

道路政策の質の向上に資する技術研究開発  
成果報告レポート  
No. 28-3

研究テーマ

複数のデータを活用した  
道路のストック効果の計測技術の再構築

研究代表者： 広島大学准教授	塚井 誠人
共同研究者： 熊本大学准教授	円山 琢也
広島大学教授	藤原 章正
金沢大学教授	高山 純一
金沢大学教授	中山 晶一郎
宮崎大学准教授	嶋本 寛
神戸大学准教授	瀬谷 創
筑波大学教授	堤 盛人
東京大学教授	布施 孝志
岡山大学准教授	橋本 成仁
鳥取大学教授	桑野 将司
広島大学准教授	力石 真
東洋大学准教授	加知 康範
復建調査設計	佐藤 啓輔
広島大学研究員	山本 航

令和元年7月

新道路技術会議

# 目次

研究概要.....	1
第1章 はじめに.....	3
1. 1 データとモデル.....	3
1. 2 本研究のテーマ.....	4
1. 3 交通状態の質的指標の算出と事故リスク分析.....	4
1. 4 観光地のトラベルコスト評価.....	6
1. 5 工業団地地価のヘドニック分析.....	7
第2章 交通状態の質的指標の算出.....	8
2. 1 暫定2車線区間を対象とする速度低下検出.....	8
2. 2 データセットの構築.....	8
2. 3 測位精度の検証.....	10
2. 4 分析手順.....	12
2. 5 決定木分析.....	13
2. 6 目的変数と説明変数.....	14
2. 7 速度低下の検定.....	14
2. 8 決定木分析による速度低下区間の特定.....	15
第3章 交通状態の質的指標と事故リスク.....	17
3. 1 交通ビッグデータと事故リスク.....	17
3. 2 分析手法.....	17
3. 3 データの加工手順.....	18
3. 4 事故発生モデルの推定結果.....	21
3. 5 事故台数モデルの推定結果.....	23
第4章 観光地のトラベルコスト評価.....	26
4. 1 対象地域の選定.....	26
4. 2 データの概要.....	26
4. 3 メッシュ別滞在者に関する時系列分析.....	29
4. 4 トラベルコストによる道路ストック効果.....	34

第5章 工業団地地価のヘドニック分析 .....	36
5. 1 研究の狙い .....	36
5. 2 対象地域とデータベースの構築 .....	36
5. 3 分析手法 .....	37
5. 4 ヘドニック分析の結果 .....	38
第6章 統計手法による道路ストック効果の検証 .....	40
6. 1 本テーマの狙い .....	40
6. 2 既往研究と本研究の位置付け .....	40
6. 3 ストック効果の探索手法 .....	41
6. 4 分析対象及び使用データ .....	42
6. 5 因果グラフの推定と考察 .....	43
第7章 おわりに .....	45
7. 1 本課題の成果 .....	45
7. 2 道路のストック効果計測フレームの整理 .....	47
参考文献 .....	50

**「道路政策の質の向上に資する技術研究開発」(平成 28 年度採択)  
研究概要**

番号	研究課題名	研究代表者
No.28-3	複数のデータを活用した道路のストック効果の計測技術の再構築	広島大学准教授 塚井誠人

ETC2.0 データから得られる交通状態の質的指標を用いた新たな事故リスクの分析, 携帯電話ビッグデータを用いた観光地のトラベルコスト評価, ならびに工業団地等の地価分析を実施して, 道路のストック効果算出のための計測技術を再構築する。

**1. 研究の背景・目的 (研究開始当初の背景・動機・目標等)**

道路整備は, “地域間の結びつきの変化” をもたらす。したがって, 事業評価時点 (事前評価時点) の地域間の結びつきや立地分布を前提とする現在のストック効果計測手法には, 方法論的限界がある。本課題では, 新しい理論の社会実装に先立って, ETC/ETC2.0 や, 携帯の位置情報を用いたモバイル空間統計などの新しい交通ビッグデータの利用可能性の検討を行う。ビッグデータを活用する際に, 小標本データ用の解析手法をそのまま適用すると, 自明あるいは冗長な結果しか得られない。そこで本課題では, 多様な交通・土地利用ビッグデータの特徴を活かした計測技術を産官学一体で開発する。

**2. 研究内容 (研究の方法・項目等)**

**テーマ1：交通状態の質的指標の算出と事故リスク分析**

ETC2.0 データを活用した, 道路の新たな質的指標を提案する。具体的には, 高速道路の走行車両から得られる速度と加速度を, それらが観測された地点に紐づけたデータベースを構築する。本研究では, これらを交通状態の質的指標: Quality of Traffic (QOT) と呼ぶ。QOT は, 車線数や道路線形を反映した道路の「走りやすさ指標」として活用できる。事故リスクの算出には, オッズ比モデルを用いる。

**テーマ2：観光地のトラベルコスト評価**

NTT ドコモが提供するモバイル空間統計を用いて, 観光地を訪問する価値は出発地から観光地までのトリップ費用以上となることに着目したトラベルコスト法を適用する。本課題では, 観光地データベースと流動データ, および他の土地利用に対する集中交通量を説明する統計モデルを推定する。

**テーマ3：工業団地地価のヘドニック分析**

インフラの整備効果が最終的に地価に帰着するという資本化仮説に基づいて, 地価のヘドニック分析を行う。データとして, モバイル空間統計を用いる。具体的には, 高速道路アクセスを期待して整備される工業団地・流通業務団地への発着交通量の分布特性を得る。さらにヘドニック分析により, 高規格道路による流動の価値の高い地域を特定して, 地価に及ぼす影響を明らかにする。

**テーマ4：統計手法による道路ストック効果の検証：中間評価の指摘事項に対応して追加**

標準的な経済モデルでは, 関係主体の経済学的な行動指針を踏まえた最適化問題から演繹的にモデル構造を導出しており, 実データに基づくモデル構造の抽出は行っていない。この課題では, 実際に発現した効果がどのような地域経済特性変数に起因して生じているかを, データに基づいて検証する手法を開発する。そのため, 統計的因果推論の枠組みを援用する。

**3. 研究成果 (図表・写真等を活用し分かりやすく記述)**

**テーマ1：交通状態の質的指標の算出と事故リスク分析**

速度分布や加速度の計測結果を交通流の質的指標ととらえると, 交通事故に寄与する統計的に有意な要因が抽出できた。すなわち, 交通事故は純粋に確率的に発生すると考えるのではなく, 交通流を整流することによって, 交通事故を削減できる可能性がある。また, 交通流の状態が交通事故発生を左右することを明らかにできた。したがって, 交通流の状態をモニタリングすることによって, 潜在的に交通事故の危険性が高い地点と交通流況を同定できる可能性がある。これらの地点の道路特性と交通流特性を改善することによって, 交通事故の削減に寄与できると思われる。

## テーマ2：観光地のトラベルコスト評価

本分析から、トラベルコストを算出することによって、観光地の状況を定量的、かつ視覚的にモニタリングできることが明らかとなった。本アプローチは、ストック効果が見込まれる地域の集客実態をトリップの発地と結びつけて検証する上で、有効なことが明らかとなった。

## テーマ3：工業団地地価のヘドニック分析

ヘドニック分析によって、高速道路アクセスに加えて、夜間および週末にピークを持つ取引先が多様なほど、工業団地地価が上昇するという結果が得られた。この成果は、ビッグデータを用いたヘドニック分析として世界的に希少な成果であり、地価に着目したストック効果計測の可能性を示している。

## テーマ4：統計手法による道路ストック効果の検証：中間評価の指摘事項に対応して追加

統計的なストック効果計測手法に関しては、最新の統計的因果推論手法である統計的因果探索手法を、アクセス性の改善が見られた尾道松江線沿線の市町村に適用した。初期仮説として、人口統計社会指標（人口、就業者、事業所数）が経済活動指標（地価と地域総生産）に影響すること、アクセス指標は、これら2グループの指標の片方、または両方に影響を及ぼす要因として現れることを想定した。その結果、アクセス指標は経済活動指標のうち、地価にのみ影響し、地域総生産には直接影響しないことが明らかとなった。本成果は、統計的因果推論のインフラストック効果に対する適用例としては、筆者の知る限り世界初の成果である。

### 4. 主な発表論文（研究代表者はゴシック、研究分担者は下線）

杉原豪，塚井 誠人：社会基盤の統計的因果探索によるストック効果の検証，土木計画学研究・講演集，Vol.59，2019。

塚井 誠人，山本 航，円山 琢也，佐藤 啓介，瀬谷 創，嶋本 寛：統計的手法による交通インフラストック効果の計測：課題と展望，土木計画学研究・講演集，Vol.58，2018。

高山莉那，塚井 誠人，山本 航，山本優樹：ETC2.0 データを用いた交通事故の要因解析，土木計画学研究・講演集，Vol.57，2018。

山本 優樹，塚井 誠人，山本 航，小山田 哲郎：ETC2.0 データを用いた高速道路上の速度低下区間の検出法，土木学会論文集 D3，vol.74-5，I\_693-I\_702，2018

嶋本寛，黒江真樹：インフラのストック効果計測に向けたモバイル空間統計データの特性把握，土木計画学研究・講演集，Vol.55，2017。

宇田 俊亮，瀬谷 創，塚井 誠人，堤 盛人：日本における工業地の分譲価格の分析，土木計画学研究・講演集，Vol.55，2017。

Wataru YAMAMOTO，Makoto TSUKAI：An Analysis of Vehicle Speed Distribution by Using Traffic Counter Big Data，THE 12 INTERNATIONAL CONFERENCE OF EAST，2017。

### 5. 今後の展望（研究成果の活用や発展性、今後の課題等）

交通状態の質的指標と事故リスクに関する研究では、特に地点別に詳細な情報が得られるプローブデータの活用法が開発できた。交通状態の質的指標と事故リスクを関連付けた分析に関しては、現在投稿論文を準備中である。観光地魅力度のトラベルコスト法による評価、ならびに工業団地地価に関するヘドニック分析については、実務的なデータ処理に関する課題が存在することが明らかとなったため、引き続き検討を行う。既存のストック効果評価手法と統計的評価手法の関係に関しては、神戸大学の研究グループ、ならびに地方整備局と協力して、既存手法との位置づけを整理する。さらに統計的な検証手法と Wider Economic Impact 計測手法との関連性についても、整理する。

### 6. 道路政策の質の向上への寄与（研究成果の実務への反映見込み等）

高速道路における速度低下区間の検出については、実務的な応用も容易なため、地方整備局担当者や実務者とともに、他路線への適用について検討中である。既存のストック効果評価手法と統計的評価手法の関係に関しては、本課題の研究メンバーとともに、実務的な応用を検討する。

### 7. ホームページ等（関連ウェブサイト等）

なし

# 第1章 はじめに

## 1.1 データとモデル

リアルタイムであれ蓄積したデータであれ、様々な GPS 計測機器からの移動軌跡や、携帯電話基地局通信による移動・滞在地点のビッグデータの蓄積によって、我々は社会・経済活動を網羅的に、また継続的にモニタリングできるようになった。この情報革命は、都市計画や交通計画、地域計画にも多大なインパクトをもたらすと考えられてきた。この研究の狙いの一つは、土木計画分野で入手できる最新データから有益な知見を搾り取る方法、つまりデータの料理法を、インフラのストック効果というテーマで考えることにある。この課題をはじめめるにあたって、まず土木計画におけるデータと理論の関係を整理しておこう。

数理最適化モデルの構造を経済学の枠組みに翻案した計量経済モデルでは、基本的にデータは理論に従うというスタンスがとられてきた。つまり、計画の根本となる因果関係や都市・経済現象を支配するパラメータが、「どのような形で」分析モデルに埋め込まれているかは、分析が始まる以前に経済理論から導出される。このとき、新しいデータや考慮すべき要素の登場が分析者に理論の改訂を促すこともあるはずだが、少なくとも経済学の範疇ではデータが理論の改訂を迫ることは稀だった。

これに対して、データの持っている情報を絞り出すことによって、意味のある知見を見出そうとするアプローチは、マイニング志向のデータサイエンスと呼ばれる。このアプローチでは、そもそも計画「理論」が存在するというよりも、統計理論（統計モデル）の比重が重い。

たとえば時系列のデータを眺める場合、どうしてもトレンドや変化点などに目を引かれるだろう。地図上にプロットした空間データの場合では、周辺と異なって何らかの集積が見られるホットスポットや、その逆にデータが低い値をとる地点が集塊するコールドスポットなどがこれに該当する。

このとき分析者は、無意識のうちに平均や分散、外れ値や局所値といった物差しを当てながらデータを眺めているのだが、単に記述ベースの統計手法では、解釈性や操作性に難がある。たとえば、既に古典的な手法の範疇に入る多変量解析は、データの持つ情報の圧縮を指向する手法である。一

般にデータマイニング手法とは、情報圧縮のために確率分布や固有空間、特異値やノンパラメトリックながら何らかの分布の特性値に着目した数理統計手法によって、データ記述・圧縮のガイドラインが得られる。計量経済モデルとは異なるものの、これらの統計的手法も、モデルと呼ばれることが多い。

土木計画学の数理モデルは、政策による効果をシミュレートする上で、適切な操作性を備えるべきとする計画論に立脚して推定される。このため、これまでは統計的な規範しか持たないデータマイニングよりも、政策変数やマクロ経済変数を含む計量経済モデルの方が好まれてきたように思われる。あるいは、パーソントリップ調査や地理情報などを除いて、データマイニングするほど膨大なデータが無かったのかもしれない。しかし上述したように、データマイニングであっても、完全にフリーハンドでデータと向き合うのでは、意味のある知見は得られない。データマイニングをすとしても、シンプルであれ高度であれ、何らかの規範やモデルが導入され、その下でデータを眺めなくてはならない。

この状況では、データとモデルの関係はどうなるのだろうか？ 少なくともデータは何かに従属しているわけではなさそうだ。といって、データがあたかも自ら「語るように」分析が進むわけではない。分析者は、何らかの規範や仮説、あるいは予断を持ってデータを眺めている。つまりデータは、分析手法（あるいは統計的仮定）というフィルタを通してのみマイニングされる、そのことには注意を払っておきたい。

いささか迂遠な確認から始めたのは、テーマとして掲げた「複数のデータを活用した道路のストック効果の計測技術の再構築」では、まさにデータと理論の関係が問われていると感じているためだ。すでに事業評価手法としてマニュアル化されている道路3便益（走行時間短縮、走行費用縮減、交通事故減少）に対して、ビッグデータが加えられるような新しい知見はあるか？ 申請者は、データと理論の関係に立ち返りながら、このテーマに取り組もうと思う。

よく知られているように、道路の便益は合理的な交通行動の仮定の下で、経済理論に基づいて発生側で計測されている。この便益計測について、新しいデータがどのような知見を付け加えるか、少し考えてみよう

移動体に関するビッグデータは、時空間解像度が高い。そこでまずは、従来の現象をより精細にとらえる、というデータ活用があり得よう。ただし、単に現象を精細にとらえただけならば、単によく知られている現象を再確認するだけになることもあるだろう。さらに逆説的には、データが精細なことが原因で、分析者が想定した結果が得られないこともあるだろう。

交通現象が「時間的に連続観測される」ことから、土木計画学として得られる新しい知見は何だろうか？ この特長からは、「稀事象」や「災害」といった、通常想定していない状況のデータを集めて分析するスタイルが考えられる。つまり、交通事故減少便益の議論もここに含まれる。交通事故をはじめとする発生頻度の低い事象は、そもそも従来はほとんどデータが無かったはずであり、したがって分析は困難となるだろう。その一方で、通常時に見られる交通現象を分析するために、稀事象を除いて分析するアプローチからは、「既存の交通調査を代替・補完する」という研究が生じるばかりのように思える。これは実務的には重要なテーマだが、そのままではデータと理論の関係に改訂を迫るほどのインパクトとはならないだろう。

一方で、「多地点が同じフォーマットで同時に」連続して計測されることは、何をもたらすか？ しばしば経済理論やモデルで検討の範囲外とされてきた、ネットワークなどの空間的なシステム境界を同定する研究が考えられる。ビッグデータを用いて経済モデルが想定する空間的な経済行為の範囲、ならびに時間的な経済行為の期間を同定するアプローチは、実務的には新たな知見をもたらすばかりでなく、経済地理学の分野へのインパクトは大きい。

「時間的に連続観測される」ことから、「稀事象」の分析が容易になることが考えられる。特に移動体データと地理データを統合利用すれば、たとえば公共施設の容量設計や緊急時の備蓄の議論に直結する成果となり、より政策的なアウトプットにつながるだろう。交通事故対策を論じる分析も、この範疇に含まれる。

通常現象を分析するアプローチではどうだろうか？ なお、時間的に連続観測される移動体データが意味を持つ時間スケールと、その周辺の地理特性を観測することが意味を持つ時間スケールが異なることに注意すべきである。比較的短期では、多くの移動体データにとって地理情報は、単に外生変数や、調整すべき局外母数に過ぎない。

しかし長期ではどうなるだろうか？たとえば、移動体データを時間、または空間単位で集計し、地理情報の計測スケールに合わせて用いる場合を考える。そのようなデータも用いた分析から得

る知見は、上述した短期の移動体データの時空間情報の議論と類似しているものの、その含意はより重要となるだろう。なぜなら時間スケールを長期にすれば、各エリア内の地理情報、すなわち立地も変化するからだ。つまりリッチな移動体データを加えた経済分析からは、「その便益がどの地域との交通によって生じているか」をめぐって、新しい知見が得られる可能性がある。これは帰着側の便益計測を行う計量経済モデルに対して大きなインパクトを与える可能性がある。

土木計画のデータとモデルに関する営みは、長らく都市と交通という、測りがたき対象を測るための方法論を産み出すことであった。理論とデータと社会の3ポイントを踏まえた実践ニーズの検討によって得られるメリットは、PDCAサイクルを通じた計画の合理化やアカウントビリティの向上と、何よりも衆知を結集しやすくするための現象理解をもたらすことであり、それこそが、次の実務や研究を始めるアイディアの母胎となるはずだ。今回の研究プロジェクトが、「複数のデータを活用した道路のストック効果の計測技術の再構築」にどのように寄与するか、これまで述べてきた観点を深めながら、じっくりと考えを重ねていきたい。

## 1.2 本研究のテーマ

これまでの議論を踏まえて、本研究では3つの初期テーマに加えて、データに基づくインフラストック効果の枠組みに関して、1つの統合的テーマに取り組むこととした。

交通状態の質的指標の算出と事故リスク分析では、道路交通状態の質的指標を算出し、交通事故リスク分析において活用する。

観光地のトラベルコスト評価では、自動車依存度の高い地方の移入産業として重要な、地点別の観光入込客に着目し、道路整備によるストック効果を観光ビッグデータと地理情報の結合によるトラベルコスト評価により算出する。これらを踏まえたうえで、観光ビッグデータ、トラベルコスト評価に関する研究動向を整理した。

工業団地地価のヘドニック分析では、道路整備がアクセス条件を改善する効果を、地価への波及効果に着目したヘドニック（地価）分析により明らかにする。特に、高速道路アクセスの重要性がしばしば言及される工業団地に焦点を当てる。これを踏まえたうえで、研究動向を整理した。

## 1.3 交通状態の質的指標の算出と事故リスク分析

本テーマに関しては、交通状態の把握に関する研究と、詳細な地点データを用いた交通事故に関する研究と、詳細な地点データを用いた交通事故

に関する研究の2つの観点からレビューを行った。

## 1) 交通状態の把握に関する研究

プローブデータや車両感知器による交通ビッグデータを活用することにより、交通状態を定量的指標化によって記述し、評価する研究が行われている。

瀬尾ら<sup>1)</sup>は、自動車工学の分野において開発が実用段階に達し、今後普及する可能性を有している車間距離計測機器に着目した。交通状態の観測機器としてGPSと車間距離計測機器を搭載した新たなプローブカーを想定し、そのプローブカーによる移動体観測の情報のみを用いて、交通状態を推定する手法を提案した。既存手法では、 $q-k$  関係・ $k-v$  関係を仮定することによって交通状態を推定していたが、提案手法ではこれらの仮定なしに既存手法とほぼ同精度で交通状態が推定できることを示した。

和田ら<sup>2)</sup>は、複数のプローブ車両軌跡データのみを用いて交通の量的把握が可能な手法を提案した。この手法は、2本のプローブ車両（先行プローブと後続プローブ）の間の交通状態をKinematic Wave (KW) 理論により補完するものであり、車両の詳細な軌跡データを縮約することなく活用することができる。データによる検証結果から、プローブ車両間の交通状態が混雑している状況では提案手法の推定精度が良いこと、および10分~20分程度の交通量推定が可能なことを明らかにした。

Heら<sup>3)</sup>は、都市の道路ネットワークにおける交通混雑を評価した。この研究では、各道路区間での最高速度に対する平均速度の割合によって、道路の混雑状態を4分類した。北京の高速道路ネットワークに提案した分類基準を適用して、時間帯別・季節別に各道路区間の混雑状況を明らかにした。

Hohmannら<sup>4)</sup>は、高速道路や都市の道路での交通状態分類手法を開発した。高速道路のデータを対象に区間別、地点別の平均速度を比較すると、区間平均速度は、地点平均速度（設置場所が固定された検出器による測定）よりも、より妥当な交通状態を表現できることを示した。また、GPSのテスト走行の結果に基づいて分類した交通状態の分類数について、道路利用者のわかり易さの観点から検討した結果、3分類が妥当であることを確認した。

## 2) 詳細な地点データを用いた交通事故に関する研究

事故（突発事象）の検出、事故が旅行時間に与える影響の把握、また事故のリスク分析に関して、様々な研究が行われている。

朝倉ら<sup>5),6),7)</sup>は、プローブカーデータを用いて、

突発事象検出する2手法の提案を行った。これらは、プローブカーの旅行時間の変化と交通容量の低下を考慮して突発事象を検出する手法と、衝撃波速度と交通容量の低下から突発事象を検出する手法である。その結果、後者の衝撃波速度を用いる手法は、最低3台のプローブカーデータがあれば事故等の突発事象を検出できるという利点を有することを示した。

Parkら<sup>8)</sup>は、事故発生後に起こる二次的な事故について、ベイジアンニューラルネットワーク理論を適用した検出方法を提案した。この手法では、プローブカーから得た速度情報によって、ロードクリアランス、待ち行列の状態、そして交通量の回復の3指標を関数とする二次的な事故の蓋然性（尤度）を算出するシステムであり、その値は対象とするネットワークの各区間において、前時点の尤度と新たに観測した速度を組み合わせた値として、再帰的に得られる。

Kinoshitaら<sup>9)</sup>は、事故などの非日常的な混雑と慢性的な渋滞などの日常的な混雑を分類することにより、プローブカーから交通事故を検知する手法を提案した。この研究が提案する検知手法は、まず一定区間内を走行している車両の速度情報を大量に収集して、日常的な交通状態を表す交通状態トピックを学習する。さらにある区間について最新のプローブデータに基づいて得られる交通状態が、学習した日常的な交通状態の分類のいずれにも合致しない場合に、非日常的な混雑と判定する。提案手法を東京の首都高速道路の実データに適用して交通状態トピックとの合致/非性能を判定するアルゴリズムの性能を比較したところ、その判定に用いる閾値は、カルバックライブラー情報量を用いて定義する場合は、最も良好なことを示した。

D'Andreaら<sup>10)</sup>は、スマートフォン等のGPSデータを用いて交通混雑や事故を検知するエキスパートシステムを提案した。このシステムでは、区間別の速度と区間内の車両数に基づいて、交通状態を5分類（車両未観測、通常速度、速度低下、大幅な速度低下、完全ブロック状態）で判定する。さらに、検出された現象の空間的な連続性に基づいて、警報を発するアルゴリズムを提案した。ピサの道路ネットワークにおいて提案システムを検証したところ、実際の事故や渋滞を正確に検知できることを示した。

Liら<sup>11)</sup>は、事故継続時間に注目して、これに影響する様々な要因の効果や継続時間予測を分析するためのモデルを提案した。この研究は、事故に関するテキスト情報にトピックモデルを適用して、各サンプルのトピック別の負荷量として定量化する。事故継続時間の記述には、競合ハザードモデルを用いた。シンガポールで収集



された2年間に及ぶ約1万件の事故データに対して提案手法を適用したところ、時間依存共変量を含む競合ハザードモデルによる事故継続時間の予測が最も精度が高く、また算出した25トピックのうち10~14トピックは、事故継続時間に有意に影響していることを明らかにした。

Hojatiら<sup>12)</sup>は、高速道路の旅行時間信頼性における交通事故のインパクトを定量化する手法を提案した。この研究では、事故時に実際に計測された区間別の所要時間の分布情報を単一指標として定量化した合成指標を提案し、その値に対して、ランダム変数を仮定したトビットモデルを推定して要因分析を行った。クイーンズランド州において収集した1年間に及ぶ高速道路の事故データを、事故分類別に分析した結果、有意な事故要因として、複数台事故、事故時間、重大事故などが挙げられた。

事故の発生リスクについて、交通状態との関係を整理した研究も行われている。藤井ら<sup>13)</sup>は、車両感知器から得られる定点観測データと、勾配の影響を適切に考慮した交通流モデルで構成されるフィードバック型交通状態推定システムを構築して、平常時、および交通事故発生前後におけるモデルパラメータの変動傾向を比較した。その結果、平常時と比較して、事故発生直前の自由流速度が低下する傾向や、勾配変化に対する速度変動の感度が高くなるケースが存在することを指摘した。

兵頭ら<sup>14)</sup>は、5分間のQ-K関係に基づいた交通流状態を、自由流、渋滞流以外に第3の状態として“混合流”を考慮し、3つの交通流状態が事故発生リスクに与える影響について比較分析を行った。その結果、混合流よりも全車線が渋滞流の方が、追突事故および車両接触事故の発生リスクが高くなる傾向を明らかにした。

#### 1.4 観光地のトラベルコスト評価

本テーマでは、自動車依存度の高い地方の移入産業として重要な、地点別の観光入込客に着目し、道路整備によるストック効果を観光ビッグデータと地理情報の結合によるトラベルコスト評価により算出する。これらを踏まえたうえで、観光ビッグデータ、トラベルコスト評価に関する研究動向を整理した。

##### 1) 観光ビッグデータに関する研究

Önder<sup>15)</sup>は、Flickrのジオタグ付けされた写真に基づいてオーストリアの周遊トリップを分類する手法を提案した。オーストリアの旅行者を対象に提案手法による分類を行った結果、2パターンに分類できた。また、それぞれの周遊トリップ毎にトリップ箇所に対するクラスター分析を行った結果、第一の都市は大都市を含み、第

二のクラスターはオーストリア西部の都市となった。これらは、同様のクラスター内の都市間を巡回する共同販売キャンペーンや新たな観光事業商品の創出に向けた知見を与えられると結論付けている。

Raunら<sup>16)</sup>は、時空間トラッキングデータを用いて旅行者の目的地までの流動を測定する手法を開発した。2011年から2013年までのエストニアにおける外国からの訪問者の携帯電話のポジショニングデータを用いて、空間、時間、および旅行者(出身地域)の組み合わせについて分析を行い、国内の旅行者との傾向の違いを明らかにした。

Tooleら<sup>17)</sup>は、オープンデータやクラウド地理データ、センサス記録や調査、また携帯電話からの通話記録を用いてOD表を推定する手法を提案した。提案手法では、オープンデータやクラウドソースのデータリポジトリを使って構成した道路ネットワークを通過したトリップ経路を生成する。またこのアウトプットを研究者、政策作成者、また一般の人と相互で共有できるオンラインプラットフォームも作成した。これらのシステムは、世界のいくつかの都市で分析が行われ、有効性が示された。

日本のビッグデータの一つであるモバイル空間統計の活用に向けた初期研究として、清家ら<sup>18)</sup>はモバイル空間統計の今後の可能性や統計の信頼性の評価を行った。その結果モバイル空間統計の時間帯別推計、人口属性別推計、居住地別推計については、一定の信頼性が得られることが分かった。一方で、空間解像度については、一定の人口密度を前提に1kmメッシュ以上の活用が望ましく、特定の施設や狭い地区内における詳細な人口変動把握には、必ずしも向いてないことが明らかになった。

清家ら<sup>19)</sup>は、モバイル空間統計の初期検討の一環として、モバイル空間統計に他統計を組み合わせた拠点地区の分析モデルと、モバイル空間統計単体での複数拠点地区の比較分析モデルを検討した。従来はアンケートや目視など労力のかかる調査方法しかなかった情報をモバイル空間統計によって簡易に把握できることを示した。また、モバイル空間統計のデータのみを活用し、指標化することで、人口流動に着目した拠点の性格分類を行う方法を示した。

田中ら<sup>20)</sup>は、「モバイル空間統計」の特徴を生かして、観光地での人の集中性及び変動制を把握し、観光地の特性を捉えるとともに人口比等の統計値を用いて、温泉地、テーマパークを有する自治体、主要観光地の傾向の差異を把握した。

