

2013 年度修士論文

時空間メッシュ集計データを用いた  
同化手法による人の流れの推定に関する研究  
Estimation of Actual People Flow from Spatiotemporal  
Aggregated Data Using Data Assimilation

戸井田 亮祐

Toida, Ryosuke

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

## 要旨

近年、交通状況の把握やマーケティング活動の支援、さらには災害発生時の緊急対応といった観点から非集計の位置情報に対するニーズはますます高いものとなっている。特に、東北地方太平洋沖地震の例からもわかるように、発災直後には大規模な鉄道の運休や道路の通行止めといった問題だけでなく、帰宅困難や交通渋滞といった問題が連鎖して発生する可能性があり、その事前、および事後の対応が重要となってくる。このための資料として「人の流れ」がある。人の流れは、1人1人の移動軌跡を把握することができるため、救援エリアや避難ルートの設定といったミクロな対応策に加えて、都市全体を対象とした問題にも適用することができる。さらに、都市で時々刻々と変化するリアルタイムな人の流れを把握ができれば、孤立した要救援者やなどに対して、迅速な対応可能となるであろう。現在、その利用可能性が様々な分野で検討されているが、個人個人の移動軌跡である非集計データは、データの性質上、非常に機密性が高く、多くの人が使えるようになるには、まだまだ時間がかかる可能性がある。

こうした背景を踏まえて、本研究では、比較的利用の容易な人の行動データから疑似的な非集計位置情報データを構築することとした。具体的には、静岡都市圏において、パーソントリップ調査データから人の流れのシミュレーションモデルを構築し、そのシミュレーションモデルに観測値として、GPSのメッシュ集計データを同化することで、モデルを補正し、より現実に近い人の流れデータの構築を試みた。

本研究では、まず静岡都市圏のPT調査データから1つのODに対して複数の経路候補を生成し、それをパーティクルとすることで、人の流れのシミュレーションモデルを構築した。次に、観測値データとして携帯GPSのメッシュ集計データである流動人口データを用いて、データ同化手法により、モデルの最適化を行った。データ同化のフィルタリングには、パーティクルフィルタを用い、それぞれパーティクルの最終的な尤度を求め、それを経路の選択確率とすることで、静岡都市圏における人の流れの全体像を推定した。

その結果、500m、1kmメッシュにおいては、比較的良好な結果を得ることができたが、時間帯により誤差が大きくなっていることが確認できた。特に通勤・通学、および帰宅の時間帯において、この傾向が顕著であり、これは実際の移動パターンの多様性をシミュレーションモデルで表すことができなかったためであると推察した。さらに、メッシュ交通量の可視化したところ、主要道路や含むメッシュにおいて、値の過大評価がみられ、主要道路の尤度が高くなりすぎていることが明らかとなった。また、研究手法を評価するために、クロスバリデーションを行ったところ、良好な結果を得ることができ、本手法の有効性を示すことができた。今回明らかとなった経路探索やパーティクル生成の手法の問題点については、今後の検討課題とした。

# 目次

1	序論	1
1.1	研究の背景	1
1.2	研究の目的	2
2	人の流れに関するデータ	3
2.1	パーソントリップ調査	3
2.2	動線解析プラットフォーム	6
2.3	GPS	10
2.4	携帯電話基地局データ (CDR)	13
3	本研究における人の流れ推定の概要	15
3.1	本研究の全体概要	15
3.2	本研究におけるデータ同化の位置づけ	16
3.3	パーティクルフィルタ	19
3.4	パーソントリップ調査の問題点	20
3.5	ネットワーク上の経路探索	21
4	人の流れ推定の手法	24
4.1	使用したデータ	24
4.1.1	パーソントリップ調査データ	24
4.1.2	携帯電話 GPS	25
4.2	パーティクルの生成によるシミュレーションモデルの構築	27
4.3	データ同化によるモデルの改善	34
4.4	推定結果の評価	36
5	推定結果の評価	37
5.1	計算結果について	37
5.2	メッシュ交通量の真値との比較	38
5.3	RMSE による評価	40
5.4	メッシュ交通量の可視化	41
5.5	クロスバリデーションによる手法の評価	44
6	NL モデルの適用によるパーティクル生成手法の改善	47
6.1	NL モデルについて	47
6.2	本研究で用いた NL モデルの概要	47
6.3	最尤法によるパラメータの推定結果	50
6.4	NL モデルの適用によるパーティクルの生成	51
6.5	NL モデルの適用による人の流れの推定結果	52
7	結論	53

7.1	まとめ .....	53
7.2	今後の課題 .....	54
	参考文献 .....	55

# 1 序論

## 1.1 研究の背景

近年、交通状況の把握やマーケティング活動の支援、さらには災害発生時の緊急対応といった観点から非集計の位置情報に対するニーズはますます高いものとなっている。

東北地方太平洋沖地震の例では、発災直後の大規模な鉄道の運休や道路の通行止めといった複数の問題により、都市内を移動する人々は大混乱に陥った。さらに、これに起因して、人々はその交通手段をバスや自家用車、徒歩に求めこととなり、交通渋滞や帰宅困難といった二次的な災害をも引き起こすことになった。こうした状況を踏まえ、首都圏では、災害時における鉄道の運行に関する協議会の開催（鉄道事業者、国土交通省）や帰宅困難者の抑制の責務を明示した条例の策定（港区）、幹線道路の通行規制等の防災訓練の実施（東京都等）など様々な取組みが進められている。

事前対応のための重要な資料の1つに「人の流れ」がある。人は都市を構成する重要な要素であり、人に着目した対応策を打つ場合には非常に有用なものであるといえる。特に、非集計の位置情報は、1人1人の移動軌跡を把握できるため、救援エリアや避難ルートの選定などのミクロな対応策に役立てることができる。さらに、それらをはるか上空から俯瞰することで、都市全体の動きとして把握することもでき、マクロな対応策にも用いることができるようになる。また、災害においては、事前対応に加えて事後、特に直後の対応が重要となってくるが、都市で時々刻々と変化する人の流れをリアルタイムで把握することができれば、孤立した要救援者や人・自動車の局所的集中に対して、効果的な対応ができるようになるだろう。

近年の非集計位置情報の状況としては、GPSの携帯電話への標準搭載やカーナビの普及等により利用価値の高まりを見せ、現在様々な分野での研究が行われている。携帯電話GPSをメッシュ集計した「混雑統計」（ゼンリンデータコム）は、携帯利用者から許諾を得て取得した位置情報を統計処理したもので、地域に流入する人口を時系列で可視化できるマーケティング・データとして提供されている。震災発生後には、本田技研工業やパイオニアなど各社によるプローブカー情報をまとめた「通行実績情報マップ」をITS Japanが公開して被災地内での移動・救援・支援活動に大きく役に立ったことから、その重要性が改めて認識された。また、学術方面では、関本ら（2012）が行った東京都市圏の震災時行動の分析や生形・関本（2012）が行った石川県の観光実態の分析などがあり、産学官が連携して、その可能性が検討されている。しかし、個々人の移動軌跡である非集計データは、データの性質上、非常に機密性が高く、多くの人が使えようになるには、まだまだ時間がかかる可能性がある。

一方、関本らの「人の流れプロジェクト」では、パーソントリップ（PT）調査デー

タを基にして、独自処理した人の流れデータの提供を行っている。パーソントリップ調査とは、「どのような人が」、「どのような目的で」、「どこからどこへ」、「どのような交通手段で」移動したかなどを調べるアンケート調査で、こうした調査データを用いることで、都市における人の流れをある程度再現することが可能となるであろう。しかし、属性ごとに与えられた拡大係数による単純な全数拡大では、同じ属性を持つ数十人の人が1日中同じ行動をしたことになり、現実と乖離した状況となる可能性が非常に高い。

## 1.2 研究の目的

こうした背景を踏まえて、本研究の目的を、比較的利用することが容易な人の行動データを用いて、疑似的な非集計位置情報データを構築することとする。それにより、こうしたデータの利活用が少しでも促進され、様々な分野で役に立つものとなることを目指す。具体的には、静岡都市圏において、パーソントリップ調査データから人の流れのシミュレーションモデルを構築し、そのシミュレーションモデルに観測値として、GPSのメッシュ集計データを同化することで、モデルを補正し、より現実に近い人の流れデータの構築を試みる。

## 2 人の流れに関するデータ

過去から現在にかけて、人の流れに関する大量のデータが観測、そして蓄積されている。それは、従来の紙ベースの統計調査から、GPS やカメラ、IC タグといった最新技術による計測データまで様々である。東京大学空間情報科学研究センター「人の流れプロジェクト」では、これら大量のデータをビジネスレベルで流通可能にすべく、研究者や実務者などを対象とした時空間サービスの実現を目指している。そこで本章では、「人の流れプロジェクト」で扱われている人の流れデータを中心にいくつかの事例を紹介する。

### 2.1 パーソントリップ調査

パーソントリップ調査とは、一定の地域における人の動きを調べ、交通機関の実態を把握する調査のことで、交通実態調査とも呼ばれている。昭和 42 年に広島都市圏で大規模に実施されて以来、これまでに日本各地の都市圏において数多く実施されている(図 2-1)。

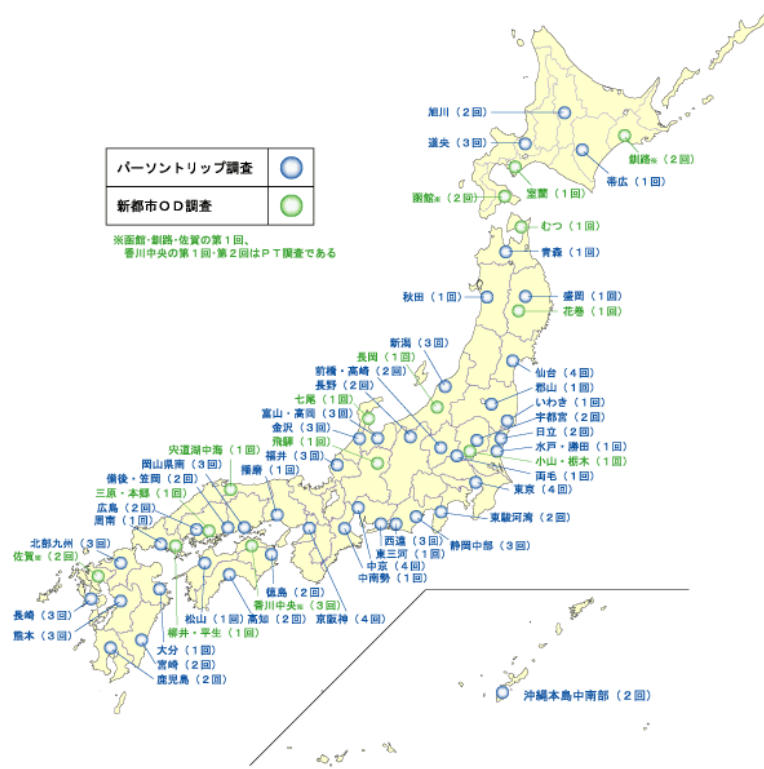


図 2-1 これまでの PT 調査実施都市とその回数  
(国土交通省都市交通調査・都市計画調査より引用)

調査方法は、無作為に選ばれた調査対象世帯に対して、調査票を配布することで行われ、調査対象者に一日の移動行動を回答してもらうことにより、「どのような人が」「どのような目的で」「どこからどこへ」「どのような交通手段で」移動したかなどを調べることができるものである。調査により得られた、交通行動の起終点、目的、利用手段、行動時間帯などの1日の詳細な交通データを用いることで、各交通手段の利用割合や駅や道路の交通量などを把握することができる。また、調査票に記載された性別・年齢などの個人の属性情報により、個人属性ごとのデータ抽出も可能となっている。

調査結果は、鉄道、バス、自家用車、自転車、徒歩などの交通機関相互の関連を含めた現在および将来の総合的な交通計画・マスタープランの策定や都市交通における個別課題への対応といった各種検討のための重要な基礎資料として活用することができる。

「パーソントリップ」における重要な概念として「トリップ」がある。「トリップ」とは、一人の人間がある目的を持ってある地点からある地点へ移動する際の、一方の移動を表す概念であり、同時にその移動を表す定量的な単位でもある（図 2-2）。そのため、移動の目的が変わるごとに1つのトリップとして数え、例えば、朝に自宅を出て会社に到着し、夕方に会社を出て自宅に帰った場合には、出勤1トリップ、帰宅1トリップの合計2トリップとなる。

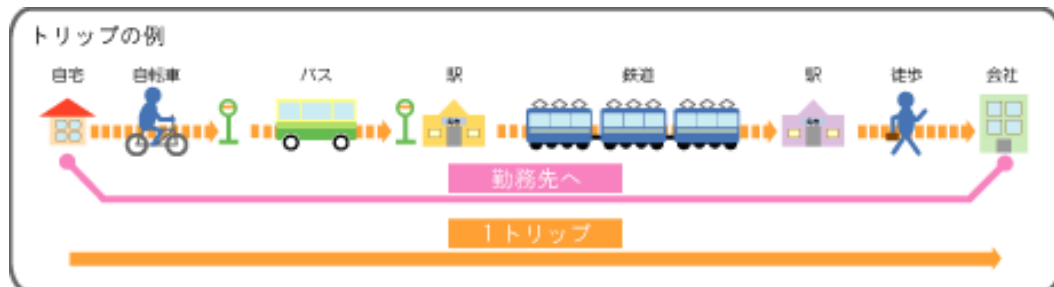


図 2-2 パーソントリップ調査におけるトリップの例  
(福井都市圏パーソントリップ調査より引用)

さらに、パーソントリップ調査における1トリップは1~複数個の「サブトリップ」で構成されている。例えば、図 2-2 のように、自宅から勤務先までの移動である「出勤」トリップを構成する交通手段が「自転車・バス・鉄道・徒歩」という4つに細分化される場合、このトリップは4つのサブトリップから構成されていることになる。このような交通手段ごとの細分化により、従来から行われている単一の交通手段を対象とした調査に比べ、鉄道や自動車、徒歩といった各交通手段の利用割合や交通量などが求められるだけでなく、ある目的をもった移動（トリップ）についても把握できるようになる。

ここで、パーソントリップ調査結果について具体例を示す。東京都市圏交通計画協議



会はホームページ上で第5回東京都市圏パーソントリップ調査結果として、いくつかの分析事例を公開している。

図2-3は、都市圏居住者の総トリップ数である。昭和53年から平成20年にかけて、そのトリップ数は1.07倍に増加しており、東京都市圏交通計画協議会のホームページによれば、これは都市圏の人口が増加したことによる影響だという。図2-4は地域別の発生集中量を示したものである。発生集中量とは、ある地域から出発したトリップの数（発生量）とその地域に到着したトリップの数（集中量）の合計のことであり、東京区部での増加が顕著に表れていることがわかる。これも、都市圏居住者の総トリップ数と同様に、都市圏の人口が増加したことにより、東京区部を目的としたトリップが増加したためと推察される。

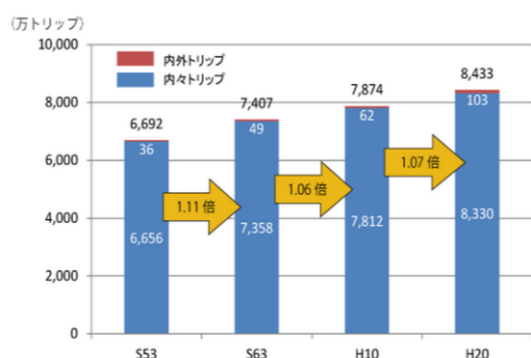


図2-3 都市圏居住者による総トリップ

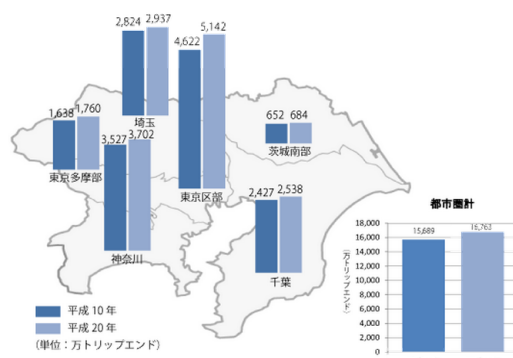


図2-4 地域別の発生集中量

(東京都市圏交通計画協議会より引用)

地域間の総トリップ数では、東京区部を発着地とするトリップが最も多く、なかでも東京多摩部、埼玉南部、横浜市、川崎市、千葉西北部間のトリップが大半を占めている(図2-5)。これは、主には通勤トリップだと考えられるが、これらの区間において、多くの人々が頻繁に移動していることを示している。地域別代表交通手段構成比では、東京区部や政令市などで鉄道の割合が増加し、自動車の割合が減少していることがわかる(図2-6)。それ以外の交通手段には、それほどの変化が見られないことから、都市圏における移動手段が自動車から鉄道へと徐々に移行していることを示した結果といえる。なお、都市圏の郊外部では、埼玉北部、千葉西南部、千葉東部、茨城南部などの一部地域で、自動車の割合の増加を見ることがもできる。

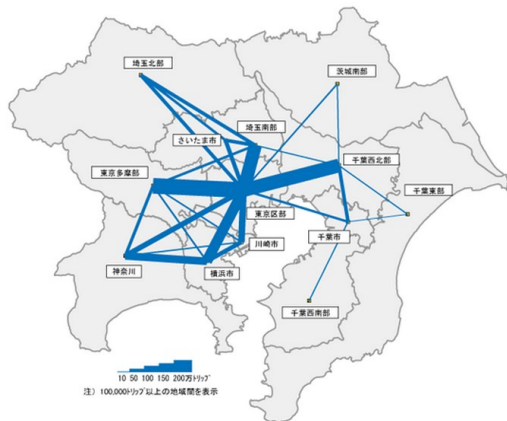


図 2-5 地域間の総トリップ数

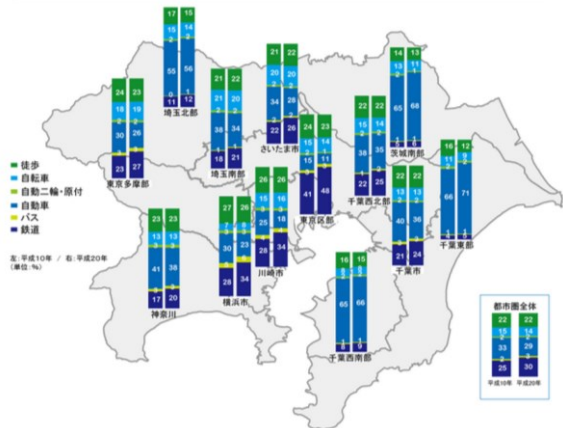


図 2-6 地域別代表交通手段構成比

(東京都市圏交通計画協議会より引用)

最後に、図 2-7 は鉄道利用者の着時間帯集中度であり、東京都市圏の鉄道を利用した通勤トリップの約半数が 8 時～9 時に集中していることが確認できる。

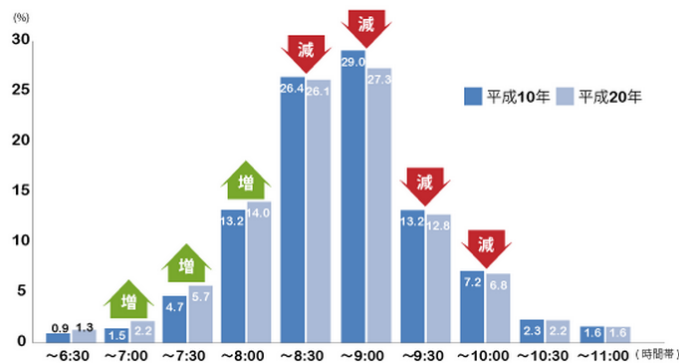


図 2-7 鉄道利用者の着時間帯集中度

(東京都市圏交通計画協議会より引用)

## 2.2 動線解析プラットフォーム

動線解析プラットフォームは、東京大学空間情報科学研究センター (CSIS: Center for Spatial Information Science) の人の流れプロジェクトで提供されているサービスで、その目的は、近年の技術革新によって、生み出された多種多様かつ大量のデータの品質確保と、その処理のために共通基盤の整備である。近年、GPS やカメラ、IC タグなど各種センサの発達により、従来のパーソントリップ調査のような統計調査による静的データに加えて、多種多様かつ大量の動的データが生み出されるようになった。これらのデータを広く一般利用として、流通可能にするためには、その品質をある程度そろえる

必要があり、調査・計測からデータの加工・蓄積、表示・提供に至るまで一連の流れをバランスよく効率的に処理する必要がある。動線解析プラットフォームでは、その共通基盤として機能することで、研究者や実務者などを対象とした時空間サービスの実現を目指している。図 2-8 は動線解析プラットフォームの役割を示したもので、各種の観測・統計データがプラットフォームを介することで、シミュレーションや行動分析、可視化などに利用できる時空間データとして扱うことができるようになる。



図 2-8 動線解析プラットフォームの役割  
(東京大学「人の流れプロジェクト」より引用)

動線解析プラットフォームでは、「データクリーニングサービス」と「データ提供サービス」という大きく 2 つのサービスを提供している。

まず、データクリーニングサービスには、「時空間内挿サービス」と「GPS データクリーニングサービス」「時空間内挿サービス」があり、時空間内挿サービスは、個人が所有するデータの位置情報に対して、道路や鉄道のネットワークをもとに起終点間の経路の推定を行い、1 分毎に時空間内挿した経路情報を取得することができるサービスである（図 2-9）。

