Identifying The Key Variables for Assessing The Reclamation Success on Early Growth Vegetation in Ex-exploration of Oil and Gas Mining Areas

Tirta Negara¹, Cecep Kusmana², Irdika Mansur², Nitya Ade Santi¹, I Nengah Surati Jaya³

¹Graduate Program of Forest Management Science, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680
²Department of Silviculture, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680
³Department of Forest Management, Faculty of Forestry and Environment, IPB University, Academic Ring Road, Campus IPB Dramaga, Bogor, Indonesia 16680

Received January 24, 2020/Accepted October 7, 2020

Abstract

This paper examines the identification of key indicators that could be used to measure the success of reclamation plants in post-exploration oil and gas mining areas. The main objective of this research was to find key indicators or variables for evaluating the level success of reclamation results in the post-mining of oil and gas area. In this study, 44 environmental variables of the physical, biological, soil, water and air indicators were analyzed from 70 field plots of 6 reclamation and 2 natural forest sites. The analysis methods included (1) cluster analysis using the Agglomerative Hierarchical Clustering method with the Ward's method, and (2) quadratic discriminant analysis. The results of the clustering analysis showed that there were some clusters due to variation of biomass, water, soil and air conditions. The three clusters developed based on water and/or air variables provided high cophenetic correlation (0.80) with low within-cluster (14.5%) and high between-cluster variations (85.5%). Based on the multicollinearity analysis, average vector difference test, variance matrix variance test, unidimensional test of each variable and quadratic discriminant function, this study found that there were 3 key indicators determining variations of the quality of the reclamation plantations within the study sites, namely, biological indicator of biomass volume (Bio_B), soil indicator of P content in the soil (Tnh_P), saturation base of soil (Tnh_Kb), Manganese (Mn) content in the soil (Tnh_Mn), Sulfur content in the soil (Tnh_S), percentage of ash in the soil (Tnh_Ab), percentage of clay in the soil (Tnh_Li), and water indicator of chloride content in the surface water (Air_Cl). The examination on four classes of the reclamation quality showed that the classes were successfully classified having excellent cross-validation error matrix with overall accuracy more than 90%.

Keywords: key variable, agglomerative hierarchy clustering, ward's method, quadratic discriminant analysis, reclamation success, oil and gas mining

*Correspondence author; email: ins-jaya@apps.ipb.ac.id

Introduction

The role of the mining sector is contributing to the economic growth of the nation, and the mining activities always bring impacts on the biophysical, economic, and socio-cultural environment. In the biophysical aspect, mining activities cause changes in land cover directly. The level of change in land cover caused by mining activities is highly dependent on the type, method and stages of mining activities applied (Vasuki et al., 2019). Therefore, every exploration and/or mining activity is always followed by a forest reclamation activity, which is an effort to restore the former mine into more productive lands (Singh et al., 2010). Reclamation activities through vegetation planting is always followed by monitoring and evaluation of the success of reclamation plants. The monitoring activities need to be performed periodically to monitor the progress of the work; while the evaluation activities are intended to measure the degree of success of forest and/or land reclamation activities (Block et al., 2001).

By considering the varied site conditions such as mining material, mining methods and local biophysical conditions, the reclamation success evaluation techniques need to be formulated, so that the effectiveness of monitoring and the reliability of evaluation methods could be achieved. The monitoring technique need to be accurate, fast, transparent, and efficient. Although the reclamation activities would not be possible to precisely restore the biophysical environmental conditions as they were in the past, the reclamation results are expected to improve environmental quality conditions in a more productive direction.

Forest reclamation activities provide a large number of benefits for the sustainability of forest functions through revegetation (Meng et al., 2012; Bauman et al., 2015; Zhang et al., 2015; Clark & Zipper, 2016). Revegetation aims to
increase biodiversity, canopy cover, canopy stratification, and soil fertility, to accelerate the colonization and entry of wild animals, and to improve the condition of the forest environment.

