

全国規模の推定世帯データの高精度化に向けたモデル構築

梶原健人*・瀬戸寿一**・関本義秀***

Model building for improving the accuracy of estimated household data in Japan

Kento KAJIWARA*, Toshikazu SETO** and Yoshihide SEKIMOTO***

The city planning policy of local governments is reviewed due to the population decline and aging of infrastructure facilities. A concrete study of city planning requires the use of open data like the result of National Census, which is conducted by the Ministry of Internal Affairs and Communications. It is indispensable to grasp the long-term outlook of local governments using data for precise city planning, and it is important to grasp the trends of the residents who are the target of public services. Therefore, in this research, household agents are estimated from open data such as census, and a long-term simulation is built focusing on members, household attribute and residence selection. Finally, the accuracy of this model is evaluated.

Keywords: データ精度 (data accuracy), 地域分析 (regional analysis), 自治体 (local government)

1. はじめに

1.1. 研究の背景

近い将来訪れる人口減少や社会基盤老朽化を踏まえ、地方都市を中心に都市のコンパクト化が求められている。国土交通省が主導している立地適性計画には令和2年3月時点で全国1724市町村中522もの自治体が参加し、これからの都市構造のあり方を模索している。こうした都市の転換期において、市民参加型のまちづくりが改めて求められている。江守(2009)によると、1992年の都市計画法改正をきっかけに都市計画への市民参加が重要視されるようになり、現在では行政主導型から市民主導型へと移行してきた。また、産業構造や人口動態、人の流れを可視化したRESAS(日本・内閣府のまち・ひと・しごと創生本部)、各府省が公表する統計データをまとめたe-Stat(独立行政法人統計センター)など、行政データのオープン化が進んでいる。しかしながら、市民参加型の都市計画事業では、地域データに基づいた議論が十分なされていない。特に地方自治体では、市民目線から将来の都市像を把握する上で重要な町丁字や世帯単位での分析、シミュレーションは整備

されていない。

1.2. 先行研究

上記のような中、行政のオープンデータを活用した研究がなされてきた。Hasegawa et al.(2018)はWebベースの都市の将来像シミュレーションシステム“MyCityForecast”を開発した。このシステムでは、500mメッシュ単位で将来人口分布、各地域の公共施設・商業施設へのアクセシビリティや環境指数等の変化を2040年まで5年単位で表示している。海外の都市シミュレーションの代表格としては、Paul Waddell (2002)によって開発されたUrbanSimが挙げられる。土地利用や交通計画に活用されており、シアトル、パリ、チューリッヒなど大都市での政策検討に活用されている。一方国内では、北詰ら(2008)によると、UrbanSimのようなマイクロシミュレーションは政策立案で活用されておらず、都市政策を考案する上で有効なツールの早急な導入が求められている。国内の比較的細かなシミュレーションシステムとして、山際ら(2017)は市民の動向を捉えた世帯を単位にしたマルチエージェントシミュレーション

* 学生会員 東京大学大学院 工学系研究科 (University of Tokyo, School of Engineering)

〒153-8505 東京都目黒区駒場 4-6-1 Ce508 Tel : 080-2014-1517 または E-mail : kento@iis.u-tokyo.ac.jp

** 正会員 東京大学 空間情報科学研究センター (University of Tokyo, Center for Spatial Information Science)

*** 正会員 東京大学 生産技術研究所 (University of Tokyo, Institute of Industrial Science)

を構築した。ここではマクロモデルとミクロモデルの中間的な粒度のメゾスコピックモデルが採用されているが、マイクロシミュレーションと比べると精細さで劣る。また、大都市である金沢市と横浜市を対象に、各自治体が独自に公開しているデータを活用している。

マイクロシミュレーションで活用できる世帯データとして、Akiyama et al.(2013)による 2010 年における市町村毎の世帯推定データが挙げられる。建物データと各種統計調査結果に基づき、住宅の戸別に世帯構成や構成員の年齢、性別を推計したデータを作成し、町丁目別での世帯数と人口は実測値に近い数値を実現した。しかしながら、性別、年齢別の人口分布では実態と大きく乖離する結果となっていた。

1.3. 研究の目的

以上のように、国内の地方自治体を中心に活用できる、国勢調査結果など低廉で汎用的に入手可能なデータを前提とした高精細な都市のマイクロシミュレーションは開発されていない。そこで本研究では、富山県、静岡県全域を対象に世帯単位の推移を捉えたエージェントシミュレーションの構築、高精度化を目的とする。Akiyama et al.(2013)による市町村毎の世帯推定手法を改良し、2010 年のみならず 1980 年でも推定した。市町村単位で性別、年齢別人口を集計し、その精度を検証した。この世帯データを元に世帯推移モデルを構築し、1980 年から 2010 年までのシミュレーション結果と国勢調査を比較した。市町村単位では性別、年齢別人口分布、家族類型別世帯数の精度を検証した。さらに、2010 年から 2040 年までのシミュレーション結果の 500m メッシュ別人口と 500m メッシュ別将来推計人口データ（平成 29 年国土交通省国土政策局）を比較し、妥当性を検討した。

2. 手法

ここでは市町村毎の世帯推計データの作成方法と世帯推移モデルの構築について記す。起点にした 1980 年、2010 年で手法が若干違うため、それぞれの違い、さらには Akiyama et al.(2013)の手法との違い

にも言及する。

2.1. 市町村毎の世帯推計データの作成

2.1.1. 概要

国勢調査の世帯に関する項目、人口動態調査の各種統計データから、世帯毎に世帯人員数、家族類型、構成員の世帯主との関係、年齢、性別を推定していく。2010 年版では 2009 年の建物データ(Z-map TOWN II,ゼンリン株式会社)が入手可能なため、世帯の住居も推定する。基本的に Akiyama et al.(2013)の手法を参考にし、富山県、静岡県全域で作成した。また、1980 年版は世帯推移モデルの精度検証のために作成したものである。本来は 2010 年の国勢調査結果を元に作成する手法のため、統計データの少ない 1980 年国勢調査ではやや粗い推計となっている。

2.1.2. 使用するデータ

図 1 に推定で用いたデータをまとめた。統計データである国勢調査、人口動態は独立行政法人統計センターが運営する e-Stat(<https://www.e-stat.go.jp>)から主にダウンロードしている。1980 年国勢調査の調査区別集計結果のみ、東京大学空間情報科学研究センターが運営している JoRAS(<https://joras.csis.u-tokyo.ac.jp>)から入手している。また、建物データは 2010 年データが作成されていなかったため、最も近い 2009 年版を使用した。テレポイント Pack!は建物の使用用途を推定する際に用いた。

