

市町村合併による統計データの集計単位変更 に対する方策の提案： 空間計量経済モデルを用いた分析への対処法

堤盛人¹・村上大輔²

¹正会員 博士(工学) 筑波大学准教授 大学院システム情報工学研究科(〒305-8573 つくば市天王台1-1-1)

²学生会員 博士前期課程 筑波大学大学院 システム情報工学研究科(〒305-8573 つくば市天王台 1-1-1)

いわゆる平成の大合併により、市町村単位の統計データの集計区分も大きく変わり、パネル分析において何らかの工夫が必要となる。分析の対象期間中にそのような市町村合併が起こった場合、通常は、合併後の市町村を単位とした集計に基づき分析が行われる。しかしながら、この方法では、より細かい集計単位である合併前の市町村を単位としたデータが持っている情報の一部が失われてしまう危険がある。本研究では、空間計量経済モデルを用いた計量分析において、合併前の市町村を集計単位とした分析を行う方法を提案し、その実用可能性について検討を行う。

Key Words: *areal interpolation, spatial econometrics, pycnophylactic property, EM-algorithm*

1. はじめに

社会経済データの多くは、空間的な単位（以下、集計単位と呼ぶ）毎の集計データとして提供されている。言うまでもなく、集計単位として我が国で最も広く用いられているのは行政区分であり、国勢調査データをはじめ多くの社会経済データは、市町村毎または都道府県毎に集計された後、一般に提供される。しかし、行政区分は変更されることも多く、それに伴う集計単位の変更は時系列での比較やパネルデータ分析を行う上での障害となる。特に近年、市町村区分は、いわゆる平成の大合併に伴い大きく変更されてきた。総務省 合併相談コーナー（URL：<http://www.soumu.go.jp/gapei/>）によると、2000年に存在した約3,200の市町村は、平成の大合併後の2007年には1,700程度にまで減少した。従って、我が国の市町村データを用いたパネルデータ分

析において、集計単位変更による障害は、大きな問題である。

時点によって集計単位の異なるデータの集計単位を全時点について統一するためには、集計されたものとは異なる集計単位（以下、本論文では「配分単位」と呼ぶこととする）へのデータの変換が必要である。そのような変換は、一般に面補間（Areal Interpolation（例えば、Tobler (1979)¹³⁾））と呼ばれる。

合併による市町村区分の変化に際しては、合併後の市町村区分を配分単位とした面補間により対処されることが多い。しかし、それでは合併前の市町村データについての情報は損失してしまう。一方で、合併前の市町村区分を配分単位とした対処も考えられる。それならば、合併前の市町村データについての情報を損失することなく集計単位の統一を行うことができる。

一般に、配分単位の取り方が、その後の分析結

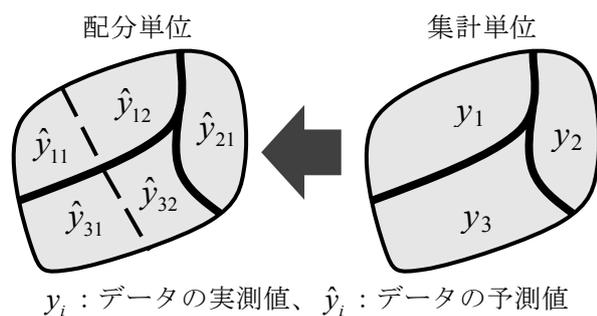


図 1 : 面補間の概要

果に影響することは、可変単位地区問題 (MUAP: Modifiable Areal Unit Problem (Openshaw *et al.* (1979)¹⁰⁾)) として知られている。筆者の知るところ、理論的に十分と思われる可変単位地区問題の解決方法は、現在のところ確立されておらず、配分単位の取り方に優劣をつけることは難しい。しかし、大きな配分単位毎に集計されたデータに対する回帰モデルの適用は、見せかけの相関や推定値の分散の増大を招くことが指摘されている (Reynolds *et al.* (1979)¹¹⁾、Tagashira and Okabe (2002)¹²⁾)。後者の方法の適用は、そのような指摘への対処ともなり得るだろう。

もちろん、後者の方法は、合併後の市町村データを合併前の市町村データに変換するための面補間が必要であり、その際の予測誤差が避けられない。予測誤差はその後の分析結果に影響を与えることから、十分に小さくする必要がある。

予測誤差を小さくするには、データが持つ特性を十分に考慮したモデルを用いて面補間を行うことが望ましい。市町村データのような空間データの代表的特性として、空間従属性が知られている。空間従属性とは、空間的に近接したデータ間に、強い類似傾向が見られるという特性であり、当然のことながら、面補間においても考慮すべき性質である。しかし、面補間において満たされるべき性質とされている体積保存則 (詳しくは次章を参照) を満足し、かつ空間従属性を考慮した手法の提案を行った研究は、筆者らの知る限り見当たらない。

そこで本研究では、まず、体積保存則及び空間従属性を考慮した面補間法の提案を行う。次に、

提案手法を実際の市町村データに適用し、予測誤差の検証を行う。

そのために、面補間に関連する既存研究の整理を 2 章で行い、提案手法のベースとなる空間計量経済モデルとそのパラメータ推定に必要である EM アルゴリズムについての説明を、3 章と 4 章でそれぞれ行う。そして、それらを元に体積保存則及び空間従属性を考慮した面補間法の提案を 5 章で示し、最後に、実証実験を通じた提案手法の有用性の確認を 6 章で行う。

2. 面補間に関する既存研究

集計データの一般的な性質として、体積保存則 (Pycnophylactic property) が知られている。これは、特定の空間領域内の変数の総量は、集計単位の取り方に依存せず、一定であることを表す性質である。例えば、体積保存則が満足されるならば、B 地区と C 地区で構成される A 市の総人口は、B 地区と C 地区の人口の和に等しくなる。

本研究では、「体積保存則の満足を前提とした、集計単位から配分単位へのデータの再配分」を面補間と定義し、以下の議論を進める。

代表的な面補間法として、面積按分法と代表点内包法が知られている (例えば、貞広 (2000)¹⁸⁾)。前者は面積に応じて集計単位のデータを配分単位へと配分することで、予測を行う方法であり、後者は集計単位のデータをポイントデータに置き換え、配分単位に含まれるポイントデータの値から配分単位のデータの予測をする方法である。しかし、配分単位の大きさが集計単位と同程度あるいはそれ以下の場合、面補間の予測誤差は、非常に大きくなることが知られており (貞広 (2000)¹⁸⁾)、面積按分法や代表点内包法も例外ではない。従って、合併前の市町村を配分単位とした面補間のための手法の選択には注意が必要である。

その一方で、近年、補助的なデータを用いた手法が数多く提案されており、合併前の市町村への面補間に対しても一定の予測精度が期待される手法も存在する。例えば、土地利用図や人工衛星画

像を用いるdasymetric法 (Wright (1936)¹⁴⁾) や、配分単位の属性データを用いる線形回帰モデル法 (Frowerdew and Green (1992)⁵⁾) などが存在する。特に後者の手法は、合併前の集計単位に関する複数の情報を用いた面補間が可能であることから、市町村合併による問題への適用に際しても、比較的、精度の高い面補間の実行が期待できる。

また、空間従属性を考慮することによる予測精度の向上も考えられる。空間従属性を面補間に考慮した研究は、これまでも僅かに存在する (Mugglin *et al.* (1999)⁹⁾) もの、空間従属性を考慮し、かつ体積保存則を満足する面補間法は未だ提案されていない。

そこで、本研究では、線形回帰モデルに基づいた面補間法 (Flowerdew and Green (1992)⁵⁾) の拡張により、空間従属性及び体積保存則を考慮した面補間を行うことのできる、新たな手法の提案を行う。

3. 空間計量経済モデルとデータ補間

(1) 空間計量経済モデルの概要

空間計量経済学では、空間従属性を考慮した多くのモデルがこれまで提案されてきた。空間計量経済学において、空間従属性の強さは、距離の関数 (例えば距離の逆数) により表現される。ゾーンを単位とした分析においては、距離はゾーンの代表点 (例えば幾何的重心) 間のユークリッド距離で与えられる場合が多く、そのような距離の関数によって地点間の空間従属性の強さが記述された行列は、空間重み行列 (Spatial Weight Matrix) と呼ばれる。通常、空間重み行列の対角要素は 0 とされる。また、各行の要素の総和が 1 となるよう基準化して用いられることが多く、本研究においてもそれらを採用する。

空間計量経済学の代表的なモデルとして空間ラグモデル (SLM: Spatial Lag Model) が知られている。SLM は被説明変数に空間従属性を仮定したモデルであり (1) 式で表される。

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (1)$$

観測地点数を s とすると、 \mathbf{y} は $s \times 1$ の被説明変数ベクトル、 \mathbf{X} は $s \times t$ の説明変数ベクトル、また、 $\boldsymbol{\beta}$ は $t \times 1$ のパラメータベクトルで表される。 \mathbf{W} は空間重み行列であり。また、 ρ は空間従属性の強さを表すパラメータであり、 -1 から 1 の間の値を取る。

通常、空間計量経済モデルのパラメータ推定の一つとして、最尤法が用いられており、SLMでは対数尤度関数(2)の最大化によりパラメータを推定する (Anselin (1988)¹⁾) 。

$$L = -\frac{s}{2} \ln(2\pi) - \frac{s}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \rho \mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \ln |\mathbf{I} - \rho \mathbf{W}| \quad (2)$$

