FleXScan User Guide

for version 3.0

Kunihiko Takahashi
Tetsuji Yokoyama
Toshiro Tango

National Institute of Public Health

May 2009
http://www.niph.go.jp/soshiki/gijutsu/index_e.html

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

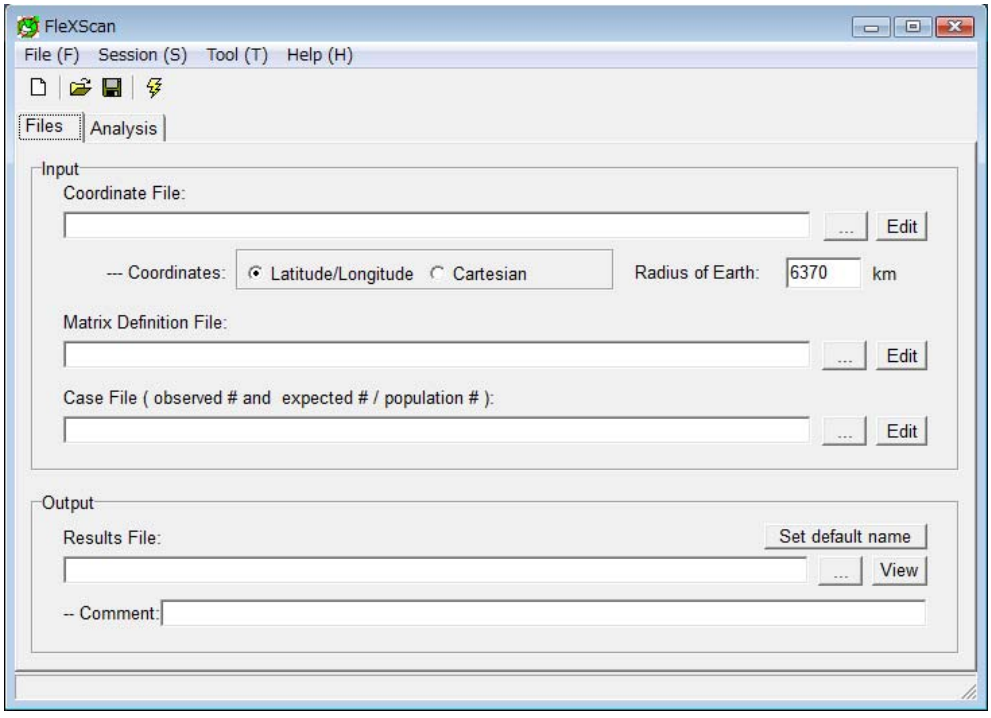
To install FleXScan, go to the Department of Technology Assessment and Biostatistics, National Institute of Public Health Web site at :

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:

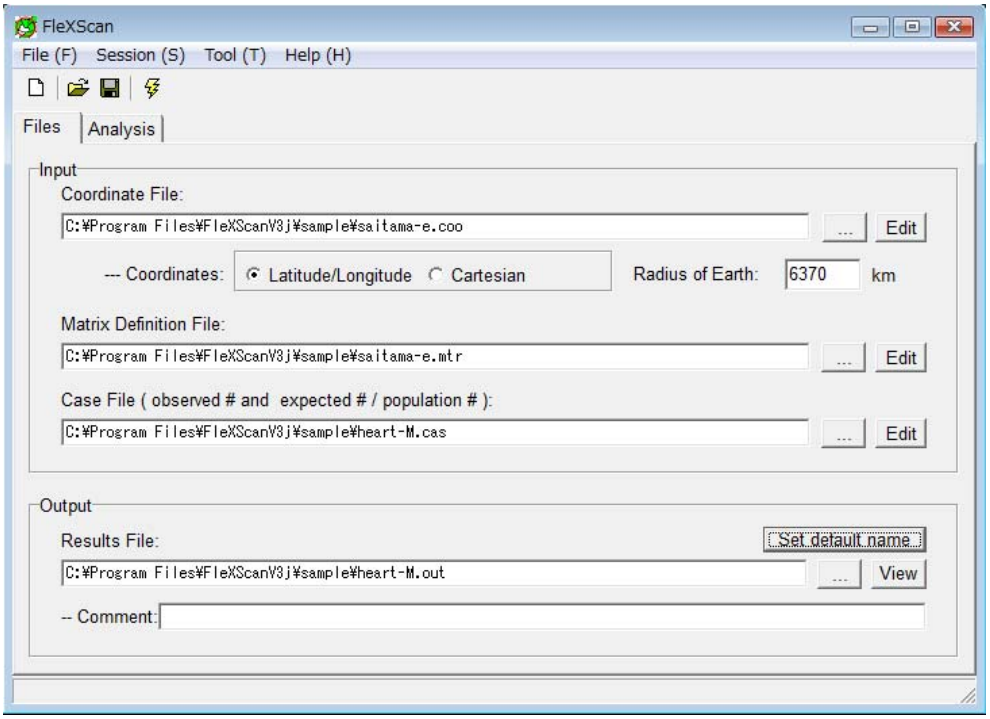
- Intel® Pentium® processor III or higher
- Microsoft® Windows® 2000,Windows XP or Windows VISTA
- 256MB of RAM (recommended)



