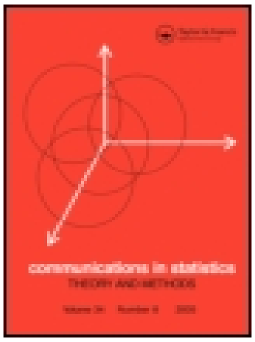

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

A kalman filter in the presence of outliers

Nihal Yatawara , Bovas Abraham & John F. MacGregor

COMMUN. IN STATIST.-THEORY METH., 20(5&6) 1803-1820 (1991)

A KALMAN FILTER IN THE PRESENCE OF OUTLIERS

Nihal Yatawara, Bovas Abraham, and John F. MacGregor*

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario
N2L 3G1

Keywords and Phrases: Robust estimation, State estimation, Kalman Filtering, Outliers, Markov Chains, Bayesian inference

ABSTRACT

A Kalman Filtering algorithm which is robust to observational outliers is developed by assuming that the measurement error may come from either one of two Normal distributions, and that the transition between these distributions is governed by a Markov Chain. The resulting algorithm is very simple, and consists of two parallel Kalman Filters having different gains. The state estimate is obtained as a weighted average of the estimates from the two parallel filters, where the weights are the posterior probabilities that the current observation comes from either of the two distributions. The large improvements obtained by this Robust Kalman Filter in the presence of outliers is demonstrated with examples.

*Nihal Yatawara is now at the Institute of Fundamental Studies, Kandy, Sri Lanka. Bovas Abraham is Professor, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. John F. MacGregor is Professor, Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4L7.

1. Introduction

The Kalman Filter is a well known recursive estimator for the state of a linear system, and has been widely used in the fields of forecasting and control. (For example see Anderson and Moore (1979) Duncan and Horn (1972), and Priestly (1980)). It has been derived as a least squares estimator, and also, under the assumption of Normality, as a Bayesian estimate. However, as with most least squares estimators, it is very sensitive to observational outliers. This sensitivity to outliers is a major drawback of the Filter because it is generally used where data from on-line sensors etc. are being processed sequentially. Bad observations arising from periodic sensor problems, etc. can seriously bias the Filter estimates for a considerable period of time thereafter.

In this paper a Bayesian approach is used to derive a recursive state estimator which allows for the fact that occasional spurious observations will arise. The new state estimator is shown to have the structure of two parallel Kalman Filters in which the final state estimate is a weighted average of the estimators from the two filters. The weights correspond to the posterior probabilities that the current observation is or is not an outlier, and therefore the filter effectively downweights any outliers.

2. The State Variable Model

Consider the state variable model given by

$$x_t = \phi x_{t-1} + G a_t \tag{2.1}$$

$$z_t = H x_t + \epsilon_t \tag{2.2}$$

where x_t is an $(n \times 1)$ vector of state variables defining the dynamic behaviour of a system, ϕ, G are known matrices (possibly time varying) and H is a known row vector, of the appropriate dimensions. The measurement z_t taken at time t is related linearly to the states through the measurement equation (2.2). The state and observation noise sequences $\{a_t\}$ and $\{\epsilon_t\}$ are usually assumed to be i.i.d. Gaussian random variables with known variances σ_a^2 and σ_ϵ^2 respectively. Under the assumption that the system is observable the Kalman Filter (Jazwinski, 1970) provides a set of recursive equations for the one-step ahead predictions of the state $\hat{x}_{t+1|t}$, and the filtered state estimates $\hat{x}_{t|t}$, given information up to and including time t. However, the optimality of the estimator depends upon knowledge of the variances σ_a^2 and σ_ϵ^2 at each point in time. The occurrence of infrequent outliers can be characterized by an increase in the variance of a_t (an innovational outlier) which leads to a real change in the state of the system, or in the variance of ϵ_t (an observational outlier) which has no effect on the true state (x_t), but will have a pronounced effect on the Kalman Filter estimates. In the following section we use a Bayesian approach to develop a robust filter to handle observational outliers.

3. Derivation of the Robust Filter

We characterize the observational outliers by replacing ϵ_t in equation (2.1) by $(1-i_t)\epsilon_t + i_t\gamma_t$, where $\{i_t,\ t = 0,1,...\}$ is a sequence of $\{0,1\}$ random variables following a Markov Chain (M.C.). Suppose that the initial probability (Π_{00}, Π_{01}) of the M.C. is known and that the transition probability-matrix is (λ_{i_{t-1},i_t}). γ_t is a Gaussian random variable with zero mean and variance σ_γ^2 which is very large compared with σ_ϵ^2. We also approximate the conditional density $p(x_{t-1}|i_{t-1},Z^{t-1})$ by $p(x_{t-1}|Z^{t-1})$ given all the observations $Z^{t-1} = (z_1,z_2,\ldots,z_{t-1})'$. We further assume that this is a Gaussian distribution with mean $\hat{x}_{t-1}$ and covariance matrix P_{t-1}. Then

