

200840023A

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

地理・時間情報を加味した突発的・集中的な健康危機事象の
発生を早期発見するための統計手法に関する研究

平成20年度 総括研究報告書

研究代表者 高橋邦彦

平成21(2009)年3月

目 次

I. 総括研究報告	
地理・時間情報を加味した突発的・集中的な健康危機事象の発生を早期発見するための 統計手法に関する研究	----- 1
(資料) FleXScan 3.0 ユーザーマニュアル	
II. 研究成果の刊行に関する一覧表	----- 22
III. 研究成果の刊行物・別刷	----- 24

総括研究報告

地理・時間情報を加味した突発的・集中的な健康危機事象の発生を
早期発見するための統計手法に関する研究

研究代表者 高橋邦彦 国立保健医療科学院技術評価部研究員
研究分担者 丹後俊郎 国立保健医療科学院技術評価部部長

研究要旨：本研究では、症候サーベイランスのように日々の観察から突発的・集中的な事象（症候、有害事象など）の発生をいち早く発見し、その有意性を客観的に判断するための方法に適用できる統計的検定手法について主として検定手法の理論の開発・評価を行った。本年度は米国ニューヨーク州の担当者との共同研究によりニューヨークのサーベイランスデータにおいて提案する方法の有用性を確認した。さらに統計量の改良を試み、その検討を行った。新たな統計量および新たなモデルを組み込んだアプリケーションソフト FleXScan version 3 を開発し、無料で公開を行った。

A. 研究目的

近年、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カルドルフ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るようになってきている。本研究では我々の提案する flexible scan 法をサーベイランスに適用するため時間変化を考慮した手法への拡張を行う。またその研究を通してより精度よく事象を同定できる統計的な手法の改良を検討しさらに実際のサーベイランスへの適用について検討を行う。

B. 研究方法

まずニューヨーク市の実務担当者との連携によって、米国でのサーベイランスの実情を把握する。その解析ツールとして用いられる集積性の検定法（従来法である SaTScan 法）の利用のされ方、およびその結果について検討を行った。さらにその実情を踏まえ、我々の提案する FleXScan 法を実際のニューヨーク市のサーベイランスデータに適用し、SaTScan 法との比較

を行った。

次に、平面および時間・空間における集積性の検定において、より精度よく集積地域を同定できるように統計量の改良を試みた。その上で、これらの改良した統計量や、また国内外の利用者からの要望に応え、新たなモデルのもとでの解析が行えるよう、よりユーザーフレンドリーになるよう工夫した FleXScan version 3 の開発を行い、無料で公開を行った。

(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) ニューヨーク市のサーベイランスデータの解析

ニューヨーク市 New York City Department of Health and Mental Hygiene (NYC DOHMH)での実際のサーベイランス（救急患者 ED）データに FleXScan を適用し、従来用いられている SaTScan と結果の比較を行った。実際にニューヨーク市でアウトブレイクがあったとされている日のデータについて解析を行った結果、両法とも同様の地域を同定したが、FleXScan がより相対リスクの高い非円状の地域を同定した（図1）。この結果はニューヨーク市の担当者が見ても、よりリーズナブルな結果であると考えられた。今後、ニューヨーク市の解析においても FleXScan での解析も検討されるよう引き続き共同研究を続けていくことになった。

