

固有ベクトル空間フィルタリングを用いた 時空間内挿

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 システム情報系社会工学域 (〒305-8573 茨城県つくば市天王台1-1-1)
E-mail: tsutsumi@sk.tsukuba.ac.jp

空間データの特性を明示的に考慮する比較的新しいアプローチに固有ベクトル空間フィルタリングがある。これは、データの空間パターンを近接行列の固有ベクトルを用いて説明しようというアプローチであり、実用的である点や、複数のスケールの空間パターンを考慮できる点で有用である。しかしながら、その時空間データへの適用は極めて限定的である。一方で、近年、状態空間モデルを用いて時系列データをモデル化するアプローチが気象学を中心に注目されている。以上を踏まえ、本研究では、固有ベクトル空間フィルタリングと状態空間モデルを組み合わせることで新たな時空間モデルを構築した。この手法は、長方形の空間を仮定しなければいけない点などの課題があり、必ずしも完全ではないものの、そのモデルはよく用いられる状態空間モデルの一形態と同一となり、また、状態空間モデルに基づいて時空間を考慮したその他の手法とは異なりパラメータ推定のための繰り返し計算が必ずしも必要とはならない点で実用的といえる。構築されたモデルの精度は、茨城県内の地価の時空間内挿を行うことで検証した。

Key Words : *Eigenvector spatial filtering, State space model, Land price, Kalman filter, Multicollinearity*

1. はじめに

GIS (Geographic Information System) や GPS (Global Positioning System) の発展に伴う時空間データの多様化は著しく、それらを適正に取り扱うことのできるモデルの構築は(時)空間統計学(Geostatistics; Cressie, 1993; Cressie and Wikle, 2011)の重要な研究テーマの一つである。

時空間統計学に基づいて時空間を考慮するアプローチは、(i)空間(ここでは2次元と仮定)と時間(1次元)からなる3次元空間上でデータをモデル化するアプローチと、(ii)時間経過に伴う空間の状態遷移をモデル化するアプローチに分類できる(Cressie et al., 2010)。例えば、共分散を時空間上の距離の関数で与えることでデータ相互の依存関係をモデル化する時空間kriging (e.g., Cressie and Huang, 1999)は(i)の代表的な手法である。一方で、空間統計モデルと状態空間モデル(詳しくは第3章)を組み合わせたKriged Kalman filter (Mardia et al., 1998)やSpace-time Kalman filter (Wikle and Cressie, 1999)は(ii)に基づいた代表的なアプローチである。(ii)に基づいた各手法は、データのダイナミクスを明示的にモデル化できる点、データ同化(Data

assumulation; e.g., 樋口ら, 2011)に基づいた柔軟なモデルが構築できる点、計算負荷が小さい点などで(i)の手法よりも優れており(Douglas, 2010)、時空間統計学のみならず、自然科学の幅広い分野で着目されている(樋口ら, 2011)。以上を踏まえ、本研究では時間経過に伴う状態遷移をモデル化する(ii)のアプローチに着目することとする。

空間を考慮する比較的新しいアプローチにSpatial Eigenvector Mappingがある。これは、近接行列の固有ベクトルを用いてデータの空間パターンを表現しようというアプローチであり、例えば、Trend surface modelに基づいたPCNM (Principal Coordinates of Neighbor Matrices; Dray et al, 2006)や、空間統計モデルに基づいたESF (Eigenvector Spatial Filtering: 固有ベクトル空間フィルタリング; e.g., Griffith, 2003)とECSF (Eigenvector Continuous Space Filtering; 村上, 2012 (Appendix.1 参照))がある。ここで、PCNMとESFは離散空間上のデータ、ECSFは連続空間上のデータをモデル化する手法である。各手法のモデルはいずれも従来の線形回帰モデルと同一であり、拡張が容易であるものの、その時空間への拡張を行った研究は極めて限定的である。これは、あまり知られていないアプ

ローチであることや、(i)の方法で拡張した場合にモデルに投入すべき固有ベクトルの数が膨大となり、自由度の損失と計算量が増大が問題となることが原因と考えられる。一方で、(ii)の方法に基づいた Moran Eigenvector Mappingの拡張は、自由度の節約や計算量の節約の観点からは非常に有効である（詳しくは4章2節）。

一般に、状態空間モデルは連続空間上のデータの予測で成果を挙げてきた。従って、本研究でも、離散空間を前提としたESFではなく、連続空間を前提としたECSFを状態空間モデルに基づいて拡張することで、新たな時空間モデルを構築することとする。