$$p(x_t, i_t|Z^{t-1}) = \sum_{i_{t-1}=0}^{1} \int p(x_{t-1}, i_{t-1}|Z^{t-1})p(x_t, i_t|x_{t-1}, i_{t-1},Z^{t-1})dx_{t-1}$$

which on further simplification (Yatawara, 1986) yields

$$p(x_t, i_t|Z^{t-1}) \doteq (2\pi)^{-n/2} \sum_{i_{t-1}=0}^{1} \lambda_{i_{t-1}i_t} k_{i_{t-1}} |P_{t|t-1}|^{-1/2}$$

$$\times \exp[-1/2\{(x_t-\hat{x}_{t|t-1})'P_{t|t-1}^{-1}(x_t-\hat{x}_{t|t-1})\}] \tag{3.1}$$

where $\lambda_{i_{t-1}i_t} = Pr(i_t|i_{t-1}, Z^{t-1})$, $P_{t|t-1} = \phi P_{t-1}\phi' + GG'\sigma_a^2$,

$\hat{x}_{t|t-1} = \phi\hat{x}_{t-1}$, and $k_{i_{t-1}} = Pr(i_{t-1}|Z^{t-1})$.

Also, by Bayes theorem

$$p(x_t\,,\,i_t\,|Z^t) = \frac{p(x_t\,,\,i_t\,|Z^{t-1})p(z_t\,|x_t\,,\,i_t\,,\,Z^{t-1})}{\sum_{i_t}\int p(x_t\,,\,i_t\,|Z^{t-1})p(z_t\,|x_t\,,\,i_t\,,\,Z^{t-1})dx_t}$$

$$\doteq k(2\pi)^{-n/2}\Big(\sum_{i_{t-1}=0}^{1}\lambda_{i_{t-1}i_t}k_{i_{t-1}}\Big)\,|\,P_{t\,|t-1}\,|^{-1/2}$$

$$\times\,\exp[-1/2\{(x_t-\hat{x}_{t\,|t-1})'P_{t\,|t-1}^{-1}(x_t-\hat{x}_{t\,|t-1})\}]$$

$$\times\,(\sqrt{2\pi}\sigma_{i_t})^{-1}\exp[-1/2\{(z_t-Hx_t)'\sigma_{i_t}^{-2}(z_t-Hx_t)\}] \tag{3.2}$$

where $\sigma_{i_t}^2 = (1-i_t)\sigma_\epsilon^2 + i_t\sigma_\gamma^2$, and $k^{-1} = \sum_{i_t=0}^{1}\int\, p(x_t\,,\,i_t\,|Z^{t-1})p(z_t\,|x_t\,,\,i_t\,,\,Z^{t-1})dx_t$. This can be further simplified (see Yatawara (1986)),

$$p(x_t\,,\,i_t\,|Z^t) = k_{i_t}(2\pi)^{-n/2}\,|P_{t\,|t}^{(i_t)}|^{-1/2}\,\exp\{-1/2(x_t-\hat{x}_{t\,|t}^{(i_t)})P_{t\,|t}^{(i_t)-1}(x_t-\hat{x}_{t\,|t}^{(i_t)})\} \tag{3.3}$$

where

$$\hat{x}_{t\,|t}^{(i_t)} = \hat{x}_{t\,|t-1} + R_{i_t}^{-1}P_{t\,|t-1}H'(z_t-H\hat{x}_{t\,|t-1}). \tag{3.4}$$

$$P_{t\,|t}^{(i_t)} = P_{t\,|t-1} - R_{i_t}^{-1}(P_{t\,|t-1}H'HP_{t\,|t-1}), \tag{3.5}$$

$$R_{i_t} = \sigma_{i_t}^2 + HP_{t\,|t-1}H' \tag{3.6}$$

and

$$k_{i_t} = C(\sqrt{2\pi}\;\sigma_{i_t})^{-1}\,|P_{t\,|t}^{(i_t)}|^{1/2}\Big(\sum_{i_{t-1}=0}^{1}\lambda_{i_{t-1}i_t}k_{i_{t-1}}\Big)\,|P_{t\,|t-1}\,|^{-1/2} \tag{3.7}$$

$$\exp[-1/2\{z_t-H\hat{x}_{t\,|t-1})'R_{i_t}^{-1}(z_t-H\hat{x}_{t\,|t-1})\}]$$

The k_{i_t}'s are the posterior probabilities of the current observation at time t being an outlier ($i_t = 1$) or not being an outlier ($i_t = 0$). C is a normalizing constant. Note that the term $\sum_{i_{t-1}=0}^{1}[\lambda_{i_{t-1}i_t}k_{i_{t-1}}] = \Pi_{ti_t}$ is the prior probability of being in state i_t at time t. We can consider a suitable transition probability matrix and carryout the computations. However in our examples we take Π_{ti_t} to be fixed in advance. For example we might select $\Pi_{t0} = 0.99$ and $\Pi_{t1} = 0.01$ to allow for a 1% probability of having an outlier. Thus

$$p(x_t \,|Z^t) = \sum_{i_t=0}^{1} k_{i_t} N(\hat{x}_{t\,|t}^{(i_t)} \,,\, P_{t\,|t}^{(i_t)}) \tag{3.8}$$

