

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

柳原 宏和¹・吉本 敦²

(受付 2002年9月17日;改訂 2003年2月4日)

要 旨

一般化多変量分散分析モデル(GMANOVAモデル)は経時測定データに関する解析に適したモデルとして幅広く用いられ、多くの応用研究に適用されている。このモデルは成長データによく適用されることから成長曲線モデルとも呼ばれている。森林計画学の分野においては森林資源の将来的な成長予測など経営計画における基礎情報として直径成長データに基づく分析が行われているが、GMANOVAモデルの適用による分析はほとんど行われていない。本論文では、成長データへのあてはめ、仮説検定、変数選択を通して、森林計画学における林木直径成長の解析法としてのGMANOVAモデル適用の可能性を検討した。

キーワード：森林計画学，成長曲線モデル，多項式曲線，平均の同等性検定，変数選択。

1. はじめに

林分の経営は植林から始まり、地拵え、下草刈り、徐伐、間伐、主伐という一連の作業から形成され、それらの作業量・時期を適切に決めることにより効率的な経営の実現が達成できる。昨今では、林業経営に関わる経済環境が悪化しており、間伐放棄あるいは主伐後の植林放棄が目立ちはじめ、将来的な森林資源管理の荒廃が危惧されている。人工林が森林資源のほとんどを成す日本においては、人工林管理の遅れが森林環境を悪化させており、管理体制の見直しが急務となっている。こうした状況下、林野庁は循環型森林資源管理を目指し、間伐促進のための補助政策を施行している。

一般に、間伐は密集した立木本数を減少させ、残存木の成長を良くするために行われるものであるが、その方法は必ずしも一意に決まるものではない。すなわち、林分の経営には目的が存在し、その目的を達成するため間伐方法、時期、量が決定される。しかしながら、そのような意思決定を行うには、間伐が如何に林分の成長に影響を及ぼすかを把握しなくてはならない。現在は林分密度管理図(安藤(1982))が間伐効果をシミュレーションできるモデルとして代表的であり、その他にも地域に限定したモデルはあるものの、林分密度管理図のように汎用性に富んだモデルは未だ構築されていない。ただ、林分密度管理図は汎用性はあるものの、その反面、特定の林分への適用には様々な修正が必要となる。

林分成長に関するデータは、実験の対象になっている各個体について経時的、または繰り返

¹ 統計数理研究所(現 筑波大学 社会工学系：〒305-8573 茨城県つくば市天王台 1-1-1)

² 統計数理研究所(現 東北大学大学院 環境科学研究科：〒980-8579 宮城県仙台市青葉区荒巻字青葉 01)

し測定することによって得られる経時測定データである．このような経時測定データについては，一般的に良く知られている一元配置多変量分散分析モデル (one-way MANOVA モデル) を用いることで各林における木の成長の時点毎の平均の傾向を見ることはできる．しかしながら，この種のモデルでは異なる林分間の成長曲線の違いを表現することは出来ない．林分成長に関する分析では，各林木の成長の変化を経時的に記述し，かつ対象とするグループ毎に成長の変化を総合的に比較する必要がある．すなわち，同一林木の繰り返しデータに対して多項式等の曲線を適合し，さらに測定間の相関を考慮した多変量モデルを適用する必要がある．Potthoff and Roy (1964) によって提案された一般化多変量分散分析モデル (GMANOVA モデル) はこのような経時測定データに関する解析に適したモデルとして広く認められ，多くの応用研究結果が発表されている．例えば，von Rosen (1991)，藤越 (1993)，Kshirsagar and Smith (1995)，Srivastava and von Rosen (1999) 等の総合報告がある．なお，この GMANOVA モデルは，成長データによく適用されていることから成長曲線モデルとも呼ばれている．

本論文で取り上げる GMANOVA モデルは，観測時点がすべての標本において同じであるバランスデータに対して，それぞれのグループに関して同一の多項式をあてはめるモデルである．なお，同一でない多項式をあてはめるモデルとして Verbyla and Venables (1988)，観測値の観測時点がすべて一定ではないアンバランスデータに関しては Vonesh and Carter (1987)，アンバランスデータに非線形曲線をあてはめるモデルとして Vonesh and Carter (1992) 等によって提案された GMANOVA モデルから拡張されたモデルがあるが，今回はデータの特性から GMANOVA モデルを用いることとする．

本論文では，森林計画分野における成長解析に対し，ここで考慮する GMANOVA モデルの適用が有効か否かを分析することを目的とする．分析では，成長曲線のあてはめ，仮説検定，変数選択問題の統計的手法を通して，林分の成長データ解析を試みる．特に間伐効果の把握を念頭に，無間伐・適正間伐・強間伐それぞれが施行された林分の直径成長データを用いて分析を行う．以下，第 2 章では GMANOVA モデルに基づいた多項式曲線のあてはめ，第 3 章では GMANOVA モデルに基づいた仮説検定問題とその手法，第 4 章では GMANOVA モデルに基づいた説明変数の選択問題とその選択法とその手法での解析結果について述べる．

2. 成長曲線モデル

2.1 GMANOVA モデル

互いに相関のある p 個の観測値 y_1, y_2, \dots, y_p を一つの観測値の組として，そのようなデータが n 個あるようなデータを考える．つまり観測値が以下のような $n \times p$ の行列 Y で得られているとする．

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_n \end{pmatrix}.$$

ここで \mathbf{y}_i ($i = 1, 2, \dots, n$) はそれぞれの観測値の組を表す p 次元ベクトルであり，

$$\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ip})', \quad (i = 1, 2, \dots, n).$$

一般的にはこれらの観測値ベクトルはそれぞれ独立に得られたもので，互いに独立であると仮定している．

Potthoff and Roy (1964) によって提案された GMANOVA モデルは、二つの説明変数行列、

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix} = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_n \end{pmatrix}, \quad \text{rank}(A) = k (< n),$$

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pq} \end{pmatrix}, \quad \text{rank}(X) = q (\leq p),$$

を使い、観測値行列 Y に以下のような関係式を想定したモデルである。

$$(2.1) \quad Y = A\Xi X' + \mathcal{E}.$$

ただし \mathcal{E} は

$$(2.2) \quad \mathcal{E} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1p} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{np} \end{pmatrix} = \begin{pmatrix} \varepsilon'_1 \\ \varepsilon'_2 \\ \vdots \\ \varepsilon'_n \end{pmatrix},$$

のような誤差行列であり、それぞれの行ベクトルは独立に同一な平均 0、共分散 Σ を持つ p 次元正規分布に従うものである。観測値ベクトルで書き直すと、それぞれの y_i は独立に以下の p 次元正規分布に従うことになる。

$$y_i \sim N_p(X\Xi' \mathbf{a}_i, \Sigma), \quad (i = 1, 2, \dots, n).$$

上記より 行列 A はそれぞれの観測値ベクトルに対して個別に影響を与える変数であり、行列 X は共通に影響する変数であることがわかる。このことから 行列 A は個体間計画行列、行列 X は個体内計画行列と呼ばれる。以下、上記の正規性の仮定の下で分析を進める。なお、仮に真の誤差分布が正規分布ではない場合は、その非正規性が解析結果に少なからず影響を与える (Yanagihara(2001))。

