

固有ベクトル空間フィルタリング に基づく面補間

村上大輔¹・堤盛人²

¹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1-1-1)
E-mail: muraka51@sk.tsukuba.ac.jp

²正会員 筑波大学准教授 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1-1-1)
E-mail: tsutsumi@sk.tsukuba.ac.jp

社会経済データの多くは、空間的な単位毎に集計されたデータとして提供されており、データの持つ集計単位の変換が必要となることも多い。そのような変換は面補間と呼ばれる。これまでに、様々な面補間法が提案されており、空間データの一般的な性質として知られる空間従属性を考慮した面補間法の提案も行われつつある。しかしながら、空間従属性を考慮した既存手法は標準的な統計パッケージを用いた実装が難しく、実用性の面で課題も残されている。そこで本研究では、まず、固有ベクトル空間フィルタリングに基づいて空間従属性を考慮した面補間法を提案する。提案手法は最小二乗法に基づいて面補間ができる実用的な手法である。次に、提案手法の拡張として、地理的加重回帰、及び一般化線形モデルの援用を行う。最後に、提案手法を高齢化率の面補間に適用し、その予測能力を検証する。

Key Words : *eigenvector spatial filtering, areal interpolation, spatial dependence, ordinary least squares*

1. はじめに

幅広い用途に対応した空間データハンドリングのためのシステムの構築は、地理情報システムの担う重要な課題であり、集計データを用いた分析のための方法論の整備もまた、その中の一つに位置づけられる重要な課題である。

集計データを用いた分析に際しては、データ間の集計単位の違いが問題となる。例えば、我が国では、社会経済データの多くは行政単位毎、土地利用データの多くはメッシュ毎に集計されており、両データを同一の分析に対してそのまま適用することはできない。データ間の集計単位の違いへの対処として、データの空間的な単位の変換は有用である。そのような変換は面補間 (Areal interpolation) と呼ばれる (例えばSadahiro (1999))。面補間は、国勢調査の

メッシュ別データの作成など、実務でも多くの適用が見られる実用的手法である。以降では、面補間前の空間単位を『集計単位(Source zone)』, 面補間後の集計単位を『配分単位(Target zone)』と呼び議論を進める。

面補間法が満足すべき最も基本的な性質 (Lam (1983)) として、体積保存則 (Pycnophylactic property) が知られている。これは、同一領域内の集計値は面補間前後で不変であるという性質を表す。以後、本研究では、体積保存則を満足した空間単位間の変換を面補間と定義することとする。

これまでに、様々な面補間法が提案されてきた。例えば、面積に応じた比例配分に基づいて面補間を行う面積按分法 (Wright (1936)) と、ポイントデータに置き換えられた集計データを集計することで面補間を行う代表点内包法 (例えば貞広 (2000)) は、実務での幅広い適用がみられる。

一方で、補助的なデータ (例えば衛星画像データ) を用いることで、面補間の対象となる変数の空間分布を考慮することのできる手法も多数提案されており、その補間精度の高さが複数の論文で指摘されている (例えば, Howley, 2005, Saporito et al., 2007)。

また、補助的な情報として空間従属性 (Spatial dependence : 例えば, Cressie (1993)) を考慮した面

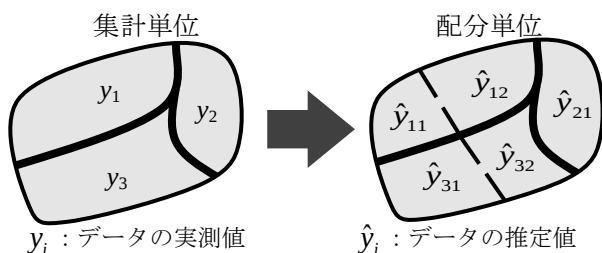


図1 面補間の概要

補間法の提案も行われつつある（例えば、堤・村上 (2010)）。空間従属性とは空間的に近接したデータ間には類似した性質が見られるという空間データの基本的な性質の一つである。Kyriakidis (2004), Gotway and Young (2007), Mugglin et al.(1999)は空間統計学（例えば、Cressie (1993)）に基づいて、堤・村上 (2010)は空間計量経済学（例えば、LeSage and Pace (2009)）に基づいて、空間従属性を考慮した面補間法を提案しており、それぞれ、提案手法による精度の向上を示している。しかしながら、空間従属性を考慮した各手法はいずれも、標準的な統計パッケージを用いた計算が難しく、実用性に課題も残されている。また、各手法のモデルを、補間するデータの特性に応じて修正（拡張）することは、一般に難しい。

その一方で、空間計量経済学に基づいて空間従属性を考慮した手法の一つに固有ベクトル空間フィルタリング（Eigenvector Spatial Filtering：例えば Griffith, 2003）がある。同手法は標準的な統計パッケージを用いて実装することができ（Griffith (2003)）、かつデータの特性に応じたモデルの拡張が比較的容易である（詳しくは2章）。

以上を踏まえ、本研究では、補助的なデータと空間従属性を考慮でき、かつ実用性・拡張可能性のある面補間法を固有ベクトル空間フィルタリングに基づいて構築する。

以降では、第2章で固有ベクトル空間フィルタリングの理論を概観し、第3章で面補間の一般的な計算手順について説明する。第4章では、統計モデルに基づいた代表的な面補間である Flowerdew and Green (1992)の手法について議論し、それを踏まえ、第5章では、固有ベクトル空間フィルタリングに基づいて空間従属性を考慮した面補間法を提案する。第6章では、提案手法が容易に拡張できることを示した上で、拡張の一例として、空間的な異質性を考慮した計量地理学のモデルである GWR（Geographical Weighted Regression：Fotheringham et al. (2002)）、及びデータに様々な分布を仮定することができるモデルである GLM（General Linear Model, Nelder and Wedderburn (1972)）を提案手法に援用する方法について議論する。最後に第6章において、提案手法を実際のデータに適用することで、その有用性を検証する。

