【原著論文】

Kriging を用いた実用的な面補間法

村上大輔*・堤盛人**

A kriging based practical approach for areal interpolation

Daisuke Murakami*, Morito Tsutsumi**

Abstract: Spatial unit convergence, called areal interpolation, is often applied to address this problem. Of the many areal interpolation methods that have been proposed, a few consider spatial autocorrelation which is the general property of spatial data. Here, a new areal interpolation method that considers spatial autocorrelation is constructed by combining linear regression based areal interpolation method and kriging: a spatial statistical method. The proposed method, whose implementation and extension is easy, is practical among spatial statistics based areal interpolation methods. Then, a case study showed that the suggested method succeeds in improving the predictive accuracy and indicated that considering spatial autocorrelation is important for accurate areal interpolation.

Keywords: 面補間 (areal interpolation), 空間統計学 (spatial statistics), 空間的相関 (spatial autocorrelation), 体積保存則 (pycnophylactic property), クリギング (kriging), 回帰法 (regression based areal interpolation method)

1. はじめに

空間データの多くは、行政区やメッシュなどの空間的な単位毎に集計された後に提供される。しかしながら、提供されるデータの集計単位は、データ作成者や配布者の都合で決定されることが多く、必ずしも分析者の望む空間的単位であるとは限らないため (貞広, 2000), 必要に応じ、データの空間的な単位の変換をしなければならない。そのような変換は面補間 (Areal interpolation) と呼ばれる (例えば Sadahiro, 1999)。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。

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より面補間を行う面

積按分法 (Wright, 1936) は、最も代表的な面補間法の一つである。また、衛星画像等の入手が容易となった近年においては、補助的なデータを考慮できる手法の研究が盛んであり、その精度の高さが報告されている (例えば Fisher and Langfold, 1995)。

ところで、空間データの持つ基本的な性質として、空間的相関 (Spatial autocorrelation, 例えば Cressie, 1993) が知られている。空間的相関とは、空間的に近接したデータ間には類似した傾向が見られるという性質であり、面補間に際しても考慮されるべき性質である。しかしながら、空間的相関を考慮し、かつ面補間法が満足すべき最も基本的な性質とされる体積保存則 (詳しくは2章) を満足した手法は非常に限られている。また、空間的相関と体積保存則を考慮した既存の面補間法は、いずれも実用性の面で課題が残されている。

そこで本研究では、空間的相関を考慮する方法として空間統計学を応用し、なおかつ体積保存則を満たす、実用的な面補間法を提案する。

以降、まず第2章で面補間について概説した後、第3章では面補間法の一つであり本研究で扱う回帰

図1 面補間の概要

* 学生会員 筑波大学大学院システム情報工学研究科 (University of Tsukuba)
〒304-8753 茨城県つくば市天王台1-1-1 E-mail: muraka51@sk.tsukuba.ac.jp

** 正会員 筑波大学大学院システム情報工学研究科 (University of Tsukuba)

法について詳しく説明する。そして第4章では、本研究において空間的相関を考慮するために用いる空間統計学の基本的な手法である kriging について説明した後、第5章において、kriging を回帰法に援用した新たな面補間法を提案し、第6章において提案手法を実際のデータに適用してその実用性を確認する。

2. 面補間の概要

空間データの補間に関する研究は、地点で観測されたデータに基づく地点の補間（本研究では、点補間と呼ぶこととする）と、空間単位内の集計データに基づく空間単位内の集計データの補間、即ち、面補間の二つに大別することができる。面補間では、計数データ (count data)、即ち、個々の要素を数え上げたデータへの適用が前提とされることが多い。

一般に、点補間では (1) 式が、面補間では (2) 式が、補間の対象となる変数（以後、対象変数と称する）を表現する関数に仮定される。

$$y_s = f(s, \mathbf{X}) \quad (1)$$

$$y_B = \int_{s \in B} f(s, \mathbf{X}) ds \quad (2)$$

ここで、 $s \in \mathbf{R}^2$ は地点、 y_s は地点 s の対象変数、 $\mathbf{X}$ は対象変数を説明する補助的なデータからなる行列である。また、 B は配分単位の領域、 y_B は配分単位 B の対象変数を表す。

集計データへの点補間の適用に際しては、まず、集計単位毎に与えられたデータを各集計単位の代表点のデータに置き換え、次に、代表点毎の対象変数に関数 $f(s, \mathbf{X})$ をあてはめ、最後に、推定された関数 $f(s, \mathbf{X})$ を用いて、地点 s の対象変数 y_s を補間することが多い。一方、面補間の適用に際しては、まず、集計単位毎のデータ $\bar{y}_A$ (実測値) を、関数 $f(s, \mathbf{X})$ の集計値とみなした上で $f(s, \mathbf{X})$ を推定し、次に、推定された $f(s, \mathbf{X})$ を配分単位毎に集計することで、配分単位のデータ y_B を補間する。従って、配分単位毎の補間値からの逆算により与えられる集計単位毎の値は、各集計単位の実測値と等しくならなくてはならない。例えば、 a 市の人口を用いた面補間により、同市を構成する b 地区と c 地区の人口を求めた場合、

両地区の補間値の和は a 市の人口の総計と等しくならなくてはならない。このような性質は『体積保存則 (Pycnophylactic property ; Tobler, 1979)』と呼ばれ (3) 式で表される。

$$\bar{y}_A = \int_{s \in A} f(s, \mathbf{X}) ds \quad (3)$$

ここで、 A は集計単位の領域である。体積保存則は面補間で満足すべき最も基本的な性質であり (Tobler, 1979)、実用上のニーズも高いことから、本論では、「体積保存則が満足されることを前提とした、集計単位から配分単位へのデータの再配分」、即ち、(2)・(3) の両式に基づいた補間を面補間と定義して議論を進める。

