Applied Science Innovation (ASI)
Published by College of Physical and Applied Sciences, MOUAU

- *Received:* 22 June 2026
- *Revised*: 14 July 2026
- *Accepted:* 30 July 2026
- *Published:* 24 August 2026

ADOPTING THE ROBUST AMMI FOR GENOTYPE-BY-ENVIRONMENT DATA ON CASSAVA STARCH YIELD IN THE EASTERN REGION OF NIGERIA

Ezenwanyi Uba[1*], Chisimkwuo John[2], , E. J. Ekpenyong[3], E.U. Ohaegbulem[4] and Felix N. Nwobi[5].

[1]Biosciences, National Roots Crops Research Institute, Umudike, Nigeria.
[23]Department of Statistics, Michael Okpara University of Agriculture, Umudike, Abia State, Nigeria
[45]Department of Statistics, Imo State University, Owerri, Nigeria

***Corresponding author:** john.chisimkwuo@mouau.edu.ng

ABSTRACT

Cassava starch has multiple importance as used in the food, beverages, candy, chemical, glue, paper making, textile, biodegradable materials, cosmetic, pharmaceutical, and animal feed industries. Various cassava genotypes across different environments produce different starch yield. One of the most widely used models to analyze genotype-by-environment data in order to select the best breed by environment is the additive main effects and multiplicative interaction (AMMI) model. The multiplicative terms are usually obtained by the application of the Singular Value Decomposition (SVD) to the residuals since the SVD system is capable of handling both square and rectangular data. However, performing the ordinary linear model upon which the SVD systems are evaluated on its interaction residuals are sensitive to the presence of outliers, small samples, and other contaminations. Hence, the need for a Robust AMMI that utilizes the robust residuals. Twelve cassava genotypes with codes NR070220, NR150008, NR150014, NR150025, NR150031, NR150040, NR150060, NR150085, NR150105. NR150113, NR150123, TMS30572, were experimented upon in three locations namely, Igbariam, Umudike, and Otobi in Nigeria in 2025 and the data for the cassava starch yield were obtained. The ordinary AMMI model and the Robust AMMI model selected the TMS30572 breed as the most stable and in addition, the Robust AMMI analysis slightly outperformed the ordinary AMMI in terms of reliability of the model and thus, need to be explored further for turbulent genotype-by-environment data sets. Also, the calibrated AMMI biplots capable of predicting stable cassava breeds for specific locations was introduced to make a graphical prediction of the best breed across the various locations.

Keywords: Cassava; starch yield; AMMI models; multi-environment trials SVD. biplots

1. Introduction

As Cassava is consumed by more than a billion people, the world cassava production stood at about 278 million tonnes (FAO, 2018), and FAOSTAT, (2019) recorded that Africa total production was about 170

million tons (about 56% of world production) of which, Nigeria produced about 60 million tonnes (FAOSTAT, 2019). It is comparatively more affordable in Nigeria than other staple foods and consumed often on daily basis, and sometimes more than once per day (Okoye *et al.*, 2021). The success of cassava in Africa, as a food security crop is largely because of its ability and capacity to yield well in drought-prone, marginal wastelands under poor management where other crops would fail. Despite cassava's ability to grow in marginal areas, large differential genotypic responses occur under varying environmental conditions. This phenomenon is referred to as genotype by environment interactions (GE), which is a routine occurrence in plant breeding programmes. The eastern region of Nigeria is made of diverse environmental conditions accompanied by different soil types located therein. Hence, a study of the cassava GE variations to ascertain stable breeds with high starch yield within each environment or location is inevitable.

