

令和 2 年 6 月 2 日現在

機関番号：15301

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K00050

研究課題名（和文）エシェロン樹木構造による時空間情報の可視化と構造分析の新展開

研究課題名（英文）New Development of Structural Analysis for Spatio-temporal Data based on Echelon Structure

研究代表者

栗原 考次（Kurihara, Koji）

岡山大学・環境生命科学研究所・教授

研究者番号：20170087

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本研究では、マルコフ確率場における各種の格子型データに対して、エシェロン樹木構造パターンに基づき可視化を行うとともに、空間クラスター（ホットスポット）や複雑性、さらに、信頼性・安定性を評価する新手法を提案した。小規模格子型データについては、従来モンテカルロ法により近似的に計算されていた統計量の確率分布を直接計算により正確に計算した。また、応用例として福島第一原子力発電所近隣の空間線量率ポイントデータに対して、クリギング手法を用いて未観測値の補間を行うとともに、時空間的観点からホットスポット検出を行った。また、各分野が直面する問題解明のため、本研究成果のソフトウェア化を行うとともに一般公開した。

研究成果の学術的意義や社会的意義

本研究は、小規模格子型データに対するホットスポットや複雑性の評価法の新たな展開として、安易に漸近的・近似的な理論やモンテカルロ法等の計算機集約法に依存せず、ノンパラメトリック接近法により直接計算、分割表空間への拡張を図るものである。また、時空間的観点からエシェロン手法を用いた空間クラスター（ホットスポット）を検出する手法を提供している。ホットスポットの検出は、環境の現状を把握するとともに、将来の環境や健康への影響を早期に発見するためにも重要である。さらに、ソフトウェア公開により、時空間データの階層構造の可視化が可能なエシェロン解析の利用が促進され、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 proposed a new method to evaluate hotspots and complexity based on the pattern of the Echelon tree structure for a model of lattice type data in Markov random field. For small-scale lattice data, the probability distribution of statistics, which has been conventionally calculated approximately by the Monte Carlo method, was accurately calculated by direct calculation.

In addition, unobserved values were interpolated using the kriging method and hotspots were detected from the spatio-temporal point of view for the air dose rate point data in the vicinity of the Fukushima Daiichi Nuclear Power Station. We have developed and released the software of research results.

研究分野：統計科学関連

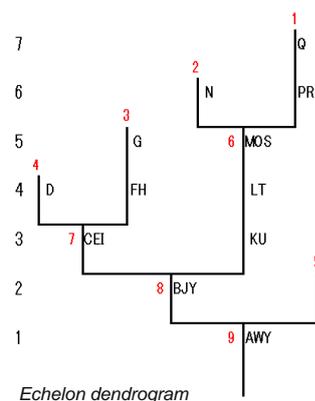
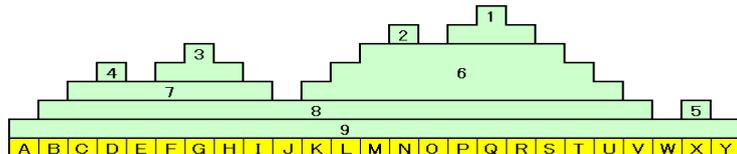
キーワード：エシェロン解析 時空間データ ホットスポット 空間スキャン統計量 可視化

様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

各種の分野で得られる多様な情報に対して、データの可視化やデータに含まれる不確定要素やばらつきを客観的かつ科学的に分析し、データに内存するメカニズムを的確に表現するモデルの構築や理論的な性質を調べるのが重要である。特に、多くの分野のデータは、時間・空間的位置情報をともなって得られることが多く、データが得られた分野の専門知識を最大限に活用し、これらのメカニズムを解明していくことが中心的課題となる。

海外共同研究者の Myers 博士は、定型格子データに対して同位相分類を行い樹木構造で可視化するエシェロン解析法(1997)を提案した。申請者は、この考え方を応用し、州や都道府県別に得られた地域データを含めた各種格子データに対して、下図のように観測値を縦軸に示したマーク付き階層構造の可視化(Kurihara et al., 2002)へ拡張するとともに、空間データを同位相分類により階層構造化するアルゴリズム構築とその応用(Kurihara, 2004)を公表し、空間データのエシェロン構造分析の基礎を築いている。