GPS データクリーニングサービスは、GPS データを道路ネットワークに吸着させることで、観測誤差を取り除き、尤もらしいデータに補正するサービスである。現在、GPS 測位が容易になったことにより、そのデータ活用が課題となっているが、GPS 観測データは、その性質上、誤差や欠損値が多く含まれている。本サービスでは、これらができる限り取り除くことで、尤もらしいデータに補正し、GPS 観測データの活用を推進する目的がある。なお、GPS データの道路ネットワークへの吸着はマップマッチングと呼ばれる手法で、最近隣の道路ネットワークに位置が補正されるため、道路以外を移動する移動体には適さないとされている。この 2 つのサービス（「時空間内挿サービス」と「GPS データクリーニングサービス」）は、動線解析プラットフォームにユーザー登録することで利用可能となり、前者は Web-API の公開もしている。

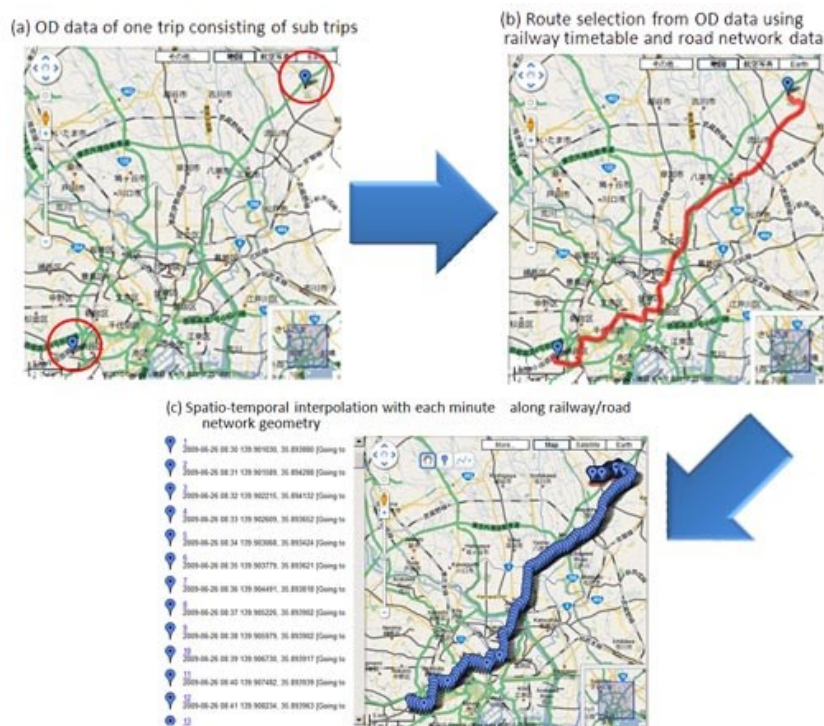


図 2-9 起終点間の経路探索と時空間内挿の例  
(東京大学「人の流れプロジェクト」より引用)

一方、データ提供サービスは、指定した検索条件によって、「人の流れデータ」から人の位置・時間情報を取得することができるサービスである。この人の流れデータは、東京大学 CSIS が提供する人の時空間位置データで、各都市圏のパーソントリップ調査データを基にして、CSIS が独自処理したものである。本データを利用するには、CSIS との共同研究申請が必要である（東京都市圏、中京都市圏、京阪神都市圏、マニラ都市圏以外については、各都市圏交通計画協議会にも利用申請が必要）。なお、現在、東京都市圏や京阪神都市圏などの国内データだけでなく、ダッカやマニラなどの国外データも利用可能であり、その数は約 30 種類である（図 2-10、図 2-11）。一部の都市圏では、空間配分版のデータが提供されており、より詳細な人の流れデータを利用することが可能である。通常版のデータが移動の起終点位置を含まれるゾーンの代表点としていたのに対して、空間配分版では、住宅地図データなどの利用により、その位置が建物に確率的に再配置されている。これにより、ゾーンの代表点という同一の地点から、何人もの人々がぞろぞろと出発するというような現実では起こりえない状況を回避でき、より現実に即したデータとして利用することができる。

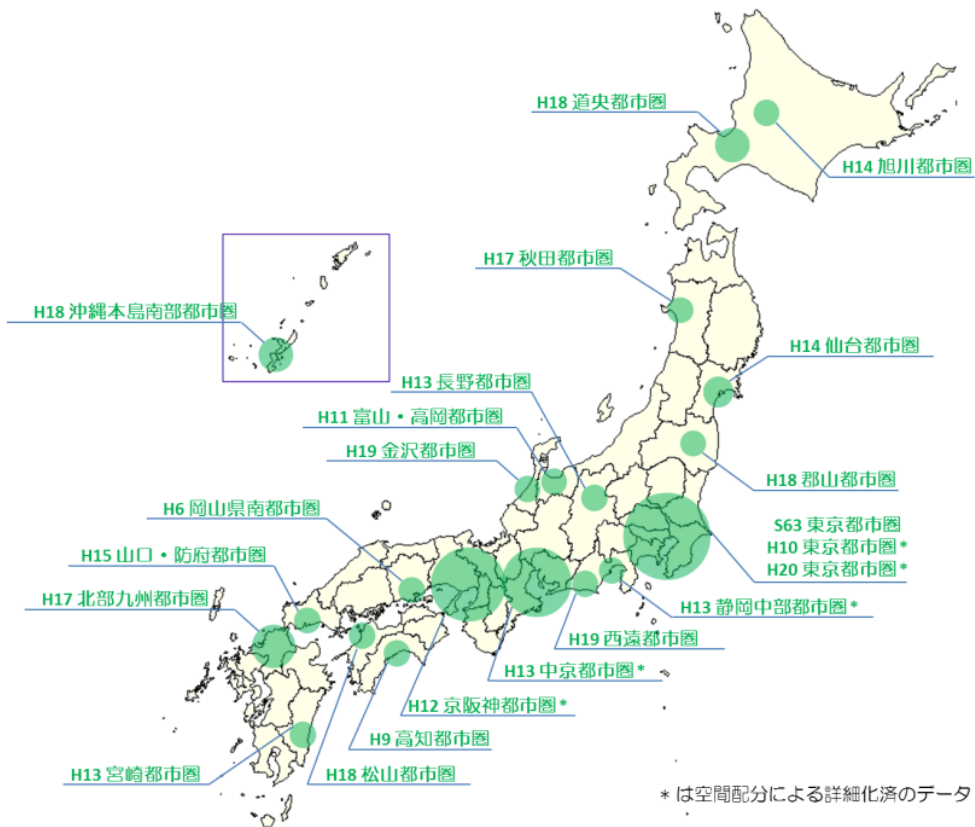


図 2-10 国内の利用可能データ  
(東京大学「人の流れプロジェクト」より引用)

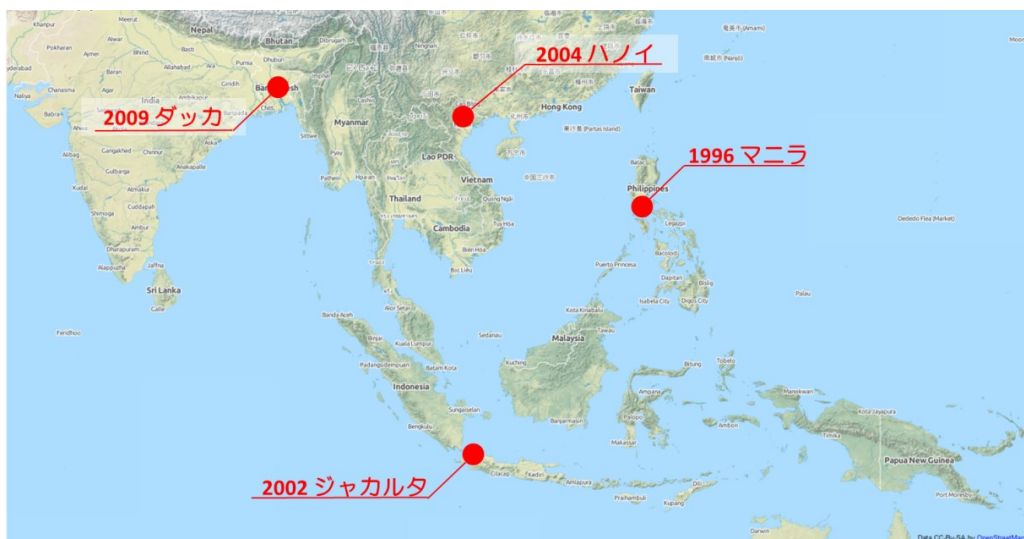


図 2-11 国外の利用可能データ  
(東京大学「人の流れプロジェクト」より引用)



## 2.3 GPS

GPS (Global Positioning System) とは、アメリカ合衆国によって運用される衛星測位システムのことで、上空にある複数の衛星からの信号を GPS 受信機で受け取ることにより、受信者は自身の現在位置を知ることができる。現在、20 個以上の衛星が地球を周回しており、現在位置を正確に知るためには、最低 3 個、より正確な位置であれば 4 個以上の衛星から電波を受ける必要がある。高層ビルなどの影響により衛星の電波を受信できないことがあり、3 つの衛星しか見えない位置では緯度・経度しか割り出せず、さらに電波の乱反射により、きわめて不正確な値となることがある。

しかし、2018 年に運用開始予定の準天頂衛星システム (QZSS: Quasi-Zenith Satellite System) により、この測位精度は大きく向上するとされている。前述したように、正確な位置測位を行うためには 4 機以上の衛星が見える位置から電波を受信する必要があるが、高層ビルの立ち並ぶ都市部や地方の山間部が多い日本では、これらが障害となり、低仰角の衛星からの信号では受信するのが難しかった。しかし、準天頂衛星システムは、少なくとも 1 機は常に日本のほぼ真上に位置するような特殊な軌道をとるため、高い仰角をとることができ、GPS とあわせることで、より高い精度の位置測位ができるようになる (図 2-12)。特に、東京では常に 70 度以上の高い仰角で 1 機以上の準天頂衛星を見通すことができるようになる予定である。すでに、2010 年に初号機であるみちびきが打ち上げられ、2016~17 年度には追加で 3 機の打ち上げが決定しており、宇宙航空研究開発機構 (JAXA) は 4 基体制でのシステム構築を目指している。



図 2-12 準天頂衛星と GPS 衛星の違い  
(JAXA より引用)

GPS は、前述したように軍事技術の 1 つであったが、現在では民生的用途でも多く利用されており、航空機や船舶などの管制システムはもちろんのこと、カーナビゲーションや携帯端末、デジタルカメラといった身の回りのものにまで搭載されるようになってきている。その利用も徐々に始まっており、ゼンリンデータコムでは、携帯利用者から許諾を得て取得した位置情報を、独自の分析プログラムにより「マーケティング・データ」として統計処理し、「混雑統計」という名で有償提供を行っている。ゼンリンデータコムのホームページでは、地域に流入する人口を、任意の期間を指定して日別・曜日別・時間帯別に見ることができる次世代マーケティング・データとされており、これを用いることで、国勢調査などの従来の調査だけでは難しかった時間帯別・曜日別の人の動きを把握することができ、従来のエリアマーケティング・データを補完するものとして活用できると説明されている。また、ゼンリンデータコムの運営する「いつも NAVI ラボ」では、メッシュ化された各エリアの混雑度を見ることのできる「混雑度マップ」が一般公開されており、リアルタイムで人の多いエリアを確認することができる（図 2-13）。

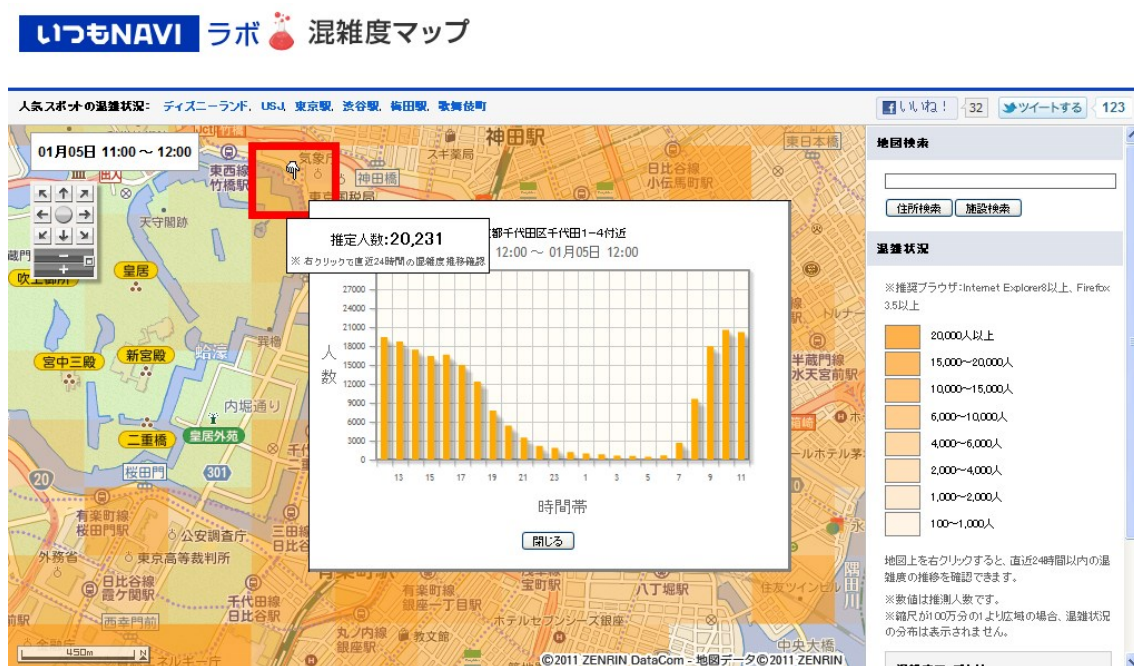


図 2-13 いつも NAVI のラボサイトで公開されている混雑度マップ  
(いつも NAVI ラボより引用)

2011 年東北地方太平洋沖地震発生直後には、インタープレミアムクラブ（本田技研工業）やスマートループ（パイオニア）、G-BOOK（トヨタ自動車）、カーウィングス（日産自動車）といった各社が運営するサービスから収集したプローブ交通情報をまとめた「通行実績情報マップ」を ITS Japan が公開し、その ITS 技術を活用した支援は災害時の ITS 活用のあり方を示した（図 2-14）。なお、プローブカーシステムとは、実

際に走行している自動車をセンサとして、データを取得し、交通流通や車両挙動、気候や路面状況などをモニタリングできるシステムのことである。すでに、シンガポール政府によるロードプライシングの料金設定のための走行速度計測（1998/04～）やトヨタやデンソーなどによるインターネットを使った交通情報配信とタクシー利用者へのコンテンツ配信などによる事業化評価およびデータ利用技術の開発（2002/01～）など、国内外を問わずに様々な実験がなされており、日本全国の車両に搭載されれば、新たな情報インフラとして活用される可能性がある。

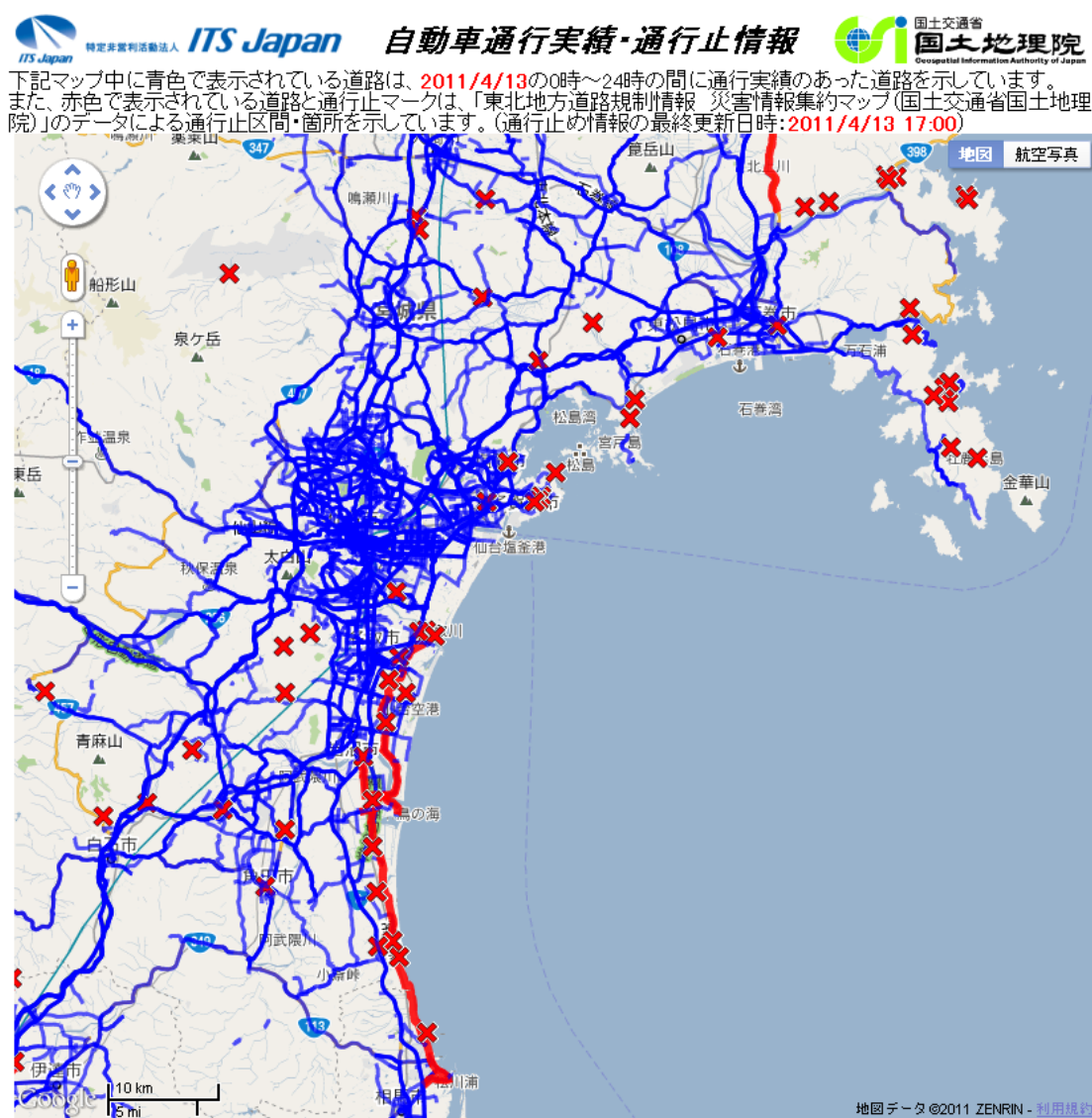


図 2-14 自動車通行実績・通行止情報  
(ITS Japan より引用)



これらの民間事例に対して、官公庁でも同様の取り組みが広がっている。観光庁では、GPS の位置情報を活用することにより、観光地における観光行動について調査・分析を進めている。これまであまり明らかでなかった観光来訪者の行動・動態について、GPS を使って、出発地や交通手段などのマクロ的な分析と滞在時間や地域内での動線などのミクロ的分析をあわせて実施することで、観光地域の課題把握、さらには対策の検討につながらないかと試みているようだ。

## 2.4 携帯電話基地局データ（CDR）

CDR とは、携帯電話の通信履歴のことで、1 回の通信要求ごとに 1 レコードずつ基地局側に蓄積されていくデータである。一般的に、CDR には、発着新卒番号、発信・着信・切断日時、切断理由コードなどが含まれており、携帯電話利用者の従量課金額の計算に用いるために記録されている。CDR は通信の秘密等に配慮して慎重に取扱う必要があるデータであるが、近年では、国内外において人の移動データとしての研究が様々な分野で行われており、その利活用方法が盛んに検討されはじめている。

NTT ドコモは、自社が所有する携帯電話ネットワークで周期的に把握している位置登録情報を利用し、基地局エリアごとの携帯電話台数を利用者の属性ごとに集計することで、人口の地理的分布を推計している（図 2-15）。この推計データは「モバイル空間統計」と呼ばれ、このデータを用いれば、エリアごとの人口分布や性別・年齢別・居住エリア別の人口構成を知ることができるようになる。現在、NTT ドコモでは、モバイル空間統計の活用を目指して、公共分野における機関や専門家と共同研究を行っている。

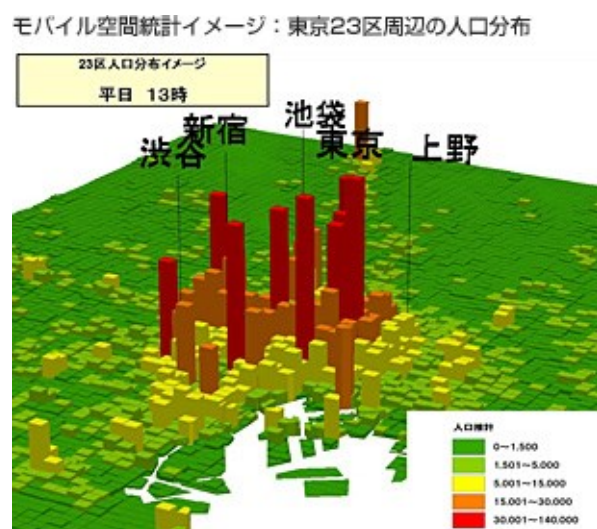
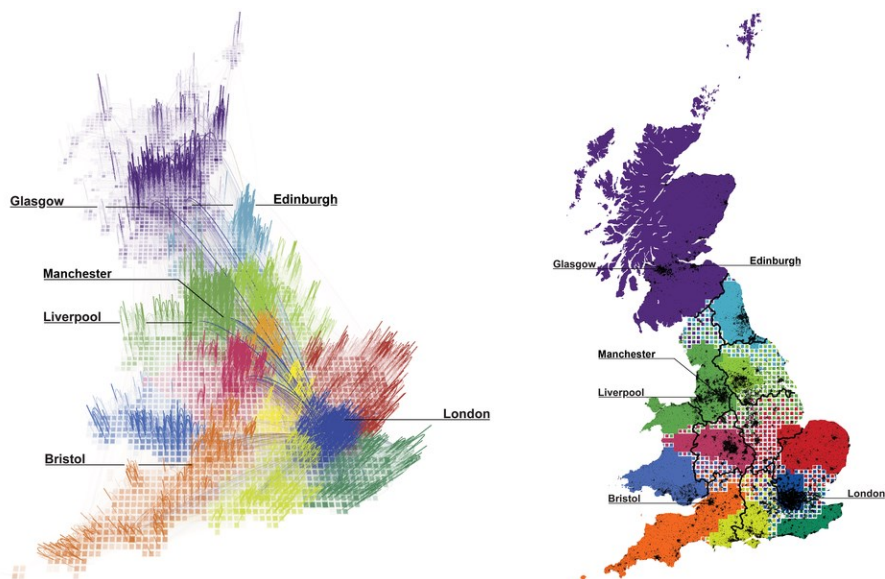


図 2-15 モバイル空間統計による人口分布のイメージ  
(NTT ドコモより引用)

一方、国外の取り組み事例としては、英国の BT がユニバーシティ・カレッジ・ロンドン(UCL)、マサチューセッツ工科大学(MIT)と共同で、CDR の地域間通信量に基づく通信圏と行政境界を比較することで、行政境界の再検討を行っている (図 2-16)。(a) は、それぞれのリンクの不透明度が 2つの地域間の通信時間を、色はそれぞれの地域を表している。



(a) 地域間通信量 (b) 通信圏と行政境界の比較

図 2-16 CDR から求めた地域間通信量と行政界の再検討  
([Carlo Ratti, 2010.12.]

