Nigerian Journal of Technology (NIJOTECH)
Vol. 32, No. 1, March, 2013, pp. 81–92.

ISSN 1115-8443

MINIMUM VARIANCE ESTIMATION OF YIELD PARAMETERS OF RUBBER TREE WITH KALMAN FILTER

A.C. Igboanugo[a], U.E. Uche[b]

Department of Production Engineering, Faculty of Engineering, University of Benin, Benin-City, Nigeria.
Emails: *[a]dracigboanugo@yahoo.com; [b]emmanuel.uche@rubberestates.com*

Abstract

Although growth and yield data are available in rubber plantations in Nigeria for aggregate rubber production planning, existing models poorly estimate the yield per rubber tree for the incoming year. Kalman filter, a flexible statistical estimator, is used to combine the inexact prediction of the rubber production with an equally inexact rubber yield, tree girth, tapping height, stimulation and tapping system measurements to obtain an optimal estimate of one year ahead rubber production. Six rubber clones-GT1, PB260, PB217, PB28/59, PB324 and RRIM703 were studied using 12-year data, generated from permanent experimental plots. Stochastic autoregressive model was fitted to the data to identify optimal management strategy that accounts for risk due to seasonality. STAMP, an OxMetric modular software system for time series analysis, was used to estimate the yield parameters. Our results show that significant test of actual yield to model forecast is less than 1.96. Hence, the null hypothesis that the actual yield is within the forecasted value is accepted at 5% significant level. Based on the impulse response function of the lead equations, the long-run elasticity of yield was estimated to be highest for PB324 (2211gm/tree) and lowest for RRIM703 (1053gm/tree). PB260 is the best short term clone with the highest dynamic multiplier of 0.59. More important, the estimator minimized the variance of estimation errors from 55% of plantation prevision to 10%. It is our opinion that Kalman filter is a robust estimator of the biotechnical dynamics of rubber exploitation system.

Keywords: Kalman filter, parameter estimation, rubber clones, Chow failure test, autocorrelation, STAMP, data characterization

1. Introduction

Obtaining reliable information for the predictive yield of rubber clones that is needed for aggregate rubber production planning and for the determination of the long run elasticity of yield, required for rubber clone selection from historical yield and growth data, has remained a great challenge for rubber plantation managers: the information emanating therein is found to be replete with imprecision for purposes of planning due to substantial error in measurements, heterogeneity of plants population in age and size, non independence of observations and multiple observations with serial correlation. The current prediction models which include use of mean [1], ratio [2] and the linear regression predictors [3] are considered rather inappropriate because several characteristics of the modelling data violate statistical assumptions underlying regression techniques. Further, the objective of forecasting is to minimize uncertainty, to identify and evaluate risks. As it stands today, most yield projections do not provide estimates of the variance of the predictions which is crucial in updating procedure, to assess the precision of predictions, to evaluate the general performance of the models and to calculate confidence intervals for the estimates. The problem therefore goes beyond estimation of conditional means, variances and covariance based on the properties of multivariate normal distribution because of the serial correlation between the observations nor is it just estimating model parameter values from a set of field data. It demands the use of field data and information about growth and development to estimate model parameters.

2. Review of Literature

In this paper, the objective is to modify the population model to make a better prediction for a specific agricultural field not for a general situation. A filter is an algorithm that is applied to a time series to improve it in some way- filtering out noise. The task of filtering is to eliminate by some means as much of the noise as possible through processing of the measurements. This task is achieved by using measured values to update the model state variables each time an observation is available. Hence, Kalman filter is not just a linear estimator where the slope (A) is a fixed matrix and the intercept (b) a fixed vector but it is a linear minimum variance estimator where the slope (A^0) and the intercept (b^0) in the prediction equation are chosen to minimize the expected mean square error such that[4]:

$$E\| X - A^0Y - b^0 \|^2 \leq E\| X - AY - b \|^2 \forall A, b \quad (1)$$

The basic key to this process is the structural time series state space model in which the state of the time series is represented by its various unobserved components such as trend and seasonality.

In using Kalman filter as an estimator, one of the independent estimates is a current estimate or monitoring measurement and the other is a previous estimate that is updated for expected changes overtime using a prediction model (Fig. 1 refers.) The variance for the updated estimate includes the effects of the errors in the previous inventory that are propagated over time and the model prediction errors between previous and current estimates. Errors in the composite estimator are typically less than errors in either of the prior estimates alone [5]. The drawback of the Kalman filter is that it assumes that the model estimates are unbiased [6].The simple Kalman filter is therefore an optimal estimator, only if the system is linear and reasonable methods are available to accurately determine the covariances P, Q, and R [7]. Further, when the normality assumption is dropped there is no longer any guarantee that the Kalman filter will give the conditional mean of the state vector though it remains an optimal estimator because it minimizes the mean square error within the class of all linear estimators [8].

The rubber tree yield model is treated as a time series in state-space form. The structure of the general state space models depends on the assumptions that are used to specify the state space vector components. In contrast to ARMA models, the specification of structural time series model requires judgements about the structures that are present in the sequence of observations. Following the presentation in [8] and [9], structural time series models consists of components that capture (deterministic or stochastic) trend, seasonality, and random error. Using μ_t and β_t to denote the trend and slope at time t, the relation

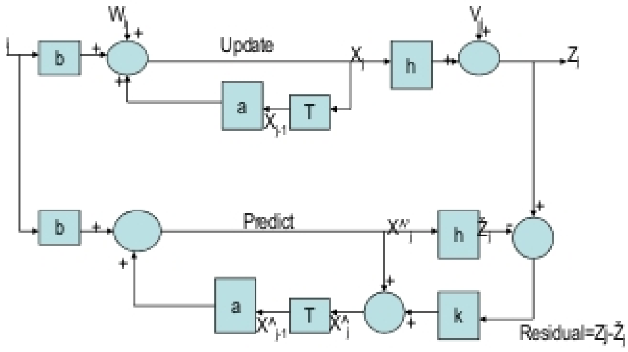

Figure 1: Kalman filter process.

between trend and slope is described as follows:
Trend:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad (2)$$

Slope:

$$\beta_t = \beta_{t-1} + \zeta_t \quad (3)$$

Seasonality:

$$\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t \quad (4)$$

System equation:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \mu_t + \gamma_t + \varepsilon_t, \quad t = p+1 \ldots T \quad (5)$$

Substituting:

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + (\mu_{t-1} + \beta_{t-1} + \eta_t) + (\omega_t - \gamma_{t-1} - \gamma_{t-2} - \ldots - \gamma_{t-s+1}) + \varepsilon_t$$

Measurement equation

$$z_t^i = \lambda_i x_t + \xi_t^i \quad i = 1, 2, \ldots, k, \; t = 1, 2, \ldots, T \quad (6)$$

The equation is written in the form of the general state-space model with ε_t, η_t, ζ_t assumed independent, normally distributed random variable with zero mean and finite variance. Thus the parameters to be estimated in structural time series formulations are: $\phi \ldots \phi_p$, σ_ε^2, σ_ε^2, σ_η^2, σ_ζ^2, σ_ω^2, $\lambda_1, \ldots, \lambda_k$, Σ_ξ.

