Using python to automatically draw the landslide susceptibility map of earthquake-induced landslides

Presenter: 沈楷庭(Kai-Ting SHEN)
Advisor: Pro. Jia-Jyun Dong
Date: 2022 / 09 / 30

Outline

1. Introduction
2. Methodology
3. Result
4. Conclusion
5. Future Work

1. Introduction

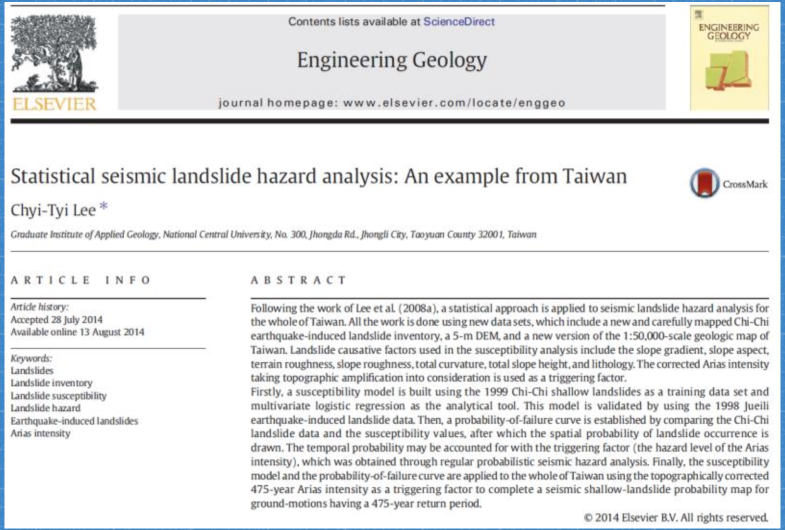
My target

Statistical seismic landslide hazard analysis: An example from Taiwan (Chyi-Tyi Lee,2014)

Case : CHI-CHI earthquake

Construct a Logistic regression model to draw susceptibility map

Use same factor and case to get the same result with Pro. Lee


Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

Statistical seismic landslide hazard analysis: An example from Taiwan

Chyi-Tyi Lee *

Graduate Institute of Applied Geology, National Central University, No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan

ARTICLE INFO

Article history:
Accepted 28 July 2014
Available online 13 August 2014

Keywords:
Landslides
Landslide inventory
Landslide susceptibility
Landslide hazard
Earthquake-induced landslides
Arias intensity

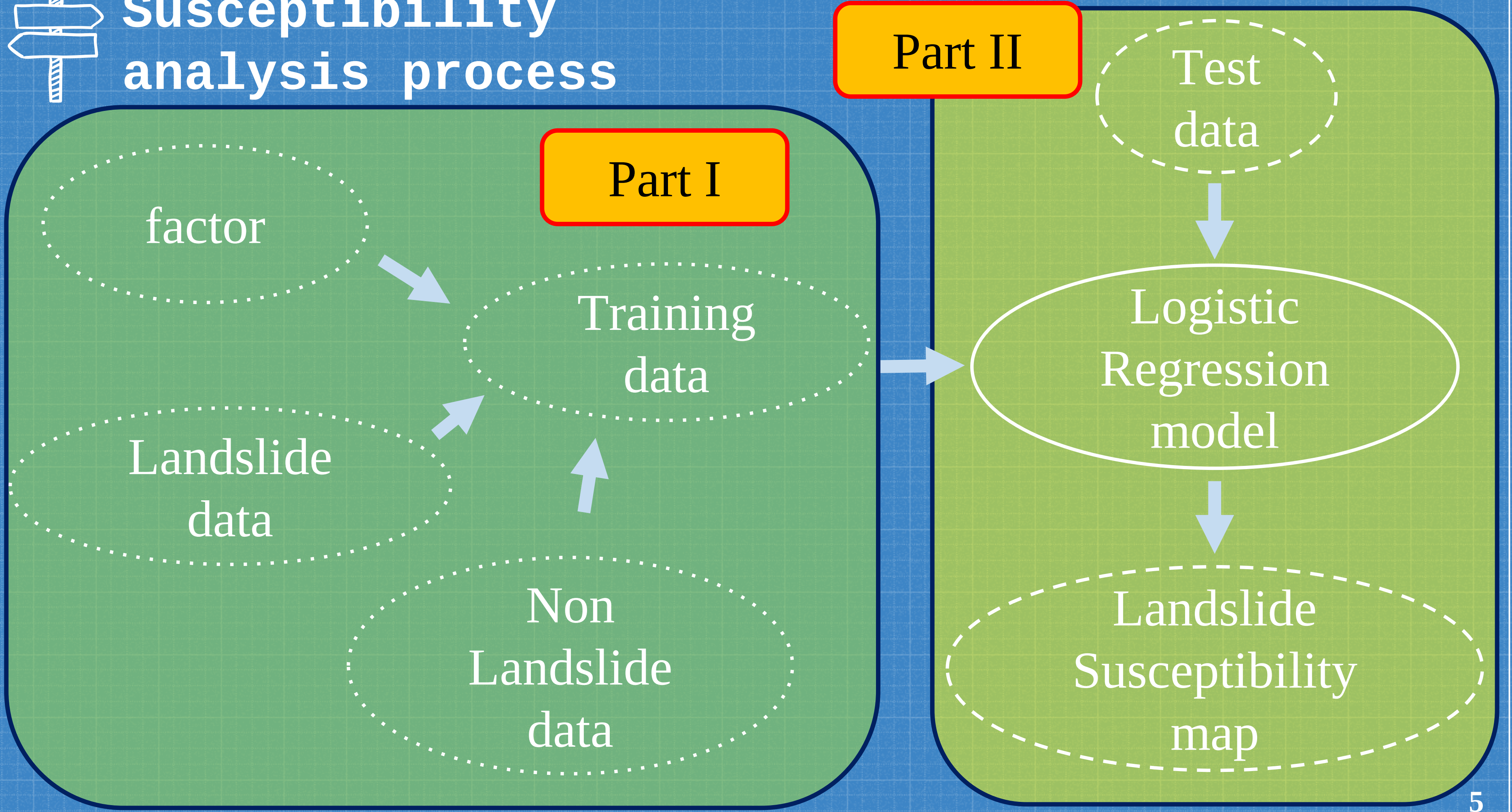
ABSTRACT

Following the work of Lee et al. (2008a), a statistical approach is applied to seismic landslide hazard analysis for the whole of Taiwan. All the work is done using new data sets, which include a new and carefully mapped Chi-Chi earthquake-induced landslide inventory, a 5-m DEM, and a new version of the 1:50,000-scale geologic map of Taiwan. Landslide causative factors used in the susceptibility analysis include the slope gradient, slope aspect, terrain roughness, slope roughness, total curvature, total slope height, and lithology. The corrected Arias intensity taking topographic amplification into consideration is used as a triggering factor.
Firstly, a susceptibility model is built using the 1999 Chi-Chi shallow landslides as a training data set and multivariate logistic regression as the analytical tool. This model is validated by using the 1998 Jueili earthquake-induced landslide data. Then, a probability-of-failure curve is established by comparing the Chi-Chi landslide data and the susceptibility values, after which the spatial probability of landslide occurrence is drawn. The temporal probability may be accounted for with the triggering factor (the hazard level of the Arias intensity), which was obtained through regular probabilistic seismic hazard analysis. Finally, the susceptibility model and the probability-of-failure curve are applied to the whole of Taiwan using the topographically corrected 475-year Arias intensity as a triggering factor to complete a seismic shallow-landslide probability map for ground-motions having a 475-year return period.

Factor

1. Slope(percentage)
2. Terrain roughness
3. Slope roughness
4. Total curvature
5. Total slope height
6. Arias Intensity
7. Aspect
8. Lithology

Susceptibility analysis process



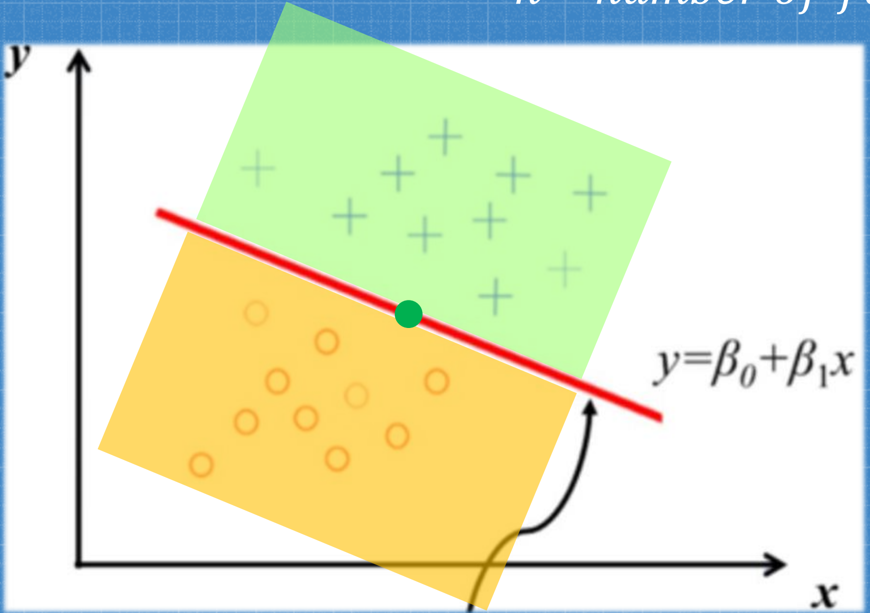
2. Methodology

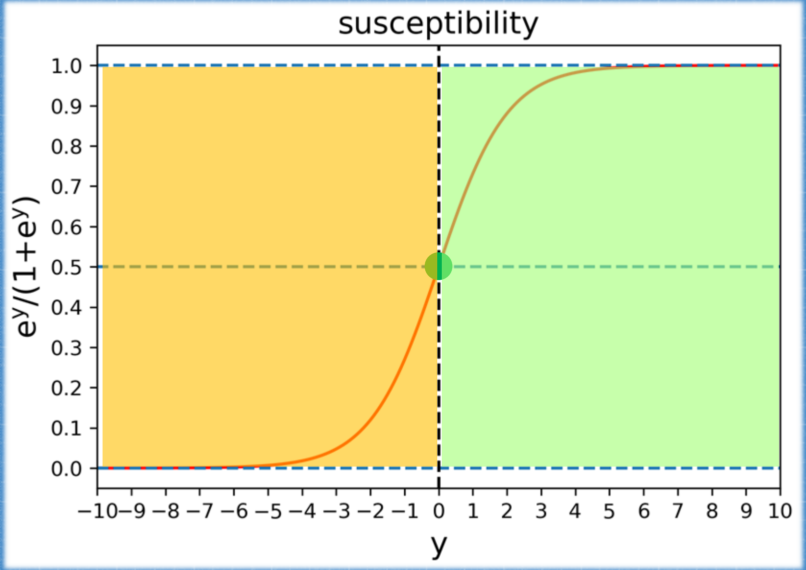
Logistic Regression

$$y = a_0 + \sum_{k=1}^{n} a_k x_k$$

a_0: *constant intercept*
a_k: *weight of factor*
x_k: *the value of factor*
n : *number of factor*

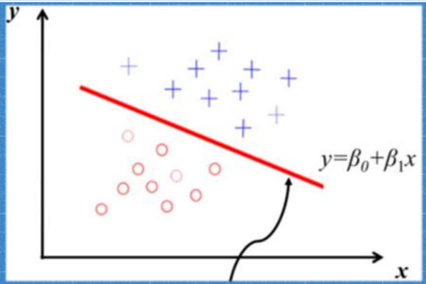
$$susceptibility = \frac{e^y}{1+e^y}$$

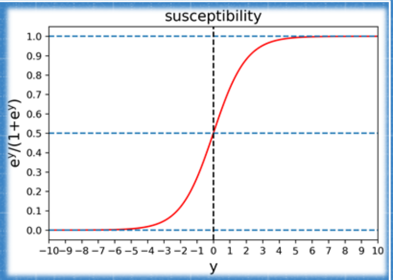




Find a line that distinguishes two groups

Logistic Regression





EX： *logistic regression model*:

$$y = 3x_1 - x_2 - 2$$

Point A : $x_1 = 2, x_2 = 3 \ \rightarrow y = 1$

Susceptibility A : $Sus = 0.73$

Point B : $x_1 = 1, x_2 = 1 \ \rightarrow y = 0$

Susceptibility B : $Sus = 0.5$

Point C : $x_1 = -1, x_2 = -4 \ \rightarrow y = -1$

Susceptibility C : $Sus = 0.27$

$$\boldsymbol{y = a_0 + \sum_{k=1}^{n} a_k x_k}$$

a_0: *constant intercept*

a_k: *weight of factor*

x_k: *the value of factor*

n : *number of factor*

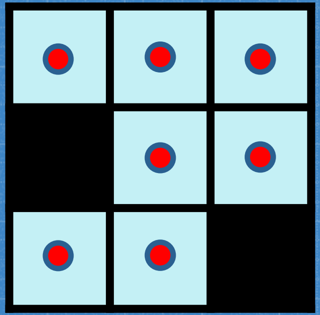
$$\boldsymbol{susceptibility = \frac{e^y}{1 + e^y}}$$

