Available online at www.isas.org.in/jisas
Journal of the Indian Society of Agricultural Statistics 73(3) 2019 181–188

Spatio-Temporal Models for Forecasting of Rice and Wheat Yields in India

Dipankar Mitra, Ranjit Kumar Paul, A.R. Udgata, A.K. Paul and L.M. Bhar

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 01 April 2019; Revised 22 April 2019; Accepted 30 April 2019

SUMMARY

The analysis of spatio-temporal data has been an increasingly demanding area of methodological and applied statistical research. These data are very common in many applications of agriculture. For analysis of spatial time series data Generalized Linear Mixed Models (GLMMs) are most commonly used. A special case of GLMM, Linear Mixed Model (LMM) is used to forecast continuous data. A modified version of LMM with spatial effects, trend and outliers for spatio-temporal time series data has been considered in this study. A linear trend, a binary method for outliers and a Multivariate Conditional Autoregressive (MCAR) model for spatial effects are adopted. A Bayesian method using Gibbs sampling in Markov Chain Monte Carlo (MCMC) is used for parameter estimation. The model is applied to forecast rice and wheat yields in India and compared with an LMM with MCAR, and a log transformed LMM with MCAR. It has been found that the modified LMM model is the most appropriate, using the Mean Absolute Error (MAE) criterion. The model performs well for fitting and validation for both rice and wheat yields.

Keywords: Forecasting, Gibbs sampling, Linear mixed model, Multivariate conditional autoregressive model, Spatio-temporal data.

1. INTRODUCTION

In recent times the analysis of spatio-temporal data has become an important branch of scientific research. Spatial time-series data are collected across both time and space. Thus, the data analysts should consider the correlations across the time and across the areas in order to capture the spatio-temporal effect. These kinds of data can be found in various fields such as agriculture, economics, medicine and the environment. The increased computational power helps to deal with such data. Annual yields of different agricultural crop collected by government in each area are typical examples of spatio-temporal data. These data sets naturally accommodate spatial as well as time series components such as trends, seasonality and outliers.

The spatio-temporal models are normally constructed by combining time series models with spatial models. Such models are usually based on Generalized Linear Mixed Models (GLMM). The common time series structures are linear trends and dummy seasonality, and one of the common spatial structures is the Conditional Autoregressive Model (CAR). For time series data, Yelland (2010) used a Bayesian statistical model to forecast the parts demand time series data for Sun Microsystems, Inc.; Box *et al*. (2007) proposed ARIMA and its component models which are the most important and widely used techniques in time-series analysis, Tongkhow and Kantanantha (2013) proposed a forecasting model that can detect trend, seasonality, auto- regression and outliers in time series data related to some covariates. For spatial data, the spatial effects can be done in a number of ways (Wakefield, 2006); one of the common approaches is a Conditional Auto- Regressive model (CAR) first introduced by Besag (1974); Clayton and Keldor (1987) extended the CAR model and proposed empirical Bayesian methods building from Poisson regression with random intercepts defined with CAR spatial correlations. 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 MCAR

Corresponding author: Ranjit Kumar Paul
E-mail address: ranjitstat@gmail.com

proposed by Carlin and Banerjee (2003) 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. A Bayesian method using the Markov Chain Monte Carlo (MCMC) procedure is extensively applied for parameter estimation in this complex model.

For spatial time-series data, Diaconoa *et al*. (2012) used geo-statistical approach to analyze the yearly data collected from 100 geo-referenced locations and studied the spatial and temporal variability of attributes related to the yield and quality of durum wheat production; Saengseedam and Kantanantha (2014) 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 GLMMs. A special case of GLMM, LMM is used to forecast when the data are continuous. LMMs are usually used when responses are correlated data which may be due to repeated measurements on each subject over time (West *et al*., 2014). The LMMs allow fixed effects and spatial effects to be included. Saengseedam and Kantanantha (2017) proposed a LMM with spatial effects, trend, seasonality and outliers for spatio-temporal time series data and applied this model to forecast rice and cassava yield in Thailand.

