

Research Application Summary

Investigation of the Sweetpotato Trial Network for Efficient and Cost-effective Variety Development in Uganda

Chelangat, D.M.,^{1,2} Anyanga, M.O.¹, Musana, P.¹, Oloka, B.³, Alajo, A.¹, Osaru, F.¹, Aboyo, R.¹, Asiimwe, J.¹, Katungisa, A.¹, Adero, J.¹, Letia, S.¹, Ssali, R.T.⁴, De Boeck, B.⁵, da Silva Pereira, G.⁶, Campos, H.⁵, Yencho, G.C.³ & Yada, B.¹

¹National Crops Resources Research Institute (NaCRRI), Namulonge, P.O. Box 7084, Kampala, Uganda

²Makerere University Regional Centre for Crop Improvement (MaRCCI), P.O. Box 7062, Kampala, Uganda

³Department of Horticultural Science, North Carolina State University, 214 Kilgore Hall, Box 7609, Raleigh, NC 27695-7609, USA

⁴International Potato Center (CIP), Kampala, Uganda

⁵International Potato Center (CIP), Lima, Peru

⁶Department of Agronomy, Federal University of Viçosa, Brazil

Corresponding author: yadabenard21@gmail.com

Abstract

Late-stage evaluation of sweetpotato breeding trials is routinely conducted at multiple locations, years, or seasons to provide a better understanding of the genotype-environment interaction and test for genotype stability. In addition, increased testing serves to reduce the experimental error and therefore provide a better estimate of the actual genetic component. This increased testing comes with a huge financial burden for the breeding program. The aim of this study was to determine the variance components of multi-environment Advanced yield trials (AYT) and identify the most appropriate sweetpotato breeding trial network. The optimization of the trial network in terms of number of replications, locations, and seasons would ensure efficient resource allocation to further streamline the breeding program strategy. The storage root yield data obtained from 13 AYT of the sweetpotato breeding program in Uganda was used for this study. In each trial, a total of 23 genotypes were established in three replications using Randomised complete block design (RCBD) and evaluated for three seasons. The data was analysed using mixed models to estimate the variance components used for the simulation of the trial network. The genotype effect (0.21) provided the highest proportion of variance, followed by genotype by season (0.14). Based on the simulation results, the best trial network for about the same number of genotypes is 4 locations, 2 seasons, and 3 replications. This network minimizes the standard error of the difference and, at the same time, maintains the one-year (2-season) evaluation for quicker variety development. This article has also provided a guideline for a trial network for earlier and later-stage gates with genotypes in the range of 06–120 evaluated in the same target population of environments.

Keywords: multi-environment trials, resource optimisation, stage gate, sweet potato breeding, trial network, Uganda

Résumé

L'évaluation à un stade avancé des essais de sélection de la patate douce est systématiquement réalisée sur plusieurs sites, années ou saisons pour mieux comprendre l'interaction génotype-environnement et tester la stabilité des génotypes. En outre, l'augmentation des tests permet de réduire l'erreur expérimentale et de fournir une meilleure estimation de la composante génétique réelle. Cette augmentation des tests représente une lourde

charge financière pour le programme de sélection. L'objectif de cette étude était de déterminer les composantes de la variance des essais de rendement avancés (AYT) en environnement multiple et d'identifier le réseau d'essais de sélection de la patate douce le plus approprié. L'optimisation du réseau d'essais en termes de nombre de répétitions, de sites et de saisons permettrait d'assurer une allocation efficace des ressources pour rationaliser davantage la stratégie du programme de sélection. Les données de rendement des racines de stockage obtenues à partir de 13 AYT du programme de sélection de la patate douce en Ouganda ont été utilisées pour cette étude. Dans chaque essai, un total de 23 génotypes ont été établis en trois répétitions selon un dispositif en blocs complets randomisés (RCBD) et évalués sur trois saisons. Les données ont été analysées à l'aide de modèles mixtes pour estimer les composantes de la variance utilisées pour la simulation du réseau d'essais. L'effet génotype (0,21) a fourni la plus grande proportion de variance, suivi par le génotype par saison (0,14). D'après les résultats de la simulation, le meilleur réseau d'essais pour un nombre similaire de génotypes est de 4 sites, 2 saisons et 3 répétitions. Ce réseau minimise l'erreur standard de la différence tout en maintenant l'évaluation d'un an (2 saisons) pour un développement plus rapide des variétés. Cet article a également fourni des lignes directrices pour un réseau d'essais pour les phases précoces et avancées, avec des génotypes allant de 06 à 120 évalués dans la même population cible d'environnements.

