

# 森林被覆率の非線形回帰モデリング

宮田 大毅<sup>1</sup>・西井 龍映<sup>2</sup>・田中 章司郎<sup>3</sup>

(受付 2011 年 7 月 12 日；改訂 11 月 11 日；採択 12 月 27 日)

## 要 旨

森林減少は人間の活動が主要因であるが、同時に地形的な制約も受けている。ここでは森林減少の定量的評価のため、森林被覆率を人口密度および土地の起伏量により説明する精密な回帰モデルを考察する。Tanaka and Nishii (2009, *IEEE Transactions on Geoscience and Remote Sensing*) は森林被覆率のロジット変換を目的変数、人口密度と起伏量を説明変数としたパラメトリックな非線形回帰モデルを探索した。ここでは平均構造を加法的な自然スプライン関数で表現した非線形回帰モデルを探索する。また誤差として 1 次近傍、2 次近傍での空間依存性も考慮した精密なモデル化を行う。広島県で観測した実データでの検証により、提案モデルが従来モデルよりはるかに優れていることが示された。

キーワード：ロジスティック回帰分析、自然スプライン、森林減少、人口密度、起伏量。

## 1. はじめに

オゾン層濃度、土壌水分、地表面温度、土地被覆状況など、テラバイト・オーダーの情報が日々、人工衛星・自動化測定機器から集められている。しかし、それらのデータ相互の関係について、特に社会経済的要因がどのように地表面に影響しているかについて、数理モデルを通じた研究が不足していると Lambin (1997) が指摘している。森林減少もその例外ではない。Myers (1990) は森林減少の主要因は発展途上国における換金作物栽培、牛などの放牧、過度の焼き畑、材木の採取や燃料需要の増加、先進国における空气中汚染物質や酸性雨であると述べている。つまり、これらの要因はすべて人為的な活動と関連している (Holdren and Ehrlich, 1974; Chown et al., 2003 を参照せよ)。森林減少の定量的評価の研究として、時系列モデルにおけるロジスティック曲線 (Esser, 1989)、線形回帰モデル (Grainger, 1999; Mather and Needle, 2000; Scricciu, 2007)、またロジスティック型曲線 (Tanaka and Nishii, 2005, 2009) が議論された。また森林の成長モデルとして、Candy (1997) and Nathoo (2010), Kneib et al. (2009) がある。

本研究は田中・西井(2003)、Tanaka and Nishii (2005, 2009)の研究に続くものである。観測した各メッシュの森林被覆率を、当該メッシュにおける人口密度、および地形的要因により森林被覆率を説明することを試みる。約  $1\text{km}^2$  の観測区画の座標  $s = (u, v)$  において、森林被覆率(森林面積比率)を  $F_s$  ( $0 \leq F_s \leq 1$ , 生態学的な疎林を含む)、人口密度を  $N_s$ ,  $R_s$  をそのメッシュでの起伏量、すなわち最高標高と最低標高の差とする。図 1, 図 2 を参照せよ。各観測区

<sup>1</sup>九州大学大学院 数理学府：〒819-0395 福岡市西区元岡 744

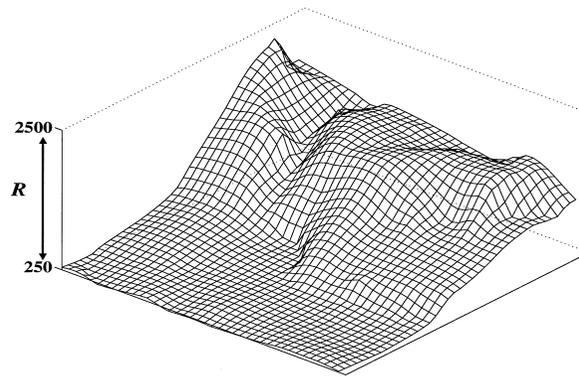
<sup>2</sup>九州大学 マス・フォア・インダストリ研究所：〒819-0395 福岡市西区元岡 744

<sup>3</sup>島根大学 総合理工学部：〒113-8654 島根県松江市西川津町 1060

$(u-1, v-1)$	$(u, v-1)$	$(u+1, v-1)$
$(u-1, v)$	<b><math>(u, v)</math></b>	$(u+1, v)$
$(u-1, v+1)$	$(u, v+1)$	$(u+1, v+1)$

- $s = (u, v)$ : 位置
- $F = F_s$ : 森林被覆率 ( $0 \leq F \leq 1$ )
- $N = N_s \geq 0$ : 人口密度
- $R = R_s \geq 0$ : 起伏量 = 最大標高 - 最小標高

図 1. 観測区画の座標と観測変数.

図 2. 起伏量  $R$ .

画に  $1, 2, \dots, n$  と通し番号をつけ, 番号  $s$  により平面上の位置を表すことにする. また全区画の集合を  $\mathcal{D} = \{1, 2, \dots, n\}$  で表す.