	1980年	2010年
世帯に関するデータ	国勢調査	
人口データ	人口動態調査	
建物データ		Z-map TOWN II (2009年) テレポイントPack! (2009年)
行政区域データ	国勢調査	

図 1：市町村毎の世帯推計での使用データ

2.1.3. 1980 年版の推計フロー

世帯データ推計の流れは図 2 のようになる。はじめに国勢調査の調査区別集計結果の世帯数から世帯を生成し、市町村毎の家族類型別世帯数(B)から各世帯に家族類型を確率配分していく。

その後、家族類型×世帯人数(C)世帯数のクロス

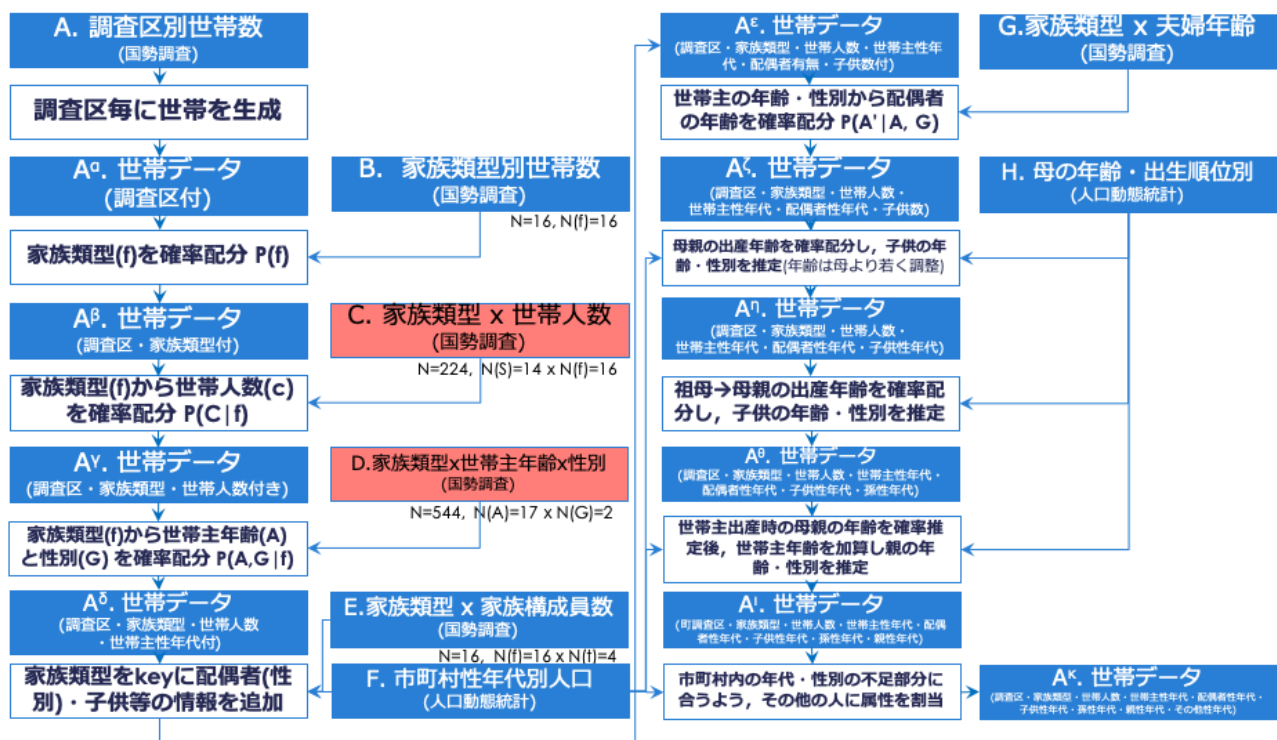


図 2：世帯データ推計のフロー

(1980 年版において、赤いクロス集計データは都道府県統計、別の市町村データから推計)

集計データを用いて、各世帯の家族類型を基準に世帯人数を確率配分していく。同様の手法で家族類型×世帯主年齢×性別(D)、家族類型×家族構成員数(E)、家族類型×夫婦年齢(G)から各属性を確率配分していく。使用しているデータは基本的に市町村別の国勢調査結果であるが、家族類型×世帯人数(C)、家族類型×世帯主年齢×性別(D)は 1980 年の市町村別での統計結果が公表されていなかったため、県別のクロス集計、市町村別の家族類型別世帯数、世帯人数別世帯数、世帯主年齢別世帯数から推定した。また、夫婦と親が同居している家族類型 5, 6, 7, 8, 11, 12 では、世帯主が夫婦世代か親世代か判別できる統計データがなかった。それぞれの世代が 5:5 の確率で世帯主であるとし、図 3 のように家族類型パターン 1, 2 を作成した。

次に、人口動態調査の母の年齢・出生順位別(H)出生数を用いて世帯主の年齢から子供、親の年齢を推定する。この段階で図 3 における世帯主、配偶者、子供、孫の年齢性別が全て配分し終わっている。最

後に、市町村性年代別人口(F)と今までの配分結果を照らし合わせ、不足分の年齢性別を補うようにその他に割り振り推計が完了する。

また、今回の推計で調査区別に世帯数を割り振ったが、当時の調査区と現在の行政区域の対応が不明

家族類型パターン 1					
番号	家族類型	配偶者	子供	親	その他
1	夫婦のみの世帯	1	0	0	0
2	夫婦と子供から成る世帯	1	1-9	0	0
3	男親と子供から成る世帯	0	1-9	0	0
4	女親と子供から成る世帯	0	1-9	0	0
5	夫婦と両親から成る世帯	1	0	2	0
6	夫婦とひとり親から成る世帯	1	0	1	0
7	夫婦、子供と両親から成る世帯	1	1-9	2	0
8	夫婦、子供とひとり親から成る世帯	1	9	1	0
9	夫婦と他の親族(親、子供を含まない)から成る世帯	1	0	0	1-9
10	夫婦、子供と他の親族(親を含まない)から成る世帯	1	1-9	0	1
11	夫婦、親と他の親族(子供を含まない)から成る世帯	1	0	1	1-9
12	夫婦、子供、親と他の親族から成る世帯	1	1-9	1	1
13	兄弟姉妹のみから成る世帯	0	0	0	1-9
14	他に分類されない世帯	0	0	0	1-9
15	非親族を含む世帯	0	0	0	0
16	単独世帯	0	0	0	0