2. 固有ベクトル空間フィルタリング:ESF/ECSF

2.1. モラン統計量

空間的相関の検定統計量にモランI 統計量(1)式 (Moran, 1948) がある。

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

ここで、 n はデータ数、 I_{moran} はモラン I 統計量、 \mathbf{y} は空間的相関を計測する変数の $n \times 1$ 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 $n \times n$ の近接行列（ここでは隣接行列とする）である。 I_{moran} は -1 から 1 までの値をとり、1 に近いことは \mathbf{y} が強い正の空間的相関を持つことを、-1 に近いことは \mathbf{y} が強い負の空間的相関を持つことを、それぞれ表わす。モラン I 統計量は近接性を $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ で表現する統計量とみなすことができる。

ESFあるいはECSFは、隣接行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ の固有ベクトル(2)式を用いて、空間的相関に起因した、空間データの空間的なパターン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の演算子である。 \mathbf{E}_1 は、最大の固有値を持つ固有ベクトルであり、モラン I 統計量で説明可能な最も大域的な空間パターンを表す。 \mathbf{E}_2 は、2番目に大きな固有値を持つ固有ベクトルであり、 \mathbf{E}_1 と無相関、かつモラン I 統計量で説明可能な最も大域的な空間パターンを表す。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相関、かつモラン I 統計量で説明される最も大域的な空間パターンを表す。

2.2. 隣接行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ の特性

いま、 X 軸方向に P 個、 Y 軸方向に Q 個のメッシュが敷き詰められた長方形の空間 (図1) を考える。 $X \cdot Y$ 軸方向のメッシュの添え字をそれぞれ $p = \{1, 2, \dots, P\}$, $q = \{1, 2, \dots, Q\}$ とすれば、隣接行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ の固有ベクトル $\mathbf{E}_{p,q}$ は、 p, q の取り方に応じて(3)式で与えられる (Griffith, 2000)。

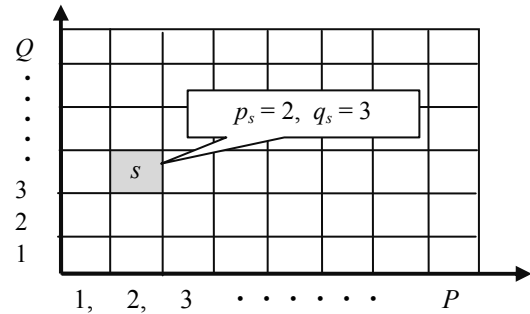


図1 メッシュの敷き詰められた空間

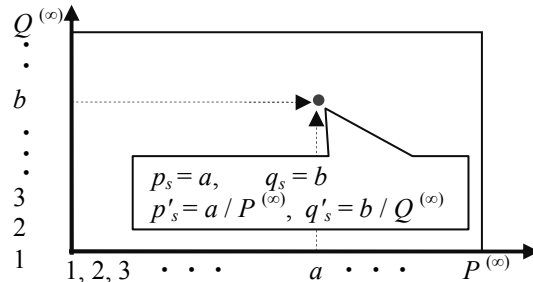


図2 メッシュの極限まで細分化された空間

$$\mathbf{E}_{p,q} \propto \begin{cases} (1/\sqrt{n})\mathbf{1} & \text{if } p = q = 1 \\ k_{p,q}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{E}_{p,q}^* & \text{otherwise} \end{cases} \quad (3)$$

\propto は比例の演算子、 $\mathbf{1}$ は 1 を要素に持つ $n \times 1$ のベクトル、 $\mathbf{E}_{p,q}^*$ は p, q に対応した、行列 \mathbf{C} の固有ベクトル ($n \times 1$)、 $k_{p,q}$ は $\mathbf{E}_{p,q}^*$ の要素の標準偏差である。 $\mathbf{E}_{p,q}^*$ の第 s 要素は(4)式で与えられる。

$$E_{s,p,q}^* = \frac{2}{\sqrt{(P+1)(Q+1)}} \sin\left(\frac{p_s p \pi}{P+1}\right) \sin\left(\frac{q_s q \pi}{Q+1}\right) \quad (4)$$

$p_s = \{1, 2, \dots, P\}$ 及び $q_s = \{1, 2, \dots, Q\}$ は第 s メッシュの X 軸及び Y 軸方向のメッシュ番号を表す。例えば、地点 s の $X \cdot Y$ 方向の各メッシュ番号が 2, 3 であれば、 p, q に対応した固有ベクトル $\mathbf{E}_{p,q}^*$ の第 s 要素は $p_s = 2, q_s = 3$ とした $E_{s,p,q}^*$ (4)式で与えられる。

