Analysis of Grouped Growth Patterns in Even-Aged Sugi Forest Stand within the Framework of Mixture Model

Yanagihara, H.[1], Ninomiya, Y.[2] & Yoshimoto, A.[3]
[1]Department of Mathematics, Graduate School of Science, Hiroshima University
[2]Graduate School of Mathematics, Kyushu University
[3]Graduate School of Environmental Studies, Tohoku University

Keywords: Growth analysis, Information criterion, k-means clustering, Multivariate Gaussian mixture model, Model selection.

Abstract: Even in the same even-aged forest stand, trees grow differently due to their own growth capacity, relative spatial conditions and other growth environments. In this paper, we investigate the growth patterns in an even-aged forest stand by a Gaussian mixture model (GMM) in order to identify groups of trees for the same growing process. Given the Richards growth function for the growth process, the derived coefficients of the function are used as new response variables in the multivariate GMM for identification of growth patterns. The optimal number of the grouped growth patterns is searched by minimizing the cross-validation (CV) criterion. We demonstrate the use of the proposed method to the growth data from a sugi (*Cryptomeria japonica*) sample plot in Hoshino village, Fukuoka prefecture, Japan. The resulted number of the patterns became three in our sample plot.

1. Introduction

Even in the same even-aged forest stand, trees grow differently due to their own growth capacity, relative spatial conditions and other growth environments. From the management viewpoint, if these growth differences can be captured, it would be beneficial to consider them in growth prediction. In this paper, we use a Gaussian mixture model (GMM; see e.g., Everitt & Hand, 1981) in order to investigate differences in the growth process in an even-aged forest stand. We focus on the growth patters in the forest stand.

A clustering method through the GMM has been widely used in several research

fields, including forestry, e.g., Zhang *et al.* (2004). Since tree growth data are repeated measures, time interval for measurement sometimes becomes unequal and the number of measures becomes large. The data over unequal time interval is called "unbalanced data", and that with the large number is "high dimensional data". Although a multivariate GMM (MGMM) is preferred for clustering, it is inefficient for the unbalanced and high dimensional data. This is mainly because an estimated covariance matrix in MGMM becomes singular or sometimes becomes non-obtainable for the unbalanced or high dimensional data.

In order to overcome the above difficulty, we reduce and equate the dimension of all data by assuming the Richards growth function (Richards, 1958) for the target growth data as in Yanagihara and Yoshimoto (2005a), and Yoshimoto *et al.* (2005). In this approach, the derived coefficients of the function are regarded as new response variables in the MGMM for the classification of growth patterns. Since a cluster for each individual is unknown, we apply the expectation-maximization (EM) algorithm proposed by Dempster *et al.* (1977) in order to have the maximum likelihood (ML) estimators of unknown parameters in the MGMM. The optimal number of the grouped growth patterns is searched by minimizing the cross-validation (CV) criterion proposed by Stone (1974). This is because the number of sampled trees is not so large, so that other information criteria are not appropriate.

The paper is organized as follows: In Section 2, we elaborate the classification method through MGMM and the method to determine the optimal number of clusters by minimizing the information criterion, followed by explanation of MGMM applied to the classification of the growth patterns. In Section 3, we demonstrate the use of the proposed method for the data obtained from the forest stand owned by Hoshino village in Fukuoka prefecture, Kyushu, Japan.

2. Theoretical Background

2.1. Classification through MGMM

Let $\boldsymbol{x}_i=(x_{i1},\ldots,x_{ip})'$ be a $p\times 1$ observation vector of the i-th individual ($i=1,\ldots,n$), and $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ be mutually independent, where n is the sample size, and "$'$" denotes a transposition of matrix or vector. Suppose that each individual belongs to one of k

populations (or clusters) $\Pi_1,\ldots,\Pi_k$ and the distribution of Π_j (j=1,…,k) is the p-dimensional normal distribution with the mean $\boldsymbol{\mu}_j$ and the variance-covariance matrix $\boldsymbol{\Sigma}$. It is well known that a probability density function for the case of $\boldsymbol{z}\sim N_p(\boldsymbol{\mu}_j,\boldsymbol{\Sigma})$ is given by

$$\phi(\boldsymbol{z};\boldsymbol{\mu}_j,\boldsymbol{\Sigma})=\left(\frac{1}{2\pi}\right)^{p/2} \mid \boldsymbol{\Sigma} \mid^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_j)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_j)\right\}. \quad [1]$$

Since we do not know which population the individual belongs to in general, we treat unknown partition of clusters as a $k\times1$ random variable vector $\boldsymbol{\delta}$. Here we assume that $\boldsymbol{\delta}$ distributes according to the k-dimensional multinomial distribution $MN_k(1,\boldsymbol{\rho})$, where $\boldsymbol{\rho}=(\rho_1,\ldots,\rho_k)'$ is the cell probabilities restricted to $\rho_1+\cdots+\rho_k=1$. Let $\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_n$ be $k\times1$ independent random vectors from $\boldsymbol{\delta}$, where $\boldsymbol{\delta}_i=(\delta_{i1},\ldots,\delta_{ik})'$ is an index vector denoting the population for the i-th individual. The following conditional probability density function of $\boldsymbol{x}_i$ given $\boldsymbol{\delta}_i$ (i=1,…,n) is derived:

$$g(\boldsymbol{x}_i;\boldsymbol{\Xi},\boldsymbol{\Sigma} \mid \boldsymbol{\delta}_i)=\phi(\boldsymbol{x}_i;\boldsymbol{\Xi}'\boldsymbol{\delta}_i,\boldsymbol{\Sigma})=\sum_{j=1}^{k}\delta_{ij}\phi(\boldsymbol{x}_i;\boldsymbol{\mu}_j,\boldsymbol{\Sigma}), \quad [2]$$

where $\boldsymbol{\Xi}$ is a $k\times p$ location parameter matrix defined by $\boldsymbol{\Xi}=(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_k)'$. Since equation [2] is defined by mixing Gaussian densities, the cell probabilities $\boldsymbol{\rho}$ is called a contaminate ratio. Let $\boldsymbol{e}_j$ (j=1,…,k) be a $k\times1$ vector such that the j-th element is 1 and the others are 0. Notice that

$$P(\boldsymbol{x}_i\in\Pi_j)=P(\boldsymbol{\delta}_i=\boldsymbol{e}_j)=\rho_j,\ (i=1,\ldots,n;j=1,\ldots,k). \quad [3]$$

From equations [2] and [3], a marginal probability density function of $\boldsymbol{x}_i$ is given by

$$f(\boldsymbol{x}_i;\boldsymbol{\rho},\boldsymbol{\Xi},\boldsymbol{\Sigma})=\sum_{j=1}^{k}g(\boldsymbol{x}_i;\boldsymbol{\Xi},\boldsymbol{\Sigma} \mid \boldsymbol{e}_j)P(\boldsymbol{\delta}_i=\boldsymbol{e}_j)=\sum_{j=1}^{k}\rho_j\phi(\boldsymbol{x}_i;\boldsymbol{\mu}_j,\boldsymbol{\Sigma}). \quad [4]$$

Therefore, a statistical model to identify clusters becomes as follows:

$$M_k:\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\sim i.i.d.\ f(\boldsymbol{x};\boldsymbol{\rho},\boldsymbol{\Xi},\boldsymbol{\Sigma}), \quad [5]$$

where an abbreviation "*i.i.d.*" stands for "independently and identically distributed". Let us call the model [5] the MGMM with k clusters.

Regarding parameter estimation, for obtaining ML estimates of $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$, we would prefer to utilize the following full log-likelihood:

$$\log\left\{\prod_{i=1}^{n} f_{\mathrm{F}}(\boldsymbol{x}_i,\boldsymbol{\delta}_i;\boldsymbol{\rho},\boldsymbol{\Xi},\boldsymbol{\Sigma})\right\} = \log\left[\prod_{i=1}^{n}\left\{g(\boldsymbol{x}_i;\boldsymbol{\Xi},\boldsymbol{\Sigma}\mid\boldsymbol{\delta}_i)\prod_{j=1}^{k}\rho_j^{\delta_{ij}}\right\}\right] = \sum_{i=1}^{n}\log g(\boldsymbol{x}_i;\boldsymbol{\Xi},\boldsymbol{\Sigma}\mid\boldsymbol{\delta}_i) + \sum_{i=1}^{n}\sum_{j=1}^{k}\delta_{ij}\log\rho_j. \quad [6]$$

However, we cannot use the full likelihood for estimating $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$ because $\boldsymbol{\delta}_i$ is unobserved. We thus apply the EM algorithm for estimating unknown parameters. The EM algorithm is widely used algorithm for obtaining the ML estimates from incomplete data by regarding $\boldsymbol{\delta}_i$ as missing values. In what follows, the estimation steps of the algorithm for the ML estimates are elaborated:

The EM Algorithm for Obtaining ML Estimates in MGMM

Step 1. We determine the initial clusters for the observations. In our algorithm, this is specified by the k-means method (MacQueen, 1967) using an determinant as the cluster criterion. For the detailed description of k-means using the determinant, refer to Yanagihara & Yoshimoto (2005). Let $\hat{\boldsymbol{\delta}}_1^{(0)},\ldots,\hat{\boldsymbol{\delta}}_n^{(0)}$ denote the given initial partition and the corresponding matrix be $\hat{\boldsymbol{\Delta}}^{(0)} = (\hat{\boldsymbol{\delta}}_1^{(0)},\ldots,\hat{\boldsymbol{\delta}}_n^{(0)})'$. Then the initial contaminate ratio, location and covariance matrices are determined as follows:

$$\begin{aligned}\hat{\boldsymbol{\rho}}^{(0)} &= (\hat{\rho}_1^{(0)},\ldots,\hat{\rho}_k^{(0)})' = \frac{1}{n}\hat{\boldsymbol{\Delta}}^{(0)\prime}\mathbf{1}_n,\\ \hat{\boldsymbol{\Xi}}^{(0)} &= (\hat{\boldsymbol{\mu}}_1^{(0)},\ldots,\hat{\boldsymbol{\mu}}_k^{(0)})' = (\hat{\boldsymbol{\Delta}}^{(0)\prime}\hat{\boldsymbol{\Delta}}^{(0)})^{-1}\hat{\boldsymbol{\Delta}}^{(0)\prime}\boldsymbol{X},\\ \hat{\boldsymbol{\Sigma}}^{(0)} &= \frac{1}{n}\boldsymbol{X}'\{\boldsymbol{I}_n - \hat{\boldsymbol{\Delta}}^{(0)}(\hat{\boldsymbol{\Delta}}^{(0)\prime}\hat{\boldsymbol{\Delta}}^{(0)})^{-1}\hat{\boldsymbol{\Delta}}^{(0)\prime}\}\boldsymbol{X},\end{aligned} \quad [7]$$

where $\boldsymbol{X}$ is an $n\times k$ matrix given by $\boldsymbol{X}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)'$, $\boldsymbol{I}_n$ is the n-th unit matrix and $\mathbf{1}_n$ is an $n\times 1$ vector with all elements equal to 1. For simplicity, we write $(p+1)(k+p/2)\times 1$ vector stacking estimates of $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$ as $\hat{\boldsymbol{\theta}}^{(0)} = (\hat{\boldsymbol{\rho}}^{(0)\prime},\mathrm{vec}(\hat{\boldsymbol{\Xi}}^{(0)})',\mathrm{vech}(\hat{\boldsymbol{\Sigma}}^{(0)})')'$, where $\mathrm{vec}(\boldsymbol{A})$ operator is to transform a matrix to a vector by stacking the 1st to the last column sequentially, and $\mathrm{vech}(\boldsymbol{B})$ is an operator to transform the lower triangle matrix of symmetric matrix to a vector by stacking the 1st to the last column (see, Harville, 1997).

Step 2. (Expectation-Step; E-Step): In the m-th repetition we have $\hat{\boldsymbol{\rho}}^{(m)} = (\hat{\rho}_1^{(m)},\ldots,\hat{\rho}_k^{(m)})'$, $\hat{\boldsymbol{\Xi}}^{(m)} = (\hat{\boldsymbol{\mu}}_1^{(m)},\ldots,\hat{\boldsymbol{\mu}}_k^{(m)})'$ and $\hat{\boldsymbol{\Sigma}}^{(m)}$, which are estimates of $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$, respectively. We set $\hat{\boldsymbol{\theta}}^{(m)} = (\hat{\boldsymbol{\rho}}^{(m)\prime},\mathrm{vec}(\hat{\boldsymbol{\Xi}}^{(m)})',\mathrm{vech}(\hat{\boldsymbol{\Sigma}}^{(m)})')'$. Let $\hat{\boldsymbol{w}}_j^{(m)}$ denote an estimated posterior probability in the m-th repetition by

$$\hat{w}_{ij}^{(m)} = \frac{\hat{\rho}_j^{(m)} \phi(\boldsymbol{x}_i; \hat{\boldsymbol{\mu}}_j^{(m)}, \hat{\boldsymbol{\Sigma}}^{(m)})}{\sum_{l=1}^{k} \hat{\rho}_l^{(m)} \phi(\boldsymbol{x}_i; \hat{\boldsymbol{\mu}}_l^{(m)}, \hat{\boldsymbol{\Sigma}}^{(m)})}. \quad [8]$$

Then we calculate the following conditional expectation:

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) &= \sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}^{(m)}} \left[\log f_{\mathrm{F}}(\boldsymbol{x}_i, \boldsymbol{\delta}_i; \boldsymbol{\Xi}, \boldsymbol{\Sigma}) \middle| \boldsymbol{x}_i \right] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{k} \frac{\hat{\rho}_j^{(m)} \phi(\boldsymbol{x}_i; \hat{\boldsymbol{\mu}}_j^{(m)}, \hat{\boldsymbol{\Sigma}}^{(m)})}{\sum_{l=1}^{k} \hat{\rho}_l^{(m)} \phi(\boldsymbol{x}_i; \hat{\boldsymbol{\mu}}_l^{(m)}, \hat{\boldsymbol{\Sigma}}^{(m)})} \log \left\{ \rho_j \phi(\boldsymbol{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}) \right\} \\ &= \sum_{j=1}^{k} \mathrm{tr}(\hat{\boldsymbol{D}}_j^{(m)}) \log \rho_j - \frac{np}{2} \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Sigma}| \\ &\quad - \frac{1}{2} \sum_{j=1}^{k} \mathrm{tr} \left\{ (\boldsymbol{X} - \mathbf{1}_n \boldsymbol{\mu}_j')' \hat{\boldsymbol{D}}_j^{(m)} (\boldsymbol{X} - \mathbf{1}_n \boldsymbol{\mu}_j') \boldsymbol{\Sigma}^{-1} \right\}, \end{aligned} \quad [9]$$

where $\boldsymbol{\theta} = (\boldsymbol{\rho}', \mathrm{vec}(\boldsymbol{\Xi})', \mathrm{vech}(\boldsymbol{\Sigma})')'$ and $\hat{\boldsymbol{D}}_j^{(m)} = \mathrm{diag}(\hat{w}_{1j}^{(m)}, ..., \hat{w}_{nj}^{(m)})$.

Step 3. (Maximization-Step; M-Step): By maximizing the conditional expectation [9], we obtain estimates of $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$ in the $(m+1)$-th repetition. Since $\boldsymbol{\rho}$ has the restriction $\rho_1 + \cdots + \rho_k = 1$, the estimates of $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$ in the $(m+1)$-th repetition are obtained by maximizing the following Lagrange function:

$$Q_\lambda(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) = Q(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) + \lambda \left(\sum_{j=1}^{k} \rho_j - 1 \right), \quad [10]$$

where λ is the Lagrange multiplier. For λ=0, we have $Q_0(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) = Q(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)})$. Let us set $\hat{\boldsymbol{w}}_j^{(m)} = (\hat{w}_{1j}^{(m)}, ..., \hat{w}_{nj}^{(m)})'$. From the first order condition for equation [10], we have

$$\begin{aligned} &\frac{\partial}{\partial \rho_j} Q_\lambda(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) = \frac{1}{\rho_j} \mathrm{tr}(\hat{\boldsymbol{D}}_j^{(m)}) + \lambda, \\ &\frac{\partial}{\partial \boldsymbol{\mu}_j} Q_\lambda(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) = \boldsymbol{\Sigma}^{-1} \left\{ \boldsymbol{X}' \hat{\boldsymbol{w}}_j^{(m)} - \mathrm{tr}(\hat{\boldsymbol{D}}_j^{(m)}) \boldsymbol{\mu}_j \right\}, \\ &\frac{\partial}{\partial \boldsymbol{\Sigma}} Q_\lambda(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) = \frac{1}{2} \left\{ \boldsymbol{\Sigma}^{-1} \sum_{j=1}^{k} (\boldsymbol{X} - \mathbf{1}_n \boldsymbol{\mu}_j')' \hat{\boldsymbol{D}}_j^{(m)} (\boldsymbol{X} - \mathbf{1}_n \boldsymbol{\mu}_j') \boldsymbol{\Sigma}^{-1} - n \boldsymbol{\Sigma}^{-1} \right\}. \end{aligned} \quad [11]$$

