Application of Procrustes Analysis in the Comparison of Two Data Matrices of Green Cherry Fruit Production of Pagaralam Robusta Coffee

Irmeilyana[1], Bambang Suprihatin[2], Sri Indra Maiyanti[3], Dita Irfanul Khasanah[4], Ngudiantoro[5], Anita Desiani[6]

{irmeilyana@unsri.ac.id[1], bambangs@unsri.ac.id[2], sri_indra_maiyanti@mipa.unsri.ac.id[3], 08011182126015@student.unsri.ac.id[4], ngudiantoro@unsri.ac.id[5], anita_desiani@unsri.ac.id[6]}

University of Sriwijaya, Faculty of Mathematics and Natural Sciences, Department of Mathematics, Jl. Raya Palembang-Prabumulih Km. 32, Indralaya, South Sumatra Province, Indonesia[1,2,3,4,5,6]

Corresponding author: bambangs@unsri.ac.id

Abstract. One factor that influences the productivity of Robusta coffee trees is the characteristics of tree branches. The purpose of this study was to compare the production of green cherries in terms of the characteristics of tree branches in two fields with different maintenance management in Pagaralam, South Sumatra. The data matrix entries of each field are the values of 7 variables in 30 coffee tree samples. These variables consist of the total number of cherries, the number of branch categories, and the number of cherries on each branch. Each data matrix was represented graphically by biplot analysis. The biplot results of each data matrix were compared using Procrustes analysis. The similarity comparison between the two data matrices included a translation process of 1.449, a rotation process of 0.191, and a dilation process of 0.180, resulting in a similarity measure of 83.7%. This indicates a similarity between the two biplots.

Keywords: Biplot analysis, Pagaralam, points configuration, procrustes analysis, robusta coffee.

1 Introduction

Coffee trees (*Coffea spp.*) are a major plantation commodity in several tropical countries, including Indonesia [1]. Coffee is a food agro-industrial product that makes a significant contribution to the Indonesian economy, both domestically and internationally, playing a crucial role in foreign exchange earnings and driving the economy for both farmers and other economic actors [2]. Indonesia is the fifth-largest coffee producer in the world after Brazil, Vietnam, Columbia, and Ethiopia, with a total contribution of approximately 5% [3]. Indonesian coffee production consists of 77.84% robusta coffee and 27.16% arabica coffee [4]. The quality and quantity of coffee harvests are highly dependent on various factors, including tree characteristics and land management [5].

SICBAS 2025, November 06, Palembang, Indonesia

DOI 10.4108/eai.6-11-2025.2364262

Interventions in land maintenance through partnership programs are one strategy that can significantly impact coffee cherry production. Through partnerships, farmers are provided with skills in sustainable coffee cultivation management [6]. To optimize land maintenance, partnership programs often involve collaboration between farmers, field agricultural extention, and related institutions to increase coffee yields [7].

In addition to land maintenance, other factors that can affect coffee plant productivity include climate change and pruning. Pruning is a tree maintenance method used to reduce leaf and branch growth and increase fruit yield [8]. Coffee plants are generally pruned four times a year [9]. Pruning is related to the classification of Robusta coffee tree branches, based on branch position, shape, and production year [10]. There are three types of pruning: water shoot pruning to create balanced branching, production pruning to generate new primary branches to increase fruit production, and post-harvest pruning, which is performed on older plants due to declining yields [11]. According to Khayati et al. [12], pruning branches at this time will impact the number of branches and production in the following year.

Research by Zulkarnain et al. [9] using descriptive analysis methods showed a significant difference between the income of coffee farmers who pruned and those who did not. Land maintenance through pruning can increase coffee fruit yield [13]. In this case, pruning and land maintenance are related to the classification of branch categories on Robusta coffee plants. Cherry production across branch categories will vary across coffee trees.

The relationship between the number of cherries and the number of fruit clusters in each branch category on a sample tree can be represented in the form of a data matrix. A data matrix represents each object with multiple variables (multivariate). Multivariate data can be reduced to a data matrix with fewer variables. One multivariate analysis for variable reduction is PCA (Principal Component Analysis) [14]. Then, the objects and their variable values can be represented as configurations of points in a space whose dimensions correspond to the number of variables. The comparative measure of suitability between two configurations of data sets is a numeric value known as Procrustes analysis [15]. The comparison of two data matrices, consisting of the original data matrix and the reduced data matrix, is performed through the translation, rotation, and dilation stages of Procrustes analysis.

The points configuration of a data matrix can be visualized in a biplot graph, which is a continuation of the results of multivariate data reduction through PCA. The interesting thing is how to analyze the structural similarity between two configurations from two data groups, so that the similarity of the relationship patterns between objects in the two configurations can be further analyzed. The purpose of this study was to compare the configuration of biplot analysis results of green cherry production of Robusta coffee on two crops with different maintenance management using Procrustes analysis. This study used secondary data from the data matrix of two plots with 30 tree samples each obtained from Irmeilyana et al. [16]. The two plots were distinguished based on whether the farmers participated in the partnership program or not, thus differing in maintenance management. Some forms of sustainable coffee cultivation management activities through the partnership program are weed control procedures, types of fertilizers used, procedures and schedules for fertilization, and pruning implementation. The variables used were green cherry production according to branch position categories, namely the number of primary branches, secondary branches, and tertiary branches, along with the total number of fruits on each branch. By using Procrustes analysis, the configuration of the results of the biplot analysis on 2 data matrices of green cherry coffee production can be compared, so that the similarity or dissimilarity of fruit production from two different crop maintenance systems can be analyzed.

2 Methods

The data source used in this study was secondary data obtained from Irmeilyana, et al. [16] on green cherry production of Robusta coffee trees related to tree branch categories on two coffee farmer's plantations with different maintenance management. There were 7 variables used, namely number of primary branches (notated as X_1), number of secondary branches (as X_2), number of tertiary branches (as X_3), total fruits (as X_4), number of fruits on primary branches (as X_5), number of fruits on secondary branches (as X_6), and number of fruits on tertiary branches (as X_7). There were 30 sample trees selected from each coffee field, which were assumed to be representative of all coffee trees on each field. So, the size of a matrix data is 30 × 7, namely $n = 30$ and $p = 7$.

