CA_Ch2: Exploratory spatial analysis of clustering of FMD in Myanmar villages

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

2.1 Background

"This section continues that from the earlier teaching material on SRA for incursion and spread of FMD in Myanmar. Some of the spatial files created in that teaching will be reused in this section.

The authors acknowledge the Livestock Breeding and Veterinary Department (LBVD), Myanmar for the use of the following data sets provided by the NZ MFAT-OIE SEACFMD project used in this section.

The datasets for Cluster Analysis exercises can be downloaded HERE"

CONTINUE

2.2 Case study- exploratory spatial data analysis to detect clusters of FMD in villages in Myanmar

2.2.1 Study data files

The original data consisted of two files located in "Geo-Features" subfolder:

1. The locations and administrative regions of FMD outbreak villages between March 2015 and Feb 2016, provided in the file “all_outbreaks_rev.csv”
2. The locations and administrative regions of all villages in Myanmar in the file “mmr_crty-level_pplp2_250k_mimu_Jan_2018_V0.csv”

2.2.2 Set up a new project in QGIS

“We set up a new project in QGIS (and later for use of SaTScan) to keep track of the many files we need to work with, just as we did with the GIS component of SRA. However, to reduce the burden of creating a new folder structure, and because we will need several files from

> **the SRA analysis, we will source the files in the previous data folders."**

- Create a new QGIS project as for MMR FMD Spatial Risk Assessment and save the QGIS project file in the same directory as you did previously (Review SRA_Chatpter 3)
- All of the raw data files for this analysis were already provided for the previous SRA exercise and are located in subfolders with the "RawData" folder
- The new results files for this project can be added to subfolders within the existing "ResData" folder

2.2.3 Visualise the spatial data

A first step in a spatial analysis is to map the data, so we start by mapping the village locations for Myanmar (mmr_crty-level_pplp2_250k_mimu_Jan_2018_V0) located in the "RawData\Geo-Features" folder:

- Click "Open Data Source Manager" on Tool bar Menu -> Dialog box … (Figure 2.1)

1. Select "Delimited Text"
2. In file name select the above file
3. Check Geometry Definition is correct

4. Click “Add”

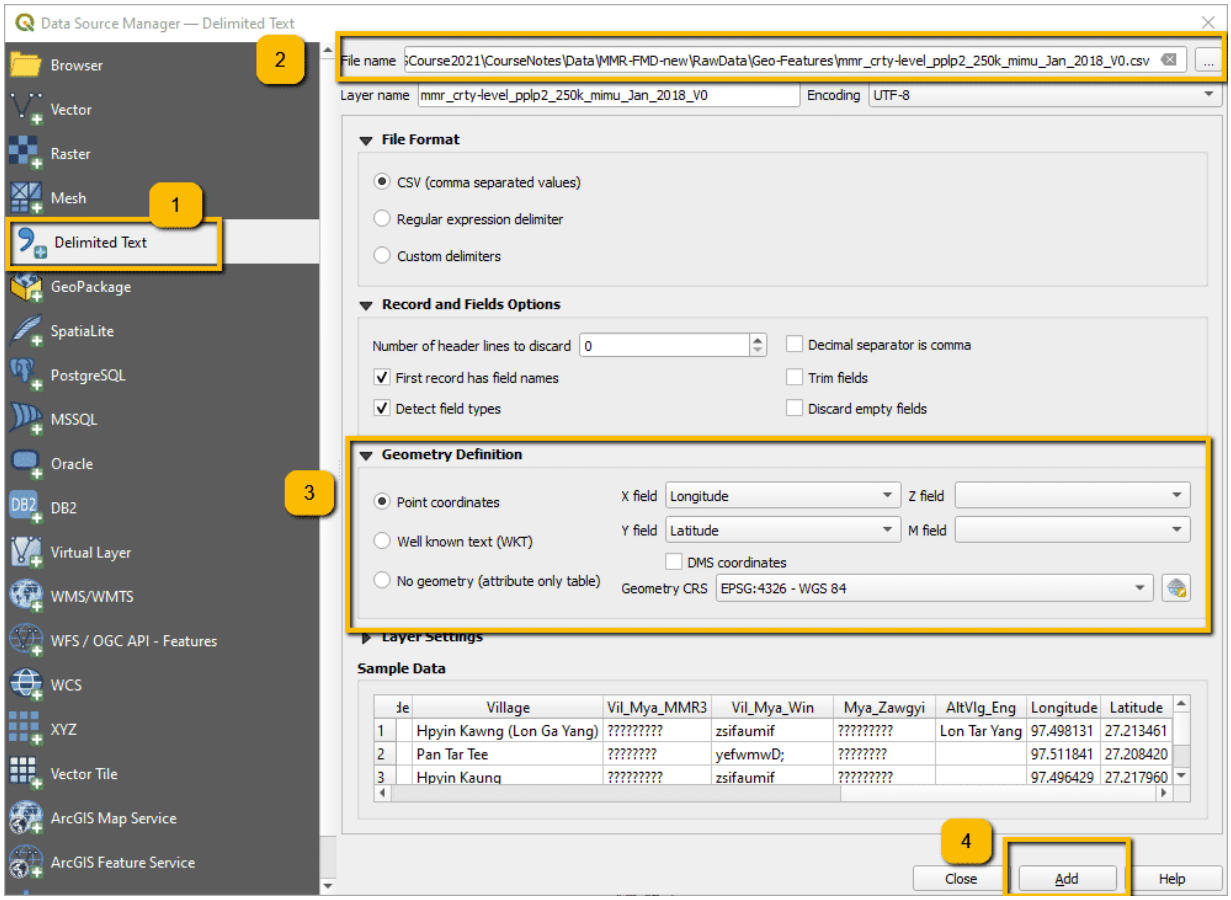


Figure 2.1: Create spatial file of Myanmar village locations

CONTINUE

We then project and save the file to the Project CRS:

- Right click mmr_crty-level_pplp2_250k_mimu_Jan_2018_V0 in Layers Pane -> Export -> Save Features As … -> Dialog box (Figure 2.2)

1. Click folder button to far right of File name and Select “ResData-Features” and name file “MMRVillages”
2. In CRS combo box select “Project CRS:32646 - WGS 84 / UTM zone 46N”
3. Click “OK”

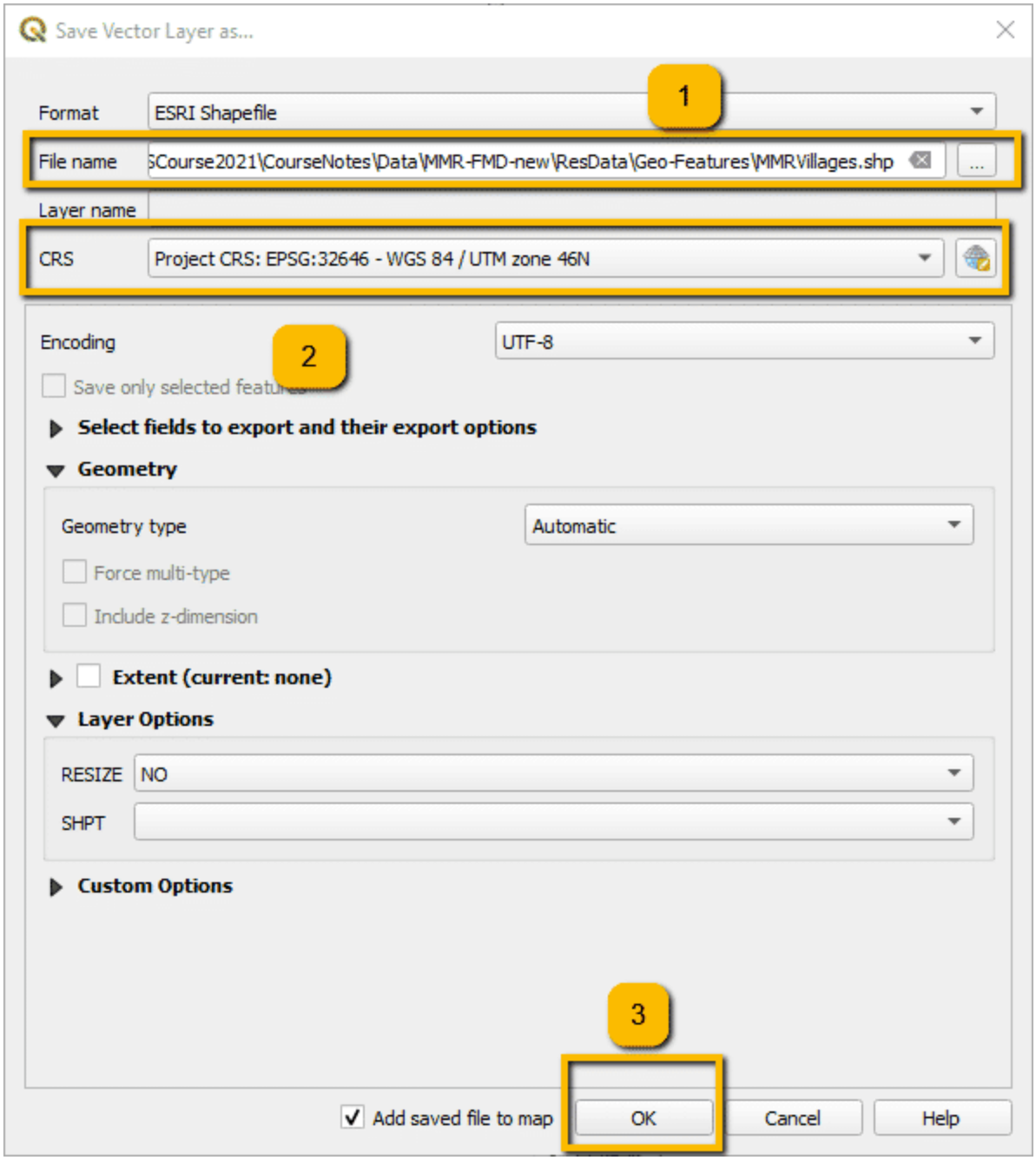


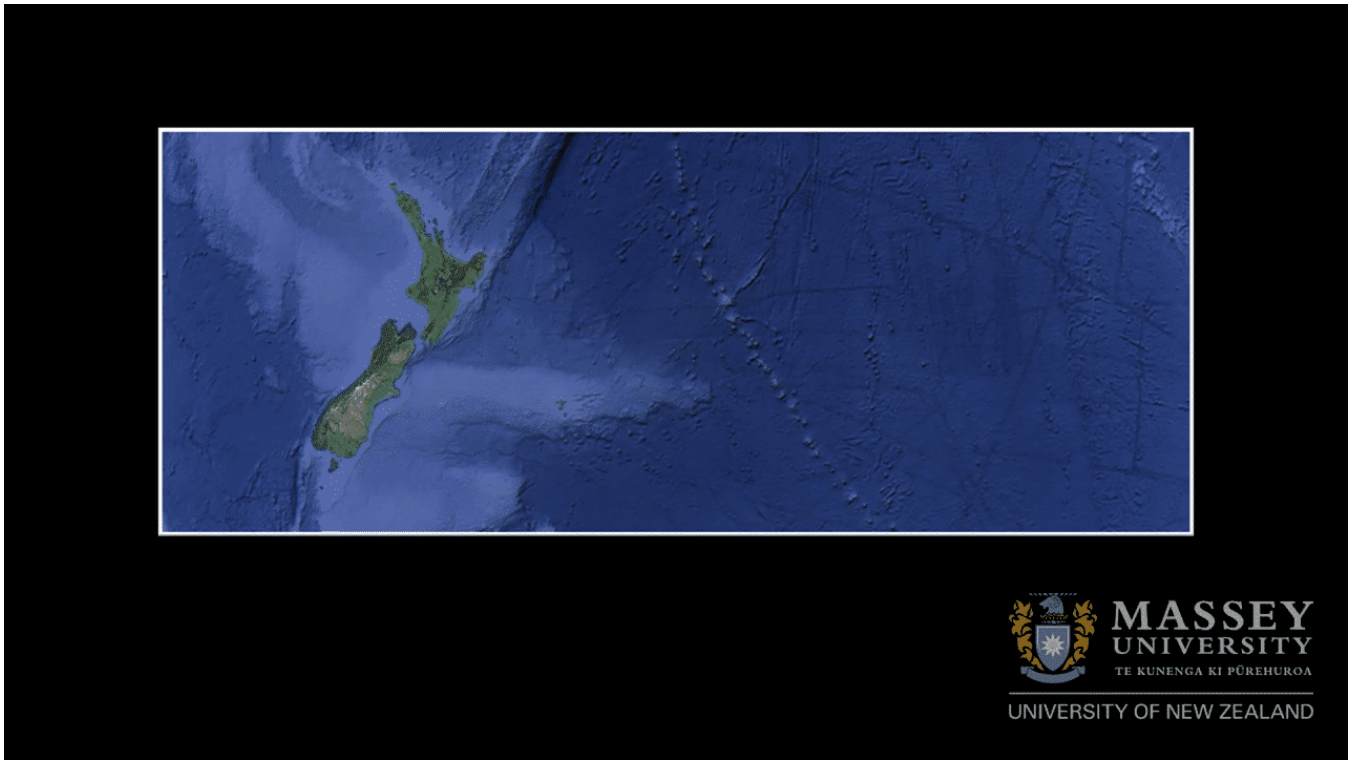
Figure 2.2: Save projected spatial file of Myanmar village locations

