

201036034A

厚生労働科学研究費補助金

健康安全・危機管理対策総合研究事業

健康危機事象発生の検出を目的とした症候サーベイランスに
おける統計解析法とその利用に関する研究

平成22年度 総括研究報告書

研究代表者 高橋邦彦

平成23（2011）年 3月

目 次

I. 総括研究報告		
健康危機事象発生の検出を目的とした症候サーベイランスにおける統計解析法とその利用に関する研究	-----	1
(資料)	-----	11
・ FleXScan User Guide for version 3.1 (日本語版)	-----	13
・ FleXScan User Guide for version 3.1 (英語版)	-----	29
・ 発表資料 : Takahashi K and Tango T. A maximum scan score-type statistic based on Anscombe's variance stabilization transformation for disease clustering. XXVth International Biometric Conference.	-----	47
II. 研究成果の刊行に関する一覧表	-----	59
III. 研究成果の刊行物・別刷	-----	63
・ Tango T, Takahashi K and Kohriyama K. A space-time scan statistic for detecting emerging outbreaks. Biometrics.		

総括研究報告

厚生労働科学研究費補助金（健康安全・危機管理対策総合研究事業）
総括研究報告書

健康危機事象発生の検出を目的とした症候サーベイランスにおける統計解析法と
その利用に関する研究

研究代表者 高橋邦彦 国立保健医療科学院技術評価部主任研究官

研究要旨：本研究では健康危機事象の早期検出（early detection）を目的とした症候サーベイランスに用いることができる統計手法「疾病集積性の検定法」の開発とその利用について、特に本研究代表者・分担者らが提案するFlexScan法を中心とした研究、検討を行った。アプリケーションソフトの改良と国内での利用促進に向けたデータの提供を行った。また健康危機事象のアウトブレイクを的確に早期検出するための新たな統計モデルに基づく検定法を開発を行い、シミュレーションによる比較を行い従来の統計量よりも精度よく検出ができることが確かめられた。

研究分担者

- ・丹後俊郎(国立保健医療科学院技術評価部前部長・昭和女子大学客員教授・医学統計学研究センター長)
- ・山岡和枝(国立保健医療科学院技術評価部開発技術評価室長)

A. 研究目的

近年のバイオテロリズムやSARS、新型インフルエンザの発生などのように、その発生をいち早く検出するためのサーベイランスシステムの必要性が世界的に高まってきている。ある症候が突発的・集中的に発生するなどの事象をいち早く検出するためには、日々それらの症候の発

生状況を継続的に観察・監視し、その中で事象の発生が突発的・集中的に発生したと疑われる場合に、それが偶然なのか、または意味のあるものなのかを、客観的に判定することが必要となる。その方法として集積性の検定という統計手法が利用できる。中でも米国ハーバード大学のKulldorff博士の提案する手法とそのソフトウェア「SaTScan」は世界的に広く利用されており実際米国でのいくつかの監視システムにも組み込まれ日々解析が行われている。この方法は簡便ではあるが、一方でいくつかの問題があることが最近指摘されてきている。その問題の一つを解決するため我々が開発したflexible scan法とそのソフトウェア「FlexScan」も注目され利用され、海外の専門書（Pfeiffer et al. “Spatial Analysis in Epidemiology”；Shekhar and Xiong(eds.) “Encyclopedia of GIS “など）にも取り上げられている。実際、一般的な研究者やユーザーが容易に、無料で解析

を行えるソフトウェアは世界においてKulldorfによる「SaTScan」と我々の「FlexScan」の2つだけであると認識されている状況であり、近年もこの2つのソフトウェアを用いた疫学研究などが世界的に増えてきている。さらにニューヨーク市の保健局（New York City Department of Health and Mental Hygiene (DOHMH)）においても保健局のGIS Center of Excellenceを中心にFlexScanの利用に関する検討が始められている。そこで本研究では、FlexScan法の更なる改良に関する理論的研究とともにアプリケーションによる解析ツールの提供、サーベイランスならびに関連分野における実際の適用と、その普及に向けた検討を行うことを目的とする。

B. 研究方法

まずアプリケーションソフトFlexScan V3.1を日本語および英語マニュアルも含めて正式に公開した。その上で国内外の利用者から寄せられる要望等に応じ、随時改良を行った。

さらに健康危機事象のアウトブレイクをより適切に検出できるように、空間・時間解析における新たな統計モデルの検討を行い、SaTScanなどで用いられている従来の統計モデルとの比較を行った。

また、時点ではなく時間変化を考慮したspace-time解析のバージョンの試行版を作成し、今後に向けて検討を行った。

(倫理面への配慮)

本研究で解析に用いたデータは個別データではなく集計されたものであり、倫理面で問題が生じることはない。

C. 研究結果

(1) アプリケーションソフト「FlexScan」の開発

Windows上でflexible scan法を利用できるアプリケーションソフトFlexScanについては、海外の疫学の専門書（Pfeiffer D et al, “Spatial Analysis in Epidemiology” 2008, Oxford）などに、そのダウンロード先もあわせて紹介されている。今年度、国内外の利用者からの要望にも応えながら、アプリケーションFlexScan V3.1と日本語版マニュアル、英語版のマニュアルを作成し公開した。その後も、随時利用者からの要望に応えながら公開を行っている：

