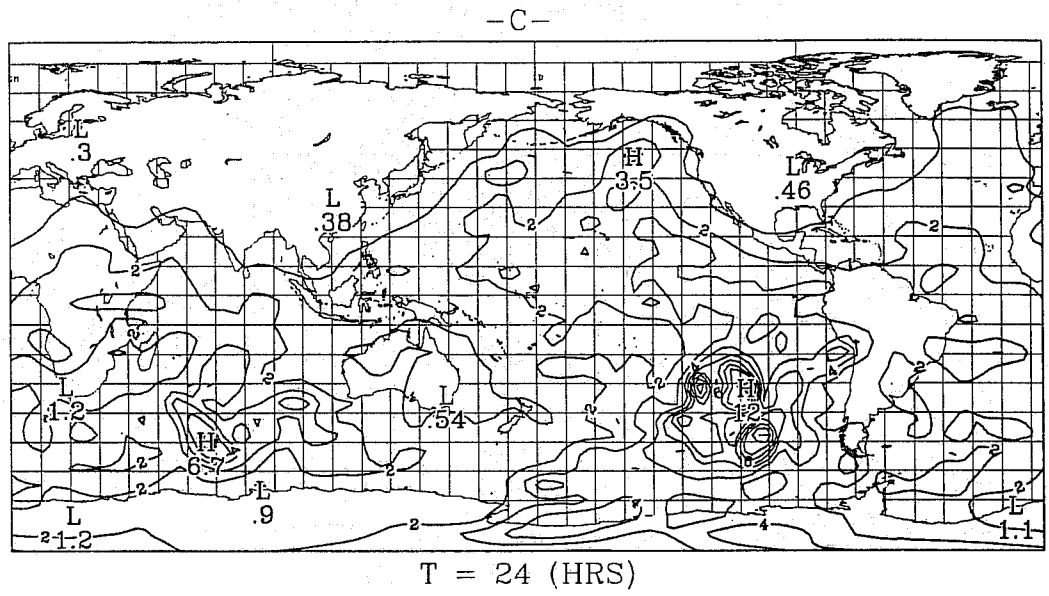

Figure 2c. Analysis error variance resulting from an assimilation with the EKF at t = 24 hrs. Units used are 1×10^{-12} s^{-2} with contour levels of 1 unit and the characteristic length of the first-guess error is 1200 km.