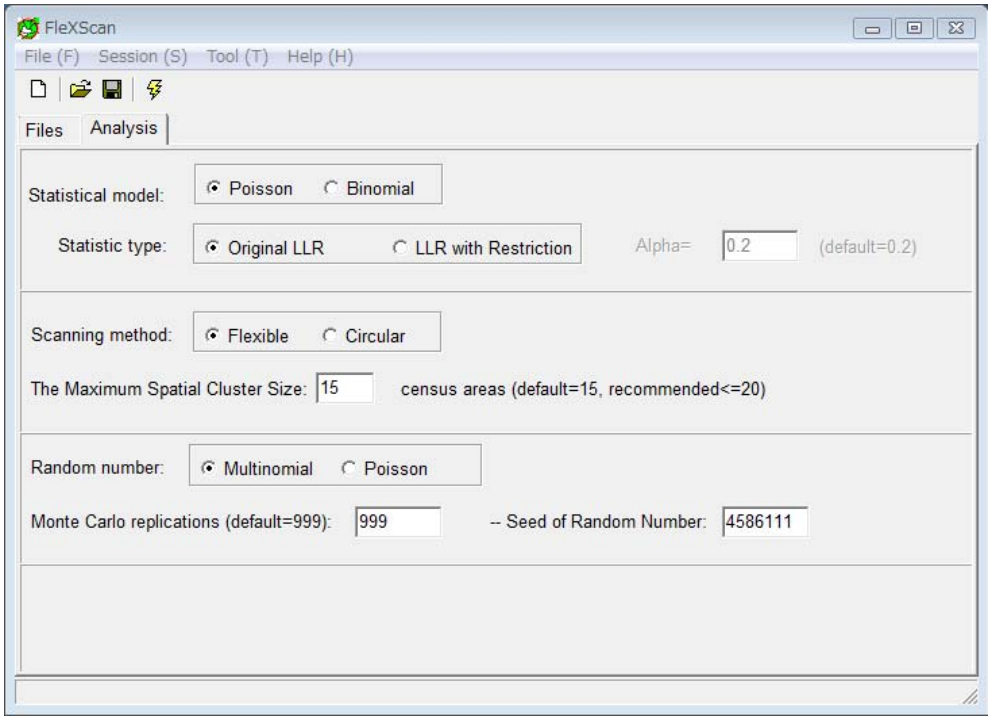
Test Run

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

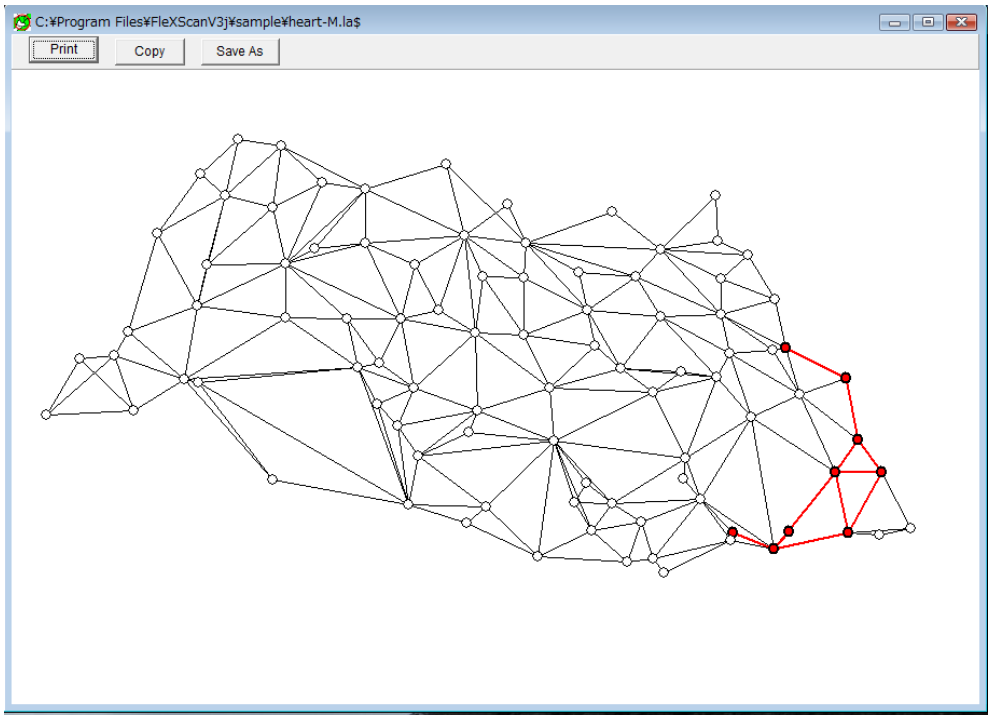
P-value .....: 0.001

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: 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)
```



Sample Data Sets and Data Input Form

To analyze your own data using FleXScan, you need to prepare at least three data files: 1) Coordinate File, 2) Matrix Definition File, and 3) Case File. The detailed structure of each file is explained in the following section showing sample data files for Saitama prefecture in Japan (automatically installed in your 'sample' folder).