http://www.niph.go.jp/soshiki/gijutsu/download/flexscan/index_j.html（日本語ページ）

<http://www.niph.go.jp/soshiki/gijutsu/download/flexscan/index.html>（英語ページ）

さらに一時点ではなく時間変化を考慮したspace-time解析のバージョンの試行版を作成し、今後の開発に向けた準備を行った（図1, 2）

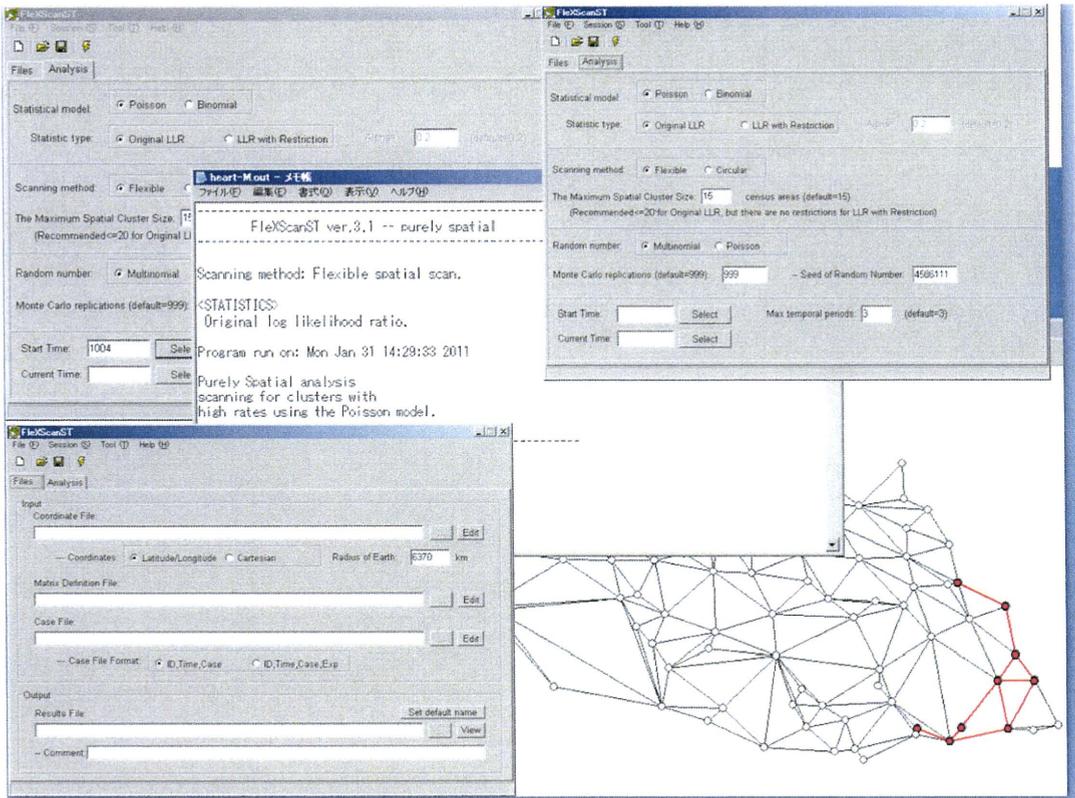


図1 FlexScan space-time解析版の試行版の画面例

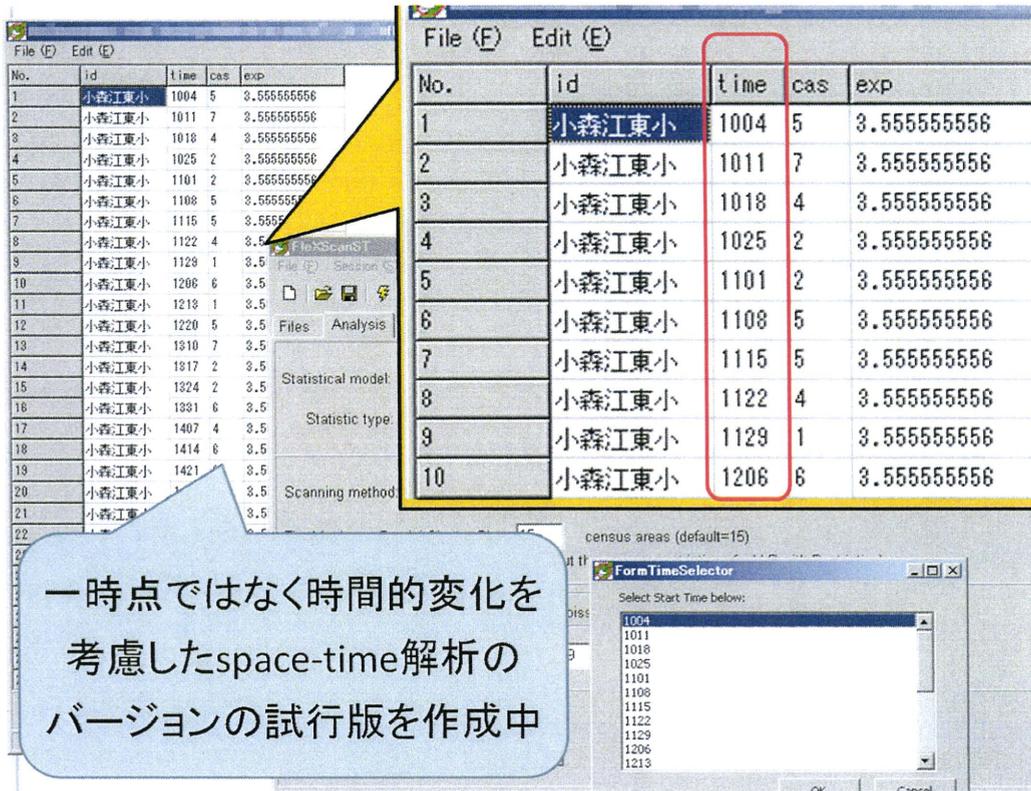


図2 FlexScan space-time解析版の試行版のデータ入力例

(3) 新たな統計量の提案

サーベイランス解析において空間・時間 (space-time) データを用いた統計量としては平面における通常の統計量と同様のポアソン分布に基づく尤度比統計量が用いられている。本研究では、健康危機事象の立ち上がりをより適切に評価し検出するための新たな統計モデルを検討、

提案した。実際、北九州市における小学校の学童欠席数のサーベイランスを想定したシミュレーション評価によって、本提案法が、従来の統計モデルよりも早期に、また適切な検出を行うことが確認でき、サーベイランスのための統計モデルとしてより適切であると考えられた。研究成果は国際学術専門誌Biometricsに掲載された。