where $N(\hat{x}_{t\,|t}^{(i_t)} \,,\, P_{t\,|t}^{(i_t)})$ denotes a normal distribution with mean $\hat{x}_{t\,|t}^{(i_t)}$ and covariance matrix $P_{t\,|t}^{(i_t)}$. Recursive formulae for updating these are given in (3.4) and (3.5). However, this leads to a computational difficulty: a single Gaussian distribution at time $t-1$ leads to a posterior which is the sum of two Gaussian distributions at time t. Continuing on to time $t+1$, the posterior would then be a sum of 4 Gaussian densities, and so on. Such a procedure requires a prohibitively large number of computations to be carried out at a much later point in time. For computational simplicity we thus approximate the posterior density (3.8) by a single Gaussian distribution with matching moments. We write this posterior density as

$$p(x_t \,|Z^t) = N(\hat{x}_t \,,\, P_t)$$

$$\text{where } \hat{x}_t = \sum_{i_t=0}^{1} k_{i_t} \hat{x}_{t\,|t}^{(i_t)} \text{ and } P_t = \sum_{i_t=0}^{1} k_{i_t} P_{t\,|t}^{(i_t)}. \tag{3.9}$$

Thus, the collapsing procedure is equivalent to setting $\hat{x}_t$ to be the conditional expectation of the states obtained from the two parallel filters and P_t to be the conditional expectation of the covariance matrices obtained from the two filters.

On line outlier detection can also be performed using this scheme by monitoring the k_{i_t}'s which are the posterior probabilities of no outlier ($i_t = 0$) or an outlier ($i_t = 1$) at each time point t. A large value of k_1 indicates the presence of a discrepant observation.

The gain of a filter determines how alert it is to track changing properties of the system. A larger gain implies quicker adaptation to a new environment but also more sensitivity to observational outliers. Thus a good choice would be a compromise between the two.

In the proposed robust filter (see (3.4), (3.5) and (3.9)), the gain corresponding to i_t is given by

$$K_{i_t} = R_{i_t}^{-1} \hat{P}_{t\,|t-1} H'$$

where R_{i_t} is as given in (3.6). Thus the state estimate is computed as a linear combination of two parallel filters having a lower and a higher gain, depending on the current observation z_t. Thus it has the capability to adapt to changing environments quickly, while maintaining resistance to observational outliers.

A summary of the robust filter is given in the form of a flow chart in Fig. 3.1.

4. Simulation Examples and Sensitivity Studies

4.1 ARIMA (0,1,1) model with outliers

Consider the following special case of the system of equations given by (2.1) and (2.2)

$$x_t = x_{t-1} + a_t \tag{4.1}$$

$$z_t = x_t + \epsilon_t \tag{4.2}$$

where $\{a_t\}$ and $\{\epsilon_t\}$ are i.i.d. Gaussian random variables with mean zero and variances σ_a^2 and σ_ϵ^2 respectively. This is equivalent to an ARIMA (0,1,1) model:

$$\nabla z_t = (1-\theta B)b_t \tag{4.3}$$

for $\theta > 0$, where θ is the moving average (MA) parameter, B is a backshift operator such that $Bz_t = z_{t-1}$ and b_t is an i.i.d. sequence of Gaussian random variables with mean zero and variance σ_b^2. The relationships among the parameters in the state model form (4.1), (4.2), and the difference equation (4.3) are given by Box and Jenkins (1976) as

$$(1-\theta)^2/\theta = \sigma_a^2/\sigma_\epsilon^2 \text{ and } \sigma_b^2 = \sigma_a^2/(1-\theta)^2. \tag{4.4}$$

The procedure in the previous section is illustrated with a time series, series A, given in Box and Jenkins (1976) with 4 outliers introduced at specific points. It has been shown in Box and Jenkins (1976) that an ARIMA (0,1,1) model adequately fits these data. From the first 75 observations the following estimates of the parameters were obtained: $\hat{\theta} = 0.7$ and $\hat{\sigma}_b^2 = 0.101$. It follows from (4.4) that the equivalent state model parameters are $\hat{\sigma}_a^2 = 0.009$ and $\hat{\sigma}_\epsilon^2 = 0.071$. The subsequent 96 observations were then considered as the sample period and