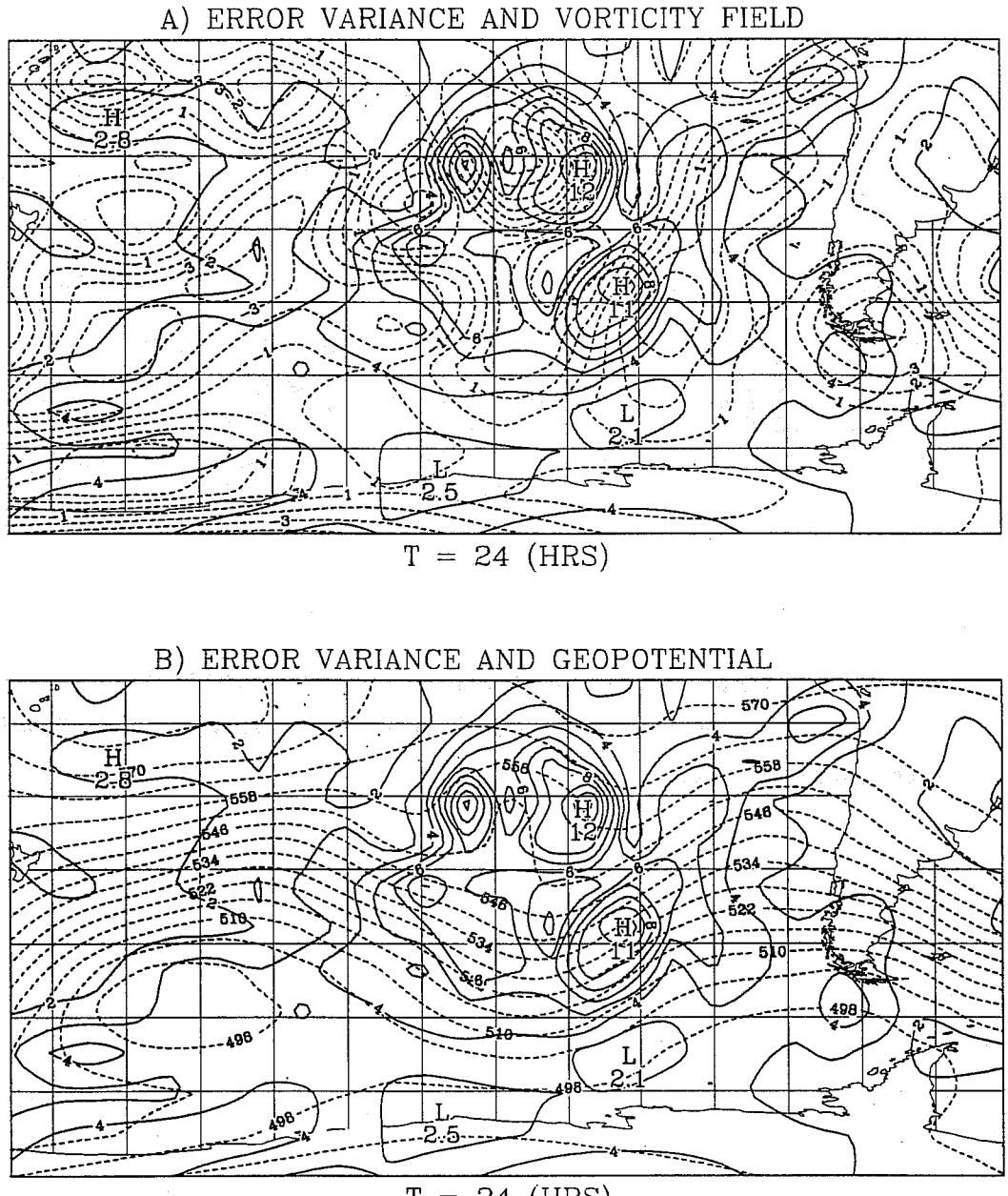

Figure 3. Analysis error variance field (solid line) at t = 24 hrs superposed over a) the vorticity field (dashed), b) the geopotential field (dashed). The domain is a close-up over the region West of South America.

Finally, the geopotential (shown on Fig.3b) indicates that the circulation tends to skirt the area and propagation of information from the outside is therefore impaired.

Without the time integration of the forecast error covariances, one has only to solve (2-3a,c) to update the analysis error covariances according to information brought in by new observations. Proceeding in this way is an improvement over optimal interpolation in that the analysis error covariance is updated and that no data selection algorithm is involved. The resulting analysis error variance is represented at the final time on Fig.4. A comparison with the variance field of Fig.2-c obtained with the EKF shows that this "improved" optimal interpolation is not giving a good estimate of the accuracy of the analysis. The conclusion is then that the EKF brings important and new information about the actual accuracy of the analysis which is seen to be quite different from the picture obtained from optimal interpolation that uses an incorrect estimate of the forecast error covariance. The reason for this is that the EKF takes into account the dynamics of error growth and propagation of information by integrating in time the forecast error with the tangent linear model. The latter being obtained from a linearization of the model's equations around a complex time dependent flow configuration, its dynamics embeds instability processes: this leads to error growth (even in the absence of model error) that depends on the local flow configuration. Propagation of information is also more accurate and the presence of areas of closed circulations would tend to block information from entering in those regions. These mechanisms are seen to be quite different from those in an operational data assimilation cycle where error growth is taken into account usually by increasing the uniform error variance linearly in time which increases it globally regardless of the particularities due to local instabilities, lack of observations, etc. .