現在までに様々な面補間法が提案されており、近年においては、補助的なデータ（即ち、 $\mathbf{X}$ ）を考慮することのできる手法の構築が積極的に試みられている。例えば、衛星画像データを考慮する dasymetric 法 (Fisher and Langfold, 1995) と、統計モデルに基づいて集計データを考慮する回帰法 (Flowerdew and Green, 1992) はその代表的な例である。回帰法に代表される統計学に基づいた面補間法は補間の精度を統計学的に示すことができる点で優れている。

近年、空間的相関を明示的に扱う応用統計学の分野として、空間統計学（ここでは鉱山学を起源とした狭義の空間統計学 (Spatial statistics: 例えば Cressie, 1993)）と空間計量経済学 (Spatial econometrics: 例えば LeSage and Pace, 2009) が目覚ましい発展を遂げている。しかしながら、両分野に基づいた面補間の研究は非常に限定的である (Gelfand, 2010)。例外として、Kyriakidis (2004)、Wikle and Berliner (2005)、Yoo and Kyriakidis (2006)、Gotway and Young (2007)、Yoo et al. (2010) は (a) 空間統計学に基づく面補間法を、堤・村上 (2010) は (b) 空間計量経済学に基づく面補間法を提案した。(a) の長所には期待二乗誤差の最小化に基づいた面補間ができる点、(b) の長所には空間的相関のモデリングに幅広い距離指標が適用できる点等がある。(a)、(b) の手法の多くは、両補間で対象とされることの多い計数データではなく密度を表すデータ（密度データと呼称することとする）を対象とする。これは、空間統計学、空間計量経済学が、通

常、密度データのような連続確率分布に従うデータを対象とするためである。ここで、密度データに対する体積保存則満足条件式は、密度データの元となる計数データが体積保存則を満足するための条件式である (3) 式で与えられる。従って、例えば、人口密度が体積保存則を満足する十分条件は、その元となる人口が体積保存則を満足することである。

(a) 空間統計学に基づく面補間法の短所の一つに実用性の欠如がある。近年においては、空間統計学の手法を適用するための多数の統計パッケージが提供されており (例えば、フリーの統計ソフトである「R」であれば `gstat`, `geoR` 等), 空間統計学の標準的な手法の実装は容易となりつつある。しかしながら、(a) の各手法は、いずれも、空間統計学の標準的なモデルとは異なるモデルを基本モデルするため、従来の空間統計モデルと同様の手順でそれらを実装・拡張することはできない。

本研究では、空間統計学の最も基本的なモデルである `kriging` (例えば, Cressie, 1993) を回帰法に援用することで、空間的相関と体積保存則を考慮し、かつ従来の `kriging` と同様の手順で実装・拡張を行うことのできる実用的な面補間法を提案する。以降で提案する手法は、空間統計学に基づいた多くの既存研究と同様に、密度データを対象とすることとする。また、密度データの予測値が非負となること、及びの予測値からの変換で得られる計数データが非負の整数となることは、今回提案する手法では必ずしも保証されないことを予め述べておく。

3. 回帰法

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』と定義し、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、各空間単位の対象変数ベクトルは $\bar{\mathbf{y}}_i = \{\bar{y}_1, \bar{y}_2, \dots, \bar{y}_I\}$, $\mathbf{y}_j = \{y_1, y_2, \dots, y_J\}$, $\hat{\mathbf{y}}_k = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_K\}$ 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 , $K \times 1$ である。

回帰法 (Flowerdew and Green, 1992) は、欠損のあるデータを用いて最尤推定を行うことのできる EM (Expectation-Maximization) アルゴリズム (Dempster et al., 1977) を用いて密度データの面補間を行う手法である。ここで、EM アルゴリズムとは、以下の手順に従って尤度を最大化する手法である。

[i] 未知パラメータに適当な初期値 θ^0 を与える。

[ii] 以下のステップを、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 を元に各欠損データの条件付き期待値を計算し、擬似的な観測データとして欠損部分に代入する。

M-Step

L^h を最大化するようなパラメータ θ^{h+1} を求める。

以上より、最尤推定量と予測値が算出される。通常、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。

回帰法は以下の手順で密度データの面補間を行う。

[1] 集計単位及び配分単位を元に細分単位を生成。

[2] 線形回帰モデルに基づいたモデルに EM アルゴリズム (手順 [i], [ii] に基づいた計算) を適用することで、細分単位のデータを予測。

[3] [2] の結果得られた細分単位の予測値を集計することで、配分単位の予測値を算出。

手順 [2] で算出される予測値が体積保存則を満足するための条件式は (4) 式で与えられる。

$$\bar{y}_i = \sum_k \frac{n_{ik}}{n_i} \hat{y}_k \quad (4)$$

図2 回帰法の計算の概要

n_i は集計単位 i の重みである。 n_{ik} は、集計単位 i に対する細分単位 k の重みを表し、細分単位 k が集計単位 i に包含される場合に正值、それ以外の場合に 0 をとる。例えば、面積を分母に持つ変数（人口密度など）を補間する場合の n_i と n_{ik} は面積（従って n_{ik}/n_i は A_i の総面積に占める C_k の面積の割合）、人口を分母に持つ変数（老年人口など）を補間する場合の n_i と n_{ik} は人口（従って n_{ik}/n_i は A_i の総人口に占める C_k の人口の割合）で与えられる。(4) 式は、密度データが体積保存則を満足するための必要十分条件が、「 n_{ik}/n_i で重みづけられた細分単位毎の予測値の加重平均と、集計単位毎の実測値が等しくなること」、であることを示している。

