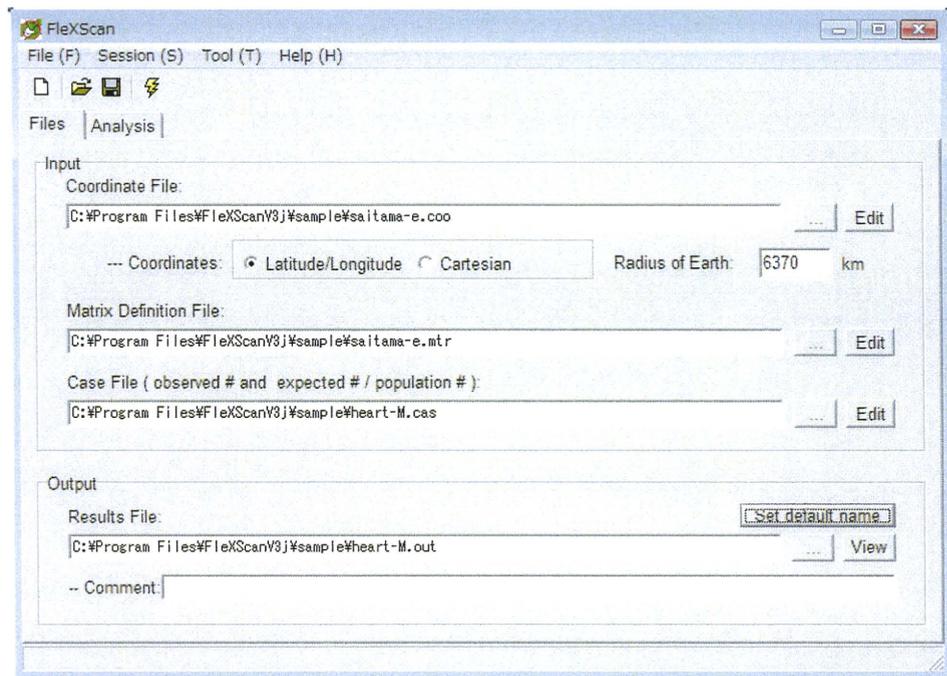


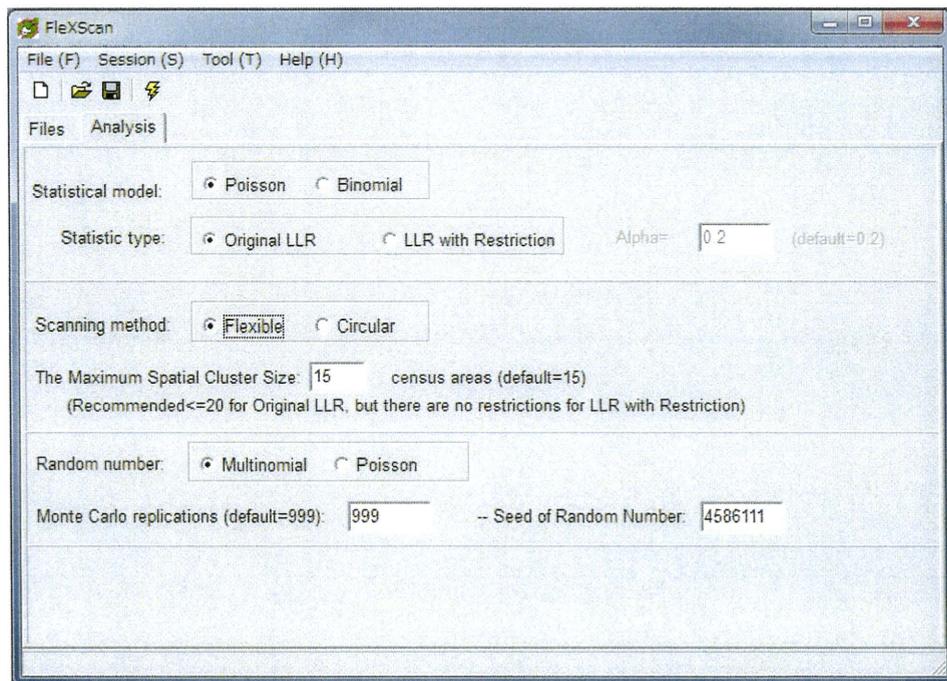
動作確認

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。

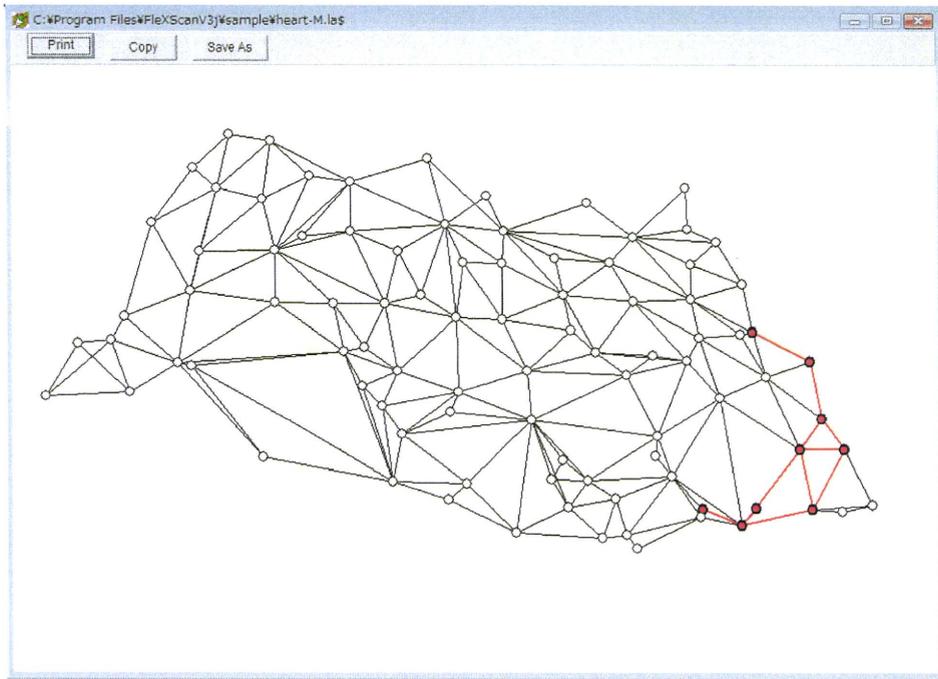
さらに下の