家族類型パターン 2					
番号	家族類型	配偶者	子供	孫	その他
5	夫婦と両親から成る世帯	1	2	0	0
6	夫婦とひとり親から成る世帯	0	2	0	0
7	夫婦、子供と両親から成る世帯	1	2	1-9	0
8	夫婦、子供とひとり親から成る世帯	0	2	1-9	0
11	夫婦、親と他の親族(子供を含まない)から成る世帯	0	2	0	1-9
12	夫婦、子供、親と他の親族から成る世帯	0	2	1-9	1

図 3：家族類型(16)別の家族構成員数

なため、市町村内での世帯の位置分布は把握できない。国勢調査の調査区別集計結果には調査区番号のみ記載されていたが、国勢調査を執り行っている総務省統計局統計調査部国政統計課に問い合わせたところ、当時の調査区域は統計図書館で地図を閲覧できるのみでデータとして管理していないので、研究での活用は困難であるとの回答いただいた。

2.1.4. 1980年版と2010年版, Akiyama et al.(2013)と本研究の相違点

2010年版では世帯に住宅の属性を付与する。つまり、1980年版と異なり世帯の具体的な位置情報も推計している。建物データから町丁目別世帯数に合うように住宅を抽出し、世帯に割り振る。次に延床面積×住居種類、住居種類×家族類型のクロス集計から家族類型を確率配分し、その後の流れは1980年版と同様である。また、市町村別の家族類型×世帯人数、家族類型×世帯主年齢×性別の統計データは2010年の国勢調査では公開されているため、そのまま使用している。

Akiyama et al.(2013)と本研究の手法で異なる点は4点ある。1点目は、家族類型を推定するために使用していたクロス集計データを、延床面積×家族類型（都道府県別）から延床面積×住居種類（市町村別）、住居種類×家族類型（市町村別）に変更したことである。2点目は、夫婦と親が同居しており世帯主の世代が分からない家族類型5, 6, 7, 8, 11, 12において、夫婦世代：親世代＝5:5の確率で世帯主であるとした点だ。Akiyama et al.(2013)の手法では夫婦世代：親世代＝10:0となっていた。3点目は、その他に割り振る年齢性別は市町村の人口分布に従っていたので、途中までの配分結果と照らし合わせ不足分を補う形に変更した。以上3点により、精度は大幅に改善された。最後の4点目は、1980年でも推計を可能にしたことである。

2.2. 世帯推移モデル

2.2.1. 概要

作成した市町村毎の世帯推計データを元に、世帯をエージェントとする将来人口分布を推計するモデル

を構築した。推計手法には、将来人口推計で一般的なコーホート要因法を用いる。人口変動要因は出生、死亡、移動の3要因を考慮し、さらに移動は転出、転入、進学、結婚、離婚に細分化した。各要因の確率データを用い、1980～2010年、2010～2040年での世帯の変化を各市町村独立して5年毎シミュレーションした。

2.2.2. 使用するデータ

図4に使用したデータの一覧を示す。一段目のデータについて、1980～2010年は国勢調査結果から作成し、2010～2040年は国立社会保障・人口問題研究所（以下社人研とする）の地域別将来推計人口で活用され、現在公開されているデータを使用した。全て市町村毎のデータである。二段目のデータは全て国勢調査結果から作成し、2010～2040年は2010～2015年の傾向が継続されると仮定した。結婚率、離婚率、夫婦年齢分布、年齢別・出生順位別子供数、嫡出子率は全国データであり、親との同居率、離婚後親権率、転出世帯構成は都道府県別データとなっている。

	1980年～2010年	2010年～2040年
移動率(転出率、転入率) 生残率、出生率 0～4歳性別仮定値	国勢調査結果より作成	国立社会保障・人口問題研究所の推計値
結婚率、離婚率 夫婦年齢分布 親との同居率 離婚後親権率 転出世帯構成 年齢別・出生順位別子供数 嫡出子率	国勢調査結果より作成	
建物データ		Z-map TOWN II (2009年) テレポイントPack! (2009年)

図4：世帯推移モデルでの使用データ

2.2.3. シミュレーションの流れ(1980～2010年版)

1980～2010年のシミュレーションはモデルの精度を検証する目的があるため、世帯推計データは年齢、性別を補正することで男女年齢別人口を実測値に近付けたものを使用する。シミュレーションの流れを図5に沿って以下に示す。

1) 生残者：

生残率（年齢別、性別）に応じて推定する。

2) 転出世帯、転出者：

転出率（年齢別、性別）から転出者の人口分布を推定

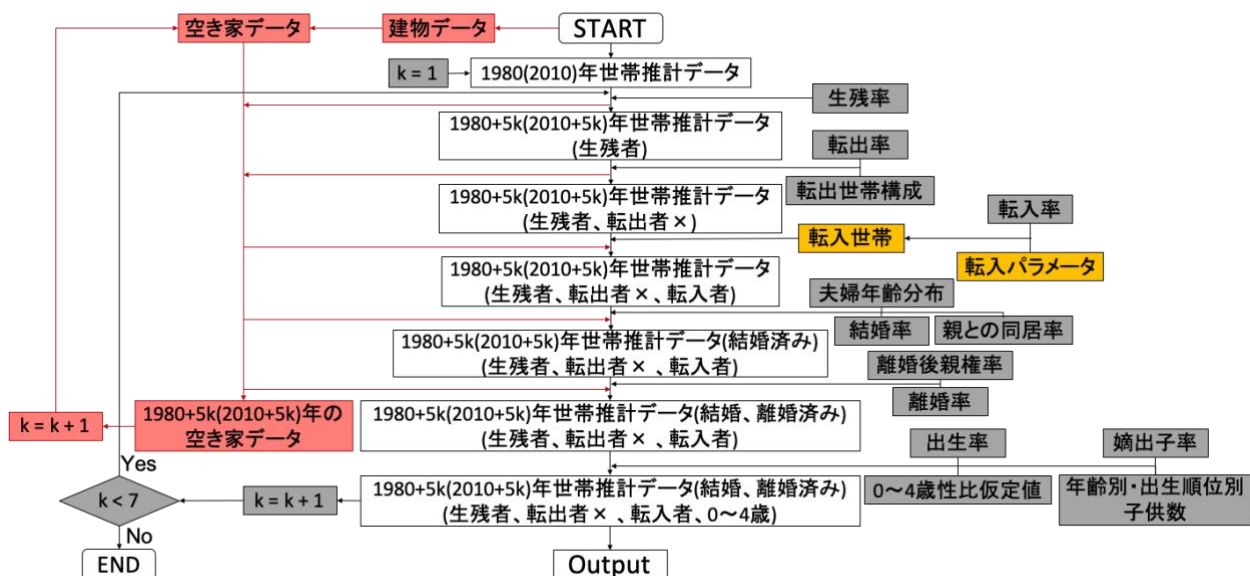


図5：世帯推移モデルのフロー
(赤色は2010～2040年のみ)

