

## Improved weather indices based Bayesian regression model for forecasting crop yield

M. YEASIN, K. N. SINGH, A. LAMA\* and B. GURUNG

ICAR-Indian Agricultural Statistics Research Institute, New Delhi – 110 012

(Received 11 June 2021, Accepted 14 September 2021)

\*e mail : achal.lama@icar.gov.in

**सार** – चूँकि कृषि भारतीय अर्थव्यवस्था की रीढ़ है, इसलिए सरकार को नई योजनाओं की योजना बनाने के लिए फसल उपज के एक विश्वसनीय पूर्वानुमान की आवश्यकता है। फसल उपज के पूर्वानुमान के लिए सबसे व्यापक रूप से इस्तेमाल की जाने वाली तकनीक प्रतिगमन विश्लेषण है। मापदंडों का महत्व प्रतिगमन विश्लेषण की प्रमुख समस्याओं में से एक है। गैर-महत्वपूर्ण पैरामीटर बेटुके पूर्वानुमान मूल्यों की ओर ले जाते हैं और ये पूर्वानुमान मान विश्वसनीय नहीं होते हैं। ऐसे मामलों में, मॉडलों में सुधार की जरूरत है। मॉडलों को बेहतर बनाने के लिए, हमने बायेसियन तकनीक के माध्यम से पूर्व ज्ञान को शामिल किया है और बायेसियन ढांचे के तहत इन मॉडलों की श्रेष्ठता की जांच की है। बायेसियन तकनीक सांख्यिकी के आधुनिक युग में सबसे शक्तिशाली विधियों में से एक है। हमने विभिन्न प्रकार के पूर्व (सूचनात्मक, गैर-सूचनात्मक और संयुग्म पूर्वज) पर चर्चा की है। बायेसियन ढांचे के तहत मापदंडों के आकलन के लिए मार्कोव श्रृंखला मॉटे कार्लो (एमसीएमसी) पद्धति पर संक्षेप में चर्चा की गई है। इन मॉडलों को स्पष्ट करने के लिए केला, आम और गेहूँ की उपज के आंकड़ों को ध्यान में रखा जाता है। हमने पारंपरिक प्रतिगमन मॉडल की तुलना बायेसियन प्रतिगमन मॉडल से की और निष्कर्ष निकाला कि बायेसियन ढांचे के तहत अनुमानित मॉडल ने शास्त्रीय दृष्टिकोण के तहत अनुमानित मॉडल की तुलना में बेहतर परिणाम प्रदान किए।

**ABSTRACT.** As agriculture is the backbone of the Indian economy, Government needs a reliable forecast of crop yield for planning new schemes. The most extensively used technique for forecasting crop yield is regression analysis. The significance of parameters is one of the major problems of regression analysis. Non-significant parameters lead to absurd forecast values and these forecast values are not reliable. In such cases, models need to be improved. To improve the models, we have incorporated prior knowledge through the Bayesian technique and investigate the superiority of these models under the Bayesian framework. The Bayesian technique is one of the most powerful methodologies in the modern era of statistics. We have discussed different types of prior (informative, non-informative and conjugate priors). The Markov chain Monte Carlo (MCMC) methodology has been briefly discussed for the estimation of parameters under Bayesian framework. To illustrate these models, production data of banana, mango and wheat yield data are taken under consideration. We compared the traditional regression model with the Bayesian regression model and conclusively infer that the models estimated under Bayesian framework provided superior results as compared to the models estimated under the classical approach.

