WOAH SRR-SEA capacity building on risk analysis for transboundary animal disease control purposes in Southeast Asia

World Organisation for Animal Health

Australian Government

Department of Agriculture, Fisheries and Forestry

EXTRA UNIT 1
SPATIAL ANALYSIS

Department of Emerging diseases and Global health

Animal Health Research Centre (CISA)

Institute for Agronomic and Food Research (INIA)

Spain's Research Council (CSIC)

Contact: iglesias@inia.csic.es

CISA

INIA CSIC

Outline

- Introduction to Spatial Analysis in Veterinary Epidemiology and Basics of Geographic Information Systems (GIS)
- Data visualization in Geographic Information Systems (GIS)
- Exploratory spatial data analysis (ESDA)
- Introduction to SaTScan: Cluster analysis
- Interpolation and spatial smoothing techniques
- Knowledge check questions
- Resources

PART 1
INTRODUCTION AND GIS BASICS

Spatial epidemiology in veterinary science

Spatial epidemiology focuses on the **geographic location of diseases to identify distribution patterns and factors influencing spread**.

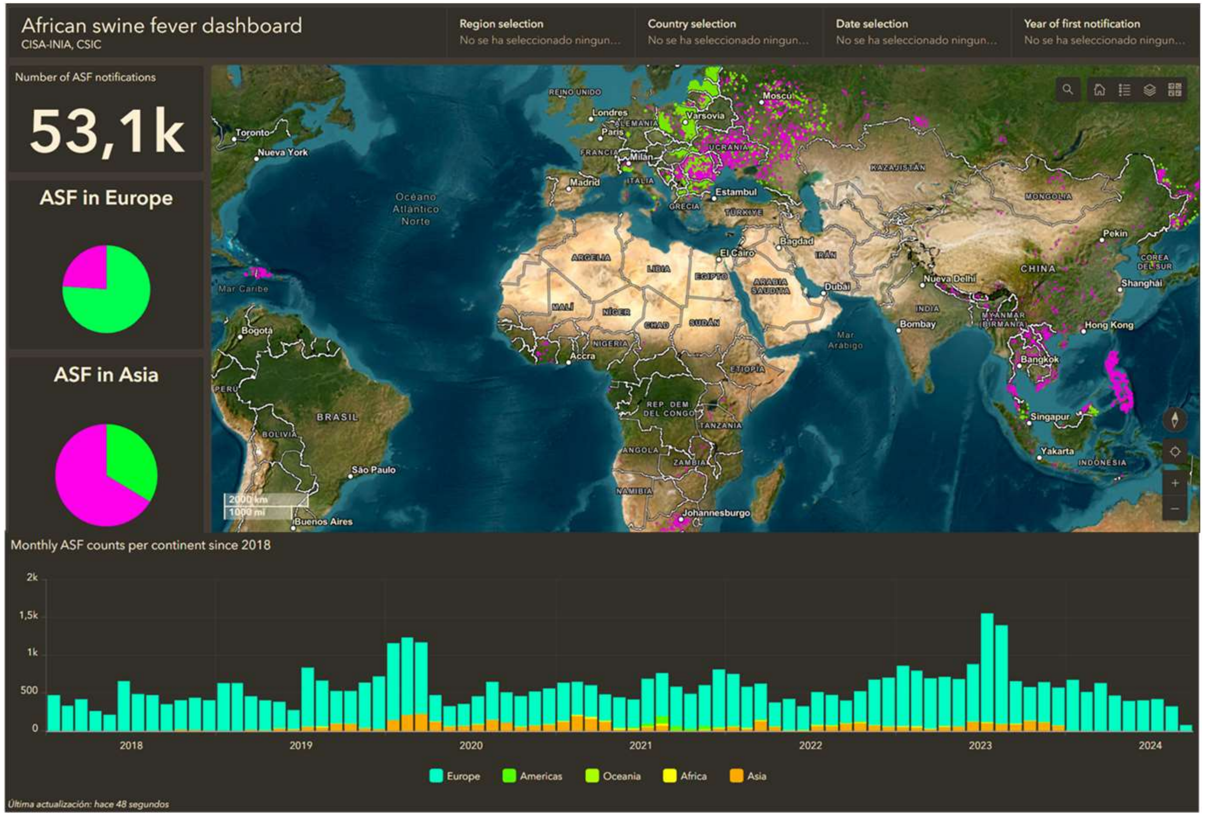
This approach is crucial in veterinary medicine for diseases which require constant monitoring and identification of risk areas (such as ASF or HPAI).

Objectives of spatial epidemiological analysis are (*Pfeiffer, 2008*):

- **the description of spatial patterns,**
- **the identification of clustering of disease cases**
- **the description or prediction of disease risk**

Spatial epidemiology in veterinary science

Geographic Information Systems (GIS) provide spatial analysis tools that, along with other spatiotemporal cluster analysis tools, enable detailed studies of disease patterns across different areas and time periods, facilitating decision-making in the implementation of control measures, surveillance, and prevention in affected areas.

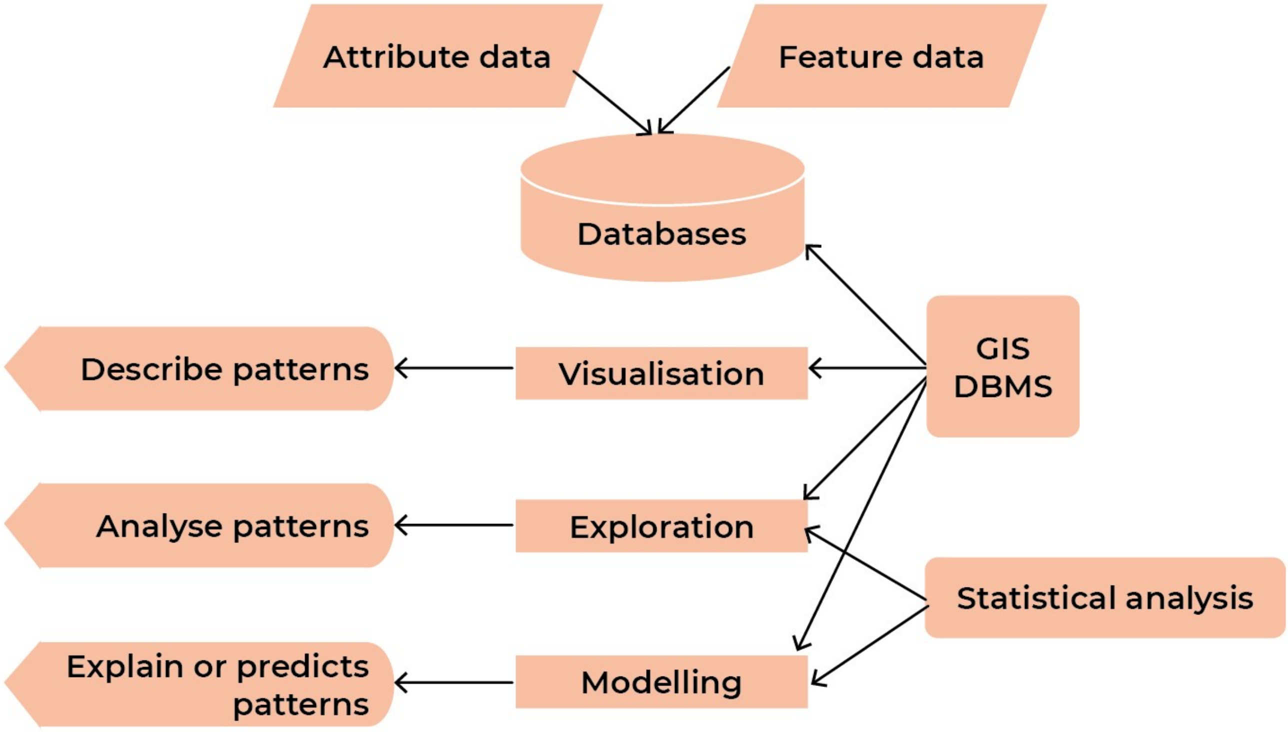


CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

Spatial epidemiology in veterinary science

The spatial analysis process comprises 3 steps: **visualization, exploration** and finally **modelling.** The last steps are helped by statistical methods



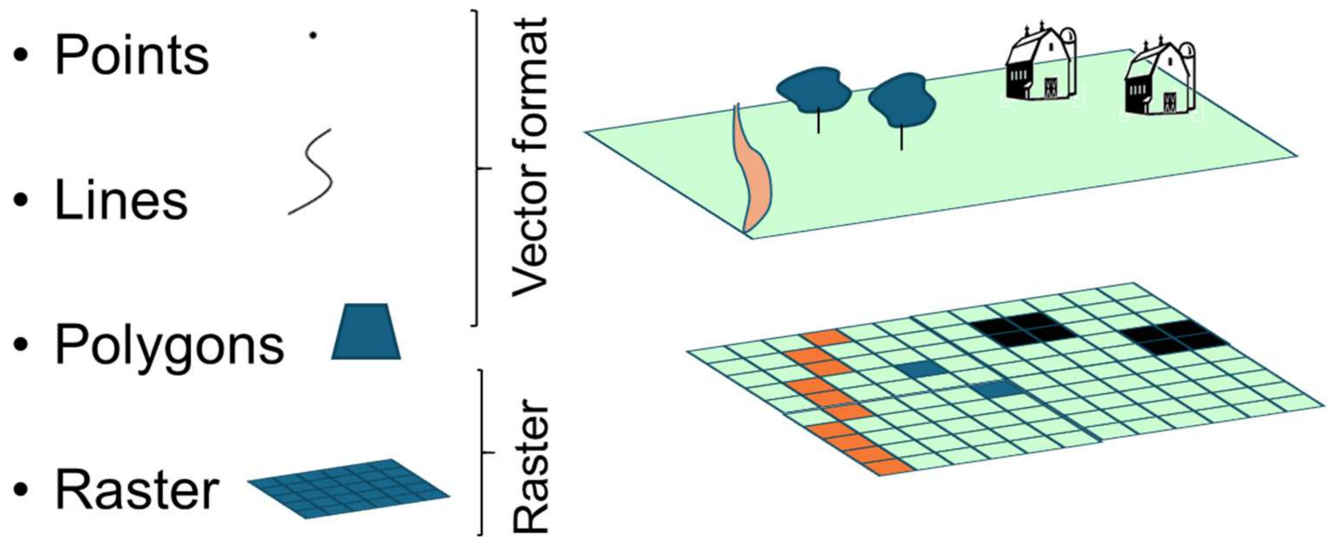
Spatial Analysis in Epidemiology,
Pfeiffer *et al.* 2008

History of spatial analysis in health science

Pumps
Cholera cases

Geographic Information Systems (GIS): Definition and components

GIS are essential tools in spatial analysis, capable of capturing, storing, managing, analyzing, and visualizing geospatial data. They facilitate understanding of disease distribution and patterns through a wide array of data represented in vector and raster formats.



GIS Software and applications

GIS software, including paid and free programs like ArcGIS and QGIS, provides powerful tools for spatial analysis. These tools are crucial for detailed studies of disease patterns and aiding in decision-making for control, surveillance, and prevention measures.:

Exercise 1:Exploring SARS-CoV-2 data in animals using SARS-BOARD

The SARS-BOARD, hosted by INIA-CSIC, Spain, provides an interactive tool for exploring SARS-CoV-2 outbreaks in various types of animals (captive, domestic, wild, and pets).

Purpose: To enable detailed examination of outbreak data across different species and regions using a dynamic dashboard.

<u>Dashboard Features:</u>

- Historical records of outbreaks since 2020.
- Time slider for selecting specific periods and visualizing outbreaks.
- Interactive maps and graphics detailing outbreak regions.

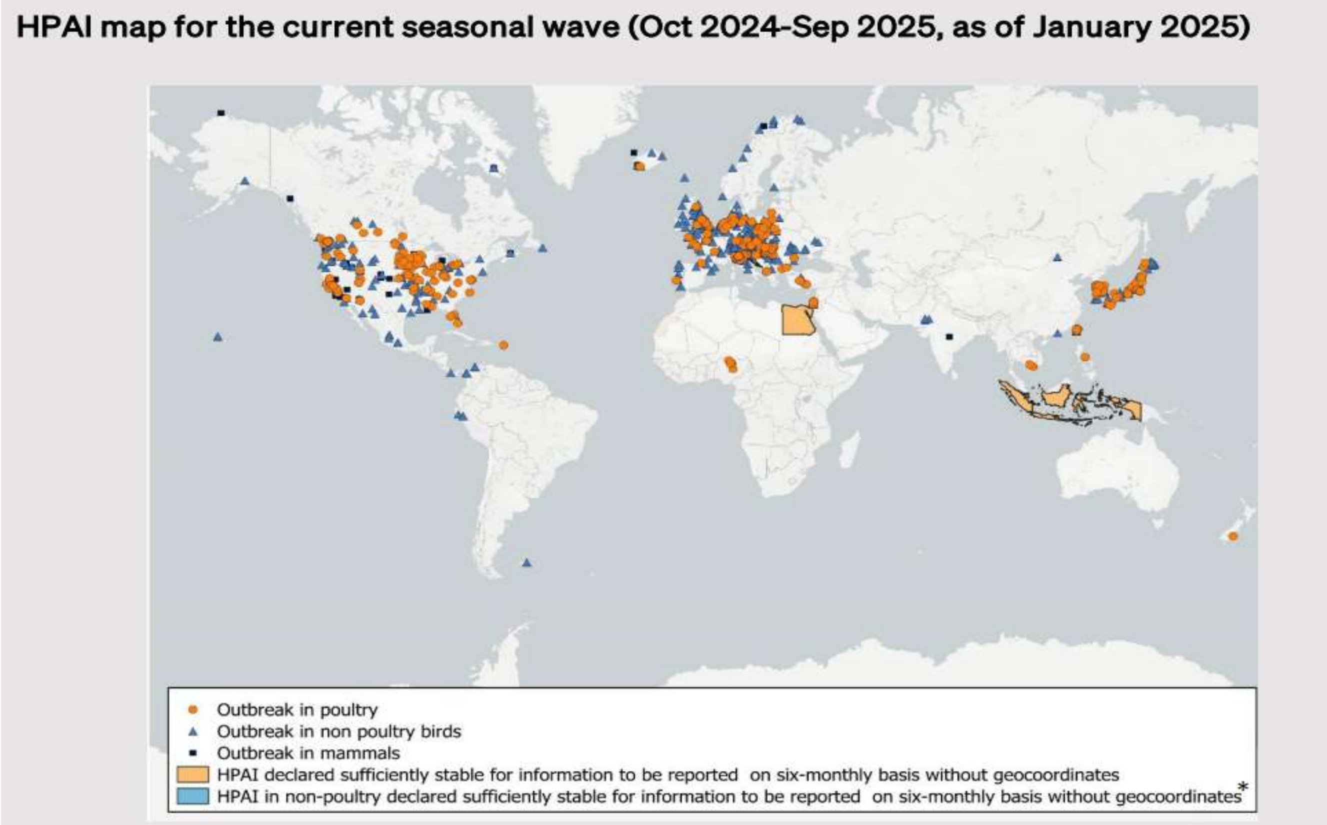
<u>Data Sources:</u>

- Outbreak data from the World Animal Health Information System (WAHIS).
- Climate data from Climate Data Online (CDO).
- Socioeconomic factors from World Bank Data.

PART 2
DATA VISUALIZATION

Data visualization on maps

- **Thematic maps** are graphical representations that highlight a specific variable (e.g., case density, farm distribution, risk areas) within a geographic area.
- They are **essential for identifying geographical patterns** of diseases
- Example: *A thematic map of high pathogenic avian influenza disease presence by region/province can help identify the most affected areas.*

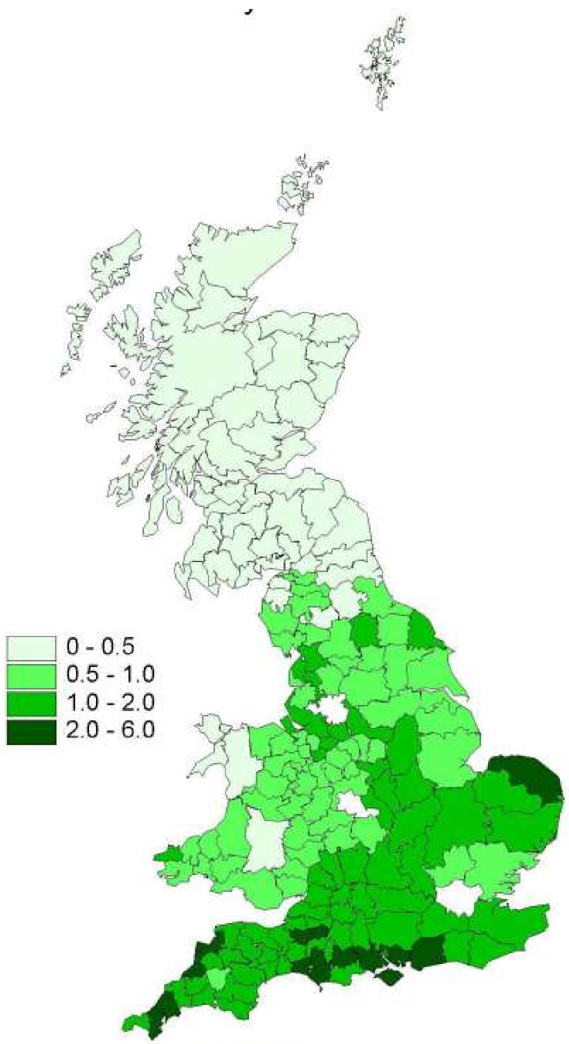


Source: **High Pathogenicity Avian Influenza (HPAI) – Situation Report 67 WOAH. January 2025**
https://www.woah.org/app/uploads/2025/02/hpai-report-67.pdf

Symbolization in GIS

Symbolization encodes information through colors, sizes, and shapes.

- Colors by area: Could represent outbreak severity



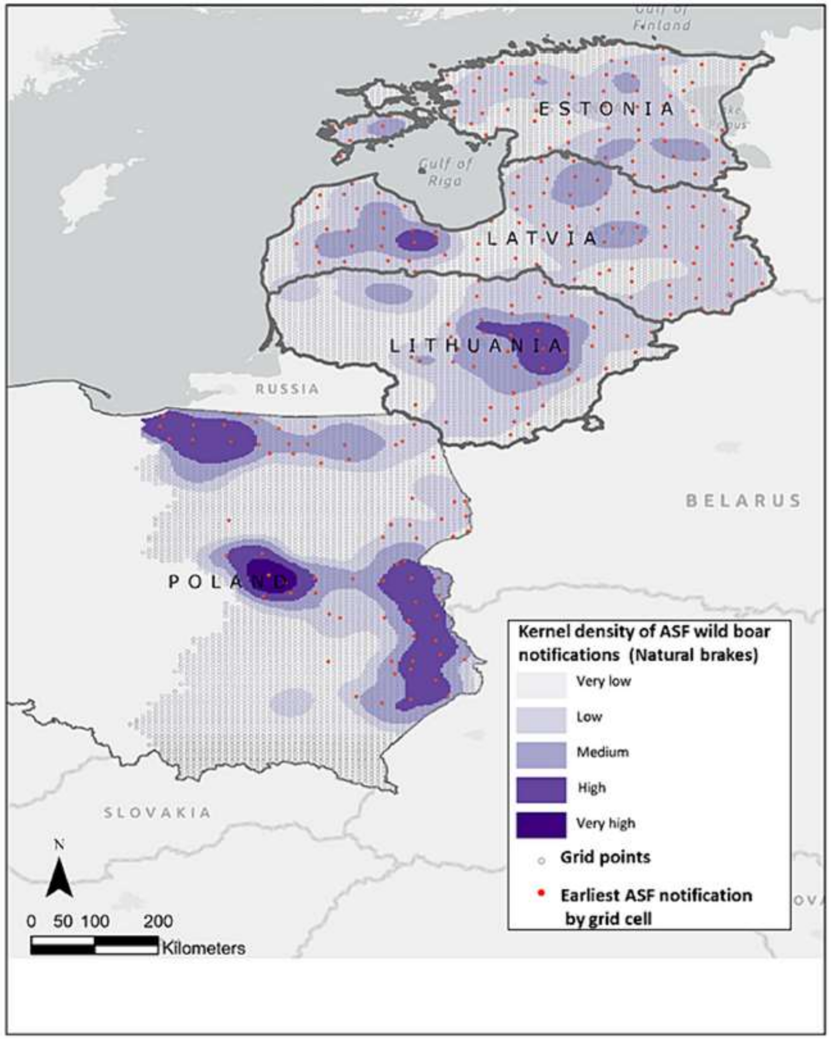
Standardised morbidity ratios for BSE in British cattle born before 30 June 1988. *Source:* "Analysis of disease count data" of OIE Sub-Regional training on applying Geographic Information Systems (GIS) for advanced spatial analysis of animal health data 16 – 19 October, 2018

Symbolization in GIS

Symbolization encodes information through colors, sizes, and shapes.

- Colors continuous : Could represent outbreak density

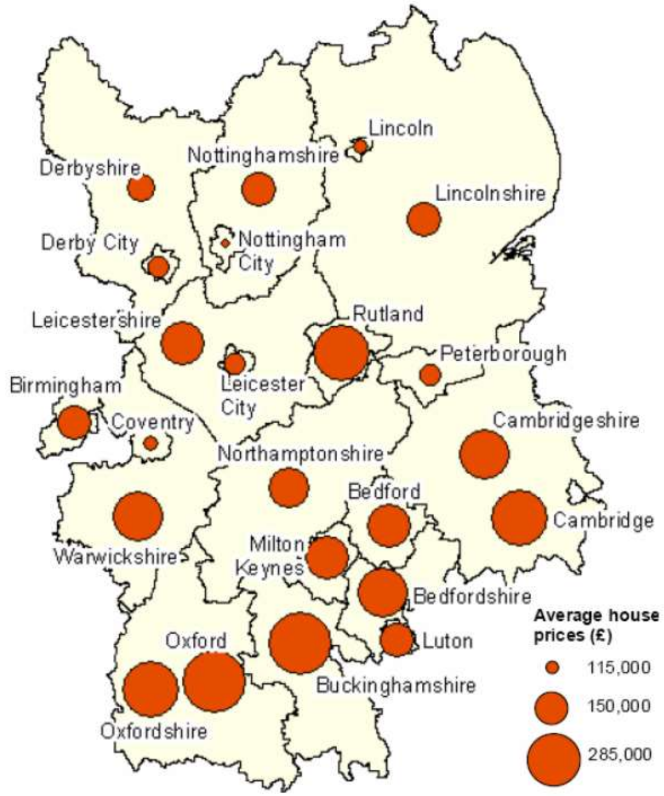
Kernel density of ASF wild boar notifications from January 2014 to January 2022 in Estonia, Latvia, Lithuania and Poland (Europe) *Source:* Martínez Avilés M, Montes F, Sacristán I, de la Torre A and Iglesias I (2024) Spatial and temporal analysis of African swine fever frontwave velocity in wild boar: implications for surveillance and control strategies.
Front. Vet. Sci. 11:1353983. doi: 10.3389/fvets.2024.1353983



Symbolization in GIS

Symbolization encodes information through colors, sizes, and shapes.

- Point sizes: Could indicate number of infected animals at each location.



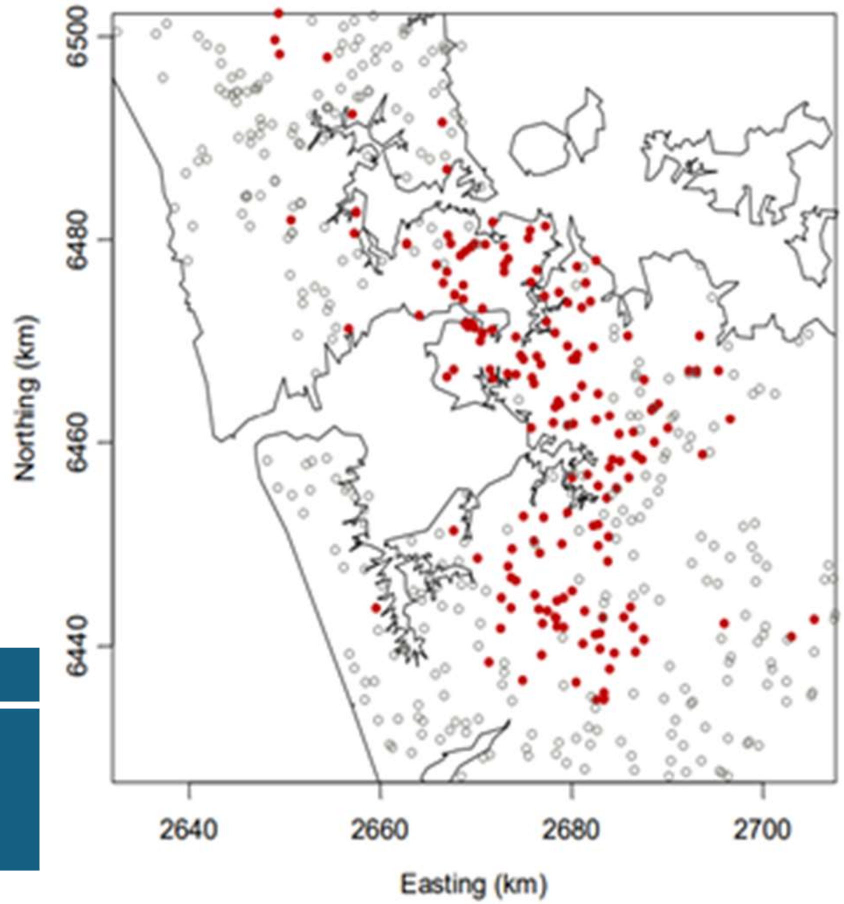
Liver fluke in dairy herds in Victoria, Australia. Point map showing the location of study herds around Maffra. The size of each point is proportional to the number of animals tested in each herd. Colour indicates the apparent fluke prevalence in each study herd. *Source:* "Analysis of disease count data" of OIE Sub-Regional training on applying Geographic Information Systems (GIS) for advanced spatial analysis of animal health data 16 – 19 October, 2018

Symbolization in GIS

Symbolization encodes information through colors, sizes, and shapes.

- Point color : Could indicate type of outbreaks , presence of disease..etc

Varroa destructor in honey bees in the greater Auckland area of New Zealand. *Source:* "Analysis of disease count data" of OIE Sub-Regional training on applying Geographic Information Systems (GIS) for advanced spatial analysis of animal health data 16 – 19 October, 2018

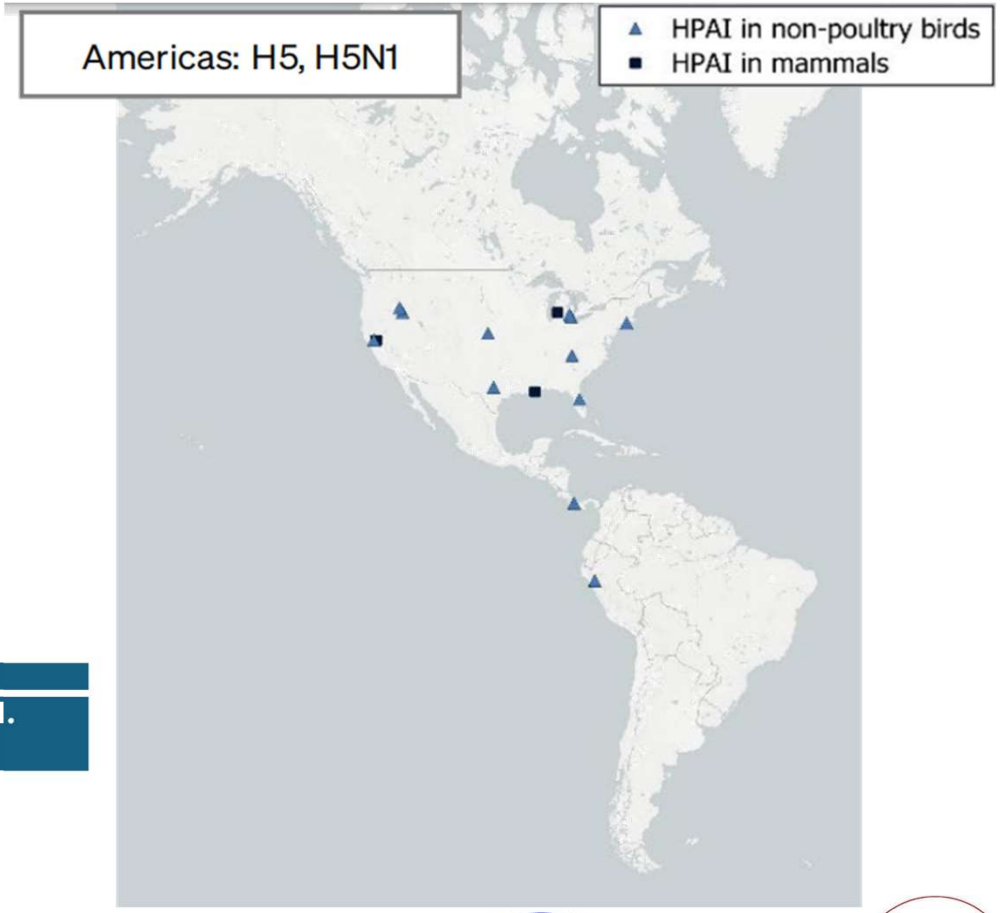


Symbolization in GIS

Symbolization encodes information through colors, sizes, and shapes.

