

交通事故の空間分布パターンの分析

山 田 晴 利

公益財団法人 交通事故総合分析センター

常務理事 研究部長

1 はじめに

交通事故の発生場所の経度・緯度データを用いて、交通事故の分布パターンの分析を行った。分析対象としたのは、次のふたつのテーマである。

- 交通事故の多発箇所（ホット・スポット）の分布パターンと重大事故（死亡・重傷事故）のパターンの間に差があるかどうか。
- 交通事故が集中している場所があるかどうか、あるとすればそれはどこか。

最初の分布パターンの分析では、交通事故総合分析センター（以下、ITARDA と記す）で設定した ITARDA 区間（詳細については後述する）をもとにして抽出した事故多発箇所と死亡・重傷事故の発生箇所の分布パターンが同じと見なせるかどうかを統計的に分析した。

二番目の交通事故の集中している場所のクラスタリング分析では、1 km メッシュごとに集計した交通事故件数を用いて、事故が集中的に発生しているメッシュのかたまり（クラスター）を見出すことを試みた。

2 交通事故の分布パターンの分析

死亡事故、重傷事故および多発箇所の空間的な分布パターンの間に差があるかどうかを統計的に検証することを目的として、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は、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。これは、事故多発箇所が一般幹線道路で発生した事故をもとにして選定されているためである。

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。

事故多発箇所の分析には「イタルダ区間」を用いた。イタルダ区間とは、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。

ここでは、年間 4 件以上の事故が発生している区間（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していない。多発箇所の数は全国で 16,856 箇所となった。

多発箇所の場所を表す代表点として、交差点についてはデジタル道路地図（DRM）の交差点の代表ノードを用いた。

単路については、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ている。

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。分析の対象とした東京都市圏、京阪神都市圏においては、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。

最初に、東京都市圏と京阪神都市圏の交通事故を対象にしてカーネル密度推定（KDE: Kernel Density Estimation）を行った結果を図 1 と図 2 に示した。

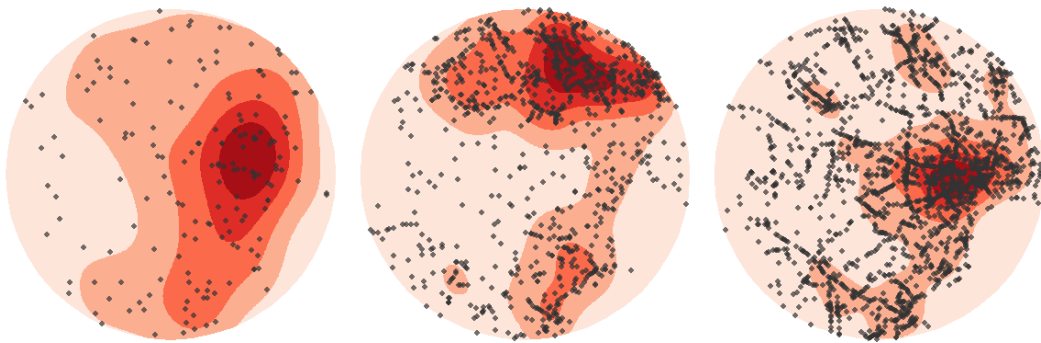


図 1 東京都市圏でのカーネル密度推定結果（2013 年事故）：左が死亡事故，中が重傷事故，右が多発箇所

これらカーネル密度推定の結果から、次の結論を導くことができる。

- 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。
- 一方、重傷事故は東京都心では密度が低く、むしろ埼玉県の南部で密度が高くなっている。これは東京都の重傷事故件数が周辺の県と比べ著しく低いためである。しかし、東京都の重傷事故件数が他県と比べ少ない原因は現時点では不明である。
- 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

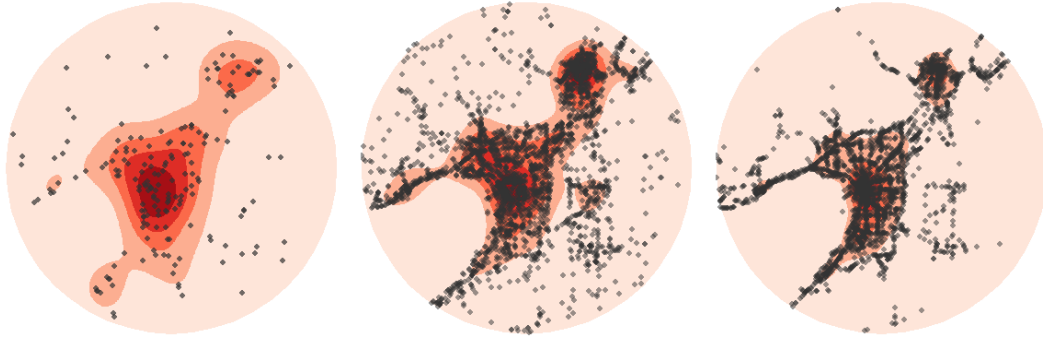


図 2 京阪神都市圏でのカーネル密度推定結果（2013 年事故）：左が死亡事故，中が重傷事故，右が多発箇所

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，大阪市と京都市の密度が高くなっている。

次にマーク付き点過程の分析を行う。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 つの分析を行った結果を示す。分析には統計ソフト R とクロス 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 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，

$$\text{Cross } K_{ij}(r) = \lambda_j E [K_{ij}(r)] \quad (1)$$

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 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。

$$\text{Cross } L_{ij}(r) = \sqrt{\frac{\text{Cross } K_{ij}(r)}{\pi}}. \quad (2)$$

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。

しかしながら，Brunsdon and Comber²⁾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している。

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，Repley³⁾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。

検定の対象となる帰無仮説 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。

以上の記述を踏まえて、東京都市圏と京阪神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中に統計的に有意な差があるかどうかを検定した。マーク付き点過程の分析を行った結果から次の結論が導かれる。

