

空間統計モデルを用いた公示地価・都道府県地価調査の評価地点の配置問題

村上大輔・堤盛人

Geostatistical Sampling Design for Land price Assessment Systems in Japan

Daisuke MURAKAMI and Morito TSUTSUMI

Abstract: 本研究では地価公示・都道府県地価調査の調査地点を効率的に削減する方法を議論する。具体的には、まず、観測点配置の議論が活発な空間統計学のモデルを用いて地価公示と都道府県地価調査の評価地点を適正に削減する方法を議論し、次にそこでの議論を踏まえて地価公示・都道府県地価調査の評価地点の削減し、その結果を議論する。

Keywords: サンプリングデザイン (Sampling design), 空間統計学 (Geostatistics), 地価公示 (Publication of Land Price), 都道府県地価調査 (Prefectural land price survey)

1. はじめに

近年の財政の逼迫を契機に地価調査事業の縮小への関心が高まりつつある。地価調査のための予算の削減は地価の評価地点を減らすことで容易に達成できる。しかしながら、地価データは土地取引や経済の主要な指標の一つであるため、評価地点の削減に伴うデータの質の低下は避けなければならない。

一般に、データ観測地点の効率的削減のような観測点配置を考える問題はサンプリングデザインと呼ばれ、近年、空間統計学（例えば Cressie, 1993）を中心に議論が進められている。そこで本研究では、空間統計学の知見に基づいて地価の評価地点を効率的に削減する方法を議論する。

2. 評価地点配置の最適化

2.1 配置の良さとは

観測点配置の良さは(i)観測点配置の持つ説明力で評価されることが多く（詳しくは次節）、また、これに加えてデータの(ii)需要や(iii)取得コストが考慮されることもある。地価の評価地点を配置する場合であれば、(i)は地価の特性をできるだけよく説明するような配置とするために重要であり、(ii)は利用者の需要に適合した配置とするために重要である。一方、地価データの取得コスト(iii)には大差はないと考えられる。従って、次節以降では(i)説明力と(ii)需要を考慮して地価データの評価地点配置を効率化する方法を議論する。

2.2 配置の説明力の評価

地価は空間統計モデル(1)式に従うものとする。

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{C}) \quad (1)$$

\mathbf{z} は地価実測値を要素に持つ $N \times 1$ のベクトル、 \mathbf{X} は $N \times K$ の説明変数行列、 $\boldsymbol{\varepsilon}$ は $N \times 1$ の乱項ベクトル、 $\mathbf{0}$ は $N \times 1$ のゼロベクトル、 \mathbf{C} は $N \times N$ の共分散行列、 $\boldsymbol{\beta}$ は $K \times 1$ のパラメータベクトルである。(1)式は共分散行列の各要素を距離減衰関数で与えることで地価データ相互間の空間的な依存性

村上大輔 〒304-8753 茨城県つくば市天王台 1-1-1

筑波大学 システム情報工学研究科

Phone: 029-853-5572

E-mail: muraka51@sk.tsukuba.ac.jp

を考慮する．距離減衰関数には例えば球型モデル(2)式が適用できる．

$$c(d_{ij}) = \begin{cases} \tau^2 + \sigma^2 & \text{if } r < d_{ij} \\ \tau^2 + \sigma^2 \left[\frac{3}{2} \left(\frac{d_{ij}}{r} \right) + \frac{1}{2} \left(\frac{d_{ij}}{r} \right)^3 \right] & \text{if } 0 < d_{ij} < r \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

(1)式を用いることで，地価の予測値や予測誤差分散を任意地点について算出できる．いま，地点0の地価予測値 z_0 が(3)式に従うと仮定する．

$$z_0 = \mathbf{x}_0' \boldsymbol{\beta} + \varepsilon_0 \quad (3)$$

「 $'$ 」は転置の演算子， \mathbf{x}_0 は地点0の説明変数から成る $K \times 1$ のベクトル， ε_0 は同地点の攪乱項である．不偏かつ線形という制約の下で地価 z_0 の予測誤差分散を最小化することで， z_0 の予測量 \hat{z}_0 とその予測誤差分散はそれぞれ(4)，(5)式となる．

$$\hat{z}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}} + \mathbf{c}' \mathbf{C}^{-1} (\mathbf{z} - \mathbf{X} \hat{\boldsymbol{\beta}}) \quad (4)$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{C}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{C}^{-1} \mathbf{z}$$

$$E[(z_0 - \hat{z}_0)^2] = \mathbf{C}(0) - \mathbf{c}' \mathbf{C}^{-1} \mathbf{c} +$$

$$(\mathbf{x}' - \mathbf{c}' \mathbf{C}^{-1} \mathbf{X})(\mathbf{X}' \mathbf{C}^{-1} \mathbf{X})^{-1} (\mathbf{x}' - \mathbf{c}' \mathbf{C}^{-1} \mathbf{X})' \quad (5)$$