2. Literature Review

The additive main effects and multiplicative interaction (AMMI) models remains one of the most widely used models to analyze genotype-by-environment data sets (Agyeman *et al*., 2015). In multi-location trials, Genotype-by-environment (GE) data are usually organized in two-way tables with genotypes in the rows and environments (location-year combinations) in the columns. The AMMI model applies singular value decomposition (SVD) to the residuals of a specific linear model, to decompose the GE interaction into a sum of multiplicative terms. Cox and Cox (2001) argued that SVD systems are sensitive to outliers, which in turns contaminates the data. In South Eastern part of Nigeria, inadequate plots of trial sites centered in a particular location might contribute to contaminated data sets and thus, need for a robust AMMI analysis. In addition, the traditional AMMI biplot was not calibrated, making it difficult or complex to interpret or predict. In fact, the AMMI procedure should address the following four questions: "(1) Can the target environment be divided into meaningful mega-environments so that some of the GE can be exploited or avoided? Multi-year data are essential to address this question. (2) What are the causes of GE? Data of genetic and environmental covariates are required to address this question. (3) What are the best test environments (representative and discriminating)? (4) What are the superior genotypes (both high and stable performance within a mega-environment)?" Given sufficient data, biplot analysis implemented by GE biplot, which is obtainable from the AMMI analysis can help address these questions effectively and conveniently (Yan and Tinker, 2006). The aim of this research will focus on comparing the ordinary AMMI and the Robust AMMI using the robust linear residuals for GE data on cassava starch yield in Nigeria. A second objective is to construct the calibration of the axis of the AMMI model and the robust AMMI model biplot using the approach summarized by Gower *et al.* (2011) for easy prediction. Mbe et al. (2024) evaluated 203 diverse cassava clones across three Nigerian locations and two years under optimum- and low-nitrogen conditions, using AMMI and GGE to identify stable/high-yielding genotypes and representative/discriminating testing environments. Also, Amelework et al. (2023) evaluated 11 cassava genotypes at six South African sites**,** using AMMI and the Genotype Stability Index (GSI) to identify cultivars combining high fresh-root yield and dry-matter content with stability and to identify suitable testing/mega-environments. Other works like that of Bakare et al. (2022) evaluated 36 elite cassava clones in 11 Nigerian locations over three cropping seasons**,** comparing AMMI, GGE and Finlay–Wilkinson

approaches to characterize G×E, identify stable/high-yielding clones and delineate mega-environments; AMMI-2 identified six mega-environments for fresh-root yield.

3. Materials and Methods

The research focused on twelve cassava genotypes with codes NR070220, NR150008, NR150014, NR150025, NR150031, NR150040, NR150060, NR150085, NR150105. NR150113, NR150123, TMS30572, experimented upon in three locations namely - Igbariam, Umudike, and Otobi in Nigeria in 2020. The data of the cassava starch yield were obtained after harvesting and processing. The AMMI model is given by the equation:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \sum_{i-1}^{k} \lambda_i u_{ij} v_{ij} + e_{ij} \qquad (1)$$

where Y_{ij} the measured mean yield of genotype *i* (1, 2, …, n) in environment *j* (1,2,…,*p*)

μ is the grand mean, α_i is the main effect of environment $i\left(\mu+\alpha_i\right)$, β_j is the main effect of environment $j\left(\mu+\beta_j\right)$ being the mean yield of environment j, λ_i is the singular values of the *i*th principal component (PC), u_{ij} is the eigenvector of genotype *i* for PCs, v_{ij} is the eigenvector of environment *j* for PCs and e_{ij} the residual associated with genotype *i* in environment *j* .

The AMMI model represented in (1) above is fitted in two stages. First, the main effects of the model are estimated using the additive two-way analysis of variance (ANOVA) by least squares given by $Y_{ij} = \mu + \alpha_i + \beta_j + \delta_{ij}$, with $\delta_{ij} = \sum_{i-1}^{k} \lambda_i u_{ij} v_{ij} + e_{ij}$. The second stage then involves the SVD, which is applied on the part $\delta_{ij} = \sum_{i-1}^{k} \lambda_i u_{ij} v_{ij} + e_{ij}$ that contains the δ_{ij} with the residuals e_{ij}, i.e. to the interaction, to obtain the estimates for the multiplicative terms of the AMMI model. Following the approach discussed by Rodrigues *et al*. (2015), the problem of the data contamination with outlying observations could be curbed by the use the robust regression based on the M-Huber estimator (Huber, 1981) to replace the ANOVA model.