2. 固有ベクトル空間フィルタリングの概要

計量地理学等において良く用いられる、空間従属性の有無を検定することのできる統計量に次式で表されるモラン I 統計量がある。

$$I_{\text{moran}} = \mathbf{y}'(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{y}, \quad (1)$$

I_{moran} はモラン I 統計量、 n はデータ数、 $\mathbf{y}$ は $n \times 1$ の変数のベクトル、 $\mathbf{I}$ は $n \times n$ の単位行列、 $\mathbf{1}$ は 1 を要素に持つ $n \times 1$ のベクトル、 $\mathbf{C}$ は $n \times n$ の隣接行列である。モラン I 統計量は -1 から 1 の間の値をとり、 -1 に近ければ $\mathbf{y}$ は強い負の空間従属性を持ち、反対に 1 に近ければ $\mathbf{y}$ は強い正の空間従属性を持つとされる。モラン I 統計量は行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n) \mathbf{C} (\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ を用いて $\mathbf{y}$ の近接性を定義した統計量である。

固有ベクトル空間フィルタリングとは、(2)式より算出される固有ベクトルを用いて、空間従属性に起因したデータの変動をモデル化するアプローチである。

$$\{\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_n\} = \text{evec}[(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)], \quad (2)$$

ここで、 $\text{evec}[\]$ は $[\]$ 内の行列の固有ベクトルを表す演算子である。 $\mathbf{E}_1$ は、行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n) \mathbf{C} (\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ の、最大固有値を持つ固有ベクトル ($n \times 1$) であり、モラン I 統計量で説明することのできる最大の空間変動を表す。また、 $\mathbf{E}_2$ は、2番目に大きな固有値を持つ固有ベクトル ($n \times 1$) であり、 $\mathbf{E}_1$ と無関連な成分のうちで、モラン I 統計量で説明することのできる最大の空間変動を表す。同様に、 $\mathbf{E}_n$ は $(n-1)$ 番目までの全ての固有ベクトルと無関連な成分のうちで、モラン I 統計量で説明することのできる最大の空間変動を表す。

(2)式より与えられる各固有ベクトルを線形回帰モデルに投入することで、(3)式が得られる。

$$y_s = \sum_p x_{s,p} \beta_p + \sum_q E_{s,q} \gamma_q + \varepsilon_s, \quad \varepsilon_s \sim N(0, \sigma^2), \quad (3)$$

ここで s, p, q はそれぞれ、地点、説明変数、固有ベクトルを表す添え字である。 y_s は地点 s の被説明変数、 $x_{s,p}$ は地点 s の p 番目の説明変数、 ε_s は攪乱項、 $\beta_p, \gamma_q, \sigma^2$ はパラメータ、 $E_{s,q}$ は、(3)式に投入された固有ベクトルのうちの、 q 番目の固有ベクトルの第 s 要素を表す。固有ベクトルの選択はステップワイズ法等により行うことができる。(2)式の右辺第一項は、補助的なデータより説明される y_s の変動、右辺第二項は潜在的な空間変動または空間パターンより説明される y_s の変動を表す（Tiefelsdorf and Griffith (2007)）。

一般に、空間統計学、空間計量経済学の殆どのモデルは、近接行列を用いて空間従属性を考慮した結果として内生性の問題が生じるため、パラメータの推定に通常最小二乗法が適用できない（Anselin (2007)）。しかしながら、固有ベクトル空間フィルタリングは、近接行列自体ではなく、その固有ベクトルを用いて空間従属性を考慮する手法であり、内生性の問題が生じないため通常最小二乗法が適用できる。従って、固有ベクトル空間フィルタリングは、標準的な統計パッケージを用いて容易に実装できるという特徴を有する。

また、(3)式は本質的には通常線形回帰モデルと同一であるため、様々な拡張が容易である。例えば、ポアソン回帰モデルの説明変数に上述の固有ベクトルを加えることで、ポアソン分布に従うデータのためのモデルに拡張できる（Griffith (2010a)）。また、(3)式の誤差項の分散に、何らかの変数で重み付けした上で重み付き最小二乗法を適用することで、空間従属性と分散不均一を考慮したモデルに拡張することもできる。さらに、固有ベクトル空間フィルタリングはデータ補間手法としての拡張も試みられており、その精度の高さが示されている（Griffith (2010b)）。

このように、固有ベクトル空間フィルタリングは、実用性・拡張可能性に富み、精度の良い面補間法を構築するために有用な手法となり得ると考えられる。

3. 面補間の一般的な手順

集計単位、配分単位、さらに必要に応じて補助データ等の空間単位を重ね合わせることで得られる空間単位を『細分単位（Partial zone）』と定義し（図2）、集計単位、配分単位、細分単位の各領域を以下のように表現する。

集計単位の各領域： A_i ($i = 1, 2, \dots, I$)

配分単位の各領域： B_j ($j = 1, 2, \dots, J$)

細分単位の各領域： C_k ($k = 1, 2, \dots, K$)

また、各空間単位についての被説明変数ベクトルは y_i, y_j, y_k と表現する。各空間単位の数値は、それぞれ I, J, K である。

一般に、面補間は、以下の手順に基づいて行われることが多い。

[1] 集計単位毎のデータ $\bar{y}_i$ の按分（面補間）による、細分単位毎のデータ $\hat{y}_k$ の算出

[2] [1]の結果得られた細分単位の推定値 $\hat{y}_k$ の集計（面補間）による配分単位毎の推定値 $\hat{y}_j$ の算出

手順[1]では、推定誤差の発生が避けられないため、精度良く面補間を行うことのできるモデルが必要となる。手順[1]では、 $\bar{y}_i$ が密度を表すデータ（例：人口密度）であれば(4)式、 $\bar{y}_i$ がカウントデータ（例：人口）であれば(5)式が、体積保存則を満足するための条件式となる。

$$\bar{y}_i = \sum_{\{k|C_k \subset A_i\}} \frac{n_k}{n_i} \hat{y}_k, \quad \bar{y}_i = \sum_{\{k|C_k \subset A_i\}} \hat{y}_k, \quad (4) \quad (5)$$

n_i は配分単位 A_i の重み、 n_k は細分単位 C_k の重みを表す。例えば、 $\bar{y}_i$ が面積で割られた変数であれば、 n_i 及び n_k の各要素は面積、人口で割られた変数であれば、 n_i 及び n_k の各要素は人口が与えられる。

