Conference on Applied Statistics in Agriculture 2016 - 28th Annual Conference Proceedings

BAYESIAN ESTIMATION OF STABILITY INDICES OF SORGHUM VARIETY TRIALS

Siraj Osman Omer
Agricultural research Corporation, sirajstat@yahoo.com

Abdel Wahab Hassan Abdalla
Department of Agronomy, Faculty of Agriculture, University of Khartoum, Sudan, abdelwahab.hassan@yahoo.com

Mohammed Hamza Mohammed
Cereal Research Center, Agricultural Research Corporation (ARC), P.O. Box 126, Wad, Sudan, elgadahamza@yahoo.com

International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan

See next page for additional authors

Recommended Citation

Omer, Siraj Osman; Abdalla, Abdel Wahab Hassan; Mohammed, Mohammed Hamza; and International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan (2016). "BAYESIAN ESTIMATION OF STABILITY INDICES OF SORGHUM VARIETY TRIALS," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1483

Author Information

Siraj Osman Omer; Abdel Wahab Hassan Abdalla; Mohammed Hamza Mohammed; and International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan

BAYESIAN ESTIMATION OF STABILITY INDICES OF SORGHUM VARIETY TRIALS

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

[1.] Experimental Design and Analysis Unit, Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan, Tell: + 249908246491, E-mail: sirajstat@yahoo.com

2. Department of Agronomy, Faculty of Agriculture, University of Khartoum, Sudan

3. Cereal Research Center, Agricultural Research Corporation (ARC), P.O. Box 126, Wad, Sudan

4. International Center for Agricultural Research in the Dry Areas (ICARDA), Amman, Jordan

Abstract: Multiple–environmental trials are routinely conducted by crop improvement programs for developing desired genotypes. Over a long run, these programs gather information on genotypic performance and variability. Bayesian approach can be used to utilize prior information to identify genotypes for high and stable yield. A set of 18 sorghum genotypes were evaluated in randomized complete block designs (RCBD) with four replications during three seasons, 2009-2012 at diverse locations, North-Gedarif and South-Gedarif, in Sudan. Data on grain yield was analyzed. The aim of this paper was to estimate stability indices such as regression coefficient, coefficient of variation (CV %) and coefficient of determination (R^2) using a Bayesian approach. R2WinBUGS and R packages have been used. The results of these different stability indices agreements and suggesting that this approach produces reliable estimates of the stability of crop variety. In general, Bayesian compared to frequentist approach gave higher precision in terms of standard error of genotypes means, regression coefficient and coefficient of determination. Moreover, Bayesian has a broader inference-base to allow an integration of prior information about the current data and is recommended for use following the steps illustrated with the example datasets.

Keywords: Stability Analysis, Genotypes by Environment Interaction (GEI), Bayesian Method, R2WinBUG.

1. Introduction

A wide range of methods is available for the analysis of genotypes by environment interaction (GEI) and can be broadly classified into four groups: the analysis of components of variance, stability analysis, multivariate methods and qualitative methods (Motamedi *et al*., 2013). For a long time, the term stability has been characterized as genotype which gave consistently high yield (Yan *et al.,* 2006). Often stability implies that a stable variety has a low variance across environments (Glymour, 2011). This idea of stability is in agreement with the concept of homeostasis widely used in genetics and was called a 'biological concept' of stability (Bhatia, 2012). Stability analysis, based on individual indices, led many plant breeders to wonder which stability statistics should be used for their particular problem (Mut et al., 2010). Stability indices are generally estimated using frequentist approach and little work has been carried out for their estimation in context of the Bayesian framework (Josse *et al.,* 2014).

Burgueño (2012) investigated a Bayesian analysis of linear–bilinear models that had the advantage of incorporating prior information. His approach yielded shrinkage estimates of the eigenvalues on the bilinear effects of GEI using additive main effect and multiplicative interaction (AMMI) model. While Bayesian computational methods of the Markov Chain Monte Carlo (MCMC) variety (Gilks *et al*., 1996; Gelman *et al*., 2003; Robert and Casella, 2004) led to an explosion of Bayesian methods in the 1990s, linear-bilinear models produced unique problems due to the orthogonal bases used in singular value decomposition. Oral et al. (2012) studied Bayesian estimation using stable distributions of different priors examined through

simulations, where the mean square error of Bayesian estimates was found smaller than the mean square error from the other methods considered (Berger, 2006). This paper reviews selected stability analysis concepts and illustrates them on sorghum variety frequentist and Bayesian approaches. The Bayesian approach in this study considers a range of priors for standard deviation and components from the uniform, positive half-t, positive half-normal and inverse-gamma distributions. The most coherent prior was used to for assessing the stable genotypes in terms of responsiveness (regression coefficient) to the environment, coefficient of variation, coefficient of determination and mean square error. The priors for standard deviation components were screened from the class of recommended priors (Gelman, 2006).

2. Concepts of Stability

The term stability is used to characterize a genotype, which shows a relatively constant yield, independent of the changing environmental conditions (Akinwale et al., 2011). There are various concepts of adaptability and stability designed to evaluate a group of genotypes tested in a series of environments (Lin *et al.*, 1986; DeLacy *et al.*, 1996). Among these, the most widely used are the ones based on linear regression of a genotype's mean response in individual environments on the environment mean (Finlay and Wilkinson, 1963; Eberhart and Russel, 1966 and Perkins and Jinks, 1968). See also, Verma et al. 1978, and Cruz et al. 1989. Francis and Kannenberg (1978) used the coefficient of variation (CV) of the genotype as a measure of its stability where a genotype with an average high yield and low CV was considered stable. There are different concepts of stability: static (biological) and dynamic (agronomic). With the static concept, a stable genotype possesses an unchanged performance regardless of variation in the environmental conditions (Lin *et al*. 1986). Lin *et al*. (1986) identified three concepts of stability.

Type 1 stability is based on among-environment variance and can be measured by variance across environments (S^2) and coefficient of variation (CV) for each genotype. Type 2 stabile genotype has its response to environments parallel to the mean response of all genotypes in the trial. A regression coefficient (b) can be used to measure this type of stability. A genotype is stable in the sense of Type 3 if the residual mean square representing deviation from the linear model on the environmental index is small. All stability procedures based on quantifying GEI effects belong to the dynamic stability concept. Parametric stability analysis provides a general summary of the response patterns of genotypes to environmental change (Becker and Leon, 1988).

3 Materials and Methods

3.1 Experimental Data Set

A series of trials were conducted in randomized complete block design (RCBD) to evaluate 18 genotypes of sorghum during the rainy seasons of 2009-2010, 2010-2011 and 2011-2012 at each of the two contrasting locations, North-Gedarif and South-Gedarif, Sudan. As a result, there were a total of six environments (location - year combinations). For each of the trials the design used was RCBD with four replications. The response variable analyzed was grain yield (kg/ha). The stability analysis using frequentist approach was carried out using a MTLV program coded in Genstat software (Dr. Murari Singh, ICARDA, 2013, reassured communication).

