

THE GGE BIPLLOT ON RCIM MODEL FOR ASSESSING THE GENOTYPE-ENVIRONMENT INTERACTION WITH SIMULATING OUTLIERS: ROBUSTNESS IN R-SQUARED PROCRUSTES

Alfian Futuhul Hadi¹, Halimatus Sa'diyah², Dimas BC Wicaksono³

¹ Statistical Laboratory, Department of Mathematics, University of Jember, Indonesia

² Biometrics & Plant Breeding Laboratory, Department of Agronomy, University of Jember, Indonesia

³ Biograph-i Research Group, Faculty of Public Health, University of Jember, Indonesia

e-mail: sadiyah@unej.ac.id

DOI: 10.14710/medstat.15.2.209-219

Article Info:

Received: 30 January 2022

Accepted: 29 March 2023

Available Online: 4 April 2023

Keywords:

GGE Biplot; AMMI; RCIM; outliers; Procrustes.

Abstract: The genotype by environment interaction (GEI) analysis was usually done by Additive Main Effects and Multiplicative Interaction (AMMI) model with Biplot features, and recently there was a Row Column Interaction Model (RCIM) alternatively. In the Biplot of genotype (G) and genotype by environment (GE) interactions, known as the GGE Biplot, the main effect of environment (E) was deleted, while the main effect of G and the interaction effect of GE is kept and combined. Subsequently, continuing our recent research of the robustness of the GGE Biplot in RCIM models, this paper aims to develop the GGE Biplot by RCIM model to analyze the GEI with outlying observations. We used the RCIM model with Asymptotic Laplace Distribution (ALD) that was applied on the simulated data with scattered and single environment outliers to evaluate the robustness of the GGE Biplot. In addition, the robustness was evaluated using the R-squared statistic of the Procrustes analysis. It is shown that the GGE Biplot of RCIM with the ALD family function provides better robustness than the Gaussian. A noticeable superiority of the GGE Biplot with RCIM ALD appeared as the percentage of single environment outliers reach the number of rows of the data matrix.

1. INTRODUCTION

Plant breeding efforts are typically accomplished through two key elements: (i) agronomically conducting a multi-environment trial (MET); (ii) doing a statistical assessment of the genotype-environment interaction (GEI). The MET is a common experiment that breeders do before releasing a new genotype. The emergence of the GEI makes statistical analysis of the interaction become more difficult (Hadi et al., 2018b). In 1988, Gauch introduced a model that combined additive components for main effects (genotype-environment) and multiplicative components for genotype-environment interaction called AMMI (Additive Main effects and Multiplicative Interaction). The interaction term in AMMI model can be modeled with Singular Value Decomposition (SVD). The SVD is a technique commonly used for reduces dimensionality but retains most of the variability. The SVD will visualize the interaction terms graphically by Biplot, making

the GEI analysis easier. Biplot analytics is a descriptive method with two dimensional that visualized the interaction of genotype and environment (Yasin & Rusgiono, 2013).

In multi-environment trials (MET), genotypes (G) are applied to different environments (E) to assess interactions between G and E. The AMMI model is commonly used to analyze stability and adaptability in the GEI studies. Since G and GE must be considered simultaneously when making cultivar selection decisions, Yan et al. (2000) Evaluation of GEI and stability performance by retaining the main effect (G) and the environmental genotypic interaction effect (GE) and combining them as the GGE, while the environmental main effect (E) being removed. The GGE biplots are useful for (i) showing patterns in the data that lead to the identification of different mega environments, who-won-where, and (ii) identifying high-yielding, stable cultivars and which ones are adaptable (Yan et al, 2001).

Alternatively, the Row Column Interaction Model (RCIM) (Yee & Hadi, 2014) can be used. From the perspective of the RCIM, the AMMI or generalized AMMI model was a model with one or more multiplicative interaction components in addition to the row and column main effects. The singular value corresponding to each multiplicative component is often factorized as a measure of the strength of the association between the row and column values, indicating the component's or axis's importance. Biplots produced using SVD have a vulnerability to the existence of outliers, as reported by Hadi (2011). But the proposed GGE Biplot has the potential to be robust against outliers (Hadi et al., 2018). While the VGAM family functions are probably beneficial together with the RCIM model including the asymptotic Laplace distribution (ALD) (Yee, 2015). Hadi et al (2018a) showed that the ALD provides a better robust fitted value than the Normal distribution in the RCIM modeling of the GEI.

For these reasons regarding the robustness of RCIM and GGE biplots, this paper is intended to provide a final overview of the previous working GGE biplot functions of the RCIM model. There are several features of the GGE biplot using the RCIM model applied to the simulation data with a single environmental outlier to provide GEI analysis in the event of outliers. We will discuss some interesting topics. The first is about the outliers: (i) the magnitude and the number (%) of outliers; (ii) the placement of the outliers. The second is about the statistics of the Biplot robustness: proportion of total variance explained by Biplot and the similarity of Biplot with Procrustes analysis. For an illustrative example of the application of the GGE Biplot, it is recommended to see Farshadfar et al (2012), or the precursor paper of the GGE Biplot of Yan et al (2007).

