

2019年6月

東京都における流動人口データの有効性の検証

菅 愛子*
飯島 信也**
兵頭 大史***
藤原 直哉****
水野 貴之*****
松本 裕介*
武藤 杏里**
瞿 雪吟*
伊藤 武真*
松井 伸司*****
五十嵐 盛仁*****
上田 聖*****

- * 慶応義塾大学 大学院経営管理研究科
- ** 独立行政法人 統計センター
- *** 株式会社 Agoop
- **** 東北大学 大学院情報科学研究科
- ***** 国立情報学研究所
- ***** 総務省 統計委員会担当室

総務省 統計委員会担当室

〒162-8668 東京都新宿区若松町 19 番 1 号

総務省統計委員会担当室ワーキングペーパーは、統計委員会担当室スタッフ又はスタッフと外部研究者との共同による調査・研究の成果をまとめたもので、公的統計の整備に係る各種施策に役立てることを企図としている。ただし、ワーキングペーパーの内容や意見は、執筆者個人に属し、総務省の公式見解を示すものではない。

東京都における流動人口データの有効性の検証

要旨

本ワーキングペーパーは、東京都において GPS 方式で収集したメッシュ型の流動人口データ（GPS データ）を国勢調査や基地局方式で収集したデータ（基地局データ）と比較し、統計的に分析した結果をまとめたものである。この結果から GPS データは、国勢調査や基地局データと相関があり、一定レベルの信頼性が確保されているとともに、利活用にあたり国勢調査を補完しうるものとして有効であることを示すことができた。更に以下の GPS データの特徴に留意して使用することで、より効果的な活用が可能であると考えられる。

- 国勢調査 5 年間における差率の標準偏差を基準とした評価から、GPS データは流動人口 2,000 人以上のメッシュ（500m 四方）において一定レベルの信頼性が確保されており、そのエリアは東京都 23 区を中心に居住地域の面積の約半分に該当する。
- 時間帯別の GPS データは、国勢調査では把握対象外である通勤・通学以外の勤務中の移動や余暇・消費活動による人の動きを、ビジネス街や住宅地などの地域の特性と整合する形で捉えている。
- GPS データは解像度が高いため、大量の人の動線のハブとなる都心ターミナル駅や海岸沿い等、隣接するメッシュ間の人口差が大きいエリアにおいて、より強みを発揮できる。
- GPS データはアプリユーザーの属性に依存することから、標本の偏りが発生しやすいことを理解した上で、利活用を進めていくことが肝要である。

キーワード：ビッグデータ、公的統計、GPS、基地局、国勢調査、メッシュ、相関係数、差率、標準偏差、流動人口

本稿は、総務省において開催している「ビッグデータ等の利活用推進に関する産官学協議のための連携会議」配下のワーキンググループとして立ち上げた「メッシュ型流動人口検証 WG」で行った検証の結果をまとめたものである。本稿の作成にあたっては、高橋大志（慶應義塾大学）、櫻川幸恵（総務省）の各氏から有益な助言を頂いた。また上述の会議において、構成員の皆様より有意義な意見をいただいたことを、ここに記して感謝したい。ただし、本稿の内容と意見は筆者ら個人に属し、所属組織の公式見解を示すものではない。また、ありうべき誤りはすべて筆者ら個人に属する。

目次

1. はじめに
2. 検証データと検証方法
 - 2.1 検証データ (GPS データ)
 - 2.2 国勢調査との比較方法
 - (1) 比較データ (国勢調査)
 - (2) 国勢調査常住地人口との比較方法
 - (3) 国勢調査昼間人口との比較方法 (市区町村単位)
 - 2.3 基地局データとの比較方法
 - (1) 比較データ (基地局データ)
 - (2) 比較方法
3. 検証結果
 - 3.1 国勢調査との比較結果
 - (1) 国勢調査常住地人口との比較結果
 - (2) 国勢調査昼間人口との比較結果 (市区町村単位)
 - 3.2 基地局データとの比較結果
 - (1) 分析の前提
 - (2) 相関係数と差率の分析
 - (3) エリア面からの分析
 - (4) メッシュサイズ拡大効果の分析
 - (5) 年齢階級別の相関分析
 - (6) 人口規模別の差率分析
4. 考察
 - 4.1 国勢調査との比較
 - 4.2 基地局データとの比較
 - (1) 標本の偏りや作成方式に関する考察
 - (2) メッシュサイズに関する考察
 - (3) 年齢別考察
 - (4) 人口規模別考察
 - (5) エリア面からの考察
5. まとめ

(参考) 流動人口データによる市区町村別人口変動の分析

1. はじめに

国・地方公共団体における政策立案や民間における商品戦略立案等のマーケティングの基礎情報として、地理的な人口分布を示すデータは非常に有用であり、その中で公的統計として提供しているものの代表が国勢調査¹である。国勢調査結果の中に、常住地人口を地域メッシュ²や町丁字別の小地域集計として統計化したものがあり、そのデータは5年に1度の調査が行われる度に更新され、詳細な結果が公表されるまでに調査の期日から約2年が必要とされている。この国勢調査結果をより有効なものとするには、人口分布のタイムリーな把握が課題であるが、現在それを補う民間データとして、携帯端末の基地局情報やGPS³アプリから取得した位置情報に基づき推計された流動人口⁴のデータが提供されている。しかし、これらに対するデータの信頼性や利活用にあたっての有効性について、公的に評価した事例はまだ少ない。

「公的統計の整備に関する基本的な計画⁵」（平成30年3月6日閣議決定）では「行政記録情報⁶等及び民間企業等が保有するビッグデータ⁷等を統計の作成に活用することは、統計調査における報告者の負担軽減のみならず、正確で効率的な統計の作成にも寄与することから、各府省における積極的な活用が必要」とされている。この趣旨を踏まえ、携帯端末を活用した情報について、公的統計と同様に広く利活用していく仕組みを構築していくことは有効なことであり、当該データの信頼性や有効性の評価を行うことはそのための必要なプロセスであると考えられる。

本稿は、民間から提供される携帯端末情報のうち、GPSによる流動人口データに焦点を当て、信頼性、有効性の評価の第一歩として、東京都における国勢調査の人口や基地局情報に基づく流動人口と比較した結果について報告するものである。

¹ 総務省統計局から公表される。

² 緯度・経度に基づき地域を隙間なく網の目（メッシュ）の区域に分けたもの。

³ 全地球測位システム（Global Positioning System）。

⁴ 一時的にある場所に滞在している人口。

⁵ 統計法（平成19年法律第53号）に基づき、公的統計の整備に関する施策の総合的かつ計画的な推進を図るため、総務省より策定された計画。

⁶ 行政機関の職員が職務上作成し、又は取得した情報であって、当該行政機関の職員が組織的に利用するものとして、当該行政機関が保有しているもののうち、行政文書に記録されているもの。

⁷ 一般的なデータ管理・処理ソフトウェアで扱うことが困難なほど巨大で複雑なデータの集合であり、ICTの進展により生成・収集・蓄積等が可能・容易になる多種多量なデータ。

2. 検証データと検証方法

携帯端末に搭載されたアプリの GPS 位置情報を基にメッシュ統計化した人口データ（GPS データ）を検証対象のデータとし、公的統計である国勢調査や、膨大な数の携帯端末とつながる基地局の位置情報を基にメッシュ統計化したデータ（基地局データ）など高い信頼性が確保されていると想定されるデータと比較することにより、GPS データの信頼性の検証を行った。

なお、基地局データは秘匿処理として人口の少ないメッシュが削除されているため、それと比較するときには、国勢調査と GPS データも基地局データと同じ基準で少人口メッシュを削除している。またメッシュ比較において相手方に対応するメッシュが存在しなかった場合は、該当データを削除した上で統計数値を算出している。

2.1 検証データ（GPS データ）

株式会社 Agoop よりメッシュ型流動人口データ⁸を入手し、検証データとして使用した。これはスマートフォン向けアプリケーションにおいて許諾を得たユーザーから GPS 位置情報を取得し、地域メッシュ単位に推計した人口を収録したものである。この推計人口は、国勢調査の市区町村人口を基に1台の携帯端末アプリのGPSデータ（「1セッション」という。）あたりの換算係数（1セッションあたり何人分の代表となっているかを示す係数。）を日別に算出⁹し、各時点・各地点のセッション数にその換算係数を乗じて算出されている。今回はこの方法により時間帯別、100mメッシュ単位に集計された全国版の推計人口が、年月別、休日／平日別に平均化されて提供されている。

本検証においては、2015年、2016年、2017年10月の東京都¹⁰のデータを500mメッシュ単位に集計した時間帯別のGPSデータ（図表1参照。）と、それを24時間平均したGPSデータ（以降、単に「GPSデータ」と表す。図表2参照。）の2種類を検証対象のデータとして使用した。

なお、GPS方式での地理的な誤差は約10mであり、高い解像度を保有している一方、アプリのGPSから取得したデータは、一般的にユーザー数が少なくなりがちであり、その少ないユーザーの属性に依存することから、標本の偏りが発生しやすいという特性があることに留意が必要である。

図表1：時間帯別GPSデータ（東京都分）

	対象年	休日/平日	メッシュ数	平均人口	標準偏差	最大人口
1	2015	休日	130,134	2,530	3,092	73,992
2	2015	平日	131,621	2,796	6,084	328,734

(*) メッシュの最少人口は1人。

図表2：GPSデータ（24時間平均、東京都分）

	対象年	休日/平日	メッシュ数	平均人口	標準偏差	最大人口
1	2015	休日	5,552	2,470	2,912	45,163
2	2015	平日	5,479	2,799	5,161	175,029
3	2016	休日	5,415	2,545	2,903	43,299
4	2016	平日	5,291	2,951	6,570	334,115
5	2017	休日	5,302	2,666	3,090	44,541
6	2017	平日	5,236	3,116	7,713	418,064

(*) 少人口メッシュを削除。

⁸ 全国で21万セッションのアプリユーザー情報から作成されており、データには個人情報情報は含まれていない。

⁹ 2015年、2016年のデータは2010年の国勢調査、2017年のデータは2015年の国勢調査を使用して換算係数を算出している。

¹⁰ 隣県と接しているメッシュには隣県分の人口も含んでいる。

2.2 国勢調査との比較方法

GPS データと基地局データそれぞれについて、国勢調査との相関係数と差率（計算式の詳細は後述）の標準偏差（相違の度合い）を計算し比較することで、GPS データの信頼性に関しての分析を行った。なお、常住地人口を表す国勢調査と、その時々の流動人口を推計し 24 時間平均した基地局データや GPS データは、異なる状態を表したデータであることを前提としているが、より差異が小さい条件で比較を行うために、基地局データや GPS データは昼間帯の人口変動が小さいと想定される休日のデータを使用することとした。

また、時間帯別 GPS データを市区町村毎に合算したデータと、国勢調査の市区町村別昼間データとの相関の確認を行うとともに、時間経過に伴う人口変動の分析を行った。

(1) 比較データ（国勢調査）

国勢調査常住地人口との比較において、2010 年、2015 年 10 月の常住地人口（夜間の実在人口に近い値を示すと考えられる）を 500m メッシュ単位に集計したデータ¹¹（以降、「国勢データ」と表す。図表 3 参照。）を使用した。

また昼間人口の比較においては、通勤・通学者の人口を常住地ではなく、通勤・通学先の市区町村に算入したデータ（以降、「国勢昼間データ」と表す。図表 4 参照。）を使用した。

