

GoogleMaps を用いた地理統計データの可視化

久保田 貴文¹・垂水 共之²

(受付 2006 年 11 月 8 日; 改訂 2007 年 3 月 29 日)

要 旨

気象データや環境データなどの地理統計データをあつかう場合、その特性値とともに位置情報が重要となってくる。予測を行う場合には空間構造をあらわす指標の 1 つであるバリオグラムのモデルを求める必要がある。また、可視化の場面では、地図上に観測点をプロットし、バリオグラムと連携してインタラクティブな操作をし、予測値を地図上に描画する要求がうまれてくる。そこで本研究では、GoogleMaps を用いて、観測値のプロットや空間構造さらには予測値などの表示を行う。

キーワード：GoogleMaps, 地理統計解析, AJAX, SVG.

1. はじめに

気象データや環境データなどの地理統計データをあつかう場合、その特性値とともに位置情報が大きな意味をもち、それらを可視化することが重要となってくる。緯度・経度の位置情報を可視化する方法は様々あるが、地図上に観測値を表示することで、例えば道路や工場などといったその地点に及ぼす影響や、衛星写真(航空写真)上に表示することで、緑地度や河川・湖沼などの地理的要因を類推することもできる。そこで、まずはじめに、観測値を地図上もしくは衛星写真(航空写真)上に表示することを考える。その手段として Google 社が公開している GoogleMaps を使用するための API を用いる。

また、地理統計解析を行う上では位置情報と観測値(特性値)の関係をモデル化することが重要である。その指標の 1 つにバリオグラムがある。R のライブラリーである「gstat」や「geoR」などを用いれば探索的に可視化を行うことは可能である。しかし、バリオグラムと観測値との対応や推定するモデルの選択などをリアルタイムに、そしてインタラクティブに行うことは容易ではない。そこで、第 2 に、バリオグラムをモデル化するためのシステムを作成し、さらに、バリオグラムと GoogleMaps との連携も行えるようにする。バリオグラムを可視化するためには、XML 形式に準拠した 2 次元ベクターグラフィックスを表現するための形式である SVG 形式を用いる。

さらに、求めたバリオグラムのモデルを用いることによって、通常クリギングにより空間予測を行うことができ、これにより、未観測地点における特性値を予測できる。そこで、第 3 に、求められたバリオグラムのモデルを用いて GoogleMaps 上に予測値を表示するシステムを作成する。

なお、地図上に観測値や予測値を表示するシステムは様々なものが提案されており、例えば

¹ 岡山大学大学院 社会文化科学研究科：〒700-8530 岡山県岡山市津島中 3-1-1

² 岡山大学 アドミッションセンター：〒700-8530 岡山県岡山市津島中 3-1-1

James Reserve の Data Management Systems (<http://dms.jamesreserve.edu/>) では, GoogleEarth 上に森林のデータや統計グラフを重ね合わせている. また, クリギングおよびその可視化については, ArcGIS のエクステンションである Geostatistical Analyst でも実現可能である.

これらのシステムと比べて, 本研究で開発したシステムでは, Web 上で, バリオグラムモデルのパラメータ推定, およびモデル選択がインタラクティブに実行でき, さらに, 求めたモデルを用いて予測し, その結果を地図上に表示するまでの一連の地理統計データ解析が行える点で優れている.

2章では, 地理統計解析について述べ, 3章ではシステムについて述べる. それを受けて, 4章では GoogleMaps 上に地理統計解析の結果を載せる方法について述べ, 最後に5章でまとめと今後の展望を述べる.

2. 地理統計解析について

本研究で採用している地理統計解析については, Cressie (1993), 間瀬・武田(2001)に詳しい. ここでは概要を示す. 地理統計学では, データは確率場からの実現値とみなす. 領域 $D \subset R^d$ 上の確率場 $Z = \{Z(\mathbf{s}); \mathbf{s} \in D\}$ を考え, 観測地点 $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$ における確率変数 $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ の実現値である $z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)$ が観測データである. 通常クリギングを行う場合, $Z(\mathbf{s})$ に対して, 式(2.1) (2.2)の本質的2次定常性を仮定する必要がある.

$$(2.1) \quad E\{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\} = 0$$

$$(2.2) \quad \text{Var}\{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\} = 2\gamma(\mathbf{s}_i - \mathbf{s}_j)$$

γ をバリオグラムとよび, γ のモデルを推定し, それを用いて空間補間することで予測を行うのがクリギングである.

