

セミパラメトリック GWR モデリングによる空間分析

—社会関係と主観的健康の関連性にみられる地域差—

中 谷 友 樹

I. はじめに

地理的加重回帰法 (Geographically weighted regression: GWR)¹⁾ は、複数の変量間の共変動関係にみられる地域差を推定するローカル回帰分析手法の 1 種である。回帰分析では、被説明変数と説明変数との関係を、(偏) 回帰係数の推定を通して確認するが、GWR ではこの回帰係数の地理的な変動 (変異) を推定する。例えば、地理的単位毎に観察される死亡率の水準を、地域の富裕度に関連した所得水準指標から推定する回帰分析の課題では、死亡率が被説明変数であり、所得指標が説明変数、説明変数 1 単位の増加がもたらす被説明変数の変化の期待値が回帰係数として推定される。ここで地域の所得水準を測る指標が小地域では得にくく代替的な指標が利用される場合がある。例えば、英国では 1 世帯あたりの自動車保有台数がしばしば利用されてきた²⁾。しかし、都市的地域内ではそのような関係がみられたとしても、生活交通手段として自家用車の保有が必須であるような農村的域では、自動車の保有と所得の相関関係はみられないかもしれない。また、都市的地域にくらべて、農村的域では、生活水準の維持に必要な所得水準が低く、低所得でも相互扶助的な仕組みで健康が維持されるような状況が仮にあるとすれば、都市圏内部で所得と死亡率の関係はより明瞭であるかもしれない³⁾。こうした指標の内容や指標間の関連性に潜む地理的文脈性を反映して生じる変数間の「関連性の地域差」を、局所的な重み付け回帰を繰り返し、回帰パラメーターの変異として推定する手法が GWR である。推定された回帰係数の地域差は、分布図に視覚化できる新たな空間データとして生成されるため、GIS とも親和性の高いローカルな空間分析の代表的手法の 1 つに位置づけられている⁴⁾。

同様に回帰分析におけるパラメーターの地理的な変動 (変異) をモデル化する手法には、座標変数のべき級数を利用する空間的展開法 (spatial expansion method)⁵⁾ や、近年では階層ベイズの枠組みを利用した地理的な変動係数モデル⁶⁾ が提案されてきた。これらに比べると、GWR は地理的に変動する係数の推定方法が直感的であり、また事前にパラメーターの分布について明確な仮定をおかずに推定するノンパラメトリック回帰としての特性を備えている点が特徴的である。そのため GWR は、指標間の関連性に関する地理的な文脈性を検討するための探索的な空間分析 (Explanatory spatial data analysis: ESDA) 手法として、普及してきた⁷⁾。

一方で、GWR ではパラメーターの空間変異を許すことで、モデルの複雑さは増大し、説明変数が多いと結果の解釈が難しくなる。そのため、モデルを単純化させるために、地理的な変異のない (固定された) 回帰係数を伴う説明変数の項を一部含むように修正された GWR を、mixed GWR (混合 GWR)⁸⁾ あるいはセミパラメトリック GWR (Semiparametric GWR: S-GWR)⁹⁾ と呼ぶ。これを実行するソフトウェアとして、筆者らが開発したソフトウェアが GWR4.0 (本稿執筆時点はバージョン 4.09)¹⁰⁾ であり、アイルランド国立大学メイヌース校の国立ジオコンピューテーション・センター等か

ら、フリーソフトウェアとして配布している¹¹⁾。S-GWR および GWR4.0 について日本語で解説した記事や応用事例が乏しいことを踏まえ、本稿では、この分析環境を利用した S-GWR モデリングの実例を提示し、その基本的考え方と有用性を検討する。

S-GWR による空間分析の経験的事例として、本稿では居住する近隣地区内での社会関係と主観的健康の関連性に着目する。健康地理学や社会疫学では、社会関係資本として概念化される近隣居住者間の信頼関係や規範意識、結合関係が、直接的・間接的に居住者の健康に関与する近隣の社会環境として機能していることが議論され、多くの事例研究が実施されてきた¹²⁾。ただし、研究の蓄積が進む一方で、実施されてきた過去の対象や指標、分析された時期・状況も多様であり、社会関係と健康の社会的文脈性に着目した検討の必要性を指摘しうる。とくに都市社会において社会関係のあり方には、居住者のとる生活様式に応じて近隣コミュニティ間の特徴的な違いが存在する¹³⁾。生活様式によっては、近隣との関係が生活を支えるものと感じられる場合もあれば、精神的負担と感じられる状況も想定され、これを踏まえた居住地の選択もなされていると考えられる。そのため、一元的に近隣との社会関係が健康に寄与するのではなく、そうした居住地の社会的性格が、社会関係と健康との関連性に違いを生んでいるかもしれない。そこで、本稿では、S-GWR を利用して、都市圏内の居住地の位置に着目しながら近隣の社会関係と主観的健康の関連性にみられる地域差を検討してみたい。

以下本稿では、II では S-GWR の定式化、III では S-GWR を利用した分析の有用性、IV では S-GWR を利用するために開発された GWR4.0 の機能の概要を示す。V において、S-GWR を利用した健康地理学の分析事例を示し、VI において本稿の内容を総括する。

II. S-GWR モデル

係数の地理的変異を想定しない通常の重回帰モデルは、以下のように示される。

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \varepsilon_i = \sum_k \beta_k x_{k,i} + \varepsilon_i, \quad (1)$$

ここで、 i はサンプルに関する添字であり、分析単位が住所地のような地点で示されるものであれば、 i 番目のサンプル地点となる。 y_i は地点 i の被説明変数、 x_{ki} は同地点の k 種類目の説明変数であり、 β_k がこの説明変数に関する（偏回帰）係数である（ここでは $x_{0i} = 1$ としている）。また、 ε_i は誤差項であり、互いに独立な平均 0、分散 σ^2 の正規分布 $N(0, \sigma^2)$ に従うと仮定する。

これに対して GWR では、以下のように係数が位置座標の関数として定式化される¹⁴⁾。

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1,i} + \beta_2(u_i, v_i)x_{2,i} + \cdots + \varepsilon_i = \sum_k \beta_k(u_i, v_i)x_{k,i} + \varepsilon_i, \quad (2)$$

ここで、 (u_i, v_i) は地点 i の地理座標であり、この位置に応じた係数の滑らかな地理的変異が想定されている。ただし、各地点（の位置）固有のパラメーターを推定するには、自由度が足りないため、GWR では地点 i の係数 $\{\beta_k(u_i, v_i)\}$ を推定するために、地点 i のみならずその周辺の地点データを含めたサンプルのサブセットを利用して回帰分析を行う。より詳細には、地点 i で最大値をとり、地

