CBCS
HTTP://AGROBIOL.SGGW.WAW.PL/CBCS

COMMUNICATIONS IN BIOMETRY AND CROP SCIENCE
VOL. 10, NO. 2, 2015, PP. 82–95

INTERNATIONAL JOURNAL OF THE FACULTY OF AGRICULTURE AND BIOLOGY,
WARSAW UNIVERSITY OF LIFE SCIENCES – SGGW, POLAND

REGULAR ARTICLE

Bayesian estimation of genotype-by-environment interaction in sorghum variety trials

Siraj Osman Omer[1*], Abdel Wahab Hassan Abdalla[2], Mohammed Hamza Mohammed[3], Murari Singh[4]

[1]Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan
[2]Department of Agronomy, Faculty of Agriculture, University of Khartoum, Sudan
[3]Cereal Research Center, Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan
[4]International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan
*Corresponding author: Siraj Osman Omer; E-mail: sirajstat@yahoo.com

CITATION: Omer S.O., Abdalla A.W.H., Mohammed M.H., Singh M. (2015). Bayesian estimation of genotype-by-environment interaction in sorghum variety trials. *Communications in Biometry and Crop Science* 10, 82–95.

Received: 12 January 2015, Accepted: 2 September 2015, Published online: 27 September 2015

ABSTRACT

Genotype × Environment Interaction (GEI) plays an important role in identifying genotypes for high and stable yield for broad and specific adaptations. It continues to be a challenging issue among plant breeders and agronomists when conducting crop performance trials across diverse and unpredictable environments. Normally, the analysis of GEI is carried out under the frequentist paradigm, even though ongoing crop improvement programs gather information on genotypic and experimental error parameters that could be incorporated using a Bayesian approach. The objective of this paper was to estimate, for sorghum (*Sorghum bicolor*) in Sudanese environments, genotypic and GEI variances, heritability, genetic advance attributable to selection, and genotype means using Bayesian and frequentist approaches. Eighteen genotypes of sorghum were evaluated in randomized complete block designs with four replicates in six environments, during 2009/10 - 2011/12, at South-Gedarif and North-Gedarif in Sudan. Priors were obtained from a previous set of multi-environment trials in sorghum during 2006/7 – 2008/9 at Rahab, Sudan. Estimates of heritability and genetic advance under the Bayesian approach were higher than those under the frequentist approach. Precision of means of genotypes and heritability estimates were also higher under the Bayesian approach. The Bayesian approach provides a wider coverage for statistical inference and incorporates prior information with the likelihood of current data. For this approach, an illustrative step-by-step procedure is presented and recommended for use in statistical analysis of crop genotypes from multi-environment trials.

Key Words: *Bayesian analysis; crop variety trials; genotype-by-environment interaction; heritability; genetic gain.*

INTRODUCTION

Genotype × Environment Interaction (GEI) in crop traits arises due to a complex interplay among physiological and genetic factors, and biotic and abiotic environmental factors. Detection and exploitation of GEI is essential for identifying genotypes for high and stable yield. Vast literature is dedicated to this subject in relation to general methodologies and applications on specific crops. Among others, Lin et al. (1986), Gauch (1988), Cooper and Hammer (1996), DeLacy et al. (1996), Yan et al. (2007), Smith et al. (2005) and Sarker et al. (2010) deal with various aspects of the analysis of multi-environment trials (METs). Selection methods for genotypes for specific responses to environments, e.g. broad and specific adaptations, were examined by searching for underlying patterns of GEI. Kempton (1984) proposed to use bi-plots in explaining GEI in crop variety trials. Gauch (1988) modelled the GEI matrix as a low-rank approximation by singular value decomposition, which is known as Additive Main Effects and Multiplicative Interaction (AMMI) model. Later on, adding the genotype mean to the GEI to measure genotype performance in each environment, Yan et al. (2000) applied another type of multiplicative decomposition and displayed its results a bi-plot (GGE-biplot). A review of this technique is given in Yan (2011).

The majority of GEI investigations in the above studies have been carried out by using the frequentist approach, which bases on the likelihood of current data but does not make use of any prior information. On the other hand, the Bayesian approach uses such prior information available in the data collected in past/ongoing crop improvement programs and, therefore, possesses a much higher potential for statistical inference on GEI and other parameters of interest.

To understand the difference more clearly, suppose we wish to estimate a parameter θ using an observed data vector $\underline{y} = (y_1, ..., y_n)'$ where the likelihood of observing $\underline{y}$ is a function of θ, i.e., $f(\underline{y}|\theta)$. A series of available estimates of θ may provide prior information or belief in θ, which may be expressed as a probability distribution function, i.e., $g(\theta)$. Under the Bayesian approach and using Bayes' Theorem, inference on θ is obtained as conditional probability distribution of θ given the data $\underline{y}$, which is given by $f(\theta|\underline{y}) \propto g(\theta) f(\underline{y}/\theta)$. This expression, called the posterior density function of θ, is numerically obtainable by using the methods described in standard texts (Gelman et al. 2004). In the context of multi-environment trials, the parameter θ would be a vector of various main effects, interactions and their variance components, which are to be inferred. In Bayesian formulation, all model parameters are treated as random variables, and a prior probability distribution is specified for each of them.

