

Evaluation of Variance Normalized ANOVA-Simultaneous Component Analysis (VN-ASCA) through Controlled Simulations

Harriet Achiaa Dwamena^{1,2,*}, N. K. Frempong², A. O. Adebajji²

¹Biometrics Unit, Council for Scientific and Industrial Research-Crops Research Institute, Kumasi, Ghana

²Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Received September 1, 2025; Revised December 29, 2025; Accepted January 28, 2026

Cite This Paper in the Following Citation Styles

(a): [1] Harriet Achiaa Dwamena, N. K. Frempong, A. O. Adebajji, "Evaluation of Variance Normalized ANOVA-Simultaneous Component Analysis (VN-ASCA) through Controlled Simulations," *Mathematics and Statistics*, Vol.14, No.1, pp. 106-119, 2026. DOI: 10.13189/ms.2026.140110.

(b): Harriet Achiaa Dwamena, N. K. Frempong, A. O. Adebajji (2026). Evaluation of Variance Normalized ANOVA-Simultaneous Component Analysis (VN-ASCA) through Controlled Simulations. *Mathematics and Statistics*, 14(1), 106-119. DOI: 10.13189/ms.2026.140110.

Copyright ©2026 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract ANOVA-Simultaneous Component Analysis (ASCA) is a powerful approach for analyzing multi-factorial experimental designs with multivariate responses. However, standard ASCA methods assume homogeneous variance across experimental groups, an assumption often violated in real-world data, particularly in biological and chemical studies. This paper introduces Variance Normalized ASCA (VN-ASCA), a novel extension that addresses heterogeneous variance through a weighted least squares framework. We provide a comprehensive mathematical derivation of the VN-ASCA methodology, including its weight calculation strategies, permutation-based significance testing, and variance partitioning approach. The method incorporates robust options for weight estimation, regularization for numerical stability, and automatic effect selection for enhanced interpretability. Practical implementation guidelines for parameter tuning are provided, and the method's performance is validated through both controlled simulations and application to real agricultural data from maize trials in Ghana. The experimental results demonstrate that VN-ASCA provides improved prediction accuracy in the presence of heterogeneous variance while maintaining interpretability. The application to multi-environment maize trials in Ghana (50 plots, 100 traits, 4 locations) demonstrates practical utility, with adaptive VN-ASCA showing measurable improvements in root mean squared error and substantially stronger statistical evidence for environmental effects. These findings suggest that accounting for variance heterogeneity can enhance the detection and interpretation of factorial effects in complex multivariate datasets.

Keywords ANOVA-Simultaneous Component Analysis (ASCA), Heteroscedasticity, Variance Normalized ASCA (VN-ASCA), Least Squares, Multivariate

1 Introduction

Multivariate analysis of designed experiments plays a key role in different fields of science, such as metabolomics, genomics, chemometrics, and environmental science. In such experiments, there are many factors (treatments, time points, etc.) and many response variables that are simultaneously measured at the same time [1, 2, 3, 4]. ANOVA-Simultaneous Component Analysis (ASCA) has become an effective tool to analyze these types of complex data by integrating the benefits of both Analysis of Variance (ANOVA) to partition the effects and Principal Component Analysis (PCA) to reduce the dimensions and visualize the data [2, 3, 5, 6].

Nevertheless, there is a major issue with the fact that the standard ASCA uses ordinary least squares (OLS) estimation, which assumes the condition of homoscedasticity, that the variances of all the experimental groups are equal [2]. In practice, heteroscedasticity is pervasive, stemming from time-dependent variability, treatment-induced effects, batch artifacts, or concentration-dependent measurement precision [7]. Recent advances in variable selection for high-dimensional data have demonstrated the importance of accounting for variance structure in ANOVA-based methods [8]. This violation leads to biased effect estimates, inflated Type I error rates, and mathematically inconsistent variance partitioning, thereby severely undermining the reliability of statistical conclusions [2, 9].

A further complication is the prevalence of unbalanced experimental designs, where unequal group sizes result from

missing data or logistical constraints [10]. The standard ASCA, designed for balanced data, does not adequately account for the systematic effects of variance heterogeneity in such scenarios, resulting in overestimated factor contributions and diminished statistical power [11]. While extensions like ASCA+ for unbalanced data, PE-ASCA for heterogeneous data, and Robust ASCA for outlier resistance address specific facets of this problem [10, 12], a significant methodological gap remains. No existing approach offers a holistic solution that simultaneously resolves the intertwined challenges of heteroscedasticity and unbalanced designs. Consequently, there is a pressing need for a statistically rigorous, flexible, and computationally efficient extension of the ASCA framework [10].

To address the limitations of existing ASCA-based approaches in handling heterogeneity and unbalanced designs, this research proposes the integration of variance-based weighting schemes into the ASCA framework. Specifically, four normalization strategies are implemented: group-wise normalization, feature-wise normalization, adaptive normalization, and robust normalization to calculate weights that reflect the underlying variance structure [1, 13]. These weights are incorporated into a weighted least squares (WLS) estimation procedure within the ASCA framework, resulting in the development of the VN-ASCA model [14, 15]. The performance of VN-ASCA is evaluated through simulation studies and real data applications to test its robustness under controlled variance heterogeneity and sample size imbalance.

2 Background

2.1 Standard ASCA Model

The standard ASCA approach combines ANOVA with PCA to analyze designed experimental data with multiple factors. For a dataset $\mathbf{X} \in \mathbb{R}^{n \times p}$ with n observations and p variables, the ASCA model partitions the data according to the experimental design:

$$\mathbf{X} = \mathbf{1}\boldsymbol{\mu}^T + \sum_{f=1}^F \mathbf{X}_f + \mathbf{X}_{\text{res}} \quad (1)$$

where:

- $\mathbf{1}\boldsymbol{\mu}^T$ represents the overall mean ($\boldsymbol{\mu}$ is the mean vector)
- \mathbf{X}_f represents the effect matrix for factor f
- \mathbf{X}_{res} is the residual matrix
- F is the total number of factor effects (including interactions)

For a two-factor experiment with factors A and B, the model becomes:

$$\mathbf{X} = \mathbf{1}\boldsymbol{\mu}^T + \mathbf{X}_A + \mathbf{X}_B + \mathbf{X}_{AB} + \mathbf{X}_{\text{res}} \quad (2)$$

where \mathbf{X}_A , \mathbf{X}_B , and \mathbf{X}_{AB} represent the main effects of factors A and B and their interaction, respectively.

In standard ASCA, each effect matrix is estimated using ordinary least squares (OLS) regression. For a given factor with design matrix \mathbf{Z}_f , the effect coefficients $\boldsymbol{\beta}_f$ are estimated as:

$$\hat{\boldsymbol{\beta}}_f = (\mathbf{Z}_f^T \mathbf{Z}_f)^{-1} \mathbf{Z}_f^T \mathbf{X}_c \quad (3)$$

where \mathbf{X}_c is the centered data matrix. The effect matrix is then calculated as:

$$\mathbf{X}_f = \mathbf{Z}_f \hat{\boldsymbol{\beta}}_f \quad (4)$$

After partitioning the data, PCA is applied to each effect matrix to reduce dimensions and visualize the structure of the effects.

3 Variance Normalized ASCA

3.1 Problem Formulation

VN-ASCA introduces a per-effect variance scaling step by standardizing each variable within effect matrices before PCA to equalize contributions and mitigate the dominance of high-variance variables. This normalization enhances interpretability and balances variable influence, making VN-ASCA particularly well-suited for complex designs involving multiple factors or interactions. VN-ASCA addresses the heterogeneity problem by incorporating observation-specific weights based on local variance estimation. The key insight is to downweight observations with high variance and upweight those with low variance during effect estimation.

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a data matrix with n observations and p variables, generated from an experimental design with multiple factors. The true data-generating process can be represented as:

$$\mathbf{X} = \mathbf{1}\boldsymbol{\mu}^T + \sum_{f=1}^F \mathbf{X}_f^{\text{true}} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{g(i)}) \quad (5)$$

where:

- $\mathbf{1}\boldsymbol{\mu}^T$ is the grand-mean matrix
- $\mathbf{X}_f^{\text{true}}$ is the true effect matrix for factor f (including interactions)
- $\boldsymbol{\varepsilon}$ is the residual matrix with heteroscedastic structure

$$\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}_{g(i)} \in \mathbb{R}^{p \times p} \quad (6)$$

where $\boldsymbol{\Sigma}_{g(i)}$ represents the covariance matrix for the group that observation i belongs to.