ここで、 $X \cdot Y$ 軸方向の各メッシュの数 P, Q を極限まで大きくすることで連続空間を定義すると、同空間内で定義される各極小メッシュについての隣接行列 $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ が定義でき、その固有ベクトル $\mathbf{E}_{p,q}^{(\infty)}$ の第 s 要素は(5)式で与えられる (村上, 2012)。

$$E_{s,p,q}^{(\infty)} \propto \begin{cases} (1/\sqrt{\infty})\mathbf{1} & \text{if } p = q = 1 \\ \sin(p'_s p \pi) \sin(q'_s q \pi) & \text{otherwise} \end{cases} \quad (5)$$

ここで、 p'_s は、長方形空間の、 X 軸方向の総長に占める、 X 軸の端点から地点 s の X 座標までの長さの割合、 q'_s は Y 軸方向の同割合を表わす。 $\mathbf{E}_{p,q}^{(\infty)}$ の各固有値 $\lambda_{p,q}^{(\infty)}$ は(6)式で表現できる。

$$\lambda_{p,q}^{(\infty)} \propto -\frac{\pi^2}{2} \left(p^2 + q^2 \frac{P^{(\infty)}}{Q^{(\infty)}} \right) \quad (6)$$

$P^{(\infty)} / Q^{(\infty)}$ = (X 軸の総長) / (Y 軸の総長) である。

2.3.ESF/ECSFの基本モデル

ESFあるいはECSFの基本式は、ともにデータの空間パターンを固有ベクトルの線形和で表現した(7)式で与えられる。

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (7)$$

\mathbf{X} は $n \times k$ の説明変数行列、 $\boldsymbol{\varepsilon}$ は $n \times 1$ の攪乱項ベクトル、 $\boldsymbol{\beta}$ ($k \times 1$)、 $\boldsymbol{\gamma}$ ($r \times 1$) はパラメータベクトル、 σ^2 はパラメータ (スカラー) を表わす。 \mathbf{E} は近接行列の固有ベクトルを要素に持つ $n \times r$ の行列であり、ESFであれば、 \mathbf{E} の中の r 個の固有ベクトルは、モランI統計量(1)式が0.25以上を示す各固有ベクトル(2)式の中から変数選択により決定する。一方で、ECSFであれば、固有ベクトル (各要素は(5)式で与えられる) の中から、モランI統計量が0.25以上というESFの考え方の類推に基づいて導出される閾値である、大きな固有値を持つ上位 $0.4196 \times n$ 個の固有ベクトルの中から、変数選択により決定する。

ECSFは分析結果が長方形の空間の与え方に依存する点でESFに劣る。また、ECSFで用いる各固有ベクトルは互いに完全に無相関となるとは限らず、この点でも、完全に無相関な固有ベクトルを与えるESFに劣る。一方で、ECSFは、ESFとは異なり連続空間を考慮できる。また、ESFとは異なり固有ベクトル算出のために特性方程式の計算が不要であり、ESFひいてはMoran Eigenvector Mappingの短所として指摘されることの多い計算量の問題 (e.g., Griffith and Peres-Neto, 2006; Dray et al., 2006) を緩和している。さらに、ECSFは行列演算が不要であるため、中規模程度までのデータであればExcel等を用いて実装することもできる。即ち、ECSFは実用に特化した空間モデルといえる。

2.4.ESF/ECSFを用いた(時)空間内挿の先行研究

ESFあるいはECSFを空間内挿に応用した研究はわずかに存在する。ESFを空間内挿に応用した研究にGriffith (2010b; 2010c)がある。それらは、いずれも不完全データの尤度を最大化することのできるアルゴリズムであるEM (Expectation -Maximization) アルゴリズムを用いた繰り返し計算を行うことで離散空間上のデータを内挿した。Griffith (2010b)は、このアプローチの精度が、空間内挿への幅広い適用のみられる空間統計モデルの精度と同等であることを確認した。一方で、村上 (2012)はECSFを連続空間上のデータ内挿に応用し、代表的な空間統計モデルであるKriging (e.g., Cressie, 1993) を相当程度上回る精度が得られることを確認した。

ESFを時空間内挿に応用した研究に、2次元空間上の隣接行列の固有ベクトル \mathbf{E} と、(9)式で与えられ

る時間軸上の隣接行列の固有ベクトル \mathbf{E}_t を用いて、3次元時空間上で定義された隣接行列の固有ベクトル \mathbf{E}_{st} を用いるGriffith (2010a)の手法がある。

$$\mathbf{E}_{st} = \mathbf{E} \otimes \mathbf{E}_t \quad (8)$$

$$E_{i,r,(t)} = \cos\left(\frac{t(i)}{T} r \pi\right) \quad (9)$$

ここで、 $t\{=1,2,\dots,T\}$ は時点、 i はデータの添え字、 t_i は i 番目のデータの時点である。また、 $r=\{1,2,\dots,T\}$ は固有ベクトル \mathbf{E}_t の持つ固有値の大きさの順位を表す。 $E_{i,r,(t)}$ は時間軸上で定義された隣接行列の第 r 固有ベクトルの第 i 要素であり、 \mathbf{E}_t の i 行 r 列の要素を表す。彼は、この手法を (空間的に離散なデータの) 時空間データの内挿に応用し、一定の内挿精度で内挿が行われることを確認した。ここで、この方法は第1章で議論したアプローチの(i)に分類される。