Crop variety field trials are normally conducted by using block designs (Cochran and Cox 1957, Patterson and Williams 1976, Hinkelmann and Kempthorne 2005). Theobald et al. (2002) applied a Bayesian approach and predicted regional and local-area yields from crop variety trials. In another application of the Bayesian approach to the data from a small complete block design experiment, Forkman and Piepho (2012) found a smaller mean-square error and more accurate coverage of prediction intervals for means, compared to best linear unbiased predictors. Singh et al. (2015) presented a step-by-step procedure for Bayesian analysis of trials conducted in complete or incomplete block designs and provided the computational codes. In MET data analysis, heterogeneity in GEI variances and/or experimental error variances were addressed by Edwards and Jannink (2006), with the exponential of an additive model with random effects for genotypes and environments, where suitable priors for the variance components were assigned and fitted under a Bayesian framework. Crossa (2012) reviewed a number of frequentist models to assess GEI and used a Bayesian approach with a set of priors for variance parameters to model gene (in terms of markers) × environment interaction. In the context of exploring GEI pattern using an AMMI model, Crossa et al. (2011) applied a Bayesian method for the estimation of model

parameters and found its usefulness for unbalanced datasets with heterogeneous variances. Perez-Elizalde et al. (2012) successfully applied the Bayesian approach for modelling maize yield data using AMMI and considering Mises-Fisher distributed priors for the multiplicative part.

Recently, Josse et al. (2014) proposed a Bayesian treatment of linear-bilinear models to deal with the problem of over-parameterization by ignoring it at the prior level, but applying an appropriate processing at the posterior level. Couto et al. (2014) estimated stability and adaptability parameters in popcorn using a Bayesian approach for fitting Eberhart and Russell's (1966) model and implemented the method by using the 'MCMCregress' function available in the 'MCMCpack' package in the open source R software (R Core Team 2015). They found Bayesian techniques efficient for cultivar recommendation in more or less favorable environments, leading to high accuracy of the parameter estimates for cultivar evaluation. Credible intervals obtained by de Oliveira et al. (2014) for data from maize trials allowed them to identify genotypes and environments that did not contribute to GEI; thank to this, the authors constructed homogeneous subgroups of genotypes and environments for specific adaptation.

Modelling real data from multi-environment crop variety trials generally requires computing with large models. With availability of current level of computing power, the Bayesian approach is reasonably suited to handle complex statistical models without involving large-sample approximations like in the frequentist scenario. For general model fitting, a Bayesian approach can be implemented by using the Markov-Chain Monte Carlo (MCMC) procedure, which is highly flexible, works for an arbitrary number of random effects, and provides high accuracy in the estimates of parameters of interest. The necessary algorithms are available in the WinBUGS (Spiegelhalter et al. 2003) and R (R Core Team 2015) software. If various plausible priors are available, selection criteria such as the Bayesian Deviance Information Criterion (DIC) can be used to assess the degree of compromise between data and priors. These criteria are discussed in standard texts such as Gelman et al. (2004) or Carlin and Louis (2009).

In an ongoing crop improvement program, prior information on variance components is generally available. Using the Bayesian approach, this study aims to assess GEI from one multi-environment genotype experiment in sorghum (*Sorghum bicolor*), by incorporating prior information from a set of similar trials conducted in the past, in Sudanese environments. For both the Bayesian method and the traditional frequentist method, we will 1) estimate genotypic, GEI and error variances, 2) predict genotype means and rank genotypes, and 3) estimate parameters such as heritability, i.e., the proportion of phenotypic variation explained by the genotypes, and genetic advance due to selection, i.e., the extent to which breeding progress takes place in advancing the population through selection for the quantitative characters.

MATERIALS AND METHODS

EXPERIMENTAL DATA

Eighteen genotypes of sorghum were evaluated in a randomized complete block design (RCBD) with four replicates in six environments. These trials were conducted by the Cereal Research Program, Agricultural Research Corporation (ARC), Wad Medani, Sudan during the seasons 2009/10, 2010/11 and 2011/12 at each of the two contrasting locations North-Gedarif and South-Gedarif, which have a substantial difference in long-term rainfall, among others. Grain yield was recorded in kg ha^{-1}.

There are two possible approaches to the MET analysis: 1) locations and years are combined and treated as environments; 2) environments are partitioned into locations, years and their interaction. In the present study, as commonly done in many MET studies, we will

follow the first approach. This is because the number of years and locations is too small to analyze the structure of these environment components.

FREQUENTIST APPROACH

From the frequentist point of view, we modelled plot grain yield data collected from multiple environments in terms of environment effects, block effects within environments, genotype effects, and genotype-by-environment interactions:

$$Y_{ijk} = \mu + E_j + R_{kj} + G_i + GE_{ij} + e_{ijk} \quad (1)$$

where Y_{ijk} is yield from the plot under the i^{th} genotype in the k^{th} block of the j^{th} environment, μ is the general mean, E_j is the effect of the j^{th} environment, R_{kj} is the effect of the k^{th} block in environment j, G_i is the effect of the i^{th} genotype, GE_{ij} is the interaction between the i^{th} genotype and the j^{th} environment, and e_{ijk} is the random error from this plot. The environments were the combination of years and locations, and we assumed the effects of the six environments were fixed. Since we also aimed to estimate heritability and genetic gain due to selection, we also assumed the genotypic effects and G × E interaction as random (Kempthorne 1983). Overall, apart from μ, the following random-effects assumptions were made:

$R_{jk} \sim N(0, \sigma_R^2)$
$G_i \sim N(0, \sigma_G^2)$
$GE_{jk} \sim N(0, \sigma_{GE}^2)$
$e_{ijk} \sim N(0, \sigma_e^2)$,

where the symbol ~ means 'is distributed as' and $N(0, \sigma^2)$ is the normal distribution with mean 0 and variance σ^2 , where σ^2 stands for each of the above variance components. The indices $i = 1,\ldots,n_G(=18)$ refer to genotypes, $j = 1,\ldots,n_P(=6)$ refer to the environments, and $k = 1,\ldots,n_B(=4)$ refer to the replications. In METs, the variance components due to replications, genotypes, and errors may vary with the environment (Smith et al. 2005). Based on the analysis of environment-wise data under the present study and using the Bartlett's test for homogeneity of variances from the three environments, we found an acceptable level of homogeneity among the variances due to replications (P = 0.708) and genotypes (P = 0.035). However, the error variances differed significantly (P < 0.001) but were ignored to enable the use of their pooled variance in the standard expressions for the heritability and genetic gain due to selection (Kempthorne 1983). For a more general situation with variance-covariance structures, however, one may pursue the estimation of heritability on the lines of Cullis et al. (2006) and Piepho and Möhring (2007). To fit the mixed model in equation (1), we used the restricted maximum likelihood (REML) method based on the REML directive in Genstat software (Payne 2013).

BAYESIAN APPROACH

From a Bayesian perspective, model in equation (1) can be re-written as

$$Y_{ijk}|E_j, R_{jk}, G_i, GE_{ij}, \sigma_e^2 \sim N(\mu + E_j + R_{kj} + G_i + GE_{ij}, \sigma_e^2)$$

The variance components of the effects and interactions in equation (1) will be assumed to be random variables having distributions, called *a priori* distributions, with known parameters. Such *a priori* distributions have been studied and recommended for variance components or for the corresponding standard deviation components or scale parameters by Gelman (2006), who also commented on the limitations of gamma/inverse-gamma as the prior distribution for variance components. Based on the estimates of standard deviation components (SDC) obtained from a series of trials conducted in past, Singh et al. (2015) found the truncated normal distributions as satisfactory priors for those SDCs. The various *a priori* distributions of the scale parameters, $\sigma_R, \sigma_G, \sigma_{GE}$ and σ_e for the effects and errors of the model (1) were, therefore, taken from the families of positively truncated normal distributions.

Priors for the SDCs: In this study, priors were obtained by using data on sorghum yield (kg ha^{-1}) from three similar experiments conducted to evaluate 18 genotypes in RCBDs with four replications during 2006/07- 2008/09 at Rahab station in Sudan. The various components of variance were estimated using restricted maximum likelihood (REML) estimation, by taking data from 1) the first two –years and 2) all three years. Of the many choices for building prior information, the first two years data were the earliest to enable an estimation of GEI. The estimates of variance components along with their standard errors were obtained by using REML and associated directives in the Genstat software. The variance parameters of the SDCs were estimated by using the approximation for variance of square-root of a random variable, say X, as Var(X)/(4X). For each of the two priors sets, Table 1 gives the estimates of the variance components (σ^2, say) and values of the precision parameter (τ) of the SDC (σ), defined as the inverse of its variance. The *a priori* distribution for the SDC (σ), may be denoted as the positively-truncated-normal: $N(0, \tau^{-1})^+$. In the notation used in the WinBUGS software, this distribution will be written as "dnorm(0, τ)I(0,)". The precision (τ) of the SDCs based on the three-year data was much higher than that from the first-two-year data. The two priors thus taken were:

1) P_1: Priors set with $\sigma_R \sim N(0, \tau^{-1} = 0.000993^{-1})^+$, $\sigma_G \sim N(0, \tau^{-1} = 0.000199^{-1})^+$, $\sigma_{GE} \sim N(0, \tau^{-1} = 0.000731^{-1})^+$ and $\sigma_e \sim N(0, \tau^{-1} = 0.000473^{-1})^+$ (using data from 2006/7-2007/8 and the values of the variances are from Table 1), and
2) P_2: Priors set with $\sigma_R \sim N(0, \tau^{-1} = 0.00184^{-1})^+$, $\sigma_G \sim N(0, \tau^{-1} = 0.000687^{-1})^+$, $\sigma_{GE} \sim N(0, \tau^{-1} = 0.00142^{-1})^+$ and $\sigma_e \sim N(0, \tau^{-1} = 0.00760^{-1})^+$ (using data from all the years 2006/7-2008/9)

We selected a better of the two priors set with the deviance information criterion (DIC) criterion. In each of the above sets of priors, the prior for each parameter was assumed independent of the other priors.

Table 1. Estimates of variance components and precision of the standard deviation components from data on sorghum genotype yields from trials in the three environments during 2006/07 - 2008/09 at Rahab, Sudan.

