

機関番号：15301  
 研究種目：基盤研究（C）  
 研究期間：2008～2010  
 課題番号：20500258  
 研究課題名（和文） エシェロン階層情報に基づく時空間データの分析とその応用  
 研究課題名（英文） Spatio-Temporal data analysis based on echelon hierarchical structure and its applications

## 研究代表者

栗原 考次（KURIHARA KOJI）

岡山大学・大学院環境学研究科・教授

研究者番号：20170087

## 研究成果の概要（和文）：

本研究では、時間的・空間的に相互作用を受けながら得られるデータに対して、エシェロン階層情報に基づき領域分割、時空間分割を行い、時空間データの有意に集積性がみられるホットスポットの検出に関わる新手法を確立した。さらに、多変量空間データ、各種地域データ、SNP データ等の時空間データに対して開発手法の拡張を行い、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について検討した。

## 研究成果の概要（英文）：

In this research, we explore the cluster analysis for spatio-temporal data based on echelon hierarchical structure. We newly develop the technique to detect the candidates of hotspot as the top echelon in the dendrogram. In addition, the development technique was extended to the data of multi-variate space data, various regional data, SNP data and so on. Thus we can detect the time changes, expansion, reduction, movement and merger of a hotspot based on our technique.

## 交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,200,000	360,000	1,560,000
2009年度	1,000,000	300,000	1,300,000
2010年度	1,200,000	360,000	1,560,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：エシェロン解析、時空間情報、空間位相構造、分類、ホットスポット、空間スキャン統計量

## 1. 研究開始当初の背景

ある地域における病気の発生率や環境汚染物質の分布のように空間的構造をもつデータに対して、これらの関連を調査したり、有意に高い値を示す地域(ホットスポット)の検出をおこなうことは、環境の状況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。また、森林学、疫学、動物学、天文学などの分野において、植物、疾病、動物、星の地域的な集積性を検出することが大きな課題である。集積性については、空間的自己相関や移動窓による移動窓による空間スキャン統計量(Kulldorff 1997)や Flexible スキャン法(Tango, 2005)など研究が進められているが、横断的(cross sectional)な2次元空間における1変量データに関する解析であった。また、ホットスポット検出に使用される空間スキャン統計量のp値は、統計量の分布が複雑であるためにモンテカルロ法により計算されているのが現状であった。よって、(1)時空間データのホットスポット検出、(2)多変量データのホットスポット検出、(3)空間スキャン統計量の直接計算、などに対する研究の成果が期待されていた。

## 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得られるデータに対して、エシェロン階層情報に基づき領域分割、時空間分割を行い、時空間上での集積性や複雑性に関わる新手法を確立するとともに各種統計量の確率分布の計算を行い、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ことを意図している。具体的には、以下のような問題について取り扱う。

- (1) 時空間データのホットスポットの検出
- (2) 地域データに対するピークを中心とした圏による分類手法の開発
- (3) 格子数が少ない格子空間データに対するスキャン統計量の直接確率分布の計算
- (4) 多変量空間データの分類とホットスポットの検出
- (5) ソフトウェアの開発と公開

## 3. 研究の方法

時間的な広がり中で観測される空間データのメカニズムを解明するために、研究の目的で記述した5つの問題に以下のような研究の方法で積極的に取り組んだ。また、新たに開発したホットスポット検出法に対する解析結果の信頼性・安定性を評価するために、(6)ホットスポット検出法の性質評価、についても研究を行った。

- (1) 時空間データのホットスポットの検出

本研究では、時空間的に変動するデータを横断的に2次元空間上でモデル化し、ホットスポットを検出する従来の方式(栗原考次, 2003)を時空間上に拡張し、時間及び空間上でモデル化を行いホットスポットを検出する方式について研究する。まず、時間的・空間的位置情報を利用し時空間立体間の隣接情報を定義し、エシェロン情報に基づき時空間上の立体全体の階層構造を求める。その階層構造に基づきホットスポットの時間的な遷移を調べる。

- (2) 地域データに対するピークを中心とした圏による分類手法の開発

地域データの分類を行う場合、ファウンデーションに属する地域は、ピークとなる地域と独立であると考えたよりは、上位に位置するピークの地域に分類するほうが自然であ

る。本研究では、エシェロン解析で得られた階層構造とピークとの隣接情報を基に、ピークに属する地域及びピークへの寄与が大きいファウンデーションの地域から構成される、ピークを中心とした圏による分類手法を新たに開発する。

