

# 時間帯別滞留人口データを用いた街頭犯罪発生リスクの地理的要因の研究

足立 浩基\*・中谷 友樹\*\*

## A Study on Geographical Factors Determining Street Crime Risks Using Ambient Population Data

Hiroki M. ADACHI\* and Tomoki NAKAYA\*\*

According to environmental criminological theories, such as the routine activity and the crime pattern theory, ambient population affects crime risks in various settings. In recent years, several ambient population data disaggregated by daily time zone have become available at a relatively high spatial resolution enabling the detailed analysis of relationships between ambient population and crime risks. Further subdividing the time-specific volume of ambient population into gender and age groups may provide a new insight on the crime risk analysis, since the spatio-temporal patterns of crime differ depending on the gender and age of potential victim, offender, and guardian. In this study, we examine the relationships between the street crime risks including snatch-and-run and indecent assault and indices of detailed ambient population data in the southern part of Kyoto Prefecture using spatial statistical models.

**Keywords:** 街頭犯罪 (street crime), 滞留人口 (ambient population), モバイル空間統計 (Mobile Spatial Statistics), 空間統計解析 (spatial data analysis), 非負値行列因子分解 (NMF: Nonnegative Matrix Factorization), 条件付自己回帰モデル (conditional autoregressive models)

### 1. はじめに

犯罪は特定の場所と時間に発生することが知られている。犯罪発生リスクが高くなる時空間を明らかにできれば、法執行機関や市民に具体的な防犯策などを提言することができ、犯罪認知件数の削減に貢献できる。犯行機会は犯行者と犯行の対象（ターゲット）が同じ時空間に介在する際に生じると考えられるが、さらに日常活動理論は、「動機付けられた犯行者」、「適切なターゲット」、「有能な監視者の不在」が収束する時空間を犯行の成立条件としている (Cohen and Felson, 1979)。また、犯罪パターン理論は、人々の日常的な活動パターンの形成に着目し、犯行者やターゲットの活動と犯罪機会の関係を説明する (Brantingham and Brantingham, 1984)。これらに鑑みると、特に人をターゲットとした犯罪の場合は時空間的な人口分布の変化に基づいて犯罪発生リスクを評価できると考えられる。最近では、携帯電話や SNS による時空間的な解像度の高い時空間別

滞留人口データを利用できるようになり、こうしたデータに基づいた犯罪発生リスクの定量的評価研究が見られるようになった (Bogomolov et al., 2014; Malleson and Andresen, 2015; Amemiya et al., 2018; Hanaoka, 2018)。

しかし、潜在的に犯行者、監視者、ターゲットとなる人の性別や年齢はそれぞれ異なると考えられる。例えば、強制わいせつは周囲の監視者となる人口が少ない夜間に 10~20 歳代の女性が被害に遭う割合が高く、犯行者はほぼ男性である (警察庁, 2020; 内閣府, 2012)。性別や年齢別の時間帯別滞留人口を利用した分析によって、これら犯行者、監視者、ターゲットに関連した犯行パターンの時空間的な成立条件を考察できると考えられる。

そこで本研究では、2015 年~2019 年に京都府内で発生したひったくりと強制わいせつ（ただし屋外で発生したものに限り、以下、屋外強制わいせつとする）を対象に性年齢階級別時間帯別滞留人口

\* 学生会員 東北大学大学院環境科学研究科 (Graduate School of Environmental Studies, Tohoku University) 〒980-8572 宮城県仙台市青葉区荒巻字青葉 468-1 E-mail : adachi.hiroki.p2 @dc.tohoku.ac.jp

\*\* 正会員 東北大学大学院環境科学研究科 (Graduate School of Environmental Studies, Tohoku University)

データを用いて犯罪認知件数の空間回帰モデリングを行う。具体的には、時間帯別滞留人口データに非負値行列因子分解を適用することで得た各因子と犯罪発生リスクの関係を空間回帰モデルに基づいて議論する。

## 2. 方法

### 2.1. モバイル空間統計

株式会社ドコモ・インサイトマーケティングが提供する「モバイル空間統計」は、株式会社 NTT ドコモの携帯電話ネットワークの運用データを用いて作成される人口統計である(寺田ほか, 2012; Okajima et al., 2013)。

