

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

Hirokazu Yanagihara, Yoshiyuki Ninomiya and Atsushi Yoshimoto

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 environmental factors. In this paper, we use a Gaussian mixture model (GMM) to investigate growth patterns in an even-aged forest to identify groups of trees with the same growth characteristics. Given the Richards growth function for the growth process, the estimated coefficients are used as new response variables in a multivariate GMM to identify growth patterns. The optimal number of grouped growth patterns is searched by minimizing the cross-validation (CV) criterion. We demonstrate our proposed method using growth data from a sugi (*Cryptomeria japonica*) sample plot in Hoshino Village, Fukuoka Prefecture, Japan. The model identified three separate growth patterns in our sample plot.

## **1. Introduction**

Even in the same even-aged forest stand, trees grow differently because of varying growth capacity, relative spatial conditions, and other environmental factors. From a management perspective, if these growth differences could be captured, it would be beneficial to consider them in growth prediction. In this paper, we use a Gaussian mixture model (GMM; e.g., Everitt and Hand, 1981) to investigate differences in the growth patterns found in an even-aged forest stand.

A clustering method through GMM has been widely used in several research fields, including forestry (e.g., Zhang *et al.*, 2004). Because tree growth data are repeated measures, the time interval for measurement can become unequal and the number of measures becomes large. Data collected over an unequal time interval is called “unbalanced data”, and data with a large number of data points is “high dimensional data”. Although a multivariate GMM (MGMM) is preferred for clustering, it is inefficient for unbalanced and high dimensional data mainly because an estimated covariance matrix in MGMM can become singular or sometimes non-obtainable for the unbalanced or high dimensional data.

To overcome the difficulties associated with unbalanced and high dimensional data, we reduce and equate the dimension of all data by assuming the Richards growth function (Richards, 1958) for the target growth data (see Yanagihara and Yoshimoto, 2005a; Yoshimoto *et al.*, 2005). In this approach, for the classification of growth patterns, the coefficients derived from the Richards growth function are regarded as new response variables in the MGMM. 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 grouped growth patterns is searched by minimizing the cross-validation (CV) criterion proposed by Stone (1974). The number of sampled trees is not so large; thus other information criteria are inappropriate.

The paper is organized as follows. In the next section, we explain our classification method through MGMM and the method used to determine the optimal number of clusters by minimizing the information criterion, followed by an explanation of MGMM applied to the classification of growth patterns. In the third section, we demonstrate our proposed method using data obtained from a forest stand owned by Hoshino Village in Fukuoka Prefecture, Kyushu, Japan.

## 2. Theoretical Background

### 2.1. Classification through MGMM

Let  $\mathbf{x} = (x_{i1}, \dots, x_{ip})'$  be a  $p \times 1$  observation vector of the  $i$ -th individual ( $i=1, \dots, n$ ), where  $\mathbf{x}_1, \dots, \mathbf{x}_n$  are mutually independent,  $n$  is the sample size, and  $'$  denotes a transposition of the matrix or vector. Suppose that each individual belongs to one of  $k$  populations (or clusters)  $\Pi_1, \dots, \Pi_k$  and the distribution of  $\Pi_j$  ( $j=1, \dots, k$ ) 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  $\mathbf{z} \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$  is given by

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

Since we do not know which population the individual belongs to in general, we treat the unknown partition of clusters as a  $k \times 1$  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  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_k)'$  and  $\rho_1 + \dots + \rho_k = 1$ . Let  $\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_n$  be  $k \times 1$  independent random vectors from  $\boldsymbol{\delta}$ , where  $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{ik})'$  is an index vector denoting the population for the  $i$ -th individual. The following conditional probability density function of  $\mathbf{x}$  given  $\boldsymbol{\delta}_i$  ( $i=1, \dots, n$ ) is derived:

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

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

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

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

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

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

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

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

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

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

However, we cannot use the full likelihood for estimating  $\boldsymbol{\rho}$ ,  $\boldsymbol{\Xi}$  and  $\boldsymbol{\Sigma}$  because  $\boldsymbol{\delta}$  is unobserved. Thus, we 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  $\boldsymbol{\delta}$  as missing values. The following is an explanation of the estimation steps of the algorithm for the ML estimates:

### 1. The EM Algorithm for Obtaining ML Estimates in MGMM

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

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

where  $\mathbf{X}$  is an  $n \times k$  matrix given by  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ ,  $\mathbf{I}_n$  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

$\hat{\boldsymbol{\theta}}^{(0)} = (\hat{\boldsymbol{\rho}}^{(0)'}, \text{vec}(\hat{\boldsymbol{\Xi}}^{(0)'})', \text{vech}(\hat{\boldsymbol{\Sigma}}^{(0)'})' )'$  , where the  $\text{vec}(\mathbf{A})$  operator transforms a matrix to a vector by stacking the first to the last column sequentially, and the  $\text{vech}(\mathbf{B})$  operator transforms the lower triangle matrix of the symmetric matrix to a vector by stacking the first to the last column (see Harville, 1997).

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

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

Then we calculate the following conditional expectation:

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

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

**Step 3.** (Maximization-Step; E-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  $\rho_1 + \dots + \rho_k = 1$ , the estimates of  $\boldsymbol{\rho}$ ,  $\boldsymbol{\Xi}$  and  $\boldsymbol{\Sigma}$  in the  $(m+1)$ -th repetition are obtained by maximizing the following Lagrange function:

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

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

condition for equation [10], we have

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

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

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

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

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

**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  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . Here  $a$  is any given tolerance for convergence. Expressing the optimal  $\boldsymbol{\theta}$  by  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\rho}}', \text{vec}(\hat{\boldsymbol{\Xi}})', \text{vech}(\hat{\boldsymbol{\Sigma}})')$ , we calculate the following estimated posterior probability:

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

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

$$[15] \quad \hat{\delta}_i = e_{j_i}, \hat{j}_i = \arg \max_{j=1, \dots, k} \hat{w}_{ij}.$$

### 2.2. Choosing the number of clusters

Choosing 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  $M_1, \dots, M_k$ , 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 in mind, we search for the best model with minimum risk defined by the predictive discrepancy between the candidate model  $M_k$  in [5] and the true model  $M^*$ , given by

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

where  $\varphi$  is the unknown marginal probability density function of  $\mathbf{x}$ . The model with the least risk is regarded as the best among all candidate models.

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

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

where  $E_{\mathbf{X}}^*$  and  $E_{\mathbf{U}}^*$  are the expectation with respect to  $\mathbf{X}$  and  $\mathbf{U}$ , 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

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

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

needed, where  $\hat{B}_k$  is a consistent estimator of the bias given by

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

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.,

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

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

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  $\mathbf{X}$  without its  $i$ -th row vector  $\mathbf{x}_i$ . The CV criterion is given by

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

where  $\hat{r}_{ij[-i]}^2 = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{j[-i]})' \hat{\boldsymbol{\Sigma}}_{[-i]}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{j[-i]})$ . Stone (1977) pointed out that the CV criterion is an asymptotically unbiased estimator for the risk [17]. From Stone’s (1977) results, we can see that TIC and the CV criterion are asymptotically equivalent, i.e.,  $\text{CV} = \text{TIC} + O(n^{-1})$ . Therefore, the order of bias of CV is equal to that of TIC. Through an investigation of the asymptotic expansions of biases for the risk [17], Yanagihara (2006) demonstrated that the CV criterion in normal regression models has smaller bias than TIC. 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 a moderate sample size. As a result, TIC tends to have a large bias. By contrast, we can obtain the CV criterion without estimating higher-order cumulants. Thus, CV is more efficient than TIC and AIC when the



sample size is not too large. Therefore, we use CV to select the best number of clusters in this paper. We used the following procedure to determine the best number of clusters:

### 2. The Algorithm to Determine the Number of Clusters

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

**Step 2.** We calculate  $CV(k)$ , 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  $k_{\text{opt}}$ , i.e.,  $k_{\text{opt}} = \arg \min_{k=1, \dots, K} CV(k)$ .