Mots-clés: essais en environnement multiple, optimisation des ressources, porte d'étape, sélection de la patate douce, réseau d'essais, Ouganda

Introduction

Multi-environment trials of sweet potato breeding clones along the stage gates aim to get accurate results for key traits that inform the breeding objectives. Attaining high levels of precision requires the control of extraneous experimental error using modern experimental designs that capture spatial trends to reduce the unexplained error (Bernardeli *et al.*, 2021). In addition, when using traditional routine experimental designs, increasing testing locations, replications, seasons, and years reduces the unexplained variability. The accuracy of plant breeding trials is significantly influenced by the number of observations achieved through the replications involved. The relationship between replication and experimental error is intertwined, as the computation of experimental error requires replications (Brereton, 2018). However, it's important to consider the costs associated with each additional replication. Therefore, it is necessary to find an optimal balance in the number of replications for crop variety trials. (Yan, 2021; de Resende and Alves, 2022).

The solution to identifying widely adapted cultivars within a mega-environment is to "test widely" to accommodate unrepeatable genotype by environment (GE) variation. Increasing the number of test locations in plant breeding trials can lead to better adaptation to different environments, increased testing capacity, increased selection intensity, and reduced variability (Yan, 2021). However, each additional location, and year involves considerable cost, calling for the need to test adequately. In sweetpotato, for advanced trials like Advanced Yield trials (AYTs), it is recommended that four or more locations be used for replicated multi-environment testing (Grüneberg *et al.*, 2019). However, the accuracy of these trials still calls for experimental improvement because of the high coefficient of variation of traits like yield (Grüneberg *et al.*, 2010). In addition, trait heritability and the genotypic correlation across environments affect the optimum testing required in breeding experiments (de Resende and Alves, 2022). Therefore, the objective of this study was to determine the minimum number of replications, testing sites,

and seasons for sweet potato late-stage multi-environment trials to optimize resource use and trial accuracy for efficient genotype selection.

Materials and Methods

Advanced yield trials were conducted for three seasons: 2016A, 2016B, and 2017A. In 2016, trials were conducted in NaCRRI, NaSARRI, and AbiZARDI only due to limitations in planting material. However, during 2016B and 2017A, trials were established at five on-station sites: NaCRRI, BuZARDI, RwebitabaZARDI, NaSARRI, and AbiZARDI. A total of 23 clones were evaluated during the AYT stage. The experiments were established in a randomized complete block design (RCBD) and replicated three times. In all replicates, the plot size was 3 m x 4 m. This plot area consisted of five 3 m x 1 m row plots with a total of 50 plants.

Storage root yield data were computed from the commercial and non-commercial root data parameters collected. Storage root yield was a factor of both the marketable/ commercial root weight and commercial root weight. Total storage root yield in tons per hectare (RYTHA) was calculated using the formula;

$$RYTHA = \frac{\text{Commercial root weight} + \text{non commercial root weight}}{\text{Net plot area (m}^2\text{)}} \times 10$$

Storage root yield data were analyzed using R version 4.2.3 (R Core Team, 2023). The “asreml” package was used for analysis across locations, replications and seasons (Butler, 2022).

