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

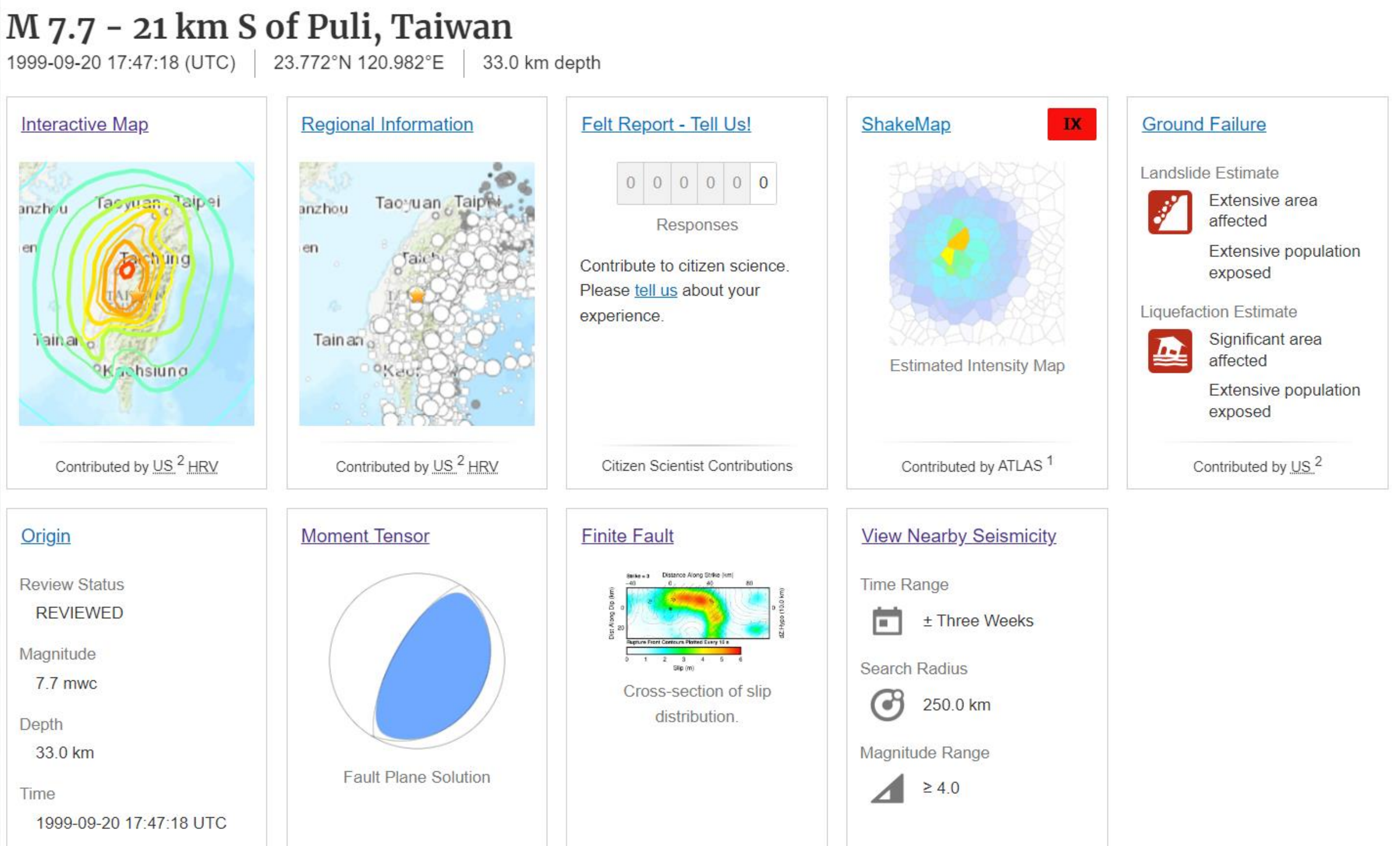
Outline

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

Introduction

Introduction

1999-09-21 01:47 (UTC+8) CHI-CHI earthquake



Cause :

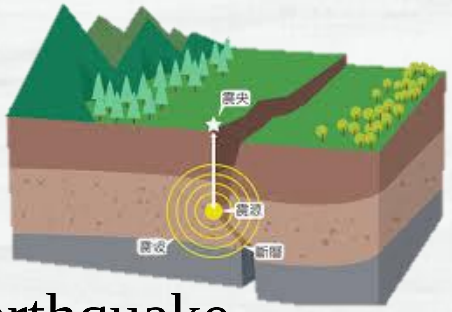
Nearly 10,000 landslides

Nearly 2500 fatality

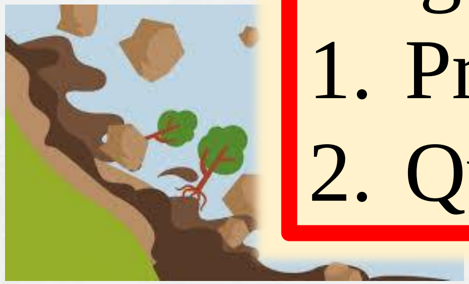
Report from USGS

jiu fen er shan

Introduction



Earthquake



landslides

caused

fatality

Rapid identification of landslides is very important:
1. Assessment of earthquake impacts
2. Hazard mitigation

In recent year:

Two kinds of methods

ased

een key

nd landslide

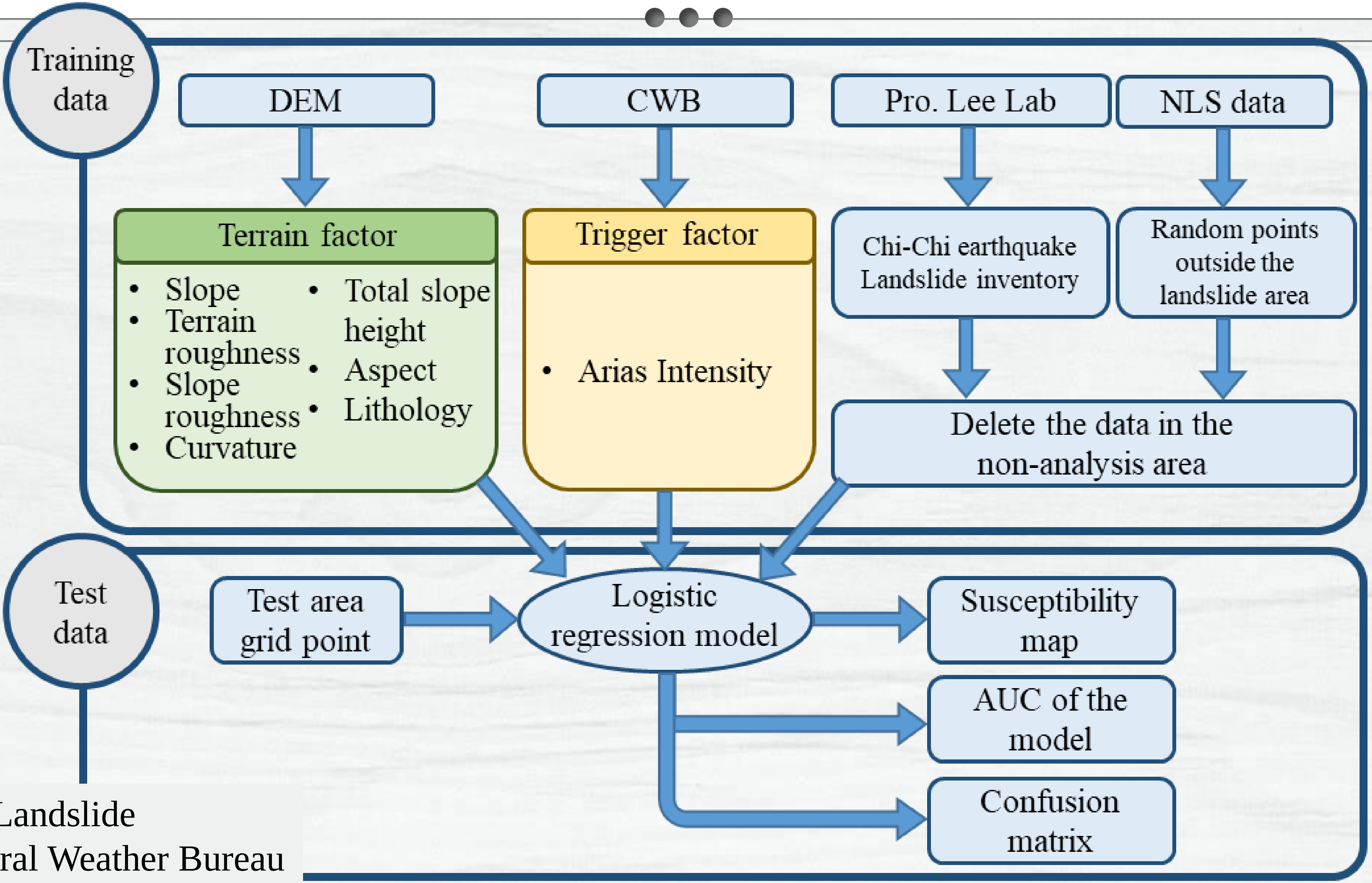
Target :