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

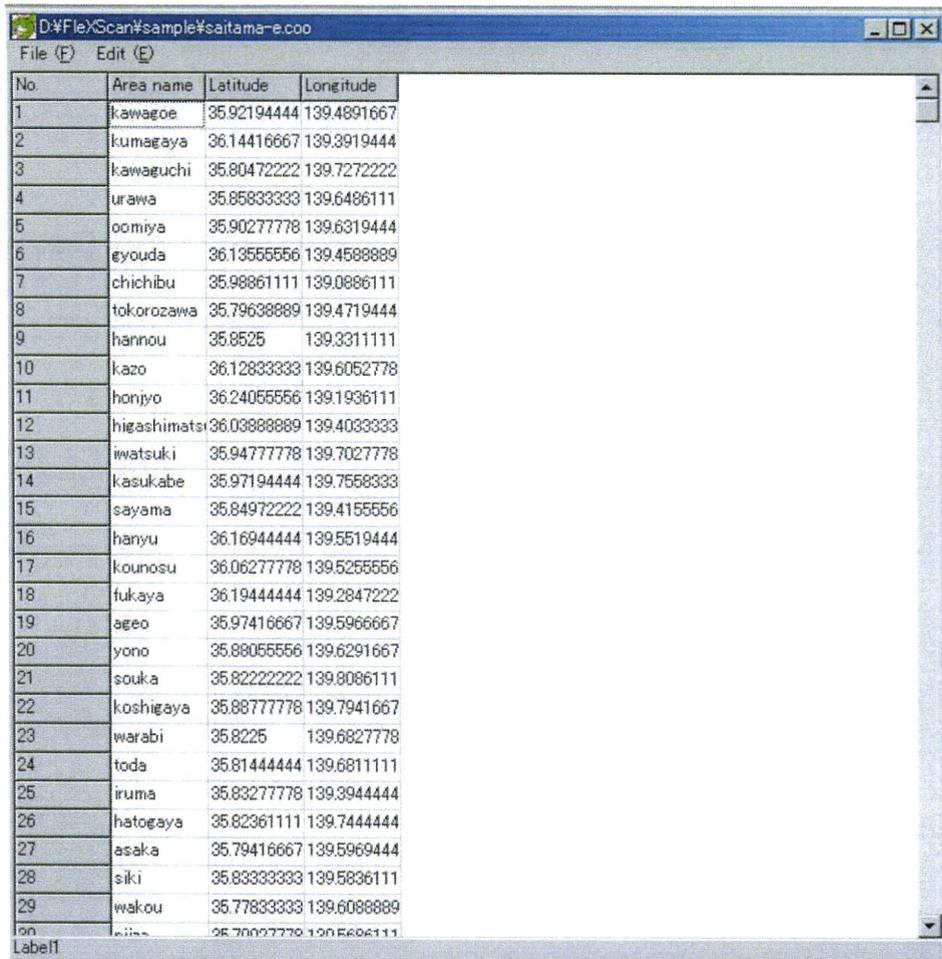
サンプルファイルは埼玉県のデータをもとに作られています。市区町村の区分および市区町村の位置（役所の場所）は「平成12年度版 日本の市区町村 位置情報要覧」（国土地理院、日本地図センター発行）に従っています。