3.2 Regression Methods of Stability Analysis

3.2.1 Finlay and Wilkinson Stability Index as Regression Slope

Finlay and Wilkinson (1963) introduced a stability index as slope of linear regression of yield of a genotype on the environment index using the model:

$$Y_{ij} = \mu + b_i I_j + \delta_{ij} \quad (1)$$

Where Y_{ij} is the mean of the i^{th} genotype in the j^{th} environment (i = 1, 2, ..., NG; j = 1, 2, ..., NP), μ is the mean of all the genotypes over all environments, b_i is the regression coefficient, that measures the response of the i^{th} genotype to the varying environments, δ_{ij} is the deviation from regression of the i^{th} genotype at the j^{th} environment, I_j is the environment index obtained as the mean of all genotypes at the j^{th} environment minus the overall mean. I_j is computed as

$$\mathrm{I_j} = \sum_i \frac{y_{ij}}{NG} - \sum_i \sum_j \frac{y_{ij}}{NG*NP} \quad (2)$$

Note that $\sum_j I_j = 0$

where NG and NP are number of genotypes and environments respectively.

3.2.2 Eberhart and Russell Method (1966): Deviation Mean Square ($\sigma_{d_i}^2$)

This parameter of stability is calculated as mean square deviations (S_d^2) from linear regression

$$S_d^2 = \frac{\sum_j \delta_{ij}^2}{(NP-2)} - \frac{S_e^2}{NB} \quad (3)$$

where $\sum_j \delta_{ij}^2 = \left[\sum_j y_{ij}^2 - \frac{y_{i.}^2}{NP}\right] - \frac{\left(\sum_j y_{ijI_j}\right)^2}{\sum_j I_j^2}$ and S_e^2 is the pooled error mean square, and NB is number of replications or complete blocks.

3.3 Bayesian Approach for Stability

In frequentist approach, REML method is widely used in breeding programs data analysis particularly for estimation of variance components (Cotes *et al.*, 2006). In an ongoing breeding

program, prior information on components of variance for genotypes, genotypes x environment interaction and experimental error are normally available. Bayesian analysis can use such prior information and may prove advantages over the frequentist method (Congdon, 2003). However, in the present sudy, a range of prior distributions of standard deviation components and variance components was considered with assumed fair values for their parameters. Deviance information criterion (DIC), an often used discrepancy statistic, was used to select the best prior out of the set of five priors considered. An R package and WinBUGS software were used to perform the Bayesian stability analysis in terms of the three phenotypic stability parameters: regression coefficient (b), coefficient of determination (R^2) and coefficient of variation (CV) for each genotype from the data under model (1).

3.4 Prior Distribution and Posterior Density of the Parameters

In Bayesian estimation, posterior density of the parameters of interest is based on fitting the model for observation Y_{ijk}in the light of priors

$$Y_{ijk} = \mu + R_{ik} + G_i + E_j + GE_{ij} + e_{ijk} \quad (4)$$

where the index i is for genotypes, j for environment and k for the replication with environment, i=1,…, NG, j=1,…, NP and k=1,…,NB. The symbols y, μ, R, G, E, GE and E stand for response, general mean, replication effect within environment, genotypes effect, environment effect, GE interaction and plot error respectively, for the associated subscript. Equation (4) can be written in matrix notation as

$$\underline{y} = \mu\underline{J} + \underline{\mathrm{H}_1\,\rho} + \underline{\mathrm{H}_2\,\gamma} + \underline{\mathrm{H}_3\,\xi} + \underline{\mathrm{H}_4\,\varphi} + \underline{\varepsilon} \quad (5)$$

where vector $\underline{y}$ includes all $Y_{ijk}'s$, $\underline{J}$ is a vector of one's, $\underline{\rho}$ for $R_{ik}'s$, $\underline{\gamma}$ for G_i, $\underline{\xi}$ for E_j, $\underline{\varphi}$ for GE_{ij} and $\underline{\varepsilon}$ for e_{ijK} and $\underline{\mathrm{H}_1}$, $\underline{\mathrm{H}_2}$, $\underline{\mathrm{H}_3}$ and $\underline{\mathrm{H}_4}$ are incidence matrices associated with these effects vectors. Various effects are assumed to be independently and normally distributed as follows

$R_{ij}\sim N(0,\sigma_R^2)$, $G_i\sim N(0,\sigma_G^2)$, $GE_{ij}\sim N(0,\sigma_{GE}^2)$ and $e_{ijk}\sim N(0,\sigma_e^2)$. Effects E_j s have been assumed fixed. The variance covariance of vector $\underline{y}$ is

$$D(\underline{y}) = \underline{\mathrm{H}_1}\,\underline{\mathrm{H}_1'}\sigma_R^2 + \underline{\mathrm{H}_2}\,\underline{\mathrm{H}_2'}\sigma_G^2 + \underline{\mathrm{H}_4}\,\underline{\mathrm{H}_4'}\sigma_{GE}^2 + \underline{\mathrm{I}}\sigma_e^2$$

where $\underline{\mathrm{I}}$ is identity matrix of appropriate order.

Priors will be considered for variance components $\sigma_R^2, \sigma_G^2, \sigma_{GE}^2$ and σ_e^2 or associated standard deviation components. Writing the prior density for $\omega = \sigma_R^2, \sigma_G^2, \sigma_{GE}^2, \sigma_e^2$ as $f_\omega(\omega|\theta_{\omega_0})$, we can write the joint posterior density of $\mu, \underline{\rho}, \underline{\gamma}, \underline{\xi}, \underline{\varphi}, \sigma_R^2, \sigma_G^2, \sigma_{GE}^2$ and σ_e^2 as

$$p(\mu, \underline{\rho}, \underline{\gamma}, \underline{\xi}, \underline{\varphi}, \sigma_R^2, \sigma_G^2, \sigma_{GE}^2, \ \sigma_e^2 \ |\underline{y}\ ) \propto$$

$$N\left(\underline{y} \mid \mu\,\underline{J} + \underline{\mathrm{H}_3}\,\underline{\xi}\ ,\ \underline{\mathrm{H}_1}\,\underline{\mathrm{H}_1'}\sigma_R^2 + \underline{\mathrm{H}_2}\,\underline{\mathrm{H}_2'}\sigma_G^2 + \underline{\mathrm{H}_4}\,\underline{\mathrm{H}_4'}\sigma_{GE}^2 + \underline{\mathrm{I}}\,\sigma_e^2\right)$$

$$\times \left(N(\underline{\rho}|\underline{0}_\rho,\ \underline{\mathrm{I}}_\rho\ \sigma_R^2)\ f_{\sigma_R^2}(\sigma_R^2|\ \theta_{0\sigma_R^2})\right)$$

$$\times \left(N(\gamma|0_\gamma, \underline{\mathrm{I}}_\gamma\ \sigma_G^2) f_{\sigma_G^2}(\sigma_G^2\ |\theta_{0\sigma_G^2})\ \right)$$

$$\times \left(N\left(\varphi\middle|\underline{O}_\varphi, \underline{\mathrm{I}}_\varphi\sigma_{GE}^2\right) f_{\sigma_{GE}^2}\left(\sigma_{GE}^2\middle|\theta_{0\sigma_{GE}^2}\right)\right)$$

$$\times f_{\sigma_e^2}(\sigma_e^2|\theta_{0\sigma_e^2}).$$

where $\underline{0}$ is a vector of zeros of appropriate size and θ_{w0} stands for a known value (Cotes*et al.*, 2006).

