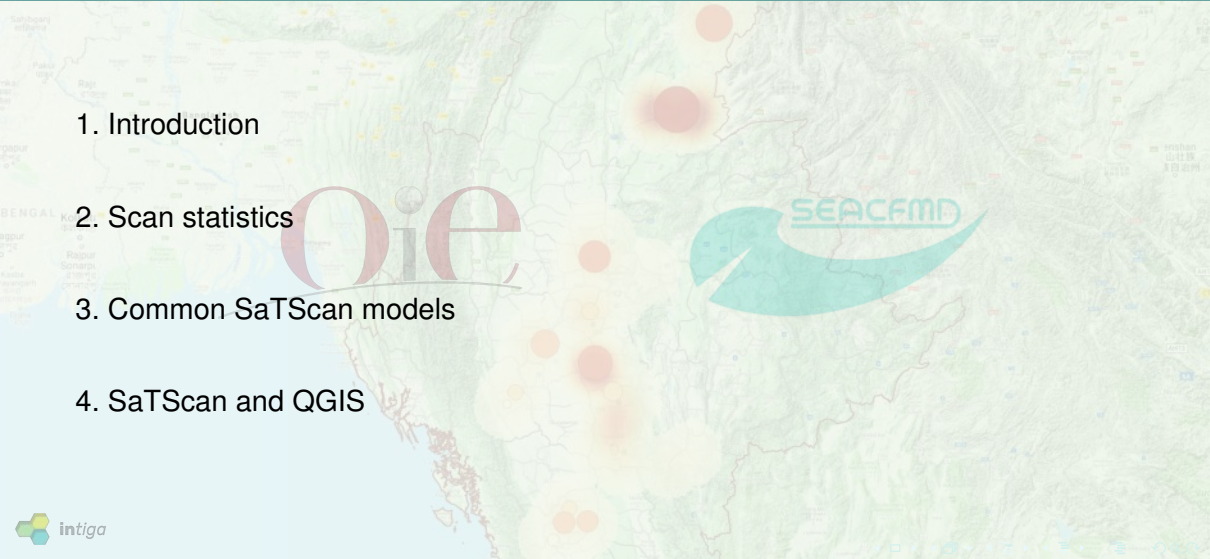


Local cluster detection using spatial scan statistics



Daan Vink

17 October 2018

A map of India showing several hotspots of animal health data. The hotspots are represented by semi-transparent circles of varying sizes and colors, ranging from small yellow circles to large red circles. The largest hotspots are located in the northern and central regions of the country. The map also shows state boundaries and major cities.

1. Introduction

2. Scan statistics

3. Common SaTScan models

4. SaTScan and QGIS

Oie



What is SaTScan?



- SaTScan is free software that analyses spatial, temporal and space-time data using the spatial, temporal, or space-time scan statistics.
- It is designed for any of the following interrelated purposes:

What is SaTScan?



- SaTScan is free software that analyses spatial, temporal and space-time data using the spatial, temporal, or space-time scan statistics.
- It is designed for any of the following interrelated purposes:
 - Perform **geographical surveillance** of disease, to detect spatial or space-time disease clusters, and to see if they are statistically significant.

What is SaTScan?



- SaTScan is free software that analyses spatial, temporal and space-time data using the spatial, temporal, or space-time scan statistics.
- It is designed for any of the following interrelated purposes:
 - Perform **geographical surveillance** of disease, to detect spatial or space-time disease clusters, and to see if they are statistically significant.
 - Test whether a disease is **randomly distributed** over space, over time or over space and time.

What is SaTScan?



- SaTScan is free software that analyses spatial, temporal and space-time data using the spatial, temporal, or space-time scan statistics.
- It is designed for any of the following interrelated purposes:
 - Perform **geographical surveillance** of disease, to detect spatial or space-time disease clusters, and to see if they are statistically significant.
 - Test whether a disease is **randomly distributed** over space, over time or over space and time.
 - Evaluate the statistical significance of **disease cluster alarms**.

What is SaTScan?

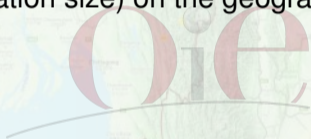


- SaTScan is free software that analyses spatial, temporal and space-time data using the spatial, temporal, or space-time scan statistics.
- It is designed for any of the following interrelated purposes:
 - Perform **geographical surveillance** of disease, to detect spatial or space-time disease clusters, and to see if they are statistically significant.
 - Test whether a disease is **randomly distributed** over space, over time or over space and time.
 - Evaluate the statistical significance of **disease cluster alarms**.
 - Perform repeated time-periodic disease surveillance for **early detection** of disease outbreaks.