モバイル空間統計は2分の1地域メッシュ(500mメッシュ)などの空間単位における24時間365日の人口分布を総数、性年齢階級別、居住地(市区町村)別、居住地(都道府県)別の各パターンで集計している。本研究では、500mメッシュで集計された性年齢階級別人口分布を使用し、分析対象エリアを京都府内の京都市以南の地域として2019年1月1日から2019年12月31日までの祝日を除く平日(244日)における時間帯別滞留人口の平均値を分析対象エリアの一般的な時間帯別滞留人口とする。なお、性年齢階級別データは、性別(男性, 女性)の2カテゴリと年齢階級別(15-19歳, 20-29歳, 30-39歳, 40-49歳, 50-59歳, 60-69歳, 70-79歳)の7カテゴリのそれぞれの組み合わせで集計されている。本研究では、対象のメッシュごとに時間帯×性×年齢階級の計336列に展開し、いずれかのフィールドにゼロが含まれるメッシュは除外する。その結果1,520メッシュが分析対象となった。

### 2.2. 犯罪発生データ

本研究は京都府で2015年1月1日から2019年12月31日の間に京都府警察に認知されたひったくりと屋外での強制わいせつ事案を対象とする。両罪種とも犯罪発生時間は夜間の発生が顕著に多い(図1)。したがって、犯罪発生時間が日中(6~17時台)と夜間(18~5時台)のデータに分割する。ひったくりの認知件数は、日中が70件、夜間が267件。屋外

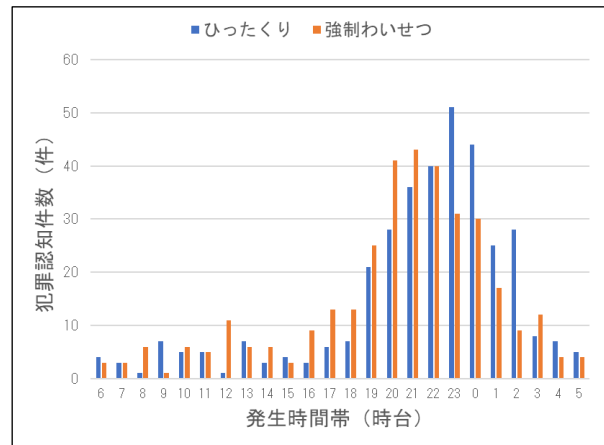


図1 2015年1月1日~2019年12月31日の平日に京都府京都市以南で発生したひったくりと屋外強制わいせつの発生時間帯別認知件数

強制わいせつは日中が46件、夜間が300件であった。それぞれの犯罪認知件数は500mメッシュごとに集計する。

### 2.3. 非負値行列因子分解による次元圧縮

非負値行列因子分解 (Nonnegative Matrix Factorization: NMF) は非負行列を非負行列の積で近似する行列分解法の一つである。非負行列の成分は非負値の足し合わせによって再構成されるため、行列分解の結果を解釈しやすい利点を持ち、画像の特徴抽出、文章クラスタリング、音源分離などでしばしば利用されている (Gillis, 2020)。

具体的には $X$ を $n \times k$ の非負行列(データ行列)、 $W$ を $n \times r$ の非負行列(基底行列)、 $H$ を $r \times k$ の非負行列(係数行列)とすると、

$$X \sim WH \quad (1)$$

のように分解する。これは $X$ と $WH$ 間の距離を最小化する以下の問題を解くことによって達成される

$$\min D(X, WH). \quad (2)$$

$D$ は距離関数であり、Kullback-Leibler ダイバージェンス (KL ダイバージェンス) や Euclidean 距離などが採用される。 $r$ はあらかじめ設定される圧縮後の

次元 (rank) であり,  $r \leq \min(n, k)$  である必要がある。

NMF は  $W$  と  $H$  の初期値によって異なる結果を与えるが, 本研究では,  $X$  の特異値分解によって得られる特異ベクトルの非負成分を初期値に採用する Non-Negative Double Singular Value Decomposition (NNDSVD) 法を用いる。NNDSVD 法は初期値を一意に決定でき, 特異ベクトルの負値の要素をゼロに置き換えることでスパース性を高める利点を持つ。また, 距離関数に KL ダイバージェンスを利用し, スパースな結果を期待できる nonsmooth nonnegative matrix factorization (nsNMF) アルゴリズム (Pascual-Montano et al., 2006) を採用する。NMF の実行には, R の NMF パッケージを用いた。