When $\hat{\boldsymbol{\theta}}^{(m+1)} = (\hat{\boldsymbol{\rho}}^{(m+1)'}, \mathrm{vec}(\hat{\boldsymbol{\Xi}}^{(m+1)})', \mathrm{vech}(\hat{\boldsymbol{\Sigma}}^{(m+1)})')'$ becomes an optimal solution, the following equations need to be satisfied:

$$\begin{aligned} &\frac{\partial}{\partial \rho_j} Q_\lambda(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(m+1)}} = 0, \quad \frac{\partial}{\partial \boldsymbol{\mu}_j} Q_\lambda(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(m+1)}} = \mathbf{0}_p, \\ &\frac{\partial}{\partial \boldsymbol{\Sigma}} Q_\lambda(\boldsymbol{\theta} \mid \boldsymbol{X}, \hat{\boldsymbol{\theta}}^{(m)}) \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(m+1)}} = \boldsymbol{O}_{p,p}, \end{aligned} \quad [12]$$

where $\mathbf{0}_p$ and $\boldsymbol{O}_{p,p}$ are $p\times 1$ vector and $p\times p$ matrix with all elements equal to 0, respectively. From the above equations, the Lagrange multiplier λ becomes $-n$. Therefore, estimates of $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$ in the $(m+1)$-th repetition are given by

$$\begin{aligned}\hat{\rho}_j^{(m+1)} &= \frac{1}{n}\mathrm{tr}(\hat{\boldsymbol{D}}_j^{(m)}), \quad \hat{\boldsymbol{\mu}}_j^{(m+1)} = \frac{1}{\mathrm{tr}(\hat{\boldsymbol{D}}_j^{(m)})}\boldsymbol{X}'\hat{\boldsymbol{w}}_j^{(m)}, \\ \hat{\boldsymbol{\Sigma}}^{(m+1)} &= \frac{1}{n}\sum_{j=1}^{k}\boldsymbol{X}'\left\{\hat{\boldsymbol{D}}_j^{(m)} - \frac{1}{\mathrm{tr}(\hat{\boldsymbol{D}}_j^{(m)})}\hat{\boldsymbol{w}}_j^{(m)}\hat{\boldsymbol{w}}_j^{(m)\prime}\right\}\boldsymbol{X}.\end{aligned} \quad [13]$$

Step 4. We repeat Steps 2 and 3 if $||\,\hat{\boldsymbol{\rho}}^{(m+1)} - \hat{\boldsymbol{\rho}}^{(m)}\,||\,/\,||\,\hat{\boldsymbol{\rho}}^{(m)}\,|| \geq a$. Otherwise, we regard $\hat{\boldsymbol{\theta}}^{(m+1)}$ as an optimal solution for $\boldsymbol{\theta}$ to maximize the marginal log-likelihood of $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$. Here a is any given tolerance for convergence. Expressing the optimal $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\rho}}', \mathrm{vec}(\hat{\boldsymbol{\Xi}})', \mathrm{vech}(\hat{\boldsymbol{\Sigma}})')'$, we calculate the following estimated posterior probability:

$$\hat{w}_{ij} = \frac{\hat{\rho}_j\phi(\boldsymbol{x}_i;\hat{\boldsymbol{\mu}}_j,\hat{\boldsymbol{\Sigma}})}{\sum_{l=1}^{k}\hat{\rho}_l\phi(\boldsymbol{x}_i;\hat{\boldsymbol{\mu}}_l,\hat{\boldsymbol{\Sigma}})}. \quad [14]$$

We assign the i-th individual to the population under the highest $\hat{w}_{ij}$ $(j=1,\ldots,k)$, i.e.,

$$\hat{\boldsymbol{\delta}}_i = \boldsymbol{e}_{\hat{j}_i},\ \hat{j}_i = \arg\max_{j=1,\ldots,k}\hat{w}_{ij}. \quad [15]$$

2.2. Choice of the Number of Clusters

Choice of the number of clusters plays an important role in the classification through MGMM. Choosing the best number of clusters is equivalent to choosing the best model among $M_1,\ldots,M_K$, where K is the maximum number of clusters in the analysis. In order to search for the best model, we apply such an idea that a model fitted to the data well with the small number of parameters is regarded as a "good" model among all candidates. With this idea, we search for the best model with a minimum risk defined by the predictive discrepancy between the candidate model M_k in [5] and the true model M^*, given by

$$M^*: \boldsymbol{x}_1,\ldots,\boldsymbol{x}_n \sim i.i.d.\,\varphi(\boldsymbol{x}), \quad [16]$$

where φ is unknown marginal probability density function of $\boldsymbol{x}$. The one with the smallest risk is regarded as the best model among all candidate models.

Let $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n$ be $p\times 1$ independent random vectors from $\boldsymbol{u}$ with the same distribution of $\boldsymbol{x}$, but independent of $\boldsymbol{X}$, and $\boldsymbol{U}=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n)'$. Then we define the risk based on the predictive Kullback-Leibler (KL) discrepancy (Kullback & Leibler, 1951) between M_k and M^* by

$$R_k = -2\sum_{i=1}^{n} E_{\boldsymbol{X}}^{*} E_{\boldsymbol{U}}^{*}\left[\log f(\boldsymbol{u}_i;\hat{\boldsymbol{\rho}},\hat{\boldsymbol{\Xi}},\hat{\boldsymbol{\Sigma}})\right], \qquad [17]$$

where $E_{\boldsymbol{X}}^{*}$ and $E_{\boldsymbol{U}}^{*}$ are the expectation with respect to $\boldsymbol{X}$ and $\boldsymbol{U}$, respectively, and f is the marginal density given by [4].

By obtaining an unbiased estimator of R_k, we can correctly evaluate the discrepancy between the data and the model. The simplest estimator of R_k is the sample KL discrepancy function by

$$\begin{aligned} -2\ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{X}) &= -2\sum_{i=1}^{n} \log f(\boldsymbol{x}_i;\hat{\boldsymbol{\rho}},\hat{\boldsymbol{\Xi}},\hat{\boldsymbol{\Sigma}}) \\ &= np\log 2\pi + n\log|\hat{\boldsymbol{\Sigma}}| - 2\sum_{i=1}^{n} \log\left\{\sum\nolimits_{j=1}^{k} \hat{\rho}_j \exp(-\hat{r}_{ij}^2/2)\right\}, \end{aligned} \qquad [18]$$

where $\hat{r}_{ij}^2 = (\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_j)'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_j)$. Note that the sample KL discrepancy function underestimates R_k generally, so that an information criterion by $-2\ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{X}) + \hat{B}_k$ is needed, where $\hat{B}_k$ is a consistent estimator of the bias given by

$$B_k = R_k - E_{\boldsymbol{X}}^{*}\left[-2\ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{X})\right]. \qquad [19]$$

Akaike (1973; 1974) evaluated B_k by "2 times the number of independent parameters" and proposed Akaike's information criterion (AIC) by adding the evaluated B_k to the sample KL discrepancy function, i.e.,

$$\mathrm{AIC}(k) = -2\ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{X}) + (p+1)(2k+p) - 2. \qquad [20]$$

In the above evaluation, if f is not equal to φ in [16], AIC has a constant bias. This is mainly because Akaike derived AIC only under the assumption that φ and f are equal. Takeuchi (1976) revaluated the bias correction term of AIC, $(p+1)(2k+p)-2$, under the inconsistency with φ and f, and proposed the Takeuchi's information criterion (TIC) by his revaluation.

In contrast to AIC and TIC, Stone (1974) proposed the CV criterion in the following way. Let $\hat{\boldsymbol{\rho}}_{[-i]}$, $\hat{\boldsymbol{\Xi}}_{[-i]}$ and $\hat{\boldsymbol{\Sigma}}_{[-i]}$ be the MLEs of $\boldsymbol{\rho}$, $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}$, which are evaluated from such a sample set consisting of $\boldsymbol{X}$ without its i-th row vector $\boldsymbol{x}_i$. The CV criterion is given by

$$\begin{aligned}\mathrm{CV}(k) &= -2\sum_{i=1}^{n}\log f(\boldsymbol{x}_i;\hat{\boldsymbol{\rho}}_{[-i]},\hat{\boldsymbol{\Xi}}_{[-i]},\hat{\boldsymbol{\Sigma}}_{[-i]}) \\ &= np\log 2\pi+\sum_{i=1}^{n}\log|\hat{\boldsymbol{\Sigma}}_{[-i]}|-2\sum_{i=1}^{n}\log\left\{\sum_{j=1}^{k}\hat{\rho}_{j[-i]}\exp(-\hat{r}^2_{ij[-i]}/2)\right\},\end{aligned} \quad [21]$$

where $\hat{r}^2_{ij[-i]}=(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_{j[-i]})'\hat{\boldsymbol{\Sigma}}^{-1}_{[-i]}(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}_{j[-i]})$. Stone (1977) pointed out that the CV criterion is an asymptotically unbiased estimator for the risk [17]. From the result in Stone (1977), we can see that TIC and the CV criterion are asymptotically equivalent, i.e., CV=TIC+$O_p(n^{-1})$. Therefore, the order of bias of CV is the same as that of TIC. Yanagihara (2006) however showed that the CV criterion in normal regression models has smaller bias than TIC by investigating the asymptotic expansions of biases for the risk [17]. That is, in order to obtain TIC, we must estimate higher-order cumulants, of which the ordinary estimators tend to underestimate too much even for the moderate sample size. This results in the fact that TIC tends to have a large bias. By contrast, we can obtain the CV criterion without estimating higher-order cumulants, so that CV for selecting the best model is more efficient than TIC and AIC when the sample size is not so large. Form these points, we use CV for selecting the best number of clusters in this paper. The following is to determine the best number of clusters:

The Algorithm to Determine the Number of Clusters

Step 1. We determine the maximum number of clusters K.