点 i から離れるほど値が小さくなるカーネル関数を利用した重み w_{ij} を利用した重み付き最小二乗法により、地点 i のローカルな係数を求める。すなわち、次のような最小化問題の解となる係数を求める。

$$\min_{\{\hat{\beta}_k(u_i, v_i)\}} \sum_j \left(y_j - \hat{y}_j(\{\hat{\beta}_k(u_i, v_i)\}) \right)^2 w_{ij}, \quad (3)$$

これをすべての地点で繰り返すことにより、局所的な係数の地理的分布を得る。ただし、 $\hat{\cdot}$ は推定値であることを示す。典型的な重みには、次のようなガウス型関数がある。

$$w_{ij} = \exp \left(-\frac{1}{2} \left(\frac{d_{ij}}{G} \right)^2 \right), \quad (4)$$

ここで d_{ij} は地点 i と j の間の距離、 G は推定に利用する実質的な地理的範囲を制御するパラメータであり、バンド幅と呼ばれる。バンド幅が大きいほど、局所的な係数の推定に利用される地理的な範囲は広くなり、結果として、得られる係数の地理的変異は小さくなる。 G を全体で1つの数値に固定すると、重み付けに利用する局所的な範囲は地理的に常に一定の大きさとなる。これを固定カーネル **fixed kernel** と呼ぶ。この方法では、推定に利用する局所的なデータ利用範囲が明確であるため解釈が容易となる反面、対象とする地域の周辺部では利用できるデータ数が乏しくなり、推定が不安定化する懸念がある。

これに対して、データの分布状況に応じて可変的にカーネル関数による重み付けの範囲を変更する手法は、適応的カーネル **adaptive kernel** と呼ばれる。典型的には、次のようなバイスクエア関数を利用したカーネルがよく利用される。

$$w_{ij} = \begin{cases} \left[1 - \left(d_{ij} / D_{i(B)} \right)^2 \right]^2 & \text{if } d_{ij} < D_{i(B)} \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

ここで、 $D_{i(B)}$ は、地点 i からみて B 番目に近い地点までの距離であり、この距離によって示される範囲が局所的な重み付けの範囲となる。バンド幅は、パラメータ B によって制御され、 $B=100$ であれば、回帰モデルの係数の推定にあたって常に周囲100地点程度の範囲のデータが利用されることになる。バンド幅を決定するパラメータ G や B の決定には、CV（クロスバリデーション）スコア、AIC、AICcなどのモデル比較の指標を利用して異なるバンド幅パラメータ値を持つモデルを複数比較し、最適なバンド幅パラメータを選択すればよい。なお、GWRの適合度評価には、通常の決定係数や逸脱度（残差二乗和）等の適合度診断指標が利用できる。

この基本的なGWRを拡張したS-GWRでは、予測値を構成する説明変数の項に、地理的な変動を許される係数を伴ったローカル項と、係数の地理的変異を想定しないグローバル項を混在させる。

$$y_i = \sum_k \beta_k(u_i, v_i) x_{i,k} + \sum_l \beta_l x_{i,l} + \varepsilon_i, \quad (6)$$

ここで、 $x_{i,l}$ は地点 i の l 種類目の説明変数であり、 β_l はその地理的な変異を想定しない係数である。このような形式のモデルは、カーネル関数によるノンパラメトリックなローカル項と通常のパラメトリックな線形回帰と対応するグローバル項を混在させているため、より一般にセミパラメト

リックな回帰モデル、あるいは部分線形モデルのように呼ばれるものと同型である。S-GWR の係数の推定には、ローカル項の係数推定と、グローバル項の係数推定を交互繰り返すバックフィッティング・アルゴリズムが利用できる¹⁵⁾。また、一般化線形モデル (Generalised linear model: GLM) のフレームワークを利用した拡張 (GWGLM および S-SGWGLM) も可能である¹⁶⁾。

Ⅲ. S-GWR モデルの有用性

S-GWR の利点として、第 1 には、重回帰モデルや GWR のいずれよりも統計学的に優れた予測力を持つモデルが得られる可能性を指摘できる。GWR は総じて重回帰モデルよりも適合度や未知データに対するモデルの予測力の高さが確認されるが、すべての係数の地理的変異を許すことにより過剰に複雑なモデルが推定されてしまう懸念もある。そのため、ローカル項とグローバル項の適当な組み合わせによって、グローバル項のみを利用する通常重回帰モデルよりも現実的で、かつローカル項のみを利用する GWR よりも簡潔な統計学的モデルとして、より統計学的な評価に優れたモデルが得られると期待される。

これに関連した第 2 の利点として、S-GWR を利用することにより、係数の地理的変異に関する統計学的な診断が容易なものとなる¹⁶⁾。ここで、係数の地理的診断を試みる特定のローカル項が 1 つあると考え、これを κ 種類目の変数と仮定する。ここで、現在関心をもっているモデル (κ 種類目の項がローカル項となっているモデル) をピボット・モデルと呼ぶことにする。これに対し、 κ 種類目の項がグローバル項である以外は、ピボット・モデルと同じモデルを、比較モデルと呼ぶことにする。ここで問題とすべき係数変異の診断は、この 2 つのモデルの比較問題に帰着する。例えば、AICc をモデル比較の基準とした場合、ピボット・モデルの AICc が、比較モデルのそれよりも小さければ、係数の地理的変異を想定したピボット・モデルの方が好ましく、 κ 項の係数は地理的に変異すると考える方が適当と診断されることになる。逆に、ピボット・モデルの AICc が比較モデルのそれよりも大きければ、 κ 項の係数は地理的に変異すると考えない方が適当と診断される。ただし、AIC や AICc の差の絶対値が 1 ないし 2 程度の場合には、有意な差は乏しく、経験則的には 4 以上の差がある場合に (とくに 10 位上の場合には確実に) 明確な判断が可能である点に留意すべきであろう¹⁷⁾。

より仮説検定の枠組みに基づく診断が好ましい場合には、以下に示すような F 統計量による F 検定を利用した診断も可能である。

$$F_{\kappa} = \frac{(RSS_{\text{比較モデル}\kappa} - RSS_{\text{ピボット・モデル}}) / (DOF_{\text{ピボット・モデル}} - DOF_{\text{比較モデル}\kappa})}{RSS_{\text{ピボット・モデル}} / DOF_{\text{ピボット・モデル}}}, \quad (7)$$

