

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. 25-3

研究テーマ

事故発生位置情報を用いた事故分析
統合システムの研究開発

研究代表者：	東京大学特任教授	山田 晴利
	兼交通事故総合分析センター研究部長	
共同研究者：	東京大学教授	柴崎 亮介
	東京大学客員研究員	Teerayut Horanont
	交通事故総合分析センター	西田 泰
	交通事故総合分析センター	下村 静喜
	交通事故総合分析センター	塩田 誠
	株式会社ゼンリン	知花 要

平成 2 8 年 7 月

新道路技術会議

目次

第 1 章	はじめに	3
1.1	研究の背景と目的	3
1.2	研究内容	3
第 2 章	経度・緯度の検証と修正システムの開発	5
2.1	事故原票データの概要	5
2.2	事故の属性データに関わる検討	7
2.3	経度・緯度の値の検証	8
2.4	事故位置修正システムの開発	12
第 3 章	交通事故・道路統合データベースシステムの作成およびデータ整備	14
3.1	交通事故・道路統合データベースシステム	14
3.2	事故分析のためのデータ整備	20
第 4 章	交通事故の発生要因とクラスタリングの分析	24
4.1	交通事故の発生要因の分析	24
4.2	交通事故のパターン分析	35
4.3	交通事故のクラスタリング分析	39
第 5 章	本研究のまとめ	49
5.1	総合的な事故分析システムの開発	49
5.2	交通事故の分析	50
5.3	研究成果の普及	50
5.4	今後の課題	50
参考文献		52

「道路政策の質の向上に資する技術研究開発」(平成 25 年度採択)
研 究 概 要

番号	研究課題名	研究代表者
No.25-3	事故発生位置情報を用いた事故分析統合システムの研究開発	東京大学 空間情報科学研究センター 特任教授 山田 晴利

交通事故位置情報の品質管理、沿道の建物用途、土地利用、プローブデータ、天候等と連携した多角的、複合的な事故分析を任意の地点・地域で行えるようにするため、事故データの品質管理から分析までを実行できる総合的な事故分析システムを新たに開発することを目的とし、さらに事故発生場所の経度・緯度情報を使った事故分析を実施する研究開発。

1. 研究の背景・目的 (研究開始当初の背景・動機、目標等)

2012 年から事故原票に事故発生位置の経度・緯度が付与されるようになり、「すべての道路と地域」を対象にして、従来は不可能だった詳細な事故分析を行うことが原理的には可能になった。しかし、既存の事故分析システムは事故発生場所の位置情報を扱えず、また毎年 50 万件を超える人身事故に付与された位置情報の品質管理・修正支援も困難である。本研究開発では、GIS を援用して、事故データの品質管理から分析までを実行できる総合的な事故分析システムを新たに開発することを目的とする。これによって、事故位置情報の品質管理、沿道の建物用途、土地利用、プローブデータ、さらには天候や運転者の属性、車両属性等と連携した多角的、複合的な事故分析などこれまで局所的にしか行えなかった総合的な事故分析を任意の地点・地域で行うことが可能となる。こうした分析結果を交通安全対策に反映させることにより、より効果的な対策の実施、ひいては交通の安全性向上につながる。

2. 研究内容 (研究の方法・項目等)

- (1) 経度・緯度の正確さを確認し、正確ではないデータを修正するシステムを開発すること
- (2) 交通事故を道路のリンク・交差点単位に集計するシステムを開発すること
- (3) 集計された事故データをもとにした事故分析用のシステムを開発すること
- (4) 事故分析に用いる種々の地理空間関連データを整備すること
- (5) 交通事故の発生要因とクラスタリングについて分析すること
- (6) 分析結果を広く社会に公開し交通安全意識の啓発をはかり、安全性の向上に努めること

3. 研究成果 (図表・写真等を活用し分かりやすく記述)

- (1) 毎年提供される交通事故発生場所の経度・緯度データの正確さをチェックし、位置が不正確なデータについては開発した修正システムを使ってデジタル道路地図ネットワーク上の正しい位置を求めた。
- (2) 過去の事故も含めて ITARDA 区間(平均長 300m)と交差点ごとに交通事故を集計するシステムを開発し、分析に必要なデータを整備した。
- (3) 集計した事故データを使って走行台キロ当たりの交通事故件数等を算出し、事故対策の企画・立案に使うためのシステムを道路管理者に提供した。
- (4) 事故分析に使うために必要なデータを幅広く収集した。
- (5) 事故要因の分析では、一般化線型モデル (GLM)、ゼロ強調モデル、さらには CAR モデルの推定を行った。交通事故のクラスタリングについては、死亡事故、重傷事故と多発箇所の分布パターンを比較し、これらのパターンの間に差があることを見出した。また、交通事故が集中しているクラスターの抽出を行った。
- (6) 事故多発箇所データをカーナビメーカー・車メーカーに提供したところ、カーナビで利用されている。また、生活道路の事故を 500m メッシュで集計した結果をホームページに掲載した。

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）

- 山田晴利: 高齢者の交通事故に関する基礎的分析, 第33回交通工学研究発表会論文集CD-ROM, pp. 23-27, 2013年9月, 交通工学研究会
- 山田晴利, Horanont T, 田中祥夫, 柴崎亮介: 交通事故発生場所の経度・緯度の精度検証と事故分析システムの開発, 土木計画学研究・講演集CD-ROM, 第49巻, 9頁, 2014年6月, 土木学会
- 山田晴利, Horanont T, 田中祥夫, 柴崎亮介: 交通事故の発生場所情報を用いた事故分析, CSIS全国共同利用研究発表大会, CSIS Days 2014
- 山田晴利, 柴崎亮介: 交通事故の空間分布パターン, CSIS全国共同利用研究発表大会, CSIS Days 2015
- Songpatanasilp P, Yamada H, Horanont T & Shibasaki R: Traffic accidents risk analysis based on road and land use factors using GLMs and zero-inflated models, Paper #320, 26 pages, CUPUM 2015 CD-ROM, 2015
- Chen Q, Song X, Yamada H & Shibasaki R: Learning deep representation from big and heterogeneous data for traffic accident inference, *The proceedings of AAAI 2016 CD-ROM*, The Association for the Advancement of Artificial Intelligence, 2016
- 山田晴利, 下村静喜, 田中祥夫, 柴崎亮介: 交通事故発生地点の経度・緯度情報を用いた事故分析, 土木計画学研究・講演集CD-ROM, 第53巻, 4頁, 2016年5月, 土木学会

5. 今後の展望（研究成果の活用や発展性、今後の課題等）

交通事故の分析には、発生場所の情報を付与された交通事故データが必須である。発生場所の経度・緯度の正確さについては漸次改善されてきてはいるものの、今後とも継続的にチェックを続け正確さを確保していく必要がある。

また分析に使われる国勢調査、経済センサス等の社会・経済データ、POIやデジタル道路地図（DRM）等の空間情報についても定期的な更新が必要であり、こうした更新業務を継続して実施していく。

交通事故分析に使われるデータの中には、プライバシー保護、知的財産権等の絡みで広く一般に公開できないものが含まれている（事故の個票データ、DRM等がこれに該当する）。こうした状況に鑑み、個人情報秘匿された分析結果をWeb等を通じて公開するようにしていきたい。

交通事故についてはこれまで幹線道路での事故に重点が置かれていたが、近年は生活道路での事故にも関心が持たれるようになってきているので、生活道路での事故と地域特性、地理空間特性との分析をさらに深化させていく考えである。特に交通事故発生場所の空間的なクラスタリングに注目した研究が行われるようになってきているので、こうした分野の研究も実施し効果的な交通安全対策の実施につながる成果を目指したい。

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）

本研究の成果は、すでに道路管理者に提供する交通事故統合データベースの作成とその分析システムに反映されており、道路交通安全の実務において広く活用されている。さらに道路行政の観点からは、従来警察から提供を受けて地図上に落としていた事故発生場所を発生場所の経度・緯度を使ってほぼ自動的に取得することができるようになったことから、コストの削減につながっている。

また近年関心が高まってきている生活道路での事故については、メッシュ単位での集計を行い、その結果を都道府県に提供しただけではなく一般にも公開した。こうした取り組みは従来の分析の枠を越えるものであり、提供するデータの種類と数を拡大していきたい。

7. ホームページ等（関連ウェブサイト等）

- www.itarda.or.jp/service/webmap.php : ITARDAのホームページに生活道路の事故を500mメッシュで集計した結果を掲載した

第 1 章

はじめに

1.1 研究の背景と目的

2012 年 4 月に京都府亀岡市で登校中の児童と保護者の列に軽自動車がつっこみ、3 人が亡くなり 7 人が重軽傷を負うという痛ましい事故が起き、通学路での事故対策に注目が集まった。しかしながら、これまでの事故データには発生位置情報が付与されていないため、通学路における事故を体系的に集計・分析することは困難であった。

しかるに、2012 年から人身事故データに事故発生位置の座標（経度、緯度）が付与されるようになったため、「すべての道路、地域」と「すべての人身事故」を対象にして交通事故の分析を行うことが可能になり、従来は実施できなかった詳細な事故分析を行う環境が整いつつある。たとえば、上で述べた通学路における事故分析、ゾーン 30 設置前後の事故の比較はその一例であるが、これだけにとどまらず車道に自転車通行レーンを設置したことによる事故の件数や形態の変化、事故多発地点における対策前後の事故の比較等も対象路線あるいは地点を指定することで簡便に行うことができるようになる。また、沿道土地利用・建物用途と事故との関連、さらにはプローブカーやドライブレコーダーによって、あるいは住民から収集されたヒヤリ・ハットデータと事故データとの関連性の分析も可能になる。ヒヤリ・ハットの多発地点が事故多発地点と重なっていることがわかれば、ヒヤリ・ハットデータを使って予防的な事故対策をとることも可能になる。一方、ヒヤリ・ハットの発生件数は多くないが、事故が多発している地点が存在するのであれば、そうした地点での対策を急ぐ必要がある。

しかしながら、従来の事故分析システムは事故発生地点の位置情報を扱えず、また毎年約 50 万件超に上る人身事故に付与された位置情報の品質管理・修正支援も困難である。本研究開発では GIS を援用し、データの品

質管理から分析までをカバーする総合的な事故分析システムを新たに開発することを目的とする。これによって沿道の建物用途、土地利用、プローブカーやドライブレコーダー等のデータ、さらには天候や運転者の属性、車両属性と連携した多角的、複合的な事故分析等、従来ごく局地的にしか行えなかった統合的な事故分析が全国の任意の地域と道路で可能となる。

この分析システムを用いれば、道路リンク、地点あるいは地域における事故特性を詳細に把握することができ、たとえば当該箇所に接近しつつある運転者にカーナビ等を通じて事故の多発する曜日、時間帯、天候等に応じた警告を発することも可能になるので、事故の低減につながると期待できる。

このシステムはわが国で発生した人身事故を網羅する大規模なシステムであり、道路管理者が行う事故分析、事故対策の立案と評価における利用をターゲットとするだけではなく、交通事故の研究者や自治体の事故対策担当者をはじめとする利用者からの要請にもとづいて集計したデータを提供する仕組みについても検討を行い、試行を実施する予定である。

これらの目的を達成するため、本研究では、道路管理者、警察の事故分析担当官、車メーカー、カーナビメーカー等の協力を得て、事故低減に向けて実効性の高い分析を行うことができるシステムを開発することを目的としている。

1.2 研究内容

これまでわが国では、交通事故の発生場所に関するデータは限られた形でしか利用できなかった。警察署の協力を得て発生場所の情報を収集するという方法はあったが、限定された地域であればともかく、全国の事故を対象にして発生場所の情報を収集することはほとんど不可能であった。

一方道路管理者は、一般都道府県道以上の道路を対象に警察から発生位置の情報を取得し、公益財団法人交通事故総合分析センター（以下、ITARDA と略記する）が有する事故原票データにある事故属性情報との突き合わせ（「マッチング」と称する）を行って事故データを整備してきた。ただし、このデータには高速道路、自動車専用道路で発生した事故の情報が含まれていない、事故の発生位置が ITARDA 区間と呼ばれる集計用の区間で与えられているがこの区間の情報が一般には公開されていない、など使い勝手の悪い面があった。

しかしながら、2012 年から事故発生場所の経度・緯度が事故原票に記載されるようになり、事故分析を取り巻く状況は一変した。すなわち、全ての人身事故について発生場所の経度と緯度が与えられるようになったため、地理空間情報と連携させてさまざまな分析を行うことが可能になり、従来道路管理者が実施してきたマッチング作業は基本的に不要になる。しかも、任意の地域と道路を対象にこうした分析が行えるのである。

ただしこれらを実現するためには、事故位置として正確な経度・緯度の値が与えられていることが大前提である。したがって、提供された経度・緯度の正確さを検証し、必要に応じて経度・緯度の値の補正を行うとともに、補正が困難なほど正確さの劣るデータを異常値として抽出するといったデータ修正システムの構築が必要である。

さらに、事故分析のためには種々の地理空間情報が必

要であり、地理空間情報と事故発生位置情報を結びつけて分析するためのシステム構築も求められる。地理空間情報はデータ量が多く必ずしも取扱いが容易ではないため、使いやすい利用環境の整備が望まれる。

なお事故分析用のシステムとしては、これまで道路管理者用に提供してきたシステムを代替するシステムに加えて、より汎用的に利用可能なシステムも開発する予定である。

以上を踏まえて、本研究開発では以下の項目について調査・検討を行うこととした。

- 事故原票に付与された経度・緯度の正確さを確認し、正確ではないデータを修正するシステムを開発すること
- 交通事故を道路のリンク・交差点単位に集計するシステムを開発すること
- 集計された事故データをもとにした事故分析用のシステムを開発すること
- 事故分析に用いる種々の地理空間関連データを整備すること
- 交通事故の発生要因とクラスタリングについて分析すること
- 事故分析の結果を広く社会に公開し、交通安全意識の啓発を図るとともに安全性の向上に努めること

第2章

経度・緯度の検証と修正システムの開発

本章では主に経度・緯度の正確さの検証方法と修正システムの開発について記載する。この検証の目的は、交通事故の発生場所の座標をもとにして事故位置をデジタル道路地図（以下、DRM と略記する）の正しい場所に落とすことである。これは、事故と DRM のもつ道路の情報（道路種別、路線番号、道路管理者、センサス交通量等）とを結びつける上で不可欠なプロセスである。道路管理者に DRM ネットワークの上に「正しく」のせられた事故データを提供するためには、経度・緯度の値を入手したあとでこの操作を 1 回行えばよい。

この検証では、事故原票に記載されている事故の属性（とくに道路種別）と DRM の道路種別の対応を考慮する必要があるので、最初にこうした道路種別の対応関係を整理する（第 2.1 節）。

さらに、道路管理者に提供する事故の属性データを事故原票のどの情報から入手するかについて、第 2.2 節で記載する。

次に事故発生場所の経度・緯度の値の正確さを検証するためのシステムについて述べ、あわせて検証結果を提示する（第 2.3 節）。

2.1 事故原票データの概要

事故原票で用いられている道路種別と DRM の道路種別との間には差異がある。以下に示すように、事故原票の道路種別は細分化されているが、DRM の道路種別はそれほど細分化されていない。このため、両者の間の対応づけを検討する必要がある。

事故原票で用いられている道路分類は下記のとおりであり¹⁾、括弧内の数字は原票における路線番号がとる値を表す。

- 一般国道（0001～0999）：番号は国道の路線番号に対応
- 主要地方道
 - 都道府県道（1000～1499）

- 市道（1500～1999）
- 一般都道府県道（2000～2999）
- 一般市町村道（3000～3999）
- 高速自動車国道（4000～4999）
- 自動車専用道路
 - 指定区間（5000～5499）
 - その他区間（5500～5999）
- 道路運送法上の道路（6000～6999）
- 農（免）道（7000～7999）
- 林道（8000～8499）
- 港湾道（8500～8999）
- 私道（9000～9499）
- その他（9500）
- 一般の交通の用に供するその他の道路（9900）：広場、駐車場等を指す

最後の「その他」と「一般の交通の用に供するその他の道路」は、9499 以下のいずれのコード番号にも該当しない道路であり、広場、サービスエリア、パーキングエリア、道の駅、駐車場等を意味する。こうした場所で発生した交通事故を道路ネットワークに紐付けても無意味なので、コード番号が 9500 と 9900 の事故については、マッチング対象としないこととした。

事故原票ではそれぞれの道路種別に対して上記のように 500 個あるいは 1000 個のコードが用意されているが、すべての道路種別において個別の道路にユニークな番号が付与されているわけではない。幹線系の道路（高速道路、一般国道、主要地方道等）ではそれぞれの道路に対してユニークなコードが与えられているが、非幹線系の道路では個別の道路にユニークなコードが与えられておらず、例えば一般市町村道についてはすべて同じ番号（たとえば 3100）が与えられているケースも存在する。

事故原票では DRM よりも細分された道路種別が用い

られているが、この分類は「相互に排他的」になっていないという問題を抱えている。たとえば、「自動車専用道」は他の分類と重複する道路を含んでいる。「道路運送法上の道路」や「高速道路」は「自動車専用道」でもあり、また「一般国道」、「主要地方道」の中には「自動車専用道」がある。新直轄方式で整備された道路は「高速道路」であり「自動車専用道」でもある。このため、DRMの道路種別と一対一に対応づけることはむずかしい。

一方、DRMの道路種別は次のように分類されている²⁾。括弧内の数字はDRMの種別コードである。なおDRMでは、道路の路線番号はそれぞれのリンクの属性情報としてリンク毎に与えられている。

- 高速自動車国道 (1)
- 都市高速道路 (2)
- 一般国道 (3)
- 主要地方道：都道府県道 (4)
- 主要地方道：指定市道 (5)
- 一般都道府県道 (6)
- 指定市の一般市道 (7)
- その他の道路 (9)
- 未調査 (0)

DRMの道路種別は原票用の道路分類より大括りになっていることがわかる。ただしDRMでは指定市の市道については、指定市道と一般市道に別々の番号が与えられておりこの種別については細かな分類となっている。

さらにDRMでは、市町村道、農（免）道、林道、港湾道、私道などはすべて「その他道路」としてひとまとめになっているため、これらの道路を区分することができない。また、DRMの「その他道路」では道路の属性情報が与えられておらず、単にリンクの始点・終点のノードとリンクが与えられているのみである。「道路管理者名」もっていないため、DRMネットワークから市町村道を抽出することができない。こうした点は、生活道路での事故を分析しようとする場合に問題となる。

「市町村道」の扱いについては、DRMを作成している日本デジタル道路地図協会の担当者とも相談したが、現状では市町村道を特定することは難しいという回答であった。このため、市町村道で発生した事故については、便宜的にDRMの「その他道路」と対応させることとした。

しかしながら、この方式では道路交通センサスの調査対象となっている政令指定市の市道で発生した事故をDRMとマッチングさせる場合に、センサス対象外の道路リンク・交差点とマッチングしてしまうという問題が

発生することが判明した。これは、市道のネットワークが稠密で、経度・緯度で与えられる事故発生場所の近傍に市道のリンクが複数存在するためである。

この問題を解決するために、

- 車道幅員が5.5m未満の道路はマッチング対象から外す
- 交差点で交差する道路の幅員がどちらも5.5m未満の交差点をマッチング対象から外す

という対応策をとることとした。これによって、マッチング候補となる道路リンク・交差点の数を減らすことができる。この方法でも問題の根本的な解決策にはなっていないが、これより優れた手段を見いだせなかった。

なお道路種別を判断する際に手がかりになる情報として、路線種別コードの他に道路名称がある。事故原票では道路名称テーブルが準備されているが、都市計画道路名が多く通称名はほとんどない（大都市の道路の一部には通称名が与えられている）。DRMでも道路名称が属性として与えられているが、通称名が主体となっており都市計画道路名称は少ない。さらに通称名も網羅的に与えられているわけではなく、都市計画道路名は普通の人にはなじみがない。こうしたことを考慮すると、道路名称を使って道路を識別することは難しいと判断した。

以上の検討結果を踏まえて、事故原票の道路種別とDRMの道路種別の対応づけを表2.1に従って行うこととした。ただし、この対応表を一律に適用するとマッチングできない道路が生じるため、道路種別を適宜拡大して（つまり、原票で与えられている道路種別とは別の道路も対象として）マッチングを行うこととした。

表 2.1: 事故原票と DRM の道路種別対応表

道路種別の対応	
DRM	事故原票
高速自動車国道	高速自動車国道
都市高速道路	自動車専用道－指定
一般国道	一般国道
主要地方道（都道府県道）	主要地方道（都道府県道）
主要地方道（指定市道）	主要地方道（市道）
一般都道府県道	一般都道府県道
指定市の一般市道	主要地方道－市道
その他の道路	一般市町村道
	道路運送法上の道路
	農（免）道
	林道
	港湾道 私道

交差点事故の扱い

交差点についても事故原票データと DRM 道路網との対応を考えておく必要がある。事故原票では、たとえ小規模な交差点であっても交差点で発生した事故は交差点事故として扱われている。一方、道路管理者に提供されている事故統合データベースでは、小規模交差点での事故は単路事故として扱われているので注意が必要である。

事故原票では交差点の規模が大・中・小に分類されているが、この分類は事故当事者が進入してきた道路の車道幅員（上り・下り合計の値）をもとにして行われているため、実際の交差点の規模を正確に反映しているわけではないことに注意する必要がある。たとえば、幅員の広い（幹線的な）道路と狭幅員道路が交差している交差点で、事故当事者がどちらも狭幅員道路から交差点に進入して事故が起きた場合には、小規模交差点で発生した事故とされる。したがって、事故原票に記載されている交差点規模をもとにして交差点を特定しようとする誤りがある。

また、事故原票では交差点の側端から 30 m 以内の道路の部分で発生した事故は「交差点付近」の事故として扱われるのに対して、道路管理者に提供されている統合データベースでは交差点中心から 35 m 以内で発生した事故は交差点事故として扱われているという違いがある。本研究では、事故原票にいう「交差点付近」の事故と「交差点」での事故をまとめて交差点事故として扱うこととした。これは、交差点側端の位置が通常の地図には記載されておらず、交差点側端から 30 m 以内という基準で交差点付近の事故かどうかを判定することができないためである。

交差点で発生した事故の場合、まず DRM の基本道路網の交差点ノードから 35 m 以内にあるかどうかをチェックし、次に基本道路網と細道路網の交差点ノードから 35 m 以内にあるかどうかをチェックするという手順をとった。いずれの場合も DRM の行き止まり点ノード、ダミー点ノード、区画辺交点ノード、属性変化点ノード等²⁾は交差点ノードと見なさず、交差点ノードのみを交差点として扱っている。

2.2 事故の属性データに関わる検討

これまで ITARDA から道路管理者に提供していた「交通事故統合データベース」には、事故属性として以下の項目が含まれていた。

- 事故発生場所の「上下区分」
- 発生場所のオフセット（道路端からの距離）
- 事故が発生した場所の道路部位
 - ▷ 横断歩道
 - ▷ 歩道等
 - ▷ 中央帯
 - ▷ 路肩
 - ▷ その他車道
 - ▷ その他
 - ▷ 不明
- 道路構造
 - ▷ 交差点大
 - ▷ 交差点中
 - ▷ 交差点小
 - ▷ 交差点付近
 - ▷ 本線（非分離）
 - ▷ 本線（分離）
 - ▷ ランプ
 - ▷ 側道
 - ▷ 踏切
 - ▷ 分岐
 - ▷ 合流
 - ▷ 不明
- 交差点番号

これらの属性について、事故原票に含まれているデータ、事故発生場所の経度・緯度の値をもとにしてマッチングした DRM リンクの属性情報等を使って作成できるかどうかを検討した。この結果、上で列挙した事故属性の中には作成可能なものと作成困難なものがあることが判明した。

「オフセット」は事故原票には含まれていない情報であり、作成は困難である。事故発生場所の経度・緯度の値が高精度で入力されており、さらに道路の幅員と線形についての精細な情報があれば、これらの値をもとにしてオフセットを計算することは可能であるが、現在のところこうした情報は存在しないため、オフセットの値を求めるのは不可能である。

しかし道路の交通安全担当者へのヒアリングによれば、オフセット情報が利用されるのは「衝突地点がどこなのか」を特定しようとする場合が多いということであった。衝突地点については事故原票の「衝突地点」情報からほぼ作成可能なので、あえてオフセットデータを作成する必要はないと考えられる。

「道路部位」のうち、「横断歩道」と「中央帯」以外の部位については、事故原票の「衝突地点」情報を使って

作成可能である。「横断歩道」については、原票の「事故類型」の中に「横断中」という項目が設けられており、この項目がさらに「横断歩道」「横断歩道付近」「横断歩道橋付近」「その他」に細分されているので、このデータを使えば取得ができる。残念ながら、「中央帯」については事故原票に記載されていないため、作成できない。

「道路構造」のうちの「交差点の大／中／小」と「交差点付近」「踏切」については事故原票に同じ項目があるので、作成が可能である。「本線（非分離／分離）」「ランプ」「側道」については、マッチングした DRM のリンク情報を使って作成可能である。一方、「分流」「合流」については、事故原票に記載がないため作成できない。

「上下区分」については、事故原票では直轄国道で発生した事故についてのみ情報が付与されている。これ以外の道路種別については、上下区分の情報は存在しないため、直轄国道での事故についてのみ「上下の別」の情報を作成することとした。

「交差点番号」は従来同一交差点で発生した事故を集約するために利用されていた。しかしこの情報が取得できなくなるため、ITARDA で定義している交差点番号で代用することとした。

最後に道路交通センサスの調査結果と交通事故データとの関連性について述べておく。これまで道路管理者に対して、道路交通センサスによって得られた自動車類の交通量（12 時間、24 時間交通量）を提供してきた。しかし、道路交通センサスではより詳細な交通量が計測されているので、これらと交通事故とを対応づけ、事故分析に利用できるようにした。

DRM の道路リンクには道路交通センサスの交通調査単位区間番号が付与されているので、経度・緯度の値をもとにしてマッチングした道路リンクからこの番号を取得し、道路交通センサスで計測された交通量と事故とを対応づけることとした。

この対応付けを行ったことによって、以下の交通量データの利用が可能になった。

- 時間帯ごとの自動車類交通量（乗用車／小型貨物車／バス／大型貨物車、12 時間 and/or 24 時間；平日休日別，上り下り別）
- 時間帯ごとの歩行者類交通量（12 時間 and/or 24 時間；平日休日別，上り下り別）
- 時間帯ごとの自転車類交通量（歩道／車道、12 時間 and/or 24 時間；平日休日別，上り下り別）
- 時間帯ごとの動力付き二輪車類交通量（12 時間 and/or 24 時間；平日休日別，上り下り別）

