

令和 5 年 6 月 22 日現在

機関番号：34305

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K11710

研究課題名（和文）位相的データ解析による時空間情報の構造解析と可視化の新展開

研究課題名（英文）New development of structural analysis and visualization for spatio-temporal data based on topological data analysis

研究代表者

栗原 考次（Kurihara, Koji）

京都女子大学・宗教・文化研究所・教授

研究者番号：20170087

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究では、時空間データに対して同位相分類に基づくエシェロン階層的構造を利用し、以下のような位相的データ解析による時空間情報の構造解析と可視化に関する研究を行った。

1．空間疫学における地域に関する空間リスク指標の計算と可視化が可能なソフトウェアの開発、2．常に最新日が含まれる前向きホットスポット検出のための手法の開発とCOVID-19データへの適用、3．位相的データ解析の観点からエシェロン解析による同位相分類のアルゴリズム、4．エシェロン解析の基礎となる考え方やアルゴリズムも含めた研究成果を反映したテキストの出版およびソフトウェアの公開、に関する研究を集中的に行った。

研究成果の学術的意義や社会的意義

本研究では、常に最新日が含まれる前向きホットスポット検出に対してスポットの持つ形状の制限を排除しリスクの低い領域を含まない新たな手法の開発と日本でのCOVID-19データのホットスポットの発生場所と発生時期を特定した。また、本研究の研究成果も含めたエシェロン解析に関する世界初のテキストとして共立出版から「エシェロン解析」を出版するとともに、時空間データの階層構造の可視化やホットスポット検出が可能なソフトウェアを公開することにより、関連する多くの分野に多大の貢献がある。

研究成果の概要（英文）：In this study, we investigate the echelon hierarchical structure based on homotopic classification for spatio-temporal data and conducted research on structural analysis and visualization of spatio-temporal information through topological data analysis.

The research concentrated on: 1. Development of software for the calculation and visualization of regional spatial risk indicators in spatial epidemiology; 2. Development of methods for the detection of prospective hotspots that always include the latest date and their COVID-19 application to data; 3. An algorithm for echelon analysis in view of topological data analysis; 4. Publication of texts and release of software reflecting research results, including basic concepts and algorithms of echelon analysis.

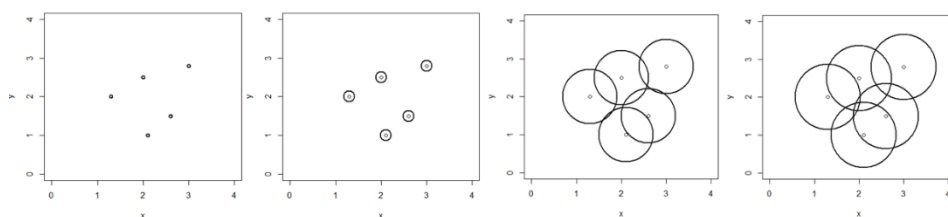
研究分野：データサイエンス

キーワード：エシェロン解析 時空間情報 可視化 位相的データ解析

1. 研究開始当初の背景

(1) 本研究の学術的背景、研究課題の核心をなす学術的「問い」

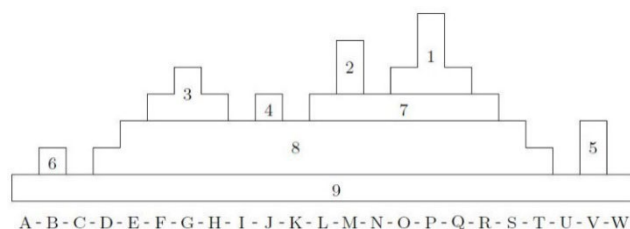
近年、Deep learningをはじめとする機械学習の技術が多くの分野で活用されている一方、こうした技術の基礎となっている統計的手法は、データが保有する情報を十分に活用されていないことが危惧されている。こうした環境下、位相的データ解析(Topological Data Analysis : TDA)が注目されている。位相的データ解析はデータの位相的・幾何的情報を抽出するための手法で、下図のようにデータの各点を中心とした円の半径を増加させた時の



