

# 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～

2019年5月22日

メッシュ型流動人口検証WG



## 目次

1. はじめに
2. 検証データと検証方法
  - 2.1 検証概要
  - 2.2 国勢調査との比較方法
  - 2.3 基地局データとの比較方法
3. 検証結果
  - 3.1 国勢調査との比較結果
  - 3.2 基地局データとの比較結果
4. まとめ

## 1. はじめに

- 各種マーケティングにおいて公的な人口統計として国勢調査を活用している
  - 長い調査周期（5年）と公表までのリードタイム（1年）が課題
  - 携帯端末の位置情報から取得した流動人口で補完
- 「公的統計の整備に関する基本的な計画」において、ビッグデータの活用を推奨している
  - 報告者の負担軽減
  - 正確で効率的な統計の作成
- 民間から提供される流動人口データの信頼性と利活用にあたっての有効性を検証する（本WGでの検証範囲）
  - 東京都
  - 携帯端末に搭載されたアプリGPSによる流動人口

3

## 目次

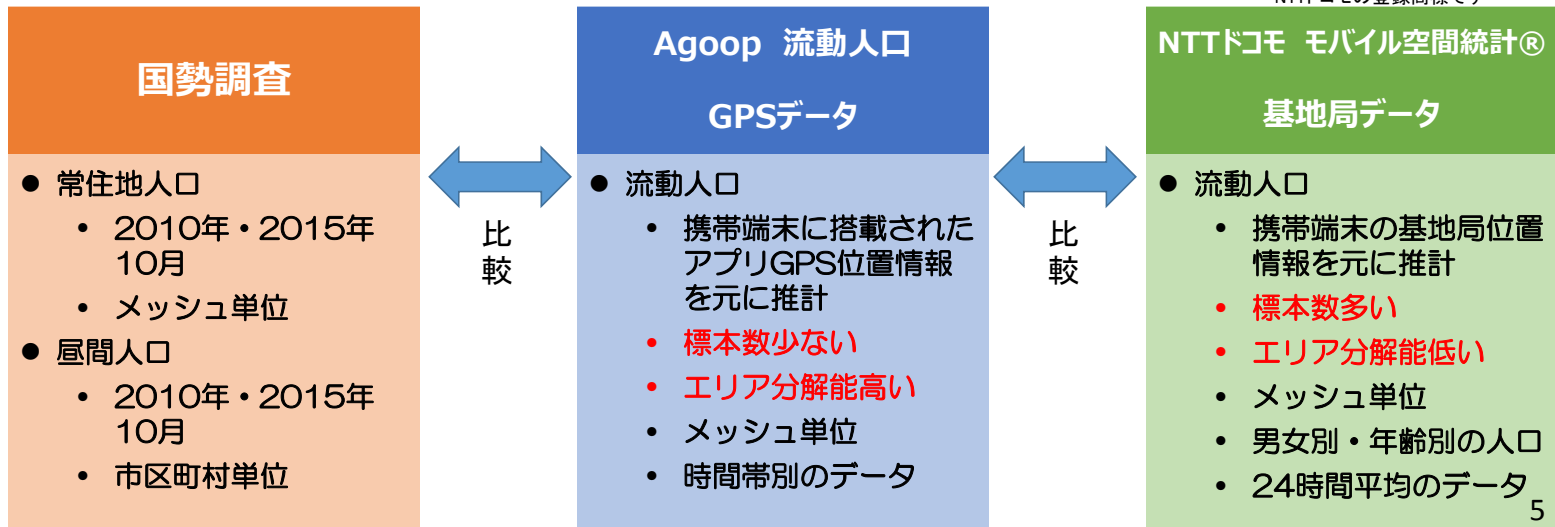
1. はじめに
2. 検証データと検証方法
  - 2.1 検証概要
  - 2.2 国勢調査との比較方法
  - 2.3 基地局データとの比較方法
3. 検証結果
  - 3.1 国勢調査との比較結果
  - 3.2 基地局データとの比較結果
4. まとめ

4

## 2.1 検証概要

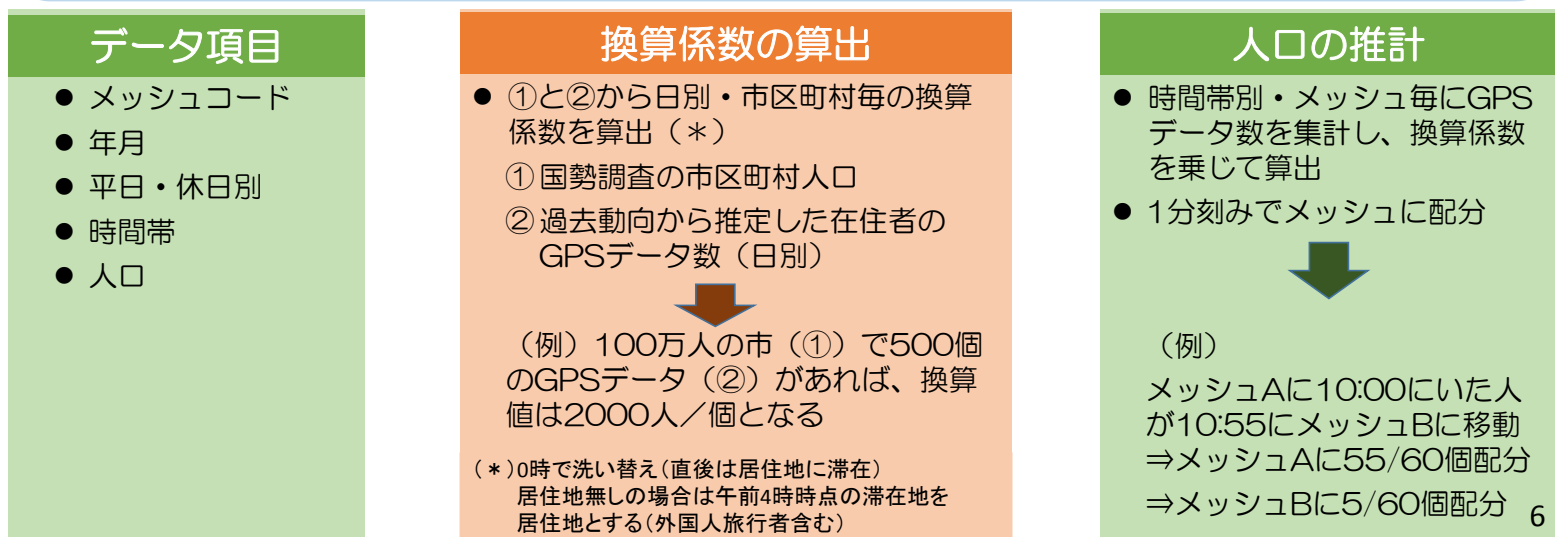
東京都において、**標本数の少ないアプリGPSデータ**（Agoop社のメッシュ型流動人口）の**信頼性・有効性を検証**するために、公的統計である**国勢調査データ**、及び精度が高いと想定される**NTTドコモのモバイル空間統計®**との乖離の状況を、**散布図・ヒストグラム等**を使用して比較分析する。