- 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）。



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	eyouda	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	honjo	36.24055556	139.1936111
12	higashimats	36.03888889	139.4033333
13	iwatsuki	35.94777778	139.7027778
14	kasukabe	35.97194444	139.7568333
15	sayama	35.84972222	139.4155556
16	hanyu	36.16944444	139.5519444
17	kunosu	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 9 columns and 23 rows. The columns are labeled "No.", "Area name", and then eight "Connected" columns. The rows contain area names and their connections to other areas.

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	ageo	yono	warabi	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	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になります。

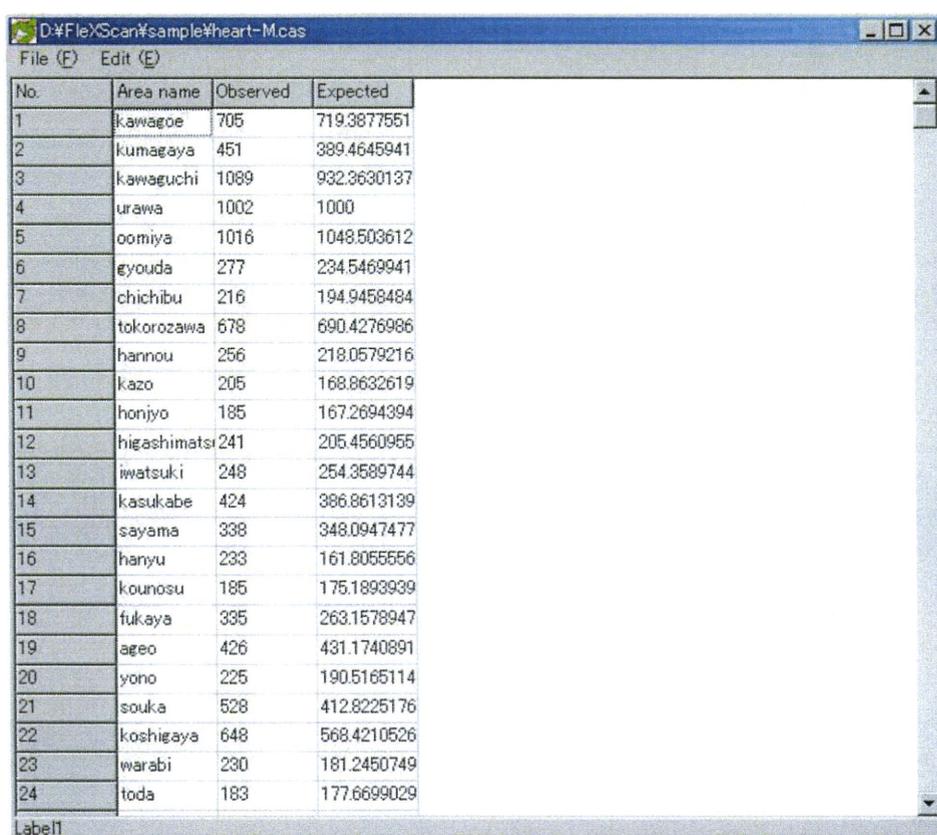
Poisson モデルでの解析を行います。

◎ 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	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。