2. LITERATURE REVIEW

2.1. The RCIM Model in the RR-VGLM

This section will discuss a framework for developing the GGE Biplot in RCIM. The RCIM was developed by Yee & Hadi (2014). The RCIM is an extension of the Reduced-Rank Vector Generalized Model (RR-VGLM) where the first linear predictor is modeled by the sum of row effect, column effect, and the row by column interactions effect. Yee & Hadi (2014) define RCIMs as an RR-VGLM as Equation (1).

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{r=1}^R c_{ir} a_{jr} \quad (1)$$

where $E(Y_{ij})$ is the expected value of the observed value of the i^{th} row and the j^{th} column, the μ is the grand mean, α_i is the main effect the i^{th} row; β_j is the j^{th} column main effect, and

$\sum_{r=1}^R c_{ir}a_{jr}$ is the decomposition of the interaction effects. We now pay more attention to the interaction effect of $\sum_{r=1}^R c_{ir}a_{jr}$. If we turn to the decomposition of SVD and use the sum-to-zero constraints rather than the set-to-zero constrain or the corner constrain, the RCIM will turn to the AMMI model (Hadi et al, 2017).

The basic model of the GGE was constructed by following the model of Yan et al (2000). The mean yield of the i^{th} genotype in the j^{th} environment is commonly approximated by \hat{Y}_{ij} and described in general as Equation (2),

$$\hat{Y}_{ij} = \mu + \alpha_i + \beta_j + \phi_{ij} \quad (2)$$

where α_i is the main effect of i^{th} genotype, β_j is the main effect of j^{th} environment, and ϕ_{ij} interaction between genotype i and environment j . Subjecting the ϕ_{ij} to SVD results in the AMMI model. Avoiding the main effect of genotype, α_i in Equation (2) allows the variation of genotype to be absorbed into the interaction term of ϕ_{ij} .

2.2. The GGE Biplot in RCIM Model

Based on the RCIMs model as the form of RR-VGLM in Equation (1) and by subjecting the ϕ_{ij} in the Equation (2) to SVD (with first two biggest singular values) will turn the form of GGE Biplot explicitly, as described in Equation (3),

$$\hat{Y}_{ij} - \mu - \beta_j = \phi_{ij} = \xi_{i1}\lambda_1\eta_{j1} + \xi_{i2}\lambda_2\eta_{j2} \quad (3)$$

where λ_1 and λ_2 are the singular value for PC1 and PC2, ξ_{i1} and ξ_{i2} are the score of PC1 and PC2 of the i^{th} genotype, η_{j1} and η_{j2} are the score of PC1 and PC2 of the j^{th} environment.

The GGE Biplot will display the PC1 and PC2 by rewriting Equation (3) as below:

$$\begin{aligned} \hat{Y}_{ij} - \mu - \beta_j &= \xi_{i1}\lambda_1\eta_{j1} + \xi_{i2}\lambda_2\eta_{j2} \\ &= \xi_{i1}\lambda_1^{0.5}\lambda_1^{0.5}\eta_{j1} + \xi_{i2}\lambda_2^{0.5}\lambda_2^{0.5}\eta_{j2} \\ &= \xi_{i1}^*\eta_{j1}^* + \xi_{i2}^*\eta_{j2}^* \end{aligned} \quad (4)$$

Finally, the GGE Biplot constructed by plotting ξ_{i1}^* and η_{j1}^* against ξ_{i2}^* and η_{j2}^* . Therefore, with Equation (4), it is now clear that the RCIM model of Equation (1) with the rank of 2 can also provide the GGE Biplot.

3. MATERIAL AND METHOD

Two-way table data randomly generated in 18 rows (genotypes) x 9 columns (environments) under a Normal distribution according to the AMMI2 model described by Rodrigues et al (2016). Row-column interactions are modeled by two multiplicative terms with RCIM of rank = 2, according to their equivalence with the AMMI2 model. For more information, you can see what we've done before in Hadi et al (2018a).

As shown in Figure 1, this statistical framework and methodology were conducted throughout three phases. The first phase is about generating randomly the data matrix, then followed by the second, simulating the outliers. The last phase was including modeling the RCIM, the GEE Biplot, and evaluating the robustness.