		Based on years 2006/07-2007/08			Based on years 2006/07-2008/09		
Random terms	VC[a]	VC estimate	SE[b] (VC estimate)	Precision (τ)[c] of the SDC	VC[a] estimate	SE[b](VC estimate)	Precision (τ)[c] of the SDC
Replications within seasons	σ^2_R	959	1965	0.00099	1332	1702	0.00184
Genotypes	σ^2_G	2760	7446	0.00020	9288	7354	0.00069
Season × Genotype	σ^2_{GE}	17045	9659	0.00073	22207	7912	0.00142
Error	σ^2_e	43085	6033	0.00473	40274	4605	0.00760

[a] VC: Variance component estimate.
[b] SE: Standard error estimate. SDC: standard deviation component.
[c] Precision (τ) of the SDC =1/((SE (VC estimate))^2/(4 × VC estimate))

A step-wise approach will have the following main steps:
1) the dataset may be read using R-codes saved in a file;
2) create another text file with extension "bug" to contain codes specifying statements to model data, description of priors, and parameters of interest;

3) the file with R-codes may contain codes to specify data variables, initial values of random variables to be generated (one statement per chain, the list of parameters whose distributional summaries are to be printed, and a statement which calls the Bayesian analysis function ("bugs") with links to data, initial values, parameters, numbers of iterations, chains and simulations); 4) run those codes and fix any errors as they come.

The example files contents of the WinBUGS and R codes are given in Appendices A1-2. The number of iterations was set at one million, the number of chains was set at three, and the last 10,000 simulated values of the parameters were taken for evaluating the posterior distributions. These settings resulted into very small Monte Carlo error values of the estimated parameters, reflecting high accuracy of the estimates.

HERITABILITY AND GENETIC ADVANCE DUE TO SELECTION

Using the variance components from the METs, the broad-sense heritability on mean-basis (h^2) and genetic advance or gain due to selection of top 100q% lines (GA(q)%) are given as:

$$h^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_{GE}^2 / n_P + \sigma_e^2 / (n_P n_B)) \text{, and}$$

$$GA(q)\% = 100C(\sigma_G^2 / \bar{Y}) / (\sigma_G^2 + \sigma_{GE}^2 / n_P + \sigma_e^2 / (n_P n_B))^{1/2}$$

where $C = \frac{1}{q\sqrt{2\pi}} e^{-z_q^2/2}$ for $0 < q < 1$ and the truncation point z_q for the standard normal distribution is given by the equation $\int_{z_q}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - q$ and $\bar{Y}$ is the grand mean. For q = 0.20, C = 1.4 (Kempthorne 1983; Singh et al. 2012).

RESULTS

SELECTION OF PRIORS

Table 2 gives discrepancy statistics for the priors used in modelling the data. A smaller value of DIC reflects a better suitability of the chosen priors and model to the data (Gelman et al. 2004). The DIC value for prior set P_2 (2861.26) was lower than that for P_1. Therefore, P_2 was used to estimate the genetic parameters.

Table 2. Discrepancy statistics for selection of the priors for sorghum grain yield data from the trials in six environments (2009/10 -2011/12 at North-Gedarif and South-Gedarif), Sudan.

Priors[a]	Statistics[b]			
	$\bar{D}$	$\hat{D}$	p_D	DIC
P_1	5587.9	7966.48	-2378.58	3209.32
P_2	5588.42	8315.58	-2727.16	2861.26

[a] P_1: Priors set with $\sigma_R \sim N(0, \tau^{-1} = 0.000993^{-1})^+$, $\sigma_G \sim N(0, \tau^{-1} = 0.000199^{-1})^+$, $\sigma_{GE} \sim N(0, \tau^{-1} = 0.000731^{-1})^+$ and $\sigma_e \sim N(0, \tau^{-1} = 0.000473^{-1})^+$

P_2: Priors set with $\sigma_R \sim N(0, \tau^{-1} = 0.00184^{-1})^+$, $\sigma_G \sim N(0, \tau^{-1} = 0.000687^{-1})^+$, $\sigma_{GE} \sim N(0, \tau^{-1} = 0.00142^{-1})^+$ and $\sigma_e \sim N(0, \tau^{-1} = 0.00760^{-1})^+$

[b] $\bar{D}$ = posterior mean of (- 2 × log-likelihood). $\hat{D}$ = - 2 × log-likelihood at posterior means of parameters. p_D = effective number of parameters, DIC = Deviance information criterion.

ESTIMATES OF COMPONENTS OF VARIATION, HERITABILITY AND GENETIC ADVANCE

Table 3 shows the results. It was found that the posterior mean (i.e., the Bayesian estimates) was higher than the corresponding frequentist estimates for the genotypic variance and lower for the other two components (GEI interaction and the error variances). For all the estimates, the posterior standard deviations in the Bayesian approach were smaller than the corresponding standard errors in the frequentist approach. The ratios of

variance component estimates to the respective standard deviation or standard error estimates were higher for the Bayesian than for the frequentist approach. This showed that the Bayesian approach had higher power in detecting the variability due to genotypes and GEI. This increase in genotypic variability and reduction in GEI and error variances has led to a considerable increase (over three times) in heritability of the sorghum grain yield: 64.2 ± 8.3% (the Bayesian approach) vs. 20.3 ± 29.9% (the frequentist approach). The heritability estimate was smaller in the frequentist approach, with a very high standard error. The standard error is based on the large sample approximation of variance of the ratio of linear functions of variance components, i.e., the delta method (Harville and Fenech 1985), and may result in a negative lower limit. For the frequentist case of multi-environment trials in RCBDs, Singh et al. (1993) evaluated the probability of observing invalid estimates of heritability, which may provide guidance on selection of an appropriate experimental design for a given level of genotypic variability. The simulated distribution of heritability under the Bayesian approach provides a better protection against the invalid estimates. The estimated genetic advance by the Bayesian approach was over three times higher than that by the frequentist approach. The WinBUGS software provided standard deviation of the posterior mean of genetic advance. However, we could not find any approximation for the standard error of this index in its frequentist version.

