

# Geographically weighted regressionモデルを用いた 建築物ストックの詳細推計

村上 大輔<sup>1</sup>・瀬谷 創<sup>2</sup>・山形 与樹志<sup>3</sup>・堤 盛人<sup>4</sup>

<sup>1</sup>国立環境研究所 地球環境研究センター (〒305-8506 つくば市小野川16-2)  
E-mail: murakami.daisuke@nies.go.jp

<sup>2</sup>広島大学 大学院国際協力研究科 (〒739-8529 東広島市鏡山1-5-1)  
E-mail: hsey@hiroshima-u.ac.jp

<sup>3</sup>国立環境研究所 地球環境研究センター (〒305-8506 つくば市小野川16-2)  
E-mail: yamagata@nies.go.jp

<sup>4</sup>筑波大学 システム情報系 社会工学域 (〒305-8573 つくば市天王台1-1-1)  
E-mail: tsutsumi@sk.tsukuba.ac.jp

将来の気候・社会シナリオの構築を適正に行うには都市分布の推計が必要となる。我々は、都市化指標の一つである建築物ストックに着目し、同データを詳細推計するためのダウンスケール手法を、geographically weighted regression (GWR) モデルを拡張することで構築してきた。しかしながら、この手法には、ダウンスケールの際の各地域の重みが既知でなくてはならない点や、従来のGWRモデルと同様、多重共線性の影響を受けやすい点などの課題が残されている。本研究では、まず、以上の問題の議論を踏まえた上で、GWRモデルに基づいた手法を再構築する。次に、修正した手法の有用性を既存のダウンスケール手法との比較により検証する。続いてこの手法を用いて市区町村別の建築物ストック量を推計する。また、推計された詳細建築物ストックデータの一適用例として電力需要推計についても議論する。

**Key Words :** *geographically weighted regression, building stock data, downscaling, electricity demand*

## 1. はじめに

将来の気候・社会シナリオの構築に際しては、例えば人口や建築物ストック（本研究では建物床面積と同意）といった、都市分布を記述する空間データの整備が必要となる。しかしながら、そういったデータは必ずしも空間詳細な単位で入手可能とは限らない。

この点を踏まえ、村上ら(2013)は空間詳細な建築物ストックデータを推計するためのトップダウン（ダウンスケール）・アプローチを、geographically weighted regression (GWR : 例えば Fotheringham et al., 2002) を拡張することで構築している。この手法は、従来の GWR と同様、回帰係数を空間的に可変としており、広域を対象とすることに起因する空間的な異質性を捉える点で優れている。一方で、この手法にはいくつかの課題も残されている。まず、推計精度向上のための補助データを十分に活用することができない点である。ダウンスケールを議論した既往研究の多くが、ゾーン毎の重みを、可能であれば複

数の補助データを考慮ながら、柔軟に与えることが重要であると指摘している（第4章参照）。しかしながら、彼らはこの重みを一変数で先見的に与えることを仮定しており柔軟性に欠ける。また、(ii) 従来の GWR と同様、多重共線性の影響を強く受けるという点もまた課題の一つである。

そこで本研究では、(i)に対処するために、複数の補助データを考慮しながら各ゾーンの重みを柔軟に決定するダウンスケール・アプローチを構築する。ここでは、GWR を応用することで各変数の影響を場所毎に可変とする。加えて、(ii)を踏まえ、多重共線性への対処法についても議論する。

また、本手法を『建築物ストック統計』データのダウンスケールに応用することで、各市区町村のストック量を推計する。本データは国土交通省により平成 22 年 1 月 1 日より毎年整備・公開されているものである。この統計は、住宅・土地統計調査、法人建築調査、及び建築着工統計を元に作成された加工統計であり、「住宅・非

住宅 / 木造・非木造 / 昭和 25～（平成 3 年以降は 1 年毎、それ以前はおおよそ 10 年毎）」の各区分について、各都道府県内の床面積の総量を推計している（国土交通省、2010）。

なお、建築物ストックは将来シナリオ分析をはじめ様々な分析における重要な基礎データであり、ダウンスケール手法で整備することには当然批判もあろう。しかし、研究者・実務者が各々の方法で床面積を按分するよりは、一定の統計学的基盤を持つ手法で床面積データを按分し、それを統一的に用いる方が、応用研究のために有益であると筆者は考える。また、ダウンスケール法によるデータ整備は、[1] 広域への適用性、[2] 建築物ストック統計との整合性、[3] 複数年度での整備の可能性という点で利点もある。

以下、第 2 章では建築物ストックの重要性について整理し、第 3 章では GWR モデルを、第 4 章では既往のダウンスケールリング手法を、それぞれ概観する。第 5 章では、GWR をダウンスケール手法に拡張する。第 6 章では、本研究で開発した手法の特性を議論する。第 7 章では、提案手法の推計精度を既往の代表的な手法と比較する。その後、実際に用途別（住宅・非住宅）・構造別（木造・非木造）の市区町村建築物ストックを推計した例を紹介し、続いて推計データの適用の一例として電力需要推計を議論する。最後に第 8 章で本研究をまとめる。

## 2. 建築物ストックデータ

### (1) 建築物ストックデータの重要性

国土交通省 (2010)によると、建築物ストックは、(i)日本全体の建築物の維持修繕費の予測、(ii)建築物に関連する消費エネルギーの分析、(iii)取り壊しに伴う廃棄物の将来予測、(iv)法律改正の際の影響の推測など、様々な分析の基礎資料として重要となる。(ii)に関しては、実際に建築物ストックをエネルギー需要推計に応用した既往研究 (e.g., 平野ら, 2008 ; 谷川ら, 2010 ; Yamagata et al., 2013) もみられる。また、治水施設の整備によってもたらされる経済的な便益や費用対効果の計測の手法を整理した国土交通省の『治水経済調査マニュアル (案)』では、家屋の期待被害額を、床面積×都道府県別家屋 1 m<sup>2</sup> 当たり評価額で算出する。

以上を踏まえると、建築物ストック統計の有用性は明らかである。一方で、例えば詳細な将来シナリオの検討やコンパクトシティ政策といった市以下の領域での検討が重要になる各種計画に資するためには、例えば市区町村単位のような、より詳細な単位での建築物ストックデータの整備が必要となる<sup>1</sup>。

市区町村レベルの建築物ストックの推計方法には、我々の着目するダウンスケール・アプローチのほかに、ボトムアップ・アプローチがある。次節では、市区町村単位の建築物ストック量をボトムアップにより推計した既往研究をレビューした上で、ダウンスケールによる建築物ストック推計の意義を明らかにする。

### (2) 市区町村単位の建築物ストック推計のためのボトムアップ手法のレビュー

市区町村レベルでの建築物ストックデータの整備のためのソースとしては、「固定資産課税台帳」（自治税務局固定資産税課）と「都市計画基礎調査」（各自治体）が挙げられる（阪田・吉川, 2001）。

前者は、作成にあたって自治体職員らによる実地調査が行われており、毎年更新がなされている。また、固定資産税上の建物用途は課税を目的としているため正確であり、現状では、個別の家屋に関する用途別建物床面積を高精度かつ悉皆的に入手するには最良の方法であるとされる（阪田・吉川, 2001）。しかし、全国市区町村について、複数年度で固定資産課税台帳データを整備するのは短期間には困難である。

一方後者は、「都市計画基礎調査」に基づいて作成される GIS データである。これは、都市計画基礎調査における土地利用現況調査に基づいて 5 年毎に作成されているものであり、基本的に航空測量と現地調査に拠っている。阪田・吉川 (2002)は東京都の GIS 建物ポリゴンデータから 1km メッシュ毎の床面積データを推計している。しかしながら、我々が宮城 (2009)、堤ら (2012)において各自治体に問い合わせた経験では、東京都市圏において都市計画基礎調査に基づくデータの整備・提供を行っている自治体は、現状では千葉県と神奈川県のみであり、同データが全国で使用できるようになるには、まだまだ時間がかかるように思われる。

用途を住宅に限れば、建築物ストック統計の元データの一つともなっている、住宅土地統計調査により、市区町村別に住宅数および住宅あたり延べ面積が取得できる。また、国勢調査で得られる一人（世帯）当たり床面積に、人口（世帯数）を乗じることで総量の推計を行うことも可能である。しかし、これらの調査では、非住宅の床面積は把握されていない。非住宅については、上述の都市計画基礎調査や固定資産課税台帳、株式会社ゼンリンの Zmap-TOWN II 等にソースが限られており、筆者らの知る限り、全国横断的に統計情報として整備するような取り組みは行われていない<sup>2</sup>。従って、非住宅ストックの

照されたい。

<sup>2</sup>日本建設情報総合センターが、治水経済計算用に固定資産課税台帳をベースとした 100m メッシュ延床面積データを提供しているが、本データでは用途が一色丹になっており、住宅・非

<sup>1</sup>諸外国の建築物データの整備状況については伊藤ら(2005)を参

詳細推計は、ダウンスケールが有用な一例といえよう。

建築物ストック推計の既往研究としては、奥村ら (2012)、小川ら (2013)などが挙げられる。奥村ら (2012)は建築物ストック統計の非住宅データを用途地域別に按分している。また、小川ら (2013)は、個別建物毎の属性（木造/非木造、築年数等）を推計し、災害リスク分析に応用している。しかし、住宅・非住宅それぞれの建築物ストックデータベースを全国市区町村で整備した例は、筆者らのレビューした範囲では存在しない。

以上を踏まえ、本研究では、以降で構築する手法を用いて全国市区町村の住宅・非住宅ストックデータを整備する。

### 3 地理的加重回帰モデル：GWR

GWR は(1)式を用いて地点  $s \in D \subset \mathcal{R}^2$  の観測データ  $y_s$  をモデル化するアプローチである。

$$y_s = \sum_p x_{s,p} \beta_{s,p} + \varepsilon_s \quad E[\varepsilon_s] = 0 \quad \text{Var}[\varepsilon_s] = \sigma^2 \quad (1)$$