Since the success of reclamation activity is highly related to the treatment of land preparation, species planting, maintenance, and type of mining activities, it is necessary to have monitoring and evaluation by the typology of the mining activities. Success in post-mining forest reclamation requires an assessment mechanism through monitoring activities (Macdonald et al., 2015). There have been several studies on the success rates of revegetation including Widdowson, (1990) the assessments of diversity index, canopy index, nutrient retention index (soil nutrient retention), colonization index, fauna index, and environmental index, and they are very influential indicators in monitoring the success of revegetation. Syaufina & Ikhsan, (2013) considered that the higher the success rate of revegetation activities in the ex-mining area, the greater the potential for carbon sequestration or biomass storing from the land. Based on the aspects of biodiversity, the canopy cover, layer of the canopy, and land quality, the success of revegetation activities in the ex-mining area could be achieved.

Improvement of the site condition on the ex-mining areas could be conducted by rehabilitating the damaged ecosystems. The reclamation activity is expected to restore the damaged ecosystems including the soils and its vegetation into their better condition. The reclamation in post-mining area is vulnerable to failure. Since the reclamation cost is quite expensive, there is a need to develop a monitoring system. The monitoring and evaluation system to assess the success of post-mining should be reliable i.e. to provide consistent result, easily implemented, and have measurable variables with a reasonable monitoring cost. Therefore, the identification of key factors of post-mining success may ensure the effectiveness of the monitoring and evaluation system. Iskandar & Suryaningtyas, (2012) noted that the fertility aspect of the planting media is a key factor that influences the success of revegetation of ex-mining land. Other physical, chemical, and biological qualities of the soil have significant effects on the plant growth (Taylor et al., 2010; Mansur, 2012).

Evaluation and monitoring of the success of reclamation in ex-gas and oil mining exploration area are very typical, where the area of land clearing is very narrow, but changes in the bio-physical environment, such as land clearing were very intensive, especially at the location around drilling site (Rathore & Wright, 1993). In addition, the mining exploration activities are not centralized and/or not compact that may cause the reclamation and monitoring activities to be time consuming and costly for mobility of ground monitoring. In addition, if the assessment of reclamation plants is conducted on reclamation plants that are still in early growth (young plantation), evaluation of the indicators of success is difficult to measure. The conventional method used to assess the vegetation success (Chambers et al., 1994; Hendrychová, 2008) is the percentage of plant growth, while indicators of its biophysical environmental aspects have not yet been measured. Since the reclamation activity requires high costs, and continuous maintenance time over a period of 2 consecutive years are required; therefore, an early and structured evaluation mechanism for reclamation success assessment needs to be conducted.

Based on the above conditions, to obtain more complete information in assessing the success of young reclamation plants, aged 12 years, this study established statistical analysis to identify key indicators or variables of success reclamation plants in ex-exploration oil and gas mining. In this study, the examined environmental variables were vegetation, soil, water, and air conditions. The soil variables included pH, total nitrogen, available phosphorus and potassium, exchangeable calcium, iron and aluminum, so they must be considered in evaluating the quality improvement of ex-mining areas (Dimitriu et al., 2010). The objective of this study was to identify key indicators and/or variables related to the success of reclamation plant. These key variables would be then used as variables for establishing the criteria and mechanisms for assessing the success of reclamation in the gas and oil exploration area.

**Methods**

**Study sites** The study was conducted at ex-wells of PetroChina International Jabung Ltd Jambi Province and 2 sites of natural forest. An overview of sampling sites is presented in Figure 1 and Table 1.

The sites included ex-oil and gas mining exploration for the period of 2013–2019. The reclamation in PetroChina sites was conducted under the Letter of Jambi Governor Number S-525/1457/SETDA-EKBang-4: V/2013 dated May 13, 2013 concerning the Rehabilitation and Reclamation of ex-exploration area. The activities in the ex-exploration wells were managed by PetroChina International Jabung Ltd. For the research, the field data collection was carried out from February 2019, and the analyses on soil data, water, fauna, canopy structure, and environmental quality as well as spatial analysis were conducted at the Soil Laboratory of the Faculty of Agriculture IPB University, Soil Physics & Chemistry and Soil Fertility Laboratory of the University of Jambi, Faculty of Agriculture, Regional Environment Laboratory of the Jambi Provincial Environment Office and Remote Sensing and GIS Laboratory of the Faculty of
Forestry and Environment of IPB University, Bogor from February 2019 to May 2019.

