

第7回日本統計学会春季集会

平成25年3月3日(日)

モバイル空間統計の信頼性評価

独立行政法人 統計センター
株式会社 NTTドコモ

○木村正一 榎田直木 坂下信之 上田聖
寺田雅之 ○小林基成 大藪勇輝

はじめに

【研究の背景】

- ICT技術の進化により大量データを処理するハード・ソフトの環境が整備
- ICT技術の利用などにより得られた各種データの統計情報作成への活用



NTTドコモにおけるモバイル空間統計の作成

NTTドコモの作成するモバイル空間統計の活用方策の検討 するため共同研究を実施

【研究の目的】

- 時代の変遷に伴って発生しつつある新たな統計ニーズの高まりに対応するために、国勢調査などをはじめとする各種の官庁統計とモバイル空間統計を組み合わせることにより、新たに有用な統計情報を創出すること



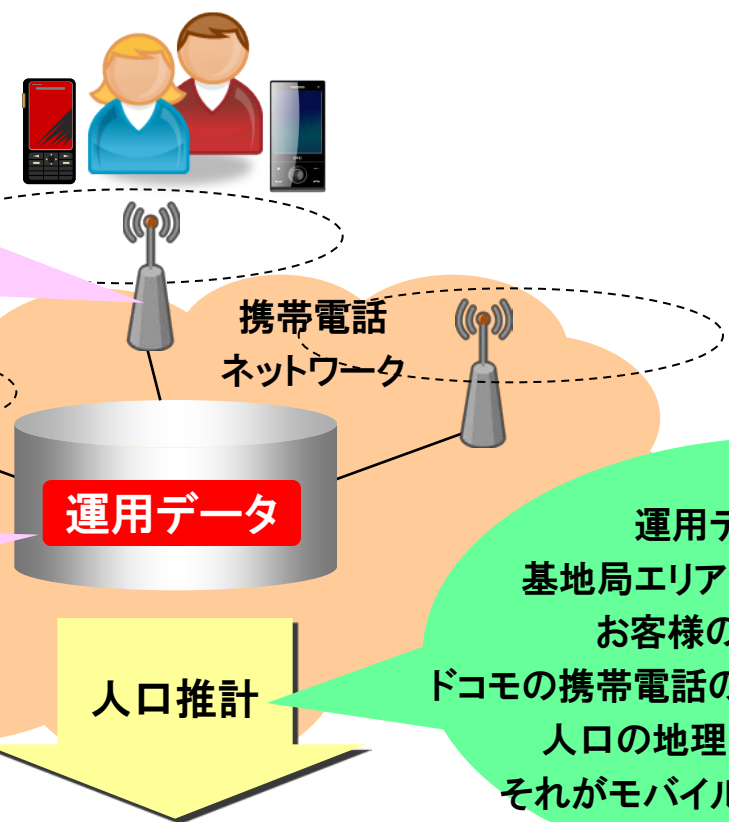
- モバイル空間統計を新たな統計として活用していく上で、十分な精度が確保されていることを検証
- モバイル空間統計を活用した新たな統計に関するニーズ把握と新たな統計作成方法の検討

モバイル空間統計の説明

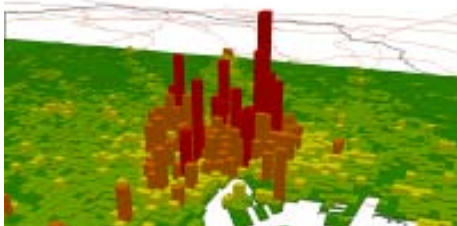
携帯電話ネットワークは、携帯電話がいつでもどこでも電話やメール等を着信できるように、各基地局のエリア毎に所在する携帯電話を周期的に把握しています。
(GPSではありません)

