Geological disaster analysis toolkits of LAN Research Group

ALSA

Automatic Landslide Susceptibility Analyst

User Manual

Hengxing LAN, Langping LI

Laboratory of Resources and Environmental Information System, LREIS
Institute of Geographic Sciences and Natural Resources Research, CAS
Contact: lanhx@lreis.ac.an, lilp@lreis.ac.cn

October 10, 2017

Content

1. Theories

Theories of ALSA can also be found in a paper published by the authors (Li L.P. and Lan H.X., et al., 2017. ***A modified frequency ratio method for landslide susceptibility assessment***. Landslides, 14: 727–741, DOI: 10.1007/s10346-016-0771-x).

1.1 Landslide susceptibility assessment

Landslide susceptibility is a measurement of the occurrence probability of landslides under certain geo-environmental conditions. Landslide susceptibility assessment is generally regarded as a prior step towards an assessment of landslide hazard and risk.

The methods of landslide susceptibility assessment can be categorized into three fundamental types, namely the qualitative "knowledge-driven methods", quantitative "data-driven methods" and quantitative "physically based methods". Knowledge-driven methods assess landslide susceptibility by ranking and weighting different landslide-related factors based on the knowledge of experts. Data-driven methods evaluate landslide susceptibility by referring to the geo-environmental characteristics of those locations where landslides had occurred. Physically based methods predict landslide susceptibility based on the mechanisms and processes that control the initiation (failure) of landslides.

All the three types of methods have both advantages and drawbacks. Generally, more complex methods that require larger amount of data are applied at larger scales of landslide susceptibility mapping, and data-driven methods have become standard in regional scale landslide susceptibility assessment. Data-driven methods broadly include bivariate methods and (multivariate) machine learning methods. Main bivariate methods commonly used are: frequency ratio method, weight of evidence method, fuzzy logic method, information value method, Dempster-Shafer method, GIS matrix method, and so on. Main machine learning methods commonly used are: logistic regression method, artificial neural network method, support vector machine method, random forest method, and so on.

Bivariate methods quantify landslide susceptibility through calculating the weight values of each class of individual landslide-related factors. Among bivariate methods, frequency ratio method is one of the most popular, and can have higher accuracies compared with other methods as shown by several case studies. Although many case studies had shown that machine learning methods generally perform better than frequency ratio method in landslide susceptibility assessments, frequency ratio method is still commonly used by researchers and practitioners. There are two reasons for the ongoing popularity of frequency ratio method. The first reason is that the frequency ratio method is friendly to end users because of the simplicity and clarity of the principles behind. The understandability of the input, calculation and output procedures, as well as the ease of implementation on a GIS environment, make the frequency ratio method be an acceptable simple tool of landslide susceptibility assessment when sufficient data are available. The second reason is that the vulnerabilities to landslide failure of individual landslide-related factors can be investigated by the frequency ratio values calculated for each factor class. Like other bivariate methods, the frequency ratio method not only produce landslide susceptibility maps, but also serve to inspect the correlations between landslides and landslide-related factors.

The classification of landslide-related factors with continuous factor values is usually the first step of the conventional frequency ratio method (see the following section "Conventional frequency ratio method"). However, the classifications of factors will induce a discontinuity problem of the frequency ratio values and a subjectivity problem. The discontinuity of the frequency ratio values means that all the factor values in the same factor class will have the same frequency ratio value, which will eventually result in a discontinuity of the spatial distribution of landslide susceptibility. The subjectivity problem means that the choices of the number and the bounds of the classes of factors are more or less subjective. Another problem faced in the practice is that the factor classifications and the calculations of frequency ratios for different factors need much manual labors. The subjectivity and manual labor problems can be moderated by adopting statistics in the classification of landslide-related factors. However, whether the classes yielded by statistics reflect reality or not remains a problem.

The modification on the conventional frequency ratio method, and an implementation of this modified method in a GIS environment to get a handy landslide susceptibility assessment tool, are done with considering the above backgrounds.

1.2 Conventional frequency ratio method

Let L and F stand for landslides and a certain landslide-related factor, respectively. Given that the factor F is categorized into n types or subdivided into n classes, the frequency ratio (FR) for the ith type or the ith class of factor F (F_i) can be written as:

$$\begin{aligned} FR_i &= \frac{PL_i}{PF_i} \\ &= \frac{the\ frequency\ of\ landslides\ in\ the\ F_i\ area}{the\ frequency\ of\ the\ F_i\ area} \\ &= \frac{the\ area\ of\ landslides\ in\ the\ F_i\ area / the\ area\ of\ landslides\ in\ the\ study\ area}{the\ area\ of\ the\ F_i\ area / the\ area\ of\ the\ study\ area} \end{aligned} \tag{1}$$

A frequency ratio FR_i larger than 1 indicates that "*the frequency of landslides in the F_i area*" (PL_i) is larger than "*the frequency of the F_i area*" (PF_i) and further indicates that the ith type or the ith class of factor F (F_i) favors the occurrence of landslides. On the contrary, a frequency ratio FR_i smaller than 1 indicates that F_i does not favor the occurrence of landslides. A transformation of Eq. (1) is:

$$
\begin{aligned}
FR_i &= \frac{\textit{the area of landslides in the } F_i \textit{ area} / \textit{the area of the } F_i \textit{ area}}{\textit{the area of landslides in the study area} / \textit{the area of the study area}} \\
&= \frac{\textit{the probability of landslides in the } F_i \textit{ area}}{\textit{the probability of landslides in the study area}} \\
&= \frac{p(L \mid F_i)}{p(L)}
\end{aligned} \tag{2}
$$

Since "*the probability of landslides in the study area*" $p(L)$ is predetermined given that the landslide and factor data are provided, the frequency ratio FR_i is totally determined by "*the probability of landslides in the* F_i *area*" $p(L|F_i)$, which is in fact "*the conditional probability of L given* F_i". A larger conditional probability $p(L|F_i)$ means that the occurrence probability of landslides is larger in the ith type or the ith class of factor F (F_i).

Consider an arbitrary landslide-related factor $F^{(j)}$ (j = 1, 2, 3, …, m), its frequency ratios with regard to different types or different classes, namely $FR_i^{(j)}$ (i = 1, 2, 3, …, n; j = 1, 2, 3, …, m), can be calculated according to Eq. (1). If the type or the class of $F^{(j)}$ at a certain location is $F_i^{(j)}$, the frequency ratio of this factor at this location $FR^{(j)}$ will be $FR_i^{(j)}$. Then, the Landslide Susceptibility Index (LSI) at this location will be the summation of the frequency ratios of different landslide-related factors at this location:

$$
LSI = \sum_{j=1}^{m} FR^{(j)} \tag{3}
$$

It is worth noting that, if landslide areas are not available, the counts of landslides can be used instead of the areas of landslides in Eq. (1) and Eq. (2).

1.3 Modified frequency ratio method

The modified method also calculates a landslide susceptibility index by summing the frequency ratios of different landslide-related factors. The framework of calculating the frequency ratio of a certain landslide-related factor is shown in the figure below. There are three essential procedures in this modified method.

The procedures of the modified frequency ratio method

1) Normalization. This procedure normalizes the continuous factor values of each factors to 0-1 range. The normalization of factor values make it possible to use the same parameters for different factors in the following precision setting and frequency statistics procedures.

2) Precision setting. The purpose of precision setting is to reduce the calculation loads on the condition that adequate precision is guaranteed so that the calculation can be accelerated. For example, if the parameter "precision" is set to be 3, the normalized factor values will have only 3 digits after the decimal point. This means that there will exist at most 1001 identical normalized factor values. This further means that the variety of the frequency ratios of each landslide-related factor will be characterized by at most 1001 values, since frequency ratios will be calculated with regard to each identical normalized factor value in the frequency statistics step.

3) Frequency statistics. For the *k*th identical normalized factor value I_k (k = 1, 2, 3, …, l), we mark those regions covered by normalized factor values within a certain neighborhood around this identical normalized factor value (I_k) as F_k. Then, F_k can be analogized to the *i*th type or the *i*th class of factor F (F_i) in the conventional frequency ratio method. Given that the area of F_k (AF_k) and the area (or the count) of landslides in F_k (AL_k) are calculated, "*the frequency of* F_k" (PF_k) and "*the frequency of landslides in* F_k" (PL_k) as well as "*the frequency ratio of* I_k" (FR_k) can be calculated according to Eq. (1). The size of the neighborhood is determined by a parameter "bin width", which is often used in histogram statistics. This bin width is a value between 0 and 1 since the factor values have already been normalized. The bin width can be analogized to the widths of different factor classes in the conventional frequency ratio method.

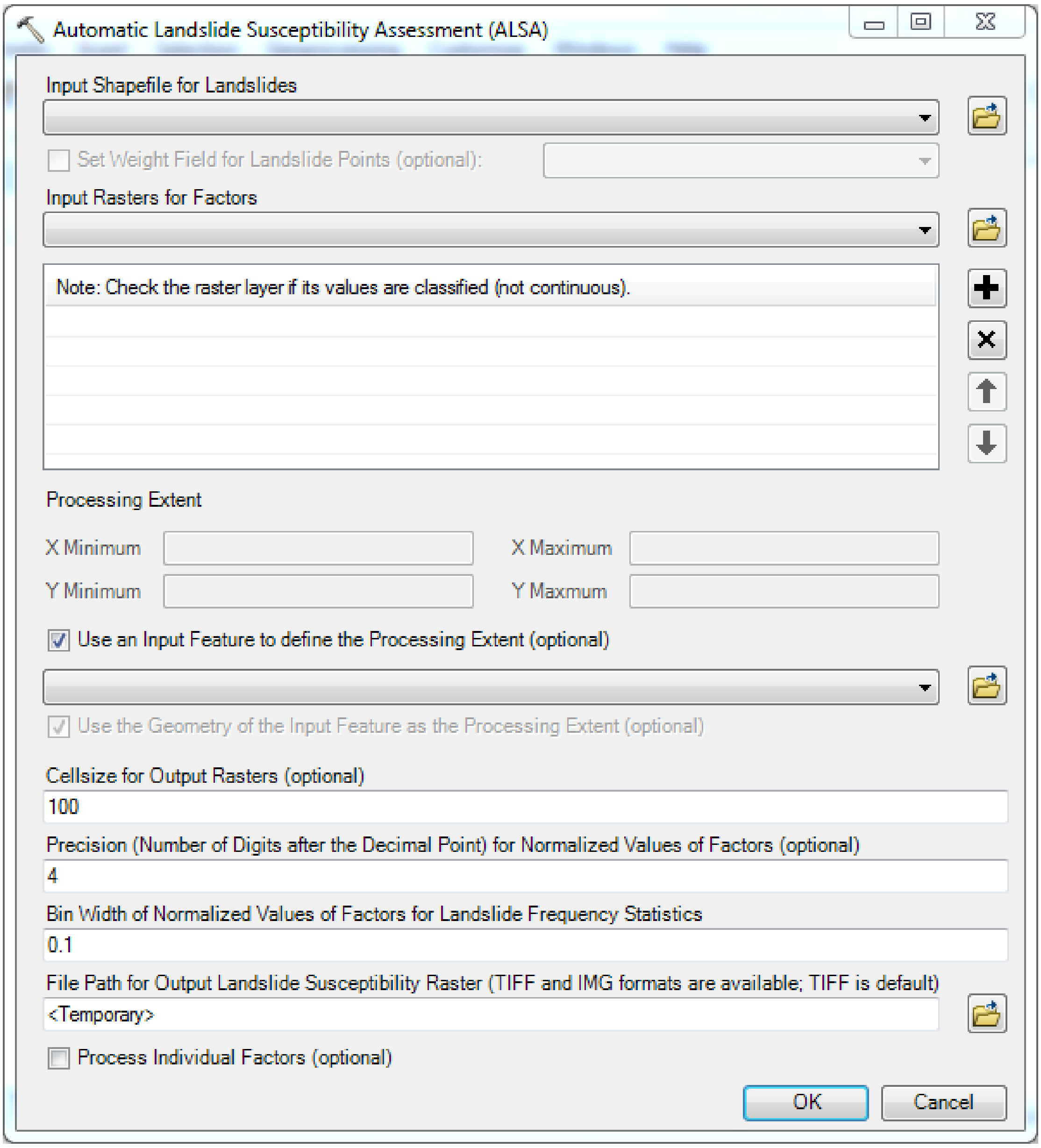
The conventional frequency ratio method calculates the frequency ratios for only a few factor classes (or types), while the modified method calculates the frequency ratios for quite a few identical normalized factor values so that the variation of frequency ratio with factor value are more differentiable. The number of identical normalized factor values is determined by the original factor values and the precision. Yet, it must be emphasized that precision setting is not obligatory. Theoretically, if precision is not considered, the modified method can calculate a frequency ratio for each of the factor values within the study area. This modified frequency ratio method can be regarded as applying "moving frequency statistics" for every identical normalized factor value using a uniform neighborhood window size. The neighborhoods of different identical normalized factor values can have overlaps, and can also have gaps if a low precision and a small bin width are adopted.

For factors that are already categorized into types, e.g. geological units, the frequency ratios are still calculated using the conventional method since continuous factor values are not available. For factors with continuous factor values, through normalizing factor values, the modified method makes the calculations of frequency ratios for different factors constrained by only two uniform parameters (precision and bin width). Because precision setting is not obligatory, the bin width becomes the only obligatory parameter that need to be input by the users. Therefore, the subjectivities associated with the manual classifications of factors in the conventional method can be reduced. Moreover, the simplicity of the inputs of the improved method significantly reduces manual labor and therefore favors an automatic and quick assessment of landslide susceptibility.

2. ALSA tool

2.1 Introduction

A GIS extension called "Automatic Landslide Susceptibility Analyst (ALSA)" that implements the modified frequency ratio method was developed in ArcGIS using ArcObjects and C#. The interface of ALSA in ArcMap is shown in the following figure.



The interface of the ALSA extension in ArcMap

A simple introduction of ALSA is given below, detailed introductions of configuring settings are given in the case study.

The input data needed by ALSA include: 1) Landslides. The landslide data can be points or polygons. Setting weight is an option for point data so that the magnitudes of landslides can be represented if landslide areas (polygons) are not available. 2) Landslide-related factors. The landslide-related factor data must be in raster format. If a factor is already classified, the checkbox in front of its corresponding data layer is needed to be checked. If all the factors are classified, i.e. all the checkboxes are checked, the input textboxes for precision and bin width will become disabled since there are no factors with continuous factor values needed to be processed using the modified method. 3) Processing extent. The four coordinates of a rectangular processing extent can be manually input and modified by the users. This rectangular processing extent can also be automatically obtained by selecting vector or raster data. In this way, the extent of the data (a rectangular) will be accepted as the processing extent, and manual modification of processing extent is not allowed. Using the geometry of a polygon feature as the processing extent is also an option. The coordinate systems of the landslide data, landslide-related factor data and the data defining the processing extent must be the same.