Tanaka and Nishii (2009) は森林被覆率  $F$  をロジスティック変換を含む巾変換後の値を目的変数とし, 人口密度と起伏量を説明変数としたパラメトリックな非線形回帰モデルを探索した. ここでは各説明変数の効果を加法的な自然スプライン関数で表現した非線形回帰モデルを探索する. また周辺からの影響を, 1 次近傍, 2 次近傍それぞれからの影響を取り入れた正規モデルを推定する. ただし 1 次近傍  $U_s^1$  は  $s$  に隣接した区画番号の集合, 2 次近傍  $U_s^2$  はななめに位置する区画番号の集合を表す. 図 3 を参照せよ.

## 2. パラメトリックな非線形回帰関数

区画  $s$  における森林被覆率  $F_s$  は 0, 1, またはその中間の値をとる. そこで全観測区画の集合  $\mathcal{D}$  を次のように 3 つに分割する.

$$(2.1) \quad \mathcal{D}_0 = \{s \in \mathcal{D} \mid F_s = 0\}, \quad \mathcal{D}_1 = \{s \in \mathcal{D} \mid F_s = 1\}, \quad \mathcal{D}_2 = \{s \in \mathcal{D} \mid 0 < F_s < 1\}.$$

$\mathcal{D}_0$  は森林が全くない領域,  $\mathcal{D}_1$  は森林に完全に覆われている領域を表す.

Tanaka and Nishii (2005) は広島県のメッシュデータ (4 節の数値解析例でデータ諸元を紹介)

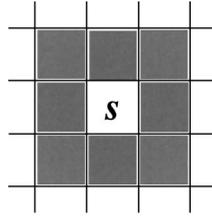


図 3. 観測区画  $s$  とその 1 次近傍  $U_s^1$  と 2 次近傍  $U_s^2$ .

表 1. パラメトリックな非線形回帰関数 (Tanaka and Nishii, 2005).

$g_{1a}(N) = -\alpha n$ ただし $n \equiv \log(N+1)$ $g_{1b}(N) = -\alpha n^\beta$ $g_{2a}(N) = -\frac{\beta_1}{1 + \exp(\beta_2 - \beta_3 n)} + \frac{\beta_1}{1 + \exp(\beta_2)}$ $g_{2b}(N) = -\frac{\beta_1}{1 + \exp(\beta_2 - \beta_3 n^{\beta_4})} + \frac{\beta_1}{1 + \exp(\beta_2)}$ $g_{3a}(N) = \gamma \{1 - \exp(\alpha n)\}$ $g_{3b}(N) = \gamma \{1 - \exp(\alpha n^\beta)\}$	$h_{s_1}(R) = I(R > \theta) \cdot \delta \log(R - \theta + 1)$ $h_{s_2}(R) = I(R > \theta) \cdot \delta \log\left(\frac{R}{\theta}\right)$ $h_1(R) = \beta_3 \exp(-\beta_1 e^{-\beta_2 R}) - \beta_3 \exp(-\beta_1)$ $h_2(R) = \frac{\beta_1}{1 + \exp(\beta_2 - \beta_3 R)} - \frac{\beta_1}{1 + \exp(\beta_2)}$ $h_3(R) = \frac{\beta_1}{1 + \exp(\beta_2 - \beta_3 R^{\beta_4})} - \frac{\beta_1}{1 + \exp(\beta_2)}$
---	--

表 2. AIC によるモデルの比較 (太字は上位 3 モデル, Tanaka and Nishii, 2005).

回帰関数	$h_{s1}$	$h_{s2}$	$h_1$	$h_2$	$h_3$
$g_{1a}$	-10184.9	-10330.0	-8868.5	-8253.4	-9426.3
$g_{1b}$ (spatial)	-10260.9	-10431.5 <b>(-10553.7)</b>	-10328.3	-10089.4	-10306.8
$g_{2a}$ (spatial)	-10288.4	-10443.8 <b>(-10557.4)</b>	-10367.4	-9944.0	-10395.9
$g_{2b}$ (spatial)	-10288.7	-10445.6 <b>(-10559.6)</b>	-10368.9	-9989.3	-10403.3
$g_{3a}$	-10245.3	-10417.5	-10304.1	-10005.5	-10236.1
$g_{3b}$	-10248.2	-10422.7	-10343.6	-10114.1	-10264.9

で  $0 < F_s < 1$  を満たす被覆率のロジット変換が, 平均が非線形関数  $\beta + g(N_s) + h(R_s)$  で与えられ, 誤差が空間依存性を持つ正規回帰モデルを適応した. 最初に  $g(N_s)$ ,  $h(R_s)$  それぞれを階段関数で推定することにより人口密度  $N_s$ , 起伏量  $R_s$  の効果の概形を知り, その階段関数に近いものとして, 表 1 の連続関数の組み合わせを評価した. 表 2 は非線形回帰関数の組み合わせを AIC で評価したものであり, (spatial) は 1 次近傍の影響を考慮したモデルである. その結果, 平均構造としては  $\beta + g_{2b}(N) + h_{s2}(R)$  が最適であり, 1 次近傍での空間従属性を持つ正規誤差を考慮すると AIC を大きく改良できることが示された. 最適モデルは同論文で, 日本全体の広域データ, Harbin, Wuhan, Tientsin のメッシュデータでも最適, あるいは準最適と

して選ばれた. ここでは森林被覆率  $F_s$  ( $s \in \mathcal{D}_2$ ) のさらなる精密なモデル化を試みる.