### (3) 格子数が少ない格子空間データに対する スキャン統計量の直接確率分布の計算

空間スキャン統計量をはじめとして空間データに関わるデータの統計量は複雑の従属性を持ち、統計量の理論分布を直接求めることは困難とされ、その計算にはモンテカルロ法が用いられてきた。本研究では、領域のパターン数を分類の観点から考察するとともに、統計量の理論的な確率分布を安易に計算機支援によるモンテカルロ法に頼ることなく、直接計算するアルゴリズムを開発する。

### (4) 多変量空間データの分類とホットスポットの検出

多変量空間データの分類に対して、2つの接近法を考えている。最初の方法は、空間的な位置情報は保持しながら、多変量データの次元の縮約を行い、分類を行う方法である。すなわち、空間データとして与えられる多変量データを主成分分析に基づきデータの次元を縮約し、主成分得点に基づき空間データの分類の方法である。もう一方の方法は、空間的な位置情報を破棄し、多変量空間における類似度を基に新たな空間構造を考え、分類を行う方法である。

### (5) ソフトウェアの開発と公開

1変量空間格子データに対して、データの隣接情報と1変量のデータ値を与え、エシェロン解析を利用し空間データの階層構造を求めるソフトウェアの開発を行う。

### (6) ホットスポット検出法の性質評価

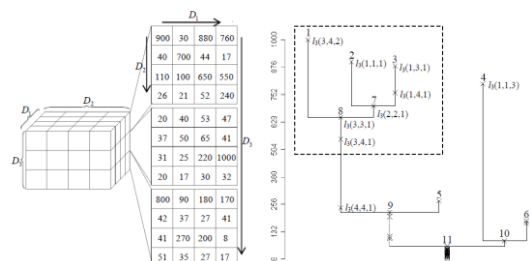
本研究で開発を行った種々の手法の信頼性・安定性を検討するために、母集団や観測

誤差などを考慮した真の空間的な階層構造について考察する。

## 4. 研究成果

### (1) 時空間データのホットスポットの検出

本研究では、ホットスポットの検出にあたり、横断的に2次元空間上でモデル化する従来の方式を拡張し、下図のような3次元時空間上でモデル化しホットスポットを検出する方式を提唱した。すなわち、**cross sectional data** のホットスポット検出から **longitudinal data** のホットスポット検出への拡張について研究を行った。また、開発された方式を、(1) 韓国地震データ、(2) 都道府県別自殺データ、(3) SNP データ、(4) 森林における樹木データ、等の時空間データに対して、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について検討した。



### (2) 地域データに対するピークを中心とした 圏による分類手法の開発

本研究では、各種の地域データに対して、エシェロン解析で得られた階層構造とピークとの隣接情報を基にエシェロンクラスター表を構成し、ピークを中心とした圏による分類する方法について新たに提唱した。また、都道府県別人口密度データに適用し、日本の都道府県を10領域の圏に分類した。

### (3) 格子数が少ない格子空間データに対する スキャン統計量の直接確率分布の計算

本研究では、代数学におけるポリオミノ理論を応用し、隣接したセルパターンを検出し、

そのパターンを領域全体に埋め込む方式のアルゴリズムを構築することに成功している。このアルゴリズムを用いることにより、先行研究で行われてきた総当たり法に比べ大幅な時間短縮を行うとともに、18個のセルパターン数の計算、4×4 Lattice 領域でのパターン数の計算まで可能にしている。

#### (4) 多変量空間データの分類とホットスポットの検出

本研究では、日本の上位4大死亡原因「悪性新生物」、「心疾患」、「脳血管疾患」、「肺炎」による都道府県別死亡データに対して、年齢構成を調節した指標 SMR を用いたエシェロン解析で求められた階層構造に基づき、疾病毎の定時点及び時系列的に地域が変化するホットスポットを検出した。さらに、4変量データのホットスポット検出のために、寄与率の高い低次元主成分空間においてボロノイ分割に基づき隣接情報を定義し、死亡総合指標を用いて、疾病に関する主成分空間におけるホットスポットの検出を行い、地域と疾病との関係を統計的に説明した。

#### (5) ソフトウェアの開発と公開

1変量空間格子データに対して、データの隣接情報と1変量のデータ値を与え、エシェロン解析を利用し空間データの階層構造を求めるソフトウェアの開発を行い、デモ版の作成を行った。