お客様に携帯電話サービスを提供するために必要となるこれらのデータやご契約のデータを運用データと呼びます。

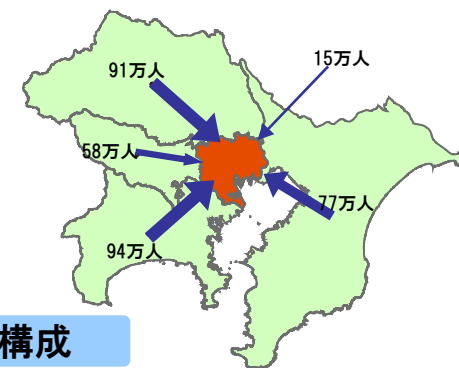
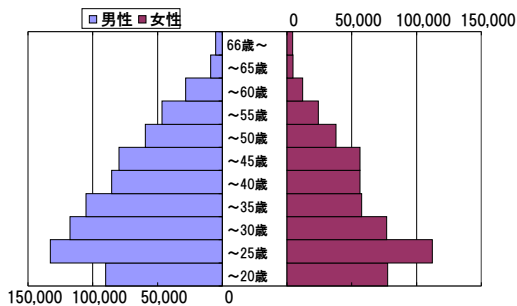
運用データを使って基地局エリア毎の携帯電話台数をお客様の属性別に数え、ドコモの携帯電話の普及率を加味することで、人口の地理的分布を推計でき、それがモバイル空間統計となります。



モバイル空間統計



人口分布



人口構成

人口分布

- 全国の人口分布のイメージ
- 東京都23区の人口分布のイメージ

人口の地理的な分布
がわかる

人口構成

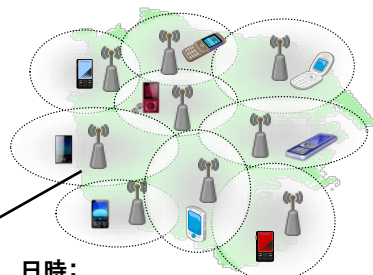
- 秋葉原駅周辺・原宿駅周辺の人口構成のイメージ
- 千代田区への流入人口のイメージ

性別・年齢別の人口
がわかる

居住エリア別の人口
がわかる

お客様のプライバシーを保護するために、非識別化処理、集計処理、秘匿処理といった3段階の処理を適切に実施して、モバイル空間統計を作成します。モバイル空間統計は、集団の人数のみをあらわす人口統計情報であるため、モバイル空間統計からお客様個人を特定することはできません。

運用データ



基地局#1、日時:
090-XXXX-、...、男性、S.43.10.11生、埼玉県A市B町1-2、....
090-YYYY-、...、男性、S.32.04.10生、千葉県C市D町3-4、....
....

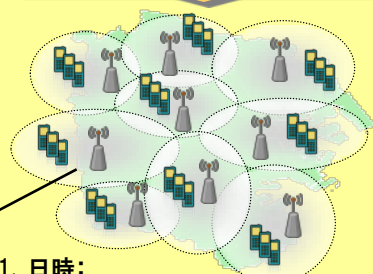
モバイル空間統計

(40歳台男性)

75人	120人	30人
90人	135人	105人
45人	60人	データ無し

非識別化処理

個人識別性の除去



基地局#1、日時:
男性、40歳台、埼玉県A市B町
男性、50歳台、千葉県C市D町

秘匿処理

少人数の除去

(40歳台男性)