How does the scan statistic work?

SaTScan:

- Imposes circles of different sizes (from zero up to a defined proportion of the population size) on the geographic area



How does the scan statistic work?

SaTScan:

- Imposes circles of different sizes (from zero up to a defined proportion of the population size) on the geographic area
- For each circle, computes a likelihood ratio statistic 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.

For each window, SaTScan calculates the likelihood, proportional to:

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

where

- n = number of cases inside circle
- N = total number of cases
- μ = expected number of cases inside circle

The scan statistic

SaTScan:

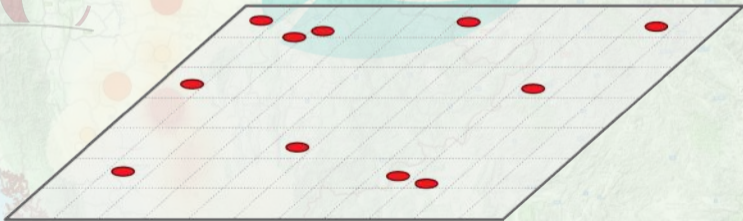
- Maximises the likelihood over all possible circles – this is the **scan statistic**.
- Tests significance by performing Monte Carlo hypothesis testing:
 - Redistributes cases randomly and recalculates the scan statistic many times (e.g.999).
 - Computes a p -value, which is the proportion of scan statistics from the Monte Carlo replicates which are greater than or equal to the scan statistic for the true cluster.

The spatial scan statistic (grid points)

- Study area with observed events.

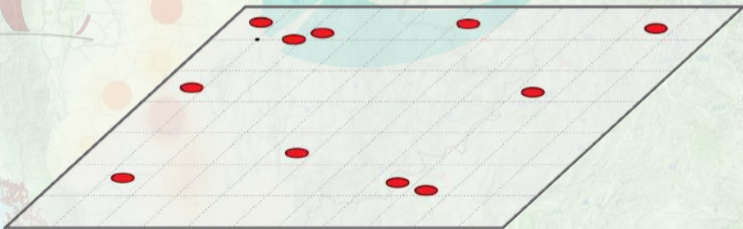
Oie

SEACFMD



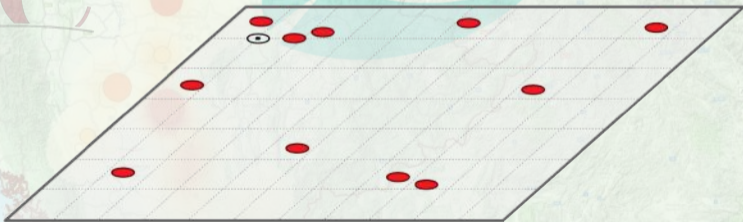
The spatial scan statistic (grid points)

- Place a point on the grid.



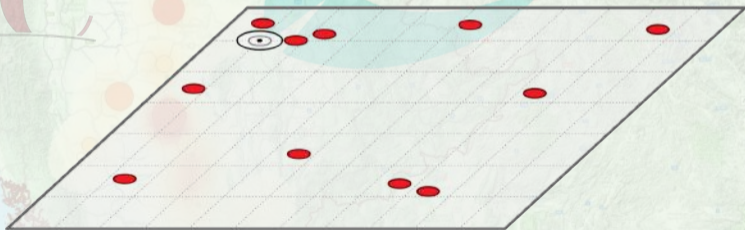
The spatial scan statistic (grid points)

- Circular scanning window with radius zero to upper limit.
- Observed and expected numbers of events are recorded.



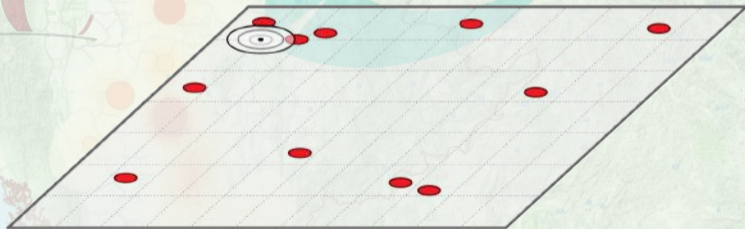
The spatial scan statistic (grid points)

- Circular scanning window with radius zero to upper limit.
- Observed and expected numbers of events are recorded.