Reclamation at this study sites was performed at the location where oil-wells have been closed due its non-economical value. The study sites were at 6 reclamation sites of Ripah, Kajen, Sarinah, Kabul, Gerbang and NW Lambur, and 2 natural forests of UNJA Education Forest and peat land forest. The locations had plants aged of 2 years, with the conditions as presented in Figure 2. The planting space was done with a spacing of 4 m × 4 m or with a density of approximately 625 stems ha⁻¹, where the plants were planted in plant-holes with the size of 80 cm × 80 cm × 80 cm. The species of plants included 9 species of local plants such as swamp jelutung (Dyera lowii Hookmeta), jelutung of dry land (Dyera costulata), meranti (Shorea sp), ketapang (Terminalia catappa), trembesi (Samanea saman), and fast-growing plant species such as jabon (Neolamarckia cadamba), pulai (Alstonia scholaris) and saga hutan (Abrus precatorius) and lamtorogung (Leucaena leucocephala). The composition of local species (endemic) and multi-purpose tree species (MPTS) was 90% and 10%. In Figure 2, it could be seen that the grass land, shrubs, cover crops and or low vegetation were among the main plants mentioned above.

Supporting data To support study, the main data used were from the observation data of 70 field plot measurements. The study also used data derived from the unmanned aerial vehicle (UAV) imageries with a spatial resolution of 15 cm. The list of data is shown in Table 2.

In general, the pre-plantation conditions are mostly bare land with grasses and/or low vegetation (Figure 3a), while 2 years after planting, the reclamation areas were covered with small trees having crown diameter of 1.5 meter (Figure 3b). For comparison purposes, condition of vegetation in natural forest of Jambi Education Forest (UNJA) is shown in Figure 3c.

Measurement method The measured water quality parameter included pH, temperature, electrical conductivity, alkalinity, acidity, and dissolved oxygen. The variable measurements of water samples were conducted using a

Table 1 Location of ex-mining wells of PetroChina International Jabung Ltd

<table>
<thead>
<tr>
<th>Descriptions</th>
<th>Ripah</th>
<th>Kajen</th>
<th>Sarinah</th>
<th>North Kabul</th>
<th>Gerbang</th>
<th>NW Lambur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planting Year</td>
<td>2016</td>
<td>2016</td>
<td>2016</td>
<td>2016</td>
<td>2016</td>
<td>-</td>
</tr>
<tr>
<td>Villages</td>
<td>Toman River</td>
<td>Simpang Tuan</td>
<td>Bungur River</td>
<td>Purwodadi</td>
<td>Nilau Bay</td>
<td>Kota Baru</td>
</tr>
<tr>
<td>Districts</td>
<td>Mendahara Ulu</td>
<td>Mendahara Ulu</td>
<td>Mauaro</td>
<td>Tebing Tinggi</td>
<td>Pengahuan</td>
<td>Geragai</td>
</tr>
<tr>
<td>Regencies</td>
<td>West Jabung</td>
<td>East Jabung</td>
<td>Jambi</td>
<td>Jabung</td>
<td>Jabung</td>
<td>Jabung</td>
</tr>
<tr>
<td></td>
<td>East Jabung</td>
<td>East Jabung</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2 The photo of the research field.
thermometer, a pH meter, a water level meter, or rope equipped with ballasts and GPS. Other complementary tools included coolant (cool storage box, 4 °C ± 2 °C) and filter device with suction/compressed pump holding the filter with a pore size of 0.45 µm. The samples analyzed in the laboratory were preserved with HNO₃, H₂SO₄, NaOH, and Zn Acetate (Rice et al. 2017).

Soil quality parameters were also derived from field measurements taken from 6 typologies within the study sites. Based on the field observations, there were 6 typologies of dry mineral soil piled with vegetation, dry mineral soil without piled vegetation, wet mineral soil with vegetation, forest/natural mineral soil, wet forest/natural mineral soil, and clay soil. Types of soil samples taken were intact and damaged soils. The amounts of intact soil samples and damaged soil samples per exploration well were in line with the typology of land in the area.

