

世帯マイクロデータ構造の改良と 推定手法間の精度検証

杉木 直¹・鈴木 温²・宮本 和明³

¹正会員 豊橋技術科学大学 工学部 (〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘1-1)
E-mail: sugiki@ace.tut.ac.jp

²正会員 名城大学 理工学部 (〒468-8502 愛知県名古屋市天白区塩釜口1-501)
E-mail: atsuzuki@meijo-u.ac.jp

³フェロー パシフィックコンサルタンツ株式会社 (〒101-8462 東京都千代田区神田錦町3-22)
E-mail: kazuaki.miyamoto@tk.pacific.co.jp

都市マイクロシミュレーションモデルは、その多様な属性と時空間的な変化の記述性能より、人口減少時代の居住地選択および都市構造の分析に有効であると考えられる。本研究では、その実行のために必要とされる初期世帯マイクロデータについて、既存研究による手法の将来シミュレーションへの適用を通じて明らかとなったデータ構造上の問題点を整理し、改良を行う。また、改良されたデータ構造を持つ初期世帯マイクロデータの推定について、アンケート調査により取得したサンプルマイクロデータを用いるエージェントベースの手法と、国勢調査等の入手可能な統計データのみを用いて簡易的に行う手法を検討する。その上で、富山市を対象として両手法による初期世帯マイクロデータの推定を実施し、推定結果の比較から精度の検証を行う。

Key Words: *Micro-simulation, Micro-Data, Data Structure, Estimation Accuracy, Open Data*

1. はじめに

人口減少下の都市圏においては、従来の画一的な政策手段の実施では、多様化する世帯状況に対する効果は限定的である。限られた資源のもとで効率的に公共サービスを提供するためには、従来よりもきめ細かな視点で、より詳細な世帯属性を考慮に入れた政策手段の選択が不可欠である¹⁾。これは、市民が求める公共サービスの質と量はその人の年齢や属する世帯構成等の個人属性に大きく依存し、その住居と施設の立地分布と交通状況がその費用を大きく規定するためである。今後増大する公共サービス需要に対して、受動的にその供給を行うことは明らかに社会的な効率性が悪く、「公共サービス需要管理 (Public Service Demand Management)」の視点から、適切な立地誘導を含めた効率的な政策手段の実施が求められる。

このような視点に立った政策手段の効果予測においては、従来用いられてきた立地主体をグループ化したメゾ

スケールのモデルでは原理的に対応が困難であり、個人を考慮した世帯単位でのマイクロシミュレーションモデルが不可欠である。マイクロシミュレーション型都市モデルは、近年欧米諸国を中心として積極的なモデル開発、適用に関する研究が進められている²⁾³⁾。居住立地にマイクロシミュレーションを導入した代表的な都市モデルとしては、UrbanSim⁴⁾、ILUTE⁵⁾、ILUMASS⁶⁾、PUMA⁷⁾、SelfSim⁸⁾などが挙げられ、研究事例および実際の都市への適用事例の蓄積が進められている。世帯の多様な属性（高齢世帯、子育て世帯など）によって異なる都市サービス需要の予測、ライフステージ進行に伴う居住世帯の時空間的な属性の変化の表現などの性能において、人口減少時代の居住地選択および都市構造の分析に有効であると考えられる。

本研究グループでは、このような背景を踏まえ、中長期的に変化する世帯属性変化に基づいて、要求するサービスや満足度、評価の変化を予測することを目的とした世帯マイクロシミュレーションモデルの開発を行ってき

た⁹⁾¹⁰⁾¹¹⁾。また、世帯を対象としたマイクロシミュレーションモデルを実行するためには、全ての世帯に対して世帯構成、住宅タイプ、自動車保有状況等の多様な属性を定義したシミュレーション初期年次データを用意する必要があるが、住民基本台帳などから個人や個別世帯に関するデータを入手することは一般的に困難であり、またプライバシー保護の観点からも望ましくない。このような課題に対し、国勢調査などの入手可能な集計データと、個別世帯の属性情報を追加的に提供するサンプル調査を組み合わせ、世帯や個人のデータ推計を行う手法に関して、体系的な手法の構築¹²⁾¹³⁾や、入手可能な公表データのみを用いた推計手法の開発¹⁴⁾¹⁵⁾を行ってきた。しかしながら、これらの一連の研究で構築された世帯マイクロデータの構造においては、時間経過に伴い世帯構成員の離脱や新規加入が生じた際に、世帯構成員の続柄の更新処理が困難であるという課題があった。