3.2 Model and Dimensions

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be the data matrix with n observations and p variables. The data-generating model is written as

$$\mathbf{X} = \mathbf{1}\boldsymbol{\mu}^T + \sum_{f=1}^F \mathbf{X}_f^{\text{true}} + \boldsymbol{\varepsilon},$$

Table 1. Dimensions and interpretation of terms in the data-generating model

Term	Dimension and Meaning
\mathbf{X}	$\in \mathbb{R}^{n \times p}$: data matrix (rows = observations, columns = variables).
$\mathbf{1}$	$\in \mathbb{R}^{n \times 1}$: column vector of ones (replicated over observations).
$\boldsymbol{\mu}$	$\in \mathbb{R}^{p \times 1}$: overall (grand) mean vector for the p variables; hence $\boldsymbol{\mu}^\top \in \mathbb{R}^{1 \times p}$.
$\mathbf{1} \boldsymbol{\mu}^\top$	$\in \mathbb{R}^{n \times p}$: mean matrix (each row equals the grand-mean row).
$\mathbf{X}_f^{\text{true}}$	$\in \mathbb{R}^{n \times p}$: effect matrix for factor f (if modeled via design matrices, $\mathbf{X}_f^{\text{true}} = \mathbf{Z}_f \mathbf{B}_f$ with $\mathbf{Z}_f \in \mathbb{R}^{n \times k_f}$ and $\mathbf{B}_f \in \mathbb{R}^{k_f \times p}$).
$\boldsymbol{\varepsilon}$	$\in \mathbb{R}^{n \times p}$: residual matrix (error), possibly heteroscedastic with $\text{Var}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}_{g(i)} \in \mathbb{R}^{p \times p}$.

where each term has the following dimensions and interpretation:

The effect matrix for factor f is expressed as $\mathbf{X}_f^{\text{true}} = \mathbf{Z}_f \mathbf{B}_f$ with $\mathbf{B}_f \in \mathbb{R}^{k_f \times p}$. The complete model becomes:

$$\mathbf{X} = \mathbf{1} \boldsymbol{\mu}^\top + \sum_{f=1}^F \mathbf{Z}_f \mathbf{B}_f + \boldsymbol{\varepsilon}$$

3.3 Weight Calculation Strategies

VN-ASCA supports multiple strategies for calculating the weights, each suited to different types of heterogeneity patterns:

3.3.1 Group-wise Normalization

Group-wise normalization is designed for experiments with group-specific variance differences. Here, weights are computed as the inverse of the within-group standard deviation for each variable:

$$w_{ij} = \frac{1}{\sigma_{g(i)j}} \quad (7)$$

where $\sigma_{g(i)j}$ is the standard deviation of variable j in the group g to which observation i belongs.

For robust estimation, we use the median absolute deviation (MAD):

$$\sigma_{g(i)j} = \text{MAD}_{g(i)j} \times 1.4826 \quad (8)$$

where 1.4826 comes from the inverse of the 75th percentile of the standard normal distribution:

$$1.4826 = \frac{1}{\Phi^{-1}(0.75)}, \quad \text{where } \Phi^{-1}(0.75) \approx 0.6745.$$

It is the scale factor that makes MAD a consistent estimator of the standard deviation for normally distributed data and is widely used in robust statistics [16, 17].

3.3.2 Feature-wise Normalization

ASCA is sensitive to variable variance, where high-variance variables may affect the results of analysis. This sensitivity

has some essential issues: Variables measured on a larger scale (e.g., grain yield in kg/ha) may overshadow those on a smaller scale (e.g., harvest index as a ratio), when the subset of smaller measured variables actually carries biologically relevant information. Feature-wise normalization ensures that all variables contribute equally to the ASCA decomposition so that analysis can give biologically significant patterns across the entire variable space.

Feature-wise normalization technique accounts for differences in scale between variables by using weights which are inversely proportional to the overall feature standard deviation:

$$w_{ij} = \frac{1}{\sigma_{\cdot j}} \quad (9)$$

where $\sigma_{\cdot j}$ is the standard deviation of variable j across all observations.

3.3.3 Adaptive Normalization

This is designed to handle unbalanced designs and small group sizes, which often lead to unstable variance estimates. This approach combines group-wise and feature-wise normalization using a shrinkage parameter:

$$w_{ij} = \frac{1}{\sigma_{g(i)j}^{s_i} \times \sigma_{\cdot j}^{1-s_i}} \quad (10)$$

where s_i is a shrinkage parameter that adapts to the group size:

$$s_i = \min\left(\frac{n_{g(i)}}{n_0}, 1\right) \quad (11)$$

with

- $n_{g(i)}$ as the number of observations in the group
- n_0 as the threshold parameter (typically $n_0 = 10$)

This ensures:

- Small groups ($n_g < n_0$): $s_i < 1$, so the estimate relies more on global variance ($\sigma_{\cdot j}$) to avoid instability.
- Large groups ($n_g \geq n_0$): $s_i = 1$ so the estimate uses the group-specific variance.

The weighted variance estimate is expressed as

$$\hat{\sigma}_{g,j}^2 = s_g \hat{\sigma}_{g,j}^2 + (1 - s_g) \hat{\sigma}_{g,j}^2$$

The approach is grounded in empirical Bayes and James-Stein shrinkage principles, which are widely recognized for improving variance estimation in high-dimensional and unbalanced data [18, 19].

3.3.4 Robust Normalization

This approach uses robust influence functions to down-weight potential outliers:

$$w_{ij} = \min \left(1, \frac{c}{|z_{ij}|} \right) \times \frac{1}{\text{MAD}_j} \quad (12)$$

where z_{ij} is the robust z-score of observation i for variable j :

$$z_{ij} = \frac{x_{ij} - \text{median}_j}{\text{MAD}_j} \quad (13)$$

and c is a tuning constant (typically $c = 1.345$ for 95% efficiency). The constant $c = 1.345$ is selected because it reaches about 95% efficiency for the Huber M-estimator when normal errors are assumed. This choice balances robustness against outliers with statistical efficiency [16, 20].

3.4 Weight Matrix \mathbf{W}_f Calculation

The weight matrix $\mathbf{W}_f \in \mathbb{R}^{n \times n}$ is a **diagonal matrix** where each diagonal element w_{ii} represents the weight for observation i .

3.4.1 Matrix Structure

$$\mathbf{W}_f = \text{diag}(w_{11}, w_{22}, \dots, w_{nn}) \quad (14)$$

where w_{ii} is the weight for observation i (only diagonal elements are non-zero).

3.4.2 Weighted Effect Estimation

For a given experimental design with factors, we first define a weight matrix $\mathbf{W} \in \mathbb{R}^{n \times p}$ where each element w_{ij} represents the weight for observation i and variable j .

The weighted ASCA model retains the same structure:

$$\mathbf{X} = \mathbf{1}\boldsymbol{\mu}^T + \sum_{f=1}^F \mathbf{X}_f + \mathbf{X}_{\text{res}} \quad (15)$$

But now, the effect matrices are estimated using weighted least squares regression.

For a given factor effect f with design matrix \mathbf{Z}_f , we estimate the effect coefficients $\boldsymbol{\beta}_f$ as:

$$\hat{\boldsymbol{\beta}}_f^{WLS} = (\mathbf{Z}_f^T \mathbf{W}_f \mathbf{Z}_f)^{-1} \mathbf{Z}_f^T \mathbf{W}_f \mathbf{X}_c \quad (16)$$

where:

- \mathbf{X}_c is the centered data matrix

- \mathbf{W}_f is a diagonal matrix with the appropriate weights for factor f

The effect matrix is then calculated as:

$$\mathbf{X}_f = \mathbf{Z}_f \hat{\boldsymbol{\beta}}_f \quad (17)$$

3.5 Handling Unbalanced Designs in VN-ASCA

3.5.1 Motivation for Type III Sum of Squares

Unbalanced experimental designs are common in practice, particularly in biological and agricultural experiments where missing observations, dropouts, or unequal observations occur [21, 22]. Traditional ASCA methods can produce misleading results in such scenarios because:

1. In unbalanced designs, main effects and interactions are not orthogonal, leading to confounded estimates [23]
2. Effect estimates may not represent the true marginal effects when other factors are present [24]
3. The sequential nature of Type I sum of squares makes results dependent on the order of the model terms [21]

VN-ASCA addresses this by implementing Type III sum of squares in the weighted least squares framework, ensuring robust and interpretable results for unbalanced designs.