Results and Discussion

The analysis of the data revealed that the genotype contributed greatly to the variation in the AYT_s. This was followed up by the genotype by season and, lastly, the genotype by location (Table 1). These results reveal an acceptable level of precision in the experimentation. However, the unexplained experimental error was way higher than the explained variation, suggesting the need for the breeding program to use row-column designs to capture spatial trends and improve accuracy (Shah and Sinha, 1996; Piepho and Williams, 2016).

Source	Variance estimate
VG	0.209
V GL	0.034
V GS	0.135
V GLS	0.095
residual	1.565

Based on the variance obtained above, the optimum number of replications and years, which in this case are the seasons and locations, were obtained as presented in Figure 1 and Table 2. The location trial network showed that it required more than 20 locations to obtain 100% precision, which is impractical. In addition, five seasons were required to obtain a lower mean variance compared to three years.

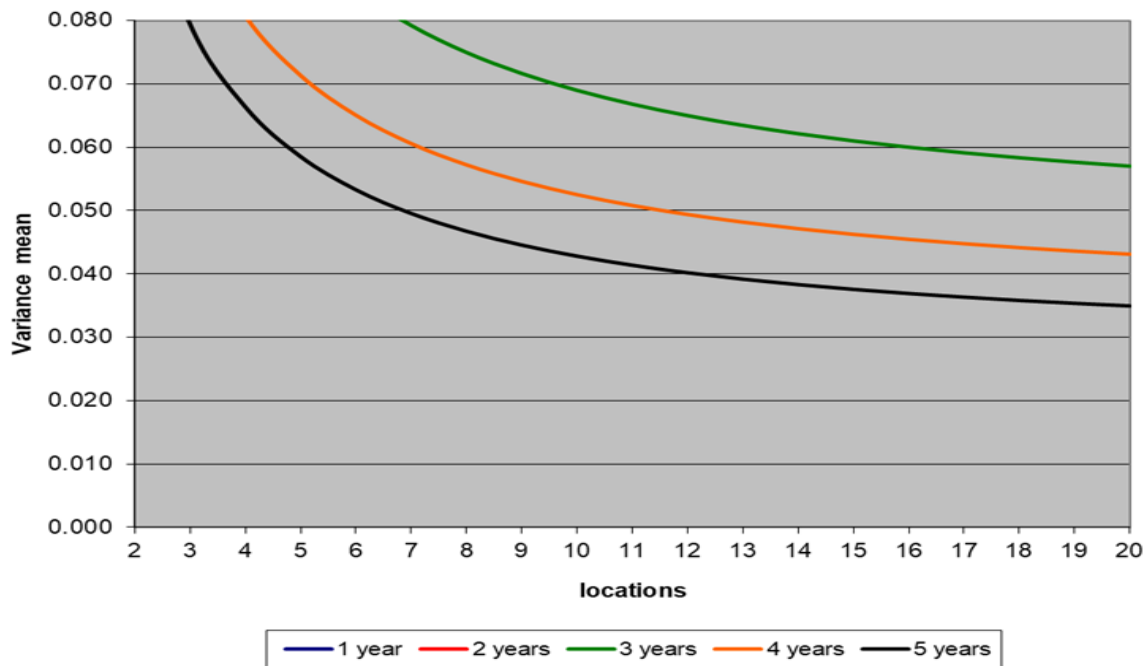


Figure 1. Trial network alternatives based on variance mean for yield

Table 2. Optimum replications and years of sweet potato yield trials

nLoc	nYear	nRep	VarMean	sed	LSD	Size
1	2	2	1.0465	1.4467	2.8934	4
2	2	2	0.5908	1.0870	2.1739	8
3	2	2	0.4388	0.9368	1.8737	12
4	2	2	0.3629	0.8519	1.7038	16
5	2	2	0.3173	0.7966	1.5932	20
6	2	2	0.2869	0.7575	1.5150	24
7	2	2	0.2652	0.7283	1.4566	28
8	2	2	0.2489	0.7056	1.4112	32
9	2	2	0.2363	0.6874	1.3749	36
10	2	2	0.2262	0.6725	1.3451	40
11	2	2	0.2179	0.6601	1.3202	44
12	2	2	0.2110	0.6496	1.2991	48
13	2	2	0.2051	0.6405	1.2810	52
14	2	2	0.2001	0.6326	1.2652	56
15	2	2	0.1958	0.6257	1.2515	60
16	2	2	0.1920	0.6196	1.2393	64
17	2	2	0.1886	0.6142	1.2284	68
18	2	2	0.1856	0.6093	1.2187	72
19	2	2	0.1830	0.6049	1.2099	76
20	2	2	0.1806	0.6010	1.2019	80
1	3	2	0.5403	1.0395	2.0789	6
2	3	2	0.3039	0.7796	1.5592	12
3	3	2	0.2251	0.6709	1.3419	18
4	3	2	0.1857	0.6094	1.2188	24
5	3	2	0.1621	0.5693	1.1386	30
6	3	2	0.1463	0.5409	1.0818	36
7	3	2	0.1350	0.5197	1.0394	42
8	3	2	0.1266	0.5032	1.0064	48
9	3	2	0.1200	0.4900	0.9799	54
10	3	2	0.1148	0.4791	0.9582	60
11	3	2	0.1105	0.4701	0.9401	66
12	3	2	0.1069	0.4624	0.9248	72
13	3	2	0.1039	0.4558	0.9115	78
