

令和 6 年 6 月 13 日現在

機関番号：62603

研究種目：若手研究

研究期間：2020～2023

課題番号：20K13261

研究課題名（和文）多様なデータから多様な効果を推定するための空間回帰モデリング

研究課題名（英文）Spatial regression modeling estimating a wide variety of effects from diverse data

研究代表者

村上 大輔（Murakami, Daisuke）

統計数理研究所・統計基盤数理研究系・准教授

研究者番号：20738249

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：(a)多様なデータから(b)多様な効果を推定するための時空間回帰手法の開発と、フリーの統計ソフトウェアRパッケージへの実装を行った。(a)では合成変換関数の導入により、データ分布を自動推定することができるように空間回帰を拡張した。同手法により、Box-Cox分布やTukey g-and-h分布を含む幅広い分布のデータが扱えることを確認した。(b)では、空間相関と複数時間軸上の時間・時空間相関を同時に扱える時空間回帰手法を新たに開発した。以上で開発した手法を、住宅地価、犯罪件数、Covid-19の陽性者数を含む幅広いデータに応用してその有用性を確認した。

研究成果の学術的意義や社会的意義

既存の空間統計手法の多くは、ガウス過程に依拠したものであり結果的に計算コストや柔軟性に課題が残されていた。本研究では、それらの課題に対処して、計算コストを維持しながら幅広いデータと効果を扱えるように空間統計手法を拡張するものであり、同分野の発展に寄与する学術的意義の大きい研究である。また開発手法をフリーの統計ソフトウェアRのパッケージを通して公開しており、地理情報に関する実務者・研究者を、手法提供の観点で支える社会的意義の大きな研究である。

研究成果の概要（英文）：This study developed spatio-temporal regression models for estimating (a) diverse effects from (b) diverse data, and implemented the developed methods in packages in a free statistical software R. Regarding (a), spatial regression was extended to allow automatic estimation of data distributions by introducing a compositional transformation function. It was confirmed that the developed method flexibly handle a wide range of data distributions, including Box-Cox and Tukey g-and-h distributions. Regarding (b), we developed a new spatio-temporal model that can handle spatial, temporal, and spatio-temporal correlations on multiple time axes simultaneously. The usefulness of the methods developed in (a) and (b) was confirmed by applying them to a wide range of real-world data, including residential land prices, the number of crimes, and the number of positive Covid-19 cases.

研究分野：空間統計学

キーワード：時空間回帰 非ガウスデータ 高速化 spmoran

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

空間相関(近接した場所同士が互いに強く影響しあう)や空間的異質性(場所毎に特性値やパラメータが異なる)を考慮する回帰である空間回帰は、生態学、環境科学、計量経済学、空間疫学などの幅広い分野で用いられている。特に多種多様な空間データ(位置情報を持つデータ)が利用可能な今日では、多様なデータから多様な効果を推定するための柔軟な空間回帰法が必要とされている。しかしながら、空間回帰法には、(a)多様なデータのモデリング、(b)多様な効果のモデリングの双方に以下のような課題が残されており、昨今の多様な空間データを柔軟に解析するための手法が十分に確立されてきたとは言いがたい。具体的な課題を以下に示す

(a) 多様なデータのモデリングの課題

センサ技術の発達に伴う(時)空間データの多様化・大規模化は著しい。しかし、標準的な空間回帰の計算量は N^3 (N はサンプルサイズ)のオーダーで指数的に増大するため大規模データの解析には向かない。また、大規模データのための空間回帰法も開発されているが、それらのほとんどはガウス分布に従うデータを仮定しており、様々な分布に従う大規模時空間データを計算効率よく扱う手法は発展途上にある。

(b) 多様な効果のモデリングの課題

空間データには、上述の空間効果以外にも様々な非空間効果(時間効果、時空間効果、グループ効果、非線形効果など)の影響が考えられる。しかし、空間回帰に関する研究では空間効果のみに焦点をあてることが多く、非空間効果も含む幅広い効果を捉えるような(時)空間回帰手法はいまだ発展途上にある。

2. 研究の目的

以上を踏まえ、本研究では多様・大規模な空間データから空間効果・非空間効果を推定・識別する方法の開発を行う。また、開発した手法をフリーの統計ソフトウェア R のパッケージとして整理して公開することで、地理情報に関係した幅広い分野の研究者・実務者を手法の観点で支援する。