- Point shape : Could indicate type of outbreaks (species, diseases, etc)

Source: **High Pathogenicity Avian Influenza (HPAI) – Situation Report 67 WOAH. January 2025** https://www.woah.org/app/uploads/2025/02/hpai-report-67.pdf



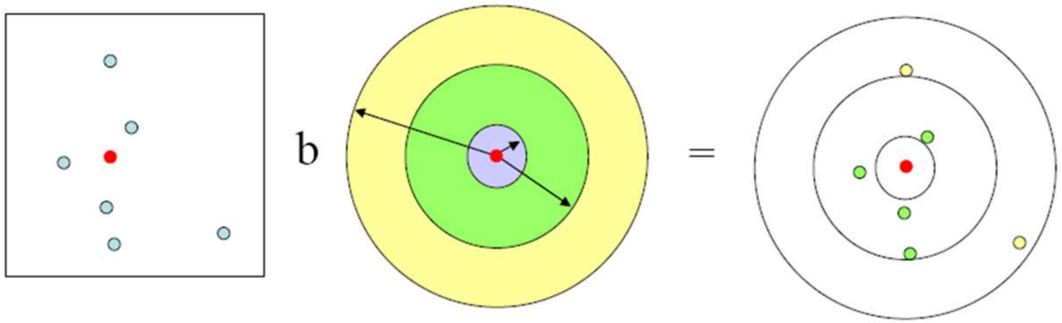
Proximity tools - Distances and buffers

❖ Distance calculation:

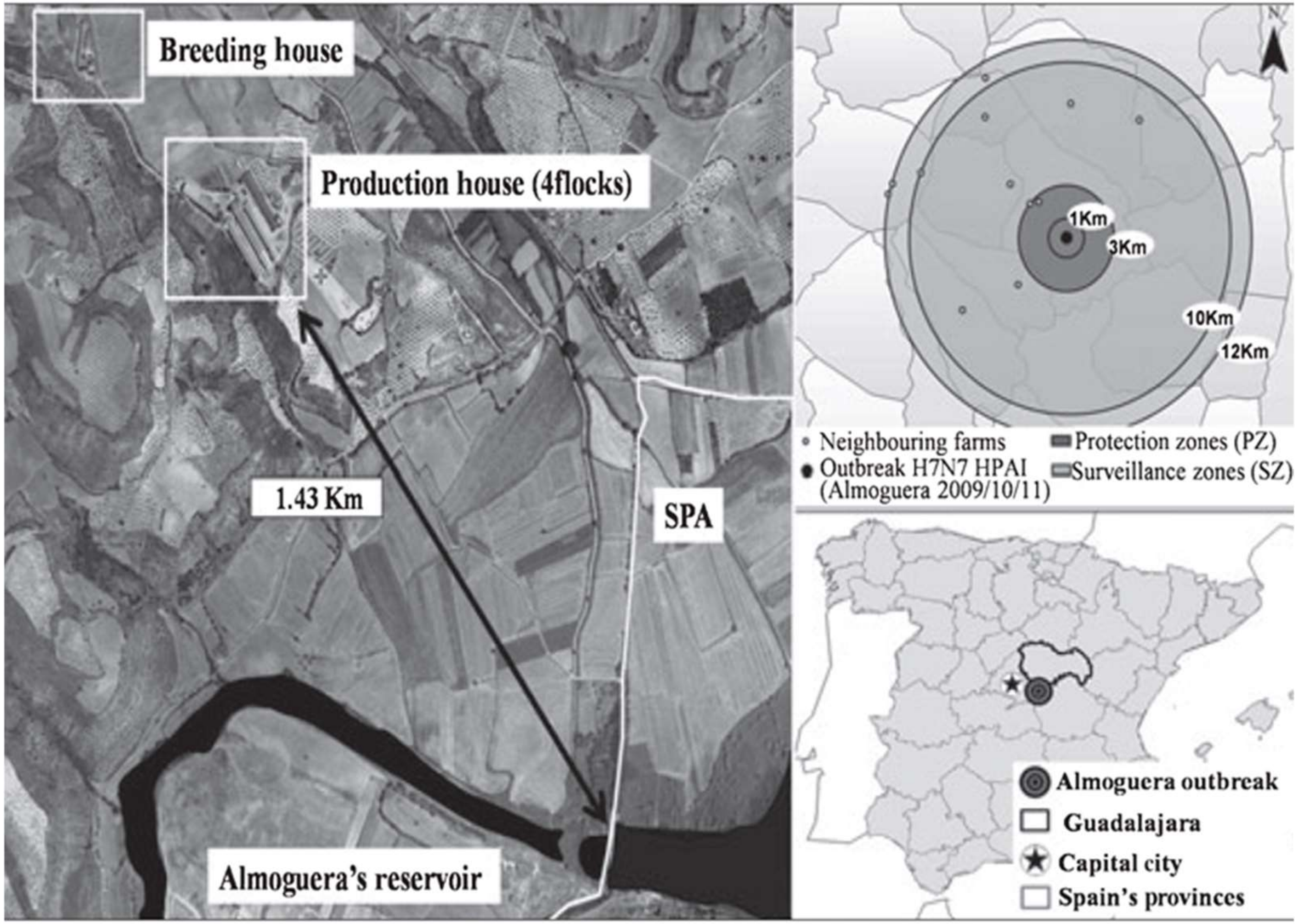
- Measures distances between points (e.g., infected farms) or between geographic elements (e.g., wetlands, roads, rivers) and animal population areas (e.g., farms or live animal markets).
- Essential for assessing potential disease transmission pathways.

❖ Buffers:

- Buffer zones are areas of influence created around a point, line, or polygon at a specified distance.
- They help identify which geographic features lie within a set proximity to an outbreak.



Proximity tools - Distances and buffers

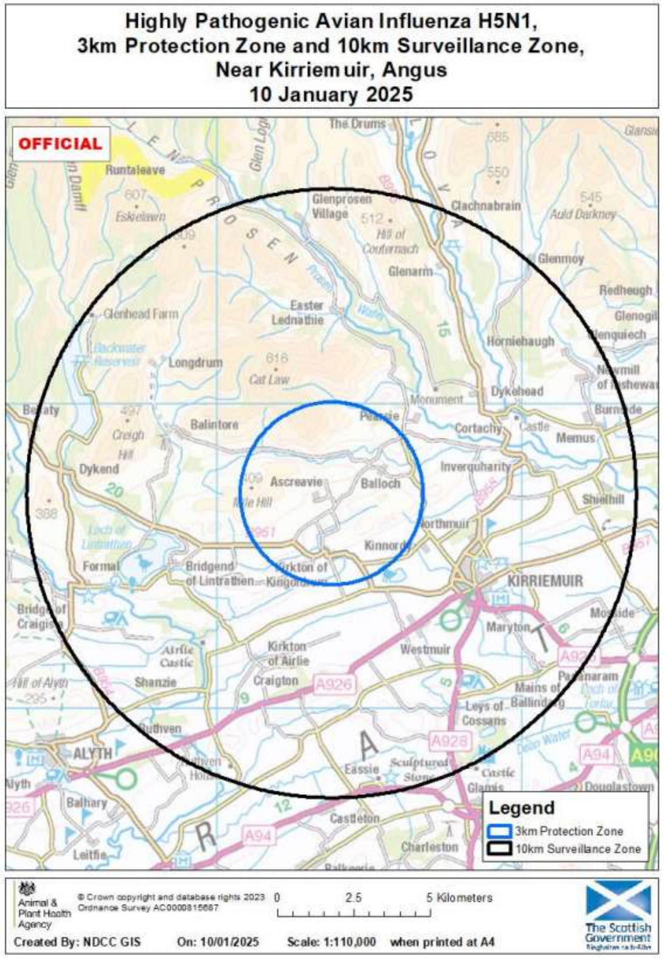


*Left :Ortophoto of HPAI-affected holding (production house) and **distance** to the nearest reservoir, located within a SPA (Special Protection Area for birds, delimited with a white line).*
*Right : Neighboring farms (white dots) within the **buffers** of protection and surveillance zones (top) and the location of the HPAI outbreak in Spain (bottom).*
Source: Iglesias, I., Martínez, M., Muñoz, M. J., De La Torre, A., & Sánchez-Vizcaíno, J. M. (2010). First case of highly pathogenic avian influenza in poultry in Spain. Transboundary and emerging diseases, 57(4), 282-285.

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTER

INIA CSIC

Application of buffers in veterinary epidemiology

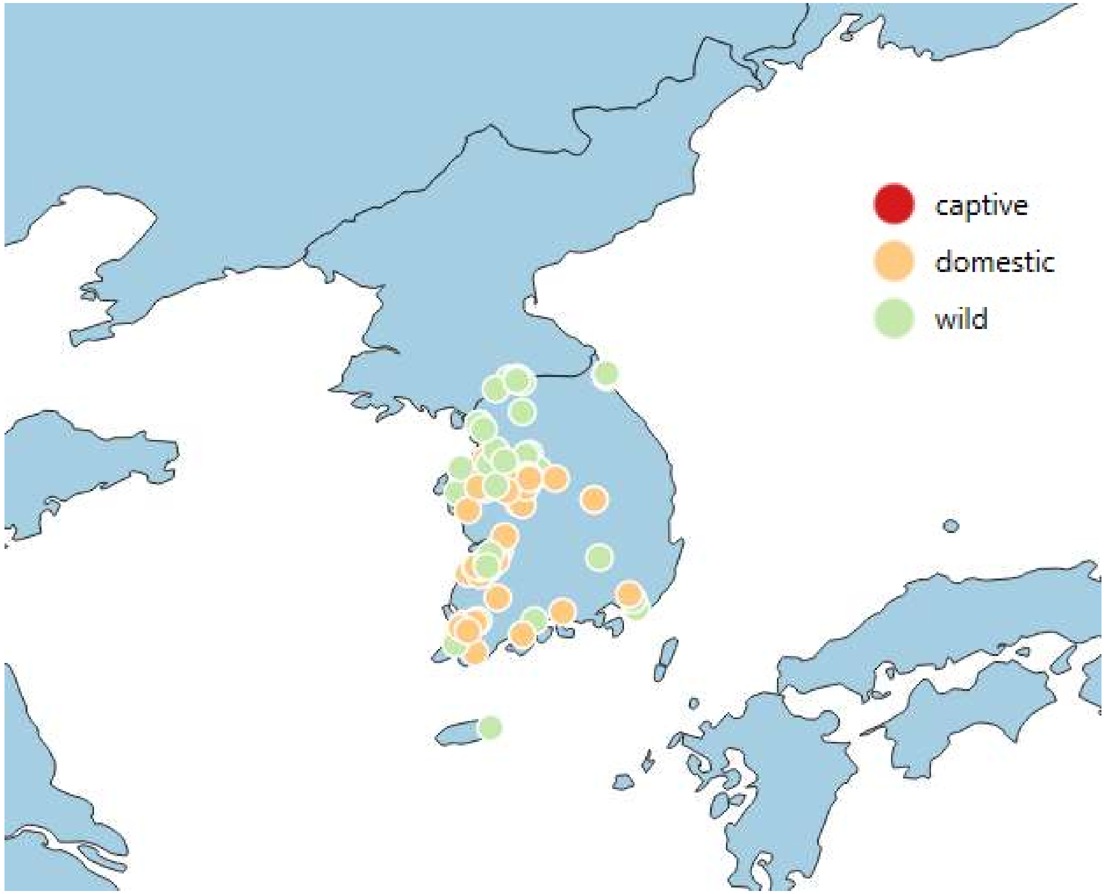


Map Showing the HPAI Protection Zone and Surveillance Zone in Kirriem uir. Source: [Scottish Government's website](#)(January 2025)

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

Exercise 1:Simbology



DATASET

DATASET: Download Link: https://saco.csic.es/s/bcYcx8oAqoymPyd

The dataset consists of **two spatial layers**:

1) Countries Layer (Shapefile "**countries**"): polygons

A global shapefile containing the geographical boundaries of all countries worldwide.

Used as a reference layer for spatial analysis.

2) HPAI H5N1 Outbreaks Layer (Shapefile " **Outbreaks**"): points

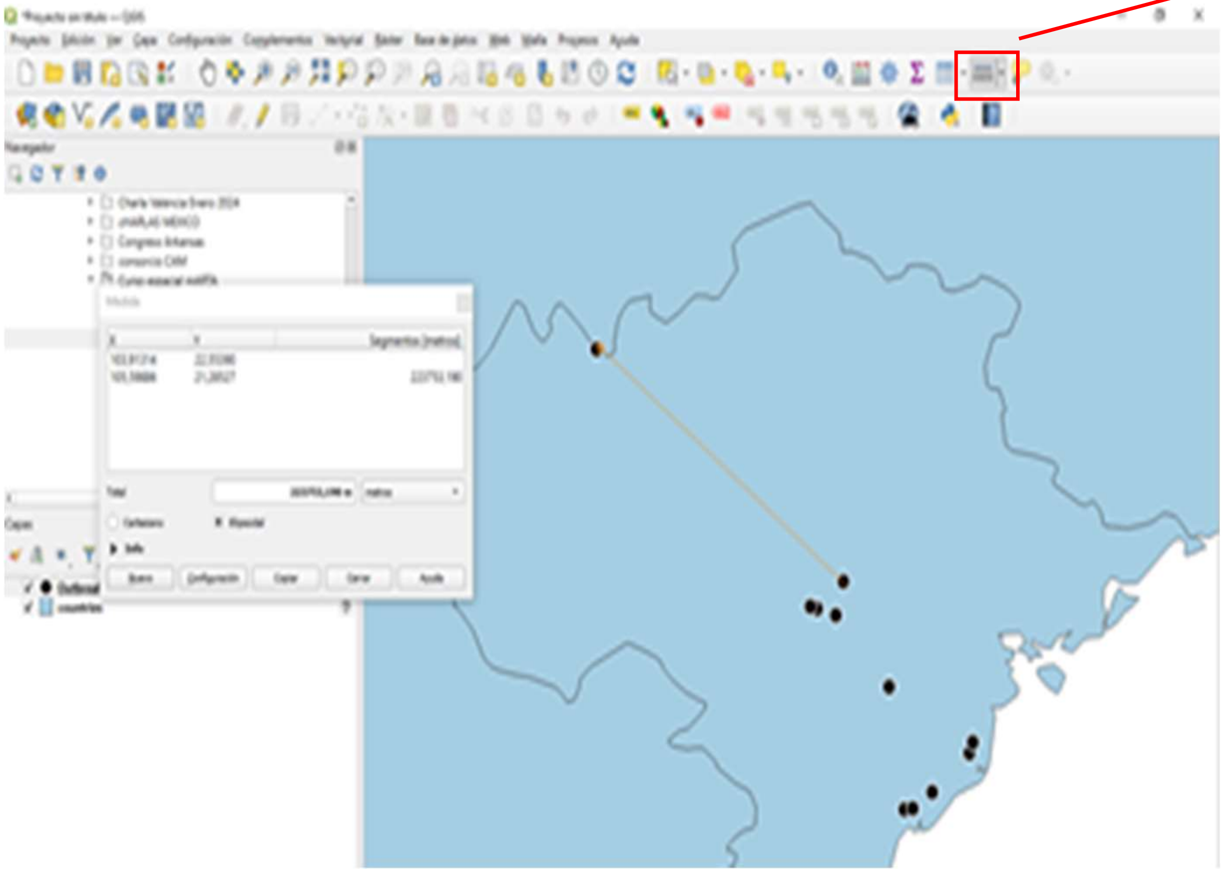
Contains 396 reported outbreaks of Highly Pathogenic Avian Influenza (HPAI) H5N1 in five Asian countries (Cambodia, South Korea, Laos, Philippines, Vietnam) during 2021 and 2022. The data includes outbreak notifications from the World Organisation for Animal Health (OMSA). Each outbreak point is georeferenced and represents a reported case of HPAI H5N1 in poultry or wild birds.

This dataset can be used for Exercises

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

Exercise 2: Distance



Exercise 3: Buffer

PART 3
EXPLORATORY DATA ANALYSIS

Introduction ESDA

Exploratory

Spatial

Data

Analysis

<u>Questions that ESDA helps to answer</u>

- Is the variable I am looking at concentrated in space?
- Do similar values tend to be located close to each other?
- Can I identify a specific area where certain values cluster together?

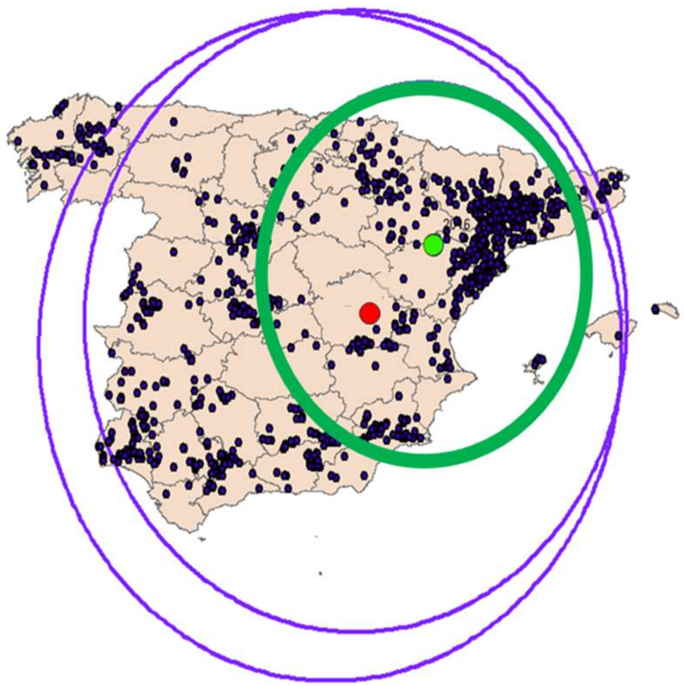
↓

What is behind this pattern?
What could be generating the process?
Why do certain clusters appear in space?

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTER

INIA CSIC

Descriptive spatial statistics:



- Poultry farms in Spain

Mean center

$$\bar{X} = \sum_{i=1}^{N} \frac{X_i}{N}$$

$$\bar{Y} = \sum_{i=1}^{N} \frac{Y_i}{N}$$

Spatial Standard Deviation

$$SD_{x,y} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{N-1} + \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{N-1}}$$

Weighted by census
Mean Center

$$\bar{X} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

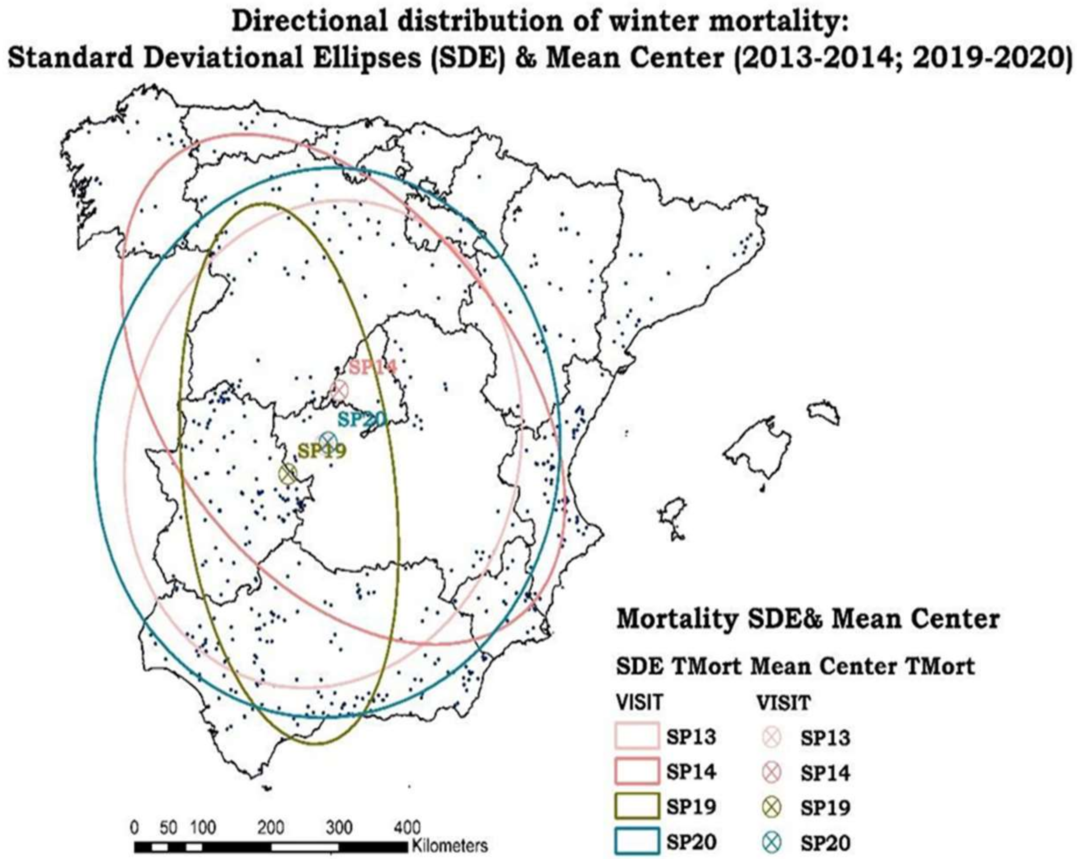
$$\bar{Y} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i}$$

Weighted by census
Spatial Standard Deviation

$$\mathrm{WSD} = \sqrt{\frac{\sum_{i=1}^{n} w_i \left[(x_i - \bar{x}_w)^2 + (y_i - \bar{y}_w)^2\right]}{\sum_{i=1}^{n} w_i}}$$

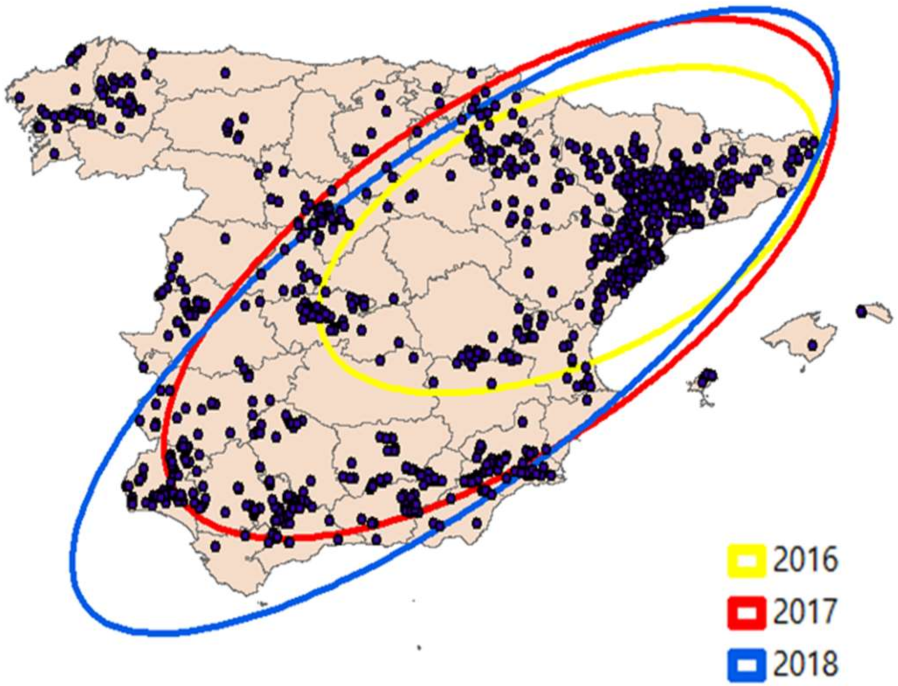
Descriptive spatial statistics:

Standard deviation ellipses (SDEs) and mean center (MC) of apiary winter mortality (SP) occurring between the springs of 2013-14 and 2019-2020 reflect a pronounced change of pattern, showing a shift to a North-South orientation in the western region of Spain in SP19. Each ellipse is drawn from data collected during the annual spring visits (SP). Source: Perez-Cobo et al, 2025



Descriptive spatial statistics:

Example of Standard deviation ellipses (SDEs) *by year using simulated annual data on poultry census variation in farms in Spain over three years (2016-2018)*



Spatial autocorrelation

The definition of spatial autocorrelation measures the extent to which nearby objects are similar.

Why is spatial autocorrelation important?
One of the main reasons why **spatial autocorrelation** is important is that **statistical analyses rely on the assumption of independent observations**. When spatial autocorrelation is present in a map, this violates the assumption that observations are **independent from one another**, potentially leading to **biased results and incorrect inferences**.

Tobler

Spatial autocorrelation

Statistical representation of Tobler's Law

In traditional statistics is equivalent to correlation.

To measure spatial autocorrelation, we use Moran's I test.

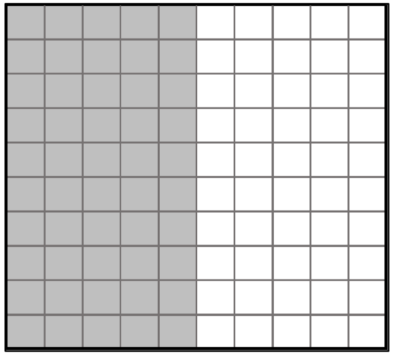
Spatial autocorrelation

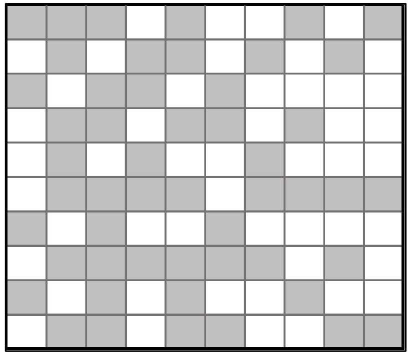
Moran's I can be classified as positive, negative, or with no spatial autocorrelation.

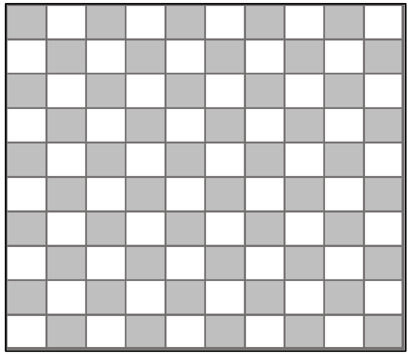
- Positive spatial autocorrelation occurs when similar values cluster together on a map.
- No spatial autocorrelation happens when similar values are randomly distributed.
- Negative spatial autocorrelation occurs when similar values are far apart from each other.

Income
Poverty
Vegetation
Temperature

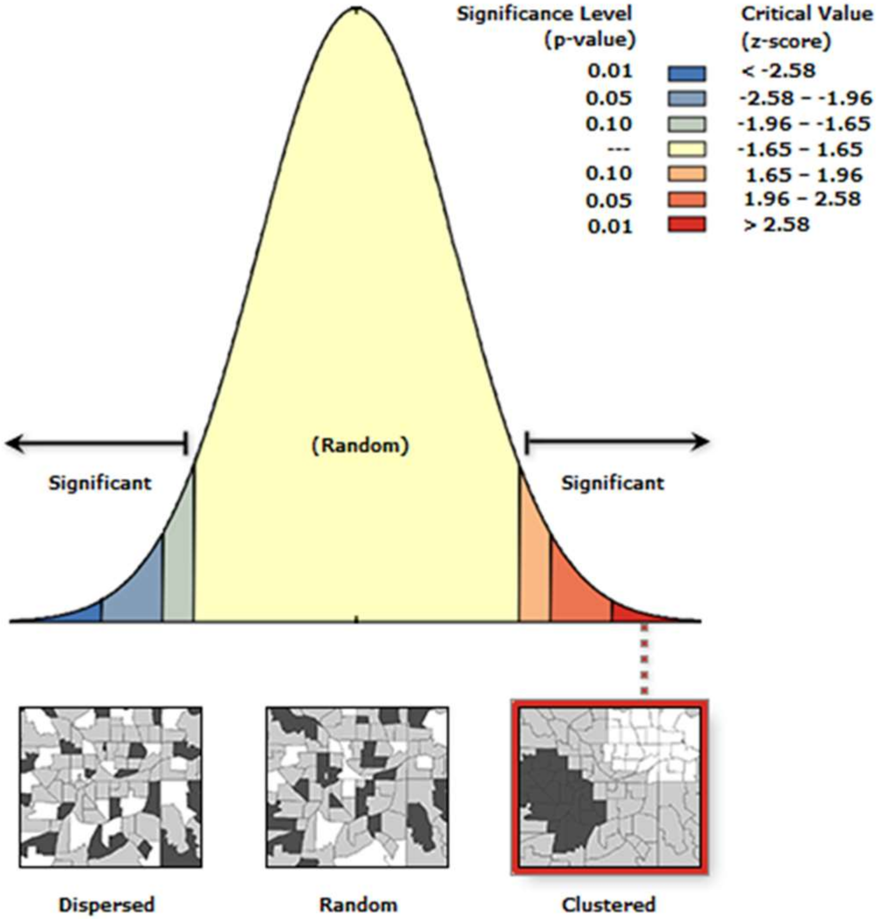
Supermarkets
Police Stations
Fire Stations
Hospitals


Positive autocorrelation


Negative autocorrelation


No spatial autocorrelation

Spatial autocorrelation



$$I = \frac{N}{W} \times \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

Where:

- I = Moran's I index
- N = Total number of spatial units (observations)
- X_i = Value of the variable at location i
- $\bar{X}$ = Mean of the variable
- w_{ij} = Spatial weight between locations i and j
- W = Sum of all spatial weights ($W = \sum_i \sum_j w_{ij}$)

The null hypothesis for spatial pattern analysis tools is complete spatial randomness. The null hypothesis is typically rejected when the p-value is less than 0.1.

