

# 小地域データにおける地理統計解析と SVG による可視化

亀川 佳美<sup>1</sup>・藤野 友和<sup>2</sup>・垂水 共之<sup>3</sup>

(受付 2006 年 7 月 18 日; 改訂 2006 年 10 月 2 日)

## 要 旨

インターネット接続環境が整備され、国勢調査に代表される官庁統計の小地域集計データについても入手が容易となってきたが、新たな知見を得るためには、単に公開されているデータを眺めるだけでは不十分である。それは、デスクトップ GIS などを利用し、個々の環境において分析することで、初めて達成されると考えられる。しかしながら、一般ユーザーにとって、高度な分析機能を有するデスクトップ GIS は高価であり、利用方法を習得するには時間もかかる。このような問題を解決するためには、Web 上にそのような空間データを対象とした分析ツール(WebGIS)をデータと共に公開することが効果的であると考えられる。WebGIS においては、一般的な Web 上の統計解析システムなどに比べ、地図などのグラフィックスをサーバ側で動的に生成し、さらにクライアント側でそのグラフィックスに対話的機能を持たせるというように、グラフィックスに対してより充実した機能が求められる。本論文では、国勢調査の小地域集計データに関する WebGIS を紹介するが、グラフィックスに対するそのような要求を満たすため、開発コストや操作性の面で優れた性質を持っている SVG を画像フォーマットとして採用した。

一方、小地域集計データの統計解析においては、各小地域の代表点を 2 次元の連続確率場からの標本データと見なして、空間予測のための理論(クリギング)を適用することができる。各小地域の特異性の度合いを可視化するため、特異性の指標として、小地域ごとの特性値に対するクリギングによる予測値と観測値の予測誤差をコロプレス図により表現することを考案する。さらにこれを WebGIS に実装した例を示す。

キーワード：空間統計解析，小地域データ，インターネット，GIS，SVG。

## 1. はじめに

近年、Web やメディアの配布を通じた各種統計調査データの入手が容易となってきた。欧米では、個票レベルでの調査結果の公表も進んでおり、アメリカのセンサスデータに関しては、個人の識別が困難になるような処理が施された上で、一部の標本データが Public Use Microdata Sample (PUMS) として、Web で公開されている。日本でも、個票レベルでの公開には至っていないものの、国勢調査の小地域(町丁・字等別)に関する集計データが CD-ROM に

---

<sup>1</sup> 株式会社 数理システム：〒160-0022 東京都新宿区新宿 2-4-3

<sup>2</sup> 福岡女子大学 人間環境学部：〒813-8529 福岡市東区香住ヶ丘 1-1-1

<sup>3</sup> 岡山大学 アドミッションセンター：〒700-8530 岡山市津島中 3-1-1

より入手可能になっている。このようなデータの多くは独自形式のデータファイルとして提供されており、一般に入手可能なソフトウェアで手軽に分析できるというものではない。実際には、これらのファイルを入手した研究者や専門家たちが、独自の処理を行って、個々の環境で分析を実行する必要がある。特に、地域ごとの集計データなどについては、単なる統計処理だけではなく、GIS を利用した分析を行うのが一般的である。ただ、現在のようにインターネット接続環境が充実する以前は、ESRI 社の ArcInfo のような、単独のアプリケーションソフトとして動作する、いわゆるデスクトップ GIS を利用するのが主流であった。しかし、このようなソフトウェアは高機能ではあるが、高価であり、一般のユーザーが利用するには敷居が高い。一般の利用者にとっては、入手したデータが価値の高いものであっても、分析ツールが伴わなければ、その価値は損なわれる。したがって、公開されたデータを有効活用するという観点から考えると、Web 上で動作する GIS、いわゆる WebGIS は、そのような状況に対する一つの解決策となりうる。

Web 上に構築される分析ツールには、特定データを対象としたものと、汎用的な解析ツールを提供するものがある。特に、WebGIS においては、一般のユーザーの利用や、広くデータを公開して、分析してもらうことが前提となっているため、特定データを利用して、そのデータに特化した分析ツールを、一般ユーザーに利用しやすい形にインターフェースなどを工夫して構築されていることが多い。また、近年の Web 技術の進化に伴って、そのような WebGIS を低コストで、短期間に、かつ軽量なアプリケーションとして構築できるようになったというのも、特定データに関する WebGIS が多く公開されている理由の一つであろう。低コストという点に関しては、近年、サーバ OS やアプリケーションにおいて、オープンソースソフトウェアの利用が進んでおり、オープンソースソフトウェアによる Web アプリケーションの典型的な構成例を示す LAMP (Linux, Apache, MySQL, PHP) や、MySQL を PostgreSQL に置き換えた LAPP という言葉を目にする機会が増えてきた。そのような状況の中で、特に WebGIS を構築する上で問題となっていたのは、グラフィックスの表示であった。WebGIS においては、地図をはじめとするグラフィックスをサーバ側で動的に生成し、さらにクライアント側でそのグラフィックスに対話的機能を持たせることが求められる。自由に用いることのできる、Web 上の画像フォーマットとしては、これまで JPEG や PNG をはじめとするラスターグラフィックスであった。これらのフォーマットを利用して WebGIS を構築すると、開発コストや操作性の面で不利な点が多かった。しかし、近年 Web 上で利用可能なベクターグラフィックスの標準規格として、Scalable Vector Graphics (SVG) が公開された。我々は、SVG が WebGIS における画像フォーマットとして優れた性質を持っていることに着目し、データベースと連携した WebGIS において、SVG を利用したシステムに関して研究を進めてきた。

一方、冒頭で述べた小地域データは、国勢調査をはじめ、各省庁が行っている様々な統計調査に関するものが、今後公開されていくことが期待され、それに伴って、それらに対する分析環境の整備も望まれる。小地域データは、一般の市区町村や県単位のような地域データに比べて標本数が非常に多いため、各小地域の代表点を 2 次元の連続確率場からの標本データと見なすことで、地理統計学におけるクリギングをはじめとする予測理論を適用することができる。WebGIS の利用者にとって、興味の対象の 1 つとなるのは、どの地域が周辺地域と比較して、特異な値を示しているのかということ、つまりホットスポットの検出である。我々は、小地域ごとの指標に対するクリギングによる予測値と観測値の予測誤差を、特異性の指標として採用し、WebGIS に実装することを考案した。本論文では、この手法に関する紹介を行うと共に、この手法を実装した SVG による WebGIS に関する提案を行う。

