

位置登録情報を利用した長崎の観光地間の 相互影響力評価

一藤 裕¹・村上 大輔²

(受付 2024 年 6 月 30 日；改訂 11 月 11 日；採択 11 月 11 日)

要 旨

観光は、日本の主要産業の一つである。特に地方都市では、地域活性化の役割を担っている。新型コロナウイルスにより大きなダメージを受けたが、コロナ前の状況、もしくはそれ以上の観光客数の増加により観光が活性化している。今後は、観光客をさらに誘致する方策や、オーパーツーリズムとなっている地域では観光客の抑制や制御する方策が必要となっている。しかし、個人旅行が増加しているため、全体的な動きや観光地間の移動に関する分析は難しい。そこで本研究では、通信キャリアの位置登録情報を利用して、観光地間の関係性を評価し解釈する方法を提案し、定量的に評価することで観光客誘致のための基礎的なデータを創出することを目的とする。本提案手法は、各観光地のユニークユーザ数と観光地間の移動数、ユーザの居住地情報を利用して、観光地間の相互影響力を評価するものであり、各地における来訪者数の増加の影響が、位置登録情報から推定された移動者数に比例的に拡散していくことを仮定する多変量時系列モデルがベースとなっている。この手法を長崎県の観光地に適用し、観光地間の関係性を評価した結果、居住地・年代ごとに行動が異なることが示され、居住地・年代ごとに広告宣伝方法を変えることで効率的に観光客の誘致ができる可能性が示された。

キーワード：観光、位置登録情報、個人属性情報、相互相関、ブースティング。

1. はじめに

観光は日本の主要産業の一つとなっている。東京や京都といった主要都市だけでなく、特に地方都市にとっては地域活性化を実現するための大きな柱の一つと言える。観光客誘致のために、地方自治体は様々な活動を積極的に行っている。例えば、観光地ごとに観光地域づくり法人(DMO)が立ち上がり、観光庁(2024)が登録を管理し、令和6年4月時点で全国301団体が登録されている。DMOは、観光庁によると「地域の多様な関係者を巻き込みつつ、科学的なアプローチを取り入れた観光地域づくりの司令塔となる法人」とされている。よって、観光客誘致のための科学的根拠となるデータを準備することが求められている。

従来の観光は団体旅行が主流だったため、観光客の行動を把握することが容易であった。しかし、現在の日本国内の観光旅行は個人旅行が多様化したため個人旅行が主流となっており、画一的な行動分析では観光客誘致の根拠として十分な対応が出来なくなっている。よって、個人旅行の行動を把握し、観光客が訪れる要因や行動分析に対応することが必要となる。個人旅行

¹ 長崎大学 大学院総合生産科学研究科：〒852-8521 長崎県長崎市文教町 1-14

² 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

の観光客の行動を把握する手段として、スマートフォンに搭載される Wi-Fi 機能や GPS、携帯基地局のログデータが挙げられる。

Wi-Fi 機能に注目した行動分析結果を利用したサービスとして、GMO (2024) や Cinarra Systems (2024) が提供する来店分析が挙げられる。これは、Wi-Fi の特性を利用したものである。店にアクセスポイントを置いておくと、Wi-Fi の機能を ON にしている端末を持った利用者が入店し通信可能範囲に入ると、通信のためのやり取りを行い通信した端末 ID がログに残る。通信した端末 ID と通信契約時に同意した個人属性情報を紐づけることで、個人の属性別に来店傾向を分析でき、集客につなげるための根拠を提示する。このサービスは、店のアプリと連携することでポイント付与などのインセンティブを用意できるため、利用者の行動分析には向いている。しかし、観光地のような広大な場所に設置することは難しく、また、観光地は定期的に訪れる場所ではないため、アプリと紐づけるといったことが難しく、データ数が十分に得られない問題がある。また、Wi-Fi パケットセンサーを使ってデータを収集し、来場者の推定や Origin-Destination (OD) 票の作成、観光地間の移動パターンを推定する関連研究が多くされている (Gao and Schmocker, 2022; 中西 他, 2018; 寺部 他, 2019; 神谷, 2018; 遠藤 他, 2019)。これらの研究は、観光地の限定的なエリアに Wi-Fi パケットセンサーを設置し、端末の MAC アドレスを利用して観光客数や OD データの生成、得られたデータから移動パターンの分析を行っている。観光客の細かな誘導・行動分析に関しては有益であるが、Wi-Fi パケットセンサーが捉えられるデータ範囲が理想値で半径 100m であり、対象観光スポットが広域の場合、設置場所や数を考慮しなければならない。また、MAC アドレスを接続先などでランダム化する機能が搭載されたスマートフォンの出現や、観光客がどこから来たのかなど観光客誘致に必要な個人の属性情報がこのデータには紐づかないため、観光客誘致の根拠として利用するには十分ではない。

次に、GPS のログデータを利用した行動分析サービスとして、Agoop (2024) の流動人口データが挙げられる。これは、スマホアプリから取得した GPS などの位置情報を秘匿化・統計加工した情報を使い、エリア毎の時間経過による来訪・滞在人口の推移や、OD などの人流を可視化したものである。マーケティング、需要予測、観光調査など様々な分野で活用できる。しかし、GPS データはアプリで取得することが主で、そのアプリをインストールされていない観光客の情報は取得できない。特に地方で分析を行う場合、数が十分にないと統計化された時点で具体的な数値が分からず分析できない可能性がある。また、GPS データを使った観光客の行動分析研究も多く行われている。例えば、Liu et al. (2022) は、登山に関する GPS データを用いた行動パターンのクラスター分析を行った。この GPS データはオープン化されているものであり、また、登山道にも限りがあるため、データ量が少なくてもデータの補完が可能でクラスター分析には耐えられる。しかし、県全体の観光地を想定した場合、十分なデータが集まらないという問題がある。D'Angelo et al. (2023) は、イタリアとクロアチアでクルーズ客に GPS 追跡デバイスを配布し、データ収集およびアンケート調査を実施し、観光客の移動パターンを収集した。観光地とそのルートを直線で結び空間点過程モデルを用いて分析し、観光客の滞留パターンを定量的に評価した。その結果、観光地からの距離やクルーズ客同士の相互作用が最も行動に影響力があることを明らかにした。個人を特定できる範囲で同じ属性を持った集団を対象とする分析は、観光行動の要因を明らかにするために非常に重要であるが、本研究のように広域かつ不特定多数を長期間対象として分析を行うにはあまり適さない。特に GPS と個人属性情報が紐づいたデータセットを日々収集することは難しいからである。

最後に、位置登録情報を利用した行動分析として、NTT docomo (2024) のモバイル空間統計を筆頭に、各社、携帯電話端末が事前準備として、携帯電話網に接続した際に登録される基地局単位的位置情報を用いた統計サービスが挙げられる。これらは、システム上ほぼリアルタ