2.3 経度・緯度の値の検証

本節では、事故原票に記載された経度・緯度の精度を検証する方法とその結果について述べる。

正確さの検証は次の 2 段階に分けて行った。これは、2012 年にいくつかの県のデータを用いて予備的に経度・緯度の正確さを検証したところ、各都道府県のポリゴンから外れる事故が存在することが判明したためである。

- (1) 都道府県別に全ての人身事故を対象に、事故原票に記載された経度・緯度の値をもとにして事故発生箇所が各都道府県のポリゴンの中にあるかどうかを確認する。これは各都道府県の経度・緯度情報の大まかな精度をチェックするための検証である。
- (2) それぞれの事故を対象に、事故発生箇所の近傍に当該事故の発生した道路種別に該当する道路（単路上の事故の場合）あるいは交差点（交差点事故の場合）があるかどうかを検証する。近くに同じ種別の道路あるいは交差点が存在する場合には、DRM 上で事故発生場所から最近傍の位置にあるリンクあるいは交差点に事故を紐付けた。

ただし、道路以外の場所で発生した事故（事故発生場所が「その他道路」「一般交通の用に供するその他道路」となっている事故）については、この処理は行わない。

経度・緯度の精度を検証するに当たっては、事故原票に記載された経度・緯度の値（度分秒）を十進数に変換する、各都道府県のポリゴンを準備するなどの作業を行った。

また、事故が各都道府県のポリゴンの中に落ちるかどうかの判定には統計ソフト R の `maptools` ライブラリの `over()` 関数を用いた。

2.3.1 都道府県別の事故発生場所の検証

事故発生場所の経度・緯度の計測が始まった 2012 年には全国の 47 都道府県のうちの 43 都道府県で計測が行われた。残りの 4 県では計測が行われなかった。その後 2014 年に 1 県、2015 年に新たに 2 県でそれぞれ計測が行われるようになり、2015 年事故では合わせて 46 都道府県で経度・緯度の計測が行われた。

事故発生位置がそれぞれの都道府県のポリゴン内に落ちるかどうかを検証した結果は以下のとおりである。

図 2.1 には、都道府県別に 2012 年と 2013 年の事故を対象に経度・緯度の値から求めた事故発生場所がそれぞれのポリゴン内に落ちる割合をグラフ化した結果を示

した。

このグラフから以下がわかる。

- 全体としてみると、2012 年に比べ 2013 年のデータでは各都道府県のポリゴンの中に事故が落ちる割合が向上している。とくに、北海道、東北、関東、信越、北陸（図の上方部分）、および近畿地方（図の真ん中より少し下の部分）ではポリゴン内に落ちる割合が高い。
- これに対して、中国、四国、九州地方（図の下方部分）ではポリゴン内に落ちる割合が低い。とはいえ、これらの県でもポリゴン内に落ちる割合は 95% を超えている。
- ポリゴン外に落ちたデータは以下の三つに分類す

ることができる。すなわち

- ▷ 明らかにポリゴン外に位置するもの
- ▷ ポリゴンの境界近くに位置するが微妙にポリゴンから外れているもの
- ▷ 海上の橋の上等に位置するもの

図 2.2 には微妙にポリゴンから外れた事故と橋の上に落ちた事故の例を掲げた。

ポリゴン内に落ちる割合の低い県では、事故発生場所の経度・緯度の取得に GPS レシーバーあるいは 110 番マップを利用していることが多い。GPS レシーバーについては、マルチパスの影響や読み取り時の読み誤り、転記ミス等によって誤差が発生している可能性がある。

事故発生場所が各県のポリゴン内に落ちる割合

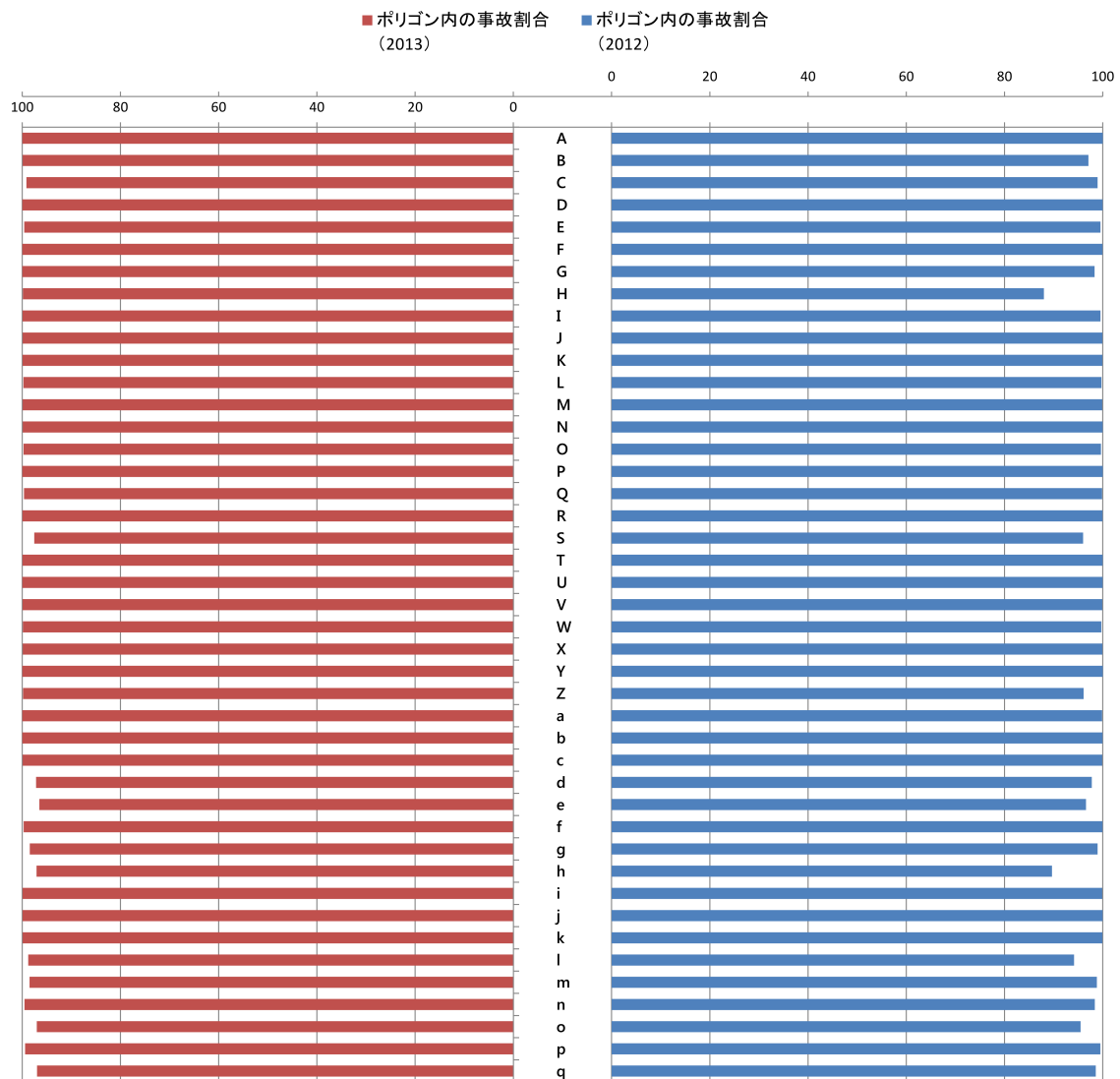
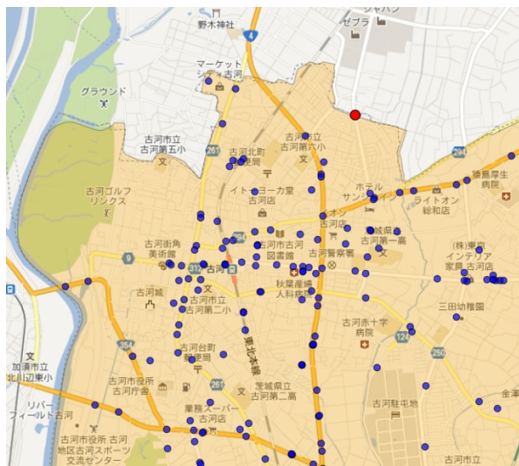


図 2.1: 事故位置が当該都道府県のポリゴン内に落ちる割合（左：2013 年，右：2012 年）



(a) 微妙にポリゴンから外れた事故の例（茨城県）：赤丸はポリゴン外と判定された事故。青丸はポリゴン内と判定された事故。
(b) 海上の橋の上で発生した事故の例（神戸市）：赤丸の事故はポリゴン外と判定されているが、その多くは橋上にある。

図 2.2: 微妙にポリゴンから外れた事故および海上の橋の上で発生した事故の例

「ポリゴンの境界近くに位置するが微妙にポリゴンから外れている」事故は、ポリゴンの誤差によってポリゴン外と判定された事故である可能性を排除できない。また「海上の橋の上等に位置する」事故は、海上の橋等が都道府県のポリゴンに含まれていないためポリゴン外と判定された可能性が高く、実際には正しい経度・緯度の値が附与されて事故も多いと考えることができる。

したがって、この二つのカテゴリーに分類された事故については一概にポリゴン外と判断せず、地図と照合した上でポリゴン外かどうかを決定する必要がある。つまり、事故が都道府県のポリゴンの中に位置しているかどうかの判定をすべて自動で行うことはできず、一部目視で確認しなくてはならないということである。

ただし、「海上の橋の上等に位置する」事故に関しては、DRM ネットワークへのマッチングでは問題が生じない。というのは、DRM ネットワークでは海上の橋も道路の一部として記載されているからである。

2.3.2 DRM ネットワークを使った検証

次に経度・緯度で与えられる事故位置の近傍に同種別の道路・交差点が存在するかどうかを都道府県別に検証した。この検証には ArcGIS を用いた。

表 2.2 には、一般国道および都道府県道で発生した交通事故（2014 年）を対象に、経度・緯度で与えられる事故発生位置の近傍に同種別の道路が存在するかどうかを検証した結果を示した。（都市部で発生した事故については、半径 50m の円内に同種別の道路が存在するかど

うかを、地方部で発生した事故については半径 100m の円内に同種別の道路が存在するかどうかで判定した結果である。）

表 2.2 によると、道路種別ごとのマッチング率には都道府県によってばらつきがあることがわかる。経度・緯度で与えられる事故発生場所が高い割合で当該都道府県のポリゴンに入っていた県であっても、同種別への道路のマッチング率が低いところが存在している（岩手県、山形県、栃木県）。この原因については、現時点でも不明である。

一方、多くの都道府県では道路種別ごとのマッチング率が高く、一般国道では 99% を超えるマッチング率を示す府県も存在している（北海道、青森県、宮城県、福島県、茨城県、千葉県、神奈川県、静岡県、愛知県、三重県、滋賀県、京都府、兵庫県、奈良県、和歌山県、鳥取県、島根県、岡山県、徳島県、高知県、佐賀県、および大分県）。

これらの結果を踏まえて、2015 年の業務では表 2.2 で網掛けをした 32 都道府県について、従来警察から事故発生場所の情報を入手して行っていた「マッチング作業」を事故発生場所の経度・緯度情報を使って代替することとした。これによってマッチング作業にかかるコストの削減につながる。選ばれた 32 都道府県は一般国道、主要都道府県道および一般都道府県道のマッチング率がいずれも 90% を超えている。

表 2.2 の中の記号 L で示される県については、一般都道府県道のマッチング率が低い、これは一般都道府県

道から主要地方道に区分変更した道路が適切に分類されていなかったことが原因であった。道路の種別を適切に設定し直したところマッチング率が向上したため、対象に含めた。

2014 年業務で「マッチング作業」を省略した県の数が 9 県であったことと比べると、2015 年業務では対象都道府県の数が大幅に増加した。2016 年業務で「マッチング作業」を省略した県数は 36 とさらに増加したことを付記しておく。

2015 年度にはマッチング作業を省略する都道府県の数が増加し、対象都道府県の中には政令指定市を含むところも存在する結果となった。このため、政令市の市道を抽出し、当該道路で発生した事故を統合データ

ベースに含めることが求められた。これは従来から道路管理者に提供している事故統合データベースに、政令指定市の市道で発生した交通事故が含まれているためである。しかし、市道のネットワークはかなり密に構成されており、この作業のすべてを自動化することができず、一部目視によって対象道路の探索を行わざるをえなかった。

なお、マッチング作業の省略に伴って従来道路管理者に依頼して実施していた道路種別・管理者の確認作業も省略できるようになった。この確認作業によって、道路種別・道路管理者の変更、新設道路の追加、廃止となった道路の削除等の道路ネットワークの更新に必要なデータを収集し統合データベースに反映していたのである。

表 2.2: 都道府県別の一般国道，都道府県道のマッチング率（2014 年）

都道府県	一般国道 (%)	主要都道府県道 (%)	一般都道府県道 (%)
A	99.3	95.2	90.5
B	99.7	98.0	97.3
C	40.5	33.6	40.8
D	99.3	97.2	97.6
E	56.7	57.3	46.6
F	99.6	97.8	95.3
G	99.6	97.7	96.8
H	92.5	88.4	84.8
I	97.3	98.2	92.6
J	98.9	96.0	95.4
K	99.0	96.1	94.3
L	98.8	90.4	68.8
M	99.1	98.1	97.0
N	97.7	96.1	94.1
O	94.7	95.0	95.4
P	98.5	98.2	94.0
Q	81.0	77.3	75.4
R	98.9	94.7	94.4
S	99.2	94.6	94.4
T	99.6	98.6	97.2
U	99.0	98.3	96.6
V	99.9	98.7	98.3
W	99.8	98.9	95.9
X	96.2	95.1	93.6
Y	99.7	97.2	92.6
Z	99.5	99.1	98.5
a	99.8	98.3	95.8
b	99.1	97.1	97.0
c	99.9	99.5	97.3
d	99.5	97.5	96.9
e	69.1	64.8	63.9
f	65.3	70.0	60.8
g	99.5	98.7	98.8
h	99.0	96.1	86.5
i	86.3	83.1	57.6
j	99.9	98.8	99.3
k	96.9	94.6	93.1
l	99.9	99.7	98.8
m	96.1	95.7	91.8
n	35.0	27.9	39.8
o	99.1	98.3	68.1
p	80.0	76.4	74.4
q	94.9	95.0	92.4
r	80.4	79.2	80.7
全 国	91.4	89.8	85.6

しかしこうした情報の収集が困難になったため、統合データベースの作成に用いる DRM を毎年更新し、DRM から情報を収集することで対応する必要が生じた。従来は道路交通センサスの実施にあわせて 5 年ごとに DRM を更新していたのであるが、5 年間隔の更新では道路ネットワークの最新情報が入手できないため、DRM も毎年更新することを予定している。

2.4 事故位置修正システムの開発

事故位置修正システムは、事故原票に記載されている経度・緯度の値をもとにしてその近傍に存在する DRM のリンクあるいは交差点を探索し、事故と紐付けることを目的としている。本節では、ここまでの記述でも何度か引用した「事故位置修正システム」がどのようにして事故位置を修正し、事故分析に必要な情報を取得しているのかを解説する。

DRM は、「当初は国土地理院の 1/25000 の地形図をもとに作成され、その後は全国の地方整備局、都道府県、政令指定都市、高速道路会社、道路関係公社など全国の道路管理者の資料と新刊地形図によりデータの更新」が行われている（DRM 協会のサイトの解説³⁾による）。

一方、事故の発生場所の経度・緯度の値は地図メーカーの手になる地図あるいは GPS レシーバーを使って取得されているため、そのままでは必ずしも国土地理院の地図をもとにした DRM のネットワークの上にのらない。また、地図上では道路は幅をもった帯で表現されているが、DRM では道路ネットワークは線で表現されていることも事故の発生場所が DRM の上にのらない理由の一つである。

しかしながら、DRM のリンクには種々の属性情報が付与されており、とくに交通量のデータは事故分析に欠かすことができない。このため、事故を DRM のリンクと紐付け、こうした属性情報を取得することは事故分析のために不可欠である。

DRM の上に事故をのせるための基本的な方法は、事故の近傍に存在するリンクあるいは交差点で事故に付与された「道路種別」と同じ種別のものを探索するというやり方である。しかしこの作業を行う際には、DRM と事故原票の仕様の違いを把握しておく必要がある。道路種別が DRM と事故原票の間で異なっているという問題については、すでに第 2.1 節で述べたのでここでは繰り返さない。もうひとつ考慮する必要があったのは、道路の上り・下りの区分である。

事故原票では、国道で発生した事故に限ってではある

が、上り・下りの情報が付与されている。一方、DRM では上下分離された幹線道路を二条線で表現している。こうしたことから、原票に付与された上り・下りの情報をできる限り生かすようにした。このときに問題となったのは、DRM では上り・下りのペアとなるリンクの情報が与えられていないという点である。より具体的に述べれば、事故の近傍に存在する同じ道路種別のリンクが見つかったとしても、上り・下りの区分が一致しなかった場合には上り・下りの方向が一致するリンクを再度探索しなくてはならない。上り・下りのペアとなるリンク情報があれば再探索は不要となるので、DRM 協会に対してそれぞれのリンクに上り・下りのペアとなるリンクの情報を付与してくれるよう要請したが、まだ実現していない。上り・下りのペアリンクの問題は、上り・下りの事故件数や交通量を合計して路線としての値を求める場合にも発生するので、早急な解決が望まれる。

2.4.1 システムの概要

本システムでは、日本全国の交通事故を対象にしてそれぞれの事故の近傍に存在する同種別の道路リンクあるいは交差点をまとめて計算している。これはこの計算を 1 回行って交通事故と道路ネットワークを紐付けておけば、以後同じ計算を行う必要はないからである。

ただし第 2.3.1 節で述べたように、それぞれの都道府県の事故が当該都道府県の中に収まっておらず他の都道府県に入っていることもあるので、他県に存在すると判断された事故については、計算の対象から除外した。

本システムは ArcGIS のアドオンソフトとして動作する。システムを起動すると図 2.3 に示した初期設定画面が現れる。

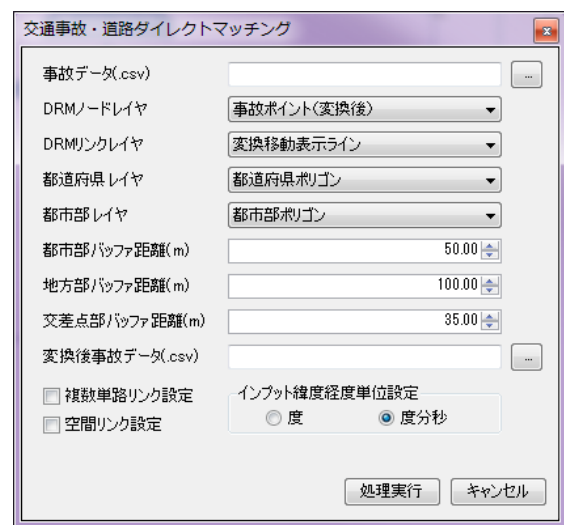


図 2.3: 事故位置修正システムの初期設定画面

入力として、それぞれの事故の発生場所の経度・緯度と道路種別等の情報を収めた csv 形式のファイルを与える。道路ネットワークデータは、DRM の基本道路網および全道路網のデータをシェープファイルに変換したものを利用する。DRM ネットワークデータは独自の固定長テキスト形式を採用しており、しかも測地系として旧測地系が使われているため、そのままでは ArcGIS を含む一般的な GIS では使いにくいので、シェープファイルに変換した。

初期設定画面では、近傍の道路リンク、交差点を探索する際のバッファの値を指定する。デフォルト値は都市部で 50m、地方部で 100m となっているが、自由に変更可能である。

それぞれの事故に紐付けられた DRM リンクあるいは交差点の情報は、csv 形式で出力される。出力ファイル名も初期設定画面で入力する。

図 2.4 に交通事故と DRM とを紐付けた結果を掲げた(新宿駅東側の部分である)。この図に示された結果からは、事故の多くで原票の経度・緯度の値は DRM ネットワークに近接する位置にあり、かなり正確な値が取得さ

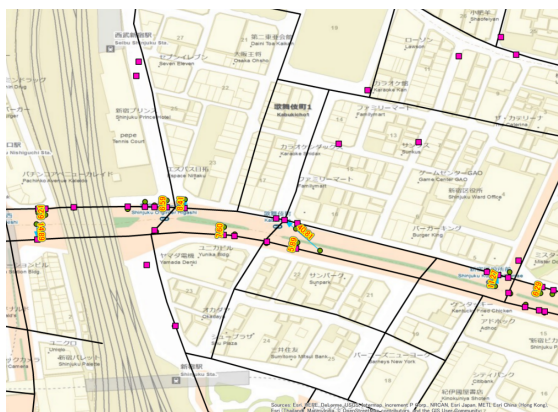


図 2.4: 事故位置修正システムの出力画面：緑丸は原票の経度・緯度の与える位置、赤い四角は DRM 上に紐付けられた事故位置、水色の矢印は事故位置の移動方向を表す。黄色の地に赤字で示された数字は移動距離 (m) である。DRM リンクと紐付けられていない赤い四角は近傍に同種別の道路が存在しなかった事故である。

れていることがわかる。とくに幹線的な道路で発生した事故についてはこのことが当てはまる。一方、非幹線道路で発生した事故については、近傍に同種別の道路が見つからない例がある。これは、原票では道路種別が市町村道(区道を含む)となっているにもかかわらず、DRM では市町村道という道路種別がないことに起因する。

2.4.2 交通事故データのもつ属性

以上の操作によって、交通事故のデータに DRM のデータが付加される。こうしてできあがった事故データのもつ主な属性は表 2.3 に示すとおりである⁴⁾。この属性は様々な事故分析に使われる。

表 2.3: 交通事故データのもつ属性

項目名	内容
事故番号	警察署番号+事故番号
死者、重傷者、軽傷者数	
発生年月日、昼夜の別	昼、夜、不明の三分
年齢	0~15, 16~24, 25~64, 65 以上
路面状態	乾燥, 湿潤, 凍結, 積雪, 非舗装
当事者種別	自動車, 二輪車, 自転車, 歩行者
事故類型	人対車両, 車両相互, 車両単独他
通行目的	観光, ドライブ, その他, 不明
行動類型	発進, 直進, 追越・追抜, 左折, 右折
道路線形	カーブ・屈折, 直線, これら以外
道路形状	交差点, 単路, 踏切, これら以外
法令違反	信号無視, 通行禁止違反他
道路種別、路線番号	
道路管理者	
現旧新道区分	
キロ程種別、キロ程	
上り、下りの区分	
道路構造	交差点, 同付近, 単路, 踏切, 分合流部
単路、交差点の区分	
2 次地域メッシュ番号	すぐ下のノード番号と組合せる
ノード 1, 2 番号	リンクの始点, 終点のコード
事故発生場所の座標	WGS 84 による
交通調査基本区間番号	道路交通センサスに対応
区間延長	
車道幅員、車線数	
歩道設置率	歩道代表幅員も含まれる
中央帯の設置状況	
交差点密度	信号機あり, 信号機なし
沿道状況	DID, その他市街部, 平地部, 山地部
用途地域区分	住居系, 商業系, 工業系, 用途地域外
時間帯別車種別交通量	センサスの一般交通量調査結果

第 3 章

交通事故・道路統合データベースシステムの作成およびデータ整備

道路管理者のための事故分析システムについては、第 2 章で述べたように、事故発生場所の経度・緯度情報を使って事故と道路とをマッチングする方式への転換を図った。2014 年度はこの方式を経度・緯度が高い精度で入力されている 9 県に適用し、新しい方式で作成したシステムに問題がないことを確認した。さらに 2015 年度はこの方式を 32 都道府県に拡大して適用し、2016 年度には 36 都道府県へと一層の拡大を図った。対象都道府県の拡大に伴い交通事故・道路統合データベースの作成に要する時間とコストの低減を図ることができた。

本章では、最初に第 3.1 節で本システムがどのような機能をもっているかを紹介するとともに、道路管理者からの要望を踏まえて改善を行った点をまとめる。次に、第 3.2 節で交通事故の分析で用いられる各種データの整備状況について触れる。3.2 節で扱うデータは道路管理者向けの分析システムを対象としたものではなく、一般的な事故分析で使われることを想定している。

3.1 交通事故・道路統合データベースシステム

道路管理者向けの交通事故・道路統合データベースシステム（以下、統合 DB と略記する）は、第 2 章で述べた方法で作成された交通事故のマッチングデータを基本データとして利用し、道路の区間別に事故を集計し走行台キロ当たりの事故率を求めることを基本的な機能としている。台キロ当たりの事故率は、整備の優先順位を判断する際に重要な指標として使われている。

統合 DB は、一般道路のうちの幹線的な部分（原則として、道路交通センサスの対象区間）を分析対象としている。高速道路、都市高速道路、自動車専用道路は統合

DB には含まれておらず、市区町村道、農（免）道、港湾道路、私道なども含まれていない。ただし、新直轄区間の一部は含まれている。これは、関連する道路管理者からの要請にもとづいて含めるようにした結果である。また、統合 DB には過去 10 年分の交通事故データが保存されており、交通事故の経年的な変化を分析することもできる。

以下では、3.1.1 節で統合 DB で使われている道路区間について述べ、3.1.2 節で分析事例を示す。最後の 3.1.3 節で道路管理者からの要望を踏まえて行った統合 DB 関連の改善について記す。

3.1.1 統合 DB の道路区間

統合 DB では、交通事故を分析するために従来から ITARDA 区間と呼ばれる集計用の区間を使ってきた。ITARDA 区間は交差点と単路の二つに分かれており、基本的に道路交通センサの調査に使われているセンサス区間を細分して作られている。これは、走行台キロ当たりの事故率を計算するさいにセンサス交通量を使う必要があるからである。

ただし、従来の道路交通センサスではセンサスを実施するたびにセンサス区間の見直しが行われており、この見直しに合わせて ITARDA 区間も改めて設定し直す必要があった。設定の見直しは 5 年に一度実施されるセンサスに合わせて行っていたが、対象道路の延長が長いため膨大な作業であった。さらに、ITARDA 区間の見直しに合わせて、ITARDA 区間別に集計されている事故についても新しい ITARDA 区間に再度振り分ける必要があるため、かなりの手間がかかった。

しかし 2010 年度に実施されたセンサスで「基本的にセンサス区間（交通量調査単位区間）は原則として固定

する」という方針が示され、2015 年度のセンサスでもこの方針が引き継がれたため、上記の見直し作業を大幅に縮小することができた。すなわち、2010 年度のセンサスで設定されたセンサス区間にもとづいて作成した ITARDA 区間を 2015 年以降も引き続き使用することが可能であった。もちろん、新設道路等については新たに ITARDA 区間の設定を行う必要があったが、従来の作業に比べれば大幅に軽減された。