Fraunhofer IGD や西イングランド大学(UWE)などの産学9組織が参加する EU の Urban API Project では、オーストリアの携帯電話事業者 A1 が商品として販売している CDR を利用して、携帯電話利用者の分布や移動先を推定している (図 2-17)。

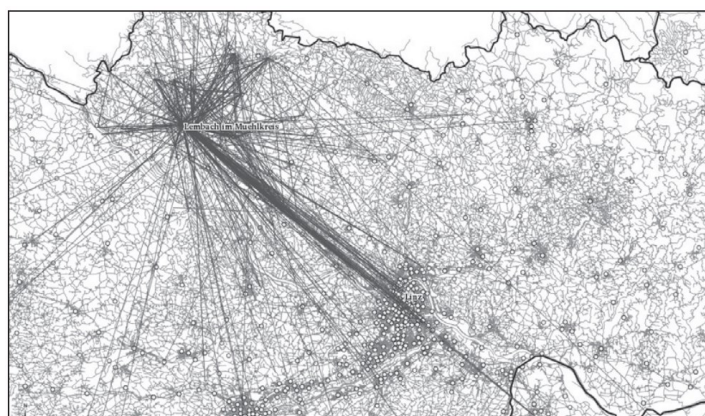


図 2-17 Lembech 市に住む携帯電話利用者の移動先の分布  
([Wolfgang LOIBL, 2012.]より引用)

## 3 本研究における人の流れ推定の概要

### 3.1 本研究の全体概要

本研究では、PT 調査データから人の流れのシミュレーションモデルを構築し、そのシミュレーション値に観測値を同化することで、そのモデルを補正し、より現実に近い人の流れデータの構築を試みた。その研究の流れを図 3-1 に示す。本研究は、シミュレーションモデルの構築、データ同化、推定結果の評価と大きく 3 つに分かれている。

まず、パーティクルの生成によるシミュレーションモデルの構築では、静岡都市圏の PT 調査データから 1 つの OD に対して複数の経路候補を生成し、それをパーティクルとすることで、人の流れのシミュレーションモデルを構築した。次に、観測値データとして携帯 GPS のメッシュ集計データである流動人口データを用い、データ同化手法により、モデルの最適化を行った。なお、データ同化のフィルタリングには、パーティクルフィルタという手法を用いた。パーティクルフィルタとは、パーティクルと呼ばれる解候補を複数生成させ、それぞれの最終的な尤度を求めることで、その尤度に従って解を求める手法である。ここでは、解候補を OD 間の経路候補とし、フィルタリングにより、経路の選択確率を求めることで、静岡都市圏における人の流れの全体像を推定する。最後に、推定値と真値の道路交通量を比較することで、推定結果の評価を行う。本章では、本研究で用いたデータ同化やパーティクルフィルタなどの手法について述べる。

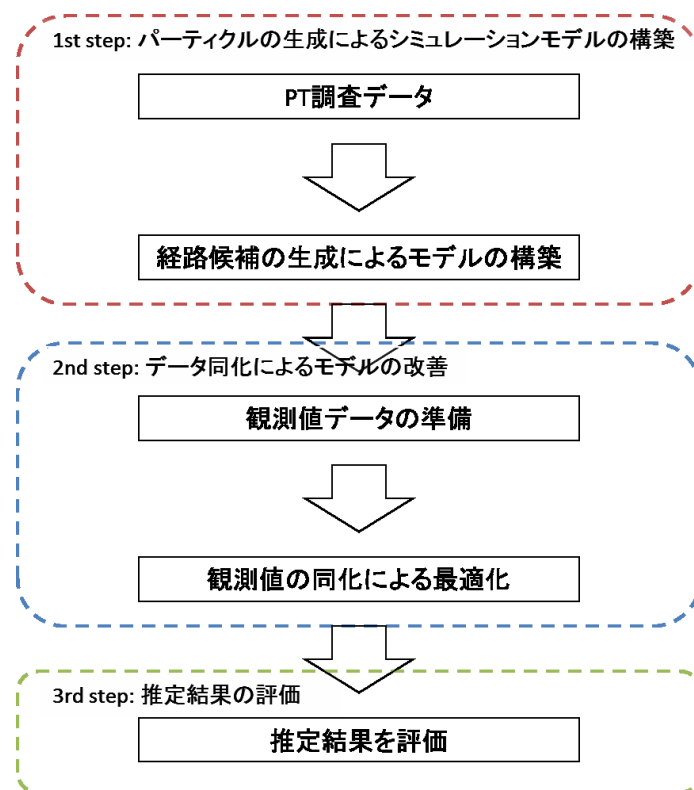


図 3-1 研究の流れ

### 3.2 本研究におけるデータ同化の位置づけ

本研究で用いたデータ同化とは、主に気象学や海洋学の分野においてたびたび用いられる手法である。これらの分野では、物理法則などに基づいて時空間シミュレーションモデルを構築し、実際の地球システムなどの再現、解明、予測を行なうことになる。その際に、システムの持つ非線形性が原因となり、初期条件や境界条件さらにはモデルに含まれるパラメータの与え方によって、シミュレーション結果が現実の現象から著しく乖離したものとなることがある。そこで、このシミュレーションモデルに実際の観測データを統合し、より精度の高いシステムの再現、適切な諸変数の推定や初期値の構成を目指すようになったのが、データ同化である。このようにして得られた諸変数を初期値として用いることにより、予測の精度が向上することが見込まれ、実際のシステムの理解に役立つことになる。

近年では、人の流動推定においても、データ同化手法を用いた研究が多くみられる。大瀧ら（2005）は携帯型装置による店舗内の人間の移動行動の推定を行っている。店舗内の移動行動をセンシングする際には、GPS などの測位システムやカメラ等の光学システムでは技術・コスト的な観点から難しいため、慣性航法（デッドレコニング）という手法を提案している。デッドレコニングでは、歩幅やその角度から移動距離と進行

方向を逐次推定し、それらを積算することで、移動軌跡を計算することができるが、誤差が累積していくという大きな欠点がある。大瀧らは、デッドレコニングによる推定結果に対して、パーティクルフィルタにより状態推定を行うことで、正確な移動経路を推定している。図 3-2 は提案手法による推定結果で、実線はパーティクルフィルタによる推定結果、破線は慣性航法のみによる推定結果を示している。デッドレコニングにより計算した移動軌跡にパーティクルフィルタを用いることで、安定した移動経路の推定が行えることが確認されている。

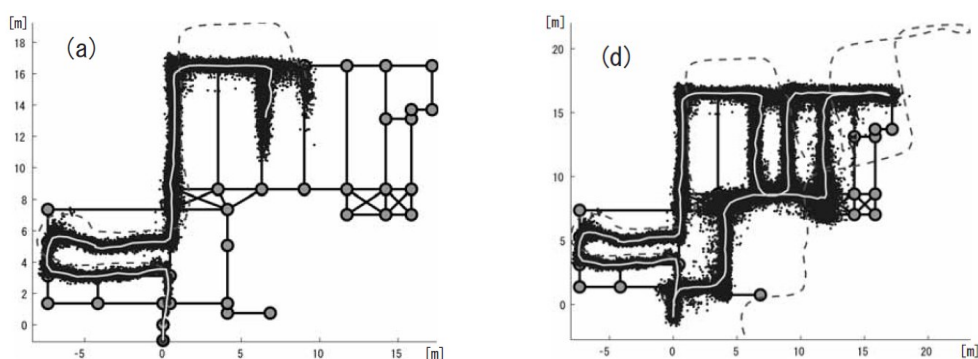
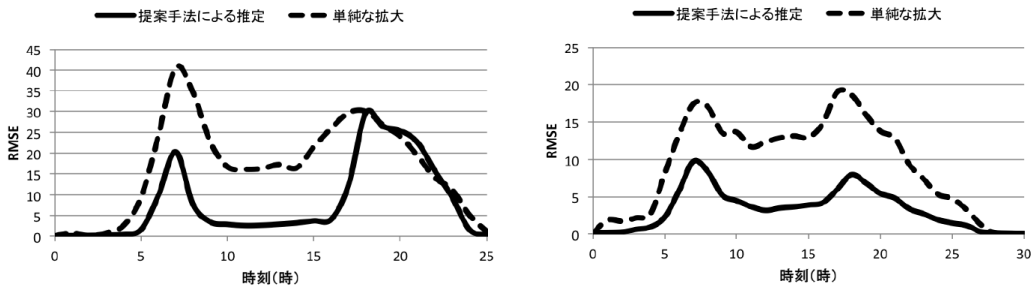


図 3-2 移動経路の推定結果  
 (「大瀧保明, 2005.6.」より引用)

一方、中村ら (2013) は都市圏レベルの大規模な人の流れの推定手法の構築を行っている。ここでも、パーティクルフィルタによるデータ同化手法が提案されており、その動的モデルに PT 調査データ基にしたトリップ集合からトリップ候補を生成する操作を、その観測モデルのトリップ候補を観測値により重み付けする操作をあてはめている。動的モデルでは、1 つのトリップにつき約の 10 のパーティクルをその重みに基づいて生成し、観測値として駅乗降客数や道路交通量といった統計データを同化することで、パーティクルそれぞれの重みを更新し、その最終的な重みから都市圏の人の流れを推定する。図 3-3 はその推定結果で、実線が提案手法による推定結果、破線が従来の単純な拡大による推定結果を示している。本論文では、提案手法によって、推定結果の大幅な改善を確認することができ、高精度な人の流れの推定を行えた結論付けられている。





(a) 時間帯別の駅乗車数 (b) 道路交通センサ対象道路の交通量

図 3-3 推定結果の RMSE による評価

(「中村敏和, 2013.」より引用)

図 3-4 に、本研究におけるデータ同化のイメージを示す。本研究では、PT 調査のサブトリップ OD 間にいくつかの経路候補を生成し、それをシミュレーションモデルと位置付けている。ここでは、OD 間の経路候補をパーティクルとして扱い、観測値として携帯 GPS の時空間メッシュ集計データを用いたデータ同化により、フィルタリングすることで、シミュレーションモデル全体の最適化を図り、尤もらしい人の流れデータの構築を目指した。

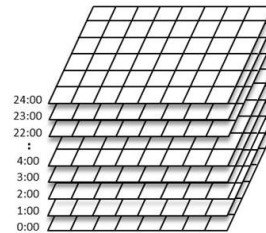
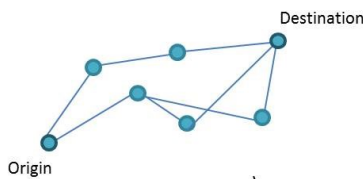
input

シミュレーションモデル

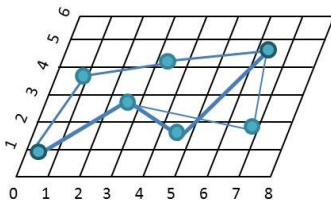
- PT調査データ
- 状態変数 (x, y)
- 遷移項 PT調査OD間の経路

観測値

- 携帯GPSのメッシュ集計データ
- 集計レベル: 250mメッシュ
- タイムステップ: 1時間



output



各パーティクルの重みは、観測データに基づいて更新され、重みが高いほど選ばれやすくなる。

図 3-4 本研究におけるデータ同化のイメージ

### 3.3 パーティクルフィルタ

パーティクルフィルタ（粒子フィルタ）は、モンテカルロ法による時系列フィルタの一種である。その特徴は、あらゆる種類の高次元の非線形・非ガウス型状態空間に適用できる点で、原理的に万能であることである。一般状態空間モデルにおける高次元の状態ベクトルの推定手法として 1996 年に提案されて以来、その実装の容易さから様々な分野で応用され、普及している。近年では、コンピュータービジョンの分野でも多く研究されており、対象トラッキングや輪郭抽出などのパーティクルフィルタを用いた画像解析手法がいくつも提案されている。そのアルゴリズムは非常にシンプルで、以下の式のように予測分布（昨日までのデータに基づく今日の状態）とフィルタ分布（今日までのデータに基づく今日の状態）のそれぞれの条件付き分布を、そこから得られた独立な実現値とみなせる多数のパーティクルを用いて近似できる。

$$\text{予測分布} \quad p(x_t | y_{1:t-1}) \cong \frac{1}{N} \sum_{i=1}^N \delta(x_t - x_{t|t-1}^{(i)}) \quad (1)$$

$$\text{フィルタ分布} \quad p(x_t | y_{1:t}) \cong \frac{1}{N} \sum_{i=1}^N \delta(x_t - x_{t|t}^{(i)}) \quad (2)$$

ただし、 $\mathbf{x}_t$  は状態、 $y_{1:t}$  は観測値、 $\delta$  はデルタ関数である。

これらのパーティクルの集団は、以下のアルゴリズムに従って、逐次的に求めることができる。なお、図 3-5 にパーティクルフィルタのアルゴリズムの概念図を示す。

- 1) 初期化 :  $i=1, \dots, N$  について  $k$  次元の乱数  $x_{0|0}^{(i)} \sim p_0(x)$  を生成する。ただし、 $p_0(x)$  は  $\mathbf{x}$  の時刻  $t=0$  における分布、つまり初期分布を示す。
- 2) 予測 :  $t=1, \dots, T$  について、パーティクルの状態を予測する。
  - a) 各  $i$  ( $i=1, \dots, N$ ) について、1次元の乱数  $v_t^{(i)} \sim q(v)$  を生成する。 $q(v)$  は、あらかじめ設定したシステムノイズ  $v$  の密度関数である。
  - b) システムモデル  $x_{t|t-1}^{(i)} = F_t(x_{t-1|t-1}^{(i)}, v_t^{(i)})$  に従って、各パーティクルの状態を予測する。
- 3) 尤度推定 : 予測された各パーティクル  $x_{t|t-1}^{(i)}$  について、その尤度を計算する。

尤度 
$$\tilde{w}_t^{(i)} = \frac{R_t(y_t | x_{t|t-1}^{(i)})}{\sum_{i=1}^N R_t(y_t | x_{t|t-1}^{(i)})} \quad (3)$$

- 4) フィルタ :  $X_{t|t-1} = \{x_{t|t-1}^{(1)}, \dots, x_{t|t-1}^{(N)}\}$  からそれぞれ  $x_{t|t-1}^{(i)}$  をそれぞれの尤度に比例する割合で復元抽出し、 $X_{t|t} = \{x_{t|t}^{(1)}, \dots, x_{t|t}^{(N)}\}$  を生成する。ただし、復元抽出とは、一度取り出したパーティクルを繰り返しサンプルしてもよい、重複を許したサンプル抽出である。

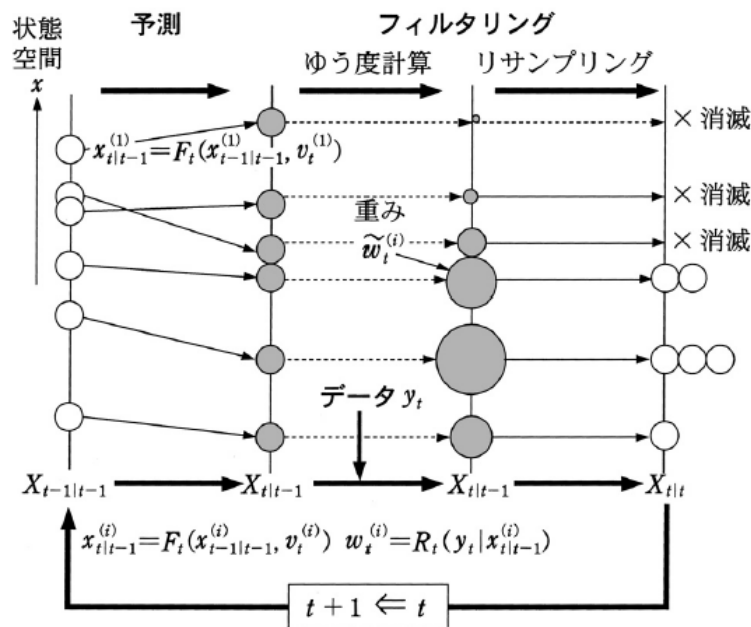


図 3-5 パーティクルフィルタのアルゴリズム概念図  
([樋口知之, 2013.]より引用)

### 3.4 パーソントリップ調査の問題点

パーソントリップ調査は、一定の地域における人の動きを調べるために、無作為に選ばれた調査対象世帯に対して、調査票を配布することで行われるサンプル調査である。そのため、調査対象の都市圏におけるすべての人の移動を把握することはできない。そのサンプル率はおよそ 2~3%程度であり、全体を把握するためには、統計的手法を用いてサンプル全体を人口ベースの全数に拡大する必要がある。最も一般的な方法としては、有効サンプルを属性ごとのカテゴリに分け、それぞれのカテゴリごとの有効サンプル率



の逆数を拡大係数として設定し、その係数により拡大を行う。カテゴリ分けする基準の属性としては、性別、年齢、職業、産業、世帯規模、居住地などがある。このようにサンプルを属性のカテゴリ別に拡大することで、調査対象者の死亡や転居、不在による調査不能や調査拒否などによって発生する個人属性間の調査率（回収率）の差異を補正することができ、推定量の偏りを取り除くことができる。これにより、単純な人口総数による比率で全員を等しく拡大しただけでは再現することができない現実の人の種類と比率を再現できるようになる。

個人属性のカテゴリ分けの具体例としては、個人が4変数(a, b, c, d)の属性を持ち、各変数が4種類(1, 2, 3, 4)の値をとり得るとすれば、この地区で抽出されたサンプルは $4^4=256$ 種類に分類することができる。この各属性のサンプル率の逆数を属性ごとに異なる拡大係数として設定することで、現実には即した調査データの拡大が行えるようになる。(a, b, c, d) = (0, 1, 2, 1)の値を持つ人が現実には100人、サンプルには4人いるとすれば、その抽出率は4%であり、その逆数である拡大係数は25となる。拡大係数25とは、その属性を持つ1人あたりが25人分の重みをもっているという意味であり、他の属性のデータと統合することで、調査対象地区の推定された人の流れや、交通機関の利用実態として扱うことができるようになる。

しかし、パーソントリップ調査のこの拡大手法には問題点もある。先ほどの具体例でいえば、属性(a, b, c, d) = (0, 1, 2, 1)の移動軌跡を持つサンプルが25人分の重みをもつということは、25人の人々が1日中全く同じ行動をとったということである。これは、非常に単純化された考え方であり、現実と大きく乖離した状況と言わざるを得ず、拡大手法にはまだ検討の余地があるといえる。

### 3.5 ネットワーク上の経路探索

本研究では、シミュレーションモデルを構築するに際して、PT調査のそれぞれのOD間について、経路探索を行うことで、いくつかの経路候補を生成した。ここでは、既存の経路探索手法について、簡単に説明する。

ネットワーク上の経路を探索する手法として最も有名なものに、Dijkstra法がある。Dijkstra法は、2点間の最短経路を効率的に求めることができるグラフ理論のアルゴリズムで、その単純なアルゴリズムがゆえに、OSPFなどのインターネットルーティングプロトコルや道路・鉄道などの経路案内など様々な分野で利用されている。図3-6のようにノードが接続されたグラフを例に、Dijkstraのアルゴリズムについて簡単に説明する。

- 1) 初期化：起点(s)のコストを0、他のノードのコストを未定義として設定する。
- 2) 全てのノードに対して、起点からの最短コストを確定させていく。
  - a) まだ確定されていないノードのうち、最小の値を持つノードを見つけ、確定ノ

- ードとする。
- b) 確定ノードから伸びているリンクをそれぞれチェックし、「確定ノードまでのコスト+リンクのコスト」を計算し、そのノードの現在値よりも小さければ更新し、接続するノードの情報もあわせて更新する。
  - 3) 全てのノードに対して、最短のコストを計算したら、処理を終了する。起点 (s) から終点 (g) までの経路は、それぞれのノードが持つ最短経路の隣接接続ノードを終点から辿ることによって得ることができる。