## 2) トラベルコスト評価に関する研究

旅行費用法は、訪問地までの旅行費用と訪問回数との関係をもとに間接的に訪問地の利用価値を評価する方法と想定される利用者の訪問の意向を考慮して推定される方法に分かれる。いずれも評価すべき対象が「訪問するだけの価値」を持つことが前提となり、訪問されない対象については、評価が困難であるといわれている。本研究では、少なくとも総移動支払額以上の魅力度が目的地にあると考える、最もシンプルな方法で分析を行う。

菊池<sup>21)</sup>は、公の施設に係る費用便益分析を議論するため、公立図書館に注目し、旅行費用法を用いた便益計測を検討し、公立図書館の費用便益手法としての旅行費用法の有効性、移転に伴う便益の変化とその要因を明らかにしたが、旅行費用法の問題点として一般化費用の精度の問題等も明らかとなった。

渡邊<sup>22)</sup>は、旅行費用法において生じる代替地価格問題に着目し、対処法を提案した。代替地価格を正確に計測することは非常に困難であるため、代替地価格には計測誤差が生じ、その結果、推定便益が誤って計測されてしまうことを指摘した。提案した方法と、通常行われている方法を比較することで影響の大きさを実証的に明らかにした。

## 1.5 工業団地地価のヘドニック分析

本テーマでは、道路整備がアクセス条件を改善する効果を、地価への波及効果に着目したヘドニック（地価）分析により明らかにする。特に、高速道路アクセスの重要性がしばしば言及される工業団地に焦点を当てる。これを踏まえたうえで、研究動向を整理した。

旧通産省の定義によれば、工業団地とは、複数の工場等の立地を想定して、計画的に用地の取得、あるいは敷地の造成・整備等が行われ、分譲された一団の工業用地等（一体的に取得・造成・整備される道路・緑地・住宅用地等の公共施設を含む）とされる。工業団地については、1960年～70年代、地理学で盛んに研究が行われ、立地動向、工業構成、吸引要因、雇用、地域経済へのインパクト、などに関する知見が蓄積された。やや古い文献であるが、百瀬<sup>23)</sup>によれば、工業団地の主な目的は、立地制約の解消（都市公害、従業員の確保難の解消）、産業再編成による都市計画の促進、産業再配置による低開発地域の開発、低開発地域の開発のための企業誘致となっている。

土木計画学における工業団地に関する研究は非常に少なく、工業団地の立地選択理由や工業団地販売価格の推移について考察した伊藤<sup>24)</sup>など、数えるほどしかない。工業地価のヘドニック分析に関する研究も我々がレビューした範囲では存

在しない。一方で、海外ではRyan<sup>25)</sup>、Ambrose<sup>26)</sup>、Beekmans et al.<sup>27),28)</sup>、del Saz-Salazar and García-Menéndez<sup>29)</sup>、Clark and Pennington-Cross<sup>30)</sup>といった事例がある。これらの研究で地下に及ぼす影響が大きいとして特定された要因は、交通（空港、港湾、インター、鉄道、および都市へのアクセス）、市場（人口、都市化率）、工業団地固有要因（リース期間、分譲割合、規模、年）であり、インフラのストック効果が示唆されている。本研究においても、データが入手できる限り、これらの変数の導入について検討する。

## 第2章 交通状態の質的指標の算出

### 2.1 暫定2車線区間を対象とする速度低下検出

ETC2.0 データの利用を念頭に、モデルケースとする高速道路ネットワーク上のデータ取得状況を確認する分析用のデータベースを構築した。モデルケースの選定にあたっては、中国地方整備局と十分に協議を行い、鳥取姫路自動車道（以下、鳥取道）の佐用～鳥取とした。

ETC2.0 データの特徴を考慮し、時間別・区間別の速度分布を抽出することを目的とし、モデルケースである鳥取道を対象に速度低下に関する分析を行った。まず、本テーマで用いた ETC2.0 データから対象区間を走行した自動車のみを抽出した。そして、抽出したデータの速度分布から低下区間を整理する。さらに、複数の候補要因の中から、実際の速度分布の違いをもたらす要因を抽出できる決定木を用いることで速度低下区間と合わせて同時にその要因を明らかにした。

本分析では、まず将来的に4車線化が予定されているが、現在は片側1車線の対面通行で供用されている高速道路は暫定2車線道路を対象とする。暫定2車線区間は、日本の高速道路の約3割(31)を占めている。2016年6月以降、国土交通省は暫定2車線道路のうち速度低下が著しい一部の区間に追い越し車線を設ける方針(32)を示した。これは地方からの強い要望に応じた方針ではあったが、具体的にどの区間に追い越し車線を付加するかについて、定量的な根拠は乏しいといわざるを得ない。従来の4車線以上の区間に関して追い越し車線を設ける区間やその区間長を決定する指針を示す研究(33)では、常時観測によらない線形などが基準として言及されているばかりであった。事業を実施する上で、直接的な根拠となる道路のパフォーマンス指標としては、当該区間の走行速度情報の活用が考えられる。追い越しができない暫定2車線区間では、4車線区間と比べて前方の低速車の影響が後方車群に与える影響が大きいと考えられる。しかし、これまで低速車による速度低下の伝播を明らかにする研究の蓄積は、ほとんど行われてこなかった。これは、区間別の速度変化を常時把握する必要があるためである。現在、高速道路上の車両の速度は断面交通量を把握する目的で設置されているトラフィックカウ

ンターを用いて得られているが、同機器の設置間隔はおおよそ10kmである。この程度の観測密度では、いつどこで発生するかわからない低速車による速度変化の影響を捉えることはできない。

この分析では、ETC2.0 データの特徴を考慮し、暫定2車線区間の追い越し車線整備区間を特定するために時間別・区間別の速度分布を抽出することを目的とする。また、ETC2.0 データ処理上の留意点を明らかにすることで、今後の研究に活用できるようにする。

### 2.2 データセットの構築

ETC2.0 データにはデータフォーマットが13種類存在する。これらのうち、本テーマでは、詳細な分析が可能な様式1-2を用いた。様式1-2には、表2.1に示す33項目のデータが格納されている。緯度・経度情報には①ローデータ（観測した緯度・経度そのもの）のほかに、②マップマッチング後の緯度・経度がある。

No.15とNo.16に記録された①前者は、車両に搭載されたGPSから得られたデータであり、GPSの精度によっては走行中の道路外で観測されることもある。No.22とNo.23に格納された後者のデータは、最寄りの道路への位置補正処理を、道路位置情報を参照して事後に施した結果を示している。

②後者については補正アルゴリズムが不明であって精度検証ができないため、以下ではマップマッチング前のNo.15、No.16の経度・緯度データを使用する。

対象エリアと分析期間は、図2.1に示す鳥取道<sup>36)</sup>を対象に研究を行った。鳥取道は兵庫県姫路市の佐用JCTから岡山県を経由して、鳥取県鳥取市の鳥取ICを結ぶ全長62.3kmの暫定2車線で供用されている高速道路である。2009年に最初の区間が開通し、2013年に全線開通した。西粟倉ICと智頭ICの間の一部区間が一般国道（厳密には志度坂峠道路起点（西粟倉ICの南0.5km）と智頭ICの間が高速自動車国道に並行する一般国道自動車専用道路）である。全長62.3kmのうち、約30kmがトンネル区間である。

表 2.1 ETC2.0 様式 1-2 データフォーマット

No	情報項目		単位	データ型	桁数	小数部 桁数	備考	
1	運行情報	RSU-IDコード		文字列型	14	—	センター間インタフェース 装置番号: センターコード+機器種別+連番	
2		受信時刻		日付型	14	—	YYYYMMDDHHMMSS	
3		運行日		日付型	8	—	YYYYMMDD	
4		運行ID		文字列型	12	—	変換後ASL-IDを日毎の通し番号(連番)に変換したコード	
5	車種情報	自動車の種別		整数型	2	0	0: 軽二輪 1: 大型 2: 普通 3: 小型 4: 軽自動車 5~15: 未定義	
6		自動車の用途		整数型	2	0	0: 未使用 1: 乗用 2: 貨物 3: 特殊 4: 乗合 5~15: 未定義	
7	トリップに関する 情報	マップマッチン グ前の情報	GPS時刻		日付型	14	—	(YYYYMMDDHHMMSS) 年月日時分秒(時間は24時間とする) 電波ビーコン5.8GHz帯データ形式仕様書 【走行履歴情報1】4-2-2 + 4-3-4
8			通し番号		整数型	5	0	運行日毎のGPS時刻順の通し番号
9			トリップ番号		整数型	3	0	
10			トリップの起点時刻		日付型	14	—	
11			トリップの終点時刻		日付型	14	—	
12			トリップの完全性		整数型	1	0	0: 完全トリップ 1: 上流側不完全 2: 下流側不完全 3: 両側不完全
13			トリップ起終点フラグ		整数型	1	0	0: 起点 1: 終点 2: 通過点 3: その他
14			タグ番号		整数型	2	0	電波ビーコン5.8GHz帯データ形式仕様書 アップリンク編
15			経度	deg	小数型	10	7	差分データから通常の経度に戻した情報
16			緯度	deg	小数型	9	7	差分データから通常の緯度に戻した情報
17			蓄積条件		整数型	1	0	0: 走行距離による蓄積 1: 進行方向変化による蓄積 2: 発動履歴による蓄積
18			道路種別コード		整数型	1	0	未使用 0: 高速 1: 市内高速 2: 一般道路 3: その他 電波ビーコン5.8GHz帯データ形式仕様書 【走行履歴情報1】4-2-5 + 4-3-7
19		速度	km/h	小数型	5	2	電波ビーコン5.8GHz帯データ形式仕様書 【走行履歴情報1】4-2-6 + 4-3-8	
20		高度	m	整数型	4	0	未使用 電波ビーコン5.8GHz帯データ形式仕様書 【走行履歴情報1】4-2-7 + 4-3-9	
21		マップマッチン グ後の情報	マッチングフラグ		整数型	1	0	0: 成功 1: 経路探索失敗(基本道路を通らなかった) 2: 回転角許容外 3: 平均時速許容外 4: 経過時間許容外 5: その他 (緯度経度付近にリンクなし・前後の走行履歴が無いなど)
22			マッチング後経度	deg	小数型	10	7	世界座標系
23			マッチング後緯度	deg	小数型	9	7	世界座標系
24	DRMバージョン			整数型	4	0		
25	2次メッシュコード			整数型	6	0		
26	流入ノード			文字列型	6	—	区画コード(1桁)+ノード(5桁)	
27	流出ノード			文字列型	6	—	区画コード(1桁)+ノード(5桁)	
28	流入ノードからの距離		m	整数型	4	0		
34	確定フラグ		整数型	1	0	0: 未確定、1: 確定		
35	交通調査基本区間番号		文字列型	11	—	都道府県CD(2)+道路種別CD(1)+路線番号(4)+順番号(4)		
36	上り・下りコード		整数型	1	0	1: 上り、2: 下り		
37	管理者コード		整数型	1	0			
38	更新情報	更新日時		日付型	14	—	YYYYMMDDHHMMSS	

ETC2.0データは国土技術政策総合研究所から提供を受けた。取得期間は2015年4月から2016年3月である。以下の分析では、データの基本的な処理手順について整理するため、夏期休暇や降雪の影響を受けない2015年4月から6月までの3か月分のデータを用いた。

ETC2.0データはファイルサイズが巨大なため、日付ごと、1次メッシュごとのcsvファイルに記録されている。これらのファイル名は、4月1日のメッシュ番号5334のデータなどのように名づけられている。本テーマの対象区間である鳥取道は1

次メッシュ5234と5334にまたがっている。ただし、1次メッシュ5234と5334には鳥取道以外の道路も存在するため、鳥取道の走行記録データを抽出する必要がある。

抽出アルゴリズムを作成し、データを抽出した結果、鳥取道におけるETC2.0搭載車両の1か月分の位置情報として約40万レコードが得られた。これは約60kmを200m間隔でデータ取得したとして300区間あるので、200m当たり1月で平均1300点の位置(速度)データが得られていることを意味する。つまり1日の観測数は40点程度であ



図 2.1 鳥取道の位置

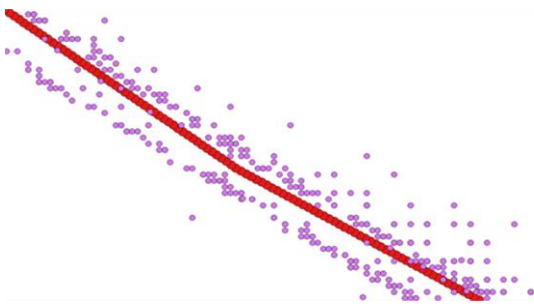


図 2.2 ETC2.0 データの測位点と DRM リンク (佐用 JCT~佐用平福 IC 間)

り、1時間あたりでは2点未満にすぎない。分析に必要なデータを抽出し、データベースを構築するプログラムを作成した。今回の鳥取道を例に、データベース構築までの手順を図 2.4に示す。

3か月分のデータ抽出では、3か月×30日×2メッシュ (5234と5334)、の180ファイルを扱わなければならない。ファイル数が多いため、C言語プログラムを用いて、以下の手順でデータ処理を行った。まず、DRMビューワを用いて対象路線の二次メッシュコード (7つ)、流入ノード、流出ノードを読み取り、対象路線のリンクのリストを作成した。このリストを用いて対象路線のデータのみを抽出するC言語プログラムを実装して、データを抽出した。

### 2.3 測位精度の検証

ETC2.0はGPS機器によって観測されているため、走行地点の上空視野の大きさによって、衛星

表 2.2 DRM リンクと ETC2.0 データの測位誤差 (データ数約 20 万)

パーセンタイル	測位誤差 (m)
5%	0.58
25%	2.52
50%	5.60
75%	9.81
95%	18.35

表 2.3 測位方法別データ割合

測位方法	該当レコード数	割合 (%)
未定義	2,313	0.607
GPS	115,167	30.206
GPS+ジャイロス コープ+スピード パルス	472	0.124
Map Matching	263,316	69.063
計	381,268	100

表 2.4 DRM リンクと ETC2.0 データの測位誤差

パーセン タイル	測位誤差 (m)	
	GPS	Map Matching
5%	0.58	0.58
25%	2.55	2.46
50%	5.74	5.25
75%	9.90	9.64
95%	18.56	17.66

の捕捉可能数が異なる。衛星の捕捉数が機器性能に左右されるのは当然だが、そのほかに重要な要因として、車内の機器設置場所も電波状態を左右する。機器間の測位精度に大きな違いがないと考えると、設置条件によってETC2.0データの位置情報の精度は、大きく低下する可能性がある。さらに速度が距離差分と時間差分から求められているとしたら、位置の精度は速度にも影響を及ぼす恐れがある。そこで、ETC2.0データとデジタルロードマップのマッチング前のデータを用いて、測位誤差の大小による速度分布の違いを検証した。また、合わせてデータの取得状況や速度分布状況を確認した。

図 2.2に、得られた位置情報 (点) をGISソフトにより地図上へ描画した結果の一部を示す。赤い点はDRMリンクと呼ばれる鳥取道上り線の道路中心から作成した情報であり、紫の点はETC2.0データを表す。同図に示すように、実際の道路中心から約10m離れた点も存在する。そこで、測位誤差のばらつきを確認することとした。まず、各点の測位誤差をGISソフトの最近接ツールを用いて求め、それらを昇順に並べ替えた。その上で、位置情報の観測総数に対するパーセンタイル値ごとに測位誤差を求め、その結果を表 2.2に示す。

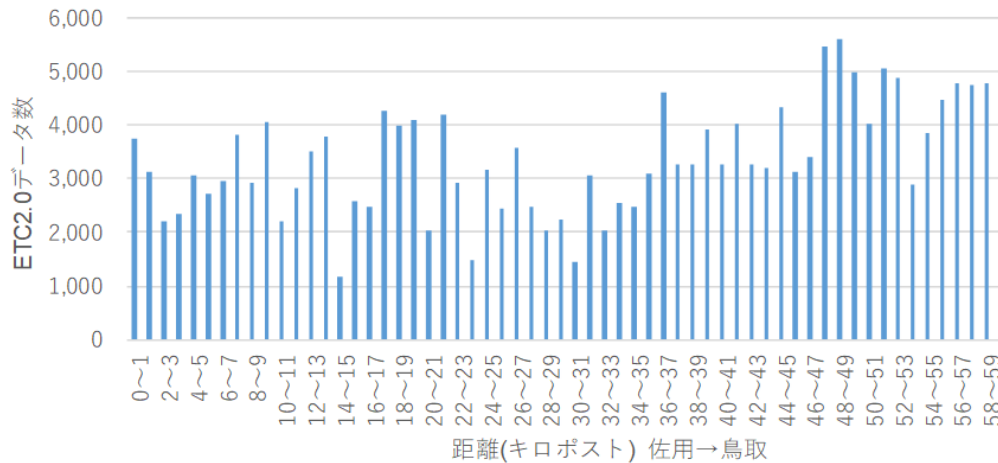


図 2.3 データ取得状況\_全データ

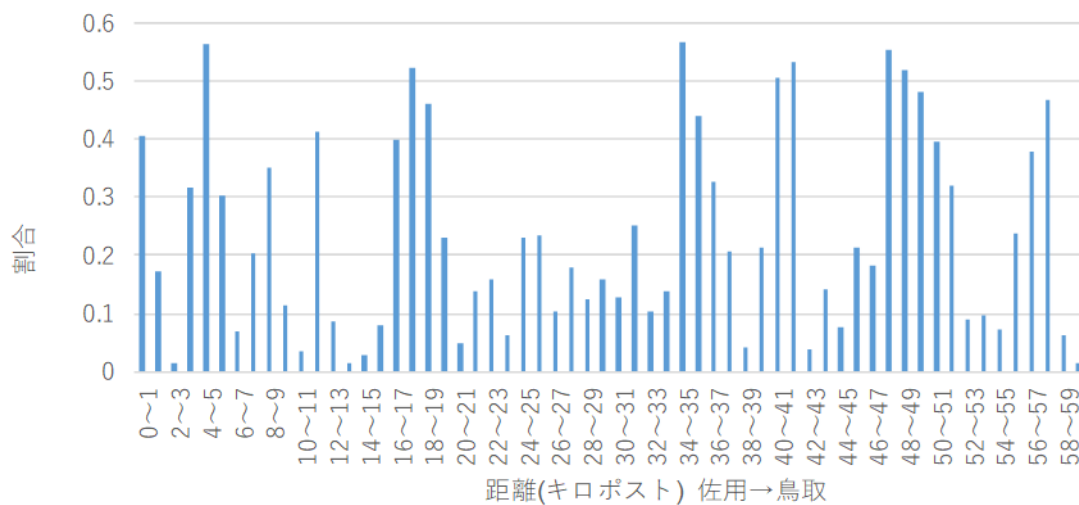


図 2.4 データ割合\_測位誤差 25%タイル以内 (2.52m 以内) データ/全データ

鳥取自動車道は車線幅が両側で7mのため、表2.2より、25%タイル値の2.52mであれば、両側の車線幅に収まる。そこで以下では、測位誤差2.52m (25%タイル) 以内の点に着目する。

4月の鳥取道走行データについて集計した結果を表2.3に示す。割合をみるとGPSとMapMatchingが全体の99%を占めているため、両者による測位誤差の確認を行った表2.4に測位方法別の測位誤差を示す。両者を比較すると測位誤差について大きな差は見られない。

図 2.3に、1kmごとの4月上り線の全ETC2.0データ数と測位誤差25%以内のデータ数を示す。同図より区間ごとにETC2.0データ数は異なることがわかる。また図2.4に、区間別の即誤差25%タイル以内のデータ割合を示す。両図の比較から、全データでは1km当たり最低でも1000データ取得できているが、25%以内のデータに限定すると、一部ではほとんどデータが取得できない区間が生じることが分かる。ETC2.0では車速パルスに

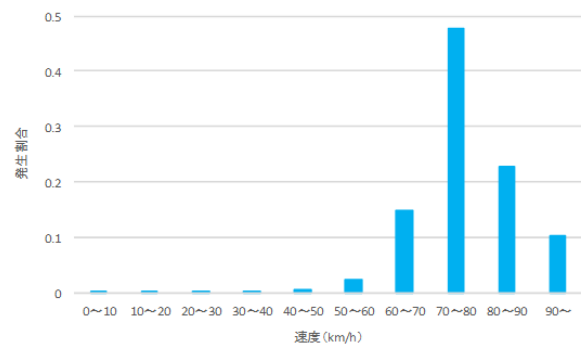


図 2.5 GPS 速度分布図 (全データ)

よる速度と、GPS測位による速度が収集されている。本テーマではこのGPS測位点の速度をGPS速度と呼ぶ。GPS速度の利用可能性を検討するため、計算速度同様に全データを図2.5に示す。また、パーセント別の速度分布を集計した結果を図2.6に示す。GPS速度の速度分布は、測位誤差の

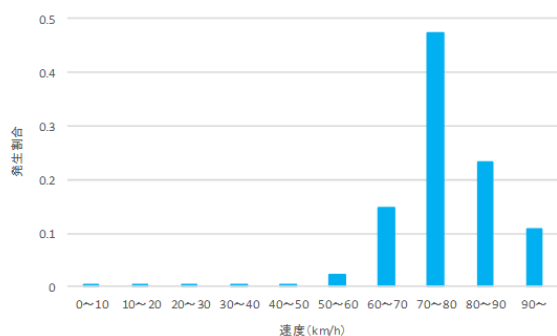


図 2.6 GPS速度分布図(測位誤差95%  
マイル以内)

大小による有意差はない。一方で、標本数の増加により、最頻値の出現が下がる（分散が小さくなる）傾向が見られるなど、統計分布として自然な形状を示していることがわかる。なお、この性質は計算速度ではみられなかった。本テーマではGPS速度を用いて分析を行う。

## 2.4 分析手順

本研究では、軽微な速度低下を検出するために、事前データの加工手順と検定手法に工夫を凝らす。既存の手法として、速度分布データをそのまま用いて、速度低下が疑われる区間と全データの分布の違いに関して検定（コルモゴロフ・スミルノフ検定や $\chi^2$ 検定など）を行う方法が考えられる。しかしこの検定手順を試行したところ、データのサンプル数が検定結果に大きな影響を及ぼすため、分布中央部にほんの少しの速度分布の違いが表れる場合に、分布の違いすぐに検出していまい、結果として決定木で算出した全てのリーフについて、統計的な有意差を検出してしまうという限界があった。つまりこの方法では、全て速度分布が異なる、という結果しか得られず、政策的な判断を行う手法として有用ではない。

一方で提案手順は、以下に述べるようにパーセントイル値に基づく離散化と決定木分析、さらに後述するコ克蘭・アーミテージ検定による一様分布からの乖離に関する傾きの検定を組み合わせることにより、1)離散化した各カテゴリを公平に扱うことができる。また、2)リーフとして検出したサブグループのサンプルサイズの影響を受けないように配慮する。以上から提案手順では、3)従来のデータ処理手順では検出が困難だった軽微な速度低下を目指す。他方でこの手順は、分布の裾で発生する極端な速度低下（重大な交通障害に沿うと）に関連するサンプル数が少ない場合は、一様分布では明確な検出が困難となることが予想される。ただしそのような顕著な速度低下は、本手法によらずとも検出可能なため、実用的な問

題は少ないと思われる。

なお本結果に基づいて、道路や交通制御、TDMなど道路に何らかの対策が行われると速度分布は変化するため、同一の手順を繰り返しても、いつまでも対策が必要な区間が抽出され続ける。このような場合、速度低下区間の改良効果検証手順を工夫する必要がある。対策前の状態に対して速度分布が改善（つまり分布が、速度の低い方から高い方に寄る形状）か否かを検証するためには、1)対策前の速度分布の閾値の情報を用いて、対策後のデータを離散化したデータベースを作成し、このデータベースに決定木を適用してリーフを検出し、それぞれにコ克蘭・アーミテージ検定を適用して速度低下側だけ検出すればよい。この手順の最後に得られる傾きの正負によって、対策前の速度分布（一様分布）よりも、速度分布が悪化する区間が表れるか否かを検証できる。

## 2.5 決定木分析

決定木<sup>31)</sup>は意思決定ツリーやデシジョンツリーとも呼ばれ、樹形図、またはツリー構造と呼ばれる図によって、着目した目的変数の分布が、ツリーの分岐点で条件付けしたときに、最も異なるような分岐条件を探索するアプローチである。複数の説明変数の候補集合の中で、どの変数が分岐条件として適切か、つまり最も目的変数の分布が異なるサブグループを生成するかを明らかにする手法である。決定木分析では、より大きな部分集合に影響する説明変数から順に分岐していく。決定木において分岐要因を示す点をノード、ある分岐要因と次の分岐要因まとの間を結ぶ線分をリンク、さらにノードの末端部分の、目的変数分布をリーフと呼ぶ。以下の分析では分岐数が2つ以下の2分木を用いるが、アルゴリズムのパラメータ設定によっては、分岐数が3以上の結果も得られる。

決定木を求めるアルゴリズムはC5.0やID3、CHAID、CARTなど複数種類存在する。本研究では、CARTを基に作られたRプログラムのパッケージ、rpartを用いて決定木分析を行う。rpartはCARTに基づいたプログラムである。同パッケージの特長は一度分岐に現れた説明変数が、それ以下の分岐において再度出現ができる点と、最終的なリーフの目的変数分布をグラフ化して出力する点である。分析によって得られる決定木において、根からリーフへのリンクをたどれば、各リーフの速度分布をもたらす分岐条件の組み合わせが得られる。

CART<sup>32)</sup>は以下の3ステップによってツリーモデルが生成される。

- 1) 樹木の成長
- 2) 樹木の刈り込み
- 3) 最適な樹木の決定

1)はノードを構成されるサンプルを2群に分割し、ある停止基準に達するまで分割を繰り返す手順である。2)は分析の精度を向上させるために1)で生成したツリーから、重要ではない分岐を刈り取る。3)は交差確認法などを用いて2)で得られた樹木列の中から最適基準に合う樹木を選択する。

### 1) 樹木の成長

目的変数 $y_n$  ( $n=1, \dots, N$ )を連続変数とする。通常、回帰樹木では、誤差の均一性を基準におき、ノード内の目的変数の平均を回帰予測値に用いたときの回帰残差をノード内不均一性測度として用いる。すなわち、ノード $t$ の回帰残差 $r(t)$ は、残差平方和

$$r(t) = \frac{1}{N(t)} \sum_{n \in t} \{y_n - \bar{y}(t)\}^2 \quad (2.1)$$

である。ここに、 $\sum_{n \in t}$ はノード $t$ に属する個体のすべての和を表し、 $\bar{y}(t)$ はノード $t$ 内の目的変数の標本平均(ノード $t$ に属する個体の目的変数の予測値)

$$\bar{y}(t) = \frac{1}{N(t)} \sum_{n \in t} y_n \quad (2.2)$$

を表し、 $N(t)$ はノード $t$ に属する個体数を表す。これを用いて、ノード $t$ で(真のモデルと推定モデル間の乖離を表す)不均一性指標

$$R(t) = p(t)r(t) = \frac{1}{N} \sum_{n \in t} \{y_n - \bar{y}(t)\}^2 \quad (2.3)$$

が定義される。ここに、 $p(t) = N(t)/N$ であり、データの全個体が $t$ に属する確率の再代用推定値と呼ばれる。

ノード $t$ において、ある分岐規則 $s_t$ をもって2分岐される子ノードを $t_L, t_R$ とする。分岐規則 $s_t$ は、説明変数 $x_p$ が順序付き変数の時、 $s_t = \{x_p < c\}$ の形式をもち、 $x_p$ が名義変数の時 $s_t = \{x_p \in c\}$ である。ここで、 $c$ は変数 $x_p$ である分岐候補を表す。

ノード $t$ に属する個体 $n$ は、分岐規則 $s_t$ を満たせば、 $t_L$ に送られ、そうでなければ $t_R$ に送られる。分岐規則 $s_t$ に対する分岐速度は、リスクの再代用推定値の減少量

$$\Delta R(s_t, t) = R(t) - R(t_L) - R(t_R) \quad (2.4)$$

により測られる。このとき、ノード $t$ で選択される分岐規則 $s_t$ は、分岐規則 $s_t$ 集合をと $U_t$ おくと、 $U_t$ の中で最大の減少量を与える分岐規則

$$s_t^* = \operatorname{argmax}_{s_t \in U_t} \Delta R(s_t, t) \quad (2.5)$$

として選択される。すべての分岐対象のノードにおいて、このような分岐が実行され新しい子ノードが次々に生成される。このようにして、樹木 $T$ は

根ノードから出発して順に分岐(成長)し、一定の停止基準に達するまで行われる。

### 2) 樹木の刈り込み

樹木 $T$ が大きくなるにつれて、式(3)の $R(t)$ は小さくなるが、大きな樹木に基づく予測性能は必ずしもよいとはいえない。これはデータの過学習と呼ばれ、分析結果の安定性が損なわれる。そこで決定木から過学習を除くために樹木の減少列を作る方法<sup>32)</sup>として、樹木 $T$ のリスクの再代用推定値に複雑度のペナルティを課すことで、複雑度コストと呼ばれる

