A BAYESIAN APPROACH TO NORMAL REGRESSION MODELLING WITH AGGREGATE CLIMATIC DATA

[1]AGADA, P.O.[*], [2]OGWUCHE, O. I., [2]ABAGWALATU, C.F.

[1,*] Department of Mathematics/Statistics/Computer Science, University of Agriculture, Makurdi. Benue State, Nigeria.
[2] Department of Mathematics and Computer Science, Benue State University, Makurdi, Benue State, Nigeria.
*Corresponding author's email: gadexx@yahoo.com

ABSTRACT

Modeling the relationship between some climatic determinants in the wet or cropping season of Makurdi, Benue State, Nigeria requires data aggregation. The consequence of this aggregation is the reduction in data sample size. This poses serious challenge of lack of model-fit when the Classical Linear Regression Modeling Approach is employed. The Bayesian Normal Regression Modeling Approach was therefore employed in surmounting this problem. Three Bayesian Normal Regression Models were fitted namely; the Solar radiation, Total Rainfall Amount and the Number of Dry Days model. Each model result was compared with that of its Classical model counterpart. The discrepancies observed were blamed on the sample size reduction. The results of the Bayesian models revealed that; Solar radiation increases by 0.791 MJ/m^2 for each unit increase in the natural logarithm of Relative humidity. While it increases by 0.895 MJ/m^2 for each unit increase in the natural logarithm of Wind speed. Total Rainfall Amount increases by 66.280 mm for each unit increase in the natural logarithm of the Number of Dry Days while it increases by 2.912 mm for each unit increase in the natural logarithm of the Number of Wet Days. Furthermore, the Number of Dry Days decreases by 6.905 days for each unit increase in the natural logarithm of Number of Wet days, while it increases by 2.028 days for each unit increase in the natural logarithm of Total Rainfall Amount. The study affirmed that the cropping season climate of Makurdi is becoming warmer and drier and that the Bayesian Normal Regression Modelling approach be employed when sample size reduction due to data aggregation is a concern.

Keywords: Bayesian, Aggregate, Regression, Wet season

INTRODUCTION

Statistical modeling using individual – level data is paramount for obtaining accurate estimation most especially in Normal Regression Modeling (Moineddin and Urquia, 2014). However, there are some circumstances where individual – level data are not available due to confidentiality concerns or the research problem requires that data be aggregated. The latter is the case with modeling the relationship between some climatic determinants in the wet or cropping season of Makurdi, Benue State, Nigeria – the focus of this research. The wet season is the cropping season of the area whose inhabitants are predominantly farmers. These farmers can control for most productivity variables, except for climatic determinants, which have been stated to be out of their control due to the impact of climate change (Agada, Imande and Amedu, 2018). It therefore becomes paramount to understand how these climatic determinants interplay in the cropping season. This we believe can be achieved by modeling the relationship among them in the aforementioned season. Moreover, seasonal relationship among these determinants will require aggregating (averaging) data over each rainy month of the year (March – October) and over the entire data period of 34 years (1977-2010). The consequence of this aggregation is the reduction in the sample size to eight (8). This poses serious challenge of lack of model-fit or incorrect fit when the Classical Linear Regression Modeling Approach is employed. A way of

surmounting this obstacle is by employing the Bayesian Normal Regression Modeling Approach; a method proven to be efficient in handling cases of modeling with incomplete and sparse data (Taeryon, Mark, Ketra and Mitchell, 2008; Agada, Udoumoh and Gboga, 2019; Tripathi, Singh and Singh, 2019). This was implemented on the Windows Bayesian Inference Using Gibbs Sampling (WinBUGS) platform.

Some researchers (Moineddin *et al,* 2014) have demonstrated that Normal Regression Modeling using aggregate data yield similar regression coefficients as the individual-level data based models. Other researchers such as Aymen and Mohammed (2019) argued that aggregating both the dependent and independent variable values smoothes out variation at the individual level and therefore, synthetically inflate the correlation coefficient values. They also opined that, assuming that relationships observed for groups necessarily holds for individuals is an ecological fallacy, therefore models fit into group (aggregate) level data cannot be applied to datasets at the individual level. Other supporters of this claim include; Dias, Sutton, Welton and Aides (2013). Aymen and Mohammed (2019) therefore recommended that partial aggregation of data should be done to wade off the aforementioned problem. In the same vein, Robinson (1950) stated that the correlation between the properties of individual-level data have no bearing on the properties of groups or aggregates. Furthermore, Eric, Hanushek, John,

Jackson and Kain (1974) expanded the scope of Robinson's independent variables of Negro or White race to include Mexican and Indian born, as well as elementary school age population at the state level. They proved Robinson's conclusion wrong stating that individual- level data or not is not the problem, but improper model specification. This is because their expanded model better estimated the literacy rate of a race than that of Robinson. The authors emphasized that since individual-level data may not always be available due to confidentiality concerns, difficulties associated with the use of aggregate data should be dealt with through proper model specification.

A recent research in the study area by Agada *et al.* (2018) revealed that the area is becoming warmer and drier. The authors could neither identify and specify the climatic determinants nor quantify the magnitude of their effects on the warming and dry climate. This gap in literature we have been able to identify and intend to fill. From the aforementioned researches, the problems we must surmount for using aggregate data is that of inflated correlations among independent variables and that of improper model specification. This we have handled by proper data transformation and the use of correlation matrix of the climatic determinants in identifying independent determinants that correlate significantly with the dependent determinant and or with themselves. Pathan (2015) emphasized the use of the correlations among climatic variables in accomplishing the aforementioned task. The implication of this study to crop farming in the area, led us to fitting three normal regression models namely; the Solar Radiation, Total Rainfall Amount and the Number of Dry Days Model. Results of the Classical and Bayesian models employed in the study were compared. The rest of the paper is sectioned into; Methodology, Results, Discussion, Conclusion and Recommendations.

MATERIALS AND METHOD

Source of data and transformation

The data for this work is a 34 year (1977 – 2010) daily data on some climatic variables or determinants of the city of Makurdi, Nigeria. The climatic variables or determinants are; Solar radiation (MJm^{-2}day^{-1}), Relative humidity (%), Air temperature (^{0} C), Wind speed (m/s), Sunshine Duration (hrs), Total Rainfall Amount (mm), Number of wet days and Number of dry days. The number of Wet and Dry days were respectively determined by counting the number of times the rainfall amount of a particular day is greater than or equal to 0.85mm and less than 0.85mm respectively. This correspond to rainfall state transformations 1 and 0 respectively which were used for the counts. The minimum and maximum Air temperature and Relative humidity data primary were averaged and used as Air temperature and Relative Humidity data in the study. Using the entire dataset, the climatic variables or determinants were aggregated (averaged) over the rainy months of Makurdi, Nigeria (March, April, May, June, July, August, September and October). This was done to reflect their association or interplay in the wet or cropping season of the area.