バリオグラムのモデルを観測値から推定する手順は,

- (1) すべての観測値のペアからバリオグラムクラウドを求める
- (2) バリオグラムクラウドから経験バリオグラムを求める
- (3) 経験バリオグラムに理論バリオグラムを最小2乗法で当てはめ, パラメータ推定する

である.

2.1 バリオグラムクラウドについて

$Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ が与えられたとき, $Z(\mathbf{s}_i), Z(\mathbf{s}_j)$ 間の非類似度すなわち2つの値の差分を2乗したものの1/2の値は, 式(2.3)となる.

$$(2.3) \quad \gamma_{ij}^* = \frac{(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2}{2}$$

2地点 $\mathbf{s}_i, \mathbf{s}_j$ はベクトル $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ により結びつけられるが, 本研究においては γ は距離のみに依存する, すなわち等方的であるとする. ここですべてのペアにおいて横軸に2点間の距離, 縦軸に式(2.3)をとりプロットしたものをバリオグラムクラウドと呼ぶ.

2.2 経験バリオグラムについて

式(2.3)を \mathbf{h} についてまとめると式(2.4)となる.

$$(2.4) \quad \hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2$$

ここで、 $N(\mathbf{h}) = \{(i, j); \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$, $|N(\mathbf{h})|$ はそのようなペアの総数である。式(2.4)は、固定した \mathbf{h} に対する推定量である。しかし実際には \mathbf{h} を 1 つの値に固定すると、対応するデータのペア $N(\mathbf{h})$ が殆どなくなり、推定が不安定になる。そこで次のように、最大距離を決め、さらにクラス分けを行う。

- (a) 推定に用いる最大距離 h_{max} を決め、それより大きいデータは推定に使用しない。ここで、 h_{max} のことをカットオフ (cutoff) と呼ぶ。
- (b) $0 < R_1 < \dots < R_K (= h_{max})$ を決め、 K 個の区間 $I_1 = (0, R_1], I_2 = (R_1, R_2], \dots, I_K = (R_{K-1}, R_K]$ を考える。
- (c) $d_{i,j} = \|\mathbf{s}_i - \mathbf{s}_j\| \in I_k$ となる (i, j) のペアの集合を N_k , その元の個数を $|N_k|$ とする。
- (d) N_k に属するペア (i, j) に対する $d_{i,j}$ の平均を h_k とする。
- (e) N_k に属するペア (i, j) に対する $\gamma_{ij} = (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2/2$ の平均を γ_k とする。

これらすべての k についての、 h_k に対する γ_k を経験バリオグラムと呼ぶ。

(b) の区切り方については、 $(0, h_{max}]$ を距離で均等に分割する方法が実用的であり、その幅を「width」と呼ぶ。一方、ペア数で均等に分割し、それぞれの範囲 (width) が均等でない方法も考えられている。また、パラメータ h_{max} については注意が必要である。なぜなら、すべてのペアでの最大距離を d_{max} とすると、 $h_{max} = d_{max}$ の場合、距離の大きいクラスではそのクラスに属するペア数が少なくなり、想定している領域において端の方のデータの影響が大きいため、パラメータ推定の際に不安定になってしまうためである。

2.3 理論バリオグラムについて

求められた経験バリオグラムに最小 2 乗法によりパラメータ推定を行う。地理統計解析では、式(2.5)の球型モデルがよく用いられる。

$$(2.5) \quad \gamma(h; \boldsymbol{\theta}) = \begin{cases} \theta_0 + \theta_1 \left(\frac{3}{2} [h/\theta_2] - \frac{1}{2} [h/\theta_2]^3 \right) & (0 < h \leq \theta_2) \\ \theta_0 + \theta_1 & (h > \theta_2) \\ 0 & (h = 0) \end{cases}$$

