



BAYESIAN ANALYSIS OF AGRICULTURAL EXPERIMENTS USING PROC MCMC

Mehmet Ziya FIRAT^{1*}


¹Akdeniz University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, 07058, Antalya, Turkey

Abstract: The purpose of this study is to present the general concept of Bayesian analysis and the Markov chain Monte Carlo (MCMC) algorithm and to make some numerical comparisons with frequentist analyses. A factorial randomized complete-block (RCB) experiment is used to analyze the cowpea data set that has four separate single-column replicates, each containing 9 combinations of 3 varieties and 3 spacings. Response is the yield of cowpea hay. Point estimates of variance components obtained in the Bayesian analysis under the priors presented some differences with the Restricted Maximum Likelihood (REML) estimate. The Bayesian method overestimates the variance component compared with the REML estimate. Bayesian method to agricultural experiments is a very rich and useful tool. It provides in depth study of different features of the data which are otherwise hidden and cannot be explored using other techniques. Moreover, SAS software has a power and efficiency to deal with the numerical as well as graphical features of data sets from agricultural experiments.

Keywords: Bayesian analysis, Agricultural experiments, Markov Chain Monte Carlo, PROC MCMC

*Corresponding author: Akdeniz University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, 07058, Antalya, Turkey

E mail: mziyafirat@gmail.com (M.Z. FIRAT)

Mehmet Ziya FIRAT  <https://orcid.org/0000-0002-0091-4713>

Received: February 04, 2021

Accepted: March 19, 2021

Published: July 01, 2021

Cite as: Firat MZ. 2021. Bayesian analysis of agricultural experiments using PROC MCMC. BSJ Agri, 4(3): 88-96.

1. Introduction

A main objective of most of agricultural experiments is to determine the effect of different treatments on a particular crop variety. Data collected from agricultural experiments have been analyzed using different statistical methods, such as analysis of variance, maximum likelihood, REML and Bayesian methods. Despite a great number of literatures on the design and analysis of agricultural field experiments, dating back to the early 1900s, and a large uptake of Bayesian methods in many other scientific fields in the last 30 years (Firat, et al., 1997a; Firat, et al., 1997b; Firat, 2001; Karaman, et al., 2014; Cemal, et al., 2016; Firat et al., 2016), Bayesian analysis for agricultural field experiments has not received very much attention.

Statistical methods for the design and analysis of agricultural field experiments were essentially developed by R. A. Fisher, F. Yates and many others. However, most of the modern courses and text-books focus on industrial and medical applications. Field trials are rather different, because a researcher always knows that generally she/he will obtain similar yields on two experimental units that are close together than on units that are further apart (Besag and Higdon, 1999), and also fields trials are conducted real-world settings.

Bayesian methods improve upon frequentist methods by expressing uncertainty regarding the unknown parameters and simplifies the interpretation of the results, especially in ranking and selection of crop

varieties. Moreover, an analysis of complex formulations can be carried out with comparative ease, and computation of complicated functionals of high dimensional posterior distributions can be done by using MCMC methods. Although, faster computers and increasing popularity of MCMC methods have allowed Bayesian methods to become widely used in complex data analysis problems, the Bayesian approach has yet to provide a completely satisfactory answer in the analysis of agricultural experiments, since there has been a lack of application in this area.

Besag and Higdon (1993, 1999) and Besag et al. (1995) discussed Bayesian approaches for analyzing agricultural field experiments. They proposed complex formulations for situations when spatial effects were considered, while our approach is for the standard additive mixed model. Our approach has some advantages over other Bayesian approaches; the marginals of complex functions of the unknown parameters can be easily obtained and implementation of high dimensional posterior probability functions can be conveniently done.

In this research, the general concept of Bayesian analysis and the MCMC algorithm are presented. The implementation of the MCMC algorithm using the PROC MCMC procedure of SAS software package (SAS Institute, 2004) is demonstrated through a real data set from an agricultural experiment. Some numerical comparisons with frequentist analyses are made. We shall have rather little to say about MCMC sampling. The analyses in this



paper were all carried out using simple Gibbs samplers. In organizing the paper, we decided to present the methodology first, followed by an application. We describe some Bayesian formulations for a particular field trial, in which (a) the treatment or variety effects have no special structure, and (b) Gaussian assumptions are appropriate, both in the likelihood and in the prior. We contrast Bayesian and frequentist formulations. Finally, the last section contains some discussion.