その一方で、手順[2]では、 $\hat{y}_j$ が密度を表すデー

タであれば(6)式、カウントデータであれば(7)式を用いることで、推定誤差なしに面補間ができる。

$$\hat{y}_j = \sum_{\{k|C_k \subset B_j\}} \frac{n_k}{n_j} \hat{y}_k, \quad \hat{y}_j = \sum_{\{k|C_k \subset B_j\}} \hat{y}_k, \quad (6) \quad (7)$$

n_j は配分単位 B_j の重みを表す。

従って、以降では、手順[1]の面補間を精度よく行うためのモデルについて議論する。なお、面補間はカウントデータを対象とすることが多いが、密度を表すデータを対象とした面補間が必要とされることも多い。本研究では、カウントデータと密度を表すデータの両方の面補間に対する面補間法を、それぞれ構築することとする。

4. 統計モデルに基づいた面補間法

統計モデルに基づいた代表的な面補間法として、EM（Expectation-Maximization）アルゴリズムを援用した Flowerdew and Green (1992) の手法が知られている。本章では、まず、EMアルゴリズムの概要について述べ、次に、Flowerdew and Green (1992) で提案された手法とその問題点について議論する。

(4-1) EMアルゴリズム

EMアルゴリズムとは、観測データ $\mathbf{z}$ と欠損データ $\mathbf{z}_0$ から成る完全データ $\mathbf{Z}\{\mathbf{z}, \mathbf{z}_0\}$ の対数尤度 $l(\mathbf{Z}, \boldsymbol{\theta})$ の期待値を最大化することのできるアルゴリズムである。ここで $\boldsymbol{\theta}$ は $\mathbf{Z}$ を表現するパラメータのベクトル ($\theta \times 1$) である。EMアルゴリズムは、以下のステップを、対数尤度が収束するまで交互に繰り返すことで実行される（例えば小西ほか(2008)）。

・E（Estimation）ステップ

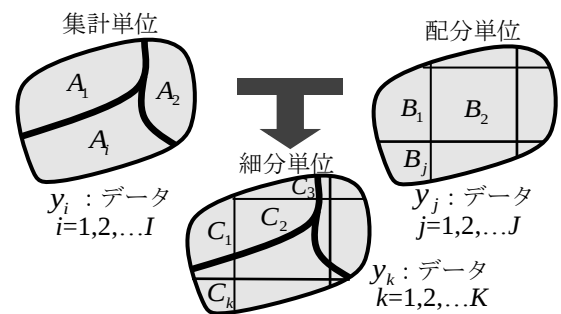


図2 細分単位の概要

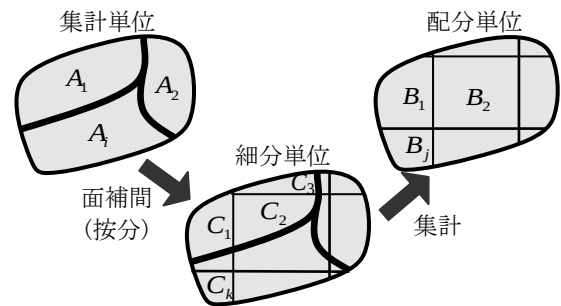


図3 面補間の手順

t 回目の繰り返しで得られたパラメータ $\theta^{(t)}$ が与えられた下での、完全データ $\mathbf{Z}$ の対数尤度 $l(\mathbf{Z}, \theta)$ の期待値 $L(\theta | \theta^{(t)})$ を求める。

$$L(\theta | \theta^{(t)}) = E[l(\mathbf{Z}, \theta) | \mathbf{z}, \theta^{(t)}], \quad (8)$$

• M (Maximization) ステップ

$L(\theta | \theta^{(t)})$ を最大化する θ を求め、 $\theta^{(t+1)}$ とする。

$$\theta^{(t+1)} = \arg \max_{\theta} L(\theta | \theta^{(t)}), \quad (9)$$

(8), (9)式より、両ステップが同一の(対数)尤度関数に基づいて行われることが、理論に整合したEMアルゴリズムの必要条件であるとわかる。

(4-2) Flowerdew and Green (1992) の概要

Flowerdew and Green (1992)は、集計単位から細分単位への面補間（手順[1]：詳しくは3章）を、以下に示す2つのステップを交互に繰り返すことで実行した。

• Eステップ

(10)式に基づいた $\hat{y}_k$ の期待値の計算

$$\hat{y}_k = \sum_p x_{k,p} \beta_p + \left[\bar{y}_i - \sum_{\{k|C_k \subset A_i\}} \sum_p \frac{n_k}{n_i} x_{k,p} \beta_p \right] + u_k \quad (10)$$

ここで、 $x_{k,p}$, u_k は細分単位 k の p 番目の説明変数、 q 番目の固有ベクトル、期待値が 0 の攪乱項を表す。同ステップは(10)式の対数尤度関数の期待値を計算するステップとみなすことができる。

• Mステップ

通常の線形回帰モデルの尤度最大化に基づいたパラメータの推定

Flowerdew and Green (1992)は以上のアルゴリズムがEMアルゴリズムであることを前提に手法を構築した。しかしながら、上記のEステップとMステップは互いに異なる尤度関数に基づいており、EMアルゴリズムの理論と整合していない。

そこで本研究では、彼らの手法を参照しながら、本来のEMアルゴリズムの理論と整合した面補間を行うことのできる手法を構築することとする。

5. 固有ベクトル空間フィルタリングの面補間への応用

集計単位から細分単位への面補間（手順[1]：詳しくは3章）で用いるモデルの構築に際し、細分単位毎に想定された被説明変数 $\hat{y}_k$ が下式に従うものと仮定する。

$$\hat{y}_k = \sum_p x_{k,p} \beta_p + \sum_q E_{k,q} \gamma_q + u_k, \quad (11)$$

