多変量線形モデルによる林分成長要因探索のための変数選択

吉本 敦^{*1}・柳原宏和²・二宮嘉行³

吉本 敦・柳原宏和・二宮嘉行：多変量線形モデルによる林分成長要因探索のための変数選択 日林誌 87：504~512, 2005 本研究では、林分の成長を左右する要因探索のための統計的手法を提示した。その方法は、まず、個々の林木の成長データにそれぞれ成長曲線をあてはめ、曲線の係数を推定する。次に、推定された成長曲線の係数値を新たな観測値とし、それらの値を被説明変数、考えられ得る要因を説明変数とした多変量線形モデルを構築する。説明変数の組み合わせを変えることにより、さらに組み合わせの数だけ多変量線形モデルを構築する。最後に、得られるモデル群に対し、交差確認法に基づく情報量規準 (CV 規準) を用いて、その規準量を最小にするモデルを最適なモデルとして決定し、最終的に成長に影響を与える要因の組み合わせを探索する。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 本の胸高直径および樹高の成長データを用いた。分析の結果、胸高直径および樹高の成長に影響を与える要因は、現時点での胸高直径の大きさと半径 5 m 基準円内における周辺林木の距離による重み付き平均胸高直径となった。

キーワード：交差確認法、情報量規準、正規多変量線形モデル、林分成長分析

Yoshimoto, A., Yanagihara, H., and Ninomiya, Y.: Finding Factors Affecting a Forest Stand Growth through Multivariate Linear Modeling. J. Jpn. For. Soc. 87: 504~512, 2005 We proposed a statistical method to search for factors affecting the target forest stand growth. The method consists of multivariate linear modeling and selection of an optimal model for such factors based on the information criterion in the following way. First, by fitting a growth curve model to the growth data for each tree, coefficients of the growth curve model are estimated. Then, by setting the estimated coefficients as new response variables, several multivariate linear models are constructed with different sets of explanatory variables possibly affecting the growth. Among a set of the derived multivariate linear models, an optimal model is sought by minimizing the cross-validation criterion for the final set of factors. We applied the proposed method to both the diameter and height growth data from a sugi (*Cryptomeria japonica*) sample plot set in one of Hoshino village forests in the Yame region, Fukuoka Prefecture. Our example showed that the current DBH (diameter at breast height) of the target tree and the distance-weighted average DBH of the surrounding trees are found to be the final explanatory variables.

Key words: cross-validation, growth analysis, information criterion, multivariate linear modeling

I. はじめに

管理を目的とした林分あるいは林木の成長を予測することは森林計画学の分野において重要な課題の一つであり、予測如何により計画の妥当性が左右される。これまで森林計画学の分野では、実務的に古くから規範的な収穫表の作成、あるいは安藤 (1966) によって提唱された林分密度管理図の作成が行われてきた。これらは表あるいは図面上で管理計画を調整するものである。それに対し、Konohira (1995) にあるように、コンピュータの普及に伴い、プログラムシステムとしてのシステム収穫表が開発されている。その例には、白石 (1986) が開発した LYCS (Local Yield Table Creation System)、田中 (1995) の“シルブの森”、稲田 (2005) の“穂の国『創造』”等があり、それぞれが特有の方法論の展開によりシステムを構成している。

成長予測のモデルとして様々なシステム収穫表が開発される一方で、システムを実務的に利用するための地域特有の係数推定が同時に行われている (Matsumoto, 1997; 嘉戸・田中, 1995)。すなわち、個々の林分の成長は地域特有の様々な要因に左右されるため、それらの特定によるシステムの調整も必要不可欠となる。さらに、同一林分内であっても、個々の林木は局所的な地理条件の違いや林木間の競争などの影響を受け成長しているため、SPS (Arney 1985), PROGNOSIS (Wykoff *et al.*, 1982), Vettenranta 1999), ORGANON (Hann, 2003), SILVA (Pretzsch *et al.*, 2002) などのように、距離従属あるいは独立の単木レベルの成長モデルを構築する際には、それらの影響を考慮できるようにモデルの変数およびパラメータの調整も必要となる。

林木の成長データは経時測定データであり、統計的には

* 連絡・別刷請求先 (Corresponding author) E-mail: yoshimoo@mail.kankyo.tohoku.ac.jp

¹ 東北大学大学院環境科学研究科 (980-8579 仙台市青葉区荒巻字青葉 6-6-20)

Graduate School of Environmental Studies, Tohoku University, 6-6-20 Aoba-Aramaki, Aoba-ku, Sendai 980-8579, Japan.

² 筑波大学大学院システム情報工学研究科 (305-8573 つくば市天王台 1-1-1)

Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8573, Japan.

³ 九州大学大学院数理学研究院 (812-8581 福岡市東区箱崎 6 丁目 10-1)

Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan.

(2005 年 6 月 2 日受付; 2005 年 11 月 11 日受理)

観測時点が一致していないアンバランスデータ、あるいは観測時点が非常に多い高次元データのカテゴリに分類されるため、成長データに対して従来の多変量分散分析モデル、一般化多変量分散分析モデル、あるいは多変量非線形回帰モデルなどをそのままあてはめて成長に影響を与える要因を探索することはモデルの構造上不可能あるいは困難である(柳原・吉本, 2003; 柳原ら, 2004)。その場合、成長曲線を決定する係数パラメータに個体変動を加味した一般化混合効果モデル(Vonesh and Carter, 1992)を用いることで、データのアンバランス性と観測次元の増大に関する問題点を回避することができる。しかしながら、一般化混合効果モデルに考えられる要因を説明変数として要因探索を行うことは、モデル推定法の煩雑性および最終的な説明変数の組み合わせを選択するための計算量に問題が出てくる。