ここで  $p$  は説明変数の添え字、 $x_{sp}$  は説明変数、 $\sigma^2$  は分散パラメータである。 $\beta_{sp}$  はトレンドパラメータを表し、これを地点毎に変えることで空間的な異質性を考慮する。

$\beta_{sp}$  の推定には weighted least squares (WLS) を用いる。 $\beta_s = [\beta_{s,1}, \dots, \beta_{s,p}, \dots, \beta_{s,p}]'$  とすると、その推定量は次式で与えられる。

$$\hat{\beta}_s = (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{X}'_s \mathbf{y}_s \quad (2)$$

$$\mathbf{X}_s = [g(h_{s,\tilde{s}})^{1/2} x_{\tilde{s},p}]_{\tilde{s},p} \quad \mathbf{y}_s = [g(h_{s,\tilde{s}})^{1/2} y_{\tilde{s}}]_{\tilde{s}}$$

ここで「 $'$ 」は転置、 $[\bullet]_{\tilde{s},p}$  はその第  $(\tilde{s}, p)$  要素が  $\bullet$  で与えられる行列、 $[\bullet]_{\tilde{s}}$  はその第  $\tilde{s}$  要素が  $\bullet$  で与えられるベクトル、 $g(h_{s,\tilde{s}})$  は地点  $s$  に対する地点  $\tilde{s}$  の重みを表す地点  $s$ - $\tilde{s}$  間の距離  $h_{s,\tilde{s}}$  の関数である。 $g(h_{s,\tilde{s}})$  は例えばガウス関数(3)式でモデル化できる。

$$g(h_{s,\tilde{s}}) = \exp\left[-\left(\frac{h_{s,\tilde{s}}}{b}\right)^2\right] \quad (3)$$

ここで  $b$  はバンド幅パラメータである。(2)式は空間的に近接した観測値により大きな重みを与えるという条件の下での WLS 推定量である。

$b$  は leave-one-out cross-validation (LCV) で最適化できる。LCV の手順は次の通りである：(i)  $y_s$  以外の全サンプルを

住宅を分類することができない。

用いて  $y_s$  を予測する；(ii)手順(i)の予測を全サンプルについて行う；(iii) mean squared prediction error (MSPE) を評価する；(iv) 手順(i)~(iii)を繰り返すことで MSPE を最小化する  $b$  を探索する。

なお、各地点のローカルモデル(1)式をまとめると下式のように表現できる。

$$\mathbf{y} = \boldsymbol{\mu}_{(x)} + \boldsymbol{\varepsilon} \quad E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I} \quad (4)$$

$$\boldsymbol{\mu}_{(x)} = \left[ \sum_p x_{s,p} \beta_{s,p} \right]_s$$

ここで  $\boldsymbol{\varepsilon}$  は攪乱項ベクトル、 $\mathbf{0}$  はゼロベクトル、 $\mathbf{I}$  は単位行列である。GWR についてより詳しくは Fotheringham et al. (2002) を参照されたい。

### 4 ダウンスケール手法のレビュー

ダウンスケールは、地理学では集計単位の変換を表す面補間 (areal interpolation) (例えば、貞広, 2000) の一種に位置づけられる。一方、ダウンスケール法/面補間法は、空間統計学の分野では change of support problem のための一手法に位置づけられ、現在までに膨大な知見が蓄積されている (Gotway and Young, 2002)。

いま、 $\mathbf{y} = [y_1, \dots, y_d, \dots, y_D]'$  を非集計単位 (例：市区町村) 毎の未知のストック量、 $\bar{\mathbf{y}} = [\bar{y}_1, \dots, \bar{y}_a, \dots, \bar{y}_A]'$  を集計単位 (例：都道府県) 毎の既知のストック量とする。面補間とは、(5)式で定義される集計メカニズムを考慮した  $\mathbf{y}$  の補間を意味する。

$$\bar{\mathbf{y}} = \mathbf{A} \mathbf{y} \quad (5)$$

ここで  $\mathbf{A}$  は  $A \times D$  の集計行列であり、その  $(a, d)$  要素は、集計単位  $a$  が非集計単位  $d$  を包含する場合に 1、そうでない場合に 0 とする。(5)式は、単に市区町村別のストック推計値 (予測値) を都道府県に集計した値は、属する都道府県別のストック量 (観測値) に一致しなくてはならないことを意味する。上式で定義されるこの原則は体積保存則 (pynophylactic property) (Goodchild and Lam, 1980) と呼ばれ、面補間法が満たすべき最も基本的な性質のひとつとされている。

体積保存則を満足する基本的な面補間式の一つに(6)式がある。

$$y_{d(a)} = \frac{w_{d(a)}}{\sum_{d(a)} w_{d(a)}} \bar{y}_a \quad (6)$$

$d(a)$  は集計単位  $a$  に包含される非集計単位の添え字である。 $w_{d(a)}$  は各非集計単位  $d(a)$  の重みを表し、(6)式はこれ

に応じて  $\bar{y}_a$  を比例配分する。  $w_{da}$  を面積で与えた場合、(6)式は実務でも幅広く用いられる面積按分法の予測式となる。一方、  $w_{da}$  を何らかの補助データで与える手法は *dasymetric* 法と呼ばれる（例えば都道府県から市区町村への住宅ストックのダウンスケールの場合であれば、  $w_{da}$  は市区町村別居住人口で与えることができる）。 *dasymetric* 法の精度の良さは地理学の多くの既往研究で報告されている（例えば Fisher and Langford, 1995）。

*dasymetric* 法の高度化や応用は近年でも活発に議論されている<sup>3</sup>。例えば道路ネットワークを活用した *dasymetric* 法（Xue, 1995）や専門家の知識を活用した Mennis and Hultgren (2006)、ポイントデータを考慮した Zhang and Qui (2011) はその一例である。これらの既往研究は、適切な補助データで重み  $w_d$  ( $w_{da}$ ) が与えた場合の *dasymetric* 法の精度の良さを指摘している。一方で、 *dasymetric* 法をはじめとした地理学の手法は、統計学的見地からは、MSPE を最小化しない（即ち、予測量が、予測誤差の最小化から得られる理論上の最適解と必ずしも一致しない）という短所がある。

空間統計学では MSPE の最小化に基づいた手法が活発に議論されてきた。例えば、空間的従属性（空間的に近接した観測値が類似した傾向を示すという性質）を考慮したポイントデータの補間法である *kriging* (e.g., Cressie, 1993) をダウンスケールに応用した *area-to-point kriging* (ATP *kriging* : Kyriakidis, 2004) はその一例である。ATP *kriging* の有用性は数多くの既往研究で指摘されている。

以上を踏まえると、複数の補助データを（各地域の重みとしてではなく）説明変数として考慮し、かつ MSPE を最小化する *area-to-point kriging with external drift* (ATP KED : e.g., Yoo and Kyriakidis, 2009) は有効なダウンスケール法の一つといえよう。しかしながら、この手法もまた説明変数の効果は対象地域で様という強い仮定を持つ。本研究では日本全国という広域を対象とするため、説明変数の効果は地域毎に異なると考えることが妥当といえよう。また、ATP KED、及び殆どの空間統計手法には、各空間単位の重み  $w_d$  ( $w_{da}$ ) を何らかの一変数で先見的に与えなくてはならないという制約がある。

そこで本研究では次の条件を満たす手法を構築する：(A) MSPE を最小化する；(B) 体積保存則を満足する；(C) 複数の重み変数／説明変数の影響を、それらの空間的な異質性も考慮しながら、推定する。ここで我々が村上ら (2013) で提案したダウンスケール法は、ATP KED と同様、重みを一変数で予め与える必要があり、(C) を満たして

いない。重みを柔軟に与えることの重要性が *dasymetric* 法の既往研究で数多く指摘されていることから、複数の変数を考慮して重みの影響を推定することは重要である。

なお、以降では重み変数  $w_d$  は地域毎の重みを表す *extensive (volume / count) variables* として、説明変数は地域毎の属性を表す *intensive (density / ratio) variables* として、それぞれ定義した上で議論を進める<sup>4</sup>。

## 5 GWRに基づくダウンスケール手法

### (1) モデル

いま、 $\mathbf{W}$  を重み変数  $\{w_1, \dots, w_d, \dots, w_D\}$  を要素にもつ対角行列とすると、非集計レベルの密度変数ベクトルは  $\mathbf{W}^{-1}\mathbf{y}$  と書ける（例えば  $\mathbf{y}$  の要素を人口、 $\mathbf{W}$  の要素が面積とすると、 $\mathbf{W}^{-1}\mathbf{y}$  の要素は人口密度）。

GWR は密度変数のモデル化に用いられることが多いことから、ここでも、まずは密度変数  $\mathbf{W}^{-1}\mathbf{y}$  を以下に示す GWR でモデル化することを考える。

$$\mathbf{W}^{-1}\mathbf{y} = \boldsymbol{\mu}_{(x)} + \mathbf{W}^{-1}\boldsymbol{\varepsilon}$$

$$E[\mathbf{W}^{-1}\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\mathbf{W}^{-1}\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{W}^{-2} \quad (7)$$

$$\boldsymbol{\mu}_{(x)} = \left[ \sum_p x_{d,p} \beta_{d,p} \right]_d \quad (8)$$

$x_{dp}$  と  $\beta_{dp}$  は非集計単位  $d$  の説明変数と回帰係数である。ここでは、攪乱項を重み  $w_{dp}$  で除すことで基準化している。

(7)式は直ちに(9)式に展開できる。

$$\mathbf{y} = \mathbf{W}\boldsymbol{\mu}_{(x)} + \boldsymbol{\varepsilon} \quad E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I} \quad (9)$$