### 2.1 自然スプラインによる平均構造の推定

表 1 で議論したパラメトリックな回帰関数は, 推定した母数が不安定であり, また細かい変動を取り込めない欠点がある. そこで回帰関数  $g(N_s)$ ,  $h(R_s)$  を自然 3 次スプラインで近似して, 回帰関数を柔軟に推定する. Hastie et al. (2009) の 5 章を参照せよ.

一般に自然 3 次スプラインは基底関数  $b_1(x), \dots, b_K(x)$  の 1 次結合で関数  $f(x)$  を次のように近似する.

$$(2.2) \quad f(x) = \sum_{k=1}^K \gamma_k b_k(x)$$

ただし  $K$  個の基底関数は  $x$  の 3 次以下の区分多項式:

$$(2.3) \quad b_1(x) = 1, \quad b_2(x) = x, \quad b_{k+2}(x) = d_k(x) - d_{K-1}(x) \quad (k=1, 2, \dots, K-2)$$

で与えられ,  $d_k(x)$  は次で定義される.

$$(2.4) \quad d_k(x) = \frac{(x - \xi_k)_+^3 - (x - \xi_{K-1})_+^3}{\xi_{K-1} - \xi_k}$$

ただし  $(x)_+ = x$  ( $x > 0$  のとき);  $= 0$  ( $x \leq 0$  のとき) を表し,  $\xi_1 < \xi_2 < \dots < \xi_K$  は変数  $x$  の定義域を分割した節点である. 自然 3 次スプラインは 2 回微分が連続となる区分多項式であり, 両端の区間  $(-\infty, \xi_1]$ ,  $[\xi_K, \infty)$  では 1 次式となる. 基底関数を固定すると式(2.2)は線形モデルとなるため, 回帰係数  $\gamma_1, \dots, \gamma_K$  の推定は容易である.

また森林被覆率のロジット変換を巾変換に拡張したモデルも考察する. すなわち  $0 < x < 1$  を満たす値  $x$  に対して巾変換型のロジット変換 (p-logit 変換) を定義する.

$$(2.5) \quad \psi_\theta(x) = \begin{cases} \log\{x/(1-x)\} & \theta = 0 \text{ のとき,} \\ \frac{\{x/(1-x)\}^\theta - 1}{\theta} & \theta \neq 0 \text{ のとき.} \end{cases}$$

p-logit 変換は, 全ての  $\theta$  に対して  $x$  の単調増加関数となっていることに注意せよ.

## 3. 森林被覆率のモデル化とモデル評価

森林被覆率  $F_s$  ( $s \in \mathcal{D}_2$ ) に対して 1) 目的変数の変換, 2) 平均構造の再検討, 3) 誤差の空間依存性, の観点から精密なモデル化を試みる.

### 3.1 平均構造の推定の精密化

区画  $s$  における説明変数ベクトルを  $\mathbf{x}_s = (N_s, R_s)'$  とおく. さて  $s$  での森林被覆率  $F_s$  は近傍の区画における被覆率, 人口密度, 起伏量すべてに影響されていると考えられる. そのため本稿では, 森林被覆率の p-logit 変換は,  $\mathbf{x}_s$  や  $s$  の近傍の目的変数  $F_t$ , 説明変数  $\mathbf{x}_t$  が与えられたとき, 次の正規分布で与えられると仮定する.

$$(3.1) \quad \psi_\theta(F_s) | \{\mathbf{x}_s, (F_t, \mathbf{x}_t); t \in U_s^1 \cup U_s^2\} \sim N\left(\beta(s) + g(N_s) + h(R_s) + \sum_{r=1}^2 \rho_r \sum_{t \in U_s^r \cap \mathcal{D}_2} e_t, \sigma^2\right)$$

ここで定数  $\beta(s)$  は  $\beta(s) = \beta + \beta_0 c_0(s) + \beta_1 c_1(s)$  ( $c_i(s) = |\mathcal{D}_i \cap U_s^1|$ : 1 次近傍における  $\mathcal{D}_i$  の区画数,  $i=0, 1$ ) と分解され,  $s$  の近傍内で  $\mathcal{D}_0, \mathcal{D}_1$  に含まれる区画による影響を考慮した定数である. また  $g(N)$  は人口密度による影響を  $k_N$  個の自然 3 次スプライン  $g_1(N), g_2(N), \dots, g_{k_N}(N)$  の 1

次結合で、 $h(R)$  は起伏量による影響を  $k_R$  個の自然 3 次スプライン  $h_1(N), h_2(N), \dots, h_{k_R}(R)$  の 1 次結合で近似した関数を表す。  $U_s^1, U_s^2$  はそれぞれ  $s$  の 1 次近傍, 2 次近傍,  $e_t = \psi_\theta(F_t) - \beta(t) - g(N_t) - h(R_t)$  ( $t \in \mathcal{D}_2$ ) は誤差,  $\rho_r$  は  $\psi_\theta(F)$  と近傍  $U_s^r$  ( $r=1, 2$ ) での説明変数等との空間依存性を示すパラメータである。

