International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:8, No:7, 2014

Spatial Time Series Models for Rice and Cassava Yields Based On Bayesian Linear Mixed Models

Panudet Saengseedam, Nanthachai Kantanantha

Abstract**—**This paper proposes a linear mixed model (LMM) with spatial effects to forecast rice and cassava yields in Thailand at the same time. A multivariate conditional autoregressive (MCAR) model is assumed to present the spatial effects. A Bayesian method is used for parameter estimation via Gibbs sampling Markov Chain Monte Carlo (MCMC). The model is applied to the rice and cassava yields monthly data which have been extracted from the Office of Agricultural Economics, Ministry of Agriculture and Cooperatives of Thailand. The results show that the proposed model has better performance in most provinces in both fitting part and validation part compared to the simple exponential smoothing and conditional auto regressive models (CAR) from our previous study.

Keywords**—**Bayesian method, Linear mixed model, Multivariate conditional autoregressive model, Spatial time series.

I. Introduction

SPATIAL time series are data collected across both time and space. Thus the data analysis should consider the correlations across the time and across the areas. These kinds of data are found in many applications especially in agriculture. For example, the Office of Agricultural Economics, an organization under the Ministry of Agriculture and Cooperatives of the Kingdom of Thailand [1], releases annual reports for common agricultural product yields such as rice, rubbers, cassava, sugar cane, and pineapples, in each province. These reported data and a study of forecasting, which helps decision making and planning in the present, motivated us to investigate and develop a proper forecasting model.

There have been a number of approaches to model time series data, spatial data, or spatial time series data. For example, for time series data, [2] used a Bayesian statistical model to forecast the parts demand time series data for Sun Microsystems, Inc., [3] proposed ARIMA models that could be used to make efficient forecast for boro rice production in Bangladesh from 2008-09 to 2012-13, and reference [4] proposed a forecasting model that can detect trend, seasonality, auto regression and outliers in time series data related to some covariates. Their proposed model was applied to vegetable prices in Thailand.

For spatial data, the spatial effects can be done in a number of ways [5]; one of the common approaches is a conditional auto regressive model (CAR) first introduced by [6]. Reference [7], extending the model of [6], proposed empirical Bayesian methods building from Poisson regression with random intercepts defined with CAR spatial correlations. Reference [8] extended the models of [7] to fully Bayesian setting for mapping the risk from a disease.

For spatial time series data, [9], using geostatistical approach to analyze the yearly data collected from 100 georeferenced locations, studied the spatial and temporal variability of attributes related to the yield and quality of durum wheat production. Reference [10] presented spatial time series models, based on Bayesian linear mixed models with CAR spatial effects, for rice yields in Thailand. Most models for spatial time series data are based on generalized linear mixed models (GLMMs). In this paper we focus on a linear mixed model (LMM), a special case of the GLMMs, since the product yields are continuous data.

LMMs are usually used when responses are correlated data which may be due to repeated measurements on each subject over time [11]. The LMMs allow fixed effects and spatial effects to be included. Recently, a Bayesian approach using Markov Chain Monte Carlo (MCMC) is becoming increasingly popular as techniques for parameter estimation in complex models due to its extreme flexibility, so it is adopted for parameter inference in this paper.

In particular, a CAR model is used for univariate spatial data; the data involve a single response variable. For multivariate spatial data which involve more than one response variables, a multivariate conditionally autoregressive model (MCAR) proposed by [12] is commonly applied. An advantage of an MCAR model is that it can handle the correlations between the response variables as well as well as the spatial correlations between areas. Reference [13] used MCAR for multivariate areal boundary analysis. They illustrated the methods using Minnesota county-level esophagus, larynx, and lung cancer data.

The study of forecasting and the agricultural reported data motivated us to do this work as mentioned earlier. We select to forecast rice and cassava yields because they are major crops of Thailand. Rice production has long played a vital role in Thailand's socio-economic development, making the country the world's largest rice exporter in the last 3 decades. Thailand has the fifth-largest amount of land under rice cultivation in the world [14]. It has planned to increase the rice-growing areas available for rice production by adding 500,000 hectares to it is already 9.2 million hectares [15].