また、集計行列  $\mathbf{A}$  を用いて(9)式の各要素を集計すると、(10)式が得られる。

$$\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)} + \mathbf{A}\boldsymbol{\varepsilon} \quad E[\mathbf{A}\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\mathbf{A}\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{A}\mathbf{A}' \quad (10)$$

体積保存則の条件式  $\mathbf{A}\mathbf{y} = \bar{\mathbf{y}}$  を代入することで、(10)式は(11)式に展開できる。

$$\bar{\mathbf{y}} = \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)} + \mathbf{A}\boldsymbol{\varepsilon} \quad E[\mathbf{A}\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\mathbf{A}\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{A}\mathbf{A}' \quad (11)$$

以上で定義された非集計モデル(9)と集計モデル(11)は次の式でまとめることができる。

<sup>3</sup> 例えば *Geographical Analysis* 誌の 45 巻 3 号(2013)で“*Areal interpolation and dasymetric modeling*”という特集号が組まれ、*dasymetric* 法の様々な拡張が議論されている。

<sup>4</sup> 厳密には、*extensive variable* は観測値が集計単位のサイズに応じて増加する変数を表す。例えば人口はその一例である。一方で、*intensive variable* は観測値が集計単位のサイズに依存しない変数を表す。例えば人口密度はその一例である。

$$\begin{bmatrix} \mathbf{y} \\ \bar{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \mathbf{W}\boldsymbol{\mu}_{(x)} \\ \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{A}\boldsymbol{\varepsilon} \end{bmatrix} \quad \boldsymbol{\mu}_{(wx)} = \begin{bmatrix} \sum_q w_{d,q} \gamma_{d,q} \end{bmatrix}_d \quad (19)$$

$$E \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{A}\boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad \text{Var} \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{A}\boldsymbol{\varepsilon} \end{bmatrix} \sim \sigma^2 \begin{bmatrix} \mathbf{I} & \mathbf{A}' \\ \mathbf{A} & \mathbf{A}\mathbf{A}' \end{bmatrix} \quad (12)$$

(12)式の前での  $\mathbf{y}$  の best linear unbiased predictor (BLUP) は次式で与えられることが知られている。

$$\hat{\mathbf{y}} = \mathbf{W}\boldsymbol{\mu}_{(x)} + \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}(\bar{\mathbf{y}} - \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)}) \quad (13)$$

BLUP である  $\hat{\mathbf{y}}$  は MSPE を最小化する。また(13)式は次のように展開できる。

$$\begin{aligned} \mathbf{A}\hat{\mathbf{y}} &= \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)} + \mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}(\bar{\mathbf{y}} - \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)}) \\ &= \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)} + \bar{\mathbf{y}} - \mathbf{A}\mathbf{W}\boldsymbol{\mu}_{(x)} \\ &= \bar{\mathbf{y}} \end{aligned} \quad (14)$$

(14)式は  $\mathbf{A}\hat{\mathbf{y}} = \bar{\mathbf{y}}$  であること、即ち  $\hat{\mathbf{y}}$  が体積保存則を満足することを意味する。

以上より、(13)式より得られる  $\hat{\mathbf{y}}$  は第4章の末尾で述べた条件(A)、(B)を満足する面補間式である。一方で、(13)式は重み変数を一つしか仮定しておらず、(C)は満たしていない。そこで、非集計レベルモデル(9)を複数の重み変数(添え字:  $q = 1, \dots, Q$ )を考慮するモデル(15)に拡張することを考える。

$$\mathbf{y} = \mathbf{W}_1\boldsymbol{\mu}_{(x)1} + \dots + \mathbf{W}_q\boldsymbol{\mu}_{(x)q} + \dots + \mathbf{W}_Q\boldsymbol{\mu}_{(x)Q} + \boldsymbol{\varepsilon} \quad E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{I} \quad (15)$$

$$\boldsymbol{\mu}_{(x)q} = \begin{bmatrix} \sum_p x_{d,p} \beta_{d,p,q} \end{bmatrix}_d \quad (16)$$

$\mathbf{W}_q$  は非集計単位  $d$  における  $q$  番目の重み変数  $w_{dq}$  を第  $d$  要素に持つ対角行列、 $\beta_{dpq}$  は回帰係数である。ここで  $\mathbf{W}_q\boldsymbol{\mu}_{(x)q}$  の第  $d$  要素は  $\sum_p w_{d,q} x_{d,p} \beta_{d,p,q}$  のように書ける。

そのため、(15)式は(17)式に展開できる。

$$\mathbf{y} = \boldsymbol{\mu}_{(wx)} + \boldsymbol{\varepsilon} \quad E[\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{I} \quad (17)$$

$$\boldsymbol{\mu}_{(wx)} = \begin{bmatrix} \sum_q \sum_p w_{d,q} x_{d,p} \beta_{d,p,q} \end{bmatrix}_d \quad (18)$$

(17)式は  $w_{dq}x_{dp}$  を説明変数とするだけで、複数の重み変数が考慮できることを示している。(18)式は(19)式のようにも書ける。

ここで  $\gamma_{d,q} = \sum_p x_{d,p} \beta_{d,p,q}$  は重み変数  $w_{dq}$  の係数となる。

非空間的な異質性を捉える  $x_{dp}$  と空間的な異質性を捉える  $\beta_{dpq}$  の両方を考慮したこの係数は、複数の重み変数の影響を柔軟に捉える上で重要である。

(17)式を(9)式と同様に展開することで、基本式(20)式が得られる。

$$\begin{bmatrix} \mathbf{y} \\ \bar{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{(wx)} \\ \mathbf{A}\boldsymbol{\mu}_{(wx)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{A}\boldsymbol{\varepsilon} \end{bmatrix} \quad E \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{A}\boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad \text{Var} \begin{bmatrix} \boldsymbol{\varepsilon} \\ \mathbf{A}\boldsymbol{\varepsilon} \end{bmatrix} \sim \sigma^2 \begin{bmatrix} \mathbf{I} & \mathbf{A}' \\ \mathbf{A} & \mathbf{A}\mathbf{A}' \end{bmatrix} \quad (20)$$

また、(20)式より  $\mathbf{y}$  の BLUP(21)式もまた得られる。

$$\hat{\mathbf{y}} = \boldsymbol{\mu}_{(wx)} + \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}(\bar{\mathbf{y}} - \mathbf{A}\boldsymbol{\mu}_{(wx)}) \quad (21)$$

(21)式は、第4章末尾で議論した条件(A)、(B)、及び(C)を満足する面補間式である。

## (2) パラメータ推定

$\mathbf{y}$  が未知のため、非集計モデル (20式内1段目) を用いてパラメータを推定することはできない。そこで、空間統計学の既往研究 (e.g., Gotway and Young, 2007; Nagle et al., 2011) に倣い、(20)式内2段目の集計モデル(22)を用いてパラメータを推定する。

$$\bar{\mathbf{y}} = \mathbf{A}\boldsymbol{\mu}_{(wx)} + \mathbf{A}\boldsymbol{\varepsilon} \quad E[\mathbf{A}\boldsymbol{\varepsilon}] = \mathbf{0} \quad \text{Var}[\mathbf{A}\boldsymbol{\varepsilon}] = \sigma^2\mathbf{A}\mathbf{A}' \quad (22)$$

ここで、 $\mathbf{A}\mathbf{A}'$  が集計単位  $a$  内の非集計単位数  $N_a$  を第  $a$  要素に持つ対角行列となることは容易に確認できる。

(22)式は  $D$  個の局所モデルを集約した式と考えることができる。その第  $d(a)$  モデルは被説明変数が  $\bar{y}_a$ 、説明

変数が  $\bar{w}x_{a,p,q} = \sum_{d(a)} w_{d,q} x_{d,p}$ 、かつ攪乱項が  $1/N_a$  で重

みづけられたモデルである。このことを利用すると、 $\beta_{dpq}$  を要素に持つベクトルを  $\boldsymbol{\beta}_d$  の推定量は、従来のGWRモデルと同様にWLSを適用することで、次式で与えることができる。

$$\hat{\boldsymbol{\beta}}_d = (\bar{\mathbf{X}}_d' \bar{\mathbf{X}}_d)^{-1} \bar{\mathbf{X}}_d' \bar{\mathbf{y}}_d \quad (23)$$

$$\bar{\mathbf{X}}_d = \begin{bmatrix} \frac{\bar{g}(h_{d,\ddot{a}})^{1/2}}{N_{\ddot{a}}} \bar{w}x_{\ddot{a},p,q} \end{bmatrix}_{\ddot{a},(p-1)Q+q} \quad \bar{\mathbf{y}}_d = \begin{bmatrix} \frac{\bar{g}(h_{d,\ddot{a}})^{1/2}}{N_{\ddot{a}}} \bar{y}_{\ddot{a}} \end{bmatrix}_{\ddot{a}}$$

ここで  $\bar{g}(h_{d,\ddot{a}})$  は非集計単位  $d$  と集計単位  $\ddot{a}$  の空間的な近接性を表す。これは、バンド幅パラメータを持つ非集計単位間の距離  $h_{d,\ddot{a}}$  の関数を  $g(h_{d,\ddot{a}})$  とすると、(24)式で与えることができる<sup>5</sup>。

$$\bar{g}(h_{d,\ddot{a}}) = \sum_{\ddot{a}(\ddot{a})} g(h_{d,\ddot{a}}) \quad (25)$$

(25)式のように非集計単位間の近接性の和で集計単位間の近接性を定義した先行研究には例えばYoo and Kyriakidis (2006)がある。

$\hat{\beta}_d$  は次の手順で推定できる：(i) LCVでバンド幅を最適化する；(ii) (i)で得られたバンド幅を(23)式に代入して  $\hat{\beta}_d$  を求める。手順(i)のLCVは集計モデルのMSPEである

$$(\bar{y} - \mathbf{A}\hat{\mu}_{(wx)})'(\bar{y} - \mathbf{A}\hat{\mu}_{(wx)})/A \text{ を最小化する。}$$