ここで $E_{k,q}$ は第 q 固有ベクトルの第 k 要素、 u_k は期待値が 0 の攪乱項、 β_p , γ_q はパラメータを表す。なお、ここでは u_k に正規分布を仮定しない。

被説明変数が密度を表すデータであれば、体積保存則満足条件式(4)式に(11)式を代入することで、(12)式が与えられる。

$$\bar{y}_i = \sum_{\{k|C_k \subset A_i\}} \sum_p \frac{n_k}{n_i} x_{k,p} \beta_p + \sum_{\{k|C_k \subset A_i\}} \sum_q \frac{n_k}{n_i} E_{k,q} \gamma_q + \varepsilon_i \quad (12)$$

ここで、 ε_i は集計単位毎の攪乱項を表す。

被説明変数がカウントデータである場合についても同様にモデルを構築することで、(13)式が与えられる。

$$\bar{y}_i = \sum_{\{k|C_k \subset A_i\}} \sum_p x_{k,p} \beta_p + \sum_{\{k|C_k \subset A_i\}} \sum_q E_{k,q} \gamma_q + \varepsilon_i \quad (13)$$

(12)式または(13)式より与えられる $\bar{y}_i$ の予測値がその実測値と等しくなるという制約を(11)式に与えることで、密度を表すデータのための面補間の推定式(14)式、及び、カウントデータのための推定式(15)式が導出される。

$$\hat{y}_k = \sum_p x_{k,p} \beta_p + \sum_q E_{k,q} \gamma_q + u_k + \left[\bar{y}_i - \sum_{\{k|C_k \subset A_i\}} \sum_p \frac{n_k}{n_i} x_{k,p} \beta_p - \sum_{\{k|C_k \subset A_i\}} \sum_q \frac{n_k}{n_i} E_{k,q} \gamma_q - \varepsilon_i \right], \quad (14)$$

$$\hat{y}_k = \sum_p x_{k,p} \beta_p + \sum_q E_{k,q} \gamma_q + u_k + \frac{1}{N_i} \left[\bar{y}_i - \sum_{\{k|C_k \subset A_i\}} \sum_p x_{k,p} \beta_p - \sum_{\{k|C_k \subset A_i\}} \sum_q E_{k,q} \gamma_q - \varepsilon_i \right], \quad (15)$$

ここで、 N_i は集計単位 i に含まれる細分単位の数を表す。(14), (15)式は、Flowerdew and Green (1992) のモデルの一般形とみなすことができる。

ここで、(14), (15)式は未知変数 $\hat{y}_k$ を含むため、そのままではパラメータ β_p , γ_q が推定できない。

一方で、(12), (13)式は未知変数を含まないため、最小二乗法等に基づいた β_p と γ_q の推定ができる。

また、密度を表す変数であれば(14)式、カウントデータであれば(15)式に β_p と γ_q の推定値を代入することで、 $\hat{y}_k$ の面補間ができる。

以上より算出される $\hat{y}_k$ は体積保存則を必ず満足する。また、 $E_{k,q}$ を考慮した場合、空間従属性が考慮される。提案手法は、筆者らの知る限り、最小二乗推定に基づいて空間従属性を考慮した面補間を行うことのできる唯一の手法であり、標準的な統計パッケージを用いて実装できるという利点を持つ。また、提案手法のパラメータ推定に用いるモデルである(12), (13)式は、本質的には、通常の線形回帰モデルと同一であるため、通常の固有ベクトル空間フィルタリングと同様に幅広い拡張可能性を持つ（詳しくは2章）。従って、提案手法は実用性、拡張可能性の両観点から、空間従属性を考慮した他の面補

間法よりも優れている。

6. 固有ベクトル空間フィルタリングに基づいた面補間法の拡張

既述のように、提案手法は様々な手法の援用に基づいた幅広い拡張ができる。そこで本章では、拡張の一例として、GWR (Geographic Weighted Regression) 及びGLM (Generalized Linear Model) を、提案手法に援用する方法について示す。なお、両手法の固有ベクトル空間フィルタリングに対する援用の有用性は、Griffith (2010a)で示唆されている。

(6-1) GWRに基づいた提案手法の拡張

(6-1-1) GWRとは

GWRは、回帰係数のパラメータを地点毎に与えることで空間な異質性を考慮したモデルであり、(16)式が基本モデルとなる。

$$y_s = \sum_p x_{s,p} \beta(r)_{s,p} + \varepsilon_s, \quad (16)$$

ここで、 r はバンド幅と呼ばれるパラメータである。 $\beta(r)_{s,p}$ は地点 s の p 番目の説明変数のパラメータを表し、その各要素は(17)式より与えられる。

$$\beta(r)_s = [X'W(r)X]^{-1}X'W(r)y, \quad (17)$$

ここで、 $\beta(r)$ は $\beta(r)_{s,p}$ を要素に持つ $p \times 1$ のベクトル、 y は y_s を要素に持つ $n \times 1$ のベクトル、 X は $x_{s,p}$ を要素に持つ $n \times p$ の行列である。 $W(r)$ は $n \times n$ の対角行列であり、 s' 番目の対角要素 $w_{s,s'}(r)$ は、地点 s に対する地点 s' の重みを表す。 $w_{s,s'}(r)$ は、例えばガウス型関数(18)式で与えることができる。

$$w_{s,s'} = \exp(-d_{s,s'}^2 / r^2), \quad (18)$$

GWRについて、より詳しくは、Fotheringham et al. (2002)等を参照されたい。

(6-1-2) 提案手法に対するGWRの援用

密度を表す変数の面補間を前提とした場合、(12)、(14)式に(16)式を援用することで、GWRを援用した面補間のパラメータ推定のためのモデル(19)式、及補間を行うためのモデル(20)式が導出される。

$$\bar{y}_i = \sum_{\{k|C_k \subset A_i\}} \sum_p \frac{n_k}{n_i} x_{k,p} \beta(r)_{i,p} + \sum_{\{k|C_k \subset A_i\}} \sum_q \frac{n_k}{n_i} E_{k,q} \gamma(r)_{i,p} + \varepsilon_i, \quad (19)$$

$$\hat{y}_k = \sum_p x_{k,p} \beta(r)_{i,p} + \sum_q E_{k,q} \gamma(r)_{i,p} + u_k + \left[\bar{y}_i - \sum_{\{k|C_k \subset A_i\}} \sum_p \frac{n_k}{n_i} x_{k,p} \beta(r)_{i,p} - \sum_{\{k|C_k \subset A_i\}} \sum_q \frac{n_k}{n_i} E_{k,q} \gamma(r)_{i,p} - \varepsilon_i \right], \quad (20)$$