#### (6) ホットスポット検出法の性質評価

エシェロン解析は、空間的な階層構造を客観的かつ一意的に表現する記述的な方法として利用されてきた。本研究ではこれらを拡張し、母集団や観測誤差などを考慮した真の空間的な階層構造を検討した。具体的には、得られた観測値に基づき信頼区間を計算し、信頼区間から抽出された標本に基づきエシェロン解析を繰り返し行い、その傾向を調べる方法を提唱している。

これらの研究結果は、統計関連学会連合、日本計算機統計学会、科学研究費研究集会、日中統計会議等で講演するとともに、京都大学や北海道大学における特別講演、国際計算機統計会議 IASC(2008, 横浜)における招待論文、*Computational Statistics & Data Analysis* (2008)、国際計算機統計学会欧州地区大会 COMPSTAT(2008, 2010)の査読付論文として公表した。

#### 5. 主な発表論文等

[雑誌論文] (計17件)

- ① Al-Sultan, Ahmed Thanyan, M. Ishioka, F. and Kurihara, K., Airlines scheduling model and solution algorithms, *Communications of the Korean Statistical Society*, 査読有, accepted, 2011.
- ② Tomita, M. and Kurihara, K., Comprehensive Assessment on Hierarchical Structures of DNA markers Using Echelon Analysis, *Proceedings of COMPSTAT2010*, (Edited by Lechevallier, Y. and Saporta, G.), Physica-Verlag, 査読有, 2010, 1645-1652.
- ③ Oda, M. Ishioka, F. and Kurihara, K., Forestry Structure and Distribution of Trees, *Proceedings of JKSC2010* (Edited by Tanaka, Y. et al.), 査読有, 2010, 145-152.
- ④ Ishibashi, Y., Hara, A., Okayasu, I., and Kurihara, K. (2010), Neural Network Approach for Histopathological Diagnosis of Breast Diseases with Images, *Proceedings of COMPSTAT2010*, (Edited by Lechevallier, Y. and Saporta, G.) Physica-Verlag, 査読有, 2010, 1151-1158.

- ⑤ Oda, M. and Kurihara, K., Statistical Analysis of the Birds Data around the Final Waste Disposal Site, *Journal of Environmental Science for Sustainable Society*, 査読有, 2009, Vol.3, 1-8.
- ⑥ Kurihara, K. and Ishioka, F., Hotspot Detection for health and environmental data, *Proceedings of IASC2008* (Edited by M. Mizuta), 査読有, 2008, 953-961, (Invited paper).
- ⑦ Han, S., Ishioka, F. and Kurihara, K., Detection of Space-Time Hotspots for Korean Earthquake Data Using Echelon Analysis, *Proceedings of COMPSTAT2008*, (Edited by P. Brito) Physica-Verlag, 査読有, 2008, 785-792.
- ⑧ Ishioka, F. and Kurihara, K., A New Approach to Spatial Clustering Based on Hierarchical Structure, *Proceedings of COMPSTAT2008*, (Edited by P. Brito) Physica-Verlag, 査読有, 2008, 193-200.
- ⑨ Kurihara, K. and Ishioka, F., Statistical approach for environmental problems based on spatial structure, *Hokkaido University Technical report series in Mathematics*, 査読無, 2008, No.136, 10-19. (Invited paper)
- ⑩ Han, S., Ishioka, F. and Kurihara, K., Detection of Hotspot for Korea Earthquake Data using Echelon Analysis and Seismic Wave Energy, *Journal of the Faculty of Environmental Science and Technology*, 査読無, 2008, 13(1), 51-56.
- ⑪ 羅明振, 小野雄策, 小野芳朗, 栗原考次, 最終処分場に関する住民の意識調査, 岡山大学環境理工学部研究報告, 査読無, 2008, 13(1), 35-42.
- ⑫ Tomita, M., Hatsumichi, M. and Kurihara, K., Identify LD blocks based on hierarchical spatial data, *Computational Statistics & Data Analysis*, 査読有, 2008, 52(4), 1806-1820.
- [学会発表] (計32件)
- ① 惟高裕一、石岡文生、栗原考次、Echelon解析による統計的観点からの疾病集積性の検討、日本計算機統計学会第24回シンポジウム、2010.11.11、箕面市。
- ② 羅明振、石岡文生、文勝浩、栗原考次、一般廃棄物最終処理場の立地候補地の選定について、日本計算機統計学会第24回シンポジウム、2010.11.11、箕面市。
- ③ Makiko Oda, Takashi Masaki and Koji Kurihara, Patch Detection of Forest based on Spatial Structure, 科学研究費研究集会「時空間統計解析の理論と応用」、2010.11.4、鎌倉市。
- ④ 栗原考次、石岡文生、時空間データの位相構造による分類とその応用、2010年度統計関連学会連合大会、2010.9.7、新宿区(早稲田大学)。
- ⑤ 石岡文生、栗原考次、空間データの階層構造の安定性評価について、2010年度統計関連学会連合大会、2010.9.7、新宿区(早稲田大学)。
- ⑥ 小田牧子、正木隆、栗原考次、樹木の空間位相構造、2010年度統計関連学会連合大会、2010.9.7、新宿区(早稲田大学)。
- ⑦ 石岡文生、矢木良、栗原考次、空間データの階層構造の信頼性について、日本計算機統計学会第24回大会、2010.5.14、立川市。
- ⑧ 矢木良、石岡文生、栗原考次、Echelon