Training data selection

Landslide polygon

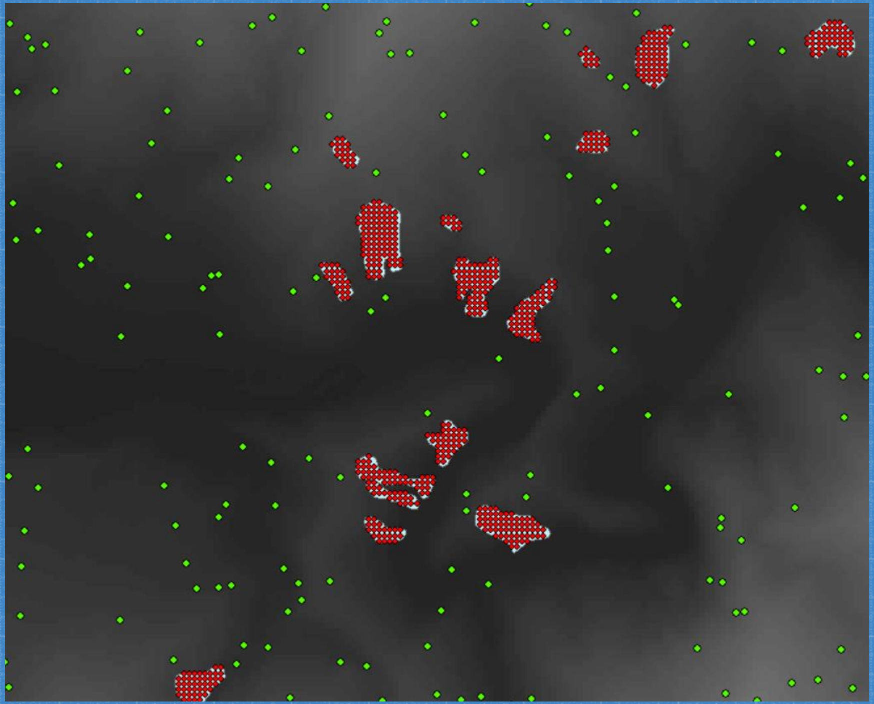
Landslide data

Non-Landslide data



<u>Landslide data :</u> total landslide grids

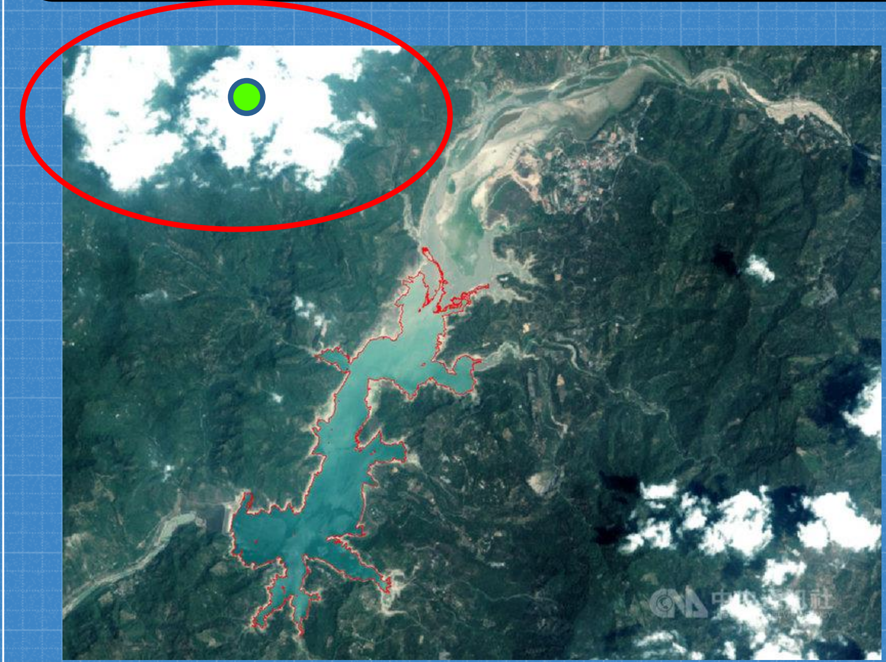
<u>Non-Landslide data :</u> randomly selected data of similar size with landslide data



Training data select

Important thing for Non-Landslide data

Delete data on : 1. Shallow area 2. Stable area 3. History landslide area



2. Stable area :
Slope gradient below 10% and
Continuation area above 1 hectare

3. History landslide area :
The geological conditions in
this area are unstable.

Arias Intensity

$$I_A = \frac{\pi}{2g}\int_0^{T_d} a(t)^2 dt$$

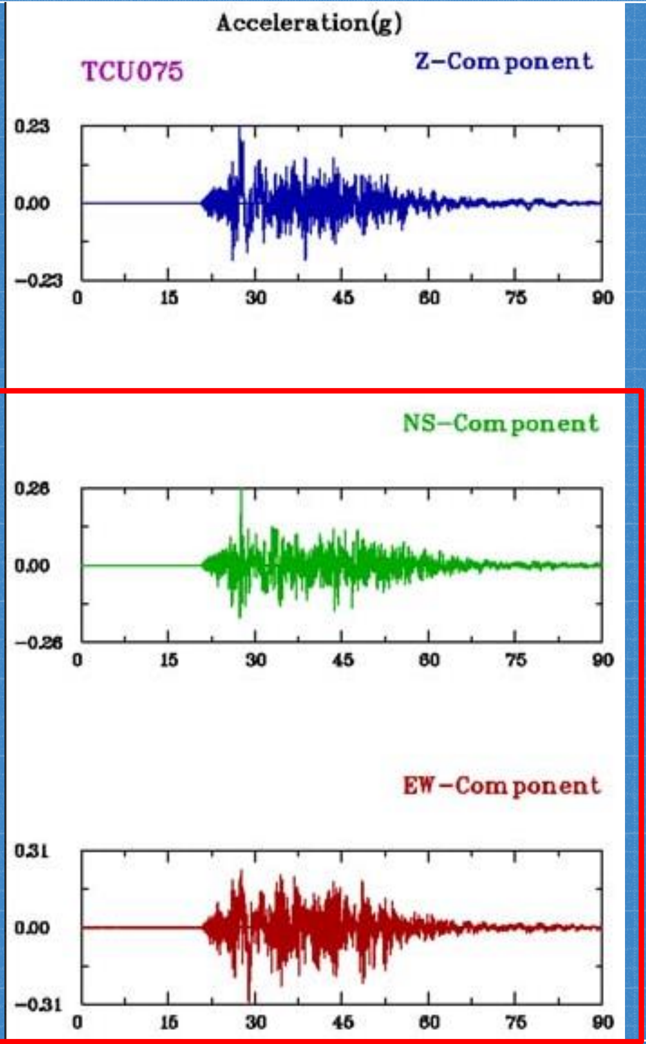
g : *Acceleration of Gravity*

T_d : *duration of signal*

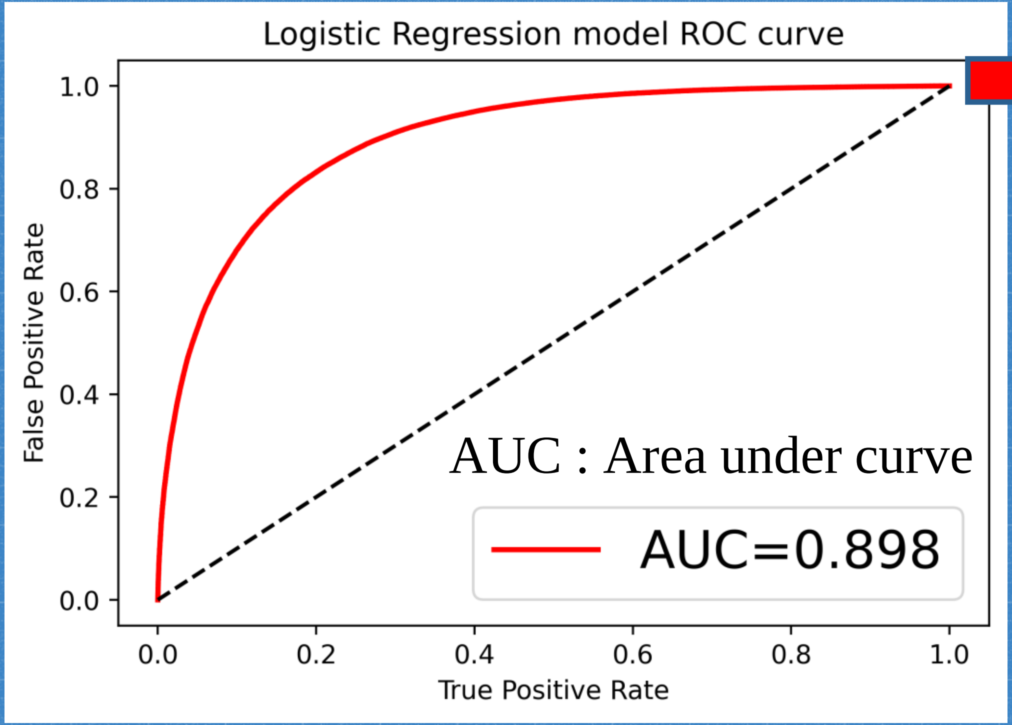
a : *earthquake acceleration*

Calculate the earthquake acceleration in the horizontal direction and take the geometric mean (Chyi-Tyi Lee,2014)

$$I = \sqrt{I_{NS} * I_{EW}}$$



AUC(Area Under Curve)

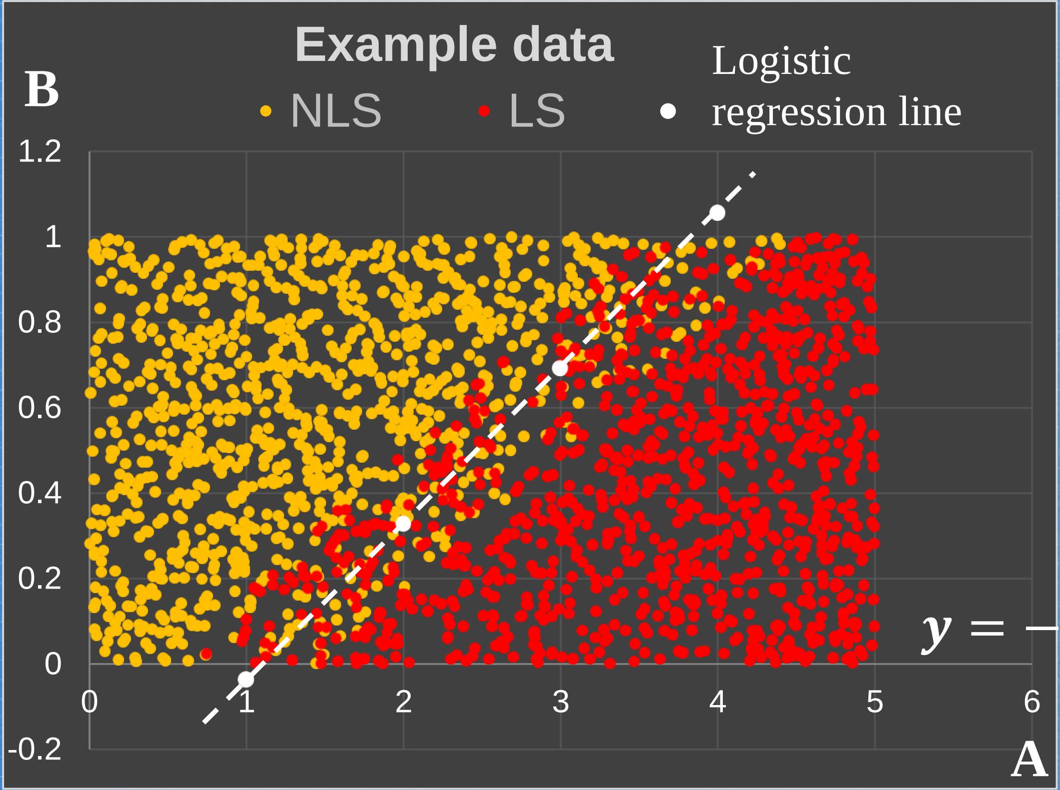


ROC :
Receiver Operating
Characteristic Curve

AUC>0.9 as excellent,
0.9 > AUC > 0.8 as good,
0.8 > AUC > 0.7 as fair,
0.7 > AUC > 0.6 as poor,
AUC < 0.6 as very poor

4. Result

Verify logistic model correctness



LS : 1003 cases
NLS : 997 cases

Coefficient:

A	B	Intercept
-2.95	8.1	3.247

Logistic Regression model:

$$y = -2.95 \times x_A + 8.1 \times x_B + 3.247$$

Model coefficient

L1	L2	L3	L4	L5	L6	L7	A1	A2	A3	A4
-0.273	0.068	0.395	0.674	0	-1.396	-0.341	-0.802	-0.139	0.260	0.472

A5	A6	A7	A8	F1	F2	F3	F4	F5	F6	C
0.428	0	-0.345	-0.956	0.782	0.393	0.198	0.129	0.263	0.914	-0.874

Lithology :

L_1: *Terrace deposits*
L_2: *Pleistocene series*
L_3: *Pliocene Series*
L_4: *Upper Miocene Series*
L_5: *Lower Miocene Series*
L_6: *Slate and schist*
L_7: *Quartzite and other hard rocks*

Aspect :

A_1: *N*
A_2: *NE*
A_3: E
A_4: SE
A_5: S
A_6: *SW*
A_7: *W*
A_8: *NW*

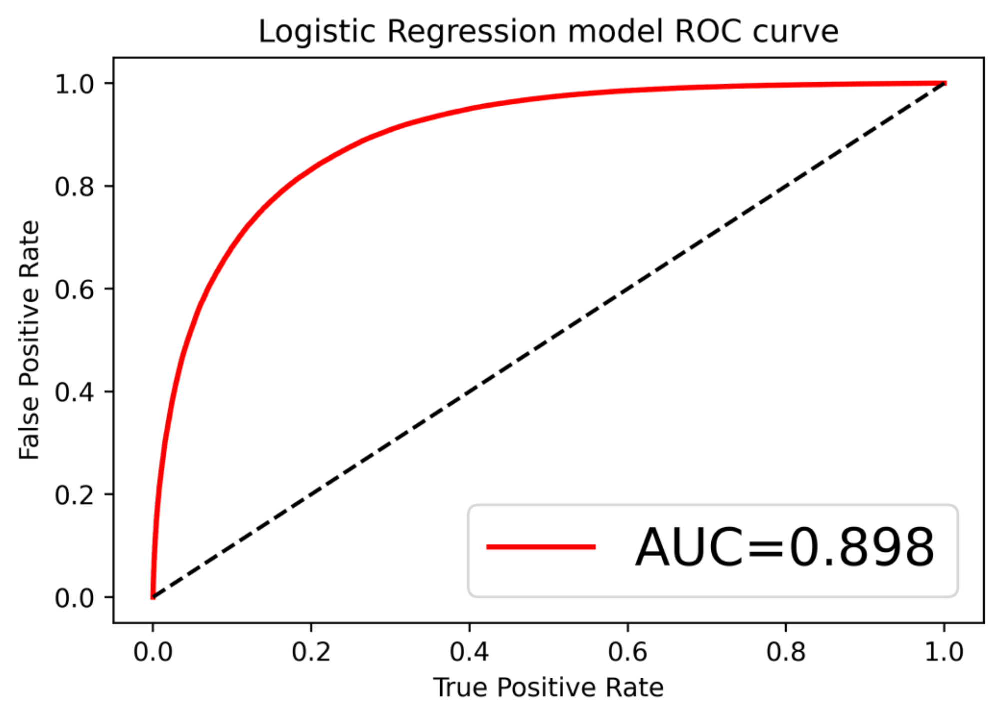
C : intercept

Factor:

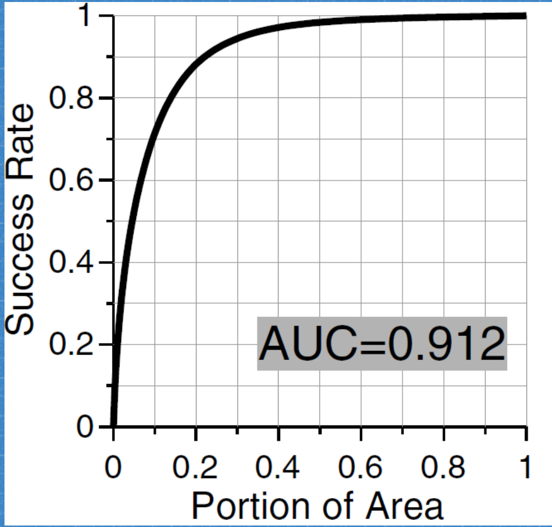
F_1: *Slope gradient*
F_2: *Terrain roughness*
F_3: *Slope roughness*
F_4: *Total curvature*
F_5: *Total slope height*
F_6: *Arias intensity*

Model AUC

AUC>0.9	as excellent,
0.9 > AUC > 0.8	as good,
0.8 > AUC > 0.7	as fair,
0.7 > AUC > 0.6	as poor,
AUC < 0.6	as very poor



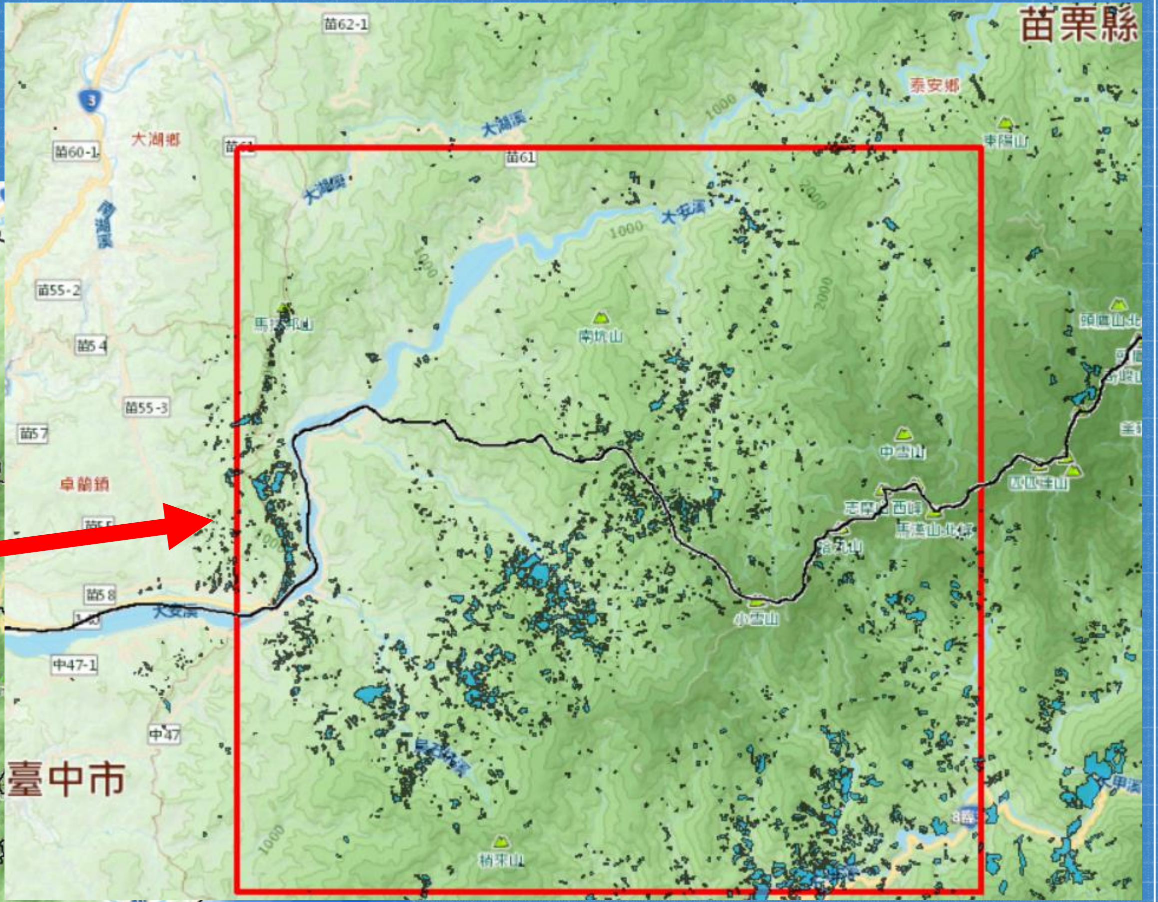
My logistic regression model AUC



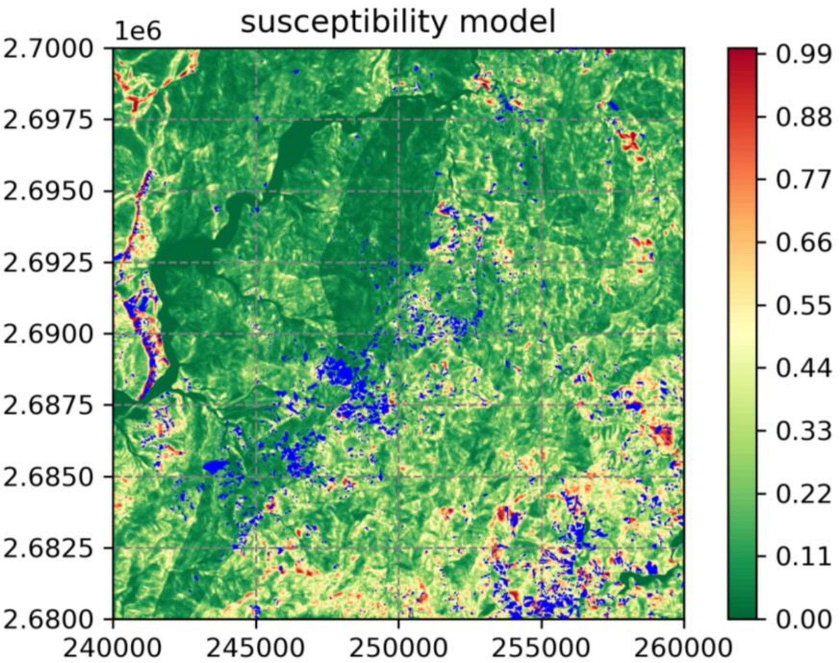
Pro. Lee AUC

Test Area


Miaoli county
Taichung county
新竹市
新竹縣
苗栗縣
臺中市
南投縣
彰化縣

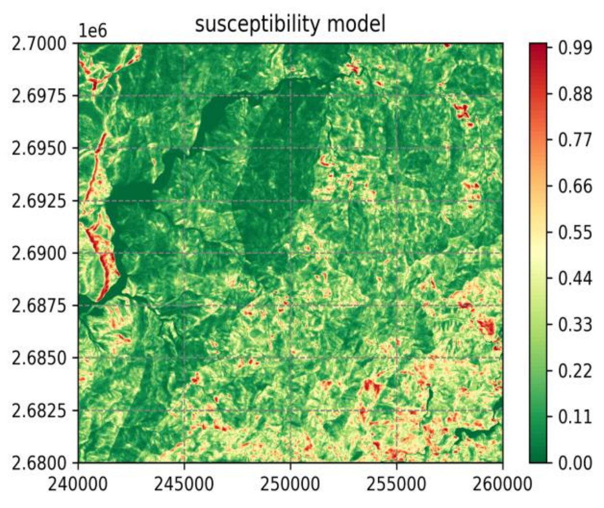

苗栗縣
臺中市
大湖鄉
卓蘭鎮
泰安鄉

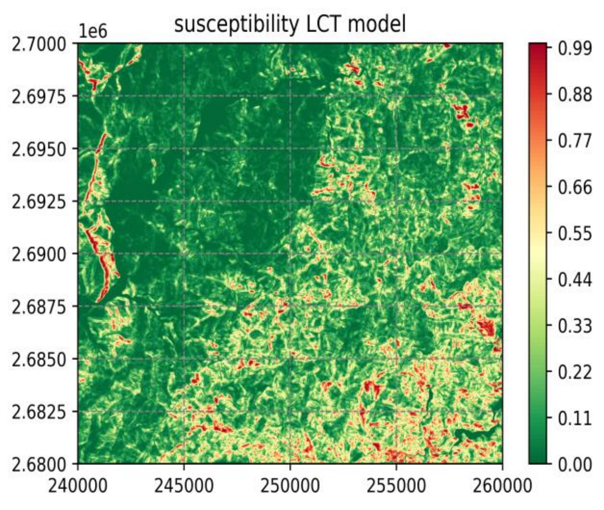
Susceptibility Map



Landslide polygon

My logistic regression model



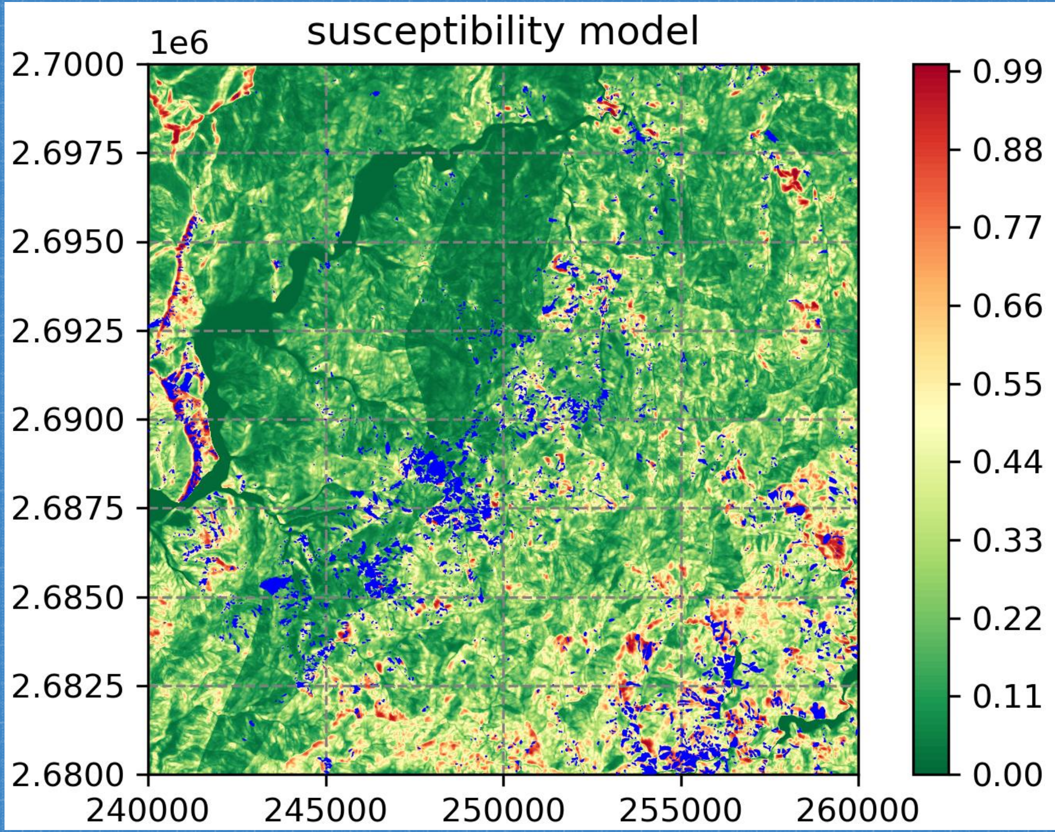


My

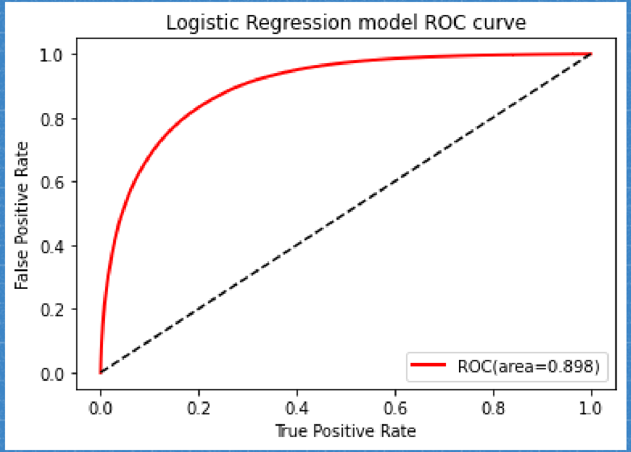
Pro. Lee

5. Conclusion

Conclusion



1. Can roughly find where the landslide is
2. The AUC of this model is 0.898



6. Future Work

Future Work

Convert it into automatically analysis and post on the website

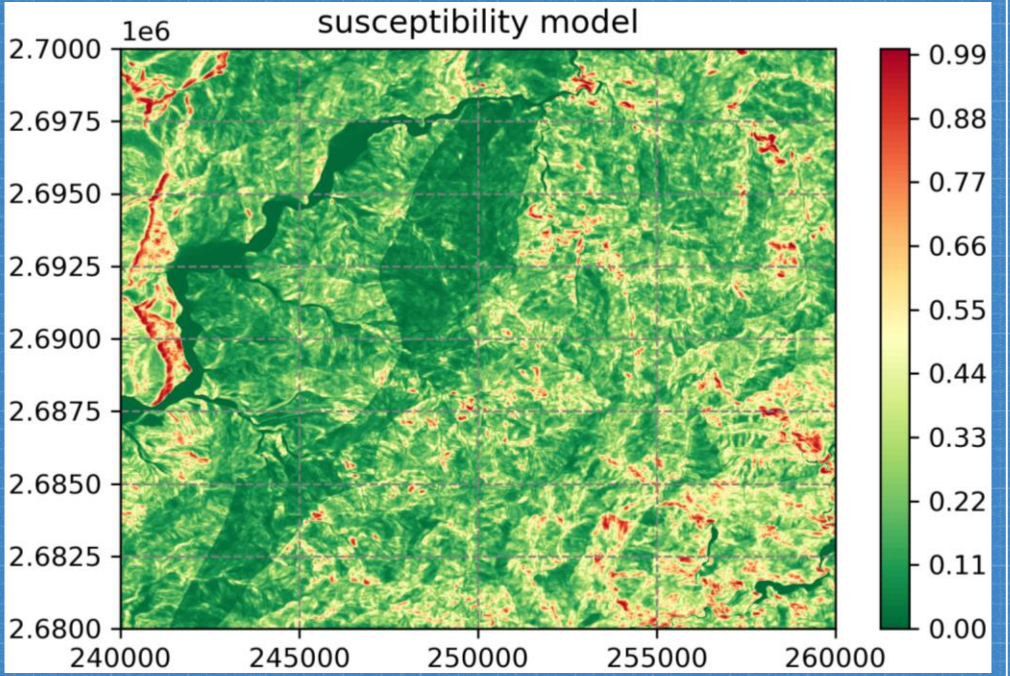
It will not trigger landslide before earthquake occurred