The input parameters needed by ALSA include: 1) cell size for output rasters; 2) precision of identical normalized factor values; and 3) bin width for frequency statistics. The full path of the output *LSI* raster will be automatically defined according to the configured settings or can be manually defined by the users. Furthermore, if the checkbox "Process Individual Factors (optional)" is checked, the *LSI* will not be calculated according to Eq. (3), and only the frequency ratio rasters for individual landslide-related factors will be output.

2.2 Requirements

This program is an extension of ArcGIS. It was developed using ArcObjects and .NET framework. The following prerequisites are needed for running this program.

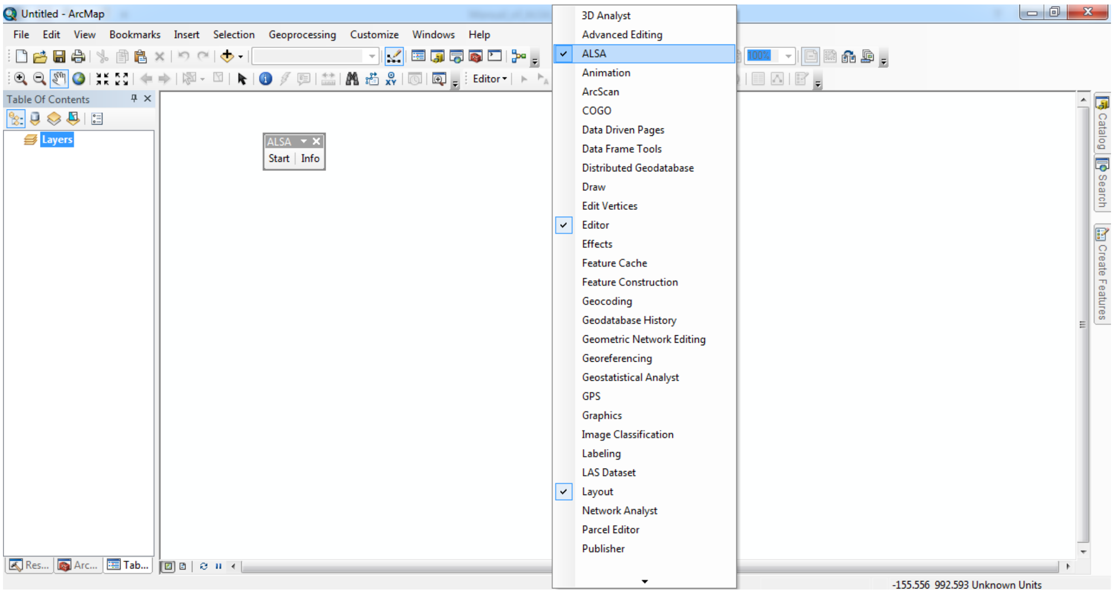
1) ***ArcGIS 10.3 or higher*** with spatial analyst extension. This program had been tested and run successfully in ArcGIS 10.3 and ArcGIS 10.4.

2) ***Microsoft .NET Framework 4.5 or higher***.

2.3 Installation

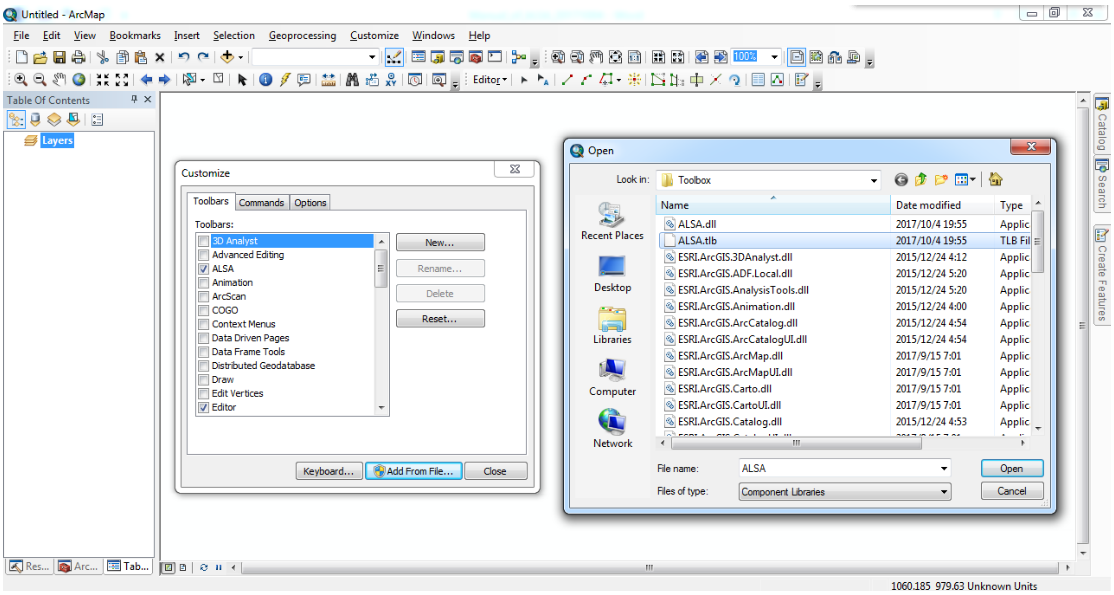
Locate to the folder "ALSA\Setup", click Setup.exe or Setup.msi, then the setup processes will start. The default installation folder is "C:\Program Files (x86)\ALSA\". Follow the onscreen instruction and close the setup form when the setup is complete.

Start ArcMap, there is a toolbar named "ALSA" to access the functions of the ALSA program. Right click the upper empty space of ArcMap and check the "ALSA" toolbar if it is not shown up.



Active the ALSA extension in ArcMap

If setup is unsuccessful, an alternative option can be adopted. Copy the folder “Toolbox” to the local disk. Click “Customize >> Customize Mode” in ArcMap to open the Customize form. Click “Add From File”, browse to “ALSA.tlb” in the folder “Toolbox” and open it. Then, “ALSA” will appear in the “Toolbars” list.



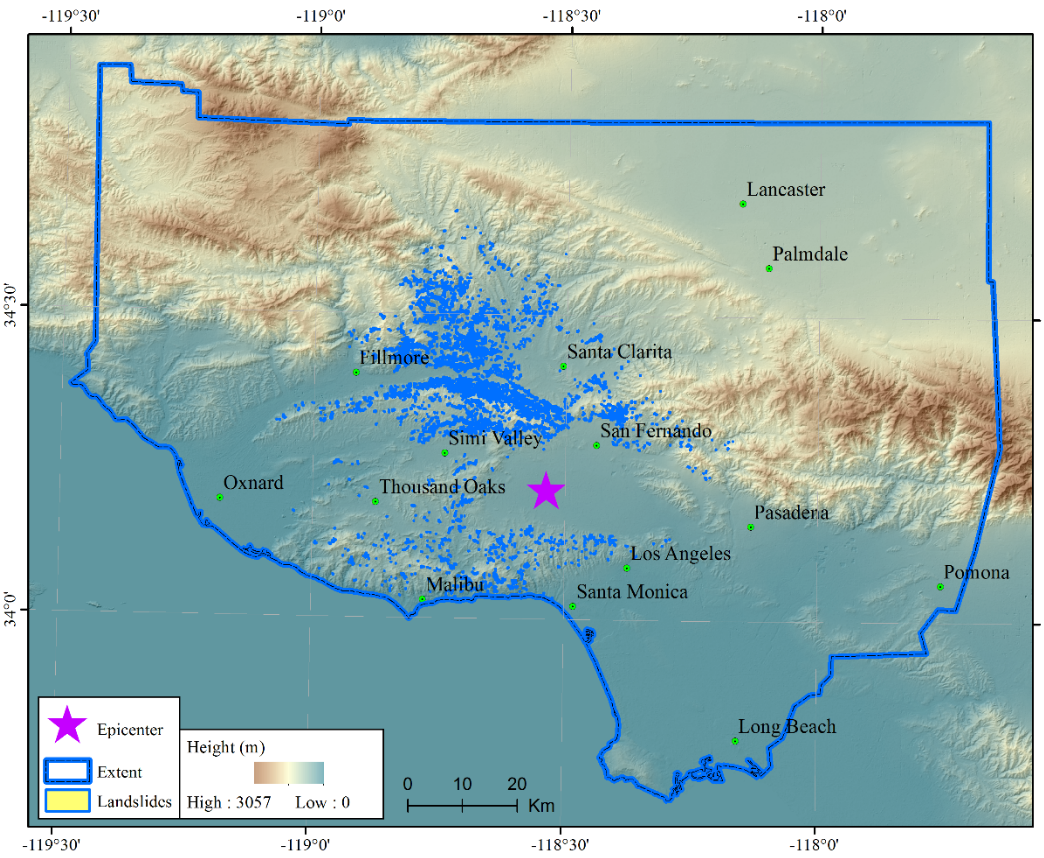
Install ALSA by adding toolbox in ArcMap

3 Case study

3.1 Study area and data

3.1.1 Study area

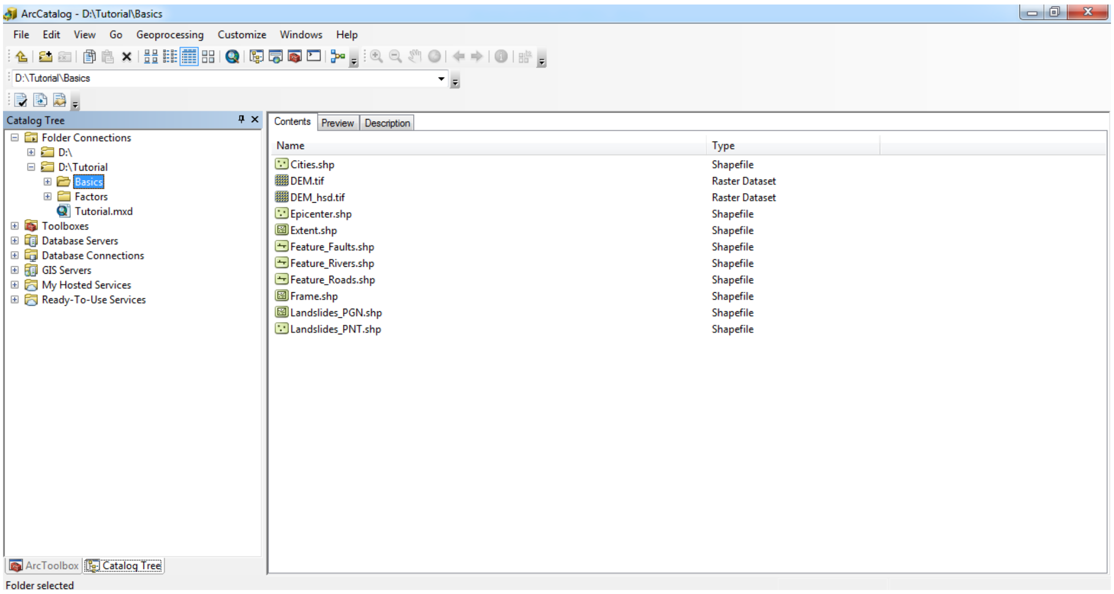
The study area is located in the adjacent areas of the 1994 Northridge Earthquake, which was occurred on January 17th, 1994 and had its epicenter near Los Angeles, California, USA. This Mw 6.7 earth quake had caused 57 deaths, and more than 8,700 injured. More than 10,000 landslides were triggered in the seismic area as shown in the following figure. The landslide data can be accessed via USGS "Open-File Report 95-213" (Harp and Jibson, 1995, https://pubs.usgs.gov/of/1995/ofr-95-0213/).



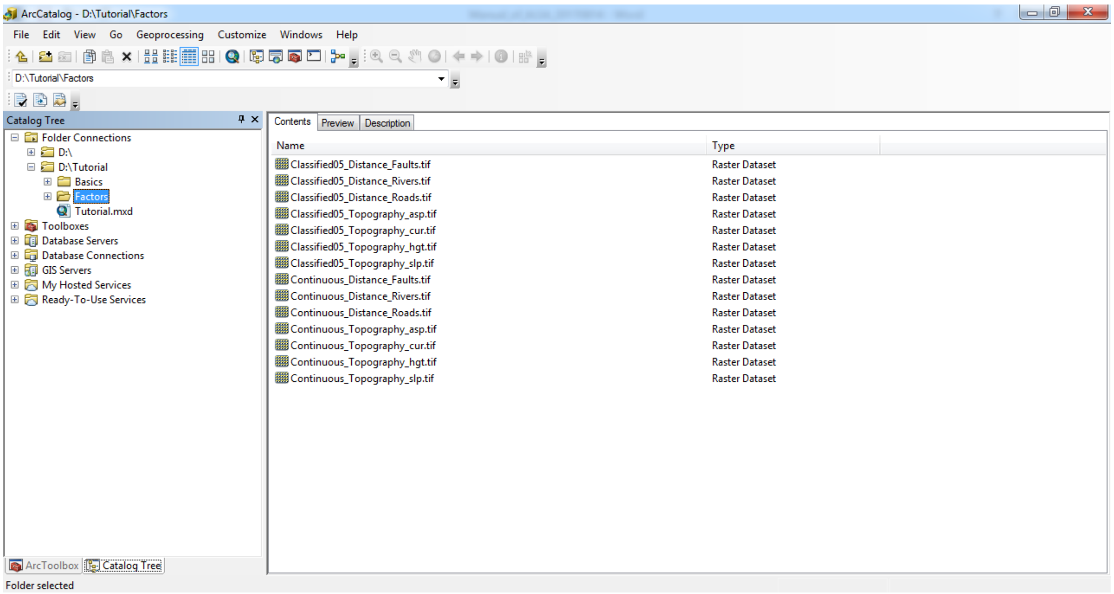
The study area

3.1.2 Input data

The input data in this tutorial are shown and listed in the following figure and table. They are put in two folders (Basics and Factors). Those three "distance factors" were derived from corresponding polyline features. Those four "topography factors" were derived from DEM. Those classified factors were derived from "continuous factors". All the input data have a projected coordinate system "WGS_1984_UTM_Zone_11N".



Input data in the folder “Basics”



Input data in the folder “Factors”

File	*Model*	*Description*
Folder Basics		
Cities.shp	Point Vector	Major cities in the study area.
DEM.tif	Raster	DEM with 30m resolution.
DEM_hsd.tif	Raster	Hillshade of DEM.
Epicenter.shp	Point Vector	Epicenter of the 1994 Northridge Earthquake.