そのために、(2) 式を $\boldsymbol{\beta}$ に関し偏微分することで、 $\hat{\boldsymbol{\beta}}$ が得られる。

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_1 - \rho \hat{\boldsymbol{\beta}}_{11} \quad (3)$$

ただし、 $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ 、 $\hat{\boldsymbol{\beta}}_{11} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{y}$ である。次に、対数尤度関数を σ^2 に関し偏微分することで、 $\hat{\sigma}^2$ が得られる。

$$\hat{\sigma}^2 = \frac{(\hat{\boldsymbol{\varepsilon}}_1 - \rho \hat{\boldsymbol{\varepsilon}}_{11})' (\hat{\boldsymbol{\varepsilon}}_1 - \rho \hat{\boldsymbol{\varepsilon}}_{11})}{s} \quad (4)$$

ただし、 $\hat{\boldsymbol{\varepsilon}}_1 = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_1$ 、 $\hat{\boldsymbol{\varepsilon}}_{11} = \mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{11}$ である。

(2) 式に (3)、(4) 式を代入することで、 ρ に関する集中化対数尤度関数(5)式が得られる。

$$L_c = -\frac{s}{2} \ln \left[\frac{1}{s} (\hat{\boldsymbol{\varepsilon}}_1 - \rho \hat{\boldsymbol{\varepsilon}}_{11})' (\hat{\boldsymbol{\varepsilon}}_1 - \rho \hat{\boldsymbol{\varepsilon}}_{11}) \right] + \ln |\mathbf{I} - \rho \mathbf{W}| + const. \quad (5)$$

上の式を元に、以下の手順でパラメータが推定される。

- [1] OLS を用いて \mathbf{y} を \mathbf{X} に回帰し $\hat{\boldsymbol{\beta}}_1$ を算出する
- [2] OLS を用いて $\mathbf{W}\mathbf{y}$ を \mathbf{X} に回帰し $\hat{\boldsymbol{\beta}}_{11}$ を算出する
- [3] 残差 $\hat{\boldsymbol{\varepsilon}}_1$ 、 $\hat{\boldsymbol{\varepsilon}}_{11}$ を求める
- [4] L_c を最大にする ρ を求める
- [5] $\hat{\boldsymbol{\beta}}$ と $\hat{\sigma}^2$ を上の等式により求める

(2) 空間計量経済モデルを用いた補間の既存研究

通常の空間計量経済モデルでは、観測データ間のみの空間従属性を考慮した空間重み行列により、

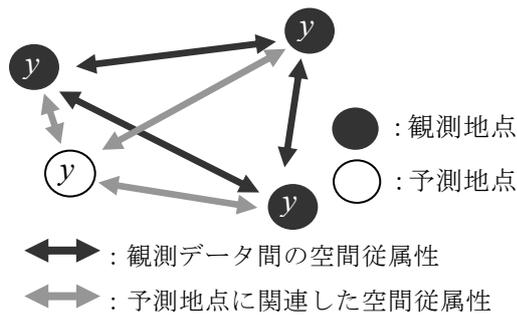


図2：地点間の空間従属性

モデル構築及びパラメータ推定が行われる。従って、補間地点に関連した空間従属性も考慮しなくてはならない補間に対し、通常モデルを用いることはできない（堤他 (2000)²⁰⁾（図2）。

それに対し、LeSage and Pace (2004)⁵⁾では、補間地点も含む全地点についての空間重み行列を用いた空間計量経済モデルの構築と、EMアルゴリズム（次章を参照）によるパラメータ推定により、理論に整合した点補間を行っている。また、瀬谷 (2008)¹⁹⁾では、EMアルゴリズムの類似手法であるSimultaneous approach（Martin (1984)⁸⁾）を用いた点補間を行い、Kriging（Cressie (1993)³⁾）など、他の分野の手法による結果と比較することで、空間計量経済モデルにより、精度の良い点補間が行われることを確認している。

しかし、いずれも点補間ための手法であり、それらを面補間に適用しても、前章で説明した体積保存則は一般的には満足されない。

4. EM アルゴリズムと面補間

(1) EM アルゴリズムの概要

EM(Expectation Maximization)アルゴリズムは、Dempster *et al.* (1977)⁴⁾により提案された、不完全データを用いて最尤推定を行うことのできるアルゴリズムであり、その簡潔さと柔軟な適用可能性のために幅広い分野での応用が見られる。通常、EMアルゴリズムでは以下の手順に従い最尤推定を行う。

[1]パラメータに適当な初期値 θ^0 を与える。

[2]以下の2つのステップを収束するまで交互に

繰り返す。

E-Step

h 回目の繰り返しで得られたパラメータ θ^h を元に、データの欠損部分に疑似的な観測値を代入し、尤度関数 L^h を構築する。

M-Step

尤度関数 L^h を最大化するようなパラメータ θ^{h+1} を求める。

以上の計算により、尤もらしいパラメータと予測値の算出を行うことができる。繰り返し計算における尤度の単調増加は保証されており、通常は、パラメータまたは尤度の変化が十分に小さくなれば、計算は収束したと見なす。ただし、EMアルゴリズムは局所的最適解への収束の可能性が知られており、パラメータの初期値は適切に設定される必要がある。

(2) EM アルゴリズムを用いた面補間

本節以降では、集計単位、配分単位に加え、集計単位及び配分単位の両方の空間分割を施すことで得られる空間単位（図3）を細分単位と定義し、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、各空間単位内における被説明変数ベクトルを \mathbf{y} は \mathbf{y}_i 、 \mathbf{y}_j 、 \mathbf{y}_k と表現する。各ベクトルのサイズは、それぞれ $I \times 1$ 、 $J \times 1$ 、 $K \times 1$ であり、 $I \leq K$ 及び $J \leq K$ が成り立つ。

Flowerdew and Green (1992)⁵⁾では、線形回帰モデルをベースとしたモデルによる面補間を、EMアルゴリズムを用いた以下の手順で行っている（図4）。

[1]集計単位及び配分単位を元に細分単位を生成する。

[2]線形回帰モデルをベースとしたモデルにより、集計単位から細分単位へのデータの按分を行う。

[3][2]の結果得られた細分単位の予測値を元に、配分単位の予測値を算出する。

以降では、量を表す変数により基準化された変数（例えば高齢化率）を対象とした面補間の手順を、

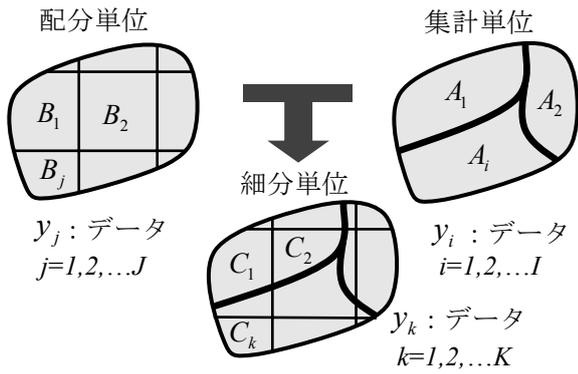


図3：細分単位の概要

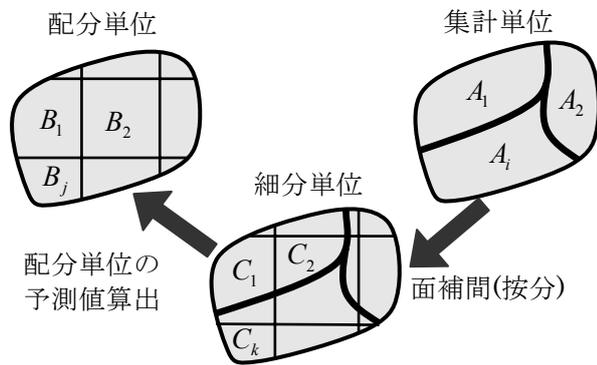


図4：Flowerdew and Green (1992)⁵⁾の計算手順

[2]及び[3]について説明する。以下では基準化に用いる変数は「基準化変数」と呼ぶ。

手順[2]では、面補間の結果得られる細分単位の予測値が、体積保存則を満足しなくてはならない。一般に、基準化された変数が体積保存則を満足するためには、(6)式が成り立たなくてはならない。

$$\bar{y}_i = \sum_k n_{ik} \hat{y}_k \quad (6)$$

\bar{y}_i は、集計単位*i*における被説明変数の実測値である。また n_{ik} は集計単位*A_i*における基準化変数の観測のうち、細分単位*C_k*で観測された同観測量の割合を表す。従って、 $A_i \supseteq C_k$ の場合についてのみ n_{ik} は0から1の間の値を取り、それ以外の場合は0である。

(6)式を、細分単位の被説明変数の予測値ベクトル $\hat{\mathbf{y}}_k$ が体積保存則を満足するための条件式に直すと(7)式が得られる。

$$\bar{\mathbf{z}}_i = \mathbf{N}_{ik} \hat{\mathbf{y}}_k \quad (7)$$

ここで、 $\bar{\mathbf{z}}_i$ は各細分単位が所属する集計単位の被説明変数を表す、 $K \times 1$ の実測値ベクトルである。

また、 \mathbf{N}_{ik} は n_{ik} を要素とする $K \times K$ の正方行列であり、その第*p*行*q*列目の要素は、細分単位*C_p*が所属する集計単位の基準化変数観測のうち、細分単位内*C_q*で得られる同観測値の割合である。 \mathbf{N}_{ik} を模式的に表すと、以下で表される。

$$\mathbf{N}_{ik} = \begin{pmatrix} n_{11} & n_{12} & 1-n_{11}-n_{12} & 0 & 0 & 0 \\ n_{11} & n_{12} & 1-n_{11}-n_{12} & 0 & 0 & 0 \\ n_{11} & n_{12} & 1-n_{11}-n_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & n_{31} & 1-n_{31} \\ 0 & 0 & 0 & 0 & n_{31} & 1-n_{31} \\ \vdots & & & & & & \ddots \end{pmatrix} \quad (8)$$