This study conducted an analysis to compare the configuration of two biplots of two data matrices for green cherry production of Robusta coffee trees. The following are the steps taken in this study:

1. Perform statistics descriptive on each variable including mean and standard deviation.
2. Compile a data matrix for each Ismail's coffee field and Wahid's coffee field, denoted by $\boldsymbol{A}$ and $\boldsymbol{B}$.
3. Conduct a biplot analysis on the data matrix for each coffee field:
   3.1 Compile the correlation matrices $\boldsymbol{R_A}$ and $\boldsymbol{R_B}$.
   3.2 Calculate the eigenvalues λ_1, λ_2, $\cdots, \lambda_p$ (where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$) and eigenvectors $\boldsymbol{a_1}$, $\boldsymbol{a_2}$, $\cdots$, $\boldsymbol{a_p}$ of matrices $\boldsymbol{R_A}$ and $\boldsymbol{R_B}$.
   3.3 Select the two largest eigenvalues λ_1 and λ_2 in each matrix $\boldsymbol{R_A}$ and $\boldsymbol{R_B}$.
   3.4 Calculate the goodness of fit measure of the two largest eigenvalues using

$$\rho = \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^{p} \lambda_j}. \tag{1}$$

4. Interpret the results of the biplot analysis, namely the relationship between variables, the proximity between objects, and the relative relationship between variables and objects.
5. Perform Procrustes Analysis [17-21]. Procrustes analysis is a numerical method to compare two configurations that obtained by moving one configuration so that it aligns best with the other.
   5.1 Create a configuration of coordinates of points in first biplot results as an observation matrix $\boldsymbol{X} = (x_{ij})$ and a configuration of coordinates of points in second biplot results as a target matrix $\boldsymbol{Y} = (y_{ij})$. The configuration of these points is in 2-dimensional space. Both matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same dimension $n \times 2$ or $p \times 2$. The row entries of $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be the coordinates of points in component scores of biplot result such that $i = 1, 2, \cdots, n$ or the terminal points of variable vectors such that $i = 1, 2, \cdots, p$. In this study, the configurations compared are the terminal points of variable vectors.

5.2 Perform a data set translation process on matrix $\boldsymbol{Y}$ using

$$M^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \left((x_{ij} - \bar{x}_j) - (y_{ij} - \bar{y}_j) \right)^2, \tag{2}$$

where

M^2 : distance to the corresponding points.

x_{ij} : element of matrix $\boldsymbol{X}$ on row$-i$ and column$-j$.

y_{ij} : element of matrix $\boldsymbol{Y}$ on row$-i$ and column$-j$.

$\bar{x}_j$: mean of variable X_j based on entries of matrix $\boldsymbol{X}$ on column$-j$.

$\bar{y}_j$: mean of variable Y_j based on entries of matrix $\boldsymbol{Y}$ on column$-j$.

5.3 Perform a rotation adjustment process on matrix Y using

$$M^2 = tr(\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}} + \boldsymbol{Y}\boldsymbol{Y}^{\boldsymbol{T}} - 2\boldsymbol{X}\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{Y}^{\boldsymbol{T}}), \tag{3}$$

where M^2 is distance to the corresponding points in rotation process and $\boldsymbol{Q}$ is orthogonal matrix by SVD (Singular Value Decomposition) of matrix $\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X}$,

$$\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{U}^{\boldsymbol{T}}. \tag{4}$$

SVD of matrix $\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X}$ is a method of decomposing the matrix $\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X}_{p\times p}$ into three matrix components, namely the $\boldsymbol{U}$, $\boldsymbol{L}$, and $\boldsymbol{V}^{\boldsymbol{T}}$ matrices, so $\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X} = \boldsymbol{U}\,\boldsymbol{L}\,\boldsymbol{V}^{\boldsymbol{T}}$.

where $\boldsymbol{U}_{p\times p} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}}\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X}\,\boldsymbol{v}_{\boldsymbol{1}} & \frac{1}{\sqrt{\lambda_2}}\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X}\,\boldsymbol{v}_{\boldsymbol{2}} & \cdots & \frac{1}{\sqrt{\lambda_p}}\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X}\,\boldsymbol{v}_{\boldsymbol{p}} \end{bmatrix}$

$$\boldsymbol{L}_{p\times p} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sqrt{\lambda_p} \end{bmatrix}$$

$\boldsymbol{V}_{p\times p} = [\boldsymbol{v}_{\boldsymbol{1}} \quad \boldsymbol{v}_{\boldsymbol{2}} \quad \dots \quad \boldsymbol{v}_{\boldsymbol{p}}]$

λ_j is j – th largest eigenvalue of the matrix $(\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X})^T(\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X})$; with $j = 1, 2, \dots, p$.

$\boldsymbol{v}_{\boldsymbol{j}}$ is the eigenvector of the matrix $(\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X})^T(\boldsymbol{Y}^{\boldsymbol{T}}\boldsymbol{X})$ that correpons to the eigenvalue λ_j.

Matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices such that $\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^{\boldsymbol{T}} = \boldsymbol{I}$, $\boldsymbol{V}^{\boldsymbol{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\boldsymbol{T}} = \boldsymbol{I}$, and also $\boldsymbol{Q}\boldsymbol{Q}^{\boldsymbol{T}} = \boldsymbol{I}$.

5.4 Equalize the scales of both data sets by performing a dilation process on matrix Y using

$$M^2 = tr(\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}) - c^2 tr\,(\boldsymbol{Y}\boldsymbol{Y}^{\boldsymbol{T}}), \tag{5}$$

where M^2 is distance to the corresponding points in dilation process and scalar

$$c = \frac{tr(\boldsymbol{X}\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{Y}^{\boldsymbol{T}})}{tr(\boldsymbol{Y}\boldsymbol{Y}^{\boldsymbol{T}})}. \tag{6}$$

5.5 Select the smallest distance value in the transformation process.

5.6 Calculate the similarity of the configuration of two matrices (R^2) using

$$R^2 = \left(1 - \frac{M_{min}^2}{tr(\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}})}\right) \times 100\%, \tag{7}$$

where R^2 is determination coefficient or similarity measure and M_{min}^2 is minimum distance on transformation process. The R value is also known as the Procrustes correlation.

5.7 Interpret the results of the Procrustes analysis.

The research steps were carried out using Minitab version 19 and R Studio version 4.3.1.