求められたモデルを理論バリオグラムと呼び、予測値を求める際に必要となる。

2.4 クリギングについて

確率場 Z が本質的 2 次定常であり、バリオグラム γ をもつとする。位置 \mathbf{s}_0 における値 $Z(\mathbf{s}_0)$ の最良線形不偏推定量

$$(2.6) \quad \hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n w_i Z(\mathbf{s}_i) = \mathbf{w}' \mathbf{Z}$$

を通常クリギングと呼ぶ。ここに $\mathbf{w} = [w_1, \dots, w_n]'$ は適当な重みであり、平均 2 乗予測誤差を最小化することにより、

$$(2.7) \quad \mathbf{w} = \Gamma^{-1} \boldsymbol{\gamma}_0 + \left(\frac{1 - \mathbf{1}' \Gamma^{-1} \boldsymbol{\gamma}_0}{\mathbf{1}' \Gamma^{-1} \mathbf{1}} \right) \Gamma^{-1} \mathbf{1}$$

を得る。ここで、 $\boldsymbol{\gamma}_0 = [\gamma(\mathbf{s}_1 - \mathbf{s}_0), \dots, \gamma(\mathbf{s}_n - \mathbf{s}_0)]'$, $\Gamma = [\gamma(\mathbf{s}_i - \mathbf{s}_j)]_{ij}$, $\mathbf{1} = [1, \dots, 1]'$ とする。

本システムにおいては、この通常クリギングを地図上に可視化することを実現している。すなわち、予測するエリアをメッシュに区切り、それぞれのメッシュの中心座標における予測値でそれぞれのメッシュを色分けし、可視化を実現している。

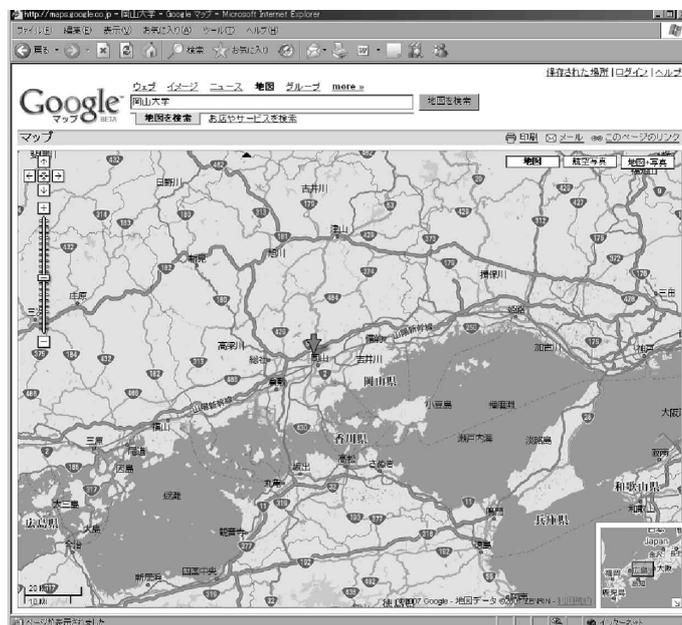


図 1. GoogleMaps による地図の表示の例.

3. システムについて

本システムにおいては、観測地点の描画や予測値のメッシュの描画などの地図を用いる部分には GoogleMaps を、バリオグラムクラウドや経験バリオグラムの統計グラフの部分は SVG を用いている。

3.1 GoogleMaps について

Web 上での地図のサービスについては、決してめずらしいものではなく、Mapion (<http://www.mapion.co.jp/>) などさまざまなページでかなり以前から行われてきた。それに対して、GoogleMaps (<http://maps.google.co.jp/>) (図 1) は、表示されている地図をマウスのドラッグで自由に位置を変更でき、左上にあるスライダにより拡大縮小することもできる。さらに、右上の「地図」、「航空写真」そして「地図+写真」のボタンにより、それぞれ地図、航空写真、そして、航空写真の上に地図を重ね合わせたマップを選択変更できる。