本研究では測定データに成長曲線をあてはめ、得られる係数の推定値を新たな観測値として多変量線形モデルの被説明変数とする方法を採用し、考えられる要因を説明変数とする多変量線形モデルを構築する。次に、構築に関わる情報量規準を最小化することにより説明変数の絞り込みを行い、成長に影響を与える要因の探索を行う統計手法を提示する。最終的なモデルの選択には、交差確認法に基づく情報量規準(cross-validation criterion: CV 規準)を用いる。この手法の適用事例に、最後に福岡県八女郡星野村の村有林(無間伐林)から得られた23年生のスギ(*Cryptomeria japonica*) 30本の胸高直径および樹高の成長データを用いる。説明変数としては、対象林木の胸高直径および位置情報、周辺林木の胸高直径および本数情報を用いる。

II. 正規多変量線形モデルと最尤推定量

今 n 本の林木に対し、それぞれが相関を持つ p 個の観測値が独立に得られたものとし、以下のベクトル表示を用いる。

$$\mathbf{y}_i = (y_{i1}, \dots, y_{ip})' \quad (1)$$

y_{ij} は第 i 番目の林木の第 j 番目の観測値である。なお、表記 ' は転置を表す。観測値ベクトル $\mathbf{y}_i$ は確率変数で変動を持ち、その平均変動 $E[\mathbf{y}_i]$ は第 i 番目の林木に対し k 個の関連のある説明変数 $(x_{i1}, \dots, x_{ik})$ の変動によって記述されるものとする。すなわち、

$$E[\mathbf{y}_i] = x_{i1}\boldsymbol{\xi}_1 + \dots + x_{ik}\boldsymbol{\xi}_k = \boldsymbol{\Xi}'\mathbf{x}_i, (i=1, \dots, n) \quad (2)$$

ただし $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})'$, $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_k)'$ であり、各 $\boldsymbol{\xi}_i$ は $p \times 1$ ベクトルである。また、観測値ベクトル $\mathbf{y}_i$ は分散共分散行列 $\text{Cov}[\mathbf{y}_i] = \boldsymbol{\Sigma}$ を持つため、平均変動 $\boldsymbol{\Xi}'\mathbf{x}_i$, 分散共分散行列 $\boldsymbol{\Sigma}$ を持つ p -変量正規分布に従う正規多変量線形モデルを $\mathbf{y}_i$ に仮定する。

$$\mathbf{y}_i \sim i.d.N_p(\boldsymbol{\Xi}'\mathbf{x}_i, \boldsymbol{\Sigma}), (i=1, \dots, n). \quad (3)$$

ここで、 $\sim i.d.$ は“独立な分布に従う (independently distributed)”の意味を表す。

このとき、観測値ベクトル $\mathbf{y}_i$ の確率密度関数は以下となる(参照: 塩谷, 1990, p. 7-12)。

$$f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\Xi}, \boldsymbol{\Sigma}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \times \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\Xi}'\mathbf{x}_i)' \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\Xi}'\mathbf{x}_i)\right\} \quad (4)$$

なお、 $|\cdot|$ は行列式を表す。従って、 $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)'$ ($n \times p$ 行列), $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ ($n \times k$ 行列, $\text{rank}(\mathbf{X}) = k$ を仮定) とすると、与えられた $(\mathbf{X}, \boldsymbol{\Xi}, \boldsymbol{\Sigma})$ に対し、 $\mathbf{Y}$ の同時分布の確率密度関数 $g(\mathbf{Y} | \mathbf{X}, \boldsymbol{\Xi}, \boldsymbol{\Sigma})$ は以下となる。

$$g(\mathbf{Y} | \mathbf{X}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}) = \prod_{i=1}^n f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\Xi}, \boldsymbol{\Sigma}) \\ = (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \times \exp\left\{-\frac{1}{2} \text{tr}((\mathbf{Y} - \mathbf{X}\boldsymbol{\Xi})' \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Xi}))\right\} \quad (5)$$

なお $\text{tr}(\cdot)$ は対角成分の和を示すトレースを表す。以下、上式を候補のモデルと呼ぶことにする。候補のモデルの数は、説明変数 $\mathbf{x}_i$ の組み合わせの数により決定され、説明変数の数が K 個あった場合、それらをすべて含めないモデルを除いて候補のモデルの数は $(2^K - 1)$ となる。

候補のモデルの係数推定には最尤法を用いる。(5) 式に対する対数尤度関数は

$$l(\boldsymbol{\Xi}, \boldsymbol{\Sigma} | \mathbf{Y}, \mathbf{X}) = \sum_{i=1}^n \log f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\Xi}, \boldsymbol{\Sigma}) \\ = \log g(\mathbf{Y} | \mathbf{X}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}) \\ = -\frac{1}{2} \{ np \log 2\pi + n \log |\boldsymbol{\Sigma}| \\ + \text{tr}((\mathbf{Y} - \mathbf{X}\boldsymbol{\Xi})' \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Xi})) \} \quad (6)$$

となり、最適化に基づく第一次導関数に課される条件式(7)を解くことにより、それぞれの係数の最尤推定量 $(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}})$ は(8)式で与えられる。

$$\frac{\partial}{\partial \boldsymbol{\Xi}} l(\boldsymbol{\Xi}, \boldsymbol{\Sigma} | \mathbf{Y}, \mathbf{X}) = \mathbf{0}, \quad \frac{\partial}{\partial \boldsymbol{\Sigma}} l(\boldsymbol{\Xi}, \boldsymbol{\Sigma} | \mathbf{Y}, \mathbf{X}) = \mathbf{0} \quad (7)$$

$$\hat{\boldsymbol{\Xi}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n} \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_X) \mathbf{Y} \quad (8)$$

ただし、 $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ であり、 $\mathbf{I}_n$ は $n \times n$ の単位行列を表す。得られる対数尤度は以下のとおりとなる。

$$l(\hat{\boldsymbol{\Xi}}, \hat{\boldsymbol{\Sigma}} | \mathbf{Y}, \mathbf{X}) = -\frac{1}{2} \{ np(\log 2\pi + 1) + n \log |\hat{\boldsymbol{\Sigma}}| \} \quad (9)$$