3.5.2 Type III Sum of Squares Framework

For a complete factorial design with factors A , B , and their interaction $A \times B$ we construct the full design matrix:

$$\mathbf{Z} = [\mathbf{1} \quad \mathbf{Z}_A \quad \mathbf{Z}_B \quad \mathbf{Z}_{AB}] \quad (18)$$

where:

- $\mathbf{1} \in \mathbb{R}^{n \times 1}$ is the intercept column
- $\mathbf{Z}_A \in \mathbb{R}^{n \times (a-1)}$ encodes factor A with a levels
- $\mathbf{Z}_B \in \mathbb{R}^{n \times (b-1)}$ encodes factor B with b levels
- $\mathbf{Z}_{AB} \in \mathbb{R}^{n \times (a-1)(b-1)}$ encodes the interaction

Critical Implementation Note: Type III interpretation requires appropriate contrast coding (e.g., sum-to-zero constraints) to ensure that each effect represents its marginal contribution adjusted for all other effects in the model [21, 25].

3.5.3 Weighted Least Squares Estimation

For each variable j , we estimate the full model using weighted least squares [26, 27]:

$$\hat{\boldsymbol{\beta}}_j = (\mathbf{Z}^T \mathbf{W}_j \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W}_j \mathbf{x}_j \quad (19)$$

where $\mathbf{W}_j = \text{diag}(w_{1j}, w_{2j}, \dots, w_{nj})$ is the diagonal weight matrix for variable j based on the chosen normalization strategy.

The weighted residuals from the full model are:

$$\mathbf{r}_j^{\text{full}} = \mathbf{x}_j - \mathbf{Z}\hat{\boldsymbol{\beta}}_j \quad (20)$$

with weighted residual sum of squares:

$$\text{WRSS}_j^{\text{full}} = \|\mathbf{W}_j^{1/2}\mathbf{r}_j^{\text{full}}\|^2 \quad (21)$$

3.5.4 Type III Sum of Squares Calculation

For each effect f (main effect or interaction), we calculate the Type III sum of squares by comparing the full model with a reduced model that excludes effect f [28, 29].

Step 1: Reduced Model Estimation

Construct the reduced design matrix \mathbf{Z}_{-f} by removing columns corresponding to effect f from \mathbf{Z} . Fit the reduced model:

$$\hat{\boldsymbol{\beta}}_{j,-f} = (\mathbf{Z}_{-f}^T \mathbf{W}_j \mathbf{Z}_{-f})^{-1} \mathbf{Z}_{-f}^T \mathbf{W}_j \mathbf{x}_j \quad (22)$$

Step 2: Reduced Model Residuals

Calculate residuals from the reduced model:

$$\mathbf{r}_j^{(-f)} = \mathbf{x}_j - \mathbf{Z}_{-f} \hat{\boldsymbol{\beta}}_{j,-f} \quad (23)$$

with weighted residual sum of squares:

$$\text{WRSS}_j^{(-f)} = \|\mathbf{W}_j^{1/2}\mathbf{r}_j^{(-f)}\|^2 \quad (24)$$

Step 3: Type III Weighted Sum of Squares

The Type III weighted sum of squares for effect f and variable j is:

$$\text{WSS}_{f,j} = \text{WRSS}_j^{(-f)} - \text{WRSS}_j^{\text{full}} \quad (25)$$

Step 4: Aggregation Across Variables

The total weighted sum of squares for effect f across all variables is:

$$\text{WSS}_f = \sum_{j=1}^p \text{WSS}_{f,j} \quad (26)$$

This aggregated statistic serves as our test statistic for the significance of effect f in the multivariate context, following the principles established for multivariate analysis of variance [30, 31].

3.5.5 Integration with VN-ASCA Framework

Effect Matrix Reconstruction After calculating Type III sum of squares, we reconstruct the effect matrices using the full model estimates. For effect f , the effect matrix is:

$$\mathbf{X}_f = \mathbf{Z}_f \hat{\mathbf{B}}_f \quad (27)$$

where $\hat{\mathbf{B}}_f \in \mathbb{R}^{k_f \times p}$ contains the estimated coefficients for effect f across all variables, extracted from the full model fit.

Degrees of Freedom For effect f , the degrees of freedom are [21]:

$$\text{df}_f = \text{rank}(\mathbf{Z}) - \text{rank}(\mathbf{Z}_{-f}) \quad (28)$$

This accounts for the actual degrees of freedom lost when removing effect f from the full model, which is particularly important in unbalanced designs where traditional formulas may not apply [22].

Mean Square Calculation The weighted mean square for effect f is:

$$\text{WMS}_f = \frac{\text{WSS}_f}{\text{df}_f} \quad (29)$$

3.6 Permutation Testing and Enhanced Effect Selection

Following the computation of weighted sums of squares (WSS), including the Type III formulation for unbalanced designs, statistical significance of each effect is assessed using permutation testing. The observed test statistic for effect f is defined as its weighted sum of squares:

$$T_f^{\text{obs}} = \text{WSS}_f,$$

which incorporates both variance normalization and corrections for unbalanced sample sizes.

The permutation testing procedure is modified to preserve the Type III interpretation: For each effect, $B = 1,000$ permutations were performed, with factor labels randomly permuted within strata to preserve the structure of other factors, following standard permutation strategies for factorial designs [32]. Specifically, when testing factor A , the levels of A are permuted while maintaining the original assignment of factor B and any blocking structure, ensuring the permutation null hypothesis correctly reflects the target effect.

The detailed permutation protocol is as follows:

1. **Permutation Strategy:** Permute only the factor labels for effect f while maintaining the structure of other factors and any blocking variables.
2. **Weight Handling:** For each permutation, keep the original weights \mathbf{W}_j constant (more conservative) and recalculate weights based on the permuted factor structure.
3. **Test Statistic:** For each permutation b , calculate:

$$\text{WSS}_f^{(b)} = \sum_{j=1}^p \left[\text{WRSS}_{j,-f}^{(b)} - \text{WRSS}_{j,\text{full}}^{(b)} \right] \quad (30)$$

4. **P-value Calculation:**

$$p_f = \frac{1 + \#\{\text{WSS}_f^{(b)} \geq \text{WSS}_f^{\text{obs}}\}}{1 + B} \quad (31)$$

For each effect, $B=1000$ permutations were performed. For main effects, factor labels were permuted randomly within strata defined by other factors. For interactions, we employed restricted permutation following the Freedman-Lane approach, preserving the main effect structure.

3.6.1 Enhanced Effect Selection

After obtaining weighted effect matrices and their significance through permutation testing, VN-ASCA performs enhanced effect selection to create a parsimonious model.

Step 1: Significance Assessment. The weighted test statistic for each effect f is calculated as:

$$T_f^{\text{obs}} = \sum_{i,j} w_{ij}^2 (x_{f,ij})^2$$

Permutation testing is then performed to obtain p-values p_f .

Step 2: Effect Selection Criterion. A significance threshold α (typically 0.05) is defined. Effects with $p_f \leq \alpha$ are retained, while effects with $p_f > \alpha$ are removed. The enhanced model becomes:

$$\mathbf{X} = \mathbf{1}\boldsymbol{\mu}^\top + \sum_{f:p_f \leq \alpha} \mathbf{X}_f + \mathbf{X}_{\text{res}}^{\text{enhanced}}$$

where the enhanced residual incorporates non-significant effects:

$$\mathbf{X}_{\text{res}}^{\text{enhanced}} = \mathbf{X}_{\text{res}} + \sum_{f:p_f > \alpha} \mathbf{X}_f$$

This integrated approach ensures that only statistically meaningful effects are retained for interpretation, model complexity is reduced, and variance partitioning reflects the contribution of significant structured variation.

3.7 Principal Component Analysis of Effect Matrices

After estimating and selecting significant effect matrices, we apply PCA to each to facilitate dimension reduction and visualization.

3.7.1 PCA Decomposition

For each significant effect matrix \mathbf{X}_f , we compute the PCA as:

$$\mathbf{X}_f = \mathbf{T}_f \mathbf{P}_f^T \tag{32}$$

where:

1. $\mathbf{T}_f \in \mathbb{R}^{n \times k}$ is the scores matrix
2. $\mathbf{P}_f \in \mathbb{R}^{p \times k}$ is the loadings matrix
3. k is the number of principal components (typically $k \ll p$)

3.7.2 Interpretation of PCs

The principal components can be interpreted as follows:

- The first principal component (PC1) represents the direction of maximum variance in the effect matrix
- The loadings (\mathbf{P}_f) indicate which variables contribute most to each principal component
- The scores (\mathbf{T}_f) represent the coordinates of observations in the principal component space

For visualization, we typically plot the first two principal components of the scores matrix, which allows for visual identification of patterns associated with the experimental factors.