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

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

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

$$[23] \quad \boldsymbol{\eta}(\mathbf{t}_i; \boldsymbol{\beta}) = \begin{pmatrix} \eta(t_{i1}; \boldsymbol{\beta}) \\ \vdots \\ \eta(t_{iq}; \boldsymbol{\beta}) \end{pmatrix},$$

while  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  is a  $p \times 1$  unknown parameter vector,  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iq})'$  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  $\boldsymbol{\Sigma}_i$ .

Because  $\mathbf{y}_i = (y_{i1}, \dots, y_{iq})'$  is an unbalanced and high dimensional data, the generalized non-linear mixed effect model proposed by Vonesh and Carter (1992) might be preferable to  $\mathbf{y}_i$  for the growth analysis. The non-linear mixed effect

model used for the forest growth analysis can be found in Fang and Bailey (2001), 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 unbalanced and high dimensional data, the model is computationally and practically difficult to extend for MGMM. To overcome this difficulty, we use the estimated  $\beta$  in each tree as the response variables in MGMM, as in Yanagihara and Yoshimoto (2005a), and Yoshimoto *et al.* (2005), i.e.,

$$\begin{aligned}
 \mathbf{x}_i &= \hat{\beta}_i = (\hat{\beta}_{i1}, \dots, \hat{\beta}_{ip})' \\
 [24] \quad &= \arg \min_{\beta_i} \{ \mathbf{y}_i - \boldsymbol{\eta}(t_i; \beta_i) \}' \{ \mathbf{y}_i - \boldsymbol{\eta}(t_i; \beta_i) \}, \quad (i = 1, \dots, n).
 \end{aligned}$$

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

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

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 the regions of the estimated coefficients from  $[0, \infty)$  to  $(-\infty, \infty)$ .

### 3. Numerical Example

We used growth data obtained from a survey conducted at Hoshino Village of Fukuoka Prefecture in Kyushu, Japan. The data consists of growth measurements from thirty trees. The study plot is shown in Figure 1. There were 136 trees in the study plot was 136. In Figure 1-1,  $\bullet$  and  $\circ$  denote the location of sampled trees and the remaining trees, respectively. The size of each circle corresponds to the relative DBH at the time of survey, and the number on the right side of each circle denotes the ID number assigned to each tree. The curved lines in Figure 1-1 represent topographic contour lines of the area. The area has higher elevation at the lower left of the figure (near P3) and lower elevation at the upper right of the figure (near P1).

Figure 1-2 shows a complete three-dimensional shape of the study plot obtained by a spatial smoothing. In this figure,  $\circ$  denotes the location of trees and darker colors indicate lower elevation. The contour lines in Figure 1-1 are based on the three-dimensional topography of the study plot in Figure 1-2. We conducted a stem analysis (see Philip 1994) of the sampled trees to obtain the growth data of DBH ( $cm$ ), height ( $m$ ) and volume ( $m^3$ ).

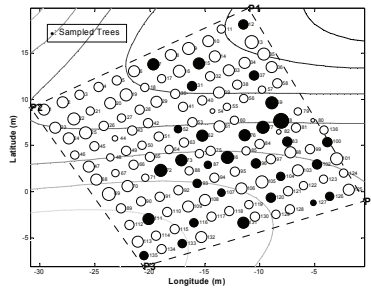


Figure 1-1. Study plot

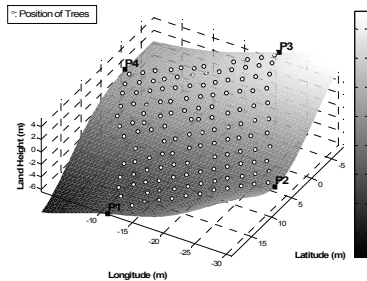


Figure 1-2. Three-dimensional sharp of study plot

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

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 are chosen as the optimal number of clusters, respectively, in all growth data. Based on 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 is recommended for determining the number of clusters in MGMM when the sample size is not so large.

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,  $\circ$ ,  $\triangle$  and  $\square$  denote the data belonging to clusters 1, 2, and 3, respectively, and the curved lines are contour lines of probabilities based on the fitted marginal density function  $f$  from equation [4]. The darker the color of a contour line, the lower the probability is.