Flowerdew and Green (1992)⁵⁾では、通常の線形回帰モデルに対して補正項を加えた(9)式による面補間法を提案している。(9)式から得られる予測値 $\hat{\mathbf{y}}_k$ は体積保存則満足のための条件式(7)を必ず満足する。

$$\mathbf{y}_k = \mathbf{X}_k \boldsymbol{\beta} + [\bar{\mathbf{z}}_i - \mathbf{N}_{ik}(\mathbf{X}_k \boldsymbol{\beta})] + \boldsymbol{\varepsilon}_k \quad (9)$$

\mathbf{X}_k は $K \times t$ の細分単位の説明変数ベクトルであり、 $\boldsymbol{\beta}$ は $t \times 1$ のパラメータベクトルである。

(9)式は、 $\bar{\mathbf{z}}_i$ と \mathbf{y}_k (ともに $K \times 1$ のベクトル)との乖離を、 \mathbf{X}_k^* により予測することで、 $\hat{\mathbf{y}}_k$ を算出するモデルである。従って、同一の集計単位に所属する細分単位ならば必ず同一の値を示す変数を説明変数とした場合、同一集計単位に属する全ての細分単位の予測値は等しくなる。そのような場合、体積保存則を満足するためには、 $\bar{\mathbf{z}}_i = \hat{\mathbf{y}}_k$ とならなくてはならないことが、(7)式より確認できる。そのため、集計単位毎に同一の値を示す変数を説明変数に用いて面補間をすることはできない。また、同様の理由で定数項を用いることはできない。

(9)式による予測値の算出とパラメータの推定を、EMアルゴリズムにより行うことで、各細分単位の予測値 $\hat{\mathbf{y}}_k$ が算出され、その要素を(10)式に代入することで、配分単位*B_j*における予測値 \hat{y}_j を算出することができる。

$$\hat{y}_j = \sum_k n_{jk} \hat{y}_k \quad (10)$$

n_{jk} は、配分単位*B_j*における基準化変数の観測のうち、細分単位*C_k*で観測された同観測量の割合である。従って、 $B_j \supseteq C_k$ の場合のみ n_{jk} は0から1の

間の値を取り、それ以外の場合は0である。

5. 空間重み行列を用いた EM アルゴリズムによる面補間手法の提案

(1) 空間重み行列を用いた面補間のためのモデルの導出

本研究では、LeSage and Pace (2006)⁷⁾における空間計量経済モデルによる点補間法と、Flowerdew and Green (1992)⁵⁾におけるEMアルゴリズムを用いた面補間法を組み合わせることで、空間従属性を考慮した面補間のための手法の提案を行う。まず、(9)式に倣い、体積保存則のための補正項を加えた(11)式を定義する。

$$\mathbf{y}_k = \rho \mathbf{W}_k \mathbf{y}_k + \mathbf{X}_k \boldsymbol{\beta} + [\bar{\mathbf{z}}_i - \mathbf{N}_{ik} (\rho \mathbf{W}_k \mathbf{y}_k + \mathbf{X}_k \boldsymbol{\beta})] + \boldsymbol{\varepsilon}_k \quad (11)$$

(9)式と同様に、(11)式についても、所属する集計単位毎に同一の値を示す説明変数及び定数項を用いてはいけない。

ここで、 $\mathbf{W}_k^* = (\mathbf{I} - \mathbf{N}_{ik}) \mathbf{W}_k$ 、 $\mathbf{X}_k^* = (\mathbf{I} - \mathbf{N}_{ik}) \mathbf{X}_k$ とすることで(12)式のように整理される。

$$\mathbf{y}_k = \rho \mathbf{W}_k^* \mathbf{y}_k + \mathbf{X}_k^* \boldsymbol{\beta} + \bar{\mathbf{z}}_i + \boldsymbol{\varepsilon}_k \quad (12)$$

(12)式は、(1)式で示した SLM と構造が同じであり、対数尤度関数は次式で与えられる。

$$L^* = -\frac{K}{2} \ln(2\pi) - \frac{K}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y}_k - \rho \mathbf{W}_k^* \mathbf{y}_k - \mathbf{X}_k^* \boldsymbol{\beta} - \bar{\mathbf{z}}_i)' (\mathbf{y}_k - \rho \mathbf{W}_k^* \mathbf{y}_k - \mathbf{X}_k^* \boldsymbol{\beta} - \bar{\mathbf{z}}_i) + \ln |\mathbf{I} - \rho \mathbf{W}_k^*| \quad (13)$$

対数尤度関数を $\boldsymbol{\beta}$ と σ^2 それぞれについて偏微分することで、 ρ に関する集中化対数尤度関数(14)式が導かれる。

$$L_c^* = -\frac{n}{2} \ln \left[\frac{1}{n} (\hat{\boldsymbol{\varepsilon}}_i - \rho \hat{\boldsymbol{\varepsilon}}_{ii} - \hat{\boldsymbol{\varepsilon}}_{im})' (\hat{\boldsymbol{\varepsilon}}_i - \rho \hat{\boldsymbol{\varepsilon}}_{ii} - \hat{\boldsymbol{\varepsilon}}_{im}) \right] + \ln |\mathbf{I} - \rho \mathbf{W}_k^*| + const. \quad (14)$$

ただし、 $\hat{\boldsymbol{\varepsilon}}_{im} = \bar{\mathbf{z}}_i - \mathbf{X}_k \hat{\boldsymbol{\beta}}_{im}$ 、 $\hat{\boldsymbol{\beta}}_{im} = (\mathbf{X}_k' \mathbf{X}_k)^{-1} \mathbf{X}_k' \bar{\mathbf{z}}_i$ であり、

(13)式内の $\boldsymbol{\beta}$ は $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_i - \rho \hat{\boldsymbol{\beta}}_{ii} + \hat{\boldsymbol{\beta}}_{im}$ で与えられる。

(14)式から、尤度算出の必要条件は、 $|\mathbf{I} - \rho \mathbf{W}_k^*| = |\mathbf{I} - \rho (\mathbf{I} - \mathbf{N}_{ik}) \mathbf{W}_k| > 0$ となる。 $\mathbf{D} = |\mathbf{I} - \rho \mathbf{W}_k^*|$ と

$$\mathbf{D}_1 = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{O} & \mathbf{D}_{13} \end{pmatrix}$$

(row labels: $n-m$, m ; column labels: $n-m$, m)

図5：ブロック三角行列の構造

すると、 \mathbf{D} の、任意の集計単位と完全に一致する細分単位についての行は、対角項以外の全ての要素が零となる。従って、行列の基本変形及び行列式の公式により、 \mathbf{D} は(15)式のように展開される。

$$|\mathbf{D}| = \begin{vmatrix} \mathbf{D}_1 & \mathbf{D}_2 \\ \mathbf{O} & \mathbf{I} \end{vmatrix} = |\mathbf{D}_1| \quad (15)$$

そこで、以下では $|\mathbf{D}_1| > 0$ 、即ち、 \mathbf{D}_1 の正定値性を検証することで $|\mathbf{D}| > 0$ を示す。Burga(1972)²⁾によると、正方行列が正定値行列であるための十分条件は、以下の条件を満たすことである(塚井、奥村(2007)²¹⁾参照)。

- [I] 既約行列であること
- [II] 狭義優対角行列であること
- [III] 全ての対角項が実数であること

[I]の既約行列とは、行、列の同時入れ替えにより、ブロック三角行列への変換が不可能な行列である(例えば、児玉・須田(1978)¹⁶⁾)。ブロック三角行列とは、三角行列の対角要素を正方行列で置き換えた行列であり、行列 \mathbf{D}_{11} 、 \mathbf{D}_{12} 、 \mathbf{D}_{13} 、及び零行列 \mathbf{O} を用いて \mathbf{D}_1 を図5のように表すことができる場合、 \mathbf{D}_1 はブロック三角行列である。ただし、 \mathbf{D} 及び \mathbf{D}_{13} のサイズはそれぞれ $n \times n$ 、 $m \times m$ ($1 \leq m < n$) とする。

そこで、 \mathbf{D}_1 のブロック三角行列への変換の可否を検証する。まず、細分単位 C_p 、 $C_{p'}$ 、 C_q 、 $C_{q'}$ を定義し、 C_q 及び $C_{q'}$ は、 C_p とは異なる集計単位に所属していると仮定する。すると、 \mathbf{D}_1 の要素 $[p, q]$ は(16)式で表される。

$$-\rho(1 - n_{pq})w_{pq} - \sum_{p \neq q} n_{pq} w_{pq} \quad (16)$$

ここで、 n_{pq} は \mathbf{N}_{ik} の要素 $[p, q]$ 、 w_{pq} は \mathbf{W}_k の要素 $[p, q]$

である。次に、 \mathbf{D}_1 の要素 $[p, q']$ についても同様に式を構築し、両式が同時に0となる連立方程式を解くことで(17)式を得る。

$$w_{pq} = w_{p'q'} \quad (17)$$

また、同一集計単位に所属する C_p 、 C_p 及び、それらとは異なる集計単位に所属する C_q を仮定し、 \mathbf{D}_1 の要素 $[p, q]$ 及び $[p', q]$ が同時に0となるための連立方程式を解くことで、条件式(18)を得る。

$$w_{pq} = w_{p'q} \quad (18)$$

ここで、 \mathbf{D}_1 がブロック三角行列であるためには、 \mathbf{M} と同様に $(n-m) \times m$ の零行列を持つことが必要条件である