The associations among these determinants were determined from their matrix of bivariate correlation coefficient. This was non – parametrically obtained using the Spearman Rank Correlation Statistic. The reason for adopting the Spearman Rank Correlation Statistic is that unlike its Pearson counterpart, it is not distribution dependent. The specification of the right variables in the Bayesian Normal model employed in this work is greatly eased by the correlation between the climatic determinants. This is because, predictor variables or determinants that relate significantly with the response variables can be easily identified and those that correlate with themselves identified for logarithm transformation in order to reduce multicolinearity.

Normal regression model specification in the Bayesian framework

Bayesian regression notations according to Ioannis (2009) were adopted. The response variable Y is considered to be a continuous random variable defined on the whole set of real numbers following the normal distribution with mean μ and variance σ^2. The model is summarized as follows;

$$Y \sim N(\mu(\beta, X_1, X_2, \dots X_p), \sigma^2) \tag{1}$$
With
$$\mu(\beta, X_1, X_2, \dots X_p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \dots + \beta_p X_p \tag{2}$$
Where σ^2 and $\beta = (\beta_0, \beta_1, \beta_2 \dots \beta_p)^T$ are regression parameter under estimation and $X_1, X_2, \dots X_p$ are the explanatory or predictor variables.
Alternatively, the model in (1) can be specified as;
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \dots + \beta_p X_p + \varepsilon, \ \varepsilon \sim N(0, \sigma^2) \tag{3}$$
where ε represent the error term or residuals.

Specifying the likelihood

Observe a sample of size n with response values $\mathbf{y} = (y_1, y_2, \dots y_p)^T$ and $x_{i1}, x_{i2}, \dots x_{ip}$ the values of the explanatory variables $X_1, X_2, \dots X_p$ for individuals $i = 1, 2, \dots, n$. Then the likelihood is expressed as
$$Y_i \sim N(\mu_i, \sigma^2) \tag{4}$$
$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}, \dots + \beta_p x_{ip} \quad \text{for } i = 1, 2, \dots, n \quad \text{(Ioannis, 2009)}$$
Within WinBUGS the normal distribution is defined in terms of its precision $\tau = \sigma^{-1}$. Details of the implementation are shown in the program codes.

Specifying the independent prior distributions

According to Ioannis (2009), the simplest approach to normal regression modeling is to assume that all parameters are a priori independent having the structure;

$$f(\beta, \tau) = \prod_{j=0}^{p} f(\beta)f(\tau),$$

$\beta_j \sim N(\mu_{\beta_j}, c_j{}^2)$ $for\ j = 0, 1, \dots p$ and   (5)

$\tau \sim gamma\ (a, b)$

where $\mu_{\beta_j}, c_j{}^2$,a and b are respectively; the mean of the $\beta_j's$, the variance of the $\beta_j's$, the location and scale parameter of the gamma distribution.

The gamma prior of the precision parameter τ induces prior mean and variance given by

$E[\tau] = \frac{a}{b}$ $and\ Var[\tau] = \frac{a}{b^2}$ respectively.

In this prior setup, we ensure compatibility with the WinBUGS notation by substituting the variance σ^2 by the corresponding precision parameter τ. The gamma prior used for τ corresponds to an inverse gamma prior distribution for the original variance parameter with prior mean and variance given by;

$E[\sigma^2] = \frac{b}{a-1}$ $and\ Var[\sigma^2] = \frac{b^2}{(a-1)^2(a-2)}$

respectively. When no information is available, a usual choice for the prior mean is the zero value ($\mu_{\beta_j} = 0$). This prior choice centers our prior beliefs around zero, which corresponds to the assumption of no effect of X, on Y.

The prior variance $c_j{}^2$ of the effect β_j is set equal to a large value (e.g., 10^4) to represent high uncertainty or prior ignorance. Similarly, for τ, equal low prior parameter values is used, setting in this way its prior mean equal to one and its prior variance large. If for instance $a = b = 0.01$, this result in to $E[\tau] = 1$ $and\ Var[\tau] = 100$.   (Ioannis, 2009)

Interpretation of the regression coefficients

Each regression coefficient β_j pertains to the effect of explanatory variable X_j on the mean of the response variable Y adjusted for the remaining covariates. The inference concerning the model parameters were made by giving answers to the following questions posed by Ioannis (2009).

Is the effect of X, important for the prediction or description of Y?

To find answer to this question, posterior distribution of β_j, is examined to see if it is scattered around zero (or not). Posterior distributions far away from the zero value indicate an important contribution of X, on the prediction of the response variable. This can be judged by examining the proportion of times that β_j exceed zero [p($\beta_j > 0$)] or the proportion of times that β_j is less than zero [$p(\beta_j < 0)$]. This analysis offers a first and reliable tool for tracing important variables in the model.

What is the association between Y and X, (positive or negative)?

The task here is to identify whether the relationship is positive or negative. This we base on the signs of the posterior summaries of central and relative location (e.g., mean, median, 2.5% and 97.5% percentiles). If it happens that all of them are positive or negative, then the corresponding association can be concluded. Positive association means that changes of the explanatory variable X, cause changes in the same direction for variable Y while negative association means that changes of the explanatory variable X, cause changes in the opposite direction for variable Y. Within this analysis, we a posteriori calculate the posterior probability:

$\pi_0 = \min\{f(\beta_j < 0 \backslash y),\ f(\beta_j > 0 \backslash y)\}$ (6)

When the zero value lies at the center of the posterior distribution, then the value shown above will be close to indicating that there is no clear positive or negative effect of X, on Y. When π_0 is low (e.g., lower than 2.5%, 1%, or 0.5%), then we may conclude positive or negative association depending on the sign of the posterior location summaries. Within WinBUGS we calculate the posterior probability $f(\beta_j > 0 \backslash y)$ using the syntax

$p.beta\ j < -step(beta\ j)$ (7)

this creates a binary node $p.beta\ j$ taking values equal to one when β_j, is positive and zero otherwise. Obtaining the posterior mean via the sample monitor tool provides us the estimate of the posterior probability $f(\beta_j > 0 \backslash y)$.

