Fundamental aspects of the Kalman filter

Report WD 88-03

Fundamental Aspects of the Kalman Filter

J. Molenaar
H. Visser

mei 1988

Wiskundige Dienstverlening
Faculteit W&N
Toernooiveld
6525 ED Nijmegen

080-613138

FUNDAMENTAL ASPECTS OF THE KALMAN FILTER

J.Molenaar(*)

and

H.Visser(**)

(*) Mathematics Consulting Department, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

(**) Environmental Department, Joint laboratories and other services of the Dutch electricity supply companies, P.O.Box 9035, 6800 ET, Arnhem, The Netherlands.

Contents

1 Introduction.

In experiments, it is rather the rule than the exception that instead of the parameters of interest only some related parameters can be measured. This immediately leads to the need of reliable regression methods. For example, this is the situation in the study of the influence of acid rain on tree growth. Here, tree ring widths are measured in order to obtain information about the variations in tree behaviour with the weather conditions. This might eventually yield insight in the way acid rain affects tree growth. This phenomenon is poorly understood yet. In the regression of tree ring series on the weather, the application of the so-called Kalman approach appears to be very useful. Several results have been presented in references [Visser and Molenaar 1988, Visser 1986]. This report is meant to give an overview of the theoretical backgrounds of this method. The main purpose is to clarify the ideas, but the mathematical details will be included in order to present a self-contained derivation of the theory. Only elementary knowledge of statistics and linear algebra is presupposed.

This report is organised as follows. In sections 2 and 3 we introduce the state space model to be studied. In section 4 we deal with the important concept of 'innovations' and the 'orthogonality principle'. For the theory, it is not necessary to know the distribution of the stochastic terms in the model in advance. The general version of the theory is derived in section 5. If the disturbances are Gaussian, the derivations of the formulae can be considerably simplified. This is the subject of section 6. One of the main aspects of the theory is the recursive form of the calculation procedure. It is interesting to note that in this respect a great analogy exists between the Kalman procedures and the ordinary least squares method in recursive form. This relation is the subject of section 7. In the Gaussian case, unknown parameters may be estimated by maximum likelihood estimation, which is dealt with in section 8. There, the maximum likelihood function is expressed in terms of innovations, which leads to remarkably simple formulae. In section 9, the general form of linear regression models, which are also called structural models, is presented. There, we also point out how the general methods, developed in this report, are used in the analysis of tree-ring series.

In this report we pay no attention to the estimation of non-linear models, which need application of the so-called Extended Kalman filter. For an example of this extension we refer to [Molenaar and Visser 1987].

2 State Space Model.

In this report, we study the following state space model:

$$y_t = \mathbf{m}'_t \mathbf{x}_t + v_t, \ t = 1, ..., T \tag{1}$$

The state vector $\mathbf{x}_t$ of dimension N is not directly observable and is to be estimated from T successive observations of y_t . We call (1) the measurement equation. The vector $\mathbf{m}_t$ of length N is, in the first instance, assumed to be known. The disturbance v_t represents the measurement error. The vectors $\mathbf{x}_t$ are assumed to be related by the linear transition equation

$$\mathbf{x}_t = \mathbf{T}_t \mathbf{x}_{t-1} + \mathbf{w}_t \tag{2}.$$

$\mathbf{T}_t$ is a known $N \times N$ matrix and the vector $\mathbf{w}_t$ of length N represents some internal stochastic process in the system. For the disturbances v_t and $\mathbf{w}_t$ we assume

$$\begin{aligned}
cov(v_{t_1}, v_{t_2}) &= R_{t_1} \delta_{t_1 t_2} && \forall t_1 t_2 \\
cov(\mathbf{w}_{t_1}, \mathbf{w}_{t_2}) &= \mathbf{Q}_{t_1} \delta_{t_1 t_2} && \forall t_1 t_2 \\
cov(\mathbf{w}_{t_1}, v_{t_2}) &= 0 && \forall t_1 t_2 \\
cov(\mathbf{w}_t, \mathbf{x}_0) &= 0 && \forall t \\
cov(v_t, \mathbf{x}_0) &= 0 && \forall t
\end{aligned} \tag{3}$$

with δ the Kronecker symbol and $\mathbf{x}_0$ the initial value of the state.

For the model contained in expressions (1)-(3) the following remarks are important:

a). The parameters $\mathbf{m}_t, \mathbf{T}_t, R_t$ and $\mathbf{Q}_t$ are assumed to be known in advance. In most cases of interest this is a real problem and some of them have also to be estimated from the data. This point is further dealt with in section 8.

b). If v_t, $\mathbf{w}_t$, and the initial distribution $\mathbf{x}_0$ are normally distributed Gaussian processes, then it is clear that $\mathbf{x}_t$ and y_t are also jointly Gaussian for all t.

c). In most applications one takes R and $\mathbf{Q}$ constant. Moreover, $\mathbf{Q}$ is often assumed to be a diagonal matrix.

d). In the univariate case studied here (i.e. y_t is a scalar), there is no need to deal with R_t (a scalar) and $\mathbf{Q}_t$ separately, because all formulae in the Kalman theory depend on the quotient $\mathbf{Q}_t / R_t$ only.

3 State Estimation.

For convenience, we introduce the notation

$$\mathbf{Y}_t \equiv (1, y_1, ..., y_t)' \tag{4}$$

We want to find an estimator for the state $\mathbf{x}_t$ in terms of the components of the vector $\mathbf{Y}_{t'}$.We shall denote this estimator by $\mathbf{x}_{t/t'}$. The corresponding estimate $\bar{\mathbf{x}}_{t/t'}$ is then directly obtained by replacing $\mathbf{Y}'_t$ by the corresponding vector containing measured values in the expressions for the estimator. In linear regression theories, of which the Kalman approach is an example, the estimator is assumed to be linear in $\mathbf{Y}_{t'}$, i.e. an affine transformation of the first t observations, and thus of the form

$$\mathbf{x}_{t/t'} = \mathbf{A}_t \mathbf{Y}_{t'} \tag{5}$$

The matrix $\mathbf{A}_t$ is to be determined. We have still to specify in which sense the estimator $\mathbf{x}_{t/t'}$ should be optimal. It is appropriate to demand that $\mathbf{x}_{t/t'}$ should minimise the variance $var(\mathbf{x}_t - \mathbf{x}_{t/t'})$ and we shall call it the minimum variance estimator (MVE). Because the rows of $\mathbf{A}_t$ are to be determined independently, minimisation of this variance implies the minimisation of the variance of each component of $\mathbf{e}_{t/t'} \equiv \mathbf{x}_t - \mathbf{x}_{t/t'}$ separately.