3 Results and Discussion

Thirty tree samples were collected from each coffee field, with seven variables, so the two data matrices can be written in a 30 × 7 matrix as follows:

$$\boldsymbol{A} = \begin{bmatrix} 3 & 15 & 4 & \cdots & 106 \\ 7 & 21 & 0 & \cdots & 0 \\ 9 & 11 & 0 & \cdots & 0 \\ 11 & 18 & 0 & \cdots & 0 \\ 3 & 8 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 8 & 32 & 5 & \cdots & 65 \\ 13 & 43 & 0 & \cdots & 0 \end{bmatrix}$$

and

$$\boldsymbol{B} = \begin{bmatrix} 3 & 12 & 27 & \ldots & 861 \\ 7 & 30 & 16 & \ldots & 702 \\ 4 & 12 & 18 & \ldots & 1248 \\ 4 & 12 & 36 & \ldots & 2466 \\ 5 & 8 & 26 & \ldots & 2090 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 6 & 15 & 2 & \ldots & 12 \\ 3 & 12 & 30 & \ldots & 1792 \end{bmatrix}$$

The statistics descriptive of the data for each variable shown in **Table 1** only include the mean and standard deviation (that denoted as StDev).

Based on **Table 1**, number of tertiary branches (X_3), number of fruits on tertiary branches (X_7), and total fruits (X_4) in data matrix $\boldsymbol{B}$ have significantly higher means and standard deviations than in data matrix $\boldsymbol{A}$. The apposite is true for number of fruits on primary branches (X_5) in data matrix $\boldsymbol{B}$, which has a much lower mean and standard deviation than in data matrix $\boldsymbol{A}$. Meanwhile, number of primary branches (X_1), number of secondary branches (X_2), and number of fruits on secondary branches (X_6) in the two data matrices have almost the same (or do not have a large difference) mean and standard deviation.

Table 1. Descriptive statistics of each variable in both data matrices

Variable	Data matrix $\boldsymbol{A}$		Data matrix $\boldsymbol{B}$	
	Mean	StDev	Mean	StDev
X_1	7.2	3.78199	5.03	1.650148
X_2	15.7	8.199804	17.96	8.547971
X_3	3.5	5.276711	30.5	19.04215
X_4	967.7	588.7554	2113.87	1057.148
X_5	224.9	292.4462	31.4	63.10396
X_6	684.0	425.1348	766.4	614.7395
X_7	87.2	147.2089	1326.93	889.1701

The variables in the data matrix have different values, so standardization is necessary by converting the data into Z-scores. The correlation between 2 variables on initial every data matrix are shown as $\boldsymbol{R_A}$ and $\boldsymbol{R_B}$, namely

$$\boldsymbol{R_A} = \begin{bmatrix} 1 & 0.384 & -0.284 & 0.411 & 0.656 & 0.278 & -0.341 \\ 0.384 & 1 & 0.016 & 0.395 & 0.081 & 0.831 & 0.001 \\ -0.284 & 0.016 & 1 & 0.247 & -0.207 & 0.089 & 0.954 \\ 0.411 & 0.395 & 0.247 & 1 & 0.582 & 0.713 & 0.251 \\ 0.656 & 0.081 & -0.207 & 0.582 & 1 & 0.156 & -0.224 \\ 0.278 & 0.831 & 0.089 & 0.713 & 0.156 & 1 & 0.110 \\ -0.341 & 0.001 & 0.954 & 0.251 & -0.224 & 0.110 & 1 \end{bmatrix}$$

$$\boldsymbol{R_B} = \begin{bmatrix} 1 & 0.603 & -0.205 & 0.064 & 0.662 & 0.437 & -0.282 \\ 0.603 & 1 & -0.044 & 0.409 & 0.426 & 0.820 & -0.115 \\ -0.205 & -0.044 & 1 & 0.686 & -0.488 & -0.014 & 0.869 \\ 0.064 & 0.409 & 0.686 & 1 & -0.189 & 0.586 & 0.795 \\ 0.662 & 0.426 & -0.488 & -0.189 & 1 & 0.280 & -0.497 \\ 0.437 & 0.820 & -0.014 & 0.586 & 0.280 & 1 & -0.021 \\ -0.282 & -0.115 & 0.869 & 0.795 & -0.497 & -0.021 & 1 \end{bmatrix}$$

Based on the entry of row 3, column 7 in the matrices, namely 0.954 and 0.869, this shows that the highest correlation occurs between tertiary branches (X_3) and the number of fruits on tertiary branches (X_7). This means that in both matrices. the variables influence each other.

PCA (Principal Component Analysis) is used to reduce the original data into fewer main components but does not eliminate the original data information. PCA as an initial analysis to represent data in biplot graphs in 2-dimensional space. PCA (Principal Component Analysis) in this study was conducted on two data, namely green cherry production data in data $\boldsymbol{A}$ and in data $\boldsymbol{B}$ which consists of 7 variables and 30 research objects using a correlation matrix. The first step to determine the PC (Principal Component) is to determine the eigenvalues and eigenvectors of the correlation matrices $\boldsymbol{R_A}$ and $\boldsymbol{R_B}$.

Based on the solution of the characteristic equation, the eigenvalues are obtained from the largest to the smallest. The first two eigenvalues of the correlation matrix $\boldsymbol{R_A}$ are $\lambda_1 = 2.8218$ and $\lambda_2 = 2.2808$. Whereas, the eigenvalues of the correlation matrix $\boldsymbol{R_B}$ are $\lambda_1 =$

$3.0195,$ and $\lambda_2 = 2.6824$. By equation (1), the first two eigenvalues indicate the variance of the 2 new variables, so that the variance percentage result of the 2 PC is obtained.

$$\Pr(\boldsymbol{A}) = \frac{\lambda_1 + \lambda_2}{tr(\boldsymbol{R_A})} \times 100\% = \frac{2.8218 + 2.2808}{7} \times 100\% = 72.9\%$$

$$\Pr(\boldsymbol{B}) = \frac{\lambda_1 + \lambda_2}{tr(\boldsymbol{R_B})} \times 100\% = \frac{3.0195 + 2.6824}{7} \times 100\% = 82\%$$

The cumulative proportion of the first two eigenvalues from the correlation matrix explains the level of suitability of the PCA-reduced variable subspace. The proportion of sum of the first two eigenvalues indicates the biplot fit measure. So, the biplot fit measure for data $\boldsymbol{A}$ and $\boldsymbol{B}$ are repectively 72.9% and 82%.

