

空間データベースを用いた隣接情報の作成と 自殺データの集積性への応用

久保田 貴文¹・藤野 友和²・冨田 誠³・石岡 文生⁴・藤田 利治⁵

(受付 2012 年 10 月 12 日; 改訂 12 月 11 日; 採択 12 月 25 日)

要 旨

空間データの一つである Lattice データを解析に用いる場合には、その隣接情報が肝要ではあるが、不規則的な Lattice であれば、それを作成するのは困難である。そこで、本研究では空間データベースを用いて隣接情報を作成する方法を提案した。さらに、作成された隣接情報および日本における自殺データを用いて集積性の検出へと適用し、先行研究にて検出されている同データの集積性の結果と比較した。また、空間データベースを用いれば、1つの Lattice に含まれる領域の数の修正などにも容易に対応できるので、同自殺データにおいて、二次医療圏での集積性と市区町村での集積性についても考察した。

キーワード：空間データベース、隣接情報、空間集積性、自殺データ。

1. はじめに

空間データの一つである Lattice データを解析に用いる場合には、特にその隣接情報が重要である。空間自己回帰、空間的集積性の検定、地理空間相関分析など、ある地域ごとの周辺の情報解析に利用する場合、隣接情報が必要となる。リモートセンシングデータなどの規則的な Lattice データであれば、隣接情報も当然ながら規則的になるため、容易に得ることができるとは、市区町村ごとのデータなどの不規則的な Lattice データについては容易に得ることができない。

これまででは、利用可能なデータが紙媒体の地図などアナログなデータであったため、その隣接情報の算出方法は図を目視にて判断する方法で行われてきた。また、その過程で誤りを避けるために塗分け地図などを用いて行われてきた。しかし近年では、例えば日本における国土数値情報は国土交通省から入手可能であり(国土交通省, 2012)、それを取り扱うためのデータベースや描画するためのソフトウェアも開発されており、本研究で対象とする隣接情報についても容易に求められるようになってきている。特に 2012 年 4 月より SHAPE 形式も提供されるようになり、より利便性が高まったといえる。

ところが、空間データの解析の場面で、このように隣接情報をデータベースから利用するよ

¹ 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

² 福岡女子大学 国際文理学部：〒813-8529 福岡県福岡市東区香住ヶ丘 1-1-1

³ 東京医科歯科大学病院 臨床試験管理センター：〒113-8519 東京都文京区湯島 1-5-45

⁴ 岡山大学大学院 法務研究科：〒700-8530 岡山県岡山市北区津島中 3-1-1

⁵ 故人

うな研究は十分になされていない。例えば、本研究で取り扱う自殺のデータ解析においても、富田 他(2010)で隣接情報が使用されているが、その導出については十分に議論されていない。

そこで本研究では、まずはじめに、日本における二次医療圏ごとの隣接情報を空間データベースから作成し、「自殺対策のための自殺死亡についての地域統計」(藤田, 2009)を利用して日本における二次医療圏ごとの自殺死亡の集積性を検出し富田 他(2010)の結果と比較検証を行った。

さらに、領域の大きさを変更して隣接情報を作成することが、空間データベースを用いる上での利点でもあるため、二次医療圏と市区町村ごとの領域において、先述のデータを用いて、同様に自殺死亡の集積性を比較した。ただし、領域の大きさを変更した影響をより確認しやすいように、関東地方に限定して解析を行った。

本論文においては、2節において空間データベースを用いて隣接情報の作成について述べ、3節においてそれらの隣接情報の自殺死亡の空間集積性について述べ、4節において考察を行う。

2. 隣接情報の作成

2.1 環境

空間データベースを用いて隣接情報を作成するための環境として、Ubuntu Linux 11.10 (Ubuntu Japanese Team, 2012)の環境のもと、Postgresql 8.4.10(The PostgreSQL Global Development Group, 2012)において国土数値地図のうち行政区域(面)のデータを格納し、PostGIS 1.5.1(Project Steering Committee, 2012)により空間データベースの機能を利用した。