一方で、ECSFを時空間に拡張した手法は存在しないものの、Griffith (2010a)と同様の方法で容易に時空間に拡張できる。具体的には、ECSFが仮定する2次元空間上の固有ベクトル(5)式と、(9)式に基づいて与えられる時間軸上の隣接行列の固有ベクトルのクロネッカー積をとることで3次元時空間上の固有ベクトルは(10)式で定義できる。

$$E_{p,q,r}^{(\infty)} \propto \begin{cases} (1/\sqrt{\infty})1 & p = q = r = 1 \\ \sin(p'_i r \pi) \sin(q'_j r \pi) \cos(r'_k r \pi) & \text{otherwise} \end{cases} \quad (10)$$

ここで $r'_i = t(i)/T$ である。(10)式で要素の与えられた固有ベクトルを(7)式に代入することで、ECSFを時空間に拡張できる。

しかしながら、(i)のアプローチに基づいたこの方法、は第1章で議論した(i)に基づいたアプローチの短所に加え、時空間統計学の代表的な内挿手法である時空間krigingなどで考慮される、時間に依存しない空間パターンや空間 (位置) に依存しない時間パターンが考慮できないという短所を持つ。即ち、(10)式で考慮できるのは時間と空間の両方に依存する時空間パターンのみである。これは、(10)式2段目の、空間パターンを表す最初の2つの三角関数は位置に応じて、時間軸上のパターンを表す最後の三角関数は時点に応じて、それぞれ変化するためである。このことより、このアプローチを用いた時空間内挿の精度は低くなる可能性がある。実際にこのアプローチをつくば市の地価の時空間内挿に適用したところ、その精度は通常最小二乗法 (そもそも内挿手法ではないが) と同等となってしまった。

従って、以降では、方法(i)ではなく、状態空間モデルを用いる方法(ii)に基づいてECSFを時空間に拡張する方法について議論する。

3. 状態空間モデル

3.1. モデル

状態空間モデル (State Space Model : e.g., 谷崎, 1993 ; 樋口ら, 2011) とは、時点 $t\{=1,2,\dots,T\}$ に n_t 個の標本持つデータ (総標本数 : $\sum_t n_t$) を表現するモデルであり、観測方程式(11)式と状態方程式(12)式、及び状態方程式の初期状態を表わす(13)式から成る (ビショップ, 2007)。

$$\mathbf{y}_t = \mathbf{A}\mathbf{z}_t + \mathbf{u}_t \quad \mathbf{u}_t \sim N(0, \mathbf{U}) \quad (11)$$

$$\mathbf{z}_t = \mathbf{B}\mathbf{z}_{t-1} + \mathbf{v}_t \quad \mathbf{v}_t \sim N(0, \mathbf{V}) \quad (12)$$

$$\mathbf{z}_1 = \mathbf{z}_0 + \mathbf{v}_0 \quad \mathbf{v}_0 \sim N(0, \mathbf{V}_0) \quad (13)$$

但し、ここではモデル内の各攪乱項に正規分布を仮定した。 $t\{=1,2,\dots,T\}$ は時点、 \mathbf{y}_t は時点 t に観測された被説明変数のベクトル ($n_t \times 1$)、 \mathbf{z}_t は \mathbf{y}_t の潜在的な状態を表わす非観測の変数 (潜在変数) のベクトル ($S \times 1$)、 \mathbf{u}_t は観測方程式の攪乱項ベクトル ($n_t \times 1$)、 \mathbf{v}_t は状態方程式の攪乱項ベクトル ($S \times 1$)、 \mathbf{z}_0 は潜在変数 \mathbf{z}_t の初期状態を表わすベクトル ($S \times 1$)、 \mathbf{v}_0 は \mathbf{z}_0 の攪乱を表わすベクトル ($S \times 1$) をそれぞれ表わす。 \mathbf{A} ($S \times n_t$) と \mathbf{B} ($S \times S$) は \mathbf{z}_t の \mathbf{y}_t への影響、 \mathbf{z}_t の \mathbf{z}_{t-1} への影響を表す行列である。 \mathbf{u}_t ($S \times 1$)、 \mathbf{v}_t ($T \times 1$)、 \mathbf{v}_0 ($T \times 1$) は正規分布に従う攪乱項であり、各々の分散共分散行列は、それぞれ \mathbf{U} ($S \times S$)、 \mathbf{V} ($T \times T$)、 \mathbf{V}_0 ($T \times T$) に従う。ここで、(11)~(13)式内 \mathbf{A} 、 \mathbf{B} 、 \mathbf{z}_0 、 \mathbf{U} 、 \mathbf{V} 、 \mathbf{V}_0 は尤度最大化に基づいた推定により、 \mathbf{z}_t はカルマンフィルタ (後述) と呼ばれる逐次更新により、それぞれ内生的に与えることができる。