Hot Spot analysis (Getis-Ord)

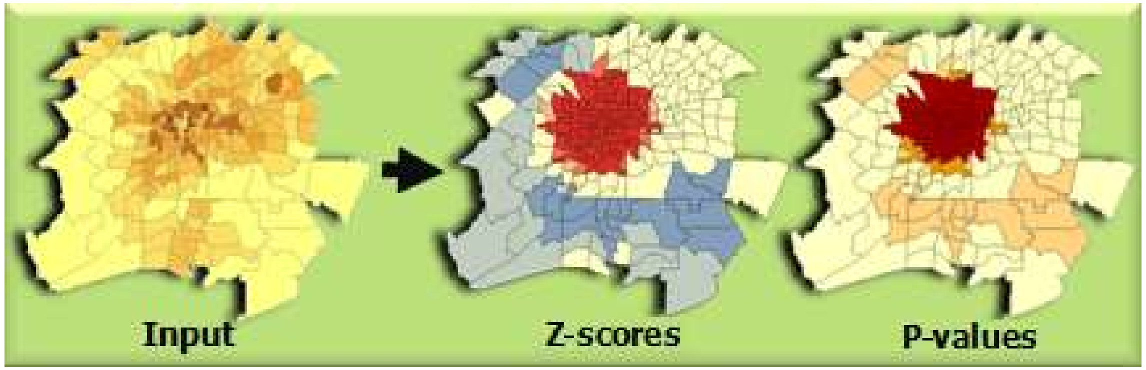
Measures spatial autocorrelation at a local level (as if locally aggregating intensity values).

Indicates areas of high and low clustering.

Requires an intensity level of the event, for example, the number of cases per province. An input field such as **year** can be used for temporal analysis.

The *Gi statistic* → *Z-score interpretation**: *If Gi is significant**:

- **Z > 1** → The larger the Z-score, the stronger the clustering of high values (**hot spot**).
- **Z < 1** → The smaller the Z-score, the stronger the clustering of low values (**cold spot**).

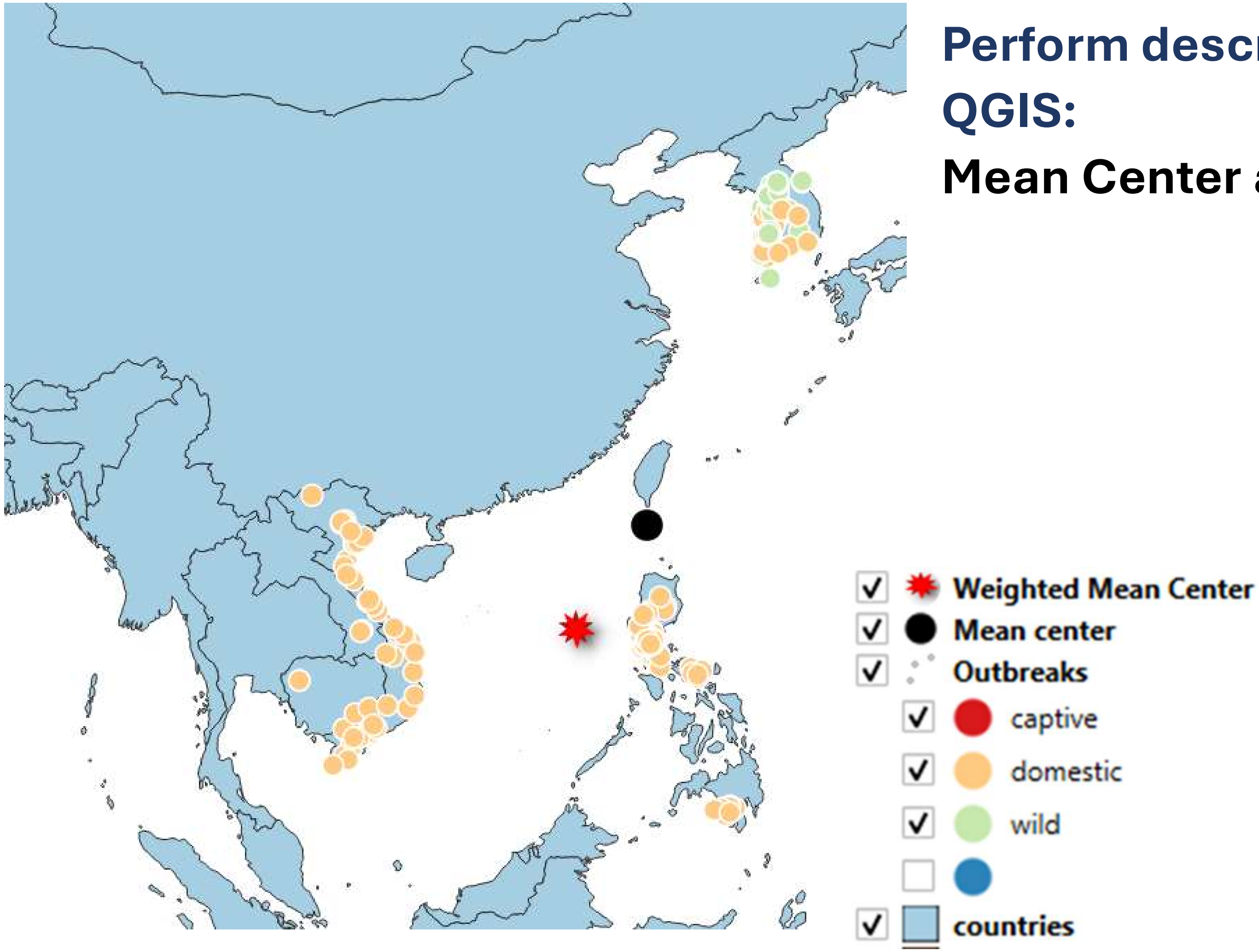


https://desktop.arcgis.com/es/arcmap/latest/tools/spatial-statistics-toolbox/hot-spot-analysis-with-rendering.htm

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTER

INIA CSIC

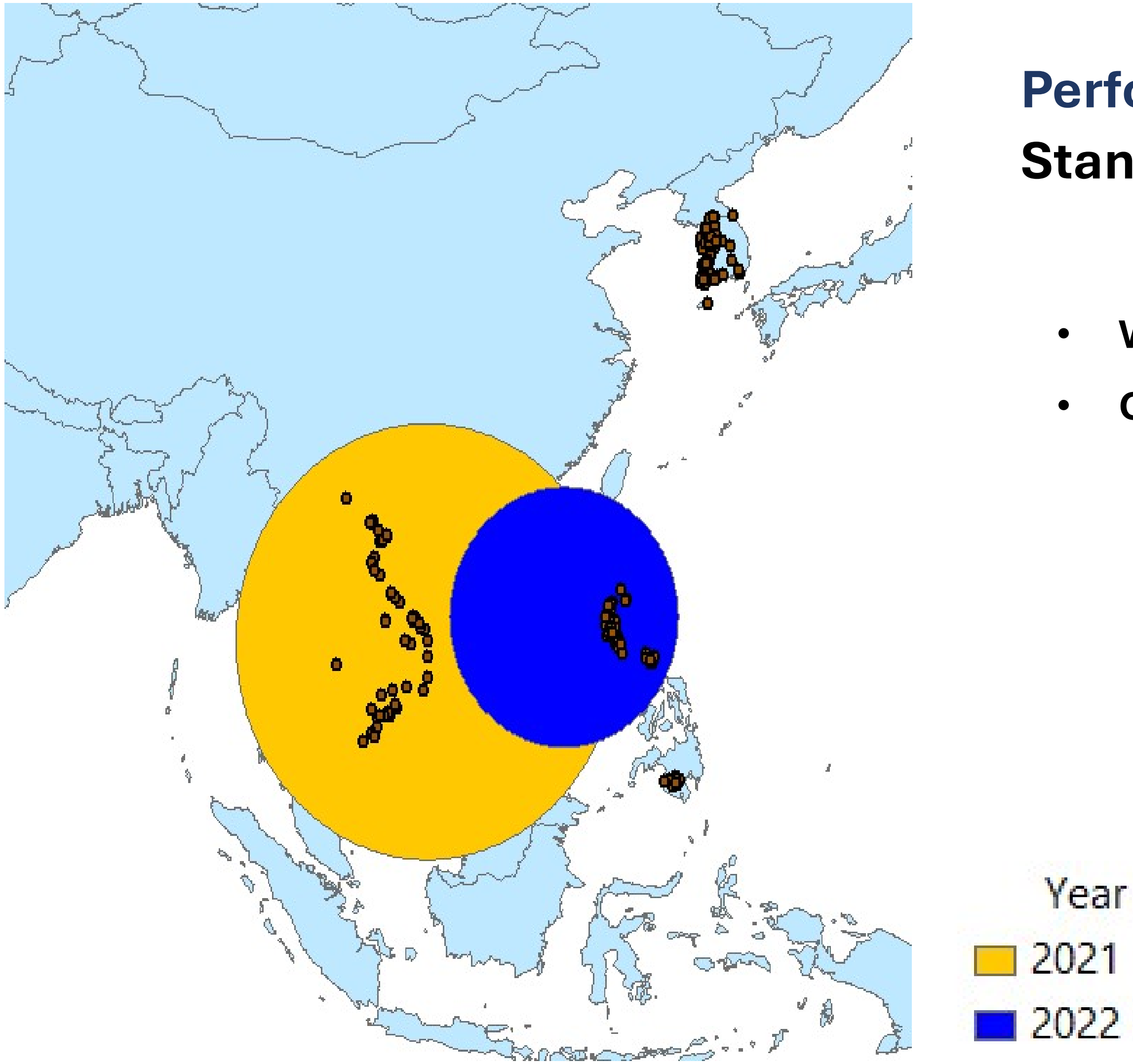
Exercise 3 (Optional): Perform descriptive spatial statistics



Perform descriptive spatial statistics in QGIS:

Mean Center and Weighted Mean Center

Exercise 3 (Optional): Perform descriptive spatial statistics



Perform descriptive spatial statistics

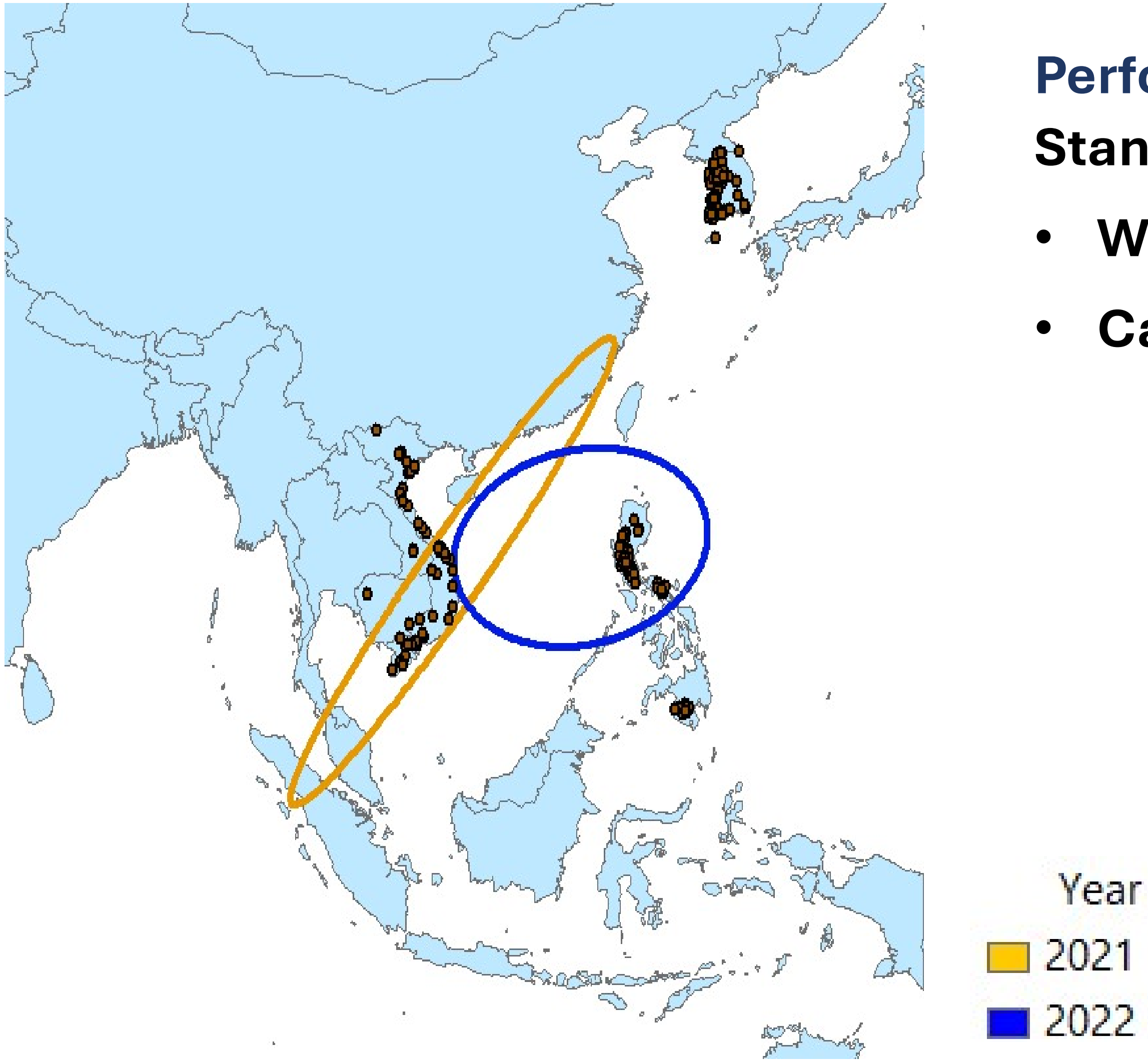
Standard deviation (ArcGIS)

- **Weight field : Cases**
- **Case field: year**

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTER

INIA CSIC

Exercise 3 (Optional): Perform descriptive spatial statistics



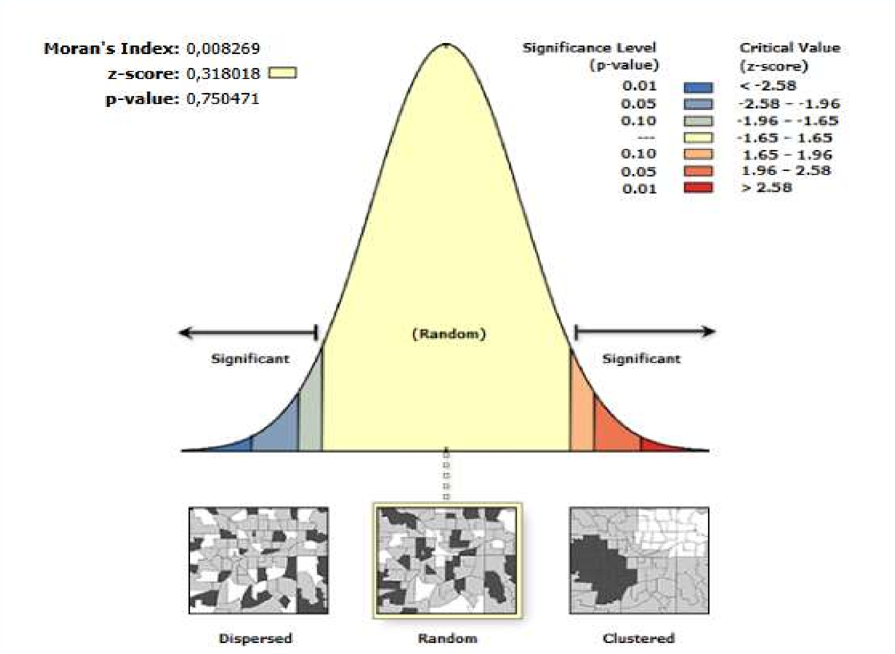
Perform descriptive spatial statistics:

Standard Deviation Ellipse (SDE) (ArcGIS)

- **Weight field : Cases**
- **Case field: year**

Exercise 3 (Optional): Autocorrelation and pattern analysis

Spatial Autocorrelation Report



Given the z-score of 0.318018236983, the pattern does not appear to be significantly different than random.

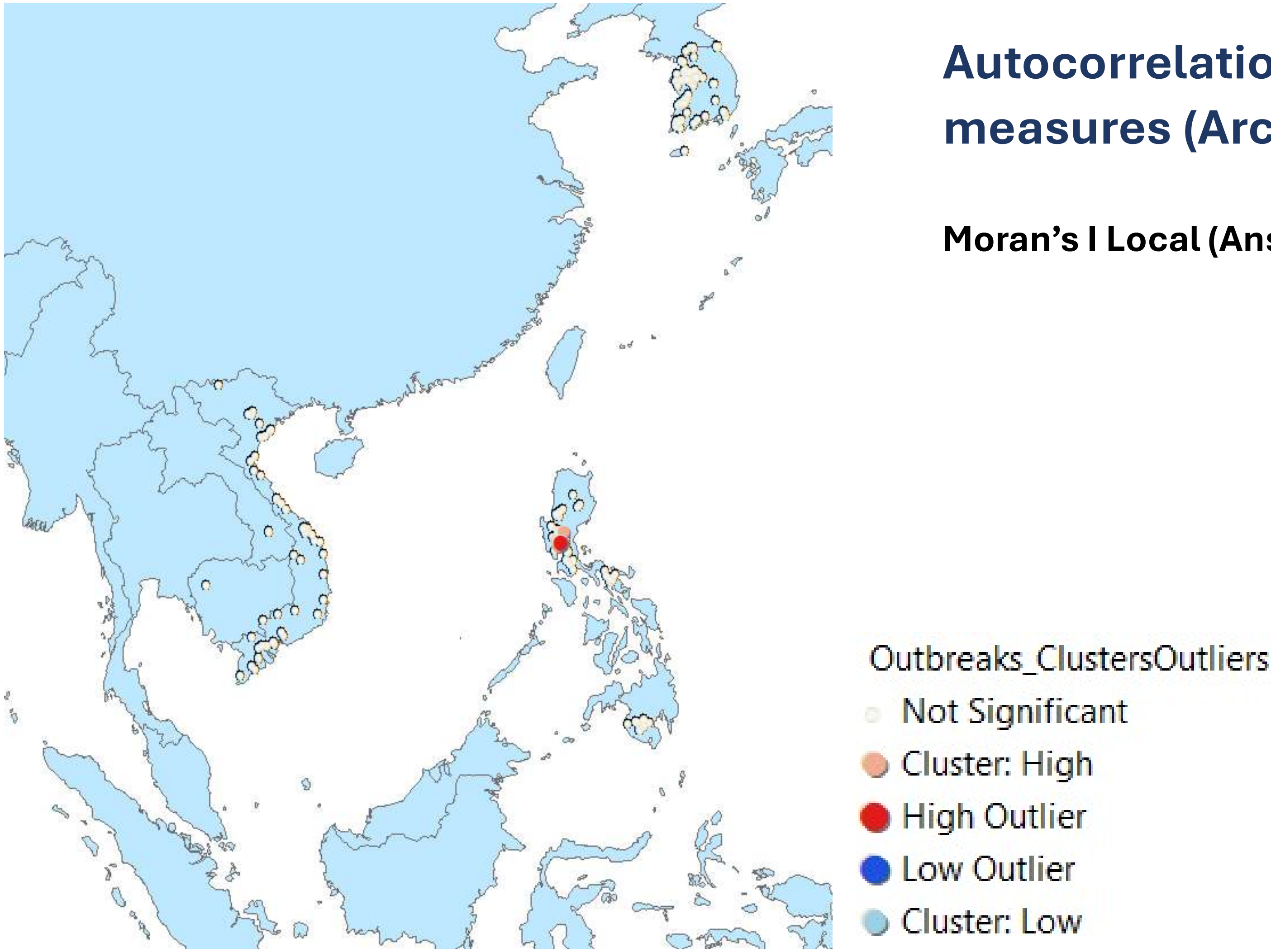
Global Moran's I Summary

Moran's Index:	0,008269
Expected Index:	-0,002538
Variance:	0,001155
z-score:	0,318018
p-value:	0,750471

Autocorrelation and pattern analysis measures (ArcGIS):

Moran's I - Global Spatial Autocorrelation

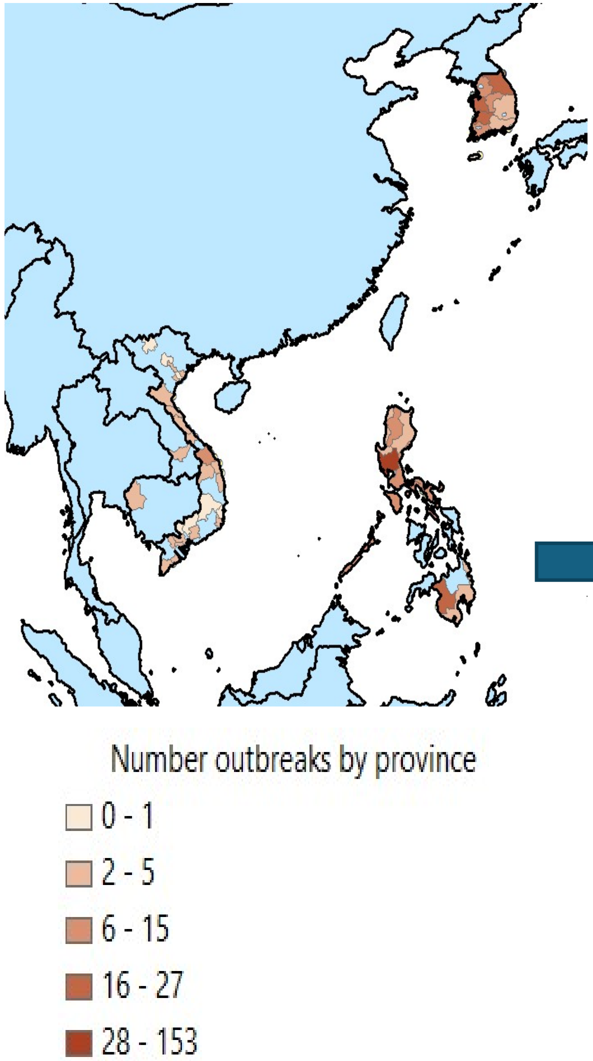
Exercise 3 (Optional): Autocorrelation and pattern analysis



Autocorrelation and pattern analysis measures (ArcGIS):

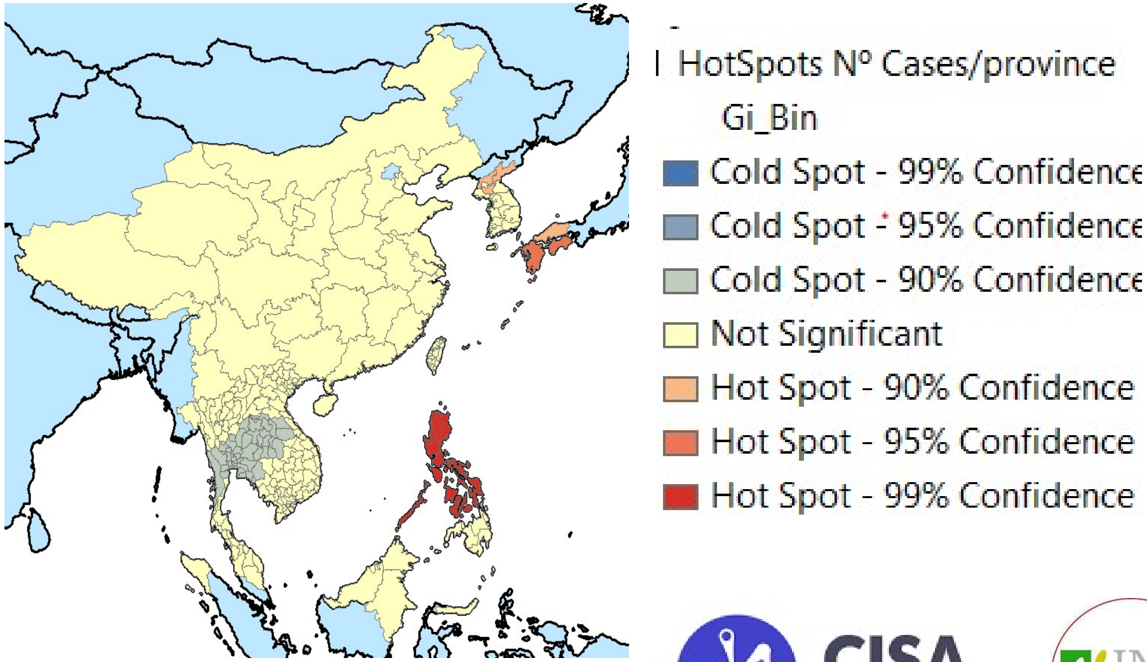
Moran's I Local (Anselin)

Exercise 3 (Optional): Autocorrelation and pattern analysis



Autocorrelation and pattern analysis measures (ArcGIS):

Hot Spot Analysis (Getis-Ord Gi*)



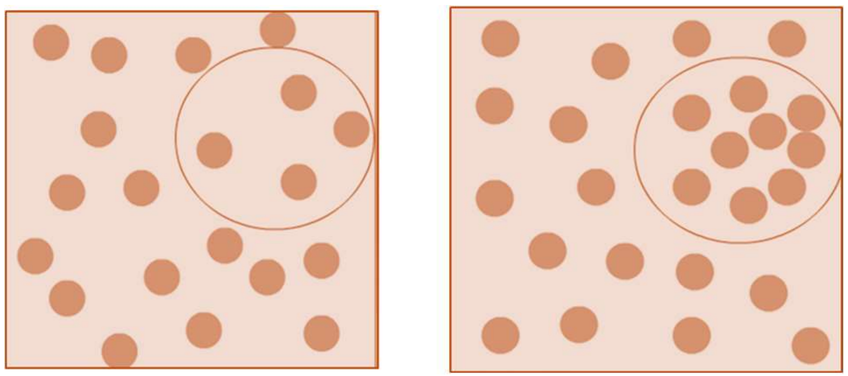
PART 4
CLUSTER ANALYSIS

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

Introduction to SaTScan

- Detects the most probable and secondary clusters
- Uses Monte Carlo simulation to generate the p-Value
- Allows the identification of spatial, temporal, and spatiotemporal clusters



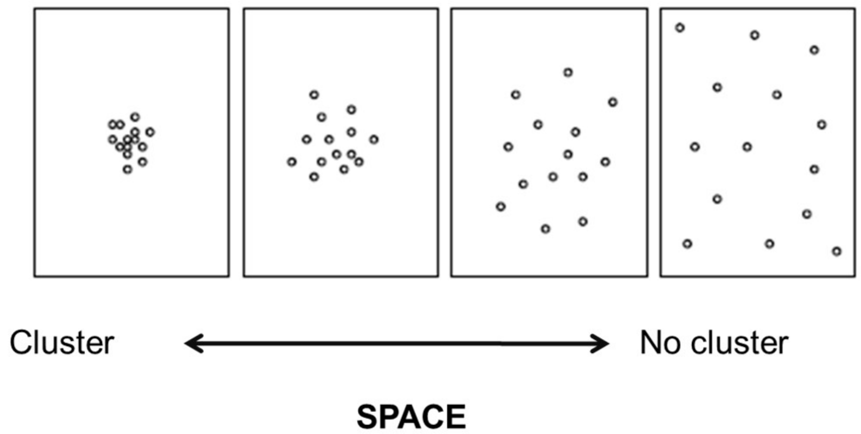
Observed = Expected
NO CLUSTER

Observed > Expected
CLUSTER

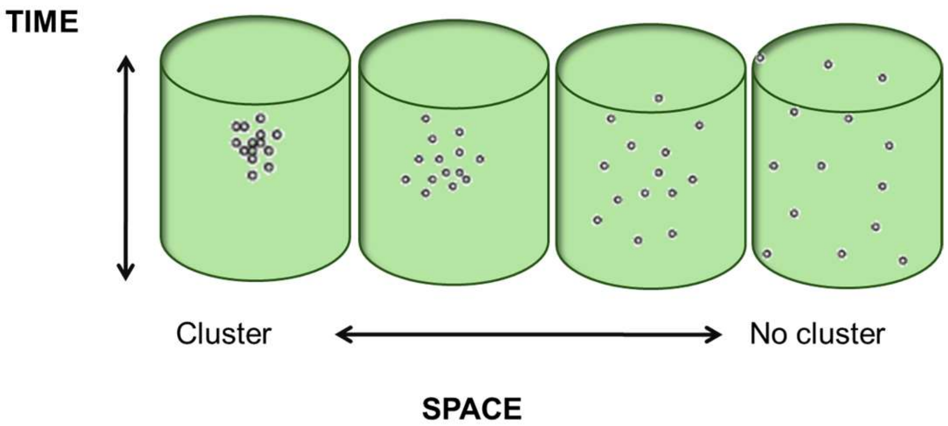
Introduction to SaTScan

SaTScan™
Software for the spatial, temporal, and space-time scan statistics

Spatial analysis



Spatiotemporal analysis



CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

Main statistical models in SaTScan

Poisson Model

- Cases / Total Population
- Requires: Case counts & population data
- Used when population-at-risk is available
- Spatial, temporal and spatiotemporal

