地理的重み付き回帰分析と Kriging 法による都市モデルのための空間データの補間

栃木県　正会員　○ 川合 健司
東北大学　正会員　Varameth Vichiensan
東北大学　正会員　Antonio Paez
東北大学　フェロー　宮本 和明

1 はじめに

都市モデルを適用する場合、必ずしも必要な分析単位によるデータが存在するとは限らない。また、同じデータであっても、年度によって集計単位が異なる場合も少なくない。そのような場合、観測されているデータを利用し、異なる地点もしくは領域における未知データを補間する必要が生じる。精度の高い補間を実現するためには、誤差相関や分散不均一性といった空間的諸問題を適切にモデル化することが必要である。

そこで本研究では、最良線形不偏予測による一般化補間モデルに基づき、分散共分散行列の定義による空間データ補間の統一化を行う。そして、誤差の相関を Kriging のコバリオグラム関数によって、分散不均一性を Geographically Weighted Regression によって同時に考慮した、新しい空間データ補間モデルを提案する。そして実データを用いた補間を行い、既存モデルと推定精度を比較する。

2 既存補間手法

(1) 一般化補間モデル

空間内の任意地点 o における空間データ y_o の最良線形不偏推定量は、観測値の分散共分散行列 $\mathbf{V}$、観測値と推定値の共分散ベクトル $\mathbf{c}$、観測値の誤差 $\boldsymbol{\varepsilon}$ を用いて式(1)のように与えられる[1)]。

$$\hat{y}_o = \mathbf{x}'_o\hat{\boldsymbol{\beta}}_{GLS} + \mathbf{c}'\mathbf{V}^{-1}\boldsymbol{\varepsilon} \tag{1}$$

本研究では、最良線形不偏予測に基づく式(1)を一般化補間モデルと位置付ける。そして、既存の手法を分散共分散行列の定義によって整理することを試みる。

(2) Kriging

Kriging の特徴は、誤差項に対する定常性の仮定であり[2)]、誤差項 $\boldsymbol{\varepsilon}$ の共分散に関して式(2)を仮定する。

$$Cov(\varepsilon_i, \varepsilon_j) = C(R, d_{ij}) \tag{2}$$

ここで、d_{ij} とは任意地点 i、 j 間の距離であり、共分散を定義する関数 C はコバリオグラムと呼ばれる。観測データを用いて推定された共分散関係は、定常性の仮定により全ての地点間において成り立つため、式(1)による補間を行うことができる。しかし、その仮定のために、全ての地点における分散は均一となる。

キーワード：空間データ補間、Kriging、GWR
〒980-8576　仙台市青葉区川内
東北大学大学院工学研究科土木工学専攻
TEL 022-217-7568　FAX 022-217-7477

(3) GWR

GWR とは、例えば式(3)のように移動平均型に構造化された誤差項の確率項 $\boldsymbol{\mu}$に対し、推定地点 o と観測地点 i との距離 d_{oi} を用い、式(4)、(5)のような分布を仮定する手法である[3)]。

$$\boldsymbol{\varepsilon} = \lambda\mathbf{W}\boldsymbol{\mu} + \boldsymbol{\mu} \tag{3}$$

$$\boldsymbol{\mu} \sim N(0, \boldsymbol{\Omega}) \tag{4}$$

$$\omega_{ij} = \begin{cases} \sigma_o^2 g_{oi}(\gamma_o, d_{oi}) = \sigma_o^2 exp(\gamma_o d_{oi}^2) & (i = j) \\ 0 & (i \neq j) \end{cases} \tag{5}$$

式(3)による誤差項の構造化の結果、観測値の分散共分散行列 $\mathbf{V}$ の各要素 v は、空間重み付き行列 $\mathbf{W}$ の各要素 w によって、式(6)のように定式化される。

$$v_{ij} = \begin{cases} \sigma_o^2\left\{g_{oi} + \lambda^2\sum_{k=1}^{N} w_{ik}w_{ki}g_{ok}\right\} & (i = j) \\ \sigma_o^2\left\{\lambda(w_{ij}g_{oi} + w_{ji}g_{oj}) + \lambda^2\sum_{k=1}^{N} w_{ik}w_{kj}g_{ok}\right\} & (i \neq j) \end{cases} \tag{6}$$

