International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

http://ijpsat.ijsht-journals.org
Vol. 22 No. 1 August 2020, pp. 82-92

An Efficient Analysis of Cassava and Rubber Yields in Thailand using GEE and LMM models with Spatial Effects

Rachadasak Supengcum, Watcharin Sangma, Pitsanu Tongkhow and *Sompong Chueaprakha

Department of Industrial Engineering, Faculty of Engineering,
Rajamangala University of Technology Phra Nakhon, Thailand.10800
E-mail: rachadasak.s@ rmutp.ac.th, pitsanu.t@ rmutp.ac.th, and
*Corresponding Author E-mail: sompong.c@rmutp.ac.th

Abstract – The objectives of this research are to propose an efficient and proper model that fits the cassava and rubber yields data. A generalized estimating quation (GEE) and a Linear mixed model (LMM) with spatial correlation following the conditional autoregressive model (CAR) were adopted. The dependent variables are the cassava and rubber yields collected each month in every province of Thailand. The factors considered are rainfall, averaged temperature, and regions. The results from GEE and LMM showed that the factors influencing on the cassava and rubber yields were rainfall, averaged temperature, and regions. Both GEE and LMM fitted the correlated data. The GEE is used to explain the influence of factors on the yields in all provinces while the LMM is used to explain the influence of factors on the yields in each province.

Keywords – Crop yields, Linear mixed model (LMM), generalized estimating equation (GEE), Conditional autoregressive model (CAR), Spatial relationship, Bayesian estimation.

I. INTRODUCTION

Cassava is one of the important agricultural products in the world with yields volume of over 200 million tons per year. The yields tend to expand continuously. Cassava is cultivated in South America, Africa and Asia. For Thailand, cassava is considered an important commercial economic crop in which the yield volume is more than 20 million tons each year. About 30 percent of the yields are used domestically and the rest is exported to the world market. Rubber is another important economic crop in Thailand where the southern region has the highest potential for rubber planting.

In 2017, the southern region had more than 9 million rai (3,558,317 acre) of rubber tapping or 72 percent of the country's total area of tapping resulting in the total 2.7 million tons or 75% of total output. Data for cassava and rubber products that are continuously collected every year are published on the website of Office of Agricultural Economics (2019) using descriptive statistics for presentation, such as tables, graphs, and percentages. Such data will be more useful if they are analyzed in depth using mathematical and statistical models. The most widely used model for repetitive measures is a generalized estimating equation (GEE) model presented by Liang and Zeger (1986). The model shows the relationship between some independent variables and a dependent variable without the assumption that the dependent variable must be independent of each other. The correlation can occur when the measure is repeated in the same sample unit. The dependent variable can be both continuous and discrete. There are some correlation structures, such as independence, exchangeable, first-order autoregressive and unstructured. For independence, we assume that the relationship between the repeated measures is zero. For exchangeable, the correlation is constant. For autoregressive, we assume that the relationship decreases over time and for the unstructured, we assume that the relationship of each pair of data is not formatted. The GEE has the same regression coefficient, therefore an explanation of the magnitude of the influence of factors is on the overall population and. This model is called population averaged model. Limmun and Ingsawang (2010) used the GEE model to find factors that affected road traffic accidents. Amnattrakul and Ingsrisawang (2010) used the GEE model to predict the efficiency of the diaphragm flow cooling tower. Lekdee and Ingsrisawang (2010) used the GEE model to analyze the risk of malaria. Another suitable model for analyzing this type of data is taking spatial relationships into consideration. The most widely used spatial data analysis model is based on a linear mixed model or LMM. The LMM model is the model that shows the relationship of the dependent variable and some independent variables. It is highly flexible so that some variables can be easily added to the model such as variables showing spatial relationships of data. The spatial relationships occur, based on the fact that things that are closer to one another are more related than those that are far away. There are many methods for estimating parameters in an LMM model, but the most widely used method when the model is complex, for example, the addition of variables that represent both spatial and time relationships , is Bayesian methods. Saengseedam and Kantanantha proposed a linear mixed model (LMM) with spatial effects, trend, seasonality and outliers for spatio-temporal time series data. Sammatat and Lekdee [6] used LMM with spatial effects for estimation and detection of rice yield in Thailand. Fox, Veen, and Klotzke (2019) used linear mixed models for randomized responses of the main concerns when asking sensitive questions about behavior and attitudes.

II. Methods

Theories and basic principles used in the development and design of experiments on a generalized estimating quation (GEE) and linear mixed model (LMM) models that have spatial influences are included that the following.

The GEE model

Islam and Chowdhury(2017) has the following form.