① Coordinate File (coo)

The coordinate file provides the geographic coordinates for each area. Coordinates may be specified either using the standard 'Cartesian coordinates' system or in 'latitude and longitude.' 'Cartesian' is the regular planar x,y-coordinate system. Each line of the file represents an area name (or code) and its geographical location.

For the Cartesian coordinates system

- Format : <Area name or code> <X-coordinate> <Y-coordinate>

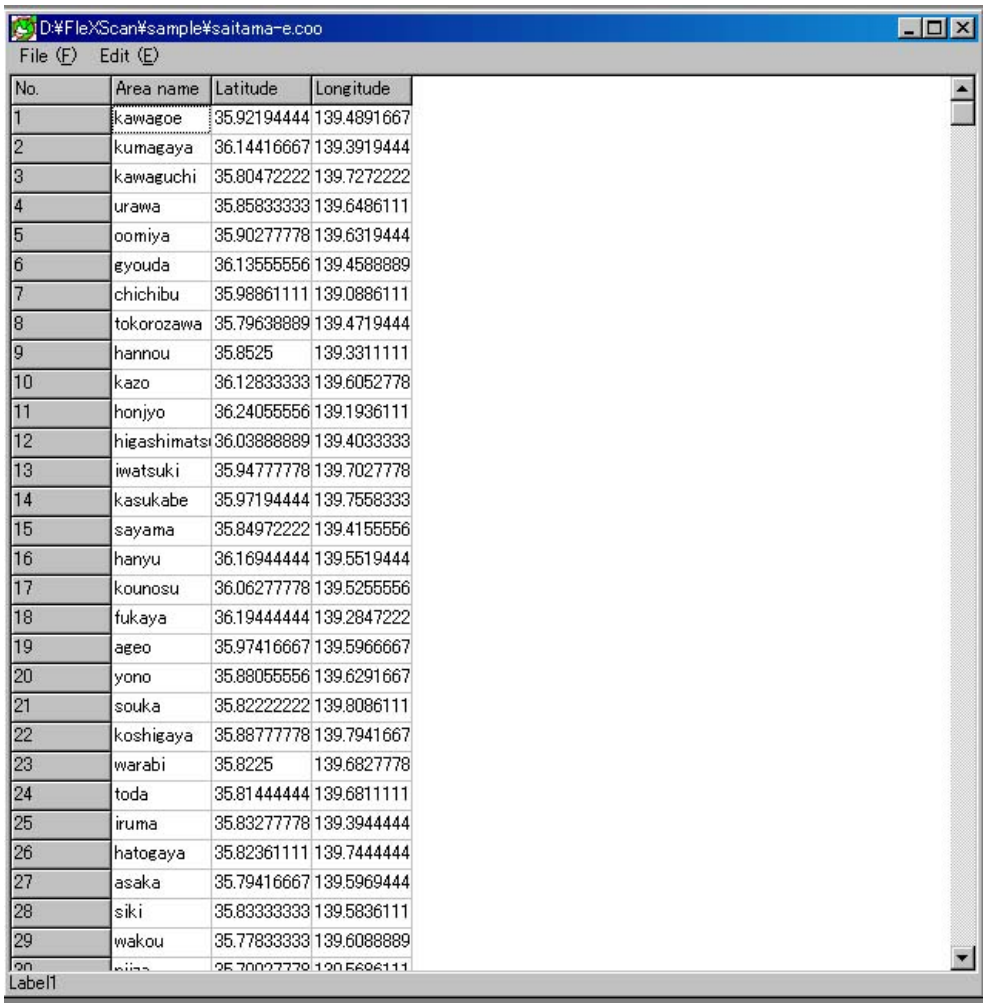
For the Latitude and Longitude coordinates system

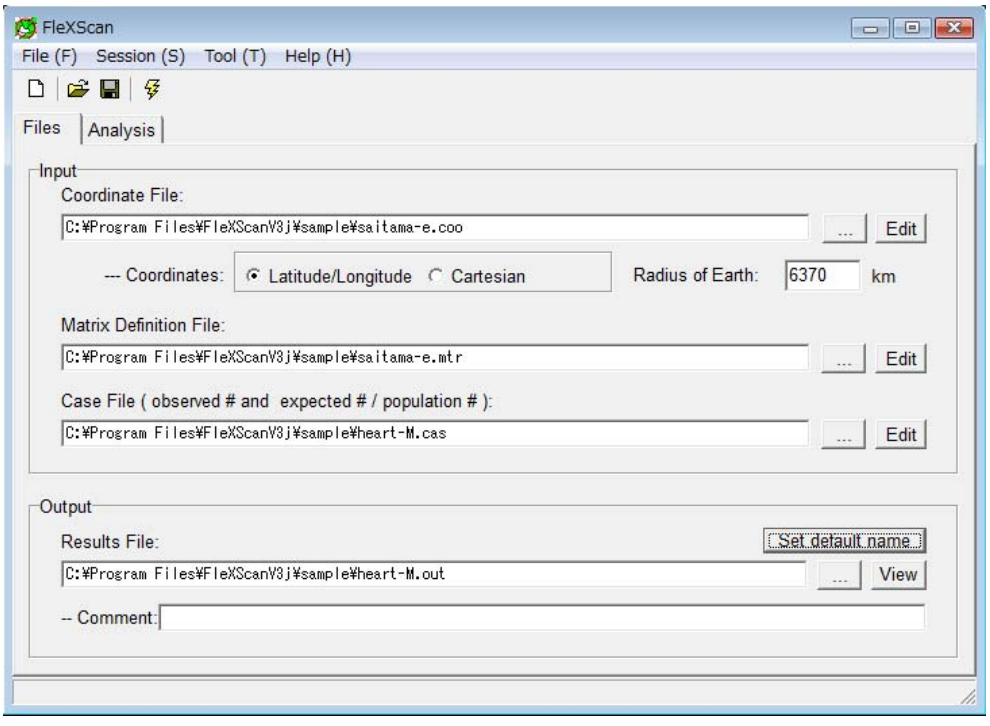
- Format : <Area name or code> <Latitude> <Longitude>

Latitudes and longitudes should be entered as decimal numbers of degrees. You can convert latitudes and longitudes expressed in degrees, minutes, and seconds to decimal number of degrees by the following formula:

xx (degrees) yy (minutes) zz (seconds) → xx + yy/60 + zz/3600 (degrees).

When coordinates are specified in latitudes and longitudes, FleXScan calculates the distance between two points on the surface of the spherical earth with a radius given in the 'Radius of Earth.'


D:¥FleXScan¥sample¥saitama-e.coo
File (F) Edit (E)
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 higashimatsu 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
Label1


