

空間効果を考慮した組成データ解析

吉田 崇紘¹・堤 盛人²

¹学生非会員 筑波大学大学院システム情報工学研究科 (〒305-8573 茨城県つくば市天王台一丁目1-1)

E-mail: yoshida.takahiro@sk.tsukuba.ac.jp

²正会員 筑波大学システム情報系 (〒305-8573 茨城県つくば市天王台一丁目1-1)

E-mail: tsutsumi@sk.tsukuba.ac.jp

例えば交通機関分担率データや岩石内の化学組成データのような複数の属性の割合をもつデータは組成データと呼ばれる。組成データの分析に際しては、変数の総和が一定であるという定数制約を考慮する必要があり、地質学分野を中心に組成データのための統計解析手法が発達してきた。ところで近年、空間的な従属性や異質性といった空間効果を考慮した組成データのための手法も議論されており、その有用性は多くの分野で確認されている。しかしながら、その社会経済データへの適用は限定的である。そこで本研究では、空間効果を考慮する組成データの解析手法について議論し、実証研究として、社会経済データを用いた土地利用の推定を行うことでその適用可能性を検証した。

Key Words : *compositional data, unit-sum constraint, log-ratio transformation, spatial effects, multivariate conditional autoregressive model*

1. はじめに

データのサンプルがそれぞれ多次元の非負の属性をもち、各属性の値の総和が全サンプルを通じて一定という定数制約 (unit-sum constraint) をもったデータを、組成データ (compositional data) という^{1,2)}。いま、 n サンプル、 D 次元で与えられた $n \times D$ の組成データ行列 \mathbf{Y} は、式 (1) のように、行和が常に定数 \bar{a} で、各要素 y_{ik} が非負のデータと表すことができる。

$$\sum_{k=1}^D y_{ik} = \bar{a}, \quad y_{ik} > 0 \quad (K=1, \dots, D, \quad i=1, \dots, n), \quad (1)$$

組成データの例としては、岩石内の化学組成、市町村の年代別人口比率、交通機関の分担率が挙げられる。なお、compositional data の訳語には比率データ、構成比データなどが挙げられるが、本研究では、異なる訳語を用いることで生じうる混乱をなるべく避けるため、これまでに compositional data を扱った研究の蓄積が多い地質学 (geology) の記述に倣い、「組成データ」を用いることとする。

組成データは定数制約に起因して変数の独立性が保たれないという問題がある³⁾。この問題は、たとえば 3 次元の組成データの場合、2 次元の属性の値が得ることができれば、残り 1 次元の属性の値が定数制約から一義的に決定されてしまうため、変数の独立が前提にある

統計解析手法を組成データに直接適用することを不適切にする。この問題の対処として、Aitchison (1986)²⁾ 以降、組成データの統計学的な性質の議論、扱い方についての基礎的研究が蓄積されてきた³⁾。

しかし、これまでのところ、組成データの統計モデル (CDM, compositional data model) についての研究とその実証研究は、組成データの解析手法が地質学分野の方法論として発展してきた経緯から、鉱物資源の組成調査などの分析に偏っており、社会経済データを対象に分析を行った研究は、筆者の知る限りほとんど蓄積されていない。

ところで、近年、近しい事物は似た属性をもつ傾向があるという正の空間的自己相関 (spatial autocorrelation) に代表される空間効果 (spatial effects) を考慮する必要性が空間統計学 (spatial statistics) の発展に伴って意識されるようになってきた⁴⁾。位置情報をもつデータ (空間データ: spatial data) を分析する際は、データの空間的な情報を利用する、すなわち空間効果を考慮することで、統計モデルの推定・予測精度を向上させ、信頼性を高めることができる。

組成データの分析に対する空間効果の考慮は、地質学に関連が深い鉱山学に端を発し、また、空間上の連続量として扱う地球統計学 (geostatistics) の手法を用いた例⁵⁾が多いが、空間データを、ある集計単位で区切られたという意味で、離散量として扱う空間計量経済学 (spatial

econometrics) の手法を用いた研究は、筆者の知る限り、数例に限られている^{6),7),8)}。

これらの現状を踏まえ本研究は、CDM がこれまで十分に対象にしていなかった社会経済データを用いた分析を行う。CDM の社会経済データへの適用可能性の観点から、研究の蓄積を図ることの意義は、決して小さくないと考えられる。また、空間計量経済学的手法により空間効果を考慮するCDM も社会経済データに適用し、空間効果の考慮有無の比較を行うことでその有用性を示す。