The soil samples taken in the field included composite soil samples from the location of 0–60 cm depth and intact soil samples from the locations of 0–30 cm and 30–60 cm as well as the chemical materials for the laboratory analysis. The equipment used for soil sample collection included Global Positioning System (GPS) receiver, compass, mineral soil drill, clay drill, meter, field knife, Soil Munsell Color Chart book, sample ring, plastic bag, label paper, rubber band, description card, equipment stationery, cameras, and equipment for analyzing the soil samples in the laboratory.

The observational data were analyzed descriptively by comparing the morphological changes, some physical properties of soil and nutrients contained in the soil between the reclamation land of the Petrochina wells and the land in natural forests (mineral soil forests and clay) (Lembaga Penelitian Tanah, 1983).

Data processing and analysis The primary analytical methods used in this study were cluster and discriminant analyses. The cluster analysis was used to find out how the natural hierarchical structure of the classes of reclamation quality and their environmental conditions, while discriminant analysis was used to find out the variables that...
most influenced the grouping of each reclamation plant condition.

Prior to performing other analyses, the cluster analysis was preceded by evaluating normality of the data, multicollinearity between independent variables, and by evaluating the mean difference of each variable independently (F-test). The steps on applied cluster analysis were (1) determining the measure of the degree of similarity between sample groups, i.e. Euclidean Distance; (2) conducting the clustering process with the Agglomerative Hierarchical Clustering (AHC) approach using the Ward's method; (3) calculating the Cophenetic Distance, Cophenetic correlation and variance decomposition within-cluster variance, between-cluster variances; (4) Labelling each cluster name and identifying the profile descriptions for each cluster. In the cluster analysis, all 70 samples were treated as initial clusters.

Cluster analysis Grouping was conducted using the Agglomerative Hierarchical Clustering (AHC) approach (Davidson & Ravi, 2005) and the Cophenetic distance on the dendrogram using the Ward's algorithm (Murtagh & Legendre, 2014). Ward method is based on variance, where the cluster will be combined if it contributes to the smallest variance. Merging process was not based on distance considerations as in the single, complete, or average method but on the basis of the minimum variance increased. In AHC method, the initial cluster continues to be combined with other clusters so that one cluster has multiple members and eventually become one large cluster. Each cluster was separated by their distance. Clusters with the closest distance would be merged using linkage algorithm. The distance between clusters was calculated using the Euclidean formula as shown in Equation [1].

\[ d_{ij} = \left( \sum_{k=1}^{n} (x_{ik} - x_{jk})^2 \right)^{1/2} \]  

note: \( d_{ij} \) = distance between \( i-th \) and \( j-th \) cluster, \( x_{ij} \) = variable \( k-th \) at the \( i \) and \( j \) class

In the clustering process, the success of class merging was evaluated using the Cophenetic correlation value, which expressed the relationship between the distance matrices, and between the original Euclidean Distance with the distance between clusters after merging with the Ward's method. The correlation between distance matrix and Cophenetic matrix was computed as shown in Equation [2].

\[ CC = Correl(DM, CM) \]

note: \( CC \) = Cophenetic correlation, \( DM \) = distance matrix, and \( CM \) = merging cophenetic matrix. The formula for calculating correlations (Correl) was the same as the conventional formula for calculating correlations between paired values. In this case, the paired value was the location of the same cell from two different matrices (Saraçlı et al., 2013).

Discriminant analysis In discriminant analysis, the steps of analysis include (1) defining the category of each class that derived from cluster analysis and identifying the independent variables to be selected; (2) conducting multicolinearity tests with tolerance and variance inflation factor (VIF) measures; (3) conducting mean vector similarity tests using the Wilks’ Lambda test, Pillai’s test, Hotelling-Lawley test, Roy’s greatest root; (4) variance-covariance difference test using Box test, either with Chi-sq. or Fisher’s asymptotic approximation and Kullback’s test; (5) constructing discriminant function equations the most significant variables; (6) establishing an error matrix to measure the accuracy of discriminant functions, both using training area sample data, and using cross-validation methods to measure the stability of the function and its ability to be generalized.