CONTINUE

Add map layer of FMD case village locations:

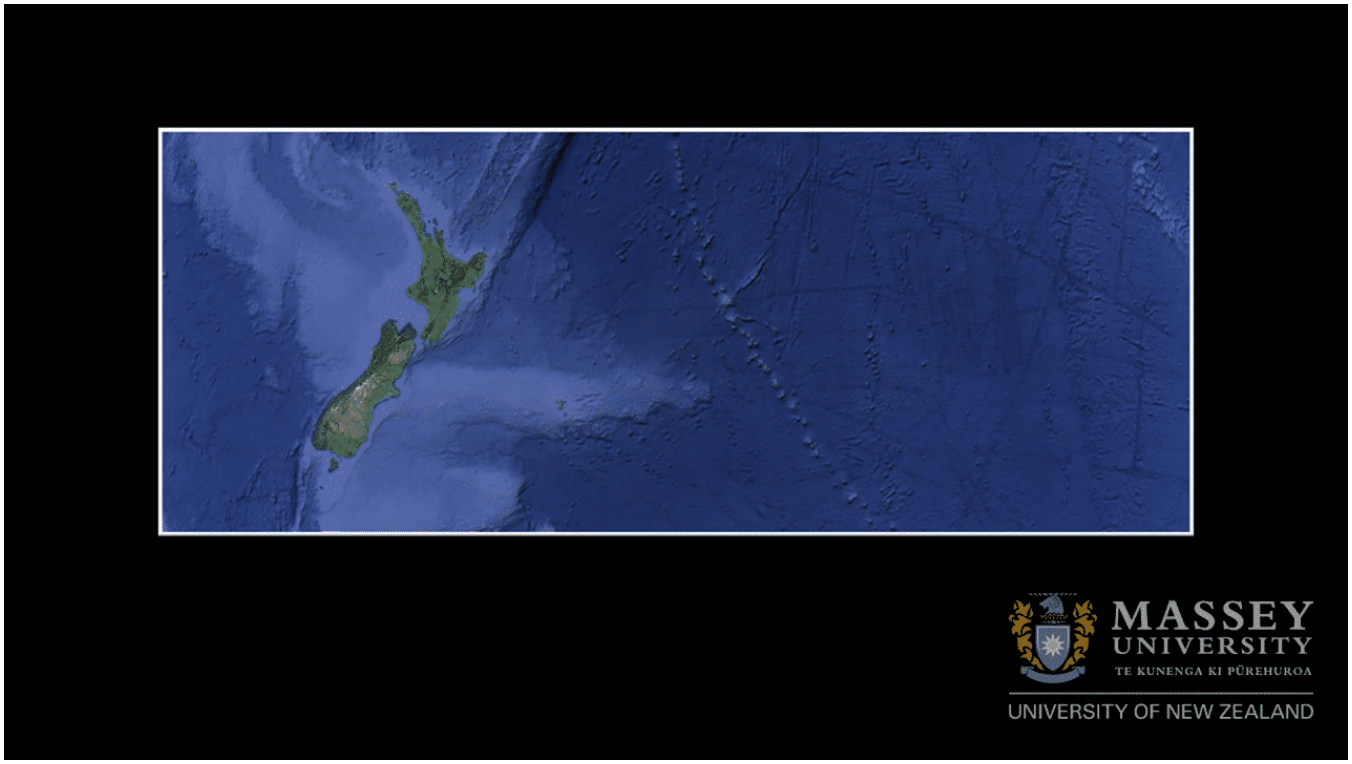
- Drag file “all_outbreaks_rev.shp” from “ResData-Features” folder created in the previous SRA project onto the Map palette

Click ▶ to play the video



CONTINUE

Click ▶ to play the video



CONTINUE

7.2.4 Create a kernel-smoothed density map (heatmap) of FMD outbreak villages

The Heatmap (Kernel Density Estimation)" function in QGIS works on only 1 set of points at a time. In our example, we are interested in both the outbreak villages, but also in the background population of all villages. Each data set

can only be smoothed into one heatmap, but it is of course possible to create different layers of spatial objects and view them together

- Highlight “all_outbreaks_rev” in the Layers Pane
- In Processing Toolbox search bar type “kernel density” and select “Interpolation -> double click”Heatmap (Kernel Density Estimation) -> Dialog box (Figure 2.3)

1. Set radius to 100000 (100 km) to begin with (this setting is subjective only)
2. Set number of rows in output to 200 (the other boxes self-fill)
3. Click “Run”
4. Click “Close”

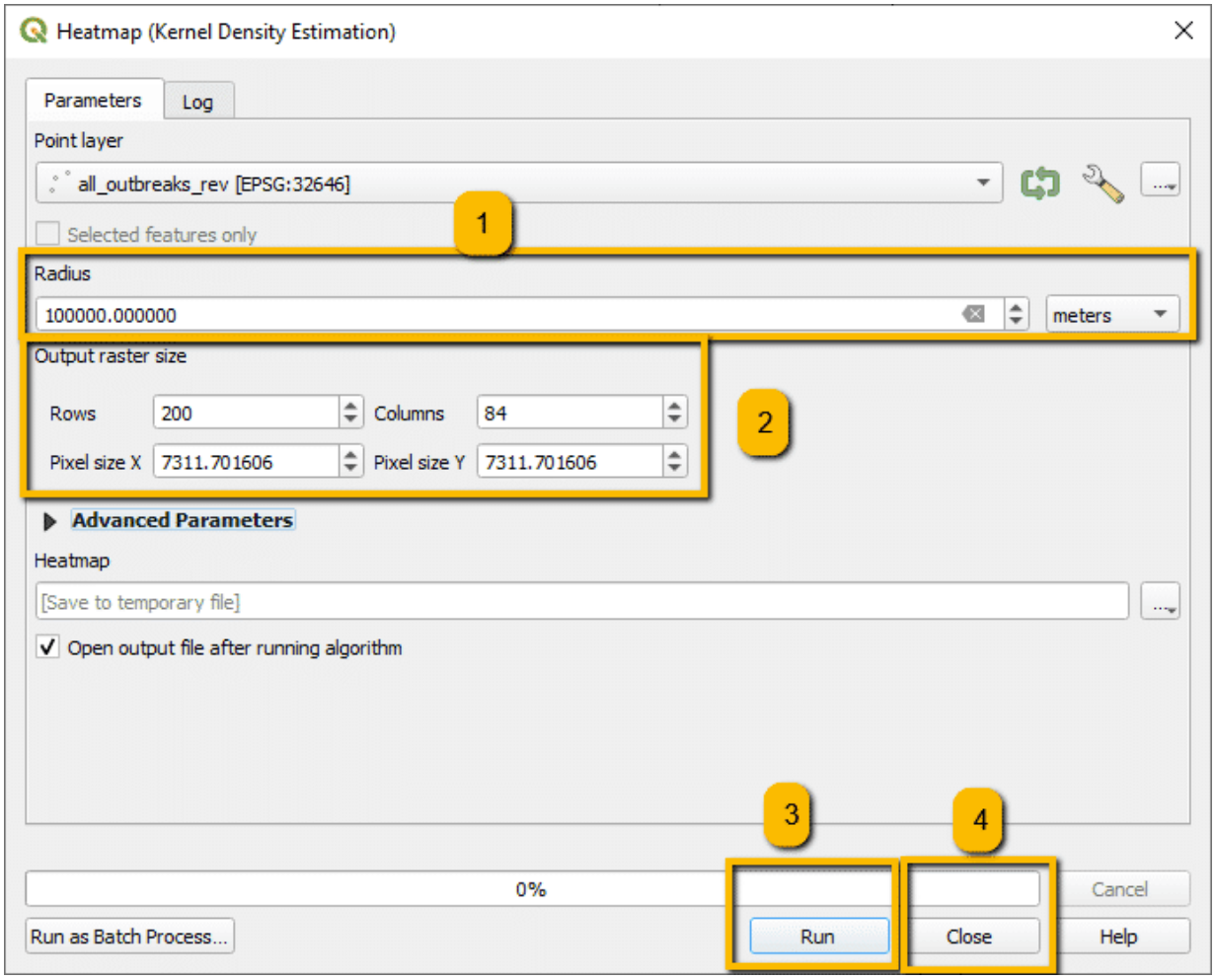


Figure 2.3: Create heat map of FMD outbreak villages in Myanmar

CONTINUE

Save your heatmap as raster layer:

- Right-click “Heatmap” in the Layers Pane -> Export -> Save As … -> Dialog box (Figure 2.4)

1. Select folder “-Features” and create a new name for the file “Outbreaks_Heatmap” in the GeoTIFF format
2. Check that the CRS is correct for the project and set the extent to that of the study region (MMR_0)
3. Click OK

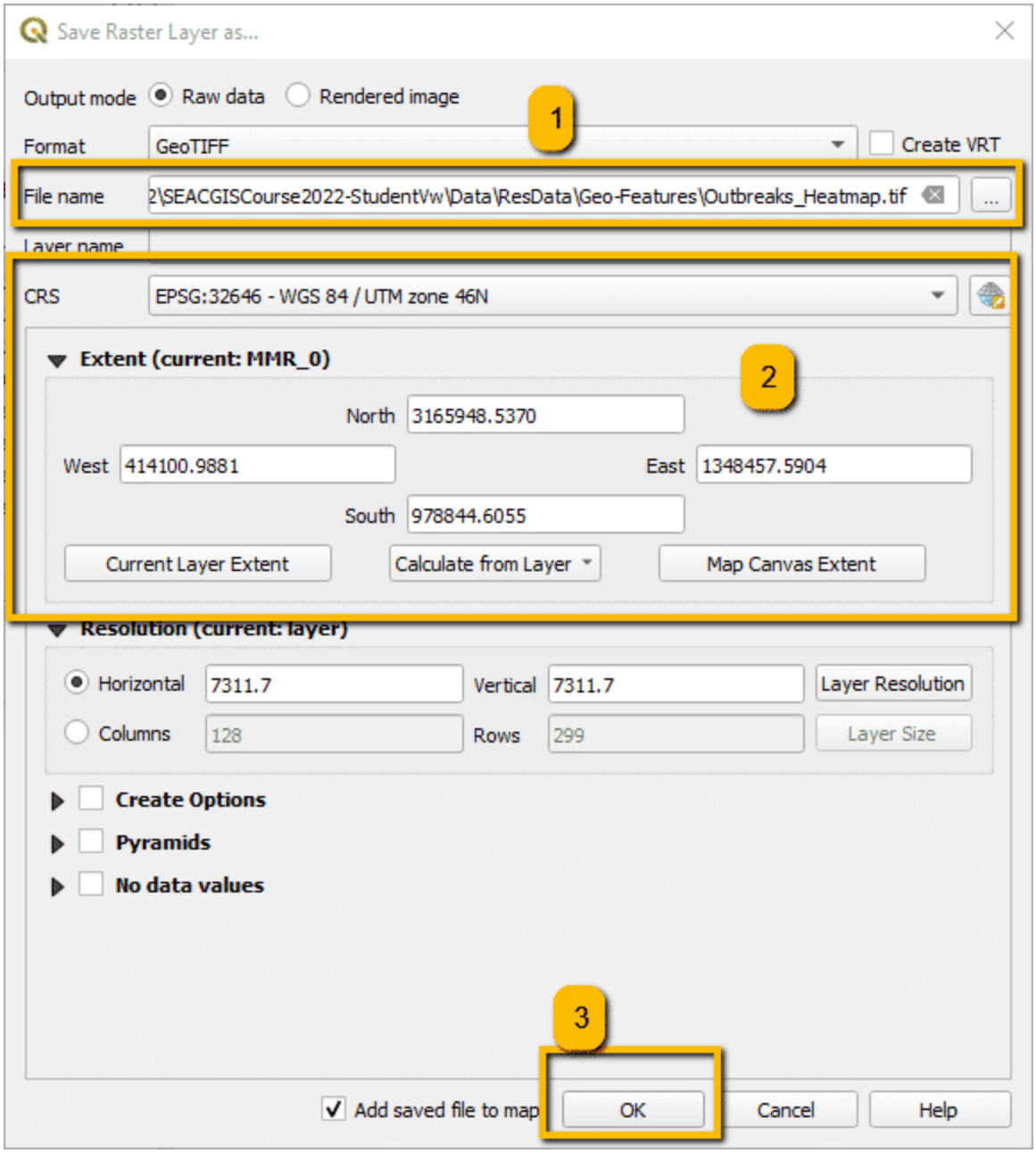


Figure 2.4: Save heat map of FMD outbreak villages in Myanmar as a raster file

Remove “Heatmap” from Layers Pane (it is a temporary file only and is not saved when a session is closed)

