# CHAPTER 8
# LIDAR INVERSION AND NONLINEAR
# KALMAN FILTERING

The inversion of extinction and backscatter profiles from lidar return-signals can be tackled by means of nonlinear Kalman filtering theory. It is proved that it is the optimal estimator under any reasonable criterion.

The potential advantage of Kalman filtering over nonmemory algorithms is the goal of retrieving both extinction and backscatter profiles at virtually every observation cell. This is a result of making the most of correlation amongst past lidar return-signals. Since the training capacity of the filter enables to give different weights to the actual and past measurements, nonmemory algorithms (in paricular, least-squares) can be understood as the simplest version of the Kalman filter, where the maximum confidence measurement is the actual one.

Philosophically, the Kalman filter represents a challenging and ambitious goal to the general problem of lidar inversion. Since the possibilities of the filter depend too much on how well the theorical model represents the atmosphere, whose physical model cannot exactly be known, the solution of the Kalman filter will be suboptimal. That is why the natural flow of the chapter leads to different filters and stochatic models for the atmosphere based on simulated data, leaving for Chap.9 its incipient application to real data.

Finally, the last part of the chapter addresses the observability problem, where prerequisites for the Kalman-lidar estimator are stated.

*"... the Kalman filter represents the most widely applied and demonstrably useful result to emerge from the state variable approach of modern control theory"* (H.W. Sorenson [203]).

## 1. INTRODUCING THE KALMAN FILTER

### 1.1 Linear discrete Kalman filter

Let us consider a stochastic discrete time vector process $x_k$ (dimensions are indicated in brackets) modelled by

$$x_{k+1} = \Phi_k \, x_k + w_k \tag{1}$$

where:    $x_k$ is the time-state vector at time $t_k$ *(nx1)*

$\Phi_k$ is the transition state matrix from time $t_k$ to $t_{k+1}$ *(nxn)*

$w_k$ is the state noise vector *(nx1)*.

The measurement or observation also takes place at discrete times $t_k$, according to the following linear relationship:

$$z_k = H_k \, x_k + v_k \tag{2}$$

where:

$z_k$ is the measurement vector at time $t_k$ *(nx1)*
$H_k$ is the measurement matrix *(mxn)*
$v_k$ is the measurement error vector *(mx1)*.

To illustrate somehow the physical meaning of the variables introduced, let us take the case of *xy-speed monitoring* of an uniform bidimensional motion. As far as it concerns to the system under study, $x_k$ would be the mobile's xy-speed components $(v_x, v_y)$, $\Phi_k$ would be the bidimensional matrix describing a simple uniform 2-D motion that relates vector speed components from one sample time to the next one (as the motion is expected to be truly uniform over the entire time, $\Phi_k = I$ can be assumed), finally, $w_k$ would be the noisy process modelling the velocity drift in the mobile's motion. As for the observation system, $H_k$ would be the linear equations relating the mobile's xy-coordinates to the vector speed components. Were these equations not to be linear, it would be necessary to resort to the extended Kalman filter that will be commented in Sect.1.2. Finally, $v_k$ would be the measurement noise, for instance, electronic thermal noise.

The noisy vectors, $w_k$, $v_k$ must be white sequences with known covariance matrices. In addition, $w_k$ and $v_k$ must be uncorrelated. In instances where these conditions are not fulfilled, one can augment the state vector [169] and estimate the correlated noisy samples $w_k'$, $v_k'$, from others $w_k$, $v_k$ that form an uncorrelated, orthogonal base. This enables the Kalman filter to work with coloured noise, even if it is non-stationary. Noise covariance matrices are given by:

$$E[w_k w_i^T] = \begin{Bmatrix} Q_k & i=k \\ 0 & i \neq k \end{Bmatrix} \tag{3}$$

$$E[v_k v_i^T] = \begin{Bmatrix} R_k & i=k \\ 0 & i \neq k \end{Bmatrix} \tag{4}$$

$$E[w_k v_i^T] = 0 \quad \forall\ k,i \tag{5}$$

At this point, it is assumed that an initial estimate of the process at the same point in time, $t_k$, is known. This estimation, called *a priori estimate*, will be denoted $\hat{x}_k^-$, where the superscript *minus* is a reminder that it is the best estimate prior to assimilating the measurement at $t_k$. The estimation error can be written as:

$$e_k^- = x_k - \hat{x}_k^- \tag{6}$$

and its associated error covariance matrix as:

$$P_k^- = E[e_k^- e_k^{-T}] = E[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T] \tag{7}$$

Now, a linear combination of the measurement $z_k$ is sought to improve the prior estimate $\hat{x}_k^-$:

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H_k \hat{x}_k^-) \tag{8}$$

where:

$\hat{x}_k$ is the updated estimate and,

$\mathbf{K_k}$ is the Kalman gain (yet to be determined).

The problem now is to find the particular blending factor $K_k$ that yields an optimal updated estimated under some reasonable criterion. Just as in the Wiener solution, a minimum mean-square error criterion is used. Toward this end, let us write the error covariance matrix associated with the *a posteriori* (updated) estimate:

$$P_k = E[e_k e_k^T] = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] \tag{9}$$

If eq.(2) is substituted into eq.(8), and then into eq.(9) a final result for eq.(9) is given by:

$$P_k = (I - K_k H_k)P_k^-(I - K_k H_k)^T + K_k R_k K_k^T \tag{10}$$

*The optimization problem is now equivalent to finding the optimum gain at each time $t_k$, $K_k$, that minimizes the error variances for the elements of the state vector being estimated.* As the individual terms along the major diagonal of $P_k$ do represent these error variances, the optimization problem can be solved by minimizing the trace of $P$. If we follow straightforward matrix differentiation analysis and the next two formulas are considered:

$$\frac{d[trace(\mathbf{AB})]}{dA} = \mathbf{B}^T \quad (\mathbf{AB} \text{ must be square}) \tag{11}$$

$$\frac{d[trace(\mathbf{ACA}^T)]}{dA} = 2\mathbf{AC} \quad (\mathbf{C} \text{ must be symmetric}) \tag{12}$$

it can be formulated that:

$$\frac{d[tr(P)]}{dK} = -2(HP^-)^T + 2K(HP^-H^T + R) \tag{13}$$

Now, the derivative can be set equal to zero and the optimal gain, the Kalman gain, found. The result is:

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \tag{14}$$

Once the Kalman gain is known, if it is substituted in eq.(10), the *a priori* and *a posteriori* error covariance matrices can be related:

$$P_k = (I - K_k H_k)P_k^- \tag{15}$$

Now, note that both eq.(8) and the Kalman gain given in eq.(14) provide a means of assimilating the measurement at $t_k$ in a recursive relation that makes use of the *a priori* variables $\hat{x}_k^-$, $P_k^-$. For that reason a similar need for $\hat{x}_{k+1}^-$ can be anticipated at next step

in order to assimilate the next measurement $z_{k+1}$. This new *a priori* estimate can be computed if we project ahead the old estimate $\hat{x}_k$ via a transition matrix:

$$\hat{x}_{k+1}^- = \Phi_k \hat{x}_k \tag{16}$$

Finally, the new *a priori* error covariance matrix $P_{k+1}^-$ is computed in this way:

$$P_{k+1}^- = E[e_{k+1}^- e_{k+1}^{-T}] = E[(\Phi_k e_k + w_k)(\Phi_k e_k + w_k)^T] = \Phi_k P_k \Phi_k^T + Q_k \tag{17}$$

The needed quantities at time $t_k$ given, the measurement $z_{k+1}$ can be assimilated just as in the previous step. Thus, it is a routine matter to cycle through the Kalman loop comprising eqs.(8), (14), (15), (16) and (17) (see Fig.1). The loop begins with the *a priori* estimates of the state vector $\hat{x}_0^-$ and the error covariance matrix $P_0^-$.
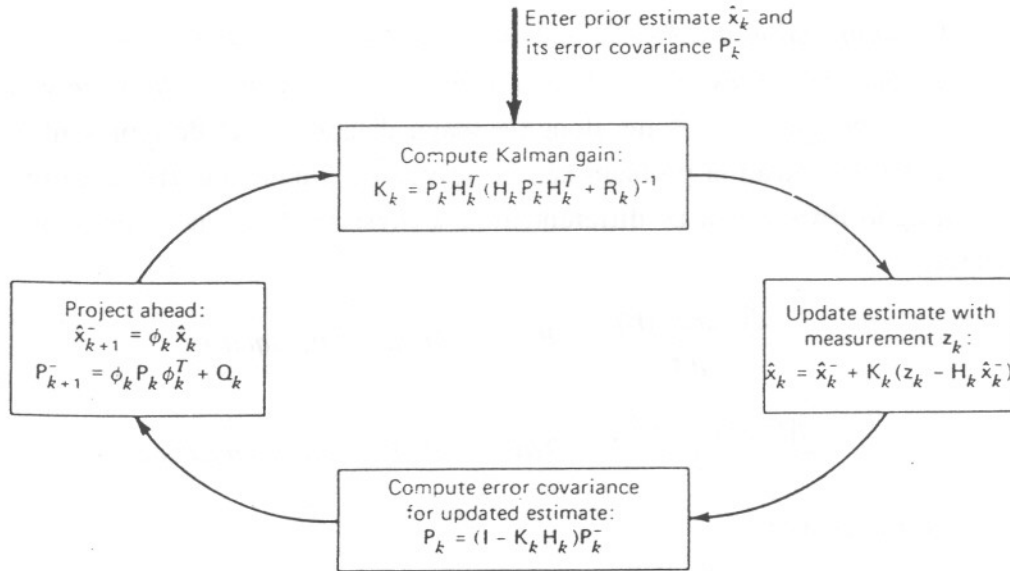


*Fig.1 Kalman filter loop.*

## 1.2 Extended discrete Kalman filter

Some of the most successful applications of the Kalman filter arise in situations where nonlinear dynamics and measurements have to be considered. This is the present case. In these instances, the stochastic process to estimate and the measurement relation can be written in its most general form as:

$$x_{k+1} = f_k(x_k) + w_k \tag{18}$$

$$z_k = h_k(x_k) + v_k \tag{19}$$

where $f_k$ and $h_k$ are nonlinear functions. The same constraints given to $w_k$ and $v_k$ in Sect.1.1 also apply here.

To begin with, let us approximate these functions by their Taylor's series expansion around the *a posteriori* $\hat{x}_k$ and *a priori* $\hat{x}_k^-$ estimates, respectively, and let us retain first-order terms only. It is found that:

$$f_k(x_k) \approx f_k(\hat{x}_k) + \frac{\partial f_k(x)}{\partial x}\Big|_{x=\hat{x}_k} (x_k - \hat{x}_k) \tag{20}$$

$$h_k(x_k) \approx h_k(\hat{x}_k^-) + \frac{\partial h_k(x)}{\partial x}\Big|_{x=\hat{x}_k^-} (x_k - \hat{x}_k^-) \tag{21}$$

Note that these developments assume the following conditions:

$$|x_k - \hat{x}_k| \ll 1 \tag{22}$$

$$|x_k - \hat{x}_k^-| \ll 1 \tag{23}$$

as it is expected that there will not be large differences among the three variables. Yet, careful attention should be drawn to the fact that the extended Kalman filter is a risky one as the linearization process takes places about the filter's estimated trajectory of the state vector rather than about a precomputed nominal trajectory (Fig.2). This is to say, that the partial derivatives are evaluated along a trajectory that has been updated with the filter's estimates; these, in turn, depend on the measurements, so the filter gain sequence will depend on the sample measurement sequence realized on a particular run of the experiment. Thus, the gain sequence is not predetermined by the process model assumptions as in the linear Kalman filter.

A general analysis of the extended Kalman filter is difficult because of the feedback of the measurement sequence into the process model. However, qualitatively it would seem to make sense to update the trajectory that is used for the linearization (after all, one may well wonder, why use the old trajectory when a better one is available ?). The flaw of this argument is this: the *better* trajectory is only better in an *statistical sense*. There may be a chance that the updated trajectory be poorer than the nominal one. In that event, the estimates will be poorer and this, in turn, will lead to further error in the trajectory, which causes further errors in the estimates, and so forth and so forth, leading to eventual divergence of the filter. The net result is that it should be cautioned that the *extended Kalman filter is a risky one, especially in situation where the initial uncertainty and measurement errors are large.* Yet, it may be better *on the average* that the regular filter.

Following a similar mathematical analysis to Sect.1.1, one can bridge gulfs with the classical linear filter if the equivalent matrices $F_k$ and $H_k$ are defined in the following way:

$$F_k = \frac{\partial f_k(x)}{\partial x}\Big|_{x=\hat{x}_k} \tag{24}$$

$$H_k = \frac{\partial h_k(x)}{\partial x}\Big|_{x=\hat{x}_k^-} \qquad (25)$$

If they are identified with the first-order terms of eqs.(20) and (21), they yield to:

$$x_{k+1} \approx f_k(\hat{x}_k) + F_k(x_k - \hat{x}_k) + w_k \qquad (26)$$

$$z_k \approx h_k(\hat{x}_k^-) + H_k(x_k - \hat{x}_k^-) + v_k \qquad (27)$$

These equations represent the linearized version of the filter and look very much like eqs.(1) and (2), except for the fact that rather that presenting to the filter total quantities, incremental ones are considered. In relation to eqs.(26) and (27), these are:

$$\Delta x_k = x_{k+1} - f_k(\hat{x}_k) \qquad (28)$$

$$\Delta z_K = z_k - h_k(\hat{x}_k^-) \qquad (29)$$

For further insight into how the filters keeps track of the total estimates see [169].
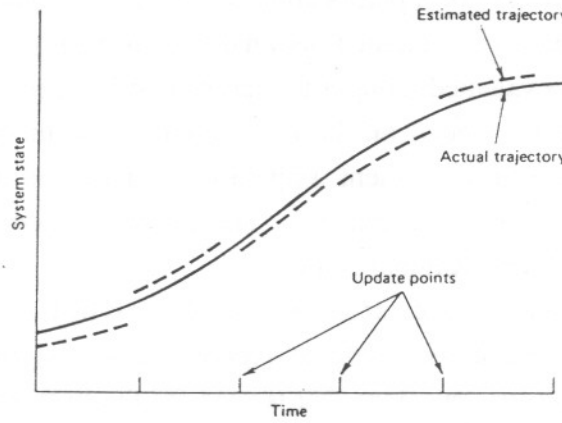


*Fig.2 Reference and actual trajectories*
*of an extended Kalman filter [169].*

To sum up, extended Kalman filter's recursive equation set becomes:

$$\hat{x}_k = \hat{x}_k^- + K_k \cdot [z_k - h_k(\hat{x}_k^-)] \qquad (30)$$

$$P_k = (I - K_k H_k) P_k^- \qquad (31)$$

$$\hat{x}_{k+1}^- = f_k(\hat{x}_k) \qquad (32)$$

$$P_{k+1}^- = F_k P_k F_k^T + Q_k \qquad (33)$$

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \qquad (34)$$

For completeness, let us give a similar example to that of Sect.1.1: Consider the case of plane's position radar monitoring in a pulse position indicator (PPI). As far as it concerns to the system under study, $x_k$ would be the plane's space coordinates $(x, y, z)$, $\Phi_k$ would be the tridimensional matrix describing the simple uniform 3-D movement that relates aircraft's position form one sample time to the following and $w_k$ would be the noisy process modelling the drift in the plane's trajectory. As for the observation system, $H_k$ would be the nonlinear equations that relate the tridimensional spatial coordinates to the PPI ones. $z_k$ would be the polar coordinates of the PPI (note how this involves a change in dimension) and finally, $v_k$ would be the electronic thermal noise added (that is, the measurement noise).