2.2 林木直径成長データへのあてはめ

ここでは GMANOVA モデルを用いて、林分の林木直径成長データに多項式曲線をあてはめてみる。表 1 にはここで用いたデータ、林齢 28, 31, 35, 39, 47, 53, 58, 63, 68 年での林木の直径データの一部を示す。これらのデータは東京大学演習林矢竹沢試験地における強間伐、適間伐、無間伐を施した 3 種類の林分から得られたものである。それぞれの林分での立木本数の減少割合の推移を図 1 に示す。自然枯死が発生しているのわかり難いが、それぞれの林分において間伐を施した時期は 31 年時と 68 年時である。表 1 をもとにした林齢と直径の関係は図 2 の通りである。は無間伐、は強間伐、は適間伐を施した林分での直径成長過程を表している。

これらの成長データに対し、それぞれの間伐を行った林分別に林齢に相関を持たせた林齢に関する 2 次曲線を同時にあてはめるモデルを考える。曲線のあてはめに用いたデータは 68 年時まで伐られずに残っていたものの中から成長データの不備なものを取り除いたものである。最終的な立木本数は無間伐林で 187 本、強間伐林で 96 本、適間伐林で 86 本となり、合計で標

表 1. 間伐強度の異なった林分における直径成長 (cm) データ.

林分の種類	ID	林齢									
		28	31	35	39	47	53	58	63	68	
無間伐	1	30.5	32.3	34.2	35.8	37.2	38.7	39.1	40.2	40.2	
	2	31.4	33.0	34.7	36.5	39.4	42.2	43.9	47.1	49.6	
	3	20.4	21.6	23.2	24.5	26.8	28.7	30.0	31.5	32.3	
	7	33.7	35.3	36.9	38.7	40.6	42.5	43.5	45.1	47.3	
	9	25.6	27.3	29.2	31.6	33.9	36.2	37.6	39.3	41.0	
強間伐	2	31.1	32.9	35.5	38.3	42.6	46.4	49.4	51.9	53.8	
	9	30.1	31.7	33.5	34.9	37.4	39.3	40.8	42.0	42.6	
	10	23.1	25.5	27.7	30.5	33.5	36.4	38.5	40.0	40.8	
	16	36.4	38.1	41.9	45.6	50.5	55.0	59.6	62.6	64.7	
	22	27.1	27.5	29.7	32.4	35.0	37.0	38.1	39.8	41.0	
適間伐	3	26.6	27.8	30.1	33.3	37.5	40.0	41.5	44.7	45.7	
	5	29.2	30.1	32.5	35.6	40.0	41.6	45.5	46.7	49.9	
	8	23.8	25.7	28.9	32.4	35.3	38.2	39.7	40.3	41.0	
	12	26.9	28.3	30.0	33.0	37.0	39.9	41.6	42.0	43.0	
	19	26.8	26.9	28.4	29.8	31.9	34.4	35.5	37.4	37.9	

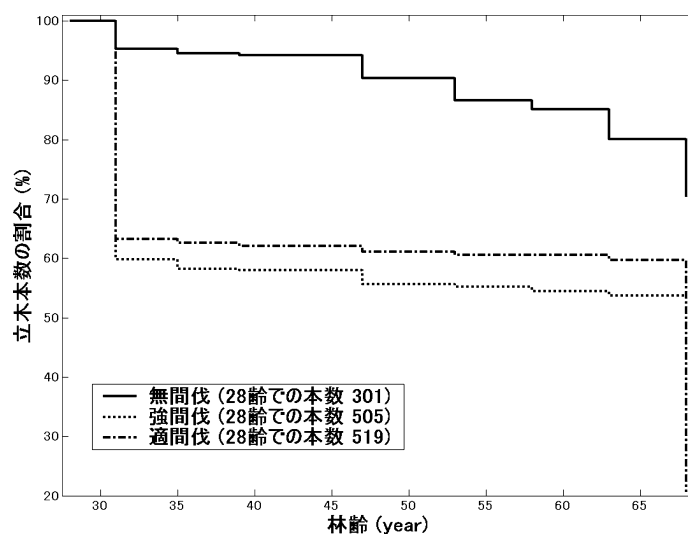


図 1. 間伐強度の異なる林分での立木本数減少推移.

本数は $n = 369$ である. 従って標本数が 369 と十分大きく, 非正規性の影響はそれほど問題にはならない (Yanagihara (2001)). なお, 本論文では取り扱わないが, 間伐された林木や自然枯死の林木のデータ, また欠測値を伴う林木のデータを使って解析を行いたい場合, Vonesh and Carter (1987) で提案されているアンバランスデータに基づく成長曲線モデルを用いると良い.

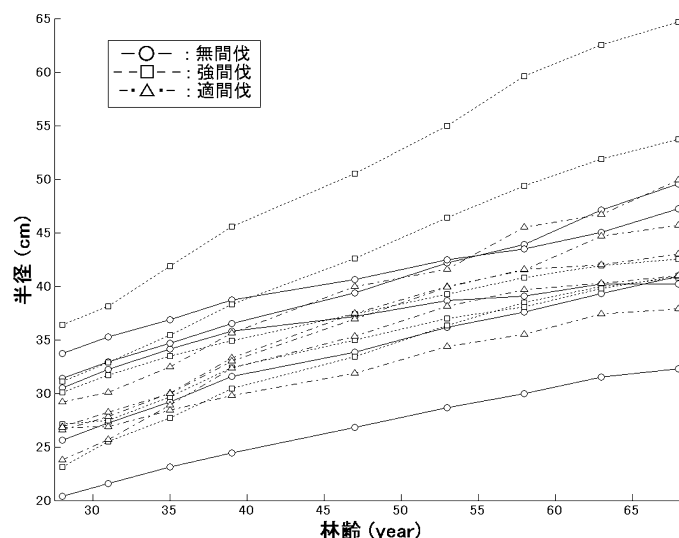


図 2. 表 1 のデータでの直径成長 .