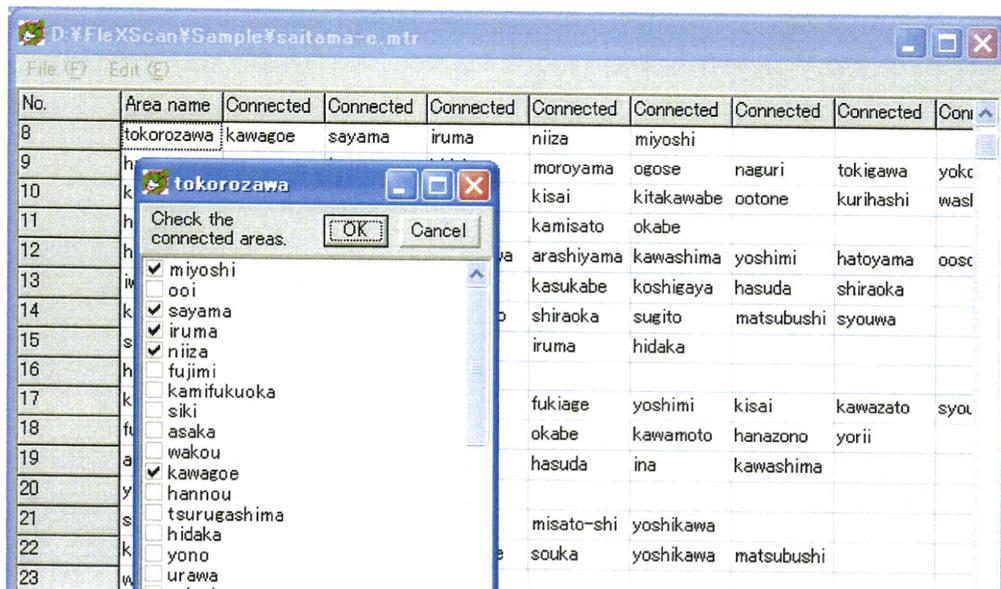
- **Coordinate** ファイルで、XY 座標を入力した場合には、「Files」タブの
Coordinates - Cartesian
をチェックしてください。

- **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。



(補足情報) 実際の計算では、この mtr ファイルの情報をもとに、接続情報行列ファイル (mt0 ファイル) が自動的に作成されます。行列ファイル (mt0) から接続情報ファイル (mtr) への変換も可能です。FlexScan のメニューバー上の「Tool」に変換ツールが入っています。

- cas ファイルの編集

coo ファイルと同様に編集することができます。

パラメータの値

FleXScanにはいくつかのパラメータが用意されています。それらは「Analysis」タブの項目で設定します。

- 統計モデル：Statistical model
 - ① Poisson：用いるデータが「観測度数」と「期待度数」の場合に、その比（O/E比）に基づいた解析を行う Poisson モデル
 - ② Binomial：データが「観測度数」と「対象者数」（人口）の場合にその割合に基づいた解析を行う二項モデル

- 用いる統計量：Statistic type
 - ① Original LLR：Kulldorffによる従来の尤度比統計量
FleXScan version 2 まで用いられていたものはこの統計量です。
 - ② LLR with Restriction：Tangoによる制限付尤度比統計量
この場合制限のパラメータ Alpha を事前に定める（default は 0.2）
この統計量を用いることで多くの地域を同定してしまうことを防ぎ、また計算時間も大幅に速くなります。詳しくは参考文献を参照下さい。