Accurate and reliable forecasting of crop yield is essential for crop production, marketing, storage, and transportation decisions and also helps in managing the risk associated with these activities (Lee, 1999; Potgieter *et al*., 2005). Foodgrain production forecasting is central to make food policy decisions. Almost all major food security programmes, such as imports, strategic food reserves, granting of licenses to public firms to import and export, local procurement by the government and donors, emergency food assistance and distribution of stored grains rely on crop forecasts for strategic planning. The availability of yield forecast information of different crops over different locations is highly sought after by industry and government agencies to use as an objective tool to assist in decision-making.

Rice is the staple food for more than half of the total population and its importance in the country cannot be negated. India is not only leading consumer of rice crop but also it is the second largest producer in the world (112.91MT), lagging only behind China according to the annual report by Public Investigation Bureau of 2017-18. Next to rice, wheat is the most important food-grain of India and is the staple food of millions of Indians, particularly in the northern and north-western parts of the country. The production of wheat in the country has increased significantly from 75.81 MT in 2006-07 to an all-time record high of 99.70 MT in 2017-18. Therefore, in rainfall-dependent and highly variable agriculture systems, forecasting of rice and wheat yields is important from consumer as well as producer point of view. In the present study, the modified LMM is applied to forecast the rice and wheat yields in India over different states.

2. METHODOLOGY

2.1 Linear Models (LMs) and Linear Mixed Models (LMMs)

Linear model for the observation y_{ij} with mean μ_{ij} is defined as

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (1)$$

where μ, α_i, β_j are fixed unknown constants and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

A variant of LMs is where parameters in an LM are treated not as constants but as (realizations of) random variables. Model (1) under the assumption that μ, β_j are fixed unknown constants; $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$, is called as Linear Mixed Model (LMM).

2.2 Generalized Linear Models (GLMs) and Generalized Linear Mixed Models (GLMMs)

LMs and LMMs are extended to Generalized Linear Models (GLMs) and Generalized Linear Mixed Model (GLMMs). The essence of this generalization is of two-fold: 1. the data are not necessarily assumed to be normally distributed and 2. that the mean is not taken as a linear combination of parameters but that function of mean.

Generalized linear models (GLMs) represent a class of fixed effects regression models for several types of dependent variables (i.e., continuous, dichotomous, counts). Common GLMs include linear regression, logistic regression, and Poisson regression. There are three specifications in a GLM. First, the linear predictor, denoted as η_i, of a GLM is of the form

$$\eta_i = \boldsymbol{x}_i' \boldsymbol{\beta} \quad (2)$$

where $\boldsymbol{x}_i$ is the vector of regressors for unit i with fixed effects $\boldsymbol{\beta}$. Then, a link function g(·) is specified which converts the expected value μ_i of the outcome variable Y_i (i.e., μ_i = E[Y_i]) to the linear predictor η_i, g(μ_i) = η_i.

Fixed effects models are based on the assumption that all observations are independent of each other. But this assumption is not appropriate for analysis of several types of correlated data structures, in particular, for clustered and/or longitudinal data. For analysis of such multi-level data, random cluster and/or subject effects can be added into the regression model to account for the correlation of the data and the resulting model is called as **Generalized Linear Mixed Models (GLMMs).**

Let *i* denote the level-2 units (e.g., subjects) and let *j* denote the level-1 units (e.g., nested observations). Assume there are $i = 1,2,\dots,N$ subjects (level-2 units) and $j = 1,2,\dots,n_i$ repeated observations (level-1 units) nested within each subject. A random-intercept model, which is the simplest mixed model, augments the linear predictor with a single random effect for subject *i*,

$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \upsilon_i \tag{3}$$

where υ_i is the random effect (one for each subject) distributed as N(0, σ_υ^2).

2.3 Linear Mixed Models for Time -Series Data

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

$$y_{it} = \boldsymbol{X}_{it}'\boldsymbol{\beta} + \boldsymbol{Z}_{it}'\boldsymbol{b}_i + \varepsilon_{it}, i = 1,2,\dots,m; t = 1,2,\dots,T \tag{4}$$

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

2.4 Conditional Autoregressive (CAR) model

A key assumption in GLM models is that each response variable is independent from all others, after accounting for the covariate effects. When the response variables are collected in space, it is very common for the residuals resulting from a regression or GLM analysis to show spatial autocorrelation. Instead of assuming independence, spatial statistical models directly account for spatial autocorrelation through modelling the covariance matrix Σ of the residuals as a function of the locations where the response variable, contained in the vector y, were collected. For example, when the observations are point-referenced (i.e., each y was collected at a location with known GPS coordinates), geo-statistical methods are often used (Turner *et al*., 1991). For a real data such as quadrats or pre-specified spatial polygons, one could use a geo-statistical model, such as the exponential covariance model, but this requires specifying a point to represent each areal unit, for example the centroid of each areal unit. While this is possible, another class of spatial covariance models has been developed specifically to take advantage of the characteristics of areal data, the autoregressive spatial models. In these models, a network of connections between neighboring areal units (like- city, states etc.) is specified, and spatial dependence is specified through a model that conditions on observations at neighboring locations.