**Key words** – Bayesian technique, MCMC, Prior distribution, Simple regression model, Weather indices.

### 1. Introduction

The agriculture sector is one of the most significant contributors to the Indian economy. Agriculture enables the development of more densely populated and stratified societies by creating food security. India is called an agricultural country, as the agriculture sector of India has occupied almost 43 percent of India's geographical area and almost seventy percent of people of India are engaged in the agriculture sector directly or indirectly. In agricultural countries like India, planning formulation and implementation of several policies dealing with food procurement are very much dependent on the forecast of crop yield before the harvest. To forecast crop yield before harvest, we have to know about the factors which affect the crop yield. Weather variables are one of the

crucial factors for the growth and production of crops. Modeling and forecasting crop yield before harvesting is a matter of concern to the researcher. Multiple Regression analysis is one of the powerful tools for agro-metrological crop yield forecasting (Gommes, 1998). Regression Analysis is a multivariate technique used to analyze the environmental factors as explanatory variables and their inflection on crop yield to obtain a decision (Sellamand Poovammal, 2016). In most cases, the traditional regression model gives a very efficient estimate value of the parameters, but sometimes parameters of the model may not be statistically significant (Chatterjee and Hadi, 2015). So, there lies a scope for further improvement of regression parameter estimates. Thus, in this manuscript, we have employed the Bayesian technique in this context.

Bayesian analysis is based on the basic assumption that all parameters of the model are random in nature. We can incorporate prior knowledge into the estimated value of parameters to get improved forecast values. Bayesian statistics is contrary to classical frequentist statistics where all the parameters are treated as fixed and unknown quantities. Bayesian analysis is based on the theories of probability. Bayesian regression is not only an algorithm but also a different mode of approach for various statistical inferences. By using the Bayesian framework, we obtain a range of inferential solutions instead of the point estimate (Marin *et al.*, 2014). Usually, a non-Bayesian regression model or simply a linear regression model tends to over fit the data. Bayesian Linear Regression helps to overcome the problem by incorporating “Predictive Distribution” (Gregory, 2005). The Bayesian technique is a subjective procedure to estimate the unknown parameters of the linear model and this model results more efficient estimate of parameters (Bunn, 1975). The advantage of Bayesian regression estimation can be attributed to the fact that one can incorporate a prior distribution, or use assumed knowledge about the present state of “beliefs” and make the estimated value more precise and efficient. There are various analytical and numerical techniques available for implementing the Bayesian method with appropriate prior to solve a wide class of problems (Bernardo and Smith, 2001). The Bayesian technique can be successfully applied to Generalized Linear Model (GLM) for improvement of the model and this improvement depends on the choice of the prior distribution (Das, 2008).

In this paper, we attempt to improve the crop yield forecast using the Bayesian technique. We could not find similar work in the literature and hope our approach will enrich the existing literature of weather indices based crop yield forecasting models. In subsequent sections, we discuss material methods, illustration, results and followed by conclusions.

## 2. Materials and method

### 2.1. Data sources

The daily data on weather parameters such as maximum and minimum temperature, morning and evening relative humidity, amount of rainfall for 23 years (from 1984 to 2007) has been collected from a weather station located at IARI, New Delhi. Three production data, *viz.*, wheat, banana and mango data series have been collected to illustrate the models. Wheat yield data were collected from IARI, New Delhi. The wheat data sets contain one dependent variable with 11 independent weather variables. Production data of banana and mango has been collected from the National Horticulture Board.

Production data of banana and mango contains 12 and 10 independent weather variables respectively.

### 2.2. Transformation of datasets

We assume that  $w$  denotes weeks ( $w = 1, 2, \dots, n$ ) at which the pre-harvest forecast of the crop yield needs to be released. If we use the weekly data on  $m$  weeks in  $p$  variables, now new weather variables and interaction components can be generated with respect to each of the weather variables using the below-mentioned procedure. A forecast model has been developed by considering all the generated variables simultaneously, including the time trend ( $T$ ) (Agrawal *et al.*, 2001).