- 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（いずれの場合も $p < 0.01$ ）。
 - 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。
- 京阪神都市圏
 - 京阪神都市圏においても、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（いずれの場合も $p < 0.01$ ）。
 - 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。

3 メッシュ単位の交通事故多発クラスター分析

本節では、1 km メッシュ単위에集計した人身事故件数を対象にして事故多発メッシュのクラスターリングを分析した結果を示す。ここで「クラスターリング」といっているのは、事故件数の多い（または少ない）メッシュが地理空間的に集まっていることを意味する。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、前節と同じく 2013 年に発生した人身交通事故のうち一般幹線道路で発生した事故である。

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、兵庫県を対象にクラスターリングの分析を行った結果を R のコードとともに示す。このコードは Brunson and Comber²⁾ と Chang⁴⁾ をもとにしている。なお、以下に示すコード、出力結果そしてコメントは、RStudio に実装されている `rmarkdown` を利用してこれらを $\text{X}_{\text{L}}\text{A}_{\text{T}}\text{E}_{\text{X}}$ に変換した結果をもとにして、作成した。

```
# 必要なパッケージを読み込む
```

```
library(GISTools)
```

```
library(spdep)
```

```
library(rgdal)
```

```
library(maptools)
```

```
library(ggplot2)
library(grid)
```

```
# 1 km メッシュファイルを読み込む
proj4UTM53N <-
"+proj=utm +zone=53 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no_defs"
CRS_UTM53N <- CRS(proj4UTM53N)
Hyogo_1km_UTM53N <- readShapePoly("Hyogo_1kmMesh_UTM53N.shp",
                                proj4string=CRS_UTM53N)
```

それぞれのポリゴンの事故件数、道路延長などの属性を格納してある csv ファイルを読み込む。事故件数は一般道路で発生した件数を平日、休日別に集計した値となっている。また、道路地図から読み込んだメッシュ毎の道路延長が国道 (croad_length), 主要地方道 (local1_length), 一般都道府県道 (general_length), 細道路 (thin_length) 別に記録されている。

ポリゴンの属性データをポリゴンのシェープファイルの属性として付与する。

```
# csv ファイルを読み込む
Hyogo_acc2013_G <- read.csv("Hyogo_final.csv", header=T)
# 属性データの付与
Hyogo_1km_UTM53N@data <- merge(x=Hyogo_1km_UTM53N@data, y=Hyogo_acc2013_G,
                              by="CODE", all.x=T)
```

メッシュ毎に一般道路の総延長の値を計算し、総延長がゼロのメッシュは分析対象から除外する。一般道路の総延長がゼロのメッシュは 1658 ある。

```
# メッシュ内の道路延長を計算する
attach(Hyogo_1km_UTM53N@data)
Hyogo_1km_UTM53N@data$GRoad_Total <- croad_length + local1_length +
                                general_length + thin_length
detach(Hyogo_1km_UTM53N@data)
# GRoad_Total = 0 のメッシュを検索する
r0_index <- Hyogo_1km_UTM53N@data$GRoad_Total == 0
sum(r0_index)
```

```
## [1] 1658
```

```
# 一般道路の総延長が 0 のメッシュを除外する
Hyogo_1km_UTM53N_R <- Hyogo_1km_UTM53N[!r0_index,]
length(Hyogo_1km_UTM53N_R@data$CODE)
```

```
## [1] 6807
```

```
# シェープファイルを書き出す
```

```
writePolyShape(Hyogo_1km_UTM53N_R, "Hyogo_1kmM_UTM53N_acc_withR")
```

すぐ上で書き出したシェープファイルをもう一度読み込んで、`ggplot2` で作図するのに用いる。読み込んだシェープファイルの属性名が一部短縮されているので注意する必要がある。これはシェープファイルの属性名が 10 文字に制限されているためである。

```
# シェープファイルを読み込む
```

```
# いくつかの変数名が短く変換されているので注意
```

```
Pref_1kmM_acc <- readShapePoly("Hyogo_1kmM_UTM53N_acc_withR.shp",  
                               proj4string=CRS_UTM53N)
```

```
# id 番号を付与する
```

```
Pref_1kmM_acc$id <- rownames(Pref_1kmM_acc@data)
```

`ggplot2` でプロットするデータはデータフレームでないといけない。しかし、シェープファイルはデータフレームではないので、`ggplot2` パッケージの `fortify()` 関数を使ってシェープファイルをデータフレームに変換する。ただし、`fortify()` 関数はポリゴンデータを変換するだけで属性データは変換されないで、ポリゴンの属性データを結合する。

```
# データフレームに変換する
```

```
Pref_1kmM_pre <- fortify(Pref_1kmM_acc)
```

```
## Regions defined for each Polygons
```

```
# 属性データを付与する
```

```
Pref_1kmM_attr <- merge(x=Pref_1kmM_pre, y=Pref_1kmM_acc@data, by="id",  
                       all.x=T)
```

```
# 不要となったデータフレームを消去する
```

```
rm(Pref_1kmM_pre)
```

これで必要なデータが準備できたので、平日・休日別に事故件数のコロプレス図を描く。

ポリゴン（メッシュ）の描画には、`ggplot` パッケージの `geom_polygon()` 関数を使い、引数 `color` に “grey60” を指定した。 `group=CODE` と指定することで、同じ番号をもつデータをメッシュ毎にひとつにまとめている（`CODE` はメッシュ番号である）。 `fill` には色分けに使う属性情報を与える（ここでは事故件数）。

ポリゴンの色分けは `scale_fill_continuous()` で与えている。ここでは連続値を `break` で与える値で区切って色分けする指定をした。 `low` と `high` の二つの引数で色を指定し、事故件数が高いほど色が濃くなるようにした。 `name` で与えた文字列が凡例に表示される。 `coord_equal()` 関数は縦横比を固定する関数である。

