

RE-ESF-SVC と Fused LASSO を融合した空間的異質性分析手法の提案

彭 湛*・井上 亮**

Proposal of a fusion model combining RE-ESF-SVC and Fused LASSO for extracting multi-scale spatial heterogeneity

Zhan PENG*, Ryo INOUE**

Abstract: Spatial heterogeneity can drive socioeconomic phenomena at different spatial scales. Taking the urban real estate market as an example, housing prices are affected by environmental and social determinants that are continuously changing across the region, such as the valuation of green spaces, proximity to transport, and distance to the central business district. Furthermore, housing prices could also change significantly and discontinuously at a local spatial scale, such as district boundaries and areas along this railway line. Therefore, multi-scale spatial heterogeneity needs to be considered to avoid the misspecification of regional characteristics. In this study, we propose a fusion model that combines the random effects eigenvector spatially filtering-based spatially varying coefficient (RE-ESF-SVC) model with the sparse modeling approach to extract continuous and discontinuous spatial heterogeneity simultaneously. The proposed method is evaluated through the application to simulation data.

Keywords: 空間的異質性 (spatial heterogeneity), 空間可変パラメータモデル (spatially varying coefficient model), 固有ベクトル空間フィルタリング (eigenvector spatial filtering), スパースモデリング (sparse modeling)

1. はじめに

情報通信技術の発展やオープンデータ施策の進展に伴い、社会経済現象の実態を詳細な時空間解像度で記録した地理空間データが蓄積されている。これらデータの分析を通じた地域モニタリングに基づく政策決定への活用が都市・地域計画分野における重要な研究課題の一つである。

さて、地域モニタリング上の注目事項の一つは、社会経済現象の生成要因が場所により異なることに着目した、空間的異質性分析である。社会経済現象には、空間全域で連続的に変化する異質性と特定

の地域で不連続的に変化する異質性の二種類が存在する。前者に対して、近接した地域ではパラメータが近い値として推定され、位置によって生成要因の影響が徐々に変化する空間構造を捉える Spatially Varying Coefficient (SVC) モデルが提案されている。地理的加重回帰モデル (Brunsdon et al. 1996; Fotheringham et al. 1997) や Bayesian SVC モデル (Gelfand et al. 2003), Eigenvector spatial filtering-based SVC (ESF-SVC) モデル (Griffith, 2008) はその代表的手法である。また後者に対しては、特定の地域区分に基づき各地域に異なる係数を設定したモデル

* 学生会員 東北大学 大学院情報科学研究科 (Tohoku University)

〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06 E-mail : peng.zhan.p3.@dc.tohoku.ac.jp

** 正会員 東北大学大学院情報科学研究科 (Tohoku University)

を用い、不連続に変化する空間構造を捉える分析が検討されてきた。例えば、Goodman and Thibodeau (2003) は小学校区を空間単位としてパラメータ推定を行った。また、Parker et al. (2016) はスパースモデリング手法を活用した特定の領域を抽出する手法を提案した。さらに、Inoue et al. (2020) は町丁目の単位で地理的市場分割分析を行った。

しかし、大都市圏のように多様な特徴を有する地域では、連続的・不連続的な空間的異質性が同時に存在する場合がある。不動産市場を例にとると、環境的・社会的な決定要因が住宅価格への影響が場所によって異なる。都心と郊外では、物件の広さや交通機関への近接性、都心への距離などの属性の影響は、大域的に連続的に変化する一方で、行政境界、鉄道沿線などのローカルな空間スケールでは、周辺と著しく違って物件が評価され、価格が不連続に変化する異質性が見られる。この場合には、単一種類の空間的異質性しかないとの仮定が適切ではない。従来の分析手法ではその片方しか考慮できないため、偏りのある推定結果に基づき地域特徴の解釈を誤る可能性がある。