To be specific, we chose the following classes of distribution for priors:

1) P_1: the priors for the standard deviation components $\sigma_R, \sigma_G, \sigma_{GE}$ and σ_e each follow Uniform (0, 1000).
2) P_2: the priors for the standard deviation components $\sigma_R, \sigma_G, \sigma_{GE}$ and σ_e each follow Half-t distribution $dt(0, c, v)I(0,)$ = Half-t (0, 4, 3). Here, c is non-centrality parameter and υ is the degree of freedom of the t-distribution. The values of c and υ are set at 4 and 3 respectively. I(0,) stands for the positive part of distribution.
3) P_3: the priors for the standard deviation components $\sigma_R, \sigma_G, \sigma_{GE}$ and σ_e each follow Half-normal distribution $N(0, \tau^{-1} = 10000) * I(0,)$ and $N(0, \tau^{-1}, 100) * I(0)$, = Half-normal (0, 0.001)I(0,). Here, τ is precision parameter, $\tau = \sigma^{-2}$, inverse of the variance.
4) P_4: the priors for the variance components were taken as inverse- Gamma distributions. Thus, the inverse (τ) of the variance components, $\sigma_R^2, \sigma_G^2, \sigma_{GE}^2$ and σ_e^2 each were assumed to follow Gamma distribution ~ dgamma (alpha=0.05, beta=0.05) (Gelman et al., 1995).

We believe that the above values for the parameters of the prior distributions in P_1 to P_4 fairly cover the respective parameter spaces suited for the experiments analyzed. The marginal posterior density of a chosen parameter is obtained by integrating the joint a posterior over the remaining parameters (Gelman, 2006). This integration is numerically done by MCMC methods run with R2WinBUGS, where, for the present study, the number of iterations was set at 50000, number of chains at three and simulation runs to 5000 for reporting the posterior distributions and means. These settings resulted into acceptably small Monte Carlo error values in the estimated parameters, reflecting high accuracy of the estimates. The R2WinBUGS and R codes, given in Appendix-A.1 and Appendix-A.2, were used for the computations.

4 Results

4.1 Bayesian Approach - Selection of Best Priors

The choices of priors for Bayesian analysis were made using the often reported discrepancy statistics: $\bar{D}$ = posterior mean of (- 2 × log-likelihood), $\hat{D}$ = - 2 × log-likelihood at posterior means of parameters, p_D = effective number of parameters and DIC = deviance information criterion (Table 1). The values of the DIC and P_D varied with the priors sets P_1, P_2 , P_3 and P_4. However, the prior set P_3 has the lowest value of DIC (2400.12). We took P_3 for further estimation of the predicted means and the three stability indices.

Table 1. Discrepancy statistics values for selection of the priors for data on sorghum hybrids over six environments for grain yield.

Priors model	$\bar{D}$	$\hat{D}$	p_D	DIC
P_1	5606.82	5517.48	89.333	5696.18
P_2	5610	5524.04	86.072	5696.15
P_3	5638.26	8876.4	-3238.14	2400.12
P_4	5597.97	5471.73	126.246	5724.22

Footnotes: Priors set are:

P_1: σ_R , σ_G , σ_{GE} and σ_e independently $\sim$ uniform$(0, 1000)$.

P_2: σ_R , σ_G , σ_{GE} and σ_e independently $\sim$ positive half $-$ t$(0, 4, 3)$.

P_3: σ_R , σ_G , σ_{GE} and σ_e independently $\sim$ positive half $-$ normal $(0, 0.01)$.

P_4: $\sigma_R^{-2}, \sigma_G^{-2}, \sigma_{GE}^{-2}$ and σ_e^{-2} independently ~ Gamma $(0.05, 05)$.

4.2 Bayesian Approach -Stability Analysis and Evaluation of Genotypes Means

Table 2 gives predicted means of each genotype under frequentist and Bayesian approaches. For Bayesian approach, it gives posterior means, standard deviations (SD) and 2.50% percentile, median (50% percentile) and 97.50% percentile of the genotypes mean: G10, G13 and G15 were three high yielding under both frequentist and Bayesian approaches. The average estimated standard deviation (SD) predicted means for Bayesian approach was lower than the standard error (SE) for the frequentist approach. Monte Carlo errors were acceptably small.

Table 2. Stability analysis under frequentists and Bayesian approach of means parameters for sorghum for grain yield. Entries were evaluated for three cropping years (2009-2012) at two locations in Sudan.

Genotypes	Frequentist approach	Bayesian approach (priors model : P_3)					
	Mean	Mean	SD	MC error	Percentile		
					2.50%	50%	97.50%
G1	401.1	416	24.2	0.35	367.7	416.2	462.9
G2	505.8	503.2	23.21	0.34	457.7	503.1	548.3
G3	440.8	449.7	23.16	0.31	403.6	450.1	494
G4	416.2	428.6	23.65	0.35	383.4	428.6	475.3
G5	465	469.3	23.02	0.29	423.5	469.5	513.8
G6	383.6	401.9	24.19	0.33	354.2	401.7	449.1
G7	499.6	498.2	23.32	0.33	452.7	498.1	543.4
G8	517.8	514.1	23.5	0.36	468.3	514	560.1

G9	536.1	528.4	22.94	0.32	483.2	528.5	573.3
G10	574.1	560.2	23.22	0.29	516.2	559.9	605.9
G11	517.4	512.8	22.44	0.33	469.7	512.9	556.8
G12	500.1	498.9	22.72	0.33	455.2	498.7	543.2
G13	597.5	579.8	24.15	0.36	533.3	579.9	626.2
G14	535.2	528.1	23.19	0.33	482.3	528	573.6
G15	574.8	560.7	23.33	0.34	516.1	560.4	608.1
G16	420.3	432.5	23.56	0.34	386.1	433.1	477.1
G17	517	513.1	22.87	0.32	468.4	513.2	557.4
G18	527.5	533.2	26.7	0.41	481.4	533	585.1
AvSE	±30.4		±23.5				