(\*)モバイル空間統計は株式会社NTTドコモの登録商標です



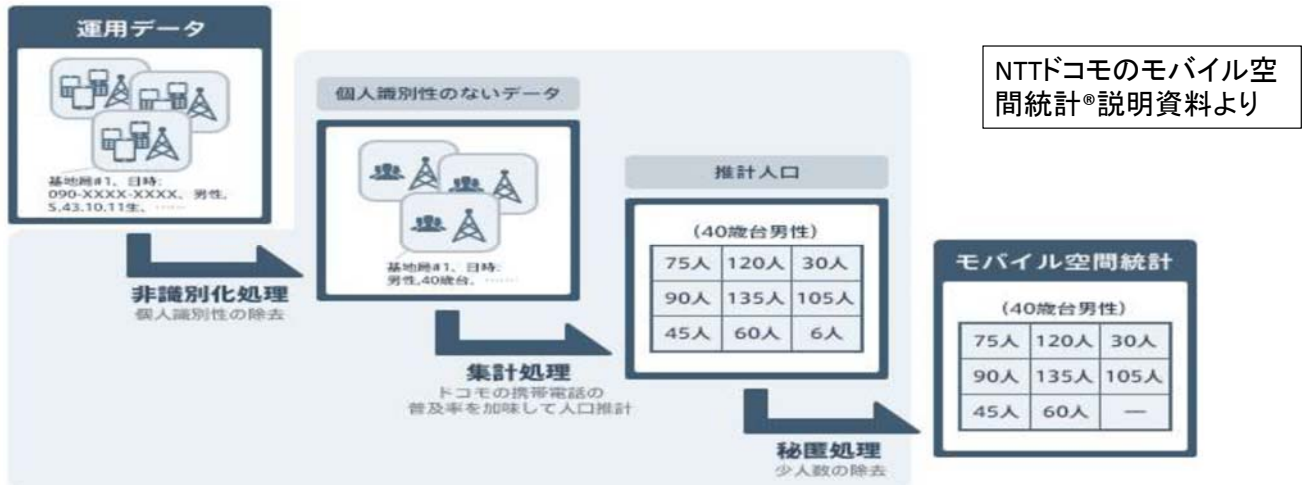
## 2.1 検証概要\_GPSデータ

- 国勢調査の市区町村別人口をベンチマーク（基準人口）として、携帯アプリから取得した**GPSデータ（約21万セッション）**の存在数に応じてベンチマーク（基準人口）を配分した推定値（メッシュ型流動人口）である
- 500mメッシュは**高分解能の位置情報（100mメッシュ）**を積みあげて作成している



## 2.1 検証概要\_基地局データ

- NTTドコモの基地局データの仕組みを使用して携帯端末数（約7600万台）を把握する
- 住民基本台帳の性別・年齢階級（5歳）別の市区町村人口を元に拡大推計する
- 日本全国の1時間毎の人口分布を24時間365日、性別・年齢階級別に分けて提供する
- 500mメッシュは低分解能の位置情報（基地局単位）の重み付け分割で作成している

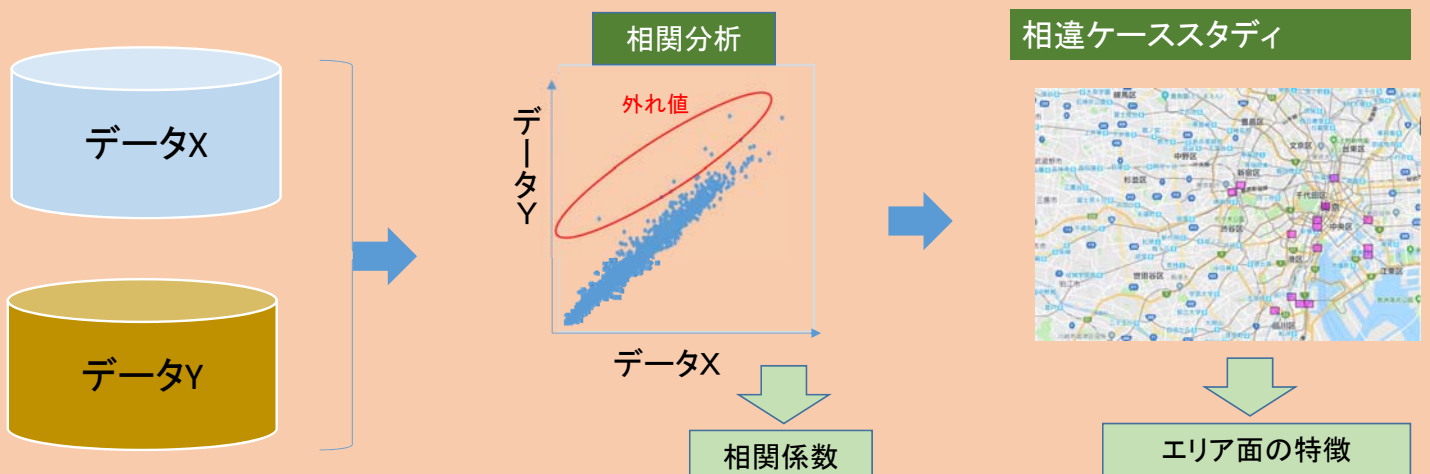


7

## 2.1 検証概要\_相関分析

- 二つのデータを散布図にプロットし、相関係数を算出する
- ケーススタディとして、相関から外れる15個のデータを外れ値として地図上で可視化し、エリア面からの特徴を分析する