従って、(19)式のパラメータ推定を、従来のGWRと

同様の手順で行い、その結果として与えられるパラメータ推定値を(20)式に導入することで、空間従属性に加え、空間的な異質性を考慮した面補間が可能となる。

(6-2) GLMに基づいた提案手法の拡張

(6-2-1) GLMとは

GLMとは、連続変数、離散変数に対する回帰分析を統一的に行えるように、線形回帰モデルを系統的に拡張したモデルであり、以下の2つの条件を満足するモデルとして定義される。

[1] y_s の条件付き期待値 μ_{ys} が (21)式に従う。

$$\eta(\mu_{ys} | \beta_p) = \sum_p x_{s,p} \beta_p, \quad (21)$$

$\eta(\mu_{ys})$ はリンク関数と呼ばれる関数である。

[2] y_s の条件付き確率密度分布が指数分布族、即ち、(22)式で表される分布に従う。

$$f(y_s | \varphi) = \exp\left(\frac{y_s \varphi - a(\varphi)}{\psi}\right) + b(y_s, \varphi) \quad (22)$$

φ , ψ はパラメータ、 $a(\cdot)$, $b(\cdot)$ は関数を表す。

指数分布族には、正規分布、ポアソン分布、2項分布など幅広い分布が存在し、GLMは幅広いデータに適用できる。GLMのパラメータ推定は最尤法に基づいて行う。

なお、GLMについてより詳しくは、Dobson and Barnett (2008)を参照されたい。

(6-2-2) 提案手法に対するGLMの援用

例えば、各空間単位の個体の数を表すカウントデータの面補間に際しては、所与の空間内で発生する現象の数を表す際に良く用いられる分布であるポアソン分布をデータの分布に仮定して用いることが望ましい場合も多い。ポアソン分布を持つデータの面補間のための手法として提案手法を拡張するためには、まず、 $\bar{y}_i$ の条件付き期待値 μ_{yi} が(23)式、 $\bar{y}_i$ の条件付き確率密度分布が(24)式で示されるポアソン分布にそれぞれ従うと仮定した上で、通常のGLMと同様の方法でパラメータ推定を行う。

$$\eta(\mu_{yi} | \beta_p, \gamma_q) = \sum_{\{k|C_k \subset A_i\}} \sum_p \frac{n_k}{n_i} x_{k,p} \beta_p + \sum_{\{k|C_k \subset A_i\}} \sum_q \frac{n_k}{n_i} E_{k,q} \gamma_q, \quad (23)$$

$$f(\bar{y}_i | \beta_p, \gamma_q, \phi) = \frac{\lambda^{\bar{y}_i} \exp(-\lambda)}{\bar{y}_i!}, \quad (24)$$

ここで、 λ はパラメータである。

次に、 $\hat{y}_k$ の条件付き期待値 μ_{yk} が(25)式で与えられることを前提に、(15)式に基づいて構築される予測式(26)式を用いて、面補間を行うことができる。

$$\eta(\mu_{yk} | \beta_p, \gamma_q) = \sum_p x_{k,p} \beta_p + \sum_q E_{k,q} \gamma_q, \quad (25)$$

$$\hat{y}_k = \hat{\mu}_{\hat{y}_k} + \left[\bar{y}_i - \frac{1}{N_i} \sum_{k|C_k \subset A_i} \hat{\mu}_{\hat{y}_k} \right], \quad (26)$$

$\hat{\mu}_{\hat{y}_k}$ は, (23), (24)式から得られたパラメータ推定値を(25)式に代入することで与えられる, $\mu_{\hat{y}_k}$ の推定値である.

7. 実証分析: 提案手法の高齢化率の面補間への適用

(7-1) 検証の概要

北関東 3 県 (茨城県, 栃木県, 群馬県) のデータを用いた実証により提案手法の有用性を検証する. ここでは, 2007 年度の 116 市町村区分を集計単位, 1995 年度の 203 市町村区分を配分単位とした高齢化率の面補間を行う.

本実証の予測精度の検証のためには, 1995 年度の市町村区分で集計された 2007 年度のデータが必要であるため, ここでは, 2007 年度の町丁目データからそれを作成することとする. なお, 対象地域内の各配分単位は, いずれかの集計単位に包含されるため, 配分単位と細分単位は一致し, (6)式を用いた予測値の集計作業は不要である.

本実証では, 提案モデルである(14)式を[SFilter], Flowerdew and Green (1992)のモデル, 即ち, (14)式から固有ベクトルの項を取り除いたモデルを[Non-Spatial], 及び, 提案モデルに GWR を援用したモデルである(20)式を[GWR-SFilter]と呼称することとし, それぞれを面補間に適用することで, 予測精度の比較等を行うこととする.

説明変数は, 東京駅までの距離 (km) [都心距離], 最寄りの県庁所在市町村の主要駅までの距離 (km) [県庁距離]及び人口密度 (千人/km²) を用いる. これらは, 町丁目単位でのデータ入手・作成が容易であり, かつ, 高齢化率との相関が考えられるという理由により採用しており, 社会学あるいは計量地理学等の知見に基づくものではないことを断っておく. また, [SFilter], [GWR-SFilter]の計算に必要な固有ベクトルは, 隣接行列Cを(2)式に投入

することで抽出することとする. 但し, 本研究では, 最大固有値から20番目に大きな固有値を持つ各固有ベクトルまでのみを抽出し, それらから, ステップワイズ法を用いた固有ベクトルの選択を行うこととする. なお, 以上の全ての変数を用いたVIF診断を行った結果として, 多重共線性が存在しないことを確認している.