AvSE=average standard error (SE) or SD, MC=Monte Carlo

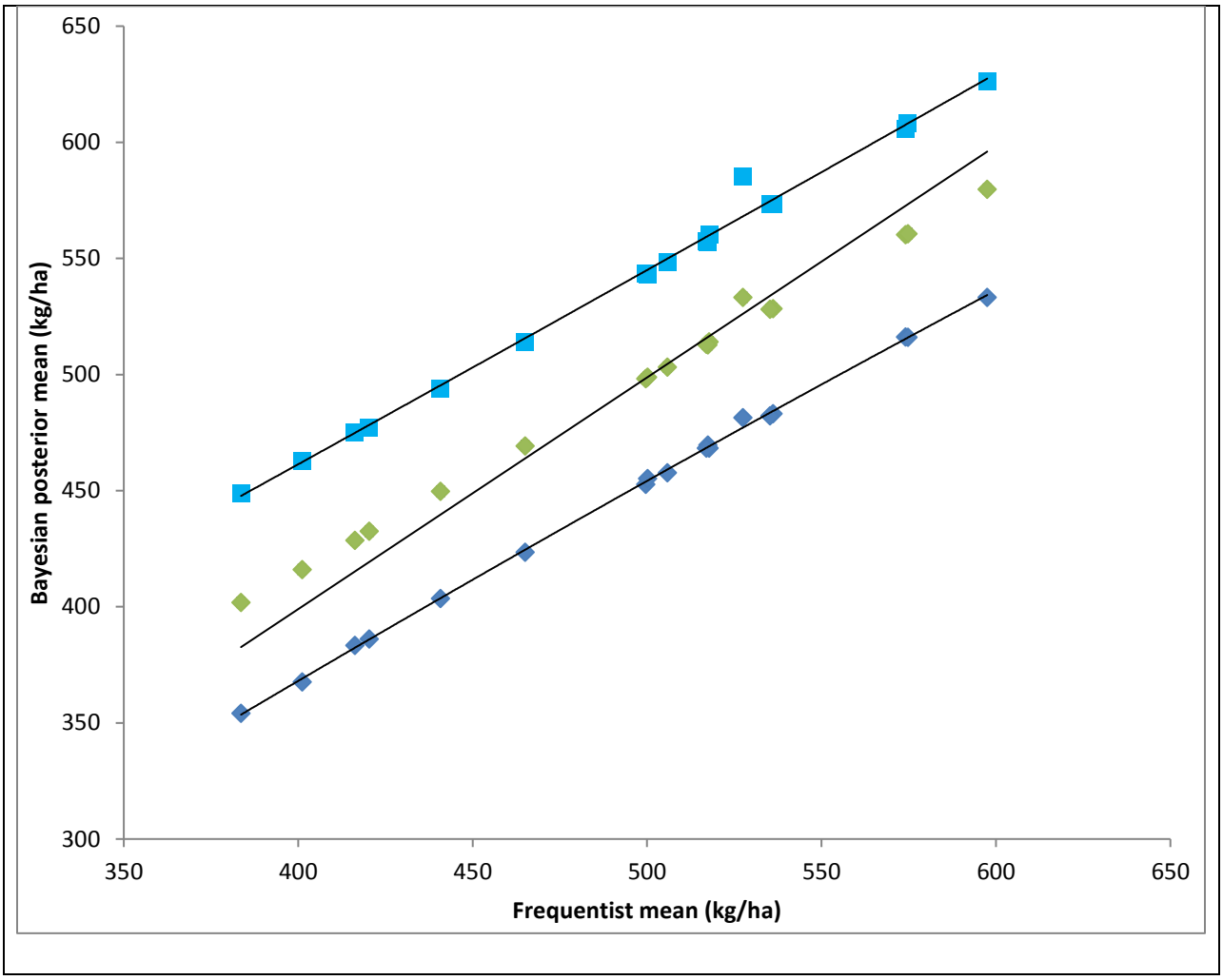

Figure 1. Scatterplot of predicted means of grain yield (kg/ha) of the 18 sorghum genotypes across the six environments during 2009-2012, Sudan, for comparing frequentist mean, Bayesian posterior mean and Bayesian credible bands using 95% credible interval values for individual genotypes.

The Figure 1 indicates comparison of Bayesian and frequentist estimates of predicted mean of each genotypes, The correlation between frequentist and Bayesian approaches was highly significantly positive (0.998, $P<0.001$). Bayesian estimates showed shrinkage effect for value higher than the mean under frequentist approach, and were enhanced for the genotypes below the mean under frequentist approach.

4.3 Stability Analysis Evaluation Using Regression Coefficient

Table 3 gives estimates of regression coefficient (slope) for each genotype under frequentist and Bayesian approaches. For Bayesian approach, it also gives its standard deviations, MC error and percentiles. Average standard error of regression coefficients under Bayesian approach is lower than that under the frequentist approach.

Table 3. Stability analysis under frequentist and Bayesian approaches of regression coefficient (b_i) parameters for sorghum for grain yield. Entries were evaluated for three cropping years (2009-2012) at two locations in Sudan.

Genotypes	Frequentist approach	Bayesian approach						
	bi	b_i	SD(bi)	P(bi)	MC error	Percentile		
						2.50%	50%	97.50%
G1	0.729	0.751	0.067	0.0099	0.00095	0.621	0.751	0.884
G2	0.989	0.998	0.065	0.9765	0.00084	0.873	0.998	1.125
G3	0.807	0.804	0.066	0.0250	0.00092	0.677	0.804	0.936
G4	0.769	0.760	0.065	0.0102	0.00092	0.633	0.761	0.887
G5	1.019	1.024	0.065	0.7246	0.00086	0.897	1.024	1.152
G6	0.945	1.014	0.067	0.8414	0.00089	0.883	1.014	1.145
G7	1.112	1.047	0.064	0.4904	0.00087	0.922	1.046	1.172
G8	1.058	1.036	0.064	0.5942	0.00080	0.912	1.036	1.162
G9	1.095	1.082	0.066	0.2604	0.00100	0.954	1.082	1.211
G10	0.952	0.975	0.065	0.7138	0.00089	0.847	0.975	1.105
G11	0.851	0.838	0.066	0.0495	0.00097	0.709	0.840	0.964
G12	1.065	1.089	0.066	0.2262	0.00098	0.959	1.088	1.219
G13	0.955	0.935	0.067	0.3694	0.00090	0.803	0.935	1.069
G14	1.362	1.369	0.066	0.0014	0.00101	1.238	1.370	1.499
G15	1.461	1.443	0.067	0.0006	0.00098	1.314	1.442	1.575

G16	0.797	0.814	0.066	0.0304	0.00096	0.682	0.815	0.944
G17	0.729	1.082	0.066	0.2604	0.00095	0.952	1.082	1.213
G18	0.989	0.938	0.076	0.4458	0.00096	0.790	0.938	1.090
AvSE	±0.127		±0.066					

Footnote: AvSE=average standard error (SE) or SD, MC=Monte Carlo. For a choseen genotypes i, P(bi)= 2(Prob ($t_{df} > \frac{|b_i-1|}{SE_{b_i}}$), df=number of environment -2.