Table 2 shows estimated posterior probabilities of all growth data. The highest probability in each tree is marked in bold, and the tree belongs to the corresponding cluster. 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 the clusters for volume, tree numbers 52, 83, 107 and 127 did not match. As for the comparison of DBH with height, tree numbers 7, 12, 15, 31, 37 and 127 did not match. If the difference in the growth pattern were caused by differences in the trees themselves, we would expect all cluster partitions to be the same (or very similar). Therefore, we can expect that trees in the study plot are not different in themselves. Because almost all the highest estimated posterior probabilities of the growth data were very close to 1, we can conclude that the resultant clusters are clearly divided. The highest posterior of tree number 83 in the volume data was only slightly lower than expected. Essentially, this was because the data was on the middle of the centers of clusters 2 and 3. Moreover, from Figure 4, three peaks were clearly observable in the fitted density function of the DBH data. As a result, the DBH data was divided into three clusters. Not so clear as the DBH data, we observed three peaks in the fitted density function of the volume data. However, there were only two peaks in the fitted density function of the height data, even though the analysis concluded three clusters, suggesting that the distribution of  $\boldsymbol{x}$  would not be normal.

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

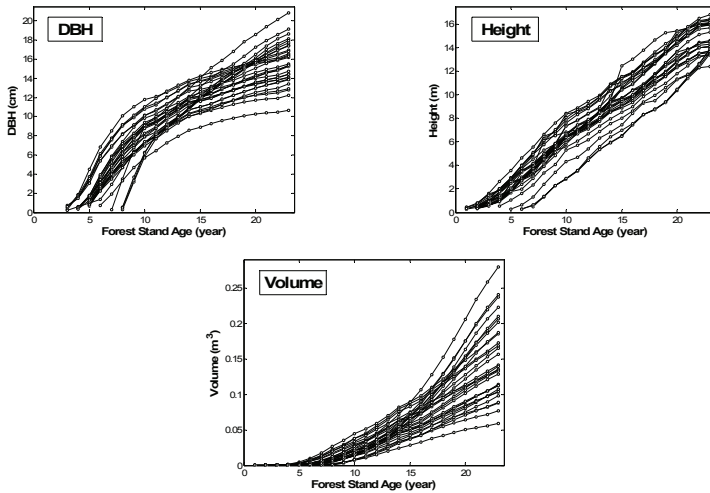


Figure 2. Real growth data

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	<b>0.78</b>	3.96
	AIC	33.34	1.47	-10.97	<b>-15.13</b>
Height	CV	47.21	5.31	<b>-55.41</b>	-53.74
	AIC	41.83	-1.73	-65.87	<b>-69.76</b>
Volume	CV	-60.18	-67.67	<b>-75.13</b>	-48.98
	AIC	-62.55	-81.18	-99.02	<b>-105.01</b>

The smallest value in each information criterion is marked in bold

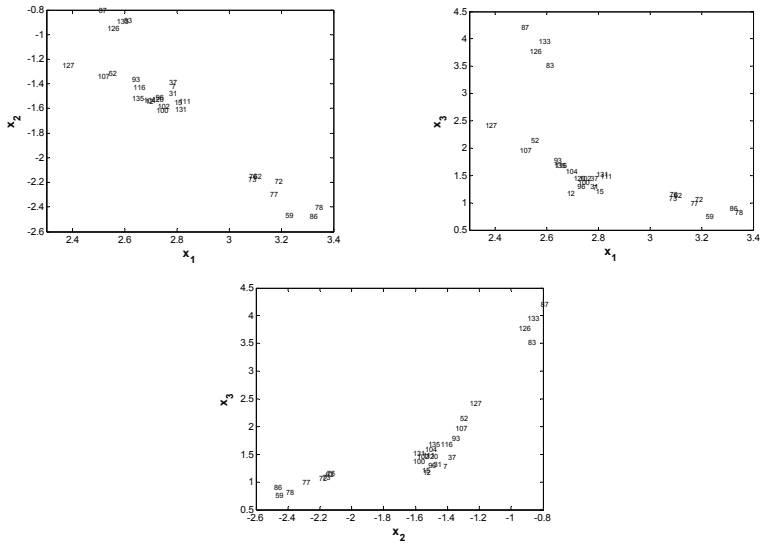


Figure 3-1. Scatter plot of response variables (DBH)