Let Y_i is the observed number of cases of a certain disease in region *i,* $i = 1,2,\dots,p,$ and modeled by a spatial Poisson regression of the form

$$Y_i|\mu_i \sim P(E_i e^{\mu_i}) \tag{5}$$

where Y_i's are independent and $\mu_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \theta_i + \phi_i$. The $\boldsymbol{x}_i$'s are explanatory variables, representing region-level spatial covariates, θ_i captures region-wide heterogeneity via the normal prior $\theta_i \sim iidN(0,1/\tau)$, where the precision τ controls the magnitude of θ_i.

Finally, the ϕ_i are the parameters that make this a truly spatial model by capturing regional clustering. While common geo-statistical (e.g. exponential, spherical, Gaussian, etc.) models could be used as priors here, the most common approach is to adopt a conditionally autoregressive (CAR) prior of the form

$$\phi_i|\boldsymbol{\phi}_{-i} \sim N(\bar{\phi}_i, 1/(\lambda m_i))$$

where $\boldsymbol{\phi}_{-i} = (\phi_1,\dots,\phi_{i-1},\phi_{i+1},\dots,\phi_p)^T, m_i$ is the of "neighbours" (adjacent regions) of region *i,* and $\bar{\phi}_i = m_i^{-1}\sum_{j\ adj\ i}\phi_j$, the average of the neighbouring values. This model corresponds to a multivariate normal distribution $\boldsymbol{\phi} = (\phi_1,\dots,\phi_p)^T$ with a less-than-full-rank covariance matrix.

It is more convenient to deal with matrix notations. Let, for brevity, consider a vector $\boldsymbol{\phi} = (\phi_1,\dots,\phi_p)^T$

of p components that follows a multivariate Gaussian distribution with mean 0 and $\boldsymbol{B}$ as the inverse of the dispersion matrix, so that $\boldsymbol{B}$ is $p \times p$ and positive definite. The density for $\boldsymbol{\phi}$ is given by

$$p(\boldsymbol{\phi}) = (2\pi)^{-p/2}|\boldsymbol{B}|^{1/2} exp\left(-\frac{1}{2}\boldsymbol{\phi}^T \boldsymbol{B}\boldsymbol{\phi}\right) \quad (6)$$

For such a distribution, it is of interest to look at the conditional distribution of a particular component given the remaining components. In terms of the elements of the matrix $\boldsymbol{B} = ((b_{ij}))$, it is well-known from normal theory that ϕ_i has the full conditional distribution

$$p(\phi_i|\boldsymbol{\phi}_{-i}) \propto \exp\left(-\frac{1}{2}b_{ii}(\phi_i - \sum_{j\neq i}\frac{-b_{ij}}{b_{ii}}\phi_j)^2\right) \quad (7)$$

which is a $N(\sum_{j\neq i}\frac{-b_{ij}}{b_{ii}}\phi_j, \frac{1}{b_{ii}})$ distribution.

Suppose that the full conditional distribution of ϕ_i is specified as $N(\sum_{j\neq i} c_{ij}\,\phi_j, \sigma_i^2)$ so that

$$(\phi_i|\boldsymbol{\phi}_{-i}) \propto \exp\left(-\frac{1}{2\sigma_i^2}(\phi_i - \sum_{j\neq i} c_{ij}\,\phi_j)^2\right) \quad (8)$$

when compared with (7), this reveals that $c_{ij} = \frac{-b_{ij}}{b_{ii}}$ and $\frac{1}{b_{ii}} = \sigma_i^2$. Now, form a matrix $\boldsymbol{C}$ with $c_{ii} = 0$ and $c_{ij} = \frac{-b_{ij}}{b_{ii}}$, and another matrix $\boldsymbol{M} = Diag(\sigma_i^2)$ (so that $\boldsymbol{M}^{-1} = Diag(b_{ii})$). Then $\boldsymbol{B}$ is related to $\boldsymbol{M}$ and $\boldsymbol{C}$ as

$$\boldsymbol{B} = \boldsymbol{M}^{-1}(\boldsymbol{I} - \boldsymbol{C}) \quad (9)$$

where $\boldsymbol{I}$ is the identity matrix. Thus the joint distribution of $\boldsymbol{\phi}$ is $N(\boldsymbol{0}, (\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M})$.