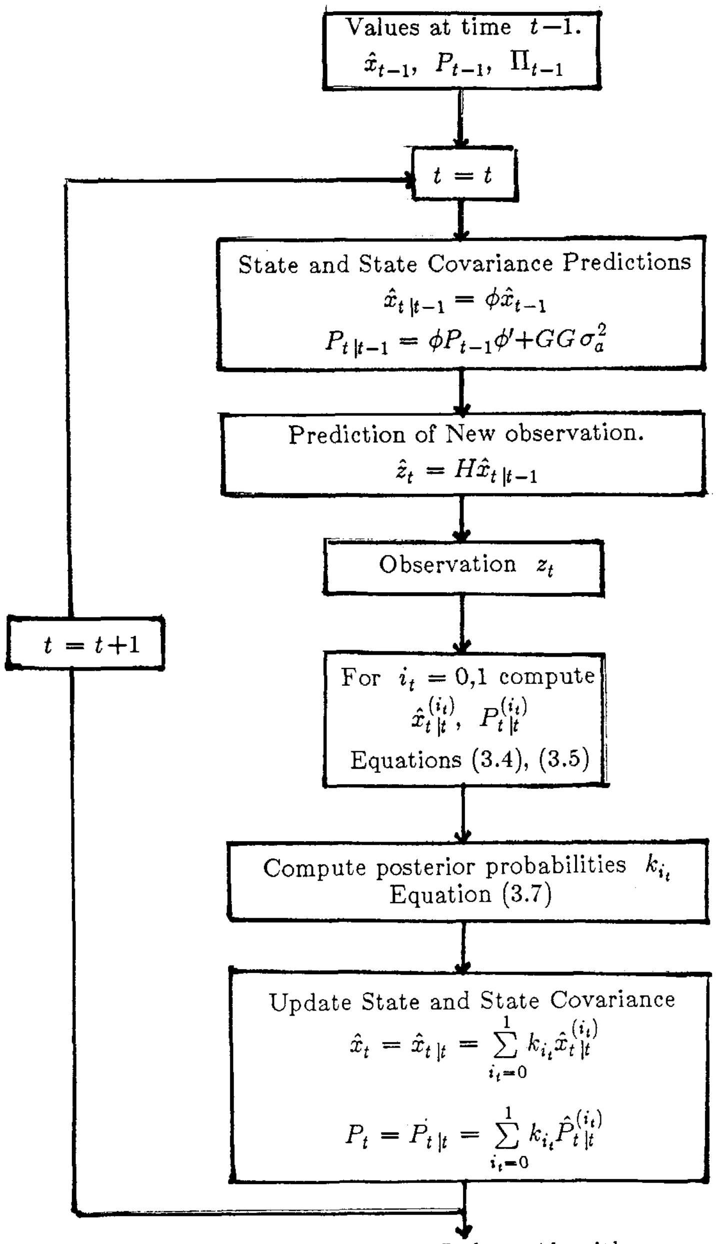

Figure 3.1: Robust Algorithm

outliers were introduced at the 25th, 50th, 65th and 75th observations by adding $N(0\ ,\ 1.75)$ random variables to them.

Initially the ordinary Kalman filter (ie; with no outlier protection) was run with $\hat{x}_0 = 17$ (1st observation in the sample period). $\hat{P}_0 = 0.009$, $\hat{\sigma}_a^2 = 0.009$ and $\hat{\sigma}_\epsilon^2 = 0.071$. The predictions obtained and the actual observations are given in Fig. 4.1. This clearly shows that the ordinary Kalman filter estimates are sensitive to spurious observations. For example the predictions after the spurious observation at $t = 50$, remain too high for a considerable period of time, before aligning with the data again. A similar behaviour can be noticed at time points subsequent to the other discrepant observations.

The robust filter was then run assuming $\hat{x}_0 = 17$, $\hat{P}_0 = 0.009 = \hat{\sigma}_a^2$, $\sigma_\epsilon^2 = 0.071$, $\sigma_\gamma^2 = 1.75$, $\Pi_{t0} = 0.9$ and $\Pi_{t1} = 0.1$. The predictions and the actual observations are plotted in Fig. 4.2. These indicate that this filter is robust to observational outliers. For example, the prediction at $t = 50$ and at the subsequent points do not seem to be unduly affected by the aberrant observation at $t = 50$, unlike in the previous case. Similar behaviour can be seen at $t = 25$, 65, and 75.

In addition, the robust filter also provides the posterior probability (k_1) of any observation being an outlier. These values of k_1 are plotted in Fig. 4.3. They clearly indicate the true positions of the outliers.

The above plots represent a simulation where the four outliers were all quite large. This simulation was chosen simply to illustrate the behaviour of the filter. However, in the following section some of the properties of the filter based on a large number of simulations will be summarized.

4.2 Sensitivity of the Filter to Changes in some Parameters

In the above example, the value of σ_γ was assumed known. However, in practice its value is generally unknown. Thus it is important to know how sensitive the filter is to changes in $\sigma_\gamma/\sigma_\epsilon$. This is investigated in the following simulation study.