By PCA, new variables (Principal Components; PC) can be formed from both data matrices. The values of the eigenvectors $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ are as coefficients of first PC (notated as PC1) and second PC (notated as PC2). The eigenvectors $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ will be obtained which correspond to the first two eigenvalues. The first two eigenvectors of correlation matrices $\boldsymbol{R_A}$ and $\boldsymbol{R_B}$ can be seen in **Table 2**.

Table 2. Eigenvector coefficients (loadings) of PCA results for both data matrices

Variable	Data matrix $\boldsymbol{A}$		Data matrix $\boldsymbol{B}$	
	PC1 ($\boldsymbol{a_1}$)	PC2 ($\boldsymbol{a_2}$)	PC1 ($\boldsymbol{a_1}$)	PC2 ($\boldsymbol{a_2}$)
X_1	*0.432*	0.263	*0.380*	0.296
X_2	*0.434*	-0.119	0.269	*0.486*
X_3	-0.046	*-0.619*	*-0.464*	0.259
X_4	*0.489*	-0.216	-0.275	*0.521*
X_5	*0.384*	0.214	*0.466*	0.114
X_6	*0.483*	-0.211	0.184	*0.509*
X_7	-0.056	*-0.629*	*-0.492*	0.257

Note: Numbers in italics indicate the variable that dominantly represents PC.

At the PC coefficients in data $\boldsymbol{A}$, PC1 is dominantly represented by X_1, X_2, X_4, X_5 and X_6 variables. PC2 is dominantly represented by X_1, X_3, X_4 and X_7 variables. Meanwhile, the PC1 coefficient in data $\boldsymbol{B}$ is more dominantly represented by X_1, X_3, X_4, X_5, and X_7 variables. Then, its PC2 is dominantly represented by X_1, X_2, X_4 and X_6 variables. The coefficient values of every PC form a linear combination of each variable. For an example PC1 in data $\boldsymbol{A}$ as follows:

$$PC1_A = 0.432X_1 + 0.434X_2 - 0.046X_3 + 0.489X_4 + 0.384X_5 + 0.483X_6 - 0.056X_7$$

Furthermore, to determine the component scores (namely PC1 and PC2 coordinates) for each object's data, we can substitute the values of the $X_1, X_2, X_3, \ldots, X_7$ variables of each object in the standardized matrix into PC1 and PC2. For example, for the component scores of first object in data $\boldsymbol{A}$ (as I1) are

$$PC1_A = 0.432(-1.11053) + 0.434(0.951242) - 0.046(-0.60644) + 0.489(1.195403) + 0.384(-0.02462) + 0.483(1.406142) - 0.056(0.671155) = 1.1777$$

$$PC2_A = 0.263(-1.11053) - 0.119(0.951242) - 0.619(-0.60644) - 0.216(1.195403)$$

$$+0.214(-0.02462) - 0.211(1.406142) - 0.629(0.671155) = -1.0111$$

The component scores of first object in data ***B*** (as W1) are

$$PC1_B = 0.380(-1.23221) + 0.269(-0.69802) - 0.464(1.153581) - 0.275(-0.92784) + 0.466(-0.49759) + 0.184(-0.80424) - 0.492(-0.52401) = -1.0582$$
$$PC2_B = 0.296(-1.23221) + 0.486(-0.69802) + 0.259(1.153581) + 0.521(-0.92784) + 0.114(-0.49759) + 0.509(-0.80424) + 0.257(-0.52401) = -1.4894$$

So, the component score of I1 is point $(1.0111\,, 1.1777)$ and the component score W1 is point $(-1.0582\,, -1.4894)$. By the same way, it can be obtained points of the other objects in each data matrix.

After conducting PCA, the next step was to conduct a biplot analysis to create a map of green cherry production for 30 trees on each data matrix. The object coordinates were determined based on the component score values, with the variable vector determined based on the coefficient values (loadings) in **Table 1**. The biplot of green cherry production on data ***A*** represents 72.9% of the data diversity, as can be seen in **Figure 1**.

Figure 2 and **Figure 3** respectively represent graphs of loadings plot and coordinate points of the objects based on the biplot of **Figure 1**. Based on **Figure 1** and **Figure 2**, it can be seen that variable X_1 has a very strong positive correlation with variable X_4. Variable X_2 has a strong positive correlation with variable X_6. Meanwhile, variable X_3 has a very strong positive correlation with variables X_4 and X_7. Based on the relative position of the objects (trees) and the variable vector, the position of objects in the same direction as the variables has a relatively high value compared to objects in the opposite direction.

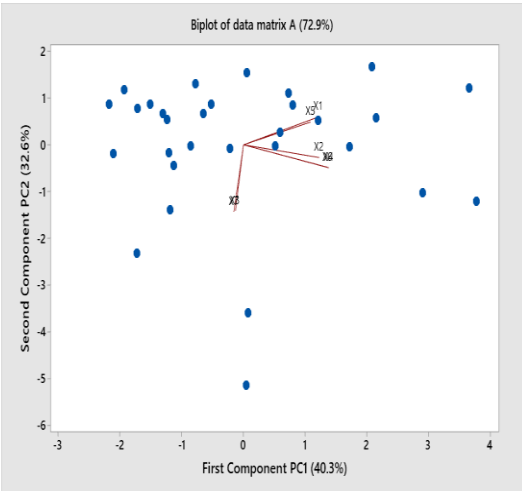

Fig. 1. Biplot of data matrix ***A***

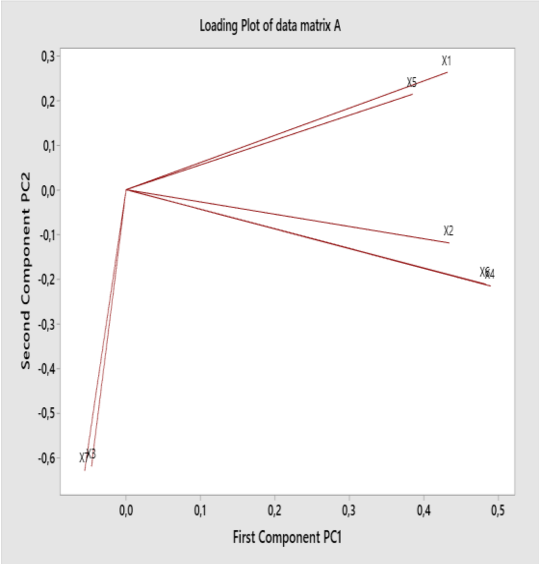

Fig. 2. Loadings plot of data matrix ***A***

Based on **Figures 1** and **Figure 3**, tree I13 and tree I29 have higher values of X_2 and X_6 compared to other trees. Trees I4, I19, I8, I22, I27, and I25 have higher values of X_1 and X_5 than other trees. Meanwhile, trees I23 and I28 have higher values of X_3, X_4, and X_7 than other trees.