FleXScan
File (F) Session (S) Tool (T) Help (H)
Files Analysis
Input
Coordinate File:
C:¥Program Files¥FleXScanV3j¥sample¥saitama-e.coo
... Edit
--- Coordinates: Latitude/Longitude Cartesian
Radius of Earth: 6370 km
Matrix Definition File:
C:¥Program Files¥FleXScanV3j¥sample¥saitama-e.mtr
... Edit
Case File (observed # and expected # / population #):
C:¥Program Files¥FleXScanV3j¥sample¥heart-M.cas
... Edit
Output
Results File:
Set default name
C:¥Program Files¥FleXScanV3j¥sample¥heart-M.out
... View
-- Comment:

② Matrix Definition File (mtr)

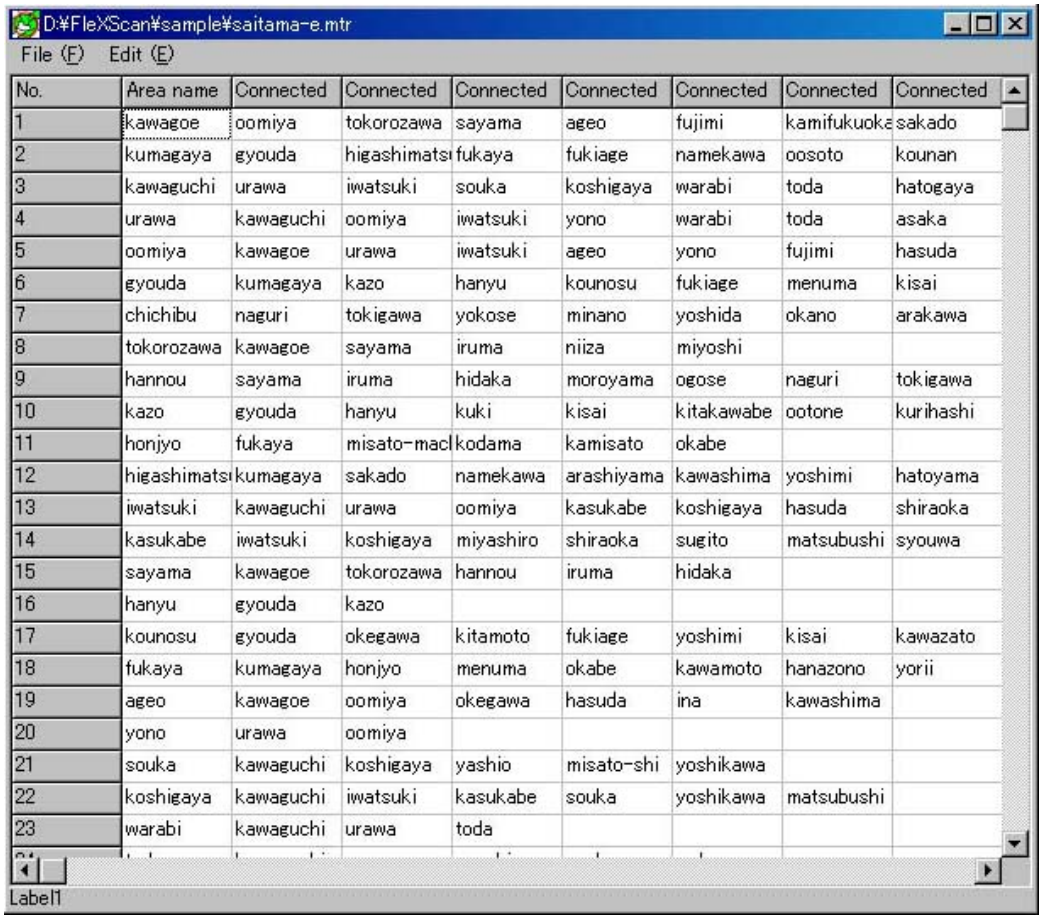
- Format : <Area name or code> <Area 1> <Area 2> …

The first column of each line is the area name, which must be identical to that in Coordinate File. The following columns specify the area name(s) that are adjacent to (i.e., border on) the area described in the first column. For example, kawagoe, sayama, iruma, niiza, and miyoshi areas are adjacent to tokorozawa (see row No.8 of the figure below).

When Area1 is adjacent to Area3 and Area5, the mtr file should be:

Area1 Area3 Area5
Area2 ….
Area3 Area1 …
…

Note that "Area1" also appears in the line of "Area3" in this case (and vice versa), and the matrix must be symmetrical, otherwise an error occurs. The 'Check symmetry' tool in the File menu is available to check the symmetry of the matrix.