計算はフリーの統計ソフトウェア「R」を用いた独自のプログラムにより行い, 地図の描写にはESRI社の「ArcGIS」を用いる.

(7-2) パラメータ推定結果

各モデルのパラメータ推定結果を表 1 に示す. 但し, [GWR-SFilter]の各推定値は, 各集計単位に与えられた推定値の平均値を表す. 表 1 から, 人口密度が各モデルで有意な負の効果を持つこと, 県庁距離が[Non-Spatial], [GWR-SFilter]で有意な正の効果を持ち, [SFilter]の正の推定値もまた比較的高い t 値を示していることがわかる. 以上の結果は直感と整合する. その一方で, 都心距離は, 各モデルで有意ではないが, これは, 対象地域が都心から離れており有意な効果が検出されなかった可能性がある.

次に, 各固有ベクトルに対するパラメータ推定値をみると, [SFilter], [GWR-SFilter]の各推定値は類似した傾向を示していることがわかる. 各固有値と対応するパラメータ推定値の積の和を, 図 5 に示す. 図中の黒い集計単位は, 空間従属性が高齢化率を引き上げている地域, 白い集計単位はその逆を表す. 図 5 から, 対象地域西部の高齢化率が空間従属性により引き上げられている可能性が示唆された.

各モデルの予測精度は, (27)式で示される AIC(Akaike Information Criterion)を用いて比較する.

$$AIC = -2l(\theta) + 2P \quad (27)$$

ここで $l(\theta)$ は θ をパラメータに持つモデルの対数尤度, P はモデルのパラメータ数を表す. 各モデルの AIC (表 1) をみると, [SFilter] 及び [GWR-SFilter] の精度が, [Non-Spatial] の精度を大きく上回っており, 固有ベクトル空間フィルタリングの援用による

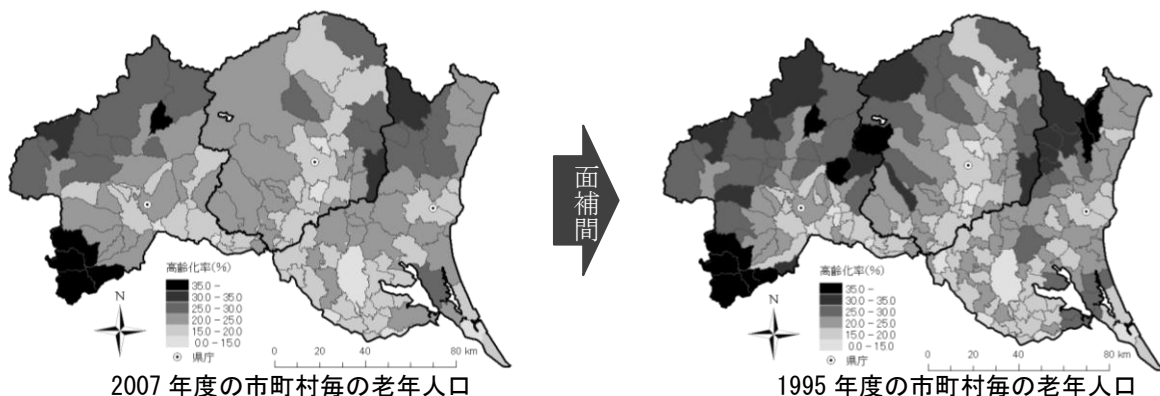


図4 実証における面補間の概要

表1 パラメータ推定結果

変数	Non-Spatial			SFilter			GWR-SFilter (平均値)		
	推定値	t 値		推定値	t 値		推定値	t 値	
定数項	18.3	7.90	***	23.3	9.39	***	23.3	9.34	***
都心距離(km)	4.61×10^{-2}	1.56		-5.58×10^{-2}	-1.57		-5.58×10^{-2}	-1.51	
県庁距離(km)	6.30×10^{-2}	3.70	***	2.86×10^{-2}	1.59		2.86×10^{-2}	1.65	*
人口密度(千人/km ²)	-5.40	-5.33	***	-3.78	-3.53	***	-3.78	-3.57	***
第 1				-10.0	-1.67	*	-10.1	-1.67	*
第 3				10.1	1.68	*	9.73	1.68	*
第 5				7.21	1.13		6.83	1.11	
第 6 固有				9.97	1.32		9.20	1.30	
第 9 ベクトル				24.2	3.59	***	23.6	3.58	***
第 10				-18.7	-3.24	***	-19.1	-3.24	***
第 14				-26.1	-3.78	***	-26.3	-3.78	***
第 15				-13.9	-2.35	**	-13.8	-2.33	**
バンド幅(km)							280		
AIC	692.3			665.7			666.5		

※※ ***は 1%, **は 5%, *は 10%水準で有意であることを表す



図5 抽出された空間フィルタ

モデルの精度向上が確認できる。その一方で、[SFilter]と[GWR-SFilter]のAICには差がみられない。GWRを援用することがモデルの精度を向上させないのは、対象地域に顕著な空間的な異質性が存在しないためである可能性がある。

(7-3) 予測精度の検証

各モデルを用いた面補間を実際に行うことで、予測精度の高さを確認する。予測精度の検証は(28)式の平均 2 乗平方根誤差 (RMSE: Root Mean Square Error) により行う。

$$RMSE = 100 \times \sqrt{\sum_{k'} \frac{(\hat{y}_{k'}^{obs} - \hat{y}_{k'})^2}{K'}}, \quad (28)$$

ここで、 k' は合併により消滅した細分単位 (全 K' 箇所)、即ち、面補間の対象となる細分単位の添え字、 $\hat{y}_{k'}^{obs}$ は細分単位 k' の高齢化率実測値を表す。