そこで本研究では、初期世帯マイクロデータについて、既存研究による手法の将来シミュレーションへの適用を通じて明らかとなったデータ構造上の問題点を整理し、改良を行う。また、改良されたデータ構造を持つ初期世帯マイクロデータの推定について、アンケート調査により取得したサンプルマイクロデータによりモンテカルロ法を用いて行うエージェントベースの手法と、国勢調査等の入手可能な統計データのみを用いて簡易的に行う手法を検討する。その上で、富山市を対象として両手法による初期世帯マイクロデータの推定を実施し、推定結果の比較から精度の検証を行うとともに、各手法の利点と欠点、および適用可能範囲等の視点から考察を行う。

2. 世帯マイクロデータの課題と改良

(1) 既存世帯マイクロデータの構造

既存研究において、マイクロデータは年齢、性別、世帯主との続柄、その他の個人属性（免許保有有無、就業就学状況、通勤通学地等）、その他の世帯属性（住宅タイプ、居住ゾーン、世帯収入、自動車保有台数等）の集合として定義している。m人からなる世帯sのデータ h_{ms} を式(1)で表わす。

$$h_{ms} = \{c_{ms}, x_{ms}, b_{ms}, z_s\} \quad (1)$$

ここで、 c_{ms} は m 人からなる世帯 s の世帯構成、 x_{ms} は年齢構成、 b_{ms} はその他の個人属性、 z_s はその他の世帯属性である。世帯データは、世帯構成員カテゴリの一般形に対する年齢ベクトルとして表現される。世帯構成員カテゴリは、世帯主との続柄および性別によって定義される。マイクロデータセット全体の一般的な表現は式(2)で表わされる。

$$A = \{a_i = (a_{i1}, a_{i2}, \dots, a_{iR}) \mid 1 \leq i \leq N\} \quad (2)$$

ここで、 R は世帯構成員タイプ数、 N は対象地域の世帯数、 a_{ik} は世帯構成員 k の年齢である。世帯 i に世帯構成員 k が存在しない場合、 a_{ik} には不在を表すダミー変数(999)を割り当てる。

実都市を対象とした具体的な初期世帯マイクロ世帯の推定においては、世帯構成員タイプ表-1 に示す 20 タイプを設定している。これらの世帯構成員タイプは世帯主を基準とした続柄により定義されるため、世帯構成員のうち誰を世帯主とするかについて明示的なルールが必要であり、若いほど優先される以下の 3 つのシンプルな基準により世帯主を定めている。

- ①世帯内に 1 組以上の夫婦がいる場合は、夫婦の男性のうち最年長の男性
- ②世帯内に夫婦がおらず、23 歳以上の男性がいる場合は最年長の男性
- ③世帯内に夫婦がおらず、23歳以上の男性がいない場合は、最年長の女性

表-1 世帯主との続柄による世帯構成員タイプの設定

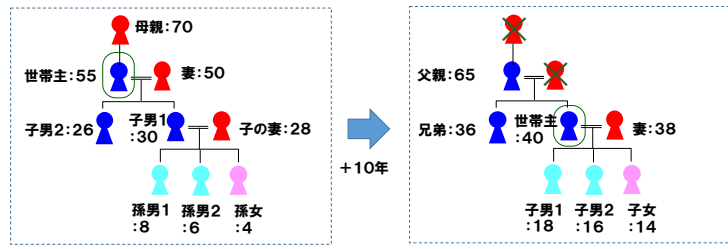
1	世帯主・男	11	妻
2	子供・男1	12	子供・女1
3	子供・男2	13	子供・女2
4	子供・男3	14	子供・女3
5	孫・男	15	孫・女
6	兄弟	16	姉妹
7	父	17	母親
8	その他・男1	18	子の妻
9	その他・男2	19	その他・女1
10	世帯主・女	20	その他・女2