する。この分布、転出世帯構成（核家族世帯、核家族以外の親族世帯、単独世帯）に沿う形で世帯を選び転出させる。転出者に過不足が生じた場合は転入者の人口分布に足し合わせ調節する。また、10～14歳→15～19歳、15～19歳→20～24歳はそれぞれ高校、大学進学があるので、単身での転出も考慮する。

3) 転入世帯：

転入率、転入パラメータを用いて転入世帯を推定する。詳しくは後述する。

4) 結婚者：

女性の結婚率（年齢別）から女性の結婚者を推定する。夫婦年齢分布から夫の年齢を決定し、該当する未婚男性を市町村内からランダムに抽出する。親との同居率に合わせて家族類型1（夫婦のみの世帯）を形成するか、どちらかの家に住むか決める。親と同居する場合、優先的に男性の家に住む。また、親との同居率（年齢別）は都道府県の統計データであるため使用する際は、市町村別親と同居する夫婦世帯率/都道府県別親と同居する夫婦世帯率を乗じて補正する。

5) 離婚者：

女性の離婚率（年齢別、子供数別）から離婚する夫婦を推定する。男女半々の確率で片方が家を離れ同一

市町村内で新たに世帯を形成する。子供がいる場合は、離婚後親権率に基づいて父母どちらの世帯に引き取られるか推定する。

6) 0～4歳出生者：

15～49歳女性人口に出生率を掛け合わせ、0～4歳出生者数を算出する。さらに年齢別・配偶関係別・既存子供数別女性数、嫡出子率、年齢別・出生順位別子供数を用い、出生率（女性年齢別、配偶関係別、既存子供数別）を算出する。この確率に基づき0～4歳児を出生させ、性別は0～4歳性比仮定値から推定する。

転入世帯は図6、7のように決定する。1980(2010)年の家族類型別世帯割合 x と転入パラメータ a から、転入者の子供（20歳未満）、大人（20歳以上）がどの家族類型に属するか確率を算出する。この確率を転入率（年齢別、性別）から作成した転入者の人口分布に用いることで、転入世帯を推定する。

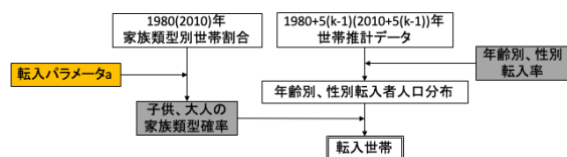


図6：転入世帯の構築

	家族類型	確率
子供 (20歳未満)	家族類型3 (夫婦と子供からなる世帯)	P_{child_3}
	家族類型7 (夫婦、子供と両親からなる世帯)	P_{child_7}
	家族類型4 (母親と子供からなる世帯)	$P_{child_16} = 1 - P_{child_3} - P_{child_7}$
(子供のいる世帯を生成した後の) 残りの大人 (20歳以上)	家族類型1 (夫婦のみの世帯)	P_{adult_1}
	家族類型4 (母親と子供からなる世帯)	P_{adult_4}
	家族類型16 (単独世帯)	P_{adult_16}
	家族類型6 (夫婦とひとり親からなる世帯)	$P_{adult_6} = 1 - P_{adult_1} - P_{adult_4} - P_{adult_16}$

$$P_i = a_{i,1}x_1 + a_{i,2}x_2 + a_{i,4}x_4 + a_{i,78}x_{78} + a_{i,16}x_{16} + a_i$$

(i = child_3, child_7, adult_1, adult_4, adult_16)

$a_{i,1}$ $a_{i,2}$ $a_{i,4}$ $a_{i,78}$ $a_{i,16}$ a_i : 転入パラメータ
 x_1 : 1980年(2010年)の家族類型1の世帯割合
 x_2 : 1980年(2010年)の家族類型2の世帯割合
 x_4 : 1980年(2010年)の家族類型4の世帯割合
 x_{78} : 1980年(2010年)の家族類型7, 8の世帯割合
 x_{16} : 1980年(2010年)の家族類型16の世帯割合

図7：子供、大人の家族類型確率

転入パラメータ a は手動でチューニングした P を用いて算出する。世帯構成を多様にするため富山県黒部市、南砺市、射水市、静岡県熱海市、裾野市の5都市を選出し、各都市で P の値をチューニングした。この際、2010年の家族類型別世帯数が家族類型1 > 2 > 7, 8 > 16 > 4の優先順位で実測値に近づくよう調節した。求められた P と1980(2010)年の家族類型別世帯割合 x を重回帰分析することで転入パラメータ a を算出した。

2.2.4. 1980~2010年版と2010~2040年版の相違点

2010~2040年版は1980~2010年版と比べて各世帯の住居も推定する。建物データから建物属性が一般住宅(民家)、共同住宅、住商混合(住宅中心)、住商混合(事業所中心)の建物を抽出し、2010年の世帯推計データで配分されていない住居を空き家リストとしてまとめる。死亡、転出により空き家ができた場合は空き家リストに追加する。転入、結婚、離婚で世帯が発生した際は、同じ家族タイプの世帯数に応じて転入する500mメッシュを確率的に選択し、メッシュ内の空き家から住居をランダムに選択する。

3. 結果

世帯推計データ、世帯推移モデルの結果を国勢調査結果と比較することで精度検証する。ただし、世

帯推移モデルのシミュレーション結果である2040年に関しては国勢調査結果が存在しないため、社人研の市町村別将来推計人口、国土交通省国土政策局総合計画課の500mメッシュ別将来推計人口と比較することでモデルの妥当性を検証する。