図表 3：国勢データ（東京都分）

	対象年	休日/平日	メッシュ数	平均人口	標準偏差	最大人口
1	2015	—	5,430	2,489	2,213	11,605
2	2010	—	5,458	2,407	2,041	10,707

(*) メッシュの最少人口は 1 人。

図表 4：国勢昼間データ（東京都分）

	対象年	区人口 (区数23)	市人口 (市数26)	町村人口 (町村数13)
1	2015	12,033,592	3,798,280	88,533
2	2010	11,711,537	3,776,318	88,275

¹¹ 政府統計の総合窓口（e-Stat）より取得。

(2) 国勢調査常住地人口との比較方法

メッシュ単位の国勢データとの比較のために、以下の相関分析と差率分析を行った。

<相関分析、外れ値分析>

メッシュコード(i)に対応する2015年の国勢データ(x_i)と2015年休日の基地局データ(y_i)を散布図に描くとともに相関係数を算出した。更に、メッシュコード(j)に対応する2015年の国勢データ(x_j)と2015年休日のGPSデータ(z_j)でも同様のことを行い、その上で、GPSデータと基地局データそれぞれについて、国勢データとの散布図、相関係数を比較した。

更にケーススタディとして、GPSデータと国勢データとの相関から大きく外れているデータ(「外れ値」と表す。詳細は散布図に示す。)を15個抽出¹²し、地図上にプロットして、GISソフトを用いたビジュアル的観測によりエリア面からの分析¹³を行った。

<差率分析>

メッシュコード(i)に対応する2015年の国勢データ(x_i)と2015年休日の基地局データ(y_i)に関して、差率(i)を以下の式で算出し、更にその平均と標準偏差を算出した。

$$\text{基地局データの差率： } R_{(\text{基地局}, i)} = \frac{y_i - x_i}{y_i + x_i}$$

同様にメッシュコード(j)に対応する2015年の国勢データ(x_j)と2015年休日のGPSデータ(z_j)に関して、差率(j)を以下の式で算出し、更にその平均と標準偏差を算出した。

$$\text{GPSデータの差率： } R_{(\text{GPS}, j)} = \frac{z_j - x_j}{z_j + x_j}$$

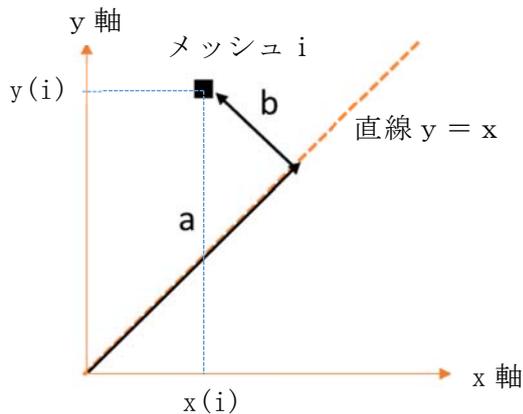
これらGPSデータと基地局データそれぞれについて、上記で算出した国勢調査との差率の平均及び標準偏差の比較を行った。

なお、この式で計算した「差率」という値の幾何学上の意味を図表5に示す。

¹² 具体的には回帰直線から残差を求めて、大きいものから15個抽出している。

¹³ 地図上にプロットして、エリア面から分析するGISツールとして、総務省統計局が提供するjSTAT MAPを使用している <https://www.e-stat.go.jp/gis>。

図表 5 : 「差率」の幾何学上の意味



$$\text{差率} : R_{(i)} = \frac{y_i - x_i}{y_i + x_i} = \frac{b}{a}$$

(3) 国勢調査昼間人口との比較方法（市区町村単位）

市区町村単位の国勢調査昼間人口との比較のために、以下の相関分析と時間経過に伴う人口変動分析を行った。

<相関分析>

メッシュ単位の 2015 平日の GPS データを市区町村単位に集計し、市区町村 (i) 毎に、時間帯 (t) に対応する GPS データ (z_{it}) と国勢昼間データ (x_i) との相関係数を算出し、時間帯別に相関係数がどう変化するか、すなわち、どの時間帯の GPS データが国勢昼間データとのフィッティングがよいかについて確認を行った。

なお GPS データの集計にあたり、複数の市区町村に跨がるメッシュに関しては、そのメッシュに含まれる市区町村数で人口を除算した結果を該当市区町村に算入した。

<時間経過に伴う人口変動分析>

市区町村 (i) 別に、時間帯 (t) を横軸、GPS データ (z_{it}) を縦軸にしてグラフ化し、2010 年の国勢調査常住地人口及び昼間人口と比較することで時間経過に伴う GPS データの人口変動の分析を行った。

2.3 基地局データとの比較方法

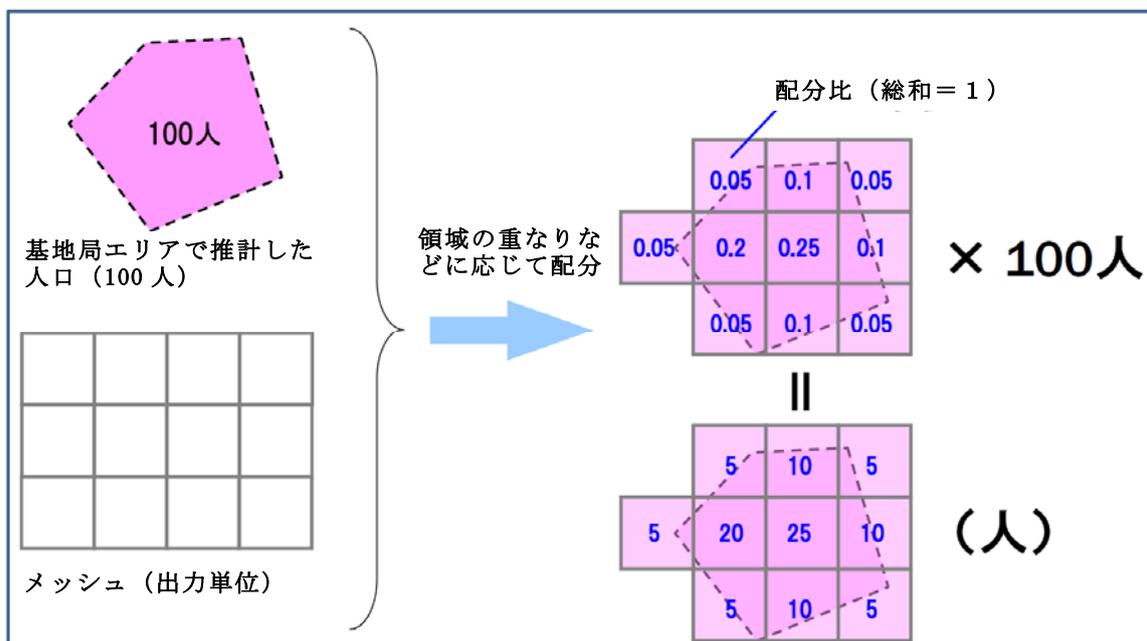
7,600万台という膨大な数の携帯端末のデータから人口推計を行っている基地局データは、精度が確保された流動人口が推計されているという前提のもと、GPSデータと基地局データとの相関や偏差を分析することにより、GPSデータの信頼性の分析を行った。分析にあたり、公的な統計として活用されている国勢調査の2010年と2015年との相関と偏差のレベルが一般的に許容されているものと仮定し、それと比較することにより、GPSデータの信頼性の評価を行った。

更に、メッシュサイズを拡大したときの効果、年齢階級別のデータのフィッティング、人口規模別の相関と偏差の分析及び地図上での可視化など、多面的に検証することで、その特徴を明らかにした。

(1) 比較データ（基地局データ）

株式会社NTTドコモが提供しているモバイル空間統計¹⁴を入手し、比較データとして使用した。これはNTTドコモ社が保有する携帯電話ネットワークの仕組みを利用して、携帯端末の台数を基地局毎に把握し基地局がカバーするエリアの実在人口を、普及率を加味し重み付け分割して地域メッシュ単位に推計したものである（図表6参照）。

図表6：基地局における推計人口のメッシュへの変換イメージ¹⁵



¹⁴ モバイル空間統計は株式会社ドコモの商標登録である。

入手したデータには個人情報含まれておらず、秘匿処理により少人数のメッシュは削除されている。

¹⁵ 独立行政法人統計センター、株式会社NTTドコモ「官庁統計とモバイル空間統計に基づく新たな統計の創出に関する共同研究」を基に作成。

この流動人口は、住民基本台帳の性別・年齢階級（5歳）別の市区町村人口を基準値として1端末あたりの換算係数を日別・時間帯別に算出し、端末数にその換算係数を乗じて推計されている。今回は、性別、年齢階級（10歳）別、年月別、休日／平日別に、500mメッシュ単位に集計された24時間平均の人口データが提供されており（以降、「基地局データ」と表す。）、本検証では、2015年、2016年、2017年10月の東京都¹⁶のデータを使用した。

図表7：基地局データ（東京都分）

	対象年	休日/平日	メッシュ数	平均人口	標準偏差	最大人口
1	2015	休日	6,383	1,837	1,886	32,105
2	2015	平日	6,393	1,945	2,328	28,281
3	2016	休日	6,385	1,846	1,908	31,617
4	2016	平日	6,361	1,972	2,404	27,970
5	2017	休日	6,355	1,870	1,918	31,697
6	2017	平日	6,360	1,981	2,421	28,715

(*) 少人口メッシュは秘匿処理により削除されている。対象データは15歳以上80歳未満。

図表8：2017平日の年齢階級別の基地局データ（東京都分）

	10代	20代	30代	40代	50代	60代	70代
メッシュ数	5,869	6,005	5,929	6,084	5,982	6,206	6,122
平均人口	94	318	399	443	323	284	226
標準偏差	83	492	586	576	378	266	191

(*) 少人口メッシュは秘匿処理により削除されている。10代の対象は15歳以上20歳未満。

¹⁶ 隣県と接しているメッシュには隣県分の人口も含んでいる。

(2) 比較方法

メッシュ単位の基地局情報との比較のために、以下の相関分析と差率分析を行った。

<相関分析、外れ値分析>

i. 500m メッシュにおける相関分析

年別、休日／平日別に、メッシュコード(i)に対応する基地局データ(y_i)と GPS データ(z_i)を散布図に描くとともに相関係数を算出した。

また、国勢調査のメッシュデータが更新されるのは5年毎であることを踏まえ、メッシュコード(j)に対応する2010年の国勢データ(x_{2010j})と2015年の国勢データ(x_{2015j})との相関係数をデータの相違が許容される基準値とし、前述の相関係数をこの数値と比較し分析を行った。

更にケーススタディとして、GPS データと基地局データの相関から大きく外れているデータ(外れ値)を15個抽出し、地図上にプロットして、GISソフトを用いたビジュアル的観測によりエリア面からの分析を行った。

ii. メッシュサイズ拡大による効果

2017 休日／平日の基地局データと GPS データに関して、500m メッシュのデータを1km メッシュ単位に集計し、1km メッシュでの相関係数を500m メッシュのそれと比較し分析を行った。

iii. 年齢階級別人口との相関

本検証で使用している GPS データは、特定のアプリから情報を得ており、かつ年齢に関する情報を保有していないため、特定の年齢階級に偏っている可能性がある。これを検証するため2017 休日／平日のデータに関して、メッシュコード(i)と年齢階級(a)に対応する基地局データ(y_{ia})と GPS データ(z_i)との相関係数を算出し、基地局データの特定の年齢階級に対し、GPS データがどのような相関を示すかの分析を行った。この数値を評価するため、メッシュコード(j)と年齢階級(a)に対応する基地局データ(y_{ja})と全基地局データ(y_j)との相関係数を算出し、比較対象として使用した。