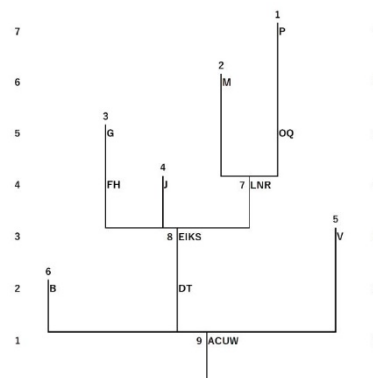
図形の穴に着目したパーシステントホモロジーを中心に研究開発されて

おり、アモルファスの構造解析、タンパク質の解析、ウイルスの遺伝的進化の解析、センサーネットワークの解析などの分野で成果を挙げている。

研究代表者は、海外共同研究者 Myers 博士が開発した定型格子データの構造を階層的に表現するエシエロン解析法を応用し、各種空間データに対して、観測値を縦軸に示した階層構造の可視化へ拡張している。さらに、空間データを階層構造化するアルゴリズム構築とその応用を公表し、空間データのエシエロン構造分析の基礎を築いている。1次元格子データは、下図のように高度の異なる地域 A から W からの地形の断面図として与えられ、その構造はエシエロンデンドログラムとして階層的に表現することができる。図の中に記されている数字は、位相的に構造が同じ領域に分類した時の番



号で、1から6はピーク、7から9はファウンデーションを示している。これまで、研究代表者を中心に各種の格子データに対して同位相分類の観点による階層構造化とその応用について研究を進めてきた。1次元空間データの同位相分類は視覚的にも容易に行うことができるが、2次元さらに3次元以上空間データの



位相的構造、特に、可視化の研究は困難性を伴う。また、エシエロン解析は、ソフトウェアが整備されていないこともあり、エシエロン開発グループが中心となり研究開発を進めてきた。従って、世界中で注目されているにもかかわらず、エシエロン技法の使用は限られていた。よって本研究では、これまでの研究成果に対する R および Shiny-server で開発されたソフトウェアをリリースするとともにエシエロン解析に関するテキストを出版する。こうした一連の研究により、格子型空間データに対するエシエロン解析は全世界で行われることが期待される。

2. 研究の目的

本研究では、時間的・空間的に相互作用を受けながら観測される時空間データに対して、

位相的データ解析の観点からエシェロン解析による同位相分類の考え方を拡張するとともに、位相的 Mapper 技法を医薬品テキスト及び化学記述子データに適用しデータの位相的構造を可視化する。また、感染症などのサーベイランス問題で応用されている前向きホットスポット検出および空間疫学における地域空間リスク指標の計算と可視化を行い、エシェロン解析の適用分野をさらに広げる。さらに、本研究の期間中に発生した 2019 年に発生した新型コロナウイルス感染症 (COVID-19) の時空間統計データ解析も行う。加えて、これまで研究を進めてきたエシェロン解析に関するテキストを出版するとともにソフトウェアを公開する。こうした研究は日本ではあまり活発に行われていない状況であり、本研究により国内外における時空間データを中心とした位相的データ解析に関する研究の発展を目指す。

3. 研究の方法

本研究では、時空間データに対して同位相分類に基づくエシェロン階層的構造を利用し、以下のような位相的データ解析による時空間情報の構造解析と可視化に関する研究を行う。

(1) 時空間データの位相的データ解析への展開

① 位相的データ解析の観点からエシェロン解析による同位相分類の考え方を拡張

2次元格子に対して、同位相を持つような順番に1次元格子に並び変え、視覚的に同じイメージで同位相分類を行う。たとえば、5×5の格子データに対して、同じ位相を持つ分類に基づき25個の1次元格子に並び替えることにより、視覚的に1次元格子データと同じイメージで取り扱うことが可能となる。

② 位相的 Mapper 技法を医薬品テキスト及び化学記述子データに適用しデータの位相的構造を可視化

医薬品テキストデータ及び化学記述子データに対して、位相的データ解析 Mapper 技法を使用して、定量的および定性的データの統合された主成分スコアに基づいてデータの視覚化を行い、薬剤のクラスターで構成されるネットワークを求める。こうした研究はドラッグリポジショニングにつながる。

