Local cluster detection using spatial scan statistics

Daan Vink

17 October 2018

1. Introduction
2. Scan statistics
3. Common SaTScan models
4. SaTScan and QGIS

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**.

intiga

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.

intiga

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

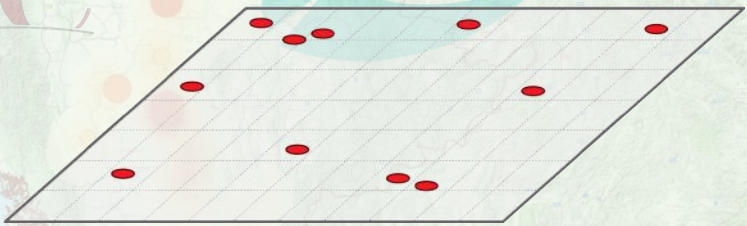
SaTScan:

- 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.

intiga

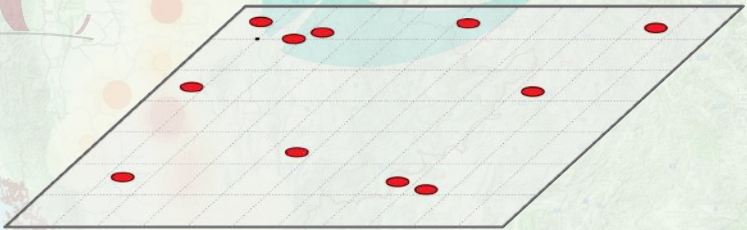
The spatial scan statistic (grid points)

- Study area with observed events.



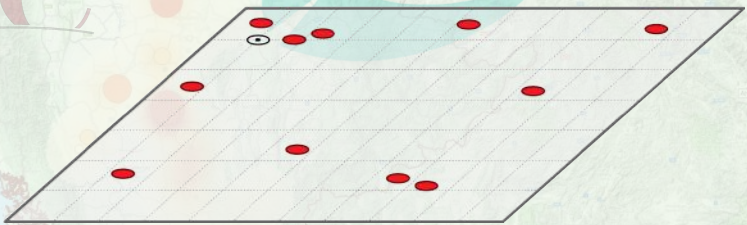
The spatial scan statistic (grid points)

- Place a point on the grid.



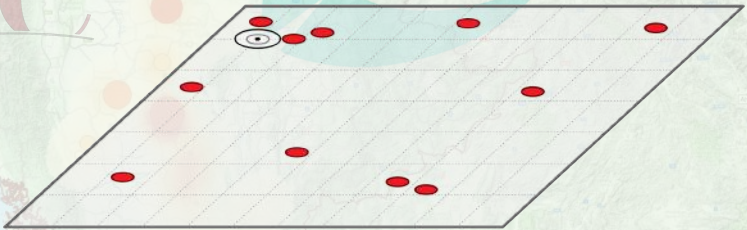
The spatial scan statistic (grid points)

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



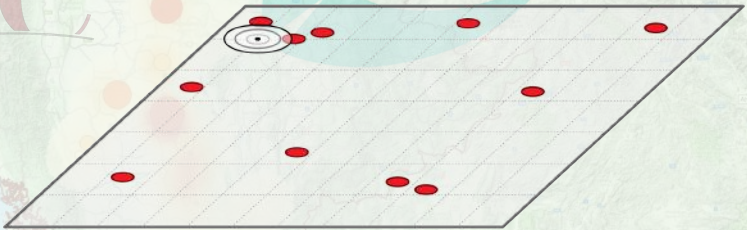
The spatial scan statistic (grid points)

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



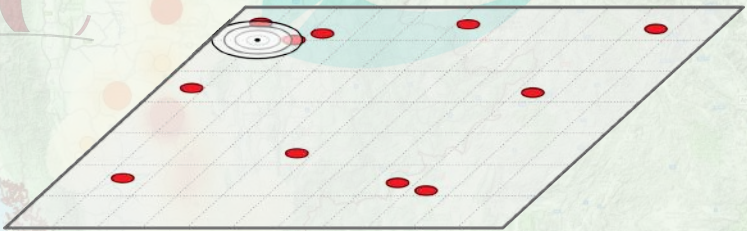
The spatial scan statistic (grid points)

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



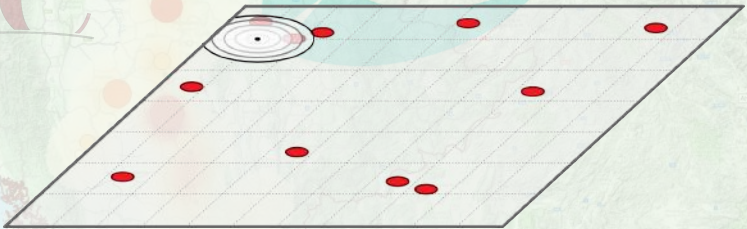
The spatial scan statistic (grid points)