<差率分析>

年別、休日／平日別に、メッシュコード(i)に対応する基地局データ(y_i)とGPSデータ(z_i)に関して、差率(i)を以下の式で算出し、更にその平均と標準偏差を算出した。

$$\text{差率} : R_{(i)} = \frac{z_i - y_i}{z_i + y_i}$$

また、メッシュコード(j)に対応する2010年国勢データ(x_j)と2015年国勢データ($x_{2015 j}$)との差率(j)を以下の式で算出し、更にその平均と標準偏差を算出¹⁷した。

$$\text{差率} : R_{(j)} = \frac{x_{2015 j} - x_{2010 j}}{x_{2015 j} + x_{2010 j}}$$

これは相関係数で行ったケースと同様に、データの相違が許容される基準として算出したものであり、GPSデータと基地局データとの差率の標準偏差をこの数値と比較して分析を行った。

更に、ここで算出したGPSデータと基地局データとの差率を地図上にプロットして、GISソフトを用いたビジュアル的観測によりエリア面から分析を行った。

<人口階級別メッシュ数分析>

2017平日のGPSデータ、基地局データそれぞれに関して、人口階級別のメッシュ数をカウントし、横軸を人口階級、縦軸をメッシュ数としたヒストグラム(形態は折れ線グラフ)を作成し、人口階級に対するメッシュ数の分布の相違の確認を行った。

<人口規模別の差率分析>

<差率分析>で算出した2017休日／平日の基地局データ(y_i)とGPSデータ(z_i)との差率(i)に関して、メッシュコード毎にGPSデータの人口を横軸、差率を縦軸にプロットした散布図を作成した。

更に、このデータをGPSデータの人口でソートしたうえで、大きい方から100メッシュ毎にグループ化し、グループ毎にGPSデータの平均人口と差率の標準偏差を算出し、前者を横軸、後者を縦軸にプロットし、人口規模の拡大による差率の標準偏差の変化を確認した。

¹⁷ 2015国勢データと2010国勢データとの比較においては少人数のメッシュの削除は行っていない。

3. 検証結果

3.1 国勢調査との比較結果

(1) 国勢調査常住人口との比較結果

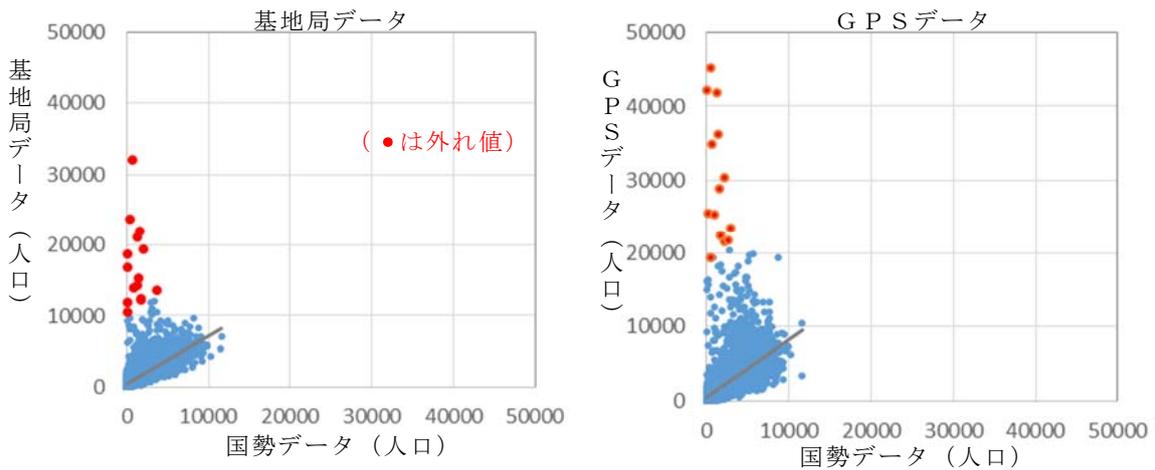
2015 休日の基地局データと GPS データそれぞれに関して、国勢データとの統計値、散布図、差率ヒストグラム及び外れ値マップを、それぞれ図表 9、図表 10、図表 11、図表 12 に示す。

図表 9：2015 休日データの国勢データとの統計値

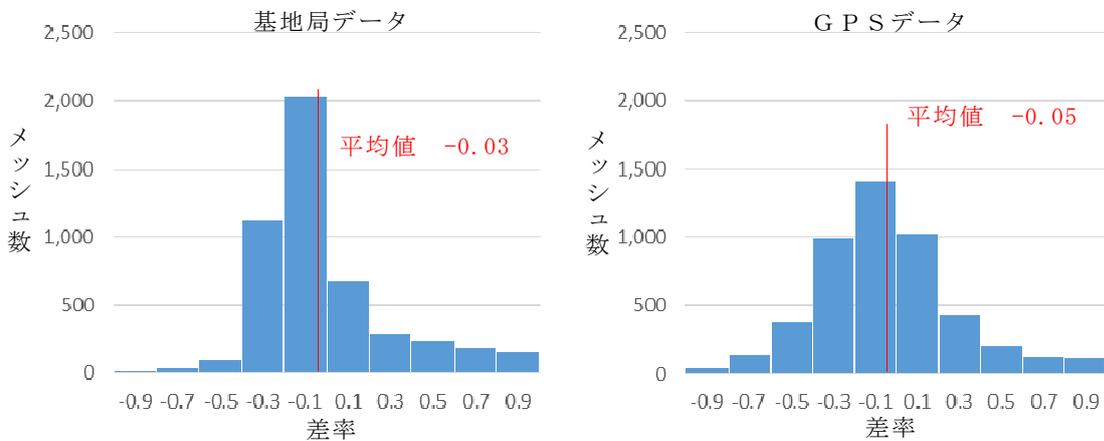
	比較データ	メッシュ数	相関係数	差率平均	差率標準偏差
1	基地局データ	5,011	0.71	-0.03	0.32
2	GPS データ	4,833	0.52	-0.05	0.33

(*) 少人口メッシュ、相手なしメッシュを削除。

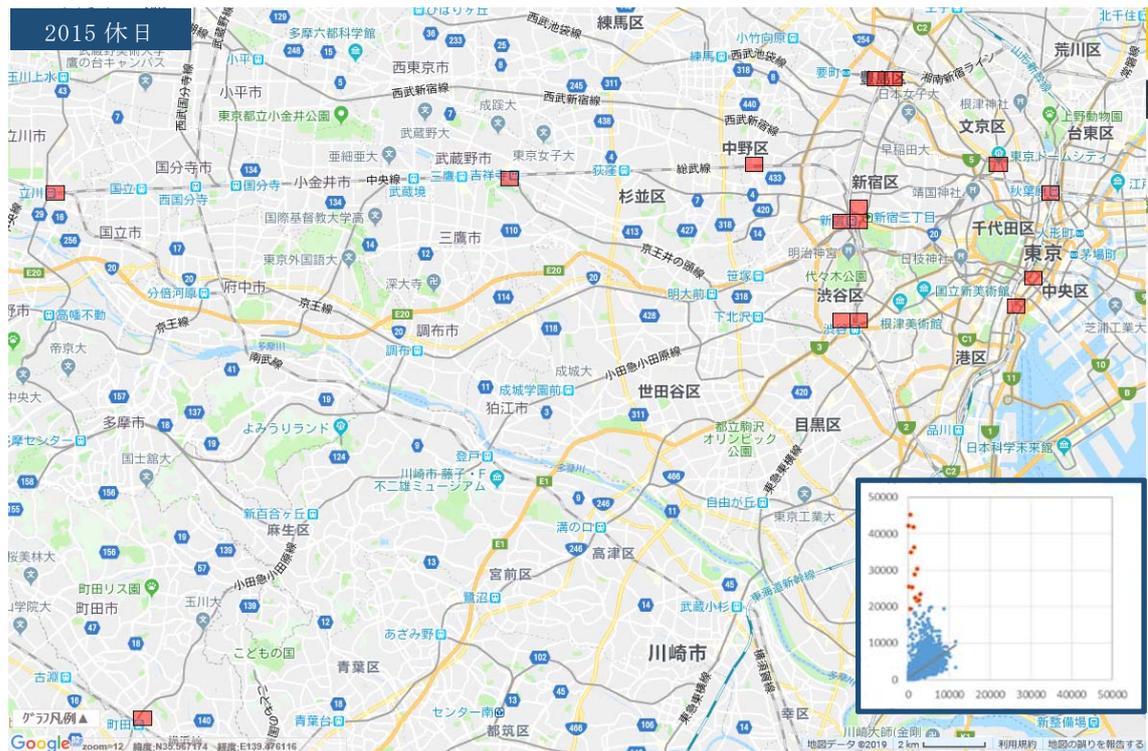
図表 10：2015 休日データと国勢データとの散布図



図表 11：2015 休日データと国勢データとの差率ヒストグラム



図表 12 : 2015 休日の GPS データと国勢データとの外れ値マップ



図表 9、図表 10、図表 11 で、GPS データと国勢データは相関係数が 0.52 と正の相関が示されており、差率の標準偏差は 0.33 となった。図表 10 で国勢データの人口が約 3000 人未満のメッシュで GPS データの人口が 20,000 人以上のメッシュが外れ値として現れており、これは図表 12 で都心ターミナル駅周辺に加えて、吉祥寺、立川、町田などの鉄道沿線ターミナル駅に該当していた。

GPS データと基地局データに関して国勢データとの関係を比較してみると、相関は GPS データの方が弱いですが、差率の偏差は同レベルであった。

(2) 国勢調査昼間人口との比較結果（市区町村単位）

詳細は文末の「(参考) 流動人口データによる市区町村別人口変動の分析」参照

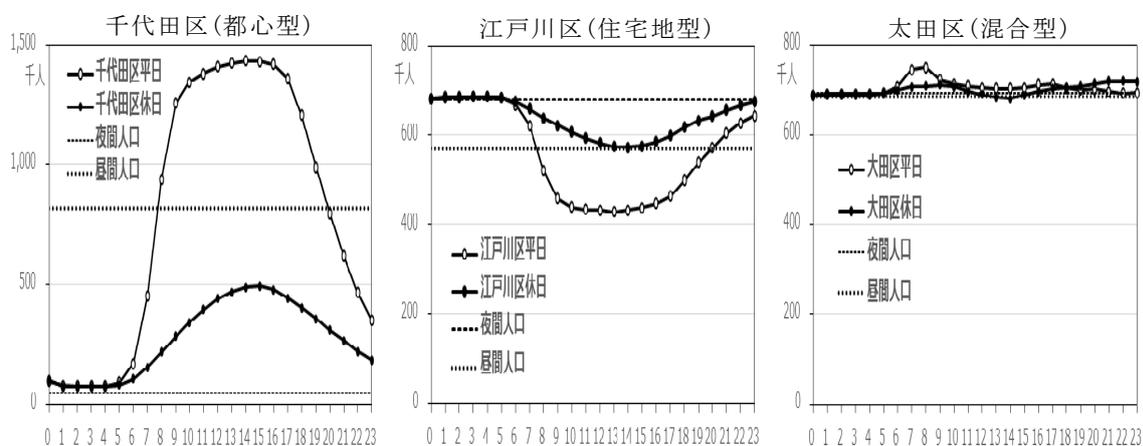
2015 平日の市区町村単位での時間帯別 GPS データと国勢昼間データとの相関係数を図表 13 に示す。GPS データは 10 時から 16 時の間の全ての時間帯において国勢昼間データと強い正の相関が示されており、時間帯による相関係数の違いはなかった。