2.2 手順

まずはじめに、国土数値地図のうち行政区域(面)のデータを都道府県別に取得し、データベースに格納した。このデータのうち、緯度・経度などのジオメトリ型の情報としては、ポリゴンとして格納されていて、一つの市区町村は複数のポリゴンから構成されている。そのために、次に同一の市区町村コードをもつポリゴンをPostGISの関数のうちST_Unionにより統合する。さらに、二次医療圏ごとの隣接情報も必要なため、二次医療圏ごとに市区町村を統合する。ここで、二次医療圏は2008年3月31日現在の情報を利用する(市町村要覧編集委員会, 2008)。

得られた、市区町村もしくは二次医療圏ごとのジオメトリ型のデータにおいて、それぞれの領域が他の領域と隣接しているか否かを、PostGISの関数のうちST_Intersectsによりそれぞれ確認する。

ただし、データベース上で一連の作業は複雑さが伴うため、スクリプト言語Ruby(1.9.1)(まつもと, 2012)およびそのライブラリーであるpg(Granger et al., 2012)を用いて、PGconnによりPostGISと接続し、処理を行う。具体的には、 n 個の領域からそれ自体の領域を除く $(n-1)$ 個のすべての領域に対して、隣接しているかどうかを、都合 $n(n-1)/2$ 回繰り返し判断する。

さらに、描画のためには、PostGISの関数のうちST_Simplifyにより、境界を単純化する。

3. 自殺データの集積性への応用

3.1 データについて

「自殺対策のための自殺死亡についての地域統計」(藤田, 2009)は厚生労働省の人口動態調査死亡票を使用して改訂され、2008年3月末現在での二次医療圏および市区町村ごとに1973年まで遡れるかたちで自殺死亡の実態を整理している。また、二次医療圏では年齢階級・性別での自殺死亡数・死亡率・対全国比をまとめている(年齢は10歳以上である)。

本研究においては、二次医療圏ごとおよび市区町村ごとの4期間(1988–1992年(90)、1993–1997年(95)、1998–2002年(00)、2003–2007年(05))に着目し、地域ごとに人口が異なるため

自殺率(人口 10 万人あたりの自殺者数)を用いた。

3.2 手法

二次医療圏ごとおよび市区町村ごとの自殺率のデータに対して、どこに集積地域(有意に高い値を示す地域)が存在しているのかを検討する。集積性検出のために、空間スキャン統計量(Kulldorff, 1997)を用い、スキャンの方式として Echelon 解析(Myers et al., 1997; Kurihara, 2004)を利用する。

3.3 日本の二次医療圏ごとの自殺データへの適応

空間データベースにより作成された隣接情報により、富田 他(2010)の隣接情報が一部更新された。県境等で識別できなかった隣接、地図では襷がけのように描かれている隣接、小さな島や埋め立て地に境界がある場合など、目視では確認できない点が更新された。一方、湖については、上記数値地図には登録されていないため、逆に目視の方が正しい隣接を示していた。なお、富田 他(2010)では、陸地が物理的に隣接していなくても、航路、トンネル、橋等が存在すれば隣接としていたが、その点は引き続き利用している。この更新された隣接情報で求めた日本における自殺死亡の地域統計(二次医療圏)の集積エリアを検出した結果、空間スキャン統計量ともに大きな違いは見受けられなかった。

3.4 関東地方の自殺データへの適応

次に、関東地方の二次医療圏ごとおよび市区町村ごとの自殺死亡率のデータに対して同様に集積エリアを検出した。ただし、東京都の島しょについては、地理的な属性が大きく異なるため、本研究では除外して、二次医療圏ごとでは 65 の地域、市区町村ごとでは 340 の地域を対象とした。集積エリアとしての最大クラスター数を全体のエリアの約 10, 15, 20, 25%をそれぞれで求めて比較を行った(エリアの数はそれぞれ、二次医療圏: 7, 10, 13, 16, 市区町村: 34, 51, 68, 85 である)。また、男女については別々に取り扱った。

関東地方における二次医療圏ごとの集積エリアのうち、最も統計量の高い集積エリア(第一ホットスポット、黒色で塗り分け)、2 番目に統計量の高い集積エリア(第二ホットスポット、灰色で塗り分け)およびその他のエリア(白色で塗り分け)のコロプレスマップのうち、図 1 に男性を図 2 に女性を示す。それぞれの図は行ごとに最大クラスター数(上から順に全体の約 10, 15, 20, 25%)、列ごとに年(左から順に 90, 95, 00, 05)の図を示している。