III. モデル選択のためのリスクと情報量規準 CV 規準

本研究では以下の K-L 情報量 (Kullback and Leibler, 1951) の最小化により、候補のモデルから“最適な”モデルを選択する。今、 $\varphi(\mathbf{Y})$ を観測値ベクトル $\mathbf{y}_1, \dots, \mathbf{y}_n$ が真に従う同時分布の確率密度関数とする。このとき、真のモデルは

$$\mathbf{Y} \sim \varphi(\mathbf{Y}) \quad (10)$$

となり、モデル $g(\mathbf{Y} | \mathbf{X}, \boldsymbol{\Xi}, \boldsymbol{\Sigma})$ に対する K-L 情報量は

$$I(\varphi, g) = E^*E^*[\log \varphi(U) - \log g(U|X, \hat{\Sigma}, \hat{\Sigma})] \\ = E^*[\log \varphi(U)] - E^*E^*[l(\hat{\Sigma}, \hat{\Sigma}|U, X)] \quad (11)$$

となる。なお、 E^* と E^* は真のモデルに基づく U と Y に対する期待値を意味する。また、 U は $n \times p$ 行列 $U = (u_1, \dots, u_n)'$ であり、その行ベクトル $u_i = (u_{i1}, \dots, u_{ip})'$ はそれぞれ独立、かつ観測値ベクトル y_i と独立で同一のモデルに従う確率変数である。すべての候補のモデルにおいてK-L情報量(11)式の右辺第一項は共通であり、K-L情報量の性質として、 $I(\varphi, g) \geq 0$ であるので、K-L情報量を比較する場合、(11)式の右辺第二項目を比較すれば良い。すなわち、

$$\min_g I(\varphi, g) = \max_X E^*E^*[l(\hat{\Sigma}, \hat{\Sigma}|U, X)] \quad (12)$$

実際に比較する場合は、 -2 倍したものをモデルのリスク R として、その最小化により最適なモデルが選択される。

$$R = -2E^*E^*[l(\hat{\Sigma}, \hat{\Sigma}|U, X)] \\ = np \log 2\pi + nE^*[\log |\hat{\Sigma}|] \\ + E^*E^*[\text{tr}((U - X\hat{\Sigma})\hat{\Sigma}^{-1}(U - X\hat{\Sigma})')] \quad (13)$$

Stone (1974) は交差確認法により候補のモデルに対するリスク R を直接推定しようとする情報量規準としてCV規準を提案した。今、 $Y_{(-i)}$ と $X_{(-i)}$ を Y と X からそれぞれ第 i 番目の観測値 y_i とそれに対する説明変数 x_i を取り除くことにより得られる $(n-1) \times p$ 行列と $(n-1) \times k$ 行列とする。このとき、 $\hat{\Sigma}_{(-i)}$ と $\hat{\Sigma}_{(-i)}$ を $Y_{(-i)}$ と $X_{(-i)}$ により推定される最尤推定量、つまり

$$\hat{\Sigma}_{(-i)} = (X_{(-i)}' X_{(-i)})^{-1} X_{(-i)}' Y_{(-i)}, \\ \hat{\Sigma}_{(-i)} = \frac{1}{n-1} Y_{(-i)}' (I_{n-1} - P_{X_{(-i)}}) Y_{(-i)} \quad (14)$$

とする。これらを用いて、候補のモデルに対するCV規準は以下のように定義される。

$$CV = -2 \sum_{i=1}^n \log f(y_i | x_i, \hat{\Sigma}_{(-i)}, \hat{\Sigma}_{(-i)}) \\ = np \log 2\pi + \sum_{i=1}^n \{ \log |\hat{\Sigma}_{(-i)}| \\ + (y_i - \hat{\Sigma}_{(-i)} x_i)' \hat{\Sigma}_{(-i)}^{-1} (y_i - \hat{\Sigma}_{(-i)} x_i) \} \cdot \quad (15)$$

CV規準の推定については、それぞれ第 i 番目の観測値を除いたデータを用いて、未知パラメータの推定値を計 n 回求める必要がある。しかしながら、Fujikoshi *et al.* (2003) の正規多変量線形モデルに対する結果

$$y_i - \hat{\Sigma}_{(-i)} x_i = \frac{1}{1 - x_i' (X' X)^{-1} x_i} (y_i - \hat{\Sigma} x_i) \quad (16) \\ \hat{\Sigma}_{(-i)} = \frac{1}{n-1} \left\{ n\hat{\Sigma} - \frac{1}{1 - x_i' (X' X)^{-1} x_i} \right. \\ \left. \times (y_i - \hat{\Sigma} x_i) (y_i - \hat{\Sigma} x_i)' \right\} \quad (17)$$

を利用すれば、CV規準は下記により算出することができる。

$$CV = np \log \left(\frac{2\pi n}{n-1} \right) + n \log |\hat{\Sigma}| \\ + \sum_{i=1}^n \left\{ \log \left(1 - \frac{\hat{r}_i^2}{nh_i} \right) + \frac{(n-1)\hat{r}_i^2}{h_i(nh_i - \hat{r}_i^2)} \right\} \quad (18)$$

ただし、

$$h_i = 1 - x_i' (X' X)^{-1} x_i, \\ \hat{r}_i^2 = (y_i - \hat{\Sigma} x_i)' \hat{\Sigma}^{-1} (y_i - \hat{\Sigma} x_i) \quad (19)$$