図表 13：2015 平日の時間帯別 GPS データと国勢昼間データとの相関係数

	10時	11時	12時	13時	14時	15時	16時
相関係数	0.92	0.92	0.91	0.91	0.91	0.91	0.91

2015 休日／平日の代表区における時間経過に伴う人口変動グラフを図表 14 に示す。千代田区や中央区など昼間帯に人口が増加する「都心型」、江戸川区や狛江市など昼間帯に人口が減少する「住宅地型」、大田区や武蔵野市など昼間帯と夜間帯で大きな変動がない「混合型」及び人口変動の傾向が定まらない「その他」の地域に分類できる。夜間帯については、全ての地域で GPS データの人口は国勢調査の常住人口と同じレベルとなっていた。一方、昼間帯においては、「都心型」の市区町村では GPS データの人口が国勢調査昼間人口を上回り、「住宅地型」ではそれを下回っており、いずれも国勢調査昼間人口のレベルを超えて変動していた。

図表 14：2015 休日／平日の時間経過に伴う GPS データ人口変動グラフ
(横軸：時間帯、縦軸：流動人口)



3.2 基地局データとの比較結果

(1) 分析の前提

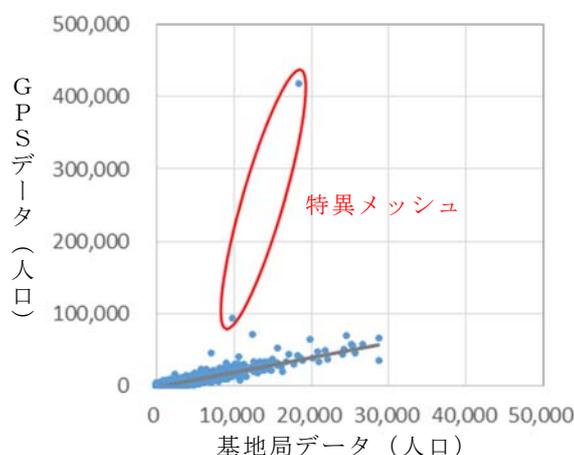
2015～2017 休日／平日の GPS データと基地局データとの統計値を図表 15 に、2017 平日の GPS データと基地局データとの散布図を図表 16 に示す。GPS データと基地局データは正の相関があり、特に休日では 0.9 程度の高い相関係数となっている一方、平日では相関を非常に大きく外れるメッシュ（以降、「特異メッシュ」と表す）が 2 個存在し、相関係数の低下に大きな影響を与えていた。この特異メッシュを除くことで、相関係数は 2015 年平日は 0.80→0.87、2016 年平日は 0.69→0.90、2017 年平日は 0.65→0.89 となり、いずれも相関係数が大幅に上昇した。これは GPS データのアプリユーザーの属性に起因する標本の偏りであると判断できたため、以降、平日 GPS データを使用した検証については、全て特異メッシュを除いて実施している。

図表 15：GPS データと基地局データとの統計値

	対象データ	メッシュ数	相関係数	差率平均	差率標準偏差
1	2015 休日	5,331	0.86	-0.03	0.33
2	2015 平日	5,283	0.87 (0.80)	-0.05	0.34
3	2016 休日	5,258	0.89	-0.04	0.34
4	2016 平日	5,163	0.90 (0.69)	-0.05	0.33
5	2017 休日	5,172	0.88	-0.04	0.35
6	2017 平日	5,129	0.89 (0.65)	-0.06	0.35

(*) 少人口メッシュ、相手なしメッシュを削除。相関係数の () 内は特異メッシュ削除前の数値。

図表 16：2017 平日の GPS データと基地局データとの散布図



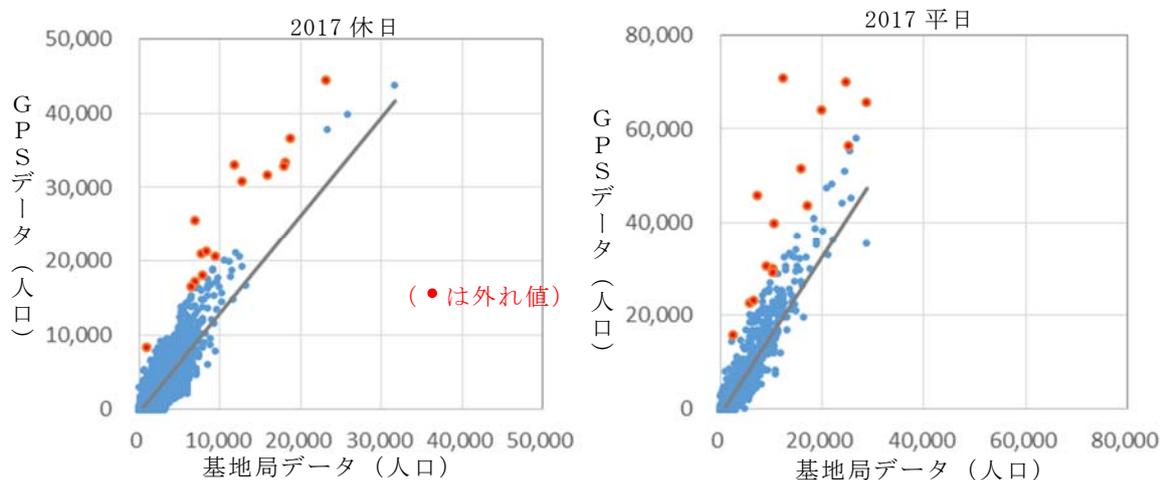
(2) 相関係数と差率の分析

2017 休日／平日の GPS データと基地局データとの散布図、差率ヒストグラムを、それぞれ図表 17、図表 18 に示す。

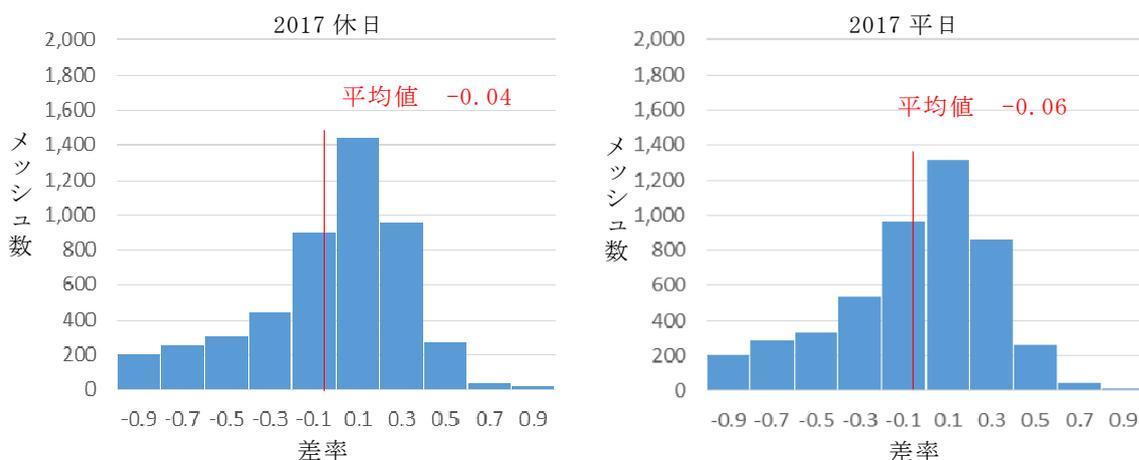
図表 15 も合わせてみると、GPS データと基地局データは 2015～2017 休日／平日の全てのケースで、相関係数が 0.90 程度と強い正の相関が示されており、差率の標準偏差は約 0.35 であった。

休日と平日の比較で、相関係数や差率の標準偏差には大きな違いはなかったが、平日の方が人口の多いメッシュを含んでおり、人口の集中度合いが高かった（図表 17）。また年別（2015 年、2016 年、2017 年）の比較では相関係数を含む全ての統計値において大きな違いはなかった。

図表 17：GPS データと基地局データとの散布図



図表 18：GPS データと基地局データとの差率ヒストグラム



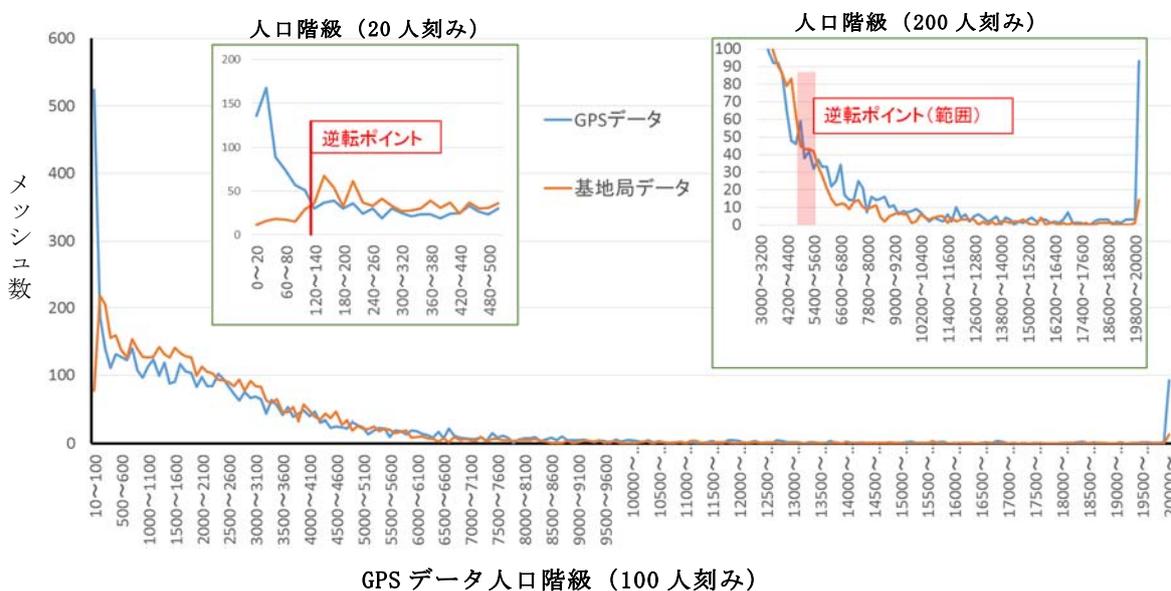
データの信頼性を判断するための基準とした国勢調査 2010 年と 2015 年との相関係数等の統計値を図表 19 に示す。この相関係数は 0.99、差率の標準偏差は 0.22 であり、これと比べると GPS データと基地局データとの相関係数は低く、差率の標準偏差は大きかった。

図表 19：国勢調査間隔 5 年間の比較における統計値

比較名	メッシュ数	相関係数	差率平均	差率標準偏差
2015 と 2010 との比較	5,546	0.99	-0.01	0.22

メッシュ人口の大きさに対するメッシュ数の分布を明らかにするために、2017 平日の GPS データ、基地局データそれぞれにおけるメッシュ人口階級別のメッシュ数ヒストグラム（折れ線）を図表 20 に示す。メッシュ人口が約 100 人未満と約 5000 人以上では GPS データの方が基地局データよりもメッシュ数が多くなり、人口が 100 人～5000 人のメッシュでは逆に基地局データの方が多くなる傾向があることがわかった。