In order to study the individual effect of each weather variables, two new variables from each weather variable can be generated as follows:

Let  $X_{iw}$  be the value of the  $i^{\text{th}}$  [ $i = 1(1) p$ ] weather variable at  $w^{\text{th}}$  week ( $w = 1, 2, \dots, n$ ),  $r_{iw}$  be the simple correlation coefficient between weather variable  $X_i$  at the  $w^{\text{th}}$  week and yield over a period of  $k$  years. The generated variables are given by:

$$Z_{ij} = \frac{\sum_{w=1}^n r_{iw}^j X_{iw}}{\sum_{w=1}^n r_{iw}^j}; j=0,1$$

For  $j = 0$  we have unweighted variables are generated as:

$$Z_{i0} = \frac{\sum_{w=1}^n X_{iw}}{n}$$

And weighted variables are generated as:

$$Z_{i1} = \frac{\sum_{w=1}^n r_{iw} X_{iw}}{\sum_{w=1}^n r_{iw}}$$

### 2.3. Traditional regression model

Regression analysis is a very important statistical tool for modeling, investigating and determining the relationship between the variables. The regression model involves three components. Those are dependent variable, independent variables and unknown coefficient. Dependent variables are also known as response or outcome or regressed variables. Independent variables are also known as predictors or regressors. In the regression model, there is one dependent variable, one or more than one independent variable and one or more than one unknown regression coefficient. A simple regression model can be expressed as  $y$  is a function of  $x$  and  $\beta$ .

$$x \text{ and } \beta : y \approx f(x, \beta)$$

where  $y$  is denoted as a dependent variable,  $x$  is a dependent variable and  $\beta$  is unknown coefficients.

In a multiple regression model dependent variable is related to more than one independent variables.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon$$

where  $y$  denotes a dependent variable,  $x_i$ 's denotes independent variables and  $\varepsilon$  is the error component.  $\beta_0$  is the intercept parameter and  $\beta_i$ 's are the slope parameters.

Our main interest lies in unknown parameters. The unknown parameters of the model should be estimated unbiasedly and efficiently. We estimate the unknown parameters of the model in such a manner that the sum of square of the residuals is the least. Where residuals are the difference between the observed values and the corresponding expected value of a dependent variable. This method is called Ordinary Least Square, invented by Carl Friedrich Gauss, but it was first published by Adrien-Marie Legendre (Plackett, 1972).

In our cases, the traditional models are :

(i) For banana data,  $k = 11$

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_{11}x_{11} + \varepsilon$$

where  $x_i$ 's are the weather indices,  $\beta_0$  is intercept and the total number of parameters is 12.

(ii) For mango data

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_{12}x_{12} + \varepsilon$$

where  $x_i$ 's are the weather indices,  $\beta_0$  is intercept and the total number of parameters is 13.

(iii) For wheat data

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_{12}x_{12} + \varepsilon$$

where  $x_i$ 's are the weather indices,  $\beta_0$  is intercept and the total number of parameters is 11.

All the parameters of these models are estimated by the OLS estimations technique.

#### 2.4. Bayesian approach

Bayesian inference is one of the most powerful techniques of estimation. This technique of estimation has various advantages over point estimation. The main advantage of this technique of estimation is that it uses the

prior information. By incorporating prior information, we get a posterior distribution. The posterior distribution contains more information due to the incorporation of extra information in the form of prior distribution.

The prior distribution is a probability distribution that expresses the possible uncertainty before examination of current data. The main problem of Bayesian estimation is to find a suitable prior with its parameters. Irrelevant prior misleads the researcher as it gives spurious results. So, one should be careful at the time of selection of priors. There are various types of priors proposed in the literature. These are informative or non-informative, conjugate or non-conjugate priors. Our interest lies in the conjugate priors because the form of posterior distribution remains the same as its prior distribution, only hyper-parameters are updated (Chen and Ibrahim, 2003). The exponential family of distribution is commonly used conjugate priors (Morris, 1983). In this study, the most common distribution of exponential family, *i.e.*, normal distribution is taken as a prior distribution. Steps undertaken for the Bayesian approach of regression estimation are as follows:

- (i) Model specification
- (ii) Selection of prior distribution
- (iii) Find the likelihood function
- (iv) Apply Bayes theorem and generate posterior distribution with the help of Markov Chain Monte Carlo (MCMC) Method
- (v) Find the expectation of the posterior distribution

Brief description of the following steps:

(a) *Model specification*

Under the Bayesian approach, our study models will be :

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \varepsilon$$

here  $\beta_i$ 's are not constant, they follow some specific distribution (prior distribution) with specific parameters.