式 (3.1) は中心区画  $s$  および近傍区画における残差  $e_s, e_t$  が空間的な相関を持つモデルであるため、2 次近傍までの説明変数も中心区画の森林被覆率に影響するモデルとなっている。なお Tanaka and Nishii (2005) でのモデル化では 1 次近傍における  $\mathcal{D}_2$  の区画だけを用いた。

### 3.2 同時分布と母数推定

森林被覆率の p-logit 変換  $\psi_\theta(F_s)$  ( $s \in \mathcal{D}_2$ ) の同時分布を導くため、 $\mathcal{D}_2$  における  $n$  個の目的変数ベクトルを  $\mathbf{z}_\theta = (\psi_\theta(F_1), \dots, \psi_\theta(F_n))^T$  とおく。このとき、森林被覆率の同時分布は  $\mathbf{z}_\theta$  を目的変数とした線形回帰モデルで表現できる。ただし  $\beta(s)$  の加法的分解には定数項  $\beta$  が含まれているため、自然 3 次スプラインの基底関数のうち  $g_1(N) = h_1(R) \equiv 1$  は不要である。その結果、計画行列  $X: n \times (k_N + k_R + 1)$  と回帰係数ベクトル  $\beta$  は次で与えられる。

$$(3.2) \quad X = \begin{pmatrix} 1 & c_0(1) & c_1(1) & g_2(N_1) & \dots & g_{k_N}(N_1) & h_2(R_1) & \dots & h_{k_R}(R_1) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & c_0(n) & c_1(n) & g_2(N_n) & \dots & g_{k_N}(N_n) & h_2(R_n) & \dots & h_{k_R}(R_n) \end{pmatrix}.$$

$$(3.3) \quad \beta = (\beta, \beta_0, \beta_1, \gamma_2^N, \dots, \gamma_{k_N}^N, \gamma_2^R, \dots, \gamma_{k_R}^R)'$$

ただし  $c_i(s) = |U_s^1 \cap \mathcal{D}_i|$  ( $i=0, 1$ ) を表す。このとき、式 (3.1) より同時分布

$$(3.4) \quad \mathbf{z}_\theta \sim N_n(X\beta, \sigma^2(I - \rho_1 B_1 - \rho_2 B_2)^{-1})$$

が導かれる。ただし  $B_1$  の  $(s, t)$  成分は区画  $t$  が  $s$  の 1 次近傍にあれば 1, その他は 0 の  $n$  次対称行列であり、 $B_2$  は 2 次近傍から同様に得られる対称行列である。

母数  $(\theta, \rho_1, \rho_2)$  が既知のとき、式 (3.4) から回帰係数ベクトルと分散の最尤推定量は

$$(3.5) \quad \hat{\beta}(\theta, \rho_1, \rho_2) = \{X'(I - C)X\}^{-1} X'(I - C)\mathbf{z}_\theta,$$

$$(3.6) \quad \hat{\sigma}^2(\theta, \rho_1, \rho_2) = \{\mathbf{z}_\theta - X\hat{\beta}(\theta, \rho_1, \rho_2)\}'(I - C)\{\mathbf{z}_\theta - X\hat{\beta}(\theta, \rho_1, \rho_2)\}/n$$

( $C \equiv \rho_1 B_1 + \rho_2 B_2$ ) と求められる。これを式 (3.4) から与えられる同時密度関数に代入した  $(\theta, \rho_1, \rho_2)$  に関するプロファイル尤度関数は次となる。

$$L(\theta, \rho_1, \rho_2) = \frac{|I - \rho_1 B_1 - \rho_2 B_2|^{\frac{1}{2}}}{\{2\pi\hat{\sigma}^2(\theta, \rho_1, \rho_2)\}^{\frac{n}{2}}} \exp\left(-\frac{n}{2}\right)$$

これを用いると  $\beta, \sigma^2$  の最尤推定量  $\hat{\beta} = \hat{\beta}(\hat{\theta}, \hat{\rho}_1, \hat{\rho}_2)$ ,  $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\theta}, \hat{\rho}_1, \hat{\rho}_2)$  が得られ、AIC は次で与えられる。

$$(3.7) \quad \text{AIC} = n \log(2\pi e) - \log(|I - \hat{\rho}_1 B_1 - \hat{\rho}_2 B_2|) + n \log \hat{\sigma}^2 \\ + 2 \sum_{s \in \mathcal{D}_2} \left[ \log\{F_s(1 - F_s)\} - \hat{\theta} \log\left(\frac{F_s}{1 - F_s}\right) \right] + 2(k_N + k_R + 1 + 4)$$

ここで (3.7) 式 2 行目の第 1 項は、確率変数  $F_s$  を巾変換したことによる変数変換のヤコビアンに由来する。またペナルティ項のうち  $k_N + k_R + 1$  は未知回帰係数ベクトル (3.3) の次元数、4 は未知母数  $\{\theta, \rho_1, \rho_2, \sigma^2\}$  の数を表す。

注意：式 (3.7) はフルモデルに対する AIC であり、 $\theta = 0$  (ロジット変換),  $\rho_2 = 0$  (1 次近傍だけの空間相関を考慮),  $\rho_1 = \rho_2 = 0$  (空間相関のない独立モデル) 等の部分モデルを考える場合は、未知母数の数を減らす必要がある。