Flowerdew and Green (1992) では行列表記を用いず体積保存則を示しているが、第 5 章で示す本研究の提案手法を簡潔に説明するための準備として、ここでは (4) 式を (5) 式のとおり行列表記する。

$$\bar{\mathbf{y}}_i = \mathbf{N}_{ik}^* \hat{\mathbf{y}}_k \quad (5)$$

$\mathbf{N}_{ik}^*$ は $I \times K$ の集計行列であり、その i 行 k 列目の要素は、集計単位 i に対する細分単位 k の重み n_{ik}/n_i で与えられる。 n_{ik}/n_i は、集計単位 i が細分単位 k を包含する場合に正值、それ以外の場合に 0 をとる。図 2 に示す面補間の場合の $\mathbf{N}_{ik}^*$ を模式的に表すと (6) 式となる。

$$\mathbf{N}_{ik}^* = \begin{pmatrix} n_{11}/n_1 & n_{12}/n_1 & n_{13}/n_1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & n_{24}/n_2 & n_{25}/n_2 & \\ \vdots & & & & & \ddots \end{pmatrix} \quad (6)$$

(5) 式は、体積保存則を満足するためには、細分単位毎に与えられた $\hat{\mathbf{y}}_k$ を集計単位毎に集計した値が、集計単位毎の観測値と等しくならなくてはならないということを表す条件式である。

細分単位毎の $\hat{\mathbf{y}}_k$ を予測するためには、集計単位毎の I 本の条件式から成る (5) 式を、細分単位毎の K 本の条件式に変換する必要がある。ここで、細分単位 k の予測値が体積保存則を満足することは、即ち、細分単位 k を包含する集計単位 i の観測値が体積保存則を満足することである。従って、細分単位 k の予測値 $\hat{\mathbf{y}}_k$ に対する条件式を (5) 式の i 行目の式で与えることで、細分単位毎の K 本の条件式 (7) 式が構築される。

$$\bar{\mathbf{y}}_i | (A_i \supseteq C_k) \neq \emptyset = \mathbf{N}_{ik} \hat{\mathbf{y}}_k \quad (7)$$

$\bar{\mathbf{y}}_i | (A_i \supseteq C_k) \neq \emptyset$ は、細分単位 k を包含する集計単位 i の観測値 $\bar{\mathbf{y}}_i$ から成る $K \times 1$ のベクトルである（例えば、 $\bar{\mathbf{y}}_i | (A_i \supseteq C_k) \neq \emptyset$ の第 3 要素は 3 番目の細分単位を包含する集計単位 i で得られた観測値 $\bar{y}_i$ ）。 $\mathbf{N}_{ik}$ は $K \times K$ の集計行列であり（集計単位 i 毎の観測値を細分単位 k 毎に与えるための行列であるため $\mathbf{N}_{ik}$ と表記することとした）、その k 行 k' 列目の要素は、細分単位 k' を包含する集計単位 i' の、細分単位 k に対する重み $n_{i'k}/n_{i'}$ で与えられる。 $n_{i'k}/n_{i'}$ は、細分単位 k が集計単位 i' に包含される場合に正值、それ以外の場合に 0 をとる。以上の定式化の結果として、 $\mathbf{N}_{ik}$ は I 個のブロックを持つブロック対角行列となり、その第 i ブロック内の第 k'' 列の各要素は、一様に $n_{i,k''}/n_{i''}$ 、即ち、細分単位 k'' に対する集計単位 i の重みで与えられる。図 2 に示す面補間の場合の $\mathbf{N}_{ik}$ を模式的に表すと、(8) 式のように表される。

$$\mathbf{N}_{ik} = \begin{pmatrix} n_{11}/n_1 & n_{12}/n_1 & n_{13}/n_1 & 0 & 0 & \cdots \\ n_{11}/n_1 & n_{12}/n_1 & n_{13}/n_1 & 0 & 0 & \\ n_{11}/n_1 & n_{12}/n_1 & n_{13}/n_1 & 0 & 0 & \\ 0 & 0 & 0 & n_{24}/n_2 & n_{25}/n_2 & \\ 0 & 0 & 0 & n_{24}/n_2 & n_{25}/n_2 & \\ \vdots & & & & & \ddots \end{pmatrix} \quad (8)$$

(7) 式の第 k 番目の式は、細分単位 k の予測値 $\hat{\mathbf{y}}_k$ が体積保存則を満足するための条件式を表す。

手順 [2] では、線形回帰モデルに対する制約として体積保存則の条件式 (7) 式を与えた (9) 式を用いて細分単位毎の密度データ $\hat{\mathbf{y}}_k$ を予測する。

$$\hat{\mathbf{y}}_k = \mathbf{X}_k \boldsymbol{\beta} + [\bar{\mathbf{y}}_i | (A_i \supseteq C_k) \neq \emptyset - \mathbf{N}_{ik} \mathbf{X}_k \boldsymbol{\beta}] + \boldsymbol{\varepsilon}_k \quad \boldsymbol{\varepsilon}_k \sim N(\mathbf{0}, \mathbf{I}_k) \quad (9)$$

p を説明変数の数とすると、 $\mathbf{X}_k$ は細分単位に関する $K \times p$ の説明変数ベクトル、 $\boldsymbol{\beta}$ は $p \times 1$ のパラメータベクトルである。(9) 式を用いた予測は EM アルゴリズムに基づいて行う。具体的には、まず、体積保存則の条件式 (7) を満足する適当な初期値を各細分単位に与え、次に、((9) 式ではなく) 通常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と、(9) 式を用いた $\hat{\mathbf{y}}_k$ の更新を、収束するまで交互に繰り返すことで、予測を行う。以上の手順により得られる予測値 (量) $\hat{\mathbf{y}}_k$ は体積保存則満足のための条件式 (7) を必ず満足する。しかしながら、一般に、EM アルゴリズムとは、同一の式から構築される尤

度関数と予測式を用いるアルゴリズムである (小西ほか, 2008). 従って, 通常線形回帰モデルの尤度関数と (9) 式に示す予測式を用いる回帰法は, 従来の EM アルゴリズムとの理論的な整合性に問題もある.