以上より、 C_q 及び C_q とは異なる集計単位に C_p が所属しているならば、(17)式が成り立つような q 及び q' の組み合わせが、少なくとも $(n-m-1) \times m$ 組存在しなくてはならない。さらに、 C_q を含まない同一集計単位に C_p 及び C_p が所属しているならば、 p, p' のどのような組み合わせに対しても(18)式が成り立つような q が少なくとも $n-m$ 通り存在しなくてはならない。

以上の条件は、同一の代表点を持つ細分単位が複数存在する場合には成立しうるが、異なる市町村が同一の代表点を持つことは通常ない。従って、 \mathbf{D}_1 は、ブロック三角行列への変換が不可能であることから、既約行列である。

[II]の狭義優対角行列とは、全ての行について、対角要素が非対角要素の総和よりも大きい行列である。以降では、最初に \mathbf{D} の性質を調べ、それを元に \mathbf{D}_1 の性質を調べることで狭義優対角行列であることを示す。

まず、 \mathbf{D} の性質を調べる。そのために、 \mathbf{D} を \mathbf{I} 、 $-\rho \mathbf{W}_k$ 及び $\rho \mathbf{N}_{ik} \mathbf{W}_k$ に分解し、各行列の非対角要素の和を調べる。行列 \mathbf{M} の第 p 行の非対角要素の和を $sum[\mathbf{M}]_p$ とすると、

$$sum[\mathbf{I}]_p = 0 \quad (19)$$

は自明である。また、 \mathbf{W}_k の行基準化を仮定していることから、

$$sum[-\rho \mathbf{W}_k]_p = -\rho \quad (20)$$

である。さらに、 $sum[\rho \mathbf{N}_{ik} \mathbf{W}_k]_p$ は(21)式で表される。

$$sum[\rho \mathbf{N}_{ik} \mathbf{W}_k]_p = \rho [n_{p1}(w_{12} + w_{13} \cdots w_{1K}) + \cdots + n_{p,p-1}(w_{p-1,1} + \cdots + w_{p-1,K}) + n_{p+1,p}(w_{p+1,1} + \cdots + w_{p+1,K}) + \cdots + n_{pK}(w_{K1} + \cdots + w_{K,K-1})] \quad (21)$$

$$\text{s.t. } n_{pq} \geq 0, w_{pq} \geq 0, \sum_{q=1}^n n_{pq} = 1, \sum_{q=1}^n w_{pq} = 1$$

制約式により(21)式は(22)式に展開される。

$$sum[\rho \mathbf{N}_{ik} \mathbf{W}_k]_p = \rho(1 - n_{pp}(w_{p1} + w_{p2} + \cdots + w_{pK})) \quad (22)$$

(19)(20)(22)の各式を足し合わせることで \mathbf{D} の各行の非対角要素の和を求めると、(23)式となる。

$$sum[\mathbf{D}]_p = -\rho n_{pp}(w_{p1} + w_{p2} + \cdots + w_{pK}) \quad (23)$$

次に、行列 \mathbf{M} の第 p 行の対角要素を $diag[\mathbf{M}]_p$ とすると、 $diag[\mathbf{D}]_p$ は、その要素である \mathbf{I} 、 $-\rho \mathbf{W}_k$ 及び $\rho \mathbf{N}_{ik} \mathbf{W}_k$ の各対角要素を元に(24)で表される。

$$diag[\mathbf{D}]_p = 1 + \rho n_{pp}(w_{p1} + w_{p2} + \cdots + w_{pK}) \quad (24)$$

$diag[\mathbf{D}]_p > sum[\mathbf{D}]_p$ を解くことで、少なくとも ρ が0以上であれば、 \mathbf{D} は狭義優対角行列となると分かる。

次に、 \mathbf{D}_1 の性質を調べる。 \mathbf{D}_1 は、 \mathbf{D} の小行列である。また、 \mathbf{D} の3つの各成分中の要素の符号は単一である。従って、 ρ が0以上であれば(25)が成り立つ。

$$sum[\mathbf{D}]_p < sum[\mathbf{D}_1]_p < 0 \quad (25)$$

さらに、 \mathbf{D}_1 の対角要素は \mathbf{D} の対角要素に一致する。従って、 ρ が0以上であれば(26)が成り立つ。

$$diag[\mathbf{D}]_p = diag[\mathbf{D}_1]_p > 0 \quad (26)$$

(25)、(26)より、 \mathbf{D}_1 は狭義優対角行列である。

[III]は自明である。

以上より、 \mathbf{D}_1 の正定値行列は示され、(15)式より \mathbf{D} の正定値性も示されるため、 $[\mathbf{I} - \rho \mathbf{W}_k]^* > 0$ は満足され、尤度関数(13)式を最大化するための必要条件は満たされた。

(2) 提案手法の計算手順

以下に示す手順により、提案モデルを用いた面補間を行うことができる。

[1]各々の細分単位が所属する集計単位の実測値 y_i を、各細分単位内における被説明変数 y_k の初期値とすることで、細分単位の被説明変数ベクトル \mathbf{y}_k を構築する。

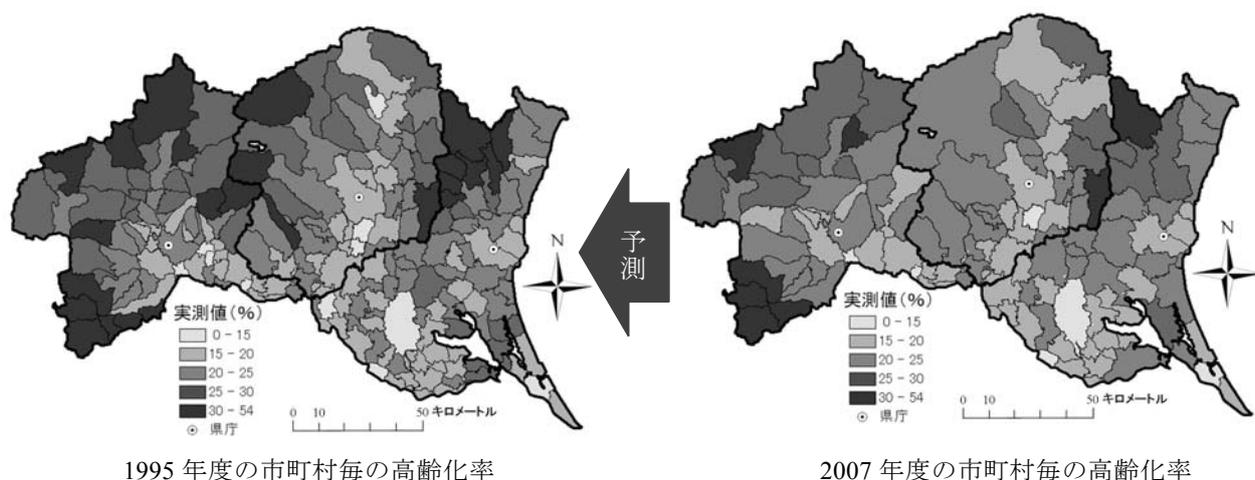


図 6：実証における面補間の概要

[2][1]で得られる y_k を用いたSLMにより、各パラメータの初期値を求める。

[3]以下の2ステップを収束するまで交互に繰り返す。

E-Step

与えられたパラメータを元に、(11)式を用いて予測値を算出する。

M-Step

(13)式で示される尤度関数 L^* を最大化する。ここで、(13)式は補正項付きの、SLMに対する対数尤度関数であることから、(LeSage and Pace (2006)⁷⁾における、予測地点を含む全地点により空間的な均衡状態に至るという点補間の際の仮定の引用により、通常のSLMの場合と同様に、以下の手順でM-STEPを遂行することができる。

- [1]OLSを用いて y_k を $(I-N_{ik})X_k$ に回帰し $\hat{\beta}_1$ を得る
- [2]OLSを用いて $(I-N_{ik})W_k y_k$ を $(I-N_{ik})X_k$ に回帰し $\hat{\beta}_m$ を得る
- [3]OLSを用いて \bar{z}_i を $(I-N_{ik})X_k$ に回帰し $\hat{\beta}_m$ を得る
- [4]残差 $\hat{\epsilon}_1$ 、 $\hat{\epsilon}_m$ 、 $\hat{\epsilon}_m$ を求める
- [5] L_c を最大にする ρ を求める
- [6] $\hat{\beta}$ と $\hat{\sigma}^2$ を得る

以上の計算を、 L_c が収束するまでの間繰り返し行うことで、体積保存則を満足し、かつ、空間従属性を考慮した面補間が細分単位 C_k に対してなされる。最後に、導かれた予測値 \hat{y}_k を(10)式に代入

することで、配分単位 B_j への面補間を行うことができる。

ただし、5章で述べたように、定数項及び所属する集計単位毎に同一の値を示す変数は、提案モデルの説明変数として用いてはいけない。

6. 実際のデータを用いた検証

(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年度の市町村区分を集計単位、1995年度の市町村区分を配分単位とした面補間（図6）を、通常の線形回帰モデル（線形モデル）、Flowerdew and Green (1992)⁵⁾による補正項付きの線形回帰モデル（FGモデル）（5章第2節を参照）及び提案手法（提案モデル）それぞれにより行い、予測精度の比較等から、提案手法の有用性を検証する。