このようなページを実現するためには新たな技術的枠組みが必要となる。ただ単に HTTP の通信と HTML 形式のファイルを組み合わせただけでは、レスポンスのたびにその都度リロードする必要があり、うまくいかない。Perl や Ruby などに代表されるサーバーサイドスクリプトと組み合わせて実現したとしても、地図などを効率よく作成したり、データベースとの接続をする点では優れているが、サーバー上で動作し画像などを含む HTML 形式のファイルを生成しブラウザに送信するという過程は既存の HTML 形式のファイルを送信する場合と同様である。それに対し、GoogleMaps では描画エリアおよびその周辺の地図情報をクライアント側で持ち、クライアント側でのイベントとは別に必要な部分の地図等の情報をサーバー側からダウンロードしている。そのため、描画エリアを変更するさいに、いったんページをストップし画面を切り替えるような必要がない。

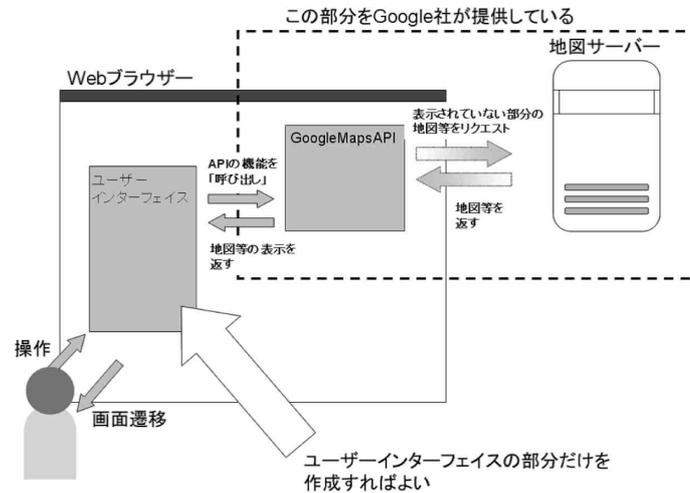


図 2. GoogleMapsAPI の構成.

3.2 AJAX について

この GoogleMaps の技術的枠組みを支えているのが、AJAX である。AJAX とは「Asynchronous JavaScript + XML」の略で、非同期的にクライアント側の HTML とサーバー側の XML 形式のデータをやり取りすることで、ユーザーはクライアントとサーバーの通信を意識せずに画面の遷移やデータのやり取りなどを行うことができる。そのさい、現在主要なブラウザに実装されている JavaScript 言語を用いて互換性を保ち、ユーザーの環境に依存しないシステムを実現している。

3.3 GoogleMapsAPI について

図 2 に示すように、ユーザーが自身の Web サイトに GoogleMaps を設置するための JavaScript の API が Google 社から提供されている。GoogleMapsAPI を利用することで、誰でも Web サイト上に GoogleMaps を利用したコンテンツを作成することが可能となっている。このため、ユーザーはインターフェイスを作成することだけに専念できる。

3.4 SVG について

SVG とは、「Scalable Vector Graphics」の略で、XML ベースの 2 次元ベクターグラフィックスの記述言語である。地理統計システムにおける SVG の利用については久保田・垂水 (2004)、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 の特徴は、XML ベースであるため、すなわちグラフィックスがプレーンテキストで書かれているため、SVG ファイルに対してスクリプトを追記もしくは修正することによりインタラクティブに操作可能な SVG ファイルを作成することができる。また、ベクターグラフィックスであるために、線や面などの図形を、見る環境に応じて最適なサイズに変形し表示することが可能となる。SVG の閲覧方法としては、専用のソフトもしくはプラグイン、例えば Adobe 社の SVGViewer などが必要となる。