- 検定法の選択：Scanning method
 - ① Flexible：Tango and Takahashi による flexible scan statistic
 - ② Circular：Kulldorffによる scan statistic

- The Maximum Spatial Cluster Size：
検定で用いる統計量の最大連結地域数です。この数を大きくすると広い地域を同定することができるようになりますが、Original LLR では計算時間が長くなります。
詳しくは参考文献を参照下さい。

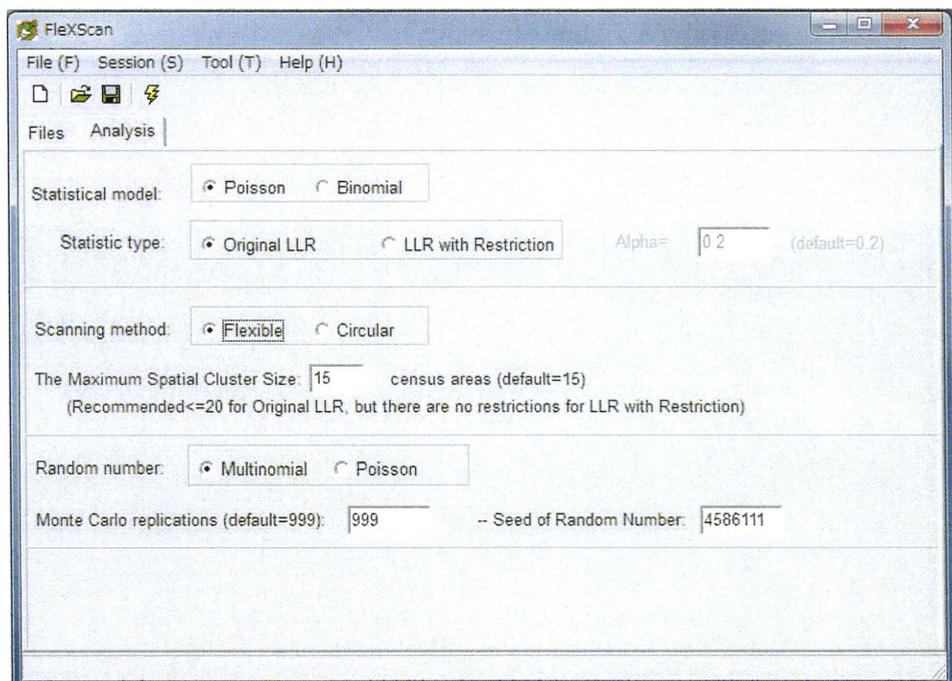
- Random number：モンテカルロシミュレーションに用いる乱数
 - ① Multinomial：多項乱数（観測数の総数を固定した乱数）
 - ② Poisson：ポアソン乱数（総数を固定しない乱数。Poisson モデルで選択可）
 - ③ Binomial：二項乱数（総数を固定しない乱数。Binomial モデルで選択可）

- Monte Carlo replications

検定に用いる p 値を計算するためのモンテカルロシミュレーションの回数です。例えば 999 に設定した際は、999 回のシミュレーションの値と実データからの値の $999+1=1000$ 個の統計量から p 値を求めることになります。

- Seed of Random Number

モンテカルロシミュレーションの乱数を発生させるパラメータです。



参考データの入手

FleXScan を使った解析を体験するために、公開されている統計数値などを用いることができます。以下のようなものを参考にするとよいでしょう。

- 「日本の市区町村 位置情報要覧」
国土地理院作成、(財) 日本地図センター複製発行
- 政府統計の総合窓口 (e-Stat)
<http://www.e-stat.go.jp/>
- 厚生労働統計
<http://www.mhlw.go.jp/toukei/index.html>
- 総務省統計局 統計データ
<http://www.stat.go.jp/>
- 「統計でみる市区町村のすがた」
電子媒体 (財) 統計情報研究開発センター
- 「住民基本台帳人口要覧」 (財) 国土地理協会