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。そこで、2007年度の町丁目データからそれを作成する。ここでは、町丁目データの入手が容易であった高齢化率 [%] の面補間を例に検証を行う。説明変数には、同じくデータ入手や加工が容易であった、東京駅までの距離（東京距離）[km]、最寄りの県庁所在市町村の主要駅までの距離（県庁距離）[km]、人口密度 [千人/km²]、市町村

表 1：パラメータ推定結果

ケース	1			2	
モデル	線形モデル	FG モデル	提案モデル	線形モデル	空間ラグモデル
定数項	21.8*** (7.59)			18.53*** (2.87)	2.32*** (0.35)
東京距離(km)	0.04*** (2.68)	0.11*** (7.37)	-0.06*** (-4.02)	0.07*** (4.69)	0.02 (1.34)
県庁距離(km)	0.05* (1.90)	0.15*** (5.69)	0.11*** (4.17)	0.06*** (2.27)	0.04 (1.52)
人口密度(千人/km ²)	-0.07 (-1.58)	-0.15*** (-2.48)	-0.08 (-1.32)	-0.33*** (-5.45)	-0.23*** (-3.80)
域外従業率(%)	-0.04 (-1.35)	0.03 (0.87)	0.05 (1.58)	-0.06*** (-2.03)	-0.05 (-1.70)
ρ			7.29		0.92
AIC		624.6	588.1		

※括弧内はt値 ※※***は1%水準、**は5%水準、*は10%水準で有意であることを示す
 ※※※提案モデルの ρ は0以上、空間ラグモデルの ρ は-1から1の間の値をそれぞれとる

外での従業率（域外従業率） [%] を用いる。

また、空間重み行列 \mathbf{W}_k は、要素が $(1/d_{pq}^2)$ である p 行 q 列の行列を、行基準化することで与える。ただし、 d_{pq} は p 番目及び q 番目の各細分単位の幾何的重心間のユークリッド距離とする。

なお、以下の計算は、フリーの統計ソフトウェア「R」により行う。また、予測結果等の空間分布図作成には ESRI 社の「ArcGIS」を使用する。

(2) 空間従属性の検定

空間従属性の検定量としてモランI統計量が知られている。モランI統計量は-1 から 1 までの値を取り、値が 1 に近い程空間従属性が強いと判断される。本研究では、先に述べた被説明変数及び説明変数を用いた線形モデルから得られる残差 ε_k に対するモランI統計量(27)式の適用により、空間従属性の検証を行う。

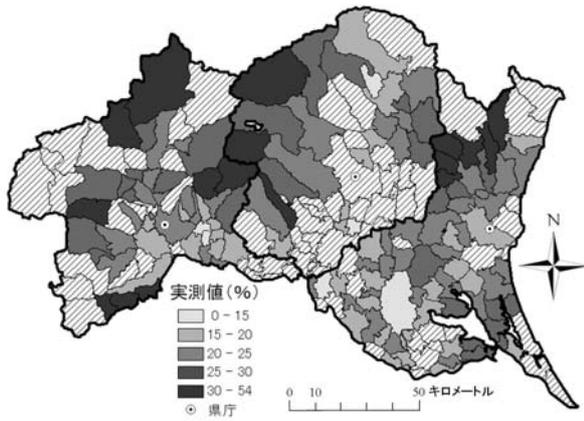
$$\text{モラン I 統計量} = \frac{\sum_p \sum_q w_{pq} (\varepsilon_p - \bar{\varepsilon})(\varepsilon_q - \bar{\varepsilon})}{\sum_p \sum_q w_{pq} \sigma^2} \quad (27)$$

計算の結果、モラン I 統計量は 0.16 であり、空間従属性は存在するという仮説が有意水準 1%で

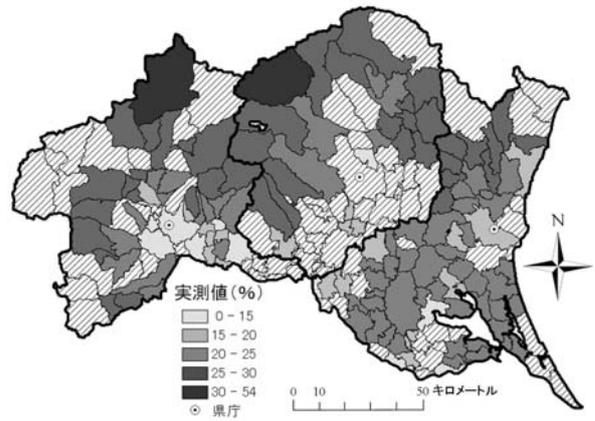
採択された。

(3) パラメータの推定

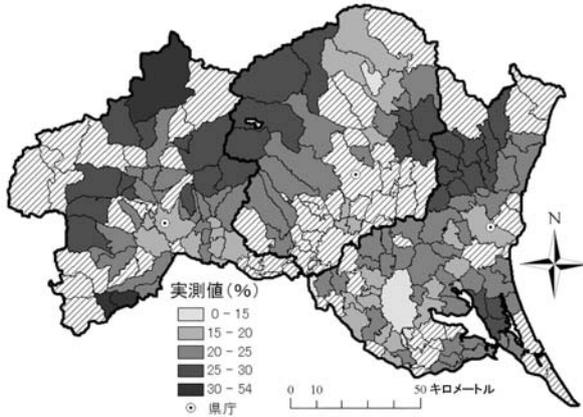
次に、各モデルを用いたパラメータ推定を行う。推定は、面補間における推定(ケース 1)に加え、合併前区分の高齢化率を既知とした、完全データによる推定(ケース 2)を行い、両者を比較することで面補間におけるパラメータ推定値の妥当性の検証を行う。ケース 1 で用いるモデルは線形モデル、FGモデル及び提案モデル、ケース 2 で用いるモデルは線形モデル及び空間ラグモデルである(表 1)。パラメータ推定の結果、ケース 1 において、線形モデルにより得られたパラメータの符号は、ケース 2 の両モデルにおけるそれと完全に一致しているのに対し、FGモデル及び提案モデルにより得られたパラメータの符号は、完全には一致していないことが分かる。このことから、FGモデル及び提案モデルによるパラメータ推定結果には注意が必要であることが示唆される。また、 ρ の値が大きいため、強い空間従属性が示唆される。なお、ケース 1 における提案モデルとケース 2 における SLMモデルの各空間重み行列は、それぞれ \mathbf{W}_k^* 及



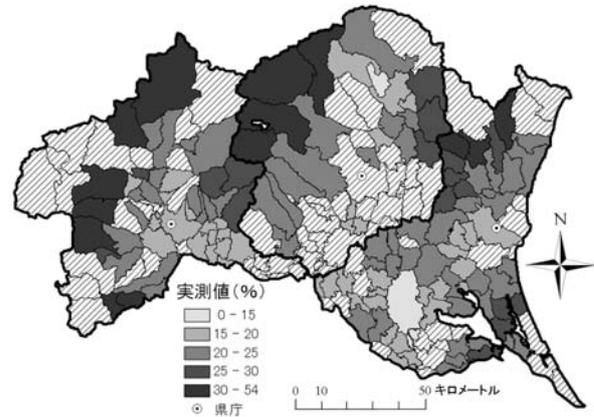
図：7：予測対象の各配分単位の高齢化率実測値



図：8：線形モデルによる高齢化率予測値



図：9：FGモデルによる高齢化率予測値



図：10：提案モデルによる高齢化率予測値

表 2：RMSE の算出結果

	線形モデル	FGモデル	提案モデル
RMSE	6.11	3.78	3.46

び \mathbf{W}_k であり異なるため、 ρ の推定値の比較はできない。

提案モデルは ρ を用いている分パラメータが多いため、提案モデル及び、提案モデルと入れ子構造の関係にあるFGモデルのAICを算出することで、両モデルの予測能力を比較した。その結果、提案モデルのAICはFGモデルのAICに比べて小さく、提案モデルの予測能力高さが確認された。

(4) 予測精度の検証

予測精度の検証は平均2乗平方根誤差(RMSE: Root Mean Square Error) (28)式により行う。

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

ここで、 k' は合併により消滅した細分単位(全 K' 箇所)の添え字である。

表2はRMSEの算出結果であり、FGモデルと線形モデルとの差は、体積保存則の考慮による効果、提案モデルとFGモデルとの差は、空間従属性の考慮による効果をそれぞれ表す。表2から、前者だけでなく、後者についても、一定の差が存在することがわかり、空間従属性の考慮が、面補間においても重要であることが示唆される。

各モデルによる予測結果のプロットは図7～図10である。各図から、茨城県北部や栃木県北部において、提案モデルによる予測値の分布が実測値の分布に近いことがわかる

次に、FGモデルと提案モデルとの予測誤差の差

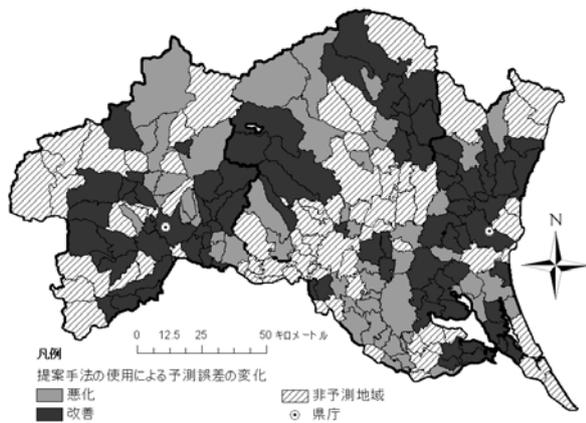


図 12 : FG モデルと提案モデルの予測精度比較

をプロットすることで (図 12)、両モデルの予測精度を詳細に比較した。図 12 から、FG モデルに比べ、提案モデルは茨城県東部や群馬県南部において、比較的優れた予測結果が得られていることが分かる。また、予測結果を行った 134 市町村のうち、83 の市町村で、空間従属性の考慮による予測精度の向上が見られた。