2. 組成データモデル (CDM)

本章では、組成データの対数比変換とMardia (1988)⁹⁾が提案した multivariate conditionally autoregressive (MCAR) モデルを簡単に紹介する。

(1) 対数比変換

組成データは、定数和制約によって標本空間が実空間に内包される単体空間上に属するため、実空間で定義された統計解析手法を直接適用することができない。組成データに実空間で定義された統計解析手法を用いるには、組成データを単体空間から実空間に写像する操作である、対数比変換が用いられることが多い²⁾。

対数比変換は、組成データの各属性にある共通の規格化変数で除し、各属性の値を比に置き換え、その比に対数をとる変換操作のことである。ここで、比をとることは、定数和制約を解くため、また、組成データ間の比が対数正規分布によく近似することを利用するための操作である^{2),5)}。そして、対数をとることは、変換後の組成データを解析のしやすい正規分布に近似するための操作である^{2),3),5)}。

本研究では、代表的な対数比変換法の1つである、alr (additive log-ratio) 変換を用いる。alr 変換は、 \mathbf{Y} のある1つの次元を規格化変数として利用する対数比変換法である (式(2))。

$$Z_{ik} = \ln(y_{ik}/y_{id}) \quad (k=1, 2, \dots, d), \quad (2)$$

ここで、 $d = D - 1$ 、 Z_k は $n \times d$ の潜在変数行列 \mathbf{Z} の i 行 k 列の要素、 y_{ik} は \mathbf{Y} の i 行 k 列の要素、 y_{id} は規格化変数

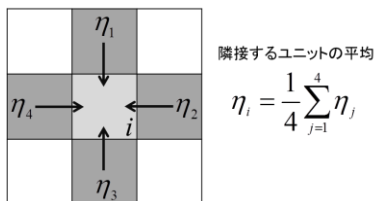


図-1 MCAR モデルの空間効果の考慮する方法のイメージ

で \mathbf{Y} の i 行 D 列の要素である。 \mathbf{Y} の予測値は、 \mathbf{Z} の i 行である Z_i を被説明変数とする統計モデルから予測値を得た後、式(2)の逆変換 (式(3),(4)) を行うことで得られる。

$$y_{ik} = \frac{\exp(Z_{ik})}{1 + \sum_{k=1}^d \exp(Z_{ik})} \quad (k=1, \dots, d), \quad (3)$$

$$y_{id} = 1 - \sum_{k=1}^d y_{ik} = \frac{1}{1 + \sum_{k=1}^d \exp(Z_{ik})}. \quad (4)$$

(2) MCAR モデル (Mardia, 1998)⁹⁾

本研究では、Billheimer *et al.* (1997)⁹⁾, Allen *et al.* (2013)⁷⁾ に倣い、ランダム効果で空間効果を簡便に考慮できる、Mardia (1998)⁹⁾ が提案したMCARモデルを用いる。

いま、 \mathbf{Z}_i を d 次元の属性をもつ被説明変数とおくとき、 \mathbf{Z}_i が d 次元正規分布 N_d に従う固定効果とランダム効果の混合モデルであると仮定する (式(5))。

$$\mathbf{Z}_i \sim N_d(\mathbf{B}^T \mathbf{x}_i + \boldsymbol{\eta}_i, \mathbf{V}) \quad (5)$$

ここで、 \mathbf{B} は $(p+1) \times d$ のパラメータ行列、 \mathbf{x}_i はサンプル i における $(p+1)$ 個の説明変数で構成される $(p+1) \times 1$ ベクトル、 $\boldsymbol{\eta}_i$ はサンプル i のランダム効果で $d \times 1$ ベクトル、 \mathbf{V} は $d \times d$ の分散共分散行列、 T は転置を表す記号である。また、空間効果の考慮のため、 $\boldsymbol{\eta}_i$ に自己回帰型の事前分布 (式(6)) を仮定する。

$$\boldsymbol{\eta}_i | \{\boldsymbol{\eta}_j\}_{j \neq i} \sim N_d\left(\frac{1}{S_i} \sum_{j=1}^n w_{ij} \boldsymbol{\eta}_j, \frac{1}{S_i} \boldsymbol{\Sigma}\right), \quad (6)$$