また、表 1, 2 にはそれぞれ、前述の二次医療圏ごとの男性・女性の検出された第一ホットスポット、第二ホットスポットの空間スキャン統計量を示す。なお、P 値については、すべての第一・第二ホットスポットにおいて、0.01 以下であった。

これより、図 1 のそれぞれの行ごとの 4 つのコロプレスマップより、年によって集積エリアは異なり、特に 05 は大きく異なる。また、この結果は、Ishioka et al. (2010)と比較しても同等のエリアだといえる。女性においても、図 2 より同様のことがいえる。

さらに、図 1 のそれぞれの列ごとの 4 つのコロプレスマップと表 1 のそれぞれの列ごとの空間スキャン統計量より、最大クラスターサイズと検出されるエリアの大きさが関係することがわかった。すなわち、クラスターサイズが大きくなるほど、大きすぎるエリアが検出されることがわかった。

二次医療圏と同様に、関東地方における市区町村ごとの集積エリアのうち、最も統計量の高い集積エリア(第一ホットスポット、黒色で塗り分け)、2 番目に統計量の高い集積エリア(第二ホットスポット、灰色で塗り分け)およびその他のエリア(白色で塗り分け)のコロプレスマップのうち、図 3 に男性を図 4 に女性を示す。それぞれの図は行ごとに最大クラスター数(上から順に全体の約 10, 15, 20, 25%)、列ごとに年(左から順に 90, 95, 00, 05)の図を示している。