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}| \quad (2.6)$$

を定義する方法がある。式(2.6)では、樹木 $T$  ( $\ll T_{max}$ )の複雑度をリーフ数 $|\tilde{T}|$ とおき、複雑度を実数 $\alpha (\geq 0)$ で規定している。 $T_{max}$ は成長過程によって最終的に得られる最大樹木を表し、 $T \ll T_{max}$ の記法は、 $T$ が $T_{max}$ の部分樹木であることを表す。この $R_\alpha(T)$ に基づいて任意の $\alpha$ 値に対して、

$$T(\alpha) = \operatorname{argmin}_{T \ll T_{max}} R_\alpha(T) \quad (2.7)$$

で定められるある部分樹木 $T(\alpha)$ を見つけられる。ここに、同じ $R_\alpha(T)$ の値をもつ樹木が存在すれば、最も小さな部分樹木を $T(\alpha)$ とする。以上の手順で、任意の $\alpha$ に対し $R_\alpha$ を最小にする部分樹木系列が構成できる。

### 3) 最適な樹木の決定

前節の刈り込み過程により得られる部分樹木列 $T_j$ のなかで、どの樹木が最良であるかを決定する必要がある。通常、CARTでは交差確認法を用いて予測性能の高い決定木を決定する。 $k$ 分割交差確認法では、全体標本 $N$ を $k$ 分割し、そのうちの1つをテストデータにし、残りの $k-1$ 個のデータを訓練データとして使用する。交差確認推定値は式(2.8)と表せられる。なお、 $d_j^k$ は予測子を表し、 $L(y_n, a)$ は $y = y_n$ のとき $a$ の行動を取る損失関数である。この交差確認値が最小値をもつ樹木 $T_j$ が選択される。

$$R(T_j) = \frac{1}{N} \sum_{i=1}^k \sum_{(y_n, x_n) \in N} L(y_n, d_j^k(x_n)) \quad (2.8)$$

### 2.6 目的変数と説明変数

本研究では決定木分析の目的変数は、ETC2.0データに含まれる各点の瞬間速度(以下、GPS速度)とする。説明変数として、以下の6種を設定した。

- ・ 平休(平日, 休日), ダミー変数
- ・ 交通量(1時間当たり), 連続変数
- ・ HVR(大型車混入率), 連続変数



- ・ 天候(雨, 晴れ), ダミー変数
- ・ キロポスト(0.1km 刻みの距離)
- ・ 時間帯(朝: 7,8 時台, 昼: 9~16 時台, 夕方: 17,18 時台), カテゴリ変数

鳥取道の多くの区間では最高速度が 70km/h だが, 一部区間に 40~60km/h の区間が存在する。つまり, 後者の区間では, 必然的に速度が低下するため, 分析の前提であるサンプルの観測条件が異なる。そこで以下では, あらかじめ後者の区間から得たサンプルを除き, 最高速度 70km/h の区間のみのデータを用いて分析を行う。また, HVR は%ではなく小数点表記である。さらに交通量などが大きく異なる夜の時間帯のデータは除外した。なお速度に影響を与える道路構造の情報として勾配や分流点, カーブなども考えられるが, これらはいずれも特定の場所に依存する情報のため, キロポスト情報を事後確認することで, 十分と考えられる。よって, 以下ではキロポストのみを道路構造を表す説明変数として扱う。

## 2.7 速度低下の検定

速度低下の判定には, 全データの速度分布とリーフの速度分布が異なっており, かつ後者が前者よりも遅いことを統計的に示す必要がある。しかし GPS 速度そのものの分布は対数正規分布のような形状である。そのような分布が 2 つあったとして, それらが統計的に有意であるか, 検定を行うことは困難である。そのため GPS 速度そのものを目的変数として検定を行うことはできない。これは, ビッグデータに対して通常の仮説検定を適用する場合に共通する問題である。例えば, 全データとリーフの分布にそれぞれ正規分布を仮定して, それぞれの平均と分散から正規検定を行う場合を考えよう。正規検定統計量  $z$  は, 全データのサンプル数を  $N$ , リーフのサンプル数を  $n_i$ , 全データの平均と分散を  $\mu, \sigma$ , リーフ  $i$  の平均と分散を  $\mu_i, \sigma_i$  とすると,

$$z = \frac{\mu - \mu_i}{\sqrt{\frac{\sigma^2}{N} + \frac{\sigma_i^2}{n_i}}} \quad (2.9)$$

となる。すなわち  $N$  が非常に大きいとき,  $z$  は簡単に 5%水準を超えてしまう。同様に, 両分布の違いを適合度検定に基づいて算出する場合は, 全データの速度 1km/h ごとの度数を  $N_i$  ( $i: 1 \sim r$ ), リーフの同度数を  $n_i$  とし,

$$p_i = \frac{N_i}{\sum_{i=1}^r N_i} \quad (2.10)$$

$$n = \sum_{i=1}^r n_i \quad (2.11)$$

とすると検定統計量  $\chi^2$  は

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} \quad (2.12)$$

となる。あるいはノンパラメトリック検定の一種である, コルモゴロフ・スミルノフ検定<sup>33)</sup>の検定統計量  $D$  は

$$D = \frac{Nn_i}{N + n_i} \sup |F_n(x) - F(x)| \quad (2.13)$$

である。ここで,  $F_n(x)$  はリーフの分布関数,  $F(x)$  は全データの分布関数を表す。上記のいずれの検定も, 統計検定量がサンプル数の影響を受けることに注意してほしい。ネイマン・ピアソン流の検定では,  $N$  や  $n_i$  が比較的小さな場合を想定して有意水準が定められている。したがって大きなサンプル数では, 有意水準を非常に小さくとらない限り, あらゆる分布間で有意差が検出されてしまう。

この問題に対してベイズ統計に基づく仮説検定を採用することもできるが, 妥当な事前分布を仮定した上で事後分布を算出する必要があるため, その設定が容易でない。また, あまり馴染みのない有意水準  $10^{-3}$  以下のオーダーを採用する積極的な理由がないので, この方法は恣意的である。また, 事前分布が非正規な場合は, 数値積分が必要となるため計算の簡便さが損なわれてしまう。

本研究では, サンプル数を検定統計量に含まない方法として, 2 群のカテゴリ変数の傾向性を検定するコ克蘭・アーミテージ検定<sup>34)</sup>を用いる。同検定は, 薬の投与量が異なる 2 群において投与量と治療の程度を表すカテゴリカル変数の分布間の有意差を検出する方法である。統計検定量は順序付きのカテゴリカル変数を  $v$ , 各カテゴリカル変数の該当サンプル数を  $n_k$  とすると,  $n_k$  を  $v$  に回帰した時の傾きに関数として得られる。なお  $k$  はカテゴリ数である。この検定は  $n_k$  と  $X_k$  に線形性が存在する場合のみ適用できる手法である。そのため, 例えば正規分布に従う 2 群のデータを確率変数  $X$  の任意の区間で分割して  $X$  を順序付きのカテゴリ変数とすると, 分布が単峰性を示すため適用できない。

そこで以下の分析では, 軽微な速度分布を検出するという目的を踏まえて, 全データのパーセンタイル値(以下, %ile)に基づいてサンプルを分割し, 順序付きのカテゴリ変数と, それぞれに対応するサンプル数を得る。これにより, 全データの期待分布は一様分布となる。

検定に当たって, カテゴリ変数  $k$  を設定しなくてはならない。ただし, 本データに特有の制約条件として, 速度データが整数値でしか得られないため, %ile カテゴリ間の閾値も整数で設定せざるを得ない。ここで分割をあまり細かくすると, 異なる %ile カテゴリが同整数の範囲内に収まってしまうという問題がある。一方で分割を粗くする

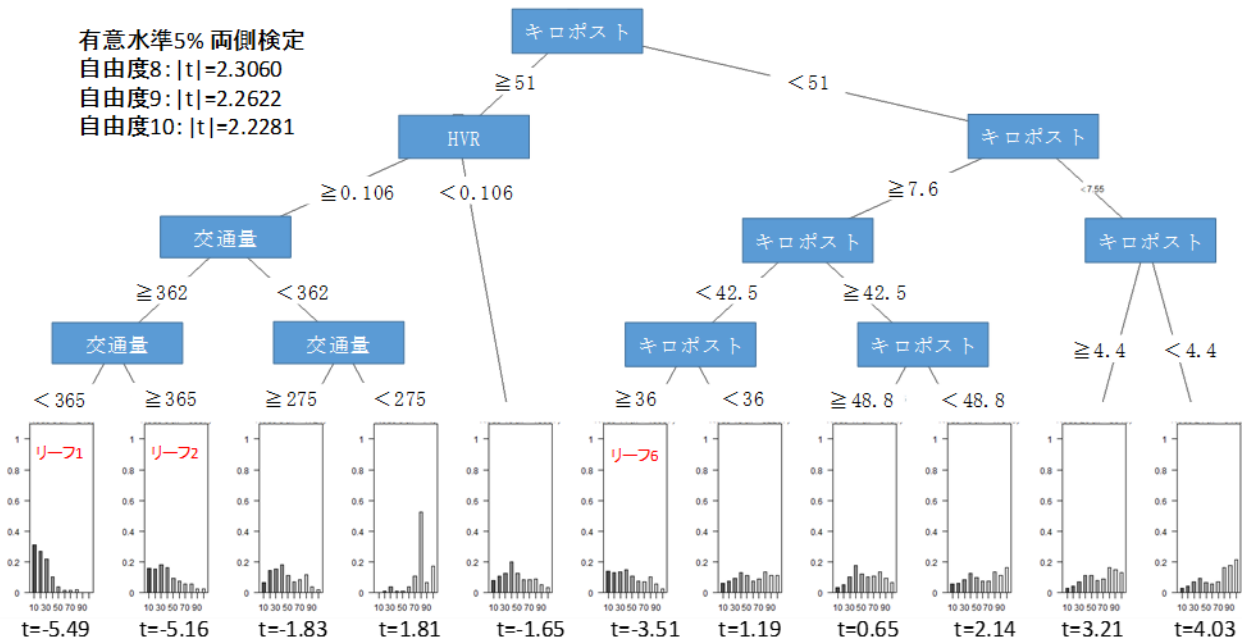


図 2.7 4月の決定木分析の結果

と、速度低下の検出能力が低くなる(たとえば 2 分割などを想定すると自明)。

これらの条件を踏まえて以下の分析では、 $k=10$  とし、%ile 値を用いて決定木を適用する。全データとリーフ  $i$  の偏差平方和を  $S_0, S_i$ 、両データを  $y$  軸,  $x$  軸を  $x=1\sim 10$  とした重回帰係数を  $R_0^2, R_i^2$  とすると、両データの分散はそれぞれ

$$s_0^2 = S_0(1 - R_0^2), \quad s_i^2 = S_i(1 - R_i^2) \quad (2.14)$$

と表せるため、両データの分散  $s$  は

$$s = \sqrt{\frac{s_0 + s_i}{N + n_i - 4}} \quad (2.15)$$

から求められる。上述より統計検定量は、全データの  $N_k$  の  $V_k$  に対する単回帰の傾きを  $b$ 、リーフの  $n_k$  に対する  $v_k$  に対する単回帰の傾きを  $\beta_k$  とおくと、

$$t = \frac{b - \beta_k}{s \sqrt{\frac{1}{N_k} + \frac{1}{n_k}}} \quad (2.16)$$

となる。この  $t$  値により、全データの分布 とリーフの分布に有意水準 5% で有意差があり、かつ  $t$  値が負の値(リーフの傾きが負)であれば、そのリーフに至る道に現れる条件の組み合わせが、速度低下をもたらすと判断できる。リーフの傾きは負であることを判定の条件とするのは 10%ile や 20%ile の割合のみが大きい渋滞現象ではなく、渋滞には至らないが、慢性的な微小な速度低下の検出に重点を置いているためである。また鳥取道は交通量が少なく渋滞が頻発していないことも考

慮したためである。ただし本体なら元の分布は一様分布に近いものとなるはずであるが、ETC2.0 の速度データは整数値であるため、完全な一様分布にはならない。しかし近似的な判定は可能と考えて、上述の手順による分析を行う。

## 2.8 決定木分析による速度低下区間の特定

決定木分析は複数の条件の組み合わせを議論できる。そのため、キロポストと共に検出された条件を考察することで、様々な対策を検討できる。一方で速度分布を一定間隔ごと(例えば 1km ごと)に切り口を変えて複数描く。それらと比較することで、特に速度が低い区間を見つけ、同区間への対策を検討する方法が考えられる。しかしこの方法では追い越し車線を設置することが優先されてしまうし、間隔の分け方によっては大変な労力を伴う。またキロポスト以外の条件を検出できないこの方法では、速度が通常よりも遅いのに、同区間で交通量が一定以上のデータに絞ると、速度が通常よりも速くなるといったパラドックスが生じる可能性がある。決定木分析ではキロポストと共に条件を抽出しているため、この現象は発生せず、また複数の図を描かずに 1 枚の図で理解できるため、実務において労力削減といったメリットが期待できる。

4 月の決定木分析の結果を図 2.7 に示す。式 (2.16) により算出された  $t$  値の絶対値が図 2.6 中に示す  $t$  値より大きく、かつ  $t$  値の符号が負のときに速度低下と判定する。



図 2.8 鳥取自動車道の速度低下区間

本分析ではカテゴリ数  $k$  を 10 としたため、自由度は 10、あるいは 9 と考えられるが、さらに小さい 8 を使う考え方もある。そのため、8~10 の 3 つの自由度で  $t$  値を比較する。しかし、図 2.6 からいずれのリーフも自由度の違いによって、帰無仮説の棄却判定に違いはなかった。

この検定により赤字で示した 3 つのリーフで、速度低下が発生していると判断できる。図 2.7 の結果は、例えば、リーフ 1 はキロポストが 51km 以上、HVR が 0.106 以上かつ交通量が 362 以上、364 未満の条件で速度低下が発生していると思われることができる。リーフ 6 のようにキロポストのみで分岐していれば、他の条件に関係なく速度低下しているため、その区間では慢性的に速度低下が発生していると判断できる。12 か月間、月別に決定木分析を行ったところ、各月とも共通して 2 つの区間(図 2.8 の矢印で示す区間)が速度低下区間として検出された。この 2 つの区間の各月の分岐条件を検討したところ、36~42km と 52~60km であることが判明した。

表 2.5 に、これらの 2 区間に関して月別に最も速度低下が顕著なリーフの分岐条件をまとめる。なお、表中にキロポストについて記入がない個所は、分岐条件がキロポスト 36~42km、あるいは 52~60km のみで構成されることを表している。表 2.5 から交通量、HVR、キロポストは分岐条件に頻繁に表れているが、天気、時間帯、平休はほ

表 2.5 月別の速度低下主要因

月	36~42km	52~60km
4		交通量 $\geq$ 363, HVR $\geq$ 0.106
5	交通量 $\geq$ 838, 夕方, 雨	
6		キロポスト $\geq$ 59, HVR $\geq$ 0.067
7		キロポスト $\geq$ 58.2, 昼・夕方
8		交通量 $\geq$ 705.5
9	交通量 $\geq$ 551	交通量 $\geq$ 869
10	交通量 $\geq$ 702	
11		キロポスト $\geq$ 58.2
12		夕方, 休日
1	HVR $\geq$ 0.057, 朝・夕方	
2	HVR $\geq$ 0.156, 雨	交通量 $\geq$ 267
3		キロポスト $\geq$ 58.2, 交通量 $\geq$ 588

とんど現れないことがわかる。つまり、交通流の特徴が速度低下に影響することがわかった。一方は智頭 IC の北側 36~42km の区間である。7 つの月でキロポストのみの分岐が見られ、その他の条件がない。つまり、慢性的に速度低下が発生している可能性が高い。速度低下要因が交通量 500 以上の月が 3 つあり、また出現数の少ない雨、夕方である月がそれぞれ 5 月と 1 月、5 月と 2 月にある。これらのことから、この区間では交通量が 500 台以上のとき、また雨天時、夕方に速度低下が起りやすい。

他方は河原 IC の北側 52km~60km である。この区間を指定するリーフには、キロポスト以外に HVR や交通量が含まれている月が多く、これらの条件が揃ったときに速度低下が起きる。分岐には 58.2km 以上が含まれていて、この区間で特に大きな速度低下が発生している。交通量は 36~42km よりも低い。時間帯は夕方が 7 月と 12 月にある。これらのことから、この区間では 36~42km 区間よりも少ない交通量でも、速度低下が生じやすいことがわかる。

鳥取道に限らず各地の路線へ本手法を用いることで速度低下区間と共にその原因を抽出できる。キロポストのみで分岐している場合、その区間には一般的な速度低下対策である、追い越し車線を設置する対策が適当であろう。しかし交通量や時間帯が原因であるならば、追い越し車線のような膨大なコストと時間のかかる対策だけでなく、料金施策や交通量規制、注意喚起といった対策を検討できる。ただし、本研究は単に速度低下区間を特定しているにすぎないので、その区間に対策することで、隣接する区間に影響を与えてしまうといった点は、別途対策後のデータを入手して、先述した手順に従って、検定する必要がある。

## 第3章 交通状態の質的指標と事故リスク

### 3.1 交通ビッグデータと事故リスク

国内の交通事故は、2016年には死者数3,904人、死傷者数は62万2,757人と12年連続で減少傾向にあるが、高速道路における交通事故は、増減を繰り返している。交通事故は直接的な人命損失ばかりではなく、経済損失、エネルギー損失や排気ガスによる大気汚染といった間接的な損失を社会に引き起こす。交通事故によるこれらの被害を軽減するためには、渋滞の交通対策をさらに進めて、動的に変わる交通状態を、新たな交通ビッグデータから抽出し、その指標と事故リスクの関係を明らかにする必要がある。

近年「交通ビッグデータ」が注目を浴びている<sup>34)</sup>。ビッグデータは従来のデータと比べて情報量が多いばかりでなく、その種類が多様で、更新頻度が高いという特徴がある。交通事故研究や事故対策の観点からは、新旧の交通データ、つまり、事故データに交通分野のビッグデータを組み合わせることで、新たな分析につながる可能性がある。交通分野で代表的なビッグデータであるトラフィックカウンター（以下、トラカン）データは、交通量の変動特性を分析する研究<sup>35)</sup>などに用いられてきた。さらに近年では、2011年より本格導入されたETC2.0データが注目されている。ETC2.0対応車載機搭載車両からは、GPSによる車両の走行経路や瞬間速度などの詳細な走行履歴（プローブ情報）が得られるようになった。その取得間隔は100m、または200mである<sup>36)</sup>。この取得間隔は、出入のインターチェンジしかわからない従来のETCや、路側の固定点でしか情報が得られないトラカンよりもはるかに短い。また既往研究<sup>38)</sup>では、ETC2.0は他のプローブ情報のように車種や用途の限定がないという利点が指摘されている。その一方で、2018年3月末時点で、ETC2.0の搭載率は全国で約3.19%と低いため、データ活用に際して分析者の工夫が求められる。

以上の動向をふまえると、ETC2.0対応車載機搭載車数の乏しさには留意する必要があるものの、ETC2.0の情報を活用することによって、従来のトラカンデータよりも、時間や場所に関してより詳細に交通状態を把握できる可能性がある。本研究では、複数の統計データを併用して、交通状況の質的指標

に関する指標を作成して、交通状態と事故リスクの関係について分析を行う。具体的には、事故データ、トラカンデータ、ETC2.0データ、道路線形データの4種類の統計データを用いて、高速道路上の区間ごとの特徴を表す変数を作成し、事故の発生に結びつく交通状態を明らかにする。また事故データから、事故規模の指標として事故台数に着目して、事故規模と交通状態の質的指標の関係を明らかにする。

事故分析に関する既往研究では、事故発生に関して走行台キロベースの事故率が論じられることが多かった。その一方で、交通ビッグデータから得られる交通状態の質的指標と、交通事故発生との関係が明らかにされることは、ほとんどなかった。走行台キロが交通事故発生に直結することは疑いないが、他方で交通事故をもたらす交通流の状況やその特性が明らかとなれば、交通状態のモニタリングによる事故対策の立案など、新たな交通事故対策の可能性を探ることができる。以上の観点から本研究では、ETC2.0データから得られる交通状態の質的指標と、トラカンデータから得られる走行台キロ情報を組み合わせて、交通事故発生前の交通状況の中で、事故発生に結びつく状況の抽出を試みる。さらに発生した交通事故の規模を表す変数として、1事故当たりに関係する台数に着目して、より深刻な事故に結び付きやすい交通状態を明らかにする。

### 3.2 分析手法

事故発生に関して、走行台キロベースの発生確率そのものをモデル化すると、発生確率として極めて小さな値を示す稀事象の分析を行うことになり、統計的に有意な結果が得られない可能性が高い。そこで本研究では交通事故発生に関して、比較的長期間にわたって観測した交通状態に基づいて、その質的状態を表す指標を作成したうえで、同期間内における事故発生の有無を目的変数とするロジスティック回帰モデル<sup>39)</sup>を用いる。変数 $X$ は潜在的な事故要因への曝露の有無を表す（例えば $X=1$ の場合は曝露群、 $X=0$ の場合は非曝露群）。

上述したようにモデルの目的変数は、事故発生の有無とするが、空間方向の分割を一定値として、サンプルを得る。なおここで高速道路区間を一定間隔で分割することにより、区間長（走行キロ）が異なる

ることによる事故発生への影響を除く。ただし走行台数については、トラカンデータを用いる。またこの均等な区間分割と対応するように ETC2.0 データの空間集計単位をとることにより、データ加工の妥当性を担保する。事故発生に関する尤度関数  $L$  は、以下のように得られる。

$$R(x_i) = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \quad (3.1)$$

$$L = \prod_{i=1}^N \left\{ \left( R(x_i)^{n_i} \right) \times \left( 1 - R(x_i)^{1-y_i} \right) \right\} \quad (3.2)$$

$$y_i = \begin{cases} 1 & n_i \geq 1 \\ 0 & n_i = 0 \end{cases} \quad (3.3)$$

ここで  $\alpha$  は定数項、 $\beta$  はパラメータ、 $y_i$  は区間  $i$  の事故の有無、 $n_i$  は区間  $i$  の事故発生件数である。式 (3.2) が示すように、同一区間で複数回事故が発生した場合でも、交通状態の質的指標は同一値をとるため、区間別発生頻度については、重みを与えることによって、その違いを考慮する。

事故台数の分析には、オーダードプロビットモデル<sup>40</sup>を用いる。オーダードプロビットモデルは、順序付きの質的変数を目的変数とする統計分析に用いられる。オーダードプロビットモデルは、説明変数とパラメータで定義した関数に加えて、カテゴリ間の閾値パラメータを推定することにより、カテゴリ間隔が一定ではない目的変数の段階に影響を及ぼす要因を明らかにする。オーダードプロビットモデルのカテゴリ一選択確率  $P_j$  は式(3.4)、確定項  $V_j$  は式(3.5)、対数尤度関数  $\ln L$  は式(3.6)となる。

$$P_j = \phi(\theta_k - V_j) - \phi(\theta_{k-1} - V_j) \quad (3.4)$$

$$V_j = \sum_i \beta_i x_{ij} \quad (3.5)$$

$$\ln L = \sum_j \sum_k A_{jk} \left( \phi(\theta_{k-1} - V_j) \right) \quad (3.6)$$

ここで  $\phi$  は標準正規分布の累積分布関数、 $\theta_k$  はカテゴリ間の閾値、 $\beta_i$  はパラメータ、 $x_{ij}$  は説明変数、 $A_{jk}$  は観測カテゴリーが  $k$  のとき 1 となるダミー変数である。

### 3.3 データの加工手順

#### 1) 速度平均効果量と分散変化量

事故発生に関係する交通状態として、速度平均効果量と分散変化量に着目する。効果量<sup>41</sup>とは、2 群間の違いを定量化した指標である。Cohen<sup>42</sup>は効果量の目安を 0.2 : 小, 0.5 : 中, 0.8 : 大とした。本研究では、速度平均効果量を、連続する 2 区間 (当該区間 (1km) とその上流区間 (1km)) の速度平均の変化を表す指標として定義し、式(3.7)で算出する。

$$\delta_i = \frac{|\mu_i - \mu_{i-1}|}{S_p} \quad (3.7)$$

ここで、 $i$  は区間を示しており、 $\delta_i$  は効果量、 $\mu_i$  は平均値である。 $S_p$  は  $i$  とその直前の  $i-1$  区間でプールした標準偏差であり、式(3.8)により算出する。

$$S_p = \sqrt{\frac{n_i S_i^2 + n_{i-1} S_{i-1}^2}{n_i + n_{i-1}}} \quad (3.8)$$

ここで、 $n$  はサンプル数、 $S$  は標準偏差を示す。また、分散変化量 $v$  は、連続する 2 区間の分散の変化を表す。 $v$  は式(3.9)で算出する。

$$v_i = |S_i^2 - S_{i-1}^2| \quad (3.9)$$

#### 2) 速度分布の乖離度

事故台数に関係する交通状態として、連続する区間間の速度分布の乖離度に着目する。乖離度が値が大きいほど、連続する 2 区間の速度分布が乖離していることを示す。分布の乖離度は、L1 ノルムと L2 ノルムで算出する。L1 ノルムを式(3.10)で、L2 ノルムを式(3.11)で、それぞれ示す。また、2 つの速度分布の乖離度算出時の領域を図 3.1 に示す。

$$L1_i = \sum_{j=1}^{180} |p_i(j) - q_i(j)| \quad (3.10)$$

$$L2_i = \sum_{j=1}^{180} \{p_i(j) - q_i(j)\}^2 \quad (3.11)$$

ここで、 $i$  ( $=1, 2, \dots, N$ ) は事故を、 $j$  ( $=1, 2, \dots, 180$ ) は速度をそれぞれ示す。また、 $p_i(j)$  は事

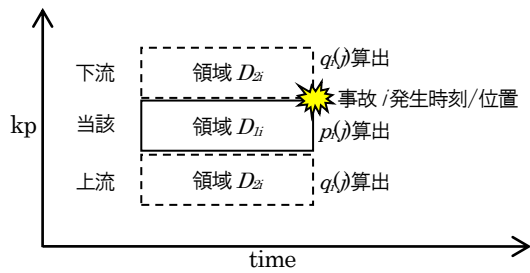


図 3.1 乖離度算出時の領域

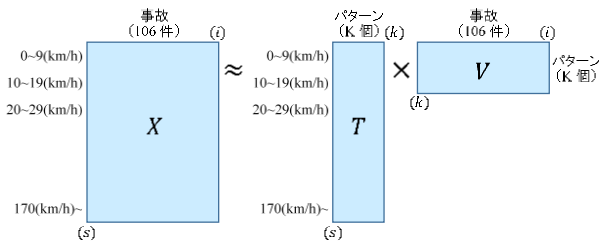


図 3.2 NMF のイメージ

故  $i$  当該区間 (領域  $D_{1i}$ ) の速度  $j$  の確率を,  $q_i(j)$  は事故  $i$  当該区間の下流区間もしくは上流区間 (領域  $D_{2i}$ ) の速度  $j$  の確率を, それぞれ示す. なお図 3.1 に示すように, 交通量の算出範囲は, 全て事故以前の時点とした. またここで下流とは, 車両の進行方向を前方とすると事故発生地点の前方を表し, 上流はその後方を表す. 分析に際して, 下流に影響を及ぼすと考えられる交通流の質的な違いをもたらす上流側の特徴だけでなく, 下流側の質的な違いにも着目する理由は, 前方の交通流の乱れが上流側に伝播する可能性を考慮するためである.

### 3) 速度分布パターン

事故台数に関係する説明変数として, 事故発生時の速度分布に関するパターンに着目する. 速度分布パターンは, 図 3.1 に示す領域  $D_{1i}$  内の速度分布から, NMF<sup>43)</sup> を用いて分類する. NMF は, 非負値行列から統計的に特徴のあるパターンを抽出する多変量解析手法の 1 つである. NMF のイメージを図 3.2 に示す.  $X$  は入力データ行列,  $T$  はパターンの特徴を示す特徴行列,  $V$  はパターンの重みを示す重み行列である. なお, 速度は, ETC2.0 の速度を 10km/h ずつ丸めた速度帯とした.  $X_{si}$  は, 領域  $D_{1i}$  内の速度帯  $s$  のデータ件数を示す. 重み行列  $V$  を事故の深刻度要因モデルの説明変数として採用する. NMF では, 分析者がパターン数を設定する必要がある. 本研究では, パターンの安定性の尺度として Brunetrt ら<sup>44)</sup> が提案した共役相関係数

(cophenetic) が減少し始める  $k$  や, Hutchins ら<sup>45)</sup> が提案した観測行列  $X$  と推定行列  $X'$  (特徴行列  $T$  と重み行列  $V$  の積) との残差平方和 (rss) が変曲点を示す最小値  $k$  を参照して, 最適な値を定める.