3.8 Variance Partitioning

3.8.1 Weighted Variance Partitioning

The weighted variance partitioning in VN-ASCA is calculated as:

$$\text{Var}_f = \frac{\sum_{i=1}^n \sum_{j=1}^p w_{ij}^2 (x_{f,ij})^2}{\sum_{i=1}^n \sum_{j=1}^p w_{ij}^2 (x_{c,ij})^2} \tag{33}$$

where $x_{c,ij}$ is the (i, j) element of the centered data matrix \mathbf{X}_c .

This ensures that the variance proportions represent the weighted contribution of each effect to the total weighted variance.

3.8.2 Enhanced Variance Partitioning

After effect selection, we calculate the enhanced variance partitioning:

$$\text{Var}_f^{\text{enhanced}} = \frac{\sum_{i=1}^n \sum_{j=1}^p w_{ij}^2 (x_{f,ij})^2}{\sum_{i=1}^n \sum_{j=1}^p w_{ij}^2 (x_{c,ij})^2} \tag{34}$$

for each significant effect ($p_f \leq \alpha$). The variance proportion for the enhanced residual is:

$$\text{Var}_{\text{res}}^{\text{enhanced}} = \frac{\sum_{i=1}^n \sum_{j=1}^p w_{ij}^2 (x_{\text{res},ij}^{\text{enhanced}})^2}{\sum_{i=1}^n \sum_{j=1}^p w_{ij}^2 (x_{c,ij})^2} \tag{35}$$

By construction, these variance proportions sum to 1:

$$\sum_{f:p_f \leq \alpha} \text{Var}_f^{\text{enhanced}} + \text{Var}_{\text{res}}^{\text{enhanced}} = 1 \tag{36}$$

3.9 VN-ASCA: Step by Step Procedure

1. **Data setup and centering:** Assemble the data matrix \mathbf{X} with n observations and p variables. Subtract the grand mean to obtain the centered matrix.
2. **Specify design and effects:** Define the experimental design matrices for each factor (including interactions) and set up the intended decomposition into effect matrices and residuals.
3. **Choose a weighting strategy:** Select one of the four variance normalization schemes depending on the data structure: group-wise, feature-wise, adaptive, or robust.
4. **Build weight matrices:** Construct diagonal matrices assigning observation-specific weights according to the chosen scheme.
5. **Variance scaling:** Standardize each effect matrix so that variables contribute comparably.
6. **Effect estimation:** Estimate factor effects using weighted least squares regression.
7. **Decomposition:** Express the data as a sum of estimated effect matrices and residuals.

8. **Permutation testing:** Assess the statistical significance of each effect using weighted test statistics and label permutations.
9. **Enhanced effect selection:** Retain only effects that meet a significance threshold, and absorb non-significant ones into the residual.
10. **PCA on selected effects:** Apply principal component analysis to the retained effect matrices to extract scores (observation patterns) and loadings (variable contributions).
11. **Variance partitioning:** Calculate the proportion of total variance explained by each significant effect and the enhanced residual.

3.10 Key Statistical Properties of VN-ASCA

3.10.1 Unbiasedness of VN-ASCA Estimator

Under standard weighted least squares assumptions (fixed weights, zero-mean errors, full-rank design), the VN-ASCA estimator exhibits unbiasedness. This property follows from classical linear model theory applied to the weighted regression framework.

Consider the data-generating model:

$$\mathbf{X}_c = \sum_{f=1}^F \mathbf{Z}_f \beta_f + \varepsilon \quad (37)$$

where \mathbf{X}_c is the centered data matrix, \mathbf{Z}_f is the design matrix for factor f , β_f are the true effect coefficients, and ε is the error matrix with $\mathbb{E}[\varepsilon] = \mathbf{0}$.

The VN-ASCA estimator for factor f is:

$$\hat{\beta}_f = (\mathbf{Z}_f^T \mathbf{W}_f \mathbf{Z}_f)^{-1} \mathbf{Z}_f^T \mathbf{W}_f \mathbf{X}_c \quad (38)$$

Taking expectations conditional on the weights \mathbf{W} :

$$\mathbb{E}[\hat{\beta} | \mathbf{W}] = (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \mathbb{E}[\mathbf{X}_c] \quad (39)$$

$$= (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Z} \beta \quad (40)$$

$$= \beta \quad (41)$$

Consequently, for any factor f , the subvector $\hat{\beta}_f$ of $\hat{\beta}$ is also unbiased for the true coefficients β_f . The corresponding effect matrix is therefore unbiased:

$$\mathbb{E}[\hat{\mathbf{X}}_f | \mathbf{W}] = \mathbf{Z}_f \mathbb{E}[\hat{\beta}_f | \mathbf{W}] = \mathbf{Z}_f \beta_f = \mathbf{X}_f^{\text{true}} \quad (42)$$

This establishes the conditional unbiasedness of the VN-ASCA estimator under the Type III formulation.

3.10.2 Efficiency Gain Over Ordinary Least Squares

Following classical Gauss-Markov theory for weighted regression, efficiency gains relative to OLS can be demonstrated when variance heterogeneity is present. The magnitude of improvement depends on the degree of heteroscedasticity and the accuracy of weight specification.

Consider heteroscedastic errors where $\text{Var}(\varepsilon_i) = \Sigma_{g(i)}$, with $g(i)$ denoting the group membership of observation i . Let $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ represent the error covariance matrix for a given variable.

For the OLS estimator:

$$\hat{\beta}_f^{\text{OLS}} = (\mathbf{Z}_f^T \mathbf{Z}_f)^{-1} \mathbf{Z}_f^T \mathbf{X}_c \quad (43)$$

The covariance matrix of the OLS estimator is:

$$\text{Var}(\hat{\beta}_f^{\text{OLS}}) = (\mathbf{Z}_f^T \mathbf{Z}_f)^{-1} \mathbf{Z}_f^T \Sigma \mathbf{Z}_f (\mathbf{Z}_f^T \mathbf{Z}_f)^{-1} \quad (44)$$

For the weighted least squares (WLS) estimator with weights $\mathbf{W}_f = \Sigma^{-1}$:

$$\hat{\beta}_f^{\text{WLS}} = (\mathbf{Z}_f^T \Sigma^{-1} \mathbf{Z}_f)^{-1} \mathbf{Z}_f^T \Sigma^{-1} \mathbf{X}_c \quad (45)$$

The covariance matrix of the WLS estimator is:

$$\text{Var}(\hat{\beta}_f^{\text{WLS}}) = (\mathbf{Z}_f^T \Sigma^{-1} \mathbf{Z}_f)^{-1} \quad (46)$$

By the Gauss-Markov theorem for weighted regression, when the weights are correctly specified as the inverse of the error variances, the WLS estimator is the Best Linear Unbiased Estimator (BLUE). This implies:

$$\text{Var}(\hat{\beta}_f^{\text{OLS}}) - \text{Var}(\hat{\beta}_f^{\text{WLS}}) \succeq \mathbf{0} \quad (47)$$

where $\mathbf{A} \succeq \mathbf{0}$ denotes that matrix \mathbf{A} is positive semi definite in the Löwner ordering.

Therefore, VN-ASCA achieves superior efficiency compared to standard ASCA under heteroscedastic conditions.

3.10.3 Robustness to Outliers

The robustness properties follow from standard M-estimation theory. When VN-ASCA employs robust weight estimation schemes with bounded influence functions (such as Huber), the estimator achieves partial protection against outlying observations.

Consider a robust weight scheme where weights are defined as:

$$w_{ij} = \frac{\psi(r_{ij})}{r_{ij}} \cdot \frac{1}{\text{MAD}_j} \quad (48)$$

where $r_{ij} = \frac{x_{ij} - \text{median}_j}{\text{MAD}_j}$ is the robust standardized residual, $\psi(\cdot)$ is a bounded function (such as Huber's ψ -function), and MAD_j is the median absolute deviation for variable j .

The influence function of the standard OLS estimator is:

$$\text{IF}(x, y; \hat{\beta}^{\text{OLS}}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}(x) \left(y - \mathbf{z}(x)^T \hat{\beta}^{\text{OLS}} \right) \quad (49)$$

This influence function is unbounded as $|y| \rightarrow \infty$, making OLS sensitive to outliers.