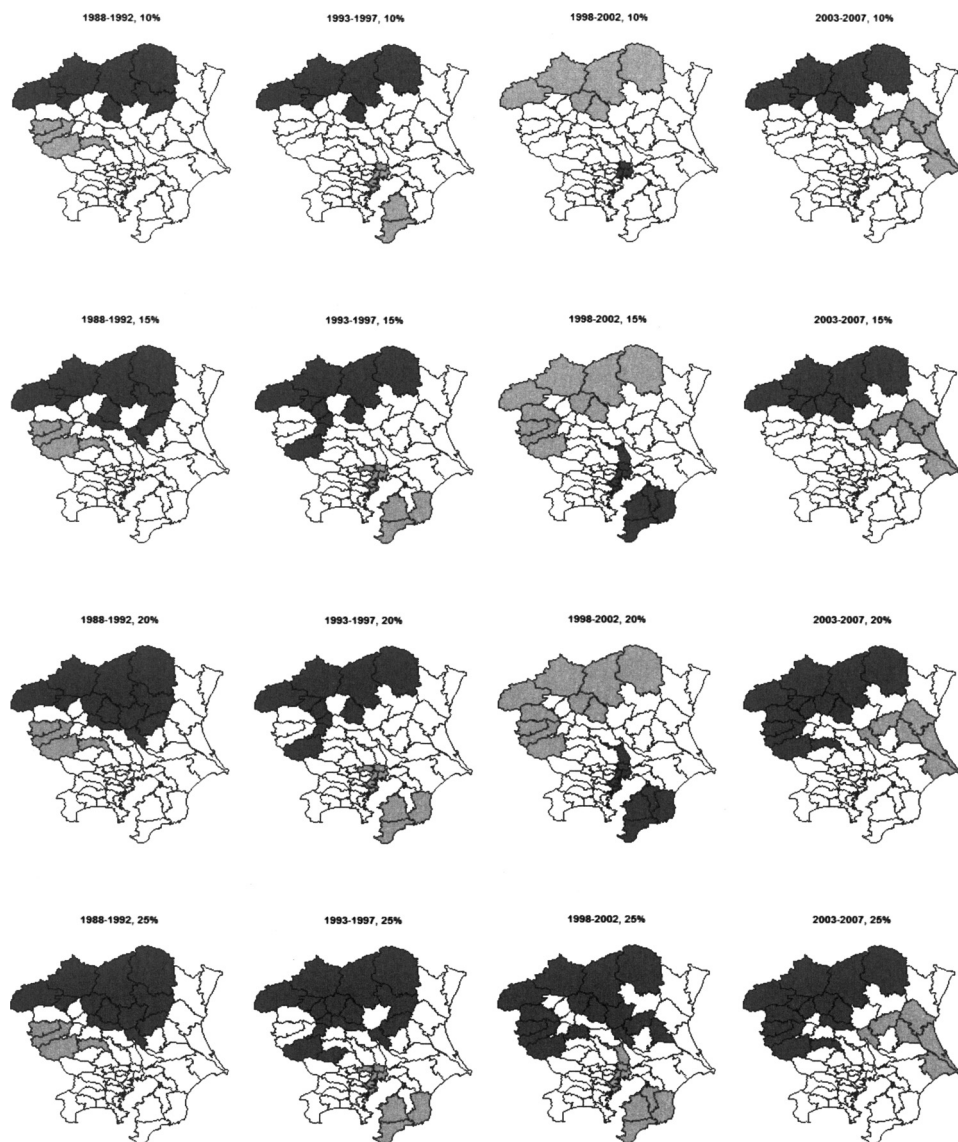


図1. 関東地方における二次医療圏ごとのホットスポット（男性）（黒色；第一ホットスポット，灰色；第二ホットスポット，行ごとに最大クラスター数（上から順に全体の10，15，20，25%），列ごとに年（左から順に90，95，00，05）を示す）。

また，表3，4にはそれぞれ，前述の市区町村ごとの男性・女性の検出された第一ホットスポット，第二ホットスポットの空間スキャン統計量を示す。なお，P値については，すべての第一・第二ホットスポットにおいて，0.01以下であった。