ここで、 S_i は外生的に与える $n \times n$ の隣接行列 \mathbf{W} の i 行の行和、 w_{ij} は \mathbf{W} の i 行 j 列の要素、 $\boldsymbol{\Sigma}$ は $d \times d$ の分散共分散行列である。MCAR モデルは、図-1 のように、隣接するユニットのランダム効果が与えられた下で、それらの平均値をランダム効果に与えることで空間効果を考慮する。

パラメータの推定には、ベイズ推定が用いられる。いま、ベイズ推定にあたり、それぞれ独立と仮定するパラメータ $\mathbf{B}, \mathbf{V}, \boldsymbol{\Sigma}$ の事前分布を式(7)~(9)のように設定する。

$$\text{vec}(\mathbf{B}) \sim N_{d(p+1)}(\text{vec}[\boldsymbol{\mu} \quad \mathbf{0}_{d \times p}], \lambda \mathbf{1}_{d \times (p+1)}), \quad (7)$$

$$\mathbf{V} \sim IW_d(m_V, \mathbf{M}_V), \quad (8)$$

$$\boldsymbol{\Sigma} \sim IW_d(m_\Sigma, \mathbf{M}_\Sigma), \quad (9)$$

ここで、 IW_d は d 次元逆ウィシャート分布、 $\text{vec}(\bullet)$ は行列 \bullet を列ベクトルに変換する演算子、 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)^T$ は $\mu_k = \mathbf{x}_i^T \boldsymbol{\beta}_k$ を要素にもつ $d \times 1$ ベクトル、 $\boldsymbol{\beta}_k$ は \mathbf{B} の k 列の $(p+1) \times 1$ ベクトル、 $\mathbf{1}_{d \times (p+1)}$ は 1 を全要素にもつ $d \times (p+1)$ ベクトル、 $\lambda, m_V, \mathbf{M}_V, m_\Sigma, \mathbf{M}_\Sigma$ は後述するハイパーパラメータ

である。ベイズの定理より、式 (5), (6) の尤度関数と式 (7) ~ (9) の事前分布から、式 (10) ~ (13) の条件付事後分布が導出できる。

$$\text{vec}(\mathbf{B}) | \mathbf{Z}_i, \mathbf{X}, \mathbf{V}, \boldsymbol{\eta}_i$$

$$\sim N_i \left(\boldsymbol{\Omega} (\mathbf{V}^{-1} \otimes \mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \sum_{i=1}^n \mathbf{x}_i^T (\mathbf{Z}_i - \boldsymbol{\eta}_i), \boldsymbol{\Omega} \right), \quad (10)$$

$$\mathbf{V} | \mathbf{Z}_i, \mathbf{X}, \mathbf{B}, \boldsymbol{\eta}_i \sim IW_d \left(m_v + S_i, \mathbf{M}_v + \sum_{i=1}^n \mathbf{E}_i \mathbf{E}_i^T \right), \quad (11)$$

$$\boldsymbol{\eta}_i | \mathbf{Z}_i, \mathbf{X}, \mathbf{B}, \{\boldsymbol{\eta}_j\}_{j \neq i}, \boldsymbol{\Sigma}$$

$$\sim N_d \left(\mathbf{A}^{-1} \left(\mathbf{V}^{-1} (\mathbf{Z}_i - \mathbf{B}^T \mathbf{x}_i) + \boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^n w_{ij} \boldsymbol{\eta}_j \right) \right), \mathbf{A}^{-1} \right), \quad (12)$$

$$\boldsymbol{\Sigma} | \boldsymbol{\eta}_i \sim IW_d \left(m_\Sigma + S_i, \mathbf{M}_\Sigma + \sum_{i=1}^n \sum_{j=1}^n (\mathbf{D}_w - \mathbf{W})_{ij} \boldsymbol{\eta}_i \boldsymbol{\eta}_j^T \right), \quad (13)$$