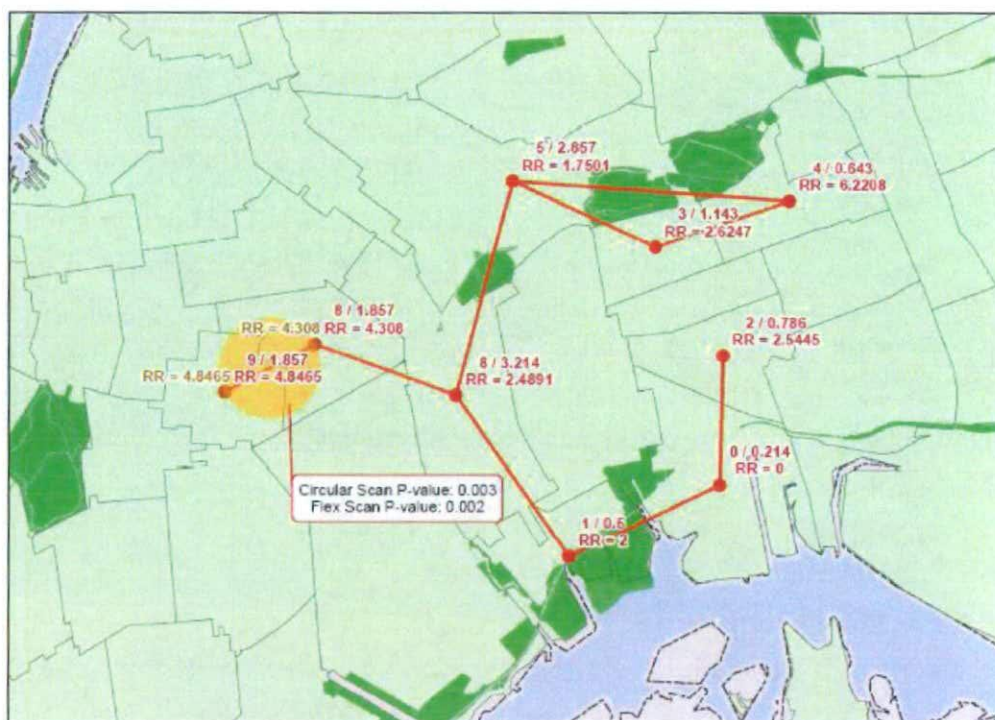


図1：ニューヨーク市のデータの解析結果 Circular scan significant cluster and flexible scan significant cluster. Relative risk is also shown.

(2) 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 FlexScan の改良を行った。ユーザーとして利用するような人たちからの意見を取り入れ、新たなモデル（2項モデル）、新たに提案された統計量を組みこんだ FlexScan v3.0 の開発を行った（図2）。一般に死亡などの比較的稀な事象についてはポアソンモデルで解析が行えるが、人口と観測数を用いた「受診率」や比較的観測数が大きくなる事象については2項モデルが利用できる。実際、国内外の研究者から、このモデルを組み込んだアプリケーションの要望が寄せられ、今回、それを含めた改良ができた。また、新たな統計量を組み込むことで、より適切なク

ラスターを、計算時間も短く検出・同定できるようになった。さらにソフトウェアのインターフェイスもよりユーザーフレンドリーになるよう心がけ、実務家でも利用しやすいものとなった。なお、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た。本研究成果の一部として、日本語および英語のマニュアルも作成し、無料でダウンロードできる形で公開した。