### 検証方法

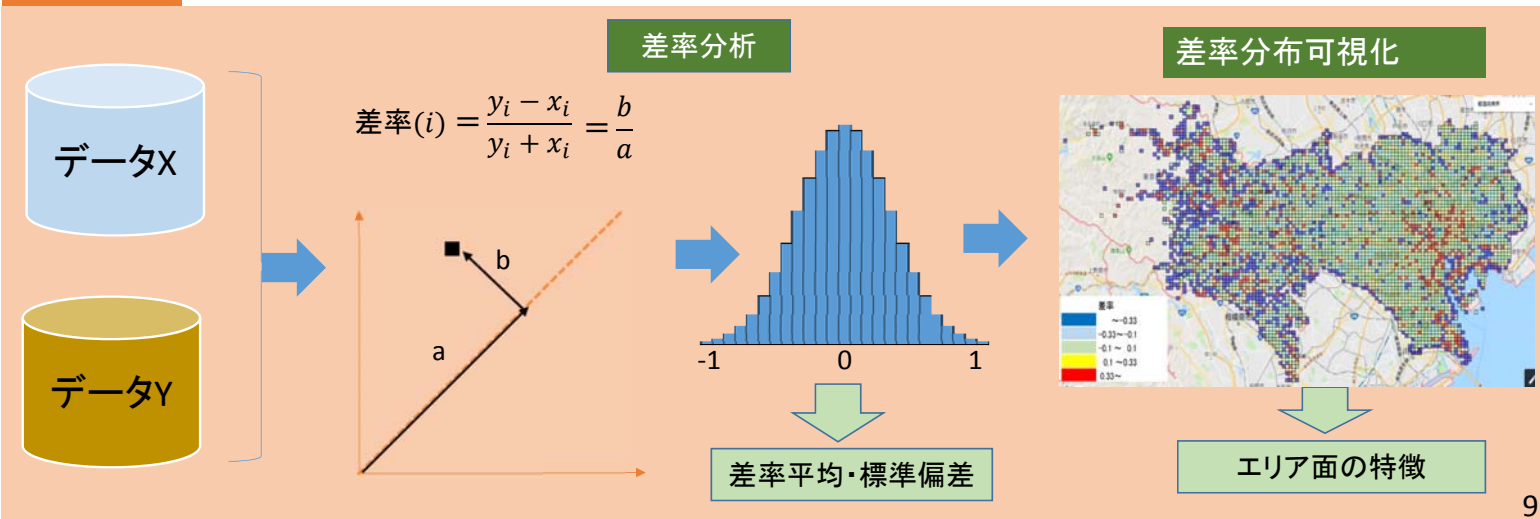


8

## 2.1 検証概要\_差率分析

- メッシュコード(i)に対応するデータX(i)とデータY(i)に関して、**差率(i)**を算出し、その**平均と標準偏差**を算出し、ヒストグラムを作成する
- 差率分布を地図上で可視化し、エリア面からの特徴を分析する

### 検証方法



## 目次

1. はじめに
2. 検証データと検証方法
  - 2.1 検証概要
  - 2.2 国勢調査との比較方法
  - 2.3 基地局データとの比較方法
3. 検証結果
  - 3.1 国勢調査との比較結果
  - 3.2 基地局データとの比較結果
4. まとめ

## 2.2 国勢調査との比較方法\_国勢調査常住人口

- 500mメッシュで“GPSデータと国勢調査との相関係数・差率”を算出し、“**基地局データと国勢調査との相関係数・差率**”と比較する

※基地局データは概ね正しいという前提

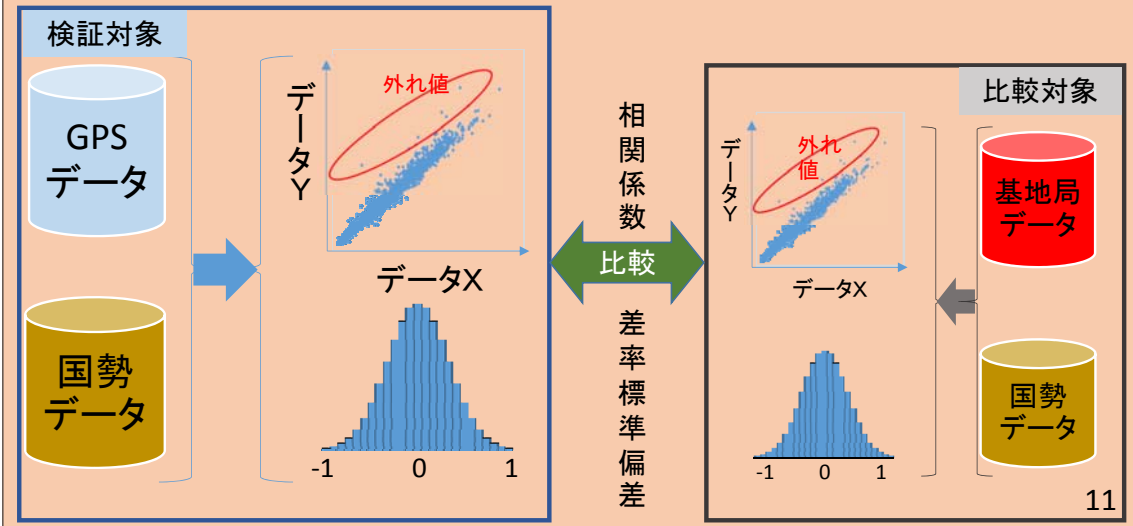
### 検証項目

- 相関分析、外れ値
- 差率分析

(比較年月)

