**Local cluster detection for point-process data**

This practical analyses incident cases of *P. falciparum* malaria and population controls from a prospective health facility-based, unmatched case-control study between December 2012 - July 2014. Cluster analysis will be performed using Kulldorff’s spatial scan statistic, implemented in SaTScan™. Since we are interested in analyzing the locations of malaria cases and controls, we will be using the Bernoulli model.

The model requires: case and control counts plus geographical coordinates for each location (each with household\_id to link the files). These have been generated for you but can be easily exported as .csv files from R.

**Data:**

1. Satscan\_cases.csv: columns “household\_id”, “ncase”

2. Satscan\_controls.csv: columns “household\_id”, “ncontrol”

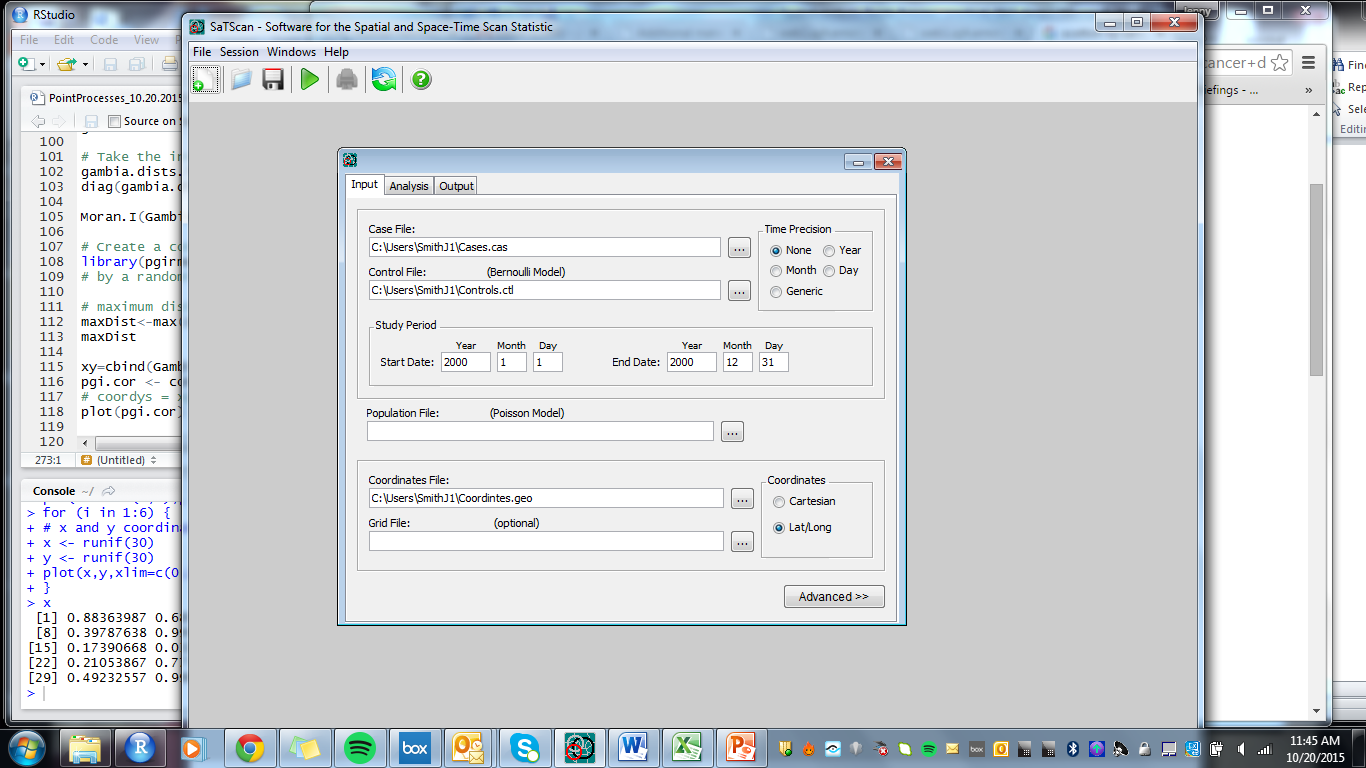
3. Satscan\_locations.csv: columns “household\_id”, “lat”, “long”

* Load SaTScan and click on **Create New Session**

**Step 1 – Input tab**

* We need to import the case data. Click on the right hand Untitled-2 symbol next to the Case File: section. Select the **SatScan\_cases.csv** file in your folder.
* The model should be set to Bernoulli. For Location IDselect **household\_id**; for Number of Cases select **ncase**. We do not need to add any optional information. Click **Next**.
* SaTScan will generate a case (.cas) file. Create a new folder to save this file in. Then click **Import.**
* We now need to import our control data, by clicking on the right hand Untitled-2 symbol next to the Control File. Select that the “First row is column name” and assign the Location ID as **household\_id**; for Number of Controls select **ncontrol.** Click next, choose the folder to save the newly generate .ctl file and then click **Import**.
* Finally, we need to import the location co-ordinates (SatScan\_locations). Set location ID to household\_id. Interestingly, the longitude and latitude are switched in the program, therefore set the X as “lat” and Y as “long”. Make sure the coordinates are set to Lat/Long

The input tab should look like this:



**Step 2 – Analysis tab**

Under the analysis tab, set the type of analysis to “Purely Spatial” and use the Bernoulli as the probability model. We will scan for areas with high rates. Click **Advanced**, and note that the maximum percent of the population that can be included in a single cluster is set with a maximum no greater than 50% by default.

**Step 3 – Output tab**

Under the Output files tab, specify the name of the file to which the results of the analysis should be written to. Ensure that both the KML file and Shapefile formats are selected for Geographical Output.

Under Column output files, make sure that dBase files are made for **Cluster Information** and **Location Information**. SatSCan will now produce a .gis file that contains a list of the locations in each cluster, and a .col file, that gives details of the size of each cluster and the relative risk associated with that cluster.

* Hit Execute on the main toolbar Untitled-3.jpg. Examine the results of the analysis.

Questions

**Q1:** How many significant clusters do you observe, and what is the relative risk (RR, ratio of the probability of the event occurring inside the cluster to occurring outside the cluster) associated with each one? Which households are in each cluster, and what is the radii of any significant clusters?

**Q2:** Is there much difference between the clusters you visually identified and the clusters identified by SaTScan?

**Q3:** What are some possible determinants for the observed spatial clustering?

**Step 4 - Visualising results**

SatScan exports a shapefile of the highest ranked clusters, which can be imported into R or a GIS program, and a kml file of any statistically significant clusters. View these files in Google Earth and R.