ここで、 RSS はモデルの残差 2 乗和、 DOF はモデルの自由度である¹⁸⁾。 κ 番目の項に関する比較モデルとピボット・モデルとの間に適合度の差がないとする帰無仮説 (κ 番目の項の係数について、地理的変異がないとする仮説) が正しければ、この F 統計量は、自由度 ($DOF_{\text{ピボット}} - DOF_{\text{比較モデル}\kappa}$, $DOF_{\text{ピボット・モデル}}$) の F 分布に従う。すなわち、当該の F 統計量が適当な F 分布から得られる有意水準 $\alpha\%$ の検定棄却域に含まれない場合、帰無仮説を棄却し、 $\alpha\%$ 水準で係数の地理的変異を有意と判断する。

第3に、係数の地理的な変異に関する先験的な知見や分析上の仮定を反映させることで、より焦点を絞った空間分析が可能となる。例えば、死亡率のような健康関連の指標を被説明変数とする場合、性・年齢による生物学的な違いが広く観測され、その影響を調整した上で分析したい場合がある。この場合には、性・年齢に関する説明変数の項をグローバル項とする S-GWR により、特定の説明変数に焦点をあて、より解釈の容易な形で、係数の地理的な変異の検討が可能となる。

IV. GWR4.0 とその分析機能

GWR4.0 は、S-GWR とその GLM 拡張版 (S-GWGLM) を利用するために、筆者らによって開発された MS-Windows 用のアプリケーション・ソフトウェアである¹⁹⁾。標準的な GWR は、現在、様々な環境で利用可能であり、例えば、汎用的な GIS ソフトウェアとして普及している ArcGIS (ESRI 社) には GWR の分析ツールが実装されている。その一方、S-GWR/S-GWGLM については、利用できる環境は限定的である。

プログラムのコードは C および C++/CLI 言語によって記述され、.NET framework 上で動作する。GWR は局所的な係数の推定に関する繰り返し処理の多さから、計算負荷の高い分析手法の 1 つであり、GWR4.0 では、係数推定アルゴリズムの並列化処理とともに、行列計算ライブラリ Eigen v3²⁰⁾ を利用して高速化をはかっている。

インストール後に登録されるスタートメニューやデスクトップ上の起動アイコンからは、ウィンドウ・モードで GWR4.0 を起動できる。このウィンドウ・モードでは、全体は Data, Model, Kernel, Output, Execute の 5 つのタブに分かれており (第1図)、データの読み込み、モデルの指定、カーネル・バンド幅の選択方法、出力結果の保存、実行の、分析に必要な一連のステップを、順を追って遂行できるように設計されている。なお、バッチ処理用にコマンドライン上で動作させることも可能である。以下、ウィンドウ・モードにおける各タブ (分析ステップ) の概要を示す。

(1) Data

データ・ファイルは、CSV ファイルや dbase IV の形式に対応する。GWR 分析では、サンプルの位置座標が必要であり、被説明変数、説明変数に関するフィールドに加えて、2次元の座標値を与える 2 つのフィールドが必要である (経度・緯度でもよいが、距離計算上の負荷を考えると、適当な投影座標に基づいた x 座標、 y 座標の数値が望ましい)。また、この分析課題 (セッション) の名称を与え、これを出力結果に残すことができる。

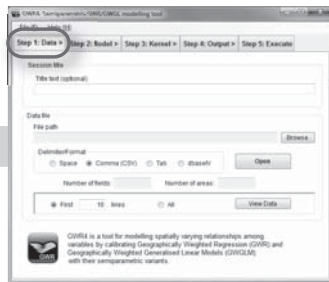
(2) Model

Data タブで適切にデータが読み込まれた場合には、フィールド (変数) のリストが画面中央の Variable (Field) List ボックスに表示される。ここから、S-GWR モデルのために必要な各変数のボックスへ、個別の変数フィールドを「移す」ことで、モデルの変数設定が可能となる (第2図)。

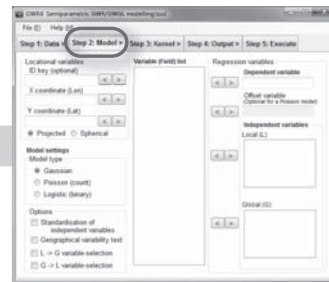
変数ボックスには、サンプルの位置に関する、識別用 ID (オプション)、 x 座標、 y 座標のボックス群 (画面左側) と、被説明変数、説明変数 (ローカル項とグローバル項用に 2 つのボックスに分かれている) の回帰モデル中の変数に関するボックス群 (画面右側) がある。なお、定数項 (“000 Intercept”) は、デフォルトではローカル項のボックスに含まれる。利用する説明変数がすべてローカル項である場合には、通常の GWR モデルに、グローバル項に 1 つでも変数を加えると、S-GWR モデルにな

Step1: Data>

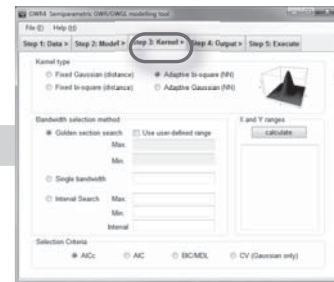
分析の開始（課題名を記録用につけることが可能）
データファイルを指定して開く。

**Step2: Model>**

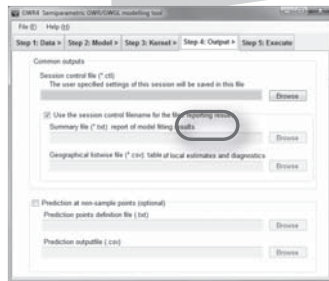
モデルの種類、変数の指定、係数の地理的変異に関する診断など、GWRモデルに関する詳細を設定する。