その後、申請者や共同研究者を中心に、アメリカ、イタリア、韓国、中国、インドネシアなどの国で、時空間データのエシェロン樹木構造を利用したホットスポット検出及び環境、生態、医学、公衆衛生などの分野への応用に関する最先端の研究が推進されてきた。研究成果は多岐にわたり、リモートセンシングデータに基づく植生の生態分布と人口密度の関連(Kurihara

et al., 2000)、空間スキャン統計量とエシェロン解析を利用した空間データのホットスポット検出(Kurihara, 2003)、エシェロン構造と尤度に基づくホットスポット検出(Myers et al., 2006)、3次元時空間データのホットスポット検出と環境汚染データの影響評価(Ishioka et al., 2010)、エシェロン階層構造を利用したゲノムデータの LD ブロック同定(Tomita et al., 2007)、韓国地震データのホットスポット検出(Han et al., 2008)、エシェロン階層構造を利用した森林の分類(Oda et al., 2012)、自殺データのホットスポット検出(Kubota et al., 2013)、多次元時空間データの分類(Kurihara et al., 2014)等の研究成果が公表されている。

近年、申請者らは、小規模格子型データである全国の都道府県のホットスポット検出の際必要となる都道府県の組合せからなる隣接ブロックを網羅的に数え上げに初めて成功し、組合せ隣接ブロックの総数が 1098 億 5712 万 5702 通りであること公表している (Kawahara et al, 2016)。こうした状況下、各種の統計量に対して、漸近的・近似的な理論に依存せず、直接確率分布を計算することによる理論と応用の新しい研究が期待されている。

### 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情報が観測されるマルコフ確率場において、空間的相互作用を明確にした各種の格子型データに対して、マーク付きエシェロン樹木構造に基づき可視化を行うとともに、集積性 (ホットスポット) や複雑性、さらに、信頼性・安定性を評価する新手法を構築する。特に、小規模格子型データについては、樹木構造のパターンに対して ノンパラメトリック接近法により各種統計量の確率分布を直接計算することにより、新たに複雑性や信頼性・安定性を評価することを目的としている。さらに、本研究成果のソフトウェア化を行うとともに一般公開し、各分野が直面する問題の解明を行い、その発展に寄与する。

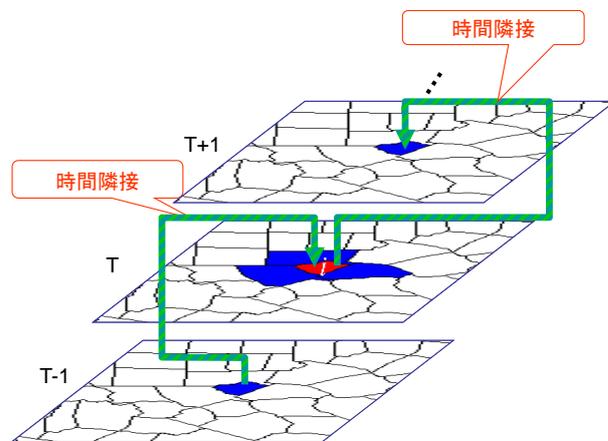
### 3. 研究の方法

本研究では、マルコフ確率場で得られた平面内の格子上的位置  $x$  に確率変数  $W(x)$  を与えた確率場を考え、その実現値として得られる大規模及び小規模格子型空間データに対して、エシェロン樹木構造に基づく時空間データの構造解析の新展開として、以下のような課題の解決を推進する。

#### (1) 格子型空間データに対する構造の単純性・複雑性の評価

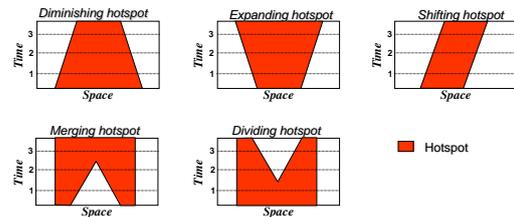
本研究では、空間の規模に応じたエシェロン樹木構造の単純性・複雑性の評価を行う。すなわち、小規模格子データに対して、エシェロン樹木を利用したノンパラメトリック接近法について考察する。具体的には、エシェロン樹木の階層やピークの数や階層のレベル数などの確率分布を直接計算により求め、空間格子データの単純性・複雑性に関する構造分布について検討する。