7. おわりに

本研究では、市町村合併に伴う市町村区分の変化により生じるパネルデータ分析の困難への対処方法について議論した。その中で、全時点におけるデータの集計単位を合併前時点における集計単位へと精度良く面補間するために、体積保存則及び空間従属性を考慮した面補間法を提案した。そして、体積保存則のみを考慮した手法及び通常の線形回帰モデルによる手法との実データを用いた比較実験から、提案手法により精度の良い面補間が行われることを確認した。

本研究における提案手法は、市町村データのみならず、多くの集計データへの集計単位変更への適用が考えられる。そのためには、集計データの持つ特性と、提案手法による面補間の精度の関係は明らかにすべきであろう。また、本研究における提案手法は、合併前の市町村区分を用いたパネル分析を行うための手法であり、本来は、3 時点以上の複数時点のデータを同時に面補間できることも望ましい。そのために、空間方向だけでな

く、時間方向の従属性も考慮した手法の拡張についても今後検討していきたい。

謝辞: 第 23 回 ARSC 研究発表大会において、討論者の広島大学の塚井誠人准教授から、有益なコメントを頂きました。ここに記して感謝を申し上げます。無論、本稿に関する有りうべき誤りは全て筆者らに属するものである。

参考文献

- 1) Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- 2) Burga, R. S. (1972) (渋谷政昭 訳) : 『計算機による大型行列の反復解法』, サイエンス社.
- 3) Cressie, N. (1993), *Statistics for Spatial Data*, Revised Edition, John Wiley & Sons.
- 4) Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. Royal Statistical Soc.*, 39, 1, pp.1-38, 1977.
- 5) Flowerdew, R. and Green, M. (1992), Developments in areal interpolation methods and GIS, *Annals of Regional Science*, 26, pp.67-78.
- 6) Goodchild, M. F. and N. N-S. Lam (1980), Areal Interpolation: a Variant of the Traditional Spatial Problem, *Geo-processing 1*, pp.297-312.
- 7) LeSage, J. P. and Pace, R. K. (2004), Models for Spatially Dependent Missing Data, *Journal of Real Estate Finance and Economics*, 29: pp.233-254.
- 8) Martin, R. J. (1984), Exact maximum likelihood for incomplete data from a correlated gaussian process, *Communications in Statistics: Theory and Methods*, 13, pp.1276-1288.
- 9) 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.
- 10) Openshaw, S., and Taylor, P. (1979), A Million or so

Correlation Coefficients, *Statistical Methods in the Spatial Sciences*, Wrigley, London: Pion, pp.127-144.

11) Reynolds, H., and Amrhein, C. G. (1998), Some effects of spatial aggregation on multivariate parameters. *Econometric Advances in Spatial Modeling and Methodology: Essays in the Honour of Jean Paelinck*, Griffith, D., Amrhein, C. G. and Huriot, J-M. (Eds), Kluwer Academic Publishers.

12) Tagashira, Y., and Okabe, A. (2002), The modifiable areal unit problem in a regression model whose independent variable is a distance from a predetermined point, *Geographical analysis*, vol.34(1), pp.1-20.

13) Tobler, W. R. (1979), Smooth Pycnophylactic Interpolation for Geographical, Regions, *Journal of the American Statistical Association*, 74, pp.519-530.

14) Wright, J. K. (1936), A Method of Mapping Densities of Population with Cape Cod as an Example, *Geographical Review* 26, pp103-110.

15) 川上一郎 (2007) : 『数値計算の基礎』, <http://www7.OCN.ne.jp/kawai/>.

16) 児玉慎三・須田信英 (1978) : 『システム制御のため

のマトリクス理論』, コロナ社.

17) 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, pp.69-78.

18) 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, pp.75-81.

19) 瀬谷創 (2008) : 「ヘドニック・アプローチを用いた便益評価への空間統計モデルの適用に関する基礎的研究」, 筑波大学大学院システム情報工学研究科修士論文.

20) 堤盛人・清水英範・井出祐史 (2000) : 「誤差要素モデルに基づく Kriging を用いた空間内挿」, 『応用力学論文集』, 3, pp.125-132.

21) 塚井誠人・奥村誠 (2007) : 「事業所立地の地域間依存構造のグローバル化に関する統計的検証」, 第15回応用地域学会研究発表会,

<http://www.cneas.tohoku.ac.jp/labs/dsese/kaken0.pdf>.

(2010.XX.XX 受付)

(2010.XX.XX 受理)

A countermeasure against a change in the areal unit for statistical data brought about by municipal mergers: An analysis by spatial econometric models

Abstract: The Great Merger of the Heisei Era has resulted in great changes in the areal units for Japanese municipal data. These changes have led to difficulties in panel data analysis. The most conventional measure today is to build a data set on the basis of the merged municipalities. However, such a measure may destroy some information that the smaller areal units before the municipal mergers might have conveyed. This paper proposes SLM based method that enables us to analyze municipal data by employing spatial econometric models on the basis of the areal units before the Great Merger, and examines the practicability of the proposed method.

市町村合併による統計データの集計単位変更 に対する方策の提案

空間計量経済モデルを用いた分析への対処法

筑波大学大学院システム情報工学研究科
博士前期課程 村上大輔
准教授 堤盛人

1

社会経済データの特徴

背景

- 空間的な集計単位毎のデータとしての提供が多い
 - 都道府県、市町村など

しばしば変更される

我が国の市町村区分は平成の大合併に伴い大きく変化

市町村数の変化

3232 → 1772

(1999年3月31日) (2011年12月1日)



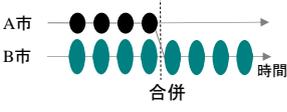
茨城県周辺の合併の有無
(黄色: 合併有り、白: 合併なし)

集計単位変更への対処

背景

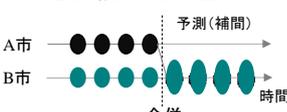
- 分析を容易にするために、集計単位を統一することが多い

(1): 合併後の行政区分への統一



利点: 作業が容易
欠点: 情報量が減少

(2): 合併前の行政区分への統一



利点: 細かい単位での分析が可能
欠点: 予測誤差が発生

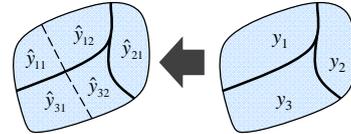
3

合併前の行政区分への統一に必要な処理

背景

- 合併後に集計されたデータを合併前の集計単位へと再配分するような手法が必要

面補間 (Areal Interpolation)



体積保存則: 合併前後で変数の総量は保存されなくてはならない

$$\hat{y}_{11} + \hat{y}_{12} = y_1$$

$$\hat{y}_{21} = y_2$$

$$\hat{y}_{31} + \hat{y}_{32} = y_3 \quad *yが人口の場合の条件式$$

4

面補間で考慮すべき空間データの性質

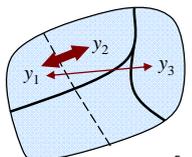
背景

□空間従属性を考慮した手法

↑

近接したデータは相互に似た傾向を示すという空間データの基本的性質

体積保存則と空間従属性の両方を考慮した面補間法は提案されていない



5

研究目標

研究目標

体積保存則及び空間従属性を考慮した
面補間法の提案

□提案の流れ

線形回帰モデルによる面補間法
体積保存則○
空間従属性×

空間計量経済モデルによる点補間法
体積保存則×
空間従属性○

↓

空間計量経済モデルによる面補間法
体積保存則○
空間従属性○

↑

面補間実行のための前提条件満足の確認

6

空間計量経済学

空間計量経済モデルによる点補間

- 空間従属性を考慮することのできるモデルが多数存在

■ 空間重み行列Wにより地点間の空間従属性を考慮する

$$W = \begin{pmatrix} 0 & w_{12} & w_{13} & \dots & w_{1n} \\ w_{21} & 0 & w_{23} & \dots & w_{2n} \\ w_{31} & w_{32} & 0 & \dots & w_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \dots & 0 \end{pmatrix}$$

- 各要素は距離の関数
(例) $w_{ij} = 1/d_{ij}^2$
- 行和が1になるように基準化
 $\sum_{j=1}^n w_{ij} = 1$
- 対角項は全て0

地区3から地区1への影響力

代表的な空間計量経済モデル

空間計量経済モデルによる点補間

□ 空間ラグモデル(SLM: Spatial Lag Model)

- 被説明変数に対して空間従属性を仮定

$$y = \rho W y + X \beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I)$$

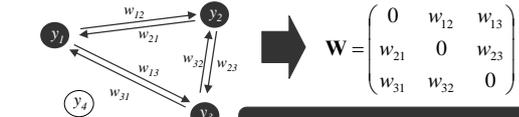
観測地点間のみについての空間重み行列

y : 被説明変数 X : 説明変数 ε : 誤差項
 I : 単位行列 ρ, β, σ^2 : パラメータ

空間計量経済モデルによる点補間の問題

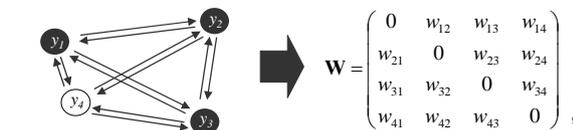
空間計量経済モデルによる点補間

□ 観測地点間のみについての空間重み行列



● : 観測地点 ○ : 予測地点

□ (観測+予測)地点についての空間重み行列



空間計量経済モデルによる点補間

空間計量経済モデルによる点補間

□ Pace and Lesage.(2006)