[N.B : For banana data  $k = 12$ , for mango data  $k = 13$  and wheat data  $k = 11$ ]

(b) *Selection of prior distribution*

We have already discussed selection of prior. In our case, we select normal distribution as prior because it

gives more accurate results in the case of Bayesian regression analysis (Bunn 1975; Evans 2012).

Let  $\beta$  is the vector of parameters. For banana, mango and wheat data sets it consists of 12, 13 and 11 parameters respectively.

For banana data,  $\beta \sim N(\text{Mean} = M, \text{Variance} = V)$

Where  $M$  is a vector of order  $12 \times 1$  with elements  $\{M_1, M_2, \dots, M_{12}\}$  and  $V$  is a diagonal matrix of size  $12 \times 12$  with diagonal element  $\{V_1, V_2, \dots, V_{12}\}$ .

The values of  $M_1, M_2, \dots, M_{12}$  are computed by iterative procedure with a vague idea from the parameter estimates of the traditional regression model.  $M_i$ 's values are generally very near to the estimated values of the corresponding  $\beta_i$ 's. In the same procedure, we also compute the value of  $V_i$ 's which takes a value close to the variance of estimated  $\beta_i$ 's.

On similar lines, we specified prior distributions to the model parameters of mango and wheat datasets.

(c) *Find the likelihood function*

For the given datasets, we computed the likelihood function which is the joint probability density function of normal distribution.

(d) Apply Bayes' theorem and generate posterior distribution with the help of Markov Chain Monte Carlo (MCMC) Method.

Let  $\theta$  is the parameter of interest and  $\pi(\theta)$  is the prior distribution which gives the prior knowledge about the population and  $Y$  is the sample collected under study. Then the posterior distribution can be calculated with the help of Bayes' theorem and it is given by :

$$\pi(\theta|y) \propto L(Y|\theta) \pi(\theta)$$

where  $L(Y|\theta)$  is the likelihood function.

In the case of Bayesian inference, the MCMC (Markov Chain Monte Carlo) method is mainly used as a parameter estimation technique. MCMC is a restricted type of stochastic process. The Markov chain Monte Carlo (MCMC) method is used to generate values from a transition kernel in such a way that the collected sample from that kernel converges to a specified distribution which is targeted previously. This method simulates the Markov chain with a predefined distribution as the distribution of equilibrium or convergence of the chain. A

Markov chain can be defined as any sequence of states or values generated from the domain of pre-specified distribution; in such a way that the distribution at any stage depends on only its current state of the chain and each state only depends only on its immediate predecessor. The probability of convergence of the chain is directly proportional to the chain length of the MCMC method. The oldest version of the MCMC method was the Metropolis algorithm which was proposed by Metropolis and Ulam (1949) and Metropolis *et al.* (1953). In this algorithm, a lot of sequences of states are generated and each of the states can be obtained from only the previous state. Hastings (1970) presented a simpler and more general version of the MCMC algorithm, which is now known as the Metropolis-Hastings (MH) algorithm. To understand the process let us assume that we want to have information regarding a distribution  $\pi^*$ , of which we have information upto the point  $C$ , where an assumption that the state space  $E$  is either finite or countable. Then the distribution of  $\pi^*$  will be  $\pi(\theta) | C$ , as its probability mass function. The main purpose of using the MCMC method is to obtain the posterior distribution as :

$$\pi^*(\theta | y) = \frac{f(y|\theta)p(\theta)}{\sum_{\theta \in E} f(y|\theta)p(\theta)}$$

For obtaining the posterior distribution, the following steps are followed :

(i) An ergodic Markov Chain  $\theta_0, \theta_1, \theta_2, \dots$  is set up which results in a stationary posterior distribution.

(ii) Using Markov Chain simulate  $\theta_0, \theta_1, \theta_2, \dots, \theta_{l+k}$  for large  $l$  and  $k$ .