Step 2. We calculate $\mathrm{CV}(k)$, where k is the number of clusters.

Step 3. We search for k providing the smallest value of CV as the optimal number of clusters k_{opt}, i.e., $k_{\mathrm{opt}}=\arg\min_{k=1,\ldots,K}\mathrm{CV}(k)$.

2.3. Application of MGMM to Classification of Tree Growth Patterns

Let y_{il} be observation from the i-th individual tree at the l-th time t_{il} (i=1,…,n; l=1,…,q_i), and let the corresponding vector $\boldsymbol{y}_i=(y_{i1},\ldots,y_{iq_i})'$ be a $q_i\times 1$ vector of the repeated measurement for the i-th individual tree and $\boldsymbol{t}_i=(t_{i1},\ldots,t_{iq_i})'$ be a $q_i\times 1$ chronological vector of t_{il}. Note that q_i is the number of observation of the i-th individual tree. Given such growth data, we apply the following non-linear growth curve model to $\boldsymbol{y}_i$:

$$\boldsymbol{y}_i=\boldsymbol{\eta}(\boldsymbol{t}_i;\boldsymbol{\beta})+\boldsymbol{\varepsilon}_i,\ (i=1,\ldots,n), \quad [22]$$

where $\boldsymbol{\eta}(\boldsymbol{t}_i;\boldsymbol{\beta})$ is a $q_i \times 1$ mean vector to specify a chronological non-linear trend as a function of $\eta(t_{il};\boldsymbol{\beta})$, i.e.,

$$\boldsymbol{\eta}(\boldsymbol{t}_i,\boldsymbol{\beta}) = \begin{pmatrix} \eta(t_{i1},\boldsymbol{\beta}) \\ \vdots \\ \eta(t_{iq_i},\boldsymbol{\beta}) \end{pmatrix}, \qquad [23]$$

while $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_p)'$ is a $p \times 1$ unknown parameter vector, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1},\ldots,\varepsilon_{iq_i})'$ is a $q_i \times 1$ random vector identically and independently distributed according to the q_i-dimensional normal distribution with the mean $\mathbf{0}_{q_i}$ and the variance-covariance matrix $\boldsymbol{\Sigma}_i$.

Since $\boldsymbol{y}_i = (y_{i1},\ldots,y_{iq_i})'$ is unbalanced and high dimensional data, the generalized non-linear mixed effect model proposed by Vonesh and Carter (1992) might be preferable to $\boldsymbol{y}_i$ for the growth analysis. The non-linear mixed effect model used for the forest growth analysis can be found in Fang and Bailey (200), Hall and Bailey (2001), Garber and Maguire (2003), Leites and Robinson (2004), Yanagihara *et al.* (2004), and Yanagihara and Yoshimoto (2005b). When applying the non-linear mixed effect model to the unbalanced and high dimensional data, the model is computationally and practically difficult to extend for MGMM. To overcome difficulty, we use the estimated $\boldsymbol{\beta}$ in each tree as the response variables in MGMM, as in Yanagihara and Yoshimoto (2005a), and Yoshimoto *et al.* (2005), i.e.,

$$\boldsymbol{x}_i = \hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{i1},\ldots,\hat{\beta}_{ip})' = \arg\min_{\boldsymbol{\beta}_i} \left\{\boldsymbol{y}_i - \boldsymbol{\eta}(\boldsymbol{t}_i;\boldsymbol{\beta}_i)\right\}' \left\{\boldsymbol{y}_i - \boldsymbol{\eta}(\boldsymbol{t}_i;\boldsymbol{\beta}_i)\right\}, (i = 1,\ldots,n). \qquad [24]$$

The growth patterns can be classified only through the estimated parameters, $\hat{\boldsymbol{\beta}}_i$, which control the shape of growth curves for $\boldsymbol{y}_i$. In our analysis, we use the Richards growth function, i.e.,

$$\eta(t,\boldsymbol{\beta}) = e^{\beta_1}\left\{1 - \exp\left(-e^{\beta_2}t\right)\right\}^{\exp(\beta_3)}. \qquad [25]$$

Note that the growth curve is constrained to be sigmoid by exponential parameter transformations. The unknown parameter $(\beta_1,\beta_2,\beta_3)$ in the original function is transformed to $(e^{\beta_1},e^{\beta_2},e^{\beta_3})$ in order to extend regions of estimated coefficients from $(0,\infty)$ to $(-\infty,\infty)$.

3. Numerical Example

We used the growth data obtained from the survey at Hoshino village of Fukuoka

prefecture in Kyushu, Japan. The growth data of thirty trees were obtained. This study plot is shown in Figure 1. The total number of trees in the study plot was 136. In Figure 1-1, ● and ○ denote the sampled trees and the remaining trees, respectively. The size of each circle corresponds to the relative size of DBH at the time of survey, and the number on the right side of each circle denotes the ID number of trees. Curved lines in Figure 1-1 are topological contour lines of the area. The area has higher latitude at the lower left of the figure (near P3) and lower latitude at the upper right of the figure (near P1). A complete three-dimensional shape of the study plot obtained by a spatial smoothing is shown in Figure 1-2. In this figure, ○ denotes the location of tree and darker color means lower latitude of the area. The contour lines in Figure 1-1 are based on the three-dimensional topography of the study plot in Figure 1-2. To obtain the growth data of DBH (cm), height (m) and volume (m^3), we conducted the stem analysis (see e.g., Philip, 1994) for the sampled trees.

Please insert Figure 1 around here

Figure 2 shows real growth data of DBH, height and volume. Figure 3 shows the scatter plots of estimated coefficients of the Richards growth function. In these figures, the number also denotes the ID number of each tree. From these figures, there seems to be three clusters in DBH data, no clusters in height data, and two clusters in volume data.

Please insert Figures 2 & 3 around here

Table 1 gives values of AIC and CV in each candidate model when the maximum number of clusters K is 4. In the table, the smallest value in each information criterion is marked in bold. The table implies that based on CV and AIC, 3 and 4 were chosen as the optimal number of clusters, respectively, in all growth data. Form these results, it is more likely that AIC overestimates the number of clusters in MGMM when the sample size is not so large. Therefore, CV can be recommended for determining the number of clusters in MGMM when the sample size is not so large.

Please insert Table 1 around here

Figure 4 shows the optimal cluster partition chosen by the CV criterion. The optimal number of clusters in each growth data was 3. In the figures, ○, △ and □ denote the data belonging to the clusters 1, 2 and 3, respectively, and the curved lines are contour lines of probabilities based on the fitted marginal density function f of equation [4]. Darker color of a contour line means lower probability. Table 2 shows

estimated posterior probabilities of all growth data. The highest probability in each tree is marked in bold, and the corresponding cluster is the one that the tree belongs to. From the figures and table, we can see that the cluster partition is slightly different depending upon the growth data. Comparing the clusters for DBH with that for volume, trees numbers 52, 83, 107 and 127 did not match. As for comparison of DBH with height, trees numbers 7, 12, 15, 31, 37 and 127 did not match. If the difference in the growth pattern would be caused by the difference in tree itself, all cluster partition should have become the same (or very alike). Therefore, we could expect that trees in the study plot are not different in themselves. Since almost all the highest estimated posterior probabilities of the growth data were very close to 1, we can conclude that the resulted clusters were clearly divided. The highest posterior of only the tree of No. 83 in the volume data became slightly low. This was mainly because the data was on the middle of the centers of cluster 2 and 3. Moreover, from Figure 4, three peaks were clearly observed in the fitted density function of the DBH data. This resulted in the fact that the clusters of DBH data were clearly divided into three. Not as clearly as in the case of DBH data, three peaks in the fitted density function of the volume data were observed. However, there were only two peaks in the fitted density function of the height data although the analysis concluded three clusters. This could suggest that the distribution of $\boldsymbol{x}_i$ would not be normal.

Please insert Figure 4 and Table 2 around here

Finally, we show the cluster partition in the study plot in Figure 5. In the figure, ●, ▲ and ■ denote the sampled trees belonging to the clusters 1, 2 and 3, respectively. From the figure, trees belonging to the cluster 3 were observed around the central part of the study plot. This could imply that the difference in the growth pattern is caused by geographical and topographical factors.

Please insert Figure 5 around here

4. Conclusion and Discussion

In this paper, we utilized GMM in order to identify the growth patterns in an even-aged forest stand. Assuming the Richards growth function for the growth data, the estimated coefficients of the function were used as the response variables of MGMM. The number of clusters was chosen by minimizing the CV criterion.