イムに様々な人口統計を作成可能となっており、社会基盤の一つとしてデータ利活用が期待されている。例えば、大井 (2022) は、KDDI が公表している全国主要観光地の人流データを用いて 23 の観光地の日時データを作成した。そのデータを用いて観光に対し新型コロナウイルス感染症の影響が縮小している傾向を示した。また、観光需要の変動要因として時間軸に着目し、その要因として 1 週間内の変動が最大の影響を与えていることを明らかにしている。また、Yamaguchi and Nakayama (2023) は、モバイル空間統計を使って、新型コロナウイルス感染症の感染拡大時期における長距離移動パターンの変化を分析し、観光目的の移動が大幅に減少したがビジネス目的の移動はある程度維持されていることを明らかにしている。Nakanishi et al. (2018) もモバイル空間統計データを利用し、日本全国を 9 つの地域に分割し、その地域間の旅行者数の変化の要因を分析した。その結果、季節や週ごと、経済状況および気候が要因であることを明らかにしている。大井 (2020) はモバイル空間統計データの訪日外国人データを用いて、大阪市と和歌山市のインバウンドの観光の季節変動を分析した。その結果、大阪市では観光需要の安定化が進んでいるが、地域格差が存在していることを示した。その一方で和歌山市はデータ数が限られていたため、特定のエリアでのみ分析できなかったことを述べている。Qian et al. (2021) は、観光地、休憩場所、交通ハブの空間的な相関を明らかにし、観光施設の相互関係を評価する枠組み、特に観光地間の結びつきを強化するためのバスの路線の追加について検討した。このように観光客の行動を分析するために、携帯電話の位置情報を利用することは非常に重要であり、観光客の滞在数や移動に着目しそのパターン化や要因を分析する研究が活発に行われている。また、観光地間のつながりについても移動手段や OD データによって、隣接する観光地や交通のハブなどの結びつきを評価し、ネットワークのハブとして整備する都市計画の効率化に主眼が置かれている研究が多く、特定の観光地の観光客数を観測し、その結果から他の観光地の観光客数を推定するような研究は行われていない。加えて、Aasa et al. (2021) が携帯電話の位置情報を利用した統計情報の有用性について報告しているが、その一方で、個人属性情報が位置登録情報に完全に紐づくため、その扱いに関する議論が継続的に行われており、データ利活用の障壁が高いことも併せて報告している。また、詳細に広く追跡できてしまうため、可視化した後に自治体側がどう活用してよいか分からないという問題もあり、位置登録情報の社会での利活用方法の実装は議論の余地が多くある。

そこで本研究では、位置登録情報の観光客誘致への利用に向けた検討として、位置登録情報から得られる観光地間のつながりが、各観光地の来訪者数に及ぼす影響を分析する。具体的には、観光地別・週別来訪者数の時間変化傾向を解析し、週を追う毎に変化する観光地間の結びつきが、翌週以降の来訪者数に及ぼす影響を評価することを目的とする。もし観光地間の結びつきの強さが来訪者数に影響しているのであれば、個別地域ごとの誘致だけでなく、結びつきの強い観光地も含めた一体的な誘致も有効である可能性がある。また、他の観光地に強く影響するような観光地が特定できれば、同観光地への誘致を積極的に行うことで、周囲の観光地への波及効果も期待できる。反対に、観光地間の結びつきの強さが来訪者数に影響していないのであれば、個別観光地ごとに誘致を行えば十分ということになる。以上のように、各観光地間の相互影響関係とその影響を明らかにすることは、誘致を効果的に行う上で重要である。

しかしながら、相互影響関係を評価する方法は必ずしも明らかではない。観光地別・週別の時系列データが利用できる場合であれば Vector autoregressive (VAR) モデルなどの多変量時系列モデルが応用できる。しかし、本研究のように数十の観光地が分析対象の場合、仮に全ての時点で観光地間の結びつきの強さは同じと仮定したとしても、観光地間のつながりを表すパラメータの数は (観光地数)² と多くなり効果の識別は困難である。さらに、結びつきの強さが時点によって変わるというより現実的な仮定をおいた場合、同パラメータ数は (地域数)² × 時点数となり、効果の識別はより困難となる。時点毎にパラメータを推定するために、パラメータ

を時間の関数で与える VAR モデルの拡張 (Christopoulos and Leon-Ledesma, 2008; Bringmann et al., 2018) や、直近の標本により大きな重みを与えることで局所推定を行おうという VAR モデルの拡張 (Casas and Fernandez-Casal, 2019; 村上・松井, 2022) もあるが、それらは必ずしも推定結果の解釈性を担保するものではない。

以上を踏まえ、本研究の章立ては次の通りとした。2 章では位置登録情報および個人属性情報を利用して、長崎県内の観光地毎の属性別観光客の来訪者数データを作成する。また、観光地間の結びつきの強さを表すデータとして OD データも併せて作成する。3 章では、同 OD データで表される観光地間のつながりが各地の来訪者数に及ぼす影響を評価するための新たな手法を提案する。4 章では、同手法を用いて観光地間の相互影響力を評価し、ある観光地に人が増加した場合、他の観光地にどのような影響を及ぼすか、居住地によって行動がどう異なるかを定量評価する。最後に 5 章で結果を要約し今後の展望について論じる。なお、周遊行動をはじめとした個別観光客の時間帯毎の移動パターンなども考慮したより詳細な解析については今後の課題としたい。

2. 利用データとプライバシー保護の取組み

2.1 位置登録情報の詳細

本稿では、大手携帯電話事業者の一つである株式会社ソフトバンクの位置登録情報を利用する。今回利用した位置登録情報には、居住地 (関東エリア、福岡県、長崎県)、毎週日曜日のユニーク検知数、年代 (5 歳刻み。紙面の制限のため、今回は 30–34 歳、60–64 歳のみを比較) の属性情報が付随する。位置情報が記録されるタイミングは、基地局の通信カバー範囲に入ったタイミングであり、同じ範囲に居続ける限り情報は更新されない。通信カバー範囲と観光地の対応付けはされているため、観光地ごとの検知数を算出できる。また、居住地の関東エリアは、東京・千葉・神奈川・埼玉・茨城・栃木・茨城・長野・群馬の羽田空港を利用して長崎を訪れる可能性の高い地域をまとめたものである。これらのデータからユニークユーザ数を集計した来訪者数データと OD データの 2 種類を利用する。

来訪者数データは、観光地毎に居住エリア別、日別かつ年代別で集計したものである。また、OD データは、検知された観光地に来る前に、最後にどこで検知されたかの情報を利用して、2 地点間の移動検知数を集計したものである。ただし、直前の位置情報が検知された日時については考慮しないものとする。例えば、A 地点で 6 月 1 日 8 時～10 時に検知され、直前に B 地点にいた場合を考える。B 地点で検知された日時は考慮しないため、B 地点で 6 月 1 日 6 時～8 時に検知されていても、もしくは、それ以前に検知されていた場合でも同様に「B→A」の検知数に集計される。対象期間は、2022 年 8 月 1 日から 2023 年 7 月 31 日までとした。また、今回は観光客を分析対象とするため、毎週日曜日データのみを用いることとした。

2.2 位置登録情報のプライバシー保護の取組みについて

位置登録情報には、通信の秘密には該当しないものの、プライバシーの観点で強い保護が求められることから、十分な匿名化の実施が必要となる。「十分な匿名化」には、位置情報と付帯情報とを結合して作成したデータを用いることができる。ただし、結合することのできる付帯情報 (電気通信事業者が保有する契約者に係る情報) は、性別、年齢、市区町村までの住所や利用者の趣味趣向情報である。また、「十分な匿名化」により加工した位置情報について、次のすべての条件を満たせば、事前の包括的同意に基づいて活用をすることができるとしている。