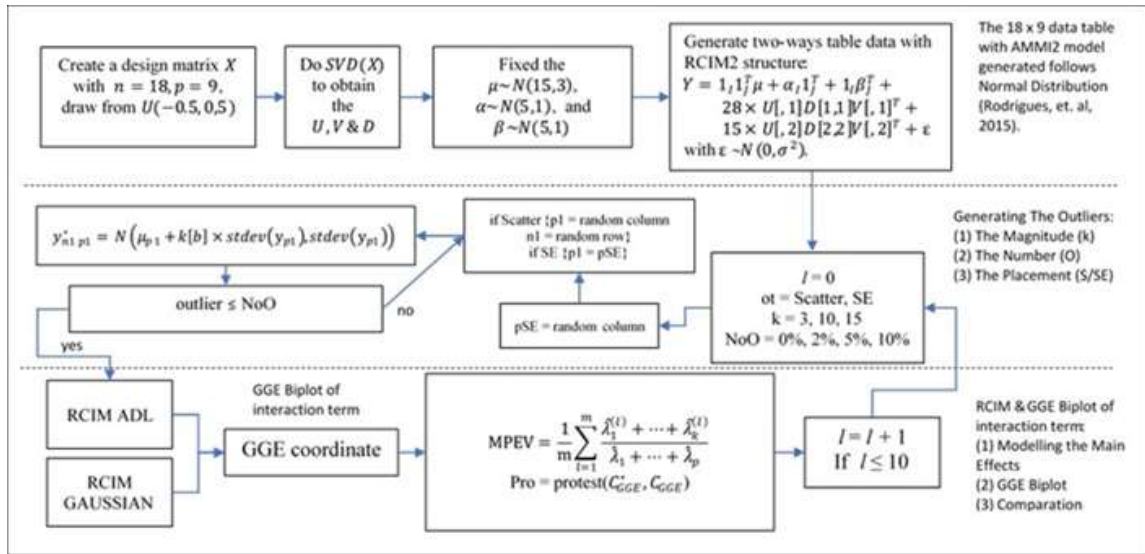


Figure 1. Statistical Framework and Methodology

3.1. Simulation Study

Outliers are defined as parts of the observation that are different from most of the corresponding observation data sets. If the standard deviation of k times is greater than the original mean (k is greater than 3), the observations are considered outliers. The outlier was mathematically expressed as Equation (5) (Rocke & Woodruff, 1996). Outliers were randomly generated according to the normal distribution $N(\mu_j + k\mu_j, \sigma_j^2)$, where the μ_j is the mean of the data in column j , and σ_j^2 is the variance of the error term (or the variance of the certain environment), and k are constant values of outlier size, $k = 3, 10, 15$. The value of $k = 3$ chosen for small magnitude and other k is chosen by assumption that GEI data sometimes have a huge outlier value (Nascimento et al., 2021).

$$y_i^* \geq \mu_j + k\sigma. \quad (5)$$

Next, we ran a simple simulation scheme to add outliers to the data matrix. Outliers are added to the generated data, randomly placed as we did before in Hadi et al (2018a), sets the number of outliers to 4, 9, and 17 for the entire 2%, 5%, and 10% of whole cells in the data matrix, respectively. These outliers are placed in the RCIM2 data table in two different arrangements: (i) scattered outliers, and (ii) single environment outliers. The scattered outlier was a simple random placement at the whole data matrix, while the single environment outlier placement was done by systematic column-wise placement. For both placement's detail, see Hadi et al (2018b). This simulation scheme will be repeated about ten times to get randomness from each replication for better evaluation.

3.2. RCIM Model Fitting & The GGE Biplot

In R programming, the RCIM can be estimated by the `rcim` command from `VGAM` package (Yee & Hadi, 2014) with Gaussian and Asymptotic Laplace distribution family function below:

```
rcim(data, family = gaussianff(), Svd.arg=TRUE, Rank=0)
rcim(data, family = alapace1(tau=0.5), Svd.arg=TRUE, Rank=0)
```

Then, use SVD from both models to decompose the rest and get the first two major components (PC1 and PC2) plotted through the `ggplot2` package.

3.3. Evaluating The Influence of Outliers and The Robustness

In our simulation study, we use the Goodness of Fit (GoF) statistics of Biplot to evaluate the influence of outliers and the robustness of the GGE Biplot. We use the Biplot Mean Percentage of Explained Variability (BMPEV) together with the Procrustes rotation methods to evaluate GoF. Here we will talk about those two methodologies as follow:

1. The Biplot Mean Percentage of Explained Variability (BMPEV)

The Biplot MPEV was used formally in a study of the GoF (Gabriel, 2002). It has the form in Equation (5) as follow

$$\frac{1}{m} \sum_{l=1}^m \frac{\hat{\lambda}_1^{(l)} + \hat{\lambda}_2^{(l)}}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (5)$$

where λ_j , $j = 1, \dots, p$, are the eigenvalues of the original data matrix (two-way data tables), the $\hat{\lambda}_1^{(l)}$ and $\hat{\lambda}_2^{(l)}$, for $l = 1, \dots, m$, with $m = 10$, are the eigenvalues of contaminated data matrix for each replication of simulation. Siswadi & Bakhtiar (2011) have revealed the interconnection between Biplot, principal component, and Procrustes analyses in calculating the goodness-of-fit coefficients of two configuration matrices. Here we use the percentage of explained variability from the Procrustes, as an average from some replicates.