The results may also depend on what is used to initialize the forecast error covariance. For instance, the correlation length used in the experiments reported above was 1200 km which means that observations had an impact over a very large region. To assess the impact of this particular choice, the experiment was repeated using 400 km as the characteristic scale of the initial first guess forecast errors. The results show that the error level is considerably higher even at the initial time. This is a direct consequence of the reduced characteristic scale: by keeping the geopotential error field (and hence the streamfunction) to be the same as in the previous experiments (21 m), the corresponding vorticity error is increased because it is obtained by differentiating the streamfunction twice thereby making the characteristic scale to appear at the denominator. However, the general features of the variance field are very similar to those reported earlier (Fig.3), in particular the regions of intense error growth are still present. Fig.5 shows the time evolution of the analysis error variance averaged over the globe when the inital characteristic length is respectively 400 and 1200 km; to emphasize the impact

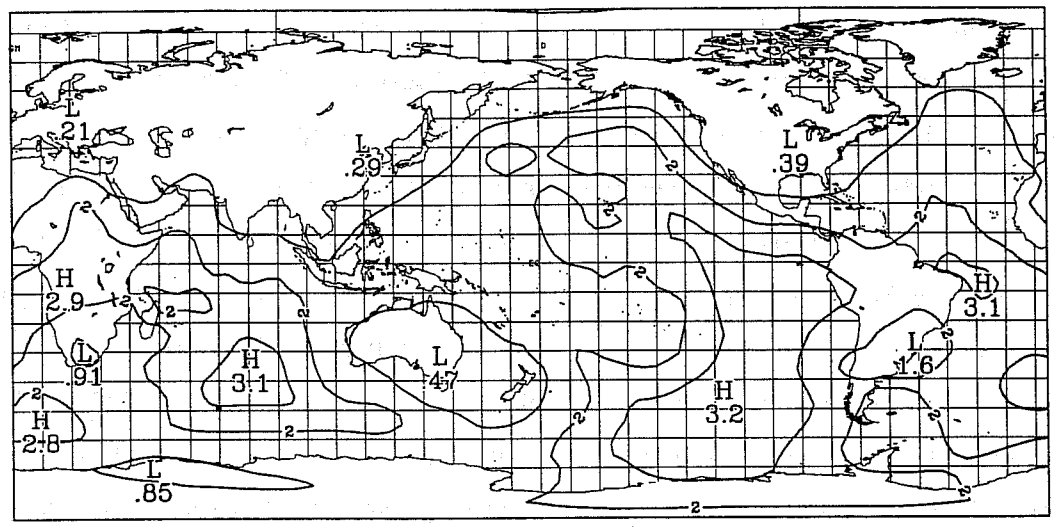

Figure 4. Analysis error variance at the final time for an assimilation without the time integration of the forecast error covariances.