Applying the proposed method to our growth data of a sugi plantation forest, we found that there were three growth patterns in the study plot.

The k-means method is well known as one of clustering methods, and also widely used in several research fields. This method has an advantage that the cluster partition is obtained more easily than the clustering method through MGMM proposed here. By using the k-means method, Yanagihara and Yoshimoto (2005a) analyzed the growth pattern of the stem volume data which was also used in this paper. The clustering partition obtained by them was almost the same as our result. However, there is a disadvantage in this method. When the number of clusters is searched by the k-means method, we need the information criterion defined under the assumption that the cluster partition is explicitly assigned. In other words, the cluster partition to be searched is not treated as random variables. As a result, the problem of choosing the number of clusters is replaced with the problem of choosing the number of groups in multivariate analysis of variance (MANOVA) models. Therefore, AIC, TIC or CV in MANOVA models is used for selecting the number of clusters when we use the k-means method for the cluster analysis. However, such an information criterion does not work well for selecting the number of clusters. One reason for this is that the bias-correcting term in the information criterion is underestimated from the actual value when the MANOVA model is applied to data partitioned by the k-means method. Yanagihara and Yoshimoto (2005a) placed a burden on the bias-correcting term by assuming heteroscedasticity in the MANOVA models. Although such a criterion worked well in their paper, there is no theoretical guarantee. To the contrary, the clustering method through MGMM does not have such a disadvantage. Moreover, the clustering method through MGMM has an advantage to obtain the estimated posterior probabilities as in Table 2 or the contour lines as in Figure 4. These probabilities were used as likeliness for classifying the individual to the corresponding cluster. Not to mention, we need to choose an appropriate method to use according to the purpose of analysis.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd. International Symposium on Information Theory* (B. N. Petrov & F. Csáki Eds.), 267–281, Akadémia Kiado, Budapest.

[2] Akaike, H. (1974). A new look at the statistical model identification. *Institute of*

Electrical and Electronics Engineers, Transactions on Automatic Control, **AC–19**, 716–723.

[3] Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood function from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series* **B**, **39**, 1–38.

[4] Everitt, B. S. & Hand, D. J. (1981). *Finite Mixture Distributions*. Chapman & Hall/CRC, London.

[5] Fang, Z. & Bailey, R. L. (2001). Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *Forest Science*, **47**, 287–300.

[6] Garber, S. M. & Maguire, D. A. (2003). Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. *Forest Ecology and Management*, **179**, 507–522.

[7] Hall, D. B. & Bailey, R. L. (2001). Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *Forest Science*, **47**, 311–321.

[8] Harville, D. A. (1997). *Matrix Algebra From A Statistician's Perspective*. Springer, New York.

[9] Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79–86.

[10] Leites, L. P. & Robinson, A. P. (2004). Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. *Forest Science*, **50**, 204–212.

[11] MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Ed. J. Neyman), **1**, 281–298, Berkeley.

[12] Philip, M. S. (1994). *Measuring Trees and Forests* (2nd. ed.). CABI Publishing, Wallingford.

[13] Richards, F. J. (1958). A flexible growth function to empirical use. *Journal of Experimental Botany*, **10**, 290–300.

[14] Stone, M. (1974). Cross-validatory choice and assessment of statistical prediction. *Journal of the Royal Statistical Society, Series* **B**, **36**, 111–147.

[15] Stone, M. (1977). An asymptotic equivalence of choice of model by

cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, Series* **B**, **39**, 44–47.

[16] Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Sciences*, **153**, 12–18 (in Japanese).

[17] Yanagihara, H. (2006). Corrected version of AIC for selecting multivariate normal linear regression models in a general nonnormal case. *Journal of Multivariate Analysis*, **97**, 1070–1089.

[18] Yanagihara, H. & Yoshimoto, A. (2005a). Clustering individual growth patterns in an even-aged forest stand. *Forest Resources Management & Mathematical Modeling* Vol. **4**, 49–70, Japan Society of Forest Planning Press, Tokyo (in Japanese).

[19] Yanagihara, H. & Yoshimoto, A. (2005b). Statistical procedure for assessing the amount of carbon sequestrated by sugi (*Cryptomeria japonica*) plantation. In *Multipurpose Inventory for the Aged Artificial Forest* (Eds. Y. Nobori, N. Takahashi & A. Yoshimoto), 125–140, Japan Society of Forest Planning Press, Utsunomiya.

[20] Yanagihara, H., Yoshimoto, A. & Nomoto, M. (2004). A generalized non-linear mixed-effects model for forest growth analysis. *Forest Resources Management & Mathematical Modeling* Vol. **3**, 14–46, Japan Society of Forest Planning Press, Tokyo (in Japanese).

[21] Yoshimoto, A., Yanagihara, H. & Ninomiya, Y. (2005). Finding factors affecting a forest stand growth through multivariate linear modeling. *Journal of Japanese Forest Society*, **87**, 504–512 (in Japanese).

[22] Vonesh, E. F. & Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, **48**, 1–17.

[23] Zhang, L., Liu, C. & Davis, C. J. (2004). A mixture model-based approach to the classification of ecological habitats using forest inventory and analysis data. *Canadian Journal of Forest Research*, **34**, 1150–1154.

Table 1. Results of information criteria for selecting the number of clusters

Data	Information Criterion	The number of clusters			
		1	2	3	4
DBH	CV	35.12	6.34	**0.78**	3.96
	AIC	33.34	1.47	-10.97	**-15.13**
Height	CV	47.21	5.31	**-55.41**	-53.74
	AIC	41.83	-1.73	-65.87	**-69.76**
Volume	CV	-60.18	-67.67	**-75.13**	-48.98
	AIC	-62.55	-81.18	-99.02	**-105.01**

The smallest value in each information criterion is marked in bold

Table 2. Estimated posterior probabilities

ID	DBH			Height			Volume		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
7	**1.00**	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	0.00
12	**1.00**	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	0.00
15	**1.00**	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	0.00
31	**1.00**	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	0.00
37	**1.00**	0.00	0.00	0.00	0.00	**1.00**	**1.00**	0.00	0.00
52	**1.00**	0.00	0.00	**1.00**	0.00	0.00	0.00	**1.00**	0.00
59	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
62	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
72	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
73	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
76	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
77	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
78	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
83	0.00	**1.00**	0.00	0.00	**1.00**	0.00	**0.61**	0.39	0.00
86	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**
87	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00
93	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
96	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
100	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
102	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
104	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
107	**1.00**	0.00	0.00	**1.00**	0.00	0.00	0.06	**0.94**	0.00
111	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
116	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
120	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
126	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00
127	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00
131	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00
133	0.00	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**1.00**	0.00
135	**1.00**	0.00	0.00	**1.00**	0.00	0.00	**0.99**	0.01	0.00

The highest probability in each tree is marked in bold.