2. Procrustes Analysis

Multivariate analysis often gives coordinates of the inner point cluster multi-dimensional space. Specifically, this is obtained from the attempt to represent data as the distance between points of objects in space multidimensional, such as PCA or Biplot which involves the distance (Euclidean or Mahalanobis). The distance between points does not change with changing point's origin, as well, if the coordinate axis is being rotated. Unfortunately, the orientation and scale of the axes are usually slightly different, making it difficult to find such a match. Ordination is best compared by using Procrustes Rotation (Oksanen, 2015).

Procrustes analysis was developed to compare the observed data matrix with the target matrix. In the target matrix, both matrices have the same number of samples (as rows) and sometimes different numbers of variables (as columns) (Siswadi et al, 2012). As a general rule, in order to achieve shape and size similarity between the two configurations, one configuration remains fixed and the other configuration is transformed to match the first configuration. Procrustes rotates one configuration to maximize its similarity to another and also tests the non-randomness between the two configurations. The rotation of Procrustes causes the matrix to rotate, maximizing its similarity to the target matrix and minimizing the sum of the squares (SS) of the differences. The `protest` function inherited from Procrustes then returns the SS and measures the difference between X and (rotated) Y. In other words, it is a similarity error SS. The `protest` uses a correlation-like statistic derived from the symmetric Procrustes SS as $R\text{-squared} = 1 - SS$ (Peres-Neto & Jackson, 2001). R-squared statistics are used to describe the similarity of the two forms of configuration compared. This value indicates what percentage of the observations in both configurations can be considered the same. If this value is equal to 1 (100%), it means that both configurations are equal.

4. RESULTS AND DISCUSSION

4.1. Influence of The Outliers in the GGE Biplot

The data were randomly generated in the error term of the Normal distribution, so no outliers are displayed. We see that the percentage of total variance explained of GGE Biplot was originally 86.11% without outlier (Figure 2). This variance explained goes down with the increasing number of scattered outliers, to about 53% for a small magnitude of $k = 3$. The worst was fell steeply down to only 39% – 40%, by 17 outliers for value magnitude of $k = 10$ and 15 (Figure 2a). It shows that there was a loss of capability to explain the total variance of the GGE interaction matrix, affected by a scattered outlier. Figure 2b shows that the single environment outliers reduce GGE's biplot MPEV from 86.11% to about 69%. This decline looks a bit flatter than a steeper decline due to the scattered outliers (Fig. 2a).



Figure 2. The Biplot Mean Percentage of Explained Variability (Biplot MPEV) by The GGE Biplot of RCIM With Gaussians Family Function, Affected by (a) The Scattered Outlier, and (b) The Single Environment Outliers.

Figure 3 shows that the GGE Biplot RCIM with ALD family function has better fitting when the magnitudes of $k = 3$ and $k = 10$, as its higher MPEV than the Gaussian in case of scattered outliers. However, for large $k = 15$, its goodness of fit is a little bit lower than the Gaussian. The situation is different for contaminated data from a single environment outlier. As shown in Figure 4, for a small magnitude of outlier ie. $k = 3$, ALD has a better fit than Gaussian. But for a larger magnitude, ALD cannot perform well. Only when $k = 10$ with 10% outliers ALD does give higher Biplot MPEV.

