

Introduction to SaTScan, seasonality and time series analysis

Hu Suk Lee

GIS training course for animal health workers
Jakarta, Indonesia
28 November 2016



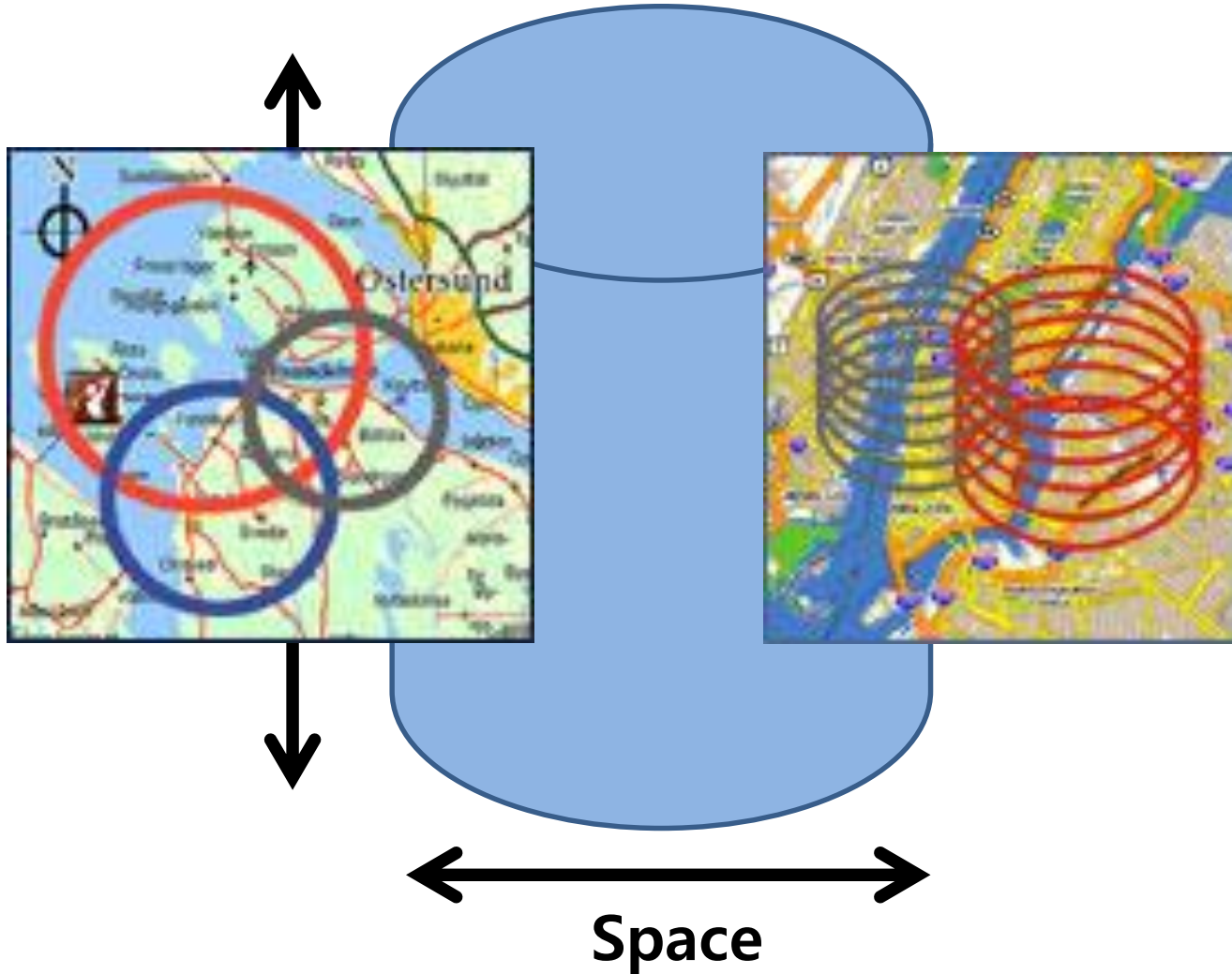
Outline

- Introduction of SaTScan
- Examples
 - Leptospirosis in the USA
 - FMD in Korea
- Introduction of seasonality analysis
- Example
 - Viral Encephalitis in Vietnam
- Introduction of time series analysis
- Example
 - Brucellosis in South Korea

What is SaTscan?

- SaTScan is a freely available software that uses the scan statistic to detect clusters (www.satscan.org)
- To test whether a disease is randomly distributed over space, over time or over space and time.
- To perform geographical surveillance of disease, to **detect areas of significantly high or low rates.**

What is SaTscan?



Scan Statistics

For each distinct window, calculate the likelihood, proportional to:

$$\left(\frac{n}{\mu}\right)^n \left(\frac{N-n}{N-\mu}\right)^{N-n}$$

n = number of cases inside circle

N = total number of cases

μ = expected number of cases inside circle

Scan Statistics

- Circles of different sizes (from zero up to 50 % of the population size)
- For each circle a likelihood ratio statistic is computed based on the number of **observed and expected cases within and outside the circle** and compared with the likelihood L_0 under the null hypothesis.

Scan Statistics

- The **scan statistic** is the maximum likelihood over all possible circles
 - Identifies the most unusual clusters
- To find p-value, use Monte Carlo hypothesis testing
 - Redistribute cases randomly and recalculate the scan statistic many times
 - Proportion of scan statistics from the Monte Carlo replicates which are greater than or equal to the scan statistic for the true cluster is the p-value

What SaTScan can/can't do?

- **CAN**

- Identify spatial, temporal, spatial-temporal clusters
- Provide flexible geographic units

- **CANNOT**

- Display maps of events and clusters locations
- Need GIS or mapping software (such as ArcGIS)
- Create other statistical and regression models

Introduction of Statistical models in SaTScan

Bernoulli Model

- There are animals with or without a disease (represented by a 0/1 variable)
 - A set of cases and controls
- Purely temporal/spatial or the space-time scan statistics

Discrete Poisson Model

- The number of cases in each location is Poisson-distributed.
- Under the null hypothesis, and when there are no covariates, the expected number of cases in each area is proportional to its population size
- Purely temporal, purely spatial and space-time
- This model a very good approximation to the **Bernoulli model** if few cases VS controls (less than 10%)

Space-Time Permutation Model

- Requires only cases data with information about the spatial location and time for each case (**No information needed for population at risk**)
- If the population increase (or decrease) is the same across the study region, that is okay, and will not lead to biased results
- The user is advised to be very careful when using this method for data spanning several years
 - population in some areas grows faster than in others

Screenshot of SaTScan

File Atumn_cases - 메모장

파일(F) 편집(E) 서식(O) 보기(V)

1001	42.062	-72.6255
1002	42.3818	-72.4542
1007	42.2788	-72.4028
1008	42.1881	-72.9663
1020	42.1759	-72.5676
1028	42.0593	-72.4973
1033	42.2579	-72.5066
1035	42.3547	-72.5724
1036	42.0669	-72.4127
1039	42.4043	-72.6883
1040	42.2133	-72.6408
1057	42.0895	-72.3216
1060	42.3161	-72.6276
1069	42.1849	-72.3098
1077	42.0508	-72.7769
1085	42.1511	-72.7653
1095	42.1297	-72.4284
1096	42.4204	-72.776
1104	42.1351	-72.5702
1118	42.0951	-72.5259
1119	42.1239	-72.511
1201	42.4531	-73.2856
1230	42.1906	-73.406
1238	42.2951	-73.2268
1247	42.703	-73.0702
1301	42.628	-72.6009
1351	42.5383	-72.5154
1370	42.5933	-72.727

Input Analysis Output

Case File: ...

Control File: (Bernoulli Model) ...

Study Period

Start Date: Year: 2000 Month: 1 Day: 1 End Date: Year: 2010 Month: 12 Day: 31

Population File: (Poisson Model) ...

Coordinates File: ...

Grid File: (optional) ...

Time Precision

None Year

Month Day

Generic

Coordinates

Cartesian

Lat/Long

Advanced >>

Input

Input Analysis Output

Type of Analysis

Retrospective Analyses:

- Purely Spatial
- Purely Temporal
- Space-Time
- Spatial Variation in Temporal Trends

Prospective Analyses:

- Purely Temporal
- Space-Time

Probability Model

Discrete Scan Statistics:

- Poisson
- Bernoulli
- Space-Time Permutation
- Multinomial
- Ordinal
- Exponential
- Normal

Continuous Scan Statistics:

- Poisson ...

Scan For Areas With:

- High Rates
- Low Rates
- High or Low Rates

Time Aggregation

Units: Year

Month

Day

Length: Days

Advanced >>

Input – Advanced Analysis

The screenshot displays the 'Advanced Analysis Features' dialog box, which is divided into several tabs: 'Spatial Window', 'Temporal Window', 'Space and Time Adjustments', 'Inference', 'Border Analysis', and 'Power Evaluation'. The 'Spatial Window' tab is currently active.

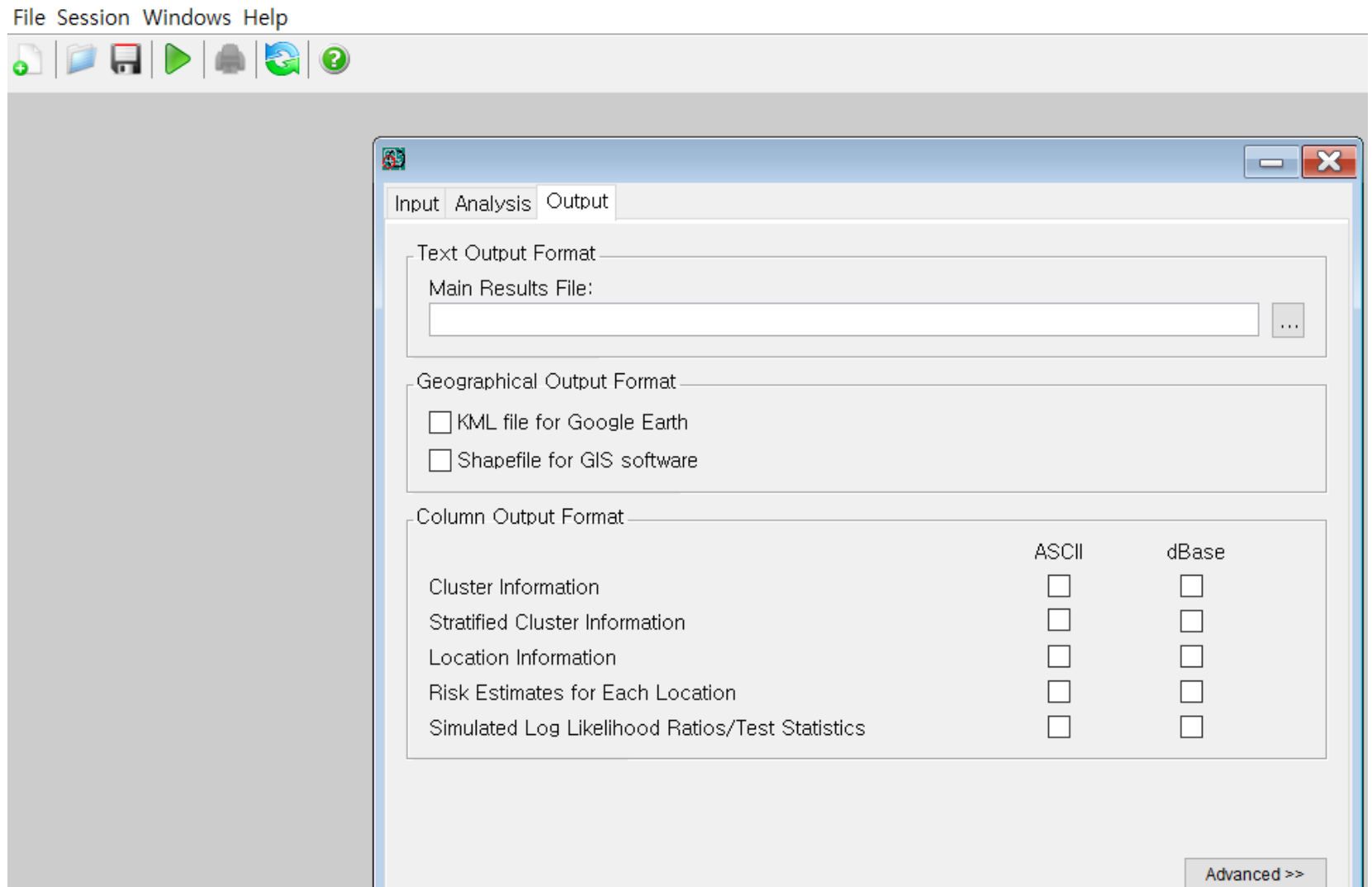
On the left side of the dialog, there are three main sections:

- Input**: Contains sub-tabs for 'Input', 'Analysis', and 'Output'.
- Type of Analysis**: A dropdown menu.
- Retrospective Analyses**: Includes radio buttons for 'Purely Spatial' (selected), 'Purely Temporal', 'Space-Time', and 'Spatial Variation in Temporal Trends'.
- Prospective Analyses**: Includes radio buttons for 'Purely Temporal' and 'Space-Time'.

The main area of the dialog is titled 'Advanced Analysis Features' and contains the following settings:

- Maximum Spatial Cluster Size**: A section highlighted with a red box. It contains a text input field with the value '50.0' and the text 'percent of the population at risk (<= 50%, default = 50%)'. Below this, there is another checkbox option with a text input field set to '50.0' and the text 'percent of the population defined in the max circle size file (<= 50%)', followed by a file selection button (...). A third checkbox option is labeled 'is a circle with a' followed by a text input field set to '1.0' and the text 'kilometer radius'.
- Include Purely Temporal Clusters (Spatial Size = 100%)**: A checkbox that is currently unchecked.
- Spatial Window Shape**: A section containing radio buttons for 'Circular' (selected) and 'Elliptic'. To the right of the 'Elliptic' option is a dropdown menu for 'Non-Compactness Penalty' set to 'Medium'.
- Use Isotonic Spatial Scan Statistic**: A checkbox that is currently unchecked.

Output



Executing

SAIScan - Software for the Spatial and Space-Time Scan Statistic

File Session Windows Help



Input Analysis Output

Case File:
C:\Cluster\All_FMD\Case_FMD.txt ...

Control File: (Bernoulli Model)
... ..

Time Precision
 None Year
 Month Day
 Generic

Study Period
Start Date: Year Month Day End Date: Year Month Day
2010 11 1 2015 4 30

Population File: (Poisson Model)
C:\Cluster\All_FMD\Pop_FMD.txt ...

Coordinates File:
C:\Cluster\All_FMD\Geo_FMD.txt ...

Grid File: (optional)
... ..