CONTINUE

Edit the default map properties to better visualise the density of cases:

- Highlight the newly-created file in the Layers Pane “Heatmap”
- Right click -> Properties -> Dialog box … (Figure 2.5)

1. Select “Symbology” tab
2. In “Render type” select “Singleband pseudocolor”
3. Select a “Color ramp” option (orange to red is a reasonable choice)
4. Click “OK”

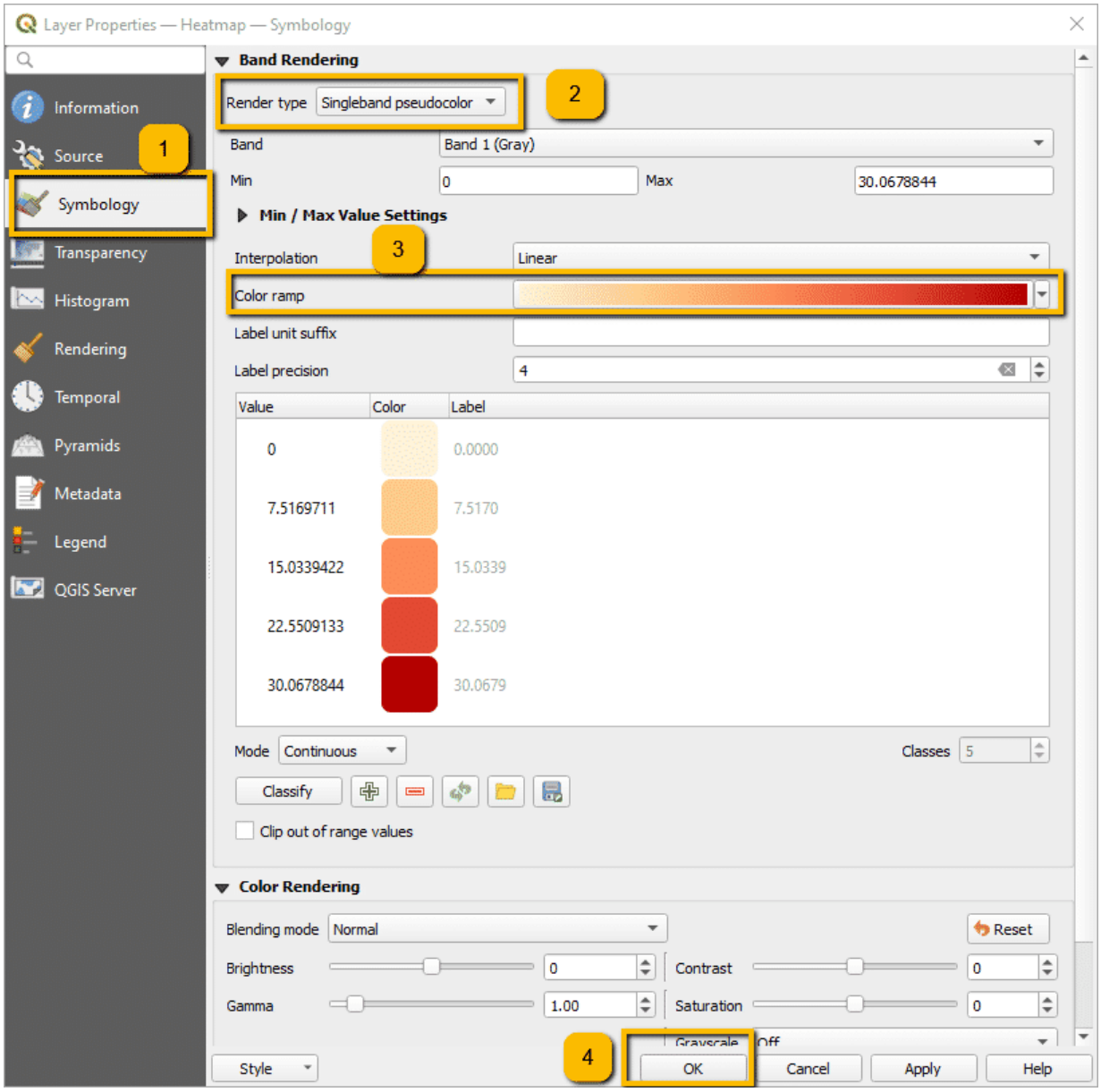
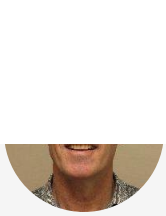


Figure 2.5: Edit symbology of heat map of FMD outbreak villages in Myanmar

CONTINUE

"Experiment with different radius dimensions to visualise



how the heat map and its interpretation varies. A heatmap of the density of outbreak villages plotted alongside their locations and the background population of all villages might look something like Figure 2.6).”

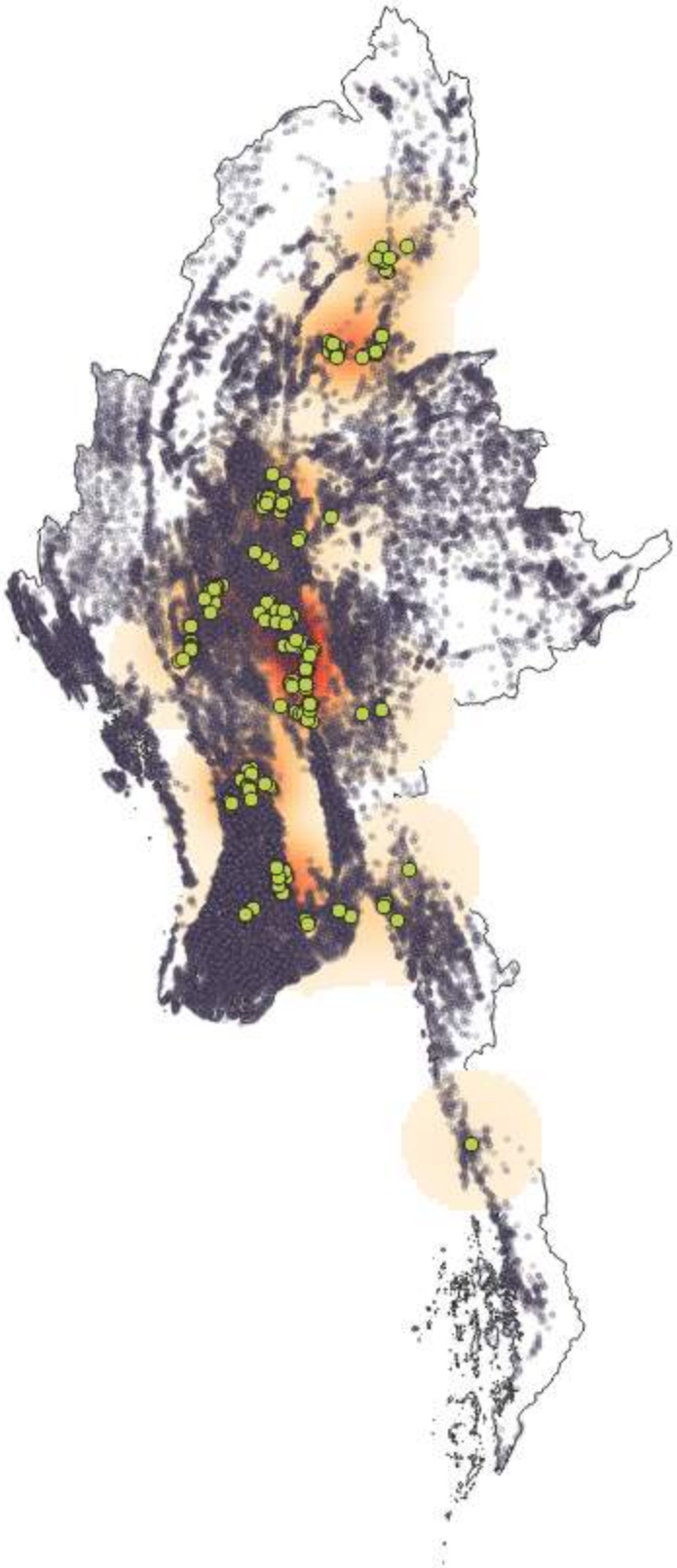
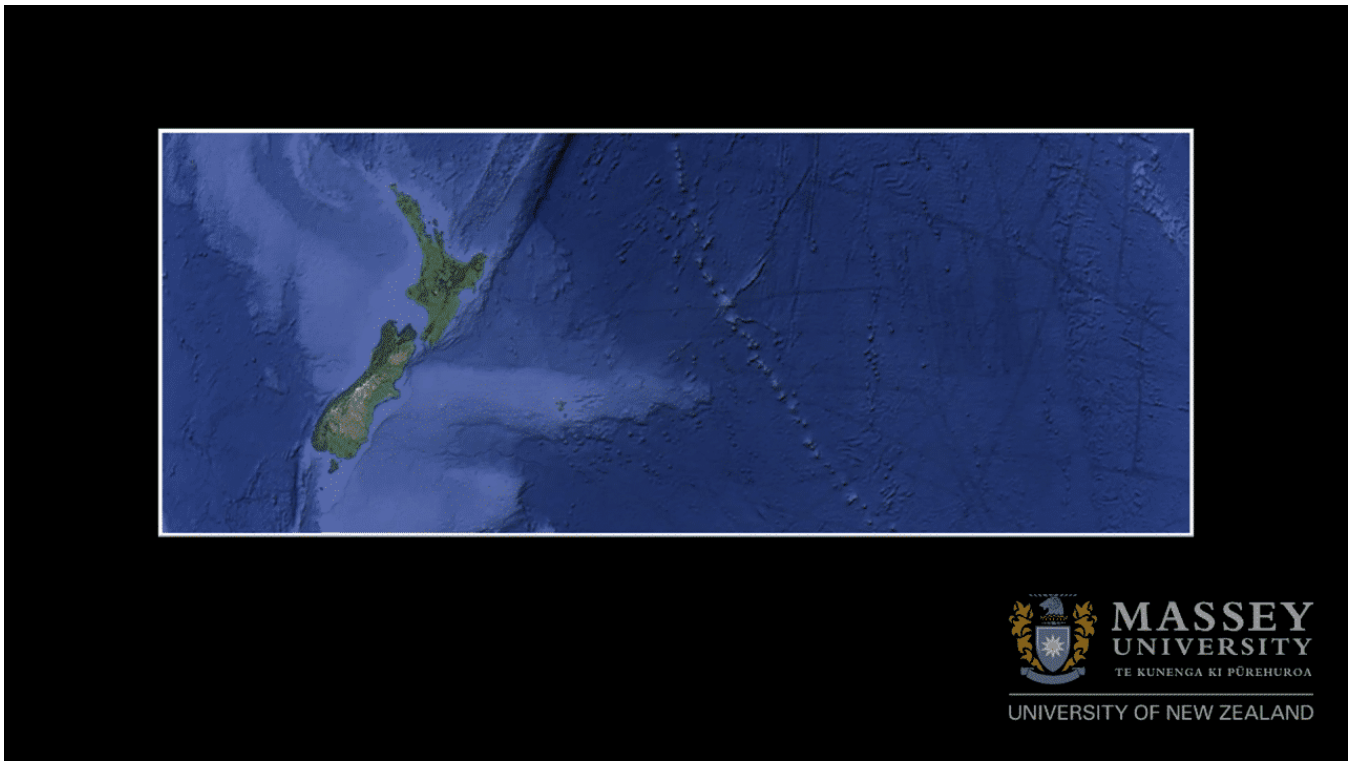


Figure 2.6: Heat map (red shading indicates areas of increased density) of FMD outbreak villages (green points) against population of all villages (black points) in Myanmar

CONTINUE

Click ▶ to play the video



CONTINUE

Exercise 2.1: Exploratory spatial data analysis of villages with FMD outbreaks in Myanmar

"**Exploratory spatial data analysis begins with mapping and describing the raw data.**

1. **Describe in words the density of village outbreaks only within the study region in terms of both large and small scale patterns**
2. **What additional information is needed to adequately explain this pattern?**
3. **Does it appear that the density of outbreak villages is approximately the same as that for all outbreak and non-outbreak villages combined?**
4. **Could the pattern of density of FMD outbreaks just be reflecting the underlying density of the population of villages at risk of an outbreak, or could other factors be responsible?**

CONTINUE

Have you answered all the questions in Exercise 2.1?

- ◯ Yes
- ◯ No

SUBMIT

CONTINUE

Answer keys

1. The large scale distribution of outbreak villages shows them mainly in the central areas of Myanmar with few towards the extremities of the compass points. At a smaller scale, in the areas where outbreaks were recorded, the outbreak villages appear to cluster together in groups.