まず平日の事故件数のコロプレス図を描く。結果は図 3 に示した。この図によると、神戸市、芦屋市、尼崎市に線状に事故の多いメッシュが集中している。さらに姫路市、明石市に事故の多いメッシュが認められる。それ以外の地域では事故の件数はきわめて少ない。

```
# 軸には何も表示しないようにするための関数を定義する
# R グラフィックスブックを参考にした (W Chang 著, オライリー・ジャパン)
theme_clean <- function(base_size=12) {
  theme_grey(base_size) %+replace%
  theme(axis.title      = element_blank(),
        axis.text       = element_blank(),
        panel.background = element_blank(),
        panel.grid       = element_blank(),
        axis.ticks       = element_blank(),
        panel.margin     = unit(0, "lines"),
        plot.margin      = unit(c(0, 0, 0, 0), "lines"),
        complete         = TRUE
  )
}
# コロプレス図 (平日の事故件数)
ggplot(Pref_1kmM_attr) +
  geom_polygon(aes(long, lat, group=CODE, fill=NUM_ACC_WO),
              color="grey60") +
  scale_fill_continuous(low="#fefefe", high="#dd6600",
                        breaks=c(10, 20, 30, 40, 50, 75, 100), name=" 事故件数") +
  labs(title="1 km メッシュごとの交通事故件数\n兵庫県 2013 年：平日，一般道路") +
  coord_equal() + theme_clean()
```

つぎに休日の事故件数のコロプレス図を描く。結果は図 4 に示した。事故件数が多いのは、平日の場合と同じ地域であるが、事故件数そのものは平日と比べ少なくなっている。

Brunsdon and Comber のやり方²⁾に従って、メッシュの間の事故件数の空間的な分布を分析する。事故が特定のメッシュ (の集合) に集中しているかどうかを検証することが目的である。

こうした分析に使われるのは、ローカル・モラン I, Ord の G 統計量であるが、ここでは Ord の G 統計量を使った。これは次式で定義される。

$$G_i = \frac{\sum_{j=1}^n z_j v_{ij}(d)}{\sum_{j=1}^n z_j}. \quad (3)$$

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。

1 kmメッシュごとの交通事故件数
兵庫県 2013年:平日, 一般道路

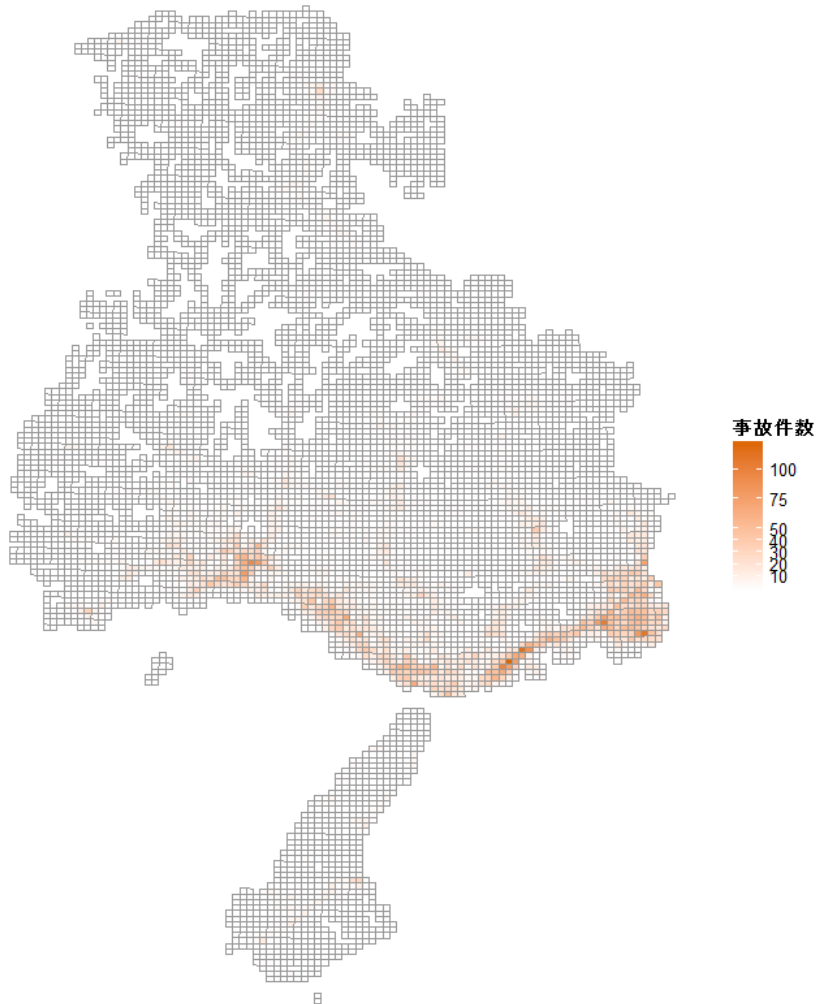


図3 兵庫県における一般道路の人身事故件数(2013年, 平日, 1 kmメッシュ)

以下では「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, その結果から p 値を求めるといった手順をとった.

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. ここでは, FDR法(False Discovery Rate)を使って補正した.

平日の事故についての分析に使ったRのコードを以下に掲げる. 計算結果については, 平日と休日双方の結果を提示した.

まず2 km以内にあるメッシュを近傍のポリゴンと見なし(すなわち, 式(3)で $d = 2$ km), この距離をもとにして近傍メッシュのリストを作る.