## 2. 小地域データについて

本研究で取り扱う「小地域」とは、研究対象とする領域全体に対する各領域の相対的な大きさのことを指している。また、非常に多くの小地域の存在を前提としている。特に、本研究では、研究対象とする領域全体を各都道府県とし、各領域を町丁・字等のレベルとしている。町丁単位とは、例えば岡山市いずみ町 1 番 1 号など〇町〇番〇号で表される地域を町、津島中 1 丁目 1 番 1 号など〇(町)〇丁目〇番〇号というように丁目が住所につく地域は丁である。字単位とは、岡山市湊 1 番地など地域名〇番地で表される地域をいう。町丁・字とその他の小地域単位によって町丁・字等单位という。

### 2.1 国勢調査小地域集計データ

本研究で用いたデータは、「平成 12 年度国勢調査 小地域集計データ(岡山県)」のうちの第一次基本集計ファイルである。これらのデータは(財)統計情報研究開発センターから購入したものであり、各都道府県別のデータを入手することができる。表 1 は、第一次基本集計に関する集計項目の一覧である。小地域集計データにおいては、人口が数人程度しかいない小地域が存在し、このような地域に対しても他の地域と同様にデータを開示すると、調査対象者の個人情報漏洩してしまう可能性がある。このようなことを防ぐため、この小地域集計データには一定の秘匿処理が施されている。秘匿処理の方法としては、秘匿対象となった地域の集計データを、隣接地域のデータに足し合わせる形をとっており、小地域集計データを利用して、複数の小地域の合計を求める際に、問題が生じないように配慮されている。この点に関しては、小地域データに特有のものであるため、データの取り扱いの際には注意が必要である。

### 2.2 数値地図

国土地理院が刊行する数値情報は、国土の自然的な状況や開発、利用状況など各種の地理的な情報を電子計算機で処理できるようにコード化・数値化した数値地図情報である。大量の数値地図情報を処理するには、大型計算機並みの処理能力が必要とされ、国や都道府県等の機関に限られてきたが、近年高性能のワークステーションやパソコンが普及していることから、ディスプレイ画面を介して地図情報が扱える「数値地図」としてのデータが提供されるようになった。「数値地図」は、国土地理院で設定した、共通のファイル形式で FD や CD-ROM 等の記録媒体に保管されており、地形図の図葉単位、あるいは市区町村や都道府県等の行政区単位の解

表 1. 国勢調査小地域集計データの集計項目.

集計項目
男女別人口
世帯数
年齢別人口
年齢別人口(日本人)
配偶関係(15歳以上)
配偶関係(65歳以上)
世帯の種類
世帯人員
世帯の家族類型
一般世帯数・世帯人員・延べ面積
住宅の建て方(世帯数)
住宅の建て方(世帯人員)
住宅に住む主世帯数・主世帯人員・延べ面積
住宅の所有関係(世帯数)
住宅の所有関係(世帯人員)
母子世帯数・父子世帯数・高齢者世帯
65歳以上親族のいる一般世帯数・人員・65歳以上親族人員・延べ面積

析が簡単に行えるという特徴をもっている。元となる地形図や、数値化されたデータの種類によって、数種類の数値地図が CD-ROM や FD に記録され、(財)日本地図センターが販売している。一部の数値地図(市区町村単位の行政界データなど)は、国土交通省の Web サイト「国土数値情報ダウンロードサービス」などを通じて入手可能である。

本研究で用いた数値地図は平成 12 年度国勢調査町丁・字等別境界データ(岡山県)であり、数値地図 10000 (総合)準拠形式ファイル仕様である。データ形式は総合ファイルである。総合ファイル形式とは、ディスプレイ画面を介して地図情報が扱える「数値地図」への要望が高まり、国土地理院がこれまでに作成された多種類のファイルを共通の形式にまとめたものである。このデータは、境界を表す線データのみから構成されており、個々の町丁・字等の境界線がどの線で構成されているかを表すエリア情報が含まれていないため、このままでは、コロプレス図などの作成には適さない。したがって、予め線データから必要なポリゴンデータを作成しておく必要がある。

### 3. 小地域データに対する地理統計学の利用

非常に多くの小地域に対してデータが得られている場合、これらを面域データとして処理するよりも、点データ、つまり 2 次元連続確率場からの標本データとして扱い、地理統計学の手法を適用することが合理的であると考えられる。はじめに述べたように、本論文で提案するシステムでは、クリギングにより小地域のホットスポットを検出する機能を実装している。地理統計学における空間予測の理論に関しては、Cressie (1991)、間瀬・武田(2001)などが詳しい。本論文に関連する部分を付録にまとめている。

ここでは、実際の小地域データに対して、上で述べた手法を適用した例を示す。利用するデータは、岡山県の小地域集計データであり、各小地域の老人化率を分析対象とする。老人化率とは、各小地域の人口に対する 65 歳以上の人口の割合である。データの一部を表 2 に示す。ここで、 $x$  座標、 $y$  座標は、各小地域の代表点を表しており、それらの値は、平面直角座標系において、岡山県が属する V 系の原点(北緯 36 度 0 分、東経 134 度 20 分)からの相対位置(単位: m)となっている。これをプロットしたものが図 1 である。

老人化率に関する経験バリオグラム(図 2)から、バリオグラムモデルのパラメータを決定する。この図より、シルの存在が示唆されることから、シルを持つバリオグラムモデルである球型、指数型、ガウス型に対してパラメータ推定を行う。その結果が表 3 であり、推定された各モデルを、経験バリオグラム上に描いたのが、図 3 である。この図から、球型モデルと指数型モデルが経験バリオグラムの構造をよく捉えていること、ガウス型モデルは全体的に当てはまりが悪いことなどが分かる。これらのモデルを用いて、通常型クリギングによる予測を行っ

表 2. 岡山県の小地域集計データ(老人化率)の一部.