$\hat{\beta}_d$  が推定されれば  $\gamma_{d,q} = \sum_p x_{d,p}\beta_{d,p,q}$  は直ちに求まる。また、推定された  $\beta_{d,p,q}$  を(18), (21)式に代入することで予測値  $\hat{y}$  もまた得られる。

ATP krigingに倣い、本章で構築した手法はATP GWRと呼称することとする。

## 6 提案手法の特性

### (1) 長所

提案手法は体積保存則を考慮しながらMSPEを最小化する。このことは空間統計手法にもあてはまる。しかしながら、提案手法はいくつかの面で空間統計手法よりも優れている。

まず、補助データの影響を柔軟にモデル化できる点である。標準的な空間統計手法と異なり、提案手法は重み変数の影響を非集計単位毎に可変としている(19式参照)。重み変数の影響を適切に捉えることの重要性を論じてきた地理学の既往研究(例えばMennis and Hultgren, 2006)を鑑みると、この点は、提案手法の精度の良さを示唆する特性の一つといえよう。

Dasymetric法との類似もまた提案手法の長所の一つである。(6)式に示すように、dasymetric法は  $y_a$  が重み変数  $w_a$  に比例することを仮定する。一方で、(17), (19)式に示すように、提案手法は  $y_a$  が  $\sum_q w_{a,q}y_{a,q}$  に比例することを仮定する。即ち、提案手法はdasymetric法の拡張と考えることが

<sup>5</sup>第  $d$  要素が  $g(h_{d,\ddot{a}})$  で与えられた  $D \times 1$  のベクトルを  $\mathbf{g}_d$  とすると、 $\bar{g}(h_{d,\ddot{a}})$  は  $\bar{\mathbf{g}}_d = \mathbf{A}\mathbf{g}_d$  の第  $a$  要素となる。

できる。dasymetric法の精度の良さは地理学の多くの既往研究で明らかにされていることから、dasymetric法との類似もまた提案手法の精度の良さを示すものと言えよう。なお、量を表す変数  $y_a$  を補間するdasymetric法と、密度を表す変数 (e.g.,  $y_a/w_{a,q}$ ) を補間する標準的な空間統計手法との関係は、不明瞭である。

提案手法のもう一つの長所に、分析結果が集計単位の取り方に依存する問題であるModifiable Areal Unit Problem (MAUP: Openshaw and Taylor, 1979)への頑健性がある。MAUPが空間統計手法に及ぼす影響は必ずしも十分には議論されていない。一方で、GWRのMAUPへの頑健性はFotheringham et al (2002)などで示唆されている。また、Murakami and Tsutsumi (2013)によれば、集計メカニズムを考慮することで、GWRはMAUPにより頑健になる。提案モデルもまた集計メカニズムを考慮したGWRの一種であり、MAUPの有効な対処法となる可能性がある。

以降では、提案手法の持ついくつかの短所と、それらへの対処法を議論する。

### (2) 多重共線性

GWRは多重共線性の影響を強く受けることが知られており、このことが面補間の結果を不安定にする可能性がある。本研究では、Wheeler (2007)に倣い、リッジパラメータを導入することでこの問題に対処する。

バンド幅を既知とすると、 $\beta_d$  のリッジ推定量は(26)式となる。

$$\hat{\beta}_d^{ridge} = (\bar{\mathbf{X}}_d' \bar{\mathbf{X}}_d + \lambda_d \mathbf{I})^{-1} \bar{\mathbf{X}}_d' \bar{\mathbf{y}}_d \quad (26)$$

$\lambda_d$  は第  $d$  ローカルモデルのリッジパラメータ、 $\mathbf{I}$  は単位行列である。(26)式は多重共線性がある場合に不安定化する  $(\bar{\mathbf{X}}_d' \bar{\mathbf{X}}_d)^{-1}$  をより安定的な  $(\bar{\mathbf{X}}_d' \bar{\mathbf{X}}_d + \lambda_d \mathbf{I})^{-1}$  に置き換えた推定量である。(26)式は  $\lambda_d$  を導入した Weighted Least Squares (WLS)推定量である。MSPE最小化に基づくリッジパラメータの解析解 (Hoerl et al., 1975) を Generalized Least Squares (WLS 含む) 推定量に拡張した Firinguetti (1989)によれば、 $\lambda_d$  の解析解は次のように与えることができる。

$$\hat{\lambda}_d = \frac{A \hat{\sigma}_d^2}{\hat{\beta}_d' \hat{\beta}_d} \quad (27)$$

ここで  $\hat{\sigma}_d^2$  は次式で与えられる。

$$\hat{\sigma}_d^2 = \frac{(\bar{\mathbf{y}}_d - \bar{\mathbf{X}}_d \hat{\beta}_d)' (\bar{\mathbf{y}}_d - \bar{\mathbf{X}}_d \hat{\beta}_d)}{A - PQ} \quad (28)$$

(27)式より与えられる  $\hat{\lambda}_d$  は説明変数のスケールに依

存するため（例えば説明変数 1 の分散が説明変数 2 よりも大きい場合、 $\lambda_d$  は説明変数 1 により強く依存する）、ここでは説明変数を基準化する次の手順で  $\hat{\beta}_d^{ridge}$  を推定する (Wheeler, 2007 参照) : (i)  $\bar{y}_d$  を中心化 (平均 0) し、 $\bar{X}_d$  内の各変数を基準化 (平均 0, 分散 1) する ; (ii) (i) で中心化/基準化した各変数を(26), (27), (28)式に代入することで、基準化されたリッジ推定量  $\hat{\beta}_d^{ridge(scale)}$  を求める ; (iii)  $\hat{\beta}_d^{ridge(scale)}$  の各要素に  $\bar{X}_d$  内の各変数の分散をかけてリスケーリングすることで、 $\hat{\beta}_d^{ridge}$  を求める。

一方、バンド幅が既知の場合は次の手順で  $\hat{\beta}_d^{ridge}$  を推定する : (i) 5章2節で議論したLCVで最適なバンド幅を探索する ; (ii) 上述の方法で  $\hat{\beta}_d^{ridge}$  を推定する。バンド幅とリッジパラメータを別々に推定する手順は、両パラメータを同時最適化するものではない。しかしながら、Wheeler (2007)によれば、この手順より得られる推定値は、同時最適化で得られる推定値と殆ど変わらないはずである。

リッジパラメータありのGWRはgeographically weighted ridge regression (GWRR : Wheeler, 2007) と呼ばれることから、リッジパラメータ付きのATP GWRはATP GWRRと呼称する。

### (3) 負の補間値

面補間の対象は、多くの場合非負の変数 (例 : 人口) であり、予測式(21)への非負制約の導入が必要となる場合も多い。

ここで、予測式(21)が二次計画問題(29), (30)の解と同一であることは容易に示すことができる。

$$\underset{\mathbf{y}}{\operatorname{argmin}} (\mathbf{y} - \boldsymbol{\mu}_{(wx)})'(\mathbf{y} - \boldsymbol{\mu}_{(wx)}) \quad (29)$$

$$\text{s.t. } \mathbf{A}\mathbf{y} = \bar{\mathbf{y}} \quad (30)$$

従って、(21)式による予測を(29), (30), (31)式で定義される二次計画問題の計算に置き換えることで、非負制約下での予測値が算出できる。

$$\mathbf{y} > \mathbf{0} \quad (31)$$

幸い二次計画問題への非負制約(31)の導入は容易である (Simon, 2010参照)。

### (4) 計算量

提案手法のLCVは、非集計レベル回帰係数ベクトル  $\beta_d$

を繰り返し推定しなくてはならず計算負荷が大きい。そこで、 $\beta_d$ を集計レベルの回帰係数ベクトル  $\beta_a$ で置き換えることで、このLCVを近似することを考える。 $\beta_d$ の推定量は、 $\hat{\beta}_d$ 内の  $\bar{g}(h_{d,\ddot{a}})$  を集計単位間の空間的近接性を記述する変数  $\bar{g}(h_{a,\ddot{a}})$  に置き換えることで、次式で定義される。

$$\hat{\beta}_a = (\bar{X}'_a \bar{X}_a)^{-1} \bar{X}'_a \bar{y}_a \quad (32)$$

$$\bar{X}_a = \left[ \begin{array}{c} \bar{g}(h_{a,\ddot{a}})^{1/2} \\ N_{\ddot{a}} \end{array} \right]_{\ddot{a}, (p-1)Q+1} \quad \bar{y}_a = \left[ \begin{array}{c} \bar{g}(h_{a,\ddot{a}})^{1/2} \bar{y}_{\ddot{a}} \\ N_{\ddot{a}} \end{array} \right]_{\ddot{a}}$$

$\bar{g}(h_{a,\ddot{a}})$  は集計単位  $a$  内の各非集計単位と集計単位  $\ddot{a}$  内の各非集計単位との空間的近接性を足し合わせた(33)式で定義される。

$$\bar{g}(h_{a,\ddot{a}}) = \sum_{d(a)} \sum_{\ddot{a}} g(h_{d,\ddot{a}}) \quad (33)$$

(32) を用いることで、前述のLCVを  $(\bar{\mathbf{y}} - \bar{X}\hat{\beta}_a)'(\bar{\mathbf{y}} - \bar{X}\hat{\beta}_a)/A$  を最小化するLCVに置き換えることができる。ここで  $\bar{X}$  は集計レベルの説明変数行列である。この置き換えによりLCVの計算負荷は約  $A/D$  になる。なお両LCVがほぼ同じ結果を与えることは確認済みである。