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

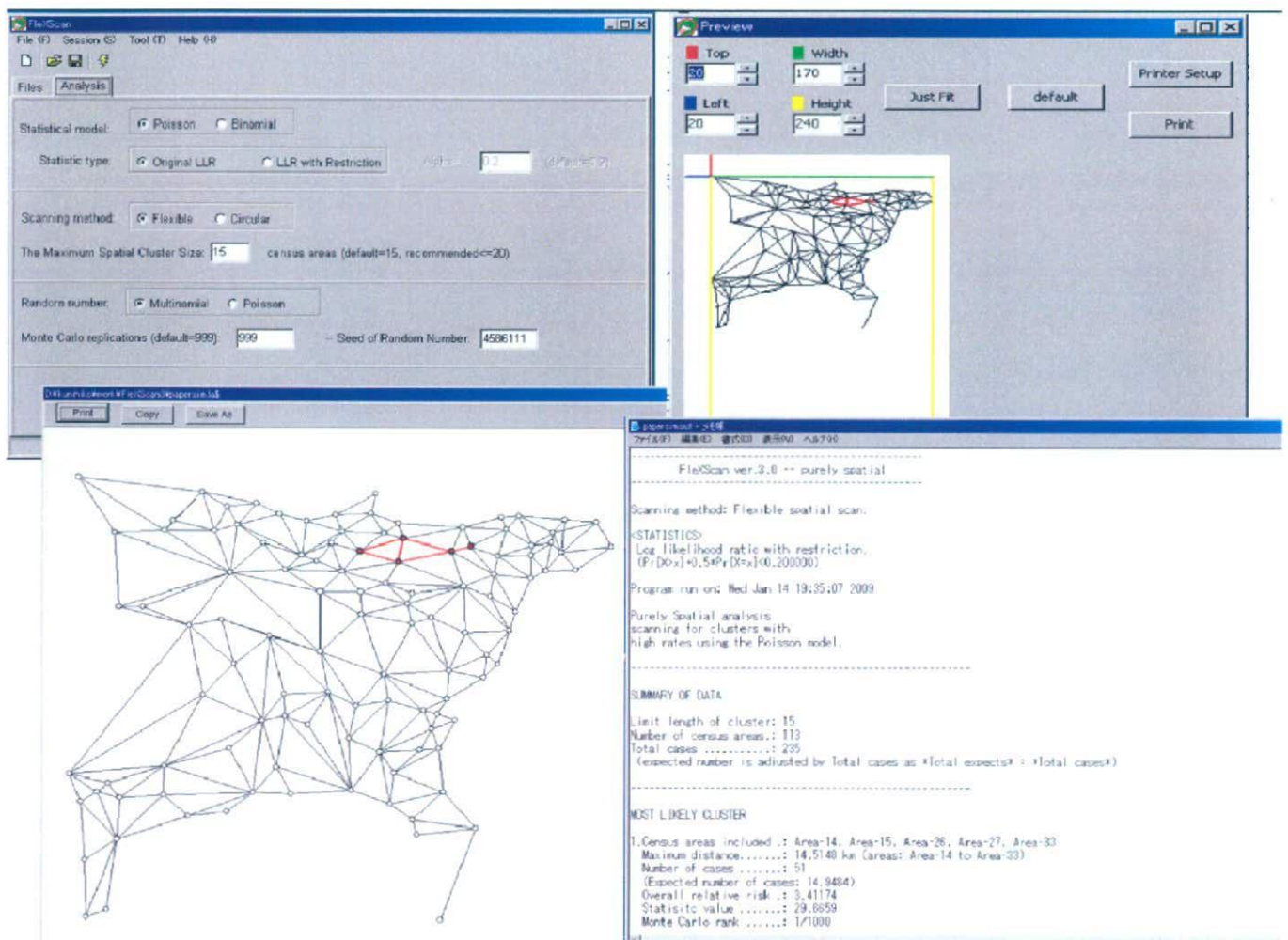


図2： FlexScan version 3.0 の画面

(3) 新たな統計量の提案

従来の SaTScan および FleXScan, またそれ以外の提案されている方法でのほとんどが、いずれも同じ尤度比統計量に基づいた統計量での方法となっている。しかし、この統計量では時に大きすぎる地域や、本来検出すべきでない地域までも同定してしまうという問題が指摘されている。そこで、より精度よく、適切に、健康危機事象の発生をその地域とともに検出するため、従来用いられる統計量の改良について検討を行った。特に(a)制限された尤度比統計量、(b)分散安定化変換に基づく統計量などを検討した。シミュレーションによる数値結果などから、従来の方法よりもより真のクラスターを精度よく同定できる様子が確かめられた。この方法およびその結果についても研究発表をおこなった。

D. 考察

上記の研究を通して、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た。特にニューヨーク市のサーベイランスデータへの適用についても、よい結果を示すことがわかった。その結果も、理論的、数値的にはもちろん、実務家が見ても解釈できる妥当なものであることが確かめられた。しかしその一方でその理論的検討から、統計量に関する新たな問題点があることがわかった。これらは従来の方法を含め全ての scan 法による同定で共通の問題となるが、これまでその点に関する議論はほとんど行われていない。そこでその解決を目指して、新たな検定統計量の検討を行った。いくつかの方法について、その精確性が確かめられた。今後さらに改良を行った新たな統計量の開発により、適切なサーベイランスが行えることになると考えられる。