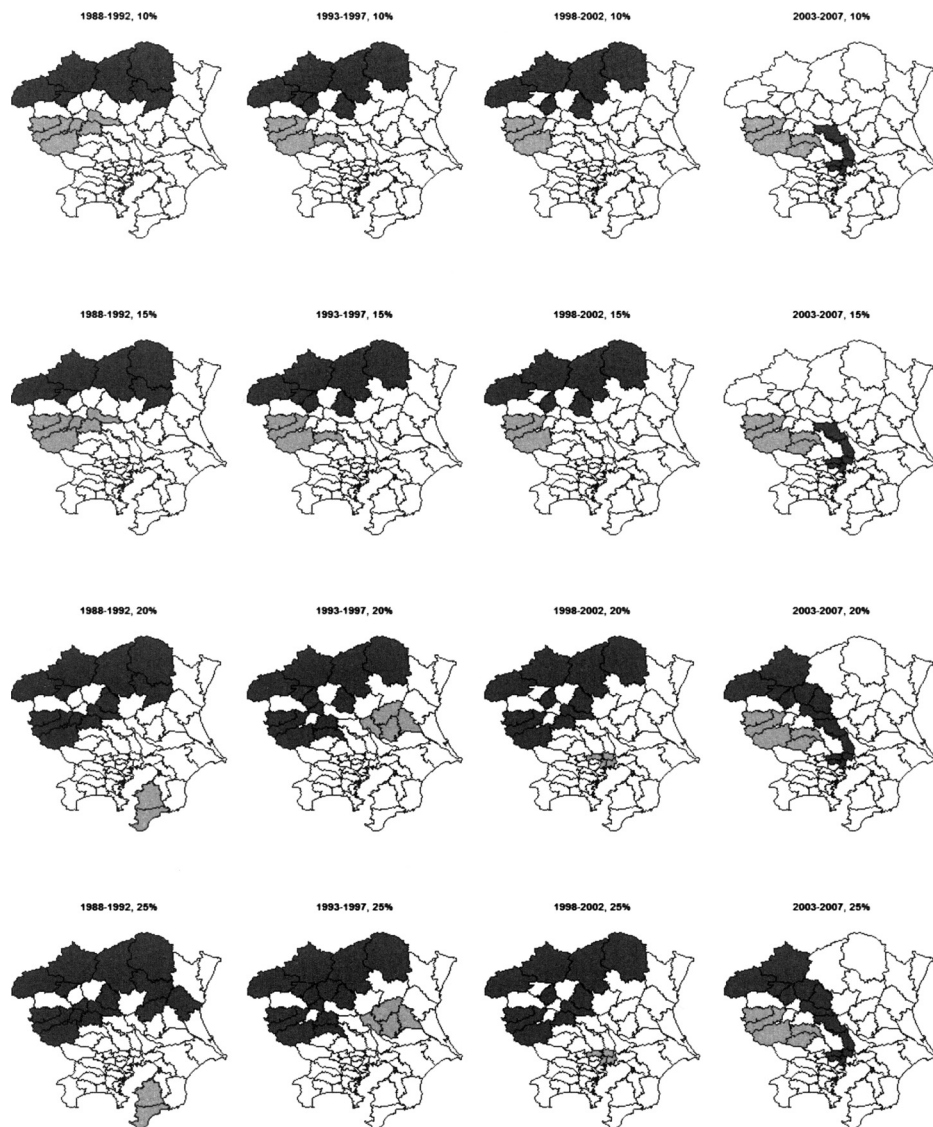


図 2. 関東地方における二次医療圏ごとのホットスポット（女性）（黒色；第一ホットスポット，灰色；第二ホットスポット，行ごとに最大クラスター数（上から順に全体の 10, 15, 20, 25%），列ごとに年（左から順に 90, 95, 00, 05）を示す）。

4. 考察

空間データベースを用いて作成した隣接情報で集積エリアを求め、比較することにより（響がけや県境に多い）山間部や、小さい島の部分の隣接情報の変更が、日本の自殺率の集積性には影響を与えないことが確認された。

また、関東地方における自殺データの集積性を求めることにより、時期（年）や性別の違いにより集積エリアの場所が異なることも確認できた。具体的には、群馬県から栃木県にかけての

表 1. 検出された第一ホットスポット, 第二ホットスポットの空間スキャン統計量 (二次医療圏ごとの男性).

最大クラスター数	1988 - 1992		1993 - 1997		1998 - 2002		2003 - 2007	
	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster
10%	58.498	19.016	44.493	32.946	37.556	37.135	66.452	51.487
15%	68.31	19.016	56.322	36.527	54.607	51.924	68.372	51.487
20%	71.943	19.016	56.322	36.527	54.607	51.924	83.279	51.487
25%	71.943	19.016	60.487	36.527	65.461	54.607	83.295	51.487

表 2. 検出された第一ホットスポット, 第二ホットスポットの空間スキャン統計量 (二次医療圏ごとの女性).

最大クラスター数	1988 - 1992		1993 - 1997		1998 - 2002		2003 - 2007	
	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster
10%	67.478	46.769	44.95	28.03	23.315	15.441	14.353	11.433
15%	67.478	46.769	44.95	28.03	23.315	15.441	14.353	11.433
20%	121.829	8.655	78.956	6.423	37.414	10.865	20.164	11.433
25%	127.783	8.655	79.299	6.423	37.414	10.865	20.164	11.433

表 3. 検出された第一ホットスポット, 第二ホットスポットの空間スキャン統計量 (市区町村ごとの男性).

最大クラスター数	1988 - 1992		1993 - 1997		1998 - 2002		2003 - 2007	
	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster
10%	86.776	31.533	74.612	45.848	63.959	56.105	89.867	76.972
15%	103.597	31.533	92.711	45.848	66.537	56.105	93.715	76.972
20%	103.715	31.533	110.182	45.848	66.537	56.105	127.223	76.972
25%	141.53	24.172	122.388	45.848	90.989	56.105	127.223	76.972

表 4. 検出された第一ホットスポット, 第二ホットスポットの空間スキャン統計量 (市区町村ごとの女性).

最大クラスター数	1988 - 1992		1993 - 1997		1998 - 2002		2003 - 2007	
	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster	Most likely cluster	Secondary cluster
10%	88.252	69.555	64.051	40.264	36.819	30.181	20.739	14.46
15%	88.252	69.555	81.18	40.264	36.819	30.181	23.819	20.739
20%	88.252	69.555	125.613	12.415	36.819	31.655	40.172	20.739
25%	168.988	16.071	128.406	12.415	36.819	31.655	41.475	20.739