The system equation in matrix form is as shown below.

$$X_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \cdot \\ x_{t-p+1} \\ \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \vdots \\ \gamma_{t-s+2} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p & 1 & 1 & -1 & -1 & \cdots & -1 \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ & & & & 1 & 1 & & & & \\ & & & & 0 & 1 & & & & \\ & & & & & & -1 & -1 & \cdots & -1 \\ & & & & & & 1 & 0 & \cdots & 0 \\ & & & & & & 0 & 1 & \cdots & 0 \\ & & & & & & \vdots & \vdots & \vdots & \vdots \\ & & & & & & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \cdot \\ x_{t-p} \\ \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \gamma_{t-3} \\ \vdots \\ \gamma_{t-s+1} \end{bmatrix} + \begin{bmatrix} \eta_t + \omega_t + \varepsilon_t \\ 0 \\ \cdot \\ 0 \\ \eta_t \\ \varsigma_t \\ \omega_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (7)$$

where: X_t; $d \times 1$ vector used to represent the yield state of the clone at the start of period t, Z_t; $k \times 1$ vector used to represent the current yield, $\varnothing$: AR coefficients, ε_t: random error terms, ω: seasonal random error, λ: measurement transition matrix, η: trend error or irregularity, t: the lag.

The unknown variances and the components are estimated by casting the model into state space form and employing the Kalman filter in which the variances are obtained using the expectation maximization (EM) algorithm. The EM algorithm is a recursive method of obtaining maximum likelihood estimates of the unknown elements of the state and error system matrices of the state space form. Estimates of the component are then calculated using a smoothing procedure which first estimates the components recursively for $t = 1, \ldots, T$ and then runs a second "backward" recursion over $t = T, , 1$ to obtain the smoothed estimates[10]. Full details of the estimation procedure may be found in [11]. Autocorrelation and partial autocorrelations functions are traditionally used for selecting between ARMA models but does not apply to time series models in state-space form [12].

3. Material and Method

Growth data were obtained from a trial plot located at Osse River, Ovie South West Local Government Area of Edo State, Nigeria, established since 1990 for the purpose of monitoring and analyzing the dynamics of rubber clones of rubber plantation belonging to Rubber Estate Nigeria Limited (RENL). The site is planted with different standard rubber clones for long term periodic observation. The rubber clones are GT1, PB260, PB217, PB28/59, PB324 and RRIM703. Data are collected on budding and planting success, girth, resistance to disease, evolution of tapping density and yield per clone and by replicate. Diameter measurements are made at 100cm height on 40 trees per clone per Fisher plots biannually.

Once a clone is of the right girth (50cm at 1m) and tapping density (40%) the clone is opened for tapping. The trunk circumference is thereafter measured using a measuring tape at 1.7 metre height previously marked with white paint. The production is weighed in replicate and expressed in gram per tree (g/t) and in kilogram per hectare (kg/ha). The trial has been opened since 1998 and tapped on S2D4 6d/7 tapping intensity. In other words, the tapping system used is the half spiral cut (S2) from tapping year one to nine and alternate quarter cut (S4) and half spiral cut thereon at fourth daily frequency (D4), six days tapping, followed by one day rest(6D/7). Hence trees are tapped 6 times per month. Ethrel 10% is routinely used to enhance production- 1 gram of stimulant is applied on 1 cm band of the tapping panel after dilution with water to 2.5, 3.3 or 5% concentrations depending on the clone. The stimulation frequencies are varied depending on the clone and the year of tapping. A fairly standard and consistent tapping policy was followed over the period under study. The aforementioned clones were used to determine yield response to girth, tapping height, tapping system, concentration and rounds of stimulation. The yield data used consist of daily data averaged per tree in gram and summed by quarter. Total quarterly dry rubber production per year provides adequate data for yield parameter estimation and for forecasting the seasonal breakdown of the incoming year rubber production per tree using Kalman filter. The one year-ahead estimate is used to prepare the aggregate rubber production plan for the plantation.

Although there are many ways of estimating trends, the focus here is on signal extraction by filtering using the crop record itself. Prediction of one-step-ahead yield per tree is within the frame work of Gauss-Markov Discrete-Time Kalman filter model where the underlying system equations are difference equations rather than differential equations. The impetus for the choice of discrete-time systems is that tree growth and yield observations are made and used at discrete time instants. The general framework deployed is to de-

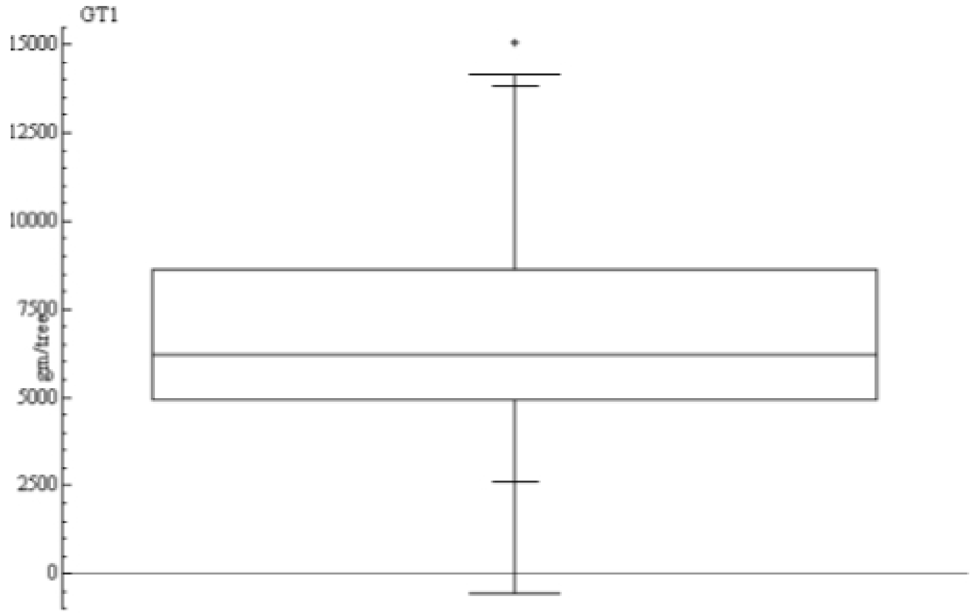

Figure 2: Box Plot of GT1 rubber yield data.

compose the series of 48 observed rubber yield values into unobserved trend and noise components. This is to facilitate the description of the series in terms of its component of interest such as the seasonal behaviour of yield and the trend movement. The predictive capacity of the model is also enhanced by the decomposition. The stochastic trend model used here is the smooth trend in which the trend and its slope component evolve as random walk. The slope innovation and the noise component are assumed to be independent, zero mean white noises with constant variances.