連続的な空間的異質性と不連続的な空間的異質性を同時に検出するために、ESF-SVC とスパースモデリング手法の 1 つである Fused LASSO を融合したモデルが提案され、両者の空間的異質性を分離した推定ができることが示されている (田・井上, 2020)。また近年、ESF-SVC モデルに変量効果を導入した random effects ESF-SVC (RE-ESF-SVC) モデルに拡張され、連続的な空間的異質性の構造をより高い精度で分析できることが示されている (Murakami and Griffith, 2017)。そこで本研究は、より適切な連続的・不連続な空間的異質性構造の抽出を目指し、RE-ESF-SVC モデルと Fused LASSO の融合モデルを提案し、その効果を確認する。

以降、第 2 章で提案した融合モデルとその推定方法を示し、第 3 章ではシミュレーション実験で提案モデルの有用性を示し考察する。最後に、第 4 章において本研究の結論を示す。

2. 提案手法

2.1. RE-ESF-SVC モデル

RE-ESF-SVC の基本は、空間相関の検定統計量である Moran's I 統計量 (MC) に基づいた空間モデリングアプローチ Eigenvector spatial filtering (ESF) である (Griffith, 1996)。 N 地点で得られた空間データ \mathbf{z} ($N \times 1$ ベクトル) の空間相関を表す MC は、地点間の $N \times N$ の空間近接性行列を \mathbf{C} 、 $N \times N$ の centering matrix を $\mathbf{M} = \mathbf{I} - \mathbf{1}\mathbf{1}'/N$ 、 $N \times N$ 単位行列を \mathbf{I} 、すべての要素が 1 の $N \times 1$ のベクトルを $\mathbf{1}$ と表すと、式 (1) で定式化できる。

$$MC = \frac{\mathbf{z}'\mathbf{M}\mathbf{C}\mathbf{M}\mathbf{z}}{\mathbf{z}'\mathbf{M}\mathbf{z}} \quad (1)$$

MC の値は、 \mathbf{z} に正の空間的相関がある場合は正、負の空間的相関がある場合は負である。 $\mathbf{M}\mathbf{C}\mathbf{M}$ の各固有ベクトル \mathbf{e}_l は空間近接行列 \mathbf{C} の有する空間相関構造を表し、 \mathbf{e}_l に対応する固有値 λ_l は空間的相関のスケールを表す (Griffith, 2003)。固有ベクトルを線形モデルの説明変数として導入し、データの空間的相関を考慮する ESF モデルの基本式は式 (2) で与えられる。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \quad (2)$$

\mathbf{y} は被説明変数の $N \times 1$ のベクトル、 \mathbf{X} は切片と説明変数を含む $N \times K$ の行列、 \mathbf{E} は空間隣接行列の正の空間的相関を表す固有ベクトルの $N \times L$ ($L < N$) の行列、 $\boldsymbol{\beta}$ と $\boldsymbol{\gamma}$ はそれぞれ回帰係数の $K \times 1$ と $L \times 1$ のベクトルである。

$\boldsymbol{\beta}$ の空間的異質性を考慮するために、Griffith (2008) は ESF を拡張して ESF-SVC モデル (式(3)) を提案した。

$$\mathbf{y} = \sum_{k=1}^K \mathbf{x}_k \circ \boldsymbol{\beta}_k^{ESF} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \quad (3)$$

$$\boldsymbol{\beta}_k^{ESF} = \boldsymbol{\beta}_k\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_k$$

" \circ " は要素ごとの積、 $\boldsymbol{\beta}_k^{ESF}$ は k 番目の説明変数 \mathbf{x}_k の空間可変パラメータベクトルである。定数 $\boldsymbol{\beta}_k$ と $\mathbf{E}\boldsymbol{\gamma}_k$ はそれぞれ、 $\boldsymbol{\beta}_k^{ESF}$ の平均値と空間的変動を表す。ESF-SVC モデルは他の SVC モデルに比べ、変数ご