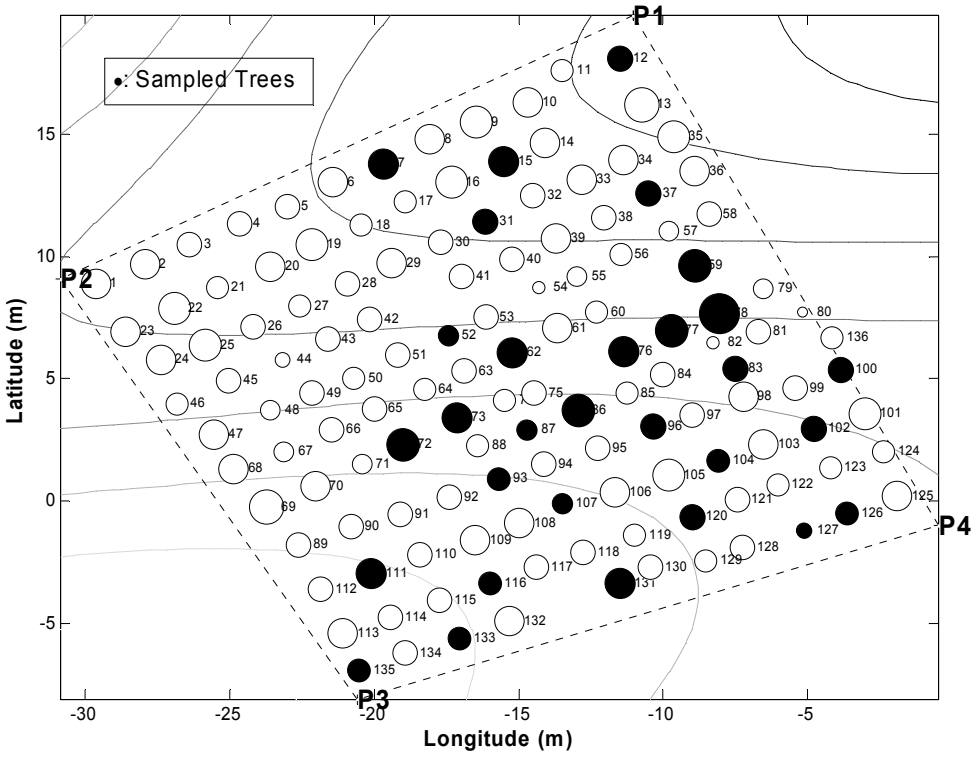

Figure 1-1. Study plot

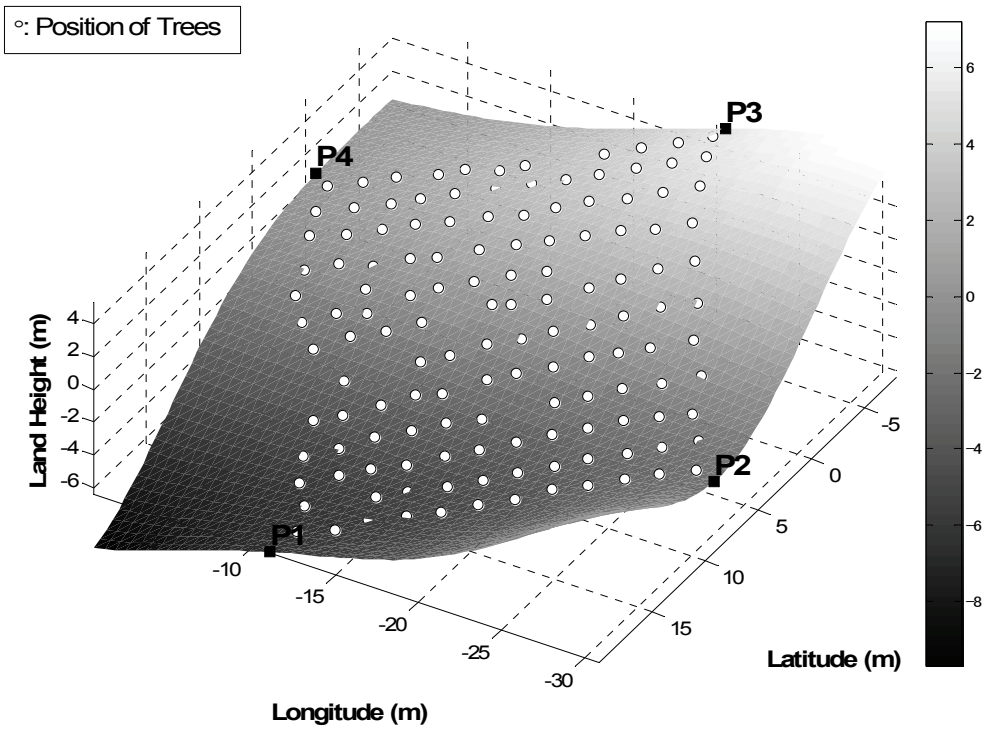

Figure 1-2. Three-dimensional sharp of study plot

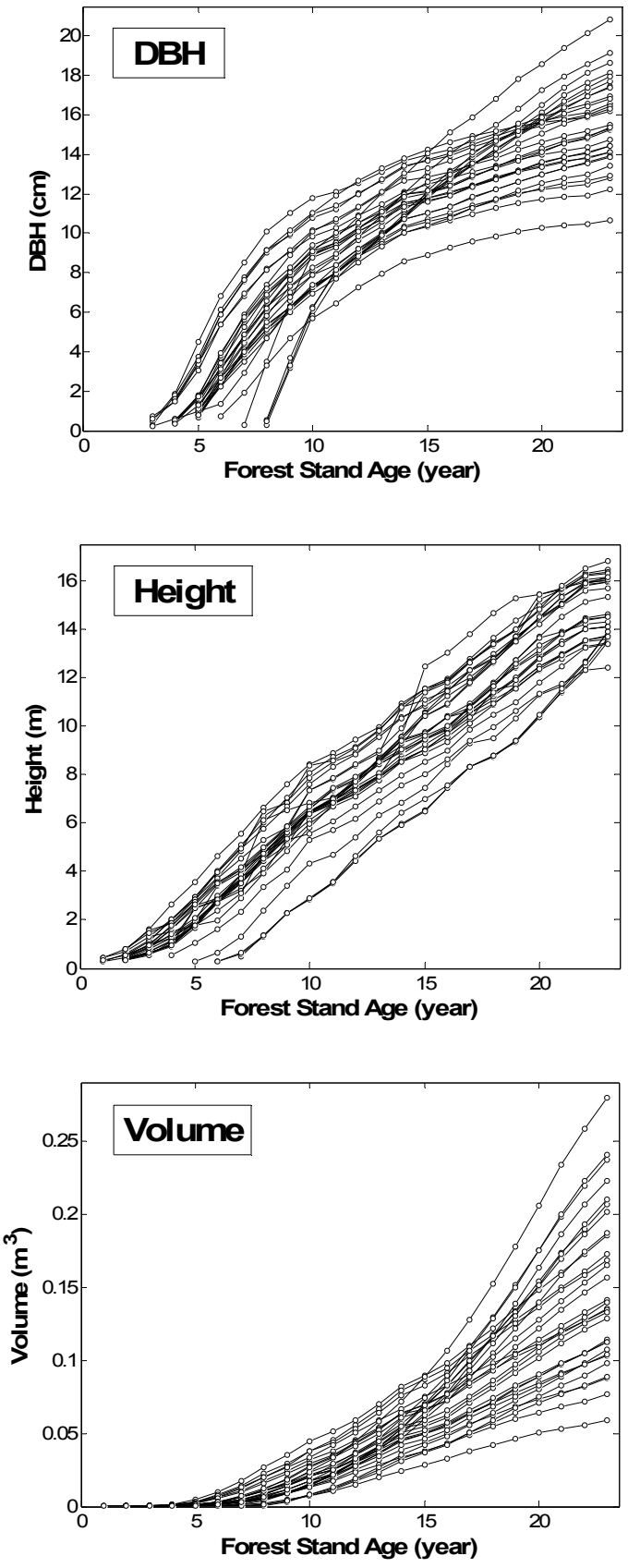

Figure 2. Real growth data

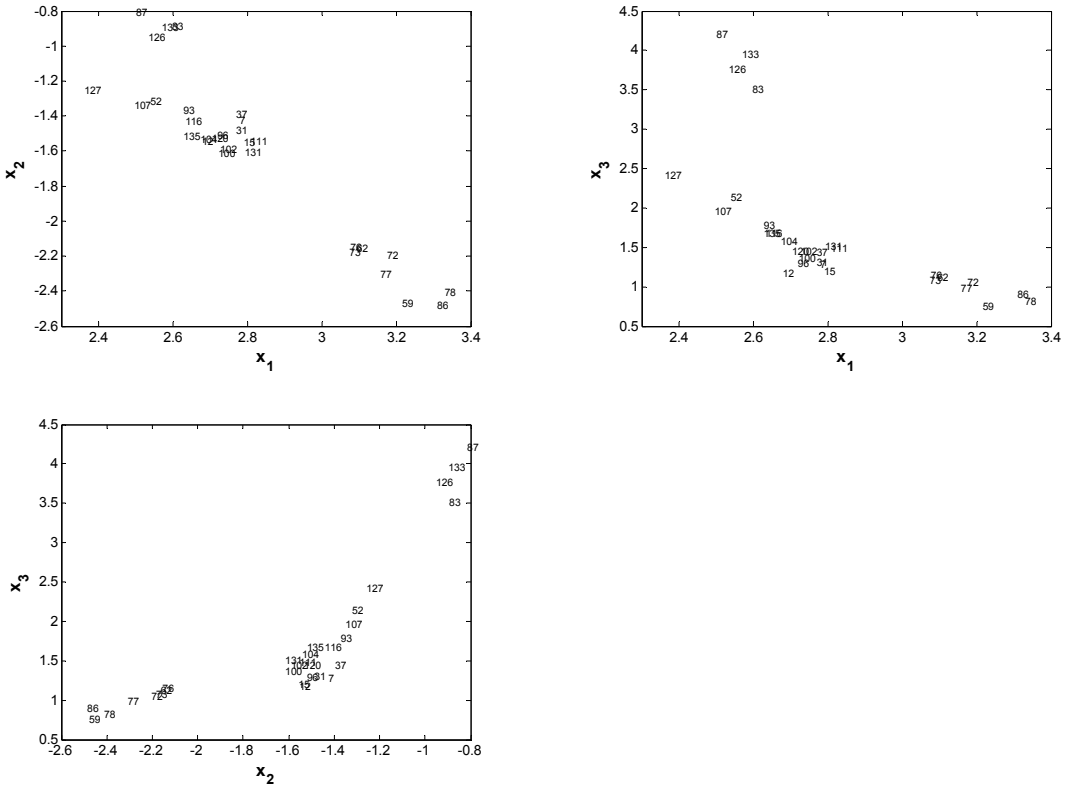

Figure 3-1. Scatter plot of response variables (DBH)