(2) データ構造上の課題

推定された初期マイクロ世帯データを用いて将来シミュレーションを実行する場合、時間経過に伴う世帯構成員の変化を表現する。その際、構成員の死亡、離婚、離家（構成員の独立、世帯の分離）といったイベントが発生した場合は、マイクロ世帯データからの該当構成員の削除を行う必要がある。また、結婚や出産等のイベントが発生した場合は、マイクロ世帯データへの構成員の追加が生じる。世帯構成員タイプが世帯主を基準とした続柄であるデータ構造上、世帯構成員の削除や追加が生じた場合、前述の基準に基づいた世帯主の変更や、各世帯構成員の続柄の更新を行う必要がある。

図-1 に続柄の更新の例を示す。母親と妻の 2 名が世帯から離脱したことにより、世帯主が子世代に移動し、その他すべての世帯構成員の続柄を変更しなければならない。このような続柄に関するデータの更新は非常に複雑であり、プログラム化が困難である。また、孫世代において性別が同じ場合、2 人目以降はその他男性、または

その他女性として扱われるため、子世代へのデータ更新が不可能となる。ここで示した例はやや極端な例ではあるが、世帯構成員が少ない場合でもあっても、またシミュレーション期間が長くなるほどシミュレーション結果に与える影響が大きい。



(3) 世帯マイクロデータ構造の改良

以上の課題を踏まえ、本研究では世帯マイクロデータ構造の改良を行う。新たなデータ構造では、 m 人からなる世帯 s のマイクロ世帯データを、式(3)のように各世帯構成員の属性ベクトル \mathbf{d}_{ms} と、その他の世帯属性 \mathbf{z}_s で定義する。

$$h_{ms} = \{\mathbf{d}_{ms}, \mathbf{z}_s\} \quad (3)$$

各世帯構成員の属性 \mathbf{d}_{ms} は、世代と性別の組み合わせによる変数 g_{ms} 、年齢 x_{ms} 、婚姻状態を表す変数 r_{ms} 、その他の個人属性 \mathbf{b}_{ms} を情報として持つ個人データとして、式(4)のように定義する。

$$\mathbf{d}_{ms} = \{g_{ms}, x_{ms}, r_{ms}, \mathbf{b}_{ms}\} \quad (4)$$

ここで世代と性別の組み合わせによる変数 g_{ms} は、1桁目を性別（男性=1、女性=2）、2桁目を世代（1世代目=1、2世代目=2、3世代目=3、…）とする2桁の変数である。また、婚姻状態を表す変数 r_{ms} は未婚の場合は0であり、既婚の場合は婚姻相手の個人IDが格納される。以上のような改良により、続柄を含まないデータ構造となった。図-1の例について、改良後のデータ構造においてデータ更新する場合のイメージを図-2に示す。T時点における1世代目の母親が離脱することにより、T+10時点では各人の世代が1つ繰り上がっている。また、T時点の妻が離脱することにより、T+10時点では婚姻状態が変化するとともに、世帯主が変更となる。世帯主の判定のみ前述のルールで行うこととすれば、世帯構成員の離脱による属性の更新は、1世代目が空集合となった場合の各世帯構成員の世代の繰り上げと、婚姻状態変数の更新のみで行うことができ、単純な処理で実行することが可能となった。

3. 初期世帯マイクロデータの推定手法

ここでは、前節で提案した世帯マイクロデータに対する初期データの推定手法について、アンケート調査により取得したサンプルマイクロデータを用いるエージェントベースの手法と、国勢調査等の入手可能な統計データのみを用いて簡易的に行う手法を検討する。

性別が同じ2人目以降の孫はその他扱い

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	本人・男	子・男1	子・男2	子・男3	孫・男	兄弟	父	その他・男1	その他・男2	本人・女	妻	子・女1	子・女2	子・女3	孫・女	姉妹	母	子の妻	その他・女1	その他・女2
T	55	30	26	999	8	999	999	6	999	999	50	999	999	999	4	999	70	28	999	999
T+10	40	18	999	999	36	65	16	999	999	38	14	999	999	999	999	999	999	999	999	999