14	3	2	0.1013	0.4500	0.9001	84
15	3	2	0.0990	0.4450	0.8900	90
16	3	2	0.0970	0.4406	0.8811	96
17	3	2	0.0953	0.4366	0.8732	102
18	3	2	0.0938	0.4330	0.8661	108
19	3	2	0.0924	0.4298	0.8597	114
20	3	2	0.0911	0.4269	0.8539	120

Several simulated trial networks were explored for this study. These included networks with 1–20 locations, 1–5 replications, and 1–5 seasons. Presented herein are the 2-season and 2-3 replication network results in Table 2. This was because of the variety release requirements in Uganda to have at least two seasons of testing for later stages. In addition, the use of two replications has been reported to suffice for trialing at later stages (Yan, 2021).

From this trial network, for around the same number of genotypes evaluated in AYT, 4-5 locations and 3 replications provided a lower sed value (0.56-0.54) compared to the 0.72-0.70 for the 2-replication network. Four (04) locations and three (3) replication networks were considered optimum because they account for resources, accuracy, and time. This result differs from the two replication trial networks recommended for MET testing in oats (Yan, 2021).

In statistics, the lower the sed value, the better the accuracy; however, even with 20 locations, 0.41 sed was achieved, which agrees with the fact that trait heritability greatly influences the breeding trial networks. This implies that, as a breeding program, there is a need to explore the use of spatial modeling for increased heritability and accuracy (Bernardeli *et al.*, 2021). The increased testing could be achieved through sparse testing (Jarquin *et al.*, 2020; Atanda *et al.*, 2022; Montesinos-López *et al.*, 2023) while accounting for spatial trends.

Conclusion

The optimal number of replications for late-stage multi- environment sweetpotato breeding trial evaluation can vary depending on several factors. Key among them is the variance of the mean, which is related to phenotyping accuracy. Current research has shown that to maximize accuracy while accounting for cost and time, a 4-location, 3-replicate trial network can suffice. It is imperative to acknowledge that these numerical values are not inherently fixed and may exhibit variability contingent upon the circumstances surrounding a given trial.