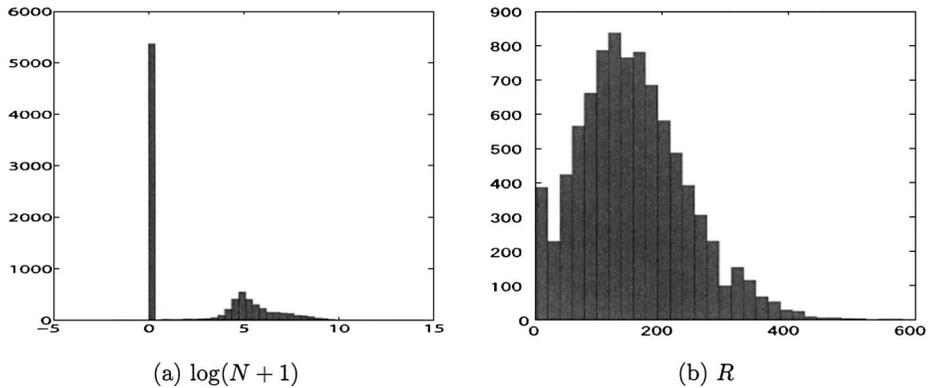


図 4. 広島県の人口密度と起伏量の頻度分布.

#### 4. 解析例：広島県

図 4 は広島県の観測領域の (a) 人口密度  $N$ , (b) 起伏量  $R$  の頻度分布である. 観測メッシュは  $1\text{km}^2$  であり, 被覆率で 3 グループに分けたときの  $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$  の標本数はそれぞれ 159, 1713, 6825 である. 地形的に見ると, 沿岸部の平坦な領域に広島市や福山市等の都市部があり, 県東北部や県西部は山間部となっている.

##### 4.1 平均構造のスプラインによる推定

$0 < F_s < 1$  を満たす ( $s \in \mathcal{D}_2$ ) 被覆率を p-logit 変換し, その平均を 2 つの説明変数の加法モデルで表現する. 回帰関数として自然 3 次スプラインを用いるが, その引数としては人口密度や起伏量を対数変換した  $\log(N_s + 1)$ ,  $\log(R_s + 1)$  を採用する.

まず人口密度に関する回帰関数  $g(N_s)$  を  $\log(N_s + 1)$  の多項式を用いて次のように近似する.  $\log(N_s + 1)$  の変域を  $[0, \xi_1], [\xi_1, \xi_2], \dots, [\xi_{k_N-1}, \xi_{k_N}], [\xi_{k_N}, \infty]$  と等分割し, 対応する自然 3 次スプラインの  $k_N$  個の基底関数  $g_i(N)$ ,  $i = 1, \dots, k_N$  により  $g(N_s)$  を近似した. 同様に  $h(R_s)$  に対して,  $\log(R_s + 1)$  の多項式  $h_i(N)$ ,  $i = 1, \dots, k_R$  で近似した.

次に節点の個数  $k_N, k_R$  を決定する. まず被覆率のロジット変換  $\mathbf{z}_0 = (\psi_0(F_1), \dots, \psi_0(F_n))'$  に対して, 等分散で独立な誤差を持つ正規線形回帰モデル:

$$\mathbf{z}_0 = X\beta + \mathbf{e}$$

を考え, 次の手順で  $k_N, k_R$  を決定した.

- (1) 3 以上 10 以下の節点数  $k_N, k_R$  に対して, 計画行列 (3.2) を用いた  $\mathbf{z}_0$  の回帰モデルを考える.
- (2) 回帰モデル (1) について, AIC に基づく自然 3 次スプラインの係数に対する変数減少法で, より小さいモデルが選ばれなかった (全ての係数が選ばれた)  $(k_N, k_R)$  を記録する.
- (3) (2) で選ばれた節点数の組み合わせのなかで, AIC 最小のものを最適な節点数とする.

この探索により  $k_N = 5$ ,  $k_R = 7$  が選ばれた. そこで今回考察する全てのモデルで, この節点数を用いた. もちろん回帰係数はモデルごとに再推定している. また  $\rho_1, \rho_2$  の最尤推定量は分散共分散行列  $I - \rho_1 B_1 - \rho_2 B_2$  が正定値行列となる条件下でグリッドサーチにより求めた.

表 3 は 1) ロジット変換と p-logit 変換, 2) 独立モデルと空間従属モデルを比較している. 表 2

表 3. 自然スプラインに基づく加法モデルによる母数推定とモデル比較 (太字は上位 3 モデル).

	空間従属性	$\theta$	$\rho_1$	$\rho_2$	$\beta$	$\beta_0$	$\beta_1$	$\sigma^2$	AIC
$\theta = 0$ ロジット 変換	独立モデル	0	0	0	0.6777	-0.5619	0.5623	1.33750	-11257.6
	1 次近傍だけ	0	0.1725	0	0.8448	-0.5607	0.5718	1.18274	-11689.2
	2 次近傍まで	0	0.1349	0.0616	0.8972	-0.5524	0.5721	1.18167	<b>-11735.7</b>
$\theta \neq 0$ p-logit 変換	独立モデル	-0.02563	0	0	0.6111	-0.5819	0.5201	1.22728	-11314.7
	1 次近傍だけ	-0.03257	0.1791	0	0.7557	-0.5918	0.5182	1.05031	<b>-11789.4</b>
	2 次近傍まで	-0.03297	0.1388	0.0634	0.8011	-0.5834	0.5175	1.04891	<b>-11839.2</b>

表 4. 自然 3 次スプラインに基づく加法モデルによる回帰係数の推定値.