- オプトアウトが適切に提供されている。
- 加工及び運用管理体制が適切である。

- ・プライバシー影響評価(PIA)が適切に運用されている。

株式会社ソフトバンクでは、位置登録情報を含む運用データの取得・保有・利用に関する取扱いを Web 上に掲載している (SoftBank, 2024)。これに基づき、個人を特定可能な項目を符号や番号等へ置き換える仮名化と、個人を特定可能な項目を要約等する一般化の方法を用いて匿名加工処理を施す。その後、匿名加工情報を、性別、年代、居住エリア毎に集計し、かつ対象となるデータ内に同じ属性を持つ数値データが k 件以上存在するようにデータを変換する k -匿名化処理 ($k=10$) も併せて実施し、十分に匿名化した上で、統計データを作成する。このように処理されたデータを用いて観光地間の相互影響力を評価する。

2.3 観光地の選定と観光客の傾向

本稿で選定した長崎県のエリア(観光地や交通ハブ)とその特徴について述べる。長崎県内の観光地は県内に散らばっており、すべての観光地を 1 日で回るのは難しい。そこで、図 1 に示す通りに観光エリアを選定した。具体的には、長崎市(長崎駅、浜町、平和公園、グラバー園、出島、ペンギン水族館)、時津町、佐世保市(佐世保駅、ハウステンボス)、諫早市、西海市、平戸市、波佐見町、松浦市、島原市、雲仙市、九十九島エリア、大村空港、口之津の計 20 エリアである。この中で、長崎駅、佐世保駅、大村空港、口之津は交通のハブである。また、雲仙市は小浜温泉エリアと雲仙温泉、島原市は世界遺産群の一部、九十九島エリアはリゾート地、波佐見町は焼き物、諫早市は V・ファーレン長崎のホームスタジアム、その他のエリアは歴史的遺産があるという特徴がある。

上記エリアの来訪者数は、エリアと紐づく基地局のログから算出される。そのログから 2 時間ごとに集計しているため、2 時間ごとの集計では同一端末は重複してカウントされないが、1 日単位では重複してカウントされる場合もある。例えば、別のエリアに移動し、その後、元のエリアに戻ってきた場合は、2 時間を超えていれば再度カウントされることになる。この集計ルールに基づいてエリア別の来訪者数を居住地(長崎県、福岡県、関東エリア)と年齢(30-34 歳、60-64 歳)について図 2 にプロットした。図 2 から、来訪者数の週毎の変動は比較的小さいことや、エリア毎の差は大きいことが確認できる。長崎県居住者は各エリアを比較的満遍なく訪問している。一方で、福岡県民と関東エリアの住民の訪問先は、長崎駅、浜町、ハウステンボス、佐世保駅といった中心エリアに集中していることがわかる。また、福岡県や関東エリ

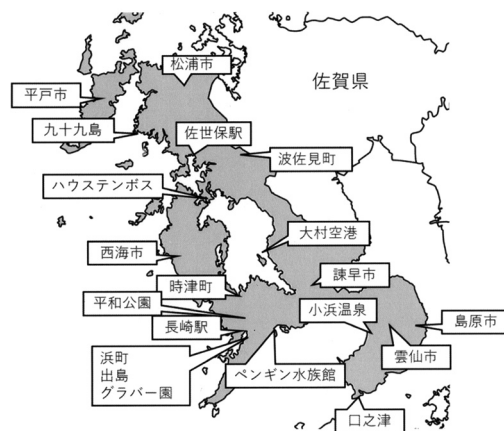


図 1. 長崎県内の観光地および交通ハブの位置関係。

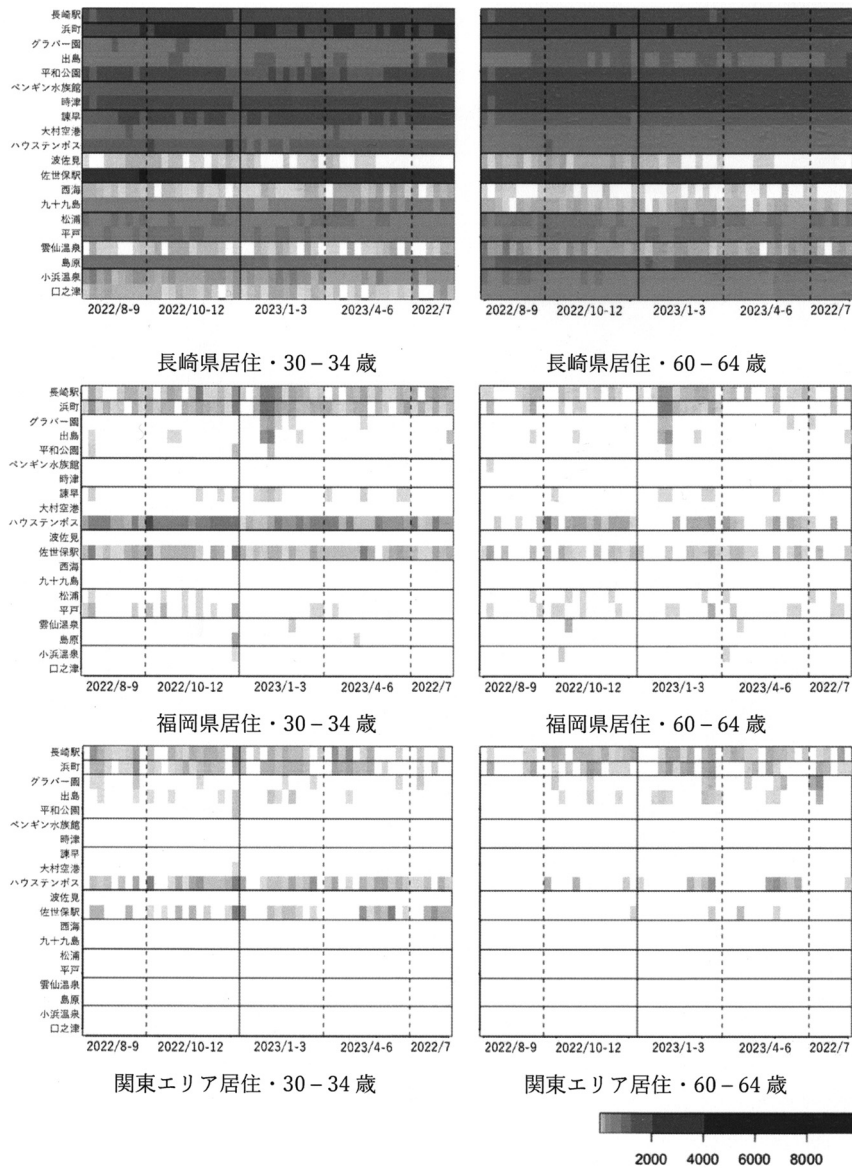


図 2. 居住エリア・年齢別の滞在者数.