- 201510休日

### 検証方法



11

## 2.2 国勢調査との比較方法\_国勢調査昼間人口

- 時間帯への依存関係を調べるため、**時間帯別のGPSデータ**（平日昼間の市区町村別）と**国勢調査昼間人口**（市区町村別）を比較する

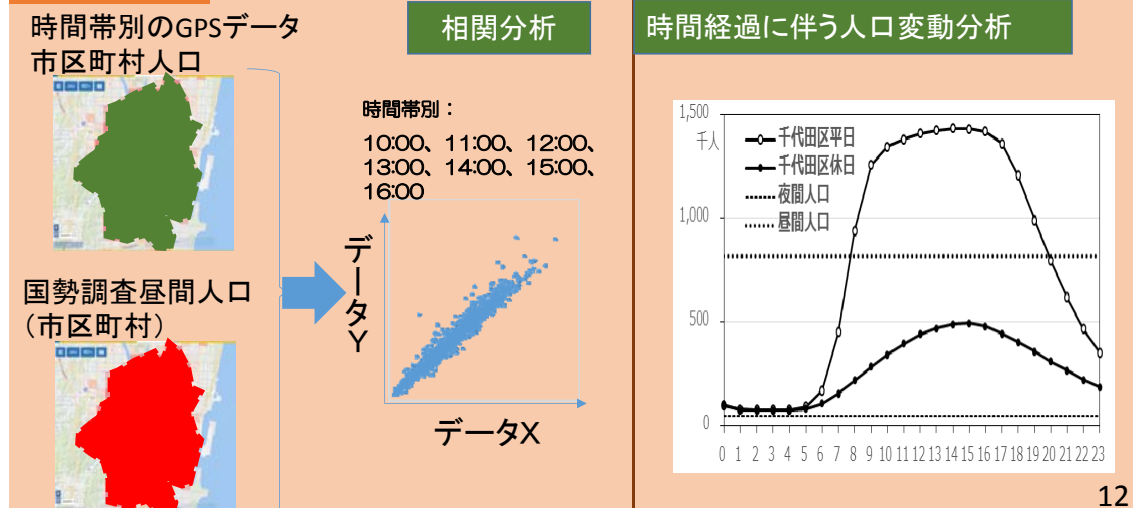
### 検証項目

- 相関分析
- 時間経過に伴う人口変動分析

(比較年月)

- 201510休日／平日

### 検証方法



12

# 目次

1. はじめに
2. 検証データと検証方法
  - 2.1 検証概要
  - 2.2 国勢調査との比較方法
  - 2.3 基地局データとの比較方法
3. 検証結果
  - 3.1 国勢調査との比較結果
  - 3.2 基地局データとの比較結果
4. まとめ

13

## 2.3 基地局データとの比較方法\_基本

- 500mメッシュで“GPSデータと基地局データとの相関係数・差率”を算出
- “国勢調査の5年間の相関係数・差率”を相違基準として比較する

※ 基地局データは概ね正しい前提下で、“国勢調査の5年間の関係”をメッシュ統計が許容していた相違基準とした

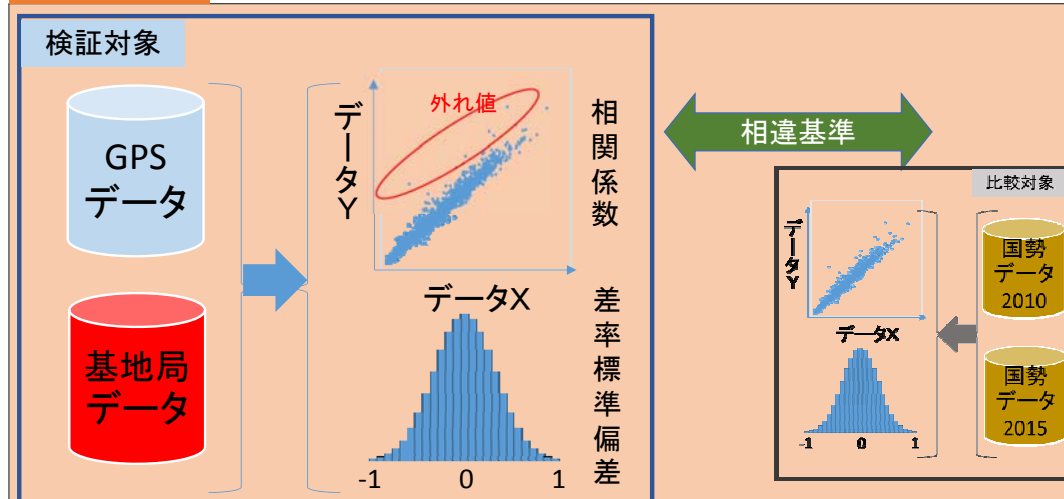
### 検証項目

- 相関分析、外れ値
- 差率分析、地図マッピング

(比較年月)

- 2015、2016、2017
- 休日／平日

### 検証方法



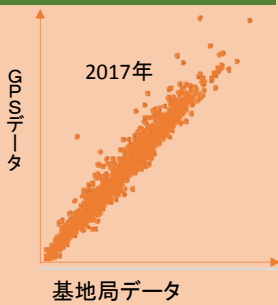
14

## 2.3 基地局データとの比較方法\_メッシュサイズ・年齢別

- “GPSデータと基地局データ” の比較に関して、メッシュサイズ拡大してその影響を分析する：500m→1km
- “GPSデータと年齢階級別の基地局データ” を比較して、年齢階級への依存関係を分析する：10代(15歳以上)、20代、30代、40代、50代、60代、70代

### 検証方法

#### メッシュサイズ拡大の影響 500m→1kmメッシュ

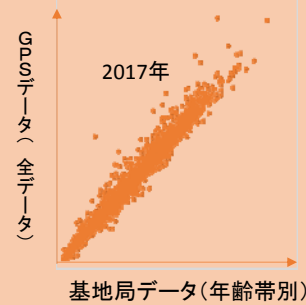


相関分析

相関係数

#### 年齢階級への依存

10代(15歳以上)、20代、30代、40代、50代、60代、70代



相関分析

相関係数  
(年齢階級別)