Coordinates
 Cartesian
 Lat/Long

Advanced >>

Processing

File Session Windows Help



The main window shows the following options:

- Text Output Format**
 - Main Results File: C:\Cluster\WAIL_FMD\FMD_50_50.txt
- Geographical Output Format**
 - KML file for Google Earth
 - Shapefile for GIS software
- Column Output Format**
 - Cluster Information
 - Stratified Cluster Information
 - Location Information
 - Risk Estimates for Each Location
 - Simulated Log Likelihood Ratios/Test Statistics

The 'Running Session' dialog box displays the following data:

SaTScan log likelihood ratio for # of 999 replications:	Value
SaTScan log likelihood ratio for #585 of 999 replications:	6,14
SaTScan log likelihood ratio for #586 of 999 replications:	7,14
SaTScan log likelihood ratio for #587 of 999 replications:	8,81
SaTScan log likelihood ratio for #588 of 999 replications:	9,35
SaTScan log likelihood ratio for #589 of 999 replications:	9,25
SaTScan log likelihood ratio for #590 of 999 replications:	8,98
SaTScan log likelihood ratio for #591 of 999 replications:	4,80
SaTScan log likelihood ratio for #592 of 999 replications:	8,26
SaTScan log likelihood ratio for #593 of 999 replications:	6,36
SaTScan log likelihood ratio for #594 of 999 replications:	7,40
SaTScan log likelihood ratio for #595 of 999 replications:	7,34
SaTScan log likelihood ratio for #596 of 999 replications:	5,99
SaTScan log likelihood ratio for #597 of 999 replications:	7,80
SaTScan log likelihood ratio for #598 of 999 replications:	5,03

Warnings/Errors:

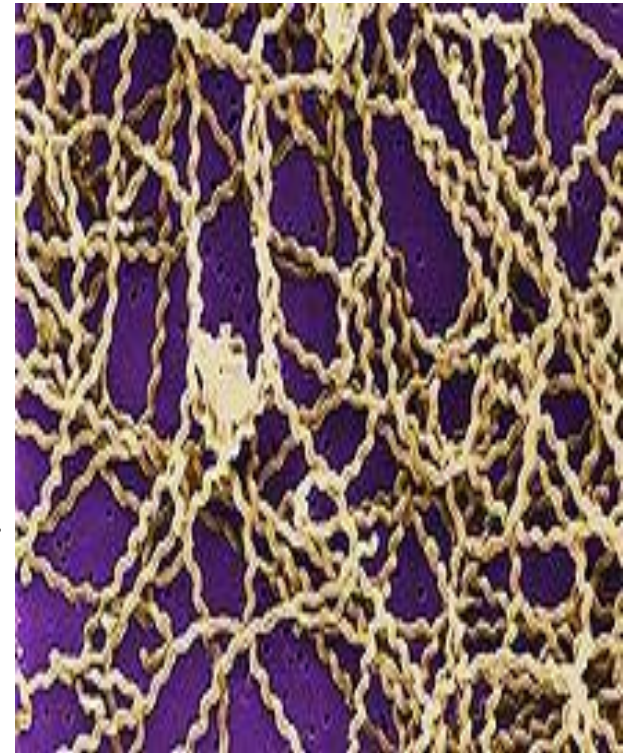
Buttons: Email, Cancel

Advanced >>

Example: Lepto

Leptospirosis

- A bacterial zoonotic disease caused by spirochetes of the genus *Leptospira*
- Pathogenic:
 - *L. interrogans* and *L. kirschneri*
 - More than 200 serovars



Epidemiology

- Zoonotic disease
 - Transmitted to humans from a variety of wild and domesticated animal hosts
 - Most common reservoirs: rodents (rats), wild animals and farm animals in the US
 - Occupations that involve animal handling are more likely to contract disease
- Transmitted through damaged skin or mucus membranes of exposed humans and animals
- Indirect contact (water, soil and feed) with infected urine from an animal with leptospirosis



Host animals

Species	Common infections	Possible others
Dogs	Canicola, Icterohemorrhagiae, Gri ppotyphosa, Pomona	Bratislava, Autumnalis
Cats		rarely identified
Cattle (and deer)	Hardjobovis, Pomona, Grippotyph osa, Icterohemorrhagiae	Australis, Autumnalis, Canicola, Bataviae, Hebdomadis, Kremat osis, Tarassovi, Sejroe, Bratislav a
Pigs	Pomona, Bratislava, Canicola, Tar assovi, Icterohemorrhagiae	Grippotyphosa, Sejroe
Sheep	Pomona, Grippotyphosa, Bratislav a, Hardjo	
Horses	Pomona, Bratislava, Canicola, Icter ohemorrhagiae, Sejroe	

Reference: Bharti, A.R., Nally, J.E., Ricaldi, J.N., Matthias, M.A., Diaz, M.M., Lovett, M.A., Levett, P.N., Gilman, R.H., Willig, M.R., Gotuzzo, E., 2003. Leptospirosis: a zoonotic disease of global importance. *Lancet Infect. Dis.* 3, 757-771.

Objective of study

- To evaluate spatial clustering by each serovar in the United States between 2008 and 2010 using data from IDEXX Laboratories Inc.

Materials and methods

- Procedures

- Spatial clustering analysis (2008-2010)

- Overall (any titers ≥ 1600) / each serovar
 - Spatial window was set at maximum 20%
 - P -value < 0.05

- Geocoding was conducted from hospital zipcodes into geographical coordinates (latitude and longitude)

- Sample locations with clusters were visualized on the map

Materials and methods

- Which model is appropriate?
 - *Discrete Poisson Model*
 - *May consider... Bernoulli Model*
- Software programs
 - Microsoft Excel & Note
 - STATA version
 - ArcGIS (ESRI, CA, USA)
 - SaTScan

Results from Poisson and Bernoulli Models

- 50% of the population at risk area (serovar Autumnalis)
- Only primary cluster is significant in both models

```
77043, 45828, 75071, 75092, 43207, 75042, 75013, 75070, 74070, 770
75228, 77035, 43204, 74006, 77042, 77053, 75081, 43026, 75218, 32
75238, 77079, 34293, 43206, 27560, 75023, 43318, 75009, 26150, 75
27513, 43212, 75025, 75243, 75080, 77077, 27501, 43064, 77084, 43
77072, 43221, 43232, 77433, 75214, 75035, 75231, 75223, 43227, 34
45822, 77082, 77477, 61614, 73446
```

```
Overlap with clusters. No Overlap
Coordinates / radius.. (33.227300 N, 87.578900 W) / 856.51 km
Gini Cluster..... Yes
Population..... 519
Number of cases..... 107
Expected cases..... 54.61
Annual cases / 100000. 20573.9
Observed / expected... 1.96
Relative risk..... 2.08
Log likelihood ratio.. 21.144354
P-value..... 0.00000094
```

```
.Location IDs included.: 1824, 1886, 1852, 1741, 1821, 1879, 1876, 1730, 1460, 1826, 1720,
1742, 1827, 1887, 1803, 1450, 1810, 1719, 1432, 2420, 1844, 2421,
1843, 3060, 3051
```

Poisson Model



```
77043, 45828, 75071, 75092, 43207, 75042, 75013, 75070, 74070, 770
75228, 77035, 43204, 74006, 77042, 77053, 75081, 43026, 75218, 32
75238, 77079, 34293, 43206, 27560, 75023, 43318, 75009, 26150, 75
27513, 43212, 75025, 75243, 75080, 77077, 27501, 43064, 77084, 43
77072, 43221, 43232, 77433, 75214, 75035, 75231, 75223, 43227, 34
45822, 77082, 77477, 61614, 73446
```

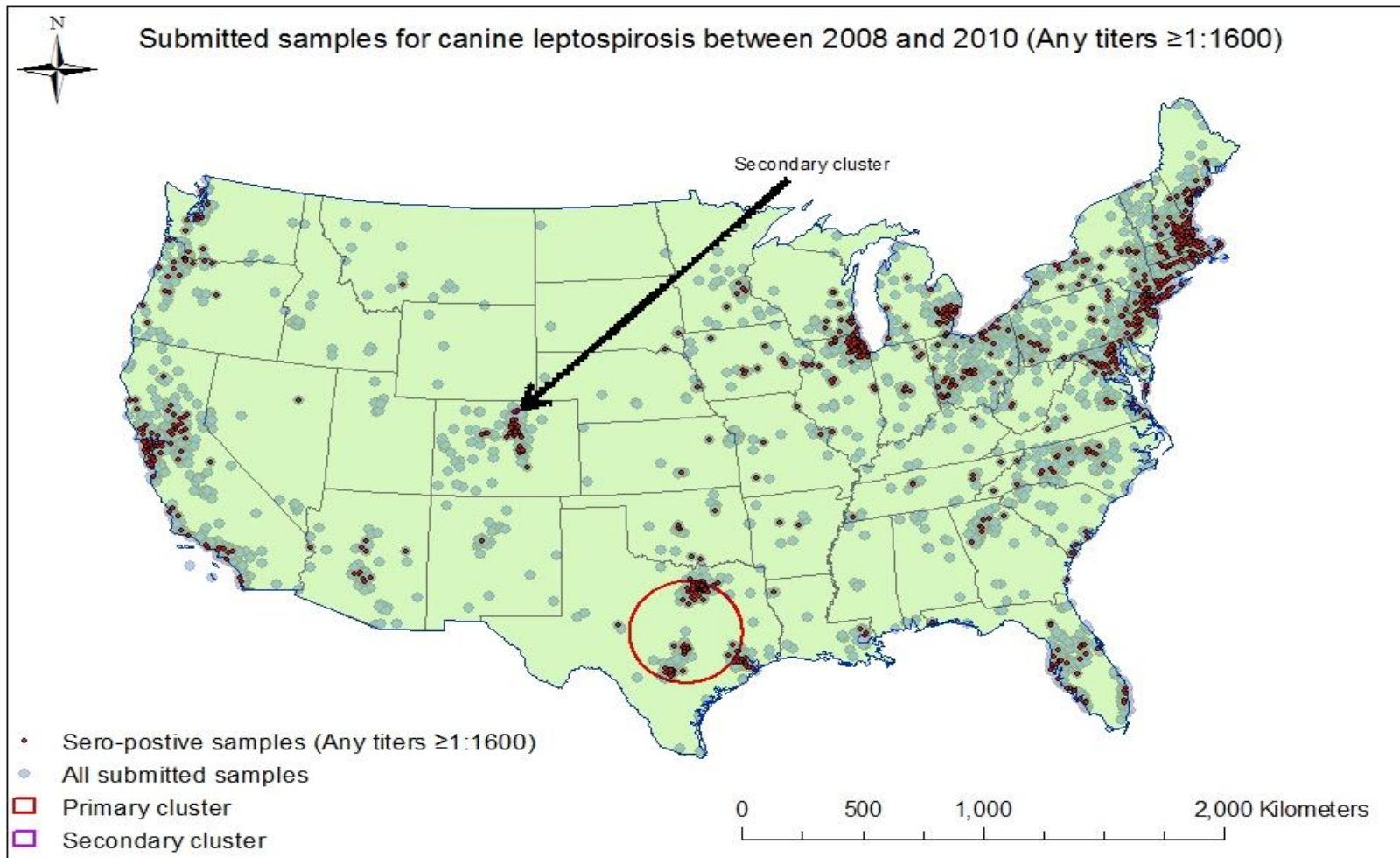
```
Overlap with clusters. No Overlap
Coordinates / radius.. (33.227300 N, 87.578900 W) / 856.51 km
Gini Cluster..... Yes
Population..... 626
Number of cases..... 107
Expected cases..... 59.60
Observed / expected... 1.80
Relative risk..... 1.90
Percent cases in area. 17.1
Log likelihood ratio.. 18.675230
P-value..... 0.000017
```

```
2.Location IDs included.: 59068, 59044, 82414, 59102, 59101, 59718, 59714, 597
59754, 59601, 59405, 59725, 82520, 82604, 83201, 832
82636, 83204, 59802, 59840, 82901, 59875, 59808, 833
57713, 84000, 80000, 80000, 80000, 50000, 57701, 800
```