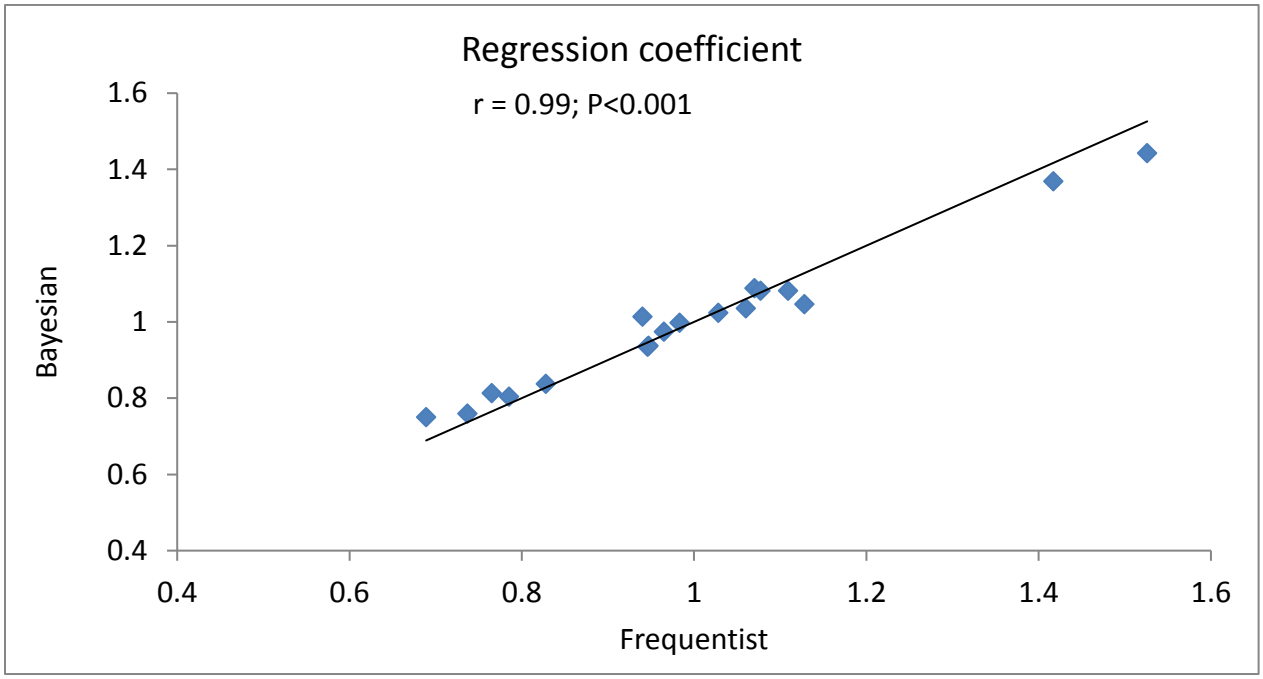

Figure 2. Scatterplot of regression coefficients for the 18 sorghum genotypes for comparing frequentist (x-axis) and Bayesian (y-axis) approaches.

The Figure 2 indicates comparison of Bayesian and frequentist approaches on regression coefficient for each genotype. Correlation between frequentist and Bayesian approaches was 0.989% (P<0.001). Stability indices using Bayesian estimation of regression coefficient reflected in genotypes with slopes significantly different from unit, and thus implying more

number of sensitive/unstable genotype under Bayesian approach compared to frequentist approach. Also the slopes were estimated with higher precision in Bayesian approach relative the frequentist approach.

4.4 Stability Analysis Evaluation Using Coefficient of Determination (R^2)

Table 4 gives estimates of coefficient of determination (ratio of sum of squares due to regression on the environment index to the total sum of squares) under frequentist and Bayesian approaches. Bayesian estimate of R^2 is generally higher than the frequentist approach, for example, frequentist vs Bayesian values were 0.89 vs. 0.94 for G10, 0.92 vs. 0.96 for G11 and 0.87 vs. 0.95 for G13.

Table 4. Coefficient of determination (R^2) under frequentist and Bayesian approaches for the sorghum genotypes in stability analysis on grain yield. Entries were evaluated for three cropping years (2009-2012) at two locations in Sudan.

Genotypes	Frequentist approach	Bayesian approach (priors model : P_3)					
	R^2	R^2	SD	MC error	Percentile		
					2.50%	50%	97.50%
G1	0.96	0.97	0.02	0.00	0.91	0.97	1.00
G2	0.91	0.97	0.02	0.00	0.93	0.97	0.99
G3	0.80	0.93	0.03	0.00	0.85	0.94	0.98
G4	0.80	0.92	0.04	0.00	0.82	0.92	0.98
G5	0.95	0.98	0.01	0.00	0.94	0.98	1.00
G6	0.78	0.93	0.03	0.00	0.87	0.93	0.97
G7	0.89	0.96	0.02	0.00	0.92	0.97	0.99
G8	0.95	0.98	0.01	0.00	0.94	0.98	1.00

G9	0.93	0.97	0.01	0.00	0.94	0.98	0.99
G10	0.89	0.97	0.02	0.00	0.92	0.97	0.99
G11	0.92	0.96	0.03	0.00	0.90	0.96	0.99
G12	0.86	0.95	0.02	0.00	0.90	0.95	0.98
G13	0.87	0.95	0.02	0.00	0.90	0.96	0.99
G14	0.93	0.96	0.02	0.00	0.92	0.96	0.98
G15	0.94	0.96	0.01	0.00	0.93	0.96	0.98
G16	0.96	0.97	0.02	0.00	0.92	0.98	1.00
G17	0.58	0.87	0.04	0.00	0.80	0.88	0.94
G18	0.71	0.76	0.06	0.00	0.64	0.77	0.87
AvSE	±0.0329		±0.0210				

Footnote: AvSE=average standard error (SE)/SD, MC=Monte Carlo. *SE (R^2) was computed as R/d.f. where d.f. is the associted degree of freedom (6-2=4).