	空間従属性	$\gamma_2^N$	$\gamma_3^N$	$\gamma_4^N$	$\gamma_5^N$	—	—
$\theta = 0$ ロジット 変換	独立モデル	-0.1187	-0.1779	0.3006	-0.1071	—	—
	1 次近傍だけ	-0.1505	-0.0977	0.1446	-0.0454	—	—
	2 次近傍まで	-0.1520	-0.1009	0.1561	-0.0587	—	—
		$\gamma_2^R$	$\gamma_3^R$	$\gamma_4^R$	$\gamma_5^R$	$\gamma_6^R$	$\gamma_7^R$
$\theta = 0$ ロジット 変換	独立モデル	-1.7499	6.3174	-14.6364	12.7253	-5.8127	1.4482
	1 次近傍だけ	-1.7222	5.9741	-13.6352	11.4420	-4.8645	1.1366
	2 次近傍まで	-1.7816	6.1927	-14.1451	11.8107	-4.9174	1.1146
		$\gamma_2^N$	$\gamma_3^N$	$\gamma_4^N$	$\gamma_5^N$		
$\theta \neq 0$ p-logit 変換	独立モデル	-0.1052	-0.1768	0.2928	-0.1022	—	—
	1 次近傍だけ	-0.1337	-0.0976	0.1392	-0.0434	—	—
	2 次近傍まで	-0.1340	-0.1024	0.1537	-0.0579	—	—
		$\gamma_2^R$	$\gamma_3^R$	$\gamma_4^R$	$\gamma_5^R$	$\gamma_6^R$	$\gamma_7^R$
$\theta \neq 0$ p-logit 変換	独立モデル	-1.9132	6.9151	-15.9835	13.7523	-6.1707	1.5293
	1 次近傍だけ	-1.9149	6.7078	-15.3092	12.7384	-5.3142	1.2282
	2 次近傍まで	-1.9785	6.9442	-15.8575	13.1382	-5.3839	1.2125

から 1 次近傍までの従属性を考えた森林被覆率をロジット変換した最適パラメトリックモデルの AIC は -10559.6 であった. 同じ設定で平均構造をスプラインで近似すると -11689.2 (表 3 の 3 行目) となり改善された. これは平均構造を柔軟な回帰関数に変えたことによる効果である. さらに p-logit 変換したモデルは単なるロジット変換したモデルより AIC が小さいことがわかる. また空間依存性を考慮すると, AIC は大幅に改善されている.

最適モデルでは空間相関が  $0 < \rho_2 < \rho_1$  と推定されていることや,  $\beta_0 < 0$  (森林被覆率ゼロの近傍区画の影響),  $\beta_1 > 0$  (森林で覆われた区画の影響) と推定されていることから, モデルが有効に機能していることがわかる.

表 4 はスプラインにかかる回帰係数をまとめたものである. 独立モデルと従属モデルには違いがあるが, その他のモデルでは, 係数に大きな差は見られない. 表 5 は 2 次近傍での被覆率 0 または 1 の区画からの影響を考慮したモデルの評価表であり, 1 次近傍で同様の考察を行った表 3 の AIC をわずかに改良している.

さてロジット変換かつ独立モデルの場合と, p-logit かつ 2 次近傍までを考慮した場合のスプラインによる平均構造を比較する (表 3 における最悪モデルと最良モデル).

図 5 は被覆率に対する人口密度の影響をみたものである (横軸の節点の設定位置は表 6 の 2 行目を参照). 左図は表 4 の 2 行目, 右図は表 4 の 10 行目で与えられている回帰係数を用いてグラフ化したものである. 人口密度の増加に伴い, 森林被覆率を変換した値が減少することが

表 5. 空間相関が 2 次近傍までの自然スプラインによる 6 つの推定モデルと AIC (周辺画素の平均への影響を定数にモデル化, 太字は上位 3 モデル).

$\theta = 0$ ロジット 変換		モデル番号	$\theta$	$\rho_1$	$\rho_2$	$\beta$	$\beta_0^1$
	独立モデル	1	0	0	0	0.7059	-0.5164
	1 次近傍だけ	2	0	0.1742	0	0.8658	-0.5431
	2 次近傍まで	3	0	0.1372	0.0599	0.9091	-0.5281
		モデル番号	$\beta_1^1$	$\beta_0^2$	$\beta_1^2$	$\sigma^2$	AIC
	1	0.5089	-0.1153	0.0854	1.3343	-11264.1	
	2	0.5318	-0.1162	0.1091	1.1763	-11709.0	
3	0.5304	-0.1083	0.1055	1.1762	<b>-11754.8</b>		
$\theta \neq 0$ p-logit 変換		モデル番号	$\theta$	$\rho_1$	$\rho_2$	$\beta$	$\beta_0^1$
	独立モデル	4	-0.02551	0	0	0.6404	-0.5358
	1 次近傍だけ	5	-0.03234	0.1804	0	0.7797	-0.5751
	2 次近傍まで	6	-0.03283	0.1409	0.0617	0.8165	-0.5594
		モデル番号	$\beta_1^1$	$\beta_0^2$	$\beta_1^2$	$\sigma^2$	AIC
	4	0.4710	-0.1171	0.0789	1.2249	-11322.6	
	5	0.4834	-0.1213	0.0989	1.0457	<b>-11810.2</b>	
6	0.4807	-0.1133	0.0950	1.0447	<b>-11859.1</b>		