番号	コード	$x$ 座標	$y$ 座標	小地域名	人口	65 歳以上人口	65 歳以上男性	老人化率
1	332010050	-146700	-40513	万成東町	707	173	80	0.24
2	332010060	-147048	-40276	万成西町	909	172	72	0.19
3	332010070	-145660	-41111	三門東町	371	87	32	0.23
4	332010080	-145863	-41176	三門中町	320	81	31	0.25
5	332010090	-146077	-41083	三門西町	443	107	37	0.24
6	332010100	-146048	-40868	岩井宮裏	308	36	11	0.12
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5439	33665005003	-111275	-29242	吉ヶ原下の町	126	38	17	0.30
5440	33665005004	-110350	-28300	吉ヶ原坂	75	27	9	0.36

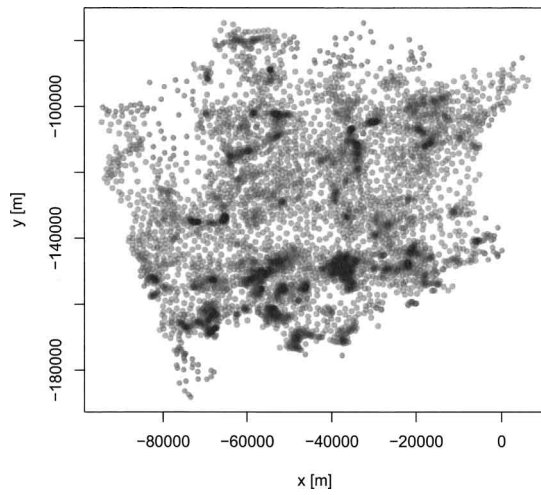


図 1. 各小地域の代表点.

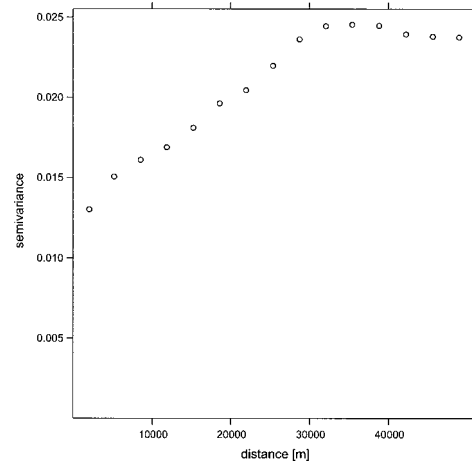


図 2. 老人化率に関する経験バリオグラム.

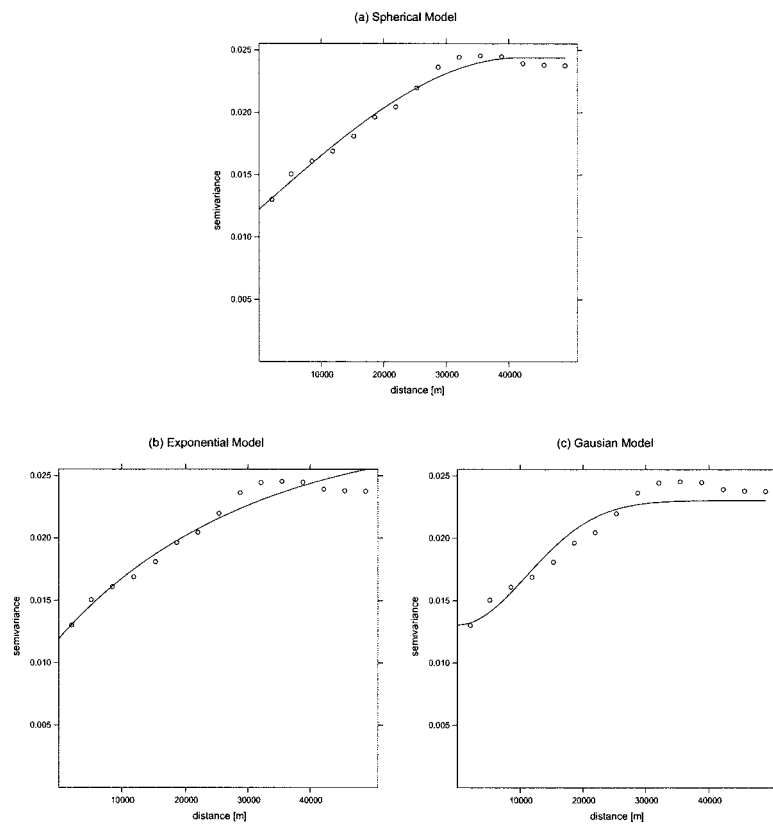


図 3. 標本バリオグラムへのバリオグラムモデルの当てはめ. (a) 球型, (b) 指数型, (c) ガウス型モデルに対応.

表 3. モデルパラメータの推定値.

モデル	球型	指数型	ガウス型
$\theta_0$ (ナゲット)	0.0136	0.0126	0.0139
$\theta_1$	0.0091	0.0105	0.0074
$\theta_2$ (レンジ)	24747	9982	6786
$\theta_0 + \theta_1$ (シル)	0.0227	0.0231	0.0213

表 4. 予測値と実際のデータとの予測誤差.

モデル	球型	指数型	ガウス型
$\sum(\hat{z}(\mathbf{x}_0) - z(\mathbf{x}_0))^2$	81.473	80.64336	84.8284

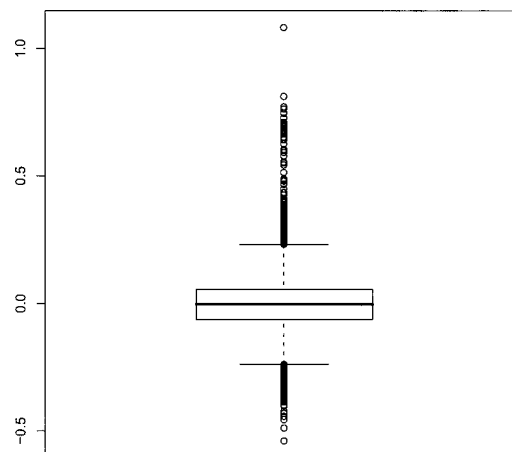


図 4. 予測誤差の分布.