アからの来訪者数は、30-34 歳の方が 60-64 歳よりも多いことなども確認できる。

3. 観光地の相互影響力評価モデルの提案

本研究では、あるエリアでの来訪者の増加が、同エリアのその後の来訪者数や、周辺エリアのその後の来訪者数に及ぼす影響を推定する。それにより、エリア別来訪者数の相互依存関係を明らかとする。

まず、エリア $s \in \{1, \dots, S\}$ で週 $t \in \{1, \dots, T\}$ に来訪者が 1 単位増えることを、第 s 要素が

1でその他が0のone-hotベクトル $\delta(s, t) (S \times 1)$ を表す．本研究では，この1単位増加に伴う L^* 週先のエリア別来訪者数の増分は下式に従うと仮定する：

$$(3.1) \quad z(s, t + L^*, L) = \prod_{l=1}^{L^*} \rho_{(L)}^l W_{t+l} \delta(s, t).$$

W_t は第 t 週におけるエリア間の移動者数の行列 $(S \times S)$ であり，その要素はODデータから与える． L を来訪者数増加の影響が残る期間とすると，パラメータ $\rho_{(L)}$ は期間内での時間減衰を表現するために $\rho_{(L)}^L = 0.1$ を満たすように与える($\rho_{(L)} = 0.1^{1/L}$)．ただし L は未知とする．上式はエリア s ，時点 t における来訪者増加の影響が，移動者数に従う形で他地域に拡散していき，その影響の90%が L 期先までに消失することを表している．例えば $\rho_{(L)} = 0.01$ とすれば L 期までの影響の消失率が99%となるなど， $\rho_{(L)}$ の設定には恣意性が残るが，影響の及ぶ期間は L で調整されるため， L との識別の難しさも勘案して消失率は90%で固定することとした．

本研究では，エリア s ，週 t ，期間 L を変えながら来訪者数増加の影響を足し合わせていくことで，単純集計データから得られるエリア別，週別の来訪者数行列 $Y (S \times T)$ に対する予測行列 $\hat{Y}$ を最適化する．(3.1)式で記述される $t+1$ 週から $t+L$ 週までの来訪者数の増分を対象期間全体で並べた行列を $Z(s, t, L) (S \times T)$ と表すと，その推定手順は以下のとおりである：

- (1) 予測行列の初期値 $\hat{Y}_0$ を設定． $i = 0$ ．
- (2) 以下の手順で予測行列を更新：
 - (a) 週 t_i をランダムに選択
 - (b) エリア $s \in \{1, \dots, S\}$ と拡大期間 $L \in \{1, \dots, L_{max}\}$ の全組み合わせについて，以下の手順で目的関数を評価(グリッドサーチ)：
 - i. 週 t_i ，エリア s で1単位増加した影響が L 期先までの続いた場合の，来客数の増分 $Z(s, t_i, L)$ を(3.1)式を用いて評価．
 - ii. $vec(Z(s, t_i, L))$ を残差 $vec(Y - \hat{Y}_i)$ に回帰して回帰係数 $\hat{\beta}_i$ を推定．新たな予測行列 $\hat{Y}_i + \hat{\beta}_i Z(s, t_i, L)$ を得る．ただし $vec(\cdot)$ は行列をベクトル化する演算子である．効果の識別を容易にするために，ここでは $\hat{\beta}_i \geq 0$ を満たすように制約を課して最小2乗推定している (Lee and Seung, 2000)．
 - iii. 同予測値の目的関数を評価．
 - (c) 手順(2)(b)で目的関数を最良にしたエリア s_i と拡大期間 L_i を用いて，予測行列の候補 $\hat{Y}_{i+1}^* = \hat{Y}_i + \hat{\beta}_i Z(s_i, t_i, L_i)$ を与える．
 - (d) $\hat{Y}_i$ よりも目的関数が改善する場合は $\hat{Y}_{i+1}^*$ を採択($\hat{Y}_{i+1} = \hat{Y}_{i+1}^*$)，さもなくば $\hat{Y}_{i+1} = \hat{Y}_i$ のままとする．
- (3) i を $i+1$ に更新．
- (4) 手順(2)–(3)を，目的関数が改善しなくなるまで繰り返す．

同手法は誤差を減らすように弱学習器を合成していくブースティングの一種である (Freund and Schapire, 1996)．なお，週 t_i をランダムに与えたのは，週数が大きくグリッドサーチにした場合に計算コストの増大が見込まれた点，並びに局所解への収束を避けるためである．目的関数として本研究ではBayesian Information Criterion (BIC)を用いた．以上で得られる時点 T の予測来訪者数の予測値ベクトル $(S \times 1)$ は下式となる：

$$(3.2) \quad \hat{Y}_T = \sum_{i=1}^l W_{T,i}^* \delta(s_i, t_i)$$

ただし $W_{T,i}^* = \hat{\beta}_i \delta(t_i + 1 \leq T \leq t_i + L_i) \prod_{l_i=1}^{T-t_i} \rho_{(L_i)}^{l_i} W_{t_i+l_i}^{L_i}$ である．

$W_{T,i}^*$ 同行列を全反復について足し合わせた $W_T^* = \sum_{i=1}^I W_{T,i}^*$ は時点 T におけるエリア間の相互相関関係を表す行列になっている。その第 (s, s') 要素は、エリア s' の来訪者数が増えた場合に、翌週以降にエリア s の来訪者数が増える傾向がある場合には正、さもなくば 0 となる。従って、 W_T^* の列和が大きいエリアは各地の来訪者数に影響するエリア (以後、主要エリアと呼称)、 W_T^* の行和が大きいエリアは主要エリアに依存して来訪者数が上下する従属エリアと考えることができる。以上のような解釈をもとに、4 章では W_T^* の推定結果をもとに、エリア間の主従関係をみていく。ここで、本研究における主要エリアとは、あくまで他地域への強い影響が推定された地域のことであり、観光客の多くが同エリアを経由して周遊していることを意味しているわけではない点には注意されたい。

なお、一連の推定ではパラメータ β_i の符号を非負に制約することで識別の問題を緩和している。具体的には、極端に大きな正の推定値が、極端に小さな別パラメータの負の推定値で相殺されることにより生じる特異推定を回避している (村上・松井, 2022)。また、予測式は観光地における来訪者増加の影響の拡散を明示的にモデル化した (3.1) 式の線形和に制限されているため特異になりにくく、かつ解釈しやすい。以上より、提案手法は推定の安定性や解釈性を重視して観光地間の相互影響関係を評価しようという手法となっている。

4. 提案手法を利用した長崎県における観光地間の相互影響力評価

本章では、毎週日曜日のエリア別来訪者数 (30–34 歳と 60–64 歳) を提案モデル ((3.2) 式) に適用することで、観光地間間の相互影響力を評価し、関係性を明らかとする。

4.1 年間の観光地間の相互影響力評価結果

提案モデルから推定されたエリア間の相互相関行列を見ていく。まずは、全期間の平均値を図 3 に示す。行列の第 (i, j) 要素の色が濃いことは、エリア i の傾向が、エリア j の翌週以降の傾向をよりよく説明することを意味する。図 3 上段から、長崎県民にとっての主要エリアは、長崎駅、浜町、平和公園、時津、諫早、佐世保駅などであり、長崎県内の各エリアの集客状況は、それら主要エリアの集客状況で説明されとの示唆を得た。

福岡県民にとっての主要エリアは、長崎駅、浜町、ハウステンボスエリアなどであり、それらのエリアの傾向から、翌週以降の長崎市内 (長崎駅～平和公園) の傾向が説明されること、ならびに、ハウステンボスの傾向から、翌週以降の佐世保周辺エリア (ハウステンボス含む) の傾向が説明されとの推定結果となった。関東エリアからの来訪者にも類似した傾向がみられたが、長崎駅・浜町エリアの来訪者数は翌週のより広域の来訪者数を説明する傾向がみられた。