Bernoulli Model

- Cases / Total Controls (1 = case, 0 = control)
- Requires: Case-control data
- Used in clinical & epidemiological studies
- Spatial, temporal and spatiotemporal

Permutation Model

- Cases per unit area/time
- Assumes uniform distribution
- Used when no population data is available
- Spatiotemporal

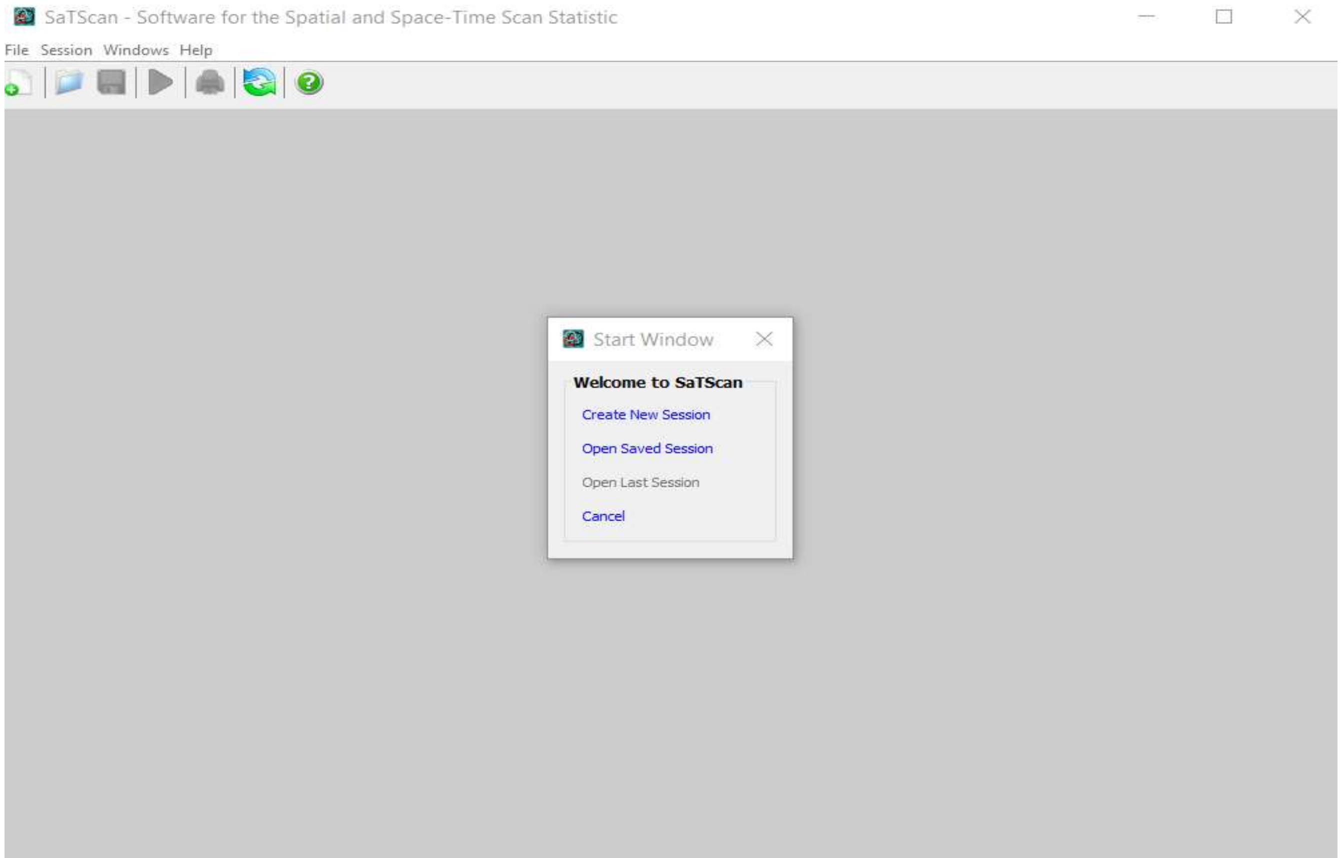
Other Models

- Normal: Continuous data (e.g., weight gain)
- Exponential: Time-to-event (e.g., disease detection)
- Ordinal: Ordered categories (e.g., mortality levels)
- Multinomial: Multiple categories without ranking

CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

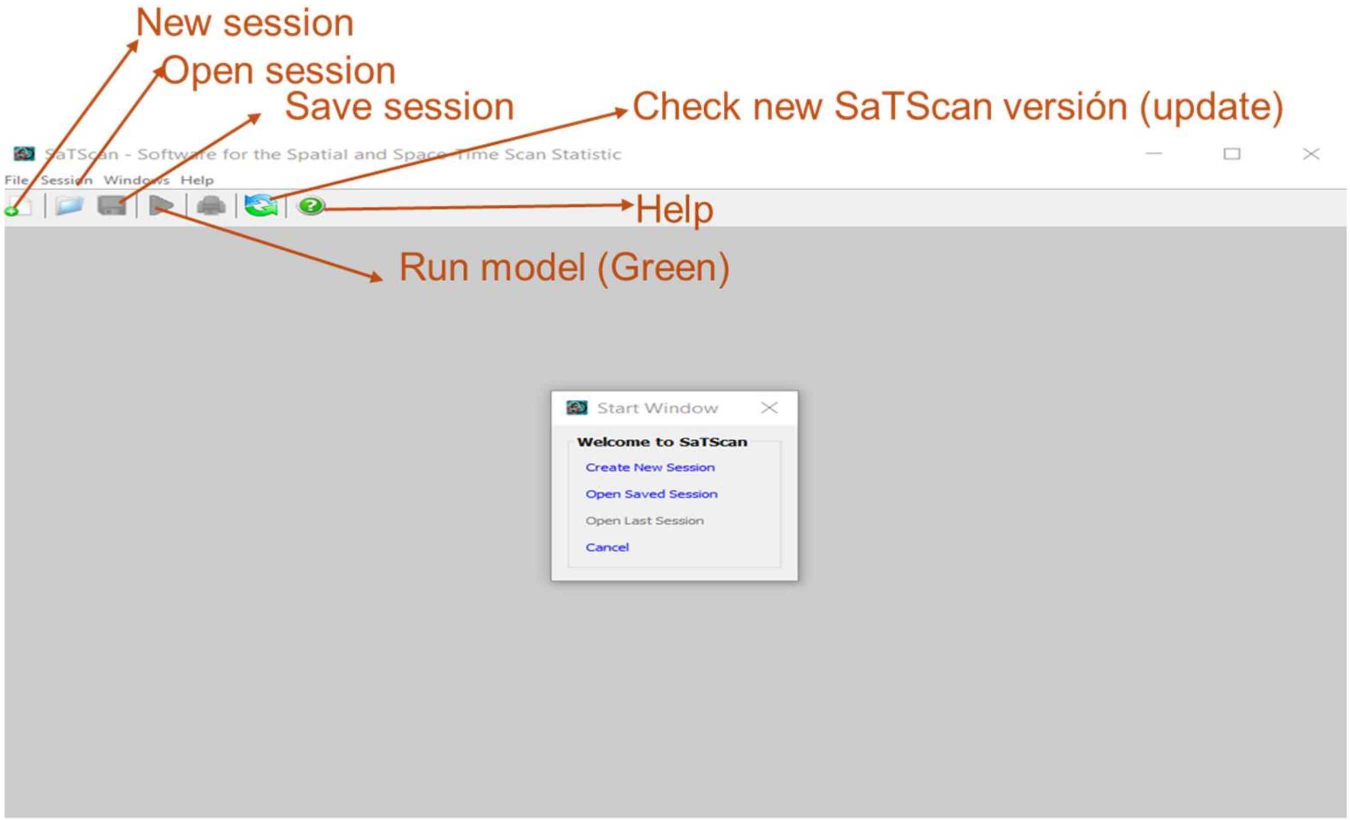
INIA CSIC

Data configuration and spatial-temporal scanning Windows



Interface

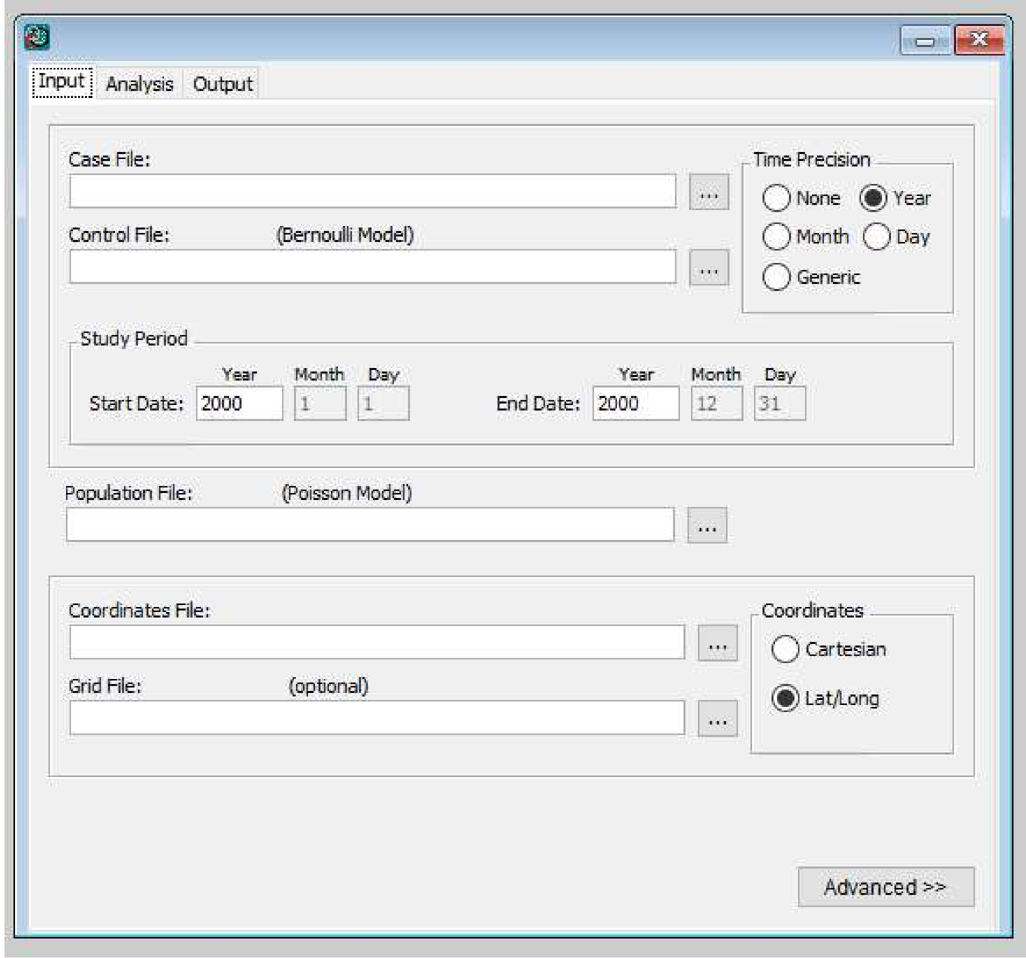
Data configuration and spatial-temporal scanning Windows



Interface

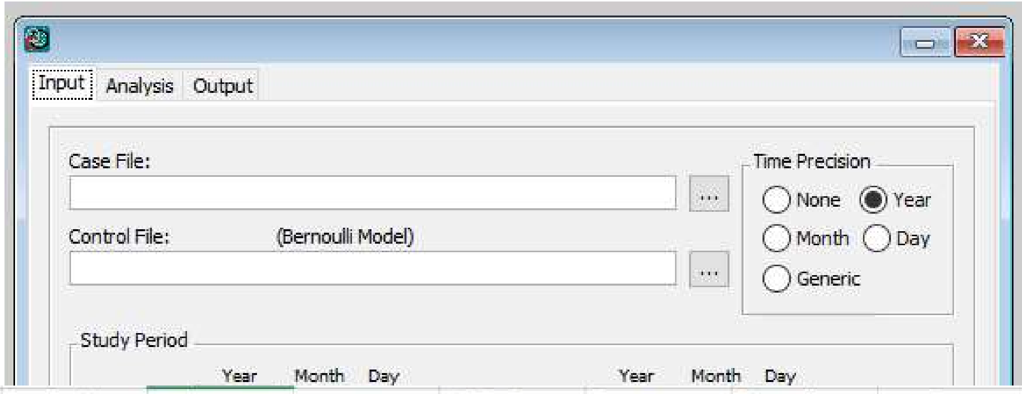
Data configuration and spatial-temporal scanning Windows

Input data



Data configuration and spatial-temporal scanning Windows

Input data

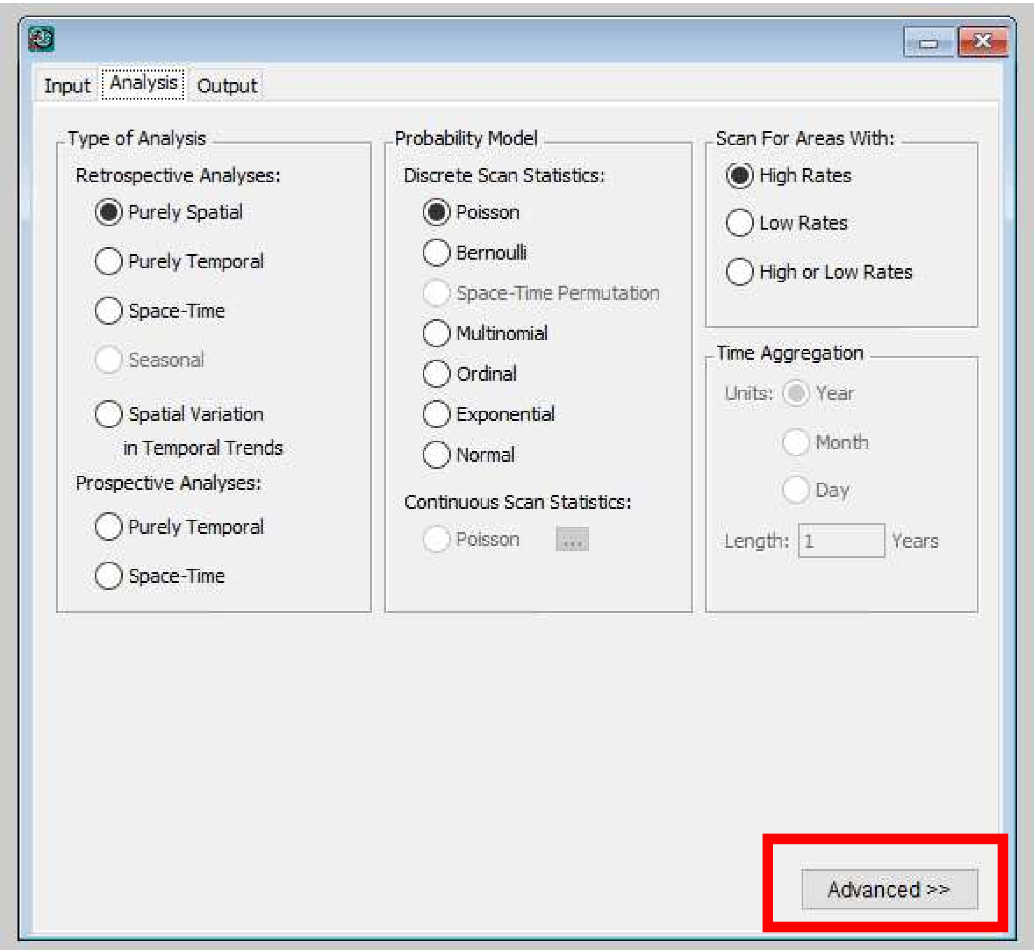


Id	source	latitude	longitude	region	country	admin1	Satscan	dom_silv	sumCases
278641	National auth	52,09	14,66	Europe	Germany	Brandenburg	2020/9/18	wild boar	6
278618	OIE	53,257791	50,084607	Europe	Russian Feder	Samarskaya C	2020/9/11	domestic	1
278617	OIE	53,2486	49,5708	Europe	Russian Feder	Samarskaya C	2020/9/11	wild boar	1
278616	OIE	53,263896	49,280197	Europe	Russian Feder	Samarskaya C	2020/9/11	wild boar	1
278615	OIE	53,376765	49,35858	Europe	Russian Feder	Samarskaya C	2020/9/11	wild	
278588	OIE	50,468611	22,744167	Europe	Poland	Lubeiskie	2020/9/15	domestic	
278587	OIE	50,668611	23,6375	Europe	Poland	Lubeiskie	2020/9/15	domestic	1
278586	OIE	50,439722	22,729722	Europe	Poland	Lubeiskie	2020/9/15	domestic	2
278585	OIE	53,736395	20,578624	Europe	Poland	Warminsko-M	2020/9/15	domestic	34
278584	OIE	50,569722	23,644722	Europe	Poland	Lubeiskie	2020/9/15	domestic	5
278583	OIE	50,4475	23,261667	Europe	Poland	Lubeiskie	2020/9/15	domestic	1
278582	OIE	50,531718	23,673204	Europe	Poland	Lubeiskie	2020/9/15	domestic	2
278581	OIE	51,530278	22,391667	Europe	Poland	Lubeiskie	2020/9/15	domestic	5
278580	OIE	50,344167	22,9875	Europe	Poland	Lubeiskie	2020/9/15	domestic	10
278579	OIE	50,447222	23,266389	Europe	Poland	Lubeiskie	2020/9/15	domestic	13

Attention! Check date format: YY/MM/DD

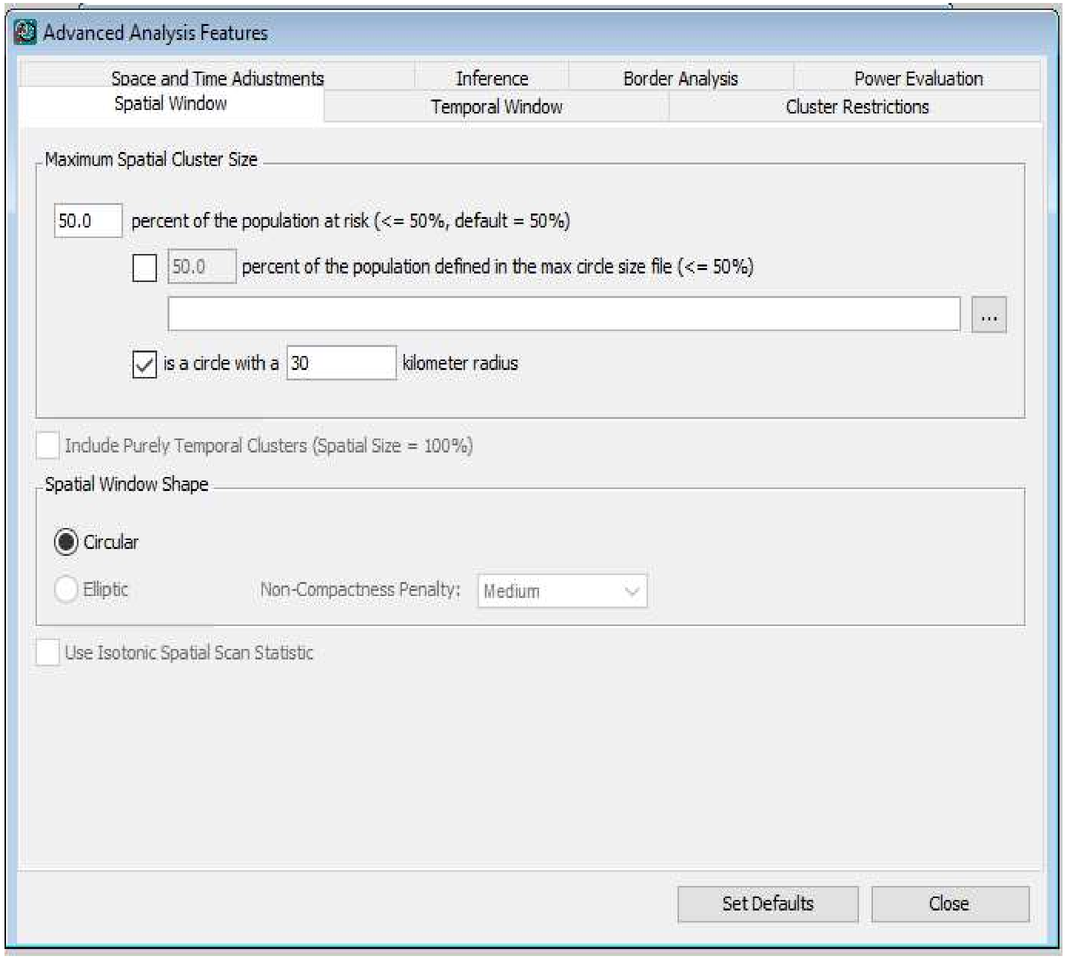
Data configuration and spatial-temporal scanning Windows

Analysis



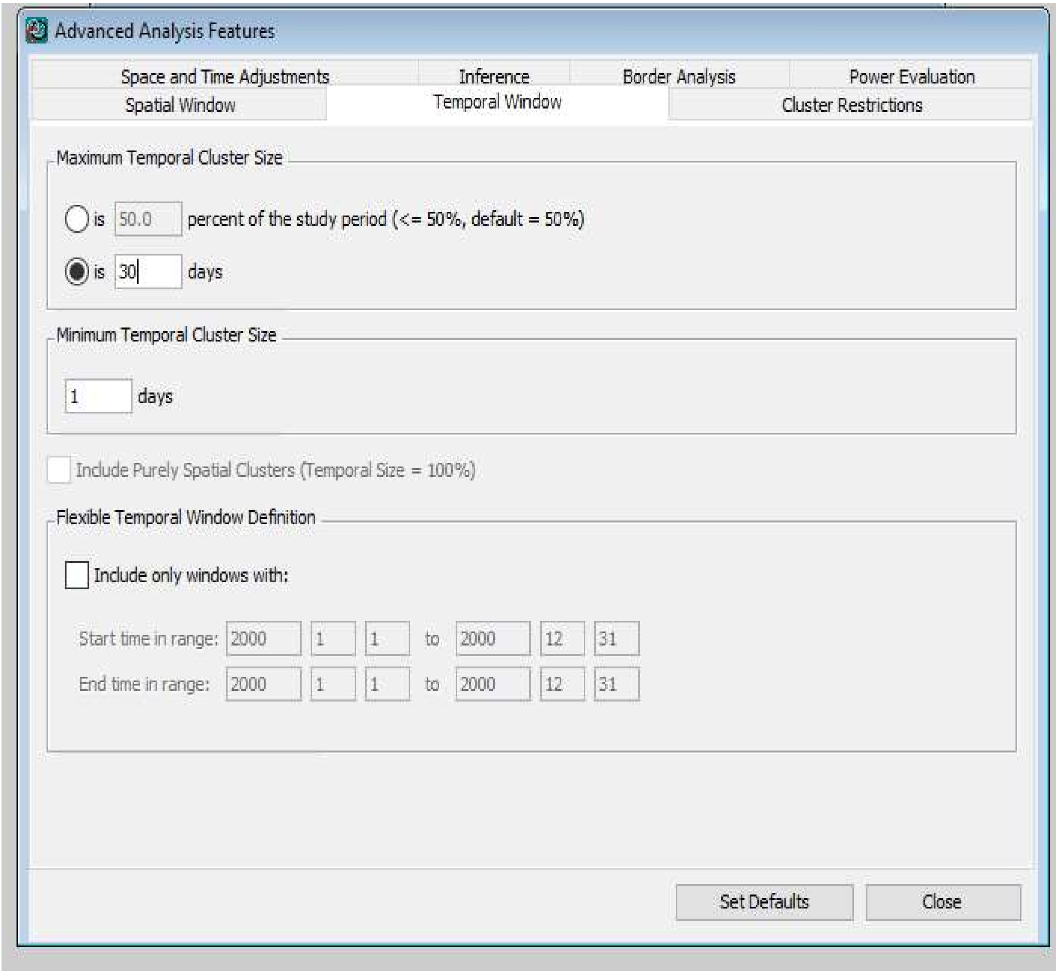
Data configuration and spatial-temporal scanning Windows

Spatial window



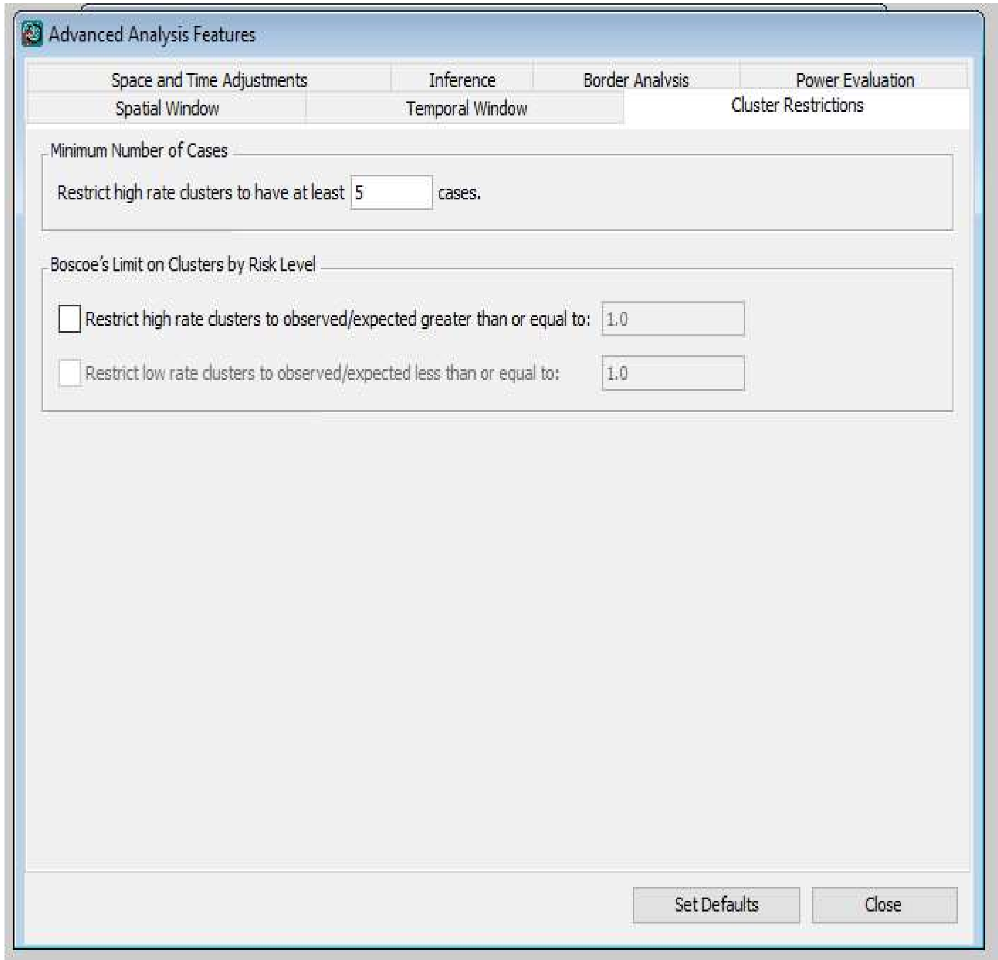
Data configuration and spatial-temporal scanning Windows

Temporal window



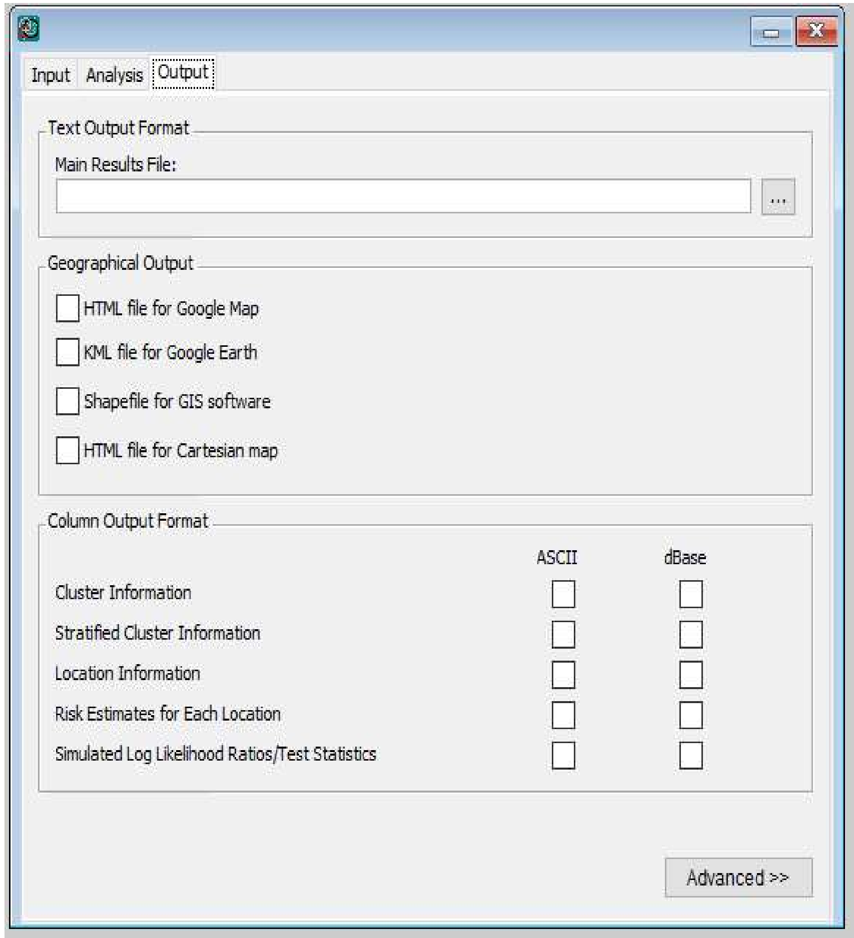
Data configuration and spatial-temporal scanning Windows