A well-known result of estimation theory states that the minimum variance estimate is given by the conditional mean of $\mathbf{x}_t$ given $\mathbf{Y}_{t'}$, which is denoted by $E(\mathbf{x}_t|\mathbf{Y}_{t'})$ [Bagchi 1982]. However, because $\bar{\mathbf{x}}_{t/t'}$ is bound to be of the form (5), this result does not hold in general for $\bar{\mathbf{x}}_{t/t'}$. It is only the case if $\mathbf{x}_t$ and $\mathbf{Y}_t$ are jointly Gaussian, because then a famous theorem states that the MVE of $\mathbf{x}_t$ is always of the form (5) [Sage and Melsa 1971]. In that case, $\bar{\mathbf{x}}_{t/t'}$ is not only the best linear estimate of $\mathbf{x}_t$ (in the sense of minimum variance) but even the best of all possible estimates and we may write

$$\bar{\mathbf{x}}_{t/t'} = E(\mathbf{x}_t|\mathbf{Y}_{t'}) \tag{6}$$

4 Innovations and the Orthogonality Principle.

Equation (5) states that each component of $\mathbf{x}_{t/t'}$ is a linear combination of the components of $\mathbf{Y}_{t'}$. This suggests to interprete these components as elements of a linear vector space. This space contains all stochastic scalar processes with finite variances and all constants. A natural inner product for elements y_t and $y_{t'}$ in this space is given by

$$(y_t, y_{t'}) \equiv E(y_t y_{t'}) \tag{7a}$$

which induces the norm

$$\|y_t\| = \sqrt{(y_t, y_t)} \tag{7b}$$

This allows for a nice geometric interpretation of the MVE. In the preceding section we concluded that each component $x_{t/t'}$ of $\mathbf{x}_{t/t'}$ is chosen such that $var(x_t - x_{t/t'}) \equiv var(e_{t/t'})$ is minimised with x_t the corresponding component of $\mathbf{x}_t$. Because $E(e_{t/t'}) = 0$, it holds that $var(e_{t/t'}) = \|e_t\|^2$. So, we conclude that $x_{t/t'}$ is that element of the subspace spanned by the components of $\mathbf{Y}_{t'}$, which has minimal distance to $\mathbf{x}_t$. In other words, it is the projection of $\mathbf{x}_t$ on that subspace. Denoting this projection operator by $Pr_{t'}$ we may write for all components of $\mathbf{x}_{t/t'}$ at the same time :

$$\mathbf{x}_{t/t'} = Pr_{t'}\mathbf{x}_t \tag{8a}$$

The property that all components of $\mathbf{x}_t - \mathbf{x}_{t/t'}$ are orthogonal to the subspace spanned by the components of $\mathbf{Y}_{t'}$ is called the *Orthogonality Principle*, which can also be expressed in the form :

$$E((\mathbf{x}_t - \mathbf{x}_{t/t'})\mathbf{Y}'_{t'}) = 0 \tag{8b}$$

In the following, we shall often use this kind of expression. For conciseness, we introduce for stochastic vectors $\mathbf{x}_1, \mathbf{x}_2$, with components in the linear vectorspace introduced above, the notation

$$(\mathbf{x}_1, \mathbf{x}_2) \equiv E(\mathbf{x}_1 \mathbf{x}'_2) \tag{8c}$$

If $\mathbf{x}_1$ and $\mathbf{x}_2$ are scalar processes, this notation agrees with definition (7a). In analogy with the Orthogonality Principle, we call $\mathbf{x}_1$ and $\mathbf{x}_2$ orthogonal if $(\mathbf{x}_1, \mathbf{x}_2) = 0$. Note that, with definition (8c), the matrix $(\mathbf{x}_1, \mathbf{x}_2)$ is nothing else but the covariance matrix of $\mathbf{x}_1$ and $\mathbf{x}_2$, provided that $E(\mathbf{x}_1) = E(\mathbf{x}_2) = 0$. The

matrix $(\mathbf{x}_1, \mathbf{x}_2)$ can also be read as a tensor product. The inner or scalar product, commonly used in connection with such a tensor product, is the sum of the squares of all matrix elements. The Orthogonality Principle expresses the orthogonality of two vectors with respect to this inner product.

For the general derivation of the Kalman formulae it is very useful to construct a orthonormal basis in the following way:

$$\begin{aligned} \nu_0 &= 1 \\ \nu_1 &= y_1 - Pr_0 y_1 \\ \nu_2 &= y_2 - Pr_1 y_2 \end{aligned} \tag{9}$$

The orthogonal basis elements $\nu_i, i = 0, 1, ..., t$ are thus obtained from the elements of $\mathbf{Y}_t$ by a Gram-Schmidt procedure. If we define a vector $\mathbf{N}_t$ by

$$\mathbf{N}_t = (\nu_0, \nu_1, ..., \nu_t)', \tag{10}$$

this procedure is equivalent with the transformation

$$\mathbf{N}_t = \mathbf{L}_t \mathbf{Y}_t \tag{11}$$

with $\mathbf{L}_t$ a lower triangular matrix with ones at the diagonal. This unique orthogonalisation procedure is also called Choleski decomposition [Harvey 1981b]. We call the ν_i *innovators* and the corresponding realisations are known as *innovations*.The innovators can be expressed in terms of the estimators $\mathbf{x}_{t/t-1}$:

$$\begin{aligned} \nu_t &= y_t - Pr_{t-1}(\mathbf{m}'_t \mathbf{x}_t + v_t) \\ &= y_t - \mathbf{m}'_t \mathbf{x}_{t/t-1} \end{aligned} \tag{12}$$