1 kmメッシュごとの交通事故件数
兵庫県 2013年:休日, 一般道路

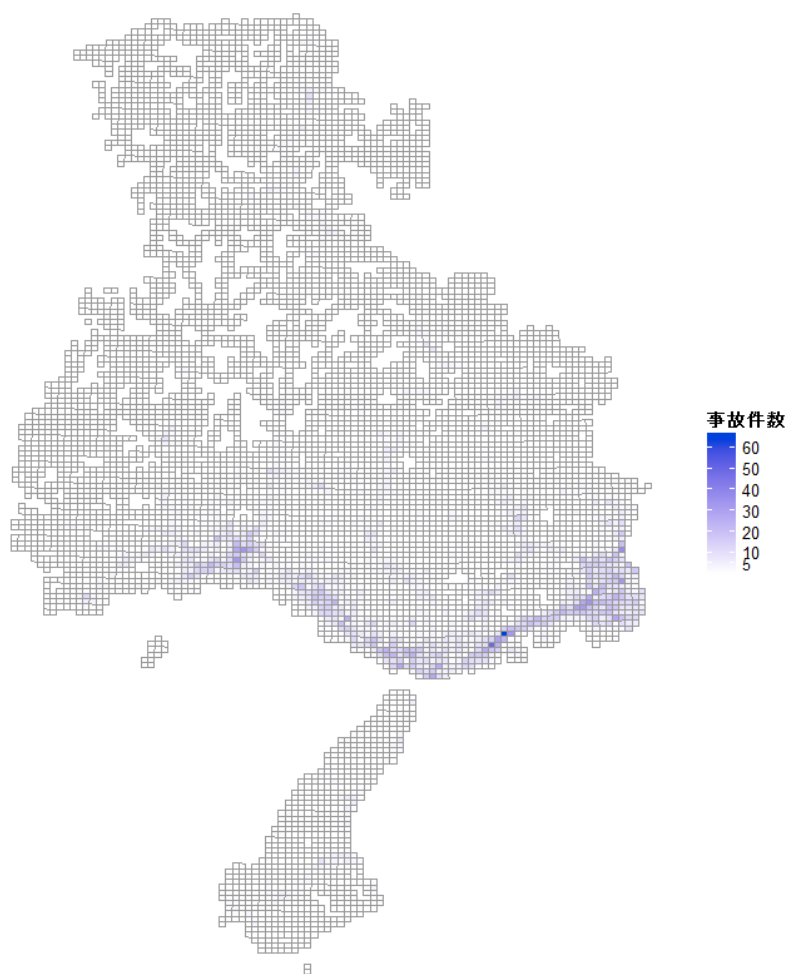


図4 兵庫県における一般道路の人身事故件数(2013年, 休日, 1 kmメッシュ)

```
# 近接するメッシュを求める  
# 距離を 2 km に設定する (距離の単位は m なので 2000 を指定する)  
# style="B" はバイナリの意味  
near.dist <- 2000  
near.poly <- dnearneigh(coordinates(Pref_1kmM_acc), 0, near.dist)  
near.poly.lw <- nb2listw(near.poly, style="B")
```

メッシュ毎の事故件数を使って local G の値を計算する.

```
acc.mesh.work <- Pref_1kmM_acc@data$NUM_ACC_WO  
accm.localG.work <- localG(acc.mesh.work, near.poly.lw)
```

```
summary(accm.localG.work)
```

```
##      Min.   1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.00100 -0.87220 -0.74950  0.08101 -0.40710 19.12000
```

計算された local G の値をコロプレス図に表す (図 5 と図 6)。ここで描いているのは「標準化された」 local G の値である (すなわち、平均が 0、分散は 1 になっている)。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されている。

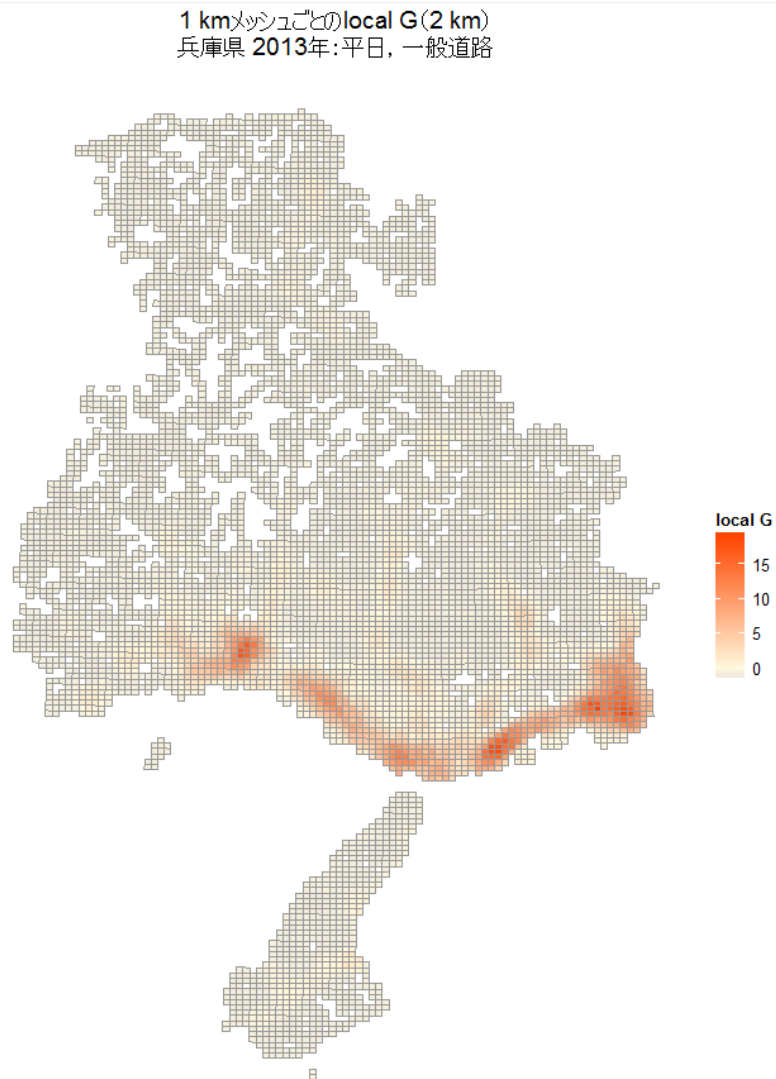


図 5 兵庫県 Local G の値 (2013 年, 平日, 1 km メッシュ)