2.5 Multivariate Conditional Autoregressive (MCAR) Model

Let $\boldsymbol{\Phi}^T = \left(\phi_1^T, \phi_2^T, \ldots, \phi_p^T\right)$ where $\boldsymbol{\Phi}$ is $np \times 1$ with each ϕ_i being an n-dimensional vector. Consider a multivariate Gaussian distribution for $\boldsymbol{\Phi}$ of the form

$$p(\boldsymbol{\Phi}) = (2\pi)^{-np/2}|\boldsymbol{B}|^{1/2} exp\left(-\frac{1}{2}\boldsymbol{\Phi}^T \boldsymbol{B}\boldsymbol{\Phi}\right) \quad (10)$$

Here $\boldsymbol{B}$ is a $np \times np$ symmetric positive definite matrix. In fact it is easier to visualize $\boldsymbol{B}$ as a $p \times p$ block matrix with $n \times n$ blocks B_{ij}. Analogous to the univariate situation, it follows that the full conditional distribution.

2.6 Extension of an LMM for spatio-temporal data

2.6.1. Adding spatial effects

Model (4) can be further adjusted and extended to include spatial random effects for spatio-temporal time series data. For a univariate, a conditional autoregressive (CAR) model is a common approach to explain the spatial correlation. For a multivariate, the CAR is modified to be MCAR model; therefore, Model (4) with an MCAR has the following form,

$$y_{ijk} = \alpha_k + b_{kt} + \phi_{ik} + \varepsilon_{ikt} \quad (11)$$

where y_{ijk} is the response at area i of the product k at time t, α_k is a random effect representing the baseline of product k, b_{kt} is a random effect of representing the baseline of product k and time t, ε_{ikt} is a random error, and ϕ_{ik} follows MCAR model whose details are as follows.

Let $\boldsymbol{\phi}^T = (\phi_1^T, \phi_2^T)$ where $\phi_1^T = (\phi_{11}, \ldots, \phi_{m1})$, $\phi_2^T = (\phi_{12}, \ldots, \phi_{m2})$ and m is the number of areal units. The bivariate spatial random effect is defined as the conditional distribution,

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

where $\boldsymbol{\phi}_{-(i1,i2)}$ is the collection of all ϕ_{il} except ϕ_{i1} and ϕ_{i2}. $\bar{\phi}_{i1} = \sum_l \frac{w_{il}\phi_{l1}}{w_{i+}}$ and $\bar{\phi}_{i2} = \sum_l \frac{w_{il}\phi_{l2}}{w_{i+}}$ are the averages of the random for area i's neighbours specific to variables ϕ_{i1} and ϕ_{i2}, respectively. $\boldsymbol{\Lambda}$ is a scaled conditional precision for (ϕ_{i1} and ϕ_{i2}) and w_{i+} is a scale parameter.

$\boldsymbol{\Lambda}$ is common for all areas $i = 1, \ldots, m$; therefore, it controls the conditional precision for each pair of variables at the same site averaged over all areas. Let $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^{-1}$ then $(w_{i+}\boldsymbol{\Lambda})^{-1} = \frac{1}{w_{i+}}\boldsymbol{\Sigma}$ which is the conditional covariance matrix with $\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$, the conditional correlation between ϕ_{i1} and ϕ_{i2}, $i = 1, \ldots, m$. For MCAR, the multivariate joint distribution

$$p(\phi) \propto exp\left(-\frac{1}{2}\boldsymbol{\phi}^T[\boldsymbol{\Lambda} \otimes (\boldsymbol{D}_w - \boldsymbol{W})]\boldsymbol{\phi}\right) \quad (12)$$

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

$$w_{ij} = \begin{cases} 1, if\ \ regions\ \ i\ \ and\ \ j\ \ are\ \ neighbours \\ 0,\ \ otherwise \end{cases}$$

And $\boldsymbol{D}_w = Diag(w_{i+})$ is a diagonal matrix whose (i, i) entry equals to $w_{i+} = \sum_j w_{ij}$.