Cluster restrictions



Data configuration and spatial-temporal scanning Windows

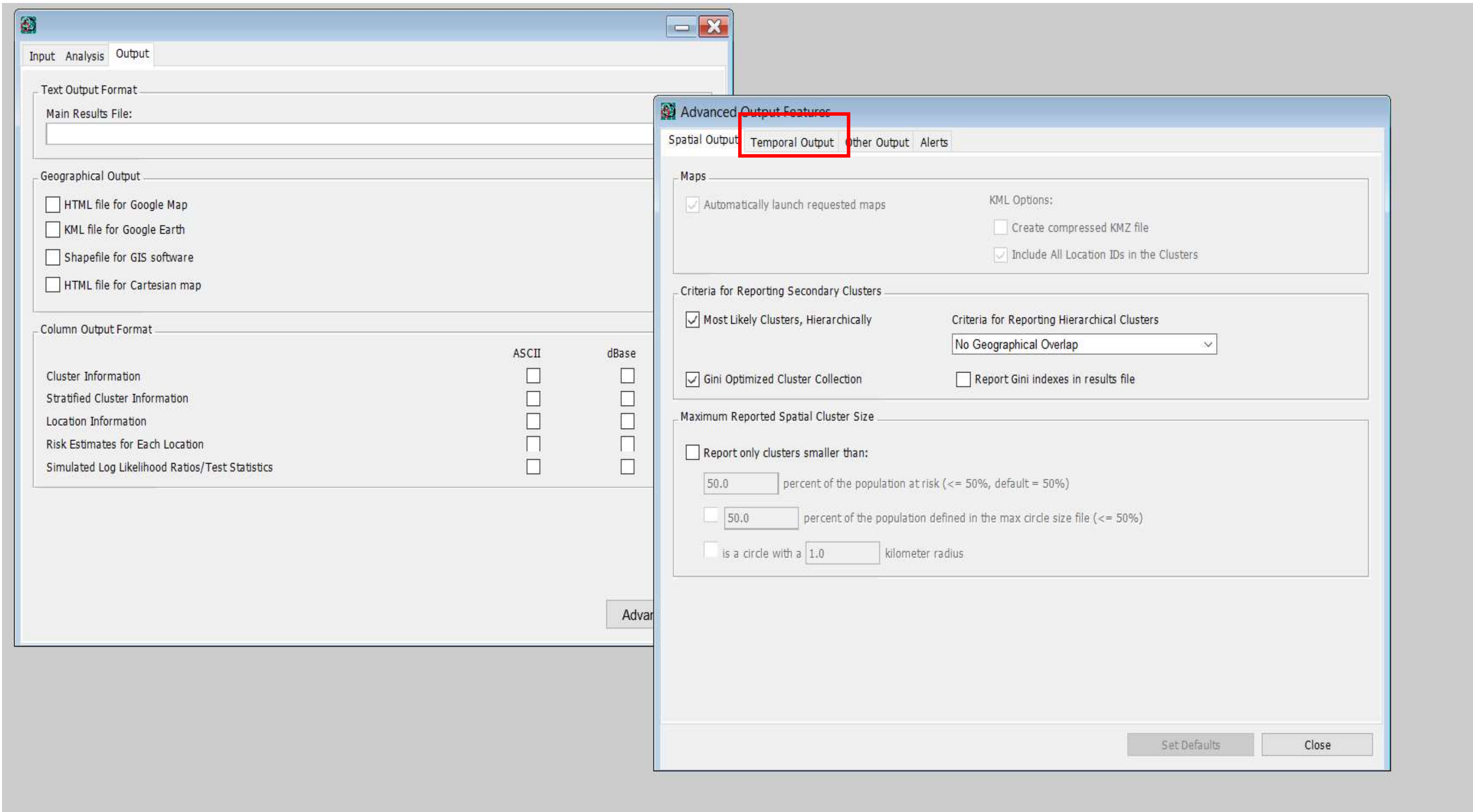
Outputs:
check all of them



Data configuration and spatial-temporal scanning Windows

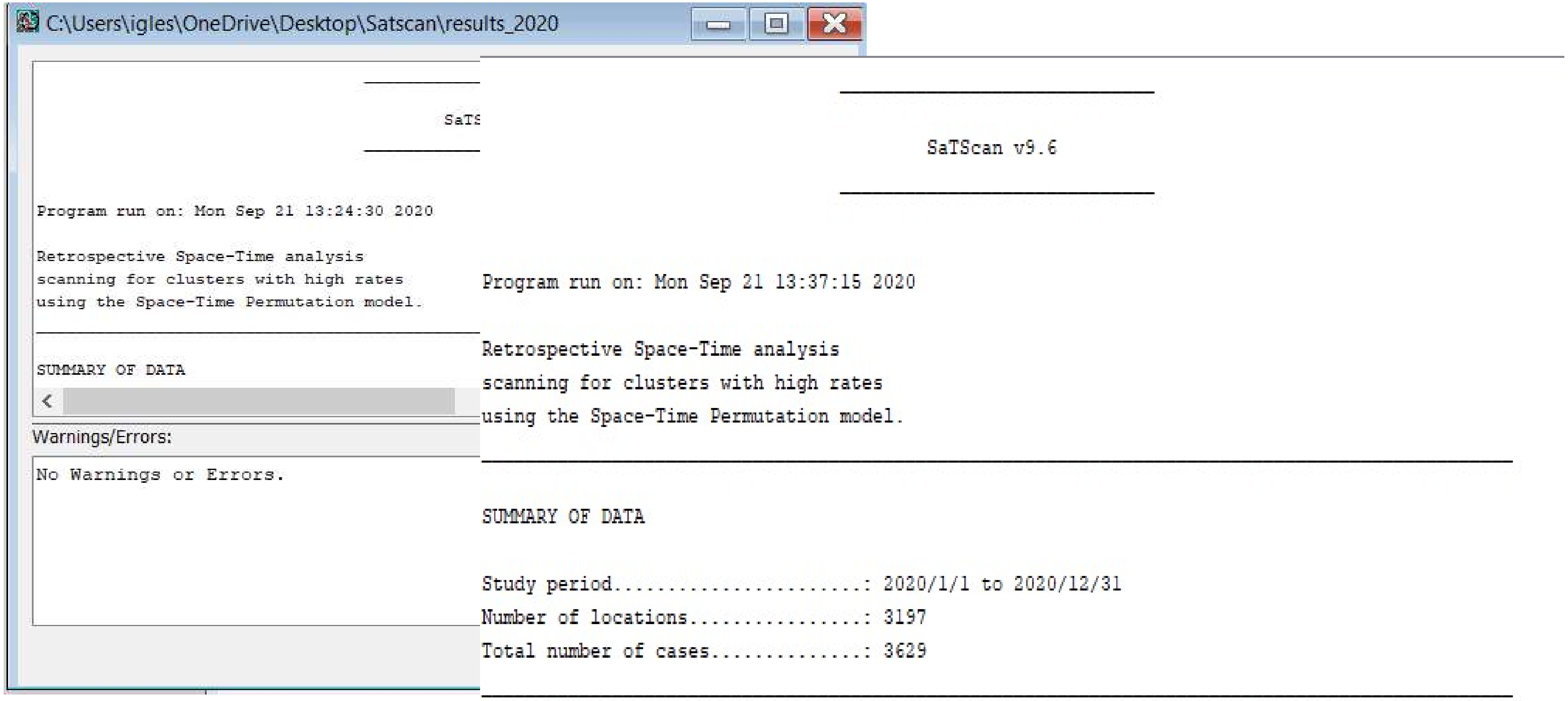
Outputs:
check all of them..

..including temporal output



Data configuration and spatial-temporal scanning Windows

Results Txt


C:\Users\igles\OneDrive\Desktop\Satscan\results_2020

SaTS

Program run on: Mon Sep 21 13:24:30 2020

Retrospective Space-Time analysis
scanning for clusters with high rates
using the Space-Time Permutation model.

SUMMARY OF DATA

Warnings/Errors:

No Warnings or Errors.

SaTScan v9.6

Program run on: Mon Sep 21 13:37:15 2020

Retrospective Space-Time analysis
scanning for clusters with high rates
using the Space-Time Permutation model.

SUMMARY OF DATA

Study period.......................: 2020/1/1 to 2020/12/31
Number of locations................: 3197
Total number of cases..............: 3629

Data configuration and spatial-temporal scanning Windows

Results Txt

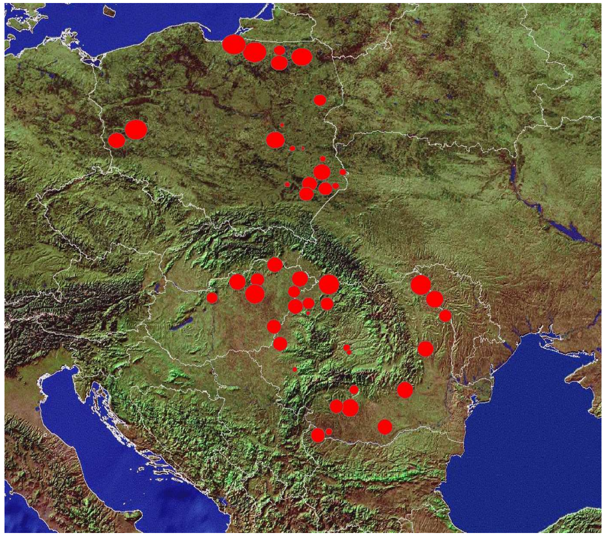
```
PARAMETER SETTINGS

Input
-----
  Case File        : C:\Users\igles\OneDrive\Desktop\Satscan\Cases_2020
  Time Precision   : Day
  Start Time       : 2020/1/1
  End Time         : 2020/12/31
  Coordinates File : C:\Users\igles\OneDrive\Desktop\Satscan\Coordinates_2020
  Coordinates      : Latitude/Longitude

Analysis
--------
  Type of Analysis        : Retrospective Space-Time
  Probability Model       : Space-Time Permutation
  Scan for Areas with     : High Rates
  Time Aggregation Units  : Day
  Time Aggregation Length : 7

Output
------
  Main Results File       : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020
  Google Earth File       : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.kml
  Google Maps File        : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.clustermap.html
  Shapefile               : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.clustermap.col.shp
  Cartesian Graph File    : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.cluster.html
  Cluster File            : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.cluster.col.dbf
  Stratified Cluster File : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.cluster.sci.dbf
  Location File           : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.cluster.gis.dbf
  Simulated LLRs File     : C:\Users\igles\OneDrive\Desktop\Satscan\results_2020.cluster.llr.dbf
```

```
Spatial Window
--------------
  Maximum Spatial Cluster Size : 50 percent of population at risk
  Maximum Spatial Cluster Size : 30 km
  Window Shape                 : Circular

Temporal Window
---------------
  Minimum Temporal Cluster Size : 1 Day
  Maximum Temporal Cluster Size : 30 Days

Cluster Restrictions
--------------------
  Minimum Cases in Cluster for High Rates : 5
  Restrict High Rate Clusters             : No

Space And Time Adjustments
--------------------------
  Adjust for Weekly Trends, Nonparametric : No

Inference
---------
  P-Value Reporting                  : Default Combination
  Number of Replications             : 999
  Adjusting for More Likely Clusters : No
```

Data configuration and spatial-temporal scanning Windows

Results:

KML

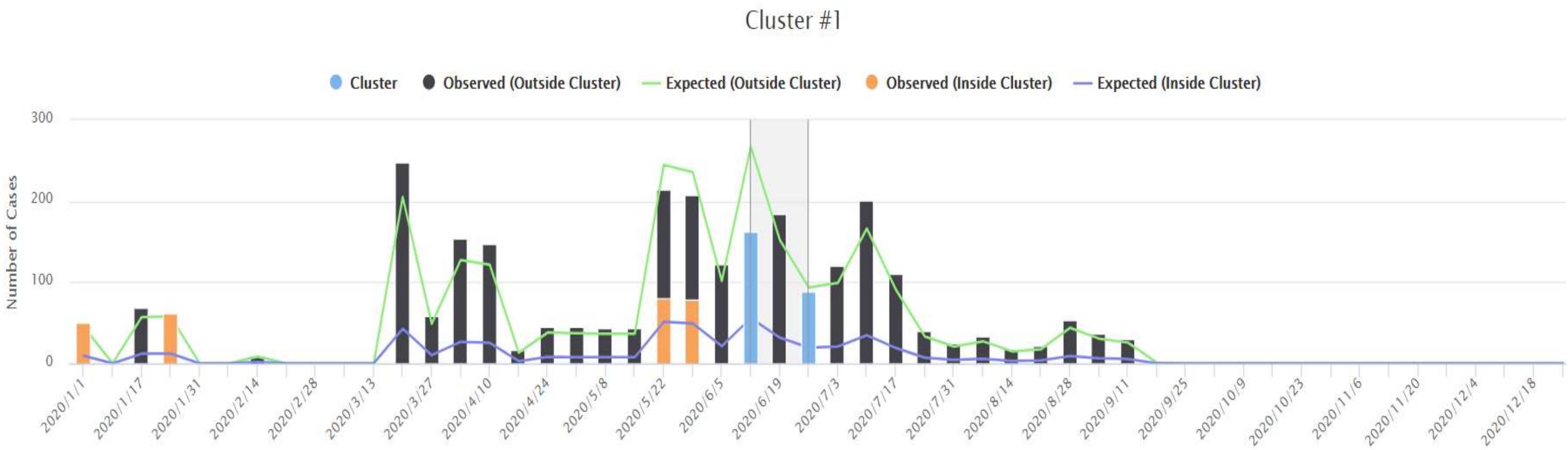


Shapefile

- Number of Clusters
- Duration
- Time of occurrence
- Radius
- Number of Cases per Cluster
- Observed/Expected Ratio
- P-Value

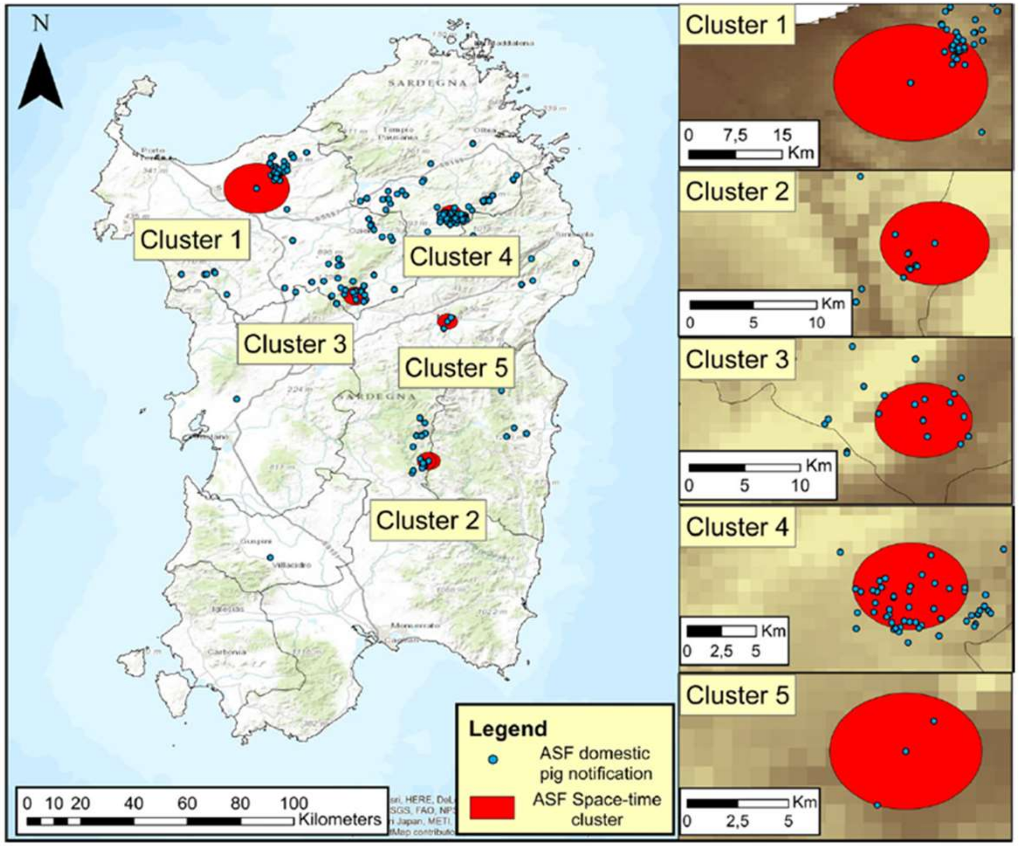
Data configuration and spatial-temporal scanning Windows

Temporal Results
Html



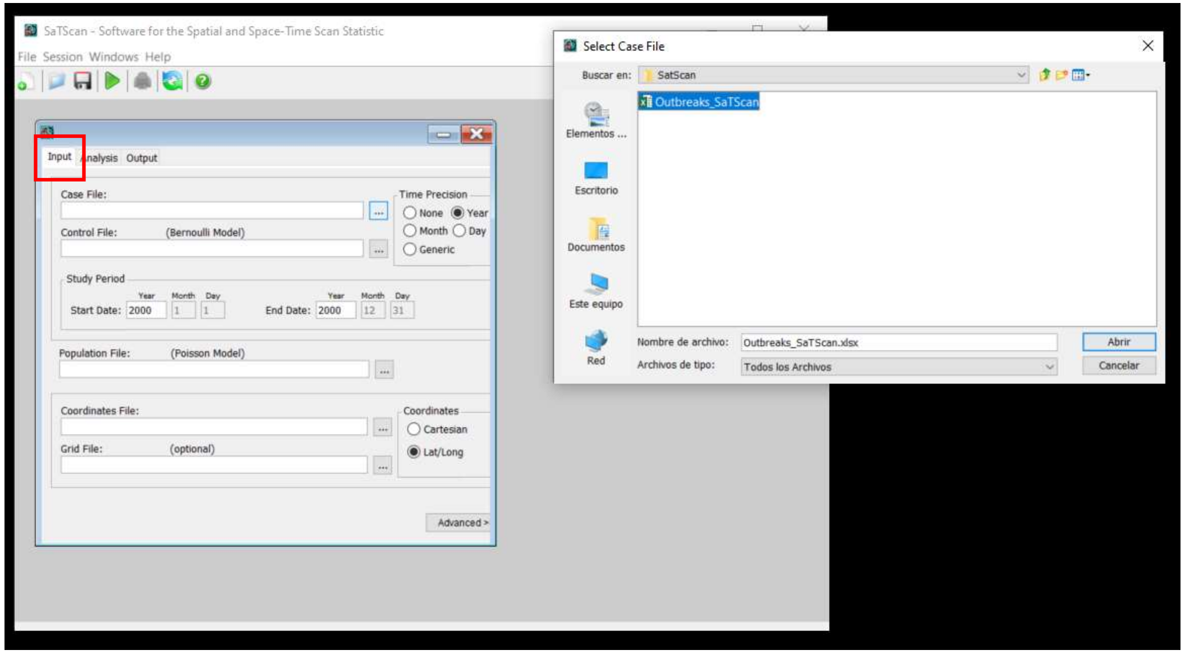
Examples of use in animal health disease analysis

Space-time clusters of ASF domestic pigs notifications



Space–time clusters of domestic pig African swine fever (ASF) notifications detected with the space–time permutation statistic (Kulldorff et al., 2005) in Sardinia (Italy) Significant clusters have a P-value <0.05. Source: Iglesias, I., Rodríguez, A., Feliziani, F., Rolesu, S., & De la Torre, A. (2017). Spatio-temporal analysis of African Swine Fever in Sardinia (2012–2014): trends in domestic pigs and wild boar. Transboundary and emerging diseases, 64(2), 656-662.

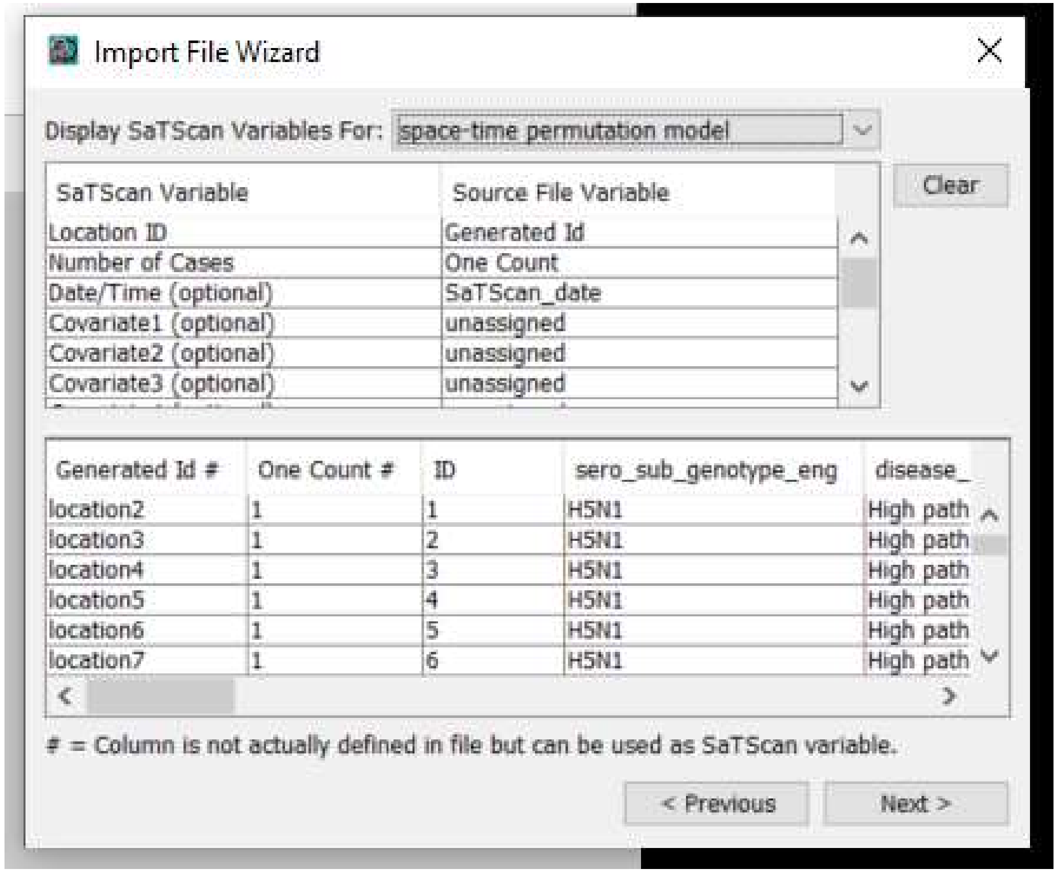
Exercise:Conducting a Space-Time Permutation Cluster Analysis of ASF Outbreaks in SaTScan



Import case file from excel

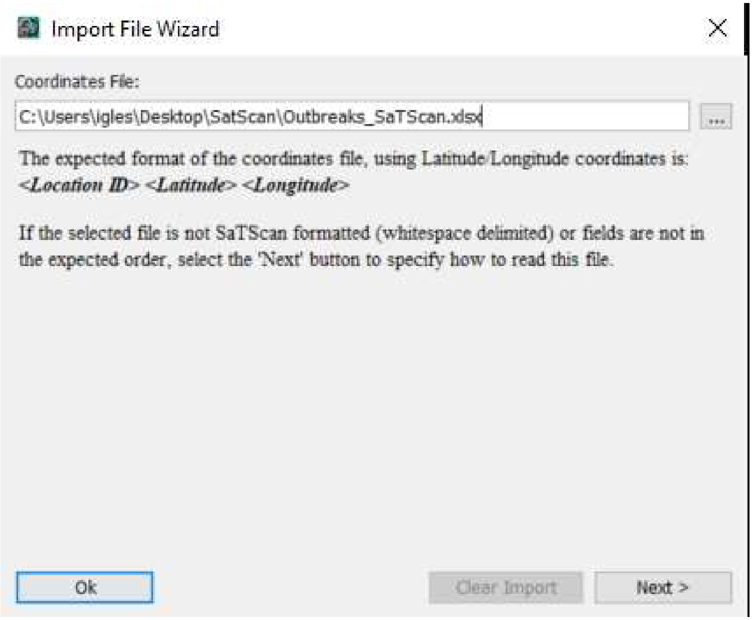
SaTScan Folder 📥 Download Link: https://saco.csic.es/s/bcYcx8oAqoymPyd
✅ All exercise steps and results are included in this folder.