アプリケーションについても身近なユーザーの意見を取り入れながら改良を行った。特に要望のあった2項モデルについても解析が行うことができるようになり、また実際の利用者にと

って必要なアウトプットも行えるように改良を行った。今後ひきつづき、さらにユーザーの意見を取り入れながら、広く、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。

E. 結論

本年度の研究では、米国・ニューヨーク市のサーベイランス解析の担当者との連携によって我々の提案する FleXScan 法をサーベイランスデータに適用することができ、実際の観点からもその有効性が確認できた。統計量の問題点に対していくつかの検討を行い改良された統計量を提案できたが、依然として問題点は残っている。今後、これを解決するため更なる統計量の改良の検討が必要であり、理論的側面、実際の側面から様々な検討を行い、より適切な手法を提案したい。また、アプリケーションソフト FleXScan についても、国内外で徐々に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な改良を続け、公開されることが望まれる。

F. 健康危機情報

なし

G. 研究発表

1. 論文発表

- Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics* 2008, 7:14.
- Tango T. A spatial scan statistic with a restricted likelihood ratio. *Japanese Journal of Biometrics* 2008, 29;75-95.

・高橋邦彦, 横山徹爾, 丹後俊郎. 疾病地図から疾病集積性へ. 保健医療科学 2008, 57(2);86-92.

・相田潤, 森田学, 安藤雄一, 丹後俊郎, 高橋邦彦, 青山旬, 小坂健. 歯科疾患の地域差の検討. 保健医療科学 2008, 57(2);92-98.

・高橋邦彦, 丹後俊郎. 疾病集積性の検定を用いた症候サーベイランス解析. 保健医療科学 2008, 57(2);122-129.

・郡山一明, 片岡裕介, 竹中ゆかり, 浅見泰司, 高橋邦彦, 丹後俊郎. 健康危機管理と小学校欠席状況サーベイランス. 保健医療科学 2008, 2008, 57(2);130-136.

・高橋邦彦. 疾病地図と統計解析. 数理解析研究所講究録 2008, 1603;11-24.

2. 学会発表

・丹後俊郎. 空間疫学—稀少生起事象の集積性の検出のための統計的方法—. 応用統計学会 2008年度年会, 筑波大学, 2008年6月7日.

・Takahashi K and Tango T. A comparison of SaTScan and FleXScan for outbreak detection and monitoring. The 7th Annual International Society for Disease Surveillance Conference, December 3-5, 2008, Raleigh, North Carolina, USA.

・Tango T and Takahashi K. A new space-time scan statistic for timely outbreak detection taking overdispersion into account. The 7th Annual International Society for Disease Surveillance Conference, December 3-5, 2008, Raleigh, North Carolina, USA.

・Goranson C, Takahashi K, Tango T, Cajigal A, Paladini M, Murray EL, Nguyen T, Konty K and Harfisty F. Cluster detection comparison in syndromic surveillance. The 7th Annual International Society for Disease Surveillance Conference, December 3-5, 2008, Raleigh, North Carolina, USA.