100 observations were generated using the model (4.1) and (4.2) with $\sigma_a = 1$, $\hat{x}_1 = 0$, $\hat{P}_1 = 1$, $\sigma_\epsilon = 5$. Outliers were introduced at $t = 25, 50, 65, 75$ by taking $\sigma_\gamma = 25$. The robust filter was run with

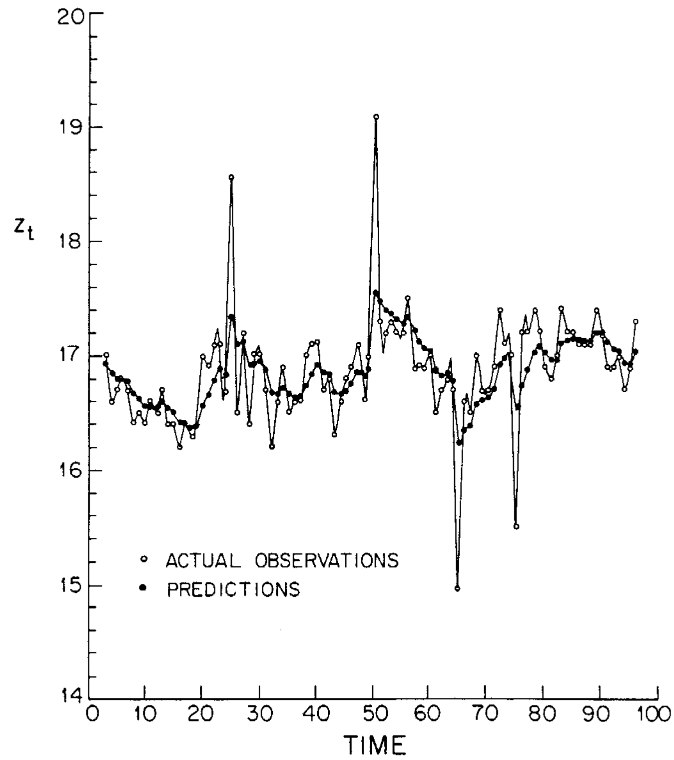

FIG 4.1 Predictions from the ordinary Kalman Filter

$\Pi_{t1} = Pr(i_t = 1) = 0.1$ and the average of k_1 (prob. of an outlier) at these four points was computed, assuming $\sigma_\gamma/\sigma_\epsilon = 2,\ 3,\ 5$ and 10 (the true value $= 5$). The whole process was repeated 100 times and the averages ie; $D = \frac{1}{100}\sum_{j=1}^{100}(\sum_{t\epsilon U}\frac{k_{1t}}{4})$, where $U = \{25, 50, 65, 75\}$ were computed. The average detectability D is quite large and roughly uniform for all these values of $\sigma_\gamma/\sigma_\epsilon$ considered (See Table 1 for $\Pi_{t1} = 0.1$). This indicates that the method is not very sensitive to the value of the ratio $\sigma_\gamma/\sigma_\epsilon$, over the range considered.

In addition the mean squared error:

$$MSE = \frac{1}{100}\sum_{t=1}^{100}(i_t - k_{1t})^2$$

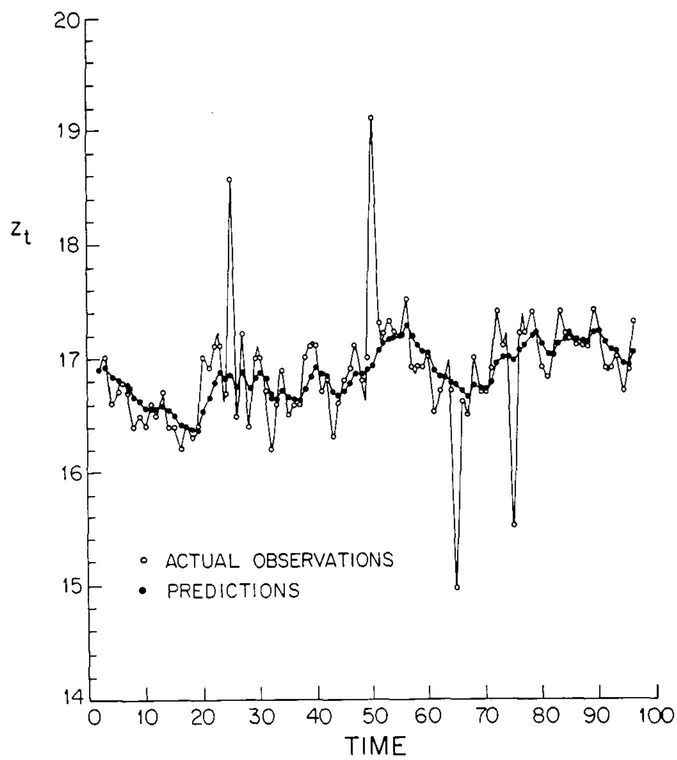

FIG 4.2 Predictions from the Robust Kalman Filter

was computed for each repetition of the simulations. The average of these over the 100 repetitions is also shown in the last column of Table 1. Although the mean squared errors decrease with $\sigma_\gamma/\sigma_\epsilon$, they are roughly of the same order of magnitude. This further confirms that the procedure is not very sensitive to changes in $\sigma_\gamma/\sigma_\epsilon$, over the range considered.

A related question that arises in connection with the scheme is how sensitive it is to $\Pi_{t1} = Pr(i_t=1)$, which in the previous simulation study was held constant at 0.1. To answer this we repeated the entire simulation for $\Pi_{t1} = 0.001$, 0.01, and 0.05 . The detectability (D) and the MSE for each case are given in Table 1. For a given Π_{t1}, D is roughly uniform for all values of $\sigma_\gamma/\sigma_\epsilon$. The table also indicates that the

Table 1: Detectability and the mean squared error.