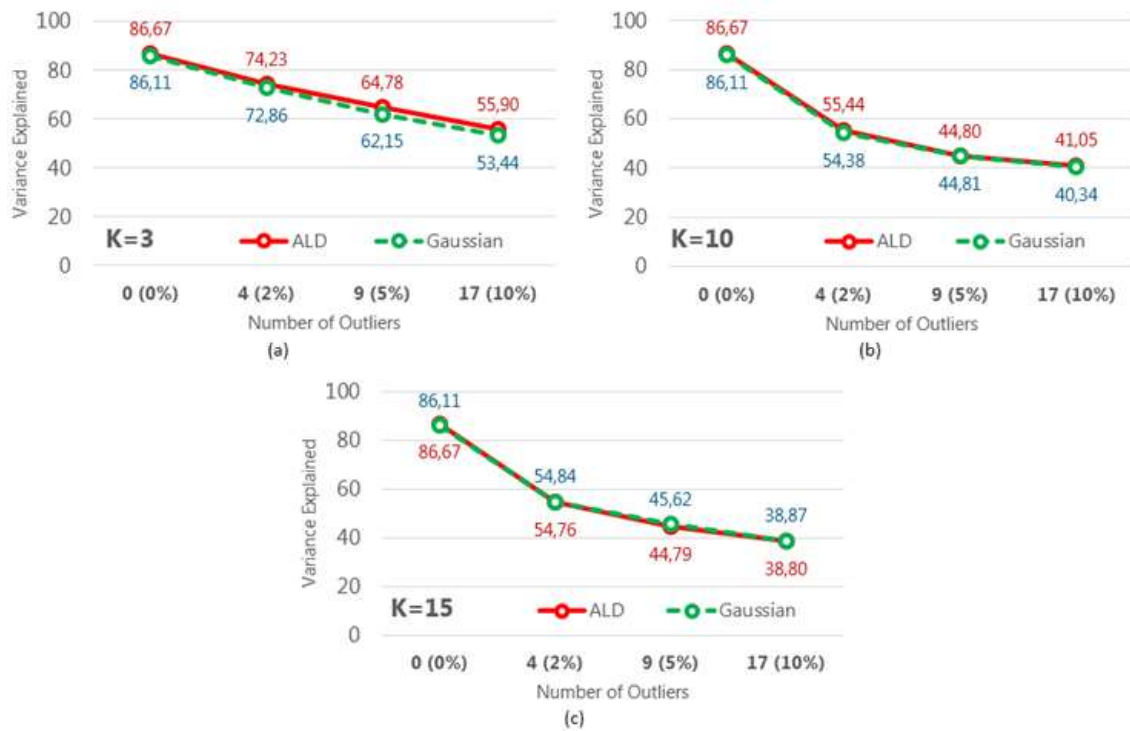


Figure 3. The Biplot MPEV of the GGE Biplot of RCIM with Gaussians and ALD Family Function, Affected by The Number of Scattered Outliers with The Magnitude of $k = 5, 10, 15$, for (a), (b), and (c) respectively

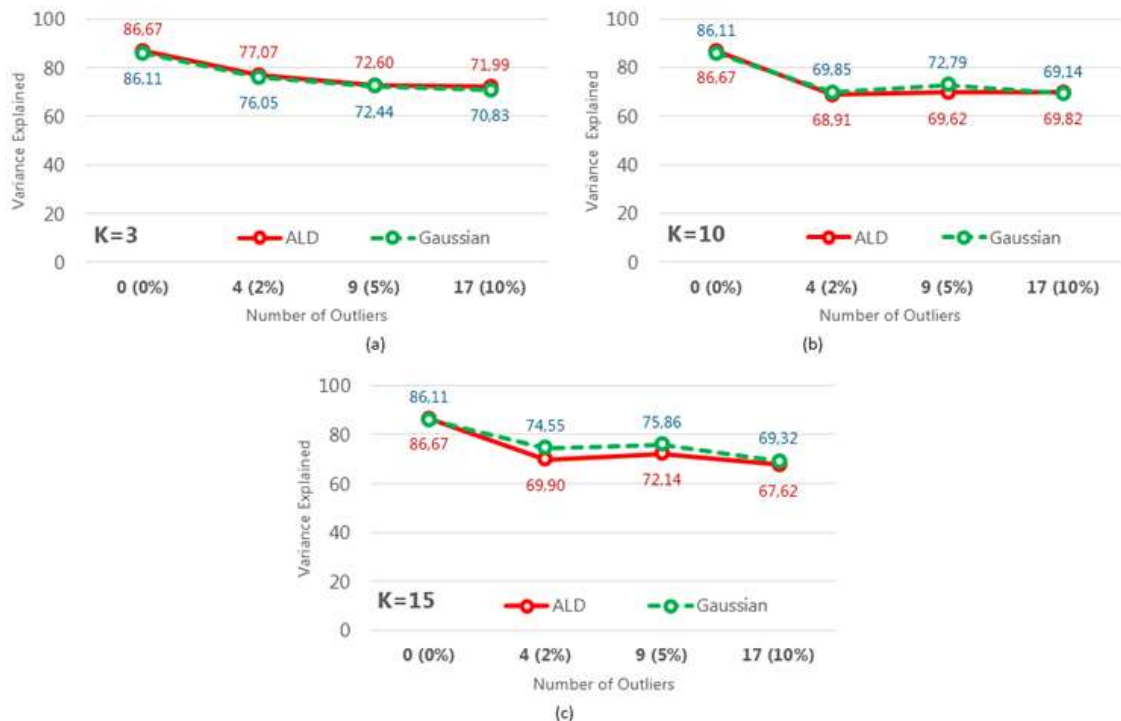


Figure 4. The Biplot MPEV by The GGE Biplot of RCIM with Gaussians and ALD Family Function, Affected by The Number of Single Environment Outliers with The Magnitude of $k = 5, 10, 15$ for (a), (b), and (c), respectively

4.2. The Robustness of GGE Biplot

The following describes the robustness of the GGE Biplot in the RCIM model in the case of the two types of outliers. We used R-square Procrustes statistics to assess its robustness. Procrustes analysis was performed to measure the similarity between the two biplot coordinates, to check if the outliers affected the change in the biplot configuration compared to the original configuration.