GWR によって定式化される観測値の分散および共分散は、着目する 2 点間の直接的な影響だけでなく、他の全ての地点 k を介した影響によって表現される。そのため観測値によって推定された共分散関係は、着目した観測値間のみに成立する関係であり、式(1)を用いた任意地点の補間を行うことができない。

3 誤差相関と不均一分散を考慮した補間モデルの提案

そこで、本研究では共分散をコバリオグラム関数で、分散を GWR によって直接定義するモデルを提案する。すなわち任意地点 o に対する観測値の分散共分散行列 $\mathbf{V}_o$ を、2 つの関数を用いて式(7)のように定義する。

$$\mathbf{V}_o = \sigma_o^2\begin{bmatrix} g_{o1} & & C_{ij} \\ & \ddots & \\ C_{ij} & & g_{on} \end{bmatrix} \tag{7}$$

GWR によって対角項を直接的に定義することによって、推定地点ごとに誤差構造を変化させることができ、より詳細な空間的特徴を捉えることが可能となる。また、コバリオグラム関数を用いることによって、共分散が任意 2 地点間の距離のみの関数となり、観測値と推定値の共分散 c_o も式(8)のように与えることが可能

となる。そのため、式(1)による補間が可能となる。

$$\mathbf{c}_o = \sigma_o^2 \left[C(R_o, d_{o1}), \cdots, C(R_o, d_{on}) \right] \tag{8}$$

4 実データを用いた検証

(1) 適用対象

平成 3 年度札幌市都市計画基礎調査データから 883 個の小ゾーン切り出し、さらに小ゾーンを結合し、ゾーン数が等しい(202 ゾーン)2 つのゾーンシステムを作成した。補間対象は人口データとし、住宅棟数、商業棟数、工業棟数をトレンドの説明変数として用いる。

適用するモデルは誤差構造の異なる、OLS、GWR(誤差相関なし)、Kriging、提案モデルの 4 つとする。一般化モデルにより各手法を分類すると、**表-1** のようになる。GWR と提案モデルの分散項は式(5)の関数によって定義し、Kriging と提案モデルの共分散項は spherical 型のコバリオグラムによって定義する。

表-1 適用するモデル

	V	c	予測式
OLS	$\sigma^2 \begin{bmatrix} 1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 1 \end{bmatrix}$	$\mathbf{0}$	$\mathbf{x}_o\hat{\boldsymbol{\beta}}$
GWR	$\sigma_o^2 \begin{bmatrix} g_{o1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g_{on} \end{bmatrix}$	$\mathbf{0}$	$\mathbf{x}_o\hat{\boldsymbol{\beta}}_o$
Kriging	$\sigma^2 \begin{bmatrix} 1 & & C_{ij} \\ & \ddots & \\ C_{ij} & & 1 \end{bmatrix}$	$\sigma^2[C_{o1},\cdots,C_{on}]$	$\mathbf{x}_o\hat{\boldsymbol{\beta}} + \hat{\mathbf{c}}\hat{\mathbf{V}}^{-1}\boldsymbol{\varepsilon}$
提案モデル	$\sigma_o^2 \begin{bmatrix} g_{o1} & & C_{ij} \\ & \ddots & \\ C_{ij} & & g_{on} \end{bmatrix}$	$\sigma_o^2[C_{o1},\cdots,C_{on}]$	$\mathbf{x}_o\hat{\boldsymbol{\beta}}_o + \hat{\mathbf{c}}_o\hat{\mathbf{V}}_o^{-1}\boldsymbol{\varepsilon}$

(2) 空間問題の診断

提案モデルの漸近的共分散行列に対する LM 検定によって、空間問題を診断した結果を**図-2** に示す。有意水準 5%において、少なくとも誤差相関を示す共分散パラメータ R か、不均一分散を示す分散パラメータ γ の存在が検出された。また、不均一分散と誤差相関を同時に考慮することは、全 202 ゾーン中 124 ゾーンにおいて有意であると判定された。