75人	120人	30人
90人	135人	105人
45人	60人	6人

集計処理

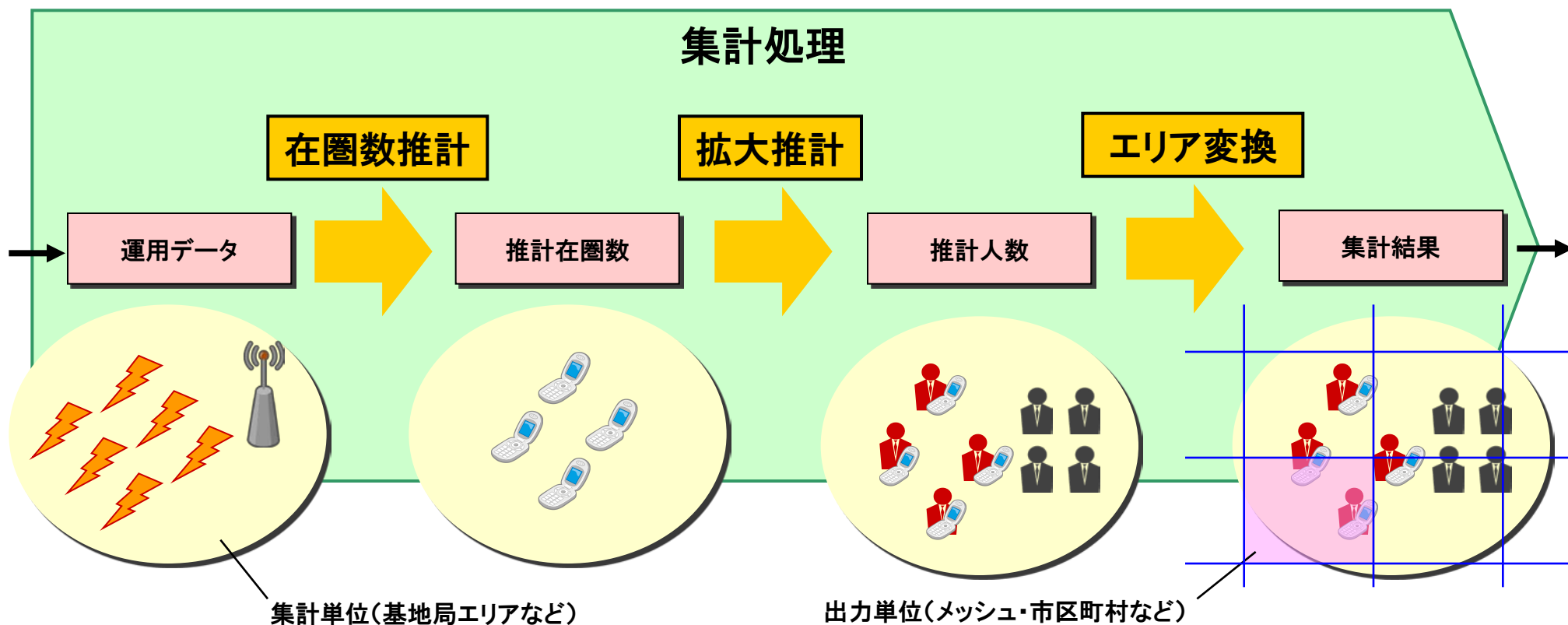
ドコモの携帯電話の普及率を
加味して人口推計

個人識別性のないデータ

推計人口

本作成処理の実施基準を、社外有識者による検討結果を参考にして策定した自主ルール「モバイル空間統計ガイドライン」としてコーポレートサイトにて公開しています

- 集計処理は、「在圏数推計」「拡大推計」の2段階の統計的推定(＋「エリア変換」)により実施される。
 1. 運用データからの、**集計単位**(基地局エリアなど)ごとの携帯電話数の推計(在圏数推計)
 2. 推計在圏数からの、在圏率(※1)や普及率(※2)を加味した、集計単位ごとの人数の推計(拡大推計)
 - ※1 在圏率:在圏数の契約台数に対する比率
 - ※2 普及率:契約台数の人口に対する比率
 3. 集計単位ごとの推計人数からの、**出力単位**(メッシュ, 市区町村など)ごとの集計結果への変換(エリア変換)

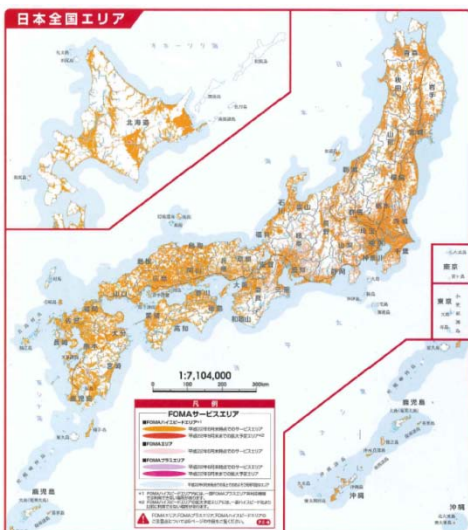


モバイル空間統計は、携帯電話ネットワークに依存したカバレッジ、空間解像度、時間解像度などの特性を有する人口統計情報です。

カバレッジ

全国市区町村役場を
100%カバー

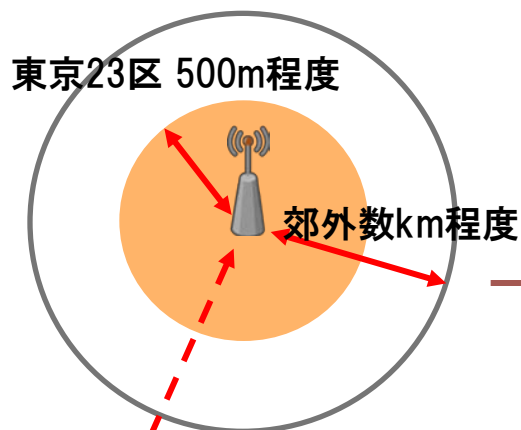
携帯電話のサービスエリアに依存



空間解像度

500mメッシュ(東京23区)
～数kmメッシュ(郊外)