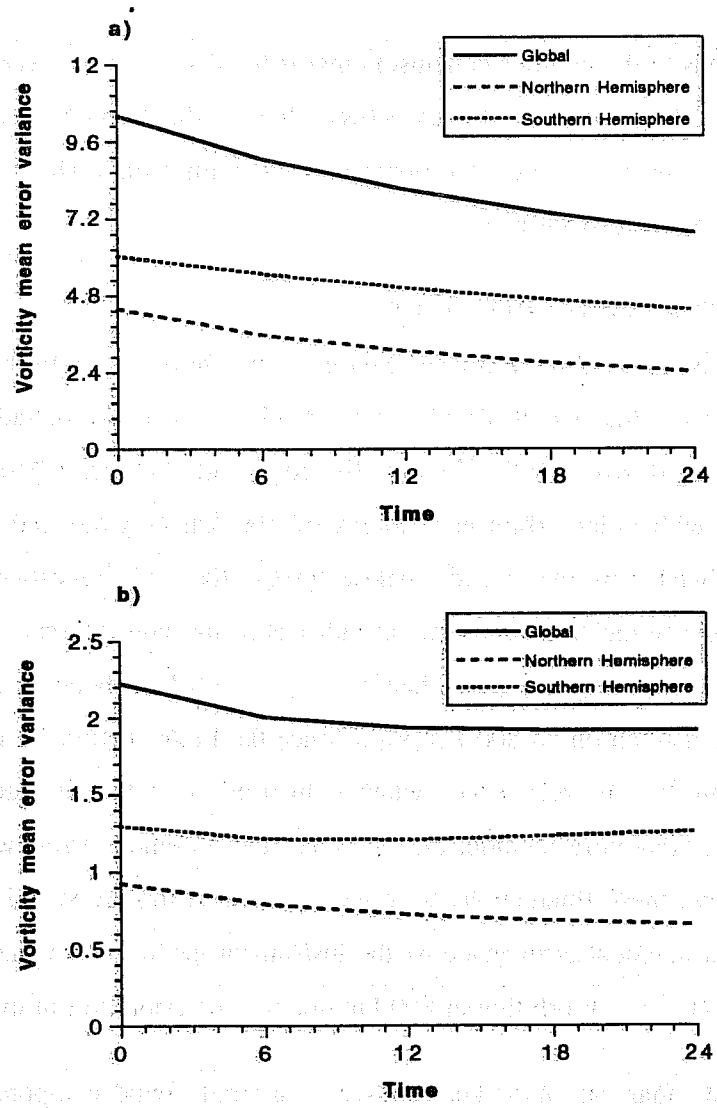

Figure 5. Time evolution of the global mean analysis error variance (solid line) for a characteristic length of the first-guess error of a) 400 km, b) 1200 km. On both figures, the contributions from the Northern and Southern Hemispheres has been plotted separately to emphasize the differences existing between the two hemispheres.

of having a denser observation network, the contributions to the global mean coming from the Northern and Southern hemispheres have also been plotted. This clearly shows the analysis to be poorer in the Southern Hemisphere in both cases and the global reduction in analysis error is done mostly in the Northern Hemisphere.

This last result leads to the obvious conclusion that if the data coverage were more uniform, it is to be expected that few differences should exist between the Southern and Northern hemispheres. This is investigated in the next section where wind observations coming from a simulated satellite based Lidar system are added to the radiosonde network.

4. IMPACT OF THE WIND LIDAR DATA

In recent years, both NASA and the European Space Agency have proposed to launch a satellite based Doppler Lidar instrument that would provide a fairly global coverage of radial wind measurements along the line of sight (Curran, 1989; Betout, Burridge and Werner, 1989). Fig.6 shows the data coverage associated with polar orbits at altitudes of 400 km (Fig.6a) and 800 km (Fig.6b). This particular choice is dictated by the fact that below 400km, the orbit is unstable while at 800 km, the Lidar instrument is becoming less accurate. In both cases, the conical scan has a 45° aperture. The main feature emerging from Fig.6 is that when orbiting at 400 km, there are significantly larger data gaps than those created when on an 800 km orbit. Since the Lidar instrument provides a measurement over a volume of air that is very small when compared with the resolution of the model, the observation error must therefore include, on top of the measurement error, wind fluctuations on the unresolved scales: here, these fluctuations have been estimated to correspond to the radiosonde wind error. Finally, the measurement error made by the instrument has been estimated to be of 1 ms^{-1} for a 400 km orbit and 4.3ms^{-1} when orbiting at 800 km due to a deterioration in the signal-to-noise ratio.

The Doppler wind Lidar data used here consist of a single wind component measured along the direction of the line-of-sight: the orbitography simulator provided the location of the measurement and the direction of the line-of-sight that were used in the simulation and assimilation of the data. In all the following experiments, the methodology is the same as in the previous section but now, Doppler wind Lidar data have been simulated and added on top of the same radiosonde data used in the preceding section (the same initial conditions situation were used to simulate the data). The Lidar data are inserted at every hour while the radiosonde network comes in at every six hours as before. In the first experiment, the satellite is on a polar orbit at an altitude of 400 km and the Lidar shots have a 1 Hz firing frequency (3,600 wind measurements per hour): an observation error of 2.45 ms^{-1} on the radial wind component has been considered. Over a 24 hour period, the volume of data corresponds to 86,400 single wind component observations that must be added to the 10,845 scalar observations of