### 4) 横断線形と縦断線形

事故に影響すると思われる構造的な要因として, 道路線形を示す横断線形と縦断線形を用いる. 横断線形  $s_i$  は, 起終点間の直線距離  $l_i$  と道路線形に沿う実距離  $r_i$  の比であり, カーブの少ない道路線形ほど 1 に近い値をとる. 縦断線形  $g_i$  は, 起終点間の垂直距離  $v_i$  と水平距離  $h_i$  の比であり, 正の値は上り坂, 負の値は下り坂を示す. ここで, 水平距離  $h$  は, 分析時の区間幅とする. 横断線形  $s_i$  を式(3.12)に, 縦断線形  $g_i$  を式(3.13)に, それぞれ示す.

$$s_i = \frac{r_i}{l_i} \quad (3.12)$$

$$g_i = \frac{v_i}{h_i} \times 100 \quad (3.13)$$

### 5). 事故台数の定義

事故台数の分析では, 序数型の変数を用いる. まず, 事故深刻度を第 1 当事者対象物の自動車台数により 3 段階のカテゴリーで定義した (表 3.1). 同表より, 段階 1 が最も事故台数が少なく, 段階 3 が最も被害が大きいことを表す. そのうえで本分析では, 事故発生時に現場周辺にいた周辺車両のデータを抽出することにより, 事故につながりやすい交通状態を抽出する. たとえば, ETC2.0 データに記録された速度や車種などの特性が事故深刻度に及ぼす影響を分析する.

周辺車両の抽出条件の設定には, 以下のトレードオフがある. すなわちより事故発生時の交通流の状況を正確に捉えるには, 抽出条件を厳しく (短時間または狭い区間に) しなくてはならないが, 一方で特に ETC2.0 データの普及率が低い地方では, 時間や空間を短く, 狭く設定すると, 十分なデータ数が得られない, という問題が起こる.

そこでこれらの設定条件について, いくつかの組み合わせを算出して検討を行った. 区間幅を 400m とし, 時間幅を 30 分, 1 時間, 1 時間 30 分, 2 時間の条件下で, 周辺車両が 10 台以上確保できた事故件数を比較した. 表 3.2 に比較結果を示す. 同表より, 周辺車両を 10 台以上確保できた事故件数が 100 件を超える時間幅の 2 時間を, 抽出条件として採用した.

表 3.1 事故台数とカテゴリ定義

	第1当事者対象物の自動車台数		
	0台 <sup>※1</sup>	1台	2台以上 <sup>※2</sup>
分類	深刻度1	深刻度2	深刻度3

※1: 第1当事者対象物が自動車以外の単独事故

※2: 多重事故は先頭事故のみを対象

表 3.2 条件別周辺車両 10 台以上となる事故件数

時間幅	周辺車両が 10 台以上となる事故件数
30分	10
1時間	38
1時間30分	74
2時間	106

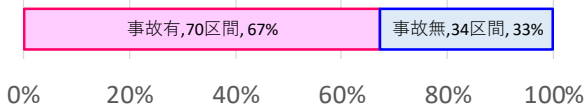


図 3.3 対象区間の事故発生の有無割合

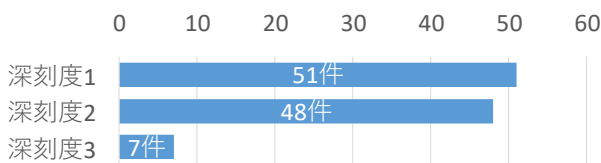


図 3.4 カテゴリ別事故件数

### 6) 分析対象データ

本研究では、NEXCO 西日本中国支社管轄エリア内の山陽自動車道入り本線の廿日市 JCT～笠岡 IC の約 130km を対象に分析を行った。また、対象期間は 2015 年 4 月～2015 年 12 月の 9 か月間とした。

事故データには、NEXCO 西日本中国支社管内の高速道路上で発生した事故の発生日時、発生場所、事故時の周辺状況など様々な項目が記録されている。

既往研究<sup>10)</sup>では、事故に影響する要因は、自由流と渋滞流でそれぞれ異なると報告されている。事故データに記録された“交通渋滞”を参照して、渋滞なしの事故のみを抽出したところ、322 件が得られた。これは、分析対象期間区間の全事故件数 351 件の約 9 割を占めており、渋滞流よりも自由流で発生した事故のほうが圧倒的に多いことがわかる。以下では、事故件数が多かった自由流の事故に着目する。

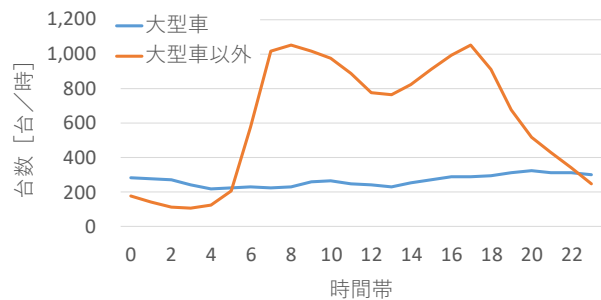


図 3.5 車種別交通量

ETC2.0 データは、利用者が少ない夜間や GPS 電波が受信できないトンネル内の取得率が低い可能性がある。そこで、これらのデータが入手できる昼間（7 時～19 時）、かつトンネル外で発生した事故を対象とした。さらに、事故データには、先行車両の事故に引き続いて発生する後続事故（2 台以上の車両による多重事故等）も記録されている。こちらも自由流の事故とは発生状況が異なるため、後続事故と考えられる事故データを除去した。後続事故と扱う時間帯ウィンドウは、全事故データからその発生傾向を確認して、30 分以内かつ 0.5km 以内とした。事故発生分析では、上記条件を満たす 150 件の事故を扱う。一方事故台数の分析では、周辺車両が 10 台以上確保できた 106 件の事故を対象に分析を行う。

対象区間の事故発生の有無割合を図 3.3 に示す。分析対象区間を 1km で区切り、トンネル区間を除くと、104 区間となった。このうち事故が発生した区間は 70 区間であった。事故規模別の事故件数を図 3.4 に示す。深刻度 1=単語区事故が 51 件、深刻度 2=車 2 台が関係する事故が 48 件、深刻度 3=車 3 台以上が関係する事故が 7 件となった。すなわち、事故台数の多い事故は少ない傾向にある。

分析対象区間に設置された全 16 台のトラカンデータを用いる。トラカンデータには、日時、位置、交通量（方向別、車種別）が記録される。本研究で用いたトラカンデータは 1 時間単位で集計されている。つまり、各交通量の単位は [台/時] である。

対象トラカン全 16 台の 2015 年 4 月～12 月の時間帯別平均交通量を図 3.5 に示す。時間帯により、大型車以外の交通量の変化は大きい。大型車の交通量の変化は小さい。特に後者の通行については時間帯により、交通状況が変化していることが分かる。

ETC2.0 データには 13 種の様式が存在する。本研究では、これらのうち 200m 間隔（機器によっては 100m 間隔）で速度などを記録した様式 1-2 と、加速度が閾値を超えるとのみ加速度などを記録した様式 1-4 を使用する。ETC2.0 データの種別と用途により、本研究で定義した車種を表 3.3 に、2015 年

表 3.3 本研究の車種分類

		用途					
		未使用	乗用	貨物	特殊	乗合	未定義
種別	普通	—	普通車	大型車	大型車	普通車	—
	小型	—	小型車	小型車	小型車	—	—
	軽自動車	—	軽自動車	軽自動車	軽自動車	—	—
	未定義	—	—	—	—	—	—

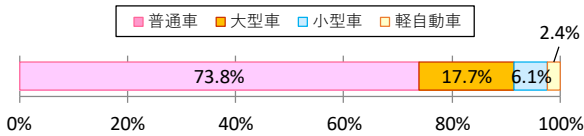


図 3.6 車種別データ割合

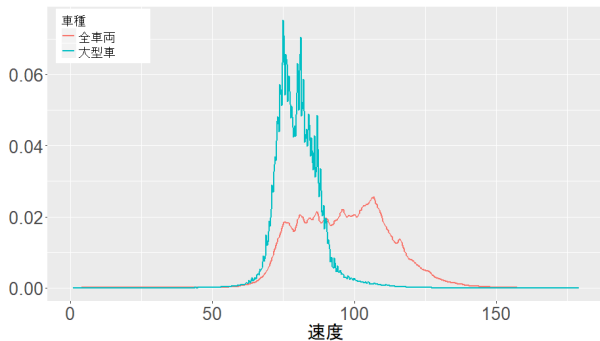


図 3.7 2015年4月～12月の車種別速度分布

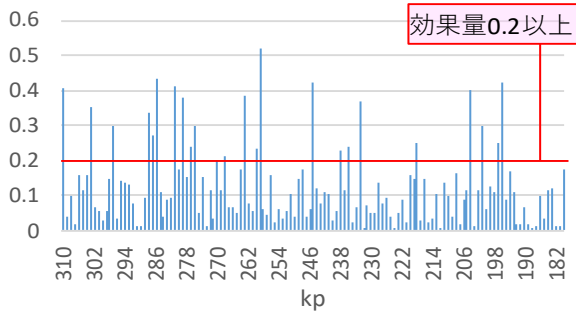


図 3.8 区間別速度平均効果量

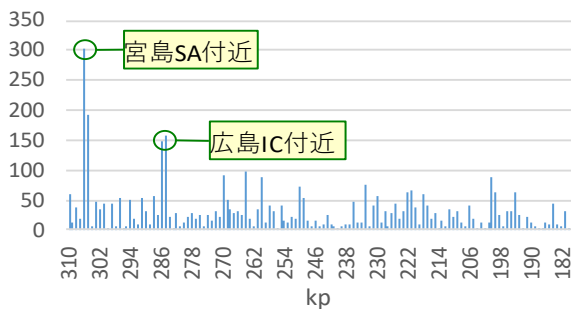


図 3.9 区間別分散変化量

4月～12月の車種別のデータ割合を図3.6に、それぞれ示す。図3.6より小型車、軽自動車はそれぞれ6.1%、2.4%に留まる一方で、大型車は17.7%を占めている。2015年4月～12月の車種別の速度分布を図3.7に示す。同図より、大型車と全車両の速度分布に違いがみられる。これは、大型車は運行管理を目的<sup>23)</sup>として、デジタルタコグラフを搭載しているためと考えられる。デジタルタコグラフは、車両の運行速度・時間などを自動的に記録する運行記録装置であり、運行会社はこの記録に基づいてドライバーの法定速度、休憩時間などの遵守を確認する。貨物輸送やバス運行が大半を占める大型車は、時間通りに目的地へ到着する必要があるため、法定速度を超えないぎりぎりの速度を保ちながら走行する、と想定される。したがって、一般車と大型車を比較すると速度分布は異なると予想される。

式(3.5)より算出した連続2区間の速度平均効果量を図3.8に、式(3.7)より算出した連続2区間の速度分散変化量を図3.9に、それぞれ示す。図3.8より、効果量が0.2以上の区間は全体の18.3%であった。図3.9より、分散変化量は、宮島SAや広島IC付近で特に大きいことが分かる。この辺りは対象区間の中でも利用者が多く、混雑が発生しやすい。また、走行速度の高い本線車両と、走行速度が比較的低い宮島SAや広島ICを利用する車両が混在するためと考えられる。

NMFによるパターン数別の共役相関係数(cophenetic)と残差平方和(rss)を図3.10に示す。残差平方和(rss)の変曲点と、共役相関係数(cophenetic)の極大点を踏まえて、NMFのパターン数は、 $K=9$ とした。特徴行列 $T$ より、事故時の速度分布パターンの特徴を図3.5に示す。図3.5より、低速域にピークを持つパターンは、30km/h台に速度分布のピークを持つパターンと、60km/h台に速度分布のピークを持つパターンであった。その他は、80km/h台、90km/h台、100km/h台、110km/h台、120km/h台、130km/h台、130km/h台以上に、それぞれ速度分布のピークを持つパターンが得られた。以上より、交通事故発生時の速度分布は交通状況に応じて様々なパターンが存在することわかる。

道路線形データには、10m間隔の緯度、経度、標高が記録されている。このデータに基づいて、横断線形 $s_i$ や縦断線形 $g_i$ を式(3-12)、(3-13)より算出した。

### 3.4 事故発生モデルの推定結果

事故発生モデルの推定結果を表-4に示す。モデル1は大型車の特徴を表す説明変数を含めて分析を行



表 3.4 事故発生モデルの推定結果

説明変数	モデル 1		モデル 2		モデル 3	
	Estimate	Std.E	Estimate	Std.E	Estimate	Std.E
交通量	-2.E-04	0.001	-3.E-04	0.001	—	—
大型車交通量	-0.011	0.011	—	—	—	—
速度平均効果量_0.2 以上ダミー	2.790 **	1.271	2.661 **	0.771	2.815 *	1.193
速度分散変化量	-0.029 **	0.010	-0.023 **	0.008	-0.025 **	0.009
速度分散	0.005	0.009	4.E-04	0.006	—	—
速度中央値	-0.277 **	0.127	-0.040	0.053	-0.166 *	0.085
大型車速度分散	0.012	0.017	—	—	—	—
大型車速度中央値	0.573 **	0.233	—	—	0.407 *	0.167
前後加速度件数	3.E-04	2.E-04	2.E-04	2.E-04	—	—
左右加速度件数	0.005 .	0.003	0.003 **	0.001	0.004 .	0.002
横断線形	2.626	0.002	1.229	5.373	—	—
縦断線形	0.171	5.956	0.055	0.214	—	—
定数項	-19.290	0.251	3.976	7.948	-14.608	8.891
初期尤度	-120.608		-120.608		-120.608	
最終尤度	-72.290		-77.514		-72.440	
AIC	170.579		175.028		156.881	
R <sup>2</sup> McFadden	0.401		0.357		0.399	
R <sup>2</sup> McFadden 自由度調整済み	0.301		0.282		0.358	
サンプル数	104		104		104	

\*\* : 1%有意 \* : 5%有意 . : 10%有意

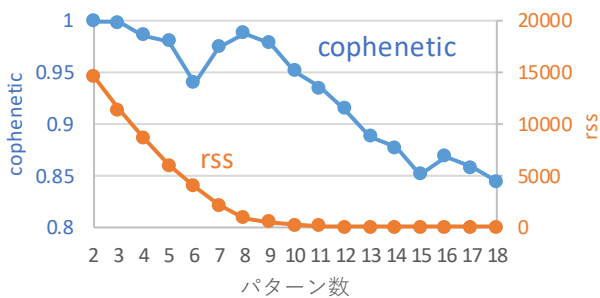


図 3.10 共役相関係数 (cophenetic) と残差平方和 (rss)

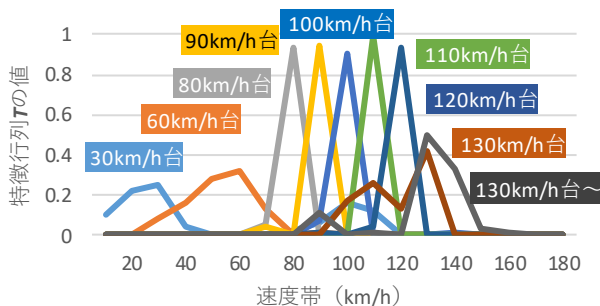


図 3.11 事故時の速度分布パターンの特徴

った結果である。自由度調整済みの疑似決定係数 (R<sup>2</sup>) は 0.301 を示した。全車両の連続 2 区間の速度平均効果量が 0.2 以上となった場合を 1 とするダミー変数、および大型車速度中央値は、それぞれ正で有意となった。一方で、全車両の連続 2 区間の速度分散変化量、ならびに速度中央値は、それぞれ負で有意となった。モデル 2 は大型車の特徴を表す説明変数を除いて、全車両の特徴を表す説明変数のみで分析を行った結果である。

モデル 2 の自由度調整済み R<sup>2</sup> は 0.282 を示しており、AIC を比較するとモデル 1 の当てはまりが良い。つまり、大型車の特徴を表す説明変数を含めて分析を行った方が、データへの当てはまりは良い。そこでモデル 1 の説明変数を用いて、ステップワイズ法により AIC が最小となる説明変数の組合せを選択し、分析を行った結果をモデル 3 に示す。自由度調整済み R<sup>2</sup> は 0.358 を示し、データへの当てはまりは良好である。モデル 3 では、モデル 1 と同様に全車両の連続 2 区間の速度平均効果量が 0.2 以上となった場合を 1 とするダミー変数、大型車速度中央値はそれぞれ正、全車両の連続 2 区間の速度分散変化量、速度中央値はそれぞれ負で有意となった。さら

に、左右加速度件数が正で有意となった。

AIC が最小のモデル 3 より、全車両の連続 2 区間の速度平均効果量が 0.2 以上となった場合を 1 とするダミー変数、全車両の連続 2 区間の速度分散変化量、速度中央値、大型車速度中央値、左右加速度件数の 5 変数が、事故の発生に影響を与えていることが明らかとなった。それぞれの説明変数の含意を考察する。全車両の連続 2 区間の速度平均効果量が 0.2 以上となる場合に事故発生が増加する状況は、上流区間 (1km) と対象区間 (1km) の速度平均の変化が大きいかほど事故は発生することを表す。これは、速度が変化することで車の密度が変化している。一方で全車両の連続 2 区間の速度分散変化量のパラメータは負の値となった。つまり、上流区間 (1km) と対象区間 (1km) の速度分散の変化が小さいほど事故が発生しやすくなる。

対象区間の上下流間で速度分散が変化しない状況には、車間距離が空いている状況と、車間距離が詰まった状態の 2 状態が考えられる。このパラメータが負で有意ということは、これら 2 状態のいずれかが上流～対象区間で類似している（継続している）状況で事故が発生しやすくなることを表している。速度中央値のパラメータは負の値となった。つまり、全体の速度が遅いほど事故が発生しやすくなる。この結果は、一般に速度が速いと車間距離が広くなり、目の前で起きる突発的な危険事象に対処できる余裕が生じるためと考えられる。大型車速度中央値のパラメータは正の値となり、全車両の速度中央値のパラメータと符号が逆となった。つまり、大型車の速度は速いほど事故は発生しやすくなる。これは、以下の状況をあらわすと考えられる。全車両の速度中央値が低くなり、大型車速度中央値が高くなると、すべての車両が同じような速度分布となる。しかし、大型車以外の車両は、より高い速度を目標として走行することが多いため、大型車を追い越そうとして、事故が発生しやすくなると思われる。左右加速度件数のパラメータは正の値となった。つまり、閾値を超える左右加速度が多く観測されるほど事故は発生する。なお左右加速度は閾値を超える件数で定義した点に留意する必要がある。この値は急な車線変更の多さを表していると考えられるので、交通流の横断方向の乱れが、事故発生を招くと思われる。

### 3.5 事故台数モデルの推定結果

式(3.2)による事故台数モデルの推定結果を表 3.5 に示す。各モデルは、ステップワイズ法により AIC が最小となる説明変数の組合せを選択した。説明変数として、ETC2.0 から得た交通状況を表す変数と

して速度と加速度に加えて、事故状況、交通量、および道路線形を加えてモデル推定を行った。

モデル 1 は、L1 ノルムの速度分布乖離度を説明変数に含めて分析を行った結果である。モデル 1 の自由度調整済み  $R^2$  は 0.095 を示した。また、速度分布の乖離度、速度パターン、前後加速度件数、ヨー角速度件数、ヨー角速度件数差分、時間帯ダミー、天候ダミー、第 1 当事者車種ダミー、交通量、道路線形、そして閾値が、それぞれ有意となった。モデル 2 は、L2 ノルムの速度分布乖離度を説明変数に含めて分析を行った結果である。自由度調整済み  $R^2$  は 0.079 を示した。モデル 2 では、モデル 1 と概ね同様の変数が有意となった。モデル 3 は、ETC2.0 の速度、加速度情報を除き、事故、交通量そして道路線形に関する情報のみ、説明変数に含めて分析を行った結果である。モデル 3 の自由度調整済み  $R^2$  は 0.022 に留まり、モデルの説明力は低い。

自由度調整済み  $R^2$  は、モデル 1 やモデル 2 の方がモデル 3 よりも高い。すなわち、ETC2.0 による速度や加速度を表す説明変数を含む方がデータへの当てはまりが良く、これらの情報を含まないモデル 3 の当てはまりは悪い。この結果は、事故台数が交通量や道路線形といった、交通状況を含まない変数のみではうまく説明できないこと、すなわち道路線形等だけでは説明できない事故台数に対する影響を、速度や加速度の情報によって有意に補えることを示している。

ここからは、データの当てはまりのよかったモデル 1 について、1%と 5%で有意となった変数が示す交通状況を、それぞれ考察する。速度に関する変数では、事故時の下流区間との速度分布乖離度パラメータが負の値となった一方で、通常時の速度分散は正の値となった。なお下流とは事故発生地点よりも前方を指している。また速度の集計対象時間帯は、事故に先立つ時間帯に限定しているため、事故そのものの影響が現れることはない。これらの定義を考え合わせると、通常時は下流区間との速度分布が乖離しているが、事故発生の直前では上下流間の速度分布の乖離が小さい区間で、事故台数が増える傾向にある。

信号や交差点のない高速道路では、そもそも速度分布乖離度が小さい区間が大半を占めることを念頭に考えると、渋滞などを原因とする速度低下が起こっている区間が下流側にあり（速度分布乖離度が大きい）、その速度低下が上流側に延びる状況となったとき（速度分布の乖離度が小さい）に起こる事故は、より事故台数が増えることを示している、と考えられる。

表 3.5 事故台数モデルの推定結果

説明変数				L1 ノルム		L2 ノルム		ETC2.0 データなし	
				Estimate	Std.E	Estimate	Std.E	Estimate	Std.E
速度	中央値	事故時	当該と下流			0.053 *	0.025		
	分散	通常時	当該			-0.007 .	0.004		
	分布乖離度	事故時	当該と下流	-1.942 *	0.875				
		通常時	当該と下流	7.096 *	3.123	821.649 *	356.235		
	速度パターン	事故時	60km/h 台	0.036 *	0.014	0.073 **	0.021		
			120km/h 台	0.086 **	0.026	0.112 **	0.032		
			130km/h 台			-0.060 .	0.032		
		130km/h 台以上	-0.062 **	0.020	-0.082 **	0.023			
加速度	件数_前後	通常時	当該	0.003 *	0.001	0.003 *	0.001		
	件数_左右	通常時	当該	-0.032	0.021	-0.049 .	0.027		
	件数差分_左右	通常時	当該-下流			-0.044	0.028		
	件数_ヨー角	通常時	当該	0.059 **	0.020	0.096 **	0.031		
	件数差分_ヨー角	事故時	当該-下流	1.394 **	0.521	2.176 **	0.619		
		通常時	当該-下流 上流-当該	-0.006 .	0.003	-0.010 **	0.003		
事故状況	7時~10時			0.971 **	0.372	1.028 *	0.408	0.561 .	0.063
	晴れ			0.633 *	0.284	0.483	0.305	0.279	0.245
	第1当事者車種_大型車(普通貨物)			0.736 *	0.300	0.790 *	0.320	0.587 *	0.030
交通量	交通量			-0.001 **	4.E-04	-0.002 **	5.E-04	-2.E-04	4.E-01
	大型車交通量			0.004 **	0.002	0.005 *	0.002	0.001	0.303
道路線形	横断線形			-84.384 .	45.923	-133.021 *	54.314	-29.796	0.473
	縦断線形			-0.428 **	0.124	-0.568 **	0.144	-0.236 *	2.E-02
閾値	1   2			-85.133 .	45.952	-131.425 *	54.758	-29.800	0.473
	2   3			-82.933 .	45.893	-129.042 *	54.676	-28.064	0.499
初期尤度				-94.362		-94.362		-94.362	
最終尤度				-68.358		-63.875		-85.253	
AIC				174.716		177.750		188.506	
R <sup>2</sup> McFadden				0.276		0.323		0.097	
R <sup>2</sup> McFadden	自由度調整済み			0.095		0.079		0.022	
サンプル数				106		106		106	

\*\* : 1%有意 \* : 5%有意 . : 10%有意

事故時の速度分布のパターンは、60km/h 台との120km/h 台の変数のパラメータが正の値となった。つまり、速度分布のピークが60km/h 台と120km/h 台のときに発生する事故台数は多い。一方、事故速度分布のパターンが130km/h 台以上の変数のパラメータは、負の値となった。速度分布のピークが130km/h 台以上では周辺に車両がない状況が多く、他車両を巻き込む事故にはなりにくいためと考えられる。

ETC2.0 データにおいて加速度に関する変数は、

その常時観測に基づいて、閾値以上の大きな値（異常値）が現れた場合のみ、データ上に残される。したがってこれらの件数が多いことは、加速度の乱れが多いことを表す。通常時の前後加速度の件数とヨー角速度の件数のパラメータは正の値となった。この結果は、普段から前後加速度やヨー角速度が多く記録される区間では事故台数が増えることを表しており、想定通りの結果である。また、事故時の下流区間とのヨー角速度の件数差分のパラメータは正の値となった。これは

下流区間と比較して、横断方向の加速度の乱れが多い区間では、事故台数が多くなることを表している。

事故状況に関する変数では、7時から10時を1としたダミー変数のパラメータは正の値となり、この時間帯に発生する事故台数は多い。これは、通勤ラッシュで交通量が多く、一部に急いでいる車が含まれるためと考えられる。天候が晴れのダミー変数のパラメータは正の値となり、晴れているときに発生する事故台数は大きい傾向となった。この傾向は、たとえば事故発生に関する潜在的ハザードを考慮すると、通常の設定に反する結果だが、以下の傾向を示していると考えられる。すなわち、雨など天候が良くない場合は視界不良となって運転が慎重になるため、いったん事故が発生しても事故台数が少なくなると思われる。第1当事者車種が大型車を1としたダミー変数のパラメータは正の値となった。これは、大型車が第1当事者車種の事故は衝突エネルギーが大きく、関連する台数の多い事故に発展しやすいためと考えられる。

交通量に関する変数に着目する。交通量のパラメータは負の値となり、大型車交通量は正の値となった。つまり交通量が少なく、大型車の割合が大きいほど、事故台数は多くなる。道路線形に関する変数では、縦断線形の変数のパラメータは負の値となり、下り坂で発生した事故は、事故台数は多いことが分かった。

# 第4章 観光地のトラベルコスト評価

## 4.1 研究の狙いと対象地域

観光地のトラベルコスト評価は、モバイル空間統計などの交通ビッグデータを利用して、入れ込み地ベースで出発地を考慮しつつ、地域魅力度を定量的に算出することを目的として実施している。このため、モバイル空間統計が入手できる2014年以降の最近で、高速道路アクセスが改善した地域を選定することが望ましい。そこで本研究では、大分市～宮崎市間の東九州道が2015年3月に開通し、また延岡市～高千穂町間の自動車専用道路も順次開通している点から、分析対象地域を宮崎県延岡市と高千穂町を対象地域として選定する(図4.1)。

観光施設の立地特性と日別・時間帯別の集客量を分析する。特に域内の高速道路からの交通アクセスと集客量に着目する。まず、使用したデータを示し、次に、集客量の変化を確認する。

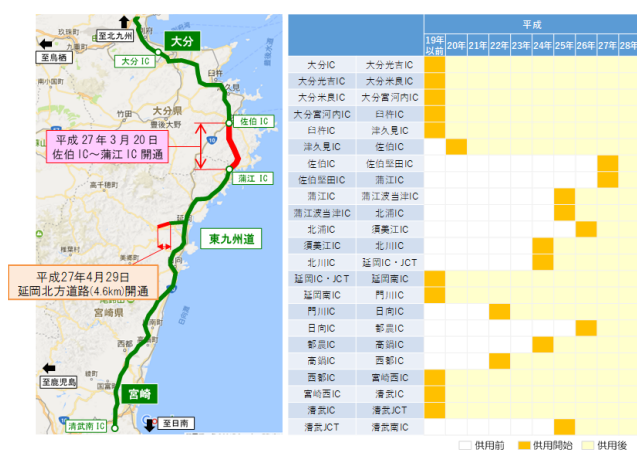


図 4.1 対象地域と東九州自動車道

## 4.2 データの概要

### 1) モバイル空間統計

モバイル空間統計<sup>11)</sup>とはドコモの携帯電話ネットワークの仕組みを利用した人口の統計情報で、日本全国において1時間毎の性別・年代別、居住地域別

の人口分布、また訪日外国人の場合は国別の人口分布を365日24時間把握することが可能である。このデータはプライバシー保護の観点から、携帯電話の持つ個人情報を除去する「非識別化処

表 4.1 データ概要

		開通前年	開通年	開通翌年
対象期間	繁忙期	2014年8月	2015年8月	2016年8月
	閑散期	2015年1月 13日～ 2月12日	2016年1月 12日～2月 11日	2017年1月 10日～2月9 日
出発地情報		市町村単位		
到着エリア		3次メッシュ単位		
総流動数		360,902,615		