Acknowledgement

We thank the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) for the opportunity to share our research findings and Makerere University Regional Centre for Crop Improvement (MaRCCI) for the 2023 RUFORUM AGM sponsorship. The research was funded by the Bill & Melinda Gates Foundation (BMGF) through the PEARL I: Biochemistry-Based Selection and Development of Nutrient Rich, Weevil Resistant Sweetpotato Varieties in Uganda Sweetpotato project (Contract ID: OPP1112515). This paper is a contribution to the 19th RUFORUM Annual General Meeting and Scientific Conference held 28th October -02 November 2023 in Yaounde, Cameroon

References

- Atanda, S.A., Govindan, V., Singh, R., Robbins, K.R., Crossa, J. and Bentley, A.R. 2022. Sparse testing using genomic prediction improves selection for breeding targets in elite spring wheat. *Theoretical and Applied Genetics* 135 (6):1939–1950. <https://doi.org/10.1007/S00122-022-04085-0/FIGURES/5>
- Bernardeli, A., Rocha, J.R.A.S. de C., Borém, A., Lorenzoni, R., Aguiar, R., Silva, J.N.B., Bueno, R. D., Alves, R. S., Jarquin, D., Ribeiro, C. and Lamas Costa, M.D.B. 2021. Modeling spatial trends and enhancing genetic selection: An approach to soybean seed composition breeding. *Crop Science* 61 (2):976–988. <https://doi.org/10.1002/CSC2.20364>

- Brereton, R.G. 2018. Degrees-of-freedom, errors, and replicates. *Journal of Chemometrics* 32 (6):e3016. <https://doi.org/10.1002/CEM.3016>
- Butler, D. 2018. Package “asreml”. Fits the Linear Mixed Model. www.vsni.co.uk
- de Resende, M.D.V. and Alves, R.S. 2022. Statistical significance, selection accuracy, and experimental precision in plant breeding. *Crop Breeding and Applied Biotechnology* 22 (3):e42712238. <https://doi.org/10.1590/1984-70332022V22N3A31>
- Grüneberg, W. J., Eyzaguirre, R., Díaz, F., Boeck, B. de, Espinoza, J., Swanckaert, J., Dapaah, H., Andrade, M.I., Makunde, G.S., Agili, S., Ndingo-Chipungu, F.P., Attaluri, S., Kapinga, R., Nguyen, T., Kaiyung, X., Tjintokohadi, K., Ssali, R. T., Carey, T., Low, J. W. and Mwanga, R.O.M. 2019. Procedures for the evaluation of sweetpotato trials. Manual. Procedures for the Evaluation of Sweetpotato Trials. <https://doi.org/10.4160/9789290605225>
- Jarquín, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J. W. R., Pazaran, G. C., Burgueño, J., Pacheco, A., Grondona, M., Wimmer, V. and Prasanna, B.M. 2020. Genomic prediction enhanced sparse testing for multi-environment trials. *G3 Genes|Genomes|Genetics* 10 (8):2725–2739. <https://doi.org/10.1534/G3.120.401349>
- Montesinos-López, O. A., Saint Pierre, C., Gezan, S. A., Bentley, A. R., Mosqueda-González, B. A., Montesinos-López, A., van Eeuwijk, F., Beyene, Y., Gowda, M., Gardner, K., Gerard, G. S., Crespo-Herrera, L. and Crossa, J. 2023. Optimizing sparse testing for genomic prediction of plant breeding crops. *Genes* 14 (4):927. <https://doi.org/10.3390/GENES14040927/S1>
- Piepho, H.P. and Williams E.R. 2016. Augmented row-column designs for a small number of checks. *Agronomy Journal* 108 (6):2256–2262. <https://doi.org/10.2134/agronj2016.06.0325>
- R Core Team. 2023. R: The R Project for Statistical Computing. In R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- Shah, K.R. and Sinha, B.K. 1996. 24 Row-column designs. *Handbook of Statistics* 13:903–937. [https://doi.org/10.1016/S0169-7161\(96\)13026-1](https://doi.org/10.1016/S0169-7161(96)13026-1)
- Yan, W. 2021. Estimation of the optimal number of replicates in crop variety trials. *Frontiers in Plant Science* 11:590762. <https://doi.org/10.3389/FPLS.2020.590762/BIBTEX>