2. Material and Methods

2.1. Data Set and Descriptive Statistics

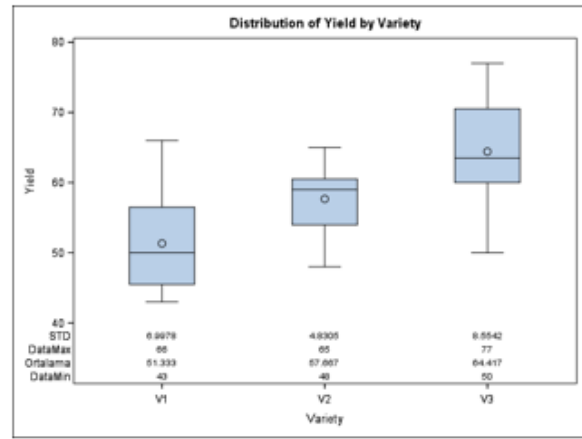
In this section, we analyze a factorial randomized complete-block (RCB) experiment on cowpea. A data set is reported in Snedecor and Cochran (1989), page 308. There are four separate single-column replicates, each containing 9 combinations of 3 varieties and 3 spacings. Response is the yield of cowpea hay (lb/100 morgen plot) and the corresponding yields are given in Table 1. Descriptive statistics of Cowpea data set and boxplots of main factor effects and interactions are displayed in Table 2 and Figure 1, respectively, for the purpose of illustration. We note here that the basic Bayesian formulation, with Gaussian assumptions for variety, spacing and yield but a vague prior applied to varieties and spacings, produces variety and spacing effects that agree closely with those from a superficially similar frequentist extended first-differences analysis. The same holds for the standard deviations and corresponding standard errors for variety differences.

Table 1. Data from Cowpea variety trial. The original dataset was published by Snedecor and Cochran (1989, p. 308). This data set is called 'Cowpea data' in this study

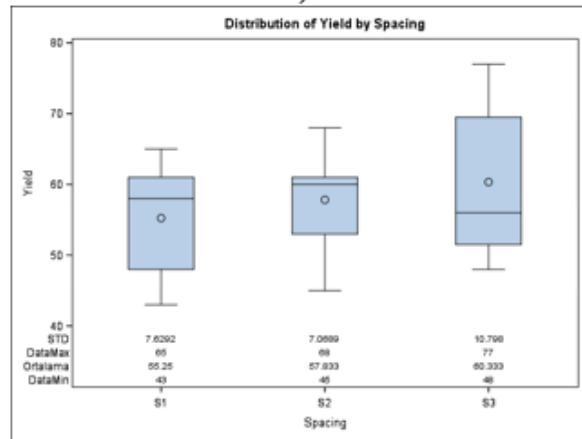
Variety	Spacing	Blocks			
		B1	B2	B3	B4
V1	S1	56	45	43	46
	S2	60	50	45	48
	S3	66	57	50	50
V2	S1	65	61	60	63
	S2	60	58	56	60
	S3	53	53	48	55
V3	S1	60	61	50	53
	S2	62	68	67	60
	S3	73	77	77	65

Table 2. Descriptive statistics of Cowpea data set

Variety	Spacing	Mean	Stdev.	Min.	Max.
V1	S1	47.50	5.802	43	56
	S2	50.75	6.500	45	60
	S3	55.75	7.588	50	66
V2	S1	62.25	2.217	60	65
	S2	58.50	1.915	56	60
	S3	52.25	2.986	48	55
V3	S1	56.00	5.354	50	61
	S2	64.25	3.862	60	68
	S3	73.00	5.657	65	77



a)



b)

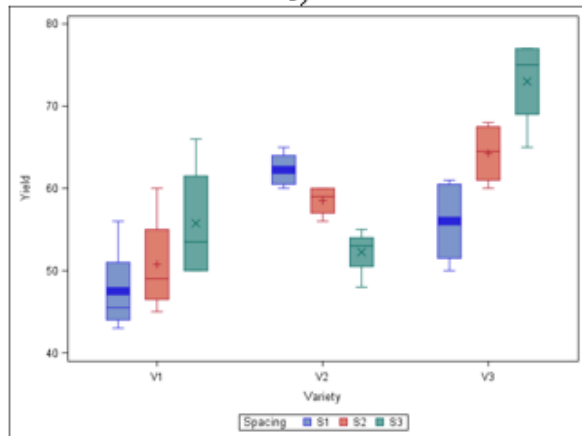


Figure 1. Boxplots of main factor effects and interactions: (a) boxplot of variety effect, β_j 's, (b) boxplots of treatment (spacings) effects, γ_k (c) boxplots of interactions $(\beta\gamma)_{jk}$'s.

2.2. Bayesian Formulation of Field Experiments

For definiteness, we focus on a full factorial Randomized Block Design (RBD) with two three-level treatment factors (variety and spacing) occurring in a factorial structure. The randomized block design is used to control and reduce experimental error. It is intended to make yield comparisons between the levels of treatment factors of a crop. We assume that measurements are effectively continuous, resulting in an n -vector y of yields over n rectangular plots. However, our formulation

extends to more complicated treatment structures and to discrete observations, such as litter size or number of laid eggs per month.