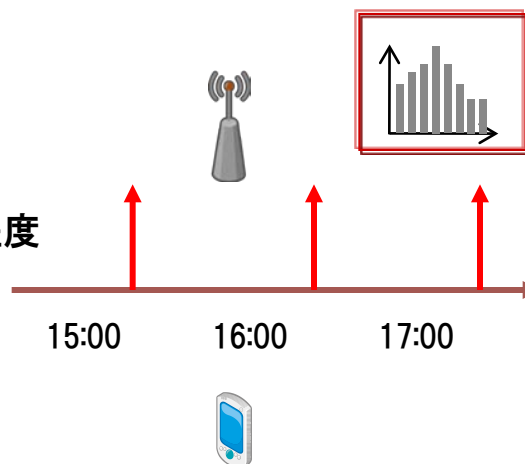
基地局の設置間隔に依存



時間解像度

1時間毎の平均人口を
継続的に

基地局がエリア内に所在する
携帯電話を把握する頻度に依存



対象

15歳～79歳

十分なサンプル数が
確保できる年齢層

～14歳

×対象外



15歳～79歳
○対象



80歳～

×対象外



※1つの集計単位において少ない人口データが生じないようにする秘匿処理のために、意図的に空間解像度および時間解像度を荒くすることがあります。

信頼性評価

モバイル空間統計が、実際の人口に対してどの程度の誤差で推計可能か評価する

モバイル空間統計

5800	1250	1020
5000	1330	1580
2080	6070	2000

実際の人口

6000	1000	1200
4700	1020	1580

—

=

誤差

-200	250	-180
300	310	0
200	-500	-10



既存の人口統計の中で、信頼性が高い国勢調査を実際の人口に見立てて、国勢調査とモバイル空間統計の「差異」を見ることで評価を行う
モバイル空間統計は、多くの人々が在宅している夜間の時間帯のものを比較する

国勢調査

すべての人は自宅でカウントされる



自宅



外出先

モバイル空間統計 (夜間時間帯)

多くの人々が在宅しており自宅でカウントされる

夜間業務、旅行等一部の人は外出先でカウントされる



自宅



外出先

モバイル空間統計 (昼間時間帯)

多くの人々が外出しており外出先でカウントされる



自宅



外出先

比較



国勢調査とモバイル空間統計午前4時人口を比較する

■ 比較データ

- 国勢調査:2010年国勢調査地域メッシュ統計
 - 15~79歳人口
 - 1kmメッシュ、500mメッシュ
- モバイル空間統計:2010年10月1日午前4時時点人口
 - 15~79歳人口
 - 1kmメッシュ、500mメッシュ

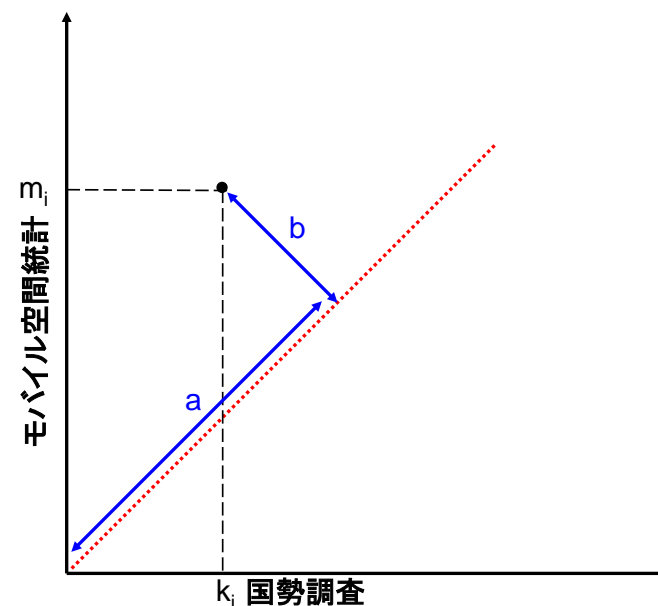
■ 比較範囲

- 全国のメッシュ
 - ただしモバイル空間統計100人以上、国勢調査1人以上

■ 比較指標

- 国勢調査とモバイル空間統計のメッシュあたりの人口の差異を下記の式で表される「偏差率」で評価
- モバイル空間統計と国勢調査がどのくらいの割合で違うか -1~1 の値で表現

$$\text{偏差率} = \frac{\text{モバイル空間統計} - \text{国調人口}}{\text{モバイル空間統計} + \text{国調人口}} = \frac{b}{a}$$