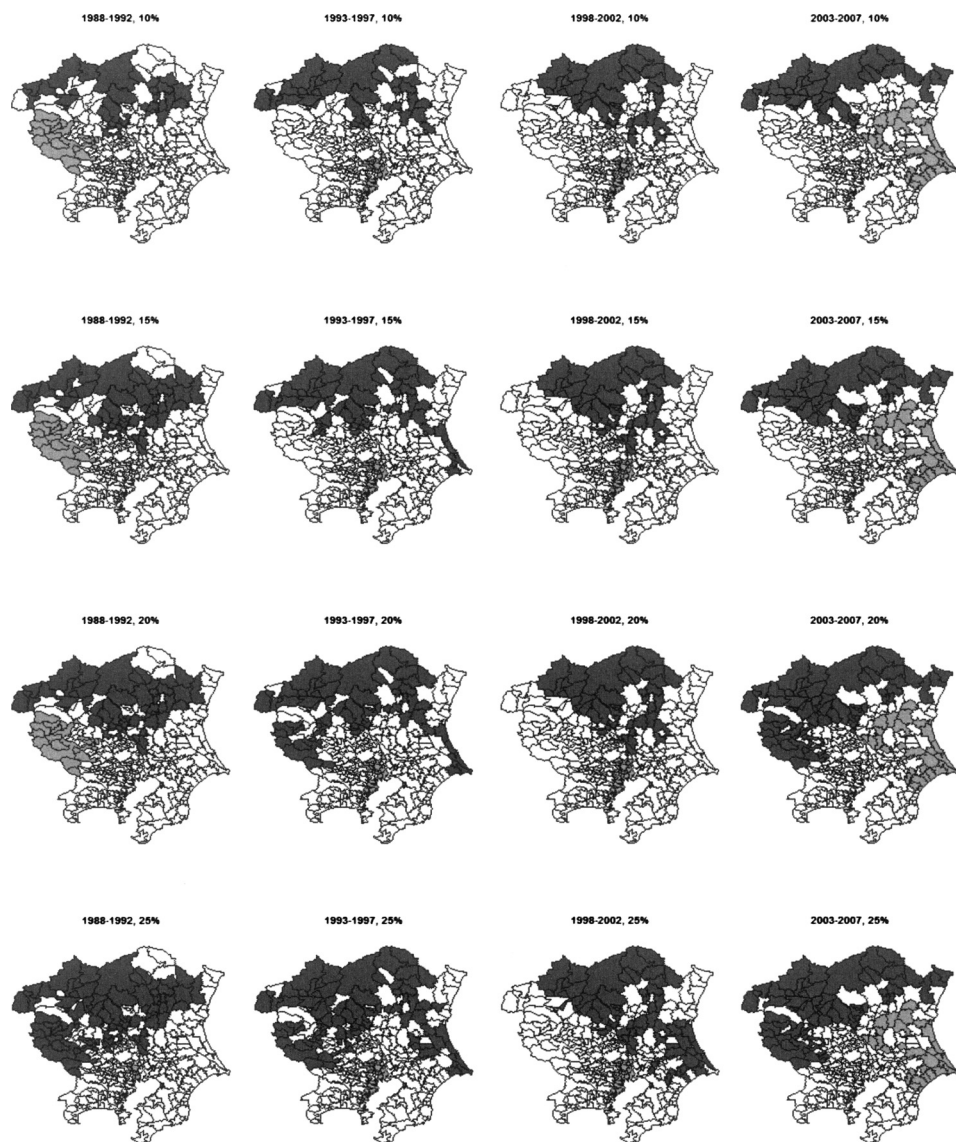


図3. 関東地方における市区町村ごとのホットスポット(男性)(黒色;第一ホットスポット, 灰色;第二ホットスポット, 行ごとに最大クラスター数(上から順に全体の10, 15, 20, 25%), 列ごとに年(左から順に90, 95, 00, 05)を示す).

エリアは(女性の05を除けば)集積エリアとして検出されているが, 東京都23区から神奈川県
の沿岸部にかけてのエリアは検出されている時期とそうではない時期が存在する点である。

さらに, 最大クラスター数が大きくなると, 空間スキャン統計量は大きくなるが, その分エ
リアが大きくなり解釈が難しくなることもわかった. 一方, それにともなって, 第一ホット
スポットのエリアが大きくなることにより, 例えば図2の1列を見ると, 1988-1992年の女性で
は, 房総半島南部においてこれまで発見できていなかった第二ホットスポットを発見できた