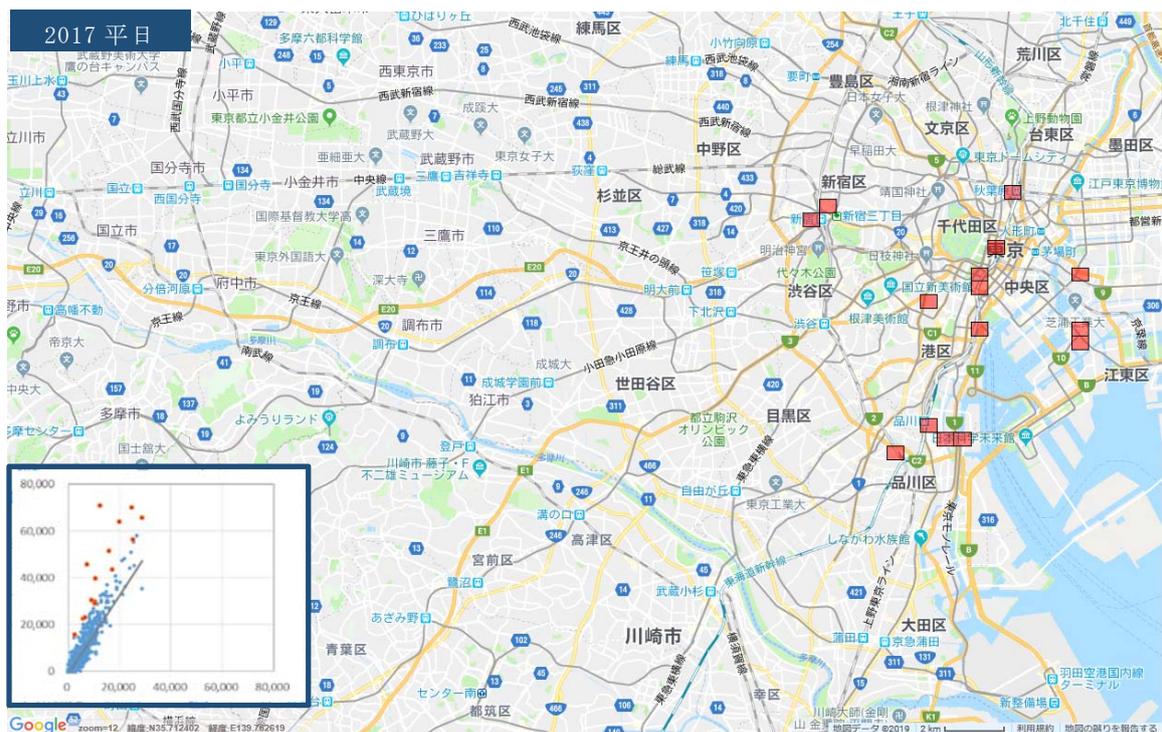
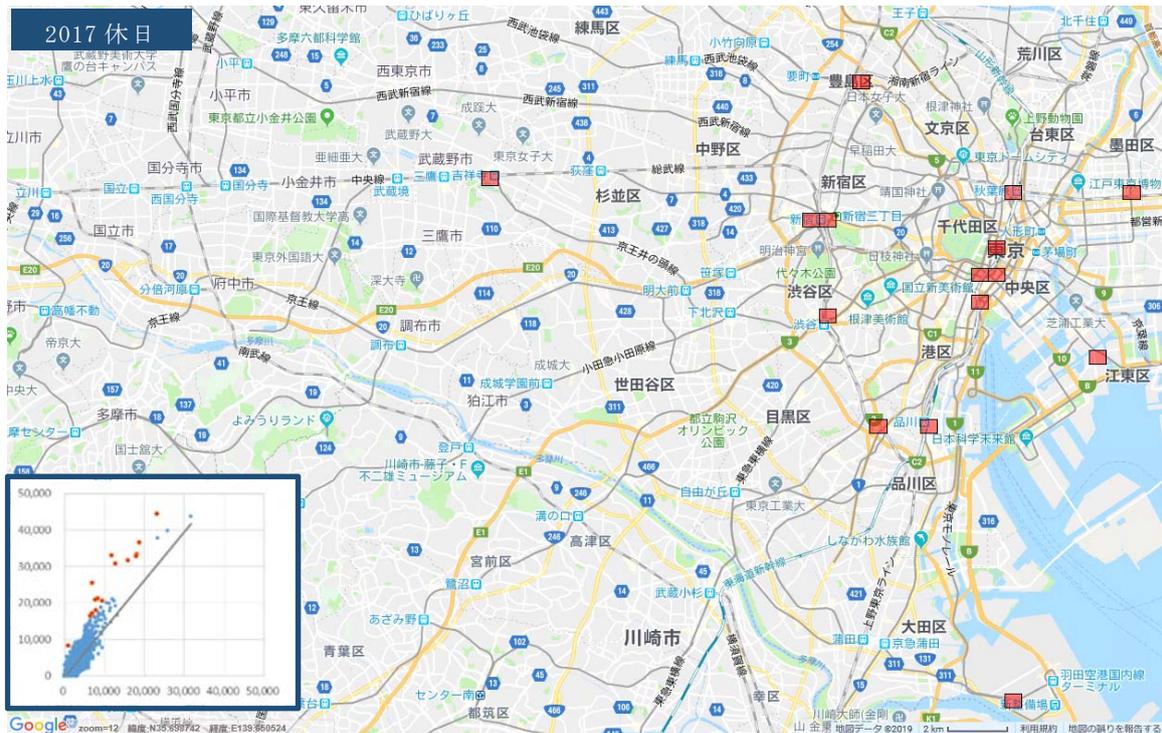
図表 20：2017 平日の人口階級別のメッシュ数ヒストグラム（折れ線）



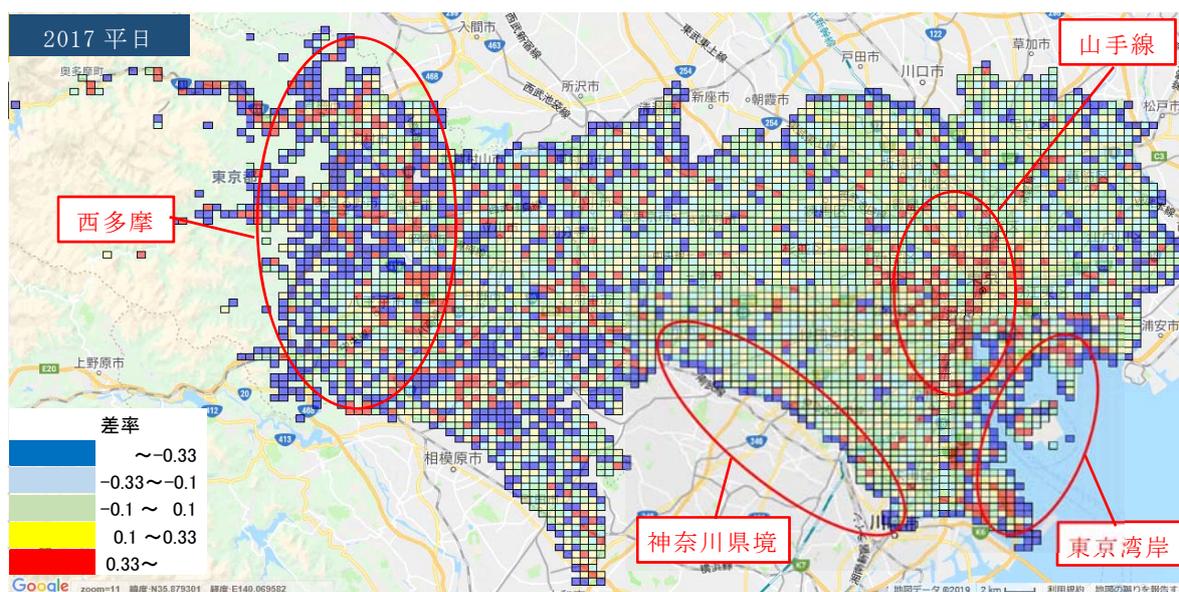
(3) エリア面からの分析

2017 休日／平日の GPS データと基地局データとの比較において、図表 17 で外れ値となったメッシュのマップを図表 21 に、2017 平日の GPS データと基地局データとの差率マップを図表 22 に示す。

図表 21：GPS データと基地局データの外れ値マップ



図表 22 : 2017 平日の GPS データと基地局データとの差率マップ



図表 22 で差率の分布を俯瞰してみると、東京都全体としては大きな偏りはなく分布しているが、各エリアを詳細に見ると、山手線内で差率がプラス（GPS データの人口の方が大きい）のメッシュが偏在している一方、西多摩地区の山間エリアでは差率が大きなプラス（0.33 以上）と大きなマイナス（-0.33 未満）のメッシュが多数隣接して存在していた。更に隣県との県境や東京湾に面している海沿いのメッシュには差率がマイナスとなっている（GPS データの人口の方が小さい）メッシュが多く存在していた。

図表 21 で外れ値となったメッシュは全て GPS データの人口が基地局データより大きく、都心ターミナル駅や鉄道沿線の主要ターミナル駅を含むエリアであった。ここは社会活動が活発になる昼間帯に、平日ではビジネス街、休日では繁華街への動線となる、時間的にも空間的にも著しく人口が集中するエリアであり、図表 22 で示した山手線内の差率がプラスのメッシュの一部に対応するものであった。

(4) メッシュサイズ拡大効果の分析

図表 17 で使用した 2017 年休日／平日の GPS データと基地局データのメッシュサイズを 1km に拡大したときの相関係数を図表 23 に示す。500m メッシュと比べて相関係数は高くなり、差率の標準偏差は小さくなった。

図表 23：1km メッシュでの GPS データと基地局データとの統計値

	対象データ	メッシュ数	相関係数	差率平均	差率標準偏差
1	2017 休日	1,476	0.94	-0.02	0.30
2	2017 平日	1,448	0.94	-0.03	0.31
(参考：500m メッシュ)					
1	2017 休日	5,172	0.88	-0.04	0.35
2	2017 平日	5,129	0.89	-0.06	0.35

(5) 年齢階級別の相関分析

2017 平日の GPS データ（年齢区分はなく総数データ）と年齢階級別の基地局データとの相関係数と、比較対象として全基地局データと年齢階級別の基地局データとの相関係数を図表 24 で示す。GPS データは全年齢階級で相関が示されているが、10 代（15 歳以上）と 70 代の相関係数は 20～60 代のそれよりも低かった。また比較対象としての全基地局データと年齢階級別の基地局データとの相関係数も、同様の傾向であった。