Let $Y_{ij}, j=1,...,n_i$, $i=1,...,M$, be the observation of unit i at time j and $N=\sum_{i=1}^{K} n_i$ denotes the total number of observations. Y_{ij} is related to the independent variables $\mathbf{X}_{ij} = (1, X_{ij.1},..., X_{ij.p})^T$ and the parameter $\boldsymbol{\beta} = (\beta_0, \beta_1,..., \beta_p)^T$, and has an identity link: $g(a)=a$, natural log link: $g(a)=\log(a)$, or logit link: $g(a)=\text{logit}(a)=\log(a/(1-a))$. The relationship between Y_{ij} and $\mathbf{X}_{ij} = (1, X_{ij.1},..., X_{ij.p})^T$ is expressed as

$$g(E(Y_{ij} \mid \mathbf{x}_{ij})) = g(\mu_{ij}) = \mathbf{x}_{ij}{}^T\boldsymbol{\beta} \qquad (1)$$

which is the marginal regression model. The covariance Matrix of $\mathbf{Y}_i = (Y_{i1},...,Y_{ij})^T$ is

$$\mathbf{V}_i = \phi \mathbf{A}_i{}^{1/2}\mathbf{R}(\rho)\mathbf{A}_i{}^{1/2} \qquad (2)$$

where ϕ is dispersion parameter, $\mathbf{A}_i$ is a diagonal matrix of a variance function and $\mathbf{R}(\rho)$ is a correlation matrix. The correlation matrix shows the pattern of data relations of $\mathbf{Y}_i = (Y_{i1},...,Y_{ij})^T$, which has one of the following forms.

Exchangeable structure:

$$\mathbf{R}(\rho) = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix},$$

First-order autoregressive (AR(1)) structure:

$$\mathbf{R}(\rho) = \begin{bmatrix} 1 & \rho & \cdots & \rho^{j-1} \\ \rho & 1 & \cdots & \rho^{j-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{j-1} & \rho^{j-2} & \cdots & 1 \end{bmatrix}$$

Unstructured structure:

$$\mathbf{R}(\rho) = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1j} \\ \rho_{21} & 1 & \cdots & \rho_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{j1} & \rho_{j2} & \cdots & 1 \end{bmatrix}.$$

For parameter estimation of $\boldsymbol{\beta}$, we use a numerical iterative method to find the answer of the score function of $\boldsymbol{\beta}$ which is represented as $U(\boldsymbol{\beta}) = \sum \mathbf{D}_i{}^T \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$, where $\mathbf{D}_i = \dfrac{\partial \mu_{ij}}{\partial \beta_k}, k = 1, 2, ..., p$, which has the following steps:

1) Determine the format of $\mathbf{R}(\rho)$ then calculate values $\mathbf{V}_i$ and $\mathbf{b}$ which are the estimates of $\boldsymbol{\beta}$.

2) Calculate the error of $r_{ij} = Y_{ij} - \mu_{ij}$.

3) Recalculate the value of $\mathbf{V}_i$ from the tolerances r_{ij} in item 2).

4) Recalculate the value of $\mathbf{b}$ from the tolerances $\mathbf{V}_i$ in item 3).

5) Repeat step 2) - 4) until $\mathbf{b}$ converges to a constant.

A linear mixed model (LMM)

$$y_{ij} \mid \mathbf{b} \sim N(\mu_{ij}, \sigma^2), i = 1, ..., \mathrm{m}$$

, $j = 1, ..., \mathrm{n}$, where μ_{ij} is shown in Eq. (3),

$$\begin{aligned}\mu_{ij} &= \beta_0 + \beta_1(\mathrm{x}_1) + \\ &\beta_2(\mathrm{x}_2) + \beta_3(\mathrm{x}_3) + \mathrm{b}_\mathrm{i} + \phi_i,\end{aligned} \quad (3)$$

where $\beta_0, \beta_1, \beta_2, \beta_3 \sim N(0, 1.0E09)$,

$$\mathrm{b}_\mathrm{i} \sim N(0, \tau_b^{\,2})$$

$$, \phi_i \mid \mathbf{\Phi}_{(-i)} \sim N\left(0, \frac{\tau_\phi^{\,2}}{w_{i+}}\right),$$

$$\mathrm{var}(\mathbf{\Phi}) = (\mathbf{D}_w - \rho \mathbf{W})^{-1}, \rho \sim \mathrm{Uniform}(0,1),$$

$$\tau_b^{\,2}, \sigma^2, \tau_\phi^{\,2} \sim \mathrm{IG}(0.0001, 0.0001).$$

A linear mixed model with spatial effects is shown in Eq.(4).

$$\mu_{ij} = \mathbf{x}_{ij}^{\,T} \boldsymbol{\beta} + \mathbf{z}_{ij}^{\,T} \mathbf{b} + \phi_i \quad (4)$$

The most widely used model for spatial influence is a conditional autoregressive (CAR) model which has the following details: (Banerjee et al. (2004)).

$$\phi_i \mid \mathbf{\Phi}_{(-i)} \sim N\left(\sum_{j=1}^{m} a_{ij}\phi_j, \tau_i^{\,2}\right),$$

where $\phi_{(-i)} = \{\phi_j : j \neq i\}$.