(iii) Discard the first  $l-1$  samples with  $l+k$  sufficiently large to obtain.

(iv) Obtain the expectation and other statistics using the  $l+k$  samples, this is done to obtain stationary values.

The two very widely used MCMC algorithms are the Metropolis-Hastings (MH) algorithm and Gibbs sampling. Gibbs sampling is considered to be a special sampler of the MH algorithm. Metropolis-Hastings (MH) algorithm and Gibbs sampling involve complex computation but as in the modern era, computers are very advanced, Bayesian technique can be used over the traditional method (Evans, 2012). We use the Gibbs sampling algorithm because this algorithm is simple, easily implemented and can handle the problem of high dimensionality (Smith and Roberts, 1993).

TABLE 1

Parameter estimates of simple regression model for banana

Coefficients	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value	Significance
B1	334.13	227.48	1.47	0.16	No
B2	68.28	14.98	4.56	0.00	Yes
B3	120.92	352.97	0.34	0.74	No
B4	-85.72	120.31	-0.71	0.49	No
B5	-236.56	417.00	-0.57	0.58	No
B6	-209.81	375.03	-0.56	0.58	No
B7	-7389.48	6851.06	-1.08	0.30	No
B8	3702.16	1487.62	2.49	0.02	Yes
B9	315.71	166.71	1.89	0.08	No
B10	3.55	12.42	0.29	0.78	No
B11	1799.32	936.62	1.92	0.07	No
B12	-630.70	364.02	-1.73	0.10	No

TABLE 2

Parameter estimates of simple regression model for mango

Coefficients	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value	Significance
B1	37.84	5.70	6.64	0.00	Yes
B2	39.87	87.59	0.46	0.65	No
B3	0.97	42.08	0.02	0.98	No
B4	-112.78	76.65	-1.47	0.16	No
B5	53.94	23.37	2.31	0.03	Yes
B6	4006.64	1682.53	2.38	0.03	Yes
B7	-2269.65	980.86	-2.31	0.03	Yes
B8	12.95	16.74	0.77	0.45	No
B9	-20.53	9.30	-2.21	0.04	Yes
B10	-663.12	341.90	-1.94	0.07	No
B11	141.50	269.66	0.53	0.60	No
B12	53.04	92.43	0.57	0.58	No
B13	-10.28	51.51	-0.20	0.84	No

(e) Find the expectation of the posterior distribution

The estimate of  $\theta$  can be computed easily from the posterior distribution. The estimate of  $\theta$  is the expectation of the posterior distribution and given by

$$\hat{\theta} = \int \theta \pi(\theta | y) d\theta$$

$$= \frac{\sum_{\theta \in E} \theta f(y | \theta) p(\theta)}{\sum_{\theta \in E} f(y | \theta) p(\theta)}$$

### 3. Results and discussion

We illustrate the above discussed models with help of mentioned data sets of banana, mango and wheat. We first use traditional regression methods to examine the

TABLE 3

## Parameter estimates of simple regression model for wheat

Coefficients	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value	Significance
B1	62.32	13.83	4.51	0.00	Yes
B2	3.21	4.50	0.71	0.49	No
B3	10.20	36.03	0.28	0.78	No
B4	1.48	7.12	0.21	0.84	No
B5	223.21	94.70	2.36	0.03	Yes
B6	-0.20	3.21	-0.06	0.95	No
B7	10.25	18.34	0.56	0.58	No
B8	-1.04	1.32	-0.79	0.44	No
B9	13.13	17.42	0.75	0.46	No
B10	-1.17	2.45	-0.48	0.64	No
B11	7.52	12.32	0.61	0.55	No