2. We need to understand the density of villages to determine whether the density of outbreaks merely represents

the density of the population of villages at risk of an outbreak, or whether there are true clusters of outbreaks due to local risk factors after accounting for the underlying population at risk

3. No. There appear to be fewer outbreaks in the high-density areas of villages, such as the south delta regions.

4. It is likely that factors other than the village density are associated with FMD outbreaks in villages"

CONTINUE

2.3 Localised non-focused cluster detection

"A formal statistical test is a useful additional step to detect spatial clusters of disease because it removes some of the issues about subjective assessment.

Spatial clustering may be investigated in three different dimensions (linear, point and area) and from a range of different study and data types, for example case-control or surveillance studies with case and population count data, and dichotomous, categorical, rank or continuous data types. Individual statistical tests were primarily developed to be used with one data type, but it is possible to aggregate point to areal data, and possibly areal to point data (by using areal centroids). However, the scanning methods for point data are less suitable for areal data as sub-regions may not neatly fall within the scanning circle."

2.3.1 Kulldorff's spatial scan test

- Spatial data type: Areal or point

- Data needed: Polygons of areal units, counts of cases and either controls or population at risk, or dichotomous infection status of study units e.g. farms or villages
- How the test works:
    - A theoretical circular window is placed on a map of all study locations, for example the coordinates of the villages or centroids of administrative regions
    - A scanning window of increasing radius is placed around one of many possible centroids by sequentially aggregating the nearest neighbour areas to create zones
    - The window radius may vary to a defined upper limit (no more than 50% of study population is recommended)
    - For each window the risk of disease is compared with that of the study area outside the window
    - If using case-control data, controls should be selected from same source population as the cases
    - Significance testing is estimated by Monte Carlo sampling
    - The disease data may be either Bernoulli (zero for cases and one for controls) or Poisson (the number of cases and the population at risk)
    - The test adjusts for the heterogeneity of the population at risk by indirect adjustment to calculate the expected number of cases for each location
    - This test may be used as complement to a global clustering test
    - The test can be used to detect clusters with increased, decreased or both increased and decreased incidence of disease
    - The test reports the most significant primary and secondary clusters

    - A case study on the use of the spatial scan test is described in Section 7.4

- References: Kulldorff and Nagarwalla (1995), Kulldorff (1997)

7.3.2 Temporal scan test

- Data needed: Count of cases by time
- How the test works:
    - Originally proposed by Naus (1966) for use in stable population and analogous to spatial scan statistic
    - The test statistic is the maximum number of cases in a predefined "window" of time found by scanning all time series of that interval in the study
    - The test can be generalised to account for temporal trends in the population size and incorporated in the SaTScan software
    - The test is most sensitive when the the scanning window is a similar interval as the duration of the clusters
    - It is recommended to set the scanning window on basis of known disease patterns, but the subjectivity of this setting can affect test results

- References: Kulldorff (2005)

CONTINUE

2.4 Use of SaTScan software to investigate spatial clusters of FMD in Myanmar

2.4.1 Background

"We set out below steps for a statistical analysis for spatial clusters of FMD in Mandalay and Nay Phi Taw States of Myanmar in the 2015-2016 period using the open-source software "SaTScan". These data are a subset of the national data used in the previous exercises.

The data for this case study of the use of SaTScan were sub-setted so as to include only data in the Mandalay and Nay Pyi Taw state regions. The records of outbreaks in these two state regions are provided for you in the file "MndlyNPTVillStatus.csv" and the variables are described in Table 2.1."

Variable name	Description

Variable name	Description
Vill_Pcode	Unique village post code
State_Regi	State or Region name
Latitude	Latitude (decimal degrees)
Longitude	Longitude (decimal degrees)
Detect_date	Date of first detection of FMD in village
NCase	Indicator variable for case village
NCon	Indicator variable for control village

CONTINUE

2.4.2 Analysis steps

2.4.2.1 Software and file management

- Download and install SaTScan software on your PC
    - SaTScan software is available from https://www.satscan.org/

    - You will need to register yourself to install the software, but this process is free

- Google Earth Pro from https://www.google.com/earth/download/gep/agree.html?hl=en-GB
    - This application will allow you to visualise the locations and geographic features of clusters identified by SaTScan using ‘Google Earth’

- Folder and data management
    - Create the following folder and subfolder structure at the level below your course project folder
        - SaTScan
            - BernoulliModel
                - InputFiles
                - OutputFiles

- Copy the following study file into the “InputFiles” folder:
    - “MndlyNPTVillStatus.csv”

CONTINUE

2.4.2.2 Import the data into the SaTScan application

- Launch the SaTScan software
- Select “Create New Session” from the Start Window -> Opens tabbed dialog box
- You need to create the three files needed for this analysis (Figure 2.7) from the NPTVillStatus.csv file using the Import File Wizard of the application
- The three files will be linked through a location ID, which represents the name or code for a geographical entity such as a state, province, county, zip code, postal code area, census tract or dwelling
- In the Mandalay Province data, the unique location variable used is “Vill_Pcode”, with each village represented by a set of 6 integers
    - The three files you will create are:
        - A Case File
        - A Control File
        - A Coordinates File

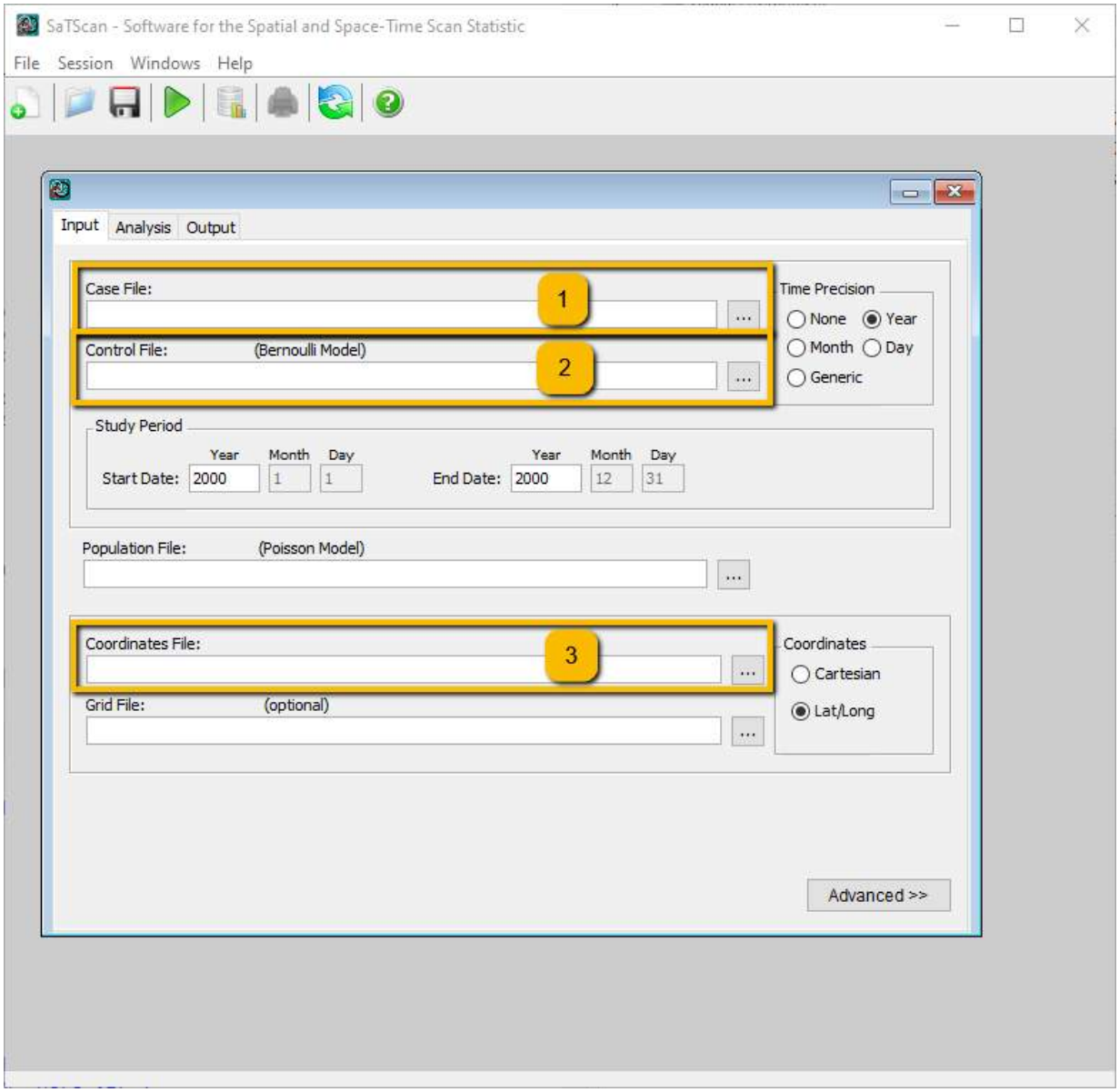


Figure 2.7: Dialog box on opening screen of SaTScan software

Case File:

- This file will contain data on the case villages and have the filename extension “*.cas”
- Click button on right of Case File combo box to open the Import File Wizard -> Click Next at bottom of window:

- Select Case File -> Dialog box opens
- Locate “MndlyNPTVillStatus.csv” and double-click to add it to the File name box, click Open
- Import File Wizard opens
- Select “Next” -> Dialog box opens: (Figure 2.8)

1. Tick “First row is a column name”
2. Select Field Separator: Comma
3. Group Indicator: Double quotes
4. Click Next

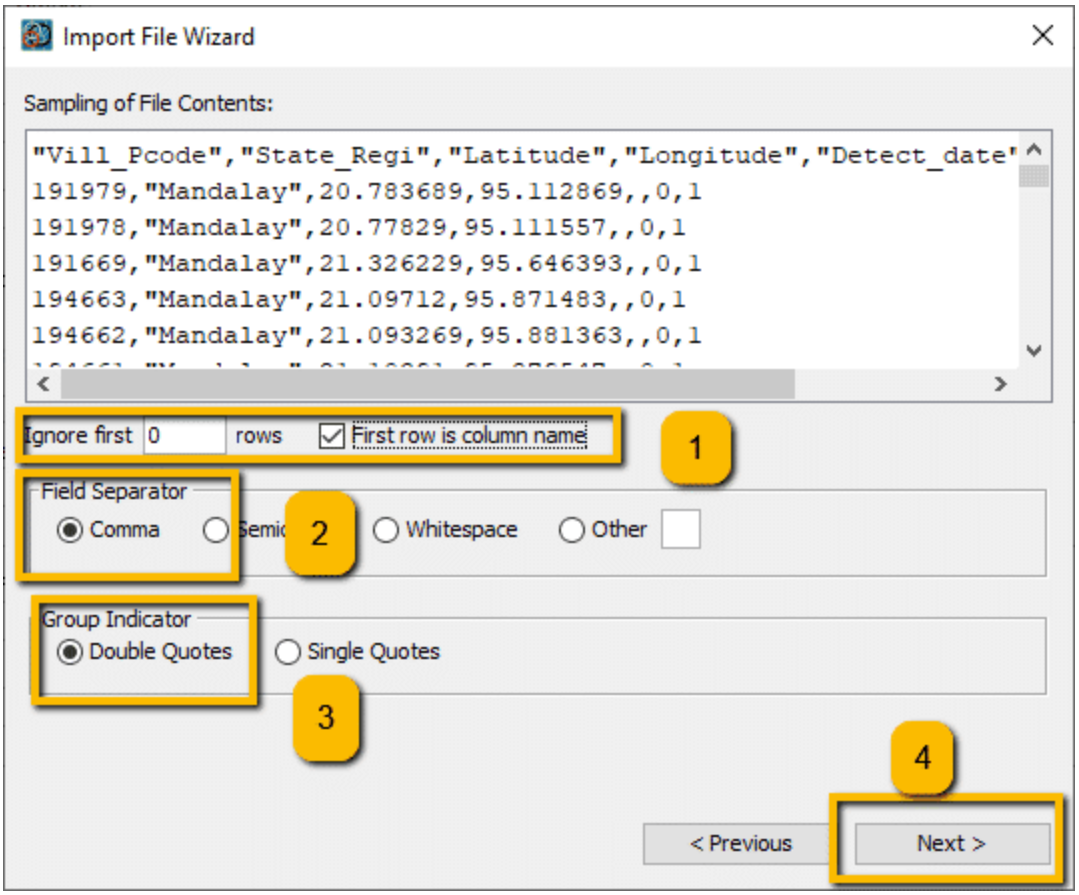


Figure 2.8: Dialog box of Import File Wizard in SaTScan software to define format of study file from which importing case data

CONTINUE

- Define model and variables (Figure 2.9)

1. Display SaTScan variables for: “Bernoulli model
2. Location ID: “Vill_Pcode”
3. Number of cases: “NCase”
4. Click “Next”