○ 発生の「立ち上がり」を反映したモデル化に基づく手法の開発

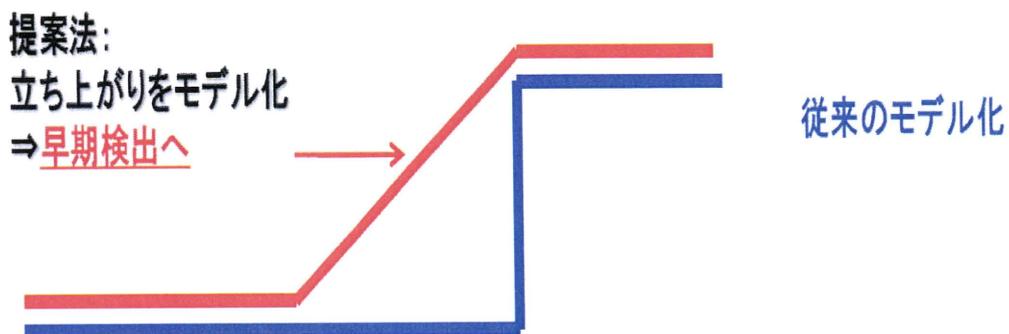
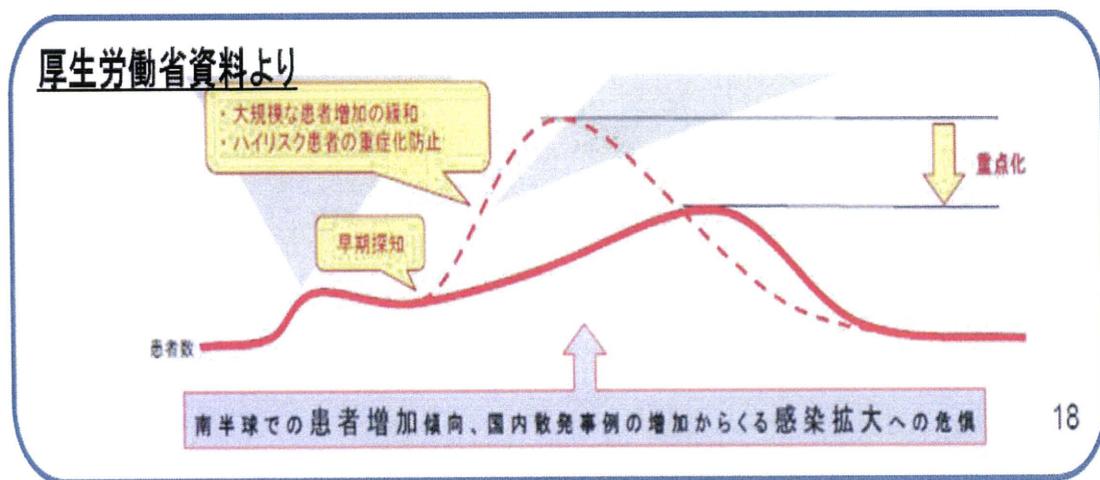
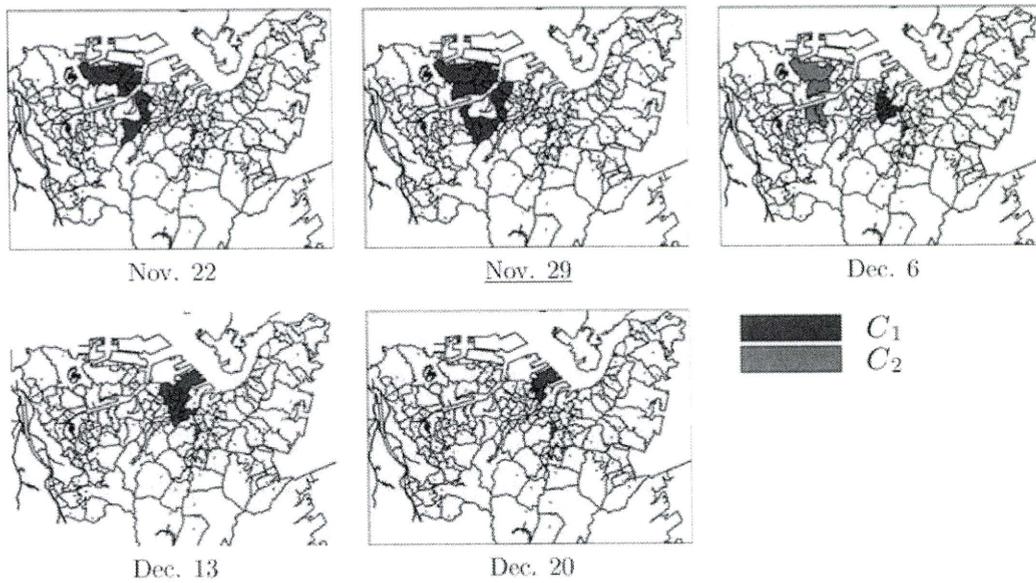


図3：アウトブレイクをモデル化した新たな統計モデルのイメージ図

a) Kulldorff's cylindrical scan ($T = 2$)



b) Proposed cylindrical scan ($T = 2$)