In this approach, the AMMI biplot that will be constructed would be obtained from the ordinary least squares residuals (denoted by e_{ij}) and the robust regression based on the M-Huber estimator (denoted by e_{ij}^{*}). With n the number of genotypes, and $k \le p$, where p represents the number of environments or locations, let $\underset{n\times p}{\mathbf{X}}$ represents the residual matrix e_{ij} or e_{ij}^{*} as the case may be, the SVD decomposition of the residual matrix is given by

$$\mathbf{X} = \mathbf{U\Sigma V'} \qquad (2)$$

In (2), $\underset{n\times p}{\mathbf{U}}$ represents the ($n\times k$) orthonormal matrix consisting of the first columns of **U** and $\underset{p\times k}{\mathbf{V}}$ is a ($p\times k$) orthonormal matrix consisting of the first k columns of **V** and $\mathbf{\Sigma}$ is a ($k\times k$) diagonal matrix. For the matrix notation, the SVD given in (2) when plugged into the AMMI model in (1) will yield the matrix

$$\mathbf{Y} = \mathbf{1}_I\mathbf{1}'_J\mu + \mathbf{1}'_I\alpha_i + \mathbf{1}'_J\beta_j + \mathbf{U}^*\mathbf{\Sigma}^*\mathbf{V}^{*\prime} + \mathbf{e} \quad (3)$$

where $\mathbf{Y}$ is the two-way $(I \times J)$ table of genotypic means across locations or environments. $\mathbf{1}_I\mathbf{1}'_J\mu$ is the $(I \times J)$ matrix with the grand mean μ in all positions. $\mathbf{1}'_I\alpha_i$ is a $(I \times J)$ matrix of genotype main effects (equal rows), and $\mathbf{1}'_J\beta_j$ is a $(I \times J)$ matrix of environmental main effects (equal columns). $\mathbf{e}$ remains the vector of the residuals, e_{ij}.

Cox and Cox (2001) expressed the quality of the biplot as

$$\Psi = \frac{\sum_{k=1}^{q}\sigma_q}{\sum_{k=1}^{p}\sigma_q} = \frac{\sum(\sigma_1 + \sigma_2 + \ldots + \sigma_k)}{\sum(\sigma_1 + \sigma_2 + \ldots + \sigma_p)} \quad (4)$$

where σ are the eigen values, (see John *et al.* 2023).

To assess the stability of the biplots in 2D configuration, the algorithm will incorporate the principal component analysis biplot calibration procedure in Gower *et al., (*2011) with calibration factor for the q^{th} axis expressed as

$$\alpha = \frac{\mu^*}{\mathbf{e}'_q\mathbf{V}^*_{[k]}\mathbf{V}^{*}_{[k]}{}'\mathbf{e}_q} \quad (5)$$

where μ^* is a marker on the axis and $\mathbf{e}_q$ is the unit vector with zeros except for a one in the q^{th} position (See Gower *et al.*, 2011). In addition to the Biplot Quality measures, a major inferential selection method for the AMMI model is the use of the AMMI stability index. Let N′ be the number of significant interaction principal component axis (IPCs) that were retained in the AMMI model via F-tests); λ_n is the singular value for the IPC and correspondingly λ_n its eigen value; γ_{in} is the eigenvector value for ith genotype; δ_{jn} is the eigenvector value for the jth environment. PC1, PC2, …, PCn are the scores of 1st, 2nd, and nth IPC respectively; θ_1, θ_2, and θ_n are percentage sum of squares explained by the 1st, 2nd, and nth IPC, respectively. Based on stability and computational efficiency of the AMMI indexes, some simple indexes for the selection of the best genotype are utilized. For representativeness of the stability indexes, a possible criterion is to select indexes that capture, in at least one of the cases, the use of the eigenvalue-eigenvector pairs, the IPCs, and the percentage sum of squares explained by the 1st, 2nd, and nth IPC. Four out of these indexes, but to mention a few, are:

(i) Modified AMMI Stability Value (MASV) (Ajay et al. 2019)

$$MASV = \sqrt{\sum_{n=1}^{N^+}\left(\frac{SSIPC_n}{SSIPC_{n+1}}\right)\times\left(PC_n\right)^2 + \left(PC_{N'}\right)^2} \quad (6)$$

(ii) Sums of the Absolute Value of the IPC Scores (SIPC) (Sneller et al. 1997)

$$SIPC = \sum_{n=1}^{N^+}\left|\lambda_n^{0.5}\gamma_{in}\right| \quad (7)$$

(iii) AMMI Based Stability Parameter (ASTAB) (Rao and Prabhakaran 2005).

$$ASTAB = \sum_{n=1}^{N^+}\lambda_n\gamma_{in}^2 \quad (8)$$

(iv) Weighted average of absolute scores (WAAS) (Olivoto et al. 2019)

$$WAAS_i = \sum_{n=1}^{N^+} \left| IPCA_{ik} \times \theta_k \Big/ \sum_{k=1}^{P} \theta_k \right| \quad (9)$$

In the usage of equations (6) to (9) for genotype recommendation, a useful guide to get a desirable genotype is to seek is to seek the genotype with high mean performance as well as low stability index. Olivoto (2019) opined that WAAS is especially useful because it explicitly incorporates the **relative importance of the IPCs**, while WAASB goes a step further by combining stability with genotype mean performance. MASV is also very useful when several significant IPCs are retained.

4. Result and Discussion

Given the Cassava data set with the Starch yield y, the 12 genotype or species (GEN) with many replicates and the three environments or locations (ENV), the residual matrices is showcased in Table 1(a) and Table 1(b).

Table 1: Regression residuals

	(a) Residuals			**(b) Robust residuals**		
Genotype	igbariam	otobi	umudike	igbariam	otobi	umudike
NR070220	4.521944	1.0825	2.004444	3.909761	-0.27137	0.709264
NR150008	-0.18806	0.1225	2.754444	-0.80024	-1.23137	1.459264
NR150014	0.211944	-2.8375	-14.5556	-0.40024	-4.19137	-15.8507
NR150025	2.716944	6.483333	-3.72694	2.799865	4.55917	-4.00025
NR150031	-1.87306	3.753333	1.643056	-1.79013	1.82917	1.369749
NR150040	-8.66306	2.193333	-1.33694	-8.58013	0.26917	-1.61025
NR150060	-1.32222	-2.58139	0.628056	-1.80959	-4.00139	1.049854
NR150085	3.507778	5.968611	3.258056	3.02041	4.548615	3.679854
NR150105	1.087778	5.758611	4.548056	0.60041	4.338615	4.969854
NR150113	2.7675	-0.38639	-4.42111	0.718521	-1.11128	-4.5696
NR150123	5.8975	1.483611	1.308889	3.848521	0.75872	1.160398
TMS30572	-19.4625	-0.44639	-1.90111	-21.5115	-1.17128	-2.0496

Table 2: AMMI model ANOVA

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Proportion
ENV	2	2333.444	1166.722	33.87	7.57E-11	NA
REP(ENV)	6	75.156	12.526	0.36	0.899	NA
GEN	11	1570.73	142.7936	4.14	0.00012	NA
GEN:ENV	22	1087.761	49.4437	1.44	0.132	NA
PC1	12	953.129	79.4274	2.31	0.0159	87.6
PC2	10	134.633	13.4633	0.39	0.946	12.4
Residuals	66	2273.681	34.4497	NA	NA	NA

--- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 3: Robust AMMI model ANOVA