3. 研究の方法

申請者がこれまでに開発を進めてきた空間加法モデルを拡張することで研究を進めてきた、以下、(a)と(b)に分けて説明する：

(a) 多様なデータのための時空間回帰の開発

上記の空間加法モデルと compositionally-warped Gaussian process を融合した。それにより、非線形変換を繰り返すことで、あらゆる連続確率分布に従う被説明変数が扱えるように空間加法モデルを拡張した。さらに、カウントデータに対する回帰モデルを近似するための非線形変換についても新規に導出することで、幅広いカウントデータに対しても、同手法が適用できるように拡張を行った。開発手法のパラメータ推定法は、大規模データへの応用を見据えて、サンプルサイズに対して線形以上のオーダーとならないような形で開発した。

(b) 多様な効果を捉える時空間回帰の開発

幅広い効果を扱うために、計算効率を維持しながら幅広い空間効果・非空間効果が選択できるように上記の空間加法モデルを拡張した。具体的には、サンプルサイズが増えなくても計算コストが増えないようにデータの内積をとったうえで、内積を用いて一連のモデル選択を行う方法を開発した。以上に加え、交互効果(特に時空間効果)を考慮する場合は、この方法を用いても計算コストが大きくなる場合があったため(内積の巨大化が原因)、reluctant interaction modeling のアイデアを踏襲して、非交互効果の選択を最初に行い、その後逐次的に交互効果をモデルに加えていくという選択法を開発した。同手法を用いることで、多様な効果の選択が大規模データから行えることを確認した。

以上、(a)で開発した手法は統計ソフトウェア R のパッケージ spmoran に実装し、(b)に関しても交互効果の対応部分以外は実装した。また、開発手法を住宅地価、犯罪件数、COVID-19 陽性者数などの幅広い実データに応用して、それらの有用性を検証した。

4. 研究成果

以下、(a)と(b)に分けて説明する：

(a) 多様なデータのための時空間回帰の開発

様々な尖度、歪度を持つ被説明変数データをシミュレーションで生成し、それらに対する予測誤差をみることで、提案手法が本当に幅広い分布に従うデータに適用できるかを確認した。下図は結果の一例である。ここでは、Tukey g-and-h (TGH) 分布を用いて上段に示すような分布を持つ被説明変数 2 パターンを生成し、それらに対する予測誤差を評価・比較した。結果は下段に示すとおりであり、下段(中央)の一番左(緑)が被説明変数がガウス分布に従う通常の場合に相当し、この場合であれば通常の間加法モデルでも予測誤差は小さい。一方で、同モデルの誤差は分布の裾が厚くなるにつれて増大していることも確認できる。対照的に、提案手法であれば、裾が厚くなっても精度があまり悪化せず、また裾が厚く、また歪んでいても精度が維持されることが下段(右)より確認できる。また、被説明変数を TGH 分布から生成したため、当然同分布を仮定したモデル(青)が誤差最小となっているが、提案手法は TGH 分布を明示的に仮定していないにもかかわらず、それに近い予測誤差になっていることが確認できる。

次に、同手法を犯罪件数データに応用した。その結果、例えば万引きは、山手線の内側エリアは、同じ地域で繰り返される傾向が強いことなどを確認した(図2)。

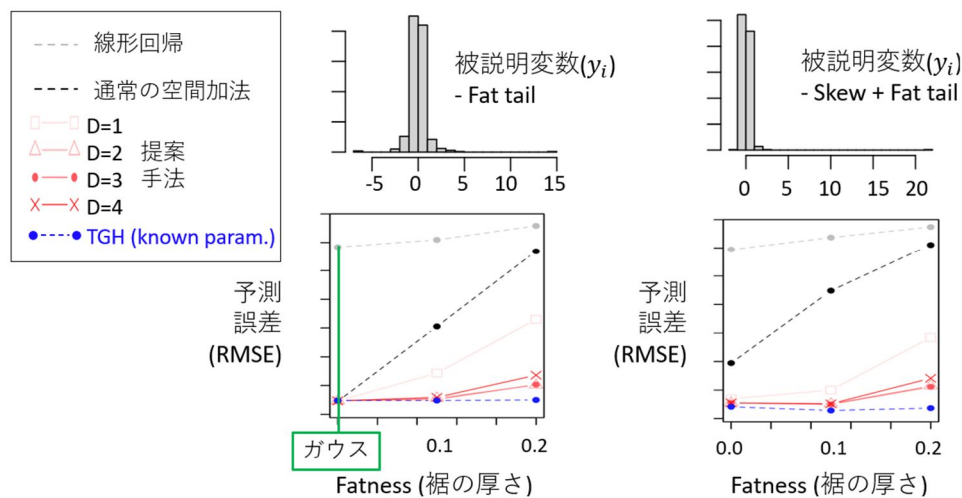


図1：非ガウスデータに対する予測誤差の比較 (Dは提案手法で被説明変数を変換する回数)