**Step 3: Kernel>**

地理的重み付けを行うカーネルの種類やバンド幅選択の方法を指定する。

**Step 4: Output>**

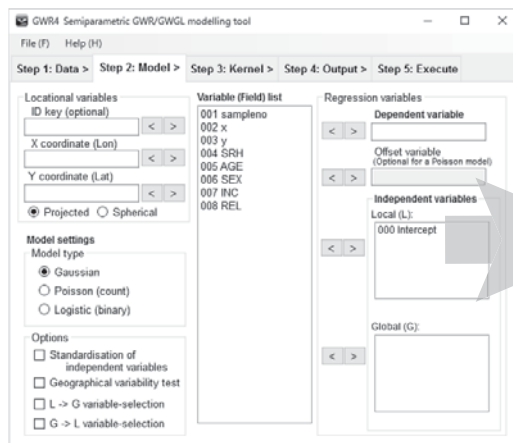
分析結果を保存するファイル名を指定する。

**Step 5: Execute>**

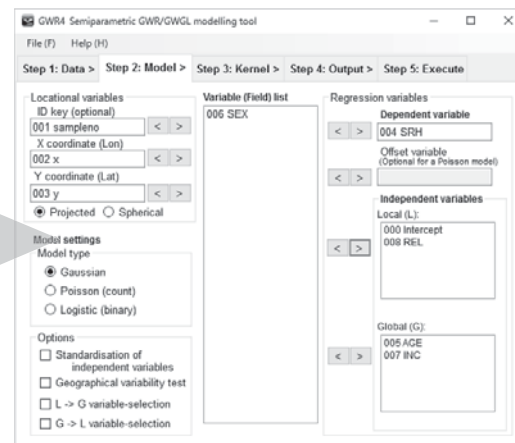
分析を実行し、その結果を確認する。



第 1 図 GWR4.0 の全体的構成



(a) データ読み込み後の初期状態の例



(b) S-GWR モデル設定の例

第 2 図 GWR4.0 によるモデル設定例

る。第 2 図には、次節で利用する S-GWR の分析時に使用した設定を示してあり、定数項と 1 つの変数（"008REL"）がローカル項に、2 つの変数（"005AGE" および "007INC"）が、グローバル項に設定されている。

モデルの種類としては、Gaussian（通常の回帰モデル）、Poisson（ポアソン回帰モデル）、Logistic（2 項ロジスティック回帰モデル）に対応した S-GWR が利用でき、いずれかを指定する（デフォルトは Gaussian）。その他、先に説明したローカル項の係数の地理的変異に関する診断については、

“Geographical variability test” オプションを有効とすることで実行される。モデル比較を利用して、自動的にローカル項とグローバル項を選択する変数選択オプションも利用可能だが²¹⁾、本稿ではその詳細は省略する。

(3) Kernel

ここでは、地理的重み付けを行うカーネル関数の種類とバンド幅選択の方法を指定する。デフォルトでは、適応型のバイスクエア・カーネル関数を利用し、AICc を最小化するバンド幅を、黄金分割探索アルゴリズムにより求める。データ件数が多い場合には、このバンド幅選択に多大な時間を要するため、解釈の容易な空間スケールを想定して単一のバンド幅をあらかじめ指定する、あるいは一定の刻み幅で、ある範囲内のバンド幅から最適な値を選択することも可能である。

(4) Output

出力されるファイルは、制御ファイル（分析作業を再現するための設定ファイル: 拡張子 .ctl）、モデル実行結果の要約ファイル（実行したモデルの内容や利用されたカーネル・バンド幅、モデルの適合度等の情報を格納する: 拡張子 .txt）、ローカル項の係数推定値リスト・ファイル（ローカル項の係数の推定値や局所的な適合度に関する指標値を格納する: 拡張子 .csv）の3種類である。ローカル項の係数推定値を地図化したい場合、GWR4.0 には地図描画の機能が含まれないため、ローカル項の係数推定値リスト・ファイルを GIS ソフトウェアで読み込み視覚化する。また、データ・サンプルのない位置座標での係数推定値を予測するオプション機能もある。

(5) Execute

Data から Output までのタブで設定された内容にもとづき、S-GWR モデルのデータへの当てはめが実行される。結果の要約が画面に表示されるが、同一の内容は Output タブで設定したモデル実行結果の要約ファイルに保存される。オプションとして、ここで設定された S-GWR モデルの分析をコマンドラインで利用するためのバッチファイルを作成する機能もある²²⁾。

V. 主観的健康感と社会関係との関連性に関する空間分析

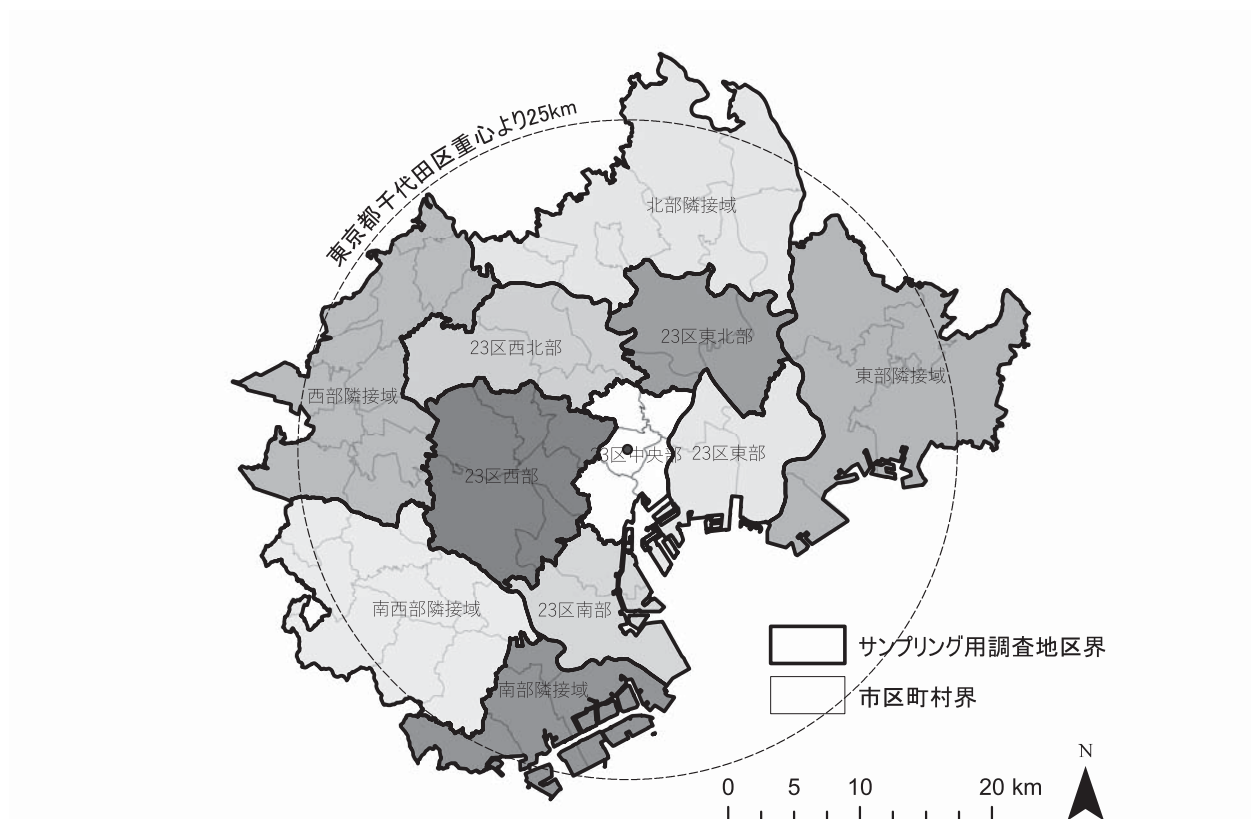
S-GWR による空間分析例として、本稿では都市圏内部に居住する人の主観的健康感（self-rated: SRH）の規定因として、居住する近隣での社会関係との関連性を想定した分析課題をとりあげる。SRH は自覚的健康度と訳される場合もあり、調査対象者による自身の健康度の主観的評価を指す。様々な縦断調査の結果から、SRH が良好であるほど、その後の死亡リスクが低下する傾向が、既往疾患等の既知の諸要因を調整した上で確かめられている²³⁾。SRH は、質問紙調査によって比較的容易に得られることもあり、健康の社会的決定因をめぐる調査研究でよく利用されてきた。