まず, t を下記のようにそれぞれの林齢を成分に持つ 9 次元ベクトルとする .

$$t = (t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8, t_9)' = (28, 31, 35, 39, 47, 53, 58, 63, 68)'$$

直径の成長を林齢の 2 次曲線でそれぞれの林分別にあてはめれば,

$$(2.3) \quad y_{ij} = \alpha_i t_j^2 + \beta_i t_j + \gamma_i + \varepsilon_{ij}, \quad (i = 1, 2, \dots, 369, j = 1, 2, \dots, 9, l = 1, 2, 3),$$

となる . ここで $l = 1, 2, 3$ はそれぞれの林分を識別するための添え字であり, $l = 1$ であれば無間伐林, $l = 2$ であれば強間伐林, $l = 3$ であれば適間伐林を表している . 次にグループ分けに関するダミー変数を導入し, 無間伐林であれば $(1, 0, 0)$, 強間伐林であれば $(0, 1, 0)$, 適間伐林であれば $(0, 0, 1)$ となるようなベクトル a_i を考えると (2.3) は以下のように書き換えられる .

$$(2.4) \quad y_{ij} = a_i' \begin{pmatrix} \alpha_1 t_j^2 + \beta_1 t_j + \gamma_1 \\ \alpha_2 t_j^2 + \beta_2 t_j + \gamma_2 \\ \alpha_3 t_j^2 + \beta_3 t_j + \gamma_3 \end{pmatrix} + \varepsilon_{ij}, \quad (i = 1, 2, \dots, 369, j = 1, 2, \dots, 9).$$

従って, あてはめには前述した GMANOVA モデルを以下のようにすれば良いことが分かる .

$$(2.5) \quad A = \begin{pmatrix} a_1' \\ a_2' \\ \vdots \\ a_{369}' \end{pmatrix}, \quad X = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_9 & t_9^2 \end{pmatrix}, \quad \Xi = \begin{pmatrix} \gamma_1 & \beta_1 & \alpha_1 \\ \gamma_2 & \beta_2 & \alpha_2 \\ \gamma_3 & \beta_3 & \alpha_3 \end{pmatrix},$$

上記の GMANOVA モデルにおいて, 成長の経時的な変動を推論するには Ξ, Σ を推定する必要がある . それらの推定に関しては, 尤度関数の最大化によって未知パラメータを決定する最尤推定法を用いることが一般的である . 尤度関数はデータの同時分布の確率密度関数のことであり, 最尤推定法では, 与えられたデータの下でその出現確率が最大になるように Ξ と

Σ の推定値が決定される．実際には，尤度関数の対数値である対数尤度関数の最大化によって Ξ, Σ の推定値を決定する．その方法は以下の通りである．

本論文で用いる GMANOVA モデルではそれぞれの観測値ベクトルに独立に p 次元正規分布を仮定しているので， Y の同時確率密度関数は以下ようになる．

$$f(Y; \Xi, \Sigma) = \prod_{i=1}^n \left(\frac{1}{2\pi} \right)^{p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_i - X\Xi'\mathbf{a}_i)' \Sigma^{-1} (\mathbf{y}_i - X\Xi'\mathbf{a}_i) \right\}.$$

よって対数尤度関数は

$$\begin{aligned} (2.6) \quad \ell(\Xi, \Sigma | Y) &= \log \{f(Y; \Xi, \Sigma)\} \\ &= -\frac{n}{2} \log |\Sigma| - \frac{np}{2} \log(2\pi) \\ &\quad - \frac{1}{2} \text{tr} \{ (Y - A\Xi X')' \Sigma^{-1} (Y - A\Xi X') \}, \end{aligned}$$

となる． Ξ と Σ の推定量 $\hat{\Xi}, \hat{\Sigma}$ は対数尤度関数の最大化によって決定されるので，

$$\frac{\partial}{\partial \Xi} \ell(\Xi, \Sigma | Y) = 0, \quad \frac{\partial}{\partial \Sigma} \ell(\Xi, \Sigma | Y) = 0,$$

を同時に満たすことにより求められる．すなわち，

$$(2.7) \quad \hat{\Xi} = (A'A)^{-1} A' Y S^{-1} X (X' S^{-1} X)^{-1}, \quad \hat{\Sigma} = \frac{1}{n} (Y - A\hat{\Xi} X')' (Y - A\hat{\Xi} X'),$$

である．ただし

$$(2.8) \quad S = Y'(I_n - P_A)Y.$$

ここで P_A は行列 A で張る空間への射影行列であり， $P_A = A(A'A)^{-1}A'$ で定義される．

実際に 369 個の直径成長のデータに 2 次曲線をあてはめた時の Ξ と Σ の推定量は以下の通りとなった．

$$\hat{\Xi} = \begin{pmatrix} 11.74 & 0.523 & -0.003 \\ 8.758 & 0.716 & -0.004 \\ 11.21 & 0.631 & -0.003 \end{pmatrix},$$

$$\hat{\Sigma} = \begin{pmatrix} 17.20 & 17.65 & 18.77 & 19.98 & 20.94 & 22.06 & 23.73 & 24.59 & 25.29 \\ 17.65 & 18.57 & 19.86 & 21.36 & 22.43 & 23.72 & 25.54 & 26.55 & 27.34 \\ 18.77 & 19.86 & 21.49 & 23.20 & 24.57 & 26.10 & 28.18 & 29.36 & 30.32 \\ 19.98 & 21.36 & 23.20 & 25.80 & 27.55 & 29.55 & 32.03 & 33.56 & 34.78 \\ 20.94 & 22.43 & 24.57 & 27.55 & 30.34 & 32.84 & 35.80 & 37.61 & 39.17 \\ 22.06 & 23.72 & 26.10 & 29.55 & 32.84 & 35.95 & 39.27 & 41.43 & 43.29 \\ 23.73 & 25.54 & 28.18 & 32.03 & 35.80 & 39.27 & 43.28 & 45.67 & 47.77 \\ 24.59 & 26.55 & 29.36 & 33.56 & 37.61 & 41.43 & 45.67 & 48.70 & 50.95 \\ 25.29 & 27.34 & 30.32 & 34.78 & 39.17 & 43.29 & 47.77 & 50.95 & 54.13 \end{pmatrix}.$$

図 3 に 2 次曲線をあてはめた結果を示す．図からも分かるように間伐を行うことにより成長速度が速くなる傾向がある．さらに強間伐よりも適間伐の方がその傾向が強くなっていることもわかる．またデータの共分散の推定量より，林齢が大きくなる，つまり直径が大きくなればなるほど，その成長のばらつきが大きくなっていることもわかる．