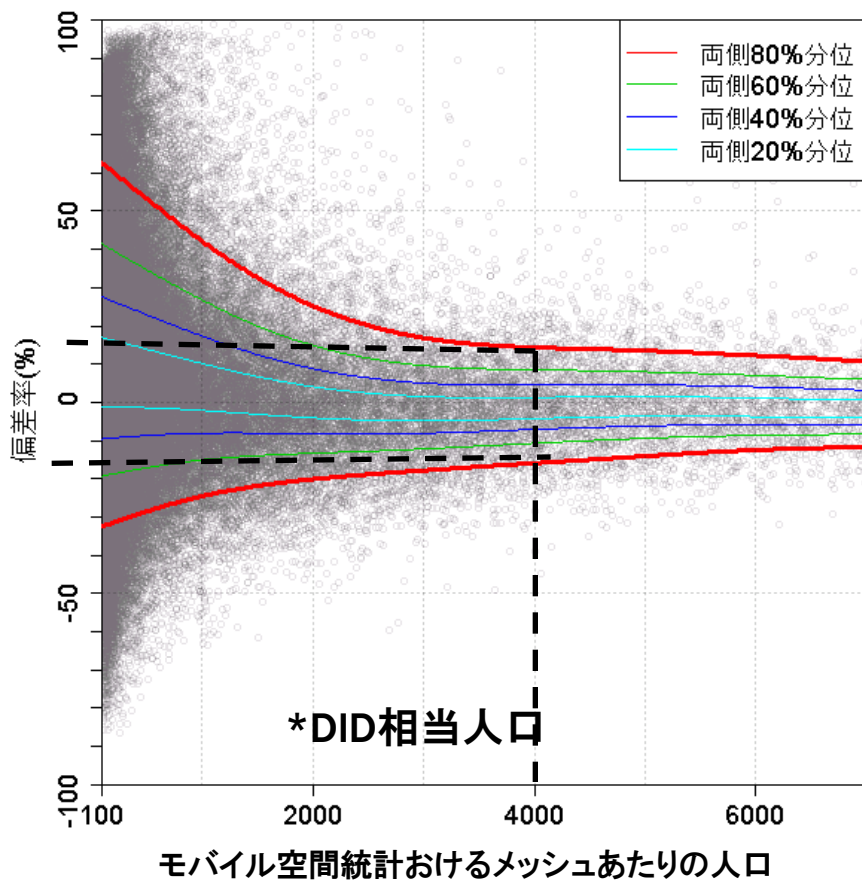


- 人口密度による偏差率の傾向
 - 空間解像度が基地局の設置間隔に依存し、人口密度が多いエリアほど密に基地局が存在する
 - そのため人口の大小によりモバイル空間統計の信頼性は異なる
 - 人口の大小により国勢調査とモバイル空間統計の差異の大きさにどのような傾向が見られるか検証する