2010 年度の交通センサスをもとに設定したイタルダ区間の数は、交差点部で 247,862、単路部で 599,774、合計 847,636 である。平均区間長は 310m、最大区間長は 1km である。最大区間長については、途中で他の道路との交差等がなければ 1km より長い距離を設定することも可能であるが、あまり区間長が長くなると細かな事故分析に対応できなくなるので、1km を最大長として分割することにした。図 3.1 には埼玉県熊谷市付近の ITARDA 区間の設定状況を掲げた。

最後に、統合 DB に含まれるデータについて簡単に記述しておく。道路管理者に提供している統合 DB は以下のデータを含んでいる。

- 区間別データ
 - ▷ ITARDA 区間別データ（都道府県別、道路管理者別）
 - 事故率データ（単路、交差点別）
 - ▷ センサス区間別データ（都道府県別）
 - 死傷事故、死亡事故別データ
- 事故別データ（道路管理者別、過去 10 年間の事故データ）

センサス区間別データは集計単位が大きく細かな事故分析には向いていないので、以下では ITARDA 区間別データと事故別データについての分析結果を述べる。

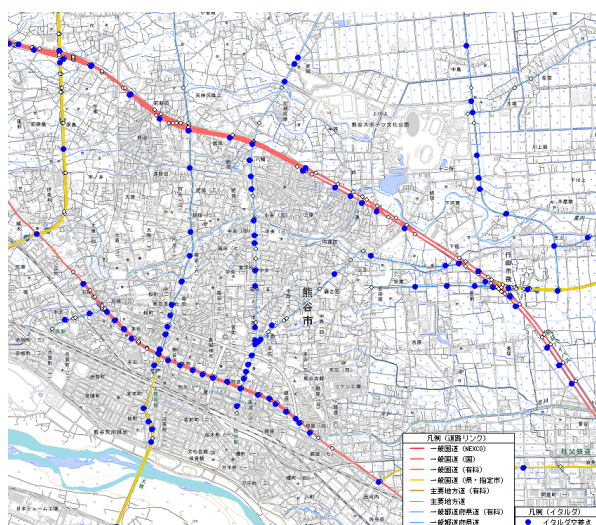


図 3.1: ITARDA リンクとノード：青丸は DRM 基本道路網上のノード（交差点）、白抜きの丸は DRM 細道路網上のノードである。青丸と青丸の間が 1 本の ITARDA リンクに対応する。背景は国土地理院地図。

3.1.2 統合 DB による分析事例

ITARDA 区間別データを使った分析では、各都道府県の道路管理者別（国道事務所、都道府県、および政令市）、路線別に事故率の比較を行うことができる。事故率は、過去 4 年間に当該区間で発生した事故件数とセンサス交通量をもとにして台キロ当たりの事故件数として計算されている。事故の種類は死亡事故、重大事故（死亡事故＋重傷事故）、死傷事故（死亡事故＋重傷事故＋軽傷事故）の三つである。

図 3.2 には ITARDA 区間（単路）別の死亡事故率を大きい順に並べたグラフを、また図 3.3a には ITARDA 区間別の重大事故率を大きい順に並べたグラフを表示した（いずれも仮想データを使ったグラフである）。どちらの種類の事故でも事故率の高い区間が存在することがわかる。この区間番号は、グラフ上で当該区間の棒をクリックすることで表示される。（図 3.3a のグラフにはこの区間情報が表示されている。）

さらに、それぞれの ITARDA 区間で発生した事故類型別の事故件数をグラフに表示することも可能であり、図 3.3b にはそれぞれの区間における追突事故の件数を表示した結果を示した。図 3.3b の ITARDA 区間の並びはその上の図 3.3a と同じである。この図によれば、追突事故の件数は必ずしも台キロ当たりの事故率と同じ傾向にはなっていないことがわかる。

こうしたグラフは交差点についても単路と同じように作成・表示することができる。（紙面の都合で交差点についてのグラフは省略する。）

図 3.3b に示した事故類型別の件数は、死亡事故、重大事故、死傷事故のそれぞれについて以下の事故類ごとにグラフ化することができるので、事故類型別の事故件数を ITARDA 区間単位で容易に比較することができる。なお、1 当とは「第一当事者」の略で、最初に交通事故に関与した車両等の運転者又は歩行者のうち、当該交通事故における過失が重い者をいい、また過失が同程度の場合には人身損傷程度が軽い者をいう。

- 夏季・冬季別の事故件数
- 昼夜別の事故件数
- 人対車の事故件数
- 自転車対車の事故件数
- 人対車の事故のうち、通行中、横断中、およびその他の件数
- 車両相互事故のうち、正面衝突、追突、出会い頭、右折時、左折時、その他の件数
- 車両単独事故のうち、工作物衝突、路外逸脱、転倒、そ

の他の件数

- 1 当が自動車，二輪車，自転車，歩行者の件数
- 1 当の年齢が0 歳～15 歳，16 歳～24 歳，25 歳～64 歳，65 歳以上の件数
- 1 当が65 歳以上の人対車両事故について，通行中，横断中，その他の件数
- 1 当が65 歳以上の車両相互事故について，正面衝突，追突，出会い頭，右折時，左折時，その他の件数

- 1 当が65 歳以上の車両単独事故について，工作物衝突，路外逸脱，転倒，その他の件数
- 2 当が65 歳以上の人対車両事故について，通行中，横断中，その他の件数
- 2 当が65 歳以上の車両相互事故について，正面衝突，追突，出会い頭，右折時，左折時，その他の件数
- 2 当が65 歳以上の車両単独事故について，工作物衝突，路外逸脱，転倒，その他の件数

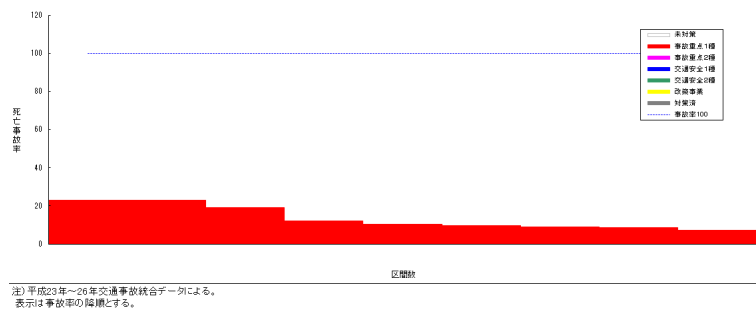
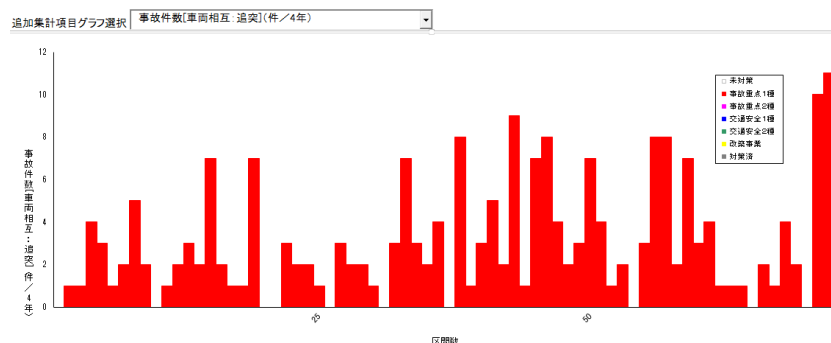


図 3.2: ある国道の ITARDA 区間別の走行台キロ当たり死亡事故率（仮想のデータ）



(a) ある国道の ITARDA 区間別の走行台キロ当たり重大事故率（仮想のデータ）：最も事故率の高い ITARDA 区間の属性情報がウィンドウに表示されている。



(b) ある国道の ITARDA 区間別の追突事故件数（仮想のデータ）：区間の並びは上の図と同じになっている。追突事故件数は，台キロ当たりの事故率と同じ傾向にはないことがわかる。

図 3.3: ある国道の ITARDA 区間別の走行台キロ当たり重大事故率と追突事故件数（仮想のデータ）

図 3.3a のウィンドウの中に示されている番号をもつ ITARDA 区間がどこに存在するかは、図 3.1 に示したシステムを使って検索することができる。結果は地図上に表示される。

ここまでは ITARDA 区間別データについて述べてきたところであるが、前述したようにこれとは別に事故別データも整備されている⁵⁾。これは区間別データと同じく、一般道路のうちの幹線的な部分で発生した事故についてのデータであり、表 2.3 に示した事故属性が付与さ

れている。

事故別データは事故の属性を使ったクロス集計に使われるほか、発生場所の座標を使って DRM 上に事故を落とすことができる。事故を地図に落とす場合も、属性を使って事故を絞り込んだ上で表示することができる。図 3.4 には、死亡・重傷・軽傷別に事故を表示した例を掲げた。この場合も、DRM 上の事故をマウスでクリックすることによって事故の属性が表示される。

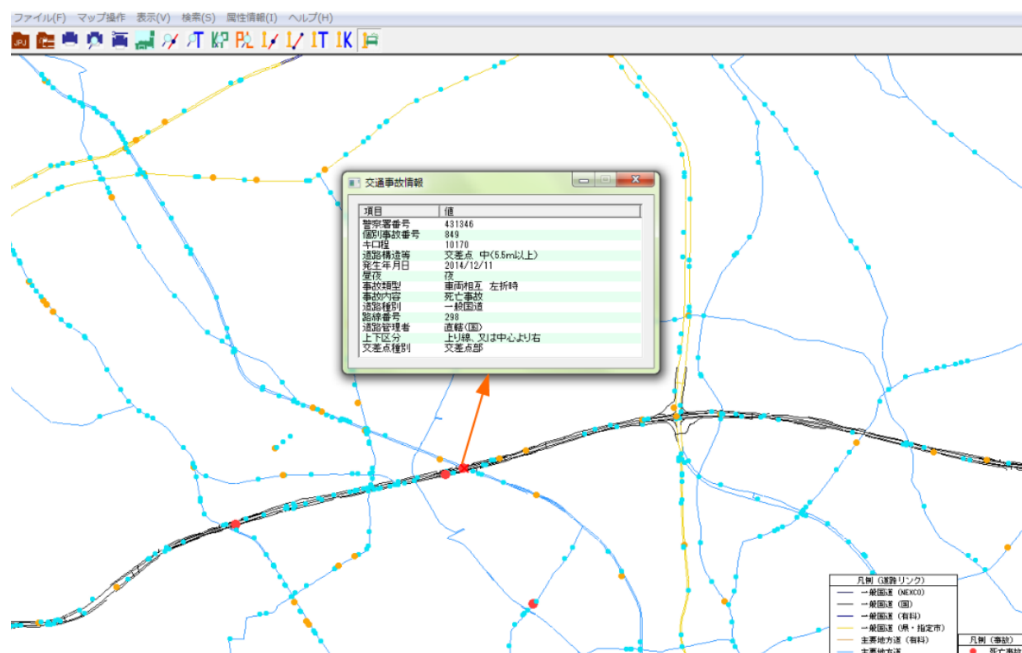


図 3.4: 個別の交通事故の DRM 上での表示：赤色の丸が死亡事故、橙色の丸が重傷事故、水色の丸が軽傷事故、中央のウィンドウの中には矢印で示した事故の属性が示されている。

3.1.3 統合 DB の改善

前節で説明した統合 DB システムに対して次のような改善要望が寄せられた。

- (1) (道路管理者から) 交通事故が発生した場所を地図上で確認できるようにしてほしい。
- (2) (道路管理者から) 地域メッシュ (1 km あるいは 500 m メッシュ) 単位に事故件数を集計してほしい。
- (3) (道路管理者から) 生活道路での交通事故を集計してほしい。
- (4) (道路管理者から) 交通事故データと ETC 2.0 で収集されたプローブデータを同じシステムに載せて両者の関連性を分析したい。
- (5) (県の担当者から) 事故発生場所の経度・緯度を

使った事故分析を行いたいので、協力してほしい。

最初の要望については、事故発生場所と ITARDA 区間とを対応させるシステム (図 3.1) を作成して解決した。

二番目と三番目の要望については 2014 年度からメッシュデータベースでの集計システムの開発を開始し、2015 年度には地域メッシュを使って集計したデータを作成し道路管理者に提供を行ったほか、ITARDA のサイトで公開した (詳細は第 3.2.6 節で述べる)。

四番目の要望については必要性が高いことは理解できるが、プローブデータ分析システムと事故分析システムはこれまで別々に開発されてきており、しかも双方のデータのボリュームがかなり大きいので、どのように二つのデータを統合して分析を行うか今後検討することとしたい。

最後の五番目の要望については、県側でどのような分析を行いたいのか十分詰まっていないところがあるが、少なくとも事故発生場所の経度・緯度を使った一般的な分析であれば、本プロジェクトで開発してきたシステムによって対応可能である。

これまでの道路管理者向けの交通事故・道路統合事故データベースは一般道（一般国道，主要地方道，一般都道府県道，および指定市の市道）を対象に作成されており，高速道路，都市高速道路は対象となっていなかった。しかし上記の 1～5 まで要望に加えて，高速道路，都市高速道路の管理者から，管理者として交通事故をすべて把握しているわけではなく，とくに重大でない事故については十分な把握ができていないため，一般道路を対象にして作成されている交通事故・道路統合データベースの高速道路版を作成してほしいという要請を受けた。

そこで，2015 度に高速道路（NEXCO 管理部分）と首都高速道路を対象にして道路管理者向けの交通事故・道路統合データベースを作成し，各道路管理者に提供した。データの内容は一般道路のものと同様である。ただし事故類型については，事故原票の高速道路追加項目票の記載内容を用い以下のように整理した。

- 人対車両
 - ▷ 故障修理中
 - ▷ 路上作業中
 - ▷ その他
- 車両相互
 - ▷ 追突
 - 追越・追抜車に
 - 走行車に
 - 流入車に
 - 流出車に
 - 車線停止中の故障・事故車に
 - 路肩停止中の故障・事故車に
 - その他
 - ▷ 衝突・接触
 - 対向車に
 - 追越・追抜車に
 - 走行車に
 - 流入車に
 - 流出車に

- 車線停止中の故障・事故車に
- 路肩停止中の故障・事故車に
- その他
- ▷ その他車両相互
- 車両単独
 - ▷ 転倒
 - ▷ 路外逸脱
 - ▷ 中央分離帯
 - ▷ 防護柵
 - ▷ その他

交通事故・高速道路統合データベースの概要は次のとおりであり，NEXCO 3 社にこのデータベースを提供した。

- 対象データ：2014 年に高速道路 3 会社が管理する道路上で発生した人身事故と 2010 年の道路交通センサスデータ
- 集計単位：IC 区間
- 集計対象：事故件数，死者数，死傷者数
- データ形式：IC 区間別に集計されたエクセルxlsx形式データと事故別のxlsx形式データの2種類

この統合データベースでは，事故発生場所の経度・緯度をもとにして発生場所のキロポスト（100 メートル単位）と該当する IC 区間を推定した。これは，高速道路で発生した交通事故を集計するにはこうした情報が有用だと考えられたからである。

これらの情報を得るために，事故発生場所の経度・緯度情報と事故原票の補充票に記載されているキロ程の情報，さらには DRM のセンサス対象区間に記載されているキロ程の情報をもとにしてキロポストを推定するプログラムを開発した。図 3.5 に本システムを使ってキロ程を推定している状況を示した。

なお，首都高版の統合データベースでは，集計単位を道路交通センサス区間および路線コードとした。発生場所のキロポストは付与していない。図 3.6 に首都高で発生した人身事故を死亡，重傷，軽傷の別に表示した結果を示した。死亡，重傷事故の件数は少なく，ほとんどが軽傷事故である。また，交通渋滞がよく発生する箇所では事故が多いことも読み取れる。

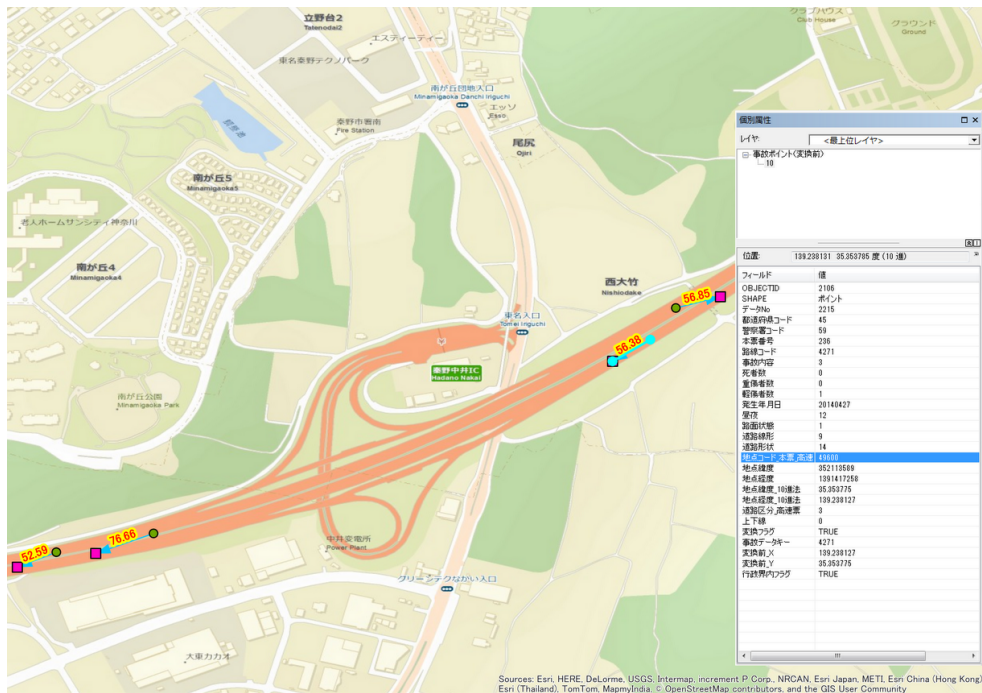


図 3.5: 高速道路上で発生した交通事故のキロ程の推定状況（緑丸が原票の経度・緯度による事故発生位置，赤い四角は推定されたキロポストの位置，この二つの点を結ぶ矢印の上の数字は事故を移動させた長さ (m).)

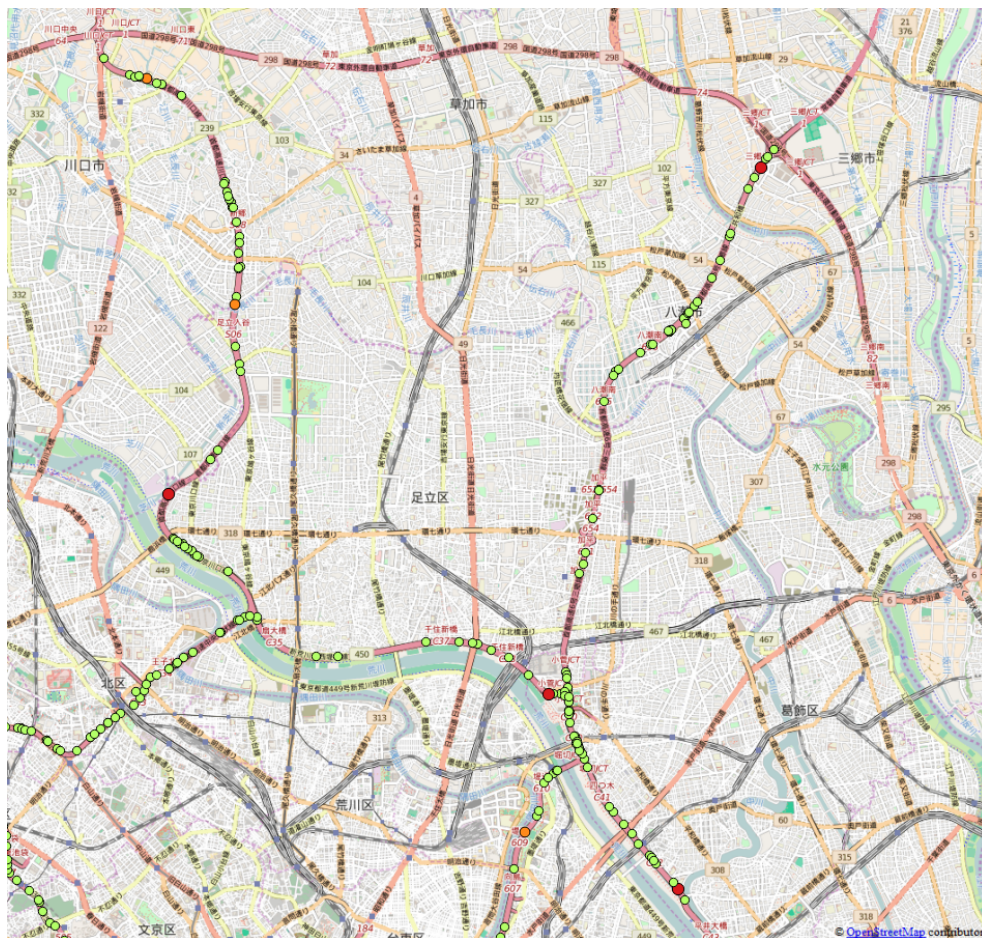


図 3.6: 首都高速道路上で発生した交通事故（●：死亡事故，○：重傷事故，●：軽傷事故）

3.2 事故分析のためのデータ整備

道路環境さらには周辺環境と交通事故とを関連づけて分析するために、以下のデータを整備した。以下のリストでメッシュといているのは第3次地域メッシュ（辺長は約1km）あるいは第4次地域メッシュ（辺長は約500m）のことである。

- 行政区域（市区町村および都道府県ポリゴン）
- 小学校区
- バス路線・停留所
- 豪雪地帯
- 過疎地域
- 都市地域
- 用途地域：メッシュ毎に用途地域別の面積を算出した
- DID 地区：メッシュ毎に DID 地区に「含まれる」「一部含まれる」「DID 地区に含まれない」の3カテゴリに分類した
- デジタル道路ネットワークおよびカーナビ用道路ネットワーク：メッシュ毎に道路ネットワークの道路種別毎の道路延長を計算した
- POI（Points of Interest）データ：テレポイント（電話帳）データを使って、メッシュ毎にそれぞれの業種の店舗数を集計した。ただし、元データの業種分類は細かすぎるので業種を統合した。
- センサス交通量：2010年に実施された道路交通センサスによる12時間／24時間交通量をリンク毎に紐付けた。
- 国勢調査による人口（全人口、高齢人口）、世帯数（メッシュ単位）
- 国勢調査による手段別通勤・通学者数（メッシュ単位）
- 経済センサスによる事務所数、従業員数（全産業、第2次産業、第3次産業別；メッシュ単位）
- 交通シミュレーション結果：i-Transport Labが行った交通シミュレーションによって求められた幹線道路の交通量を DRM リンクと紐付けた。このシミュレーションでは、DRMの基本道路網の上でシミュレーションが行われており、道路交通センサスの対象となっていない道路の交通量を把握できる。
- ゾーン30のポリゴンデータ
- 気象データ（メッシュ単位）
 - ▷ 年間降雨日数（1mm以上の降雨があった

日数）

- ▷ 年間降雪日数
- ▷ 最大積雪深
- ▷ 日最高気温が0度未満の日数

これらのデータのうち、行政区域から DID 地区までは国土数値情報である。以下、国土数値情報以外のデータについて概要を記述する。

3.2.1 DRM データ

本報告書でも再三にわたって DRM の名前が出てくることからわかるように、交通事故分析にあたって DRM は不可欠である。しかし、DRM は特殊なデータ形式（固定長テキスト形式）で記録されており、しかも測地系として旧測地系が使われており、図郭ごとに別ファイルになっている。そのため地理情報システム（GIS）でそのまま読み込むことができない。

本プロジェクトでは DRM を GIS で広く使われている shape ファイル形式に変換し、利用した。shape 形式への変換は1回行えばすむ。

3.2.2 国勢調査、経済センサスのデータ

国勢調査関係のデータについては、2010年に実施された国勢調査の結果を地域メッシュ単位に集計したデータを整備した。地域メッシュとしては3次メッシュと4次メッシュの二つを用いた。これらのメッシュデータは、東京大学空間情報科学研究センターの共同研究利用システム（JoRAS）に登録されているものを利用させていただいた。

国勢調査の関係で整備したデータ項目は以下のとおりである^{6),7)}。

- 人口総数
- 65歳以上の高齢者人口
- 世帯総数
- 持ち家世帯数（所得の代理変数として用いる）
- 利用交通手段別15歳以上自宅外就業者・通学者数
 - ▷ 徒歩のみ
 - ▷ 鉄道・電車
 - ▷ 乗合バス
 - ▷ 自家用車
 - ▷ オートバイ
 - ▷ 自転車

経済センサスのデータについては、2009（平成21）年に実施された経済センサスの結果を地域メッシュ単位に集計したデータを整備した。地域メッシュとしては3次

メッシュを用いた。

経済センサスの関係で整備したデータ項目は以下のとおりである⁸⁾。

- 事業所数（全産業）
- 従業者数（全産業，第 2 次産業，第 3 次産業；第 1 次産業のデータは利用していない）

3.2.3 ゾーン 30 のデータ

ゾーン 30 とは，生活道路における歩行者等の安全な通行を確保することを目的として区域を定めて最高速度 30 km/h の速度規制を行うとともに，そのほかの安全対策も必要に応じて組み合わせ，ゾーン内における速度抑制，ゾーン内を抜け道として通行する行為の抑制等をはかる生活道路対策である（警察庁資料⁹⁾から引用）。

ゾーン 30 を実施した区域の情報については，一部の自治体のホームページで公開されている。たとえば，埼玉県警察では県内のゾーン 30 規制区域の情報をネットで公開しており，非常に有用である。しかしながら，ここまで詳細な情報を提供している自治体は限られており，埼玉県警察の情報も紙ベース（画像ベース）でしか入手できない。GIS ソフトを使ってゾーン 30 での事故を分析するには，電子化されたゾーン 30 の情報が必要である。

こうしたことから，警察庁からゾーン 30 の情報を入手しポリゴンをシェープファイル化した。ただし，入手したデータは，2013 年 4 月から 2015 年 8 月の間に新設，更新，あるいは廃止されたゾーン 30 区域のデータに限定されており，ゾーン 30 区域がすべて網羅されているわけではないので留意していただきたい。

3.2.4 交通シミュレーションで求められた交通量

5 年に一度実施されている道路交通センサスでは，主要な幹線道路において交通量の観測が行われているが，すべての道路で交通量観測が行われているわけではない。とくに幹線道路以外の細道路の交通量が観測されていないことが道路交通センサスの交通量を事故分析に利用しようとする場合にネックとなる。

そこで本研究では，道路交通センサスに代えて i-Transport Lab が行った交通シミュレーション^{10),11)}の結果を用いることとした。このシミュレーションでは DRM の基本道路網を構成するすべてのリンクを対象にして i-Transport Lab が開発したプログラム SOUND