上に示した状態空間モデルは、尤度最大化と潜在変数 \mathbf{z}_t の更新を収束するまで繰り返さなければならないため、実際には、(11)~(13)式を簡略化した次のモデルが用いられることも多い (川口, 1991)。

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{z}_t + \mathbf{u}_t \quad \mathbf{u}_t \sim N(0, \sigma_u^2 \mathbf{I}) \quad (14)$$

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \mathbf{v}_t \quad \mathbf{v}_t \sim N(0, \mathbf{V}) \quad (15)$$

$$\mathbf{z}_1 = \mathbf{0} + \mathbf{v}_0 \quad \mathbf{v}_0 \sim N(0, \sigma_{v0}^2 \mathbf{I}) \quad (16)$$

ここで σ_u^2 と σ_{v0}^2 はパラメータ (スカラー)、 $\mathbf{0}$ ($S \times 1$) はゼロベクトルである。 \mathbf{A}_t 、 σ_u^2 を所与とすれば、上述の繰り返し計算なしに各時点の潜在変数 \mathbf{z}_t を求めることができるため、(14)~(16)式から成る状態空間モデルは(11)~(13)式から成るものよりも実用的といえる。既述のようにECSFもまた実用に特化した手法であることから、両モデルの結合は、時空間を考慮した実用的な状態空間モデルの構築に有効である可能性がある。従って、以降では(14)~(16)式から成る状態空間モデル (及びECSF) に焦点を絞ることとする。

3.2. 潜在変数 \mathbf{z}_t の予測

t 期の潜在変数が $t-1$ 期より前の潜在変数と独立

であるというマルコフ性を仮定した場合、 $t-1$ 期の潜在変数 \mathbf{z}_{t-1} で条件付けられた、時点 t の潜在変数の期待値 $\bar{\mathbf{z}}_{t|t-1}$ と分散共分散行列 $\bar{\mathbf{V}}_{t|t-1}$ は、それぞれ(17)、(18)式で与えられる。

$$\bar{\mathbf{z}}_{t|t-1} = \mathbf{z}_{t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{A}_t \mathbf{z}_{t-1}) \quad (17)$$

$$\bar{\mathbf{V}}_{t|t-1} = \mathbf{V}_{t-1} - \mathbf{K}_t \mathbf{A}_t \mathbf{V}_{t-1} \quad (18)$$

$$\mathbf{K}_t = \mathbf{V}_{t-1} \mathbf{A}_t' (\mathbf{A}_t \mathbf{V}_{t-1} \mathbf{A}_t' + \sigma_u^2 \mathbf{I})^{-1} \quad (19)$$

ここで、 \mathbf{I} は $S \times S$ の単位行列である。(17)、(18)式を用いて $\bar{\mathbf{z}}_{t|t-1}$ と $\bar{\mathbf{V}}_{t|t-1}$ を逐次更新していくことで、各期の潜在変数 \mathbf{z}_t の予測量が $\bar{\mathbf{z}}_{t|t-1}$ で与えられる。以上のような潜在変数を逐次更新するアルゴリズムはカルマンフィルタ (Kalman filter ; Kalman, 1960) と呼ばれる。

4. ECSFの状態空間モデルに基づいた拡張

4.1. モデル

(14)~(16)式からなる状態空間モデルとECSFの基本式(7)式を組み合わせることで、次のモデルが定義できる。

$$\mathbf{y}_t = \tilde{\mathbf{A}} \tilde{\mathbf{z}}_t + \mathbf{u}_t \quad \mathbf{u}_t \sim N(0, \sigma_u^2 \mathbf{I}) \quad (20)$$

$$\tilde{\mathbf{z}}_t = \tilde{\mathbf{z}}_{t-1} + \mathbf{v}_t \quad \mathbf{v}_t \sim N(0, \mathbf{V}) \quad (21)$$

$$\tilde{\mathbf{z}}_1 = \mathbf{0} + \mathbf{v}_0 \quad \mathbf{v}_0 \sim N(0, \sigma_{v0}^2 \mathbf{I}) \quad (22)$$

$$\tilde{\mathbf{A}} = [\mathbf{X}_t, \mathbf{E}_t], \quad \tilde{\mathbf{z}}_t = \begin{bmatrix} \boldsymbol{\beta}_t \\ \boldsymbol{\gamma}_t \end{bmatrix}$$