ここで、 \mathbf{X} は $n \times (p+1)$ の説明変数行列、 $\boldsymbol{\Omega} = (\boldsymbol{\Lambda}_{d \times (p+1)} + \mathbf{V}^{-1} \otimes \mathbf{X}^T \mathbf{X})^{-1}$ 、 $\mathbf{E}_i = \mathbf{Z}_i - \mathbf{B}^T \mathbf{x}_i - \boldsymbol{\eta}_i$ 、 $\mathbf{A} = \mathbf{V}^{-1} + \boldsymbol{\Sigma}^{-1}$ 、 \mathbf{D}_w を対角項が S_i でそれ以外の要素が 0 の $n \times n$ 行列、 \mathbf{W} を $n \times n$ 隣接行列とおき、 \otimes はクロネッカー積である。各パラメータの条件付き分布が (1 次元) 正規分布、多次元正規分布、逆ウィシャート分布という標準的な分布になるため、ギブス・サンプラーを用いた MCMC 法によるパラメータ推定が可能になる。

3. 実証分析

本章では、前章で説明したモデルを社会経済データに適用し、組成データとみなすことができる集計された土地利用データの推定を行う。

(1) 用いるデータの概要

a) 被説明変数

本研究で用いる被説明変数は、国土交通省国土政策局の国土数値情報ダウンロードサービスより無償で入手可能な、茨城県における 2006 年の土地利用 3 次メッシュデータである。このデータは、人工衛星 (TERRA (Aster), ALOS) リモートセンシングによる画像データと数値地図図を用いて、土地分類基準表に従い作成されたもので、各メッシュには、土地利用種それぞれの占有面積が格納されている。

本研究では、式 (2) の対数比変換を行うことから 0 値の属性は許容されないため、Allen et al. (2013)⁷⁾ に倣い、各土地利用種に微小値 (0.01) を加えた上で、メッシュ内に占める各土地利用種の割合を算出し、定数和が 1 の組成データとした。しかし、この 0 値の対処法は、真の 0 値 (essential zero) と丸め誤差に起因する 0 値 (rounded

表-1 土地利用種の統合

国土数値情報	本研究
田	田
その他の農用地	畑
森林	森林
荒地	
ゴルフ場	
建物用地	
幹線交通用地	都市域
その他の用地	
河川地及び湖沼	水域

表-2 説明変数

変数名	内容
lnPOP	人口密度 (人/km ²) の自然対数値
Avg_Elv	平均標高 (m)
Avg_Slope	平均傾斜 (度)
TRL	道路総延長 (km)
Dist_Sta	最寄鉄道駅までの距離 (km)
Dist_River	最寄一級河川までの距離 (km)
D_AF	扇状地 (該当: 1, 該当しない: 0)
D_NL	自然堤防 (該当: 1, 該当しない: 0)
D_BM	後背湿地 (該当: 1, 該当しない: 0)
D_Delta	三角州 (該当: 1, 該当しない: 0)
D_SD	砂州・砂礫洲 (該当: 1, 該当しない: 0)
D_Lake	湖沼内 (該当: 1, 該当しない: 0)

zero) を区別していない、また、対処前後で属性間の比率が異なってしまうという問題がある。0 値の処理には、新井・太田 (2006)¹⁰⁾ が述べるように統計学的に適切な処理を施す必要があるが、この点は今後の課題とした。

また、本研究では、モデルで採用する説明変数による土地利用種の判別を容易にするため、また、推定パラメータの増加による計算時間の増加を避けるため、土地利用種を 5 種に統合している (表-1)。

b) 説明変数

モデルに採用する説明変数は、表-2 に示すとおりである。以下、各説明変数の出典を記述する。lnPOP は、総務省統計局の統計GISより 2005 年国勢調査の人口総数を入手し、各メッシュの面積で除し、データの特性から、自然対数をとった変数である。Avg_Elv, Avg_Slope, TRL, Dist_Sta, Dist_River は、被説明変数と同様、国土数値情報ダウンロードサービスより入手可能な、それぞれ標高・傾斜度 3 次メッシュデータ (2011 年度)、同、道路密度・道路延長メッシュデータ (2010 年度)、鉄道時系列データ (2006 年度)、河川データ (2008 年度) を出典とした変数である。なお、括弧内は変数に用いたデータの年度である。ダミー変数 (D_**) は、若松ら (2005)¹¹⁾ の微地形分類データを用いて作成した変数である。ここ

で、距離変数 (Dist_**) とダミー変数 (D_**) は、メッシュの重心点を基準とした値を与えている。また、ダミー変数 (D_**) 以外の変数は、丸め誤差を小さくするため、標準化を行った上でパラメータの推定を行っている。

(2) 分析の概要

本研究では、前章で説明した、ランダム効果で空間効果を考慮するMCAR モデルをspatialモデル、ランダム効果の項をゼロとするモデルをnonspatialモデルとおき、空間効果の考慮有無の比較を行う。