As described previously, the AHC approach shown that there were variations in environmental conditions of the reclamation. By considering reclamation condition obtained from obtained from the AHC analysis, then this study developed four classes expressing the reclamation quality. Since the plant spacing and species of plant have been determined in accordance with government policy, then study the success of reclamation was primarily categorized on the basis of the biomass, stem diameter size, proportional of living tree, crop coverage, then followed by and other bio-physical variables. Table 3 summarize the criteria of 4 classes using biomass and number of living trees. The densest class, describing the climax condition was belonged to natural forest class. The classes used in this discriminant analysis were tabulated in Table 3.

The discriminant analysis applied in this study was conducted statistically to find out what variables affected class differentiation, which variables were the most dominant and to what extent accuracy could be achieved using these selected variables. This discrimination is a classic method introduced by Fisher, (1936) useful as an explanatory and predictive method and can be used to check whether the classes differ clearly. The classifier can also identify the class variables affecting grouping accuracy and predict whether certain observations belong to a certain class or not. The discriminant analysis may include basic tests, such as normality test, variance test, and collinearity test. Discriminant analysis requires statistical parameters such as the mean value of each variable in each class, variance-covariance matrix of each class, determinant of matrix, pooled matrix of variance, and prior probability.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Biomass (ton ha(^{-1}))</th>
<th>Number of samples</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>&lt; 13.0</td>
<td>20</td>
<td>28.6</td>
</tr>
<tr>
<td>C2</td>
<td>13.0 - 21.6</td>
<td>14</td>
<td>20.0</td>
</tr>
<tr>
<td>C3</td>
<td>21.6 - 100.0</td>
<td>25</td>
<td>35.7</td>
</tr>
<tr>
<td>C4</td>
<td>100.0 up</td>
<td>11</td>
<td>15.7</td>
</tr>
</tbody>
</table>

Table 3 The description of each cluster obtained from the AHC
use a quadratic model (Srivastava et al., 2007). Initially, the type of discriminant function examined was a linear model, but when the variance-covariance between classes were different, the quadratic discriminant function was applied. The LDA & QDA functions were same as the following functions:

\[ P(x|y = k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)\right) \]

In QDA, within-class covariance \(\Sigma_k\) is assumed to be different, whereas in LDA, all covariance is assumed to be the same so that the discriminatory formula can be simplified. The value \((x - \mu_k)\) is a vector of the difference in the value of the variable \(x\) with the mean value in class \(k\) \((\mu_k)\), \(t\) is the transpose of the matrix and \(\Sigma\) is the inverse of the variance-covariance matrix, \(\Sigma^{-1}\) is the value of matrix determinant. In QDA, the model can be written as follows:

\[
\log P(y = k|x) = \log P(x|y = k) + \log P(y = k) + Cst
\]

\[
= -\frac{1}{2} \log|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k) + \log P(y = k) + Cst
\]

\(P\) is the prior probability obtained from the proportion per class and \(Cst\) is a constant. In LDA, because the variance-covariance matrix is assumed to be the same, the formula is simplified to:

\[
\log P(y = k|x) = -\frac{1}{2}(x - \mu_k)^T \Sigma^{-1}(x - \mu_k) + \log P(y = k) + Cst
\]

In other words, classes in LDA will be differentiated by linear lines, whereas in QDA, classes will be limited by curved or non-linear lines.

Multicollinearity Both in the linear model and quadratic model, the multicollinearity between dependent variables will cause zero diversity. Multicollinearity test could be conducted by searching the high correlation coefficient values (larger than 0.7) or by looking at Tolerance values greater than 0.1 or by using the Variance Inflation Factor (VIF) values smaller than 10. In the XLSTAT used in this study, variables that had a close correlation was automatically excluded in the analysis. The variables that caused multicollinearity were eliminated automatically by XLSTAT.

Evaluate the equality of mean vector of variables and within-class variance in each class To ascertain whether the classes built have significant differences and to find out the assumptions that would be used in developing the discriminant function, the mean vector difference test and the within-class variance difference test between classes were performed. The mean vector similarity test values using the Wilks’ Lambda, Pillai’s Trace, Hoteling’s Trace and Roy’s Largest Root tests indicated that the classes established had significant differences with \(p\)-values smaller than 0.05. It was a clear finding that both the mean vector and within-class were significantly different from each other (Huberty & Olejnik, 2006).