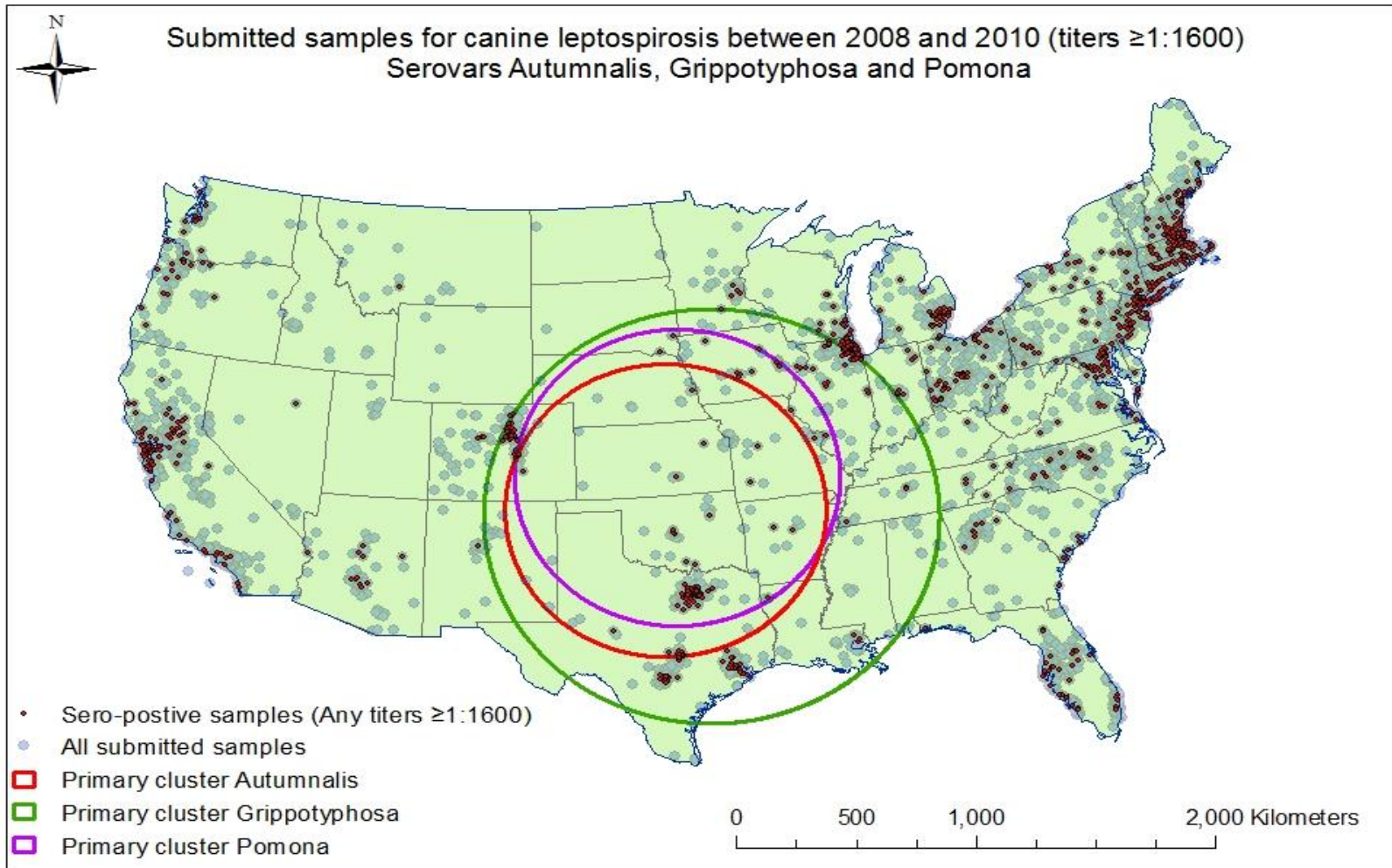
Bernoulli Model Model



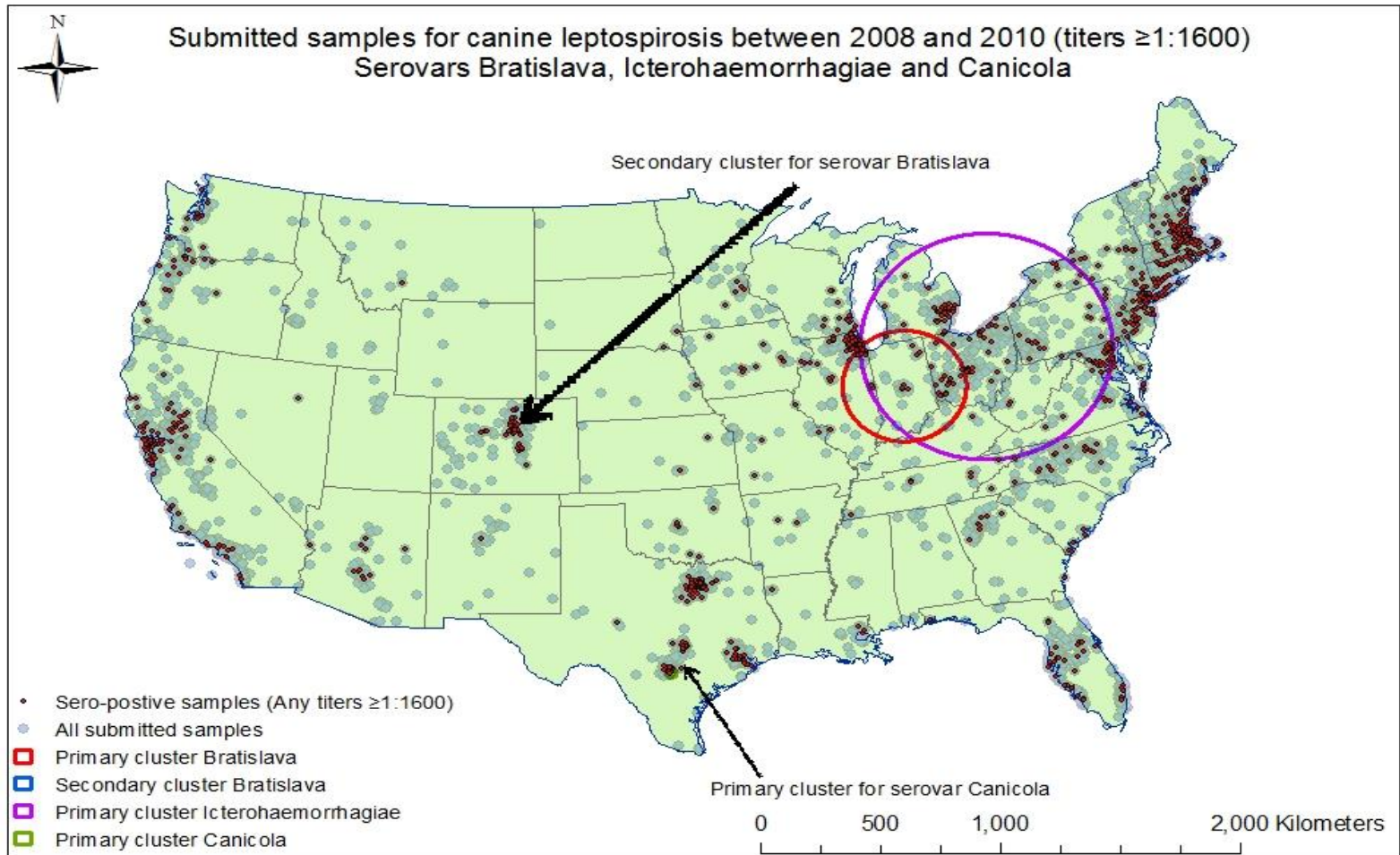
Cluster map from Poisson Model



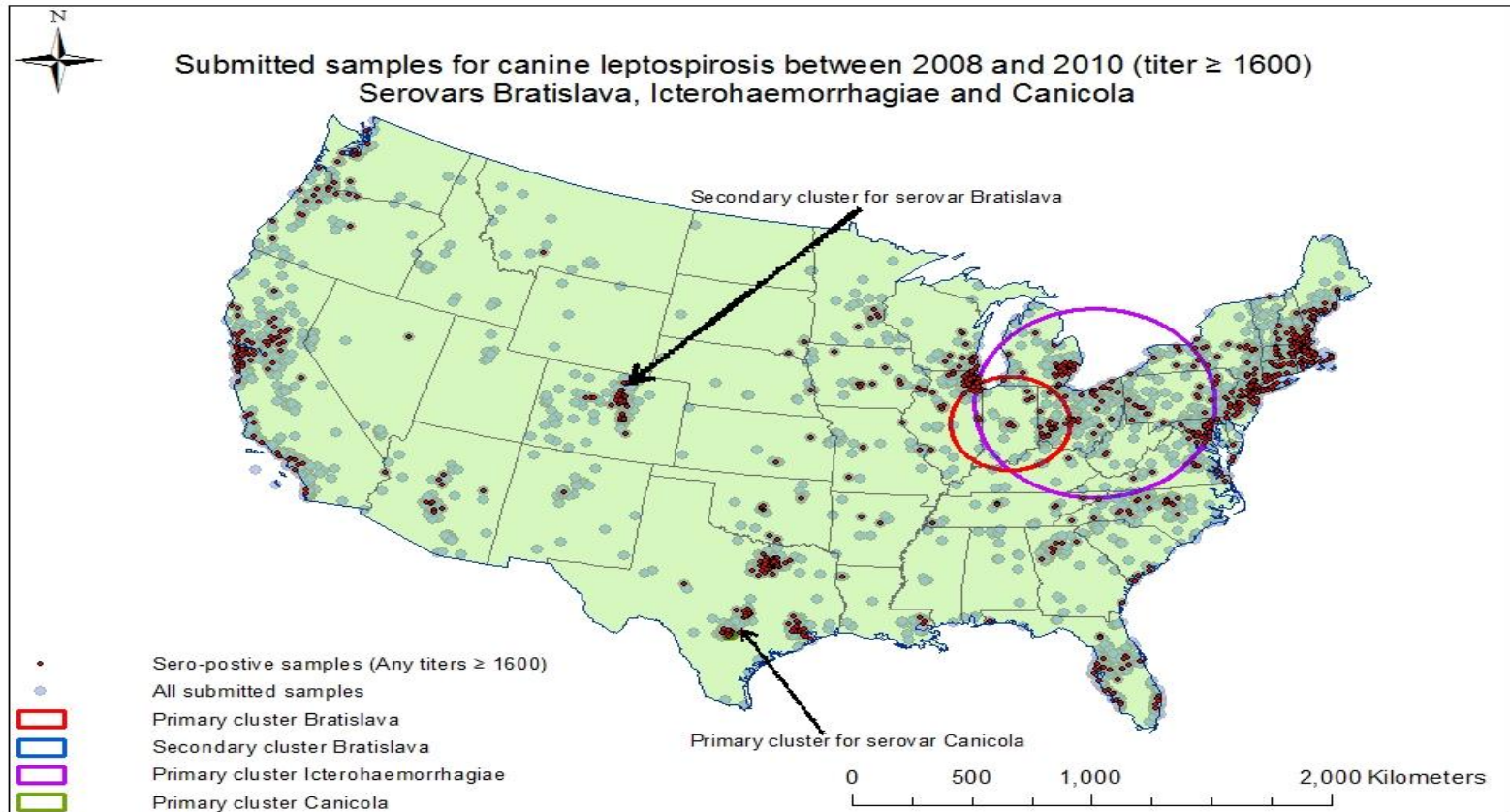
Cluster map from Poisson Model



Discrete Poisson Model



Cluster map from Poisson Model



Spatial clusters of *Leptospira* seropositivity (MAT titers $\geq 1:1,600$) between 2008 and 2010 in dogs in the United States (spatial window: max 20%)

Cluster	Geographical area	Population	Radius (km)	Obs/exp = ratio	P-value
Any titers ≥ 1600					
N=18,717; cases:1,487					
Primary cluster	Central Texas	2,139	232.87	262 / 162.02 = 1.62	<0.001
Secondary cluster	North-central Colorado	83	7.94	22 / 6.29 = 3.50	0.011
Autumnalis					
N=18,632; cases:949					
Primary cluster	Central Oklahoma	2,164	667.02	156 / 102.99 = 1.51	0.011
Grippotyphosa					
N=18,595; cases:795					
Primary cluster	Central Oklahoma	5,970	941.48	358 / 234.14 = 1.53	<0.001
Pomona					
N=18,646; cases:494					
Primary cluster	Central Kansas	2,358	675.28	115 / 59.58 = 1.93	<0.001
Bratislava					
N=18,678; cases:425					
Primary cluster	Central Indiana	919	249.73	44 / 18.69 = 2.35	0.006
Secondary cluster	North-central Colorado	65	4.86	10 / 1.32 = 7.56	0.020
Icterohaemorrhagiae					
N=18,699; cases:172					
Primary cluster	North-central Ohio	5,111	501.41	72 / 38.75 = 1.86	0.001
Canicola					
N=18,698; cases:118					
Primary cluster	South-central Texas	116	22.42	12 / 0.57 = 21.21	<0.001

Example: FMD

FMD

- Foot-and-mouth disease is a highly contagious, viral disease of domestic cloven-hoofed and many wild animals

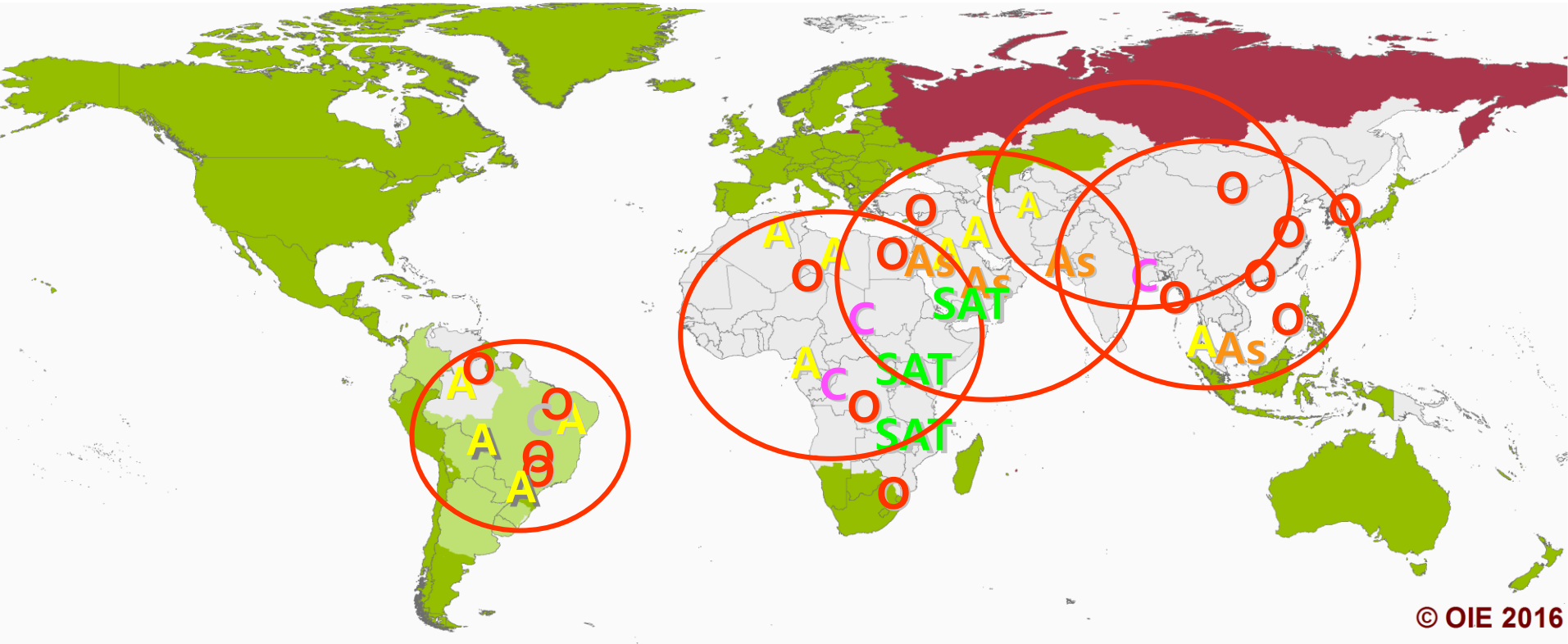


Worldwide occurrence of FMD





OIE Member Countries' official FMD status map

Last update October 2016

[Click on a specific region to zoom in](#)



© OIE 2016

- | | | | |
|--|--|---|--|
|  | Member Countries and zones recognised as free from FMD without vaccination |  | Suspension of FMD free status without vaccination |
|  | Member Countries and zones recognised as free from FMD with vaccination |  | Countries and zones without an OIE official status for FMD |

FMD in Korea

- Data (2010-2015): South Korea

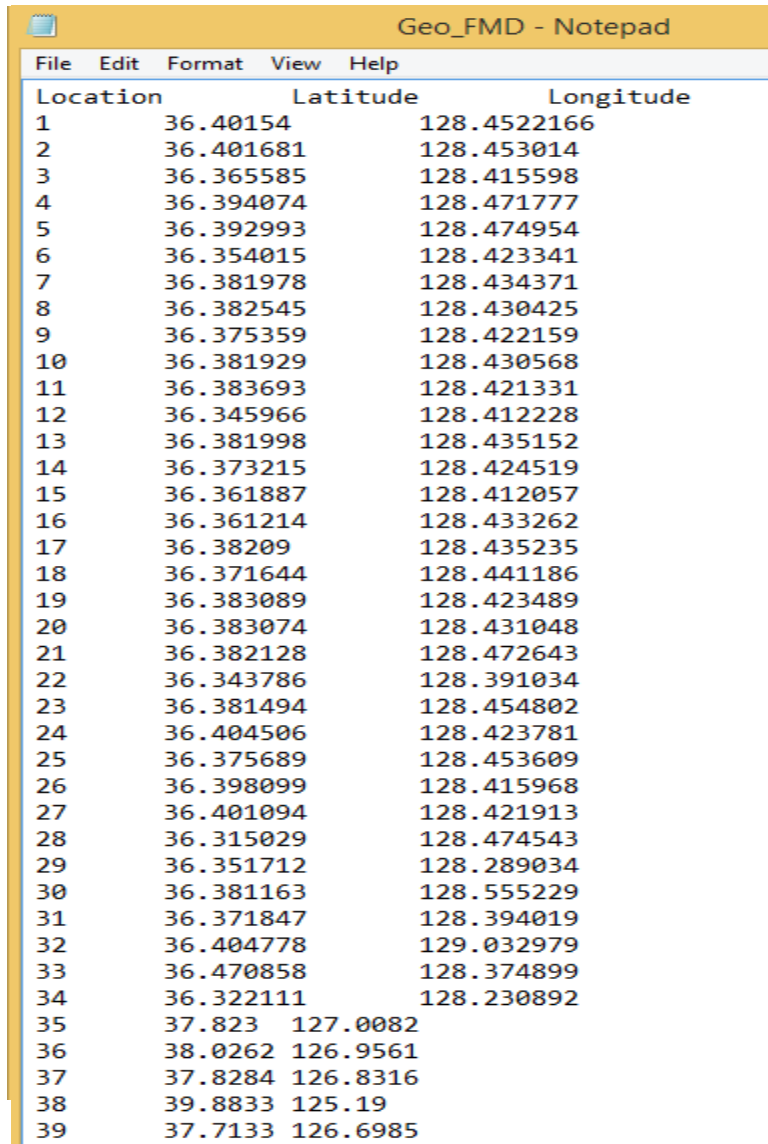
<http://www.kahis.go.kr/home/lkntscrinfo/selectLkntsOccrrncList.do?openFlag=Y>

- Data (2010-2015): North Korea

http://www.oie.int/wahis_2/public/wahid.php/Wahidhome/Home

- Number of infected/susceptible animal at farm levels (with geographical information)
- Outbreak date

Discrete Poisson Model



The image shows a Notepad window titled "Geo_FMD - Notepad" with a menu bar containing "File", "Edit", "Format", "View", and "Help". The main text area contains a table with three columns: "Location", "Latitude", and "Longitude". The table lists 39 numbered locations with their corresponding latitude and longitude values. Locations 1 through 34 have two columns for latitude and longitude, while locations 35 through 39 have three columns for latitude and longitude.

Location	Latitude	Longitude
1	36.40154	128.4522166
2	36.401681	128.453014
3	36.365585	128.415598
4	36.394074	128.471777
5	36.392993	128.474954
6	36.354015	128.423341
7	36.381978	128.434371
8	36.382545	128.430425
9	36.375359	128.422159
10	36.381929	128.430568
11	36.383693	128.421331
12	36.345966	128.412228
13	36.381998	128.435152
14	36.373215	128.424519
15	36.361887	128.412057
16	36.361214	128.433262
17	36.38209	128.435235
18	36.371644	128.441186
19	36.383089	128.423489
20	36.383074	128.431048
21	36.382128	128.472643
22	36.343786	128.391034
23	36.381494	128.454802
24	36.404506	128.423781
25	36.375689	128.453609
26	36.398099	128.415968
27	36.401094	128.421913
28	36.315029	128.474543
29	36.351712	128.289034
30	36.381163	128.555229
31	36.371847	128.394019
32	36.404778	129.032979
33	36.470858	128.374899
34	36.322111	128.230892
35	37.823	127.0082
36	38.0262	126.9561
37	37.8284	126.8316
38	39.8833	125.19
39	37.7133	126.6985

Results

1. Location IDs included: 174, 128, 103, 228, 85, 65, 218, 202, 372, 400, 405, 118, 463, 51
Space(50%)-time(50%) 379, 50, 395, 398, 157, 351, 357, 139, 318, 53, 98, 339, 366, 407
348, 353, 324, 383, 183, 375, 72, 54, 414, 62, 346, 454, 370, 341
367, 307, 340, 337, 347, 342, 352, 349, 363, 362, 328, 336, 369,
64, 338, 331, 368, 425, 359, 314, 179, 365, 106, 335, 312, 424, 3
423, 435, 57, 409, 329, 356, 309, 358, 125, 350, 301, 303, 305, 3
300, 302, 401, 317, 396, 164, 315, 299, 298, 313, 297, 61, 147, 4
135, 316, 310, 132, 144, 133, 321, 344, 345, 332, 408, 381, 142,
126, 404, 394, 420, 410, 419, 161, 155, 306, 467, 308, 311, 361,
219, 322, 163, 140, 433, 478, 304, 261, 173, 320, 63, 83, 33, 323
280, 374, 417, 360, 124, 472, 119, 327, 371, 42, 250, 481, 468, 2
27, 26, 439, 1, 2, 440, 247, 11, 19, 20, 8, 4, 10, 29, 31, 7, 17,
5, 245, 9, 23, 14, 21, 25, 3, 18, 15, 16, 445, 34, 6, 22, 47, 121
12, 30, 121, 91, 376, 102, 334, 447, 28, 441, 399, 46, 196, 453,
411, 172, 460, 249, 466, 479, 99, 465, 442, 35, 459, 448, 109, 28
127, 473, 474, 156, 100, 326, 248, 343, 476, 145, 43, 449, 457, 3
44, 469, 246, 462, 378, 416, 470, 354, 377, 182, 37, 477, 40, 430
59, 56, 274, 36, 120, 143, 39, 180, 277, 60, 52, 279, 97, 95, 422
104, 137, 96, 276, 458, 84, 283, 45, 265, 456, 138, 438, 446, 319
451, 480, 437, 443, 450, 426, 429, 235, 475, 452, 55, 403, 436, 4
427, 413, 389, 415, 406, 382, 418, 397, 434, 384, 391, 428, 393,
402, 388, 380, 385, 390, 392, 464, 387, 386, 373, 471, 444, 234,
110, 288, 92, 431, 287, 289, 82, 229, 208, 432