When Arias Intensity = 0,
susceptibility should Approach to 0

Turn coefficients into functions of AI

Original:	New:
F1	F1
0.765	$f(AI)$

※AI : Arias Intensity

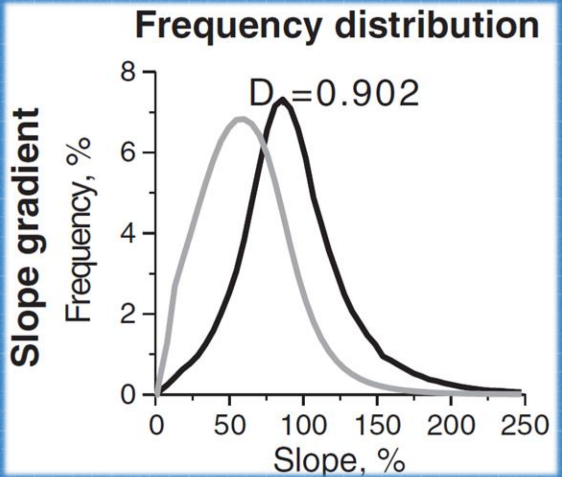


When Arias Intensity = 0

Thank you for listening

ANY QUESTIONS?

Factor Statistics

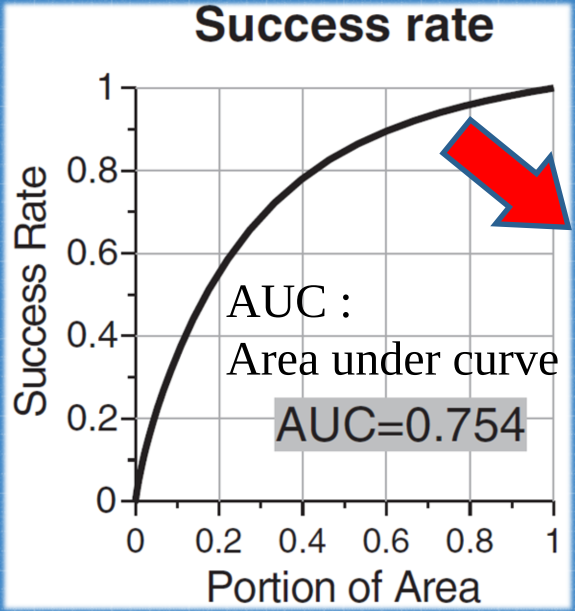


standardized difference D:

$$D_J = \frac{\overline{A_j} - \overline{B_j}}{S_{pj}}$$

$\overline{A_j}$: *mean of factor for landslide*
$\overline{B_j}$: *mean of factor for Non − landslide*
S_{pj}: *pooled standard deviation of j*

The larger the standardized difference, the more effective the factor



AUC : Area under curve

ROC : Receiver Operating Characteristic Curve

AUC>0.9	as excellent,
0.9 > AUC > 0.8	as good,
0.8 > AUC > 0.7	as fair,
0.7 > AUC > 0.6	as poor,
AUC < 0.6	as very poor