以上のステップを経ることで、起点から各ノードまでの最短コストとその経路を得ることができ、同時に起点から終点までの経路も得ることができるようになる。

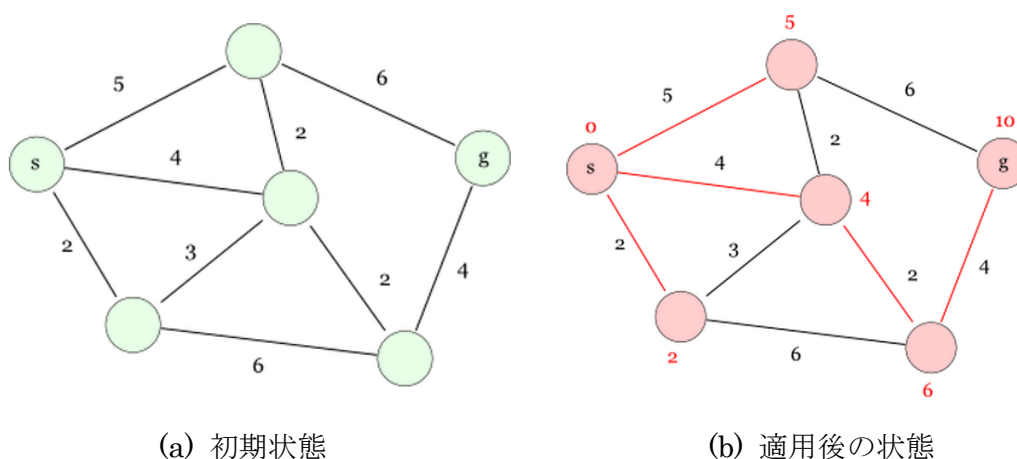


図 3-6 Dijkstra の適用例

A\*法は、各ノードから終点までのコストの推定値が既知であった場合に、適用できる手法で、最短経路を効率的に求めることができる。そのアルゴリズムは、Dijkstra法を応用したもので、簡単に言えば、推定値が小さい方ものから順に Dijkstra を適用していくものである。なお、この推定値は、実際の値と同等、ないしはそれより小さい必要があり、実際に適用する場合においては、各ノードから終点までの直線距離を用いることが多い。

複数の経路を求める手法としては、最も単純な手法に Penalty 法がある。Penalty 法も、基本的には Dijkstra を用いた手法であり、はじめに Dijkstra により最短経路を求めた後に、設定したコスト増分に基づき、その経路を構成するリンクのコスト（距離や時間）を増加させる。この処理を施すことで、再度 Dijkstra により経路を探索した時に、同一の経路が選ばれないようになり、異なる経路を複数取得できるようになる。ただし、幹線道路などの主要なネットワークも同様にコストを割り増して計算するため、生成される経路の多様性が失われる、あまり妥当でない経路を採用してしまうなど課題

も多くある。

一方、移動者の行動モデルを考慮した確率的な経路生成手法として、Dial 法がある。Dial 法は、ロジットモデルを用いる方法であり、効用最大化理論に基づいて、確率的に複数の経路を生成することができる。しかし、パラメータ算出の必要性があるため、計算コストが大きくなってしまふ、類似した経路をいくつも生成してしまふなどの問題点がある。

## 4 人の流れ推定の手法

本研究の手法は、大きく3つ分けられることは前述した。本章では、使用データと合わせて、その詳細について述べる。

### 4.1 使用したデータ

#### 4.1.1 パーソントリップ調査データ

都市内の人の流れを推定するに際して、静岡中部都市圏における平成13年の調査結果を使用した。データはcsv形式で、性別や年齢、職業といった個人属性、出発地や目的地ゾーン、移動の目的、拡大係数といったトリップ属性、さらに詳細なトリップの属性であるサブトリップ数や交通手段、出発・到着の位置座標、時刻など、様々な情報が含まれている。本調査結果の対象者は約6.5万人であり、対象地域は図4-1のようになっている。なお、本研究では、道路ネットワーク上の移動のみを対象としたため、元のデータから非鉄道利用ODのみ抽出して用いた。

非鉄道利用OD抽出後のデータの概要は表4-1に示すとおりである。ここでいう滞在トリップとは、ODの緯度経度が一致するものとしており、データに含まれる滞在か否かがわかる経路探索フラグの項目は用いなかった。表4-2からは、トリップ数 $\approx$ サブトリップ数であることが読み取れ、つまり目的をもった移動が単一の交通手段により行われているということである。この現象は、公共交通手段に頼らない地方都市特有の状況であり、東京などの大都市圏では見られないものである。また、滞在トリップは全トリップの半数以上を占めている。図4-3は、サブトリップOD間の距離の頻度分布であり、データのほとんどが0~10kmの間にあることが確認できる。

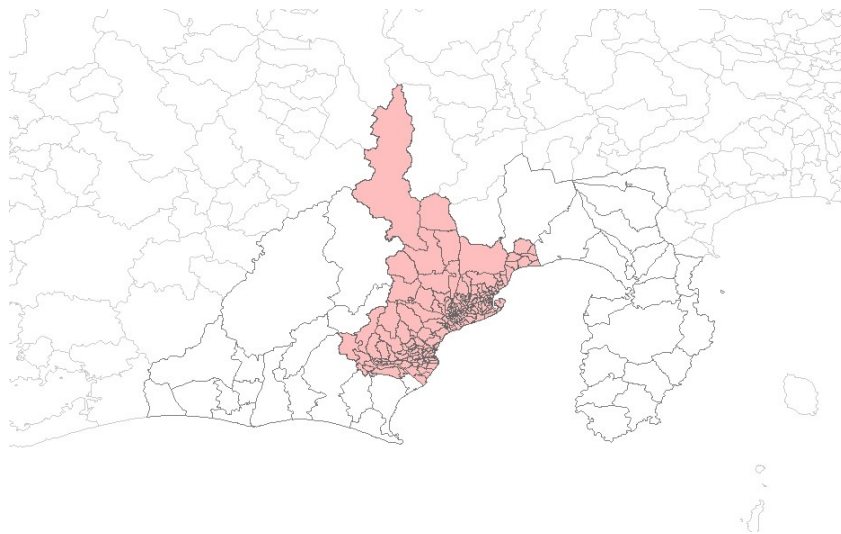


図4-1 パーソントリップ調査の対象地域

表 4-2 非鉄道利用 OD 抽出後の諸数

項目	人数	トリップ数	サブトリップ数	滞在トリップ
数量	64,984	412,060	423,271	237,186

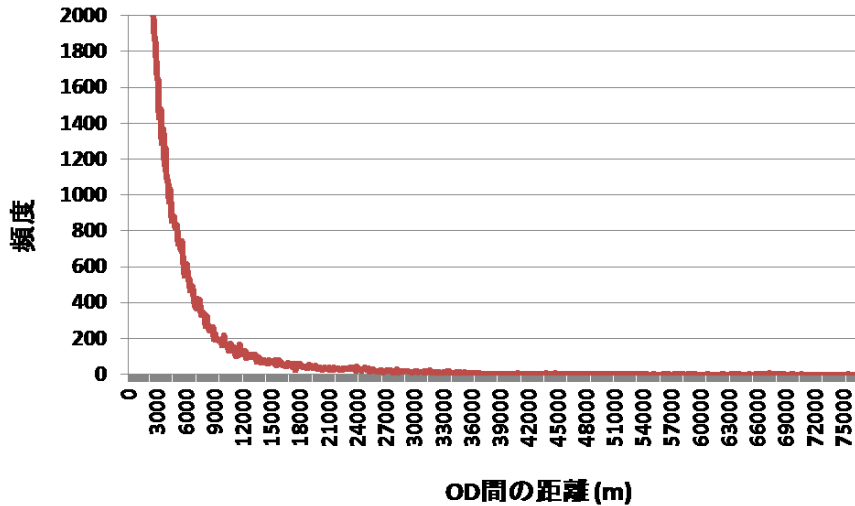


図 4-3 サブトリップ OD 間距離の頻度分布

#### 4.1.2 携帯電話 GPS

データ同化に用いる観測値として、ゼンリンデータコムに提供いただいた「流動人口」データを使用した。流動人口データとは、個人の所有する携帯電話から取得した GPS データをもとにして、メッシュ集計された統計データで、個人が特定されないよう秘匿処理が行われたものである。本データは、県 10 万円、全国 100 万円（年間使用価格）で実際に販売されているデータであるが、今回はゼンリンデータコム社の協力により、研究用途として特別に提供いただいた。

データのもととなるサービスは、「あらかじめ設定しておくだけで、お客様の位置情報をバックグラウンドで定期的に測位して、自動でサービス提供者に提供し続ける機能」と説明されており、これまでの利用者が GPS 機能进行操作することにより行われる測位とは異なり、該当サービスが提供される際に、利用者の承諾のもと最短 5 分間隔で自動測位し、圏外時を除けば、常に位置の取得・送信が行われる。位置測位については、衛星からの電波を利用する GPS 測位と基地局の情報を利用した測位の 2 種類を用いており、後者では誤差が 300m 以上になる場合がある。そのため、測位レベルをレベル 1（概ね 300m 以上）、レベル 2（概ね 300m 未満）、レベル 3（概ね 50m 未満）と 3 段階に分類している。データの該当者については、その元となるサービス利用者が 70~80 万人と言われていることから、人口の約 0.5%程度だと推測される。

流動人口データの仕様については、図 4-4 に示すとおりである。本データは、自宅人口、職場人口、流動人口という大きく分けて 3 つの項目からなっている。自宅人口はそのエリア内に自宅があると推定される人の総数、職場人口はそのエリア内に職場があると推定される人の総数、流動人口はエリア内に自宅・職場のどちらももたないと推定される人の総数であり、この 3 項目の値を合計することで、エリア内の通過人数を算出することができる。流動人口の通過人数の定義では、エリア内に GPS 記録が一瞬でもあれば、1 人として人数をカウントする。そのため、例えば新幹線に乗った人が、滞在することなく通過するだけの場合でも、1 人としてカウントされ、外部からの流入者が多く計測されてしまう可能性がある。さらに、そのデータの値には推定係数がかけられているため、人口ベースの全数に近い値でとらえることができる反面、データ上の 1 人の移動が 100 人単位の移動とみなされてしまうことがある。なお、データは、0~1 時、1~2 時といった具合に、1 時間ごとの通過人数のメッシュ集計値で示されているが、そのメッシュレベルは 250m である。

図 4-5 は、流動人口データの可視化例として 2011 年 3 月 11 日の 12~13 時の結果を示したものである。色が赤に近いほど、メッシュ内の通過人数が多いことを示しており、JR 静岡駅を中心として、人が多く移動していることが確認できる。なお、図は昼間の時間帯であるため、こうした結果となったが、夜などの別の時間帯であれば、全く異なる結果として示される。

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	2011/3/1	124740000	496001250	124747500	496012500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2011/3/1	124747500	496057500	124755000	496068750	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2011/3/1	124775000	496090000	124785000	496091250	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	2011/3/1	124792500	496293750	124800000	496305000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2011/3/1	124725000	497081250	124732500	497092500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	2011/3/1	124725000	497137500	124740000	497148750	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	2011/3/1	124740000	497103750	124747500	497115000	174	0	0	174	0	0	174	0	0	174	0	0	174	0	0	174
8	2011/3/1	124775000	497162500	124785000	497193750	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	2011/3/1	124785000	497250000	124792500	497305000	0	174	0	0	174	0	0	174	0	0	174	0	0	0	0	0
10	2011/3/1	124792500	496991250	124800000	497002500	0	348	0	0	348	0	0	348	0	0	348	0	0	348	0	0
11	2011/3/1	124590000	497430000	124597500	497441250	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	2011/3/1	124627500	497508750	124635000	497520000	174	0	0	174	0	0	174	0	0	174	0	0	174	0	0	174
13	2011/3/1	124635000	497565000	124642500	497576250	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	2011/3/1	124642500	497517500	124650000	497528750	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	2011/3/1	124650000	497283750	124657500	497295000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	2011/3/1	124650000	497373750	124657500	497385000	348	174	0	520	174	0	348	174	0	348	174	0	174	174	0	174
17	2011/3/1	124725000	497441250	124732500	497452500	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	2011/3/1	124725000	497283750	124740000	497295000	0	0	0	0	0	0	0	174	0	0	174	0	0	174	0	0
No.	年月日	最少緯度	最少経度	最大緯度	最大経度	流動人口 (0時)	自宅人口 (0時)	職場人口 (0時)	...	流動人口 (23時)	自宅人口 (23時)	職場人口 (23時)									
形式	yyyymm-dd	ミリ秒単位 (世界測地系)	ミリ秒単位 (世界測地系)	ミリ秒単位 (世界測地系)	ミリ秒単位 (世界測地系)	人数 (小数点切り下げ)	人数 (小数点切り下げ)	人数 (小数点切り下げ)	...	人数 (小数点切り下げ)	人数 (小数点切り下げ)	人数 (小数点切り下げ)									
データ型	Date	integer	integer	integer	integer	integer	integer	integer	...	integer	integer	integer									

図 4-4 流動人口データの仕様

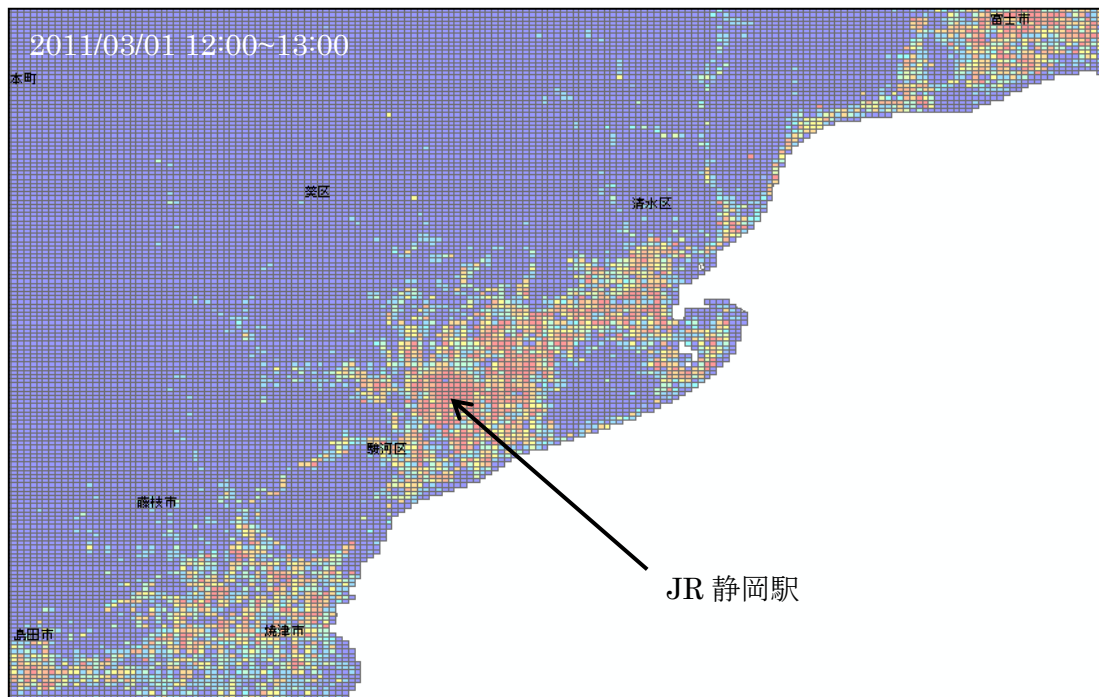


図 4-5 流動人口データの可視化例（12~13 時）

## 4.2 パーティクルの生成によるシミュレーションモデルの構築

まず、人の流れを推定するにあたり、静岡都市圏の PT 調査データからシミュレーションモデルの構築を行った。具体的には、PT 調査データのサブトリップ OD 間にそれぞれ 5 経路ずつの経路候補を生成させた。前述したように、一般的に、PT 調査データを全数拡大する場合には、層別拡大法という手法により、属性ごとに与えられた拡大係数を用いて拡大する。しかし、本研究では、1 つの OD 間に複数の経路候補を生成させ、多様性をもたせることで、より現実に近い人の流れを再現できるようにした。

本研究では経路の探索手法に、Penalty 法と Dial 法の 2 つの手法を検討した。Penalty 法とは、最短経路から順に複数の経路を探索するもので、最も単純な手法である。そのアルゴリズムは、以下のステップで説明できる。

- 1) 起点・終点間の最短経路を Dijkstra により探索する。
- 2) n-1 回以下のステップを繰り返す。n は探索する経路の数。
  - a) 取得した経路を構成するリンクのコストを、あらかじめ設定したコスト増分  $r$  に従って割り増す。

$$C(L[i \rightarrow j]) = C(L[i \rightarrow j]) * (1 + r) \quad (1)$$

$C(L[i \rightarrow j])$  はノード  $i$  から  $j$  の間にあるリンクのコスト



b) 更新後のリンクコストに基づいて、Dijkstraにより最短経路を探索する。

一方、移動者の行動モデルを考慮した Dial 法は、単純な最短経路ではなく、各リンクのコストから、分岐確率を算出し、確率的に経路を生成させる方法である。アルゴリズムは以下に示すステップで行われる。

1) 起点  $r$  から全てのノードまでの到達最小コストを計算する。

$$c(i) \leftarrow C_{min}[r \rightarrow i] \quad (1)$$

$c(i)$ は、起点からノード  $i$  までの最短コスト

2) 全リンクについて、リンク尤度  $L[i \rightarrow j]$  を計算する。

$$L[i \rightarrow j] = \begin{cases} \exp[\theta\{c(j) - c(i) - t_{ij}\}] & c(i) < c(j) \\ 0 & otherwise \end{cases} \quad (2)$$

リンク  $i \rightarrow j$  が、起点からノード  $j$  までの最短経路上にあるなら、そのリンク尤度は最大値 1 であり、遠回りであるほど尤度は小さくなる。なお、リンクを後戻りする場合、その尤度は 0 となる。

3) 各ノード  $i$  から流出するリンクのリンクコスト  $W[i \rightarrow j]$  を計算する。

$$W[i \rightarrow j] = \begin{cases} L[i \rightarrow j] & for\ i = r \\ L[i \rightarrow j] \sum_{m \in I_i} W[m \rightarrow i] & otherwise \end{cases} \quad (3)$$

4) 各ノード  $j$  に流入するリンクの交通量  $x_{ij}$  を算出し、分岐確率を計算する。

$$x_{ij} = (q_{rj} + \sum_{m \in O_j} x_{jm}) \frac{W[i \rightarrow j]}{\sum_{m \in I_j} W[m \rightarrow j]} \quad (4)$$

5) リンクに分岐確率を用いて、確率的に経路を探索する。

Dial は、単純なリンクコストのみによらない確率的な手法であるため、必ずしも最短経路が選ばれるわけではない、経路を探索するごとに生成される経路が異なるなどの特徴がある。なお、要求した数の経路が必ずしも得られるとは限らないことに注意する必要がある。Dial による経路探索の例を図 4-6 に示す。これは、代々木上原駅から東大生産技術研究所までの経路を探索した例で、左図の赤のネットワークは計算時に読み込まれたネットワーク範囲（デフォルトでは半径 1km）、右図は色の濃淡により生成された 5 つの経路を表している。単純な最短経路探索である Penalty による結果と比較すると、Dial では回り道することなく、それらしい経路が選ばれていることがわかる（図 4-7）。そのため、本研究では Dial による経路探索を用いて、モデルの構築を行うこととする。なお、Dial では、パラメータを変化させることで、生成される経路も変化させられるため、本研究においては、Penalty よりも優位な手法であると考えられる。



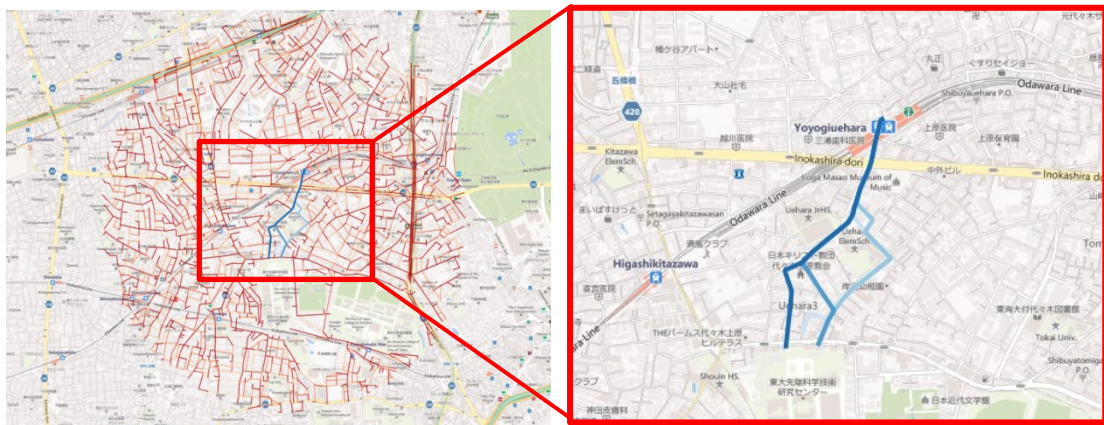


図 4-6 Dial 法による経路探索例

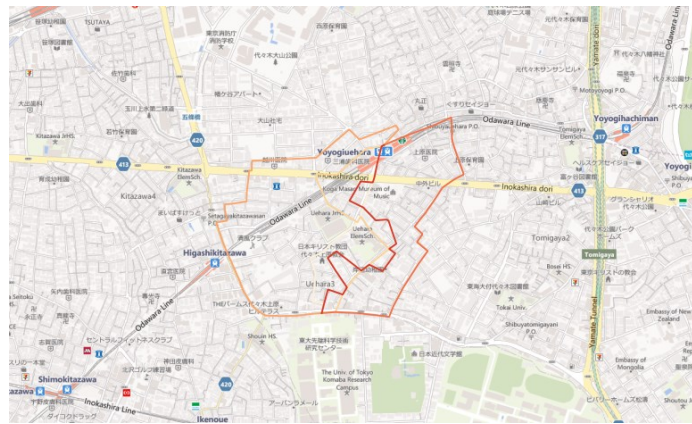


図 4-6 Penalty 法による経路探索例

Dial 法では、リンク尤度計算式のコストパラメータ  $\theta$  の値を変えることで、その生成経路に影響をもたらす。具体的には、 $\theta$  の値が大きくなるにつれ、尤度の高いリンクほど、その尤度が割り増しされる、つまり  $\theta$  の値が大きくなるにつれ、似たような経路が選ばれる確率が上がるということである。そのため、いくつかのパターンを試行し、実験で用いる  $\theta$  の値について検討した。