We consider the additive fixed model, with one observation per cell. The model for a two-factor factorial in a randomized complete block (equation 1) is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + e_{ijk} \quad \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \end{cases} \quad (1)$$

where y_{ijk} is the observed response when factor A is at the j th level and factor B at the k th level in the i th block, μ is the overall mean effect, α_i is the effect of i th block, β_j is the effect of j th level of factor A, γ_k is the effect of k th level of factor B, $(\beta\gamma)_{jk}$ is the effect of the interaction between factor A and B, and e_{ijk} is a random

$$f(\{y_{ijk}\} | \mu, \{\alpha_i\}, \{\beta_j\}, \{\gamma_k\}, \{(\beta\gamma)_{jk}\}, \sigma_e^2) = f(\mathbf{y} | \boldsymbol{\theta})$$

$$\propto (\sigma_e^2)^{-\frac{abc}{2}} \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_k - (\beta\gamma)_{jk})^2}{\sigma_e^2} \right] \right\} \quad (2)$$

Prior distributions for variety or treatment effects: In order to carry out Bayesian analysis, in addition to the likelihood function we need prior distributions for each parameter of the model, μ , α_i , β_j , γ_k , $(\beta\gamma)_{jk}$ and σ_e^2 . We consider a conjugate prior for each of these parameters. For the overall mean, μ , a flat prior was used, so that $f(\mu) \propto \text{constant}$, indicating no prior knowledge about this parameter. The simplest choice of prior for the variety effect, α_i , in a single trial is either a uniform distribution or a Gaussian distribution, when the data provide little evidence of differences between varieties. In this paper, our prior is the simple Gaussian or uniform distribution, $\{\alpha_i\} \propto \text{constant}$: the latter is useful mainly in making numerical comparisons with standard frequentist analyses. For the priors, a Uniform distribution was also assumed for other fixed effects β_j , γ_k and $(\beta\gamma)_{jk}$, $\{\beta_j\} \propto \text{constant}$, $\{\gamma_k\} \propto \text{constant}$ and

error component. Both factors are assumed to be fixed. Similarly, the interaction effects are fixed. The experimental errors are assumed independent and normally distributed with zero means and common variance σ_e^2 . To our knowledge, an analysis of the agricultural field data using PROC MCMC has never been done before, and it will be the first time to use the model in equation 1 under new priors to make Bayesian inferences.

The conditional distribution of $\{y_{ijk}\}$ given μ , α_i , β_j , γ_k , $(\beta\gamma)_{jk}$ and σ_e^2 is

$$\{y_{ijk}\} | \mu, \{\alpha_i\}, \{\beta_j\}, \{\gamma_k\}, \{(\beta\gamma)_{jk}\}, \sigma_e^2 \sim N(y_{ijk} | \mu - \alpha_i - \beta_j - \gamma_k - (\beta\gamma)_{jk}, \sigma_e^2)$$

Under this assumption of normality, the likelihood function (equation 2) is as follows:

$\{(\beta\gamma)_{jk}\} \propto \text{constant}$, respectively. Finally, for the variance component σ_e^2 , a diffuse but proper $\sigma_e^2 | \nu_e, S_e^2 \sim IG(\nu_e/2, \nu_e S_e^2/2)$ prior (i.e. an inverse gamma distribution) was assigned (equation 3).

$$f(\sigma_e^2 | \nu_e, S_e^2) \propto (\sigma_e^2)^{-\frac{1}{2}(\nu_e+2)} \exp \left\{ -\frac{\nu_e S_e^2}{2\sigma_e^2} \right\} \quad (3)$$

where S_e^2 and ν_e are scale and shape (degrees of freedom) parameters for variance component respectively.

Posterior distribution: By multiplying likelihood function with the prior distributions of all the parameters, the joint posterior density of parameters (equation 4) is obtained as:

$$f(\boldsymbol{\theta} | \mathbf{y}) \propto L(\mathbf{y} | \boldsymbol{\theta}) f(\mu) f(\{\alpha_i\}) f(\{\beta_j\}) f(\{\gamma_k\}) f(\{(\beta\gamma)_{jk}\}) f(\sigma_e^2) \propto (\sigma_e^2)^{-\frac{(abc+\nu_e+2)}{2}} \times \exp \left\{ -\frac{1}{2} \left[\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_k - (\beta\gamma)_{jk})^2 + \nu_e S_e^2}{\sigma_e^2} \right] \right\} \quad (4)$$