What is the magnitude of the effect of X, on Y?

The magnitude of the effect of variable X_j on Y is given by the posterior distribution of β_j $(for\ j = 1, 2, \dots, p)$ since

$$\triangle \mu_{X_j} = \mu(\beta, X_1, X_2, \dots X_{j-1}, X_j = x + 1, X_{j+1}, \dots, X_p) - \mu(\beta, X_1, X_2, \dots X_{j-1}, X_j = x, X_{j+1}, \dots, X_p)$$

$= \beta_0 + \beta_j(x + 1) + \sum_{k \neq j; k=1}^{p} \beta_k X_k - \beta_0 - \beta_j x - \sum_{k \neq j; k=1}^{p} \beta_k X_k = \beta_j$

It therefore follows that the posterior mean or median of β_j, will correspond to the corresponding posterior measures of the expected change of the response variable Y. Hence, an increase of one unit of X_j, given that the remaining covariates will remain stable, induces an a posteriori average change on the expectation of Y equal to the posterior mean of β_j (Ioannis, 2009).

Structural assumptions of the Bayesian normal model and general goodness of fit

Structural assumptions of the Bayesian normal model

We verify graphically, the assumptions of the Bayesian normal models using the history, density and autocorrelation plots of the residuals computed from the WinBUGS program code. The normality of errors assumption is verified by examining the density plot obtained for each month of the season to see if they follow a normal distribution about mean zero and constant variance. The homogeneity or equal variance assumption of errors is checked, by observing the history plots of the residuals for each month. A rectangular band with approximately constant width is an indication of equal or constant variance. The independence assumption of errors is checked by observing the autocorrelation plots for each month to see if the bars cut-off for each lag. If this happens, it is an indication of zero autocorrelation at each lag. This implies independence of the errors. The idea of using this graphical approach in the check of model assumption was stimulated from the Stationerity, Autocorrelation function (ACF) and Partial Autocorrelation (PACF) concepts in time series modeling. Stationariety of time series is graphically determined by plotting the actual or historic data (history plot) to see if it exhibits a constant mean and variance while the PACF and ACF plots are used to graphically determine the order (p) of the AR(p) and the order (q) of the MA (q) components of an ARMA (p,q) or ARIMA (p,q) process. This principle is based on the dependency property of the series (Klaus, 2016).

General goodness of fit.

The general goodness of the Bayesian model-fit is determined using the model's coefficient of determination (R_B^2) computed using the WinBUGS code. Since the precision parameter τ and the variance σ^2 indicates the precision of the model (if the precision τ is high and its variance low, then the model can accurately predict (or describe) the expected values of Y. Using the coefficient of determination (R_B^2), this can be rescaled as:

$$R_B^2 = 1 - \frac{\tau^{-1}}{s_Y^2} \qquad (8)$$

where s_Y^2 is the sample variance of Y. The quantity R_B^2 can be interpreted as the proportional reduction of uncertainty concerning the response variable Y achieved by incorporating the explanatory variables X_j in the model. Moreover, it can be regarded as the Bayesian analog of the adjusted coefficient of determination R_{adj}^2 (used in the frequentist approach of the normal regression model). It is given by:

$$R_{adj}^2 = 1 - \frac{\hat{\sigma}^2}{s_Y^2} \quad \text{where} \quad \hat{\sigma}^2 = \frac{1}{n-p}\sum_i^n (y_i - \hat{y}_i)^2 \text{ and}$$

$$\hat{y}_i = \hat{\beta}_0 + \sum_{i=1}^{p} x_{ij}\hat{\beta}_j \qquad (9)$$

The $\hat{\beta}_j's$ are the maximum likelihood estimates of β_j . In order to calculate R_B^2 in WinBUGS, we use the codes

$$sy2 < - pow\,(sd(y[\,],2) \text{ and } R2B < -(1 - S^2/sy2) \qquad (10)$$

or we can directly incorporate the precision parameter and use it as follows

$$R2B < - (1 - 1/(tau * sy2)) \qquad \text{(Ioannis, 2009)} \qquad (11)$$

We state here that the model equation (9) is the classical version of the Bayesian normal model. We implement the model on the datasets using the Statistical Package for Social Science (SPSS) version 21 to enable us compare its results of and those of the Bayesian normal model.

The Bayesian normal model convergence diagnostics

Model convergence diagnostics was done using history plots, density plots and autocorrelation plots of the beta coefficients. The plots were produced when the model parameters and measures were monitored on the run of the WinBUG program. Our approach for investigating convergence issues is by inspecting the mixing and time trends within the chains of individual parameters. The history plots are the most accessible convergence diagnostics and are easy to inspect visually. The history plot of a parameter plots the simulated values for the parameter against the iteration number. The history plot of a well-mixing parameter should traverse the posterior domain rapidly and should have nearly constant mean and variance. The density plots of the model parameters were checked against their actual probability distributions to see whether the right distribution is simulated.

Samples simulated using MCMC methods are correlated. The smaller the correlation, the more efficient the sampling process. Though, the Gibbs, MCMC algorithm typically generates less-correlated draws, there is a need to monitor the autocorrelation of each parameter to ensure samples are independent. The autocorrelation plot that comes from a well-mixing chain becomes negligible fairly quickly, after a few lags. This was achieved for each of the model parameters and measures.

RESULTS AND DISCUSSION

As earlier mentioned, the specification of the right variables in the Bayesian Normal model employed in this work is greatly eased by the correlations among the climatic determinants. The reason been that, the independent variables or determinants that relate significantly with the dependent variable can be easily identified and those that correlate with themselves identified for logarithm transformation in order to reduce multicolinearity. Table 1 below shows that Solar radiation only correlate with Relative humidity and Wind speed, while Relative humidity correlates with Wind speed and Solar radiation. Observe also that Total Rainfall Amount correlate with Number of Wet Days and Number of Dry Days while Number of Dry Days correlates with Total Rainfall Amount and Number Wet Days. Air temperature and Sunshine hours did not correlate with any climatic determinant. This correlation structure, informed the specification of the three (3) climatic models; Solar radiation model as a function of Relative humidity and Wind speed, Total Rainfall Amount model as a function of Number of Wet Days and Number of Dry Days and finally the Number of Dry Days model as a function of Total Rainfall Amount and Number of Wet Days. Correlation structure was used in specifying model variables as suggested by Pathan (2015).