The yield data analysis therefore involved selecting a model and fitting it to the generated data so as to estimate the parameters using Structural Time Series Analyser, Modeller and Predictor–STAMP software, which executes the process as described earlier. Further, the normal probability plots of the residual are obtained to support the result and assumptions on error terms. Finally, Kalman filter is applied to minimize the noise associated with the forecast of the time series model. In selecting between deterministic (risk free) and stochastic model, Akaikes Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to measure the trade-off between prediction error and number of explanatory variables with smaller values being preferable:

$$AIC = \frac{-2 \times \log lik + 2 \times (n + d)}{T} \quad (8)$$

$$BIC = -2 \times \log lik + \log(T) \times n \quad (9)$$

where $\log lik$ is the logarithm of the model likelihood, n is the number of estimated parameters and T is number of data.

The estimated parameters were subjected to significance analysis to decide which explanatory variable is important in determining the dynamic of the rubber production. Backward elimination was used to refine the model by removing the statistically insignificant parameters from the model specification. Further, the selected model was tested using prediction error variance (p.e.v.) which is the basic measure of goodness of fit in time series model in the form of mean deviation and Ljung-Box statistic, the Q statistics that measures lack of fit. In addition to this goodness of fit test, models are further diagnosed by generating in-sample and out-of sample predictions. In-sample predictions were made from $t = d + 1$ to $t = T$ and compared with the observations thereby studying the models capability to reproduce observations and to capture the underlying mechanism generating the data sequence. Out of sample prediction start from $t = T + 1$ and the length of prediction is arbitrary. This is the most powerful test for a time series model. Post sample observations are used as a yardstick by which to judge the forecasting accuracy of the model estimated on the basis of the sample observations. After selecting the model, diagnosis check is carried out by examining the residual i.e. the difference between one step predictions and observations. The residual analysis is to check that serial correlations of the residuals are independent and identically distributed as initially assumed. Finally, is a check on the convergence of iterative procedure because when the number of parameters is high relative to the number of available data the algorithm fail to converge to the true optimum indicating an intrinsic problem of

Table 1: 12-Year yield data of six RENL rubber clones (gm/tree).

Year	Age	GT1	PB217	PB260	PB28/59	PB324	RRIM703
1998(1)	7	2605.4	3430.5	3229.4	3598.5	2278.5	2573.5
1998(2)	7	3001.0	4985.4	4426.1	5150.7	5070.6	4368.1
1998(3)	7	6214.9	6866.3	7560.1	9191.4	8695.5	6559.3
1998(4)	7	4606.6	6086.4	6066.0	8576.7	5486.4	5425.0
1999(1)	8	2684.4	3081.4	3476.8	2658.1	2333.1	3294.2
1999(2)	8	5722.8	6353.1	7141.5	8150.9	7085.2	6680.8
1999(3)	8	6228.1	6341.3	8045.7	9276.5	7710.0	7412.2
1999(4)	8	8671.5	8978.1	10795.4	10649.6	10677.6	11640.8
2000(1)	9	5141.7	4908.5	6363.0	5185.6	4417.6	4997.9
2000(2)	9	4165.8	3930.7	7319.0	5908.2	5031.5	4533.7
2000(3)	9	8618.7	6970.2	11167.0	9505.2	10073.5	9559.0
2000(4)	9	10326.6	9251.2	11952.9	12117.7	9526.6	12370.0
2001(1)	10	2994.9	3345.8	3408.4	2844.3	3067.2	2716.5
2001(2)	10	4446.7	5177.7	4973.3	5153.4	5934.6	5842.0
2001(3)	10	4860.0	5542.6	5389.3	4538.6	6448.5	6490.6
2001(4)	10	4987.3	5567.7	4884.5	4658.1	5855.6	5905.4
2002(1)	11	4259.1	4011.4	3767.6	2956.8	3189.2	3416.2
2002(2)	11	10500.4	10489.2	10949.4	10948.8	11099.7	10043.9
2002(3)	11	10530.5	10092.9	13587.2	11111.2	9889.6	11297.8
2002(4)	11	9472.9	10916.2	11299.9	10559.8	11110.3	11151.7
2003(1)	12	3551.0	4348.0	3234.2	3000.5	4088.2	3236.7
2003(2)	12	4948.9	6364.3	4742.2	4390.7	7679.6	6050.1
2003(3)	12	6716.7	8409.9	5854.7	5176.6	9513.7	7807.7
2003(4)	12	6496.1	8902.6	6117.5	5832.0	9265.8	7809.6
2004(1)	13	5251.7	6153.0	5446.8	4885.0	5341.0	5062.1
2004(2)	13	5073.5	6188.9	5240.3	5192.9	6470.1	5669.0
2004(3)	13	6232.0	6705.9	6128.6	5707.6	6978.9	7101.6
2004(4)	13	7548.9	8134.5	7867.7	7408.0	7771.5	7919.9
2005(1)	14	6165.2	8773.1	5783.5	5355.9	6622.1	6072.8
2005(2)	14	6097.1	6880.3	5205.3	5316.8	7719.1	5978.1
2005(3)	14	7452.3	10052.5	7291.9	7211.7	9158.7	7530.9
2005(4)	14	7084.7	10430.6	7432.0	7983.1	9281.3	7628.8
2006(1)	15	4186.2	5896.1	3380.0	4107.3	5102.0	4039.2
2006(2)	15	6738.2	9024.0	6819.4	6389.5	9906.7	6848.4
2006(3)	15	8384.9	9205.9	8111.2	8203.4	9734.1	8293.7
2006(4)	15	8625.5	11160.7	8306.6	8636.5	9459.4	8127.1
2007(1)	16	5229.6	7642.2	6755.5	4685.9	5687.8	4982.9
2007(2)	16	7011.6	7749.7	7931.3	6415.2	7656.5	5781.2
2007(3)	16	10375.4	10231.0	12209.9	9831.8	10090.1	10566.5
2007(4)	16	11644.7	11408.5	13136.5	10390.4	10381.5	9529.7
2008(1)	17	2801.4	6863.7	3420.5	2938.6	4324.1	2981.9
2008(2)	17	6665.1	9191.3	7671.2	6667.4	8554.4	7104.5
2008(3)	17	11201.9	14020.3	11332.5	9075.2	13065.3	10327.8
2008(4)	17	8523.0	13240.7	9853.6	9090.6	10609.3	9350.4
2009(1)	18	5049.5	5805.1	3583.3	3483.3	3994.0	3686.5
2009(2)	18	9951.1	10792.9	8230.0	7203.7	11042.9	6443.1
2009(3)	18	13833.1	14865.2	10544.9	9146.9	14575.1	8864.7
2009(4)	18	15078.5	17777.3	11791.0	10337.7	16609.8	10932.8

matching the model to data. Hence solutions should be viewed with caution in that case.