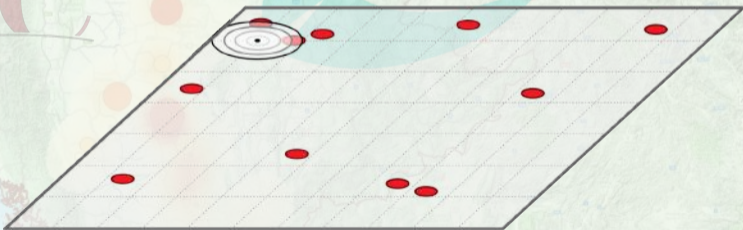
The spatial scan statistic (grid points)

- Circular scanning window with radius zero to upper limit.
- Observed and expected numbers of events are recorded.



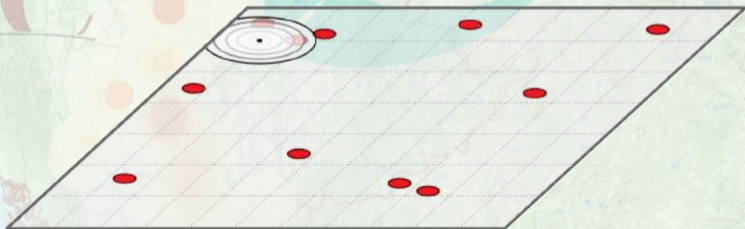
The spatial scan statistic (grid points)

- Circular scanning window with radius zero to upper limit.
- Observed and expected numbers of events are recorded.



The spatial scan statistic (grid points)

- Circular scanning window with radius zero to upper limit.
- Observed and expected numbers of events are recorded.

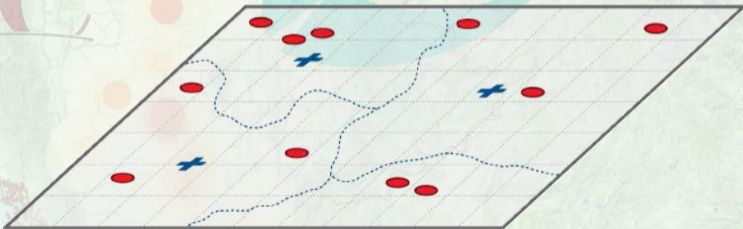


The spatial scan statistic (location ID points)

- Study area with observed events.

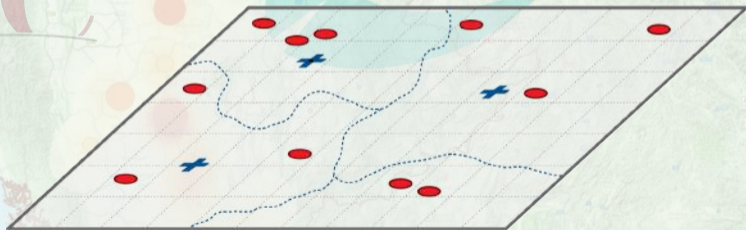
Oie

SEACFMD



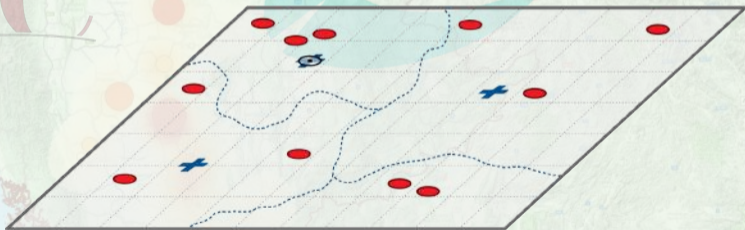
The spatial scan statistic (location ID points)

- Coordinates of a location ID (centroid).



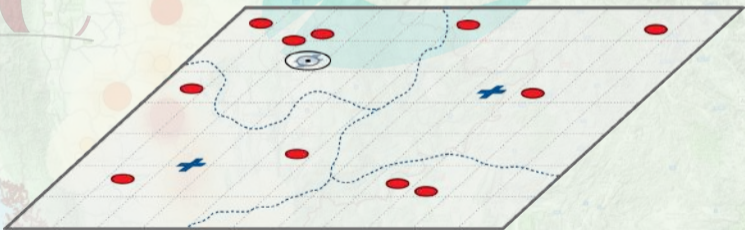
The spatial scan statistic (location ID points)

- Circular scanning window.
- Observed and expected numbers of events are recorded.