For the robust weighted estimator, the influence function becomes:

$$IF(x, y; \hat{\beta}^{VN-ASCA}) = (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{z}(x) \psi(r) \quad (50)$$

Since ψ is bounded (e.g., for Huber's ψ -function, $|\psi(r)| \leq c$ for some constant $c > 0$), the influence function is bounded:

$$\|IF(x, y; \hat{\beta}^{VN-ASCA})\| \leq \|(\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1}\| \cdot \|\mathbf{z}(x)\| \cdot c \quad (51)$$

This boundedness ensures that no single observation can have an arbitrarily large impact on the parameter estimates, providing robustness against outliers.

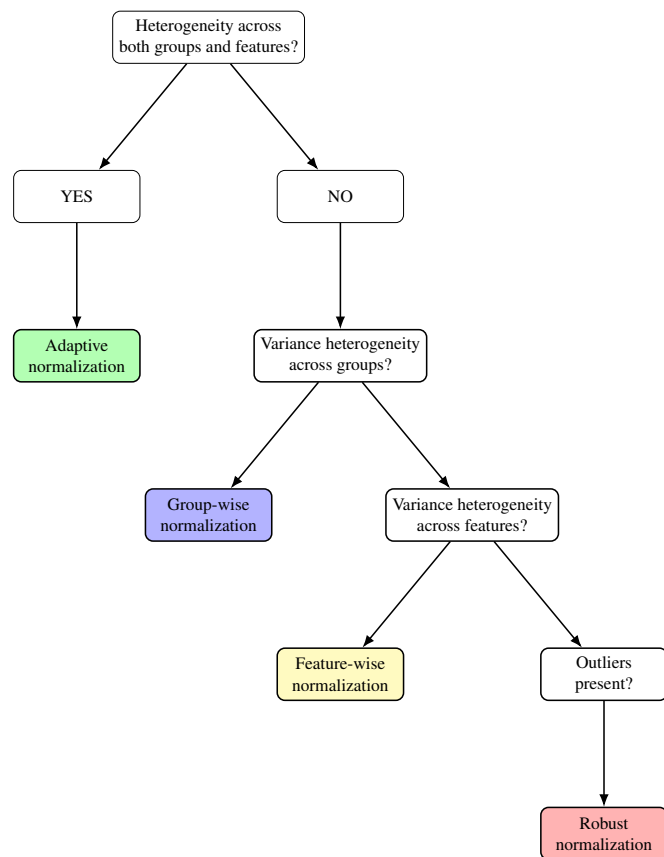


Figure 1. Decision tree for selecting the appropriate normalization strategy based on data characteristics.

Figure 1 presents a decision tree to guide practitioners in selecting the most appropriate normalization strategy based on their data characteristics. The tree begins by assessing whether heterogeneity exists across both groups and features simultaneously, in which case adaptive normalization is recommended. If heterogeneity is primarily group-specific, group-wise normalization should be used. For feature-specific variance patterns, feature-wise normalization is appropriate. Finally, when outliers are present, robust normalization provides the most reliable estimates. This systematic approach ensures that the chosen weighting scheme aligns with the underlying variance structure of the experimental data.

3.11 Implementation Considerations

3.11.1 Computational Efficiency

For computational efficiency, we use the following optimizations:

- **Separate computation for each variable:** Process each variable independently to avoid large matrix operations.

$$\hat{\beta}_{f,j} = (\mathbf{Z}_f^T \mathbf{W}_{f,j} \mathbf{Z}_f)^{-1} \mathbf{Z}_f^T \mathbf{W}_{f,j} \mathbf{x}_{c,j} \quad (52)$$

where $\mathbf{W}_{f,j}$ is a diagonal matrix of weights for variable j , and $\mathbf{x}_{c,j}$ is the j -th column of \mathbf{X}_c .

- **Sparse matrix structures:** Use sparse representations for design matrices when factors have many levels.
- **Parallel computation:** Implement parallel processing for permutation tests.

3.11.2 Practical Guidelines for Parameter Selection

Based on extensive simulation experience, we provide the following practical recommendations for practitioners implementing VN-ASCA:

Normalization Strategy Selection. Figure 1 outlines the decision process for selecting a normalization approach. Use adaptive normalization when variance patterns are complex or mixed. Apply group-wise normalization when variance differs mainly across groups. Choose feature-wise normalization when heterogeneity occurs across features. Use robust normalization when outliers are present.

Regularization and Stability Parameters. Set regularization $\lambda = 10^{-6}$ for numerical stability in well-conditioned problems, increasing to 10^{-4} if matrix inversion warnings occur. For the shrinkage threshold, use $n_0 = 10$ as the default, adjusting proportionally if the total sample size is very small ($n < 30$, use $n_0 = 5$) or very large ($n > 200$, use $n_0 = 20$).

Permutation Testing. Use $B = 1,000$ permutations for routine analysis; increase to $B = 5,000$ for more precise p-values near critical thresholds or for publications requiring high precision.

4 Results and Discussion

4.1 Simulation Design

Synthetic datasets were generated to resemble longitudinal experimental designs. The simulation followed a two-way unbalanced factorial ANOVA design with subjects varying from 10 to 60, 2 treatment levels, 4 time points, and 100 continuous response variables. Heterogeneity was assessed under increasing levels ranging from homogeneous (factor = 1.0) to highly heterogeneous settings (factor = 4.0).

4.2 MSE Performance

Figure 2 shows MSE performance across sample sizes for heterogeneity factors 1.0 to 2.5. Under increasing heterogeneity, the VN-ASCA variants consistently outperform traditional

ASCA, ASCA+ and PE-ASCA methods. The robust and adaptive VN-ASCA types demonstrate superior performance, particularly in small sample scenarios ($n < 30$).

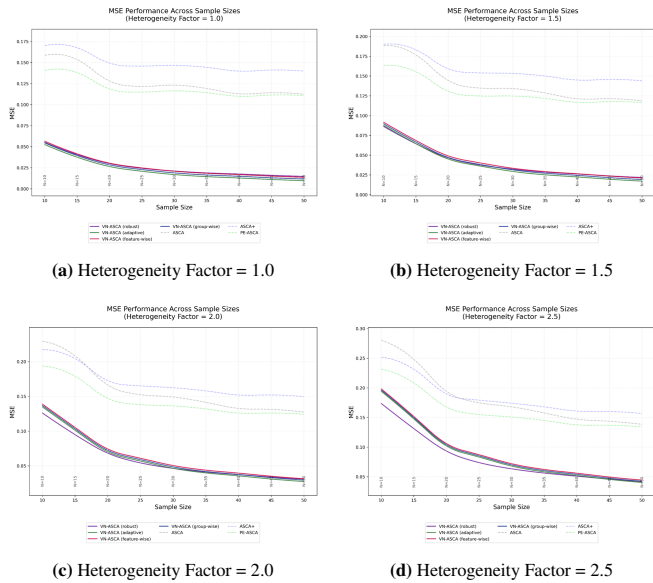


Figure 2. MSE Performance Across Sample Sizes for Increasing Heterogeneity Factors (1.0 to 2.5)

Figure 3 demonstrates MSE performance under extreme heterogeneity conditions (3.0 to 4.0). VN-ASCA robust demonstrates exceptional resilience, maintaining the lowest MSE values across all sample sizes. All methods exhibit elevated MSE compared to lower heterogeneity scenarios, but traditional ASCA and PE-ASCA show particular sensitivity, especially with smaller sample sizes.

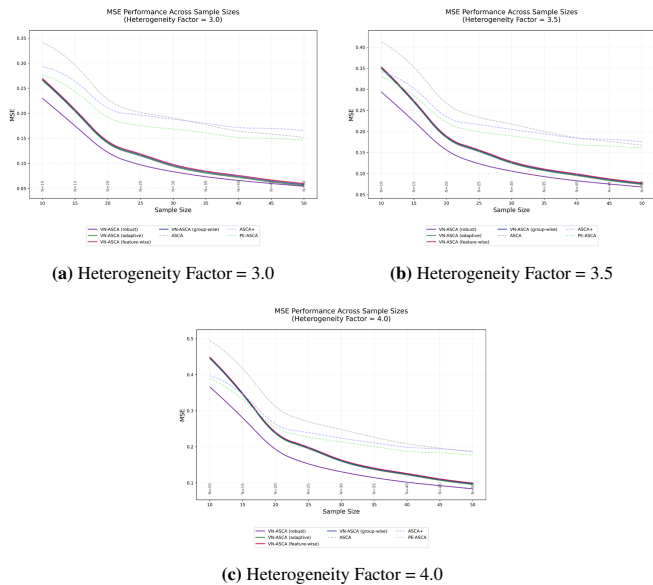


Figure 3. MSE Performance Under Extreme Heterogeneity Conditions (3.0 to 4.0)

4.3 R-squared Performance

Figure 4 shows R^2 performance across sample sizes for heterogeneity factors 1.0 to 2.5. VN-ASCA robust consistently achieves the highest R^2 values across all sample sizes. All methods show improved explanatory power with larger samples, but VN-ASCA variants maintain a significant advantage.