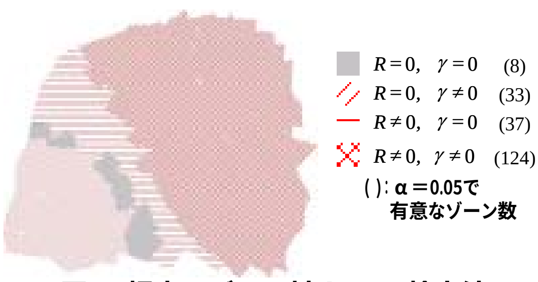

図-2 提案モデルに対する LM 検定結果

(3) モデルの推定

4 つのモデルを最尤法によってパラメータ推定した結果を**表-2** にまとめる。また、提案モデルの推定においては、分散共分散行列 **V** は正値定符号性を満足する必要があり、$\gamma \geq 0$ を制約条件とした。

表-2 パラメータ推定結果

	住宅棟数	商業棟数	工業棟数	定数項	σ^2	γ	R
OLS	4.29	24.16	-28.22	316.46	3.68×10^5	-	-
GWR	3.63	18.26	-28.94	255.55	1.20×10^5	-0.086	
	4.48	28.28	-20.95	399.90	8.00×10^5	0.108	-
	4.07	24.22	-24.37	315.76	3.24×10^5	0.018	
Kriging	5.32	17.69	-7.81	-48.09	8.50×10^5	-	1.79
提案モデル	4.61	12.09	-13.20	-48.28	6.58×10^5	0	1.79
	5.53	20.68	-4.60	162.18	8.50×10^5	0.045	2.65
	5.10	16.66	-7.44	6.69	5.37×10^5	0.018	2.01

注) GWRと提案モデルについては上段から，最小値・最大値・平均値

(4) 誤差評価

補間精度の評価指標として、クロスバリデーション、AIC、実データとの誤差平方和という 3 つの指標によってモデルを評価する。評価結果を**表-3** にまとめる。

空間問題を考慮しない OLS に対して、GWR によって異質性のみを考慮した場合、観測値に対するモデルの当てはめは改善されたが、未知データの補間精度は改善されなかった。しかし、Kriging によって誤差相関のみを考慮した場合には、補間精度は大きな改善を示した。さらに、提案モデルによって誤差相関と不均一分散を同時に考慮した場合、実データと比較した誤差分布は Kriging と同じ傾向を示したが、誤差の大きさ自体は減少し、誤差平方和で判断して最も良い結果を与えることができた。

表-3 誤差評価結果

	C.V.Score	AIC			誤差平方和
		min.	max.	average	
OLS	7.68×10^7	-	-	3183.8	6.86×10^7
GWR	7.30×10^7	3146.6	3185.6	3168.4	7.17×10^7
Kriging	6.27×10^7	-	-	3170.6	3.72×10^7
提案モデル	6.20×10^7	3109.7	3172.8	3151.1	3.35×10^7

5 おわりに

本研究では、最良線形不偏予測による一般化補間モデルによって補間モデルの統一化を試み、既存の Kriging と GWR の問題点を整理した。そして、誤差の相関をコバリオグラム関数によって、空間的異質性を GWR の不均一分散によって考慮した、新しい空間データ補間モデルを構築した。

実データに対して適用した結果、提案モデルは既存モデルに比べて精度の改善を示すことができた。しかし、補間精度は対象とする空間データの特性に大きく依存するため、関数や変数を適切に選択する必要があることも明らかになった。

参考文献

1) Arthur S.: Best Linear Unbiased Predictor in the Generalized Linear Regression Model, *American Statistical Association Journal*, 369-375, 1962
2) Chiles J.: *Geostatistics*, A Wiley Publication, 1998
3) Paez A.: A general framework for estimation and inference of geographically weighted regression models, *Environment and planning A*, 2002, forthcoming