Table 3. Bayesian posterior means and frequentist estimates of variance components, heritability and genetic advance for sorghum grain yield (kg ha^{-1}) from the trials in six environments (2009/10 -2011/12 at North-Gedarif and South-Gedarif), Sudan.

Parameters[a]	Bayesian approach (Priors set = P_2) [b]					Frequentist approach	
			95% Credible interval				
	Posterior mean	SD[c]	Median	Lower	Upper	Estimate	SE[d]
σ^2_G	4520	1654	4239	2146	8529	1169	2095
σ^2_{GE}	9621	1738	9477	6655	13330	21342	4249
σ^2_e	17820	1079	17770	15820	20020	24659	1994
h^2	0.642	0.083	0.646	0.470	0.793	0.203	0.299
GA(0.2)%	15.09	3.61	14.78	9.034	23.2	4.35	
CV%	26.9	0.81	26.9	25.4	28.5	31.7	
Mean (kg ha^{-1})	496	6.366				496	

[a] σ^2_G, σ^2_{GE} and σ^2_e are variance components due to genotypes, genotype × environment interaction and error respectively. h^2 = broad sense heritability on mean-basis. GA (.2)% = genetic advance at 20% selection.

[b] The prior distributions were $\sigma_R \sim N(0, \tau^{-1} = 0.00184)^+$, $\sigma_G \sim N(0, \tau^{-1} = 0.000687)^+$, $\sigma_{GE} \sim N(0, \tau^{-1} = 0.00142)^+$ and $\sigma_e \sim N(0, \tau^{-1} = 0.00760)^+$

[c] SD= posterior standard deviation.

[d] SE= standard error.

PREDICTED VALUES AND RANKS OF THE GENOTYPES ACROSS ENVIRONMENTS

Table 4 shows the posterior means of genotypes, 95% credible interval, and predicted means for the frequentist approach. Denoting the genotype numbers 1 to 18 as G1, G2 …G18, posterior means gave small differences of -1% (for G6) to 1.3% (for G18) over the frequentist estimate. The range of predicted means was almost the same for the Bayesian (380 – 600 kg ha^{-1}) and the frequentist approach (384 – 598 kg ha^{-1}). Genotypes G13 and G6 were found the highest- and lowest–yielding, respectively, under both approaches. The posterior means had much lower standard deviation (25.5 kg ha^{-1}) than that in the REML best linear unbiased predictor (BLUP) standard error estimates (30 kg ha^{-1}).

Using the Bayesian simulation, it was possible to obtain ranks of the genotypes for predicted means for each simulation run. Using the 10,000 values, R codes were used to compute the posterior means of the ranks of the genotypes and their credible intervals (Table 4). High-yielding lines will be those which are high-ranking genotypes most of the times. In this case one looks for high rank within say 95% credible interval. The rank of genotype G13 varied from 1 to 4 in this interval, while the rank of the worst-yielding genotype G6 varied from 15 to 18. The next two best lines under Bayesian approach were G15 and G10, whose ranks varied from 1 to 7 with 95% probability.

Table 4. Predicted values of the genotype means and their ranks, under Bayesian and frequentist approaches, for sorghum grain yields (kg ha^{-1}) from the trials in six environments (2009/10 -2011/12 at North-Gedarif and South-Gedarif), Sudan.

	BA[a]			FA[b]	BA[a]	FA[b]	BA[a]		
		95% credible interval						95% credible interval for ranks	
Genotypes	Posterior mean	Lower	Upper	Pred-mean	Rank	Rank	Mean	Lower	Upper
G1	398	347	449	401	17	17	16.5	14	18
G2	505	456	554	506	10	10	9.3	4	13
G3	439	388	489	441	14	14	14.3	11	17
G4	413	363	462	416	16	16	15.7	13	18
G5	464	414	514	465	13	13	12.8	9	16
G6	380	329	430	384	18	18	17.3	15	18
G7	500	452	549	500	12	12	9.8	4	14
G8	518	468	567	518	7	7	8.0	3	13
G9	537	488	587	536	4	4	6.1	2	11
G10	577	527	628	574	3	3	2.8	1	7
G11	518	469	566	517	8	8	8.0	3	13
G12	500	450	549	500	11	11	9.8	5	14
G13	600	550	649	598	1	1	1.7	1	4
G14	536	485	584	535	5	5	6.2	2	12
G15	577	526	628	575	2	2	2.8	1	7
G16	418	368	466	420	15	15	15.5	13	18
G17	517	468	566	517	9	9	8.1	3	13
G18	534	481	587	528	6	6	6.3	2	12
Av SD/ SE[c]	25.5			30					
Mean	496			496					

[a]BA= Bayesian approach. [b]FA= Frequentist approach. [c]Av SD/SE = Posterior standard deviation / average standard error.