Space(10%)-time(10%) 379, 50, 395, 398, 157, 351, 357, 139, 318, 53, 98, 339, 366, 407,
348, 353, 324, 383, 183, 375, 72, 54, 414, 62, 346, 454, 370, 341,
367, 307, 340, 337, 347, 342, 352, 349, 363, 362, 328, 336, 369, 364,
64, 338, 331, 368, 425, 359, 314, 179, 365, 106, 335, 312, 424, 333,
423, 435, 57, 409, 329, 356, 309, 358, 125, 350, 301, 303, 305, 330,
300, 302, 401, 317, 396, 164, 315, 299, 298, 313, 297, 61, 147, 49,
135, 316, 310, 132, 144, 133, 321, 344, 345, 332, 408, 381, 142, 355,
126, 404, 394, 420, 410, 419, 161, 155, 306, 467, 308, 311, 361, 325,
219, 322, 163, 140, 433, 478, 304, 261, 173, 320, 63, 83, 33, 323,
280, 374, 417, 360, 124, 472, 119, 327, 371, 42, 250, 481, 468, 24,
27, 26, 439, 1, 2, 440, 247, 11, 19, 20, 8, 4, 10, 29, 31, 7, 17, 13,
5, 245, 9, 23, 14, 21, 25, 3, 18, 15, 16, 445, 34, 6, 22, 47, 421,
12, 30, 121, 91, 376, 102, 334, 447, 28, 441, 399, 46, 196, 453, 412,
411, 172, 460, 249, 466, 479, 99, 465, 442, 35, 459, 448, 109, 281,
127, 473, 474, 156, 100, 326, 248, 343, 476, 145, 43, 449, 457, 32,
44, 469, 246, 462, 378, 416, 470, 354, 377, 182, 37, 477, 40, 430,
59, 56, 274, 36, 120, 143, 39, 180, 277, 60, 52, 279, 97, 95, 422,
104, 137, 96, 276, 458, 84, 283, 45, 265, 456, 138, 438, 446, 319,
451, 480, 437, 443, 450, 426, 429, 235, 475, 452, 55, 403, 436, 455,
427, 413, 389, 415, 406, 382, 418, 397, 434, 384, 391, 428, 393, 461,
402, 388, 380, 385, 390, 392, 464, 387, 386, 373, 471, 444, 234, 101,
110, 288, 92, 431, 287, 289, 82, 229, 208, 432

Coordinates / radius.: (37.224964 N, 128.081455 E) / 167.39 km
Time frame.....: 2014/12/1 to 2015/4/30
Population.....: 2172
Number of cases.....: 155767
Expected cases.....: 6523.33
Annual cases / 100000.: 8923072.9
Observed / expected...: 23.88
Relative risk.....: 254.51
Log likelihood ratio..: 457679.909303
P-value.....: < 0.0000000000000001

Coordinates / radius.: (37.224964 N, 128.081455 E) / 167.39 km
Time frame.....: 2014/12/1 to 2015/4/30
Population.....: 2172
Number of cases.....: 155767
Expected cases.....: 6523.33
Annual cases / 100000.: 8923072.9
Observed / expected...: 23.88
Relative risk.....: 254.51
Log likelihood ratio..: 457679.909303
P-value.....: < 0.0000000000000001

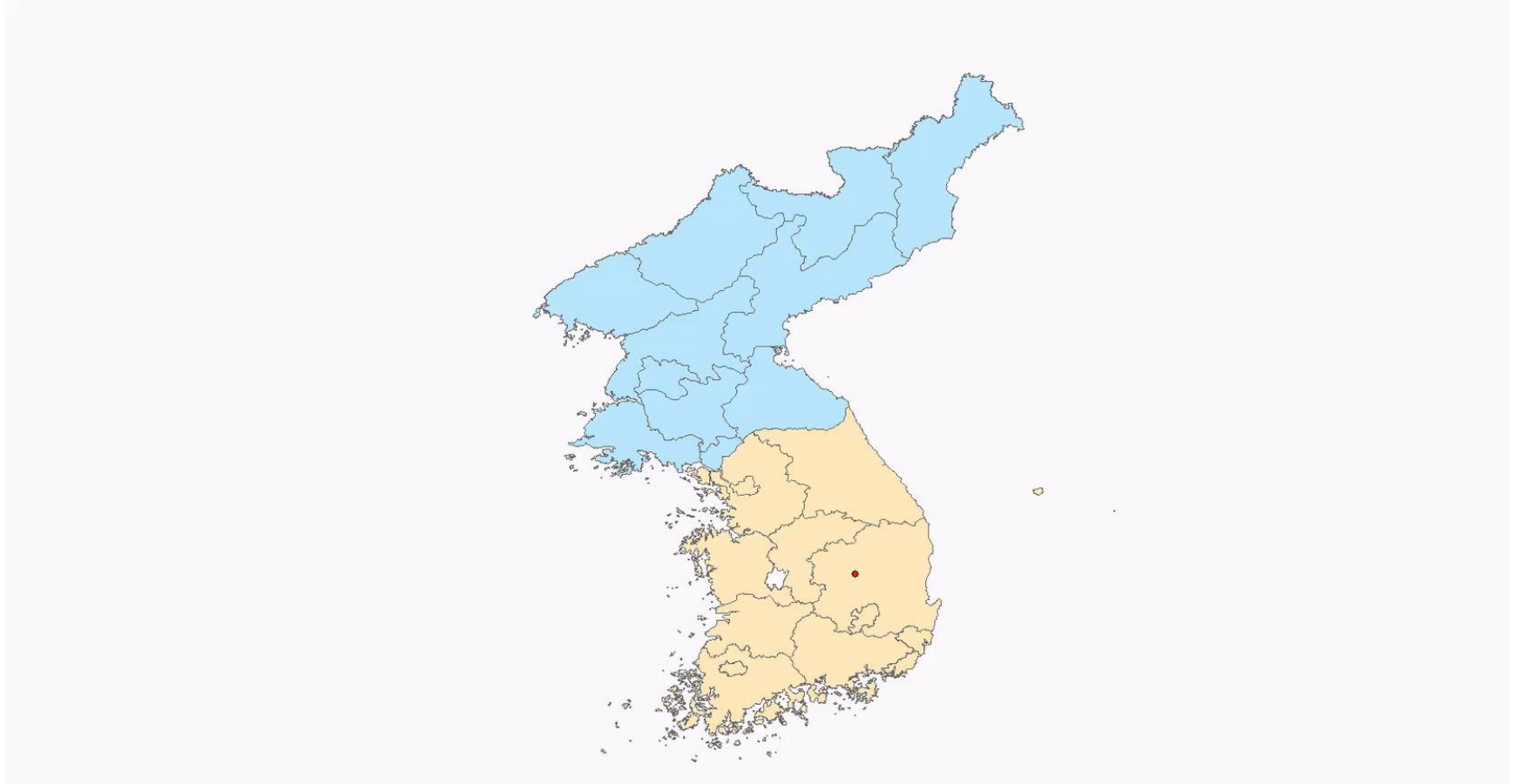
2. Location IDs included.: 165, 129, 225, 105, 223, 232, 166, 206, 222, 184, 226, 254, 195,
273, 257, 130, 216, 220, 210, 136, 186, 215, 189, 188, 268, 153,
150, 152, 259, 160, 159, 117, 148, 77, 256, 149, 213, 192, 123, 1
294, 193, 66, 94, 212, 86

2. Location IDs included.: 165, 129, 225, 105, 223, 232, 166, 206, 222, 184, 226, 254, 195, 122,
273, 257, 130, 216, 220, 210, 136, 186, 215, 189, 188, 268, 153, 253,
150, 152, 259, 160, 159, 117, 148, 77, 256, 149, 213, 192, 123, 197,
294, 193, 66, 94, 212, 86

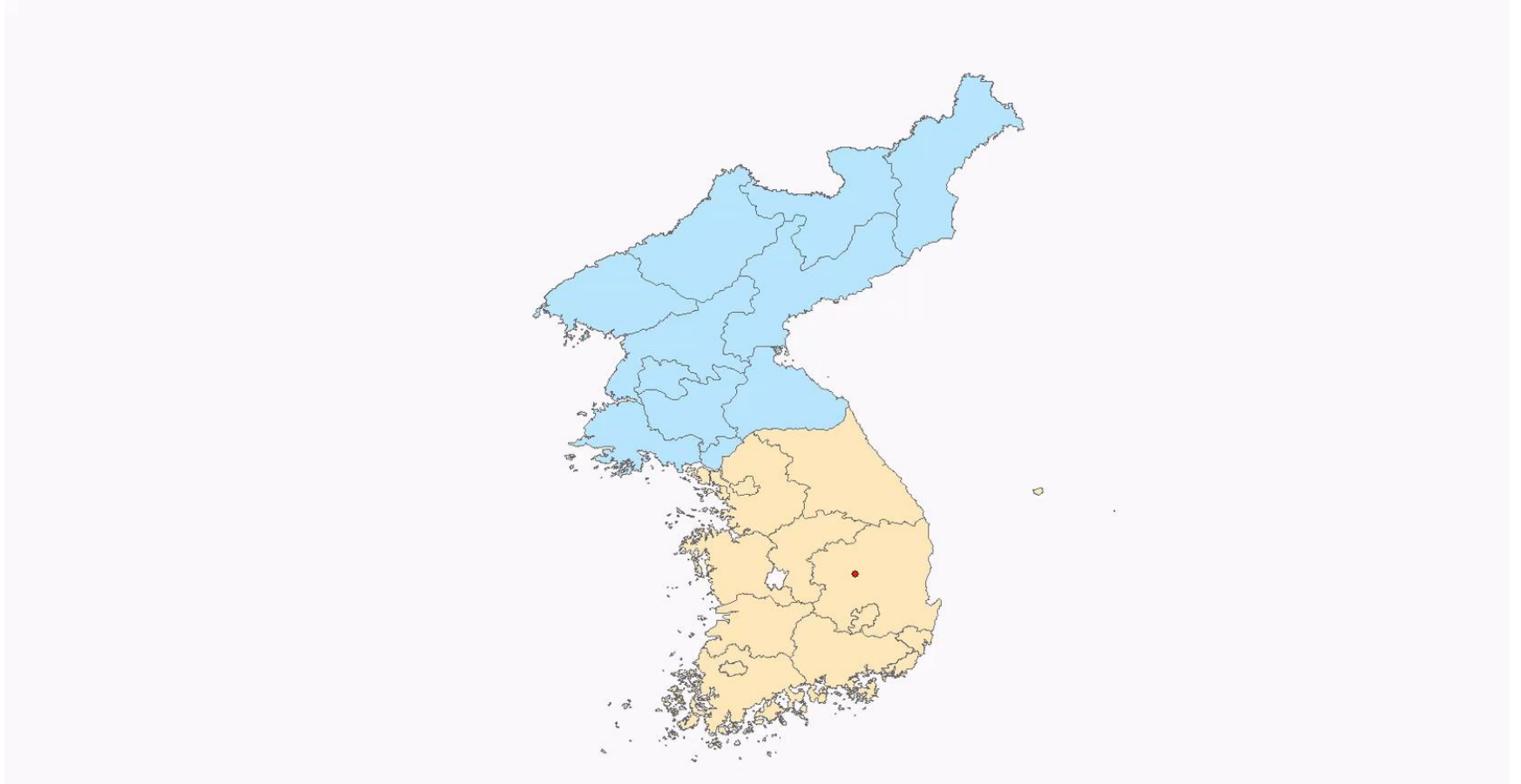
Coordinates / radius.: (38.755300 N, 126.120000 E) / 34.16 km
Time frame.....: 2010/12/1 to 2011/1/31
Population.....: 20
Number of cases.....: 6172
Expected cases.....: 73.77
Annual cases / 100000.: 31265973.8
Observed / expected...: 83.67
Relative risk.....: 86.76
Log likelihood ratio..: 21334.342555
P-value.....: < 0.0000000000000001

Coordinates / radius.: (38.755300 N, 126.120000 E) / 34.16 km
Time frame.....: 2010/12/1 to 2011/1/31
Population.....: 20
Number of cases.....: 6172
Expected cases.....: 73.77
Annual cases / 100000.: 31265973.8
Observed / expected...: 83.67
Relative risk.....: 86.76
Log likelihood ratio..: 21334.342555
P-value.....: < 0.0000000000000001

ArcGIS – import (outbreak locations)



Outbreak locations (2010 Nov-2015 April)



ArcGIS – buffer (50%-50%)

File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help

Scale: 1:5,000,000

Table of Contents

- Layers
 - C:\Users\dell\Desktop\Cl
 - XYSecond_cluster_50
 - XYPrim_cluster_50_50
 - prim_50_50
 - C:\Users\dell\Desktop\Cl
 - Prim_cluster_50_50.bt
 - Outbreak_location.bt
 - Outbreak_location.bt
 - Prim_cluster_50_50.bt
 - C:\Users\dell\Desktop\M
 - Korea
 - polbnda2013

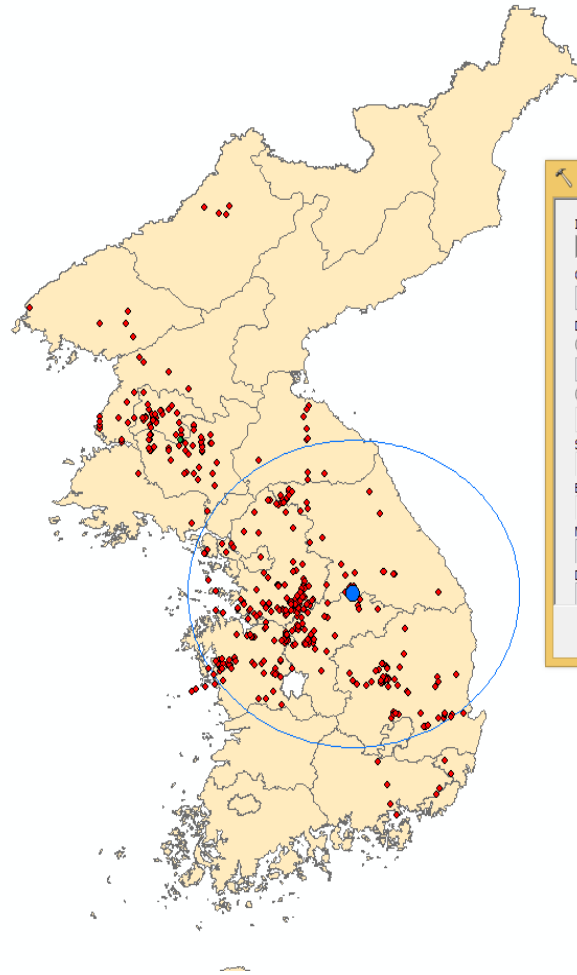
1. Location IDs included: 174, 128, 103, 228, 85, 65, 218, 202, 37, 379, 50, 395, 398, 157, 351, 357, 139, 348, 353, 324, 383, 183, 375, 72, 54, 41, 367, 307, 340, 337, 347, 342, 352, 349, 64, 338, 331, 368, 425, 359, 314, 179, 423, 435, 57, 409, 329, 356, 309, 358, 300, 302, 401, 317, 396, 164, 315, 299, 135, 316, 310, 132, 144, 133, 321, 344, 126, 404, 394, 420, 410, 419, 161, 155, 219, 322, 163, 140, 433, 478, 304, 261, 280, 374, 417, 360, 124, 472, 119, 327, 27, 26, 439, 1, 2, 440, 247, 11, 19, 20, 5, 245, 9, 23, 14, 21, 25, 3, 18, 15, 16, 12, 30, 121, 91, 376, 102, 334, 447, 28, 411, 172, 460, 249, 466, 479, 99, 465, 127, 473, 474, 156, 100, 326, 248, 343, 44, 469, 246, 462, 378, 416, 470, 354, 59, 56, 274, 36, 120, 143, 39, 180, 277, 104, 137, 96, 276, 458, 84, 283, 45, 265, 451, 480, 437, 443, 450, 426, 429, 235, 427, 413, 389, 415, 406, 382, 418, 397, 402, 388, 380, 385, 390, 392, 464, 387, 110, 288, 92, 431, 287, 209, 82, 229, 2

Coordinates / radius... (37.224964 N, 128.081455 E) / 167.39 km
Time frame... 2014/12/1 to 2015/4/30
Population... 2172

Number of cases... 155707
Expected cases... 6523.33
Annual cases / 100000... 8923072.9
Observed / expected... 23.86
Relative risk... 254.51
Log likelihood ratio... 457679.909303
P-value... < 0.000000000000000001