3. 著書

・なし

4. その他

・FleXScan v3.0: 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 3.0

高橋邦彦 横山徹爾 丹後俊郎

国立保健医療科学院

2009年3月

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

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

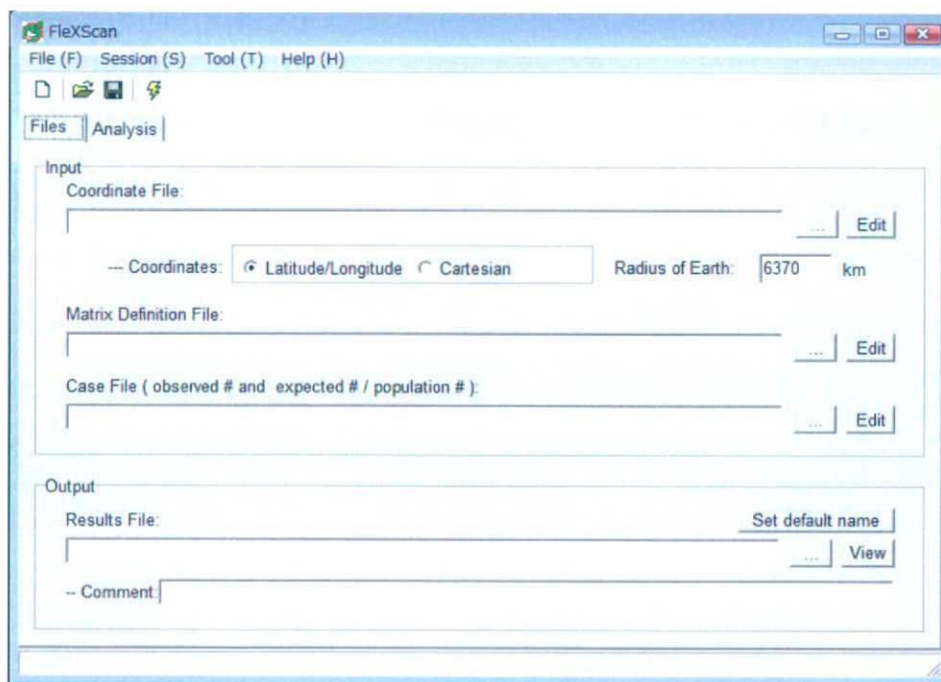
- 丹後俊郎, 横山徹爾, 高橋邦彦 (2007). 空間疫学への招待 (医学統計学シリーズ 7). 朝倉書店.
- Tango T. and Takahashi K. (2005). A flexibly shaped spatial scan statistic for detecting clusters, *International Journal of Health Geographics* 4:11.
- Kulldorff M. and Nagarwalla N. (1995). Spatial disease clusters: Detection and Inference. *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 3.0 (for Windows) は以下の環境を推奨します。

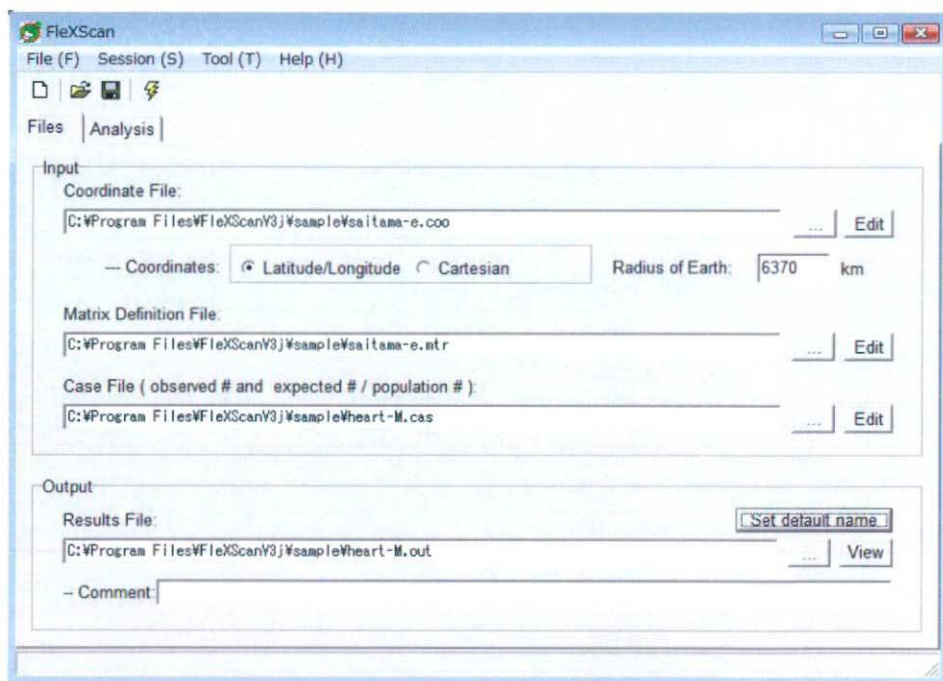
- OS : Microsoft Windows 2000, Windows XP, Windows Vista
- メモリ : 256MB 以上
- CPU : Pentium III 以上