#### 2.4. 空間統計モデル

分析対象エリアに定義されたメッシュのうち犯罪が記録されているメッシュは稀である。したがって, 本研究では, 犯罪認知件数がポアソン分布に従うと仮定する。また, 未知要因に基づくランダム変動に空間相関を考慮する統計モデルである Conditional Autoregressive モデル (CAR モデル) を採用し, 隣接するランダム効果を平滑化する強さを推定できる Leroux モデル (Lawson, 2021) を用いる。Leroux モデルは以下の通り定式化される

$$y_i \sim \text{Poisson}(\lambda_i) \quad (3)$$

$$\ln(\lambda_i) = \mathbf{x}_i^T \boldsymbol{\beta} + u_i \quad (4)$$

$$u_i | u_{-i} \sim \text{Normal} \left( \frac{\rho \sum_j u_j w_{ij}}{\rho \sum_j w_{ij} + 1 - \rho}, \frac{\tau_u^2}{\rho \sum_j w_{ij} + 1 - \rho} \right). \quad (5)$$

$y_i$  は犯罪認知件数,  $\mathbf{x}_i$  は説明変数ベクトル,  $\boldsymbol{\beta}$  は係数ベクトル,  $u_i$  はデータ  $i$  の空間相関を持つランダム効果,  $u_{-i}$  はデータ  $i$  以外の空間相関を持つランダム効果,  $w_{ij}$  は近接行列,  $\rho$  は空間相関の強さ,  $\tau_u^2$  は空間相関変数の分散である。Leroux モデルは, 空間相関がない場合 ( $\rho = 0$ ) から強い空間相関がある場合 ( $\rho = 1$ ) までの空間相関の強さを柔軟に推定できる。

説明変数には, 人口分布変数 (性年齢階級別の時

間帯別滞留人口に NMF を適用して得られた係数ベクトル), 地域形態変数 (建物面積, 道路総延長), 施設変数 (鉄道駅の数, バス停の数, 高等学校の数, 大学・短大の数, コンビニエンスストアの数), 事業所変数 (卸売業・小売業数, 生活関連サービス業・娯楽業数) を用いる。これらの変数はあらかじめ標準化 (平均値を 0, 標準偏差を 1 に変換) する。なお, 説明変数の多重共線性を回避するため, あらかじめ, VIF が 8 以上となったコンビニエンスストアの数, 卸売業・小売業事業所数は除外した。CAR モデルの推定には R の CARBayes パッケージ (Lee, 2000) を使用する。

### 3. 結果

NMF の rank は, 毎回ランダムな初期値を与えた NMF を 30 回実行した結果から得られる Cophenetic 相関係数を参考に 8 とした。また, 基底ベクトル (以下, 因子と呼ぶ) の要素の和が 1 になるよう基準化した。NMF によって得られた各因子の解釈を表 1 に示す。また, これらの中から「因子 1」と「因子 4」を抜粋して図 2 に示す。

表 1 各因子の解釈

因子	解釈
1	郊外の住宅地
2	オフィス街やショッピング施設の多い地域
3	高校・大学付近の学生が多い地域
4	ナイトライフ施設やショッピング施設の多い地域
5	市街地に位置する古い住宅地
6	新興住宅地
7	工場などの就業地
8	病院・宗教施設・福祉施設が多い地域

「因子 1」は鉄道駅の周辺から少し離れた場所で高い値を持つ (図 2 (a))。21~7 時台は比較的高い値が続いている様子から, この時間は多くの人が自宅に滞在していると考えられる。8~18 時台は全体的に減少するが, 年齢が高いほど減少の程度は小さくなる。したがって, 「因子 1」は広い性年齢階級が居

