

サーフェスモデルによる町丁字別人口の面補間

小西純, 関根智子, 高阪宏行

Areal Interpolation for Small Area Statistics of Census Data Using Surface Model

Jun KONISHI, Tomoko SEKINE and Hiroyuki KOHSAKA

Abstract: Direct comparison of populations using small area statistics of census data for different time periods is usually problematic because of change in zonal units used to report counts. Areal interpolation is used where the data comprise measurements made over areas and the desire is to convert between the existing zonal system and another zonal system. A variety of routines for generation of population surfaces have been developed. In this paper the approach of Martin (1989, 1996) is applied and assessed. The present paper makes use of 40m gridded counts provided as small area statistics data of the 2005 Population Census in Tokyo's 23 wards to assess the performance of methods for generation of population surfaces.

Keywords: 小地域統計 (small area statistics), 面補間 (areal interpolation), サーフェスモデル (surface model), 地区制約サーフェスモデル (zonally-constrained surface model)

1. はじめに

国勢調査の小地域の集計結果による人口の時系列比較は, 集計の地域単位が変化するため容易ではない. このようなある区域のデータを他の区域やグリッド単位に変換する方法として面補間がある (Lloyd, 2010a).

村上, 堤 (2010) は, 空間統計学の空間内挿手法である Kriging をベースとし, 空間従属性と体積保存則を考慮した新たな面補間法を提案している.

また, 村上, 堤 (2009) は, 空間計量経済モデルである EM アルゴリズムを用いた面補間を行っている. これらの分析は, いずれも高齢化率などの割合

を示す変数を対象として補間を行っている.

本稿では, Martin (1989, 1996) によるサーフェスモデルと地区制約サーフェスモデルを採用し, 東京都区部について, 平成 17 年国勢調査の小地域統計データを利用して 40m のグリッド上に補間した. 本手法は, 人口のデータを町丁目・字の人口重心から, 地域内のグリッドに再配分するため, 補間後は直接人口増減を計算できるという特徴がある. 平成 17, 12 年国勢調査の町丁目・字等別境界データを用い, 境界変化の影響について分析し, 小地域の将来人口推計などへの利用可能性について考察する.

2. 面補間法

2.1 面補間の局所的方法

面補間に対しては, 二つの方法が知られている. 一つは, 面補間が外部的で, 全域的な空間構造を用

小西純 〒101-0051 東京都千代田区神田神保町 3-6 能楽書林ビル 5F

財団法人統計情報研究開発センター

Phone: 03-3234-7471

E-mail: junkonishi@sinfonica.co.jp

いる方法である。この方法によると、短い範囲の局所的な変動はランダムで構造化されていないノイズとして除去される。直観的にみて、数値は近くの数値と類似するので、この方法は賢明なものではない。したがって、面補間に対して近隣のデータ地点からの情報を利用する局所的な方法が考えられている (Burrough and McDonnell, 1998)

局所的方法により、空間補間を行う場合、

- ①予測地点の周りに探索エリア（窓）を設定する
- ②探索エリア内のデータ地点を検出する
- ③この限られた数のデータ地点上での変動を表現する関数を設定する
- ④規則グリッド上の地点に対しその数値を評価する

ことが関係してくる。これに関連して、①では、探索エリアの規模、形状、指向を検討する必要がある。

②では、データ地点数とデータ地点がどのように分布しているか（規則的か不規則か）が考察される。

③では、利用する補間関数の種類が検討される。④では、傾向に関する外部情報や異なった領域からの外部情報を組み込む可能性が検討される。

2.2 サーフェスモデルと地区制約サーフェスモデル

1) サーフェスモデル

面とは、複数の頂点を結ぶ直線の辺に囲まれた図形である。このモデルでは、グリッド上のセル内の人口を推定する。人口分布は、セントロイドの位置で最高で、距離が離れるにつれて減衰する。人口密度は陰影で表示され、濃い陰影ほど高密度を示す。この方法は、中心位置によって与えられる分布情報を一定範囲内の人口密度や無人口地域の再構築のため利用する点で優れている。人口数保存から見ると (Lloyd, Martin and Shuttleworth, 2007)、このモデルでは、対象地域全体の人口の合計と人口面上の人口の合計が一致するので、人口数は全域的に