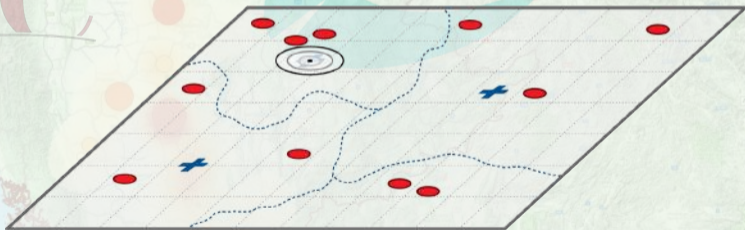
The spatial scan statistic (location ID points)

- Circular scanning window.
- Observed and expected numbers of events are recorded.



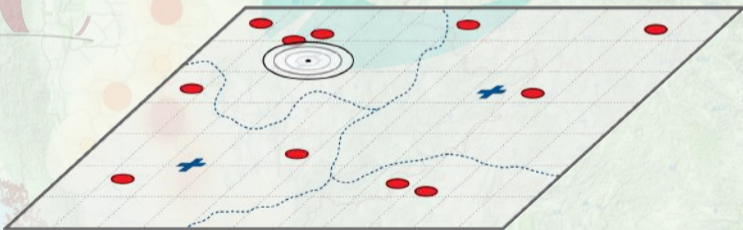
The spatial scan statistic (location ID points)

- Circular scanning window.
- Observed and expected numbers of events are recorded.



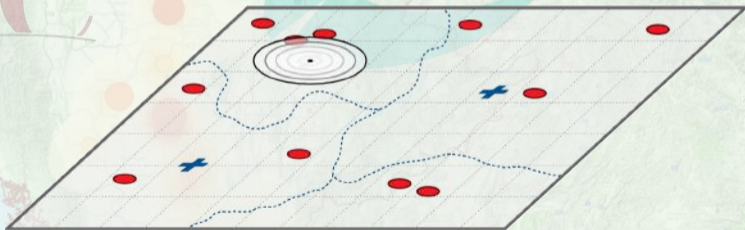
The spatial scan statistic (location ID points)

- Circular scanning window.
- Observed and expected numbers of events are recorded.



The spatial scan statistic (location ID points)

- Circular scanning window.
- Observed and expected numbers of events are recorded.



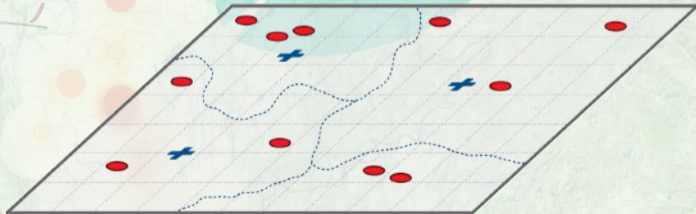
Space-time scan statistic

- Study area with observed events, at times t_1 to t_4 .

Oie

SEACFMD

t = 1



Space-time scan statistic

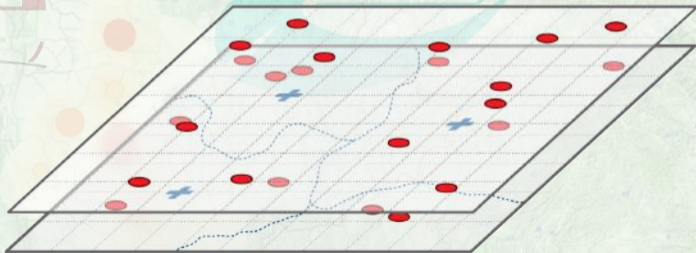
- Study area with observed events, at times t_1 to t_4 .

Oie

SEACFMD

$t = 2$

$t = 1$



Space-time scan statistic

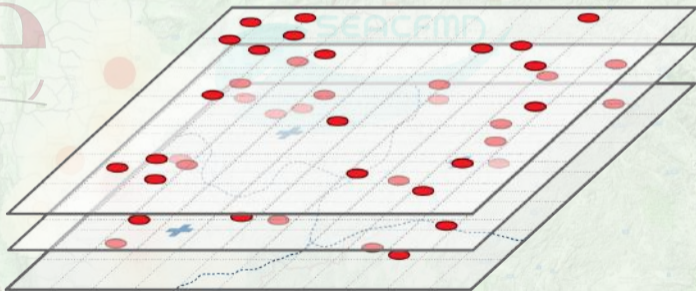
- Study area with observed events, at times t_1 to t_4 .