データ更新が非常に複雑
子男2に更新できない

図-1 既存世帯マイクロデータ構造における属性更新の例

T	男性	女性	T+10	男性	女性
1世代目		母親 : 70	1世代目	父親 : 65	
2世代目	世帯主 : 55	妻 : 50	2世代目	兄弟 : 26	世帯主 : 30
3世代目	子男2 : 26	子男1 : 30	3世代目	子男1 : 18	子男2 : 16
4世代目	孫男1 : 8	孫男2 : 6	4世代目	子女 : 14	

図-2 改良された世帯マイクロデータ構造における属性更新のイメージ

(1) エージェントベース手法

アンケート調査により取得したサンプルマイクロデータを用いるエージェントベースの推計手法は、世帯タイプの設定を、表-1の世帯構成員タイプの組み合わせではなく、世代と性別の組み合わせによる変数 g_{ms} および各世代の婚姻関係にある夫婦の組数の組み合わせによる定義で行う点異なるが、既存研究¹²⁾¹³⁾で提案した手法と同様の手法で推定可能である。

世代、性別、年齢、婚姻状態に関する推定手法の基本的な考え方は次のとおりである。

- ・世帯人数別の世帯数に対して、世帯サンプルより各世帯の構成員の質的属性である性別、続柄を決定する。
- ・十分なサンプルが得られる世帯タイプに対しては、属性（各世帯構成員の年齢）間の相関性を考慮して年齢を決定し、出現頻度の低い世帯タイプ（ C_{rare} ）については世帯サンプルと同様の年齢を与える。
- ・人口データ推計においては、周辺分布（性別5歳年齢階層別人口）に一致するように世帯構成および各世帯構成員の年齢に関して調整を行う。

・推定，データ生成，調整はすべてモンテカルロ法等を用いて確率的に行う。

本推定手法の最も特徴的な点は，連続変数である属性間の相関性の処理方法である．まずサンプルにおける m 人世帯データの属性変数 ($X_{is} = (x_{1s}, \dots, x_{ms})$) を，主成分分析を用いて無相関変数 ($P_{is} = (p_{1s}, \dots, p_{ms})$) に変換する．

$$P = aX \tag{5}$$

世帯サンプルの無相関変数 p_{is} の値に基づいて，図-3に示すように， $i = 1 \sim m$ について累積曲線を作成する．式(5)より次式が導出される．

$$X = A^{-1}P = BP \tag{6}$$

世帯を生成する際には属性 i に対して乱数 ran_i を発生させ，図-3の累積曲線より $i = 1 \sim m$ について生成世帯の各構成員に対する p_{is} を求める．生成世帯の x_i は式(6)より $i = 1 \sim m$ について求められる．同様の処理を対象地域内の全ての世帯人別世帯数に対して実行することで，世帯推定の初期データセットが作成される．この初期データセットは，周辺分布である性別5歳階級人口を満たさないため，モンテカルロ法を用いて世帯をランダム抽出し，世帯構成員の年齢が更新された新たなデータによって置き換える．このような調整処理を，周辺分布を満たすまで繰り返し実行する．本研究の初期マイクロデータ世帯データ推定に関するシステムの処理フローを図-4に示す．

就業状態等のその他の個人属性，住宅タイプ等のその他の世帯属性については，総合的属性からなる世帯マイクロデータを対象として開発された既存研究による手法¹²⁾を用いて行うものとする．

(2) オープンソースデータを用いた推計手法

国勢調査等の入手可能な統計データのみを用いて簡易的に行う手法については，既存研究¹⁵⁾で構築された手法において，世帯類型別世帯数および世帯類型別5歳階級別人口を周辺分布として用いるプロセスについて，世代と性別の組み合わせによる変数 g_{ms} および各世代の婚姻関係にある夫婦の組数の組み合わせと世帯類型を対応させることで推定可能である．