1. Precisely : most landslides can be predicted
2. Quickly : near real-time (maybe in few minute)

References	(Jibson et al., 2000) (Gallen et al., 2017)	(Nowicki Jessee et al., 2018) (Robinson et al., 2018)
Apply to	Small Area	Large Area

Methodology

Flow Chart



NLS : Non-Landslide
CWB : Central Weather Bureau

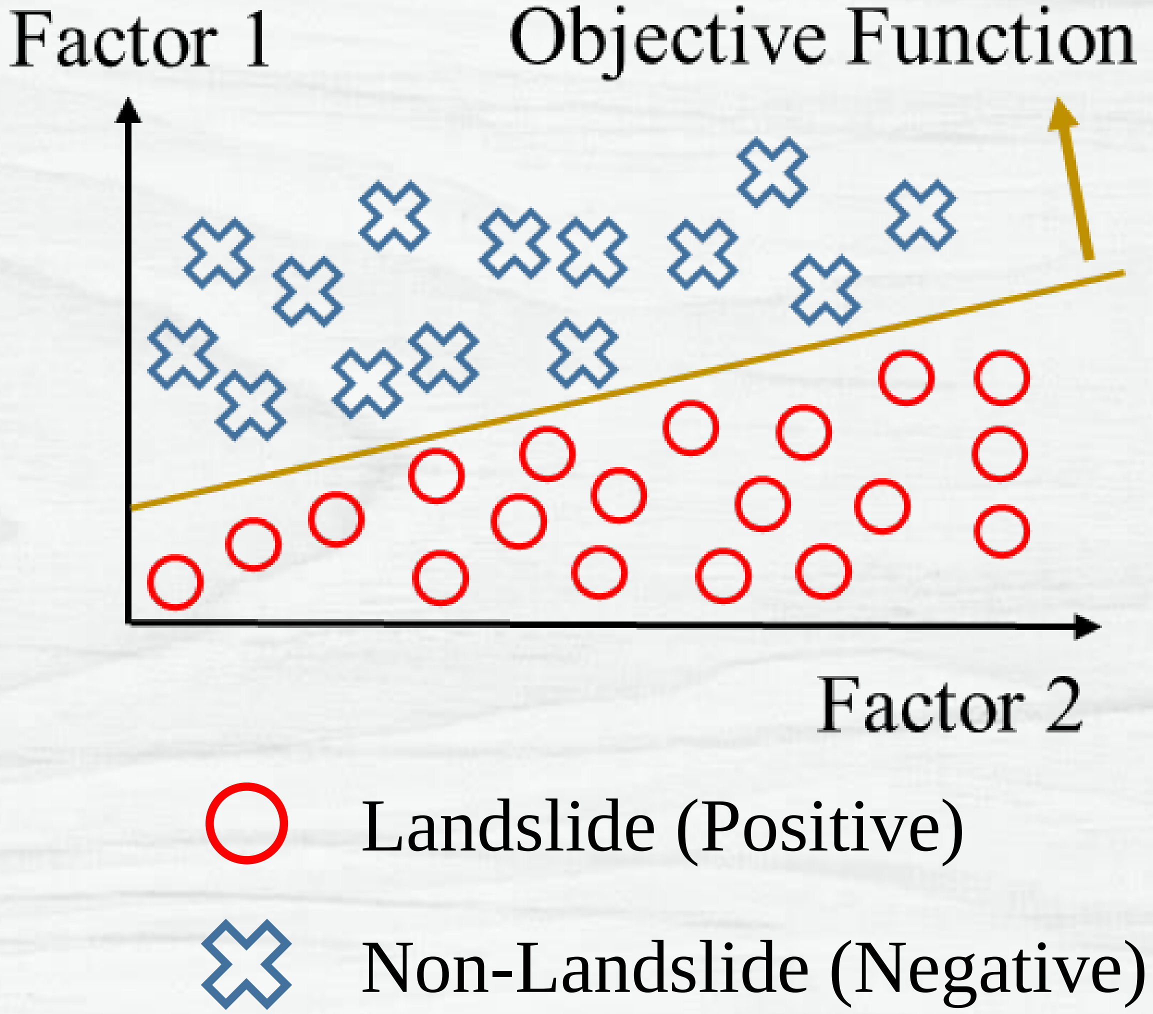
Logistic Regression

Target :

Find objective function to separate two groups of data

$$f(x) = a_0 + \Sigma_{n=1}^{k} a_n x_n$$

a_0 : $Intersection$
a_n : $coeficient\ of\ the\ factor$
x_n : $value\ of\ the\ factor$
k : $number\ of\ factors$



Apply to:

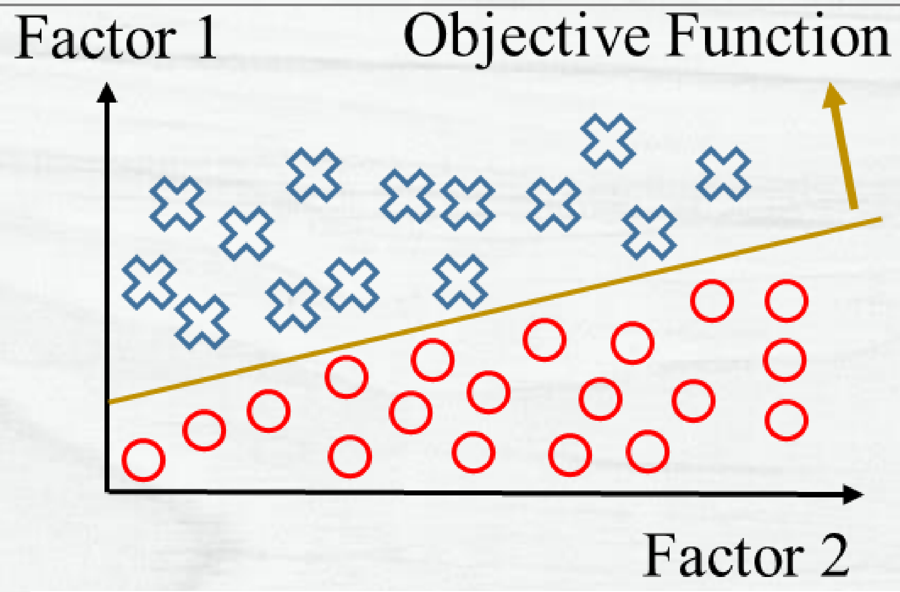
1. Binary case
2. Discontinuity variable