Confusion matrix and cross validation To find out whether the discriminant function was reliable (stable) as well as could be generalized, the model validation was performed. The validation was commonly conducted by two methods: (1) the subset validation method and (2) the cross-validation method. The subset validation method was carried out by dividing the observation data into a subset of data to establish a discriminant mode (training data set) and a subset of data for the reliability test (data set test). Since the number of samples per class was limited (only 1419 samples per group), the subset validation method was not carried out, then the cross-validation method was selected, where all training data sets were treated as test data sets.

In cross-validation, each data set was excluded from the training data set to test whether the group was classified as the correct group or not. This cross validation was important in discriminant analysis, because very often the model built must be able to be measured for its reliability (reliability) and generalized ability (generalizability). In this cross validation, the samples were divided according to important variables. If each subset of variables produces the same results, then the function is stable (reliable) and can be generalized. This cross-validation approach is adopted when researchers suspect the finding that some variables do not apply to all levels and variables.

The number of correctly and miss classified data was tabulated in the form of error matrix. This matrix would assess the accuracy of discriminant function, both using training data sets and cross-validation results. The more thorough the observations in the diagonal element of the matrix, the better the discriminant function obtained. From the confusion matrix, the values of overall accuracy (OA), producer’s accuracy (PA) and user’s accuracy (UA) would be derived. With cross-validation, we could be able to know the prediction from an observation, if the observation is removed from the estimation sample. The higher the value of OA, UA, and PA in the cross-validation matrix, the better the reliability of the discriminant function that is built based on the selected key variables.

Results and Discussion Exploration of reclamation conditions based on AHC The initial analysis using the AHC approach found that the reclamation conditions were clustered following the variation of the environmental variables considered in this study. On the basis of the dissimilarity measure, from the 70 sample plots of 8 sites, it was shown that the similarity or dissimilarity between clusters was affected by the variables applied.

The study found that of all 44 measured variables, 9 variables were not used due to the multicollinearity reason, having VIF more than 10 or tolerance values larger than 0.1. Only the remaining 35 variables were examined in the further steps: Bio_B, Tnh_Nt, Tnh_P, Tnh_K, Tnh_Ca, Tnh_Mg, Tnh_Kg, Tnh_Kik, Tnh_Kb, Tnh_AL, Tnh_H, Tnh_Fe, Tnh_Mn, Tnh_Psr, Tnh_Db, Tnh_Li, Tnh_S, Tnh_Ab, Air_Tds, Air_kkr, Air_wrn, Air_Shu, Air_ph, Air_DO, Air_NO3, Air_NO2, Air_NH3, Air_Cu, Air_Fe, Air_Mn, Air_Zn, Air_Cl, Air_Mg, Air_CCO3, and Air_KMn. It was noted that when the 44 variables were
used, then no significant clusters were found due to within-class variance value of 57.7%, much larger than between-class variance, i.e. 42.3%. However, when the analysis used only 35 variables (variables with no collinearity) included in the AHC, the decomposition of variance provided the between-class variance was 52.3% larger than the within-class variance 40.7%. These findings expressed that there are variations of biomass, soil, water, and air variables among clusters (Figure 4a).

Furthermore, since the biomass would be treated as a response (dependent) variable that described the class of reclamation success, the biomass variable was omitted from the AHC. With only 34 variables where biomass was omitted, the decomposition of the variance of the AHC method provided very clear cluster having very homogenous the within-class variance, of only 14.5% while the between-class variance was 85.5%. These findings expressed that the variations of the non-biomass variables strongly affected the dissimilarity among clusters (Figure 4b).

Figure 4 Dendrogram with Ward Method.

Table 4 Decomposed variance obtained from using 44, 35, and 34 variables using AHC with Ward method.

<table>
<thead>
<tr>
<th>Variables</th>
<th>44 variables</th>
<th>35 variables</th>
<th>34 variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within class</td>
<td>57.70%</td>
<td>40.7%</td>
<td>14.50%</td>
</tr>
<tr>
<td>Between class</td>
<td>42.30%</td>
<td>59.27%</td>
<td>85.50%</td>
</tr>
<tr>
<td>Total</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Remarks: *) initial variables, **) non-multicollinearity variables, & ***) non-biomass & non-multicollinearity variables

Environmental quality of reclamation plants To assess the quality of reclamation condition on the basis of the biomass content, following the results obtained from the AHC analysis, the study developed 4 classes of reclamation plants conditions as summarized in Table 5.

