

**2011 年度 修士論文**

**発展途上国の諸都市における交通統計を用いた  
人の流れデータセット構築手法の開発**

**Development of People Flow Data Set Using Transportation Statistics  
in Cities in Developing Countries**

**渡邊 淳人**

**Watanabe, Atsuto**

**東京大学大学院 新領域創成科学研究科  
社会文化環境学専攻**

## 概要

近年、防災・環境・マーケティング・交通計画等の観点から、都市空間における日々の人々の流れを把握することが重要になってきている。そこで、東京大学空間情報科学研究センター（以下、CSIS）では、「人の流れプロジェクト」を立ち上げ、パーソントリップ調査（以下、PT 調査）等の時空間データから人の流れの再現に取り組んでいる。特に、近年では、独立行政法人国際協力機構（以下、JICA）が行った PT 調査（以下、JICA-PT 調査）を用いた時空間内挿をすることにより、アジア・アフリカなどの発展途上国での人の流れの再現にも取り組んでいる。しかし、国内で確立された手法を発展途上国の諸都市に適用すると、「人の流れデータセット」の精度が低下することがある。それは、これらの都市では、日本国内のように、時空間内挿時に必要な道路ネットワークデータ等が充実していないことが多く、時空間内挿が困難であるからだ。そこで、本研究では、グローバルで利用可能なデータと汎用的な統計的手法を用いて、高精度な「人の流れデータセット」を構築する手法を開発することを目的としている。

# 目次

<b>第1章</b>	<b>はじめに</b>	<b>1</b>
1.1	研究の背景	1
1.2	研究の目的	2
1.3	本研究のアプローチ	2
1.4	本論文の構成	3
<b>第2章</b>	<b>既往の研究事例</b>	<b>4</b>
2.1	人の流れデータセットに関する既往研究	4
2.2	時間的平滑化に関する既往研究	6
2.3	空間的平滑化に関する既往研究	7
2.4	道路ネットワーク整備に関する既往研究	7
<b>第3章</b>	<b>研究対象都市と使用するデータ</b>	<b>9</b>
3.1	研究対象都市	9
3.2	使用するトリップデータ	10
3.2.1	JICA-PT データ	10
3.2.2	JICA-PT データの拡大係数	10
3.3	道路ネットワークデータ	13
3.3.1	デジタル道路地図 (DRM)	13
3.3.2	商用地図データセット	14
3.3.3	GoogleMaps	14
3.3.4	OpenStreetMap	14
3.3.5	まとめ	17
3.4	土地被覆・人口に関するデータ	17
3.4.1	LandScan	17
3.4.2	MCD12Q1	20

<b>第4章 人の流れデータセット構築手法</b>	<b>21</b>
4.1 人の流れデータセット構築手法のフレームワーク	21
4.2 時間的平滑化手法	22
4.2.1 カーネル密度推定	23
4.2.2 バンド幅選定手法	23
4.3 空間的平滑化手法	25
4.3.1 土地被覆データと人口データを用いた重み算出	26
4.3.2 重みつきジオコーディング	27
4.4 道路ネットワーク補正手法	28
4.4.1 道路ネットワークデータの品質評価	30
4.4.2 ノードマッチング処理	30
4.5 時空間内挿	31
4.6 データセットの構築	32
<b>第5章 適用実験</b>	<b>35</b>
5.1 時間的平滑化手法の適用	35
5.2 空間的平滑化手法の適用	38
5.3 道路ネットワークデータ補正手法の適用	39
5.4 人の流れデータセット構築実験	42
5.4.1 人の流れデータセット構築の概要	42
5.4.2 人の流れデータセットの考察	43
5.4.3 活用事例	43
5.4.4 まとめ	55
<b>第6章 結論と今後の展望</b>	<b>57</b>
6.1 本研究の成果	57
6.2 今後の課題	58
6.2.1 時間的平滑化手法の改良	58
6.2.2 空間的平滑化手法の改良	59
6.2.3 道路ネットワーク補正手法の改良	59
6.2.4 時空間内挿処理手法の改良	59
6.2.5 人の流れデータセットの精度検証	59
6.3 今後の展望	60



6.3.1	他都市への展開 . . . . .	60
6.3.2	他データとの統合 . . . . .	60
6.3.3	活用範囲の拡大 . . . . .	60
<b>謝辞</b>		<b>61</b>
<b>参考文献</b>		<b>62</b>

# 表 目 次

2.1	人の流れデータセットの仕様 . . . . .	5
2.2	カーネル関数と効率性 . . . . .	6
3.1	JICA-PT データの概要 (サンプル数・トリップ数など) . . . . .	12
3.2	道路ネットワークデータの特徴 . . . . .	17
4.1	土地被覆ポイントの持つ属性情報 . . . . .	27
4.2	OSM のリンク数・ノード数と接続状況 . . . . .	28
4.3	OSM データの品質評価の結果 . . . . .	31
4.4	時空間内挿の結果得られる情報 . . . . .	33
4.5	人の流れデータセットサンプル . . . . .	34
5.1	各バンド幅選定手法で算出されたバンド幅と処理時間 . . . . .	36
5.2	各バンド幅選定手法の比較 . . . . .	36
5.3	ノードマッチング処理前後の比較 . . . . .	39
5.4	経路探索の結果の評価 . . . . .	41
5.5	OSM データの品質評価の結果 (補正処理後) . . . . .	41
5.6	データセットの構築を行った Case . . . . .	42
5.7	補正済み JICA-PT データの概要 . . . . .	42
5.8	LandScan データとの比較 (相関係数・総数) . . . . .	44
5.9	トリップデータに含まれる人口 . . . . .	56

# 目次

2.1	人の流れデータセット提供状況	4
2.2	OSM で定義されているノード・リンク・クローズドリンク	8
3.1	Manila の調査エリア (航空写真)	9
3.2	Cairo の調査エリア (航空写真)	9
3.3	Jakarta の調査エリア (航空写真)	9
3.4	Ho Chi Minh の調査エリア (航空写真)	9
3.5	Hanoi の調査エリア (航空写真)	10
3.6	拡大係数の算出フロー (Ho Chi Minh)	11
3.7	Potlatch を使った OSM 編集画面	16
3.8	OSM のノード数・リンク数の変遷	16
3.9	LandScan データ (東京)	18
3.10	国勢調査と LandScan 人口分布の比較	19
3.11	国勢調査と LandScan 人口の比較	19
3.12	MCD12Q1 (東京)	20
4.1	本研究のアプローチ	21
4.2	出発時刻のヒストグラム (ハノイ)	22
4.3	時間的平滑化のイメージ	22
4.4	午前 9:00 における人口分布 (マニラ, 既存手法)	25
4.5	空間的平滑化のイメージ	26
4.6	OSM グループ分けの手順	29
4.7	OSM を用いた経路探索の結果例	29
4.8	OSM グループ分けの結果	30
4.9	ノートマッチング処理の例	31
4.10	経路探索の手順	32
4.11	時空間内挿の手順	32

5.1	出発時刻の平滑化の結果 (Manila)	35
5.2	出発時刻の平滑化の結果 (Ho Chi Minh)	35
5.3	空間的平滑化前後でのトリップ 発生場所の分布の比較 (Manila)	38
5.4	OSM グループ分けの結果 (左図：補正前, 右図：補正後)	39
5.5	補正前後 OSM を用いた経路探索の結果の比較 (例 1)	40
5.6	補正前後 OSM を用いた経路探索の結果の比較 (例 2)	40
5.7	処理前・後における迂回率の比較 ((a) Hanoi, (b) Manila, (c) Jakarta)	42
5.8	LandScan メッシュによる 24 時間平均の人口集計 (Manila)	44
5.9	LandScan メッシュによる 24 時間平均の人口集計 (cairo)	45
5.10	LandScan メッシュによる 24 時間平均の人口集計 (jakarta)	45
5.11	LandScan メッシュによる 24 時間平均の人口集計 (hcmc)	46
5.12	LandScan メッシュによる 24 時間平均の人口集計 (hanoi)	46
5.13	交通機関別に色分けした人の分布 (Manila)	48
5.14	交通機関別に色分けした人の分布 (Cairo)	49
5.15	交通機関別に色分けした人の分布 (Jakarta)	50
5.16	交通機関別に色分けした人の分布 (Ho Chi Minh)	51
5.17	交通機関別に色分けした人の分布 (Hanoi)	52
5.18	選択した場所 (Cairo)	53
5.19	選択した場所 (Hanoi)	53
5.20	選択された場所での人数の時間変化 (Cairo)	54
5.21	選択された場所での人数の時間変化 (Hanoi)	54

# 第1章 はじめに

## 1.1 研究の背景

近年，防災・マーケティング・交通など様々な観点から，人の流れを把握することが重要になってきている．また，人の流れに関するデータも増えてきている．例えば，携帯電話の通信ログデータや GPS ロガーで取得されるログデータがあげられる．主に携帯電話などのモバイル通信機器の小型化・汎用化や携帯電話に GPS の搭載を義務づける等の動向もあり，これらのログデータを大量に取得できるようになってきている．これらのログデータを活用すれば，理論的には，人の流れをかなり正確にかつ絶えず，把握することが可能である．また，ログデータをすぐに取得でき，データ処理を行うことができれば，リアルタイムに人の流れを把握することも可能である．しかし，これらのログデータは個人情報であるため，データの使用にあたりユーザーの許可が必要となる．すべてのユーザーから許可を得ることは不可能に近く，実際にはこのようなログデータを用いて，すべての人の流れを把握することは難しい．

また，交通統計も人の流れを把握することのできる有効なデータである．交通統計は調査によって取得されるため対象が限られてしまううえに，調査をしてからデータとして整理されるまでに時間がかかることが多い．これらの要因から，リアルタイム性に乏しいといった欠点があげられる．しかし，多くの交通統計では確立されたサンプリング手法を用いており，拡大係数などが設けられている．この拡大係数を用いて，統計を拡大することで，理論的には調査対象とする地域の全ての人口をカバーでき，かなり現実に近い人の流れが把握することも可能であると考えられる．

このような交通統計として，パーソントリップ調査（以下，PT 調査）がある．パーソントリップ調査は，都市内での人の交通を把握するために実施される調査であり，国内では，東京をはじめとする多くの都市で，50 年以上の間実施されている．この調査から得られるパーソントリップデータ（以下，PT データ）は，都市内での人の流れを把握する研究に多く利用されている．この PT 調査は国内だけでなく海外でも実施されている．JICA がアジア・中東・アフリカを中心に PT 調査を実施しており，1966 年から開始され，2004 年までにアジア・アフリカなどの 52 都市で実施された．

このような様々なデータ・統計は、どれも有効なデータであるが、形式・品質・内容などがそれぞれで異なる。そこで、これらのデータを処理・統合し、人の流れを把握に特化した「人の流れデータセット」を構築することが有益であると考えられる。この取り組みは既に日本国内では始まっており、CSIS では、研究目的に限ったデータセットの提供が開始されている。[25]

## 1.2 研究の目的

既にデータ提供も開始している日本国内の PT データに適用されている既存の人の流れデータセット構築の手法をそのまま JICA-PT 調査から得られる PT データ（以下、JICA-PT データ）に適用するといくつかの問題が発生する。PT データを用いた人の流れデータセット構築には、PT データの他に、道路ネットワークデータや人口分布のデータ住宅地図などが必要である。国内においては、これらのデータが高品質な上に充実している。しかし、JICA-PT データの対象地域であるアジア・アフリカなどの発展途上国諸都市においては、これらのデータは品質が保証されていなかったり、整備がされていないことがある。これらのデータの品質の低さや不足により、都市内での人の流れを詳細に正確に把握することのできる人の流れデータセットの構築が難しいことがある。

人の流れデータセットとは、人の流れを詳細に様々な側面から把握することができるデータセットであり、日本国内では伝染病、需要予測、災害リスク予測など様々な活用がなされている。本研究では、グローバルで利用可能なデータを用いてデータ補間・補正を行うことで、JICA-PT データから人の流れデータセットの構築をする手法を提案することを目的としている。

## 1.3 本研究のアプローチ

既存の手法を JICA-PT データに適用すると発生する問題は、大きく分けて以下の 3 点があげられる。

1. 時間的な偏り
2. 空間的な偏り
3. 道路ネットワークデータの品質

これら3点の問題を解消し、より人の流れを正確に把握することのできる人の流れデータセットを構築する。時間的な偏りを解消するために時間的平滑化を、空間的な偏りを解消するために空間的平滑化を、道路ネットワークデータの品質を解消するために、道路ネットワークデータ補正を行った。各手法の詳細については第4章で述べる。

## 1.4 本論文の構成

本論文は以下のように構成されている。

第2章：既往の研究事例

第3章：研究対象都市と使用するデータ

第4章：人の流れデータセット構築手法

第5章：適用実験

第6章：結論と今後の展望

第2章では、人の流れデータセット、時間的平滑化、空間的平滑化、道路ネットワークデータ補正に関する既往の研究事例について述べる。第3章では、本研究の対象都市と使用するデータについて述べる。そして、第4章で本研究で提案する人の流れデータセットの適用手法について述べ、第5章で適用実験を行う。最後に、第6章で本研究の結論と今後の課題について述べる。

## 第2章 既往の研究事例

### 2.1 人の流れデータセットに関する既往研究

CSIS がたちあげている人の流れプロジェクト [25] では PT データを使った人の流れデータセットの開発が行われてきた。例えば、関本ら [21] は、PT データをもとに、様々なデータ処理を行い時空間的に詳細化することにより、人の流れを知りたい様々なニーズに対応することを目的に、その処理のプロセス、精度、時間的なコストを明らかにしつつ、動画画像や Web API の作成を行い、人の流れデータセットの利用の方向性を示している。また、東京都市圏をはじめとする日本国内の 14 の都市圏における人の流れデータセットの提供するデータ提供サービスを開始している。このサービスにより、指定した検索条件における人の位置・時間情報を取得することができる。近年では、人の流れデータセット構築の取り組みの場を海外にも広げている。[16]

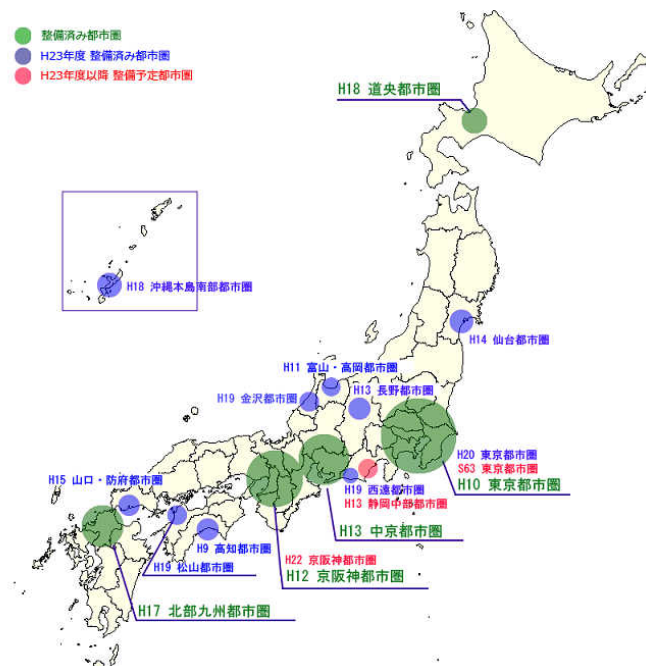


図 2.1: 人の流れデータセット提供状況

また、この人の流れデータセットは、人の流れを様々な側面から把握することのできる



表 2.1: 人の流れデータセットの仕様

名称	フィールド名	形式	概要
パーソン ID	PID	Char	パーソン ID (一意に決まっている)
トリップ番号	TNO	Integer	トリップ番号
サブトリップ番号	SNO	Integer	トリップ番号
日時	PDATE	Date	日付の形式は yyyy/mm/dd HH:MM:SS
経度	LON	Double	経度: 小数点含む [世界測地系 (WGS84) ]
緯度	LAT	Double	緯度: 小数点含む [世界測地系 (WGS84) ]
性別コード	SEX	Char	性別コード (API 仕様書参照)
年齢コード	AGE	Char	年齢コード (API 仕様書参照)
住所コード	PADD	Char	住所コード (各調査に依存)
職業	WORK	Char	職業コード (API 仕様書参照)
目的コード	PURPOSE	Char	移動の目的コード (API 仕様書参照)
拡大係数	MAGFAC	Integer	拡大係数
拡大係数 2	MAGFAC2	Integer	拡大係数が二つある場合に記載
交通手段コード	DATUM	Char	移動の交通手段コード (API 仕様書参照)