Table 1: Degree of Association Between Climatic Determinants in the Wet Season

		Solar radiation	Air temperature	Relative humidity	Sunshine hours	Number of dry days	Wind speed	Rainfall total amount	Number of wet days
Solar radiation	Correlation Coefficient	1.000	.595	-.714*	-.262	.240	.810*	-.024	-.228
	Sig. (2-tailed)	.	.120	.047	.531	.568	015	.955	.588
Air temperature	Correlation Coefficient	.595	1.000	-.310	-.262	.132	.571	.167	-.156
	Sig. (2-tailed)	.120	.	.456	.531	.756	.139	.693	.713
Relative humidity	Correlation Coefficient	-.714*	-.310	1.000	.452	.347	-.810*	-.476	-.383
	Sig. (2-tailed)	.047	.456	.	.260	.399	.015	.233	.349
Sunshine hours	Correlation Coefficient	-.262	-.262	.452	1.000	-.096	-.310	-.024	.012
	Sig. (2-tailed)	.531	.531	.260	.	.821	.456	.955	.978
Number of dry days	Correlation Coefficient	.240	.132	.347	-.096	1.000	-.060	-.946**	-.988**
	Sig. (2-tailed)	.568	.756	.399	.821	.	.888	.000	.000
Wind speed	Correlation Coefficient	.810*	.571	-.810*	-.310	-.060	1.000	.214	.036
	Sig. (2-tailed)	.015	.139	.015	.456	.888	.	.610	.933
Rainfall total amount	Correlation Coefficient	-.024	.167	-.476	-.024	-.946**	.214	1.000	.946**
	Sig. (2-tailed)	.955	.693	.233	.955	.000	.610	.	.000
Number of wet days	Correlation Coefficient	-.228	-.156	-.383	.012	-.988**	.036	.946**	1.000
	Sig. (2-tailed)	.588	.713	.349	.978	.000	.933	.000	.

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

The results of each Bayesian Normal Regression Model coefficients and those of its Classical counterpart are captured in table 2, 3 and 4 for the Solar radiation, Total Rainfall Amount and the Number of Dry Days model respectively. Table 2 shows for the Bayesian approach, that the Solar radiation model relates positively and significantly with Relative humidity and Wind speed (with prob (beta > 0) equals 1 for each variable). For this approach it can be inferred from the table that, in the wet or cropping season of Makurdi, Solar radiation increases by 0.791 MJ/m^2 for each unit increase in the natural logarithm of Relative humidity, while, it increases by 0.895 MJ/m^2 for each unit increase in the natural logarithm of Wind speed. The result for its Classical counterpart shows that Solar radiation neither relate significantly with Relative humidity nor Wind speed (p values > 0.05). Although the coefficient of determination is over 65 % for both models, this discrepancy still exist. Observe for the Bayesian model on table 3 shows that Total Rainfall Amount relates positively and significantly with Number of Wet Days and Number of Dry Days (with prob (beta > 0) equals 1 for each variable).

Further result reveal for the Bayesian model that, Total Rainfall Amount increases by 66.280 mm for each unit increase in the natural logarithm of Number of Dry Days while it increases by

Model in table 2 indicate that increase in Relative humidity and Wind speed increases Solar radiation while the model in table 3 shows that the frequency of dry days increases with increase in

2.912 mm for each unit increase in the natural logarithm of the Number of Wet Days. The result for its Classical counterpart on the same table shows that Total Rainfall Amount relates negatively and significantly with Number of Wet days (p value < 0.05) but has no significant relationship with Number of Dry Days (p value > 0.05). This result differs from that of the Bayesian model despite the fact that its coefficient of determination is as high as 96 %. A good coefficient of determination may not always mean a good – fit.

Table 3 reveals different results for the Bayesian and Classical models despite their coefficient of determination of over 90%. For the Bayesian model, the Number of Dry Days relates negatively and significantly with the Number of Wet Days (prob (beta < 0) equals 1) and positively and significantly with Total Rainfall Amount (prob (beta > 0) equals 1). Further result on this shows that in the wet or cropping season of Makurdi, the Number of Dry Days decreases by 6.905 days for each unit increase in the natural logarithm of Number of Wet days while it increases by 2.028 days for each unit increase in the natural logarithm of Total Rainfall Amount. The result for its Classical counterpart on the same table shows that Number of Dry Days neither relate significantly with the Number of Wet Days nor Total Rainfall Amount.

Total Rainfall Amount in the area. This result is consistent with Agada et al. (2018) and Fabiyi et al. (2013).

Table 2: Classical and Bayesian Regression statistics for Solar Radiation (SR) model

	Classical Multiple Linear Regression Model				Bayesian Multiple Linear Regression Model			
Model	Beta Coefficients	P value	R^2 (%)	95 % Confidence interval	Beta Coefficients	Prob (Beta >0)	R^2 (%)	95 % Credible interval
Constant	267.728	0.050		(0.220, 535.236)	13.310	1.000		(-46.900, 73.76)
Ln(REL)	-58.371	0.610	(66.300)	(-120.560, 3.818)	0.791	1.000	(65.400)	(-13.300, 14.810)
Ln(WS)	0.695	0.167		(-0.410, 1.800)	0.859	1.000		(-0.560, 2.264)

Table 3: Classical and Bayesian Regression statistics for Total Rainfall Amount (TRA) Model

	Classical Multiple Linear Regression Model				Bayesian Multiple Linear Regression Model			
Model	Beta Coefficients	P value	R^2 (%)	95 % Confidence interval	Beta Coefficients	Prob (Beta >0)	R^2 (%)	95 % Credible interval
Constant	1170.345	0.0120		(388.644, 1952.045)	8.506	1.000		(-50.900, 67.780)
Ln(NW)	-339.563	0.0120	(96.100)	(-565.081, -114.044)	66.280	1.000	(63.840)	(30.160, 90.020)
Ln(ND)	10.359	0.610		(-38.620, 59.339)	2.912	1.000		(-20.370, 30.420)

**Table 4: Classical and Bayesian Regression statistics for Number of Dry Days (ND) Model**

Model	Classical Multiple Linear Regression Model				Bayesian Multiple Linear Regression Model			
	Beta Coefficients	P value	R^2 (%)	95 % Confidence interval	Beta Coefficients	Prob (Beta >0)	R^2 (%)	95 % Credible interval
Constant	27.730	0.023		(5.615, 49.946)	25.770	1.000		(-4.425, 44.070)
Ln(NW)	-6.147	0.154	(90.400)	(-15.667, 3.268)	-6.905	0.000	(94.160)	(-15.960, 0.982)
Ln(TRA)	1.306	0.706		(-7.085, 9.695)	2.028	1.000		(-4.097, 10.09)

We also state that the reason for the discrepancies in the results of the Bayesian models and their Classical counterparts lie on the aggregation of the entire dataset of 34 years into an eight (8) - Sample size dataset. This amounts to an incorrect fit with the Classical approach. As earlier mentioned, researchers have affirmed that the Classical or Frequentist Regression Modeling approach may not yield good results with sparse or incomplete datasets (Taeryon, *et al,* 2008; Agada, *et al,* 2019; Tripathi, *et al,* 2019). According to these authors, this is not the case with the Bayesian modeling approach as it does not depend chiefly on data sampling but on model parameter sampling.