を使って交通量が求められており，時間帯別，車種別（大型車，小型車の 2 区分），上り・下り別の交通量を知ることができる。このため，道路交通センサスよりも広い範囲の道路の交通量を使うことができる。このシミュレーションでも幹線道路以外の細道路の交通量が得られないという欠点は残っているが，細道路まで含めて交通量をシミュレーションすることは困難な作業なので，今回は幹線道路の交通量を分析に使うこととした。

このシミュレーションデータでは，道路リンクごとに時間帯別，車種別，上り・下り別に交通量が csv 形式で記録されている。ベースデータとしては 2005 年の道路交通センサスによる OD 交通量が使われている。道路リンクとしてはデジタル道路地図（DRM）の 2007 年版が使われている。

csv 形式の交通量データには，位置座標が付与されていないので，DRM を利用して次の手順で道路リンクの端点の座標を求め，さらに各メッシュに含まれるリンクを特定し，それぞれのリンクの交通量を足し合わせてメッシュの交通量とした。下記の手順 4 では ArcGIS を使ったが，その以外の手順の計算は統計ソフト R で行った。

- (1) cvs ファイルから道路リンクの ID を読み込む。リンク ID は端点 1 と端点 2 のノード ID と 2 次メッシュ番号を組み合わせて作られているので，メッシュ番号と端点 1 と 2 のノード ID を求める。
- (2) ノード ID は「2 次メッシュコード＋メッシュ内でユニークに振られた一連番号」で構成されている。一連番号はそれぞれのメッシュ内においてのみユニークなので，一つのリンクの端点 1 と 2 が異なる 2 次メッシュに属していないことを確認する。
- (3) リンクの端点の座標は DRM の基本ノードファイルに含まれているので，このファイルと端点 1, 2 の ID およびメッシュ番号を使ってリンクの端点の座標を求める。
- (4) 上で求めた端点の座標は旧座標系にもとづく正規化座標なので，これらを世界測地系（WGS 84）の経度・緯度に変換する。この変換作業には ArcGIS を用いた。
- (5) ここまでの作業で各リンクの端点の座標が求められる。交通量については，端点のノード番号をもとにして上り・下りの交通量を足し合わせて時間帯別，車種別の交通量を計算した。これらの結果を使ってリンクとその属性情報を GIS ソフトで読み込むことができる形式に変換する。（WKT 形式に準拠した形式でリンクとその属性情報を作成した。）
- (6) リンク（と属性情報）をメッシュのポリゴン情報と組合せ，それぞれのリンクが含まれるメッシュを求める。（line_in_polygon の計算を行い，各ポリゴンに含まれるリンクを探し出す。）
- (7) メッシュ毎にその中に含まれているリンクの交通量を時間帯別，車種別に集計する。

3.2.5 気象データ

交通事故の発生には気象条件が影響することが知られている。しかしながら、気象条件が交通事故に与える影響は複雑であり、必ずしも「気象条件が悪ければ事故が増える」とはいえない。

たとえば、雨や雪が降っている時には路面の摩擦係数の値が小さくなり、車はスリップしやすくなる。これは交通事故の発生につながる要因である。一方、こうした悪天候時には歩行者、自転車の交通量は少なくなり、また車も低速で走行するようになるので、交通事故の発生件数は少なくなる。このように、気象条件が事故の発生に与える影響は複雑である。

こうした気象条件が事故に与える影響を定量的に分析できるようにするために、日最高気温が 0 度以下の日数、年間総降水量、年間総降雨日数（日雨量が 1 mm 以上の日数）、年間降雪日数（降雪量が 1 mm 以上の日数）、累積積雪深をとりあげた。

AMeDAS の観測点は全国に約 1,300 カ所存在しているが、その数はメッシュの数に比べ圧倒的に少ない^{12),13)}。このため、それぞれのメッシュの気象データを整備するには、何らかの補間が必要である。

気象データの補間については、Hofstra *et al.*¹⁴⁾ が様々な手法の比較を行って、点推定値については手法の間の差は小さいが精度は観測点の密度に依存するという結論を得ている。ここでは、この結論を参考にして補間には通常クリギングを用いるとともに、対象都道県の周辺の都府県の AMeDAS データも用いて当該県の補間を行った。すなわち、クリギングによってそれぞれのメッシュ中心点での気象データを求めた。ただし、北海道については周辺に隣接する県が存在しないため、北海道の AMeDAS データだけを使って補間を行った。

クリギングで用いた都道府県のグルーピングは以下のとおりである。

- 北海道
- 東北地方
 - ▷ 青森県
 - ▷ 秋田県
 - ▷ 岩手県
 - ▷ 宮城県
 - ▷ 福島県
- 関東地方
 - ▷ 茨城県
 - ▷ 群馬県
 - ▷ 埼玉県
 - ▷ 千葉県
 - ▷ 東京都

- ▷ 神奈川県
- 中部地方 1
 - ▷ 新潟県
 - ▷ 富山県
 - ▷ 石川県
 - ▷ 岐阜県
 - ▷ 福井県
- 中部地方 2
 - ▷ 長野県
 - ▷ 静岡県
 - ▷ 愛知県
 - ▷ 三重県
- 関西地方
 - ▷ 滋賀県
 - ▷ 京都府
 - ▷ 大阪府
 - ▷ 兵庫県
 - ▷ 奈良県
 - ▷ 和歌山県
- 中国地方
 - ▷ 鳥取県
 - ▷ 島根県
 - ▷ 岡山県
 - ▷ 広島県
- 四国地方
 - ▷ 徳島県
 - ▷ 香川県
 - ▷ 愛媛県
 - ▷ 高知県
- 九州地方
 - ▷ 福岡県
 - ▷ 佐賀県
 - ▷ 長崎県
 - ▷ 大分県
 - ▷ 熊本県
 - ▷ 宮崎県
 - ▷ 鹿児島県

なお、降雨日数、降雪日数、最高気温が 0 度未満の日数を求めるに当たっては、それぞれのメッシュを対象にこれら気象データの 1 年 365 日ごとの値を求め、その結果を使って条件を満たす日数を計算した。

気象庁からはメッシュごとの平年値の値が公表されているが、これは 30 年間の観測値の平均なので特定の年に発生した交通事故の分析に使うには適切ではないと考え、2013 年一年間の AMeDAS 観測データを使って、メッシュ毎に補間を行い必要な値を求めた。なお、クリギングの計算には統計ソフト R を用いた。

3.2.6 メッシュ単位の集計

交通事故の分析ではメッシュを使った分析が行われることも多く、3 次または 4 次の地域メッシュ単位に事故

件数を集計することが求められる機会も多い。ただし分析の都度、対象とする事故を抽出しメッシュごとに抽出された事故の件数を集計するのは手間と時間がかかるので、以下のようなデータを作り、メッシュベースの集計を簡便に行えるようにした。

すなわち、事故に与えられている経度・緯度の値をもとにしてそれぞれの事故が属すメッシュのコードを求め、このメッシュコードを事故データの属性として付与しておくのである。本研究ではそれぞれの事故に対して、その事故が属する4次地域メッシュ（辺長約500m）の9桁のコードを付与した。このコードを使え

ば、4次メッシュだけではなく、3次メッシュ他のメッシュを使った集計を簡単に行うことができる（GISのpoint-in-polygonなどの機能を使うまでもなく、クロス集計を行えばよい）。

このデータを使って生活道路での事故件数を4次地域メッシュ単位で集計した結果をITARDAのサイト（<http://www.itarda.or.jp/service/webmap.php>）で公開した（図3.7）。ただし「生活道路」については統一的な定義が定まっていないので、ここでは便宜的に一般市町村道で発生した事故件数を集計している。



図 3.7: 東京・神奈川地域の生活道路での事故を4次メッシュで集計した結果

第 4 章

交通事故の発生要因とクラスタリングの分析

本章では、交通事故の発生要因を一般化線形モデル (GLM), CAR モデル (Conditional Autoregression model) 等を使って分析した結果と交通事故の空間的な分布パターンとクラスタリングについて分析した結果を提示する。

交通事故の発生要因の分析では、2013 年に一般道路で発生した人身交通事故を 3 次メッシュ (辺長約 1 km) 単位に集計したデータを基礎データとして用い、計数モデルを推定した。高速道路、都市高速道路、自動車専用道路で発生した事故、SA・PA、広場、駐車場等で起きた事故は分析から除外した。これは、こうした道路と場所で発生した事故では一般道路で発生した事故と発生要因が異なっていると考えられるからである。

一般道路で発生した事故を 3 次メッシュ単位に集計すると、東京都、大阪府等の大都市を含む都府県以外では事故件数が 0 というメッシュが大多数を占める。このため通常広く用いられているポアソン回帰モデルではよい推定結果が得られなかった。この点を解決するために、負の二項分布回帰モデル、ゼロ強調モデル等の適用を行った。さらにメッシュの間の空間的な相関を考慮したモデルについても推定を試みた。

これらのモデルの概要と推定結果については第 4.1 節で述べる。

交通事故の空間的な分布パターンの分析では、一般道路の中でも幹線的な道路である国道、主要地方道と一般都道府県道で発生した事故をとりあげ、死亡事故、重傷事故、そして事故多発箇所の空間分布パターンに差があるかどうかを分析した。この結果は第 4.2 節で提示する。この分析では事故をメッシュ単位に集計せず、個々の事故そして事故多発箇所の場所そのものを分析対象としている。

さらに、3 次メッシュ単位に集計した人身事故件数を使って、事故件数の多いメッシュが集中 (クラスタリング) しているかどうかを分析した。この結果は第 4.3 節

に記載した。

なお、以下の分析は統計ソフト R を使って行った。これは、R がオープンソースとして提供されていることに加え、空間情報データを自由に扱うことができ、さまざまなモデルを推定するためのパッケージがそろっていることが大きな要因である。R での空間情報データの処理については、Bivand *et al.*¹⁵⁾ および Brunson & Lex¹⁶⁾ に詳しく書かれているので省略する。

4.1 交通事故の発生要因の分析

本節では、交通事故の発生要因をさまざまなモデルを使って分析した結果を示す。

交通事故は 0 または正の整数値をとる「計数データ」なので、分析にあたっては計数データを扱うことのできるモデルを選択する必要がある。交通事故のようなまれに発生する事象のモデルとして広く使われているのはポアソン回帰モデルである^{17), 18)}。

しかしながら、ポアソン分布は「期待値と分散が等しい」という性質をもっており、実際の観測データではこの性質が満足されないことがある。とくに交通事故などのデータでは分散の値が期待値より大きくなることが多く、「過大分散 (overdispersion)」と呼ばれている。過大分散に対処するための方策のひとつは、ポアソン分布ではなく負の二項分布を使うことである。負の二項分布の分散の値は期待値より大きくなることが知られている。

過大分散に対処するためのもう一つの方法として「ゼロ強調モデル (zero-inflated models)」がある¹⁹⁾。ゼロ強調モデルでは、値 0 をとる確率モデルとポアソン回帰モデルあるいは負の二項分布回帰モデルとを組み合わせたモデルが使われる。

ポアソン回帰モデルと負の二項回帰モデルは、どちらも「一般化線形モデル (GLM: Generalized Linear Models)」と呼ばれるモデル族に含まれ、その分布の期

待値が説明変数の線形関数の指数関数で与えられる（すなわち、期待値 $= \exp(x^\top \beta)$ ； x は説明変数のベクトル、 β は係数ベクトルで、 $(\cdot)^\top$ は転置を表す）。これに対してゼロ強調モデルは一般化線形モデルには含まれない。このためモデルを相互に比較する場合には注意が必要である。

さらに、事故の分析では地理空間的な相関についても考慮する必要があると考えられる。交通事故のモデルとして従来広く用いられているポアソン回帰モデル、負の二項分布回帰モデルでは、それぞれのメッシュは互いに独立と仮定されているが、交通事故データではこの独立性の仮定が必ずしも成り立たない可能性があるからである。つまり、近隣のメッシュの事故件数は類似した値をとることが多く、メッシュの間に相関があると仮定したモデルの方がより適切かもしれない。こうした空間的な相関関係を考慮したモデルもいくつか提案されている²⁰⁾ ので、本章ではこうしたモデルの一つである CAR モデルについてベイズ推定を行った。CAR モデルには空間的なランダム効果項が組み込まれている^{21), 22)}。

また、空間的な重み付けを考慮したモデル (GWM: Geographically Weighted Models) も提案されている²³⁾。このモデルは空間的な相関関係のモデル化を目指したものではなく、空間的な非定常性をモデル化することを目的としている。このモデルの推定では、対象としている（ある一つの）メッシュのモデルの係数を推定するさいに、そのメッシュから離れたメッシュの重みが小さくなるような重み付けを行い、メッシュごとに係数の推定を行う。こうして推定されたメッシュごとの係数の値を、すべてのメッシュのデータを（重み付けを行わないで）使って推定したグローバルモデルの係数と比較することによって空間的な非定常性を検出する。

最初に GWM ポアソン回帰モデルの推定結果を述べておくと、意味のあるモデルの推定はできなかった。これは、交通事故が各都道府県の都市地域に集中して発生しており、非都市部ではほとんどのメッシュの事故件数

が 0 となることが原因である。非都市部のメッシュでは当該メッシュの事故件数が 0 であることに加えて、その周辺の（相対的に大きな重みをもつ）メッシュの事故件数がいずれも 0 となってしまう、モデルが推定できないのである。無理にモデルを推定しようすると、推定に使われるメッシュの数を増やして都心部のメッシュ（その事故件数は 0 ではない）までも含めて推定が行われ、結局グローバルなモデルとの差がほとんどなくなってしまう。したがって、以下では GWM モデルにはこれ以上触れない。

以下では、島嶼部を除く東京都の交通事故（2013 年の平日に発生した事故）を対象にして分析を行った結果を主体に述べる。これ以外の道府県についてもモデル推定を行ったが、紙幅が限られているので必要最小限の結果のみを示す。まず第 4.1.1 節でポアソン回帰モデル (Po と略記する)、負の二項分布回帰モデル (NB と略記する)、ゼロ強調ポアソン回帰モデル (ZiPo) とゼロ強調負の二項分布回帰モデル (ZiNB と略記する) の推定について述べ、次に第 4.1.2 節で CAR モデルの推定結果を提示する。

4.1.1 ポアソン回帰モデル、負の二項分布モデル、およびゼロ強調モデル

本節では、次の四つのモデルを用いて 3 次地域メッシュ単位に集計した平日の事故件数をモデル化し、モデル相互の比較を行った結果を述べる。なお説明変数については第 3.2 節で述べたので、特に必要がない限り細かな説明は省く。

- (1) ポアソン回帰モデル (Po)
- (2) 負の二項回帰モデル (NB)
- (3) ゼロ強調ポアソン回帰モデル (ZiPo)
- (4) ゼロ強調負の二項回帰モデル (ZiNB)

それぞれのモデルは以下のように表わされる。ゼロ強調モデルでは、過度なゼロをとる確率 π_i あるいは $\hat{\pi}_i$ はロジットモデルで表現されている。

$$\text{ポアソン回帰モデル } \text{Prob}(Y_i = y) = \frac{\lambda_i^y \exp(-\lambda_i)}{y!} \quad y = 0, 1, 2, \dots, \quad (4.1)$$

$$\lambda_i = \exp \left\{ \beta_0^{(\text{Po})} + \beta_1^{(\text{Po})} x_1^{(\text{Po})} + \dots \right\},$$

$$\text{負の二項分布回帰モデル } \text{Prob}(Y_i = y) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \frac{\alpha_i^y}{(1+\alpha_i)^{y+k}} \quad y = 0, 1, 2, \dots, \quad (4.2)$$

$$Y_i \text{ の期待値 } E(Y_i) = \alpha_i k = \exp \left\{ \beta_0^{(\text{NB})} + \beta_1^{(\text{NB})} x_1^{(\text{NB})} + \dots \right\},$$

$$\text{ゼロ強調ポアソン回帰モデル } \text{Prob}(Y_i = y) = \delta_y(0) \cdot \pi_i + (1 - \pi_i) \frac{\check{\lambda}_i^y \exp(-\check{\lambda}_i)}{y!} \quad y = 0, 1, 2, \dots, \quad (4.3)$$

$$\check{\lambda}_i = \exp \left\{ \check{\beta}_0^{(\text{ZiPo})} + \check{\beta}_1^{(\text{ZiPo})} \check{x}_1^{(\text{ZiPo})} + \dots \right\},$$

$$\pi_i = \frac{1}{1 + \exp(-z_i)} \quad z_i = \zeta_0^{(\text{ZiPo})} + \zeta_1^{(\text{ZiPo})} u_1^{(\text{ZiPo})} + \dots,$$

$$\text{ゼロ強調負の二項回帰モデル } \text{Prob}(Y_i = y) = \delta_y(0) \cdot \tilde{\pi}_i + (1 - \tilde{\pi}_i) \frac{\Gamma(y + \tilde{k})}{\Gamma(\tilde{k})\Gamma(y + 1)} \frac{\tilde{\alpha}_i^y}{(1 + \tilde{\alpha}_i)^{y + \tilde{k}}} \quad y = 0, 1, 2, \dots, \quad (4.4)$$

$$\text{負の二項分布の期待値} = \tilde{\alpha}_i \tilde{k} = \exp \left\{ \tilde{\beta}_0^{(\text{ZiNB})} + \tilde{\beta}_1^{(\text{ZiNB})} \tilde{x}_1^{(\text{ZiNB})} + \dots \right\},$$

$$\tilde{\pi}_i = \frac{1}{1 + \exp(-\tilde{z}_i)} \quad \tilde{z}_i = \tilde{\zeta}_0^{(\text{ZiNB})} + \tilde{\zeta}_1^{(\text{ZiNB})} \tilde{u}_1^{(\text{ZiNB})} + \dots$$

変数の意味は次のとおりである。

Y_i :	メッシュ i の事故件数 (確率変数)
y :	事故件数のとる非負の整数値 ($y = 0, 1, 2, \dots$)
λ_i :	Po におけるメッシュ i の事故件数の期待値
$\Gamma(\cdot)$:	ガンマ関数
$x_j^{(\text{Po})}$:	Po の期待値の説明変数 ($j = 1, 2, \dots$)
$\beta_m^{(\text{Po})}$:	Po の期待値の説明変数の係数 ($m = 0, 1, 2, \dots$)
$\alpha_i k$:	NGB の期待値
$x_j^{(\text{NB})}$:	NGB の期待値の説明変数 ($j = 1, 2, \dots$)
$\beta_m^{(\text{NB})}$:	NGB の期待値の説明変数の係数 ($m = 0, 1, \dots$)
$\delta_y(0)$:	$y = 0$ のとき値 1, $y \geq 1$ のとき値 0 をとる関数
π_i :	ZiPo において過度なゼロをとる確率
$\check{\lambda}_i$:	ZiPo のポアソン分布部分の期待値
$\check{x}_j^{(\text{ZiPo})}$:	ZiPo のポアソン分布部分の期待値の説明変数 ($j = 1, 2, \dots$)
$\check{\beta}_m^{(\text{ZiPo})}$:	ZiPo のポアソン分布部分の期待値の説明変数の係数 ($m = 0, 1, \dots$)
$u_l^{(\text{ZiPo})}$:	ZiPo のロジットモデルの説明変数 ($l = 1, 2, \dots$)
$\zeta_n^{(\text{ZiPo})}$:	ZiPo のロジットモデルの説明変数の係数 ($n = 0, 1, \dots$)
$\tilde{\pi}_i$:	ZiNB において過度なゼロをとる確率
$\tilde{\alpha}_i \tilde{k}$:	ZiNB の負の二項分布部分の期待値
$\tilde{x}_j^{(\text{ZiNB})}$:	ZiNB の負の二項分布部分の期待値の説明変数 ($j = 1, 2, \dots$)
$\tilde{\beta}_m^{(\text{ZiNB})}$:	ZiNB の負の二項分布部分の期待値の説明変数の係数 ($m = 0, 1, \dots$)
$\tilde{u}_l^{(\text{ZiNB})}$:	ZiNB のロジットモデルの説明変数 ($l = 1, 2, \dots$)
$\tilde{\zeta}_n^{(\text{ZiNB})}$:	ZiNB のロジットモデルの説明変数の係数 ($n = 0, 1, \dots$)

モデルの構築は次の手順で行った。

- 目的変数と他の変数との間の相関係数を計算する。
- さらに、目的変数と他の変数の散布図を描く。
- 相関係数と散布図をもとにして、説明変数の候補を選択する。
- ツリーモデルを使って、目的変数に対する影響力の大きな変数を選択する。
- 目的変数に対する影響力の大きな変数の組を選び、一般化加法モデルのノンパラメトリックな平滑化関数を使って目的変数との間に非線形な関係

が存在するかどうかを確認する。

- 樹木モデルを使って、説明変数に複雑な交互作用がないかを調べる²⁴⁾。

この手順にしたがって説明変数の絞り込みを行い、最尤法を使ってモデルを推定した。

まず目的変数である平日の事故件数とそれ以外の変数との間の相関係数の値を求めたところ、結果は表 4.1 に示したようになった。この表によれば、事故件数との間の相関件数が高い変数として「店舗数 (r=0.681)」「娯楽施設数 (r=0.583)」「高校・大学数 (r=0.600)」「旅行者関連施設数 (r=0.611)」「一般道路延長 (r=0.609)」「交

差点数 ($r=0.701$)」「合計交通量 ($r=0.717$)」「駐車場数 ($r=0.599$)」「道路の総延長 ($r=0.695$)」等があげられる (括弧内の数字は平日の事故件数との間の相関係数である)。一方、「小学校数 ($r=0.464$)」「銀行数 ($r=0.497$)」「デパート数 ($r=0.479$)」「病院数 ($r=0.319$)」「鉄道関連施設数 ($r=0.443$)」「主要地方道延長 ($r=0.481$)」「国道延長 ($r=0.291$)」等の変数は事故件数との間の相関係数が低い。

目的変数と説明変数の間の散布図も描いた (図は省略)。その結果によると商業系の用途地域で事故件数が高いことがわかる。これに次ぐのは、工業系の用途地域である。またメッシュ内の小学校数が増加しても事故件数は上昇する傾向にはない。これは、小学校数の多い

メッシュは住宅系の用途に特化しており、幹線的な道路が少なく交通量も少ないことを反映しているためであると考えられる。ただし、このことは小学校の周辺での交通安全対策が不要であることを意味するものではない。高校・大学数やスポーツ施設数が増加すると事故件数も増える傾向が読み取れる。駐車場数が増えると事故件数も増加する。道路関係の変数と事故件数との関係を見てみると、道路延長が増加するにともない事故件数も増える傾向にあることがわかる。同じく総道路延長、交通量合計の値が増えると事故件数が増加することがわかる。一方、センサスの平均旅行速度と事故件数との間には負の相関があるように見える。これは、平均旅行速度が低いメッシュでは交通量が多く、したがって事故件数も増えることを反映しているためであろう。

表 4.1: 平日の一般道路における事故件数と説明変数との間の相関係数の値 (東京都, 2013 年)

変数	事故件数	店舗数	百貨店数	娯楽施設	大学数	SP 施設	旅行施設	一般道長	細道路長	交差点	交通量	駐車場数
店舗数	0.68											
百貨店数	0.48	0.77										
娯楽施設数	0.58	0.87	0.69									
大学等数	0.60	0.78	0.62	0.69								
スポーツ施設数	0.57	0.77	0.65	0.77	0.75							
旅行施設数	0.61	0.78	0.57	0.81	0.67	0.70						
一般道長	0.61	0.49	0.31	0.39	0.41	0.37	0.44					
細道路長	0.51	0.33	0.19	0.15	0.35	0.25	0.14	0.38				
交差点数	0.70	0.54	0.33	0.36	0.49	0.39	0.39	0.70	0.88			
交通量計	0.72	0.58	0.39	0.55	0.47	0.47	0.58	0.53	0.24	0.50		
駐車場数	0.60	0.55	0.39	0.59	0.52	0.50	0.65	0.40	0.09	0.32	0.53	
道路延長	0.69	0.49	0.30	0.32	0.46	0.38	0.33	0.73	0.90	0.97	0.48	0.28

次にツリーモデルを使って目的変数と説明変数との関係を分析した。この結果によると、事故件数に対して大きな影響をもつ変数として、「店舗数」「道路総延長」「合計交通量」「デパート数」「駐車場数」が抽出された。「店舗数」「道路総延長」「駐車場数」は、ツリーモデルの中に複数回現れており、これらの変数と事故件数との間には非線形な関係がある可能性がある。そこで、一般化加法モデルの平滑化関数を使ってこれらの変数と事故件数との関係を分析した。

一般化加法モデルの結果を見ると、事故件数と店舗数、道路総延長の間には複雑な非線形関係はない。しかし合計交通量の値の大きな部分では非線形の関係があるように見える。また、事故件数と駐車場数との間にも非線形の関係があるように見える。

これらの結果を踏まえて、まずポアソン回帰モデルの推定を行った。以下で紹介するモデルでは、オフセット項として各メッシュの道路総延長の対数をとった項が組み込まれている。したがって以下のモデルは道路延長あたりの事故件数を推定していることになる。オフセット項の係数は 1 に固定されており、推定結果をまとめた表の中にはこの項は現れない。

ポアソン回帰モデルの推定結果は表 4.2 に示したと

りである。用途地域を表すカテゴリー変数のいくつかが有意ではないが、残りの変数は有意となっている。とくに、低層・中層の住居専用地域では事故件数は (基準となっている) 商業地域より少ない。これに対して DID 地区では事故件数は多くなる。さらに、店舗数、デパート数、交通量、駐車場は正の係数を持ち、これらの要因の増加は事故件数の増加を意味する。

しかしこのポアソン回帰モデルのデータへの当てはまりについて χ^2 検定を行ったところ、当てはまりがよいとはいえない結果であった。すなわち、全てのメッシュの事故件数をすべて正しく予測する「フルモデル」(メッシュの数と同じ数のダミー変数を使って構築されるモデル—したがって、統計的に意味のあるモデルではない)とは有意な差があるという結果が得られた。この検定は表 4.2 の下部に示されている「残差逸脱度」の値とその自由度を用いて行った。残差逸脱度は推定されたモデルの逸脱度を使って次式で計算できる²⁵⁾。

$$\text{逸脱度 } D = -2 \log L^*, \quad (4.5)$$

$$\begin{aligned} \text{残差逸脱度} &= D - \text{フルモデルの逸脱度} \\ &= -2 (\log L^* - \log L_{\text{Full}}) \end{aligned} \quad (4.6)$$