In the case of a scattered outlier, the GGE Biplot of RCIM with the ALD family function has a higher similarity than the Gaussian. In addition, Figure 5c shows that when $k = 3$, both of the Gaussian and ALD models hold more than 90% similarity even though 10% of the data has been contaminated with scattered outliers. But at $k = 15$, the highest R-square similarity was only 50.71% with 2% of scattered outliers. And with 10% of scattered outliers, the R-square goes down to the smallest one, 27.6% on GGE Biplot RCIM with ALD and 23.88% on Gaussian. Overall, Figure 5 shows us that the ALD family function gives better similarity than the Gaussian in the case of scattered outliers, including the worst in Figure 5c. This means that the ALD has good robustness in the GGE Biplot of RCIM in the case of scattered outliers.

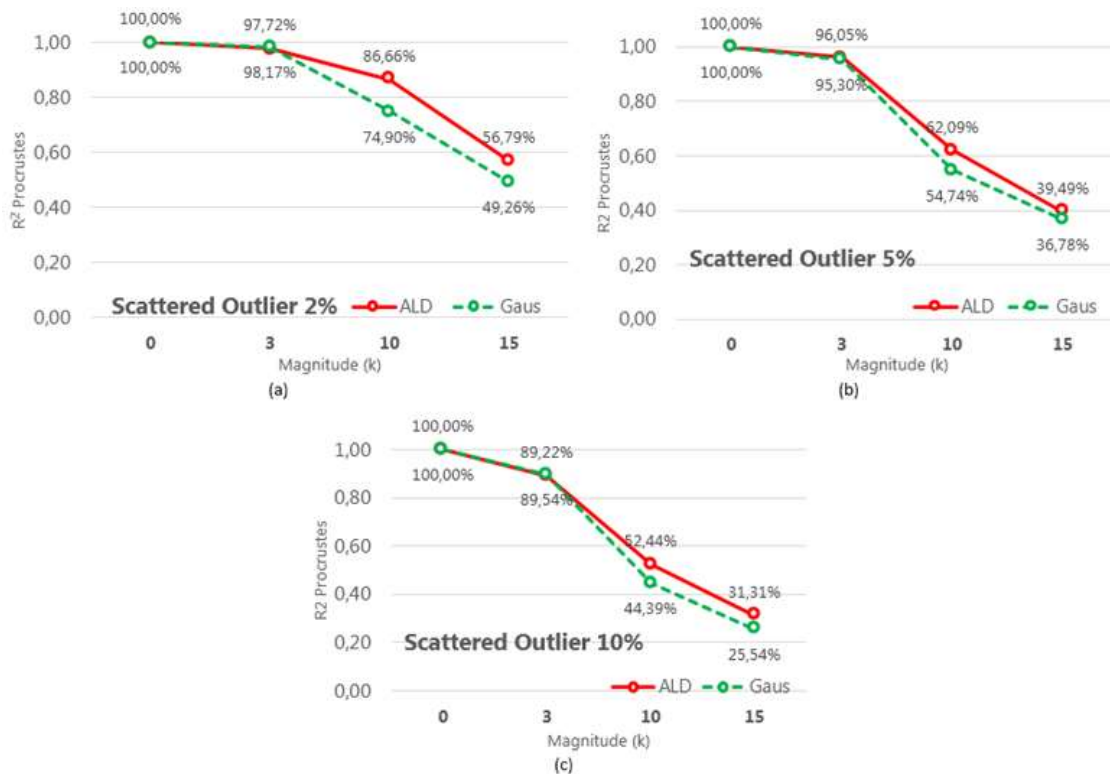


Figure 5. The R-Square Procrustes of The GGE Biplot of RCIM with Gaussian and ALD Family Function, Affected by The Magnitude of an Outlier, with The Number of The Scattered Outlier of 2%, 5%, and 10% for (a), (b), and (c) respectively