TABLE 4

## Parameter estimates of Bayesian regression model for banana

Coefficients	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value	Significance
B1	317.04	214.32	1.48	0.16	No
B2	65.75	14.00	4.70	0.00	Yes
B3	400.94	335.20	1.20	0.25	No
B4	-91.16	112.83	-0.81	0.43	No
B5	-481.95	390.82	-1.23	0.24	No
B6	-117.01	351.92	-0.33	0.75	No
B7	-1431.38	6528.48	-0.22	0.83	No
B8	3492.54	1384.81	2.52	0.02	Yes
B9	323.66	157.84	2.05	0.05	Yes
B10	2.31	11.82	0.20	0.84	No
B11	2033.01	891.70	2.28	0.04	Yes
B12	-628.52	344.26	-1.83	0.09	No

performance of these models. We calculate the estimated value of the parameters by using Ordinary Least Square (OLS) methods. We checked the significance status of the estimated value at 5% level of significance. The estimated value along with its standard errors, *t*-value are enlisted in the following table.

From Table 1, we found that only 16.67% of the total parameters are significant and 83.33%, a large percentage of parameters are non-significant at 5% significance level.

It can also be observed that the standard error of parameters is very high.

In Table 1, 38.46% of total parameters are significant whereas the rest of the parameters (61.54 %) are not only non-significant but also its standard error is very high.

In Table 3, it can be observed that the non-significance rate of the parameters is very high (81.82 %) and the standard error of the parameters is high. Only 18.18 % of the parameters are significant.

TABLE 5

## Parameter estimates of Bayesian regression model for mango

Coefficients	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value	Significance
B1	37.58	5.73	6.56	0.00	Yes
B2	8.14	92.02	0.09	0.93	No
B3	-0.41	42.14	-0.01	0.99	No
B4	-91.99	78.13	-1.18	0.26	No
B5	54.41	23.20	2.35	0.03	Yes
B6	3366.10	1819.72	2.25	0.04	Yes
B7	-2256.69	987.31	-2.29	0.04	Yes
B8	11.04	16.71	0.66	0.52	No
B9	-19.76	9.39	-2.10	0.05	Yes
B10	-680.15	337.38	-2.02	0.05	Yes
B11	87.23	271.66	0.32	0.75	No
B12	45.83	91.77	0.50	0.62	No
B13	-4.10	51.44	-0.08	0.94	No

TABLE 6

## Parameter estimates of Bayesian regression model for wheat

Coefficients	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value	Significance
B1	62.89	13.74	4.58	0.00	Yes
B2	3.01	4.56	0.66	0.52	No
B3	10.09	35.53	0.28	0.78	No
B4	1.71	7.18	0.24	0.81	No
B5	224.23	94.41	2.38	0.03	Yes
B6	-0.09	3.34	-0.03	0.98	No
B7	9.75	18.71	0.52	0.61	No
B8	-1.13	1.48	-0.76	0.46	No
B9	12.17	17.88	0.68	0.51	No
B10	-1.42	2.85	-0.50	0.62	No
B11	8.41	12.88	0.65	0.52	No

From the following result, we found that most of the estimate of the parameters is not significant at 5% level of significance. Thus, there is a scope to improve the parameter estimates by increasing the level of significance rate and reducing the standard error of the parameter estimates. Hence, we apply linear regression under the Bayesian framework for the given datasets to examine the improvement of the model. The Bayesian technique uses some prior knowledge about the parameters. To incorporate the prior knowledge to parameter, we select normal distribution, (a conjugate and informative prior) as

a prior distribution. To analyze the Bayesian linear regression model, we have used R software (version R-3.3.3) (<http://www.R-project.org/>) (R Core Team, 2013). The results of the posterior are enlisted in the following tables.

In Table 4, the non-significance rate of the parameters are reduced to 66.67 % from 83.33 % and the increase in significance rate to 33.33 % from 16.67 %. Standard errors of the parameters have been also reduced.

TABLE 7

RMSE comparison for banana

Data sets	RMSE for simple regression	RMSE for Bayesian regression
Banana	223.14	216.87
Mango	75.33	72.99
Wheat	266.35	265.85

In the case of mango data sets, in Table 5, we found that the non-significance rate of the parameter slightly decreases from the traditional regression model. The non-significance rate of the parameters of the Bayesian regression model is 53.84 %, earlier it was 61.54 % for the regression model. The standard error of the parameters has been also reduced.