また、非負制約ありの二次計画問題もまた大きな計算負荷を抱える。幸い、我々の二次計画問題(29), (30), (31)は、異なる集計単位内の  $y_d$  が独立であることを仮定しており、下式で定義される集計単位毎の二次計画問題をまとめたものに等しい。

$$\underset{\mathbf{y}_{d(a)}}{\operatorname{argmin}} (\mathbf{y}_{d(a)} - \boldsymbol{\mu}_{(wx)d(a)})'(\mathbf{y}_{d(a)} - \boldsymbol{\mu}_{(wx)d(a)}) \quad (34)$$

$$\text{s.t. } \sum_{d(a)} \mathbf{y}_{d(a)} = \bar{\mathbf{y}}_a \quad (35)$$

$$\mathbf{y}_{d(a)} > \mathbf{0} \quad (36)$$

ここで、 $\mathbf{y}_{d(a)}$  と  $\boldsymbol{\mu}_{(wx)d(a)}$  は、 $\mathbf{y}$  と  $\boldsymbol{\mu}_{(wx)}$  の集計単位  $a$  内の各非集計単位  $d(a)$  についての要素からなるベクトルである。上式で定義される  $A$  個の二次計画問題を解くことで、(29), (30), (31)式を用いた場合と同一の予測値を、高速に算出できる。

Value (km<sup>2</sup>/100km<sup>2</sup>)

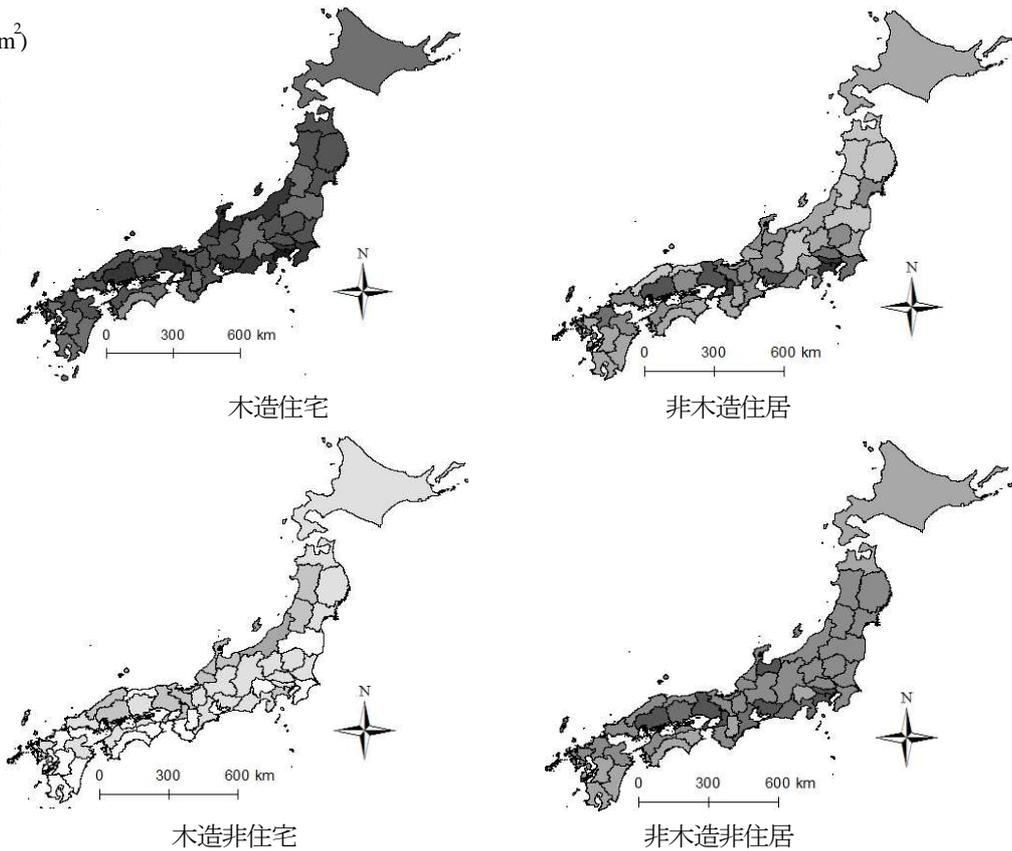
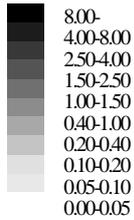


図1：都道府県別建築物ストックの実測値

## 7 実証分析

### (1) 概要

本章では、2005年度までに竣工された46都道府県別（沖縄県は、他県から非常に離れているため、バンド幅パラメータの推定の不安定さから、今回対象から除くこととした）の木造/非木造住宅ストック（図1；出典：建築物ストック統計）を、2005年時点の市区町村（標本数1,803）にダウンスケールし、その予測精度を比較する。

精度の検証には市区町村毎の住宅ストックデータが必要となる。住宅土地統計調査の市区町村別住宅ストックデータ（2章参照）は全国で入手できるものの、おおむね2割程度をしめる共同住宅の共有部分が含まれていないため、今回は宮城（2009）、堤ら（2012）において申請で取得した2005年の東京都市圏の固定資産課税台帳を用いるものとする。無論、このデータも検証データとしては限界がある。第一に、地域が限定されている点である。この点については、密な東京都心部から、多摩地域のような比較的疎な地域も含まれており、重要な傾向はつかめると判断した。第二は、建築物ストック統計公開のデータには公共の公務員住宅を含む一方で固定資産課税台帳はこれを含まないため、市区町村によっては若干の齟齬が生じる可能性があるという点である。この点については、固定資産課税台帳より算出される総住宅ストック

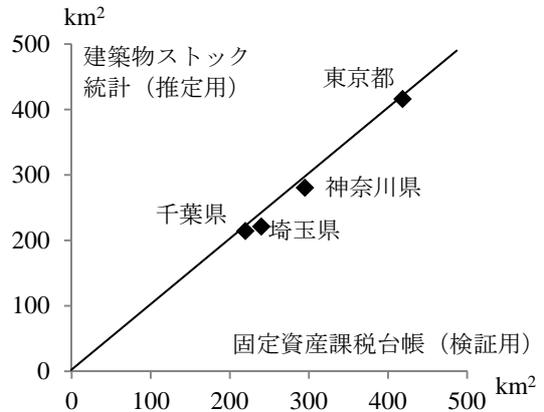


図2：推定用データと検証用データの比較

を県レベルに集計すると、建築物ストック統計におおむね一致し（図2）、誤差率は最大で7.1%程度にすぎないため、その影響は十分に小さいと判断した。

比較する手法は、ATP GWR, ATP GWRR, dasymetric法、及びATP KEDとした。なお、標準的な面補間法の一つである面積按分法については、精度の低さがかねてから指摘されているため、ここでは比較対象としていない。

各手法で用いる補助データは表2に要約した。この表に示すとおり、常住人口に応じた比例配分を行うDasy\_Pと建物用地面積に応じた比例配分を行うDasy\_Bの2通りのdasymetric法を適用する。空間的従属性をモデル化するためにATP KEDで用いる距離減衰関数は指数関数

表1：手法の概要

手法	重み ( $w_{dp}$ )	説明変数 ( $x_{dp}$ )	空間的従属性	空間的異質性
DM_P	人口	N.A.	×	
DM_B	建物用地面積			
ATP KED	面積 <sup>1</sup>	定数項 人口密度 <sup>2</sup> 建物用地面積の割合 <sup>2</sup> 面積あたり鉄道数 道路密度 政令指定都市までの距離 (対数値)	○	×
ATP GWR	人口 建物用地面積	定数項 面積あたり鉄道数 道路密度 政令指定都市までの距離 (対数値)	×	○

<sup>1</sup>ほぼ全ての ATPkriging の既往研究が各非集計単位の重みを面積に応じて与えている

<sup>2</sup> ATP KED は  $y_d$  自体ではなく密度変数を補間するため  $y_d/w_{dq}$ 、人口と建物用地面積ではなく、それらを面積で除した人口密度と建物用地面積の割合を用いる

(Cressie, 1993参照)、空間的異質性をモデル化するために ATP GWR, ATP GWRR で用いる距離減衰関数はガウス関数とする。市区町村間の距離は各市区町村の重心点間の直線距離で与える。なお、VIF 診断により

ATP GWR と ATP GWRR については、非負制約を導入したうえで、6章4節の高速計算アルゴリズムによりダウンスケールを行う。また ATP KED については非負制約下でダウンスケールを行う。

## (2) パラメータ推定結果

ATP GWR より推定されたバンド幅は 463km (木造住宅)、401km (非木造住宅) であり<sup>6</sup>、木造/非木造住宅ストックを説明する係数  $\beta_{dp}$  (あるいは  $w_{dq}$ ) は、空間上で比較的緩やかに変化しているとの示唆を得た。なお、重み変数  $w_{dq}$  を適切に考慮することの重要性を指摘した dasymetric 法の既往研究を踏まえ、以降では  $w_{dq}$  の係数  $\gamma_{dq}$  の推計結果 (木造住宅) について詳細にみていくこととする。

ATP GWR と ATP GWR より推定された  $\gamma_{dq}$  を図3にプロットした。この図によると、単位建物用地面積あたりの木造住宅ストック量は、主要都市周辺で比較的少ない。これは、都市部の建物用地において非木造住宅の割合が少ないためである可能性がある。また、西日本の木造住宅ストックは人口でより良く説明でき、東日本の木造ストックは建物用地面積でより良く説明できるとの結果も得られている。

ただし、以上の結果は多重共線性の影響を受けている可能性がある。実際に、人口と建物用地面積の係数間の相関係数は -0.87 (ATP GWR) と -0.73 (ATP GWRR) であり、ATP GWRR を用いることによる一定の改善が

確認できたものの、その値は依然として (負に) 大きい。リッジパラメータを導入したにもかかわらずパラメータ間に強い相関がみられたのは、重み変数である人口と建物用地面積の間に強い相関 (相関係数: 0.88) が存在し、両変数の効果の識別が困難であったためと考えられる。強い相関を持つ複数の重み変数の影響を、多重共線性に配慮しながら適切に捉える面補間法の構築は今後の重要な課題といえよう。