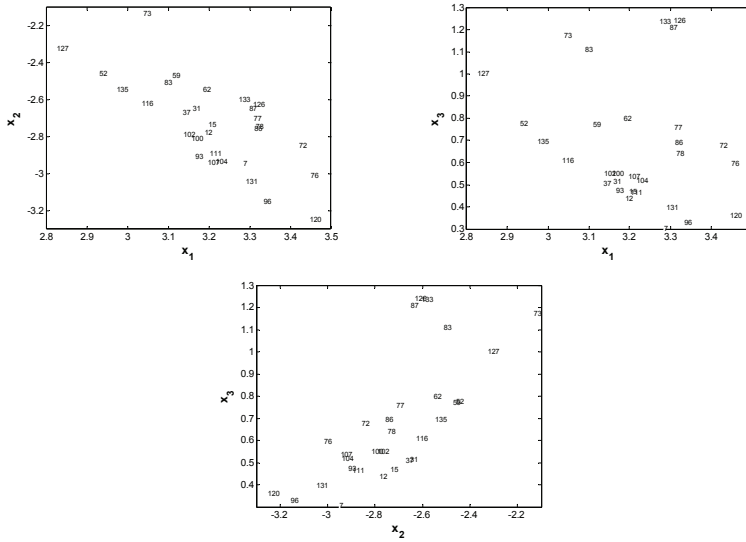


Figure 3-2. Scatter plot of response variables (Height)

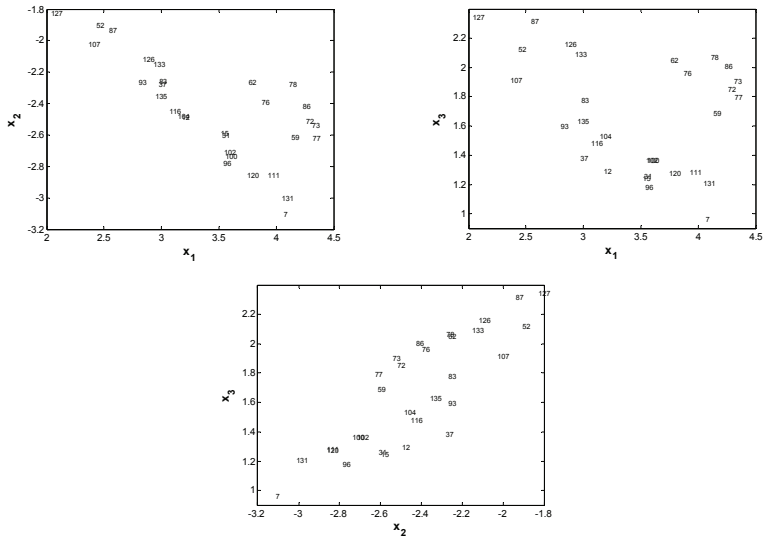


Figure 3-3. Scatter plot of response variables (Volume)

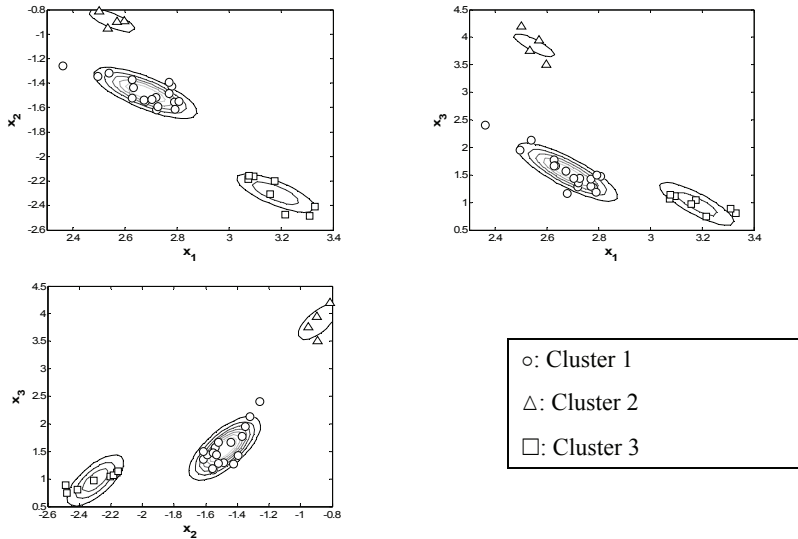


Figure 4-1. Clustering result (DBH)