$\hat{r}_i^2$ は推定残差のノルムの平方値である。

IV. 林木成長分析への応用

平成15年2月福岡県八女郡星野村の村有林で行った林分調査から得られたデータを用いて上記で提示した手法の適用例を示す。調査林分の立木位置図は図-1に示す。この林分は面積0.0466 ha、高度328 mの北斜面に位置する調査プロットであり、これまで間伐が施されていなかった林分である。調査林内の立木本数は136本、その内30本の間伐対象木について樹幹解析を行い、成長データを収集した。分析対象林木の平均胸高直径は15.66 cm (標準偏差2.25 cm)、平均樹高14.82 m (標準偏差1.22 m)、平均幹材積0.15 m³ (標準偏差0.055 m³)であった。なお、図-1に示されている点○の大きさは観測時点におけるそれぞれの林木の胸高直径の相対的な大きさに対応している。ここで、黒く塗りつぶした点は解析に使用したサンプル木(30本)、白い点はそれ以外を表している。また図中の番号は個々の林木に付けたID番号である。30本のサンプル木から得られた胸高直径と樹高の成長データを図-2に示す。これらの成長データに成長曲線をあてはめ、推定される係数の値を新たな観測値とし分析を進める。

ここでは、成長曲線にリチャーズの成長関数(Richards, 1958)を仮定し、柳原ら(2004)の推定方法(修正スパイダルアルゴリズム)によりそれぞれの林木に対する係数を同時に推定した。このアルゴリズムはLine Search法(Bazaraa and Shetty, 1979)に類似したもので、関数値をそのまま最適化に用いるため、ニュートン法などにみられる微分計算を用いず、探索点を動かしていくことにより最適化を行うものである。なお、リチャーズの成長関数をS字形(シグモイド型)に制約するため、 z_j を各林木の j 時点の林木の成長量とし、下記のとおり係数に指数変換を行った。

$$z_j = e^{\theta_1} \{1 - \exp(-e^{\theta_2} t_j)\}^{\exp(\theta_3)} \quad (20)$$

表-1に推定された係数を示す。新たな観測値となる第 i 番目の林木に対する係数の各変量 θ_1 , θ_2 , θ_3 は、それぞれ成長の飽和密度、内的自然増加率、成長曲線の形状を決定する数値である。ここで得られた係数の推定量を新たな観測値 y (被説明変数)とし、成長に影響を与えるだろう要因を説明変数とした正規多変量線形モデルをあてはめる。ここで、第 i 番目の林木に対し、推定された係数をそれぞれ $\hat{\theta}_{i1}$, $\hat{\theta}_{i2}$, $\hat{\theta}_{i3}$ とすると、下記を得る。

$$y_i = (y_{i1}, y_{i2}, y_{i3})' = (\hat{\theta}_{i1}, \hat{\theta}_{i2}, \hat{\theta}_{i3})' \quad (21)$$

図-1. 分析対象林分における立木位置図

考慮した説明変数は12個で、対象林木が持つ直接的要因とその周辺林木から推定される要因から選んだ。まず、同一林分内においても微少な位置の変化（空間分布の違い）により成長が左右されるかどうかを調べるため、対象林木固有の空間構造の要素として位置情報（ Ltd (m)：対象林木の緯度、 Lgd (m)：対象林木の経度、 Ald (m)：対象林木の高度）を用いた。仮に、林木の成長が空間分布の違いにより異なれば、Pukkala and Miina (1998) にもあるように距離従属単木モデルの最適化において、間伐対象となる林木を位置により特定することができる。次に、成長に対する自身の大きさの影響をみるため、対象林木の観測時点における胸高直径（ DBH_0^l (cm)）を用いた。距離依存の林木成長の競合に対する分析では、例えば Bella (1971), Daniels (1976), Biging and Dobbertin (1992), Pretzsch (1997), Vettenranta (1999), Canham *et al.* (2004) などで見られるように、周辺林木の大きさや距離が取り上げられている。ここでは、対象林木に一番近い林木の胸高直径（ DBH_0^l (cm)）および距離（ Dst_0^l (m)），そして、以下で算出される周辺林木の競合を考慮した要因を用いた。すなわち、対象林木を中心とした半径 r の基準円内部での立木密度（ Dns_r (本/㎡)），基準円面積に対する周辺林木の胸高断面積合計の占有率（ Shr_r (%)），さらに周辺林木の対象林木からの距離および胸高直径の大きさを考慮した距離による重み付き平均胸高直径（ $wDBH_0^l$ (cm)）を要因として用いた。それぞれの算出式は下記のとおりである。

$$Dns_r = \frac{\text{半径 } r \text{ の基準円内の立木本数}}{\pi r^2} \quad (\text{本/㎡}) \quad (22)$$

$$Shr_r = \frac{\text{半径 } r \text{ の基準円内の周辺林木の胸高断面積合計}}{\pi r^2} \times 100 (\%) \quad (23)$$

$$wDBH_0^l = \sum_{i=1}^n w_0^l \times DBH_i^0 \quad (24)$$

ただし、

$$w_0^l = \frac{\tau(d_0^l, r)}{\sum_{j=1}^n \tau(d_0^j, r)},$$

$$\tau(d_0^l, r) = \begin{cases} 1 + \cos(d_0^l \pi / r) & (0 < d_0^l \leq r), \\ 0 & (d_0^l = 0, d_0^l > r), \end{cases} \quad (25)$$

ここで、 d_0^l は対象林木から第 l 番目の林木までの距離、 DBH_i^0 は第 i 番目の林木の現時点の胸高直径、 n は基準円内の他の林木の総数である。上記の重み (25) 式は、基準円内の他の林木の対象林木への距離による重みを表すもので、対象林木に近ければ大きく、遠ければ小さくなるように設定した。なお、 Dns_r 、 Shr_r を算出する際、試験林の境界線と基準円が交差する場合、試験林と基準円が交わる部分の面積で対応した。また、基準円半径 r については、プロット内において最小林木間距離が1m以上であったため、2mから5mまで基準円半径を1m刻みで設定し、それぞれの値を計測した結果、変数値の分散が最も大きくなった半径2mと最も小さくなった5mを、変数の多重共線性を加味し基準円半径 r の影響の違いを顕著にするために採用した（ $r=2, 5$ ）。