入手可能な公表データを用いた初期世帯マイクロデー

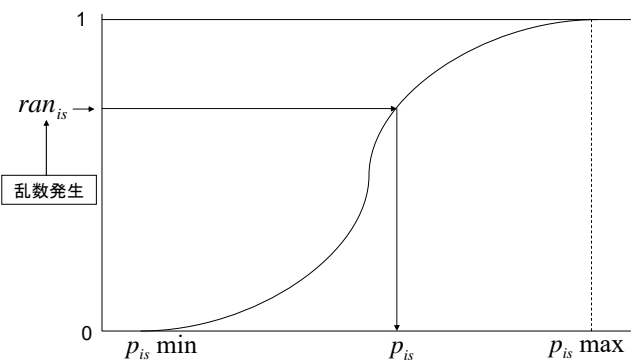


図-3 無相関変数を用いた相関の統合

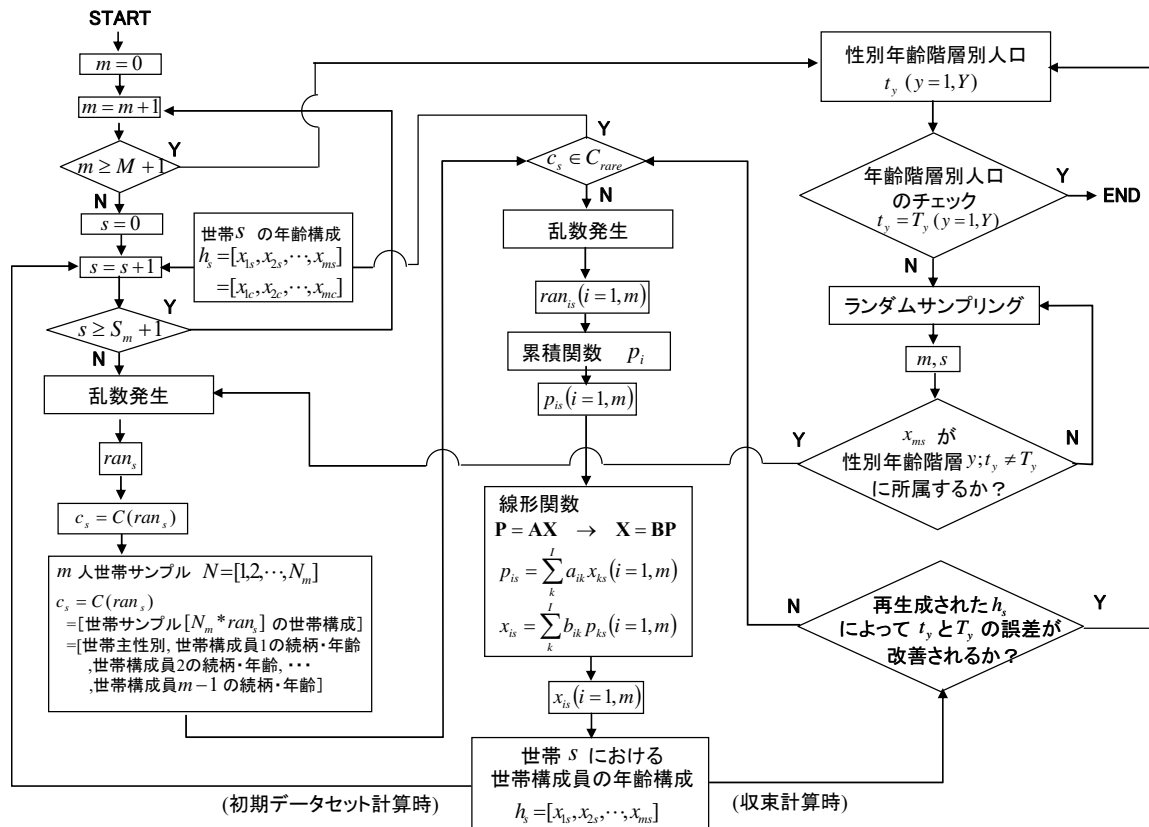


図-4 エージェントベース手法による初期世帯マイクロデータ推定システムのフロー

タの推計フローを図-2に示す。

まず、対象地域のゾーン別の年齢・性別分布をもとにゾーン別個人データを作成する。このゾーン別個人データは性別5歳階級別に対象地域内の全人口を対象とする。次に、世帯人数、世帯タイプに応じた世帯データを作成する。続いて、生成した世帯データに対し個人データから確率的に個人を割り当てる。このとき、個人データに適切な個人が存在しない場合は、繰り返し計算を行い、確率的に適切な個人を同一ゾーンの個人データから割り当てる。全ての個人を割り当てた後、出生順位や婚姻関係、就業形態、住宅タイプを付加していく。この時個人データと世帯データの属性はそれぞれ周辺分布に従って与える。個人データのリストから各世帯タイプへの割り当て順は、夫婦のみ、夫婦と子供、夫婦とひとり親、夫婦と両親、夫婦と子供とひとり親、夫婦と子供と両親、単身、女親と子供、男親と子供、その他、施設等とする。