理」を施し、性別や年代別などの属性別に携帯電話の台数を数えドコモの携帯電話の普及率を加味することで母集団への拡大を行う「集計処理」を行った後、再びプライバシー保護の観点から個人の特定を防ぐため、少人数エリアの数値を除去、もしくは丸めこみを行う「秘匿処理」の3段階を経て作成されている。

本研究で使用するデータ項目を表4.1に示す。夏季における期間を繁忙期、冬季を閑散期と置く。対象期間は、「開通前年」「開通年」「開通翌年」が対象となる3年間のデータを対象としている。データは前述した処理を経ており、1時間毎の滞在人口から1時間毎の人口流動が把握できる。なお、起点は携帯電話の契約者の市町村で記録されており、終点は当該時間における滞在場所を1kmメッシュ単位で記録したものとなっている。日別時間帯別に集計した結果を図4.2～図4.4に示す。

本研究で用いるデータの特性として来訪者が同一メッシュに数時間滞在する際には複数回移動したように記録されてしまう点、市町村境のメッシュに居住する住人が隣接する市町村に訪問していないにもかかわらず来訪者としてカウントされてしまう。また1時間以上の滞在において、複数回の移動があれば、ダブルカウントされる点にも注意が必要である。

モバイル空間統計データは、個人を特定できないようにするため、出発地は市町村単位、着地は3次メッシュ単位で公開されている。本研究はアクセス性と観光施設の関係を確認するものであるが、着地に注目することで3次メッシュ単位でより詳細に関係性を確認できる。そのため来訪者に着目した分析を行う。ただし、同データを用いて来訪者の分析を行うためには、記録されているデータの出発地の分布について、注意を払う必要がある。

図4.5はあるメッシュにおける流動について

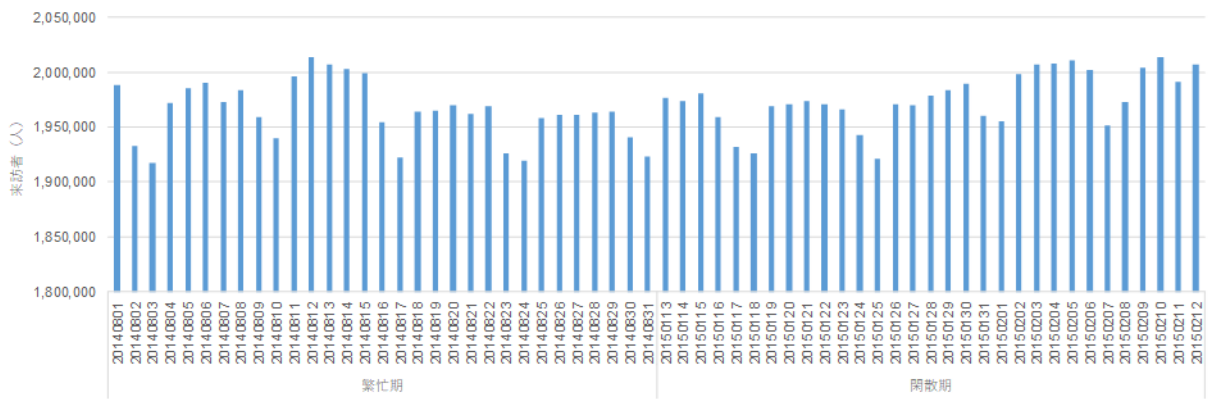


図 4.2 開通前年の日別来訪者

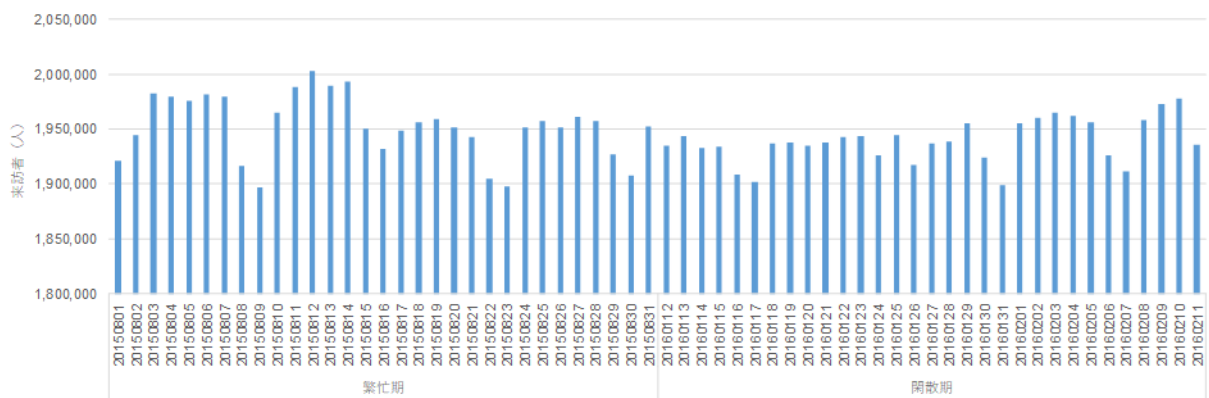


図 4.3 開通年の日別来訪者

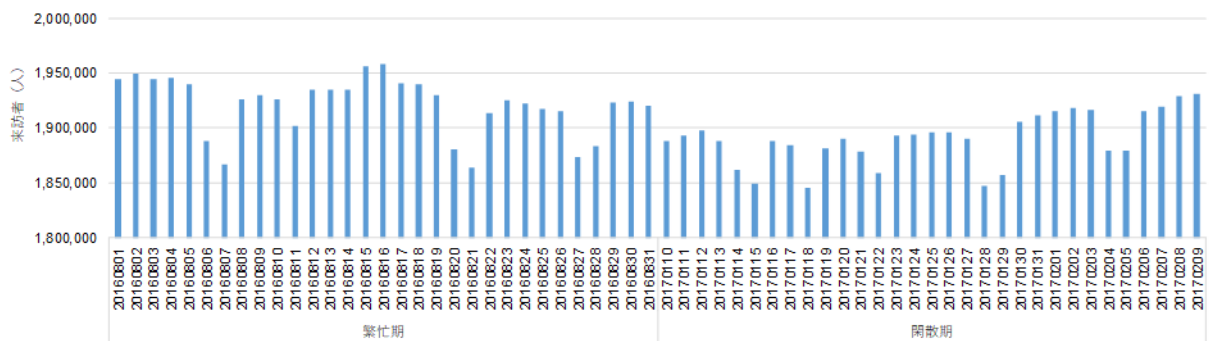


図 4.4 開通翌年の日別来訪者

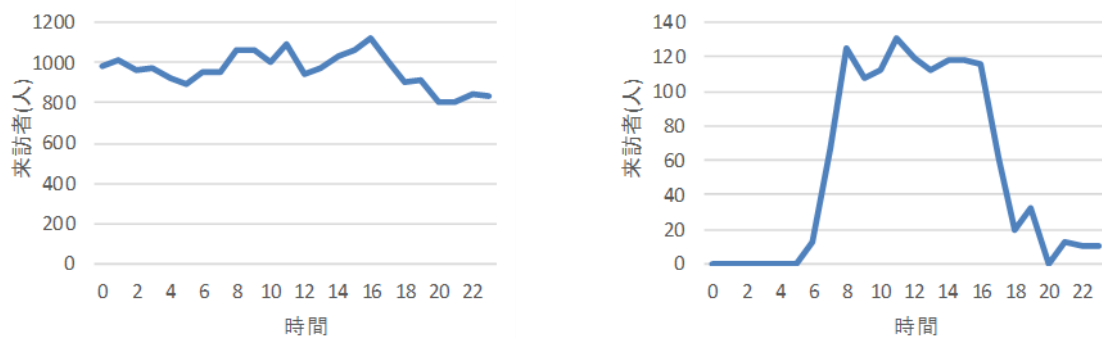


図 4.5 あるメッシュの来訪者数 (左：加工前，右，加

元データより作成した来訪者数の時間ごとの推移である。この図 4.4 の左では深夜時間帯にも関わらず 1,000 人前後の来訪者の存在が示されているが、このメッシュの宿泊施設ではこの人数を処理できないことがわかっているため居住者が来訪者として計測されている可能性がある。この原因として、隣接市町村の居住者が来訪者としてカウントされていることが考えられる。したがって、当該市町村の来訪者、複数の市町村を含むメッシュにおける隣接市町村からの流動を消去する処理を施す。同一のメッシュにおける処理後の来訪者の時間推移を図 4.4 の右に示す。これを見ると、深夜時間帯の来訪者数が大幅に減少しており、当該メッシュの来訪者数の実態におおむね近づいたと考えられる。

## 2) 交通サービス水準

対象地域内の 3 次メッシュレベルの交通サービス水準は、NITAS(National Integrated Transport Analysis System)<sup>46)</sup>はシステムによって算出した。NITAS は、道路、鉄道、航空、船舶の各交通機関を組み合わせて総合的に交通体系の分析を行うことができるシステムである。複数の起点から複数の終点間の経路探索が可能で、費用や所要時間が最小となる経路や分布を可視化することができる。全国から地方レベルの課題に対応して、交通ネットワークの現状や整備効果を視覚的・定量的に表現することが可能である。

全国幹線旅客純流動調査<sup>13)</sup>は、幹線交通機関における旅客流動の実態を定量的かつ網羅的に把握することを目的とした調査である。幹線交通機関の輸送実績を整理する他の流動調査とは異なり、個々の旅客に着目することでその旅行行動全体を捉えており、出発地・目的地、旅行目的や旅客属性が把握できることが特色である。起点終点が同一県内の移動については集計されていないため、注意が必要である。

本研究は高速道路の開通による流動の変化の分析を目的としたものであるが、モバイル空間統計はすべての移動手段を合計した流動である。そこで、2010 年に実施された幹線純旅客流動の調査における 207 生活圏間の交通手段別流動データから自動車利用比率を算出し、自動車による来訪者数を算出する。なお、幹線純旅客流動調査データには県内の流動が含まれていないため、県内の流動における自動車利用比率は次節で述べる集計ロジットモデルを構築し算出する。以下、来訪者数は自動車数による来訪者とする。

総旅行費用の算出にあたり必要となるメッシュ間の自動車による移動費用等は、NITAS において起点を各市町村の役所、終点を対象 3 次メッ

表 4.2 推定式の結果

係数	推定値	P 値
$\beta_1$	$-7.554 \times 10^{-5}$	0.0006
$\beta_2$	1.7643	$2.9661 \times 10^{-42}$
$\beta_3$	-0.95116	0.0057
調整済み R <sup>2</sup> 値 : 0.836		

シとして距離、所要時間、総費用、一般化費用の算出を行った。なお、四国からの移動はフェリーを手段とするものも多いと考えられるため、一般化費用の算出においてフェリーを考慮した。幹線純旅客流動調査データでは、起終点が同一県内の移動については集計されていない。そこで、県内間の移動における自動車の選択確率は、集計ロジットモデルを構築して算出することにした。また、東九州自動車道開通により目的地までのアクセス向上が考えられる。そこで本研究では総合交通分析システム(NITAS)を用い、新たな IC 開設に伴うアクセス時間の変化について全メッシュを対象に算出を行った。

モデルの推定には対象の地域が含まれる九州圏内の 2010 年における幹線純旅客流動調査データを用いた。九州内の流動のみを対象としたため、自動車、バス、鉄道の 3 肢選択問題としてモデルを構築した。

仮定した効用関数は以下のとおりである。表 4.2 に推定結果を示すが、旅行費用のパラメータが有意な負値をとっており、妥当なモデルであるといえる。また、自動車ダミーが有意な正値をとっており、交通機関は自動車が選択されやすいことがわかる。

なお、前節で述べたように県をまたぐ流動における自動車選択比率は、分析対象である東九州道開通前の 2010 年における幹線純旅客流動調査データを用いており、開通後の実態に即していない可能性がある。全流動に対して、東九州道開通後の一般化費用を推定した集計ロジットモデルの式(4.1)に代入することにより、自動車選択確率を更新することも考えられるが、以下では簡単のため、選択確率は更新前のデータとは大きく異ならないものとしてそのまま用いることとした。東九州道開通にあたる 2015 年の幹線純旅客流動調査データは現時点では公開されていないが、推定したモデルを用いることで、適切な自動車選択比率を算出できる。

$$P_M = \frac{\exp V_M}{\exp V_C + \exp V_B + \exp V_T}, \quad M \in \{C, B, T\} \quad (4.1)$$

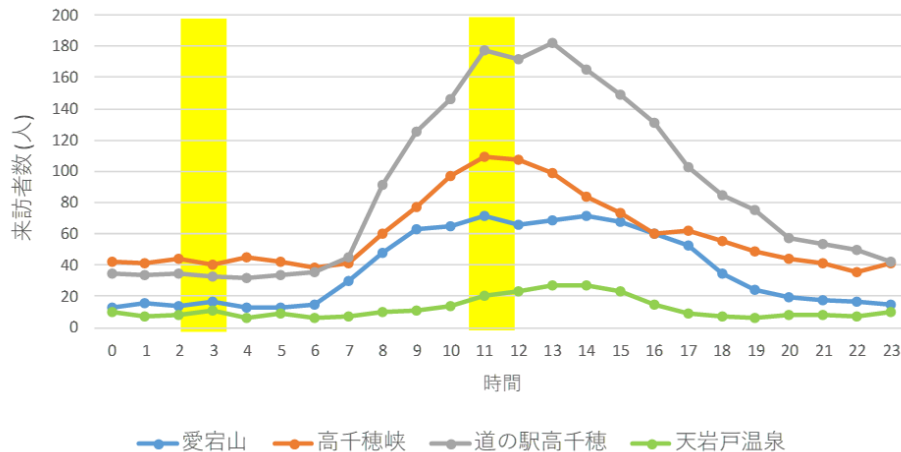


図 4.6 開通翌年の日別来訪者

$$\begin{cases} \text{自動車効用 } V_C = \beta_1 C_C + \beta_2 \\ \text{バス効用 } V_B = \beta_1 C_B + \beta_3 \\ \text{電車効用 } V_T = \beta_1 C_T \end{cases}$$

$C_C$ : 自動車使用時の一般化費用

$C_B$ : バス使用時の一般化費用

$C_T$ : 鉄道使用時の一般化費用

### 3) 施設立地関連データ

テレポイントデータ<sup>47)</sup>は電話帳データシステムに対応し、緯度経度のデータが追加されているデータベースであり、業種コードによって施設の種類が分類されている。本研究では高速道路開通により来訪者数の変化がある可能性のある 1)工業施設, 2)飲食店, 3)観光施設, 4)観光施設, 5)宿泊施設, 6)小売店の機関に施設を分類し、メッシュごとの施設数を算出した。また、同一施設であっても回線を複数持っている施設がある場合は、その中で代表となる施設のみ抽出し、ダブルカウントにならないように配慮した。

宿泊施設データ、および駐車場データは、テレポイントデータに記載されていない情報を補完する目的で収集した。宿泊施設データは施設名、住所等が記載されているほか、客室数、最大収容客数、駐車場データの情報が得られる。このデータにより、各メッシュの可能収容客数の把握ができる。駐車場データは時間貸し駐車場、ホテル駐車場、大型小売店駐車場、公営駐車場情報があり、これらの緯度経度データから三次メッシュへ変換した後、宮崎県の延岡・高千穂の対象エリアの情報を抽出した。

### 4) 時系列分析のためのデータ加工

モバイル空間統計から得られる着地別の滞在人口に関して、時系列分析によるモデル適合度の改善を図るために、各対象メッシュにおける集客施設数、曜日、アクセス時間、駐車場台数、宿泊施設詳細情報や、そして1日前および7日前の来訪者数のデータセットを用意する必要がある。以下の分析では、膨大なサンプル数を有するモバイル空間統計を、Visual Studio を用いて Python によるプログラムを構築した。具体的には、メッシュごとの1か月間の来訪者数、1日前及び1週間前の来訪者数、詳細情報が羅列されたものを延岡・高千穂に分割し、データセットを作成した。

#### 4.3 メッシュ別滞在者に関する時系列分析

高速道路開通による流動変化の時系列分析を行う。まず東九州自動車道開通前後のデータから来訪者数を単純集計したのち、対象のメッシュから無作為にいくつかのメッシュを抽出し、来訪者数の推移を把握する。次にGISを用い、施設数の増減を把握する。最後に、時系列変動や施設数、アクセス時間を考慮に入れた重回帰分析により、時間ごとの来訪者数の変動を捉える。

##### 1) 分析時刻の決定

時間ごとに来訪者数を算出することによって、累積人数を用いることによるダブルカウントを無くすることができる。今回の分析では昼夜間の来訪者の行動を分析するために、着目する時刻を検討する。全対象メッシュから観光地や重要拠点が存在するメッシュを無作為に4つ選び、一日の来訪者数の推移を把握する。4つのメッシュの1か月の来訪者数を平均し、時間ごとにプロットした結果を図4.6に示す。同図より、道の駅高千穂は13:00に再び来訪者数が増加しているものの、他のメッシュは11:00を境に



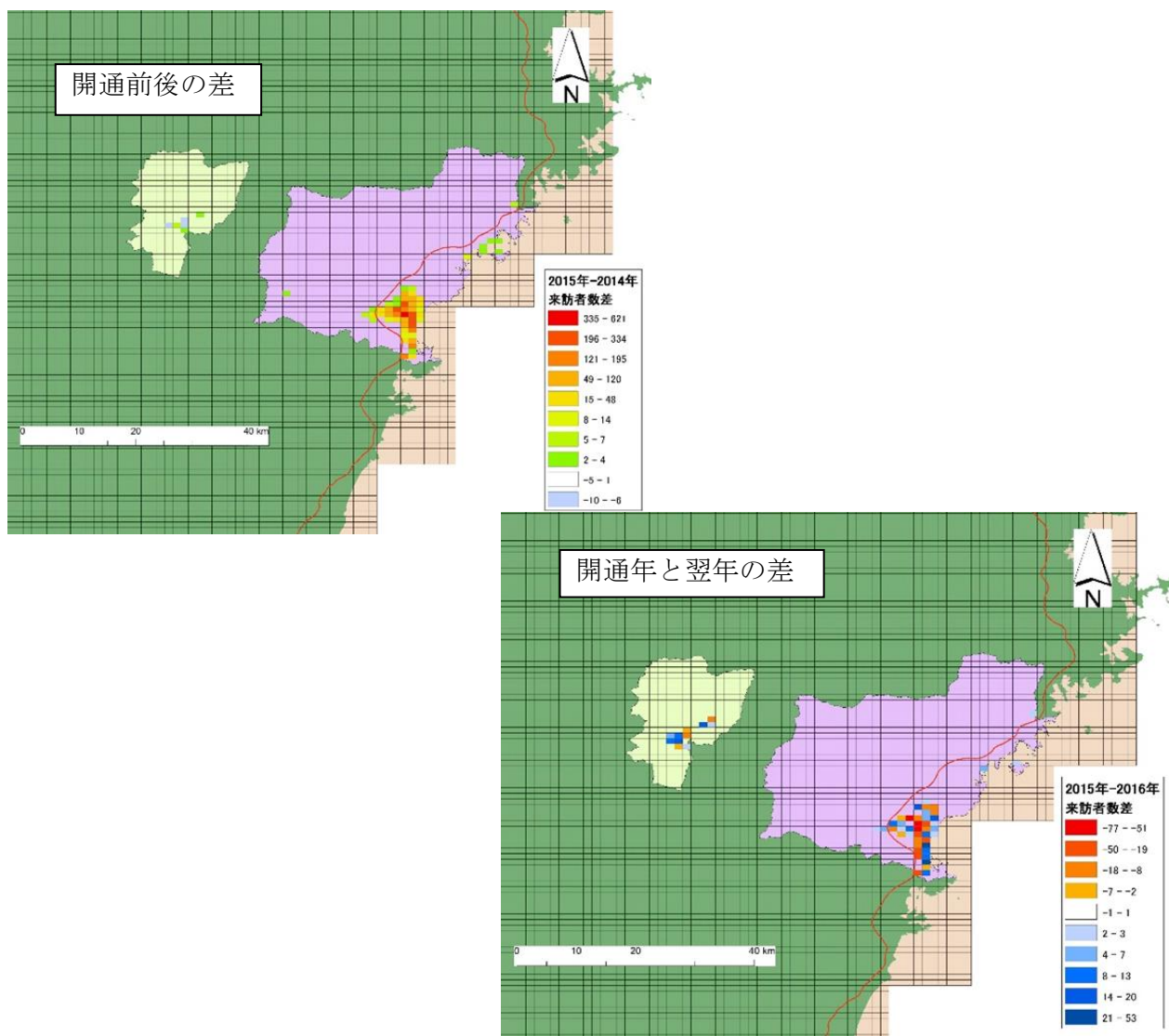


図 4.7 来訪者数の変化（左上：開通前後，右下：開通翌年—開通年）

減少傾向にある。これを考慮して、今回の研究では昼間の分析時間を 11:00 とし、夜間の分析時間を比較的流動が落ち着いている 2:00 と設定し、分析を行う。

## 2) 高速道路開通前後の来訪者数の分析

来訪者の行動に影響を及ぼす主要因として、高速道路の最寄り IC から周辺施設や観光地等とのアクセス性が挙げられる。本節では対象エリアでの高速道路開通前後の施設数や施設分類毎の変化を分析すると同時に、来訪者数の推移を把握する。

高速道路開通による来訪者数の差を分析する。図 4.7 の集計結果からは、地図内のエリアに居住地登録されているサンプルを除去しており、純粋に域外からの来訪者数を表している。また上述の検討結果に基づいて、一日の中で最も来訪者数の多い 11:00 のデータを用いている。来訪者の動向の推移が把握できるよう、左上に開通前後の、右下に開通翌年と

開通の来訪者数の差を、それぞれ示している。開通前後の来訪者の差より、全体的に来訪者は増加傾向にあり、特に延岡の中心市街地の増加が多いことがわかる。開通翌年と開通年では、開通前後と比べるとその差は小さくなっているが、依然として延岡の中心市街地は増加しており、東九州道開通による延岡へのアクセス向上によるものと推測される。一方で高速道路の IC から距離のある観光都市高千穂では、来訪者数はほとんど変化していない。

## 3) 高速道路開通前後の施設数の変化

本研究では、域外からの来訪者が訪問する可能性のある施設として、観光地や飲食店・宿泊施設だけでなく、百貨店等の小売店、延岡の中心産業である工業施設の 5 施設に着目して、分析を行った。施設はテレポイントデータに記載されている業種コードに基づいて分類をした。また、宿泊施設に関してはテレポイントデータに記載されていない施設があり、

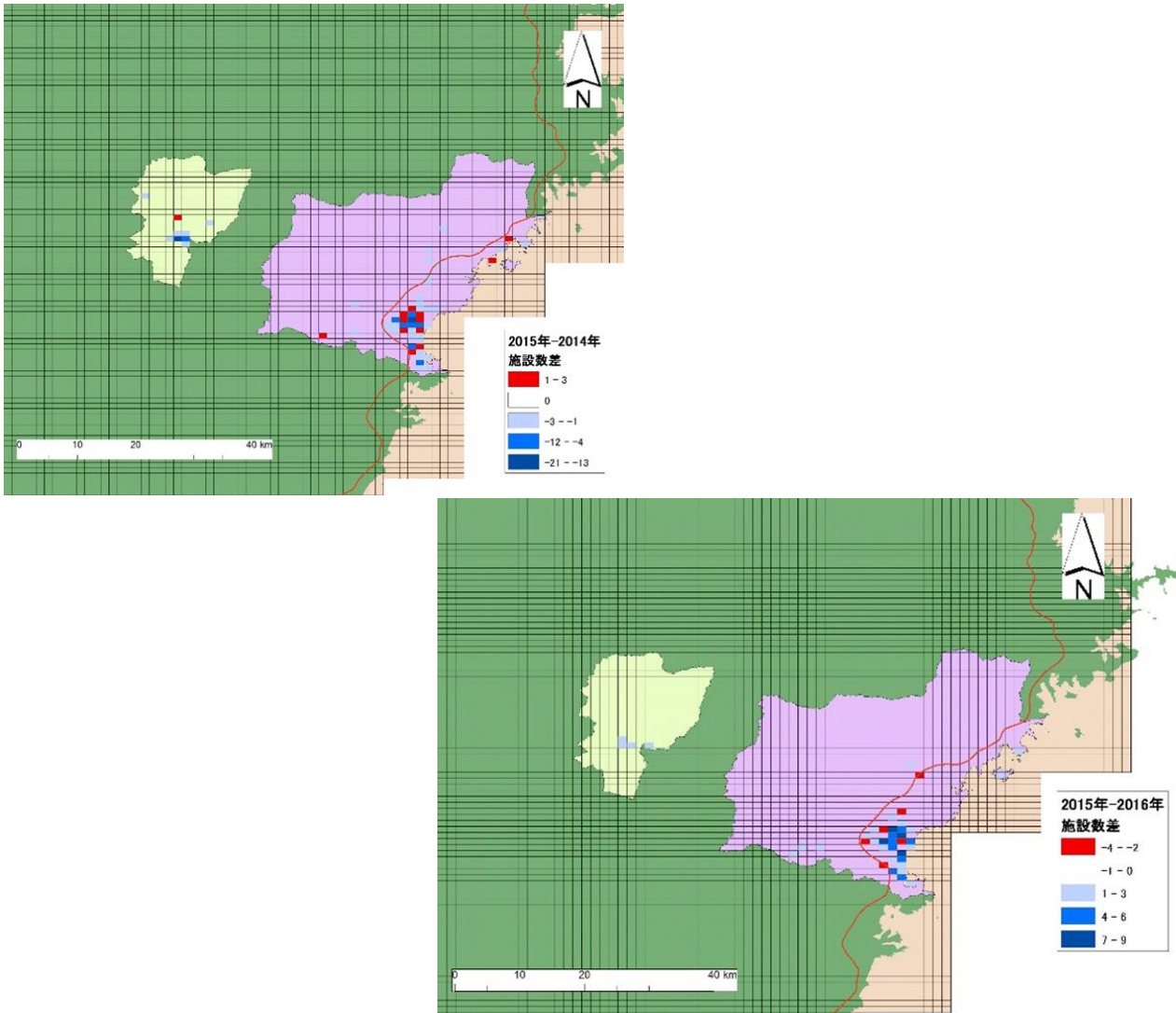


図 4.8 施設数の変化（左上：開通前後，右下：開通翌年—開通年）

表 4.3 業種別施設数変化

	施設分類				
	宿泊	工業	飲食	観光	小売
15-14	12	-17	-15	-1	-75
16-15	0	-21	-17	1	-39
合計	12	-38	-32	0	-114
14 比	15.0%	-5.5%	-5.2%	0%	-7.2%

表 4.4 延岡来訪者の重回帰分析の結果

	延岡 11:00						延岡 2:00					
	AR-1			AR-7			AR-1			AR-7		
	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率
(定数)	19.27		0.00	5.54		0.00	3.05		0.00	2.32		0.00
AR_Visitors	0.94	0.94	0.00	0.97	0.96	0.00	0.99	0.99	0.00	0.99	0.99	0.00
D_holiday	-16.96	-0.01	0.00	-7.06	0.00	0.00	-4.61	0.00	0.00			
宿泊	-3.83	-0.01	0.01	-3.47	-0.01	0.01						
工業										0.41	0.00	0.00
飲食店	-1.30	-0.03	0.00				-0.16	0.00	0.00			
観光地										-5.58	-0.01	0.00
小売店	2.74	0.07	0.00	1.63	0.04	0.00	0.27	0.01	0.00	0.22	0.01	0.00
駐車場合計										4.98	0.00	0.01
アクセス時間	-0.38	0.00	0.01									
客室数												
最大収容人数												
宿泊施設駐車場												
時間貸し駐車場台数				-9.20	-0.02	0.00						
ホテル駐車場台数												
大型小売店駐車場台数	14.98	0.00	0.00	20.49	0.01	0.00						
公営駐車場台数												
	R2 : 0.978			R2 : 0.987			R2 : 0.996			R2 : 0.994		

表 4.5 高千穂来訪者の重回帰分析の結果

	高千穂 11:00						高千穂 2:00					
	AR-1			AR-7			AR-1			AR-7		
	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率
(定数)	3.350		0.000	0.230		0.783	0.705		0.038	6.210		0.014
AR_Visitors	0.940	0.941	0.000	0.942	0.947	0.000	0.982	0.983	0.000	0.968	0.976	0.000
D_holiday	-5.029	-0.009	0.000	4.252	0.008	0.000				2.633	0.006	0.001
宿泊	3.894	0.032	0.004	4.059	0.033	0.003	1.601	0.015	0.000	2.161	0.020	0.000
工業	1.892	0.018	0.003									
飲食店												
観光地	3.843	0.011	0.001	2.592	0.007	0.032						
小売店	-1.177	-0.056	0.000	-0.872	-0.042	0.000						
駐車場合計												
アクセス時間										-0.147	-0.004	0.017
客室数	0.456	0.051	0.000	1.006	0.112	0.000						
最大収容人数												
宿泊施設駐車場				-0.510	-0.066	0.002						
時間貸し駐車場台数												
ホテル駐車場台数												
大型小売店駐車場台数												
公営駐車場台数												
	R2 : 0.982			R2 : 0.984			R2 : 0.993			R2 : 0.990		