以上のように、各地の来訪者数は比較的少数の主要エリアの傾向で説明されることや、主要エリアの数は、県外からの来訪の場合に少なくなる傾向がみられた。これは、長崎県内居住者の方が、比較的マイナーな観光地を訪れることや、県外からの来訪者が主要な観光地を愛好するための可能性がある。

参考までに、時系列相互作用を評価する標準的手法である Vector autoregressive (VAR) モデルを用いて地域間の相互相関を推定した。ここでは、安定した解を得るために SparseLag 損失 (Nicholson et al., 2017) を導入した。VAR モデルから得られた相互相関行列を図 4 に示す。図 4 から、提案手法とは異なり長崎駅や浜町がハブとなっておらず、また例えば雲仙温泉が出島エリアとだけ強い相互相関を持つなど解釈に窮する箇所が散見される。提案手法を用いることで、エリア間の主従／相互関係が、より解釈しやすい形で推定されることが確認された。

解釈しやすい結果が得られた理由には、ある観光地における来訪者増加の影響の拡散を明示的にモデル化した点や、同影響の符号を非負に限定することで効果の識別を容易にした点

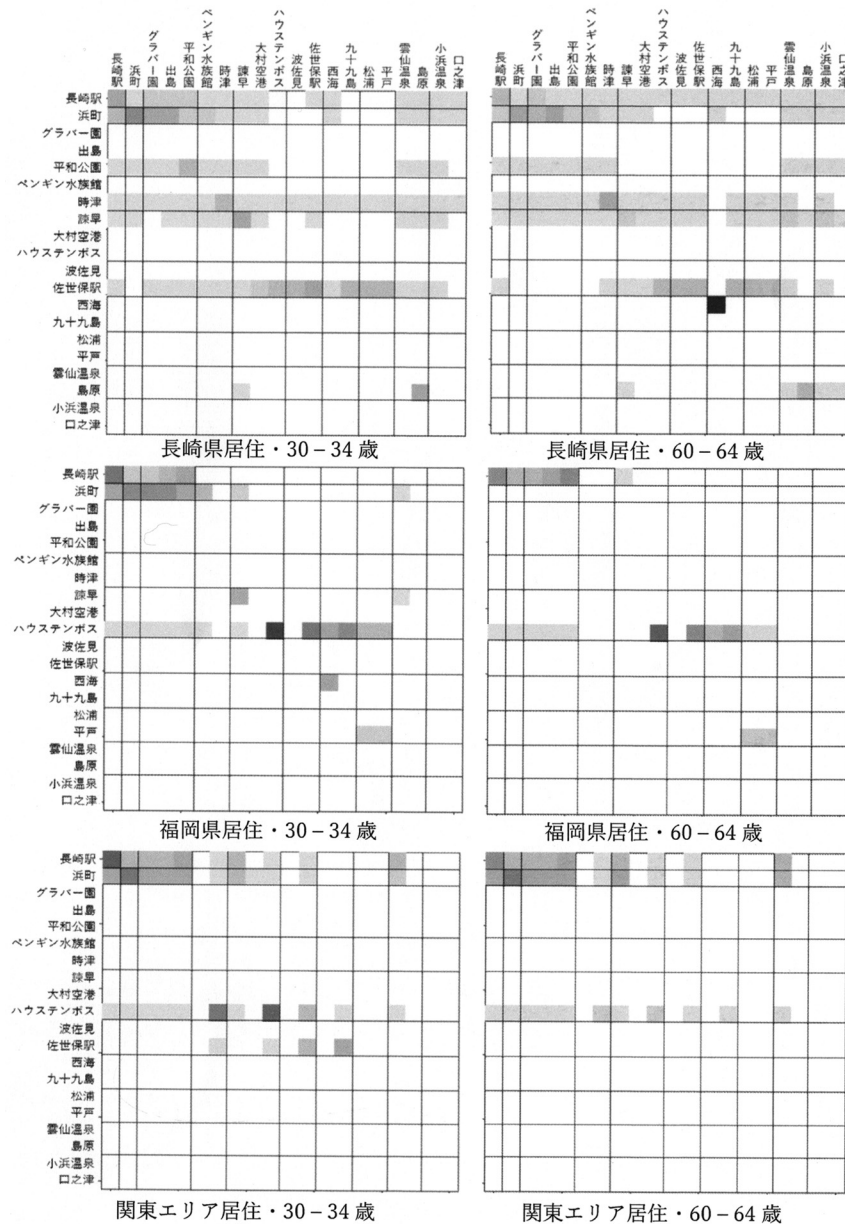


図 3. エリア間の相互作用行列の推定結果(全期間の平均).

($\hat{\beta}_i \geq 0$) などが考えられる。推定したパラメータの数が極めて多いことを踏まえると(1章参照), 妥当と思われる以上の仮定を導入することで地域間・週別の相互影響関係が解釈しやすい形で評価できるという知見は興味深い。言い換えると, おおむね予想されていた傾向が実データ分析を通して確認された。

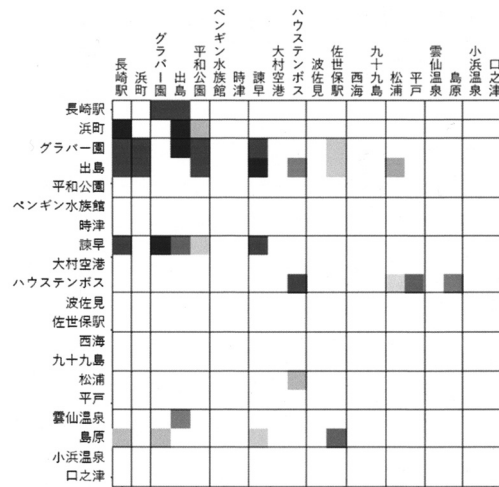


図 4. VAR モデルから推定された相互相関行列(福岡県居住・30-34 歳).

4.2 週毎の観光地間の相互影響力評価結果

週毎に推定された相互相関行列 W_T^* をみていく. 紙面の関係から, ここでは同行列の対角要素, 列和, 行和についてそれぞれ週別に見ていく.

W_T^* の対角要素の大小を濃淡で表したものを図 5 に示す. 図 5 で色が濃いことは, 同一エリア内の時系列相関で説明される来訪者数が多く, 先週までと似た来訪者数となりやすいことを意味する ($\beta_i \geq 0$ を仮定したため). 長崎県内からの来訪者数(30-34 歳)に関しては, 半数程度のエリアで系列相関がみられた. 特に長崎駅, 浜町, 諫早市, 佐世保駅, 島原市などの主要エリアでは, 期間を通して系列相関が強く, 直近と同数程度のまとまった来訪があるとの示唆を得た. 同じ長崎県民でも, 60-64 歳に関しては, 2022 年末頃までは系列相関がみられないエリアも多く, 通年で系列相関がみられたのは長崎駅・浜町エリアのみであった. これはコロナウィルスの感染を回避するための自粛をより長い期間行ったためである可能性がある.