と異なるスケールの空間変動を考慮することができ、計算や拡張が容易であるという長所を持つ。

さらに、Murakami and Griffith (2017) は \mathbf{y}_k を変量効果とした仮定に基づき、ESF-SVC をより高い推定精度を持つ RE-ESF-SVC に拡張した。RE-ESF-SVC モデルは式 (4) で表される。

$$\begin{aligned} \mathbf{y} &= \sum_{k=1}^K \mathbf{x}_k \circ \boldsymbol{\beta}_k^{R-ESF} + \boldsymbol{\varepsilon}, & \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \\ \boldsymbol{\beta}_k^{R-ESF} &= \beta_k \mathbf{1} + \mathbf{E} \boldsymbol{\gamma}_k, \\ \boldsymbol{\gamma}_k &\sim N(\mathbf{0}_L, \boldsymbol{\Sigma}) \end{aligned} \quad (4)$$

$\boldsymbol{\gamma}_k$ は平均 $\mathbf{0}$ の正規分布に従う確率変数、その分散共分散行列 $\boldsymbol{\Sigma}$ は $KL \times KL$ の対角行列である。Murakami and Griffith (2017) は $\boldsymbol{\Sigma}$ の対角要素を $\sigma_{k(y)}^2 \lambda_l(\alpha_k) = \sigma_{k(y)}^2 (\sum_l \lambda_l / \sum_l \lambda_l^{\alpha_k}) \lambda_l^{\alpha_k}$ と定義する。 $\sigma_{k(y)}^2$ と α_k はハイパーパラメータである。

2.2. Fused LASSO

Fused LASSO は、正則化項を目的関数に加えて隣接した説明変数同士のパラメータを同じ値に推定するとともに、変数選択を行うスパースモデリング (sparse modeling) である (Tibshirani et al., 2005)。Fused LASSO の推定式は式 (5) で与えられる。

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \underset{\boldsymbol{\beta}}{\text{augmin}} \left[\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right. \\ &\quad \left. + \lambda \sum_{(m,n) \in F} |\beta_m - \beta_n| \right. \\ &\quad \left. + \delta \lambda \|\boldsymbol{\beta}\|_1 \right] \end{aligned} \quad (5)$$

第 2 項は隣接パラメータ対の集合 F に属するパラメータ $\beta_m, \beta_n ((m, n) \in F)$ の差に対する L_1 正則化項、第 3 項はパラメータ自体に対する L_1 正則化項であり、モデルの過学習を防ぐために変数選択を行う。なお λ, δ は正則化の強さを定めるハイパーパラメータである。

Fused LASSO を地域分析に活用する場合、分析地域を小地域に空間分割し、各小地域に設定したダミー変数のパラメータに対して隣接関係の L_1 正則化項を加えた推定を行うことで、パラメータが同じ非 0 値を取る周辺地域と異質な小地域群を抽出できる。

2.3. 融合モデル

以上を踏まえ、本研究は、RE-ESF-SVC と Fused LASSO を融合したモデルを提案する。融合モデルの基本式は式 (6) で表される。

$$\begin{aligned} \mathbf{y} &= \sum_{k=1}^K \mathbf{x}_k \circ \boldsymbol{\beta}_k^{Fusion} + \mathbf{X}^{area} \boldsymbol{\beta}^{area} + \boldsymbol{\varepsilon}, \\ \boldsymbol{\beta}_k^{Fusion} &= \beta_k \mathbf{1} + \mathbf{E} \boldsymbol{\gamma}_k^{Fusion}, \\ \boldsymbol{\gamma}_k^{Fusion} &\sim N(\mathbf{0}_L, \boldsymbol{\Sigma}), \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \quad (6)$$

連続的な空間的異質性を、変量効果を有する固有ベクトルを導入した回帰係数 $\boldsymbol{\beta}_k^{Fusion}$ で表現し、不連続な空間的異質性を、分割した小地域のダミー変数 \mathbf{X}^{area} のパラメータ $\boldsymbol{\beta}^{area}$ で表現する。