その施設に関しては別途宿泊施設データの情報を追加し、集計した結果を図 4.8 に示す。同図において、左上は高速道路開通前後の施設数の差、右下は開通翌年と開通年の施設数の差であり、赤の実線は東九州自動車道を示している。同図より、高速道路が開通したにもかかわらず、全体的に施設数は減少傾向にあることがわかる。しかしながら延岡の市街地や高速道路沿線にある施設は微増していることが分かった。

次に、施設分類毎の開通前後の差を表 4.3 に示した。宿泊施設・観光地を除いて、全ての施設が減少していた。とりわけ、小売店の減少件数が他と比べて大きい。対 2014 年比では、他の施設とさほど変わりのない減少率を示している、つまり減少数が多いのは、そもそも 2014 年の小売店数が多いためと考えられる。

以上の結果より、開通前後で宿泊施設が増加し、他の施設が減少していることが把握できた。

#### 4) 来訪者数の空間分布の分析

本節ではメッシュ毎の来訪者数と集客に影響する要因との関係性を、時系列モデルにより分析をする。従属変数は 2014 年から 2016 年の 8 月の来訪者のデータである。説明変数は土日祝日の情報、来訪者に影響を与える可能性のある施設数、アクセス時間、駐車場件数の合計、宿泊施設の詳細データである客室数と最大客室数、時間貸し駐車場、ホテル駐車場、大型小売店駐車場、公営駐車場の各々の駐車場を合計した駐車場の台数とした。さらに、観測日の 1 日前、あるいは 7 日前の来訪者データを説明変数に加えた 2 種類のモデル(以下ではそれぞれ AR-1, AR-7 と記す)を構築した。過去の来訪者数を以上のように

表 4.6 延岡来訪者の重回帰分析の結果（居住者を含む）

	延岡 11:00						延岡 2:00					
	AR-1			AR-7			AR-1			AR-7		
	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率
(定数)	19.980		.000	5.802		.000	3.049		.000	2.325		.002
AR_Visitors	.942	.943	0.000	.967	.965	0.000	.993	.993	0.000	.989	.992	0.000
D_holiday	-16.955	-.009	.000	-7.033	-.004	.001	-4.609	-.003	.000			
宿泊	-3.903	-.007	.012	-3.008	-.005	.036						
工業										.409	.005	.001
飲食店	-1.320	-.027	.000				-.162	-.004	.001			
観光地										-5.577	-.008	.000
小売店	2.763	.073	.000	1.579	.042	.000	.275	.008	.000	.222	.006	.004
駐車場合計										4.979	.003	.007
アクセス時間	-.429	-.004	.005									
客室数												
最大収容人数												
宿泊施設駐車場												
時間貸し駐車場台数				-9.325	-.018	.000						
ホテル駐車場台数												
大型小売店駐車場台数				.014	.002	.053						
公営駐車場台数												
	R2 : 0.978			R2 : 0.987			R2 : 0.996			R2 : 0.994		

表 4.7 高千穂来訪者の重回帰分析の結果（居住者を含む）

	高千穂 11:00						高千穂 2:00					
	AR-1			AR-7			AR-1			AR-7		
	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率	非標準化係数	標準化係数	有意確率
(定数)	3.350		.000	.230		.783	.705		.038	6.210		.014
AR_Visitors	.940	.941	0.000	.942	.947	0.000	.982	.983	0.000	.968	.976	0.000
D_holiday	-5.029	-.009	.000	4.252	.008	.000	1.601	.015	.000	2.633	.006	.001
宿泊	3.894	.032	.004	4.059	.033	.003				2.161	.020	.000
工業	1.892	.015	.003									
飲食店												
観光地	3.843	.011	.001	2.592	.007	.032						
小売店	-1.177	-.056	.000	-.872	-.042	.000						
駐車場合計												
アクセス時間										-.147	-.004	.017
客室数	.456	.051	.000	1.006	.112	.000						
最大収容人数												
宿泊施設駐車場				-.510	-.066	.002						
時間貸し駐車場台数												
ホテル駐車場台数												
大型小売店駐車場台数												
公営駐車場台数												
	R2 : 0.982			R2 : 0.984			R2 : 0.993			R2 : 0.990		

に設定した理由は、日あるいは週単位で来訪者の流動傾向に周期性があることを確認するためである。モデルはそれぞれの都市における全メッシュの来訪者の時系列データを被説明変数として、ステップワイズ法を用いて推定を行った。

8月の11:00、2:00の延岡・高千穂の推定結果を表4.4～4.7に示す。表4.4、表4.5は、当該メッシュに居住しているデータを除去した結果を、表4.5、表4.6は、対象地の居住者を含む全てのデータを使用した結果である。総じてR2値が0.97を越え、モデルの適合度が非常に高い値を示した。推定された全ての標準化係数の値に注目すると、来訪者の行動には前日、あるいは1週間前の来訪者数の影響が大きいことが分かった。

表4.4に着目すると、AR-1、7ともに休日の非標

準化係数が、有意な負の値を示している。すなわち延岡では、平日よりも休日の来街者数が少ない傾向が確認できた。また小売店が有意な正の値をとっており、来訪者の目的地となっていることがわかった。11:00の結果では、AR-1、7ともに宿泊施設が有意な負値を取っていた。これは、延岡は日帰りあるいは通過する人が多いためと考えられる。2:00の結果ではAR-7で駐車場合計が有意な正値をとることが確認された。ただし、夜間の滞在施設として駐車場が現れる理由は不明である。一方で、観光地は有意な負の値をとり、夜間に来訪者が増加する施設ではないことが明らかとなった。

表4.5に着目すると、AR-1、7ともに宿泊施設が有意な正値をとることが分かった。また、昼間11:00の推定結果の中で観光地を見ると、AR-1、7ともに

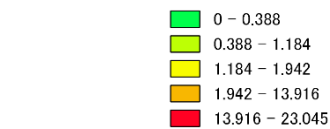
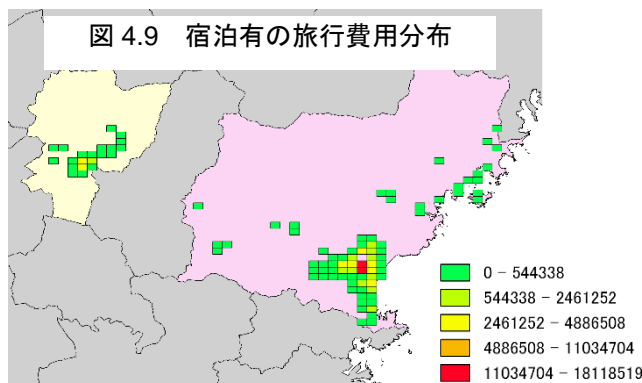
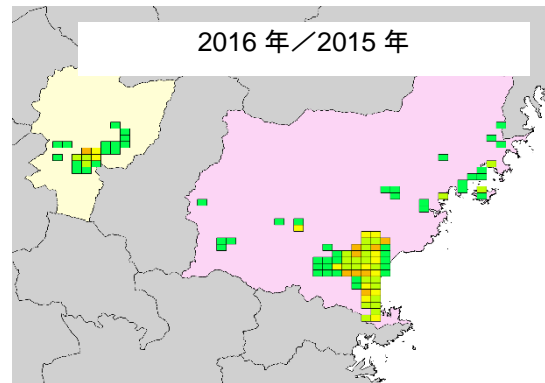
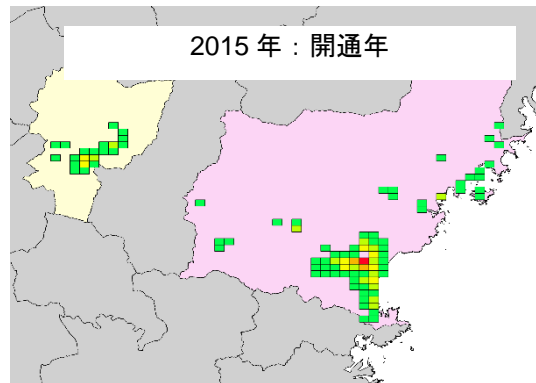
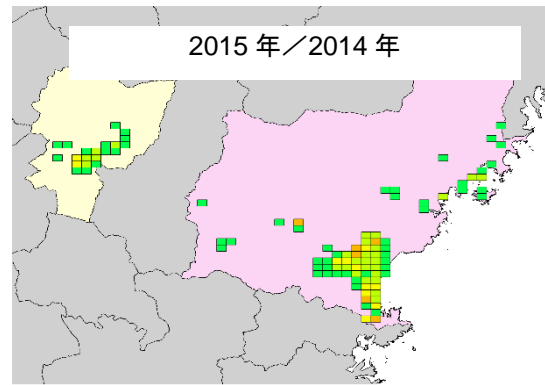
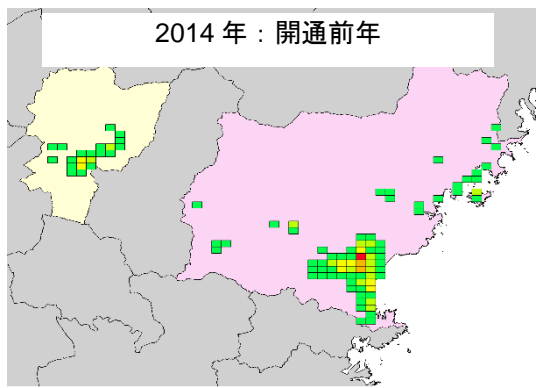


図 4.10 宿泊有の旅行費用の増減  
(2:00 滞在者)

図 4.9 宿泊有の旅行費用 (2:00 滞在者)

有意な正值を取っている。これは訪問者が観光関連施設を訪問していることを示しており、妥当な結果が得られた。

居住者を除去した場合のモデル：表 4.4, 表 4.5 と居住者を含む場合のモデル：表 4.6, 表 4.7 は、概ね同様の推定結果となることが明らかとなった。以上のより、当該地の居住者と来訪者の滞在地店分布は、おおむね、等しいことが分かった。

#### 4.4 トラベルコストによる道路ストック効果

本節では、観光地における宿泊の有無または滞在時間を考慮して、トラベルコスト評価による道路のストック効果を計測する。

本分析では、ストック効果を旅行費用として、宿泊の有無に着目して算出する。旅行費用は、来訪者数と一般化費用の積で定義されるもので、対象とす

る目的地が旅行費用を払ってでも来訪する価値があるとの想定によって、目的地の魅力度を定量的に表す指標である。これまでに分析に倣って、宿泊の有無は、時間帯を限定することで考慮する。すなわち、2 時の滞在者数から宿泊有の旅行費用を、11 時の滞在者数から、宿泊無の旅行費用を算出した。図 4.9 に宿泊有の旅行費用をまとめた結果を示す。高千穂町、延岡市ともに、それぞれの中心地である集客施設の多いメッシュでの旅行費用は大きくなっている。高千穂町は延岡市の中心地に比べて施設が集中しているため、旅行費用の大きなメッシュの広がり小さい。図 4.10 に宿泊有の旅行費用の増減をまとめ結果を示す。開通年(2015 年)から開通翌年(2016 年)にかけて延岡市では総旅行費用の増加しているメッシュが多く見られた。増加メッシュ数は限られるものの、高千穂町でも同様の結果が得られた。

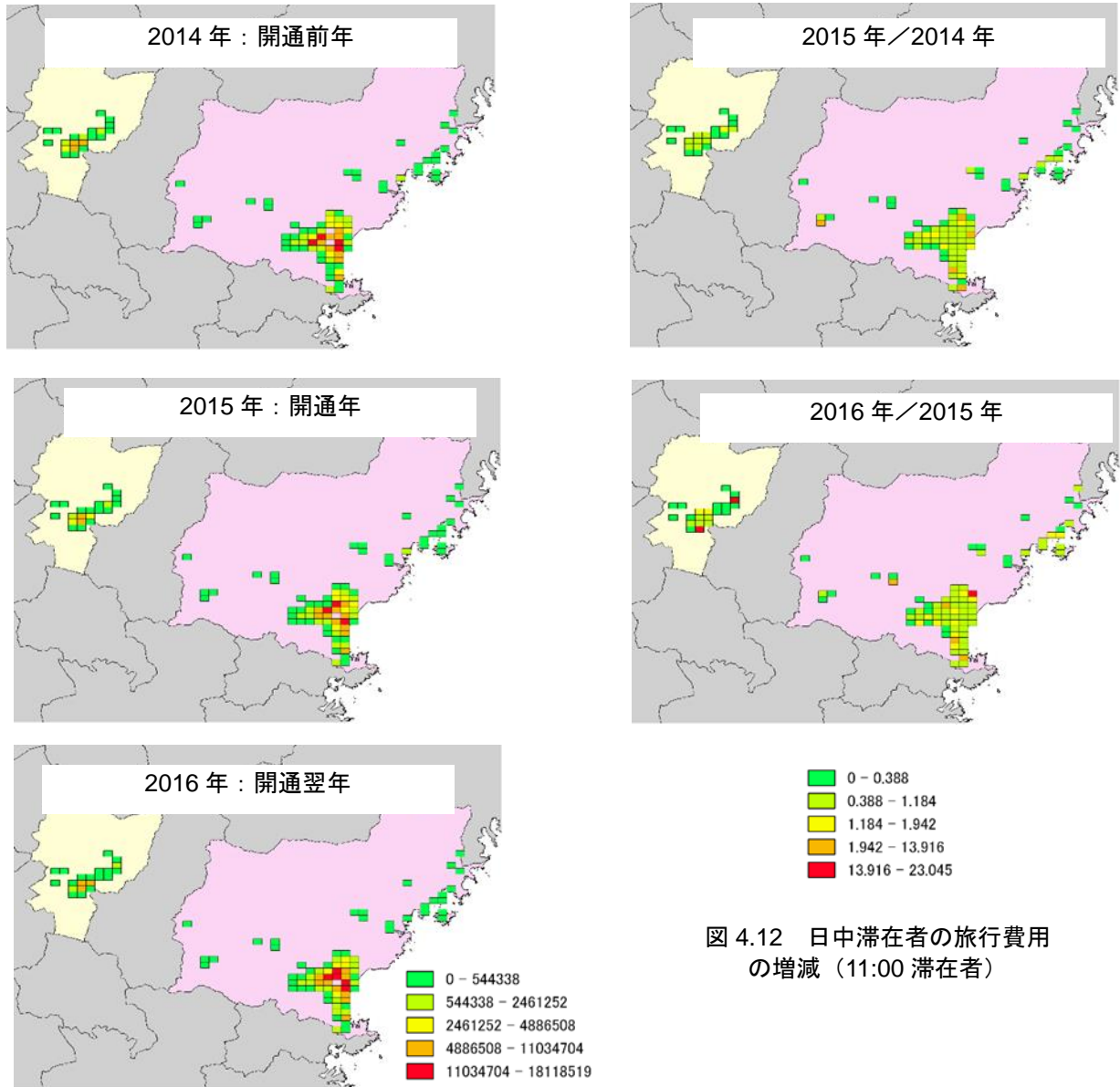


図 4.12 日中滞在者の旅行費用の増減（11:00 滞在者）

図 4.11 日中滞在者の旅行費用（11:00 滞在者）

図 4.11 は宿泊無の旅行費用の変化をまとめた結果である。宿泊有と同様、中心となる集客施設の多いメッシュで、旅行費用が高い。また昼間の方が、旅行費用の空間的な分布が広がっており、滞在者がより広い範囲に分散していることがわかる。図 4.12 は宿泊無の宿泊有の旅行費用の増減をまとめた結果である。宿泊有に比べて、トラベルコストが増加しているメッシュ数は少ない。しかし高千穂町では、中心部以外のメッシュへの来街者が多く、これらの場所への集客が多くなる傾向が見れた。

対象地域においてメッシュ単位でトラベルコストを算出した結果、開通年から開通翌年にかけて、延岡市で宿泊者のトラベルコストが増加したメッシュ

は多くなっており、東九州道の開通によって一定の影響があったと考えられる。一方で日中滞在者のトラベルコストが増加するケースは少なく、開通の影響はあまり大きくない。他方で高千穂町に関しては、宿泊によるトラベルコストの増加は小さい一方で、日中の来訪者のトラベルコストの増加が、対象地域の中心部以外のメッシュで見られることが明らかとなった。

# 第5章 工業団地地価のヘドニック分析

## 5.1 研究の狙い

本テーマでは、高速道路アクセスと工業団地地価の関係を明らかにすること、ならびに、モバイル空間統計などの交通ビッグデータから得られる工業団地別の取引の特性指標が地価に及ぼす影響を明らかにするため、工業団地の地価と、その特性を表すデータベースを作成して、ヘドニックモデルの推定を試みる。

前年度までに開発したヘドニックモデルでは、モバイル空間統計から得られる流動から、工業団地内への通勤者と取引者の分離を行う必要があることが明らかとなっていた。ただしこの分離作業を、工業団地に到着する時刻をもとに行う手順をとったところ、多数の流動パターンが混在することにより、良好な分析結果が得られないという課題があった。そこで、取引パターンを合理的に分類する方法を検討し、携帯ビッグデータをヘドニック分析において活用する方法を検討する。

## 5.2 対象地域とデータベースの構築

### 1) 工業団地データと高速道路アクセス

対象地域は、産業用地ガイドに掲載されている工業団地とした。産業用地ガイドは、一般財団法人日本立地センターが毎年発行しており、産業用地情報の提供を行っている。本研究では2015年版の産業用地ガイドを対象とした。産業用地ガイドに含まれている情報は、表 5.1の通りである。

産業用地ガイドには775の産業用地が掲載されているが、これらの産業用地には製造業や工場団地・流通団地のみでなく、オフィス・店舗を対象施設に含む用地も存在する。そこで純粋に工業団地と流通団地を抽出するために、「対象施設」に「製造業 | 工場 | 流通施設」を含み、かつ「オフィス | 店舗」を含まないという条件で、対象を抽出した。オフィス・店舗を含む用地を除外した理由としては、純粋な工業地とは傾向が違うニュータウンのような複合的な大規模開発地を排除することで、純粋に工業地における影響を取り出すためである。また譲渡単価には造成後価格と素地価格の2種類があるが、造成後価格で分譲されている工業団地の方が圧倒的に多いことから、譲渡単価は造成後価格になっているものに限定する。以上の絞り込みの結果、今回の工業地の定義に合致し、分析対象としたサンプル数は486となった。

表5.1 産業用地ガイド掲載項目

項目	掲載内容	備考
1	産業団地・用地名	—
2	所在地	—
3	事業主体	—
4	面積	全体計画, 分譲対象, 分譲可能
5	施設対象	—
6	交通アクセス	道路, 鉄道, 空港, 港湾
7	都市計画法の用途地域	—
8	用水	工業用水, 上水道, 地下水の給水量
9	排水	下水道接続の有無, 企業内処理の可否
10	電力	高压, 特別高压, 変電所距離
11	ガス	—
12	譲渡単価	—
13	賃貸区分	—
14	連絡先	—

表 5.2 モバイル空間統計の属性情報

カラム名	内容	備考
Date	日付	YYYYMMDD
day_of_week	曜日	11:月,12:火,13:水,14:木,15:金,16:土,17:日
time	時間	HHMM ※MM=00
area	エリア	3次メッシュ (1kmメッシュ):8桁
residence	居住地	市区町村コード:5桁
age	年齢	区別なし
gender	性別	区別なし
population	人口	調査実数

### 2) 高速道路アクセス・取引特性データ

ヘドニックモデルの入力変数として、詳細な地点別の高速道路アクセスデータベースを作成する。産業用地ガイドの情報を基に、工業団地別に最寄りICまでの距離を整理した。

工業団地に関する取引特性を表すデータとして、本分析ではモバイル空間統計データを用いる。同

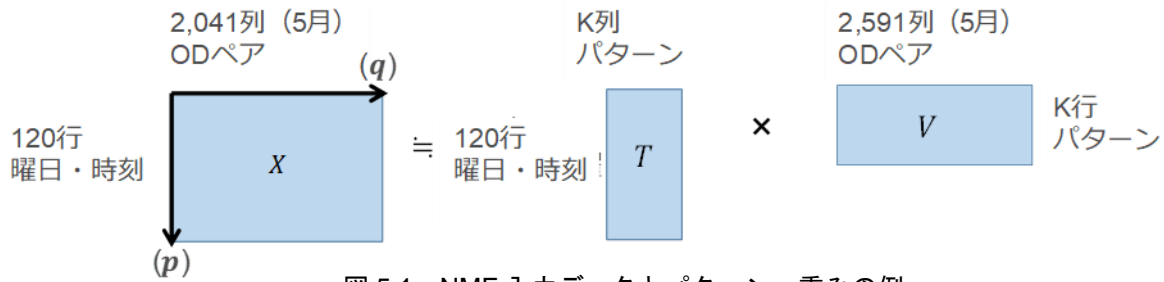


図 5.1 NMF 入力データとパターン・重みの例

データは、日本全国において、1時間ごとの人口分布の時間変動を観測することができ、その特徴において、紙ベースのアンケート調査で行われているこれまでのパーソントリップ調査を代替する手法としても期待されている。この仕組みや課題といった詳細については新階ら<sup>10)</sup>の記事に詳しくまとめられている。得られる情報を、表5.2にまとめて示す。

### 5.3 分析手法

#### 1) 通勤と取引の分離方法

工業団地（が存在する同一3次メッシュ内）への通勤と工業団地との取引の量（以降、交流量と呼ぶ）を分離するため、ODペア毎の曜日時間帯最小値を通勤と仮定し、式(5.1)で分離した。なお、ここでOはモバイル空間統計の居住地を、Dは3次メッシュに紐づけた工業団地を示す。

$$POP_{ij}^{mdwh} = pop_{ij}^{mdwh} - \min_{m,d} (pop_{ij}^{wh}) \quad (5.1)$$

POP：交流量，pop：滞在人口（モバ空），  
i：O（契約地），j：D（工業団地），x：滞在人口，  
m：月，d：平日，w：曜日，h：時刻（1時間帯）

#### 2) 取引パターンの分類

NMF<sup>48)</sup>を用いて時間帯に着目した工業団地のパターン分類を行う。NMFは、非負値行列から統計的に特徴のあるパターンを抽出する多変量解析手法の1つである。NMFのイメージを図5.1に示す。Xは入力データ行列、Tはパターンの特徴を示す特徴行列、Vはパターンの重みを示す重み行列である。データ行列を図のように配置することによって、パターンの特徴は時間軸方向の特性を表す内容に限定される。一方で空間軸方向は、ODを区別せずに列方向に配置することによって、時間パターンOD別負荷量として、重み行列を算出できる。なおこの配置をとることによって、ODベースではゼロ値が大変多くあらわれる疎行列となるデータベースからも、有効な時間分布パ

ターンが抽出できる。

祝日や休日の影響を除くため、5月、8月、11月で祝日（お盆含む）がない1週間（各月の第3週が該当）を分析対象とした。月別のパターンを算出するため、月別に入力データ行列を作成し、曜日と時刻に関する主要なパターンへ分類した。

#### 3) ハーフィンダール・ハーシュマン指標

取引先の多様性は工業地の需要の高さを表しているという仮説のもと、多様性を表す指標としてハーフィンダール・ハーシュマン指標（HH指標／HHI）<sup>49)</sup>を導入した。HH指標は、元来は市場の集中度を測る指標として登場したため主に経済・金融分野で使われていて、市場占有率（シェア）の2乗和によって定義される。たとえば10社が10%ずつシェアを分け合っている場合は $0.1^2 \times 10 = 0.1$ となり、各企業のシェアの格差が大きく、また市場に参加する企業数が少ないほど、HH指標は1に近づいていく。

本分析では、NMFから得られたパターンのうち、特に特徴的な時間分布を表すパターンごとに、取引先の多様性指標を式(5.2)で作成する。得られた指標を用いて、工業団地のパターン別取引先の多様性が地価へ及ぼす影響を確認する。

$$HH_{kj} = \sum_i \left\{ \left( \frac{v_{kij}}{\sum_i v_{kij}} \right)^2 \right\} \quad (5.2)$$

ここで $v_{kij}$ は、工業団地*i*と交流先*j*における、パターン*k*の重みを表す指標である。

#### 4) ヘドニックモデル

工業団地の単位面積当たり造成後価格を譲渡価格とみなして、回帰モデルの推定を行う。

$$\log(\text{譲渡価格}) = f \left( \begin{array}{l} \text{交通アクセス,} \\ \text{事業主体民間ダミー} \\ \text{, パターン別 HH指標} \end{array} \right) + u_1 \quad (5.3)$$



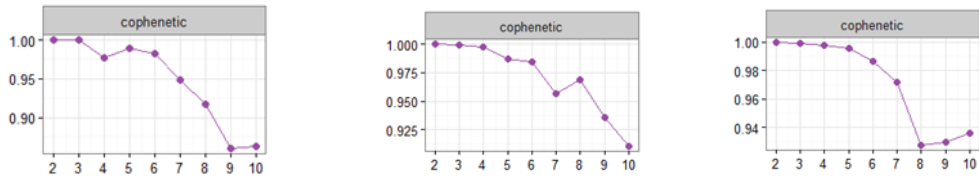


図 5.2 月別 Cophenetic とパターン数 (左 : 5 月, 真ん中, 8 月, 右 : 11 月)

表 5.3 月別パターンの特徴

パターン	時間帯 ピーク	曜日 ピーク	cos類似度 0.9以上
5月			
1	日中	月, 火	8月のパターン1
2	日中	木, 金	-
3	夜間	-	8月のパターン3
4	日中	木, 金	8月のパターン4/11月のパターン4
5	午前	月, 水	-
6	午後	金	8月のパターン6/11月のパターン5
8月			
1	日中	月, 火	5月のパターン1
2	午前	木, 金	-
3	夜間	-	5月のパターン3
4	日中	木, 金	5月のパターン4/11月のパターン4
5	8時, 17時	水	-
6	-	金	5月のパターン6/11月のパターン5
11月			
1	日中	火, 木	-
2	夜間	-	-
3	日中	月	-
4	日中	木, 金	5月のパターン4/8月のパターン4
5	8時, 17時	金	5月のパターン6/8月のパターン6
6	午前	金	-

表 5.4 パターン間の cos 類似度

	5月						8月						11月					
	パターン1	パターン2	パターン3	パターン4	パターン5	パターン6	パターン1	パターン2	パターン3	パターン4	パターン5	パターン6	パターン1	パターン2	パターン3	パターン4	パターン5	パターン6
5月	1	0.386887	0.280198	0.234791	0.559531	0.219832	0.902962	0.527818	0.243057	0.133629	0.398906	0.102997	0.75856	0.363527	0.634369	0.19228	0.216	0.259048
5月	0.386887	1	0.23303	0.691865	0.546785	0.524006	0.592798	0.620226	0.204893	0.682694	0.383606	0.477333	0.702054	0.266827	0.318062	0.756293	0.549693	0.623121
5月	0.280198	0.23303	1	0.251243	0.422933	0.422298	0.294244	0.550372	0.905722	0.280272	0.616524	0.362047	0.266486	0.878389	0.209426	0.333019	0.286558	0.272758
5月	0.234791	0.691865	0.251243	1	0.443213	0.85537	0.502687	0.444091	0.221019	0.975118	0.299277	0.865684	0.438679	0.333018	0.248569	0.939833	0.889805	0.850987
5月	0.559531	0.546785	0.422933	0.443213	1	0.454964	0.60703	0.73895	0.504446	0.384595	0.459856	0.331293	0.473124	0.569789	0.641275	0.494869	0.404445	0.368205
5月	0.219832	0.524006	0.422298	0.85537	0.454964	1	0.447565	0.458121	0.412942	0.840071	0.549888	0.907697	0.355543	0.526073	0.205994	0.755558	0.954722	0.750477
8月	0.902962	0.592798	0.294244	0.502687	0.60703	0.447565	1	0.517243	0.24202	0.399223	0.413122	0.345563	0.816017	0.400445	0.636772	0.448334	0.43375	0.505114
8月	0.527818	0.620226	0.550372	0.444091	0.73895	0.458121	0.517243	1	0.439792	0.451231	0.591488	0.307297	0.635527	0.626956	0.321546	0.551173	0.402354	0.320274
8月	0.243057	0.204893	0.905722	0.221019	0.504446	0.412942	0.24202	0.439792	1	0.221033	0.503192	0.347318	0.135516	0.818329	0.333815	0.269741	0.29587	0.252442
8月	0.133629	0.682694	0.280272	0.975118	0.384595	0.840071	0.399223	0.451231	0.221033	1	0.309542	0.848029	0.422251	0.344794	0.106767	0.960032	0.858609	0.826932
8月	0.398906	0.383606	0.616524	0.299277	0.459856	0.549888	0.413122	0.591488	0.503192	0.309542	1	0.387964	0.446402	0.650481	0.165001	0.31661	0.442273	0.33648
8月	0.102997	0.477333	0.362047	0.865684	0.331293	0.907697	0.345563	0.307297	0.347318	0.848029	0.387964	1	0.258354	0.449033	0.123561	0.747548	0.91154	0.796171
11月	0.75856	0.702054	0.266486	0.438679	0.473124	0.355543	0.816017	0.635527	0.135516	0.422251	0.446402	0.258354	1	0.352941	0.202518	0.480485	0.330624	0.389166
11月	0.363527	0.266827	0.878389	0.333018	0.569789	0.526073	0.400445	0.626956	0.818329	0.344794	0.650481	0.449033	0.352941	1	0.239987	0.379402	0.380607	0.3421
11月	0.634369	0.318062	0.209426	0.248569	0.641275	0.205994	0.636772	0.321546	0.333815	0.106767	0.165001	0.123561	0.202518	0.239987	1	0.182508	0.236259	0.199279
11月	0.19228	0.756293	0.333019	0.939833	0.494869	0.755558	0.448334	0.551173	0.269741	0.960032	0.31661	0.747548	0.480485	0.379402	0.182508	1	0.755953	0.788314
11月	0.216	0.549693	0.286558	0.889805	0.404445	0.954722	0.43375	0.402354	0.29587	0.858609	0.442273	0.91154	0.330624	0.380607	0.236259	0.755953	1	0.745528
11月	0.259048	0.623121	0.272758	0.850987	0.368205	0.750477	0.505114	0.320274	0.252442	0.826932	0.33648	0.796171	0.389166	0.3421	0.199279	0.788314	0.745528	1