- SLMを(観測+予測)地点のデータを元に定義

$$y = \rho W y + X \beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I)$$

y : (観測+予測)地点の被説明変数 X : (観測+予測)地点の説明変数
 ε : 誤差項 I : 単位行列 ρ, β, σ^2 : パラメータ

欠損があり、通常的手法によるパラメータ推定ができない

EMアルゴリズムにより計算

EMアルゴリズムによる点補間

空間計量経済モデルによる点補間

- 欠損のあるデータから最尤推定を行うことのできるアルゴリズム

□ 基本的な計算手順

- (1) 欠損の初期値を設定
- (2) (観測+欠損)データに基づく尤度の最大化でパラメータを推定
- (3) 欠損値の更新
- (4) (2), (3)を収束するまで交互に繰り返す

□ Pace and Lesage.(2006)におけるEMアルゴリズム

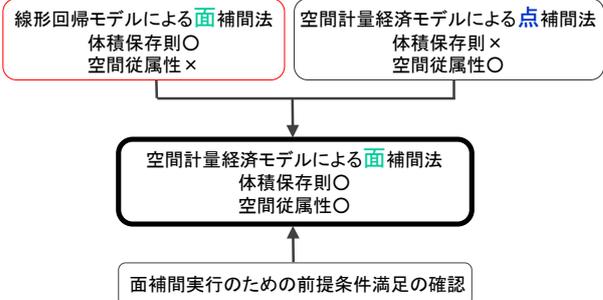
- SLMから構築される以下の尤度関数Lを手順(2)で最大化

$$L = -\frac{s}{2} \ln(2\pi) - \frac{s}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y - \rho W y - X \beta)' (y - \rho W y - X \beta) + \ln |I - \rho W|$$

s : (観測+欠損)地点数

この手法により面補間した場合、体積保存則は満足されない

面補間法提案までの流れ



体積保存則の整理

線形回帰モデルによる面補間法

面補間前後で変数の総量は保存する

- 密度を表す変数の面補間に焦点を当てる
(例) 人口に対する割合: 高齢化率、出生率、人口増加率 . . .
面積に対する割合: 人口密度、道路密度、. . .

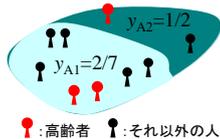
□ 密度を表す変数の体積保存則の成立条件

(例) 市町村A1及びA2の合併によりできた市町村Aの高齢化率

$$y_A = n_{A1}y_{A1} + n_{A2}y_{A2}$$

y_i : 空間単位*i*の高齢化率

n_{ij} : 合併後単位*j*の総人口のうち、
合併前単位*i*に所属する人口の割合



13

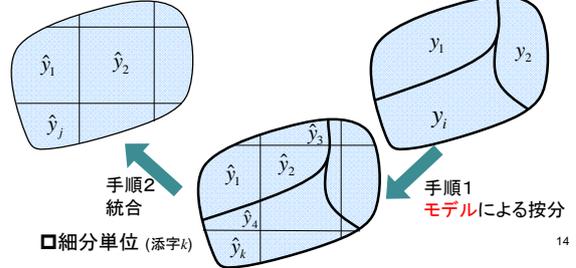
Flowerdew and Green.(1992)の面補間概要

線形回帰モデルによる面補間法

- 線形回帰モデルを用いた面補間を行った研究
- 集計単位から配分単位への面補間法を定式化

□ 配分単位 (添字*j*)

□ 集計単位 (添字*i*)

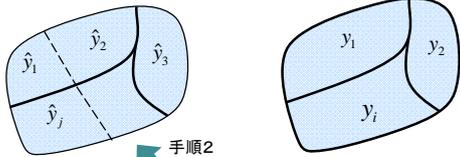


14

Flowerdew and Green.(1992)による市町村合併への対処

線形回帰モデルによる面補間法

□ 配分単位 (=合併前単位) (添字*j*) □ 集計単位 (=合併後単位) (添字*i*)



□ 細分単位 (=合併前単位) (添字*k*) □ 手順1 モデルによる按分

15

Flowerdew and Green.(1992)のモデル (以下、F&Gモデル)

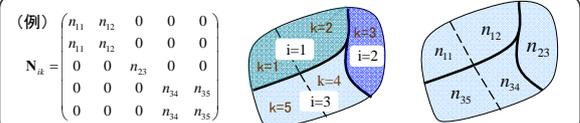
線形回帰モデルによる面補間法

$$y_k = \mathbf{X}_k \boldsymbol{\beta} + (\bar{z}_i - \mathbf{N}_{ik} \mathbf{X}_k \boldsymbol{\beta}) + \varepsilon_k$$

トレンド項 (定数項なし) 補正項

予測値 \hat{y}_i が体積保存則を満たすための補正

\bar{z}_i : 細分単位が所属する集計単位における被説明変数の実測値 \mathbf{N}_{ik} : 集計単位*i*の総人口のうち、細分単位*k*に所属する人の割合 n_{ik} を要素とする行列



■ y_k は未知であるためEMアルゴリズムで計算実行

16

面補間法提案までの流れ

線形回帰モデルによる面補間法
体積保存則○
空間従属性×

空間計量経済モデルによる点補間法
体積保存則×
空間従属性○

空間計量経済モデルによる面補間法
体積保存則○
空間従属性○

面補間実行のための前提条件満足の確認

17

空間従属性を考慮した面補間法

空間計量経済モデルによる面補間法

■ F&Gモデル

$$y_k = \mathbf{X}_k \boldsymbol{\beta} + (\bar{z}_i - \mathbf{N}_{ik} \mathbf{X}_k \boldsymbol{\beta}) + \varepsilon_k$$

トレンド項 補正項

$$y_k = \rho \mathbf{W}_k y_k + \mathbf{X}_k \boldsymbol{\beta} + (\bar{z}_i - \mathbf{N}_{ik} (\rho \mathbf{W}_k y_k + \mathbf{X}_k \boldsymbol{\beta})) + \varepsilon_k$$

トレンド項 補正項

$$\mathbf{W}_k^* = (\mathbf{I} - \mathbf{N}_{ik}) \mathbf{W}_k \quad \mathbf{X}_k^* = (\mathbf{I} - \mathbf{N}_{ik}) \mathbf{X}_k$$

$$y_k = \rho \mathbf{W}_k^* y_k + \mathbf{X}_k^* \boldsymbol{\beta} + \bar{z}_i + \varepsilon_k$$

18

尤度関数の構築

空間計量経済モデルによる面補間法

$$L^* = -\frac{K}{2} \ln(2\pi) - \frac{K}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y_k - W_k^* y_k - X_k^* \beta - \bar{z}_i)' (y_k - W_k^* y_k - X_k^* \beta - \bar{z}_i) + \ln |\mathbf{I} - \rho W_k^*|$$

ρに関する集中化対数尤度関数

$$L_C^* = -\frac{n}{2} \ln \left[\frac{1}{n} (\hat{\epsilon}_1 - \rho \hat{\epsilon}_{II} - \hat{\epsilon}_{III})' (\hat{\epsilon}_1 - \rho \hat{\epsilon}_{II} - \hat{\epsilon}_{III}) \right] + \ln |\mathbf{I} - \rho W_k^*| + const.$$

$$\begin{aligned} \hat{\epsilon}_1 &= y_k - X_k' \hat{\beta}_1 & \hat{\epsilon}_{II} &= W_k' y_k - X_k' \hat{\beta}_{II} & \hat{\epsilon}_{III} &= \bar{z}_i - X_k' \hat{\beta}_{III} \\ \hat{\beta}_1 &= (X_k' X_k)^{-1} X_k' y_k & \hat{\beta}_{II} &= (X_k' X_k)^{-1} X_k' W_k' y_k & \hat{\beta}_{III} &= (X_k' X_k)^{-1} X_k' \bar{z}_i \\ \hat{\beta} &= \hat{\beta}_1 - \rho \hat{\beta}_{II} & \sigma^2 &= \frac{(\hat{\epsilon}_1 - \rho \hat{\epsilon}_{II})' (\hat{\epsilon}_1 - \rho \hat{\epsilon}_{II})}{s} \end{aligned}$$

19

EMにおける尤度最大化の手順

空間計量経済モデルによる面補間法

- 通常のSLMと同様
- [1] OLSを用いて y_k を $(\mathbf{I} - N_{ik}) X_k$ に回帰し $\hat{\beta}_1$ を得る
- [2] OLSを用いて $(\mathbf{I} - N_{ik}) W_k y_k$ を $(\mathbf{I} - N_{ik}) X_k$ に回帰し $\hat{\beta}_{II}$ を得る
- [3] OLSを用いて \bar{z}_i を $(\mathbf{I} - N_{ik}) X_k$ に回帰し $\hat{\beta}_{III}$ を得る
- [4] [1][2][3]の回帰の残差 $\hat{\epsilon}_1, \hat{\epsilon}_{II}, \hat{\epsilon}_{III}$ を求める
- [5] L_C^* を最大にする $\hat{\rho}$ を求める
- [6] $\hat{\beta}$ と $\hat{\sigma}^2$ を得る