\mathbf{X}_t は時点 t に観測された説明変数から成る $n_t \times S_1$ の行列、 \mathbf{E}_t は時点 t の各観測地点について、(5)式に基づいて与えられた固有ベクトルから成る $n_t \times S_2$ の行列である。ここで、 $S_1 + S_2 = S$ である。また、 $\boldsymbol{\beta}_t$ ($S_1 \times 1$) と $\boldsymbol{\gamma}_t$ ($S_2 \times 1$) はパラメータベクトルである。

通常のECSFに従えば、(20)式内の \mathbf{E}_t は、 $0.4196 \times n_t$ 個の固有ベクトルを並べた (即ち $S_2 = 0.4196 \times n_t$ とした) 行列で定義する必要がある。ここで、通常のECSFと同様、 \mathbf{E}_t を構成する各固有ベクトルが、互いに完全に無相関となるとは限らない。従って、上のモデルを実装するにあたっては多重共線性に配慮する必要がある。

\mathbf{A} ($\tilde{\mathbf{A}}$) の中の変数間に多重共線性がある場合、状態空間モデルを用いた潜在変数 \mathbf{z}_t ($\tilde{\mathbf{z}}_t$) の予測結果が不安定となることが知られている (Jazwinski, 1970)。Watson and Augustine (1983)によると、(14)~(16)式から成る状態空間モデルであれば、 σ_{v0}^2 を小さな値で与えることで、多重共線性に配慮できる ((11)~(13)式から成る一般の状態空間モデルにはこれは成り立たない)。これは、 σ_{v0}^2 を小さくした結果として、(17)式で与えられる $\bar{\mathbf{z}}_{t|t-1}$ と \mathbf{z}_{t-1} の差が縮小され、多重共線性に伴う予測量 $\bar{\mathbf{z}}_{t|t-1}$ のばらつ

きが抑えられるためである（これは縮小推定の一つとみなすことができる）。従って、本研究においても、 $\sigma_{i,0}^2$ を小さな値で与えることで多重共線性に配慮することとする。

4.2.計算手順

$\sigma_{i,0}^2$ を小さな値で与えた上で、(20)～(22)式に基づいたカルマンフィルタを適用することで各時点の潜在変数 \tilde{z}_t を求めていくことで、モデルを同定できる。また、同定されたモデルを元に時空間内挿を行うことができる。例えば、地点0の t 期における被説明変数 $y_{t,0}$ の内挿値は下式で与えられる。

$$E(y_{t,0}) = \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{c}_0 \end{bmatrix} \tilde{\mathbf{z}}_t = \tilde{\mathbf{a}}_0 \tilde{\mathbf{z}}_t \quad (23)$$

ここで、 \mathbf{x}_0 は地点0で観測された説明変数から成る $S_1 \times 1$ のベクトル、 \mathbf{c}_0 は(5)式に基づいて与えられた固有ベクトルの、地点0における実現値から成る $S_2 \times 1$ のベクトルである。

構築されたモデルは、ECSFと同様に、データの持つ空間のパターンを明示的に考慮するだけでなく、通常の状態空間モデルと同様に時間変化に伴う状態遷移も考慮する。即ち、このモデルは空間パターンの状態遷移を考慮したモデルといえる。また、従来の状態空間モデルと同様に、計算量は小さい。さらに、状態空間モデルに基づいて提案した本手法は、固有ベクトルを2次元の長方形空間（各時点で同一と仮定）上から抽出すればよく、(i)の方法（1章参照）に基づいてESFを時空間に拡張したGriffith (2010a)のような、3次元空間上から固有ベクトルを抽出する手法よりも、各時点で用いられる固有ベクトルの数が少なく済むため、自由度の節約の観点も、提案手法は優れている。

一般に、空間統計モデルと状態空間モデルを組み合わせた時空間モデルは、潜在変数以外にも、空間的な依存関係を表現するパラメータを仮定する必要があり、それらと潜在変数を同時に解くために、一般に繰り返し計算が不可欠となる。一方で、提案したモデルは潜在変数のみを未知変数に仮定したモデルであり、繰り返し計算は不要である。また、実用で広く用いられている形態の状態空間モデル（川口, 1991）と同一である。従って、提案したモデルは実用に特化した時空間モデルといえる。