Cassava is considered as one of the most important economic crops of Thailand. It is utilized as raw material in

P. Saengseedam is with the Department of Industrial Engineering, Faculty of Industrial Engineering, Kasetsart University, Bangkok, Thailand (phone: 662-579-8610; e-mail: panudech_mie@yahoo.com).

N. Kantanantha is with the Department of Industrial Engineering, Faculty of Industrial Engineering, Kasetsart University, Bangkok, Thailand (phone: 662-579-8610; e-mail: nantachai.k@ku.ac.th).

various industrials of the country, such as food, feed mill and other continuous industrials [16]. It is also an important energy crop for ethanol production, which is used for gasohol production [17]. Thailand is the fourth cassava producer in the world; however, it is the world largest exporter with exported value of over THB 29 billion per year. Thailand's cassava planted area is 1.2 million hectares which is the fourth following rice, maize and rubber tree, with production yield of 26.9 million ton [1].

This study proposes the LMM with multivariate conditional autoregressive model (MCAR) representing spatial effects for rice and cassava yields in 19 northeastern provinces of Thailand which has not been proposed yet. The proposed model is compared to the simple exponential smoothing and CAR models from our previous study [10]. This paper is organized as follows. Section II briefly describes the methodology. The application is illustrated in Section III. In Section IV, the result of the study is presented. Lastly, in Sections V and VI the discussion and conclusion are drawn.

II. Methodology

A. Linear Mixed Models for Time Series Data

Linear mixed models for time series data can be expressed as:

$$\mathbf{y}_{it} = \mathbf{X}_{it}{}^T\boldsymbol{\beta} + \mathbf{Z}_{it}{}^T\mathbf{b}_i + \boldsymbol{\varepsilon}_{it} \tag{1}$$

for $i=1,...,m;\ t=1,...,T$, where $\mathbf{y}_{it}$ is the i^{th} response at time t, $\mathbf{X}_{it}$ is the regressor variables associated with the fixed effects, $\boldsymbol{\beta}$ is the parameter vector of fixed effects, $\mathbf{Z}_{it}$ corresponds to the predictor variables with random effects, $\mathbf{b}_i \sim \mathrm{MN}(\mathbf{0},\mathbf{D})$ is the random effects for the i^{th} cluster where $\mathbf{D}$ is the positive definite matrix, and $\boldsymbol{\varepsilon}_{it}$ is the random errors.

B. Multivariate CAR Models (MCAR)

Reference [12] describes multivariate CAR models as follows. Let areal random effects corresponding to the two crop yields be $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_1{}^T, \boldsymbol{\varphi}_2{}^T)$ where $\boldsymbol{\varphi}_1{}^T = (\phi_{11},...,\phi_{m1})$, $\boldsymbol{\varphi}_2{}^T = (\phi_{12},...,\phi_{m2})$, and m is the number of areal units. The bivariate spatial random effect $\boldsymbol{\varphi}$ is defined as the conditional distribution,

$$\begin{pmatrix}\phi_{i1}\\ \phi_{i2}\end{pmatrix} \Big| \boldsymbol{\varphi}_{-(i1,i2)} \sim \mathrm{N}\left(\begin{pmatrix}\bar{\phi}_{i1}\\ \bar{\phi}_{i2}\end{pmatrix}, \left(w_{i+}\boldsymbol{\Lambda}\right)^{-1}\right), \tag{2}$$

where $\boldsymbol{\varphi}_{-(i1,i2)}$ stands for the collection of all ϕ_{il} except ϕ_{i1} and ϕ_{i2}. Let $\bar{\phi}_{i1} = \sum_l \frac{w_{il}\phi_{l1}}{w_{i+}}$ and $\bar{\phi}_{i2} = \sum_l \frac{w_{il}\phi_{l2}}{w_{i+}}$, the averages of the random effects for area i's neighbors specific to variables 1 and 2, respectively. It can be seen that $\boldsymbol{\Lambda}$ serves as scaled conditional precision for (ϕ_{i1},ϕ_{i2}), where w_{i+} is a scale parameter. Areas with more neighbors have higher precision.