Let $\mathbf{\Phi} = (\phi_1, ..., \phi_m)^T$ is a vector of a random variable that changes in area. It has the influence on the variable Y_i, the measure in each area i, $i = 1, ..., m$. τ^2 is the conditional variance and a_{ij} is the constant that shows the relationship of the area i and j, where $a_{ii} = 0$, $\mathbf{A} = (a_{ij})$ and $\mathbf{M} = Diag(\tau_1^{\,2}, ..., \tau_m^{\,2})$. The joint distribution (Brook's Lemma) of $(\phi_1, ..., \phi_m)$ is shown in Eq. (5).

$$p(\mathbf{\Phi}) \propto \exp\left\{-\frac{1}{2}\mathbf{\Phi}^T \mathbf{M}^{-1}(\mathbf{I} - \mathbf{A})\mathbf{\Phi}\right\},$$

$$\mathrm{var}(\mathbf{\Phi}) = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{M}. \quad (5)$$

$\mathbf{M}^{-1}(\mathbf{I} - \mathbf{A})$ is a symmetrical matrix and $p(\mathbf{\Phi})$ is a probability distribution. The condition $\dfrac{a_{ij}}{\tau_i^{\,2}} = \dfrac{a_{ji}}{\tau_j^{\,2}}$ must be true. This condition can be true when $a_{ij} = \dfrac{w_{ij}}{w_{i+}}$ and $\tau_i^{\,2} = \dfrac{\tau^2}{w_{i+}}$. Therefore, the enumeration of $\phi_i \mid \mathbf{\Phi}_{(-i)}$ which is a probability distribution. Therefore, it has the form; $\phi_i \mid \mathbf{\Phi}_{(-i)} \sim N\left(\sum_{j=1}^{m} \dfrac{w_{ij}\phi_j}{w_{i+}}, \dfrac{\tau^2}{w_{i+}}\right)$ which is called the CAR model. The joint distribution of $(\phi_1, ..., \phi_m)$ is $p(\mathbf{\Phi}) \propto \exp\left\{-\dfrac{1}{2\tau^2}\mathbf{\Phi}^T(\mathbf{D}_w - \mathbf{W})\mathbf{\Phi}\right\}$. $\mathbf{W} = (w_{ij})$ is a weighted matrix showing the relationship between areas. $w_{ij} = 1$ if the areas are adjacent and $w_{ij} = 0$ if the areas are not adjacent to each other. $\mathbf{D}_w = Diag(w_{i+})$ is the main diagonal matrix whose members are in diagonal and $w_{i+} = \sum_j w_{ij}$. Since $(\mathbf{D}_w - \mathbf{W})$ is a singular matrix, $p(\mathbf{\Phi})$ is improper, i.e., the total area under the graph is not equal to 1. Therefore, the CAR model is improper. The solution for $(\mathbf{D}_w - \mathbf{W})$ to have inverse is to add a parameter ρ to $(\mathbf{D}_w - \mathbf{W})$, then $\mathrm{var}(\mathbf{\Phi}) = (\mathbf{D}_w - \rho\mathbf{W})^{-1}$. The distribution of $\phi_i \mid \mathbf{\Phi}_{(-i)}$ has a new form, $\phi_i \mid \mathbf{\Phi}_{(-i)} \sim N\left(\rho\sum_{j=1}^{m} \dfrac{w_{ij}\phi_j}{w_{i+}}, \dfrac{\tau^2}{w_{i+}}\right)$ and is called spatial parameter.

The estimation of parameters in the model was performed based on the Bayesian method by Congdon (2006) using the Markov Chain Monte Carlo method with Gibbs sampling. For the Bayesian paradigm, current knowledge about the model parameters is expressed by placing a probability distribution on the parameters, called the "prior distribution", often written as $p(\theta).p(\theta)$. When new data $\mathbf{y}$ becomes available, the information they contain, regarding the model parameters, is expressed in the "likelihood" which is proportional to the distribution of the observed data given the model parameters, written as

$p(y\,|\,\theta).p(y\,|\,\theta)$. This information is then combined with the prior to produce an updated probability distribution called the "posterior distribution". Bayesian inference is based on Bayes' Theorem. The posterior is proportional to the prior times the likelihood, or more precisely,

$$p(\theta|y) = \frac{p(\theta).p(y|\theta)}{\int_{\Theta} p(\theta).p(y|\theta)d\theta} \qquad (6)$$

In theory, the posterior distribution is always available, but in realistically complex models, the required analytic computations often are intractable. Over several years, in the late 1980s and early 1990s, it was realized that methods for drawing samples from the posterior distribution could be very widely applicable. The Bayesian method was obtained by writing programs in OpenBUGS and processing simulation in R program with the package of R2OpenBUGS proposed by Cowles (2013). The independent variables are rainfall, average temperature and regions of Thailand and the dependent variables are annual cassava and rubber yields in every province of Thailand collected from the Department of Agriculture, Office of Agricultural Statistics. The data for independent variables were obtained from the Meteorological Department (2019).