2. Location IDs included: 165, 129, 225, 105, 223, 232, 166, 206, 273, 257, 130, 216, 220, 210, 136, 186, 150, 152, 259, 160, 159, 117, 148, 77, 294, 193, 66, 94, 212, 86

Coordinates / radius... (38.733300 N, 126.120000 E) / 34.16 km
Time frame... 2010/12/1 to 2011/1/31
Population... 20
Number of cases... 6172
Expected cases... 73.77
Annual cases / 100000... 31265973.8
Observed / expected... 83.67



Buffer

Input Features
XYSecond_cluster_50_50

Output Feature Class
C:\Users\dell\Desktop\Cluster\Korea_map\seond_50_50.shp

Distance [value or field]
 Linear unit
34.16 Kilometers

Field

Side Type (optional)
FULL

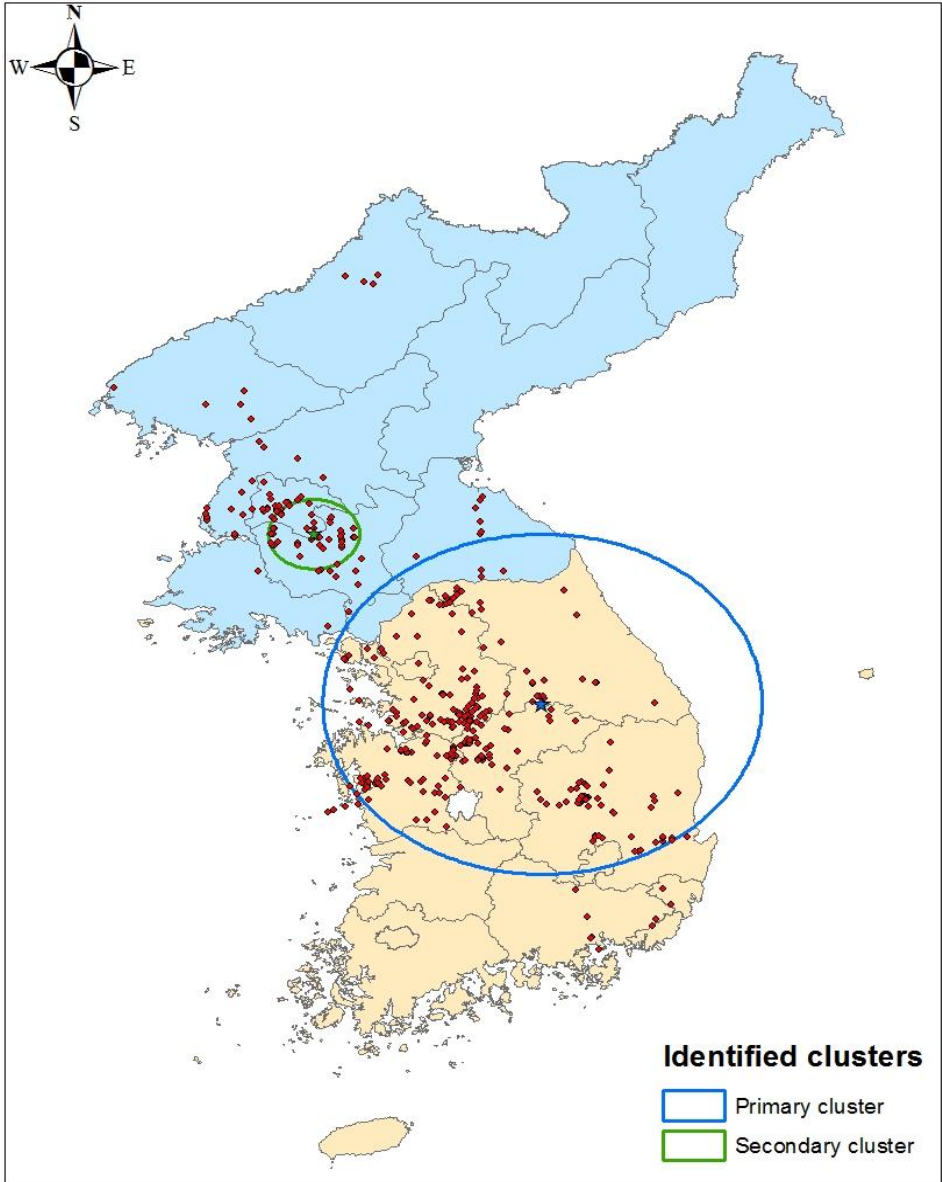
End Type (optional)
ROUND

Method (optional)
PLANAR

Dissolve Type (optional)
NONE

OK Cancel Environments... Show Help >>

Map with identified clusters

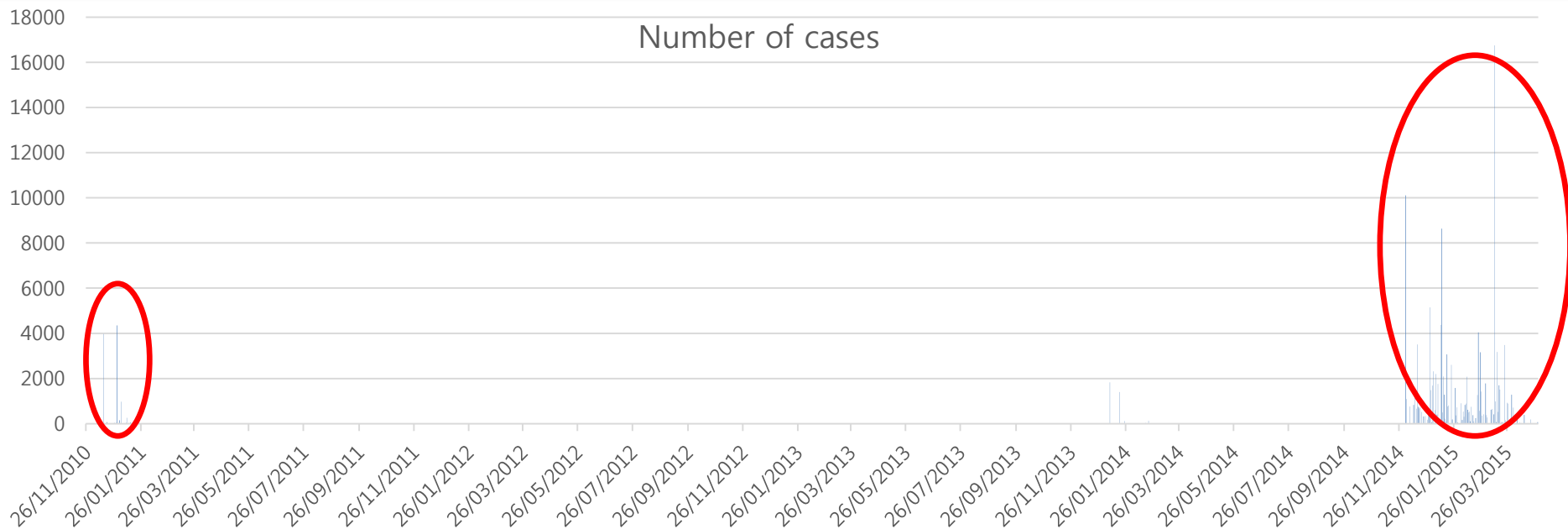


```

1.Location IDs included.: 174, 128, 103, 228, 85, 65, 218, 202, 379, 50, 395, 398, 157, 351, 357, 139, 348, 353, 324, 383, 183, 375, 72, 54, 367, 307, 340, 337, 347, 342, 352, 349, 64, 338, 331, 368, 425, 359, 314, 179, 423, 435, 57, 409, 329, 356, 309, 358, 300, 302, 401, 317, 396, 164, 315, 299, 135, 316, 310, 132, 144, 133, 321, 344, 126, 404, 394, 420, 410, 419, 161, 155, 219, 322, 163, 140, 433, 478, 304, 261, 280, 374, 417, 360, 124, 472, 119, 327, 27, 26, 439, 1, 2, 440, 247, 11, 19, 20, 5, 245, 9, 23, 14, 21, 25, 3, 18, 15, 16, 12, 30, 121, 91, 376, 102, 334, 447, 28, 411, 172, 460, 249, 466, 479, 99, 465, 127, 473, 474, 156, 100, 326, 248, 343, 44, 469, 246, 462, 378, 416, 470, 354, 59, 56, 274, 36, 120, 143, 39, 180, 277, 104, 137, 96, 276, 458, 84, 283, 45, 261, 451, 480, 437, 443, 450, 426, 429, 235, 427, 413, 389, 415, 406, 382, 418, 397, 402, 388, 380, 385, 390, 392, 464, 387, 110, 200, 92, 431, 207, 209, 02, 229, 2
Coordinates / radius...: (37.224964 N, 128.081455 E) / 167.39 km
Time frame.....: 2014/12/1 to 2015/4/30
Population.....: 2172
Number of cases.....: 155767
Expected cases.....: 6523.33
Annual cases / 100000.: 8923072.9
Observed / expected...: 23.88
Relative risk.....: 254.51
Log likelihood ratio...: 457679.909303
P-value.....: < 0.000000000000000001

2.Location IDs included.: 165, 129, 225, 105, 223, 232, 166, 206, 273, 257, 130, 216, 220, 210, 136, 186, 294, 193, 66, 94, 212, 86
Coordinates / radius...: (38.733300 N, 126.120000 E) / 34.16 km
Time frame.....: 2010/12/1 to 2011/1/31
Population.....: 200
Number of cases.....: 6172
Expected cases.....: 73.77
Annual cases / 100000.: 31265973.8
Observed / expected...: 83.67
    
```

Time clusters (2010-2015)



```
Coordinates / radius...: (37.224364 N, 120.001455 E) / 167.39 km  
Time frame.....: 2010/12/1 to 2015/4/30  
Population.....: 2172  
Number of cases.....: 155767  
Expected cases.....: 6523.33  
Annual cases / 100000.: 8923072.9  
Observed / expected...: 23.88  
Relative risk.....: 254.51  
Log likelihood ratio...: 457679.909303  
P-value.....: < 0.000000000000000001
```

```
2. Location IDs included.: 165, 129, 225, 105, 223, 232, 166, 206, 222, 184, 226, 254, 195, 122,  
273, 257, 130, 216, 220, 210, 136, 186, 215, 189, 188, 268, 153, 253,  
150, 152, 259, 160, 159, 117, 148, 77, 256, 149, 213, 192, 123, 197,  
294, 193, 66, 94, 212, 86
```

```
Coordinates / radius...: (58.755500 N, 120.120000 E) / 34.16 km  
Time frame.....: 2010/12/1 to 2011/1/31  
Population.....: 20  
Number of cases.....: 6172  
Expected cases.....: 73.77  
Annual cases / 100000.: 31265973.8  
Observed / expected...: 83.67  
Relative risk.....: 86.76  
Log likelihood ratio...: 21334.342555  
P-value.....: < 0.000000000000000001
```

Exercises (50 mins)

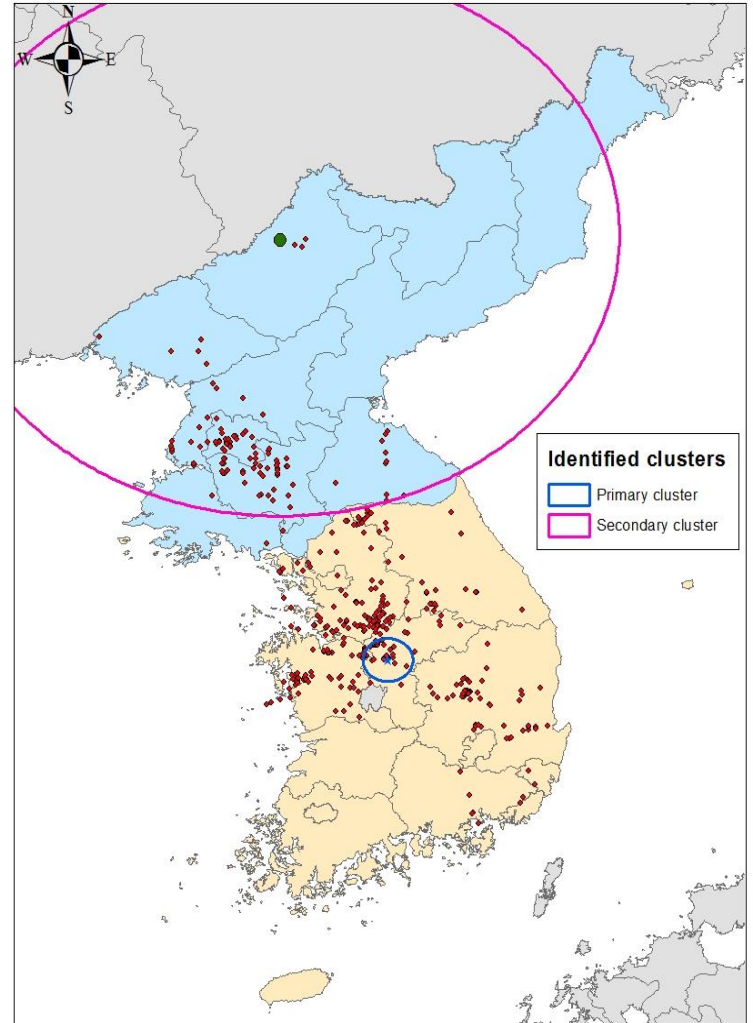
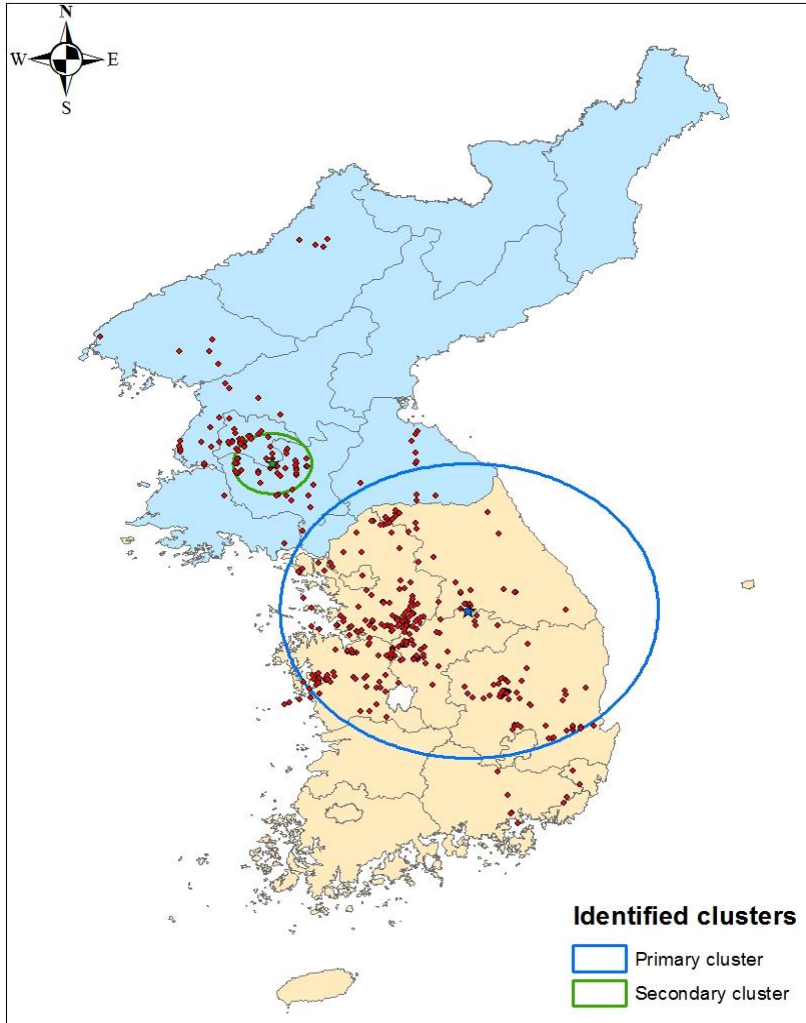
- **Leptospirosis data**

- Which model? (Bernoulli vs Poisson)
- Create files (I already created for you)
- Import data and Output
- Determine: cluster size (from zero to 50%)
- Comparison among different size (10%, 30% and 50%)

- **FMD data**

- Which model? Space or time ?
(Poisson vs Space-time permutation)
- Create files (You need to create files your own) - Notepad
- Import data and Output
- Determine: cluster size (from zero to 50%)
- Comparison among different size (50%-50%, 50%-10%, 10%-50% and 10%-10%)

Poisson vs Permutation models (50%-50%)



Seasonality analysis

Introduction

- Procedures (seasonality using R)

J Vet Intern Med 2014

Regional and Temporal Variations of *Leptospira* Seropositivity in Dogs in the United States, 2000–2010

H.S. Lee, M. Levine, C. Guptill-Yoran, A.J. Johnson, P. von Kamecke, and G.E. Moore

Background: Previous studies have reported a seasonal increased risk for leptospirosis, but there is no consistent seasonality reported across regions in the United States.