1 kmメッシュごとのlocal G(2 km)
兵庫県 2013年:休日, 一般道路

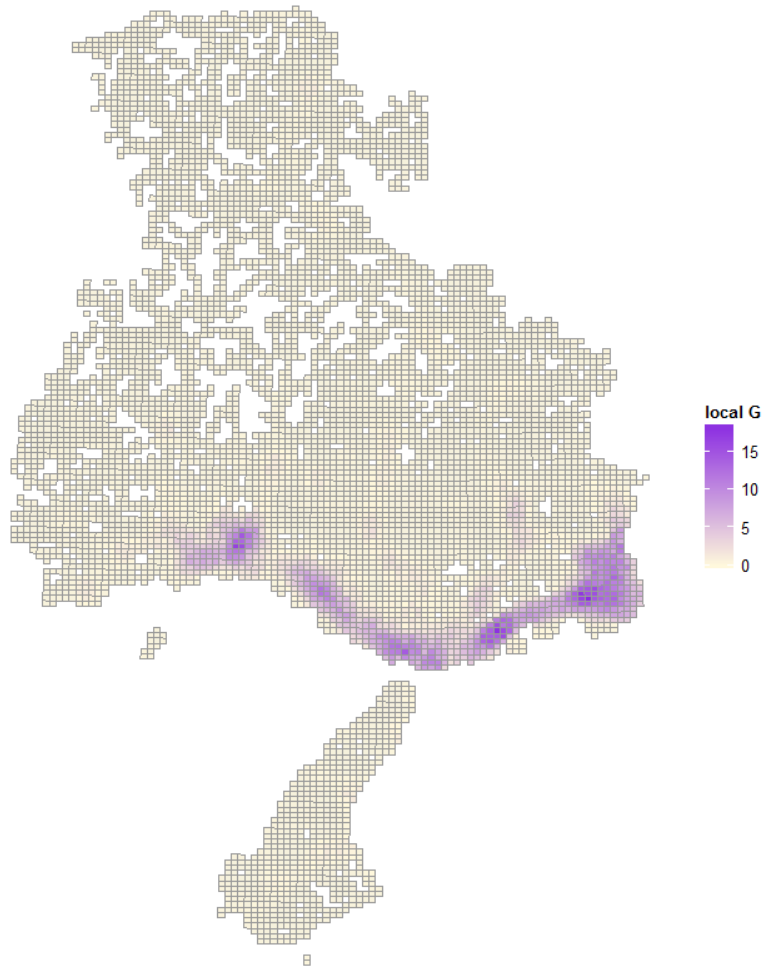


図 6 兵庫県 Local G の値 (2013 年, 休日, 1 km メッシュ)

```
# ggplot2を使って描画する
# ggplot2の描画に使うデータフレーム Pref_1kmM_attr に上で計算した
# accm.localG.work の値を左結合する
# この結合のキーとなる変数 id は, ファクターに変換されないよう I() を使って与える
# acc.localG.work は local G クラスなので, 数値に変換する
df.accm.work <- data.frame(id=I(rownames(Pref_1kmM_acc@data)),
                           localG_work=as.numeric(accm.localG.work))
Pref_1kmM_attr <- merge(x=Pref_1kmM_attr, y=df.accm.work, by="id", all.x=T)
# 平日の事故について, 標準化された Local G のコロプレス図を描く
# 中間の値 (ここではゼロ) からどのように値が変化しているかを示したいので,
```

```
# scale_fill_gradient2() 関数を使う
ggplot(Pref_1kmM_attr) +
  geom_polygon(aes(long, lat, group=CODE, fill=localG_work),
              color="grey60") +
  scale_fill_gradient2(low="#0000ff", mid="#fff8dc", high="#ff4500",
                      midpoint=0, guide="colourbar", name="local G") +
  labs(title="1 km メッシュごとの local G\n兵庫県 2013 年: 平日, 一般道路") +
  coord_equal() + theme_clean()
```

次に計算された local G の値について、正規性の検定を行う。前にも述べたように、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。ただし、メッシュの数が多数あり、すべてのメッシュについての QQ プロットを載せることはできないので、図は省略する。平日、休日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する。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もとにして 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うことで修正された 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に表した (図 7 と 図 8)。

```
MC.localG.pvals <- (colSums(sweep(simulate.G, 2, accm.localG.work, ">=")) +
                  1) / (nrow(simulate.G) + 1)
adjust.p <- p.adjust(MC.localG.pvals, method="fdr")
# 分位数を与える。分位数で p 値をカテゴリ分けする。
break.points <- c(0, 0.001, 0.01, 0.05, 0.10, 1)
p.category <- cut(adjust.p, break.points,
                 labels=c("0.1% Signif", "1% Signif", "5% Signif", "10% Signif", "Insign"),
                 include.lowest=T)
# 各カテゴリのカラーパレット (離散的) を作る
cat.palette <- colorRampPalette(c("#ff4500", "#ffdab9", "#ffffff"))(5)
# 各メッシュのカテゴリをデータフレームに左結合する
df.accm.pval <- data.frame(id=I(rownames(Pref_1kmM_acc@data)),
                          pval_work=p.category)
Pref_1kmM_attr <- merge(x=Pref_1kmM_attr, y=df.accm.pval, by="id", all.x=T)
# コロプレス図を描く
ggplot(Pref_1kmM_attr) +
```

```
geom_polygon(aes(long, lat, group=CODE, fill=pval_work), color="grey60") +
scale_fill_manual(values=cat.palette, name="p-val") +
labs(title="1 km メッシュ毎の p 値 (Local G) \n兵庫県 2013 年: 平日, 一般道路") +
coord_equal() + theme_clean()
```