データセットであり、人の流れデータセットを使った様々な分野の研究がなされている。島崎ら [23] は、人の流れデータセットから推定される時間帯別人口が、店舗数や店舗数の変化とどのような関係にあるのか、広域的かつ他業種に渡って分析している。CSIS との共同研究として、消費者行動のモデルを検討する研究や旅客流動の再現といった交通分野の研究以外にも多岐にわたる分野で人の流れデータセットを用いた研究がされている。[25] 例えば、江島ら [22] は、人の流れデータセットを用いて、実際の人の流れを想定した感染症伝播の様子をシミュレーションし、感染症が感染拡大の際の対応策や輸送制限など対応策の効果を評価している。正木ら [24] は、人の流れデータセットを用い、予想震度、液状化危険度、浸水危険度などの災害リスク情報の重ね合わせを行い、「いつ、どんな場所にいたらこんな危険にあうかもしれない」といったピンポイントな災害リスクを予測している。

## 2.2 時間的平滑化に関する既往研究

PT 調査のようなアンケート調査では時刻表記に偏りがあり，トリップの発生時刻は連続的ではなく不連続な分布となっているうえに，きりの良い時刻に極端に偏る傾向がある．このような偏りのある時間データの傾向を把握したり，この偏りを誤差として考えて真の連続的な分布を知りたい場合には，密度推定を行う必要がある．密度推定に関する研究は多く存在するが，B. W. Silverman ら [18] が行っているカーネル密度推定がよく用いられている．カーネル密度推定を行う際には，カーネル関数とバンド幅を設定する必要がある．カーネル関数には，Epanechnikov, Gaussian, Rectangular などの関数が知られている．各関数は表 2.2 のように定義され，Epanechnikov が最も効率性の高いカーネル関数として知られている．[18] しかし，実際には計算式がシンプルであり計算がしやすい Gaussian 関数が使われることが多い．このカーネル密度推定の適用分野は様々であり，経済学の分野では世界の所得の分布の推定に使われていたり [8]，神経生理学実験における，外界刺激や動物行動と神経細胞発火頻度との関係の分析や推定に使われている例もある．[17] 時空間データでは，密度地図作成のための空間的補間に使われることがあるが，時間データの偏りを平滑化するためにカーネル密度推定が使われている例はほとんどない．

表 2.2: カーネル関数と効率性

カーネル関数	$K(t)$	効率性
Epanechnikov	$\frac{3}{4}(1 - \frac{1}{5}t^2)/\sqrt{5}$ ( $ t  < \sqrt{5}$ ) 0 その他	1
Biweight	$\frac{15}{16}(1 - t^2)^2$ ( $ t  < 1$ ) 0 その他	$(\frac{3087}{3125})^{1/2} \approx 0.9939$
Triangular	$1 -  t $ ( $ t  < 1$ ) 0 その他	$(\frac{243}{250})^{1/2} \approx 0.9859$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-(1/2)t^2}$	$(\frac{36\pi}{125})^{1/2} \approx 0.9512$
Rectangular	$\frac{1}{2}$ ( $ t  < 1$ )	$(\frac{108}{125})^{1/2} \approx 0.9295$

( [18] より抜粋 )

カーネル密度推定において，バンド幅の選定は結果を大きく左右する重要な要素であり，Plug-in 法，Silverman の経験則など様々なバンド幅選定手法がある．しかし，多くの研究ではこれらの選定手法によって選定された値に，任意で設定するパラメータを掛け合わせ

てそれをバンド幅としていることが多い。この場合、任意で設定するパラメータ自体が結果に影響を大きく与えることがある。このパラメータを設定することなくバンド幅を選定する手法もいくつか提案されている。例えば、[18]では、標準偏差を用いて単峰形の分布を真の分布と想定して密度を推定する手法、またこの手法を拡張し、四分位範囲を用いて双峰形の分布に対応した手法や、クロスバリデーション法とコスト関数を用いた手法などがある。また、[17]では、ポアソン分布を真の分布として想定し、コスト関数を用いて平均二乗誤差が最小となるようなバンド幅を選定する手法が提案されている。

## 2.3 空間的平滑化に関する既往研究

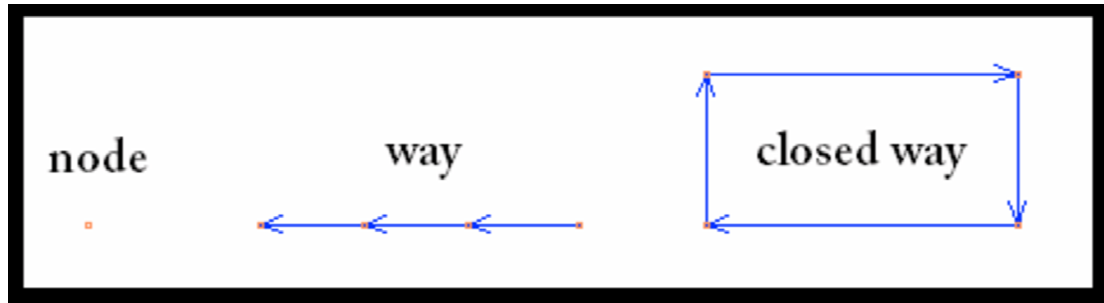
地理情報科学における点データは、基本的でありよく使われる時空間データである。この点データを扱う際、点データの数が多すぎたりノイズを含んだりしている場合、点データそのものからは知りたい情報を得られないことがある。このような場合に、空間的平滑化をすることで点データの分布をうまく把握することができる。空間的平滑化をする方法としては、ある空間的な単位で集計する方法や、カーネル法を用いる手法 ([18]) 等が知られている。

地図の平滑化の影響を分析している例として [27] があり、点分布の平滑化地図に焦点を当て、平滑化地図の表現方法の評価を行っている。このような研究は、多くの点データの傾向を把握するために作成される平滑化地図に対する人間知覚に関する研究であり、偏りがある時空間データを対象にしている研究事例はあまり見られない。

## 2.4 道路ネットワーク整備に関する既往研究

本研究で道路ネットワークデータとして使用している OpenStreetMap (以下、OSM) [13] は、ボランティアによって整備されているプロジェクトであり、道路ネットワークデータの品質を確保するためにマッピングのルールを作成し、遵守を呼びかけてはいるが、道路ネットワークデータの品質に問題があることがある。Haklay ら [7] では道路ネットワークデータの品質として、系列 (Lineage)、位置の精度 (Positional accuracy)、属性情報の精度 (Attribute accuracy)、論理的一貫性 (Logical consistency)、完全性 (Completeness)、意味的精度 (Semantic accuracy)、用途・目的および制約条件 (Usage, purpose, and constraints)、時間的品質 (temporal quality) の基準で評価できるとしている。本研究で行う経路探索に必要な道路ネットワークデータの品質は、トポロジー構造が保持されているかどうかの論理

的一貫性が最も重要である。しかし、この基準で OpenStreetMap の道路ネットワークデータの品質を評価、または補正を行っている研究事例は少ない。



(OSM ホームページより抜粋)

図 2.2: OSM で定義されているノード・リンク・クローズドリンク

## 第3章 研究対象都市と使用するデータ

### 3.1 研究対象都市

入手可能な ICA-PT データのうち、本研究では Manila (フィリピン), Cairo (エジプト), Jakarta (インドネシア), Ho Chi Minh City (ベトナム), Hanoi (ベトナム) の 5 都市を対象としている。以下に、PT 調査エリアの地図を示す。



図 3.1: Manila の調査エリア (航空写真)

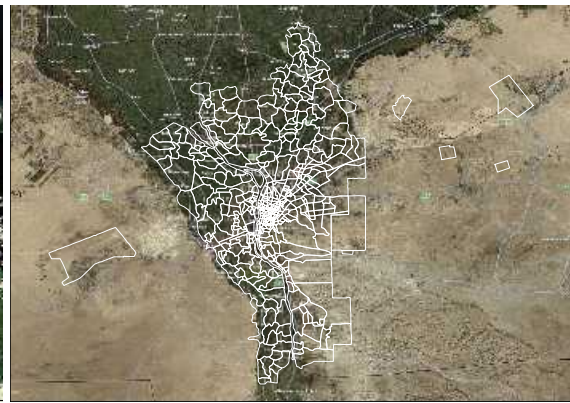


図 3.2: Cairo の調査エリア (航空写真)



図 3.3: Jakarta の調査エリア (航空写真)

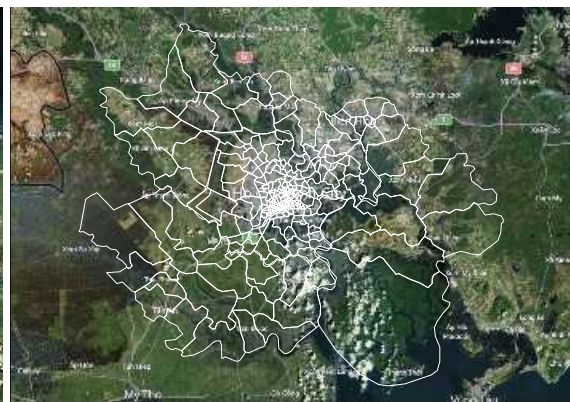


図 3.4: Ho Chi Minh の調査エリア (航空写真)

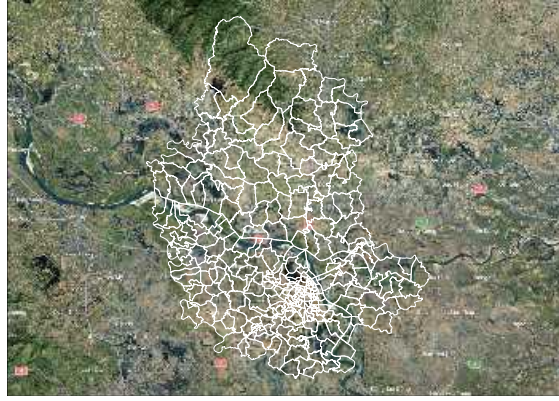


図 3.5: Hanoi の調査エリア（航空写真）

## 3.2 使用するトリップデータ

### 3.2.1 JICA-PT データ

本研究では、JICA-PT データを使用した。JICA-PT データとは、JICA が実施しているパーソントリップ調査から得られたデータである。この調査は、1966 年から開始され、2004 年までにアジア・アフリカなどの 52 都市で実施されている。そのうち 15 都市の JICA-PT データが、JICA から利用許可を受けた研究者などに提供されている。本研究では、JICA-PT 調査のうち世帯や世帯構成員の属性、トリップの情報を含む House Hold Interview の結果を JICA-PT データと呼ぶことにする。以下の表 3.1 にデータの概要を示す。JICA-PT データに含まれている拡大係数を用いて JICA-PT データを拡大すると、対象都市全体をカバーすることができる。拡大係数の詳細については後述する。

### 3.2.2 JICA-PT データの拡大係数

拡大係数とは、サンプルデータの偏りを除去するための係数であり、既知である属性情報の分布から算出される係数である。理論的には、拡大係数を使うことで限られたサンプルだけを用いて、母集団全体を再現することができる。JICA-PT データには各世帯または、各世帯・各個人両方に設定されており、各サンプル個人の拡大係数を足し合わせることで、対象地域の全人口と相当する人数が得られる。JICA-PT 調査では、対象地域にある人口データをもとに、拡大係数が算出されている。

Ho Chi Minh の JICA-PT データでは、以下のようにして拡大係数が算出されている。



- 1999 Census と statistical year book 2001 から、調査エリアごとの人口・世帯数を算出する。
- 上で求めた 1999 年と 2001 年の人口から 2002 の人口・世帯数を推定する。
- 調査エリアごとの推定された世帯数と調査が行われた有効な世帯数を比較し、世帯の拡大係数を算出する。
- 調査エリア・性別・年齢カテゴリーごとの推定された人口と調査が行われた有効なサンプル数を比較し、個人の拡大係数を算出する。

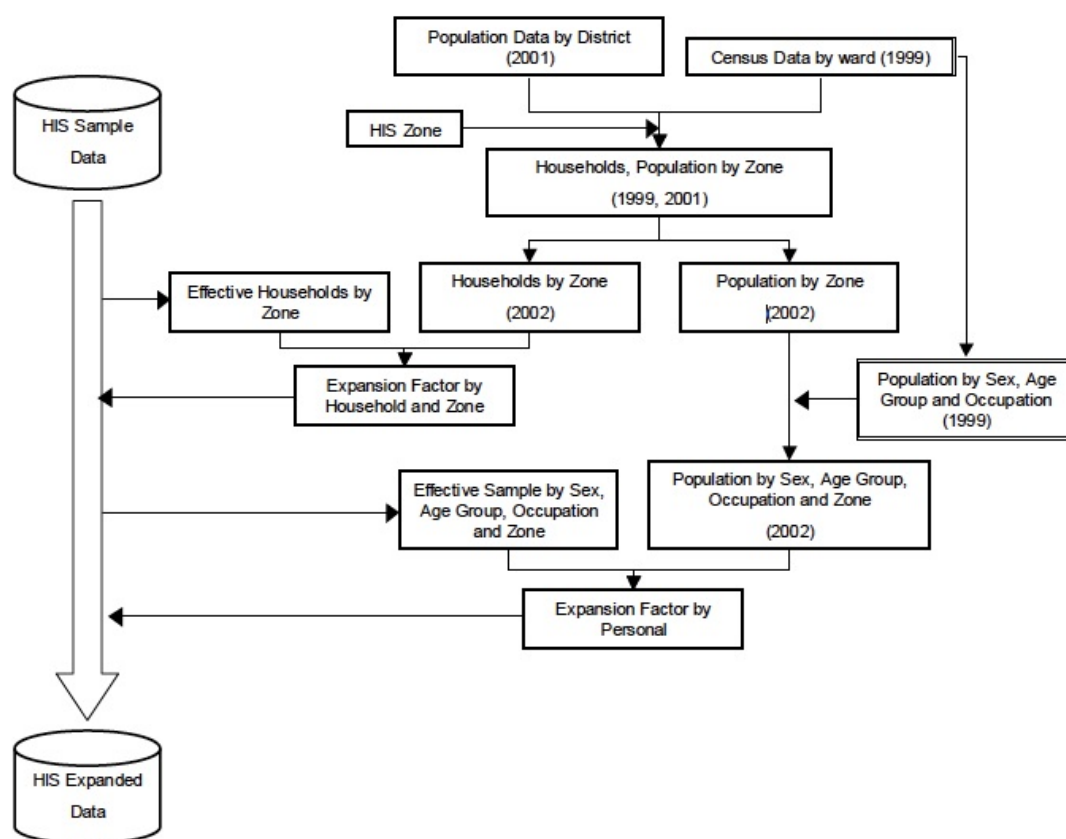


図 3.6: 拡大係数の算出フロー (Ho Chi Minh)