because v_t is, by definition, orthogonal to the subspace spanned by the elements of $\mathbf{Y}_{t-1}$. From expression (12) it is clear why the innovations are also called the *one-step-ahead prediction errors*. To normalise the ν_t, we rewrite expression (12) in the form

$$\nu_t = \mathbf{m}'_t(\mathbf{x}_t - \mathbf{x}_{t/t-1}) + v_t \tag{13}$$

Introducing a matrix $\mathbf{P}_{t/t'}$ by

$$\mathbf{P}_{t/t'} = (\mathbf{x}_t - \mathbf{x}_{t/t'}, \mathbf{x}_t - \mathbf{x}_{t/t'}), \tag{14}$$

we find that

$$(\nu_t, \nu_t) = \mathbf{m}'_t \mathbf{P}_{t/t-1} \mathbf{m}_t + R_t \equiv f_t \tag{15}$$

because, in view of assumptions (3), v_t is orthogonal to both $\mathbf{x}_t$ and $\mathbf{x}_{t/t-1}$. In the univariate case studied here, f_t is a positive scalar and we may normalise the ν_t by

$$\pi_t = \nu_t (f_t)^{-1/2} \tag{16}$$

For $t = 0$ we cannot use (15) but the normalisation is trivial in that case.

For later purposes, we note that, from the construction of the π_t and assumptions (3),

$$(v_t, \pi_{t'}) = 0 \qquad \text{if} \quad t > t' \tag{17a}$$

and

$$(\mathbf{w}_t, \pi_{t'}) = 0 \qquad \text{if} \quad t \geq t' \tag{17b}$$

Because $\pi_t, t = 1, 2, \ldots$ are orthogonal to π_0, which is constant, we have

$$(\pi_t, \pi_0) = E(\pi_t) = 0 \qquad \forall t > 0 \tag{17c}$$

5 Kalman Theory: The General Case.

In this section, we present a general derivation of the Kalman theory. In the literature, one can find many approaches to this estimation problem. For example, see [Kalman 1960, Kalman and Bucy 1961, Sage and Melsa 1971, Jazwinski 1970, Kwakernaak and Sivan 1972, Otter 1984]. We prefer to present a self-contained derivation based on the notions introduced in sections 3 and 4. There, we established that the components of the estimator $\mathbf{x}_{t/t'}$ are elements of the linear vector space spanned by the orthonormal basis vectors $\pi_0, \pi_1, ..., \pi_{t'}$. So we may expand

$$\mathbf{x}_{t/t'} = \sum_{i=0}^{t'} (\mathbf{x}_t, \pi_i)\pi_i \tag{18}$$

Using this representation we shall successively deal with the cases of prediction ($t > t'$), filtering ($t = t'$) and smoothing ($t < t'$).

It is important to realise that the expectation value $E(\mathbf{x}_{t/t'})$ is completely contained in the π_0 component of expression (18), because $(\mathbf{x}_t, \pi_0)\pi_0 = E(\mathbf{x}_t)$. This directly implies that the estimator $\mathbf{x}_{t/t'}$ is unbiased. Therefore, the variances of the components of $\mathbf{x}_t - \mathbf{x}_{t/t'}$ are given by the diagonal elements of the matrix $\mathbf{P}_{t/t'}$ and these are just the quantities we are minimising.

5.1 Prediction.

It is easy to express $\mathbf{x}_{t/t'}$ with $t > t'$ in terms of $\mathbf{x}_{t'/t'}$. To that end, we substitute the transition equation (2) into the coefficients of expansion (18). Because of property (17b), we have for $t > i$:

$$\begin{aligned} (\mathbf{x}_t, \pi_i) &= \mathbf{T}_t(\mathbf{x}_{t-1}, \pi_i) + (\mathbf{w}_t, \pi_i) \\ &= \mathbf{T}_t(\mathbf{x}_{t-1}, \pi_i) \end{aligned} \tag{19}$$

If we repeat this procedure we obtain the result

$$\mathbf{x}_{t/t'} = \mathbf{T}_t\mathbf{T}_{t-1}...\mathbf{T}_{t'+1}\mathbf{x}_{t'/t'} \tag{20a}$$

For the special case of one-step-ahead prediction we have

$$\mathbf{x}_{t+1/t} = \mathbf{T}_{t+1}\mathbf{x}_{t/t} \tag{20b}$$

Using equations (20) in definition (14) of $\mathbf{P}_{t/t'}$ we obtain a similar recurrence relation for the error variance matrix. In practice, only the one-step-ahead version

$$\mathbf{P}_{t+1/t} = \mathbf{T}_{t+1}\mathbf{P}_{t/t}\mathbf{T}'_{t+1} + \mathbf{Q}_{t+1} \tag{21}$$

is used.

5.2 Filtering.

In the case $t = t'$, it is appropriate to separate the term with $i = t' = t$ from the summation in (18):

$$\mathbf{x}_{t/t} = \mathbf{x}_{t/t-1} + (\mathbf{x}_t, \pi_t)\pi_t \tag{22}$$

To evaluate the inner product $(\mathbf{x}_t, \pi_t)$ we have to substitute expressions (16) and (13). From (3) we have

$$(\mathbf{x}_t, v_t) = 0 \tag{23}$$

From the geometrical interpretation of $\mathbf{x}_{t/t-1}$ as the projection of $\mathbf{x}_t$ on a subspace, which does not contain $\mathbf{x}_t$, it immediately follows that

$$(\mathbf{x}_t, \mathbf{x}_t - \mathbf{x}_{t/t-1}) = (\mathbf{x}_t - \mathbf{x}_{t/t-1}, \mathbf{x}_t - \mathbf{x}_{t/t-1}) = \mathbf{P}_{t/t-1} \tag{24}$$

by definition (14). So we arrive at

$$(\mathbf{x}_t, \pi_t)\pi_t = \mathbf{P}_{t/t-1}\mathbf{m}_t(f_t)^{-1}\nu_t \tag{25}$$