### 1.3 Relationship to deterministic least-squares

Both Kalman and Wiener filtering are sometimes referred to simply least-squares filtering. Though the topic is widely developed in [169], this is an oversimplification. *The criterion for optimization is* minimum mean-square error *for the Kalman filter and not the* squared error *as it is in a deterministic sense (least-squares of Chap.7).*

It is possible, however, to recapitulate the conditions under which the Kalman filter estimates coincides with the deterministic least-squares estimate [203]:

1.-     *The system state vector must be assumed a random constant* (the dynamics are thus trivial),

2.-     *the measurement sequence, $z_k$, must yield an overdetermined set of linear equations* and,

3.-     *no prior knowledge about the vector being estimated must be assumed* $(P_0 \to \infty)$.

In instances were the covariance matrices are assumed diagonal and uniform $(P_k = \sigma_p^2 I,\ Q_k = \sigma_q^2 I,\ R_k = \sigma_r^2 I)$, it has been possible for the author of this work to derive asymptotic relations about the filter's performance (Tab.1), which have been corroborated during the simulation process. They all stem from eqs.(30) to (34).

| | GAIN $K_k$ | $P_k$ | $\hat{x}_k$ | COMMENTS |
|---|---|---|---|---|
| $\sigma_p \to 0$ or $\sigma_r \to \infty$ | 0 | $P_k^-$ | $\hat{x}_k^-$ | no update possible as $K_k = 0$ filter strongly confident on past data |
| $\sigma_p \to \infty$ or $\sigma_r \to 0$ | $H_k^{-1}$ | 0 | $H_k^{-1} z_k$ | estimate only depends on present data least-squares behaviour |

*Tab.1 Analysis of asymptotic error covariance matrices.*

This latter assumption is unusual because in many situations, as it is the lidar case, there is at least some prior knowledge of the processes being estimated. *One of the things that distinguishes the Kalman filter from other estimators is the convenient way in which it accounts for this prior knowledge via the initial conditions of the recursive process. Of course, if there is no prior knowledge to use, the Kalman filter advantage is lost (in this respect), and it degenerates to a least-squares fitting under the conditions just stated.*

## 2. LIDAR INVERSION PHILOSOPHY USING THE KALMAN FILTER

The inversion of lidar signals using classical methods such as *slope-method, least-squares, Klett* or any of the approaches given in Chap.7 suffers from the basic fault of lacking memory. All the inverted profiles, does not matter whether the results were homogeneous of inhomogeneous, were based on present-data realizations of the lidar return-signal. For each data stream received, a new inversion, completely independent from those previously done, was performed. As a result, *no advantage was taken from past-time correlation* except for an increase in signal-to-noise ratio due to pulse integration, if used.

*Kalman filter major interest is to estimate for nearly every observation cell the dynamic behaviour of the above mentioned optical parameters, as it is often the case in real atmospheres. This comes as a result of making the most of correlation.* As long as different received realizations income, the filter updates itself weighted by the on-line unbalance between the *a priori* estimates (based on past realizations) and the new ones. Thus, the project ahead step in the filter or *a posteriori* estimate is performed based on a *minimum variance estimator*.

Moreover, variables in atmospheric phenomena are in general nonlinearly related. Unless microscale analysis is considered and plenty of boundary constraints from remote meteorological stations given, the struggle to physically model these optical parameters results awkward and cumbersome. Noticing that neither the visibility margin nor the signal-to-noise ratio change swiftly, it has prompted an stochastic model of the atmosphere (Sect.4) that models its macroscopic effects on the optical parameters using a *time-space correlation graph*. It devises correlation links among the cells remotely sensed within the lidar range, while keeping physic phenomena in the background.

As a result, it will be shown that not only does the filter its best at estimating the optical parameters along the observation cells but also models the nonlinear dynamics of a long-term slow-drift atmosphere.

## 3. FIRST APPROACH TO THE INVERSION: THE WHITE NOISE ATMOSPHERE

It is wished now to derive the extinction- and backscatter-coefficient over the entire lidar range. It is wished, then, to solve the functions $\alpha(R,t)$ and $\beta(R,t)$ that, in a minimum mean-square sense, best fit the observable power $P(R,t)$ at every time $t_k$. (Recall that the term *mean*, refers here to the ensemble average of lidar return observables).

To tackle the problem, a twofold study is presented: In the first place, a successful undersampling filter is explained (Sect.3.1), leaving the sampling rate filter in second place (Sect.4) as it will be shown to be not observable. As it can be guessed, they came across in the opposite direction but they are presented in reverse order for clarity reasons.

### 3.1 The undersampling filter

From the outset, there have been considered spatial samples every $\Delta R$ and lidar minimum range, $R_{min}$. Given the acquisition system sampling rate, $f_s$, it can be easily related to the former parameters by the expression

$$\Delta R = \frac{c}{2f_s} \tag{35}$$

If the spatial sampling points are

$$R_i = R_{min} + (i-1)\Delta R \qquad i=1..N \tag{36}$$

the power samples will be

$$P_i \equiv P(i\Delta R) \qquad i=1..N \tag{37}$$

Now, if the *extinction- and backscatter-functions* are decimated by a factor of 2 (that is, it is considered a sampling period $2\Delta R$, twice that of the power), *N/2 extinction samples* and *N/2 backscatter samples* will be estimated. (For simplicity, assume that $N$ is an even number). This can be expressed as

$$\alpha_i \equiv \alpha(R_{min} + [i-1]2\Delta R) \qquad i=1..\frac{N}{2}$$

$$\beta_i \equiv \beta(R_{min} + [i-1]2\Delta R) \qquad i=1..\frac{N}{2} \tag{38}$$

These points settled, it emerges that these two halves of *N/2* variables will form the state vector to solve

$$x \equiv \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{N/2} & \beta_1 & \beta_2 & \cdots & \beta_{N/2} \end{bmatrix}^T \tag{39}$$

As the discrete Kalman filter works with the estimates of each spatial cell at times $t_k$, the following notation enables to combine both time-space concepts

$$\alpha_i(k) \equiv \alpha_i(t_k) \quad i=1..\frac{N}{2}$$

$$\beta_i(k) \equiv \beta_i(t_k) \quad i=1..\frac{N}{2} \tag{40}$$

### 3.2 Modelization of the atmosphere

The atmosphere dynamics have been modelled by means of the noise state vector, $w_k$ (eq.(18)). The extinction- and backscatter-noise-state vectors are defined as follows

$$\vec{w}^\alpha \equiv \begin{bmatrix} w_1^\alpha & w_2^\alpha & \cdots & w_{\frac{N}{2}}^\alpha \end{bmatrix}^T \tag{41}$$

$$\vec{w}^\beta \equiv \begin{bmatrix} w_1^\beta & w_2^\beta & \cdots & w_{\frac{N}{2}}^\beta \end{bmatrix}^T \tag{42}$$

The noise-state vector is defined as follows where $k$ is a reminder of time, $t_k$.

$$w_k = \begin{bmatrix} \vec{w}^\alpha(k) \\ \vec{w}^\beta(k) \end{bmatrix} \tag{43}$$

At this first stage, a low-variance white-gaussian noise is used. Noting that this noise has nothing to deal with the observation noise, it means that the optical parameters of each observation cell undergo a gaussian drift in time. Because the noise-state vector is zero bias, it is equivalent to a steady-state atmosphere (on average no significative perturbations add to the optical parameter figures). As white noise involves wide-band behaviour, which is not the case of the atmosphere, low constant variances are assumed.

As noise in each cell is independent from others, a further simplification is to assume that extinction and backscatter figures are of the same magnitude. (It is known that this two values are correlated by the Koshmieder's relation introduced in Chap.3). Yet, the simplification above can be justified on account of the fact that the algorithm assumes no correlation at all between the two parameters. Contrary to what happened to non-memory algorithms, where tight correlation relations should be assumed (power-law dependency, Ricatti's differential equation and the like), the above assumption paves the way to loose correlation relations (e.g. variables having a same stochastic behaviour), if any.

Summing up, the set of transition equations for this atmospheric model become

$$\alpha_i(k+1) = \alpha_i(k) + w_i^\alpha(k) \quad i=1..\frac{N}{2}$$

$$\beta_i(k+1) = \beta_i(k) + w_i^\beta(k) \quad i=1..\frac{N}{2} \tag{44}$$

In the vectorial form next, it is equivalent to eq.(18).

$$x_{k+1} = x_k + w_k \tag{45}$$

As regards to the state-noise covariance matrix, it is in agreement with the white behaviour of the atmosphere

$$Q_k = E[w_k w_k^T] = \sigma_q^2 \cdot I \tag{46}$$

where $I$ is the identity matrix and $\sigma_q^2$ is the state noise variance. Note that $Q_k$ is diagonal because noise is assumed to be uncorrelated among the observation cells.

### 3.3 The measurement equation

We reproduce the lidar equation (Chap.3), that relates the range-received power to the sought-after optical parameters

$$P(R) = \frac{A}{R^2} \beta(R) \exp\left[-2 \cdot \int_0^R \alpha(r)\, dr\right] \tag{47}$$

where:

$A$ is the system's constant,
$\alpha$ is the extinction-coefficient in $km^{-1}$,
$\beta$ is the backscatter-coefficient in $km^{-1}$,
$R$ is the range.

If according to eq.(37), it is assumed that each power sample corresponds to a spatial increment $\Delta R$, and a rectangle approximation is used to compute the integral in eq.(47), the observable power samples become

$$P_1 = \frac{A}{R_1^2} \beta_1 \exp(-2\alpha_1 R_{min}) \tag{48}$$

$$P_2 = \frac{A}{R_2^2} \beta_1 \exp[-2\alpha_1(R_{min} + \Delta R)] \tag{49}$$

$$P_3 = \frac{A}{R_3^2} \beta_2 \exp\{-2[\alpha_1(R_{min} + \Delta R) + \alpha_2 \cdot \Delta R]\} \tag{50}$$

$$P_4 = \frac{A}{R_4^2} \beta_2 \exp\{-2[\alpha_1(R_{min} + \Delta R) + \alpha_2 \cdot 2\Delta R]\} \tag{51}$$

The last power sample results

$$P_N = \frac{A}{R_N^2} \beta_{N/2} \exp\{-2[\alpha_1(R_{min}+\Delta R)+\sum_{i=2}^{N/2} \alpha_i \cdot 2\Delta R]\} \tag{52}$$

The $N$ power samples build the measurement vector. As this relation is clearly nonlinear, it is necessary to resort to the extended Kalman filter (Sect.1.2). In relation to eq.(19), the power samples are linked to the nonlinear function $h(x_k)$ in this way

$$h(x_k) = \begin{bmatrix} P_1(x_k) \\ P_2(x_k) \\ \vdots \\ P_N(x_k) \end{bmatrix} \tag{53}$$

The observation model of eq.(19) also takes into account the observation noise. It stands for both the electronic noise of the receiver and the modelling errors (e.g. numerical approximations). Since the receiver noise is the same at every time, does not matter whether the return-signal is received or not, it equally affects all the measurement cells. For this reason, its noise covariance matrix can be written as

$$R_k = E[v_k v_k^T] = \sigma_r^2 \cdot I \tag{54}$$

It is time, now, to compute the observation matrix $H_k$ $(n \times 2n)$ that linearizes somehow the nonlinear behaviour of eq.(47). This matrix can be split between two submatrices, $H_1$ and $H_2$ according to the two sets of parameters in eq.(39)

$$H=[H_1 \; H_2] \tag{55}$$

where

$$H_{ij}^{(1} = \frac{\partial P_i}{\partial \alpha_j} \qquad H_{ij}^{(2} = \frac{\partial P_i}{\partial \beta_j} \tag{56}$$

Finally, this yields to

$$H_1 = \begin{bmatrix} -2R_{min}P_1 & 0 & 0 & \cdots & 0 \\ -2(R_{min}+\Delta R)P_2 & 0 & 0 & \cdots & 0 \\ -2(R_{min}+\Delta R)P_3 & -2\Delta R\,P_3 & 0 & \cdots & 0 \\ -2(R_{min}+\Delta R)P_4 & -4\Delta R\,P_4 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -2(R_{min}+\Delta R)P_{N-1} & -4\Delta R\,P_{N-1} & -4\Delta R\,P_{N-1} & \cdots & -2\Delta R\,P_{N-1} \\ -2(R_{min}+\Delta R)P_N & -4\Delta R\,P_N & -4\Delta R\,P_N & \cdots & -4\Delta R\,P_N \end{bmatrix} \tag{57}$$

$$
H_2 = \begin{bmatrix}
\dfrac{P_1}{x_{N/2+1}} & 0 & 0 & ..... & 0 \\[2em]
\dfrac{P_2}{x_{N/2+1}} & 0 & 0 & ..... & 0 \\[2em]
0 & \dfrac{P_3}{x_{N/2+2}} & 0 & ..... & 0 \\[2em]
0 & \dfrac{P_4}{x_{N/2+2}} & 0 & ..... & 0 \\[2em]
.... & .... & .... & .... & .... \\[1em]
0 & 0 & 0 & ..... & \dfrac{P_{N-1}}{x_N} \\[2em]
0 & 0 & 0 & ..... & \dfrac{P_N}{x_N}
\end{bmatrix}
\qquad (58)
$$

These two matrices, which represent the linearized observation dynamics, involve the evaluation of eqs.(48) to (52) at $\hat{x}_k^-$. That is how the Kalman filter estimates the observable trajectory at every time, $t_k$.

### 3.4 Initialization

Initialization is a twofold process: It comprises initial guesses of both the *a priori* estimate $\hat{x}_k^-$ and the covariance matrices, $P, Q$ and $R$. In particular, the *a priori* error covariance matrix $P_0^-$ indicates the degree of uncertainty the guesses are expected to have. Terms along the main diagonal should be interpreted as error-terms whose square-root would represent the $\pm\sigma$-*confidence interval* of the a priori estimate $\hat{x}_k^-$. For this estimate an uniform hypothesis has been assumed, since there is no prior knowledge about what the atmospheric profile will be like. This is to say that both sets of optical parameters, extinction and backscatter, are assumed constant based on a guessed visibility margin and the Koshmieder's relation [9]. Though another and quite sensible possibility would be to guess an *a priori* estimate by means of some non-memory algorithm, this has not been considered as it is wished to explore the adaptive possibilities of the filter. In particular, it is wished to investigate if the filter can keep track of the inhomogeneous atmospheric profiles departing from a simple homogeneous one.