ここに $\log L^*$ は推定されたモデルの対数尤度の最大値

を表し、 $\log_{<Full>}$ はフルモデルの対数尤度の値である。フルモデルの対数尤度は他のモデルに対し最大の値となるので残差逸脱度は 0 または正の値をとり、その値が大きいほど当てはまりが悪いことを意味する。残差逸脱度を使った検定は推定されたモデルとフルモデルを比較していることになり、この二つのモデルの間に有意な差がなければ推定されたモデルは一応のレベルに達していることを意味する。

また過大分散についても検討を行った。この検討を簡便に行うには残差逸脱度をその自由度で除してディスパージョン・パラメーター ϕ の値を計算すればよい。表 4.2 のポアソン回帰モデルの場合には、 $5467.5/1646 = 3.32$ となり 1 より大きな値である。より正確にディスパージョン・パラメーターの値を計算するにはピアソンのカイ二乗統計量を使う。同じモデルについてこの方法で計算すると 3.45 となり、表 4.2 のポアソン回帰モデルは過大分散になっている。

そこで、負の二項回帰モデルの推定を行った。説明変数はポアソン回帰モデルで用いたものと同じである。結果は表 4.3 に示されている。カイ二乗検定の結果によれば、このモデルはフルモデルと有意な差はないという結論が得られた。

表 4.3 の推定結果を見ると、ポアソン回帰モデルの結果と比べて統計的に有意な変数は少なくなっている。「デパート数」がそうである。さらに、用途地域の中の「工業系地域」「中高層住居専用地域」「住居地域」も有意ではない。用途地域の基準となっているのが「商業系地域」であるので、係数が有意に推定されなかった「工業系地域」「中高層住居専用地域」「住居地域」については、

「商業系地域」と同程度の事故の起こりやすさになっていると解釈できる。一方、「低層住居専用地域」と「その他地域」の係数は負の値であり、「商業系地域」より事故が起こりにくいことを意味する。これ以外の変数では、「合計交通量」と「駐車場数」の一次と二次の項がどちらも有意に推定されている。したがって、交通量あるいは駐車場の数が多いメッシュでは事故が起こりやすいことがわかる。

この負の二項回帰モデルについても過大分散の検討を行ったところ、ディスパージョン・パラメーター ϕ の値は、簡便な方法で 1.03、ピアソンのカイ二乗統計量を使った方法で 1.08 と求められた。したがって、負の二項回帰モデルは過大分散ではない。

表 4.2 に示したポアソン回帰モデルと表 4.3 に示した負の二項回帰モデルは入れ子の関係 (nested) にあるので、尤度比検定を使ってモデルを比較した。その結果によれば、負の二項回帰モデルの方が統計的に優れているという結論であった。さらに AIC の値を計算したところ、ポアソン回帰モデルでは 11251.6、負の二項回帰モデルでは 9185.0 であった。

次にゼロ強調モデルの推定を行った。ゼロ強調モデルとしては、ゼロ強調ポアソン回帰モデル (ZiPo) とゼロ強調負の二項回帰モデル (ZiNB) をとりあげた。

最初に ZiPo を推定した。説明変数として表 4.2 に示したポアソン回帰モデル (Po) と同じ説明変数を使うことを試みた。ただし、Po ではポアソン分布の期待値の説明変数として使われていた「DID 地区」(カテゴリ変数) をゼロ強調部分の説明変数として組み込んだ。

表 4.2: ポアソン回帰モデルの推定結果 (東京 2013 年の平日、一般道路での事故)

変数	推定値	標準誤差	z 値	Pr(> z)	
切片	-8.844e+00	8.500e-02	-104.052	< 2e-16	***
用途地域 (Indust)	-5.186e-02	3.044e-02	-1.704	0.088465	.
用途地域 (Low_Res)	-2.272e-01	3.199e-02	-7.102	1.23e-12	***
用途地域 (Mid_Res)	-1.159e-01	3.157e-02	-3.671	0.000241	***
用途地域 (Res)	-1.009e-02	3.208e-02	-0.315	0.753072	
用途地域 (Other)	-2.658e-02	6.388e-02	-0.416	0.677299	
DID 地区 (DID)	1.430e+00	7.922e-02	18.048	< 2e-16	***
店舗数	3.013e-04	5.467e-05	5.512	3.55e-08	***
デパート数	1.114e-03	3.329e-04	3.346	0.000821	***
合計交通量	2.032e-05	6.909e-07	29.413	< 2e-16	***
合計交通量の二乗	-1.357e-10	6.251e-12	-21.707	< 2e-16	***
駐車場数	7.754e-02	6.562e-03	11.817	< 2e-16	***
駐車場数の二乗	-2.486e-03	3.661e-04	-6.790	1.12e-11	***
ヌル逸脱度	11032.1	(df=1658)			
残差逸脱度	5467.5	(df=1646)			
AIC	11252				

表 4.3: 負の二項回帰モデルの推定結果 (東京 2013 年の平日, 一般道路での事故)

変数	推定値	標準誤差	z 値	Pr(> z)	
切片	-8.775e+00	1.268e-01	-69.209	< 2e-16	***
用途地域 (Indust)	-3.476e-02	8.315e-02	-0.418	0.67591	
用途地域 (Low_Res)	-2.254e-01	8.442e-02	-2.670	0.00759	**
用途地域 (Mid_Res)	-1.264e-01	8.419e-02	-1.501	0.13336	
用途地域 (Res)	-2.900e-02	8.656e-02	-0.335	0.73761	
用途地域 (Other)	-2.934e-01	1.278e-01	-2.295	0.02175	*
DID 地区 (DID)	1.284e+00	9.621e-02	13.345	< 2e-16	***
店舗数	5.753e-04	1.566e-04	3.673	0.00024	***
デパート数	3.062e-04	9.385e-04	0.326	0.74419	
合計交通量	2.561e-05	1.745e-06	14.678	< 2e-16	***
合計交通量の二乗	-1.857e-10	1.921e-11	-9.666	< 2e-16	***
駐車場数	8.107e-02	1.766e-02	4.592	4.4e-06	***
駐車場数の二乗	-2.852e-03	1.138e-03	-2.506	0.01220	*
ヌル逸脱度	3368.8	(df=1658)			
残差逸脱度	1696.4	(df=1646)			
AIC	9185				
k	5.779	(sd = 0.336)			

しかしながら, この説明変数の組み合わせでは推定がうまくいかなかった. そこで変数の組み合わせを変えて試行錯誤的にモデルを探索した. その結果選ばれたのが表 4.4 に示したモデルである. このモデルでは「総交通量」「交差点数」は有意な変数として選択されなかった.

表 4.4 に掲げた ZiPo と切片だけをもつ「ヌルモデル」とを尤度比検定を使って比較したところ, 表 4.4 の ZiPo はヌルモデルと比べ優れているという結果であった.

ただし過大分散について検討したところ, ディスパーション・パラメーターの値は 2.98 となり過大分散であるという結果になった. また, このモデルの AIC の値は 12295.3 であった.

最後に, ゼロ強調負の二項回帰モデル (ZiNB) を推定した. 説明変数は表 4.4 に掲げた ZiPo と同じである. 推定結果は表 4.5 に示したとおりである.

表 4.4 の ZiPo と表 4.5 の ZiNB とは入れ子の関係に

なっているので, 尤度比検定を行って比較した. その結果によると, ZiNB の方が優れているという結論が導かれた.

また表 4.5 の ZiNB についてディスパーション・パラメーターの値を計算したところ, 0.96 となり過大分散とはいえないという結果であった. このモデルの AIC の値は 9456.7 であった.

表 4.5 のモデルでは, 用途地域を表すすべての変数が有意にはなっていない. 「工業系地域」「住居系地域」の係数は有意ではなく, これらの地域での事故の起こりやすさは「商業系地域」と有意な差はない. 一方, 「低層住居専用地域」「中高層住居専用地域」および「その他地域」の係数は負で有意に推定されていることから, これらの地域での事故の起こりやすさは「商業系地域」と比べ有意に低いことがわかる. また, 「駐車場数」「店舗数」の係数は正で有意に推定されており, 駐車場, 店舗の多い繁華な地域では事故が起こりやすいといえる.

表 4.4: ゼロ強調ポアソン回帰モデルの推定結果 (東京 2013 年の平日, 一般道路での事故)

計数モデル (ポアソンモデル)					
	推定値	標準誤差	z 値	Pr(> z)	
切片	-6.837e+00	2.655e-02	-257.544	< 2e-16	***
用途地域 (Indust)	-2.905e-01	2.855e-02	-10.174	< 2e-16	***
用途地域 (Low_Res)	-6.394e-01	2.738e-02	-23.353	< 2e-16	***
用途地域 (Mid_Res)	-4.221e-01	2.884e-02	-14.636	< 2e-16	***
用途地域 (Res)	-2.447e-01	3.023e-02	-8.093	5.81e-16	***
用途地域 (Other)	-4.428e-01	5.890e-02	-7.517	5.62e-14	***
駐車場数	4.103e-02	2.734e-03	15.007	< 2e-16	***
店舗数	4.272e-04	2.413e-05	17.707	< 2e-16	***
ゼロ強調モデル (ロジットモデル)					
	推定値	標準誤差	z 値	Pr(> z)	
切片	1.5442	0.1820	8.486	< 2e-16	***
都市地域ダミー	-5.5269	0.3015	-18.329	< 2e-16	***
対数尤度	-6138				

表 4.5: ゼロ強調負の二項回帰モデルの推定結果（東京 2013 年の平日，一般道路での事故）

計数モデル（負の二項分布モデル）					
	推定値	標準誤差	z 値	Pr(> z)	
切片	-7.11446	0.09121	-78.002	< 2e-16	***
用途地域 (Indust)	-0.12655	0.08971	-1.411	0.15832	
用途地域 (Low_Res)	-0.46795	0.08882	-5.269	1.37e-07	***
用途地域 (Mid_Res)	-0.24926	0.08983	-2.775	0.00553	**
用途地域 (Res)	-0.06591	0.09325	-0.707	0.47972	
用途地域 (Other)	-0.42549	0.13197	-3.224	0.00126	**
駐車場数	0.06126	0.01102	5.562	2.67e-08	***
店舗数	0.00101	0.00012	8.137	4.04e-16	***
$\log(k)$	1.49613	0.05665	26.411	< 2e-16	***
ゼロ強調モデル（ロジットモデル）					
	推定値	標準誤差	z 値	Pr(> z)	
切片	1.3438	0.1941	6.923	4.43e-12	***
都市地域ダミー	-7.1449	1.1064	-6.458	1.06e-10	***
k	4.4644				
対数尤度	-4717				

さて，最後にゼロ強調モデルとそうではないモデルの比較を行う．従来この比較を行うために Vuong の検定が使われてきた．しかし Wilson²⁶⁾ は，ゼロ強調モデルとそうでない計数モデルの比較に Vuong 検定を使うのは誤りだと論じている．Vuong 検定に代わる方法として Wilson が提案しているのは，ゼロ強調モデルと非ゼロ強調モデルによって計算された個々のデータの確率をプロットすることである．すなわち，横軸にゼロ強調ではないモデルの確率（たとえばポアソン回帰モデルによって計算された確率）をとり，縦軸にゼロ強調モデルの確率（ゼロ強調ポアソン回帰モデルによって計算された確率）をとってデータ点をプロットする．もし点が直線 $y = x$ にほぼ沿うように並んでいればゼロ強調モデルを使う必要はないと判断するのである．

推定したポアソン回帰モデルとゼロ強調ポアソン回帰モデル，負の二項分布回帰モデルとゼロ強調負の二項分布回帰モデルの組合せそれぞれについてこのプロットを行った結果を図 4.1 に掲げた²⁷⁾．この図によるとポアソン回帰モデル，負の二項分布回帰モデルいずれの場合も直線 $y = x$ から外れており，ゼロ強調モデルの必要性が示されている．

以上，東京都の一般道路で 2013 年の平日に発生した交通事故を 1 km メッシュごとに集計して計数モデルを適用した結果を述べてきた．計数モデルとしては，ポアソン回帰モデル，負の二項分布回帰モデル，ゼロ強調ポアソン回帰モデル，およびゼロ強調負の二項分布回帰モデルの四つを推定し相互に比較を行った．その結果によると，事故件数が 0 のメッシュの数が多く，過大分散の状況にある事故データのモデル化にはゼロ強調負の二項分布回帰モデルが適していることがわかった．

この結果を踏まえて，東京都以外の都道府県で交通事

故の経度・緯度が高精度で計測されているところについてゼロ強調負の二項分布回帰モデルを適用した．分析に使ったのは 2013 年（暦年）の平日に発生した交通事故のうち，高速道路，都市高速道路，自動車専用道，道路運送法上の道路および一般交通の用に供するその他の道路（広場・駐車場等）で生じた事故を除いた事故である．

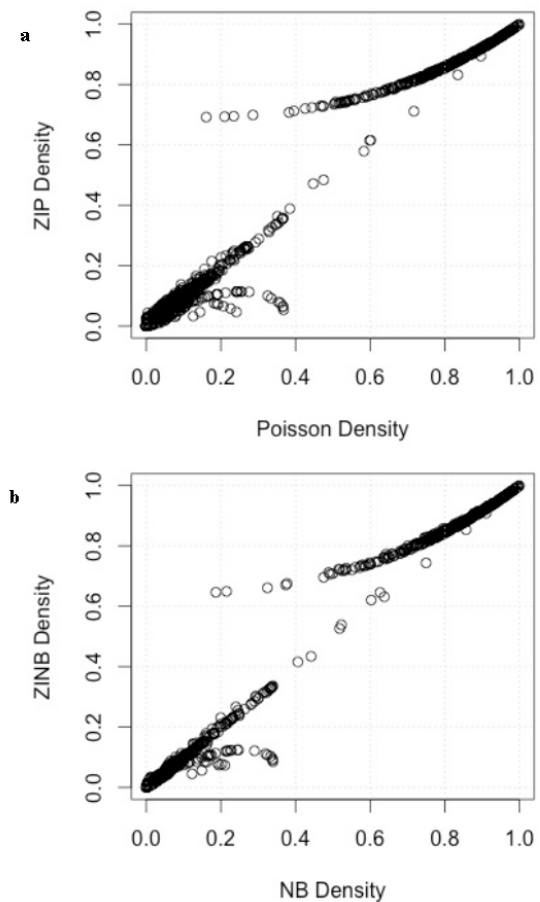


図 4.1: ゼロ強調モデルと非ゼロ強調モデルの確率プロット：上図はポアソン回帰モデル，下図は負の二項分布回帰モデルに対応している

2013 年の交通事故の発生場所の経度、緯度が高精度で与えられている次の 33 都道府県を分析対象とした。それぞれの都道府県について、1km メッシュごとに事故件数を集計しこれを目的変数とした。

- | | | | |
|--------|--------|--------|-------|
| ● 北海道 | ● 青森県 | ● 宮城県 | ● 福島県 |
| ● 茨城県 | ● 群馬県 | ● 埼玉県 | ● 千葉県 |
| ● 東京都 | ● 神奈川県 | ● 新潟県 | ● 石川県 |
| ● 福井県 | ● 長野県 | ● 静岡県 | ● 愛知県 |
| ● 三重県 | ● 滋賀県 | ● 京都府 | ● 大阪府 |
| ● 兵庫県 | ● 奈良県 | ● 和歌山県 | ● 鳥取県 |
| ● 島根県 | ● 岡山県 | ● 徳島県 | ● 香川県 |
| ● 高知県 | ● 福岡県 | ● 佐賀県 | ● 長崎県 |
| ● 鹿児島県 | | | |

第 3.2 節で解説したデータの中から、国勢調査の結果（人口、高齢者人口、世帯数、利用交通手段別就業者・通学者数）、経済センサスの結果（事業所数、従業者数）、交通量シミュレーション結果、気象データ（年間総降水量、年間総降雨日数、年間降雪日数、日最高気温が 0 度未満の日数）、DID、土地利用用途、POI データ、および道路延長と交差点数を説明変数として利用した。

都道府県別・平日休日別の事故件数モデルの推定結果のうち北海道・東北の分を以下の表 4.6 と 4.7 に掲げた。各道県のモデルの説明変数は変数選択の結果選ばれたものであり、都道府県別・平日休日ごとに異なるモデルが推定されている。表の中で推定された係数の欄に－が書かれている場合にはその説明変数が選択されなかったことを意味する。また表中の推定された係数の右側に書かれているアステリスクとピリオドは有意水準を表し、*** は 0.1% 水準で有意、** は 1% 水準で有意、* は 5% 水準で有意、. は 10% 水準で有意であることをそれぞれ意味する。

平日のモデルと休日のモデルを比較すると、平日モデルの方が有意に推定された説明変数の数が多く全体としてよいモデルになっている。これは休日の事故件数が平日と比べて少なく、多くのメッシュで事故件数が 0 になっていることの影響である。ゼロ強調部分の推定結果

で、都市地域や部分的都市地域（これらは 0, 1 のダミー変数で、該当するメッシュでは値が 1 となる）の係数がマイナスで推定されているのは、これらのメッシュでは事故件数がゼロではないことの反映である。同じくゼロ強調部分の説明変数道路総延長の係数が負で推定されているのも、道路総延長の長いメッシュ、すなわち都市的なメッシュでの件数がゼロでないことを意味している。

平日モデルの計数部分では、店舗数、スポーツ施設数、交差点数、自動車通勤・通学者数、事業所数、交差点数の係数が正で推定されている県が多い。このことは、店舗、施設、事業所が多数存在し、自動車で通勤・通学する人が多く、さらには交差点の多いメッシュで事故が発生しやすいことを示している。ただし東京都と鳥取県では事業所数の係数が負で推定されている。これは事業所の多い場所（主に都心）では歩道や信号機が十分整備されており、事故が抑止されていることを反映していると考えられる。さらに鳥取県では都心部の人口が少ないこともこの結果に影響していると考えられる。

気象関連の変数が交通事故に与える影響は様々ではない。年間降雨日数の係数が負の値として推定されている県が多いが、一方で正で推定されている県もいくつか存在する（埼玉県、東京都、福井県、愛知県、鳥取県）。この係数が負の場合には、雨の日には歩行者・自転車の交通量が少なくなり、車の運転も用心深くなるため事故が少なくなると解釈できる。一方、係数が正の値で推定されている場合には、こうした解釈は成り立たないと考えられる。年間降雪日数の係数も県によって正負がまちまちである。

休日モデルの計数部分に着目すると、多くの県で店舗数と住民数の係数が正で推定されている。これは、店舗が多く、多数の住民が居住しているメッシュで休日の事故件数が多いことを意味している。交差点数の係数も多くの県で正となっている。いくつかの県では、デパート数の係数が負で推計されているが、これは歩道や信号機が十分整備されている都心部での事故が少ないことを反映していると考えられる。

表 4.6: ZINB モデルの推定結果（北海道・東北地方）：2013 年平日の一般道路での交通事故

県名	北海道		青森県		宮城県		福島県	
計数部分								
切片	-6.5641	***	-8.5863	***	-9.3619	***	-8.5782	***
店舗数	0.0043	**	—		0.0049	***	0.0055	***
小学校数	—		—		—		—	
スポーツ施設数	0.0805	***	0.0962	***	—		—	
交差点数	0.0008	**	—		0.0005		0.0017	***
列車通勤通学者数	-0.0006	***	—		—		—	
自動車通勤通学者数	—		0.0005	***	0.0002	*	0.0003	**
徒歩通勤通学者数	—		—		—		—	
自転車通勤通学者数	0.0010	**	—		0.0008	***	—	
事業所数	0.0045	***	0.0024	***	0.0004		0.0013	*
年間降雨日数	-0.0188	***	-0.0063	***	—		-0.0064	***
年間降雪日数	—		0.0039		0.0074	**	—	
log(θ)	0.4267	***	0.8313	***	0.4622	***	0.3530	***
ゼロ強調部分								
切片	-1.1245	***	1.9310	***	1.5541	***	1.6435	***
都市地域	-17.7084		—		-6.9114		-7.2865	
部分都市地域	-8.5254		—		-6.7515	*	-6.9979	**
道路総延長	—		-0.0003	***	-0.0002	***	-0.0002	
対数尤度	-2944		-2849		-4135		-4638	
自由度	12		9		12		11	
AICc	5911.44		5715.62		8293.47		9297.85	

表 4.7: ZINB モデルの推定結果（北海道・東北地方）：2013 年休日の一般道路での交通事故

県名	北海道		青森県		宮城県		福島県	
計数部分								
切片	-8.2114	***	-8.5803	***	-10.4927	***	-9.3803	***
店舗数	0.0076	***	0.0059	***	0.0066	***	0.0057	***
デパート数	—		—		—		—	
交差点数	—		—		0.0016	***	0.0035	***
住民数	0.0002	***	0.0002	***	—		—	
年間降雨日数	-0.0107	**	-0.0071	***	—		—	
年間降雪日数	—		0.0050	*	0.0118	***	-0.0031	*
log(θ)	0.2910	*	0.5807	***	0.2892	**	0.2852	***
ゼロ強調部分								
切片	0.6940	.	1.929	***	0.5661	***	0.2007	*
都市地域	—		—		-13.6159		-13.2719	
部分都市地域	—		—		-12.0188		-14.4269	
道路総延長	-0.0002	***	-0.0003	***	—		—	
対数尤度	-2017		-2872		-2639		-4746	
自由度	7		8		8		8	
AICc	4047.92		5759.29		5294.83		9508.34	

4.1.2 メッシュ間の空間相関を考慮したモデル

前の 4.1.1 節の分析では、各メッシュはお互いに独立であるという前提が置かれていた。本節では、メッシュの間に空間的な相関関係があるという前提で事故件数の分析を行う。対象とする事故は、東京都（島嶼部を除く）の一般道路で 2013 年の平日に発生した全人身事故である。

まず、大域的な空間自己相関が存在するかどうかを Moran の I 統計量²⁰⁾ と Getis-Ord の G 統計量^{28), 29)} を使って確認した。これらの統計量を計算するには、隣接ポリゴンと空間重み付け行列を定義しておく必要がある。ここではルーク型（隣接する上下左右の四つのメッシュが隣接しているとする）とクイーン型（周囲を囲む八つのメッシュが隣接しているとする）の二つの定義を用いた。この定義を用いてバイナリータイプの空間重み付け行列を定めた。

統計量の計算結果は表 4.8 に示したとおりであり、Moran の I 統計量、Getis-Ord の G 統計量いずれを使った場合でも空間的自己相関があるという結論であった。しかし、これら大域的な統計量は空間的自己相関があるかないかを教えてくれるが、それがどこにあるかは教えてくれない。「どこにあるか」を知るためにはローカルな分析が必要であり、4.3 節で関連する分析を行う。

表 4.8: 東京の 2013 年平日事故の空間的自己相関

隣接ポリゴン	Moran's I	Getis-Ord G
ルーク型	0.693 ($p < 2.2\text{e-}16$)	4.31e-03 ($p < 2.2\text{e-}16$)
クイーン型	0.652 ($p < 2.2\text{e-}16$)	8.37e-03 ($p < 2.2\text{e-}16$)

この結果を踏まえて、空間的な自己相関を考慮に入れたポアソン回帰モデルを MCMC で推定した。推定には統計ソフト R の CARBayes パッケージを使った³⁰⁾。

推定した CAR モデルの概要は以下に述べるとおりである。分析の対象地域 $\mathcal{S}$ は互いに重複しない K 個のエリア（ここではメッシュ）に分割されているものとする。すなわち、 $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_K\}$ 。それぞれのエリアの目的変数（事故件数）を $\mathbf{Y} = (Y_1, \dots, Y_K)$ ，オフセット変数（ここでは道路延長）を $\mathcal{O} = (O_1, \dots, O_K)$ ，説明変数を $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_K)$ ， $\mathbf{x}_k^\top = (x_{k1}, \dots, x_{kp})$ ，ランダム効果を表す項を $\phi = (\phi_1, \dots, \phi_K)$ とする。モデルは一般化線型混合モデル (GLMM) を使って次のように表現される（以下の式はポアソン回帰モデルを前提としている）。

$$Y_k | \mu_k \sim \text{Poisson}(\mu_k), \quad k = 1, \dots, K \quad (4.7)$$

$$\log(\mu_k) = \mathbf{x}_k^\top \boldsymbol{\beta} + O_k + \phi_k, \quad (4.8)$$

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta, \Sigma_\beta) \quad (4.9)$$

ここに $\text{Poisson}(\mu_k)$ は期待値 μ_k をもつポアソン分布を表し、式 (4.8) は μ_k のリンク関数である。また式 (4.9) は説明変数の係数ベクトル $\boldsymbol{\beta}$ の事前分布を与え、 $\mathcal{N}(\cdot, \cdot)$ は多変量正規分布を表す。デフォルトでは $\boldsymbol{\mu}_\beta = \mathbf{0}$ ， $\Sigma_\beta = \text{diag}(1000, \dots, 1000)$ となっている。

ランダム項の分布は Leroux *et al.* の式に従う³⁰⁾。すなわち、

$$\begin{aligned} & \phi_k | \phi_{-k}, W, \tau^2, \rho \sim \\ & \mathcal{N} \left(\frac{\rho \sum_{i=1}^K w_{ki} \phi_i}{\rho \sum_{i=1}^K w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{i=1}^K w_{ki} + 1 - \rho} \right), \quad (4.10) \\ & \phi_{-k} = (\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_K), \\ & \tau^2 \sim \text{Inverse-Gamma}(a, b), \\ & \rho \sim \text{Uniform}(0, 1). \end{aligned}$$

ここに ρ は空間的自己相関の程度を表すパラメーターで、 $\rho = 0$ なら自己相関はなく、 $\rho = 1$ なら強い自己相関があることを意味する。 $\rho = 1$ は intrinsic CAR モデルに対応する。式 (4.10) に現れる重み付け行列 W は、エリア $\mathcal{S}_k$ と $\mathcal{S}_j$ が境界を共有しているならば対応する要素 $w_{kj} = 1$ ，そうでなければ $w_{kj} = 0$ で定義される。