We certified the correctness of the Bayesian Normal Model via the satisfaction of structural assumptions, goodness of fit test and model diagnostic checks. We adopt a graphical approach, in the verification of the assumptions of the Bayesian normal models. The history, density and autocorrelation plots of the residuals computed from the WinBUGS program code were employed for this purpose. For the normality of errors assumption, observe for each model, that the density plots of residuals obtained for each month of the season follow a normal distribution in the neighborhood of a zero mean as seen in **Fig. 1**. The homogeneity or equal variance assumption of errors can be checked, by examining the history plots of the residuals for each model and for each month. Observe a rectangular band with approximately constant width which characterizes each plot (see **Fig. 2**). This is an indication of equal or constant variance. The independence

assumption of errors can be checked by examining the autocorrelation plots for each model and for each month. Observe that the bars cut-off at each lag (**Fig. 3**). This is an indication of zero autocorrelation at each lag signifying the independence of the errors.

We judge the goodness of fit of the Bayesian models from the value of the Coefficient of Determination (R^2) for each model. Tables 2, 3 and 4 reveal R^2 values of 65.40, 63.84 and 94.16 percent respectively for the Solar radiation, Total Rainfall Amount and the Number of Dry Days model. This shows that good percentage variability is explained by the models. Hence a good model fit. After the development of the model, convergence diagnostic checks were conducted for each model parameter and measure in order to ascertain model adequacy. The history plot, density plots and autocorrelation plots were used for this purpose. See **Figs. 4 – 6** for the respective plots of the parameters for each model. Observe