A discriminant analysis was carried out to find out which environmental variables influenced the environmental quality of recovery plants, as described in Table 5. Examination of the assumptions to be fulfilled in this study including the normality test, the multicollinearity test, and the diversity difference test produced the following results:

The test for evaluating the mean vector and within-class variance difference Based on the test of differences in the variance-covariance matrices between classes, using the Box Test (Box's test) with Chi-square asymptotic approximation approach, Fisher's F and Kullback Test, it was known that at least a pair of the variance-covariance matrix was different with p-value smaller than 0.05. This means that the H1 hypothesis which stated the matrix of variance-covariance was statistically different was accepted. This study shows that the non-linear discriminatory functions that must be used in this study were quadratic discriminant functions, since the linear discriminant function could only be used if there is no difference between the within-class variance classes.

Quadratic discriminant function In the construction of the quadratic discriminant function, because the smallest number of observations per class from the 10 observations was K4, the maximum number of variables that could be examined was 9 variables, while the number of variables that had been screened on the multicollinearity and unidimensional tests were 34 non-biomass variables, the testing of the variables in the quadratic discriminant function (QDF) was carried out in several parts. By considering the role of each variable that had small p-value of < 0.0001, cross-validation matrix values, the study found that 9 variables that most affected the classification accuracy included Phosphor of the soil (Tnh_P), base saturation of soil (Tnh_Kb), Manganese of the soil (Tnh_Mn), percent clay in the soil (Tnh_Li), Sulfur content on the soil (Tnh_S), ash content of the soil (Tnh_Ab), Chloride content of the water (Air_Cl), Zinc content in the water (Air_Zn), Ammonia
content of the water (Air_Cl). By considering the variance of the Air_Zn & Air_NH, which was exceedingly small, those variables were omitted. Finally, only 7 variables were maintained in the quadratic discriminant functions: Tnh_P, Tnh_Kb, Tnh_Mn, Tnh_Li, Tnh_S, Tnh_Ab, and Air_Cl (Table 6).

The base saturation of soil (Tnh_Kb) is largely determined by the amount of base cation and the pH of the soil. The higher the pH of the soil, the higher base saturation and vice versa. Base saturation preferably reflects base cations that are useful as nutrients or as parameters for understanding soil fertility (Budiana, 2017). The higher the clay content in the soil (Tnh_Li), the higher the water porosity and root growth. The water porosity influences the water content of a soil, and water is a solvent for various organic molecular compounds, transport for photosynthesis, maintenance of the plumpness and a main component of photosynthesis (Salisbury & Ross, 1995). Clay particles are small and have large surfaces that support cation exchange (Hardjowigeno, 2003). One of the nutrients needed for plant growth is sulfur (Tnh_S). Sulfur is classified as a secondary macronutrient for plants, especially for young plants, needed after revising mine exploration (Herwanda et al., 2017). The element of Mn (Tnh_Mn) plays a role in various processes such as photosynthesis, assimilation and breathing (Gardner et al., 2017). The ash content (Tnh_Ab) describes the mineral content of an ingredient. Minerals play a significant role in plant growth, maintenance of cell functions, tissues, organs, and overall plant functions.

To facilitate the description of the grouping of each class, Bartlett’s statistic was developed. The F1 and F2 functions were able to explain the diversity of data from the initial variables as much as 97.8%, of which 82.7% were from F1 and 15.1% from F2 (Table 7), with the coefficient of the canonical discriminant function presented in Table 8.

By considering the standardized canonical function (F1~F3), on the F1 axis, the most affecting factor was soil sulfur having, with negative relationship followed by the variable of the percent clay in the soil (Tnh_Li) with positive relationship and manganese content in the soil (Tnh_Mn) with negative relationship (Figure 5).

By simply putting the original data and centroid on each of the K1 ~ K4 categories on the F1 and F2 axes, the scattered plots of each reclamation observation member in K1 through K4 using F1 and F2 are presented in Figure 6 and Figure 7.