$\sigma_\gamma/\sigma_\epsilon$	Π_{t1}	0.001	0.01	0.05	0.1
2	D	0.3175	0.4356	0.5468	0.5654
	MSE	0.0265	0.0207	0.0198	0.0215
3	D	0.3442	0.4666	0.5615	0.5716
	MSE	0.0254	0.0197	0.0194	0.0201
5	D	0.3491	0.4665	0.5538	0.5569
	MSE	0.0253	0.0199	0.0188	0.0189
10	D	0.3397	0.4465	0.5307	0.5275
	MSE	0.0258	0.0208	0.0186	0.0186

detectability increases as Π_{t1} approaches the true value. However, the MSE of detectability is very flat. For example when $\sigma_\gamma/\sigma_\epsilon = 5$, the values of the detectability for $\pi_{t1} = 0.001$, 0.01, 0.05 and 0.1 are 0.3491, 0.4665, 0.5538 and 0.5569 respectively. The corresponding MSE's are 0.0253, 0.0199, 0.0188 and 0.0189.

It should be noted that only 4 outliers were introduced in 100 observations. Even in the case when $\Pi_{t1} = 0.01$ (1 in 100) and $\Pi_{t1} = 0.001$ (1 in 1000), the detectability measure D is still quite large. It must also be noted that an outlier is considered to be any observation with the corresponding noise component coming from the $N(0\ ,\ \sigma_\gamma^2)$ distribution. Many of these may be quite small in magnitude and would not lead to noticeable outliers. This is reflected by the numbers in Table 1. When there are large outliers however, k_1 takes values close to 1, as shown in Fig. 4.3.

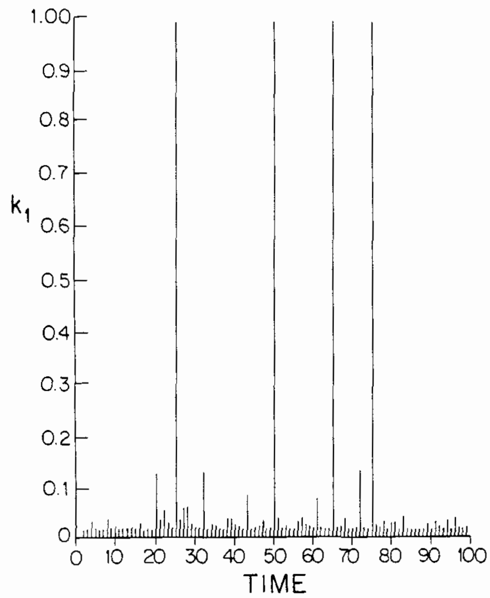

FIG 4.3 Posterior probability of an outlier at time t (π_{t1} = .1)

4.3 ARIMA (1,1,0) model with outliers

Consider the model

$$(1-\phi B)(1-B)y_t = a_t \tag{4.6}$$

where ϕ is the autoregressive parameter and $\{a_t\}$ is a white noise sequence with mean zero and variance σ_a^2. Let the measurement process be defined by

$$z_t = y_t + \epsilon_t \tag{4.7}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. Gaussian random variables with mean zero and variance σ_ϵ^2. For $\phi>0$ this model represents a very slowly drifting process. Predictions for any lead time obtained at a particular time point $t = t_0$ in this series is a linear function of z_{t_0} and z_{t_0-1}. Thus, if there is an observational outlier at t_0, it strongly influences the predictions. Thus robust Kalman Filtering and outlier detection are quite relevant in this case.

We shall now incorporate occasional outliers by replacing ϵ_t by $(1-i_t)\epsilon_t + i_t\gamma_t$, where $\{\gamma_t\}$ and $\{i_t\}$ are as defined before. A state space equivalent of this ARIMA (1,1,0) observational outlier model, can be written as

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & \phi \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} a_t \tag{4.8}$$

$$z_t = (0\ 1) \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} + (1-i_t)\epsilon_t + i_t\gamma_t$$

where x_t is a 2×1 state vector. Now, using the flow chart given in Fig. 3.1, the following equations can be obtained. Let the state at time $t-1$, x_{t-1}, be approximately normally distributed with mean $\hat{x}_{t-1}$ and covariance matrix $\hat{P}_{t-1}$.