図-2. 計測成長データ

a) 胸高直径成長, b) 樹高成長。

すべての説明変数の数は $K=12$ となるので、候補のモデルの総数は $2^{12}-1=4095$ となる。それらすべてのモデルにおいて CV 規準を計算し、最終的に選択された結果を表-2に示す。表中の GCD は一般化決定係数 (Generalized Coefficient of Determination, 柳井, 1975) と呼ばれる指数で、0 から 1 までの値をとり、観測値 Y と説明変数 X の関連の強さを表すものである。ここでは、比較のため情報量規準 AIC (Akaike, 1973), TIC (竹内, 1976) も算出した。これらの情報量規準は、前述したリスク R を標本対数尤度 $-2l(\hat{\Sigma}, \hat{\Sigma}|Y, X)$ で推定したときのバイアスを補正したもので、 AIC は候補のモデルと真のモデル $\varphi(Y)$ が一致するという条件、 TIC は一致していないという条件でそれぞれバイアス補正項を評価したものである。

$$\begin{aligned} AIC &= -2l(\hat{\Sigma}, \hat{\Sigma}|Y, X) + \hat{B}_{AIC}, \\ TIC &= -2l(\hat{\Sigma}, \hat{\Sigma}|Y, X) + \hat{B}_{TIC} \end{aligned} \quad (26)$$

AIC については、 y_i が独立に同一の多変量正規分布に従えば、 $n \rightarrow \infty$ に対し $E^*[AIC] \rightarrow R$ となり AIC のリスク R に対するバイアスはなくなるが、そうでない場合は、 $E^*[AIC] \rightarrow R - \kappa_4^{(1)}$ となり、真のモデルでの多変量尖度

表-1. 胸高直径と樹高の成長曲線の推定係数値

ID	胸高直径			樹 高		
	y_1	y_2	y_3	y_1	y_2	y_3
7	2.7801	-1.4247	1.2678	3.2854	-2.9499	0.3011
12	2.6796	-1.5522	1.1612	3.1895	-2.7803	0.4345
15	2.7897	-1.5559	1.1849	3.1998	-2.7355	0.4642
31	2.7708	-1.4857	1.2969	3.1607	-2.6519	0.5093
37	2.7712	-1.3935	1.4219	3.1363	-2.6721	0.5033
52	2.5407	-1.3201	2.1245	2.9315	-2.4583	0.7723
59	3.2168	-2.4764	0.7470	3.1110	-2.4728	0.7666
62	3.0934	-2.1634	1.1181	3.1857	-2.5511	0.7972
72	3.1761	-2.2022	1.0455	3.4225	-2.8552	0.6720
73	3.0733	-2.1838	1.0655	3.0400	-2.1325	1.1711
76	3.0763	-2.1556	1.1339	3.4507	-3.0135	0.5919
77	3.1586	-2.3086	0.9747	3.3109	-2.7084	0.7546
78	3.3300	-2.4112	0.8044	3.3159	-2.7452	0.6395
83	2.5966	-0.8948	3.5076	3.0914	-2.5102	1.1067
86	3.3112	-2.4842	0.8874	3.3127	-2.7575	0.6889
87	2.5014	-0.8125	4.1993	3.2991	-2.6498	1.2054
93	2.6282	-1.3714	1.7735	3.1664	-2.9122	0.4724
96	2.7197	-1.5198	1.2782	3.3352	-3.1539	0.3246
100	2.7218	-1.6170	1.3509	3.1574	-2.8152	0.5482
102	2.7261	-1.5914	1.4315	3.1385	-2.7911	0.5470
104	2.6727	-1.5411	1.5671	3.2162	-2.9404	0.5125
107	2.4960	-1.3475	1.9456	3.1980	-2.9438	0.5355
111	2.8088	-1.5512	1.4755	3.2026	-2.8953	0.4618
116	2.6339	-1.4392	1.6610	3.0341	-2.6258	0.6061
120	2.7035	-1.5317	1.4420	3.4477	-3.2514	0.3586
126	2.5344	-0.9529	3.7589	3.3088	-2.6318	1.2396
127	2.3623	-1.2583	2.4046	2.8270	-2.3240	1.0010
131	2.7935	-1.6146	1.4961	3.2906	-3.0460	0.3955
133	2.5695	-0.8968	3.9415	3.2733	-2.6039	1.2321
135	2.6278	-1.5216	1.6591	2.9727	-2.5466	0.6912

表-2. 各情報量規準により選ばれた最適モデルと要因

情報量規準	GCD	選ばれた説明変数 (要因)
胸高直径		
CV	0.6333	$Const.$ $DBH_0^{\frac{1}{2}}$ $wDBH_0^{\frac{1}{2}}$
AIC^*	0.6648	$Const.$ Ltd $DBH_0^{\frac{1}{2}}$ $wDBH_0^{\frac{1}{2}}$
TIC^{**}	0.7118	$Const.$ Ltd $DBH_0^{\frac{1}{2}}$ Dns_5 Shr_5 $wDBH_0^{\frac{1}{2}}$ $wDBH_5^{\frac{1}{2}}$
樹 高		
CV	0.6157	$Const.$ $DBH_0^{\frac{1}{2}}$ $wDBH_0^{\frac{1}{2}}$
AIC	0.7288	$Const.$ Lgd Ald $DBH_0^{\frac{1}{2}}$ Dns_5 Shr_5 $wDBH_0^{\frac{1}{2}}$
TIC	0.7288	$Const.$ Lgd Ald $DBH_0^{\frac{1}{2}}$ Dns_5 Shr_5 $wDBH_0^{\frac{1}{2}}$

$Const.$ は定数項。* $AIC = np(\log 2\pi + 1) + n \log |\hat{\Sigma}| + 2\{pk + p(p+1)/2\}$ 。

