

令和 3 年 5 月 5 日現在

機関番号：32704

研究種目：若手研究

研究期間：2018～2020

課題番号：18K12755

研究課題名(和文)大規模空間データに対する効率的統計解析について

研究課題名(英文)Efficient statistical analysis for large spatial datasets

研究代表者

平野 敏弘(Hirano, Toshihiro)

関東学院大学・経済学部・講師

研究者番号：10816010

交付決定額(研究期間全体)：(直接経費) 1,200,000円

研究成果の概要(和文)：本研究課題では、大規模な空間データに対する効率的統計解析手法の提案を目標として研究を実施し、主に以下の2点の成果を得た。

第一に、既存の効率的統計解析手法の解像度を増大させる形で拡張した新しい高速な推定・予測手法である「線形射影を用いた多重解像度近似」を提案した。第二に、多変量時空間データに対する時空間相関を考慮した効率的なベイズ的テンソル補完法を提案した。その他の空間統計学に関連する研究も行った。

研究成果の学術的意義や社会的意義

本研究課題において得られた2つの研究成果は、大規模空間データに対する新しい効率的統計解析手法を提案したものであり、そのうちの一部は統計学分野の国際学術誌に受理されている。そのため、本研究の成果は一定の学術的意義が認められたと言える。

また、ここであげた2つの研究成果は、空間データが非定常性を持つ場合や非ガウス性を持つ場合にもそれぞれ適用可能なので、例えば地価、交通量、気温・降水量の統計解析において柔軟に運用できる。これらに対する空間データ分析は、不動産市場の分析、交通渋滞の削減、気候変動問題といった応用と密接に結びついており、この点において本研究成果の社会的意義を見出すことができる。

研究成果の概要(英文)：In this program, I mainly conducted the following two topics of research in the efficient statistical analysis for large spatial datasets.

First, I proposed a new fast computational method for the estimation and prediction, which was referred to as multi-resolution approximation via linear projection. This proposed method was obtained by increasing the resolution of the existing efficient statistical method. Second, I considered the efficient Bayesian tensor completion for multivariate spatio-temporal data, which allows us to estimate the spatio-temporal parameters.

I also conducted other researches related to spatial statistics.

研究分野：空間統計学

キーワード：確率場 大規模空間データ 多重解像度近似 テンソル補完 多変量時空間データ データサイエンス  
並列計算

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

全球測位衛星システムなどの計測技術の発展により、センサーを通じて位置情報を伴う様々なデータが観測されている。たとえば、交通網では道路に設置されたセンサーから単位時間あたりの車両数を表す交通量が得られている。さらに、多くの地点で観測された日次最大温度・月次最小温度・月次総降水量といったデータも存在する。これらの都市・環境データは緯度・経度といった空間情報を伴って観測されており空間データと呼ばれている。空間統計学における統計解析手法を用いることで、空間データの相関構造の推定、未観測地点における空間データの予測や観測データのデノイズングを実行し、重要な意思決定につなげることができる。

### 2. 研究の目的

空間データに対して最尤法、クリギング、マルコフ連鎖モンテカルロ法といった統計解析手法を適用した場合、計算途中で現れる共分散行列の逆行列に関連する計算はサンプルサイズの3乗のオーダーの計算量となる。そのため、空間データが大量に得られた場合、計算負荷は極めて大きくなる。この問題に対処するため、本研究課題では大規模な空間データが得られた場合の効率的な統計解析手法の提案を目的とする。

### 3. 研究の方法

実際に観測される大規模空間データは非定常性、非ガウス性、もしくは時間情報を持つ場合が多い。多くの空間統計学の研究者によって大規模空間データの効率的な計算方法が活発に研究されているが、このような性質を持つ空間データに対して、先行研究で提案された手法を適用した場合の推定・予測精度はまだ改善の余地がある。本研究課題では、研究代表者が過去に提案した大規模空間データのための効率的統計解析手法を利用、もしくは拡張することで上記のような実際の空間データが持つ性質への効果的な対処ができるかどうかを調べていった。

### 4. 研究成果

本研究課題では、以下にあげる研究成果が得られた。

#### (1) 線形射影を用いた多重解像度近似の提案

ここでの研究成果は、「5. 主な発表論文等〔雑誌論文〕」における1つ目の論文に対応している。本研究では、Hirano (2017)で提案された modified linear projection の解像度を増大することによる推定・予測精度の改良を提案した。具体的には、正規確率場における共分散関数のパラメータの最尤推定量と未観測地点の予測分布を高速計算するアルゴリズムを開発した。さらに、提案アルゴリズムの時間計算量・空間計算量・並列計算方法も導出した。提案アルゴリズムについて、様々な解像度・空間相関・領域分割の設定の下でシミュレーションを行い、共分散関数の近似精度・未観測地点の予測精度・数値的安定性・スケーラビリティといった評価指標に基づいて最近の先行研究のアルゴリズムと比較した。結果として提案したアルゴリズムは、線形射影を使用している分の計算時間がかかるためスケーラビリティについては先行研究の方が若干優れていたが、それ以外の全指標については良好な性能を示していると分かった。また、空間線量率を用いた実データ分析において、提案したアルゴリズムは先行研究よりも短い計算時間で、高速計算していないオリジナルの最尤推定量・予測分布に近い推定・予測結果を導くことが示唆された。得られた研究成果について、国内学会・国際学会・国内研究会で発表した。また、論文としてまとめて投稿し、最終的に国際学術誌に受理された。さらに、受理された論文の今後の課題の1つに取り組んだ。提案アルゴリズムは共分散関数の近似精度が部分的に悪くなる場合があり、結果として、予測平面に不連続な境界を生じさせる可能性がある。この問題を回避するために、covariance tapering を解像度毎に適用することで改良を行い、一定の成果を得た。

