STATE-SPACE MODELS FOR ANALYSIS OF HYDROLOGICAL SERIES.

S.O.N AGWUEGBO and A.P. ADEWOLE***

**Department of Statistics, University of Agriculture, Abeokuta*

***Department of Computer Science, University of Agriculture, Abeokuta*

ABSTRACT

In this study, the class of state pace model for which optimal forecasts may be computed using a recursive estimation procedure called the Kalman Filter is developed for the analysis of hydrological series. The state-space formulation yields a practical means of estimation for his complex time varying dynamical process. It provided a generic and flexible framework for hydrological modeling and inference. A straight forward implementation was achieved in the software package S-Plus

1.0 INTRODUCTION

Hydrological variables such as rainfall are products of complex time-varying phenomena which can be measured by a finite number of observations. The highly non-linear and extremely sensitive process governing rainfall makes a purely deterministic physical description and forecasting impossible (Bardossy and Plate, 1991). A detailed study of hydrologic phenomena requires mathematical models that should take into account time variability. Recently, there is growing interest in non-linear models combined with a greater computational facility for describing data where the variance changes through time.

State-space models, Durbin and Koopman (2000, 2001, 2002) and Chatfield (2004) are a widely used tool in time series analysis to deal with processes which gradually change over time. The state-space model represents a physical system as n first order coupled differential equations. The state-space representation is essentially a convenient notation to estimate stochastic models in which one assumes measurement errors in the system which allows handling a large set of time series models. The recursive nature of the model and the computation techniques used for its analysis allow the direct incorporation of known breaks in the system structure over time.

Kalman (1960) estimated co-efficient of a non-linear differential equation using an optimal sequential estimation techniques often referred to as Kalman filter. Kalman's derivation took place within the context of state-space models whose core is the recursive least squares estimation. Within the state-space notation, the Kalman filter derivation rests on the assumption of normality of the initial state vector, and as well as the disturbances of the system.

The state-space formulation is useful in its ability to cast many models within the mathematical frame work of two equations; system equation and measurement equations. The linear state-space form has been demonstrated to be an extremely powerful tool in handling all linear, and many classes of non

linear time series models (Harvey, 2001). Box and Jenkins (1976) use the term non-linear estimation to describe procedures for minimizing a sum of squares function when numerical methods have to be used. Theoretically, there is no distinction between the state-space and the ARMA representations of a stationary process (Wei, 1990). Akaike (1974) expressed the state-space model parameters in terms of ARMA models. The ARMA forecast updating models can also be reformulated in the state variable form upon which the Kalman filter algorithm can be employed to obtain hydrological forecasts. It is argued that when hydrological time series model is assumed to be an auto-regressive moving average (ARMA) model, the corresponding hydrological series are considered to be free of measurement errors, and the minimum mean square error forecasts obtained by using the conditional Box and Jenkins time series forecasting methods are identical with those obtained by using the Kalman filtering technique (Chatfield 2004, Ahassan and O'connor, 1994).

It will be demonstrated in this paper that the Kalman filter technique is adopted as the best linear filter in an expected square error sense, and the filter algorithm degenerates into simpler algorithm that is identical with the conventional time series method of forecasting. A major practical advantage of the Kalman filter is that the calculations are recursive, so that, although the current estimates are based on whole past history of measurements, there is no need for an ever expanding memory (Chatfied 2004).

Kalman (1960) procedure is the most efficient category of prediction models that have an adaptive behaviour. In application of Kalman filtering theory, the mathematical formulation of the problem and the computational techniques involved may depend heavily on the computational simplicities of the system model which is used. Kalman filtering is designed to strip unwanted noise out of the stream of data (Cipra, 1993) and it addressed the question of getting accurate information out of inaccurate data.

Kalman (1960) and Kalman and Bucy (1961) papers have generated thousands of other papers on aspects and application of filtering. Their work has also simulated mathematical research in such areas as numerical methods for linear algebra (Cipra, 1993). Kalman filtering has been used for parameter optimization or sequential parameter estimation of linear rainfall- run off models as new data become available in real time (Ahsan and O'connor, 1994). Kalman filtering has been applied in non linear modeling and estimation problems related to river flow forecasting. As the original version of the Kalman filter is intended for linear estimation only, the later applications usually require prior linearization so that the linear Kalman filter can be applied. This type of applications is described as the extended Kalman filter.