In Table 6, we see that there is no reduction in the non-significance rate of the parameter but the standard errors of the parameters have reduced as compared to the traditional regression model.

From Table 7, comparison of RMSE, we find that there is a little decrease in the Root Mean Square Error (RMSE) of Bayesian regression models. Percentage in a reduction in RMSE for datasets banana, mango, and wheat are 0.03, 0.03 and 0.002 respectively.

#### 4. Conclusion

Our main objective in this study was to compare the traditional regression models with the modified regression model under the Bayesian framework. The efficiency and preciseness of regression analysis are non-questionable as it gives BLUE (Best Linear Unbiased Estimator) estimates under defined assumptions. But under certain circumstances, it fails to give an efficient estimate of parameters, or sometimes the parameter estimate of the model is not significant. In such situation, the Bayesian technique can be applied to improve the regression model and to get an efficient estimate of parameters. From our study we can empirically infer that first, regression under Bayesian reduce the non-significance rate of the parameters of the regression model, second, the Bayesian technique also helps to reduce the standard error of the parameters, third, it is also reduced the RMSE of the model which indicates that Bayesian regression provides more accurate forecast than the traditional regression model. This study will give an alternate methodology for estimating weather-based regression models where the standard regression approach fails to do so. The present investigation can be applied to various other datasets and results can be compared which will further enrich the literature of the weather-based modelling domain.

*Disclaimer* : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

#### References

- Agrawal, R., Jain, R. C. and Mehta, S. C., 2001, "Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis", *Indian Journal of Agricultural Science*, **71**, 7, 487-490.
- Bernardo, J. M. and Smith, A. F., 2001, "Bayesian theory", New York. Wiley Series.
- Bunn, D. W., 1975, "A Bayesian approach to the linear combination of forecasts", *Journal of the Operational Research Society*, **26**, 2, 325-329.
- Chatterjee, S. and Hadi, A. S., 2015, "Regression analysis by example", John Wiley & Sons.
- Chen, M. H. and Ibrahim, J. G., 2003, "Conjugate priors for generalized linear models", *Statistica Sinica*, **13**, 2, 461-476.
- Das, S., 2008, "Generalized linear models and beyond: An innovative approach from Bayesian perspective", Dissertation. University of Connecticut.
- Evans Sara, 2012, "Bayesian regression analysis", Dissertation, University of Louisville.
- Gommes, R., 1998, "Agrometeorological Crop Yield Forecasting Methods", The Netherlands : International Statistical Institute.
- Gregory, P., 2005, "Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support", Cambridge University Press.
- Hastings, W. K., 1970, "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, **57**, 1, 97-109.
- Marin, J. M., Pillai, N. S., Robert, C. P. and Rousseau, J., 2014, "Relevant statistics for Bayesian model choice", *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, **76**, 5, 833-859.
- Metropolis, N. and Ulam, S., 1949, "The Monte Carlo method", *Journal of the American statistical association*, **44**, 247, 335-341.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E., 1953, "Equation of state calculations by fast computing machines", *Journal of Chemical Physics*, **21**, 6, 1087-1092.
- Morris, C. N., 1983, "Parametric empirical Bayes inference: theory and applications", *Journal of the American Statistical Association*, **78**, 381, 47-55.
- Plackett, R. L., 1972, "Studies in the History of Probability and Statistics. XXIX : The discovery of the method of least squares", *Biometrika*, **59**, 2, 239-251.
- R Core Team, 2013, "R : A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria (<http://www.R-project.org/>).
- Sellam, V. and Poovammal, E., 2016, "Prediction of crop yield using regression analysis", *Indian Journal of Science and Technology*, **9**, 38, 1-5.
- Smith, A. F. and Roberts, G. O., 1993, "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods", *Journal of the Royal Statistical Society. Series B (Methodological)*, **55**, 1, 3-23.