Objectives: To evaluate and compare seasonal patterns in seropositivity for leptospirosis in dogs for 4 US regions (northeast [NE], midwest [MW], south-central [SC], and California-southern west coast [CS]).

Animals: Forty four thousand nine hundred and sixteen canine serum samples submitted to a commercial laboratory for microscopic agglutination tests (MAT) from 2000 through 2010.

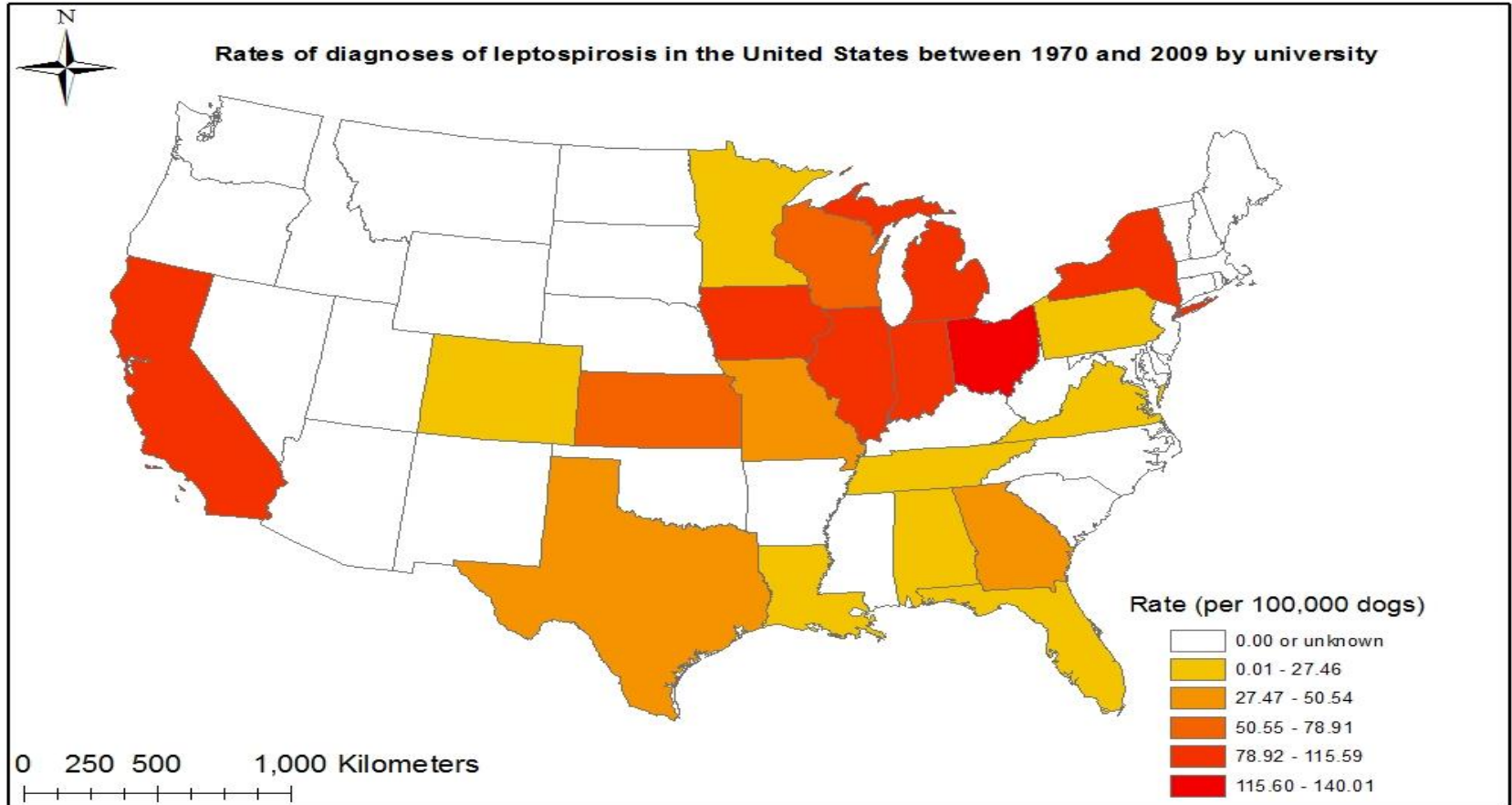
Methods: In this retrospective study, positive cases were defined as MAT titers $\geq 1 : 3,200$ for at least one of 7 tested serovars. Four geographic regions were defined, and MAT results were included in regional analyses based on hospital zip-code. A seasonal-trend decomposition method for times series was utilized for the analysis. Monthly variation in the seropositive rate was evaluated using a seasonal cycle subseries plot and logistic regression.

Results: Two thousand and twelve of 44,916 (4.48%) samples were seropositive. Compared to seropositive rates for February, significantly higher monthly rates occurred during the 2nd half of the year in the MW (OR 3.92–6.35) and NE (OR 2.03–4.80) regions, and only in January (OR 2.34) and December (OR 1.74) in the SC region. Monthly seropositive rates indicative of seasonality were observed earlier in the calendar year for both CS and SC regions.

Conclusions and Clinical Importance: Seasonal patterns for seropositivity to leptospires differed by geographic region. Although risk of infection in dogs can occur year round, knowledge of seasonal trends can assist veterinarians in formulating differential diagnoses and evaluation of exposure risk.

Key words: Dogs; *Leptospira*; Microscopic agglutination tests; Seasonal cycle subseries plot; Seasonal-trend decomposition procedure based on loess; Seropositive.

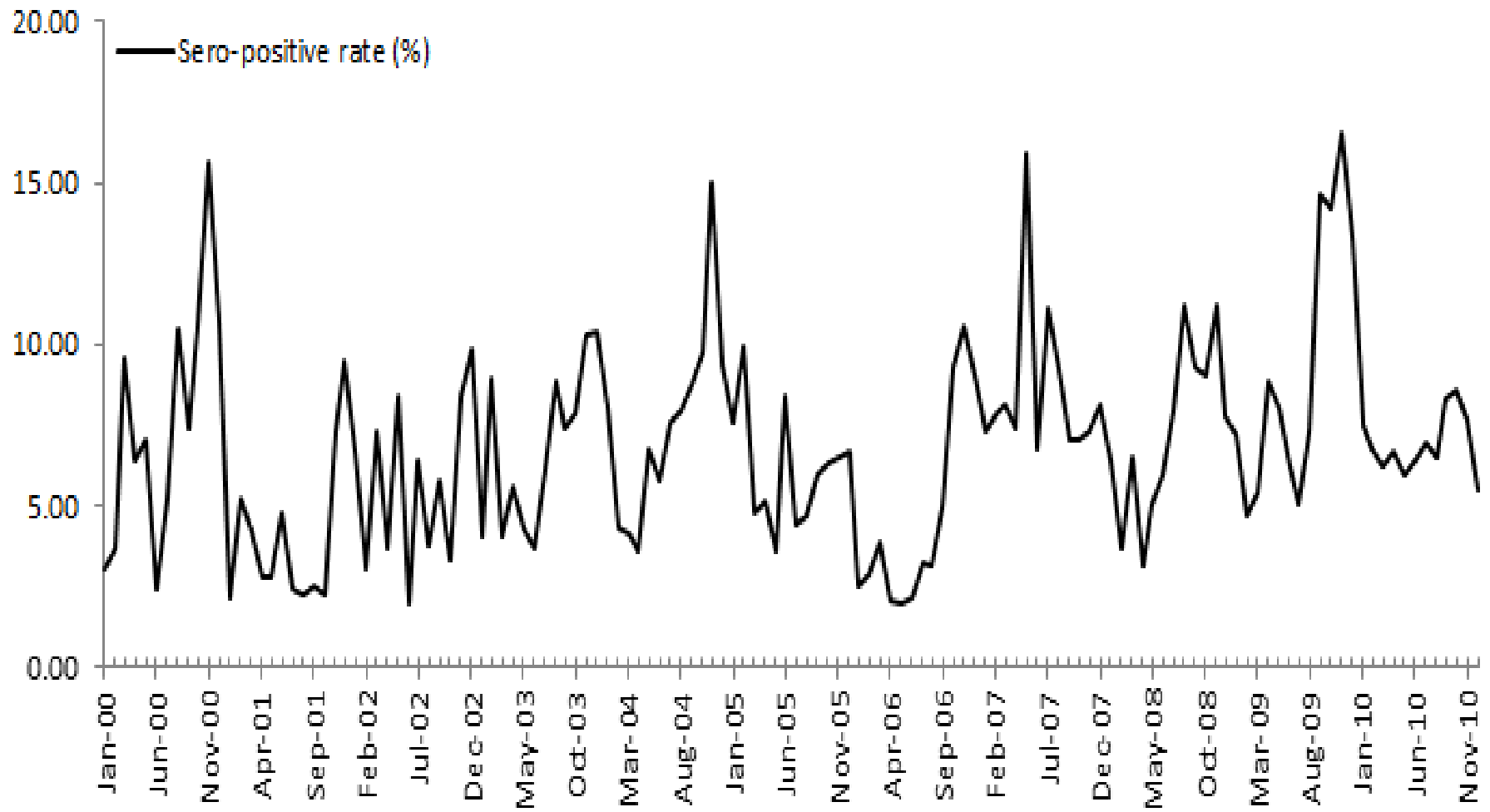
Risk map



What is STL and SCS ?

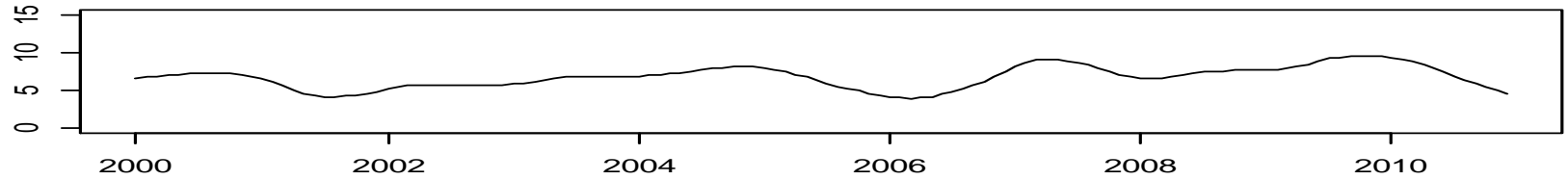
- STL is a useful tool to visualize time series datasets that consists of
 - Trend, seasonal and remainder components
- SCS helps to visualize patterns both between and with groups that consists of
 - Horizontal lines: average for each month
 - Vertical lines: individual pattern for the same month in each year