ATP KED のパラメータ推計結果 (木造住宅) は表2に示すとおりである。なお、ATP KED の説明変数間に深刻な多重共線性が生じていないことは VIF (variance inflation factor) 診断より確認している。この表に示すとおり、人口密度と建物用地面積の割合が 1%水準で正に有意となった。これらの結果は直感に整合する。一方で面積あたり鉄道駅数と道路密度は負に有意となったが、これは非都市部における木造住宅の比率の高さを反映した結果の可能性もある。表中のレンジパラメータは木造住宅ストックの持つ空間的従属性が比較的局所的であることを示している。

## (3) 面補間結果

ATP GWR と ATP GWRR の両予測値の相関係数は木造/非木造の両ケースとも 0.9997 であり、両手法がほぼ同じ予測結果を与えることが明らかとなった。そのため、以降では ATP GWR の結果についてのみ議論する。

按分精度の比較結果は、用いる指標によって大きく変化しうることが知られている。そのため、ここでは root mean squared error (RMSE)、mean absolute error (MAE)、root mean squared percentage error (RMSPE)、mean absolute percentage error (MAPE) の各々で精度を評価し、その結果を考察する。

<sup>6</sup> ATP GWRR のバンド幅はリッジパラメータ推定前に推定するため、ATP GWR と ATP GWRR のバンド幅は同じ。

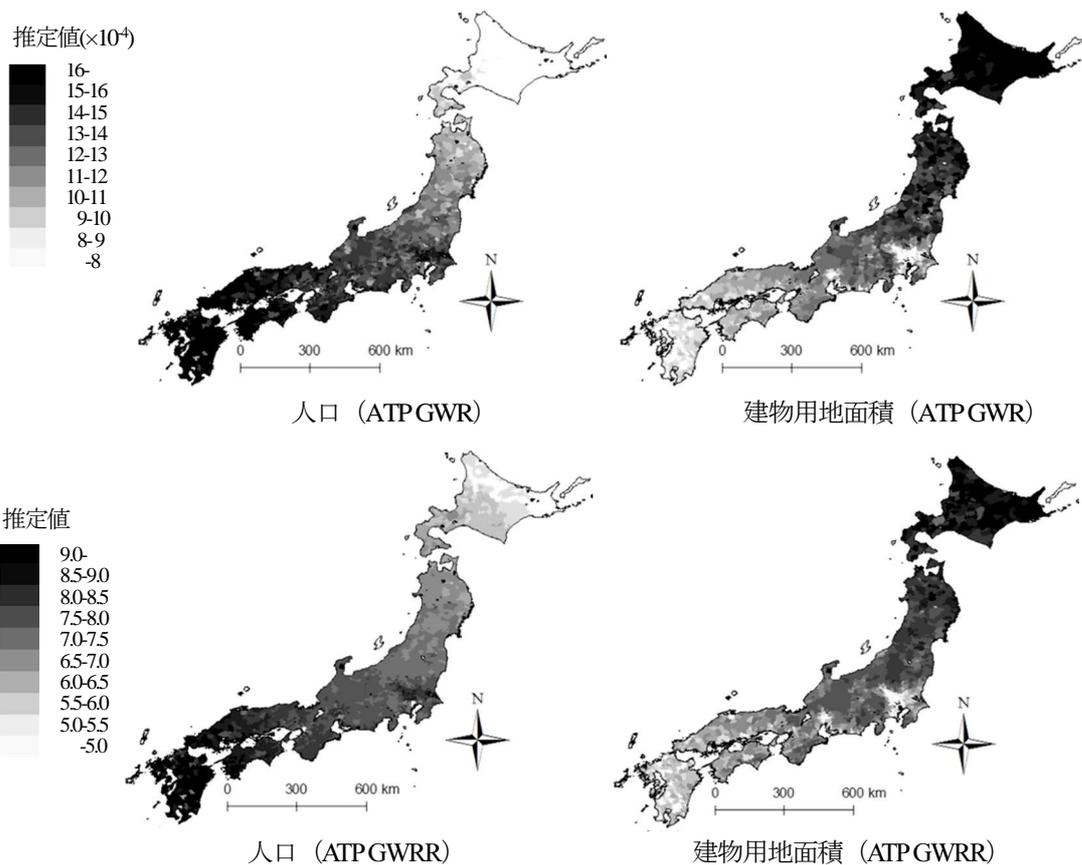


図3：木造住宅ストックに対する重み変数の係数 $\gamma_{da}$ の推定値

$$\text{RMSE} = \sqrt{\frac{1}{D} \sum_d (\hat{y}_d - y_d)^2}, \quad (37)$$

$$\text{MAE} = \frac{1}{D} \sum_i |\hat{y}_d - y_d|, \quad (38)$$

$$\text{RMSPE} = \sqrt{\frac{1}{D} \sum_d \left( \frac{\hat{y}_d - y_d}{y_d} \right)^2}, \quad (39)$$

$$\text{MAPE} = \frac{1}{D} \sum_d \left| \frac{\hat{y}_d - y_d}{y_d} \right|. \quad (40)$$

RMSEは、MAEよりも大きな誤差に敏感な指標である。またRMSPEはRMSEを、MAPEはMAEを、それぞれ基準化した指標であり、予測誤差と実測値の比率が大きければ、大きな値をとる。

各モデルの精度の評価結果を表3（木造住宅）、表4（非木造住宅）に示す。この図より、木造・非木造のいずれについても、RMSEで見るとATP KEDが最良、MAE、RMSPE、MAEで見るとATP GWRが最良という結果が得られた。提案手法（及びATP KED）の精度が、精度の良さが指摘されてきたdasymetric法を大きく上回っている点は興味深い。また、ATP GWRのRMSPEとMAEはATP

表2：ATP KEDのパラメータ推定結果

変数	推定値	t値	
定数項	$9.96 \times 10^2$	1.24	
人口密度	$9.46 \times 10^4$	16.1	1) <sup>***</sup>
建物用地面積の割合	12.7	10.0	***
面積あたり鉄道数	-4.39	-4.67	***
道路密度	$-7.12 \times 10^2$	-4.11	***
政令指定都市距離	$3.78 \times 10^2$	$9.75 \times 10^1$	
レンジ(km)	156		
パーシャル・シル <sup>2)</sup>	$5.66 \times 10^2$		

1)<sup>\*\*\*</sup>は有意水準1%で有意であることを意味する

2) レンジは空間的従属性の及ぶ範囲、パーシャル・シルは空間的従属成分の分散を表す。なお、ATP krigingは空間的従属成分の分散とそれ以外の分散（ナゲット）を識別できないため（Goovaerts, 2008）、Goovaerts (2008)に倣い後者は0としている。

KEDよりも大幅に小さく、ATP GWRを用いることで、誤差率（percentage error） $(\hat{y}_d - y_d)/y_d$ の小さな予測値が得られることを確認した。

図4は木造住宅ストックの推計におけるDasy\_P, Dasy\_B, ATP GWRの各誤差率のプロットである。この図より、Dasy\_Pには郊外部のストックを過大推計する傾向があり、反対にDasy\_Bには都心部のストックを過大推計する傾向があることがわかる。確定論的手法である

表3：精度比較結果（木造住居）

	Dasy_P	Dasy_B	ATP KED	ATPGWR
RMSE	1.55×10 <sup>6</sup>	8.39×10 <sup>5</sup>	8.98×10 <sup>5</sup>	9.16×10 <sup>5</sup>
MAE	6.89×10 <sup>5</sup>	5.56×10 <sup>5</sup>	4.88×10 <sup>5</sup>	4.22×10 <sup>5</sup>
RMSPE	3.91×10 <sup>-1</sup>	5.11×10 <sup>-1</sup>	3.89×10 <sup>-1</sup>	3.17×10 <sup>-1</sup>
MAE	2.99×10 <sup>-1</sup>	3.41×10 <sup>-1</sup>	2.62×10 <sup>-1</sup>	1.98×10 <sup>-1</sup>

灰色枠：最良

表4：精度評価結果（非木造住居）

	Dasy_P	Dasy_B	ATP KED	ATPGWR
RMSE	8.26×10 <sup>5</sup>	1.91×10 <sup>6</sup>	4.64×10 <sup>5</sup>	4.69×10 <sup>5</sup>
MAE	4.34×10 <sup>5</sup>	9.75×10 <sup>5</sup>	2.93×10 <sup>5</sup>	2.69×10 <sup>5</sup>
RMSPE	1.80	6.63	1.68	9.97×10 <sup>-1</sup>
MAE	9.56×10 <sup>-1</sup>	2.68	7.67×10 <sup>-1</sup>	5.23×10 <sup>-1</sup>

灰色枠：最良

表5：残差のモランI統計量

	Dasy_P	Dasy_B	ATP KED	ATPGWR
木造	0.203 ***	0.259 ***	0.072 **	0.084 **
非木造	0.474 ***	0.438 ***	0.143 **	0.156 **

\*\*\*は1%水準で\*\*は5%水準でそれぞれ有意であることを表す

両アプローチはこういった過大過小推計をパラメータ等で調整することはできない。一方、提案手法は、各重み変数の影響をパラメータ $\gamma_{dq}$ で調整した結果、対象地域全域のストックを概ね小さい誤差率で予測していることが確認できる。

また、残差 $\hat{y}_d - y_d$ の空間的従属性（近接する市区町村の残差が類似した値を示す傾向）をモランI統計量（e.g., Griffith, 2003）で検定した（表5）。ここで、モランI統計量が有意であることは、モデルで説明されない空間的従属成分が残差に残存していることを意味する。表5より、残差の空間的従属性は全てのモデルで有意となったものの、Dasy\_BやDasy\_Pに比べてATPGWRとATP KEDのモランI統計量は小さく、両統計手法を用いることによる一定の改善がみられた。