Test Area

沖積層

台地堆積層

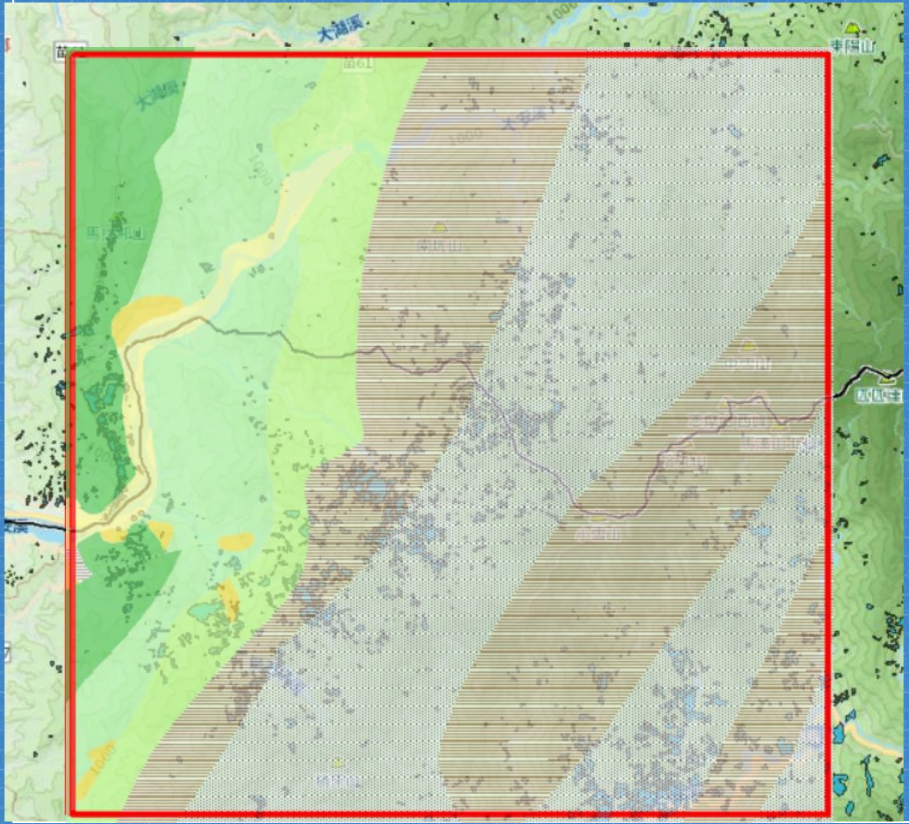
水長流層

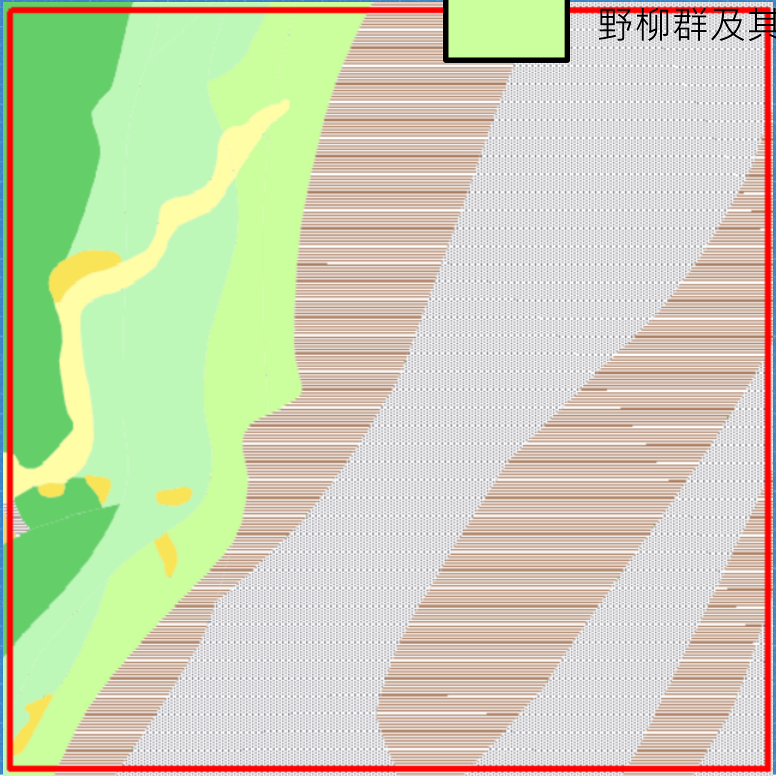
白冷層

桂竹林層及其相當地層

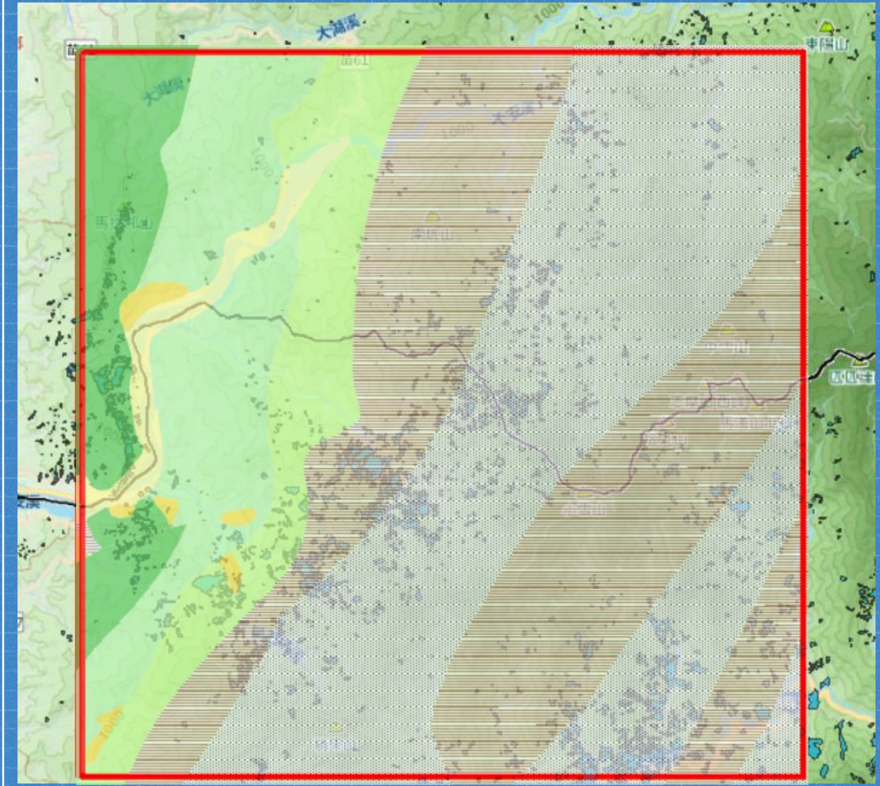
瑞芳群及其相當地層

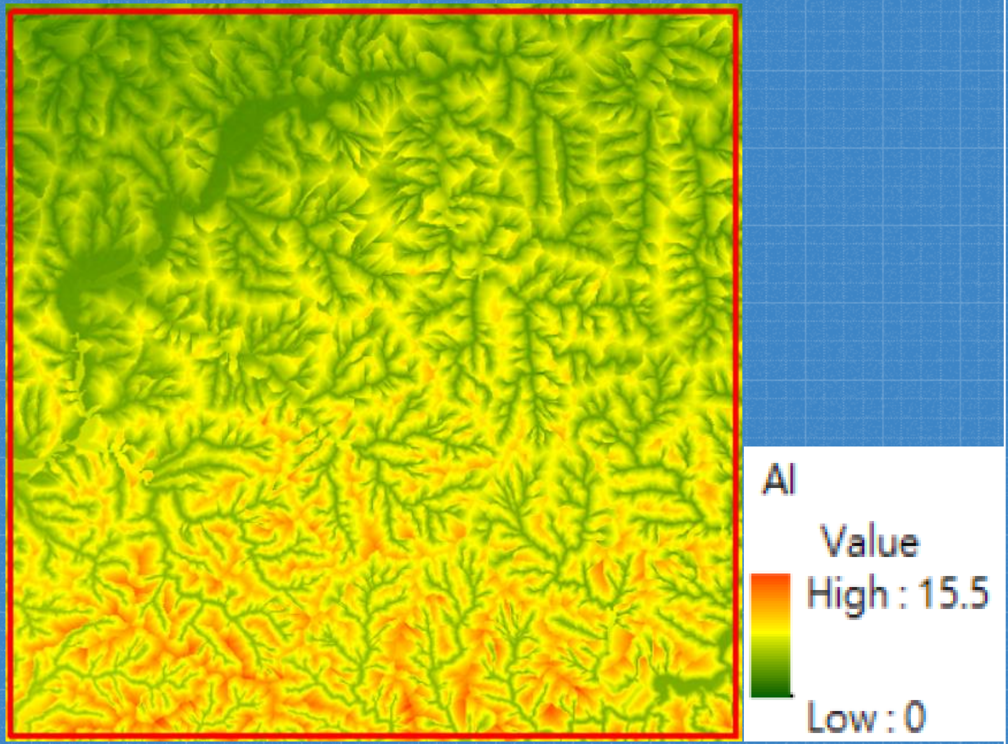
野柳群及其相當地層





Test Area





Model validation

Confusion matrix		
pre \ obs	1	0
1	TP	FP
0	FN	TN

TP : True Positive
TN : True Negative
FP : False Positive
FN : False Negative

Threshold = 0.5

Confusion matrix		
pre \ obs	1	0
1	112126	9407
0	105582	190364

Accuracy = $\frac{TP+TN}{Total}$ = 72.4%

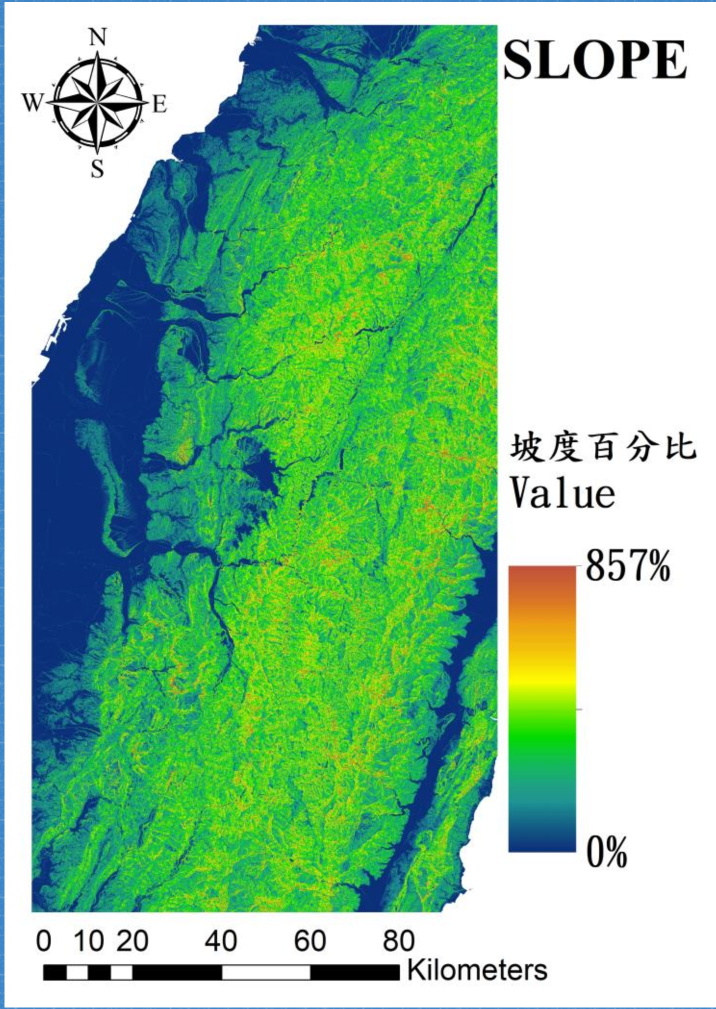
FNR = $\frac{FN}{TP+FN}$ = 48.5%

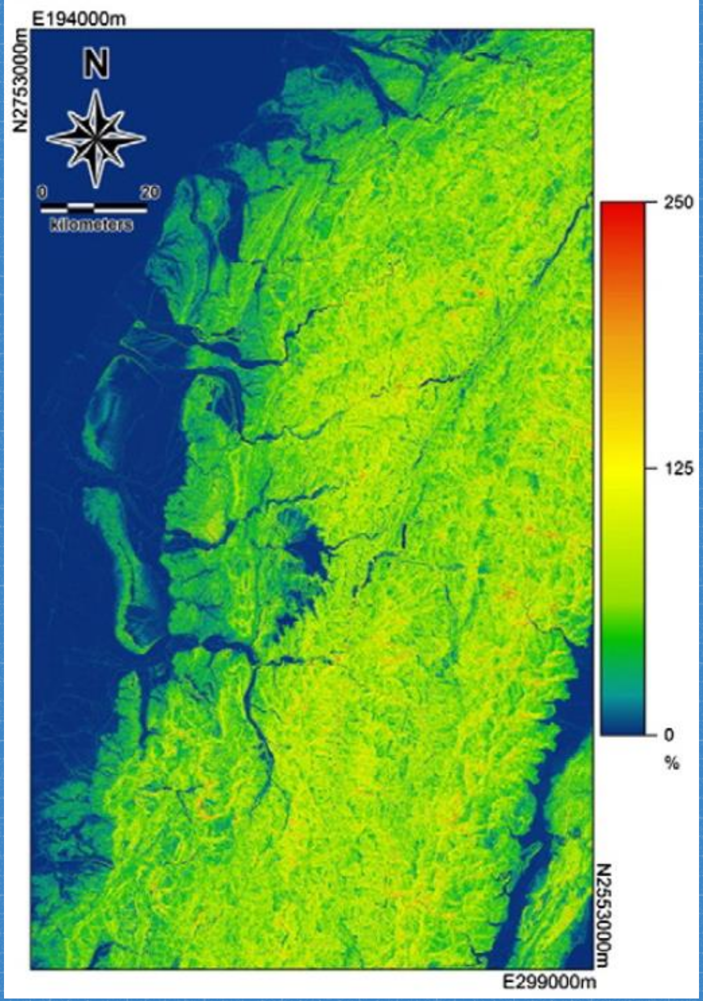
Threshold = 0.3

Confusion matrix		
pre \ obs	1	0
1	160034	25410
0	57674	174361

Accuracy = $\frac{TP+TN}{Total}$ = 80%

FNR = $\frac{FN}{TP+FN}$ = 26.5%


N
W
E
S
SLOPE
坡度百分比
Value
857%
0%
0 10 20 40 60 80
Kilometers


E194000m
N2753000m
N
0 20
Kilometers
250
125
0
%
N2553000m
E299000m

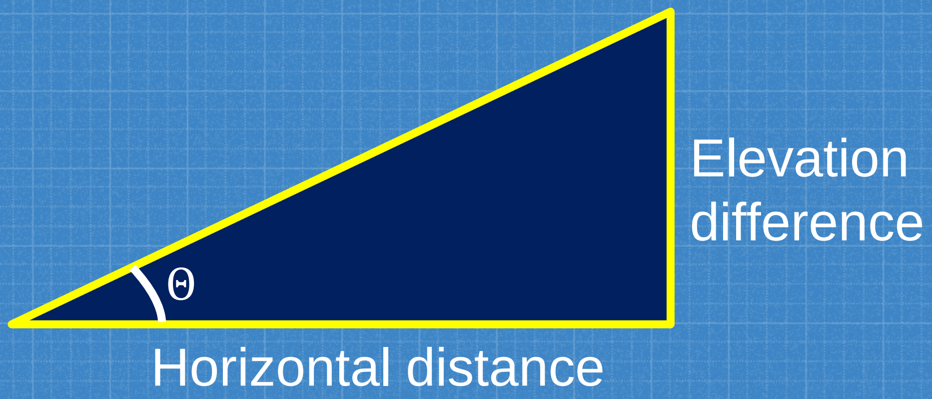
MY

Pro. Lee

1.Slope(percentage) 4.Total curvature 7.Aspect
2.Terrain roughness 5.Total slope high 8.Lithology
3.Slope roughness 6.Arias Intensity

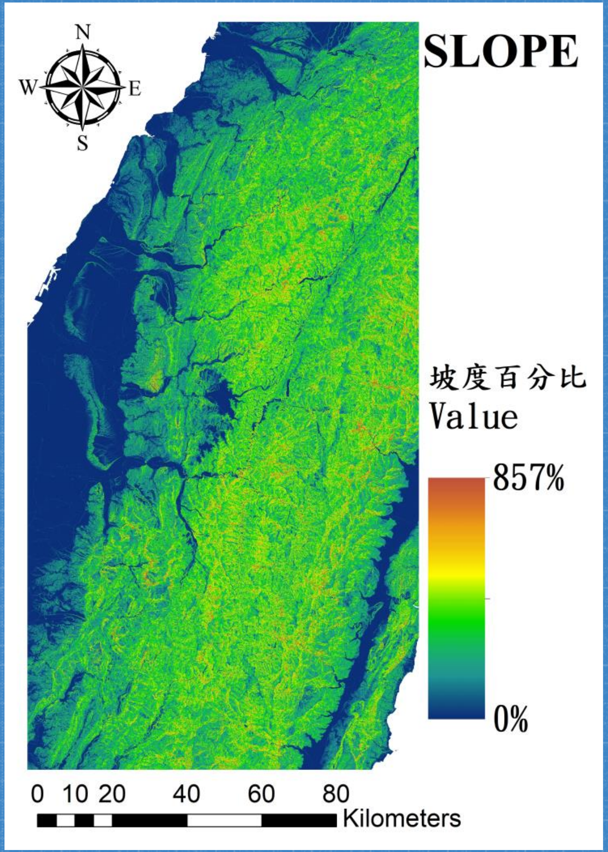
Slope(percentage)

Slope (3D Analyst) (Tool)



$$\tan\theta = \frac{Elevation}{Horizontal} \text{ ， } \theta: slope(degree)$$

$$\tan\theta = \frac{Elevation}{Horizontal} \times 100\% \text{ ， } \theta: slope(percentage)$$



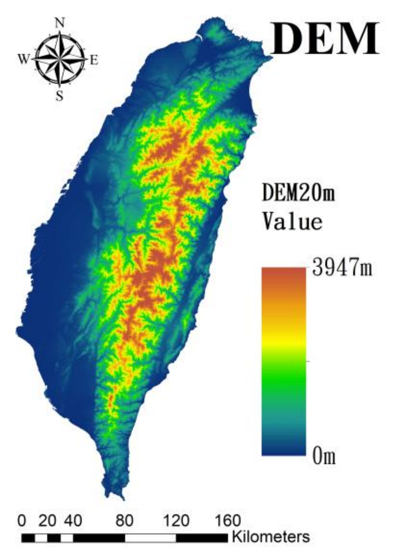
Terrain roughness

1.Slope(percentage) 4.Total curvature 7.Aspect
2.Terrain roughness 5.Total slope high 8.Lithology
3.Slope roughness 6.Arias Intensity

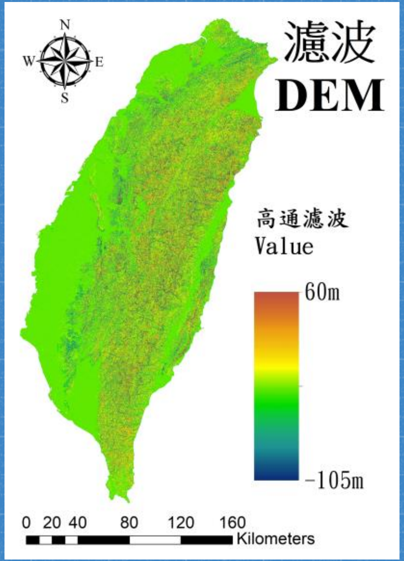
1. 將DTM資料輸入Imagine，利用高通濾波將原本屬於低頻的大區域地勢起伏去除，留下屬於高頻的地表粗糙度
2. 再利用濾波過後的數值地形使用既有gmd檔案計算地形粗糙度
3. 並將山稜線Buffer20公尺的區域，由外而內（山稜線）逐步削減為0，則完成地形粗糙度的計算。

Used a 13 × 13matrix to calculate the standard deviation of terrain heights and used this standard deviation as the terrain roughness.

(Chyi-Tyi Lee,2014)

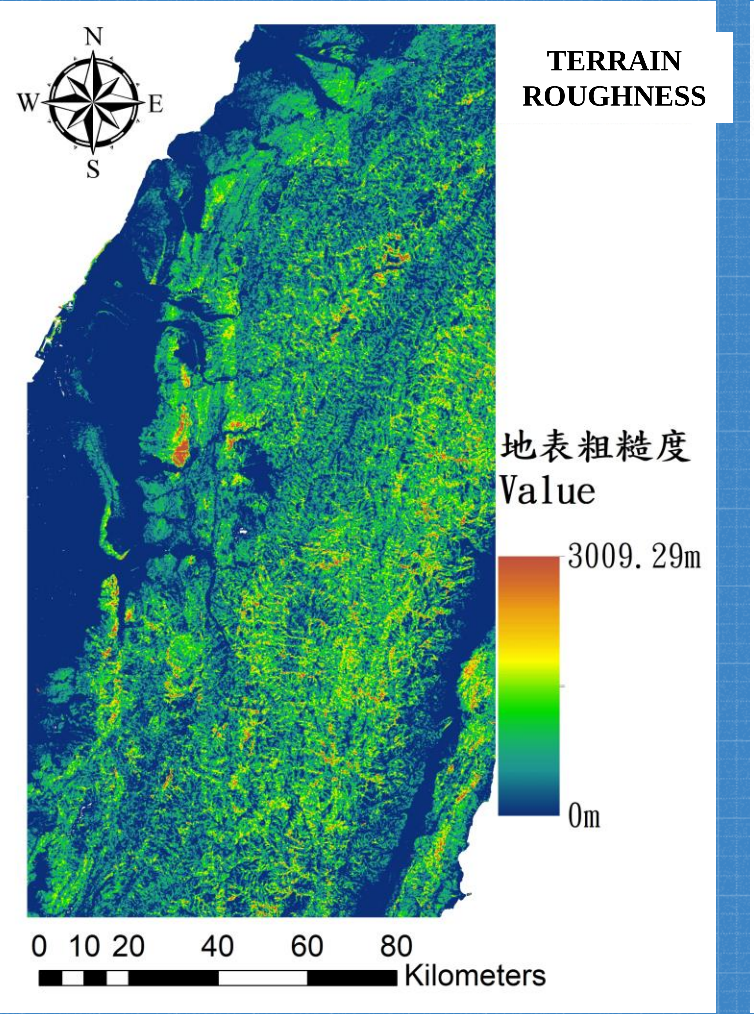


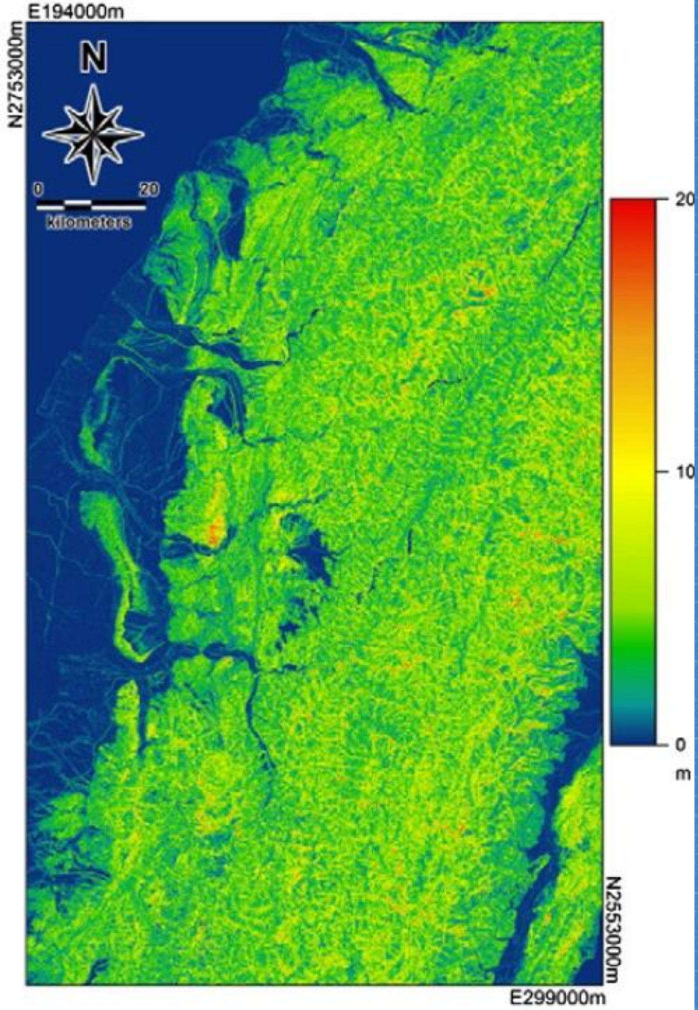
High pass
filter



Filter (Spatial Analyst) (Tool)