DISCUSSION

Interest in Bayesian methods in plant breeding research has been regularly growing. This is mainly because information collected from past/ongoing crop improvement programs, particularly on variance components involving genotypes, may be used to improve the inference on the parameters of interest for the current program. The present study compared the analyses of multi-environment trials in sorghum in Sudanese conditions, as obtained with or without prior information. We considered priors based on the standard deviation components from three previous trials in sorghum. Since a subset of data containing at minimum two environments can also be used to obtain the priors, priors based on the full dataset (three environments) were compared with those derived from a smaller dataset (two environments), by using the DIC (the smaller DIC, the better). Not surprisingly, the DIC favored the priors based on the larger dataset. Some priors obtainable from general considerations (Gelman 2006) may also be screened as candidate priors, whenever they could be seen as reasonable and have been used by others in similar studies.

Evaluation of genetic advance is of primary interest in any crop improvement programs that lead to changes in the genotype base, through crossing and selection across time, inclusion of new or promising genotypes, and removal of the poorly performing ones. In this context, the assumption of random effects of genotypes becomes appropriate, and the interest lies in the prediction of the future performances of genotypes. It is not true that Bayesian estimates will always be relatively more precise than their frequentist versions. This is because various priors and various frequentist methods can be used, depending on the researcher's choice. However, this study found that the Bayesian results showed higher genotypic variance and lower GEI and error variances. These resulted in higher heritability and genetic advance with higher precisions than those in the frequentist approach. Such over three times higher values of heritability and genetic advance may reflect the power of the suitably chosen priors as well as the Bayesian approach itself. Plant breeders desire high heritability and genetic gains for an efficient breeding program. These high values themselves should not be used to prefer the Bayesian approach over the frequentist approach; but this study should encourage plant breeders to exploit the historical data to add value and enhance the breeding progress.

The Bayesian method can be very helpful in summarizing complex data analysis, like in METs. Due to its computational complexity, the Bayesian method needs to be implemented by using tools such as R2WinBUGS. In the end, simulated values for the parameters of interest describe their distributions well, and several summary statistics that either are less accurate or not available under traditional estimation methods may be accurately obtained. For example, standard errors of ratios of linear function of variance components or heritability estimated by using REML are approximate, and so are their confidence intervals. To the best of our knowledge, a frequentist approach does not enable one to analyze the distribution of ranks of the genotypes and approximation of standard errors of genetic advance. The broader statistical inference base of the Bayesian approach is facilitated by the available computational tools, for example, for computing the credible intervals for the rank of a genotype due to inbuilt simulations in R2WinBUGS software. The distributions of ranks of genotypes in the frequentist approach also can be obtained through an exclusive simulation. For example, in the case of MET data analysis aimed at comparing varieties for their stochastic dominance (Anderson 1974), Piepho and McCulloch (2004) presented a simulation approach based on using data transformation to evaluate the relative and absolute risks in wheat varieties.

There are situations in METs that will be pursued by using priors derived from already conducted trials as well as by using uninformative priors, such as, for example, the evaluation of stability indices (Lin et al. 1986). The genotypic response to the environment may be partitioned into the response to the components of the environment, such as

locations and years, which may require more complex statistical models than those presented in this paper. The Bayesian approach to model location and year components will be pursued on other datasets to fully assess its usefulness.

CONCLUSIONS

In this study, we have compared frequentist and Bayesian approaches for detecting GEI in sorghum yield trials conducted in RCBDs in Sudan. Prior information was obtained based on a different set of multi-environment trials in sorghum. The Bayesian analysis was done using R2WinBUG. The error variances, coefficient of variation, heritability, genetic advance, and predicted means of genotypes were estimated. The Bayesian posterior means (i.e., the estimates) for heritability and genetic advance were substantially higher than those of the frequentist approach, and the posterior standard deviations of those estimates were relatively lower. Incorporating the prior information from an ongoing crop improvement program has potential for adding value to the crop breeding program. The illustrative step-by-step Bayesian procedure presented here is recommended for use in statistical analysis of multi-environment trials.

ACKNOWLEDGEMENTS

The authors thank the reviewers for their suggestions and the language editors for presentation, which led to a substantial improvement of the earlier versions of the manuscript. The first author is grateful to ICARDA and Arab Fund for Economic and Social Development (AFESD) for granting a fellowship for carrying out the research study.

REFERENCES

Anderson J.R. (1974). Risk-efficiency in the interpretation of agricultural production research. *Review of Marketing and Agricultural Economics* 42, 131–184.

Carlin B.P., Louis T.A. (2009). *Bayesian Methods for Data Analysis*. Chapman & Hall, Boca Raton, FL.

Cochran W.G., Cox. G.M. (1957). *Experimental Design*. John Wiley, New York, NY.

Couto M.F., Nascimento M., do Amaral A.T., e Silva F.F., Viana A.P, Vivas M. (2015). Eberhart and Russell's Bayesian method in the selection of popcorn cultivars. *Crop Science* 55, 571–577.

Crossa J. (2012). From genotype × environment interaction to gene × environment interaction. *Current Genomics* 13(3), 225–244.

Crossa J., Perez-Elizalde S., Jarquin D., Cotes, J.M., Viele K., Liu G., Cornelius P.L. (2011). Bayesian estimation of the additive main effects and multiplicative interaction model. *Crop Science* 51, 1458–1469.

Cooper M., Hammer G.L. (1996). *Plant Adaptation & Crop Improvement*. CAB International Distribution: Oxford University Press.

Cullis, B.R., Smith, A., Coombes, N. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological and Environmental Statistics* 11, 381–393.