Methodology

Logistic Regression Susceptibility

Objective Function : $f(x) = a_0 + \Sigma_{n=1}^{k} a_n x_n$

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



If the data point located at :

- ✓ Right side of the function

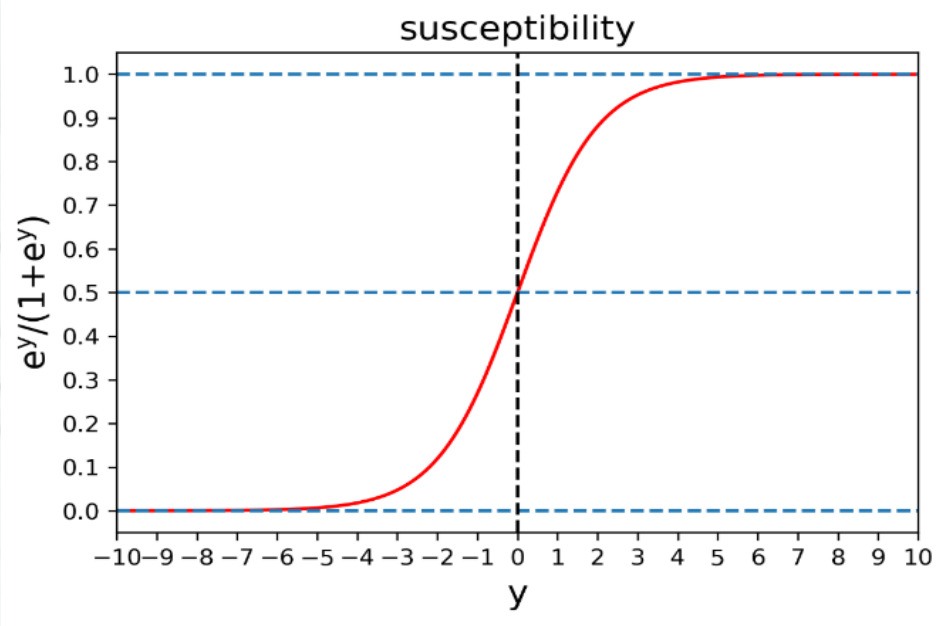
  $y > 0$，$Susceptibility > 0.5$

  Landslide may occur

- ✓ Left side of the function

  $y < 0$，$Susceptibility < 0.5$

  Landslide may not occur



Methodology

Logistic Regression Models

Use all of data to build model → Model 1

Use intensity to separate different groups and build model

- Use corrected Arias Intensity to separate into a few groups → Model 2
- Use uncorrected Arias Intensity to separate into a few groups → Model 3

Factors in my model :

1. Slope
2. Terrain roughness
3. Slope roughness
4. Curvature
5. Total Slope Height
6. Arias Intensity
7. Aspect
8. Lithology

Arias Intensity (I_a) correction equation :

$$I_{a(corrected)} = f \times I_a$$

$$f = \left(\frac{h}{93.8} + 0.287\right)^{0.5} + 0.464$$

f : *Magnification*
h : *Slope Height*

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

Confusion Matrix

Confusion matrix		Observed	
		LS	NLS
predicted	LS	TP	FP
	NLS	FN	TN

LS : Landslide
NLS : Non-Landslide
TP : True Positive
FP : False Positive
TN : True Negative
FN : False Negative

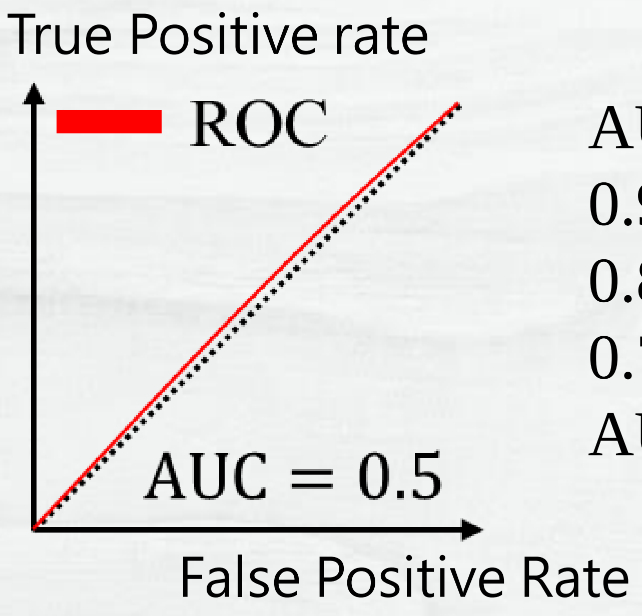
TPR : True Positive Rate
TNR : True Negative Rate
FPR : False Positive Rate
FNR : False Negative Rate

$$\mathrm{TPR} = \frac{\mathrm{TP}}{TP + FN}，\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}，FPR = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}，FNR = \frac{\mathrm{FN}}{TP + FN}$$

$$\text{準確率(Precision)} = \frac{\mathrm{TP}}{TP + FP}，\text{召回率(Recall)} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

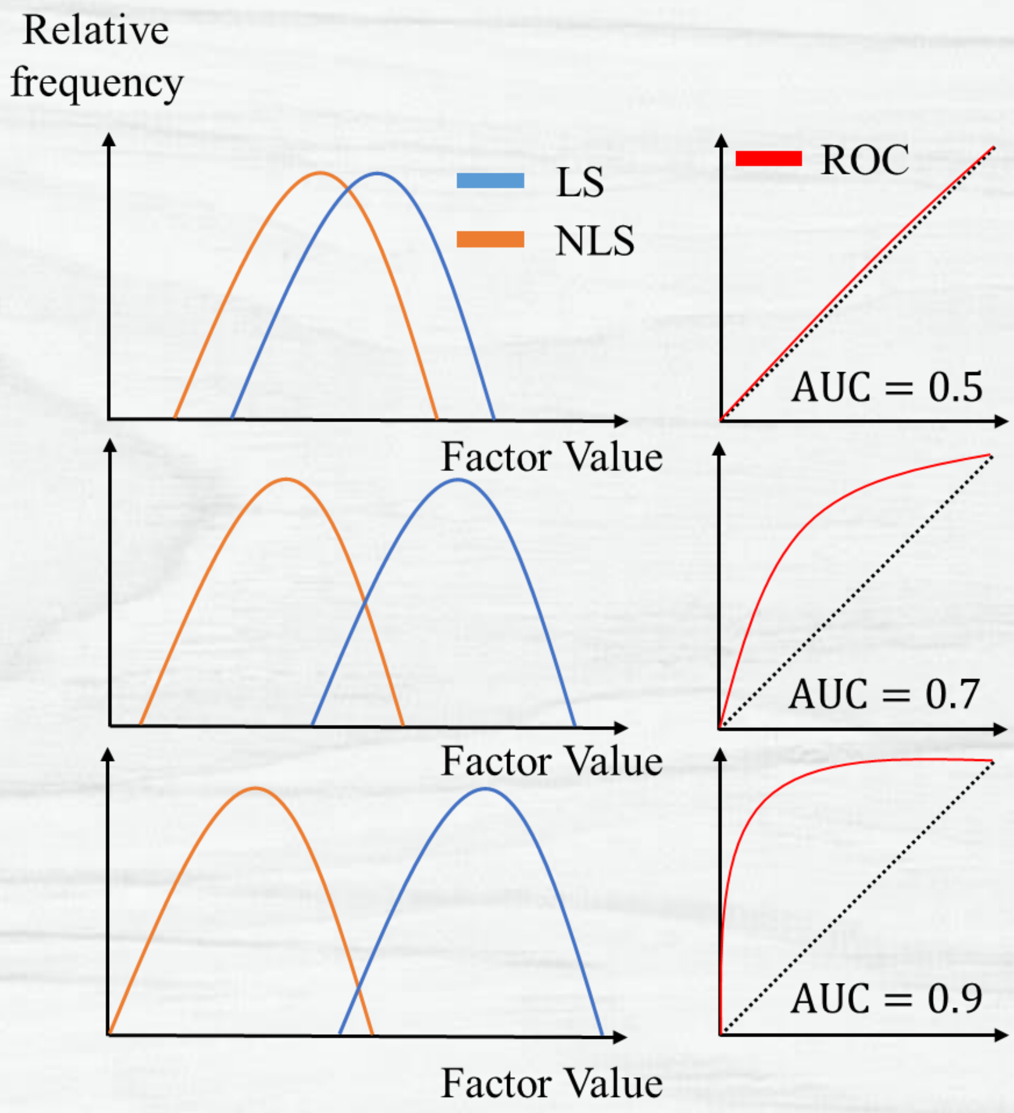
★ If the Recall is higher, it means more landslides will be predicted