(以上, Ho Chi Minh の JICA-PT 調査の報告書より抜粋)

表 3.1: JICA-PT データの概要 (サンプル数・トリップ数など)

都市	人口 (百万人)	調査年	サンプル人数	トリップデータ行数	交通機関分担率 (二輪車/四輪車/タクシー/バス/電車)
Manila	9.45	1996	231889	471035	2%/10%/25%/58%4%
Kuala Lumpur	1.39	1997	80560	218460	29%/44%2%/23%/2%
Damascus	3.08	1998	38490	81698	4%/25%/15%/56%/0%
Managua	1.2	1998	24854	54138	2%/25%/4%/69%/0%
Bucharest	2.15	1998	67509	143311	0%/19%/0%/27%/54%
Phnom Penh	1.15	2000	18664	40369	89%/11%(四輪車, タクシー)/-/
Chengdu	3.09	2000	31188	70199	81%/10%/4%/4%/0%
Belem	1.78	2000	24043	59529	15%/13%/2%/70%/0%
Jakarta	2.1	2000	423237	1083280	2%/0%/42%/56%/0%
Tripoli	0.33	2001	3608	7650	35%/29%/17%/19%/0%
Cairo	14.4	2001	136070	268360	2%/13%/46%/29%/10%
Ho Chi Minh City	3.18	2002	27412	71890	96%/2.4%(四輪車, タクシー)/1.8%/-
Hanoi	7.16	2004	63716	188949	89%/3.6%(四輪車, タクシー)/6.7%/-
Nairobi	4.04	2004	20980	26828	2%/29%(四輪車, タクシー)/68%/1%
Lima	8.04	2004	115728	270384	17.2%(二輪車, 四輪車)/13.7%/69.1%/-



### 3.3 道路ネットワークデータ

人の流れデータセット構築には、経路探索・時空間内挿をする際に道路ネットワークデータが必要である。道路ネットワークデータとは、道路などのリンクや交差点などのノード、及びそれらの関係性が含まれたデータを指す。多くの道路ネットワークデータでは、リンク・ノードの識別番号に加え、これらの位相構造が整備されている。リンク・ノードの関係性がどのようにつながっているかの構造を位相構造と呼び、経路探索を行う上でこの位相構造は必要不可欠である。また、道路ネットワークデータには、リンクの長さ等の定量的な情報に加えて様々な属性情報が含まれていることが多い。例えば、規制速度・道路の幅・交通規制（一方通行）・道路の種類などであり、より正確な経路探索に寄与している。道路ネットワークデータの作成手法はいくつか知られている。古くからある手法としては、紙地図をデジタル化し属性情報や位相構造を付加していく手法がある。また、近年では GPS のログデータやプローブ車などの測定機器によって取得されたデータを使う手法もある。世の中には様々な道路ネットワークデータが存在し、種類や作成の方法が多様化してきている。

#### 3.3.1 デジタル道路地図 (DRM)

日本国内で最も高精度な道路ネットワークデータとされているものは、デジタル道路地図（以下、DRM）である。これは、財団法人日本デジタル地図協会が、国土地理院の 1/25000 の地形図を基に作成され、全国の地方整備局、都道府県、政令指定都市、高速道路会社、道路関係公社など全国の道路管理者の資料と新刊地形図により毎年データの更新が行われている。[26] DRM の特徴としては、以下の 4 点が特徴としてあげられる。

- ネットワーク表現に適したデータ構造であり、コンピュータを用いた経路探索が可能。
- 幹線道路は供用開始などに先がけたデータの更新を行うなど、データの更新が頻繁。
- 位置的データの他に道路管理者、路線番号、幅員、橋・トンネルなどの道路構造物などの多くのデータを保有。
- データベース標準の制定・公開を行い、官民における共通基盤となるデータベースであること。

DRM は、VICS（道路交通情報システム）やカーナビゲーションシステムにも利用されている。また、ITS の今後の発展にも貢献していくであろう。CSIS による日本国内の人の流れデータセットにもこの DRM が利用されている。

### 3.3.2 商用地図データセット

民間企業が提供し、ライセンス料を得ている商用の地図データセットも存在する。例えば、世界的に有名なのは NAVTEQ[11]・TomTom[19] であり、日本国内ではゼンリン [20] がある。これらのデータを利用する際には、ライセンス料が発生し原則として無断利用・二次利用をすることが禁じられている。

### 3.3.3 GoogleMaps

Google の GoogleMaps[5] も道路ネットワークデータとして整備されており、経路探索などを行うことができる。ただし、詳細なデータ構造の公開や提供はされておらず、Web API による経路探索のサービスの提供のみ行っている。これらのサービスは一定時間内の検索回数の制限はあるものの、全て無料で提供されている。ただし、使用されている道路データは先述した NAVTEQ・TomTom、日本国内では、ゼンリンからライセンスを受けたデータであり、GoogleMaps のデータ自体を完全に自由に使うことはできない。

### 3.3.4 OpenStreetMap

OSM は、道路地図などの地理情報データを誰でも利用できるよう、フリーの地理情報データを作成することを目的としたプロジェクトである。誰でも自由に参加し、誰でも自由に編集でき、誰でも自由に利用することができる。近年では、世界中で整備が進み、Google Map に代表される他の電子地図にも遜色のない程度まで整備されている。OSM の特徴としては次の点があげられる。

- フリーのプロジェクトであり、誰でも自由にデータを使用することができる。
- 地理情報の整備をユーザーが行う。
- ネットワーク構造を持つデータをフリーで入手することができる。

地図の作成は、GPS トラックログを利用する方法と衛星・航空写真を利用する方法がある。地図の作成、編集する際の標準と慣習を以下に記す。(OpenStreetMap Wiki[13] より抜粋)

#### 1. 対象とする地物

OSM では主にポイント、ウェイ、エリア、リレーションなどの要素を対象としている。ノードは OSM スキームの基本的な要素であり、緯度・経度から構成される、一つの地点を表したものである。ウェイは 2 点以上かつ 2000 点以下のノード間を順序

付きでつながる要素であり、道路など線状の地物を表したものである。エリアは、閉じたウェイで構成される特定の閉じた空間を表したものである。リレーションは、他の要素・ノード、ウェイ、他のリレーションをグループ化する要素である。自転車ルートや方向転換制限などを表す際に使用されている。また、これらの地物を説明するタグをつけることができ、検証可能な限りどのようなタグでもつけることができる。

## 2. 道路

道路はウェイでまとめられたノード群によって表現され、`highway` タグと名前でタグ付けがされている。OSM エディターでは、入力されたタグの値に基づき、異なる色や幅で表示される。また、対象とする道路の環境に応じて編集の仕方がある。例えば、一方通行の道路、中央分離帯などによって分割された道路、交差点、橋などを識別できるように編集を行う。特に交差点を表現する際には、道路と接続した単一のノードであるべき、ということを周知している。

## 3. 正確性

OSM では道路を点と直線で表現するため、カーブする道路を表現する時の正確性が問題になることがある。カーブする道路を表現する場合は、おおまかな形ではなく、ノードを多く加えることで実際の道路の形状に近づけるよう呼びかけている。

## 4. トポロジー

OSM では、ノード・ウェイを接続することで自動的にトポロジー構造が構築される。正確なトポロジー構造を構築するために、交差点であれば道路の交差する点にノードを加えたり、立体交差のような交差している場所に橋がある場合は、橋の地物を生成したりする必要がある。

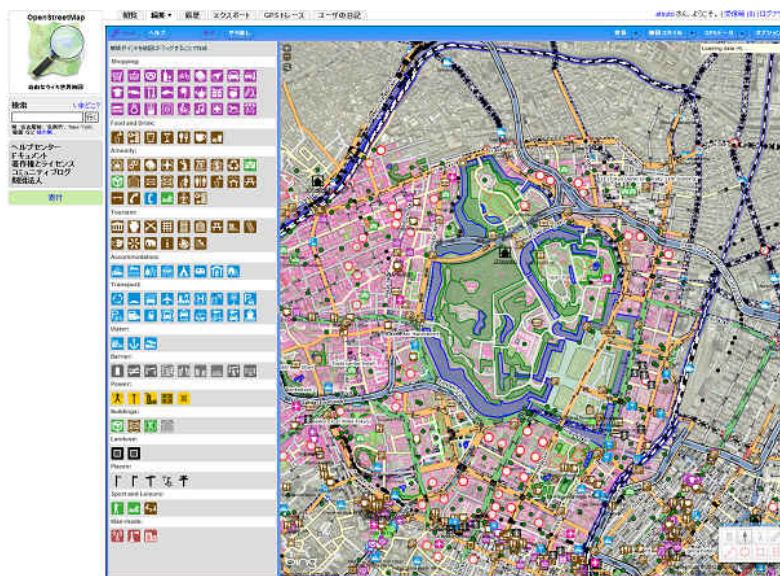


図 3.7: Potlatch を使った OSM 編集画面

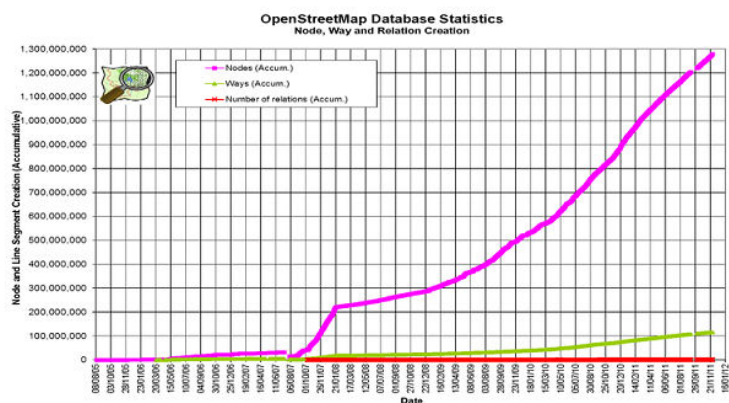


図 3.8: OSM のノード数・リンク数の変遷

表 3.2: 道路ネットワークデータの特徴

データ名	カバーする範囲	長所	短所	精度	無償
DRM	日本国内	更新が早い.	日本国内のみしかない	良	研究目的に限る
GoogleMaps	全世界	様々な API を提供	元データが入手できない	良	
OpenStreetMap	全世界	ネットワークデータを入手可能	精度がマッパーに依存する	(*)	✓

(\*) ボランティアによって整備されている OSM の精度には問題があり，詳細は後述する．

### 3.3.5 まとめ

OSM ではネットワークデータ自体を自由に入手・使用することができ，XML 形式で記述されているため，編集なども簡単に行うことができる．また，データベース管理システム PostgreSQL[6] も OSM のデータ構造に対応しており，OSM を利用した経路探索も容易に行うことができる．一方，GoogleMaps ではサービスのみ提供がされており道路ネットワークデータ自体を取得することができないうえに，ライセンスの問題もある．本研究では，グローバルに利用可能なフリーなデータを利活用できる手法を構築することが目的の一つである．そこで，本研究では OSM の道路ネットワークデータを使用する．ただし，OSM の道路ネットワークデータの品質に問題がある都市もある．詳細については後で述べる．

## 3.4 土地被覆・人口に関するデータ

### 3.4.1 LandScan

LandScan は，Oak Ridge National Laboratory[12] の Land Scan Global Population Project[9] によって整備されている人工グリッドデータであり，世界中にある非集計人口データからデータ構築が行われている．人口のグリッドへの集計は道路ネットワーク，傾斜，土地被覆，夜間光などの様々な情報をもとに構築された人口モデルが用いられている．[4] 全球で整備されている人口データセットの分解能は，30''（約 1km）であり，人口はすべて 24 時間平均の人口である．LandScan USA は，アメリカのみで整備されている人口データセットで，より高分解能で昼夜間人口も整備されている．[1]

このデータを使用する前に，LandScan データ自体の品質を調べる必要がある．上で述べたように，LandScan データは入手可能な人口の非集計データ（主にセンサスなど）を独自の人口モデルを用いて，メッシュ集計することで構築されている．LandScan データの品質を調べるために，正確な人口データが充実している東京（離島など除く）の国勢調査と LandScan の人口の比較を行う．以下に手順を記す．

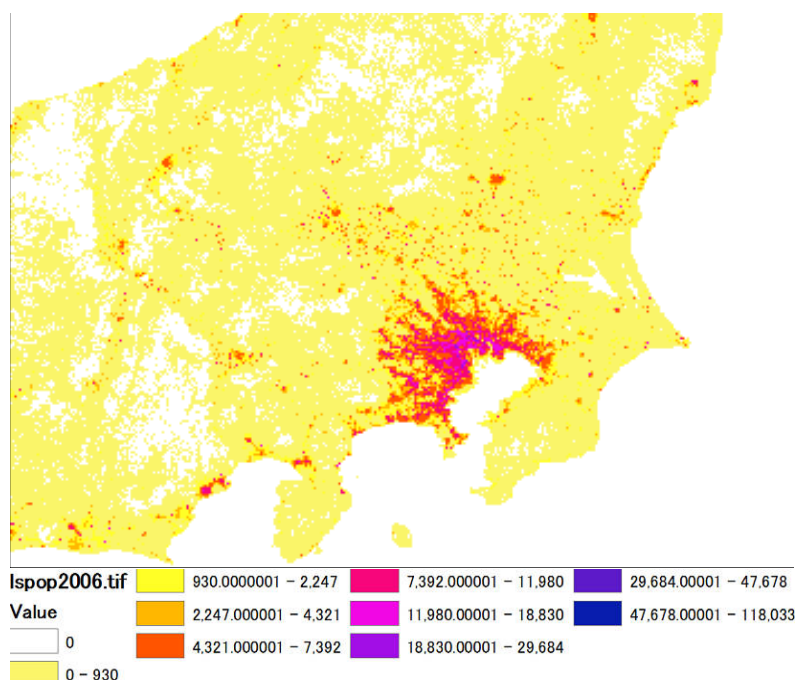


図 3.9: LandScan データ（東京）

- 平成 17 年度の国勢調査を 3 次メッシュで集計する。
- LandScan の人口データを 3 次メッシュで面積按分し、3 次メッシュ単位での LandScan 人口を取得する。
- 上の 3 次メッシュ集計された国勢調査・LandScan の人口を比較する。

その結果、対象メッシュでの人口の総計は、国勢調査では 12,547,841 人のところ、LandScan では 11,620,421 人となった。また、相関係数を見ても約 0.716 と相関も強くないことがわかる。国勢調査の人口データを最も精度の高い人口データとみなすと、LandScan の人口データは必ずしも精度が高いとはいえないことがわかる。