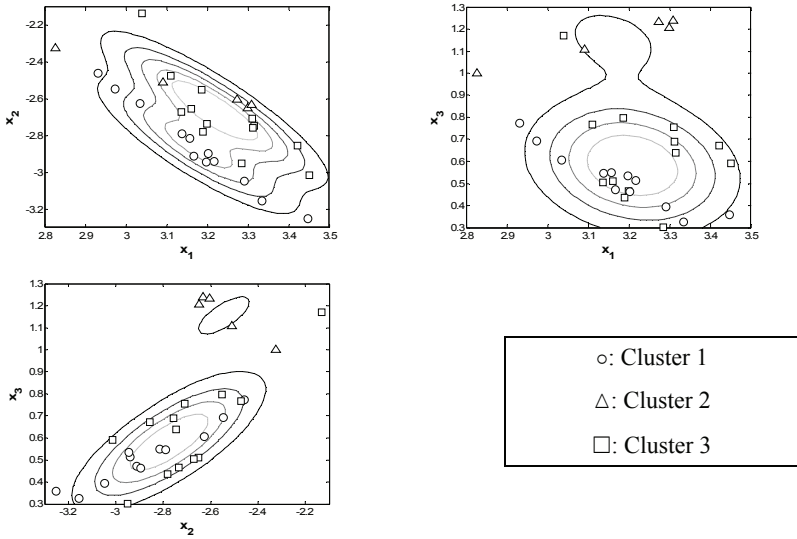


Figure 4-2. Clustering result (Height)

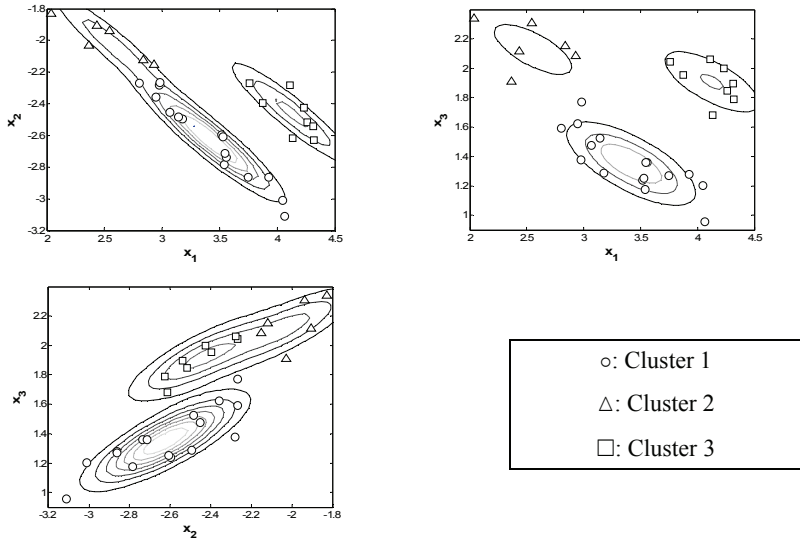


Figure 4-3. Clustering result (Volume)