3.1. 市町村毎の世帯の推計精度

3.1.1. 1980年版の精度検証

市町村毎に5歳階級別人口(85歳以上は一区分)を世帯推計データ、国勢調査で集計し、それぞれを推計値、実測値として、決定係数 R_2 (以下 R_2 とする)と平均平方二乗誤差 RMSE(以下 RMSE とする)を算出した。また、同様に家族類型16分類別も集計し、精度を検証した。図8が示すように、年齢別人口の決定係数 R_2 は0.7~0.9に収まり、家族類型別世帯数の R_2 は限りなく1に近かった。RMSEは R_2 が大きくとも、人口規模の大きい市町村であれば大きな値となっている。つまり、 R_2 が大きく RMSE も大きい市町村は人口が多く、 R_2 が小さく RMSE も小さい市町村は人口が少ない。

市町村名	1980年				2010年				メッシュ別人口	
	年齢別人口		家族類型別世帯数		年齢別人口		家族類型別世帯数		メッシュ別人口	
	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
富山市	0.829	4121	0.999	53.92	0.895	1752	0.978	2194	0.964	74.5
高岡市	0.798	2128	0.999	40.17	0.902	828.3	0.988	557.2	0.956	80.68
魚津市	0.733	643.9	0.999	29.15	0.87	235.1	0.982	183.7	0.961	56.06
氷見市	0.693	782.8	0.999	21.15	0.825	354.6	0.99	113.4	0.955	38.48
滑川市	0.748	359.5	0.998	28.89	0.87	163.2	0.992	81.21	0.945	49.53
黒部市	0.806	444.9	0.999	18.12	0.871	194	0.975	195.1	0.928	50.23
砺波市	0.65	565.9	0.998	35.37	0.731	316.7	0.941	289.1	0.934	39.84
小矢部市	0.688	455	0.998	25.81	0.825	198.9	0.981	88.26	0.948	35.9
南砺市	0.67	818.6	0.999	27.6	0.771	389.8	0.979	164	0.939	34.36
射水市	0.753	1185	0.999	31.21	0.888	454	0.991	244.7	0.959	67.58
中新川郡	0.777	570.1	0.999	26.79	0.836	294.5	0.987	162.3	0.97	27.02
下新川郡	0.708	560.7	0.999	19.02	0.827	247.3	0.985	119.3	0.946	37.49
静岡市	0.892	6683	0.999	134.1	0.909	2896	0.994	2018	0.984	98.31
浜松市	0.884	6354	0.999	110.6	0.849	3686	0.985	3432	0.972	87.98
沼津市	0.897	1899	0.999	67.27	0.923	851	0.982	988.8	0.977	102.3
熱海市	0.781	557.4	0.999	13.79	0.941	238.5	0.982	311.4	0.9	124
三島市	0.874	994.3	0.999	38.5	0.898	523.1	0.971	742.7	0.942	168.1
富士宮市	0.871	1196	0.999	27.51	0.898	594.1	0.979	596.4	0.964	66.59
伊東市	0.758	898.8	0.999	30.66	0.92	426.9	0.992	264.4	0.931	90.14
島田市	0.818	1053	0.999	26.66	0.796	525.1	0.992	247	0.981	49.9
富士市	0.893	2192	0.999	28.68	0.907	1134	0.98	1154	0.968	96.6
磐田市	0.853	1516	0.999	43.75	0.897	690.5	0.977	805.8	0.954	79.56
焼津市	0.864	1305	0.999	23.15	0.902	616.5	0.989	450.3	0.957	101.1
掛川市	0.783	1032	0.999	27.88	0.835	558.2	0.974	549.3	0.968	48.46
藤枝市	0.858	1202	0.999	53.69	0.872	713.3	0.993	374.1	0.97	80.81
御殿場市	0.944	494.3	0.999	13.7	0.968	247.3	0.969	514.7	0.942	91.36
袋井市	0.803	668.2	0.999	28.38	0.937	306.3	0.968	487.2	0.929	77.82
下田市	0.679	414.3	0.999	16.17	0.931	132.4	0.984	132.8	0.956	40.17
裾野市	0.903	413.2	0.999	18.92	0.921	226.3	0.929	556.4	0.875	149.5
湖西市	0.876	529.5	0.999	22.76	0.929	221.7	0.981	272.1	0.972	58.7
伊豆市	0.64	504.4	0.999	11.69	0.818	257.7	0.963	200	0.944	30.73
御前崎市	0.769	339.1	0.998	24.82	0.354	324.5	0.789	420.2	0.929	53.67
菊川市	0.791	401.1	0.999	10.2	0.836	231.9	0.965	241.2	0.942	53.44
伊豆の国市	0.808	519.3	0.999	30.84	0.893	246.6	0.978	250.3	0.964	62.27
牧之原市	0.746	531.9	0.998	40.16	0.754	269	0.972	196.9	0.954	48.23
賀茂郡	0.543	922.9	0.999	22.72	0.902	353	0.982	236.5	0.931	39.23
田方郡	0.781	403.2	0.999	18.06	0.907	191.1	0.991	122.3	0.885	146.9
駿東郡	0.843	976.4	0.999	22.87	0.941	337.3	0.982	433.9	0.938	135.5
榛原郡	0.744	393.1	0.999	22.32	0.587	235.9	0.963	211.2	0.965	49.93
周智郡	0.646	241.2	0.998	14.84	0.706	120.6	0.986	55.34	0.958	32.7

図8：世帯推計データの市町村別精度評価

3.1.2. 2010 年版の精度検証

図 8 のように、2010 年版の精度評価では 1980 年版の評価項目に加え、メッシュ別人口でも精度を検証している。2010 年版の世帯推計データには世帯毎の住居を建物データから推定しており、位置情報も付与されているため、500m メッシュ別に人口を集計し、実測値と精度検証が可能である。年齢別人口の R_2 はほとんどの市町村で 0.8 以上に収まり、RMSE も 1980 年版に比べ小さくなり、高い精度になっている。しかしながら、御前崎市のみ R_2 が 0.354 と非常に低い数値になっていた。家族類型別世帯数では R_2 が御前崎市以外で 0.9 以上であるものの、RMSE は 1980 年版と比較して大きな値になっている。また、メッシュ別人口は R_2 がほとんどの市町村で 0.9 以上となっていた。