これらの図によれば、平日、休日いずれにおいても山陽道沿いの都市部で p 値が有意になっており、事故の多いメッシュがここに集中していることがわかる。これ以外の地域では、有意なクラスターは存在しない。また平日と休日を比べると、休日では有意なメッシュが少なくなっていることがわかる。

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

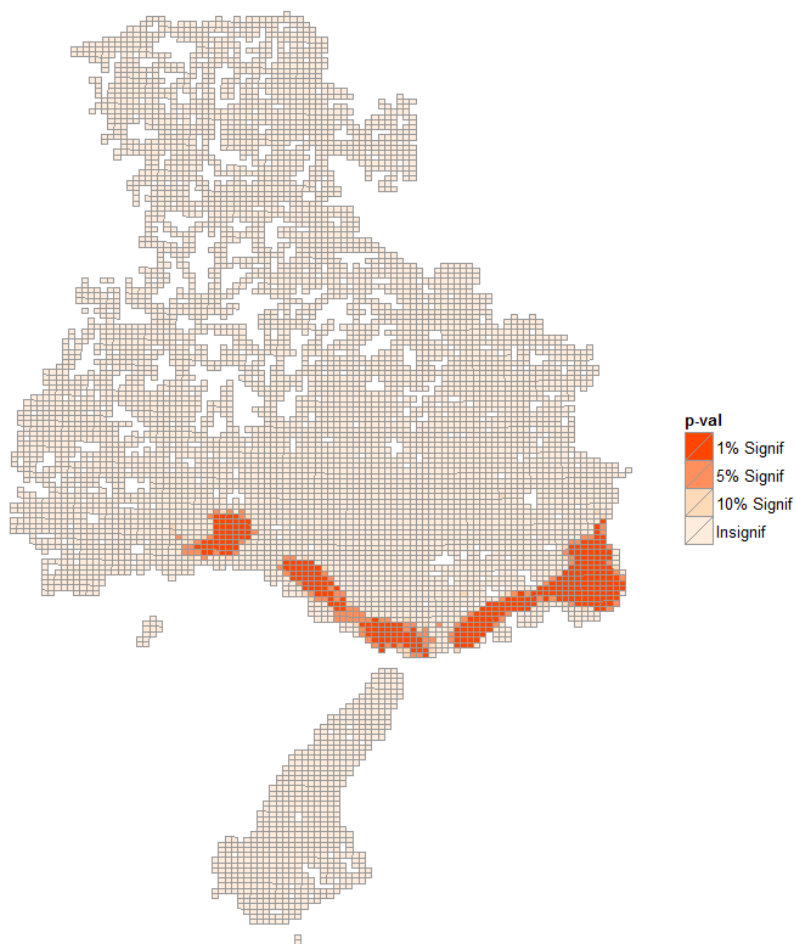


図 7 兵庫県メッシュ毎の p 値 (2013 年, 平日, 1 km メッシュ)

1 kmメッシュごとのp値(Local G(2 km))
 兵庫県 2013年:休日, 一般道路

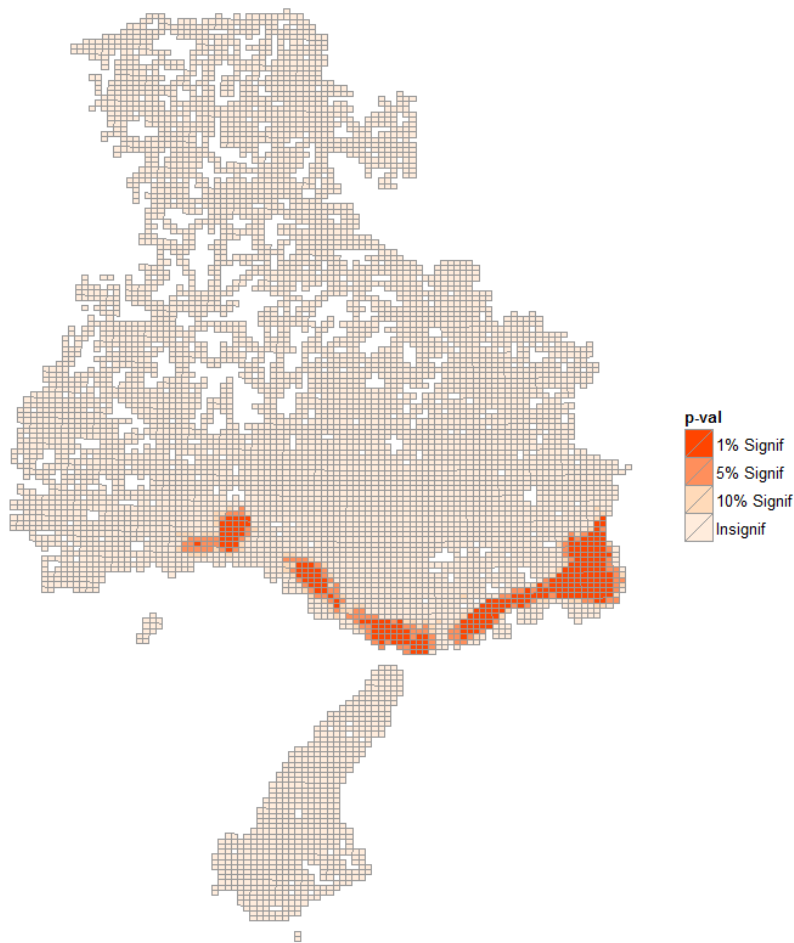
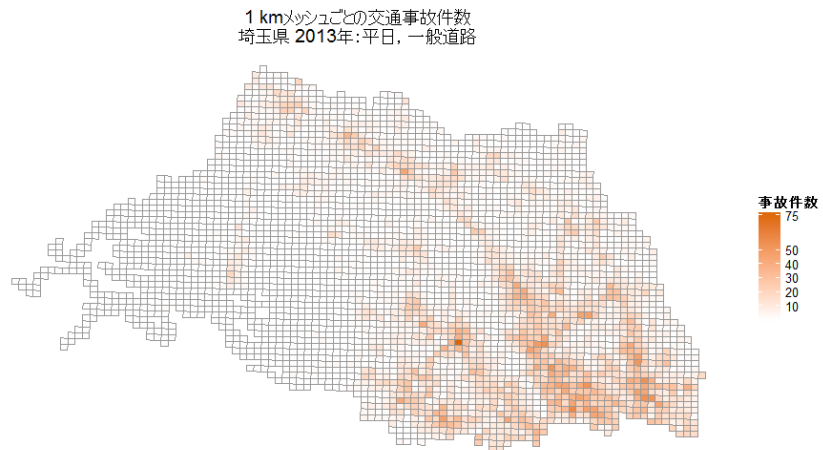