** TIC は正規多変量線形モデルに対するもの (Fujikoshi et al., 2005)。

$$TIC = AIC + 2 \sum_{i=1}^n x_i'(X'X)^{-1} x_i'(\hat{r}_i^2 - p) + \hat{\kappa}_4^{(1)}$$

$$\hat{r}_i^2 \text{ は (19) 式, } \hat{\kappa}_4^{(1)} = \frac{1}{n} \sum_{i=1}^n (\hat{r}_i^2)^2 - p(p+2)$$

$\kappa_4^{(1)}$ 分のバイアスが生じる。多変量尖度 $\kappa_4^{(1)} (\geq -2p)$ は分布の裾の重さを表す指数であり、 y_i の非正規性が強くなれば大きくなり、その結果、 $\hat{B}_{AIC}$ は小さく見積られる傾向がある。それに対し、 TIC は真のモデルの分布には依存せず、 y_i が同一な分布に従えば、 $n \rightarrow \infty$ に対し、 CV 規準同様 R に対するバイアスがなくなる。ただし、理論上は CV 規準と TIC は漸近同値であるが、 CV 規準は TIC と異なり尖度の推定量 $\hat{\kappa}_4^{(1)}$ (表-2 の注参照) を必要としないため、リスク R への収束は TIC に較べ速くなり、その結果、 CV 規準のバイアスは TIC のそれよりも小さくなる傾向がある。実際には $\hat{\kappa}_4^{(1)}$ は過小推定となるので (Yanagihara,

図-3. 観測値のそれぞれの変量に対する QQ-plot

2005), 結果的に $\hat{B}_{TIC}$ は小さく見積もられる傾向がある。また AIC と TIC は、候補のモデルのパラメータが多くなれば $\hat{B}_{AIC}$ と $\hat{B}_{TIC}$ が小さく見積もられ、リスク R に対するバイアスが大きくなる傾向がある (Fujikoshi *et al.*, 2005)。

上記を鑑み表-2 をみると、 AIC と TIC は CV 規準と比較し、説明変数の数の多いモデルを最適なモデルとして選択していることがわかる。これは観測値の非正規性と候補のモデルのパラメータ数の多さから、 AIC と TIC の $\hat{B}_{AIC}$ 、 $\hat{B}_{TIC}$ が小さく見積もられていることが原因だと考えられる。図-3 にそれぞれの変量の QQ-Plot (Quantile-Quantile Plot) を示す。QQ-Plot 図中の線上にプロット点が集中していれば正規分布に近いことを示す。図-3 からわかるように、胸高直径に対してはすべての変量で、また樹高については変量 y_1 と y_3 で非正規性が強いことがわかる。

最後に、 CV 規準により選択された最適なモデルの推定結果を表-3 に示す。表-3 の相関係数 $\hat{\rho}_{ij}$ からわかるように、成長の飽和密度 (y_1) と内的自然増加率 (y_2)、内的自然増加率 (y_2) と成長曲線の形状 (y_3) に比較的強い相関があることがわかる。説明変数については、 $DBH\%$ と $wDBH\%$ が選択された。これは自身の大きさに加え、Bella (1971) や Daniels (1976) が指摘しているように、より広い範囲での競合要素が成長に影響を与えているものといえる。次に、 $DBH\%$ の係数から、胸高直径と樹高ともに、飽和密度 (y_1) が現時点の胸高直径の大きさから強い影響を受けており、その値が大きいほど、最終的に大きく育つ傾向にある

表-3. 最適モデルでの要因とその推定値

		胸高直径 ($GCD^*=0.6333$)		樹高 ($GCD=0.5968$)	
		相関係数: $\hat{\rho}_{12}^*=-0.8711$		相関係数: $\hat{\rho}_{12}=-0.7586$	
		$\hat{\rho}_{13}=-0.4135$		$\hat{\rho}_{13}=-0.1094$	
		$\hat{\rho}_{23}=0.7661$		$\hat{\rho}_{23}=0.6945$	
説明変数	変量	推定値	P-値***	推定値	P-値
Const.	y_1	1.6958	(0.0000)	2.3814	(0.0001)
	y_2	-0.0936	(0.9442)	-0.7762	(0.4776)
	y_3	11.5737	(0.0005)	4.7088	(0.0000)
$DBH\%$	y_1	0.1037	(0.0000)	0.0335	(0.0010)
	y_2	-0.1598	(0.0000)	-0.0226	(0.2248)
	y_3	-0.2795	(0.0000)	-0.0433	(0.0216)
$wDBH\%$	y_1	-0.0369	(0.1052)	0.0170	(0.6123)
	y_2	0.0673	(0.3670)	-0.0992	(0.1043)
	y_3	-0.3306	(0.0731)	-0.2070	(0.0008)

* $GCD = \text{tr}(\mathbf{P}_X \mathbf{P}_Y) / \text{rank}(\mathbf{X}' \mathbf{Y})$ 。

** $\hat{\rho}_{ij} = (\hat{\Sigma})_{ij} / \sqrt{(\hat{\Sigma})_{ii}(\hat{\Sigma})_{jj}}$, ただし、 $(\mathbf{A})_{ij}$ は行列 $\mathbf{A}$ の第 (i, j) 成分。

***P-値 $= 2[1 - \Phi(T_{ji})]$ 。ただし、 $\Phi(x)$ は標準正規分布の分布関数、 T_{ji} は、帰無仮説 $H_0: (\hat{\Sigma})_{ji} = 0$ を検定する検定統計量

$T_{ji} = (\hat{\Sigma})_{ji} / \sqrt{((\mathbf{X}' \mathbf{X})^{-1})_{jj}(\hat{\Sigma})_{ii}}$, ($j=1, \dots, k$, $i=1, \dots, p$)

ことがわかる。また、 $wDBH\%$ の係数からは、5 m 半径内の林木の距離による重み付き平均胸高直径が大きい場合 (あるいは周辺林木の胸高直径が大きい場合)、対象の林木の胸高直径の飽和密度は小さくなり、逆に樹高については、大きくなる傾向がある。このことは、樹種の違いはあるが Hara *et al.* (1991) や Sumida *et al.* (1997) が指摘しているように、胸高直径が競合により被圧される場合、その胸高直径は大きくはならないが、樹高はより高く上に伸びようとする傾向があることを示している。