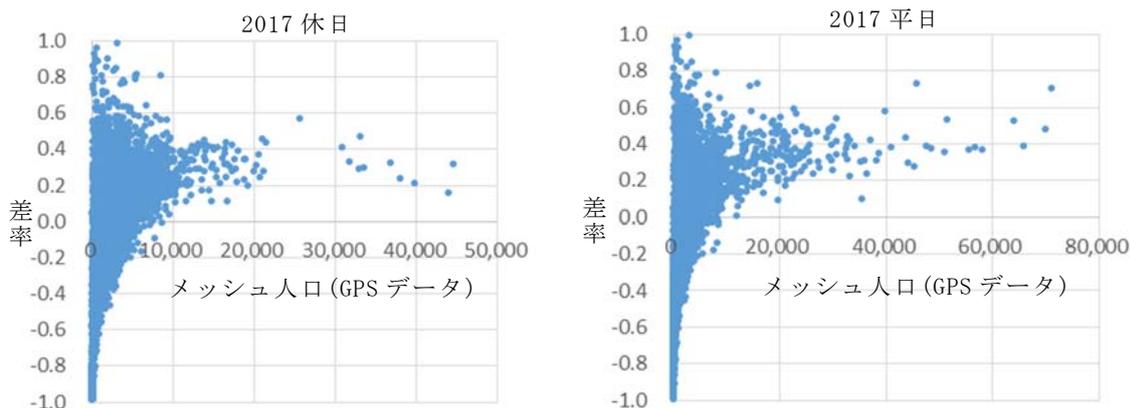
図表 24：2017 平日の GPS データと年齢階級別基地局データとの相関係数

	10 代	20 代	30 代	40 代	50 代	60 代	70 代
相関係数	0.57	0.88	0.92	0.91	0.90	0.75	0.52
(比較対象：基地局データ全体と年齢階級別基地局データとの相関係数)							
相関係数	0.75	0.96	0.98	0.99	0.99	0.93	0.76

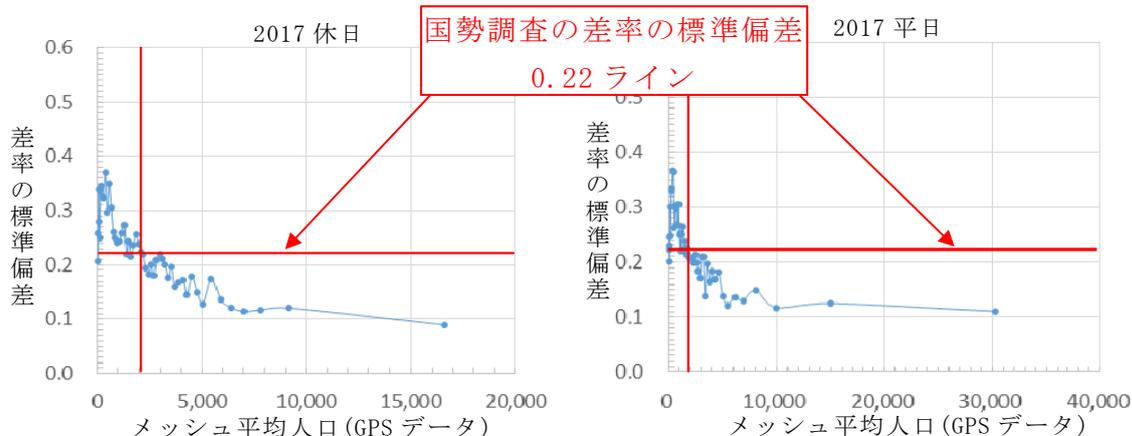
(6) 人口規模別の差率分析

図表 18 で使用した 2017 年の GPS データと基地局データとの差率を縦軸、GPS データの人口を横軸とした散布図を図表 25、更にこれを GPS データの人口でソートし 100 メッシュ毎にグループ化した GPS データの平均人口と差率の標準偏差をプロットしたグラフを図表 26 に示す。

図表 25 : GPS データと基地局データとの比較における GPS 人口と差率との散布図



図表 26 : グループ化した GPS データ平均人口と差率の標準偏差との散布図



図表 25 で、メッシュ人口の少ない層では差率がマイナスのメッシュが多数存在し、人口規模の増大とともに差率がプラスのメッシュが増加することがわかった。図表 26 で、メッシュ人口の少ない層では差率の標準偏差は大きいですが、人口規模の増大とともに差率の標準偏差が小さくなり、GPS データ平均人口が 2,000 人前後で、データの信頼性の基準として設定した国勢調査 5 年間における差率の標準偏差 (0.22) と同レベルとなった。

4. 考察

4.1 国勢調査との比較

GPS データが 24 時間平均人口、国勢データが常住人口と、この両者のデータがもともと異なる状態を表したデータであることを考慮すれば、図表 10 と図表 12 で、昼間帯に人口が集中する都心ターミナル駅が外れ値として現れていることは合理的な結果である。この両者間の相関係数は 0.52 と高くはないが、これには前述の外れ値の影響が含まれていることも考慮すると、GPS データは国勢調査との相関が示されており、利活用においても十分に有効なデータであると考えられる。

また、図表 9 で示されたように、GPS データと基地局データそれぞれの国勢データとの差率の標準偏差はほぼ同じであり、同レベルの精度が確保されていると考えられる。

更に図表 14 で、時間帯別の GPS データには時間経過に伴う人口変動に各市区町村の特徴が現れており、国勢調査では把握対象外である通勤・通学以外の勤務中の移動や余暇・消費活動による人の動きを捉えているといえる。

4.2 基地局データとの比較

(1) 標本の偏りや作成方式に関する考察

図表 15 の GPS データと基地局データとの比較において、休日では強い正の相関が示されているのに対して、それと比べると平日では相関が弱かった。この平日のデータに含まれている特異な数値を示すメッシュを除いたところ、相関係数は休日と同レベルになった。この事象は GPS データのアプリユーザーの属性に起因する標本の偏りの影響と推測できる。

図表 20 で両方のデータのメッシュ人口の分布を分析したところ、人口の少ないメッシュ（約 100 人未満）と多いメッシュ（約 5000 人以上）では GPS データが基地局データよりも多く存在していた。これは GPS データの標本の偏りが要因の一つと考えられるが、GPS データの高い解像度により、極端に少ない、又は極端に多い人口がより精度良くメッシュに反映されたことも、要因の一つであると推測される。

(2) メッシュサイズに関する考察

図表 23 の 1km メッシュサイズの比較において、メッシュサイズの拡大により相関係数が上昇したのは当然のことであるが、500m メッシュでも相関係数は 0.9 程度を示しており、500m メッシュでも十分な精度が確保されていると解釈できる。

(3) 年齢別考察

図表 24 で、比較対象として実施した全基地局データと年齢階級別の基地局データとの比較において、10 代（15 歳以上）と 70 代は 20～60 代よりも相関が弱かった。これは基地局データ自体に年齢階級に依存した標本の偏りがあるのか、又はそもそも 10 代と 70 代の人の動きが 20～60 代と異なるのかのどちらか又は両方に起因するものと考えられる。GPS データと年齢階級別の基地局データとの比較で現れた同様の事象には、この基地局データに起因するものも含んでいるため、GPS データに特定の年齢階級に偏りがあるかどうかは言及できない。

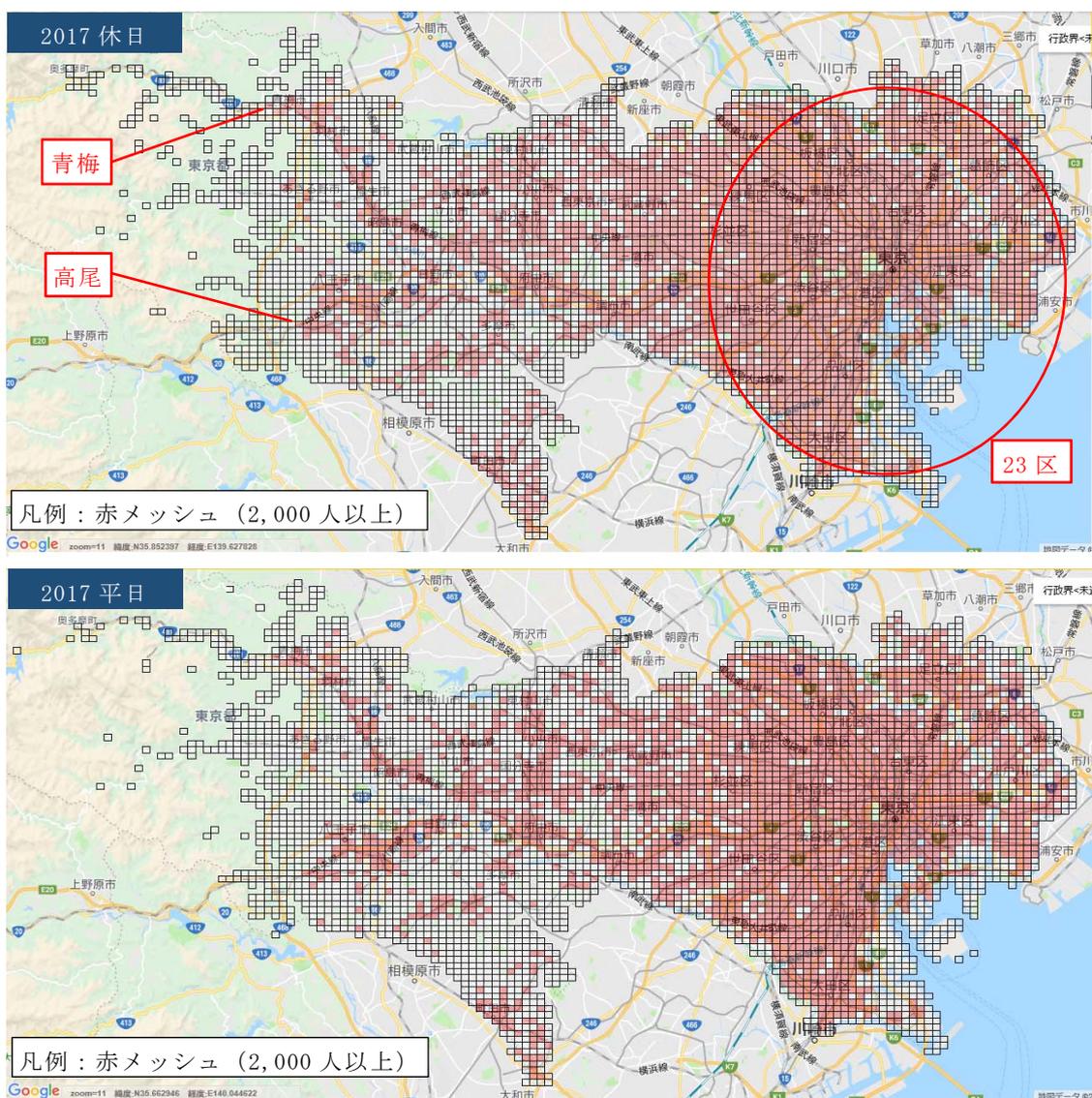
(4) 人口規模別考察

図表 26 で示した人口規模別にみた GPS データと基地局データとの差率の標準偏差の分析では、人口規模の増大とともに差率の標準偏差が小さくなり、GPS データのメッシュ平均人口が約 2,000 人で、データの信頼性の基準として設定した国勢調査 5 年間における差率の標準偏差（0.22）と同レベルになった。国勢

調査 5 年間に於ける標準偏差は時系列で比較したものであり単純な比較はできないが、GPS データは 2,000 人以上のメッシュにおいて一定レベルの信頼性が確保されているとも解釈できる。また国勢調査は信頼できる公的な人口統計として、5 年間に生じる乖離を許容して活用されていることを前提とすれば、利活用という視点においてもこのデータは有効であると考えられる。