#### (2) 格子型空間データに対するホットスポット検出に関する研究

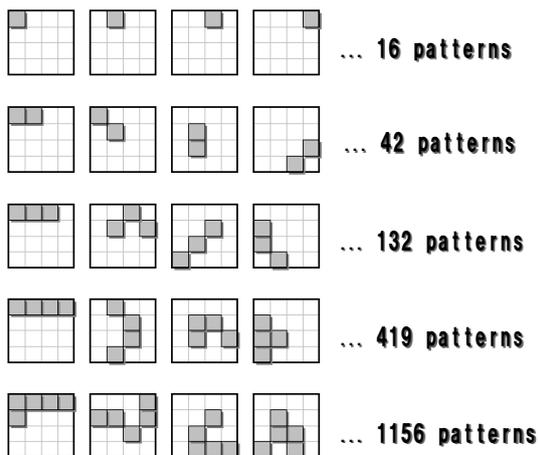


を調べる。この改良により、ホットスポットの時間的な拡大、縮小、移動、合併などの時間的な遷移（右図参照）の検出が可能となる。

本研究では、横断的に得られる格子型空間クロスセクションデータに対して、時間及び空間上でホットスポットを検出するために、同時間における空間隣接に加え、左図のように異なる時間における時間隣接を定義し、エシェロン情報に基づき時空間に渡る階層構造を求める。その階層構造に基づきホットスポットの時間的な遷移



#### (3) 統計量の理論的な確率分布



すなわち、本研究では、安易に計算機支援によるモンテカル法に頼ることなく、統計量の理論的な確率分布を求めるアルゴリズムを開発する。

集積性のある地域（ホットスポット）を検出するために用いられる空間スキャン統計量の計算は、データの間複雑な従属性があるために、統計量の理論分布を直接求めるのは困難とされ、モンテカル法による疑似計算が主に用いられてきた。しかし、左図のような  $4 \times 4$  の定型格子型空間データや都道府県別地域データなどのように格子数が少ない場合は、隣接するブロックのパターン数を数え上げ、統計量の確率分布を直接計算することが必要である。

#### (4) ソフトウェアの開発と公開

これまで、申請者は、アメリカを始め海外の研究者と共同でエシェロン樹木解析による空間構

造の可視化とその応用について研究を進めてきた。しかし、ソフトウェアが公開されていないため、その使用は一部の研究者に限られ、一般には普及されているとは言い難い。そこで、本研究で開発した手法のソフトウェアをフリーの統計処理言語Rを用いて開発し、自由にダウンロードできるように公開する。

#### 4. 研究成果

##### (1) 格子型空間データに対する構造の単純性・複雑性の評価

本研究では、1次元及び2次元の規則的な小規模格子データに対して、ノンパラメトリック接近法により、エシェロン樹木の階層やピークの数、階層のレベル数などの確率分布を直接計算により求めた。また、これらの信頼性・安定性を評価する手法についても検討した。具体的には、以下のような研究成果が得られた。

- ① 1次元10セルの場合の特性値として、エシェロン樹木は3,628,800通り存在し、8つのエシェロン樹木のパターンに分類されること、さらに、エシェロン樹木の階層やピークの数や階層のレベル数などの確率分布を直接計算により求めた。これらの結果は、2017年8月に東京で開催された国際分類学会において招待講演として公表した。
- ② 2次元5×5格子データにより構成されるエシェロン樹木に対して、その樹木構造の複雑さを計量化した指標に基づきステージを定義し、格子データの空間構造の複雑さの遷移を客観的にステージにより評価する手法を提唱した。応用例として、東京都23区における10年間の65歳以上人口の遷移に適用した。これらの研究成果は2018年7月にドルトムントで開催された第7回日独分類シンポジウムでの招待講演などで公表した。

##### (2) 格子型空間データに対するホットスポット検出等に関する研究

本研究では、大規模格子型空間データに対して、時間を固定した横断的なクロスセクションデータに加えて、経年的変化を調査するために、時系列空間データの空間スキャン統計量を用いたホットスポット検出を行った。また、都道府県別データに対して地域特性の評価に関する研究を行った。具体的には、以下のような研究成果が得られた。