一方、ここで問題とする居住地域と関係した社会関係とは、近隣において生じる居住者間での価値観の共有や信頼関係、生活上の関わり合いの深さといった地域コミュニティとの関わり合いの度合いを反映するものである。近年の社会疫学研究ではとくにその近隣における機能に着目して、社会関係資本論による概念化と定量的な指標を利用した分析が進められ、様々な社会関係と健康指標との関連が報告されてきた²⁴⁾。この社会関係という視点からみた良好なコミュニティの状態は、様々な生活上のストレスの軽減、健康に資する生活習慣上の好ましい規範の醸成・維持、情報の共有を通しての良好な物的環境の利用、政策や好ましい資源配分への政治的影響力の発揮といった多様な

経路をへて、個人の健康に寄与すると考えられている。

その一方で、社会関係と健康のより一般的な議論の中では、社会関係が果たす否定的な機能も指摘されてきた²⁵⁾。匿名性の高い都市的な生活を求める個人にとって、近隣との親密な社会関係は個人のプライバシーへの干渉や過剰な負担感などストレス要因ともなり得る。都市社会学における生活様式に関する検討でも、都市圏内部での居住地に応じて居住者の生活様式と期待する近隣との社会関係に違いがあることが指摘されてきた。Greer²⁶⁾ のロサンゼルスにおける調査では、都市化の程度の高い社会地区では近隣との社会関係を重視せず、むしろ近隣からの干渉が少ない生活に満足覚え、都市化の程度の低い郊外の社会地区では、親密な近隣との社会関係に満足覚えをしていることが報告されている。こうした指摘を踏まえれば、社会関係と健康との関連性は、都市圏内部でも居住地の特性に応じて異なるかもしれない。すなわち、ここでの基本的な仮定は、良好な近隣との社会関係は、個人の良好な健康状態（SRH）と関連する（寄与する）が、その程度には居住地の違いに基づいた地域差が存在する、というものである。ただし、どのようにその地域差が認められるのかは事前に想定が難しいため、これを S-GWR を用いて地域差の全体的な傾向を俯瞰する。

利用する資料は、東京 23 区とその周辺に隣接する自治体（千代田区重心点より 25km 圏内に完全に含まれる市区）の居住者を対象に 2009 年に実施された「居住・生活様式と健康に関する調査」から得た。対象地域を第 3 図に示す。2 次医療圏を参考に分割した 11 地区毎に目標回収数を 200 サンプルとした上で、インターネット調査機関を利用して、登録されている調査モニターから、調査時に 30 歳以上 60 歳未満の男女を抽出した。ただし、一定期間毎に回収数が目標値を超えているかを確認してモニター登録者への回答依頼を順次追加する仕組みで調査が実施されているため、目標サンプル



第 3 図 分析対象地域とサンプリング区域
11 サンプリング区域に応じて市区を塗り分けてある。

数と最終的な回収数は厳密には一致せず、結果として2,613サンプルを回収している。なお、11地区それぞれにおいて性・年齢（10歳階級）別のサンプル分布は、国勢調査人口による比率とほぼ一致するように割付を実施した上で、回収を調整した。ただし、以下では、社会関係資本の健康への影響がより強いとされる²⁷⁾ 女性サンプル（ $n=1,089$ ）に限定した分析結果のみを示す。

サンプルの地理的座標は、居住する町丁字の代表座標とし、被説明変数である *SRH* については、「あなたの全般的な健康状態は」という問に対する、「とても良い」（1点）から「かなり良くない」（5点）までの5段階の選択肢に基づく回答値を利用した。社会関係については、先行した検討を参考に、「地域の一員であると感じるか」「近所の人たちとの付き合いはどの程度か」「住んでいる地域の住民は信頼できるか」の3つの設問に対する回答結果に対して、主成分分析を利用して合成得点 *REL* を得た²⁸⁾。この主成分得点は、平均0、標準偏差1に標準化されており、値が高いほど3つの指標からみた近隣との社会関係の程度が高い（地域の一員と感じる、生活上の付き合いの程度が深い、他の地域住民への信頼度が高い）ことを意味する。統制変数として、年齢 *AGE*（30歳代0、40歳代1、50歳代2）、等価世帯所得²⁹⁾ の平方根変換値 *INC* を追加し、以下のような重回帰モデルを設定した。

$$SRH_i = \beta_0 + \beta_A AGE_i + \beta_I INC_i + \beta_R REL_i + \varepsilon_i, \quad (8)$$

このモデルによる推定結果を第1表に示す。*SRH* は小さいほど良好な健康評価である点に注意すると、年齢が高いほど健康度が悪く、所得が高いほど健康度に優れるほか、本研究に着目する社会関係の係数に着目すると、良好な近隣社会との関係を得て居住している人ほど、健康度に優れる明確な関連性を確認できる。

これに対して、上述した通り、近隣との社会関係の程度を示す *REL* の係数に地域差を想定した S-GWR モデルを同じデータに適用した。ただし、特定の説明変数をローカル項として利用すると、その係数の地域差を反映して定数項の地域差が生じることも多いため³⁰⁾、ここでは以下のような *REL* の項と定数項をローカル項としたモデルを利用した。

$$SRH_i = \beta_0(u_i, v_i) + \beta_A AGE_i + \beta_I INC_i + \beta_R(u_i, v_i) REL_i + \varepsilon_i, \quad (9)$$