図表 27 は GPS データのメッシュ人口が 2,000 人以上の分布を表している。対象となるメッシュの数は休日で 2620 個（全 5172 個）、平日で 2400 個（全 5129 個）と、全メッシュの約半数であり、その分布は 23 区内を中心に鉄道路線に沿って、西端は西多摩エリアの青梅駅や高尾駅まで広がっている。また休日と平日では、メッシュ数とその分布に関して、大きな違いはなかった。

図表 27：メッシュ人口 2,000 人以上の分布図



(5) エリア面からの考察

図表 22 で東京湾に面しているメッシュの差率がマイナス（GPS データの方が少ない）となっているのは、GPS データは解像度が高いため、人のいない海上と隣接する海沿いのメッシュの人口を精度良く把握していることが要因の一つであると考えられる。

図表 21 で外れ値となったメッシュは、都心の山手線ターミナル駅等、大量の人が移動する際の動線となる地点を含むエリアである。GPS データは高い解像度を持つとともに、前述のとおり一定人口以上のメッシュにおける精度はある程度確保されていることから、このように人口が極端に集中しているエリアの流動人口は一定レベルの信頼できる情報であると推測できる。

エリア面からの分析結果をまとめると、隣接するメッシュ間で人口差が大きいエリアでは、高い解像度を持つ GPS データのメリットが活かされて、より有効に機能するものと評価できる。

5. まとめ

本ワーキングペーパーは、東京都において GPS 方式で収集したメッシュ型の流動人口データ（GPS データ）を国勢調査や基地局方式で収集したデータ（基地局データ）と比較し、統計的に分析した結果をまとめたものである。この結果から GPS データは、国勢調査や基地局データと相関があり、一定レベルの信頼性が確保されているとともに、利活用にあたり国勢調査を補完しうるものとして有効であることを示すことができた。更に以下の GPS データの特徴に留意して使用することで、より効果的な活用が可能であると考えられる。

- 国勢調査 5 年間における差率の標準偏差を基準とした評価から、GPS データは流動人口 2,000 人以上のメッシュ（500m 四方）において一定レベルの信頼性が確保されており、そのエリアは東京都 23 区を中心に居住地域の面積の約半分に該当する。
- 時間帯別の GPS データは、国勢調査では把握対象外である通勤・通学以外の勤務中の移動や余暇・消費活動による人の動きを、ビジネス街や住宅地などの地域の特性と整合する形で捉えている。
- GPS データは解像度の高さから、大量の人の動線のハブとなる都心ターミナル駅や海岸沿い等、隣接するメッシュ間の人口差が大きいエリアにおいて、より強みを発揮できる。
- GPS データはアプリユーザーの属性に依存することから、標本の偏りが発生しやすいことを理解した上で、利活用を進めていくことが肝要である。

今回の検証において、携帯端末の位置情報を活用した民間の流動人口データは、国勢調査を補完しうるものとして利活用できる可能性が見いだされた。本稿は東京都に絞って検証を進めてきたが、更に、これを全国に拡大するとともに、より効果的かつ効率的に検証を進めていくことが必要であると考えられる。

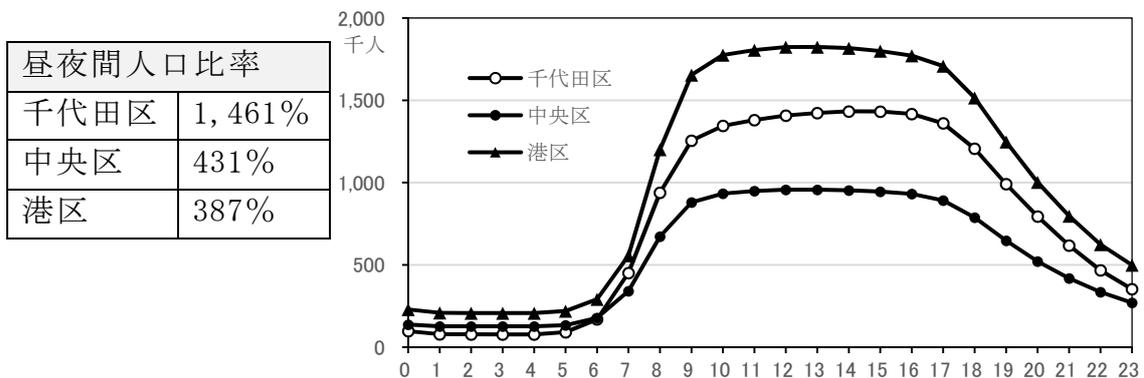
また今回の取組を展開し、各種ビッグデータの活用による正確で効率的な統計の作成に役立てるとともに、それをオープンデータ化して広く一般にも利活用できる仕組みを構築することにより、政府や民間でのより効果的な取組へとつなげていくことが必要であると考えられる。

(参考) 流動人口データ¹⁸による市区町村別人口変動の分析

1. 流動人口の時間帯別変動の分析

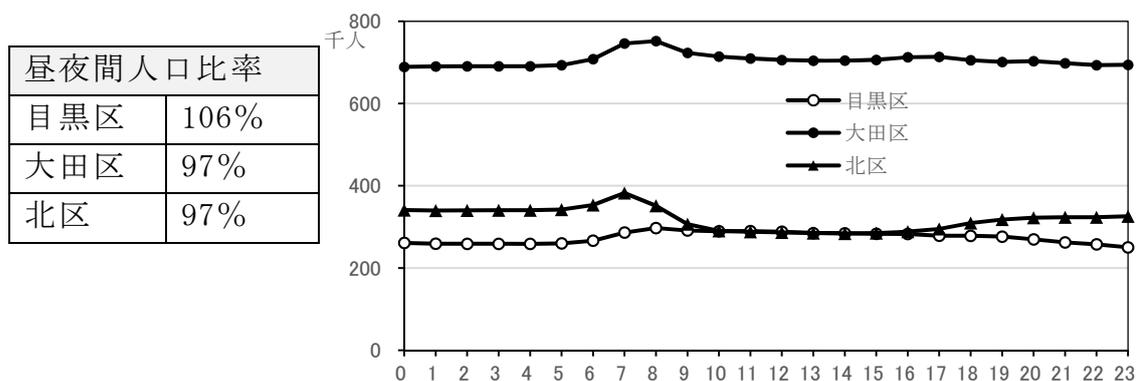
平日の流動人口について、時間帯別変動の分析を行う。2015年国勢調査の昼夜間人口比率の大きさを基準に、代表的な市区町村を抽出し、時間帯別の変動をグラフ化した。

(1) 区部で昼夜間人口比率が最も大きい3区



- ・ 夜間に比べて昼の人口が大幅に増加。「都心型」と呼ぶことができる。

(2) 区部で昼夜間人口比率が100%に近い3区

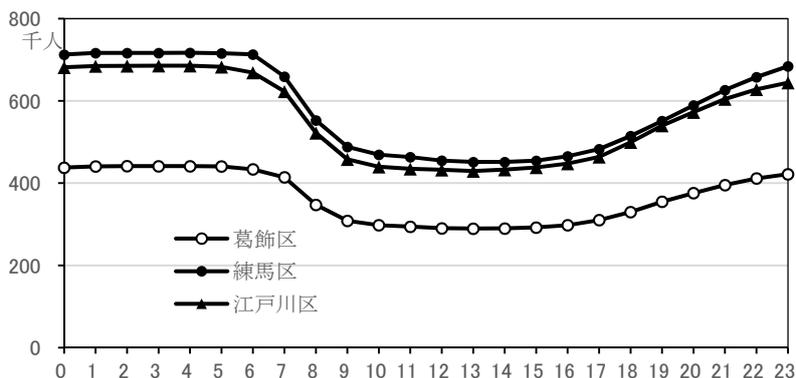


- ・ ほぼ横ばい。大田区と北区は朝の通勤時間帯に通過人口によるピークを示す。

¹⁸ スマートフォンのアプリのGPS位置情報を基に人口推計したもの。株式会社Agoopより入手。

(3) 区部で昼夜間人口比率が最も小さい3区

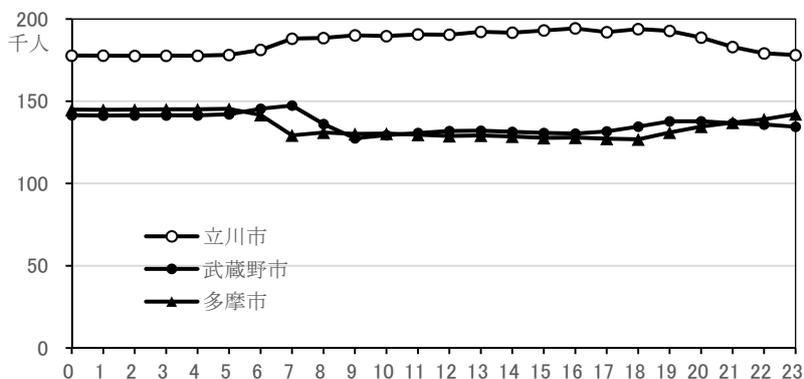
昼夜間人口比率	
葛飾区	84%
練馬区	84%
江戸川区	82%



- ・ 夜間に比べて昼の人口が減少。「住宅地型」と呼ぶことができる。

(4) 市部で昼夜間人口比率が最も大きい3市

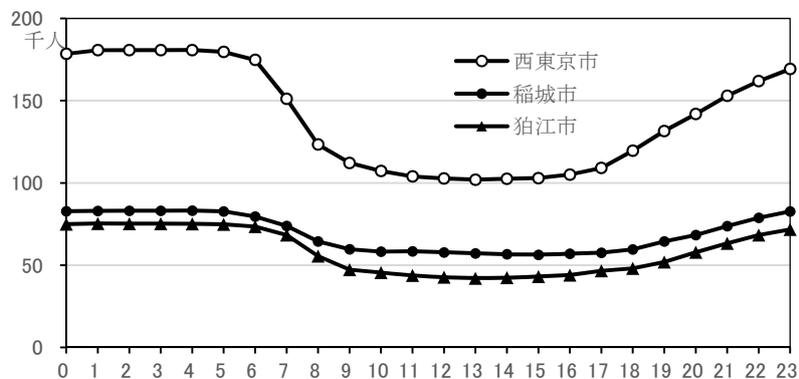
昼夜間人口比率	
立川市	114%
武蔵野市	109%
多摩市	101%



- ・ 増減幅は小さいが、立川市は昼の人口が増加し、武蔵野市と多摩市は昼の人口が減少。武蔵野市では、朝の通勤時間帯の通過人口の影響で、7時にわずかなピークあり。

(5) 市部で昼夜間人口比率が最も小さい3市

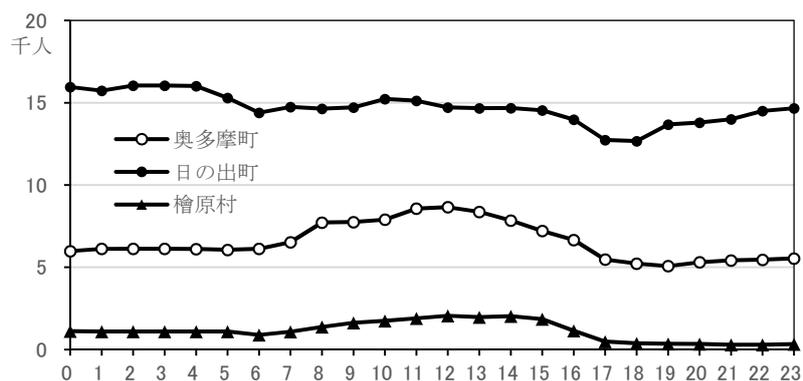
昼夜間人口比率	
西東京市	79%
稲城市	78%
狛江市	74%



- ・ 夜間に比べて昼の人口が減少する「住宅地型」。

(6) 町村部

昼夜間人口比率	
奥多摩町	100%
日の出町	99%
檜原村	95%

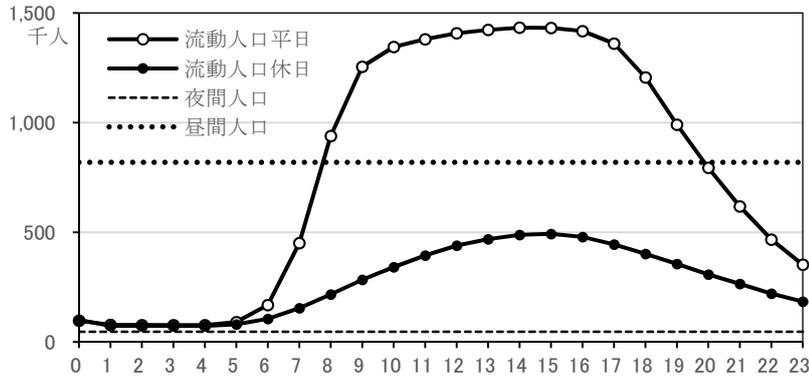


- ・ 奥多摩町と檜原村は昼に人口が増加傾向。日の出町は夕方 17～18 時に人口が最小。
- 以上のグラフを見ると、区部や市部については、流動人口の時間帯別推移は、ある程度実感と整合的な動きになっているといえる。