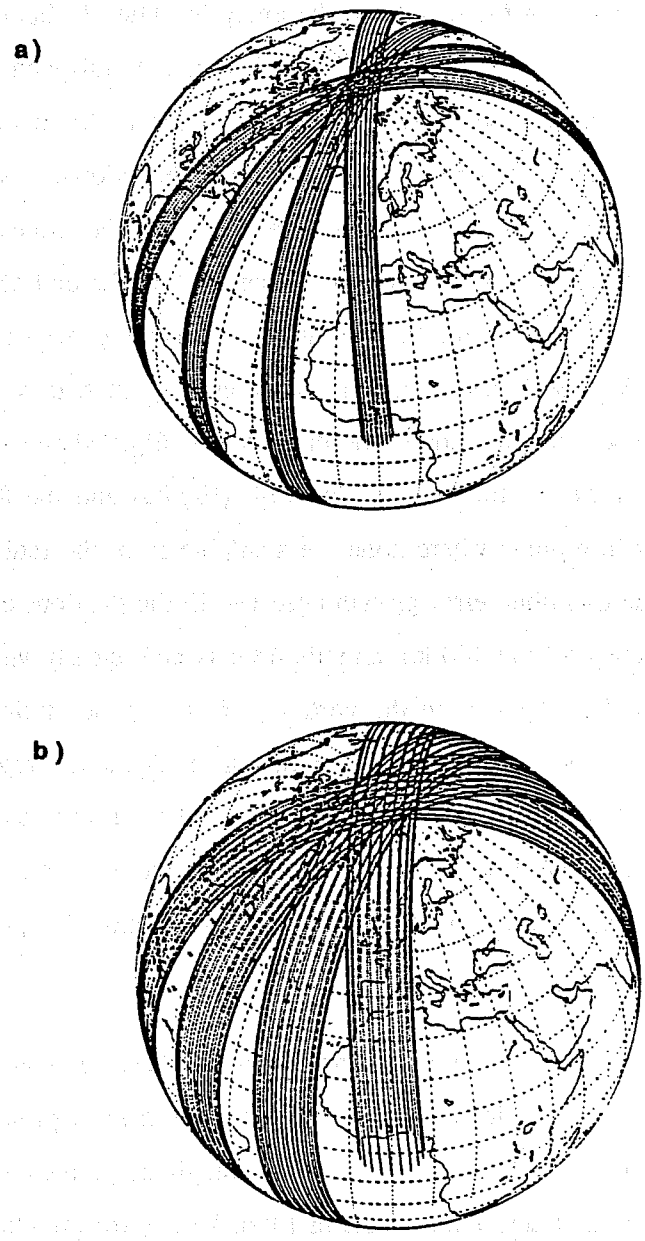

Figure 6. Typical data coverage for a satellite on a polar orbit at an altitude of a) 400 km, b) 800 km. Distribution of the points correspond to those of a conical scan with a scanning period of 10 seconds. The shot frequency was set to 2 Hz.

the radiosonde network. Fig.7a summarizes the results from an experiment in which P_0^f was initialized with a forecast error scale of 400 km. A comparison with Fig.5 shows that the Lidar data have a very beneficial impact on the quality of the analysis. After 24 hours, the data coverage being then nearly uniform, the quality of the analysis in the Southern Hemisphere is now comparable to that of the Northern Hemisphere. This experiment were repeated for an 800 km orbit and Fig.7b shows the mean error variance. Although a qualitatively similar error reduction is observed, the increase in the measurement error has a detrimental effect that overcomes the beneficial effect of having a more global data coverage. For instance, the mean error variance $\bar{\sigma}^2$ at the end of the assimilation period is found to be $\bar{\sigma}^2 = 0.45\times10^{-12}s^{-2}$ for the 400 km orbit while it goes up to $0.81\times10^{-12}s^{-2}$ for a satellite orbiting at 800 km. Both cases converge to an error level that is lower than the variance of $1.915\times10^{-12}\ s^{-2}$ obtained with the radiosonde data alone. Fig.8 shows the corresponding analysis error variance field at t= 24 hrs for the 400 km orbit (Fig.8a) and the 800 km one (Fig.8b). By providing measurements in regions where none are available from the radiosonde network, the Lidar data can put a stop to the unlimited error growth observed in the previous experiments. The impact of the non global data coverage of the 400 km altitude orbit is now clearly visible in the tropical belt as an alternation of high and low values in the vorticity error variance field. Finally, it is also worth noticing that in both cases, the error variance is higher in the equatorial region: this is a consequence of having a higher number of Lidar measurements in the polar regions. This would suggest that shot management schemes should be designed so as to provide a more uniform coverage by reducing the number on shots in the polar regions and increasing it in the equatorial region.