15

## 2.3 基地局データとの比較方法\_人口層別分布

- GPSデータ、基地局データのそれぞれが持つ人口分布に関する特徴を明らかにするため、メッシュ人口層別のヒストグラム（形態は折れ線）を作成する

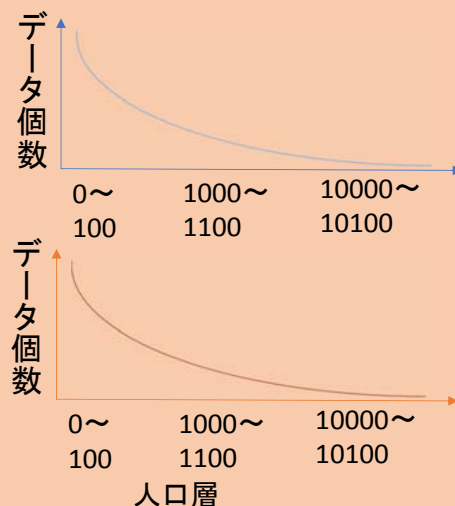
### 検証方法



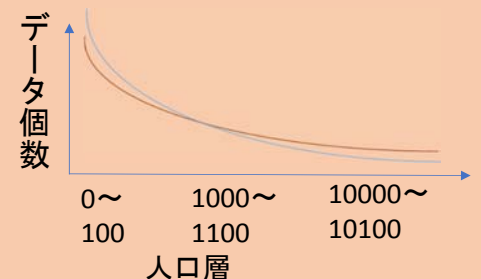
GPS  
データ



基地局  
データ



重ね合わせ



16

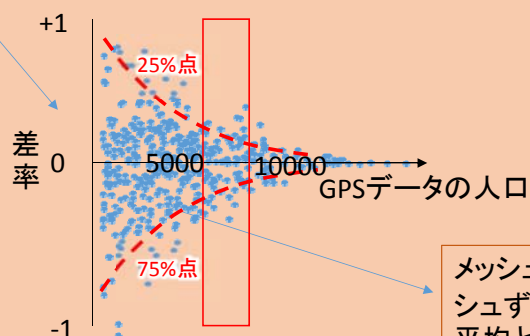


## 2.3 基地局データとの比較方法\_人口規模別偏差

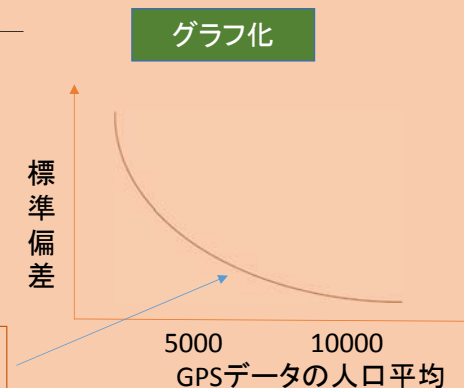
- “GPSデータと基地局データとの差率” に関して、GPSデータのメッシュ人口規模別に標準偏差を算出し、標準偏差とメッシュ人口規模の関係を可視化する

### 検証方法

$$\text{差率} = \frac{\text{GPSデータ(2017年)} - \text{基地局データ(2017年)}}{\text{GPSデータ(2017年)} + \text{基地局データ(2017年)}}$$



メッシュ人口順に並べ100メッシュずつグループ化し、人口平均と標準偏差を算出



17

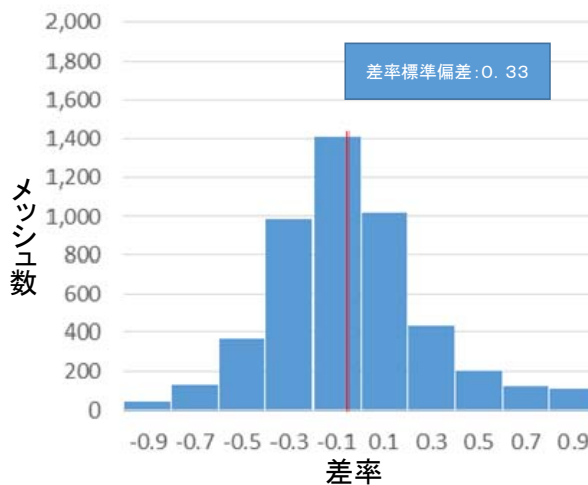
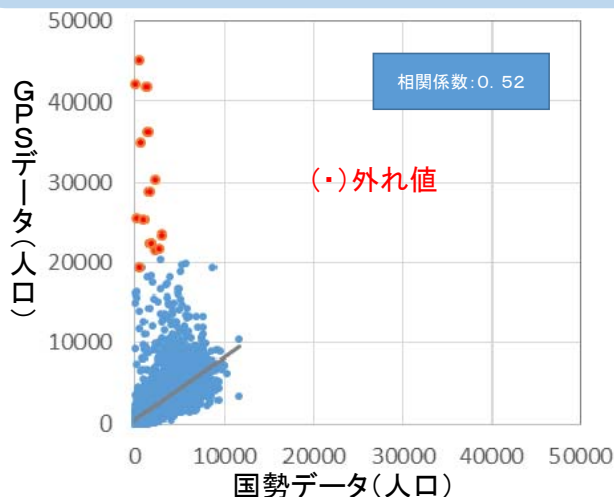
## 目次

1. はじめに
2. 検証データと検証方法
  - 2.1 検証概要
  - 2.2 国勢調査との比較方法
  - 2.3 基地局データとの比較方法
3. 検証結果
  - 3.1 国勢調査との比較結果
  - 3.2 基地局データとの比較結果
4. まとめ

18

### 3.1 国勢調査との比較結果\_常住人口

- GPSデータは24時間平均人口、国勢データは常住人口と異なるものであるが、**相関があり、一定の偏差**を持っている（相関係数：0.52 差率標準偏差：0.33）
- 散布図において、国勢データの人口が少ない（約3000人未満）層でGPSデータの人口が非常に多い（20000人以上）メッシュが外れ値として現れている



### 3.1 国勢調査との比較結果\_常住人口

- 外れ値となったメッシュは都心ターミナル駅周辺と鉄道沿線ターミナル駅である

➢ 昼間帯に人が集中するエリア

- 「GPSデータと国勢データ」は「基地局データと国勢データ」よりも相関が弱い、偏差レベルは同程度である

➢ データのまとめり具合は同レベル



2015休日GPSデータと国勢データとの外れ値マップ

2015休日におけるGPSデータ、基地局データ各々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	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

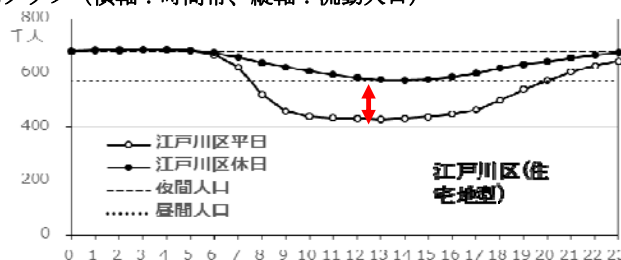
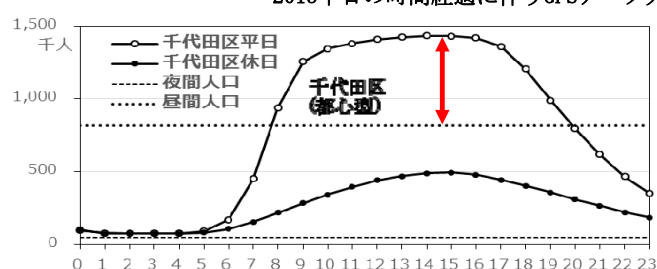
## 3.1 国勢調査との比較結果\_昼間人口