(2) エシェロン解析の適用分野の展開

① 時空間データにおける真のクラスター (ホットスポット) に近いクラスターの検出法

時空間データにおいて、より真のクラスターに近いクラスターの検出を可能にする新たな手法の提案を行う。すなわち、検出されるクラスターの持つ形状の制限を排除しリスクの低い領域を含まない手法として、データの持つ位相的な空間的階層構造に基づき上位階層の領域から構成されるクラスターを求める。

② 空間疫学における地域に関する空間リスク指標の計算と可視化が可能なソフトウェアの開発

疾病の罹患や死亡の危険度を表すリスクの指標を計算するとともに、それらの結果を地図上に可視化できるソフトウェアを R の shiny パッケージを用いて開発する。本パッケージの使用により、対話的な操作を行いながら、リスク指標の分布の様子を空間的に捉えることが可能となる。

(3) COVID-19 の時空間統計データ解析

本研究は研究開始時点では想定してなかったが、本研究の期間中の 2019 年に発生した新型コロナウイルス感染症 (COVID-19) の時空間統計データ解析は喫緊の課題であり、新たな研究課題として行う。

① COVID-19 発生から各国への初感染が観測されるまでの空間的拡大の統計的分析

COVID-19 の初感染が観測されるまでの時間に影響を与える要因を理解することは、次回パンデミックが起きた際に各国が初期段階で取るべき行動指針を考えるの一助となることが期待される。本研究では、初感染までの時間を複数の社会経済的要因の関数として表現した 2 成分混合回帰モデルを提案する。

② 常に最新日が含まれる前向きホットスポット検出のための手法の開発と COVID-19 データへの適用

常に最新日が含まれる前向きホットスポット検出に対して、ホットスポットの持つ形状の制限を排除しリスクの低い領域を含まない新たな手法の開発と日本での COVID-19 データのホットスポットの発生場所と発生時期を特定する。

(4) エシェロン解析に関するテキストを出版およびソフトウェアの公開

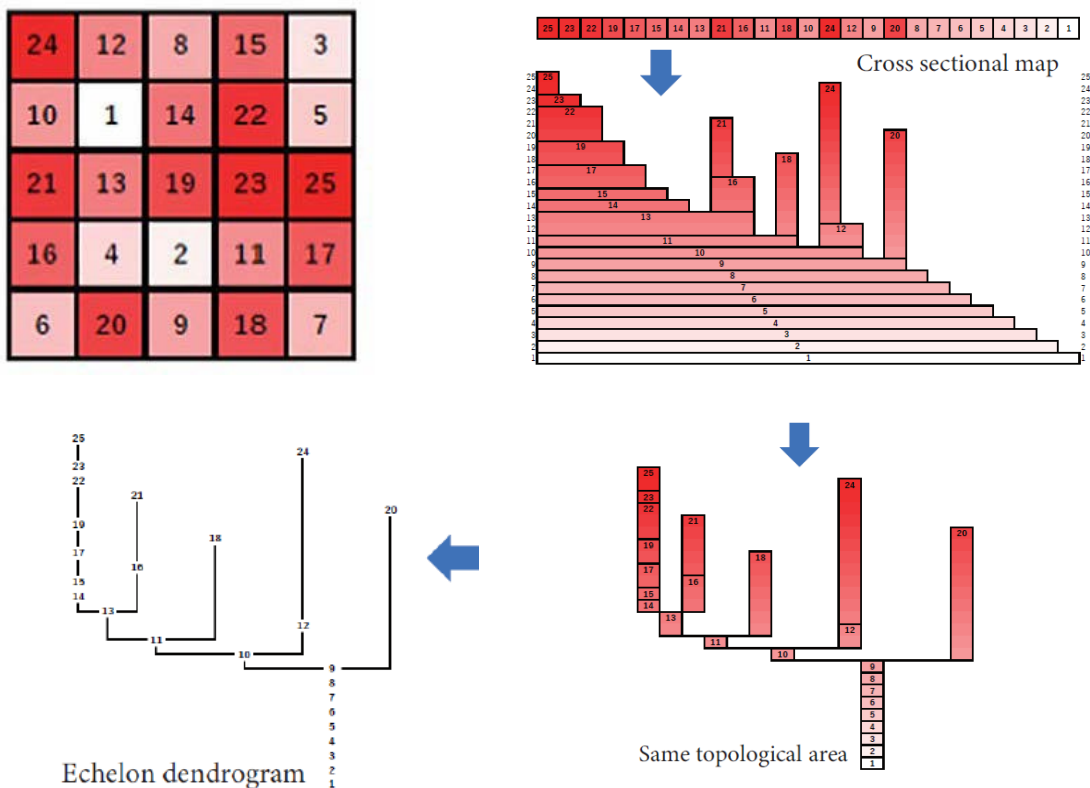
本研究の研究成果も含めたテキストとして共立出版から統計学 One Point シリーズ第 19 巻「エシェロン解析」を発刊する。また、時空間データの階層構造の可視化やホットスポット検出が可能なソフトウェアとして R の shiny パッケージを公開する。