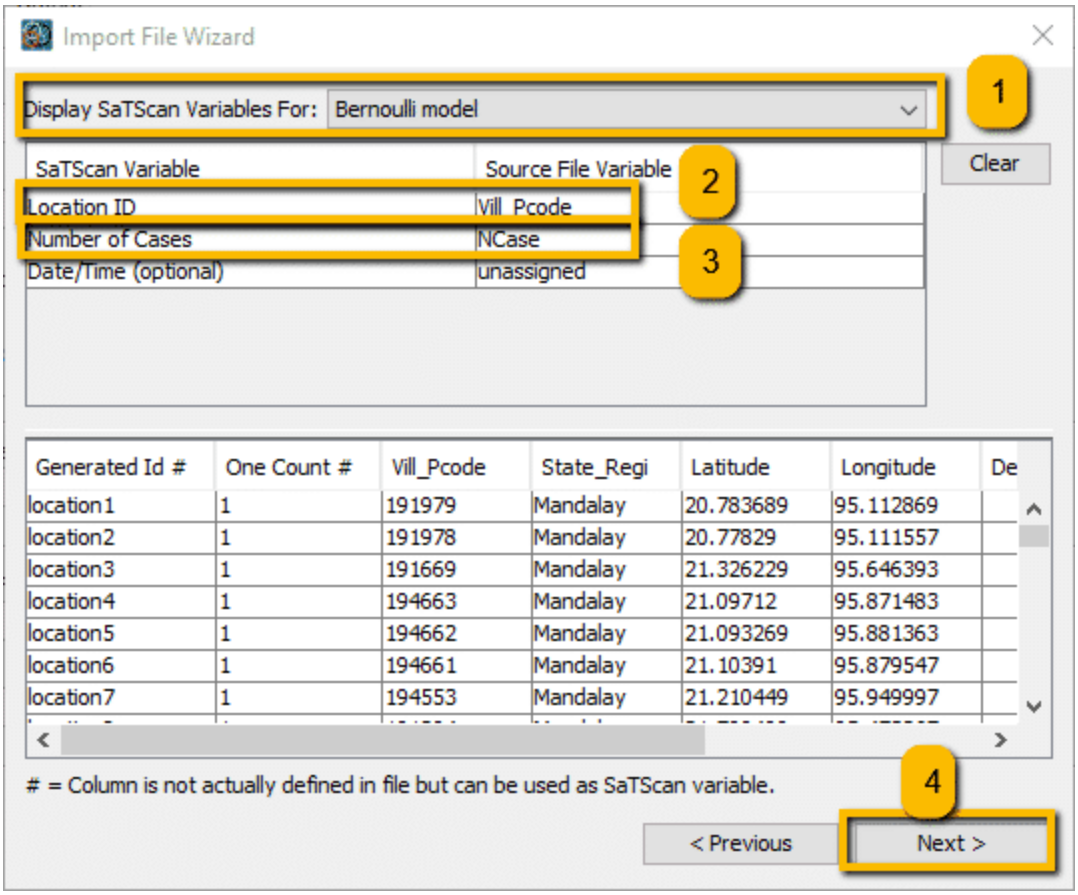


Figure 2.9: Dialog box of Import File Wizard in SaTScan software that defines variables in study file from which importing case data

- Select directory and create file name to save imported file (Figure 2.10)

1. Save imported input file as: “Data-FMD-new.cas”
2. Click “Import” to complete the step

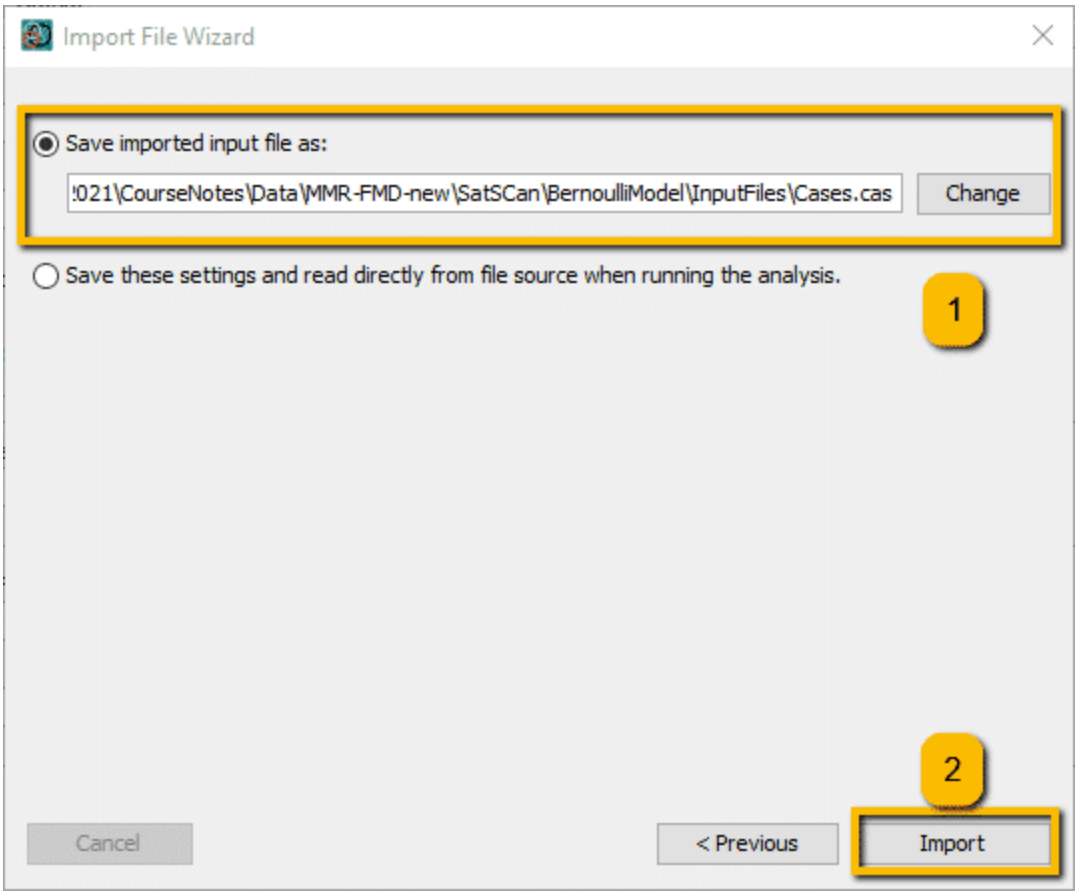


Figure 2.10: Dialog box of Import File Wizard in SaTScan software to name and save case file

CONTINUE

Control File:

- This file contains data on the control (non-case) villages and will have a *.ctl" filename extension
- Click button on right of Control File combo box to open the Import File Wizard -> Click Next at bottom of window :
- Select Case File Dialog box opens

- Locate “MndlyNPTVillStatus.csv” and double-click to add it to the File name box, click Open
- Import File Wizard opens
- Select “Next” (Figure 2.11)

1. Tick “First row is a column name”
2. Select Field Separator: Comma
3. Group Indicator: Double quotes
4. Click Next

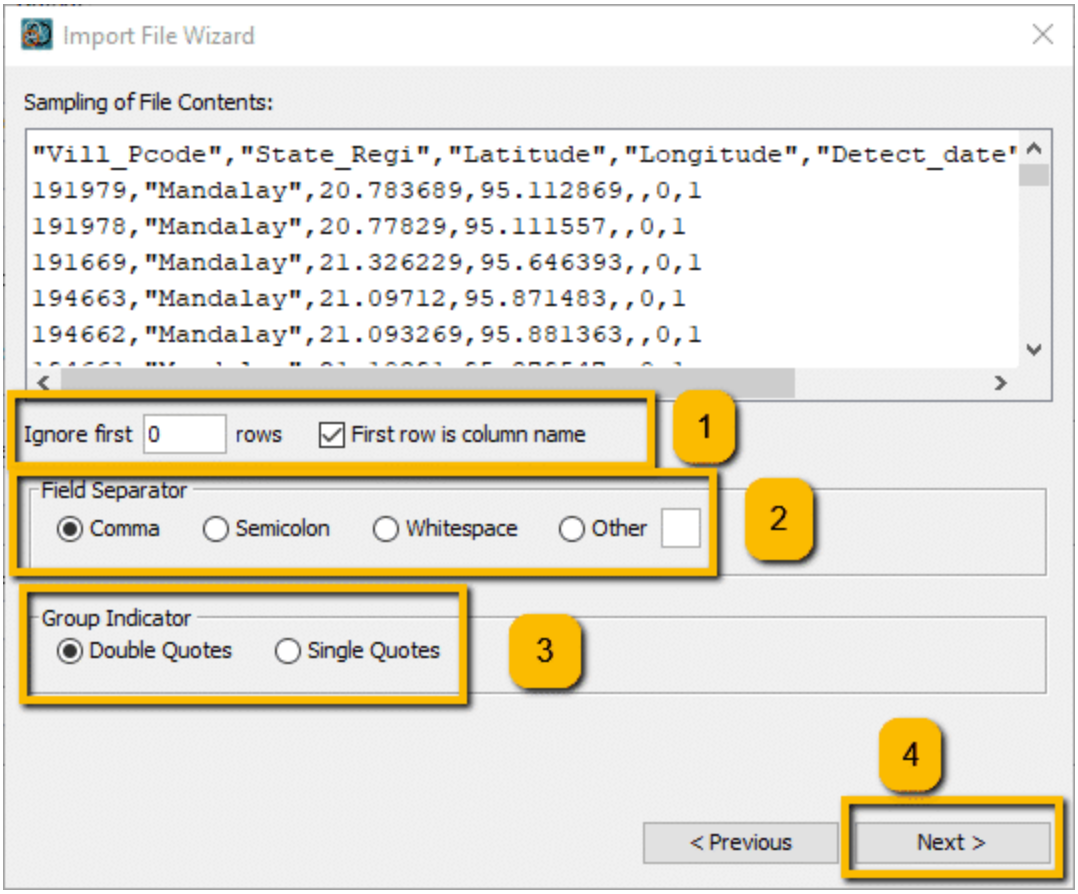


Figure 2.11: Dialog box of Import File Wizard in SaTScan software to define format of study file from which importing control data

CONTINUE

Define model and variables (Figure 7.12)

1. Location ID: “Vill_Pcode”
2. Number of cases: “NCon”
3. Click “Next”

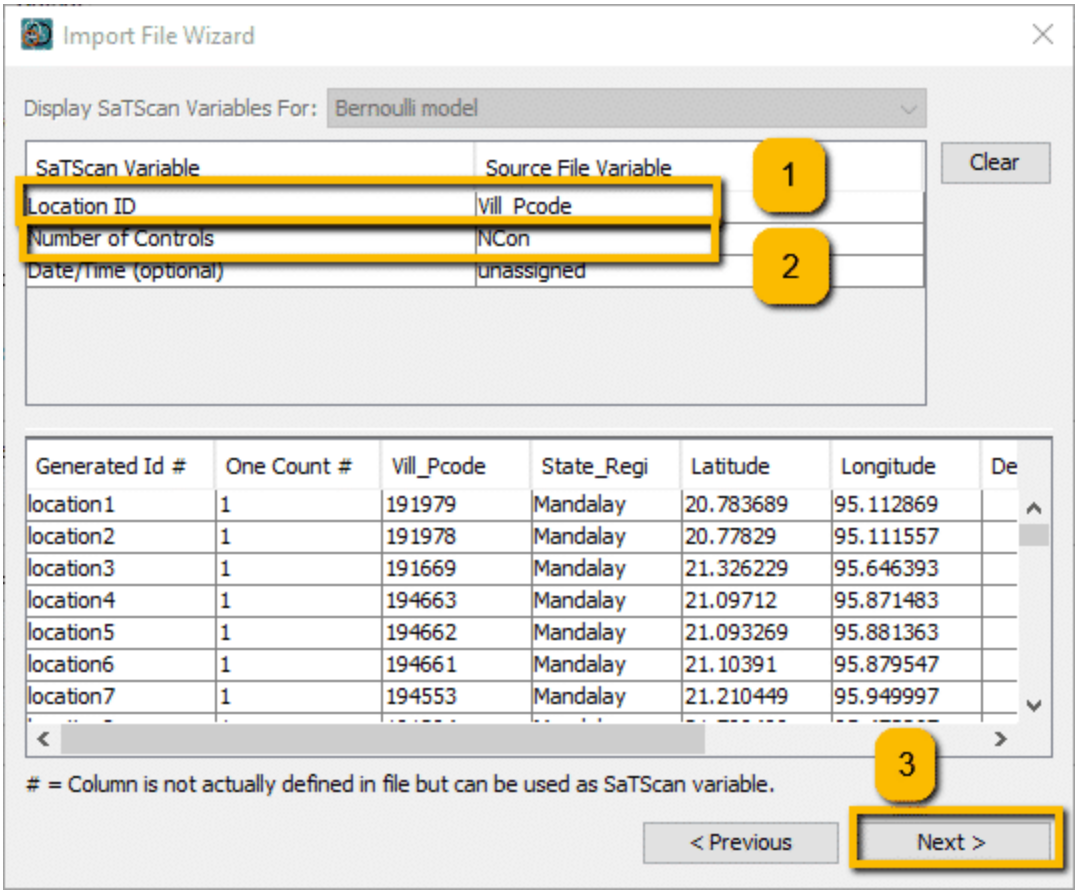


Figure 2.12: Dialog box to of Import File Wizard in SaTScan software to name and save controls file