III. Results And Discussion

For this study, an efficient analysis of cassava and rubber yields in Thailand using GEE and LMM models with spatial effects, the objectives aim at creating a suitable model for estimating cassava and rubber yields and investigating factors that influence the yields. The data used in this study are secondary data. Cassava and rubber yields were collected in every province of Thailand from the Office of Agricultural Economics, Ministry of Agriculture and Cooperatives. Rain and temperature data were collected from the Meteorological Department, Ministry of Information and Communication Technology. Data analysis was performed by programming in OpenBugs and R. The general characteristic of the data and the factors affecting the yields of cassava and rubber yields were analyzed. The mean and standard deviation were used to present the general characteristics of the data. The LMM with spatial influence was used for to estimate the cassava and rubber yields and to investigate the factors related to the yields. Then the LMM with spatial influence was compared to the GEE model. The Bayesian method was used for parameter estimation. The details are shown in Eq. (7-9).

The proposed model has details as follows:

Let Y_{ij} represent the yields in the i province, the j year, where $i = 1,...,76$ and $j = 1,...,12$.

$X_{ij.1}$ represents the amount of rain (mm) in the ith province, the j th year.

$X_{ij.2}$ is the average temperature (degree Celsius) in the i th province, the j th year.

$X_{i.3}$ represents the ith province which is in the north.

$X_{i.4}$ represents the ith province in the northeast.

$X_{i.5}$ represents the i province in the southern region.

$X_{i.6}$ represents the i province in the eastern region.

$X_{i.7}$ represents the ith province where is in the western region.

β_0 is Intercept and $\beta_1,...,\beta_7$ are the regression coefficients of rainfall, average temperature, north, northeast, south, east and west, respectively. The relation of the independent variables and the dependent variable is presented by Eq. (7):

$$\begin{aligned}\mu_{ij} &= \beta_0 + \beta_1 X_{ij.1} + \beta_2 X_{ij.2} + \beta_3 X_{i.3} \\ &+ \beta_4 X_{i.4} + \beta_5 X_{i.5} + \beta_6 X_{i.6} + \beta_7 X_{i.7}\end{aligned} \qquad (7)$$

LMM model for cassava yields with spatial effects are shown in Eq. (8-9) and rubber yields with spatial effects are shown in Eq. (10-11). , parameter estimation by Bayesian method, the model details are shown as follows:

$$y_{ij} \,|\, \mathbf{b} \sim N(\mu_{ij}, \sigma^2) \;, i = 1,...,45 \;, j = 1,...,9. \qquad (8)$$

$$\begin{aligned}\mu_{ij} &= \beta_1 + \beta_2 * rain_{ij} + \beta_3 * temp_{ij} + \\ &\beta_4 * north + \beta_5 * northeast + \\ &\beta_6 * south + \beta_7 * east + \beta_8 * west + \\ &\beta_9 * k + b_{1i} + b_{2ij} + v_i\end{aligned} \qquad (9)$$

$$y_{ij} \,|\, \mathbf{b} \sim N(\mu_{ij}, \sigma^2) \;, i = 1,...,50$$

$$, j = 1,...,9. \qquad (10)$$

$$\begin{aligned}\mu_{ij} &= \beta_1 + \beta_2 * rain_{ij} + \beta_3 * temp_{ij} + \\ &\beta_4 * north + \beta_5 * northeast + \\ &\beta_6 * south + \beta_7 * east + \beta_8 * west \\ &+ \beta_9 * k + b_{1i} + b_{2ij} + v_i\end{aligned} \qquad (11)$$

v_i is the spatial influence following the distribution of CAR model which has the following form.

$$v_i \mid \mathbf{v}_{(-i)} \sim \mathrm{N}\left(\sum_{k=1}^{m} \frac{w_{ik} v_k}{w_{i+}}, \frac{\tau_v^{\,2}}{w_{i+}}\right) \text{ and}$$

$$\mathbf{v} \sim \mathrm{N}\left(\mathbf{0}, \tau_v^{\,2} (\mathbf{D}_w - \mathbf{W})^{-1}\right) \qquad \text{or}$$

$$\mathrm{p}(\mathbf{v}) \propto \exp\left\{-\frac{1}{2\tau_v^{\,2}} \mathbf{v}^T (\mathbf{D}_w - \mathbf{W}) \mathbf{v}\right\}.$$