Oie

t = 3

t = 2

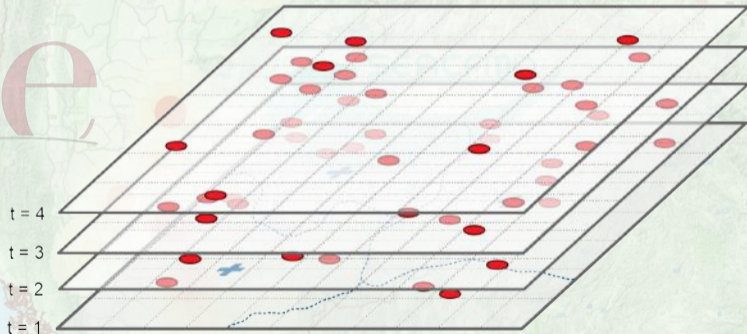
t = 1



Space-time scan statistic

- Study area with observed events, at times t_1 to t_4 .

Oie



Space-time scan statistic

- Coordinates of an observed event.

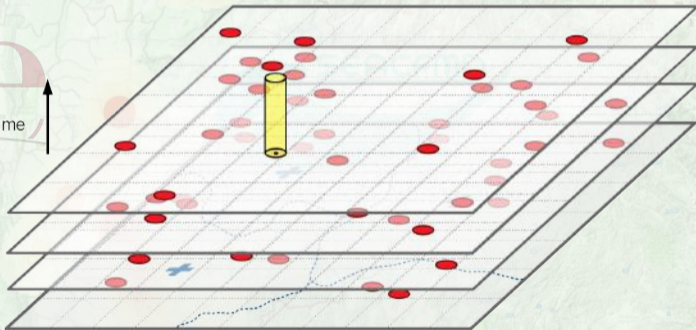
Time ↑

t = 4

t = 3

t = 2

t = 1

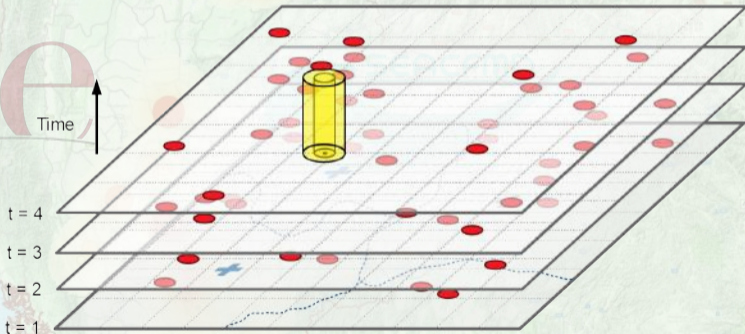


Space-time scan statistic

- Cylindrical scanning window.
- Observed and expected numbers of events are recorded.

Oie

Time ↑

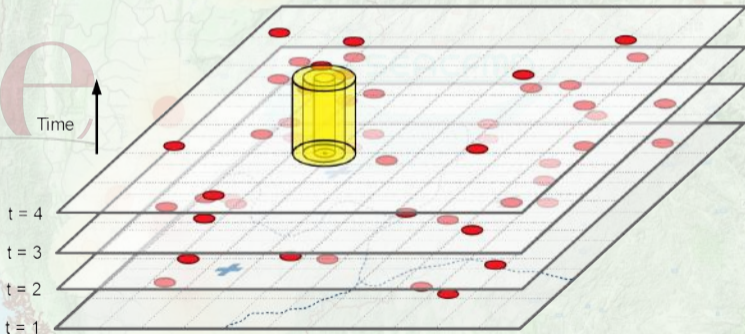


Space-time scan statistic

Oie

Time ↑

- Cylindrical scanning window.
- Observed and expected numbers of events are recorded.



Space-time scan statistic

- Cylindrical scanning window.
- Observed and expected numbers of events are recorded.