- GPSデータは国勢調査昼間人口と強い相関があり、**時間帯別（10:00～16:00）に相関の強さには違いが無い**
- GPSデータの平日昼間帯の人口が国勢調査昼間人口を超えて大きく変動している  
 ➤ **国勢調査で捉えられていない通勤・通学以外の余暇・消費活動等を把握**

2015平日の時間帯別GPSデータと国勢昼間データとの統計値

	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平日の時間経過に伴うGPSデータ人数変動グラフ（横軸：時間帯、縦軸：流動人口）



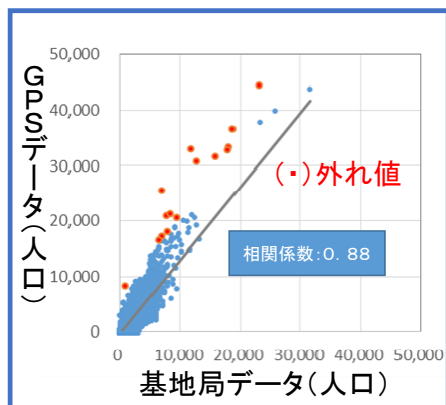
## 目次

- はじめに
- 検証データと検証方法
  - 2.1 検証概要
  - 2.2 国勢調査との比較方法
  - 2.3 基地局データとの比較方法
- 検証結果
  - 3.1 国勢調査との比較結果
  - 3.2 基地局データとの比較結果
- まとめ

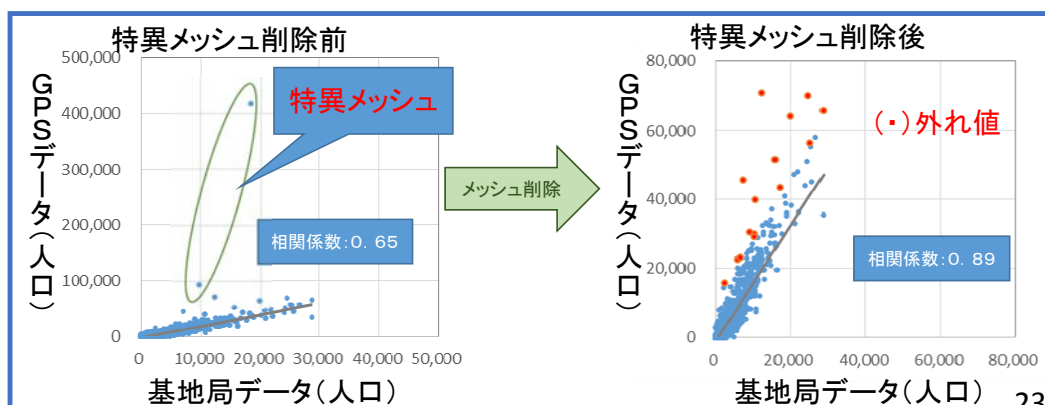
## 3.2 基地局データとの比較結果\_基本

- 休日において、GPSデータと基地局データは強い相関がある
- 平日の散布図において、大きく外れるメッシュが存在する
  - アプリ利用者属性に起因する標本の偏りがある
  - 散布図で大きく外れる“特異メッシュ”2点を除くと相関係数は大幅に高くなる

2017休日



2017平日



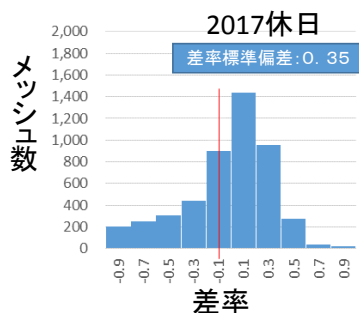
## 3.2 基地局データとの比較結果\_基本

- GPSデータと基地局データは強い相関と一定の偏差レベルを持っている
- 相関係数・差率標準偏差とも、年別（2015～2017）、平日／休日別で、質的な違いは無い

GPSデータと基地局データとの統計値

	検証名	メッシュ数	相関係数*1	差率 平均	差率 標準偏差
3.1-1	2015 休日	5,331	0.86	-0.03	0.33
3.1-2	2015 平日	5,283	0.87 (0.80)	-0.05	0.34
3.2-1	2016 休日	5,258	0.89	-0.04	0.34
3.2-2	2016 平日	5,163	0.90 (0.69)	-0.05	0.33
3.3-1	2017 休日	5,172	0.88	-0.04	0.35
3.3-2	2017 平日	5,129	0.89 (0.65)	-0.06	0.35

(\*1) ()内は特異メッシュ除外前

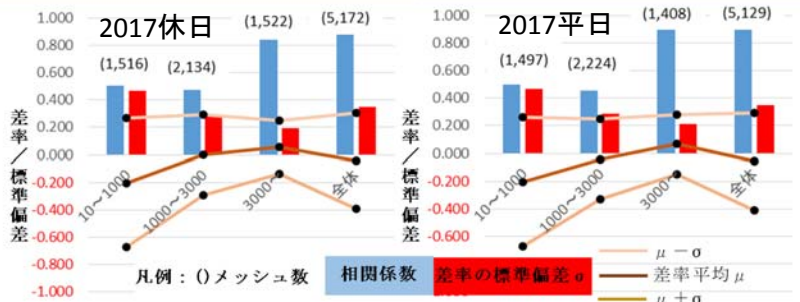


## 3.2 基地局データとの比較結果\_メッシュ人口層別分布

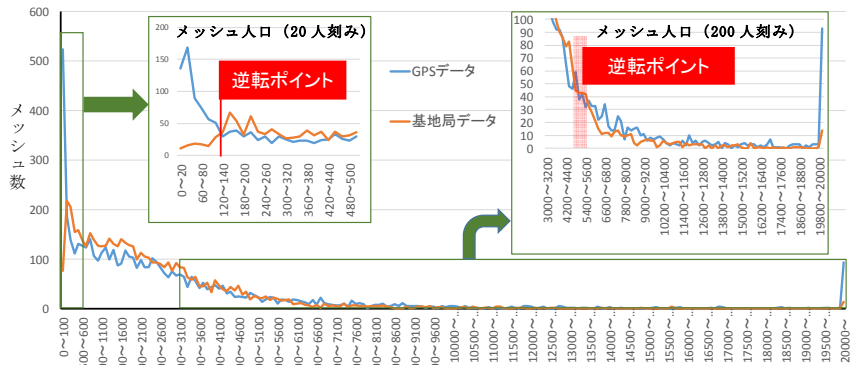
- 基地局データの人口が増加するに沿い、相関係数は大きくなり、差率標準偏差は小さくなる