一方で、多重共線性への対処を行った結果として、一般の状態空間モデル(11)～(13)式ではなく、より限定された状態空間モデル(14)～(17)式に基づいた拡張しかできていない。また、提案モデルは、従来のECSFと同様に長方形の平面（図2）を仮定しなければならない。従って、提案モデルは理論的には必ずしも完全とは言えないため、その理論の精緻化は

今後の重要な課題といえよう。

5. 実証分析:住宅地地価の時空間内挿

5.1.概要

住宅地地価（出典：国土数値情報）の時空間内挿に提案手法を適用することで、その有用性を検証する。対象年次は2000年から2009年、対象範囲は茨城県とした（標本数：4200）。比較するモデルは、(14)～(16)式で定義される状態空間モデル（Non-Spatial）で定義される提案モデル（Spatial）とした。

両モデルの被説明変数は公示地価（万円/m²）の自然対数、説明変数は表1に示す各変数である。なお、土地利用のダミーを表す変数間の完全な多重共線性を防ぐため、土地利用分類のうちの一つである「建物用地」は説明変数から除外した。

ECSFで必要となる固有ベクトルは、平面直角座標系新日本測地系第9系で定義された座標上で、対象地域内のX、Y座標の最大値、最小値を計測し、それらで囲まれる長方形の空間をから抽出した。

時空間内挿は対象地域内の町丁目に対して行う。なお、内挿結果等の空間分布は大会時に報告する。

5.2.結果

5-fold クロスバリデーションを用いることで、即ち、データをランダムに5分割し、そのうちの4つのデータを用いた残りの1つの予測を5回繰り返すことで、両モデルの予測能力を検証する。各モデルの予測精度は平均二乗平方根誤差（RMSE：Root Mean Square Error）により評価する。

$$RMSE = \sqrt{\frac{1}{4800} \sum_s \sum_t (y_{s,t} - \hat{y}_{s,t})^2} \quad (24)$$

$y_{s,t}$ と $\hat{y}_{s,t}$ は、それぞれ地点 s 、時点 t の地価（対数

表1:説明変数の概要

説明変数	出典・年度	概要
東京距離	Yahoo! 路線情報 (2009)	最寄り駅から東京、新宿、池袋、渋谷、品川のいずれかの駅までの最短所要時間(分)
駅距離	国土数値情報 (2007)	最寄り駅までの距離(km)
市街化区域	国土数値情報 (2006)	市街化区域であれば1を示すダミー変数
田	国土数値情報 (2006)	所属する3次メッシュ（約1km ² ）あたりの面積(m ² /km ²)
その他農地		
森林		
荒地		
交用地		
その他用地		
河川		
海浜		
海水域		
ゴルフ場		

表2：RMSEの比較（単位：円/m²の対数値）

	Non-Spatial	Spatial
RMSE	0.469	0.388

表3：誤差率の比較（単位：%）

	Non-Spatial	Spatial
平均値	3.545	2.894
中央値	2.795	2.205
標準偏差	2.926×10 ⁻²	2.528×10 ⁻²
最大値	26.26	23.09
最小値	2.407×10 ⁻⁶	5.247×10 ⁻⁶

値)の実測値と予測値を表す。

RMSEの計算結果を表2に示す。この表より、SpatialのRMSEがNon-Spatialよりも小さく、ECSFを援用することで空間パターンを考慮することが、時空間内挿の精度向上に資することが確認された。

次に、両モデルの誤差率(25)式を比較することで、データ観測点毎の内挿精度を比較する。

$$\text{誤差率} = \left| \frac{\hat{y}_{s,t} - y_{s,t}}{y_{s,t}} \right| \times 100 \quad (25)$$

誤差率の計算結果を表3に示す。この表の平均値と中央値の比較より、誤差率でも、Spatialの精度がNon-Spatialを相当程度上回ったことが確認できる。また、Spatialの誤差率は最大でも23.09%であり、この点でも、最大26.26%の誤差率を持つNon-Spatialよりも精度がよいことが確認できる。実際に、4,200の全観測点の約2/3にあたる2795の観測点で、Spatialの精度はNon-Spatialを上回った。

謝辞

本研究は、(財)日本学術振興会科学研究費補助金(特別研究員奨励費)の成果の一部である。

Appendix 1

村上(2012)は投稿中であり参照できないため、その概略を以下に示す。

ESFは、柔軟かつ実用的であり有用であるものの、隣接行列に基づいた手法であるために空間の連続性を考慮したモデリングに適用できない。この点に着目し、村上(2012)はESFを連続空間に拡張したECSFを提案した。