ここで, 定義上, N_{ik} の行和は必ず 1, かつその第 i 行は集計単位 i に包含される細分単位についての要素のみが正值となる. 従って, 集計単位 i に包含される各細分単位が常に同一の値を示すような説明変数を用いた場合, (9) 式内の β が掛かる項である $(\mathbf{X}_k - \mathbf{N}_{ik}\mathbf{X}_k)\beta$ は常に 0 となり, その結果, $\hat{y}_k$ の予測値は $\bar{y}_i | (A_i \supseteq C_k) \neq \phi$ に一致するため実質的な予測とならない. 従って, そのような特徴を持つ説明変数 (及び定数項) を用いることはできない.

手順 [3] では, 手順 [2] で得られた各細分単位の予測値 $\hat{y}_k$ を (10) 式に代入することで, 配分単位 B_j の予測値 $\hat{y}_j$ を算出する.

$$y_j = \sum_k \frac{n_{jk}}{n_j} \hat{y}_k \quad (10)$$

n_j は配分単位 B_j の重み, n_{jk} は配分単位 j に対する細分単位 k の重みを表す. 従って, n_{jk}/n_j は, 配分単位 j が細分単位 k を包含する場合に正值, それ以外の場合に 0 をとる.

手順 [3] の細分単位から配分単位への集計は, (10) 式より予測誤差なしに行うことができたため, 以降では, 手順 [2] の集計単位から細分単位へのデータの配分 (面補間) のみに注目する.

4. クリギング (Kriging)

空間統計学の基本的な手法である kriging は, ユークリッド空間上の地点 $s \in \mathbf{R}^2$ で観測されたデータ (サンプル数 S) の空間過程をモデル化する手法であり, (11) 式を基本式とする.

$$\mathbf{y}_s = \mathbf{X}_s \beta + \boldsymbol{\varepsilon}_s \quad (11)$$

$\mathbf{y}_s$ は $S \times 1$ の被説明変数ベクトル, $\mathbf{X}_s$ は $S \times p$ の説明変数行列, β は $p \times 1$ のパラメータベクトル, $\boldsymbol{\varepsilon}_s$ は $S \times 1$ の局所的変動を表すベクトルをそれぞれ表す.

Kriging は, $\boldsymbol{\varepsilon}_s$ の共分散を距離の関数 (共分散関数) で与えることで空間的相関を考慮する. 共分散関数には, 例えば指数型共分散関数 (12) 式がある.

$$c(d_{s,s'}) = \begin{cases} \sigma^2 \exp\left[-\left(\frac{d_{s,s'}}{w}\right)\right] & (s \neq s') \\ \tau^2 + \sigma^2 & \text{otherwise} \end{cases} \quad (12)$$

ここで, σ^2 は partial-sill, τ^2 は nugget, w は range と呼ばれるパラメータである (e.g. Cressie, 1993).

Kriging は点補間 (任意地点のデータの予測) に適用できる. ユークリッド空間上の地点 $\mathbf{s}_0 \in \mathbf{R}^2$ の被説明変数 y_0 の予測量 $\hat{y}_0$ は, $\mathbf{y}_s$ が与えられた下での $\hat{y}_0$ の条件付き期待値として (13) 式で与えられる.

$$\hat{y}_0 = E[y_0 | \mathbf{y}_p] = \mathbf{x}'_0 \hat{\beta} + \mathbf{c}'_0 \mathbf{C}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\beta}) \quad (13)$$

$$\hat{\beta} = (\mathbf{X}_s \mathbf{C}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}_s \mathbf{C}_s^{-1} \mathbf{y}_s$$

$\mathbf{x}_0$ は $p \times 1$ の説明変数ベクトル, $\mathbf{C}_s$ は観測地点間の共分散行列 ($S \times S$), $\mathbf{c}_0$ は観測地点と予測地点との間の共分散ベクトル ($S \times 1$) をそれぞれ表す.

5. Kriging に基づく新たな面補間法

細分単位を十分に細密な空間単位で定義した上で, 細分単位毎の局所的変動 $\boldsymbol{\varepsilon}_k$ の共分散を距離の関数 (共分散関数) で与えることで, 空間的相関を考慮することとする. ここで, 例えば, 細分単位をメッシュデータや衛星画像データ等に基づいて与えることで, それらを説明変数として適用できる (図 3).