た. 本来, クリギング法は, 空間補間法であるが, 本研究で用いた国勢調査小地域集計データには未観測地点が存在しない. 本研究では, 各観測地点に対して, その点以外のデータ (近傍の 100 カ所) から各観測地点のデータに対するクリギング予測を行い, その予測誤差を各観測地点, つまり小地域の特異性の指標としている. また, モデルごとに, 全観測地点についての予測誤差の二乗和を求め, モデル選択の基準とした (表 4). 二乗和が最小となった指数型モデルを用いて得られた予測誤差の分布を図 4 のボックスプロットに示す. 図 5 は, このボックスプロットにおける上側の外れ値 (左上), 中央値から上側のひげの値まで (右上), 下側のひげから中央値まで (左下), 下側の外れ値 (右下) に対応した観測点の予測誤差に関するバブルプロットである. この左上と右下の図は, 距離以外の要因による特異性を持つ小地域を検討するための材料となるだろう.

#### 4. WebGIS としての実装

##### 4.1 SVG

Scalable Vector Graphics (SVG) は, eXtensible Markup Language (XML) に準拠した 2 次元ベクターグラフィックスを表現するためのデータ形式である. SVG を利用して統計グラフを表現する方法については, Fujino et al. (2004) で提案されている. SVG は, HTML ファイルを Web ブラウザで表示するのと同様に, SVG の表示用ソフト (プラグイン) によって画像の表示を実現する. 最も多く利用されているのは, Web ブラウザのプラグインとして動作する Adobe の SVG Viewer である. SVG の特長は以下のとおりである.

- ・ XML ベースである — 他の XML アプリケーションの利用や連携が可能である.



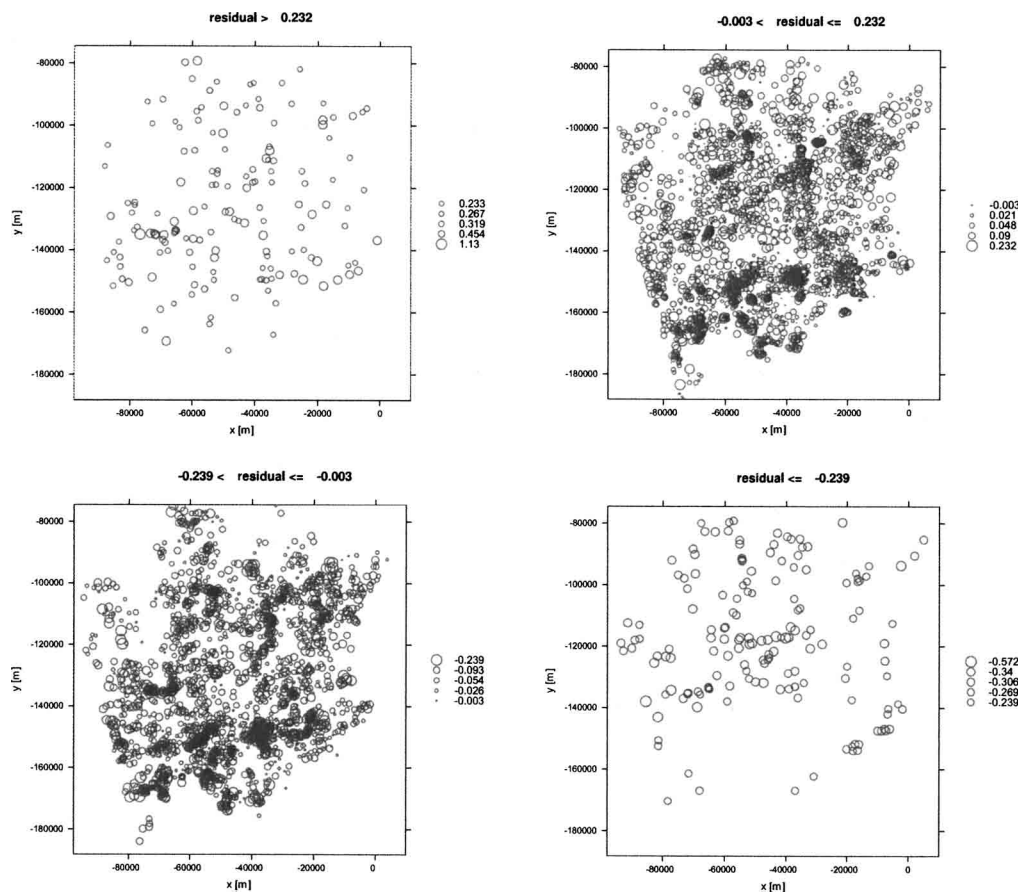


図 5. 予測誤差のバブルプロット.

- ・ベクターグラフィックスである — JPEG や GIF などのラスターグラフィックスと違い、図形情報を保持しているため、画像の大きさに依存せずに高品質な画像を表示することができる。また、図形単位での色や大きさの変更が容易である。
- ・インタラクティブな機能を実装できる — SVG ファイル内に JavaScript を記述して、プラグインが持っている機能(拡大・縮小・移動など)以外に、新たにインタラクティブな機能を実装することができる。
- ・テキストファイルである — データに基づいてグラフィックスを動的に生成する際に、プログラミング言語やそのライブラリに依存せず、単にテキストファイルを出力するだけでこれを実現できる。また、一度生成されたグラフィックスに対する変更や機能の追加も容易である。