4. Results and Discussion

4.1. Data characterization

Table 1 depicts the quarterly rubber production per tree from 1998 to 2009 for the six rubber clones under study. A box plot for the characterization of the GT1 rubber production data is shown in Figure 2, while Table 2 shows the result of data characterization of all the clones. Two clones GT1 and PB217 have outliers due to change in stimulation from 2.5 to 5% for 2009 physiological year. The interquartile range of the rubber clones sample data ranges from 3708g for RRIM703 to 4525g for PB324 rubber clones. This observation reflects the variability in the rubber clone yield response to seasonal changes. In other words, PB324 is more sensitive to season than RRIM703.

4.2. Plot inspection

Figure 3 shows the actual plot of GT1 rubber clone yield per tree from January 1998 to December 2009. The graph shows a seasonal pattern of yield usually associated with tropical climate of dual seasons- dry and wet season within a year. The overall trend of the series is constant over the years. Its salient characteristics are a trend, which represents the long-run movements in the series, a seasonal pattern which repeats itself more or less every year and the irregular components which reflects non-systematic movement in the series. The choice of model is therefore informed by the decomposition of the series into its components as shown in Figure 4. The model to be considered initially is therefore the basic structural time series model (BSM) without cycle which is given by:

$$y_t = \mu_t + \gamma_t + \varepsilon_t \tag{10}$$

where μ_t is the local level component modelled as the random walk $\mu_{t+1} = \mu_t + \xi_t$, y_t is the trigonometric seasonal components and ε_t is a disturbance term with mean zero and variance $\sigma^2\varepsilon$.

As displayed in Figure 4, the seasonal effect hardly changes and is therefore considered fixed. The estimated level does pick up the underlying movement of the series and the estimated irregular is alright.

Structural time series model that separately models all the series components -seasonality, trend and explanatory variables was therefore implemented within the framework of state-space representation. The convergence of the yield model at steady state is strong for all clones.

4.3. Comparism of rubber yield models - deterministic Vs stochastic models

In Table 3, two models are compared - deterministic and stochastic structural time series models for purposes of model specification. Goodness of fit parameters and prediction criteria used are shown in the table. From the table, the lead model is stochastic-lower Akaike information criterion (AIC), lower Q-stat, higher Rs^2 and better normality.

4.4. Parameter estimation

The result of the optimization algorithm in Table 4 is obtained from quarterly rubber yield per tree and the explanatory variable data set of the six RENL rubber clones under study as processed by STAMP. The series is modelled as a linear combination of trend, seasonal, irregular, autoregression and explanatory variables (11).

$$\begin{aligned} Y = \text{Trend} + \text{Seasonal} + \text{Irregular} + \text{AR}(1) \\ + \text{Explanatory variables} \end{aligned} \tag{11}$$

Table 2: Box plot data characterization.

Clones	Minimum	Maximum	1st Ql*	Median	3rd Qu*	IQR*	LW*	UW*	Outlier
GT1	2605	15078	4882	6230	8623	3741	-730	14235	15078
PB217	3081	17777	5828	7406	10083	4255	-555	16466	17777
PB260	3229	13587	5031	6980	99467	4436	-1623	16121	-
PB28/59	2658	12118	4736	6402	9133	4397	-1860	15729	-
PB324	2279	16610	5377	7715	9902	4525	-1411	16690	-
RRIM703	2574	12370	5014	6620	8722	3708	-548	14284	-

*QL = Lower quartile, Qu = Upper quartile, IQR = Interquartile range, LW = Lower whisker, UW = Upper whisker

Table 3: Comparism of rubber yield deterministic and stochastic models.

Clones	Pev-D	Pev-S	Nor-D	Nor-S	H(12)-D	H(12)-S	Q-D	Q-S	Rs^2-D	Rs^2-S	AIC-D	AIC-S	Pev/ Md-D	Pev/ Md-S
GT1	1.91	1.87	2.84	1.75	0.98	1.27	7.7	5.6	0.61	0.62	14.93	14.9	1.09	1.14
PB217	2.02	1.97	6.75	4.4	0.74	0.81	6.3	7.4	0.57	0.58	14.97	14.95	1.04	1.01
PB260	2.68	2.45	6.78	2.73	0.39	0.67	25.2	16.61	0.54	0.58	15.26	15.17	1.01	1.17
PB28/59	1.96	1.99	5.84	7.06	0.14	0.19	15.3	18.1	0.59	0.59	14.95	14.96	1.18	1.38
PB324	1.93	1.88	1.56	6.01	0.67	0.60	10.3	8.33	0.62	0.62	14.93	14.91	1.35	1.11
R703	2.03	2.27	3.27	3.11	0.23	0.36	12.6	9.0	0.59	0.54	14.98	15.09	1.12	1.08

Table 4: Parameter estimation.

Clone	Seasonal $\sigma^2\omega$	Irregular $\sigma^2\varepsilon$	Level $\sigma^2\eta$	Slope $\sigma^2\xi$	ϕ_1 AR(1)	Lead Indicator - Ht	Lead Indicator - Gth	Lead Indicator Stim. Conc.	Lead Indicator Stim. Round
GT1	0	115293	11659	0	0.39	25**		931*	
PB217	0	1168	0	0	0.30		-553*	687**	
PB260	0	0	0	2735	0.59	28**			531**
PB28/59	0	0	0	1161	0.51	24**	-856***		
PB324	0	0	0	0	0.32			1570***	
RRIM703	0	609	0	6078	0.33		-737**		

***0.01, **0.05 and *0.1 significant level

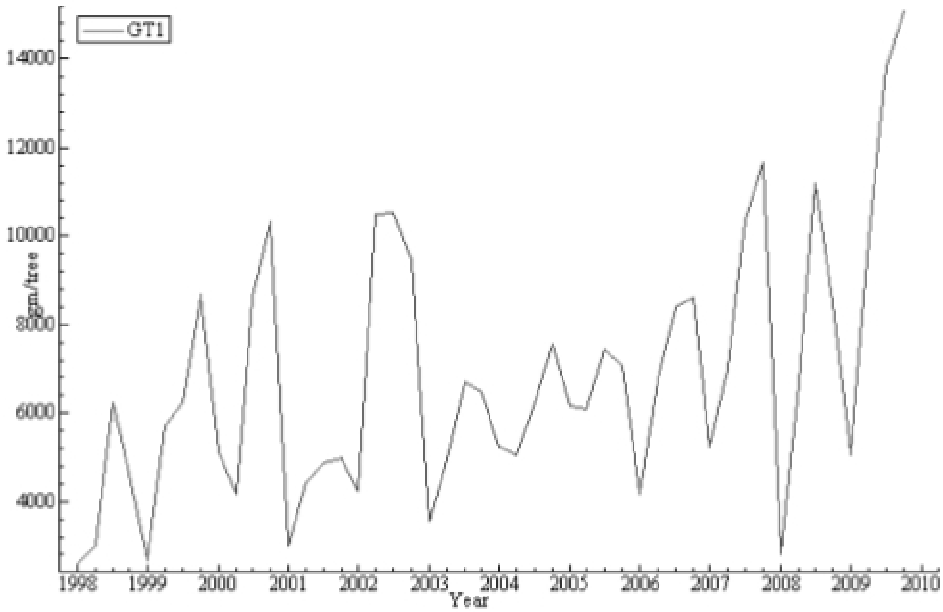

Figure 3: GT1 g/tree quarterly actual Series Plot.