Under the Bayesian method, the prior distributions are assumed to be non-informative priors that do not affect the posterior.

$$\beta_0, \beta_1, \beta_2, ..., \beta_9 \sim N(0,100000000) \; b_{1i} \sim N(0, \tau_{b1}^{\,2})$$

$$b_{2it} \sim N(0, \tau_{b2}^{\,2})$$

$$\tau_{b1}^{\,2}, \tau_{b2}^{\,2}, \tau_v^{\,2} \sim InvGamma(0.0001, 0.0001)$$

The parameters were estimated by the Bayesian method using MCMC Gibbs sampling, programming in OpenBUGS and simulating in R2OpenBUGS. The results of the analysis of the factors affecting cassava and rubber yields in Thailand using the GEE and LMM models are shown in Table 1-4. The top 10 provinces the rubber yields in descending order are Surat Thani (475,051.33), Songkhla (393,894.33), Nakhon Si Thammarat (342,159.11), Trang (328,600.56), Yala (252,339.00), Narathiwat (230,049.00), Phang Nga (154,378.11), Phatthalung (143,606.44), Krabi (138,642.00), Rayong (135,092.67), respectively. The top 5 cassava yields in descending order are Nakhon Ratchasima (6,023,831.33), Kamphaeng Phet (2,071,006.17), Chaiyaphum (1,361,139.11), Sa Kaeo (1,286,453.22), Kanchanaburi (1,228,839.10), respectively.

From Table 1, the factors affecting cassava yields are rainfall, average temperature, northern region, northeast region, southern region and western region when the central region is the reference region. If the rainfall increases 1 mm, the average cassava yield per year will decrease 115.32 tons. If the average temperature increases 1 Celsius, the average annual cassava yields will decrease 9,993.35 tons. The northern region has a low average annual cassava yield over the central region 510,025.90 tons. The northeastern region has an average yield of cassava 279,458.40 tons more than the central region.. The eastern region has an average cassava yield of 432,167.26 tons more than the central region, and the western region has an average yield of cassava less than the central region 78,096.29 tons. For the southern region, there is no cassava product. When in order of estimation, the influence of the region on cassava yields to average per year, from high to low, arranged as follows, East, Northeast, central, west, north and south respectively.

Table 1. Results of analysis of factors affecting cassava yields in Thailand using GEE model.

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	954097.763	257685.4672	449043.528	1459151.998	13.709	1	.000
rain	-115.323	36.3562	-186.580	-44.066	10.062	1	.002
temp	-9993.353	3886.3790	-17610.516	-2376.190	6.612	1	.010
north	-510025.908	177511.1495	-857941.368	-162110.449	8.255	1	.004
northeast	279458.406	359046.4256	-424259.657	983176.469	.606	1	.436
south	0a	.	.	.	.	.	.

east	432167.262	207240.0023	25984.321	838350.203	4.349	1	.037
west	-78096.289	307652.7618	-681084.622	524892.044	.064	1	.800
(Scale)	847543381741.733						

Dependent Variable: rubber

Model: (Intercept), rain, temp, north, northeast, south, east, west

From Table 2, the factors affecting the rubber yields are the amount of rain, the average temperature of the north, northeast, south, west and west, when the central region is the reference region. If the rainfall increases 1 mm, the average rubber yields will decrease 4.21 tons, if the average temperature increases 1 Celsius, the average rubber yields per year will decrease 317.80 tons. For the rubber yields, the average yield per year in the northern region is 8,680.58 tons more than in the central region. In the northeastern region, the average rubber yield per year is 12,355.25 tons less than the central region. In the southern region, the average annual rubber yield is 208,546.17 tons higher than the central region. In the eastern region, the average rubber yield is 52,691.44 tons higher than the central region, and the western region has an average annual rubber output of 7,599.88 tons higher than the central region.

When ranked in order of estimates, the influence of the region on the average rubber yield per year from highest to lowest can be arranged as follows: south, east, northeast, north, west and central respectively.

Table 2. Results of analysis of factors affecting rubber yields in Thailand using GEE model.