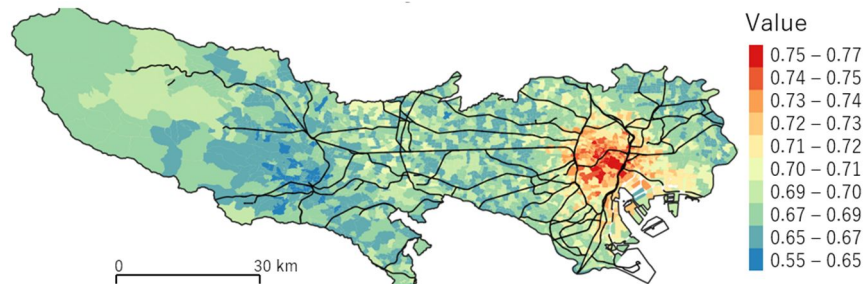


図2：万引き件数データの空間回帰への応用例 (近隣地域での反復傾向の推定結果)

(b) 多様な効果を捉える時空間回帰の開発

開発した空間・非空間効果の選択法がモデルの精度を改善させるかどうかを、シミュレーション実験で検証した。ここでは、回帰係数毎に空間、時間、時空間、時間（周期あり）、時空間（周期あり）の各パターンの有無を設定したうえで、提案手法による回帰係数の推定精度が、従来手法を上回るかどうかを検証した。結果の一部は図3に示すとおりである。図中の STC_{int} が提案手法であり、それ以外が既存手法である。この図から、提案手法の平均2乗平方根誤差（RMSE）が既存手法よりも小さくなり、精度が良好となることが確認された。なお、提案手法の計算時間が既存手法よりも短くなることも別途確認している。

次に、提案手法をサンフランシスコ市の窃盗件数データに応用した。例えば図4は結果の一例であり、図2と同様、近隣地域での犯罪の繰り返されやすさを推定した結果である。この結果から、コロナ渦の2020年4月に反復傾向が弱まるといった時間傾向や、対象地域北東部のベイサイドエリアで近接反復傾向が強まるといった空間傾向などが、感覚に合う形で推定された。

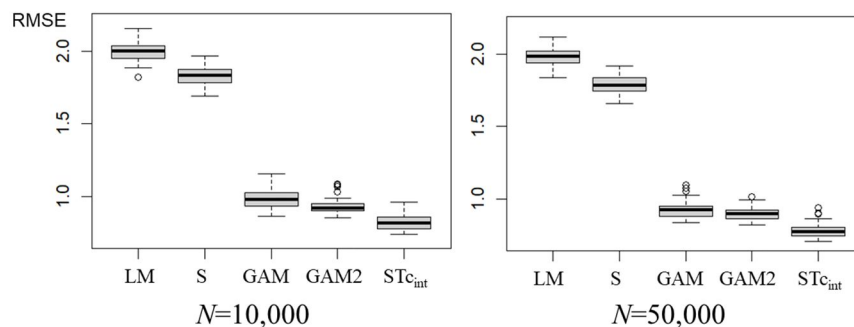


図3：提案手法の万引き件数データへの応用結果（近隣地域での反復傾向の推定結果）

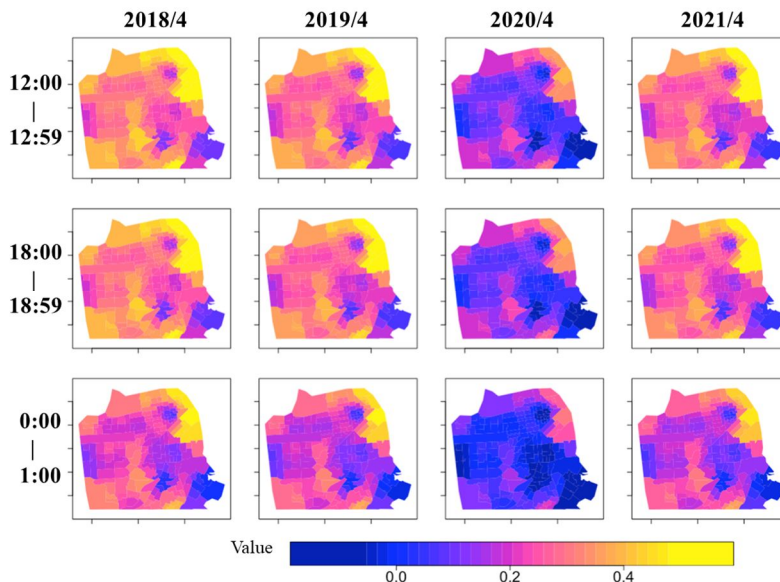


図4：提案手法の窃盗件数データへの応用結果（近隣地域での反復傾向の推定結果）

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 9件/うち国際共著 3件/うちオープンアクセス 5件）