If we combine (22) and (25), we find that the filtered estimator $\mathbf{x}_{t/t}$ is given by the prediction estimator $\mathbf{x}_{t/t-1}$ and a correction proportional to the innovator ν_t:

$$\mathbf{x}_{t/t} = \mathbf{x}_{t/t-1} + \mathbf{K}_t\nu_t \tag{26}$$

with the so-called gain $\mathbf{K}_t$ defined by

$$\mathbf{K}_t \equiv \mathbf{P}_{t/t-1}\mathbf{m}_t(f_t)^{-1} \tag{27}$$

It remains to express $\mathbf{P}_{t/t}$ in terms of $\mathbf{P}_{t/t-1}$. This is most simply performed using relations (13) and (26)

$$(\mathbf{x}_t - \mathbf{x}_{t/t}) = (1 - \mathbf{K}_t\mathbf{m}'_t)(\mathbf{x}_t - \mathbf{x}_{t/t-1}) - \mathbf{K}_t v_t. \tag{28}$$

As already noted in the derivation of expression (15), both terms in the right hand side of (28) are orthogonal, so we may write:

$$\mathbf{P}_{t/t} = (1 - \mathbf{K}_t\mathbf{m}'_t)\mathbf{P}_{t/t-1}(1 - \mathbf{m}_t\mathbf{K}'_t) + \mathbf{K}_t R_t \mathbf{K}'_t \tag{29}$$

Using definition (27) for $\mathbf{K}_t$, this equation may be reduced to

$$\mathbf{P}_{t/t} = \mathbf{P}_{t/t-1} - \mathbf{P}_{t/t-1}\mathbf{m}_t(f_t)^{-1}\mathbf{m}'_t\mathbf{P}_{t/t-1} \tag{30}$$

$$= (1 - \mathbf{K}_t\mathbf{m}'_t)\mathbf{P}_{t/t-1} \tag{31}$$

5.3 Smoothing.

Here, we are mainly interested in the fixed interval, smoothed estimator $\mathbf{x}_{t/M}$ with $t \leq M$ and M fixed. However, the derivations to be used can easily be extended to cover other cases such as the fixed lag smoothing estimator. We derive expressions for the smoothed estimator in terms of the filtering estimator dealt with above.

From expansion (18) we deduce that

$$\mathbf{x}_{t/M} = \mathbf{x}_{t/t} + \sum_{i=t+1}^{M} (\mathbf{x}_t, \pi_i)\pi_i \tag{32}$$

in which $\mathbf{x}_{t/t}$ and the π_i are already known. It remains to study the inner products $(\mathbf{x}_t, \pi_i)$ with $t < i$. In view of equation (13) we may write for $t < t'$:

$$\begin{aligned}(\mathbf{x}_t, \nu_{t'}) &= (\mathbf{x}_t, \mathbf{x}_{t'} - \mathbf{x}_{t'/t'-1})\mathbf{m}_{t'} + (\mathbf{x}_t, v_{t'}) \\ &= (\mathbf{x}_t - \mathbf{x}_{t/t-1}, \mathbf{x}_{t'} - \mathbf{x}_{t'/t'-1})\mathbf{m}_{t'} \\ &\equiv \mathbf{P}(t,t')\mathbf{m}_{t'}\end{aligned} \tag{33}$$

Note that we introduce a matrix $\mathbf{P}(t,t')$ here, which is different from $\mathbf{P}_{t/t'}$ defined in (14). They can be expressed in each other. For example, it holds that

$$\mathbf{P}(t,t) = \mathbf{P}_{t/t-1} \tag{34}$$

To derive a more general relation, we use the transition equation (2) and prediction equation (20b) and write

$$\mathbf{x}_t - \mathbf{x}_{t/t-1} = \mathbf{T}_t(\mathbf{x}_{t-1} - \mathbf{x}_{t-1/t-1}) + \mathbf{w}_t \tag{35a}$$

From filtering equation (26), we may put this in the form

$$\mathbf{x}_t - \mathbf{x}_{t/t-1} = \mathbf{T}_t(\mathbf{x}_{t-1} - \mathbf{x}_{t-1/t-2}) - \mathbf{T}_t\mathbf{K}_{t-1}\nu_{t-1} + \mathbf{w}_t \tag{35b}$$

Substitution of (13) for ν_{t-1} leads to

$$\mathbf{x}_t - \mathbf{x}_{t/t-1} = \mathbf{T}_t(1 - \mathbf{K}_{t-1}\mathbf{m}'_{t-1})(\mathbf{x}_{t-1} - \mathbf{x}_{t-1/t-2}) - \mathbf{T}_t\mathbf{K}_{t-1}v_{t-1} + \mathbf{w}_t \tag{35c}$$

If we substitute this into definition (33) of $\mathbf{P}(t,t')$, we obtain a recurrence relation under the condition $t < t' - 1$:

$$\mathbf{P}(t,t') = \mathbf{P}(t,t'-1)(1 - \mathbf{m}_{t-1}\mathbf{K}'_{t-1})\mathbf{T}'_t \tag{36}$$

To start the iteration, we have to study the case $t = t' - 1$ or $t' = t + 1$ separately. From (35a), it immediately follows that

$$\begin{aligned}\mathbf{P}(t,t+1) &= (\mathbf{x}_t - \mathbf{x}_{t/t-1}, \mathbf{x}_{t+1} - \mathbf{x}_{t+1/t}) \\ &= \mathbf{P}_{t/t}\mathbf{T}'_{t+1}\end{aligned} \tag{37}$$

Combining equations (36),(37) and (30) we find

$$\mathbf{P}(t,t') = (\prod_{i=t}^{t'} \mathbf{P}_i^*)\mathbf{T}'_{t+1} \tag{38}$$

with $\mathbf{P}_i^*$ given by

$$\mathbf{P}_i^* = \mathbf{P}_{i/i}\mathbf{T}'_{i+1}\mathbf{P}_{i+1/i}^{-1} \tag{39}$$