2.6.2. Adding trend and outlier components

Model (11) can be extended to include trend, seasonal and outlier components. A linear trend and seasonal dummy variables are common methods for capturing trend and seasonality, respectively. Following Tongkhow and Kantanantha (2013) and Yell and (2010) binary selection method is applied to capture outliers. The LMM for a multivariate time series with a linear trend and outliers can be expressed as follows:

$$y_{ijk} = \alpha_k + b_{kt} + \phi_{ik} + At + O_{ikt} + \varepsilon_{ikt} \quad (13)$$

where At is a linear trend and O_{ikt} is an outlier.

3. AN ILLUSTRATION

3.1 Data description and model specification

The modified LMM with spatial effect, trend and outlier is applied to rice and wheat yields (Unit: kgs) in 9 states of India (Andhra Pradesh, Assam, Bihar, Haryana, Karnataka, Odisha, Punjab, Uttar Pradesh, West Bengal) which are shown in Fig 1. Data are obtained from the Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare, GoI (https://eands.dacnet.nic.in website) from the year 1966 to 2015 (49 years). The data are divided into two parts, the first 43 years are used for model fitting and the last 6years are used for model validation. The neighbourhood relationship among the selected states can be shown using the matrix $\boldsymbol{W}$ which is given below

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

where the $(ij)^{th}(i, j = 1,2,\ldots,9)$ element takes on value 1 if state j shares a common geographical boundary with state i, otherwise zero.

Let z_{ikt} be the agricultural yields in state $i, i = 1,\ldots,19$, product type $k, k = 1$ for rice and $k = 2$ for wheat, and year t, t = 1, . . . , 49. The collected data are transformed using a natural logarithmic function to make the data more normally distributed (Fletcher *et al*., 2005). The results of Anderson-Darling test for normality of both original and logarithmic transformed data of rice and wheat yield are shown in Table 1 and 2 respectively. The test results indicate that the transformed data are more normally distributed for both yield data.

$$y_{ikt} = \ln(z_{ikt} + 1)$$

$$y_{ikt} = \alpha_k + b_{kt} + \phi_{ik} + At + O_{ikt} + \varepsilon_{ikt}$$

$$y_{ikt}|\alpha_k, b_{kt}, \phi_{ik}, O_{ikt} \sim N(\mu_{ikt}, \sigma^2)$$

where $\varepsilon_{ikt} \sim N(0, \sigma^2)$, α_k is a product type random effect, b_{kt} is product type and time effect, ϕ_{ik} is the area-product type spatial effect, At is a linear trend, O_{ikt} is an outlier and ε_{ikt} is a state-product type-time random effect. The estimated μ_{ikt} is used for predicting y_{ikt} and exp ($^{\mu_{ikt}}$) is used for predicting z_{ikt}.

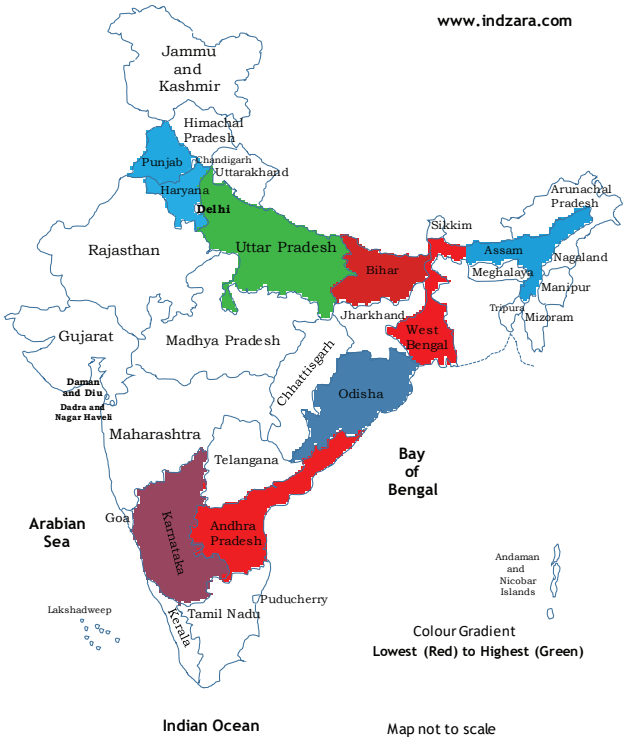

Fig. 1. Selected states of India

3.2 Model comparison

The performance of the modified LMM is compared with the LMM with MCAR, and the model with log transformation, using the Mean Absolute Error (MAE) criterion. Mathematically, MAE is defined as