胸高直径の内的自然増加率 (y_2) に対しては、 $DBH\%$ の増加に伴いその値は減少する結果となり、現時点の胸高直径が大きいほど成長の立ち上がりが遅かったことがわかる。それに対し、P-値は大きくなるものの、 $wDBH\%$ の係数から、その値の増加に伴い内的自然増加率が増加し、成長の立ち上がりが早くなる傾向があった。また、樹高成長については、 $DBH\%$ 、 $wDBH\%$ それぞれに対し、負の推定値が得られた。 $DBH\%$ については胸高直径成長と同様、現時点の胸高直径が大きいほど樹高成長の立ち上がりが遅かった。それに対し、 $wDBH\%$ については、逆に周囲の林木との競合により被圧される結果、樹高成長の立ち上がりが遅くなる結果となった。

成長曲線の形状 (y_3) については、胸高直径成長および樹高成長とも同様に、 $DBH\%$ および $wDBH\%$ の増加に伴い、その値は減少する傾向となった。成長曲線の形状 (y_3) は、その値が増加すると、より S 字形の度合いが強くなり、逆に減少すると、S 字形の度合いが弱くなり、丸みを帯びる形状になる。

上記では、パラメータ (y_1, y_2, y_3) それぞれを通して $DBH\%$ 、 $wDBH\%$ の成長曲線への影響について個別に述べたが、実際の成長曲線のパラメータ (y_1, y_2, y_3) は $DBH\%$ 、 $wDBH\%$ により同時に決定されるため、次に $DBH\%$ 、 $wDBH\%$ それぞれの変化の成長曲線に及ぼす影響について直接調べてみた。図-4, 5 に $DBH\%$ 、 $wDBH\%$ の胸高直径成長曲線およ

図-4. 最適モデルにおける被説明変数の胸高直径成長曲線への影響

び樹高成長曲線への影響を示す。 $DBH\%$ 、 $wDBH\%$ のそれぞれのデータの統計値は平均 (16.315, 16.059), 最小値 (11.25, 14.6), 最大値 (22.4, 18.0) となったので, $DBH\%$ の変化については, $wDBH\%$ を10, 15, 20と固定し, $DBH\%$ を10から19まで1ずつ増加させ, また, $wDBH\%$ の影響についても同様の变化を仮定して曲線の変化を調べた。なお, 各図の左側の図は $wDBH\%$ を10, 15, 20と固定した $DBH\%$ の影響, 右側の図は $DBH\%$ を固定した $wDBH\%$ の影響をそれぞれ示す。また図中では $DBH\%$ を DBH^0 , $wDBH\%$ を $wDBH^5$ と記した。

図-4からわかるように, $DBH\%$ の胸高直径成長曲線への影響については, $wDBH\%$ の値が小さいとき, すなわち周囲林木の胸高直径が小さいとき, $DBH\%$ の増加に伴い, S字形の度合いが強くなりながら, 成長の立ち上がりが遅く

なり, かつ飽和密度が上昇する傾向が顕著であった。それに対し $wDBH\%$ が15, 20と増加するに従い, $DBH\%$ の増加に伴うS字形の度合いおよび成長の立ち上がりへの変化が弱くなり, 飽和密度の上昇だけが観察される結果となった。それに対し $wDBH\%$ の影響については, 基本的にその増加に伴い, S字形の度合いを増しながら成長の立ち上がりが遅くなる結果となったが, その変化の度合いは $DBH\%$ が大きいものほど顕著であった。

樹高成長曲線については, 図-5からわかるように, $wDBH\%$ の大小にかかわらず, $DBH\%$ の増加に伴い, 同様の飽和密度の上昇がみられた。また, $wDBH\%$ の影響についても $DBH\%$ の大小にかかわらず, その増加に伴い, 成長の立ち上がりが遅くなり, かつS字形の度合いが強くなることがわかる。

図-5. 最適モデルにおける被説明変数の樹高成長曲線への影響

V. お わ り に

本研究では、林分の成長を左右する要因探索のための統計的手法を提示した。この手法では、採用する成長曲線の係数の推定値を被説明変数とし、成長に影響を与えていると考えられる要因を説明変数とした多変量線形モデルにより候補のモデルを構築する。次に、得られるモデル群に対し、交差確認法に基づく情報量規準 CV 規準を用いて、その規準量を最小にするモデルを選択し、得られたモデルから成長を左右する要因を特定する。なお、ここで提示した方法はあくまでも CV 規準に基づくモデル選択による要因特定であるため、特定された要因が必ずしも研究者のニーズに合わない場合がある。その場合は、モデルの選択肢として、 CV 規準に基づき複数のモデル候補を準備する

ことにより、その中から最終的にニーズに合ったモデルを選択することが望ましい。

近年、林分成長データに対する多変量モデルの開発も Fang and Bailey (2001), Hall and Bailey (2001), Garber and Maguire (2003), Leites and Robinson (2004), Yanagihara and Yoshimoto (2004), 柳原ら (2004) などで見られるように、混合効果 (mixed-effects) を取り入れたモデル構築が主流になってきている。それらのモデルでも要因を考慮に入れたモデルリングを行うことは可能ではあるが、推定法が非常に煩雑になり、要因探索のための最適な説明変数の組み合わせを選択することは計算量的に困難である。そのため、ここで示した方法により、成長を左右する外的な要因を絞ることができれば、それらを組み入れた推定結果を少ない計算量で得ることができる。今回は

手法の提示を主な目的としたが、今後は実際の成長モデルへの適用を念頭に成長データおよびそれに関わる要因データの分析を進めると同時に、混合効果モデルへの組み込みも必要となろう。