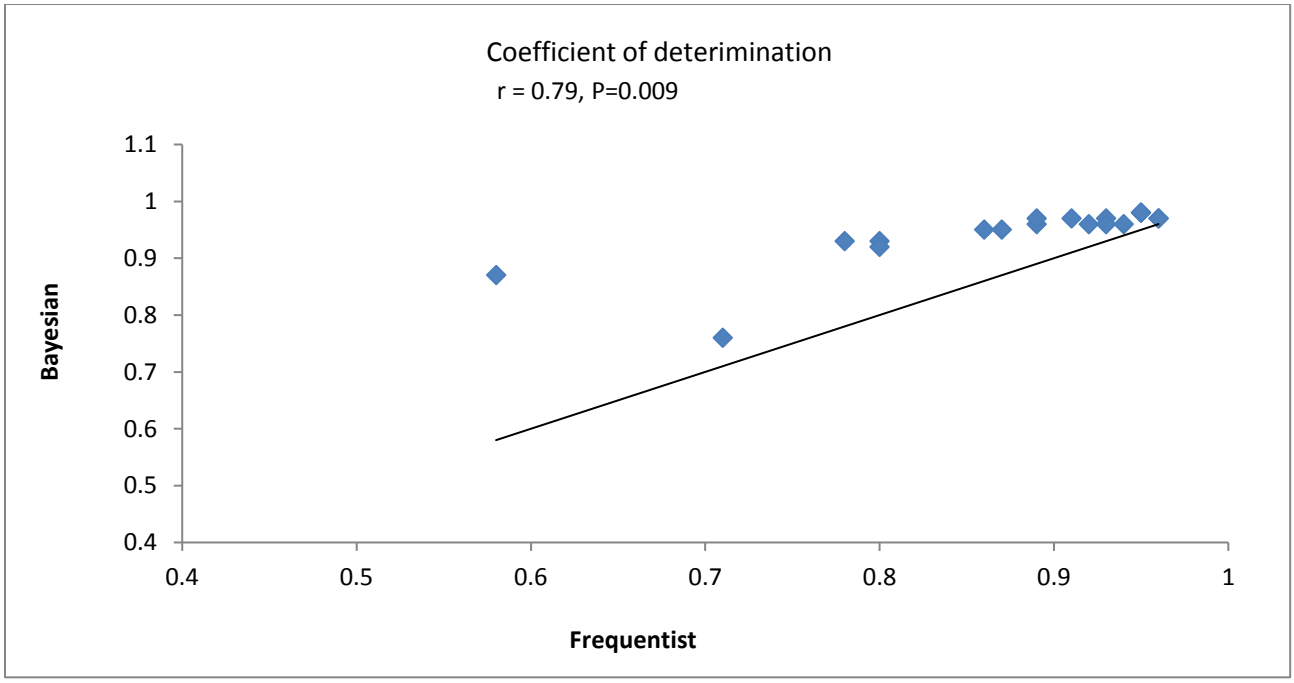

Figure 3. Scatterplot of coefficients of determination (R^2) for the 18 sorghum genotypes for comparing frequentist (x-axis) and Bayesian (y-axis) approaches.

Figure 3 exhibits comparison of Bayesian and frequentist estimate of R-Square (R^2) for each genotype. Correlation between the frequentist and Bayesian approaches was 0.778 (P = 0.009). The goodness of fit of the linear regression was much better in each genotype case. This implies that we can use slope from Bayesian approach to measure adaptability of lines with a higher confidence. Genotype G17 and G18 have lower values of R^2 under frequentist approach but much higher values under the Bayesian approach (0.58 vs 0.87 for G17 and 0.71 vs. 0.76 for G18).

4.5 Stability Analysis Using Coefficients of Variation (CV %)

Table 5 gives estimates of CV associated with each genotype under frequentist and Bayesian approaches. CV is generally lower in Bayesian case compared to frequentist case. In the sense of low CV, G13, G11 and G10 were found most stable genotypes. The overall Bayesian estimate of CV% compared to frequentist was (83.4 vs. 81.07).

Table 5. Coefficient of variation (CV) under frequentist and Bayesian approaches for the sorghum genotypes in stability analysis on grain yield. Entries were evaluated for three cropping years (2009-2012) at two locations in Sudan

Genotypes	Frequentist approach	Bayesian approach (priors model : P_3)					
	CV%	CV%	SD	MC error	Percentile		
					2.50%	50%	97.50%
1	71.61	71.69	7.63	0.11	57.25	71.53	87.20
2	78.03	78.52	6.48	0.09	66.20	78.40	91.40
3	77.56	72.35	6.86	0.10	59.71	72.10	86.76
4	78.57	72.28	6.95	0.10	59.29	72.04	86.20
5	89.08	86.04	7.08	0.09	72.95	85.83	100.60
6	111.92	102.1	9.23	0.13	85.37	101.60	121.70
7	91.62	83.45	6.49	0.09	71.32	83.24	96.68
8	80.57	79.47	6.16	0.09	68.20	79.34	92.42
9	81.64	80.88	6.13	0.09	69.62	80.63	93.26
10	66.36	69.01	5.53	0.08	58.64	68.87	80.37
11	63.6	65.02	5.73	0.08	54.23	64.83	76.64
12	88.09	87.25	6.71	0.09	74.76	86.99	100.80
13	62.77	64.36	5.29	0.08	54.00	64.30	75.17
14	104.21	103.40	7.04	0.10	90.14	103.10	117.70
15	102.72	102.40	6.57	0.10	90.03	102.30	115.60
16	74.83	74.41	7.37	0.11	60.99	74.18	89.41
17	98.7	87.95	6.61	0.10	75.50	87.75	101.10
18	78.83	78.59	6.86	0.09	65.64	78.31	92.63
Mean	83.37	81.07	6.71	0.10	68.55	80.85	94.76

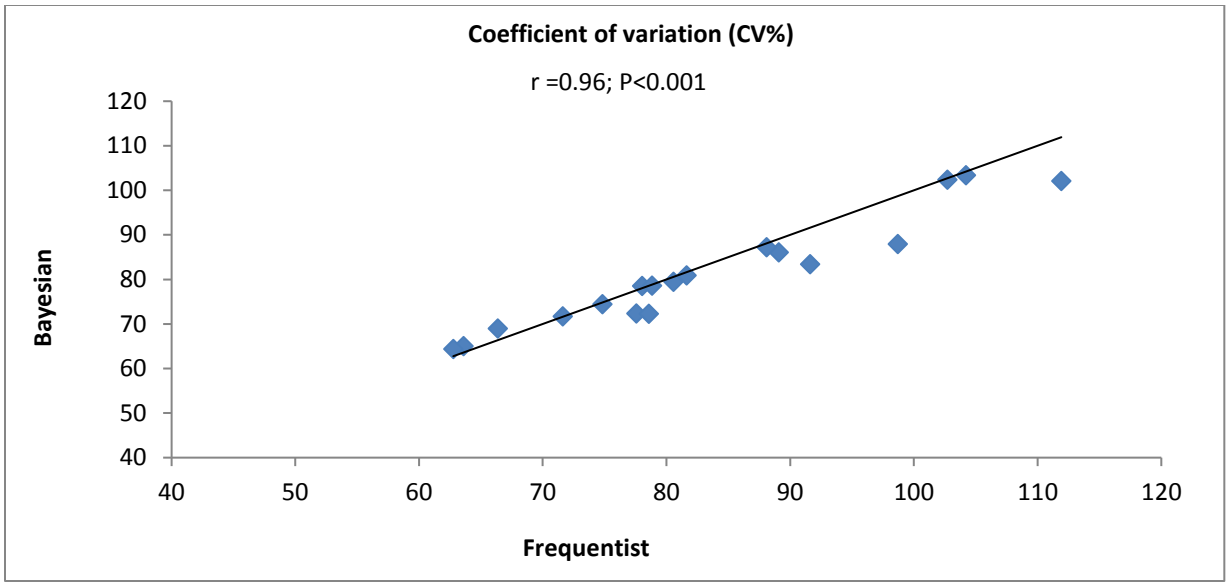

Figure 4. Scatterplot of coefficients of variation (CV) for the 18 sorghum genotypes for comparing frequentist (x-axis) and Bayesian (y-axis) approaches.