Table 1. Anderson Darling test for normality of rice yields

States	Original data		Transformed data	
	statistic	p-value	statistic	p-value
AP	0.71	0.061	0.97	0.015
AS	1.04	0.190	0.87	0.024
BR	0.44	>0.250	0.33	0.070
HR	1.58	<0.005	2.69	<0.005
KA	0.69	0.072	0.38	<0.005
OD	0.92	0.046	0.76	0.016
PB	4.88	0.015	0.97	<0.005
UP	1.51	0.015	0.97	<0.005
WB	1.45	0.024	0.86	<0.005

(*AP: Andhra Pradesh, AS: Assam, BR: Bihar, HR: Haryana, KA: Karnataka, OD: Odisha, PB: Punjab, UP: Uttar Pradesh, WB: West Bengal)

Table 2. Anderson Darling test for normality of wheat yields

States	Original data		Transformed data	
	statistic	p-value	statistic	p-value
AP	0.89	0.022	1.15	<0.005
AS	2.46	<0.005	4.10	<0.005
BR	0.48	0.226	1.12	0.006
HR	1.23	<0.005	1.46	<0.005
KA	0.14	>0.250	0.90	0.021
OD	0.44	>0.250	0.51	0.095
PB	1.07	0.008	1.38	<0.005
UP	0.97	0.015	1.37	<0.005
WB	0.84	0.03	2.39	<0.005

$$\text{MAE} = \frac{1}{h}\sum_{t=1}^{h}|y_t - \hat{y}_t|, t = 1,2,\ldots,h$$

where y_t is the actual observation for the time t and $\hat{y}_t$ is the forecast value of the series for the same time; h denotes the forecast horizon.

3.3 Results

Gibbs sampling MCMC is used to estimate parameter and it is found that MCMC for each parameter is converged. MAE values of the modified LMM, MCAR model with log transformation and MCAR model are listed for both fitting and validation dataset in Table 3 and Table 4 for rice and wheat yields respectively. Smallest MAE values are marked bold. From Table 3 it is clear that in case of rice yield the modified LMM has a better performance in 6 out of 9 states (66.67%) when compared to MCAR, but lesser performance when compared to MCAR with log transformation in 4 out of 9 states (44.44%) in the fitting part. However, in the validation part, the modified LMM is superior to the MCAR and MCAR with log transformation in all 9 of 9 states (100%). Similarly, Table 4 indicates that for wheat yield in the fitting part, the modified LMM has a better performance in 5 out of 9 states (55.56%) when compared to MCAR, but lesser performance when compared to MCAR with log transformation in 3 out of 9 states (33.33%). However, in the validation part, the modified LMM is superior to the MCAR and MCAR with log transformation in all 9 of 9 states (100%).

Table 3. Performance of the modified LMM model compared to MCAR for rice yields

States	Model	MAE (kgs)	
		Fitting	Validation
AP	Modified LMM	**975.29**	**733.02**
	MCAR with log	1,230.25	1,425.72
	MCAR	1,714.54	2,055.83
AS	Modified LMM	**677.33**	**848.03**
	MCAR with log	1,726.94	1,331.65
	MCAR	2,903.18	1,167.07
BR	Modified LMM	1,203.28	**225.69**
	MCAR with log	**798.91**	634.91
	MCAR	831.66	712.70
HR	Modified LMM	1,142.12	**895.35**
	MCAR with log	**566.18**	1,028.94
	MCAR	1,504.72	2,055.10
KA	Modified LMM	**902.79**	**965.69**
	MCAR with log	1,959.24	1,290.62
	MCAR	2,164.87	3,730.02
OD	Modified LMM	2,051.55	**726.97**
	MCAR with log	**974.39**	1,662.88
	MCAR	1,029.12	1,617.60
PB	Modified LMM	379.19	**601.82**
	MCAR with log	**215.02**	1,423.65
	MCAR	649.82	2,712.35
UP	Modified LMM	872.61	**535.83**
	MCAR with log	523.24	612.37
	MCAR	**453.55**	685.00
WB	Modified LMM	**354.95**	**982.24**
	MCAR with log	673.44	842.85
	MCAR	1,068.57	1,723.01