As to the covariance matrix, $P_0^-$, it becomes

$$P_0^- = E[e_0^- \, e_0^{-T}] = \sigma_p^2 \cdot I \; ; \qquad e_0^- = x_0 - \hat{x}_0^- \tag{59}$$

where $e_0^-$ is the *a priori* initial estimation error. The larger $\sigma_p$, the larger the initial search interval of the filter. (Tab.2 lists values of other variables of interest).

## 3.5 Results

Owing to the fact that the two sets of optical parameters, *extinction and backscatter* have been modelled independently in Sect.3.1, it has been considered that it could serve a better purpose to evaluate the filter's performance if the atmospheric test profiles were synthesized using the same orders of magnitude for both optical parameters rather than others in better agreement to the reality (Tab.2).

As regards signal-to-noise ratio, medium to large values have been used to minimize noise observation effects on the filter.

| PARAMETER | VALUE | COMMENTS |
|---|---|---|
| $\alpha$ | $10^{-2}$ Nep/km$^{-1}$ | extinction-coefficient (synthesized profiles comply with this order of magnitude) |
| $\beta$ | $10^{-2}$ km$^{-1}$ | backscatter-coefficient (synthesized profiles comply with this order of magnitude) |
| $\sigma_p$ | $10^{-2}$ | Initial error standard deviation |
| $\sigma_q$ | $10^{-3}$ | State-vector mobility (std. deviation) |
| $\sigma_r$ | $10^{-1}$ W | Equivalent noise in received power |
| A | $10^4$ W·km$^3$·sr | System's constant. |
| N | 60 | State-vector dimension |
| SNR(R) | 90-40 dB | Signal-to-noise ratio from $R_{min}$ to $R_{max}$ |

*Tab.2 Simulation figures for the undersampling Kalman filter.*

According to Tab.2, from top to bottom, Fig.5 depicts two tridimensional plots of successive atmospheric states (first one at the top) along with Kalman estimates (second one). In the first one, it can be seen how the atmosphere drifts away from its initial state consisting in a soft two-hump-profile, to a final state, where the two humps have become

sharper and ridged. Recall that the state vector should be split into extinction (*1..N/2-th* components) and backscatter (*N/2+1..N-th* components) so that the tridimensional plots must be read considering that there are two vector halves in progress. The second one shows how the filter's estimates track the atmospheric states above, little by little. Beginning at $t_0$ from an uniform state (non represented), the results are fairly encouraging five iterations later. The third and fourth plots illustrate time evolution of the inversion relative error and a map of the gain matrix, $K_k$.

   Fig.3 and Fig.4 show a time-cut graph of the $t_5$-profiles for the SNR of Tab.2 (90-40 dB in Fig.3) and for a very high one (110-60 dB in Fig.4). Close inspection of these figures would reveal backscatter-errors of 8 % at $R_{max}$ in Fig.3 down to 3 % in Fig.4 as well as extinction errors of 70 % down to 40 %, respectively. At this point, it has to be stressed that errors increase in the high atmosphere, were the SNR is much lower. For instance, if errors were read in the low atmosphere, say less than *2 km,* extinction errors would easily be under 10 %, while backscatter ones would become negligible. Another point is that extinction rather than backscatter is more likely to be misestimated. Such differences arise from substantial different behaviours of these parameters in eqs.(57) and (58). Since backscatter-variables directly appear in the denominator of the elements of $H_2$, it is possible for them to exert a *higher control over the β-trajectories of the filter.* On the contrary, extinction-variables do not benefit from this situation because $H_1$-elements depend on the extinction quite indirectly via the model power, $P_j$.
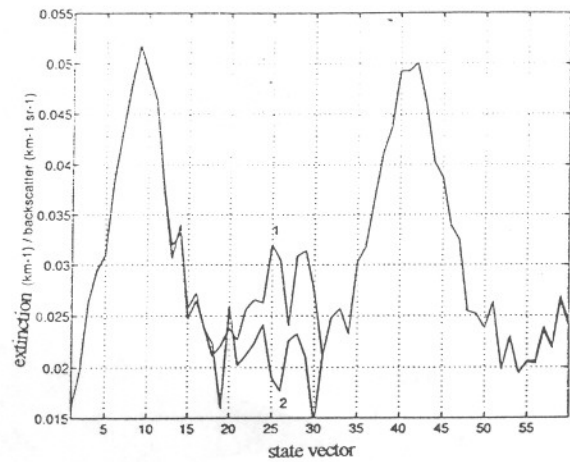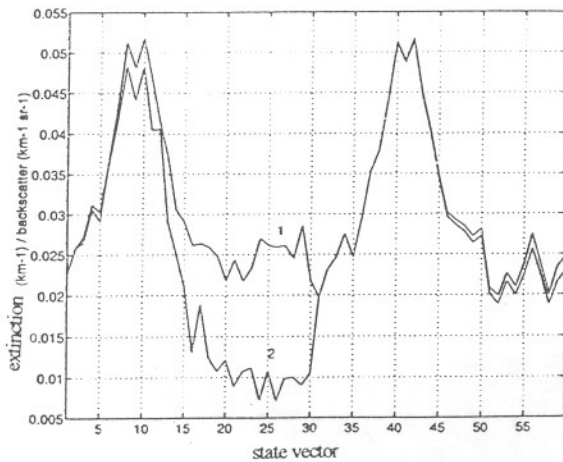


*Fig.3 State vector (1) and estimated one(2)*     *Fig.4 State vector (1) and estimated one (2)*
              *(SNR 90-40 dB).*                                          *(SNR 110-60 dB).*

state vector



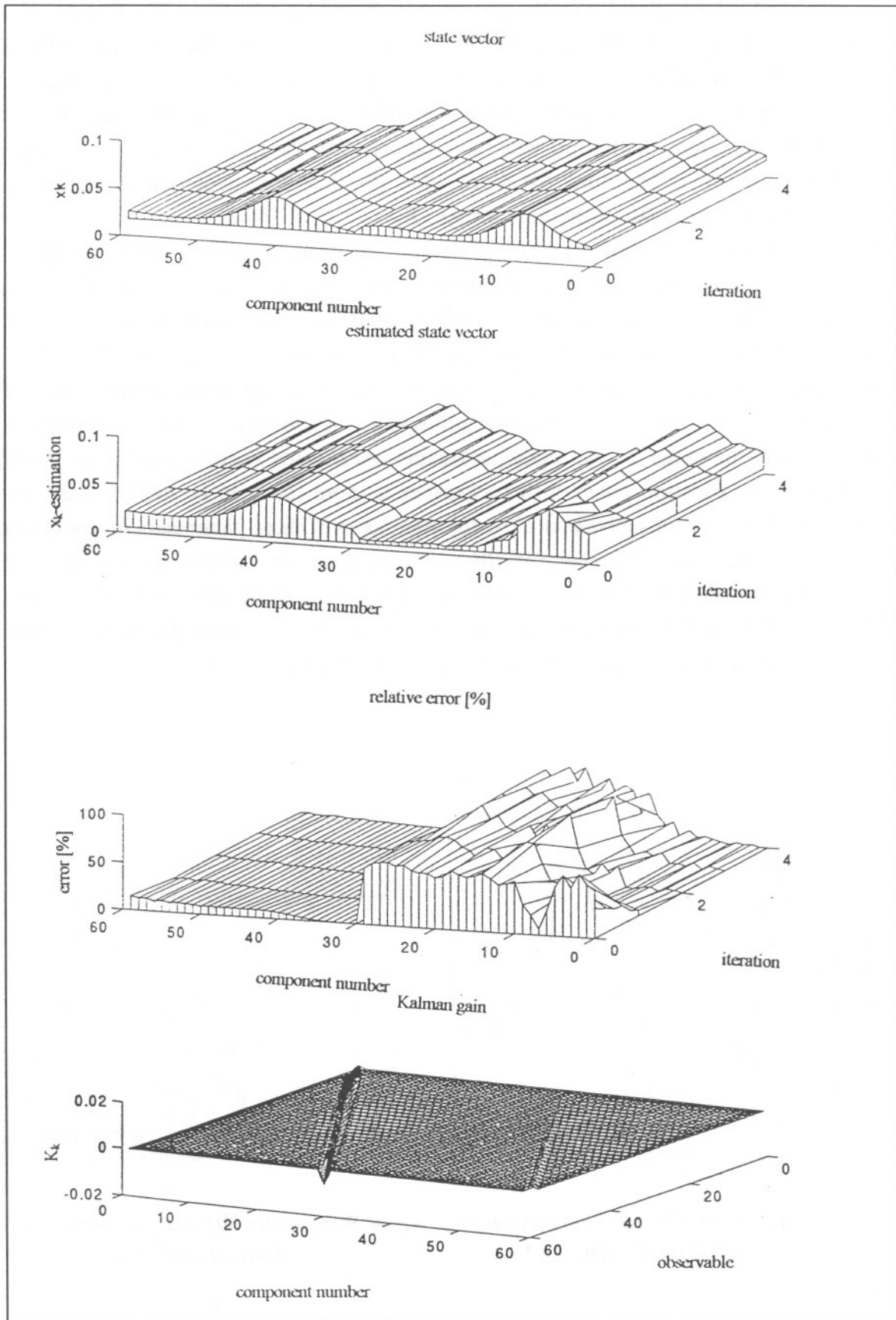estimated state vector



relative error [%]



Kalman gain



*Fig.5 Atmospheric evolution and Kalman estimates.*

Summing up, it can be said that the features of the input distribution are perfectly retrieved over the whole lidar range for the backscatter-coefficient and quite well for the extinction. Thus, *the undersampling Kalman filter fulfils lidar inversion needs fairly well, provided SNR is high enough and sensible initial choices are given.*

## 4. SECOND APPROACH TO THE INVERSION: GAUSS-MARKOV ATMOSPHERE

Kalman filter first approach to the lidar inversion has paved the way to a matched-pair of cooperating filters that model the atmosphere in a more sensible way. Noticing that there is room for improvement, the first enhancement has consisted in computing each optical parameter as the sum of two processes, one of constant behaviour plus another of stochastic nature. Thus, the first filter (Sect.4.3) will be devised to estimate the constant part, that represents a rough estimate of the atmosphere, whereas the second (Sect.4.4), taking a step further, will be aimed at estimating the stochastic increments of these parameters based on a correlation graph. It will be shown that the accuracy of the latter does only depend on having high SNRs, which even questions whether this second filter is worthwhile or not.

Not only an stochastic atmospheric model has been devised, but also there has been made room to include the intrinsic correlation between the two optical parameters, so often addressed in the literature [185][186][187]. Working on that ground, the first filter uses an adaptive linear relationship between extinction, $\alpha$, and backscatter, $\beta$. As far as it concerns the second filter, it is based on a markovian atmosphere (Sect.4.2).

### 4.1 The Gauss-Markov process

It comes as no surprise that the state vector does not undergo so swift changes in time as the white noise does. What the Gauss-Markov process [169][190], often called markovian noise, offers is a low-pass version of the white noise, equivalently, a time-correlated noise. Its autocorrelation function can be written as

$$R_y(\tau) = \sigma_m^2 \exp(-\beta |\tau|) \tag{60}$$

where $\sigma_m$ and $\beta$ are characteristic parameters of the process.

Its power can be computed by using that

$$P_y = E[y^2(t)] = R_y(0) = \sigma_m^2 \tag{61}$$

where it arises that $\sigma_m$ represents the Gauss-Markov standard deviation.

The colouring filter (shaping filter), $H(s)$, can be obtained by power spectral density decomposition as

$$S_y(s) = \frac{2\sigma_m^2\beta}{-s^2+\beta^2} = \frac{\sqrt{2\sigma_m^2\beta}}{s+\beta} \frac{\sqrt{2\sigma_m^2\beta}}{-s+\beta} = S_y^+(s) \cdot S_y^-(s) \qquad (62)$$

Thus, by identification

$$H(s) = S_y^+(s) = \frac{\sqrt{2\sigma^2\beta}}{s+\beta} \qquad (63)$$

Note that, here, it arises $\beta$ as the *3 dB* cut-off frequency. In addition, its discrete time counterpart can be written in the form of an *ARMA (AutoRegressive Moving Average)* process [27][28]

$$y_{k+1} = \exp(-\beta\Delta t)y_k + n_k \qquad (64)$$

where $y_k$ are the markovian samples, and $n_k$ the white samples.

By applying the expectancy operator to both sides of the above equation, it yields an important property,

$$m_y = \exp(-\beta\Delta t) \cdot m_y \;\Rightarrow\; m_y[1-\exp(-\beta\Delta t)]=0 \;\Rightarrow\; \left\{ \begin{array}{c} m_y=0 \\ \beta\Delta t=0 \end{array} \right\} \qquad (65)$$

Therefore, a Gauss-Markov process is a zero mean process (the mean is denoted $m_y$).

The parameter $\beta \cdot \Delta t$ can be interpreted as the inverse of the time correlation length of the process. Recalling that the *3 dB* cut-off frequency of a filter is equivalent to the inverse of its impulsional length or correlation time, $\tau_c$, in the time domain, it can be written that

$$\frac{1}{\beta\Delta t} = \frac{\dfrac{1}{\beta}}{\Delta t} = \frac{\tau_c}{\Delta t} = L_c \qquad (66)$$

where $L_c$ is the correlation length.

## 4.2 Modelization of the atmosphere
### 4.2.1 Temporal correlation

Let us note by $\alpha_c$, $\beta_c$ the extinction- and backscatter-time-constant terms, respectively, and by $\alpha_m$, $\beta_m$ the markovian terms associated. As their effects superimpose on each cell, it can be written that

$$\left\{ \begin{array}{l} \alpha(R) = \alpha_c(R) + \alpha_m(R) \\ \\ \beta(R) = \beta_c(R) + \beta_m(R) \end{array} \right. \qquad (67)$$

Many times in the literature, these two optical parameters have been correlated using linear or power relationships. Klett [185][186] joined these two assumptions into a single power-law expression

$$\beta(R) = C\alpha(R)^k \qquad 0.67 \leq k \leq 1 \qquad (68)$$

[9] relates these two parameters to the visibility margin and the Koshmieder's relation. So as to lessen the dependency, one can opt to only link their constant parts using a linear relationship *(k=1)* and *let the markovian components, which will be superimposed to the constant vectors, model the departures from the ideal case of linear correlation.* In this way a gross direct dependency between them is ensured without imposing a tight linear correlation on the total components $\alpha$ and $\beta$.

In addition, *the C-parameter has been given time adaptability as it will form part of the state-vector.* This can be expressed as

$$\alpha_k^c(R) = C_k \cdot \beta_k^c(R) \qquad (69)$$

where the subscript $k$ indicates that the estimates are given at time $t_k$.

Taking a step further, the following expression combines the above reasoning and the idea of an undersampling factor of value $M$ together. *The latter factor accounts for the fact that M cells share the same values of $\alpha$ and $\beta$.*

$$\begin{cases} \alpha_i = C \cdot \beta_i^c + \alpha_i^m & i = 1..\dfrac{N}{M} \\[2mm] \beta_i = \beta_i^c + \beta_i^m & i = 1..\dfrac{N}{M} \end{cases} \qquad (70)$$

(The time subscript $k$ has been skipped).