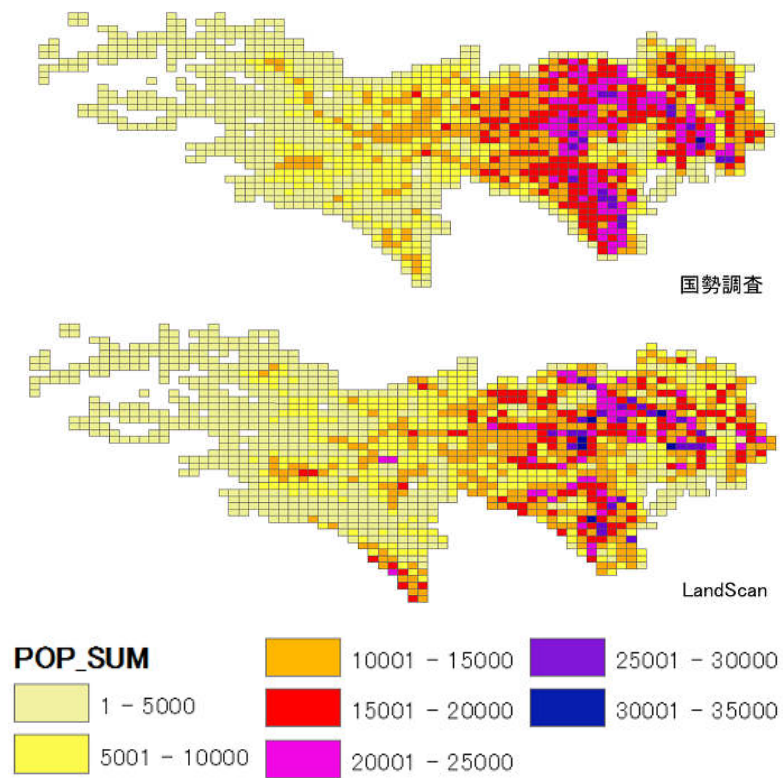


図 3.10: 国勢調査と LandScan 人口分布の比較

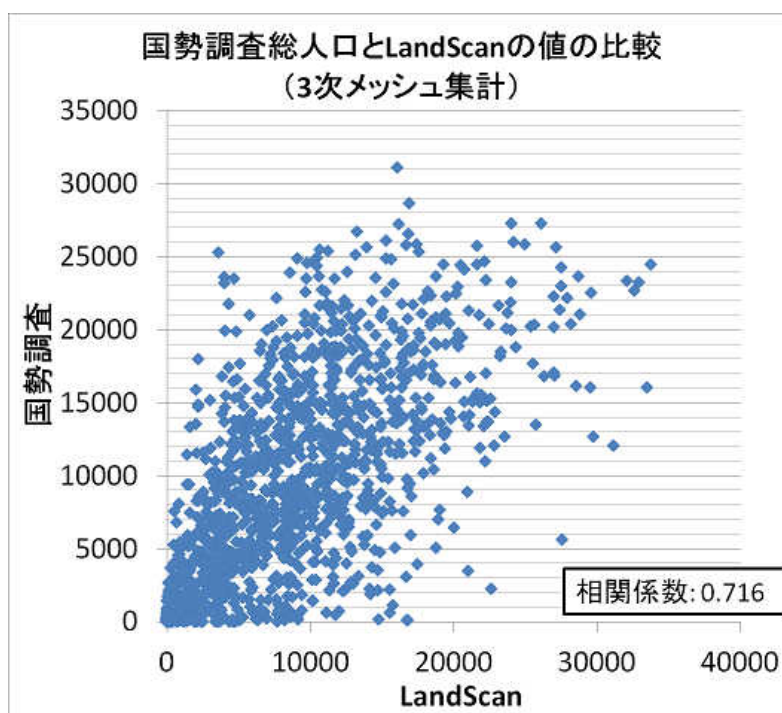


図 3.11: 国勢調査と LandScan 人口の比較



### 3.4.2 MCD12Q1

MODIS[10] によって取得されたデータをもとに整備され、Land Processes Distributed Active Archive Center[3] が提供している土地被覆データであり、IGBP の 17 クラスによって分類されている。[15]MODIS は、アメリカ航空宇宙局（NASA）によって開発された可視・赤外域の放射計であり、地球観測システム Terra に搭載されている。

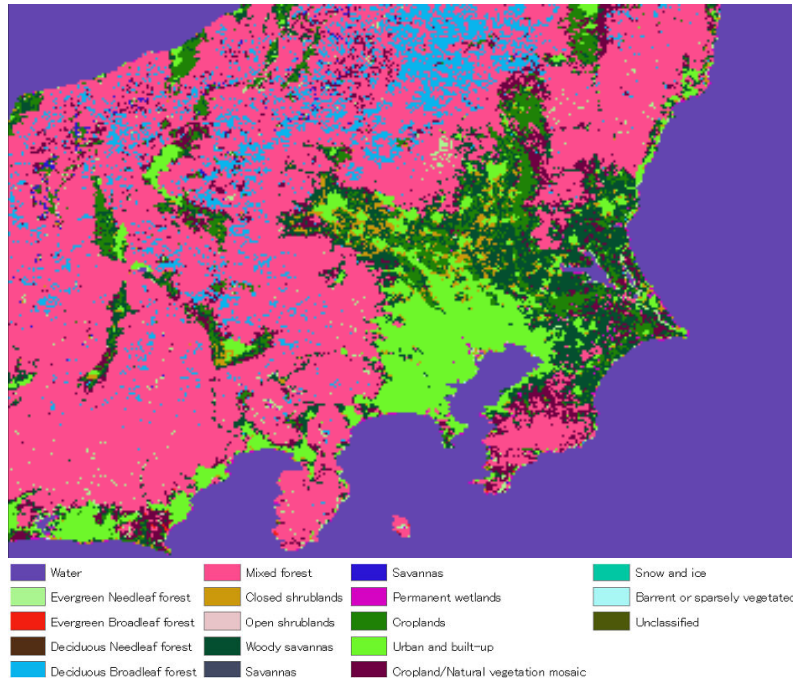


図 3.12: MCD12Q1（東京）



## 第4章 人の流れデータセット構築手法

### 4.1 人の流れデータセット構築手法のフレームワーク

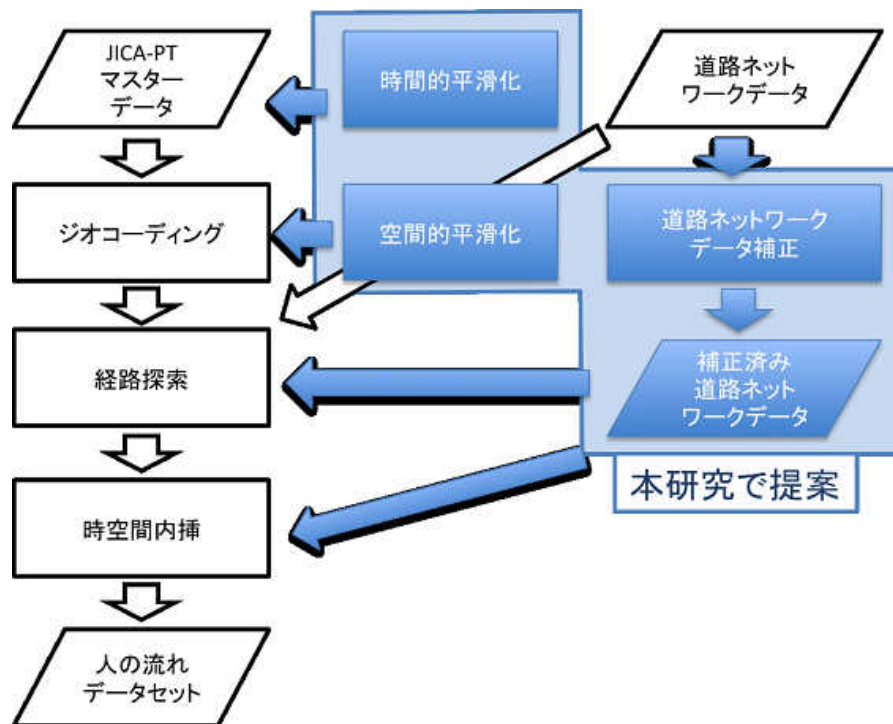


図 4.1: 本研究のアプローチ

図 4.1 にあるように、JICA-PT データにジオコーディング・経路探索・時空間内挿を施すことによって、人の流れデータセットを構築する。時間的な偏りを解消するための時間的平滑化、空間的な偏りを解消するための空間的平滑化、道路ネットワークデータの品質を解消するための道路ネットワークデータ補正をさらに行った。この 3 つの手法を既存の人の流れデータセットに統合し、JICA-PT データを用いた人の流れデータセットの構築を行った。

## 4.2 時間的平滑化手法

本研究で使用している JICA-PT データの時刻表記には偏りがある。図 4.2 からわかるように、出発時刻・到着時刻が〇〇時 00 分や〇〇時 30 分などのきりのよい時刻に集中している。これは、JICA-PT データの調査方法がアンケートであることに起因していると考えられる。このような時刻の偏りは、現実の状況と近いとは考えにくく、人の流れデータセットの再現性にも影響を与える。そこで、本研究では、一部の時刻に集中している出発時刻・到着時刻を平滑化した。

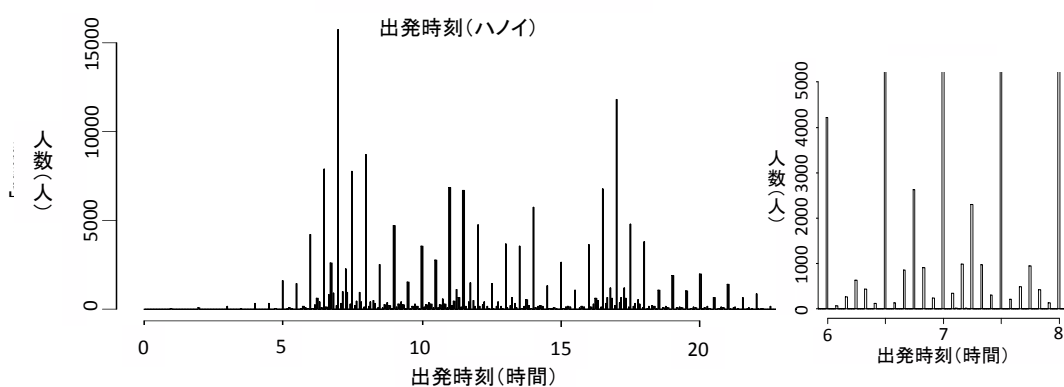


図 4.2: 出発時刻のヒストグラム (ハノイ)

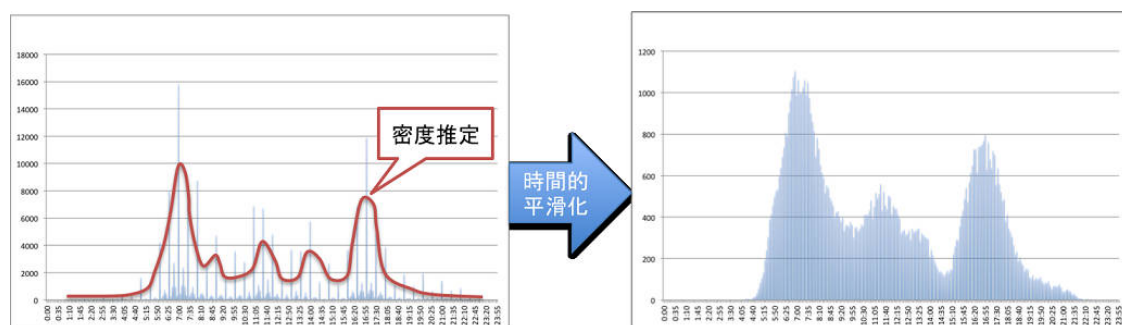


図 4.3: 時間的平滑化のイメージ

### 4.2.1 カーネル密度推定

カーネル密度推定を用いて、時間的平滑化を行った。カーネル密度推定は本来、確率密度関数を推定する手法であり、時空間データの平滑化手法としても知られている。

$x_i$  を観測値,  $n$  を観測値の個数,  $h$  をバンド幅,  $K(\cdot)$  をカーネル関数とすると, 以下のよう  
に算出される。

$$\hat{f}(x) = \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (4.1)$$

カーネル密度推定では, 4.1 式の, カーネル関数  $K(\cdot)$  とバンド幅  $h$  を設定する必要がある。カーネル関数には, Epanechnikov, Gaussian, Rectangular などがあることが知られている。本研究では, 平滑化の結果が滑らかになり, 計算過程がシンプルで計算時間のかからない Gaussian をカーネル関数に採用した。

カーネル密度推定を用いた平滑化では, 選定するバンド幅により結果が大きく異なる。本研究で提案する手法は, 様々な都市に適用していくため, 既存に存在する手法から次の二点を満たす手法を採用する。

- 本研究で使用するデータの性質に適した手法,
- 各都市に応じて, 個別にパラメータを設定する必要がない手法,

### 4.2.2 バンド幅選定手法

任意でパラメータを設定する必要のないバンド幅選定手法について述べる。

#### i: Reference to standard deviation selector

この手法は, 平均二乗誤差を最小にするというコンセプトのもとにバンド幅が算出される。算出されるバンド幅を  $h_{opt}$  とすると, 以下のよう  
に算出される。

$$h_{opt} = k_2 \left\{ \int K(t)^2 dt \right\}^{1/5} \left\{ \int f''(x)^2 dx \right\}^{-1/5} n^{-1/5} \quad (4.2)$$

真の分布が標準分布である場合, 観測値の標準偏差を  $\sigma$  とすると, 最適なバンド幅  $h_{opt}$  は, 以下のよう  
に算出される。

$$h_{opt} = 1.06\sigma n^{-1/5} \quad (4.3)$$

#### ii: Data-based selector

4.3 式は, 真の分布が標準分布であるという仮定に基づいているので, 観測値の標準

偏差を  $\sigma$ , 四分位範囲を  $R$  とすると, 最適なバンド幅  $h_{opt}$  は, 以下のように算出することもできる.

$$h_{opt} = 0.79Rn^{-1/5} \quad (4.4)$$

式 4.3 の  $\sigma$  のかわりに adaptive estimate を用い, 式 4.4 の 1.06 を減退させることにより, 式 4.3 と式 4.4 を統合して,  $h_{opt}$  は以下のように算出される.

$$A = \min\{\sigma, R/1.34\} \quad (4.5)$$

さらに, Gaussian 関数が採用される場合は,  $h_{opt}$  は以下のように算出される.

$$h_{opt} = 0.9An^{-1/5} \quad (4.6)$$

### iii : Least-squares cross-validation selector

Rudemo ら [14] や Bowman ら [2] によって提案されている, 自動的にバンド幅を選定する手法では, 以下のように真の分布と推定される分布の積分二乗誤差が最小になるように, バンド幅が算出されている.

$$\int (\hat{f} - f)^2 = \int \hat{f}^2 - 2 \int \hat{f}f + \int f^2 \quad (4.7)$$

式 4.7 の第二項中の  $f$  は, クロスバリデーション法を使い, サンプルデータ自体から推定される. この手法では, 以下のコスト関数式 4.8 を最小にする時の  $h$  が最適なバンド幅となる.

$$M_1(h) = n^{-2}h^{-1} \sum_i \sum_j K^* \{h^{-1}(X_i - X_j)\} + 2n^{-1}h^{-1}K(0) \quad (4.8)$$

なお, 上の式 4.8 中の  $K^*(t)$  は, 以下のように定義される.

$$K^*(t) = K^{(2)}(t) - 2K(t) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{4}t^2\right) - 2 \times \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{2}t^2\right) \quad (4.9)$$

### iv : Least-squares Poisson selector

この手法は, H. Shimazaki ら [17] によって, 平均二乗誤差を最小にするというコンセプトのもとにスパイクレートを推定する手法として考案された.

$$C_n(h) = \frac{1}{n^2} \sum_{i,j} \psi_h(t_i, t_j) - \frac{2}{n^2} \sum_{i \neq j} k_h(t_i - t_j) \quad (4.10)$$

$$\psi_h(t_i, t_j) = \int_b^a k_h(t - t_i)k_h(t - t_j)dt \quad (4.11)$$

これらの i~iv のバンド幅選定手法は単峰形分布，双峰形分布どちらかの真の分布として推定している．バンド幅選定手法 i, ii が単峰形分布を，手法 iii, iv は双峰形分布を想定している．また，双峰形分布を想定するバンド幅選定手法は，観測値自体または，離散的分布を推定し，それを真の分布と想定しバンド幅を選定する．