#### 4.2 データベースの構成

ここでは、WebGIS のバックエンドとなるデータベースの構成について述べる。データベースには以下の 3 種類のデータをテーブルとして格納する。

- ・小地域集計データ
- ・ポリゴンデータ
- ・index データ

小地域集計データは、1 テーブル 1 都道府県のデータ、1 レコードにつき 1 小地域におけるデータ (小地域名、小地域コード、集計項目  $1, \dots, n$ 、代表点  $x$  座標、代表点  $y$  座標) が含まれているように格納する。集計項目には表 1 の集計項目全てに対する値が含まれる。

ポリゴンデータは、1 テーブル 1 市区町村のデータ、1 レコードにつき 1 点データ ( $x$  座標,  $y$  座標, 小地域コード, 小地域名) が含まれているように格納する。

index データは、1 テーブル 1 都道府県のデータ、1 レコードにつき 1 小地域コードとその小地域における近傍の点 100 個の行番号 (小地域コード,  $v_1, v_2, \dots, v_{100}$ ) のデータが含まれているように格納する。このデータを作成するために、予め格納した小地域集計データに対して、全小地域における代表点間の距離を求めておく必要がある。

#### 4.3 システム構成

今回構築した WebGIS は、以下のような構成となっている。

- ・クライアント側：OS (Windows 2000/XP), Web ブラウザ (Internet Explorer 6.0), SVG ビューア (Adobe SVG Viewer 3.02)
- ・サーバ側：OS (Linux 2.4.22 (VineLinux 2.6r3)), Web サーバ (Apache 1.3.27), RDBMS (PostgreSQL 7.2.3), CGI プログラム (Ruby 1.6.7), 統計エンジン (R 1.9.0)

システムのトップページは <http://face.f7.ems.okayama-u.ac.jp/~ykameka/cgi-bin/system/> である。トップページでは、分析対象とする市区町村を 1 つ以上選択する。選択された市区町村のコードが Web サーバに送信され、ruby スクリプトによって記述された CGI プログラムが、RDBMS に接続し、選択された市区町村の小地域地図を描画するために必要なポリゴンデータを取得する。このポリゴンデータから地図として表示する SVG ファイルを生成し、HTML で記述された小地域分析のためのインターフェイスと共に、ユーザーの Web ブラウザに表示する。SVG による地図は、小地域選択のためのインターフェイスを兼ねており、ある小地域をクリックすれば、小地域の詳細を表示するための選択用フォームにその小地域名が自動入力される仕組みとなっている。これは、SVG の JavaScript により実現しているものである。

小地域選択後の分析画面では、以下の 3 つの処理を行うことができる。

- (1) 小地域の集計情報の表示
- (2) 集計項目の値によるコロプレス図の表示
- (3) 任意の集計項目による、クリギングの実行 (特異値の検出) および、バリオグラムの表示

(1) では、小地域コードがフォームから Web サーバに送られ、CGI プログラムがデータベースの小地域集計データを格納するテーブルに問い合わせ、必要なデータを取り出し、HTML 形式に整形して Web ブラウザに表示させる。(2) では、集計項目がフォームから Web サーバに送られ、CGI プログラムがデータベースのポリゴンデータと小地域集計データから必要なデータを取り出し、SVG によるコロプレス図を生成し、Web ブラウザに表示させる。(3) では、任意の集計項目に対する、クリギングの予測値や予測誤差についてのコロプレス図を表示する。事前に全ての場合 (集計項目や市区町村の組み合わせ) に対する分析結果を予めデータベースに登録しておくのは、その組み合わせの数を考えると現実的でないため、ユーザーの要求に応じて、統計エンジンの R によりクリギング予測を実行している。具体的には、フォームから入力された集計項目とバリオグラムモデルを CGI プログラムが受け取り、必要なデータ (小地域集



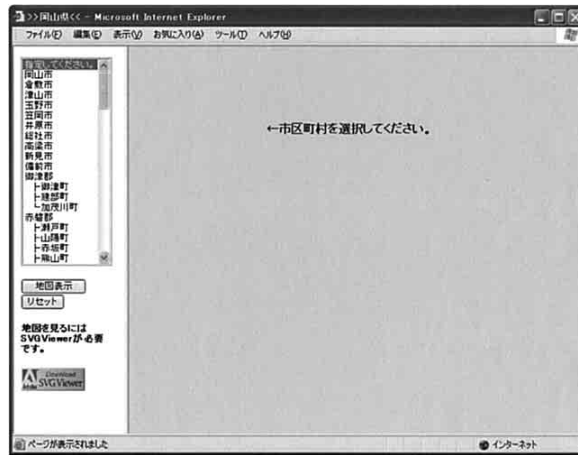


図 6. システムのトップページ.

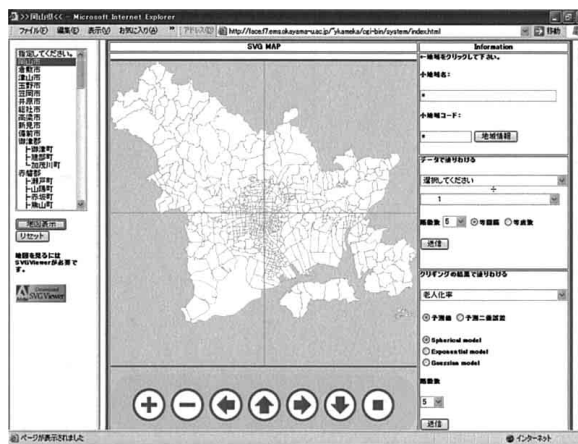


図 7. 岡山市選択時の小地域地図の表示.

計データ, ポリゴンデータ, index データ)をデータベースから取得する. これらのデータを R で分析できるようにするため, 一旦別ファイルに出力しておき, さらに R のスクリプトファイルを出力する. そして, 子プロセスとして, R を起動しクリギング予測の結果, および経験バリオグラムのプロット (SVG 形式) を取得し, HTML 形式でユーザーの Web ブラウザに表示する.