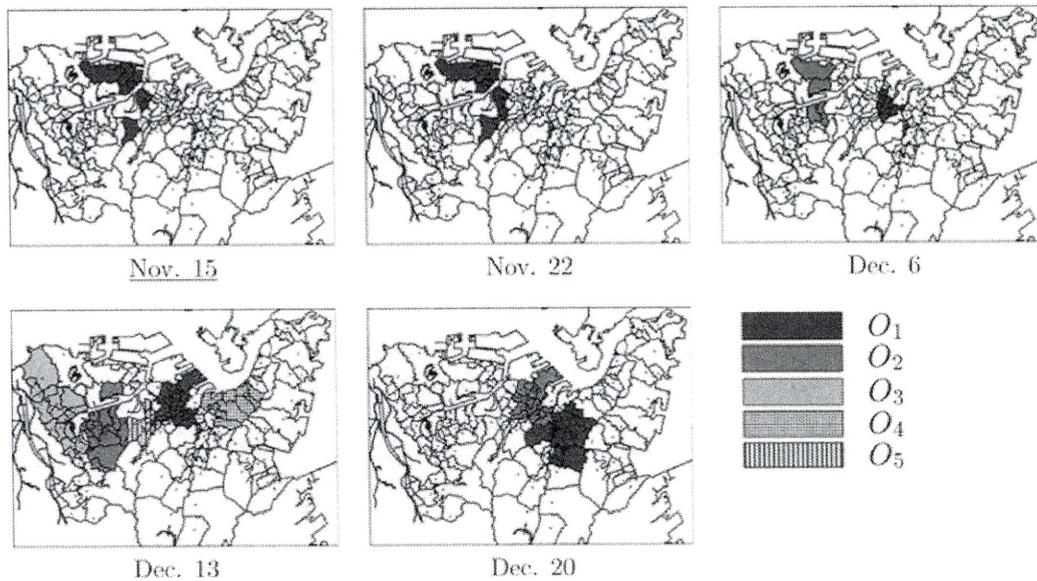


図4 提案法と従来法の比較の図（詳細は論文参照）

D. 考察

昨年までの研究を通して、提案するFleXScan法は従来法（SaTScan法）に比べ、複雑な形状の地域でのアウトブレイクをうまく検出できることが確認できている。実際、全ての集積地域が円状であるとは考え難く、FleXScanによってより複雑な形状の地域を勃発地の候補として同定することが適切であると考えられる。特にニ

ューヨーク市におけるインフルエンザA（H1N1）の発生を用いたサーベイランスデータへの適用についての研究成果に基づき、ニューヨーク市保健局（NYC-DOHMH）GISセンターでは解析手法のひとつとしてFleXScanが利用されていることがホームページにも掲載されるまでになった（図5）。現在、毎日の自動解析に組み込むべく検討が開始されている。

アプリケーションの開発としては、身近なユ

ユーザーの意見を取り入れながら改良を行った。特に海外の実務者からの問い合わせも寄せられた点を重点的にその対応および改善を行った。本研究期間内に昨年公開のバージョン3.0に続き、バージョン3.1の公開を行った。今後、実務者等に利用してもらい、その上で有用性を確かめたい。またこれらの解析とその解釈について、国立保健医療科学院での講義、また国内外の関連学会で発表を行い、普及に努めた。その成果もあり、いくつかの疫学・GISの専門書で取り上げられた。今後ひきつづきユーザーの

意見を取り入れながら、実用性を考慮しつつ、理論的側面からの新しい統計量の開発も含め、国内外で利用できるアプリケーションへ発展させ、またその普及につとめたい。

さらに提案する新たな手法について、その有用性をさらに検討するとともに、それらの空間・時間解析を組み込んだアプリケーションソフトの公開の要望も海外からいくつか寄せられており、その実現に向けて更なる開発の検討を行うことが望まれている。

図5 : FleXScanが取り上げられているニューヨーク市保健局 (NYC-DOHMH) のホームページ
<http://www.nyc.gov/html/doh/html/epi/giscenter.shtml>

E. 結論

本年度の研究では、サーベイランスに向けたより適切な解析ができるような統計量の改良の検討について、理論的側面および実際の側面から様々な検討を行い、より適切な手法・ツールを提案したが、今後も引き続き検討を行っていききたい。

アプリケーションソフトFlexScanについても国内外でこのアプリケーションの認知されるようになってきた。今後もユーザーの声を反映しサーベイランスにおいて有用なツールとして容易に利用できるよう改良を続け、解析に有用なデータを含め公開されるとともに、今後はニューヨーク市での実用化とともに国内での実用化についても検討を行っていききたい。

F. 健康危機情報

なし

G. 研究発表

1. 論文発表

・Tango T, Takahashi K, Kohriyama K. A space-time scan statistic for detecting emerging outbreaks. *Biometrics* 2011, 67:106-115

2. 学会発表

・Takahashi K and Tango T. A maximum scan score-type statistic based on Anscombe's variance stabilization transformation for disease clustering. XXVth International Biometric Conference, December 5-10, 2010, Florianopolis, Brazil.

3. 著書

・Tango T. *Statistical Methods for Disease Clustering*. Springer, New York, 2010.

・高橋邦彦. 空間スキャン統計量. In 丹後俊郎, 小西貞則 (編) *医学統計学の事典*. 朝倉書店, 2010.

・高橋邦彦. 症候サーベイランス. In 丹後俊郎, 小西貞則 (編) *医学統計学の事典*. 朝倉書店, 2010.

4. その他

・FlexScan v3.1: Software for the Flexible Scan Statistic. 国立保健医療科学院技術評価部,
http://www.niph.go.jp/soshiki/gijutsu/index_j.html

資料

FleXScan User Guide

for version 3.1

高橋邦彦 横山徹爾 丹後俊郎

国立保健医療科学院

2010年10月

http://www.niph.go.jp/soshiki/gijutsu/index_j.html

User Guide version 3.1

はじめに

FleXScan は対象地域の中で観察された疾病がある地域に集積しているかどうかという疾病集積性 (disease clustering) の検討をするための統計解析を Windows 上で行うことができるソフトウェアです。集積性の検定にはいくつかの手法が提案されていますが、FleXScan では Kulldorff's Circular Scan 法、Tango and Takahashi's Flexible Scan 法の 2 つによって検討することができます。

FleXScan (Version 3) では従来の尤度比統計量に加え、制限付き尤度比統計量による解析が行えます。さらに Version 2 までのポアソンモデルに加え 2 項モデルに基づく解析が行えるようになっています。これらの理論については参考文献を参照して下さい。