また、融合モデルは式 (7) と表せ、固定効果 $\boldsymbol{\beta}^{all}$ と変量効果 $\boldsymbol{\gamma}_k^{Fusion}$ の両方を有する線形混合モデル (linear mixed model) であることが分かる。

$$\begin{aligned} \mathbf{y} &= \mathbf{X}^{all} \boldsymbol{\beta}^{all} + \tilde{\mathbf{E}} \tilde{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}, \\ \mathbf{X}^{all} \boldsymbol{\beta}^{all} &= \sum_{k=1}^K \mathbf{x}_k \circ \beta_k + \mathbf{X}^{area} \boldsymbol{\beta}^{area}, \\ \tilde{\mathbf{E}} &= [\mathbf{x}_1 \circ \mathbf{E}, \dots, \mathbf{x}_K \circ \mathbf{E}], \\ \tilde{\boldsymbol{\gamma}} &= [\boldsymbol{\gamma}_1^{Fusion}, \dots, \boldsymbol{\gamma}_K^{Fusion}]', \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \end{aligned} \quad (7)$$

式 (7) の推定では、式 (8) の目的関数を最小化するパラメータを求める。

$$\begin{aligned} Q(\boldsymbol{\beta}^{all}, \boldsymbol{\Sigma}, \sigma^2) &= (\mathbf{y} - \mathbf{X}^{all} \boldsymbol{\beta}^{all})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}^{all} \boldsymbol{\beta}^{all}) \\ &\quad + \log |\mathbf{V}| + N \log (2\pi \sigma^2) \\ &\quad + \lambda \sum_{(m,n) \in F} |\beta_m^{area} - \beta_n^{area}| \\ &\quad + \delta \lambda \|\boldsymbol{\beta}^{area}\|_1 \end{aligned} \quad (8)$$

ただし、 \mathbf{V} は \mathbf{y} の総分散で、 $\mathbf{V} = \sigma^2 \mathbf{I} + \tilde{\mathbf{E}} \boldsymbol{\Sigma} \tilde{\mathbf{E}}'$ である。式 (8) の最初の 3 項は対数尤度関数の -1 倍、第 4 項は隣接する小地域パラメータ $\beta_m^{area}, \beta_n^{area} ((m, n) \in F)$ の差に対する L_1 正則化項、第 5 項は小地域パラメータ自体に対する L_1 正則化項、 λ と δ はハイパーパラメータである。

2.4. パラメータ推定

正則化項の導入により、式 (8) を直接最小化することは難しい。本研究では、Yu and Hong (2021) が

提案した 2 段階推定法 (two-step iteration algorithm) を用いてパラメータ推定を行う。融合モデルの推定の手順を以下に示す。

- (1) 変量効果を考慮せずに、Fused LASSO で固定効果 β^{all} の初期値 $\widehat{\beta}^{all(0)}$ を推定する (式(9)).

$$\widehat{\beta}^{all(0)} = \underset{\beta^{all}}{\text{augmin}} \left[(\mathbf{y} - \mathbf{X}^{all} \beta^{all})' (\mathbf{y} - \mathbf{X}^{all} \beta^{all}) + \lambda \sum_{(m,n) \in F} |\beta_m^{area} - \beta_n^{area}| + \delta \lambda \|\beta^{area}\|_1 \right] \quad (9)$$

- (2) 固定効果の初期値を与え、 $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}^{all} \widehat{\beta}^{all(0)}$ として得られる、式 (10) の変量効果モデルのパラメータを推定する。

$$\hat{\mathbf{u}} = \tilde{\mathbf{E}}\tilde{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad (10)$$

この推定量は以下の通りである。

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= (\tilde{\mathbf{E}}'\tilde{\mathbf{E}})^{-1} \tilde{\mathbf{E}}'\hat{\mathbf{u}} \\ \widehat{\sigma^2} &= \frac{(\hat{\mathbf{u}} - \tilde{\mathbf{E}}\hat{\boldsymbol{\gamma}})'(\hat{\mathbf{u}} - \tilde{\mathbf{E}}\hat{\boldsymbol{\gamma}})}{N - q}, \\ \hat{\boldsymbol{\Sigma}} &\approx \hat{\boldsymbol{\gamma}}\hat{\boldsymbol{\gamma}}', \\ \hat{\mathbf{V}} &= \widehat{\sigma^2} \mathbf{I} + \tilde{\mathbf{E}}\hat{\boldsymbol{\Sigma}}\tilde{\mathbf{E}}'. \end{aligned} \quad (11)$$