3.2. 世帯推移モデルのシミュレーション結果

3.2.1. 1980~2010 年版の精度検証

転入パラメータは図 9 のようになった。以下のパラメータを用い、1980 年版世帯推計データの男女年齢別人口が実測値に近づくように補正し、富山県、静岡県全市町村で 2010 年までシミュレーションした。年齢別人口、家族類型別世帯数の精度検証方法は世帯推計データと同じであり、シミュレーション終了時の 2010 年における精度を評価している。年齢別人口の R_2 は全市町村で 0.95 以上と高い精度となっている。RMSE は人口規模の大きな自治体ほど高い値となった。家族類型別世帯数の R_2 はほとんどの市町村で 0.9 以上となっていたが、RMSE にはばらつきがあった。人口規模を考慮しても RMSE が 3000 以上と特に大きな富山市を精査したところ、全世帯数の 28.8% を占める単独世帯において、世帯数を 28.7% 過少推計していた。家族類型によって精度が大きく変化する可能性が示唆されたため、家族類型別の精度を検証し以下に示す。

i \ k	1	2	4	78	16	NaN
child_3	-0.0977936	-0.3798679	12.4124957	-1.6951169	-6.0479777	1.38251712
child_7	2.66676261	-1.5873946	-5.1418307	-1.1659207	-2.3444115	1.09100777
adult_1	-0.8855686	1.35379431	-5.7397158	1.66175497	2.86681421	-0.2545823
adult_4	-0.2172224	-0.4455045	2.24489062	0.11338751	-0.5008177	0.19810461
adult_16	1.10279098	-0.9082898	3.49482514	-1.7751425	-2.3659965	1.05647773

図 9：転入パラメータ a_{ik}

市町村名	年齢別人口		家族類型別世帯数		市町村名	年齢別人口		家族類型別世帯数	
	R^2	RMSE	R^2	RMSE		R^2	RMSE	R^2	RMSE
富山市	0.986896	620.495	0.941616	3622.1	富土市	0.997402	190.1936	0.977976	1215.239
高岡市	0.994129	203.7557	0.983067	666.9627	磐田市	0.957253	445.8697	0.991329	503.932
魚津市	0.997224	34.43512	0.980846	190.1677	焼津市	0.997315	102.5175	0.941532	1046.084
氷見市	0.998825	29.0813	0.984348	142.4961	掛川市	0.975967	213.3098	0.990074	342.3313
清川市	0.995373	30.81846	0.970545	165.4337	藤枝市	0.996657	116.8506	0.927994	1212.034
黒部市	0.991792	49.05043	0.98693	141.6369	御殿場市	0.968343	249.4432	0.988229	319.0696
砺波市	0.998467	23.95018	0.980537	167.2715	袋井市	0.988713	129.8679	0.993327	224.5475
小矢部市	0.985608	57.19606	0.955638	136.8	下田市	0.999134	14.91643	0.877199	371.8179
南砺市	0.996861	45.68187	0.977776	170.3098	裾野市	0.937232	202.9992	0.869924	210.7426
射水市	0.998236	57.13285	0.988893	279.2076	湖西市	0.980127	167.1699	0.974462	322.3604
中新川郡	0.994963	51.64565	0.974044	231.5934	伊豆市	0.996742	34.50121	0.829646	430.7442
下新川郡	0.981414	81.17539	0.962053	193.8199	御前崎市	0.98732	45.47099	0.85237	111.3991
静岡市	0.991944	861.8065	0.977686	3987.004	菊川市	0.996354	34.65945	0.993814	101.7697
伊賀市	0.966572	1739.122	0.983837	3683.355	伊豆の国市	0.984817	93.10687	0.916433	490.1784
沼津市	0.998672	111.9223	0.991157	708.7441	牧之原市	0.995849	34.97856	0.963895	223.9255
熱海市	0.996378	59.23025	0.993555	186.4437	賀茂郡	0.99823	47.58443	0.725208	935.8839
三島市	0.997709	78.53945	0.995719	290.0324	田方郡	0.996745	35.79261	0.742988	655.0986
富士宮市	0.993474	150.8892	0.957075	862.0185	駿東郡	0.995899	89.26179	0.990828	315.644
伊東市	0.992375	132.3772	0.781794	1408.927	榛原郡	0.902861	114.3941	0.979766	157.906
島田市	0.996493	68.92669	0.935433	707.3902	周智郡	0.993042	18.58464	0.949957	105.4582

図 10：2010 年時点の市町村別精度評価

家族類型を核家族（家族類型 1～4）、核家族以外の親族のみの世帯（家族類型 5～14）、単独世帯（家族類型 16）に再分類し、それぞれの世帯数推計値と実測値の誤差割合を横軸に、市町村数を縦軸に取りヒストグラムを作成した。核家族はほとんどの市町村が誤差 10% 以内に収まっており、核家族以外の親族のみの世帯は誤差 -30~10% の範囲に収まっている。全体的に過少推計している様子が窺える。また、単独世帯の誤差は大きく、中には 50% 以上の自治体もある。多くの市町村で過大推計している。

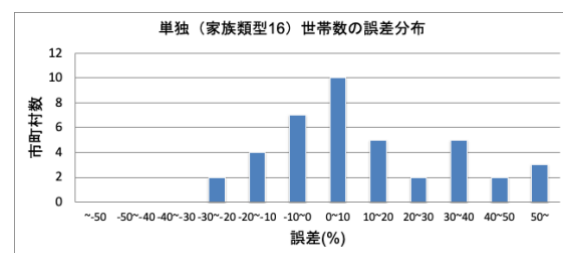
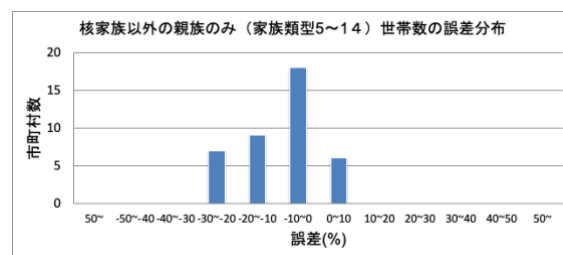
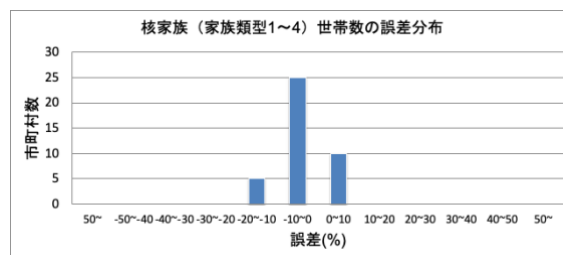