5.1. モデルの構築

十分に細密な細分単位毎に与えられたデータ $\hat{y}_k$ が kriging の基本式 (14) 式に従うと仮定する.

$$\hat{y}_k = \mathbf{X}_k \beta + \boldsymbol{\varepsilon}_k \quad \boldsymbol{\varepsilon}_k \sim N(0, \mathbf{C}_k) \quad (14)$$

ここで, 細分単位毎に得られる $\hat{y}_k$ が体積保存則を満足するためには, 条件式 (7) 式を満足しなくてはならない. (14) 式の制約として (7) 式を与えることで, (15) 式が導出される.

$$\hat{y}_k = \mathbf{X}_k \beta + [\bar{y}_i | (A_i \supseteq C_k) \neq \phi - \mathbf{N}_{ik}(\mathbf{X}_k \beta + \boldsymbol{\varepsilon}_k)] + \boldsymbol{\varepsilon}_k$$

図3 細密な細分単位

$$\epsilon_k \sim N(0, C_k) \quad (15)$$

(15) 式より算出される $\hat{y}_k$ は、空間的相関を考慮し、また、体積保存則を満足する。従って、(9) 式の代わりに (15) 式を用いることで、体積保存則と空間的相関を考慮した面補間ができる。

5.2. 尤度関数の構築

(15) 式の尤度関数を導出するために、まず、(15) 式を (16) 式のように展開する。

$$(\mathbf{I}-\mathbf{N}_{ik})\epsilon_k = (\hat{y}_k - \bar{y}_i |_{(A_i \supseteq C_k) \neq \phi}) - (\mathbf{I}-\mathbf{N}_{ik})\mathbf{X}_k\beta \quad (16)$$

ここで、(8) 式に例示したように、同一の集計単位に包含される細分単位についての $\mathbf{N}_{ik}$ の各行は同一であり、その結果として $\text{rank}(\mathbf{N}_{ik}) < K$ となるため $(\mathbf{I}-\mathbf{N}_{ik})$ は特異行列となり、 ϵ_k が一意に求まらない。そこで、ここで何らかの制約を課して (例えば、追加的に設定したコスト関数の最小化に基づいて) 逆行列を与えることとする。そのような逆行列は一般化逆行列と呼ばれる。以降では、 $(\mathbf{I}-\mathbf{N}_{ik})$ の一般化逆行列 $(\mathbf{I}-\mathbf{N}_{ik})^+$ を用いて (16) 式を (17) 式に書き換えることで ϵ_k を求めることとする。

$$\epsilon_k = (\mathbf{I}-\mathbf{N}_{ik})^+(\hat{y}_k - \bar{y}_i |_{(A_i \supseteq C_k) \neq \phi}) - (\mathbf{I}-\mathbf{N}_{ik})^+(\mathbf{I}-\mathbf{N}_{ik})\mathbf{X}_k\beta \quad (17)$$

多くの場合、一般化逆行列は、誤差平方和 ((16) 式の場合、両辺の差の誤差平方和) や、ノルム ((17) 式の場合、 $\epsilon_k' \epsilon_k$) の最小化に基づき決定される (例えば、土木学会, 2000)。 (16) 式に関しては、誤差平方和の最小化は (16) 式から (17) 式への近似の精度を高める点から、ノルムの最小化は面補間の予測誤差を小さくする点から、それぞれ重要である。そこで、本研究では両方が同時に最小化される Moore-Penrose (M-P) 型一般化逆行列で $(\mathbf{I}-\mathbf{N}_{ik})^+$ を与えることとする。行列 $\mathbf{M}$ に対する M-P 型一般化逆行列 $\mathbf{M}^+$ は (18) 式で与えられる。

$$\mathbf{M}^+ = \mathbf{M}'(\mathbf{M}'\mathbf{M}\mathbf{M}'\mathbf{M})\mathbf{M}' \quad (18)$$

ここで、(16) 式の、いずれかの集計単位と同一である細分単位に係る行に着目すると、 $(\mathbf{I}-\mathbf{N}_{ik})$ と $(\hat{y}_k - \bar{y}_i |_{(A_i \supseteq C_k) \neq \phi})$ の同行は必ず 0 となるため、(16) 式の等号は常に成立する。即ち、M-P 型一般化逆行列の目的関数の一つである誤差平方和は常に 0 となる。従って、M-P 型一般化逆行列のもう一つの目的

関数であるノルムの最小化の結果として、 ϵ_k は必ず 0 となる。必ず 0 となる ϵ_k の共分散は、共分散が距離に依存するという kriging の仮定に整合しない。従って、ここでは、5 章 1 節で議論したような十分細かい空間スケールで細分単位を定義することで、集計単位と細分単位の一致を防ぐこととする (図 3)。ここでの対処は細分単位の設定を要求するものであり、提案手法をより汎用的な手法とするためには、今後、集計単位と細分単位に一致がある場合にも適用できる手法へと、提案手法を拡張することが望ましい。

次に、(17) 式の尤度関数を構築する。従来の kriging (11) 式の対数尤度関数は (19) 式で与えられる (e.g. Cressie, 1993)。

$$L = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(\mathbf{y}_s - \mathbf{X}_s\beta)' \mathbf{C}_s^{-1} (\mathbf{y}_s - \mathbf{X}_s\beta) - \frac{1}{2} \ln |\mathbf{C}_s| \quad (19)$$

提案モデル (17) 式は、kriging の基本式 (11) 式の被説明変数ベクトルを $\mathbf{a}_k = (\mathbf{I}-\mathbf{N}_{ik})^+(\hat{y}_k - \bar{y}_i |_{(A_i \supseteq C_k) \neq \phi})$ 、説明変数行列を $\mathbf{B}_k = (\mathbf{I}-\mathbf{N}_{ik})^+(\mathbf{I}-\mathbf{N}_{ik})\mathbf{X}_k$ で与えたモデルに相当する。従って、通常の kriging の基本式 (11) 式、その尤度関数 (19) 式、及び (17) 式から、提案モデルの対数尤度関数は (20) 式となる。

$$L_k \approx -\frac{K}{2} \log(2\pi) - \frac{K}{2} \log |\mathbf{C}_k| - \frac{1}{2} (\mathbf{a}_k - \mathbf{B}_k\beta)' \mathbf{C}_k^{-1} (\mathbf{a}_k - \mathbf{B}_k\beta) \quad (20)$$