表 2 は RMSE の算出値を表す。表から、今回のデータに関しては[SFilter]、[GWR-SFilter]、[Non-

Spatial]の順で補間の精度が良いとわかる。また、固有ベクトル空間フィルタリングを援用した2つのモデルの RMSE が、[Non-Spatial]の RMSE に比べて、10%程度小さいとわかる。以上より、固有ベクトル空間フィルタリングの援用を援用することで、面補間の精度が向上することが確認された。補間の対象となる各市町村の高齢化率実測値、及び各モデルを用いた補間結果を図 6 に示す。

最後に、(29)式で定義される誤差率を用いて、モデル間の補間精度の差を細分単位毎に確認する。

$$\text{誤差率} = 100 \times \left| \frac{\hat{y}_{k'} - \hat{y}_k^{obs}}{\hat{y}_k^{obs}} \right|, \quad (29)$$

[SFilter]と[Non-Spatial]の誤差率の差を図 7 の右図に示した。図中の白い円は空間フィルタリングの援用により精度が向上した細分単位を、黒はその逆を表

表2 RMSE の算出結果

	Non-Spatial	SFilter	GWR-SFilter
RMSE	4.30	3.85	3.96

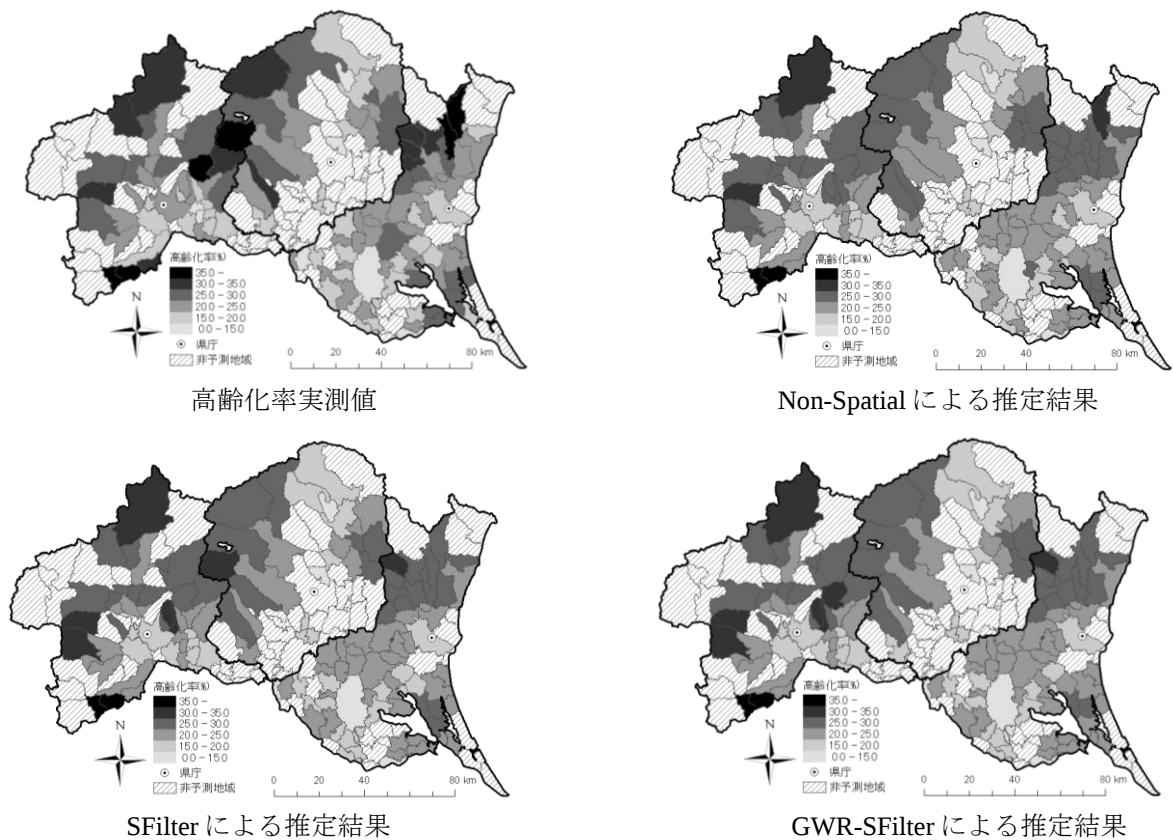


図6 面補間の結果

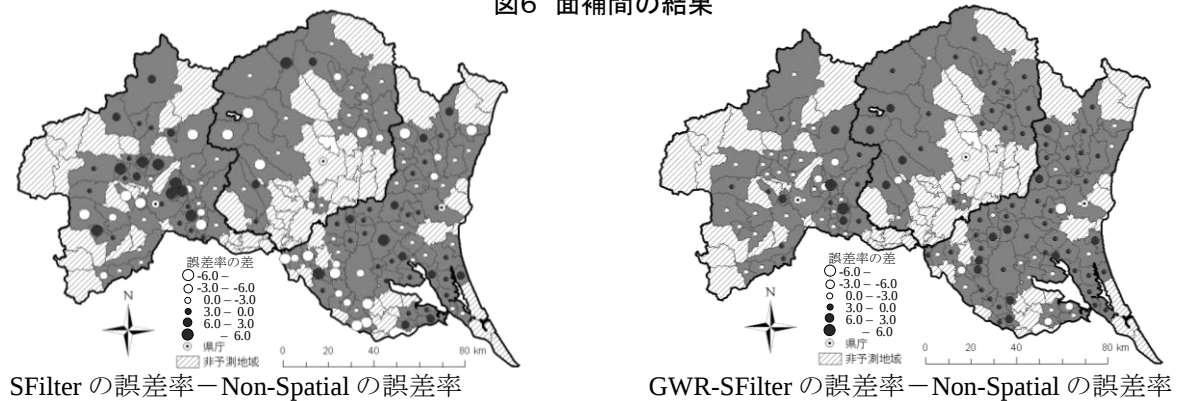


図7 誤差率の差

す。図から、茨城県南部、栃木県全域、群馬県南部等で精度が向上している一方で、群馬県中央部以北、茨城県中央部で精度が低下していることがわかる。全体としては、全細分単位の55%にあたる73市町村で、[SFilter]の補間精度が[Non-Spatial]の精度を上回った。次に、[GWR-SFilter]と[SFilter]の誤差率の差を図7の左図に示した。図中の白い円はGWRの援用により精度が向上した細分単位を、黒はその逆を表す。図から、群馬県全域で精度向上している一方で、その他の地域では精度が低下していることがわかる。全体としては、全細分単位の40%にあたる54市町村で、[GWR-SFilter]の補間精度が[SFilter]の精度を上回った。