使用上の注意

- FleXScan の著作権は高橋邦彦、横山徹爾、丹後俊郎（以下、著作者という）が有します。
- FleXScan は非営利目的であれば誰でも自由に利用することができます。ただし FleXScan の二次配布については著作者の承諾が必要です。
- FleXScan を利用して解析を行った場合には、参考資料として FleXScan を明記して下さい。その際
Takahashi K, Yokoyama T and Tango T. FleXScan v3.1: Software for the Flexible Scan Statistic. National Institute of Public Health, Japan, 2010.
のように引用して下さい。
- FleXScan は予告なしにバージョンアップを行います。最新版は国立保健医療科学院技術評価部ホームページ (http://www.niph.go.jp/soshiki/gijutsu/index_j.html) から入手できます。最近の User Guide、サンプルデータの配布、最新の情報などは全てホームページ上で行います。詳しくはホームページをご覧ください。

FleXScan User Guide

for version 3.1

Kunihiko Takahashi
Tetsuji Yokoyama
Toshiro Tango

National Institute of Public Health

October 2010

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

User Guide version 3.1e

Introduction

The FleXScan software has been developed to analyze spatial count data using the flexible spatial scan statistic developed by Tango and Takahashi (2005) and Kulldorff's circular spatial scan statistic (1997). The current version includes a spatial scan statistic with a restricted likelihood ratio proposed by Tango (2008). It is similar to the SaTScan software (2008) developed by Kulldorff together with Information Management System Inc., but the current version of FleXScan is still restricted to spatial analyses.

Kulldorff's circular spatial scan statistic uses a "circular window" with variable size to define the potential cluster area and so it is difficult to correctly detect noncircular clusters, such as those along a river. In real applications, most geographical areas are noncircular. Tango and Takahashi's flexible spatial scan statistic, therefore, is designed so that the detected cluster is allowed to be flexible in shape, while at the same time the cluster is confined within relatively small neighborhoods of each region. It should be noted that both of these scan statistics are based on maximizing the likelihood ratio. Tango (2000) showed an interesting example in which Kulldorff's circular spatial scan statistic detected an unrealistically large *most likely cluster* (MLC) consisting of 70 regions. This was much larger than expected from an observed disease map, and was due to absorption of neighboring regions with a non-elevated risk of disease occurrence in his simulated data. Furthermore, Tango and Takahashi have shown examples in which Duczmal and Assunção's non-circular scan statistic (2004) detected quite large and peculiar shaped MLC that had the largest likelihood ratio among the MLCs identified by the three different spatial scan statistics: Kulldorff's, Duczmal and Assunção's, and Tango and Takahashi's. These results cast doubt on the validity of model selection based on maximizing the likelihood ratio. Tango (2008) proposed a new spatial scan statistic free from this undesirable property by modifying the likelihood ratio so that it scans only the regions with elevated risk at the "region's significance level of alpha," specified by the user.

The FleXScan software is a free software program designed for any of the following interrelated purposes:

- 1) To evaluate reported spatial disease clusters, to see if they are statistically significant.
- 2) To test whether a disease is randomly distributed over space.

- 3) To perform geographical surveillance of disease, to detect areas of significantly high rates.

FleXScan uses the Poisson model, where the number of events in an area is Poisson distributed according to a known underlying population at risk. This version can also analyze data under the Binomial model. The data may be either aggregated at the census tract, zip-code, county or other geographical level. FleXScan can adjust for the underlying inhomogeneity of a background population and for any number of categorical covariates provided by user.

References

- Tango T and Takahashi K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005, **4**:11. [Open Access] <http://www.ij-healthgeographics.com/>
- Kulldorff M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, **26**:1481-1496.
- Tango T. (2008). A spatial scan statistic with a restricted likelihood ratio. *Japanese Journal of Biometrics* **29**:75-95.
- Kulldorff M and Information Management Services, Inc (2008). SaTScan™ v7.0: Software for the spatial and space-time scan statistics. <http://www.satscan.org/>
- Duczmal L and Assunção R. (2004). A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics & Data Analysis*, **45**, 269-286.
- Tango T. (2000). A test for spatial disease clustering adjusted for multiple testing. *Statistics in Medicine*, **19**, 191-204.