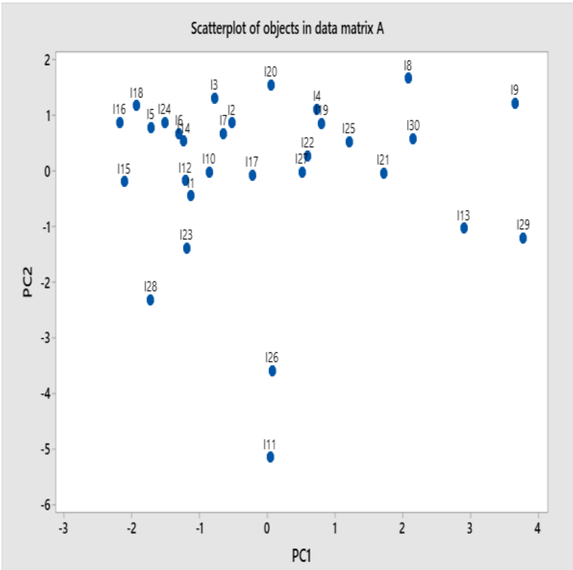

Fig. 3. Tree component scores in data matrix ***A***

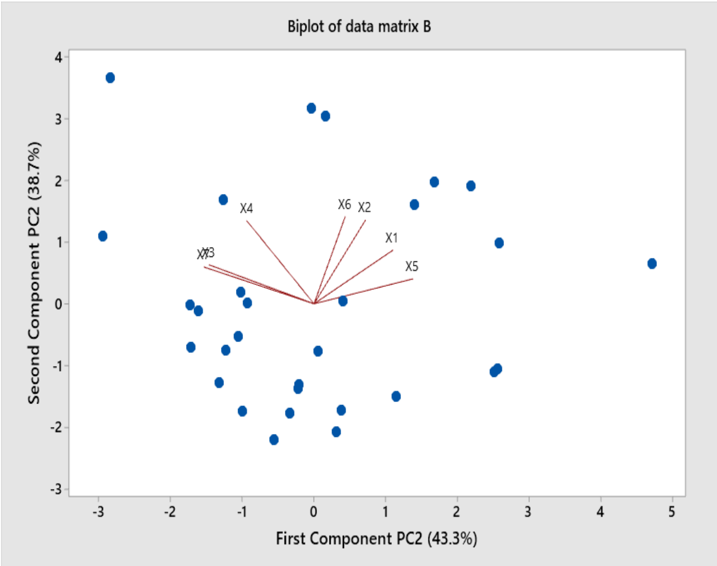

Fig. 4. Biplot of data matrix ***B***

The biplot of green cherry production on data matrix ***B*** represented 82% of the data diversity and can be seen in **Figure 4**. Based on **Figure 5**, variable X_3 has a very strong positive correlation with variable X_7. At the relative positions of the objects (trees) and the variable vector, the positions of objects in the same direction as the variables have relatively high values compared to objects in the opposite direction.

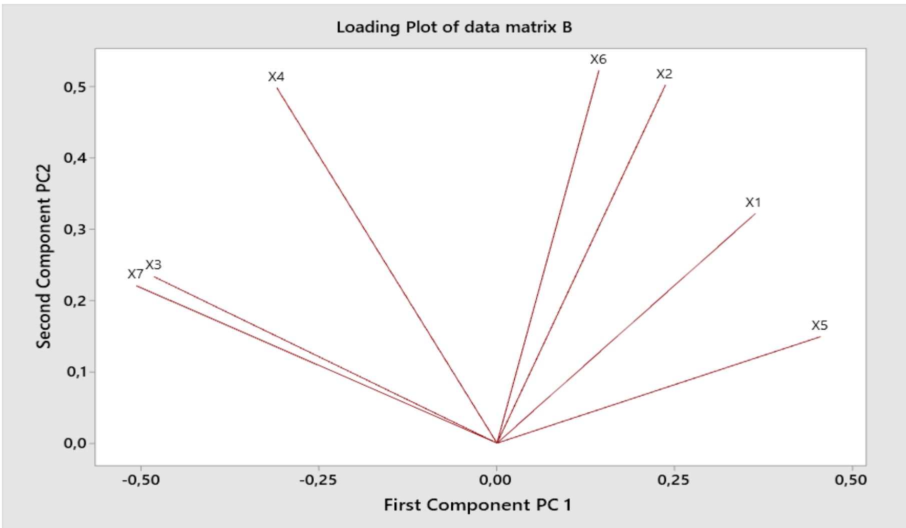

Fig. 5. Loadings plot of data matrix ***B***

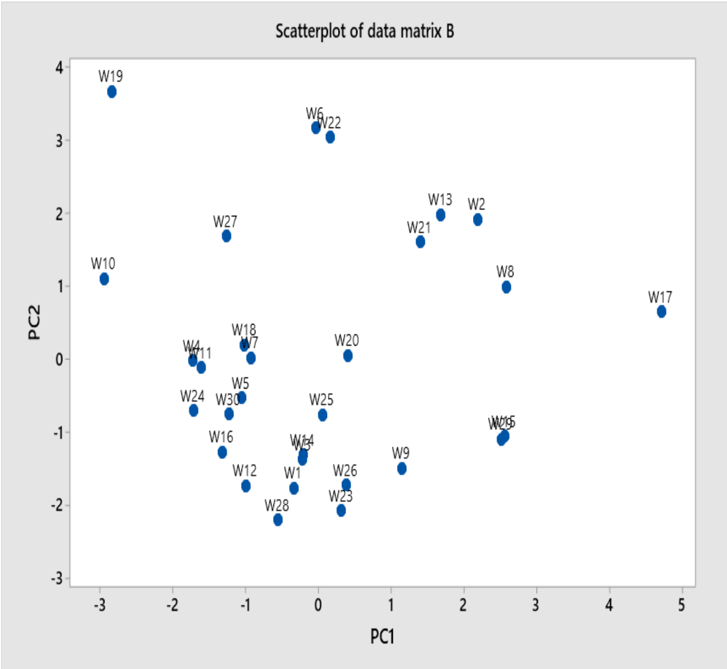

Fig. 6. Tree component scores in data matrix $\boldsymbol{B}$

Figure 6 is as a graph of the coordinate points of the objects in the biplot in **Figure 4**. Based on **Figure 6**, trees W19 and W27 have higher value of X_4 than the other trees. Trees W6 and W22 have higher values of X_2 and X_6 than the other trees. Trees W2, W8, W13, and W21 have higher value of X_1 than the other trees. Tree W17 has higher value of X_5 than the other trees. Meanwhile, trees W10, W14, and W18 have higher values of X_3 and X_7 than the other trees.