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Proportion
ENV	2	2333.444	1166.722	33.87	7.57E-11	NA
REP(ENV)	6	75.156	12.526	0.36	0.899	NA
GEN	11	1570.73	142.7936	4.14	0.00012	NA
GEN:ENV	22	1087.761	49.4437	1.44	0.132	NA
PC1	12	971.525	80.9605	2.35	0.014	89.3
PC2	10	160.449	16.0449	0.47	0.906	14.8
Residuals	66	2273.681	34.4497	NA	NA	NA

--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4: Comparing the AMMI and robust AMMI model Biplots

	AMMI Biplot			Robust AMMI Biplot		
	igbariam	otobi	umudike	igbariam	otobi	umudike
Predictivity	0.999	0.443	0.968	0.999	0.468	0.98
Quality	92.23%			93.85%		

Table 5: AMMI Stability Indexes

		(a) AMMI indexes				**(b) Robust AMMI indexes**			
GEN	Mean	MASV	SIPC	ASTAB	WAAS	MASV	SIPC	ASTAB	WAAS
NR070220	16.081	1.404	1.387	0.016	0.653	1.102	1.206	0.014	0.517
NR150008	21.136	2.364	2.451	0.057	1.113	3.071	2.924	0.054	1.410
NR150014	19.224	1.893	2.009	0.041	0.893	2.48	2.416	0.039	1.148
NR150025	22.200	1.572	1.374	0.011	0.696	1.265	1.229	0.010	0.585
NR150031	20.799	1.560	0.811	0.002	0.53	1.562	0.799	0.002	0.527
NR150040	14.170	4.948	2.872	0.023	1.78	4.688	2.800	0.021	1.711
NR150060	17.394	1.596	1.568	0.02	0.741	1.253	1.365	0.018	0.588
NR150085	15.823	2.195	1.288	0.005	0.794	2.073	1.253	0.004	0.761
NR150105	11.580	4.373	2.224	0.015	1.470	4.353	2.178	0.015	1.451
NR150113	25.359	1.699	0.984	0.003	0.610	1.612	0.959	0.002	0.587
NR150123	18.064	2.781	1.771	0.009	1.049	2.967	1.859	0.009	1.110
TMS30572	23.046	0.515	0.554	0.003	0.243	0.423	0.469	0.003	0.179

In Table 2 and Table 3, the full AMMI model ANOVA for the ordinary AMMI and the robust AMMI are displayed respectively. Focusing on the PC source, it is observed that while the two methods show significance for PC1 and PC2 representation of both the Ordinary AMMI and Robust AMMI interaction matrix, the PC1 and PC2 of the Robust AMMI seems to be more assuring. In addition, different species of cassava seem to be significant and so does the environmental (location) factor in both Tables 2 and 3. This can be seen as their p-levels are lower tan 5% significant level. Although the next step could have been to

orthogonalize or separate the main effects, a salient focus of this paper is to evaluate the quality of the AMMI biplot produced using the two approaches as well as calibrating the biplot in two dimensions (i.e. the first and second principal components). Across the locations, the cassava specie TMS30572 seems to be stable with high starch yield. This is showcased in Table 5 and evaluated using the MASV, SIPC, ASTAB, and WAAS indexes for both the ordinary AMMI and Robust AMMI. Again, observe that the indexes obtained from the Robust AMMI model seem to be more assuring as it shows lower significant values across all the tables (Tables 2,3,4, and 5) signifying the strength and suitability of the Robust AMMI approach in this application