3.5 R について

R (<http://www.r-project.org/>) は、統計計算とグラフィックスのための言語・環境である。R

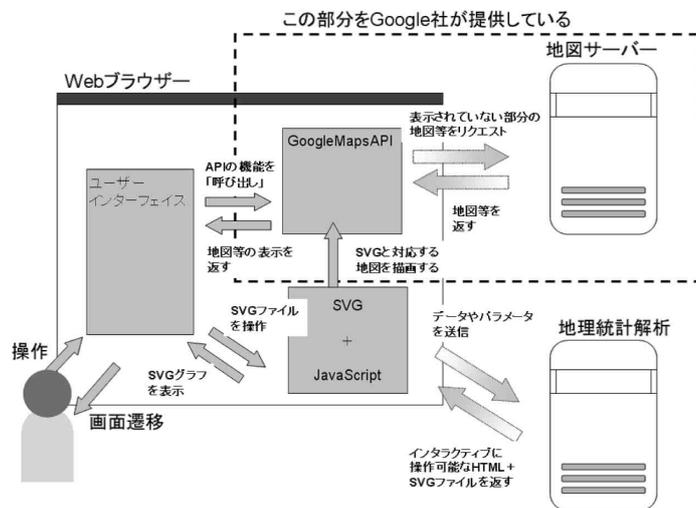


図3. 提案するシステム構成.

においては様々な統計手法が利用可能であり、そのなかでも、地理統計学に特化したものとしては、「gstat」および「geoR」などのライブラリがあげられる。本システムではそれらを用いてバリオグラムのモデルを求める部分およびクリギングを用いて予測を行う部分をシステムの中で実装している。

4. GoogleMaps を用いた地理統計データの可視化

4.1 環境について

2007年3月現在 GoogleMapsAPI (バージョン 2) は標準ではポリゴンを地図上に重ね合わせることが不可能である。そこで、gpolygon.gen.js というライブラリ (<http://www.geocities.com/goldnabr1/gmap/gpolygonsample.html>) を使用している。本研究で実装したシステム (図3) については、「国立環境研究所、環境数値データベース」 (<http://www.nies.go.jp/igreen/>) の大気環境のデータのうち 2004 年の岡山県における、60 カ所の NOX の値を CSV 形式に変形したものをサンプルのデータとして利用する。ユーザーは、クライアント側からデータをサーバ側にデータを送信する。そしてサーバ側でデータを解析し、GoogleMapsAPI を用いて作成した地図、SVG を用いて描いたグラフ、およびそれらに対してインタラクティブに操作可能なスクリプトをクライアント側に返す。

4.2 観測値の可視化について

図4にデータの入力ウィンドウを示す。データの入力ボックス、経度 (lng)、緯度 (lat) および特性値 (value) の列番号を入力するボックス、そしてデータを送信するための「submit」ボタンがある。データは 1 行目に変数名、2 行目以降は行ごとに観測地点のデータを CSV 形式で入力することを要求する。ここでは、データの例として、上記 NOX のデータを用意している。また、列選択フォームでは、データ入力ボックスのデータのうち、どの列が、lng, lat, value に対応するのかを入力する。lng=1, lat=2, value=4 がデフォルトで与えられている。

「submit」ボタンでデータをサーバ側に送信し、サーバ側で地理統計解析の計算・HTML の作成を行い、クライアント側に観測地点のプロット (図5)、バリオグラムクラウド (図6)、経

Geostatistics on Google Maps

1. Input Data

```
x, y, elev, year, max, day
133.9731, 34.60361, 5, 0.029, 0.221, 0.075
133.9347, 34.60611, 5, 0.03, 0.276, 0.078
134.0333, 34.65725, 10, 0.034, 0.25, 0.086
133.9233, 34.67139, 15, 0.041, 0.241, 0.085
133.9836, 34.69444, 15, 0.026, 0.416, 0.067
134.0031, 34.63417, 5, 0.032, 0.26, 0.093
133.9211, 34.66111, 10, 0.041, 0.32, 0.09
133.8522, 34.58917, 5, 0.033, 0.292, 0.1
134.07, 34.61593, 10, 0.015, 0.149, 0.042
```

2. Select Column

lng:

lat:

value:

3. Submit

図 4. データ入力ウィンドウ.