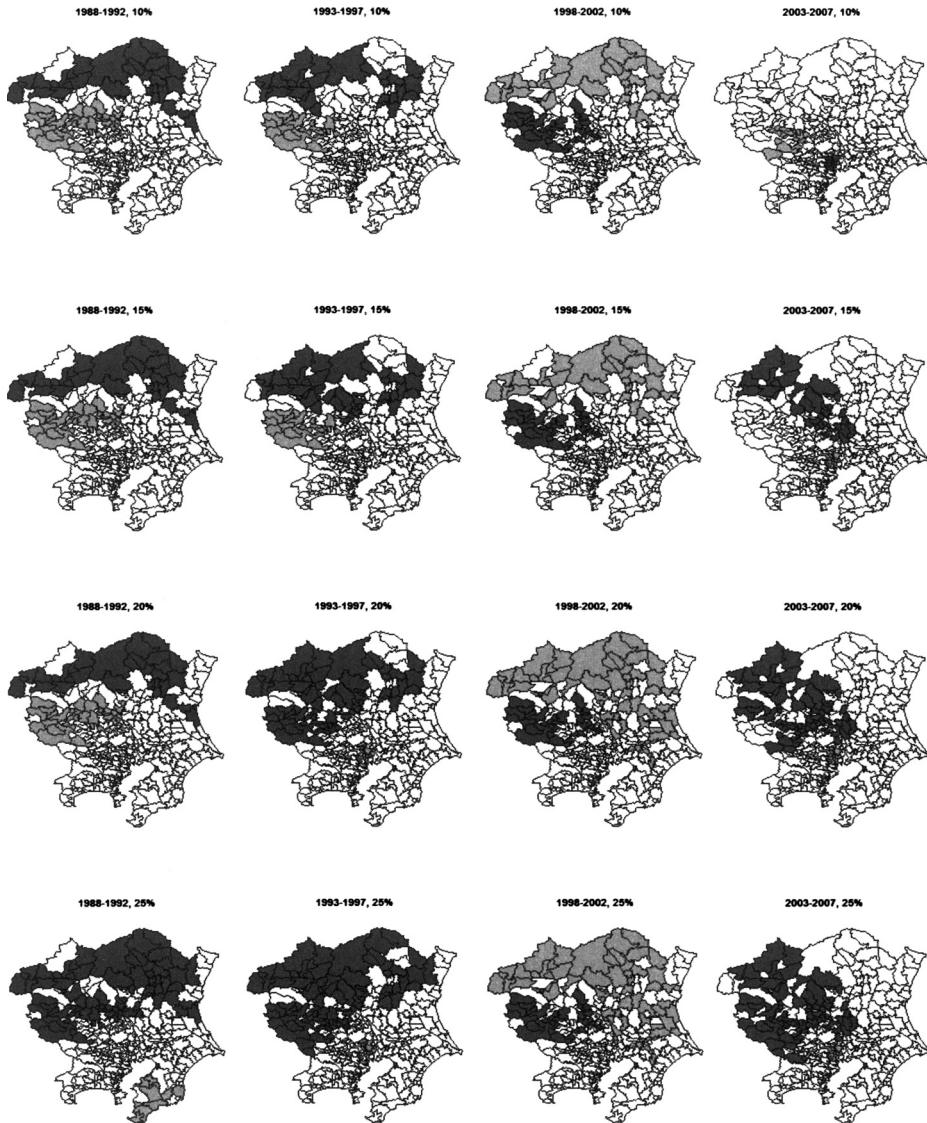


図4. 関東地方における市区町村ごとのホットスポット(女性)(黒色;第一ホットスポット, 灰色;第二ホットスポット, 行ごとに最大クラスター数(上から順に全体の10, 15, 20, 25%), 列ごとに年(左から順に90, 95, 00, 05)を示す).

もいえる。

最後に、隣接情報を二次医療圏から市区町村に変更することにより、より細かい隣接情報を用いて集積エリアを求められるだけでなく、二次医療圏単位では発見できなかった集積エリアに含まれる市区町村を検出できた。逆に、大きな二次医療圏単位では集積エリアと検出されていたが、市区町村単位で考えると、集積エリアとはならない場所も存在することもわかった。

ただし、これらの傾向は市区町村の区切りでは、人口が大きく異なるため、それにとまなっ

て自殺率が敏感に影響が出ている可能性がある。これらの対策として、自殺率の標準化死亡比 (SMR) やベイズ推定自殺率を利用することを今後の課題としたい。特に、自殺死亡についての地域統計ではベイズ推定自殺率は二次医療圏ごとの人口規模を重みとするモーメント推定値により算出されているが (丹後, 1988; 平子 他, 1999), 本研究で提案した空間データベースを用いて各市区町村に隣接する市区町村を 1 つのエリアとして、その人口規模により推定する方法も検討する予定である。