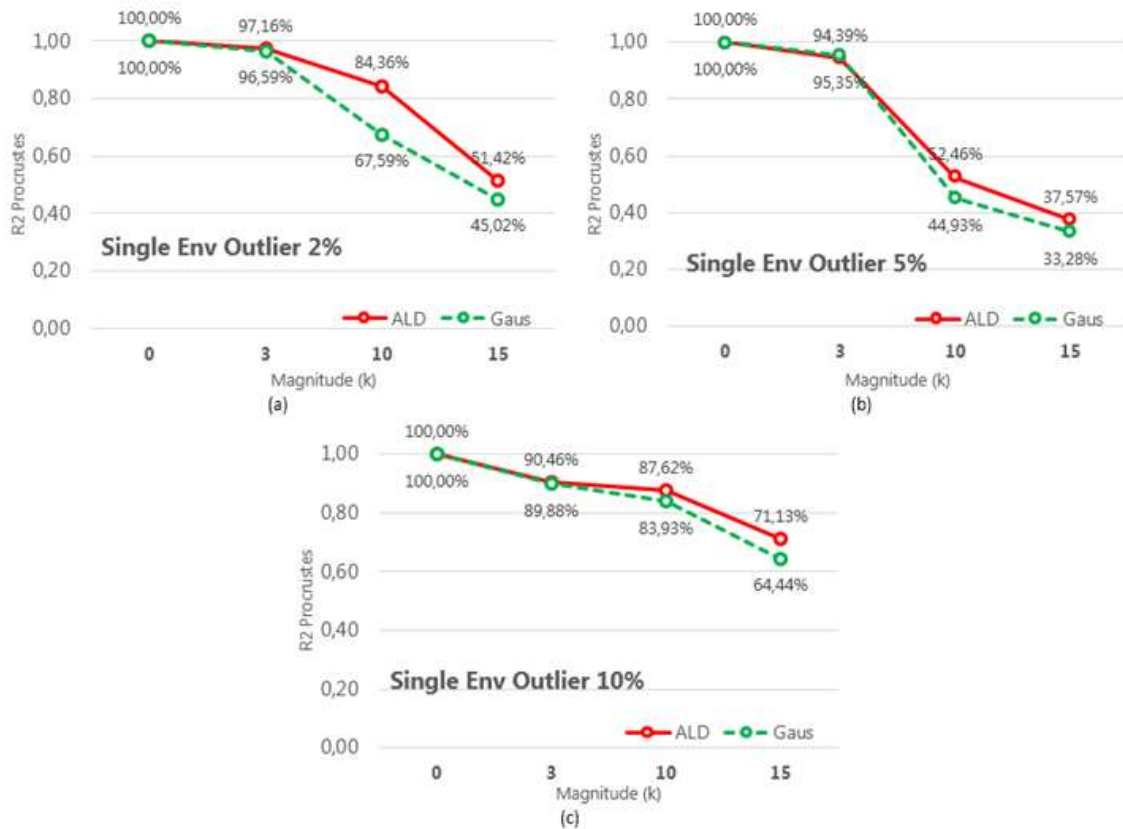


Figure 6. The R-Square Procrustes of The GGE Biplot of RCIM with Gaussians and ALD Family Function, Affected by The Magnitude of an Outlier, with 2%, 5%, and 10% of Single Environment Outlier for (a), (b), and (c) respectively

Together with the previous one, Figure 6 now shows that the GGE Biplot of RCIM with the ALD family function provides superior robustness than the Gaussian family function, for all contaminated data with any type of outliers. A noticeable difference between the ALD vs the Gaussian appears when the percentage of single environment outliers reach the number of rows, at 10% of outliers. It means that outliers fill out the entire environment (column). In this condition, the Gaussian cannot perform well. In the great magnitude of $k=15$, the Gaussian's similarity goes down linearly to 58%, while ALD can hold at least 71% similarity.

4.3. Discussion

What we did in this study was develop a statistical assessment method by establishing a G+GE interaction component. In certain cases in plant breeding, this GGE interaction is more required than GE or GE+E interactions (Yan et al., 2001; Yan et al., 2007). The second we developed here was the robustness of the GGE interaction matrix decomposition by adopting several robust decomposition algorithms that have been developed for the GE as previously done by Hadi (2011) and Hadi et al. (2018b). The study of outliers in GGE Biplot was completed in a previous academic work that developed a method to detect the presence of outliers in a two-way tabular data interaction investigation. Hadi et al. (2022) detect outlier observations using elliptic confidence region and identify environment (E) with large variance observations that can be readily detected by Robust Biplots. This reinforces the importance of handling outliers in GGE interaction analysis.

If the study of the GE Robustness, Rodrigues et al. (2016) warned that when the data have only significant main effects (i.e. no interaction), the interaction should be ignored and the family of AMMI models is not useful. So in our model, one needs to be carefully considered when not only the GE interaction is not significant, but also the main effect G is also not significant. In another field, our results here can be adapted for interaction assessment with an additional component of a single main effect. If desired, the practitioner can adapt the form of interaction that suits the purpose of the interaction assessment.

5. CONCLUSION

Here the GGE Biplot of RCIM with ALD can be said to have good robustness against the single environment outliers. Single environment outliers mean that the outliers are distributed first in a particular column, and as the number of outliers increases, it fills the entire row of cells in that particular column. Therefore, the column's mean will be much different from others. Instead, the scattered outliers are placed randomly around the whole cell of the matrix. There is no guarantee that the mean of columns (rows) will differ from others. In the case of scattered outliers, the GGE Biplot of RCIM with ALD was very similar to the Gaussian, just a little bit better than the Gaussian. It's parallel with the previous work of Hadi et al (2018b) that the GGE Biplot of RCIM with Gaussian has its robustness to scattered outlier.

ACKNOWLEDGMENT

Hadi would like to thank Thomas Yee for collaborating and join authorship in the function of RCIM, marking row and column baselines of matrix data in VGAM (Vector Generalized Linear and Additive Models), at the University of Auckland during 2010-2011.