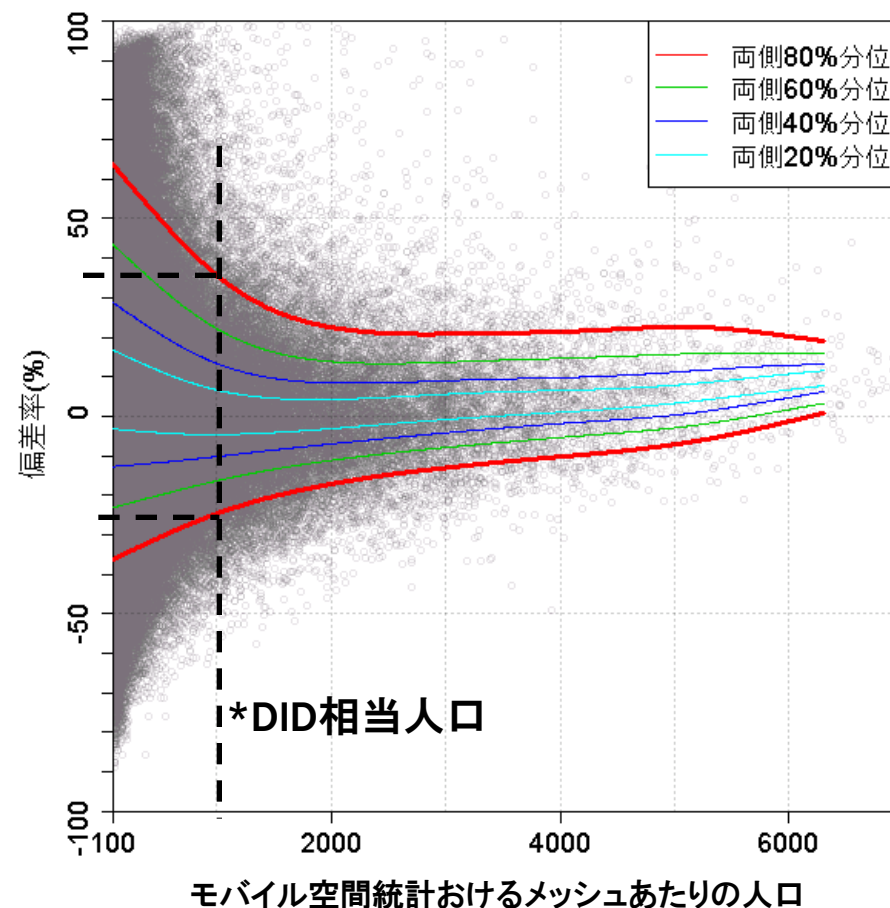
- 地理的な偏差率の分布の傾向
 - 地理的に偏差率の大きい地域、小さい地域に特徴的な偏りはあるか検証する

人口密度が高いところほど偏差率は小さくなる傾向がある
DID相当の人口のエリアにおいて500mメッシュにおいては偏差率±15%程度である
1kmメッシュにおいては、偏差率±30%程度である