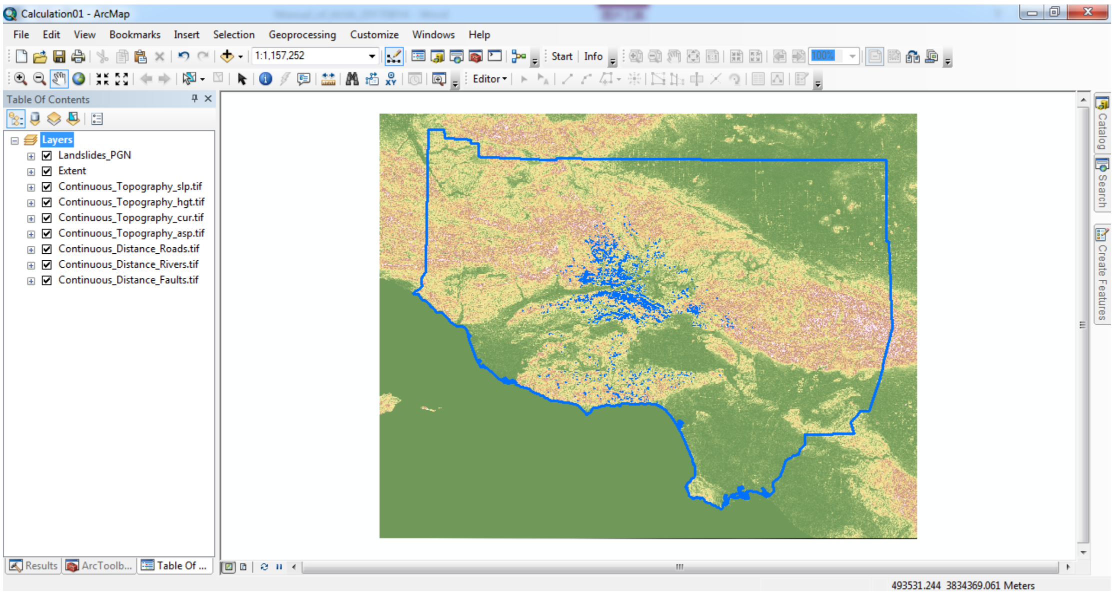
Extent.shp	Polygon Vector	The extent (study area) in which landslide susceptibility assessment will be done, which is the coverage of two counties (Los Angeles and Ventura) in this tutorial.
Feature_Faults.shp	Polyline Vector	Tectonic faults.
Feature_Rivers.shp	Polyline Vector	Rivers and streams.
Feature_Roads.shp	Polyline Vector	Primary and secondary roads.
Frame.shp	Polygon Vector	A fame used to clip input data, which must be larger than the extent.
Landslides_PGN.shp	Polygon Vector	Landslides (totally 10983) triggered by the 1994 Northridge Earthquake.
Landslides_PNT.shp	Point Vector	Landslides in point format, which is converted from polygon format.
Folder Factors		
Classified05_Distance_Faults.tif	Raster	Distance to faults classified into 5 classes using “Natural Breaks (Jenks)” method, with 30m resolution.
Classified05_Distance_Rivers.tif	Raster	Distance to rivers classified into 5 classes using “Natural Breaks (Jenks)” method, with 30m resolution.
Classified05_Distance_Roads.tif	Raster	Distance to roads classified into 5 classes using “Natural Breaks (Jenks)” method, with 30m resolution.
Classified05_Topography_asp.tif	Raster	Aspect classified into 5 classes using “Natural Breaks (Jenks)” method, with 30m resolution.
Classified05_Topography_cur.tif	Raster	Curvature classified into 5 classes using “Natural Breaks (Jenks)” method, with 30m resolution.

Classified05_Topography_hgt.tif	Raster	Height classified into 5 classes using "Natural Breaks (Jenks)" method, with 30m resolution.
Classified05_Topography_slp.tif	Raster	Slope classified into 5 classes using "Natural Breaks (Jenks)" method, with 30m resolution.
Continuous_Distance_Faults.tif	Raster	Distance to faults, with 30m resolution.
Continuous_Distance_Rivers.tif	Raster	Distance to rivers, with 30m resolution.
Continuous_Distance_Roads.tif	Raster	Distance to roads, with 30m resolution.
Continuous_Topography_asp.tif	Raster	Aspect from DEM, with 30m resolution. Aspect value -1 derived using ArcToolbox "Aspect" was set to be null.
Continuous_Topography_cur.tif	Raster	Curvature from DEM, with 30m resolution.
Continuous_Topography_hgt.tif	Raster	Height from DEM, with 30m resolution.
Continuous_Topography_slp.tif	Raster	Slope from DEM, with 30m resolution.

3.2 Calculation examples

3.2.1 Calculation01 - Using polygon landslide dataset

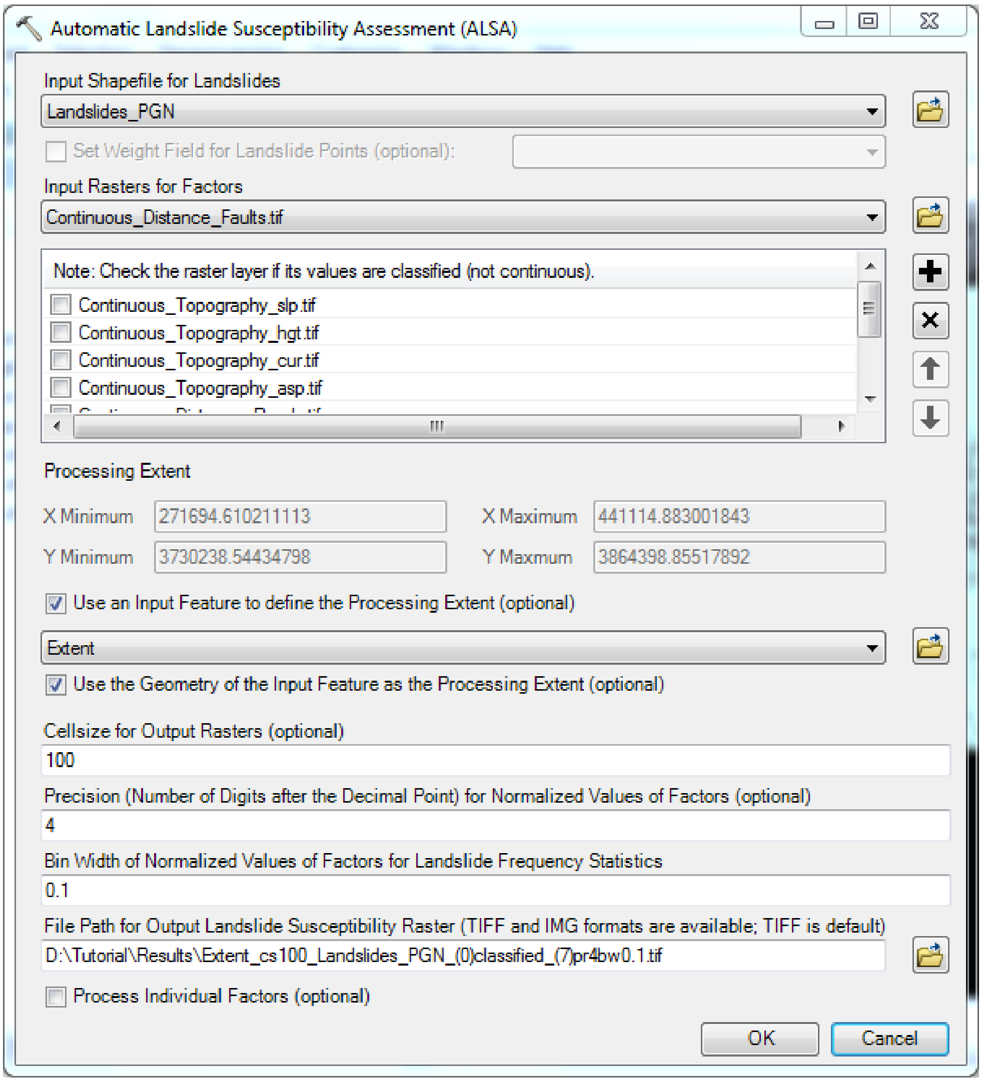
- Open ArcMap and start with a new empty map, and rename it to "***Calculation01***".
- Add all the *7 **factors with continuous value*** in the folder "Factors" into the map.
- New a folder "Results" in the folder "…\Tutorial\" to save resulted files.
- Copy "Extent.shp" and "Landslides_PGN.shp" from the folder "Basics" to the folder "Results".
- Add "***Extent.shp***" and "***Landslides_PGN.shp***" in the folder "Results" into the map.
- Click the "Start" item of the ALSA toolbar to start the setting panel.
- Get the settings configured as shown in the following figure. ***The operation procedures for configuring settings are detailed below***.
- Press the "Ok" button, then the calculation will start.



A map contains the data needed in the calculation example using polygon landslide dataset

■ Input Shapefile for Landslides

- All point and polygon shapefile layers in the map will be added as the items of the comboBox control when the setting panel is loaded. Choose the polygon layer “***Landslides_PGN***” as the input landslide shapefile.
- Immediately after a landslide layer is chosen, the “File Path for Output Landslide Susceptibility Raster” will change from “<Temporary>” to a file path, which is “D:\Tutorial\Results\cs100_Landslides_PGN_(0)classified_(0)pr4bw0.1.tif” in the case of this tutorial.
- ALSA automatically use the folder where the landslide dataset (in the case of this tutorial is “D:\Tutorial\Results\Landslides_PGN.shp”) locates in as the folder where the output landslide susceptibility raster will be saved, in this tutorial is “D:\Tutorial\Results\”. Only if a landslide dataset is defined, a folder and thus a file path for saving the output landslide susceptibility raster can be defined.
- “***cs100***” in the file path indicates the cellsizes for all output rasters, including the output landslide susceptibility raster, is 100m. “***Landslides_PGN***” indicated layer “Landslides_PGN” is chosen as input landslide dataset. “***(0)classified***” indicated the output landslide susceptibility raster is derived based on 0 classified landslide factors. “***(0)pr4bw0.1***” indicated the output landslide susceptibility raster is derived based on 0 landslide factors with continuous values, while the “Precision” and “Bin Width” are 4 and 0.1, respectively. Both classified and continuous factors count zero indicates currently no factor is selected in “Input Rasters for Factors”.
- Given the “File Path for Output Landslide Susceptibility Raster” is defined, either automatically filled by ALSA or manually edited by users, ***all the calculations will be implemented and all the output files will be saved in the folder defined by this file path***, in the case of this tutorial is “D:\Tutorial\Results\”.
- The option “Set Weight Field for Landslide Points” is disabled because the input landslide dataset is polygon in this case.



Settings for the calculation example using polygon landslide dataset

- **Input Rasters for Factors**
- All raster layers in the map will be added as the items of the comboBox control when the setting panel is loaded. ***Choose all the 7 raster layers with continuous values*** as the inputs for factors.
- Immediately after a factor raster layer is chosen, it will be filled in the listView control, and the "File Path for Output Landslide Susceptibility Raster" will also update accordingly. For example, if only one single factor raster layer is chosen, the "File Path for Output Landslide Susceptibility Raster" will change to "D:\Tutorial\Results\cs100_Landslides_PGN_(0)classified_(1)pr4bw0.1.tif", since only one factor with continuous values will be used to assess landslide susceptibility in this instant.
- Immediately after all the 7 factor raster layers are chosen, in the case of this tutorial the "File Path for Output Landslide Susceptibility Raster" will change to "D:\Tutorial\Results\cs100_Landslides_PGN_(0)classified_(7)pr4bw0.1.tif", which means all 7 factors with continuous values will be used to assess landslide susceptibility.

- The checkBox before the factor raster layer in the listView control must be checked if this factor is classified. All the checkBoxes before the factor raster layers in the listView control are left unchecked, since all factors are not classified but have continuous values in this case.

- **Processing Extent**
- ***The processing extent of the calculation is a rectangle*** defined by four coordinates: minimum x, maximum x, minimum y and maximum y.
- Several data preparing procedures will be implemented within this processing extent, including preparing landslide raster, clipping the resampled factor rasters, and sorting the clipped resampled factor rasters. That means, in this case, the output landslide raster “Extent_cs100_Landslides_PGN.tif”, the clipped resampled factor raster “Extent_cs100_Continuous_Distance_Faults.tif”, the index raster for the sorted factor raster “Extent_cs100_Continuous_Distance_Faults_Index.tif”, the sorted factor raster “Extent_cs100_Continuous_Distance_Faults_Sorted.tif”, and the “derived rasters” (including the clipped resampled factor raster, the index raster for the sorted factor raster and the sorted factor raster) for factors other than “Continuous_Distance_Faults.tif” all have the same rectangular extent as the processing extent.
- The processing extent (four coordinates) can be manually input by users or be automatically defined by ALSA through extracting the extent of the chosen layer in the comboBox control if the checkBox “Use an Input Feature to define the Processing Extent” is checked. ***The advice and also the default is to use an input feature to define the processing extent***. All data layers in the map will be added as the items of the comboBox control when the setting panel is loaded.
- Immediately after a data layer is chosen in the comboBox, the “File Path for Output Landslide Susceptibility Raster” will update accordingly, and will change to “D:\Tutorial\Results\Extent_cs100_Landslides_PGN_(0)classified_(7)pr4bw0.1.tif” in the case of this tutorial. The prefix “Extent_” is added in the file name of the output landslide susceptibility raster, of the output landslide raster, and also of the output “derived rasters” for all factors since file “Extent.shp” is used to define the (rectangular) processing extent. The meanings of “cs100”, “Landslides_PGN”, “(0)classified” and “(7)pr4bw0.1” in the file names of all the output rasters have been already explained before.
- Immediately after a polygon layer is chosen in the comboBox, the checkBox “Use the Geometry of the Input Feature as the Processing Extent” is enabled. If this checkBox is enabled and checked, ***further calculations of the frequency ratios for different factors will be implemented within the “geometrical extent” (not a rectangle) of the chosen polygon layer***. That means, the frequency ratio raster “Extent_cs100_Landslides_PGN_pr4bw0.1_Continuous_Distance_Faults.tif” and the frequency ratio rasters for factors other than “Continuous_Distance_Faults.tif” all have the same geometrical extent as the chosen polygon layer “Extent.shp” in the case of this tutorial. As a result, the output landslide susceptibility raster “D:\Tutorial\Results\Extent_cs100_Landslides_PGN_(0)classified_(7)pr4bw0.1.t

if" which is the addition of the frequency ratio rasters for all factors will have the same geometrical extent as the chosen polygon layer "Extent.shp". Similarly, ***the advice and also the default is to use the geometry of the input feature as the processing extent***.

- If the geometry of the input feature is used as the processing extent, some other datasets will be output, in the case of this tutorial, including an extent raster "Extent_cs100.tif", an extent polygon feature converted from the extent raster "Extent_cs100.shp" which is commonly different from the original extent polygon feature "Extent.shp", and 7 resampled factor rasters clipped using the geometry of the input feature among which the one for factor "Continuous_Distance_Faults.tif" is "Extent_cs100_Continuous_Distance_Faults_Mask.tif".

■ **Cellsize for Output Rasters**

- The input value will be used as the cellsizes of all output rasters. ***The advice is to input a value although this is optional***. The cell size of the first factor raster will be used as the cellsizes of all output rasters if there is no input. The input cellsize is not necessarily the same as the input factor rasters. ***A value equal to or larger than the cellsizes of the input factor rasters is commonly suggested***. Larger cellsize (lower spatial resolution) results in faster calculations.
- The input cellsize value will influence the automatically generated "file path for output landslide susceptibility raster" and the file names of output datasets. In this case, "cs100" will change to "cs50" if the input cellsize change from 100 to 50.