これらの手法以外にもバンド幅選定手法が存在する．例えば，真値と推定値の最大誤差を最小にするテストグラフ法などが知られている．[18] しかし，これらの手法を用いて，パラメータの設定なしで，バンド幅を選定する手法は提案されていない．

### 4.3 空間的平滑化手法

本研究で使用している JICA-PT データの出発・到着地点に関する位置情報は，調査で定められた調査区の識別番号（以下，エリア ID）のみである．このエリア ID を諸処理に適した緯度・経度に変換するジオコーディングを行う必要がある．例えば，各エリアの代表点を定め，エリア ID を代表点の緯度・経度に変換する方法が考えられる．しかし，この方法でジオコーディングすると，サンプルの出発・到着地点に偏りが生じ，人の流れデータセットに空間的な偏りが生じてしまう．そこで本研究では，このような空間的な偏りが生じないようなジオコーディングの方法を提案する．

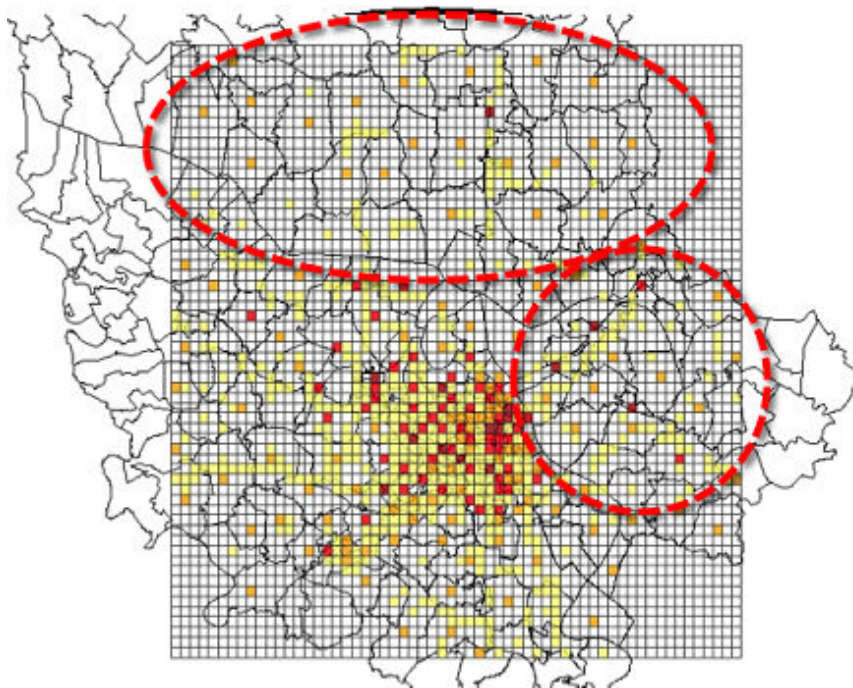


図 4.4: 午前 9:00 における人口分布（マニラ，既存手法）

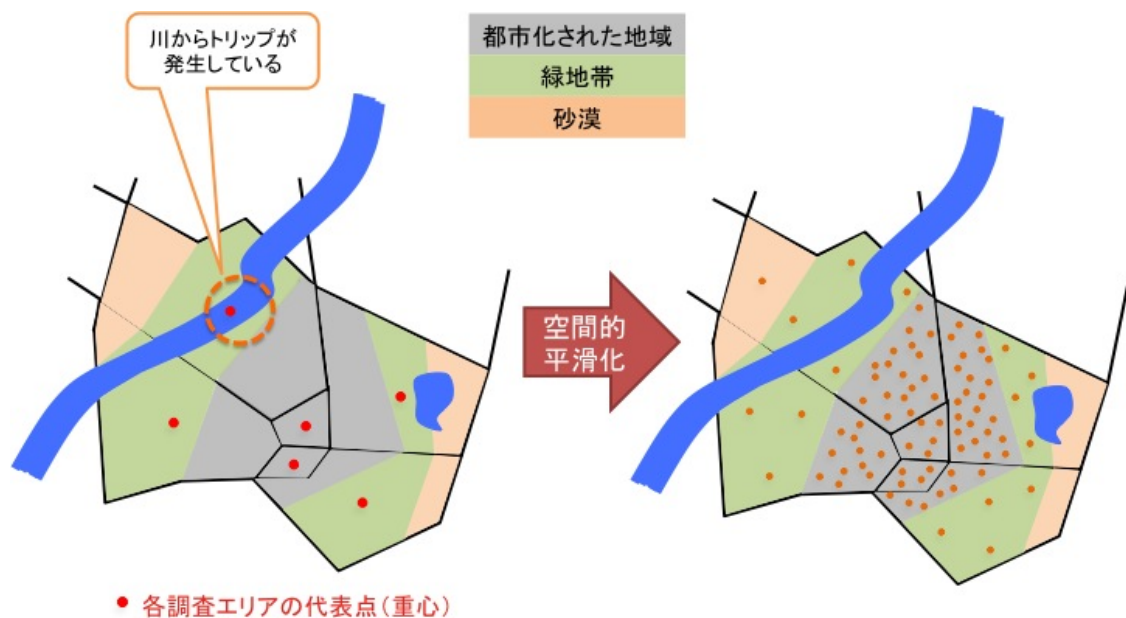


図 4.5: 空間的平滑化のイメージ

#### 4.3.1 土地被覆データと人口データを用いた重み算出

各エリア ID を各エリアの任意の地点の座標に変換してしまうと、実態に伴わない出発・到着地点が存在してしまう可能性がある。例えば、河川などが出発・到着地点となってしまうことがある。また、ほとんどのトリップは自宅や職場など建物があるところが出発・到着地点となっており、森林・牧草地よりも建物などの人工物のある場所が出発・到着地点となるべきである。これらの点を考慮してジオコーディングするためには、出発・到着地点として可能性の高い場所に重みをつけたジオコーディングを行う必要がある。

本研究では、この重み付けに土地被覆の情報を利用した。具体的には、土地被覆分類（17 分類）を全球で利用できる MCD12Q1 を利用した。以下の手順を各都市の JICA-PT エリア内で行い、重みを算出した。

##### 1. 土地被覆ポイントの生成

MCD12Q1 の各グリッドの中心にポイントを作成する。このポイントは土地被覆ポイントと呼ぶことにし、ポイント ID、緯度、経度、土地被覆分類の値、ポイントが含まれるエリア ID を属性情報として持つ。

##### 2. 各土地被覆ポイントにおける人口の取得

土地被覆ポイントと LandScan のラスターデータをオーバーレイし、各土地被覆ポイントにおける LandScan の人口の値を取得する。この人口の値を土地被覆ポイントの属性情報に追加する。

表 4.1: 土地被覆ポイントの持つ属性情報

属性	Type	内容
ポイント ID	Integer	土地被覆ポイント ID (一意に設定)
緯度	Double	土地被覆ポイントの緯度
経度	Double	土地被覆ポイントの経度
土地被覆分類の値	Integer	土地被覆分類 (IGBP)
人口 (LandScan)	Integer	LandScan から取得された人口の値

### 3. 重みの算出

土地被覆分類の値を  $l(0 \sim 16)$ ,  $LS_{il}$  を土地被覆ポイント  $i(i = 0 \sim N)$  において、土地被覆分類の値が  $l$  であるときの LandScan から得られる人口とすると、調査エリア内における土地被覆分類の値が  $l$  の地点の人口の総計は、

$$P(l) = \sum_{i=0}^N LS_{il} \quad (4.12)$$

によって与えられ、土地被覆分類の値が  $l$  のにおける重みは、以下のように算出される。

$$W(l) = \frac{Pop(l)}{\sum_{l=0}^{16} LS_{il}} (0 \leq Pop(l) \leq 1) \quad (4.13)$$

#### 4.3.2 重み付きジオコーディング

上で求めた重み  $W(l)$  を用いて、JICA-PT トリップデータの各サブトリップに対して、重み付きジオコーディングを行った。重み付きジオコーディングは以下の手順で行った。

##### 1. 出発・到着エリアに含まれる土地被覆ポイントの取得

トリップデータに含まれる出発・到着エリア ID を取得する。そして取得した出発・到着エリアに含まれる土地被覆ポイントを全て取得する。

## 2. ジオコーディング

各土地被覆分類における重み  $W(l)$  を用いて、ジオコーディング先の土地被覆ポイントを選定する。選定された土地被覆ポイントの緯度・経度をトリップデータに加える。

このようにして、「エリア ID  $\Rightarrow$  緯度・経度」のジオコーディングを行った。

## 4.4 道路ネットワーク補正手法

フリーのプロジェクトである OSM では、道路ネットワークデータを無料で入手することができる。しかし、不特定多数のボランティアによって整備されているため、道路ネットワークデータとしての問題点がある。それは、道路ネットワークデータのトポロジ構造が必ずしも保持されておらず、接続されているはずのリンクが接続されていないことがあることだ。経路探索に必要な道路ネットワークデータは、トポロジ構造が保持されている必要があり、接続されているはずのリンクが接続されていないと、経路探索ができないことがあり、経路探索を行う上で、トポロジ構造を保持した道路ネットワークデータは必須となる。

OSM の道路ネットワークデータがどの程度接続されているかを確認するために以下の手順で OSM のグループ分けを行った。OSM データの任意のノードを取り、それにつながるリンクを取っていき、接続しているリンクを全てつなげたものを一つのリンクのグループとする。

この OSM のグループ分けの結果を表 4.2 と、図 5.4 に示す。図 5.4 の色分けは、リンク数の多いグループ上位 5 グループを赤、青、緑、水色、ピンクで、それ以外を黒で色分けしている。リンクのグループが多く存在し、各グループのリンク数も少なくいうえ、接続されるべきリンクが接続されていないことも確認できる。このように、トポロジ構造が保持されていないことで、経路探索の結果の精度が下がることが予測される。

表 4.2: OSM のリンク数・ノード数と接続状況

都市	リンク数	ノード数	グループ数	最大ループ内リンク数
Manila	55282	331557	38881	2131 (5.5%)
Cairo	14057	23940	10603	66 (0.5%)
Jakarta	16928	107399	11719	157 (1.3%)
Ho Chi Minh	18310	27151	9484	241 (1.3%)
Hanoi	3552	24185	2210	91 (4.1%)



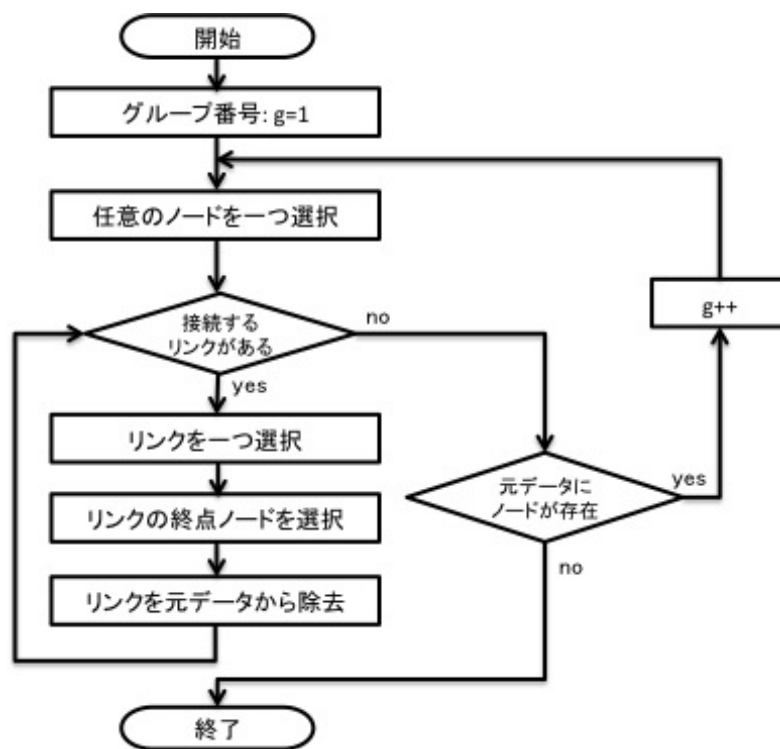


図 4.6: OSM グループ分けの手順

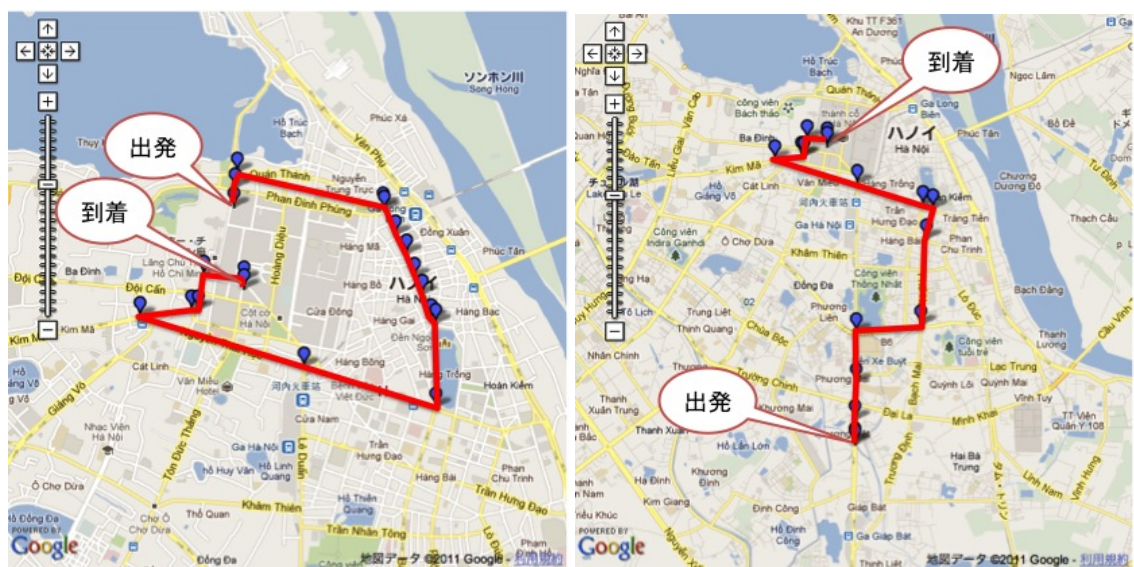


図 4.7: OSM を用いた経路探索の結果例

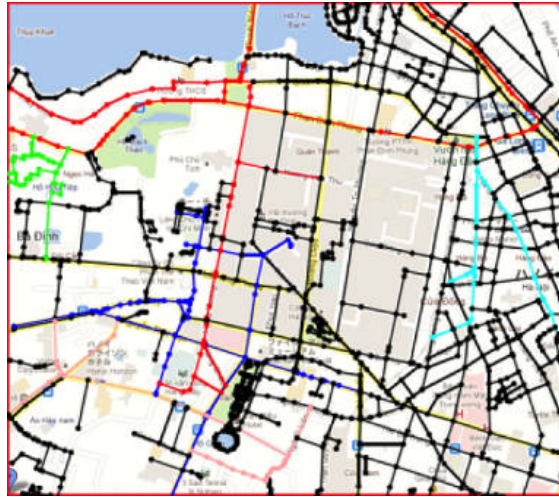


図 4.8: OSM グループ分けの結果

#### 4.4.1 道路ネットワークデータの品質評価

ここまでで、経路探索を行う際の OSM データの問題点をあげた。OSM データの品質が経路探索にどのように影響を与えるかを明らかにするために、OSM データの品質評価を行った。