From this explicit representation for $\mathbf{P}(t,t')$, we deduce that

$$\mathbf{P}(t,t') = \mathbf{P}^*_t\mathbf{P}(t+1,t') \tag{40}$$

This enables us to express $\mathbf{x}_{t/M}$ in terms of $\mathbf{x}_{t+1/M}$. From (32),(33) and (16) we have

$$\mathbf{x}_{t/M} = \mathbf{x}_{t/t} + \sum_{i=t+1}^{M} \mathbf{P}(t,i)\mathbf{m}_i(f_i)^{-1}\nu_i \tag{41a}$$

In a similar manner it holds that

$$\mathbf{x}_{t+1/M} = \mathbf{x}_{t+1/t} + \sum_{i=t+1}^{M} \mathbf{P}(t+1,i)\mathbf{m}_i(f_i)^{-1}\nu_i \tag{41b}$$

Substitution of (40) into (41a) yields

$$\mathbf{x}_{t/M} = \mathbf{x}_{t/t} + \mathbf{P}_t^*(\mathbf{x}_{t+1/M} - \mathbf{x}_{t+1/t}) \tag{42}$$

From the orthogonality of the basis elements ν_t we may immediately conclude that the two terms in the right hand side of (42) are orthogonal. This implies that the recurrence relation for $\mathbf{P}_{t/M}$ is given by

$$\mathbf{P}_{t/M} = \mathbf{P}_{t/t} + \mathbf{P}_t^*(\mathbf{P}_{t+1/M} - \mathbf{P}_{t+1/t})(\mathbf{P}_t^*)' \tag{43}$$

with $\mathbf{P}_t^*$ given by (41). So, starting from known expressions for $\mathbf{x}_{M/M}$ and $\mathbf{P}_{M/M}$ we may apply (42) and (43) and work backwards to obtain successively all smoothed estimators and thus estimates.

5.4 Summary of Algorithms.

For convenience, we summarise here the algorithms derived until now. The expressions are in recurrent form and highly appropriate for numerical implementation.

Prediction:

$$\mathbf{x}_{t+1/t} = \mathbf{T}_{t+1}\mathbf{x}_{t/t}$$

$$\mathbf{P}_{t+1/t} = \mathbf{T}_{t+1}\mathbf{P}_{t/t}\mathbf{T}'_{t+1} + \mathbf{Q}_{t+1}$$

The matrices $\mathbf{T}_t$ and $\mathbf{Q}_t$ are introduced in (2) and (3).

Filtering :

$$\mathbf{x}_{t/t} = \mathbf{x}_{t/t-1} + \mathbf{K}_t\nu_t$$

$$\mathbf{P}_{t/t} = (1 - \mathbf{K}_t \mathbf{m}'_t)\mathbf{P}_{t/t-1}$$

The gain $\mathbf{K}_t$ is defined by

$$\mathbf{K}_t = \mathbf{P}_{t/t-1}\mathbf{m}_t(f_t)^{-1}$$

and the innovators ν_t and their normalization by

$$\nu_t = y_t - \mathbf{m}'_t \mathbf{x}_{t/t-1}$$

$$f_t \equiv (\nu_t, \nu_t) = \mathbf{m}'_t \mathbf{P}_{t/t-1}\mathbf{m}_t + R_t$$

The prediction and filtering procedures work forward and have to start with initial estimates $\mathbf{x}_{0/0}$ and $\mathbf{P}_{0/0}$. These estimates should be unbiased and have minimum variance. In practice, these estimates are seldom available. See also section 9. for a discussion of this aspect.

Smoothing :

The smoothing procedure works backwards and assumes that the prediction and filtering estimates have already been determined at the time points $t = 1, 2, ..., M$.

$$\mathbf{x}_{t/M} = \mathbf{x}_{t/t} + \mathbf{P}^*_t(\mathbf{x}_{t+1/M} - \mathbf{x}_{t+1/t})$$

$$\mathbf{P}_{t/M} = \mathbf{P}_{t/t} + \mathbf{P}^*_t(\mathbf{P}_{t+1/M} - \mathbf{P}_{t+1/t})(\mathbf{P}^*_t)'$$

with the matrix $\mathbf{P}^*_t$ given by

$$\mathbf{P}^*_t = \mathbf{P}_{i/i}\mathbf{T}'_{i+1}\mathbf{P}^{-1}_{i+1/i}$$

6 Kalman Theory: The Gaussian Case.

In this section, we show that the derivations in the preceding section can be performed in an alternative, elegant manner if all stochastic processes are normally distributed. The present derivations are taken from [Meinhold and Singpurwalla 1983, 1987].An analogous approach is presented in [Rauch, Tung and Striebel 1965]. In the model contained in equations (1),(2) and (3), this implies that the initial distribution of the state $\mathbf{x}_0$ together with the disturbances v_t and $\mathbf{w}_t$ are assumed to be normally distributed. Then, the same holds for $\mathbf{x}_t$ and y_t at all times, as already noted in remark b of section 2.

In the following we shall make use of the following well-known theorems for normally distributed vectors x:

Theorem 1. If $x \sim N(\mu, \Sigma)$ then $Ax \sim N(A\mu, A\Sigma A')$ with A an arbitrary matrix.

Theorem 2. If $\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix} \right\}$
then

$$(\mathbf{x}_1|\mathbf{x}_2) \sim N\{\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\}$$