■ **Precision (Number of Digits after the Decimal Point) for Normalized Values of Factors**

- The input integer will be used as the number of digits after the decimal point for normalized values of factors. ***The advice is to input an integer although this is optional***. 0 precision will be used if there is no input. 0 precision means the frequency ratios for all the identical factor values of factor rasters within the processing extent will be calculated. Larger precision (more calculations of frequency ratio) results in longer time of calculations.
- The input precision value will influence the automatically generated "file path for output landslide susceptibility raster" and the file names of output frequency ratio rasters. In this case, "pr4bw0.1" will change to "pr5bw0.1" if the input precision change from 4 to 5.

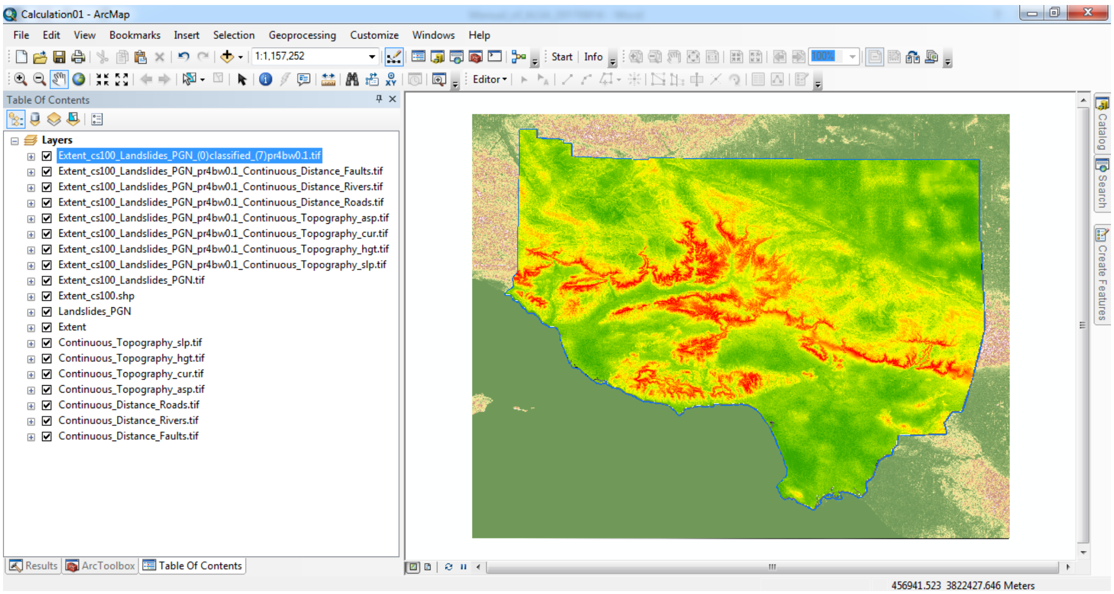
■ **Bin Width of Normalized Values of Factors for Landslide Frequency Statistics**

- The input value (0 ~ 1) will be used as the neighborhood size in the frequency statistics of identical normalized values of factors. This input is obligatory. Larger bin width (wider statistical window) results in longer time of calculations.
- The input bin width value will influence the automatically generated "file path for output landslide susceptibility raster" and the file names of output frequency ratio rasters. In this case, "pr4bw0.1" will change to "pr4bw0.2" if the input bin width change from 0.1 to 0.2.

- ■ **File Path for Output Landslide Susceptibility Raster**
- ● ***The file path for the output landslide susceptibility raster will automatically change as the settings change*** as stated before. The file path in this case is “D:\Tutorial\Results\Extent_cs100_Landslides_PGN_(0)classified_(7)pr4bw0.1.tif”, in which “Extent” means the data layer “Extent.shp” is used to define the processing extent, “cs100” means the cellsizes of the output rasters are 100m, “Landslides_PGN” means landslide dataset “Landslides_PGN.shape” is used to calculate frequency ratios of different factors, “(0)classified” means 0 classified factors are used to calculate landslide susceptibility, “(7)pr4bw0.1” means 7 continuous factors are used to calculate landslide susceptibility by adopting a precision value of 4 and a bin width value of 0.1, while the directory “D:\Tutorial\Results\” in which the output rasters will be saved is the same as the directory where the used landslide dataset “Landslides_PGN.shp” locates in.
- ● “File Path for Output Landslide Susceptibility Raster” can also be manually defined or modified if the user do not want to use the file path automatically generated. However, ***the advice is to use the automatically generated file path***.
- ● The automatically generated file path cannot distinguish whether weight field is set for point landslide dataset, cannot distinguish whether the geometry of the input feature is used as the processing extent, and also cannot tell which factors are used to calculate landslide susceptibility but can only tell the counts of used factors.

- ■ **Processing Individual Factors**
- ● If the checkBox “Processing Individual Factors” is checked, frequency ratio rasters will be derived for each factors, but the final step that adding the frequency ratios for different factors to get landslide susceptibility will not be done.

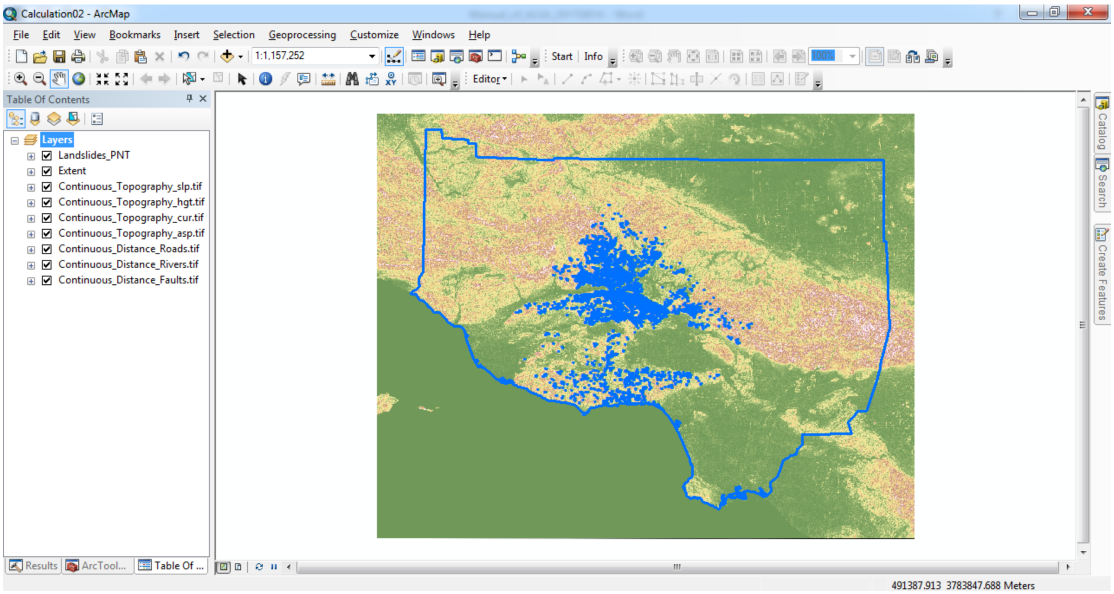
- ■ The map “Calculation01” after calculation is shown as the following figure.



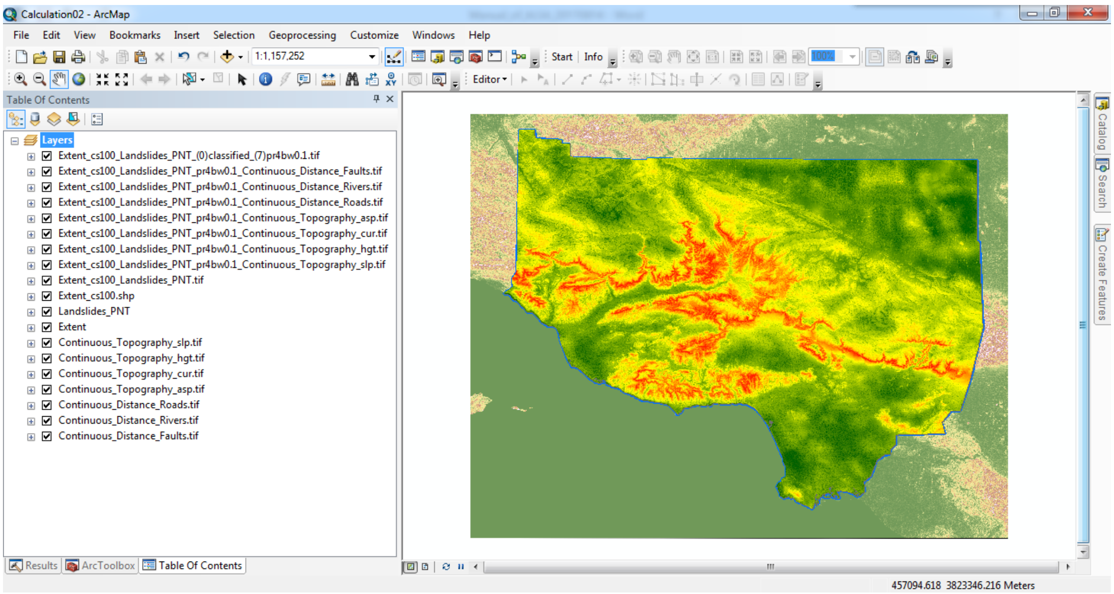
A map after calculation in the calculation example using polygon landslide dataset

3.2.2 Calculation02 - Using point landslide dataset

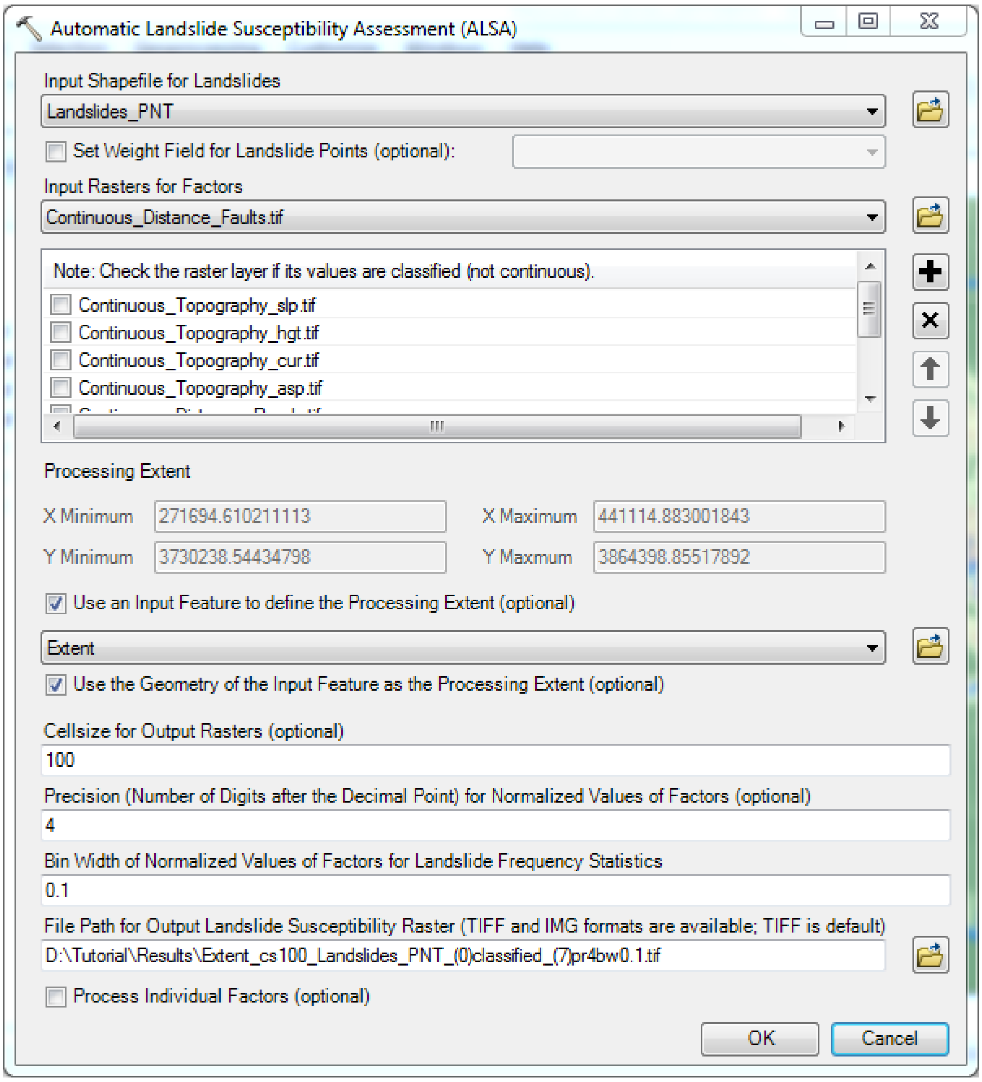
- Open ArcMap and start with a new empty map, and rename it to "***Calculation02***".
- Add all the *7 **factors with continuous value*** in the folder "Factors" into the map.
- Copy "Landslides_PNT.shp" from the folder "Basics" to the folder "Results".
- Add "***Extent.shp***" and "***Landslides_PNT.shp***" in the folder "Results" into the map.
- Click the "Start" item of the ALSA toolbar to start the setting panel.
- Get the settings configured as shown in the following figure. The differences of settings compared with the previous example result from the use of landslide point.
- Press the "Ok" button, then the calculation will start.
- The map "Calculation02" after calculation is also shown as follows.



A map contains the data needed in the calculation example using point landslide dataset



A map after calculation in the calculation example using point landslide dataset



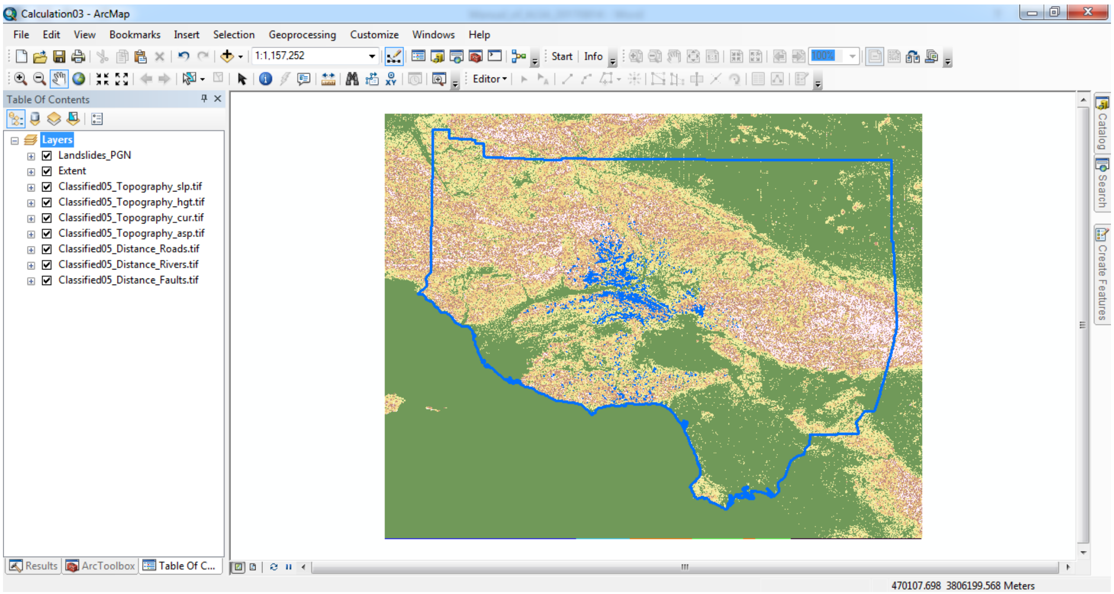
Settings for the calculation example using point landslide dataset

- The settings in this example is exactly the same as that in the example using polygon landslide dataset, except that a point landslide dataset is used. Therefore, the detailed operation procedures for configuring settings are referred to the calculation example using polygon landslide dataset.