## 5.4 ヘドニック分析の結果

### 1) NMFによる取引パターンの分類結果

NMFでは、分析者がパターン数Kを設定する必要がある。本研究では、パターンの安定性の尺度として Brunetrtら<sup>44)</sup>が提案した共役相関係数 (cophenetic) が減少し始めるKを参照して、K=6とした。

特徴行列から得られるパターンの表5.3に、月別パターン別のcos類似度を表 1.15に、それぞれ示す。3か月での全出現パターンのうち、cos類似度0.9以上は、「5月と8月のパターン1」、「5月と8月のパターン3」、「5月と8月のパターン4と11月のパターン4」、「5月と8月のパターン6と11月のパターン5」であった。以上より、工業団地への流入交通

表 5.5 ヘドニックモデルの推定結果

log(平均単価)	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	10.583	0.193	54.854	9.E-92
最寄り高速道路距離	-0.009	0.004	-2.173	0.032
最寄り空港距離	0.001	0.002	0.362	0.718
最寄り新幹線駅距離	-0.002	0.001	-1.340	0.183
最寄り一般駅距離	0.008	0.007	1.190	0.236
最寄り港湾距離	0.001	0.002	0.493	0.623
事業主体ダミー：民間	0.253	0.149	1.699	0.092
HH_01	-0.114	0.372	-0.307	0.760
HH_03	-0.565	0.291	-1.945	0.054
HH_04	0.377	0.342	1.101	0.273
HH_06	-0.699	0.305	-2.294	0.023
Multiple R-squared	0.242			
Adjusted R-squared	0.184			

量の時間分布パターンは、月別に異なる特徴を示すことが明らかとなった。

本研究では2015年5月、8月、11月のモバイル空間統計を入手したが、全体傾向として類似パターンが多かった5月と8月のデータを対象に分析を行う。なお、産業用地ガイドに掲載されている用地とモバイル空間統計の対応をとり、さらに産業用地ガイドに掲載されている用地との対応がとれた工業団地数は最終的に412となった。

う結果が得られた。これは広域的な生産活動の調整に際して、夜間や週末といった、相手先の生産活動の特性に合わせた物流や生産調整が可能な工業団地ほど、地価が高いことを表している。以上の結果より、特に高速道路アクセスが工業団地地価に対して有意な影響を及ぼしていること、ならびに取引先の多様性の中でも、夜間と週末に取引先が多様な工業団地の方が、地価が高くなる傾向が明らかとなった。

## 2) ヘドニック分析の結果

推定結果を表5.5に示す。モデルの適合度を表す決定係数は0.24程度で、あまり高くない。ただし既往研究で報告されている値とは同程度であり、一定の精度を有する結果が得られたと思われる。

距離指標の中で、最寄高速道路距離のみが負で有意となり、高速道路アクセスが工業団地地価に対して正の効果を持つことが確認できた。なお事業主体が民間の場合は、10%程度の有意水準ではあるが、地価が高くなる傾向がみられた。

HH\_01は日中/月火にピークをもつパターンの取引先の空間的な多様性指標、HH\_03は夜間にピークをもつパターンの取引先の空間的な多様性指標、HH\_04は日中/木金にピークをもつパターンの取引先の空間的な多様性指標、HH\_06は週末(金)にピークをもつパターンの取引先の空間的な多様性指標である。パターン別の取引の多様性のうち、HH\_03とHH\_06が負で有意となった。すなわち、夜間や週末にHH指標が小さい、つまり取引先に多様性がある工業団地の地価が高いとい

# 第6章 統計手法による道路ストック効果の検証

## 6.1 本テーマの狙い

前章までの分析では、近年注目されることの多い交通ビッグデータの中でも、ETC2.0とモバイル空間統計に着目して、それらを用いたインフラのストック効果を計測する手法の開発を行った。中間評価での指摘を受けて、各手法がストック効果計測の全体フレームの中でどの部分に対応するかについては、7.2において整理を試みる。一方で、社会基盤ストック効果全体の再構築を掲げた以上、既存のストック効果計測手法が存在する中で、一連の提案手法の狙いや価値を明確に位置付けることはできていないとの指摘を受けた。この点に関しては、本課題を追加して対応する。

この指摘に対する研究代表者の対応方針は、以下のとおりである：既存の費用便益手法が理想的な仮定に基づく演繹的な手法となっているのに対して、実社会を細かくモニタリングできる統計データが入手できるようになった今日では、新しいデータに基づく新しい手法の開発が必要である。すなわち、モデルの仮定に立脚した将来予測手法と並んで、過去を振り返って便益の発現状況を検証できる方法が重要、との立場に立っている。本章の提案手法は、いまのところはビッグデータを直接組み込んだ手法とはなっていないが、今後交通や土地利用に関するビッグデータを組み込む余地は大きい。既存手法との違いは、本テーマに関する成果を述べたのちに振り返ることとして、まずはストック効果の検証につながるモデルシステムと、その背景を概観する。

冒頭に提示した社会基盤整備の意思決定は、伝統的に費用便益分析に基づいて行われてきた。費用便益分析は、現在の費用便益マニュアルで標準化されている便益発生側での計測手法においても、近年実用的な計算手法の開発が進められてきた空間的な便益帰着側に着目した計測手法（応用空間一般均衡分析）においても、同一のロジックの上に組み立てられている、すなわち、経済学的な考察に基づく因果構造を前提として、統計学的な仮定の上で将来予測を行う手続き合理的な手法である。ただし、事業効果の事前予測値と事後の観測値の間では、常に乖離が生じる。

平成28年11月に国土交通省が策定した「ストック効果の最大化に向けて～その具体的戦略の提言～」<sup>48)</sup>では、社会基盤整備のストック効果の最大化のために、発現した様々なストック効果を積極的に把握し、「見える化」することが推奨されて

いる。さらにこの方針では、ストック効果に関して得られる知見を事業に有効活用するという、社会基盤のマネジメントサイクルの確立の必要性が述べられている。

社会基盤のストック効果が様々な面で発現することは論を俟たない。しかしその一方で、予測の基盤となる経済モデルのロジックやモデル構造そのものの妥当性や限界が再検討されることはほとんどなかった。標準的な経済モデルでは、関係主体の経済学的な行動指針を踏まえた最適化問題から演繹的にモデル構造を導出する。また実証分析の際は、入手可能なデータに適合するようにある程度簡略化したモデル構造を用いる。この一連のモデル同定手順では、モデル構造を省みる余地は少なく、せいぜいモデルパラメータの推定手順を見直すなどの修正が行われるに留まっていた。しかし「ストック効果に関して得られる知見を事業に有効活用する」ことには、本来モデルと現実との乖離が生じる原因をモデル構造に遡って検討し、新たな経済モデルを開発する手順も含まれるはずである。ただしその実現には、経済モデル実際に発現した効果を計測し、経済モデルが想定した因果構造の妥当性を事後的に検証する分析手法が必要である。

本研究では道路整備の効果の中でもストック効果に着目し、統計的因果探索を用いて個別地域・主体での社会基盤の整備と経済成長の因果関係を明らかにすることを目的とする。具体的には、道路アクセスに変化が見られた地域を対象に、統計的因果構造を経年的に推計する。

## 6.2 既往研究と本研究の位置付け

小池<sup>49)</sup>は、近年ストック効果が注目される理由として、費用便益分析の背景にある仮説的補償原理の仮定から逸脱するような地域開発効果を挙げている。社会基盤整備事業の優位性を判断するのであれば、個別地域・主体の効果の明示は必要ないことを指摘した上で、個別地域・主体の効果（帰着便益）は、社会的に望ましい分配に関する計画論を語る上で重要であるとしている。これを踏まえて小池は、帰着便益を明示することによって、その限界や活用方法について技術者や政策決定者がより深く議論することが重要と結論付けている。

社会基盤による生産力効果の測定には、一般的にAschauer<sup>50)</sup>によって研究がはじめられた生産

関数アプローチが用いられる。林<sup>51)</sup>は、我が国では地域間再分配の手段として、より多くの公共投資が生産や所得の低い地域へ投入されることを指摘した。江尻ら<sup>52)</sup>は、生産関数アプローチでは社会基盤と経済成長の間の因果関係が同定できないことを指摘したうえで、社会基盤の整備が経済成長の結果であるという、逆因果の可能性について言及した。その上で、因果構造の同定問題を、生産力効果計測上の課題としている。

複数の観測データ間の因果関係を明らかにする手法は統計的因果推論<sup>53)</sup>と呼ばれ、その代表的な手法の一つに操作変数法がある。Duranton & Turner<sup>54)</sup>は、高速道路ストックが1983~2003年のアメリカの都市人口・雇用の成長に与える因果効果について構造方程式を定式化して、操作変数法によってモデルを推定した。その結果、道路ストック10%の増加が都市の雇用を20年間で約1.5%上昇させるという推定結果を得ている。また、Holl<sup>55)</sup>は、1997~2007年のスペインの製造業を対象に、高速道路整備への近接性が企業の生産性に及ぼす影響について固定操作変数法を用いて分析を行い、高速道路によるアクセス性の向上効果は企業レベルでは集積の効果以上に企業の生産性を向上させたことを明らかにした。さらに、高速道路に近接する郊外地域の企業で生産性が大きく向上する一方で、その周辺地域の企業の生産性が低下することにより、道路投資による便益の一部が相殺されることを示した。Redding and Turner<sup>56)</sup>は、交通基盤整備が経済活動の空間構造に及ぼす効果を、操作変数法などを用いて推定する研究についてレビューを行った。都市内(intracity)、都市間(intercity)で交通インフラを分類し、主に立地変化に着目してそれぞれについての既往研究を整理した。その結果、道路や鉄道までの距離が2倍になると、人口や雇用が6~15%低下すること、高速道路整備は都市人口や製造業の郊外化をもたらすこと、効果は経済発展の度合いには影響を受けないが、交通機関や産業によって異なることなどを明らかにした。

本研究は、自治体単位のデータに基づいて、アクセス性の向上と経済成長の因果関係を明らかにする。操作変数法に代表される統計的因果推論の手法は、分析に先立って経済学的知見や経済理論上の妥当性に基づいて因果構造を仮定し、その効果を統計的に推定する手順をとる。しかし、こ

の手順では逆因果の可能性が否定できない。本研究では、先験的な因果構造を仮定せずに、統計学的知見に基づいて因果構造を決定する。したがって、従来の手法と併用すれば因果関係の同定問題を解決できる可能性がある。

### 6.3 インフラストック効果の探索手法

本研究では、道路整備のストック効果に関して計測されることの多いアクセス性の向上と経済成長の因果関係を明らかにするため、統計的因果探索を用いる。統計的因果探索とは、未知の因果グラフを、データが示す分布特性に基づいて推測する手法である。

本研究では、統計的因果探索の中でも因果グラフが推測できるLiNGAM (Linear Non-Gaussian Acyclic Model (線形非ガウス非巡回モデル))を用いる。LiNGAMの推定原理は、まず変数間の因果的順序を求め、その因果的順序に従う回帰分析を実行することによって、変数間の因果グラフを探索するという2段階で構成される。変数の因果的順序とは、その順序に従って変数を並べ替えると、後の変数が先の変数の原因となることが無いような、変数の順序である。なおAcyclicという表現に見られるように、因果的順序を仮定する場合は、因果が循環的に発生する変数組は、許容できない。また非ガウス分布は、「説明変数と残差が独立ならば、推計は正しい因果的順序で行われている」というDarmois-Skitovichの定理の対偶に基づく統計的性質を考慮するために必要な仮定である。

$p$ 種の観測変数 $x_1, x_2, \dots, x_p$ に関するLiNGAMモデルは、式(6.1)で定式化される。

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i \quad (i = 1, \dots, p) \quad (6.1)$$

ここで、 $b_{ij}$ は推定される係数、 $e_i$ は誤差変数をあらわす。 $x_i$ は、 $i$ 以外の観測変数 $x_j (j = 1, \dots, j \neq i)$ とその誤差変数 $e_i$ の重み $b_{ij}$ 付きの線形和である。なお、誤差変数 $e_i$ は独立で、非ガウス連続分布に従うと仮定する。行列を使うと、式(6.1)は式(6.2)と表される。

$$\mathbf{x} = \mathbf{B}\mathbf{x} + \mathbf{e} \quad (6.2)$$

$\mathbf{x}$ ,  $\mathbf{B}$ ,  $\mathbf{e}$ は、それぞれ観測変数 $x_i$ 、係数 $b_{ij}$ 、誤差変数 $e_i$ を表す。 $\mathbf{B}$ のゼロ・非ゼロパターンから、変数間の因果関係を表す因果グラフを作成する。例を図-1に示す。同図では、四角は観測変数を、矢印は因果を表しており、矢印の出る変数は原因、矢印の入る変数は結果である。

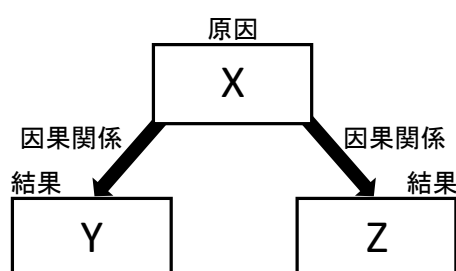


図 6.1 因果グラフの例

LiNGAM モデルの推計は、清水<sup>10)</sup>の提案した独立成分分析によるアプローチに則って行う。独立成分分析は、多変量系列を、非ガウス分布を持つ成分の混合によって記述する統計手法である。LiNGAM の推計は、本来変数間の独立性の検証によって行う方が自然だが、非ガウス成分をシステムティックに抽出できる ICA を用いることによって、効率的に因果構造を探索できるという利点がある。以下に推計手順の概要を示す。

式(2)の係数行列  $\mathbf{B}$  を左辺に移項して、左から  $(\mathbf{I} - \mathbf{B})^{-1}$  を乗じて、式(6.3)を得る。

$$\mathbf{x} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{e} = \mathbf{A} \mathbf{e} \quad (6.3)$$

ここで、 $\mathbf{A} = (\mathbf{B} - \mathbf{I})^{-1}$  である。 $p$  次元の誤差変数ベクトル  $\mathbf{e}$  の成分は独立で非ガウス分布に従うので、式(3)は独立成分分析 (以下 ICA) の混合行列  $\mathbf{A}$  から  $(\mathbf{I} - \mathbf{B})^{-1} = \mathbf{A}$  を満たす行列として求められる。 $(\mathbf{I} - \mathbf{B})^{-1} = \mathbf{A}$  を満たすには、行の順序と尺度を一意に決定しなくてはならない<sup>10)</sup>。そこで、LiNGAM モデルの仮定に基づいて行の順序を決定する置換行列  $\mathbf{P}$  と、行の尺度を決定する対角行列  $\mathbf{D}$  を探索し、 $(\mathbf{I} - \mathbf{B})^{-1} = \mathbf{DPA}$  として係数行列  $\mathbf{B}$  を得る。

ICA には推定誤差が残るため、モデルに含まれる任意の 2 変数の真の関係が独立であっても、その 2 変数に対応する  $\mathbf{B}$  の要素は、の係数が正しく 0 となるとは限らない。そこで、 $\mathbf{B}$  からゼロ・非ゼロパターンを得るため、以下の手順により、まず変数を因果的順序に並び替える。非巡回の仮定より、式(2)の  $\mathbf{B}$  は、因果的順序に従って変数が並んでいる場合、対角及び上三角成分が 0 となる。あるいは、対角及び上三角成分の絶対値の和が最小となる時、変数は因果的順序に従って並んでいる。

しかし、本研究で使用したアルゴリズムの性質上、変数順序の初期配置によって結果が変化するため、初期配置の候補全て ( $n$  変数の場合、 $n!$ 通り) で推計し、その頻度分布に基づいて因果的順序を決定する。

以上の手順で求めた因果的順序を基に、目的変数よりも因果的順序の早い全変数を説明変数にとって、Lasso による回帰を行う。Lasso とは、する手法である<sup>11)</sup>。Lasso によって推計されたパラメータに基づいて、LiNGAM モデルの推計を行う。さらに、推計した係数行列  $\mathbf{B}$  のゼロ・非ゼロパターンに基づき、因果グラフを作成する。なお、推計には Python のプログラムを使用した。

プログラムの作成では、ICA の推定法では代表的な FastICA を、置換行列の探索にはハンガリア

表 6.1 データの詳細

期間*	平成 17 年, 22 年, 27 年
範囲	鳥取県, 島根県, 岡山県, 広島県のデータ欠損のない 54~61 市町村
変数**	就業者数/国勢調査 人口/国勢調査 事業所数/経済センサス 総生産/市町村民計算 地価 (住宅) /公示地価 地価 (商業) /公示地価 域内アクセス指標/NITAS 域外アクセス指標/NITAS $\Delta$ 域内アクセス指標/NITAS (域内アクセスの時点間差分)

\*観測年次が異なるデータは直近の平成 21 年, 平成 28 年の値を用いる

\*\*以下はデータの出典を表す

線形回帰モデルの推定と変数選択を同時に実行する手法を用いた。さら求める問題は、下三角行列を割当問題を解く munkres ライブラリを、Lasso は scikit-learn ライブラリからそれぞれ使用した。本分析において、分析者が想定する因果の初期仮説は、「人口社会指標→経済活動指標」という順序の因果である。これに加えて、アクセス指標は「アクセス指標→人口社会指標」、または、「アクセス指標→経済活動指標」の順序で因果を示すと考えた。

#### 6.4 分析対象及び使用データ

本研究では、平成 27 年 3 月に全線開通した尾道松江線を分析対象とする。尾道松江線は、広島県尾道市を起点とし、中国山地を横断して島根県松江市を終点とする延長約 137km の自動車専用道路であり、新直轄方式で整備された。路線計画は昭和 62 年 9 月に始まり、平成 22 年 11 月から順次開通し、平成 27 年 3 月 22 日に全線開通した。推計は、人口社会指標→経済活動指標の因果の存在を仮説とし、それらとアクセス指標との因果関係を探索するため、表-1 に示す変数を用いる。ここで、就業者数、人口、事業所数は人口社会指標、セス (域内アクセスの時点間差分) は、アクセス指標、域内アクセス、域外アクセス、 $\Delta$  域内アクセス費用である。なお、地域  $i$  のアクセス指標  $AC_i$  は、以下の式(6.4)で定義する。

$$AC_i = \sum_{j=1}^n \frac{P_j}{t_{ij}^2} \quad (6.4)$$

ここで、 $i, j$ : 地域、 $t_{ij}$ :  $i$  から  $j$  への自動車での所要時間、 $n$ : 地域数、 $P_j$ : 地域  $j$  の人口である。

表 6.2 Lasso 回帰の結果（平成 17 年-22 年）

目的変数	決定係数	説明変数と偏回帰係数								
		就業者	域外	人口	Δ 域内	総生産	事業所数	地価(商業)	域内	地価(住宅)
就業者		0	0	0	0	0	0	0	0	0
域外	0.251	0.411	0	0	0	0	0	0	0	0
人口	0.990	0.900	0	0	0	0	0	0	0	0
Δ 域内	0.006	0	0	0.027	0	0	0	0	0	0
総生産	0.915	0	0	0.862	0	0	0	0	0	0
事業所数	0.986	0.746	0	0.152	0	0	0	0	0	0
地価(商業)	0.793	0	0.007	0	0.235	0	0.710	0	0	0
域内	0.426	0	0	0	0.561	0	0	0	0	0
地価(住宅)	0.827	0	0	0	0	0	0	0.619	0.338	0

表 6.3 Lasso 回帰の結果（平成 22 年-27 年）

目的変数	決定係数	説明変数と偏回帰係数								
		就業者	Δ 域内	人口	総生産	地価(商業)	事業所数	域外	地価(住宅)	域内
就業者		0	0	0	0	0	0	0	0	0
Δ 域内	0	0	0	0	0	0	0	0	0	0
人口	0.990	0.900	0	0	0	0	0	0	0	0
総生産	0.983	0	0	0.897	0	0	0	0	0	0
地価(商業)	0.704	0.070	0.014	0	0.674	0	0	0	0	0
事業所数	0.987	0.898	0	0	0	0	0	0	0	0
域外	0.268	0	0	0	0	0.033	0.398	0	0	0
地価(住宅)	0.685	0	0.055	0	0	0.723	0	0	0	0
域内	0.661	0	0	0	0	0	-0.308	-0.057	0.773	0

表 6.4 Lasso 回帰の結果（平成 17 年-27 年）

目的変数	決定係数	説明変数と偏回帰係数								
		就業者	Δ 域内	人口	地価(商業)	地価(住宅)	域内	総生産	事業所数	域外
就業者		0	0	0	0	0	0	0	0	0
Δ 域内	0	0	0	0	0	0	0	0	0	0
人口	0.990	0.900	0	0	0	0	0	0	0	0
地価(商業)	0.764	0.724	0.184	0	0	0	0	0	0	0
地価(住宅)	0.753	0	0.238	0	0.642	0	0	0	0	0
域内	0.677	-0.248	0.247	0	0	0.586	0	0	0	0
総生産	0.983	0	0	0.897	0	0	0	0	0	0
事業所数	0.987	0.898	0	0	0	0	0	0	0	0
域外	0.268	0	0	0	0.033	0	0	0	0.398	0

また、域内アクセス指標は推計を行う市町村間のアクセス性、域外アクセス指標は西日本の県庁所在地へのアクセス性、Δ域内アクセス指標は2時点の域内アクセス指標の差分である。以上のデータを用いてデータ行列を3時点分作成し、推計に使用する。

### 6.5 因果グラフの推定と考察

作成したデータ行列を用いて LiNGAM モデルの推計を行う。本研究では平成 17 年-22 年間、平成 22 年-27 年間、平成 17 年-27 年間の 3 回で推計を行った。また、因果的順序は、初期配置の異なる推計を繰り返し、結果的に現れた因果的順序を記録し、出現回数が最多の因果的順序を採用

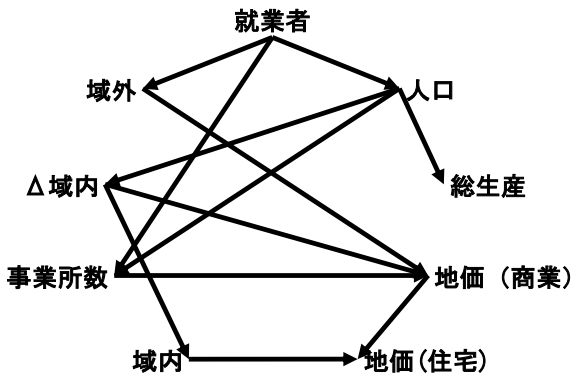


図 6.2 因果グラフ (平成 17 年-22 年)

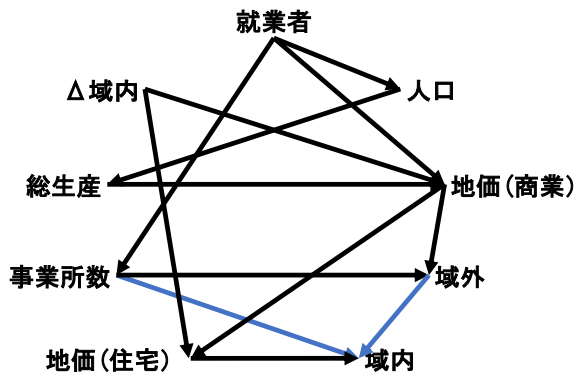


図 6.3 因果グラフ (平成 22 年-27 年)

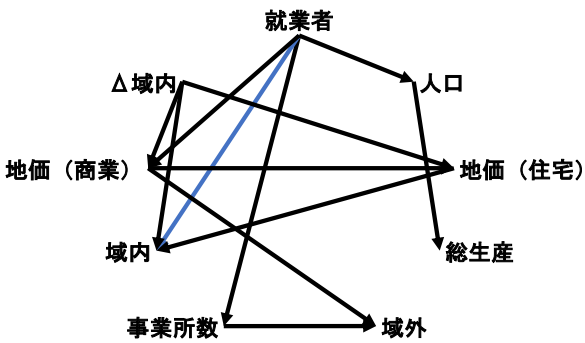


図 6.4 因果グラフ (平成 17 年-27 年)

の定義は前述の通りであり、値の増加はアクセス性の低下を表すので、図中ではアクセス指標に関わる矢印の色を逆にして示している。なお、同図は矢印を見やすくするため、変数を囲む四角を省略している。

図 6.2, 図 6.3, 図 6.4 には、「就業者→人口→総生産」、「△域内→地価」へのパスが共通して見られる。これは、変数の選択時に設定した初期の因果仮説である人口社会指標→経済活動指標に合致し、加えてアクセス指標→経済活動指標の因果関係の存在を示している。また、いずれの因果グラフも総生産には就業者と人口のみが原因となっており、アクセス指標の変化による総生産の変化が起こらない、という結果が得られた。

した。なお、出現回数が同数となっている因果的順序は、Lasso 回帰を行い、最も決定係数が高くなる因果的順序を採用した。採用した結果を表 6.2, 表 6.3, 表 6.4 に、それぞれ示す。

表の 1 列目は Lasso を適用する回帰モデルの目的変数であり、因果的順序に従って並んでいる。2 列目は決定係数, 3 列目以降は説明変数であり、表中の係数値は、対応行の目的変数に対する、対応列の説明変数に関する偏回帰係数である。すなわち同表において回帰モデルは、各行ごとに推計する。表 6.2, 表 6.3, 表 6.4 に基づいて作成した因果グラフをそれぞれ図 6.2, 図 6.3, 図 6.4 に示す。図中の黒矢印は正、青の矢印は負の効果であることを、それぞれ示す。ここで、アクセス指標

## 第7章 おわりに

### 7.1 本研究の成果

本研究では、複数データを活用した道路のストック効果計測手法の再構築というタイトルで、3つのテーマについて検討を行った。さらに、統計データを用いてストック効果を検証するアプローチを位置付けるため、1テーマを追加して検討を行った。テーマ別に得られた成果をまとめる。

「交通状態の質的指標と事故リスク分析」に関しては、まずETC2.0データから得られる速度指標の精度を検討し、速度データの推計精度が高いことを確認した。さらにETC2.0を搭載した車両の少ない（観測数の少ない）地方部の高速道路においても、速度データを用いれば、区間内で低下が起こる地点をその発生条件とともに算出できることを明らかにした。本手法は、特に地方の高速道路のうち暫定2車線で運用されている路線を対象としている。提案手法は、そのような路線の中で、走行速度が低下気味で十分なパフォーマンスが発揮できていない区間を同定して付加車線設置を検討できるばかりでなく、速度低下が発生する状況を参照することによって、交通需要管理などのソフトウェア施策の立案に有用な情報が得られる。すなわち、ストック効果最大化に向けて、交通ビッグデータを活用できる可能性が示された。

交通状態の質的指標と事故リスクについては、事故発生確率と事故台数別発生確率に関する分析を行った。この研究はこれまでマクロベースで議論されていた走行台キロあたりの事故発生に加えて、各区間の走行状態を表す質的指標をモデルの説明変数とすることによって、事故発生原因の説明変数となり得るか否かを検討する目的で実施した。事故発生モデルの推定結果からは、速度平均効果量、全車両の連続2区間の速度分散変化量、速度中央値、大型車速度中央値、左右加速度件数の5変数が、事故の発生に影響を与えていた。他方で交通量に関しては、モデル内で説明変数として設定したものの、事故発生に対して有意な影響は見られなかった（なお、本研究ではサンプルの区間長は全て同一としているため、走行台キロではなく、交通量を説明変数として用いた）。この結果が得られた理由として、本研究が高速道路の中でも限られた一部の区間を対象に行った統計モデルから得られた知見であることに注意する必要がある。すなわち対象区間をある都道府県内程度に限定すると、区間別の交通量の差が小