No.	Area name	Connected	Connected	Connected	Connected	Connected	Connected	Connected
1	kawagoe	oomiya	tokorozawa	sayama	ageo	fujimi	kamifukuoka	sakado
2	kumagaya	gyouda	higashimatsu	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-mach	kodama	kamisato	okabe		
12	higashimatsu	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				

③ Case File (cas)

The frequency of disease in each area is described in Case File. The current version of FleXScan can analyze two types of data.

① 'observed number' and 'expected number,'

- Format: <Area name or code> <Observed no.> <Expected no.>

For this data, 'Poisson model' should be selected in the 'Statistical model' for the analysis.

② 'observed number' and 'population,'

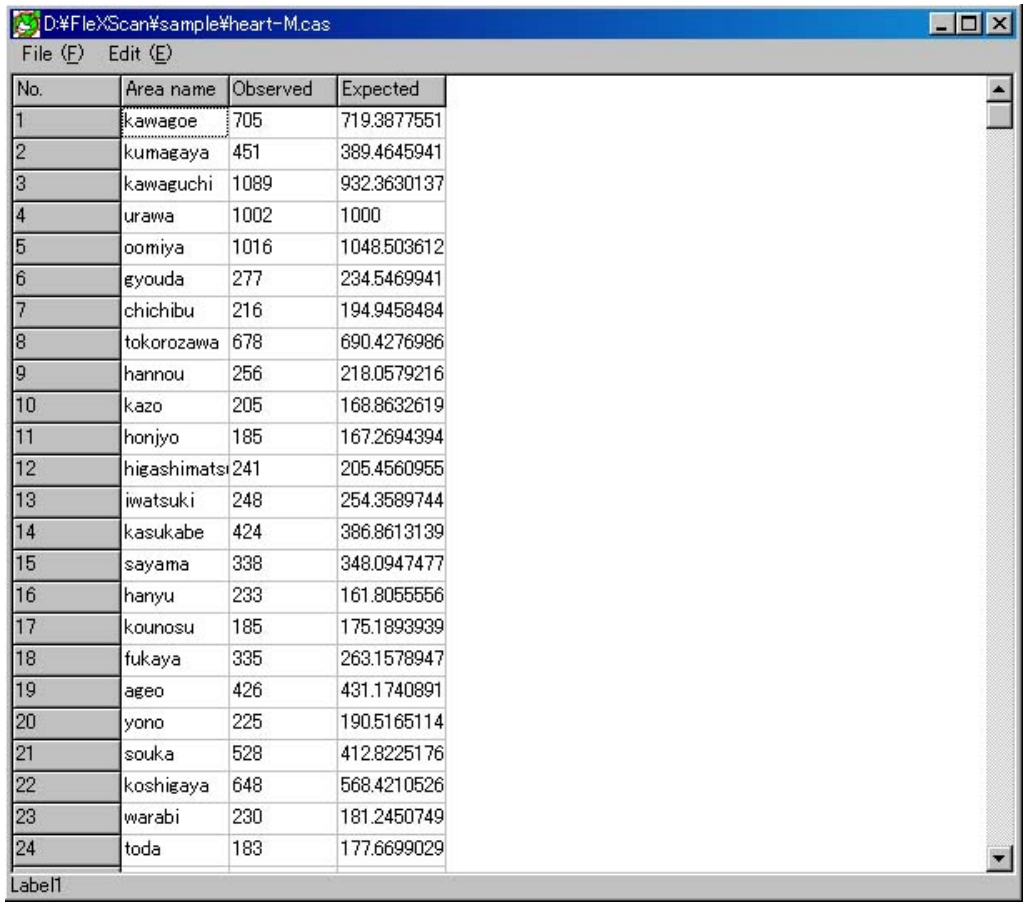
- Format: <Area name or code> <Observed no.> <Population >

For this data, 'Binomial model' should be selected in the 'Statistical model' for the analysis.

The first column of each line is the area name, which must be identical to that in the Coordinate File. The second column is the observed number of diseases, and the third column is the expected number of diseases under the null hypothesis, or the background population at risk in each area.