The markovian components (let us assume for the time being the $\alpha_m$-*components, though the same applies for* $\beta_m$*)* are expected to represent only a few percentage of the total extinction drift. As both $\alpha_c$ and $\beta_c$ are correlated, it seems sensible to relate the white noise standard deviation, $\sigma_w$, to the markovian one, $\sigma_m$, and it to the total extinction, $\alpha$, at any time. Using eq.(64) both deviations can be related like this [169]

$$\sigma_w = \sigma_m \sqrt{1 - \exp(-2\beta \Delta t)} \qquad (71)$$

Since the markovian components only represent small perturbations around the mean extinction and backscatter in each cell, the standard deviation, $\sigma_m$, will have to be some small percentage of the constant components.

In addition, as any markovian component may take negative values it is necessary to know which is the maximum amplitude of a markovian noise with standard deviation $\sigma_m$. A comprehensive collection of histograms about how markovian amplitudes distribute statistically have solved this question. Thus, if a markovian processes of power $\sigma_m^2$ is

considered, it is within $\pm 4\sigma_m$ dynamic range. A more conservative assumption recommends a $\pm 5\sigma_m$ limit. Mathematically, the total extinction will always be positive if

$$|\alpha_i^m| < 5 \cdot \sigma_i^m \qquad i=1..\frac{N}{M} \tag{72}$$

More realistically, the relationship between the markovian perturbations and their associated extinctions can be expressed by means of a factor $p$

$$|\alpha_i^m| < P \cdot \alpha_i \qquad i=1..\frac{N}{M} \tag{73}$$

If the upper bounds of eqs.(72) and (73) are equalled and $\sigma_i^m$ is substituted into eq.(71), it yields

$$\sigma_i^w = \frac{P}{5}\alpha_i \sqrt{1-\exp\left(-\frac{2}{L_c}\right)} \tag{74}$$

This important relation provides *the white noise power needed to produce a p-per-one markovian change in the extinction amplitude over a temporal correlation length of $L_c$ samples.*

### 4.2.2 Spatial correlation

In addition to the temporal correlation that exists among different samples of a same cell, nearby cells may well undergo similar changes in their optical parameters. Believing that temporal and spatial correlation are independent and recalling that markovian processes are filtered versions of white gaussian processes, one can take advantage of this and spatially correlate the white noise rather than the markovian. This is to say that two white noises or two markovian realizations, $i,j$ share a same correlation coefficient $\rho_{ij}$. Based on eq.(71), the relationship between their covariance matrices is:

$$C_m = \frac{C_w}{1-\exp(-2\beta\Delta t)} \tag{75}$$

where the subscripts $w$ and $m$ indicate *white* and *markovian*, as introduced before.

It is wished now to find the white correlated noise covariance matrix $C_w$ that will work as the state noise covariance matrix $Q_k$ of the filter. Towards this end, two different spatial correlation coefficients have to be distinguished: $\rho$, between one cell and the next one, which represents the correlation due to the physical continuity of the atmosphere, and $\rho'$, or correlation coefficient between extinction and backscatter of any single cell.
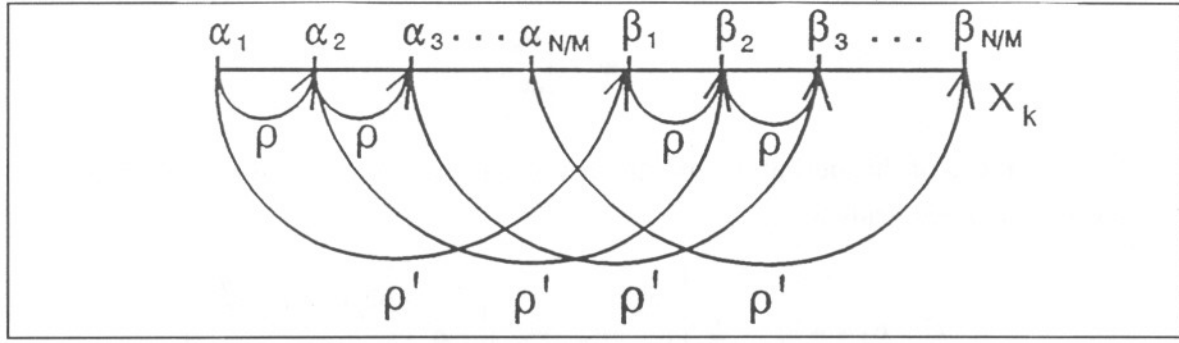
*Fig.6 Spatial correlation graph of the atmosphere.*

Mathematically, this can be expressed as

$$\rho_{\alpha_i \alpha_{i+1}} = \rho \qquad i = 1.. \frac{N}{M} - 1 \tag{76}$$

$$\rho_{\beta_i \beta_{i+1}} = \rho \qquad i = 1.. \frac{N}{M} - 1 \tag{77}$$

$$\rho_{\alpha_i \beta_i} = \rho' \qquad i = 1.. \frac{N}{M} \tag{78}$$

where $\alpha_i$, $\beta_i$ are the optical parameters of the *i-th* cell. The graph of Fig.6 sketches the correlation links amongst the cells considered. If some properties from open graph theory are assimilated, the correlation coefficient between two variables $X, Z$ can be computed by using a third variable $Y$.

$$\rho_{XY} = \rho_{XZ} \cdot \rho_{ZY} \tag{79}$$

In a similar manner, the following correlation coefficients apply for estimates located $j$ cells apart

$$\rho_{\alpha_i \alpha_{i+j}} = \rho^j \qquad i = 1.. \frac{N}{M} - 1 \quad j = 0.. \frac{N}{M} - i \tag{80}$$

$$\rho_{\beta_i \beta_{i+j}} = \rho^j \qquad i = 1.. \frac{N}{M} - 1 \quad j = 0.. \frac{N}{M} - i \tag{81}$$

$$\rho_{\alpha_i \beta_{i+j}} = \rho_{\alpha_i \alpha_{i+j}} \cdot \rho_{\alpha_{i+j} \beta_{i-j}} = \rho^j \cdot \rho' \qquad i = 1.. \frac{N}{M} - 1 \quad j = 0.. \frac{N}{M} - i \tag{82}$$

So as to give a final expression for the covariance matrix $C_w$, the correlated noise vector is defined as

$$w \equiv \begin{bmatrix} \vec{w}_\alpha \\ \vec{w}_\beta \end{bmatrix} \qquad (83)$$

The use of this definition and the fact that $C_w$ is a covariance matrix of zero mean aleatory variables leads to

$$C_w = E\{w\, w^T\} \equiv E\left\{ \begin{bmatrix} \vec{w}_\alpha \\ \vec{w}_\beta \end{bmatrix} \begin{bmatrix} \vec{w}_\alpha^T & \vec{w}_\beta^T \end{bmatrix} \right\} = E\left\{ \begin{bmatrix} \vec{w}_\alpha \vec{w}_\alpha^T & \vec{w}_\alpha \vec{w}_\beta^T \\ \vec{w}_\beta \vec{w}_\alpha^T & \vec{w}_\beta \vec{w}_\beta^T \end{bmatrix} \right\} \qquad (84)$$

or equivalently

$$C_w = \begin{bmatrix} C_{\alpha\alpha} & C_{\alpha\beta} \\ C_{\beta\alpha} & C_{\beta\beta} \end{bmatrix} \qquad (85)$$

where $C_w$ has been split in four submatrices representing the co- and crosscorrelation between $\alpha$ and $\beta$.

Using that $C_{\alpha\beta} = C_{\beta\alpha}^T$ and that for zero mean variables

$$cov(X,Y) = E(XY) = \sigma_X \cdot \sigma_Y \cdot \rho_{XY} \qquad (86)$$

a final expression for $C_w$ is reached

$$\begin{bmatrix}
\sigma_{\alpha_1}^2 & \rho\,\sigma_{\alpha_1}\sigma_{\alpha_2} & \cdots & \rho^{n-1}\sigma_{\alpha_1}\sigma_{\alpha_2} & \rho'\sigma_{\alpha_1}\sigma_{\beta_1} & \rho'\rho\,\sigma_{\alpha_1}\sigma_{\beta_2} & \cdots & \rho'\rho^{n-1}\sigma_{\alpha_1}\sigma_{\beta_n} \\
\cdots & \sigma_{\alpha_2}^2 & \cdots & \rho^{n-2}\sigma_{\alpha_2}\sigma_{\alpha_n} & \cdots & \rho'\sigma_{\alpha_2}\sigma_{\beta_2} & \cdots & \rho'\rho^{n-2}\sigma_{\alpha_2}\sigma_{\beta_n} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \sigma_{\alpha_n}^2 & \cdots & \cdots & \cdots & \rho'\sigma_{\alpha_n}\sigma_{\beta_n} \\
\rho'\sigma_{\beta_1}\sigma_{\alpha_1} & \rho'\rho\,\sigma_{\beta_1}\sigma_{\alpha_2} & \cdots & \rho'\rho^{n-1}\sigma_{\beta_1}\sigma_{\alpha_n} & \sigma_{\beta_1}^2 & \rho\,\sigma_{\beta_1}\sigma_{\beta_2} & \cdots & \rho^{n-1}\sigma_{\beta_1}\sigma_{\beta_n} \\
\cdots & \rho'\sigma_{\beta_2}\sigma_{\alpha_2} & \cdots & \rho'\rho^{n-2}\sigma_{\beta_2}\sigma_{\alpha_n} & \cdots & \sigma_{\beta_2}^2 & \cdots & \rho^{n-2}\sigma_{\beta_2}\sigma_{\beta_n} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \rho'\sigma_{\beta_n}\sigma_{\alpha_n} & \cdots & \cdots & \cdots & \sigma_{\beta_n}^2
\end{bmatrix} \qquad (87)$$

Before proceeding further, it is interesting to check if $C_w$ may represent a covariance matrix. $C_w$ does represent a covariance matrix if and only if $|\rho| < 1, |\rho'| < 1$. Then, all the angular minors of $C_w$, $\Delta$, become positive and $xC_wx^T$ is a quadratic form positive defined.

$$\Delta_N = \begin{cases} \left[\prod_{i=1}^{N}(C_w)_{ii}\right](1-\rho^2)^{N-2}(1-\rho^2)^{\frac{N}{2}} & N=even \\[2em] \left[\prod_{i=1}^{N}(C_w)_{ii}\right](1-\rho^2)^{N-2}(1-\rho^2)^{\frac{N-1}{2}} & N=odd \end{cases} \tag{88}$$

Once $C_w$ settled, the next point to tackle is the problem of building such $N$ white correlated processes from $N$ others, that are uncorrelated. The problem can be solved by finding the linear transformation, $A$, and the $N$ orthogonal white powers (in fact, independent) at its input (Fig.7).
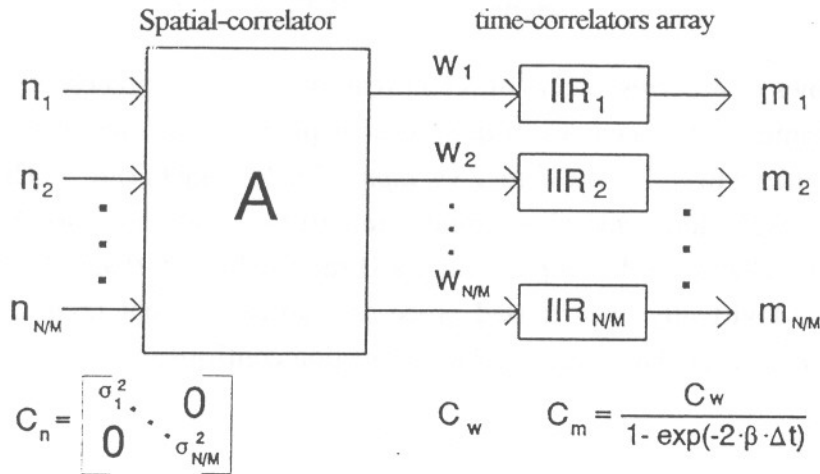


*Fig.7 Generation of spatially correlated markovian noise.*

Let $X$ be a vector of $N$ gaussian variables $X_1, X_2,\ldots, X_N$, $m_X$ the vector of means and $C_X$ its covariance matrix

$$C_X = E[(X-m_X)(X-m_X)^T] \tag{89}$$

and let $Y$ be a set of random variables $Y_1, Y_2,\ldots, Y_N$ linearly related to the $X_1, X_2,\ldots, X_N$ set via the linear equation

$$Y = AX + b \tag{90}$$

where $A$ must be a squared nonsingular matrix (i.e. invertible) and $b$ a constant vector. The following relationship can be proved [169]

$$m_Y = A\,m_X + b$$
$$C_Y = A\,C_X\,A^T \tag{91}$$

The transformation, $A$, that produces a diagonal covariance matrix $AC_XA^T$ yields a new set of normal random variables that are uncorrelated, and consequently, statistically independent. As we are interested in the correlating system which has independent samples

at its input, the linear system is represented by $A^T$, since $A$ is unitary (the bias vector $b$ is zero since $Y$ has zero mean). Hence, the solution takes the form of an eigenvalue problem: *the eigenvectors of $C_w$ are the columns of $A$ and the diagonal of $C_X$ (the vector of independent gaussian powers) is formed by the eigenvalues of $C_w$.*

So as to illustrate the physical meaning of eq.(87) and its synthesis from independent white noise processes, Fig.8 simulates the time-space behaviour of two cells (say, *cell_1* and *cell_2*) far apart $(\rho=0.05)$ with a high correlation in their in-cell optical parameters $(\rho'=0.95)$. The correlation length, $L_c$, is $L_c = 50\ samples$. Below the four processes on display there are also shown their crosscorrelation functions

$$R_{\alpha_i^m \alpha_j^m}(k) = E\left\{\vec{\alpha}_i^m(n)\cdot\vec{\alpha}_j^m(n+k)\right\}$$

$$R_{\alpha_i^m \beta_i^m}(k) = E\left\{\vec{\alpha}_i^m(n)\cdot\vec{\beta}_i^m(n+k)\right\}$$

$$(92)$$

where the rows of $R$ correspond to different realizations of the processes. If the four plots are cross-examined, two couples of quite similar processes can be identified: *beta_1* vs. *alpha_1* belonging to *cell_1* and *beta_2* vs. *alpha_2* belonging to *cell_2*. They are precisely the in-cell extinction- and backscatter-time-drifts. This is corroborated by the crosscorrelation function that is practically a Dirac's delta $(\rho=0.95)$. On the other hand, two couples of virtually uncorrelated processes can be told out (*beta_1* vs *beta_2* and *alpha_1* vs. *alpha_2*). The crosscorrelation function confirms the result $(\rho'=0.05)$ in the plot below.