2.0 LINEAR STATE SPACE MODELS

Many problems in science require estimation of the state of a system that changes over time using a sequence of noisy measurements made on the system. Systematic errors and non-homogeneities are introduced into rainfall data by changing the type or location of a rain gauge or by construction of buildings nearby or the growth of trees. These factors should be noted in the station history and the user should always be aware of this potential problem.

The prime objective of state-space modeling is to estimate the signal in the presence of noise. The state-space approach to time series modeling focused attention on the state vector of a system. The measurement vector represents noisy observations that are related to the state vector. It is assumed that the noise contaminates the signal in an additive manner so that the observations are given by the following measurement equation

$$Z_t = F_t\theta_t + v_t$$

$$v_t \sim N(0,\sigma_v^2) \qquad \text{.........................(1)}$$

where Z_t $(t = 1,2,.............N)$ is the observed noise corrupted time series, F_t is assumed to be an $((n\times 1)$ known column vector, θ_t and v_t

are the time series representing an $(n \times 1)$ state vector and the observation noise respectively. Although θ_t may not be directly observable, it is often assumed as a vector difference equation or state equation represented as

$$\theta_t = H_t \theta_{t-1} + W_t \quad \text{.......................... (2)}$$

where the $(n \times n)$ matrix H_t is assumed known, and W_t denotes an $(n \times 1)$ vector of deviations such that $W_t^T = (w_{1,t}, w_{2,t}, \ldots\ldots\ldots w_{n,t})$. In fitting the model, the state equations are permitted to wander in the form of a random walk. The model is allowed to bend so as to minimize the distance between Z_t and the prediction $\hat{Z}_t$.

The pair of equations in (1) and (2) constitute the general form of the state-space model. The errors in the measurement (or observation) equation in (1) and state (or transition) equation in (2) are generally assumed to be serially uncorrelated and also to be uncorrelated with each other at all time periods. Further, the measurement error V_t is assumed as an independent random Gaussian process while W_t is a white Gaussian noise with zero mean and variance matrix σ_w^2. Additionally V_t and W_t are assumed to be orthogonal at all pairs of time.

3.0 THE KALMAN FILTER ALGORITHM

The Kalman filter is the main algorithm to estimate dynamic systems in state-space form. Kalman (1978) defined filtering as any mathematical operation which uses past data or measurements on a given dynamical system to make more accurate statement about present, future or past variables in that system. Kalman has based the construction of the filter in probabilistic theory, more specifically, on the conditionally Gaussian properties of random variables. For the linear Gaussian estimation problem, the required probability density function (pdf) remains Gaussian at every iteration of the filter, and the Kalman filter relations propagate and update the mean and covariance of the distribution(Chatfield, 2004).

The Kalman filter recursively evaluates the estimator of the state vector conditional on the past observations up to time $(t-1)$. By considering Equation (2), where W_t is still unknown at time $t-1$, the obvious estimator for θ_t is given as

$$\theta_{t/t-1} = H_t \hat{\theta}_{t-1} \quad \text{.................................. (3)}$$

with variance covariance matrix

$$P_{t/t-1} = H_t P_{t-1} H_t^T + W_t \quad \text{.......................... (4)}$$

Equations (3) and (4) are the prediction equations. Equation (4) follows from standard results on variance -covariance matrices for vector random variables (Chatfield, 2004; Stark and Woods, 1986). When new observation has been observed, the estimator for θ_t can be modified to take account of this extra information. At time $(t-1)$, the best forecast of Z_t is given as $F_t \hat{\theta}_{t/t-1}$ so that the prediction error is given by

$$\varepsilon_t = Z_t - F_t \hat{\theta}_{t/t-1} \quad \text{.............................. (5)}$$

ε_t in (5) is called the prediction error. This quantity can be used to update the estimate of θ_t and of its variance-covariance matrix and the best way to do this is by means of the following equation

$$\hat{\theta}_t = \hat{\theta}_{t/t-1} + K_t \varepsilon_t \quad \text{.............................. (6)}$$

And

$$P_t = P_{t/t-1} - K_t F_t P_{t/t-1} \quad \text{.............................. (7)}$$

Where

$$K_t = P_{t/t-1} F_t [F_t P_{t/t-1} F_t^T + \sigma_v^2]^{-1} \quad \text{.............. (8)}$$

K_t in (8) is called the Kalman gain matrix and is a vector of size $(m \times 1)$. Equations (6) and (7) constitute the second updating stage of the Kalman filter and are called the updating equations. The solution is optimal provided the filter combines all observed information and previous knowledge about the system's behaviour such that the state estimation minimizes the statistical error.