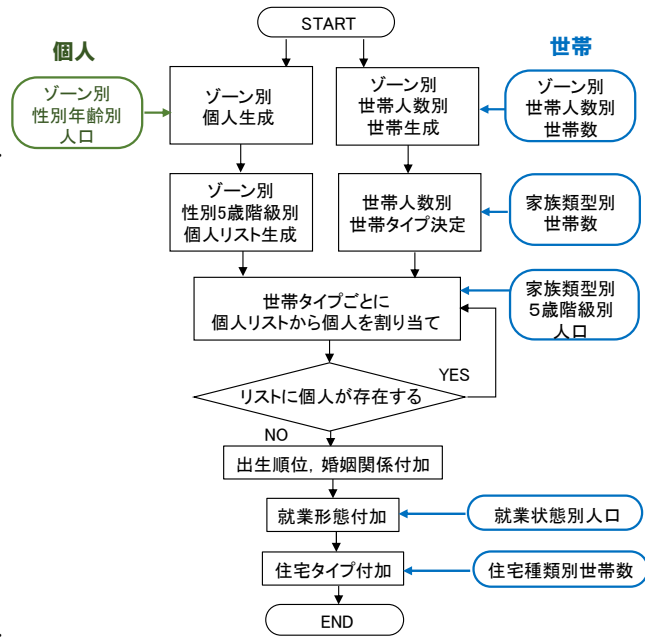


図-3 オープンソースデータを用いた推計手法による初期世帯マイクロデータ推定システムのフロー

4. 富山市における適用

(1) 対象地域の概要

本研究では、以上のエージェントベースおよびオープンソースデータによる手法を用いて、富山市を対象として初期マイクロ世帯の推計を行う。分析ゾーンは、図-3に示すような、国勢調査中ゾーンをベースとした82ゾーンを設定する。

(2) 利用データ

エージェントベースの推計手法では、平成 22 年国勢調査の人数別世帯数および性別年齢階層別人口を周辺分布として用いる。また、サンプルデータとしては、平成 23 年 12 月に富山市全域を対象に郵送配布、郵送回収形式で実施されたアンケート調査により入手された、3,864 世帯、9,747 人分のマイクロ世帯データサンプルを用いる。就業形態や住宅タイプなどその他の個人・世帯属性についてもアンケートサンプルを用いて推計を行う。

オープンソースデータによる推計手法では、上記と同様の人数別世帯数および性別年齢階層別人口に関する周辺分布に加え、国勢調査における「世帯の家族類型、世帯人員別一般世帯数」、「世帯の家族類型、世帯主の年齢、世帯主の男女別一般世帯数及び一般世帯人員」、「世帯人員、世帯主との続き柄、年齢、男女別一般世帯人員」の富山県全域のデータから、家族類型別世帯数および家族類型別 5 歳階級人口を作成して推計に用いる。また、就業状態は「職業(大分類)、従業上の地位、男女別 15 歳以上就業者数」、住宅タイプは「延べ面積、住宅の建て方、住居の種類・住宅の所有の関係別一般世帯数、一般世帯人員及び 1 世帯当たり人員」をそれぞれ国勢調査より入手して推計に持つ。

(3) 検討内容

両手法による初期世帯マイクロデータの推定結果をゾーン単位で比較し、推定精度の検証を行う。また、検証結果を踏まえ、各手法の利点と欠点、および適用可能範囲等の視点から考察を行う。

5. おわりに

本研究では、居住立地マイクロシミュレーションにおける初期世帯マイクロデータについて、既存研究による手法の将来シミュレーションへの適用を通じて明らかとなったデータ構造上の問題点を整理し、改良を行った。また、改良されたデータ構造を持つ初期世帯マイクロデータの推定について、アンケート調査により取得したサンプルマイクロデータを用いるエージェントベースの手法と、国勢調査等の入手可能な統計データのみを用いて簡易的に行う手法を検討した。その上で、富山市を対象として両手法による初期世帯マイクロデータの推定を実施するための枠組みを提示した。推定結果の比較により行う精度検証の結果、および各手法の特性の整理結果については講演時に報告予定である。