that the history plots shows that the model parameters and measures are well – mixed. This is because they traverse the posterior domain rapidly with nearly constant mean and variance. The model prior distributions for the beta coefficients are normal (0, 0.001). The normal density plots of these priors reflect these normal (0, 0.001) distributions. This further validates each model. The autocorrelation plots of each parameter and measure depict the independence of the samples generated. This is because the autocorrelations become negligible fairly quickly, after a few lags.
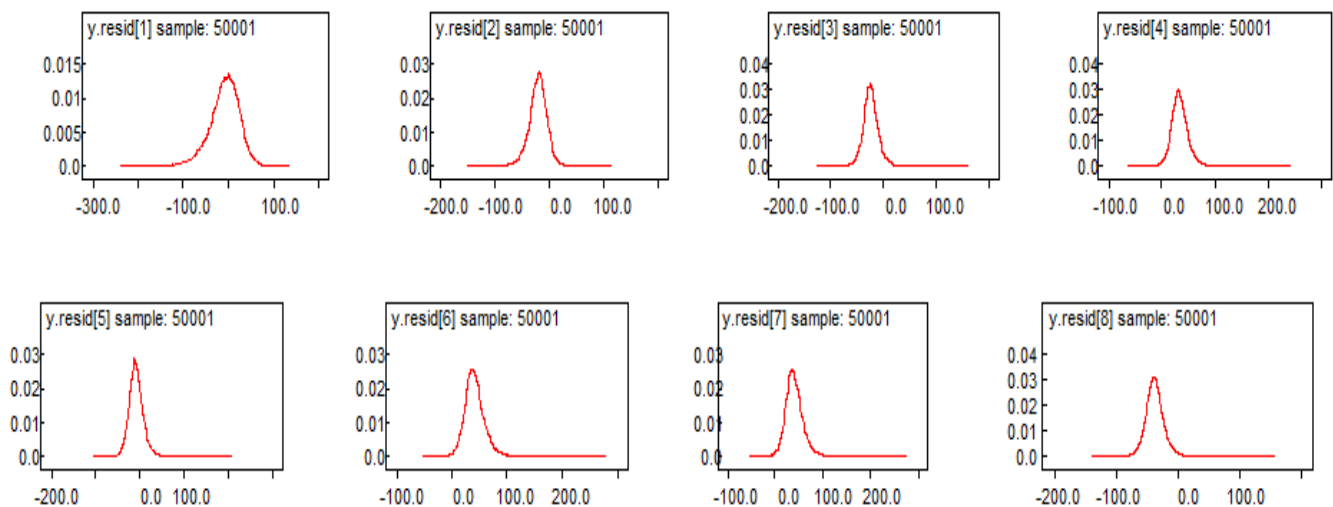
Figure 1: Normal density plots of residuals for the Total Rainfall Amount Model for each month
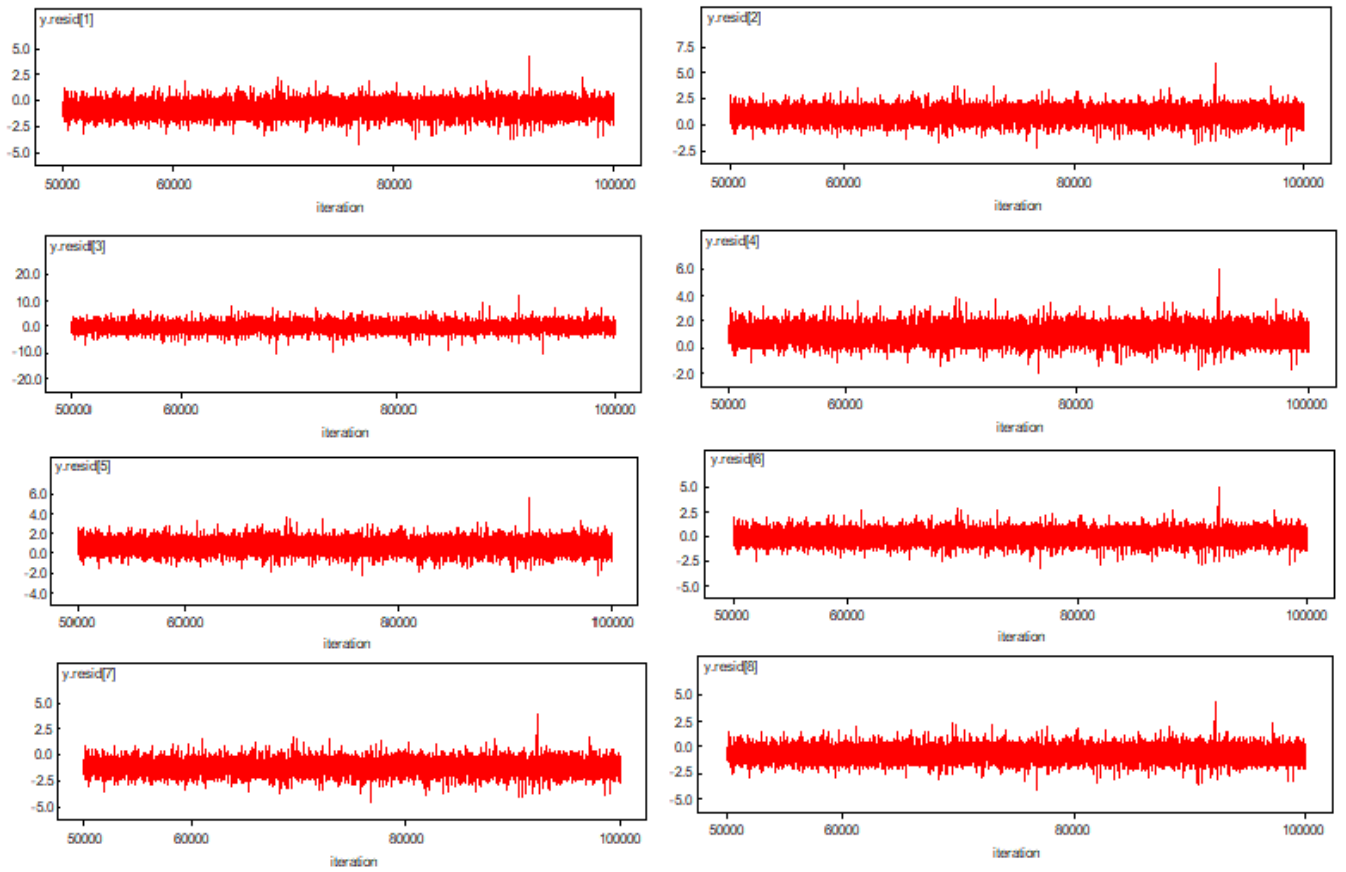
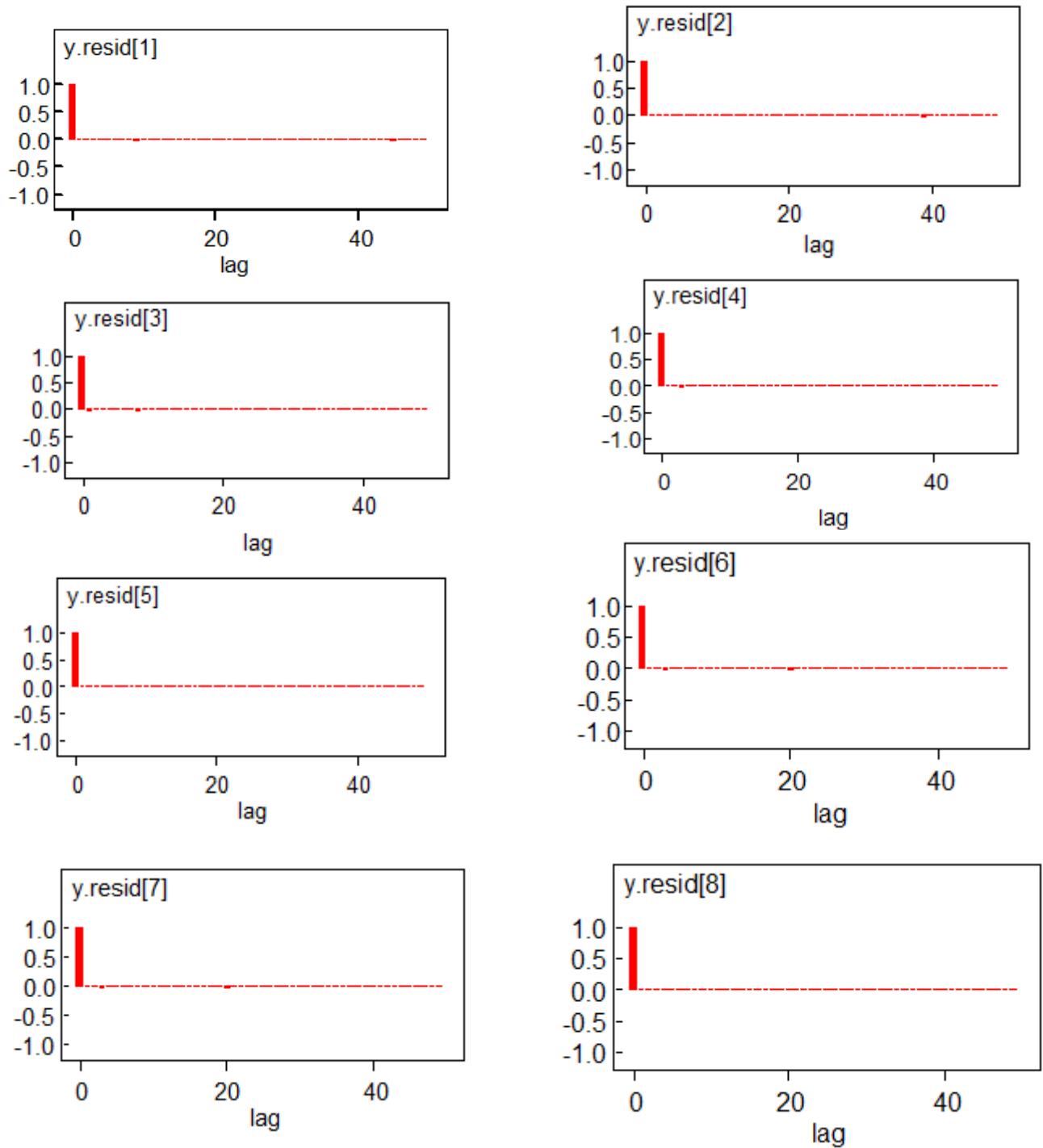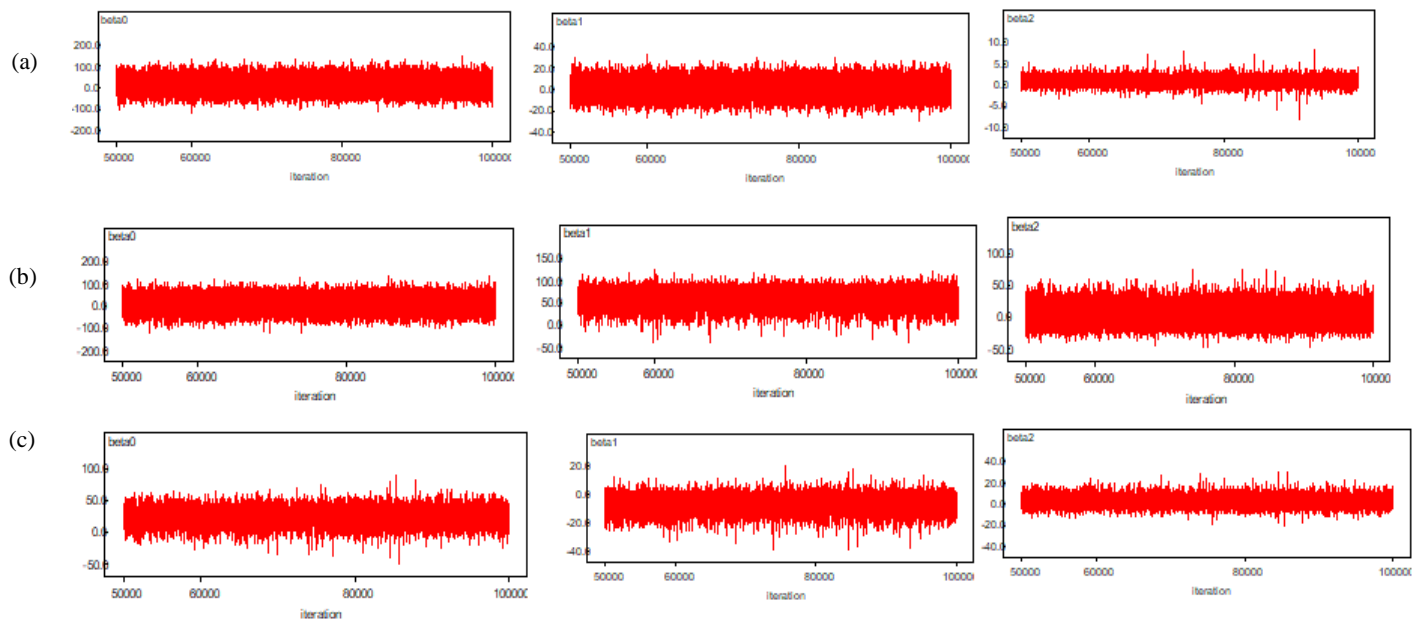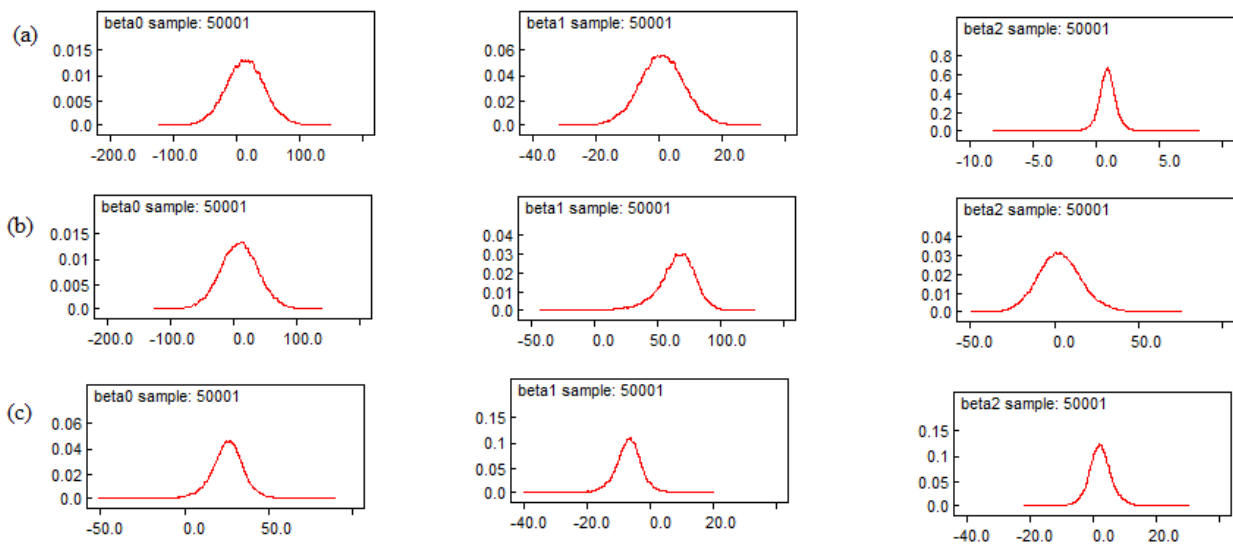**Figure 2:** History plots of residuals for the Solar Radiation Model in each month