わかる。両図の違いは、人口密度の高いときの挙動にある。

図 6 は被覆率に対する起伏量の影響をみたものである(横軸の節点の設定位置は表 6 の 3 行目を参照)。左図は表 4 の 6 行目, 右図は表 4 の最下行に対応する。両グラフは似ていて,  $R$  が大きいところでは直線的に増加している。これは Tanaka and Nishii (2005) で選ばれた  $R$  の関数形は表 1 の  $h_{s2}$  であり, 対数関数となっていることと合致している。

一方  $R$  が小さい(横軸の値が 4 以下)では図 6 は複雑な変動をしている。表 1 の  $h_{s2}$  では立ち上がりの  $\theta$  (巾指数ではない)は田中・西井(2003)により 25.84(m)と推定されている。なお図 6 の横軸の目盛り 4 は  $R = \exp(3.1824) - 1 = 23.1045(\text{m})$  であるため, 本研究は,  $R < 25$  の領域における  $R$  の森林被覆率への影響をより細かく観察したといえる。ただし,  $R \leq 23.1045$  を満たす標本数は 271 と他区間の標本数と比べて少なく, この区間での関数の変動を過度に重要視する必要はないであろう。

## 5. ディスカッション

森林被覆率が  $0 < F < 1$  の場合, 人口密度と起伏量で説明するロジット型回帰問題を次のようにモデルを精密にした。1)  $F = 0, 1$  となる周辺からの区画の影響をモデル化し, 2) スプラインによる平均構造を導入し, 3) 2 次近傍までの空間依存性を考慮し, 4) 森林被覆率の p-logit 変換を導入し, 5) 同時分布から母数推定法を確立した。以上を広島県のデータに応用して下記が判明した。

- (1) 回帰関数をパラメトリックな非線形関数から自然 3 次スプラインに変え, また  $D_0$  や  $D_1$  の近傍区画の影響をモデル化することにより, AIC が 1000 程度改善した。また自然スプラインにより, 平均構造の柔軟な変動を検出できた。
- (2) 被覆率のロジット変換から p-logit 変換に変えた変数を目的変数とすることにより, AIC が 50 - 100 程度改善した。
- (3) 空間従属性について, 独立モデルから 1 次近傍の従属モデルに変えることにより, AIC が 400 程度改善した。さらに 2 次近傍までの従属性を考えると, AIC が 50 程度改善

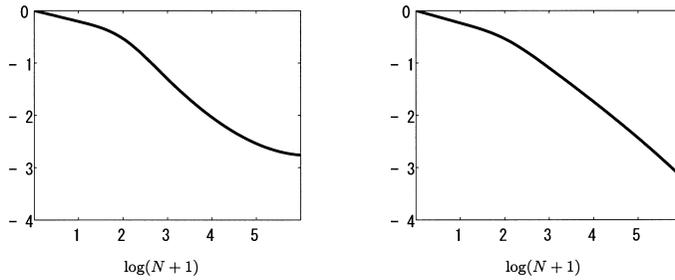


図 5. 変換した森林被覆率に対する人口密度  $N$  の回帰関数 (左図：ロジット変換 ( $\theta = 0$ ), 独立モデル, 右図：p-logit 変換 ( $\theta = -0.03297$ ), 2 次近傍まで).

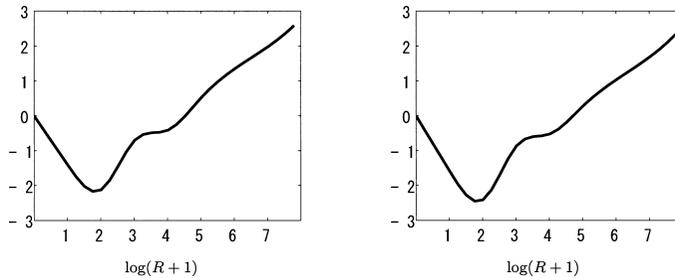


図 6. 変換した森林被覆率に対する起伏量  $R$  の回帰関数 (左図：ロジット変換 ( $\theta = 0$ ), 独立モデル, 右図：p-logit 変換 ( $\theta = -0.03297$ ), 2 次近傍まで).

表 6. 図 5, 図 6 における横軸の節点の設定位置.

節点の番号	1	2	3	4	5	6	7
$\log(N + 1)$	1.6670	3.3341	5.0011	6.6681	8.3352	—	—
$\log(R + 1)$	0.7956	1.5912	2.3868	3.1824	3.9780	4.07736	5.5692

した.