Then, the state and covariance predictions at time t can be given by

$$\hat{x}_{t|t-1} = \begin{pmatrix} 1 & 0 \\ 1 & \phi \end{pmatrix} \begin{pmatrix} \hat{x}_{1t|t-1} \\ \hat{x}_{2t|t-1} \end{pmatrix} = (\hat{x}_{1t|t-1}\ ,\ \hat{x}_{1t|t-1} + \phi\hat{x}_{2t|t-1})'$$

$$\hat{P}_{t|t-1} = \begin{pmatrix} 1 & 0 \\ 1 & \phi \end{pmatrix} \hat{P}_{t-1} \begin{pmatrix} 1 & 1 \\ 0 & \phi \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \sigma_a^2$$

$$\hat{z}_t = (0\ \ 1)\hat{x}_{t|t-1} = \hat{x}_{1t|t-1} + \phi\hat{x}_{2t|t-1}$$

A limited simulation study was conducted using 100 observations from the model (4.8) with $\phi = 0.8$, $\sigma_a = 1$, $\sigma_\epsilon = 5$. As in the previous example the outliers were introduced at the points $t = 25$, 50, 65 and 75 using a Gaussian distribution with $\sigma_\gamma = 25$. The initial values of the states which were used to generate the series were $\hat{x}_{1,0} = 20$ and $\hat{x}_{2,0} = 150$ respectively. The ordinary Kalman filter was initially used to estimate the states and to make predictions. These predictions and the actual observations are plotted in Fig. 4.4. The aberrant observations seem to have a strong influence on the predictions. For example the outlier at the 50^{th} time point pulls the predictions away from the

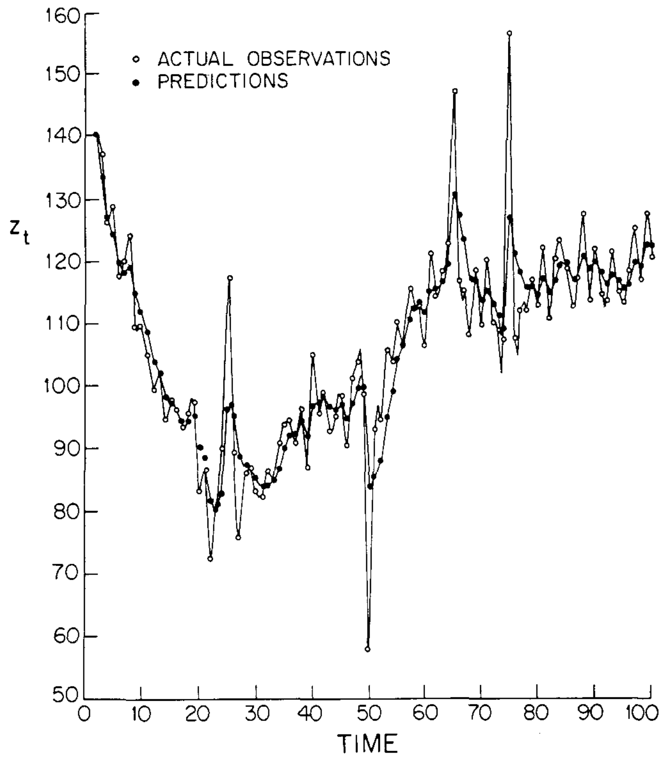

FIG 4.4 Predictions from the ordinary Kalman Filter

actual observations. This is also noticeable at the other time points where there are outlying observations. Secondly, the robust Kalman filter was run, on the assumption that $Pr(i_t = 1) = 0.1$. These predictions and the actual observations are given in Fig. 4.5. Unlike before, the predictions are not unduly influenced in this case. For example, the prediction at $t = 50$ is not very sensitive to the discrepant observation. The predictions at the subsequent points follow the observations quite closely. The probability that each point is an outlier (k_1) is plotted in Fig. 4.6. Again the four large outliers have been clearly detected.

As in the ARIMA (0,1,1) example, a further simulation study was conducted to gain some insight into the filter sensitivity towards chang-

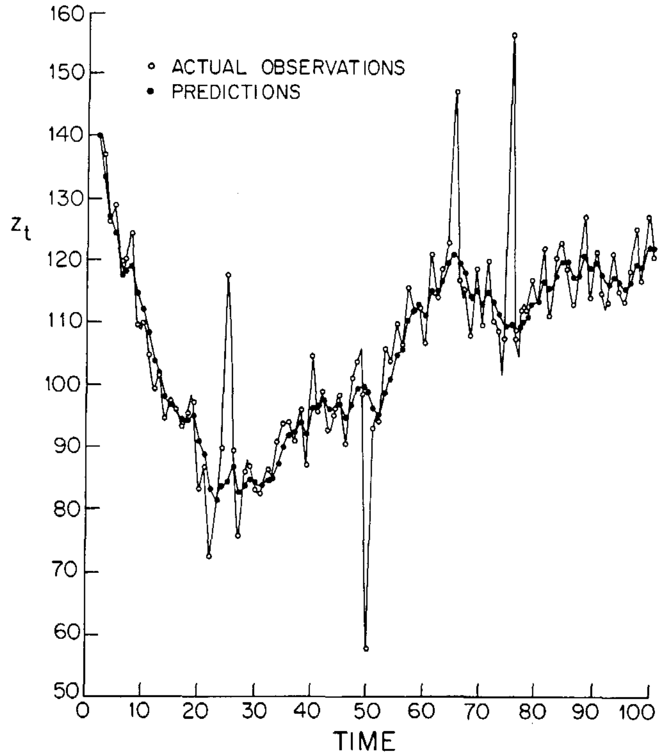

FIG 4.5 Predictions from the Robust Kalman Filter

ing $\sigma_\gamma/\sigma_\epsilon$. 100 observations were generated with the true ratio of $\sigma_\gamma/\sigma_\epsilon = 5$. Then, the detection probabilities (k_{i_t}) were computed with $\sigma_\gamma/\sigma_\epsilon = 2$, 3, 5 and 10. The detectability measure $\sum_{s\epsilon A} k_{1s}/4$ was recorded for each, where A represents the set of outliers. The whole process was repeated 100 times and the average detectability measure D, and the MSE were calculated. The results (Yatawara, 1986) were very similar to those previously shown for the ARIMA (0,1,1) example, indicating again that the robust filter is not very sensitive to moderate changes in $\sigma_\gamma/\sigma_\epsilon$.