ただし、 N は地点数、 q は $\tilde{\mathbf{E}}$ の列数である。

- (3) $\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Sigma}}, \widehat{\sigma^2}$ から得られる $\hat{\mathbf{V}}^{-1}$ のコレスキー分解の結果 \mathbf{R} を用いて、式 (9) の \mathbf{y} と \mathbf{X}^{all} を

$$\begin{aligned} \mathbf{y}^R &= \mathbf{R}\mathbf{y} \\ \mathbf{X}^{allR} &= \mathbf{R}\mathbf{X}^{all} \end{aligned} \quad (12)$$

で代替し、Fused LASSO による推定で、固定効果を新たな推定値に更新する。

収束するまで (2)(3) を繰り返す、 $\widehat{\beta}^{all}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\Sigma}}$, $\widehat{\sigma^2}$ を求める。なお、正則化の強さを決めるハイパーパラメータ λ の値は、バイズ情報量規準に基づき決定する。

3. シミュレーション実験

本章では、シミュレーション実験で提案手法の性能評価を行う。まず、連続的・不連続な空間的異質性が同時に存在するシミュレーションデータを生成して、融合モデル、RE-ESF-SVC モデル、Fused

LASSO で推定し、融合モデルは異なる空間的異質性を同時に抽出できることを確認する。さらに、連続的な空間的異質性のみ存在するデータと不連続な空間的異質性のみ存在するデータを用い、どちらか一方の空間的異質性しか存在しないときの提案手法の有用性を評価する。

なお、Fused LASSO による推定は R の `genlasso` パッケージを使用する。

3.1. シミュレーションデータの生成方法

$[0, 1] \times [0, 1]$ の正方形領域内に、一様分布 $U(0, 1)$ で座標を与えて無作為に 2,000 点を生成する。

まず、連続的な空間的異質性を有するデータ生成の設定を記す。Murakami et al. (2017) と Dray et al. (2006) に基づき、空間近接行列 \mathbf{C} の (i, j) 番目の成分 $c_{i,j}$ を指数型の距離減衰関数 (式(13)) で計算する。

$$c_{i,j} = \begin{cases} \exp(-d_{i,j}/r) & i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

$d_{i,j}$ は地点 $i \cdot j$ 間のユークリッド距離、レンジパラメーター r は全ての地点を結ぶ最小全域木の最長距離で与えた。MCM から正の空間相関を表す固有ベクトルを抽出し、さらに Griffith (2003) に従い、第 1 固有値の Moran's I 統計量との比が 0.25 以上である固有ベクトルを用いた。切片、および、説明変数として $\mathbf{x}_1 = \mathbf{1}, \mathbf{x}_2 \sim U(0, 1)$ と設定した。連続的な空間的異質性を表現する空間可変係数 β_1, β_2 を式 (14) で与える。

$$\begin{aligned} \beta_1 &= \mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_1, \quad \boldsymbol{\gamma}_1 \sim N(\mathbf{0}, 0.01\mathbf{I}) \\ \beta_2 &= (2)\mathbf{1} + \mathbf{E}\boldsymbol{\gamma}_2, \quad \boldsymbol{\gamma}_2 \sim N(\mathbf{0}, 0.01\mathbf{I}) \end{aligned} \quad (14)$$

次に、不連続な空間的異質性を有するデータ生成の設定を記す。正方領域を 10×10 の小地域に分割し、各地域に付与したダミー変数を \mathbf{X}^{area} として導入する。そして特定の小地域のダミー変数パラメータを非ゼロに設定し、周辺の地域と異なる特徴を持つ不連続な空間的異質性を発生させる。ここで、 β^{area} の 45, 46, 55, 56 番目の要素を 1, 他を 0 と設定する。

また、正規分布 $N(\mathbf{0}, \sigma^2 \mathbf{I})$ に従う攪乱項 $\boldsymbol{\varepsilon}$ を生成する。