保存されている。しかし、このモデルは、地区境界については何ら考慮していない（この点で無制約と呼ばれる）。セルは、一定範囲内の 1 つ以上の地区中心から人口数の再配分を受ける。したがって、推定された人口数は、個別の地区ごとに見るならば元の地区の人口と一致するものではなく、この意味では人口数は局所的には保存されていない。この方法は、適応カーネル推定の形式のアプローチであるとみなされている (Bailey and Gatrell, 1995)。

4) 地区制約サーフェスモデル

このモデルは、3) のモデルを改良したもので、総人口は地区内でも保存されている。地区の外側に位置するセルは、自動的に加重がゼロと設定されるので、地区の人口数と地区に当たる人口面上の人口の合計が一致し、人口数は全域的と局所的の両側面で保存される (Martin, 1966)。

3. サーフェスモデルの適用

3.1 サーフェスモデルの考察

国勢調査の小地域統計データを用いて、地区（町丁目・字界）内の人口をグリッド上のセルに再配分するサーフェスモデルについて考察する。

1) 地区のセントロイド

サーフェスモデルを応用するときにはまず直面するのは、人口を集計している地区のセントロイドをどこに位置づけるかという問題である。日本の国勢調査では、最小データ出力区は、町丁目・字であり、1995 年国勢調査から全国のデジタル境界データが作成され、小地域統計と呼ばれる多種類の統計データとともに提供されてきた。しかし、町丁目・字界の人口で加重されたセントロイドは提供されていない。本稿では、国勢調査の基本単位区の図形中心点座標データと集計データを利用し、町丁目・字界内の人口重心点を算出し、地区のセントロイドとした。人口加重されたセントロイドの座標は次式によ

り算出する.

$$x_G = \sum_i \frac{p_i x_i}{p_i}, y_G = \sum_i \frac{p_i y_i}{p_i} \quad (1)$$

ただし、人口加重されたセントロイドの座標： (x_G, y_G) ，基本単位区 i の中心点座標： (x_i, y_i) ，基本単位区 i の人口： p_i である.

2) 距離－減衰関数

地区のセントロイドの人口は、そこからの距離に従って通減する加重値に基づき周辺のセルに再配分される. Cressman (1959) は、この加重値 W を、次式のような距離－減衰関数で示した.

$$W = (N^2 - d^2)/(N^2 + d^2) \quad (2)$$

ただし、 d はセル（の中心）からデータ地点への距離、 N は W がゼロになる距離を示す. なお、 W は正の値しか使用されない.

Martin and Bracken(1991)は、Cressman のこの距離－減衰関数を基礎として、次のような適応カーネル推定を提示した(Lloyd, 2010b).

$$\begin{aligned} W_{ij} &= \{(\tau_j^2 - d_{ij}^2)/(\tau_j^2 + d_{ij}^2)\}^r & d_{ij} < \tau_j \\ W_{ij} &= 0 & d_{ij} \geq \tau_j \end{aligned} \quad (3)$$

ただし、 W_{ij} はセル i に対するデータ地点 j の加重値である。 d_{ij} はセル i からデータ地点 j への距離である。 τ_j はデータ地点 j を中心とした整合窓幅である。 r はべき乗の指数である。 r の値が大きくなるほど、データ地点に近いセルの加重は大きくなり、分布の形を支配する。

ここでいう窓とは、移動窓であり、カーネルとも呼ばれる（高阪，2002）。窓は各データ地点に順次位置づけられる。窓の規模はその周辺のデータ地点の密度に従って変わる。整合 (adjusted)，または、適応 (adaptive) とは、一般に、状況に応じパラメータを変える方式を指す。（長尾ほか，1990）。整合窓幅は、現在位置づけられているデータ地点によっ

て表される地域単位の規模を推定している。周辺にデータ地点が多く存在しているところでは、窓は小さく、データ地点がまばらにしかないところでは大きくとるような工夫がなされている。