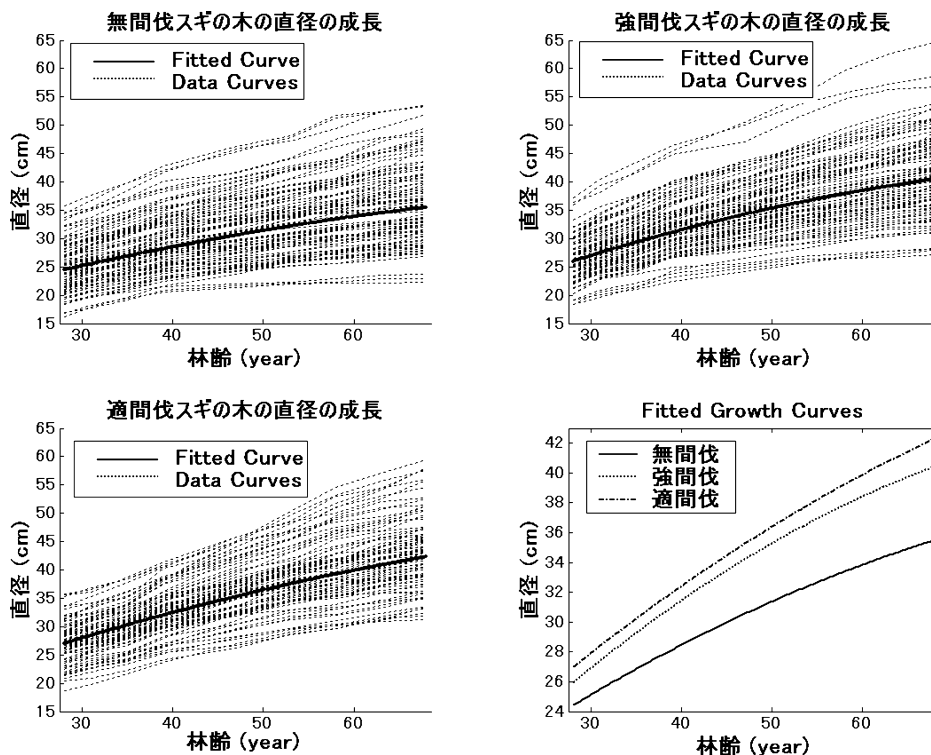


図 3. 間伐強度の異なった林分における成長曲線モデル当てはめ結果：2次曲線。

3. GMANOVA モデルに基づく仮説検定法

第 2 章では GMANOVA モデルの未知パラメーター(あてはめる多項式の係数, 共分散行列)の推定について述べたが, 実際の解析においては, 得られた値がどのような傾向を持つのかを調べることも重要な問題である. 例えば, 第 2 章で用いたデータの分析では, 実際に多項式をあてはめ, 各林分での各時点における平均成長の違いを見るのが主な目的となるが, 同時に林分間の経時的な成長に違いがあるか否かを調べることも重要な課題である. もし, 各林分での成長データに対し, 異なる強度の間伐を施しても経時的な成長に変化がなければ, 間伐を施す必要性がなくなることになる. この章では, 第 2 章で使用した各林分の直径成長データを用いてそれらの問題に対応する仮説検定法について述べる. なお, 本論文では成長の差は間伐強度の差によるものと仮定して解析を行う. 実際には様々な影響が存在するが, それらの情報が入手できないため, このような仮定に基づいた.

ここで考慮する検定では帰無仮説を

$$H_0: \text{各林分での成長はすべて同じである,}$$

とし, この帰無仮説が正しいかどうかを調べる検定を考える. 対立仮説は,

$$H_1: \text{各林分での成長はすべて同じではない,}$$

となる. なお, 仮説検定は帰無仮説が正しいかどうかを調べるものであるが, 帰無仮説が正し

くないからといって、対立仮説が正しいというわけではない。どちらかと言えば、帰無仮説が正しくない結果になった場合は、積極的ではないにしろ、対立仮説を支持しても問題はないと解釈することが妥当であろう。

数式を用いてこの帰無仮説を表してみる。帰無仮説は各林分での成長はすべて同じとしているので、この仮説はそれぞれあてはめた多項式が等しいことと同値である。第2章の2次関数(2.3)のあてはめの例で考えると、帰無仮説は

$$H_0 : \alpha_1 t_j^2 + \beta_1 t_j + \gamma_1 = \alpha_2 t_j^2 + \beta_2 t_j + \gamma_2 = \alpha_3 t_j^2 + \beta_3 t_j + \gamma_3,$$

となる。実際に多項式が等しいか否かはその係数の同等性を見れば良いので、この帰無仮説は

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 \text{ かつ } \beta_1 = \beta_2 = \beta_3 \text{ かつ } \gamma_1 = \gamma_2 = \gamma_3,$$

となる。これを行列表示にすると、

$$H_0 : \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \gamma_1 & \beta_1 & \alpha_1 \\ \gamma_2 & \beta_2 & \alpha_2 \\ \gamma_3 & \beta_3 & \alpha_3 \end{pmatrix} = 0,$$

と書き換えられる。更に上記の左の行列を C , Ξ を(2.5)のように置くと、帰無仮説と対立仮説は以下の通りとなる。

$$H_0 : C\Xi = 0, \quad \text{vs} \quad H_1 : C\Xi \neq 0.$$

従って仮説検定は、このような未知パラメーターの傾向について調べればよいことになる。この傾向が線形関係を持つのであれば、帰無仮説と対立仮説は $c \times k$ 既知行列 C ($\text{rank}(C) = c \leq k$), $q \times d$ 既知行列 D ($\text{rank}(D) = d \leq q$) を用いて、一般的に

$$(3.1) \quad H_0 : C\Xi D = 0 \quad \text{vs} \quad H_1 : C\Xi D \neq 0,$$

と表すことが出来る。

この既知行列 C と D を変えることにより、様々な仮説を表すことができる。例えば、今ここで、未知パラメーター行列を

$$\Xi = \begin{pmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1q} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{k1} & \xi_{k2} & \cdots & \xi_{kq} \end{pmatrix} = \begin{pmatrix} \xi'_1 \\ \xi'_2 \\ \vdots \\ \xi'_k \end{pmatrix},$$

とおく。ここで、 $\mathbf{1}_k$ をすべての成分が1であるような k 次元ベクトルとし、

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix} = (I_{k-1} \quad -\mathbf{1}_{k-1}),$$

とおく。このとき $C\Xi = 0$ と

$$\xi_1 = \xi_2 = \cdots = \xi_k,$$

は同値である。 ξ_j はそれぞれのグループにおける多項式の係数列であるから、上記関係式は、それぞれのグループでの多項式曲線がすべて同じであるということを意味していることになる。

また、上記の検定法を用いればそれぞれの係数を個別に比較することもできる。例えば、直線をあてはめたときに、その直線の傾きが等しいという仮定は、 D を操作することにより構成できる。すなわち、 D を第 1 成分だけ 1 で後はすべて 0 であるという q 次元ベクトル、つまり $D = (1, 0, \dots, 0)'$ と置くことにより、それぞれの ξ_i の第 1 成分だけ比較することが可能である。このとき、 $C\Xi D = 0$ は

$$\xi_{11} = \xi_{21} = \dots = \xi_{k1},$$

と同値である。もちろん、この D の形を変えることで、好きな成分を比較できる。またそれらのベクトルを複数個用いることで、いくつかの成分を同時に比較することが可能である。つまり、

$$D = \begin{pmatrix} I_2 \\ 0 \end{pmatrix},$$

であるような $q \times 2$ 行列を使うと、それぞれの ξ_i の第 1, 2 成分を同時に比較することが可能である。このとき $C\Xi D = 0$ は

$$\xi_{11} = \xi_{21} = \dots = \xi_{k1} \quad \text{かつ} \quad \xi_{12} = \xi_{22} = \dots = \xi_{k2},$$