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	**Mean**	**Std. Error**	**Lower**	**Upper**	**Wald Chi-Square**	**df**	**Sig.**
(Intercept)	16234.276	7913.8027	723.508	31745.045	4.208	1	.040
rain	-4.213	3.5404	-11.152	2.726	1.416	1	.234
temp	-317.800	315.2688	-935.715	300.116	1.016	1	.313
north	8680.582	5358.8492	-1822.569	19183.734	2.624	1	.105
northeast	12355.251	2893.9170	6683.277	18027.224	18.228	1	.000
south	208546.165	41160.0177	127874.013	289218.317	25.672	1	.000
east	52691.441	18856.5045	15733.372	89649.511	7.808	1	.005
west	7599.884	4041.8410	-321.979	15521.747	3.536	1	.060
(Scale)	6156914956.768						

Dependent Variable: rubber

Model: (Intercept), rain, temp, north, northeast, south, east, west

From Table 3, factors affecting cassava production are rainfall, average temperature, northern region, northeast region, southern region and western region when the central region is the reference region. If the amount of rainfall increases by 1 mm, the average annual cassava yield will decrease by 80.43 tons. If the average temperature rises 1 Celsius, the average cassava yield will increase 5.43 tons. In the northern region, the average annual cassava yield is 0.63 tons less than the central region. In the northeastern region, the average annual cassava yield is 0.48 tons less than the central region. The average cassava yield in the southern region is 0.91 tons higher than the central region. In the eastern region, the average yield of cassava is 1.32 tons lower than the central region. In the western region, the average cassava yield is 0.56 tons higher than the central region. When ordering the estimated influence of the region on cassava yield average per year, from high to low, they can be arranged as follows: south, west, central, northeast, north and east respectively.

Table 3. Estimates of the influence of factors affecting cassava yields from LMM model

Parameter	Mean	Std. Error	95% Credible Interval	
β_1 (Intercept)	0.14	100.20	-195.10	195.40
β_2 (rain)	-80.48	35.85	-151.90	-10.35
β_3 (temp)	5.45	99.50	-186.20	202.40
β_4 (north)	-0.63	99.73	-195.90	193.90
β_5 (northeast)	-0.48	100.40	-198.00	198.40
β_6 (south)	0.91	99.39	-194.60	195.00
β_7 (east)	-1.32	100.10	-197.30	193.70
β_8 (west)	0.56	99.51	-194.40	194.70
Central region (reference)	.	.	.	.
β_9 (tend)	10.10	99.75	-184.30	206.50

From Table 4, the factors affecting the rubber yield are the amount of rain, the average temperature, the north, northeast, south, west and west, when the central region is the reference region. If the amount of rain increases by 1 mm, the average rubber yield will increase 1.63 tons. If the average temperature rises 1 Celsius, the average rubber yield will increase 0.92 tons. The northern region has an average rubber yield of 65.91 tons more than the central region. In the northeastern region, the average rubber yield is less than the central region of 70.44 tons. The southern region has an average rubber yield of 29.09 higher than the central region. The eastern region has the average rubber yield of 48.50 tons higher than the central region, and in the western region, the average rubber yield is 16.75 tons higher than the central region.

Table 4. Estimates of the influence of factors affecting rubber yields from LMM model

Parameter	Mean	Std. Error	95% Credible Interval	
β_1 (Intercept)	8.20	33.84	-75.22	73.69
β_2 (rain)	1.63	0.05	1.55	1.71
β_3 (temp)	0.92	6.03	-7.27	5.73
β_4 (north)	65.91	72.68	-133.20	156.60
β_5 (northeast)	-70.44	48.23	-124.40	62.88
β_6 (south)	29.09	55.62	-150.50	125.50

β_7 (east)	48.50	62.55	-94.10	115.30
β_8 (west)	16.75	53.43	-111.90	125.20
Central region (reference)	.	.	.	.
β_9 (tend)	201.30	19.58	155.70	227.70

When the estimated influence of the region on average rubber yield per year ranked in descending order, they are the north, east, east, south, west, central and northeast, respectively.