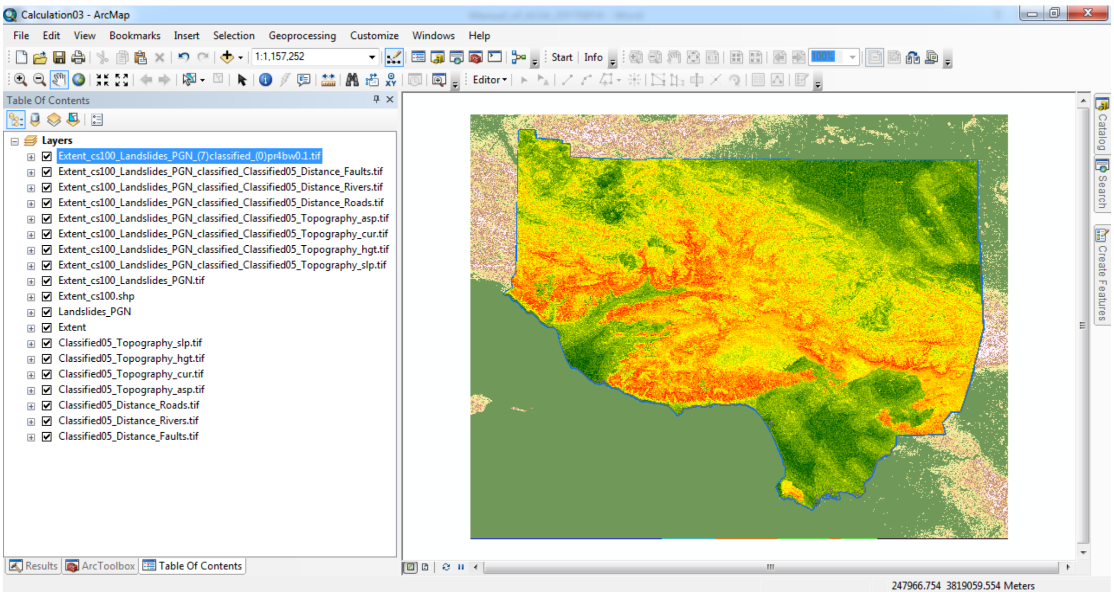
- **Input Shapefile for Landslides**
- The option "Set Weight Field for Landslide Points" is enabled because the input landslide dataset is point in this example. This checkBox is left unchecked in this calculation example.
- If this checkBox is checked, a field of landslide point shapefile can be chosen to define the weight of each feature (point). ***This considers that different landslide points may represent different magnitudes of landslide***. In this case, as the landslide point shapefile was converted from a landslide polygon shapefile with a filed recording the area of each landslide, an ideal weight field would be the "area field" which is reserved in the derived landslide point shapefile.

3.2.3 Calculation03 & 04 - Using the conventional frequency ratio method

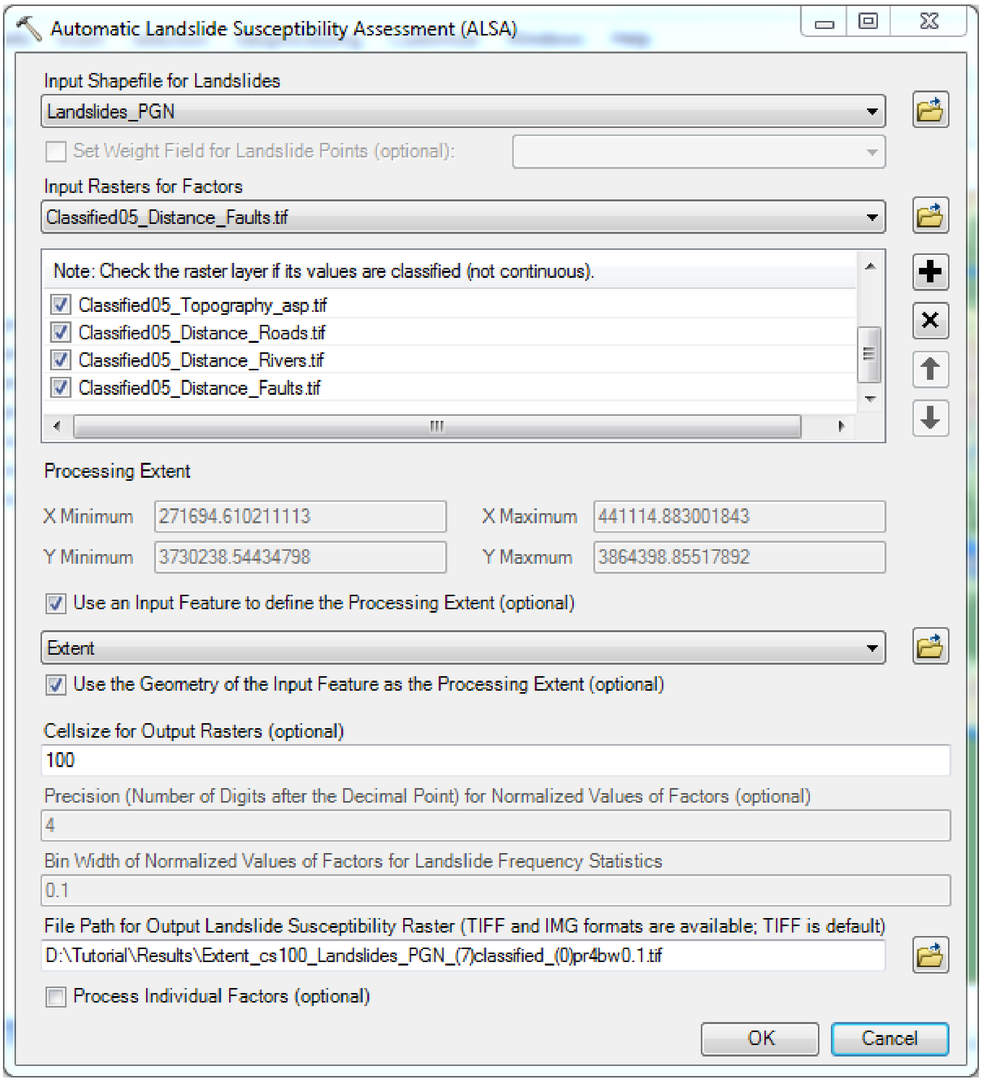
- Open ArcMap and start with a new empty map, and rename it to "***Calculation03***".
- Add all the *7* ***classified factors*** in the folder "Factors" into the map.
- Add "***Extent.shp***" and "***Landslides_PGN.shp***" in the folder "Results" into the map.
- Click the "Start" item of the ALSA toolbar to start the setting panel.
- Get the settings configured as shown in the following figure.
- Press the "Ok" button, then the calculation will start.
- The map "Calculation03" after calculation is also shown as follows.



A map contains the data needed in the calculation example using the conventional frequency ratio method and polygon landslide dataset



A map after calculation in the calculation example using the conventional frequency ratio method and polygon landslide dataset



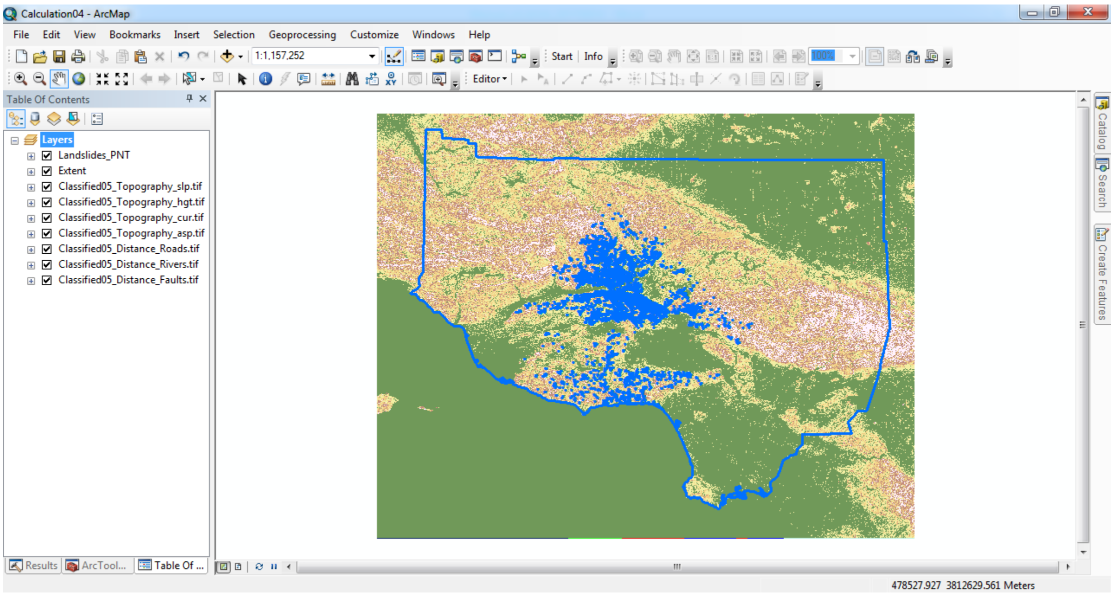
Settings for the calculation example using the conventional method and polygon landslide dataset

- ***This case shows implementing landslide susceptibility assessment using the conventional frequency ratio method*** since all used factors are classified.
- The settings in this calculation example make differences owing to the throughout use of classified factors. Differences in operation procedures for configuring settings compared with the previous calculation examples are detailed below.

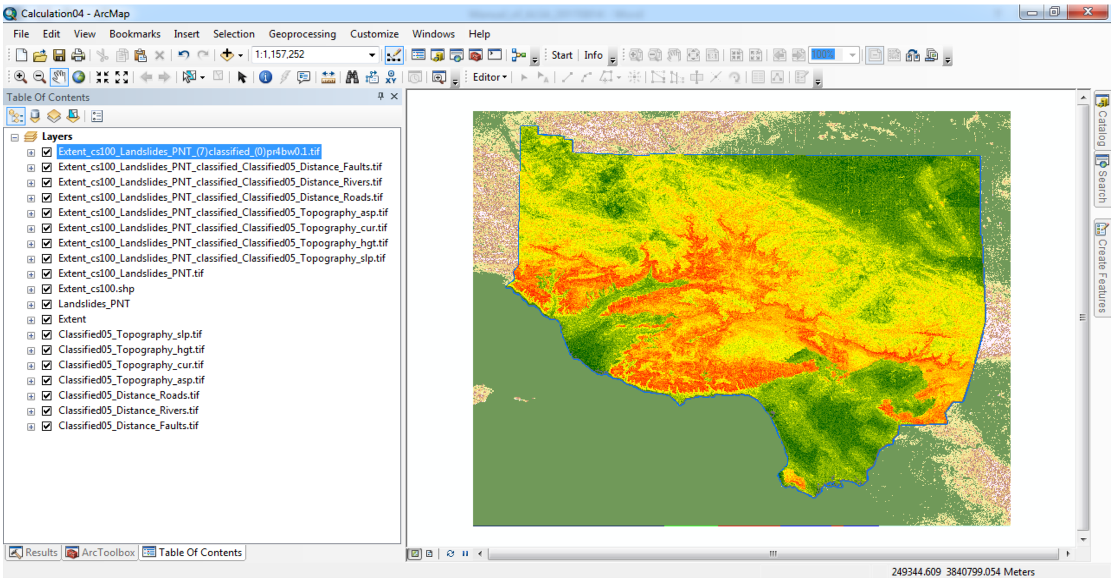
- **Input Rasters for Factors**
- ***Choose all the 7 classified factor raster layers*** as the inputs for factors.
- ***Check all the checkBoxs before the factor raster layer in the listView control***, since all factors are classified in this case.
- The controls for "Precision" and "Bin Width" will be disabled immediately after all checkBoxs for factor rasters are checked, since the conventional frequency ratio method is used in this case.
- The "File Path for Output Landslide Susceptibility Raster" will finally change to "D:\Tutorial\Results\cs100_Landslides_PGN_(7)classified_(0)pr4bw0.1.tif", since all 7 factors used to assess landslide susceptibility are classified.

- The following shows the calculation example using the conventional frequency ratio method and point landslide dataset

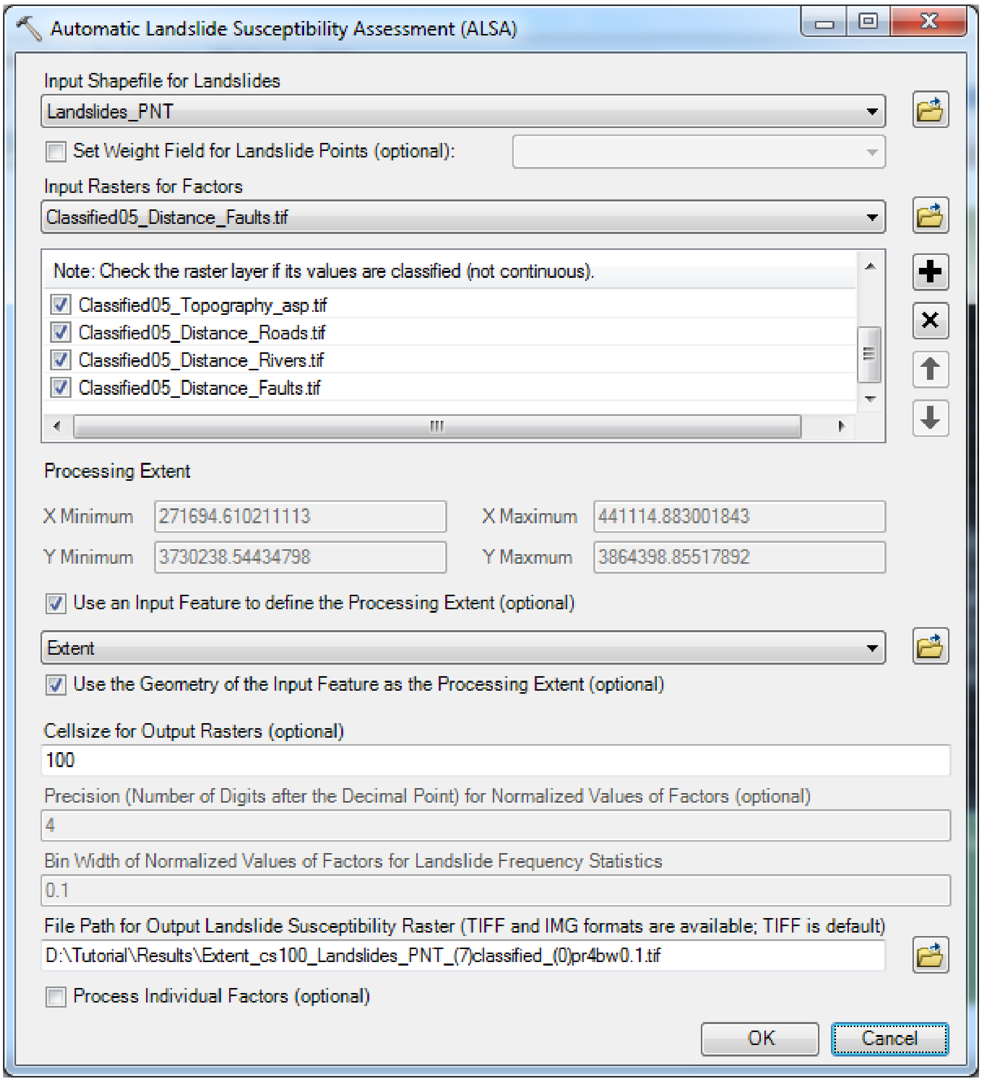
- Open ArcMap and start with a new empty map, and rename it to "***Calculation04***".
- Add all the *7* ***classified factors*** in the folder "Factors" into the map.
- Add "***Extent.shp***" and "***Landslides_PNT.shp***" in the folder "Results" into the map.
- Click the "Start" item of the ALSA toolbar to start the setting panel.
- Get the settings configured as shown in the following figure.
- Press the "Ok" button, then the calculation will start.
- The map "Calculation04" after calculation is also shown as follows.