8. おわりに

本研究では、空間従属を考慮した面補間法を、固有ベクトル空間フィルタリングを援用することで提案した。提案手法は、最小二乗法を用いた計算が可能な、実用的、かつ計算不可の小さい手法である。また、提案手法は様々な手法の援用が可能であり、援用の一例として、GWR、及びGLMの援用方法を示した。最後に、実データを用いた実証実験から、固有ベクトル空間フィルタリングの援用による空間従属性の考慮が、補間の精度を向上させることを確認した。

本研究の実証実験では、密度を表す変数のみに対し、提案手法を適用したが、カウントデータ等の幅広いデータに対する提案手法の適用も重要な課題である。提案手法の幅広い適用に際しては、本稿で示したGLMの援用も有用と考えられる。また、

Fischer and Griffith (2008)では、フローデータに対する固有ベクトル空間フィルタリングの適用が行われている。従って、提案手法の応用として、フローデータのダウンスケーリングのための手法としての拡張が考えられる。

謝辞

本研究は、(財)日本学術振興会科学研究費補助金(基盤研究(B) 23360219)の成果の一部である。

参考文献

- 1) 貞広幸雄 (2000) 空間集計データにおける面補間法の推定精度評価, 「都市計画」, **225**, 75–81.
- 2) 小西貞則・越智義道・大森祐浩 (2008) 「計算統計学の方法—ブートストラップ・EM アルゴリズム・MCMC—」, 朝倉書店, 第2部, 5章, 69–78.
- 3) 堤盛人・村上大輔 (2010) 市町村合併による統計データの集計単位変更に対する空間計量経済モデルを用いた面補間法, 「応用地域学研究」, **15**, 23–36.
- 4) Anselin, L. (2007) Spatial Econometrics, In Baltagi, B. eds. *A Companion to Theoretical Econometrics*, Oxford: Blackwell.
- 5) Cressie, N. (1993) *Statistics for Spatial Data*. Revised Edition, John Wiley & Sons.
- 6) Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM–algorithm. *Journal of the Royal Statistical Society*, **39** (1), 1–38.
- 7) Dobson, A. J. and Barnett, A. (2008) *An Introduction to Generalized Linear Models (Third Edition)*, Chapman and Hall/CRC.
- 8) Fischer, M. M. and Griffith, D. A. (2008) Modeling Spatial Autocorrelation in Spatial Interaction Data: An Application to Patent Citation Data in the European Union. *Journal of Regional Science*, **48** (5), 969–989.
- 9) Flowerdew, R. and Green, M. (1992) Developments in areal interpolation methods and GIS. *Annals of Regional Science*, **26**, 67–78.
- 10) Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Chichester: Wiley.
- 11) Gelfand, A. E. and Banerjee, S. (2010) Multivariate Spatial Process Models. *Handbook of Spatial Statistics*, Boca Raton, FL: Chapman & Hall/CRC, 495–515.
- 12) Gotway, C. A. and Young, L. J. (2007) A Geostatistical Approach to Linking Geographically Aggregated Data From Different Sources. *Journal of Computational and Graphical Statistics*, **16** (1), 115–135.
- 13) Griffith, D. A. (2003) *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*, Berlin: Springer–Verlag.
- 14) Griffith, D. A. (2010a) Spatial Filtering. In Fischer, M. M. and Getis, A. eds. *Handbook of Applied Spatial Analysis*, Springer.
- 15) Griffith, D. A. (2010b) Spatial Filtering and Georeferenced Data Imputation: A Comparison of the Getis and Griffith Methods. In Anselin, L. and Rey, S. J. eds, *Perspective on Spatial Data Analysis*, 227–233
- 16) Hawley, K. and Mollering, H. (2005) A comparative analysis of areal interpolation methods. *Cartography and Geographic Information Science*, **32** (4), 411–423.
- 17) Kyriakidis, P. C. (2004) A Geostatistical Framework for Area-to-Point Spatial Interpolation. *Geographical Analysis*, **36** (3), 259–289.
- 18) Lam, N. N-S. (1983) Spatial Interpolation Methods: a Review. *American Cartographer*, **10**, 129–149.
- 19) LeSage, J. P. and Pace, R. K. (2009) *Introduction to Spatial Econometrics*, Chapman & Hall/CRC.
- 20) Mugglin, A. S., Carlin, B. P., Zhu, L. and Conlon, E. (1999) Bayesian Areal Interpolation, Estimation, and Smoothing: An Inferential Approach for Geographic Information Systems: Population Interpolation Over Incompatible Zones. *Journal of Agricultural, Biological, and Environmental Statistics*, **3**, pp.117–130.
- 21) Nelder, J. and Wedderburn, R. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society. Series A*, **135** (3), 370–384.
- 22) Sadahiro, Y. (1999) Accuracy of Areal Interpolation: A Comparison of Alternative Methods. *Journal of Geographical Systems*, **1** (4), 323–346.
- 23) Saporito, S., Chavers, J. M., Nixon, L. C. and McQuiddy, M. R. (2007) From here to there: Methods of allocating data between census geography and socially meaningful areas. *Social Science Research*, **36** (3), 897–920.
- 24) Tiefelsdorf, M. and Griffith, D. A. (2007) Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, **39** (5), 1193–1221.
- 25) Tobler W. R. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, **46** (2), 234–240.
- 26) Tobler, W. R. (1979) Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*, **74**, 519–530.
- 27) Wright, J. K. (1936) A Method of Mapping Densities of Population with Cape Cod as an Example. *Geographical Review*, **26**, 103–110.