Area Under 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

Use different threshold of susceptibility to calculate TPR and FPR, and plot the data on the figure.



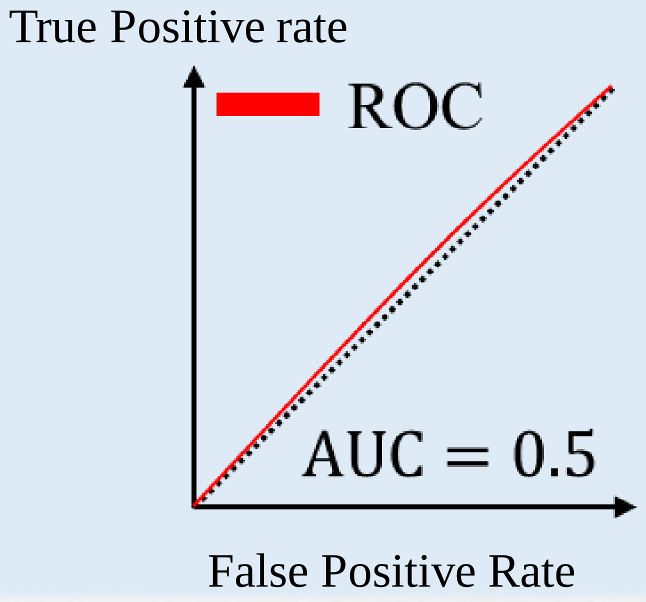
Area Under Curve

Example data :

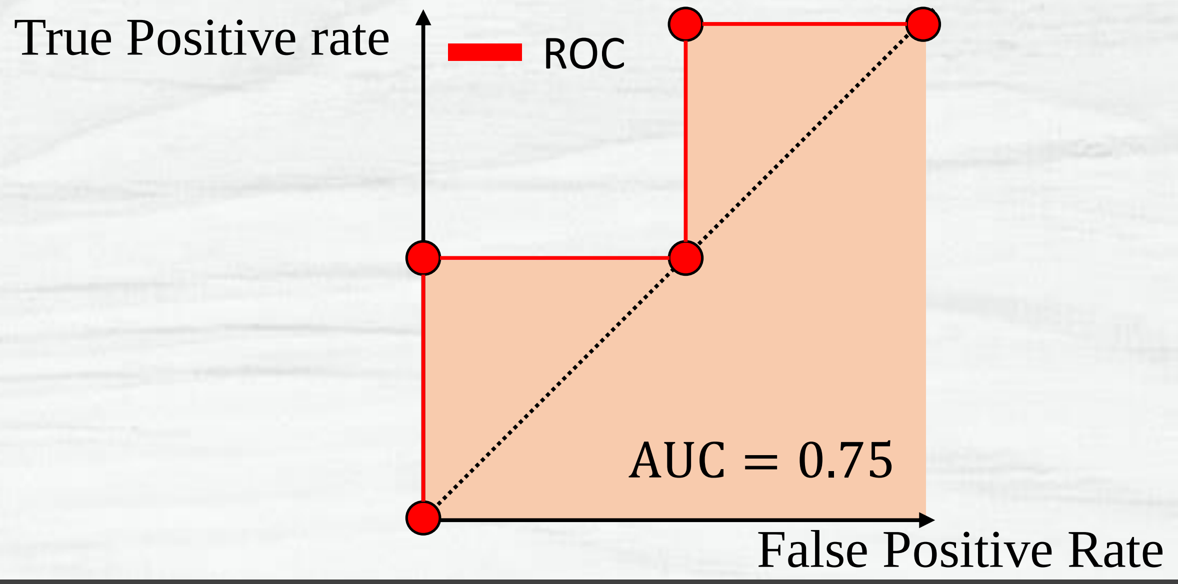
Observed	Susceptibility
1	0.8
1	0.4
0	0.6
0	0.2

Confusion matrix		Observed	
		LS	NLS
predicted	LS	TP	FP
	NLS	FN	TN

$$\mathrm{TPR} = \frac{\mathrm{TP}}{TP + FN}, \quad FPR = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}$$



Susceptibility Threshold	TPR	FPR
0	1	1
0.3	1	0.5
0.5	0.5	0.5
0.7	0.5	0
1	0	0



Cross Validation

Dataset :

90% training data	10% validation data

Build Logistic Regression Model

Compare model AUC between training data and validation data

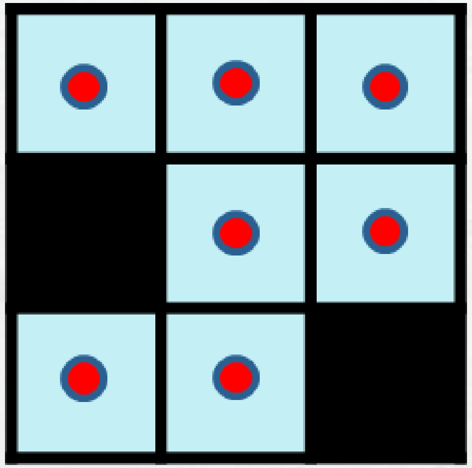
If validation data AUC different from training data