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



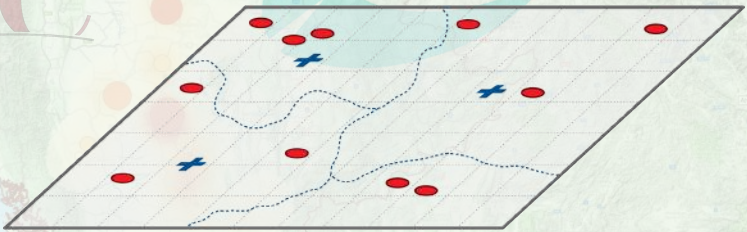
The spatial scan statistic (grid points)

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



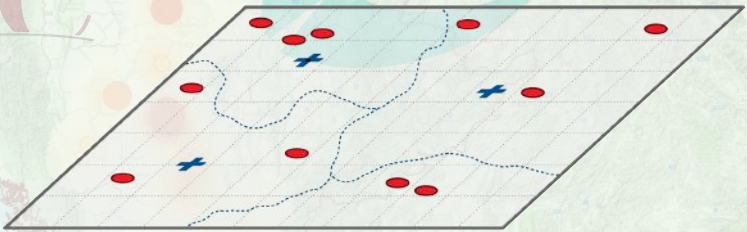
The spatial scan statistic (location ID points)

- Study area with observed events.



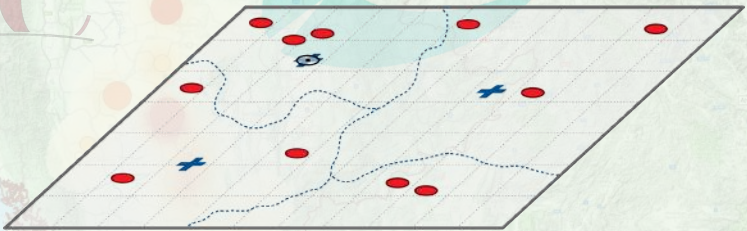
The spatial scan statistic (location ID points)

- Coordinates of a location ID (centroid).



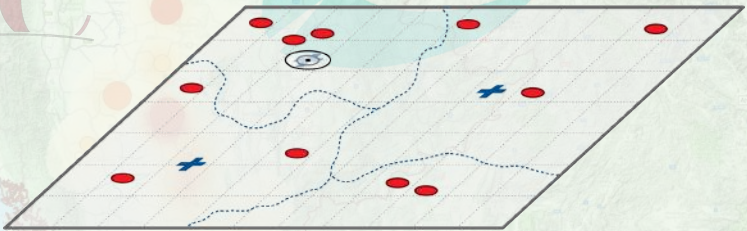
The spatial scan statistic (location ID points)

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



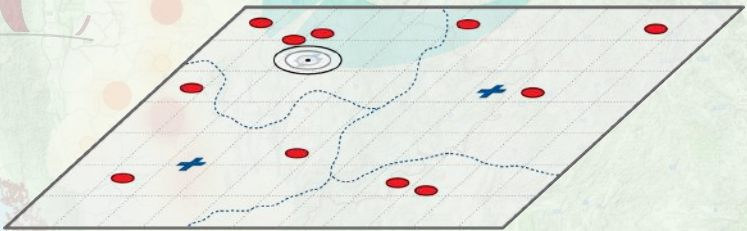
The spatial scan statistic (location ID points)

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



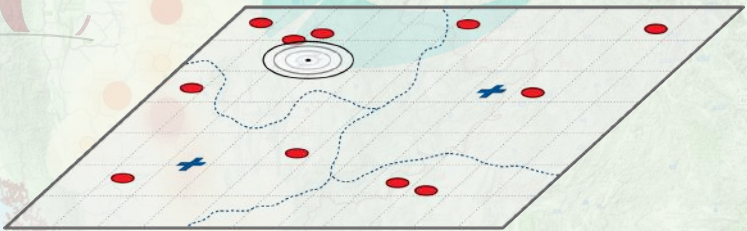
The spatial scan statistic (location ID points)