1	2	3
4	STD	5
6	7	8


N
W
E
S
TERRAIN
ROUGHNESS
地表粗糙度
Value
3009.29m
0m
0 10 20 40 60 80
Kilometers


E194000m
N2753000m
N
0 20
Kilometers
20
10
0
m
N2553000m
E299000m

MY

Pro. Lee

1.Slope(percentage) 4.Total curvature 7.Aspect
2.Terrain roughness 5.Total slope high 8.Lithology
3.Slope roughness 6.Arias Intensity

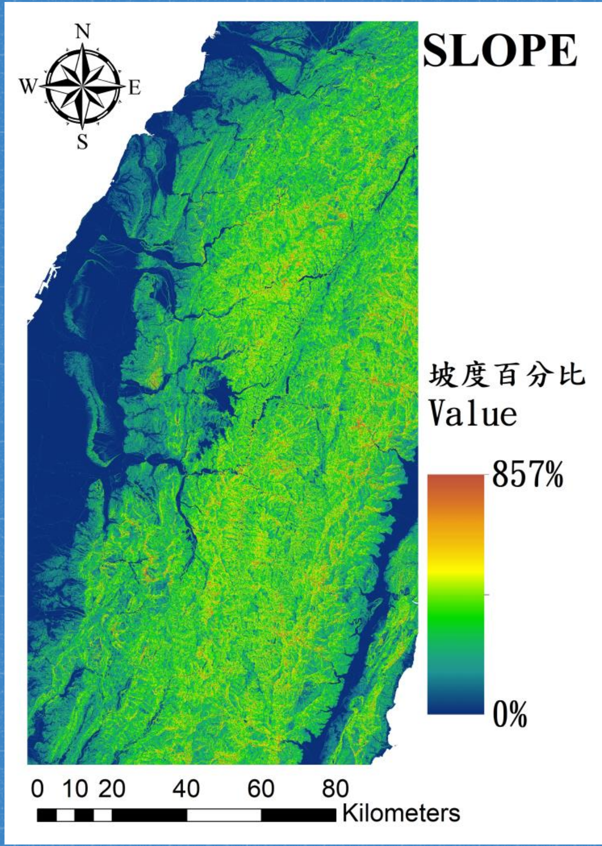
Slope roughness

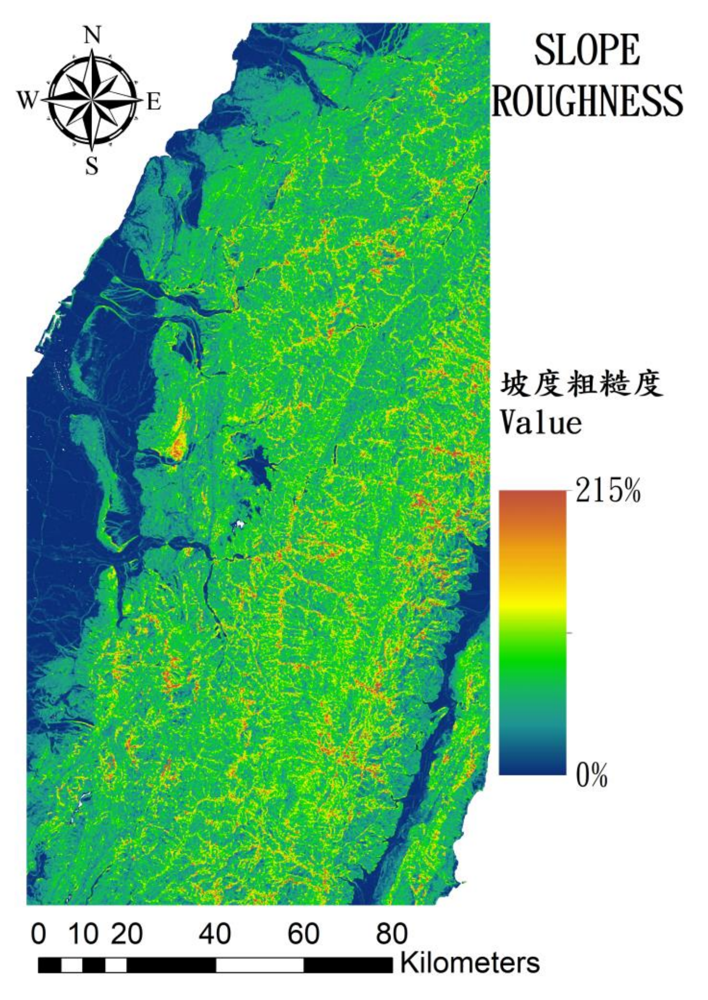
【簡易說明】

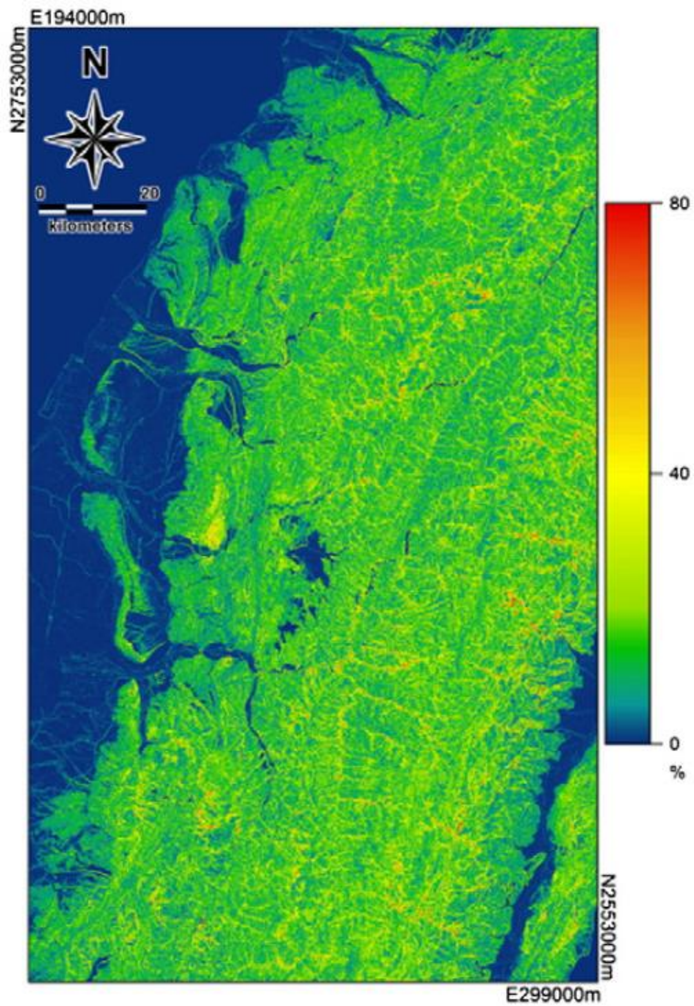
利用計算好的坡度資料，使用既有gmd檔案計算坡度粗糙度。

Used a 13 × 13matrix to calculate the standard deviation of slope and used this standard deviation as the terrain roughness. (Chyi-Tyi Lee,2014)