When you use the Poisson model, you need to calculate the expected number by yourself, for example, in the same manner as standardized mortality ratio (SMR). In kawagoe (see No.1 row of the figure) there were 705 deaths and the age-standardized expected number of deaths was 719.4 (i.e., SMR = 719.4 / 705 = 1.02).

But, if you do not need standardization (e.g., you are interested in the crude death rate), you can analyze the case file of 'observed number' and 'population' using the 'Poisson model.'



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	higashimatsı	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

Important Note:

- All area names or codes and their order must be identical among 'Coordinate File,' 'Matrix Definition File,' 'Case File,' and 'Population File' (if necessary).

- The area name or code cannot include a space character. Use an under score or a hyphen in stead of a space character.

  Good … 10001

  Good … New_York

  N.G. … New York

Editing your data set

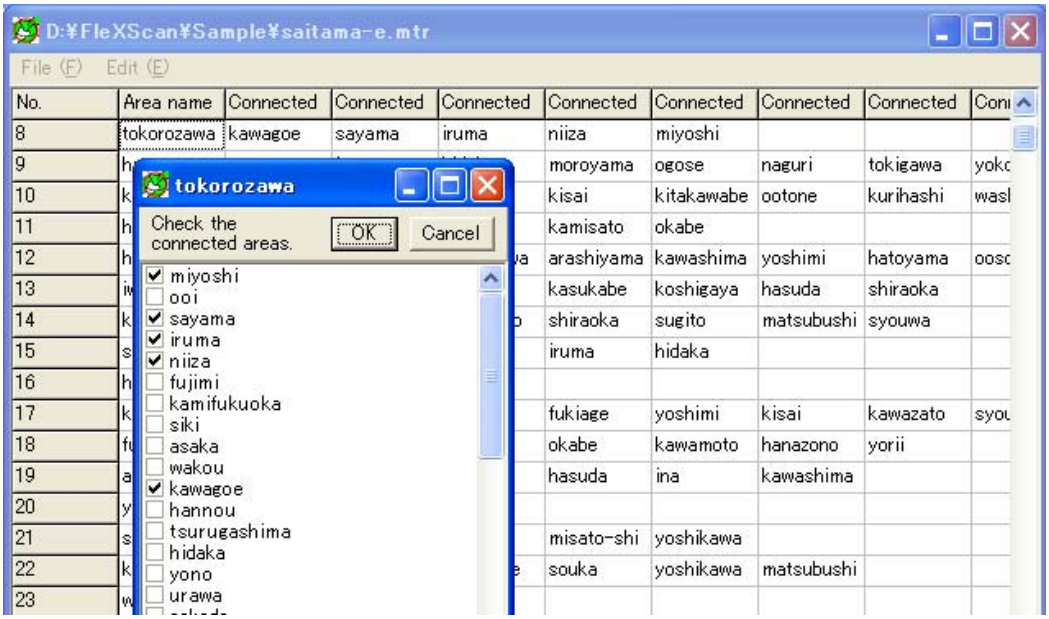
The data files 1) to 3) can be edited using FleXScan data editor. Enter the file name and click the 'Edit' button to execute the data editor. You can copy and paste your data from other software such as MS-Excel. It may be convenient to input your data on MS-Excel and copy and paste it to the FleXScan data editor.