Downloading and Installation

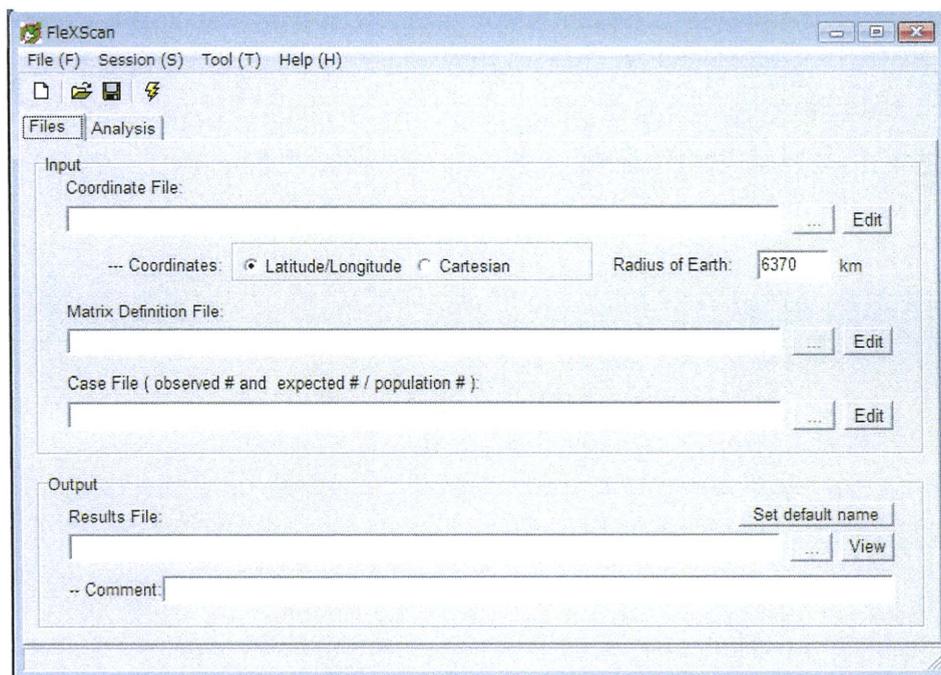
To install FleXScan, go to the Department of Technology Assessment and Biostatistics, National Institute of Public Health Web site at :

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

and select the download link. After downloading the FleXScan installation executable to your PC, click on its icon and install the software by following the step-wise instructions.

System requirements:

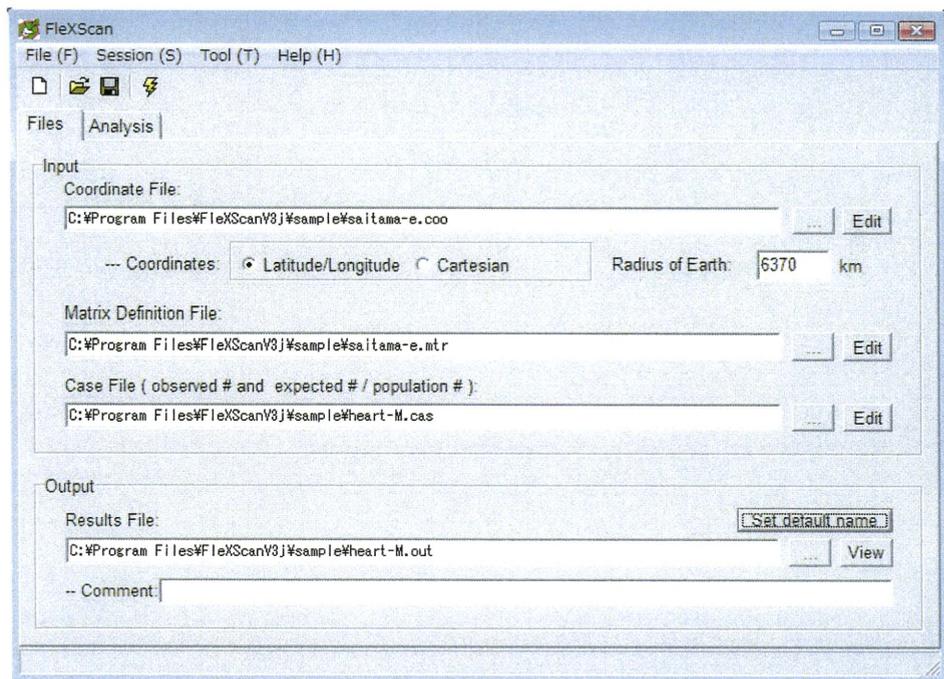
- Intel® Pentium® processor III or higher
- Microsoft® Windows® 2000, Windows XP, Windows VISTA or Windows 7
- 256MB of RAM (recommended)