3) 整合窓幅の設定

まず、探索半径が前もって決められる。各データ地点は順次取り上げられ、それを中心に探索半径が描かれ、その内部に位置する他のデータ地点が列挙される。そして、それらのデータ地点に対し、平均中心間距離が計算される。平均中心間距離は、データ地点 j から探索半径内の他のデータ地点への距離の平均が与えられ、探索半径内におけるデータ地点の分布のエリア的な広がりを表している。その値は、出力グリッド内のセルに加重値を割り振る距離－減衰関数を調整するために、整合窓幅として利用される。

4) 人口数の保存

人口総数を保存するためには、加重値が合計1になるような制約が必要である。そこで次式を用いて、データ地点 j に対し加重値の合計が1になるようにスケール化される。

$$\sum_{i=1}^k W_{ij} = 1.0 \quad (4)$$

ただし、 k は窓内のセル数である。

すると、セル i は、Cressman の距離－減衰関数から導かれた加重値に従って、次式のように各データ地点から再配分をうける。

$$Z_i = \sum_{j=1}^n p_j W_{ij} \quad (5)$$

ただし、 Z_i は i の推定人口数である。 W_{ij} はデータ地点 j の人口数をセル i に配分するための加重値である。（式（3）と（4）参照）。 n は窓内のデータ地点数である。

5) 地区制約

サーフェスモデルでは、セルは、整合窓幅内にあ

る一つあるいは複数のセントロイドから人口の加重値を受け取る。この場合、人口は大局的には保存されるが、窓が重なる場合に、必要のない地域まで人口が再分配される。このため、Martin(1996)は、応用として地区制約モデルを示した。この手法では、人口の再分配の際に、当該地区の外側にあるセルについては、自動的に重みがゼロになる。

3.1 サーフェスモデルの適用

1) 利用データおよび前処理

国勢調査小地域統計データを利用して、東京都区部において、サーフェスモデル、地区制約サーフェスモデルを適用した。本適用において、人口重心の作成及びサーフェスモデルの適用は ArcView3.3 のマクロ言語 Avenue により行った。

町丁目・字界のポリゴンデータとして、平成 17 年国勢調査の町丁・字等別境界データを利用し、区域内のセントロイド算出には、同調査の基本単位区別集計と図形中心点データを利用した。町丁・字等別境界データは、飛び地は別ポリゴンとして作成されているが、本稿ではこれらを統合し、町丁目・字界を一意にした。これらの町丁目・字界ごとに、人口加重されたセントロイドを算出した。

本稿では、サーフェスモデルと地区制約サーフェスモデルの地域単位変更にかかる精度について分析するため、地域単位が異なるポリゴンとして、平成 12 年国勢調査の町丁・字等別境界データも利用した。補間する統計データは、平成 17 年の基本単位区別集計を平成 12 年のポリゴンに対応づけて作成したセントロイドを利用した。これらの補間結果を比較することにより、境界データの変化による影響を確認することができる。

作成された人口加重されたセントロイドは平成 17 年が 3,080, 平成 12 年が 3,084 であった。町丁目・字界に重なる基本単位区の人口が全て 0 の場合は、セントロイドは作成していない。

2) 探索半径

サーフェスモデルでは、各セントロイドを中心とした探索半径の円内に列挙されるセントロイドについて、平均距離を算出する。探索半径が大きくなるほど、列挙されるセントロイドの数が増加し、平均距離も大きくなる。本稿では、当該町丁目・字界に隣接する町丁字ポリゴンに重なるセントロイドを列挙し、そのうちの最大距離を全セントロイドについて算出した。これらの平均値を算出して探索半径とする。平成 17 年の隣接町丁目・字界ポリゴンにあるセントロイドの最大距離の東京都区部における平均値は 796.9m であったため、探索半径を 800m とした。

3) 整合窓幅

サーフェスモデルでは、当該セントロイドから探索半径内にある、セントロイドを列挙し、セントロイドの距離の算術平均を整合窓幅とする。平成 17 年データによる整合窓幅の分布を図-1)。探索範囲内にセントロイドが存在しない場合は、整合窓幅を 800m とした。整合窓幅の最小値は、371.6m であった。町丁目・字界の面積が小さい地域では、整合窓幅も小さく、沿岸部など、周辺にセントロイドがない地域の整合窓幅が大きいことが分かる。