参考文献

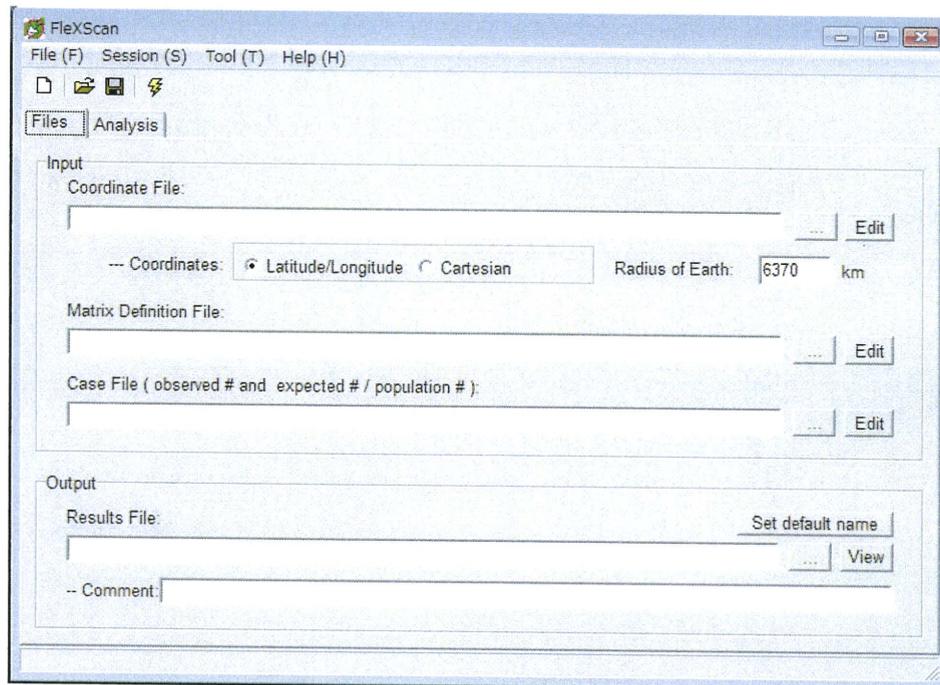
- 丹後俊郎, 横山徹爾, 高橋邦彦 (2007). 空間疫学への招待 (医学統計学シリーズ 7). 朝倉書店.
- Tango T. and Takahashi K. (2005). A flexibly shaped spatial scan statistic for detecting clusters, *International Journal of Health Geographics* 4:11.
- Kulldorff M. and Nagarwalla N. (1995). Spatial disease clusters: Detection and Inference. *Statistics in Medicine* 14:799–810.
- Tango T. (2008). A spatial scan statistic with a restricted likelihood ratio. *Japanese Journal of Biometrics* 29(2):75–95.

ダウンロードとセットアップ

FlexScan のソフトウェアおよび配布データなどは、国立保健医療科学院・技術評価部のホームページ：http://www.niph.go.jp/soshiki/gijutsu/index_j.html にアクセスし、「ダウンロード」内から入手することができます。ダウンロードしたファイルを展開してできたフォルダ内の「FlexScan」アイコンをダブルクリックして起動することができます。

FlexScan version 3.1 (for Windows) は以下の環境を推奨します。

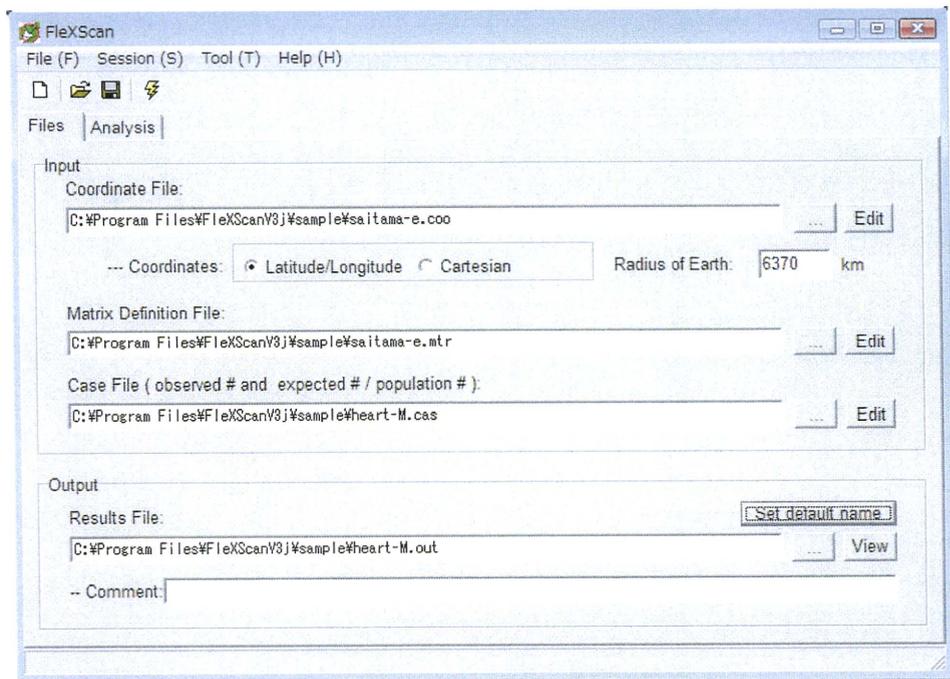
- OS : Microsoft Windows 2000, Windows XP, Windows Vista, Windows 7
- メモリ : 256MB 以上
- CPU : Pentium III 以上