To implement the Gibbs sampling algorithm, we require the full conditional posterior distributions of $\mu, \alpha_i, \beta_j, \gamma_k, (\beta\gamma)_{jk}$ and σ_e^2 . The full conditional posterior distribution of any parameter of interest can be obtained by integrating over the remaining parameters from joint posterior distribution. It is well known that the conjugate priors are very easy to work with because the posterior and prior have the same distributional form and the effect of the data is just to update the parameters from the prior to the posterior. Therefore, the resulting full conditional posterior distributions of $\mu, \alpha_i, \beta_j, \gamma_k, (\beta\gamma)_{jk}$ and σ_e^2 (equations 5-10) are summarized as follows:

$$[\mu] \sim N\left(\frac{\bar{y}_{..} - \bar{\alpha} - \bar{\beta} - \bar{\gamma} - (\bar{\beta\gamma})}{abc}, \frac{\sigma_e^2}{abc}\right) \quad (5)$$

$$[\alpha_i] \sim N\left(\frac{\bar{y}_{i.} - \mu - \bar{\beta} - \bar{\gamma} - (\bar{\beta\gamma})}{bc}, \frac{\sigma_e^2}{bc}\right) \quad (6)$$

$$[\beta_j] \sim N\left(\frac{\bar{y}_{.j} - \mu - \bar{\alpha} - \bar{\gamma} - (\bar{\beta\gamma})}{ac}, \frac{\sigma_e^2}{ac}\right) \quad (7)$$

$$[\gamma_k] \sim N\left(\frac{\bar{y}_{.k} - \mu - \bar{\alpha} - \bar{\beta} - (\bar{\beta\gamma})}{ab}, \frac{\sigma_e^2}{ab}\right) \quad (8)$$

$$[(\beta\gamma)_{jk}] \sim N\left(\frac{\bar{y}_{.jk} - \mu - \bar{\alpha} - \bar{\beta} - \bar{\gamma}}{a}, \frac{\sigma_e^2}{a}\right) \quad (9)$$

$$[\sigma_e^2] \sim IG\left(\frac{abc + v_e}{2}, \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_k - (\beta\gamma)_{jk})^2 + v_e S_e^2}{2}\right) \quad (10)$$

As can be seen from equations 5-10, the first five conditional distributions are from the normal and the last one is from the inverse chi-square distributions, therefore only two random number generators are required in this problem, the normal variable generator and the scaled inverse chi-square variable generator. Since every unknown has a closed form distribution, the Gibbs sampler algorithm can be used for the MCMC experiment. The Gibbs sampling algorithm generates random samples from the full conditional distributions of the parameters, without having to calculate the density. Gibbs sampling algorithm requires an initial starting point for the parameters. Then, one at a time, a value for each parameter of interest is sampled given the values for the other parameters and data. Once all of the parameters of interest have been sampled, the nuisance parameters are sampled given the parameters of interest and the observed data. At this point, the process is started over. The power of Gibbs sampling is that the joint distribution of the parameters will converge to the joint probability of the parameters given the observed data.

2.3. Bayesian Analysis of Cowpea Data Set Using PROC MCMC

There are two steps involved in data analysis using SAS, (1) First the data step and (2) Second the procedure step. The data step is used to input the data. The statistical analyses are performed in the procedure step via a built-in subroutine within the SAS system. Each subroutine is called a procedure performing some specific tasks. There is one particular SAS procedure called the MCMC which is designed for the MCMC implemented Bayesian analysis and handles problems with a high level of complexity. SAS Code for the data and procedure steps

```
data cowpea;
input Blok Variety $ Spacing $ Yield;
A=Blok; B=Variety; C=Spacing; y=Yield;
cards;
1 V1 S1 56
2 V1 S1 45
3 V1 S1 43
4 V1 S1 46
1 V1 S2 60
2 V1 S2 50
3 V1 S2 45
4 V1 S2 48
1 V1 S3 66
2 V1 S3 57
3 V1 S3 50
4 V1 S3 50
1 V2 S1 65
2 V2 S1 61
3 V2 S1 60
.....
1 V3 S3 73
2 V3 S3 77
3 V3 S3 77
4 V3 S3 65
run;

*nmc = specifies the number of MCMC iterations;
*nbi =specifies the number of number of burn-in
iterations;
*thin = specifies the thinning rate ;
*plot = produces plots;
*monitor = gives output of a list of symbols;
*array = gives a list of array elements;
*parms = gives a list of parameters in the model;
*prior = specifies the prior distribution of the
parameters;
*model = specifies the likelihood function;
*call = computes the statistics;

proc mcmc data=recodedb outpost=postb
propcov=quanew seed=&seed nmc=50000
nbi=100000 thin=100 plots=all
monitor = (beta1-beta&nvar sigmae diffV1V2 diffV1V3
diffV2V3 diffS1S2 diffS1S3 diffS2S3 );
array covar[&nvar] intercept &_trgind;
```

```
array beta[&nvar] ;
parms sigmae1 ;
parms (beta1-beta&nvar) 0;
prior beta:~normal(0,var=100000);
prior sigmae~igamma(shape=0.001, scale=0.001);
* Differences between Varieties ;
diffV1V3 = beta5;
diffV2V3 = beta6;
diffV1V2 = beta5 - beta6;
* Differences between Spacings ;
diffS1S3 = beta7;
diffS2S3 = beta8;
diffS1S2 = beta7 - beta8;
call mult(covar, beta, mu);
model y ~ normal(mu, var=sigmae);
run;
ods graphics off;
ods rtf close;
```

Statistical analyses for REML estimations were obtained using PROC MIXED procedure, and Bayesian analysis was conducted using the PROC MCMC procedure of SAS software in the analysis of Cowpea data. A single chain of size 500000 iterations was run. The initial 100000 iterations were discarded as a burn-in, and every 100th sample was recorded to reduce the auto-correlation. In total, 5000 samples were stored for each parameter, and means of the sample values were used as an estimate of the parameters.