1. 著者名 Murakami Daisuke, Tsutsumida Narumasa, Yoshida Takahiro, Nakaya Tomoki, Lu Binbin, Harris Paul	4. 巻 37
2. 論文標題 A linearization for stable and fast geographically weighted Poisson regression	5. 発行年 2023年
3. 雑誌名 International Journal of Geographical Information Science	6. 最初と最後の頁 1818 ~ 1839
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/13658816.2023.2209811	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Murakami Daisuke, Sugasawa Shonosuke, Seya Hajime, Griffith Daniel A.	4. 巻 1
2. 論文標題 Sub Model Aggregation for Scalable Eigenvector Spatial Filtering: Application to Spatially Varying Coefficient Modeling	5. 発行年 2024年
3. 雑誌名 Geographical Analysis	6. 最初と最後の頁 1
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/gean.12393	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Murakami Daisuke, Seya Hajime	4. 巻 4
2. 論文標題 Spatial Regression in the Presence of a Hierarchical Transportation Network: Application to Land Price Analysis	5. 発行年 2022年
3. 雑誌名 Frontiers in Sustainable Cities	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.3389/frsc.2022.905967	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yoshida Takahiro, Murakami Daisuke, Seya Hajime	4. 巻 -
2. 論文標題 Spatial Prediction of Apartment Rent using Regression-Based and Machine Learning-Based Approaches with a Large Dataset	5. 発行年 2022年
3. 雑誌名 The Journal of Real Estate Finance and Economics	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11146-022-09929-6	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Comber Alexis, Brunsdon Christopher, Charlton Martin, Dong Guanpeng, Harris Richard, Lu Binbin, Lu Yihe, Murakami Daisuke, Nakaya Tomoki, Wang Yunqiang, Harris Paul	4. 巻 55
2. 論文標題 A Route Map for Successful Applications of Geographically Weighted Regression	5. 発行年 2022年
3. 雑誌名 Geographical Analysis	6. 最初と最後の頁 155 ~ 178
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/gean.12316	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Murakami Daisuke, Griffith Daniel A.	4. 巻 NA
2. 論文標題 Balancing Spatial and Non Spatial Variation in Varying Coefficient Modeling: A Remedy for Spurious Correlation	5. 発行年 2021年
3. 雑誌名 Geographical Analysis	6. 最初と最後の頁 NA
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/gean.12310	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Murakami Daisuke, Matsui Tomoko	4. 巻 17
2. 論文標題 Improved log-Gaussian approximation for over-dispersed Poisson regression: Application to spatial analysis of COVID-19	5. 発行年 2022年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0260836
掲載論文のDOI (デジタルオブジェクト識別子) 10.1371/journal.pone.0260836	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Murakami Daisuke, Kajita Mami, Kajita Seiji, Matsui Tomoko	4. 巻 43
2. 論文標題 Compositionally-warped additive mixed modeling for a wide variety of non-Gaussian spatial data	5. 発行年 2021年
3. 雑誌名 Spatial Statistics	6. 最初と最後の頁 100520 ~ 100520
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.spasta.2021.100520	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 D. Murakami, M. Kajita, S. Kajita	4. 巻 9
2. 論文標題 Scalable model selection for spatial additive mixed modeling: application to crime analysis	5. 発行年 2020年
3. 雑誌名 ISPRS International Journal of Geo-Information	6. 最初と最後の頁 557
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/ijgi9100577	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

[学会発表] 計19件(うち招待講演 5件/うち国際学会 7件)

1. 発表者名 村上大輔, 堤田成政, 吉田崇紘, 中谷友樹
2. 発表標題 Sub-model aggregationによる地理的加重回帰の安定化・高速化
3. 学会等名 地理情報システム学会第32回研究発表大会
4. 発表年 2023年

1. 発表者名 Murakami, D., Sugasawa, S.
2. 発表標題 Sub-model aggregation for scalable Spatial spatially varying coefficient modeling
3. 学会等名 EcoSta (国際学会)
4. 発表年 2023年