#### 4.4 分析例

##### (1) WebGIS への接続と市区町村の地図表示

システムのトップページに接続すると, 図 6 のページが現れる. そこでは表示したい市区町村が選択できるようになっている. 表示したい市区町村を選択し, [地図表示] ボタンをクリックすると, 指定した市区町村の町丁・字等別地図が表示される (図 7). また市区町村の複数指



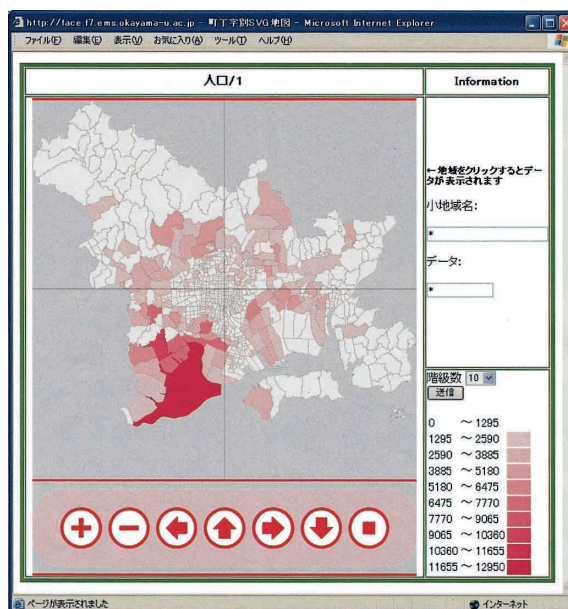


図 10. 人口のコロプレス図.

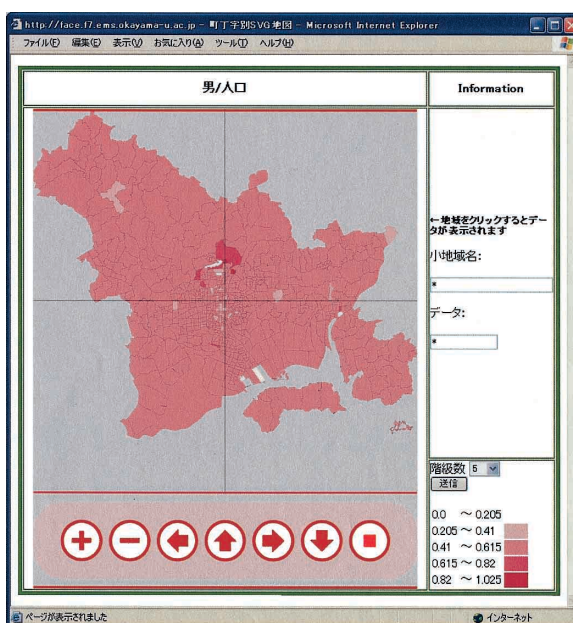


図 11. 岡山市の男性人口率.

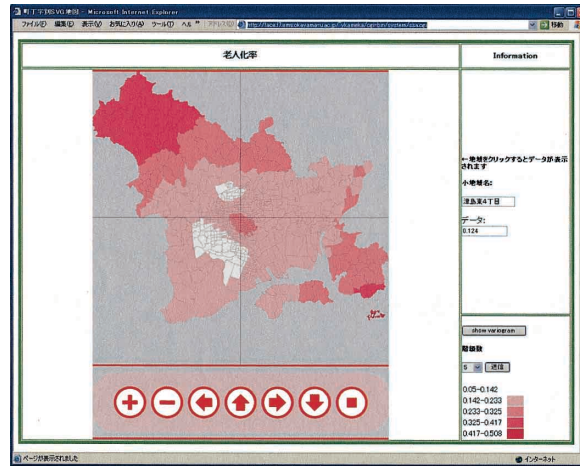


図 12. 岡山市の老人化率の予測値 (指数型モデル, 階級数 5) .

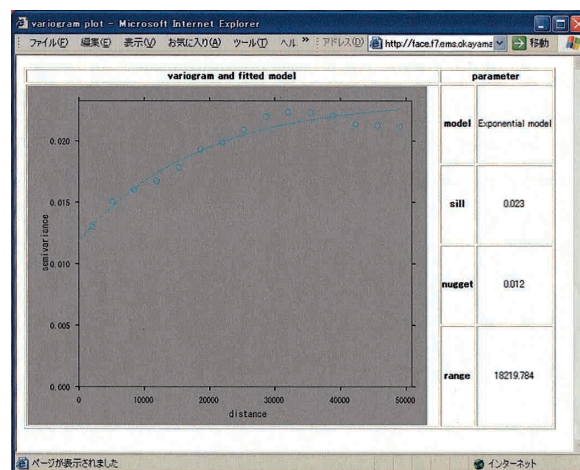


図 13. バリオグラム, モデルパラメータ.

わけたいデータの項目, 階級数, 等間隔・等個数のどちらによって塗りわけるかを選択した後, [送信] ボタンを押すと, 図 10 のページが現れる. ここでは選択した項目においてコロプレス図 (階級区分図) を表示している. 小地域の地図の表示後, データの項目を選択する. 視覚的にデータを眺めることで, 表示項目における空間的従属性の有無や他の小地域との比較が視覚的に可能となる. 階級数は 5 から 10 までを設定している. また, ある項目を他の項目で割った値を計算し, 塗りわけることも可能である. データ項目選択の上の分子部分を男性人口, 下の分母部分を人口にすると, 男性率が表示できる (図 11). 図 11 からは, 男性人口率が高い地区として理大町, いずみ町などが見られる. 理大町は, 岡山理科大学付近であり男子学生寮があることや, いずみ町は警察機動隊の合同宿舎が敷地内にあるため, 男性率が高いと思われる.