Exercise

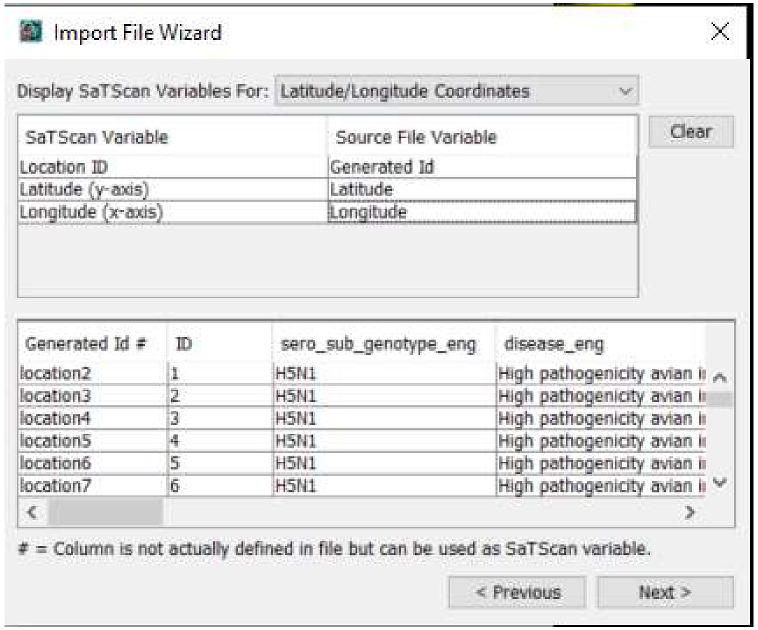


- **Import case file from excel Outbreaks SaTScan**
- **Model space-time permutation**

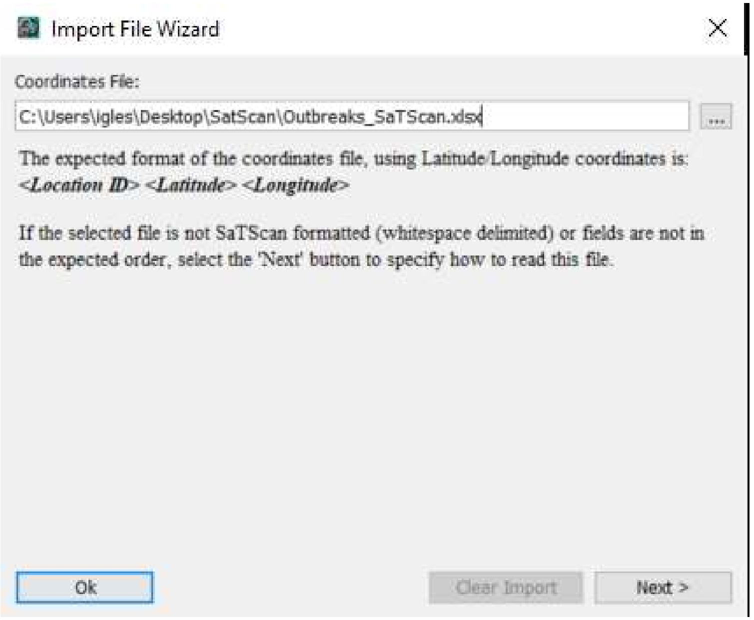
Exercise



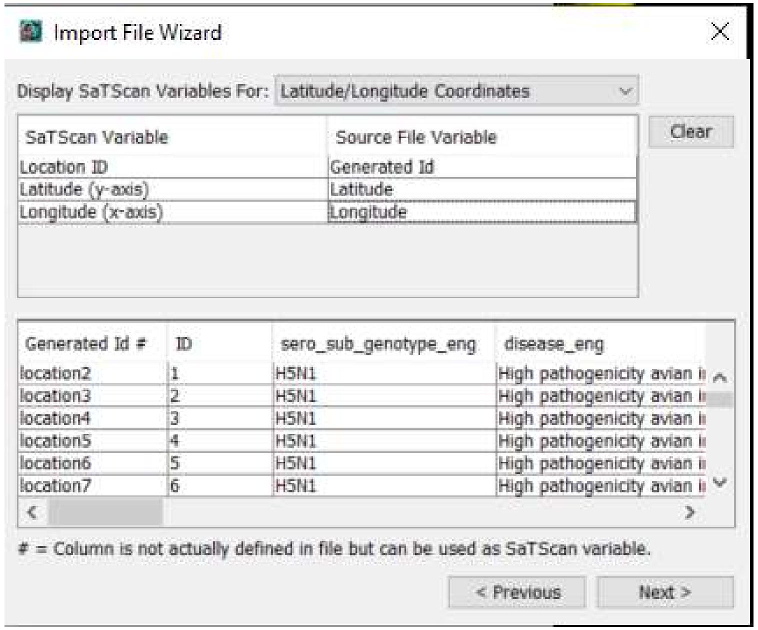
- **Import coordinates file from excel Outbreaks SaTScan**
- **Model space-time permutation**



Exercise



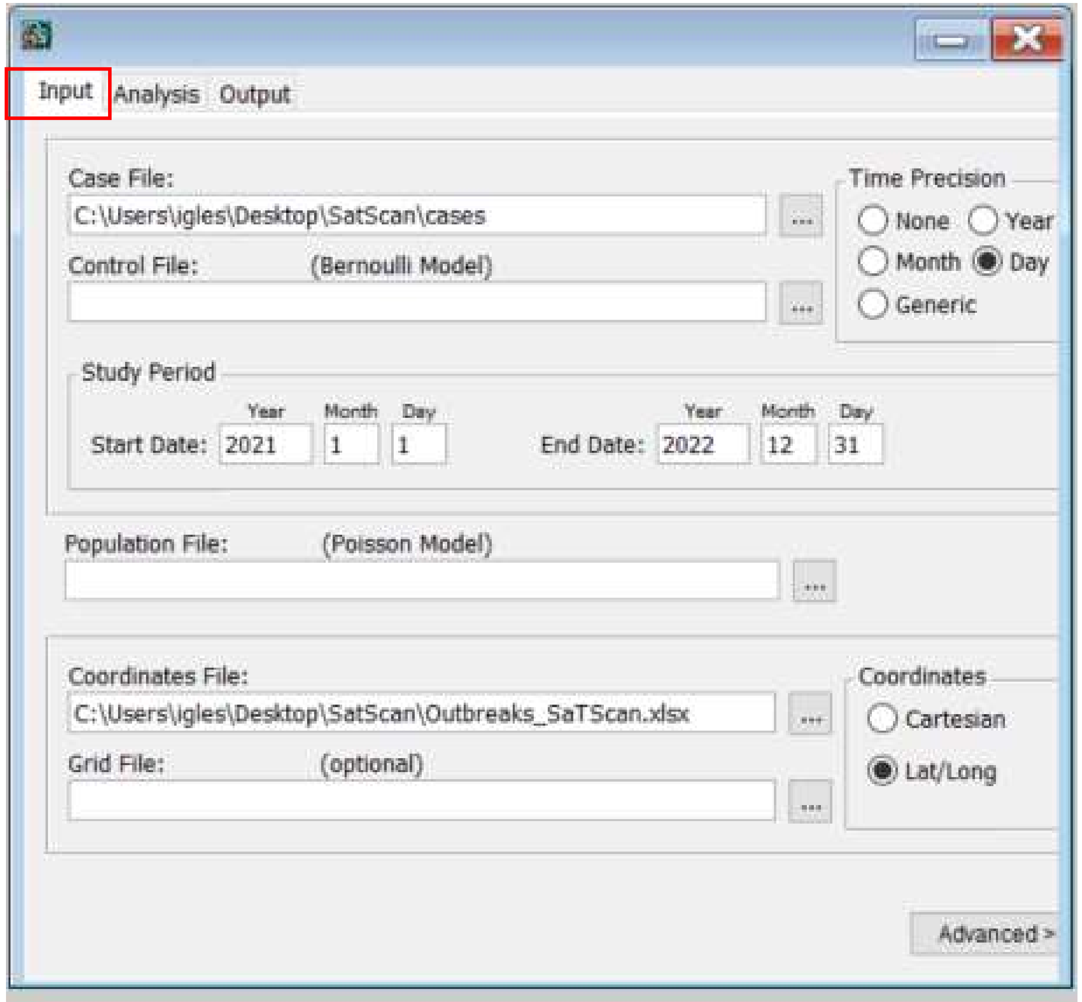
- **Import coordinates file from excel Outbreaks SaTScan**
- **Model space-time permutation**



CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTER

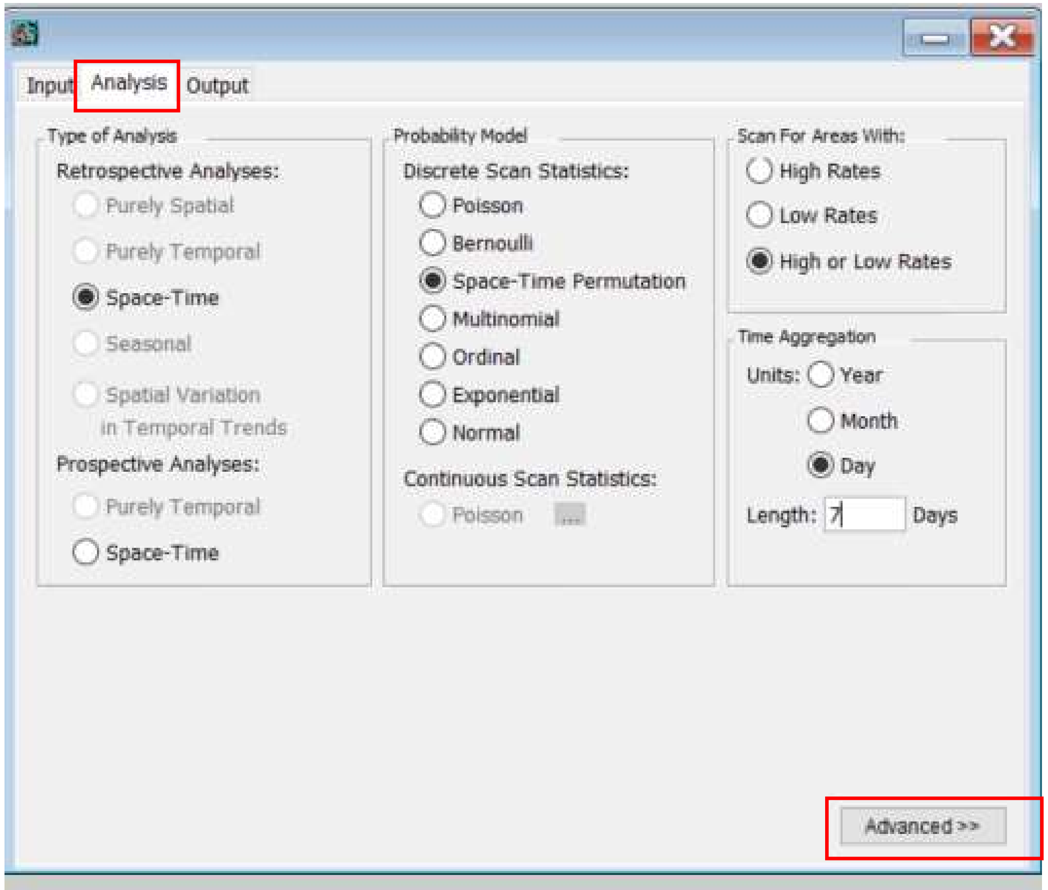
INIA CSIC

Exercise



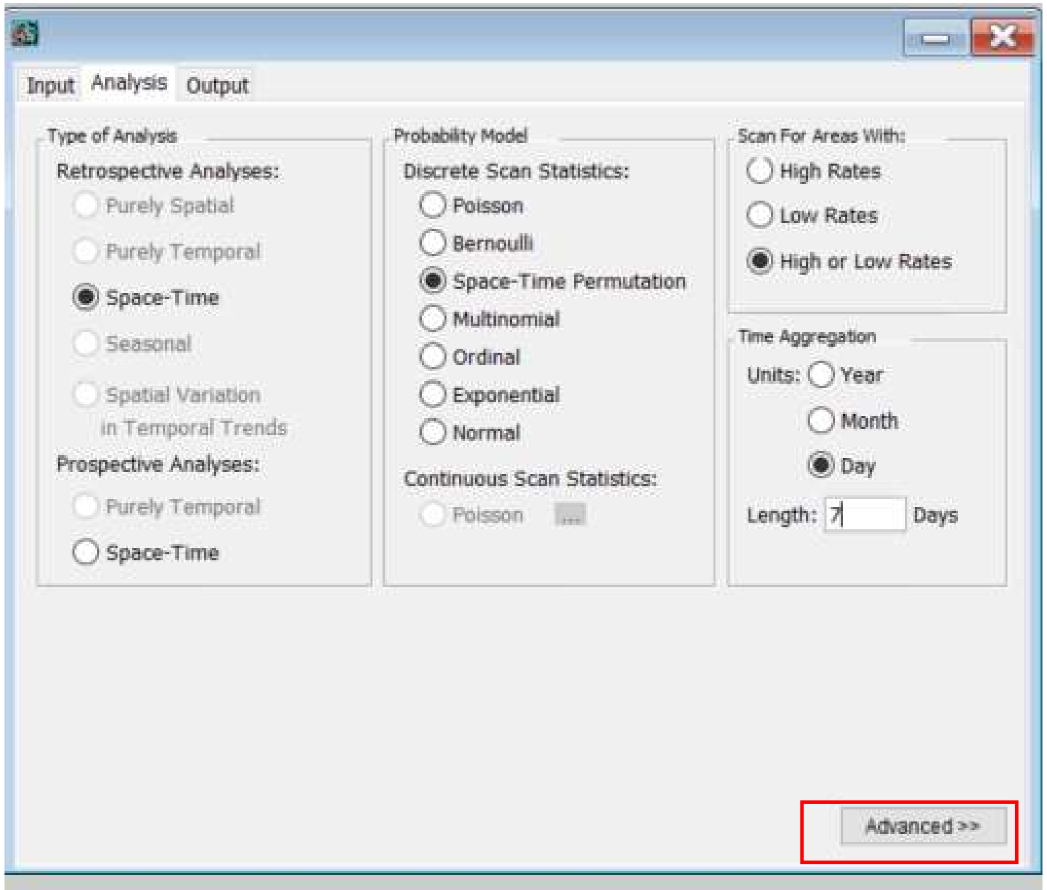
- **Check study period (2021-2022) and time precision**

Exercise



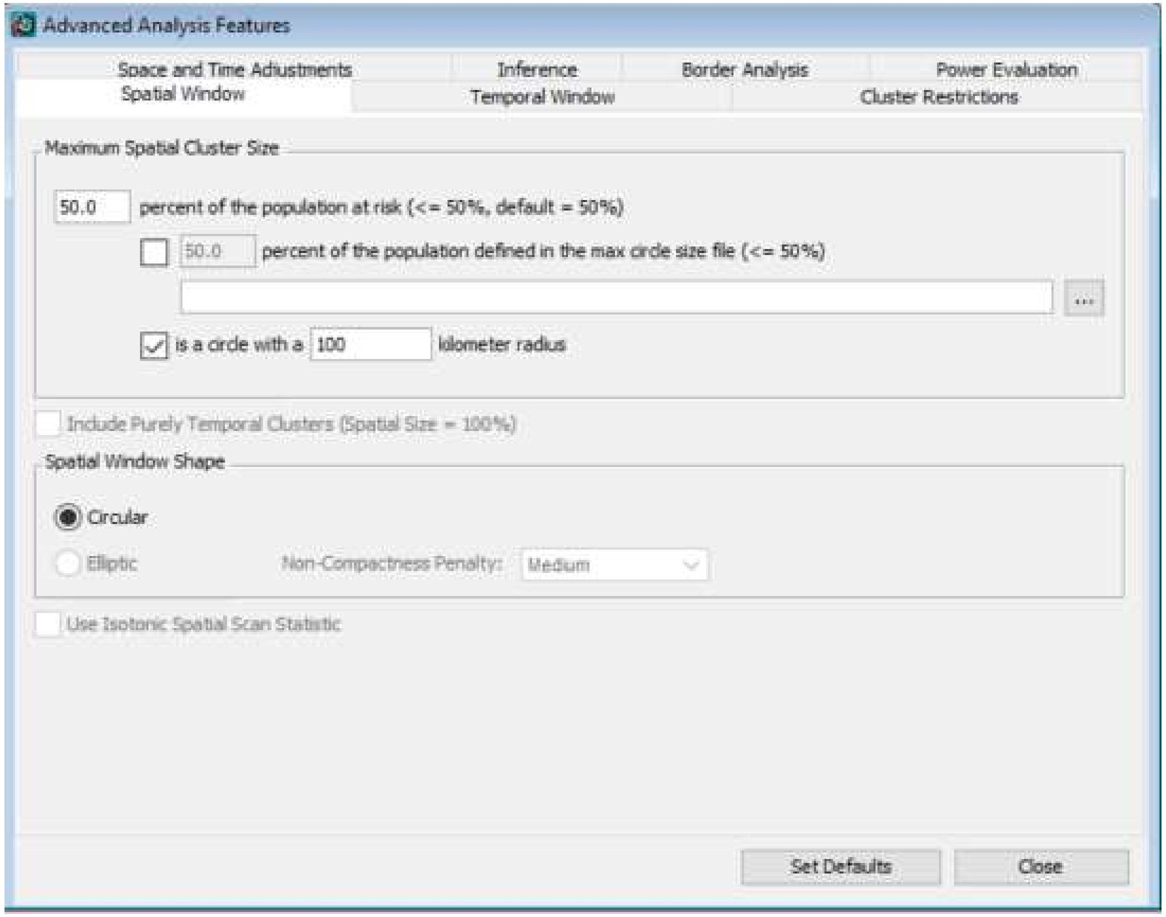
- **Analysis:**
  - **Space-time**
  - **Permutation**
  - **Scan High or Low rates**
  - **Time aggregation 7 days**

Exercise



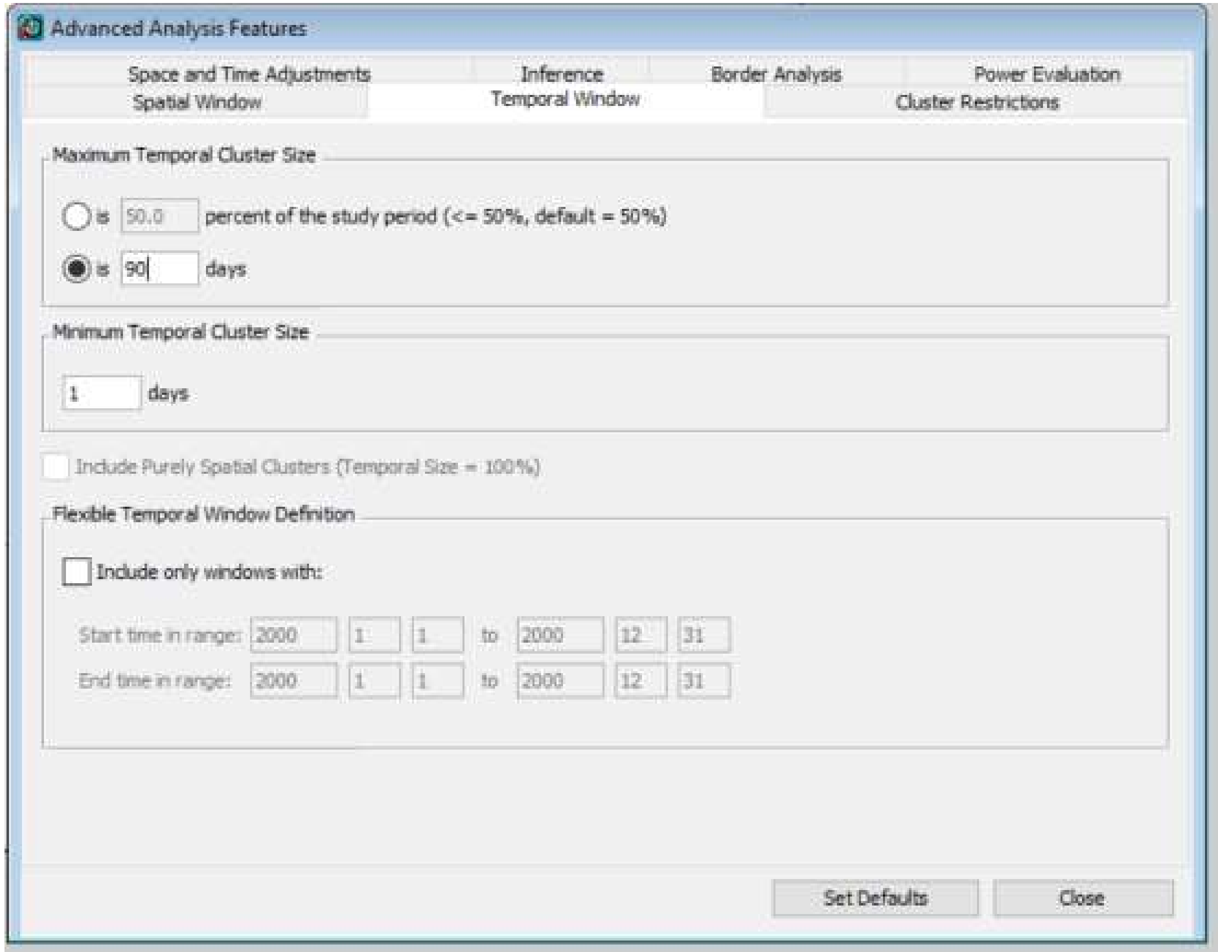
- **Analysis:Advanced**

Exercise



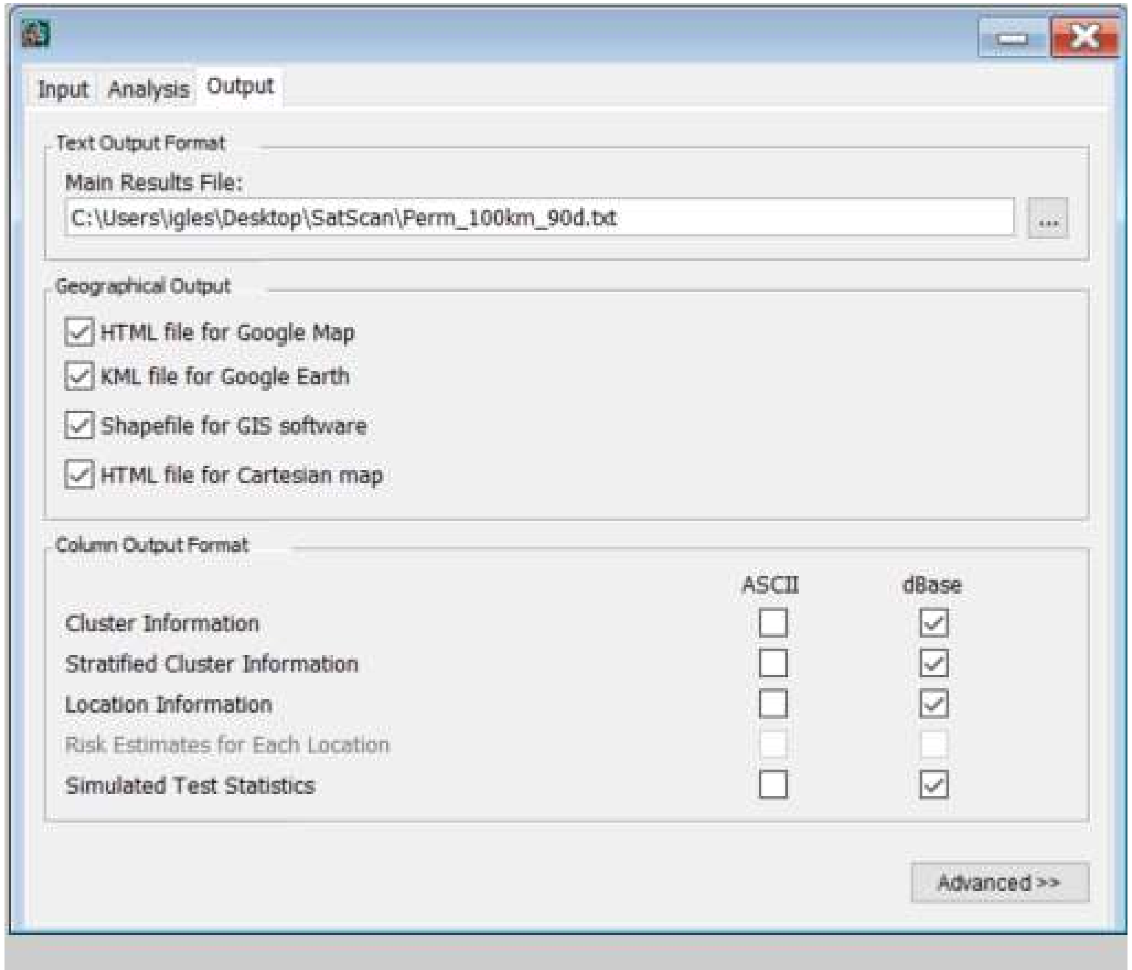
- **Analysis:Advanced**
- **Spatial window 100km**

Exercise



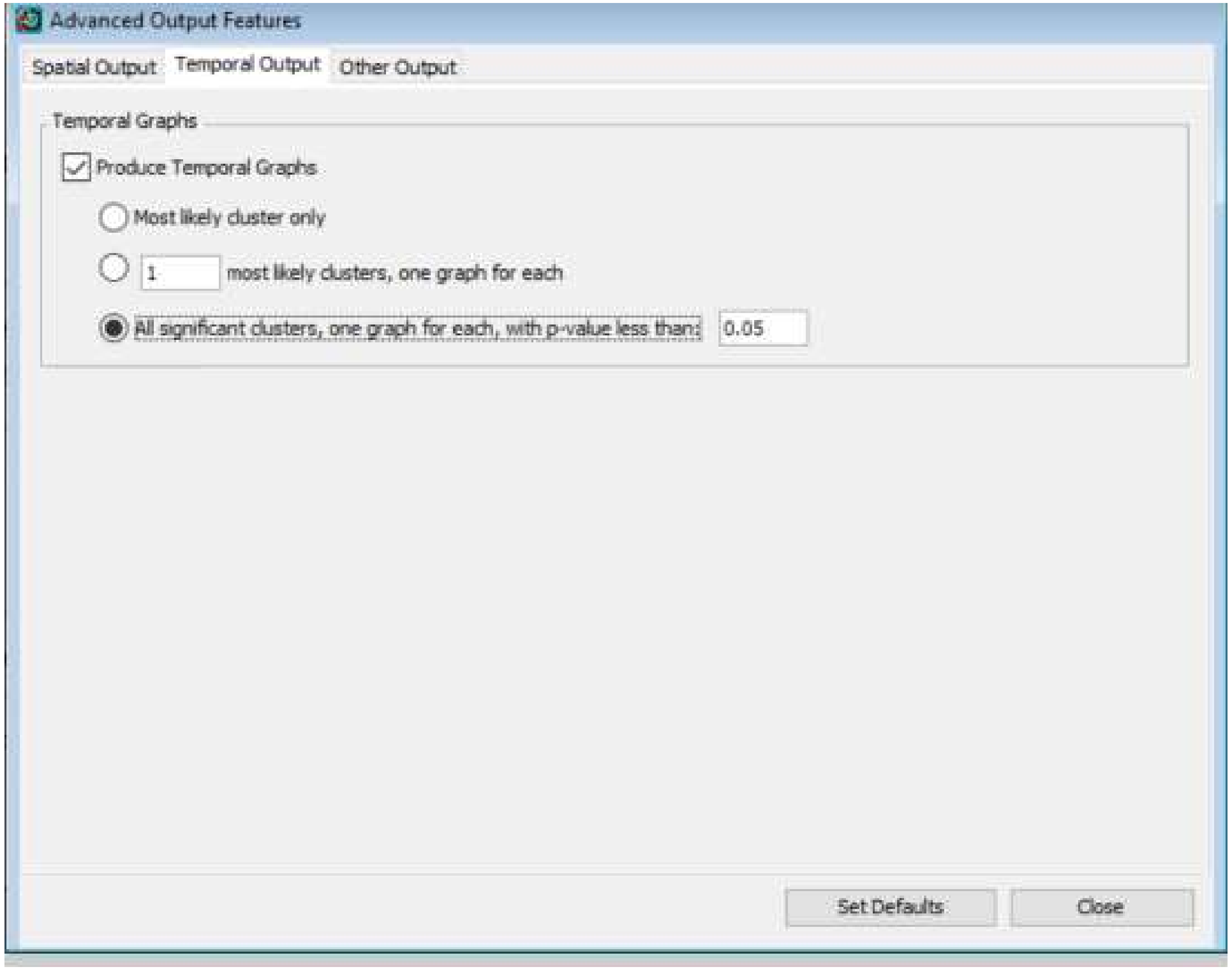
- **Analysis:Advanced**
- **Temporal window 90 days**

Exercise



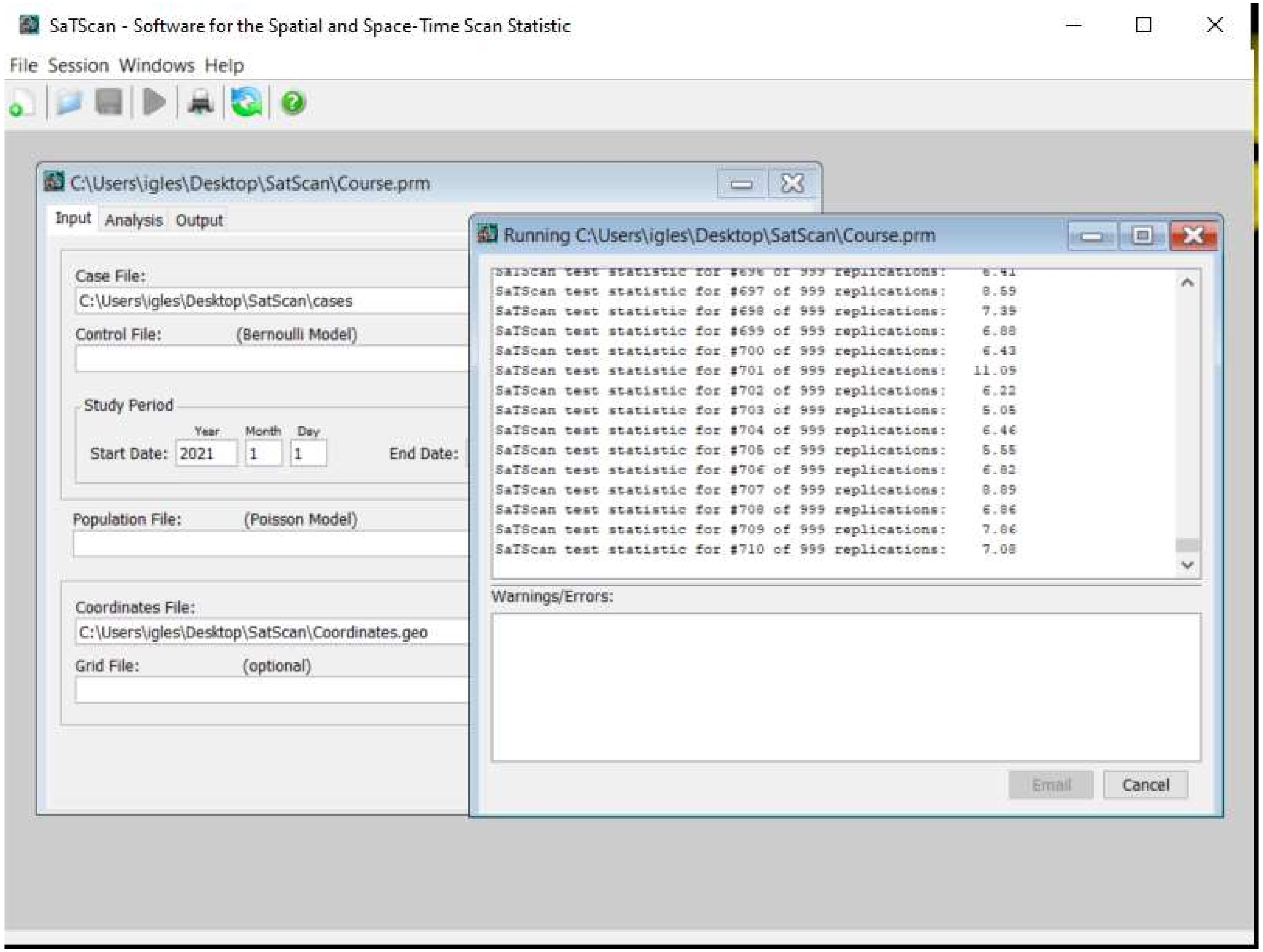
Outputs:
check all outputs in dBase format

Exercise



Outputs: Advanced Temporal all significant clusters

Exercise

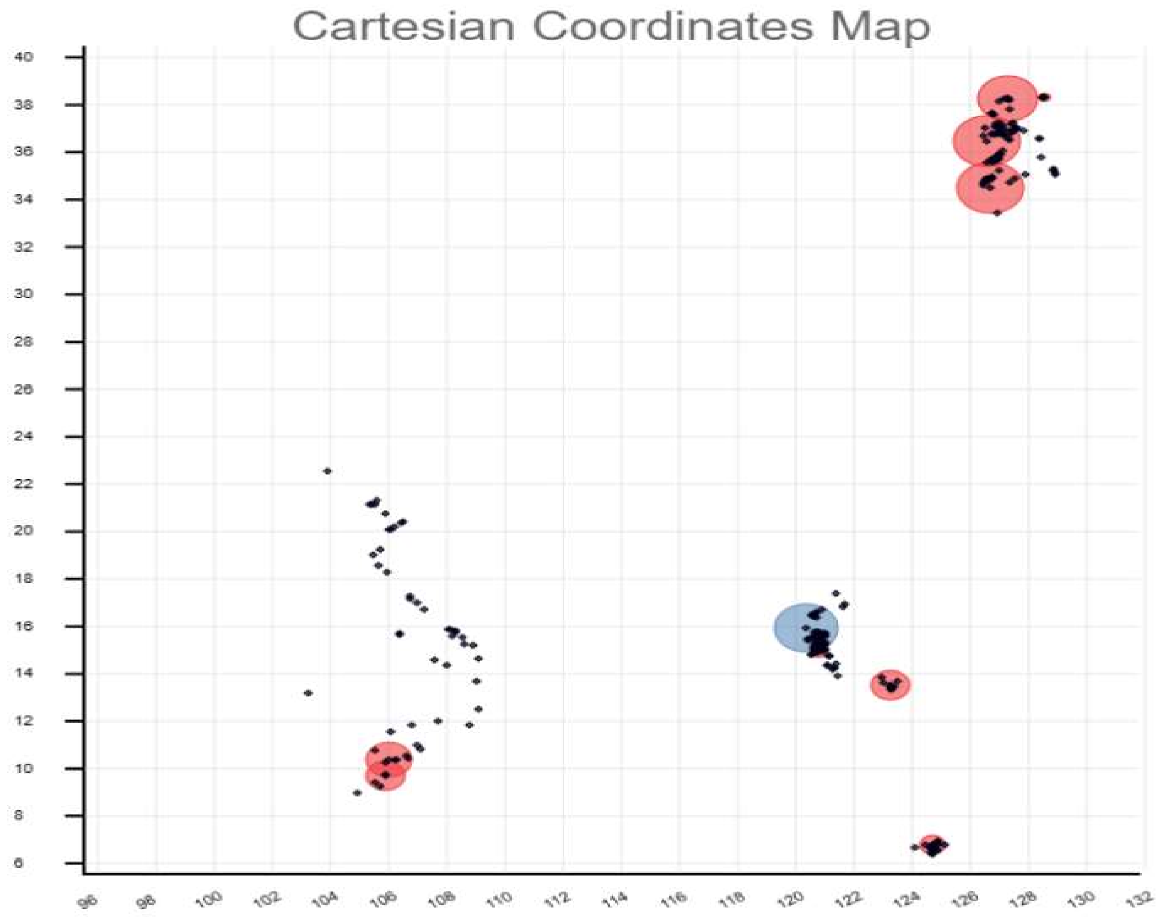


Run model

Exercise

Html map will be open when the model finish.