### 4.3 First filter: the constant backscatter estimator

As discussed in the introduction of this section, the atmospheric model of eq.(70) means estimating both the vector $\beta_c$, that represents the gross-part long-term standing backscatter, and the correlation constant $C$. As before, $\beta_c$ is estimated every $M$ cells *(M $\geq$ 2)* for observability reasons (Sect.5). If the state vector is built as follows

$$x_k \equiv \begin{bmatrix} \vec{\beta}_k^c \\ C_k \end{bmatrix}$$

$$(93)$$

the dynamics of the state vector can be expressed by

$$\begin{bmatrix} \vec{\beta}_c(k+1) \\ C(k+1) \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \vec{\beta}_c(k) \\ C(k) \end{bmatrix}$$

$$(94)$$

If this is identified with the general form of the state model (eq.(1)), $\Phi_k = I$, $w_k = 0$ results. Note here that the state noise being nil, $Q_k$ becomes nil, and this leads the filter to an idle state.
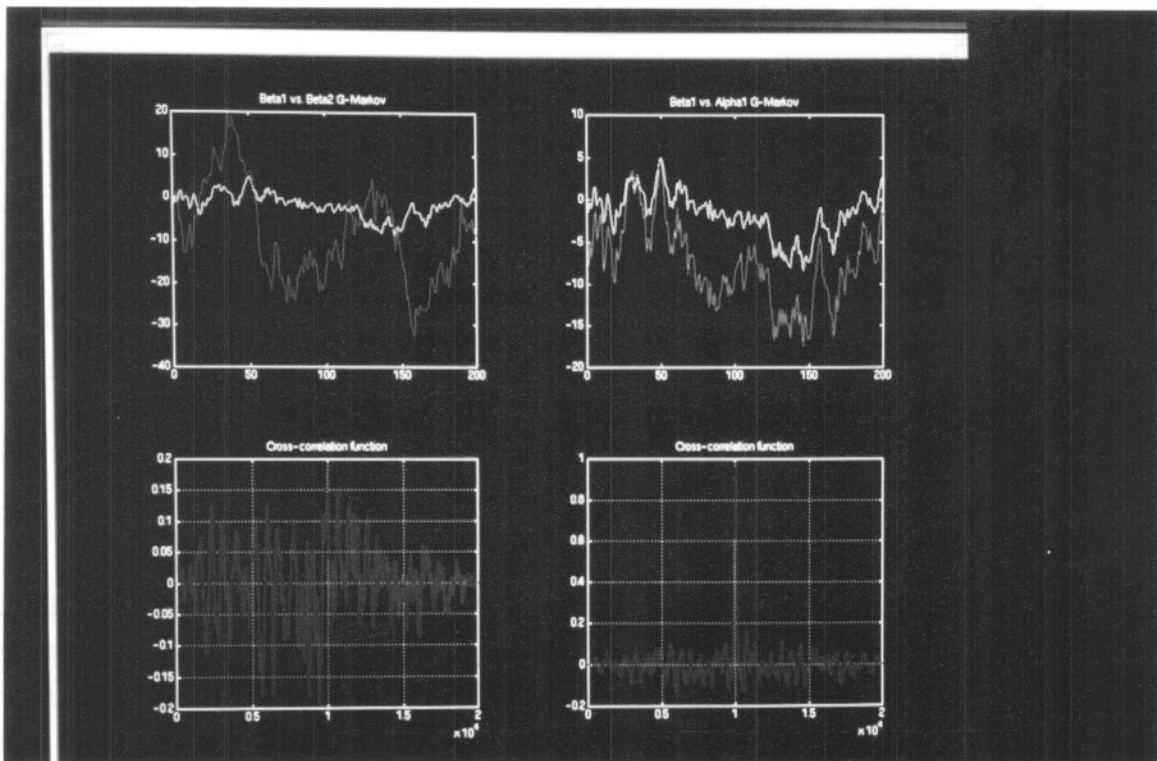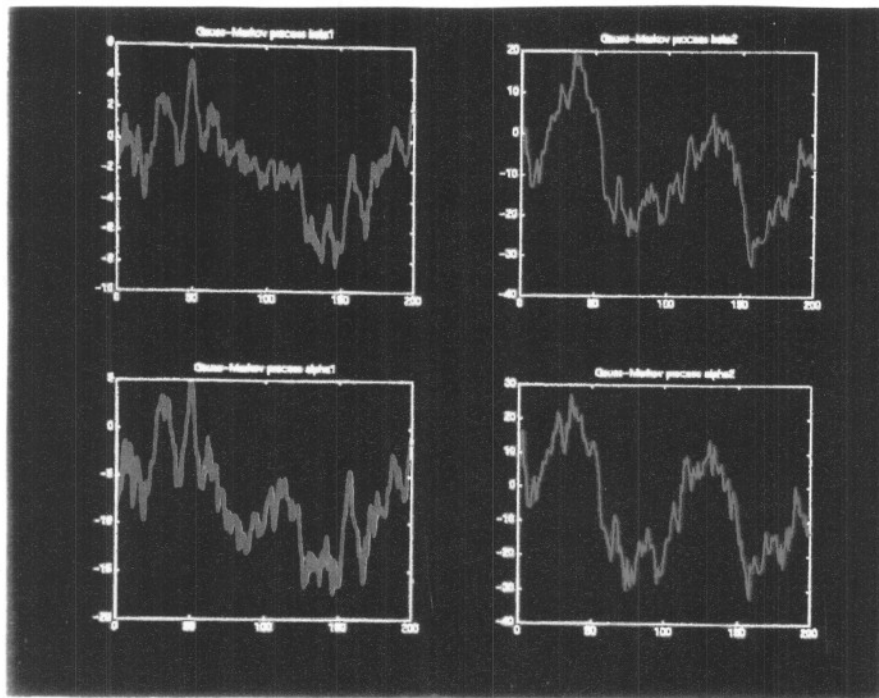
*Fig.8 Markovian noise processes modelling $\alpha_m$, $\beta_m$ of two cells (top)
and cross-correlation functions (bottom).*

So as to counteract the problem, convergence can be assured if non zero elements are set along the main diagonal of $Q_k$:

$$Q_k = \begin{bmatrix} \sigma_q^2 \cdot I & 0 \\ 0 & \sigma_{qk}^2 \end{bmatrix} \qquad (95)$$

where $\sigma_q$ and $\sigma_{qk}$ are on the order of a 10 % of the constant components $\beta_c$ and $C$, respectively ($I$ denotes the identity matrix).

As for the correlation matrix of the observables, $R_k$, a range dependent variance has been assumed, as so is the signal-to-noise ratio in a lidar system.

$$R_k = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} \qquad (96)$$

The *a priori* estimation $\hat{x}_0^-$ (whose dimension is N/M+1) becomes

$$\hat{x}_0^- = [\beta_0 , \cdots , \beta_0 , C_0]^T \qquad (97)$$

where $\beta_0$ is the estimated backscatter value and $C_0$ the initial estimate of $C$.

Its associated *a priori* error covariance matrix is

$$P_0^- = \begin{bmatrix} \sigma_p^2 \cdot I & 0 \\ 0 & \sigma_k^2 \end{bmatrix} \qquad (98)$$

where $\sigma_p^2$ is the variance of the constant backscatter components, $\sigma_k^2$ is the initial error variance of the constant $C$, and $I$ is the identity matrix.

According to (67) and (93), the state model is given by

$$\alpha_i = x_{N/M+1} x_i + \alpha_i^m \quad i=1..\frac{N}{M}$$

$$\beta_i = x_i + \beta_i^m \quad i=1..\frac{N}{M} \qquad (99)$$

where, in theory, $\alpha_i^m$ and $\beta_i^m$ should be entered to the algorithm as parameters since they do not belong to the state vector. In practice, the first filter cannot know them as the second filter is, in turn, fed by results of the first. On account of the fact that the markovian components are much lower than the constant ones and that they are zero mean processes, the following approximation applies

$$\alpha_i \approx x_{N/M+1} \, x_i \qquad i=1..\frac{N}{M}$$

(100)

$$\beta_i \approx x_i \qquad i=1..\frac{N}{M}$$

This represents a sensible approximation for the first filter as it means to replace the markovian components by their mean value, which is zero.

So as to evaluate the power derivatives with respect to the state variables, one must consider a relation of the following form together with the lidar equation (eq.(37)) that models the observables (the factor $M$ represents the undersampling factor)

$$\frac{\partial P_i}{\partial x_j} = \frac{\partial P_i}{\partial \alpha_j} \frac{\partial \alpha_j}{\partial x_j} + \frac{\partial P_i}{\partial \beta_j} \frac{\partial \beta_j}{\partial x_j} \qquad i=1..N \quad j=1..\frac{N}{M}$$

(101)

where the $\alpha_j$, $\beta_j$ derivatives are

$$\frac{\partial \alpha_j}{\partial x_j} = x_{N/M+1} \qquad \frac{\partial \beta_j}{\partial x_j}=1 \qquad j=1..\frac{N}{M}$$

(102)

Lastly, if the lidar equation is developed as it was done with eqs.(48) to (52), but following the considerations above, the equivalent observation matrix $H$ takes the form

$$H=[H_1 \quad H_2]$$

(103)

where the $H_1$ is the observation matrix of the constant backscatter components and $H_2$ that of the adaptive constant $C$. These are their closed expressions

$$H_1 = \begin{bmatrix} \left(-2C\,R_{min}+\frac{1}{x_1}\right)P_1 & 0 & 0 & ..... & 0 \\[2mm] \left[-2C(R_{min}+\Delta R)+\frac{1}{x_1}\right]P_2 & 0 & 0 & ..... & 0 \\[2mm] ...... & ..... & ..... & ..... & ..... \\[2mm] \left\{-2C[R_{min}+(M-1)\Delta R]+\frac{1}{x_1}\right\}P_M & 0 & 0 & ..... & 0 \\[2mm] -2C[R_{min}+(M-1)\Delta R]\,P_{M+1} & \left(-2C\Delta R+\frac{1}{x_2}\right)P_{M+1} & 0 & ..... & 0 \\[2mm] ............ & ..... & ..... & ..... & ..... \\[2mm] -2C[R_{min}+(M-1)\Delta R]\,P_N & -2C\,M\,\Delta R\,P_N & -2C\,M\,\Delta R\,P_N & ..... & \left(-2C\,M\,\Delta R+\frac{1}{x_{N/M}}\right)P_N \end{bmatrix}$$

(104)

$$H_2 = \begin{bmatrix} -2R_{\min}x_1\,P_1 \\ -2(R_{\min}+\Delta R)x_1\,P_2 \\ \vdots \\ -2[R_{\min}+(M-1)\Delta R]x_1\,P_M \\ -2\{[R_{\min}+(M-1)\Delta R]x_1+\Delta R x_2\}\,P_{M+1} \\ \vdots \\ -2\{[R_{\min}+(M-1)\Delta R]x_1+M\,\Delta R\,x_2+...+M\,\Delta R\,x_{N/M-1}+M\,\Delta R x_{N/M}\}\,P_N \end{bmatrix} \qquad (105)$$

*This first filter has been tested using an atmospheric simulator according to the time-space correlated atmosphere discussed in Sect.4.2 and following the computational steps presented in this Sect.4.3. This is shown diagrammatically in Fig.9.*
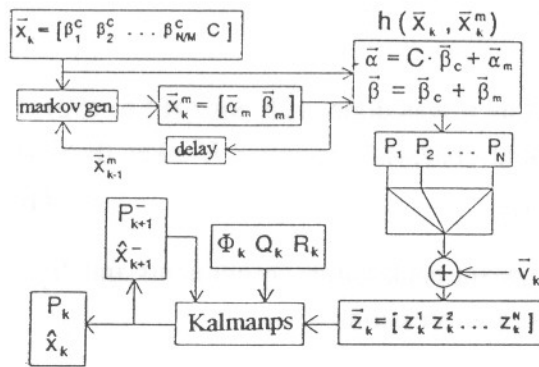


Fig.9 Block diagram of filter 1.          Fig.10 Block diagram of filter 2.
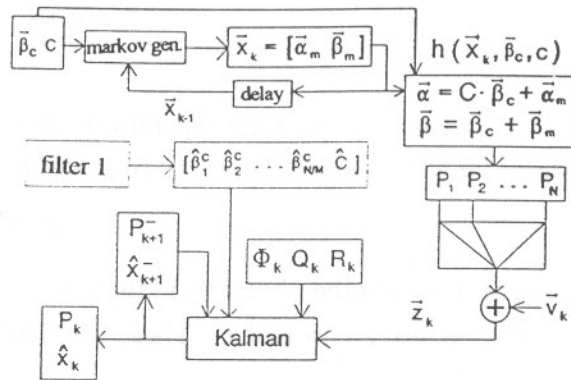
## 4.4 Second filter: the markovian estimator

The ultimate goal of this second filter is to estimate the markovian components based on the constant backscatter components estimated by the first filter. Here, they will work as parameters rather that variables of the model. Obviously, both filters must share the same undersampling factor, $M$. The following system stems from this idea

$$\alpha_i = \hat{C}\hat{\beta}_i^c + x_i \quad i=1..\frac{N}{M}$$

$$\qquad (106)$$

$$\beta_i = \hat{\beta}_i^c + x_{N/M+i} \quad i=1..\frac{N}{M}$$

where the hat (^) denotes estimated. Working on that ground and making use of the model developed in Sect.4.2. the state model is expressed by

$$\begin{bmatrix} \vec{\alpha}_m(k+1) \\ \vec{\beta}_m(k+1) \end{bmatrix} = \exp(-B\Delta t)\cdot I \cdot \begin{bmatrix} \vec{\alpha}_m(k) \\ \vec{\beta}_m(k) \end{bmatrix} + \begin{bmatrix} \vec{w}_\alpha(k) \\ \vec{w}_\beta(k) \end{bmatrix} \qquad (107)$$

where $I$ is the identity matrix.

The state covariance matrix, $Q_k$, coincides with that of the correlated noise $C_w$. Note that the model of eq.(107) is necessary because the Kalman filter model of eq.(1) has to deal with white sequences. In fact, what has been done is to augment the state vector dimension to accommodate the markovian noise.

The *a priori* error covariance at $t_0$, $P_0^-$ is assumed to be

$$P_0^- = C_w(0) \qquad (108)$$

In the same way we did in the previous section, the observation matrix $H$ can be divided into two submatrices, $H_1$ and $H_2$, one for the markovian extinctions and the other for the markovian backscatterers.

$$H_1 = \begin{bmatrix} -2R_{min}P_1 & 0 & 0 & \cdots & 0 \\ -2(R_{min}+\Delta R)P_2 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -2[R_{min}+(M-1)\Delta R]P_M & 0 & 0 & \cdots & 0 \\ -2[R_{min}+(M-1)\Delta R]P_{M+1} & -2\Delta R P_{M+1} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -2[R_{min}+(M-1)\Delta R]P_N & -2M\Delta R P_N & -2M\Delta R P_N & \cdots & -2M\Delta R P_N \end{bmatrix} \qquad (109)$$

$$H_2 = \begin{bmatrix} \dfrac{P_1}{\hat{\beta}_1^c+x_{N/M+1}} & 0 & 0 & \cdots & 0 \\ \dfrac{P_2}{\hat{\beta}_1^c+x_{N/M+1}} & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots \cdots & & \cdots \\ \dfrac{P_M}{\hat{\beta}_1^c+x_{N/M+1}} & 0 & 0 & \cdots & 0 \\ 0 & \dfrac{P_{M+1}}{\hat{\beta}_2^c+x_{N/M+2}} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots \cdots & & \cdots \\ 0 & 0 & 0 & \cdots & \dfrac{P_N}{\hat{\beta}_{N/M}^c+x_{2N/M}} \end{bmatrix} \qquad (110)$$

Each iteration must be evaluated at the a priori estimate $\hat{x}_k^-$.