Lee はさまざまなモデルを比較して Leroux *et al.* による上の式がもっともよいと結論づけている²²⁾。ここでもこの式を使って推定を行った。

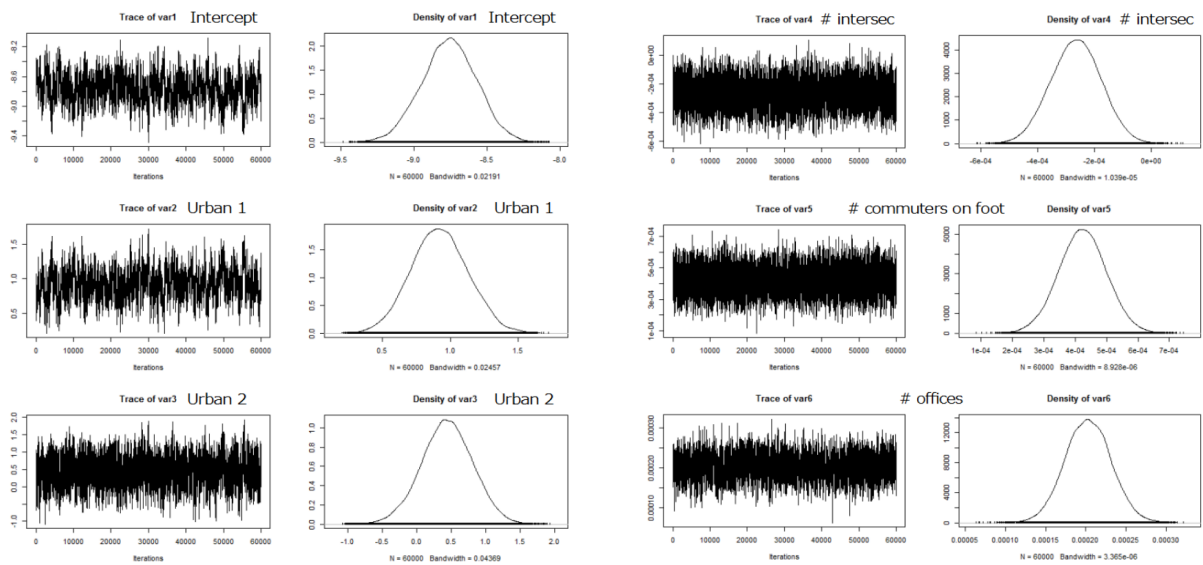
クイーン型の隣接ポリゴンを用いた場合の推定結果（事後分布の中央値他）を表 4.9 に提示した。ルーク型の隣接ポリゴンを使った場合の推定結果は省略したが、 ϕ_k の分散の大きさを表す τ^2 を除けば表 4.9 とほぼ同じ結果であった。クイーン型とルーク型では隣接するポリゴンの数が異なるので τ^2 の結果が違っていると推測される。

図 4.2 にはポアソンモデルの係数のトレースプロットを示した。さらに、図 4.3 には事後分布の中央値を使って推定した事故件数の残差の値を示した。同図には表 4.9 と同じ説明変数をもつランダム効果項を含まない単純なポアソン回帰モデルの残差の値もあわせて掲げた。

図 4.3a と図 4.3b を比べてみると、ランダム効果を含んだポアソンモデルでは残差の値が小さく、ランダム効果を含まないポアソンモデルのおおよそ 1/10 のオーダーになっていることがわかる。さらにこの二つのコロプレス図を比べると、ランダム効果をもつポアソンモデルの残差は正負の値がランダムにばらついていますが、ランダム効果をもたないポアソンモデルの残差は正または負の値を持つメッシュが隣接して存在することがわかる。

表 4.9: CAR モデルによるポアソン回帰モデルの予測結果：都市地域 1 は「非都市地域」、同 2 は「都市地域と非都市地域の混在」を意味する。Geweke は Geweke の統計量である。都市地域 2 の推定結果を見ると 95% タイル区間に 0 が含まれており、この変数は有意とはいえないことがわかる。また ρ の値はほぼ 1 になっている。

変数	中央値	2.5 %	97.5 %	Geweke
切片	-8.758	-9.132	-8.397	1.5
都市地域 1	0.917	0.514	1.331	-1.6
都市地域 2	0.440	-0.311	1.165	-2.1
交差点数	-0.00026	-0.00044	-0.000091	0.6
徒歩通勤者数	0.00042	0.00027	0.00057	-1.1
事業所数	0.00020	0.00015	0.00026	0.5
τ^2	1.08	0.957	1.215	1.1
ρ	0.995	0.977	0.999	0.9

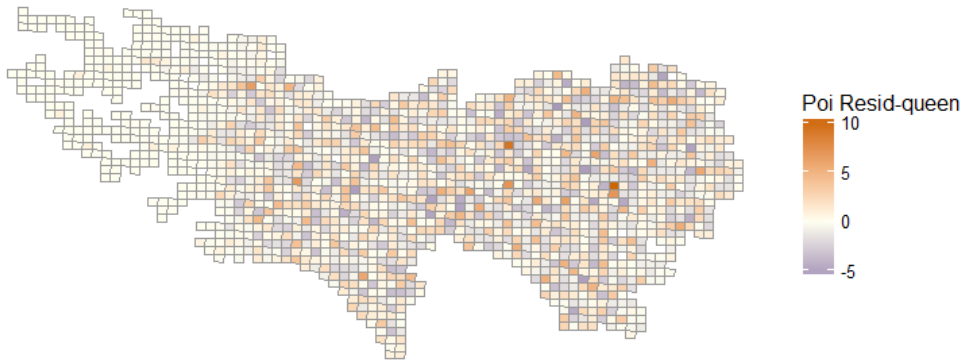


(a) トレースプロット 1：切片，都市地域ダミー変数

(b) トレースプロット 2：交差点数，徒歩通勤者数，事業所数

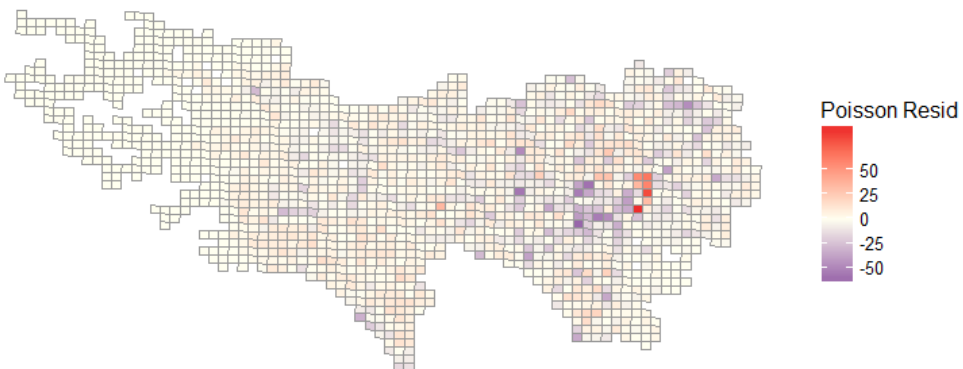
図 4.2: ランダム効果項を含んだポアソン回帰モデルの予測結果：トレースプロット

1 kmメッシュの事故件数の残差 (ポアソン・ランダムモデル・クイーン型)
東京都 2013年:平日，一般道路



(a) ランダム効果項を含んだポアソン回帰モデルの残差

1 kmメッシュの事故件数の残差 (ポアソンモデル)
東京都 2013年:平日，一般道路



(b) ランダム効果項を含まないポアソン回帰モデルの残差

図 4.3: ポアソン回帰モデルの残差:ランダム効果項の有無による比較

実際、グローバルな Moran's I の値を計算したところ、ランダム効果項をもたないモデルの残差では「空間相関がない」という帰無仮説は棄却された ($p = 0.001$) のに対しランダム効果項をもつモデルの残差については同じ帰無仮説は棄却されなかった ($p = 0.997$)。

ランダム効果項の期待値に関わるパラメーター ρ の値はほぼ 1 に近く、近隣のメッシュとの間には強い自己相関が存在するという結果が得られた。ランダム効果を含むモデルでは含まないモデルと比べて元データの値の再現性が高くなっている。この結果は、交通事故の分析において空間相関を考慮することの重要性を示していると解釈できる。

4.1.3 まとめと考察

最後に本 4.1 節の結果をとりまとめ考察を加える。

本節では交通事故の件数をモデル化することを試み、さまざまなモデルの比較を行った。交通事故は 0 または正の整数値をとるデータなので、こうしたデータをモデル化することができる計数モデルを検討対象としてとりあげた。具体的には、ポアソン回帰モデル、負の二項分布回帰モデル、そしてこれらのゼロ強調バージョンを対象として検討を行った。わが国ではこれまで交通事故の発生場所の情報が存在しなかったため、こうした分析はほとんど行われていない。

こうしたモデルを構築する第一の目的は、交通事故の件数を予測し事前に対策をとることであり、モデル化を通じて事故の発生に影響を与える要因を特定し事故の抑制を図ることを目指している。さらに、モデルによって予測される件数と実際の発生件数を比較して事故の危険性が高いエリアを特定しようとする場合にもこれらのモデルを使うことができる。

しかしながら交通事故の場合には発生件数が 0 となる地域が多く、過大分散の問題が発生することが知られている。この問題に対処するために、負の二項分布回帰モデル、さらにはゼロ強調モデルが導入されてきた。Shankar *et al.*³¹⁾ によれば、ゼロ強調モデルは 1990 年代頃から使われてきたことがわかっている。Shankar *et al.*^{31), 32)} はこうしたモデルの比較を行っているが、4.1.1 節で述べた Vuong の検定を適用可能な範囲を越えて使っているため（ゼロ強調負の二項分布モデルと負の二項分布モデルの比較などが行われている）、比較結果には意味がない。

本節で東京都の交通事故（2013 年の平日に一般道路で発生した人身事故）を対象に上述の 4 種類のモデルを推定し比較を行ったところ、ゼロ強調負の二項分布モデ

ルが最も適合度が高いという結果であった。ポアソン回帰モデルおよびゼロ強調ポアソン回帰モデルでは過大分散となっていることも判明した。

さらに、メッシュの間の空間的自己相関とランダム効果項を導入したポアソン回帰モデルも MCMC を使って予測した。このモデルはデータへの適合という点では一番よい結果を与えてくれた。

交通事故をモデル化する際に過度な 0 の存在とあわせて問題になるのが説明変数である。事故に関係すると考えられる要因には、社会経済的な要因、天候、路面条件、道路要因、交通要因がある。事故に関連するデータ整備について述べた第 3.2 節に記載したように、社会経済的な要因、天候については各種データを整備することができた。しかしながら、道路要因、交通要因については十分なデータが存在しないという問題が依然として残っている。

道路要因については、道路線形のデータが存在しないという大きな問題がある。道路の平面線形や縦断線形のデータはほとんど整備されていない状況である。とくに曲率半径、縦断線形はデータがなくこれらをモデルに組み入れることはできなかった。道路の幅員データも幹線道路以外は存在しない。

交通要因についても状況はあまり変わらない。DRM にはセンサス対象道路のセンサス交通量が組み込まれているが、交通センサス調査が 5 年に一度しか行われないため、事故発生時点の交通量を使うことができない。さらに、センサス調査の対象となっていない道路の交通量は全くわからない。こうした状況を考慮すると、スマートフォン・携帯電話の GPS データあるいはプローブデータを使って交通量を推定する方法を追究することが必要だと思われる。GPS データを使った試みについては、たとえば Chen *et al.*³³⁾ を参照されたい。

4.2 交通事故のパターン分析

本節での分析の目的は、死亡事故、重傷事故および事故多発箇所空間的な分布パターンに差があるかどうかを統計的に検証することである。検証には点パターンの分析手法を用いた。

分析に使うデータとして、2013 年に発生した人身事故のうち、一般幹線道路（一般国道、主要地方道、一般都道府県道）で発生した事故を用いた。高速道路、自動車専用道、市町村道、農林道、駐車場・広場等で発生した事故は除外した。

事故多発箇所の分析には ITARDA 区間を用いた。

ITARDA 区間とは、幹線道路のネットワークをまず単路とネットワークに分け、単路についてはリンクをさらに長さ約 300 m の区間に再分割したものであり、従来から道路管理者による交通事故の分析に広く用いられている。区間の数は単路で約 60 万、交差点で約 24.8 万である。

年間 4 件以上の人身事故が発生している ITARDA 区間（単路、交差点）を多発箇所として抽出した。ただし、抽出に当たっては単路と交差点を区別せず、件数の多い方から順に抽出した。その結果多発箇所数は全国で 16,856 箇所となった。抽出された多発箇所の中で事故件数が最も多いところでは 54 件の事故が発生していた。

多発箇所の場所を表す代表点として、交差点については DRM の交差点の代表ノードを用いた。DRM の代表ノードでは、交差点内に複数のノードが存在する場合には一番小さな番号をもつノードが選ばれている。このため、交差点の中心点の座標とは必ずしも一致しない。交差点の中心点と代表ノードの位置がずれるのは、交差点で交差する道路の中に二条線で表現されている道路が存在する場合である。交差する道路がどちらも一条線で表現されている場合にはこうした問題は生じない。

単路区間（リンク）については、始点と終点の経度・緯度を平均した単路の中心点の経度緯度で単路を代表させた。実際には単路は長さを持った線要素であるが、以下の分析では中心点で代表させた。

死亡事故・重傷事故の発生位置については、事故原票に付与された経度・緯度の値を使った。これらの経度・緯度は WGS 84 にもとづいて計測された値である。分析の対象とした東京都市圏、近畿都市圏においては、経度・緯度の値は高い精度で計測されていることが確認されているので、経度・緯度の値の修正は行っていない。

分析の対象としたのは、東京都市圏と京阪神都市圏の二つである。東京都市圏では三鷹市を中心とする半径 30 km のエリアを、京阪神都市圏では守口市を中心とする半径 50 km のエリアをそれぞれ分析対象とした。各都市圏で用いた中心の位置と半径の値は、円内に入る海域をできるだけ少なくする一方、都市域をできるだけ多く含むよう試行錯誤的に決定した。

分析に用いる死亡・重傷事故データを以下の手順で作成した。

- (1) 「国道」「主要地方道」「一般都道府県道」で発生した事故を抽出する。
- (2) 東京都市圏、近畿都市圏のデータを取り出す。東京都市圏は埼玉県、千葉県、東京都、神奈川県とし、近畿都市圏は滋賀県、京都府、大阪府、兵庫県、和歌山県、奈良県とした。

- (3) 交通事故が起きた地点の経度あるいは緯度が 0 となっているデータを除外する。
- (4) 三鷹市役所、守口市役所の経度・緯度をもつ点データを作成する。
- (5) 東京都市圏において、三鷹市役所を中心とし半径が 30 km の円内に含まれる多発箇所データと死亡・重傷事故データを取り出す。
- (6) 近畿都市圏において、守口市役所を中心とし半径が 50 km の円内に含まれる多発箇所データと死亡・重傷事故データを取り出す。
- (7) 死亡事故、重傷事故および多発箇所を対象にカーネル密度推定を行い、結果をマップ上に描く。

図 4.4 と図 4.5 にそれぞれの都市圏における死亡事故、重傷事故および多発箇所の分布と KDE の結果を示した。これらカーネル密度推定の結果から、次の結論が導かれる。

- 東京都市圏では死亡事故は東京の都心において高い密度で発生しており、周辺に行くにしたがい密度は低くなる。さらに、東京都心から南北方向に死亡事故の密度が高い地帯が見られる。
- 一方、重傷事故は東京都心では密度が低く、むしろ埼玉県で密度が高くなっている。この原因として、東京都の重傷事故の件数が周辺の県と比べ著しく低いことをあげることができる。2013 年に東京都で発生した重傷事故（発生場所は国道、主要地方道、一般都道府県道に限定されない）の件数は 288 件であるのに対し、埼玉県では 3,046 件、神奈川県では 1,201 件の重傷事故が発生しており、その差は大きい。
- 多発箇所はやはり東京都心で密度が高いが、密度の高い地域は東西に広がっているようにみえる。
- 東京都市圏と比べると、京阪神都市圏では死亡事故、重傷事故、多発箇所の間の分布パターンの差は小さい。
- 死亡事故、重傷事故、多発箇所いずれの場合も、大阪の都心部と京都の都心部の密度が高くなっている。
- ただし、重傷事故については奈良市のあたりでも密度が高くなっている。

なお、表 4.10 には、過去 5 年間の東京圏および京阪神都市圏の都府県別の死亡事故と重傷事故の件数を示した³⁴⁾。ただし、この表に示したのは各都府県の事故の総件数であり、幹線道路での事故には限定されていないので注意されたい。この表によると、死亡事故の件数に関しては、滋賀県、奈良県および和歌山県の件数が小さいことを除けば各都府県の件数に大きな差はない。しかし、重傷事故については、東京都の件数が他の府県に

比べ圧倒的に少ないことがわかる。実際、東京都の重傷事故件数は奈良県や和歌山県より少ないのである。しかも、5年間で重傷事故の件数は半分以下に減少しており、減少率も際だって大きい。

こうした東京都の重傷事故件数の特異性がどのような

要因によってもたらされているのかは現時点では不明である。考えられる要因としては、交通安全施設の整備が進んでいること、取締りが厳重に行われていること、緊急医療体制が整っていることなどをあげることができるが、憶測の域を出ない。

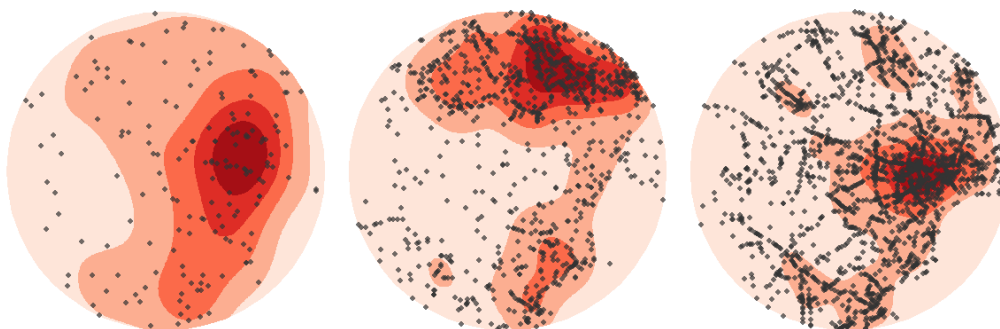


図 4.4: 東京都市圏における死亡事故（左）、重傷事故（中）及び多発箇所（右）のカーネル密度推定結果（2013 年）

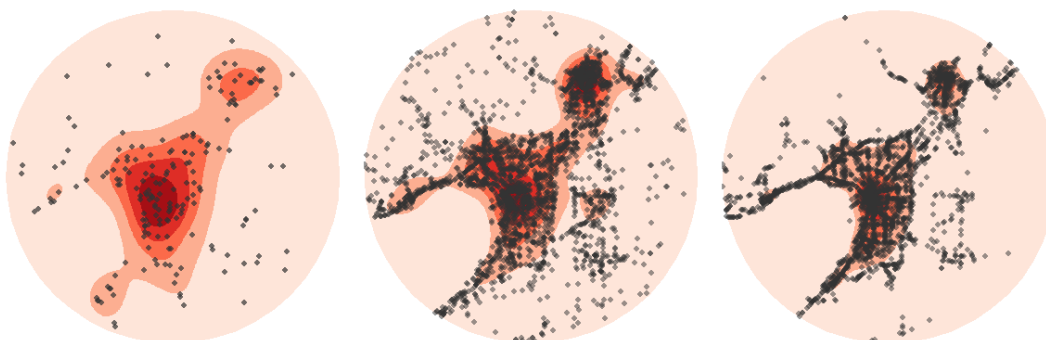


図 4.5: 京阪神都市圏における死亡事故（左）、重傷事故（中）及び多発箇所（右）のカーネル密度推定結果（2013 年）

表 4.10: 東京圏、京阪神都市圏の都府県別死亡・重傷事故件数（過去 5 年間）

	埼玉県	東京都	神奈川県	千葉県	滋賀県	京都府	大阪府	兵庫県	奈良県	和歌山県
死亡事故件数										
2010	198	214	180	179	77	95	197	187	44	51
2011	202	212	180	171	83	103	187	190	46	53
2012	192	180	177	171	74	96	178	177	48	50
2013	175	164	168	184	69	67	179	185	41	47
2014	170	169	179	175	61	67	143	181	45	39
重傷事故件数										
2010	3821	509	1543	2138	726	1298	3474	3537	777	820
2011	3584	570	1373	2043	690	1189	3394	3288	683	715
2012	3365	412	1368	2158	601	1388	3255	2995	611	742
2013	3046	288	1201	2153	598	1488	3053	2798	587	684
2014	3008	221	1206	2025	564	1464	2829	2591	589	676

以上で準備が終わったので、マーク付き点過程の分析を始める。マークは死亡事故、重傷事故、そして多発箇所（三つ）になる。以下では、死亡事故と重傷事故のパターンに差があるかどうか、死亡事故と多発箇所のパターンに差があるかどうか、そして重傷事故と多発箇所のパターンに差があるかどうかという3つの分析を順に行った結果を示す。分析にはクロス L 関数を使った。

クロス L 関数はクロス K 関数から計算される。クロス K 関数はマークが i の点からある一定の距離 r 以内にあるマーク j の点の数をもとに計算される。すなわち、

$$\text{Cross } K_{ij}(r) = \lambda_j E[K_{ij}(r)] \quad (4.11)$$

と書くことができる。ここに、 λ_j はマークが j の点の平均的な密度（単位面積あたりの点の数）、 $K_{ij}(r)$ はマークが i の点からある一定の距離 r 以内にあるマーク j の点の数で、 $E[\cdot]$ は期待値を表す。

クロス L 関数は次式で定義される。

$$\text{Cross } L_{ij}(r) = \sqrt{\frac{\text{Cross } K_{ij}(r)}{\pi}}. \quad (4.12)$$

マーク i の点のパターンと j の点のパターンが独立ならば、クロス L 関数の理論的な値は r で与えられる。すなわち、距離 r の線形関数となる。このためデータから計算されるクロス L 関数の分散はクロス K 関数よりも安定しており、包絡線の計算、仮説検定を行う上で有利である。

包絡線は R の `envelope()` 関数を使って計算できる。この関数は引数 `nsim` で与えられた回数のシミュレーションを行って包絡線を計算する。シミュレーションでは、CSR (Complete Spatial Randomness) に従う点パターン（以下の計算では、比較対象となっている二つのマークに対応する点パターン）を生成し、それらを使ってクロス L 関数の値を r の値ごとに計算してその最大値と最小値を求める。この最大値と最小値を包絡線としてプロットして、データから計算されたクロス L 関数の値と比較する。

しかしながら、Brunsdon & Lex¹⁶⁾ はクロス L 関数の値をプロットして包絡線と比較するのは、「究極の多重比較」なので避けるべきだと書いている。ただし、Baddeley *et al.*³⁵⁾ はシミュレーションによって求めた包絡線を使った検定は必ずしも全面的に否定されるべきものではなく、距離 r を固定して検定を行うのであれば有効だとしている。もっとも、距離を固定した検定にそれほど意味があるわけではないので、この陳述にそれほど意味はない。

包絡線を使った検定に代わる方法として考案されたのが DCLF 検定と MAD 検定である。DCLF はこれを考

案した Diggle, Cressie, Loosmore and Ford の頭文字である。また、MAD は Maximum Absolute Deviation（乖離の絶対値の最大値）を表し、Repley³⁶⁾ によって考案された。これらはデータに対して一つの統計量を算出してくれるので、多重比較の問題を避けることができる。

以下ではこれらの検定法の概要を述べる。以下の記述は上述の Baddeley *et al.*³⁵⁾ と Loosmore and Ford の論文³⁷⁾ による。

DCLF 統計量は次式で定義される。

$$u_i = \int_0^R [H(r) - H_{\text{thoe}}(r)]^2 dr \approx \sum_{t_k=t_{\min}}^{t_{\max}} [\hat{H}_i(t_k) - \bar{H}_i(t_k)]^2 \delta t_k. \quad (4.13)$$

ここに、 t_k は距離（の分点）、 $t_{\min}$ と $t_{\max}$ は距離の最小値と最大値、 $\hat{H}_i(t_k)$ は観測された空間データ i の検定統計量（ G , K , あるいは L 関数）の値、 $\bar{H}_i(t_k)$ は i を除いて計算した期待値、 $\delta t_k = t_{k+1} - t_k$ は区分された距離の幅である。 $\bar{H}_i(t_k)$ は次の式で計算できる。

$$\bar{H}_i(t) = \frac{1}{s-1} \left\{ \sum_{j=1, j \neq i}^s \hat{H}_j(t) \right\}. \quad (4.14)$$

s はシミュレーションの回数を表す（番号 1 は観測されたデータに対応しており、シミュレーションには 2 から s までの番号がふられているものとする）。大数の弱法則によれば、 $\lim_{s \rightarrow \infty} \bar{H}_i(t) \rightarrow H(t)$ がなりたつ（ $H(t)$ は検定統計量の理論値）。したがって、 u_i は観測された点パターンと理論的な結果の差の二乗を関心のある距離の範囲で足し合わせた量（正確には、積分した量）となっている。

（帰無仮説 H_0 を棄却するかしないかの）統計的検定は、 $s-1$ 個のシミュレーションの計算結果 u_i ($i = 2, \dots, s$) を大きさの順に並べて行う。検定の対象となる帰無仮説は、「観測された点パターンが帰無仮説にもとづいて生成された」というものではない。帰無仮説 H_0 は距離 $t \in [t_{\min}, t_{\max}]$ において計算された観測データの統計量（ G または K ）にもとづく u_i の値が仮定された空間過程のモデルからランダムに生成された値と異なっていない、というものである。

仮説検定における観測された P 値（ $\hat{p}$ と記す）は、次式で計算できる。

$$\hat{p} = 1 - \frac{\text{rank}[u_1] - 1}{s} = 1 - \frac{\sum_{j=2}^s I(u_1 > u_j)}{s} \quad j = 2, \dots, s \quad (4.15)$$

ここに $I(u_1 > u_j)$ は特性関数で、 $u_1 > u_j$ なら値 1 をとり、それ以外の場合は 0 である。この式はシミュレー

ション計算によって P 値を求めていることを意味する。これによって第 1 種の過誤の値がコントロールされる。また、MAD 統計量は次式で定義される。

$$\text{MAD} = \max_d \left| \hat{H}(d) - H_{\text{CSR}}(d) \right|. \quad (4.16)$$

ここに $H_{\text{CSR}}(d)$ は点過程が CSR であると仮定したときの関数 H の値である。

上記の手続きによって求められる u_i あるいは MAD の値は、点の分布パターンが帰無仮説にしたがう場合の検定統計量の分布を示す。シミュレーションの回数を増やせば、この値の不確実性（ばらつき）は減少するが、真の P 値を知ることはできない。しかし、シミュレーション回数と $\hat{p}$ 値のばらつきの程度を推測することはできる。

以上を踏まえて、東京都市圏を対象に事故多発箇所、死亡事故発生箇所および重傷事故発生箇所の分布パターンの間に統計的に有意な差があるかどうかを検定する。

東京都市圏の死亡事故と重傷事故の分布パターンについて Cross L 関数を使って分析した結果を図 4.6 に掲げた。図中の黒い実線は観測値から求めたクロス L 関数、赤い線は理論分布（CSR）から求めたクロス L 関数、灰色の領域の上側の線がシミュレーションで求めたクロス L 関数の最大値、下側の線が最小値である。観測値によるクロス L 関数はシミュレーションで求めた最大値の値より大きく、したがって死亡事故と重傷事故の分布パターンの間には「有意な差がある」と結論づけるのが従来のやり方であるが、上述したようにこの方法は多重比較を行っていることになるので、DCLF 検定と MAD 検定を行った。