5. CONCLUSION

The most striking feature of the results obtained with the tangent linear model is the impact of instability processes on the dynamics of error growth, which is most pronounced in regions of data voids. In these regions, forecast errors may exceed by far those predicted by optimal interpolation. The results that were presented here point to the fact that local error growth is related to the presence of strong vorticity gradients that are characteristic of shear flow instabilities. This suggests that if the EKF was based on a more complex model, other types of instabilities (e.g., baroclinic instability) are likely to lead also to error growth. In experiments with a linear Kalman filter where the linear flow does not support any instability mechanism, error growth stems from an external model error forcing (Dee, 1991). In the present study, no model error was considered to emphasize that error growth could occur even in its absence. But it must be kept in mind that a consequence of having no model error is that it usually produces overly optimistic results.

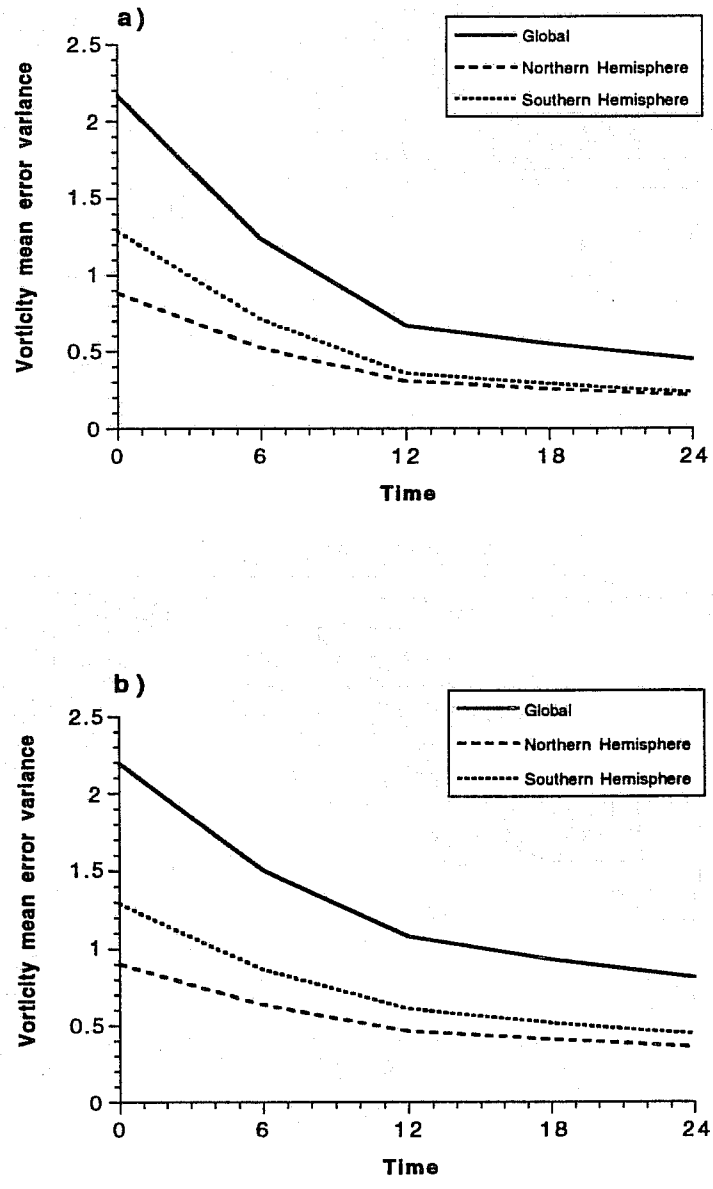

Figure 7. Time evolution of the global mean analysis error variance for an assimilation using the Lidar data coming from a satellite on a polar orbit at an altitude of a) 400 km, b) 800 km. The characteristic length of the first-guess error has been set to 1200 km.