Time series plot



STL plot

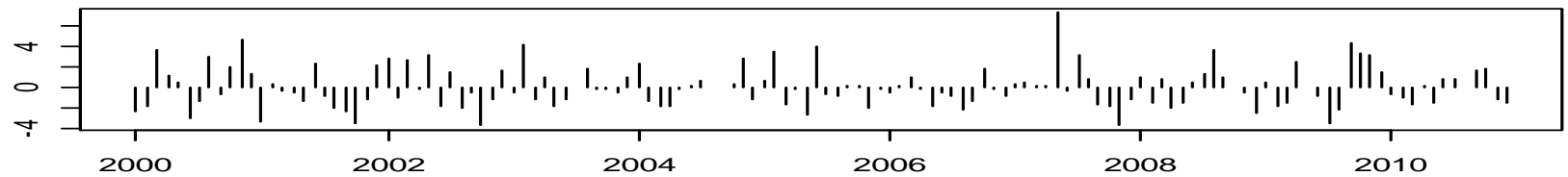
Trend



Seasonal

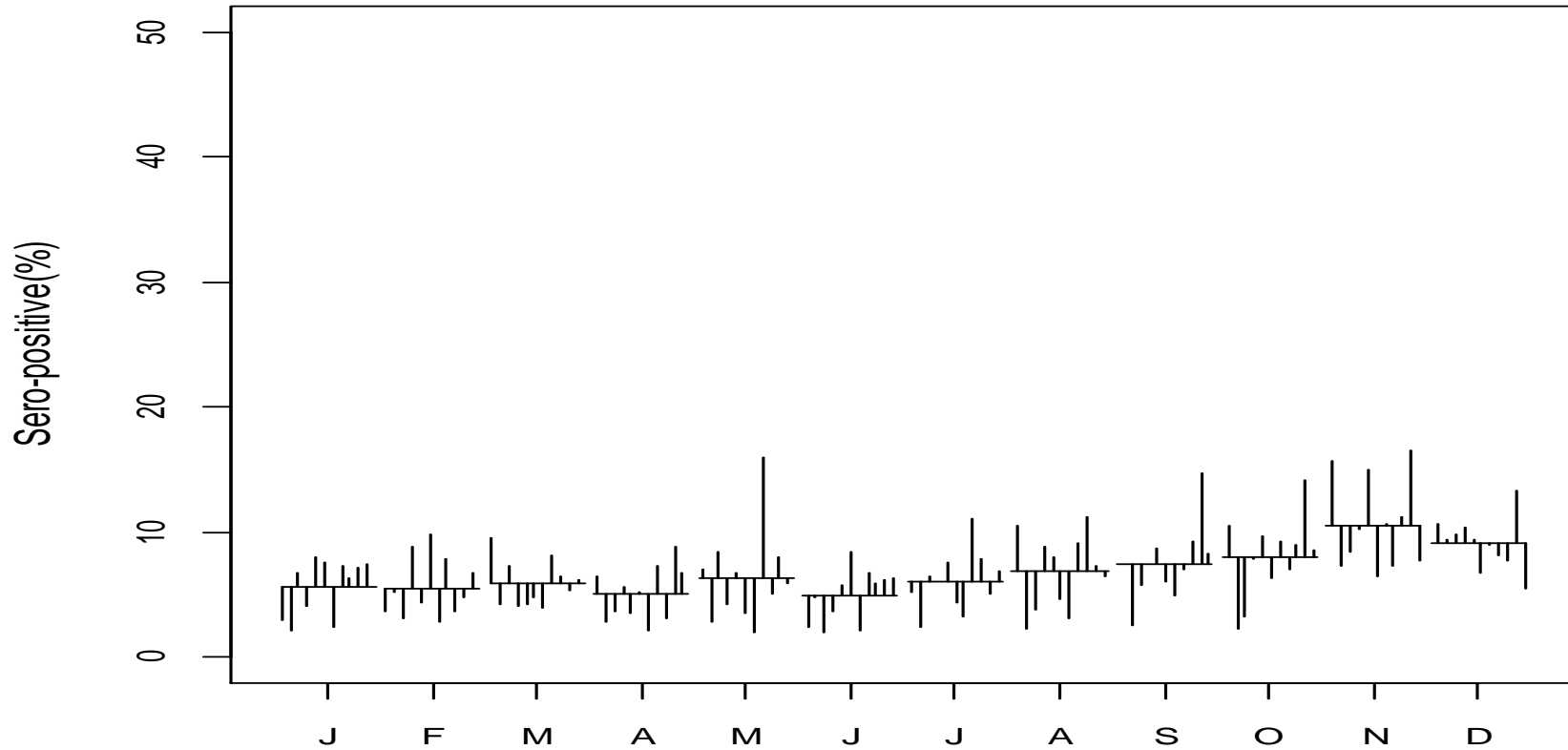


Remainder



Time

Seasonal cycle subseries plot

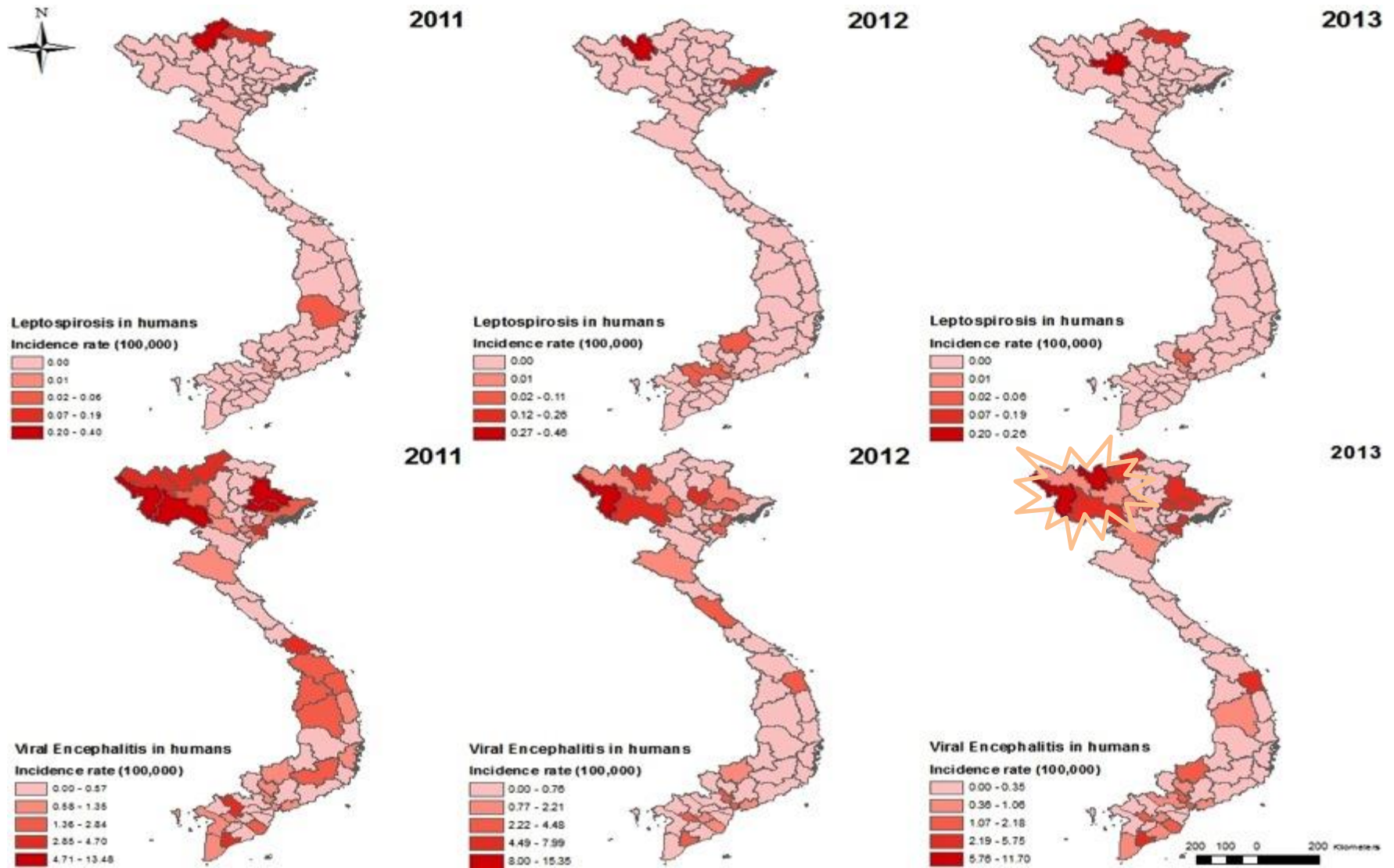


Regression model

Odds ratios with 95% Confidence interval

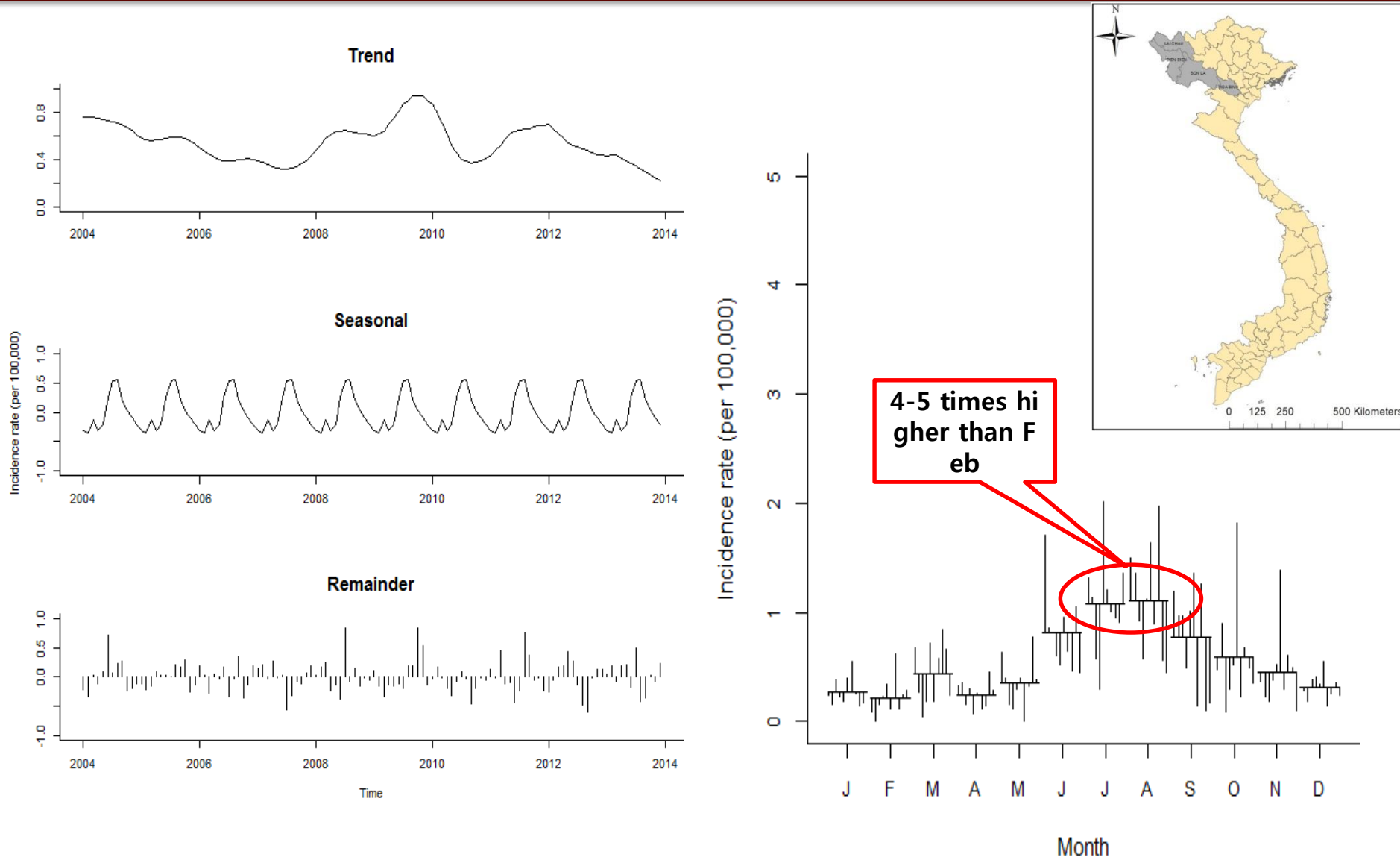
Month	Northeast	Mid-west	South central	California
Jan	1.30 (0.86-1.98)	1.18 (0.60-2.33)	1.32 (0.76-2.27)	Reference: 1
Feb	Reference: 1	Reference: 1	Reference: 1	1.93 (1.17-3.18)
Mar	0.71 (0.44-1.15)	1.26 (0.65-2.46)	1.08 (0.62-1.88)	1.98 (1.22-3.23)
Apr	0.91 (0.59-1.41)	0.87 (0.42-1.79)	1.28 (0.73-2.23)	1.60 (0.96-2.66)
May	1.06 (0.69-1.62)	1.70 (0.90-3.20)	1.45 (0.85-2.45)	1.49 (0.87-2.53)
Jun	0.90 (0.58-1.39)	1.16 (0.60-2.24)	1.08 (0.62-1.88)	1.22 (0.71-2.11)
Jul	1.23 (0.82-1.86)	1.42 (0.74-2.70)	1.35 (0.79-2.30)	0.95 (0.53-1.70)
Aug	1.61 (1.09-2.38)	1.59 (0.84-3.01)	1.07 (0.61-1.85)	1.02 (0.58-1.80)
Sep	2.04 (1.40-2.98)	1.59 (0.85-2.96)	1.30 (0.77-2.20)	0.76 (0.41-1.42)
Oct	2.29 (1.59-3.29)	1.72 (0.94-3.17)	1.30 (0.77-2.19)	0.62 (0.33-1.18)
Nov	3.40 (2.38-4.88)	2.31 (1.27-4.19)	1.47 (0.88-2.46)	1.60 (0.96-2.68)
Dec	2.39 (1.65-3.46)	2.00 (1.09-3.69)	1.74 (1.04-2.90)	1.31 (0.77-2.24)

Annual incidence rates for leptos and VE in humans



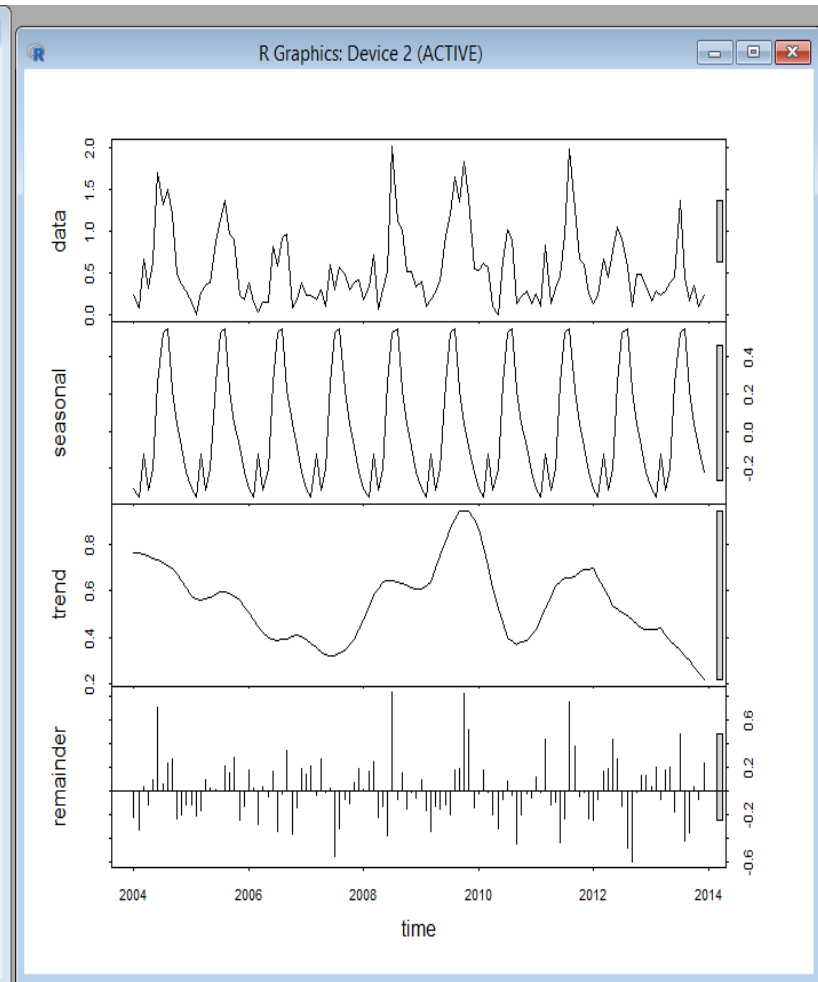
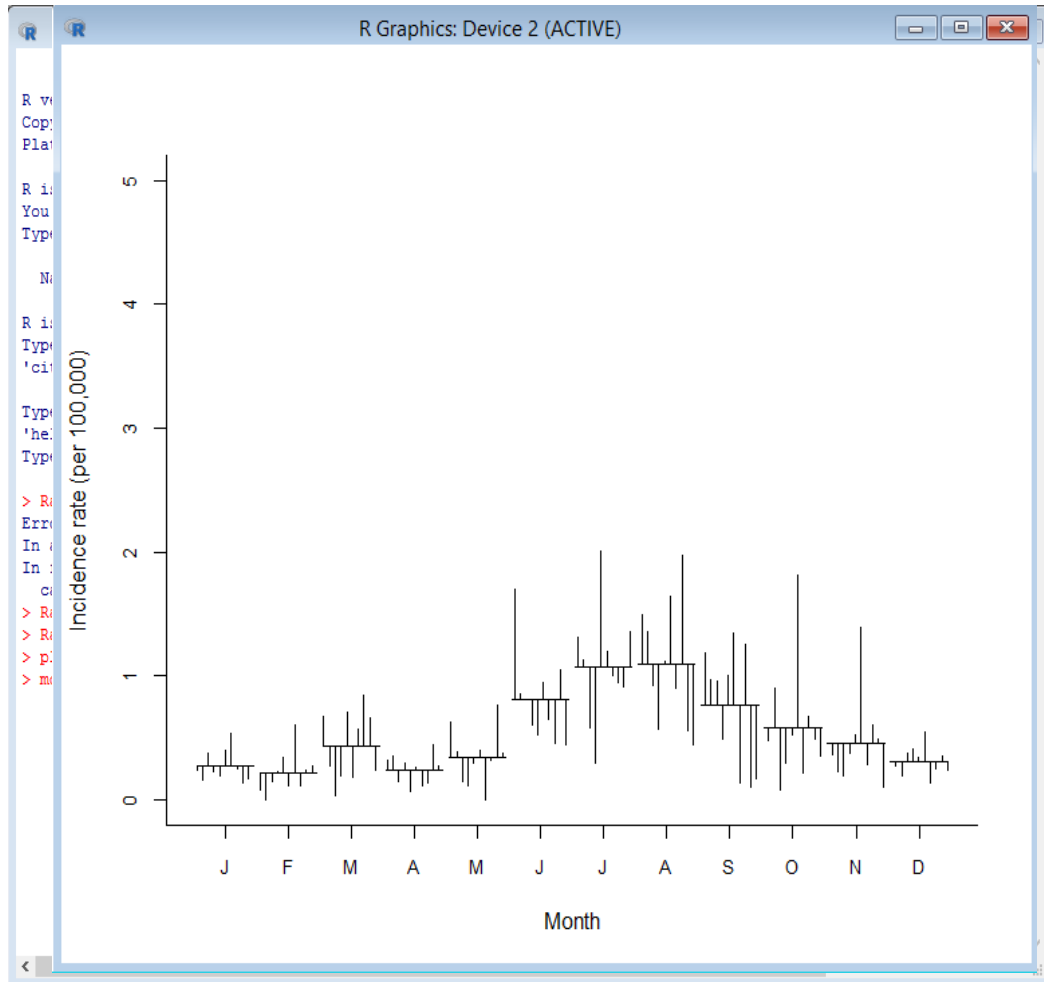
*Previous study showed that 17~71% of VE were caused by JE in Vietnam

Seasonality of VE in humans between 2004 and 2013 (Dien Bien, Hoa Binh, Lai Chau and Son La)



*Previous study showed that 17~71%

Screenshot (R package)



Exercises (15 mins)

- VE monthly incidence rate from 2004 to 2013
 - Open file: seasonality_VE and R_code_VE
 - Import into R:
 - Run the model
 - Interpretation

Time series analysis



Contents lists available at SciVerse ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Time series analysis of human and bovine brucellosis in South Korea from 2005 to 2010

Hu Suk Lee^a, Moon Her^b, Michael Levine^c, George E. Moore^{a,*}

^a Department of Comparative Pathobiology, College of Veterinary Medicine, Purdue University, 725 Harrison Street, West Lafayette, IN 47907, United States

^b OIE Reference Laboratory for Brucellosis, Bacterial Disease Division, Department of Animal and Plant Health Research, Animal, Plant and Fisheries Quarantine and Inspection Agency (QIA), 175 Anyang-ro, Manan-gu, Anyang-si, Gyeonggi 430-757, South Korea

^c Department of Statistics, Purdue University, 250 N. University Street, West Lafayette, IN 47907, United States

ARTICLE INFO

Article history:

Received 18 July 2012

Received in revised form

12 November 2012

Accepted 1 December 2012

Keywords:

Brucellosis

Time series

Lag

ARIMAX model

NBR model

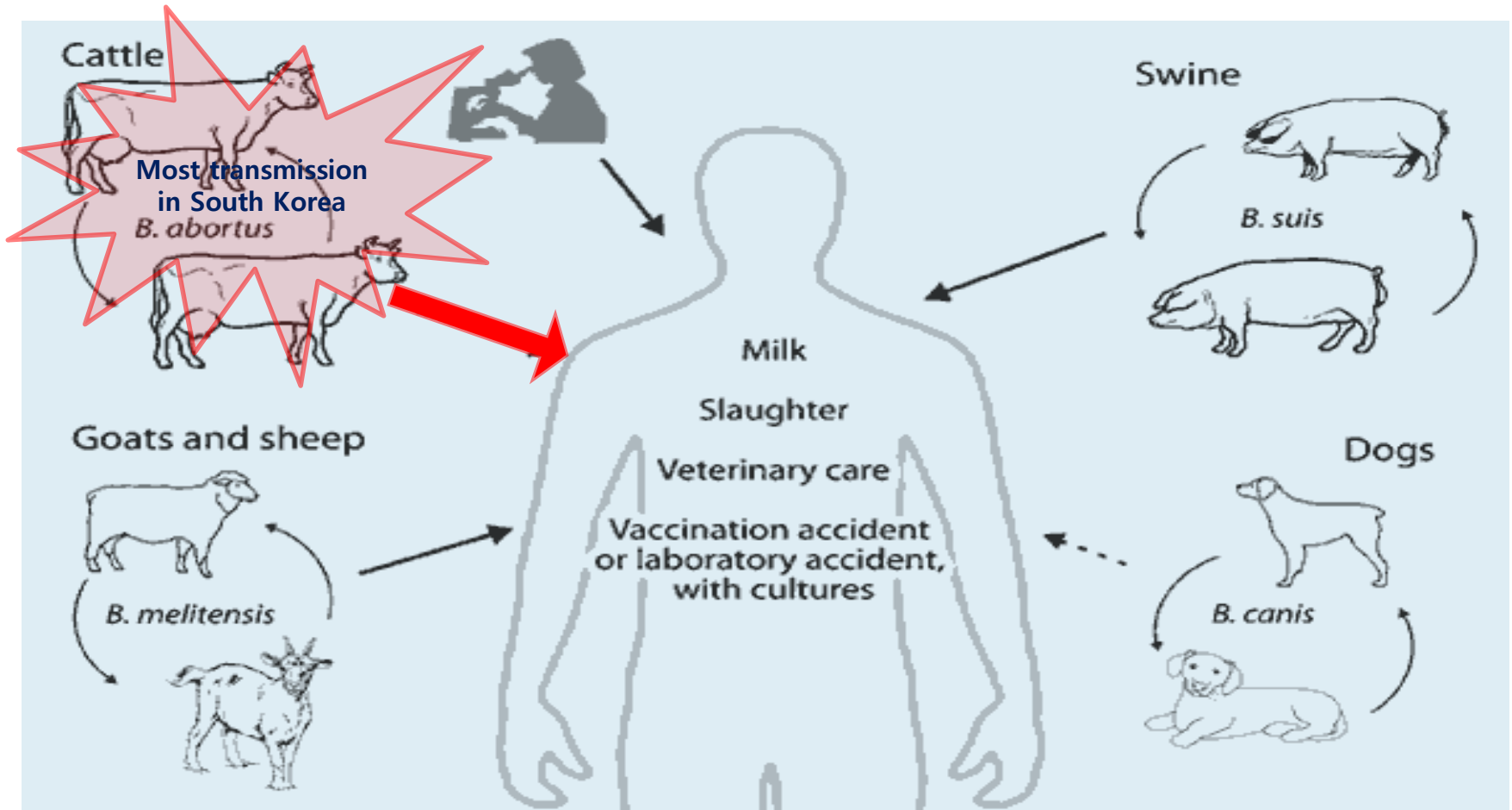
ABSTRACT

Brucellosis is considered to be one of the most important zoonotic diseases in the world, affecting underdeveloped and developing countries. The primary purpose of brucellosis control is to prevent the spread of disease from animals (typically ruminants) to humans. The main objective of this study was to retrospectively develop an appropriate time series model for cattle-to-human transmission in South Korea using data from independent national surveillance systems. Monthly case counts for cattle and people as well as national population data were available for 2005–2010. The temporal relationship was evaluated using an autoregressive integrated moving average with exogenous input (ARIMAX) model [notated as $ARIMA(p, d, q) - AR(p)$] and a negative binomial regression (NBR) model.

Human incidence rate was highly correlated to cattle incidence rate in the same month and the previous month (both $r = 0.82$). In the final models, $ARIMA(0, 1, 1) - AR(0, 1)$ was determined as the best fit with 191.5% error in the validation phase, whereas the best NBR model including lags (0, 1 months) for the cattle incidence rate yielded a 131.9% error in the

Brucellosis

Transmission to Humans



Introduction of project

- Rationale

- Most human cases are related to not wearing protection, such as gloves and protective clothing when in contact with suspected cattle or materials

- Objective

- To develop an appropriate time series model for cattle and humans in South Korea using data from independent national surveillance systems

Materials

- Study period: Jan 1, 2005 ~ Dec 31, 2010
- Data sources
 - Data collected on a yearly & quarterly basis:
 - **KOSIS** : national total population for human & cattle
(Korean Statistical Information Service)
 - Data collected on a monthly basis:
 - **KCDC** : number of human cases
(Korea Centers for Disease Control and Prevention)
 - **AIMS** : number of cattle cases
(Animal Infectious Disease Data Management System)

Methods

- Human & cattle incidence rates were calculated on a monthly basis

- Cases / national total population

The ARIMAX model is an extension of ARIMA (autoregressive integrated moving average) model

$$\hat{Y}_t = \text{ARIMA} [\text{Constant} + Y_{t-1} + \phi((Y_{t-1} - Y_{t-2})) - \theta e_{t-1}] + \text{AR} [\beta_{t-1} X_{t-1} + \beta_t X_{t-1}]$$

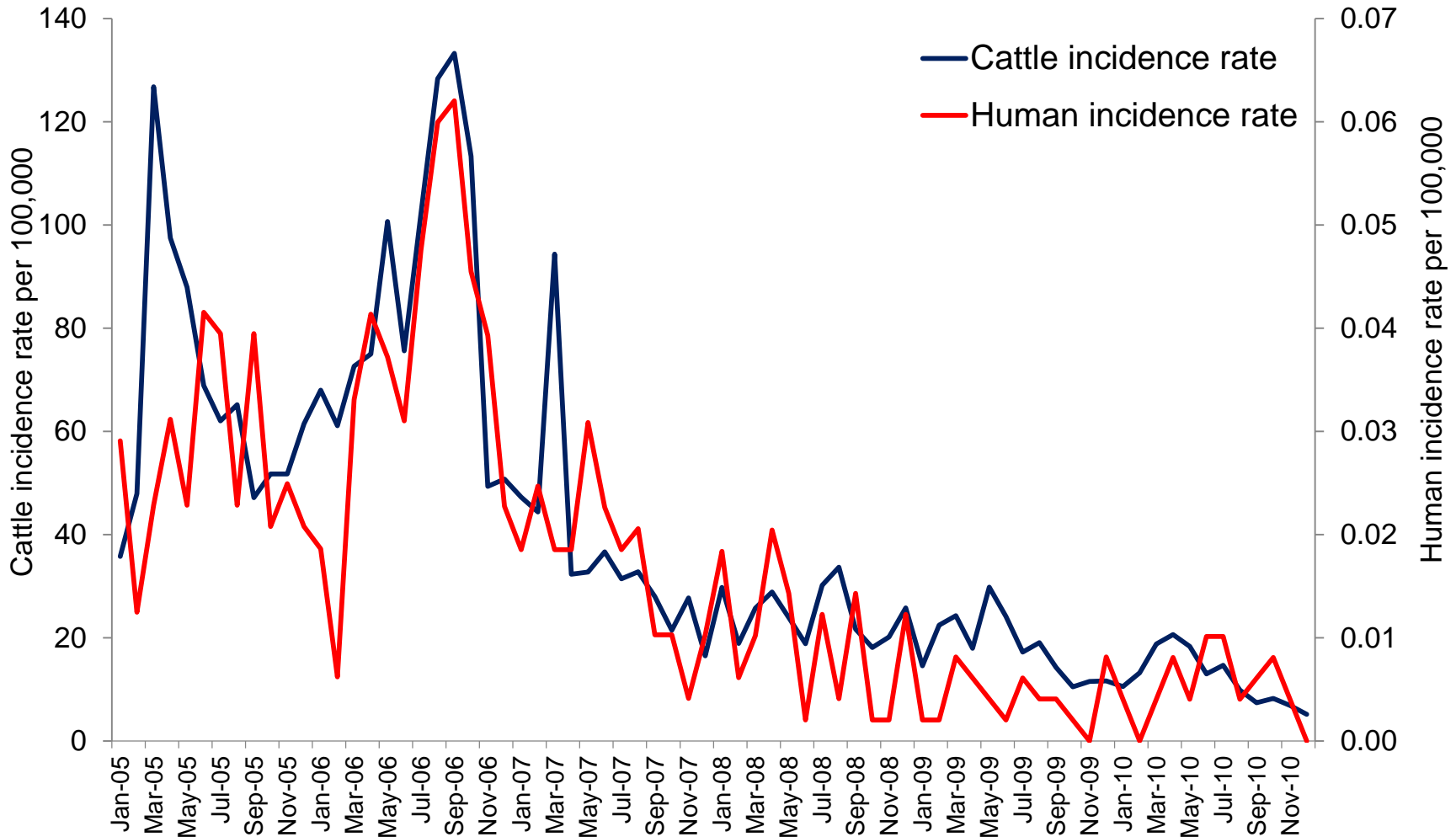
- ARIMAX (autoregressive integrated moving average with exogenous input) model
 - Cross correlation between human and cattle incidence rates
 - Value of $P < 0.05$ was considered significant
- Model validation
 - Divided into model: construction (2005-2007) and validation (2008-2010) phases

Methods

- Conducted simulation intervention scenarios (50% and 75% reductions in cattle cases)
- *MAPE = $\left(\frac{100\%}{n} \sum_{t=1}^n \left| \frac{\text{Actual cases} - \text{predicted cases}}{\text{Actual cases}} \right| \right)$
*Mean Absolute Percentage Error
- Software programs:
 - Microsoft Excel (Redmond, WA, USA)
 - STATA version 11.2 (Stata Corp., College Station, TX, USA)

Results

Descriptive statistics



Monthly incidence rates of brucellosis between 2005 and 2010 in both human and cattle

Descriptive statistics

Month	Human cases	Human (% of total)	Cattle cases	Cattle (% of total)
Jan	44	7.50	5,150	6.91
Feb	25	4.26	5,184	6.96
Mar	47	8.01	8,764	11.76
Apr	61	10.39	6,878	9.23
May	55	9.37	7,460	10.01
Jun	53	9.03	6,032	8.10
Jul	65	11.07	6,676	8.96
Aug	56	9.54	7,427	9.97
Sep	66	11.24	6,431	8.63
Oct	43	7.33	5,725	7.69
Nov	36	6.13	4,345	5.83
Dec	36	6.13	4,421	5.93
Total	587	100	74,493	100

Cases of human and cattle brucellosis reported to the KCDC and AIMS between Jan 1 2005 and Dec 31 2010

Cross-correlation results

human and cattle incidence rates between 2005 and 2010 on a monthly basis with lags of 0-12 months

Lags*(months)	Correlation(r)
0	0.82
1	0.82
2	0.79
3	0.74
4	0.67
5	0.57
6	0.52
7	0.47
8	0.44
9	0.41
10	0.40
11	0.35
12	0.40

* The time periods between two observations. For example, lag 1 is between Y_t and Y_{t-1} . Lag 2 is between Y_t and Y_{t-2} . Time series can also be lagged forward, Y_t and Y_{t+1} .

Final model

Variable	Lags(months)	Coefficient	Standard Error	P-value	MAPE
ARIMAX best model (First difference)^a : ARIMA (0, 1, 1) – AR (0, 1)					
Unpredictable factors	1	-0.86	0.10	<0.001	65.86%
Cattle incidence rate	0	1.84 x 10 ⁻⁴	0.5 x 10 ⁻⁵	<0.001	
Cattle incidence rate	1	1.85 x 10 ⁻⁴	0.5 x 10 ⁻⁵	0.001	
Constant	N/A	8.13 x 10 ⁻¹¹	1.66 x 10 ⁻⁹	0.961	

^aThe first difference of $Y_t = Y_{t-1} - Y_{t-2}$ (Y_t = human incidence rate at time t)

$$\hat{Y}_t = [8.13 \times 10^{-11} + Y_{t-1} + 1.84 \times 10^{-4} X_t + 1.85 \times 10^{-4} X_{t-1} + 0.86 e_{t-1}] \times Y_{popt}$$

\hat{Y}_t = the predicted number of human cases at time t, X_t = the cattle incidence rate at time t
 e_{t-1} = unpredictable factors at time t-1, Y_{popt} = the human population at time t

Examples

- Model Calculations

- Actual human cases: 30

- Predicted human cases in Sep 2006 : **26.12**

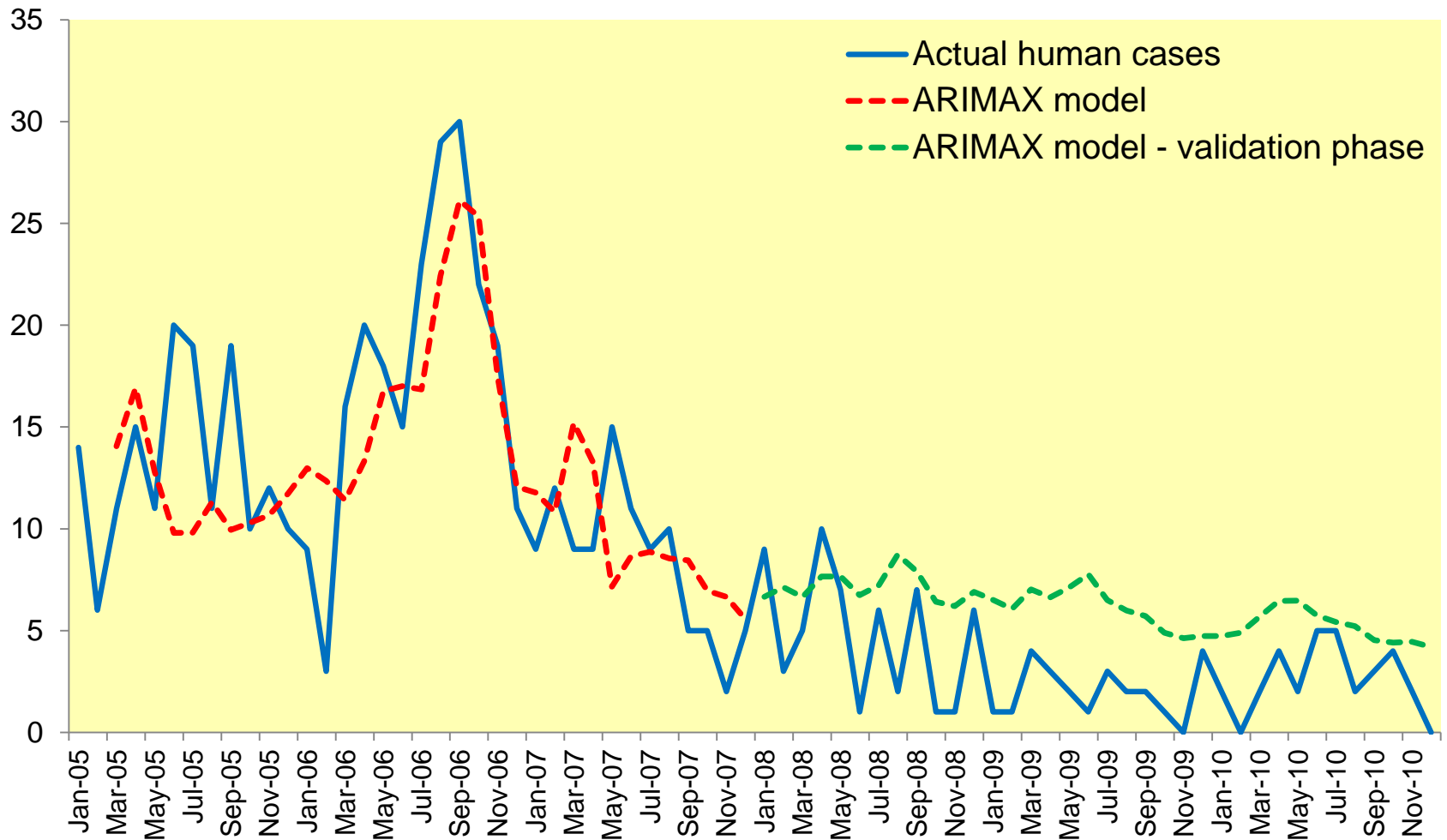
- $$26.12 = [8.13 \times 10^{-11} + 6.00 \times 10^{-7} + 1.84 \times 10^{-4} \times 1.28 \times 10^{-3} + 1.85 \times 10^{-4} \times 1.33 \times 10^{-3} + 0.86 \times e_{\text{Aug2006}}] \times 48,372,000$$

- Actual human cases: 5

- Predicted human cases in July 2010: **2.12**

- $$3.95 = [8.13 \times 10^{-11} + 1.01 \times 10^{-7} + 1.84 \times 10^{-4} \times 1.30 \times 10^{-4} + 1.85 \times 10^{-4} \times 0.15 \times 10^{-4} + 0.86 \times e_{\text{Jun2010}}] \times 49,410,000$$

Predicted versus actual human cases



Limitations

- A limitation of the study potentially lies in not being able to utilize data at the individual province level
- Exposure history of human brucellosis was not available
 - Restricting our ability to adjust for delayed recognition
- Other potential risk factors were not taken into account in the model
 - Surveillance of the wildlife has not been actively implemented to date
 - Farm levels and / or environmental factors

Thank you

ILRI
INTERNATIONAL
LIVESTOCK RESEARCH
INSTITUTE



RESEARCH PROGRAM ON
**Climate Change,
Agriculture and
Food Security**