From Table 5, it is found that the top 10 provinces with high average rubber yields, ordered from highest to lowest, are as follows: Surat Thani, 2013, (588,200), Surat Thani, 2014, (582,000), Surat Thani, 2012, (574,200), Songkhla, 2013, (493,300), Songkhla, 2014, (490,800), Songkhla, 2012, (477,000), Surat Thani, 2007, (447,100), Nakhon Si Thammarat 2013, (440,400), Nakhon Si Thammarat, 2014, (439,500), Surat Thani, 2007, (437,200), respectively. From Table 6, it is found that the top 10 provinces with high average cassava yields estimated, ordered from highest to least are as follows: Nakhon Ratchasima, 2013, (6,023,000), Nakhon Ratchasima, 2011, (6,019,000), Nakhon Ratchasima, 2008, (6,007,000), Nakhon Ratchasima, 2014' (6,007,000), Nakhon Ratchasima, 2008' (6,004,000), Nakhon Ratchasima, 2010'(6,004,000), Nakhon Ratchasima, 2012' '(5,996,000), Nakhon Ratchasima, 2007' (5,991,000), Nakhon Ratchasima, 2009' (5,990,000), Kamphaeng Phet, 2014' (2,098,000), respectively. From Table 7, it is found that the top 11 provinces with highest spatial influence on the average rubber yields in Thailand , ranked from the highest to the least, are as follows: Chumphon (0.06), Suphanburi (0.06), Bangkok (0.06), Nakhon Si Thammarat (0.05), Prachuap (0.05), Phuket (0.05), Satun (0.05), Phra Nakhon Si Ayutthaya (0.05), Nakhon Pathom (0.05), Ranong (0.05), Phitsanulok (0.05) respectively. The top 10 provinces with highest spatial influence on average cassava yields in Thailand, ranked from highest to least, are as follows: Phetchaburi (376.5), Ratchaburi (356.7), Kanchanaburi (147.8), Chiang Rai (116.8), Phayao (115.5), Lampang (103.9), Suphan Buri (84.0), Sakon Nakhon (52.1), Phrae (44.0), Chai Nat (41.8), respectively.

Table 5. Example value estimated average rubber yields per year in provinces with an average higher than 100,000 tons.

Province	**Year**	Rubber **yield** (ton(			
		Mean	**Std. Error**	**95% Credible Interval**	
Surat Thani	2013	588,200	230.0	588,000	588,300
Surat Thani	2014	582,000	223.6	581,800	582,200
Surat Thani	2012	574,200	212.1	574,000	574,400
Songkhla	2013	493,300	221.6	493,100	493,500
Songkhla	2014	490,800	218.5	490,600	490,900
Songkhla	2014	477,000	214.7	476,800	477,200
Surat Thani	2007	447,100	212.6	447,000	447,300
Nakhon Si Thammarat	2013	440,400	223.4	440,200	440,500
Nakhon Si Thammarat	2014	439,500	222.4	439,300	439,700
Surat Thani	2006	437,200	218.5	437,000	437,400

Nakhon Si Thammarat	2012	429,200	219.5	429,000	429,300
Surat Thani	2009	428,200	203.9	428,000	428,400
Surat Thani	2010	419,800	199.8	419,600	420,000
Surat Thani	2011	404,400	238.0	404,300	404,600
Surat Thani	2008	394,300	222.4	394,100	394,500
Songkhla	2009	366,600	216.4	366,500	366,800

Table 6. Example value estimated average cassava yields yields per year in provinces with an average higher than 100,000 tons.

Province	**Year**	Rubber **yield** (ton(			
		Mean	**Std. Error**	**95% Credible Interval**	
Nakhon Ratchasima	2013	6,023,000	66,490	5,891,000	6,153,000
Nakhon Ratchasima	2011	6,019,000	66,250	5,888,000	6,148,000
Nakhon Ratchasima	2006	6,007,000	65,840	5,877,000	6,135,000
Nakhon Ratchasima	2014	6,007,000	65,850	5,877,000	6,136,000
Nakhon Ratchasima	2008	6,004,000	65,810	5,874,000	6,133,000
Nakhon Ratchasima	2010	6,004,000	65,800	5,874,000	6,133,000
Nakhon Ratchasima	2012	5,996,000	65,830	5,867,000	6,126,000
Nakhon Ratchasima	2007	5,991,000	65,970	5,862,000	6,120,000
Nakhon Ratchasima	2009	5,990,000	65,980	5,861,000	6,120,000
Kamphaeng Phet	2014	2,098,000	67,380	1,965,000	2,229,000
Kamphaeng Phet	2008	2,081,000	66,190	1,951,000	2,211,000

Table7**.** Examples of the spatial effects of each province on rubber production in Thailand.

Province	**Year**	**Spatial effects**			
		Mean	**Std. Error**	**95% Credible Interval**	
Krabi	141,800	21.5	141,800	141,900	6,153,000
Chumphon	141,800	21.6	141,800	141,900	6,148,000
Trang	141,800	21.5	141,800	141,900	6,135,000
Nakhon Si Thammarat	141,800	21.5	141,800	141,900	6,136,000
Narathiwat	141,800	22.0	141,800	141,900	6,133,000

Prachuap Khiri Khan	141,800	22.0	141,800	141,900	6,133,000
Pattani	141,800	21.8	141,800	141,900	6,126,000
Phangnga	141,800	21.5	141,800	141,900	6,120,000
Phatthalung	141,800	21.5	141,800	141,900	6,120,000

The linear mixed model (LMM) model shows a relationship between more than one independent variables and the dependent variable and allows the data to be dependent. It can be used for analyzing repeated measurement data. That is a random influence can be added into the model. The regression coefficients in each unit of the sample unit is different, explaining the magnitude of the factors of the individual characteristics of each unit. Therefore, the model is classified as a subject-specific model. The GEE model, was proposed by Liang and Zeger (1986). Unlike LMM models, the GEE does not have a randomly influential term. They show the relationship between the independent variables. and the dependent variable without the assumption that the data are independent. That is to allow the data to be dependent, which can occur when the data are measured repeatedly in the same sample unit. The dependent variables can be both continuous and discrete. The equation shows the relationship between the independent variables and the dependent variable of each sample unit with the same regression coefficients. It ss an explanation of the magnitude of the influence of factors on the population as a whole, so this model is classified as a population-averaged model.