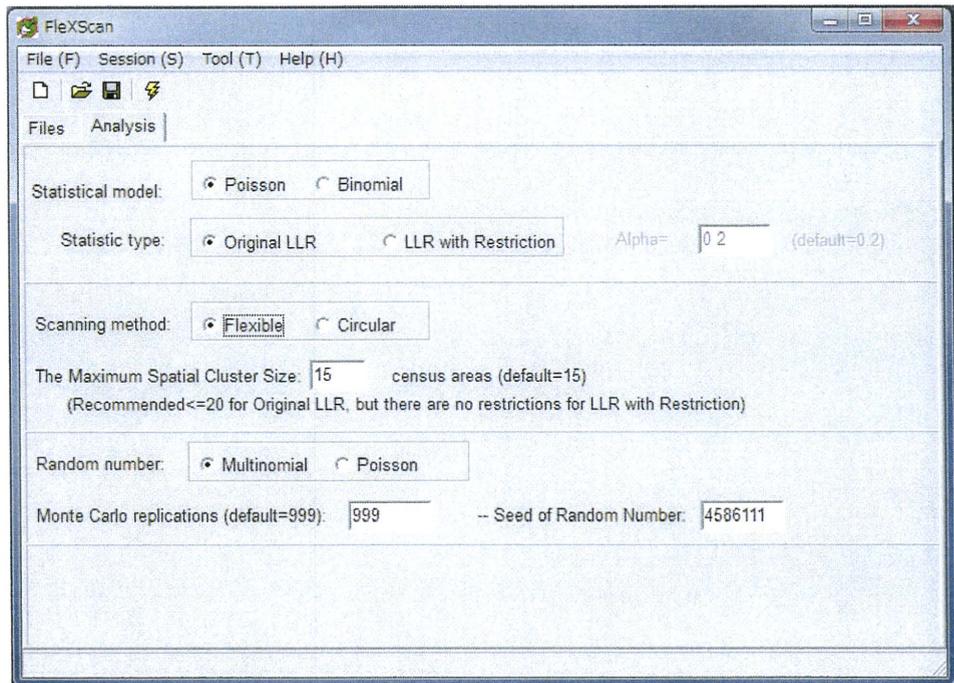
動作確認

FleXScan の中にはサンプルデータが入っていますので、それを用いて動作確認をすることができます。以下の手順を実行して下さい。

1. FleXScan を起動します。
2. Input の項目に以下の sample フォルダの中のファイルをそれぞれ選択します。
 - ① Coordinate File: “saitama-e.coo”
 - ② Coordinate – “Latitude/Longitude” をチェックする。
 - ③ Matrix Definition File: “saitama-e.mtr”
 - ④ Case File: “heart-M.cas”
3. Output の項目に出力ファイルを入力します。「Set default name」ボタンをクリックすれば自動的に出力ファイルが決まります。



- 次に「Analysis」タブを選択します。Statistical model で Poisson を選択し、Scanning method で Flexible か Circular を選択します。



- 上の Run ボタンをクリック、または「Session」－「Run」を選択し実行します。
- エラーが発生しなければ、解析終了後、結果がおよび位置情報を模式化したマップが表示されます。

結果の見方：

Output ファイルの「MOST LIKELY CLUSTER」の項の

Census areas included .: kawaguchi, souka, koshigaya,...

に挙げられている市区町村の集合が、このデータで同定された集積地域です。

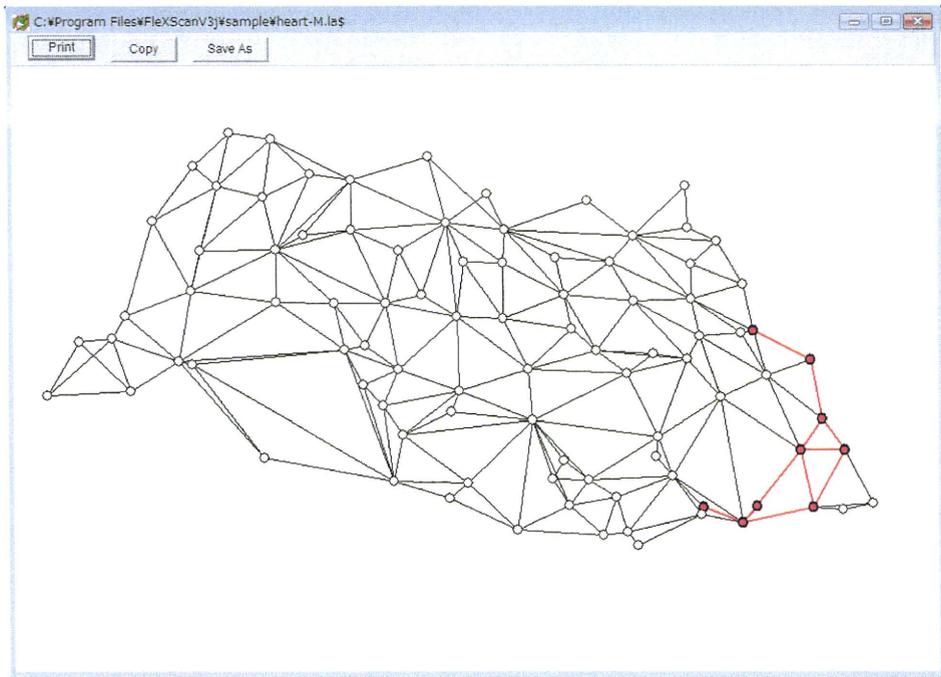
さらに下の

P-value: 0.001

がその有意性の程度を表しています。

P-value にかかわらず most likely cluster として同定された地域が模式的なマップで赤く示されています。

```
heart-M.out - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
-----
FlexScan ver.3.0 -- purely spatial
-----
Scanning method: Flexible spatial scan.
<STATISTICS>
Original log likelihood ratio.
Program run on: Wed Mar 11 14:44:28 2009
Purely Spatial analysis
scanning for clusters with
high rates using the Poisson model.
-----
SUMMARY OF DATA
Limit length of cluster: 15
Number of census areas.: 92
Total cases .....: 16462
(expected number is adjusted by Total cases as *Total expects* = *Total cases*)
-----
MOST LIKELY CLUSTER
1.Census areas included .: kawaguchi, souka, koshigaya, warabi, hatogaya, yoshikawa, sugito
Maximum distance.....: 24.2393 km (areas: kawaguchi to sugito)
```