It is convenient to summarise equations (1) and (2) in the following form:

$$\begin{pmatrix} \mathbf{x}_t \\ y_t \end{pmatrix} \Big| \mathbf{Y}_{t-1} = \begin{pmatrix} \mathbf{T}_t & I & 0 \\ \mathbf{m}'_t\mathbf{T}_t & \mathbf{m}'_t & I \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1} \\ \mathbf{w}_t \\ v_t \end{pmatrix} \Big| \mathbf{Y}_{t-1} \tag{44}$$

with I the identity matrix. As stated in equation (6), we have in the Gaussian case

$$\bar{\mathbf{x}}_{t/t} = E(\mathbf{x}_t|\mathbf{Y}_t). \tag{45}$$

Theorem 1 implies that

$$\begin{pmatrix} \mathbf{x}_t \\ y_t \end{pmatrix} \Big| \mathbf{Y}_{t-1} \sim N \left\{ \begin{pmatrix} \mathbf{T}_t\bar{\mathbf{x}}_{t-1/t-1} \\ \mathbf{m}'_t\mathbf{T}_t\bar{\mathbf{x}}_{t-1/t-1} \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma'_{xy} & \Sigma_{yy} \end{pmatrix} \right\} \tag{46}$$

with the Σ matrices given by

$$\begin{aligned} \Sigma_{xx} &\equiv \mathbf{P}_{t/t-1} = \mathbf{T}_t\mathbf{P}_{t-1/t-1}\mathbf{T}'_t + \mathbf{Q}_t \\ \Sigma_{xy} &\equiv \Sigma_{xx}\mathbf{m}_t \\ \Sigma_{yy} &\equiv f_t \quad = \mathbf{m}'_t\Sigma_{xx}\mathbf{m}_t + R_t \end{aligned} \tag{47}$$

Application of theorem 2 yields

$$(\mathbf{x}_t|\mathbf{Y}_t) \sim N\{\mathbf{T}_t\bar{\mathbf{x}}_{t-1/t-1} + \Sigma'_{xy}\Sigma_{yy}^{-1}(y_t - \mathbf{m}'_t\mathbf{T}_t\bar{\mathbf{x}}_{t-1/t-1}), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma'_{xy} \tag{48}$$

So we have obtained the probability density function of $\mathbf{x}_t$ given the data $\mathbf{Y}_t$. It follows that equation (47) contains the same information as contained in prediction equations (20b) and (21) and filtering equations (26) and (30).

To obtain the smoothing equations (42) and (43), we have to concentrate on the joined probability distribution of $\mathbf{x}_t, \mathbf{x}_{t+1}$ and y_{t+1} given the data $\mathbf{Y}_t$. Then, we have

$$\begin{pmatrix} \mathbf{x}_t \\ \mathbf{x}_{t+1} \\ y_{t+1} \end{pmatrix} \Big| \mathbf{Y}_t \Bigg) = \begin{pmatrix} I & O & O \\ \mathbf{T}_{t+1} & I & O \\ \mathbf{m}'_{t+1}\mathbf{T}_{t+1} & \mathbf{m}'_{t+1} & I \end{pmatrix} \begin{pmatrix} \mathbf{x}_t \\ \mathbf{w}_{t+1} \\ v_{t+1} \end{pmatrix} \Big| \mathbf{Y}_t \Bigg) \tag{49}$$

Successive application of the theorems 1 and 2 yields the backward recurrence relations (42) and (43). The formulae have been written out in appendix A of [Meinhold and Singpurwalla 1987]. This Bayesian approach is quite attractive, because it provides information about $\mathbf{x}_t$ through its distribution rather than just a point estimator. However, this is only possible thanks to the restrictive additional information of normality of all stochastic processes involved.

7 The Kalman Filter and Least Squares Estimation.

It is useful to point out the relationship between Kalman estimation and the familiar least squares estimation. For the Kalman theory the recursive character of all calculations is an essential feature, whereas the least squares method is usually not put into that form. However, much insight is gained if also the least squares method is approached that way. Then, it is immediately recognised that this method is a special case of Kalman estimation. This idea has also been worked out [Duncan and Horn 1972, Young 1984]. First, let us present a brief derivation of the least squares method and study the regression model

$$y = \mathbf{m}'\mathbf{x} + v \tag{50}$$

with y a stochastic scalar process, which is measured at times $i = 1, 2, ..., t$. The vector $\mathbf{m}$ of length N contains the regression variables m^j and the vector $\mathbf{x}$ of length N the regression coefficients, which are to be estimated. They are assumed to be constant in time. The noise process v represents the measurement error.

We introduce the following notation:

$$\mathbf{Y}_t = (y_1, y_2, ..., y_t)' \tag{51}$$

with y_i the value of y measured at time i;

$$(\mathbf{M}_t)_{ij} = m_i^j, \; j = 1, ..., N; \; i = 1, ..., t \tag{52}$$

with m_i^j the value of m^j at time i;

$$\mathbf{V}_t = (v_1, v_2, ..., v_t)' \tag{53}$$

with v_i the (unknown) value of v at time i.
Then, we may summarise all information up to time t in the equation

$$\mathbf{Y}_t = \mathbf{M}_t\mathbf{x} + \mathbf{V}_t \tag{54}$$

The criterium for the estimation of $\mathbf{x}$ is to minimise the length of the vector $\mathbf{V}_t$ given by

$$\|\mathbf{V}_t\|^2 = \sum_{i=1}^{t} v_i^2 \tag{55}$$

Because this length is just an estimation of the variance of the stochastic process, we may denote this as minimum variance estimation. From (54) we have

$$\begin{aligned} \|\mathbf{V}_t\|^2 &= (\mathbf{Y}_t - \mathbf{M}_t\mathbf{x})'(\mathbf{Y}_t - \mathbf{M}_t\mathbf{x}) \\ &= \mathbf{Y}'_t\mathbf{Y}_t - \mathbf{x}'\mathbf{M}'_t\mathbf{Y}_t - \mathbf{Y}'_t\mathbf{M}_t\mathbf{x} + \mathbf{x}'\mathbf{M}'_t\mathbf{M}_t\mathbf{x} \\ &= \mathbf{Y}'_t\mathbf{Y}_t - 2\mathbf{x}'\mathbf{M}'_t\mathbf{Y}_t + \mathbf{x}'\mathbf{M}'_t\mathbf{M}_t\mathbf{x} \end{aligned} \tag{56}$$

If we differentiate the latter expression with respect to $\mathbf{x}'$ and require the result to vanish, we find

$$\mathbf{M}'_t\mathbf{M}_t\mathbf{x} = \mathbf{M}'_t\mathbf{Y}_t \tag{57}$$

From this equation the estimator $\hat{\mathbf{x}}_t$ of $\mathbf{x}$ is obtained based on the information at times $i = 1, ..., t$. If the symmetric matrix $\mathbf{M}'_t\mathbf{M}'_t$ is non-singular, we arrive at the well-known ordinary least squares formula

$$\hat{\mathbf{x}}_t = (\mathbf{M}'_t\mathbf{M}_t)^{-1}\mathbf{M}'_t\mathbf{Y}_t \tag{58}$$