1) Editing Coordinate File (coo).

- Input the area name and its latitude and longitude in each column.
- If x and y-coordinates are used, select 'Cartesian' on the 'Files' tab panel.
- 'Save & return' to finish editing data.

2) Editing Matrix Definition File (mtr).

- Coordinate File must be made before starting to edit Matrix Definition File.
- Input the area name, which must be identical to that of Coordinate File. It will be convenient to Copy & Paste all the area names from Coordinate File to Matrix Definition File.
- Select an area name and execute 'Edit – Area List,' then a list of area names will appear in the order of distance from the selected area. Check the check-box of areas that are adjacent to the selected area. By clicking the 'OK' button, the checked areas will be automatically added to the 'Adjacent' columns (see the figure below).



- The symmetry can be tested by executing 'File – Check symmetry.' If the information is not symmetrical (e.g., Area3 is selected as an adjacent area to Area1, but Area1 is not selected as an adjacent area to Area3), an error message will appear.

Supplementary technical information:

The Matrix Definition File is designed in a user-friendly format. FleXScan will automatically convert it to a Connection Information Matrix File (mt0 file) and utilize it for the calculation. You can see the format of the mt0 file using a text editor. Conversely, if you already have the area connection data in the format of an mt0 file, it can be converted to an mtr file by executing 'Tool – Matrix-file converter.'

3) Editing Case File (cas).

- Coordinate File must be made before starting to edit Case File.
- Case File can be edited in the same manner as Coordinate File.

Parameters

You can change several parameters for the analysis on the 'Analysis' tab panel.

- Statistical model
  - ① Poisson: for the data of the 'observed number' and the 'expected number.'
  - ② Binomial: for the data of the 'observed number' and the 'population.'

- Statistic type
  - ① Original LLR:

    The likelihood ratio statistic by Kulldorff, which has been used in the previous FleXScan version 1 and version 2.
  - ② LLR with Restriction:

    The restricted likelihood ratio statistic by Tango, with a preset parameter for restriction 'Alpha' (default is 0.2). This statistic avoids detecting undesirably large clusters, and improves calculation time. See the reference article for details.

- Scanning method
  - ① Flexible : flexible scan statistic by Tango and Takahashi
  - ② Circular : circular scan statistic by Kulldorff

- The Maximum Spatial Cluster Size

  The number of maximum spatial cluster size to scan. See the reference article for details.

- Random number:

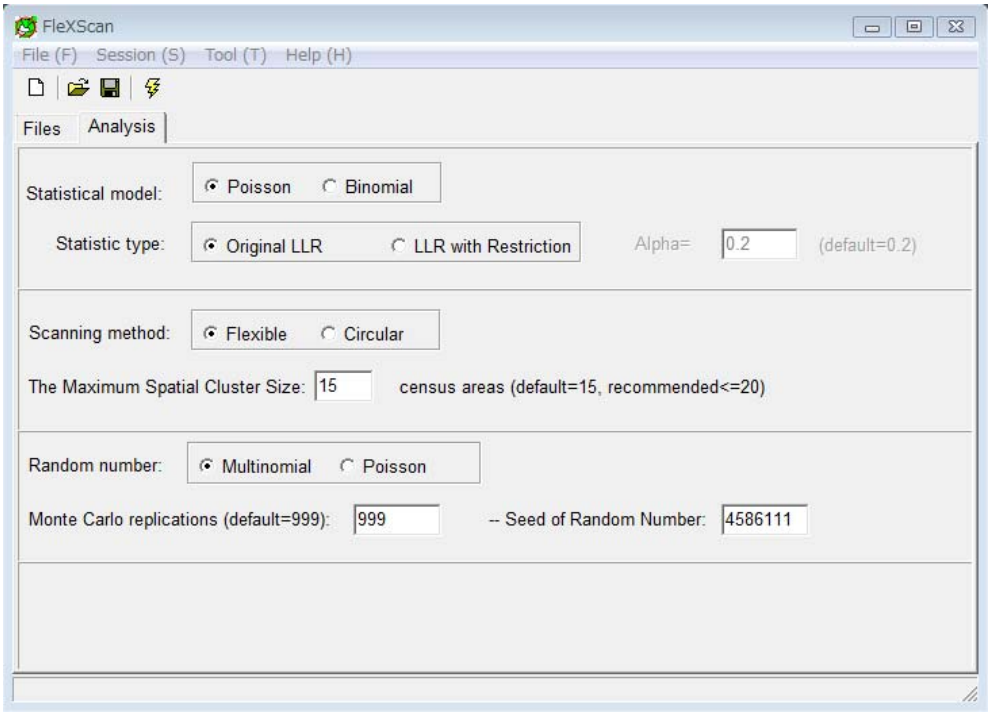
  The type of random number for Monte Carlo simulation
  - ① Multinomial: Total number of cases in whole area is fixed. It can be chosen in either 'Poisson' or 'Binomial' model.