図 11：家族類型別世帯数（核家族、それ以外の親族世帯、単身世帯）の誤差分布

3.2.2. 2010~2040 年版の妥当性検証

富山県黒部市、南砺市、射水市、静岡県熱海市、裾野市において、2010 年の世帯推計データを補正せず、転入パラメータは図 9 と同じ値を用い 2040 年までシミュレーションした。結果を社人研の市町村別将来推計人口、国土交通省国土政策局総合計画課の 500m メッシュ別将来推計人口と比較し、その妥当性を検証した。実測値との比較ではないので、精度の検証にならないことは留意しておきたい。市町村別の家族類型別世帯数の推計値は公表されていないため、人口のみを年齢別、メッシュ別に今までと同様の手法で検証した。図 1 2 のように、年齢別人口では射水市以外で R^2 が 0.8 以上と社人研と同じ傾向を持つ結果となり、妥当性を確認できた。一方、メッシュ別人口では全市町村で R^2 が 0.7 以上となり、年齢別人口には劣るが一定の妥当性が理解できる。どれも比較的同規模の人口を持つ市町村であるが、RMSE に大きな差が出ている。

市町村名	年齢別人口		メッシュ別人口	
	R^2	RMSE	R^2	RMSE
黒部市	0.813183	192.6286	0.744409	81.02321
南砺市	0.815601	289.477	0.706626	44.17423
射水市	0.695613	588.5018	0.805948	126.0514
熱海市	0.905624	223.7128	0.890454	91.22361
裾野市	0.851822	178.7268	0.795952	161.302

図 1 2 : 2040 年時点の市町村別精度評価

500m メッシュ別における社人研推計値、シミュレーション結果の人口を図 1 3 のようにプロットした。横軸にシミュレーション結果の人口、縦軸に社人研推計値を取っている。直線は回帰線を示しており、黒部市、南砺市、射水市では傾きが 0 より小さいことから、社人研の推計値に比べ本研究では人口を過大推計している傾向が読み取れる。また、 R^2 に比較して特に RMSE の大きな裾野市においては、(シミュレーション結果, 社人研推計値) = (2146, 7541) となるメッシュが RMSE に大きく寄与している様子が理解できた。

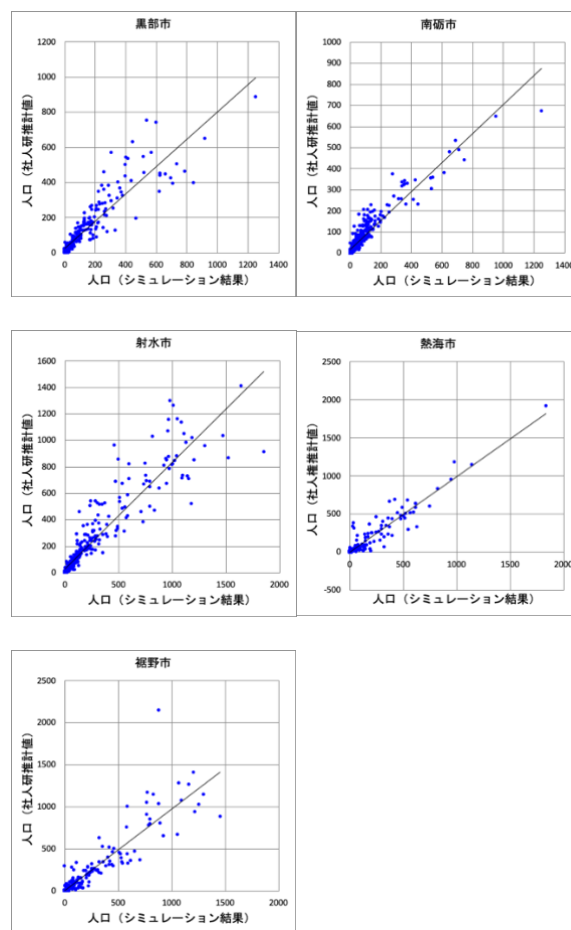


図 1 3 : 500m メッシュ別人口の散布図

4. 考察

4.1. 市町村毎の世帯推計データ

4.1.1. 1980 年版について

年齢別人口の R^2 は 0.7~0.9 に収まる自治体がほとんどであり、2010 年版に比べて精度が劣っていた。元々の推計手法が 2010 年の国勢調査で公開されている統計項目に合わせて作成されており、1980 年版で足りない項目は推計していたため、予想できた結果であった。特に、図 2 における家族類型×世帯主年齢×性別(D)で実データを用いていないことが原因だと思われる。世帯構成員の年齢、性別はいずれも世帯主を基準に推計されるので、世帯主の年齢、性別分布を精緻に推計できないと精度大きく悪くなる。その一方、家族類型別世帯数の R^2 は 1 に極めて近く、RMSE も 2010 年版と比べて大きく下回っていた。これは市町村別の家族類型別世帯数のデータを

そのまま当てはめていったので、当然の結果である。しかしながら、 R_2 が 1, RMSE が 0 にならないのは、確率配分しているために実測値と少なからず誤差が生じているためだと考えられる。

4.1.2. 2010 年版について

年齢別人口において、 R_2 , RMSE 共に 1980 年版から大きく改善されており、メッシュ別人口も高い精度となっている。それでも精度が不十分な市町村はいくつかあり、原因は大きく分けて二つあると考えられる。一つ目は、夫婦と親が同居している家族類型 5, 6, 7, 8, 11, 12 での世帯主の世代の割合である。これは推計結果に大きな影響を与える。例えば、世帯主の年齢が 50 代である際、夫婦世代が世帯主なら親の年齢は 80 代前後で世帯全体は 50~80 代、親世代が世帯主なら夫婦の年齢は 20 代前後で世帯全体は 20~50 代と年齢分布が大きく変わってくる。Akiyama et al.(2013)の手法では夫婦世代：親世代=10:0 となっていたため、高齢者を過剰に推計していた。今回の研究では統計データが得られないため、夫婦世代：親世代=5:5 としたが、本来は家族類型、世帯主年齢性別に左右されるものである。より精緻なデータが手に入れば更なる精度向上が期待できる。二つ目は、2009 年建物データと 2010 年国勢調査における町丁字コードの不一致である。2010 年版の世帯推計データでは、国勢調査の町丁字別世帯数に応じて住居系の建物データに世帯を割り振っている。この際に町丁字コードを使用しており、市町村コードの不一致により住居系の建物データに割り振られない世帯が発生している。2009 年から 2010 年にかけて一部町丁字が変更され、コードが変更されているためだと考えられる。これにより、御前崎市での推計精度が著しく低下した。2009 年、2010 年の小地域区域の対応データを手に入れ、早急に対応する必要がある。