規格化変数には、土地利用種の中で組成の変動が最も少ない水域の組成を採用した。ハイパーパラメータの初期値は、 $\lambda = 1,000$, $m_V = m_\Sigma = (d + 2)$, $M_V = M_\Sigma = 2I_d$ を与え、ギブス・サンプリングにおける乱数発生回数は 11,000 回、この内 1,000 回を burn-in 期間としている。

(3) 分析結果と考察

まず、表-3, 4 にそれぞれ spatial モデル、nonspatial モデルの説明変数のパラメータの推定結果と Geweke (1992)¹² の方法を用いた収束診断の結果を示す。ベイズ推定では、95% 信用区間 (下限: Lower, 上限: Upper) に 0 が含まれているかという観点からパラメータの有意性検定を行う。いくつかの変数 (例えば森林の Dist_Sta) は、nonspatial モデルでは有意であるが、空間効果の考慮によって、spatial モデルでは有意でないと判定されている。また、パラメータの収束判定には、Geweke (1992)¹² の方法を用い、表中の Geweke's Z の値の絶対値が 1.96 以下であれば収束を示唆する。spatial モデルにおける森林の Avg_Elv, nonspatial モデルの畑の D_Delta, 森林の lnPOP を除き、パラメータは収束している。

次に、2 つのモデルの予測精度を評価するため、組成データの各サンプル i の実測値と予測値の単体空間上の近さを表すアイチソン距離 AD (式 (14)) を算出し、その基本統計量を表-5 に、その分布を図-3 に示す。 AD はその値が 0 のとき、実測値と観測値が等しく、値が大きくなればなるほど、予測精度が悪いことを表す。表-5 より、nonspatial モデルと spatial モデルを比べると、平均値 (Mean) が 2.00 から 1.31 へと小さくなっていることから、空間を考慮したことで予測精度が向上したことが示唆される。さらに最大値 (Max.) は 6.25 から 3.43 と大幅に小さくなっており、少なくとも 2.8 ポイント以上 AD が近くなったサンプルが存在することがわかる。図-2からは、特に、nonspatial モデルの南東部の予測精度が悪いことが、このような系統的な予測精度不良は、空間効果 (空間的自己相関) を考慮することで改善できたことが見てとれる。

表-5 アイチソン距離の基本統計量

	Min.	Mean	Max.	SD
spatial モデル	0.00	1.31	3.43	0.45
nonspatial モデル	0.00	2.00	6.25	0.83

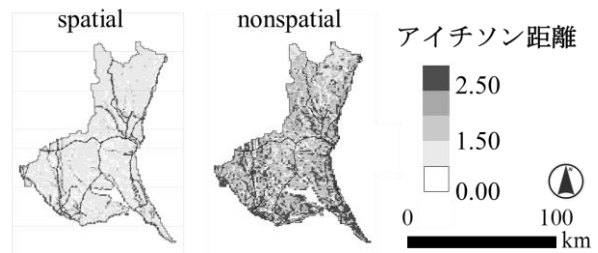


図-2 アイチソン距離の分布

$$AD = \sqrt{\sum_{k=1}^D \left[\ln \left\{ \frac{y_{ik}}{\left(\prod_{k=1}^D y_{ik} \right)^{1/D}} \right\} - \ln \left\{ \frac{\hat{y}_{ik}}{\left(\prod_{k=1}^D \hat{y}_{ik} \right)^{1/D}} \right\} \right]^2} \quad (14)$$

次に、各土地利用種の割合の実測値、spatial モデルの予測値、nonspatial モデルの予測値の分布をそれぞれ図-3 [a], [b], [c] に示す。式 (6) のように、空間効果を隣接する周辺からの平均としたことで空間ランダム効果の値が平準化され、nonspatial モデルと比べると spatial モデルの予測値は、どの土地利用種においても滑らかな分布となっていることがわかる。

4. おわりに

本研究では、組成データの統計モデルの適用がこれまで希薄であった社会経済データを用いた分析を行い、実証研究の蓄積を図った。また、MCAR モデルのランダム効果によって空間効果を考慮することで、ランダム効果をゼロとした空間効果を考慮しないモデルに比べ、推定・予測の精度が向上することを確認した。