Cross validation By using cross-validation analysis, an overall accuracy of approximately 92.9% with PA ranging from 78.6% to 100% and UA from 88.9% to 100% (Table 9) was obtained. The average of producers’ and user's accuracies were 92.4% and 93.9%. This cross-validation results ensure that the developed quadratic discriminant functions were quite reliable (stable) and could be generalized to assess the success of reclamation plants. This indicated that the variables used in this functions such as Phosphorus content on the soil (Tnh_P), base saturation of soil (Tnh_Kb), Manganese content on the soil (Tnh_Mn), percent clay in the soil (Tnh_Li), Sulfur content on the soil (Tnh_S), ash content of the soil (Tnh_Ab) as well as Chloride content of the water (Air_Cl) could be used to determine the success index of the quality of the reclamation plants.

Conclusion

From the foregoing discussion, the study concluded that the conditions of young reclamation plants, with only 2 years old, were varied following the variations of their environmental factors particularly on the soil and water
indicators. The AHC approach concluded that the non-biomass variables were significantly clustered into 3 homogenous clusters group having exceptionally low within-class variance of only 14.5% and extremely high between-class variance of 85.5%. The cophenetic correlation value provided by ward method was 0.8. The four categories of reclamation plant conditions developed in this study ranged from unsuccessful class (K1), moderately successful class (K2), successful class (K3) and natural forest class (K4) could be significantly discriminated by using seven key variables of soil and water indicators of phosphorus content on the soil (Tnh P), base saturation of soil (Tnh Kb), Manganese content on the soil (Tnh Mn), percent clay in the soil (Tnh Li), Sulfur content on the soil (Tnh S), ash content of the soil (Tnh Ab) as well as Chloride content of the water (Air Cl). The variations of those variables among categories differed significantly. There was no significant variable obtained from the air indicator. The study also

<table>
<thead>
<tr>
<th>Parameters</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>9.33</td>
<td>1.71</td>
<td>0.25</td>
</tr>
<tr>
<td>Discrimination (%)</td>
<td>82.67</td>
<td>15.12</td>
<td>2.21</td>
</tr>
<tr>
<td>Cumulative %</td>
<td>82.67</td>
<td>97.79</td>
<td>100.00</td>
</tr>
<tr>
<td>Bartlett's statistic</td>
<td>225.67</td>
<td>77.39</td>
<td>14.15</td>
</tr>
<tr>
<td>P-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 8 The standardized canonical discriminant function

<table>
<thead>
<tr>
<th>Parameter</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tnh P</td>
<td>1.129</td>
<td>6.359</td>
<td>0.689</td>
</tr>
<tr>
<td>Tnh Mn</td>
<td>4.244</td>
<td>-0.031</td>
<td>0.000</td>
</tr>
<tr>
<td>Tnh Li</td>
<td>-3.844</td>
<td>-3.495</td>
<td>-1.330</td>
</tr>
<tr>
<td>Tnh S</td>
<td>4.871</td>
<td>3.479</td>
<td>0.482</td>
</tr>
<tr>
<td>Tnh Ab</td>
<td>-5.278</td>
<td>-7.109</td>
<td>-1.085</td>
</tr>
<tr>
<td>Air Cl</td>
<td>2.112</td>
<td>-0.279</td>
<td>0.315</td>
</tr>
</tbody>
</table>

Table 9 Confusion matrix for the cross-validation results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>K1</th>
<th>K2</th>
<th>K3</th>
<th>K4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>19</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>K2</td>
<td>1</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>K3</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>K4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>12</td>
<td>27</td>
<td>11</td>
<td>70</td>
</tr>
<tr>
<td>UA</td>
<td>95.0%</td>
<td>91.7%</td>
<td>88.9%</td>
<td>100.0%</td>
<td></td>
</tr>
<tr>
<td>Avg. UA</td>
<td>93.9%</td>
<td>Avg. PA 92.4%</td>
<td>OA 92.9%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
concluded that the seven variables mentioned above were capable of classifying the success rate of reclamation with particularly good classification accuracy. The quadratic discriminant function model obtained in this study could explain the quality classes of reclamation growth with an overall average of 92.9%, with producer's ranging from 78.6% to 100% and user's from 88.9% to 100%.

References