- 解析で検出したHotSpot候補の信頼性、第4回日本統計学会春季集会、2010.3.7、青山学院大学.
- ⑨ 豆野貴之、石岡文生、栗原考次、空間スキャン統計量と隣接パターン検出法、第4回日本統計学会春季集会、2010.3.7、青山学院大学.
- ⑩ 豆野貴之、石岡文生、栗原考次、Echelon解析におけるlatticeデータ分析、科学研究費研究集会「環境・生態データの統計科学的解析およびその基礎理論の研究」、2009.11.27、横浜市.
- ⑪ Ishioka, F. and Kurihara, K., Characteristic Evaluation of hotspot detection based on Echelon Analysis, 科学研究費研究集会「時空間統計解析の理論と応用」、2009.11.19、宇治市.
- ⑫ 小田牧子、栗原考次、生態空間分布の集積性の検討、日本計算機統計学会第23回シンポジウム、2009.11.8、江別市.
- ⑬ 石岡文生、富田誠、藤田利治、栗原考次、大規模データに対する時空間集積性の解析、第1回大規模データ科学に関する研究会、2009.9.18、札幌市.
- ⑭ 富田誠、栗原考次、遺伝統計解析におけるDNAマーカーのLD集積性、第1回大規模データ科学に関する研究会、2009.9.18、札幌市.
- ⑮ 石岡文生、栗原考次、ホットスポット検出法の性質評価について、2009年度統計関連学会連合大会、2009.9.8、京田辺市.
- ⑯ 惟高裕一、石岡文生、栗原考次、空間統計学による日本の死亡率解析、日本計算機統計学会第23回大会、2009.5.16、福岡市.
- ⑰ 石岡文生、栗原考次、大規模データへのエシェロン解析の応用、2008年度新領域融合プロジェクトによる研究会「大規模データ・リンケージ、データマイニングと統計手法」、2008.12.11、東京都港区.
- ⑱ Ishioka, F. and Kurihara, K., Investigation for the stability of hotspot detection, 科学研究費研究集会「時空間統計解析の理論と応用」、2008.11.27、那覇市.
- ⑲ 石岡文生、栗原考次、地域データの階層構造と隣接・距離情報に基づく分類について、2008年度統計関連学会連合大会、2008.9.9、横浜市.
- ⑳ 栗原考次、エシェロン階層構造に基づく時空間データとその応用、京都大学経済研究所計量経済学セミナー、2008.7.30、京都市.
- 21 Oda, M. and Kurihara, K., Ecological Analysis of Birds Data Based on a Measure of Vegetation Authenticity using GIS, 2008 Korea-Japan Statistics Conference of Young Researchers, 2008.7.10, Seoul.
- 22 Han, S., Ishioka, F. and Kurihara, K., Detection of Hotspots for Korean Earthquake Data using  $c$ -value of Gutenberg-Richter relation, 2008 Korea-Japan Statistics Conference of Young Researchers, 2008.7.10, Seoul.
- 23 小田牧子、栗原考次、GISを利用した自然度と鳥類の関連性について、日本計算機統計学会第22回大会、2008.5.23、秋田市.
- [その他]  
ホームページ  
<http://www.ems.okayama-u.ac.jp/kurihara/>
6. 研究組織  
(1) 研究代表者  
栗原 考次 (KURIHARA KOJI)  
岡山大学・大学院環境学研究科・教授  
研究者番号：20170087