4.2. 世帯推移モデル

4.2.1. 1980~2010 年版について

年齢別人口の R_2 が全市町村 0.95 以上、家族類型別世帯数の R_2 はほとんどの市町村で 0.9 以上と高い

精度となっている。しかしながら、図 1-1 のように家族類型別の世帯数では大きな誤差が発生している地域もある。特に単独世帯では 50%以上の誤差となる市町村もあった。これは、元々単独世帯は結婚、出生で家族類型が変わりやすく精緻なシミュレーションが難しいことに加え、転入する単独世帯を汲み上げる際、年齢を考慮していないことが原因である。例えば、70 歳未婚男性と 25 歳未婚男性では 5 年後の結婚率が大きく異なり、単独世帯として転入してきても 5 年後に夫婦世帯となる確率は、25 歳未婚男性の方が明らかに高い。転入世帯をより現実に近い年齢性別分布で推計できれば、精度を上げられるだろう。転入パラメータを求める際に使用したサンプルも今回は 5 市町村と非常に少なく、サンプル数を増やすことでの精度改善も期待できる。また、今回の研究で世帯推移パターンを簡略化したことが精度に影響を与えていたと考えられる。例えば、結婚、離婚は市町村内でのみ発生すると仮定したが、本来は市町村、都道府県を跨いで発生するものである。全ての世帯推移パターンをモデル内で再現することは非常に困難であるが、精度に大きく寄与するパターンを見極め、モデルに取り入れていく必要がある。

4.2.2. 2010~2040 年版について

年齢別人口の R_2 は大体 0.7~0.9 に収まっているものの、ばらつきが大きい。シミュレーションで使った世帯推計データは年齢別人口に補正をかけなかったもので、誤差が広がったと考えられる。年齢別人口を補正した世帯推計データでシミュレーションした 1980~2010 年版では高い精度となっていたので、2040 年時点における年齢別人口の誤差には世帯推移モデルより世帯推計データの精度が大きく寄与していると言える。メッシュ別人口の精度には、これに加えて住居選択モデルの簡易さが強く影響していると思われる。今回のモデルでは同じ家族類型の世帯数に応じて居住するメッシュを選択していたが、住居の床面積、周辺施設へのアクセスなども考慮し、より現実的な世帯の居住地選択モデルを組んでいく必要がある。

4. おわりに

本研究では、富山県、静岡県全域を対象に世帯推計データを1980年版では国勢調査から、2010年版では建物データも加えて作成した。元々2010年国勢調査統計項目に合わせた手法であるため、1980年版では十分な精度を得られなかったが、2010年版では一定の精度を確認できた。しかしながら、町丁字コードの不一致から、精度が著しく低下している市町村も確認できた。世帯推移モデルでは1980~2010年のシミュレーションで精度確認を行い、年齢別人口では高い精度を確認できたものの、単独世帯数ではばらつきが大きい結果となった。2010~2040年のシミュレーションではメッシュ別人口で誤差の大きな地域も確認でき、世帯の居住地選択モデルの構築が望まれる。

謝辞

世帯推計データを作成するにあたり、関本研究室協力研究員の金杉洋氏、Pacific Spatial Solutions 株式会社の中川貴滋氏にはご助力いただき誠にありがとうございました。また、東京大学 CSIS 共同研究「世帯類型を考慮した世帯単位の人口分布予測シミュレーションに関する研究」および情報通信研究機構委託研究「デジタルコミュニケーションを駆使した小規模自治体における市民協働型まちづくり」の成果の一部である。皆様に心より感謝の意を表します。

参考文献

江守央・伊澤岬・横山哲 (2009) 市民参加型まちづくりの変遷に関する基礎的研究。土木計画学研究・講演集(CD-ROM), Vol.39 ROMBUNNO.184,
Hasegawa, Y., Sekimoto, Y., Seto, T., Fukushima, Y. and Maeda, M. (2019) My City Forecast: Urban Planning communication tool for citizen with National Open Data, Computers, Environment and Urban Systems
Paul Waddell (2002) UrbanSim: Modeling Urban Development for Land Use, Transportation, and Environmental Planning, Journal of the American Planning Association, Volume 68, p297-314

山際康平・藤井秀樹・吉村忍 (2017) メゾスコピックモデルを用いたマルチエージェント世帯推移シミュレーション。人工知能学会論文誌, 32 巻 5 号, p. AG16-A_1-10.

北詰恵一・杉木直・V arameth VICHENSAN・宮本和明 (2008) マイクロシミュレーション都市モデルの現状と課題, 土木計画学研究・講演集, Vol.37,

Yuki Akiyama, Hiroyuki Takada and Ryosuke Shibasaki (2013) Development of Micropopulation Census through Disaggregation of National Population Census, CUPUM 2013 conference

国立社会保障・人口問題研究所 (2011) 将来推計人口の方法と検証-平成 18 年推計の仕組みと評価-。第 12 回社会保障審議会人口部会資料 2

国立社会保障・人口問題研究所 (2018) 日本の地域別将来推計人口 (平成 30(2018)年推計)-平成 27(2015) ~ 57(2045) 年 -。 http://www.ipss.go.jp/pp-shicyoson/j/shicyoson18/1kouhyo/gaiyo_s.pdf (最終閲覧日: 2020 年 8 月 30 日)

国土交通省国土政策局総合計画課 (2019) 平成 27 年国勢調査を基準とした 500m 及び 1km メッシュ別将来推計人口の試算方法について。 https://nlftp.mlit.go.jp/ksj/gml/datalist/mesh500_1000_h30.pdf (最終閲覧日: 2020 年 8 月 30 日)

参照データサイト

e-Stat 政府統計の総合窓口, 独立行政法人統計センター, <https://www.e-stat.go.jp> (最終閲覧日: 2020 年 8 月 30 日)

JoRAS, 東京大学 空間情報科学研究センター, <https://joras.csis.u-tokyo.ac.jp> (最終閲覧日: 2020 年 8 月 30 日)

Z-map TOWN II, テレポイント Pack! (2009 年), 株式会社ゼンリン

将来推計人口・世帯数, 国立社会保障・人口問題研究所 <http://www.ipss.go.jp/syoushika/tohkei/MainMmai.asp>, (最終閲覧日: 2020 年 8 月 30 日)

国土数値情報ダウンロードサービス, 国土交通省, <https://nlftp.mlit.go.jp/ksj/index.html>, (最終閲覧日: 2020 年 8 月 30 日)