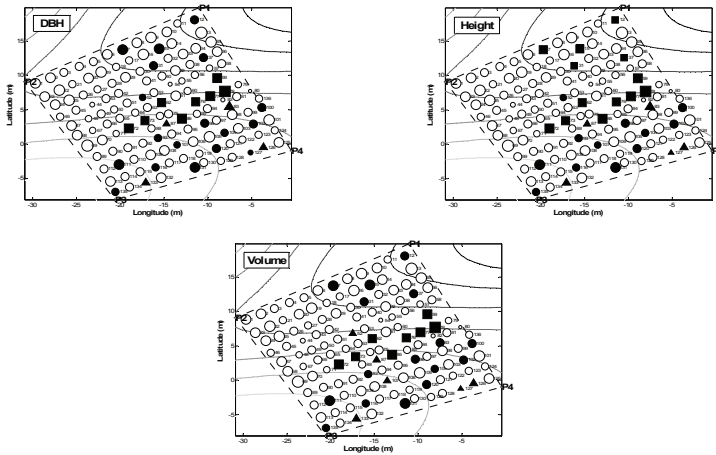


Figure 5. Clustering results via the study plot  
 (●: Cluster 1, ▲: Cluster 2, ■: Cluster 3)

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	<b>1.00</b>	0.00	0.00	0.00	0.00	<b>1.00</b>	<b>1.00</b>	0.00	0.00
12	<b>1.00</b>	0.00	0.00	0.00	0.00	<b>1.00</b>	<b>1.00</b>	0.00	0.00
15	<b>1.00</b>	0.00	0.00	0.00	0.00	<b>1.00</b>	<b>1.00</b>	0.00	0.00
31	<b>1.00</b>	0.00	0.00	0.00	0.00	<b>1.00</b>	<b>1.00</b>	0.00	0.00
37	<b>1.00</b>	0.00	0.00	0.00	0.00	<b>1.00</b>	<b>1.00</b>	0.00	0.00
52	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	0.00	<b>1.00</b>	0.00
59	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
62	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
72	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
73	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
76	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
77	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
78	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
83	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	<b>0.61</b>	0.39	0.00
86	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>
87	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00
93	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
96	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
100	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
102	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
104	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
107	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	0.06	<b>0.94</b>	0.00
111	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
116	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
120	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
126	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00
127	<b>1.00</b>	0.00	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00
131	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00
133	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00
135	<b>1.00</b>	0.00	0.00	<b>1.00</b>	0.00	0.00	<b>0.99</b>	0.01	0.00

The highest probability in each tree is marked in bold.

#### **4. Discussion and Conclusion**

In this paper, we utilized GMM to identify 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. We chose the number of clusters by minimizing the CV criterion. Applying the proposed method to growth data from a sugi plantation forest, we identified that there were three growth patterns in the study plot.

The  $k$ -means method is a well-known clustering method widely used in several research fields. This method appears to have an advantage over the clustering method through MGMM proposed here because it can more easily obtain the cluster partition. By using the  $k$ -means method, Yanagihara and Yoshimoto (2005a) analyzed the growth pattern of the same stem volume data used in this paper. The clustering partition they obtained was very similar to our results: however, there is a disadvantage to this method. When the number of clusters is searched by the  $k$ -means method, the information criterion must be defined under the assumption that the cluster partition is explicitly assigned. In other words, the cluster partition to be searched is not treated as a random variable. 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, in MANOVA models, AIC, TIC, or CV is used to select the number of clusters when the  $k$ -means method is used 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. The clustering method through MGMM does not have such a disadvantage. Moreover, the clustering method through MGMM offers the opportunity to obtain the estimated posterior probabilities as in Table 2, or the

contour lines as in Figure 4. These probabilities were used as likelihood for classifying the individual to the corresponding cluster. The appropriateness of any analysis method must be evaluated according to the purpose of the analysis.

## References

- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In 2<sup>nd</sup> International Symposium on Information Theory (Petrov, B.N. and Csáki, F. eds.), pp.267-281, Akadémia Kiado, Budapest.
- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Automat. Contr.* **AC-19**: 716-723.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood function from incomplete data via the EM algorithm, *J. Roy. Statist. Soc., Ser. B*, **39**: 1-38.
- Everitt, B.S. and Hand, D.J. (1981) Finite Mixture Distributions, Chapman & Hall/CRC, London, 123p.
- Fang, Z. and Bailey, R. L. (2001) Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments, *For. Sci.* **47**: 287-300.
- Garber, S.M. and Maguire, D.A. (2003) Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures, *For. Ecol. Manag.* **179**: 507-522.
- Hall, D.B. and Bailey, R.L. (2001) Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models, *For. Sci.* **47**: 311-321.
- Harville, D.A. (1997) Matrix Algebra From A Statistician's Perspective, Springer, New York, 630p.
- Kullback, S. and Leibler, R.A. (1951) On information and sufficiency, *Ann. Math. Statist.* **22**: 79-86.
- Leites, L.P. and Robinson, A.P. (2004) Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework, *For. Sci.* **50**: 204-212.