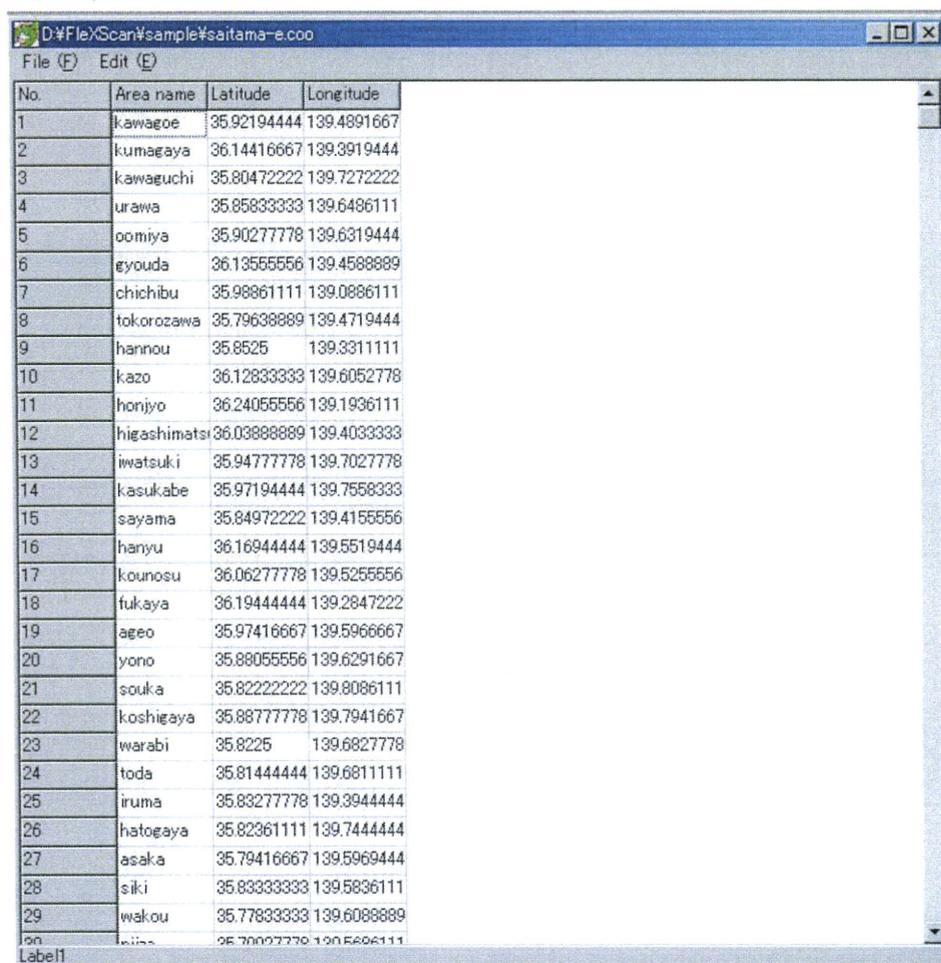
サンプルファイルとデータ形式

サンプルファイルは埼玉県のデータをもとに作られています。市区町村の区分および市区町村の位置（役所の場所）は「平成 12 年度版 日本の市区町村 位置情報要覧」（国土地理院、日本地図センター発行）に従っています。

- 位置情報ファイル：Coordinate File（.coo）
 - Format： <市区町村名> <緯度> <経度>

緯度、経度はそれぞれ「度」で表示されます。すなわち
xx 度 yy 分 zz 秒 → $xx + yy/60 + zz/3600$ （度）のように変換して入力します。

緯度、経度でなく、XY 座標を用いることも出来ます（詳しくは「注意」を参照して下さい）。



The screenshot shows a text editor window titled "D:\FileXScan\sample\saitama-e.coo". The window contains a table with the following data:

No.	Area name	Latitude	Longitude
1	kawagoe	35.92194444	139.4891667
2	kumagaya	36.14416667	139.3919444
3	kawaguchi	35.80472222	139.7272222
4	urawa	35.85833333	139.6486111
5	oomiya	35.90277778	139.6319444
6	gyouda	36.13555556	139.4588889
7	chichibu	35.98861111	139.0886111
8	tokorozawa	35.79638889	139.4719444
9	hannou	35.8525	139.3311111
10	kazo	36.12833333	139.6052778
11	honjyo	36.24055556	139.1936111
12	higashimats	36.03888889	139.4033333
13	iwatsuki	35.94777778	139.7027778
14	kasukabe	35.97194444	139.7583333
15	sayama	35.84972222	139.4155556
16	hanyu	36.16944444	139.5519444
17	kounosu	36.06277778	139.5255556
18	fukaya	36.19444444	139.2847222
19	ageo	35.97416667	139.5966667
20	yono	35.88055556	139.6291667
21	souka	35.82222222	139.8086111
22	koshigaya	35.88777778	139.7941667
23	warabi	35.8225	139.6827778
24	toda	35.81444444	139.6811111
25	iruma	35.83277778	139.3944444
26	hatogaya	35.82361111	139.7444444
27	asaka	35.79416667	139.5969444
28	siki	35.83333333	139.5836111
29	wakou	35.77833333	139.6088889
30

- 接続情報ファイル：Matrix Definition File (mtr)

➤ Format： <市区町村名> <市区町村名 1> <市区町村名 2> ...

例えば Area1 と接しているのが、Area3, Area5 の 2 つであるときには

Area1 Area3 Area5

Area2

Area3 Area1 ...

...

のようになります。Area1 と Area3 が接続している場合には、Area3 の行にも必ず Area1 が入らなければいけません。必ず全て対称になるようにして下さい。対称でない場合にはエラーが発生します。この対称性の確認は、Edit 画面の「File」－「Check symmetry」で行うことができます。詳しくは「ファイルの編集」の項目を参照して下さい。

The screenshot shows a window titled "D:\File\Scan\sample\saitama-e.mtr" with a menu bar containing "File (F)" and "Edit (E)". The main area is a table with columns "No.", "Area name", and eight "Connected" columns. The table lists 23 rows of area connections between various municipalities in Saitama.