Exercise



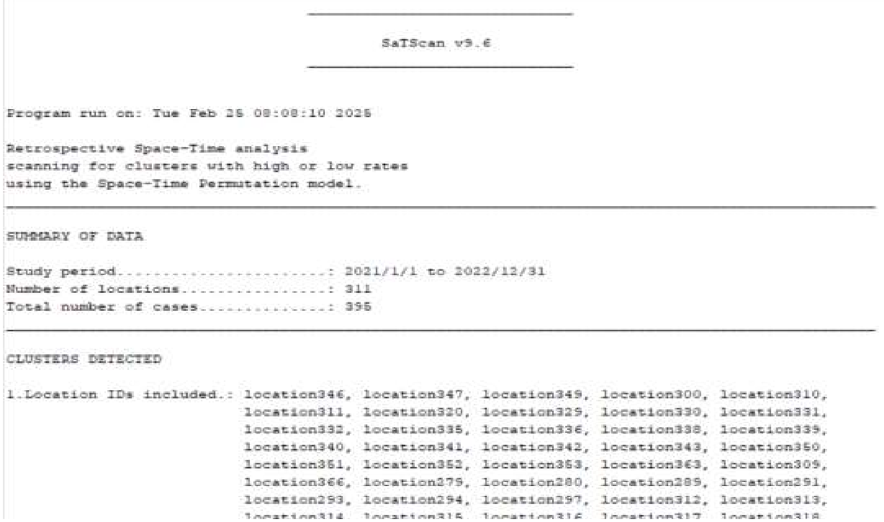
Html map will be open when the model finish.

Exercise Results

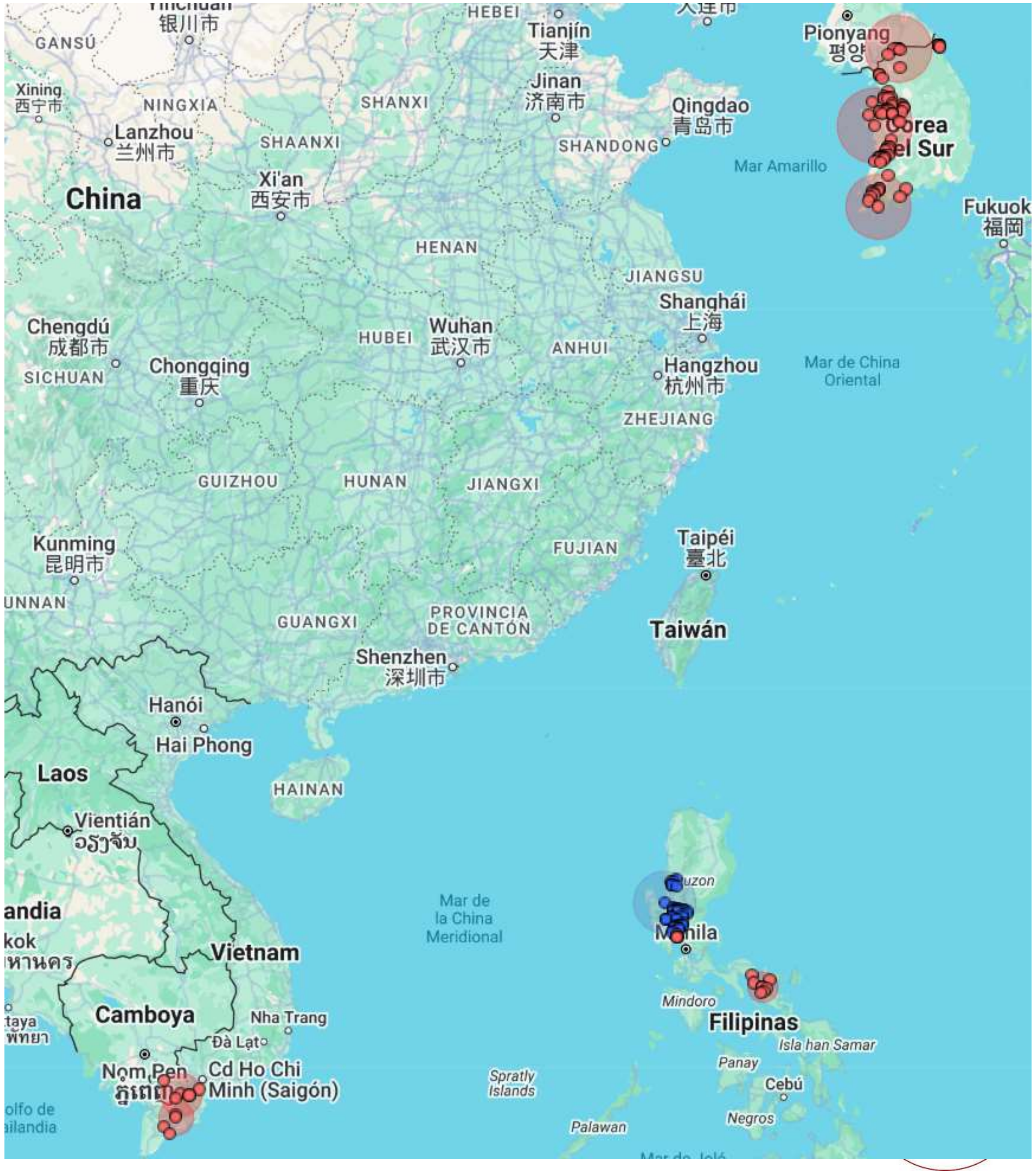
Check TXT : Perm_100km_90d

15 cluster:

- **10 significatives (p value<0.05)**
- **9 observed/expected >1 (Red)**
- **1 observed/expected <1 (Blue)**



```
                    SaTScan v9.6

Program run on: Tue Feb 25 08:08:10 2025

Retrospective Space-Time analysis
scanning for clusters with high or low rates
using the Space-Time Permutation model.

SUMMARY OF DATA

Study period.........................: 2021/1/1 to 2022/12/31
Number of locations..................: 311
Total number of cases................: 395

CLUSTERS DETECTED

1.Location IDs included.: location346, location347, location349, location300, location310,
                          location311, location320, location329, location330, location331,
                          location332, location335, location336, location338, location339,
                          location340, location341, location342, location343, location350,
                          location351, location352, location353, location363, location309,
                          location366, location275, location280, location289, location291,
                          location293, location294, location297, location312, location313,
                          location314, location315, location316, location317, location318
```



Exercise Results

Open dbf .col with Excel :

CLUSTER	LOC_ID	LATITUDE	LONGITUDE	RADIUS	START_DATE	END_DATE	NUMBER_LOC	TEST_STAT	P_VALUE	OBSERVED	EXPECTED	ODE	GINI_CLUST
1	location346	14,935704	120,755243	5,83036944	2022/7/31	2022/8/6	6	34,6517011	7,8E-16	33	5,116455696	6,44977734	FALSO
2	location88	36,456	126,5915	99,9501069	2021/12/5	2022/2/19	52	27,4606615	6,177E-12	40	9,741772152	4,10602911	FALSO
3	location123	38,313993	128,531013	3,44750347	2022/3/27	2022/4/2	14	19,5040794	1,2475E-07	11	0,744303798	14,7789116	FALSO
4	location31	34,499518	126,694278	88,5206314	2021/11/7	2022/1/1	14	15,8914703	1,1241E-05	11	1,063291139	10,3452381	FALSO
5	location344	15,940039	120,35265	99,1881234	2021/12/12	2022/3/5	50	15,5923463	1,6318E-05	0	15,28860759	0	FALSO
6	location233	10,385	106,0049	68,8975106	2021/5/23	2021/7/17	6	15,4918114	1,8496E-05	7	0,296202532	23,6324786	FALSO
7	location3	9,7231	105,8607	57,6149362	2021/1/17	2021/1/23	4	12,8289759	0,00051028	4	0,060759494	65,8333333	FALSO
8	location122	38,268104	127,287422	86,7468215	2022/2/27	2022/4/2	18	11,9779573	0,0014726	16	3,511392405	4,55659697	FALSO
9	location187	6,766535	124,705411	22,6528506	2022/4/3	2022/4/9	11	11,3788089	0,00310404	8	0,789873418	10,1282051	FALSO
10	location111	13,51452	123,2652	49,1022647	2022/2/27	2022/3/26	9	10,8650169	0,0058794	10	1,448101266	6,90559441	FALSO
11	location20	15,7227	106,3951	6,66172372	2021/7/11	2021/7/17	2	7,78073206	0,277	2	0,015189873	131,666667	FALSO
12	location229	20,0909	106,0592	74,2339733	2022/10/23	2022/10/29	6	7,70740955	0,296	4	0,227848101	17,5555556	FALSO
13	location382	18,277	105,9759	94,244301	2022/10/23	2022/10/29	3	6,93700316	0,556	3	0,113924051	26,3333333	FALSO
14	location57	11,562	106,0735	84,7604901	2022/1/2	2022/1/8	2	5,84426361	0,957	2	0,040506329	49,375	FALSO
15	location10	11,8076	108,7607	0	2021/5/23	2021/7/3	1	5,40805297	0,992	2	0,050632911	39,5	FALSO

Exercise Results

Include duration : End_date – Star_Date

CLUSTER	LOC_ID	LATITUDE	LONGITUDE	RADIUS	START_DATE	END_DATE	Duration	NUMBER_LOC	TEST_STAT	P_VALUE	OBSERVED	EXPECTED	ODE	GINI_CLUST
1	location346	14,935704	120,755243	5,83036944	2022/7/31	2022/8/6	6	6	34,6517011	7,8E-16	33	5,116455696	6,44977734	FALSO
2	location88	36,456	126,5915	99,9501069	2021/12/5	2022/2/19	76	52	27,4606615	6,177E-12	40	9,741772152	4,10602911	FALSO
3	location123	38,313993	128,531013	3,44750347	2022/3/27	2022/4/2	6	14	19,5040794	1,2475E-07	11	0,744303798	14,7789116	FALSO
4	location31	34,499518	126,694278	88,5206314	2021/11/7	2022/1/1	55	14	15,8914703	1,1241E-05	11	1,063291139	10,3452381	FALSO
5	location344	15,940039	120,35265	99,1881234	2021/12/12	2022/3/5	83	50	15,5923463	1,6318E-05	0	15,28860759	0	FALSO
6	location233	10,385	106,0049	68,8975106	2021/5/23	2021/7/17	55	6	15,4918114	1,8496E-05	7	0,296202532	23,6324786	FALSO
7	location3	9,7231	105,8607	57,6149362	2021/1/17	2021/1/23	6	4	12,8289759	0,00051028	4	0,060759494	65,8333333	FALSO
8	location122	38,268104	127,287422	86,7468215	2022/2/27	2022/4/2	34	18	11,9779573	0,0014726	16	3,511392405	4,55659697	FALSO
9	location187	6,766535	124,705411	22,6528506	2022/4/3	2022/4/9	6	11	11,3788089	0,00310404	8	0,789873418	10,1282051	FALSO
10	location111	13,51452	123,2652	49,1022647	2022/2/27	2022/3/26	27	9	10,8650169	0,0058794	10	1,448101266	6,90559441	FALSO
11	location20	15,7227	106,3951	6,66172372	2021/7/11	2021/7/17	6	2	7,78073206	0,277	2	0,015189873	131,666667	FALSO
12	location229	20,0909	106,0592	74,2339733	2022/10/23	2022/10/29	6	6	7,70740955	0,296	4	0,227848101	17,5555556	FALSO
13	location382	18,277	105,9759	94,244301	2022/10/23	2022/10/29	6	3	6,93700316	0,556	3	0,113924051	26,3333333	FALSO
14	location57	11,562	106,0735	84,7604901	2022/1/2	2022/1/8	6	2	5,84426361	0,957	2	0,040506329	49,375	FALSO
15	location10	11,8076	108,7607	0	2021/5/23	2021/7/3	41	1	5,40805297	0,992	2	0,050632911	39,5	FALSO

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTER

INIA CSIC

Exercise Results

Basic Description of SaTScan Results:

- 10 significant clusters detected:
  - 1 cluster with observed cases lower than expected
  - 9 clusters with observed cases higher than expected (Mean O/E ratio = 16)
- Seasonal Distribution:
  - 5/10 clusters occurred in winter
  - 4/10 clusters occurred in spring
  - 1/10 cluster occurred in summer
- Spatial and Temporal Characteristics:
  - Mean radius: 56 km
  - Mean duration: 27.93 days
- High Incidence Clusters:
  - 4 in South Korea
  - 3 in the Philippines
  - 2 in Vietnam
- Low Incidence Cluster:
  - 1 in the Philippines

CLUSTER	LOC_ID	LATITUDE	LONGITUDE	RADIUS	START_DATE	END_DATE	Duration	NUMBER_LOC	TEST_STAT	P_VALUE	OBSERVED	EXPECTED	ODE	GINI_CLUST
1	location346	14,935704	120,755243	5,83036944	2022/7/31	2022/8/6	6	6	34,6517011	7,8E-16	33	5,116455696	6,44977734	FALSO
2	location88	36,456	126,5915	99,9501069	2021/12/5	2022/2/19	76	52	27,4606615	6,177E-12	40	9,741772152	4,10602911	FALSO
3	location123	38,313993	128,531013	3,44750347	2022/3/27	2022/4/2	6	14	19,5040794	1,2475E-07	11	0,744303798	14,7789116	FALSO
4	location31	34,499518	126,694278	88,5206314	2021/11/7	2022/1/1	55	14	15,8914703	1,1241E-05	11	1,063291139	10,3452381	FALSO
5	location344	15,940039	120,35265	99,1881234	2021/12/12	2022/3/5	83	50	15,5923463	1,6318E-05	0	15,28860759	0	FALSO
6	location233	10,385	106,0049	68,8975106	2021/5/23	2021/7/17	55	6	15,4918114	1,8496E-05	7	0,296202532	23,6324786	FALSO
7	location3	9,7231	105,8607	57,6149362	2021/1/17	2021/1/23	6	4	12,8289759	0,00051028	4	0,060759494	65,8333333	FALSO
8	location122	38,268104	127,287422	86,7468215	2022/2/27	2022/4/2	34	18	11,9779573	0,0014726	16	3,511392405	4,55659697	FALSO
9	location187	6,766535	124,705411	22,6528506	2022/4/3	2022/4/9	6	11	11,3788089	0,00310404	8	0,789873418	10,1282051	FALSO
10	location111	13,51452	123,2652	49,1022647	2022/2/27	2022/3/26	27	9	10,8650169	0,0058794	10	1,448101266	6,90559441	FALSO
11	location20	15,7227	106,3951	6,66172372	2021/7/11	2021/7/17	6	2	7,78073206	0,277	2	0,015189873	131,666667	FALSO
12	location229	20,0909	106,0592	74,2339733	2022/10/23	2022/10/29	6	6	7,70740955	0,296	4	0,227848101	17,5555556	FALSO
13	location382	18,277	105,9759	94,244301	2022/10/23	2022/10/29	6	3	6,93700316	0,556	3	0,113924051	26,3333333	FALSO
14	location57	11,562	106,0735	84,7604901	2022/1/2	2022/1/8	6	2	5,84426361	0,957	2	0,040506329	49,375	FALSO
15	location10	11,8076	108,7607	0	2021/5/23	2021/7/3	41	1	5,40805297	0,992	2	0,050632911	39,5	FALSO

Exercise Results

<u>Basic Description of SaTScan Results: Interpretation</u>

SaTScan analysis identified spatial-temporal clusters of HPAI H5N1 outbreaks, highlighting areas with high and low case densities. The majority of high-incidence clusters were observed in South Korea, the Philippines, and Vietnam, while one low-incidence cluster was detected in the Philippines. Clusters were primarily concentrated in the winter and spring seasons, indicating potential seasonal patterns in outbreak occurrences.

CLUSTER	LOC_ID	LATITUDE	LONGITUDE	RADIUS	START_DATE	END_DATE	Duration	NUMBER_LOC	TEST_STAT	P_VALUE	OBSERVED	EXPECTED	ODE	GINI_CLUST
1	location346	14,935704	120,755243	5,83036944	2022/7/31	2022/8/6	6	6	34,6517011	7,8E-16	33	5,116455696	6,44977734	FALSO
2	location88	36,456	126,5915	99,9501069	2021/12/5	2022/2/19	76	52	27,4606615	6,177E-12	40	9,741772152	4,10602911	FALSO
3	location123	38,313993	128,531013	3,44750347	2022/3/27	2022/4/2	6	14	19,5040794	1,2475E-07	11	0,744303798	14,7789116	FALSO
4	location31	34,499518	126,694278	88,5206314	2021/11/7	2022/1/1	55	14	15,8914703	1,1241E-05	11	1,063291139	10,3452381	FALSO
5	location344	15,940039	120,35265	99,1881234	2021/12/12	2022/3/5	83	50	15,5923463	1,6318E-05	0	15,28860759	0	FALSO
6	location233	10,385	106,0049	68,8975106	2021/5/23	2021/7/17	55	6	15,4918114	1,8496E-05	7	0,296202532	23,6324786	FALSO
7	location3	9,7231	105,8607	57,6149362	2021/1/17	2021/1/23	6	4	12,8289759	0,00051028	4	0,060759494	65,8333333	FALSO
8	location122	38,268104	127,287422	86,7468215	2022/2/27	2022/4/2	34	18	11,9779573	0,0014726	16	3,511392405	4,55659697	FALSO
9	location187	6,766535	124,705411	22,6528506	2022/4/3	2022/4/9	6	11	11,3788089	0,00310404	8	0,789873418	10,1282051	FALSO
10	location111	13,51452	123,2652	49,1022647	2022/2/27	2022/3/26	27	9	10,8650169	0,0058794	10	1,448101266	6,90559441	FALSO
11	location20	15,7227	106,3951	6,66172372	2021/7/11	2021/7/17	6	2	7,78073206	0,277	2	0,015189873	131,666667	FALSO
12	location229	20,0909	106,0592	74,2339733	2022/10/23	2022/10/29	6	6	7,70740955	0,296	4	0,227848101	17,5555556	FALSO
13	location382	18,277	105,9759	94,244301	2022/10/23	2022/10/29	6	3	6,93700316	0,556	3	0,113924051	26,3333333	FALSO
14	location57	11,562	106,0735	84,7604901	2022/1/2	2022/1/8	6	2	5,84426361	0,957	2	0,040506329	49,375	FALSO
15	location10	11,8076	108,7607	0	2021/5/23	2021/7/3	41	1	5,40805297	0,992	2	0,050632911	39,5	FALSO

Exercise Results

Visualizing SaTScan Results

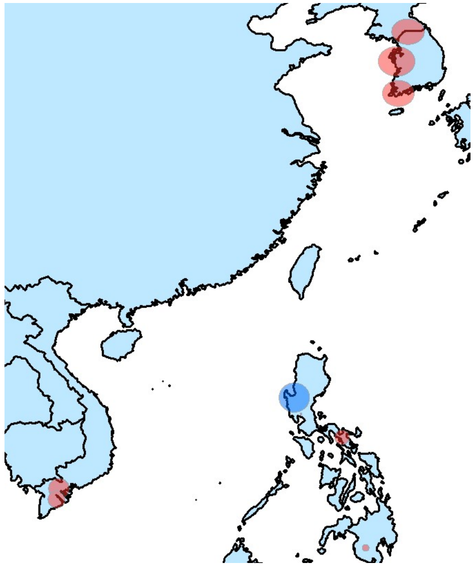
1. Load the Results in a GIS Software

•Open **QGIS**, **ArcGIS**, or any other GIS software.

•Import the **SaTScan output files .col**

Perm_100km_90d.gis.shp

Perm_100km_90d.col.shp



. **2. Display the Cluster Data**

•Load the **base map** (e.g., "countries" shapefile) for geographic reference.

•Overlay the **clusters layer** and apply a **graduated color scale** to distinguish high and low observed/expected (O/E) ratios.

•Set **symbols** to differentiate:

- •**High-incidence clusters** (Observed > Expected)
- •**Low-incidence clusters** (Observed < Expected)

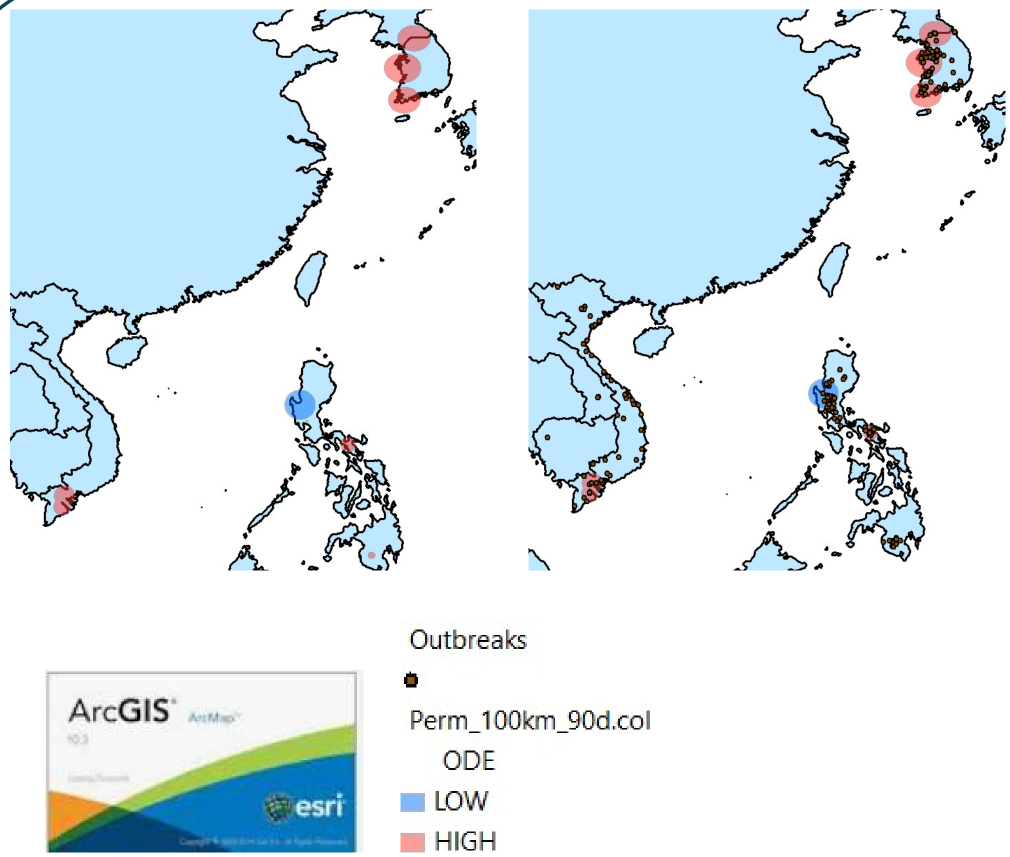
3. Analyze Spatial Patterns

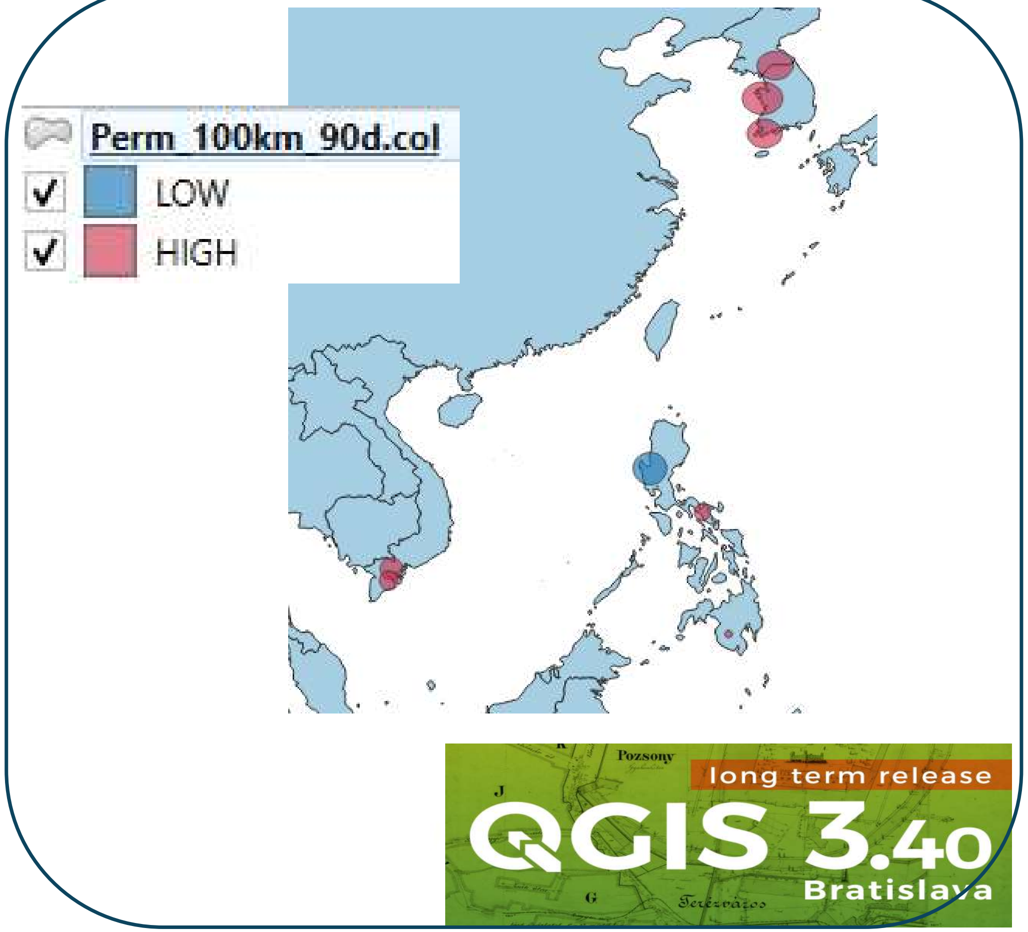
•Use **symbology settings** to categorize clusters by **season (winter, spring, summer)**.