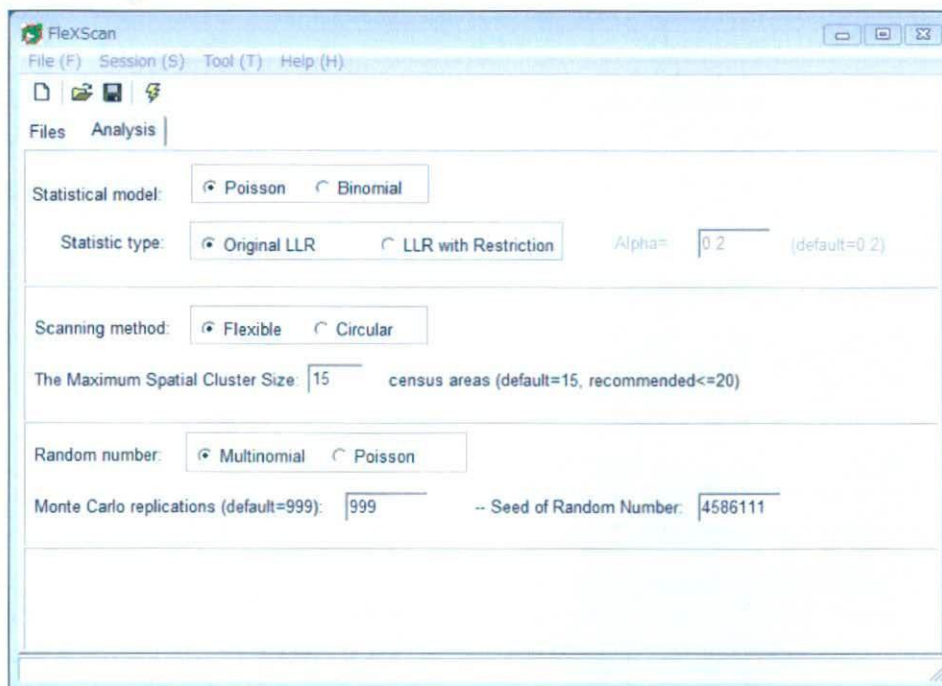
動作確認

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.