Figure 4 exhibits comparison of Bayesian and frequentist approaches for stability of the genotypes based on CV. Correlation between frequentist and Bayesian approaches was 0.964 ($P<0.001$). From frequentist approach of most of CV indicates to highest values while Bayesian to smaller values. Considering the coefficient of variation (CV) as a measure of population variability in agricultural scientists, Bayesian approach gave relatively more reliable indicates. Generally the overall Bayesian stability estimate of CV gave small value compared to frequentist was (81 vs. 83).

5. Discussion

The Bayesian approach has been used in this study for detecting a stability analysis conducted in GE means. Priors in the Bayesian analysis of this study considered uniform, positive half-t, and positive half-normal distributions for various standard deviation components and inverse-Gamma distributions for variance components of the random terms in the response model. 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. Unfortunately due to the inaccessibility of real-data for fitting the priors, we used similar values of parameters of the prior distributions used by others, e.g., Singh et al. (2015) and Omer et al (2014). The appendices provide the required R and WinBUGS codes for carrying out the Bayesian analysis of the multi-environment trials for identifying stable genotypes, and reset the number of iterations, chains and simulations for posterior distributions.

The results highlighted that there were a substantial difference between frequentist and Bayesian approaches for CV%, coefficient of determination and regression coefficient. Furthermore, the Bayesian approach can produce simulated marginal posterior of each parameter considered in the stability model. Comparing frequentist and Bayesian approach adds to knowledge. In comparing the regression coefficient based stability, we noted that the G2 (0.989 vs. 0.998) and G6 (0.945 vs. 1.014) were stable under both frequentist and Bayesian approaches. The Bayesian approach through the simulated marginal posterior of each parameter considers the joint posterior of all

other parameters in the model and their individual estimation precision. Bayesian approach can be used to address many practical questions arising when analyzing GE data and their stability analysis (Josse et al, 2014). Also the advantage of Bayesian strategy has mentioned in Crossa et al. (2011). The most comment regarding stability using two approaches, that environment factor has considered as fixed in model. Bayesian approach against frequentist approach gave higher precision in term of standard error of genotypes means (AvSE = 23.5 vs. 30.4 for Bayesian vs. frequentist), regression coefficient (0.07 vs. 0.13) and coefficient of determination a (0.02 vs. 0.03) respectively. Bayesian approach does not change the concept of stability but rather provides a wider framework for drawing inferences in presence of prior information.

Determination of an appropriate number of environments required for a regional or international trial is also important for obtaining precise estimate of genotypes yields and yield stability

Stability analysis, based on individual indices, led many plant breeders to wonder which stability statistics should be used for their particular problem they have a relatively good chance of being considered in breeding programs. In this study, the results showed that there was substantial improvement under Bayesian approach over frequentist approach for stability analysis.

6. Conclusion

In the analysis of multi-environment trials (METs), the goal was to apply Bayesian approach for assessment of stability indices such as regression coefficient effect, coefficient of variation (CV %) and coefficient of determination (R^2). The half-normal (0, 0.01) distribution was found as the best prior for stability analysis. In general, Bayesian compared to frequentist approach gave

higher precision in term of standard error of genotypes means, regression coefficient and coefficient of determination. Depending on stability parameters used in this study, Bayesian paradigm offers prospects for computing the probability of a genotype being the best performer and more stable. Bayesian approach has important advantages, it offers a more flexible way to understand the GEI and their stability indicates and provides an estimate of stability analysis profile.

Acknowledgments

Authors are thankful to Professor Perla Reyes for making constructive comments and suggestions which led to a substantial improvement in the presentation of the manuscript.

7. References

1. Akinwale MG, Akinyele BO, Odiyi AC and Dixon AGO (2011) Genotype x environment interaction and yield performance of 43 improved Cassava (Manihot esculenta Crantz) genotypes at three agro-climatic zones in Nigeria. **British Biotechnology Journal** 1(3): 68-84.
2. Becker HC and Leon J (1988) Stability analysis in plant breeding. **Plant Breeding 101**: 1-23.
3. Berger J (2006) The case for objective Bayesian analysis. **Bayesian Analysis 3**:385-402.
4. Bhatia VK (2012) some aspects of stability of crop variability. I.A.S.R.I., Library Avenue, New Delhi.
5. Burgueño J, Campos G, Weigel K and Crossa J (2012) Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and densemolecular markers. **Crop Science 52**:707–719.
6. Congdon P (2003) Applied Bayesian models. Chichester: John Wiley and Sons Limited.
7. Cotes JM, Crossa J, Sanches A and Cornelius PL (2006) Bayesian approach for assessing the stability of genotypes. **Crop Science 46**:2654–2665.
8. Cruz CD, Torres RA, Vencovsky R (1989) An alternative approach to the stability analysis proposed by Silva e Barreto. **Revista Brasileira de Genética 12**: 567-580.
9. DeLacy IH, Basford KE, Cooper M and Fox PN (1996) Retrospective analysis of historical data sets from multi-environment trials-theoretical development. In M Cooper and GL Hammer. **Plant adaptation and crop improvement**, P. 243–267. Wallingford, UK,CABI.