と同値である。一方、 $C = I_k$, $D = I_q$ とすると、 Ξ が 0 であるかどうかという仮説を表すこともできる。更に、新しく Ξ を $\Xi - \Xi_0$ と置き換えることで、 Ξ が与えられたある値 Ξ_0 であるかどうかという仮説を表すこともできる。このように C と D の表現の仕方で、様々な線形関係を表すことが出来ること(その他の仮説と行列 C , D の関係の例は Kshirsagar and Smith(1995), p. 40, Yanagihara(2001)等を参照)。

上記の帰無仮説に対する仮説検定を行う仮説検定統計量にも、尤度比検定統計量、ローレイ=ホテリングトレース規準、パートレット=ナンダ=ピライトレース規準等、様々なものがあるが、今回は尤度比検定統計量を使うことにする。この統計量は、尤度比という言葉からわかるように、帰無仮説が正しいという制約の下での標本尤度と、その制約が無い場合での標本尤度の比である。(3.1) のような仮説を検定する時のパートレット補正(Bartlett(1937))を行った尤度比検定統計量は、以下のようなものである。

$$(3.2) \quad T = -\{n - k - (p - q) - (d - c + 1)/2\} \log(\Lambda),$$

ただし、

$$\begin{aligned} \Lambda &= |S_e|/|S_e + S_h|, \\ S_e &= D'(X'S^{-1}X)^{-1}D, \quad S_h = (C\hat{\Xi}D)'(CRC')^{-1}(C\hat{\Xi}D), \\ R &= (A'A)^{-1} + (A'A)^{-1}A'Y\{S^{-1} - S^{-1}X(X'S^{-1}X)^{-1}X'S^{-1}\}Y'A(A'A)^{-1}. \end{aligned}$$

ここでの $\hat{\Xi}$ や S は第 2 章の(2.7)と(2.8)で定義したものである。帰無仮説 H_0 が正しいという下で、 Λ はウィリックスのラムダ分布に従うことが知られている。この分布を使うことで正確な α パーセント棄却域を構成することができるが、このラムダ分布のパーセント点を導出することは、なかなか困難なものである。そこで今回は、 T の漸近分布である自由度 cd の χ^2 分布のパーセント点を用いて棄却域を構成する。検定の有意水準が $\alpha\%$ になる検定では、 χ^2 分布の上側 $\alpha\%$ 点を用いる。この cd は H_0 が正しいという仮定の下で Ξ の自由に動けるパラメーターの数である。 T はパートレット補正を行った尤度比統計量であるので、補正後の帰無分布の漸近分布に対するオーダーは $O(n^{-2})$ である(Barndorff-Nielsen and Hall(1988))。つま

り $G_{cd}(x)$ を自由度 cd の χ^2 分布の分布関数とすると, その帰無分布のオーダーが,

$$P(T \leq x) = G_{cd}(x) + O(n^{-2}),$$

となる(参照: Fujikoshi(1974), Srivastava and von Rosen(1999)). 標本数が $n = 369$ と大きいことを考慮にいれると, この漸近近似は妥当なものと考えられる. また, この T の帰無分布はその非正規性の影響を受けるが, n が 369 と大きいこと, さらに平均の同等性検定であるので, このデータに関する検定は非正規性の影響を受け難い(柳原 2001)). 実際の検定結果は以下ようになる. 第 2 章のようにそれぞれの林分での直径成長データに 2 次関数をあてはめ, 2 次曲線で表した成長曲線が 3 つの林分ですべて等しいか否かを有意水準 5% で検定してみる. この仮説とモデルを表す行列は,

$$C = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \quad D = I_3, \quad A = \begin{pmatrix} 1_{187} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1_{96} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1_{86} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_9 & t_9^2 \end{pmatrix},$$

となる. このモデルと仮説に基づいて, 検定統計量を計算してみると, $T = 77.3901$ となる. $c = 2, d = 3$ より T は仮説 H_0 が正しい時に漸近的に自由度 6 の χ^2 分布に従う. 自由度 6 の χ^2 の上側 5% 点は 12.5916 であり, $T > 12.5916$ となる. 従って, 帰無仮説 H_0 を棄却する結果となる.

また間伐効果に関して言えば, 間伐強度の成長への影響も関心のある問題である. この問題, 例えば, 適間伐と強間伐を施行した林分での成長曲線が等しいか否かを検定する場合, 先ほどの 2 次曲線のあてはめで考えると, 仮説とモデルは,

$$C = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \quad D = I_3, \quad A = \begin{pmatrix} 1_{187} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1_{96} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1_{86} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_9 & t_9^2 \end{pmatrix},$$

となる. この場合, T は自由度 3 の χ^2 分布に従う. $T = 10.0726$ となり, 自由度 3 の χ^2 分布の上側 5% 点は 7.8147 であるので, 帰無仮説は棄却されることになる.

しかしながら, 仮説検定を数回繰り返す場合, 全体の有意水準が変化してしまうために, 異なった方法で棄却域を構成しなくてはならない. 今回はボンフェロニの不等式により棄却域を構成する.

$H_{0,1}$: 各林分での成長曲線は同じである,

$H_{0,2}$: 適間伐と強間伐を施行した林分での成長曲線は同じである.

このような帰無仮説を検定する検定統計量をそれぞれ T_1, T_2 とすると, ボンフェロニの不等式により,

$$P(T_1 \geq x_1, T_2 \geq x_2 \mid H_{0,1}, H_{0,2}) \leq P(T_1 \geq x_1 \mid H_{0,1}) + P(T_2 \geq x_2 \mid H_{0,2}),$$

となる. それぞれの棄却域を自由度 6 と 3 の χ^2 分布の上側 $\alpha/2\%$ 点とすると, 全体の有意水準を $\alpha\%$ に押さえることが出来る. 有意水準 5% の検定の場合, 自由度 6 と 3 の χ^2 分布の上側 2.5% 点はそれぞれ, 14.4494, 9.3484 であり, $T_1 = 77.3901, T_2 = 10.0726$ であるから, 有意水準 5% で両方の仮説を棄却することになる.