This second filter has been tested using an atmospheric simulator according to the time-space correlated atmosphere discussed in Sect.4.2 and following the computational steps presented here. This is shown diagrammatically in Fig.10.

## 4.5 Results and discussion

Several simulations have been performed with the matched-pair of Kalman filters presented. In many of them, especially in instances where large values of C $(C \geq 10)$ have been used, the error covariance matrix $P$ has turned out to be bad-conditioned (rank deficient) due to numerical round-off errors when running on a PC486DX platform. This is equivalent to say that the eigenvalue span of $P$ is far too large according to the Rayleigh criterion [28]. To counteract the problem, UD matrix factorization [169][167] was first tried, though unsuccessful results were obtained. Finally, the problem has been solved using sequential processing techniques in the kernel of the first Kalman filter. This has enabled C-values to span as large as $(C \geq 100)$. Square-root processing [165] may also bring enhanced numerical stability and possibly Cholesky factorization. In the present case, sequential processing of the measurement data, along with matrix symmetrization [169] whenever possible, has given the final thrust to the algorithms presented.

Among the comprehensive set of simulations, the four most illustrative are summarized in Tab.4 (p.8.35). Tab.3 lists the default parameters used in the simulations.

---

**KALMAN FILTER DEFAULTS**

*Atmospheric default parameters:*

$\beta_c = 2 \cdot 10^{-2} : 5 \cdot 10^{-2}$ km$^{-1} \cdot$sr$^{-1}$       C = 33.33 sr

$R_{max} = 1$ km;                         $R_{min} = 0.2$ km;                         $\Delta R = 42.1$ m

*Filter parameters:*

a) *Global sampling parameters:*            N=20,                         M=2

b) *First filter:*

Iterations $I_1 = 20$

$Q_k$ (state error covariance):       $\sigma_q = 10^{-3}$ km$^{-1} \cdot$sr$^{-1}$;           $\sigma_{qk} = 1$ sr

$P_0^-$ (a priori covariance):        $\sigma_p = 10^{-2}$ km$^{-1} \cdot$sr$^{-1}$;           $\sigma_k = 3$ sr

Initialization:                       $\beta_0 = 10^{-2}$ km$^{-1} \cdot$sr$^{-1}$;           $C_0 = 30$ sr

c) *Second filter:*

Iterations $I_2 = 20$

Correlation coeff:           $\rho$ (adjacent-cell) = 0.9        $\rho'$(inside cell) = 0.95

p (see eq.(74)) = 10 %       $L_c$ (correlation length) = 1000

---

***Tab.3 Defaults used in the simulations.***

In the first place, note that SNRs of Tab.4 are range-dependent as it is the case in lidar systems. Values are in reasonable agreement to real cases. For this reason, high values correspond to lower ranges (0.2 km) and vice versa.

In the first simulation of Tab.4, atmospheric simulated realizations vs. first filter estimations are compared (Fig.11). XY-axis show the state vector vs. time. Extinction is depicted at the top half of the figure and backscatter at the bottom one, with the exception of the last vector state component that represents the correlation constant, $C$ (eq.(93)). There, atmospheric history illustrates successive atmospheric states the filter has had to estimate departing from the initial guesses. For this reason, the simulated atmospheric vector is named *state vector* while that of the filter is named *estimated state vector*. Recall that, for the first filter, the atmospheric vector is constant with time. Also, note how after eight iterations, the filter has identified the state vector quite accurately. Constant and markovian atmospheric evolutions have been plotted in separate figures as their superimposed contributions are difficult to tell apart graphically (the markovian part is much more lower). As no adaptive capability has been given to the correlation variables $\rho$, $\rho'$, $p$, $Lc$, both the second filter and the atmospheric simulator have shared the same couple of values.

The normal running of the algorithm implies that once the first filter (the constant estimator) has successfully estimated the atmospheric state vector, results should be passed to the second filter (Fig.10) or markovian estimator. Of course, the quality of the estimation of this filter will certainly depend on how accurate these results are. Thus, simulations 3 and 4 in Tab.4, study the effect of passing to the second filter exact results (of course, known from the simulator) or the estimated ones by the first filter.

About that, Fig.13 and Fig.14 illustrate the markovian behaviour of the simulated atmosphere. Recall that markovians are understood as time-space variations that superimpose to the time-constant atmospheric state vector (no time evolutive). From top to bottom, the first plot represents the atmospheric state vector, the second one corresponds to its estimated counterpart, the third one, evaluates inversion relative error and the fourth one is the Kalman gain. The large differences in inversion error and Kalman gain of Fig.13 and Fig.14 show that while in the first simulation the filter cannot track the dynamics of the atmosphere, it can in the second one.

Though belonging to different simulations, Fig.12 and Fig.15 compare the atmospheric state vector in a colourplot. Abcissae plot the iteration number and ordinates the vector component number. Note that in Fig.12, the markovian profiles are retrieved in the low atmosphere (bottom half of the plot) but not in the high atmosphere. Simulations 1 and 2 are about cooperative filters, where results from the first filter are passed to the second. Simulations 3 and 4 focus on the performance of the second filter. In all cases, as long as the filter performs more and more estimations, it manages to keep track of the estimates.
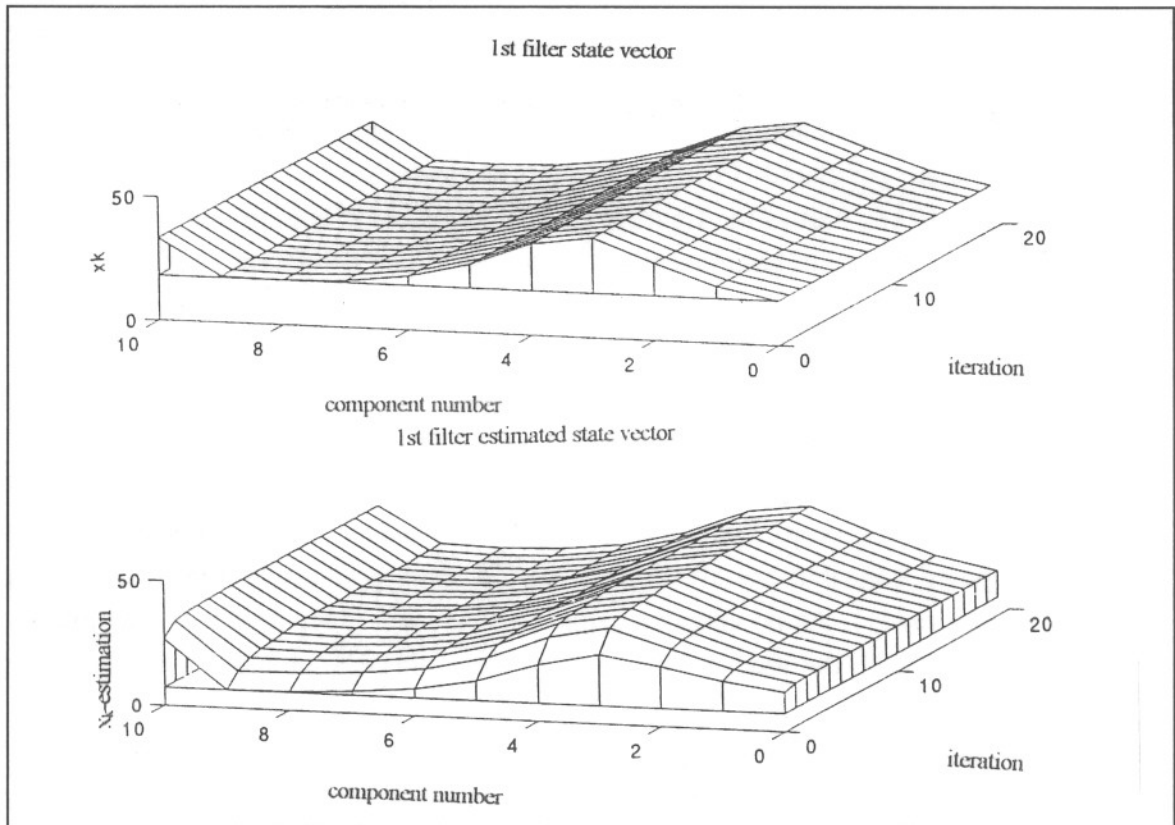
*Fig.11 Simulated atmosphere and first filter estimates (sim1).*
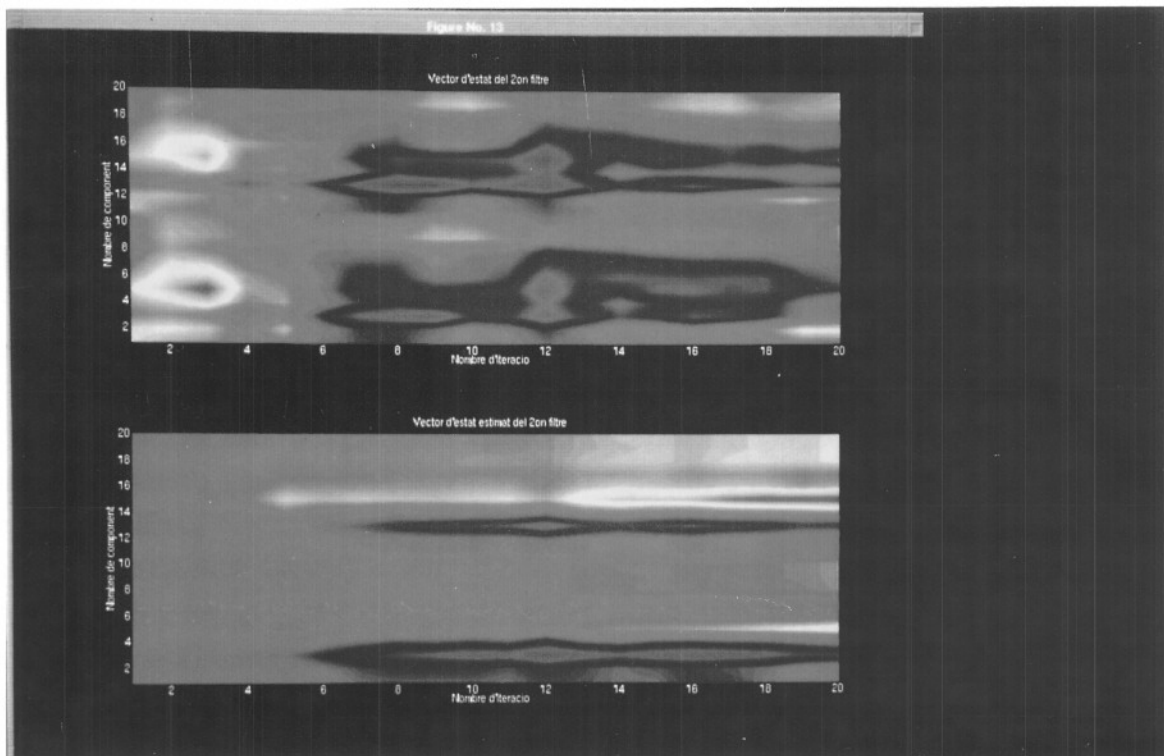


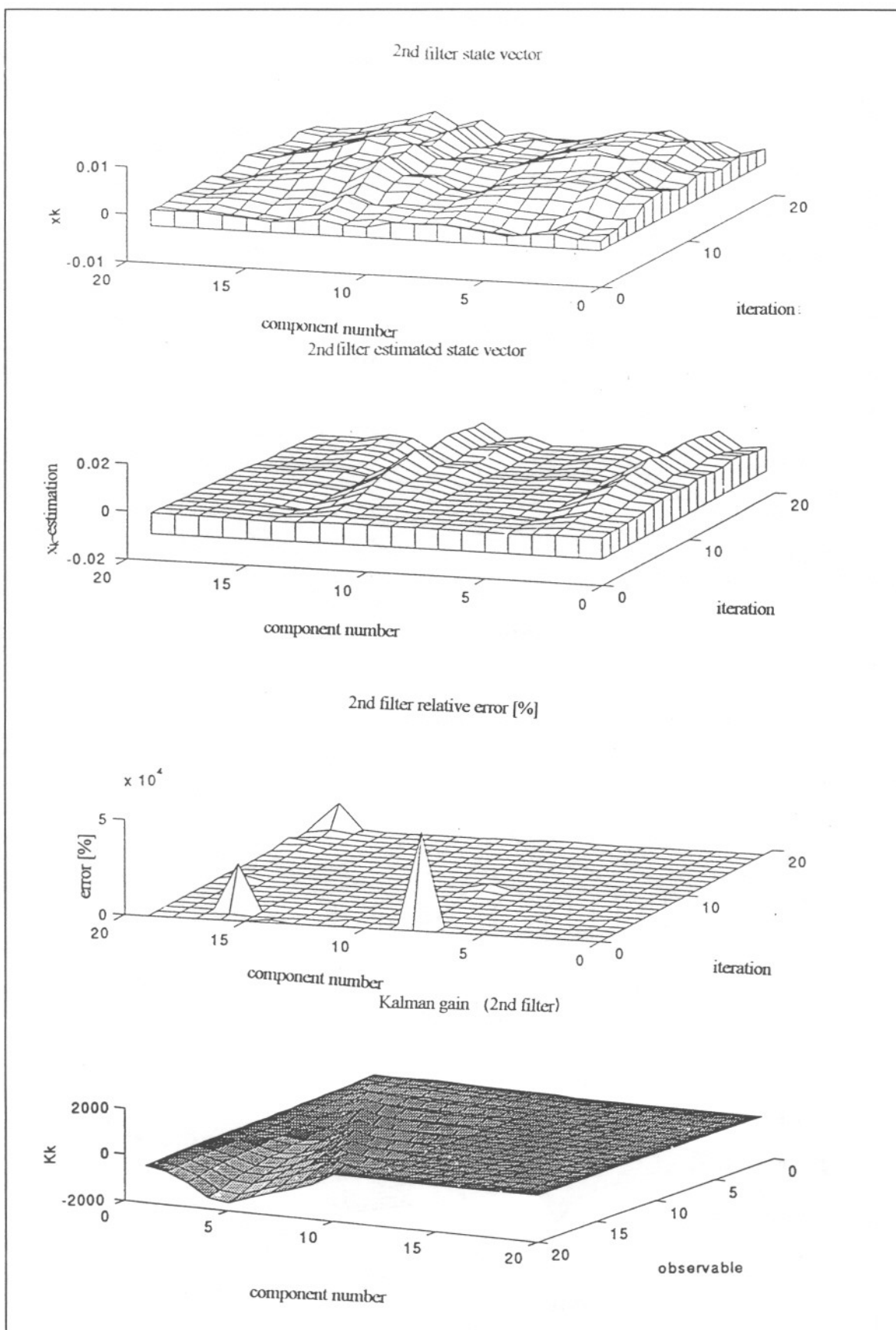*Fig.12 Time evolution of simulated atmosphere and second filter estimates (sim2).*