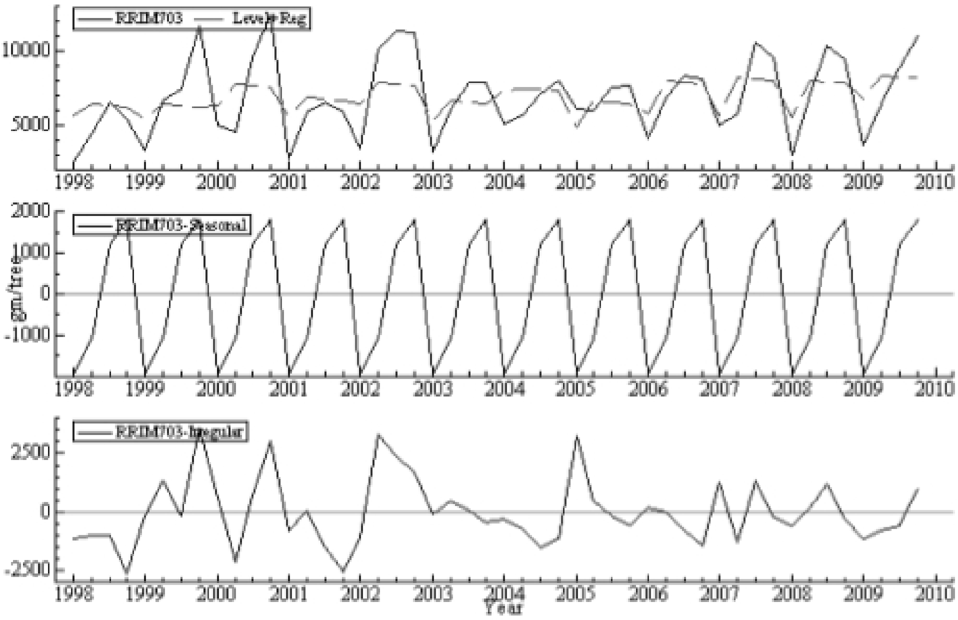

Figure 4: RRIM 703 gm/tree decomposition of actual series (black) into level (broken), seasonal and irregular.

The model parameters are estimated by maximum likelihood (exact score). Autoregression and explanatory variable coefficients estimated by STAMP are as shown in Table 4. The level of significance of each explanatory variable is depicted by *, ** and *** representing 0.1, 0.05 and 0.01 significant level of the coefficient being zero i.e. of Tcritical > Tcalculated - accepting the null hypothesis of no significant contribution in the model. Hence any of the coefficients with probability below any of these figures is considered significant and forms part of the lead equation. Tapping system is found not to be important at all three levels of significance within the limits of analysed data set and period and therefore eliminated from further consideration as yield explanatory variable. Rubber yield and marginal girth increment are inversely related.

Multivariable models are used to improve the accuracy of predictions [13].The coefficient of the explanatory variables: - girth, tapping cut height, stimulation and stimulation levels, tapping system, depicts the variables direct impact on yield per unit change in its value. It is worth noting that the estimated variance of disturbance for all the clones is zero except for GT1and PB217. This implies that the departure of the estimated trend from straight line as shown in Figure 4 (level) is due to the effect of the explanatory variables resulting from management exploitation policies on rubber tree stimulation, tapping cut height, clone selection, tapping system etc. The model estimates of negative girth relation to yield is supported by the various literature reviewed earlier which have it that rubber production is in direct competition for metabolites with the biomass of the rubber tree [14] and [15]. This observation underscores the ongoing trial in Nigeria on inorganic fertilizer supplemental application for rubber tree in tapping entitled “Interaction between Fertilization Levels of Rubber Tree Mature Crops and Yield Potential” - IFC-CIRAD-RENL Trial. Literature also supports the inference on tapping cut height as it relates to increasing magnesium content of rubber tree with height for increase in latex production [16]. Seasonality that renders the rubber business risky and therefore stochastic is significant on the yield of all the clones at 99% confidence level and therefore is of high management interest as it defines the processing plant raw material supply chain and hence marketing plan. The result matches the wet and dry spell of southern Nigeria and its effect on rubber tree latex production. The seasonal value by quarter is estimated for each clone to be used for appropriate quarterly adjustment in the aggregate rubber production planning.

The dynamic multiplier and the impulse-response function estimated by STAMP for the six clones are in consonance with [17] clonal typology of PB217 as belonging to low-medium metabolism clonal group that requires high intensity stimulation on the one hand and PB260 of high metabolism, demanding low stimulation intensity on the other end of the spectrum. It is today an important plantation practice that informs the annual campaign of latex diagnosis in which physiological status of the natural rubber tree is ascertained by the tree biochemical parameters such as dry rubber content, sucrose and inorganic phosphorus content as well as the proportion of thiols (an antioxidant) using optical density spectrophotometer in order to optimise the exploitation of rubber plantations.

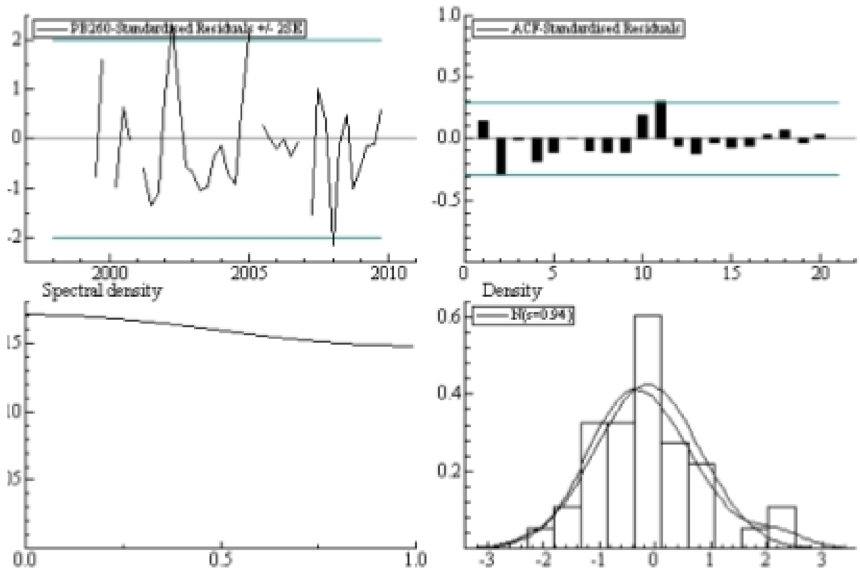

Figure 5: PB260 Residual statistics plot.