さくなるため、交通量（または区間長が異なる場合は走行台キロ）は、事故発生に対して有意な情報を持たないため、と考えられる。そのような場合、事故発生に対して有意な影響が現れるのは、たとえば本研究で定義した速度平均効果量、速度分散効果量、速度中央値、大型車速度中央値、左右加速度件数などの走行状態の質的指標の方が、事故発生に対する説明力が強くなる。なお特に速度平均効果量と速度分散効果量の2指標は、高速道路上で隣接する上流側の区間と当該区間の走行状態の差を表す質的指標であり、隣接区間との連続的な走行状態のモニタリングによって、事故発生につながる走行状態を抽出できる可能性が明らかとなった。

事故台数モデルでは、速度分布の乖離度や速度分布のパターンに関する交通状態の質的指標に着目して、モデル分析を行った。なお事故発生モデルでは、分析単位を高速道路の区間と設定したのに対して、事故台数の分析では分析単位を個別事故として、事故が発生したという条件の下での交通状態を抽出した。

事故台数に関しては、台数の多少を独自に分類して、事故台数を表す目的変数を作成したそのうえで、交通状態の質的指標を含むモデルを推計した。その結果、速度分布の乖離度、速度分布のパターン、前後加速度件数、ヨー角速度件数、ヨー角速度件数差分、時間帯（7時から10時）、天候（晴れ）、第1当事者車種（大型車）、交通量、道路線形に関する変数が事故台数に有意な影響を与えることを明らかにした。これらの変数には、事故発生モデルと同様に、交通量以外の交通状態の質的指標が含まれており、同変数によって、事故台数への影響が有意に検証できる可能性が明らかとなった。

以上より「交通状態の質的指標と事故リスク分析」の成果をまとめる。速度情報や加速度情報を活用することによって、区間別または時点別の交通状態の質的指標が得られること、その情報に基づいて付加車線の検討をより総合的に行えること、さらに事故リスクに対して交通状態の質的指標が有意に影響することが明らかとなった。

「観光地のトラベルコスト評価」に関しては、東九州自動車道開通による影響を明らかにするため、延岡市と高千穂町を対象として、これらの町域内への入れ込み客数と高速道路アクセスの関係性を分析する統計モデルと、来訪者の出発地に



基づくトラベルコスト評価を、新たな交通ビッグデータであるモバイル空間統計を用いて行った。同データは、1キロメッシュ単位で取得した。モバイル空間統計は、来街者の交通機関分担が不明のため、幹線旅客純流動調査を用いて、各出発地からの対象圏域内の来街者の自動車利用者を推計した。また、携帯電話契約地と1時間ごとの在圏人口の間で作成したOD表を、時間方向に合計すると、来訪者数をダブルカウントする問題と、移動目的が区別できないため、日常的な移動がある隣接した市町村に住んでいるデータをカウントする問題に注意して、出発地（携帯電話契約地）を参照して必要な部分のみ抽出するデータ処理を施した。その結果、昼間の来街者数として午前11時の人数を、夜間の来街者数として午前2時のデータを、それぞれ抽出した。なおデータ取得時点は、東九州自動車道の開通前（2014）、開通年（2015）、開通後（2016）の3時点である。

データを単純集計したところ、開通前と開通年を比較すると全体として来街者数が減少し、開通後は開通前と同程度の水準まで来街者が回復する傾向が明らかとなった。また開通年以降で延岡の中心市街地の来訪者数が増加していた。東九州道の最寄りICから距離のある高千穂町の来訪者数には、大きな変化が見られなかった。回帰分析に先立って域内の施設立地数の変化を確認したところ、宿泊施設が3か年で微増が確認された以外は、ほぼすべての施設において減少傾向にあった。小売店の施設数が他の施設と比較して大きく減少していたが、減少率は同等であることが確認できた。

昼間と夜間の来街者数を目的変数とする時系列モデルを推定したところ、延岡市、高千穂町とも適合度の高いモデルが推定された。延岡市では、休日のパラメータが有意な負の値を示し、平日よりも休日の来街者数が少ない傾向が確認できた。一方で小売店は有意な正の値をとり、来街者の目的地となることが確認できた。一方で延岡市については、昼間は宿泊施設の多いメッシュの滞在者は、少ない傾向がみられた。なお各メッシュから高速道路インターチェンジまでの高速道路アクセスのパラメータは、昼間の一部モデルのみ負で有意となり、高速道路までの距離と来街者数との間に、有意な関係は見られなかった。高千穂町の推定結果では、宿泊施設が有意な正值をとっていた。この傾向は、昼夜とも共通しており、宿泊施設近辺に自然観光の目的地が存在するものと思われる。またこれらとは別に、観光関連施設のパラメータも有意な正值を取っており、妥当な結果が得られた。

メッシュ別のトラベルコストを算出した。昼間の来街者について、高千穂町、延岡市ともに、それぞれの中心地である集客施設の多いメッシュで、大きな値が得られた。高千穂町は延岡市の中心地に比べて施設が集中しているため、旅行費用の大きなメッシュの広がり小さい。年次間の増減傾向に関しては、開通年（2015年）から開通翌年（2016年）にかけて延岡市で増加メッシュが多く見られた。一方で増加メッシュ数は限られるものの、高千穂町でも同様の結果が得られた。夜間の来街者数については、中心となる集客施設の多いメッシュで旅行費用が高い。また延岡市、高千穂町とも、夜間の方が滞在メッシュの広がり小さい。時点間のトラベルコストの増減傾向を算出したところ、昼間よりも増加メッシュが少ないことがわかった。しかし高千穂町では、中心部以外で増加メッシュがみられ、新たな宿泊地の開発が行われたものと思われる。

以上より、トラベルコスト評価では対象地の目的地特性を踏まえた分析が可能なことが明らかとなった。トラベルコストは、需要側から見た旅行目的地の魅力度評価値である。トリップベースの方法なので、従来の道路整備の3便益を着地側で簡易に評価する手法と考えられる。本課題ではモバイル空間統計を用いてトラベルコストを算出した結果、新たに魅力度が向上した目的地を明らかにできる。これは、道路のストック効果の可視化につながる成果と考えられる。

「工業団地地価のヘドニック分析」では、国内での分析が少ない工業団地の分譲価格について分析を行った。従来のヘドニック・アプローチの手法にとどまることなく、携帯電話の位置情報サービスを集計した人口流動データ（モバイル空間統計）を需要変数とみなし、交通インフラへの近接性以外の要素から工業団地の用地価格推定を試みた。なお本節で用いるモバイル空間統計の情報はトラベルコスト法で用いた情報と同様である。分析に当たって工業団地とその他の地域間で行われる取引情報のみを抽出する必要があったため、時間フィルタを用いて、モバイル空間統計から取引情報を取り出す方法を提案した。この取引情報から計算した時間帯パターン別HH指標を価格の説明変数としたところ、夜間や週末に交流量が多くなるパターンを持つ工業団地の取引相手地域の多様性が分譲価格に対して負に有意に働くことを示し、通常（日中や平日）に交流量の多い工業団地とは異なる工業団地でHH指標が小さい場合、（すなわち取引多様性が大きければ）用地価格も上がることをモデルから推定した。

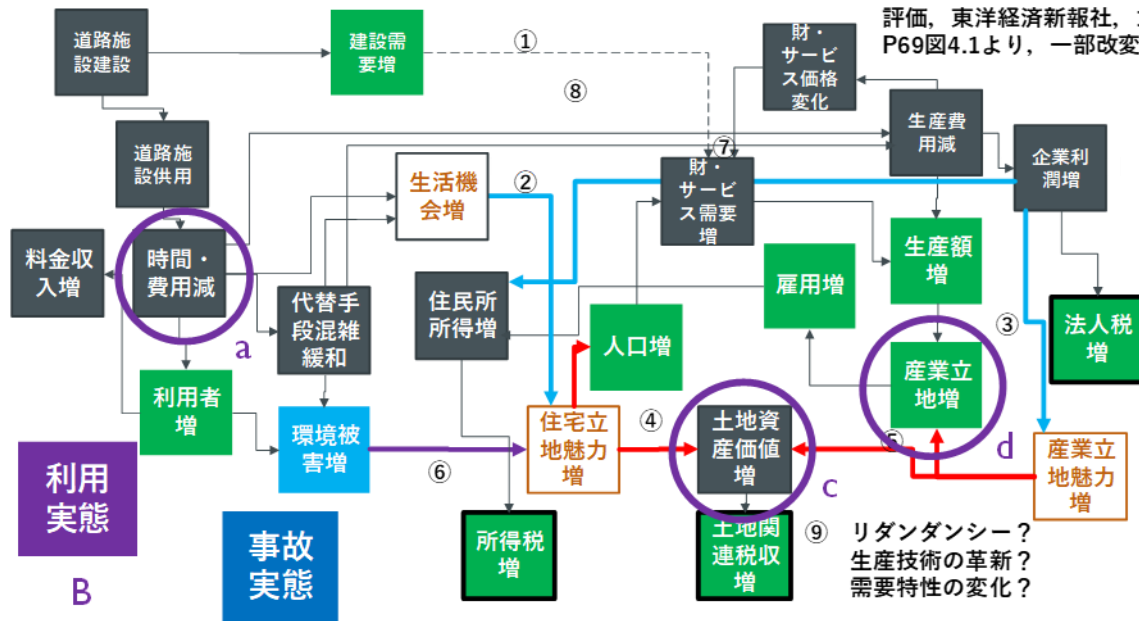


図 7.1 道路整備の効果の流れ

この研究成果について，道路のストック効果計測手法としての意義は，以下の2つと考えられる．1つはビッグデータに基づいた指標をストック効果計測に利用できたこと，もう1つは多様な取引先を実現するために全国にはりめぐらされた高速道路網が利用されているという実態に対して，その利用特性を計測できたことである．特に2点目は，アクセス可能性に基づく地価評価に対して実際の利用実績に基づく地価評価が可能なることを示した点で学術的実務的な意義もある．

「統計手法による道路ストック効果の検証」は，本研究で検討を進めてきた道路のストック効果と既存のストック効果の計測フレームを整理するために実施した．このテーマでは，主にデータの統計的特徴に着目して，回帰分析の相関を因果として抽出する上での妥当性を検証できる手法の開発を目指している．具体的には，統計的因果分析の中でも最新の研究成果である，因果探索の枠組みを応用した実証分析を行った．

中国地方を対象として，2005年，2010年，2015年にモデルを適用したところ，「就業者→人口→総生産」，「Δ域内：域内の交通アクセスの変化→地価」へのパスが共通して観測された．この結果は，変数の選択時に設定した初期の因果仮説である人口社会指標→経済活動指標に合致し，加えて域内の交通アクセス指標の改善が，経済活動指標（地価）に対して正の因果関係を有する可能性が

明らかとなった．また，いずれの因果グラフも総生産には就業者と人口のみが原因となっており，アクセス指標の変化が総生産の変化が起らない，という結果が得られた．この結果は，対象期間が短いことに起因している可能性がある．また空間的により多くの地域を含むように，分析を拡張する必要がある．

本課題に関しては，現時点では交通ビッグデータの取り込みなど，他の課題との直接的な連携はできていない．しかしながら，提案手法を通じた道路のストック効果の検証について，主に統計的な側面からの検討が可能なることが明らかとなった．そこで次節では，検討した3課題がストック効果計測上でどのように位置付けられるかを整理したうえで，従来のストック効果計測手法である費用便益分析や応用一般均衡分析と，本手法の関係の整理を試みる．

## 7.2 社会基盤ストック効果計測手法の整理

本節では，既存の経済学モデルベースの社会基盤ストック効果の計測手法と，本研究で提案した計測単用の関係を整理するとともに，新たなストック効果計測の可能性について，統計的因果推論の枠組みを位置付ける方法について検討する．なお簡単のため，以下では道路のストック効果を例に検討を進める．

道路のストック効果は、発生側の効果（例えば、道路三便益など）から帰着側の効果（例えば、地価やGDPなど）へと時間経過とともに形を変えて発現する。一方でこれまでストック効果として議論されてきた内容は、個々の計測対象に応じた個別効果に留まっており、その計測が道路ストック効果に関連する現象の因果構造のどの部分に当たるかについては、十分に整理されていなかった。以下では、各テーマで計測するストック効果と既存のストック効果全体との関連性を整理することにより、総合的な計測フレームを示す。また、社会実装に向けたパッケージ化の検討を行う。

図7.1に、中村<sup>57)</sup>によって示された道路整備の効果の流れを示す。本研究では、個別のストック効果として、事故リスク分析、観光地のトラベルコスト分析、工業団地地価へのヘドニック分析を行った。各テーマで扱う範囲は、事故リスク分析は図2.1に示すB+事故実態に、観光地のトラベルコスト分析は図2.1に示すB+dに、工業団地地価へのヘドニック分析は図2.1に示すa+B+cに、それぞれ該当する。

インフラ整備によるアクセス性の改善による生産性向上の中身は、以下の3つが考えられる。

- I. 中間財や最終財の輸送コスト削減
- II. 通勤コストの低下による労働市場アクセス圏の広がり
- III. 全要素生産性（=TFP）の上昇

IとIIIは、現行の費用便益分析のフレーム、およびSCGEのフレームのいずれでも計測できる。一方でIIIが起こるときは、費用便益分析であれSCGEであれ、モデルパラメータの変化を伴うため、将来予測の前提が崩れる。つまりIIIの効果が、費用便益分析やSCGEのシミュレーションによって便益算出されることはない。さらにIIIの具体的なケースについてについて、考えてみよう。まず、「産業構造の変化／低生産性部門から高生産性部門への雇用が転向」する状況があげられる。これは、国民経済的には重要だが、現象として雇用の部門間の転向（例えば農業従事者からIT技術者への転向）は、急速に起こるとは考えにくい。次に、「他地域からの企業転入・転出」が考えられる。企業移転が、従来のSCGEで想定するように、単にアクセス性の良い地点で生産することをあらかわすのなら、その効果は上記の1点目に相当する。TFPの上昇を伴う企業移転とは、たとえば設備更新を伴う移転によって、上記の1点目を超えた生産が行われる場合である。最後に、「移民の受け入れによる生産力向上」があげられる。これは、GDPとしては純増だが賃金上昇の抑制等を伴

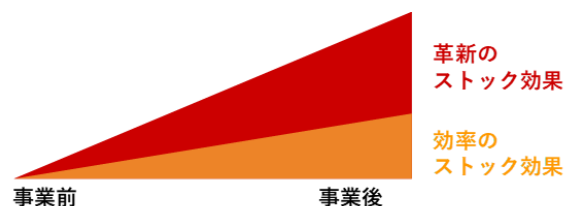


図 7.2 革新のストック効果と効率のストック効果

うのであれば、ストック効果が正でも、国民の厚生（このとき外国人は除く）は改善しないケースも考えられる。

これらのケースから想定されるように、TFPの上昇は、現実に計測された各地域の生産性の上昇と、従来のSCGEなどの帰着便益分析で予測された便益との差分として得られる、と考えられる。これらの議論をまとめて、図2.2に示す。なお、発生側便益を計測する現在の費用便益分析では、総交通量の増分と時間価値の増分が、革新のストック効果に相当すると考えられる。しかし、現行の費用便益マニュアルでは、総交通量一定、時間価値一定が仮定されているため、革新のストック効果が分離・計測されることはない。ただし事後評価で基準交通量を更新し、時間価値が更新されれば、その計測値がストック効果に含まれる可能性がある。ただしこの場合でも分離・計測されることはないため、この部分の計測には別の枠組みが必要となる。

革新のストック効果の計測方法は、まず、before時点データに基づくwith予測（=事業実施を想定したAfter時点予測）をSCGEなどによって求める。その上で、SCGEに関する感度分析を行い、地点別の帰着便益と観測便益を比較して、両者に有意差があれば、革新のストック効果が得られる。しかし、この方法には、以下の課題がある。

- I. SCGEモデルでは、分散等の指標で感度分析の精度をあらわすことが難しい。
- II. 経済理論で想定した因果構造の妥当性が検証できない。

上記の課題のうち、II.に関しては経済理論によらず、因果推論、因果探索を行う方法が必要となる。I.に関しては、サンプル数を確保した上で、誤差項を明示した回帰分析の枠組を援用し、帰着便益の予測値ではなく、観測便益の再現値に幅を持たせることによって、両者の有意差を算出できる。前章で開発したLINGAMは、これらの課題を克服できる手法になっている。

なお先述の通り、革新のストック効果を求める

には、帰着側の便益を計測する従来手法によって効率のストック効果を算出し、その結果と統計的因果推論から得られるストック効果（革新のストック効果と効率のストック効果の両方を含む値）を、比較する必要がある。

なお両手法の比較によって得られる知見は、革新のストック効果によって得られる便益を明らかにできることに留まらない。社会基盤のストック効果は、時にストック効果の空間的な移転や時間的な遅れを伴うため、両手法の推計結果が、完全に一致することは稀であろう。両者の相違を明らかにすることは、各地域のストック効果の発言状況について理解を深めるとともに、適切な地域診断と政策立案に関して、有益な知見が得られると思われる。

また統計的因果推論によって得られるストック効果の発生メカニズムを、経済学的手法によって想定していたストック効果の発生メカニズムと比較することもできる。この比較からは、経済学的な予測モデルの前提となっているわが国の産業連関構造や技術水準に関して、その変化の方向性を明確にできるため、経済学的なストック効果の計測手法の発展に寄与する知見が得られる。またストック効果をもたらすインフラの改善、たとえば地域間アクセスの向上は、そのまま各地域の経済水準の改善に結びつくわけではなく、地理・人口・産業などの様々な地域特性との交絡を経て発現する。これらの要因との交絡構造を明確にすることは、統計的なストック効果の検証手法の改善に資する。

## 参考文献

- 1) 瀬尾亨, 日下部貴彦, 朝倉康夫: 車間距離を計測するプローブカーを前提とした交通状態の推定手法, 土木学会論文集 D3, Vol.69, No.5, pp.I\_809-I\_818, 2013.
- 2) 和田健太郎, 小林桂子, 桑原雅夫: プローブデータ車両軌跡データのみを用いた交通の量的把握, 生産研究, 67 巻, 2 号, pp.143-147, 2015.
- 3) Feifei He, Xuedong Yan, Yang Liu, Lu Ma : A Traffic Congestion Assessment Method for Urban Road Networks Based on Speed Performance Index, *Procedia Engineering*, 137, pp.425-433, 2016.
- 4) Sandra Hohmann and Justin Geistefeldt : Traffic Flow Quality from the User's Perspective , *Transportation Research Procedia*, Vol.15 , pp.721-731, 2016.
- 5) Yasuo Asakura, Takahiko Kusakabe, Long Xuan Nguyen, Takamasa Ushiki : Incident detection methods using probe vehicles with on-board GPS equipment, *Transportation Research Part C*, 2016.
- 6) 牛木隆匡, 日下部貴彦, NGUYEN Xuan Long, 朝倉康夫: 交通容量低下を考慮したプローブカーデータによる突発事象検出手法, 土木学会論文集 D3, Vol.70, No.5, pp.I\_1147-I\_1157, 2014.
- 7) 日下部貴彦, 牛木隆匡, 朝倉康夫: 衝撃波速度を用いたプローブカーデータによる突発事象検出手法, 土木学会論文集 D3, Vol.71, No.5, pp.I\_827-I\_837, 2015.
- 8) Hyoshin Park, Ali Haghani : Real-time prediction of secondary incident occurrences using vehicle probe data, *Transportation Research Part C*, 70, pp.69-85, 2016.
- 9) Akira Kinoshita, Atsuhiko Takasu, Jun Adachi : Real-time traffic incident detection using a probabilistic topic model , *Information Systems*, 54, pp.169-188, 2015.
- 10) Eleonora D'Andrea, Francesco Marcelloni : Detection of traffic congestion and incidents from GPS trace analysis , *Expert Systems With Applications*, 73, pp.43-56, 2017.
- 11) Ruimin Li, Francisco C. Pereira, Moshe E. Ben-Akiva: Competing risk mixture model and text analysis for sequential incident duration prediction, *Transportation Research Part C*, 54, pp.74-85, 2015.
- 12) Ahmad Tavassoli Hojati, Luis Ferreira, Simon Washington, Phil Charles, Ameneh Shobeirinejad: Modelling the impact of traffic incidents on travel time reliability, *Transportation Research Part C*, 65, pp.49-60, 2016.
- 13) 藤井大地, 塩見康博, 宇野伸宏, 嶋本寛, 中村俊之: フィードバック型交通状態推定手法の事故発生状況分析への適用可能性, 土木学会論文集 D3, Vol.70, No.5, pp.I\_1067-I\_1076, 2014.
- 14) 兵頭知, 吉井稔雄, 高山雄貴: 車両検知器の 5 分間データを利用した交通流状態別事故発生リスク分析, 土木学会論文集 D3, Vol.70, No.5, pp.I\_1127-I\_1134, 2014.
- 15) Irem Önder, Dr: Classifying multi-destination trips in Austria with big data, *Tourism Management Perspectives*, 21, pp.54-58, 2017.
- 16) Janika Raun, Rein Ahas, Margus Tiru: Measuring tourism destinations using mobile tracking data, *Tourism Management*, 57, pp.202-212, 2016.
- 17) Jameson L. Toole, Serdar Colak, Bradley Sturt, Lauren P. Alexander, Alexandre Evsukoff, Marta C. González: The path most traveled: Travel demand estimation using big data resources, *Transportation Research Part C*, 58, pp.162-177, 2015.
- 18) 清家剛, 三牧浩也, 原裕介, 小田原亨, 永田智大, 寺田雅之: まちづくり分野におけるモバイル空間統計の活用可能性に係る研究, 日本都市計画学会都市計画論文集, No.46-3, pp.451-456, 2011.
- 19) 清家剛, 三牧浩也, 森田祥子: モバイル空間統計を活用した都市拠点地区の人口特性分析に係る研究, 日本建築学会計画系論文集, 第 80 巻, 第 713 号, 1625-1633, 2015.
- 20) 田中敦士, 岡本直久, 鈴木俊博, 浅野礼子, 白川洋司: 人口分布統計データを活用した観光地の特性把握, 土木計画学研究・講演集, Vol.54, 2016.
- 21) 菊池信輝: トラベルコスト法を用いた公立図書館の便益計測とその検証 :-岩手県立図書館の移転新築事業を事例として-, 総合政策, 第 9 巻, 第 1 号, pp.69-83, 2007.
- 22) 渡邊正英: トラベルコスト法における代替地価格問題の Multiple Indicator による解決-静岡県大井川上流部を事例として-, 農業経済研究, 第 75 巻, 第 4 号, pp.177-184, 2003.
- 23) 百瀬恵夫: 中小工業団地の理論と政策, 白桃書房, 1979.
- 24) 伊藤亮: わが国の工業団地開発戦略における現状と課題に関する一考察, 土木計画学研究・講演集, Vol.40, CD-ROM, 2009.
- 25) Ryan, S.: The value of access to highways and light rail transit: Evidence for industrial and office firms, *Urban Studies*, 42 (4), pp.751-764, 2005.
- 26) Ambrose, B.: An analysis of the factors affecting light industrial property valuation. *Journal of*

- Real Estate Research, 2009.
- 27) Beekmans, J., Beckers, P., van der Krabben, E., and Martens, K.: A hedonic price analysis of the value of industrial sites, *Journal of Property Research*, 31 (2), 108-130, 2014.
  - 28) Beekmans, J., van der Krabben, E., and Martens, K.: An indicator for decline of industrial estates, *Journal of European Real Estate Research*, 5 (3), 229-249, 2012.
  - 29) del Saz-Salazar, S., and García-Menéndez, L.: Public provision versus private provision of industrial land: a hedonic approach, *Land Use Policy*, 22 (3), 215-223, 2005.
  - 30) Clark, D., and Pennington-Cross, A.: Determinants of industrial property rents in the Chicago metropolitan area, *Regional Science and Urban Economics*, 56, 34-45, 2016.
  - 31) 豊田秀樹: データマイニング入門-R で学ぶ最新データ解析-, 東京図書, pp.83-98, 2008.
  - 32) 金明哲, 下川敏雄, 杉本知之, 後藤昌司: R で学ぶデータサイエンス 9 樹木構造接近法: 共立出版, pp.1-17, 2013
  - 33) Breimam, L. Friedman, J.H., Olshen, R.A.&Stone, C.J. : *Classification and Regression Trees* , Wadsworth, 1984
  - 34) William H. Press : *Numerical Recipes 3rd Edition: The Art of Scientific Computing* , Cambridge University Press, 2007
  - 35) 脇森浩志: ビッグデータに対するテキストマイニング技術とその適用例, *UNISYS TECHNOLOGY REVIEW* 第 115 号, pp.19-31, 2013.
  - 36) 稲田裕介, 中山晶一郎, 高山純一: 都市高速道路の時間・日交通量の変動特性と分布形状に関する一考察, *土木学会論文集 D3(土木計画学)*, Vol.67, No.5(土木計画学研究・論文集第 28 巻), pp.553-561, 2011.
  - 37) 国土交通省: ETC2.0 , [www.mlit.go.jp/road/ITS/j-html/etc2/jyouhou.html](http://www.mlit.go.jp/road/ITS/j-html/etc2/jyouhou.html)
  - 38) 末成浩嗣, 田中良寛, 橋本浩良, 高宮進: ETC2.0 プローブ情報から得られる経路データを利用した道路の利用特性把握, 第 53 回土木計画学研究発表会・講演集, pp.635-642, 2016.
  - 39) 丹後俊郎, 山岡和枝, 高木晴良: ロジスティック回帰分析 SAS を利用した統計解析の実態, 朝倉書店, 2013.
  - 40) 北村隆一, 森川高行, 佐々木邦明, 藤井聡, 山本俊行: 交通行動の分析とモデリング, 技報堂出版株式会社, 2002.
  - 41) 大久保街亜, 岡田謙介: 伝えるための心理統計, 勁草書房, 2012.
  - 42) Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*, New York: Academic Press, 1969.
  - 43) Lee, D. D. and Seung, H. S. : Algorithms for nonnegative matrix factorization, *Advances in Neural Information Processing Systems 13 (NIPS 2000)*, pp.556-562, 2000.
  - 44) Brunet, J.-P., Tamayo, P., Golub, T.R., Mesirov, J.P.: Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the national academy of sciences*, 101 (12), pp.4164-4169. 2004.
  - 45) Hutchins, L.N., Murphy, S.M., Singh, P., Graber, J.H.: Supplement Text and Figures: PositionDependent Motif Characterization Using Nonnegative Matrix Factorization. 2008.
  - 46) 吉田崇紘, 堤盛人: 人口構成比の観点からみた将来日本の縮図, *GIS 理論と応用*, Vol.25, No.2, pp.23-33, 2017.
  - 47) 緯度経度座標付き電話帳データベース, <http://www.zenrin.co.jp/product/gis/teldata/telpt.htm> (閲覧日 H31.1.29)
  - 48) 国土交通省・社会資本整備審議会計画部会専門小委員会: ストック効果の最大化に向けて~その具体戦略の提言~.
  - 49) 小池淳司: 道路のストック効果は計測可能なのか?, *高速道路と自動車*, 第 59 巻, pp.5-8, 2016.
  - 50) Aschauer, D.A.: Is public expenditure productive?, *Journal of Monetary Economics*, Vol.23, pp.177-200, 1989.
  - 51) 林正義: 社会資本の生産性と同時性, *ESRI Discussion Paper Series*, No.21, 2002.
  - 52) 江尻良, 奥村誠, 小林潔司: 社会資本の生産性と経済成長: 研究展望, *土木学会論文集 No.688/IV-53*, 75-83, 2001.
  - 53) 織田澤利守, 大平悠季: 交通インフラ整備効果の因果推論: 論点整理と展望, 第 58 回土木計画学研究発表会講演集 vol.58, S1, pp.1-13, 2018.
  - 54) Duranton, G., Turner, M.A.: Urban Growth and Transportation, *The Review of Economic Studies*, vol.79, Issue 4, pp.1407-1440, 2012.
  - 55) Holl, A.: Highways and productivity in manufacturing firms, *Journal of Urban Economics*, vol.93, pp.131-151, 2016.
  - 56) Redding, S.J. and Turner, M.A., *Transportation Costs and the Spatial Organization of Economic Activity*, *Handbook of Urban and Regional Economics*, vol.5, pp.1339-1398, 2015.
  - 57) 中村英夫: 道路投資の社会経済評価, 東洋経済新報社, 1997
- (土木学会論文集の完成版下印刷用和文原稿作成例を基本に作成)