- ① 日本原子力研究開発機構(JAEA)において、リアルタイム測定により放射線モニタリングしている福島第一原子力発電所近隣の帰還困難区域7村の2012年4月から2015年3月までの空間線量率ポイントデータに対して、クリギング手法を用いて未観測値の補間を行うとともに、時空間的観点から空間線量率が有意に高い領域(ホットスポット)の検出を行った。これらの結果は、2017年11月に杭州で開催された日中国際シンポジウムにおいて論文として公表した。
- ② 大規模格子型空間データに対して、エシェロンスキャン法による計算コストを軽減するために、制限付きスキャン統計量を提案し、従来手法と比較した。また、エシェロン解析を行うためのR-Shinyによるソフトウェア化について検討した。これらの研究成果は、2019年8月に京都で開催されたData Science, Statistics and Visualization 2019等において公表した。
- ③ 部位別死亡データの標準化死亡率及びその経験ベイズ推定値を目的変数とし、説明変数が地理的環境に依存する地理的加重回帰モデルを用いた解析を行った。特に、男性の胃癌に対する経験ベイズ推定値を目的変数とした解析では、野菜の摂取量が影響を及ぼすことなどを発見した。研究成果は2019年11月に青山学院大学で開催された日本計算機統計学会第33回シンポジウム等で公表した。

### (3) 統計量の理論的な確率分布

本研究では、小規模格子データに対して、これまで計算量が莫大で真の最大尤度を持つホットスポットを求めることが不可能であった課題に対して、ZDD 列挙法を使用したホットスポット検出方法を提案した。具体的には、以下のような研究成果が得られた。

- ① 北海道大学（現在 京都大学）の湊教授らとの共同研究において、ZDD 法を用いることにより North Carolina 州の 100 郡の組合せからなる隣接ブロックを網羅的な数え上げを行い、幼児突然死 SIDS データに対して空間スキャン統計量が最大尤度（9 セルまで）になるホットスポットの候補を検出した。これらの結果は、2017 年 12 月にオークランド（ニュージーランド）で開催されたアジア地区国際計算機統計学会の招待講演として公表した。
- ② ①の研究の続きとして、小規模格子データに対して、直接計算に基づくホットスポット検出を行った。具体的には、ZDD 法を用いることにより North Carolina 州 100 郡幼児突然死 SIDS データに対して空間スキャン統計量がすべての組み合わせの中で最大になる真にホットスポットを検出し、フレキシブルスキャンやエシエロンスキャンなどの既存の方法と比較した。これらの研究結果は、2019 SIAM International Conference および国際論文誌 *Japanese Journal of Statistics and Data Science* の第 2 巻第 1 号で公表した。

### (4) ソフトウェアの開発と公開

空間データとデータの近傍情報を与えた格子型空間データに対して、エシエロン樹木を利用し空間データの階層構造及びホットスポット検出を行うソフトウェアの開発を行った。大規模格子型空間データに対して、エシエロンスキャン法によるホットスポット検出を効率的に行う手法及びエシエロンスキャン法の R-Shiny によるソフトウェア化を行った。これらの成果は、2018 年 7 月にパーダーボルンで開催された ECDA2018 での招待講演などにおいて公表した。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 Minato, S., Kawahara, J., Ishioka, F., Mizuta, M. Kurihara, K.	4. 巻 1
2. 論文標題 A fast algorithm for combinatorial hotspot mining based on spatial scan statistic	5. 発行年 2019年
3. 雑誌名 Proceedings of the 2019 SIAM International Conference on Data Mining	6. 最初と最後の頁 91-99
掲載論文のDOI（デジタルオブジェクト識別子） <a href="https://doi.org/10.1137/1.9781611975673.11">https://doi.org/10.1137/1.9781611975673.11</a>	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 Ishioka, F., Kawahara, J., Mizuta, M., Minato, S. and Kurihara, K.	4. 巻 2(1)
2. 論文標題 Evaluation of hotspot cluster detection using spatial scan statistic based on exact counting	5. 発行年 2019年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 241-262
掲載論文のDOI（デジタルオブジェクト識別子） <a href="https://doi.org/10.1007/s42081-018-0030-6">https://doi.org/10.1007/s42081-018-0030-6</a>	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 Ishioka, F., Kurihara, K.	4. 巻 1
2. 論文標題 Visualization for radiation monitoring post data using spatial interpolation	5. 発行年 2017年
3. 雑誌名 Proceedings of 2017 Hangzhou International Statistical Symposium	6. 最初と最後の頁 55-58
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25件（うち招待講演 8件／うち国際学会 9件）