第2章で議論したように、この拡張にあたっては、まず、メッシュの敷き詰められた離散空間(図1)上で定義された、隣接行列の固有ベクトルの解析解(3)式を、 $X \cdot Y$ 軸の各方向のメッシュの数を各々 $+\infty$ とすることで連続空間(図2)上の解析解(5)式に変換した。次に、(5)式に基づいて与えられた固有ベクトルを、(通常)線形回帰モデルの説明変数として投入することで、ECSFの基本式である(7)式を構築した。(7)式の推定は、長方形空間内に分布した観測地点のデータを用いて、モデルに投入する固有ベクトルを変数選択で選定した後に、最小二乗法等を適用することで行う。

ECSFはESFと同様に柔軟かつ実用的である。また、ECSFは、空間の連続性を考慮することができる点と計算量がより小さい点でESFに勝る。但し、分析の結果が長方形空間(図2)の取り方に依存する点でESFに劣る。一方、ECSFは非定常な空間パターンを考慮することができる点、実装と拡張が容易である点で、空間統計学の代表的な手法であるKrigingに勝るが、長方形の取り方に分析結果が依存する点、パラメータ数が多くなり自由度が損なわれる点でKrigingに劣る。

ECSFの有用性を確認するために、ECSF、Kriging、通常線形回帰モデルを、それぞれ首都圏の地価の空間内挿に適用した。その結果、ECSFを用いた内挿の精度はKrigingを相当程度上回ること、残差のモランI統計量のz値は線形回帰モデルの1/100程度となり、残差の空間的相関の除外への一定の有用性があること等が確認された。

参考文献

- Cressie, N. (1993) *Statistics for Spatial Data*. Revised Edition, John Wiley & Sons.
- Cressie, N., Shi, T. and Kang, E. L. (2010) Fixed rank filtering for spatio-temporal data, *Journal of Computational and Graphical Statistics*, **19** (3), pp.724–745.
- Cressie, N. and Wikle, C.K. (2011) *Statistics for Spatio-Temporal Data (Wiley Series in Probability and Statistics)*. Wiley.
- Cressie, N. and Huang, H-C. (1999) Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. *Journal of the American Statistical Association*, **94** (448), pp.1330-1340.
- Dray, S., Legendre, P. and Peres-Neto, P. R. (2006) Spatial Modelling: A Comprehensive Framework for Principal Coordinate Analysis of Neighbour Matrices (PCNM). *Ecological Modelling*, **196** (3–4), pp.483–493.
- Douglas, W. (2010) Data Assimilation. In Gelfand, A.E., Diggle, P.J., Puentes, M. and Guttorp, P. eds. *Handbook of Spatial Statistics*, Chapman & Hall/CRC..
- Griffith, D. A. (2000) Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. *Linear Algebra & Its Applications*, **321**, pp.95–112.
- Griffith, D. A. (2003) *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin: Springer-Verlag.
- Griffith, D.A. (2010a) Modeling spatio-temporal relationships: retrospect and prospect. *Journal of Geographical Systems*, **12** (2), pp.111–123.
- Griffith, D.A. (2010b) Spatial filtering and missing georeferenced data imputation: A comparison of the Getis and Griffith methods, In Anselin, L. and Rey, S.J. eds. *Perspectives on Spatial Data Analysis*, **16**, pp.227–234.
- Griffith, D. A. (2010c) Spatial filtering. In Fischer, M. M. and Getis, A. eds. *Handbook of Applied Spatial Analysis*, Springer.
- Griffith, D. A. and Peres-Neto, P. (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, **87**, pp.2603–2613.

- Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASME, Series D* (82), 95–108.
- Mardia, K.V., Goodall, C., Redfern, E.J. and Alonso, F.J. (1998) The kriged Kalman filter. *TEST*, 7 (2), pp.217–282.
- Moran, P.A.P. (1948) The interpretation of statistical maps, *Journal of the Royal Statistical Society B*, 10, pp.243–251.
- Watson, P.K. (1983) Kalman filtering as an alternative to ordinary least squares—Some theoretical considerations and empirical results. *Empirical Econometrics*, 8, pp.71–85.
- Wikle, C.K. and Cressie, N. (1999) A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86 (4), pp.815–829.
- 川口能典 (1991) Bayesian vector autoregression—その手法の整理と予測能力の検証. 『金融研究』, 10 (1), pp.79–101.
- 谷崎久志 (1993) 『状態空間モデルの経済学への応用』, 日本評論社.
- 樋口知之・上野玄太・中野慎也・中村和幸・吉田亮 (2011) 『データ同化入門—次世代のシュミレーション技術—』, 朝倉書店.
- ビショップ, C. M. (2007) 『パターン認識と機械学習 下 (元田浩・栗田多喜夫・樋口知之・松本裕治・村田昇 監訳)』, シュプリンガー・ジャパン.
- 村上大輔 (2012) 固有ベクトル空間フィルタリングの連続空間への拡張. 『GIS-理論と応用』 (投稿中).