図 8 兵庫県メッシュ毎の p 値 (2013 年, 休日, 1 km メッシュ)

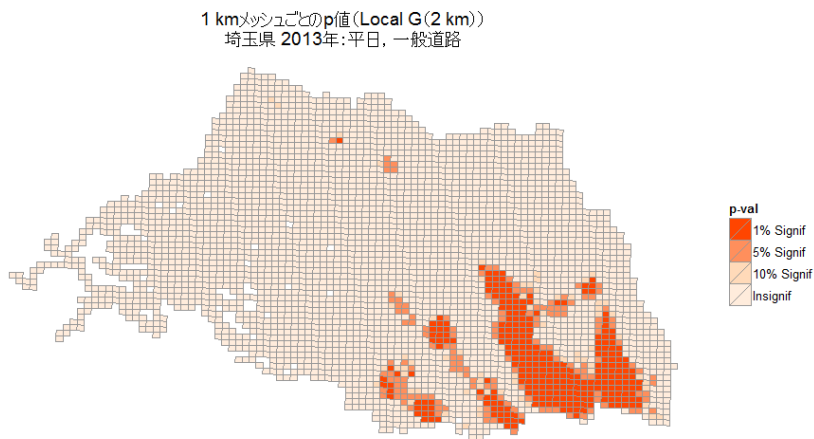
以上で兵庫県の一般道路で発生した交通事故のクラスタリングの分析が終わった。以下では兵庫県以外の都府県（埼玉県，東京都，及び大阪府）について，同じく一般道路で発生した交通事故のクラスタリングを分析した結果のコロプレス図を示す。分析は平日と休日の事故を分けて行ったが，両者の間に大きな差はないので，平日の結果のみを図 9～11に示した。

これらの図から，それぞれの都府県において交通事故が多発しているクラスターを判別することができる。こうしたメッシュにおいては，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*⁽⁵⁾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。

埼玉県については、メッシュ毎の事故件数のコロプレス図を見ると（図 9a）、東京寄りの市と国道 17 号沿いで事故件数が多いことがわかる。一方、Local G を使った検定結果からは（図 9b）、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。



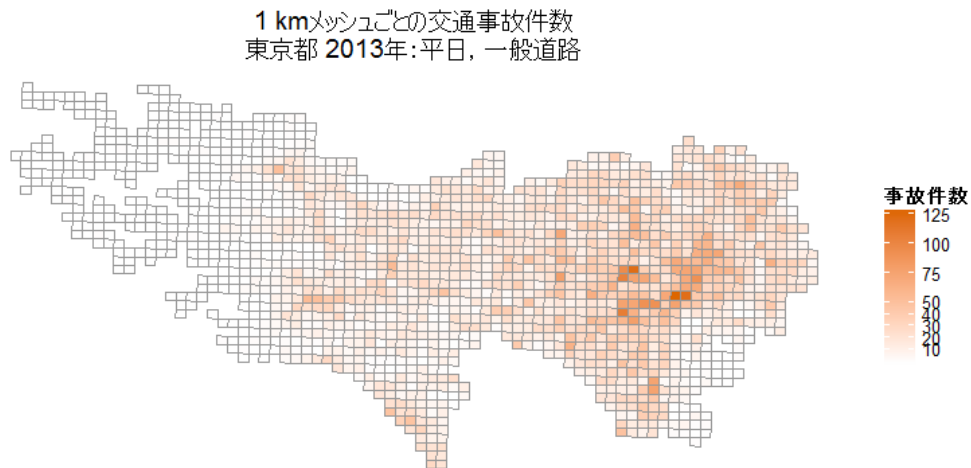
(a) 埼玉県の事故件数（2013 年，一般道路）



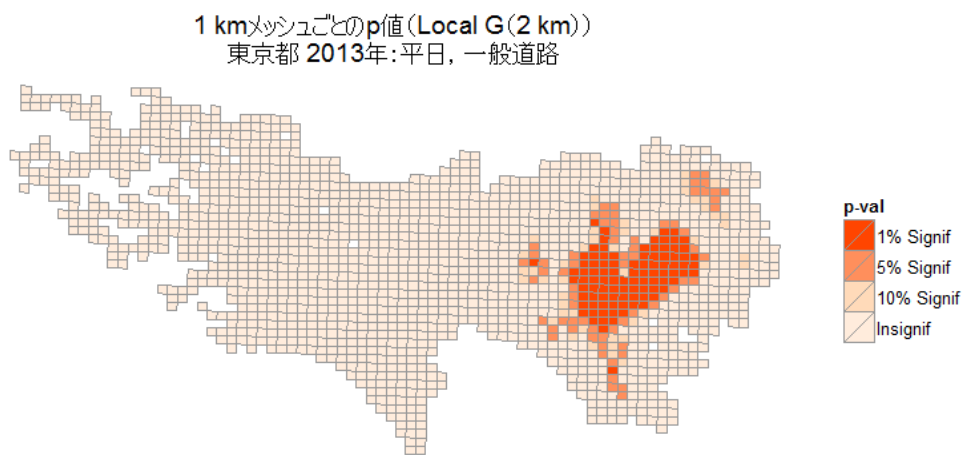
(b) 埼玉県の Local G（2013 年，一般道路）

図 9 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）

東京都では、都心部で交通事故の件数が多いことは当然として、都心から離れた郊外部でも事故が発生している（図 10a）。しかし、有意なクラスターとして抽出されているのは、都心部の他、環状 7 号の加平と駒沢付近、環状 8 号の荻窪付近、及び第一京浜の沿線である。郊外部のメッシュは有意なクラスターとはなっていない（図 10）。