4. 研究成果

(1) 時空間データの位相的データ解析への展開

① 位相的データ解析の観点からエシェロン解析による同位相分類の考え方を拡張

5 × 5 の 2 次元規則的格子データは、左上図のように格子内で観測された計量値として与えられる。この 2 次元格子に対して、右上図の同位相を持つような順番に 1 次元格子に並び替え、1 次元の場合と視覚的に同じイメージで位相的データ解析を行う。



これらの研究成果は 2 つの国際誌 Japanese Journal of Statistics and Data Science, Vol.3 No.1, WIREs Computational Statistics Vol.15 などにおいて公表した。

② 位相的 Mapper 技法を医薬品テキスト及び化学記述子データに適用しデータの位相的構造を可視化

医薬品情報がテキストデータとして記述されている Wikipedia 及び医薬品情報が化学記述子データの化合物情報として格納されている DrugBank の 2 種類のデータを定量化し、位相的 Mapper を適用する方法を提案した。提案方法により、適応疾患に関連する連続的な形状として捉えることができ、分類結果から様々な解釈及びその分類結果に基づくドラッグリポジショニングが可能となる。研究成果は *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer において公表した。

(2) エシェロン解析の適用分野の展開

① 時空間データにおける真のクラスター（ホットスポット）に近いクラスターの検出法

時空間データにおいて、クラスターの持つ形状の制限を排除しリスクの低い領域を含まずより真のクラスターに近いクラスターの検出するために、データの持つ位相的な空間的階層構造に基づき上位階層の領域から構成されるクラスターを求めた。研究成果は日本計算機統計学会の和文誌「計算機統計学」の第 34 巻第 1 号などにおいて公表した。

② 空間疫学における地域に関する空間リスク指標の計算と可視化が可能なソフトウェアの開発

空間疫学で用いられる疾病の罹患や死亡の危険度を表すリスクの指標を計算するとともに、それらの結果を対話型で操作しながら地図上に可視化できるソフトウェアを R の shiny パッケージを用いて開発・公開した。これらの研究結果は、日本計算機統計学会の和文誌「計算機統計学」の第 33 巻第 2 号において公表した。

(3) COVID-19 の時空間統計データ解析

① COVID-19 発生から各国への初感染が観測されるまでの空間的拡大の統計的分析

初感染までの時間を複数の社会経済的要因の関数として表現した 2 成分混合回帰モデルによる分析の結果、感染初期は飛行機での人の移動が原因となり中国から距離の近いアジア国々やヨーロッパやアメリカなどの先進国に感染が広がっていったことが示唆された。また、感染中期以降は、アフリカや南米などの人口が多い国に感染が広がっていったことが示された。研究結果は、日本計算機統計学会の和文誌「計算機統計学」の第 35 巻第 1 号において公表した。