間伐の影響については, 間伐により直径分布に偏りが生じ, その結果間伐強度にかかわらず初めから直径平均に差があるように見えるかもしれない. 実際に経営されている林分では,

小径木からの間伐，大径木からの間伐とその仕方は経営の目的により様々であるため，その間伐前と間伐後の林木直径分布を把握しておく必要がある．しかしながら，本論文で用いたデータではそのような分布に関する情報は無く，従って，直径に関する経時測定データ y_{ij} ($i = 1, 2, \dots, 369, j = 1, 2, \dots, 9$) を用いて 28 林齢との直径成長差のデータ $z_{ij} = y_{i, j+1} - y_{i1}$ ($i = 1, 2, \dots, 369, j = 1, 2, \dots, 8$) を新しく作り，強度の違いをその成長の変化量の違いとして捉え，検定を試みる．その際，2 次曲線，

$$z_{ij} = \mathbf{a}'_i \begin{pmatrix} \alpha_1(t_{j+1} - t_1)^2 + \beta_1(t_{j+1} - t_1) \\ \alpha_2(t_{j+1} - t_1)^2 + \beta_2(t_{j+1} - t_1) \\ \alpha_3(t_{j+1} - t_1)^2 + \beta_3(t_{j+1} - t_1) \end{pmatrix} + \varepsilon_{ij}, \quad (i = 1, 2, \dots, 369, j = 1, 2, \dots, 8),$$

をあてはめ，仮説検定を行う．個体間計画行列は共通のものであるが， Ξ と個体内計画行列 X は，

$$\Xi = \begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{pmatrix}, \quad X = \begin{pmatrix} t_2 - 28 & (t_2 - 28)^2 \\ t_3 - 28 & (t_3 - 28)^2 \\ \vdots & \vdots \\ t_9 - 28 & (t_9 - 28)^2 \end{pmatrix},$$

となることに注意していただきたい．あてはめ結果を図 4 に示す．さらに $H_{0,1}$ を検定すると

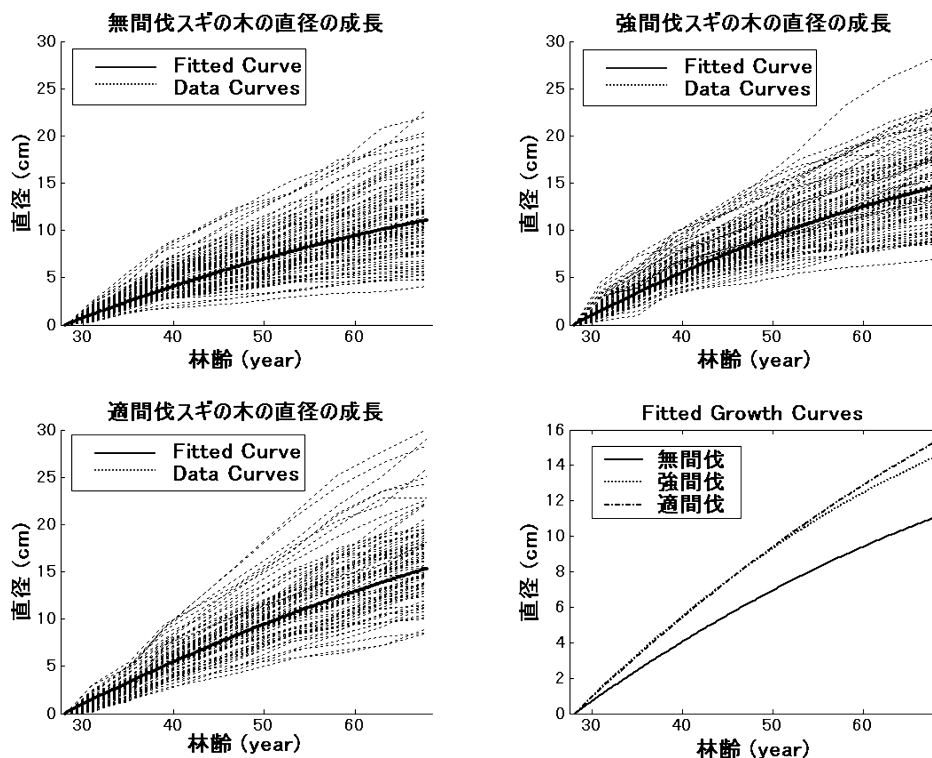


図 4. 間伐強度の異なった林分における成長差に対する曲線当てはめ結果：2 次曲線．

きには $C = (I_2 \ -1_2)$, $D = I_2$, また $H_{0,2}$ を検定するときには $C = (0 \ 1 \ -1)$, $D = I_2$, となる. このデータを用いたときの $H_{0,1}$, $H_{0,2}$ を検定する検定統計量は $T_1 = 58.1590$, $T_2 = 9.9260$ となる. 自由度 4 と 2 の χ^2 分布の上側 2.5% 点はそれぞれ, 11.1433 と 7.3778 であるので有意水準 5% で両方の仮説を棄却することになる.

以上のことから, それぞれの林分での成長曲線は同じではないということが言える. つまり強度の異なる間伐を施すことにより, 林分の平均的な成長に差が生じるばかりでなく, その経時的な成長にも差があるということになる. また適間伐や強間伐といった間伐強度による成長に有意な差があると言えることがわかる.

4. GMANOVA モデルに基づくモデル選択法

GMANOVA モデルの適用の際に問題となるものの一つに想定したモデルが望ましいものであるか否かということがある. 多項式曲線をあてはめる成長曲線モデルを例にあげれば, あてはめる多項式の次数をいくつにすれば良いかという問題である. このような問題に対し, AIC (赤池情報量規準, Akaike's Information Criterion) の最小化によって説明変数の個数を決定する方法がある. 以下その方法を紹介する.

Akaike (1973) によって提案された AIC は, カルバック=ライブラーの距離によりモデルの差を定義し, それをリスクとして提案された規準量である. 今, 候補のモデルとして,

$$(4.1) \quad M_{(j)} : Y \sim N_{pn}(A_{(j)}\Xi_{(j)}X'_{(j)}, \Sigma_{(j)}),$$

(j はそれぞれの候補のモデルに対応するインディックス)

を考える. ただし $A_{(j)}$, $X_{(j)}$ はそれぞれ $n \times k_{(j)}$, $p \times q_{(j)}$ 行列である. このモデルでは説明変数を $k_{(j)} + q_{(j)}$ 個用いていることになる. 基本的には候補のモデルはそれぞれ考慮する説明変数の組み合わせの数だけある. これらの候補のモデルに対して, AIC は以下のように定義される.

$$(4.2) \quad AIC_{(j)} = -2 \log\{f(Y; \hat{\Xi}_{(j)}, \hat{\Sigma}_{(j)})\} + 2 \times (\text{独立パラメーター数})$$

$$= n \log |\hat{\Sigma}_{(j)}| + np \{\log(2\pi) + 1\} + 2 \left\{ k_{(j)}q_{(j)} + \frac{1}{2}p(p+1) \right\}$$

ただし, $\hat{\Xi}_{(j)}$, $\hat{\Sigma}_{(j)}$ は候補のモデル $M_{(j)}$ のもとで計算される最尤推定量であり, 計算された中で AIC が最も小さくなるモデルが最適なモデルとなる.