Oie

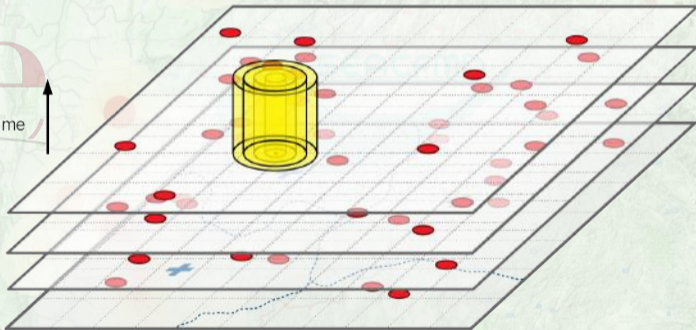
Time ↑

t = 4

t = 3

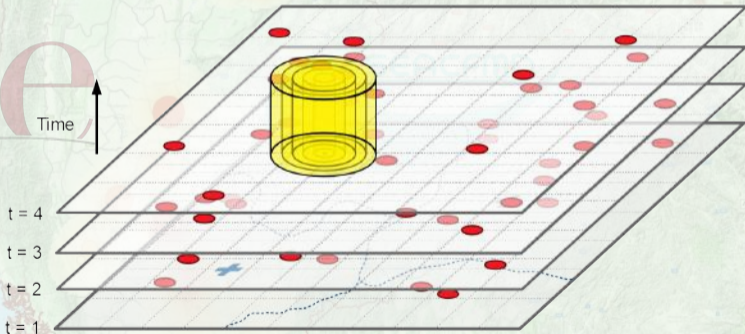
t = 2

t = 1



Space-time scan statistic

- Cylindrical scanning window.
- Observed and expected numbers of events are recorded.



Three types of models you may use:

① **Bernoulli model:**

- Animals have 0 / 1 disease status
- A set of cases and controls
- Purely temporal / spatial or space-time scan statistics

Three types of models you may use:

② Poisson model:

- The number of cases in each location is Poisson distributed
- Under the null hypothesis, the expected number of cases in each area is proportional to its population size
- Therefore, required information for the population at risk
- Purely temporal, purely spatial and space-time

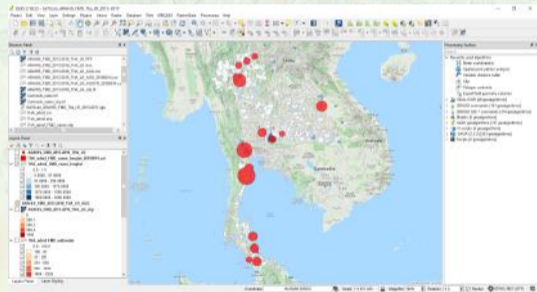
Three types of models you may use:

The logo for OIE (World Health Organization) is displayed in a large, semi-transparent font. It features the letters 'Oie' in a serif typeface, with a thin horizontal line underneath. The background of the slide is a map of South Asia with several red circular heatmaps overlaid, indicating areas of high disease prevalence.The logo for SEACFMD (South East Asia Centre for Food and Medical Diagnostics) is shown in a semi-transparent teal color. It consists of a stylized fish-like shape to the left of the text 'SEACFMD' in a bold, sans-serif font.

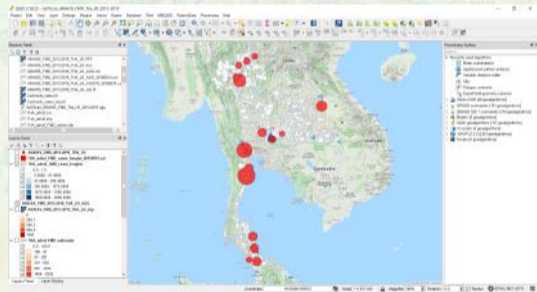
③ Space - Time Permutation model:

- Requires only case data with information about spatial location and time
- Assumes homogeneous population changes over the scanning period

- Data preparation steps:
 - The purely spatial Poisson model requires **aggregated data** (to District level)
 - We also need to prepare **population data** (expected number of cases) and **location data** (centroids of District polygons)



- Data preparation steps:
 - The purely spatial Poisson model requires **aggregated data** (to District level)
 - We also need to prepare **population data** (expected number of cases) and **location data** (centroids of District polygons)
- Running SaTScan can be fiddly!



- Data preparation steps:
 - The purely spatial Poisson model requires **aggregated data** (to District level)
 - We also need to prepare **population data** (expected number of cases) and **location data** (centroids of District polygons)
- Running SaTScan can be fiddly!
- SaTScan does not produce map output, but it generates shapefiles which can be loaded into QGIS.