### (3) 市区町村別・タイプ別ストックの推計

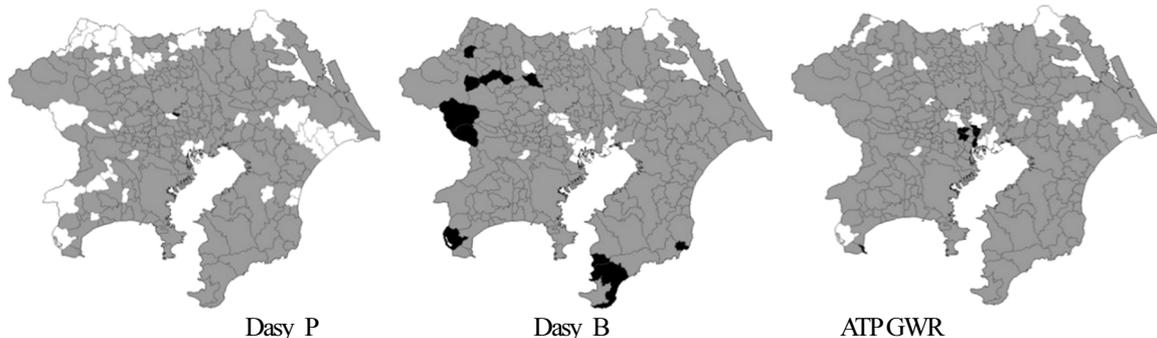


図4：誤差率の空間分布（黒：誤差率>0.5；白：誤差率<-0.5）

本節では、ATPGWRを用いてタイプ別（住宅/非住宅、木造/非木造）・市区町村別の建築物ストック（2007年）を推計した。住宅ストック推計にあたっての各市区町村の重み $w_{dq}$ は人口と建物用地面積、非住宅ストック推計での重みは事業所数と建物用地面積である。

推計結果を図5に示す。この図より主要都市周辺に建築物ストックが多く、都市部における非木造建築物の集中が木造建築物に比べて顕著であるといった、直感に整合する結果が得られていることがわかる。本結果については、今後モデルを精緻化し、最終的にはweb等で公開していきたい。

### (4) 推計されたストックデータの適用例：

#### 市区町村別の電力需要推計

本節では、推計されたストックデータの適用の一例として、市区町村別の電力需要推計を議論する。電力需要は、市区町村別建築物ストック推定値に日本エネルギー学会(2008)算出の各種原単位を乗じることで算出する。

1月の5時、12時、20時、及び7月の同時刻の1時間当たり電力需要量の推計結果を図6（住宅部門）、7（非住宅部門）に示す。この図より、住宅部門の1月の電力需要は時点間で大きな差異が見られない点、7月の電力需要が夜間にピークを迎える点、非住宅の電力需要が12時に大きくなる点、三大都市圏周辺での電力需要が大きい点、1月よりも7月の需要が大きい点などが確認できる。

## 8. まとめ

本研究ではコンパクトシティ政策や気候変動適応策の検討の際に重要となる空間詳細な単位での建築物ストックデータの整備を念頭に、GWRに基づいた新たなトップダウン・アプローチ（面補間法）を構築した。また、その有効性を標準的なダウンスケール手法との比較により確認した。次に、提案手法を用いて全国の市区町村別・タイプ別建築物ストック量を推計し、最後に推計したストックデータを電力需要推計に応用した。

今後、本研究で明らかとした知見を踏まえながら実際

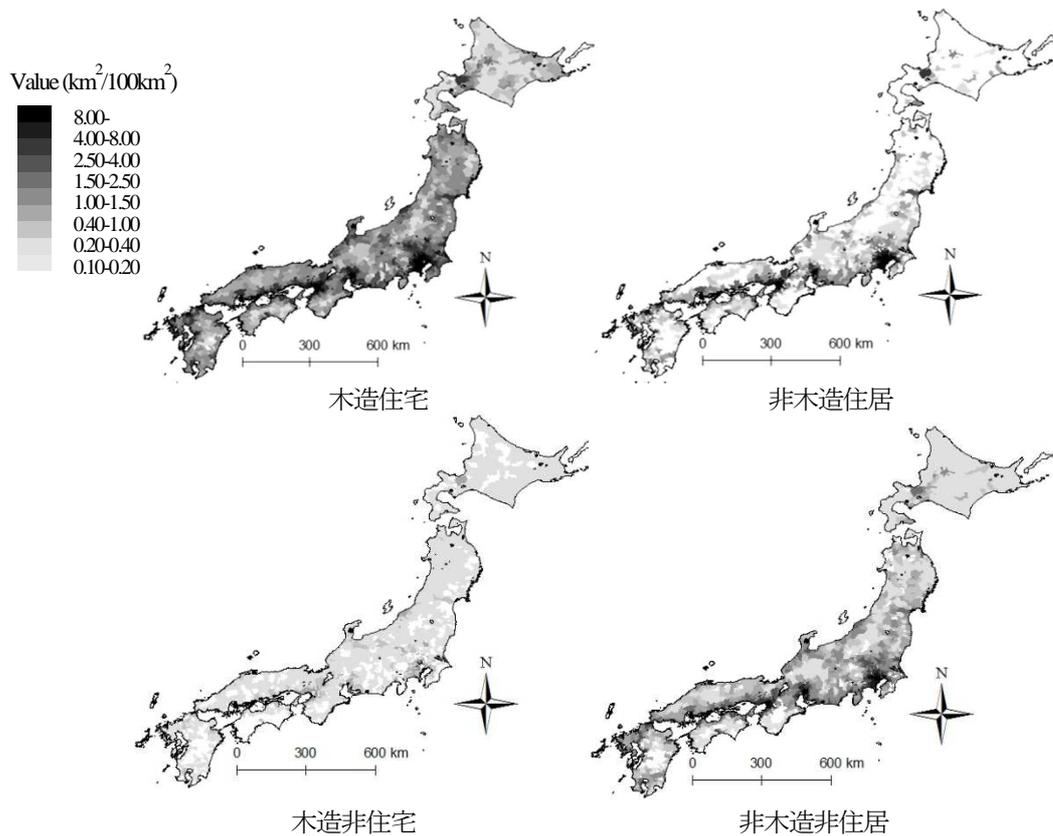


図5：市町村別ストックの推計結果

に市区町村住宅ストックを構築・公開し、さらには3次・4次メッシュ等のより詳細な空間単位毎の建築物ストックの推計に取り組みたい。

なお、本稿で構築したダウンスケール手法は、今後将来の気候・社会シナリオ構築に向けた都市分布の詳細推計に応用する予定である。

**謝辞：**本研究の前段階の原稿について、応用地域学会第27回研究発表大会にて討論者の塚井誠人准教授（広島大学）はじめ多くの方より示唆に富むコメントを頂いた。ここに記して感謝申し上げます。なお、本研究は、（財）日本学術振興会科学研究費補助金（基盤研究（B）23360219）及び環境省環境研究総合推進費 戦略的研究開発プロジェクトS-10の成果の一部である。

#### 参考文献

- 1) Cressie, N. (1993) *Statistics for Spatial Data, Revised Edition*. Wiley, New York.
- 2) Firinguetti, L. (1989) A simulation study of ridge regression estimators with autocorrelated errors. *Communications in Statistics, Simulation and Computation*, **18** (2), 673–702.
- 3) Fisher, P.F. and Langford, M. (1995) Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, **27** (2), 211–224.

- 4) Fotheringham, S., Brunsdon, C. and Charlton, M. (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, New York.
- 5) Flowerdew, R. and Green, M. (1992) Developments in areal interpolation methods and GIS. *Annals of Regional Science*, **26** (1), 67–78.
- 6) Goodchild, M.F. and Lam, N.S.-N. (1980) Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*, **1**, 297–312.
- 7) Goovaerts, P. (2008) Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geology*, **40** (1), 101–128.
- 8) Gotway, C.A. and Young, L.J. (2002) Combining incompatible spatial data. *Journal of the American Statistical Association*, **97** (458), 632–648.
- 9) Gotway, C.A. and Young, L.J. (2007) A geostatistical approach to linking geographically aggregated data from different sources. *Journal of Computational and Graphical Statistics*, **16** (1), 1–21.
- 10) Griffith, D. A. (2003) *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin, Springer-Verlag.
- 11) Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975) Ridge Regression: some simulations. *Communications in Statistics, Theory and Methods*, **4** (2), 105–123.
- 12) Kyriakidis, P.C. (2004) A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, **36** (3), 259–289.