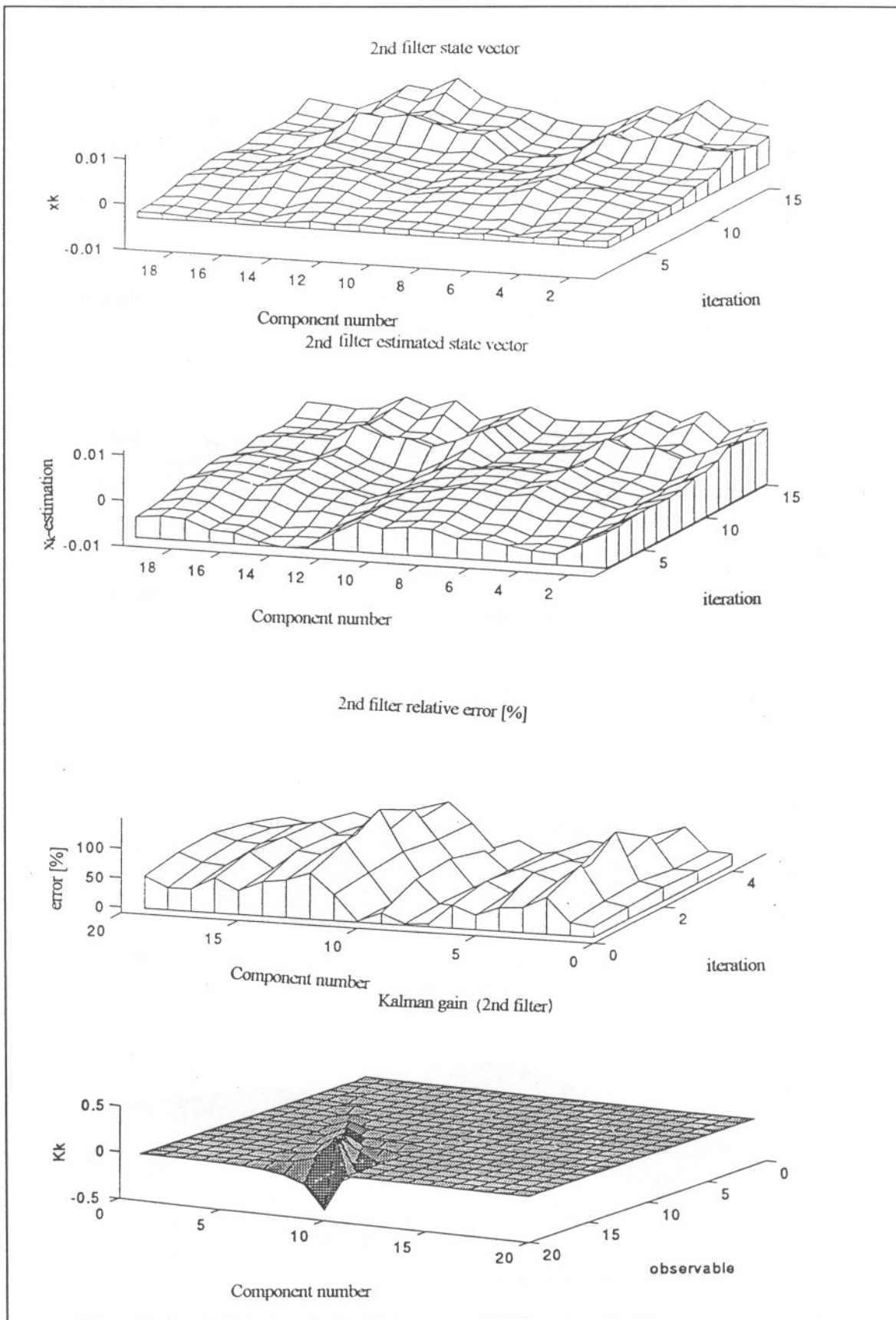Fig.13 Atmospheric dynamic evolution and second filter behaviour (sim1).

Fig.14 Atmospheric dynamic evolution and second filter behaviour (sim2).

*To conclude, as far as it concerns to the first set, it can be said that notwithstanding lower SNRs, first filter performance (the constant estimator) is always excellent, while tracking of the dynamic behaviour of the atmosphere by the second filter (the markovian estimator) is virtually impossible, unless the lidar worked with extremely high SNR that would lead to acceptable accuracies (see simulation 2).*

Low $\alpha$-*observability* together with a high in-cell correlation, $\rho'$, could possibly justify that SNR figures in *simulation 1* were not high enough to guarantee a good performance of the second filter. A high $\rho'$ could play a negative role as if $\alpha_i$ is misestimated, these errors will easily propagate to their partners, $\beta_i$, because $\rho'$ near unity virtually relates $\beta_i$ to $\alpha_i$ in a linear fashion. Nevertheless if SNR is high enough, Kalman estimates and atmospheric history look exactly alike (Fig.14).

As for the second set, if *simulation 4* is compared to *simulation 1,* it can be concluded that *first filter accuracy (the constant estimator) and SNR, as well, emerge as the dominant factors.* Thus, exact constant estimation has saved up to 30 dB of SNR. Error and contour plots will help the reader to tell out the differences.

| PARAMETERS | | FILTER 1 | | FILTER 2 | | COMMENTS |
|---|---|---|---|---|---|---|
| | | $\epsilon(C)$ | $\epsilon(\beta_c)$ | $\epsilon(\alpha_m)$ | $\epsilon(\beta_m)$ | |
| 1 | A<br>$4.71 \cdot 10^{-3}$<br>SNR<br>60:40 dB | 0.28<br>% | 4.25<br>% | 897<br>% | 14383<br>% | Filt.1 converges in 8 iterations (Fig.11)<br>Filt.2 cannot track R$\geq$0.4 km (Fig.13,Fig.12) |
| 2 | A<br>$4.71 \cdot 10^{3}$<br>SNR<br>120:100 dB | $\approx 0$<br>% | 0.5<br>% | 60<br>% | 76<br>% | Filt.1 exact<br>Filt.2 good tracking (Fig.14)<br>$I_1 = 10$; $I_2 = 15$ |
| 3 | A<br>$4.71 \cdot 10^{-3}$<br>SNR<br>60:40 dB | | | 4087<br>%<br>200 %<br>avg | 593<br>%<br>200 %<br>avg | Filt.2 given exact values of $\beta_c$, C.<br>Filt.2 rough tracking |
| 4 | A<br>4.71<br>SNR<br>90:70 dB | | | 44<br>% | 23<br>% | Filt.2 given exact values of $\beta_c$, C.<br>Extraordinary tracking<br>$I_2 = 10$ (Fig.15,Fig.16) |

*Tab.4 Summary of the simulations done using cooperative filters.*

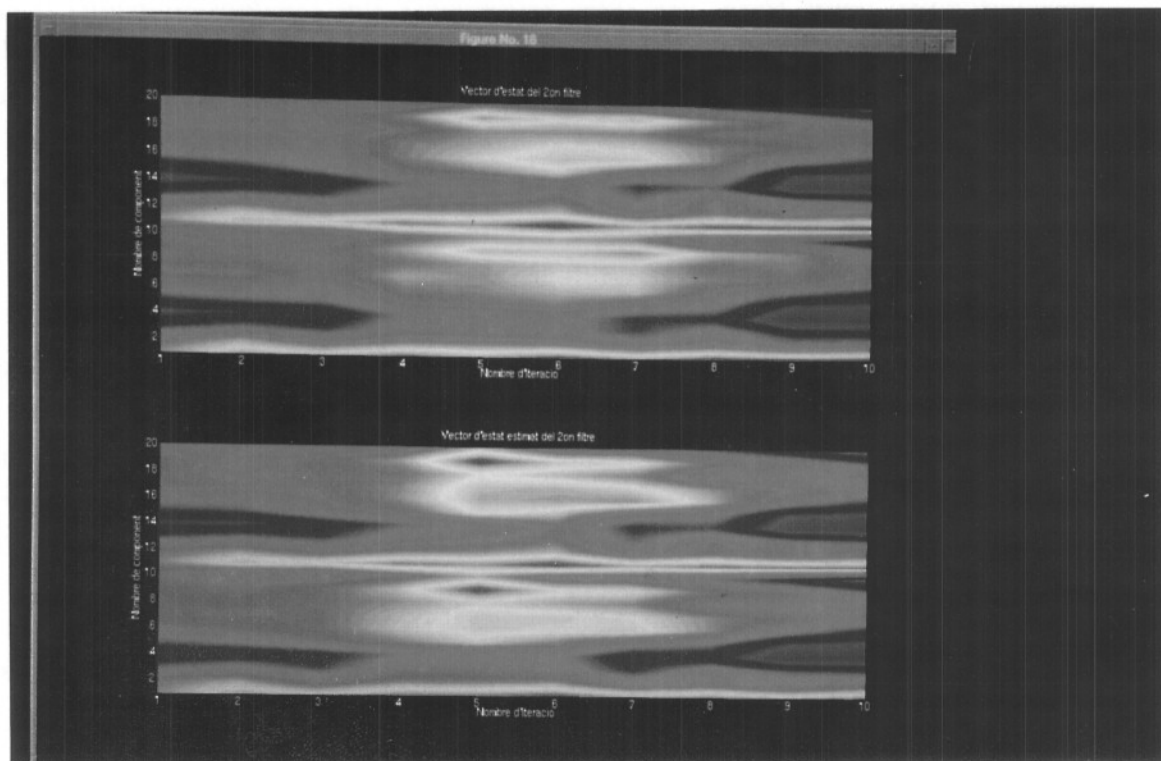*(NOTE: avg means averaged value. Peak values are shown unless otherwise indicated).*

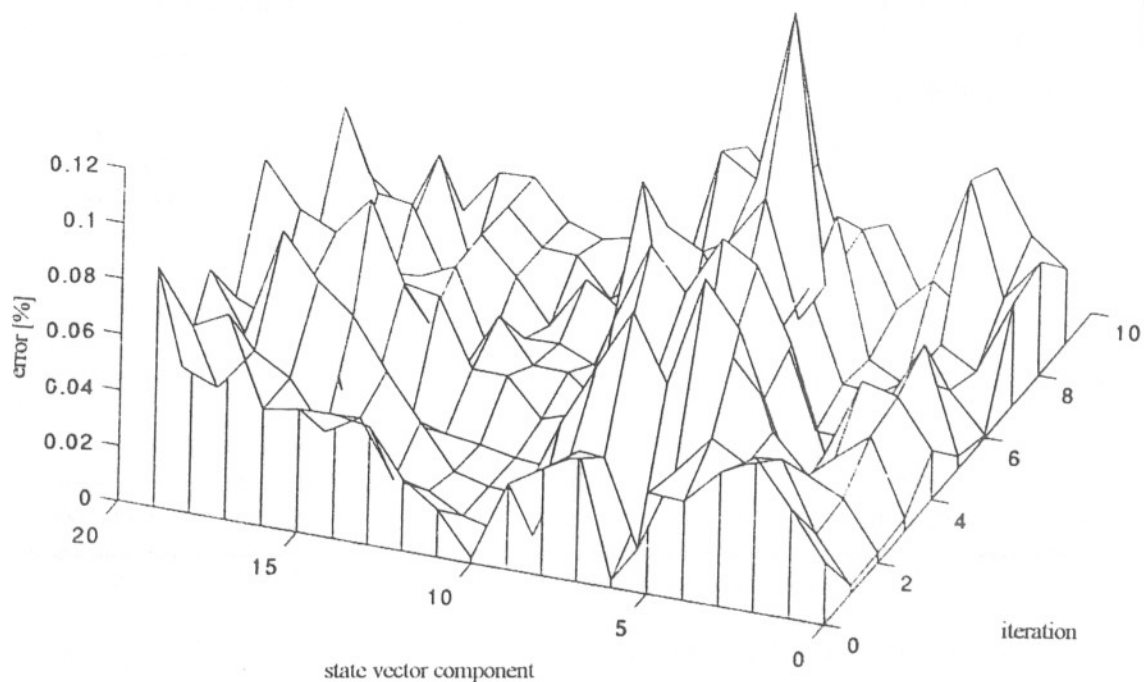*Fig.15 Atmospheric dynamic evolution and second filter estimates (sim4).*



*Fig.16 Relative error of the total estimation (sim4).*

## 5. THE OBSERVABILITY PROBLEM

Summing up the results achieved by the filters presented, one may well wonder himself which are the intrinsic hidden parameters of the model, responsible for the relative errors yielded. Form the author's viewpoint, ***Kalman filter performance is directly tied to the observability problem***. Three main factors should be mentioned:

### 1. Rank observability and undersampling factor

This accounts for situations were there are less observables (measurements) than variables to estimate. Based on these situations, the undersampling factor, $M$, is defined as the ratio of power samples to the number of cells where both optical parameters are to be estimated. Thus, $M=1$ would mean a complete coverage of the lidar range and $M=2$ would mean that only one out of every two cells is estimated.

The rank observability is a concept inherited from linear algebra. Thus, *it is necessary for a set of linear equations to be compatible with single solution to have the same number of equations as variables*. Though this is not always true for a Kalman filter as $H_k$ does not need to be square and each observation may convey some extra information, *it seems to apply in the lidar case*. Reference [169] provides formal tests of observability that may be applied to systems of low dimensionality. These test are not always practical to apply, though, in higher-order systems as it happens to the lidar case.

To illustrate this new kind of observability problem, it is worth showing some results obtained with a Kalman filter identical to that of Sect.3 except for $M=1$ (the overdetermined cases, $M \geq 3$ have proved fruitful and the $M=2$ case has been covered in Sect.3). The $M=1$ case intends to estimate both optical parameters along each observation cell, which results in $2N$ unknowns, given $N$ power samples. Matrices *H1* and *H2* are shown in eqs.(111) and (112)

$$H_1 = \begin{bmatrix} -2\Delta R\,P_1 & 0 & 0 & 0 & ..... & 0 \\ -2\Delta R\,P_2 & -2\Delta R\,P_2 & 0 & 0 & ..... & 0 \\ -2\Delta R\,P_3 & -2\Delta R\,P_3 & -2\Delta R\,P_3 & 0 & ..... & 0 \\ ..... & ..... & ..... & ..... & ..... & ..... \\ -2\Delta R\,P_N & -2\Delta R\,P_N & -2\Delta R\,P_N & -2\Delta R\,P_N & ..... & -2\Delta R\,P_N \end{bmatrix} \quad (111)$$

Two simulation sets have been performed beginning from the atmospheric state vector of Fig.17 and considering white noise perturbations in successive time steps *(the white noise atmosphere model of Sect.3)*. Thus, the first one consists of Fig.18 while the second one Fig.19 (see the simulation parameters in Tab.4).