IV. CONCLUSIONS

In general, the rubber has the highest average yield in each province in the southern region, followed by the eastern, northeast, north, western and central region, respectively. The southern region is located in a tropical region which is more suitable for rubber planting than other regions in Thailand. The soil, rainfall, relative humidity, temperature and wind speed are appropriate for rubber trees in the southern region. On average, rubber production in the southern region is the most productive. It is found that the production of rubber trees, regardless of the yield of latex or wood, depends on 3 factors: rubber varieties, suitability of the area and rubber plantation management. Therefore, for the construction of rubber plantations, in addition to considering the correct rubber varieties and managing the rubber plantations, we must also consider the suitability of the rubber plantation areas. In general, the cassava has the highest production in each province in the eastern region, followed by northeastern, central, western, northern and southern region, respectively. The suitable environment for cassava cultivation is that the height from sea level is not more than 200 meters, without flooding, loam, sandy loam or sandy soil with moderate fertility. Moreover, there should be good drainage and good ventilation. The soil surface depth being less than 30 centimeters, with the pH value between 5.5 - 7.5, is suitable for cassava cultivation.

ACKNOWLEDGMENTS

This research, an efficient analysis of cassava and rubber yields in Thailand using GEE and LMM with spatial effects, aiming to present models to analyze the cassava and rubber yields was well accomplished with support and help from many people. The research team would like to thank the president of Rajamangala University of Technology Phra Nakhon, the dean of the Faculty of Engineering and Director and all the staff of the Research and Development Institute for their support. Finally, the research team would like to thank Rajamangala University of Technology Phra Nakhon for providing financial support until this research was accomplished well.

REFERENCES

[1] Yield data. the Office of Agricultural Economics [updated 2019 Aug 19; cited 2019 Aug 9]. Available from: http://www.oae.go.th/

[2] Liang, KY. and Zeger, SL. Longitudinal data analysis using generalized linear models. Biometrika, 1986; 73: 13-22.

[3] Limmun, W. and Ingsrisawang, L. Study of factors affecting road traffic accidents using generalized estimating equations and generalized linear models. Journal of KMUTNB. 2010; 20(2): 311-321.

[4] Amnarttrakul.R and Ingsrisawang.L.2012.A Model for Prediction of Efficiency of Induced Draft Counterflow Cooling Tower Using Generalized Estimating Equations, Burapha science journal,pp.87-96.

[5] Lekdee, K. and Ingsrisawang, L. Risk factors for malaria in Thailand using generalized estimating equations (GEE) and genalized linear mixed model

(GLMM). Journal of Health Science. 2010; 19: pp. 364-373.

[6] Saengseedam, P. and Kantanantha, N. Spatio-temporal model for crop yield forecasting. Journal of Applied Statistics, 2017; 44(3): 427-440.

[7] Sammatat, S. and Lekdee K. Estimation and Detection of Rice Yield in Thailand Using Spatial and Longitudinal Data Analysis, Applied Mathematical Sciences, 2019; 13(18) : 877 - 884.

[8] Fox, J.-P., Veen, D., & Klotzke, K. (2019). Generalized linear mixed models for randomized responses. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 15(1), 1–18.

[9] Islam,M.A.and Chowdhury,R.I,2017, Analysis of Repeated Measures Data, Springer, Singapore,pp.161-167. [10] Pedroza, C. A Bayesian forecasting model: predicting U.S. male mortality. Biostatistics. 2006; 7(4): 530-550.

[10] Banerjee, S. Carlin, B.P. and Gelfand, A.E. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC Press. FL. Janeway CA, Travers P, Walport M, Shlomchik M. Immunobiology. 5th ed. New York: Garland Publishing. 2004.

[11] Congdon, P. Bayesian Statistical Modelling, 2nd ed., John Wiley & Sons: New York(2006)1-56.

[12] Cowles,MK. Applied Bayesian statistics: with R and OpenBUGS, New York, springer. 2013.

[13] Temperature data. the thai meteorological department [updated 2019 Aug 19; cited 2019 Aug 9]. Available from: https://www.tmd.go.th/en/.