1kmメッシュ



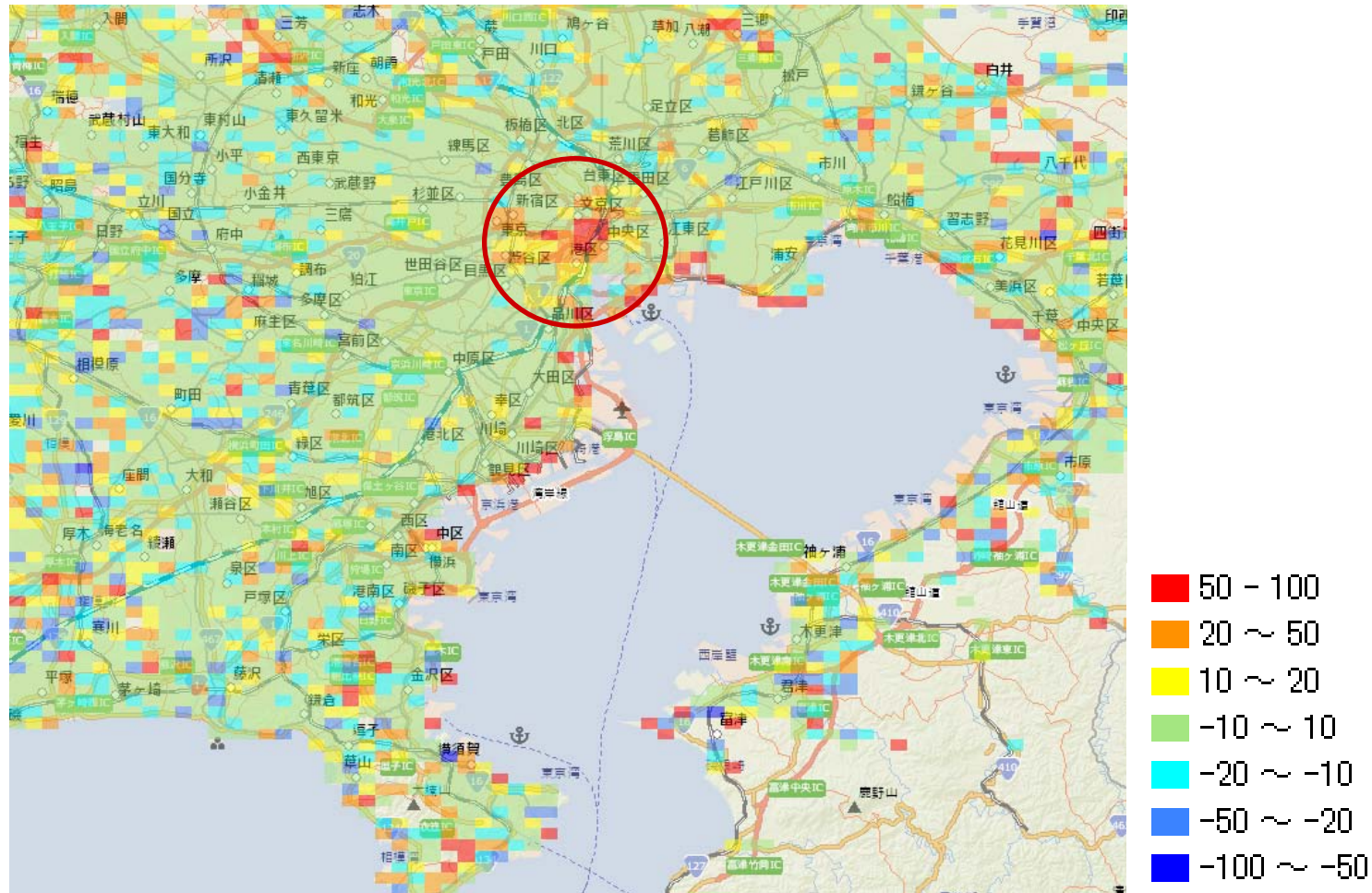
500mメッシュ



*DID(人口集中地区)相当人口:人口密度4000人/km²以上

都心部などでは、モバイル空間統計が国勢調査に比べて人口を大きく推計する傾向にある

国勢調査とモバイル空間統計の比較(1kmメッシュ)



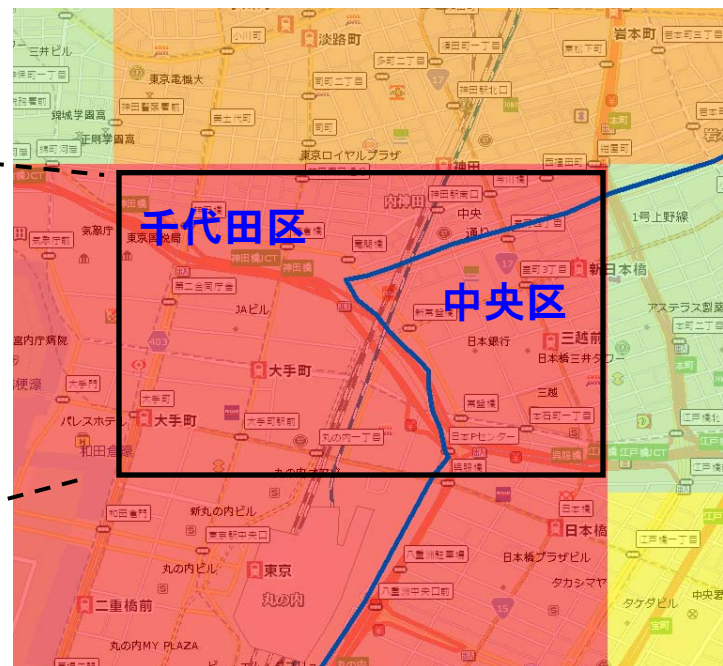
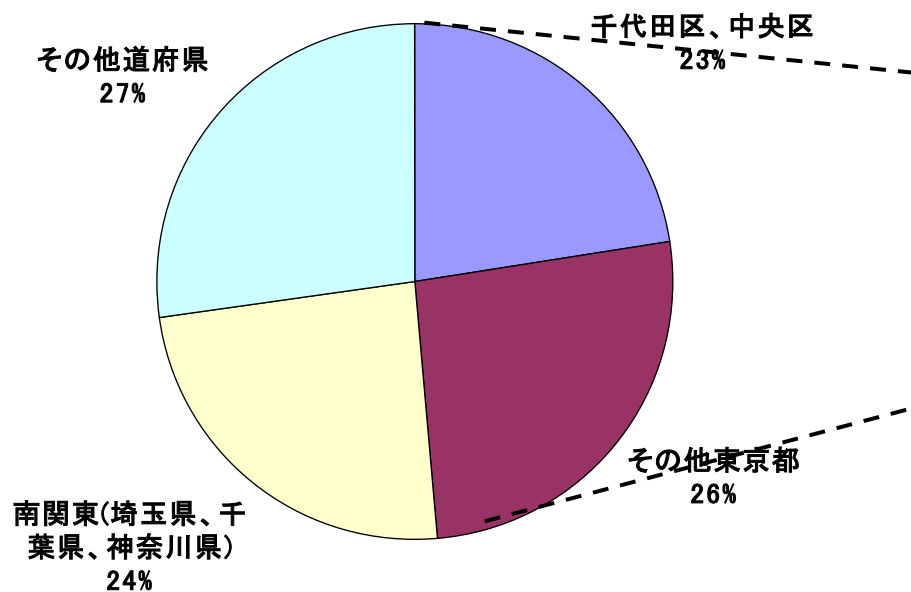
(c) Esri Japan, ZENRIN CO.,LTD.