福岡県からの来訪者数に関しては, 長崎駅, 浜町, ハウステンボス, 佐世保駅, 平戸市で系列相関がみられ, 関東からの来訪者に関しては, より少数の主要エリアでのみ系列相関が確認された. 以上より, 遠方であるほど, 複数週にわたる(系列相関で説明されるような)まとまった来訪者があるエリアは限られるとの示唆を得た. なお, 長崎県民からハウステンボスへの来訪者に関しては系列相関がみられず週毎に傾向が異なると推定された一方で, 福岡県または関東エリアからハウステンボスの来訪者に関しては強い系列相関がみられた点は興味深い.

次に, 週毎の W_T^* の列和を並べて図 6 に示す. 色が濃いことは, 翌週以降の他エリアの傾向をより良く説明する主要エリアであることを意味している. 図 6 から, 長崎県民の各地の来訪者数は, 長崎駅, 浜町, 佐世保エリアにおける傾向で説明されると推定された. それらエリアが人口密集地であり多くのトリップが発生していることを踏まえると, この結果は自然である. 福岡県民と関東エリアの住民に関しては, 上記 3 エリアに加えてハウステンボスでも色が濃くなっている. 県外からの来訪者にとっては, ハウステンボスもまた他エリアの傾向を説明する主要エリアとなっているとの示唆を得た.

最後に, 週毎の W_T^* の行和を並べて図 7 に示す. 色が濃いことは, 主要エリア(長崎駅, 浜町, 佐世保, ハウステンボスエリア等)でより多くの傾向が説明される従属エリアであることを意味する(3 章参照). 図 7 から, 長崎県民に関しては, 各地の来訪者数(30-34 歳)が主要エ

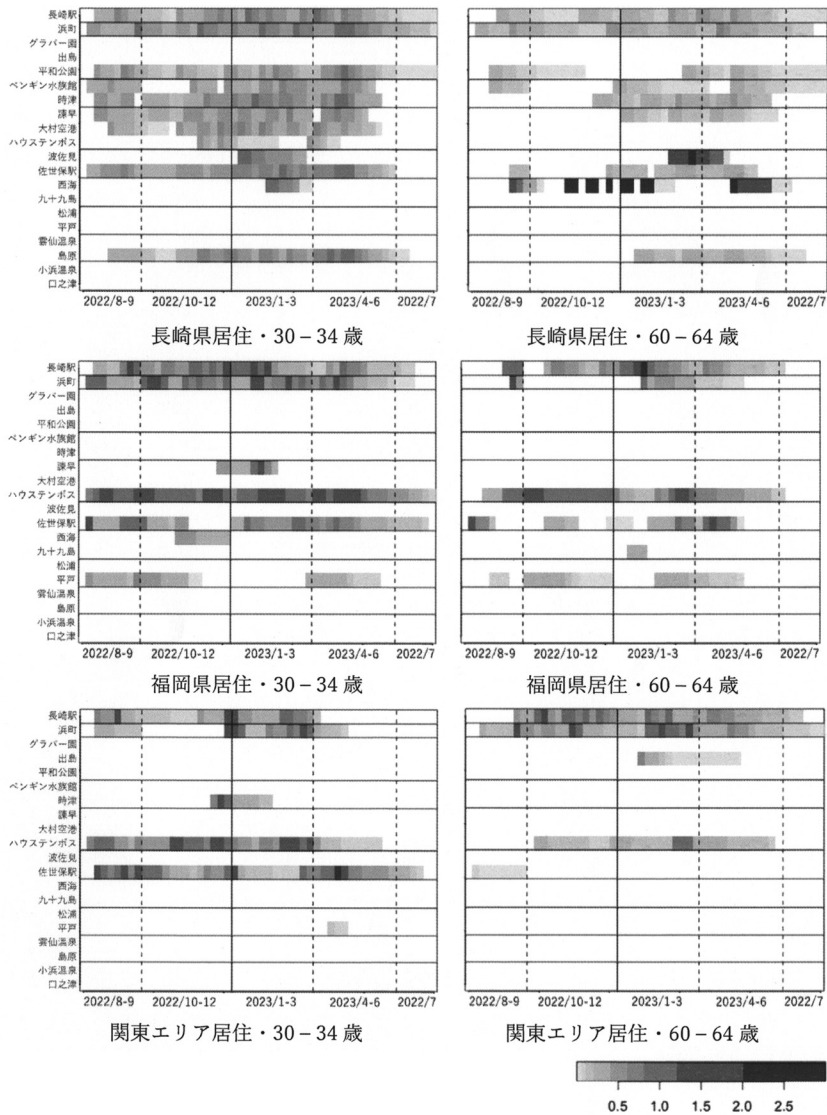


図 5. 同一エリア内の系列相関で説明される来訪者数の推定結果(相対値).

リアの傾向で説明される従属エリアとなっていることが確認できる。60-64 歳については 2022 年末までは相互相関がみられない地域もあり，コロナ渦で年齢による来訪パターンに違いがあった可能性がある。福岡県民に関してはペンギン水族館，佐世保駅，九十九島エリアなどの傾向が長崎駅エリアの傾向で説明されると推定された。関東エリアの居住者についても，長崎駅と平和公園の間で顕著な相互相関がみられたものの，福岡県民とは異なり，ペンギン水族館や九十九島エリアとは相互相関がみられず，また全体として相互相関がみられたエリアは少なかった。以上に加え，居住地によらず長崎市内の観光地間で比較的強い相互相関がみられた点や，福岡県民と関東エリアの住民で差がみられた点は興味深い。