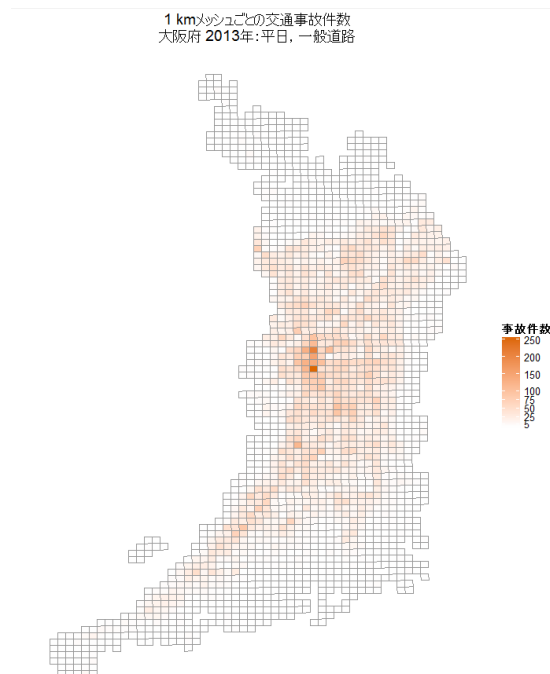
(a) 東京都の事故件数（2013 年，一般道路）



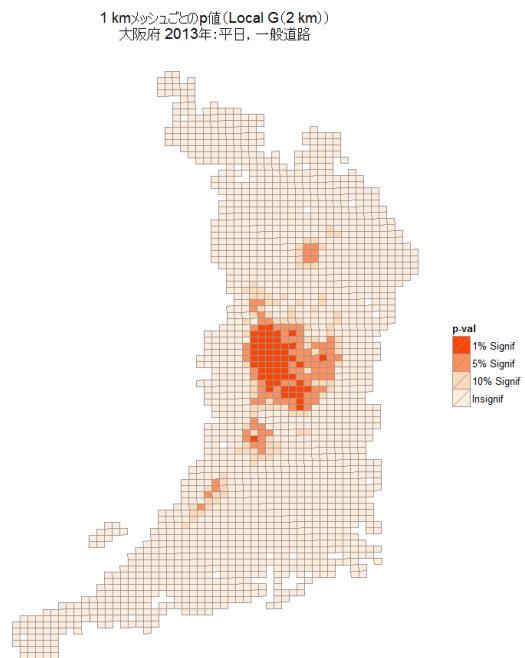
(b) 東京都の Local G（2013 年，一般道路）

図 10 東京都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 11a）。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 11）。



(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 11 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）

4 おわりに

本研究では、交通事故発生場所の経度・緯度情報を使って、交通事故の多発箇所の分布と重大事故の分布パターンの中に差があるかどうか、さらに交通事故が集中している場所があるかどうか、あるとすればそれはどこか、という二つの課題を分析した。

分析の結果、次の結論が得られた。

まず多発箇所の分布と重大事故の分布パターンについては、東京都市圏、京阪神都市圏で2013年に発生した幹線一般道路での事故データを使った分析から、多発箇所、死亡事故、そして重傷事故の分布パターンの中には統計的に有意な差があることがわかった。これは、死亡事故、重傷事故が必ずしも多発箇所が発生しているわけではないことを意味する。

さらに、1 km メッシュ単位に集計した事故データを使って事故のクラスタリングの分析を行ったところ、埼玉県、東京都、大阪府、兵庫県では、事故が集中して発生しているクラスターが存在し、それらは都市部に限定されていることが明らかになった。

こうした分析を通じて、事故の発生場所の情報を使うことによってこれまでは行えなかった交通事故の地理空間的な特徴を捉えることができるようになることを示せたと考えている。しかも、こうした地理空間的なデータの処理、分析から結果の図化までを統計ソフト R の中で完結した形でできることも明らかになった。このことは、地理空間的なデータの分析を容易にしてくれるという点で研究者にとって朗報である。

なお、本稿にはこれらの分析に使った統計ソフト R のコードを掲載した。このコードは、Brunsdon and Comber²⁾と Chang⁴⁾によるものをベースにして著者が一部改変したものである。こうしたすばらしいコードを作成してくださった著者の方々に感謝する。

謝辞

本研究では、ArcGIS に加えて以下のオープンソフトを利用させていただいた。これらのソフトの製作・改良に携わった方々に深く感謝する。とくに交通事故のパターン分析とコロプレス図等の描画では R とそのパッケージ RStudio に非常にお世話になった。

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

参考文献

- [1] (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.
- [2] Brunsdon C and Comber L: *An introduction to R for spatial analysis & Mapping*,

Sage Publications Inc, 2015.

- [3] Ripley B D: Modelling spatial patterns with discussion, *Journal of the Royal Statistical Society B*, 39, 172–212, 1977.
- [4] Chang W: *R Graphics Cookbook*, O'Reilly, 2013. (邦訳: R グラフィックスクックブック, 石井, 川内他訳, オライリー・ジャパン, 2013.)
- [5] 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.