The parameter estimates are shown in Tables 5–6. From Table 5, it can be found that there are spatial random effects which vary in each state, for example, the spatial effect of rice in AP (Spatial [1,1]) is 1.21, meaning that it increases rice yield by 1.21 kgs. The spatial effect of wheat yield in AP (Spatial [2,1]) is

Table 4. Performance of the modified LMM model compared to MCAR for wheat yields

States	Model	MAE (kgs)	
		Fitting	Validation
AP	Modified LMM	**1,076.24**	**653.52**
	MCAR with log	1,442.84	1,738.01
	MCAR	1,714.44	2,165.14
AS	Modified LMM	1,781.47	**1,049.63**
	MCAR with log	**1,456.53**	1,891.78
	MCAR	3,123.06	1,380.12
BR	Modified LMM	**715.49**	**815.08**
	MCAR with log	976.25	1,075.33
	MCAR	1,521.42	1,015.83
HR	Modified LMM	1,045.82	**525.05**
	MCAR with log	**553.48**	953.56
	MCAR	661.64	1070.99
KA	Modified LMM	**912.79**	**605.19**
	MCAR with log	1407.53	1,524.50
	MCAR	2,154.77	2,130.02
OD	Modified LMM	2,101.59	**1,229.07**
	MCAR with log	**1,456.18**	1,531.54
	MCAR	1,909.92	1,607.66
PB	Modified LMM	1,079.89	**791.22**
	MCAR with log	887.83	1,536.77
	MCAR	**719.53**	1,729.55
UP	Modified LMM	572.61	**836.43**
	MCAR with log	**306.85**	1,115.43
	MCAR	893.65	1,255.90
WB	Modified LMM	1,153.75	**684.64**
	MCAR with log	**735.73**	718.09
	MCAR	928.69	1,029.81

−1.32, meaning that it lessens wheat yield for 1.32kgs. The estimated A is 58.17, indicating that if time is increased by one year, the amount of crop yield per state per year will decrease by 58.17kgs. The estimated α1 (342.60) and α2 (413.67) are the baselines of the rice and the wheat yields per state per year, respectively, when other factors are not considered.

From Table 6, there are outliers which vary in each product-state, for example, the outlier of rice of AP (O[1,1]) is 0.02, meaning that it? increases rice yield by 0.02 kgs. The outlier of wheat yield of AP (O[2,1]) is 0.02, meaning that it increases wheat yield by 0.02 kgs.

4. CONCLUSIONS

Forecasting of crop yields of various agricultural commodities can be done using spatio-temporal data.

Table 5. Parameter estimates of spatial, trend and product type effects for rice and wheat yields

Parameter	Mean	Standard error	95% Credible interval	
			Lower bound	Upper bound
Spatial [1,1]	1.21	0.13	0.96	1.46
Spatial [1,2]	-0.18	0.12	-0.42	0.06
Spatial [1,3]	-1.67	0.12	-1.91	-1.43
Spatial [1,4]	-0.50	0.14	-0.77	-0.23
Spatial [1,5]	1.02	0.12	0.78	1.26
Spatial [1,6]	-0.65	0.12	-0.89	-0.41
Spatial [1,7]	0.82	0.13	0.57	1.07
Spatial [1,8]	-1.25	0.12	-1.49	-1.01
Spatial [1,9]	0.93	0.13	0.68	1.18
Spatial [2,1]	-1.32	0.14	-1.92	-1.38
Spatial [2,2]	0.87	0.14	1.14	1.14
Spatial [2,3]	-1.35	0.13	-1.10	-1.10
Spatial [2,4]	-1.95	0.14	-1.68	-1.68
Spatial [2,5]	0.82	0.14	1.09	1.09
Spatial [2,6]	-1.19	0.14	-0.92	-0.92
Spatial [2,7]	0.72	0.13	0.97	0.97
Spatial [2,8]	-1.54	0.13	-1.29	-1.29
Spatial [2,9]	1.01	0.12	1.25	1.25
Trend [A]	58.17	0.03	58.23	58.23
Rice [α_1]	342.60	2.73	347.95	347.95
Wheat [α_2]	413.67	1.97	417.53	417.53