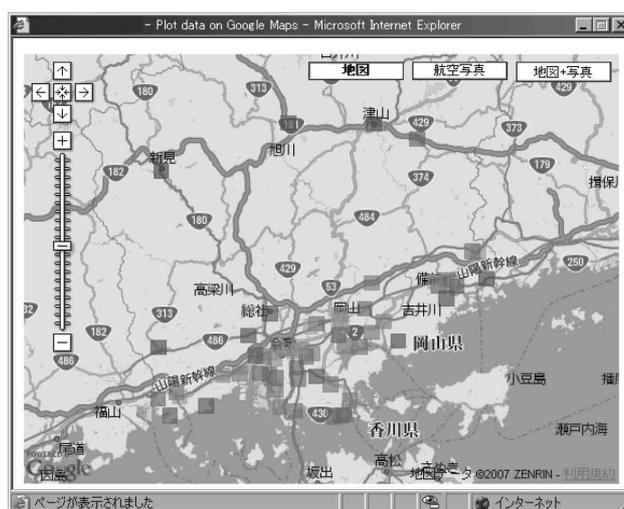


図 5. 観測地点のプロット.

験バリオグラム(図 8)を表示する.

図 5 は, GoogleMaps 上に観測データを可視化したページである. 観測地点については, 図 4 の入力ボックスで入力したデータをもとに, 列選択フォームで指定した lng, lat を用いて, その地点を中心とする正方形をプロットしている. また, ポイントの色は位置と同様に図 4 の列選択フォームで指定した value の列の特性値(NOX)の値によって, 大きいほど赤く, 小さいほど青い色で塗り分けしている.

この観測地点の色分けのプロットにより, 地理的な外れ値(外れ地)ばかりでなく, 特性値の外れ値や, 全体的な傾向, 偏りなどを発見することができる. さらに, それぞれの観測地点の

場所がどこに対応しているのかを地図の拡大によって詳細に見ることが可能である。

4.3 空間構造の可視化について

次に、図6にバリオグラムクラウドを可視化したページを示す。横軸にはユークリッド距離、縦軸には特性値の差の2乗の1/2の値をとっている。入力されたデータから2点のペアを選び、それらの lng, lat からユークリッド距離を計算する。また、それら2点の value の値から、特性値の差の2乗の1/2の値を計算し、距離に対応する点をプロットする。そして、すべてのペアについてプロットしたものがバリオグラムクラウド(図6)である。

バリオグラムクラウドのそれぞれの点は図5の地図の上にプロットされた観測地点のペアに対応している。すなわち、図6のバリオグラムクラウドの点をクリックすることで、その点に対応する図5の2つの観測地点が地図上で結ばれる(図7)。これにより、バリオグラムクラウドのなかでの外れ値を地図と対応させて検討することが出来る。

図8には入力されたデータを用いて計算した経験バリオグラムとその計算のためのパラメータである cutoff, width の値の選択及び入力フォーム、そして、経験バリオグラムに当てはめる理論バリオグラムのモデルに対応するボタンを示す。

経験バリオグラムについてはサーバー側で計算し、SVG形式に変換してクライアント側に表示している。パラメータについては、デフォルトでは $\text{cutoff} = d_{max}$, $\text{width} = h_{max}/15$ が設定されている。cutoff については、Rのライブラリ gstat では最大距離 (d_{max}) の1/3がシステムの中で採用されている。また、久保田・垂水(2005), Kubota et al. (2005) ではシミュレーションにより最大距離 (d_{max}) の1/2が最適な cutoff の値として得られている。本システムにおいては、「max/3」「max/2」のボタンを押すことにより、それぞれ $d_{max}/3$, $d_{max}/2$ に変更することが出来、入力フォームにより特定の数値を指定することも出来る。また、width も同様に cutoff (h_{max}) の1/10, 1/20に変更することも出来、特定の width を指定することも可能である。さらに、モデルを選択するためのボタン「Sph」, 「Exp」, 「Cir」そして「Gau」によりそれぞれ、「球型モデル」, 「指数型モデル」, 「円型モデル」そして「ガウス型モデル」を最小2乗法により当てはめ、そのモデルに対応する式、およびパラメータの値を表示し、SVG形式の経験バリオグラムに赤い線で重ね合わせる(図9)。