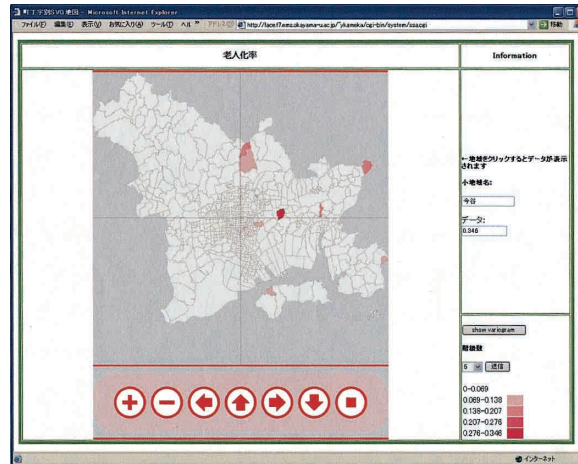


図 14. 岡山市の老人化率の予測誤差 (指数型モデル, 階級数 5)。

#### (4) 地理統計解析の実行と結果の表示, 塗りわけ

図 7 で町丁・字等別地図を表示後, クリギングの結果で塗りわける項目と表示する値(予測値又は予測誤差), バリオグラムモデル, 階級を選択し, [送信] ボタンをおすと図 12 のページが現れる。図 12 では例として岡山市の老人化率を指数型で予測した値を地図上に塗りわけている。ここで [show variogram] ボタンを押すと, 図 13 のページが現れる。図 12 の解析時に用いられたバリオグラムとそのモデルパラメータを表示する。

図 14 では岡山市の老人化率の予測誤差を表示している。例えば誤差が最も大きかった今谷には老人ホームが存在し, 近傍からでは説明できない要因があったといえる。誤差を眺めることで, 距離以外の要因でのホットスポットの検出を行うことができたといえる。

### 5. まとめと今後の課題

本論文では, 国勢調査の小地域データに焦点をあて, この種のデータに地理統計解析を適用し, この手法を WebGIS に実装するという試みについて紹介した。一般に普及している市区町村などの単位での集計データを可視化したり, WebGIS として公開する場合, 単に地域ごとのデータをそのままコロプレス図として表示させたり, 特異性の検出のために確率地図を表示させることなどが一般的であった。しかし, 本論文では, 観測点の非常に多い小地域データは, 2 次元確率場からの標本と見なし, 地理統計解析における予測手法を特異性の検出に利用することを考えた。本論文で用いたデータについては, ある小地域の予測を行う場合に, WebGIS に実装する場合の計算機の負荷のことを考慮して, その近傍 1000 カ所におけるデータを用いて予測を行ったが, この数値の妥当性に関しては今後さらに検討する必要がある。また, WebGIS に関しては, SVG を導入したことで, 数値地図や解析結果の可視化システムが容易に構築でき, 対話的機能の実装も実現した。ただし, システム全体として, 統計エンジンの R をサーバ側で CGI 経由で動作させていることなどから, 軽量の操作性を実現するには至っていない。この改善策については, Horner (2005) が, Web サーバの Apache に対するモジュールとして R を組み込むための拡張機能である mod.R を提案している。これを採用することにより, サーバ上で R が Apache の子プロセスとして起動する際のオーバーヘッドが軽減され, 解析を実施

する際の応答時間が短縮される。さらに、藤野(2005)では、WebGISにおけるサーバとクライアント間の通信方法として、Asynchronous JavaScript + XML (Ajax)を利用したシステムについて報告している。これは、JavaScriptにおけるHTTP通信コンポーネントを利用して、非同期に通信を行い、Webブラウザにすでに読み込まれたXMLコンテンツを動的に変更するための仕組みである。SVGはXMLファイルであるので、プログラムによって自動生成できるだけでなく、Ajaxによって動的更新を行うことが可能である。

#### 付録. 小地域データに対する空間予測の理論

ある領域  $D \subset \mathbf{R}^d$  上の確率場を  $Z = \{Z(\mathbf{x}); \mathbf{x} \in D\}$  とする。各小地域における属性値を、その代表点  $\mathbf{x}_1, \dots, \mathbf{x}_n \in D$  での観測値  $z(\mathbf{x}_1), \dots, z(\mathbf{x}_n)$  とし、これらは、その点における確率変数  $Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)$  の実現値と考える。ここで、この確率場  $Z$  に対し、2次の定常性

$$(A.1) \quad \begin{cases} E[Z(\mathbf{x})] = m \\ \text{Cov}\{Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})\} = C(\mathbf{h}) \end{cases}$$

を仮定すると、この確率場は、以下で定義される本質的定常性を満たすことが分かる。

$$(A.2) \quad \begin{cases} E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0 \\ \text{Var}\{Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\} = 2\gamma(\mathbf{h}) \end{cases}$$

$2\gamma$  は、バリオグラム ( $\gamma$  はセミバリオグラム) と呼ばれるものであり、典型的な推定量として次式が知られている。

$$(A.3) \quad 2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \{Z(\mathbf{x}_i) - Z(\mathbf{x}_j)\}^2$$

ただし、 $N(\mathbf{h}) = \{(i, j); \mathbf{x}_i - \mathbf{x}_j = \mathbf{h}\}$  であり、 $2\hat{\gamma}$  は、 $2\gamma$  の不偏推定量となることが示される。実際の計算においては、この推定値は不安定となるため、バリオグラムに対して等方性、つまり、

$$(A.4) \quad \gamma(\mathbf{h}) = \gamma(\|\mathbf{h}\|)$$

を仮定し、 $(0, R_K] \subset \mathbf{R}$  上の適当な区間分割  $0 = R_0 < R_1 < \dots < R_K$  に対して、経験バリオグラム

$$(A.5) \quad 2\hat{\gamma}(h_k) = \frac{1}{n_k} \sum_{\|\mathbf{x}_i - \mathbf{x}_j\| \in (R_{k-1}, R_k]} \{Z(\mathbf{x}_i) - Z(\mathbf{x}_j)\}^2, \quad k = 1, \dots, K$$

を求める。ただし、 $h_k = \sum_{\|\mathbf{x}_i - \mathbf{x}_j\| \in (R_{k-1}, R_k]} \|\mathbf{x}_i - \mathbf{x}_j\|$ ,  $n_k = \#\{\|\mathbf{x}_i - \mathbf{x}_j\| \in (R_{k-1}, R_k]\}$  である。これによって、以下に示す典型的なバリオグラムモデルのパラメータを最小二乗法などで推定する。