4.5. Long run elasticity of rubber yield per tree

The long-run height elasticity of rubber production per tree is obtained from the lead equation of each rubber clone: $x_t = 0.39x_{t-1}+25ht+931st+\mu_t+\gamma_t+\varepsilon_t$, $\varepsilon_t \backsim N(0,\ 1.2E{+}05)$ of GT1 rubber clone for instance as follows [18]:

$$\lim_{j\to\infty}\left[\frac{\partial y_{t+j}}{\partial w_t}+\frac{\partial y_{t+j}}{\partial w_{t+1}}+\frac{\partial y_{t+1}}{\partial w_{t+2}}+\cdots+\frac{\partial y_{t+j}}{\partial w_{t+j}}\right] = 1+\phi+\phi^2+\cdots = 1/(1-\phi)\forall \le 1 \quad (12)$$

where w_t = random variable and ϕ = the dynamic multiplier

$$\frac{25}{1-0.39} = 41$$

The result suggests that a permanent 1cm increase in tapping height will lead to 41g increase in rubber production per tree. PB324 has the highest long run elasticity of yield of 2211 gm per tree while RRIM 703 has the least potential of 1053gm/tree on the long run as shown in Table 5.

4.6. Test of mis-specification

R_s^2: The most suitable coefficient of determination is that around the season mean (Table 5) because of the seasonal effect, which for all the clones is positive and far from zero- greater than 50. Its relevance as a goodness of fit criterion is therefore not marginal. Normality: In Table 3.5, PB28/59 model normality is 7.06 > 5.99 at 5% critical value with 2 degree of freedom. Hence the predictive value of PB28/59 model is highly suspect with the generated data. Higher normality values are often caused by outliers.

Table 8: Chow failure test statistics.

Clones	Failure Chi2 Test	P-value	Accept Ho
GT1	11.69	0.17	> 0.05
PB217	11.81	0.16	> 0.05
PB260	9.88	0.27	> 0.05
PB28/59	3.27	0.92	> 0.05
PB324	8.54	0.38	> 0.05
RRIM703	4.21	0.83	> 0.05

Q(q,d)-stat-The Box-Ljung Q-statistics is a test for residual serial correlation based on the first q residual autocorrelations and distributed approximately as $\chi^2 d$ where d= $q-p$ is a lack of fit statistics. If model is correctly specified residuals should be uncorrelated and Q should be small with large probability value (> 0.01). The null hypothesis for the test is that all the autocorrelations of residuals for lags 1 to k, are zero. In Table 6, K = 8, GT1, PB217, PB324 and RRIM703, the null hypothesis cannot be rejected with p-values greater than 0.05 at 95% (min p=0.15) and at K=12, PB260 and PB28/59, null hypothesis cannot be rejected at 99% with values greater than 0.01(min p=0.02). In Table 6 the Q-stat probability value for GT1, PB217, PB 324 and RRIM 703 are large showing no residual correlation but PB 28/59 and PB 260 have low probability value depicting significant residual correlation at 5% but satisfactory at 10% significant levels. Autocorrelation-The statistic denoted by r(j) give the autocorrelation at lag j. An excessive amount of residual serial autocorrelation is a strong indication that the model is not adequately captur-

Table 5: Impulse response function and Long-run elasticity of yield/tree in grams.

Clones	Dynamic multiplier AR(1)	Height Impulse Response Function	Girth Impulse Response Function	Stimulation Impulse Response Function	Long-run elasticity of yield
PB324	0.32			1570	2211
PB217	0.30		-555	687	1774
PB28/59	0.51	24	-856		1600
GT1	0.39	25		931	1541
PB260	0.59	28		531	1118
RRIM703	0.33		-737		1053

Table 6: Test of mis-specification.

Clone	Rs^2	Q-Stat	Q-Prob.	Std Error	Normality	r(1)	H(12)	Durbin-Watson
GT1	0.62	5.61	0.40	1368	1.75	0.02	1.27	1.78
PB217	0.58	7.4	0.53	1404	4.4	0.05	0.81	1.80
PB260	0.58	16.6	0.06	1565	2.73	0.22	0.67	1.50
PB28/59	0.59	18.13	0.06	1410	7.06	0.21	0.19	1.56
PB324	0.62	8.33	0.21	1372	6.02	0.27	0.60	1.70
RRIM703	0.54	8.99	0.15	1506	3.11	0.16	0.36	1.59

Table 7: Test of specification result.

Clone	Ratio p.e.v/m.d in square	Residual Skewness	Residual Kurtosis	Residual Bowman Shenton Ch^2	p-value	P<0.05 = S
GT1	1.14	0.43	0,18	1.24	0.54	NS
PB217	1.01	0.73	0.15	3.41	0.18	NS
PB260	1.16	0.57	0.43	2.36	0.31	NS
PB28/59	1.39	1.01	1.9	11.92	0.003	S
PB324	1.1	0.43	0.05	1.19	0.55	NS
R703	1.08	0.64	0.16	2.63	0.27	NS

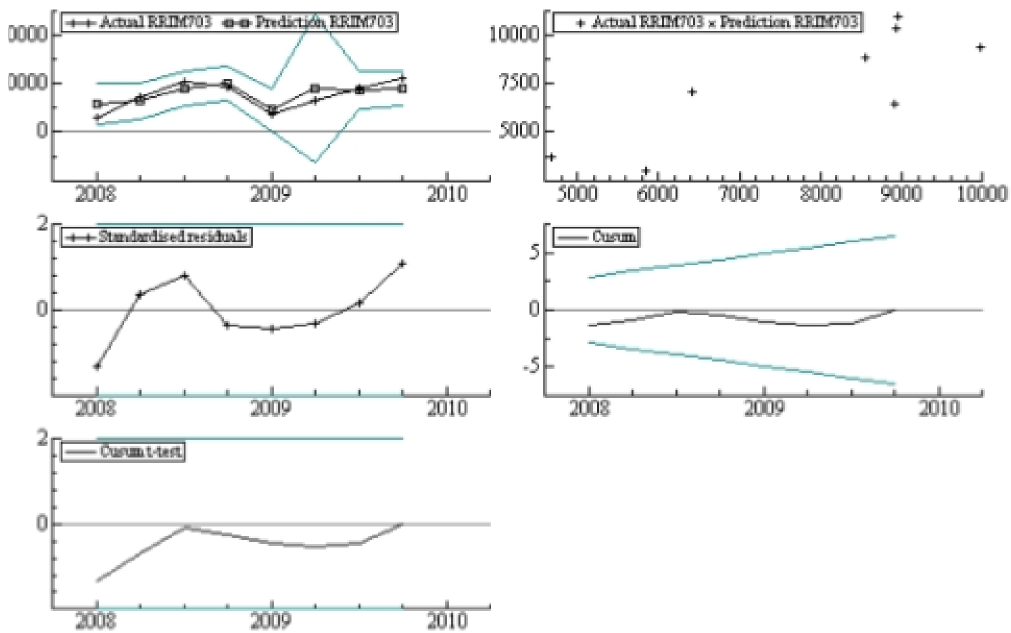

Figure 6: In sample prediction test- RRIM 703.