提案モデルの最尤推定量は、従来の kriging と同様の手順で得られる。具体的には、まず、(20) 式で与えられる β の最尤推定量が得られ、次に、最尤推定量 $\hat{\beta}$ を (21) 式に代入することで得られる集中化対数尤度化数を最大化することで、 σ^2 , τ^2 , w の最尤推定量が得られる。

$$\hat{\beta} = (\mathbf{B}_k' \mathbf{C}_k^{-1} \mathbf{B}_k)^{-1} \mathbf{B}_k' \mathbf{C}_k^{-1} \mathbf{a}_k \quad (21)$$

5.3. 予測式の構築

(11), (13), (17) 式より、提案モデルの予測量は、更新前の $\hat{y}_k$ が与えられた下での更新後の $\hat{y}_k$ を表す $\hat{y}_{0k}$ 条件付き期待値として (22) 式で与えられる。

$$E[\mathbf{y}_{0k}] = \bar{y}_i |_{(A_i \supseteq C_k) \neq \phi} + (\mathbf{I}-\mathbf{N}_{ik})\{\mathbf{X}_k\hat{\beta} + \mathbf{C}_{0k}\mathbf{C}_k^{-1}(\mathbf{a}_k - \mathbf{B}_k\hat{\beta})\} \quad (22)$$

$\mathbf{C}_{0k}$ は予測値を更新する前後の ε_k 間の相互共分散を表す $K \times K$ の行列である。 $\mathbf{C}_{0k}$ の要素を与える関数を定義するために、ここでは、更新前後の ε_k が同一の空間相関構造に従うこと、及び更新後の予測値 $\hat{y}_{0k}$ は微視的変動を表す τ^2 (nugget) を持たないことを仮定する。従来の kriging を用いた点補間に際しても、観測値と予測値の ε_k が同一の空間相関構造に従うことと、予測値が τ^2 を持たないことをそれぞれ仮定することから、上記の仮定は自然である。以上の仮定の結果として、 $\mathbf{C}_{0k}$ の各要素は、(自己) 共分散関数から nugget を排除した関数 (例えば指数型共分散関数の場合は (23) 式) で与えられる。

$$c(d_{k,k'}) = \sigma^2 \exp\left[-\frac{d_{k,k'}}{w}\right] \quad (23)$$

5. 4. 提案手法を用いた面補間の手順

$\hat{y}_k$ の初期値を $\bar{y}_i | (A_i \supseteq C_k) \neq \phi$ (細分単位を包含する集計単位の観測値から成るベクトル) で与え、それを更新することで面補間を行うことを仮定する。この仮定と (7) 式より、 $\mathbf{a}_k$ は (24) 式で与えられる。

$$\mathbf{a}_k^* = (\mathbf{I} - \mathbf{N}_{ik})^+ (\mathbf{I} - \mathbf{N}_{ik}) \bar{y}_i | (A_i \supseteq C_k) \neq \phi \quad (24)$$

その結果、対数尤度関数と予測式は (20), (22) 式内の $\mathbf{a}_k$ を $\mathbf{a}_k^*$ で置き換えた式で与えられることとなる。

上述の仮定の下で集計単位から細分単位への面補間を行う手順は以下の通りである。

- [1] $\hat{y}_k$ の初期値を $\bar{y}_i | (A_i \supseteq C_k) \neq \phi$ で与える。
- [2] $\mathbf{a}_k$ を $\mathbf{a}_k^*$ で置き換えた対数尤度関数 (20) 式の最大化を行うことでパラメータを推定する。
- [3] $\mathbf{a}_k$ を $\mathbf{a}_k^*$ で置き換えた予測式 (22) 式に推定されたパラメータを代入する。

以上より、体積保存則を満足し、かつ、空間的相関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、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。

ここで、提案手法は、 $\bar{y}_i | (A_i \supseteq C_k) \neq \phi$ を観測データ、 $\hat{y}_k$ を予測対象とした手法であり、パラメータ推定には観測データ $\bar{y}_i | (A_i \supseteq C_k) \neq \phi$ のみを用いるため、一度の計算で予測値が求まる。一方、提案手法の元となる手法である回帰法は、尤度関数と予測式を異なるモデ

ルから構築しており (詳しくは第 3 章)、その結果として、観測データ $\bar{y}_i | (A_i \supseteq C_k) \neq \phi$ と欠損データ $\hat{y}_k$ の両方がパラメータ推定に必要となるため、欠損のあるデータのための最尤推定アルゴリズムである EM アルゴリズムを用いた繰り返し計算が必要となる。

ここで、第 3 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ならない。これは従来の回帰法と同様の特徴である。

5. 5. 提案手法の特徴

Kriging に基づいた既存の面補間法 (Kyriakidis (2004), Wikle and Berliner (2005), Yoo and Kyriakidis (2006), Gotway and Young (2007), Yoo et al. (2010)) は、期待二乗誤差の最小化に基づいて面補間を行うことができる点を長所に持つ。一方で、提案手法は厳密には期待二乗誤差を最小化しない。これは、疑似的な逆行列である一般化逆行列を適用したためである。また、提案手法以外の各手法が $\hat{y}_k$ の期待二乗誤差を最小化するのに対し、提案手法は、(20) 式内の $(\mathbf{I} - \mathbf{N}_{ik})^+ (\hat{y}_k - \bar{y}_i | (A_i \supseteq C_k) \neq \phi)$ の期待二乗誤差を最小化する。従って、提案手法の優位性を予測精度の観点から示すには、期待二乗誤差最小化の基準の違いが予測精度に与える影響を理論と実証の両面から検証する必要があるが、このことについては、機会を改めて考察したい。