JICA-PT トリップデータに含まれる全ての有効なサブトリップから、全ての経路の組み合わせを抽出し、その全経路に対して経路探索を行いその結果、経路が得られた経路を成功経路とした。また、その経路が最短経路であるかを検証するために以下の式 4.14 で算出される迂回率という指標を用いた。

$$\text{迂回率} = \frac{\text{経路長}}{\text{出発・到着ノード間の直線距離}} \quad (4.14)$$

本研究では、最短経路を経路探索の結果としているため、式 4.14 で算出される迂回率が 2 より小さくなっているべきであると推測する。OSM データを用いてこの評価を行った結果を以下の表 4.3 に示す。この結果から、ほとんどの都市で経路探索の成功率がかなり低く、経路探索経路探索を行う道路ネットワークデータとして OSM データは十分でないことがわかる。

#### 4.4.2 ノードマッチング処理

経路探索を行うために OSM の補正処理を行った。上で述べたような品質の低さは、接続されているべきリンクが接続されていないことが主な原因であると考えられる。そこで、この問題を解消するために、ノードマッチング処理を行った。

表 4.3: OSM データの品質評価の結果

都市	試行トリップ数	成功トリップ数	経路探索成功率 (%)	迂回率が2未満の経路数
Manila	376341	4082	1.1	815 (57.1%)
Cairo	215815	3640	1.7	139 (62.9%)
Jakarta	922802	40621	4.4	853 (56.6%)
Ho Chi Minh	250775	2644	1.1	129 (76.8%)
Hanoi	161243	5011	3.1	135 (55.1%)

ノードマッチング処理では，図 4.9 に示す場合にノード・リンクを追加するという処理である．(a) のようにリンクが交差している場所にノードが存在しない場合は，ノードを追加する．また，(b) のようにリンクが存在し，隣接するリンクの終点とノードの距離があるいある閾値より短い場合リンクを延伸し，その交差点にノードを追加する．この閾値は，短ければ (b) の場合で，接続するべきリンクを接続できないことが多くなり，また，閾値が長ければ接続するべきでないリンクを接続することが多くなる．

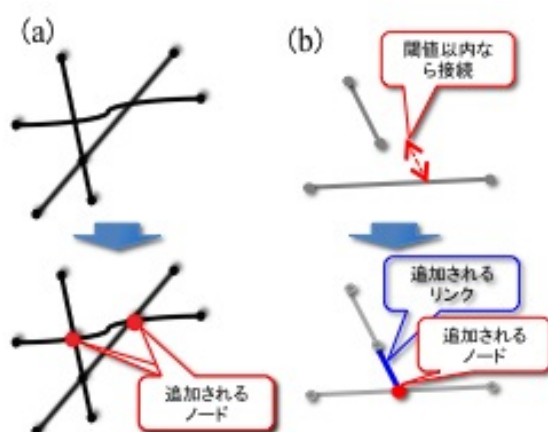


図 4.9: ノードマッチング処理の例

## 4.5 時空間内挿

本研究で使用している JICA-PT などのパーソントリップデータに含まれる位置情報は，出発・到着地点についてのみであり，出発・到着地点間の情報は全く含まれていない．そこで，人の一日の動きを再現する人の流れデータセットを構築するためには，時空間内挿を行い出発・到着地点間の情報を補間する必要がある．時空間内挿の処理は，以下の手順にそって行う．

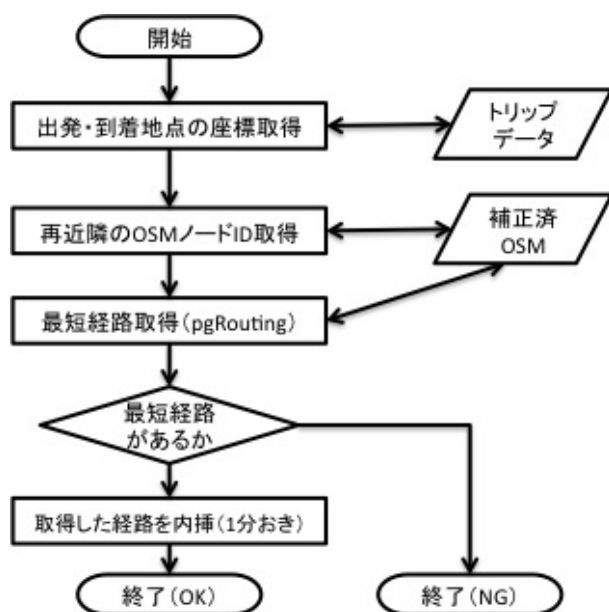


図 4.10: 経路探索の手順

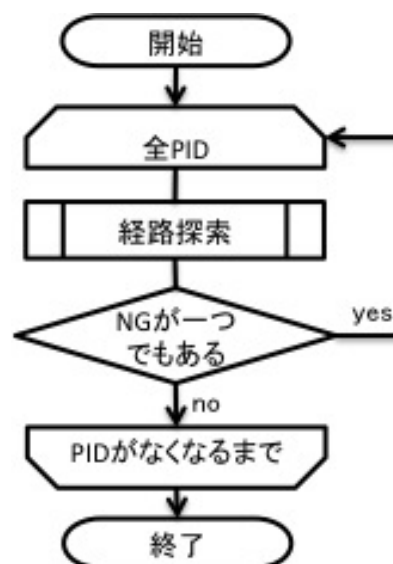


図 4.11: 時空間内挿の手順

JICA-PT トリップデータからトリップを含むサンプルを取り出し、全員分の時空間内挿を行う。その際、あるサンプルの持つトリップのうち一つでも無効なトリップや経路探索ができないトリップがある場合は、そのサンプル自体を除外する。図 4.11 内にある経路探索の処理は、図 4.10 の手順で行われる。

このように時空間内挿を行うことで、有効または経路探索のできるトリップを含む、全サンプルについて、1 分区切りで 24 時間つまり、0 分～1440 分における以下の情報を得ることができる。

## 4.6 データセットの構築

時空間内挿によって得られた表 4.4 の情報を 1 分区切り 24 時間、全サンプル分を含むものが人の流れデータセットとなる。つまり、このデータセットには、全サンプルの位置情報・使用した交通機関などの属性情報を 24 時間 1 分区切りで知ることができるデータセットである。

表 4.4: 時空間内挿の結果得られる情報

項目名	Type	内容
PID	Integer	個人ID（サンプルの識別番号）
TNO	Integer	トリップ番号
SNO	Integer	サブトリップ番号
LON	Double	緯度
LAT	Double	経度
SEX	Integer	性別コード
AGE	Integer	年齢コード（都市ごとに設定されている）
PADD	Integer	居住している調査区コード
WORK	Integer	職業コード（都市ごとに設定されている）
PURPOSE	Integer	トリップの目的コード都市ごとに設定されている）
MAGFAC	Double	拡大係数
MAGFAC2	Double	拡大係数（2つある場合の2つ目）
DATUM	Integer	使用した交通機関コード
START_TIME	Integer	時刻（分表記）
END_TIME	Integer	START_TIME に 1 分足した時刻

表 4.5: 人の流れデータセットサンプル

PID	TNO	SNO	LOX	LAT	SEX	AGE	PADO	WORK	PURPOSE	MAGFAC	MAGFAC2	DATUM	START TIME	END TIME
*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	348	349
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	349	350
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	350	351
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	351	352
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	352	353
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	353	354
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	354	355
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	355	356
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	356	357
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	357	358
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	358	359
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	359	360
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	360	361
1126503	1	1	105.8540	21.2740	2	7	148	12	6	47	47	2	361	362
1126503	1	1	105.8528	21.2694	2	7	148	12	6	47	47	2	362	363
1126503	1	1	105.8517	21.2647	2	7	148	12	6	47	47	2	363	364
1126503	1	1	105.8505	21.2601	2	7	148	12	6	47	47	2	364	365
1126503	1	1	105.8492	21.2572	2	7	148	12	6	47	47	2	365	366
1126503	1	1	105.8480	21.2545	2	7	148	12	6	47	47	2	366	367
1126503	1	1	105.8476	21.2656	2	7	148	12	6	47	47	2	367	368
1126503	1	1	105.8472	21.2717	2	7	148	12	6	47	47	2	368	369
1126503	1	1	105.8468	21.2778	2	7	148	12	6	47	47	2	369	370
1126503	1	1	105.8464	21.2839	2	7	148	12	6	47	47	2	370	371
1126503	1	1	105.8460	21.2900	2	7	148	12	6	47	47	2	371	372
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	372	373
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	373	374
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	374	375
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	375	376
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	376	377
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	377	378
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	378	379
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	379	380
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	380	381
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	381	382
1126503	2	1	105.8500	21.2900	2	7	148	12	1	47	47	2	382	383
*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****

停滞

トリップ

停滞



## 第5章 適用実験

### 5.1 時間的平滑化手法の適用

第4章で説明したバンド幅選定手法 i~iv で選定されたバンド幅・ガウス関数を使ったカーネル密度推定を用いて、出発時刻の平滑化を行った。図 5.1, 図 5.2 では、手法 i を DateBased, 手法 ii を RtoSD, 手法 iii を LSCV, 手法 iv を LSP, 観測値を raw と表現する。

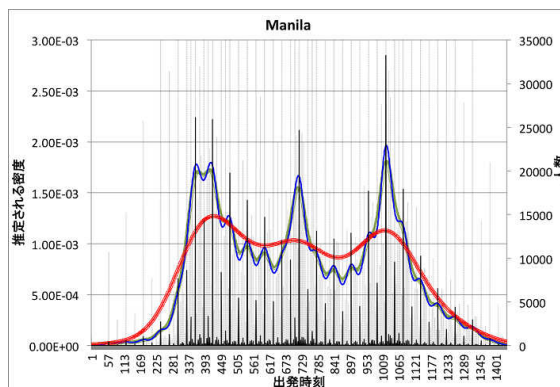


図 5.1: 出発時刻の平滑化の結果 (Manila)

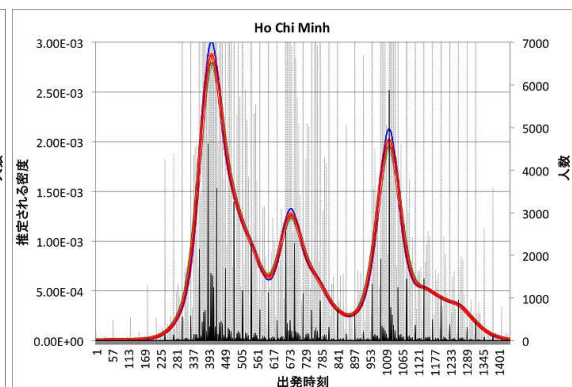


図 5.2: 出発時刻の平滑化の結果 (Ho Chi Minh)

手法 i, ii, iv で選定されたバンド幅を用いた場合うまく平滑化できているが、手法 iii で選定されたバンド幅を用いた場合、グラフの形がぎざぎざしていることからわかるように、うまく平滑化されていない。これは、各手法の原則の違いが起因していると推測される。手法 iii は、元データである観測値が真の分布であるとの原則に基づきバンド幅が選定されているため、平滑化の結果も観測値と同じ傾向を持つことになる。本研究では、JICA-PT データに含まれる出発時刻は 1 分区分切りで記録されているため、平滑化の結果もぎざぎざした不連続な分布となったと考えられる。一方、手法 i, 手法 ii では、何らかの連続的な分布を真の分布と推測しているため、平滑化の結果も連続的な滑らかな分布になっていると考えられる。

表 5.1 の処理時間を見ると、手法 i, ii と手法 iii, iv の処理時間に大きな差が見られる。これは、バンド幅算出のアルゴリズムの違いに依存している。手法 i では観測値の標準偏

表 5.1: 各バンド幅選定手法で算出されたバンド幅と処理時間

都市	サンプル数		DataBased (i)	RtoSD (ii)	LSCV (iii)	LSP (iv)
Manila	234293	バンド幅	18.843	22.172	4.03E-10	81.998
		処理時間 (ms)	2080	2027	74239	765262
Cairo		バンド幅	75.641	62.279	4.03E-10	45.934
		処理時間 (ms)	1190	1180	40803	20041
Jakarta	133926	バンド幅	20.285	23.868	4.03E-10	64.682
		処理時間 (ms)	1639	1631	62825	649407
Ho Chi Minh	36757	バンド幅	28.106	33.072	4.03E-10	30.867
		処理時間 (ms)	926	879	18405	193713
Hanoi	97244	バンド幅	21.411	25.228	4.03E-10	40.318
		処理時間 (ms)	1309	1298	40130	419484

差, 手法 ii では四分位範囲だけが必要である. 一方, 手法 iii, iv では, コスト関数を用いた最適化プロセスが組み込まれている. このコスト関数を含むアルゴリズムでは, コスト関数を最小または最大にするバンド幅を探すため, 手法 i, ii よりも処理時間がかかっている. 図 5.1, 5.2 を見ると, 手法 i, ii において, 算出されたバンド幅に大差がない場合にも関わらず, 平滑化の結果が大きく異なることがわかる. 例えば, i, ii を用いた場合の Manila と Jakarta の平滑化の結果には, 10 カ所以上に峰が確認できる. しかし, Hanoi と Ho Chi Minh の平滑化の結果には, 3~4 カ所だけの峰が確認できる. 一方, 手法 iv を用いた場合, 平滑化の結果が都市によって, 大差がないことが確認できる. よって, 手法 i, ii を用いてバンド幅を算出した場合各都市の特性によって, 平滑化の結果が異なるが, 手法 iv を用いてバンド幅を算出した場合, 各都市の特性などに左右されずに, 有効な平滑化が実現できることが考えられる.