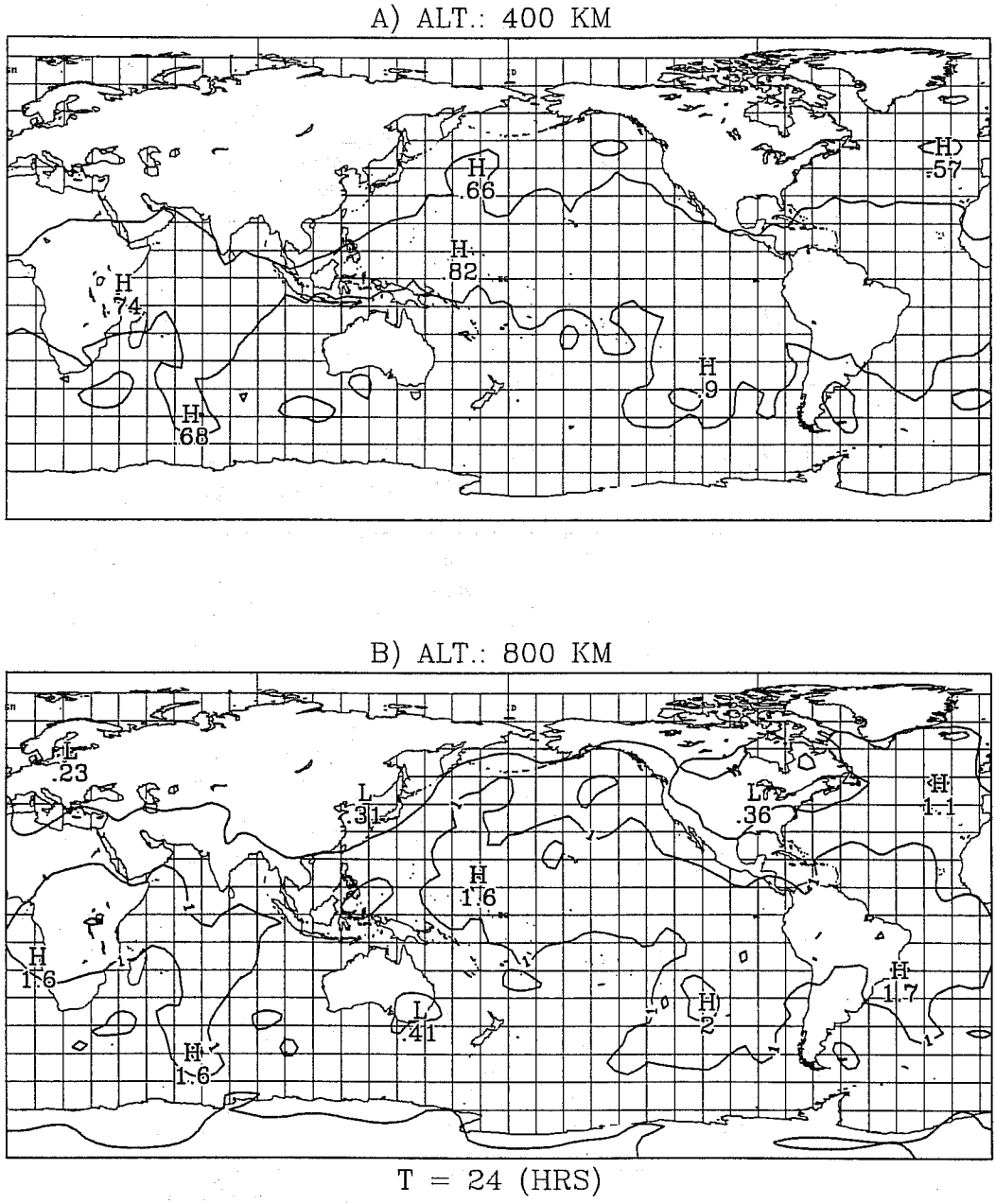

Figure 8. Analysis error variance field at t = 24 hrs for an assimilation using Lidar data obtained from a satellite at an altitude of a) 400 km, b) 800 km. In both cases, a characteristic length of the first-guess error of 1200 km has been used.

Although the methodology adopted here is correct for a study of the large scale part of a flow over a 24-hour period, the impact of having chosen the true state for the initial estimate of the atmospheric state could be questioned if the assimilation was to be carried out over a period that exceeds that over which the tangent linear model is valid. This limit of validity is determined by comparing the error obtained with the full nonlinear model against the one predicted by the tangent linear model when the initial error is of an amplitude corresponding to the estimate used for the initial error covariance. For longer time periods, the evolution of the forecast error covariance should be determined by integrating the EKF with an initial estimate of the model state that contains an error level consistent with the initial assumption made on its error covariance.

Finally, the EKF is a very useful tool to perform Observation Systems Simulation Experiments (OSSEs) since it provides not only an analysis but also information on its accuracy. In this paper, a "mini-OSSE" was conducted to make a preliminary assessment of the impact of data coming from a proposed satellite based Doppler wind Lidar instrument. Our results have shown that the availability of a global coverage of wind observations is able to put a stop to the unlimited error growth that was observed when only radiosonde data were used. The impact is even more important in the Southern Hemisphere that is poorly covered by the radiosonde network. Two scenarios have been investigated for 400 and 800 km orbits respectively. It was concluded that, due to a loss in the measurement accuracy, the 800 km gives a less accurate analysis even though the data coverage was more global.