- Select directory and create file name to save imported file (Figure 2.13)

1. Save imported input file as: “Data-FMD-new.ctl”
2. Click “Import” to complete the step

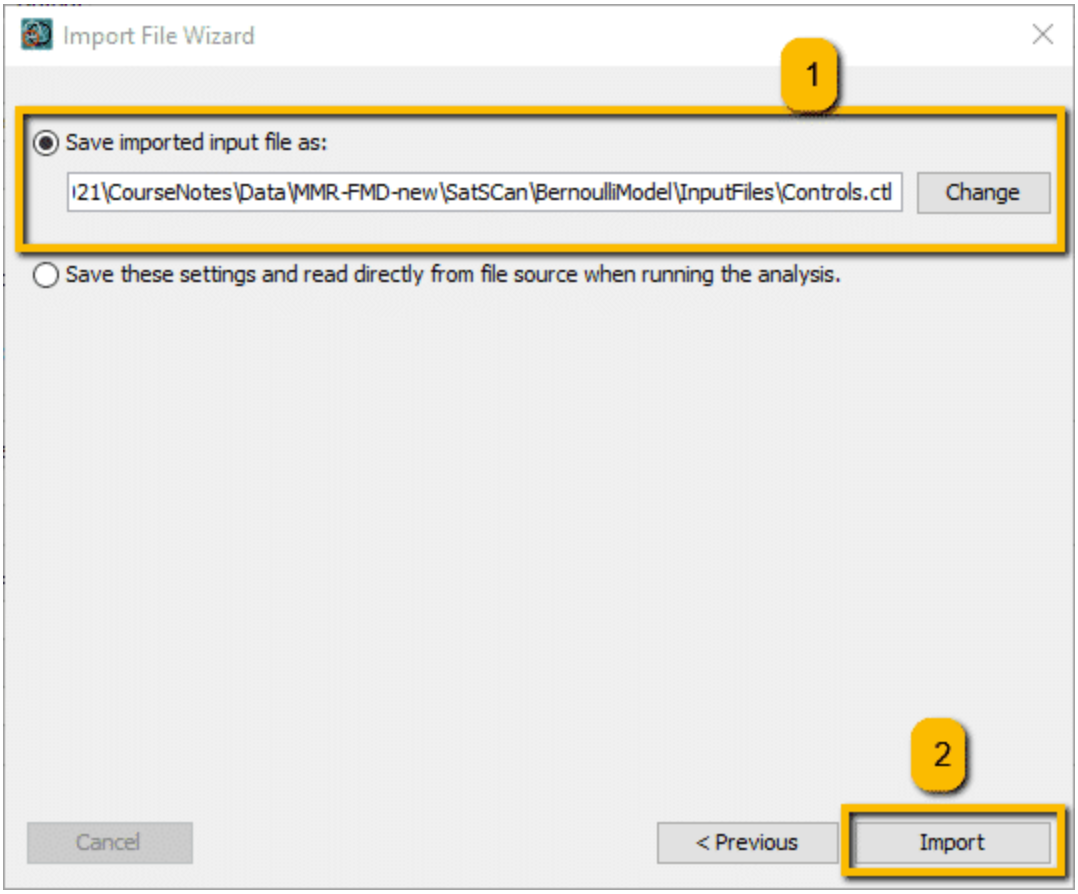


Figure 2.13: Dialog box to of Import File Wizard in SaTScan software to name and save controls file

CONTINUE

Coordinates File:

- Click button on right of Coordinates File combo box to open the Import File Wizard -> Click Next at bottom of window :
- Select Coordinates File
- Dialog box opens
- Locate “MndlyNPTVillStatus.csv” and double-click to add it to the File name box, click Open
- Import File Wizard opens
- Select “Next” -> dialog box (Figure 2.14)

1. Tick “First row is a column name”
2. Select Field Separator: Comma
3. Group Indicator: Double quotes
4. Click Next

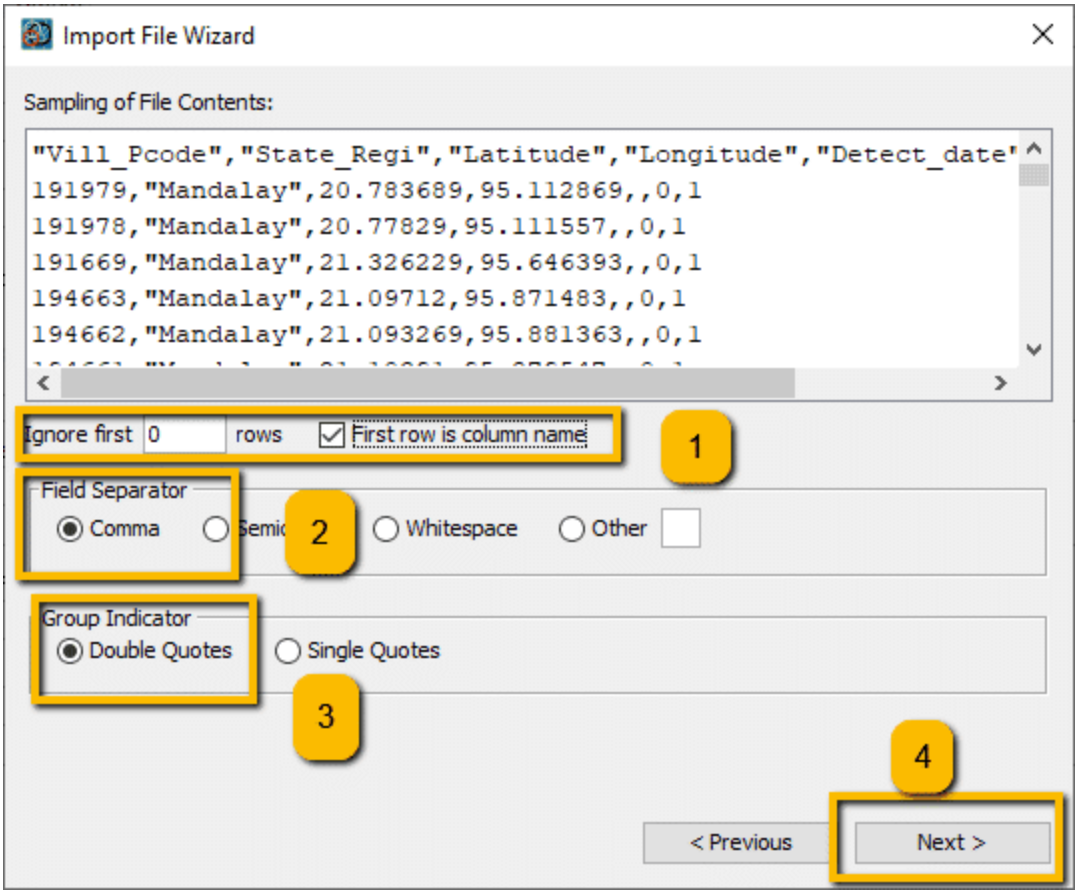


Figure 2.14: Dialog box to of Import File Wizard in SaTScan software to format of study file from which importing coordinates data

CONTINUE

- Import File Wizard Coordinates options (Figure 2.15)

1. Display SaTScan Variables for: Select Latitude/Longitude Coordinates
2. Location ID: Select “Vill_Pcode”
3. Latitude (y-axis): Select “Latitude”

4. Longitude (x-axis): Select “Longitude”

5. Click “Next”

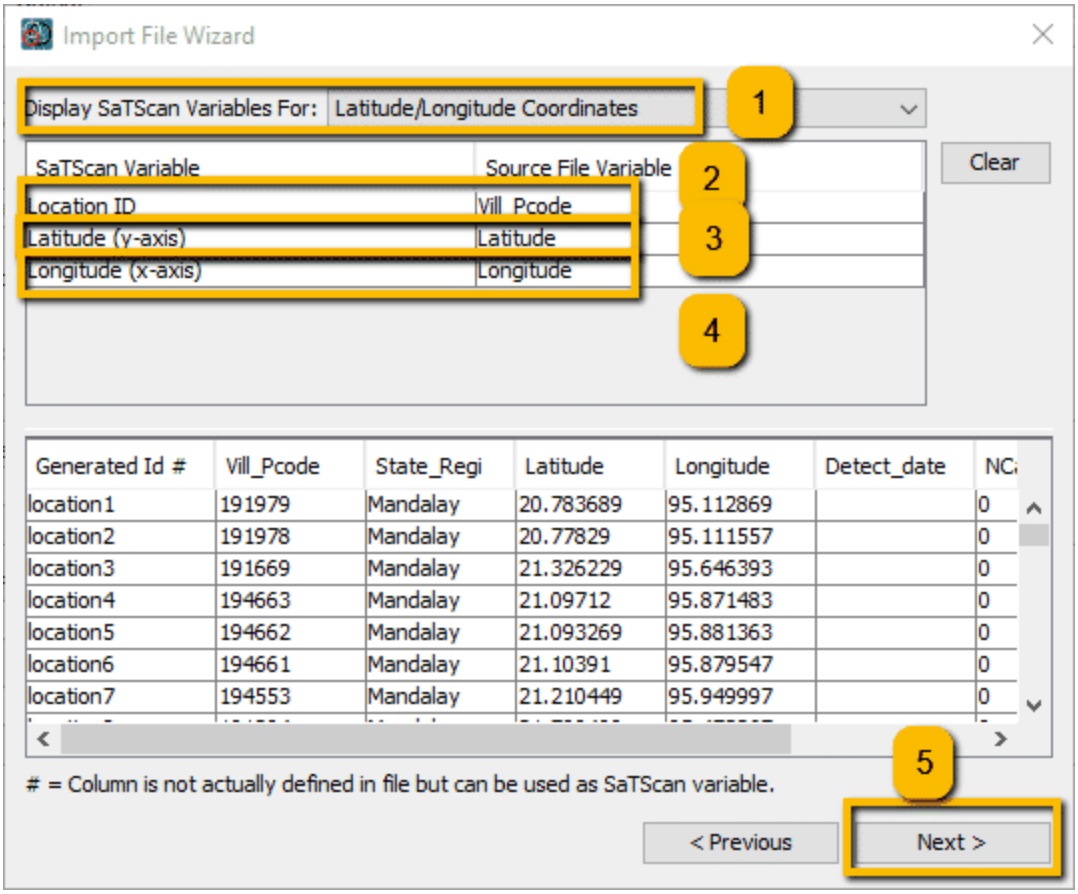


Figure 2.15: Dialog box of Import File Wizard in SaTScan software that defines variables in study file from which importing coordinates data

- Import File Wizard Coordinates combo box
- Select directory and create file name to save imported file (Figure 7.16)

1. Save imported input file as: “Data-FMD-new.geo”

2. Click “Import” to complete the step

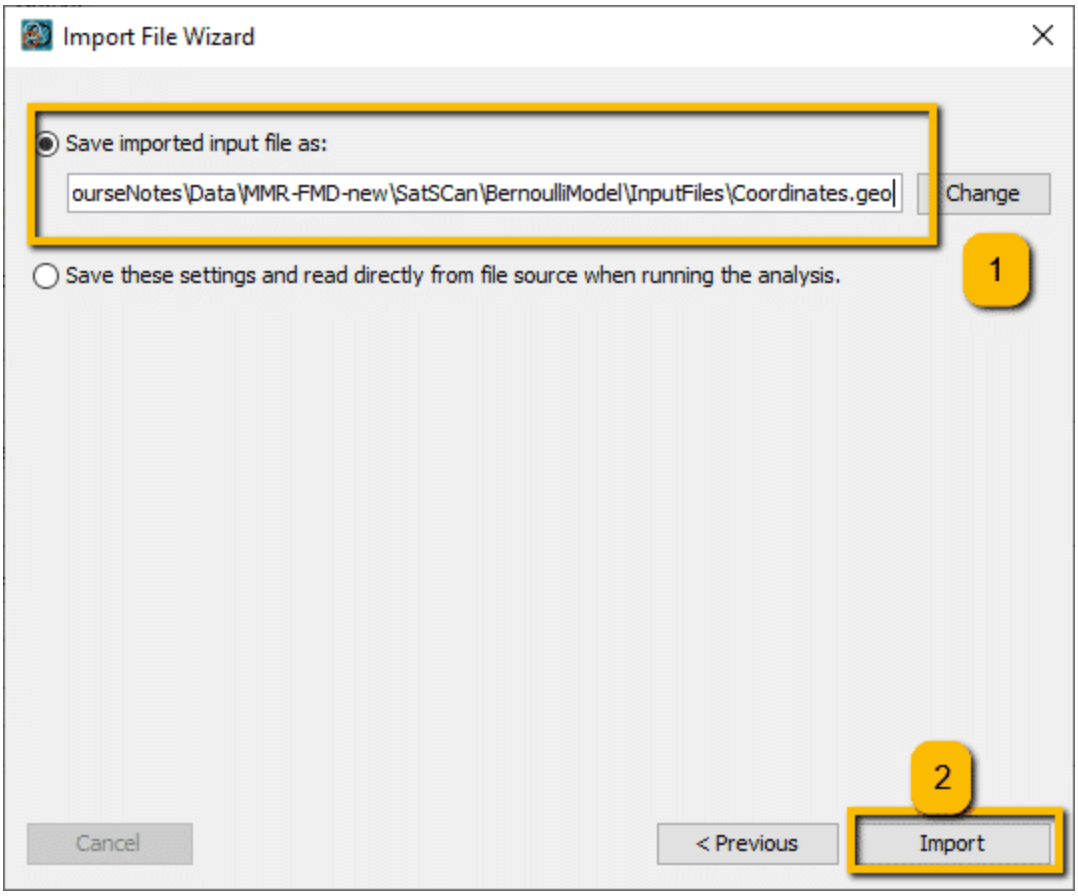


Figure 2.16: Dialog box to of Import File Wizard in SaTScan software to name and save coordinates file