本論文で取り扱った林木の直径成長のデータに関しては問題にはならないが, 標本数 n に対して, 時点数 p , 説明変数の個数 $k_{(j)}$ が大きい場合は, AIC のバイアス補正項 “ $2 \times$ 独立パラメーター数” ではリスクに対するバイアスを補正しきれず, より複雑なモデルを選ぶことがある. そのような場合, AIC のバイアスを修正した $CAIC$ (Corrected AIC , Sugiura (1978)) の適用が望ましい. GMANOVA モデルにおける多項式の次数の選択問題における $CAIC$ は Satoh et al. (1997) により以下のように提案されている.

$$CAIC_{(j)} = n \log |\hat{\Sigma}_{(j)}| + np \log(2\pi)$$

$$+ \frac{n^2(p - q_{(j)})}{n - p + q_{(j)} - 1} + \frac{nq_{(j)}(n - k_{(j)} - 1)(n + k_{(j)})}{(n - k_{(j)} - p - 1)(n - k_{(j)} - p + q_{(j)} - 1)}.$$

また, 誤差分布の正規性の仮定が満たされず実際の分布が非正規分布であった場合, AIC のバイアス項を補正するために非正規性を示す尖度パラメーターが必要となるので注意しなくてはならない(藤越 他(2002)).

林木の直径成長データを用いて AIC により適用する曲線の次数を決定してみる．あてはめる多項式として，

$$\begin{aligned} M_{(1)} : & 1 \text{ 次多項式}, & M_{(2)} : & 2 \text{ 次多項式}, \\ M_{(3)} : & 3 \text{ 次多項式}, & M_{(4)} : & 4 \text{ 次多項式}, \end{aligned}$$

の 4 つの多項式を考える．それぞれのモデルでの個体内計画行列は，

$$\begin{aligned} X_{(1)} &= \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_9 \end{pmatrix}, & X_{(2)} &= \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_9 & t_9^2 \end{pmatrix}, \\ X_{(3)} &= \begin{pmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_9 & t_9^2 & t_9^3 \end{pmatrix}, & X_{(4)} &= \begin{pmatrix} 1 & t_1 & t_1^2 & t_1^3 & t_1^4 \\ 1 & t_2 & t_2^2 & t_2^3 & t_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t_9 & t_9^2 & t_9^3 & t_9^4 \end{pmatrix}, \end{aligned}$$

となる．次元 $p = 9$ ，個体間計画行列 A はすべてのモデルで共通 $k_{(j)} = k = 3$ なので，それぞれのモデルでの AIC は以下ようになった．

$$\begin{aligned} AIC_{(1)} &= 369 \times \log |\hat{\Sigma}_{(1)}| + 3321 \times \{\log(2\pi) + 1\} + 2(3 \times 2 + 21) = 8346.9, \\ AIC_{(2)} &= 369 \times \log |\hat{\Sigma}_{(2)}| + 3321 \times \{\log(2\pi) + 1\} + 2(3 \times 3 + 21) = 8117.0, \\ AIC_{(3)} &= 369 \times \log |\hat{\Sigma}_{(3)}| + 3321 \times \{\log(2\pi) + 1\} + 2(3 \times 4 + 21) = 8022.1, \\ AIC_{(4)} &= 369 \times \log |\hat{\Sigma}_{(4)}| + 3321 \times \{\log(2\pi) + 1\} + 2(3 \times 5 + 21) = 7969.6, \end{aligned}$$

これらの結果について，横軸に多項式の次数，縦軸に AIC の値をとった折れ線グラフを図 5 に示す．この図からわかるように，1 次から 4 次多項式では 4 次多項式の当てはまりが一番

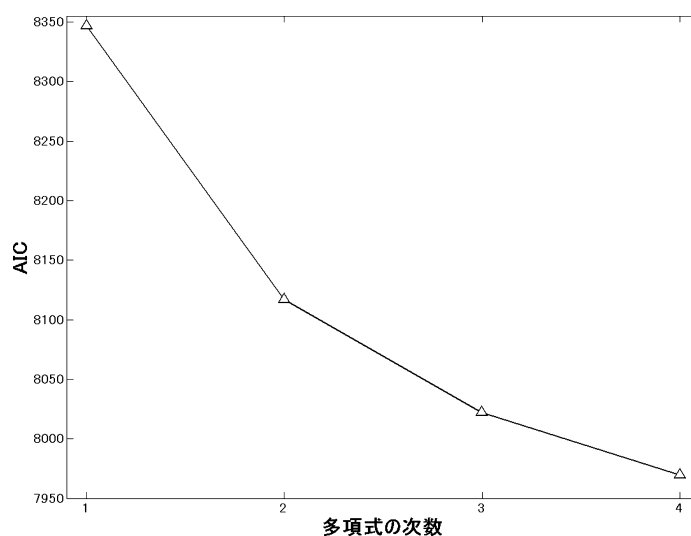


図 5. 多項式の次数と AIC の関係．

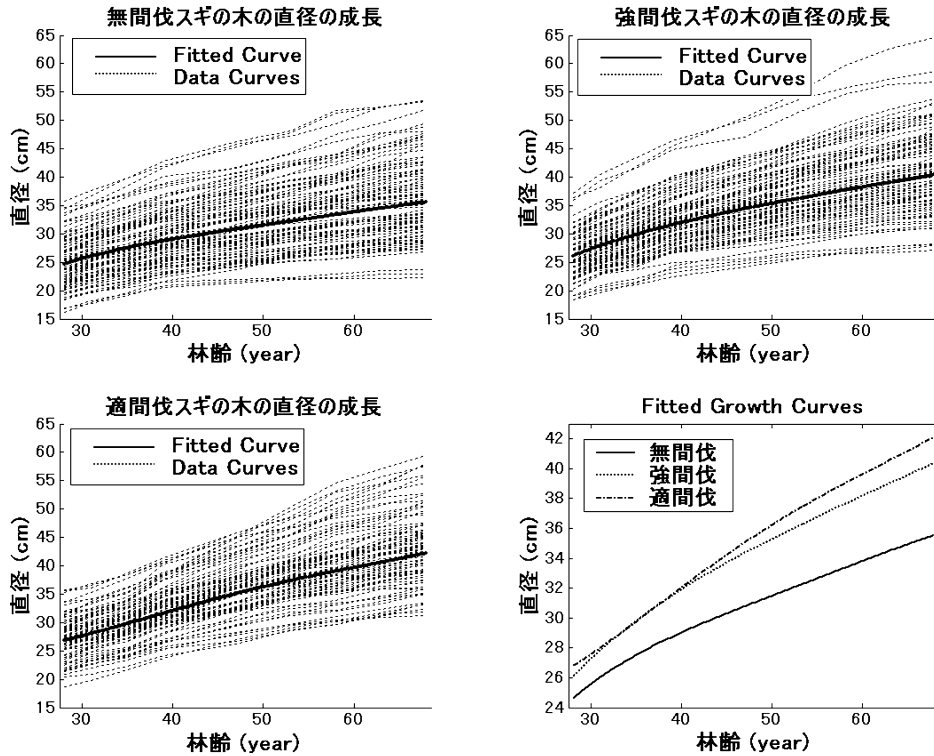


図6. 間伐強度の異なった林分における成長曲線モデル当てはめ結果：4次関数。

妥当であるという結果になった。図6に得られた4次多項式モデルの当てはめ結果を示す。図3の結果と比較し、強間伐を施した林分と適間伐を施した林分の成長の差の違いが明確に観察された。2次曲線を用いた場合は、林齢に関わらず、成長に差が観察されたが、4次多項式を用いた場合は、林齢が若いときには差がほとんどなく、高齢になると差が大きくなるのがわかる。このように、選択するモデルにより分析結果が異なってくるので、最終的に採用するモデルの決定にはここで紹介したような方法によることが望ましい。