\mathbf{c} は各観測地点と地点 i の間の共分散を表す $N \times 1$ のベクトル， $\mathbf{C}(0)$ は対角要素が $\sigma^2 + \tau^2$ ，非対角要素が0の $N \times N$ の行列である．

評価地点配置 S の説明力は予測誤差分散を元に評価できる．具体的には，例えば，対象地域内に稠密に配置した各地点（本研究では町丁目 i の各重点点を用いる）で求められた予測誤差分散の平均値や最大値は，（ある配置を持つ）データで説明されない攪乱の大きさを表わすため，これを最小化するような観測点配置 S^* はデータの特性を良く捉えた配置といえる．従って，本研究においても予測誤差分散を評価地点配置の説明力の指標として用いることとする．

2.3 配置の需要（以降，ある配置に従う各データの需要の総和を配置の需要と定義する）

配置の需要を評価需要する方法はデータに応じて変えるべきであり（Diggle, et al., 2010），観測点配置の需要を評価するための画一的な方法は

存在しない．幸い，我が国においては国土交通省提供のサイトである土地総合情報システム（<http://www.land.mlit.go.jp/webland/>）で土地取引の事例毎のデータが公表されており，このデータを元に町丁目毎の土地取引数が算出できる．従って，本研究では町丁目毎の取引数を，配置 S の需要の指標として用いることを考える．

ここで，例えば平成23年度の茨城県における住宅地取引数は3,557であり，町丁目（総数：3,945）あたりの取引数は極めて小さくなる．このような場合，町丁目毎の取引数は「たまたま対象年に取引が行われた」というような不確実な要因に大きく左右されうるため，町丁目毎の取引数を需要の指標として用いるためには，不確実性による影響をできる限り抑える必要がある．

そこで，本研究では経験ベイズ法に基づいて縮小推定を行うことでこの問題に対処することとする．まず，町丁目 i における取引数 d_i はポアソン分布(6)式の実現値であると仮定する．

$$d_i \sim \text{Poisson}(\theta_i \times \bar{d}_i) \quad (6)$$

\bar{d}_i は取引を生じうる母体の数を表わし，本研究ではこれを世帯数で与える． θ_i は町丁目 i の特性を表わすパラメータであり，上述のように \bar{d}_i を世帯数とした場合は，町丁目 i で期待される世帯あたり取引数となる．次に， θ_i はガンマ分布(7)式に従うと仮定する．

$$\theta_i \sim \text{Gamma}(a, b) \quad (7)$$

a と b はハイパーパラメータである．(6)，(7)式からなるモデルはポアソンガンマモデルと呼ばれる． θ_i の経験ベイズ推定量は(8)式となる．

$$\hat{\theta}_i = \frac{d_i + \hat{a}}{\bar{d}_i + \hat{b}} \quad (8)$$

$\hat{\theta}_i$ は町丁目 i における世帯あたりの期待取引数の推定値であり， $\hat{\theta}_i$ を与える(8)式は期待取引数 d_i が小さい程その分子を共通の水準 \hat{a} に近づけることで，取引数の少なさに起因した上述の不確実性に配慮した推定量とみなすことができる．本研究では， $\hat{\theta}_i$ に世帯数をかけることで得られる町丁

目毎の期待取引数 $\hat{\theta}_i \bar{d}_i$ を、地価データの町丁目毎の需要の指標として用いることとする。

2.4 予測誤差分散（説明力）と期待取引数（需要）を考慮した評価地点配置の最適化

予測誤差分散と期待取引数のような複数の変数を考慮して観測点配置を最適化するための普遍的な方法は必ずしも確立されておらず、最適化のため目的関数は分析の目的などを踏まえて決める必要がある (Melles et al., 2010)。本研究では、町丁目 i の重心点における予測誤差分散と同町丁目における期待取引数の積を全町丁目について足し合わせた(9)式を最小化することで、地価データの最適配置 S^* を求めることとする。

$$\sum_i \theta_i \bar{d}_i E[(z_i - \hat{z}_i)^2] \quad (9)$$

(9)式を目的関数としたのは同式が、予測誤差分散（ある配置 S で説明不能な撓乱）が土地取引に及ぼした影響の総和と解釈でき、本研究における目的関数として適切と判断したためである。なお、本研究では、予測誤差分散と期待取引数を予め算出し、それらを代入することで(9)式を求める。