マーク付き点過程の分析を行った結果は表 4.11 のようになり、以下の結論が導かれる。

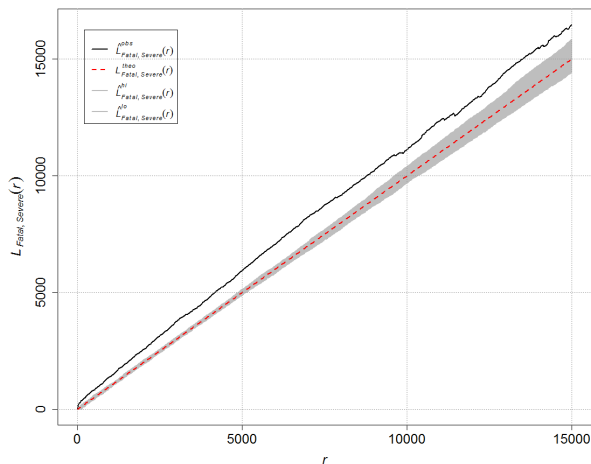


図 4.6: 東京都市圏における死亡事故と重傷事故のクロス L 関数と包絡線

表 4.11: 各都市圏における点パターンの分析結果

都市圏	DCLF 検定	MAD 検定
東京都市圏		
死亡事故－重傷事故	1.79e+10 (p=0.002)	1530 (p=0.002)
死亡事故－多発箇所	1.91e+11 (p=0.002)	5810 (p=0.002)
重傷事故－多発箇所	5.36e+10 (p=0.002)	4031 (p=0.002)
京阪神都市圏		
死亡事故－重傷事故	2.42e+12 (p=0.002)	14794 (p=0.002)
死亡事故－多発箇所	3.53e+12 (p=0.002)	17688 (p=0.002)
重傷事故－多発箇所	3.11e+12 (p=0.002)	16284 (p=0.002)

● 東京都市圏

- ▷ 死亡事故と重傷事故，死亡事故と多発箇所，そして重傷事故と多発箇所の空間パターンには統計的に有意な差がある。
- ▷ この結論は，先に示したカーネル密度推定の結果からも予想される結果である。
- ▷ 特徴的なことは，東京都における重傷事故の件数が隣接する埼玉県，神奈川県と比べ著しく少なく，このことが空間パターンの間に差を生み出す要因の一つとなっていることである。

● 京阪神都市圏

- ▷ 近畿都市圏においても，死亡事故と重傷事故，死亡事故と多発箇所，重傷事故と多発箇所の分布パターンの間には統計的に有意な差がある。
- ▷ カーネル密度推定の結果からすると，この結論はやや意外である。
- ▷ こうした結論が導かれた背景には，事故や多発箇所のデータ数が数千と多く，微妙な差であっても統計的には「有意な差がある」という結論が導かれやすいという事情があるのではないかと推察される。

4.3 交通事故のクラスタリング分析

事故の多発箇所の特定は古くから研究されてきたテーマであり，特定のための手法がいくつか考案されている。多発する場所を特定することが目的なので，当然のことながら事故の発生場所の位置情報を使う必要がある。

最も広く用いられているのは，カーネル密度推定（KDE: Kernel Density Estimation）を使う方法である。KDE は GIS ソフトにも標準的に備わっており簡便に使うことができるという利点を有するが，一方で次のような欠点があることも指摘されている。

- カーネル関数やそのバンド幅の選択によって得られる結果（見栄え）が異なる。カーネル関数の選

択は結果にそれほど大きな影響を与えないが、バンド幅は結果に大きく影響する。バンド幅が小さければ、局所的な多発箇所が強調される一方で全体の傾向はつかみにくくなる。バンド幅が大きければ、全体の傾向はつかめるが局所的な多発箇所は埋没してしまう。最適なバンド幅についての研究も行われている¹⁶⁾が、その結果が必ずしも最適ではないという指摘もある。

- 従来の KDE では二次元のカーネル密度関数が使われるため、結果も二次元の密度となる。これは道路でないところにも 0 ではない事故の発生確率が存在することを意味する。

二つ目の点については、道路の上だけで密度を計算するための手法が開発され^{38),39)}、ネットワーク KDE と呼ばれている。さらに、ネットワーク KDE を計算するためのソフトも公開されている⁴⁰⁾。

ネットワーク KDE は優れた方法であるが、しかし事故分析の観点から見ると次のような点で課題が残っている。

- 交通事故が道路の上だけで発生するという前提が置かれているが、実際には道路以外でも事故は発生する。典型的な例としては、駐車場における事故がこれに該当する。こうした事故まで含めて事故分析を行おうとする場合には、ネットワーク KDE では十分ではない。
- 単路での交通事故と交差点での交通事故は区別されずに一括して扱われている。しかし、車両の右直事故は交差点に特有の事故であって単路では発生しない。にもかかわらず、ネットワーク KDE ではこうした事故の特性を考慮することができない。
- 計算の負荷が大きく、とくに大規模な道路ネットワークを対象にして計算を行うのは難しい。

さらに、ネットワーク KDE に限らず 2 次元の KDE にも共通する課題として、

- 交通事故に対する曝露量を取り込むことができない
- 事故件数の多寡について統計的な検定を行うことができない

という 2 点を指摘することができる。

交通事故に対する曝露量としては交通量、走行台キロ等が使われることが多い。これは交通量や走行台キロの多い場所ではそれに応じて事故もより多く発生するため、単純に事故の件数だけで多発箇所を選ぶと、事故率

の高い箇所を見落とす危険があるためである。しかし、常に曝露量を使って事故率に直すことがよいとは限らない。たとえば、交通量の非常に少ない箇所でもたまたま交通事故が発生したとすると、事故率は極めて高くなってしまう。こうした箇所を多発箇所とすることには疑問がある。KDE あるいはネットワーク KDE は一件一件の事故の発生場所をもとにして密度を計算する手法なので、事故率をもとにして密度を計算することができない。ただし、ネットワーク KDE は道路を等間隔の短いリンク (lixel と呼ばれる) に区切って密度を計算しているので、実質的には道路の単位長さ当たりの事故率を比較していると解釈することができる。

上記の二番目の課題については、統計的な検定を行うための手法がいくつか考案されている^{41),42)}。これらはシミュレーション計算によって、有意性を求めるものである。

以下の分析では、1 km メッシュ単位に集計した人身事故件数を対象にしてメッシュの空間的なクラスタリングを分析した結果を示す。ここで「クラスタリング」というのは、事故件数の多い (または少ない) メッシュが空間的に集まっている箇所を意味する。ただし、交通事故の場合には事故件数の少ないメッシュは重要ではないので、事故件数の多いメッシュに着目して分析を行った。また、平日と休日では分布パターンに差がある可能性があるため、平日と休日の事故を別々に分析した。休日は、日曜日、祝日と正月三が日とした。

分析の対象としたのは、2013 年に発生した人身交通事故のうち、一般道路で発生した事故である。本節の分析では、一般国道、主要地方道、一般都道府県道、市町村道、港湾道路、農林道、私道で発生した事故が含まれている。高速道路、都市高速道路、駐車場・広場等で発生した事故は分析の対象となっていないので留意された。

1 km メッシュ単位に集計したのは、これより辺長の短いメッシュ (500 m あるいは 250 m) を使うと、とくに地方部においては事故件数が 0 のメッシュが大多数を占めるようになってしまい、事故が存在する都市部のメッシュだけを対象にした分析と同じことになってしまうからである。さらに、一般道路が存在しないメッシュは分析の対象から外した。これはこうしたメッシュでは交通事故が発生することはないと考えられるからである。

以下では東京都を対象にクラスタリングの分析を行った結果を詳述し、これ以外の都府県の分析結果はその後まとめて示すことにする。なお、分析では空間参照系として UTM 53N / 54N を使った。

まず、平日・休日別にメッシュ単位に集計した事故件数のコロプレス図を図 4.7 と図 4.8 に示した。休日の事故件数は平日に比べて少ないが、事故件数の分布パターンは平日と休日で類似している。都心部に加えて下町地

域、西側の郊外部で事故が多いことがわかる。紙幅が限られているため、以下では平日の分析結果のみを示すことにする。

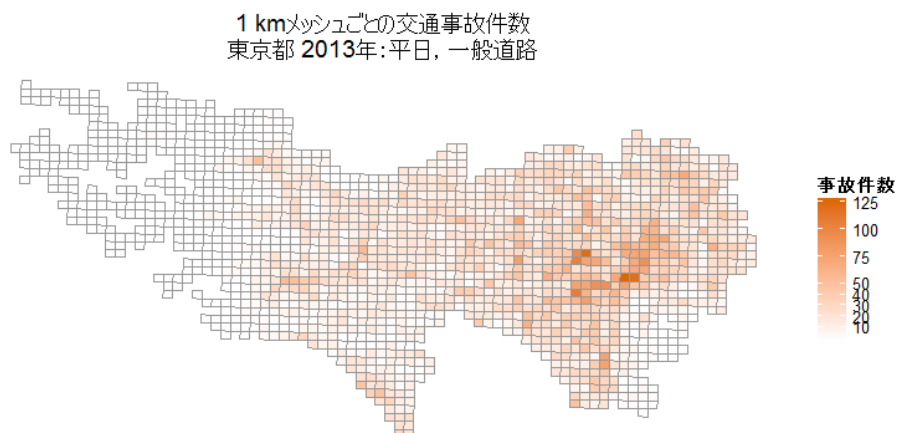


図 4.7: 東京都における一般道路での事故件数（2013 年，平日；島嶼部は除く）

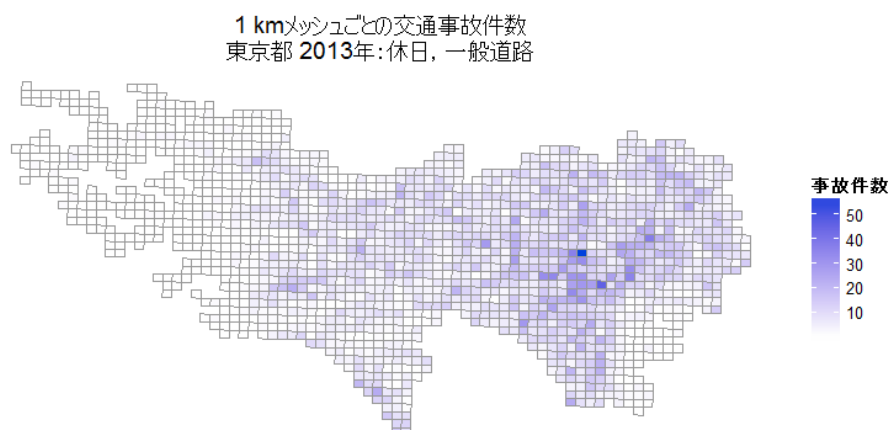


図 4.8: 東京都における一般道路での事故件数（2013 年，休日；島嶼部は除く）

Brunsdon & Lex のやり方¹⁶⁾に従って、メッシュの事故件数のクラスタリングを分析する。こうした分析に使われるのは、次の性質を満たす「空間的相関のローカル指標」(LISA: Local Indicators of Spatial Association)である⁴³⁾。

- 各観測値の LISA は、当該観測点の周辺の空間クラスタリングの有意な広がり の程度を表すことができる。
- LISA の値をすべての観測値について合計した値はグローバルな空間相関の値に比例する。

この二つの性質を満足する指標として、ローカル・モラン I, Ord の G 統計量がある。ここでは Ord の G 統計量を使って 1km メッシュごとに集計した事故件数の

クラスタリングを分析する。

Ord の G 統計量は次式で定義される。

$$G_i = \frac{\sum_{j=1}^n z_j v_{ij}(d)}{\sum_{j=1}^n z_j}. \quad (4.17)$$

ここに、 $v_{ij}(d)$ はポリゴン i と j のセントロイド間の距離が d 未満のとき値 1 を、そうでないときは値 0 をとる関数であり、 z_i は関心のある観測量（ここでは、事故件数）である。上の式からわかるように、G 統計量は観測量の値が負の値をとる場合には利用できない。

これまでクラスタリングが存在するかどうかの判定は、以下の手順に従って行われてきた。

- (1) それぞれのポリゴンごとに G 統計量の値 G_i を計算する。

- (2) 理論的に求められる G_i の期待値と分散の値を用いて、 G_i の p 値を計算する.
- (3) p 値をもとにして有意性を検討する.

しかしながらこの手順では、ポリゴンごとの G_i が正規分布すると仮定されていることに加えて多重比較が行われている、という問題が存在する. 多重比較については補正を行うことが可能であるが、正規性の仮定は必ずしも成立せず、成立しない場合効果的な補正法は存在しない.

こうした点を考慮して、ここでは「各ポリゴンの観測値をランダムに並べ替えて得られる仮想的なデータはどれも同様に確からしい」という仮説にもとづいてシミュレーションを行い、その結果を使って p 値を求めるという手順をとることにした.

ただし、その場合でも多重比較の問題は残るので補正

が必要である. 補正法として従来は Bonferroni の方法が広く使われていた. しかし、Bonferroni 法は第二種の過誤の率が高くなるという問題点をもつことが指摘されている. Bonferroni 法のもつこうした問題を解決するために Holm の方法、FDR 法 (False Discovery Rate) 等いくつかの方法が提案されている.

Brunsdon & Lex¹⁶⁾ によれば、Holm の方法はクラスターが存在するかどうかを検出したい場合に適しており、一方 FDR 法はクラスターが存在するとしてどこに存在するかを知りたい場合に適している. そこで、以下の分析では FDR 法を使って補正を行った.

さらに、式 (4.17) に現れる d の値を決める必要がある. ここではルーク型の隣接ポリゴンとクイーン型の隣接ポリゴンを選択できるように d の値を二つ選び、結果を比較した.

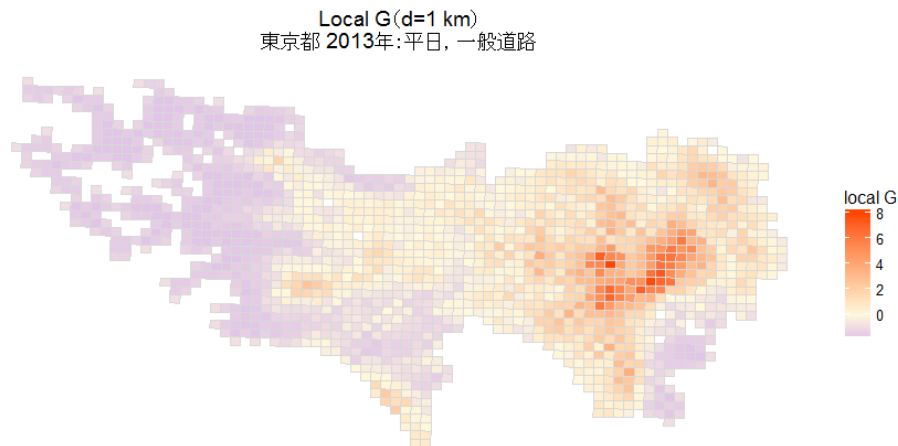


図 4.9: 東京都 Local G の値 (2013 年, 平日, ルーク型)

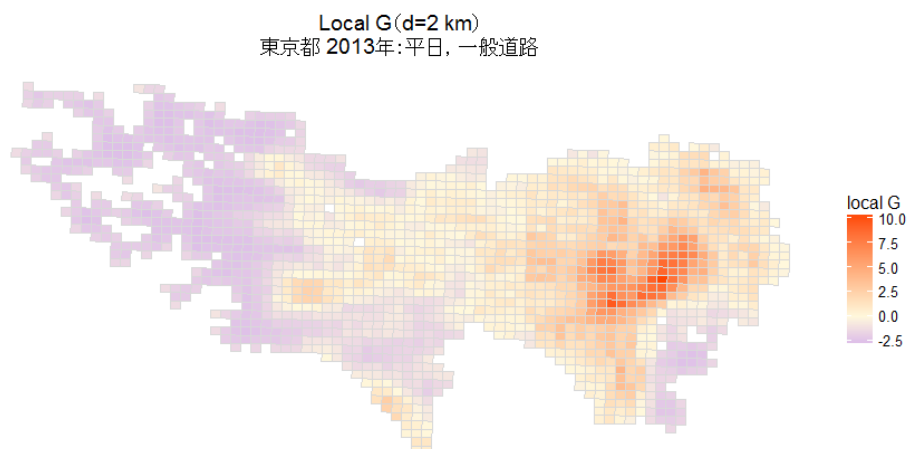


図 4.10: 東京都 Local G の値 (2013 年, 平日, クイーン型)

メッシュ毎の事故件数を使って local G の値を計算した結果をコロプレス図にして図 4.9 と図 4.10 に掲げた。この図で描いているのは「標準化された」local G の値である（すなわち、平均値が 0、分散は 1 になるように変換されている）。local G の値が大きいのは、事故件数が大きなメッシュとほぼ同じ地域であるが、事故件数よりもメッシュ間の差異が強調して表現されており、都心部から下町にかけての地域で local G の値が大きくなっていることがわかる。ルーク型とクイーン型では、クイーン型の方が local G の値の大きなメッシュが現れているが、全体的なパターンには大きな差がないことがわかる。

次に計算された local G の値について正規性の検定を行う。上で述べたように、それぞれのメッシュの事故件数を任意に並べ替えた状態がどれも等しい確率で生じると仮定し、2 万回のシミュレーションを行ってその結果をメッシュ毎に QQ プロット図に描いた。メッシュの数が多数あり、すべてのメッシュの QQ プロットを提示することはできないので、ランダムに 9 つのメッシュを抽出しそれらについての描画結果を図 4.11 と 4.12 に掲げた。

これらの図を見ると、ルーク型、クイーン型いずれの場合でも分布の右側では正規分布から外れており、裾の

長い分布になっていることがわかる。この結果は正規性の前提が成り立たないことを意味している。ここでの検定では分布の右側を使うことになるので、上述のシミュレーション結果を使って p 値を求めることにした。すなわち、2 万回のシミュレーションにおいて、実際の事故データから計算された local G の値が上から何パーセントの位置にあるのかを求め、さらに FDR 法を用いて補正を行い修正された p 値を求めた。

シミュレーション結果を用いて、メッシュ別に local G 統計量の統計的な検定を行い、その結果をコロプレス図 4.13 と 4.14 に掲げた。

これらの図によれば、都心部で p 値が有意になっており、事故の多いメッシュが都心部に集中していることがわかる。ルーク型の隣接ポリゴンを使った場合と比べ、クイーン型の隣接ポリゴンを使った場合の方が統計的に有意なメッシュの数が多くなっている。図 4.7 に示したメッシュごとの事故件数の値と比較すると、クイーン型の隣接ポリゴンを使った結果の方が事故の集中するクラスターを的確に表現できているように見える。東京都の西側の郊外部にも事故件数の多いメッシュが存在しているが、これらは有意なクラスターとして抽出されていない。

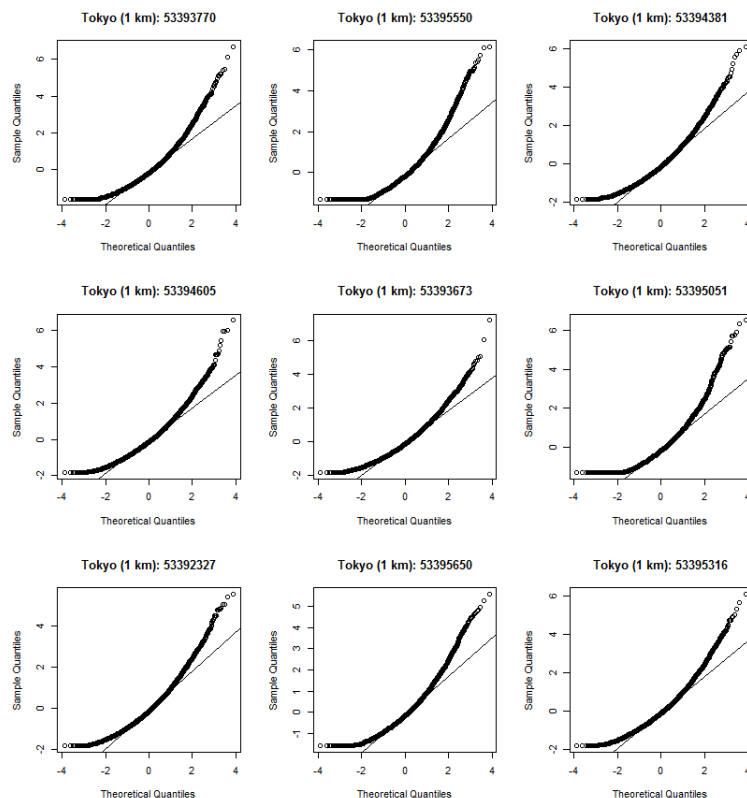


図 4.11: メッシュ別 Local G の値の QQ プロット（東京都，2013 年，平日，ルーク型）：それぞれの図の上の 8 桁の数字はメッシュの 3 次コードを表す

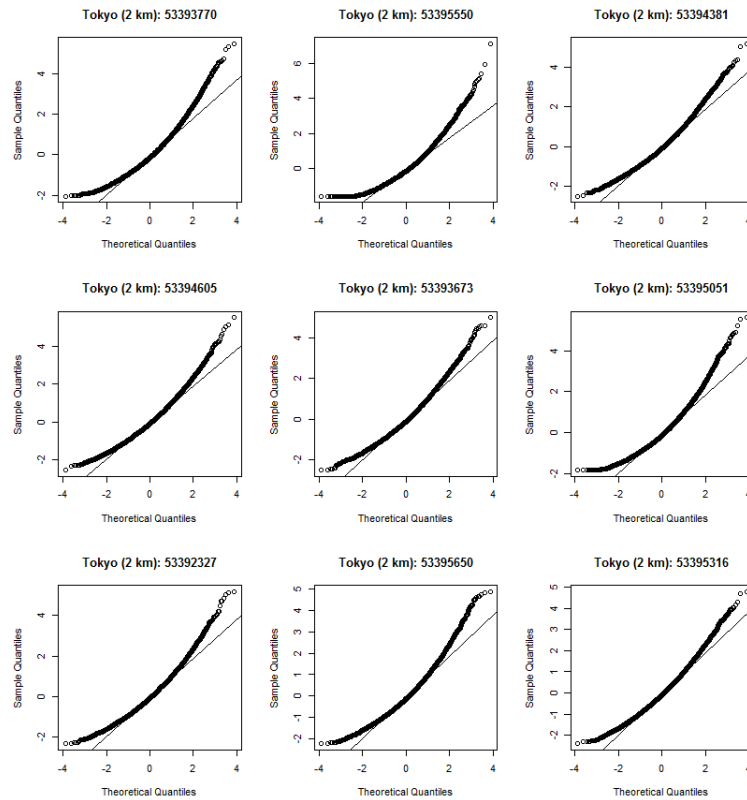


図 4.12: メッシュ別 Local G の値の QQ プロット (東京都, 2013 年, 平日, クイーン型): それぞれの図の上の 8 桁の数字はメッシュの 3 次コードを表す

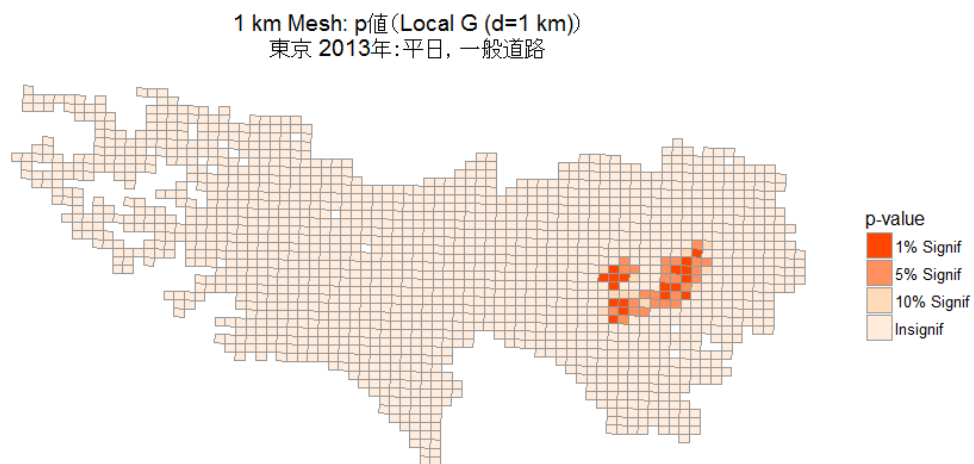


図 4.13: 東京都メッシュ毎の p 値 (2013 年, 平日, ルーク型)

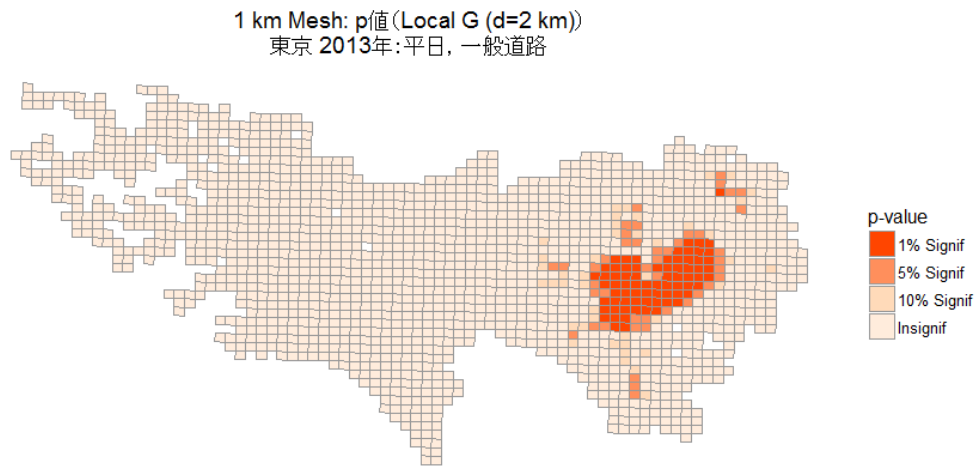


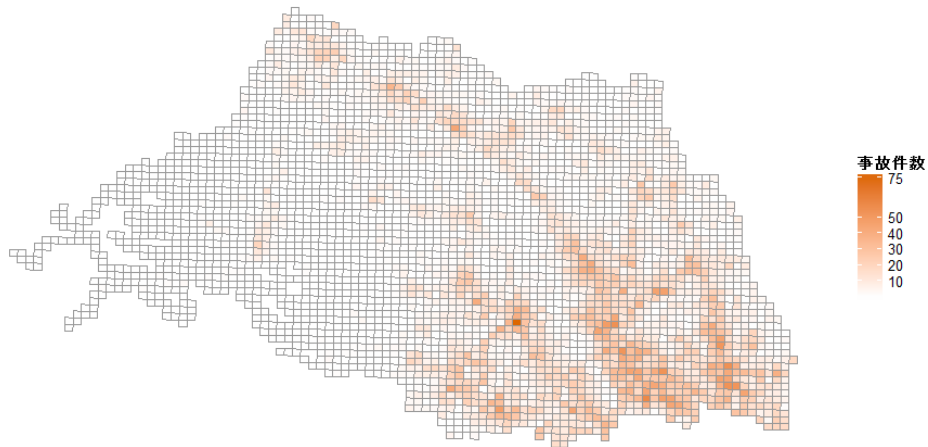
図 4.14: 東京都メッシュ毎の p 値 (2013 年, 平日, クイーン型)