② 常に最新日が含まれる前向きホットスポット検出のための手法の開発と COVID-19 データへの適用

日本国内での都道府県別 1 日単位の総人口に対する COVID-19 感染者の比率を表した大規模格子型時空間データに対して、計算コストを軽減するために制限付きエシェロンスキューン法を提案し、時空間上でのホットスポットの時系列変化を求めた。研究成果は国際誌 *Japanese Journal of Statistics and Data Science*, Vol.5 No.1、および国際学会 17th Conference of the International Federation of Classification Societies での招待講演、国際学会 The European Conference on Data Analysis 2022 での講演などにおいて公表した。

(4) エシェロン解析に関するテキストを出版およびソフトウェアの公開

本研究の研究成果も含めたエシェロン解析に関する世界初のテキストとして共立出版から統計学 One Point シリーズ第 19 巻「エシェロン解析」を発刊した。テキストでは、エシェロン解析の基礎となる考え方やアルゴリズム、さらに、エシェロンの構造を利用した応用例として、圏による地域の分類、リモートセンシングデータの分析、ホットスポットの検出、そして、エシェロン解析のためのソフトウェアを取り上げている。また、時空間データの階層構造の可視化やホットスポット検出が可能なソフトウェアとして R の shiny パッケージの公開することにより、関連する多くの分野に多大の貢献を行った。

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 0件/うちオープンアクセス 7件）

1. 著者名 梶西 将司、石岡 文生、栗原 考次	4. 巻 35
2. 論文標題 エシェロンデンドログラムを用いた空間データの構造分析とその評価	5. 発行年 2022年
3. 雑誌名 計算機統計学	6. 最初と最後の頁 17-35
掲載論文のDOI（デジタルオブジェクト識別子） 10.20551/jscswabun.35.1_17	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 高橋 正憲、羅 明振、栗原 考次	4. 巻 35
2. 論文標題 2成分混合回帰モデルを用いた COVID-19 初感染までの時間に影響を与える要因の分析	5. 発行年 2022年
3. 雑誌名 計算機統計学	6. 最初と最後の頁 1~16
掲載論文のDOI（デジタルオブジェクト識別子） 10.20551/jscswabun.35.1_1	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 Kurihara Koji、Ishioka Fumio	4. 巻 15
2. 論文標題 Echelon analysis and its software for spatial lattice data	5. 発行年 2022年
3. 雑誌名 WIREs Computational Statistics	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） 10.1002/wics.1579	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 Takemura Yusuke、Ishioka Fumio、Kurihara Koji	4. 巻 5
2. 論文標題 Detection of space?time clusters using a topological hierarchy for geospatial data on COVID-19 in Japan	5. 発行年 2022年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 279~301
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s42081-022-00159-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 FUJIWARA Mika、KAJINISHI Shoji、KURIHARA Koji	4. 巻 10
2. 論文標題 VISUALIZATION OF MULTIVARIATE DATA USING EXPANDED CONSTELLATION AND EXPANDED KANJI GRAPHS AND THEIR APPLICATION TO CLUSTERING	5. 発行年 2021年
3. 雑誌名 Journal of Environmental Science for Sustainable Society	6. 最初と最後の頁 1~8
掲載論文のDOI (デジタルオブジェクト識別子) 10.3107/jesss.10.1	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 梶西 将司、石岡 文生、栗原 考次	4. 巻 33
2. 論文標題 空間疫学におけるリスク評価とその可視化のためのソフトウェア	5. 発行年 2021年
3. 雑誌名 計算機統計学	6. 最初と最後の頁 91~105
掲載論文のDOI (デジタルオブジェクト識別子) 10.20551/jscswabun.33.2_91	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 竹村 祐亮、石岡 文生、栗原 考次	4. 巻 34
2. 論文標題 Echelon scan法による高リスクな空間クラスター検出法の提案	5. 発行年 2021年
3. 雑誌名 計算機統計学	6. 最初と最後の頁 23~43
掲載論文のDOI (デジタルオブジェクト識別子) 10.20551/jscswabun.34.1_23	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 Kurihara Koji、Ishioka Fumio、Kajinishi Shoji	4. 巻 3
2. 論文標題 Spatial and temporal clustering based on the echelon scan technique and software analysis	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 313~332
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-020-00072-1	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 Kitanishi Yoshitake, Ishioka Fumio, Iizuka Masaya, Kurihara Koji	4. 巻 1
2. 論文標題 Spatial Perception for Structured and Unstructured Data In topological Data Analysis	5. 発行年 2021年
3. 雑誌名 Studies in Classification, Data Analysis, and Knowledge Organization	6. 最初と最後の頁 103 ~ 111
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-60104-1_12	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13件 (うち招待講演 1件 / うち国際学会 2件)