以上の設定を基に、連続的・不連続な空間的異質

性が同時に存在するデータ \mathbf{y}_{GL} , 連続的な空間的異質性のみ存在するデータ \mathbf{y}_G , 不連続な空間的異質性のみ存在するデータ \mathbf{y}_L をそれぞれ式 (15) で生成する.

$$\begin{aligned} \mathbf{y}_G &= \mathbf{x}_1 \circ \boldsymbol{\beta}_1 + \mathbf{x}_2 \circ \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \\ \mathbf{y}_{GL} &= \mathbf{x}_1 \circ \boldsymbol{\beta}_1 + \mathbf{x}_2 \circ \boldsymbol{\beta}_2 + \mathbf{X}^{area} \boldsymbol{\beta}^{area} + \boldsymbol{\varepsilon} \\ \mathbf{y}_L &= \mathbf{X}^{area} \boldsymbol{\beta}^{area} + \boldsymbol{\varepsilon} \end{aligned} \quad (15)$$

3.2. シミュレーション実験の結果

提案モデルと Fused LASSO のハイパーパラメータ δ を 1 と設定し, 各データに対する推定は 50 回ずつ行う. まず, 各パラメータの設定値と推定値の空間分布に注目しモデルを比較する. また, モデルの推定精度は, 切片 $\boldsymbol{\beta}_1$, 回帰係数 $\boldsymbol{\beta}_2$ と小地域ダミーパラメータ $\boldsymbol{\beta}^{area}$ の設定値と推定値との間の RMSE (平均二乗偏差), Adjusted R^2 (自由度調整済み決定係数), 残差の Moran's I 統計量と BIC で比較する.

(1) パラメータ空間分布の考察

図 3.1 は, 50 回目の実験における 3 手法で推定したパラメータの空間分布を示す. 図 3.1 より, 3 種類のデータを用いた各実験において, 提案モデルは, 生成データとほぼ同じ空間分布を有するパラメータ推定結果を示す.

連続的・不連続な空間的異質性が同時に存在するデータに対して, 提案モデルでは, 2 種類の空間的異質性を同時に抽出でき, それぞれ切片と小地域ダミーのパラメータで表す. 一方, 連続的な空間的異質性のみ抽出できる RE-ESF-SVC モデルでは, 切片の推定結果に不連続な空間的異質性が混在している. 不連続な空間的異質性のみ抽出できる Fused LASSO は, $\boldsymbol{\beta}_1$ と $\boldsymbol{\beta}_2$ を場所に依存しない共通の値として推定し, 空間的に不連続な小地域ダミーのパラメータで連続的な空間的異質性を表そうとしていることが分かる. この結果から, 既存の二つのモデルは, 連続的・不連続な空間的異質性が同時に存在するデータに対して, 適切にその構造を推定できないことが確認できた.

連続的な空間的異質性のみ存在するデータに対

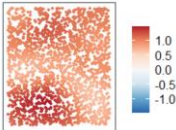
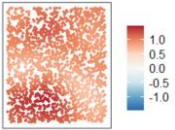
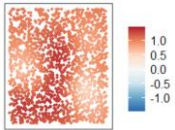
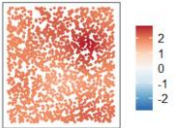
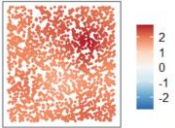
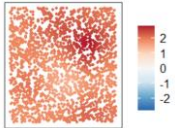
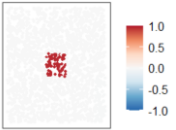
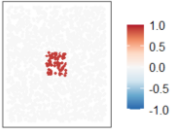
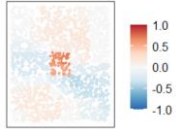
して, 融合モデルで推定したパラメータ $\boldsymbol{\beta}_1$ と $\boldsymbol{\beta}_2$ は, RE-ESF-SVC モデルとほぼ同じ空間分布を示す. また, 提案モデルは小地域ダミーのパラメータが 0 と推定する. これにより, 融合モデルは不連続な空間的異質性を抽出せず, 連続的な空間的異質性だけを抽出できることが確認できる. 一方, Fused LASSO は $\boldsymbol{\beta}_1$ と $\boldsymbol{\beta}_2$ を定数と推定し, 小地域ダミーのパラメータ推定に連続的な空間的異質性が見られ, 適切に推定できないことが明らかとなった.