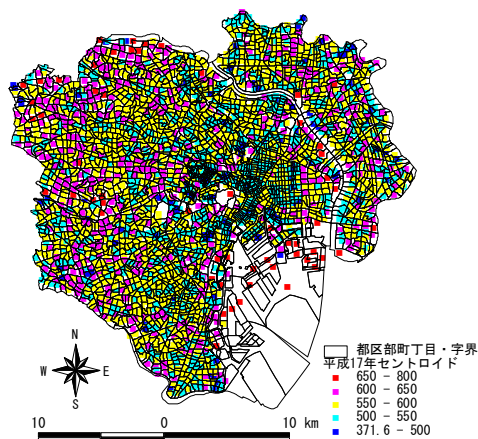


図-1 整合窓幅の分布

4) サーフェスモデル

サーフェスモデルは、この整合窓幅内にある 40m グリッド上の点全てについて、式 3 に従い、かつ式 4 の人口数の保存を満足する重みにより、セントロイド上の人口が配分される。この場合、町丁目・字界の外側のグリッドにも人口が配分されるため、整合窓幅の外縁付近にあるグリッドは、他のセントロイドからの配分も受ける。複数のセントロイドから配分されたグリッド上の推定値は合計されるためサーフェスモデルの結果は、近隣町丁目・字界の影響を受けて平滑化される。

5) 地区制約サーフェスモデル

一方、地区制約モデルは、サーフェスモデルのように町丁目・字界外のグリッドへの配分を行わず、当該町丁目・字内のグリッドについて、式 4 の人口数の保存を満足するように人口が配分される。

4. 精度の検証

4.1 検証の方法

ポリゴン境界の変化について、平成 12 年と平成 17 年のポリゴンにより、平成 17 年人口を補間し、両モデルについて、平成 17 年データによる結果を正として誤差を算出し地図化した。また、誤差の基本統計量及び RMSE (2 乗平均平方根誤差) を算出し検証した。

4.2 ポリゴン境界変化の影響

図-2 はサーフェスモデル、図-3 は地区制約サーフェスモデルについて、誤差を地図化したものである。表-1 は誤差の基本統計量と RMSE を示す。

図-2 は境界変更が行われた地域だけに誤差が表れているが、図-3 の地区制約サーフェスモデルでは、誤差が分散している。これは、ポリゴンデータの微小なズレによるものである。地区制約サーフェスモデルでは、境界の内外でデータの配分有無が決まる。地図データのズレの間にグリッドが含まれる

場合は、配分元が変化するため誤差が大きくなる。一方のサーフェスモデルは、ポリゴンを越えて推定されるため、地区単位で見ると精度は良くないが、境界が変化した際の影響は、平滑化されることによって弱まることから、その誤差は小さくなる。

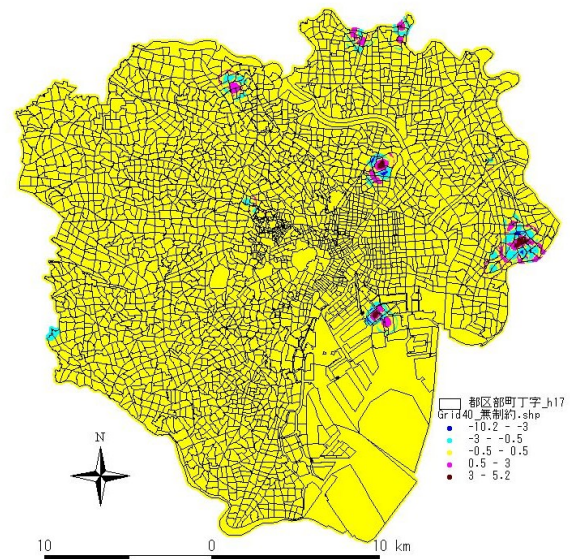


図-2 境界変化による誤差の分布
(サーフェスモデル)