(9)式の最小化には Simulated annealing 法（以後 SA 法）を用いる。SA 法は大域最適解を探索するアルゴリズムであり、観測点配置の最適化に用いられる代表的なアルゴリズムの一つである。SA 法を用いた最適化の手順は以下の通りである。

- (i) 初期配置とパラメータ T の初期値 T_0 を与える。
- (ii) 現在の配置 S_a と、そのうちの 1 地点を無作為に別の地点に移動させた配置 S_b を定義する。
- (iii) S_b と S_a について(9)式を算出し（それぞれ δ_a , δ_b とする）、 δ_b がより小さい値を示した場合は確率 1 で、そうでなければ(10)式に示す確率で S_a を S_b に置き換える。

$$\exp\left(\frac{1}{T} \delta_b - \delta_a\right) \quad (10)$$

(iv)(ii)と(iii)を交互に繰り返す

(v)(iv)を k 回試行する毎に T を pT で置き換える。

ここで p は T の減衰率を表わし、1 未満でなくてはならず、かつ 1 に近い程良い。

先行研究に倣い、本研究では無作為に初期配置を与え、かつ $k = 100$, $p = 0.9$ とした。また、(iv)を 100 回反復するという操作を、 T_0 を変化させながら繰り返し行ったところ、 T_0 が概ね 5.0×10^{-6} よりも大きい場合に (iv)をさらに反復すればいずれ(9)式は逓減し始めるはずではあるが) 最小化すべき(9)式が逓増してしまうことが明らかとなった。これは T_0 が過大であることを意味する。一方で、 T_0 が小さい程、局所最適解に陥る可能性は高まる。以上より本研究では $T_0 = 5.0 \times 10^{-6}$ とした。

3. 茨城県の地価調査データの評価地点配置

3.1 概要

平成 23 年度における地価公示データと都道府県地価調査データの評価地点を効率的に削減する方法を議論する。但し分析対象は茨城県における住宅地価に限定する。両地価データを合わせた総標本数は 1,084 であり、本研究では、それらから 200, 400, 600 の各地点数を削減する場合の最適配置を計算する。既述のように、最適配置のための目的関数(9)式は、町丁目の重心点毎に予測誤差分散を算出し、また町丁目毎に期待取引数を算出し、それらを元に評価する。

予測誤差分散(5)式の算出の際に用いるモデルの被説明変数は地価の対数値、説明変数は、最寄り駅までの距離、最寄り駅から東京駅までの鉄道ネットワーク距離、及び、町丁目の重心点を包含する 3 次メッシュにおける土地利用（田、その他農用地、森林、荒地、幹線交用地、その他の用地、河川及び湖沼、海浜、海水域、ゴルフ場）の各面積とした。

3.2 結果

最初に、完全データを用いて算出された予測誤差分散の空間分布を図 2 に示す。この図より、予測誤差に基づけば、データの少ない地域により多くの評価地点を配置すべきとわかる。一方で、期