10. Eberhart SA and Russel WA (1966) Stability parameters for comparing varieties. **Crop Science 6**: 36-40.
11. Finlay K and Wilkinson GN (1963) The analysis of adaptation in a plant breeding program. **Australia Journal of Agricultural Research 14**:742-754.
12. Francis TR and Kannenberg LW (1978) Yield stability in short-season maize. I. A descriptive method for grouping genotypes. **Canadian Journal of Plant Science 58**: 1029–1034.
13. Gelman A, Carlin JC, Stern H and Rubin DB (1995 Bayesian Data Analysis. New York: Chapman and Hall.
14. Gelman A, Carlin JB, Stern HS and Rubin DB (2003) Bayesian Data Analysis. Boca Raton: Chapman & Hall/CRC, 2nd edition.
15. Gelman A (2006) Prior distributions for variance parameters in hierarchical models. **Bayesian Analysis 3**:515–533.
16. Gilks WR, Richardson S and Spiegelhalter DJ (1996) Markov chain Monte Carlo in Practice. Chapman and Hall, London, UK.
17. Glymour B (2011) Modeling environments: interactive causation and adaptations to environmental conditions. **Philosophy of Science 78**: 448-47.
18. Josse J., van Eeuwijk F., Piepho H.-P., Denis J. (2014). Another look at Bayesian analysis of AMMI models for genotype-environment data. **Journal of Agricultural, Biological, and Environmental Statistics 19**:240–257.
19. Lin CS and Bains MR (1988) A superiority measure of cultivar performance for cultivar × location data. **Canadian Journal of Plant Science 68**: 193–198.
20. Lin CS, Bains MR and Lefkovitch, LP (1986) Stability analysis: where do we stand? **Crop Science 26**:894-900.
21. Motamedi MS and Mohammadi G (2013) Additive main effect and multiplication interaction analysis of grain yield in bread wheat genotypes across environment. **International Journal of Biosciences 8**: 218-225.
22. Mut Z, Gülümser A and Sirat A (2010) Comparison of stability statistics for yield in barley (Hordeum vulgare L.). **African Journal of Biotechnology 9**:1610-1618.
23. Omer SO, Abdalla AWH, Mohammed HM and Singh M (2015). Bayesian estimation of genotype-by-environment interaction in sorghum variety trials. **Communications in Biometry and Crop Science 10**: 82-95.
24. Oral E and Erdemir CA (2012). Bayesian estimation of stable distributions. **Journal of Statistical and Econometric Methods 3**: 39-52.
25. Robert C and Casella G (2004) Monte Carlo Statistical Methods, Springer-Verlag, New York, second edition.
26. Singh M, Al-Yassin A and 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.
27. Verma MM and Gill KS (1975) Genotypes environment interaction its measurement and significance in plant breeding. Teaching Aid. Bul. No.o1, P.AU.,Ludhiana.
28. Yan W and Tinker NA (2006) Biplot analysis of multi-environment trial data: Principles and applications. **Canadian Journal of Plant Science 86**:623–645.

Appendices

A.1: WinBUGS codes to model data from randomized complete block design (RCBD) and estimation of stability parameters

```
# stability.bug (data model)
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)
# m
m ~ dnorm(0.0, 1.0E-6)
# Block
  for (k in 1: (NB-1)){for (j in 1: NP){ b[k,j]~ dnorm(0.0, tau.b)
                                bb[k,j]<- b[k,j]
                               }
                }
   for (j in 1: NP){ b[NB,j]<-  -sum(bb[,j])  }
# Genotyptes
    for (i in 1: (NG-1)){ g[i] ~ dnorm(0.0, tau.g)   }
                    g[NG]<-  -sum(g[1:(NG-1)])
# Envirovmens it ~ 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
     # half normal priors
                sig.e ~ dnorm(0.0, 0.01)
                sig.g ~ dnorm(0.0, 0.01)I(0,)
                sig.b ~ dnorm(0.0, 0.01)I(0,)
                sig.a ~ dnorm(0.0, 0.01)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] } }
                      h2<- sig2g/(sig2g+sig2a/NP+sig2e/(NB*NP))
                      # this heritability is on  mean-basis, 28 MAY 2013
    # CV%
                  CVpc <- 100*sqrt(sig2e)/mn
                  GA20<- 100*1.4*sqrt(sig2g*h2)/mn
                    smp<- sum(p[])
                    ssp<- inprod(p[],p[])
   #CVG
                  for ( i in 1: NG){
                      bG[i] <- (inprod(PredGE[i,],p[])- mean(PredGE[i,])*smp)/ssp
                       GCV[i] <-  100*sd(PredGE[i,])/mean(PredGE[i,])
#regB
#mean square error
                           for ( j in 1: NP){ del[i,j]<- PredGE[i,j] - mean(PredGE[i,]) - bG[i]*p[j]  }
                      dms[i] <-  inprod(del[i,], del[i,])/(NP-2)- sig2e/NB
                      R2G[i]<- bG[i]*bG[i]*sd(p[])*sd(p[])/(sd(PredGE[i,])*sd(PredGE[i,]))
                     } }
# end of BUGS codes
```

A.2: R- codes for reading Datase from a randomized complete block design (RCBD) for stability analysis and calling the 'bugs' function

```
#load packs
library(lattice)
library(coda)
library(R2WinBUGS)
#data from comb.....................................
sdata<- read.table("stadata.txt", header=TRUE)
sdata
y<- sdata$GY
blk<- sdata$Rep
env<- sdata$Envi
geno<- sdata$Geno
NB<- 4
NP<- 6
NG<- 18
N<- NB*NG*NP
N
NPG<- NP*NG
NPG
NBP1<- (NB-1)*NP
NBP
# change to t/ha
#y<- y/1000
print(cbind(y,blk,env,geno))
print(cbind(NB, NP, NG, NPG, NBP1, N, NB1, NP1, NG1))
mn<- mean(y)
mn
summary(y)
100*sd(y)/mn
#---------------------------------
data<- list("y","mn","blk","env","geno","NB","NP","NG","N")
data
inits1<-  list(m=.5, b=c(rep(.01,NBP1)), g=c(rep(.21, NG)), a=c(rep(.2, NPG)), sig.e=.5, sig.b=1, sig.g=0.01, sig.a=1.1)
inits2<-  list(m=.51, b=c(rep(.01,NBP1)), g=c(rep(.22, NG)), a=c(rep(.2, NPG)), sig.e=.5, sig.b=1, sig.g=0.01, sig.a=1.1)
inits3<-  list(m=.52, b=c(rep(.01,NBP1)), g=c(rep(.02, NG)), a=c(rep(.2, NPG)), sig.e=.5, sig.b=1, sig.g=0.01, sig.a=1.1)

# for Gamma priors
#inits1<-  list(m=2.1, b=c(rep(.01,NBP)), g=c(rep(.21, NG)), a=c(rep(.2, NPG)), tau.e=.5, tau.b=3, tau.g=1, tau.a=1.2)
#inits2<-  list(m=1.1, b=c(rep(.01,NBP)), g=c(rep(.22, NG)), a=c(rep(.2, NPG)), tau.e=.5, tau.b=2, tau.g=1, tau.a=1.1)
#inits3<-  list(m=1.5, b=c(rep(.01,NBP)), g=c(rep(.02, NG)), a=c(rep(.2, NPG)), tau.e=.5, tau.b=3, tau.g=1, tau.a=1.1)
inits <- list(inits1, inits2, inits3)
inits
parameters <- c("m","g", "PredG","p", "PredE","a", "PredGE","sig2g","sig2e","sig2a",
"h2","CVpc","GCV","bG","R2G","GA20")
parameters
stability.sim <- bugs(data, inits, parameters, "stab.bug", n.chains=3, n.iter=50000, n.sims=5000, debug=TRUE)
```