1. 発表者名 梶西将司、石岡文生、栗原考次
2. 発表標題 Echelon 構造を利用した空間複雑性の評価
3. 学会等名 科学研究費・基盤研究(A)科研費シンポジウム「空間データと災害の統計モデル」、同志社大学、2019.1.26.
4. 発表年 2019年

1. 発表者名 石岡文生、川原純、水田正弘、湊真一、栗原考次
2. 発表標題 Exact countingに基づいたホットスポットクラスターの検出について
3. 学会等名 大規模・複雑化データに対する解析手法の多面的研究、札幌市、2019.2.8.
4. 発表年 2019年

1. 発表者名 Takemura, Y. Ishioka, F., Kurihara, K.
2. 発表標題 Cluster Detection Based on Restricted Hierarchical Structure for Large Scale Data
3. 学会等名 Data Science, Statistics and Visualization 2019, Kyoto 2019.8.14. (国際学会)
4. 発表年 2019年

1. 発表者名 Kajinishi, S., Ishioka, F., Kurihara, K.
2. 発表標題 Hotspot Detection for Epidemiological Data Analysis Using R Shiny
3. 学会等名 Data Science, Statistics and Visualization 2019, Kyoto 2019.8.14. (国際学会)
4. 発表年 2019年

1. 発表者名 Kitanishi, Y., Ishioka, F., Iizuka, M., Kurihara, K.
2. 発表標題 Spatial perception for structured and unstructured data in topological data analysis
3. 学会等名 16th Conference of the International Federation of Classification Societies, Thessaloniki, Greece, 2019.8.29. (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 小部未紗, 梶西将司, 石岡文生, 栗原考次
2. 発表標題 地理的加重一般化線形モデルを用いた死亡リスクの解析とその視覚化
3. 学会等名 日本計量統計学会第33回シンポジウム、東京、2019.11.31.
4. 発表年 2019年

1. 発表者名 竹村祐亮, 石岡文生, 栗原考次
2. 発表標題 大規模空間データのためのクラスター検出手法の提案
3. 学会等名 2019年度日本分類学会シンポジウム、小倉、2019.12.4.
4. 発表年 2019年

1. 発表者名 藤原美佳, 羅 明振, 石岡文生, 栗原考次
2. 発表標題 ファストファッションに関する購買行動における要因の分析
3. 学会等名 2019年度日本分類学会シンポジウム、小倉、2019.12.4.
4. 発表年 2019年

1. 発表者名 北西由武, 石岡文生, 飯塚誠也, 栗原考次
2. 発表標題 TDA (Topological Data Analysis) Mapper による構造化・非構造化データの形状認識の可能性
3. 学会等名 2019年度日本分類学会シンポジウム、小倉、2019.12.4.
4. 発表年 2019年

1. 発表者名 藤原美佳, 梶西将司, 栗原考次
2. 発表標題 空間的分析手法を用いた消費行動の評価
3. 学会等名 第40回大規模データ科学に関する研究会、札幌市、2020.2.3.
4. 発表年 2019年

1. 発表者名 栗原考次、石岡文生
2. 発表標題 エシェロン樹木による空間構造の評価
3. 学会等名 科学研究費・基盤研究(A)科研費シンポジウム「空間データと災害の統計モデル」、同志社大学、2018.1.27.
4. 発表年 2018年

1. 発表者名 梶西将司、石岡文生、栗原考次
2. 発表標題 R-Shinyによる空間データの集積性検出
3. 学会等名 大規模データ科学に関する多面的研究会、札幌市、2018.2.8.
4. 発表年 2018年