In this study, Procrustes analysis was performed on green cherry production data from Ismail's field data (data $\boldsymbol{A}$) and Wahid's field data (data $\boldsymbol{B}$). The input matrix used was the variable coordinate matrix (loadings plot on Table 1) in the biplot of each data set. The size of matrix is $p \times 2$, where each p representing the number of coordinate points in **Figure 2** and **Figure 5**. The variable coordinates in the biplot of data $\boldsymbol{A}$ are expressed as the observation matrix ($\boldsymbol{X}^* = [\boldsymbol{a_1} \quad \boldsymbol{a_2}]$) and the variable coordinates in the biplot of data $\boldsymbol{B}$ are expressed as the target matrix ($\boldsymbol{Y}^* = [\boldsymbol{a_1} \quad \boldsymbol{a_2}]$). The following are each input matrix used in the Procrustes analysis.

$$\boldsymbol{X}^* = \begin{bmatrix} 0.432 & 0.263 \\ 0.434 & -0.119 \\ -0.046 & -0.619 \\ 0.489 & -0.216 \\ 0.384 & 0.214 \\ 0.483 & -0.211 \\ -0.056 & -0.629 \end{bmatrix} \text{ and } \boldsymbol{Y}^* = \begin{bmatrix} 0.380 & 0.296 \\ 0.269 & 0.486 \\ -0.464 & 0.259 \\ -0.275 & 0.521 \\ 0.466 & 0.114 \\ 0.184 & 0.509 \\ -0.492 & 0.257 \end{bmatrix}$$

After defining the matrices $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$, the next step is to carry out the translation process, which begins by subtracting the matrices $\boldsymbol{X}^*$ and $\boldsymbol{Y}^*$ from each centroid matrix. The centroids of the matrices are obtained by $\boldsymbol{C_x} = [\boldsymbol{\bar{x}_1} \quad \boldsymbol{\bar{x}_2}]$, with

$$\bar{\boldsymbol{x}}_1 = \begin{bmatrix} \dfrac{0.432 + 0.434 - 0.046 + \cdots - 0.056}{7} \\ \vdots \\ \dfrac{0.432 + 0.434 - 0.046 + \cdots - 0.056}{7} \end{bmatrix} = \begin{bmatrix} 0.30286 \\ \vdots \\ 0.30286 \end{bmatrix}$$

$$\bar{\boldsymbol{x}}_2 = \begin{bmatrix} \dfrac{0.451 + 0.069 - 0.431 + \cdots - 0.441}{7} \\ \vdots \\ \dfrac{0.451 + 0.069 - 0.431 + \cdots - 0.441}{7} \end{bmatrix} = \begin{bmatrix} -0.18814 \\ \vdots \\ -0.18814 \end{bmatrix}$$

So, $\boldsymbol{C}_x = \begin{bmatrix} 0.30286 & -0.18814 \\ 0.30286 & -0.18814 \\ 0.30286 & -0.18814 \\ 0.30286 & -0.18814 \\ 0.30286 & -0.18814 \\ 0.30286 & -0.18814 \\ 0.30286 & -0.18814 \end{bmatrix}$.

By the same way, it can be obtained $\boldsymbol{C}_y = \begin{bmatrix} 0.00971 & 0.34886 \\ 0.00971 & 0.34886 \\ 0.00971 & 0.34886 \\ 0.00971 & 0.34886 \\ 0.00971 & 0.34886 \\ 0.00971 & 0.34886 \\ 0.00971 & 0.34886 \end{bmatrix}$.

Next, the matrices $\boldsymbol{X}^* - \boldsymbol{C}_x$ and $\boldsymbol{Y}^* - \boldsymbol{C}_y$ are obtained, so they are denoted as the new matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, namely:

$$\boldsymbol{X} = \begin{bmatrix} 0.129 & 0.451 \\ 0.131 & 0.069 \\ -0.349 & -0.431 \\ 0.186 & -0.028 \\ 0.081 & 0.402 \\ 0.180 & -0.023 \\ -0.358 & -0.441 \end{bmatrix} \quad \text{and} \quad \boldsymbol{Y} = \begin{bmatrix} 0.370 & -0.053 \\ 0.259 & 0.137 \\ -0.474 & -0.089 \\ -0.285 & 0.172 \\ 0.456 & -0.234 \\ 0.174 & 0.160 \\ -0.502 & -0.092 \end{bmatrix}$$

The next step is to calculate the sum of the squares of the distances at the corresponding points of $\boldsymbol{X}$ and $\boldsymbol{Y}$ based on equation (2), resulting in:

$$M^2 = \sum_{i=1}^{7} \sum_{j=1}^{2} \left((x_{ij} - \bar{x}_j) - (y_{ij} - \bar{y}_j) \right)^2$$
$$= ((0.129 - 0.370)^2 + (0.451 - (-0.053))^2 + \cdots + (-0.441 - (-0.092))^2) = 1.449$$

The next step after translation is the rotation process, which is used to minimize the distance between the two data sets being compared. The following is the matrix used in the rotation process using equation (3).