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



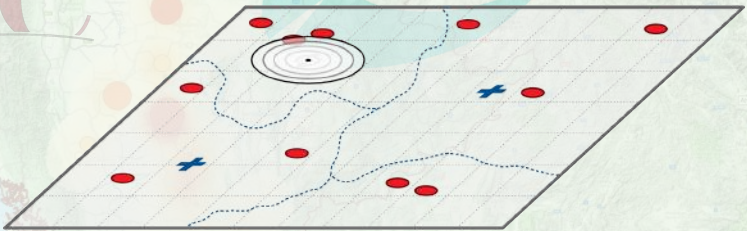
The spatial scan statistic (location ID points)

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



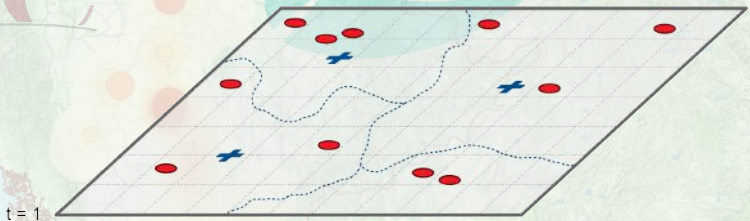
The spatial scan statistic (location ID points)

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



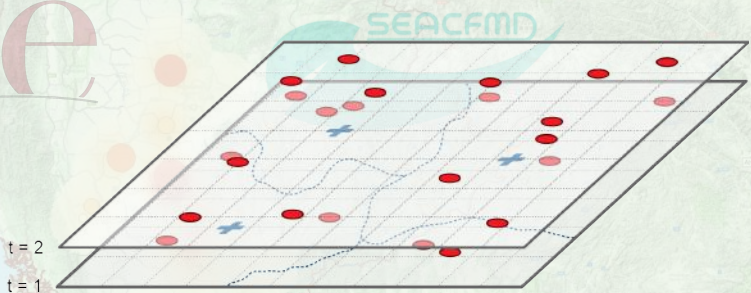
Space-time scan statistic

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



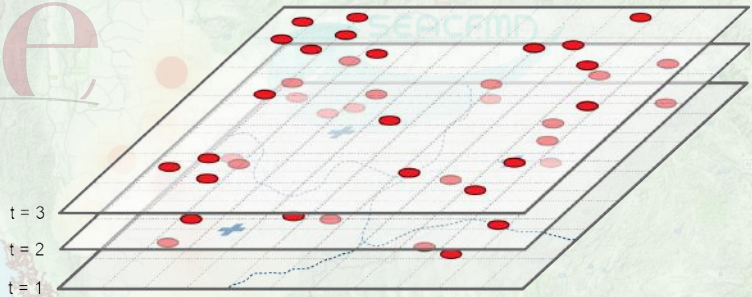
Space-time scan statistic

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



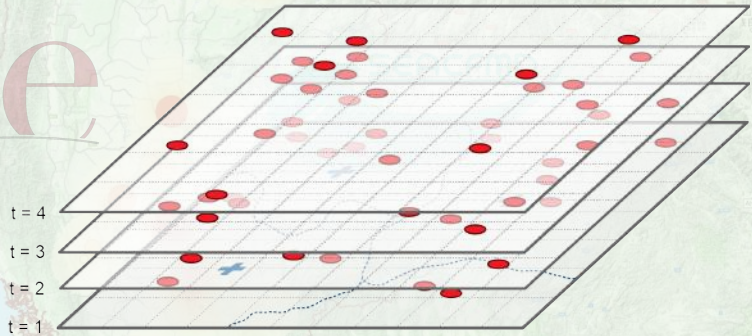
Space-time scan statistic

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



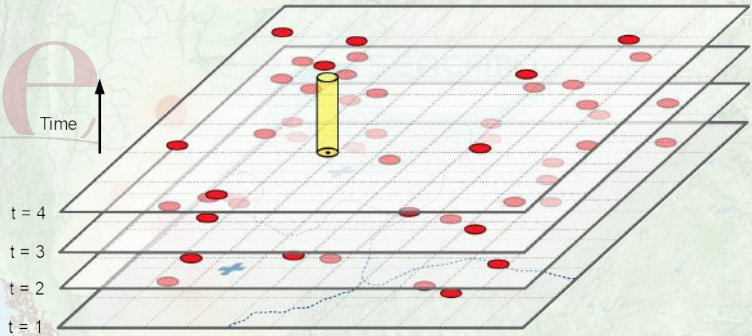
Space-time scan statistic

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



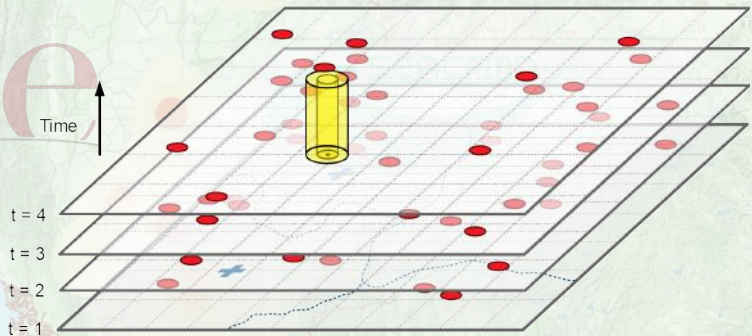
Space-time scan statistic

- Coordinates of an observed event.