In Fig.18, neither the extinction nor the backscatter have correctly been estimated by the filter. Letting alone complex mathematical reasonings, it can be explained noticing that the filter has to tackle the estimation of two parameters in each observation cell.

$$H_2 = \begin{bmatrix} \dfrac{P_1}{x_{N+1}} & 0 & 0 & ..... & 0 \\[2ex] 0 & \dfrac{P_2}{x_{N+2}} & 0 & ..... & 0 \\[2ex] 0 & 0 & \dfrac{P_3}{x_{N+3}} & ..... & 0 \\[2ex] ... & ... & ... & ..... & ... \\[2ex] 0 & 0 & 0 & ..... & \dfrac{P_N}{x_{2N}} \end{bmatrix} \qquad (112)$$

Looking back to eq.(48), let that cell be located at $R=Rmin$. It emerges that the filter is confronted by an indetermination as any misestimation in one parameter (say $\alpha$) can be offset by the other one (say $\beta$). Fig.20 shows the time-space history of the filter, which is very sluggish in the estimation for $\alpha$. This has led the author to devise an undersampling filter.

Lastly, in second simulation (Fig.19), contrary to what may be thought at first sight, the filter has been able to successfully estimate one of the two parameters, the backscatter. This surprising result can be explained considering the second type of observability.

| PARA-METER | 1st. SIMULATION Fig.18 | 2nd. SIMULATION Fig.19 |
|:---:|:---:|:---:|
| $\alpha$ | 0.2-1 | $10^{-2}$ Nep/km$^{-1}$ |
| $\beta$ | 0.2-1 | $10^{-2}$ km$^{-1}$ |
| $\hat{x}_0^-$ | $10^{-1}$ | $10^{-2}$ |
| $\sigma_p$ | $10^{-1}$ | $10^{-2}$ |
| $\sigma_q$ | $10^{-2}$ | $10^{-3}$ |
| $\sigma_r$ | $10^{-2}$ | $10^{-2}$ W |
| A | $10^4$ | $10^4$ W·km$^3$·sr |
| N | 60 | 60 |
| SNR(R) | 140-40 | 120-50 dB |

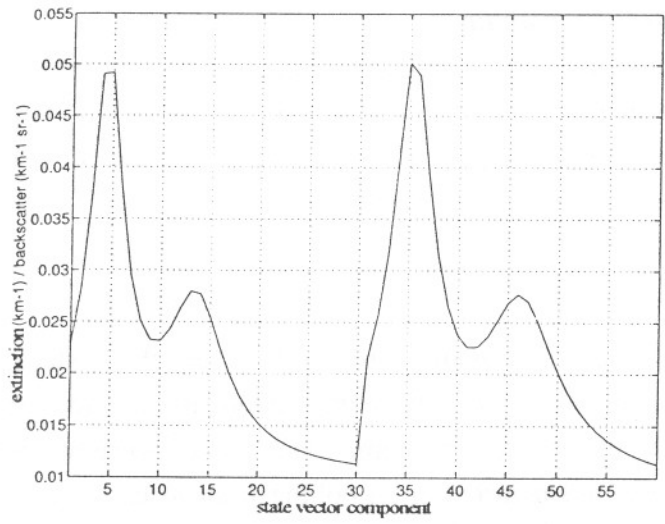*Tab.5 Parameters used in the observability simulations.*

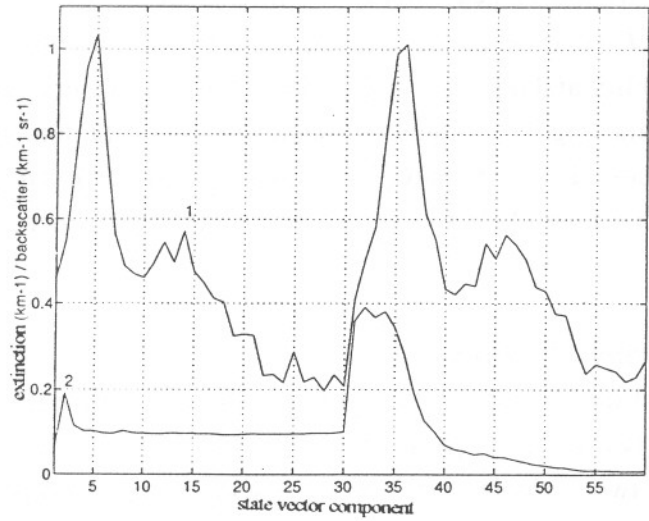*Fig.17 Initial atmospheric vector state.*



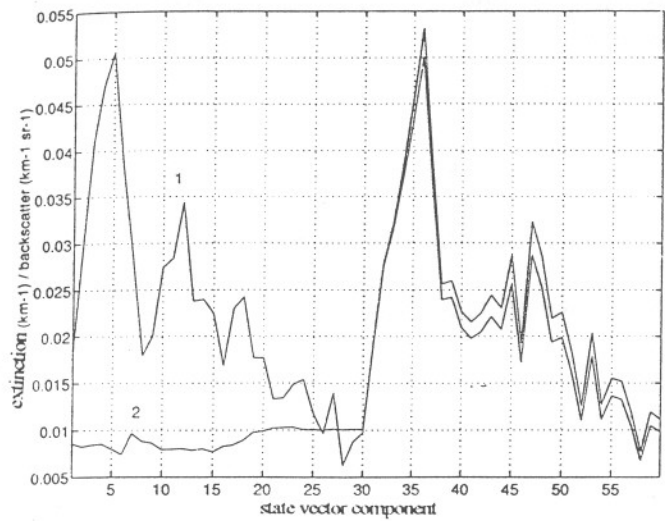*Fig.18 Final atmospheric state vector (1) and estimated one (2).*



*Fig.19 Final atmospheric state vector (1) and estimated one (2).*

## 2. Parameter observability

This second type of observability emerges as a consequence of an apparent change in the *ratio, k, number of observables to parameters,* due to a sensitivity change. Thus, referring back to eq.(48), if the second simulation is considered, $\alpha$ has little effect on the observable power as for $\alpha \ll 1$ it becomes:

$$P_i \approx \frac{A}{(i\Delta R)^2}\beta(i\Delta R)$$

Therefore, the *k-ratio* changes from 2 to 1 and the system becomes completely observable (the vector $\beta$ would act as the unknowns and the power samples as the parameters). On top of that, the most important consequence stems from the sensitivity effects on the observation matrices $H_k$. Looking back to eq.(111), $H_1$ (it is, in fact, the derivative of the observables with respect to the *$\alpha$-parameters,* therefore, a sensitivity relationship) remains unchanged. None of the *$\alpha$-project-ahead* steps will be sensitive enough to modify $H_1$ via de $P_i$ terms. As a consequence, if these trajectories are static, the filter lacks adaptability and does not update this set of parameters. This is shown by a flat portion in Fig.19 and Fig.21, which correspond to the extinction-subvector.

Besides, if eq.(111) is compared back to eq.(57) *the undersampling factor (N/M=2) shows up in an extra gain $2N/M\Delta R = 4\Delta R$ rather than $2\Delta R$,* increasing $H_1$ sensitivity to return-power, $P_i$.

## 3. Identification observability

This kind of observability problem often arises as a modelling error. The filter is taught to estimate two different parameters when, in fact, they are not. Imagine it were wished to estimate two variables *x,y* from an observable *z,* accomplishing *z=x+y.* Here, *x* and *y* could not be sorted out and the filter would only be able to identify its joint effects, say $\hat{z}=x+y$. Imagine a joint estimation of $\beta_c$, $\alpha_m$, $\beta_m$ were tried by a single Kalman filter. The state vector would be

$$\vec{x} \equiv \begin{bmatrix} \vec{\beta}_c \\ \vec{\alpha}_m \\ \vec{\beta}_m \end{bmatrix} \equiv \begin{bmatrix} x_i \\ x_{N/M+i} \\ x_{2N/M+i} \end{bmatrix} \qquad i=1..\frac{N}{M} \qquad\qquad (114)$$

and the model equations would take the form

$$\begin{bmatrix} \vec{\beta}_c(k+1) \\ \vec{\alpha}_m(k+1) \\ \vec{\beta}_m(k+1) \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & \exp(-B\Delta t)\cdot I & 0 \\ 0 & 0 & \exp(-B\Delta t)\cdot I \end{bmatrix} \begin{bmatrix} \vec{\beta}_c(k) \\ \vec{\alpha}_m(k) \\ \vec{\beta}_m(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \vec{w}_\alpha(k) \\ \vec{w}_\beta(k) \end{bmatrix} \qquad (115)$$

In terms of the state variables, this would yield to eq.(70)

$$\alpha_i = C \cdot x_i + x_{N/M+i} \quad i=1..\frac{N}{M}$$

$$\beta_i = x_i + x_{2N/M+i} \quad i=1..\frac{N}{M} \qquad (116)$$

that is not observable (e.g. assume $x=x_i$, $y=x_{2N/M+i}$ and $z$ equal to the backscatter $\beta_i$).

Evidently, the choice of state variables is largely a matter of convenience.

*All in all, the Kalman filter, linear or nonlinear, is usually the preferred algorithm under any reasonable error criterion, provided there is some knowledge of the stochastic processes to deal with. The main difficulties to overcome lie in solving the start-up problem and choosing good observable models.*
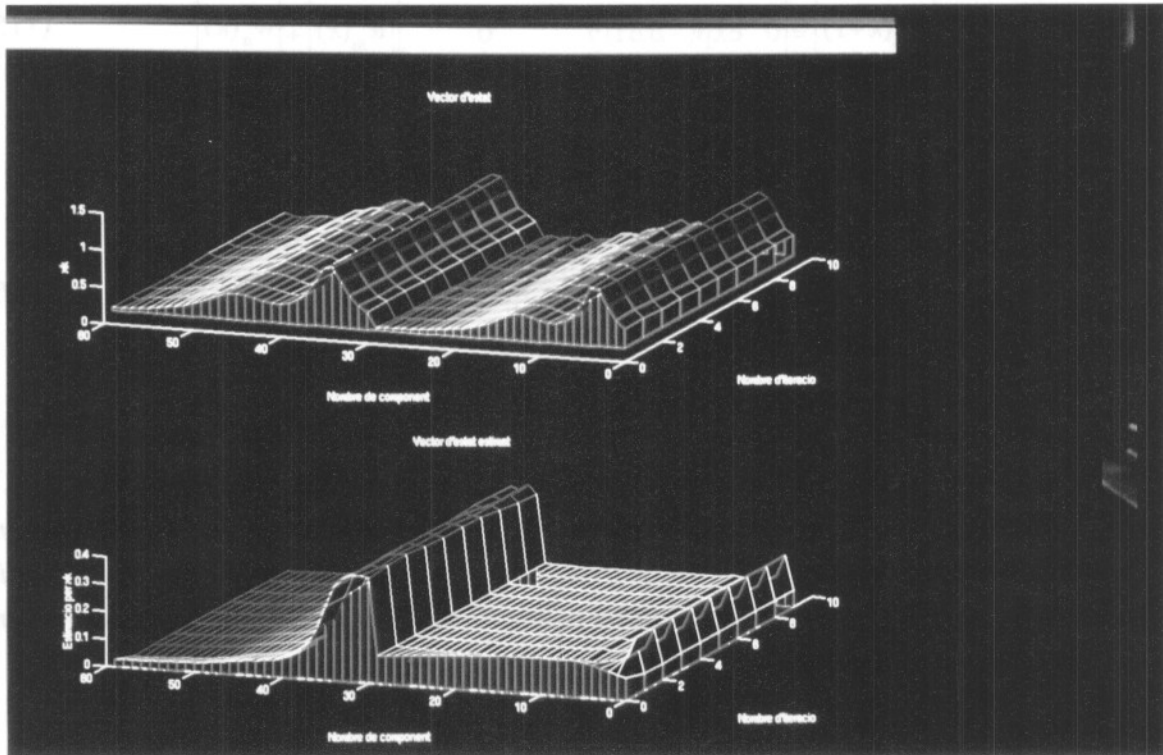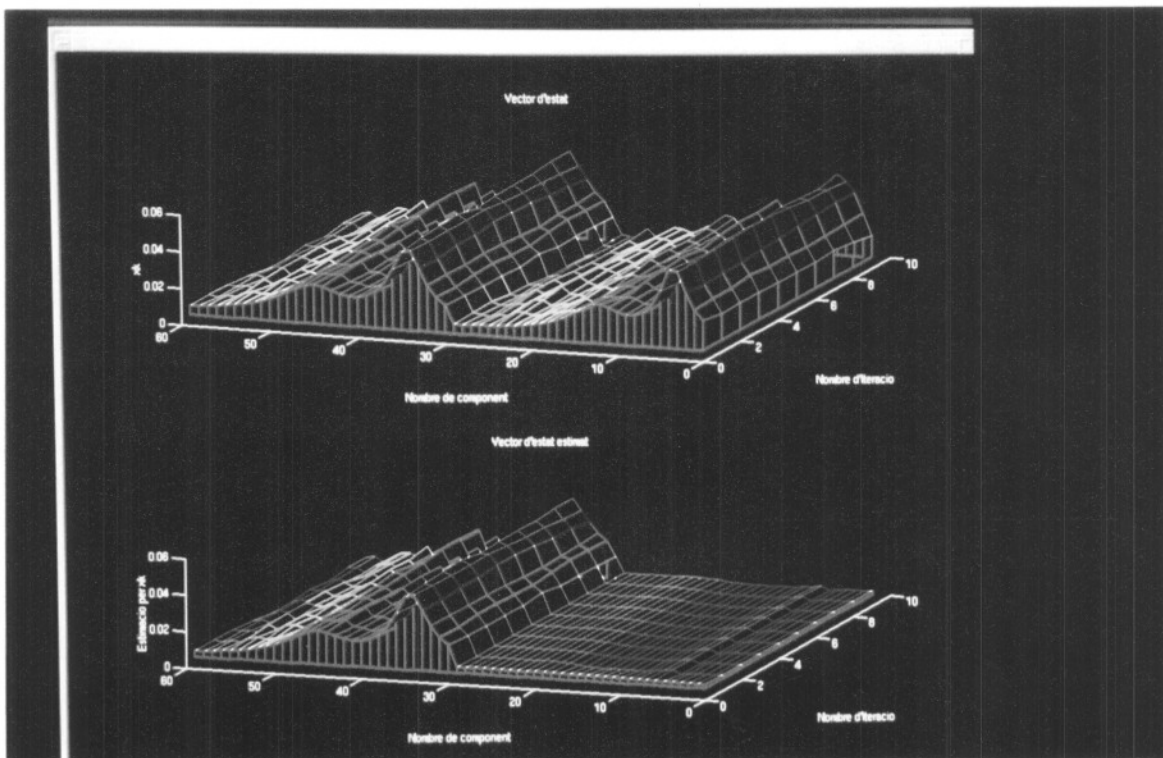
Fig.20 Rank deficiency leads to very sluggish behaviour.



Fig.21 Extinction is a very low observability parameter.