1	2	3
4	STD	5
6	7	8



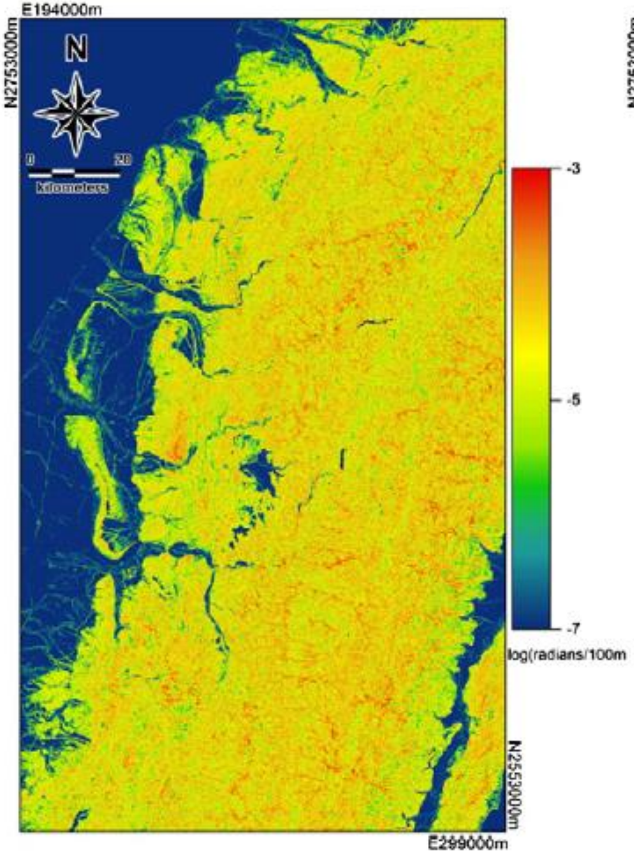

N
W
E
S
SLOPE
ROUGHNESS
坡度粗糙度
Value
215%
0%
0 10 20 40 60 80
Kilometers


E194000m
N2753000m
N
0 20
kilometers
80
40
0
%
N2553000m
E299000m

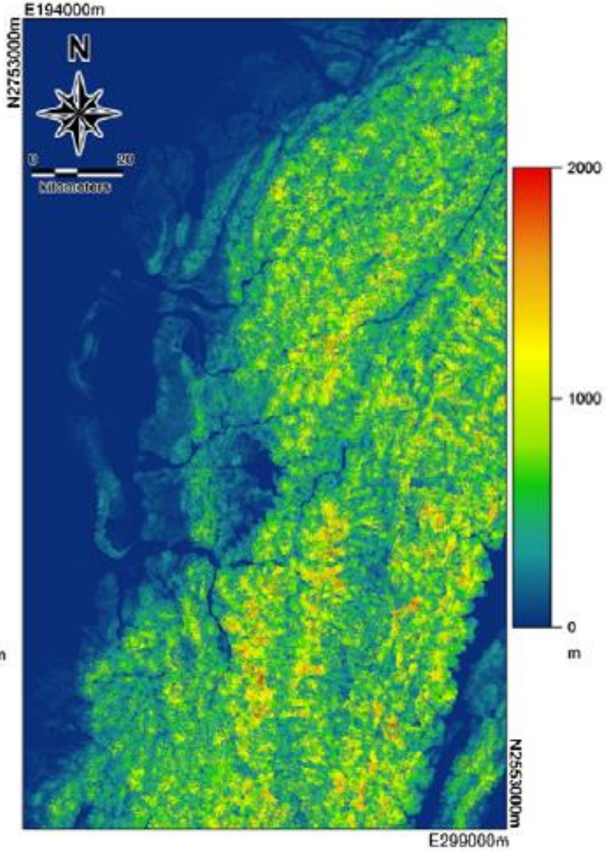
MY

Pro. Lee

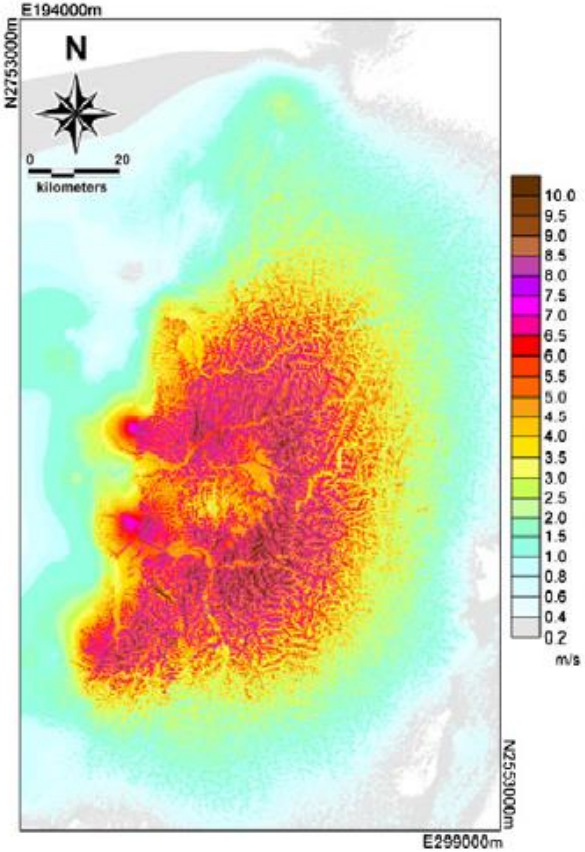
Total curvature



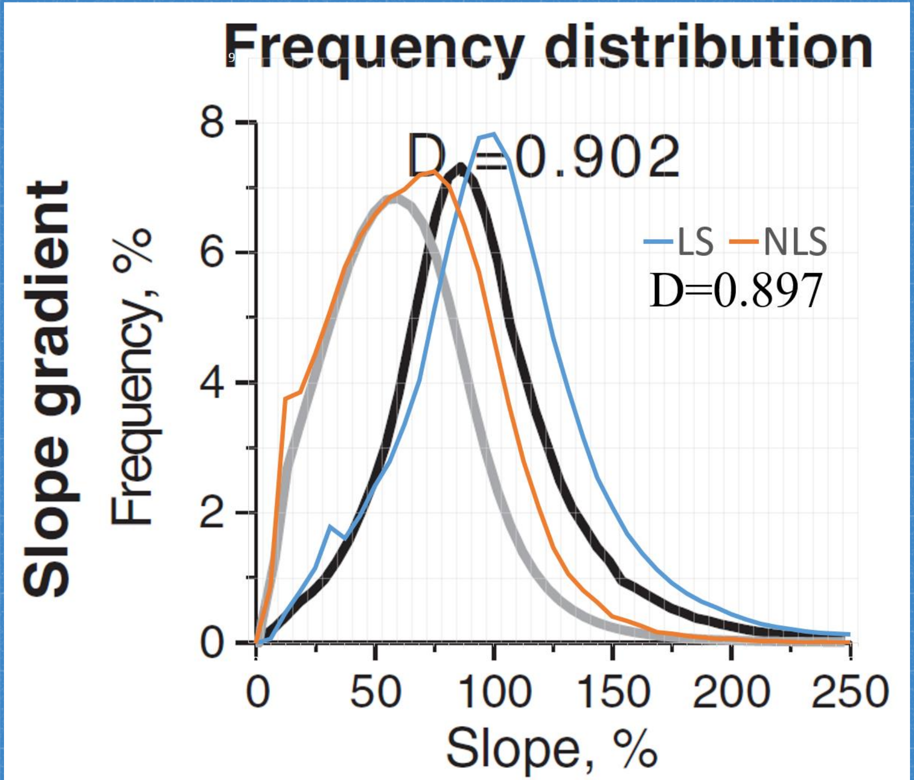
Total slope height



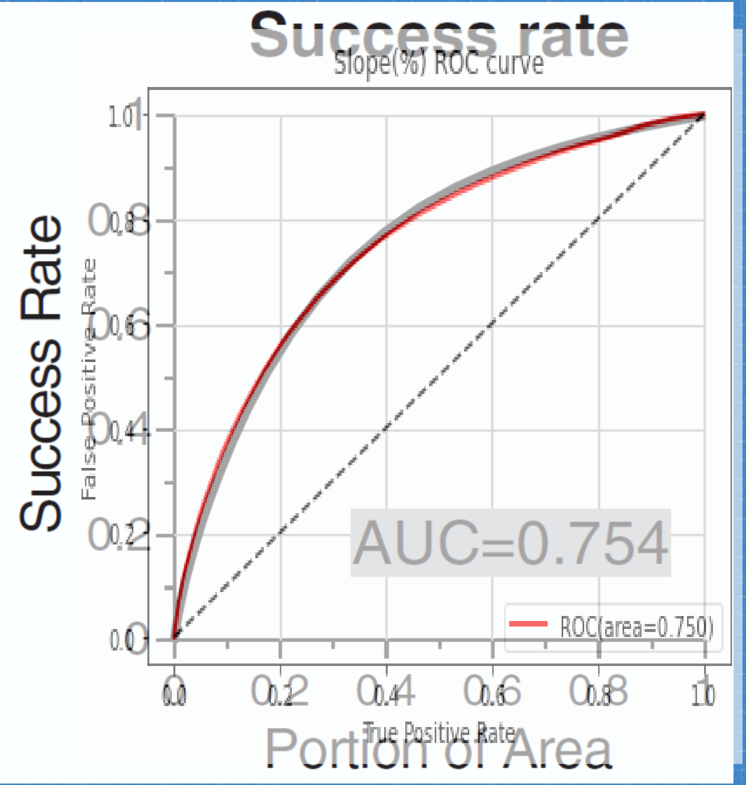
Arias Intensity



Slope

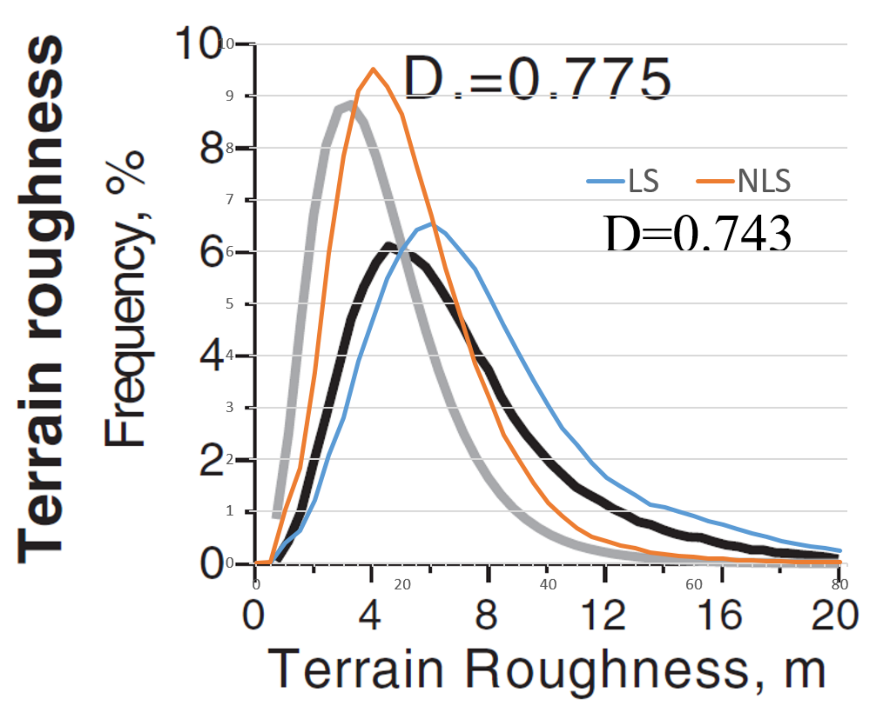


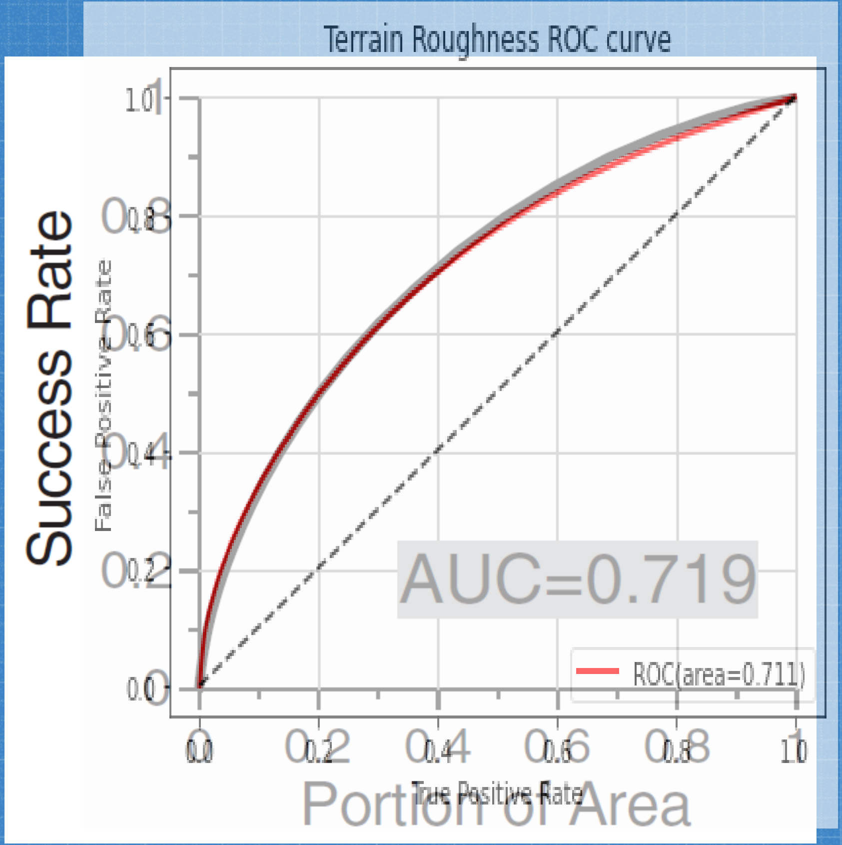
Frequency distribution
Slope gradient
Frequency, %
D =0.902
LS
NLS
D=0.897
8
6
4
2
0
0
50
100
150
200
250
Slope, %



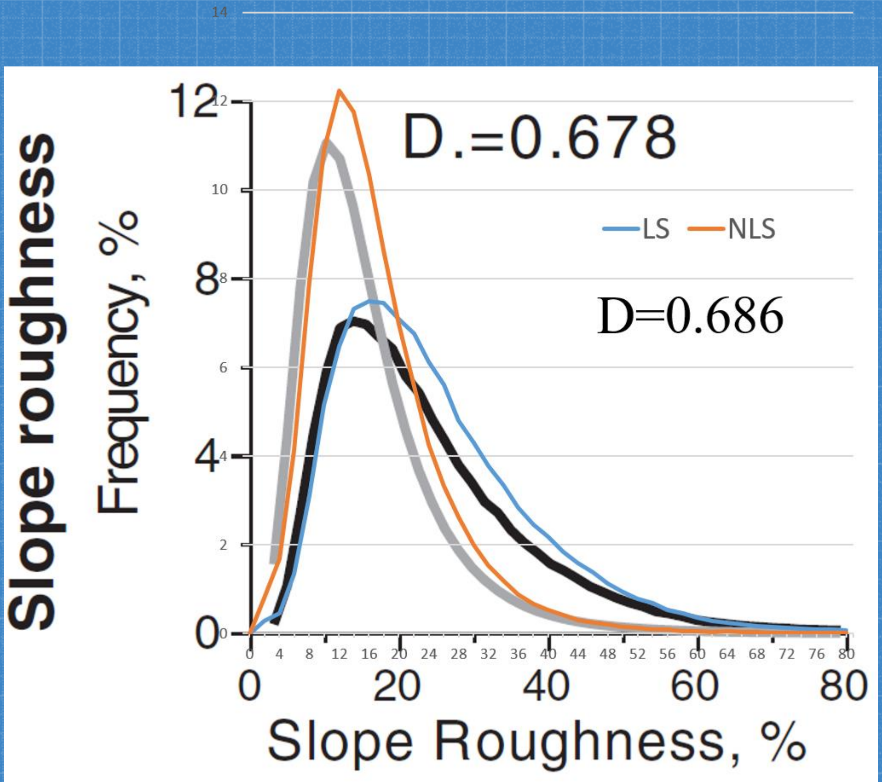
Success rate
Slope(%) ROC curve
Success Rate
False Positive Rate
AUC=0.754
ROC(area=0.750)
True Positive Rate
Portion of Area
0.0
0.2
0.4
0.6
0.8
1.0