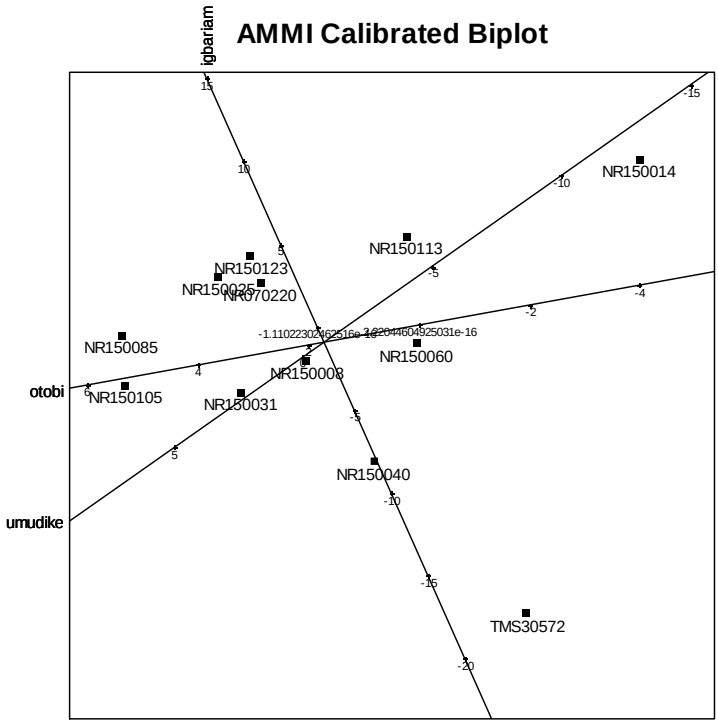


Figure 1: The AMMI calibrated biplot for the ordinary residual – PCA Quality = 0.9223

A guideline for the AMMI biplot interpretation is to note that along the mean performance, genotypes farther to the right have higher mean performance. Also, genotypes farther to the left have lower mean performance. In addition, for yield, movement to the right generally indicates better performance. Figures 1 and 2 showcase the AMMI biplots of the ordinary least squares AMMI and the robust regression AMMI procedures respectively. It is observed that the allocation of the points is optimal with the PCA reliability or quality being high at 93.85% for the robust AMMI biplot, while that of the ordinary least squares AMMI biplot is 92.23% within the 2D configuration as showcased in Table 4. The outstanding high yield of the genotypes TMS30572 and NR150113 is clearly observable on the biplots as they lie distinctively from the cluster of the other genotypes at both the lower and upper far right, indicating higher yield. This could be seen in Table 5. In addition, this may suggest more assuring predictions of the observations (environments) as the predictability for the robust regression AMMI biplot are higher or equal to that of the ordinary least squares AMMI biplot.

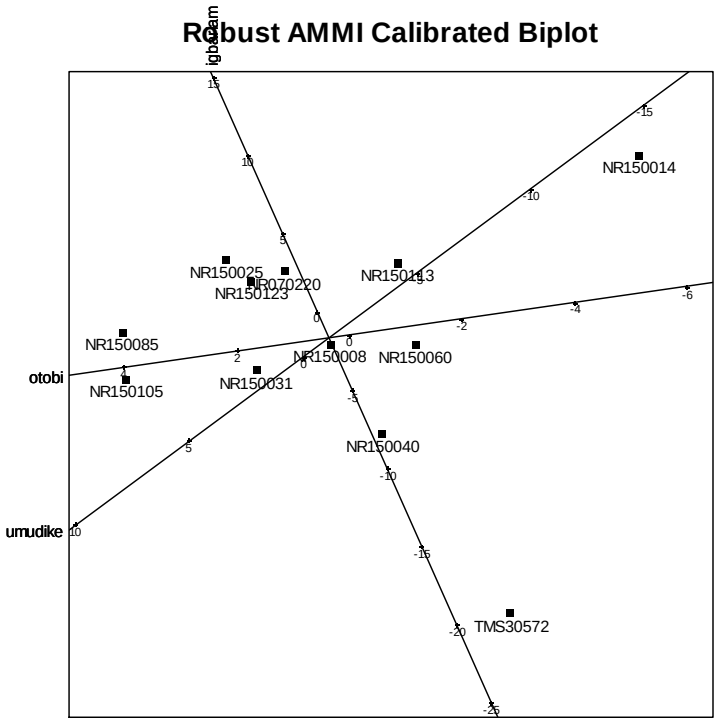


Figure 2: The AMMI calibrated biplot for robust residual – PCA Quality = 0.9385

5. Conclusion

The AMMI model biplot obtained from the both least squares residual and the robust regression residuals showcases reasonable stability of the species. However, the robust AMMI analysis biplot qualities outperformed the ordinary AMMI in terms of higher reliability or quality. This suggests that the interpretations from the robust AMMI biplot could be relied on for the investigated starch data set. In addition, this further insinuates that the genotype-by-environment data sets from the Eastern part of Nigeria was contaminated. Furthermore, a calibrated biplot, which views both the genotypes (samples) and the locations (axis) brings this research to limelight as predictions can easily be made on the 2-dimensional display as it pertains the best genotype within or across the locations and this is the apex of this research.