母数は  $0 < \rho_2 < \rho_1$ ,  $\beta_0 < 0$ ,  $\beta_1 > 0$  と期待通りの関係を満たす値として推定され, モデルが有効に機能していることがわかる.

本稿のモデリング手法は, 空間従属性を持つ目的変数, 従属変数の関係を精密に探索する新たな可能性を提示したといえる. 平均構造を自然スプラインで近似する場合, もとより区間数が重要である. また生の変数の多項式を用いるのか, あるいは単調変換した変数の多項式を用いるのかも重要なポイントである. 区間幅の取り方にも注意が必要である.

次の課題として, 近傍の説明変数も平均モデルに取り込んだ平均構造のモデル化が考えられる. また同時に他地域への応用と先行研究との比較が必要である.

## 謝 辞

2 人の査読者には拙論を丁寧に読んでいただき, 本論文を改訂する上で建設的なコメントを数多く頂戴しました. また編集者には誤りの指摘や編集上のご配慮をいただきました. あわせ

てお礼申し上げます。本研究は、科学研究費基盤研究(A)(19200020)“時空間現象データに対する統計科学モデルの構築及び解析に関する組織的研究”(研究代表者：矢島美寛(東京大学)), 及び基盤研究(B)(19300096)“超高次元データの分類手法の導出とその理論的性質の解明, および実データへの応用の研究”(研究代表者：西井龍映(九州大学))の援助をうけました。

## 参 考 文 献

- Candy, S. (1997). Estimation in forest yield models using composite link functions with random effects, *Biometrics*, **53**(1), 146–160.
- Chown, S. L., van Rensburg, B. J., Gaston, K. J., Rodrigues, A. S. L. and van Jaarsveld, A. S. (2003). Energy, species richness, and human population size: Conservation implications at a national scale, *Ecological Applications*, **13**(5), 1233–1241.
- Esser, G. (1989). Global land-use changes from 1860 to 1980 and future projections to 2500, *Ecological Modelling*, **44**, 307–316.
- Grainger, A. (1999). Constraints on modelling the deforestation and degradation of tropical open woodlands, *Global Ecology and Biogeography*, **8**, 179–190.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*, 2nd ed., Springer-Verlag, New York.
- Holdren, J. P. and Ehrlich, P. R. (1974). Human population and the global environment, *American Scientist*, **62**, 282–292.
- Kneib, T., Hothorn, T. and Tutz, G. (2009). Variable selection and model choice in geoaddditive regression models, *Biometrics*, **65**(2), 626–634.
- Lambin, F. E. (1997). Modelling and monitoring land-cover change processes in tropical regions, *Progress in Physical Geography*, **21**(3), 375–393.
- Mather, A. S. and Needle, C. L. (2000). The relationships of population and forest trends, *The Geographical Journal*, **166**(1), 2–13.
- Myers, N. (1990). The world's forests and human populations: The environmental interconnections, *Population and Development Review*, **16**(supplement), 1–15.
- Nathoo, F. S. (2010). Space-time regression modeling of tree growth using the skew-t distribution, *Environmetrics*, **21**(7–8), 817–833.
- Scrieciu, S. (2007). Can economic causes of tropical deforestation be identified at a global level?, *Ecological Economics*, **62**, 603–612.
- 田中章司郎, 西井龍映(2003). 人口増加に伴う森林減少の空間モデル, 応用統計学, **32**(1), 1–5.
- Tanaka, S. and Nishii, R. (2005). Verification of deforestation in East Asia by spatial logit models due to population and relief energy, *Proceedings of 12th SPIE Europe International Symposium on Remote Sensing* **5976**, No. 59760W-1, 1–10.
- Tanaka, S. and Nishii, R. (2009). Non-linear regression models to identify functional forms of deforestation in East Asia, *IEEE Transactions on Geoscience and Remote Sensing*, **47**(8), 2617–2626.

## Nonlinear Regression Modeling of Forest-coverage Ratios

Daiki Miyata<sup>1</sup>, Ryuei Nishii<sup>2</sup> and Shojiro Tanaka<sup>3</sup>

<sup>1</sup>Graduate School of Mathematics, Kyushu University

<sup>2</sup>Institute of Mathematics for Industry, Kyushu University

<sup>3</sup>Interdisciplinary Faculty of Science and Engineering, Shimane University

The worldwide decrease of forest area is chiefly due to the negative impact of human activities, but this is restricted to topographical circumstances. In this research, detailed regression models are proposed to elucidate forest area ratio observed with a grid-cell system by two covariates: population density and relief energy, which is the difference between the maximum and minimum altitudes at each site.

Tanaka and Nishii (2009, *IEEE Transaction on Geoscience and Remote Sensing*) explored a regression model of logit-transformed forest area ratios, where the mean function is given by the sum of two non-linear parametric functions of the covariates. Based on their results, we consider the mean structure represented by the sum of two natural spline functions of the respective covariates. The spatial dependency from the first- and second-order neighborhoods is also considered. The proposed models applied to real data in Hiroshima prefecture demonstrate outstanding superiority to the previous models in terms of AIC.