提案手法の基本モデル (17) 式は、kriging の基本式 (11) 式の被説明変数ベクトルを $\mathbf{a}_k = (\mathbf{I} - \mathbf{N}_{ik})^+ (\hat{y}_k - \bar{y}_i | (A_i \supseteq C_k) \neq \phi)$ 、説明変数行列を $\mathbf{B}_k = (\mathbf{I} - \mathbf{N}_{ik})^+ (\mathbf{I} - \mathbf{N}_{ik}) \mathbf{X}_k$ で与えたモデルとみなすことができる。即ち、提案モデルは、従来の kriging と同一である。従って、提案モデルは従来の kriging と同様の手順で実装・拡張ができる。kriging に基づいたその他の面補間法は、いずれも従来の kriging とは異なる。従って、提案手法は kriging に基づいた他の手法よりも実用的である。また、面積以外の指標 (人口など) に基づいて体積保存則満足のための制約を与えることができる点で Wikle and Berliner (2005) と Gotway and Young (2007) よりも汎用性が高く、また、説明変数を考慮できる点で Wikle and Berliner (2005) よりも

汎用性が高い。

6. 実証分析：提案手法の東京 23 区の人口データの面補間への適用

6.1. 分析の概要

東京都23区から1km メッシュ、5km メッシュへの夜間人口（出典：総務省統計局提供の平成17年度国勢調査の夜間人口）の面補間に、空間的相関、体積保存則を考慮した提案手法 [Spatial] と体積保存則のみを考慮した回帰法 [Non-spatial] をそれぞれ適用し、両手法の予測結果の比較等から、提案手法の有用性を検証する。両メッシュに対する補間に際しての細分単位は、ともに東京都23区と1km の重ね合わせで定義される空間単位で与える。なお、ここでは人口密度を対象変数として補間を行い、人口密度の予測値を変換することで人口の予測値を算出する。

補間の精度を検証するためには、(i) 1km メッシュ別の夜間人口データと (ii) 5km メッシュ別の同データが必要であり、(i) は平成17年度国勢調査の1km メッシュ別夜間人口で、(ii) は同データを5km メッシュ毎に集計することで、それぞれ与えることとする。ここで、平成17年度国勢調査のメッシュデータは、基本単位区と呼ばれる空間単位毎別のデータを面補間することで作成されており誤差を含む。しかしながら、東京都内の基本単位区（都内に172,788単位）は1km メッシュ（都内に2,549単位）に比べて十分に細密であり、この誤差は極めて小さいと示唆される。従って、本分析では、各メッシュデータを持つ誤差を無視することとする。

説明変数は、最寄り駅までの距離 [最寄り駅距離]、最寄り駅から最寄りの主要駅（ここでは東京駅、新宿駅、池袋駅、渋谷駅、品川駅で与える）までの距離 [主要駅距離]、最寄りの住宅地公示地価の対数 [対数地価] とする。各変数は、国土交通省国土計画局提供の国土数値情報から作成しており、距離はすべて直線距離を用いた。また、[Spatial] で用いる共分散関数は指数型共分散関数 (12) 式で与えた。

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、予測結果等の空間分布図作成には ArcGIS を使用する。

6.2. パラメータ推計結果

パラメータ推定結果を表1に示す。表から、[最寄り駅距離] 及び [対数地価] は、両モデルについて5% 水準以内で有意に負となり、符号は直観に整合する。一方で、[Spatial] を用いた推定では、[主要駅距離] もまた1% 水準で有意に負となるのに対し、[Non-Spatial] では有意とならないことがわかる。主要駅までの距離増加に伴うアクセシビリティの低下は、人口の低下を招くと考えられることから、[Spatial] の推定結果は、より直観に整合する。また、 w の推定値が3.80であることあることは、空間的相関の及ぶ距離が約11.4 ($=3.80 \times 3$) kmであることを示している。

次に、提案手法で適用する M-P 型一般化逆行列 $(\mathbf{I}-\mathbf{N}_{ik})^+$ の精度を、(24) 式で定義される Dirichlet スプレッド関数（例えば、メンケ、1989）を用いて検証した。

$$\text{spread}(\mathbf{I}-\mathbf{N}_{ik}) = \|(\mathbf{I}-\mathbf{N}_{ik})^+ (\mathbf{I}-\mathbf{N}_{ik}) - \mathbf{I}\|^2 \quad (24)$$

計算の結果、(24) 式の値は1.00となった。これは、 $(\mathbf{I}-\mathbf{N}_{ik})^+ (\mathbf{I}-\mathbf{N}_{ik})$ と $\mathbf{I}$ の単位要素あたりの乖離が0.0011程度であり、少なくとも、本実証分析で用いる一般化逆行列である $(\mathbf{I}-\mathbf{N}_{ik})^+$ は、精度が極めて高いとの示唆を得た。このことは、提案手法が、近似的に期待二乗誤差を最小化することも同時に示している。

6.3. 予測精度の比較

平均二乗平方根誤差 (RMSE : Root Mean Square Error) を用いて予測精度の比較を行う。両モデルの RMSE の算出結果を表2に示す。表より、1km 及び