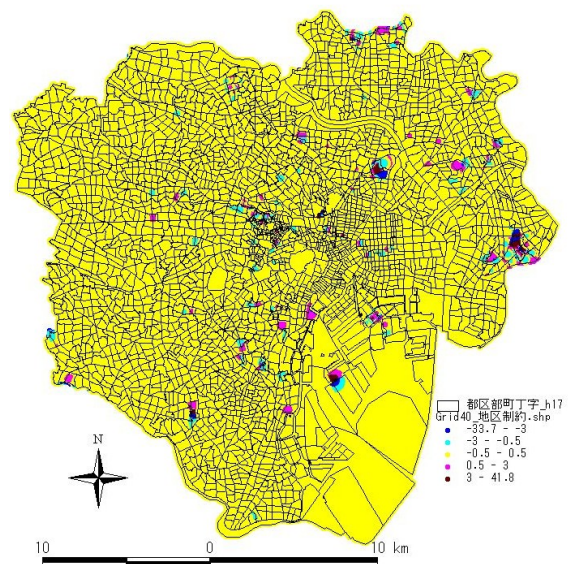


図-3 境界変化による誤差の分布
(地区制約サーフェスモデル)

また、両モデルの誤差の平均値はほぼ等しいが、最大値、最小値とも地区制約サーフェスモデルの方が大きい。これは、特に面積の小さいポリゴン等で境界データのズレが発生した場合に配分される人口の割合が大きいためと考えられる。RMSEを見ると、0.6611と0.2440であり、地区サーフェスモデルの方が大きくなっていることから、サーフェスモデルの方が境界変化の影響が小さい（表-1）。

表-1 誤差の基本統計量と RMSE

面補間モデル	グリッド数	平均	最大値
地区制約サーフェスモデル	417289	-0.00062	33.6711
サーフェスモデル	417289	-0.00062	10.2364
面補間モデル	最小値	分散	RMSE
地区制約サーフェスモデル	-41.8383	0.4370313	0.6611
サーフェスモデル	-5.2172	0.0595351	0.2440

5. おわりに

地区制約サーフェスモデルは、地区内のデータのみに補間するため、基データの精度が保存されるという利点があるが、時系列比較の際は、微小な変化にも影響を受けてしまう。一方、サーフェスモデルは、面の情報精度が悪くなるため、単年度の利用では誤用の可能性が高くなるが、境界の変化には比較的安定的であることが明らかになった。

小地域の将来人口推計には、地区制約サーフェスモデルを適用した人口に対して、サーフェスモデルの補間結果によるコーホート変化率を用いるなどの方法が考えられ、今後の課題としたい。

参考文献

高阪宏行(2002)『地理情報技術ハンドブック』朝倉書店、226。
 長尾ほか(1990)『情報科学辞典』岩波書店。
 村上大輔・堤盛人(2009)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の対処法、第23回応用地域学会発表論文。

(http://www.shiratori.riec.tohoku.ac.jp/~takita/ARSC2009/Paper/ARSC2009_03.pdf)

村上大輔・堤盛人(2010)空間統計モデルに基づく面補間法の提案、第19回地理情報システム学会(<http://ecogis.jpn.ch/GISA10/papers/3C-2.pdf>)

Bailey, T. C. and Gatrell, A. C. (1955) *Interactive Spatial Data Analysis*. Harlow: Longman.

Burrough, P. A. and McDonnell, R. A. (1998) *Principles of Geographical Information Systems*. Oxford: Oxford University Press.

Cressman, G. P. (1959) An operational objective analysis system. *Monthly Weather Review*, 87, 367-374.

Lloyd, C. D. (2010a) *Spatial data analysis*, New York: Oxford University Press, 129.

Lloyd, C. D. (2010b) *Local Models for Spatial Analysis 2nd Edition*. Boca Raton, FL: CRC Press, 177-180.

Lloyd, C. D., Martin, D. J. and Shuttleworth, I. G. (2007) Assessing population surface models using the Northern Ireland census grid square resource. *Geographical Information Science Research UK Conference*. (<http://ncg.nuim.ie/gisruk/materials/proceedings/PDF/3A4.pdf>)

Martin, D. (1989) Mapping population data from zone centroid location. *Transactions of the Institute of British Geographers*, 14, 90-97.

Martin, D. (1996) An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems*, 10(8), 973-989.

Martin, D. and Bracken, I. (1991) Techniques for modeling population-related raster database. *Environment and Planning A*, 23, 1069-1075.