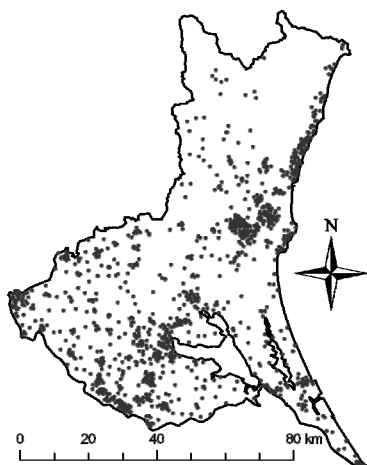


図-1 地価データの空間分布

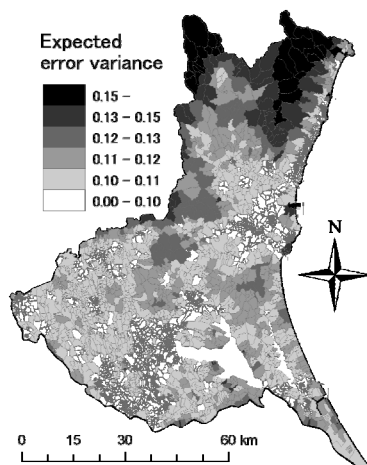


図-2 予測誤差分散の算出結果

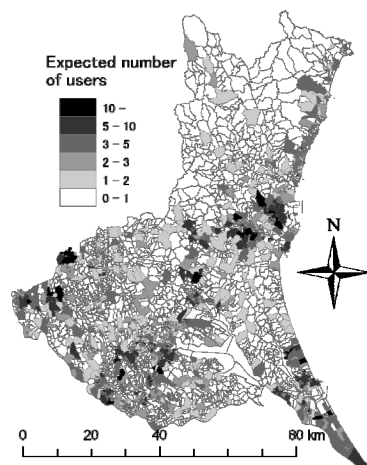


図-3 期待取引数の算出結果

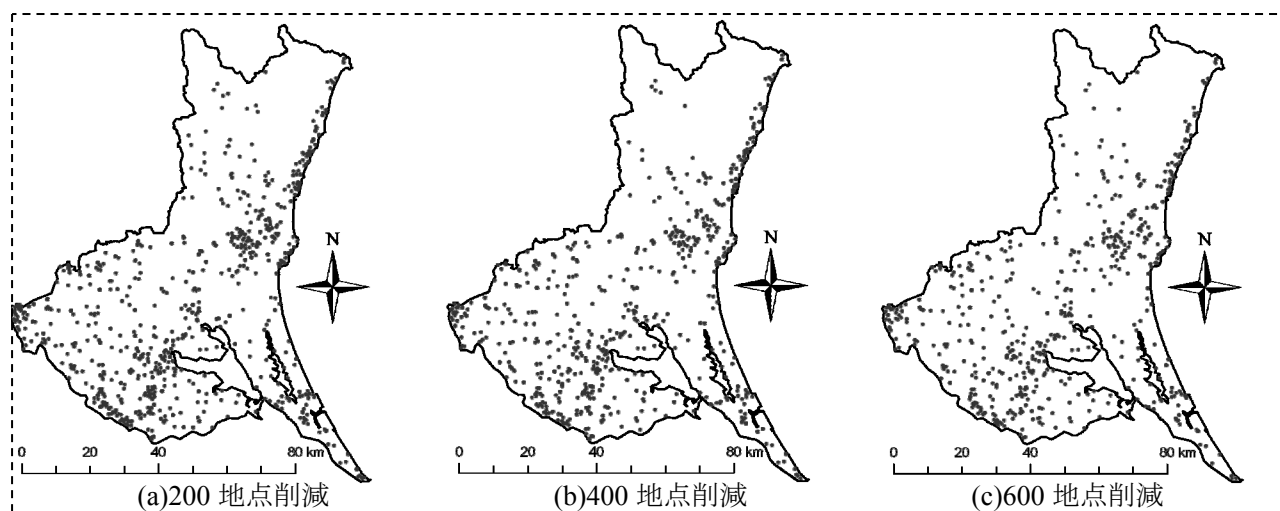


図-4 観測点配置の最適化の結果

期待取引数の算出結果（図3）に基づけば、対象地域中部の水戸市周辺やより東京に近い対象地域南西部といった、都市部により多くの観測点を配置すべきとわかる。

予測誤差分散と期待取引数を考慮した評価地点配置の結果を図4に示す。この図より、都市部の評価地点をやや優先的に減らすべきであることや、非都市部の評価地点は空間的に均一に減らすべきであるということが確認できる。

最適化により得られた目的関数(9)式の値を表1に示す。この表より、削減地点数が200であれば、その目的関数の値と元データ（削減地点数0）の

目的関数の値との間に大差は生じず、地点の削減に伴う予測誤差分散の増大によって被る負の影響は、比較的小さいことが確認できる。

参考文献

- Cressie, N.A. 1993. *Statistics for Spatial Data*. Wiley.
- Diggle, P.J., Menezes, R. and Su, T. 2010. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C*, **59**, 2, 191-232.
- Melles, S.J., Heuvelink, G.B.M., Twenhofel, C.J.W., Van Dijk, A., Himstra, P.H., Baume, O. and Stohlker, U. 2010. Optimizing the spatial pattern of networks for monitoring radioactive releases. *Computers & Geosciences*, **37**, 3, 280-288.

表-1 削減地点数と目的関数

削減地点数	0	200	400	600
目的関数	336.1	337.6	340.5	345.0