REFERENCES

- Farshadfar, E., Mohammadi, R., Aghae, M., & Vaisi, Z. (2012). GGE Biplot Analysis of Genotype \times Environment Interaction in Wheat-barley Disomic Addition Lines. *Australian Journal of Crop Science*, 6(6), 1074–1079.
- Gabriel, K. R. (2002). Goodness of Fit of Biplots and Correspondence Analysis. *Biometrika*, 89(2), 423–436. <https://doi.org/https://doi.org/10.1093/biomet/89.2.423>
- Hadi, A. F. (2011). Handling Outlier in Two-Ways Table by Robust Alternating Regression of FANOVA Models: Towards Robust AMMI Models. *Jurnal Ilmu Dasar*, 12(2), 123–131.
- Hadi, A. F., Sa'diyah, H., & Iswanto, R. (2017). On generalization of Additive Main Effect and Multiplicative Interaction (AMMI) Models: An Approach of Row-Column Interaction Models for Counting Data. *Malaysian Journal of Mathematical Sciences*, 11(S), 115–141.
- Hadi, A. F., Hasan, M., & Sadiyah, H. (2018a). Handling Outlier in Two-Ways Table Data: the Robustness of Row-Column Interaction Model. *Journal of Physics: Conference Series*, 1028(1), 012222. <https://doi.org/10.1088/1742-6596/1028/1/012222>
- Hadi, A. F., Sadiyah, H., & Hasan, M. (2018b). On the development of The GE and the GGE Interaction Biplot in the RCIM Model and the Evaluation of Its' Robustness to the Outlying Observations. *Journal of Physics: Conference Series*, 1132, 01207. <https://doi.org/10.1088/1742-6596/1132/1/012077>

- Hadi, A. F., Anggraeni, D., Sadiyah, H., and Wicaksono, D. B. C., "Multivariate outliers detection on GGE Biplot", AIP Conference Proceedings 2471, 020001 (2022) <https://doi.org/10.1063/5.0083440>
- Nascimento, G.F.M., Wurtz, F., Kuo-Peng, P. Delinchant, B.; Jhoe Batistela, N. (2021) Outlier Detection in Buildings' Power Consumption Data Using Forecast Error. *Energies*, 14, 8325. <https://doi.org/10.3390/en14248325>
- Oksanen, J. (2015). *Multivariate Analysis of Ecological Communities in R: Vegan Tutorial*. <https://www.mooreecology.com/uploads/2/4/2/1/24213970/vegantutor.pdf>
- Peres-Neto, P. R., & Jackson, D. A. (2001). How Well Do Multivariate Data Sets Match? The Advantages of A Procrustean Superimposition Approach Over the Mantel Test. *Oecologia*, 129(2), 169–178. <https://doi.org/10.1007/s004420100720>
- Rocke, D. M., & Woodruff, D. L. (1996). Identification of Outliers in Multivariate Data. *Journal of the American Statistical Association*, 91(435), 1047–1061. <https://doi.org/10.2307/2291724>
- Rodrigues, P. C., Monteiro, A., & Lourenço, V. M. (2016). A Robust AMMI Model for the Analysis of Genotype-By-Environment Data. *Bioinformatics*, 32(1), 58–66. <https://doi.org/10.1093/bioinformatics/btv533>
- Siswadi, & Bakhtiar, T. (2011). Goodness-of-fit of Biplots via Procrustes Analysis. *Far East Journal of Mathematical Sciences*, 52(2), 191–201.
- Siswadi, Bakhtiar, T., & Maharsi, R. (2012). Procrustes Analysis and the Goodness-of-Fit of Biplots: Some Thoughts and Findings. *Applied Mathematical Sciences*, 6(69–72), 3579–3590.
- Yan, W., Cornelius, P. L., Crossa, J., & Hunt, L. A. (2001). Two Types of GGE Biplots for Analyzing Multi-Environment Trial Data. *Crop Science*, 41(3), 656–663. <https://doi.org/10.2135/cropsci2001.413656x>
- Yan, W., Hunt, L., Sheng, Q., & Szlavnic, Z. (2000). Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. *Crop Science*, 40(3), 597–605. <https://doi.org/10.2135/cropsci2000.403597x>
- Yan, W., Kang, M. S., Ma, B., Woods, S., & Cornelius, P. L. (2007). GGE biplot vs. AMMI Analysis of Genotype-By-Environment Data. *Crop Science*, 47(2), 643–653. <https://doi.org/10.2135/cropsci2006.06.0374>
- Yasin, H., and Rusgiyono, A. (2013) Analisis Variabel Kanonik Biplot untuk Bank Umum di Jawa Tengah. *Media Statistika*, 6(2), 71-80. <https://doi.org/10.14710/medstat.6.2.71-80>
- Yee, T., & Hadi, A. F. (2014). Row–column Interaction Models, with an R Implementation. *Computational Statistics*, 29(6), 1427–1445. <https://doi.org/10.1007/s00180-014-0499-9>
- Yee, T. W. (2015). *Vector Generalized Linear and Additive Models, With an Implementation in R*. New York: Springer. <https://doi.org/10.1007/978-1-4939-2818-7>