-----

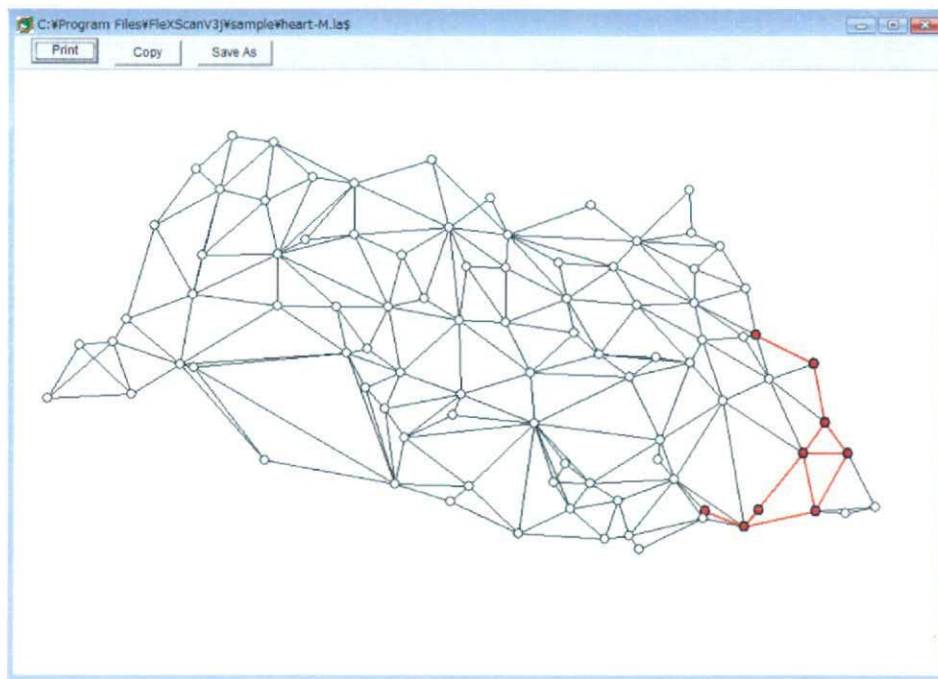
SUMMARY OF DATA

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)
```



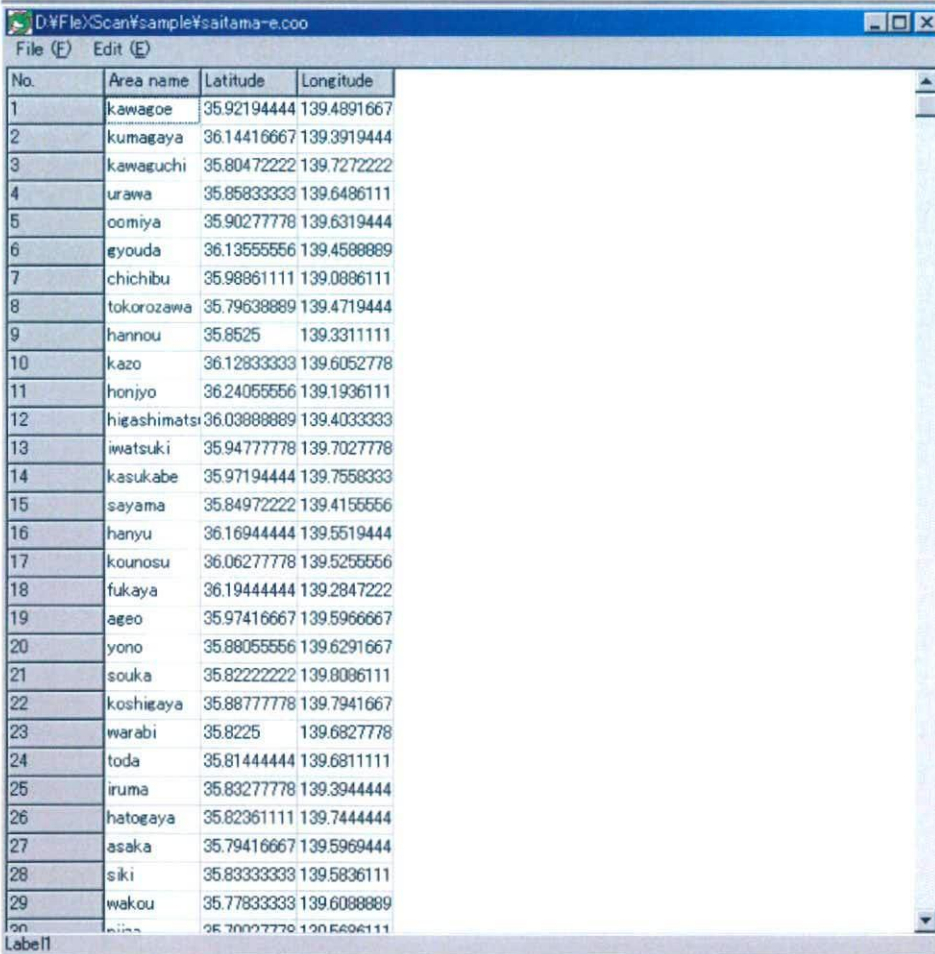
サンプルファイルとデータ形式

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 four columns: "No.", "Area name", "Latitude", and "Longitude". The table lists 29 entries, each with a number, an area name, and its corresponding latitude and longitude values.

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	139.7558333
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	...	35.70027778	139.5606111

- 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。

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-macl	kodama	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。

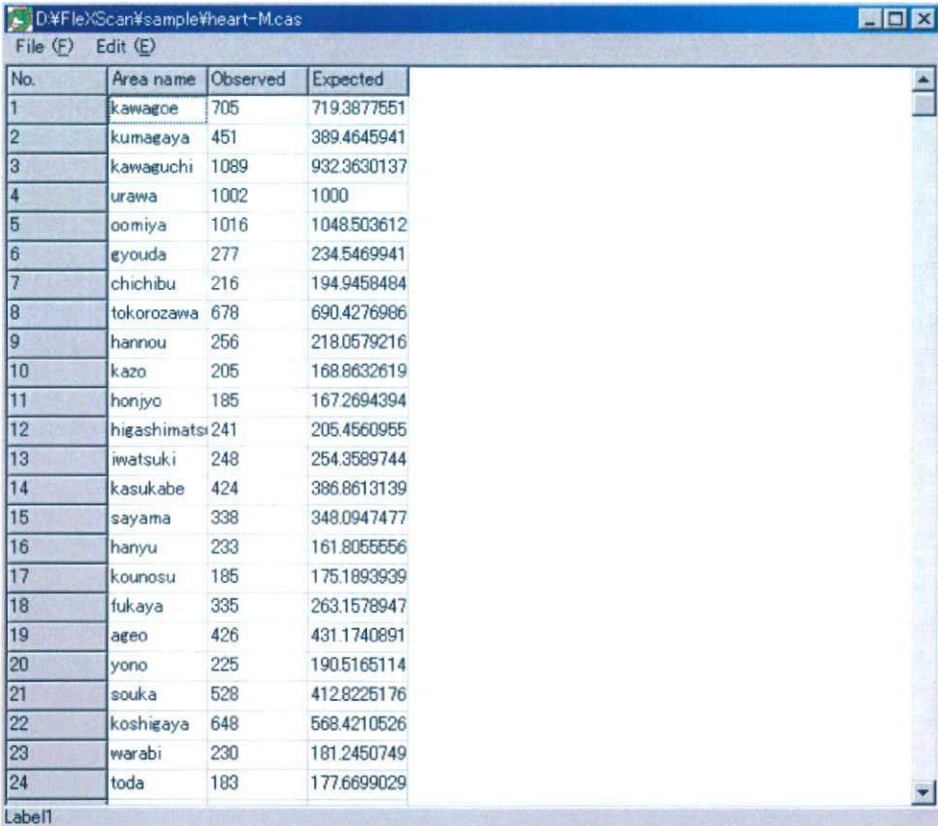
Poisson モデルでの解析を行います。

◎ 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	eyouda	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	higashimatsi	241	205.4560955