表 5.2: 各バンド幅選定手法の比較

バンド幅選定手法	計算時間	長所	短所
DataBased (i)	短	アルゴリズムがシンプル	都市の特性によって結果が大きく異なる
RtoSD (ii)	短	アルゴリズムがシンプル	都市の特性によって結果が大きく異なる
LSCV (iii)	長	—	本研究のデータに適していない
LSP (iv)	長	都市の特性に関わらず平滑化の結果が一定	アルゴリズムが複雑

適用実験の結果, 各バンド幅選定手法の特徴を整理すると, 表 5.2 のようになる. カーネル密度推定を用いた平滑化では, 各バンド幅選定手法において, 真の分布を推定して算出されたバンド幅を使い, その結果推定された密度をもとに時刻を平滑化している. この



平滑化で扱うデータは1分区分切りにトリップの発生時刻であり、この時刻を平滑化している。つまり、1分区分切りという不連続なデータの分布から、真の分布を仮定し、密度推定を行っている。その結果、手法 i, ii, iv のように、連続的な滑らかな分布を真の分布と推定した場合は有効に平滑化できるが、手法 iii は、元データの不連続な分布を真の分布として仮定しているため、推定された分布も不連続であり、その結果有効な平滑化が実現されていないと考えられる。

以上のことから、本研究では、手法 iv を用いてバンド幅を算出し、Gaussian 関数を用いたカーネル密度推定により、時間的平滑化を行う。

## 5.2 空間的平滑化手法の適用

図 5.3 は空間的補間前後におけるトリップの発生場所分布を示したものである。空間的平滑化を行うことで、トリップの発生場所の偏りが解消されていることがわかる。

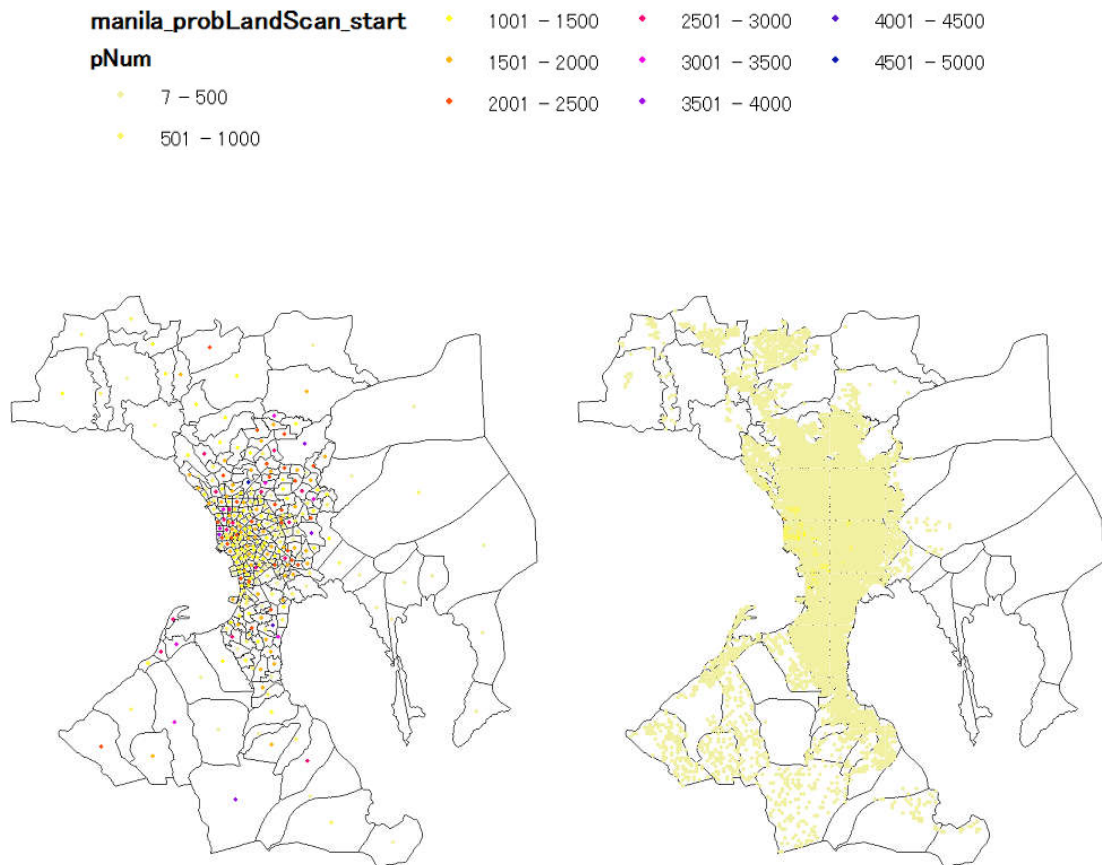


図 5.3: 空間的平滑化前後でのトリップ発生場所の分布の比較 (Manila)

(左図：空間的平滑化前, 右図：空間的平滑化後)

### 5.3 道路ネットワークデータ補正手法の適用

表 5.3, 図 5.4 からわかるように, ノードマッチング処理を行った結果, 経路探索を行うための OSM の品質が改善し, 経路探索の結果の精度が向上したことがわかる. ノードマッチング処理によって, リンク数・ノード数が増えた. リンク数が Manila では約 2.7 倍, Cairo では約 3.2 倍, Jakarta では約 2.7 倍, Ho Chi Minh では約 2.0 倍, Hanoi では約 2.4 倍増加した. また, ノード数が Manila では約 1.1 倍, Cairo では約 1.2 倍, Jakarta では約 1.2 倍, Ho Chi Minh では約 1.1 倍, Hanoi では約 1.2 倍増加した. このノードマッチング処理では, 補正されたのは, ほとんどが (a) の交差しているリンクにノード追加した場合であった. 図 5.4 からリンクが途切れている箇所が減少していることを確認できる.

表 5.3: ノードマッチング処理前後の比較

都市	結合箇所数	処理前		処理後	
		リンク数	ノード数	リンク数	ノード数
Manila	31	55282	93068	147042	106375
Cairo	–	14057	23940	44684	29499
Jakarta	757	16928	28212	45694	32653
Ho Chi Minh	–	18310	27151	36079	29419
Hanoi	23	3552	5633	8418	6472

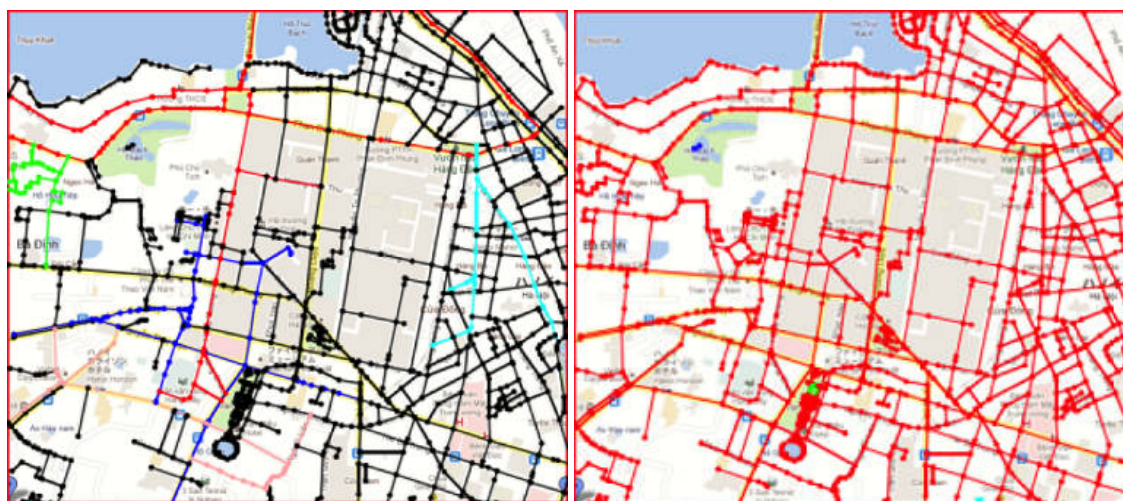


図 5.4: OSM グループ分けの結果 (左図: 補正前, 右図: 補正後)

以上のように, ノードマッチング処理をして, 補正された OSM を用いて経路探索を行った. また, 上で述べた手法で経路探索の結果の評価も行った. その結果を表 5.4, 図 5.7, 図 5.5, 図 5.6 に示す. JICA-PT データから得られる有効であるサブトリップの全経路を経路探索対象として, 実験を行った結果が表 5.5 である. 表 5.5 からノードマッチング処理をす



図 5.5: 補正前後 OSM を用いた経路探索の結果の比較 (例 1)

(左図：補正前, 右図：補正後)



図 5.6: 補正前後 OSM を用いた経路探索の結果の比較 (例 2)

(左図：補正前, 右図：補正後)



ることにより、経路探索ができるようになっていくことがわかる。図 5.7 は、補正前・後の OSM を使用して、経路探索をした際の迂回率の分布を示したものである。ノードマッチングの処理をすることによって、迂回率が小さくなり、ほぼ全ての経路（Manila: 98.2%, Cairo: 88.2%, Jakarta: 81.2%, Ho Chi Minh: 85.5%, Hanoi: 90.1%）で、迂回率が 2 未満になっていることがわかる。また、図 5.5, 図 5.6 はノードマッチング処理前・後でそれぞれ経路探索を行った結果例を地図上に描画したものである。この例においては、ノードマッチング処理前で得られる経路は、最短経路ではないことがわかる。それに対して、ノードマッチング処理後は、最短経路をとっていることがわかる。ただし、Manila においては、ノードマッチング処理によって、経路探索の結果の改善はあまり見られなかった。これは、表 4.3, 表 5.3 の結果からもわかるように、ノードマッチング処理前の OSM を使用しても経路探索の結果の精度がすでに高く、99.5%の経路で経路探索ができて、99.8%の経路の迂回率が 2 以下であった。

このように、ノードマッチング処理を行うことにより、OSM の抱えていた、経路探索を行う上での問題点を解消することができた。ただし、ここで行った経路探索の結果の評価は、経路の組み合わせ単位での集計となっているため、トリップ単位での集計となっていないことに注意されたい。

表 5.4: 経路探索の結果の評価

都市	有効トリップ	処理前			処理後		
		経路探索対象	経路探索	経路探索	経路探索対象	経路探索	経路探索
		のトリップ数	成功数	成功率	のトリップ数	成功数	成功率
Manila	395606	376341	4082	1.1	373762	367146	98.2
Cairo	244165	215815	3640	1.7	211394	186468	88.2
jakarta	1056196	922802	40621	4.4	877204	712330	81.2
Ho Chi Minh	259763	250775	2644	1.1	243010	207871	85.5
Hanoi	186608	161243	5011	3.1	149333	119015	90.1

表 5.5: OSM データの品質評価の結果（補正処理後）

都市	処理前		処理後	
	経路探索成功率 (%)	迂回率が 2 未満の経路数		
Manila	1.1	815 (57.1%)	98.2	3111 (90.5%)
Cairo	1.7	139 (62.9%)	88.2	1175 (92.3%)
Jakarta	4.4	853 (56.6%)	81.2	5099 (84.1%)
Ho Chi Minh	1.1	129 (76.8%)	85.5	1700 (81.4%)
Hanoi	3.1	135 (55.1%)	90.1	647 (77.0%)

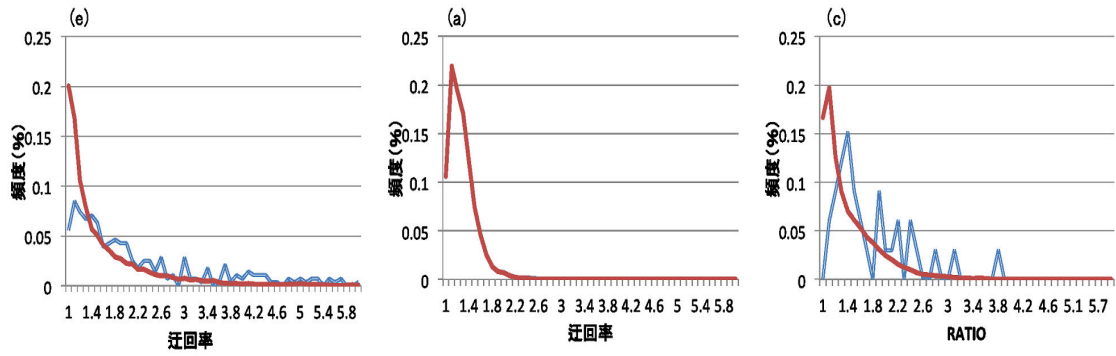


図 5.7: 処理前・後における迂回率の比較 ((a) Hanoi, (b) Manila, (c) Jakarta)

## 5.4 人の流れデータセット構築実験

### 5.4.1 人の流れデータセット構築の概要

第 4 章までで述べた手法を使い，対象都市の 5 都市でデータセット構築実験を行った．各種法の効果を確認するために，表 5.6 の 3 ケースで人の流れデータセットを構築した．また，データ補正などを行った補正済み JICA-PT データの概要を表 5.7 に示す．

表 5.6: データセットの構築を行った Case

Case	時間的平滑化	空間的平滑化	道路ネットワークデータ補正
Case1	✓		✓
Case2	✓	✓	
Case3	✓	✓	✓

表 5.7: 補正済み JICA-PT データの概要

都市	有効な出発・到着エリア ID を持つ トリップ数	全てのトリップで経路探索が 成功したサンプル数
Manila	80,824 (97.8%)	67,879 (82.1%)
Cairo	186,468 (88.2%)	110,051 (88.0%)
Jakarta	27,974 (99.7%)	23,530 (83.8%)
Ho Chi Minh	207,871 (85.5%)	78,962 (84.0%)
Hanoi	63,341 (99.4%)	49,905 (78.3%)

### 5.4.2 人の流れデータセットの考察

ここまでで、構築した人の流れデータセットの検証について述べる。国内での人の流れデータセット構築の研究では、人の流れデータセットから、12:00 と 24:00 における 3 次メッシュで集計したものと、国勢調査の昼間人口と夜間人口を同じく 3 次メッシュで集計したものとを比較し、検証を行っている。しかし、本研究の対象都市では、日本国内で得られるような正確かつ詳細な人口データは存在しない。そこで本研究では、LandScan の人口データを検証用のデータとして利用した。LandScan の人口データは、24 時間平均の人口分解能 30 分で集計したものである。以下の手順で検証を行った。

- 構築した人の流れデータセットから、0:00, 1:00, 2:00 …… 23:00 の人口を LandScan と同定義のメッシュで集計する。
- 上で集計した人口の平均を各メッシュで算出し、それを人の流れデータセットから取得した 24 時間平均のメッシュごとの人口とする。
- 人の流れデータから集計された人口と LandScan の人口をメッシュ単位で比較する。

なお、集計するメッシュは、調査エリアに含まれているメッシュのみで集計を行った。また、調査区に含まれる全てのメッシュでの集計に加え、各都市の中心部を抽出したメッシュでも集計を行った。

グローバルで利用可能なメッシュ集計された人口データである LandScan データと、構築された人の流れデータセットから集計された人口を比較した結果を以下に示す。ただし、第 3 章で精度評価を行ったように、LandScan データの精度は必ずしも高いとはいえないことに留意されたい。

### 5.4.3 活用事例

前述した通り、検証用データの入手が困難であることから、人の流れデータセットの検証を行うことは難しいが、本研究で構築したデータセットは、JICA-PT データを時空間内挿したものであり、出発・到着に関する情報しか含まれていない JICA-PT 調査の結果を、時間的に詳細な様々な側面から考察するのに、不可欠なデータセットである。以下に、その活用事例をいくつか示す。

- 都市内での人の流れの可視化

本研究で構築した人の流れデータセットからは、有効なサンプル全ての 24 時間の場合

表 5.8: LandScan データとの比較 (相関係数・総数)

都市	相関係数			人口総数 (データセット)			人口総数 (LandScan)
	case1	case2	case3	case1	case2	case3	
Manila	0.304	0.310	0.581	20,311,208	147,581	19,664,480	19,424,732
Manila (中心部)	0.235	0.246	0.339	11,385,760	113,563	11,185,698	9,206,244
Cairo	0.364	0.159	0.711	6,861,508	393,188	6,407,095	16,625,888
Cairo (中心部)	0.316	0.156	0.642	4,826,489	148,798	4,523,080	11,924,042
Jakarta	0.333	0.124	0.604	13,207,665	1,921,157	11,950,606	23,673,076
Jakarta (中心部)	0.174	0.141	0.451	6,126,106	184,131	5,919,893	8,852,126
Ho Chi Minh	0.478	0.154	0.745	5,995,090	187,860	5,064,751	7,600,481
Ho Chi Minh (中心部)	0.552	0.391	0.691	3,308	33,631	3,289,981	4,319,800
Hanoi	0.591	0.472	0.729	3,184,494	324,166	2,647,717	3,472,143
Hanoi (中心部)	0.601	0.557	0.724	2,183,079	109,886	1,806,598	2,084,831