DeLacy I.H., Ratnasiri W.G.A, Mirzawan P.D.N. (1996). Retrospective analysis of historical data sets from multi-environment trials-case studies. In: Cooper M., Hammer G.L (Eds.) *Plant Adaptation and Crop Improvement*. CAB International: Wallingford, UK, 269–290.

de Oliveira L.A., da Silva C.P., Nuvunga J.J., Da Silva A.Q., Balestre M. (2014). Credible intervals for scores in the AMMI with random effects for genotype. *Crop Science* 55, 465–476

Eberhart S.A., Russell W.A. (1966). Stability parameters for comparing varieties. *Crop Science* 6, 36–40

Edwards J.W., Jannink J.L. (2006). Bayesian modeling of heterogeneous error and genotype and environment interaction variances. *Crop Science* 46, 820–833.

Forkman J., Piepho, H.P. (2013). Performance of empirical BLUP and Bayesian prediction in small randomized complete block experiments (With Supplemental Material*), Journal of Agricultural Science (Cambridge)* 151, 381–395.

Gauch H.G. (1988). Model selection and validation for yield trials with interaction. *Biometrics* 44, 705–715.

Gelman A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1, 515–533.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (2004). *Bayesian Data Analysis*, 2nd ed. Chapman and Hall Ltd, New York.

Josse J., van Eeuwijk F.A., Piepho H.-P., Denis J.B. (2014). Another look at Bayesian analysis of AMMI models for genotype-environment data. *Journal of Agricultural, Biological and Environmental Statistics* 19, 240–257.

Harville D.A., Fenech A.P. (1985). Confidence intervals for a variance ratio, or for heritability, in an unbalanced mixed linear model. *Biometrics* 41, 137–152

Hinkelmann K., Kempthorne O. (2005). *Design and Analysis of Experiments*, *vol. 2: Advanced Experimental Design*, Wiley-Interscience, New York, NY.

Kempton R.A. (1984) The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science*, 103, 123–135

Kempthorne O. (1983). *The Design and Analysis of Experiments*. R. E. Krieger Publishing Co., Malabar, Florida.

Lin C.S., Binns M.R, Lefkovitch L.P. (1986). Stability analysis: where do we stand. *Crop Science* 26, 894–900.

Patterson H.D., Williams E.R. (1976). A new class of resolvable incomplete block designs. *Biometrika* 63, 83–92.

Payne R.W. (ed.) (2013). *The Guide to the GenStat® Command Language (Release 16). Part 2: Statistics*. VSN International, Hemel Hempstead, Herfordshire, UK.

Perez-Elizalde S., Jarquin D., Crossa J. (2012). A general Bayesian estimation method of linear–bilinear models applied to plant breeding trials with genotype × environment interaction. *Journal of Agricultural, Biological, and Environmental Statistics* 17, 15–37.

Piepho, H.P., McCulloch, C.E. (2004): Transformations in mixed models: application to risk analysis for a multienvironment trial. *Journal of Agricultural, Biological and Environmental Statistics* 9, 123–137.

Piepho, H.bP., Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177, 1881–1888.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-070. http://www.R-project.org.

Sarker A., Singh M., Rajaram S., Erskine W. (2010). Adaptation of small-seeded red lentil (*Lens culinaris* Medik. subsp. *culinaris*) to diverse environments. *Crop Science* 50, 1250–1259.

Singh M., Al-Yassin A., Omer S. (2015). Bayesian estimation of genotypes means, precision and genetic gain due to selection from routinely used barley trials. *Crop Science* 55, 501–513

Singh M., Ceccarelli S., Hamblin J. (1993). Estimation of heritability from varietal trials data. *Theoretical and Applied Genetics* 86, 437–441.

Singh M., van Ginkel M., Sarker A., Malhotra R.S., Imtiaz M., Kumar S. (2012). Increasing precision of even otherwise well-run trials by capturing heterogeneity of plot error variances. *Agricultural Research* 1, 285–294.

Smith A.B., Cullis B.R., Thompson R. (2005). The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *Journal of Agricultural Science* 143, 449–462.

Spiegelhalter D., Thomas A., Best N., Lunn D. (2003). *WinBUGS Version 1.4 User Manual*. Medical Research Council Biostatistics Unit, Cambridge, UK, 2003. (Available at http://www.mrc-bsu.cam.ac.uk/bugs).

Theobald C.M., Talbot M., Nabugoomu F. (2002). A Bayesian approach to regional and local-area prediction from crop variety trials. *Journal of Agricultural Biological and Environmental Statistics* 7, 403–419.

Yan W., Hunt L.A, Sheng Q., Szlavnic Z. (2000). Cultivar evaluation and mega-environment investigation based on GGE biplot. *Crop Science* 40, 596–605.

Yan W. (2011). GGE Biplot vs. AMMI Graphs for genotype-by-environment data analysis. *Journal of Indian Society of Agricultural Statistics* 65, 181–193.