CONTINUE

2.4.2.3 Set parameters and run the analysis

Select “Analysis” tab (Figure 2.17)

1. In “Type of Analysis” - “Retrospective Analyses” Select “Purely Spatial”

2. In “Probability Model” select: “Bernoulli”

3. In “Scan For Areas With”, select “High Rates”

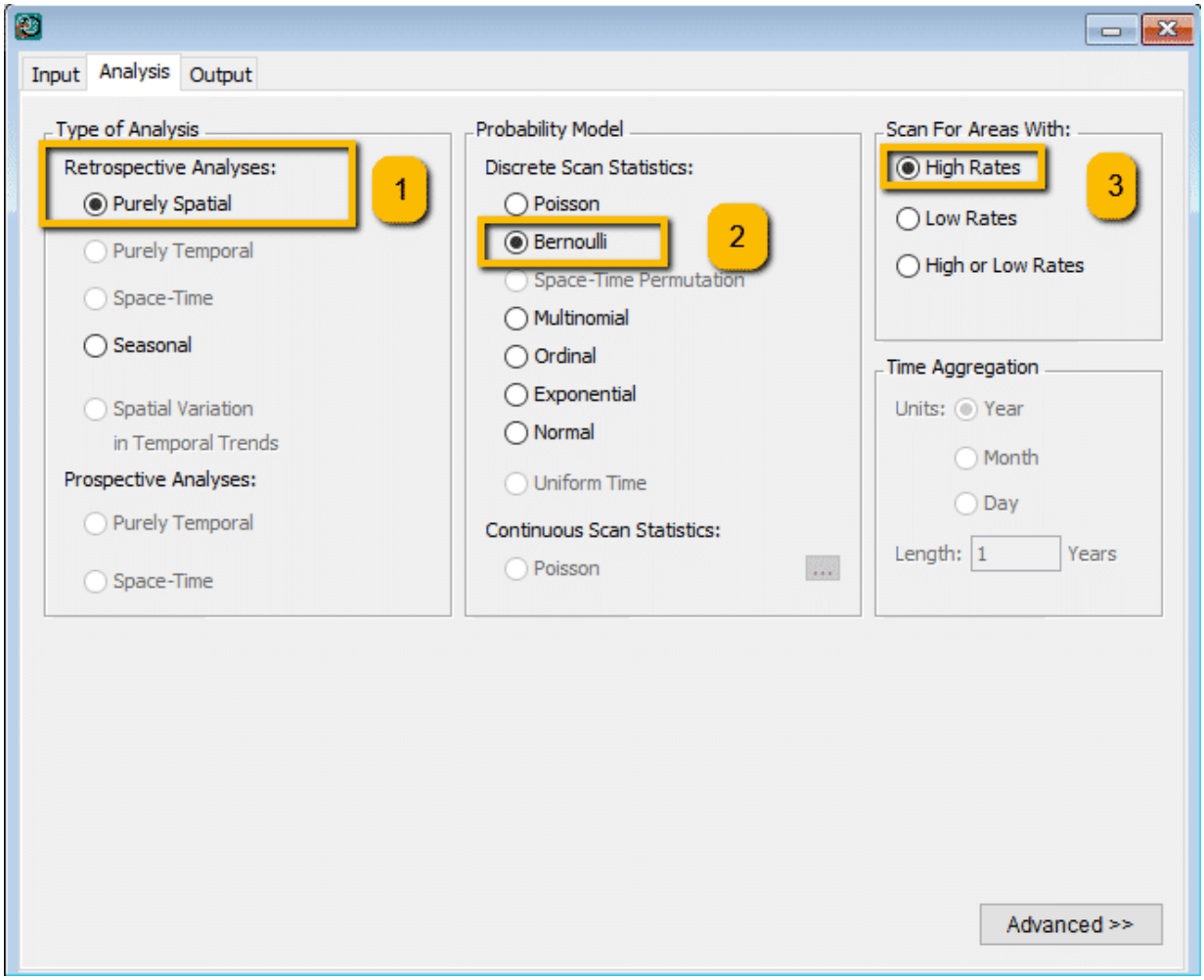


Figure 2.17: Dialog box to set analysis parameters in SaTScan software

CONTINUE

- Select “Output” tab (Figure 2.18)

1. Text Output Format - Main Results File: Create file name “Results” in OutputFiles folder
2. Geographical Output - Check “KML for Google Earth” and “Shapefile for GIS software”
3. Column Output Format- Check “Cluster Information” and “Location Information” and “Risk Estimates for Each Location”

- Click “Advanced” -> Ensure that “Automatically launch requested maps” and “Include All Location IDs in the Clusters” are selected

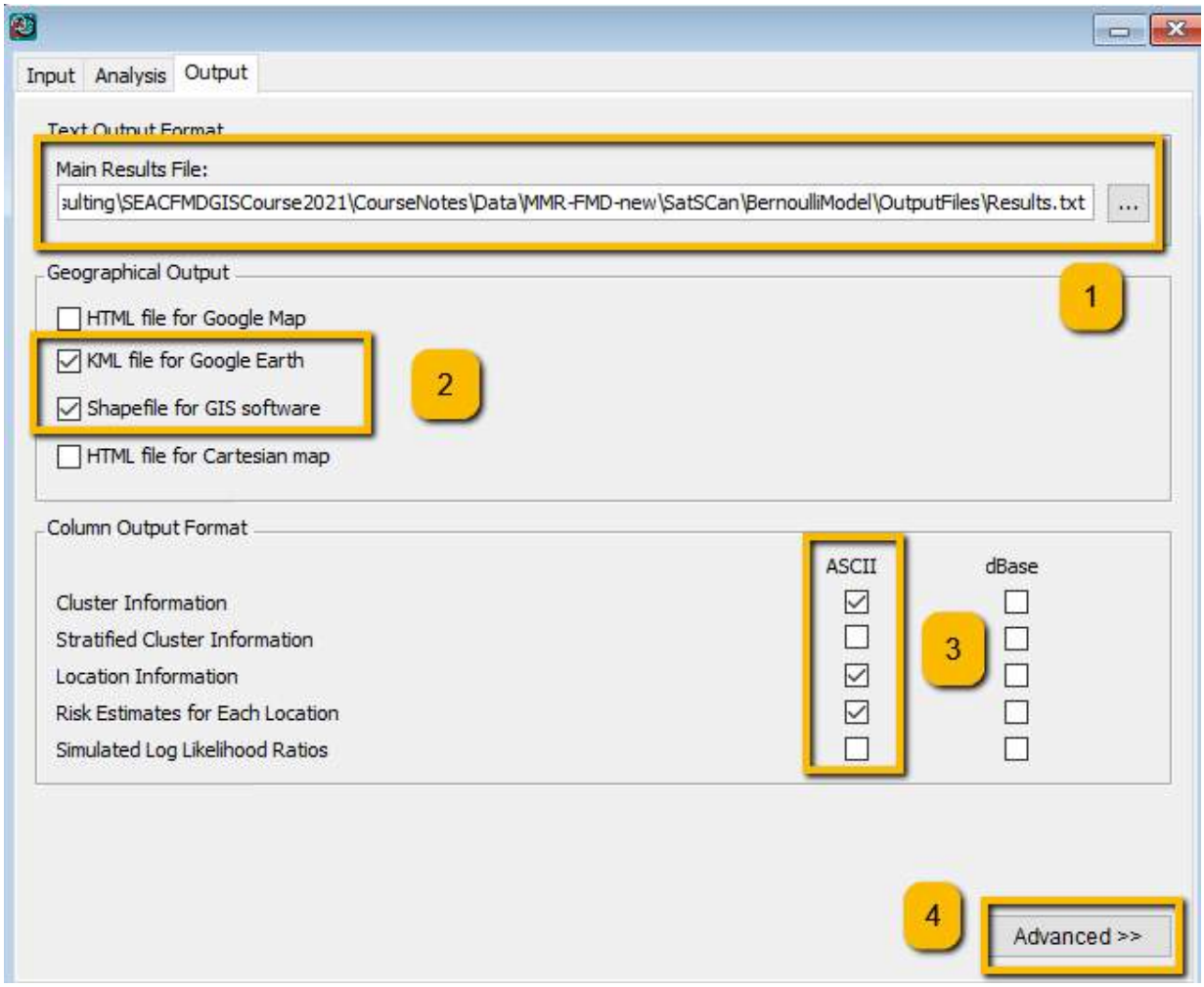


Figure 2.18: Dialog box to set analysis outputs in SaTScan software

Run analysis:

At this point you will receive error messages if mistakes were made in the creation of the files for the analysis. The error message identifies the file(s) with error(s) and may give a brief explanation of the problem. Open the file with errors in SaTScan to try to identify the problem and edit the file in another application and re-run the steps to create that file.

- Click green arrowhead button on button bar (Figure 2.19)

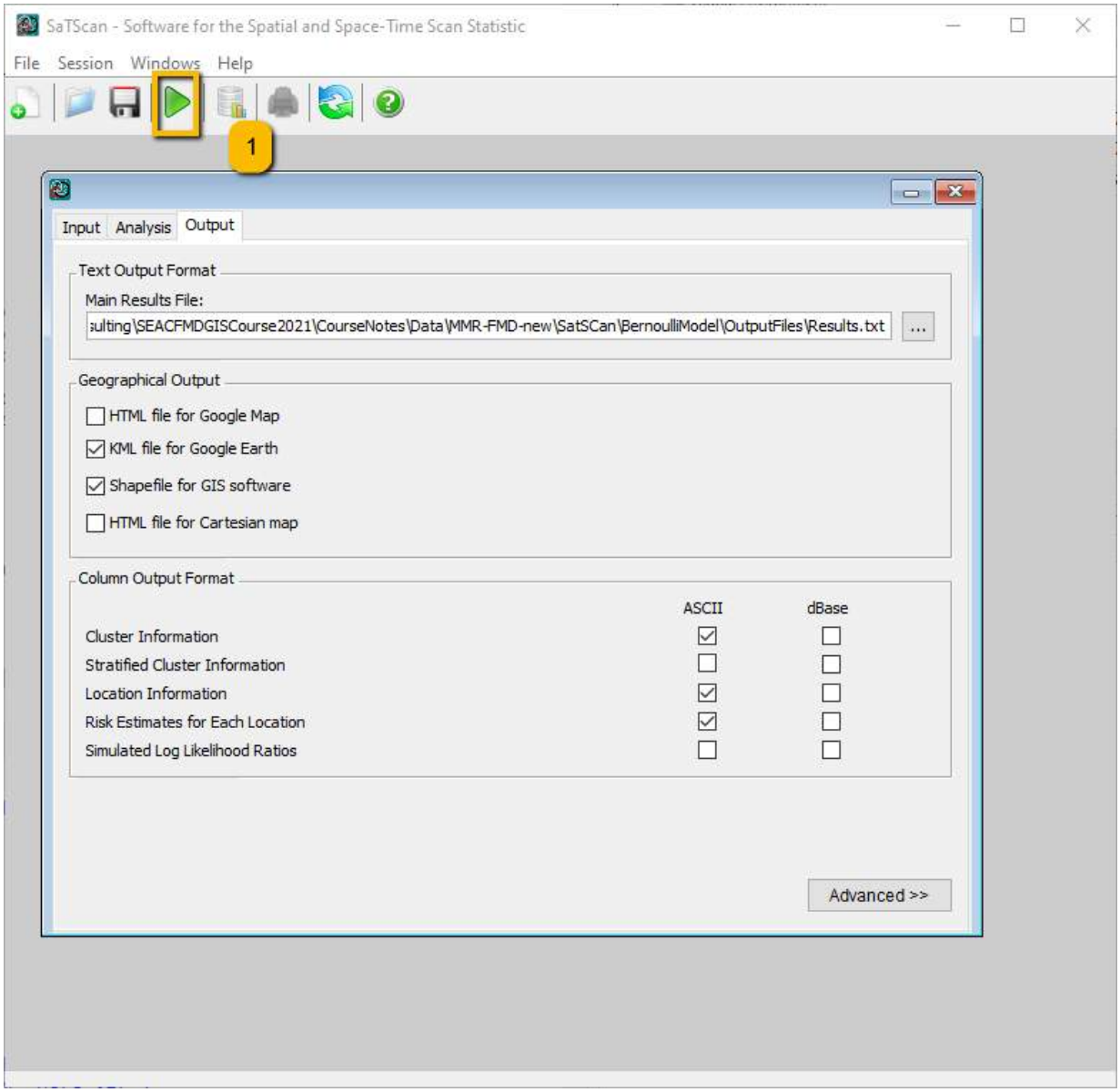


Figure 2.19: Button to run analysis

CONTINUE

2.4.2.4 View results

- If you have Google Earth Pro installed on your computer, it launches and identifies the significant clusters
- The Google Earth Layers can be expanded and checked to show more or less information and the zoom and move functions can be used to change and focus your view to investigate geographic features that might influence the patterns of disease occurrence (Figure 2.20)