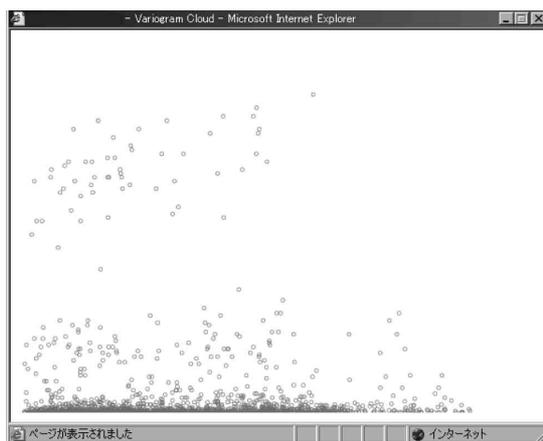


図6. バリオグラムクラウド。

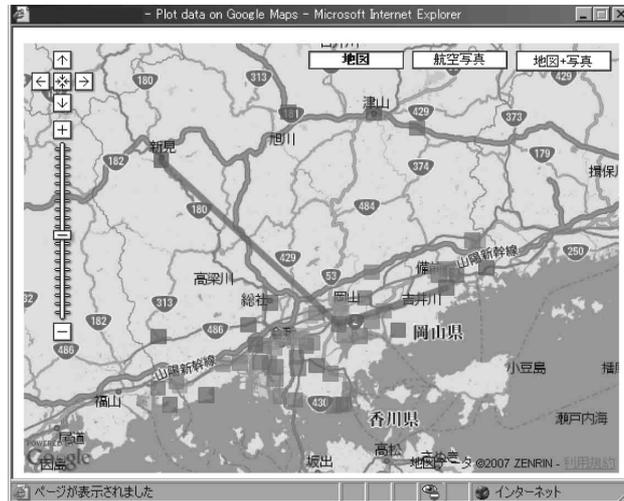


図 7. バリオグラムクラウドと観測地点の関係.

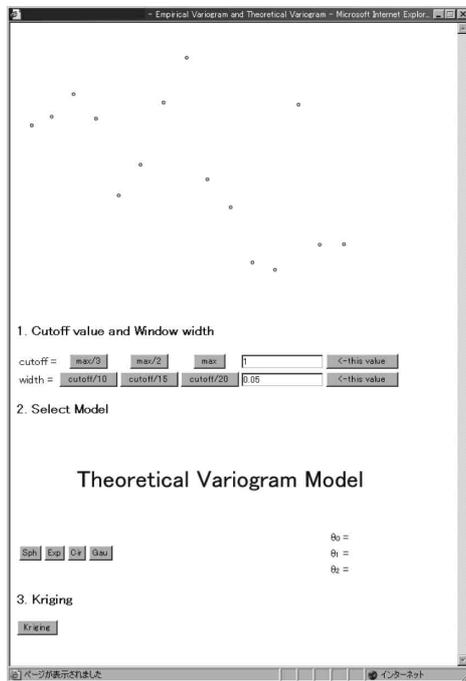


図 8. 経験バリオグラムとパラメータ (cutoff, width).

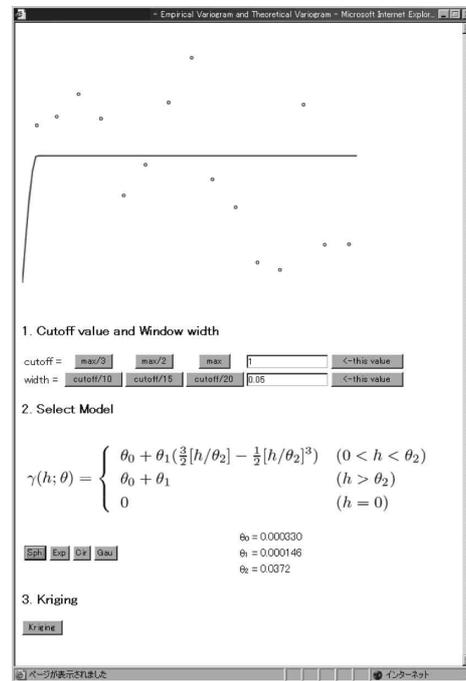


図 9. 理論バリオグラムを当てはめる.