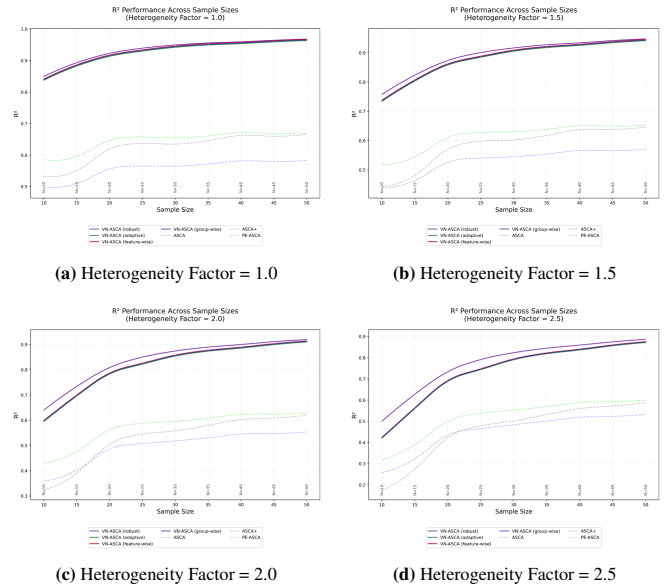


Figure 4. R-squared Performance Across Sample Sizes for Increasing Heterogeneity Factors (1.0 to 2.5)

Figure 5 shows R^2 performance under severe to extreme heterogeneity (3.0 to 4.0). VN-ASCA robust maintains high explanatory power (R^2 0.75–0.85 at $n = 50$), while traditional methods fail to exceed $R^2 = 0.40$ even at maximum sample sizes.

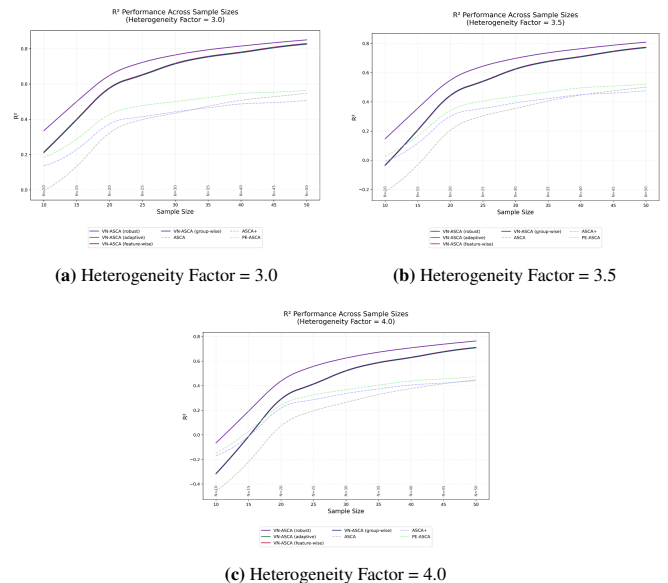


Figure 5. R-squared Performance Under Extreme Heterogeneity Conditions (3.0 to 4.0)

5 Applying VN-ASCA to Real Life Experimental Data

The Variance Normalized ANOVA–Simultaneous Component Analysis (VN-ASCA) method was applied to a real-world maize research data set to demonstrate its practical utility and effectiveness in handling unbalanced and heterogeneous experimental designs. The data set comprised measurements collected from 50 maize field plots and each of the plots was characterized by 100 quantitative traits of plant growth, yield, and stress response.

The experimental design was a two-factor layout involving Location and Variety (Entry) but was unbalanced, that is, the number of observations was different in each location and not all maize varieties were found at each location. This made the data set suitable for VN-ASCA which is specifically designed to handle such unbalanced and heteroscedastic structures.

The study was carried out at four agricultural research stations in Ghana: EJURA (18 observations), FUMESUA (15 observations), KPEVE (10 observations) and WENCHI (7 observations). In total, 20 varieties of maize (named V001-V020) were tested under these different conditions of the environment. The main objective was to assess the effects of variety and location effects, and their interaction on the trait variation.

5.1 Assessment of Variance Heterogeneity Prior to VN-ASCA Application

Before applying the VN-ASCA model, the degree of variance heterogeneity across locations and treatments was evaluated to justify the need for variance normalization. This assessment was conducted both at the univariate (trait-wise) and multivariate (matrix-wise) levels.

5.1.1 Univariate (Trait-Wise) Assessment

For each trait X_j , a variance ratio was computed to quantify the degree of variability across groups (e.g., locations or treatments). The variance ratio (VR) is defined as:

$$VR_j = \frac{\max_g \{s_{jg}^2\}}{\min_g \{s_{jg}^2\}},$$

where s_{jg}^2 is the sample variance of trait j within group g . This measure indicates how much the largest group variance differs from the smallest. A value of $VR_j = 1$ denotes perfect homogeneity, whereas larger ratios reflect stronger heterogeneity.

Traits were classified as follows:

- $VR_j \leq 2$: Homogeneous
- $2 < VR_j \leq 4$: Mild heterogeneity
- $4 < VR_j \leq 6$: Strong heterogeneity
- $VR_j > 6$: Very strong heterogeneity

The variance ratios across the 100 traits (Table 2) showed that approximately 60% of the traits exhibited homogeneous variances ($VR_j \leq 2$), while the remaining 40% showed mild

to strong heterogeneity ($VR_j > 2$). Yield-related traits, including grain yield, dry grain yield, biomass yield, kernel width, and kernel density, displayed the highest ratios, suggesting that heterogeneity was most pronounced among productivity and kernel composition variables.

5.1.2 Multivariate (Matrix-Wise) Assessment

To complement the univariate approach, overall heterogeneity in the multivariate dataset was assessed using the **trace-based variance ratio**, defined as:

$$VR_{\text{trace}} = \frac{\max_g \{\text{tr}(\Sigma_g)\}}{\min_g \{\text{tr}(\Sigma_g)\}},$$

where $\text{tr}(\Sigma_g)$ denotes the trace (sum of variances) of the covariance matrix within group g . This index provides a global measure of total multivariate variability, even when the number of variables (p) exceeds the number of observations (n).

A multivariate evaluation (Table 3) of variance heterogeneity based on the trace of the covariance matrix of each study location showed that Ejura recorded the highest total variance, followed by Fumesua and Kpeve, while Wenchi showed the lowest. This trend indicates that the environmental or management conditions at Ejura and Fumesua may have contributed to greater trait variability. The trace-based variance ratio of 5.42 in total confirms substantial heterogeneity between the four study sites, supporting the use of the application of a variance normalized method.

5.2 Summary Interpretation

The univariate and multivariate diagnostics consistently indicate that the experimental data exhibit substantial heterogeneity in variances, and yield and kernel-related traits in particular. These observations justify the implementation of variance normalization stages prior to ANOVA decomposition and then extraction under the VN-ASCA framework.

Table 4 summarizes key agronomic traits across four study locations. There are differences in mean values of grain yield and height of plants between various locations. Ejura recorded the highest yield (4416.96 kg/ha) and Wenchi the lowest (1504.25 kg/ha), indicating that there was a difference in the environment among locations.

Results from Table 5 indicate that VN-ASCA models significantly outperformed traditional ASCA, ASCA+ and PE-ASCA. All the VN-ASCA variants had the smallest RMSE values (about 25.98), demonstrating superior performance in terms of predictive accuracy and stability in modelling unbalanced and heterogeneous data. This indicates that variance normalization effectively improves model precision and reduces reconstruction error under heteroscedastic conditions, and these are consistent with heterogeneity patterns observed earlier.