地理的重み付けを行うカーネルについては、バイスクエア関数による適応型カーネルとし、AICc を最小化する基準で最適な値としてバンド幅パラメータ $B=713$ を得た。すなわち、各サンプルについてその周辺の713サンプルを利用した重み付け回帰分析によってローカルな係数の推定値を得ている。適用した S-GWR モデルの係数の推定値とローカル項の係数に関する地理的変異の診断 (F 検

第1表 重回帰分析モデルの推定結果

	推定値	標準誤差	t 値
定数項	2.717	0.088	30.805
<i>AGE</i>	0.151	0.038	3.954
<i>INC</i>	-0.013	0.004	-3.398
<i>REL</i>	-0.154	0.025	-6.069

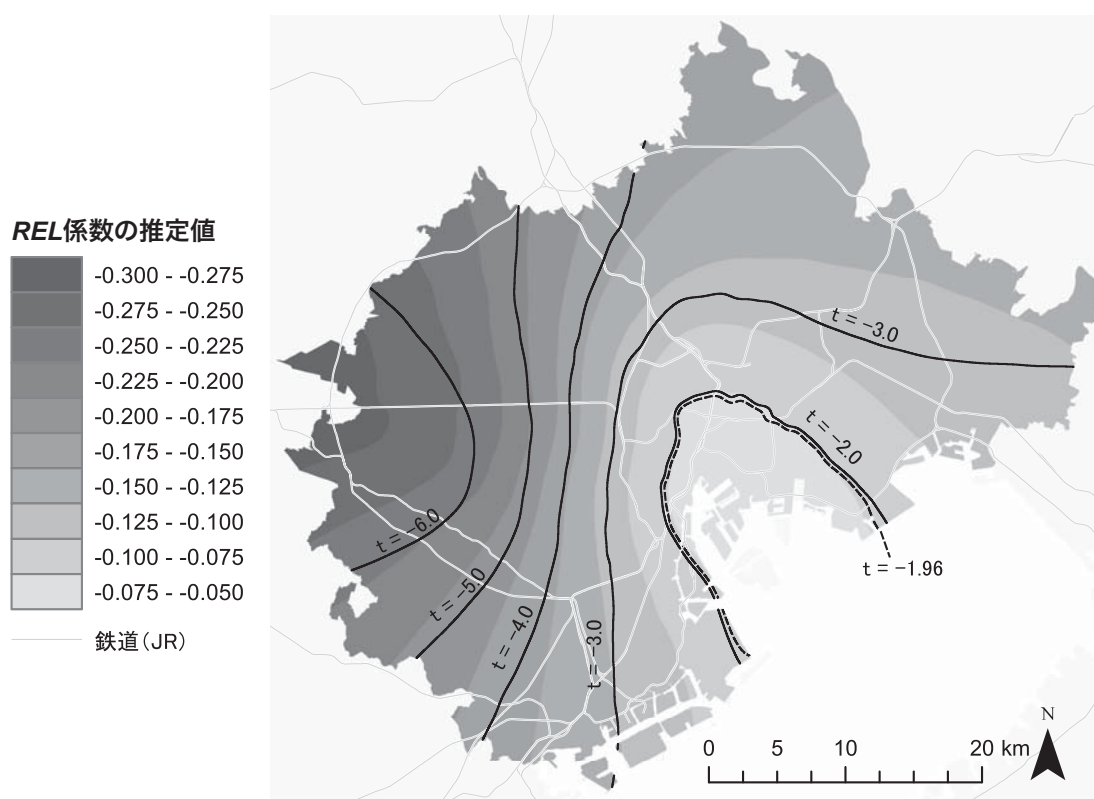
$R^2 = 0.047$, AICc = 2685.4, $n = 1,089$

第2表 S-GWR モデルの推定結果

グローバル項	推定値	標準誤差	t 値
<i>AGE</i>	0.163	0.043	3.817
<i>INC</i>	-0.009	0.006	-1.564

ローカル項	係数推定値				地理的変異検定		
	平均	標準偏差	最小値	最大値	F 値	自由度	p 値
定数	2.609	0.033	2.533	2.665	1.194	(2.807, 1079.423)	0.310
<i>REL</i>	-0.137	0.055	-0.287	-0.059	4.208	(2.743, 1079.423)	0.007

$R^2 = 0.060$, $AICc = 2682.0$, $n = 1,089$

第4図 S-GWR による社会関係指標 *REL* のローカル係数推定値分布

定) 結果を、第2表に示す。期待された通り *REL* の係数には1%水準で有意な地域差が認められる³¹⁾。

この *REL* 係数の分布を、主題図として示したものが第4図である。計算の詳細はここでは省略しているが、第4図にはローカルな係数の標準誤差に基づいた疑似 t 値の分布を示してある³²⁾。中心部付近では *REL* 係数値はゼロに近く、*SRH* への有意な寄与は認められない。その一方で中心部をとりまく多くの領域で、 t 値の絶対値は2.0を上回っており、郊外、とりわけ西側の地区ほど、*REL* 係数値が小さく、近隣との社会関係の程度が高いほど、*SRH* が優れる関連性がより明瞭となっている。

この結果からは、近隣における社会関係と *SRH* との関連性には地域差があり、中心部（都心）と

郊外という居住地に応じた社会関係の機能の違いが示唆される。例えば、郊外で子育てや長期的な暮らしを期待する人が多い地区ほど、近隣との社会関係によって居住地域での問題関心を共有し、個人の生活や健康に肯定的な機能を果たす情報の交換や社会的な支援を受けやすいといった状況が考えられる。一方で、*REL* 係数が有意でない対象地域の中心部は、かつての下町地域であり 1980 年代には居住人口の乏しさや高い高齢化率で特徴づけられた地域であったが、近年では人口増加率の著しい地域でもある³³⁾。こうした都心周辺での地域では、コミュニティの弱体化や、近隣との社会関係に依存しない生活様式をとる居住者の増加などから、当該地区では近隣の社会関係が必ずしも肯定的な機能を果たさない、近隣との社会関係を欠いていてもそれを補う別の社会関係や社会サービスが機能しているといった状況が、近隣との社会関係と健康との関連性を弱めているのかもしれない。