3. Results

The summary statistics for all the variables, including functions of the parameters and the variance component from REML and MCMC approaches for the analysis of Cowpea data set are presented in Table 3. The posterior means are based on 5000 Gibbs sampler.

Table 3. Summary statistics for all the variables from REML and MCMC methods

Parameter	Posterior Summaries								
	REML	Mean	SD ¹	Posterior Intervals				M. Carlo SEs	
				HPD ² Interval		Equal-Tail Interval		MCSE ³	MCSE/SD
beta1, μ	70.75	70.773	2.530	65.908	75.755	65.909	75.758	0.037	0.015
beta2, α_1	6.11	6.095	2.067	1.927	9.938	2.018	10.155	0.030	0.015
beta3, α_2	3.33	3.296	2.070	-0.655	7.413	-0.723	7.368	0.032	0.016
beta4, α_3	-0.44	-0.459	2.134	-4.643	3.765	-4.699	3.739	0.031	0.015
beta5, β_1	-17.25	-17.172	3.119	-23.118	-11.021	-23.238	-11.108	0.044	0.014
beta6, β_2	-20.75	-20.794	3.076	-26.639	-14.437	-26.974	-14.737	0.045	0.015
beta7, γ_1	-17.00	-17.039	3.102	-23.464	-11.234	-23.312	-11.034	0.047	0.015
beta8, γ_2	-8.75	-8.744	3.183	-14.837	-2.349	-14.967	-2.419	0.047	0.015
beta9, $(\beta\gamma)_{11}$	8.75	8.729	4.465	-0.364	17.060	-0.138	17.406	0.067	0.015
beta10, $(\beta\gamma)_{12}$	3.75	3.661	4.497	-5.169	12.468	-5.422	12.358	0.065	0.014
beta11, $(\beta\gamma)_{21}$	27.00	27.111	4.381	18.173	35.323	18.606	35.789	0.068	0.016
beta12, $(\beta\gamma)_{22}$	15.00	15.011	4.390	6.306	23.659	6.312	23.679	0.064	0.015
Sigmae, σ_e^2	17.67	19.407	6.345	9.958	32.108	10.794	34.883	0.091	0.014
diffV1V2	3.50	3.623	3.125	-2.729	9.692	-2.576	9.939	0.044	0.014
diffV1V3	-17.25	-17.172	3.119	-23.118	-11.021	-23.238	-11.108	0.044	0.014
diffV2V3	-20.75	-20.794	3.076	-26.639	-14.437	-26.974	-14.737	0.045	0.015
diffS1S2	-8.25	-8.296	3.155	-14.793	-2.364	-14.633	-2.100	0.046	0.015
diffS1S3	-17.00	-17.039	3.102	-23.464	-11.234	-23.312	-11.034	0.047	0.015
diffS2S3	-8.75	-8.744	3.183	-14.837	-2.349	-14.967	-2.419	0.047	0.015

¹SD= standard deviation, ²HPD= (95%), the 95% highest posterior density credible interval, ³MCSE= monte carlo standard error, REML= Restricted maximum likelihood, MCMC= Markov chain monte carlo.

It can be noted that the Bayesian method overestimates the variance component compared with the REML estimate. The variance component obtained by REML is

only marginal with respect to fixed effects but conditionals to other nuisance parameters of the model. The Bayesian analysis allows further marginalization via

Markov Chain Monte Carlo methods. This approach is particularly interesting for models, as the present, with high number of variance components. In consequence, point estimates of variance components obtained in the Bayesian analysis under that priors presented some differences with the REML estimate. These differences are due to the prior information.

The 95% High Probability Density (HPD) interval is the same as the equal tail intervals due to the normality of the posterior distribution. The equal-tail credibility intervals and the HPD intervals all show that all the pairwise differences are significantly different from zero (Table 3). The two varieties and two spacings have significant effect on the yield of Cowpea, i.e., $\beta_1, \beta_2, \gamma_1$ and γ_2 are significantly different from zero. Two of the interaction effects are also significant, i.e., $(\beta\gamma)_{21}$ and $(\beta\gamma)_{22}$ are different from zero.