13	iwatsuki	248	254.3589744
14	kasukabe	424	386.8613139
15	sayama	338	348.0947477
16	hanyu	233	161.8055556
17	kounosu	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

注意：

- Coordinate File、Matrix Definition File、Case File の全ての<市区町村名>はその順番も含めて完全に統一して下さい。統一されていない場合はエラーが出ます。

- データの値は「半角数字」で入力して下さい。

- Coordinate ファイルで、XY 座標を入力した場合には、「Files」タブの
Coordinates - Cartesian
をチェックしてください。

- Radius of Earth
緯度・経度から距離を求める際に用いる地球の半径です。日本付近では
6370km になります。

ファイルの編集

FleXScan には解析に必要なファイルを編集する際に使用できるツールが用意されています。各ファイルともファイル名を入力して「Edit」ボタンをクリックすることにより編集画面が立ち上がります。Excel 等で入力して、これらの編集画面でコピー・ペーストすることもできます。

- coo ファイル（位置情報ファイル）の編集
 - 各セルに<市区町村名><緯度><経度>を入力して保存します。
- mtr ファイル（接続情報ファイル）の編集
 - まず、先に coo ファイルを完成させて下さい。
 - Area name に（coo ファイルと一致した）市区町村名を入力します。
 - 市区町村を1つ選択して「Edit」－「Area list」を選択すると、他の市区町村が近い順にリストアップされます。ここで接続している市区町村をチェックし「OK」ボタンをクリックすると自動的に Connected セルに追加されます。
 - 全て入力が済んだ後で、「File」－「Check symmetry」を選択すると対称性の確認が出来ます。もし対称でない場合（一方の接続リストに入って他方に入っていない場合）にはエラーの箇所が表示されます。