1. 発表者名 Kurihara, K., Ishioka, F.
2. 発表標題 Echelon Clustering and its Applications
3. 学会等名 7th German-Japanese Symposium on Classification, Dortmund, Germany, 2018.7.3. (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 Ishioka, F., Kajitani, S., Kurihara, K.
2. 発表標題 Visualization of Cluster Detection based on Hierarchical Structure for Geospatial Data and its Application
3. 学会等名 European Conference on Data Analysis 2018, Paderborn, Germany, 2018.7.6. (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 Ishioka, F., Kurihara, K.
2. 発表標題 Cluster detection for mutli-dimensional spatial data based on hierarchical structure
3. 学会等名 COMPSTAT2018 (23rd International Conference on Computational Statistic) Uniera Hotel, Iasi, Romania, 2018.8.31. (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 竹村祐亮, 石岡文生, 栗原考次
2. 発表標題 制限付スキャン統計量に基づく空間集積性の新たな検出手法について
3. 学会等名 日本計算機統計学会第32回大会, 山口市, 2018.5.27.
4. 発表年 2018年

1. 発表者名 湊真一, 川原純, 水田正弘, 石岡文生, 栗原考次
2. 発表標題 スキャン統計量に基づく組合せホットスポット抽出を行う高速アルゴリズム
3. 学会等名 第169回アルゴリズム研究発表会, 小樽商科大学, 2018.9.
4. 発表年 2018年

1. 発表者名 北西由武、石岡文生、飯塚誠也、栗原考次
2. 発表標題 位相的データ解析を応用した医薬品テキスト情報からの特徴抽出のアプローチ
3. 学会等名 2018年度統計関連学会連合大会、中央大学、2018.9.13.
4. 発表年 2018年

1. 発表者名 竹村祐亮、石岡文生、栗原考次
2. 発表標題 制限付エシェロンスキャン法によるクラスター検出手法の提案
3. 学会等名 日本計算機統計学会第30回シンポジウム、彦根市、2018.11.11.
4. 発表年 2018年

1. 発表者名 北西由武、石岡文生、飯塚誠也、栗原考次
2. 発表標題 位相的データ解析によるデータの可視化
3. 学会等名 2018年度日本分類学会シンポジウム、沖縄青年会館、2018.11.25.
4. 発表年 2018年

1. 発表者名 Kurihara, K., Kajinishi, S., Ishioka, F.
2. 発表標題 Statistical evaluation for spatial complexity based on echelon trees
3. 学会等名 2017 Conference of the International Federation of Classification Societies, Tokyo, 2017.8.8. (招待講演) (国際学会)
4. 発表年 2017年

1. 発表者名 Ishioka, F., Kurihara, K.
2. 発表標題 Cluster detection using spatial scan statistic and its new development
3. 学会等名 2017 Conference of the International Federation of Classification Societies, Tokyo, 2017.8.8. (招待講演) (国際学会)
4. 発表年 2017年

1. 発表者名 石岡文生、川原純、栗原考次
2. 発表標題 空間データに対するエシェロン解析の新たな展開について
3. 学会等名 統計数理研究所共同研究集会「環境・生態データと統計解析」、統計数理研究所、2017.10.6.
4. 発表年 2017年

1. 発表者名 栗原考次
2. 発表標題 計算機統計学とともに歩み -空間データの構造分析を中心に
3. 学会等名 第56回大分統計談話会、大分、2017.10.12. (招待講演)
4. 発表年 2017年

1. 発表者名 Ishioka, F., Kawahara, J., Kurihara, K.
2. 発表標題 Evaluation of spatial cluster detection method based on all geographical linkage patterns
3. 学会等名 IASC-ARS2017 (The Conference of the Asian Regional Section of the International Association of Statistical Computing), Auckland, 2017.12.11. (招待講演) (国際学会)
4. 発表年 2017年

〔図書〕 計1件

1. 著者名 Miodreg Lovric編、日本統計学会訳	4. 発行年 2018年
2. 出版社 丸善出版(株)	5. 総ページ数 2200
3. 書名 統計科学百科事典	

〔産業財産権〕

〔その他〕

栗原研究室ホームページ <a href="http://www.ems.okayama-u.ac.jp/kurihara/">http://www.ems.okayama-u.ac.jp/kurihara/</a>
--

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