2. 時間帯別流動人口と国勢調査人口との比較

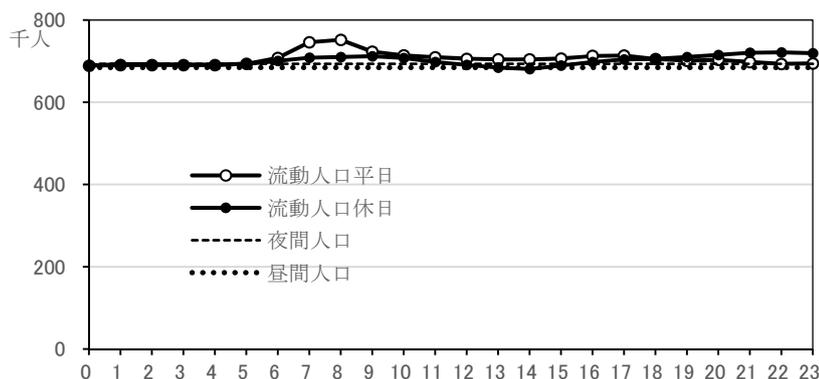
休日及び平日の時間帯別の流動人口と 2010 年国勢調査人口との比較を行う。
1. で取り上げた中の代表的な市区について、時間帯別の変動をグラフ化した。

(1) 千代田区



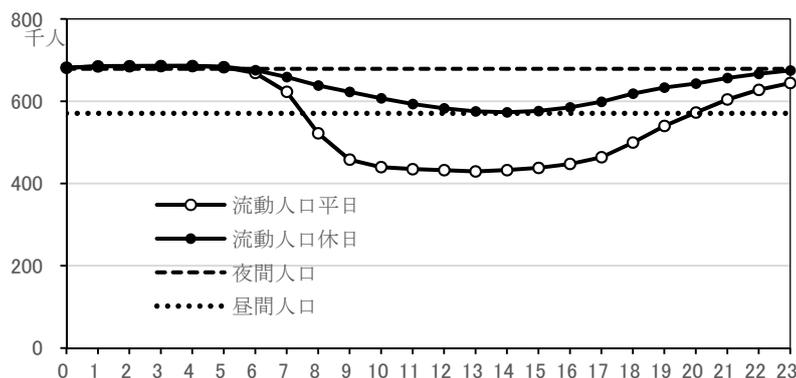
- ・ 深夜の流動人口は、休日／平日ともに、国勢調査の夜間人口よりわずかに大きい。
- ・ 昼の時間帯では、平日の流動人口の最大値は国勢調査の昼間人口より約 60 万人多く、休日の流動人口のピーク値は国勢調査の夜間人口より約 40 万人多い。これらの差は、通勤・通学以外で千代田区を訪問した人口と解釈できる。

(2) 大田区



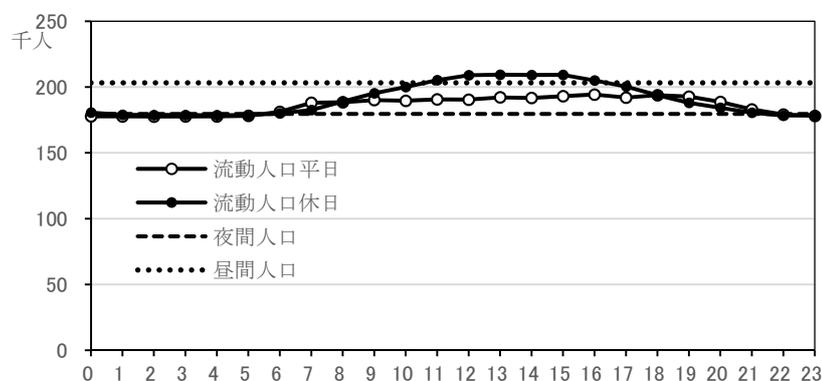
- ・ 平日朝の通勤時間帯に通過人口によるピークがあるほかは、休日／平日ともに流動人口の時間帯による変動は小さい。

(3) 江戸川区



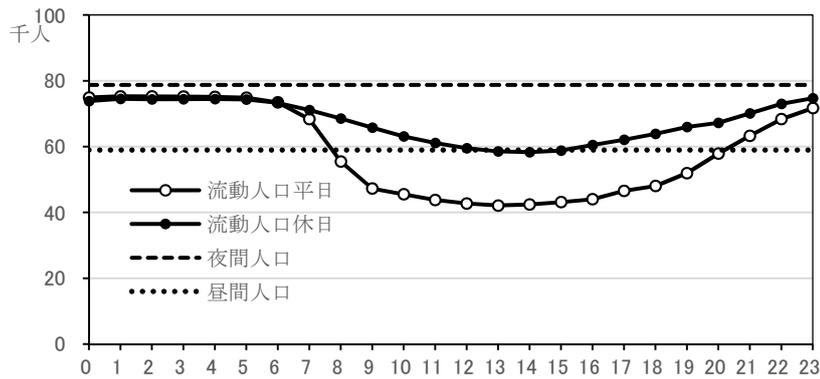
- ・ 深夜の流動人口は、休日／平日ともに、国勢調査の夜間人口とほぼ同水準。
- ・ 昼の時間帯では、平日の流動人口の最小値は国勢調査の昼間人口より約10万人少なく、休日の流動人口の最小値は国勢調査の夜間人口より約10万人少ない。これらの差は、通勤・通学以外で江戸川区外に出た人口と解釈できる。

(4) 立川市



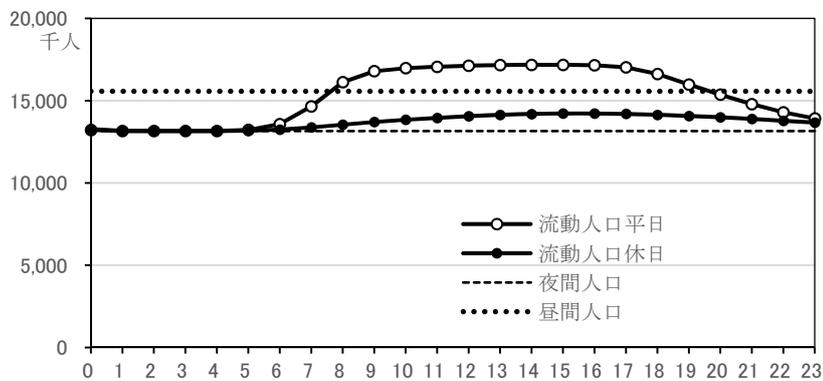
- ・ 深夜の流動人口は、休日／平日ともに、国勢調査の夜間人口とほぼ同水準。
- ・ 流動人口が昼に増加する市区では、ほとんどの場合、昼の時間帯は平日のほうが休日より人口が大きいですが、立川市では逆で、平日より休日の人口が大きい。

(5) 狛江市



- ・ 昼の時間帯では、平日の流動人口の最小値は国勢調査の昼間人口より約2万人少なく、休日の流動人口の最小値は国勢調査の夜間人口より約2万人少ない。これらの差は、通勤・通学以外で狛江市外に出た人口と解釈できる。
- ・ 深夜の流動人口は、休日／平日ともに、国勢調査の夜間人口より若干小さい。

(6) 東京都合計



- ・ 昼の時間帯では、平日の流動人口の最大値は国勢調査の昼間人口より約160万人多く、休日の流動人口の最大値は国勢調査の夜間人口より約100万人多い。これらの差は、通勤・通学以外で東京都に来訪した人口と解釈できる。
- ・ 深夜の流動人口は、休日／平日ともに、国勢調査の夜間人口と同水準。

- 以上の、流動人口と国勢調査人口との比較から、次のような状況が読み取れる。
 - ① 昼の時間帯では、平日の流動人口の最大値（又は最小値）と、国勢調査の昼間人口との間に差が見られる。この差は、通勤・通学以外の移動人口（買い物・観光などが考えられる）が反映されていると解釈することができる。
 - ② 深夜の時間帯では、流動人口の休日／平日との間には差がなく、国勢調査の夜間人口とも概ね同水準となっている。

3. まとめ

市区町村別に集計した流動人口の時間帯別データは、時系列の動きが実感ともある程度整合性があることが確認できた。また、国勢調査の昼間人口は通勤・通学者の移動を反映したデータであるが、流動人口の昼の時間帯の人口は、通勤・通学以外の移動人口（買い物・観光などが考えられる）も反映されたデータとなっている。

このような流動人口の時間帯別データは、国勢調査や推計人口などの常住地ベースの人口データでは把握できない、現在地ベースによる時間帯別の人口データであり、人口の時間変動に関する様々な分析に有用であると考えられる。ただし、人口の少ない町村部のデータについては、推計誤差について留意した上で利用する必要があると思われる。

参考文献

独立行政法人統計センター・株式会社 NTT ドコモ【2013】「官庁統計とモバイル空間統計に基づく新たな統計の創出に関する共同研究」、平成 24 年度共同研究報告書、2013 年 3 月

総務省【2014】「位置情報プライバシーレポート～位置情報に関するプライバシーの適切な保護と社会的利活用の両立に向けて」、緊急時等における位置情報の取扱いに関する検討会報告書、平成 26 年 7 月

内閣府【2015】「公的統計におけるビッグ・データの活用に関する調査研究報告書」、平成 26 年度内閣府大臣官房統計委員会担当室請負調査、平成 27 年 3 月

国土交通省【2018】「総合都市交通体系調査におけるビッグデータ活用の手引き」、国土交通省都市局都市計画課都市計画調査室、平成 30 年 6 月

閣議決定【2018】「公的統計の整備に関する基本的な計画」

総務省「地域メッシュ統計の概要」、総務省統計局ホームページ

総務省【2019】「東京都における流動人口データの有効性の検証～メッシュ型流動人口検証 WG 結果報告」、第 6 回ビッグデータ等の利活用推進に関する産官学協議のための連携会議、2019 年 5 月 22 日