1. 発表者名 村上大輔・菅澤翔之助
2. 発表標題 Sub-model aggregationによる空間可変パラメータモデルの高速化
3. 学会等名 2023年度統計関連学会連合大会
4. 発表年 2023年

1. 発表者名 Murakami, Daisuke, Tsutsumida Narumasa, Yoshida Takahiro, Nakaya Tomoki
2. 発表標題 Large-scale spatial prediction by scalable geographically weighted regression: Comparative study
3. 学会等名 The 15th International Conference on Spatial Information Theory (国際学会)
4. 発表年 2022年

1. 発表者名 Murakami Daisuke, Sugasawa Sugasawa
2. 発表標題 Sub-model Aggregation for Scalable Spatial Mixed Modeling
3. 学会等名 ISM Symposium on Environmental Statistics (国際学会)
4. 発表年 2023年

1. 発表者名 村上大輔、梶田真実
2. 発表標題 転移学習による空間予測の高精度化・犯罪データへの応用
3. 学会等名 地理情報システム学会
4. 発表年 2022年

1. 発表者名 村上大輔
2. 発表標題 マルチスケールな時空間データを活用したCOVID-19の統計モデリング
3. 学会等名 公開シンポジウム「COVID-19とデータ科学」
4. 発表年 2023年

1. 発表者名 Murakami Daisuke, Matsui Tomoko
2. 発表標題 Compositionally-warped additive mixed modeling for large non- Gaussian data: Application to COVID-19 analysis
3. 学会等名 The XV World Conference of Spatial Econometrics Association (国際学会)
4. 発表年 2021年

1. 発表者名 Murakami Daisuke, Tsutsumida Narumasa, Yoshida Takahiro, Nakaya Tomoki
2. 発表標題 Stable geographically weighted Poisson regression for count data
3. 学会等名 GIScience2021 (国際学会)
4. 発表年 2021年

1. 発表者名 村上大輔, 堤田成政, 吉田崇紘, 中谷友樹
2. 発表標題 疎なカウントデータのための地理的加重ポアソン回帰の安定化・高速化
3. 学会等名 地理情報システム学会第30回研究発表大会
4. 発表年 2021年

1. 発表者名 村上大輔
2. 発表標題 COVID-19流行の地理的要因の解明に向けた ポアソン回帰の高度化
3. 学会等名 都市経済学研究会 (招待講演)
4. 発表年 2021年

1. 発表者名 村上大輔
2. 発表標題 統計モデルで探るCOVID-19の地理的要因分析
3. 学会等名 立川商工会議所 第12回環境シンポジウム(招待講演)
4. 発表年 2022年

1. 発表者名 Murakami, D.
2. 発表標題 Compositionally-warped additive mixed modeling for large non- Gaussian data: Application to COVID-19 analysis
3. 学会等名 科研費シンポジウム "Recent Progress in Spatial and/or Spatio-temporal Data Analysis"
4. 発表年 2020年

1. 発表者名 Murakami, D.
2. 発表標題 Compositionally-warped additive mixed modeling: application to COVID19 data in Japan
3. 学会等名 ISM-Bristol Joint Seminar (国際学会)
4. 発表年 2020年

1. 発表者名 Murakami, D.
2. 発表標題 Spatial analysis of COVID18 spread using compositionally-warped Gaussian process
3. 学会等名 Global Collaboration on Data beyond Disciplines (国際学会)
4. 発表年 2020年

1. 発表者名 村上大輔
2. 発表標題 Compositionally-warped additive modelingによるCOVID19の地理的要因分析
3. 学会等名 2020年度日本保険・年金リスク学会（招待講演）
4. 発表年 2020年

1. 発表者名 村上大輔
2. 発表標題 COVID19の地理的要因の解明に向けた時空間加法モデリング
3. 学会等名 2020年度データ同化ワークショップ
4. 発表年 2021年

1. 発表者名 村上大輔
2. 発表標題 COVID-19流行の地理的要因の解明に向けた統計モデリング
3. 学会等名 情報・システム研究機構シンポジウム2020（招待講演）
4. 発表年 2021年

1. 発表者名 村上大輔
2. 発表標題 COVID-19流行の地理的要因の解明に向けた統計解析
3. 学会等名 位置情報ビジネスカンファレンス2021（招待講演）
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 村上 大輔	4. 発行年 2022年
2. 出版社 講談社	5. 総ページ数 272
3. 書名 実践Data Scienceシリーズ Rではじめる地理空間データの統計解析入門	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
米国	テキサス大学ダラス校		