Table 4 shows the Geweke z test for convergence and other diagnostic statistics for all the unknowns. If the result of the Geweke z test is significant, the chain may

not have converged. From this table it can be clearly seen that the Markov chains behave very well for all the unknowns, since the p values for the Geweke z-test are larger than 0.05 for all unknowns. Because the autocorrelation is always positive, the effective sample size is always less than the actual posterior sample size. A much smaller effective sample size than the actual size indicates poor mixing of the Markov chain. The concept of effective sample size is much the same as the effective population size in population genetics. Our results show that effective sample sizes are very close to the actual posterior sample sizes.

Figure 2 shows the posterior TAD (trace-autocorrelation-density) panels for α_1 (beta2) and β_1 (beta5), only. The Markov chain converges very well with very low autocorrelation and almost a perfect normal posterior distribution in all TAD panels representing different parameters. Overall, this dataset is sufficient to allow more precise estimates of the parameters.

Table 4. Diagnostic test statistics for the Markov chain convergence of the Cowpea data

Parameter	Geweke Diagnostics		Effective Sample Sizes			Posterior Autocorrelations		
	z	Pr> z	ESS ¹	Time	Efficiency	Lag 1	Lag 5	Lag 10
beta1, μ	0.851	0.3947	4593.1	1.0886	0.9186	0.0244	-0.0292	-0.0180
beta2, α_1	-0.289	0.7725	4601.2	1.0867	0.9202	0.0433	0.0009	-0.0092
beta3, α_2	-0.876	0.3812	4159.7	1.2020	0.8319	0.0164	-0.0015	0.0126
beta4, α_3	0.178	0.8588	4684.4	1.0674	0.9369	0.0337	-0.0001	0.0151
beta5, β_1	0.448	0.6539	5000.0	1.0000	1.0000	0.0035	0.0040	-0.0268
beta6, β_2	-0.445	0.6562	4725.4	1.0581	0.9451	0.0291	-0.0277	-0.0213
beta7, γ_1	-0.557	0.5778	4346.6	1.1503	0.8693	0.0521	-0.0078	-0.0072
beta8, γ_2	-1.542	0.1232	4536.7	1.1021	0.9073	0.0404	0.0072	-0.0072
beta9, $(\beta\gamma)_{11}$	0.322	0.7474	4463.0	1.1203	0.8926	0.0428	-0.0058	-0.0054
beta10, $(\beta\gamma)_{12}$	0.541	0.5887	4818.2	1.0377	0.9636	0.0189	0.0231	0.0024
beta11, $(\beta\gamma)_{21}$	0.476	0.6342	4160.8	1.2017	0.8322	0.0540	-0.0348	-0.0174
beta12, $(\beta\gamma)_{22}$	1.119	0.2629	4741.8	1.0545	0.9484	0.0272	0.0036	-0.0078
Sigmae, σ_e^2	0.468	0.6395	4825.4	1.0362	0.9651	0.0181	-0.0051	-0.0119
diffV1V2	0.818	0.4136	5000.0	1.0000	1.0000	-0.0055	-0.0164	-0.0016
diffV1V3	0.448	0.6539	5000.0	1.0000	1.0000	0.0035	0.0040	-0.0268
diffV2V3	-0.445	0.6562	4725.4	1.0581	0.9451	0.0291	-0.0277	-0.0213
diffS1S2	0.972	0.3309	4687.8	1.0666	0.9376	0.0333	-0.0037	-0.0088
diffS1S3	-0.557	0.5778	4346.6	1.1503	0.8693	0.0521	-0.0078	-0.0072
diffS2S3	-1.542	0.1232	4536.7	1.1021	0.9073	0.0404	0.0072	-0.0072

¹ESS= effective sample size.