Figure 3: Autocorrelation plots of residuals for the Number of Dry Days Model in each month

Figure 4: History plots of Regression coefficients for the Solar Radiation (a), Total Rainfall Amount (b) and Number of Dry Days (c) models

Figure 5 : Normal density plots of Regression coefficients for the Solar Radiation (a), Total Rainfall Amount (b) and Number of Dry Days (c) models
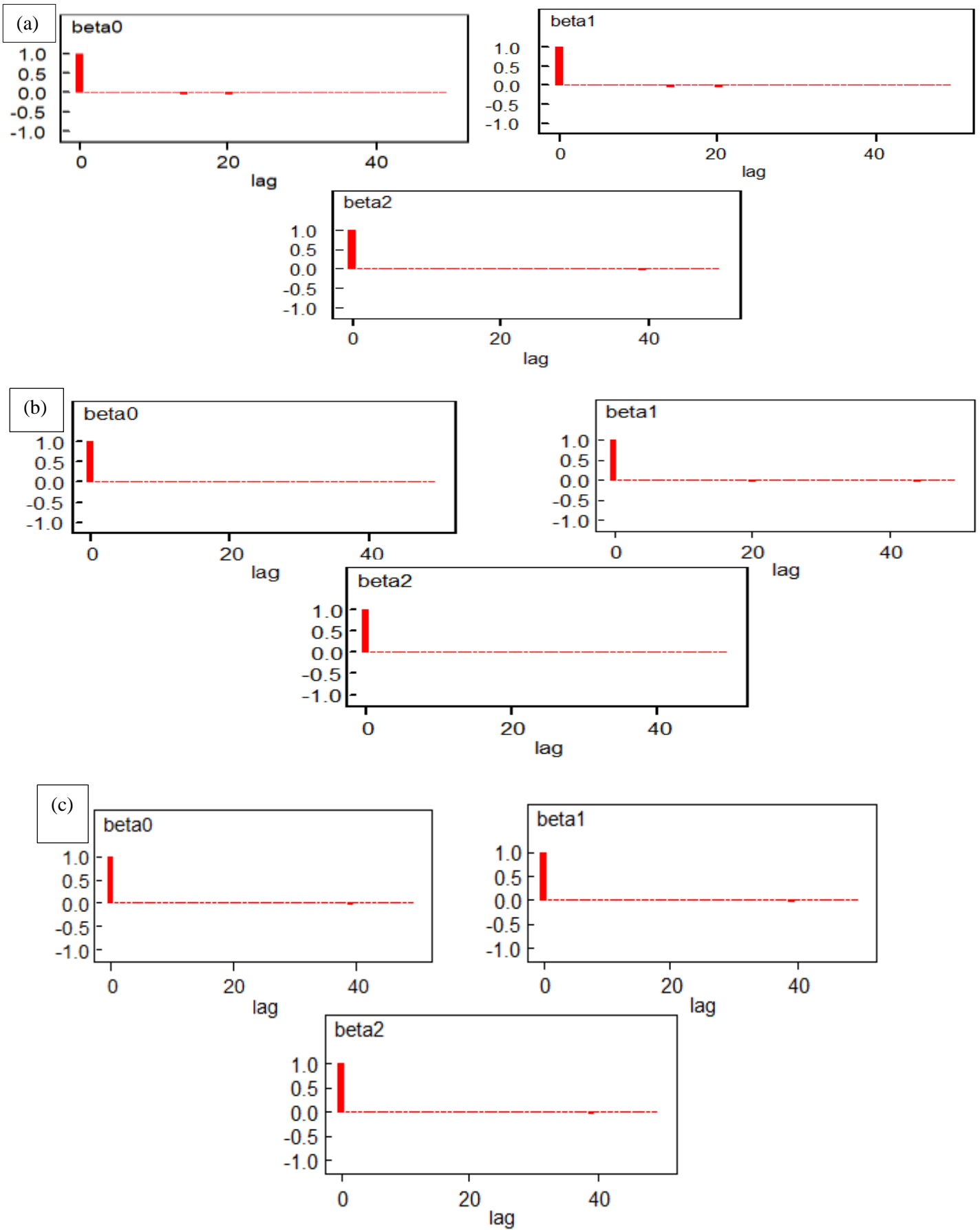
Figure 6: autocorrelation plots of Regression coefficients for the Solar Radiation (a), Total Rainfall Amount (b) and Number of Dry Days (c) models

CONCLUSION AND RECOMMENDATIONS

Conclusion

The Bayesian Normal Regression Climatic Model is more adequate in modeling the wet or cropping season climate of Makurdi than its Classical counterpart. Solar radiation increases by 0.791 MJ/m^2 for each unit increase in the natural logarithm of Relative humidity while it increases by 0.895 MJ/m^2 for each unit increase in the natural logarithm of Wind speed. Total Rainfall Amount increases by 66.280 mm for each unit increase in the natural logarithm of Number of Dry Days while it increases by 2.912 mm for each unit increase in the natural logarithm of the Number of Wet Days. The number of Dry Days decreases by 6.905 days for each unit increase in the natural logarithm of the Number of Wet days while it increases by 2.028 days for each unit increase in the natural Total Rainfall Amount. Moreover, the wet or cropping season climate of Makurdi is becoming warmer and drier.

Recommendations

The Bayesian Normal Regression Climatic Models should be employed in modeling the relationships between climatic determinants in the wet or cropping season of Makurdi, Nigeria when sample size reduction due to data aggregation is a concern. Furthermore, climate change impact on the warming and dry climate in the cropping season should be checked if the area is to retain its potentials for crop production in the long run.

REFERENCE.

Agada, P. O., Udoumoh, E. F. & Gboga, H. P. 2019. A Bayesian Simulation Modeling Approach to Predicting Maternal Age – Specific Infant Survival Outcomes with Incomplete Data. *FUDMA Journal of Science*, 3(4): 2616 – 1370.

Agada, P. O., Imande, M. T. & Amedu, M. O. 2018. Statistical Indicators of Climate Change in Makurdi, Metropolis: An Implication to Crop Production in the Area. *abacus* The journal of Mathematical Association of Nigeria. 45 (1): 198 – 213.

Aymen, R. & Mohammed, O. 2019. A Bayesian Approach to Estimate a Linear Regression Model with Aggregate Data. *Austrian Journal of Statistics*. 48: 90 – 100.

Eric, A. H., John, E. J. & John, F. K. 1974. Model Specification, Use of Aggregate Data, and the Ecological Correlation Fallacy. *Political Methodology*. 89 – 107.

Ioanasis, N. 2009. Bayesian Modeling Using WinBUGS. John Willey and Sons, Hoboken, New Jersey.

Klaus, N. 2016. Time Series Econometrics, Springer International Publishing, Switzerland.

Moineddin, R. & Urquia, M. L. 2014. Regression Analysis of Aggregate Continuous Data. *Epidemiology*, 25 (6): 929 – 930.

Pathan, M. B. 2015. Modeling of Data from Climate Science Using Introductory Statistics. *J. Climatol Weather Forecasting*. 3 :129 - 134.

Robinson, W. F. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* 63: 1183 –1196.
Taeryon, C., Mark, J. S., Ketra, A. S. & Mitchell, J.S. 2008. A Bayesian Approach to a Logistic Regression Model with Incomplete data, *Biometrics.*64: 424 - 430

Tripathi, A., Singh, G. & Singh, A. 2019. Model for Assessment of Child Mortality under Different Parity: A Bayesian Swatch. *Journal of Biosciences and Medicines*, 7: 113-126

Dias, S., Sutton A. J., Welton, N. J.,& Aides A. E., 2013. Evidence Synthesis for Decision Making 3 Heterogeneity – Subgroups, Meta – Regression, Bias, and Bias Adjustment, *Medical Decision Making*, 33 (5): 618 – 640.