A map contains the data needed in the calculation example using the conventional frequency ratio method and point landslide dataset



A map after calculation in the calculation example using the conventional frequency ratio method and point landslide dataset



Settings for the calculation example using the conventional method and point landslide dataset

3.2.4 Exploring output files

■ After the implementation of the previous four calculation examples, ***totally 110 files were output in the folder "Results"***, except for those three files copied from folder "Basics" ("Extent.shp", "Landslides_PGN.shp" and "Landslides_PNT.shp"). They are listed and explained below.

■ **Two (2) files for processing extent**

● ***Extent_cs100.tif***

➢ An extent raster converted from the original extent polygon "Extent.shp", with a spatial resolution of 100m. Only if the checkBox "Use the Geometry of the Input Feature as the Processing Extent" is checked, this file will be output.

● ***Extent_cs100.shp***

➢ An extent polygon converted from the derived extent raster, which is commonly different from the original extent polygon "Extent.shp" and be output only if "Use the Geometry of the Input Feature as the Processing Extent" is checked.

- ***The extent raster "Extent_cs100.tif" will be used by all the four calculation examples for frequency ratio statistic and as a geometrical extent for clipping the resampled factor rasters***, since all the four calculations use the geometry of "Extent.shp" as the processing extent. That is to say, the derived and output "Extent_cs100.tif" and "Extent_cs100.shp" in "Calculation01" can be directly used by the other three calculations. Those steps for deriving "Extent_cs100.tif" and "Extent_cs100.shp" were skipped during "Calculation02", "Calculation03" and "Calculation04".

■ **Two (2) raster files for 2 landslide dataset**

- ***Extent_cs100_Landslides_PGN.tif***
  - Landslide raster derived from landslide polygon shapefile "Landslides_PGN.shp", with a spatial resolution of 100m. ***The raster value of each cell is the absolute landslide area enveloped by that 100m × 100m cell***. As a result, the maximum raster value of a cell is 10000 in this case. Those raster cells not covered by landslides have no data.
- ***Extent_cs100_Landslides_PNT.tif***
  - Landslide raster derived from landslide point shapefile "Landslides_PNT.shp", with a spatial resolution of 100m. If no weight field for landslide points is set, ***the raster value of each cell will be the count of landslide points located in that 100m × 100m cell***. If a weight field for landslide points is set, ***the raster value of each cell will be the addition of the weights of all landslide points located in that 100m × 100m cell***. Those cells without a single landslide located in have no data.
- ***The landslide raster "Extent_cs100_Landslides_PGN.tif" will be used by both "Calculation01" and "Calculation03" for frequency ratio statistic***, while similarly ***the landslide raster "Extent_cs100_Landslides_PNT.tif" will be used by both "Calculation02" and "Calculation04" for frequency ratio statistic***. That is to say, those steps for deriving landslide raster were skipped during "Calculation03" and "Calculation04", since landslide rasters have been already derived from landslide polygon and landside point during "Calculation01" and "Calculation02", respectively.

■ **Twenty-eight (28) derived raster files for 7 continuous factors**

- The 4 derived raster files for factor "Continuous_Distance_Faults.tif" is presented for illustration. The meanings of those 24 derived raster files for the other 6 continuous factors are similar.
- ***Extent_cs100_Continuous_Distance_Faults.tif***
  - A raster file derived from the factor raster file "Continuous_Distance_Faults.tif" by firstly resampling it to 100m spatial resolution and then clipping it to the rectangular extent defined by the layer "Exent.shp".
- ***Extent_cs100_Continuous_Distance_Faults_Sorted.tif***
  - A raster file derived from "Extent_cs100_Continuous_Distance_Faults.tif" by sorting its cell values in the way that the smallest value locates in the left up corner while the largest value locates in the right down corner of the raster.

- ***Extent_cs100_Continuous_Distance_Faults_Index.tif***
  - The cell values of this raster file save the original indices of corresponding cells in "Extent_cs100_Continuous_Distance_Faults.tif" for each cell in the sorted raster file "Extent_cs100_Continuous_Distance_Faults_Sorted.tif".
- ***Extent_cs100_Continuous_Distance_Faults_Mask.tif***
  - A raster file derived from "Extent_cs100_Continuous_Distance_Faults.tif" by clipping it to the geometrical extent of the layer "Exent.shp". Only if the checkBox "Use the Geometry of the Input Feature as the Processing Extent" is checked, this file will be output.
- ***The derived raster files "Extent_cs100_Continuous_Distance_Faults_Sorted.tif" and "Extent_cs100_Continuous_Distance_Faults_Index.tif" will be used by both "Calculation01" and "Calculation02" for frequency ratio statistics***. That is to say, those steps for generating derived raster files for continuous factors were skipped during "Calculation02", since derived raster files have been already generated during "Calculation01".

- **Fourteen (14) derived raster files for 7 classified factors**
- The 2 derived raster files for factor "Classified05_Distance_Faults.tif" is presented for illustration. The meanings of those 12 derived raster files for the other 6 classified factors are similar.
- ***Extent_cs100_Classified05_Distance_Faults.tif***
  - A raster file derived from the factor raster file "Classified05_Distance_Faults.tif" by firstly resampling it to 100m spatial resolution and then clipping it to the rectangular extent defined by the layer "Exent.shp".
- ***Extent_cs100_Classified05_Distance_Faults_Mask.tif***
  - A raster file derived from "Extent_cs100_Classified05_Distance_Faults.tif" by clipping it to the geometrical extent of the layer "Exent.shp". Only if the checkBox "Use the Geometry of the Input Feature as the Processing Extent" is checked, this file will be output.
- ***The derived raster file "Extent_cs100_Classified05_Distance_Faults.tif" will be used by both "Calculation03" and "Calculation04" for frequency ratio statistics***. That is to say, those steps for generating derived raster files for classified factors were skipped during "Calculation04", since derived raster files have been already generated during "Calculation03".

- **Seven (7) frequency ratio raster files calculated using the modified frequency ratio method and polygon landslide dataset for 7 continuous factors**
- The 1 frequency ratio raster file for factor "Continuous_Distance_Faults.tif" is presented for illustration. The meanings of those 6 frequency ratio raster files for the other 6 continuous factors are similar.
- ***Extent_cs100_Landslides_PGN_pr4bw0.1_Continuous_Distance_Faults.tif***
  - A frequency ratio raster file for factor "Continuous_Distance_Faults.tif" clipped to the extent of "Exent.shp", with a 100m spatial resolution, and derived based on landslide dataset "Landslides_PGN.shp", a precision of 4 and a bin width of 0.1.

- **Seven (7) frequency ratio txt files calculated using the modified frequency ratio method and polygon landslide dataset for 7 continuous factors**
- The 1 frequency ratio txt file for factor “Continuous_Distance_Faults.tif” is presented for illustration. The meanings of those 6 frequency ratio txt files for the other 6 continuous factors are similar.
- ***Extent_cs100_Landslides_PGN_pr4bw0.1_Continuous_Distance_Faults.txt***
  - A txt file recording the frequency ratio values for every identical normalized factor values of factor “Continuous_Distance_Faults.tif”.
  - A row of this txt file records the data for an identical normalized factor value.
  - The 1st column records the original factor values for every identical normalized factor values.
  - The 2nd column records the areas of landslides located in the raster cells with factor values within the predefined neighborhood of the corresponding identical normalized factor values.
  - The 3rd column records the total area of landslides located in the whole processing extent.
  - The 4th column records the frequency of landslides located in the raster cells with factor values within the predefined neighborhood of the corresponding identical normalized factor values compared with landslides located in the whole processing extent, which are the ratios of the values in the 2nd column and the values in the 3rd column.
  - The 5th column records the counts of raster cells with factor values within the predefined neighborhood of the corresponding identical normalized factor values.
  - The 6th column records the counts of raster cells in the whole processing extent.
  - The 7th column records the frequency of raster cells with factor values within the predefined neighborhood of the corresponding identical normalized factor values compared with raster cells in the whole processing extent, which are the ratios of the values in the 5nd column and the values in the 6rd column.
  - The 8th column records the frequency ratios for identical normalized factor values, which are the ratios of the values in the 4th column and the values in the 7th column.
  - A “Factor value v.s. Frequency ratio” plot can be made based on the 1st column and the 8th column in this txt file, to inspect the variation of landslide susceptibility with factor values.

- **Seven (7) frequency ratio raster files calculated using the modified frequency ratio method and point landslide dataset for 7 continuous factors**
- The 1 frequency ratio raster file for factor “Continuous_Distance_Faults.tif” is presented for illustration. The meanings of those 6 frequency ratio raster files for the other 6 continuous factors are similar.
- ***Extent_cs100_Landslides_PNT_pr4bw0.1_Continuous_Distance_Faults.tif***
  - A frequency ratio raster file for factor “Continuous_Distance_Faults.tif” clipped to the extent of “Exent.shp”, with a 100m spatial resolution, and derived based on landslide dataset “Landslides_PNT.shp”, a precision of 4 and a bin width of 0.1.

■ **Seven (7) frequency ratio txt files calculated using the modified frequency ratio method and point landslide dataset for 7 continuous factors**

● The 1 frequency ratio txt file for factor "Continuous_Distance_Faults.tif" is presented for illustration. The meanings of those 6 frequency ratio txt files for the other 6 continuous factors are similar.

● ***Extent_cs100_Landslides_PNT_pr4bw0.1_Continuous_Distance_Faults.txt***

➢ The columns in this txt file records the same data for identical normalized factor values as the txt file recording the frequency ratios derived using landslide polygon shapefile, except the 2nd column and the 3rd column.

➢ The 2nd column records the counts (or weighted counts) of landslides located in the raster cells with factor values within the predefined neighborhood of the corresponding identical normalized factor values.

➢ The 3rd column records the total count (or weighted count) of landslides located in the whole processing extent.

■ **Seven (7) frequency ratio raster files calculated using the conventional frequency ratio method and polygon landslide dataset for 7 continuous factors**

● The 1 frequency ratio raster file for factor "Classified05_Distance_Faults.tif" is presented for illustration. The meanings of those 6 frequency ratio raster files for the other 6 classified factors are similar.

● ***Extent_cs100_Landslides_PGN_classified_Classified05_Distance_Faults.tif***

➢ A frequency ratio raster file for factor "Classified05_Distance_Faults.tif" clipped to the extent of "Exent.shp", with a 100m spatial resolution, and derived based on landslide dataset "Landslides_PGN.shp". "classified" is inserted in the file name to indicate that the conventional frequency ratio method is used as the function played by "pr4bw0.1" in the cases using the modified frequency ratio method.

■ **Seven (7) frequency ratio txt files calculated using the conventional frequency ratio method and polygon landslide dataset for 7 continuous factors**

● The 1 frequency ratio txt file for factor "Classified05_Distance_Faults.tif" is presented for illustration. The meanings of those 6 frequency ratio txt files for the other 6 classified factors are similar.

● ***Extent_cs100_Landslides_PGN_classified_Classified05_Distance_Faults.txt***

➢ The columns in this txt file records the same data for identical factor classes as the txt file recording the frequency ratios derived using the modified frequency ratio method, except the 1st column.

➢ The 1st column records the identical factor classes.

■ **Seven (7) frequency ratio raster files calculated using the conventional frequency ratio method and point landslide dataset for 7 classified factors**

● The 1 frequency ratio raster file for factor "Classified05_Distance_Faults.tif" is presented for illustration. The meanings of those 6 frequency ratio raster files for the other 6 classified factors are similar.

- ***Extent_cs100_Landslides_PNT_classified_Classified05_Distance_Faults.tif***
  - A frequency ratio raster file for factor “Classified05_Distance_Faults.tif” clipped to the extent of “Exent.shp”, with a 100m spatial resolution, and derived based on landslide dataset “Landslides_PNT.shp”. “classified” is inserted in the file name to indicate that the conventional frequency ratio method is used as the function played by “pr4bw0.1” in the cases using the modified frequency ratio method.

■ **Seven (7) frequency ratio txt files calculated using the conventional frequency ratio method and point landslide dataset for 7 classified factors**

- The 1 frequency ratio txt file for factor “Classified05_Distance_Faults.tif” is presented for illustration. The meanings of those 6 frequency ratio txt files for the other 6 classified factors are similar.
- ***Extent_cs100_Landslides_PNT_classified_Classified05_Distance_Faults.txt***
  - The columns in this txt file records the same data for identical factor classes as the txt file recording the frequency ratios derived using the modified frequency ratio method, except the 1st column.
  - The 1st column records the identical factor classes.

■ **Four (4) landslide susceptibility raster files**

- Only 1 landslide susceptibility raster file is presented for illustration since the naming strategies for all 4 landslide susceptibility raster files are the same.
- ***Extent_cs100_Landslides_PGN_(0)classified_(7)pr4bw0.1.tif***
  - The landslide susceptibility raster file for the extent of “Exent.shp”, with a 100m spatial resolution, which was derived based on landslide dataset “Landslides_PGN.shp”, 0 classified factors and 7 continuous factors. The precision and bin width for calculating the frequency ratios for continuous factors are 4 and 0.1, respectively.