4.0 APPLICATIONS

The study applied the model to monthly rainfall at Abeokuta, Ogun State between 1995-2004 collected from the Metrological Service, Federal Ministry of Aviation, Abeokuta. The plot of the monthly rainfall as in Figure 1 indicates that the series is stationary. The first step in state-space modeling is to find an optimal AR model that fits the data. Table 1 is the ACF and the PACF of the monthly rainfall and the correlogram is as in Figures 2 and 3. Based on the ACF and PACF of the monthly rainfall in Table 1, one may suggest an AR (1). The Splus package use the Akaike Information Criterion (AIC) to provide an optimal or best fit for the autoregressive model. The value of the AIC for the monthly rainfall data is as in Table 2. The AIC is minimum at p=1. Hence, the optimal AR order p is chosen to be one. Tables 1 and 2 indicate the behaviour of the monthly rainfall data.

Table 1: Sample ACF and PACF of the Monthly Rainfall at Abeokuta.

Lag	1	2	3	4	5	6	7	8	9	10
ACF	0.421	0.154	-0.023	-0.260	-0.450	-0.537	-0.428	-0.202	0.018	0.175
:PACF	0.421	-0.028	-0.095	-0.261	-0.311	-0.339	-0.235	-0.127	-0.127	-0.147

Lag	11	12	13	14	15	16	17	18	9	20
ACF	0.436	0.486	0.473	0.224	-0.023	-0.291	-0.417	-0.448	-0.376	-0.235
:PACF	0.095	0.056	0.219	0.035	0.032	-0.126	-0.028	-0.023	0.003	-0.105

Table 2: AIC of the Monthly Rainfall at Abeokuta.

Lag	1	2	3	4	5	6	7	8	9	10
AIC	0.000	1.200	2.562	4.287	5.890	7.356	9.298	11.219	12.685	14.630

The Gauss Markov signal model generated from the rainfall data using ARMA model is

$$\theta_t = 0.274\theta_{t-1} + W_t \qquad t \geq 0$$

with mean equal to zero and $\sigma_w^2 = 0.0078$. The Kalman gain K_t as defined in (8) is $K_t = 0.082$. The prediction error variance as defined in (5) is $\varepsilon_t = 0.089$ The Kalman filter is asymptotically given by

$$\hat{\theta}_{t/t} = 0.252\hat{\theta}_{t-1/t-1} + 0.082Z_t$$

5.0 CONCLUSION

The Kalman filter revisited the estimates by adding a correction to the preliminary estimate. The magnitude of the correction is determined by how well the preliminary estimate predicted the new observation and this will tend towards a steady state when the Kalman gain approaches a limit. The Kalman filter uses the least squares method to generate recursively a state estimator at time t, which is linear, unbiased and with minimum variance. The filter is in line with the Gauss-Markov theorem and this gives the Kalman filter a great power to solve a large range of statistical inference problems.