謝 辞

本研究は、独立行政法人国立精神・神経医療研究センター精神保健研究所からの受託研究「自殺予防対策にかかわる調査データの統計解析に関する研究」によるものであり、科研費・若手研究 (B) (21700305), 若手研究 (B) (21700317), 基盤研究 (C) (24500337) の助成の一部を受けたものである。

参 考 文 献

- 藤田利治 (2009). 自殺死亡についての地域統計, 国立精神・神経センター精神保健研究所 自殺予防総合対策センター.
- Granger, M., Bleything, B. and Davis, J. (2012). Project Info. RubyForge: ruby-pg (オンライン), <https://rubyforge.org/projects/ruby-pg/>.
- 平子哲夫, 佐伯則英, 中田 正 (1999). 人口動態市区町村別統計へのベイズ統計の応用について: (1) 標準化死亡比への応用, 厚生学の指標, **46** (10), 3–11.
- Ishioka, F., Tomita, M. and Fujita, T. (2010). Detection of spatial cluster for suicide data using Echelon analysis, *COMPSTAT2010 Proceedings in Computational Statistics*, 1159–1166.
- 国土交通省 (2012). GIS ホームページ. 国土数値情報ダウンロードサービス (オンライン), <http://nlftp.mlit.go.jp/ksj/>.
- Kulldorff, M. (1997). A spatial scan statistics, *Communications in Statistics, Theory and Methods*, **26**, 1481–1496.
- Kurihara, K. (2004). Classification of geospatial lattice data and their graphical representation, *Classification, Clustering, and Data Mining Applications* (eds. D. Banks, F. McMorris, P. Arabie and W. Gaul), 251–258, Springer-Verlag, Berlin Heidelberg, New York.
- まつもとゆきひろ (2012). オブジェクト指向スクリプト言語 Ruby, Ruby (オンライン), <http://www.ruby-lang.org/ja/>.
- Myers, W.L., Patil, G. P. and Joly, K. (1997). Echelon approach to areas of concern in synoptic regional monitoring, *Environmental and Ecological Statistics*, **4**, 131–152.
- Project Steering Committee (2012). Home. PostGIS (オンライン), <http://postgis.refractory.net/>.
- 市町村要覧編集委員会 (2008). 『全国市町村要覧 (平成 20 年版)』, 第一法規, 東京.
- 丹後俊郎 (1998). 死亡指標の経験的ベイズ推定値について: 疫学への適用, 応用統計学, **17**, 81–96.
- The PostgreSQL Global Development Group (2012). The world's most advanced open source database, PostgreSQL (オンライン), <http://www.postgresql.org/>.
- 富田 誠, 石岡文生, 藤田利治 (2010). 日本の自殺データにおける時空間解析, 計算機統計学, **23** (1), 25–43.
- Ubuntu Japanese Team (2012). Homepage, Ubuntu Japanese Team (オンライン), <http://www.ubuntulinux.jp/>.

Calculation of Neighboring Information Using Spatial Database and Its Application to Spatial Clustering of Suicide Data

Takafumi Kubota¹, Tomokazu Fujino², Makoto Tomita³, Fumio Ishioka⁴ and
Toshiharu Fujita⁵

¹The Institute of Statistical Mathematics

²The International College of Arts and Sciences, Fukuoka Women's University

³Clinical Research Center, Tokyo Medical and Dental University Hospital of Medicine

⁴School of Law, Okayama University

⁵Deceased

Neighboring information is necessary for analyzing lattice data. One of the challenging tasks is to calculate irregular lattice data. Therefore, the authors proposed using a spatial database to calculate neighboring information of irregular lattice data. With this proposed usage of the spatial database, the authors applied neighboring information to spatial clustering for small area data of suicide in Japan. One advantage of using the spatial database is that it can be used for different area sizes. Therefore, the authors calculated two kinds of neighborhood information that have different area sizes: municipalities and secondary medical care zones. Finally, they were applied to spatial clustering to discuss the effect of changing the area sizes in the case of suicide in the Kanto region of Japan.