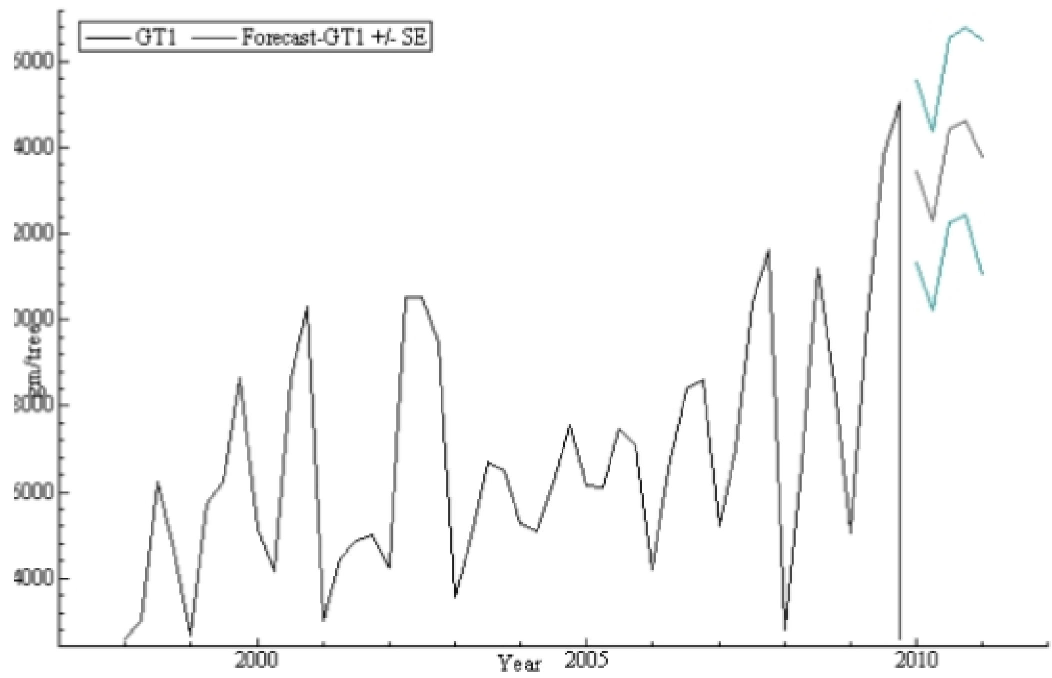

Figure 7: Kalman filter one-step ahead forecast of yield per tree for aggregate production planning.

ing the dynamic structure of the series. The r(1) is the serial correlation coefficients at the first lag (Table 6) distributed approximately as N(0,1/T) where T is the length of the time series. The acf fluctuates close to zero [r(1)] and fall within 95% confidence interval given by $\pm 1.96/\sqrt{T} \pm 0.28$ for large sample. Hence the model produced non significant residual autocorrelation.

Heteroskedasticity (H) occurs when the errors for different dates have different variances but are uncorrelated with each other. H(h) is the ratio of the squares of the last h residuals to the squares of the first h residuals where h is set to the closest integer of T/3 [11]. A high value on F-distribution indicates an increase in variance overtime. A test for Heteroskedasticity distributed approximately as F (h,h) gives 2.69. The low value for the Heteroskedasticity statistic H (12) indicates a degree of Heteroskedasticity in the residuals which is however not significant (Table 6).

In summary, a strong evidence of first-order autocorrelation is established using Durbin-Watson test statistic for the six clones which cast doubt on any inferences drawn from a least squares model currently in use. Hence, a time series model that accounts for the autocorrelation of the random error was implemented in the study. Coefficient of determination Rs^2, that measures the percentage change in the series explained by the model, is that around the season mean. It is positive for all the clones and far from zero $> 50\%$. Its relevance as a goodness of fit criterion is therefore not marginal. The result of the residual autocorrelation function (acf) test shows that the models explain the persistence by producing random residuals which satisfies the statistical assumption of the Kalman filter.

4.7. Residual test

Further, the assumptions underlying the local level model are that the disturbances ε_t and η_t are normally distributed and serially independent with constant variances. The Normality: 5% null hypothesis of normality on Bowman-Shenton χ^2 distribution is shown in Table 7. PB28/59 residual is not normally distributed with Bowman-Shenton χ^2 p-value at $0.003<0.05$. Ratio p.e.v./m.d in square- The basic goodness of fit measure is the prediction error variance (PEV) whose square root is the equation standard error. It is the variance of the residuals in the steady state. The ratio prediction error variance/mean deviation square should be equal to unity for correctly specified model. In Table 7 the goodness of fit criterion is found satisfactory for five clones PB 217, GT1, PB260, PB324 and RRIM 703 with near unity but rather on the high side for PB 28/59 at 1.38. This out come will be useful in the post sample period evaluation of the models.

In Figure 5, estimated spectral density of the residuals bears a reasonable resemblance to the theoretical spectrum of a white noise trend- flat spectrum not bumpy. Estimated probability distributions of the model residuals also closely resemble the corresponding theoretical normal distributions (density in Fig 5) with minor asymmetry for some clones. Fig 5 also

Table 9: Variance of model forecast to actual 2010 (kg/tree).

Qtr	Model Forecast	Actual	RENL Prevision	% Variance Model	% Variance Plantation
1	12.4	5.00	3.65		
2	14.6	10.86	5.68	14.2	47.6
3	14.7	13.23	7.08	9.6	46.4
4	13.9	13.99	6.40	5.8	54.2
Mean	13.9	12.67	5.73	9.5	54.7

depicts the standardized residual plot, the autocorrelation function and the cumulative sum residual plot. The standardized residual is the prediction error or innovation divided by the prediction error standard deviation while the cumulative sum is the sum of the residuals divided by the standard deviation. It is used to detect structural change in the model. The result of the residual acf plot shows that the models explain the persistence by producing random residuals which satisfies the statistical assumption of the Kalman filter (Fig 5).

4.8. In sample prediction test

Fig 6 shows the measured/calculated plotted against time as produced, in order to see how the model error varies with the dependent variable. The 8-quarter in-sample predictive test for RRIM703 shows that prediction and observed values fall within the intervals, confirming absence of any major intervention on the process. The prediction against observed values (Fig 6) is to check the agreement between measured and calculated values. The standardized plot is based on the residuals divided by the estimate of the standard deviation. The cumulative sum test is also plotted. The in-sample prediction is clearly consistent with structure of the model specification with prediction line plots falling within the specified intervals.