5. おわりに

本論文では、一般化多変量分散分析モデルの森林計画分野における成長分析への適用可能性について分析を行った。経時測定データを対象とするこのモデルは特に成長に関わるデータ解析に応用されており、成長曲線モデルとして幅広く認知されている。森林計画分野において、同様に林木の成長データの収集から分析に至る研究が行われているが、ここで考慮した成長曲線モデルの適用は試みられていない。本論文では、まずモデルの当てはめ、最尤推定法による未知パラメータの推定法を示した。次に成長過程の異なる林分における比較検定法を示し、最後にモデル選択に関わる規準について記述した。これら一連の分析法の適用により、林分成立過程の異なる様々な成長データへの成長曲線モデルの適用が可能になり、成長過程の比較検定ができることが分かった。

ここでは、異なる間伐強度を施した林分から得られた経時測定データである直径成長データ

を用いて成長データの分析への成長曲線モデルの適用を試みた。分析の結果、成長曲線モデルの適用により各林分間の成長の違いが統計的に把握できることが示された。2次曲線によるモデルを用いた場合、間伐を施した方が林木の直径成長が良くなることが分かったが、強度に間伐するよりは、ある程度適当と思われる間伐の方がその効果は増すものと考えられた。すなわち、間伐による林木1本当たりの成長に対する空間確保は重要ではあるが、ある程度の林木間の競争が成長を促進するものと考えられる。また各林分での成長過程の違いを把握するモデル選択には、*AIC* などの規準量に基づくことが重要であることも分かった。本論文では2次曲線より4次関数のモデルの方が成長の格差をより明確に表すことができるという結果になった。一般に林木の成長はシグモイド型であることが知られており、今後は更に Vonesh and Carer (1992) のような非線形の成長曲線モデルによる分析が必要である。

ここで対象としたデータは間伐強度が異なる林分から得られた成長データである。仮にこれらのデータから間伐効果の把握が容易に行われるようなモデルが構築されれば、間伐効果を考慮したモデルが容易に構築できることになる。現在は、様々なデータ収集は行われているものの、データ解析、解析結果の応用モデルへの適用などの一連の分析システムが十分に確立されているとは言い難い。昨今では林業を取り巻く経済環境の悪化に伴い、間伐の行き届かない森林が目立ってきている。人工林である以上、人為による管理が必要不可欠である。間伐という行為が如何に森林の成長過程に影響を及ぼし、経済的にどれくらいの価値があるのかといったことを把握しなければ、効率的な森林資源管理、あるいは現在林野庁が提示している循環型森林資源管理の実現は困難なものとなろう。そうした意味においても、ここで示した分析方法の開発は急務である。

謝 辞

本論文で用いた林木の成長データを提供して下さった東京大学秩父演習林の皆様、また有益なコメントを頂いた査読者の皆様に厚く御礼申し上げます。

参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémiai Kiado, Budapest.
- 安藤 貴 (1982). 『林分の密度管理』, 農林出版, 東京.
- Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic, *Biometrika*, **75**, 374–378.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests, *Proc. Roy. Soc. London Ser. A*, **160**, 268–282.
- Fujikoshi, Y. (1974). Asymptotic expansions of the non-null distributions of three statistics in GMANOVA, *Ann. Inst. Statist. Math.*, **26**, 289–297.
- 藤越康祝 (1993). 成長曲線モデル その理論と応用, 数理科学, **358**, 60–66.
- 藤越康祝, 柳原宏和, 若木宏文 (2002). 多変量非正規回帰モデルにおける変数選択について, 2002年度統計関連学会連合大会講演報告集, 405–406.
- Kshirsagar, A. M. and Smith, W. B. (1995). *Growth Curves*, Marcel Dekker, New York.

- Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51**, 313–325.
- Satoh, K., Kobayashi, M. and Fujikoshi, Y. (1997). Variable selection for the growth curve model, *J. Multivariate Anal.*, **60**, 277–292.
- Srivastava, M. S. and von Rosen, D. (1999). Growth curve models, *Multivariate Analysis, Design of Experiments, and Survey Sampling* (ed. S. Ghosh), 547–578, Marcel Dekker, New York.
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections, *Commun. Statist. Theory Methods*, **A7**, 13–26.
- Verbyla, A. P. and Venables, W. N. (1988). An extension of the growth curve model, *Biometrika*, **75**, 129–138.
- Vonesh, E. F. and Carter, R. L. (1987). Efficient inference for random-coefficient growth curve models with unbalanced data. *Biometrics*, **43**, 617–628.
- Vonesh, E. F. and Carter, R. L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures, *Biometrics*, **48**, 1–17.
- von Rosen, D. (1991). The growth curve model: A review, *Comm. Statist. Theory Methods*, **20**, 2791–2822.
- 柳原宏和 (2001). 非正規 GMANOVA モデルにおける検定統計量のパートレット補正とその特性, 第 69 回日本統計学会講演報告集, 284–285.
- Yanagihara, H. (2001). Asymptotic expansions of the null distributions of three test statistics in a nonnormal GMANOVA model, *Hiroshima Math. J.*, **31**, 213–262.

Applicability of a GMANOVA Model to Tree Diameter Growth Analysis

Hirokazu Yanagihara and Atsushi Yoshimoto

(The Institute of Statistical Mathematics)

A generalized multivariate analysis of variance model (GMANOVA model) has been widely accepted for analyses of repeatedly measured data. Because of the frequent use of the GMANOVA model for growth data analysis, it is generally called a growth curve model. In the field of forest planning, there exist many analyses on forest growth data. However, the GMANOVA model has not yet been used for the analysis. This paper investigates the applicability of the GMANOVA model to forest growth analysis in forest planning. We conduct curve fitting, hypothesis testing and variable selection to see its applicability to the analysis. Since thinning effects on forest growth are one of the main concerns of forest managers and growth analysts, we apply the model to the analysis of thinning effects on forest stand growth.