For completeness, we remark that this estimator is unbiased and if v is normally distributed around zero, i.e. $v \sim N(0, \sigma^2)$, then, it holds that $\hat{\mathbf{x}}_t \sim N(x, (\mathbf{M}'_t\mathbf{M}_t)^{-1}\sigma^2)$ .Equation (58) for $\hat{\mathbf{x}}_t$ is not in a recursive form. If $t \rightarrow t+1$ all dimensions are enlarged and the matrix inversion has to be performed anew. A recurrence relation would allow us to express $\hat{\mathbf{x}}_{t+1}$ in terms of $\hat{\mathbf{x}}_t$ and the most recent information contained in y_{t+1} and $\mathbf{m}_{t+1}$. To obtain this relation we define

$$\mathbf{B}_t = \mathbf{M}'_t\mathbf{M}_t \tag{59}$$

and remark that

$$\mathbf{B}_{t+1} = \mathbf{B}_t + \mathbf{m}_{t+1}\mathbf{m}'_{t+1} \tag{60}$$

From one of the lemmas for matrix inversion, presented in [Jazwinski 1970], we find that

$$\mathbf{B}_{t+1}^{-1} = \mathbf{B}_t^{-1} - \frac{\mathbf{B}_t^{-1}\mathbf{m}_{t+1}\mathbf{m}'_{t+1}\mathbf{B}_t^{-1}}{(1 + \mathbf{m}'_{t+1}\mathbf{B}_t^{-1}\mathbf{m}_{t+1})} \tag{61}$$

If we substitute this expression for $\mathbf{B}_{t+1}^{-1}$ in the right hand side of (58) with t replaced by $t+1$, we obtain

$$\hat{\mathbf{x}}_{t+1} = \hat{\mathbf{x}}_t + \bar{\mathbf{K}}_{t+1}(y_{t+1} - \mathbf{m}'_{t+1}\hat{\mathbf{x}}_t) \tag{62}$$

with the matrix $\bar{\mathbf{K}}$ given by

$$\bar{\mathbf{K}}_{t+1} = \mathbf{B}_t^{-1}\mathbf{m}_{t+1}/(1 + \mathbf{m}'_{t+1}\mathbf{B}_t^{-1}\mathbf{m}_{t+1}) \tag{63}$$

Note that the latter two equations strongly resemble equations (26) and (27) for the Kalman filter. This similarity reflects that the ordinary least squares method can be interpreted as a special case of the univariate Kalman approach. In the least squares regression method, the coefficients in the linear regression model are assumed to be constant. This implies that, in the transition equation (2), $\mathbf{T}_t$ should be identified with the identity matrix while $\mathbf{Q}_t = 0\ \forall t$ should be taken in equations (3). Furthermore, we have the identification $\mathbf{P}_{t+1/t} = \mathbf{P}_{t/t} \leftrightarrow \mathbf{B}_t^{-1}$ as follows from equation (21) and comparison of the gainmatrices $\mathbf{K}_t$, given in equation (27) and $\bar{\mathbf{K}}_t$, given above.

8 Maximum Likelihood Estimation.

In many problems, the model in equations (1),(2) and (3) contains some parameters which have still to be estimated, together with the state $\mathbf{x}_t$, from the measured data. This is possible via the maximum likelihood approach, if the stochastic processes y_t are jointly, normally distributed. Then, the maximum likelihood function $L(\mathbf{Y}_t)$, with $\mathbf{Y}_t$ given by (4), is [Harvey 1981b]

$$logL(\mathbf{Y}_t) = \frac{-(t+1)}{2} log(2\pi) - \frac{1}{2} log(det(cov(\mathbf{Y}_t))) - \frac{1}{2} \mathbf{Y}'_t (cov^{-1}(\mathbf{Y}_t)\mathbf{Y}_t \tag{64}$$

The calculation of $L(\mathbf{Y}_t)$ for successive t points is awkward in the form of equation (64). Therefore, we shall present a recurrence relation for L in terms of the innovations ν_t, introduced in section 4. There, we pointed out that the ν_t are obtained from the y_t by means of transformation (11). Because the ν_t are orthogonal, we have that the covariance matrix of $\mathbf{N}_t$, defined in (10), is diagonal with diagonal elements f_t, given by (15). Because transformation (11) is non-singular, we may write

$$\mathbf{Y}'_t cov^{-1}(\mathbf{Y}_t)\mathbf{Y}_t = \mathbf{N}'_t cov^{-1}(\mathbf{N}_t)\mathbf{N}_t \tag{65}$$

Further, we have

$$\begin{aligned} det(cov(\mathbf{Y}_t)) &= det\{\mathbf{L}_t^{-1} cov(\mathbf{N}_t)(\mathbf{L}_t^{-1})'\} \\ &= det(\mathbf{L}_t^{-1}) \cdot det(cov(\mathbf{N}_t)) \cdot det((\mathbf{L}_t^{-1})') \\ &= det(cov(\mathbf{N}_t)) \\ &= \prod_{i=0}^{t} f_i \end{aligned} \tag{66}$$

Here, we make use of the fact that $det(\mathbf{L}_t) = 1$, because $\mathbf{L}_t$ is triangular with unity diagonal elements. In terms of the ν_t and f_t, L is given by :

$$logL(\mathbf{Y}_t) = -\frac{t+1}{2} log(2\pi) - \frac{1}{2} \sum_{i=0}^{t} \{log f_i + \frac{\nu_i^2}{f_i}\} \tag{67}$$

This equation is known as the "prediction error decomposition" [Harvey 1981b,1984]. If $L(\mathbf{Y}_{t-1})$ has been calculated, it suffices to calculate ν_t and f_t in order to obtain $L(\mathbf{Y}_t)$ directly.