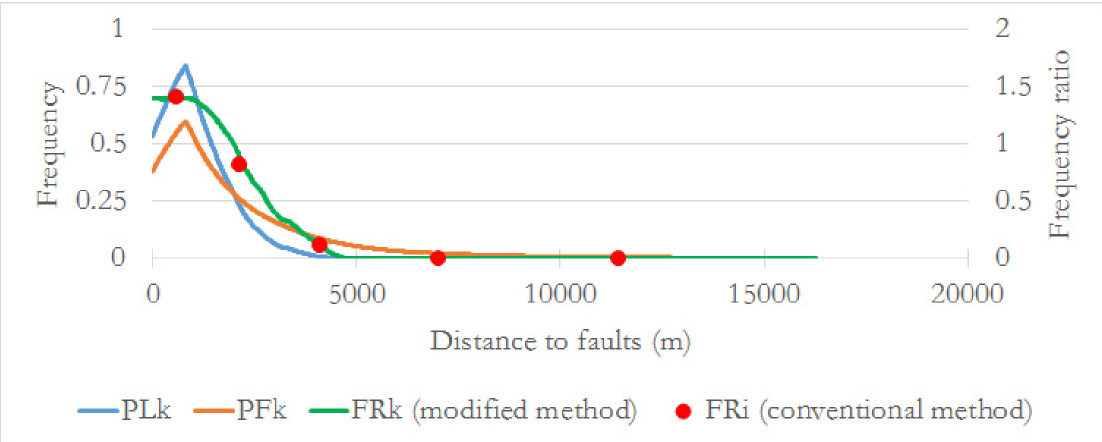
■ **Four (4) calculation logging txt files**

- Only 1 calculation logging txt file is presented for illustration since the naming strategies and logging items for all 4 calculation logging txt files are the same.
- ***Extent_cs100_Landslides_PGN_(0)classified_(7)pr4bw0.1.txt***
  - A txt file logging the calculation of the landslide susceptibility raster file for the extent of “Exent.shp”, with a 100m spatial resolution, which was derived based on landslide dataset “Landslides_PGN.shp”, 0 classified factors and 7 continuous factors. The precision and bin width for calculating the frequency ratios for continuous factors are 4 and 0.1, respectively.
  - “Input Summary” records the calculation settings.
  - Cost times and memories are recorded for every calculation steps. The 1st column records the start times of calculation steps. The 2nd column records the cost times of all calculation steps until the current calculation steps. The 3rd column records the cost times of the current calculation steps. The 4th column records the total physical memory of the machine. The 5th column records used memory. The 6th columns records the change of used memory compared the previous step.

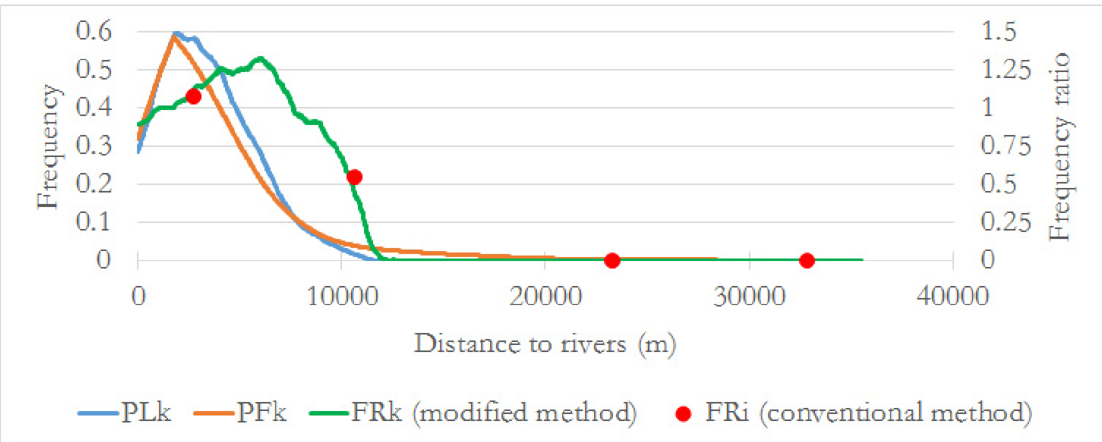
3.3 Result comparisons

3.3.1 Qualitative comparison

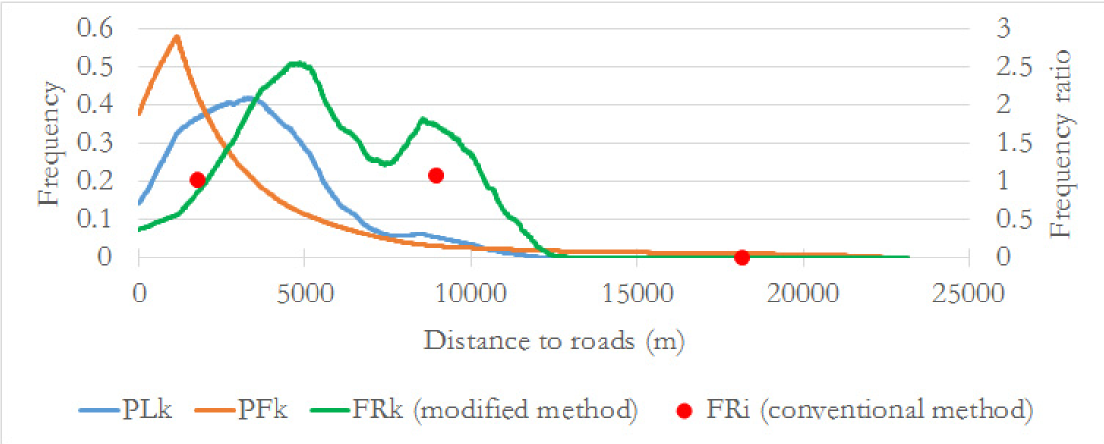
The frequency ratios of different factors calculated using the modified frequency ratio method and the conventional frequency ratio method are shown in the following figures, in which calculation examples using polygon landslide dataset are taken as examples. The results show that the frequency ratios of different factors calculated using the modified method and those calculated using the conventional method have similar fluctuating patterns. However, ***the results calculated using the modified method provide more detailed fluctuations of frequency ratio compared with the results calculated using the conventional method***. The conventional method gives less than 5 frequency ratio values for distance to rivers and distance to roads, because it may happen that factor classes located in the rectangular extent of “Frame.shp” do not present in the geometry extent of the layer “Extent.shp”.



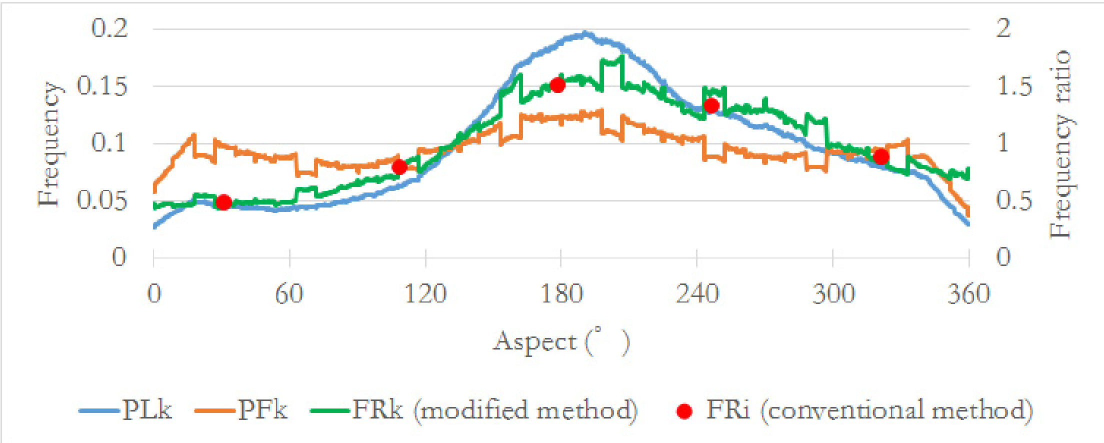
The frequency ratios of distance to faults calculated using the modified frequency ratio method and the conventional frequency ratio method



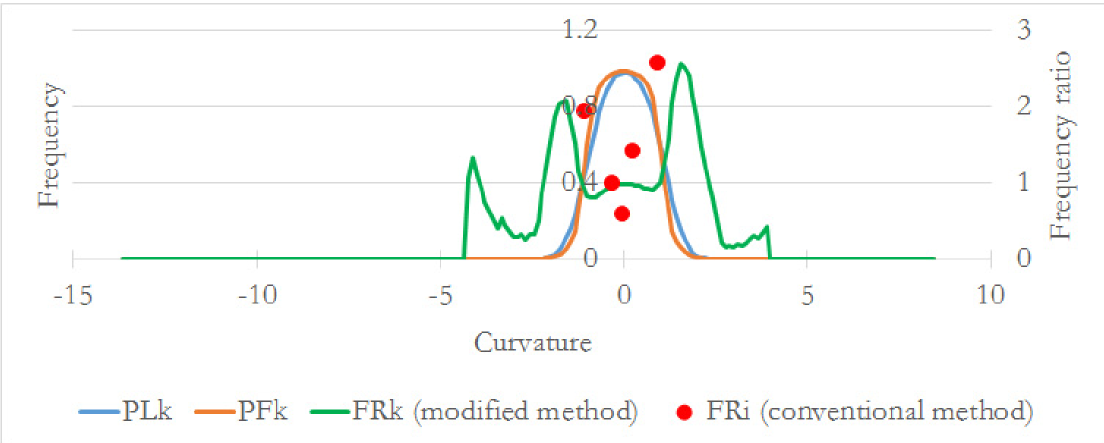
The frequency ratios of distance to rivers calculated using the modified frequency ratio method and the conventional frequency ratio method



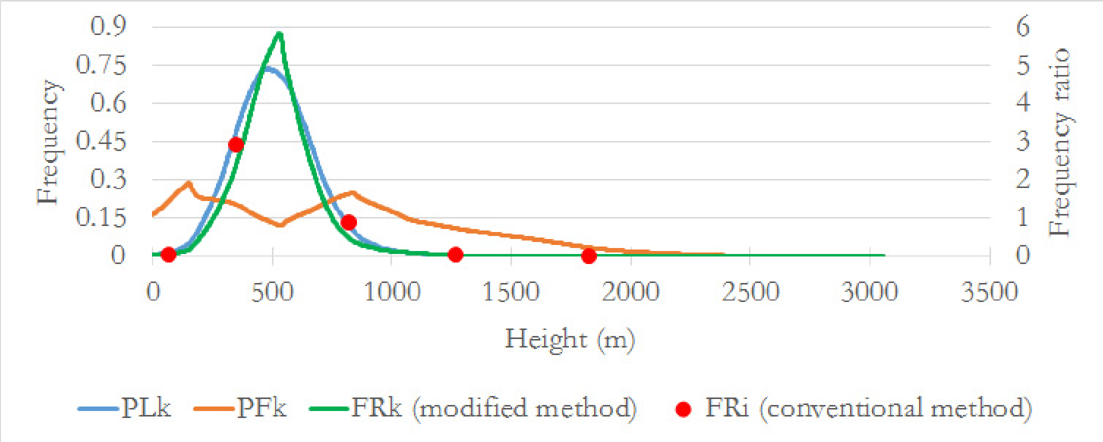
The frequency ratios of distance to roads calculated using the modified frequency ratio method and the conventional frequency ratio method



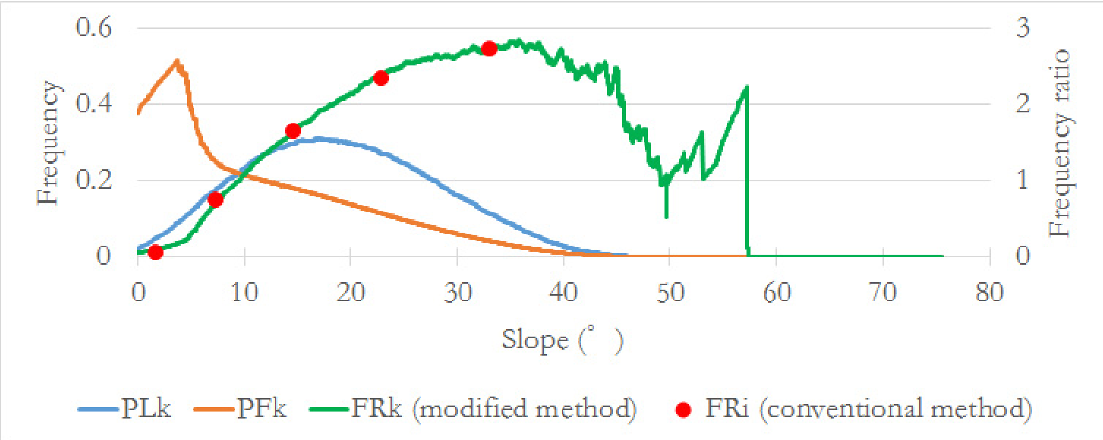
The frequency ratios of aspect calculated using the modified frequency ratio method and the conventional frequency ratio method



The frequency ratios of curvature calculated using the modified frequency ratio method and the conventional frequency ratio method



The frequency ratios of height calculated using the modified frequency ratio method and the conventional frequency ratio method

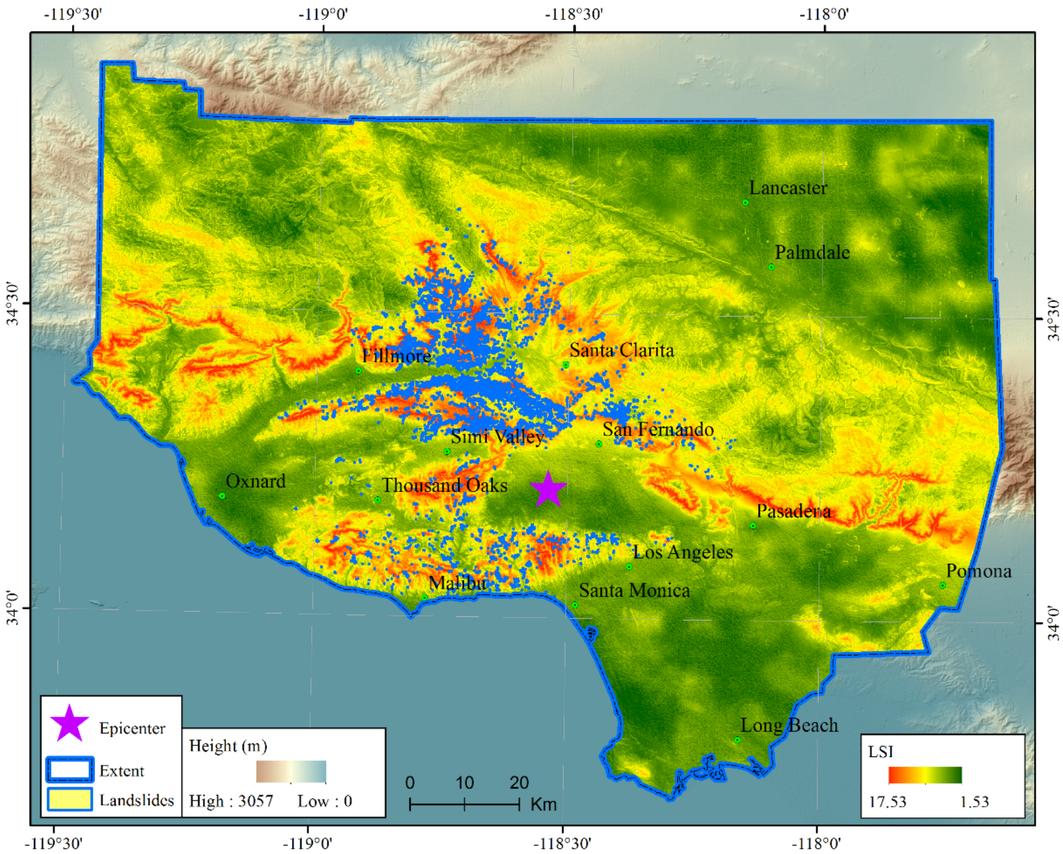


The frequency ratios of slope calculated using the modified frequency ratio method and the conventional frequency ratio method