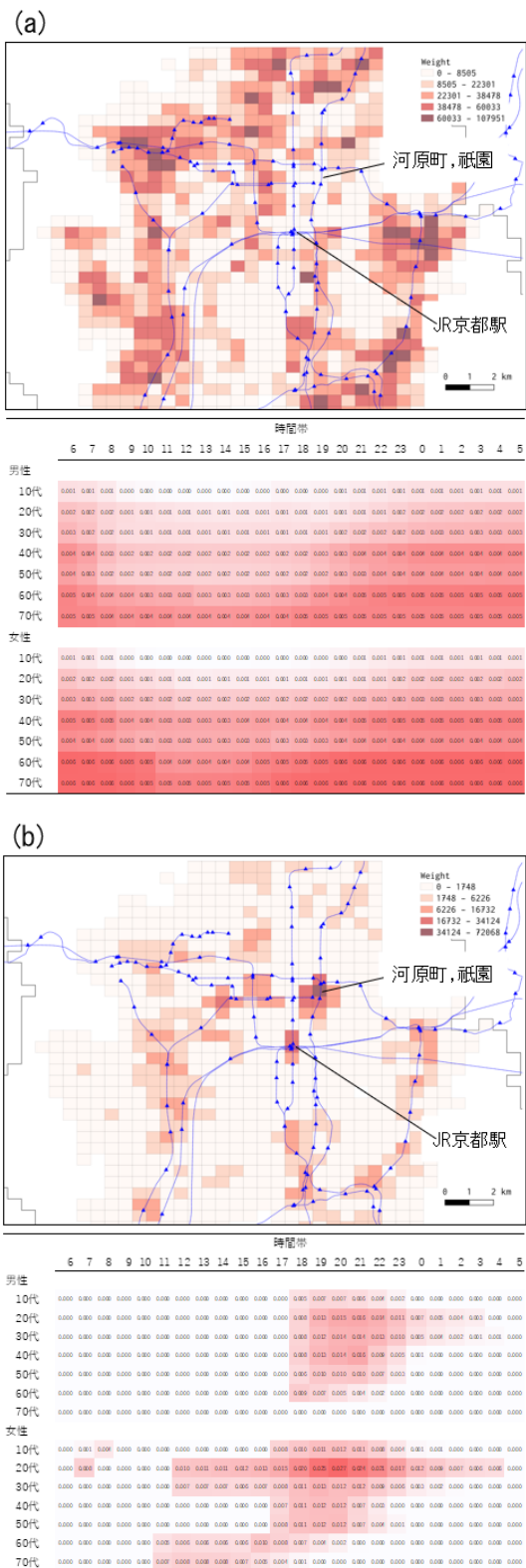


図2 各基底ベクトル (因子) の値. 地図中の三角形のシンボルは鉄道駅, 点線は鉄道線である.

(a) 因子 1, (b) 因子 4

住しており, 就学者や就業者は日中に他の地域へ移動する特徴を有すると考えられるため「郊外の住宅地」と解釈した。「因子 4」は主要駅や繁華街で顕著に高い値を持つ (図 2 (b)) 4~17 時台に男性の値はゼロであるが, 20~30, 60~70 歳代女性は 11 時台あたりから存在している. このことから日中に女性が訪れる施設が存在する地域であることが想像される. 18~23 時台になると, 主に 10~50 歳代の男女に増加が見られる. 特に 20~30 代は他の年齢階級よりも増加し, 翌 3 時台まで滞在している. このことから居酒屋などのナイトライフ施設が多い繁華街が考えられる. この解釈は, 70 歳代男性や 60~70 歳代女性が 18 時台以降に減少する様子から理にかなっている. これらの結果を基に「因子 4」を「ナイトライフ施設やショッピング施設の多い地域」とした.

空間回帰モデルの推定結果を表 2 に示す. なお, すべての説明変数は Geweke 検定統計量の絶対値が 1.96 未満であるため分布は収束していると判断した.

日中に発生したひったくりと屋外強制わいせつの人口分布変数について信用区間にゼロが含まれない係数パラメタはなかった.