$$XX^T = \begin{bmatrix} 0.220 & 0.048 & -0.239 & 0.011 & 0.192 & 0.012 & -0.245 \\ 0.048 & 0.022 & -0.075 & 0.022 & 0.038 & 0.022 & -0.078 \\ -0.239 & -0.076 & 0.307 & -0.053 & -0.202 & -0.053 & 0.315 \\ 0.011 & 0.022 & -0.053 & 0.035 & 0.004 & 0.034 & -0.055 \\ 0.191 & 0.038 & -0.202 & 0.004 & 0.168 & 0.005 & -0.206 \\ 0.013 & 0.022 & -0.053 & 0.034 & 0.005 & 0.033 & -0.055 \\ -0.245 & -0.077 & 0.315 & -0.055 & -0.055 & -0.055 & 0.323 \end{bmatrix}$$

$$YY^T = \begin{bmatrix} 0.140 & 0.088 & -0.171 & -0.115 & 0.181 & 0.056 & -0.181 \\ 0.088 & 0.086 & -0.135 & -0.050 & 0.086 & 0.067 & -0.143 \\ -0.171 & -0.135 & 0.232 & 0.119 & -0.195 & -0.097 & 0.246 \\ -0.115 & -0.050 & 0.119 & 0.111 & -0.170 & -0.022 & 0.127 \\ 0.181 & 0.086 & -0.195 & -0.170 & 0.263 & 0.042 & -0.207 \\ 0.056 & 0.067 & -0.097 & -0.022 & 0.042 & 0.056 & -0.102 \\ -0.181 & -0.143 & 0.246 & 0.127 & -0.207 & -0.102 & 0.260 \end{bmatrix}$$

$$Y^T X = \begin{bmatrix} 0.442 & 0.797 \\ 0.117 & -0.038 \end{bmatrix}$$

Next, the matrix $(Y^T X)^T (Y^T X)$ is obtained as follow

$$(Y^T X)^T (Y^T X) = \begin{bmatrix} 0.209 & 0.348 \\ 0.348 & 0.638 \end{bmatrix}$$

Using equation (4), the matrices U, L, V are obtained from the SVD of the matrix $Y^T X$, each of which is 2×2 in size.

$$U = \begin{bmatrix} -0.999 & -0.026 \\ -0.026 & 0.999 \end{bmatrix}, L = \begin{bmatrix} 0.912 & 0 \\ 0 & 0.121 \end{bmatrix}, V = \begin{bmatrix} -0.488 & 0.873 \\ -0.873 & -0.488 \end{bmatrix}$$

Furthermore, after decomposing the singular values, the matrix Q is obtained to minimize the distance obtained using equation (4).

$$Q = VU^T$$

$$Q = \begin{bmatrix} -0.488 & 0.873 \\ -0.873 & -0.488 \end{bmatrix} \begin{bmatrix} -0.999 & -0.026 \\ -0.026 & 0.999 \end{bmatrix} = \begin{bmatrix} 0.465 & 0.885 \\ 0.885 & -0.465 \end{bmatrix}$$

$$XQ^T Y^T = \begin{bmatrix} 0.175 & 0.106 & -0.209 & -0.147 & 0.232 & 0.065 & -0.222 \\ 0.041 & 0.043 & 0.065 & -0.020 & 0.036 & 0.035 & -0.069 \\ -0.196 & -0.155 & 0.267 & 0.136 & -0.223 & -0.112 & 0.282 \\ 0.013 & 0.040 & -0.045 & 0.013 & -0.013 & 0.039 & -0.047 \\ 0.152 & 0.086 & -0.176 & -0.131 & 0.207 & 0.050 & -0.187 \\ 0.014 & 0.039 & -0.045 & 0.011 & -0.011 & 0.038 & -0.047 \\ -0.200 & -0.159 & 0.274 & 0.139 & -0.228 & -0.115 & 0.289 \end{bmatrix}$$

$$2XQ^T Y^T = \begin{bmatrix} 0.350 & 0.212 & -0.418 & -0.294 & 0.464 & 0.130 & -0.444 \\ 0.082 & 0.086 & -0.131 & -0.041 & 0.072 & 0.069 & -0.138 \\ -0.391 & -0.312 & 0.535 & 0.272 & -0.445 & -0.224 & 0.566 \\ 0.027 & 0.081 & -0.090 & 0.026 & 0.027 & 0.078 & -0.095 \\ 0.304 & 0.173 & -0.352 & -0.264 & 0.413 & 0.100 & -0.374 \\ 0.029 & 0.079 & -0.090 & 0.022 & -0.022 & 0.076 & -0.095 \\ -0.400 & -0.319 & 0.547 & 0.278 & -0.456 & -0.229 & 0.579 \end{bmatrix}$$

Based on equation (5). the rotation value is obtained as follows:

$$M^2 = tr(\boldsymbol{XX^T} + \boldsymbol{YY^T} - 2\boldsymbol{XQ^TY^T}) = 1.108 + 1.148 - 2.065 = 0.191$$

The next step, after obtaining the translation and rotation results, is the dilation process, which aims to enlarge or reduce the sum of the squared distances. The value of *c* can be obtained using equation (6).

$$c = \frac{tr(\boldsymbol{XQ^TY^T})}{tr(\boldsymbol{YY^T})} = \frac{1.033}{1.148} = 0.899$$

Next, calculate the value of M^2 using equation (5). The result is

$$M^2 = 1.108 - 0.899^2(1.148) = 0.180$$

This yields an M^2 value of 0.180. The results of the sum of the squared distances for each transformation (M^2) can be seen in **Table 3**.

Table 3. The result of Procrustes analysis

Transformation Process	Value
Translation	1.449
Rotation	0.191
Dilation	0.180

Based on **Table 3**, the smallest M^2 value is 0.180. This value will be used to calculate the coefficient of determination (R^2) using equation (7).

$$R^2 = \left(1 - \frac{M^2_{min}}{tr(\boldsymbol{XX^T})}\right) \times 100\%$$

$$R^2 = \left(1 - \frac{0.180}{1.108}\right) \times 100\% = 0.837$$

Approximately 83.7% of the structural variation in one configuration can be explained by the other configuration after applying the Procrustes transformation, namely translation, rotation, and optimal scaling. Only 14% of the structural variation remains unexplained The relationship between R^2 and the Procrustes correlation R = 0.915 reflects a strong structural match between the two configurations.

The results of the Procrustes analysis indicate that approximately 83.7% of the structural variation in the cherry production data from the first field can be explained by the production structure from the second field after applying the translation, rotation, and optimal scaling transformations. This value indicates that the cherry production patterns in the two fields have a relatively high level of similarity, even though the two fields may differ in physical conditions, soil fertility, environment, cultivation practices, or crop maintenance management. So, the R^2 value is 0.837 or 83.7%, indicating that the green cherry production on Ismail's field and Wahid's field are relatively similar.

4 Conclusion

The biplot analysis on Ismail's field data (data A) shows a very strong correlation between the number of primary branches and the number of fruit on primary branches. The number of secondary branches also has a strong correlation with the number of fruit on primary branches. The number of secondary branches variable has a very strong correlation with the variable number of fruits on tertiary branches. Meanwhile, the results of the biplot analysis on Wahid's field data (data B) show a very strong correlation between the variable number of tertiary branches and the number of fruits on tertiary branches, as well as correlations with other variables.