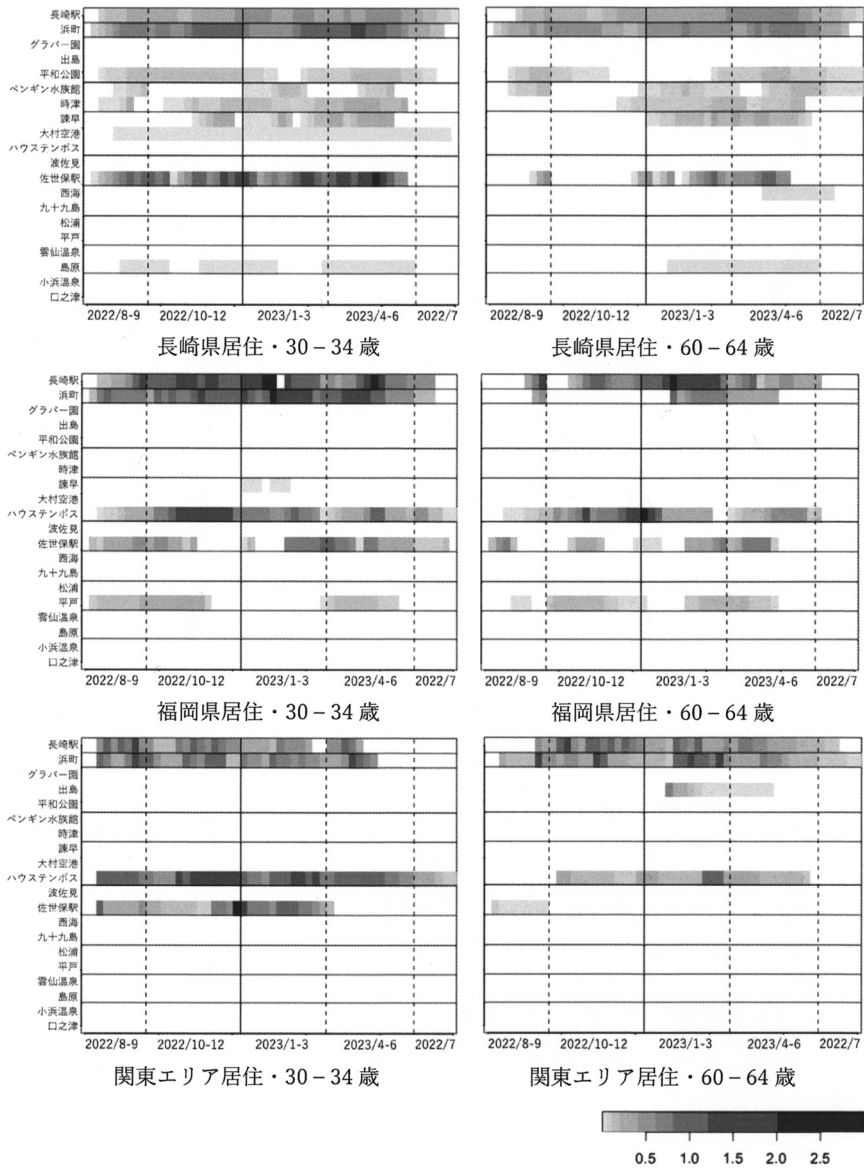


図 6. 他エリアへの影響力の推定結果.

5. まとめ

本稿は、主要産業の一つである観光を活性化するために、観光客誘致の根拠となるデータの創出を目的とした。そのため、携帯電話の位置登録情報を利用して、あるエリアでの来訪者数の増加が、同エリアおよび他のエリアのその後の来訪者数に及ぼす影響を推定する手法を提案し、それを基に、30-34 歳と 60-64 歳の年齢別かつ居住地別でエリア別来訪者数の相互依存関係を明らかにし観光地間の相互影響力について考察した。その結果、長崎県民は年齢別で見ると、コロナの影響による自粛が 60-64 歳のグループでは見られたが、30-34 歳グループには見

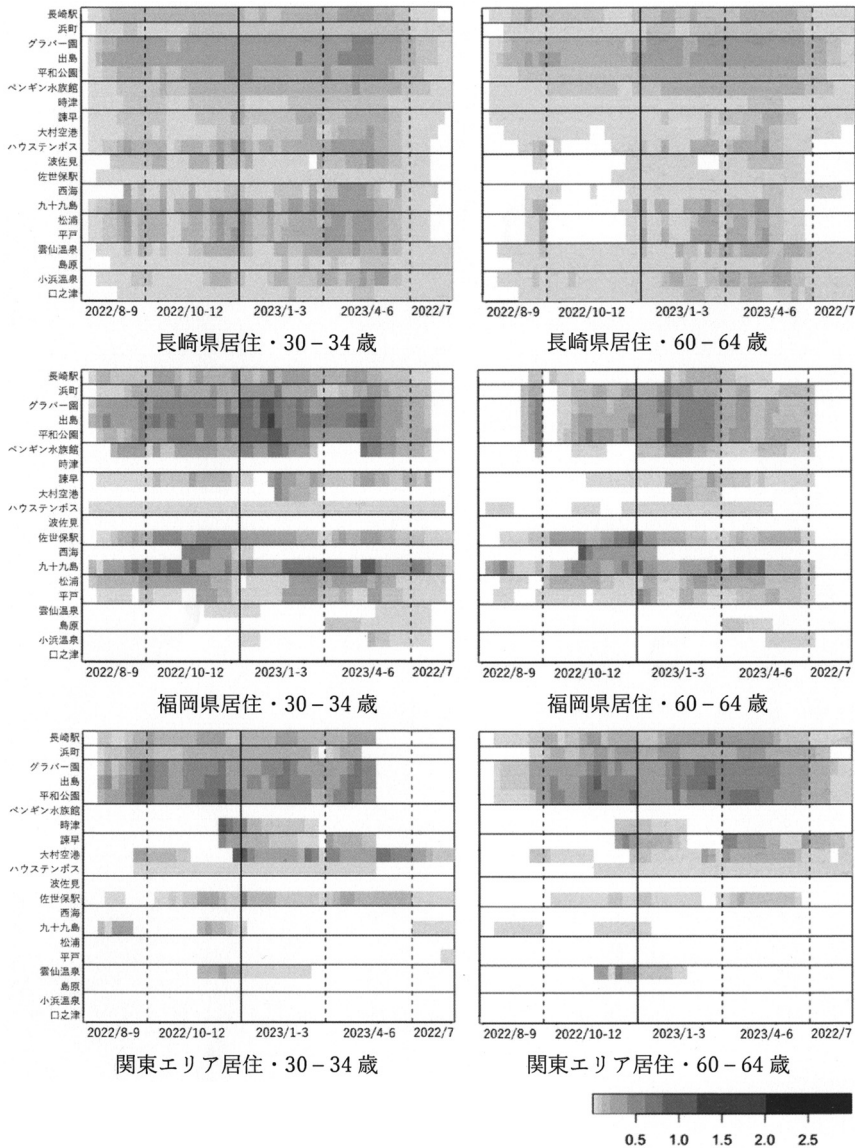


図 7. 他エリアから受ける影響力の推定結果。

られないなど年代でも行動が大きく変わることが示された。また、エリア別で比較すると、長崎県民と他県の住民の観光地の選択が異なっており、長崎県民は主要な観光地をはずし、県外からの観光客は主要観光地を目指して訪れる傾向があることが示された。さらに、他県からの観光客が訪れる先が主要観光地のため、特定の観光地に注目することで他のエリアの観光客の傾向を十分に推定できることが明らかとなった。これらを踏まえ、例えば、他県からの来訪者に対しては、マイナーな観光地の誘導が十分ではないため、情報発信を積極的にするなどし、主要観光地以外へ誘導するといった観光客誘致方法が効果的であるなどの複数の可能性が示された。

その一方で、今回の分析に利用したデータは、観光客とビジネス客を区別していない。よって、今後は観光政策の立案の根拠として役立てるために、観光客のみを対象とした集計データを使って観光地間の結びつきを追加評価することも検討したい。これに加え、日曜日だけでなく平日や時間帯別での評価方法の確立や、他県の観光地における観光客の行動との比較を行い各県の特徴を明らかにすること、観光客誘致のための応用方法についての検討も行いたい。今回は実態把握のために移動データのみで観光地間の相互相関関係を評価したが、より解釈性の高い分析を行って観光政策に役立てるためには観光地間の因果関係や観光客の移動理由を考慮してモデル化することが望ましい。時空間データに特化した因果推論法 (Reich et al., 2021) や、Directed acyclic graph を用いた柔軟な因果推論法 (Textor et al., 2016) なども提案されており、それらを応用した因果関係の特定も今後取り組みたい。また、分析期間を伸ばすことも重要な課題である。今回は1年分しかデータが得られなかったが、複数年のデータが得られれば、季節変動、過去から未来への長期的変動、パンデミックに伴う比較的短期的変動などが識別できる可能性がある。それらの識別も今後検討したい。