以上で東京都の一般道路で発生した交通事故のクラスタリングの分析が終わった。以下では東京都以外の府県（埼玉県，大阪府および奈良県）について，同じく一般道路で発生した交通事故のクラスタリングを分析した結果のコロプレス図を示す。紙幅が限られているので，平日の事故についてクイーン型の隣接ポリゴンを使って計算した結果のみを図 4.15～4.17 に掲げた。

埼玉県のメッシュ毎の事故件数のコロプレス図を見る

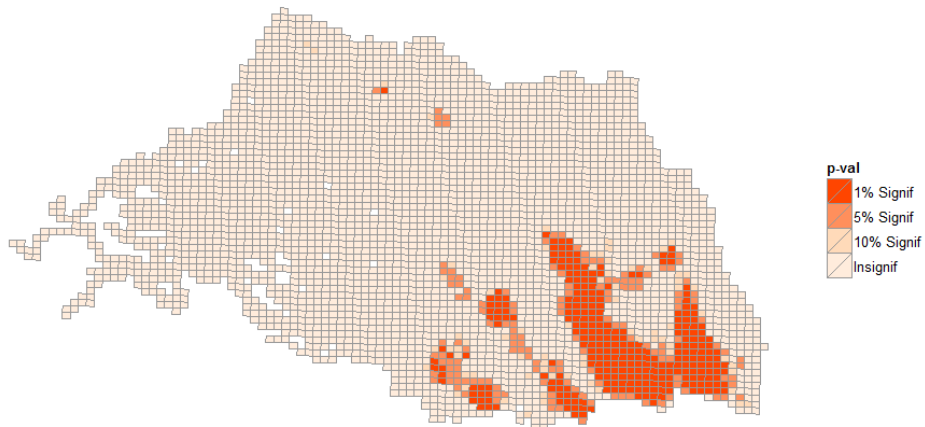
と（図 4.15a），東京寄りの地域と国道 17 号沿いで事故件数が多いことがわかる。一方，Local G を使った検定結果（図 4.15b）からは東京寄りの市（越谷，春日部，さいたま，川口，川越，和光，所沢，狭山）が有意なクラスターとして抽出されている。一方，国道 17 号沿いでは，深谷，熊谷がクラスターとして抽出されている。いずれにしても，埼玉県では交通事故は東京寄りの地域に集中していることが明らかである。

1 kmメッシュごとの交通事故件数
埼玉県 2013年:平日, 一般道路



(a) 埼玉県の事故件数 (2013 年, 一般道路)

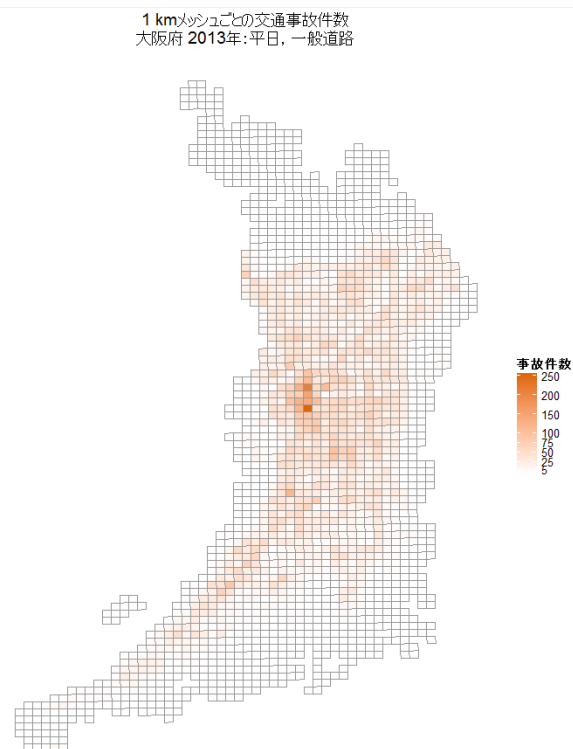
1 kmメッシュごとのp値(Local G(2 km))
埼玉県 2013年:平日, 一般道路



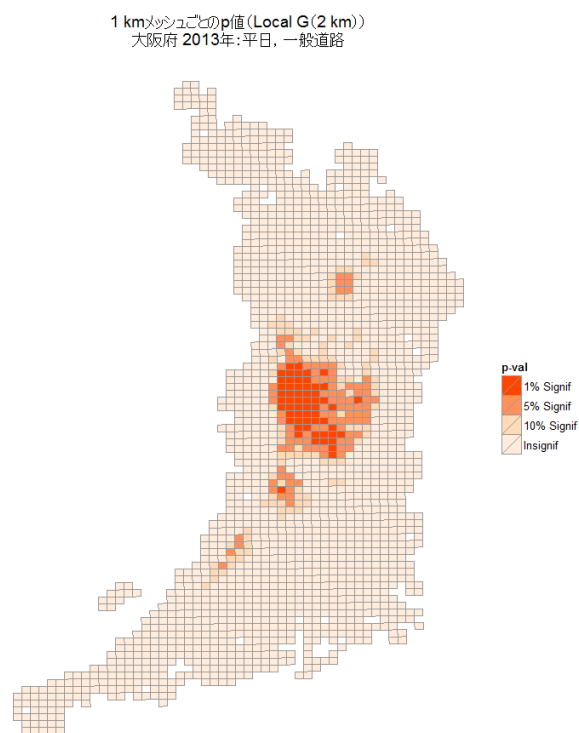
(b) 埼玉県の Local G (2013 年, 一般道路)

図 4.15: 埼玉県の交通事故のクラスタリング分析結果 (2013 年, 一般道路)

大阪府では, 交通事故は大阪市, 堺市を中心に広い地域で発生している (図 4.16a). しかし, 有意なクラスターとして抽出されたのは, 大阪市, 堺市, 岸和田市, そして高槻市にとどまる (図 4.16).



(a) 大阪府の事故件数 (2013 年, 一般道路)



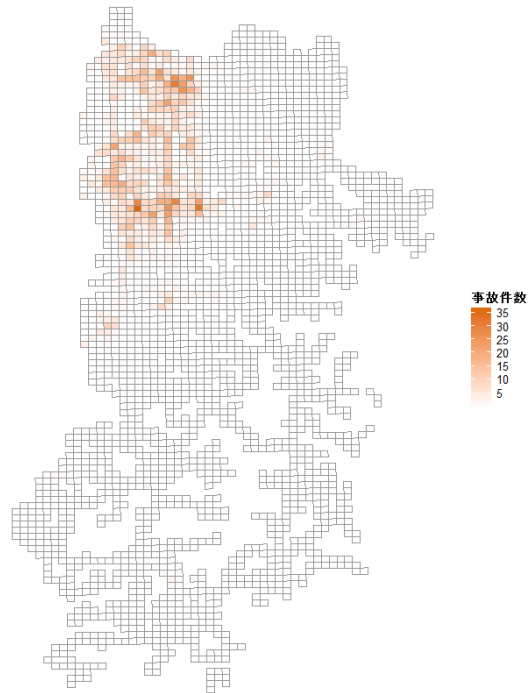
(b) 大阪府 Local G (2013 年, 一般道路)

図 4.16: 大阪府の交通事故のクラスタリング分析結果 (2013 年, 一般道路)

奈良県では, 交通事故が集中して発生しているのは北部の地域にとどまっており, 中部, 南部では事故の集中はみられない (図 4.17a). このため, 有意なクラスター

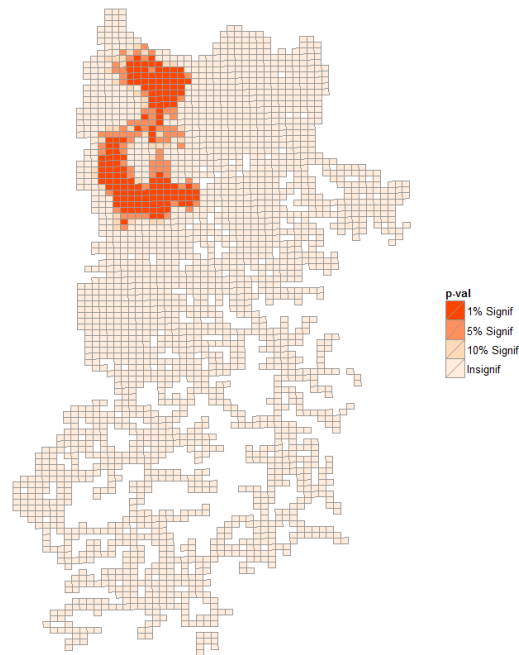
として抽出されているのは奈良, 大和郡山, 大和高田, 橿原といった北部の地域にとどまる (図 4.17b).

1 kmメッシュごとの交通事故件数
奈良県 2013年:平日, 一般道路



(a) 奈良県の事故件数 (2013 年, 一般道路)

1 kmメッシュごとのp値 (Local G (2 km))
奈良県 2013年:平日, 一般道路



(b) 奈良県 Local G (2013 年, 一般道路)

図 4.17: 奈良県の交通事故のクラスタリング分析結果 (2013 年, 一般道路)

これらの図からそれぞれの府県において交通事故が集中しているクラスターを判別することができる。こうしたクラスターにおいては、集中的な安全対策の必要性が高いといえよう。また Kuo *et al.*⁴⁴⁾ は、交通事故が多

発しているメッシュと犯罪が多発しているメッシュは重なる部分が多いことから、こうした地域で犯罪の取り締まりを重点的に行えば交通事故の抑止にも効果があると述べていることを付記しておく。

第 5 章

本研究のまとめ

2012 年から事故原票に交通事故発生位置の経度・緯度が付与されるようになり、「すべての道路と地域」を対象にして従来は不可能だった詳細な事故分析を行うことが可能になった。このことを受けて本研究開発では、GIS を援用して事故データの品質管理から分析までを実行できる総合的な事故分析システムを新たに開発することと地理空間情報を使って交通事故分析を行い事故の発生モデルを構築するとともに、事故の分布パターンを分析した。これらの研究成果は以下のようにまとめることができる。

5.1 総合的な事故分析システムの開発

GIS を援用して事故データの品質管理から分析までを実行できる総合的な事故分析システムを開発した。このシステムによって以下の課題の解決を図った。

- (1) 経度・緯度の正確さを確認し、正確ではないデータについては本来あるべき場所に修正すること
- (2) 確認・修正された事故発生場所をもとにして、事故を道路のリンク・交差点単位に集計すること
- (3) 集計された事故データをもとにした事故分析のためのシステムを構築すること

(1) については、事故データの経度・緯度の与える場所から一定の距離（原則として都市部では 50 m，地方部では 100 m；ただし、この数字は任意に設定可能）の範囲にある同種の道路（DRM リンク）を検索し、同種の道路が見つかった場合にはその場所を正しい事故発生場所とみなすようにした。この操作によって DRM リンクに付与されている様々な属性情報と事故とを結びつけることができる。

事故の集計 (2) については、DRM リンクをさらに細分した ITARDA 区間（平均長は約 300m）を単位として事故を集計するようにした。新設された道路を対象

に、ITARDA 区間を自動的に設定するツールも作成した。これまでは道路交通センサスの実施にあたってセンサス区間の見直しが行われており、この見直しにあわせて ITARDA 区間の見直しを行ってきた。しかし、2010 年の道路交通センサスから原則としてセンサス区間が固定されるようになったため、ITARDA 区間の全面的な見直しを行う必要がなくなった。（新設された道路については ITARDA 区間の設定が必要である。）このため、ITARDA 区間の設定変更にとまなう過去の事故の再集計が不要になり、ITARDA 区間を事故分析の基本的な単位として利用し続けることが可能になった。

事故分析用のシステム (3) は基本的に従前から使われていたものを踏襲しており、主たる機能として ITARDA 区間毎の事故率を計算し事故対策の優先順位を検討する機能を有している。ここに DRM ネットワーク上に事故を表示する機能を追加した。これによって、事故の発生場所を容易に確認することができる。

本システムを使った道路管理者用のシステムについては、2014 年度に 9 県に対して先行的に提供を行った。これらの県でのシステム利用において大きな問題がないことを確認した上で、2015 年度には提供対象を 32 都道府県に拡大した。2016 年度にはさらに提供対象を拡大し 36 都道府県に提供を行った。

事故分析では、最初に分析に必要と考えられる各種データを収集し、必要に応じて 3 次メッシュおよび 4 次メッシュ単位に集計した。このデータを使って事故発生要因と事故のクラスタリングについて分析を行った。

収集・整備した代表的なデータは次のとおりである。

- | | |
|------------|----------|
| ● 行政区域 | ● 小学校区 |
| ● バス路線・停留所 | ● 都市地域 |
| ● 用途地域 | ● DID 地区 |
| ● DRM | ● POIs |
| ● センサス交通量 | ● ゾーン 30 |

● 気象データ

ただし、道路の平面線形・縦断線形データは DRM に記載がなく、他のデータソースからも収集できなかった。交通量データについてはセンサス交通量を収集したが、道路交通センサスは 5 年に 1 回しか実施されておらず、また幹線道路のみで実施されているため、事故分析で使うのに十分なデータとはなっていない。

5.2 交通事故の分析

事故要因の分析では、一般道路で発生した事故の件数を都道府県別に 3 次 (1 km) メッシュで集計し、計数モデルを適用した。ただし、3 次メッシュで集計した場合でも件数が 0 となるメッシュの数が多く、特に地方部の県ではそうであった。この点を考慮して、計数モデルとして広く使われているポアソン回帰モデルに加えて、負の二項分布回帰モデル、さらにはゼロ強調モデルについても推定を行った。

推定したモデルについて比較を行ったところ、ポアソン回帰モデルとゼロ強調ポアソン回帰モデルでは過大分散となっているが、負の二項分布回帰モデルとゼロ強調負の二項分布回帰モデルでは過大分散となっていないことが判明した。さらに、負の二項分布回帰モデルとゼロ強調負の二項分布回帰モデルを比較したところでは、ゼロ強調負の二項分布回帰モデルの方が優れているという結論を得た。

この結果を踏まえて、33 都道府県を対象に 1 km メッシュごとに事故件数を集計し、ゼロ強調負の二項分布回帰モデルを推定した。この結果によると、店舗数、スポーツ施設数、交差点数、自動車による通勤・通学者数、事業所数、交差点数の係数が正の値で推定されている県が多く、これらの施設と交差点、そして自動車で通勤・通学する人が多いメッシュでは事故が発生しやすいという結果をえた。ただし気象関連の変数が事故に与える影響は様ではなく、たとえば年間降雨日数の係数は負の値として推定されている県が多いが、正の値で推定されている県もいくつか存在する。同じことは年間降雪日数の係数についても当てはまる。

さらに事故件数を対象にして Moran の I を計算したところ、「空間相関がないとはいえない」という結論が得られた。そこで、モデルにランダム効果項をとり入れ、それらの間に空間的な相関が存在する (CAR モデル) と仮定して MCMC 法でポアソン回帰モデルを予測した。この結果によれば、推定された事後分布の中央値を使って求めた各メッシュの事故件数の期待値と実際の事故件

数はかなりよく一致することが判明した。説明変数の係数の事後分布の中央値からは、都市地域では事故件数が多く、徒歩による通勤通学者の数が多く事業所数の数が多いメッシュでは事故件数が増えることが示されている (東京都の事故についての分析結果にもとづく)。ただし交差点数の係数は負になっており、交差点の多い都心部での事故が少ないことを意味している。これはおそらく東京都に固有の現象だと思われる。

最後に事故のクラスタリングについて分析結果した結果について述べる。東京都市圏の事故多発箇所と死亡事故、重傷事故の分布パターンに差があるかどうかを検討するために、多発箇所と死亡事故、重傷事故のパターンを DCLF 検定及び MAD 検定を使って相互に比較したところ、これらパターンのいずれの組合せでも有意な差が認められた。京阪神都市圏について同じ検定を行ったところ、やはりこれらパターンのいずれの組合せでも有意な差が認められた。このことは事故多発箇所について対策をとるだけでは、必ずしも死亡事故、重傷事故の低減につながらないことを意味している。

5.3 研究成果の普及

本研究の成果は、すでに道路管理者に提供する交通事故統合データベースの作成とそのシステムに反映されており、道路交通安全の実務において広く活用されている。また道路行政の観点から見ると、従来警察から提供を受けて地図上に落としていた事故発生場所を発生場所の経度・緯度を使ってほぼ自動的に取得することができるようになったことから、コストの削減にもつながっている。

さらに本研究において作成した事故多発箇所データは、車メーカー、カーナビメーカーに提供され、ウェブ上の交通安全マップ、カーナビ等で利用されている。

また、ITARDA のウェブサイトではメッシュ単位で集計した生活道路の事故件数をコロプレス図の形で提供している。

5.4 今後の課題

交通事故の分析を行うに当たっては、発生場所の情報を付与された交通事故データが必須である。発生場所の経度・緯度の正確さについては徐々に改善されてきているものの今後とも継続的にチェックを続けていく必要がある。

また分析に使われる国勢調査、経済センサス等の社会・経済データ、POI やデジタル道路地図 (DRM) 等の

空間情報データについては定期的な更新が必要であり、こうした更新業務を継続して実施していく。

交通事故分析では、これまで幹線道路での事故に重点が置かれていたが、近年は生活道路での事故にも関心が持たれるようになってきていることから、生活道路での事故と地域特性、地理空間特性との分析をさらに深化させていく考えである。とくに、交通事故発生場所の空間的なクラスタリングに注目した研究が行われるようになってきているので、こうした分野の研究も実施し効果的な交通安全対策の実施につながる成果を目指していきたい。

謝辞

本研究開発を実施する機会を与えていただいた新道路技術会議の委員の方々、ならびに国土交通省道路局と国土技術政策総合研究所の関係各位に感謝します。

また本研究プロジェクトの実施にあたって多大な協力をいただいた ITARDA および東京大学の関係各位にも深く感謝します。

本研究で利用した交通量シミュレーションの結果は株式会社アイ・トランスポート・ラボの成果です。貴重な

計算結果を快く貸与していただいた上に、細かな質問にも丁寧な回答をお寄せいただいた堀口良太様（同社代表取締役）に謝意を表します。

2010 年に実施された国勢調査および 2009 年に実施された経済センサスの成果を地域メッシュ毎に集計した結果については、東京大学空間情報科学研究センター（CSIS）の共同研究利用システム（JoRAS）のデータを利用させていただきました。感謝します。

また本研究開発では、ArcGIS, Excel, Access, Word の商用ソフトに加えて以下のフリーソフトを利用させていただきました。これらのソフトの製作・改良に携わった方々に深く感謝します。

- 統計ソフト R: version 3.2.1 and 3.3.0, by R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>, 2015.
- GIS ソフト QGIS: version 2.8.2 Wien
- 文書作成・組版: T_EX Live 2015

さらに背景地図として Open Street Map を利用させていただきました。OSM の関係者の皆様に感謝します。

参考文献

- [1] 警察庁：交通事故原票作成の手引き.
- [2] 日本デジタル道路地図協会：全国デジタル道路地図データベース標準，第 3.9 版，2015 年 4 月.
- [3] <http://drm.jp/database/structure.html>，2016 年 6 月 8 日アクセス.
- [4] 交通事故総合分析センター：平成 27 年度イタルダ区間別データ利用マニュアル，2016 年 3 月.
- [5] 交通事故総合分析センター：平成 27 年度事故別データ利用マニュアル，2016 年 3 月.
- [6] (財) 統計情報研究開発センター (Sinfonica)：データファイル説明書 平成 22 年国勢調査に関する地域メッシュ統計 (世界測地系) (その 1 人口，世帯，産業).
- [7] (財) 統計情報研究開発センター (Sinfonica)：データファイル説明書 平成 22 年国勢調査に関する地域メッシュ統計 (世界測地系) (その 2 職業，従業地・通学地，人口移動).
- [8] (財) 統計情報研究開発センター (Sinfonica)：データファイル説明書 平成 21 年経済センサス基礎調査に関する地域メッシュ統計 (世界測地系).
- [9] 警察庁：「ゾーン 30」の概要，<https://www.npa.go.jp/koutsuu/kisei/zone30.pdf>
- [10] 堀口良太，片倉正彦，桑原雅夫：都市街路網の交通流シミュレーター AVENUE - の開発，第 13 回交通工学研究発表会論文集，pp.33–36，1993.
- [11] Horiguchi R, Katakura M, Akahane H and Kuwahara M: A Development of a Traffic Simulator for Urban Road Networks: AVENUE, *Proceedings of Vehicle Navigation & Information Systems*, Yokohama, pp.245–250, 1994.
- [12] <http://www.jma.go.jp/jp/amedas/>
- [13] http://metds.co.jp/file/data/StnMap_J.pdf
- [14] Hofstra N, Haylock M, New M, Jones P and Christoph F: Comparison of six methods for the interpolation of daily, European climate data, *J of Geophysical Research*, vol 113, D21110, 19 pages, The American Geophysical Union, 2008.
- [15] Bivand R S, Pebesma E and Gómez-Rubio V: *Applied spatial data analysis with R*, 2nd ed., Springer, 2013.
- [16] Brunsdon C and Lex C: *An introduction to R for spatial analysis & Mapping*, Sage Publications Inc, 2015.
- [17] Dobson, A J: *An introduction to generalized linear models*, 2nd edition, Chapman & Hall/CRC texts in statistical science, 2002.
- [18] Faraway, J J: *Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models*, Chapman & Hall/CRC texts in statistical science, 2006.
- [19] 南 美穂子，Lennert-Cody, C E: ゼロの多いデータの解析：負の 2 項回帰モデルによる傾向の過大推定，統計数理，第 61 巻，2 号，pp. 271–287，2013.
- [20] 瀬谷 創，堤 盛人：空間統計学－自然科学から人文・社会科学まで，朝倉書店，2014 年.
- [21] Besag J, York J and Mollie A: Bayesian image restoration with two applications in spatial statistics, *Ann Inst Stat Math*, 43, 1–59, 1991.
- [22] Lee D: A comparison of conditional autoregressive models used in Bayesian disease mapping, *Spatial and Spatio-Temporal Epidemiology*, 2, 79–89, 2011.
- [23] Brunsdon C, Fotheringham S and Charlton M E: Geographically weighted regression: A method for exploring spatial nonstationarity, *Geographical Analysis*, 28(4), 281–298, 1996.

- [24] Crawley, M J: *Statistics: an introduction using R*, Wiley, 2005.
- [25] 久保 拓弥: データ解析のための統計モデリング入門—一般化線形モデル・階層ベイズモデル・MCMC, 岩波書店, 2012.
- [26] Wilson P: The misuse of the Vuong test for non-nested models to test for zero-inflation, *Economics Letters*, 127(51–53), Elsevier, 2015.
- [27] Paweenuch S, Yamada H, Teerayut H and Shibasaki R: Traffic accidents risk analysis based on road and land use factors using GLMs and zero-inflated models, *Proceedings of CUPUM 2015*, paper #320, 26 pages, 2015.
- [28] Getis A and Ord J K: The analysis of spatial association by use of distance statistics, *Geographical Analysis*, 24(3), 189–206, 1992.
- [29] Ord J K and Getis A: Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis*, 27(4), 286–306, 1995.
- [30] Lee D: CARBayes version 4.4: An R package for spatial areal unit modelling with conditional autoregressive priors, Vignette for CARBayes package, <https://cran.r-project.org/web/packages/CARBayes/vignettes/CRBayesvignette.pdf>, 2015.
- [31] Shankar V, Milton J and Mannering F: Modeling accident frequency as zero-altered probability processes: an empirical inquiry, *Accident Analysis and Prevention*, 29(6), 829–837, Pergamon, 1997.
- [32] Shankar V N, Ulfarsson G F, Pendyala R M and Nebergall M B: Modeling crashes involving pedestrians and motorized traffic, *Safety Science*, 41, 627–640, Pergamon, 2003.
- [33] Chen Q, Song X, Yamada H and Shibasaki R: Learning deep representation from big and heterogeneous data for traffic accident inference, *Proceedings of AAAI 2016*, 7 pages, Presented at the 13th AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA, 2016.
- [34] (公財) 交通事故総合分析センター: 交通事故統計年報, 平成 22 年版, 23 年版, 24 年版, 25 年版, 26 年版.
- [35] Baddeley A, Diggle P J, Hardegen, A H, Lawrence T, Milne R K and Nair G: On tests of spatial pattern based on simulation envelopes, *Ecological Monographs*, 84(3), 477–489, The Ecological Society of America, 2014.
- [36] Ripley B D: Modelling spatial patterns with discussion, *Journal of the Royal Statistical Society B*, 39, 172–212, 1977.
- [37] Loosmore N B and Ford E D: Statistical inference using the G or K point pattern spatial statistics, *Ecology*, 87(8), 1925–1931, The Ecological Society of America, 2006.
- [38] Xie Z and Yan J: Kernel density estimation of traffic accidents in network space, *Computers, Environment, and Urban Systems*, 35(5), 396–406, 2008.
- [39] Okabe and Sugihara: A kernel density estimation method for networks, its computational method, and a GIS-based tool, Discussion Paper 89, CSIS, The University of Tokyo, 2008.
- [40] SANET: Spatial analysis along networks — User guide / manual for SANET standalone beta, October 2015, Downloadable from sanet.csis.u-tokyo.ac.jp/index_jp.html
- [41] Xie Z and Yan J: Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach, *J Transport Geography*, 31(64–71), Elsevier, 2013.
- [42] Bíl M, Andrášik R and Janoška Z: Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation, *Accident Analysis and Prevention*, 55(265–273), Elsevier, 2013.
- [43] Anselin L: Local indicators of spatial association – LISA, *Geographical Analysis*, 27(93–115), 1995.
- [44] Kuo P-F, Zeng X and Lord D: Guidelines for choosing hot-spot analysis tools based on data characteristics, network restrictions, and time distributions, Presented at the 91st Annual Meeting of the Transportation Research Board, Nov 2011.