・ 球型モデル ( $\theta_0, \theta_1, \theta_2 \geq 0$ )

$$(A.6) \quad \gamma(\mathbf{h}; \theta) = \begin{cases} \theta_0 + \theta_1 \left\{ \frac{3}{2} \|\mathbf{h}\| / \theta_2 - \frac{1}{2} (\|\mathbf{h}\| / \theta_2)^3 \right\}, & 0 < \|\mathbf{h}\| \leq \theta_2 \\ \theta_0 + \theta_1, & \|\mathbf{h}\| > \theta_2 \\ 0, & \|\mathbf{h}\| = 0 \end{cases}$$

・ 指数型モデル ( $\theta_0, \theta_1, \theta_2 \geq 0$ )

$$(A.7) \quad \gamma(\mathbf{h}; \theta) = \begin{cases} \theta_0 + \theta_1 \{1 - \exp(-\|\mathbf{h}\| / \theta_2)\}, & \|\mathbf{h}\| > 0 \\ 0, & \|\mathbf{h}\| = 0 \end{cases}$$



・ ガウス型モデル ( $\theta_0, \theta_1, \theta_2 \geq 0$ )

$$(A.8) \quad \gamma(\mathbf{h}; \theta) = \begin{cases} \theta_0 + \theta_1[1 - \exp\{-(\|\mathbf{h}\|/\theta_2)^2\}], & \|\mathbf{h}\| > 0 \\ 0, & \|\mathbf{h}\| = 0 \end{cases}$$

通常型クリギングとは、確率場  $Z$  において、未観測地点  $x_0$  における値を、近傍の  $n$  個の標本点  $x_\alpha$  における値と、重み係数  $\omega$  を用いた線形結合によって推定する方法である。未観測地点の値  $\hat{Z}(x_0)$  を、

$$(A.9) \quad \hat{Z}(x_0) = \sum_{\alpha=1}^n \omega_\alpha Z(x_\alpha) = \boldsymbol{\omega}' \mathbf{Z}$$

(ただし、 $\mathbf{Z} = (Z(x_1), \dots, Z(x_n))$ ,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)'$ ) とするとき、 $\sum_{\alpha=1}^n \omega_\alpha = 1$  であれば、 $\hat{Z}(x_0)$  が  $Z(x_0)$  の不偏推定値となる。このとき、推定誤差分散  $\sigma_E^2$  は、2 次定常性のもとで、

$$(A.10) \quad \sigma_E^2 = \text{Var}(\hat{Z}(x_0) - Z(x_0)) = -\boldsymbol{\omega}' \boldsymbol{\Gamma} \boldsymbol{\omega} + 2\boldsymbol{\omega}' \boldsymbol{\gamma}_0$$

となる。ただし、 $\boldsymbol{\gamma}_0 = (\gamma(x_1 - x_0), \dots, \gamma(x_n - x_0))'$ ,  $\boldsymbol{\Gamma} = \{\gamma(x_i - x_j)\}_{ij}$  である。これを条件  $\sum_{\alpha=1}^n \omega_\alpha = 1$  のもとで、ラグランジュ法により最小化すると、最小化する重み

$$(A.11) \quad \boldsymbol{\omega} = \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0 + \left( \frac{1 - \mathbf{1}' \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0}{\mathbf{1}' \boldsymbol{\Gamma}^{-1} \mathbf{1}} \right) \boldsymbol{\Gamma}^{-1} \mathbf{1}$$

が得られ、このときの推定誤差分散が、

$$(A.12) \quad \sigma_E^2 = \boldsymbol{\gamma}_0' \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \frac{(\boldsymbol{\gamma}_0' \boldsymbol{\Gamma}^{-1} \mathbf{1} - 1)^2}{\mathbf{1}' \boldsymbol{\Gamma}^{-1} \mathbf{1}}$$

となる。

## 参 考 文 献

- Cressie, N. (1991). *Statistics for Spatial Data*, Wiley-Interscience, New York.
- 藤野友和(2005). SVG と R を利用した WebGIS の構築, 日本計算機統計学会第 20 回記念大会論文集, 163–164.
- Fujino, T., Yamamoto, Y. and Tarumi, T. (2004). Possibilities and problems of the XML-based graphics in statistics, *COMPSTAT 2004 Proceedings in Computational Statistics*, 1043–1052.
- Horner, J. (2005). Embedding R within Apache web server: What's the use?, *Proceedings of the Fourth International Workshop on Directions in Statistical Computing*, Seattle.
- 間瀬 茂, 武田 純(2001). 『空間データモデリング—空間統計学の応用』, 共立出版, 東京.

## Geostatistical Analysis for Small Area Statistics and Its Visualization Using SVG

Yoshimi Kamegawa<sup>1</sup>, Tomokazu Fujino<sup>2</sup> and Tomoyuki Tarumi<sup>3</sup>

<sup>1</sup>Mathematical Systems, Inc.

<sup>2</sup>Faculty of Human Environment, Fukuoka Women's University

<sup>3</sup>Admission Center, Okayama University

Nowadays, improvement of information infrastructure has made it easy to obtain small area statistics of official statistics including the national population census. We can find some types of insight from this statistics, but they aren't obtained by simply looking at it. They can be only achieved after statistical analysis using a desktop GIS. However a desktop GIS with advanced functions of spatial analysis are too expensive for average users and it takes a lot of time to learn how to use the software. To solve this problem, it would be effective to make public an online analysis system for the spatial data (WebGIS) together with raw data on the Internet. In the WebGIS, many features of graphics, for example, generating graphics such as maps and statistical graphs dynamically on the server side and implementing interactive functions to the graphics on the client side, are required over the ordinary statistical analysis system on the Web. This paper introduces the WebGIS about small area statistics of Japanese national population census, which adopts SVG as a graphics format of the system. SVG has many more features with respect to the development costs and user operability than existing graphics formats.

For statistic analysis of small area statistics, because the representative points in each small area can be considered as sample data from a two-dimensional continuous random field, the predictive theory using kriging can be applied. To visualize the level of singularity of each small area, we propose using prediction error between the observed value and the predicted value by kriging for each small area as the singularity index and drawing its cholopleth map. We have implemented this function in our WebGIS.