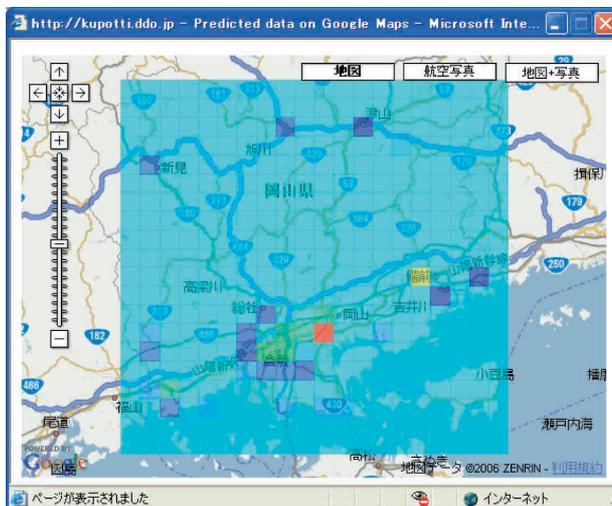


図 10. 予測値のプロット.

4.4 予測値の可視化について

図 9 において「Kriging」のボタンを押せば、求めた理論バリオグラムを用いて通常クリギングにより予測を行う。

図 10 には入力したデータをすべて含むようなメッシュを示している。メッシュは 2 次メッシュを縦横 2 等分した大きさとする。それぞれのメッシュの中心座標において図 9 で求めた経験バリオグラムを用いて通常クリギングで予測を行う。求められた値をもとに、色分けをし、それぞれの中心の値をそのセルの値として GoogleMaps 上に描画している。これにより未観測地点の NOx の値を補完し、地図上に表示出来る。

5. おわりに

本研究においては、地理統計データを、GoogleMapsAPI を用いて地図上に可視化し、SVG を用いて統計グラフを可視化するシステムを提案した。可視化においては、ただ単に観測値を地図上にプロットするだけでなく、空間相関の指標であるバリオグラムの可視化、地図との連携、インタラクティブにパラメータを変更、さらにモデルの選択及びそのモデルにおけるパラメータの推定をし、それを用いて通常クリギングにより予測を行うシステムとなっている。予測値は観測値と同様に GoogleMaps 上に描画され観測値との対応も取れるシステムとなっている。本研究ではバリオグラムは方向については考慮せず全方向バリオグラムを用いたが、一方方向バリオグラムについても取り扱う予定である。また、予測については、トレンドを考慮しておらず、さらに 1 つの特性値についてのみ対応しているが、今後普遍クリギングや、コクリギングについても取り扱い、これらの解析も可能にする予定である。さらに、経験バリオグラムを求める際のパラメータである cutoff や width についてはインタラクティブに選択・指定出来るようになっているが、計算時間も考慮しつつ、最適な値をシステムの中で得られるようにする予定である。

参 考 文 献

- Cressie, Noel (1993). *Statics for Spatial Data*, Wiley, New York.
- Fujino, T., Yamamoto, Y. and Tarumi, T. (2004). Possibilities and problems of the XML-based graphics in statistics, COMPSTAT2004 Proceedings in Computational Statistics, 1043–1052.
- 久保田貴文, 垂水共之(2004). 地理統計システムにおける SVG の利用について, 日本計算機統計学会 第 18 大会予稿集, 165–168.
- 久保田貴文, 垂水共之(2005). バリオグラム推定におけるカットオフの選択について, 日本計算機統計学会第 19 大会予稿集, 115–118.
- Kubota, T., Iizuka, M., Fueda, K. and Tarumi, T. (2005). The selection of the cutoff in estimating variogram model, The 5th IASC Asian Conference on Statistical Computing, 97–100.
- 間瀬 茂, 武田 純(2001). 『空間データモデリング』, 共立出版, 東京.

Visualizing Geostatistical Data by Using GoogleMapsAPI

Takafumi Kubota¹ and Tomoyuki Tarumi²

¹Graduate School of Humanities and Social Sciences, Okayama University

²Admission Centre, Okayama University

When we perform a geostatistical analysis, it is very important to be able to visualize the geostatistical data. By using “gstat” or “geoR”, which is the library of R, it is possible to analyze geostatistical data exploratory in R environment. However when we publish the output of geostatistical analysis, we want to handle it interactively, and to draw it on the map. Thus, we use GoogleMapsAPI and Scalable Vector Graphics (SVG) to display the plot of observed data, spatial dependence, and predicted data.