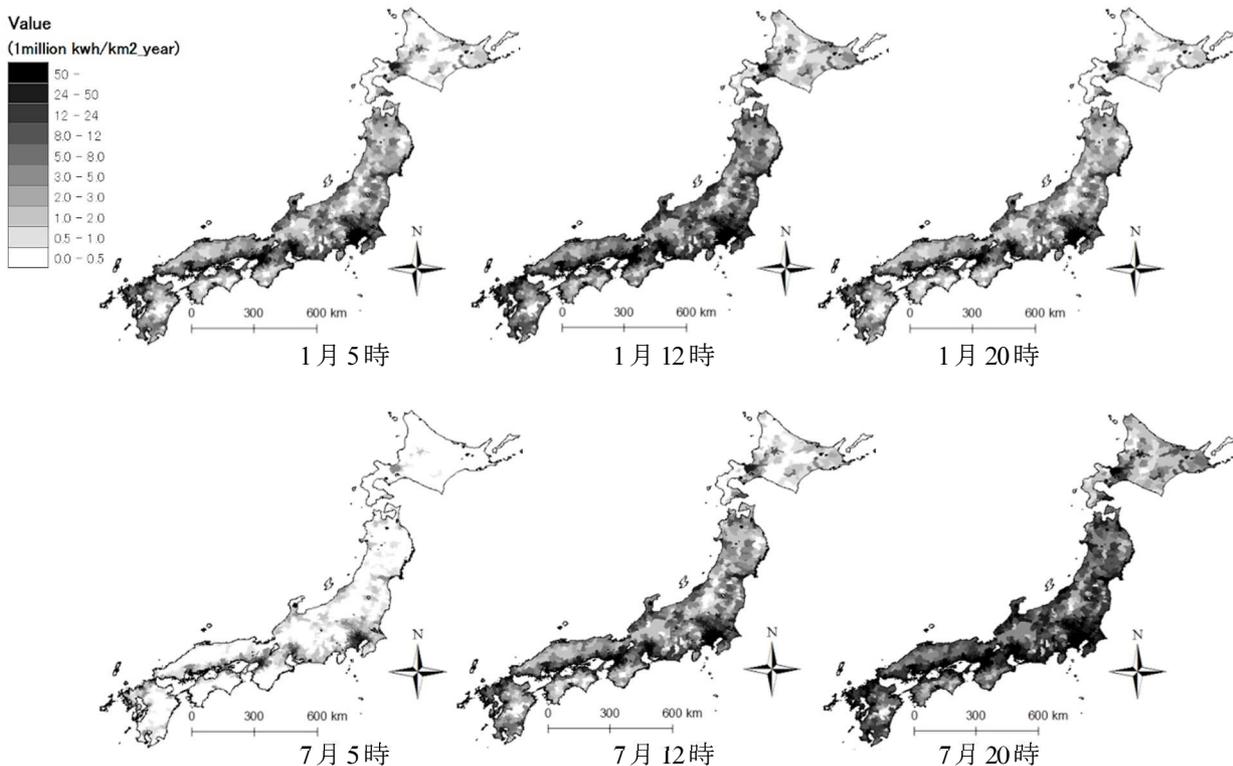


図6：電力需要推計結果（住宅部門）

- 13) Mennis, J. and Hultgren, T. (2006) Intelligent dasymmetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, **33** (3), 179–194.
- 14) Murakami, D. and Tsutsumi, M. (2013) A Geographically Weighed Regression-based Approach for the Change of Support Problem. The VII World Conference of the Spatial Economic Association, Washington, USA.
- 15) Nagle N.N., Sweeney, S.H. and Kyriakidis, P.C. (2012) A geostatistical linear regression model for small area data. *Geographical Analysis*, **43** (1), 38–60.
- 16) Openshaw, S. and Taylor, P. (1979) A million or so correlation coefficients: Three experiments on the modifiable area unit problem. In *Statistical Methods in the Spatial Sciences* (edited by Wrigley, N.), London: Pion, 127–144.
- 17) Simon, D. (2010) Kalman filtering with state constraints: A survey of linear and nonlinear algorithms. *Control Theory & Applications*, **4** (6), 1303–1318.
- 18) Wheeler, D. (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A*, **39** (10), 2464–2481.
- 19) Xie, Y. (1995) The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems*, **19** (4), 287–306.
- 20) Yamagata, Y., Seya, H. and Nakamichi, K. (2013) Creation of future urban environmental scenarios using a geographically explicit land-use model: A case study of Tokyo. *Annals of GIS*, **19** (3), 153–168.
- 21) Yoo, E-H. and Kyriakidis, P.C. (2006) Area-to-point kriging with inequality-type data. *Journal of Geographical Systems*, **8** (4), 357–390.
- 22) Yoo, E-H. and Kyriakidis, P.C. (2009) Area-to-point kriging in spatial hedonic pricing models. *Journal of Geographical Systems*, **11** (4), 381–406.
- 23) Zhang, C. and Qiu, F. (2011) A Point-Based Intelligent Approach to Areal Interpolation. *The Professional Geographer*, **63** (2), 262–276.
- 24) 伊藤徹哉, 寺木彰浩, 阪田知彦 (2005) 建築物の現況データ整備に関する海外事例調査. 『都市計画報告集』, **4** (3), 51–56.
- 25) 小川芳樹, 秋山祐樹, 仙石祐明, 柴崎亮介 (2013) 広域地震災害の被害想定のための日本全土における建物単体データの開発. GISA 学術研究発表 web 大会 (URL: <http://www.slideshare.net/yoshikiogawa1220/gisawebogawa>, アクセス 2014 年 1 月 20 日) .
- 26) 奥村公美, 伊香賀俊治, 川久保俊 (2012) 非住宅建築物の用途別, 都道府県別のストック・フロー床面積データベースの構築. 『日本建築学会技術報告集』, **18** (38), 275–280.
- 27) 国土交通省 (2010) 『建築物ストック統計検討会報告書』 (URL: <http://www.mlit.go.jp/common/000120830.pdf>, アクセス 2014 年 1 月 20 日) .
- 28) 阪田知彦, 吉川徹 (2001) 都市計画基礎調査と固定資産税データ間の建物用途の整合性に関する分析. 『GIS—理論と応用』, **9** (1), 9–18.

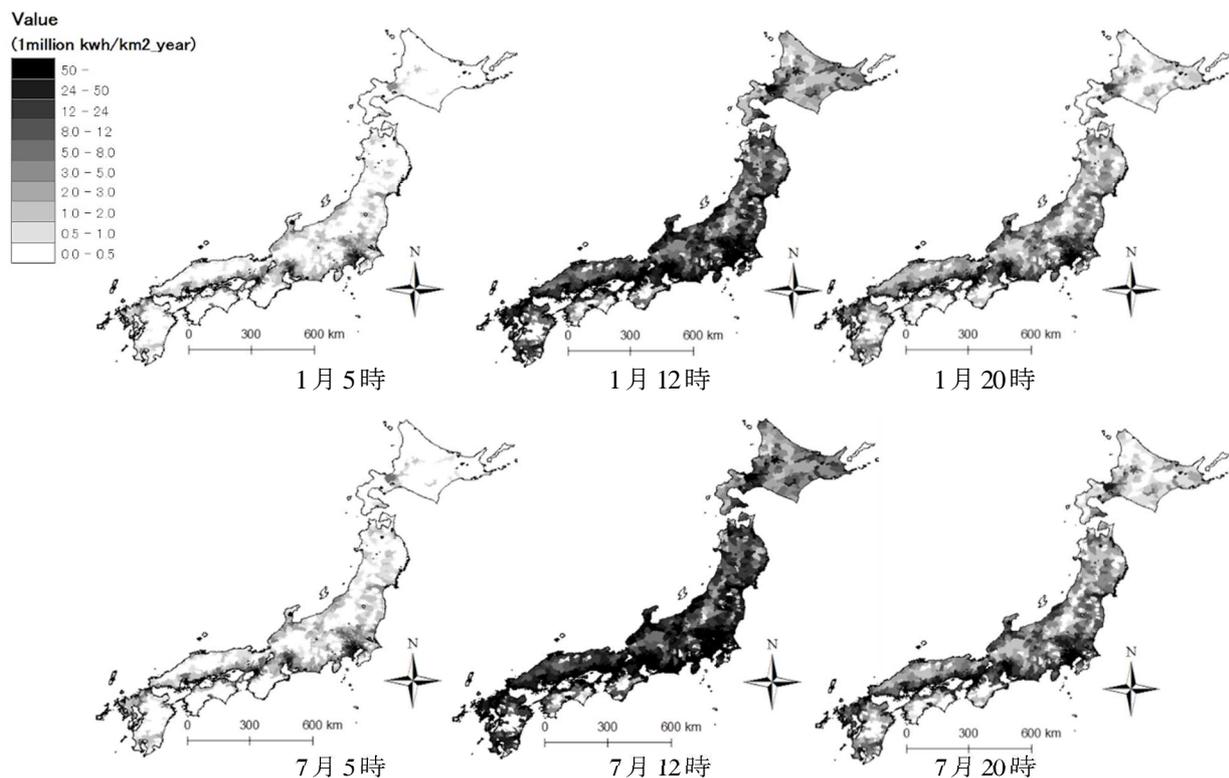


図7：電力需要推計結果（非住宅部門）

- 29) 阪田知彦, 吉川徹 (2002) 東京都 GIS 建物ポリゴンから推定した事業所系延べ床面積の補正について. 『GIS—理論と応用』, **10**(1), 85-94.
- 30) 貞広幸雄 (2000) 空間集計データにおける面補間法の推定精度評価. 『都市計画』, **225**, 75-81.
- 31) 谷川寛樹, 大西暁生, 高平洋祐, 橋本征二, 東修, 白川博章, 井村秀文 (2010) ストック型かつ低炭素型社会へ向けた都市構造物の物質・エネルギー消費の 4D マッピング: 名古屋市の建築物を対象としたケーススタディ. 『日本 LCA 学会誌』, **6**(2), 92-101.
- 32) 堤盛人, 宮城卓也, 山崎清 (2012) 建物市場を考慮した応用都市経済モデルの可能性. 『土木学会論文集 D』, **68** (4), 333-343.
- 33) 日本エネルギー学会 (2008) 『天然ガスコージェネレーション計画・設計マニュアル』. 日本工業出版.
- 34) 平野勇二郎, 外岡豊, 伊香賀俊治, 亀卦川幸浩, 藤沼康実, 下田吉之 (2008) 民生業務部門エネルギー消費原単位に関する各種資料の比較評価. 『日本建築学会環境系論文集』, **73**(633), 1331-1339.
- 35) 宮城卓也 (2009) 開発者行動を考慮した応用都市経済モデルの実用化に関する研究. 筑波大学大学院 システム情報工学研究科修士論文.
- 36) 村上大輔, 瀬谷創, 山形与志樹, 堤盛人 (2013) 空間統計モデルを用いた建築物ストックの詳細推計. 応用地域学会第 27 回研究発表大会.