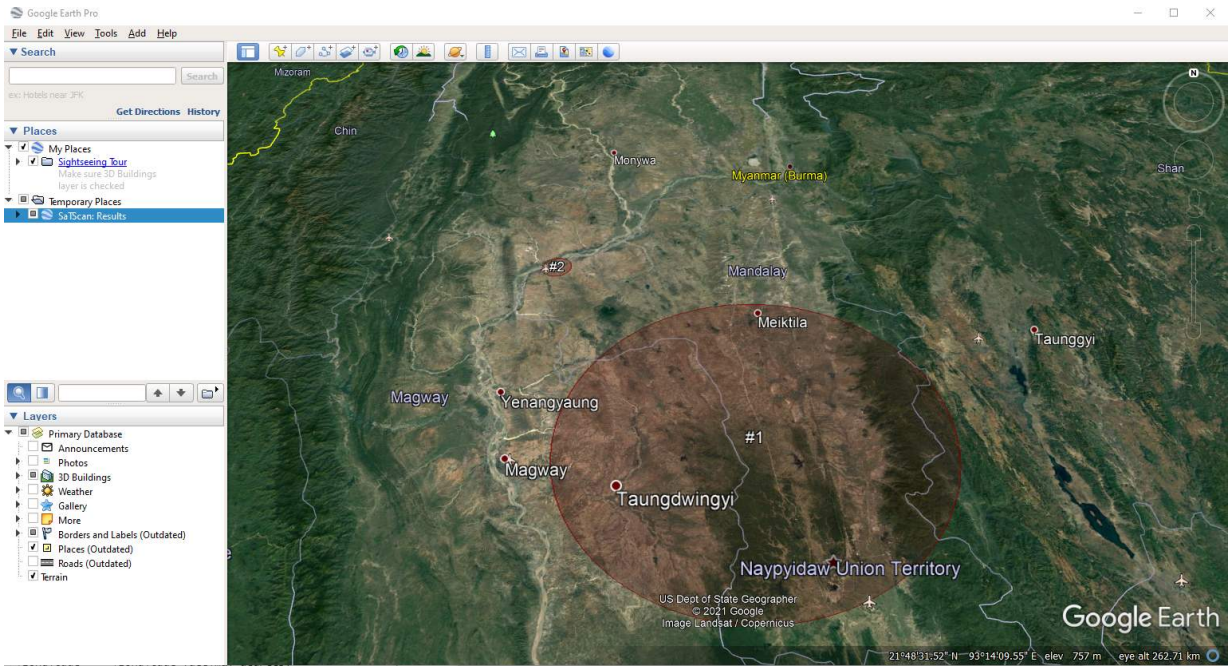


Figure 2.20: Google Earth Pro view of cluster locations

"The SaTScan detailed results can be inspected in the upper window on the application's desktop (Figure 2.21)"

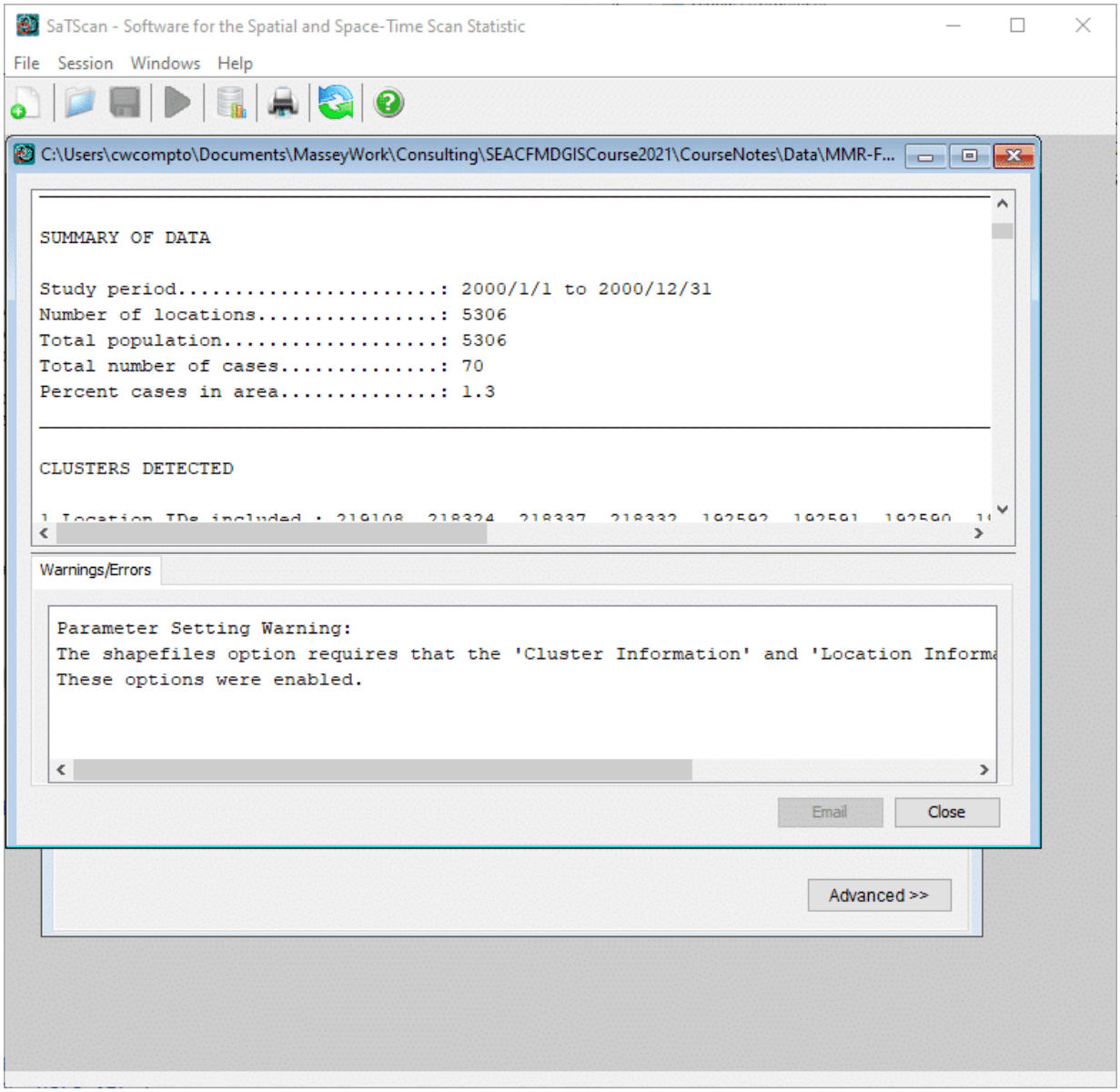


Figure 7.21: Detailed results of SaTScan cluster analysis

"Within QGIS, locate the shapefiles produced by SaTScan in the OutputFiles folder in the Browser pane and drag the shape files e.g. Results.col onto the QGIS Map palette of the Project created earlier to visualise them (Figure 2.22)"

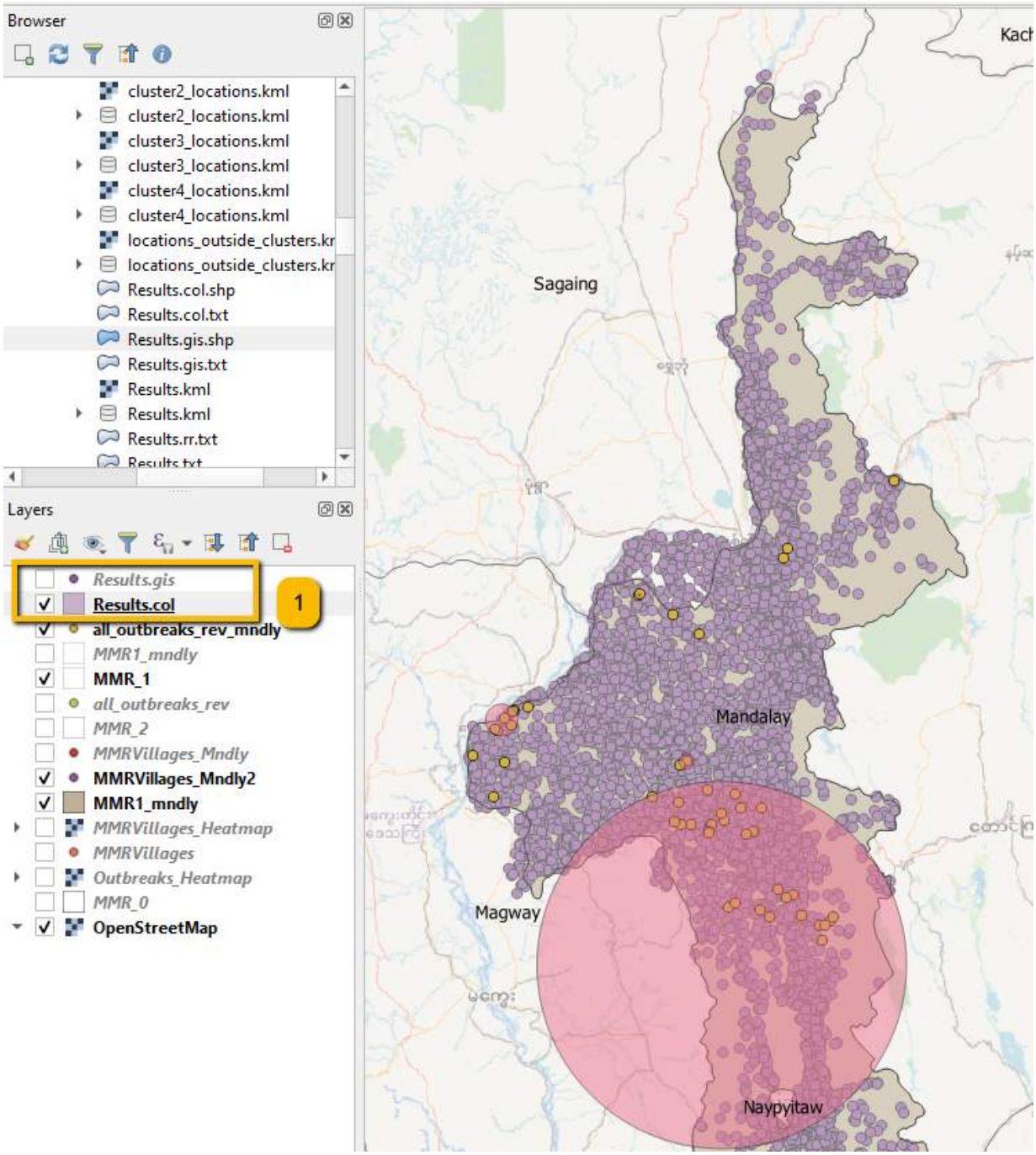


Figure 2.22: Detailed results of SaTScan cluster analysis

CONTINUE

Click ▶ to play the video

CONTINUE

Exercise 2.2: Interpretation

"SaTScan produces a range of results which can be useful for directing disease control measures

1. **What are the main findings from your analysis?**
2. **What are the limitations of your findings?**

CONTINUE

Have you written down your answers for the Exercise 2.2?

- ◯ Yes
- ◯ No

SUBMIT

CONTINUE

"**Answer keys**

1. Two statistically significant (P < 0.05) clusters were identified in the study region by SaTScan- one in the south with many outbreaks, and a smaller one in the north-west of the study area with only a few outbreaks

2. First, the statistically significant clusters are depicted as circular regions, which doesn't seem to fit the large cluster in the south particularly well. It is possible to change the spatial window shape in the Analysis tab in SaTScan from circular to elliptic, which might improve the fit. Second, we assume that the village outbreaks are reported with 100% accuracy, that is, all the true outbreaks are reported at their true locations and that no false outbreaks are reported. This may not be the case always, and it is important to understand the accuracy and completeness of the data provided for analysis when interpreting the results."

CONTINUE

Congratulations - end of lesson reached

NEW ZEALAND
FOREIGN AFFAIRS & TRADE

OIE

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Australian Government
Department of Agriculture,
Water and the Environment