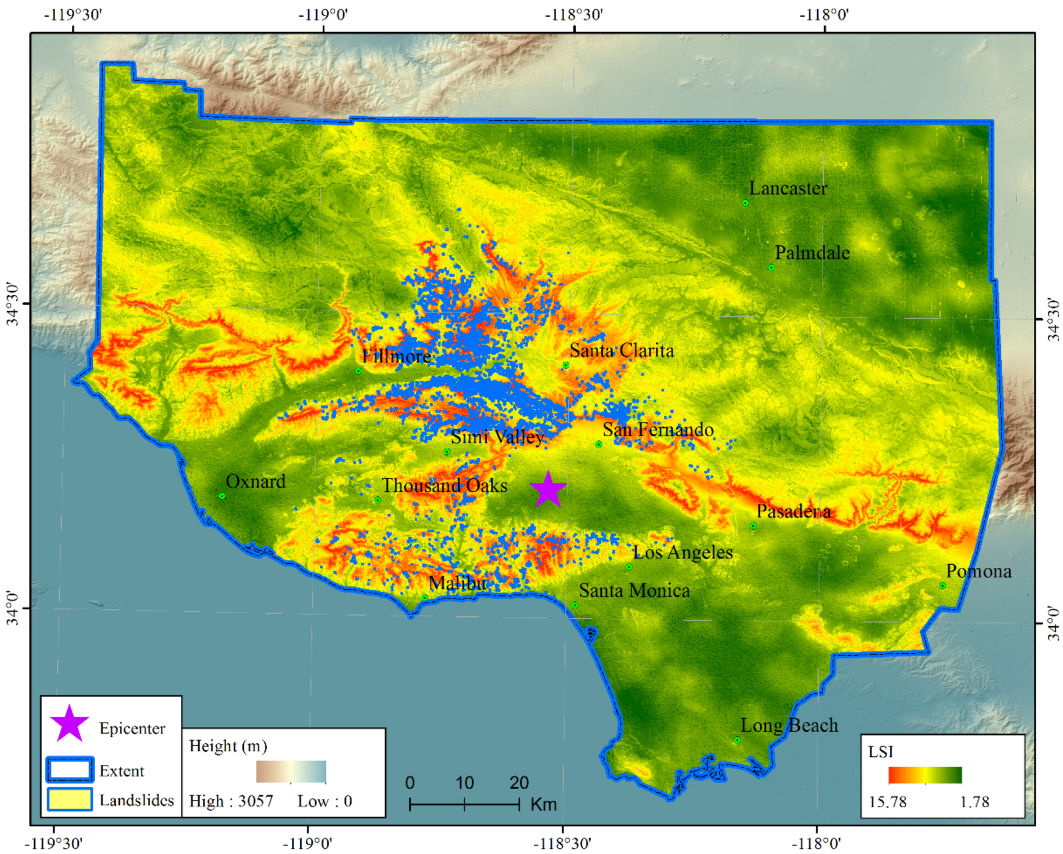
The differences in resolutions of frequency ratio curves will eventually be reflected in the spatial characteristics of landslide susceptibility maps. The landslide susceptibility maps derived in the previous four examples are shown in the following figures. From a qualitative point of view, the results calculated using polygon landslide dataset and point landslide dataset are not visually distinguishable. However, the differences between the results calculated using the modified frequency ratio method and the conventional frequency ratio method are visually remarkable. ***The high resolutions of the frequency ratios calculated using the modified method significantly reduced the discontinuity of the spatial distribution (patchy effects) of landslide susceptibility***.

3.3.2 Quantitative comparison

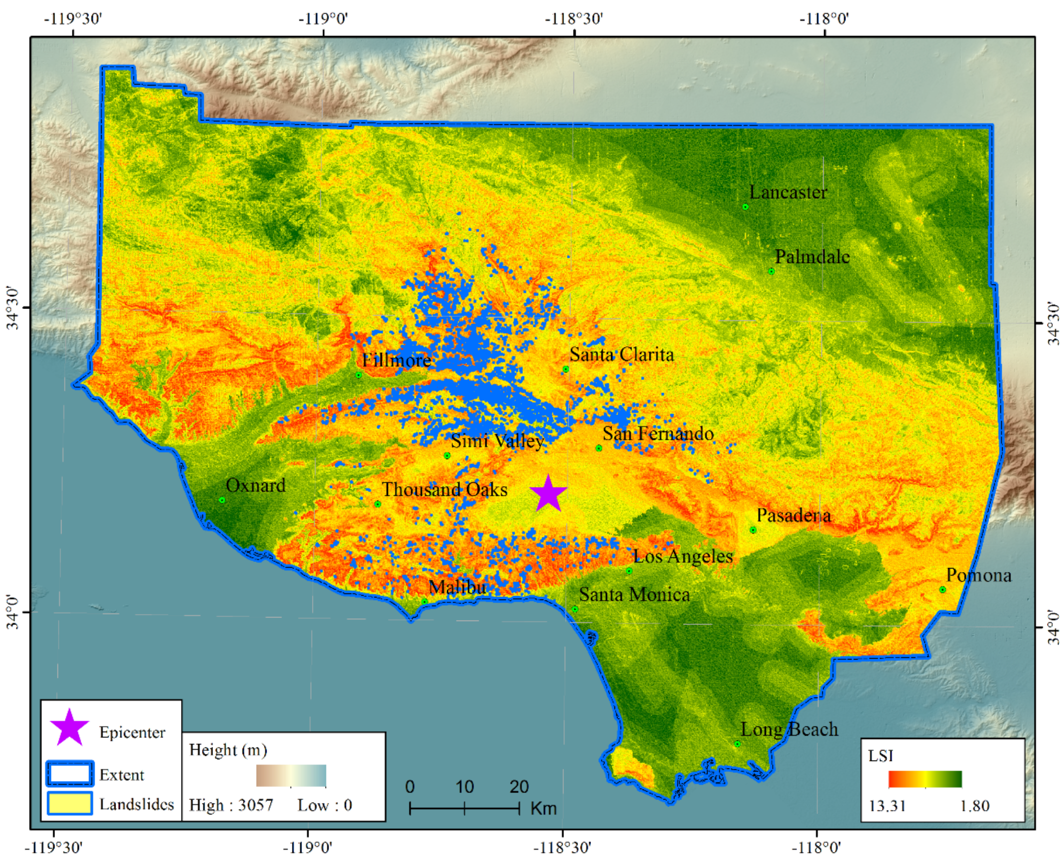
The performance of susceptibility assessments can be quantitatively evaluated using a receiver operating characteristic (ROC) curve. A model with larger AUC (Area under the ROC Curve) performs better. The following ROC curves show that ***the modified method outperforms the conventional method from a quantitative point of view***.



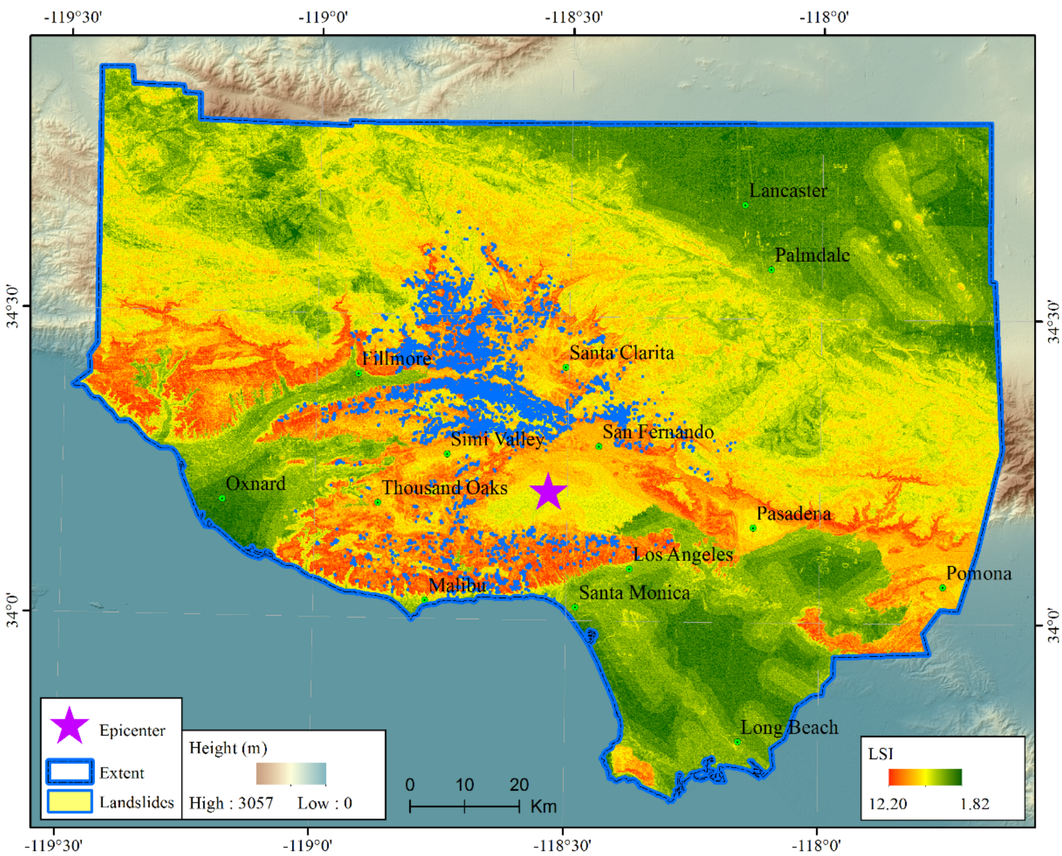
Landslide susceptibility map derived using polygon landslide dataset



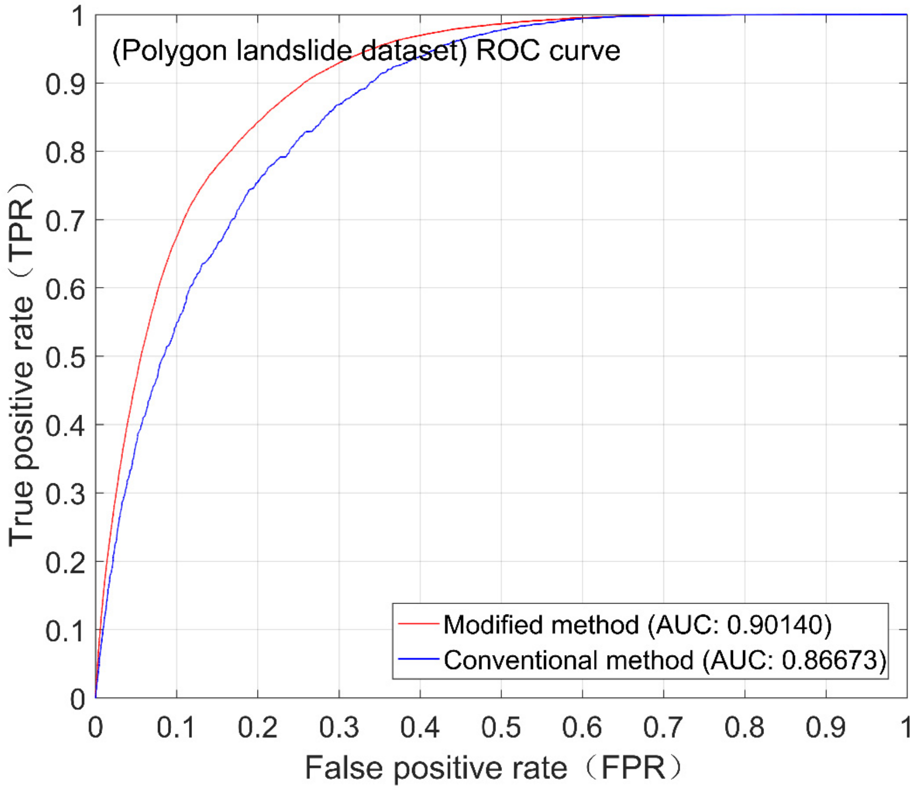
Landslide susceptibility map derived using point landslide dataset



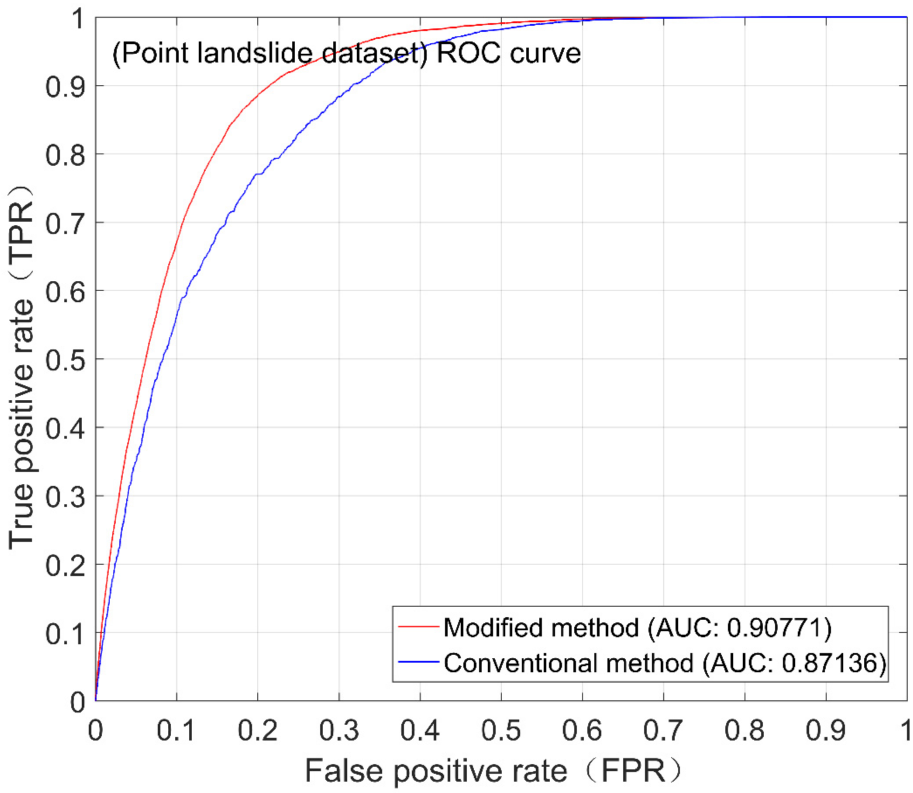
Landslide susceptibility map derived using the conventional method and polygon landslide dataset



Landslide susceptibility map derived using the conventional method and point landslide dataset



ROC curves for the landslide susceptibility models derived using polygon landslide dataset



ROC curves for the landslide susceptibility models derived using point landslide dataset