→ This model may over fitting or under fitting

If validation data AUC similar to training data

→ This model may be suitable

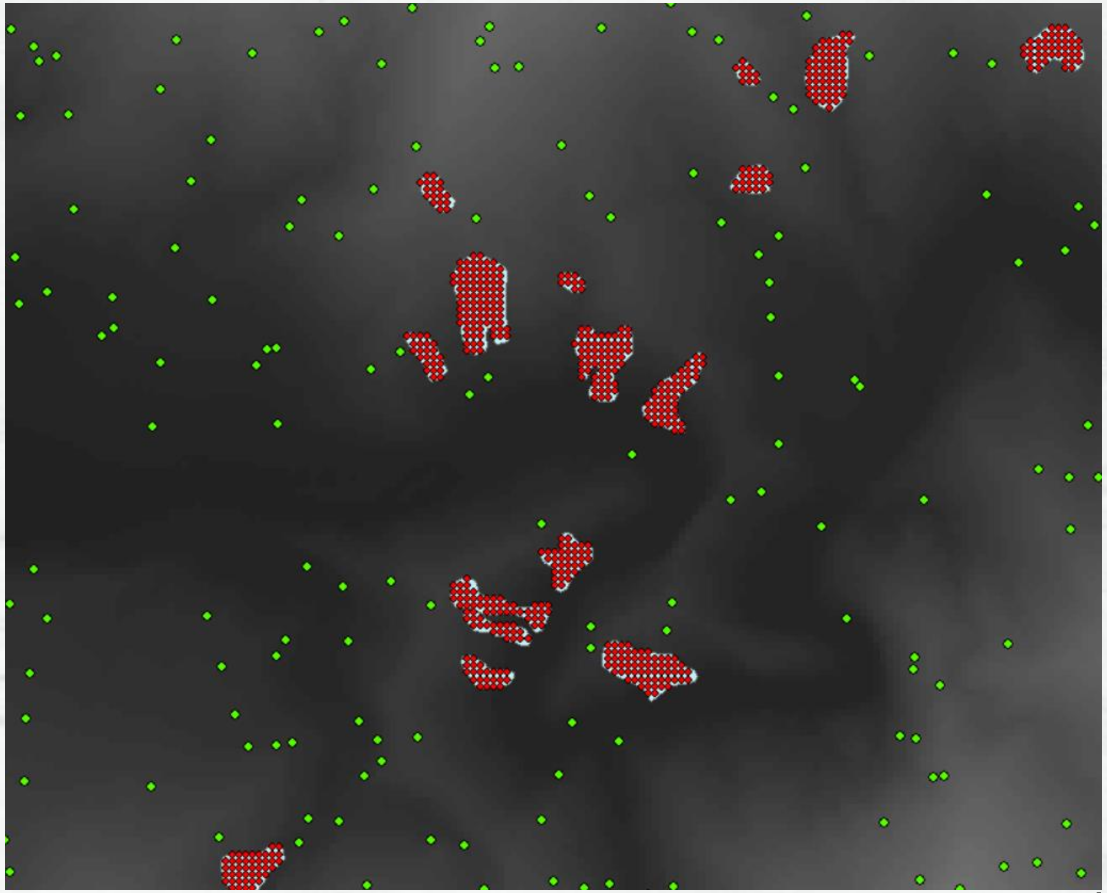
Training data selection



Landslide polygon

Landslide data

Non-Landslide data



Landslide data : total landslide grids

Non-Landslide data : randomly selected

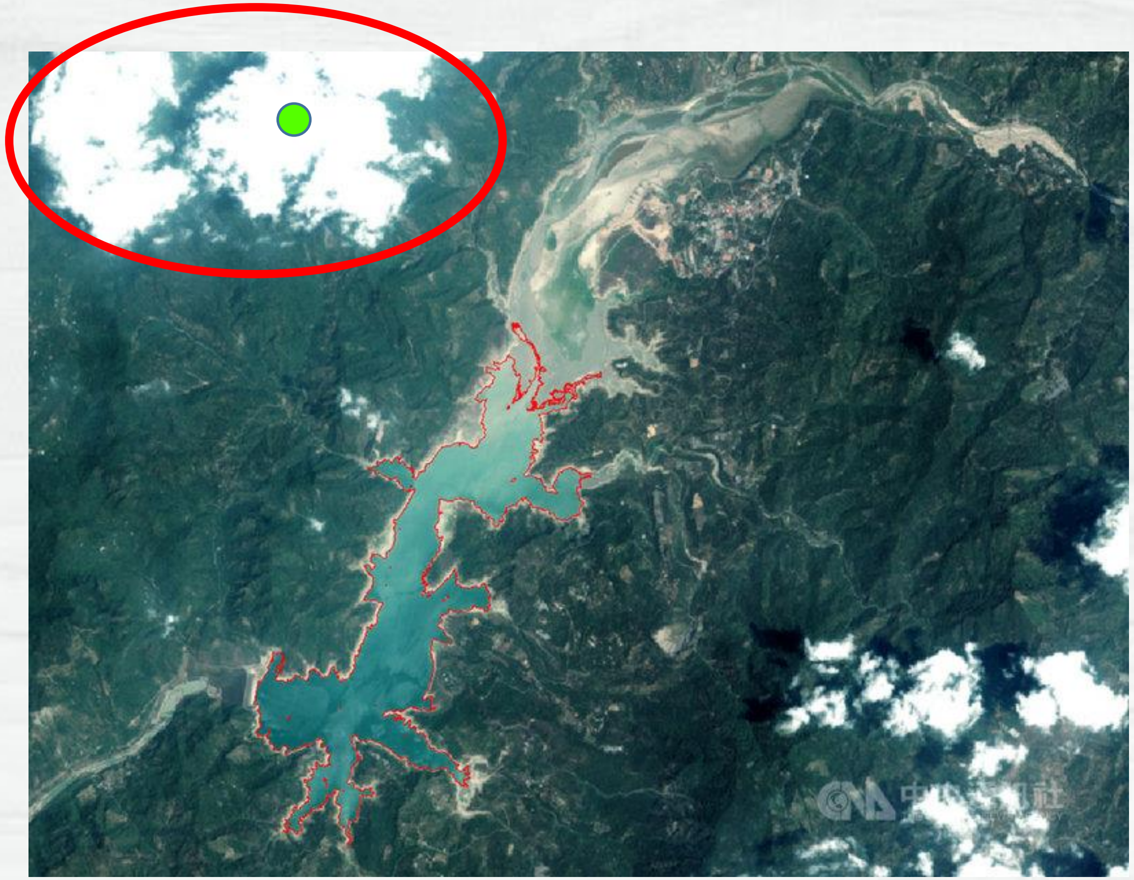
data of similar size with landslide data

Methodology

Training data selection

Important thing for Non-Landslide data :

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



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

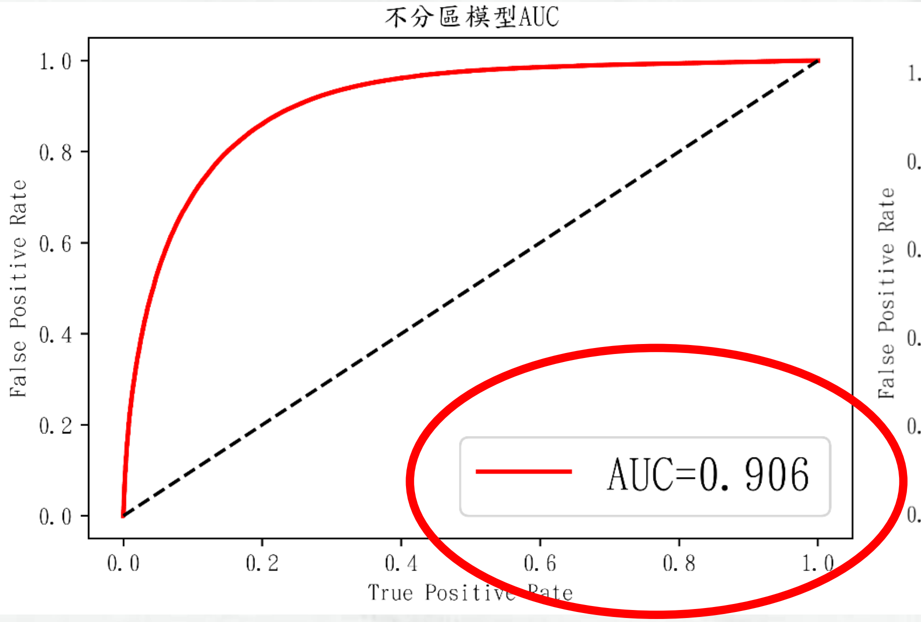
3. History landslide area :
The geological conditions in this area are unstable.