Table 6. Parameter estimates of outlier effects for rice and wheat yields

Parameter	Mean	Standard error
O[1,1]	0.02	0.12
O[1,2]	0.03	0.13
O[1,3]	0.03	0.12
O[1,4]	0.02	0.16
O[1,5]	0.03	0.18
O[1,6]	0.02	0.12
O[1,7]	0.02	0.13
O[1,8]	0.01	0.16
O[1,9]	0.03	0.15
O[2,1]	0.02	0.17
O[2,2]	0.03	0.17
O[2,3]	0.01	0.14
O[2,4]	0.03	0.18
O[2,5]	0.02	0.16
O[2,6]	0.02	0.12
O[2,7]	0.03	0.14
O[2,8]	0.04	0.16
O[2,9]	0.04	0.17

These datasets with substantial spatial and temporal details raise the issue of development of suitable forecasting models. An extension of LMM with spatial effects, trends and outliers is used as an appropriate model for multivariate spatial time series data. The model is applied to forecast the yearly spatio-temporal rice and wheat yields data in India. A MCAR model is assumed for the spatial effects, a linear trend is employed for the temporal effects and a binary method is used of the outliers. A Bayesian method using Gibbs sampling in MCMC is adopted for parameter estimation. Using the MAE criterion, the results show that the modified LMM, which considers specific time series parameters, such as trends, outliers and also both random effect and spatial effect parameters, is the most effective in both training as well as validation part for both rice and wheat yields compared to MCAR model and MCAR with log transformation. The advantage of the modified LMM is that it can predict multivariate spatio-temporal dataset.

ACKNOWLEDGEMENTS

Authors are thankful to the anonymous reviewers for providing useful comments which helped improve the paper.

REFERENCES

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc.*. Series B (Methodological), 192-236.

Box, G.E.P., Jenkins, G.M. and Reinsel G.C. (2007). *Time-Series Analysis: Forecasting and Control,* 3rd edition. Pearson Education, India.

Carlin, B.P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatially correlated survival data. *Bayesian Statistics,* **7**, 45-64.

Clayton, D. andKaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 671-681.

Diacono, M., Castrignanò, A., Troccoli, A., De Benedetto, D., Basso, B. and Rubino, P. (2012). Spatial and temporal variability of wheat grain yield and quality in a Mediterranean environment: A multivariate geostatistical approach. *Field Crops Research*, **131**, 49-62.

Fletcher, D., MacKenzie, D. and Villouta, E. (2005). Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. *Environ. Eco. Statist.*, **12(1)**, 45-54.

Lee, R.Y. (1999). Modeling corn yields in Iowa using time-series analysis of AVHRR data and vegetation phenological metrics, Doctoral dissertation, University of Kansas, Lawrence, Kansas.

Potgieter, A.B., Hammer, G.L., Doherty, A. and De Voil, P. (2005). A simple regional-scale model for forecasting sorghum yield across North-Eastern Australia. *Agril. For. Met.*, **132(1-2)**, 143-153.

Saengseedam, P. and Kantanantha, N. (2014). Spatial time series models for rice and wheat yields based on Bayesian linear mixed models. *Int. J. Mathematical, Computational, Physical and Quantum Engineering*, **8**, 1046-1051.

Saengseedam, P. and Kantanantha, N. (2017). Spatio-temporal model for crop yield forecasting. *J. App. Statist.*, **44(3)**, 427-440.

Tongkhow, P. and Kantanantha, N. (2013). Bayesian models for time series with covariates, trend, seasonality, auto-regression and outliers. *J. Comp. Sci.*, **9(3)**, 291-298.

Turner, S.J., O'Neill, R.V., Conley, W., Conley, M.R. and Humphries, H.C. (1991). Pattern and scale: statistics for landscape ecology. *Quantitative methods in landscape ecology*. 17-49.

Wakefield, J. (2006). Disease mapping and spatial regression with count data. *Biostatistics*, **8(2)**, 158-183.

West, B.T., Welch, K.B. and Galecki, A.T. (2014). *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC.

Yelland, P.M. (2010). Bayesian forecasting of parts demand. *Int. J. Forecas.*, **26(2)**, 374-396.