The Chow failure test probability at 95% confidence level is greater than 0.05 ($p > 0.05$). The difference between one step predictions and observations is therefore not significant. Table 8 shows test result for all the clones on residuals.

4.9. Incoming year rubber production forecast

Figure 7 shows the yield forecast of the incoming year. The lines on the either side of the forecast function are based on the estimated root mean square error (RMSE) and indicate the prediction interval that limits the rubber yield per tree estimate used for the aggregate production plan. As the forecast horizon increases so does the uncertainty attached to the forecasts and the prediction interval becomes wider. The one year ahead forecast from Fig 7 is shown in Table 9 compared with actual 2010 production and RENL prevision for the same year. The gap between actual rubber production and Kalman filter estimation model is reduced to 10% on the average against plantation management prevision gap of 55%. First quarter 2010 is far from the predicted but in line with data trend and closer to RENL prevision. The increase in stimulation concentration from 2.5% to 5% is a possible reason for the high model forecast value for the 1st quarter of 2010/11.

Significant test of actual yield to one year ahead model forecast in Table 10 is less than 1.96 for the rubber clones and hence the null hypothesis that the actual yield is within the forecasted value is accepted at 5% significant level (95% confidence level).

5. Conclusion

The structural time series decision models estimated by Kalman filtering provides an efficient analytical framework for tracking the yield dynamics of rubber plantations that results from various management rubber exploitation policies. And guarded by theoretical insight, dependability of empirical data and judicious choice of analytical techniques, an apparently entrenched problem of noisy rubber crop collection and prediction environment has herein been rendered more predictable and dependable as a major contribution to rubber plantation industry. However, a wider application of the Kalman filter approach to rubber yield parameter estimation requires a deeper investigation of the Kalman filter assumptions, especially the linearization of the rubber tree yield series, whose relaxation may render Kalman filter application heavy-going.

References

1. Ho, H., Chan Y., and Lim T. *Environmax Planting Recommendations- A New Concept in Choice of Clones.* Proceedings of the Rubber Research Institute of Malaysia Planters Conference, Kuala Lumpur, 1974, pp 293.
2. Leong, Y.S. and NG A.P. *Simple Method of Predicting Next Year's Yield of Clones.* Proceedings of the Rubber Research Institute of Malaysia Planters Conference, Kuala Lumpur, 1974, pp 47.

Table 10: Significant test of Actual Yield to Model Forecast.

		Actual	Forecasted	Standard error	Lower bound	Upper bound	Variate value (d)
Sum							$d = (x - \mu)/\sigma < 1.96$
Treat	Qtr	Total					
GT1	2	10857	12300	2079	10221	14379	-0.694
	3	13230	14415	2147	12268	16563	-0.552
	4	13987	14624	2179	12445	16803	-0.292
GT1 Total		38074	41339				
PB217	2	14220	13832	2070	11761	15902	0.187
	3	15096	15499	2097	13401	17597	-0.192
	4	15287	16518	2106	14412	18624	-0.585
PB217 Total		44602	45849				
PB260	2	10847	9647	2736	6911	12383	0.438
	3	13035	11695	2893	8802	14589	0.463
	4	11774	11771	2989	8782	14760	0.001
PB260 Total		35655	33113				
PB28/59	2	9224	9102	2169	6933	11271	0.056
	3	11533	10832	2283	8549	13115	0.307
	4	9016	11505	2344	9161	13849	-1.062
PB28/59 Total		29773	31439				
PB324	2	12159	13157	1859	11298	15016	-0.537
	3	13442	15059	1857	13202	16917	-0.871
	4	13503	15079	1856	13223	16936	-0.849
PB324 Total		39104	43295				
RRIM703	2	10257	7907	2116	5791	10023	1.111
	3	13119	9941	2241	7700	12183	1.418
	4	11325	10329	2360	7970	12689	0.422
RRIM703 Total		34701	28177				

3. Menon, C.M. and Watson I. *Technique of Estimating Yield for the Coming Year.* Proceedings of the Rubber Research Institute of Malaysia Planters Conference, Kuala Lumpur, 1974, pp 62.
4. Anderson, B.D.O. and Moore J.B. *Optimal Filtering.* Dover publications Inc, Mineola, New York, 1979, p 92.
5. Birdsey, R.A. and Schreuder, H.T. *An overview of forest inventory and analysis estimation procedures in Eastern United States.* Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, 1992, 11p.
6. Drecourt, J.P. *Kalman filtering in hydrological modelling.* DAIHM Technical Report, 2003, pp 5-10.
7. Michael Munroe Intel Corporation. *Sales Rate and cumulative Sales Forecasting Using Kalman Filtering Techniques.* Arizona State University, 2009.
8. Harvey, A. *Forecasting, structural time series models and the Kalman filter.* Cambridge University Press, 1989.
9. Durbin, J. and Koopman, S.J. *Time Series Analysis by State Space Methods.* Oxford University Press, 2001, pp 101-150.
10. Mills, T.C. Modelling Current Temperature Trends. *Journal of Data Science*, Vol. 7, 2009, pp 89-97.
11. Koopman, S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. *STAMP: Structural Time Series Analyser.* Modeller and Predictor. Timberlake Consultants, 2006, pp 61.
12. Chu, C.Y., Pablo L. and Cohen D. Estimation of Infrastructure Performance Models Using State-Space Specifications of Time Series Models. *Transport Research Part C*, 15, 2006, pp 17-32.
13. Mama, B.O. and Osadebe N.N. Comparative analysis of two Mathematical Models for Prediction of Compressive Strength of Sandcrete Blocks using alluvial deposit. *Nigerian Journal of Technology*, Vol. 30, No. 3, 2011, p1.
14. Webster, C.C. and Baulkwill, W.J. *Rubber.* Longman Group, UK limited 1989, pp 349.
15. Goncalves P., Martins A.L., Bortoletto N., Bataglia O. and Silva M. Age trends in the genetic control of production traits in Hevea. *Crop Breeding and Applied Biotechnology*, Vol. 3, no. 2, 2003, p.163-172.
16. Obouayeba, S., Coulibaly L.F., Gohet E., Yao T., and Ake S. Effect of tapping systems and height of tapping opening on clone PB235 agronomic parameters and its susceptibility to tapping panel dryness in south-east Cote dIvoire. *Journal of Applied Biosciences*, 24: 2009, 1535-1542.
17. Gohet E., P. Chantuma, R. Lacote, S. Obouayeba, K. Dian, A. Clement-Demange, Dadang Kurnia and J.M. Eschbach. *Physiological Modelling of Yield Potential and Clonal Response to Ethephon Stimulation.* IRRDB Annual meeting, Ho Chi Minh City, Vietnam, 2003, pp 69.
18. Hamilton, J.D. *Time Series Analysis.* Princeton University Press, 1994, pp 6.