OD 間距離が 5km、10km、15km の時の経路探索結果を図 4-7、4-8、4-9 にそれぞれ示す。 $\theta=0.0001, 0.001, 0.01, 0.1$  と 4 パターンについて試行した結果、OD 間距離 5km、 $\theta=0.1$  では、生成される経路が 2 本と極端に少なくなることが確認できた。これは、主要道路の尤度が高くなりすぎたことにより、生成された経路が重複してしまったため、OD 間距離が 5km 程度の短いトリップや、山中などそもそもの道路リンク数が少ないエリアでは、このようなことが起こる可能性が高いことが示された。逆に、 $\theta$  の値が低すぎると、主要道路の尤度が低くなりすぎてしまい、尤もらしい経路が選ばれなくなる可能性が高い。そのため本研究では、OD 間距離 10km 以下のトリップが大部分であることを鑑みて、 $\theta=0.01$  をコストパラメータの値として採用した。

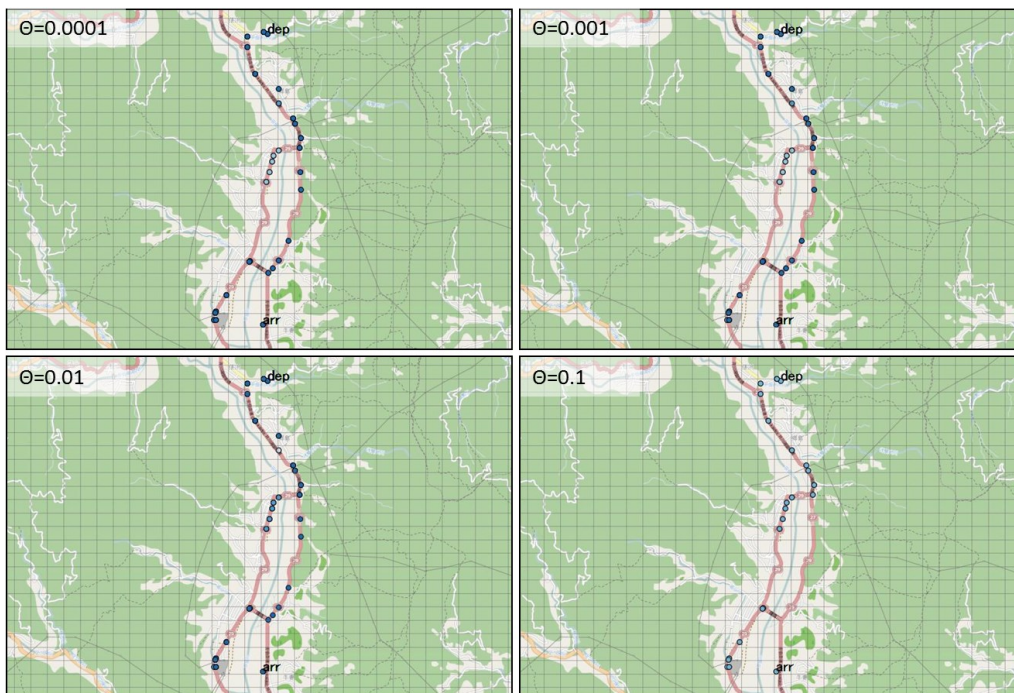


図 4-7 OD 間距離 5km の経路探索結果

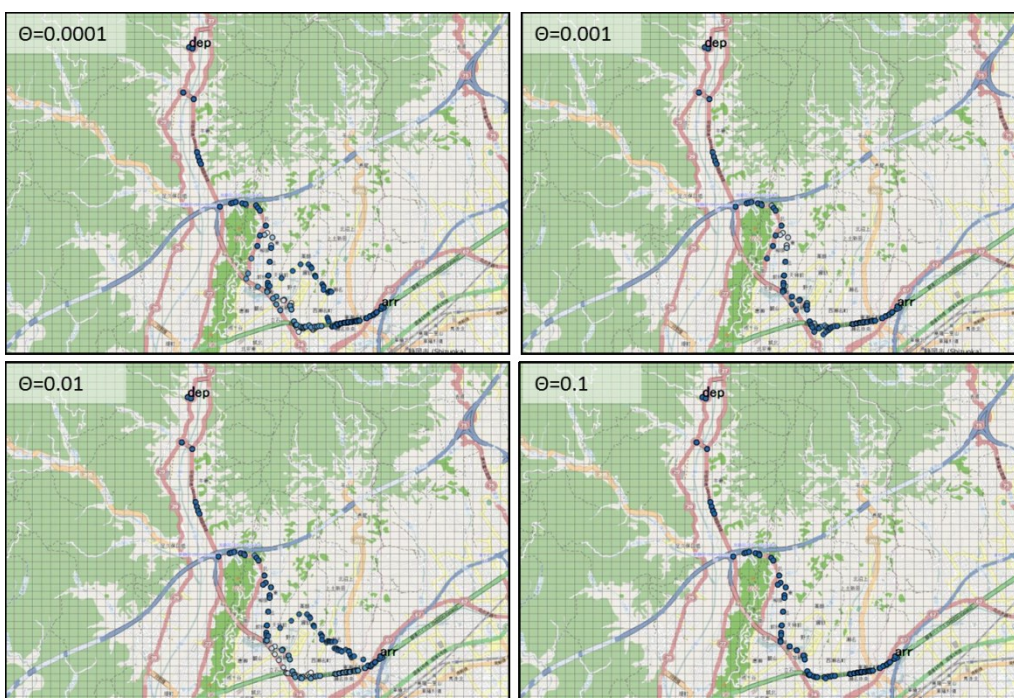


図 4-8 OD 間距離 10km の経路探索結果



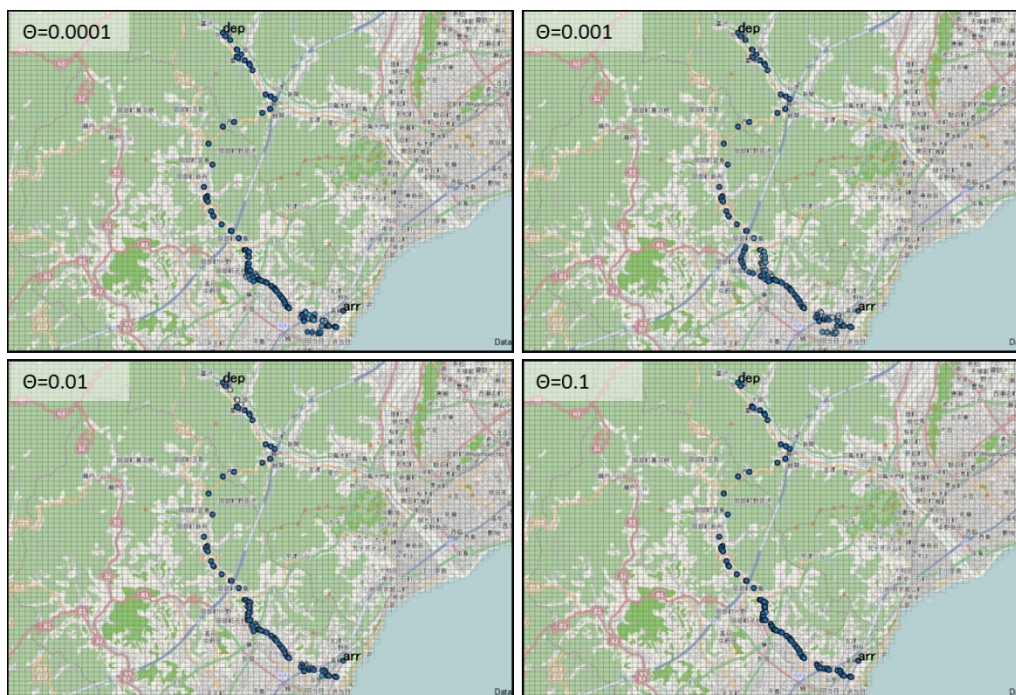


図 4-9 OD 間距離 15km の経路探索結果

なお、経路探索には、1回あたり約1~2秒程度の時間を要する。そのため、本研究においては、トリップ起終点の緯度経度を Geohash 化することで、起終点パターンを減少させ、計算時間の短縮を目指した。この Geohash 化により、道路サブトリップの起終点パターン数は 187,061 件から 116,303 に減少し、1道路あたりの探索時間を2秒とするならば、計算時間は2日弱の短縮したことになる。

Geohash とは、緯度経度に基づくジオコーディング方法の1つで、位置情報をハッシュ値で表すことができる。階層的な空間データ構造をしており、空間をメッシュ状に分割していくことで位置を表現できる。そのため、任意の精度で表現することができ、精度が高いほど、その位置を示す文字列も長くなるという特徴がある。Geohash はこのようなデータ構造をしているがゆえに、緯度経度表現とは異なり、位置の絞り込みが簡単にできる、隣接したエリア情報の取得が容易などの利点がある。しかし、最も大きな利点は、SQL などのデータベースとの親和性が非常に高いために、高速な処理が可能となることであろう。Geohash を用いることで、データベースの問い合わせ時間が大きく減少することは、小川らによって実証されている。図4-10は、例として Geohash 文字列”xn76gg”を表示したものである。データベースで、”xn76gg”を前方一致で検索することで、このようなメッシュ状の表現方法で位置を取り出すことができ、この文字列が長くなるほど、位置をピンポイントで表現できるようになる。なお、本研究では、Precision=7(WxH≒120m×150m)の Geohash を用いた。



図 4-10 Geohash”xn76gg”の表示例  
(緯度経度を文字列で表す GeoHash より引用)

さらに、一度探索した経路はデータベースにキャッシュすることで、再度同一の起終点パターンが現れた場合には、データベースから経路情報を取り出すだけの処理となり、従前に比べて高速な処理が可能となった。なお、データベースにキャッシュした項目については、表 4-11 のとおりである。

表 4-11 データベースにキャッシュした項目

項目名	内容	データ定義
gid	固有のシリアルキー	/
logic	経路探索ロジック	varchar(50) not null
org	起点ハッシュ値	varchar(10) not null
dst	終点ハッシュ値	varchar(10) not null
roads	道路種別	varchar(30)
trans	交通手段	int4
ratio	Penalty 法のコスト増分	varchar(5)
sub	補間点フラグ	int4
geom	経路の点列	geometry('LINESTRING',4326)

しかし、Geohash を用いた経路探索を行うことで、図 4-12 のような不自然な経路をとることも考えられる。(a) は、図の Geohash OD パターンにおいて、一度目に経路探索した例である。一度目に探索が行われた場合、OD 間の経路情報は点列としてデータベースにキャッシュされ、二度目以降にその経路情報が取り出されることになる。(b)

は二度目に経路探索を行った例で、同一の Geohash 内の異なる位置間で経路探索されており、その経路は一度目の経路に線形補間で接続する形で構成されていることがわかる。Geohash を用いることにより、二度目以降に探索された経路は、このような不自然な形状をとる可能性があることに留意する必要がある。

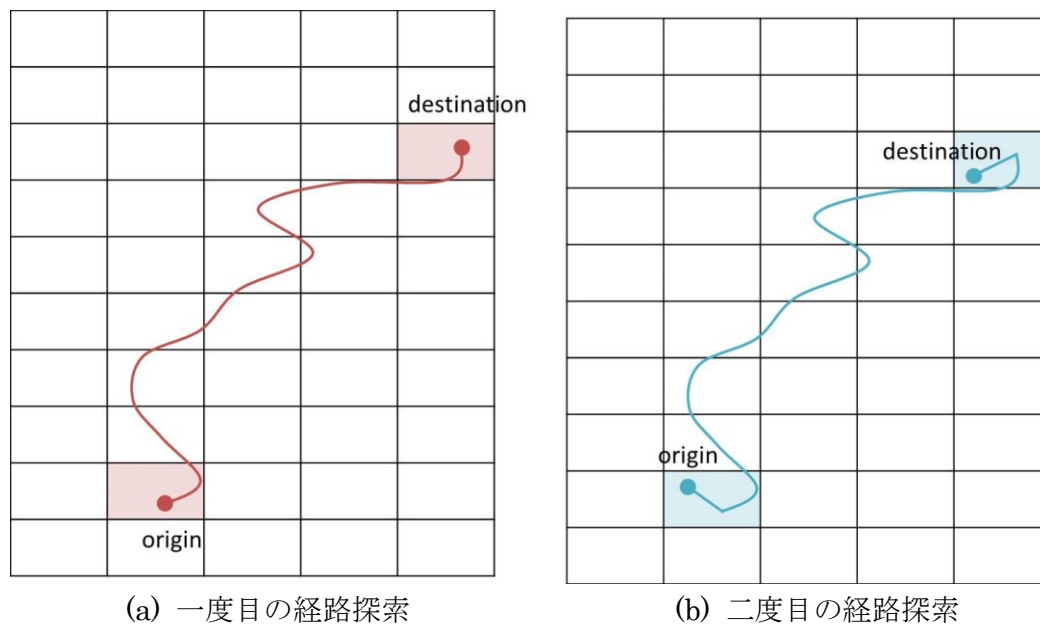


図 4-12 Geohash による経路探索の弊害

最後に、シミュレーションモデル構築の処理フローを図 4-13 に示す。これまで説明した手法により生成された経路は、構成する全ノードに時間情報が付加され、1つのパーティクルとなる。本研究では、それぞれのサブトリップ OD 間について 5 経路を生成させているので、1OD に対して 5 パーティクルが生成されることになる。なお、OD の緯度経度が一致した場合には、滞在トリップとして扱い、1 パーティクルとなるよう処理している。また、パーティクルには、重みの初期値として  $1.0 / \text{経路数}$  を付加しており、5 経路が生成された OD の場合には、それぞれのパーティクルに 0.2 の重み初期値が与えられる。次節のデータ同化では、本節で生成されたパーティクルがフィルタリングにより最適化され、尤もらしい経路候補が選ばれることになる。

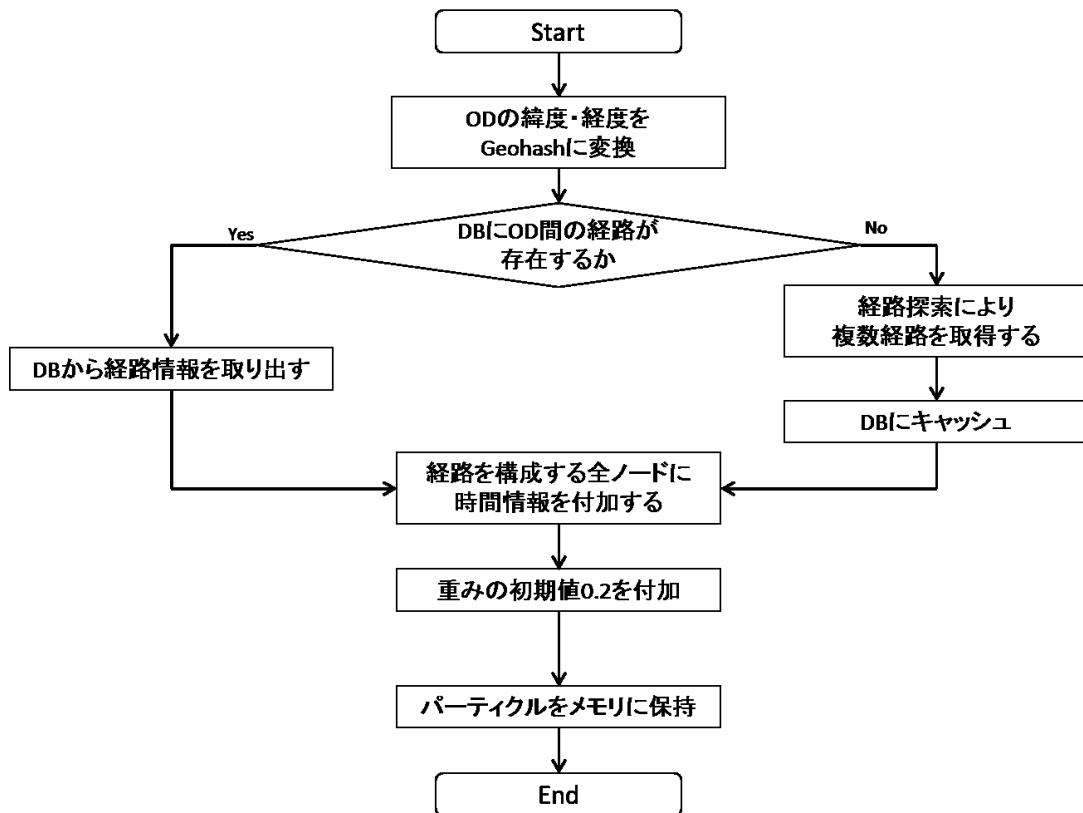


図 4-13 シミュレーションモデル構築の処理フロー

### 4.3 データ同化によるモデルの改善

本節では、人の流れの推定モデルに対して、流動人口データを観測値として同化することで、パーティクルの重み付けを行い、最終的にはモデルの改善を図る。観測値は、0~1 時、1~2 時、…、23~0 時といったように、1 時間単位のメッシュ交通量の形で表現されている。そのため、推定開始時刻から終了時刻まで、1 時間のタイムステップごとに該当する時間帯の観測値を取得しながら、パーティクルの重みを更新していく。推定終了時刻まで、パーティクルの重み更新を終えたら、各トリップの持つパーティクルの重み合計値が 1.0 となるように標準化する。ここまでがデータ同化の一通りの処理であり、全てのパーティクルの重みが収束するまで、この処理を繰り返す。例えば、0~7 までの 7 時間で同化を行う場合、同化のタイムステップは、0~1 時、1~2 時、…、6~7 時の計 7 ステップであり、7 ステップの観測値同化が終了した後に、トリップごとに重みの標準化が行われ、全てのパーティクルについて収束が確認されれば、処理を終了するという流れである。

パーティクルの重みの更新は以下の式に従って行われる。ここで、 $w_s(i)$  はメッシュ  $s$

で  $i$  番目に観測されたパーティクルの重み、 $n_s$  はメッシュ  $s$  で観測されたパーティクルの総数、 $o_s(t)$  は時刻  $t$  におけるメッシュ  $s$  の観測値である。この重み更新では、パーティクルの重みを合計した推定値と観測値が比較され、観測値の方が過大であった場合には、そのメッシュを通るパーティクルは高く評価され、過少であった場合には低く評価されるようになっている。なお、ここでいう推定値は、観測値の取得後にパーティクルを集計したもので、各パーティクルの重み×拡大係数を通過人数として計算している。ただし、同一の PID による異なるトリップが同一メッシュで 2 つ以上観測された場合には、その平均値を通過人数としてカウントする。

$$\hat{w}_s(i) = \frac{w_s(i) \times o_s(t)}{\sum_{j=1}^{n_s} w_s(j)} \quad (1)$$

図 4-14 は、パーティクルが複数のメッシュを通過する場合に、重み  $w(i)$  の更新される様子を示したものである。このパーティクルの例では、ある 1 時間に、メッシュ  $g$ 、 $d$ 、 $e$ 、 $f$  を通過したため、各メッシュで計 4 回の重み更新がされることになる。その計算は図のように行われ、各メッシュで更新された値の平均値が最終的なパーティクルの重みとなる。