不連続な空間的異質性のみ存在するデータに対して, 融合モデルで推定したパラメータ $\boldsymbol{\beta}_1$ と $\boldsymbol{\beta}_2$ は, 場所によって変化する様子が見られない一方, 小地域ダミーのパラメータはデータ生成パラメータと同様の分布を示す. この結果から, 連続的な空間的異質性が存在しない場合は, 提案モデルは Fused LASSO と同様に, 不連続な空間的異質性だけを抽出できることが確認された. 一方, RE-ESF-SVC モデルの切片 $\boldsymbol{\beta}_1$ の推定結果から見ると, 局地的な空間パターンがあり, RE-ESF-SVC モデルは不連続な空間的異質性を適切に表現できないことが確認できた.

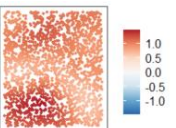
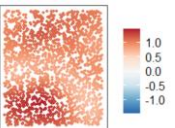
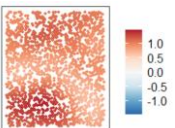
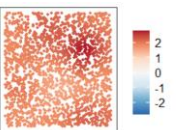
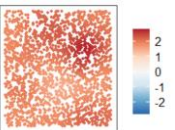
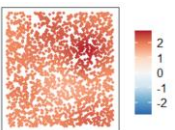
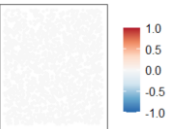
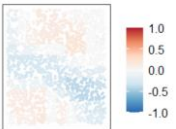
(2) 推定統計量の考察

表 3.2 は各モデルの統計量を示す. 表より, 連続的・不連続な空間的異質性が同時に存在する場合, 他のモデルに比べ, 提案モデルのパラメータ推定の RMSE と BIC が最も小さく, Adjusted R^2 が最も大きく, 高い汎化性能と説明力を示す. また, 残差の Moran's I 統計量が 0 に近く, 提案モデルは空間相関を十分に考慮できることが確認できる. 以上から, 提案モデルは異なる空間的異質性を同時に抽出できることが確認された.

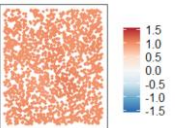
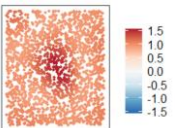
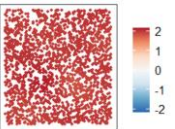
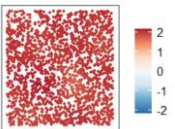
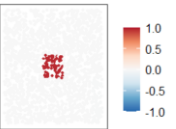
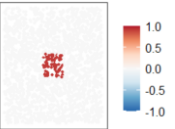
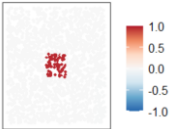
提案モデルは, 連続的な空間的異質性のみ存在する場合には RE-ESF-SVC モデルと, 不連続な空間的異質性のみ存在する場合には Fused LASSO と同等の推定精度を示す. どちらか一方の空間的異質性しか存在しないデータに対しても, 提案手法は高い推定精度を持つことも確認された.

	データ生成パラメータ	推定パラメータ		
		融合モデル	RE-ESF-SVC	Fused LASSO
β_1				定数 1.036
β_2				定数 2.006
β^{area}			なし	

(a) 連続的・不連続な空間的異質性が同時に存在するデータ

	データ生成パラメータ	推定パラメータ		
		融合モデル	RE-ESF-SVC	Fused LASSO
β_1				定数 1.066
β_2				定数 1.992
β^{area}	なし		なし	