夜間に発生したひったくりは「因子 1」, 「因子 2」, 「因子 4」, 「因子 5」, 「因子 7」, 「因子 8」で正の係数パラメタが推定された. 夜間に発生した屋外強制わいせつは「因子 1」, 「因子 4」, 「因子 5」, 「因子 8」でそれぞれ正の係数パラメタが推定された. それぞれのモデルについて, 人口分布変数の除いたモデルと偏差情報量規準 (DIC) を比較した結果, 人口分布変数のあり/なしについて, ひったくりはそれぞれ [あり: 1265.9, なし: 1274.2] となり屋外強制わいせつはそれぞれ [あり: 1098.6, なし: 1100.6] となった (表 2). いずれの罪種も人口分布変数を考慮した方が良いモデルを得られることが示された. パラメタ  $\rho$  の推定結果から, ひったくりは比較的強い空間相関を持つランダム効果を, 屋外強制わいせつは比較的弱い空間相関を持つランダム効果を持つ可能性が示された (表 2). 図 3 に犯罪認知件数と回帰分析の予測結果を示す.

表2 空間回帰モデルの結果

変数	夜間ひったくり		夜間屋外強制わいせつ	
	中央値	95% CI	中央値	95% CI
切片	-3.25	[-3.72, -2.86]	-2.91	[-3.28, -2.61]
人口分布変数				
因子1: 郊外の住宅地	0.21	[0.02, 0.42]	0.27	[0.10, 0.46]
因子2: オフィス街やショッピング施設の多い地域	0.15	[0.02, 0.31]	0.03	[-0.13, 0.18]
因子3: 高校・大学付近の学生が多い地域	-0.01	[-0.17, 0.15]	0.10	[-0.05, 0.23]
因子4: ナイトライフ施設やショッピング施設施設の多い地域	0.12	[0.02, 0.22]	0.13	[0.03, 0.23]
因子5: 市街地に位置する古い住宅地	0.18	[0.04, 0.30]	0.13	[0.00, 0.25]
因子6: 新興住宅地	0.05	[-0.07, 0.16]	-0.06	[-0.20, 0.06]
因子7: 工場などの就業地	0.20	[0.07, 0.34]	0.06	[-0.07, 0.19]
因子8: 病院・宗教施設・福祉施設が多い地域	0.16	[0.04, 0.28]	0.19	[0.07, 0.31]
建造構造変数				
建物面積	0.56	[0.24, 0.91]	0.39	[0.07, 0.73]
道路総延長	0.32	[0.07, 0.59]	0.50	[0.25, 0.78]
施設変数				
鉄道駅の数	0.12	[0.00, 0.24]	0.09	[-0.03, 0.21]
バス停の数	-0.08	[-0.22, 0.07]	0.05	[-0.08, 0.18]
高等学校の数	0.14	[0.01, 0.24]	0.10	[-0.03, 0.22]
大学・短大の数	0.06	[-0.06, 0.16]	-0.09	[-0.25, 0.05]
事業所変数				
生活関連サービス業・娯楽業施設数	0.00	[-0.19, 0.19]	0.01	[-0.19, 0.21]
$\tau^2$	3.76	[1.92, 6.35]	2.35	[1.00, 4.50]
$\rho$	0.71	[0.16, 0.98]	0.24	[0.02, 0.61]
DIC	1098.6		1265.9	
DIC (人口分布変数なし)	1100.6		1274.2	
DIC (空間相関変数なし)	1223.3		1387.8	