APPENDICES

A.1: WinBUGS codes to model data from RCBD and estimation of means predicted, GEI heritability and genetic gain

```
#Bayesian analysis of GEI. This file GEI.bug comprises the following codes
model{
 # data model
  for (i in 1 :N){  y[i] ~ dnorm(mu[i], tau.e)
                   mu[i]<-  m    +  b[blk[i],env[i]]  +  p[env[i]]  +  g[geno[i]]+
a[geno[i],env[i]]
            }
# m
m ~ dnorm(0.0, 1.0E-6)
# Block
 for (j in 1: NP){for(k in 1: (NB-1)){ b[k,j]~ dnorm(0.0, tau.b)  }
                                      b[NB,j]<-  -sum(b[1:(NB-1),j])
                 }
# Genotyptes
    for (i in 1: (NG-1)){ g[i] ~ dnorm(0.0, tau.g)   }
                        g[NG]<-  -sum(g[1:(NG-1)])
# Envirovments effects as fixed
   for (j in 1: (NP-1)){  p[j] ~ dnorm(0.0, 1.0E-6)     }
                      p[NP]<-  -sum(p[1:(NP-1)])
#GEI
for (i in 1: (NG-1)){  for (j in 1: (NP-1)){ a[i,j] ~ dnorm(0.0 , tau.a)  }  }
                  for (j in 1: (NP-1)){ a[NG,j] <-  - sum(a[1:(NG-1),j]) }
                  for (i in 1: (NG-1)){ a[i,NP] <-  - sum(a[i, 1:(NP-1)]) }
                                       a[NG,NP] <- - sum(a[NG, 1:(NP-1)])
#priors
                   sig.e ~ dnorm(0, tauE)
                   sig.g ~ dnorm(0, tauG)I(0,)
                   sig.b ~ dnorm(0, tauB)I(0,)
                   sig.a ~ dnorm(0, tauGE)I(0,)
                  tau.e <- 1/(sig.e*sig.e)
                  tau.b <- 1/(sig.b*sig.b)
                  tau.g <- 1/(sig.g*sig.g)
                  tau.a <- 1/(sig.a*sig.a)
# parameters of interest....more
                    sig2g <- (sig.g*sig.g)
                    sig2e <- (sig.e*sig.e)
                    sig2a <- (sig.a*sig.a)
 # Prediction of parameters of interest-- means, heritability, SEs
               for ( i in 1: NG) {PredG[i]<-  m + g[i]}
               for ( j in 1: NP) {PredE[j]<-  m + p[j]}
               for ( i in 1: NG) {
               for ( j in 1: NP) {
                 PredGE[i,j]<-  m + g[i]+p[j]+a[i,j]  }
                                                      }
 # Broad-sense heritability on  mean-basis, genetic advance and CV
                         h2<- sig2g/(sig2g+sig2a/NP+sig2e/(NB*NP))
                     GA20<- 100*1.4*sig2g/mn/sqrt(sig2g+sig2a/NP+sig2e/NB/NP)
                    CVpc <- 100*sqrt(sig2e)/mn
        }
# end of BUGS codes
```

A.2: R- codes for reading Dataset-RCBD OF GEI and calling the 'bugs' function

```
#load packs
library(lattice)
library(coda)
library(R2WinBUGS)
#data from comb.....................................
ndata<- read.table("stadata.txt", header=TRUE)
y<- ndata$GY ; blk<- ndata$Rep
env<- ndata$Envi ; geno<- ndata$Geno
mn<- mean(y)
NB<- 4 ; NP<- 6 ; NG<- 18
N<- NB*NG*NP ; NPG<- NP*NG ; NBP<- NB*NP
#print(cbind(y,blk,env,geno))
print(cbind(mn, NB, NP, NG, NPG, NBP, N))
# Assign tau's from the REML analysis of datasets for the priors
# Envt fixed (three years data 2006/7- 8/9)
 tauE<- c(0.00760); tauG<- c(0.000687) ; tauB<- c(0.00184) ; tauGE<- c(0.00142)
# Envt fixed (only first two seasons)
# tauE<- c(0.00473); tauG<- c(0.000199) ; tauB<- c(0.000993) ; tauGE<- c(0.000731)
#-------------------------------------------------------------------------
data<-  list("y","mn","blk","env","geno",  "NB","NP","NG","N",     "tauE",  "tauG",
"tauB", "tauGE")
data
#Envt p[] fixed
inits1<-    list(m=mn,  b=c(rep(.1,NBP)),  g=c(rep(.21,  NG)),  p=c(rep(.21,  NP)),
a=c(rep(.2, NPG)), sig.e=.5, sig.b=1, sig.g=0.01, sig.a=1.1)
inits2<-    list(m=mn,  b=c(rep(.1,NBP)),  g=c(rep(.22,  NG)),  p=c(rep(.21,  NP)),
a=c(rep(.2, NPG)), sig.e=.5, sig.b=1, sig.g=0.01, sig.a=1.1)
inits3<-    list(m=mn,  b=c(rep(.1,NBP)),  g=c(rep(.2,  NG)),  p=c(rep(.20,  NP)),
a=c(rep(.2, NPG)), sig.e=.5, sig.b=1, sig.g=0.01,  sig.a=1.1)
inits <- list(inits1, inits2, inits3)
inits
parameters  <-  c("m",  "PredG","PredE",  "tau.e",  "tau.b",  "tau.g",  "tau.a",
"sig2g","sig2e","sig2a", "h2", "CVpc", "GA20")
parameters
rcbGE.sim <- bugs(data, inits, parameters, "GEI.bug", n.chains=3, n.iter=1000000,
n.sims=10000,  debug=TRUE)
# more codes for calculating ranks and their distributions are available on request
from the corresponding author
```