•Adjust **transparency** to visualize overlapping clusters.

Exercise Results

Visualizing SaTScan Results





PART 5
INTERPOLATION AND SMOOTHING

Concept of Interpolation

- **Objective:** Estimate (predict) the value of a variable at **unsampled locations**, based on known values at sampled points.
- **Main characteristic:**
  - Assumes the existence of a **continuous field or surface** of the variable (e.g., temperature, altitude, contamination levels).
  - Many interpolation methods (e.g., **IDW, Kriging, Splines**) generate surfaces that **pass through or approximate observed values**.
- **Examples of interpolation in GIS**

❖ Estimating **altitude** between known elevation points.

❖ Predicting **temperature** in locations without meteorological stations→ survival pathogens.

❖ Estimating **contaminant concentration** based on scattered sampling sites.

Concept of Spatial Smoothing

- Spatial smoothing techniques, such as **Kernel Density Estimation (KDE)** create a **continuous density surface** based on point distribution.
- **Objective:** Identify **intensity or concentration of events** over a geographic area.
- **Main characteristic:**
  - KDE does **not estimate real values** at unknown locations but calculates how **dense** a set of points (events) is in a given area.
  - Uses a **kernel function** to smooth point concentration over space, creating a **continuous density surface**.
- **Examples of spatial smoothing in GIS**

❖ Mapping **disease hotspots** in study areas.

❖ Identifying **high-risk zones** for disease outbreaks.

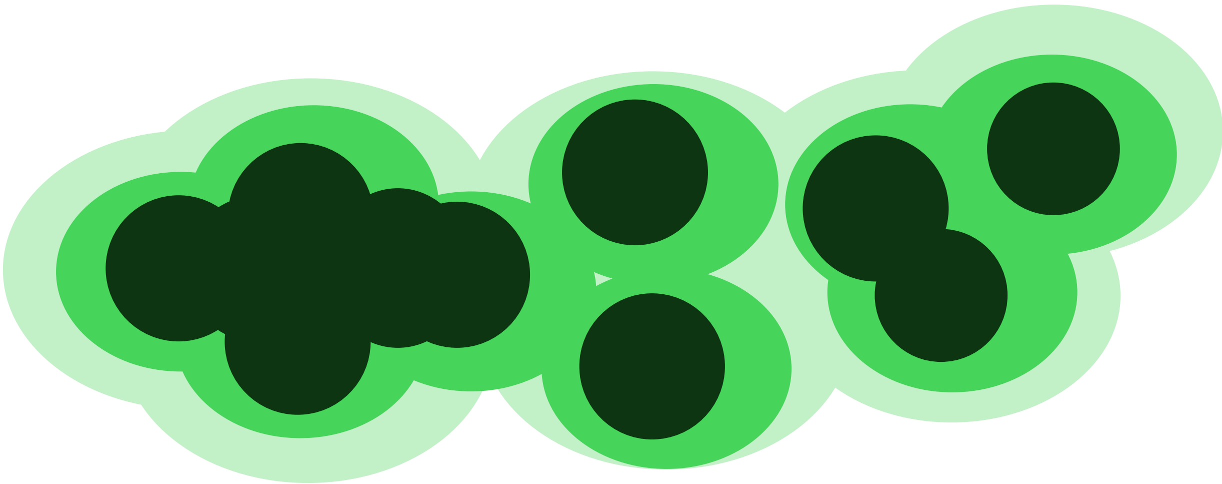
❖ Visualizing **wildlife habitat use** based on tracking data.

CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL / ANIMAL HEALTH RESEARCH CENTER

INIA CSIC

Concept of Spatial Smoothing: KDE

Transformation of point data into continuous data



CISA CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL ANIMAL HEALTH RESEARCH CENTER

INIA CSIC

Exercise 1: Self-assessment quiz on IDW vs KDE

Which of the following statements correctly describes the difference between Kernel Density Estimation (KDE) and Inverse Distance Weighting (IDW)?

A) KDE is an interpolation method that predicts values at unsampled locations, while IDW is a smoothing technique used to estimate event density.

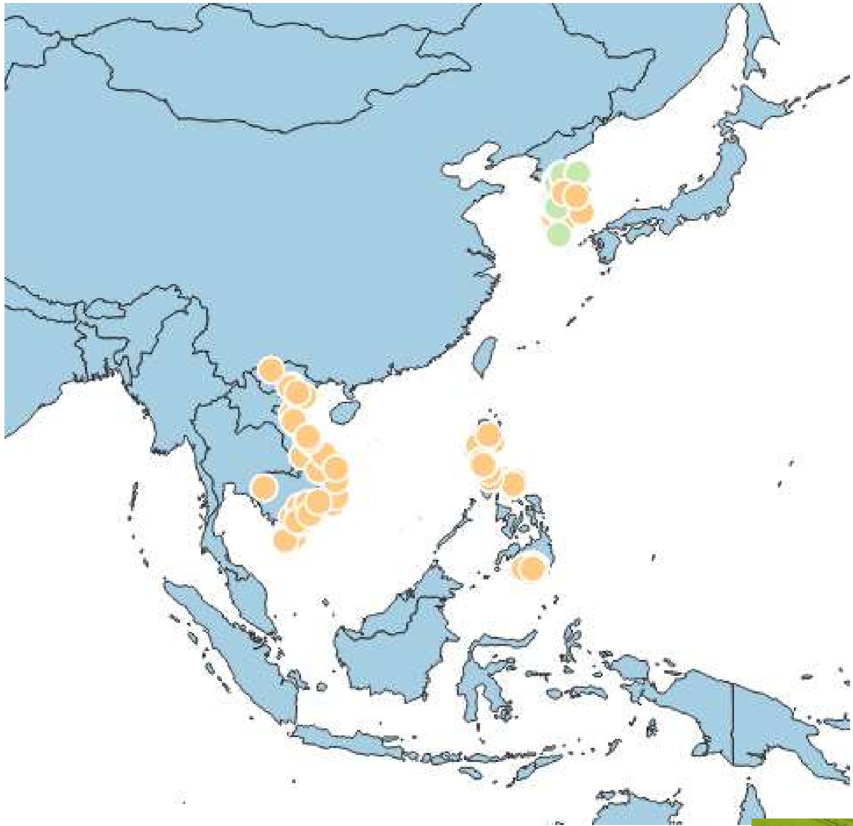
B) IDW creates a continuous density surface based on point distribution, whereas KDE estimates unknown values using weighted distances to known points.

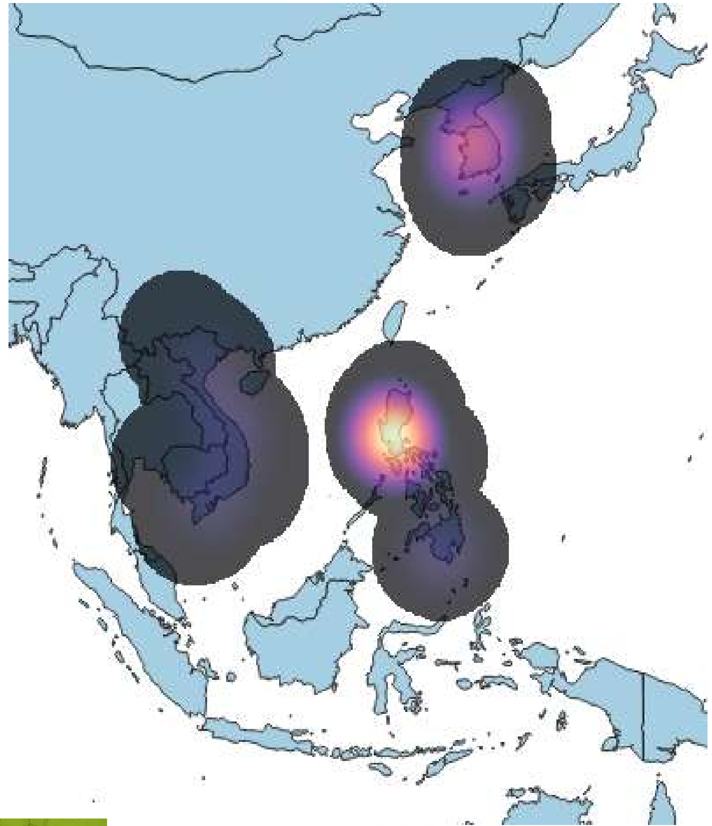
C) KDE generates a density surface representing the intensity of point events, while IDW estimates specific values at unknown locations based on weighted distances to known data points.

D) Both KDE and IDW perform interpolation by estimating missing values in a continuous spatial field.

Correct Answer: C)

Exercise 2: Spatial Smoothing: KDE QGIS



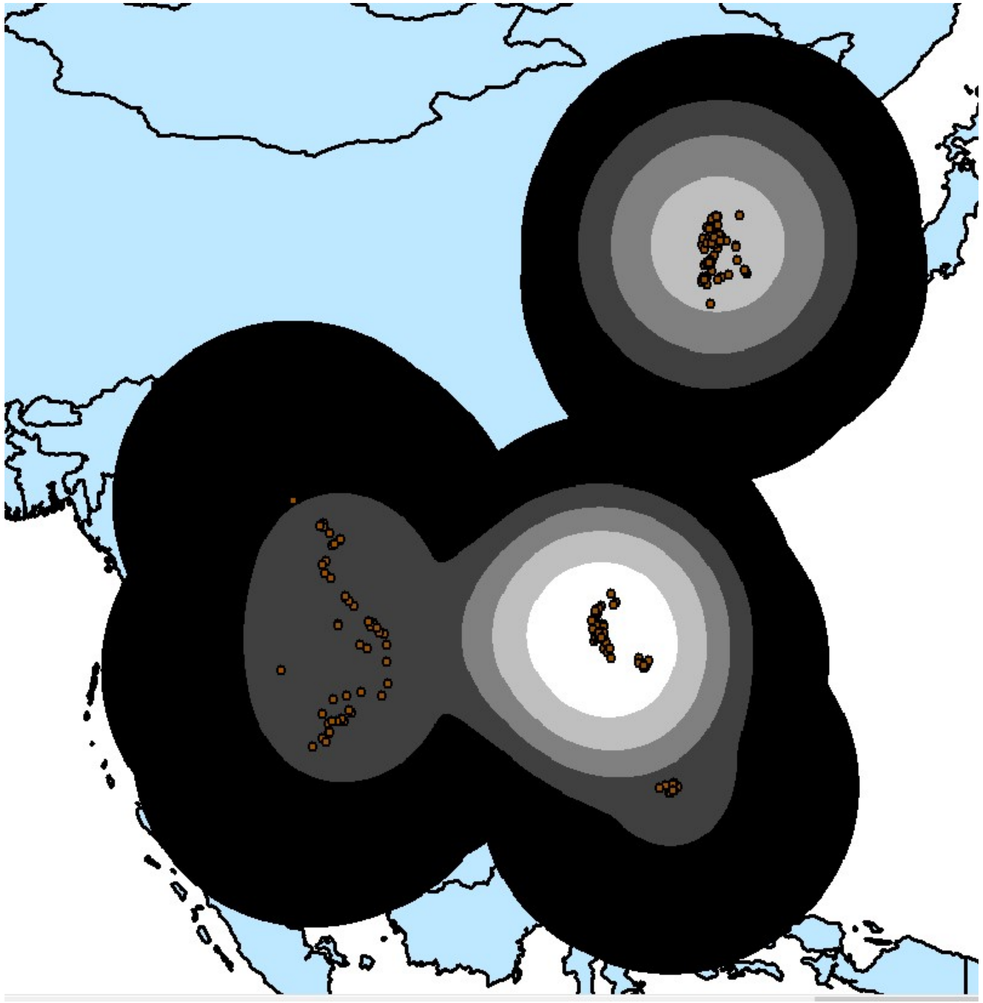


long term release
QGIS 3.40
Bratislava

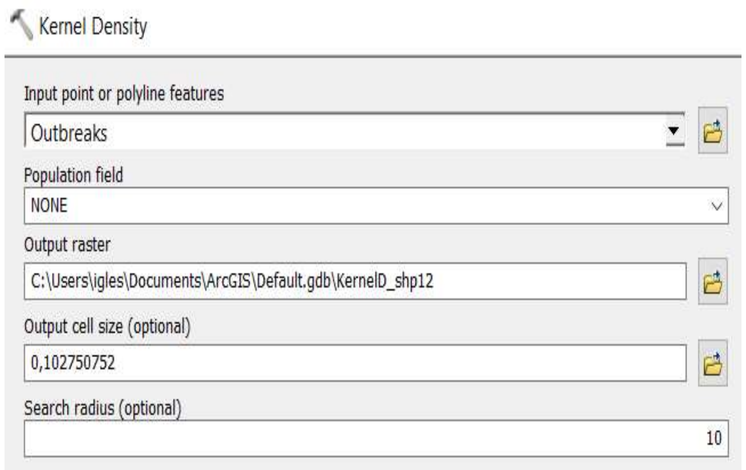
CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

INIA
CSIC

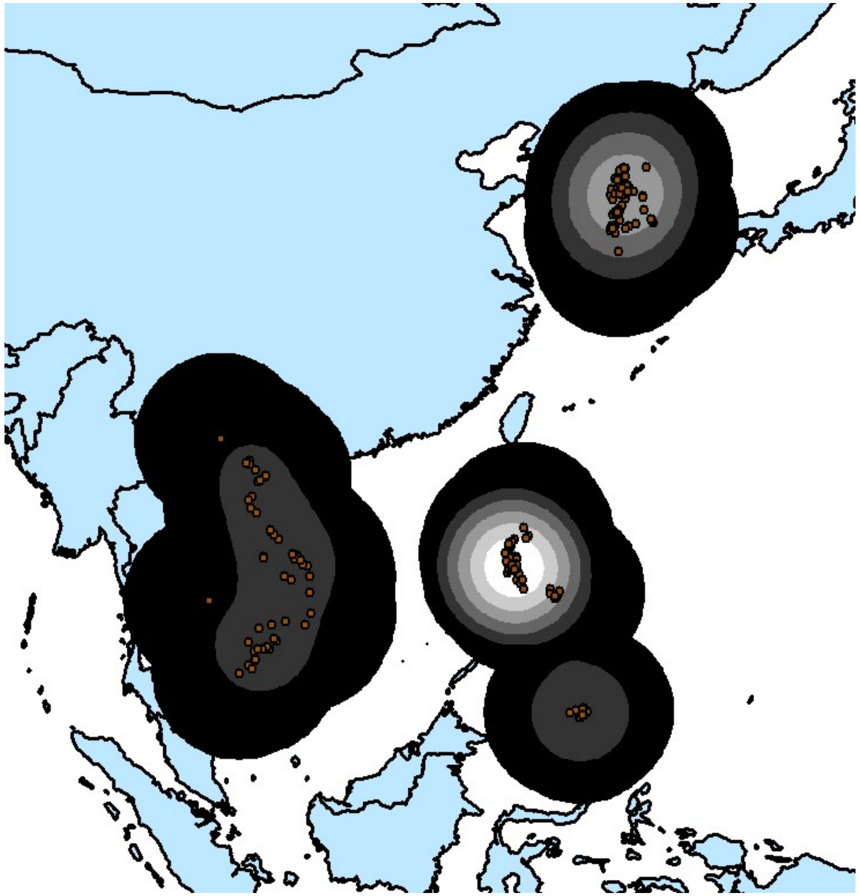
Exercise 2: KDE using ArcGIS



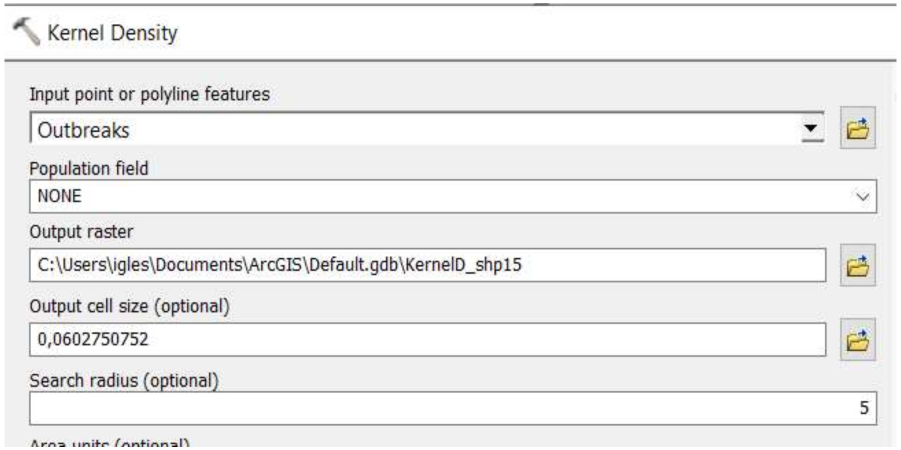
Radius 100km



Exercise 2: KDE using ArcGIS



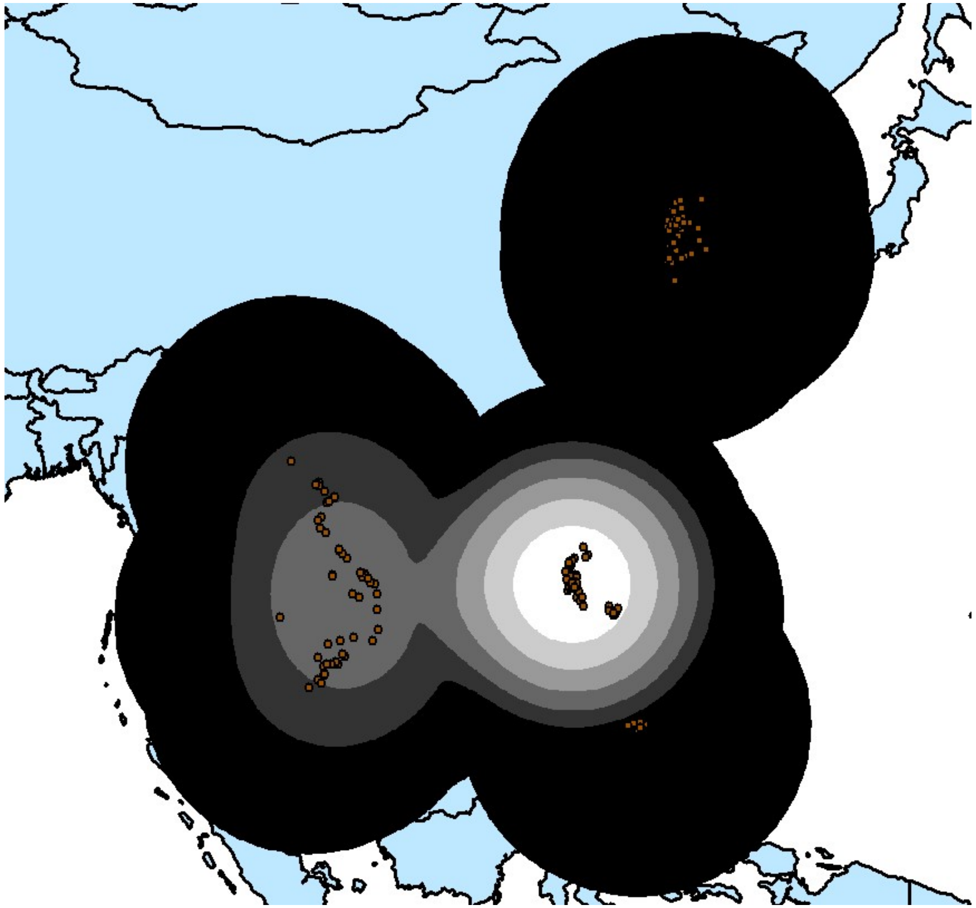
Radius 50km



CISA
CENTRO DE INVESTIGACIÓN EN SANIDAD ANIMAL
ANIMAL HEALTH RESEARCH CENTER

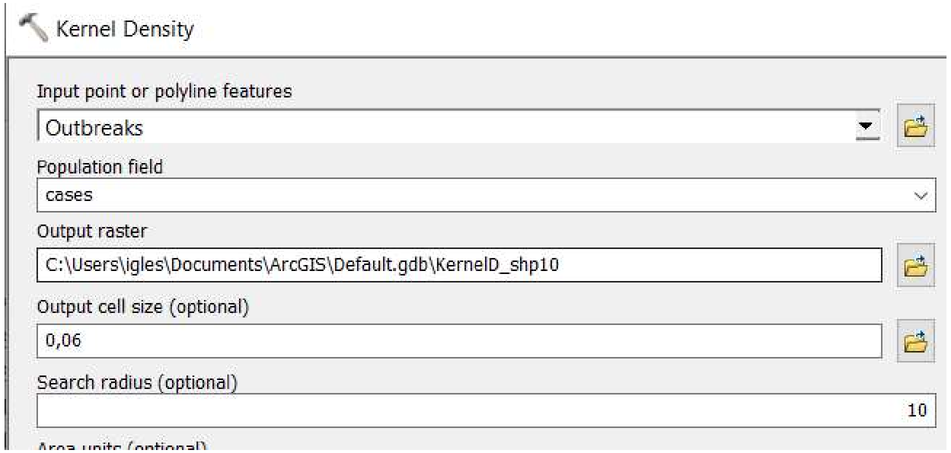
INIA
CSIC

Exercise 2: KDE using ArcGIS

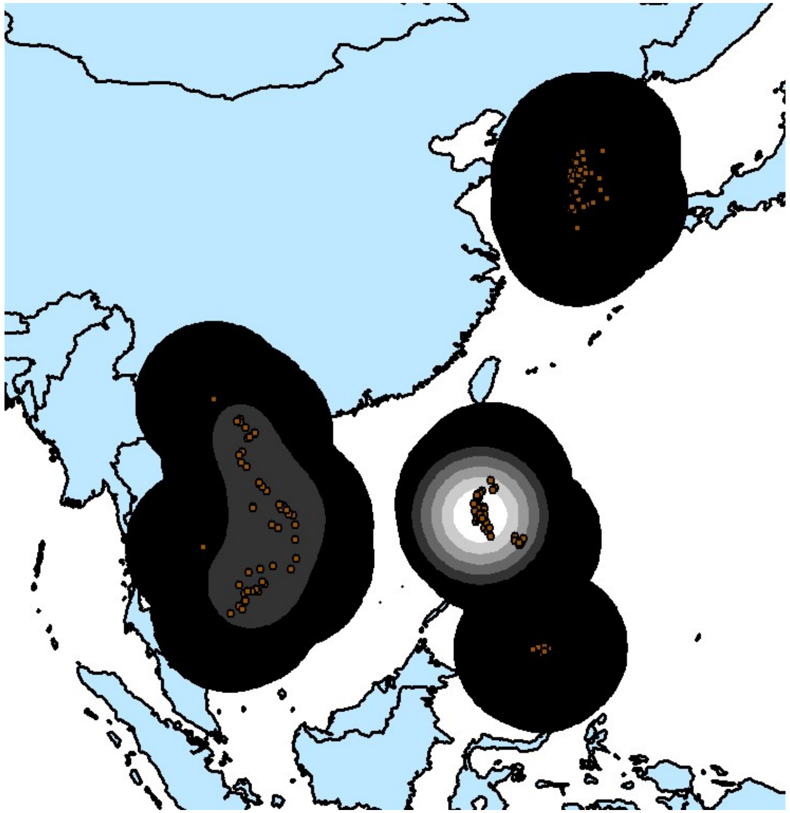


Radius 100km

Weight Field: Number of cases

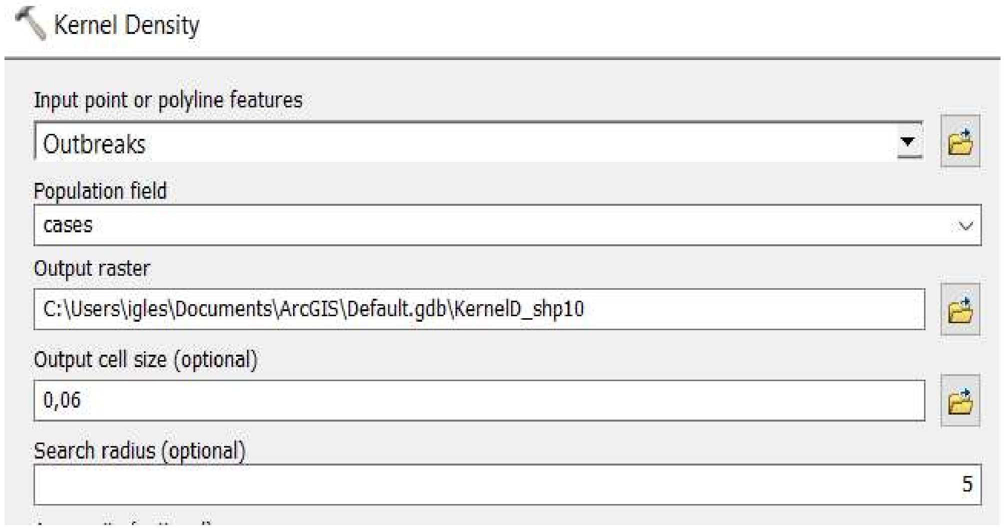


Exercise 2: KDE using ArcGIS

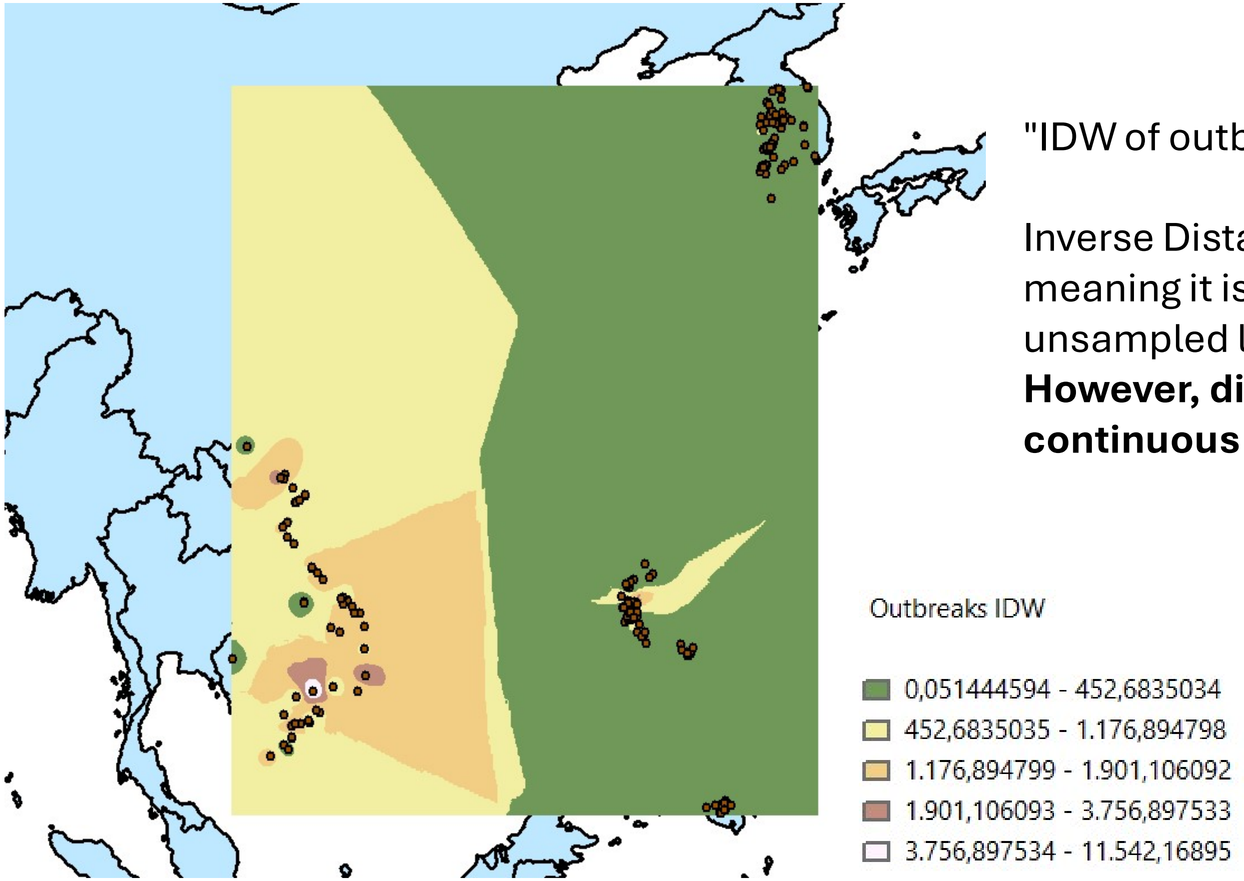


Radius 50km

Weight Field: Number of cases



Exercise 2: KDE



"IDW of outbreaks using ArcGIS“→ INCORRECT

Inverse Distance Weighting is an interpolation method, meaning it is used to estimate unknown values at unsampled locations based on known data points. **However, disease outbreaks are discrete events, not continuous variables.**