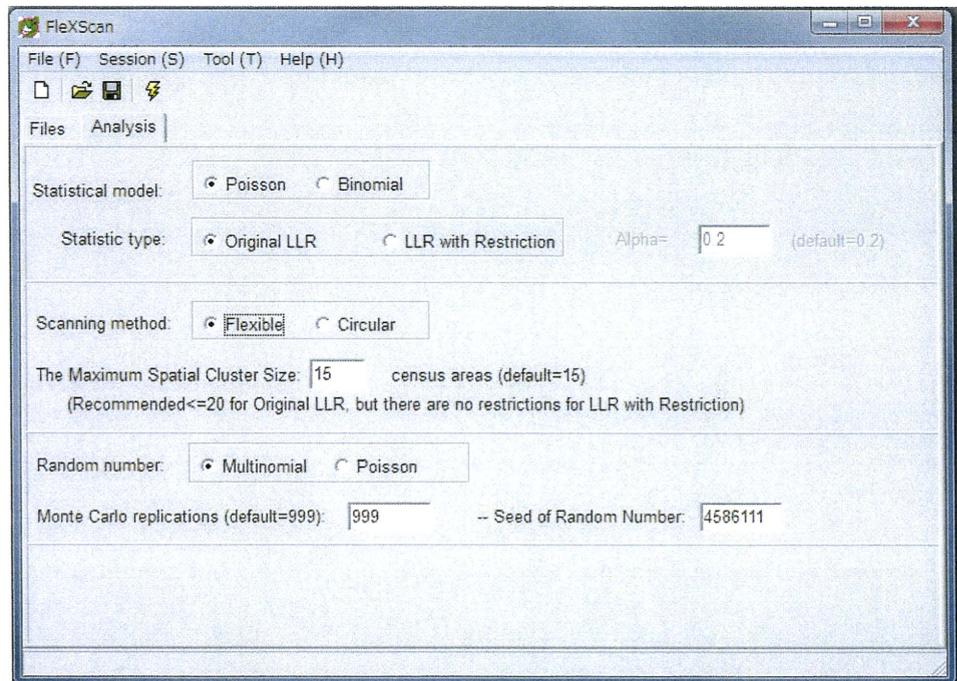
Test Run

Before using your own data, we recommend trying the sample data set provided with the software. To perform a test run:

1. Click on the FleXScan application icon.
2. Select the following files in the sample folder (C:/Program Files/FleXScan/sample, default) as input data.
 - ① Coordinate File: "saitama-e.coo"
 - ② Check 'latitude/Longitude' in the Coordinates
 - ③ Matrix Definition File: "saitama-e.mtr"
 - ④ Case File: "heart-M.cas"
3. Click on 'Set default name' or input any file name to output the results.



4. Next, click on 'Analysis' tab
 - ① Select "Poisson" in the 'Statistical model.'
 - ② Select 'Original LLR' or 'LLR with Restriction' in the Statistic type, and 'Flexible' or 'Circular' in the 'Scanning method', respectively.



5. Click on 'Run' button (⚡) or select 'Session' - 'Run'.
6. After a while, the results file and the results map will appear.

Results:

'MOST LIKELY CLUSTER'

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

are regions detected as the most likely cluster, and its significance is also given by

P-value: 0.001

Regardless of the p-value, the most likely cluster is shown in red on the map.

```
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: Fri Sep 24 17:46:34 2010
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, matsubushi, syouwa
Maximum distance.....: 24.2393 km (areas: kawaguchi to sugito)
Number of cases .....: 3097
```