Since $\boldsymbol{\Lambda}$ is common for all areas $i=1,...,m$, it controls the conditional precision for each pair of variables at the same site averaged over all areas. Letting $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^{-1}$, $\frac{1}{w_+}\boldsymbol{\Sigma}$ is the conditional covariance matrix with $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$ as the conditional correlation between ϕ_{i1} and ϕ_{i2}, $i=1,...,m$.

Under the MCAR, the multivariate joint distribution is

$$p(\boldsymbol{\varphi}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\varphi}^T[\boldsymbol{\Lambda}\otimes(\mathbf{D}-\mathbf{W})]\boldsymbol{\varphi}\right),$$

where $\boldsymbol{\Lambda}$ is 2x2 positive definite and $\otimes$ denotes the Kronecker product. $\mathbf{W}=(w_{ij})$ is a neighborhood matrix for areal units, which can be defined as

$$w_{ij} = \begin{cases} 1 \text{ if subregions } i \text{ and } j \text{ share a common boundary, } i \neq j \\ 0 \text{ otherwise} \end{cases}$$

$\mathbf{D}_w = \mathrm{diag}(w_{i+})$ is a diagonal matrix with (i,i) entry equal to $w_{i+} = \sum_j w_{ij}$.

C. Bayesian Models

The model usually consists of three levels, or stages of hierarchies. At the first stage, a linear model is set up given fixed and random effects; at the second stage, the distributions of fixed and random effects are specified given the variance components; finally, at the last stage, prior distributions are assigned to the variance components.

Reference [18] briefly describes the basic elements of Bayesian inferences. Suppose that $\mathbf{y}$ is a vector of observations, $\mathbf{y}=(y_1,...,y_m)^T$, and $\boldsymbol{\theta}$ is a vector of parameters, $\boldsymbol{\theta}=(\theta_1,...,\theta_k)^T$.

Let $\mathrm{f}(\mathbf{y}\,|\,\boldsymbol{\theta})$ represent the conditional probability density function of $\mathbf{y}$ given $\boldsymbol{\theta}$, and $\pi(\boldsymbol{\theta})$ is a prior distribution for $\boldsymbol{\theta}$. Then, the posterior probability density function of $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}\,|\,\mathbf{y}) = \frac{\mathrm{f}(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int \mathrm{f}(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{3}$$

The goal of Bayesian inference is to get the posterior distribution. Particularly, some numerical summaries are obtained from the posterior distribution. For instance, a Bayesian point estimator for a univariate θ is often obtained as the posterior mean:

$$E(\theta \mid \mathbf{y}) = \int \theta \pi(\theta \mid \mathbf{y}) d\theta = \frac{\int \theta f(\mathbf{y} \mid \theta)\pi(\theta) d\theta}{\int f(\mathbf{y} \mid \theta)\pi(\theta) d\theta}. \quad (4)$$

III. Application

The rice and cassava yields in 19 northeastern provinces of Thailand have been extracted from the Office of Agricultural Economics [1] from 2002 to 2011 (120 months). The data are divided into 2 parts, the fitting part consisting of 108 months data and the validation part consisting of the last 12 months data. The proposed model which is a special case of LMM is applied to those data and it is expressed as follows.

Let y_{ikt} be the agricultural yields in province $i, i = 1,...,19$, product type $k, k = 1$ for rice and $k = 2$ for cassava, and month $t, t = 1,...,120$.

$$y_{ikt} = V_k + \phi_{ik} + A_{kt} + \varepsilon_{ikt}, \quad (5)$$

$$y_{ikt} \mid V_k, \phi_{ik}, A_{kt} \sim N(\mu_{ikt}, \sigma^2),$$

where $\mu_{ikt} = V_k + \phi_{ik} + A_{kt}$ and $\varepsilon_{ikt} \sim N(0, \sigma^2)$. V_k are the product type random effects, ϕ_{ik} are the area-product type spatial effects, A_{kt} are the time-product type random effects, and ε_{ikt} are province-product type-time random effects. The estimated μ_{ikt} are used for prediction.

A. Model Estimation

We use a Bayesian method via Gibbs sampling MCMC in OpenBugs software [19] for parameter estimation.