謝 辞

本研究は JSPS 科研費 JP23K21816 の助成を受けたものです。

参 考 文 献

- Aasa, A., Kamenjuk, P., Saluveer, E., Šimbera, J. and Raun, J. (2021). Spatial interpolation of mobile positioning data for population statistics, *Journal of Location based Services*, **15**(4), 239–260, <https://doi.org/10.1080/17489725.2021.1917710>.
- Agooop (2024). マチレポ, <https://agoop.co.jp/> (最終アクセス日 2024 年 6 月 1 日).
- Bringmann, L.F., Ferrer, E., Hamaker, E.L., Borsboom, D. and Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model, *Multivariate Behavioral Research*, **53**(3), 293–314.
- Casas, I. and Fernandez-Casal, R. (2019). tvReg: Time-varying coefficient linear regression for single and multi-equations in R, SSRN, <http://dx.doi.org/10.2139/ssrn.3363526>.
- Christopoulos, D.K. and León-Ledesma, M.A. (2008). Testing for Granger (non-)causality in a time-varying coefficient VAR model, *Journal of Forecasting*, **27**(4), 293–303.
- Cinarra Systems (2024). 位置情報マーケティング, <https://cinarra.co.jp/> (最終アクセス日 2024 年 6 月 1 日).
- D'Angelo, N., Abbruzzo, A., Ferrante, M., Adelfio, G. and Chiodi, M. (2023). GPS data on tourists: A spatial analysis on road networks, *ASTA Advances in Statistical Analysis*, **108**, 477–499, <https://doi.org/10.1007/s10182-023-00484-w>.
- 遠藤幹大, 高橋央亘, 浅田拓海, 有村幹治 (2019). Wi-Fi パケットセンシングによる道央広域エリアの時空間周遊パターン分析, 土木学会論文集 D3(土木計画学), **75**(5), 475–483.
- Freund, Y. and Schapire, R. E. (1996). Experiments with a new boosting algorithm, *Proceedings of International Conference on Machine Learning*, **96**, 148–156.
- Gao Y. and Schmocker, J-D. (2022). Distinguishing different types of city tourists through clustering and recursive logit models applied to Wi-Fi data, *Asian Transport Studies*, **8**, <https://doi.org/10.1016/j.eastsj.2021.100044>.
- GMO (2024). Town Wi-Fi, <https://townwifi.jp/> (最終アクセス日 2024 年 6 月 1 日).
- 神谷達夫 (2018). 位置情報データを活用した観光地指標—海の京都観光圏 Wi-Fi パケットセンサーの情報解析から—, 日本観光学会誌, **59**, 41–48.
- 観光庁 (2024). 観光地域づくり法人(DMO)とは, https://www.mlit.go.jp/kankoch/seisaku_seido/dmo/dmotoha.html (最終アクセス日 2024 年 6 月 1 日).

- Lee, D. and Seung, H. S. (2000). Algorithms for non-negative matrix factorization, *Advances in Neural Information Processing Systems*, **13**, 535–541.
- Liu, W., Wang, B., Yang Y., Mou, N., Zheng, Y., Zhang, L. and Yang, T. (2022). Cluster analysis of microscopic spatio-temporal patterns of tourists' movement behaviors in mountainous scenic areas using open GPS-trajectory data, *Tourism Management*, **93**, <https://doi.org/10.1016/j.tourman.2022.104614>.
- 村上大輔, 松井知子 (2022). 世代間・地域間の時系列相互相関に着目した COVID-19 の分析, 統計数理, **70**(1), 27–39.
- 中西航, 小林巴奈, 都留崇弘, 松本拓郎, 田中謙大, 菅芳樹, 神谷大介, 福田大輔 (2018). Wi-Fi パケットセンサーによる観光周遊パターンの把握可能性: 沖縄・本部半島における検討, 土木学会論文集 D3 (土木計画学), **74**(5), 787–797.
- Nakanishi, W., Yamaguchi, H. and Fukuda, D. (2018). Feature extraction of inter-region travel pattern using random matrix theory and mobile phone location data, *Transportation Research Procedia*, **34**, 115–122.
- Nicholson, W., Matteson, D. and Bien, J. (2017). Bigvar: Tools for modeling sparse high-dimensional multivariate time series, ArXiv, <https://doi.org/10.48550/arXiv.1702.07094>.
- NTT docomo (2024). モバイル空間統計, <https://mobaku.jp/> (最終アクセス日 2024 年 6 月 1 日).
- 大井達夫 (2020). 小地域統計を利用したインバウンド観光の季節変動分析, 観光学, **22**, 13–24.
- 大井達夫 (2022). タイル指数による観光地の人流データの変動要因分析—新型コロナウイルス感染症の初期段階と現在を比較して—, 観光振興研究, **2**(4), 29–41.
- Qian, C., Li, W., Duan, Z., Yang, D. and Ran, B. (2021). Using mobile phone data to determine spatial correlations between tourism facilities, *Journal of Transport Geography*, **92**, <https://doi.org/10.1016/j.jtrangeo.2021.103018>.
- Reich, B. J., Yang, S., Guan, Y., Giffin, A. B., Miller, M. J. and Rappold, A. (2021). A review of spatial causal inference methods for environmental and epidemiological applications, *International Statistical Review*, **89**(3), 605–634.
- SoftBank (2024). プライバシーセンター, <https://www.softbank.jp/privacy/> (最終アクセス日 2024 年 6 月 1 日).
- 寺部慎太郎, 一井啓介, 柳沼秀樹, 小野瑞樹, 田中皓介, 康楠 (2019). Wi-Fi パケットセンサーを用いた歩行者行動・観光客周遊行動研究の包括的レビューとそれを踏まえた分析例示, 土木学会論文集 D3 (土木計画学), **75**(5), 669–679.
- Textor, J., Van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M. and Ellison, G. T. (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty', *International Journal of Epidemiology*, **45**(6), 1887–1894.
- Yamaguchi, H. and Nakayama, S. (2023). Pattern analysis of Japanese long-distance travel change under the COVID-19 pandemic, *Transportation Research Part A: Policy and Practice*, **176**, <https://doi.org/10.1016/j.tra.2023.103805>.

Evaluation of Interactions among Tourist Spots in Nagasaki Using Location Data

Yu Ichifuji¹ and Daisuke Murakami²

¹Graduate School of Integrated Science and Technology, Nagasaki University

²The Institute of Statistical Mathematics

Tourism is a major industry in Japan, playing a crucial role in regional revitalization, especially in local cities. Despite being heavily impacted by the novel coronavirus, tourism is starting to recover. The number of tourists is returning to pre-pandemic levels or even surpassing them. To further attract tourists and manage overtourism, it will be necessary to implement strategies and measures. However, the increase in individual travel makes it challenging to analyze overall trends and movements between tourist destinations. To address this, the study aims to propose a method for evaluating relationships between tourist destinations using location registration data from mobile carriers. Through a quantitative assessment, the study seeks to create fundamental data to support tourist attraction efforts. Our proposed method evaluates the influence between tourist destinations using the number of unique users at each destination, the number of movements between destinations, and users' residential information. It is based on a time series model assuming that the influence of an increase in the number of visitors at each location diffuses proportionally to the number of movements estimated from the location registration data. When applied to tourist destinations in Nagasaki Prefecture, the evaluation of relationships between destinations revealed differences in behavior according to residence and age group. The findings suggest that modifying advertising methods based on residence and age group could efficiently attract tourists.