Result

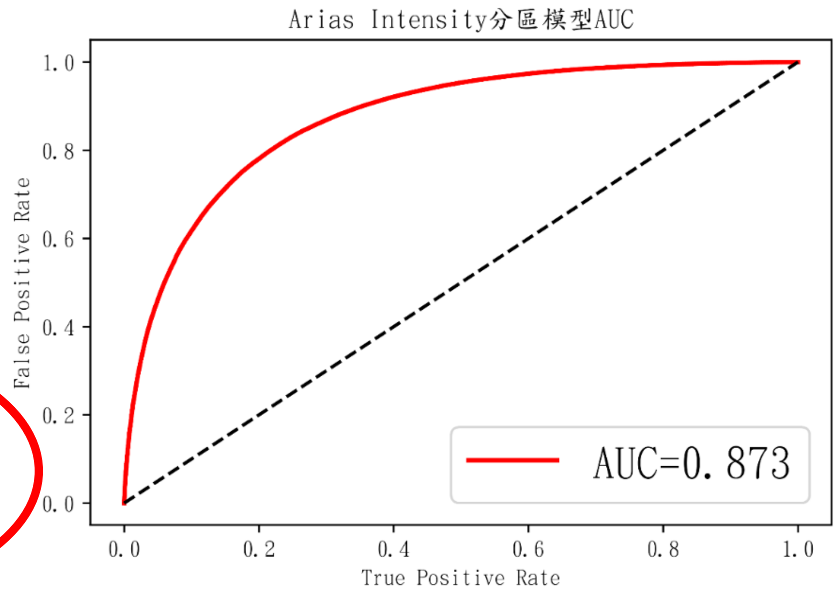
AUC of Each Model

Result

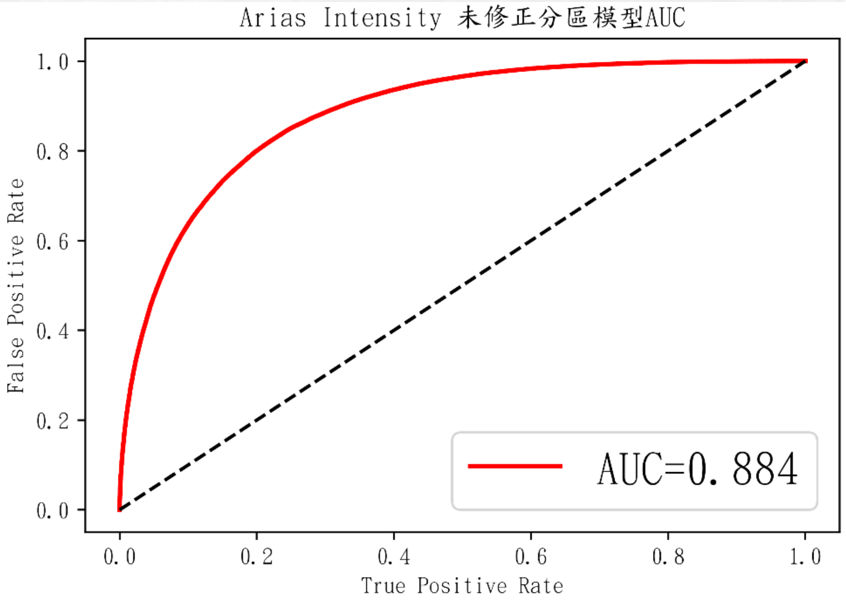
Model 1



Model 2



Model 3



Model 1 : all data
Model 2 : group by corrected Arias Intensity
Model 3 : group by uncorrected Arias Intensity

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

Result

Cross Validation

Model 1	Training data AUC = 0.906 Validation data AUC = 0.905	0.001
Model 2	Training data AUC = 0.873 Validation data AUC = 0.874	0.001
Model 3	Training data AUC = 0.884 Validation data AUC = 0.881	0.003

Model 1 : all data
Model 2 : group by corrected Arias Intensity
Model 3 : group by uncorrected Arias Intensity

The AUC between training data and validation data are similar

Three of model didn't have problem about over fitting or under fitting.

Result

Confusion Matrix

Model 1 : all data

Model 2 : group by corrected Arias Intensity

Model 3 : group by uncorrected Arias Intensity

Confusion matrix		Observed	
		LS	NLS
predicted	LS	TP	FP
	NLS	FN	TN

$$準確率(Precision) = \frac{TP}{TP + FP},$$

$$召回率(Recall) = \frac{TP}{TP + FN}$$

Confusion matrix (model 1)		observation	
		LS	NLS
prediction	LS	162083	32973
	NLS	27084	135388

TPR: 0.831
TNR:0.833
FPR: 0.169
FNR: 0.167

Precision: 0.831
Recall: 0.857

Confusion matrix (model 2)		observation	
		LS	NLS
prediction	LS	154846	40234
	NLS	40916	151776

TPR: 0.794
TNR: 0.788
FPR: 0.206
FNR: 0.212

Precision: 0.794
Recall: 0.791

Confusion matrix (model 3)		observation	
		LS	NLS
prediction	LS	158498	40135
	NLS	38565	156811

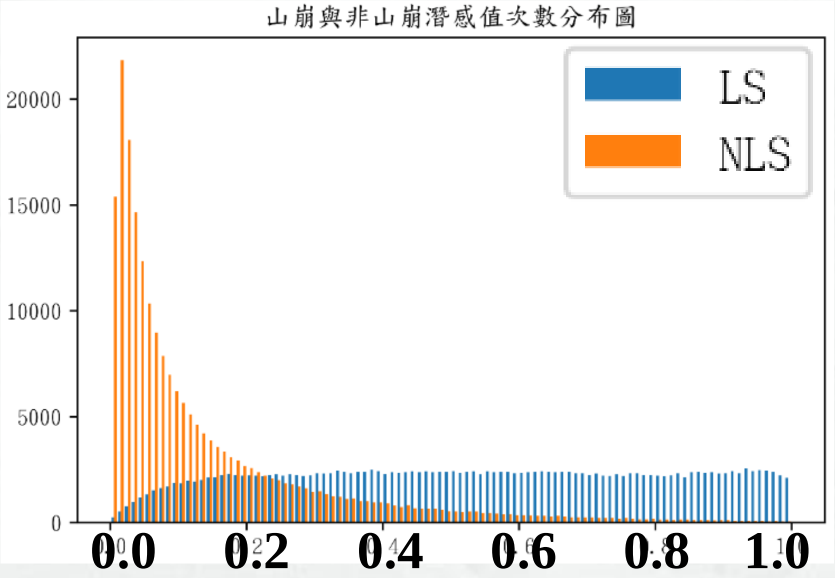
TPR: 0.798
TNR: 0.803
FPR: 0.202
FNR: 0.197