都心部においてモバイル空間統計が国勢調査に比べて大きくなるのは、夜間においても他地域から人が集まっているためである

国勢調査人口:1369人

モバイル空間統計:7071人
(内、千代田区、中央区居住者1591人)

大手町を含むメッシュの モバイル空間統計(居住エリア別人口)



午前4時のモバイル空間統計

- 国勢調査に基づく人口は居住地に基づいた人口であり、モバイル空間統計はあるときにいた場所に基づいた人口であり、位置づけが異なる
 - ただし、夜間時間帯においては、多くの人々が在宅していると考えられ、モバイル空間統計午前4時のデータと国勢調査人口を比較した
- 人口密度が高いところほどモバイル空間統計と国勢調査との差は小さい
 - 人口が密なところは基地局の設置間隔も密であるためと考えられる
- DID相当の人口のエリアにおいて500mメッシュにおいては偏差率±15%程度、1kmメッシュにおいては、偏差率±30%程度である
 - 500mメッシュの方がより高い人口密度のエリアでなければ信頼性を確保できない
- 都心部では、モバイル空間統計が国勢調査に比べて人口が多くなっている
 - 夜間も労働・娯楽・宿泊などで人が集まっており、モバイル空間統計が夜間の人の集まりを捉えているためと考えられる

モバイル空間統計を活用した新たな統計への ニーズヒアリング

■ 目的

- モバイル空間統計を活用した新たな統計に期待される要件を抽出する

■ 実施対象

- 公的統計を活用している企業 3団体
- コンサルタント 1団体
- 自治体・大学 2団体

■ 実施方法

- 各団体へ訪問
- モバイル空間統計の説明
- モバイル空間統計に対する利用法、要求ヒアリング
- モバイル空間統計+他の統計などで実現性のある統計へのニーズヒアリング

■ 総括的意見

- いずれも極めて強い関心を示し、高い潜在需要があることが示唆された
- 現在のモバイル空間統計の仕様(時間解像度、空間解像度、属性)はほぼ妥当だが、一部により細かな情報への要望があった
- 統計の精度に対する具体的な要望はなかった

■ 個別の意見等

1 精度について

- ・数値の精度への明確な要求はないが、ある程度の精度は要求される

2 空間解像度について

- ・できるだけ細かい情報がほしいという意見が多い
- ・ビル単位、ビルの階層単位、道路の路線単位などへの要求がある

3 時間解像度について

- ・1時間で十分という意見が多い
- ・休日・平日の別や季節別などへの要求がある

4 属性区分について

- ・年齢については、5歳区分、10歳区分、低／中／高年齢の3区分など利用目的による

5 居住エリア別人口について

- ・市区町村別より細かい、数km程度の範囲や町丁字別の情報が欲しい
- ・これまでにない統計なので、ある程度の精度があれば十分活用できる

6 提供形態について

- ・オンラインでの提供
- ・データだけでなく、アプリケーションを付属し簡便に利用できる形態での提供

7 新しい統計へのニーズ例

- ・ 帰宅困難者把握
- ・ 災害時の交通手段情報の提供
- ・ 商圈調査や出店判断
- ・ 地価データとの組み合わせによる町のポテンシャルの計測
- ・ 類似する他都市との比較による政策立案

まとめ・今後の課題

統計センターとNTTドコモは、官庁統計とモバイル空間統計を組み合わせ、新たに有用な統計情報を創出することを目的に共同研究を実施

モバイル空間統計と国勢調査メッシュ統計との比較検証を行った結果、概ね一致しており大きな偏りはみられない

今後、ニーズヒアリングを通して得た利用方法に適した提供のあり方について、具体的な方策を検討していく予定