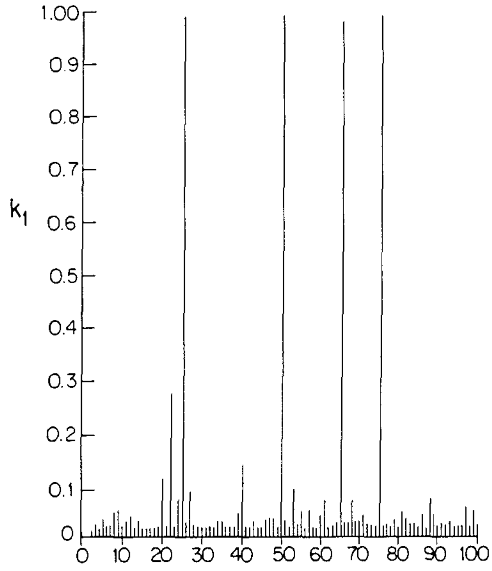

FIG 4.6 Posterior probability of an outlier at time t (π_{t1} = .1)

5. Discussion and Conclusions

Guttman and Pena (1984, 1985) derived a robust Kalman Filter using a different model for the measurement error.

In comparing the two results Yatawara (1986) showed that when, in the robust filter of this paper, the probabilities $\sum_{i_{t-1}=0}^{1} \lambda_{i_{t-1}i_t} k_{i_{t-1}}$ are fixed to a constant value, and the single state ARIMA (0,1,1) model is considered, then both filters are identical. However, when higher order models are considered, or a more general transition probability matrix is used, then the outcomes of the two filters differ. Such a difference is to be anticipated since our assumption of Gaussian transitions through a Markov Chain, and their assumption of smoothly contaminated Gaussians are quite different. However, both filters appear to give excellent protection against outliers.

The Kalman Filter is used to recursively update estimates of the states of a system using data which are collected sequentially in time. When this data is coming from on-line sensors, laboratory analyses, etc. it is not uncommon to have it contaminated by outliers or "bad" data. In these situations the Kalman Filter has been shown to perform poorly.

In this paper we have used a Bayesian approach to develop a recursive state estimator which is robust to such observational outliers. This Robust Kalman Filter reduces to the simple form of a weighted average of two parallel Kalman Filters with different gains. The posterior probability of each new observation being an outlier is also obtained, and used to weight the parallel filters.

The superior performance of this Robust Filter over the standard Kalman Filter, and the ability of the filter to clearly detect outliers was demonstrated in several simulations. The robust properties of the filter were also shown to not be seriously affected by the choice of the additional parameters, Π_{t1} and $\sigma_\gamma/\sigma_\epsilon$ that must be specified in order to use it.

7. References

Anderson, B.D.O. and Moore, J.B. (1979). *Optimal Filtering*, Prentice-Hall, N.J.

Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis Forecasting and Control* (Revised Ed.), San Francisco: Holden Day.

Duncan, D.B. and Horn, S.D. (1972). "Linear Dynamic Recursive Estimation from a Viewpoint of Regression Analysis". *Journal of the American Statistical Association, 67*, 815-821.

Guttman, I., Pena, D. (1984). "Robust Kalman Filtering and its Applications". T.R. #2766, Math. Research Center, University of Wisconsin, Madison.

Guttman, I. and Pena, D. (1985). "Comment on Robust Filtering", *Journal of the American Statistical Association, 80*, 91-92.

Harrison, P.J. and Stevens, C.F. (1976). "Bayesian Forecasting", *Journal of the Royal Statistical Society, Series B, 38*, 205-247.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering Theory*, New York: Academic Press.

Priestly, M.B. (1980). "State-dependent models: A General Approach to Time Series Analysis", *J. Time Series Analysis, 1*, 47-71.

West, M., Harrison, P.J. and Mignon, H.S. (1985). "Dynamic Generalized Linear Models and Bayesian Forecasting", *Journal of the American Statistical Association, 80*, 73-83.

Yatawara, N. (1986). Detection of Outliers and Random Events in Time Series, Unpublished Ph.D. Thesis, Department of Statistics and Actuarial Science, University of Waterloo, Canada.

Received November 1990; Revised February 1991.

Recommended by B. N. Nagarsenker, Air Force Institute of Tech., Wright Patterson AFB, OH.

Refereed Anonymously.