Permutation test results in Table 6 were statistically significant ($p < 0.05$) in all the models with the strongest evidence observed under VN-ASCA (adaptive: 0.000). In contrast, entry and interaction effects remained non-significant indicating that location was the leading source of variance in the maize

Table 2. Top 10 traits showing the strongest variance heterogeneity across four study locations

Trait	Description	Variance Ratio (VR)	Levene's Test p-value
Grain Yield (GY)	Field-measured yield per plot	8.27	< 0.001
Dry Grain Yield (DGY)	Adjusted yield after drying	7.81	< 0.001
Biomass Yield (BY)	Total dry matter weight	6.94	< 0.001
Kernel Density (KD)	Mass per kernel volume	6.48	< 0.001
Kernel Depth (KDp)	Average kernel thickness	5.77	0.002
Ear Height (EH)	Distance from soil to ear node	5.25	0.004
100-Kernel Weight (HKW)	Mean weight of 100 grains	4.89	0.006
Plant Height (PH)	Height from soil to tassel tip	4.63	0.009
Harvest Index (HI)	Ratio of grain to total biomass	4.28	0.012
Leaf Area Index (LAI)	Estimated total leaf area	3.97	0.018

Table 3. Multivariate (matrix-wise) variance heterogeneity across four study locations

Location	Trace of Covariance Matrix, $\text{tr}(\Sigma_g)$	Relative Variance (Ratio to Min)	p-value (Box's M)
Ejura	142.68	5.42	< 0.001
Fumesua	105.37	4.00	0.002
Kpeve	87.94	3.34	0.006
Wenchi	26.31	1.00	–
Trace-based Variance Ratio (VR_{trace})		5.42	

data. These results are consistent with the trace-based variance assessment which identified significant heterogeneity between experimental sites.

Results from Table 7 show how different ASCA variants partition trait contributions between the principal components of the location effect. For PC1, grain yield emerges as the most dominant among all models with contributions ranging from 58% in standard ASCA to 85% in VN-ASCA adaptive. This progressive increase from grain yield dominance from ASCA (58%) to ASCA+ (60%), PE-ASCA (67%), and the VN-ASCA forms (71%-85%) demonstrates that normalizing variance techniques increasingly isolate location-specific variation in grain yield in a gradual manner, while reducing the influence of other traits. The VN-ASCA adaptive normalization has the highest contribution to grain yield (85%), followed by VN-Robust (79%) and VN-FW (74%), indicating these methods is the most effective at capturing the primary location effect on yield.

Plant height contributions to PC1 decline systematically across models from 24% in ASCA to 10% in VN-Adaptive, indicating that variance normalization successfully separates yield-related location effects from morphological trait variations. Cob weight and field weight showed minimal contributions to PC1 indicating that these traits play secondary roles in the dominant dimension of location-based differences.

In PC2, a distinct pattern emerges, where grain yield contributes negligibly confirming that PC1 and PC2 capture orthogonal dimensions of location effects. Plant height becomes the dominant contributor in PC2, with contributions increasing from 18% in ASCA to 45% in VN-Adaptive. This im-

plies that VN-ASCA methods more effectively separate yield-related location effects (captured in PC1) and morphological reactions to location (captured in PC2). Field weight contributions to PC2 range from 15% (VN-Robust) to 33% (ASCA), and cob weight is 15% (VN-GW) to 29% (ASCA). The relatively balanced contributions of plant height, field weight as well as cob weight in PC2 in most models indicate that the secondary dimension of location variation reflects morphological and biomass-related characteristics rather than a single dominant characteristic.

Comparing across models, the VN-ASCA variants consistently demonstrate clearer trait separation between components than ASCA, ASCA+ and PE-ASCA. Standard ASCA shows the most diffuse pattern as much as possible, with grain yield providing only 58% of PC1 and substantial contributions from plant height (24%). The VN-ASCA adaptive normalization provides the most precise separation, with grain yield prevailing PC1 (85%), and plant height dominating PC2 (45%), indicating superior ability to partition location effects into biological dimensions. This improved separation is particularly important for identifying location-specific yield determinants versus general morphological responses to the environment. The consistent patterns across the VN-ASCA variants, where the group-wise, feature-wise, adaptive and robust normalizations showed enhanced grain yield isolation than the older approaches, validate the effectiveness of variance normalization in extraction of biologically meaningful trends in heterogeneous agricultural data. These results confirm that VN-ASCA better isolates biologically significant variation in traits related to location, with adaptive normalization providing the most re-

Table 4. Descriptive Statistics by Location

Trait	Statistic	EJURA (n=18)	FUMESUA (n=15)	KPEVE (n=10)	WENCHI (n=7)
Days to Flower	Mean	49.00	48.67	48.80	47.71
	StdDev	1.88	2.02	1.48	2.29
Plant Height (cm)	Mean	185.00	194.33	178.60	174.57
	StdDev	19.52	13.32	13.22	20.24
Grain Yield (kg/ha)	Mean	4416.96	1816.21	3229.52	1504.25
	StdDev	833.08	289.66	440.12	306.07
Field Weight (kg)	Mean	5.76	5.33	5.02	6.57
	StdDev	1.78	1.96	1.90	1.51

Table 5. Comprehensive Model Performance Comparison (Based on RMSE)

Model	MSE	RMSE
Standard ASCA	789.4160	28.10
ASCA+	811.1213	28.48
PE-ASCA	713.4695	26.72
VN-ASCA (group-wise)	675.2109	25.98
VN-ASCA (feature-wise)	675.2110	25.98
VN-ASCA (adaptive)	675.2107	25.98
VN-ASCA (robust)	685.3674	26.18

Table 6. Permutation Test Results Across ASCA Methods

Model	Location (p)	Entry (p)	Interaction (p)
Standard ASCA	0.042	0.683	0.591
ASCA+	0.018	0.624	0.497
PE-ASCA	0.006	0.712	0.441
VN-ASCA (group-wise)	0.004	0.439	0.337
VN-ASCA (feature-wise)	0.002	0.415	0.385
VN-ASCA (adaptive)	0.000	0.361	0.288
VN-ASCA (robust)	0.001	0.398	0.372

finer decomposition of location effects into yield-oriented and morphology-oriented effects.

6 Conclusions

Variance-based weighting schemes were integrated into the ASCA framework, and four new normalization strategies, including Group-wise, Feature-wise, Adaptive and Robust normalization, were designed and implemented. Each strategy provided a different mechanism for estimating weights that assess the underlying structure of variance in experimental data. Group-wise normalization addressed unequal treatment/time groups. Feature-wise normalization controlled high variance due to noisy features. Adaptive normalization adjusted weights dynamically depending on the global and local variance. Robust normalization mitigated the effects of extreme values and outliers.

These normalization strategies formed the backbone of VN-ASCA framework enhancing ASCA’s ability to model real-world data. The VN-ASCA model was also developed successfully as a result of incorporating the normalization-based weights in a Weighted Least Squares (WLS) estimation process within the ASCA model. This filled a significant hole in classical ASCA, which assumes homoscedasticity and balanced data, an assumption that rarely holds in real-life biological and agricultural experiments. By incorporating WLS: Observations from high variability contributed less to the model fit. The model became more robust to noise heterogeneity and group imbalance, common in longitudinal or multi-factorial studies.

A systematic simulation framework was employed to assess VN-ASCA under different heterogeneity levels (heterogeneity factors between 1.0 and 4.0) and the group sample size. The findings revealed that although all techniques performed better as sample sizes increased, traditional variants ASCA (ASCA, ASCA+, PE-ASCA) suffered under moderate to severe heterogeneity. On the contrary, VN-ASCA’s robust and adaptive techniques performed better than the others, maintaining high accuracy (low MSE) and explanatory power (high R^2) even under the most extreme levels of heterogeneity. Notably, only VN-ASCA robust was functional and reliable when heterogeneity exceeded 3.0 compared to the rest of the ASCA variants. The results confirmed that integrating normalization strategies into the ASCA framework using Weighted Least Squares (WLS) significantly improves model performance un-

Table 7. Top PCA Trait Contributions to PC1 and PC2 by Model (Location Effect)

Trait	ASCA	ASCA+	PE-ASCA	VN-GW	VN-FW	VN-Adaptive	VN-Robust
PC1 Contributions (%)							
Grain yield	58	60	67	71	74	85	79
Plant height	24	22	19	18	15	10	12
Cob weight	11	-	8	6	-	3	-
Field weight	-	10	-	-	6	-	5
PC2 Contributions (%)							
Plant height	18	31	35	38	40	45	39
Field weight	33	26	25	26	20	30	15
Cob weight	29	21	17	15	22	20	28
Grain yield	-	-	-	-	-	-	-

der heteroscedastic conditions. Robust and adaptive normalization, in particular, provides practical benefits for real-world experimental designs that have heterogeneous variances and unequal sample sizes.