- **メッシュ人口が100人未満と5000人以上では、基地局データよりもGPSデータの方がメッシュ数が多い**

➢ GPS方式はエリア分解能が高いため、両極の数値が出やすい



GPSデータと基地局データの人口層別グラフ (横軸は基地局人口層)

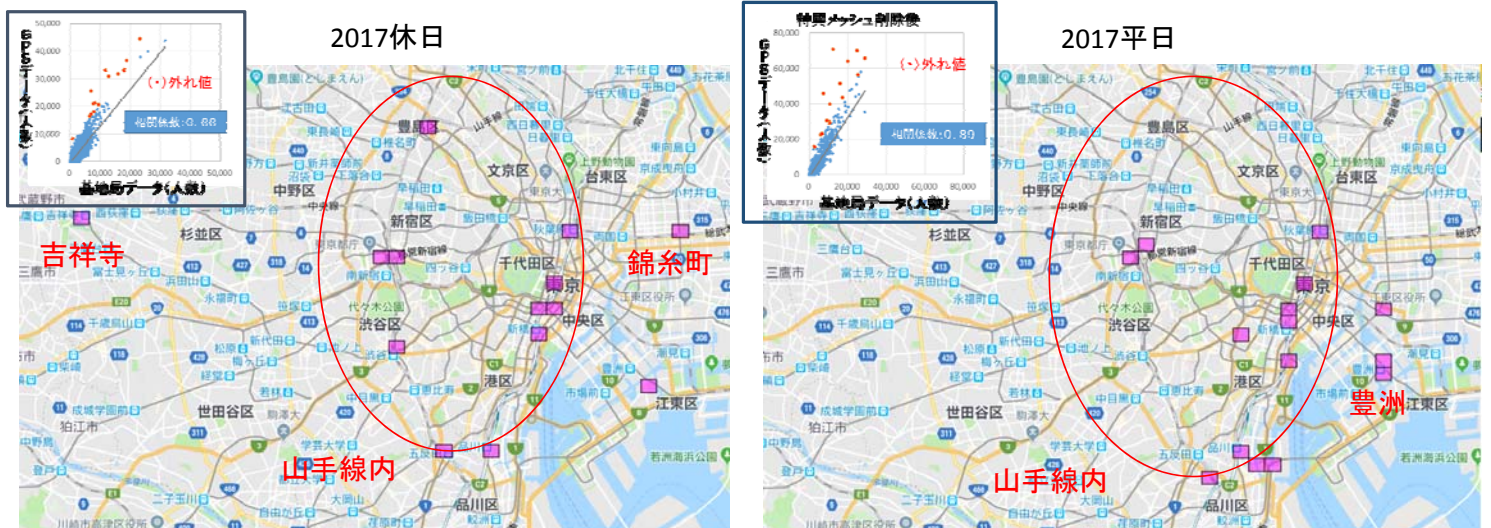


メッシュ人口 (100人刻み)  
2017平日のメッシュ人口別ヒストグラム (折れ線)

## 3.2 基地局データとの比較結果\_エリア面 (外れ値)

- 外れ値のメッシュは、平日ではビジネス街、休日では繁華街への動線となる**都心ターミナル駅、及び鉄道沿線の主要ターミナル駅**に現れている

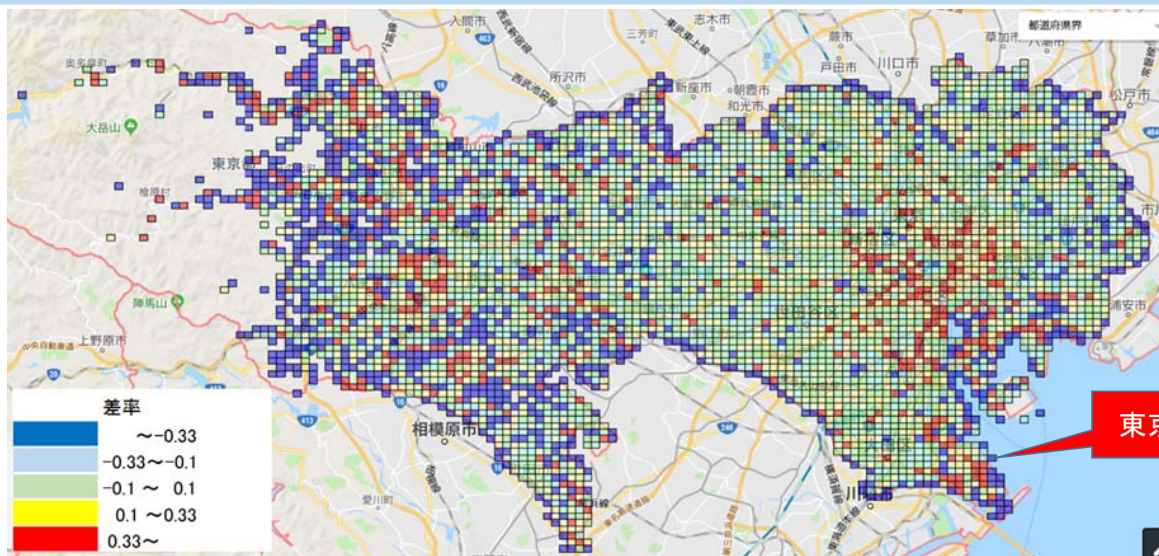
➢ 人口が集中するエリアを高分解能で捉えている



GPSデータと基地局データとの外れ値マップ

## 3.2 基地局データとの比較結果\_エリア面（差率）

- 東京都全体としては大きな偏り無く分布している
- 東京湾岸沿いのメッシュはGPSデータの方が人口が少ない（差率がマイナス）
  - 人のいない海に隣接したメッシュを高分解能で捉えている