(b) 不連続な空間的異質性のみ存在するデータ

	データ生成パラメータ	推定パラメータ		
		融合モデル	RE-ESF-SVC	Fused LASSO
β_1	定数 1			定数 0.998
β_2	定数 2			定数 2.007
β^{area}			なし	

(c) 連続的な空間的異質性のみ存在するデータ

図 3.1 三種類のデータにおける各モデルのパラメータ空間分布

表 3.1 三種類のデータにおける各モデルの統計量

データ	統計量	融合モデル	RE-ESF-SVC	Fused LASSO
連続・不連続	β_1 の RMSE	0.023	0.142	0.189
	β_2 の RMSE	0.034	0.054	0.184
	β^{area} の RMSE	0.017	NA	0.164
	Adjusted R ²	0.977	0.934	0.951
	Moran's I	-0.004	0.194	0.194
	BIC	-3561.38	-1473.41	-1702.70
連続のみ	β_1 の RMSE	0.021	0.021	0.188
	β_2 の RMSE	0.035	0.035	0.184
	β^{area} の RMSE	0.044	NA	0.259
	Adjusted R ²	0.975	0.975	0.948
	Moran's I	-0.008	-0.007	0.186
	BIC	-3533.76	-3571.55	-1568.39
不連続のみ	β_1 の RMSE	0.023	0.142	0.004
	β_2 の RMSE	0.033	0.055	0.006
	β^{area} の RMSE	0.017	NA	0.008
	Adjusted R ²	0.974	0.926	0.974
	Moran's I	-0.005	0.194	0.002
	BIC	-3486.03	-1518.85	-3485.97

4. おわりに

本研究は、RE-ESF-SVC と Fused LASSO を用いて、連続的・不連続な空間的異質性を同時に抽出した融合モデルを提案した。提案手法は RE-ESF-SVC モデルの可変係数で連続的に変化する空間構造を捉えた上で、小地域の隣接関係に L_1 正則化項を加え、特定の地域で発生する不連続的な空間構造を抽出できる。

シミュレーション実験で提案手法の有用性が確認された。単一種類の空間的異質性のみ抽出できるモデルに比べ、提案した融合モデルは、連続的・不連続な空間的異質性が同時に抽出できるだけでなく、連続的な空間的異質性のみ存在するデータと不連続な空間的異質性のみ存在するデータに対しても、高い推定性能を示す。今後については、提案手法を実データに適用して、実際の地理空間に存在

する空間的異質性を分析する。

謝辞

本研究は、JSPS 科研費 18H01552 および 21H01447 の助成を受けた。

参考文献

- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, **28**(4), 281-298.
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological modelling*, **196**(3-4), 483-493.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M.

- (2003). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
- Gelfand, A. E., Kim, H. J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, **98**(462), 387-396.
- Goodman, A. C., & Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, **12**(3), 181-201.
- Griffith, D. A. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo - referenced data. *Canadian Geographer/Le Géographe canadien*, **40**(4), 351-367.
- Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin: Springer-Verlag.
- Griffith, D. A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A*, **40**(11), 2751-2769.
- Inoue, R., Ishiyama, R., & Sugiura, A. (2020). Identifying local differences with fused-MCP: an apartment rental market case study on geographical segmentation detection. *Japanese Journal of Statistics and Data Science*, 1-32.
- Murakami, D., Yoshida, T., Seya, H., Griffith, D. A., & Yamagata, Y. (2017). A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*, **19**, 68-89.
- Parker, R. J., Reich, B. J., & Eidsvik, J. (2016). A fused lasso approach to nonstationary spatial covariance estimation. *Journal of agricultural, biological, and environmental statistics*, **21**(3), 569-587.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(1), 91-108.
- Yu, J., & Zhong, H. (2021). Time varying mixed effects model with fused lasso regularization. *Journal of Applied Statistics*, **48**(8), 1513-1526.
- 田 皓一朗, 井上 亮. (2020). 異なる空間スケールの異質性を考慮可能な分析法の提案~空間可変パラメータモデルとスパースモデリングの融合アプローチ. 地理情報システム学会 学術研究発表大会講演集, **29**: C24-1-3. CD-ROM.