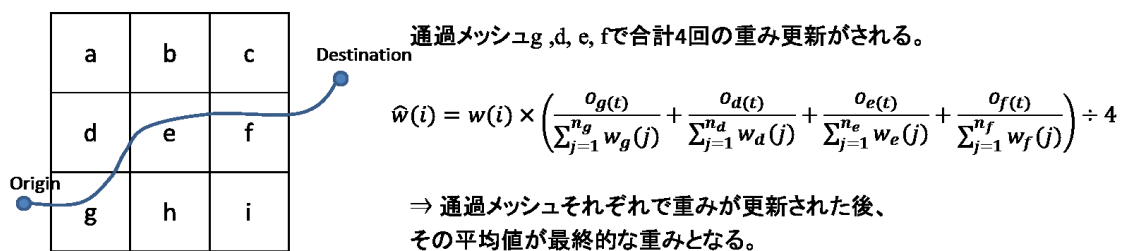


図 4-14 あるパーティクルの重み更新例

また、重みの収束判定は以下の式に従って行った。 $w_n$  は重み更新を  $n$  回施行した後の重みで、その重みは  $n-1$  回後のものと比較され、その差分が 1% 未満なら、重みは収束したと判定される。全パーティクルについて、重みが収束したと判定されると、処理が終了する。

$$\left| \frac{w_n - w_{n-1}}{w_n} \right| < 0.01 \quad (2)$$

全体の処理フローをまとめると、図 4-15 のようになる。

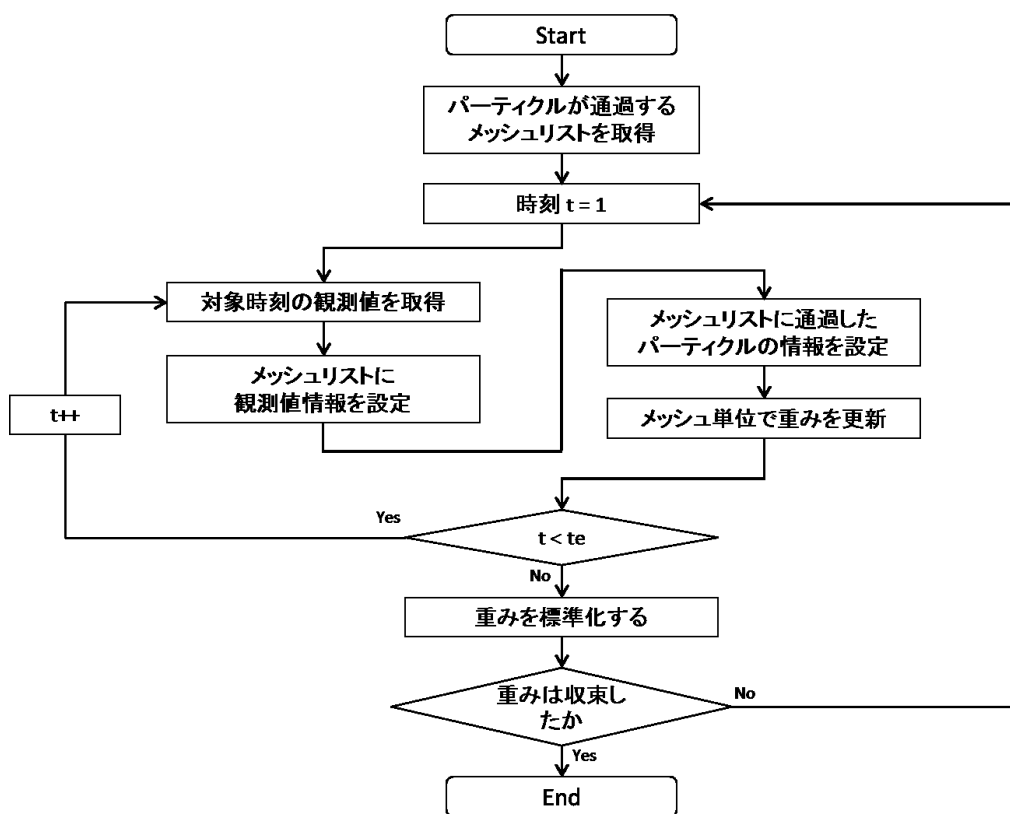


図 4-15 データ同化の処理フロー

#### 4.4 推定結果の評価

推定結果を評価する手法として、メッシュ交通量を採用した。メッシュ交通量とは、推定された人の流れから 1 時間あたりに各メッシュを通過する人数を集計したもので、真値と比較することにより、結果の精度をある程度評価することができる。ここでは、時間帯ごとのメッシュ交通量の相関関係、およびその平均二乗誤差 (RMSE: Root Mean Squared Error) により評価を行う。なお、評価対象には、観測値として用いた流動人口データを用いることとする。



## 5 推定結果の評価

### 5.1 計算結果について

本研究では、Dial 法により PT データのトリップ経路候補を生成させ、人の流れのシミュレーションモデルを構築し、それをパーティクルとしてフィルタリングすることで、モデルの最適化を図った。図 5-1 は、0~24 時までの 24 タイムステップにおいて同化した時の、パーティクルの重み差分の収束状況を示したものである。y 軸は 1 パーティクルあたりの重み差分を、x 軸は収束するまでの計算回数を示しており、計算開始直後において急激な上昇をしたのち、徐々に収束していることが確認できる。ここで、1 パーティクルあたりの重み差分とは、収束を評価したパーティクルの重み差分の平均値である。ただし、重みの収束条件を満たさないパーティクルを見つけた段階で、評価を打ち切っているため、場合により図のような重み差分の急上昇がみられることがある。なお、表 5-2 は各段階における計算時間を示しており、DB へのキャッシュや Geohash を用いたことにより、当初の数日単位から、大きく時間の短縮に成功していることがわかる。

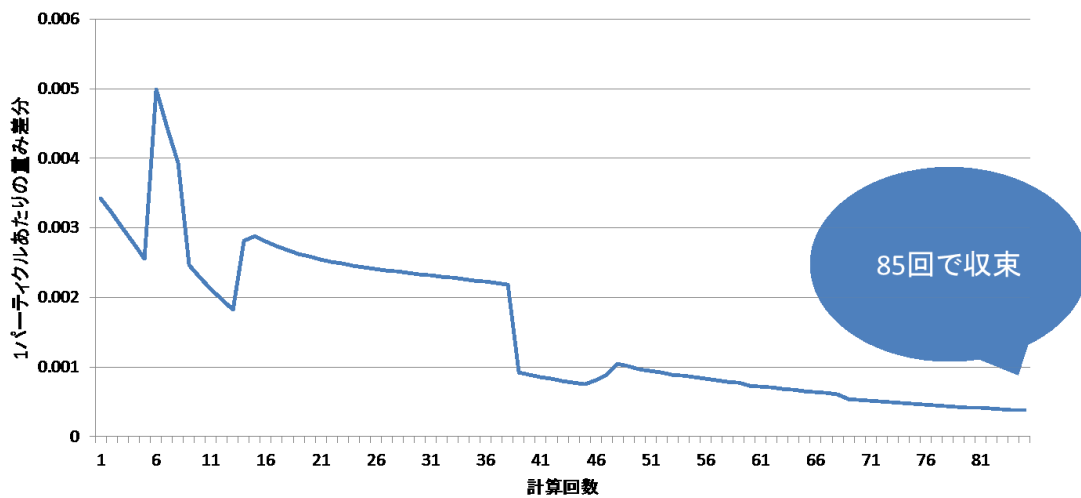


図 5-1 重み収束状況（24 時間同化、Dial）

表 5-2 各段階の計算時間

	パーティクルの生成	観測値の同化
計算時間	21.6min	20.7hour

## 5.2 メッシュ交通量の真値との比較

次に、流動人口データを真値として、1km、500m、250m のそれぞれのメッシュにおいて、メッシュ交通量を集計して、比較を行った。図 5-3、5-4、5-5 は、1~2 時、7~8 時、13~14 時、17~18 時の各時間帯の結果で、x 軸が流動人口の真値を、y 軸が本手法の結果をそれぞれ示している。まず、図 5-3 についてみると、1km メッシュでは概ね良好な結果を得ることができ、特に 13~14 時の昼間帯では、かなり高い相関が確認できた。しかし、1~2 時の夜間帯では、他の時間帯と比較して、精度が落ちているのがわかる。また、500m、250m についても同様の傾向を見ることができるが、250m ではあまり良い結果を得ることはできなかった。なお、各パターンにおける近似直線の傾きを表 5-6 に示す。最も相関の良い 1km メッシュでみると、総じて推定値が過小評価気味であり、通勤・通学や帰宅の人が多く移動すると考えられる時間帯になると、その傾きが 1 に近づく傾向が見てとれる。これは、人が多く移動した結果として、複数のメッシュでカウントされるようになったためであると考えられる。

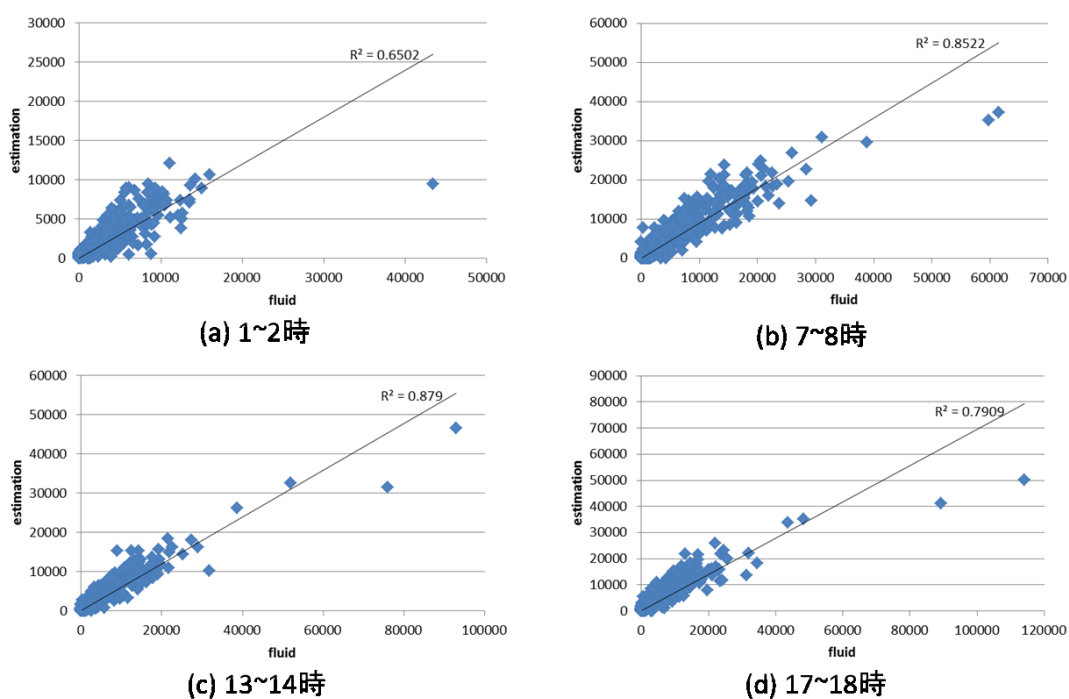


図 5-3 1km メッシュ交通量の比較

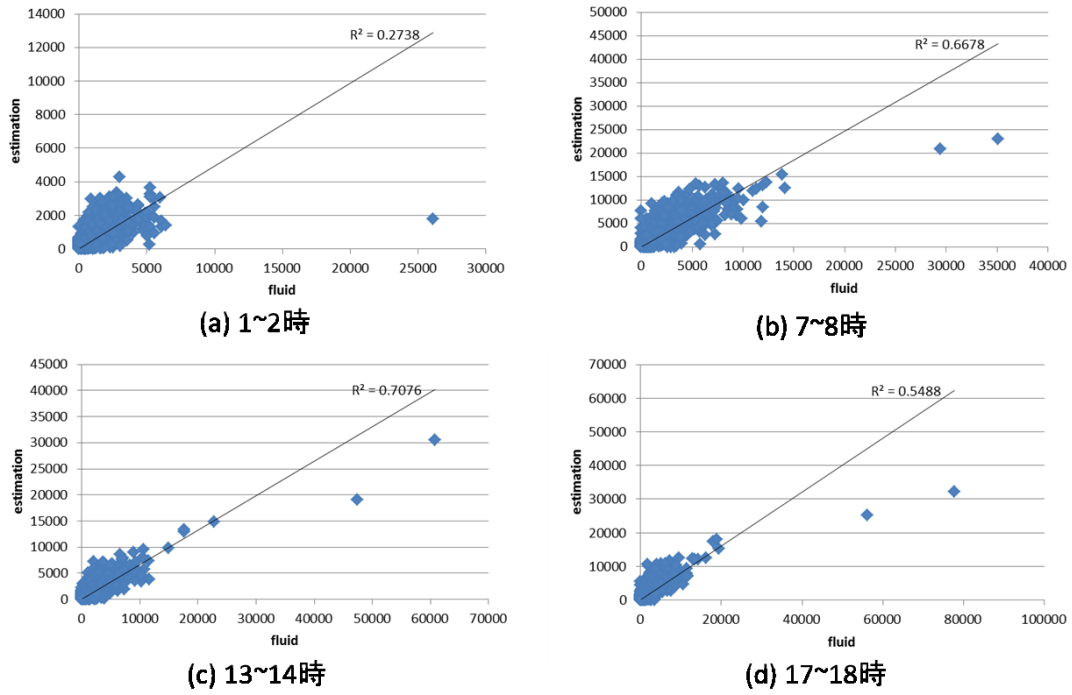


図 5-4 500m メッシュ交通量の比較

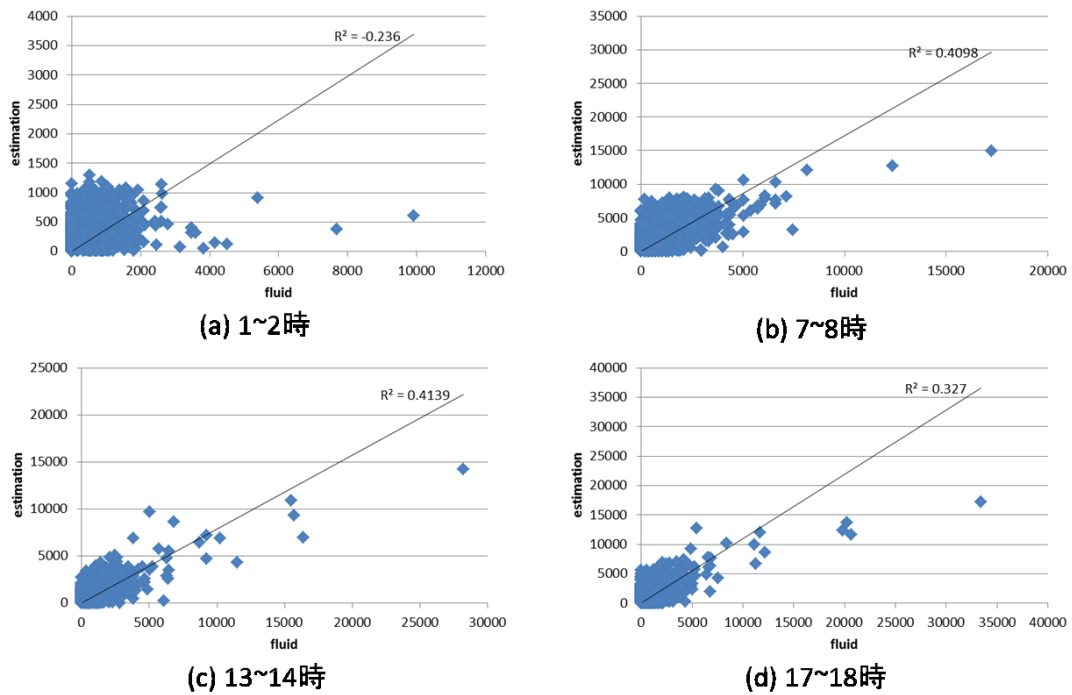


図 5-5 250m メッシュ交通量の比較

表 5-6 各メッシュサイズにおける近似直線の傾き

	1~2 時	6~7 時	13~14 時	17~18 時
1km	0.5985	0.8956	0.5965	0.6943
500m	0.4939	1.2334	0.6608	0.8018
250m	0.373	1.7213	0.7851	1.0965

### 5.3 RMSE による評価

RMSE (Root Mean Squared Error) とは、推定結果と真値の間の誤差を意味しており、値が小さいほど精度高い推定結果であるということになる。RMSE は以下の式に従って、計算することができる。メッシュ交通量の RMSE を示したものが図 5-7 である。y 軸は RMSE、x 軸は 1 時間ごとのタイムステップである。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - x_i)^2} \quad (1)$$

ここで、 $n$  は評価対象のメッシュ数を、 $X_i$ 、 $x_i$  はそれぞれ真値、推定値である。

まず、通勤・通学の時間帯、および帰宅の時間帯において、他の時間帯よりも誤差が大きくなっていることを確認できる。これらの時間帯では、絶対的な移動の量が大きく、その行動も多様であると考えられる。本手法では、PT 調査を基にして経路候補を 5 ずつ作ったが、これだけでは実際の多様性を完全には表すことができず、誤差が大きくなってしまったと考えられる。また、人の移動があまりないであろう深夜帯の誤差が小さいことから、この考察が妥当であるといえる。

次に、メッシュサイズによって、曲線の形状に差があることがわかる。1km メッシュでは、帰宅時間帯である 18 時に誤差のピークがあるが、500m、250m とメッシュサイズが小さくなるにつれて、そのピークが通勤・通学時間帯である 7 時に移行している。一般的に、朝であれば勤務地や学校などの目的地にまっすぐ向かうと考えられるが、夕方から夜にかけては、まっすぐ家に帰るとは限らず、飲食店や娯楽施設などを経由して帰宅すると考えられる。この両者の時間帯における行動パターンの違いが影響して、このような結果になったと推察される。

これらのことから、通勤・通学時間帯、および帰宅時間帯では、その移動が多様であり、トリップのもつパーティクルを増やすことで、精度向上していく必要がある。これについては、今後の検討課題としたい。

最後に、前節では 1~2 時の相関が最も悪かったにもかかわらず、RMSE ではそうではなかった。これは、1~2 時のメッシュ交通量の絶対量が少ないため、決定係数と RMSE の算出方法を比較することにより説明できる。

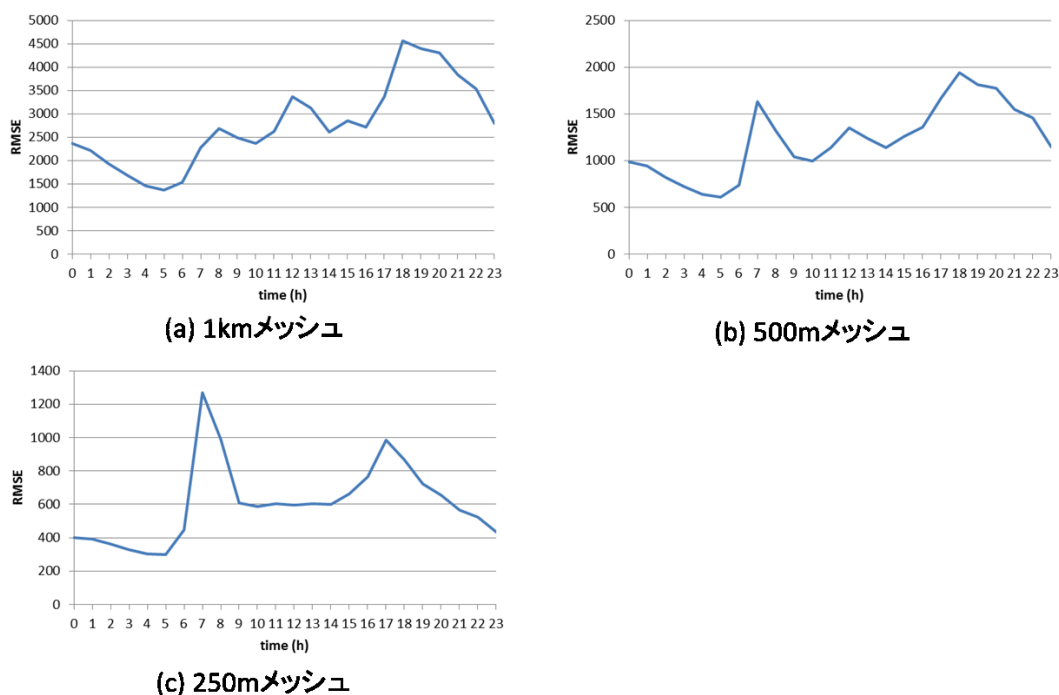


図 5-7 メッシュ交通量の RMSE

## 5.4 メッシュ交通量の可視化

メッシュ交通量の算出にあわせて、その可視化を行った。図 5-8、5-9、5-10、5-11 は、それぞれ 1~2 時、7~8 時、13~14 時、17~18 時の、500m メッシュにおける時間交通量を示したものである。なお、可視化には ArcGIS を用いた。これらの図は、メッシュ交通量を 10 段階に分類したもので、その幅は推定結果のものにあわせている。色が赤に近いほどメッシュ交通量が多いということであり、静岡都市圏の中心部に人が多く集まっていることが確認できる。真値と比較すると、概ねあっているようにも見えるが、深夜帯を除いて、都市の中心部に過剰に人が集中していることがみてとれる。特に、国道 1 動線などの主要道路や含むメッシュにおいて、交通量の過大評価がみられる（図 5-12）。これは、経路選択において、主要道路の尤度が高くなりすぎているということであり、経路探索時のパラメータ設定を再考する必要がある。

なお、深夜以外の時間帯において、中心部から離れたエリアで直線の移動が確認できるが、これは経路探索に失敗したパーティクルをとらえたものである。本手法では、経路探索した場合に、その OD 間を時空間線形補間するように処理しており、山中などの道路ネットワークが少ないエリアにおいて、このようなことが起こりうる。今後は、エリアの道路ネットワーク量に応じて階層的に処理できるようなアルゴリズムを構築し、こういった現象をできるだけ避けられるようにしていきたい。

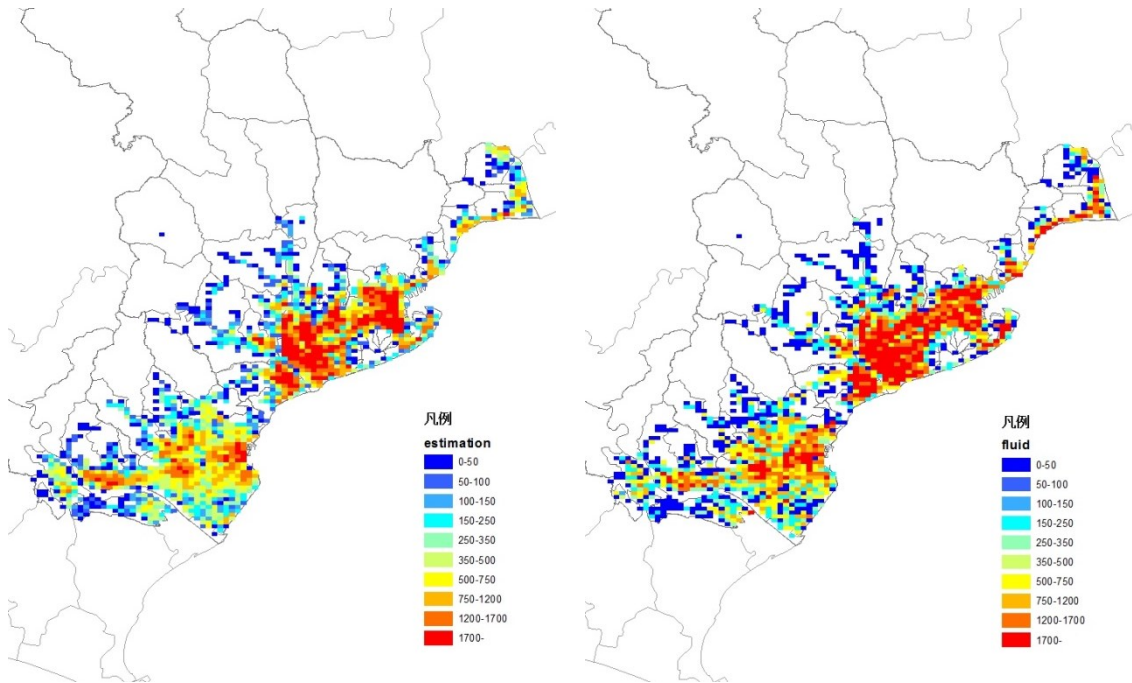


図 5-8 1~2 時の時間メッシュ交通量 (500m メッシュ)