今後の課題として、0 値を許容する組成データのモデリングを検討する必要がある。本研究では、組成データを実空間に写像する変数変換法に対数比変換を採用したため、全属性に 0 値が許容されず、ad hoc に微小値 (0.01) を加えることで対処した。しかし、この処理は推定・予測結果に歪みを生むと考えられる。Leiniger et al. (2013)⁸は、変数変換法に対数ではなく平方根を用いることで規格化変数以外の 0 値を許容するモデルを提案しており、本研究で用いた微小値を加える方法や、Martin-Fernandez et al. (2003)¹³で提案されている対処法との比較が望まれる。

表-3 spatial モデルにおける説明変数のパラメータの推定結果と Geweke (1992)¹²⁾ の収束診断結果

	田				畑				森林				都市域			
	Mean	Lower	Upper	Geweke's Z	Mean	Lower	Upper	Geweke's Z	Mean	Lower	Upper	Geweke's Z	Mean	Lower	Upper	Geweke's Z
(Intercept)	1.606	1.580	1.631	0.616	1.775	1.750	1.801	1.596	2.515	2.489	2.540	1.281	1.691	1.669	1.713	0.619
lnPOP	0.488	0.427	0.549	0.477	0.763	0.706	0.819	-0.857	0.049	-0.012	0.111	-0.758	0.868	0.819	0.918	-0.684
Avg_Elv	-0.234	-0.324	-0.136	0.545	0.187	0.095	0.281	0.810	0.532	0.439	0.629	2.191	0.400	0.315	0.486	1.549
Avg_Slope	-0.290	-0.357	-0.222	-0.851	-0.272	-0.338	-0.207	-0.404	0.459	0.391	0.529	-0.285	-0.153	-0.210	-0.095	-0.740
TRL	-0.171	-0.218	-0.123	-0.600	-0.192	-0.239	-0.144	-0.117	-0.169	-0.218	-0.121	-0.013	0.255	0.215	0.295	-0.305
Dist_Sta	0.215	0.151	0.283	0.854	0.158	0.097	0.225	0.405	0.062	-0.001	0.133	-1.601	-0.106	-0.162	-0.053	-0.041
Dist_River	0.045	-0.034	0.122	0.478	0.070	-0.004	0.153	1.254	0.281	0.207	0.383	1.637	0.311	0.245	0.385	1.550
D_AF	-3.531	-4.946	-2.131	0.577	-2.892	-4.167	-1.525	-0.080	-3.190	-4.643	-1.765	0.219	-2.867	-4.048	-1.713	0.042
D_NL	-0.909	-1.113	-0.701	-1.597	-1.819	-2.020	-1.616	-2.215	-2.674	-2.886	-2.459	-1.569	-1.524	-1.692	-1.353	-0.693
D_BM	0.502	0.327	0.671	0.378	-1.454	-1.630	-1.295	0.321	-1.766	-1.941	-1.590	0.040	-0.853	-0.995	-0.719	0.534
D_Delta	0.199	0.027	0.375	-0.773	-1.646	-1.799	-1.485	-0.986	-1.475	-1.645	-1.302	-0.874	-0.932	-1.067	-0.797	-0.141
D_SD	-0.186	-0.538	0.186	-0.491	-0.333	-0.671	0.013	-1.649	-0.859	-1.230	-0.504	-1.212	0.597	0.309	0.888	-0.865
D_Lake	-4.051	-4.291	-3.810	0.406	-3.852	-4.093	-3.604	-0.114	-4.608	-4.853	-4.360	0.391	-2.988	-3.186	-2.788	0.102