The application of Procrustes analysis in a biplot graph on data matrices A and B indicates that the configurations used are similar. This is supported by an R^2 value of 83.7%. Therefore, it can be concluded that the green cherry production results on Ismail's field data and Wahid's field data are similar and there are no significant differences between the two data sets.

Acknowledgments. The authors would like to thank the Pagaralam coffee farmers, especially Mas Ismail and Mas Wahid in Desa Rimba Candi, as well as Mr. Ahmad Surkati and the team who helped in data collection.

References

[1] Ramadhana, A. W. S., Aulia, A. D., and Ulum, T. Keunggulan komparatif ekspor kopi di Indonesia. Journal of Economics, Business, Accounting and Management. Vol. 2, no. 1, pp. 110–123 (2024)

[2] Budiarto, T., Ayun, L., Nurulhaq, M.: Pemberdayaan petani pada pengolahan pascapanen kopi Arabika (Coffea arabica L.) di Desa Kalisat Kidul, Kecamatan Kalibening, Kabupaten Banjarnegara. Jurnal Resolusi Konflik, CSR, dan Pemberdayaan. Vol. 8, no. 1, pp. 11–20 (2023)

[3] Emanuella: Nasib Kopi Indonesia: Primadona Dunia yang Tengah Terancam. Cnbcindonesia. https://www.cnbcindonesia.com/research/20240822113220-128-565458/nasib-kopi-indonesia-primadona-dunia-yang-tengah-terancam (2024)

[4] Ligar, B. W.: Review identifikasi dan klasifikasikan biji kopi menggunakan computer vision. Jurnal Sistem dan Teknologi Informasi (JustIN). Vol. 11, no. 2, pp. 243-249 (2023)

[5] Sunanto, S., Salim, S., & Rauf, A. W.: Analisis kesepakatan peningkatan produktivitas kopi Arabika pada pengembangan kawasan di Kabupaten Toraja Utara. Jurnal Sosial Ekonomi Pertanian. Vol. 15, no. 1, pp. 42-55 (2019)

[6] Harahap, P. M., Analiansari, & Nugraha, T.: Karakteristik fisik biji kopi hasil produksi unit usaha di Kecamatan Kebun Tebu Lampung Barat. Kebun Tebu. Jurnal Pengembangan Agroindustri Terapan. Vol. 2, no. 2, pp. 25-33 (2023)

[7] Sirappa, M. P., Heryanto, R., & Silitonga, Y. R.: Standardisasi pengolahan biji kopi berkualitas. Warta BSIP Perkebunan. Vol. 2, no. 1, pp. 18–25 (2024)

[8] Langobiri, C. S., Irianto, I. K., & Wirajaya, A.: Respon tanaman melon *(Cucumis melo L.)* terhadap waktu pemangkasan tunas dan interval pemberian urine kelinci. Gema Agro. Vol. 24, pp. 9–16 (2019)

[9] Zulkarnain, Rahmaddiansyah, Alpian, R., & Bagio, B.: Perbandingan tingkat produktivitas dan pendapatan petani kopi Arabika yang melakukan teknik pemangkasan rutin dan yang tidak di Kecamatan Bener Kelipah. Agrifo. Vol. 5, no. 1, pp. 78–86 (2020)

[10] Irmeilyana, Suprihatin, B., Desiani, A., Ngudiantoro, & Maiyanti, S. I.: Graphical representation and two groups analysis on data matrix of Robusta green cherries production in two harvests. Barekeng. Vol. 19, no. 2, pp. 1279–1294 (2025)
[11] Hasibuan, Z., Manumono, D., & Arum Ambarsari: Usahatani kopi Robusta (studi kasus di Desa Madigondo, Kecamatan Samigaluh). Jurnal Masepi. Vol. 1, no. 1, pp. 1–13 (2019)
[12] Khayati, N., Wachjar, A., & Sudarsono: Pengelolaan pemangkasan tanaman kopi Arabika *(Coffea arabica L.)* di Kebun Kalisat Jampit, PT Perkebunan Nusantara XII (Persero), Bondowoso, Jawa Timur. Buletin Agrohorti. Vol. 7, no. 3, pp. 295–301 (2020)
[13] Arifin, R.: Maksimalisasi hasil panen kopi melalui pemangkasan cabang pasca panen dan pengendalian hama secara alami untuk meningkatkan produksi kopi. Jurnal Pembelajaran Pemberdayaan Masyarakat (JP2M). Vol. 1, no. 1, pp. 23-33 (2019)
[14] Jolliffe, I. T.: Principal Component Analysis, 2nd ed. Springer Verlag, New York (2002)
[15] Saputri, A. E., Salma, A., Amalita, N., & Permana, D.: Biplot and Procrustes analysis of Poverty Indicators by province in Indonesia in 2015 and 2019. UNP Journal of Statistics and Data Science. Vol. 2, no. 1, pp. 1–7 (2024).
[16] Irmeilyana, Ngudiantoro, Maiyanti, S.I.: Eksplorasi Grafik Data Multivariat Cuaca, Kemitraan, dan Produktivitas Tanaman Kopi (Studi Kasus: Kopi Robusta di Candi Jaya Kecamatan Dempo Tengah Kota Pagaralam). Laporan Penelitian Hibah Kompetitif 2024. Universitas Sriwijaya, Indralaya (2024)
[17] Husson, F., Le, S., and Pagès, J.: Exploratory Multivariate Analysis by Example Using R, 2th ed. Chapman & Hall/CRC Press, London (2017)
[18] Bakhtiar, T. and Siswadi: Orthogonal Procrustes analysis: its transformation arrangement and minimal distance. International J. Applied Mathematics and Statistics. Vol. 20, No. 11, pp. 16-24 (2011)
[19] Siswadi and Bakhtiar, T.: Goodness-of-fit of biplots via Procrustes analysis. Far East Journal of Mathematical Sciences (FJMS). Vol. 52, No. 2, pp. 191-201 (2011)
[20] Krzanowski, W. J.: Principle of Multivariate Analysis, A User's Perspective. Oxford University Press, New York (1990)
[21] Johnson, R. A. and Wichern, D. W.: Applied Multivariate Statistical Analysis, 6th ed. Prentice-Hall, Inc., New Jersey (2007)