For Bayesian setting, we assume priors as follows.

$$V_k \sim N(0, \sigma_v^2),\ \sigma_v^2 \sim \text{InvGamma}(0.005, 0.005)$$

$$\begin{pmatrix} \phi_{i1} \\ \phi_{i2} \end{pmatrix} \mid \boldsymbol{\varphi}_{-(i1,i2)} \sim \text{MCAR in (2)}$$

$$A_{kt} \sim N(0, \sigma_A^2),\ \sigma_A^2 \sim \text{InvGamma}(0.005, 0.005)$$

$$\sigma^2 \sim \text{InvGamma}(0.005, 0.005)$$

B. Model Comparison

The proposed model is compared with the traditional exponential smoothing model and our previous model [10] using CAR model. The Gibbs sampling MCMC are run for 11,000 iterations, with burn-in of 1,000. We assess MCMC convergence of all model parameters by visual analysis of history and Kernel density plots.

IV. Results

The visual analysis is used for MCMC convergence diagnostics [19]. The trace plots for some estimated means are shown in Figs. 1-4 and the kernel density plots are shown in Figs. 5-8. The chains moving around the parameter spaces and the densities looking like their distributions indicate that each parameter is converged to a stationary density.

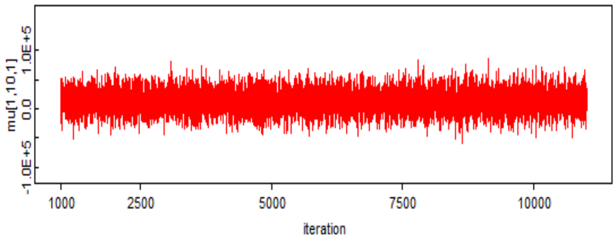

Fig. 1 History plot of the estimated mean for rice yield in January of Ubon Ratchathani

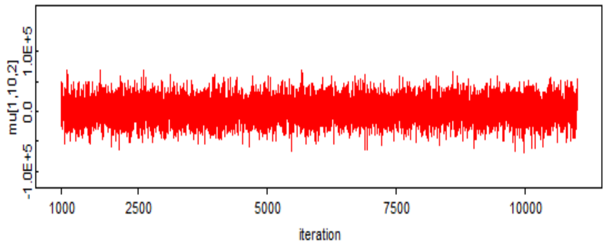

Fig. 2 History plot of the estimated mean for rice yield in February of Ubon Ratchathani

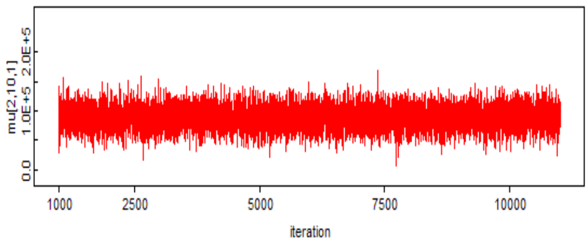

Fig. 3 History plot of the estimated mean for cassava yield in January of Ubon Ratchathani

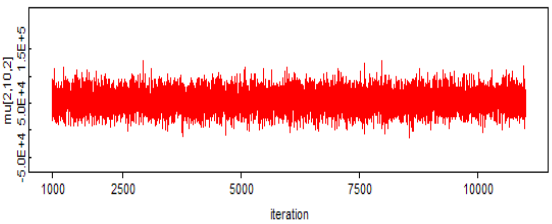

Fig. 4 History plot of the estimated mean for cassava yield in February of Ubon Ratchathani

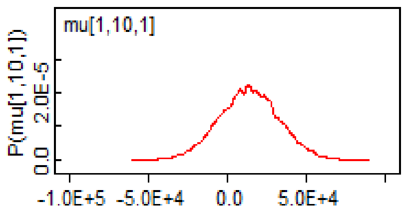

Fig. 5 Kernel density plot of the estimated mean for rice yield in January of Ubon Ratchathani