Precision: 0.798
Recall: 0.804

Result

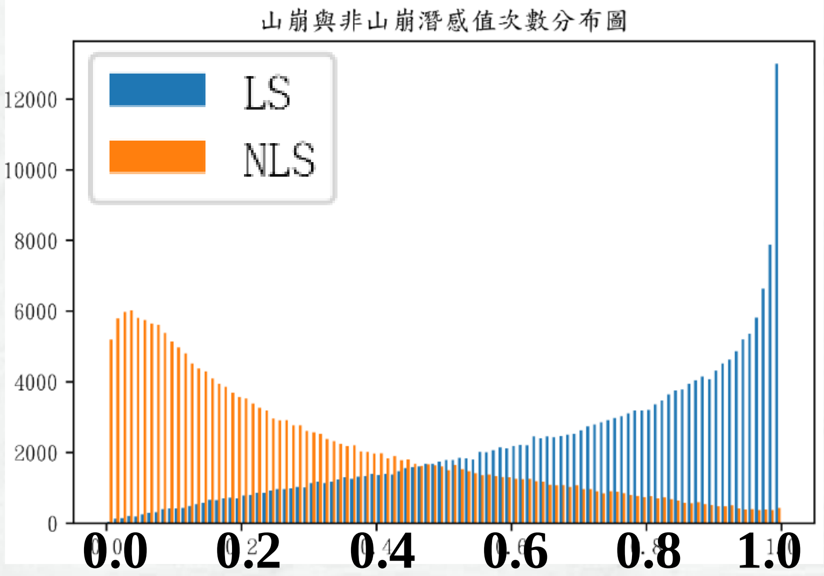
Susceptibility Distribution

Model 1



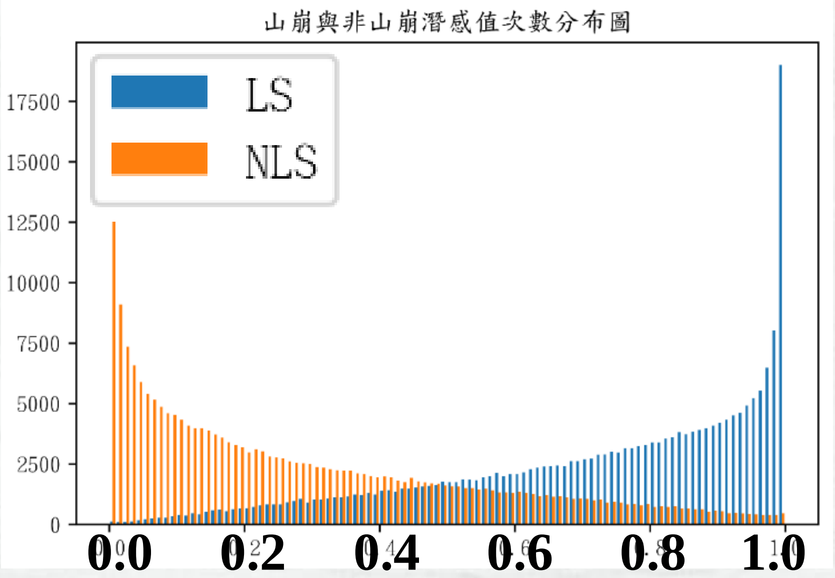
Susceptibility

Model 2



Susceptibility

Model 3



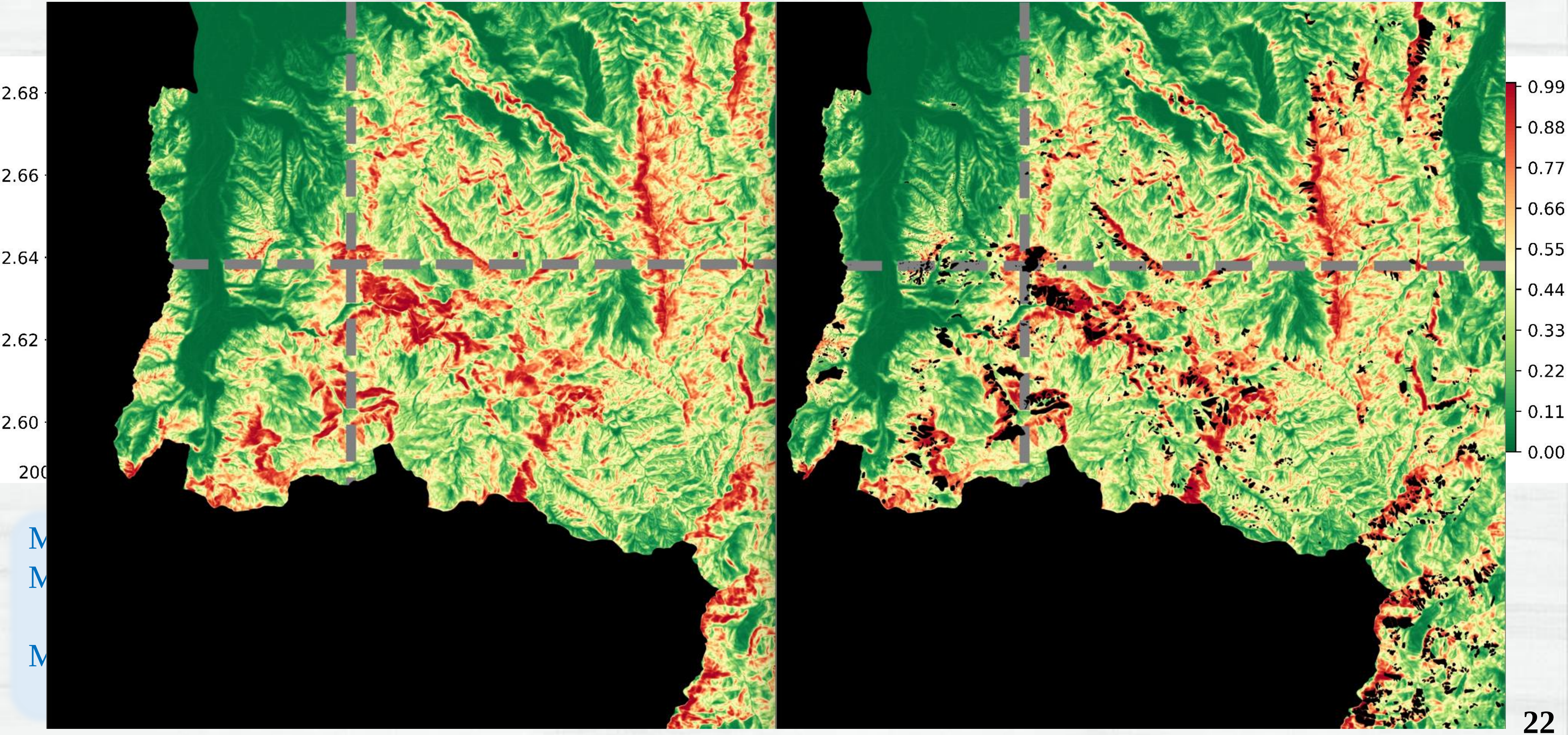
Susceptibility

Model 1 : all data
Model 2 : group by corrected Arias Intensity
Model 3 : group by uncorrected Arias Intensity

Model 2 and model 3 are better than model 1

Because it can separate landslide and non-landslide better

Susceptibility of Each Model : Nantou



M
M
M

Conclusion

Conclusion

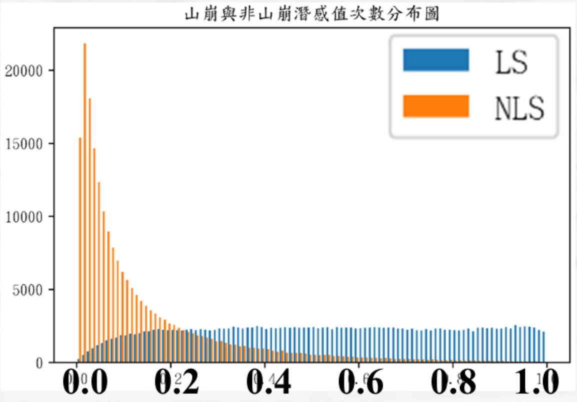
1. From the point of view of AUC, model 1 is the best model.
2. From the point of view of recall, more landslide can be predicted in model 1.
3. From the point of view of susceptibility distribution, model 2 or model 3 are better than model 1
4. Three of model didn't have problem about over fitting or under fitting.