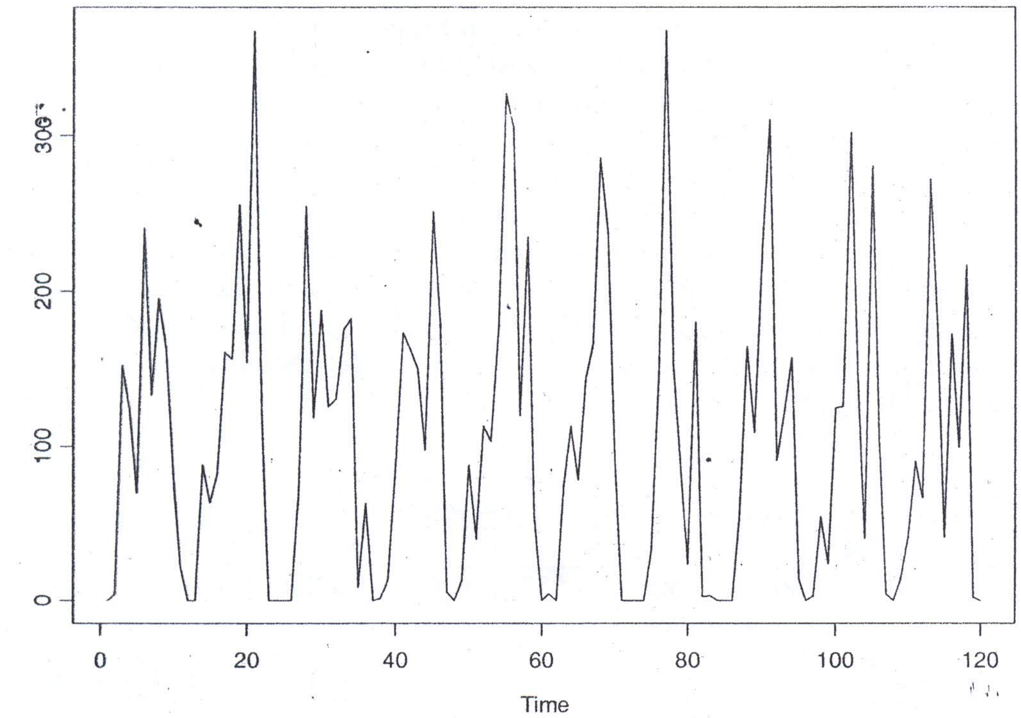

Figure1: Monthly Rainfall data at Abeokuta

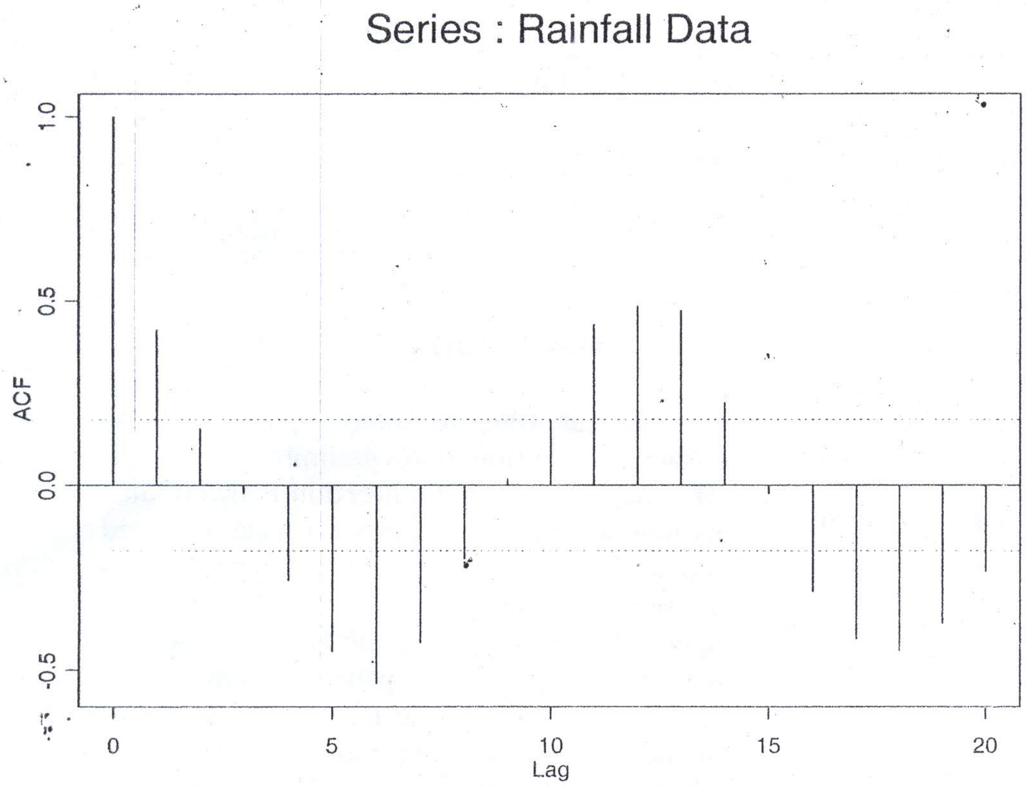

Figure 2: Sample ACF for the Monthly Rainfall Data

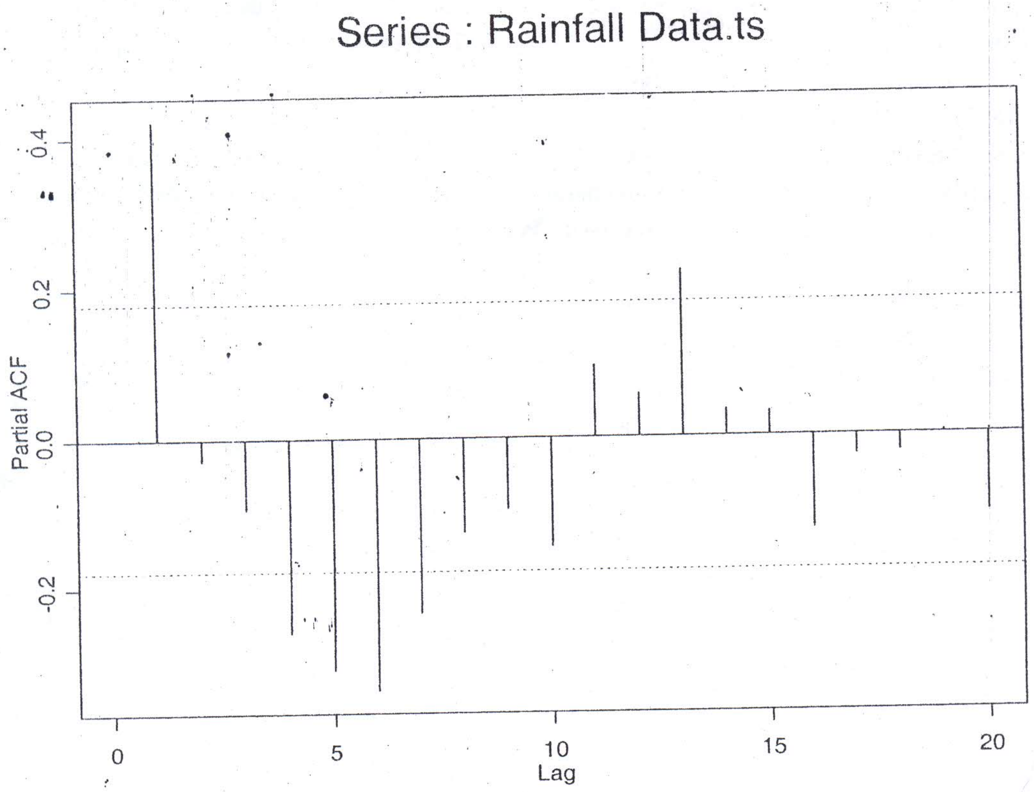

Figure 3: Sample PACF for the Monthly Rainfall Data

REFERENCES

Akaike, H (1974), "A new look at the statistical model identification". IEEE Trans. Automat. Control 19 716-723.

Ahassan, M., and O'connor, K. M (1994), "A reappraisal of the Kalman filtering technique, as applied in river flow forecasting". Journal of Hydrology, 161: 197-226.

Bardossy, A. and Plate, E. J. (1991), "Modeling daily rainfall using a semi-markov representation of circulation pattern occurrence". Journal of Hydrology, 122:33-47.

Box, G.E.P., and Jenkins, G.M (1976), "Time series analysis, forecasting and control" (rev.ed), San Franscisco, Holden-Day.

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C (1994), "Time series analysis, forecasting and control." 3rd edn. Englewood Cliffs, N.J, Prentice-Hall.

Cipra, B.A. (1993). "Engineers look to Kalman filtering for guidance." Reprinted from Society for Industrial and Applied Mathematics, Vol 26-5.

Chatfield, C. (2004), "The analysis of time series, an introduction". 6th edn, Chapman and Hall/CRC, Boca Raton.

Durbin, J. and Koopman, S. (2000), "Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives". Journal of the Royal Statistical Society, Series B 62(1), 3-56.

Durbin, J. and Koopman, S. (2002), "A simple and efficient simulation smoother for state space time series analysis". Biometrika 89,603-616.

Durbin, J. and Koopman, S. (2002), "Time Series Analysis by State Space Methods". Oxford University Press.

Harvey, A.C. (2001), "Forecasting, structural time series models and the Kalman filter". Cambridge Univ. Press, Cambridge.

Kalman, R.E (1960), "A new approach to linear filtering and prediction problems". Journal of Basic Eng, Vol.82, pp. 35-45.

Kalman R.E., and Bucy, R.S (1961), "New results in linear filtering and prediction". Theory, Journal of Basic Eng,Vol 83, pp. 95-107

Wei, W.W.S (1990), "Time Series Analysis: Univariate and Multivariate Methods". Addison-Wesley Redwood City, CA