謝辞：「本研究は JSPS 科研費 17K06597 の助成を受け、実施しました。ここに感謝の意を表します。

参考文献

- 1) 福岡裕介, 大谷紀子, 杉木直, 宮本和明: 世帯マイクロデータに基づく都市政策手段の選択方法, 第 49 回土木計画学研究発表会・講演集, CD-ROM, 2014.
- 2) Wegener, M. : Overview of land-use transport models,

- Proc. of CUPUM'03, 2003.
- 3) 宮本和明, 北詰恵一, 鈴木温 : 世界における実用都市モデルの実態調査とその理論・機能と適用対象の体系化, 平成 18~19 年度科学研究費補助金(基盤研究(C), 課題番号:18560524)研究成果報告書, 2008.
 - 4) Waddell, P. : UrbanSim: Modeling urban development for land use, transportation, and environmental planning, *Journal of the American Planning Association*, Vol.68, No.3, pp.297-314, 2002.
 - 5) Salvini, P. and Miller, E.J. : ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems, *Networks and Spatial Economics*, Vol.5, No.2, pp.217-234, 2005.
 - 6) Strauch, D., Moeckel, R., Wegener, M., Gräfe, J., Mühlhans, H., Rindsfuser, G., and Beckmann, K.J. : Linking transport and land use planning: the microscopic dynamic simulation model ILUMASS, *Geodynamics*, pp.295-311, 2005.
 - 7) Eetema, D., de Jong, K., Timmermans, H., and Bakema, A. : PUMA: multi-agent modelling of urban systems, *Modelling land-use change*, Springer, pp.237-258, 2007.
 - 8) Chengxiang, Z., Chunfu, S., Jian, G., Chunjiao, D., and Hui, Z. : Agent-based joint model of residential location choice and real estate price for land use and transport model, *Computers, Environment and Urban Systems*, Vol.57, pp.93-105, 2016.
 - 9) Otani, N., Fukuoka, Y., Sugiki, N., and Miyamoto K.: Tailor-made Selection of Policy Measures for Households Based on the Detailed Attributes by Segmentation Approach with Decision Tree Analysis, *Proceedings of CUPUM 2015*, 242-Paper, Web, 2015.
 - 10) 鈴木温・杉木直・宮本和明 : 空間的マイクロシミュレーションを用いた都市内人口分布の将来予測 - 人口 40 万人規模の富山市を対象として -, 都市計画論文集, No.51-3, pp.839-846, 2016.
 - 11) Sugiki, N., Miyamoto K., Kashimura A., and Otani, N.: Household micro-simulation model considering observed family histories in a suburban new town, S.Geertman et al. (eds.), *Planning Support Science for Smarter Urban Futures*, Lecture Notes in Geoinformation and Cartography, pp.207-229, 2017.
 - 12) Sugiki, N., Vichiensan, V., Otani, N., and Miyamoto K. : Agent-Based Household Micro-Datasets: An Estimation Method Composed of Generalized Attributes with Probabilistic Distributions from Sample Data and Available Control Totals by Attribute, *Asian Transport Studies*, Vol.2, No.1, pp.3-18, 2012.
 - 13) Sugiki, N., Muranaka, T., Otani, N., and Miyamoto K. : Agent-based Estimation of Household Micro-data with Detailed Attributes for a Real City, *Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management*, pp.231-1 - 231-18, 2015.
 - 14) 吉田慎也・鈴木温 : オープンデータを用いた初期世帯マイクロデータの生成に関する研究, 第 55 回土木計画学研究発表会・講演集, CD-ROM, 2017.
 - 15) 戸田賢寛・鈴木温 : オープンデータを用いた初期世帯マイクロデータの生成に関するシステム構築, 平成 29 年度土木学会中部支部研究発表会講演概要集, CD-ROM, 2018.

IMPROVEMENT OF HOUSEHOLD MICRO DATA STRUCTURE AND ACCURACY VERIFICATION BETWEEN ESTIMATION METHODS

Nao SUGIKI, Atsushi SUZUKI and Kazuaki MIYAMOTO