引用文献

- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *In* International Symposium on Information Theory. Petrov, B.N. and Csáki, F. eds., Akadémiai Kiado, Budapest, 267-281.
- 安藤 貴 (1966) 林分密度管理図とその使い方. 126 pp. 農林出版, 東京.
- Arney, J.D. (1985) A modeling strategy for the growth projection of managed stands. *Can. J. For. Res.* 15 : 511-518.
- Bazaraa, M.S. and Shetty, C.M. (1979) Nonlinear Programming, Theory and Algorithms. John Wiley & Sons, New York.
- Bella, I.E. (1971) A new competition model for individual trees. *Forest Sci.* 17 : 364-372.
- Biging, G.S. and Dobbertin, M. (1992) A comparison of distance-dependent competition measures for height and basal area growth of individual conifer trees. *Forest Sci.* 38 : 695-720.
- Canham, C.D., LePage, P.T., and Coates, K.D. (2004) A neighborhood analysis of canopy tree competition: effects of shading versus crowding. *Can. J. For. Res.* 34 : 778-787.
- Daniels, R.F. (1976) Simple competition indices and their correlation with annual loblolly pine tree growth. *Forest Sci.* 22 : 454-456.
- Fang, Z. and Bailey, R.L. (2001) Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. *Forest Sci.* 47 : 287-300.
- Fujikoshi, Y., Noguchi, T., Ohtaki, M., and Yanagihara, H. (2003) Corrected versions of cross-validation criteria for selecting multivariate regression and growth curve models. *Ann. Inst. Stat. Math.* 55 : 537-553.
- Fujikoshi, Y., Yanagihara, H., and Wakaki, H. (2005) Bias corrections of some criteria for selecting multivariate linear models in a general nonnormal case. *Am. J. Math. Manage. Sci.* 25 : in press.
- 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. Manage.* 179 : 507-522.
- Hall, D.B. and Bailey, R.L. (2001) Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. *Forest Sci.* 47 : 311-321.
- Hann, D.W. (2003) ORGANON User's manual Edition 7.0. Dept. of Forest Resources, Oregon State Univ., Oregon, USA.
- Hara, T., Kimura, M., and Kikuzawa, K. (1991) Growth patterns of tree height and stem diameter in populations of *Abies veitchii*, *A. Mariessi* and *Betula ermanni*. *J. Ecol.* 79 : 1085-1098.
- 稲田充男 (2005) 穂の国「創造」. (<http://www.vector.co.jp/soft/win95/business/se360304.html>).
- 嘉戸昭夫・田中和博 (1995) タテヤマスギの樹幹解析資料から推定したシステム収穫表の成長パラメータ. *日林論* 105 : 185-188.
- Konohira, Y. (1995) Definition of the system yield table. *J. For. Plann.* 1 : 63-67.
- Kullback, S. and Leibler, R.A. (1951) On information and sufficiency. *Ann. Math. Stat.* 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. *Forest Sci.* 50 : 204-212.
- Matsumoto, M. (1997) Construction of yield tables for sugi (*Cryptomeria japonica*) in Kumamoto district using LYCS. *J. For. Plann.* 3 : 55-62.
- Pretzsch, H. (1997) Analysis and modeling of spatial stand structures. Methodological considerations based on mixed beech-larch stands in Lower Saxony. *For. Ecol. Manage.* 97 : 237-253.
- Pretzsch, H., Biber, P., and Dursky, J. (2002) The single tree-based stand simulator SILVA: construction, application and evaluation. *For. Ecol. Manage.* 162 : 3-21.
- Pukkala, T. and Miina, J. (1998) Tree-selection algorithms for optimizing thinning using a distance-dependent growth model. *Can. J. For. Res.* 28 : 693-702.
- Richards, F.J. (1958) A flexible growth function to empirical use. *J. Exp. Bot.* 10 : 290-300.
- 塩谷 實 (1990) 多変量解析概論. 204 pp. 朝倉書店, 東京.
- 白石則彦 (1986) 同齢単純林の成長予測に関する研究. *東大演報* 75 : 199-256.
- Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B36* : 111-147.
- Sumida, A., Ito, H., and Isagi, Y. (1997) Trade-off between height growth and stem diameter growth for an evergreen Oak, *Quercus glauca*, in a mixed hardwood forest. *Funct. Ecol.* 11 : 300-309.
- 竹内 啓 (1976) 情報統計量の分布とモデルの適切さの規準. *数理科学* 153 : 12-18.
- 田中和博 (1995) 林分表と樹高曲線から将来の林分表と樹高曲線を予測するシステム (シルプの森). 科研報告書「システム収穫表プログラム」(No. 0456019). 文部省, 東京. 22-32.
- Vettenranta, J. (1999) Distance-dependent models for predicting the development of mixed coniferous forests in Finland. *Silva Fenn.* 33 : 51-72.
- Vonesh, E.F. and Carter, R.L. (1992) Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics* 48 : 1-17.
- Wykoff, W.R., Crookston, N.D., and Stage, A.R. (1982) User's guide to the stand prognosis model. USDA For. Serv. Gen. Tech. Report. INT-133 : 112.
- Yanagihara, H. (2005) A family of estimators for multivariate kurtosis in a nonnormal linear regression model. *J. Multivar. Anal.* : in press.
- 柳原宏和・吉本 敦 (2003) 一般化多変量分散分析モデルの林木直径成長分析への適用可能性. *統計数理* 51 : 19-35.
- Yanagihara, H. and Yoshimoto, A. (2004) Statistical procedure for assessing the amount of carbon sequestered by sugi (*Cryptomeria japonica*) plantation. Discussion Paper Series No. 1076, Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba.
- 柳原宏和・吉本 敦・能本美穂 (2004) 林分成長分析のための一般化非線形混合効果モデル. (森林資源管理と数理モデル Vol. 3, 鹿又秀聡・吉本 敦 編, 森林計画学会出版局, 東京). 14-46.
- 柳井晴夫 (1974) 一般化決定係数による多変量解析の各種技法の統一的試み. *行動計量学* 1 : 46-54.