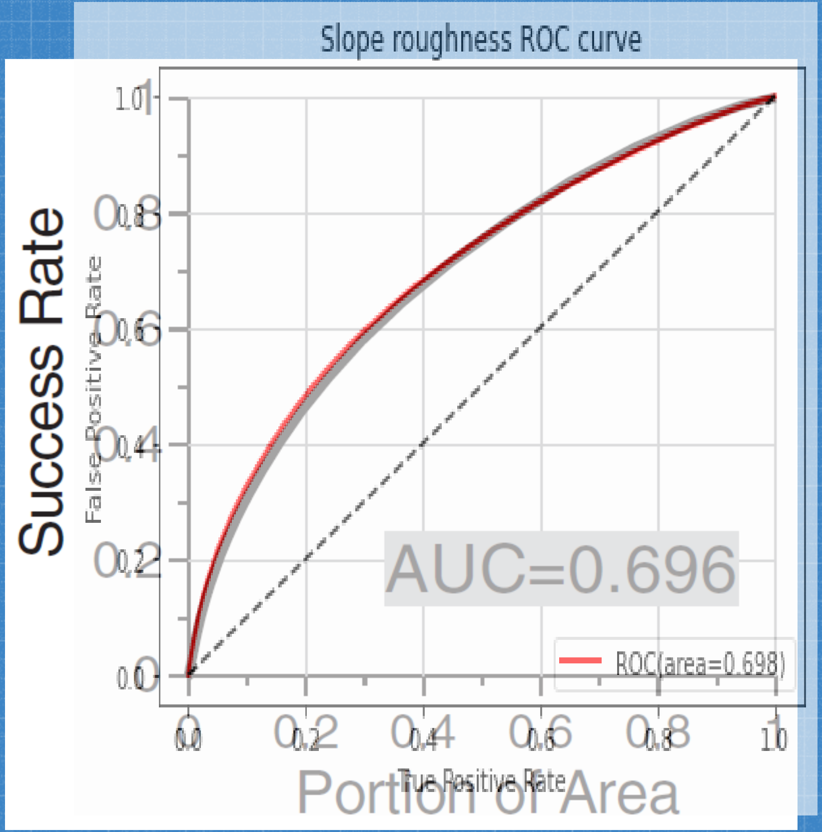
Terrain roughness



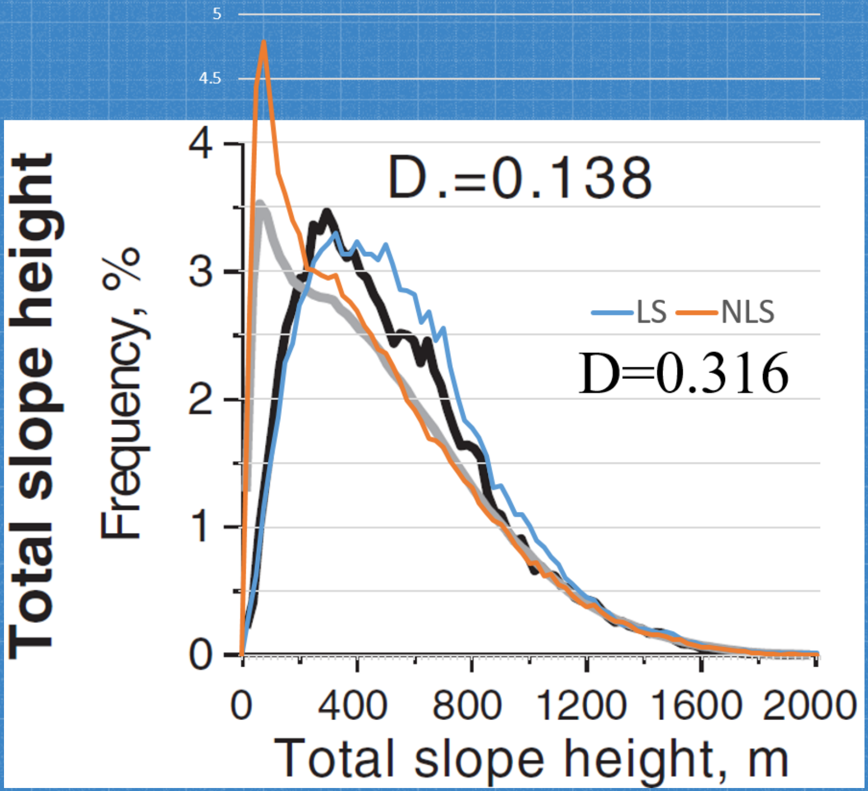


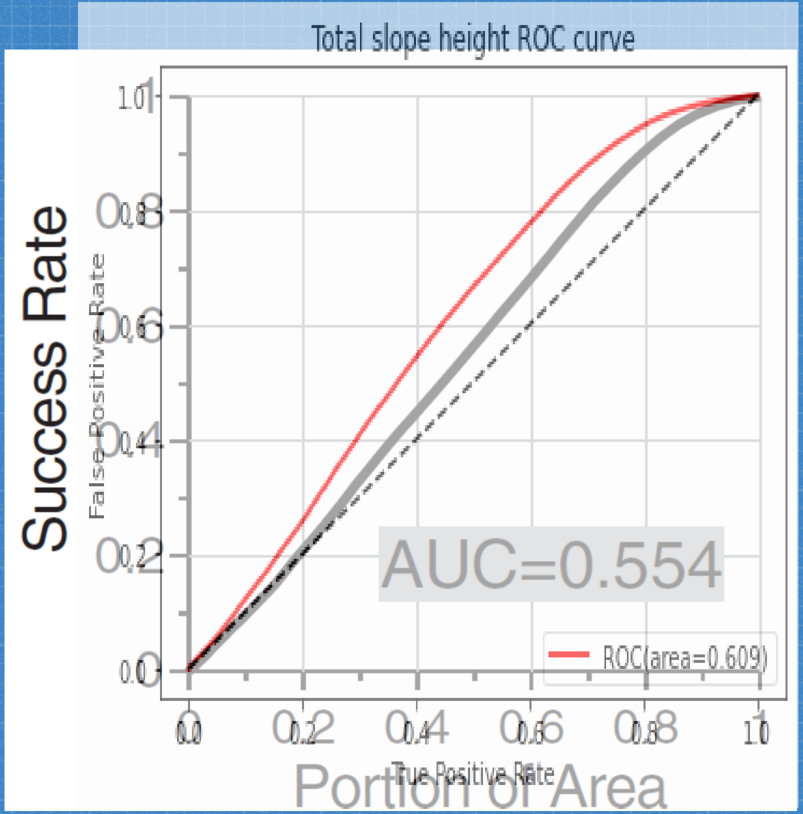
Slope roughness



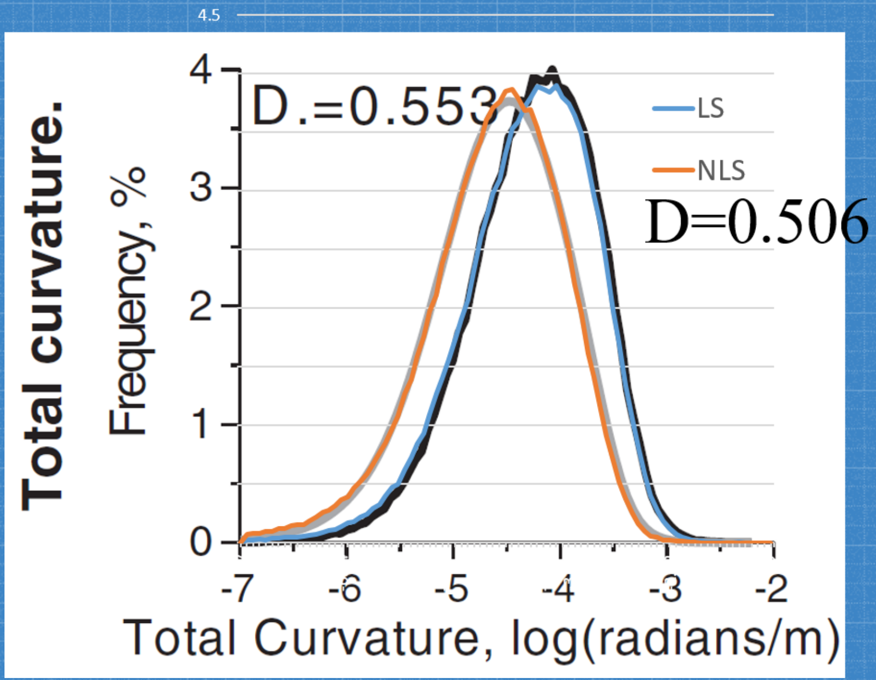


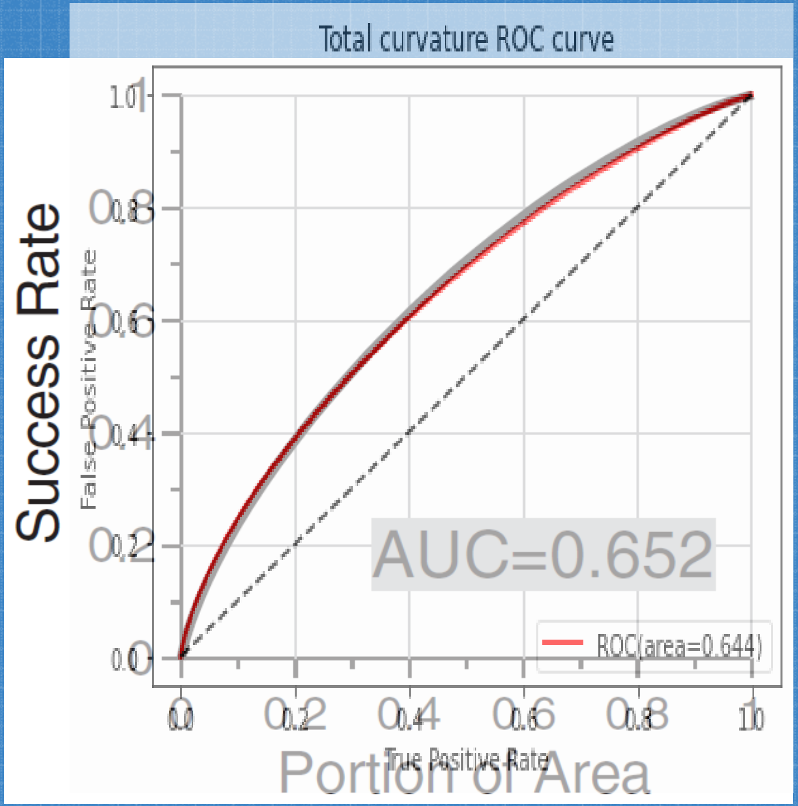
Total slope height



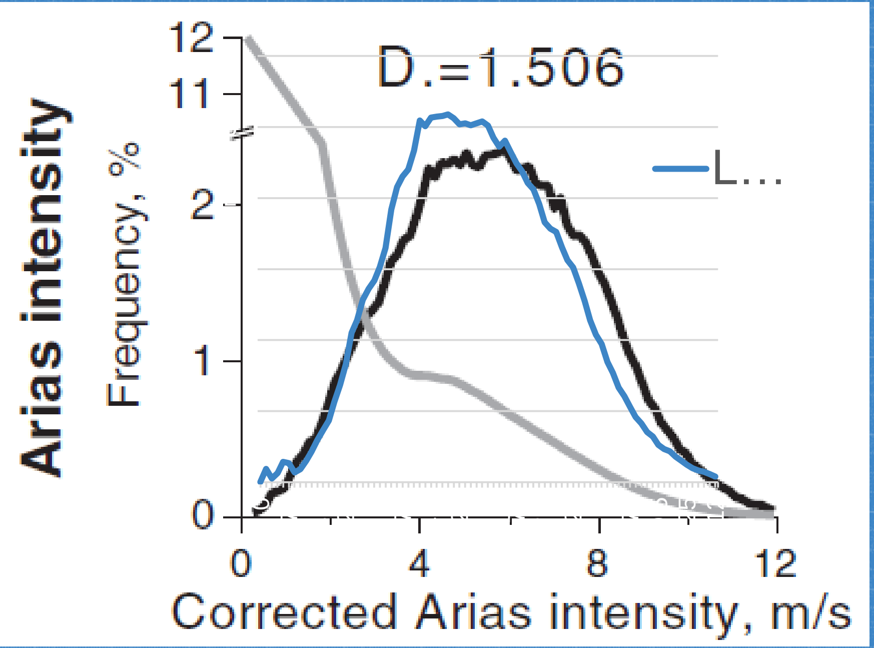


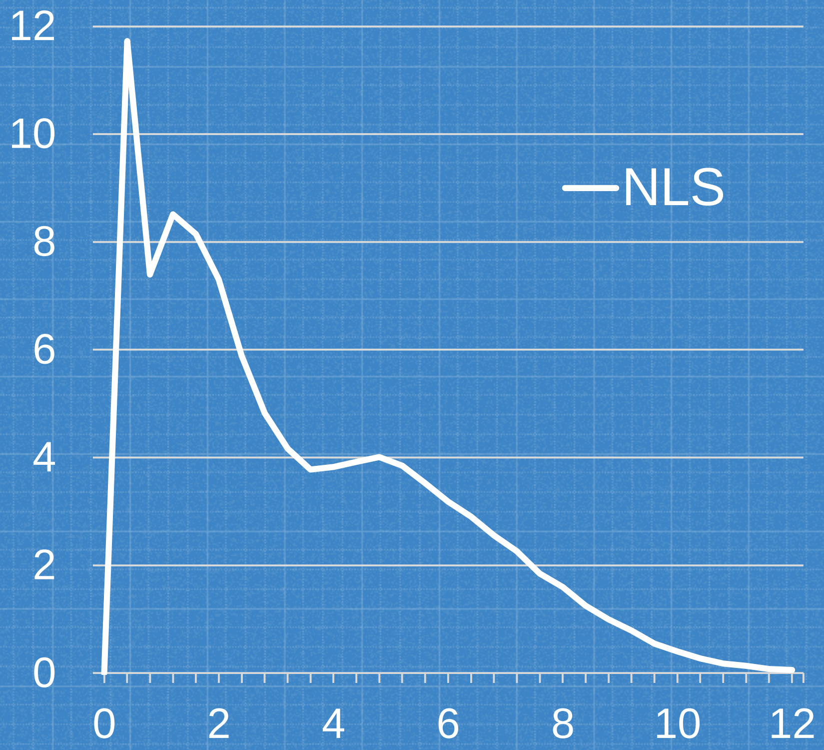
Total curvature





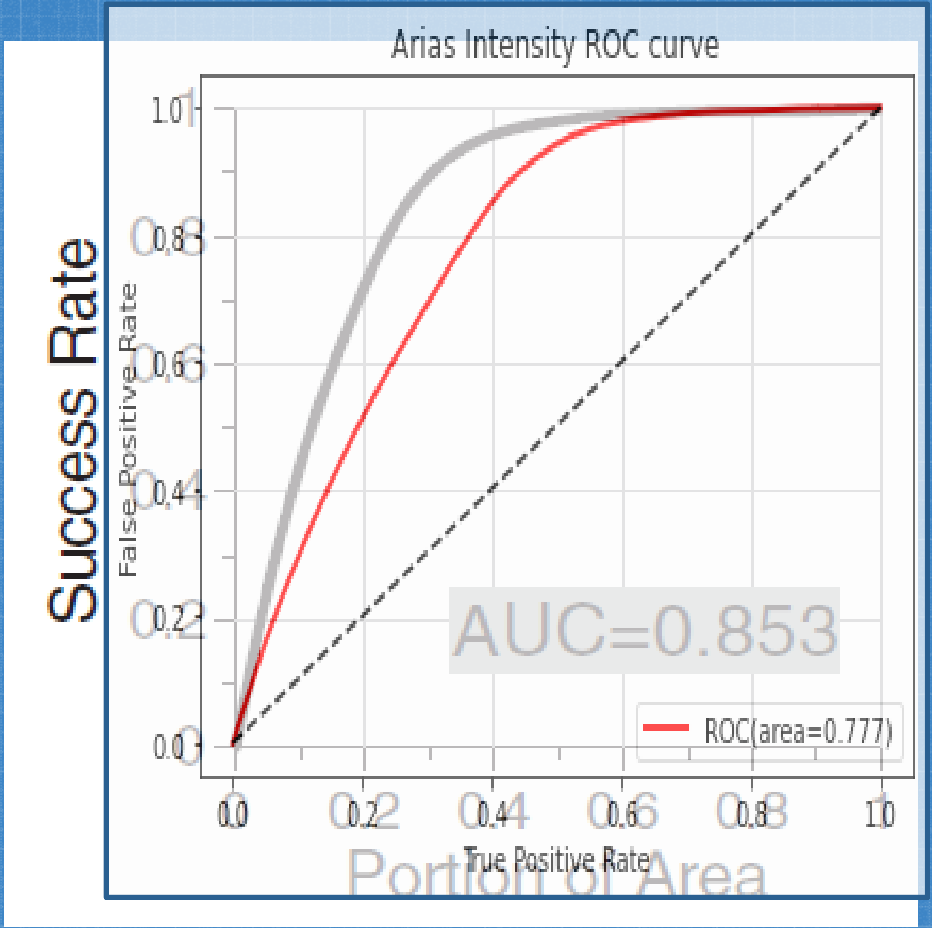
Arias Intensity





D=1.080

Arias Intensity



Appendix

AUC (Area under curve)

AUC=1→perfect model (ideal situation)

AUC>0.5 → classification effect of the classifier is better than random guessing

AUC=0.5 → classification effect of the classifier is the same as that of random guessing

AUC<0.5 → classification effect of the classifier is worse than random guessing