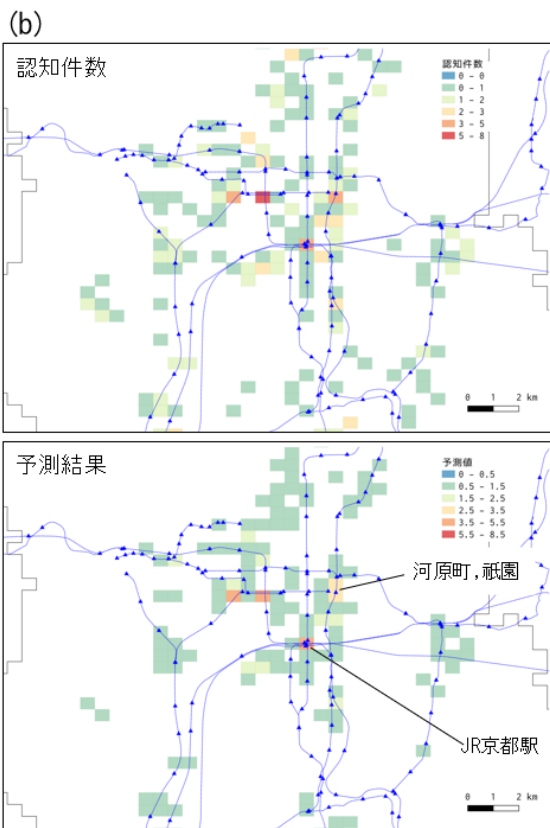
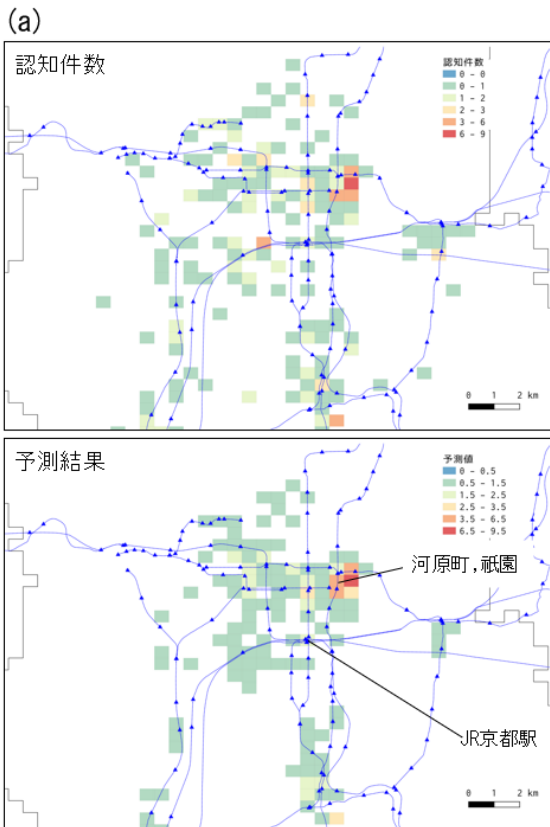


図3 夜間の認知件数と予測結果。  
(a)ひったくり、(b)屋外強制わいせつ

#### 4. 考察

夜間に発生したひったくりと屋外強制わいせつに共通して人口分布変数「因子1」、「因子4」、「因子5」、「因子8」で正の係数パラメタが推定された。このことからひったくりと屋外強制わいせつが発生する時間帯別滞留人口のパターンは類する部分があると考えられる。

「因子1」はほぼ全ての性年齢階級人口が21~7時台に屋内に滞在していると考えられるため、この時間は住民や高齢者による「有能な監視者の不在」が当てはまると考えられる。性犯罪の犯行者の行動パターンの一つに、駅やコンビニエンスストアといった多くの人が利用する施設でターゲットを選定し、尾行して犯行に及ぶケースがある(警察庁, 2019)。

「因子1」は夜遅くに帰宅してきた人が被害に遭うパターンを有する可能性がある。

「因子4」は、ひったくりと強制わいせつのターゲットになりやすい20代女性が12時台から翌3時台の間に増加が見られる。特に夜間は18~22時台の増加が特徴的な地域である。こうした地域は「適切なターゲット」となる人口の多さが犯罪発生リスクの増加に関連すると考えられる。また、この因子に特徴づけられる地域は、帰宅前のターゲットを狙った犯行パターンが多いと考えられる。

「因子5」は60~70歳代男女の人口には反応しない特徴(基底ベクトルの要素がゼロである)を有しており、高齢者による監視の目が低い可能性がある。一方、20~50歳代女性が21時台以降に増加する特徴を持っており、夜間に「適切なターゲット」の暴露が増える場所でもある。ターゲットの人口増加と直接関係するという意味では「因子4」に似るが、21~6時台まで値が一定で推移するため居住地の存在が考えられる。つまり「因子4」よりも「因子1」に近い犯行パターンを有すると考えられる。