- MacQueen, J.B. (1967) Some methods for classification and analysis of multivariate observations. In Proc. of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (J. Neyman, ed.), 1: 281-298, Berkeley.
- Philip, M.S. (1994) Measuring Trees and Forests (2<sup>nd</sup> ed.), CABI Publishing, Wallingford, 310p.
- Richards, F.J. (1958) A flexible growth function to empirical use, ***J. Expe. Bota.*** **10**: 290-300.
- Stone, M. (1974) Cross-validatory choice and assessment of statistical prediction, ***J. Roy. Statist. Soc., Ser. B***, **36**: 111-147.
- Stone, M. (1977) An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, ***J. Roy. Statist. Soc., Ser. B***, **39**: 44-47.
- Takeuchi, K. (1976) Distribution of information statistics and criteria for adequacy of models, ***Math. Sci.*** **153**: 12-18. (in Japanese).
- Yanagihara, H. (2006) Corrected version of *AIC* for selecting multivariate normal linear regression models in a general nonnormal case, ***J. Multivariate Anal.*** **97**: 1070-1089.
- Yanagihara, H. and Yoshimoto, A. (2005a) Clustering individual growth patterns in an even-aged forest stand, ***FORMATH Vol.4***: 49-70. (in Japanese with English summary).
- Yanagihara, H. and Yoshimoto, A. (2005b) Statistical procedure for assessing the amount of carbon sequestered by sugi (*Cryptomeria japonica*) plantation, In Multipurpose Inventory for the Aged Artificial Forest (Nobori, Y., Takahashi, N. and Yoshimoto, A. eds.), pp.125-140, Japan Society of Forest Planning Press, Utsunomiya.
- Yanagihara, H., Yoshimoto, A. and Nomoto, M. (2004). A generalized non-linear mixed-effects model for forest growth analysis, ***FORMATH Vol.3***: 14-46. (in Japanese with English summary).
- Yoshimoto, A., Yanagihara, H. and Ninomiya, Y. (2005) Finding factors affecting a forest stand growth through multivariate linear modeling, ***J. Jpn. For. Res.*** **87**: 504-512. (in Japanese with English summary).

- Vonesh, E.F. and Carter, R.L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures, *Biometrics* **48**: 1-17.
- Zhang, L., Liu, C. and Davis, C.J. (2004). A mixture model-based approach to the classification of ecological habitats using forest inventory and analysis data, *Can. J. For. Res.* **34**: 1150-1154.

## 混合モデルによるスギ単純同齢林における成長パターン分類分析

柳原 宏和・二宮 嘉行・吉本 敦

**要約**：例え単純同齢林内の林木であっても，林木の成長のパターンはそれら林木が元々持つ固有の個体差，相対的な空間条件またその他の環境により異なる．本論文では，そのような単純同齢林内の林木成長パターンを正規混合モデル (Gaussian Mixture Model; GMM) を用いて分類することを試みる．成長パターンを分類するために，リチャードの成長関数を成長データにあてはめその係数を推定し，その推定された係数を新たな目的変数として多変量正規混合モデル (Multivariate GMM; MGMM) をあてはめる．最適な成長パターンのグループ数はクロスバリデーション規準 (Cross-Validation Criterion; CV criterion) の最小化により行う．この手法の適用事例に，平成15年2月に福岡県八女郡星野村の村有林 (無間伐林) で行った林分調査から得られた23年生のスギ (*Cryptomeria japonica*) 30本の胸高直径，樹高そして材積の成長データを用いた．分析の結果，すべてのデータにおいて，成長パターンは3つあることが確認できた．

**キーワード**：情報量規準, 成長分析, 多変量正規混合モデル, モデル選択