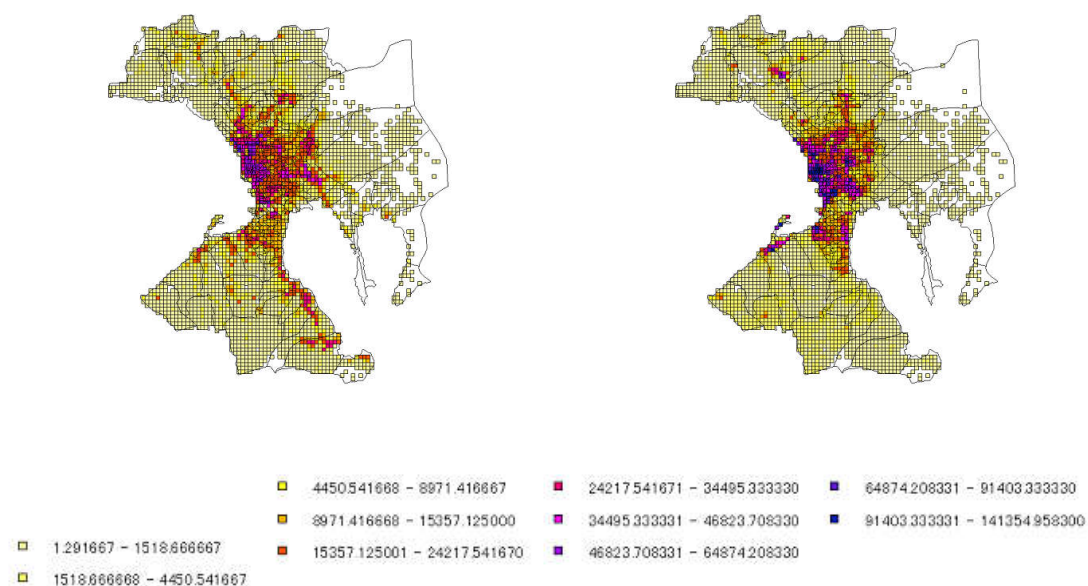


図 5.8: LandScan メッシュによる 24 時間平均の人口集計 (Manila)

(左図: LandScan, 右図: データセット)



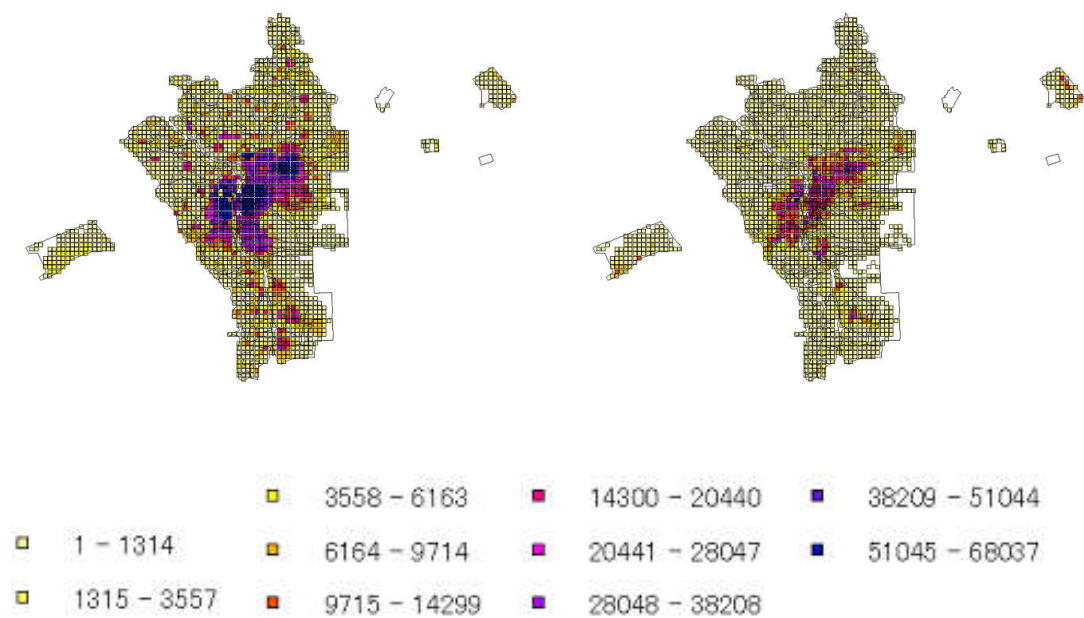


図 5.9: LandScan メッシュによる 24 時間平均の人口集計 (cairo)

(左図: LandScan, 右図: データセット)

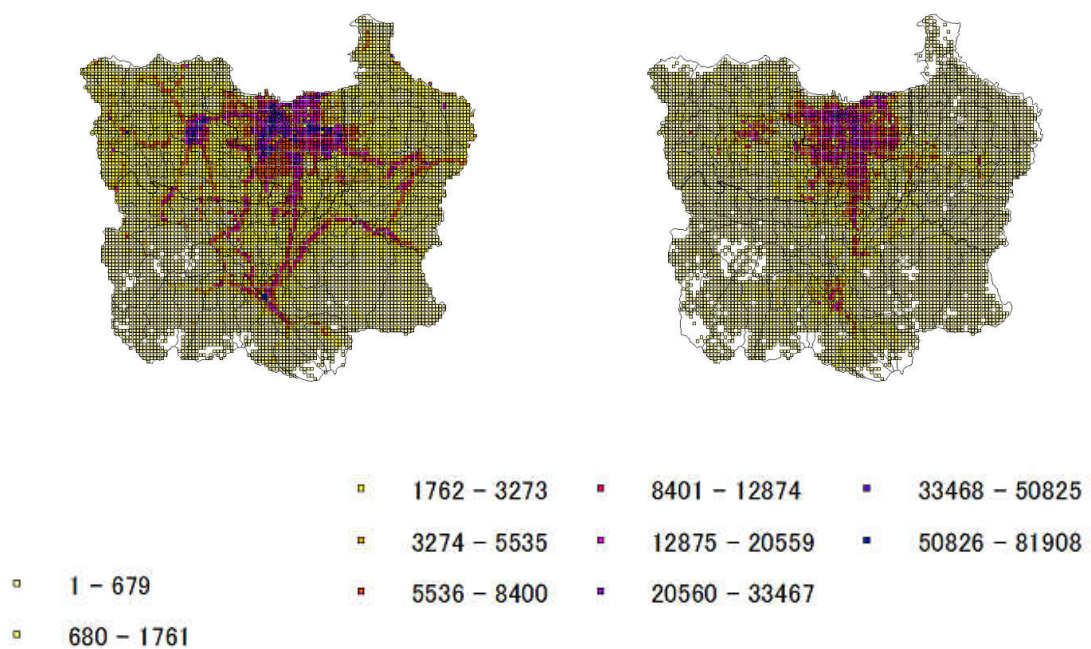


図 5.10: LandScan メッシュによる 24 時間平均の人口集計 (jakarta)

(左図: LandScan, 右図: データセット)

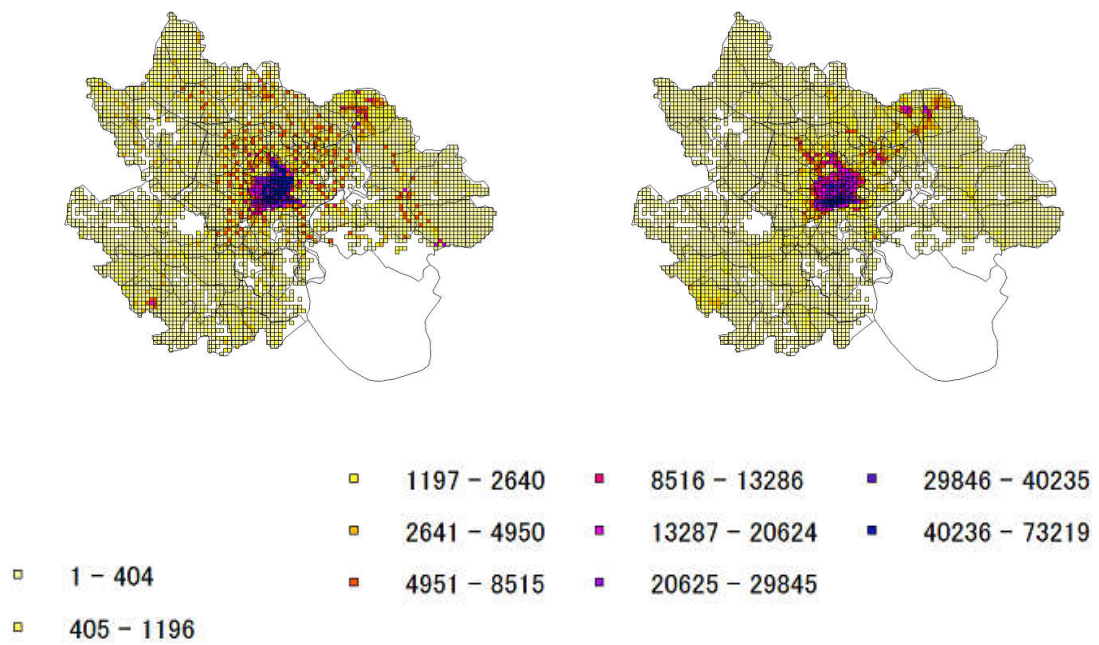


図 5.11: LandScan メッシュによる 24 時間平均の人口集計 (hcmc)

(左図: LandScan, 右図: データセット)

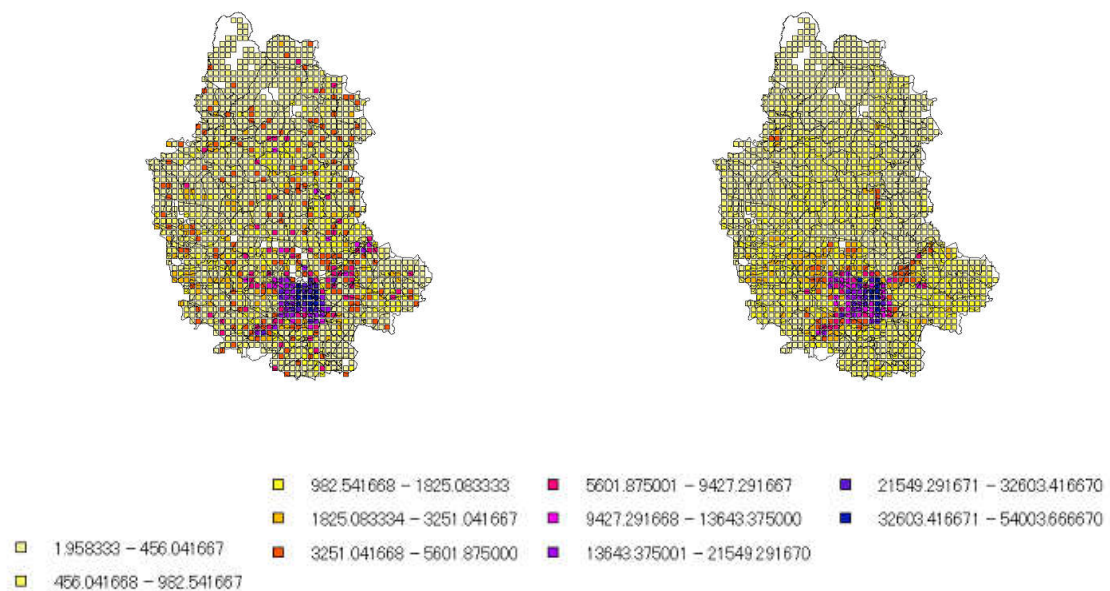


図 5.12: LandScan メッシュによる 24 時間平均の人口集計 (hanoi)

(左図: LandScan, 右図: データセット)

所を把握することができるため、人の流れデータセットを地図上に描画することで、都市内の交通を視覚的に容易に表現することができる。また、人の流れデータセットには、性別、年齢、トリップの目的、使用した交通機関など、都市内の交通において重要な属性情報も含まれているため、目的に応じた視覚を行うことができる。

一例として、以下に交通機関別に色分けした視覚化の例を示す。図5.13, 5.14, 5.15, 5.16, 5.17では、左上から、0:00, 3:00, 5:00, 6:00, 8:00, 9:00, 12:00, 15:00, 18:00, 21:00, 24:00に人の分布を示している。

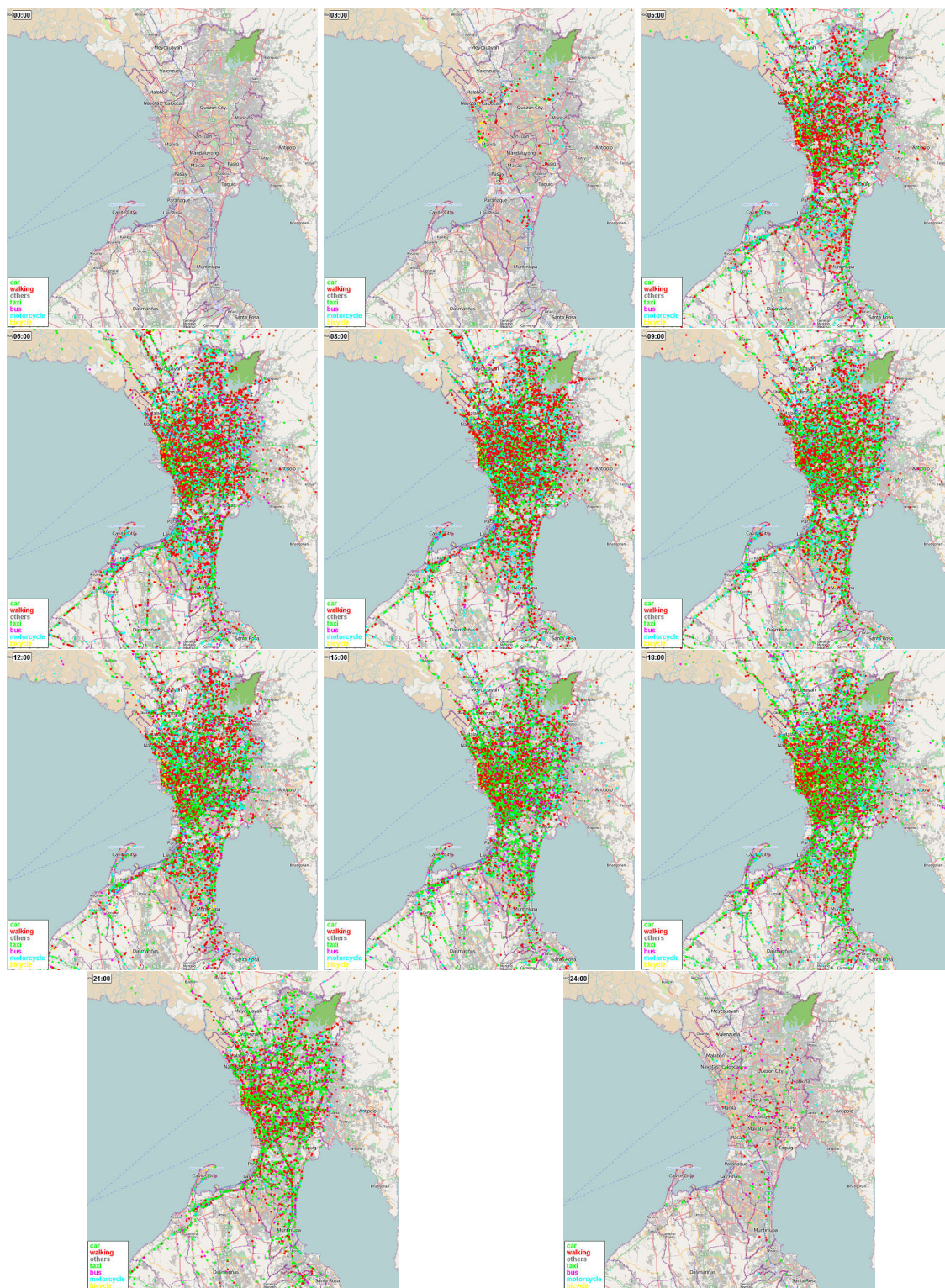


図 5.13: 交通機関別に色分けした人の分布 (Manila)