The unknown parameters can be evaluated by maximising L as a function of these parameters. This optimisation problem is strongly non-linear. In practice, this approach is tractable only if the number of

unknown parameters is restricted, e.g. by assuming that they are time independent. See also remark c in section 2. In order to gain computational speed and simplicity it is also desirable to pose, as much as is reliable, restrictions on the dimensions and parameters of the model.

9. Structural Models and the Analysis of Tree-Ring Series.

In regression analysis, the use of so-called structural models has become widespread. Such a model is based on the decomposition of the measured signal y_t into trend, seasonal component, explanatory variables and a noise term in an additive way :

$$y_t = (Trend) + (Seasonal \quad Components) + (Explanatory \quad Variables) + (Noise) \tag{68}$$

This kind of model can be put in state space form and thus analysed with the help of Kalman theory. We refer to [Harvey 1981a and b, 1984, Harvey, Henry, Peters and Wren-Lewis 1986, Harvey and Durbin 1986, Mettes and Visser 1987, Meinhold and Singpurwalla 1987]. An example is the analysis of tree-ring series [Visser and Molenaar 1988]. In this application the smoothing features of Kalman theory are utilised. The idea is to regress tree-ring data on weather data in order to detect possible variations in tree behaviour. To that end, the time dependent coefficients in a linear regression model are to be estimated. We make therefore the following identifications:

y_t represents tree-ring width or basal area increment in year t.

$\mathbf{m}_t$ contains the weather data at time t.

$\mathbf{x}_t$ is the vector of regression coefficients.

The behaviour of $\mathbf{x}_t$ in time reflects the relation between growth and specific weather conditions. An apparent and common choice for the components of $\mathbf{m}_t$ is to make use of temperature and precipitation data averaged over one month. Note that $\mathbf{m}_t$ must also contain weather information from the year preceding to year t, because a considerable delay exists between tree growth and the preparation for the growth processes.

Because hardly any biological information is available about the dynamic behaviour of the coefficients one usually takes $\mathbf{T} \equiv 1$ in transition equation (2), so that a random walk behaviour results :

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{w}_t \tag{69}$$

The prediction, filtering and smoothing formulae to be implemented in a computer program are given in section 5.4. The unknown variances R and $\mathbf{Q}$, which are assumed constant in time, can be estimated by optimising the likelihood function L given in section 8. Because starting values $\mathbf{x}_0$ are generally unknown, one usually chooses an arbitrary value in combination with large diagonal elements for $\mathbf{P}_{0/0}$. As shown by Jazwinski (1970), the prior data are eventually forgotten and a bias stemming from initial uncertainties

damps out after sufficient, say N_s, observations have been processed. The first N_s iteration steps serve as a transient period for the filtering process. The results of the smoothing process are still reliable in this period. Because the likelihood function is connected with filtering, the transient time points may not be included into this function. Unknown parameters follow from optimising the function

$$L_c = \sum_{t=N_s}^{T} \{log f_t + \frac{\nu_t^2}{f_t}\}, \tag{70}$$

which is called the concentrated likelihood function and is obtained from (67) by omitting irrelevant factors.

References.

Bagchi ,A, Stochastic Filter and Identification Theory, Report 57016, University Twente, 1982.

Duncan, D.B., Horn, S.D., Linear Dynamic Recursive Estimation From The Viewpoint Of Regression Analysis, JASA, Vol. 67, no 340, 815-821, 1972.

Harvey, A.C., The Econometric Analysis Of Time Series, P.Allan Publishers, Oxford, 1981a.

-, Time Series Models, P.Allan Publishers, Oxford, 1981b.

-, A Unified View Of Statistical Forecasting Procedures, J. of Forecasting, Vol. 3, 245-275, 1984.

Harvey, A.C., Henry, S.G.B., Peters, S. and Wren-Lewis, S., Stochastic Trends In Dynamic Regression Models : An Application To The Employment-Output Equation, The Econometric Journal, 96, 975-985, 1986.

Harvey, A.C., Durbin, J., The Effects Of Seat Belt Legislation On British Road Casualties : A Case Study In Structural Time Series Modelling, J.R.Statist.Soc.A, 149, part 3, 187-227, 1986.

Jazwinski, A.H., Stochastic Processes And Filtering Theory, Acad.Press, New York, 1970.

Kalman, R.E., A New Approach To Linear Filtering And Prediction Problems, Trans.ASME, Series D, Vol.82,35-45, 1960.

Kalman, R.E., Bucy, R.S., New Results In Linear Filtering, 1961.

Meinhold, R.J. , Singpurwalla, N.D., Understanding The Kalman Filter, The American Statistician, Vol. 37, no. 2, 123-127, 1983.

-, A Kalman-Filter Smoothing Approach For Extrapolations in Certain Dose-Response, Damage- Assessment, And Accelerated Life-Testing Studies, The American Statistician, Vol. 41, 1987.

Mettes, M.A.C., Visser, H., KALFIMAC, A Software Package To Analyse Time-Series With Trend, Cycli And Explanatory Variables, Part 1, KEMA Report 50385- MOA, 87-3129, 1987.

Molenaar, J., Visser, H., The Kalman Filter In Dendroclimatology, Proceedings ICIAM '87, Paris, 203-214, 1987.

Otter, P.W., Dynamic Feature Space Modelling, Filtering And Self-Tuning Control Of Stochastic Systems, Ph.D.Thesis, Groningen University, 1984.

Rauch, H.E., Tung, F., and Striebel T., Maximum Likelihood Estimates Of Linear Dynamic Systems, AIAA Journal, Vol. 3, no.8, 1445-1450, 1965.

Sage, A.P., Melsa, J.L., Estimation Theory With Applications To Communications And Control, McGraw Hill, New York, 1971.

Visser, H., Analysis Of Tree-Ring Data Using The Kalman Filter Technique, IAWA- Bulletin 7(4), New Series, 289-297, 1986.

Visser, H., Molenaar, J., Kalman Filter Analysis In Dendroclimatology, Biometrics (accepted), 1988.

Young, P., Recursive estimation And Time Series Analysis, Springer Verlag, 1984.