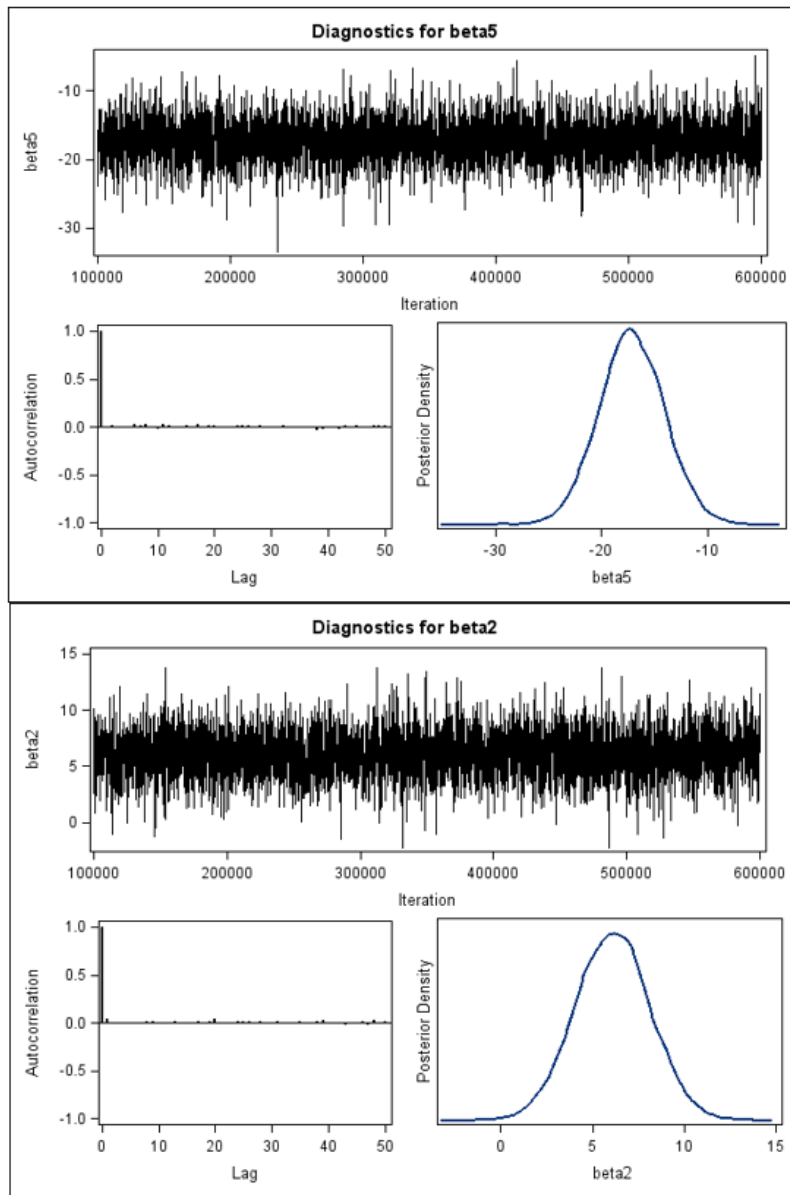


Figure 2. Posterior TAD panels for Markov chain convergence diagnosis from the model α_1 (beta2) and β_1 (beta5).

4. Discussion

In this paper, we have presented the general concept of Bayesian methodology and the MCMC algorithm for the analysis of agricultural field experiments, a subject that has received not much previous attention despite an enormous number of frequentist literatures, in a way that can be understood by agricultural practitioners. We also demonstrated the implementation of the MCMC algorithm using PROC MCMC procedure of SAS software package to obtain posterior distribution of parameters of interest through a real data set from a two-factor factorial randomized complete-block (RCB) design. Bayesian approach is compatible with factorial experiments when studying interactions. In two factors full factorial experiment, the posterior estimates of the means of mains effects and interactions were obtained and compared with those under the likelihood-based method, REML.

It is always useful to compare and contrast the results of BSJ Agri / Mehmet Ziya FIRAT

Bayesian analysis with that of the REML analysis. If they are completely different or not comparable in any way, there are at least three approaches to consider; new analyses with different models, the use of different priors and analysis of simulated data to verify the model and the priors. In the agricultural field data analysis, we used a new model under new priors, and produced results using PROC MCMC that have never been reported before. Then, we analyzed the same data using PROC MIXED to obtain the REML estimates under the mixed model. The two results do share some similarity. Based on the results from our data set, REML estimations of the unknown parameters are almost similar with MCMC posterior means. We can conclude that the estimates of REML are accurate but the posterior point estimates from the MCMC algorithm can be overestimated depending on the nature of the data set. The differences in the results of different estimation methods (REML and Bayesian) occurred the most in the estimation of error variance.

Such a comparison increases our confidence in the Bayesian analysis.

The original Bayesian method is more complicated than the classical maximum likelihood-based method REML, because multiple integrals are often involved in obtaining the posterior expectations of the unknown parameters. In most situations, an explicit form of the multiple integrals does not exist, and thus limits the application of Bayesian analysis. Although Bayesian inference was proposed earlier than the likelihood-based inference, it has only recently become popular due to the advent of high computing power and the advanced MCMC algorithms for numerical integrations. With the MCMC implemented Bayesian method, it has become much easier to adopt complicated models. Since it is often very simple to obtain the fully conditional posterior distributions, the MCMC process is much easier to understand than the maximum likelihood method. Thanks to the MCMC algorithm, which has revolutionized the field of Bayesian inference, the non-statisticians can also perform Bayesian analysis. Conducting an MCMC sampling process is no more complicated than doing an agricultural field experiment.

Frequentist approaches to making inferences about the parameters of interest in general linear models have several limitations and may not be able to handle complicated models. These include reliance on asymptotic theory and a failure to account for uncertainty for model parameters. A Bayesian approach to making inferences about the unknown parameters is proposed that circumvents many of the problems associated with alternative frequentist approaches. Markov Chain Monte Carlo (MCMC) and Gibbs sampling are used to obtain posterior point estimates from the posterior distributions. The 95% credible intervals (CI) were also obtained and finally compared with that obtained using classical approach. The Bayesian method for agricultural field experiments is useful to both researchers and students who will appreciate the importance of Bayesian approach when applied to practical statistical problems.