No.	Area name	Connected	Connected	Connected	Connected	Connected	Connected	Connected
1	kawagoe	oomiya	tokorozawa	sayama	ageo	fujimi	kamifukuoka	sakado
2	kumagaya	gyouda	higashimats	fukaya	fukiage	namekawa	oosoto	kounan
3	kawaguchi	urawa	iwatsuki	souka	koshigaya	warabi	toda	hatogaya
4	urawa	kawaguchi	oomiya	iwatsuki	yono	warabi	toda	asaka
5	oomiya	kawagoe	urawa	iwatsuki	ageo	yono	fujimi	hasuda
6	gyouda	kumagaya	kazo	hanyu	kounosu	fukiage	menuma	kisai
7	chichibu	naguri	tokigawa	yokose	minano	yoshida	okano	arakawa
8	tokorozawa	kawagoe	sayama	iruma	niiza	miyoshi		
9	hannou	sayama	iruma	hidaka	moroyama	ogose	naguri	tokigawa
10	kazo	gyouda	hanyu	kuki	kisai	kitakawabe	ootone	kurihashi
11	honjyo	fukaya	misato-ma	codama	kamisato	okabe		
12	higashimats	kumagaya	sakado	namekawa	arashiyama	kawashima	yoshimi	hatoyama
13	iwatsuki	kawaguchi	urawa	oomiya	kasukabe	koshigaya	hasuda	shiraoka
14	kasukabe	iwatsuki	koshigaya	miyashiro	shiraoka	sugito	matsubushi	syouwa
15	sayama	kawagoe	tokorozawa	hannou	iruma	hidaka		
16	hanyu	gyouda	kazo					
17	kounosu	gyouda	okegawa	kitamoto	fukiage	yoshimi	kisai	kawazato
18	fukaya	kumagaya	honjyo	menuma	okabe	kawamoto	hanazono	yorii
19	ageo	kawagoe	oomiya	okegawa	hasuda	ina	kawashima	
20	yono	urawa	oomiya					
21	souka	kawaguchi	koshigaya	yashio	misato-shi	yoshikawa		
22	koshigaya	kawaguchi	iwatsuki	kasukabe	souka	yoshikawa	matsubushi	
23	warabi	kawaguchi	urawa	toda				

解析対象のデータは以下の2つのパターンがあります。

◎ 各市区町村の期待度数を用いたモデルで解析を行う場合

- データファイル：Case File (observed # and expected #) (cas)
 - Format： <市区町村名> <観測度数> <期待度数>

解析対象全域を基準にした期待度数を用います。性・年齢調整など行った期待度数を用いることで性・年齢調整した結果が得られることとなります。

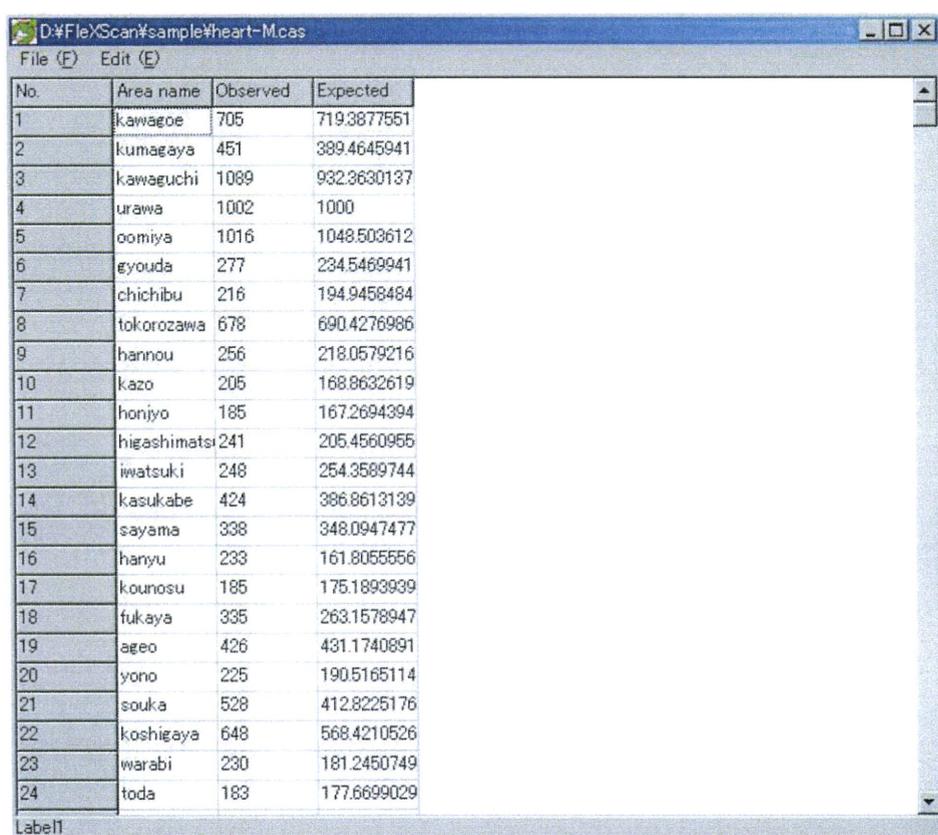
Poisson モデルでの解析を行います。

◎ 各市区町村の人口を用いたモデルで解析を行う場合

- データファイル：Case File (observed # and population #) (cas)
 - Format： <市区町村名> <観測度数> <対象者数>

各地域の対象者数（人口など）とその中での観測数の割合を用います。

二項モデル（Binomial model）での解析を行います。



No.	Area name	Observed	Expected
1	kawagoe	705	719.3877551
2	kumagaya	451	389.4645941
3	kawaguchi	1089	932.3630137
4	urawa	1002	1000
5	oomiya	1016	1048.503612
6	gyouda	277	234.5469941
7	chichibu	216	194.9458484
8	tokorozawa	678	690.4276986
9	hannou	256	218.0579216
10	kazo	205	168.8632619
11	honjyo	185	167.2694394
12	higashimats	241	205.4560955
13	iwatsuki	248	254.3589744
14	kasukabe	424	386.8613139
15	sayama	338	348.0947477
16	hanyu	233	161.8055556
17	kunosu	185	175.1893939
18	fukaya	335	263.1578947
19	ageo	426	431.1740891
20	yono	225	190.5165114
21	souka	528	412.8225176
22	koshigaya	648	568.4210526
23	warabi	230	181.2450749
24	toda	183	177.6699029