いずれにせよ *REL* 係数分布図の分布傾向から行う解釈は、あくまでも仮説的なものであり、断定的な判断はなし得ない。そのため、より確証的な分析として、Greer の研究のように居住者の生活様式の特徴的な違いをとらえる社会地区や個人の生活様式の類型を事前に設定した上で、より仮説検証的な議論を進めることが考えられる³⁴⁾。しかし、そのような事前の設定によって、注目すべき対象を見過ごしてしまうかもしれない、関連性の地域差の地図には、より確証的な分析へと進める際に必要な論点を「発見」する役割が期待される。例えば、ローカルな係数の分布と、サンプル属性や地域属性との対応関係を探索的に検討することで、新たに検討すべき課題も明らかになるかもしれない。*REL* 係数の分布には特徴的な東西の地域差が示されており、都心から離れた地区においても、西側の地区の方が *REL* 係数の絶対値が大きい (*SRH* との関連性が強い)。都心－郊外のような家族的地位に基づく居住地分化のみならず、近年の社会地図でも確認されるような、セクター的な社会経済的地位の居住地分化³⁵⁾ と関連する社会関係の機能の地域差も示唆される。

このように S-GWR とは、関連性の地域差の有無に関する仮説検証的な分析とあわせて、関連性の地域差そのものに関する探索的／発見的な分析を可能にする空間分析のアプローチと位置づけられる。

VI. 結論

本稿ではローカルな空間分析手法の 1 つとして知られる地理的加重回帰分析 (GWR) を改良したセミパラメトリック GWR (S-GWR あるいは混合 GWR) について、①その基礎的な定式化の解説、②専用ソフトウェア GWR4.0 の概要の紹介、③ S-GWR を利用した空間分析事例の提示を通して、その利用可能性と有用性を整理した。

S-GWR を利用することで先験的な仮説に応じて、問題とする関連性の地域差の有無を統計学的に診断する確証的な分析とともに、その具体的な内容の解釈にあたっては「関連性の地域差の地図」を通して探索的な分析が可能となる。本稿では、近隣との社会関係と主観的健康との関連性に絞った空間分析を試みた。社会関係が豊かであるほど自覚的な健康度の評価 (主観的健康) に優れる傾向が指摘されてきたが、東京 23 区とその周辺地域に居住する女性を対象とした分析結果からは、当該の関連性は中心部 (都心部) では観察されず、西側の郊外に行くほど明瞭である地理的な傾向が明らかとなった。それは、健康の社会的決定因としての社会関係を考えるにあたって、近隣との社会関係

が健康に寄与する機能を有するとしても、その機能を文脈づける地理的＝社会的背景が存在し、その背景に応じて社会関係と健康との関連性が異なりうることを示唆している。

ところで、本稿の事例分析では5段階のリッカート尺度に基づいた指標を被説明変数としており、順序回帰や多項ロジスティック回帰モデルの枠組みを利用した S-GWR モデルの拡張が望ましいかもしれない。統計モデリングとしては、S-GWR は一般化線形モデルの枠組みを利用した拡張 (S-GWGLM) が既に果たされており、GWR4.0 ではカウント・データの解析に適するポアソン回帰モデルの拡張 (S-GWPR)、2 値データの解析に適する 2 項ロジスティック回帰モデルの拡張 (S-GWLR) を利用できる³⁵⁾。順序回帰モデルへの拡張など、より多様な分析データの形式に対応できるモデルの実装は、外れ値や局所的な説明変数間の相関関係から生じる多重共線性問題³⁶⁾への対処とともに、空間データ解析環境の充実に向けた技術的課題として残されている。

付記

本研究における社会調査資料の作成にあたっては JSPS 科研費基盤研究 (B) 20300297 を、分析を進めるにあたっては基盤研究 (B) 15H02964 の助成を一部利用した。

注

- 1) ① Brunsdon C., Fotheringham A.S. and Charlton M.: Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28, 1996, pp. 281-289. ② Fotheringham, A.S., Brunsdon, C. and Charlton, M.: *Geographically Weighted Regression*. Wiley, 2002. ③中谷友樹「空間的共変動分析」杉浦芳夫 編『地理空間分析』朝倉書店, 2003, pp. 23-48.
- 2) Townsend, P., Phillimore, P. and Beattie, A.: *Health and deprivation: inequality and the North*, Croom Helm, 1988.
- 3) Nakaya, T.: Geographically weighted regression. Kemp, K. ed. *Encyclopaedia of Geographical Information Science*, Sage Publication, 2007, pp. 179-184.
- 4) ① Fotheringham, A.S.: Trends in quantitative methods I: Stressing the local. *Progress in Human Geography*, 21, 1997, pp. 88-96. ② Lloyd, C.D.: *Local Models for Spatial Analysis 2nd edition*. CRC Press, 2011.
- 5) Casetti, E.: Generating models by the expansion method: applications to geographic research. *Geographical Analysis*, 4, 1972, pp. 81-91.
- 6) Assunção, R.: Space varying coefficient models for small area data, *Environmetrics*, 14, 2003, pp. 453-473.
- 7) 日本の地理学に関連した学会誌でも、例えば以下のような分析事例が報告されている。①磯田 弦・渡邊 俊介「メッシュ統計と地理的加重回帰を用いた食品スーパーの立地確率の算出」統計, 65 (8), 2014, pp. 16-21. ②上杉 昌也・樋野 公宏「東京都区部における住宅侵入盗と近隣社会経済的特性の関係: 空間的異質性および物理的環境との相互作用効果に着目して」都市計画論文集, 50 (3), 2015, pp. 608-615. また、階層ベイズ・モデルと GWR を融合させてより安定したパラメータの地理的変異を推定するアプローチも提案されている。③ Alipoureshliki, S., Osaragi, T. and Oki, T.: Parameter smoothing specification in Bayesian geographically and temporally weighted regression: case study of analysing spatiotemporal distribution of population in Tokyo metropolitan area, *GIS 理論と応用*, 24-2, 2016, pp. 73-84.
- 8) 前掲 1) ②。
- 9) Nakaya, T., Fotheringham, A.S., Charlton, M. and Brunsdon, C.: Geographically weighted Poisson regression for disease associative mapping, *Statistics in Medicine*, 24, 2005, pp. 2695-2717.
- 10) ① Nakaya, T., Fotheringham, S., Charlton, M. and Brunsdon, C.: Semiparametric geographically weighted generalised linear modelling in GWR4.0. *Proceedings of Geocomputation 2009*, 2009, 5p. ②