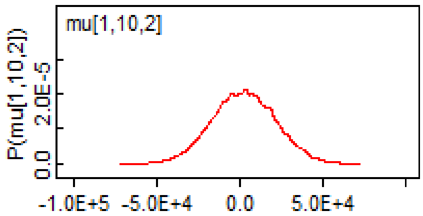

Fig. 6 Kernel density plot of the estimated mean for rice yield in February of Ubon Ratchathani

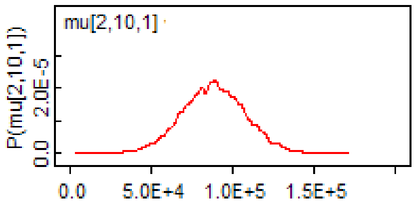

Fig. 7 Kernel density plot of the estimated mean for cassava yield in January of Ubon Ratchathani

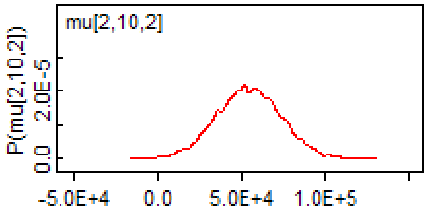

Fig. 8 Kernel density plot of the estimated mean for cassava yield in February of Ubon Ratchathani

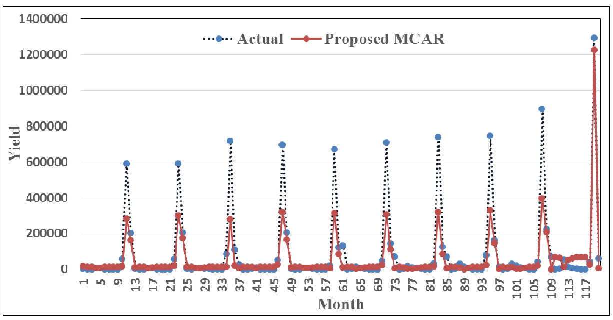

Fig. 9 Actual and predicted values of rice yield in Ubon Ratchthani province

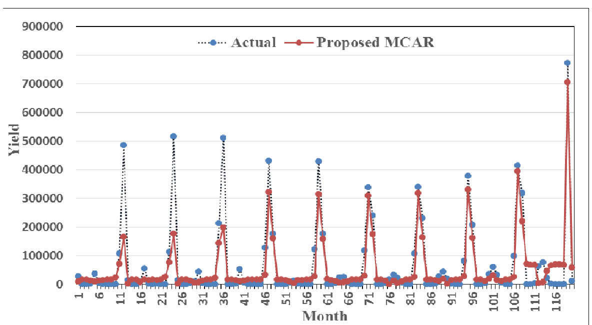

Fig. 10 Actual and predicted values of rice yield in Khon Kean province

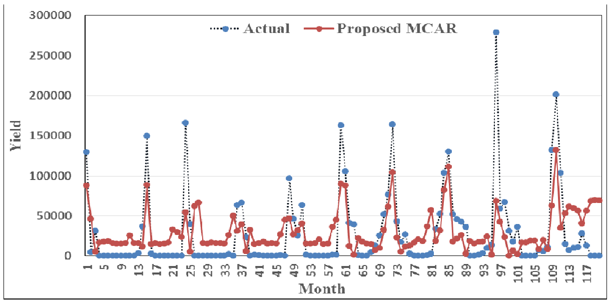

Fig. 11 Actual and predicted values of cassava yield in Ubon Ratchthani province

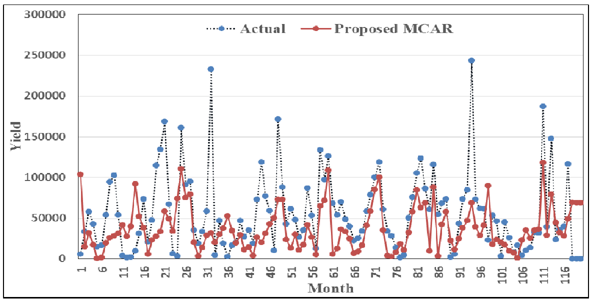

Fig. 12 Actual and predicted values of cassava yield in Khon Kaen province

Using the mean absolute error (MAE) criterion, for the rice yields, the performance of the proposed model compared to the simple exponential and the CAR models is shown in Table I. It can be seen that, in the fitting part, the proposed model has a better performance than the exponential and the CAR models in most provinces. For the validation part, the proposed model is superior to the other models in more than half of the number of provinces. The predicted and actual values of rice and cassava yields are presented in Figs. 9-12.