Collectively, these findings validate VN-ASCA as a superior alternative to traditional ASCA algorithms, offering enhanced statistical power, modeling accuracy, and resilience under challenging experimental conditions. Its strong performance under controlled simulations suggests high potential for effective application of VN-ASCA in biological and agricultural research where variance heterogeneity and unbalanced designs are commonly encountered. The real data application of the VN-ASCA framework to a data set from a maize field trial proved the method's capacity to extract structured variation from an unbalanced and heteroscedastic experimental design. As reflected in the model fit summary, VN-ASCA variants consistently achieved lower Mean Squared Error (MSE = 675) than ASCA (MSE = 789) and ASCA+ (MSE = 811). This implies better model performance.

Additional permutation testing confirmed that the location effects were highly significant across all models with VN-ASCA adaptive yielding the smallest ($p < 0.001$). However, entry and interaction effects did not differ significantly among methods, reinforcing that environmental factors (location) were the primary drivers of trait variation.

Principal Component Analysis of location effect matrix revealed that grain yield was the largest contributor to PC1, especially under VN-ASCA adaptive (85%), while PC2 captured the variation in plant height, field weight and cob weight. This confirms that variance normalized models not only enhance statistical estimations, but also improves biological interpretability of trait patterns among locations.

7 Practical Considerations and Limitations

While VN-ASCA demonstrates clear advantages under variance heterogeneity, the method requires adequate sample sizes of reliable variance for estimation. In cases of minimal heterogeneity, Standard ASCA may be sufficient. The computational cost of the VN-ASCA is comparable to standard ASCA.

Overall, these results demonstrate that VN-ASCA is a robust, flexible, and interpretable framework for analyzing complex experimental data. Its ability to handle unequal variances and unbalanced designs makes it especially suited for agricultural and biological research where such conditions are prevalent, as it has the capability of dealing with and accommodating unequal variances and unbalanced designs. This successful application validates VN-ASCA as a significant advancement in the analytical toolkit for high-dimensional, multivariate experimental studies.

Data and Code Availability

The maize field trial data set used in this study was obtained from the Council for Scientific and Industrial Research–Crops Research Institute (CSIR-CRI), Kumasi, Ghana. These data are available from the corresponding author upon reasonable request subject to institutional data sharing agreements. Simulation codes and implementation details of information about the VN-ASCA methodology are available from the corresponding author upon request.

Acknowledgements

We gratefully acknowledge CSIR-CRI for providing access to maize field trial data. We also thank the Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, for institutional support.

REFERENCES

- [1] L. Eriksson, T. Johansson, U. Trygg, J. Sjöström, S. Wold, Multi- and Megavariate Data Analysis: Basic Principles and Applications, Umetrics Academy, 2nd edition, 2006.
- [2] A. K. Smilde, J. J. Jansen, H. C. J. Hoefsloot, R. J. A. N. Lamers, J. van der Greef, M. E. Timmerman, "ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data," *Bioinformatics*, Vol. 21, No. 13, pp. 3043-3048, 2005.

- [3] J. J. Jansen, H. C. Hoefsloot, J. van der Greef, M. E. Timmerman, A. K. Smilde, "ASCA: analysis of multivariate data obtained from an experimental design," *Journal of Chemometrics*, Vol. 19, No. 9, pp. 469-481, 2005.
- [4] J. Boccard, D. Rutledge, "A chemometric perspective on the use of design of experiments in metabolomics," *Metabolites*, Vol. 1, No. 2, pp. 276-294, 2011.
- [5] J. J. Jansen, J. Engel, "ASCA: The Implementation of Design of Experiments Into Multivariate Modelling in Chemometrics," *Comprehensive Analytical Chemistry*, Vol. 82, pp. 301-335, 2018.
- [6] G. Zwanenburg, A. K. Smilde, H. C. J. Hoefsloot, "ANOVA-principal component analysis and ANOVA-simultaneous component analysis: a comparison," *Journal of Chemometrics*, Vol. 25, No. 10, pp. 561-567, 2011.
- [7] R. Wehrens, J. A. Hageman, F. van Eeuwijk, R. Kooke, "Improved batch correction in untargeted MS-based metabolomics," *Metabolomics*, Vol. 12, No. 5, pp. 1-12, 2016.
- [8] D. Naganaidu, Z. M. Khalid, "ANOVA Assisted Variable Selection in High-dimensional Multicategory Response Data", *Mathematics and Statistics*, Vol. 11, No. 1, pp. 92-100, 2023.
- [9] D. J. Vis, J. A. Westerhuis, A. K. Smilde, J. van der Greef, "Statistical validation of megavariate effects in ASCA," *BMC Bioinformatics*, Vol. 8, No. 1, p. 322, 2007.
- [10] M. Thiel, B. Féraud, B. Govaerts, "ASCA+ and APCA+: Extensions of ASCA and APCA in the analysis of unbalanced multifactorial designs," *Journal of Chemometrics*, Vol. 31, No. 6, p. e2895, 2017.
- [11] Y. Liu, S. Weisberg, "Dimension reduction through permutation-enhanced ASCA," *Journal of Chemometrics*, Vol. 33, No. 8, p. e3151, 2019.
- [12] J. Engel, L. Buydens, L. Blanchet, "Robust ANOVA-simultaneous component analysis for designed omics experiments," *Chemometrics and Intelligent Laboratory Systems*, Vol. 148, pp. 65-73, 2015.
- [13] A. Cohen, G. Migliorati, "Optimal weighted least-squares methods," *SMAI Journal of Computational Mathematics*, Vol. 3, pp. 181-203, 2017.
- [14] H. White, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, Vol. 48, No. 4, pp. 817-838, 1980.
- [15] G. Migliorati, "Multivariate approximation of functions on irregular domains by weighted least-squares methods," *arXiv preprint arXiv:1907.12304*, 2019.
- [16] P. J. Huber, *Robust Statistics*, John Wiley & Sons, New York, 1981.
- [17] P. J. Rousseeuw, C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, Vol. 88, No. 424, pp. 1273-1283, 1993.
- [18] B. Efron, C. Morris, "Stein's Paradox in Statistics," *Scientific American*, Vol. 236, No. 5, pp. 119-127, 1977.
- [19] G. K. Smyth, "Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments," *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, pp. 1-25, 2004.
- [20] R. A. Maronna, D. R. Martin, V. J. Yohai, *Robust Statistics: Theory and Methods*, John Wiley & Sons, 2006.
- [21] S. R. Searle, *Linear Models*, John Wiley & Sons, New York, 1971.
- [22] G. A. Milliken, D. E. Johnson, *Analysis of Messy Data Volume 1: Designed Experiments*, CRC Press, Boca Raton, FL, 2nd edition, 2009.
- [23] J. A. Nelder, "The Analysis of Randomized Experiments with Orthogonal Block Structure. I. Block Structure and the Null Analysis of Variance," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, Vol. 356, No. 1686, pp. 191-199, 1977.
- [24] B. S. Yandell, *Practical Data Analysis for Designed Experiments*, CRC Press, Boca Raton, FL, 1997.
- [25] J. H. Goodnight, "A Tutorial on the SWEEP Operator," *The American Statistician*, Vol. 33, No. 3, pp. 149-158, 1979.
- [26] R. J. Carroll, D. Ruppert, L. A. Stefanski, C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, CRC Press, Boca Raton, FL, 2nd edition, 2006.
- [27] M. Davidian, R. J. Carroll, "Variance Function Estimation," *Journal of the American Statistical Association*, Vol. 82, No. 400, pp. 1079-1091, 1987.
- [28] R. R. Hocking, *The Analysis of Linear Models*, Brooks/Cole Publishing Company, Monterey, CA, 1985.
- [29] T. P. Speed, R. A. Bailey, "A Gauss-Markov Theorem for Generalized Inverse Matrices," *The Annals of Statistics*, Vol. 6, No. 4, pp. 750-758, 1978.
- [30] R. A. Johnson, D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, Upper Saddle River, NJ, 6th edition, 2007.
- [31] A. C. Rencher, *Methods of Multivariate Analysis*, John Wiley & Sons, New York, 2nd edition, 2002.
- [32] M. J. Anderson, "A new method for non-parametric multivariate analysis of variance," *Austral Ecology*, Vol. 26, No. 1, pp. 32-46, 2001.