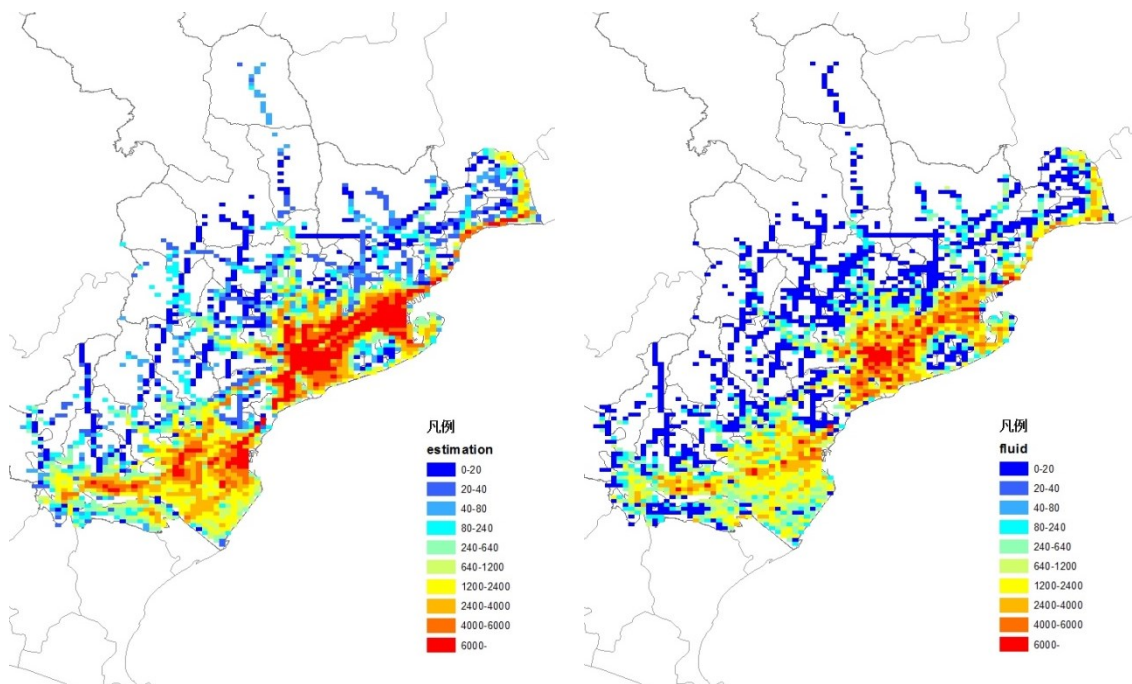


図 5-9 7~8 時の時間メッシュ交通量 (500m メッシュ)



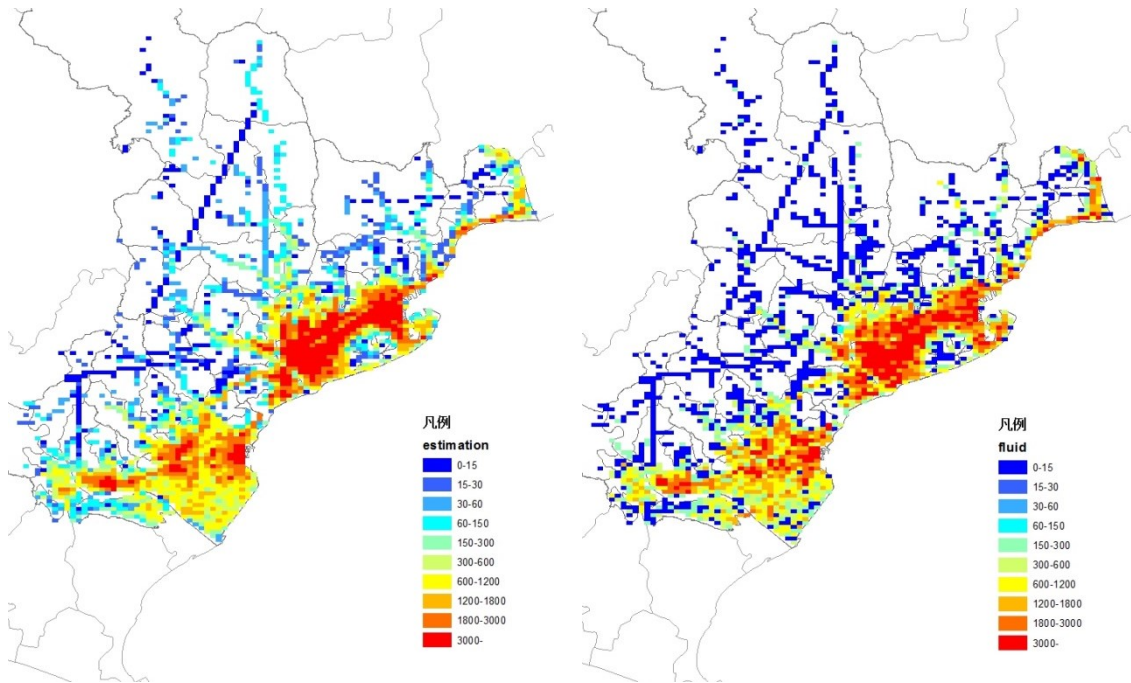


図 5-10 13~14 時の時間メッシュ交通量 (500m メッシュ)

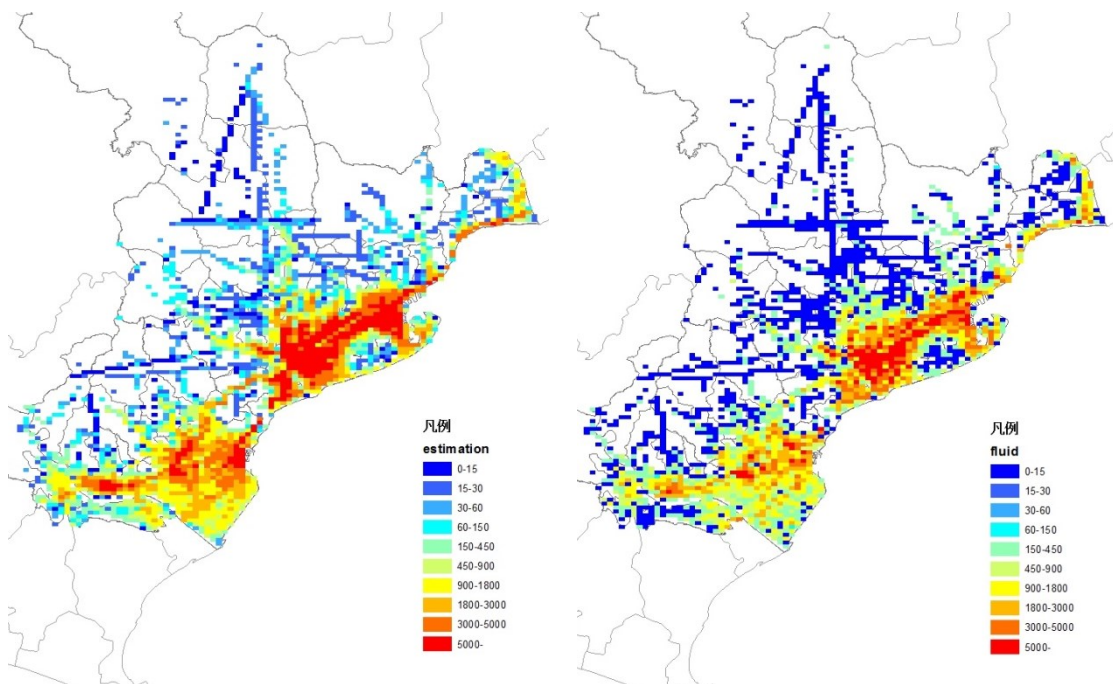


図 5-11 17~18 時の時間メッシュ交通量 (500m メッシュ)

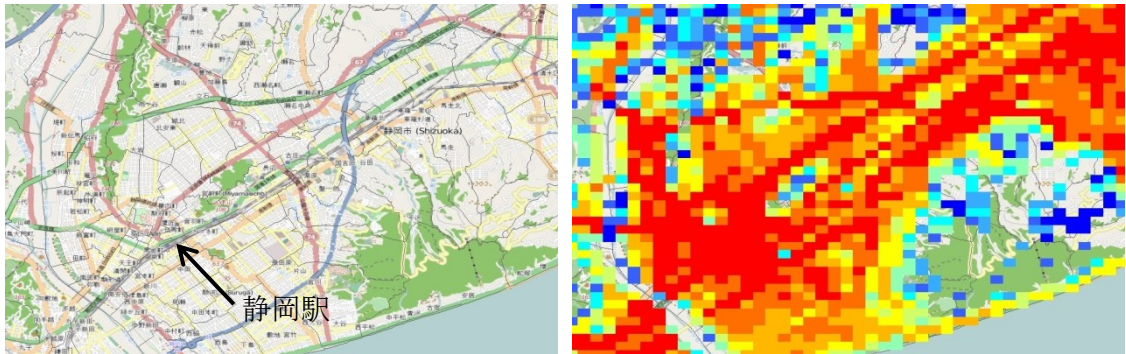


図 5-12 JR 静岡駅周辺のメッシュ交通量 (17~18 時、500m メッシュ)

### 5.5 クロスバリデーションによる手法の評価

クロスバリデーション (Cross Validation) とは、統計学においてよく用いられる、モデルの汎用性を調べることができる手法である。具体的には、標本データを  $k$  個のセットに分割し、 $k-1$  個のセット (トレーニングデータ) で解析を行い、残った 1 個のセット (テストデータ) を用いて、解析結果の評価を行う。この操作を全ての標本に対して、つまり  $k$  回繰り返して行うことで、解析手法の妥当性を確かめることができる。図 5-13 はそのイメージであり、 $k$  回実行後にそれぞれの精度の平均をとることで、推定精度を算出することができる。

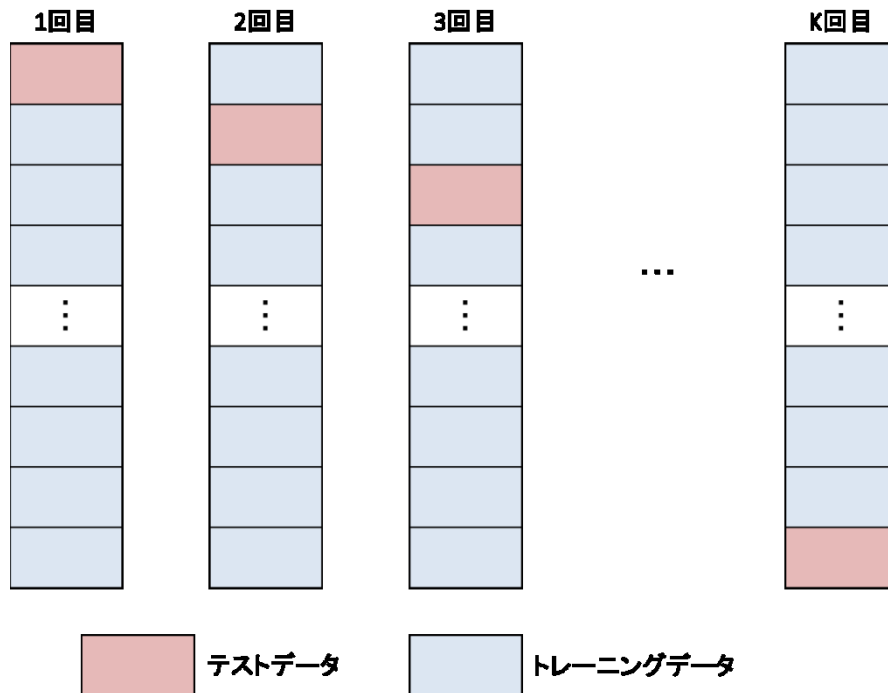


図 5-13 クロスバリデーションのイメージ



本研究においては、データ同化において評価するメッシュリストを3分割し、2個のセットをトレーニングデータ、残り1個のセットをテストデータとすることで、クロスバリデーションを行った。各セットの構成は表 5-14 に示すとおりである。なお、それぞれのセットは、オリジナルセットからランダムサンプリングにより分割したため、地理的にも偏りなく分散している（図 5-15）。

表 5-14 各セットの構成

セット番号	トレーニングデータ	テストデータ
set1	cv_mesh_0-1.csv	cv_mesh_2.csv
set2	cv_mesh_1-2.csv	cv_mesh_0.csv
set3	cv_mesh_2-0.csv	cv_mesh_1.csv

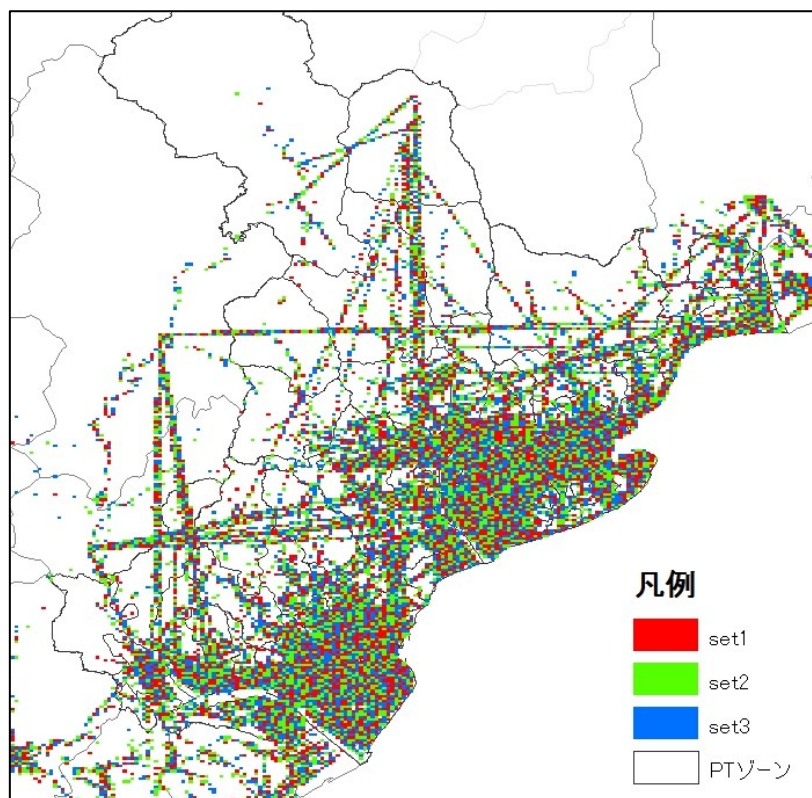


図 5-15 各セットのテストデータの位置

クロスバリデーションによる、250m メッシュ交通量（13~14 時）を真値と比較した結果を図 5-16 に示す。set1~3 はそれぞれのセットにおける結果、origin は分割前の結果である（origin は図 5-5 (c)と同等のもの）。origin が最も良い相関を示しているが、set2 を除けば、origin とあまり有意な差は見られない。一方、図 5-17 はメッシュ交通

量の RMSE を比較したものである。各セットで見ると、**origin** を中心に値が振れており、誤差が大きいのに見えるが、その平均値と比較すると、**origin** にほぼ一致していることがわかる。このことから、本研究の手法が有効的であり、頑健性の高いものであることが示された。

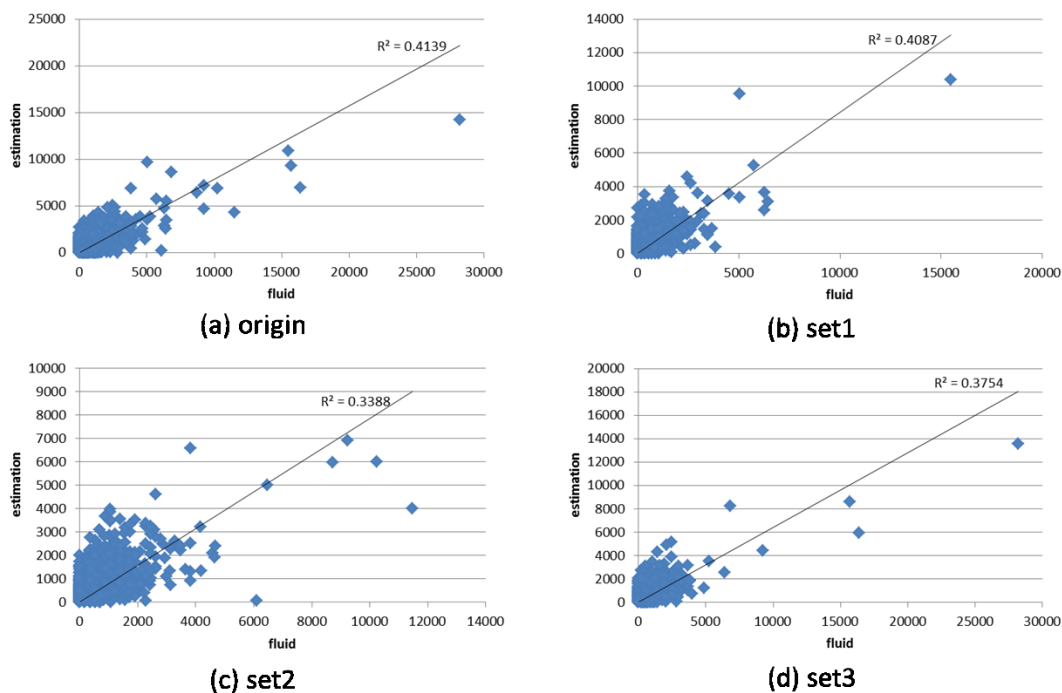
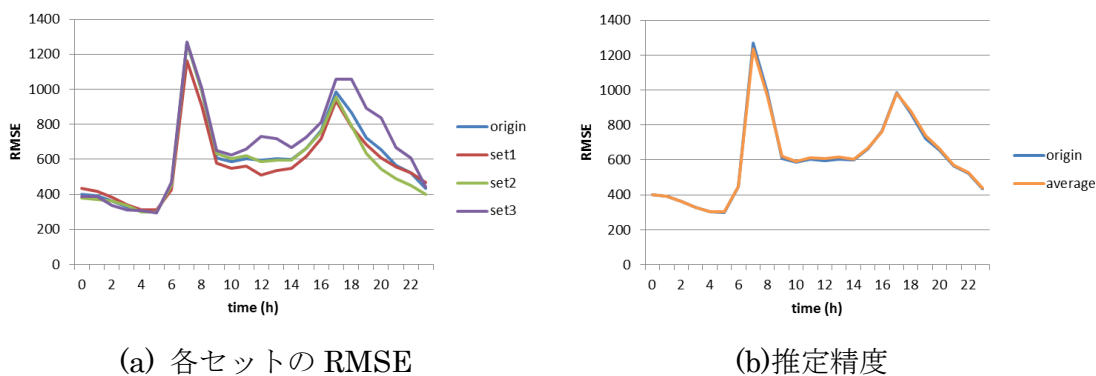


図 5-16 クロスバリデーションによるメッシュ交通量と真値の比較 (13~14 時)



(a) 各セットの RMSE

(b) 推定精度

図 5-17 クロスバリデーションによる RMSE の比較

## 6 NL モデルの適用によるパーティクル生成手法の改善

### 6.1 NL モデルについて

ロジットモデルとは、非集計モデルの中でも最も広く用いられる方法で、効用最大化理論（「個人の意思決定者が、選択肢の中から、合理的選択により、個人にとって最も望ましい選択をする」）に基づいて成り立っている。個人ごとに行動を記述することができる点や、実際の観測データを利用してパラメータを予測できる点などに長所を有し、理解しやすく、パラメータの推定も比較的容易であるため、交通分野をはじめとして様々な分野で用いられている。ロジットモデルを改良したものはいくつも提案されているが、本研究では NL モデル（Nested Logit model）を用いた。NL モデルは、IIA（Independence from Irrelevant Alternatives）特性を緩和したモデルの中でも、最も良く用いられる方法で、選択肢はいくつかの集合に分割し、集合内の選択肢の確率項間の相関を表現することが可能であり、集合内でさらに部分的な集合を設定することも可能である。IIA 特性とは、ロジットモデルにおける大きな問題とされ、赤バス・青バスを例に説明されることが多い。例えば、交通手段の選択肢としてバスト自動車の 2 つがあるとすると、その選択確率は  $1/2$  ずつになる。しかし、バスの半数を赤く、もう半数を青く塗って区別した場合に、その選択確率が  $1/3$  ずつであるとすると、バスのシェアが  $1/2$  から  $2/3$  に上昇したことになり、非現実的な結果となってしまう。このような各選択肢の確率効用項の間に相関がある場合には、IIA 特性を厳密に保つことは難しく、効用項を考慮したモデルを適用する必要がある。特に、交通行動の場合には類似した選択肢が含まれることが多く、NL モデルはこの IIA 特性を回避することができるため、用いられることが多い。