表1：パラメータ推定結果

パラメータ	Non-Spatial	Spatial
最寄り駅距離 (km)	$-2.69 \times 10^{**}$ (-2.31)	$-1.49 \times 10^{***}$ (-12.19)
主要駅距離 (km)	1.43×10 (1.07)	$-5.49 \times 10^{***}$ (-4.09)
対数地価 (万円 / m ²)	$-2.09 \times 10^{2**}$ (-2.04)	$-8.26 \times 10^{2***}$ (-8.58)
Nugget : τ^2	—	0.01
Partial-sill : σ^2	—	0.01
Range : w	—	3.80

¹⁾ 括弧内の数値はt値を表す。

²⁾ ***は1%, **は5%, *は10%水準で有意であることを表す。

5km メッシュの両方の面補間について、空間的相関を考慮することによる予測精度の向上が確認できる。

両モデルを用いた1km 及び5km メッシュへの面

補間の結果を図5、6に示す。両モデルの補間結果の傾向として、5km メッシュでは実測値に近い結果が得られたのに対し、1km メッシュでは郊外部が過度に平滑化されていることが分かる。このことから、[Spatial] 及び [Non-spatial] の補間の精度は配分単位の空間スケールに強く依存することが示唆される。

次に、(25) 式で定義される誤差率を用いて、メッシュ毎の予測精度を比較する。

表2：RMSE 算出結果 (千人)

	Non-Spatial	Spatial
5kmメッシュ	103	95.2
1kmメッシュ	1.27	1.20

図4：実際の人口の分布

図5：5km メッシュ別の人口補間結果の概要

図6：1km メッシュ別の人口補間結果の概要

$$[\text{誤差率}] = 100 \times \left| \frac{y_k - \hat{y}_k}{y_k} \right| \quad (25)$$

誤差率のプロットを図5, 6に示す。両図から、東京都西部で [Spatial] がより高い精度を示していることがわかるが、これは、同地域が比較的均質であり、kriging で仮定する定常性の仮定に整合的であったためと示唆される。その一方で、東京都中心部及び東部では [Spatial] を適用した方が、予測精度が低い、これは、同地区が丸の内地区等の異質性の強い地区を含み、定常性の仮定に整合しなかったためと示唆される。

7. おわりに

本研究では、空間統計学の代表的な手法である kriging を拡張することで、体積保存則及び空間的相関の両方を同時に考慮した面補間法を提案した。また、提案手法は、kriging に基づいた他の手法に比べて実用的であることを明らかにした。最後に、実証分析から、面補間に際しての空間的相関を考慮することの重要性を、予測精度の観点から示した。

その一方で、提案手法ではスケールの小さな配分単位への補間の結果が過度に平滑化されることが異質性の強いと示唆される地域の精度が低くなることが課題として明らかとなった。それらの原因として、空間的な異質性や局所的な空間的相関構造を考慮していないことが挙げられるため、それらを考慮して面補間を行う方法について、今後、検討が必要である。また、本研究の実証分析は、回帰法との比較分析を行ったに過ぎず、提案手法の有用性を、実用性だけでなく予測精度の観点からも示すためには、幅広い面補間法との比較、特に、空間統計学に基づいた他の面補間法との比較を行うことが必要である。

謝辞

本稿をまとめるにあたり、2名の匿名の査読者の方から、それぞれ有益なご指摘を頂いた。ここに記して感謝を申し上げます。なお、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

- 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.
- 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.
- 土木学会(2000) 『土木工学における逆問題入門』, 土木学会.
- メンケ, W. 編, 柳田俊・塚田和彦訳(1997) 『離散インバース問題』, 古今書房. Menke, W. eds. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press.
- Cressie, N. (1993) *Statistics for Spatial Data*. Revised Edition, John Wiley & Sons.
- 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.
- Fisher, P. F. and Langford, M. (1995) Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, **27**, 211–224.
- Flowerdew, R. and Green, M. (1992) Developments in areal interpolation methods and GIS. *Annals of Regional Science*, **26**, 67–78.
- Gelfand, A. E. (2010) Misaligned Spatial Data; The Change of Support Problem. In Gelfand, A. E. and Banerjee, S. eds. *Handbook of Spatial Statistics*, Boca Raton, FL: Chapman & Hall/CRC, **29**, 495–515.
- 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.
- Kyriakidis P. C. (2004) A Geostatistical Framework for Area-to-Point Spatial Interpolation. *Geographical Analysis*, **36**(3), 259–289.
- LeSage, J. A. and Pace, R. K. (2009) *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.
- Tobler W. (1979) Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Sta-*

tistical Association, **74**(367), 519–530.

Sadahiro, Y. (1999) Accuracy of Areal Interpolation: A Comparison of Alternative Methods. *Journal of Geographical Systems*, **1**(4), 323–346.

Wikle, C.K. and Berliner, L.M. (2005) Combining information across spatial scales. *Technometrics*, **47**(1), 80–91.

Wright, J. K. (1936), A Method of Mapping Densities of Population with Cape Cod as an Example. *Geographical Review*, **26**, 103–110.

Yoo, E.-H. and Kyriakidis P.C. (2006) Area-to-point kriging with inequality-type data. *Journal of Geographical Systems*, **8**(4), 357–390.

Yoo, E.-H., Kyriakidis, P.C. and Tobler, W. (2010). Reconstructing population density surfaces from areal data: A comparison of Tobler's pycnophylactic interpolation method and area-to-point kriging. *Geographical Analysis*, **42**(1), 78–98.

(2011年8月31日原稿受理, 2011年11月28日採用決定, 2011年12月15日デジタルライブラリ掲載)