GPSデータと基地局データの差率マップ

27

## 3.2 基地局データとの比較結果\_メッシュサイズ

- メッシュサイズの拡大により、**相関係数が高くなり、差率標準偏差が小さくなる**

1kmメッシュの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

28

## 3.2 基地局データとの比較結果\_年齢階級別

- GPS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代よりも相関が弱い**
  - 基地局データ全体と年齢階級別基地局データとの比較においても同様な傾向であり、**GPSデータの10代と70代の精度が悪い**とは言い切れない

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	0.40	0.65	0.68	0.68	0.67	0.54	0.38
(参考：基地局データ全体と年齢階級別基地局データとの比較)	0.75	0.96	0.98	0.99	0.99	0.93	0.76

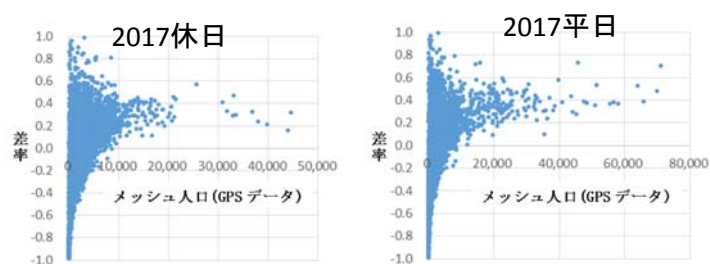
29

## 3.2 基地局データとの比較結果\_人口規模別標準偏差

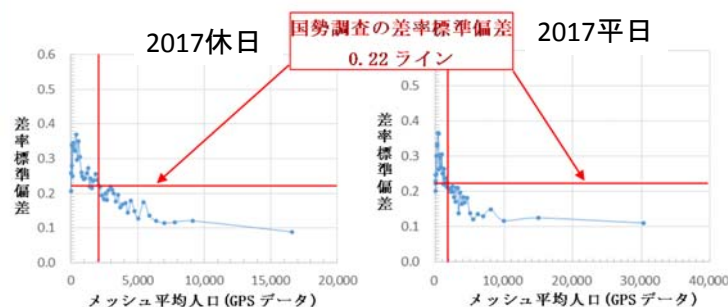
- 小人口層では差率マイナスのメッシュが多数存在し、人口規模の増大とともに差率がプラスのメッシュが増加する
- 人口規模別グループの差率標準偏差は、休日・平日ともGPSデータ平均人口が**2000人前後で、国勢調査基準（2010と2015の比較）と同レベルである**
  - **2000人以上で一定レベルの信頼性を持つ**

比較の基準とする国勢調査2010と2015の相関係数

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



GPSデータと基地局データとの比較におけるGPS人口と差率の散布図

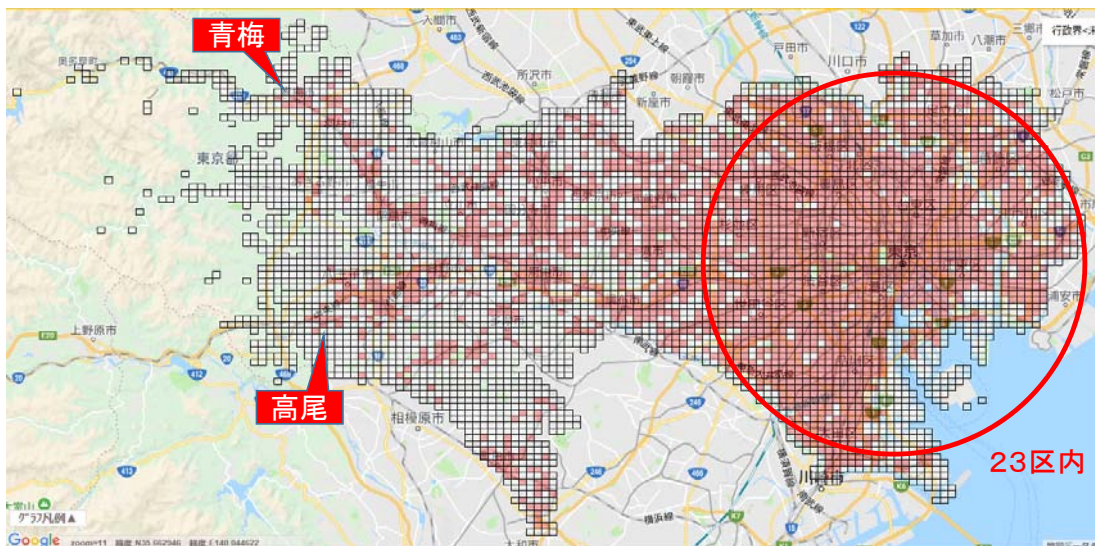


GPSグループ平均人口と差率標準偏差の散布図

30

## 3.2 基地局データとの比較結果\_2000人以上メッシュ

- 2000人以上のメッシュは23区を中心に鉄道沿線駅沿いであり、西端は西多摩エリアの青梅駅、高尾駅である
- メッシュ数は休日で2620個（全数5172）、平日で2400個（全数5129）



2017平日 2000人以上メッシュ分布

31

## 目次

1. はじめに
2. 検証データと検証方法
  - 2.1 検証概要
  - 2.2 国勢調査との比較方法
  - 2.3 基地局データとの比較方法
3. 検証結果
  - 3.1 国勢調査との比較結果
  - 3.2 基地局データとの比較結果
4. まとめ

32



## 4. まとめ～東京都の結果～

当GPSデータは国勢調査や基地局データと相関があり、国勢調査を補完するものとして有効であるが、以下の特徴に留意することで、より効果的に利活用を進めることができる

- 国勢調査の差率標準偏差を基準とすると、2000人以上のメッシュで一定レベルの信頼性を保有
- 時間帯別データは、国勢調査で捉えられていない余暇・消費活動による人の動きを捕捉
- 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で特に有効
- アプリ利用者属性に起因する標本の偏りが存在

東京都の検証結果であることに留意