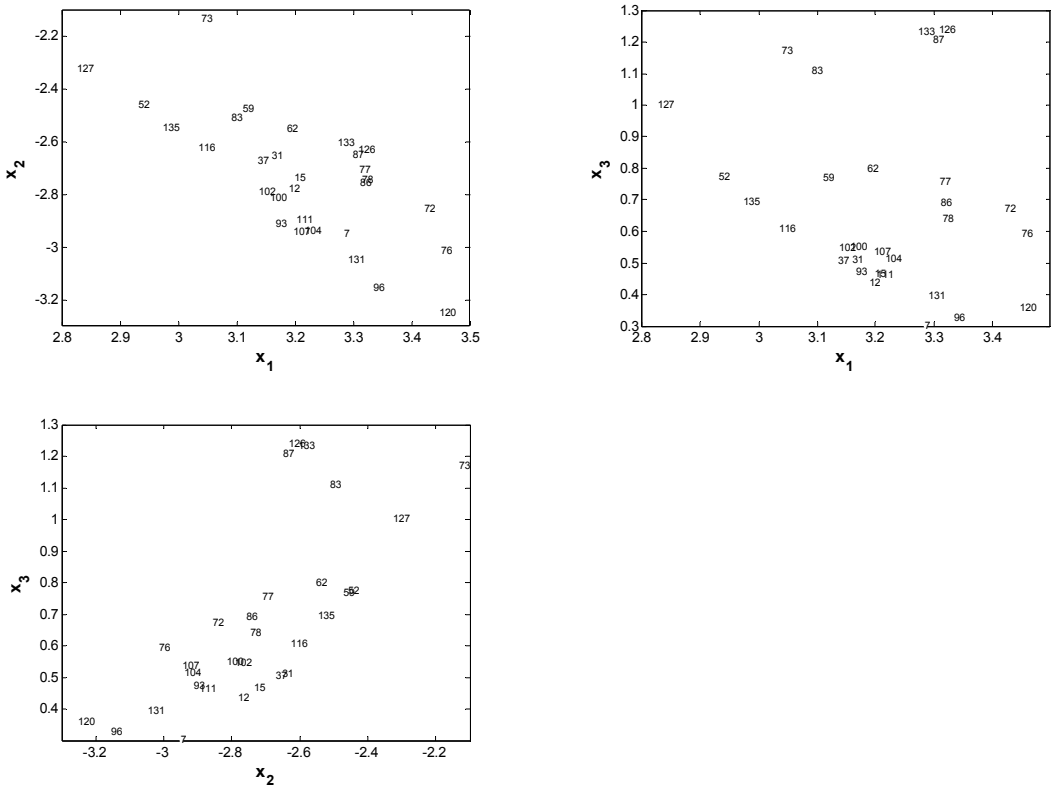

Figure 3-2. Scatter plot of response variables (Height)

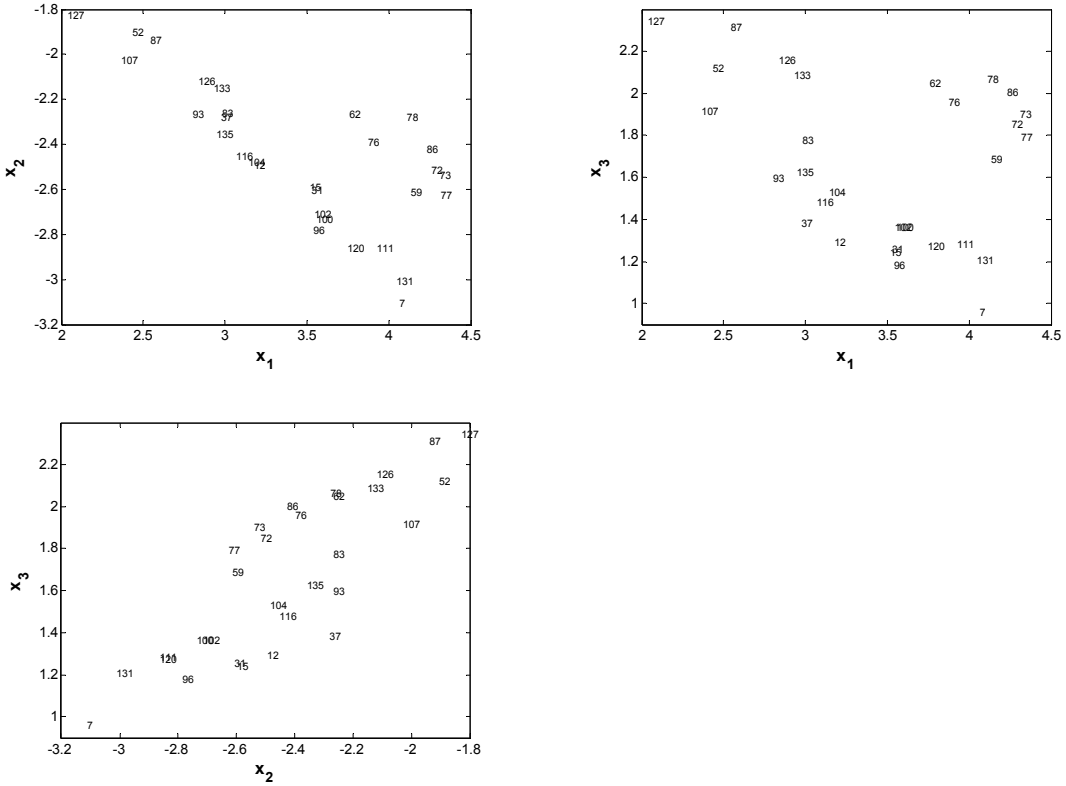

Figure 3-3. Scatter plot of response variables (Volume)

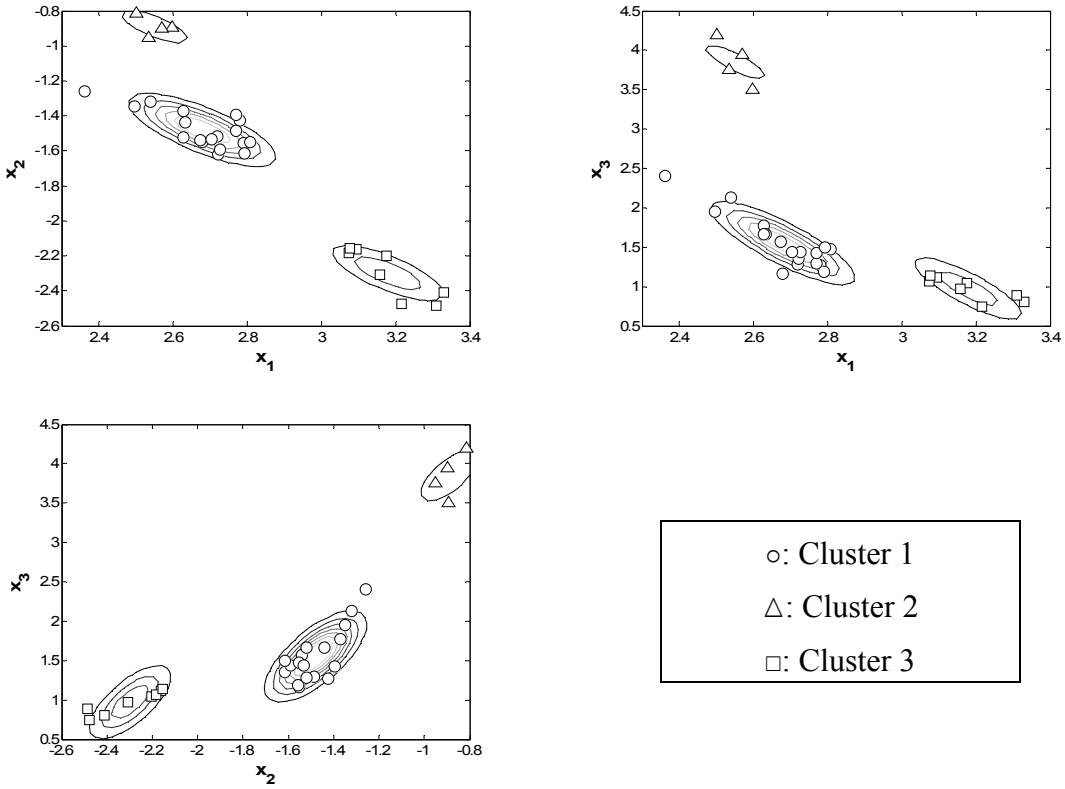

Figure 4-1. Clustering result (DBH)