One of the main differences between the Bayesian and likelihood-based approaches is the way in which they deal with nuisance parameters (Smith and Naylor, 1987). This is apparent from our results. The conditional posterior density is obtained by a Monte Carlo numerical integration method, which is known as a Gibbs Sampler, whereas the likelihood function is obtained by maximizing with respect to the nuisance parameters. In certain cases, the two operations may produce sharply contrasting results. The computations required to implement the Bayesian method are of the same order of magnitude as those required for the REML method, and therefore the Bayesian method are likely to be computationally feasible whenever the REML methods are computationally feasible.

Implementation of the Bayesian method not only simplifies the interpretation of the results, especially in

ranking and selection of the varieties, but also enables the researcher to analyze complex formulations with comparative ease, by using Markov chain Monte Carlo approaches in any agricultural field trial. Bayesian estimators depend on the information about the parameters contained in the data, and also on prior knowledge. This is one of the potential advantages of the Bayesian methods. Therefore, it is expected that the Bayesian method will do better than the classical procedures when the data contain little information about the parameters of interest. Moreover, the Bayesian method implicitly account for the uncertainty about the values at the parameters of interest.

Finally, we can conclude that the Bayesian method of estimation using the Gibbs sampling approach is suitable for estimating the unknown parameters under a full factorial Randomized Block Design (RBD) with two three-level treatment factors as compared to traditional methods, particularly for small sample data sets. It is also feasible computationally and appears to give much more sensible answer to the inferential problems than likelihood-based estimation methods. Indeed, we have maintained that Bayesian inference has some important practical advantages in analyzing field experiments. For example, the results are easier to interpret, particularly in ranking and selection of animals for next generations and when communicating with non-statisticians; the results from previous experiments can be incorporated in a rather natural manner into the prior for treatments or varieties in a subsequent trial, and there seems more freedom in using MCMC methods to analyze reasonably realistic formulations and to address model uncertainty.

5. Conclusion

It is also clear from our study that Bayesian method to agricultural experiments is a very rich and useful tool. It provides in depth study of different features of the data which are otherwise hidden and cannot be explored using other techniques. Moreover, SAS software has a power and efficiency to deal with the numerical as well as graphical features of data sets from agricultural experiments. Our Bayesian method uses Markov chain Monte Carlo (MCMC) approach and conjugate priors and balanced data. Simulating from full conditionals can also be easily done for the analysis of unbalanced data with possibly nonconjugate priors using the SAS PROC MCMC codes presented here, without leading one to consider alternative Markov chain Monte Carlo schemes.

Author Contributions

All tasks was done by the single author.

Conflict of Interest

The author declare that there is no conflict of interest.

References

Besag J, Green PJ, Higdon DM, Mengersen, KL. 1995 Bayesian computation and stochastic systems. *Stat Sci*, 10: 3-66.

- Besag J, Higdon DM. 1993. Bayesian inference for agricultural field experiments. *Bull Int Statist*, 55: 121-136.
- Besag J, Higdon, DM. 1999. Bayesian analysis of agricultural field experiments. *J Royal Stat Society B*, 61: 691-746.
- Cemal I, Karaman E, Firat MZ, Yilmaz O, Ata N, Karaca O. 2016. Bayesian inference of genetic parameters for ultrasound scanning traits of Kivircik lambs. *Animal*, 11(3): 375-381.
- Firat M.Z, Karaman E, Kaya Basar E, Narinc D. 2016. Bayesian analysis for the comparison of nonlinear regression model parameters: an Application to the growth of Japanese quail. *Brazilian J Poultry Sci*, 18: 19-26.
- Firat MZ, Theobald CM, Thompson R. 1997a. Univariate analysis of test day milk yields of British Holstein Friesian heifers using Gibbs sampling. *Acta Agri Scand Section A, Animal Science*, 47: 213-220.
- Firat MZ, Theobald CM, Thompson R. 1997b. Multivariate analysis of test day milk yields of British Holstein Friesian heifers using Gibbs sampling. *Acta Agri Scand Section A, Animal Science*, 47: 221-229.
- Firat MZ. 2001. Bayesian analysis of test day milk yields in an unbalanced mixed model assuming random herd-year-month effects. *Turk J Vet Anim Sci*, 25: 327-333
- Karaman E, Firat MZ, Narinc D. 2014. Single-Trait Bayesian analysis of some growth traits in Japanese. *Brazilian J Poultry Sci*, 16: 51-56.
- SAS Institute, 2004. SAS version 9.1.3. SAS Institute Inc., Cary, NC, USA.
- Smith RL, Naylor JC. 1987. A Comparison of Maximum Likelihood and Bayesian Estimators for the three-parameter Weibull Distribution. *J Royal Stat Society Series C*, 36: 358-369.
- Snedecor GW, Cochran WG. 1989. *Statistical methods*. Iowa State University Press, Ames, USA, 8th ed., pp 308.