「因子8」は主に40~70歳代男性、60~70歳代女性が日中に増加する特徴を持つ。20歳代女性は17~13時台に存在するが22~4時台は50~60歳代男性と20歳代女性しか反応しておらず20歳代女性が潜在的に存在する特徴を有する。したがって「因子8」は特に「適切なターゲット」と「有能な監視者の

不在」が同時に成立する特徴を持つと考えられる。

## 5. おわりに

本研究では、モバイル空間統計を利用して性年齢階級別の時間帯別滞留人口とひったくりや屋外強制わいせつの認知件数との関係性を評価した。はじめに時間帯別滞留人口は NMF によって因子を取り出し各因子ごとの地域特性を整理した。その後、これらの因子や POI 変数による犯罪認知件数の空間統計モデルを作成した。その結果、時間帯別滞留人口によって特徴づけられる地域特性と犯罪発生リスクの関係を定量的に結び付けることができた。また、各地域特性で考えられる犯行パターンを考察できた。

街頭犯罪の被害者は屋外に存在すると考えられるため、屋外の人口分布を利用することが望ましい。しかしモバイル空間統計は屋内外の人口分布を判別することができない。路上で発生した犯罪を正確に分析するためには屋外人口を考慮する必要がある。また、本研究では1年間の時間帯別滞留人口データから算出した時間帯別平均値を利用した。そのため、イベントや季節による人口変化は加味されていない。これらの課題は今後の研究で解決を図りたい。

## 謝辞

本稿をまとめるにあたり、研究のためにデータセットをご提供いただき、分析に関わる重要なアドバイスをいただいた京都府警察刑事企画課捜査支援分析センターの皆様へ感謝いたします。本研究で使用したモバイル空間統計は、株式会社ドコモ・インサイトマーケティングよりご提供いただきました。ここに記し、心より感謝を申し上げます。

## 参考文献

Amemiya, M., Ohyama, T. and Shimada, T. (2018) Analyzing Sex Crime Victimization Risk: Who, When, and Where is Riskiest? *Reports of the City Planning Institute of Japan*, **16**, 297-301.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014) Once upon a crime: Towards crime prediction from demographics and mobile data.

*ICMI 2014 - Proceedings of the 2014 International Conference on Multimodal Interaction*, 427–434.

Brantingham, P.J., and Brantingham, Patricia L. (1984) *Patterns in Crime*, New York: MacMillan.

Cohen, L.E. and Felson, M. (1979) Social change and crime rate trends: A routine activity approach. *American Sociological Review*, **44**(4), 588-608.

Gillis, N. (2020) *Nonnegative Matrix Factorization*, SIAM. 1-16.

Hanaoka, K. (2018) New insights on relationships between street crimes and ambient population: Use of hourly population data estimated from mobile phone users' locations. *Environment and Planning B: Urban Analytics and City Science*, **45**(2), 295–311.

Lawson, A.B. (2021). *Using R for Bayesian Spatial and Spatio-Temporal Health Modeling*, CRC Press.

Lee, D. (2013). CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors. *Journal of Statistical Software*, **55**(13), 1–24.

Malleson, N., and Andresen, M. A. (2015) Spatio-temporal crime hotspots and the ambient population, *Crime Science*, **4**(1), 10.

Okajima, I., Tanaka, S., Terada, M., Ikeda, D., and Nagata, T. (2013) Supporting Growth in Society and Industry Using Statistical Data from Mobile Terminal Networks – Overview of Mobile Spatial Statistic, *NTT DOCOMO Technical Journal*, **14**(3), 16-23.

Pascual-Montano, A., Carazo, J., Lehmann, D. and Pascual-Margui, R. (2006) Nonsmooth Nonnegative Matrix Factorization (nsNMF). *IEEE Trans. Pattern Analysis and Machine Intelligence*, **28**(3), 403-415.

警察庁 (2019) 子供・女性に対する犯罪等を防止するための対策に関する調査研究。

警察庁 (2020) 令和元年の刑法犯に関する統計資料。寺田雅之・永田智大・小林基成 (2012) モバイル空間統計における人口推計技術。NTT DOCOMO テクニカル・ジャーナル, **20** (3), 11-16.

内閣府 (2012) 「女性に対する暴力」を根絶するための課題と対策 ～性犯罪への対策の推進～。