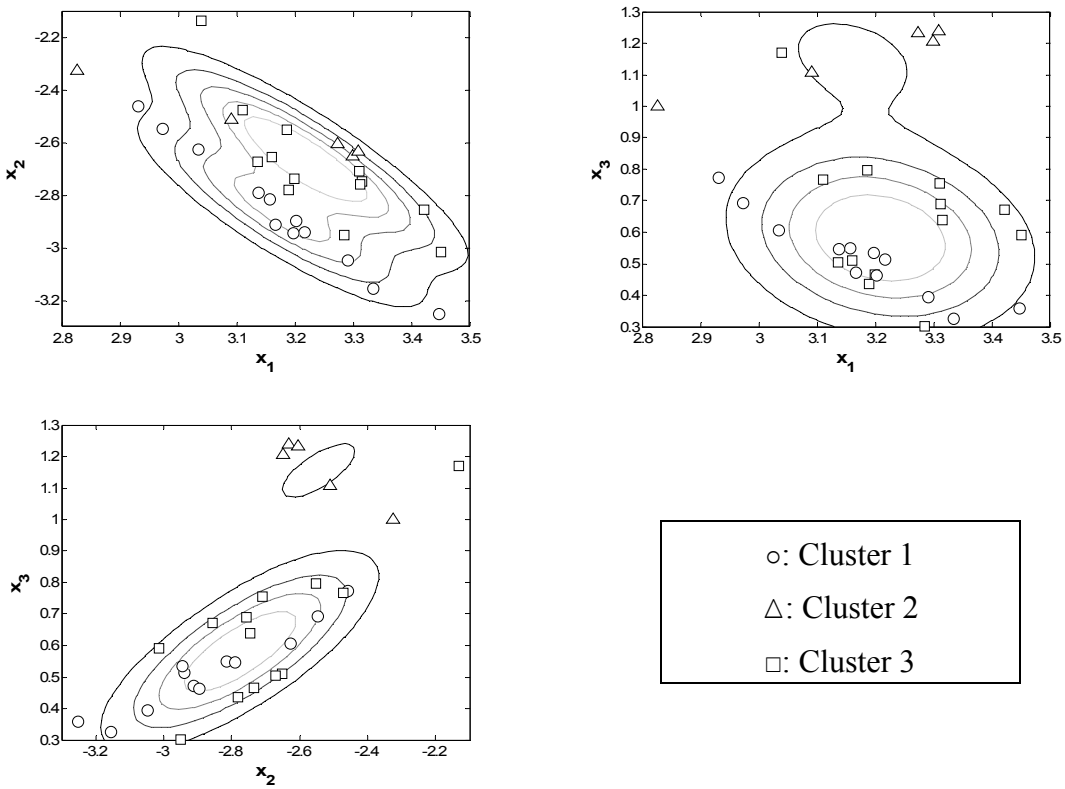

Figure 4-2. Clustering result (Height)

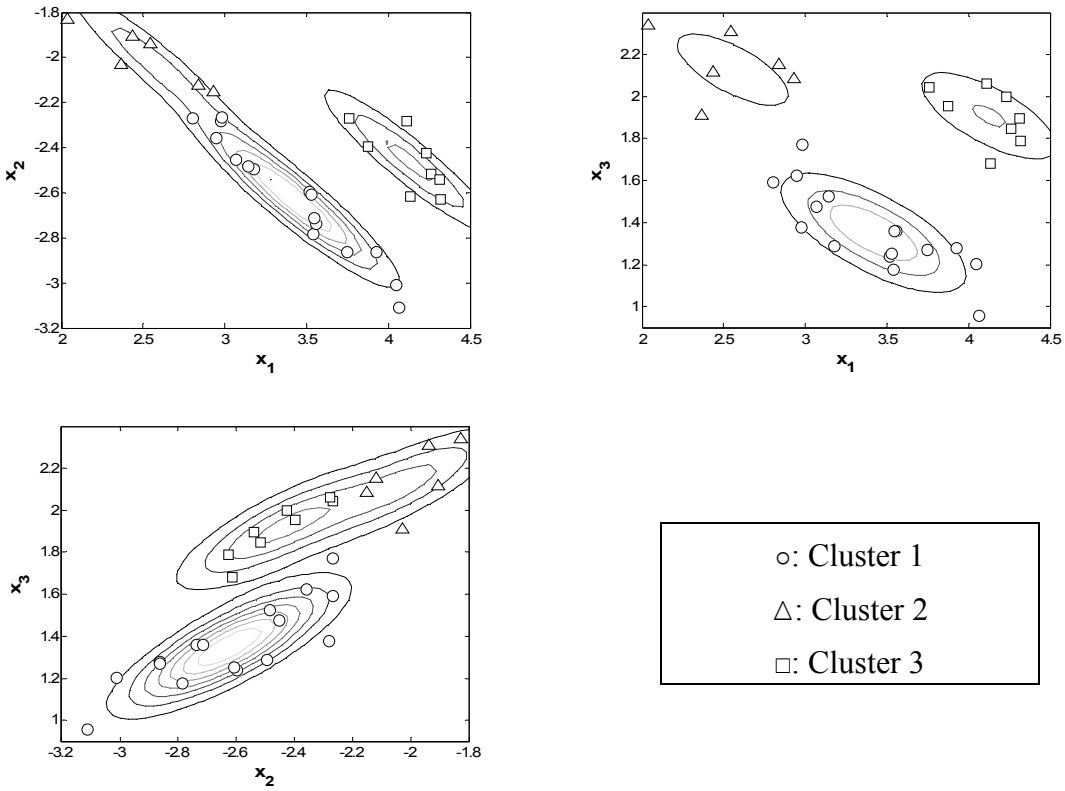

Figure 4-3. Clustering result (Volume)

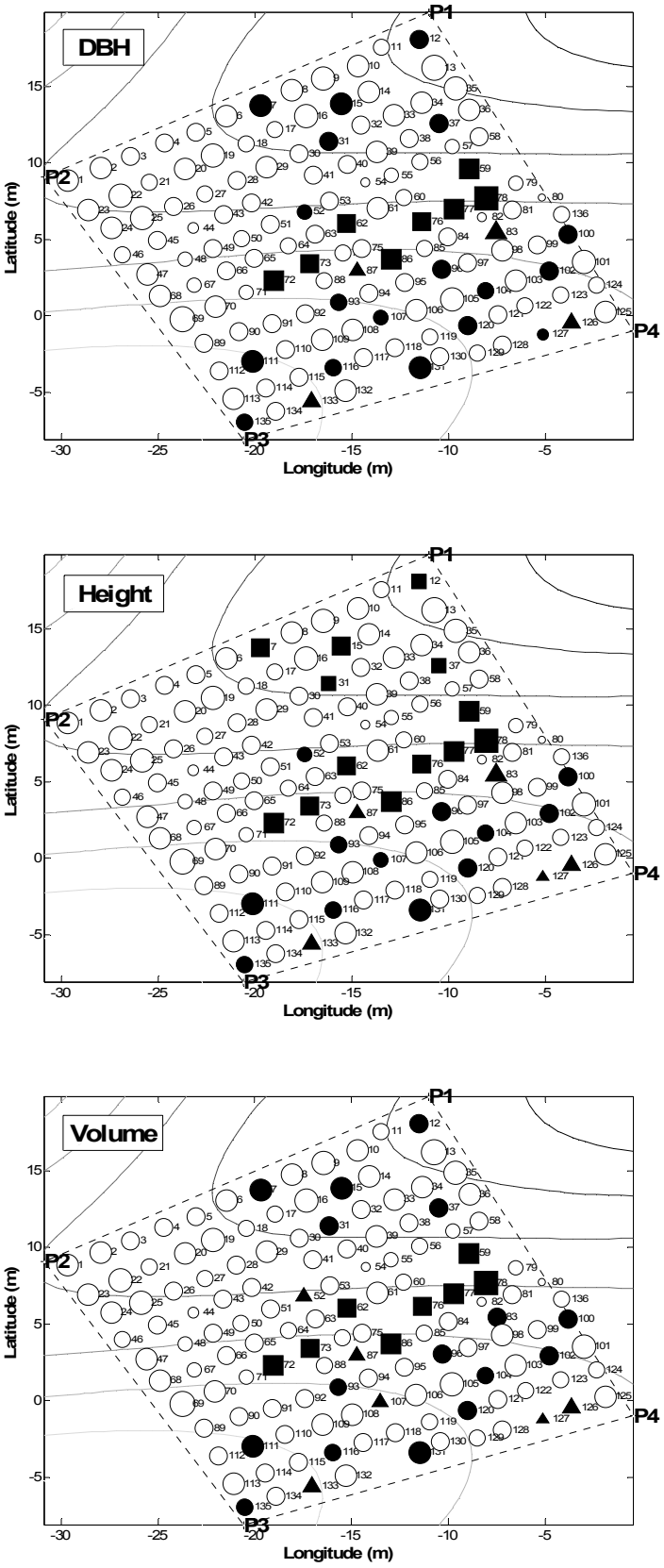

Figure 5. Clustering results via the study plot (●: Cluster 1, ▲: Cluster 2, ■: Cluster 3)