Acknowledgements
The authors acknowledge the sponsorship received from the Directorate of Research and Advancement, Michael Okpara University of Agriculture, Umudike, Abia State Nigeria.

References

Agyeman, A., Parkes, E. and Peprah, B. B. (2015). AMMI and GGE biplot analysis of root yield performance of cassava genotypes in the forest and coastal ecologies, International *Journal of Agricultural Policy and Research* (3), pp. 222-232.

Amelework, A. B., Bairu, M. W., Marx, R., Laing, M., & Venter, S. L. (2023). Genotype × environment interaction and stability analysis of selected cassava cultivars in South Africa. *Plants, 12*(13), 2490.

Ajay, B. C., Aravind, J., Abdul Fiyaz, R., Bera, S. K., Kumar, N., Gangadhar, K., & Kona, P. **(2018).** Modified AMMI stability index (MASI) for stability analysis. *ICAR-DGR Newsletter, 18,* 4–5.

Bakare, M. A., Kayondo, S. I., Aghogho, C. I., Wolfe, M. D., Parkes, E. Y., Kulakow, P., Egesi, C., Rabbi, I. Y., & Jannink, J.-L. (2022). Exploring genotype by environment interaction on cassava yield and yield related traits using classical statistical methods. *PLOS ONE, 17*(7), e0268189.

Cox, T.F. and Cox, M. A.A. (2001). *Multidimensional Scaling, 2nd edition,* New York: Chapman & Hall/CRC. Boca Raton.

FAO & IFAD, The World Cassava Economy: Facts, Trends and Outlook, Food & Agriculture Organisation, Rome, 2020.

FAOSTAT F. 2019. Food and Agriculture Data.http://www.fao.org/faostat/en/#data/

FAOSTAT F. 2018. Food and Agriculture Data.http://www.fao.org/faostat/en/#data/

Gower, J.C., Lubbe, S. and Le Roux, N. J., (2011). *Understanding Biplots*, Wiley Chichester.

Huber,P.J. (1981) *Robust Statistics.* Wiley, New York.

John, C., Ekpenyong, E.J. Omekara, C.O., and Nworu, C.C. (2023), A New Robust Bootstrapped Singular Value Decomposition Algorithm Using the Sample Myriad Estimate. *Int J Data Sci Anal* (Springer).

Mbe, J. O., Dzidzienyo, D., Abah, S. P., Njoku, D. N., Aghogho, C. I., Onyeka, J., Tongoona, P., & Egesi, C. (2024). Genotype by environment interaction effect and fresh root yield stability of cassava genotypes under contrasting nitrogen regimes. *Scientific Reports, 14*, 207-209.

Okoye, F.U., Okoye, A.C. & Umeh, S.I. (2021) Consumption behaviour analyses of cassava products among rural household in Ebonyi State, Nigeria. Agro-Science 20(2), 14 – 19.

Olivoto, T., Lúcio, A. D. C., da Silva, J. A. G., Marchioro, V. S., de Souza, V. Q., & Jost, E. (2019). Mean performance and stability in multi-environment trials I: Combining features of AMMI and BLUP techniques. *Agronomy Journal, 111*(6), 2949–2960.

Rodrigues,P.C. Monteiro, A. and Lourenco, A. M. (2015), A robust AMMI model for the analysis, of genotype-by-environment data, *Bioinformatics*, 32(1), 2016, 58–66.