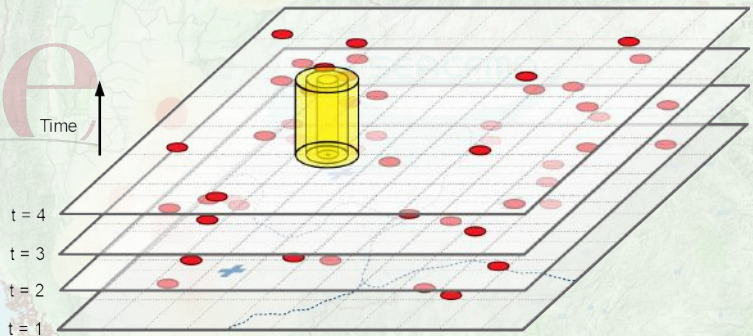
Space-time scan statistic

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



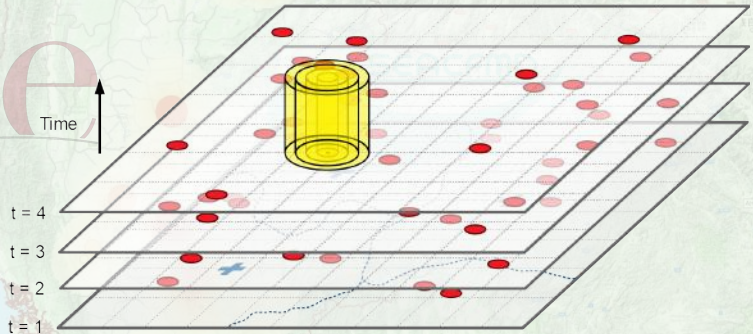
Space-time scan statistic

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



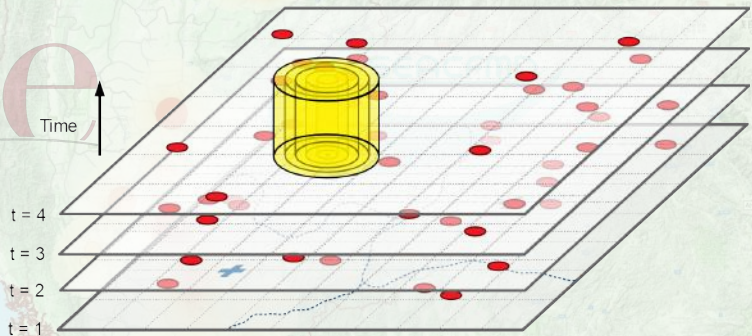
Space-time scan statistic

- 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.



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:

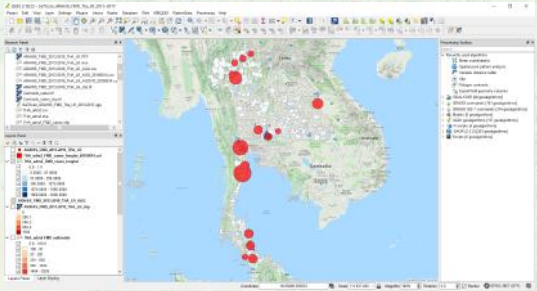
2. **Poisson model:**
   - The number of cases in each location is Poisson distributed
   - Under the null hypothesis, the expected number of cases in each area is proportional to its population size
   - Therefore, required information for the population at risk
   - Purely temporal, purely spatial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

Three types of models you may use:

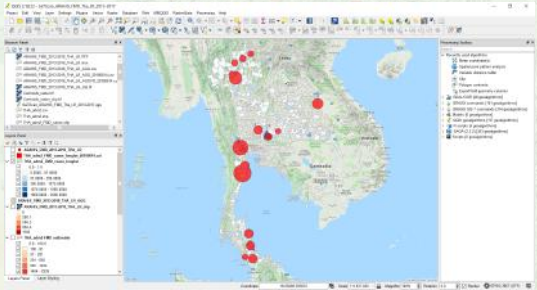
❸ **Space - Time Permutation model:**

- 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.