1. 発表者名 竹村祐亮、石岡文生、宿久洋、栗原考次
2. 発表標題 国内の COVID-19 陽性者に対する echelon 法に基づいた時空間クラスターの最近の動向について
3. 学会等名 日本分類学会第 41 回大会
4. 発表年 2022年

1. 発表者名 Ishioka, F., Kajinishi, S., Kurihara, K.
2. 発表標題 Hotspot Cluster Detection based on Spatial Hierarchical Structure and Its Software
3. 学会等名 17th Conference of the International Federation of Classification Societies, Porto, Portugal. (招待講演) (国際学会)
4. 発表年 2022年

1. 発表者名 林邦好、吉安展将、栗原考次
2. 発表標題 医療画像に対する統計的アプローチと疾患予測
3. 学会等名 2022年度統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 Takemura, Y. Ishioka, F., Yadohisa, H., Kurihara, K.
2. 発表標題 Detection of alive clusters with high-risk by using echelon hierarchical structure
3. 学会等名 The European Conference on Data Analysis 2022, Naples. (国際学会)
4. 発表年 2022年

1. 発表者名 梶西将司、石岡文生、栗原考次
2. 発表標題 パターン化されたEchelon Treeを用いた空間構造の評価
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 樋口 史、羅 明振、栗原考次
2. 発表標題 時間依存型共変量を考慮した Cox 比例ハザードモデルによるプロ野球選手登録期間の分析
3. 学会等名 2021年度日本分類学会シンポジウム
4. 発表年 2021年

1. 発表者名 三箇健太、石岡文生、栗原考次
2. 発表標題 空間補間法を利用した岡山県の確率雨量の可視化について
3. 学会等名 2021年度日本分類学会シンポジウム
4. 発表年 2021年

1. 発表者名 神原あん、竹村祐亮、石岡文生、栗原考次
2. 発表標題 Prospective echelon scan による時空間クラスター検出手法の提案
3. 学会等名 2021年度日本分類学会シンポジウム
4. 発表年 2021年

1. 発表者名 對尾卓馬、栗原考次
2. 発表標題 順序ロジットモデルを用いた仕事満足度に影響を与える要因の分析
3. 学会等名 第41回大規模データ科学に関する研究会
4. 発表年 2021年

1. 発表者名 竹村祐亮、神原あん、石岡文生、栗原考次
2. 発表標題 Echelon階層構造に基づくprospectiveな時空間クラスター検出法について
3. 学会等名 第41回大規模データ科学に関する研究会
4. 発表年 2021年

1. 発表者名 竹村祐亮、石岡文生、栗原考次
2. 発表標題 エシェロン階層構造に基づくクラスター検出法の改善について
3. 学会等名 日本計算機統計学会第34回シンポジウム
4. 発表年 2020年

1. 発表者名 竹村祐亮, 石岡文生, 栗原考次
2. 発表標題 エシエロン法によるCOVID-19の時空間クラスターについての考察
3. 学会等名 2020年度日本分類学会シンポジウム
4. 発表年 2021年

1. 発表者名 高橋正憲, 羅 明振, 栗原考次
2. 発表標題 2成分混合回帰モデルを用いたCOVID-19初感染までの時間に影響を与える要因の分析
3. 学会等名 2020年度日本分類学会シンポジウム
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 栗原 考次、石岡 文生	4. 発行年 2021年
2. 出版社 共立出版	5. 総ページ数 144
3. 書名 エシエロン解析	

〔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