V. Discussion

The LMM with MCAR spatial effects for spatial time series data is proposed. It takes into account the spatial correlations following the first law of geography stating that "Everything is related to everything else, but near things are more related than distant things" [20]. It also accounts for the correlations between product types as well as products and time. Compared to the simple exponential smoothing and our previous CAR models, the proposed model has a better performance in most provinces in both fitting part and validating part. The most advantage of the proposed model is that it can predict several product yields at each province at the same time. The limitation of this study is using secondary data which causes problems of verification. For further study the proposed model can be extended to include trend and seasonal components.

TABLE I
PERFORMANCE OF THE PROPOSED MCAR, CAR, AND ES MODELS FOR RICE YIELDS

Province	Model	MAE	
		Fitting	Validation
Loei	MCAR	43,760.09	68,576.67
	CAR	**19,663.52**	**23,324.83**
	ES	21,182.73	24,979.32
Nong Bua Lam Phu	MCAR	**32,741.48**	68,987.50
	CAR	38,540.08	44,978.08
	ES	36,668.33	**43,289.07**
Udon Thani	MCAR	**44,974.35**	**68,970.00**
	CAR	78,898.44	99,315.83
	ES	74,148.97	98,556.02
Nong Khai	MCAR	**34,566.30**	68,716.67
	CAR	39,589.53	**47,905.42**
	ES	40,616.49	53,840.55
Sakon Nakhon	MCAR	**30,262.69**	**68,781.67**
	CAR	71,723.71	83,601.67
	ES	75,126.66	90,544.82
Nakhon Phanom	MCAR	**34,867.41**	68,503.33
	CAR	39,743.58	**52,045.58**
	ES	39,952.72	58,702.04
Mukdahan	MCAR	46,269.07	68,607.50
	CAR	**15,171.42**	**21,645.50**
	ES	16,085.04	23,359.88
Yasothon	MCAR	**31,437.13**	68,802.50
	CAR	40,910.25	**53,510.50**
	ES	41,756.63	58,424.90
Amnat Charoen	MCAR	**31,961.30**	68,640.00
	CAR	37,985.24	**45,806.75**
	ES	39,371.74	49,915.28
Ubon Ratchathani	MCAR	**51,923.06**	**68,676.67**
	CAR	116,337.88	152,390.67
	ES	119,697.32	165,278.30
Si Sa Ket	MCAR	**41,355.09**	**68,463.33**
	CAR	106,546.71	140,177.33
	ES	107,194.27	159,423.68
Surin	MCAR	**52,275.19**	**68,817.50**
	CAR	128,916.47	152,628.25
	ES	132,952.98	166,862.99
Buri Ram	MCAR	**44,746.57**	**68,914.17**
	CAR	113,385.69	139,047.08
	ES	115,194.83	154,783.45
Maha Sarakham	MCAR	**23,225.65**	**69,092.50**
	CAR	76,426.60	100,950.25
	ES	78,221.27	106,847.28
Roi Et	MCAR	**44,513.70**	**68,938.33**
	CAR	113,450.50	132,047.50
	ES	115,850.40	147,864.24
Kalasin	MCAR	**32,384.81**	**68,933.33**
	CAR	61,366.79	77,158.75
	ES	62,183.73	83,444.63
Khon Kaen	MCAR	**37,123.80**	**68,550.83**
	CAR	85,891.24	103,153.17
	ES	88,218.96	112,170.31
Chaiyaphum	MCAR	**35,388.61**	68,837.50
	CAR	49,097.91	**59,031.58**
	ES	49,750.54	63,623.78
Nakhon Ratchasima	MCAR	**53,938.24**	**68,470.00**
	CAR	116,099.64	154,018.42
	ES	119,624.43	163,438.90

VI. CONCLUSION

The objective of this study is to propose an appropriate forecasting model for multivariate spatial time series data. The Bayesian inference, using the Gibb sampling MCMC method, in a LMM with MCAR spatial effects is considered. The proposed model is applied to rice and cassava yields data in 19 Northeastern provinces of Thailand in 2002 to 2011. Using the MAE criterion, the proposed model has a better performance than the simple exponential model and the CAR model from our previous study in most of provinces. The advantage of the proposed LMM with MCAR spatial effects is that it can predict several product yields in several areas at the same time.