20

面補間法提案までの流れ

線形回帰モデルによる面補間法
体積保存則○
空間従属性×

空間計量経済モデルによる点補間法
体積保存則×
空間従属性○

空間計量経済モデルによる面補間法
体積保存則○
空間従属性○

面補間実行のための前提条件満足の確認

21

尤度最大化の必要条件

面補間実行のための前提条件満足の確認

$$L_C^* = -\frac{n}{2} \ln \left[\frac{1}{n} (\hat{\epsilon}_1 - \rho \hat{\epsilon}_{II} - \hat{\epsilon}_{III})' (\hat{\epsilon}_1 - \rho \hat{\epsilon}_{II} - \hat{\epsilon}_{III}) \right] + \ln |\mathbf{I} - \rho W_k^*| + const.$$

$$|\mathbf{I} - \rho W_k^*| = |\mathbf{I} - \rho W_k + \rho N_{ik} W_k| > 0 \text{ でなくてはならない}$$

22

$\mathbf{I} - \rho W_k + \rho N_{ik} W_k \equiv \mathbf{D}$ の性質

面補間実行のための前提条件満足の確認

$$\mathbf{I} - \rho W_k + \rho N_{ik} W_k \equiv \mathbf{D} \rightarrow \begin{pmatrix} t_{11} & t_{12} & t_{13} & t_{14} & t_{15} \\ t_{21} & t_{22} & t_{23} & t_{24} & t_{25} \\ t_{31} & t_{32} & t_{33} & t_{34} & t_{35} \\ t_{41} & t_{42} & t_{43} & t_{44} & t_{45} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 \\ 0 & \mathbf{I} \end{pmatrix}$$

t_{ik} : 何らかの数

必ず単位行列となる

以下の展開が可能

$$|\mathbf{D}| = \begin{vmatrix} \mathbf{D}_1 & \mathbf{D}_2 \\ 0 & \mathbf{I} \end{vmatrix} = |\mathbf{D}_1|$$

$|\mathbf{D}| > 0$ を示すには $|\mathbf{D}_1| > 0$ を示せばよい

23

$|\mathbf{D}_1| > 0$ が満足される条件

面補間実行のための前提条件満足の確認

\mathbf{D}_1 が正定値行列であればよい

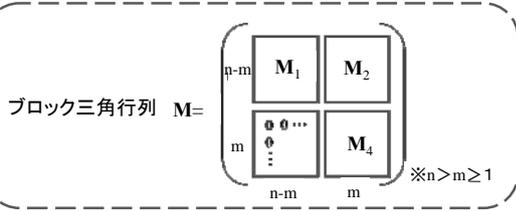
- 正定値行列の十分条件(Burga.(1972),(塚井, 奥村.(2007)参照))
 - (1) 既約行列であること
 - (2) 狭義優対角行列であること
 - (3) 全ての対角項が実数であること ← 自明

24

(1)既約行列とは

面補間実行のための前提条件満足の確認

- 行列の基本変形によりブロック三角行列への変換が不可能な行列

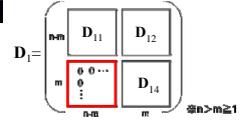


25

(1) D_1 が既約行列であることの証明(背理法)

□既約行列でない条件

- $(n-m) \times m$ の零行列が生成可能であること



■零行列生成の必要条件1

- $(n-m-1) \times m$ 組の D_1 の要素 q, q' について $w_{pq} = w_{p'q}$ が成立すること

■零行列生成の必要条件2

- 細分単位 q を含まない同一集計単位に p, p' が属しているならば、必ず $w_{pq} = w_{p'q}$ となるような q が $(n-m)$ 地点存在すること

∥

代表点の座標が同一の地区が複数存在することが既約でない条件

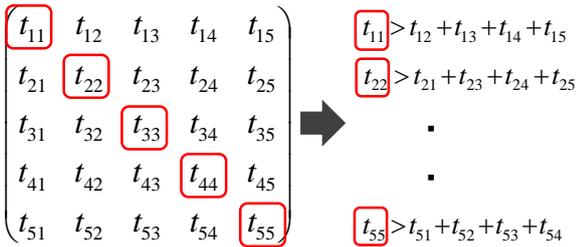
市町村データで考える際は成り立たないため D_1 は既約行列

26

(2)狭義優対角行列とは

面補間実行のための前提条件満足の確認

- 各行について、対角要素が対角要素以外の全要素の和よりも大きい値を示す行列



27

(2) D_1 が狭義優対角行列であることの証明

面補間実行のための前提条件満足の確認

- D の成分を元に、 D_1 についての検証を行う
※ただし $\rho \geq 0$ とする

Dの成分	対角要素	非対角要素の和
I	1	0
$-\rho W_k$	0	$-\rho$
$\rho N_{ik} W_k$	0以上 ρ 未満	0以上 ρ 未満

Wに基標準化を仮定するため

D	1以上 $1+\rho$ 未満	>	$-\rho$ 以上0未満
D_1	1以上 $1+\rho$ 未満	>	$-\rho$ 以上0未満

D_1 は狭義優対角行列

28

提案手法の計算手順まとめ

提案手法提案のまとめ

□提案モデル

$$y_k = \rho W_k y_k + X_k \beta + (\bar{z}_i - N_{ik} (\rho W_k y_k + X_k \beta)) + \varepsilon_k$$

□細分地区予測の手法

■EMアルゴリズム

- 手順 (1) 欠損箇所に初期値を設定
(2) (観測+欠損)データに基づく尤度最大化によりパラメータを推定
(3) 欠損箇所の値の更新

※以下の尤度関数を最大化

$$L = \frac{K}{2} \ln(2\pi) - \frac{K}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (y_k - W_k y_k - X_k \beta - \bar{z}_i) (y_k - W_k y_k - X_k \beta - \bar{z}_i) + \ln |1 - \rho W_k|$$

29

提案手法による実証

実証

2007年の高齢者のデータを
1995年時点の市町村区分へと面補間する

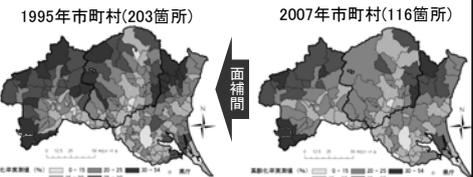
□予測精度の検証の際の真値

- 2007年の町丁目別データを1995年の市町村区分に集計して獲得

□対象地域

- 北関東3県

茨城県
栃木県
群馬県



2007年度高齢化率(ともに町丁目データから作成)

比較するモデル

実証

線形モデル $y_k = X_k \beta + \varepsilon_k$

F&Gモデル (体積保存則考慮) $y_k = X_k \beta + (\bar{z}_i - N_{ik} X_k \beta) + \varepsilon_k$

提案モデル (体積保存則+空間従属性考慮) $y_k = \rho W_k y_k + X_k \beta + (\bar{z}_i - N_{ik} (\rho W_k y_k + X_k \beta)) + \varepsilon_k$

被説明変数

- ・ 高齢化率(%)

説明変数

- ※相関が示唆される変数を投入
- ・ 東京駅までの距離[km]
- ・ 県庁所在市の主要駅までの距離[km]
- ・ 人口密度[千人/km²]
- ・ 域外従業率[%]
- (他の市町村で働く従業者の割合)

- ・ 空間重み行列Wの要素は $w_{ij} = 1/d_{ij}^2$

31

パラメータ推定結果

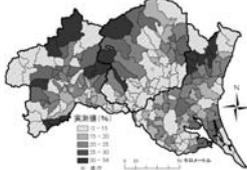
実証

ケース	1995年区分への面補間			1995年区分の全データ により推定(参考)	
	線形 モデル	F&G モデル	提案 モデル	線形 モデル	空間ラグ モデル
定数項	21.80 ***			18.53 ***	2.32 ***
東京距離[km]	0.04 ***	0.11 ***	-0.06***	0.07 ***	0.02
県庁距離[km]	0.05 ***	0.15 ***	0.11***	0.06 ***	0.04 *
人口密度[千人/km ²]	-0.07 **	-0.15 ***	-0.08	-0.33 ***	-0.23 ***
域外従業率[割]	-0.04	0.03	0.05	-0.06 ***	-0.05
ρ			7.29		0.92
AIC		624.6	588.1		

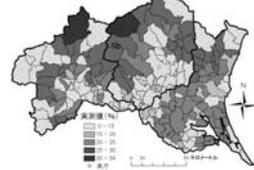
面補間結果

実証

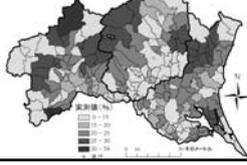
高齢化率実測値



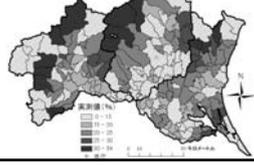
線形モデル



F&Gモデル



提案モデル



33

面補間の精度比較

実証

- ・ RMSE(平均二乗平方根誤差)により面補間の精度を比較

$$RMSE = 100 \times \sqrt{\frac{\sum_{k'} (y'_k - \hat{y}'_k)^2}{K'}}$$

結果

	線形モデル	F&Gモデル	提案モデル
RMSE (%)	6.11	3.78	3.46

- : 体積保存則
 - : 空間従属性
- の考慮による精度向上

34

まとめ

まとめ

- ・ 体積保存則と空間従属性を満足する面補間法を空間計量経済モデルを元に提案
- ・ 空間従属性を考慮することによる予測精度向上を確認

今後の課題

- ・ 尤度関数内の行列式についての証明の精緻化
- ・ 幅広いデータへの適用
- ・ 時空間への拡張

35