  - ② Poisson: Total number of cases is not fixed, and it can be chosen in 'Poisson model'
  - ③ Binomial: Total number of cases is not fixed, and it can be chosen in 'Binomial model'

- Monte Carlo replications

  The number of Monte Carlo replications to calculate a p-value for statistical test. For example, if this number is set to 999, the p-value is calculated from the simulated 999 plus 1 observed log-likelihood ratio values (999+1=1000 in total).

- Seed of Random Number

  The seed for generating random numbers in the Monte Carlo simulation.



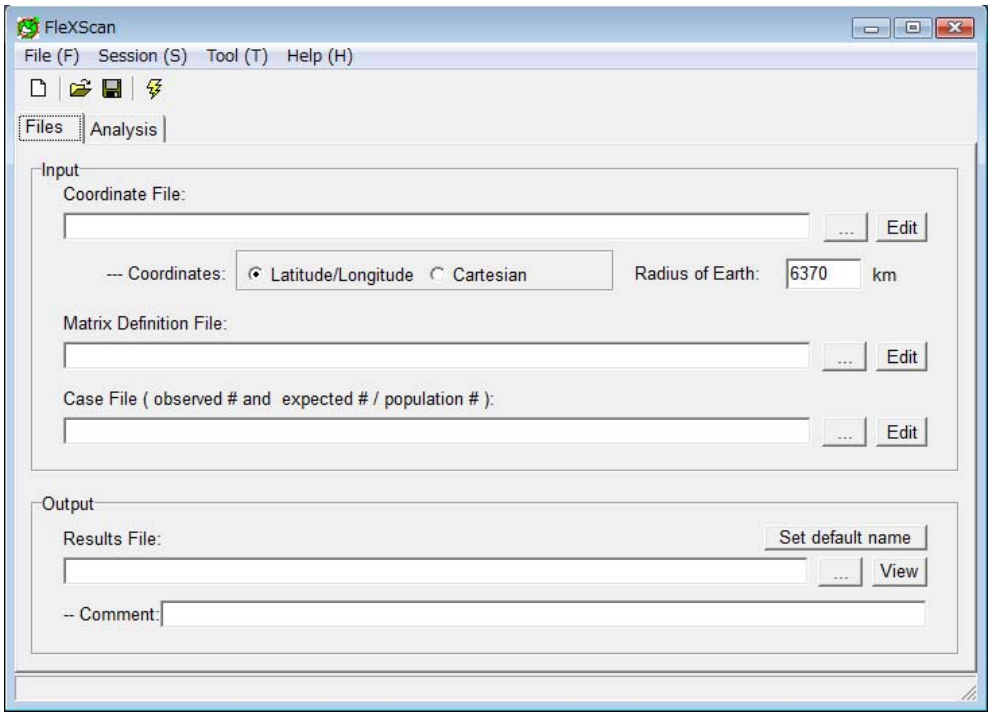
Also, on the 'Files' tab panel,

- Coordinates

  The type of coordinates used by the coordinates file.

- Radius of Earth

  Radius of Earth to calculate a distance between two sets of latitude and longitude. It is approximately 6370 km in Japan.


FleXScan
File (F) Session (S) Tool (T) Help (H)
Files
Analysis
Input
Coordinate File:
...
Edit
--- Coordinates:
Latitude/Longitude
Cartesian
Radius of Earth:
6370
km
Matrix Definition File:
...
Edit
Case File (observed # and expected # / population #):
...
Edit
Output
Results File:
Set default name
...
View
-- Comment:

Software Licenses

- The FleXScan software may be used freely, with proper references to both the software and the statistical methods papers. The suggested citations are:

  - Tango T. and Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. International Journal of Health Geographics 2005, **4**:11.

  - Takahashi K, Yokoyama T and Tango T. FleXScan: Software for the Flexible Scan Statistic. National Institute of Public Health, Japan, 2005 - 2009.

  Also for the spatial scan statistic with a restricted likelihood ratio,

  - Tango T. A spatial scan statistic with a restricted likelihood ratio. Japanese Journal of Biometrics 2008, **29**:75-95.

Copyrights

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