REFERENCES

[1] Office of Agricultural Economics, (2012, December 5). Agricultural Production. http://www.oae.go.th.
[2] P. M. Yelland, "Bayesian forecasting of part demand," *International Journal of Forecasting*, vol. 26, pp. 374–396, 2010.
[3] N. M. F. Rahman, "Forecasting of boro rice production in Bangladesh: An ARIMA approach," *Journal of the Bangladesh Agricultural University*, vol. 8, no. 1, pp. 103–112, 2010.
[4] P. Tongkhow and N. Kantanantha, "Bayesian models for time series with covariates, trend, seasonality, autoregression and outliers," *Journal of Computer Science*, vol. 9, no. 3, pp. 291–298, 2013.
[5] J. C. Wakefield, "Disease mapping and spatial regression with count data," *Biostatistics*, vol. 8, pp. 158–183, 2007.
[6] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society Series B*, vol. 36, no. 2, pp. 192–236, 1974.
[7] D. G. Clayton and J. Keldor. "Empirical Bayes estimates of age-standardized relative risks for use in disease mapping," *Biometrics*, vol. 43, pp. 671–691, 1987.
[8] J. Besag, J. York, and A. Molli, "Bayesian image restoration, with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*, vol. 43, pp. 1–21, 1991.
[9] M. Diaconoa, A. Castrignanob, A. Troccolic, D. De Benedettob, B. Bassod, and P. Rubino, "Spatial and temporal variability of wheat grain yield and quality in a Mediterranean environment: A multivariate geostatistical approach," *Field Crops Research*, vol. 131, pp. 49–62, 2012.
[10] P. Saengseedam, and N. Kantanantha, "Spatial time series forecasts based on Bayesian linear mixed models for rice yields in Thailand," in Proceedings of the International MultiConference of Engineers and Computer Scientists 2014, Hong Kong, 2014, pp. 1007–1012.
[11] B. T. West, K. B. Welch, and A. T. Galecki, *Linear mixed models: A practical guide to using statistical software*. New York: Chapman and Hall/CRC, 2007.
[12] B. P. Carlin and S. Banerjee, "Hierarchical multivariate CAR models for spatially correlated survival data," Bayesian Statistics 7. Oxford: Oxford University Press, pp. 45–64, 2003.
[13] H. Ma and B. P. Carlin, "Bayesian multivariate areal wombling for multiple disease boundary analysis," *Bayesian Analysis*, vol. 2, no. 2, pp. 281–302, 2007.
[14] ADB, (2012, June 22). The Rice Situation in Thailand. http://www.adb.org/sites/default/files/projdocs/2012/43430-012-reg-tacr-03.pdf
[15] I. Nation, (2008, April 16). Rice strain is cause of comparatively low productivity. http://nationmultimedia.com/2008/04/16/opinion/opinion_30070831.php
[16] T. Srinorakut, (2002, August 7). Research on High Grade Ethanol from Cassava for Decreasing of Chemical Importing. Research Project of Ethanol from Cassava. National Science and Technology Development Agency: NSTDA. http://www.itdoa.com/news_itda/science/doc_19.htm.
[17] K. Srirot, (2009, May 18). Situation of Raw Material for Ethanol Production of Thailand. http://www.cassava.org.
[18] P. Congdon, *Bayesian Statistical Modelling*. 2nd ed. New York: John Wiley and Sons, 2006.

[19] S. P. Brooks and G. O. Roberts, "Assessing convergence of Markov Chain Monte Carlo algorithms," *Statistics and Computing*, vol. 8, pp. 319–335, 1998.
[20] W. Tobler, "A computer movie simulating urban growth in the Detroit region," *Economic Geography*, vol. 46, no. 2, pp. 234–240, 1970.