### 6.2 本研究で用いた NL モデルの概要

本研究では、人の交通行動選択が図 6-1 のような 2 レベルのツリー構造で説明できると仮定した。交通行動の目的の選択肢は、通勤・通学、帰宅、その他の 3 種類とし、その目的地の選択肢は、パーソントリップ調査データからランダムに 20 のゾーンを選出した。なお、元の目的地ゾーンは必ず含むものとする。

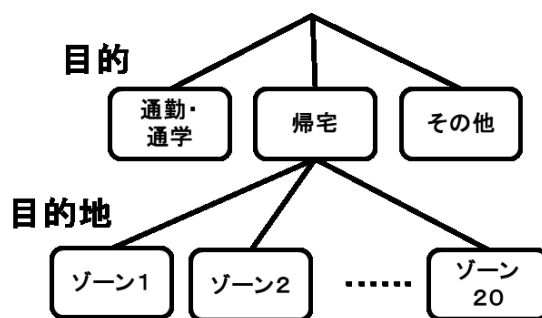


図 6-1 本絵研究の交通行動選択のモデル構造

目的  $i$ 、目的地  $s$  を選択する確率  $P(i,s)$  は以下の式により表せる。後半部は、目的  $i$  を選ぶ確率であり、 $V_i$  は目的  $i$  を選んだ時に得られる効用、 $S_i$  はログサム変数と呼ばれる目的  $i$  を選んだ状況下で、各目的地を選んだ場合に得られる効用の和、 $\mu_i$ 、 $\mu$  はそれぞれ目的地選択、目的選択におけるスケールパラメータである。一方、前半部分は目的  $i$  を選んだ時に目的地  $s$  を選ぶ確率であり、 $V_{s|i}$  は目的  $i$  を選び、かつ目的地  $s$  を選んだときに得られる効用を意味している。この両者の積により、目的  $i$ 、目的地  $s$  を選択する確率  $P(i,s)$  が求まる。

$$P(i,s) = \frac{\exp(\mu_i V_{s|i})}{\sum_j \exp(\mu_i V_{j|i})} \cdot \frac{\exp(\mu V_i + \mu/\mu_i S_i)}{\sum_j \exp(\mu V_j + \mu/\mu_j S_j)} \quad (1)$$

ログサム変数  $S_i$  は以下の式で表される。

$$S_i = \ln \sum_j \exp(\mu_i V_{j|i}) \quad (2)$$

スケールパラメータは、各選択肢集合における誤差の大きさを表現したもので、ツリー構造の上位レベルほど誤差が大きくなる。ここでは、目的地選択がレベル 1、目的選択がレベル 2 なので、以下の式を満たす必要がある。

$$\mu < \mu_i \quad (3)$$

効用項の算出に用いた説明変数は表 6-2 のとおりである。

表 6-2 効用項の算出に用いた説明変数

目的選択	出発時刻ダミー変数 (2-9 時/9-17 時/17-26 時)
目的地選択	目的地の居住者人口/1000
	目的地の通勤者人口/1000
	目的地までの距離の逆数*5000

効用項 $V_i$ 、 $V_{S|i}$ は、以下の式により表すことができる。

$$V_i = \kappa_1 I_{commute} I_{morning} + \kappa_2 I_{others} I_{noon} + \kappa_3 I_{home} I_{evening} + \kappa_4 I_{commute} + \kappa_5 I_{home} \quad (4)$$

$$V_{S|i} = \beta_1 I_{home} p_{home} + \beta_2 I_{commute} p_{commute} + \beta_3 I_{home} (1/dis) + \beta_4 I_{commute} (1/dis) + \beta_5 I_{others} \left(\frac{1}{dis}\right) \quad (5)$$

ここで、 $\kappa$ 、 $\beta$ はパラメータ、 $p_{home}$ 、 $p_{commute}$ はそれぞれ居住者人口、通勤者人口、 $dis$ は目的地までの距離である。また、 $I_{commute}$ 、 $I_{home}$ 、 $I_{others}$ は、それぞれ通勤、帰宅、その他の目的を選んだ場合に 1 となるようなダミー変数であり、 $I_{morning}$ 、 $I_{noon}$ 、 $I_{evening}$ はそれぞれ朝 (2-9 時)、昼 (9-17 時)、夜 (17-26 時)に出発した場合に 1 となるようなダミー変数である。

効用項では、定数項を除いて、基本的には全ての値が正となることを期待している。なぜなら、目的選択においては、朝は通勤、昼はその他の目的、夜は帰宅が多いと考えられるからである。同様に、目的地選択においても、帰宅目的であれば人口の多い方へ行く可能性が高く、通勤目的であれば通勤者が多い方へ行く可能性が高い、また距離に関して言えば、短いほうがより効用が高いと考えることができる。なお、 $V_i$ には定数項が存在するが、これは時刻のダミー変数が 0 になったときに効用項が 0 になることを避けるためである。

### 6.3 最尤法によるパラメータの推定結果

PT 調査データから抜き出した 64985 人、132136 トリップのデータを用いて、最尤推定法によりパラメータの推定を行った。その推定結果を表 6-3 に示す。目的選択パラメータの出発時刻ダミー変数（17-26 時）に関する符号をみると、マイナスになっているにもかかわらず、パラメータは有意であるという結果が得られた。つまり、17 時~26 時に出発したトリップは帰宅目的の方がより効用が下がるということである。これは、朝の時間帯に比べて、夕方から夜のような時間帯では、人の行動選択が多様であり、その目的が帰宅に限られないという行動実態が影響したと考えられる。

他のパラメータについても有意であり、符号に関して問題はなかった。スケールパラメータについても、上位レベルである目的地選択のスケールパラメータ  $\mu_i$  が目的選択のスケールパラメータ  $\mu$  より大きく、(3)の式が満たされており、モデル構造の妥当性が示されたといえる。

表 6-3 最尤推定法によるパラメータ推定の結果

	パラメータ	推定値	t 検定
目的選択	出発時刻ダミー (2-9)	15.17	***
	出発時刻ダミー (9-17)	9.72	***
	出発時刻ダミー (17-26)	-4.77	***
	定数項 1	-8.98	***
	定数項 2	-9.13	***
目的地選択	目的地通勤者数/1000	0.30	***
	目的地居住者数/1000	0.48	***
	距離の逆数 (通勤目的)	1.25	***
	距離の逆数 (帰宅目的)	0.78	***
	距離の逆数 (その他)	1.12	***
スケールパラメータ	$\mu_1$	1	—
	$\mu_2$	0.53	***
	$\mu_3$	0.58	***
	$\mu$	0.25	***
	サンプル数	132136	
	初期尤度	-541010	
	最終尤度	-418467	
	修正済み尤度比	0.226484	

## 6.4 NL モデルの適用によるパーティクルの生成

NL モデルの適用により、パーティクル生成段階において、経路だけでなく、その目的地も分散させることができるようになった。本節では、モデル適用により得られた目的地とそれを基にしたパーティクルの生成手法について述べる。

これまで、人の移動をサブトリップ単位で扱ってきた。しかし、NL モデルでは、ある目的を持った移動、つまりトリップに対してのみ、モデルを適用することができるため、まず元の PT 調査データをトリップ単位に再編した。モデル適用の結果、PT 調査の元のトリップは、それぞれ 20 の目的地をその選択確率とともに得ることができるようになる。ただし、滞在トリップ (OD に緯度経度が一致)、および同一ゾーン内に起点・終点があるトリップについてはモデル適用の対象外としているため、この限りではない。図 6-4 は、各トリップがもつ 20 の目的地のうち、ある閾値以上の選択確率の目的地の数とそのトリップの出現頻度の関係を示したものである。例えば、選択確率が 0.2 以上の目的地を 1 つもつトリップの数は約 6.5 万となる。また、それぞれの曲線を積分することで、トリップ総数である約 13 万となる。本手法では、目的地の選択確率が 0.1 以上あれば十分であるとして、各トリップがもつ目的地の数はその選択確率が高い順に 3 つにすることとした。なお、各目的地との間の経路数も 3 ずつとしている。つまり、各トリップ間のパーティクル数は、目的地を分散させる場合にはその積である 9 とし、それ以外の場合については 3 とした。

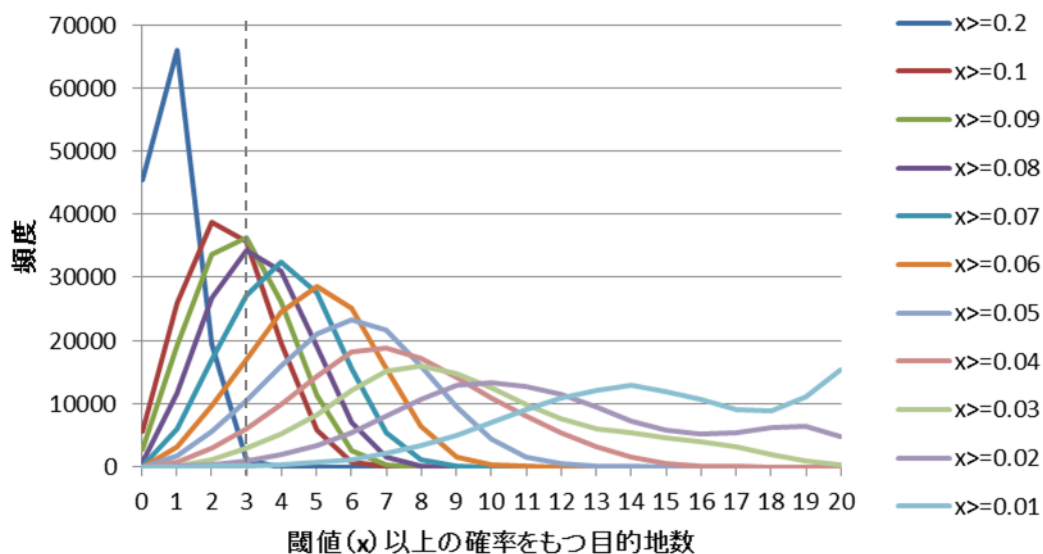


図 6-4 閾値以上の選択確率をもつ目的地数とそのトリップの出現頻度の関係

## 6.5 NL モデルの適用による人の流れの推定結果

前節で、構築したシミュレーションモデルにこれまでの手法を用いて観測値を同化させ、静岡都市圏における人の流れを推定した。ただし、結果の計算が終了しなかったため、以後差し替えるものとする。



## 7 結論

### 7.1 まとめ

本研究では、非集計の位置情報に対するニーズの高まりを背景として、その目的を、比較的利用の容易な人の行動データから疑似的な非集計位置情報データを構築することとした。具体的には、静岡都市圏において、パーソントリップ調査データから人の流れのシミュレーションモデルを構築し、そのシミュレーションモデルに観測値として、GPS のメッシュ集計データを同化することで、モデルを補正し、より現実に近い人の流れデータの構築を試みた。さらに、その推定結果を、メッシュ交通量により評価することで、本手法の精度を検証した。本論文の全体をまとめると以下ようになる。

第1章では、非集計位置情報を取り巻く現在の状況を中心に、そのニーズの高まりについて述べた。一方で、各データが抱える問題点についても触れ、本研究により解決できる可能性があるとして述べた。

第2章では、現在、数多くある人の流れデータについて、東大「人の流れプロジェクト」で扱われているものを中心に、いくつかの事例を紹介した。

第3章では、本研究の人の流れ推定の概要について述べた。本手法がシミュレーションモデルの構築、データ同化、推定結果の評価と3つに分かれていること、それぞれの手法の概要について記した。その中で、データ同化やパーティクルフィルタについて説明するとともに、人の流れデータとして用いられることの多いPT調査について、その問題点を挙げた。

第4章では、人の流れ推定手法の詳細について述べた。使用したパーソントリップ調査と流動人口データの概要とともに、本研究を構成する「パーティクルの生成によるシミュレーションモデルの構築」、「データ同化によるモデルの改善」、「推定結果の評価」の手法について記した。本研究の特出する点としては、パーソントリップ調査データを基にしてDial法により経路候補を生成させたこと、Geohashとデータベースを用いることで大幅な計算時間の短縮に成功したこと、データ同化のフィルタリング手法にパーティクルフィルタを用いたことが挙げられる。

第5章では、本手法による人の流れの推定結果について述べた。まず、本手法の重み更新方法によるパーティクルの収束状況について示し、データ同化がうまく機能したことを確認した。メッシュ交通量の比較では、1km、500mのメッシュにおいて、高い相関を確認することができた。しかし、250mメッシュにおいては、あまり良い結果は得られず、この理由として静岡駅周辺などに存在する主要道路に過剰に人が集中していることを挙げ、本手法で用いたDial法のパラメータ設定を再検討する必要があると記した。また、RMSEによる評価では、通勤・通学時間帯、および帰宅時間帯において誤

差の増大を確認することができ、これは本研究によるシミュレーションモデルが現実の人の移動の多様性を表現するためには、不十分であったと推察した。最後に、クロスバリデーションにより、手法の評価を行ったところ、その推定精度が、本手法の精度にほぼ一致していることを確認することができ、本手法の頑健性を示すことができた。

第6章では、シミュレーションモデルの精度向上案として、目的地選択 NL モデルを適用したパーティクルの生成手法について述べた。まず、一般的な NL モデルの概要について述べ、本研究で用いたモデルの構築方法について詳細に述べた。最尤推定法により、NL モデルのパラメータを推定したところ、優位な結果を得ることができ、妥当なモデル構築することができた。

## 7.2 今後の課題

本研究では、概ね良い結果を得ることはできたが、課題としていくつかの点が明らかとなった。まず、構築したシミュレーションモデルが、現実の状況を表現するものとしては、その多様性が不十分な点である。そのため、本研究で用いたような人の行動モデルを適用することで、さらなる精度の向上につなげたい。次に、経路選択において、主要道路の尤度が高くなりすぎてしまった点が挙げられる。その結果として、静岡駅付近を通る国道や東西にのびるその他の街道に人が集中してしまい、推定結果の精度の低下を招いた。本研究で用いた Dial 法は、パラメータ設定を変えることで、その出力される経路も変化させることができるため、今後は経路の生成についても検討する必要がある。さらに、経路探索について言えば、探索に失敗する例がいくつか確認することができたため、これについても改善する。

最後に、本稿では評価手法として、流動人口データを用いたが、道路交通センサスなどの実際の統計データと比較する必要があるだろう。また、本研究の優位な点は非集計の個人移動データを構築できる点にある。そのため、非集計データどうしの比較による評価は必須であり、今後の重要な課題だといえる。

## 参考文献

東京大学空間情報科学研究センター：人の流れプロジェクト、  
<http://pflow.csis.u-tokyo.ac.jp/index-j.html>

ゼンリンデータコム：混雑統計、  
<http://www.zenrin-datacom.net/business/campaign/s01.html>

ITS Japan：通行実績情報マップ、2011、<http://www.its-jp.org/saigai/>

関本義秀、中村敏和、増田祐介、金杉洋：大規模な GPS 情報をもとにした東京都市圏における震災時の行動分析、土木計画学研究・講演集、Vol.45、No.249、2012

生形嘉良、関本義秀：GPS データによる観光実態の分析、土木計画学研究・講演集、Vol.45、No.370、2012

NTT ドコモ：測位方法、<http://www.nttdocomo.co.jp/service/safety/search/usage/gps/>  
中村和幸、上野玄太、樋口知之：データ同化：その概念と計算アルゴリズム、統計数理、Vol.53、No.2、2005

中村敏和、関本義秀、薄井智貴、柴崎亮介：パーティクルフィルターを用いた都市圏レベルの人の流れの推定手法の構築、土木学会論文集 D3（土木計画学）、2013（掲載決定）

松本修一：GPS 携帯を活用した行動調査に関する基礎的研究、KEIO SFC JOURNAL、Vol.9、No.1、2009.

国土交通省：D 首都圏の交通への影響とその後の取り組み、  
<http://www.mlit.go.jp/sogoseisaku/soukou/chiebukuro/sankou08.pdf>

東京都市圏交通計画協議会：パーソントリップ調査とは、  
<http://www.tokyo-pt.jp/index.html>

東京都市圏交通計画協議会：PT データ利用の手引き、  
<http://www.tokyo-pt.jp/data/file/tebiki.pdf>

国土交通省：PT 調査とは、<http://www.mlit.go.jp/crd/tosiko/pt.html>

福井都市圏パーソントリップ調査：<http://info.pref.fukui.jp/toshi/pt/>

JAXA（宇宙航空研究開発機構）：準天頂衛星 初号機「みちびき」  
[http://www.jaxa.jp/projects/sat/qzss/index\\_j.html](http://www.jaxa.jp/projects/sat/qzss/index_j.html)

準天頂衛星システムサービス：<http://www.qzs.jp/index.html>

名古屋大学森川・山本・三輪研究室：第 2 章 プローブカーデータを用いた交通行動分析の位置づけ、[http://www.trans.civil.nagoya-u.ac.jp/~miwa/doc\\_paper/Chapter2.pdf](http://www.trans.civil.nagoya-u.ac.jp/~miwa/doc_paper/Chapter2.pdf)

観光庁：GPS を利用した観光行動の調査分析、  
<https://www.mlit.go.jp/kankocho/shisaku/kankochi/gps.html>

NTT ドコモ：モバイル空間統計に関する情報、  
[https://www.nttdocomo.co.jp/corporate/disclosure/mobile\\_spatial\\_statistics/](https://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/)

電気通信事業者協会：「携帯電話事業者の運用データ等の適正な有効利用に関する検討会」報告書、<http://www.tca.or.jp/information/ketaiunyodata/hokokusho.pdf>、2010

Carlo Ratti, Stanislav Sobolevsky, Francesco Calabrese mail, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, Steven H. Strogatz: Redrawing the Map of Great Britain from a Network of Human Interactions, PLOS ONE, 5(12), 2010.

Wolfgang LOIBL and Jan PETERS-ANDERS: Mobile Phone Data as Source to Discover Spatial Activity and Motion Patterns, GI\_Forum, 2012.

山形耕一：パーソントリップ調査における調査不能誤差と送別拡大に関する研究、土木学会論文報告集、Vol.343、1984

大瀧保明、胡丹、猪岡光、鈴木明宏：携帯型装置による人間の移動行動の推定、計素行自動制御学会東北支部第 222 回研究集会、2005

北川源四郎；モンテカルロ・フィルタおよび平滑化について、統計数理、Vol.44、No.1、

pp.31-48、1996

樋口知之：粒子フィルタ、電子情報通信学会誌、Vol.88、No.12、pp.989-994、2005

Dial, R.B. : A Probabilistic Multipath Traffic Assignment Algorithm Which Obviates Path Enumeration. Transportation Research 5, pp.83-111, 1971.

小川輝樹、伊藤嘉博、中村嘉隆、高橋修、白石陽：空間統計処理のための Geohash を用いたデータベースシステムの提案と評価、マルチメディア、分散、協調モバイルシンポジウム、2012

モデルの精度を推定する：

<http://musashi.sourceforge.jp/tutorial/mining/xtclassify/accuracy.html>

全国基準地域メッシュデータ（解説）：

[http://d.hatena.ne.jp/murakami\\_tak/20080708/p1](http://d.hatena.ne.jp/murakami_tak/20080708/p1)

原田昇：非集計行動モデルによる多次元選択行動の分析、土木計画学研究・論文集、No.4、1986

山本俊行：離散選択モデルの発展と今後の課題、交通工学、Vol.47、No.2、pp.4-9、2012

## 謝辞

修士論文を書き上げるにあたり、多くの方々から、ご指導・ご協力を頂きました。皆様に心より感謝申し上げます。

柴崎先生には、ご多忙の中、研究の指導をしていただきました。また、二度の海外学会の参加や、京大との合同ゼミ、地理空間に関するシンポジウムなど、多くの貴重な経験をさせて頂きました。心より感謝いたします。

関本先生には、毎週のようにゼミを開いていただき、密にご指導して頂きました。研究で行き詰まることも多かったですが、的確なご指導のおかげで、ここまで来ることができました。心より感謝いたします。また、先生の運営する「人の流れ研究会」では、毎回のように参加させて頂き、関連する研究についていろいろなお話を聞くことができ、非常にためになりました。

金杉さん、榎山さんにも、毎週のゼミに参加していただき、主にプログラムの面で助けて頂きました。プログラミングに関して、ほとんど知識がなかった私ですが、お二人にたくさんのお話を教えていただいたおかげで、多くの知識を身に付けることができました。また、お二人がいなければ、ここまでたどり着けなかったと思います。

秘書の秋枝さん、本間さん、川端さんには、学会参加時などに事務手続きでお世話になりました。提出書類で困っていた時にわかりやすく説明して頂き、非常に助かりました。

研究員の方には、様々な研究分野において、多くの協力を頂きました。柴崎研究室には様々な国籍の方がいらっしゃり、グローバルな環境を体験できたのは非常に良かったです。来年からも頑張ります。また、駒場にいるときには、くだらない雑談にも乗って頂き、非常に楽しかったです。

同期の池田君、仲宗根君、佐山君、Karlvin 君には、公私ともにお世話になりました。クリスマスパーティなどのイベント時には、研究室全体の人数が多く大変でしたが、みんなで協力できたおかげで、非常に楽しい会にすることができ良かったです。来年度から就職ですが、ともに頑張っていきましょう。そして、これからも公私ともに仲良くしていただければと思います。

最後に、6年間の学生生活を支えてくれた家族に最大限の感謝の気持ちを表して、謝辞といたします。

2014年1月  
戸井田 亮祐