表-4 nonspatial モデルにおける説明変数のパラメータの推定結果と Geweke (1992)¹²⁾ の収束診断結果

	田				畑				森林				都市域			
	Mean	Lower	Upper	Geweke's Z	Mean	Lower	Upper	Geweke's Z	Mean	Lower	Upper	Geweke's Z	Mean	Lower	Upper	Geweke's Z
(Intercept)	1.624	1.612	1.636	-1.322	1.932	1.920	1.944	0.404	2.635	2.623	2.646	-1.315	1.738	1.726	1.750	-0.492
lnPOP	0.566	0.548	0.584	-1.356	0.788	0.770	0.806	0.749	0.021	0.003	0.040	1.982	0.844	0.825	0.862	0.177
Avg_Elv	-0.157	-0.177	-0.138	-0.280	-0.098	-0.118	-0.079	-0.066	0.379	0.360	0.398	1.824	0.137	0.118	0.156	0.811
Avg_Slope	-0.296	-0.313	-0.280	-0.155	-0.306	-0.322	-0.289	0.220	0.542	0.525	0.558	0.219	-0.191	-0.208	-0.175	-0.087
TRL	-0.280	-0.294	-0.265	-0.201	-0.293	-0.308	-0.279	-1.263	-0.250	-0.265	-0.236	-1.136	0.352	0.337	0.366	-0.738
Dist_Sta	0.113	0.100	0.126	-0.251	0.137	0.123	0.150	-1.187	-0.021	-0.034	-0.008	-1.600	-0.013	-0.026	0.000	-0.903
Dist_River	0.128	0.116	0.140	0.095	-0.025	-0.038	-0.013	-1.027	0.141	0.129	0.154	-1.912	0.171	0.159	0.184	-0.000
D_AF	-3.085	-3.486	-2.683	-0.953	-3.048	-3.449	-2.649	1.177	-4.352	-4.740	-3.953	-1.183	-2.458	-2.855	-2.063	-0.517
D_NL	-0.847	-0.900	-0.794	-0.798	-2.394	-2.448	-2.341	-0.632	-3.292	-3.345	-3.239	0.522	-1.583	-1.635	-1.530	0.267
D_BM	0.723	0.678	0.769	-0.897	-1.899	-1.945	-1.854	0.329	-2.269	-2.314	-2.224	1.517	-0.765	-0.810	-0.720	-0.969
D_Delta	0.358	0.313	0.404	0.158	-2.452	-2.498	-2.405	-2.881	-1.905	-1.951	-1.858	0.836	-1.169	-1.214	-1.123	-0.424
D_SD	-0.911	-0.988	-0.834	-1.751	-1.437	-1.514	-1.360	-1.268	-1.167	-1.243	-1.089	0.464	0.029	-0.049	0.106	-0.709
D_Lake	-4.879	-4.944	-4.814	-1.595	-4.728	-4.792	-4.662	0.531	-5.363	-5.426	-5.297	1.408	-3.636	-3.700	-3.573	1.313