#### (2) 多変量時空間データのためのベイズ的テンソル補完の提案

ここでの研究成果は、「5. 主な発表論文等〔雑誌論文〕」における2つ目の論文に対応している。本研究では、Bahadori et al. (2014)のように多変量時空間データを3次のテンソルとしてモデリングし、CANDECOMP/PARAFAC 分解可能であると仮定して、テンソルにおける未観測地点の値を効率的に予測する MCMC アルゴリズムを導出した。具体的には、多変量時空間データが正規分布、もしくは二項分布に従う場合のベイズ的テンソル補完を実行するギブスサンプリングを提案した。提案手法はポリア・ガンマデータ拡大を用いることで、二項分布の場合でもギブスサンプリングによる効率的なベイズ推定が可能となっている。また、空間相関・時間相関のパラメータや回帰係数もベイズ推定可能である。以上の研究結果について国内研究会で発表するとともに、論文としてまとめた。さらに、多変量時空間データがガンマ・ポアソン分布に従う場合のポリア・ガンマデータ拡大を用いたベイズ的テンソル補完の効率的アルゴリズムについて検討

を行い，一定の成果を得た．

( 3 ) 空間統計学の洋書の分担執筆への参加

ここでの研究成果は，「 5 . 主な発表論文等〔図書〕」に対応している．空間統計学の理論・応用について解説している洋書の執筆において，ガウス過程に関する章の執筆に参加した．

( 4 ) 大規模空間データに対する統計解析コンペティションでの優勝

ここでの研究成果は，「 5 . 主な発表論文等〔その他〕」に対応している．2021 KAUST Competition on Spatial Statistics for Large Datasets における Sub-competition 2b の部で優勝（同点 1 位）した．コンペティションの内容は，90 万の大規模非ガウス空間データを用いて 10 万地点での空間データを予測し，その予測精度を競うというものだった．研究代表者以外には 2 名の研究者と共同で出場している．また，コンペティションの結果は次の HP

<<https://cemse.kaust.edu.sa/stsds/2021-kaust-competition-spatial-statistics-large-datasets>>

に掲載されている．

<参考文献>

Hirano, T. (2017). Modified linear projection for large spatial data sets. Communications in Statistics - Simulation and Computation, 46, 870-889.

Bahadori, M. T., Yu, Q. R. and Liu, Y. (2014). Fast multivariate spatio-temporal analysis via low rank tensor learning. In Proceedings of the 27th International Conference on Neural Information Processing Systems, 3491-3499.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Hirano Toshihiro	4. 巻 -
2. 論文標題 A multi-resolution approximation via linear projection for large spatial datasets	5. 発行年 2020年
3. 雑誌名 Japanese Journal of Statistics and Data Science	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/s42081-020-00092-x	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 Hirano, T.	4. 巻 279
2. 論文標題 Efficient Bayesian tensor completion for multivariate spatio-temporal data	5. 発行年 2020年
3. 雑誌名 経済系	6. 最初と最後の頁 60-69
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 -

〔学会発表〕 計7件（うち招待講演 1件 / うち国際学会 1件）

1. 発表者名 平野 敏弘
2. 発表標題 A multi-resolution approximation via linear projection for large spatial datasets
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 平野 敏弘
2. 発表標題 大規模空間データに対する Linear Projection を用いた多重解像度近似
3. 学会等名 経営と法セミナー (招待講演)
4. 発表年 2019年

1. 発表者名 平野 敏弘
2. 発表標題 大規模空間データに対する Linear Projection を用いた多重解像度近似
3. 学会等名 CSIS DAYS 2019「全国共同利用研究発表大会」
4. 発表年 2019年

1. 発表者名 平野 敏弘
2. 発表標題 Linear Projectionによる多重解像度近似を用いた大規模空間データに対する予測について
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 Hirano, T.
2. 発表標題 A Multi-Resolution Approximation via Linear Projection for Large Spatial Datasets
3. 学会等名 Spatial Statistics 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 平野 敏弘
2. 発表標題 大規模空間データに対する Linear Projection を用いた多重解像度近似
3. 学会等名 2018年度統計関連学会連合大会
4. 発表年 2018年

1. 発表者名 平野 敏弘
2. 発表標題 多変量時空間データに対するベイズ的テンソル補完
3. 学会等名 テクノメトリクス研究会
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 Murakami, D., Yamagata, Y., and Hirano, T.	4. 発行年 2020年
2. 出版社 Academic Press	5. 総ページ数 56
3. 書名 Geostatistics and Gaussian process models. In Y. Yamagata and H. Seya (Eds.), Spatial Analysis Using Big Data: Methods and Urban Applications.	

〔産業財産権〕

〔その他〕

<p>2021 KAUST Competition on Spatial Statistics for Large DatasetsにおけるSub-competition 2bの部で優勝（同点1位）した。研究代表者以外には2名の研究者と共同で出場している。また、コンペティションの結果は次のHP  <a href="https://cense.kaust.edu.sa/stsds/2021-kaust-competition-spatial-statistics-large-datasets">https://cense.kaust.edu.sa/stsds/2021-kaust-competition-spatial-statistics-large-datasets</a>          に掲載されている。</p>
---

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
---------	---------