6. REFERENCES

Betout, P., D. Burridge and Ch. Werner, 1989: *Doppler Lidar Working Group Report.* ESA Publications Division, Noordwijk, Netherlands, 45 pages.

Cohn, S.E. and D.F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757-1785.

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

Curran, R.J., 1989: NASA's plans to observe the Earth's atmosphere with Lidar. *IEEE Trans. on Geoscience and Remote Sensing,* **27**, 154-163.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, Cambridge, 457 pp.

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

Dee, D.P., S.E. Cohn, A. Dalcher, and M. Ghil, 1985: An efficient algorithm for estimating noise covariances in distributed systems. *IEEE Trans. on Automatic Control,* **AC-30**, 1057-1065.

Gauthier, P., P. Courtier and P. Moll, 1992: Assimilation of simulated wind Lidar data with a Kalman filter. (submitted to *Monthly Weather Review)*

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Application of estimation theory to numerical weather prediction. In *Dynamic Meteorology: Data assimilation Methods.* L. Bengtsson, M. Ghil, E. Källén, editors, Springer-Verlag, 330 pages.

Jazwinski, A., 1970: *Stochastic processes and Filtering Theory.* Academic Press, New York, 376 pp.

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

Miller, R.N., and M. Ghil, 1990: Data assimilation in strongly nonlinear currents. Pp.93-98 in *Proceedings of the WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography,* Clermont-Ferrand, France, July 9-13, 1990.

Phillips, N.A., 1986: The spatial statistics of random geostrophic modes and first-guess errors. *Tellus,* **38A**, 314-332.

Veyre, P., 1991: Direct prediction of error variances by the tangent linear model: away to forecast the uncertainty in the short range? In *New Developments in Predictability.* ECMWF Workshop Proceedings, 13-15 November 1991, Reading, U.K., 333 pages.