95%信用区間に0を含まないパラメータ、Geweke's Zの絶対値が1.96以下のパラメータを *bold italic* で表す

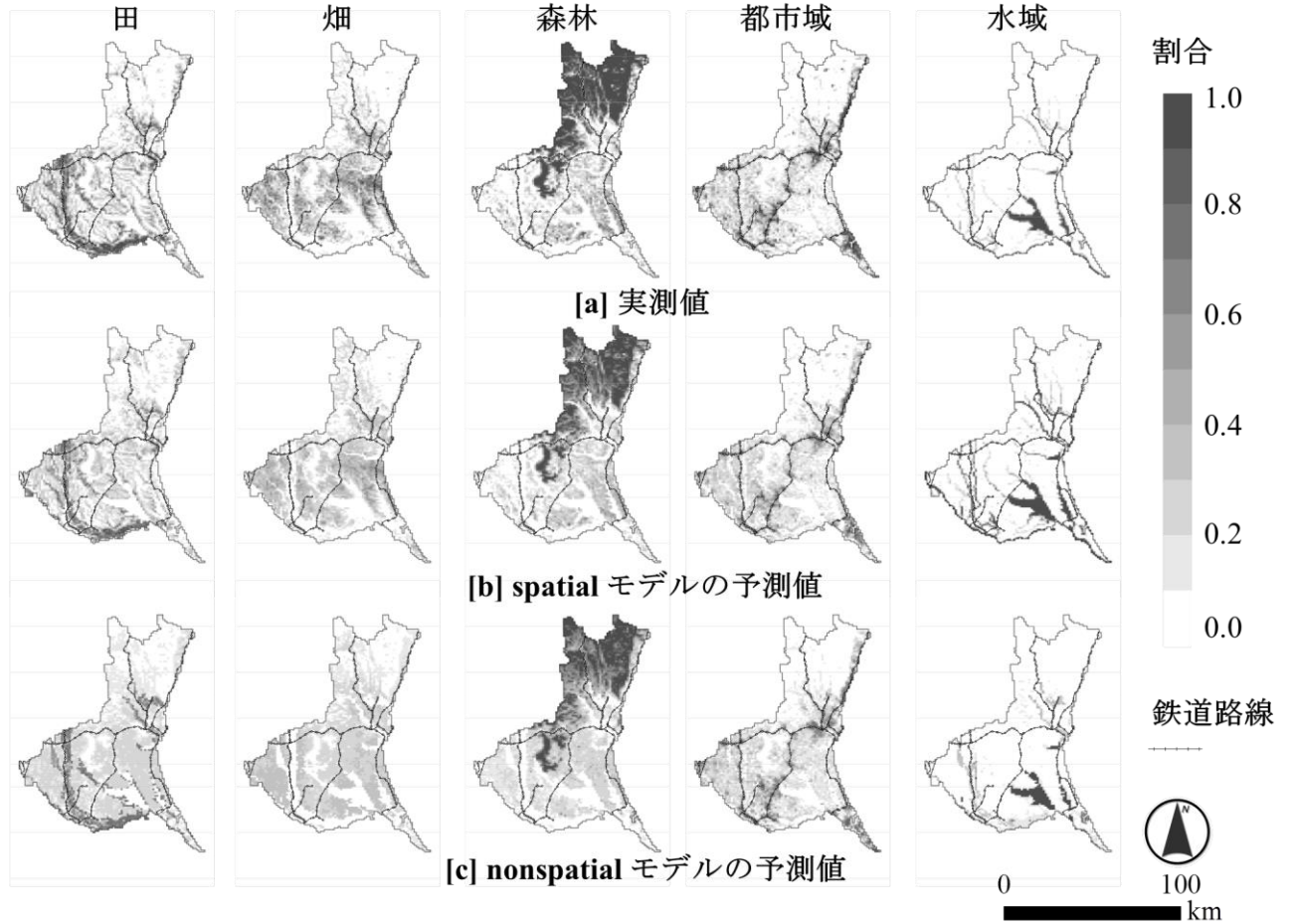


図-3 実測値と事後平均による各土地利用種の予測値

参考文献

- 1) Aitchison, J.: The Statistical Analysis of Compositional Data, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 44, No. 2, pp. 139–177, 1982.
- 2) Aitchison, J.: *The statistical analysis of compositional data*, Chapman and Hall, 1986.
- 3) 太田亨, 新井宏嘉: 組成データ解析の問題点とその解決方法, *地質学雑誌*, Vol. 112, No. 3, pp. 173–187, 2006.
- 4) 瀬谷創, 堤盛人: 空間統計学—自然科学から人文・社会科学まで—, 朝倉書店, 2014.
- 5) Pawlowsky-Glahn, V. and Olea, R.: *Geostatistical Analysis of Compositional Data*, Oxford University Press, 2004.
- 6) Mardia, V.: Multi-dimensional Multivariate Gaussian Markov Random Fields with Applications to Image Processing, *Journal of Multivariate Analysis*, Vol. 24, No. 2, 265–284, 1988.
- 7) Allen, J., Leininger, T., Hurd, J., Civco, D., Gelfand, A., and Silander, J.: Socioeconomics drive woody Invasive plant richness in New England, USA through forest fragmentation, *Landscape Ecology*, Vol. 28, No. 9, pp. 1671–1686, 2013.
- 8) Leininger, T., Gelfand, A., Allen, J., and Silander, J.: Spatial Regression Modeling for Compositional Data With Many Zeros, *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 18, No. 3, pp. 314–334, 2013.
- 9) Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H.-W., and Silkey, M.: Natural variability of benthic species composition in the Delaware Bay, *Environmental and Ecological Statistics*, Vol. 4, No. 2, pp. 95–115, 1997.
- 10) 新井宏嘉, 太田亨: 組成データ解析における0値および欠損値の扱いについて, *地質学雑誌*, Vol. 112, No. 7, pp. 439–451, 2006.
- 11) 若松加寿江, 松岡昌志, 杉浦正美, 久保純子, 長谷川浩一: 日本の地形・地盤デジタルマップ, 東京大学出版会, 2005.
- 12) Geweke, J.: Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments, In Bernardo, J., Berger, J., Dawid, A. and Smith, A. (eds.): *Bayesian Statistics 4*, Clarendon Press, pp. 169–193, 1992.
- 13) Martin-Fernandez, J., Barcelo-Vidal, C., and Pawlowsky-Glahn, V.: Dealing With Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation, *Mathematical Geology*, Vol. 35, No. 3, pp. 253–278, 2003.

(2014. 4. 25 受付)