- Nakaya, T.: Geographically weighted generalised linear modeling. Brunson, C. and Singleton, A. eds. *Geocomputation: A Practical Primer*, Sage Publication, 2016, pp. 201-220.
- 11) <http://gwr.maynoothuniversity.ie/gwr4-software/> (最終閲覧 2016 年 1 月 6 日)
- 12) ① Kawachi, I. and Berkman, L.F.: Social Cohesion, Social Capital, and Health. Berkman, L.F., Kawachi, I. and Glymour, M.M. eds.: *Social Epidemiology Second Edition*, Oxford University Press, 2014, pp. 290-319. ② Murayama, H., Fujiwara, Y. and Kawachi, I.: Social capital and health: a review of prospective multilevel studies. *Journal of Epidemiology*, 22, 2012, pp. 179-187.
- 13) ① Greer, S.: Urbanism Reconsidered: A Comparative Study of Local Areas in a Metropolis. *American Sociological Review*, 21, 1956, pp. 19-25. ② 浅川達人「東京における社会的つながり」家計経済研究, 87, 2010, pp. 12-21.
- 14) ここでの GWR の標準的な説明については、前掲 1) の文献類に基づく。
- 15) 前掲 9)。なお、当初混合 GWR を提案した前掲 1) ②の研究では、バックフッティングとは異なる推定原理による係数の推定を行っている。
- 16) 前掲 10) ②。
- 17) Burnham, K. and Anderson, D.R.: *Model selection and multimodal inference: A practical information-theoretic approach 2nd Edition*. Springer, 2002.
- 18) 前掲 10) ②。ただし、S-GWR/S-GWGLM の自由度の計算法については、前掲 9) を参照されたい。
- 19) 2011 年にアイルランド国立大学メイヌース校の国立ジオコンピューテーション・センターの専用ウェブサイトから、一般への配布を開始した。前掲 11) を参照。
- 20) Jacob, B., Guennebaud, G. and others: Eigen v3, <http://eigen.tuxfamily.org>, 2010.
- 21) 前掲 10)。
- 22) コマンドラインでは、以下のように、GWR4.0 のプログラム本体 sgwrwin.exe を、制御ファイルを引数として実行すれば、ウィンドウ・モードと同じ結果が得られる。
 >"C:\Program Files(x86)\GWR4\sgwrwin.exe" C:\workingfolder\mysession.ctl
 ただし、プログラムのインストール先や作業フォルダは、ユーザによって異なる。この実行用のコマンドライン記述を生成する機能が Execute タブに含まれている。
- 23) Idler, E. L. and Benyamini, Y.: Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, 38, 1997, pp. 21-37.
- 24) 前掲 12)。
- 25) 杉澤秀博「健康の社会的決定要因としての社会関係：概念と研究の到達点の整理」季刊社会保障研究, 48, 2012, pp. 252-265.
- 26) 前掲 13) ①。
- 27) 前掲 25)。
- 28) 類似した調査による別の社会調査資料を利用した埴淵・中谷では因子分析を利用して、近隣の社会的結束性と解釈しうる因子を抽出している。そこでは、本稿の分析で利用した 3 つの設問への回答に加えて、居住地への愛着に関する設問への回答を含めていた。ただし、地域への愛着は結束性の結果とも想定されるため、ここではこれを除いた変数の合成を行っている。
- 29) 等価世帯所得とは、世帯規模を調整した世帯所得指標であり、ここでは世帯の年間収入（万円単位）を世帯人員の平方根値で割った値で定義される。
- 30) 前掲 3)。
- 31) GWR4.0 では F 値と自由度のみを示し、検定の p 値は算出されない。ここでは、統計分析環境 R の pf 関数を利用して、 $1 - pf(F \text{ 値}, \text{自由度 } 1, \text{自由度 } 2)$ により、 F 分布の上側累積確率として p 値を得た。
- 32) 前掲 1)。
- 33) 小泉 諒・若林芳樹「バブル経済期以降の東京 23 区における人口変化の空間的パターン」地学雑誌, 123, 2014, pp. 249-268.
- 34) 社会地区類型を利用した国内での健康地理学の分析事例には、次のものがある。中谷友樹・埴淵知哉「居住地域の健康格差と所得格差」経済地理学年報, 59, 2013, pp.57-72.

- 35) 前掲 33)。
- 36) 前掲 9) には S-GWPR の分析事例が、前掲 10) ②には S-GWLR を利用した分析事例が示されている。
- 37) Wheeler, D.C.: Diagnostic tools and a remedial method for collinearity in geographically weighted regression, *Environment and Planning A*, 39, 2007, pp. 2464-2481.

(本学文学部教授)

Spatial Analysis Using Semiparametric GWR: The Effects of Social Relationships on Self-rated Health in the Tokyo Metropolitan Area

by
Tomoki Nakaya

Geographically weighted regression (GWR) is a type of local regression analysis for capturing spatial variations in relationships between variables with geographical coordinates. Semiparametric GWR (S-GWR), or mixed GWR, is a variant of GWR that incorporates explanatory terms with both fixed and geographically varying coefficients. This study aims to summarise its advantages for effective spatial analysis by (1) explaining its basic specification, (2) introducing GWR4.0, a Windows software that we developed for S-GWR modelling, and (3) demonstrating how S-GWR works, using empirical data.

In terms of statistical modelling capability, compared with both traditional GWR and conventional regression models, S-GWR can attain a better statistical performance through finding a proper combination of fixed and varying coefficients in one model. Another practical advantage of S-GWR is its flexibility in combining (a) confirmative analysis through testing local variability of specific coefficients and (b) exploratory analysis through mapping local/varying coefficients to detect spatial patterns of 'relationships between variables'.

We present a case study of an S-GWR-based spatial analysis of the associations between neighbourhood social relations and self-rated health among middle-aged women living in the Tokyo metropolitan area in Japan ($n = 1,089$). We fit an S-GWR model to the sample data, employing self-rated health as the dependent variable and a composite indicator of neighbourhood social relationships as the independent variable with varying coefficients while controlling for age and household income. Numerous studies in social epidemiology and health geography, particularly those with a special interest in social capital, have indicated that good neighbourhood social relationships were positively associated with various health outcomes, including self-rated health. However, our S-GWR modelling revealed a significant geographic variation in such associations. Whereas good neighbourhood social relationships tended to be significantly and positively associated with good self-rated health mainly in suburban areas, such associations were not observed in the central part of the study region. The finding highlights the importance of considering the socio-geographical context of neighbourhood social relationships as a social determinant of health in the urban society.

Keywords: semiparametric local regression, geographically weighted regression, GWR4.0, social capital, social relationships, self-rated health