AUC :
Model 1 = 0.906
Model 2 = 0.873
Model 3 = 0.884

Recall :
Model 1 = 0.857
Model 2 = 0.791
Model 3 = 0.804

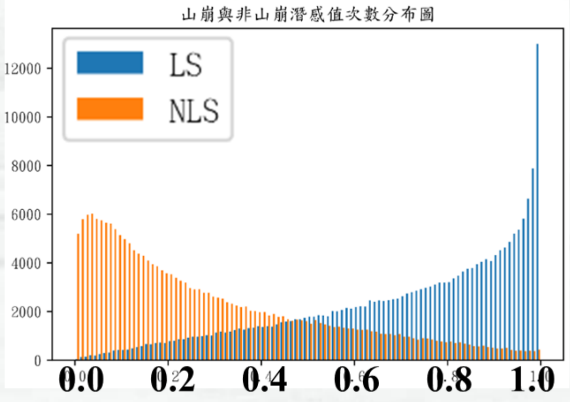
Model 1 : all data
Model 2 : group by corrected Arias Intensity
Model 3 : group by uncorrected Arias Intensity

Model 1



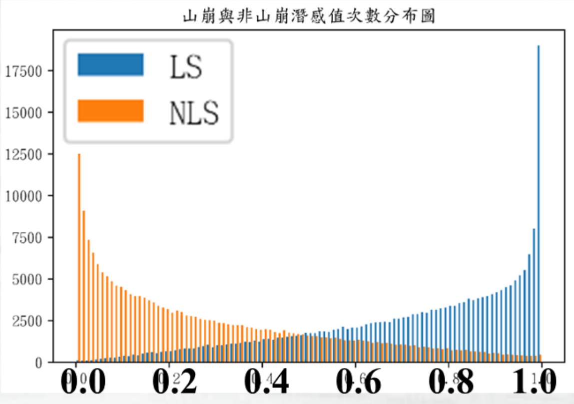
Susceptibility

Model 2



Susceptibility

Model 3



Susceptibility

Future work

Future work

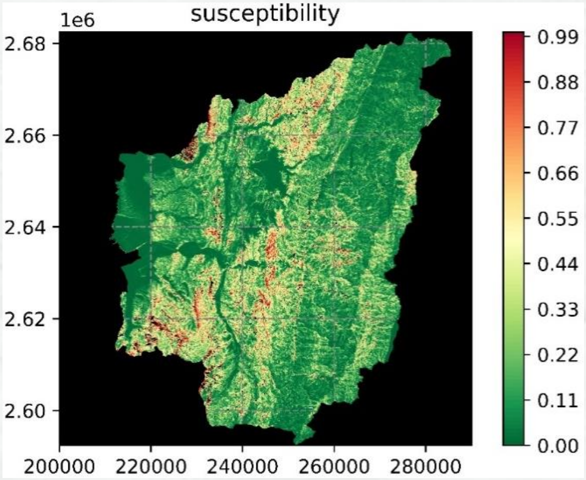
1. Modify the logistic regression model and choose the best one
2. Convert these steps into automatically program
3. Extend susceptibility map to the whole of Taiwan

earthquake signal

Get

automatically program

Export



Susceptibility Map of whole of Taiwan

Thank you

Introduction

ArcGIS

1. Collect topography information
2. Calculate the factor which the model need

Python

1. Automatically program
2. Build logistic regression model
3. Export susceptibility map

不分區

Confusion matrix (model 1)		observation LS	observation NLS
prediction	LS	169517	3998320
prediction	NLS	24820	16057343

TPR: 0.041
TNR: 0.998
FPR: 0.959
FNR: 0.002
Precision: 0.041
Recall: 0.872

Accuracy: 0.801

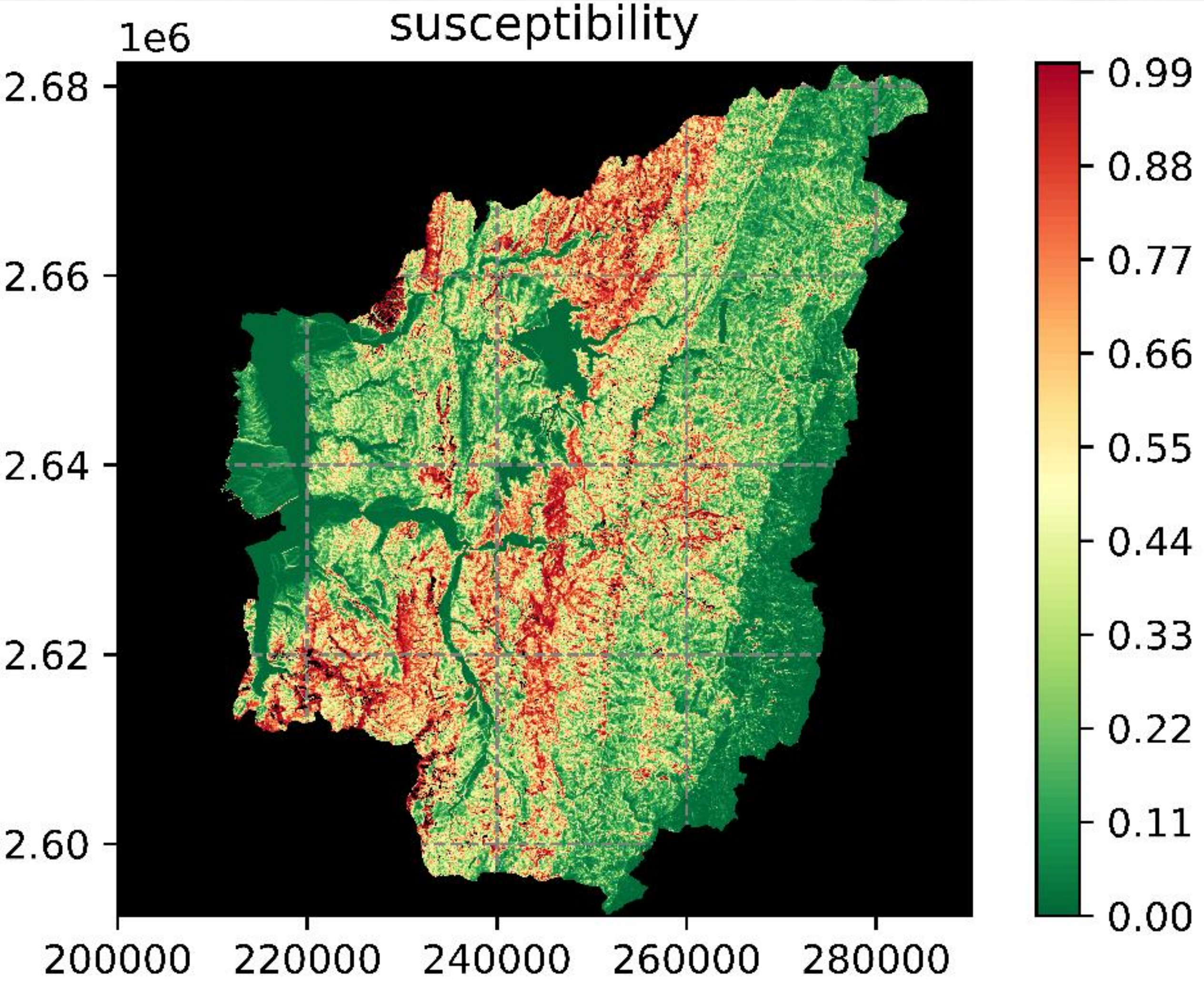
以修正後 Arias Intensity分區

Confusion matrix (model 1)		observation LS	observation NLS
prediction	LS	152799	3035719
prediction	NLS	41538	17019944

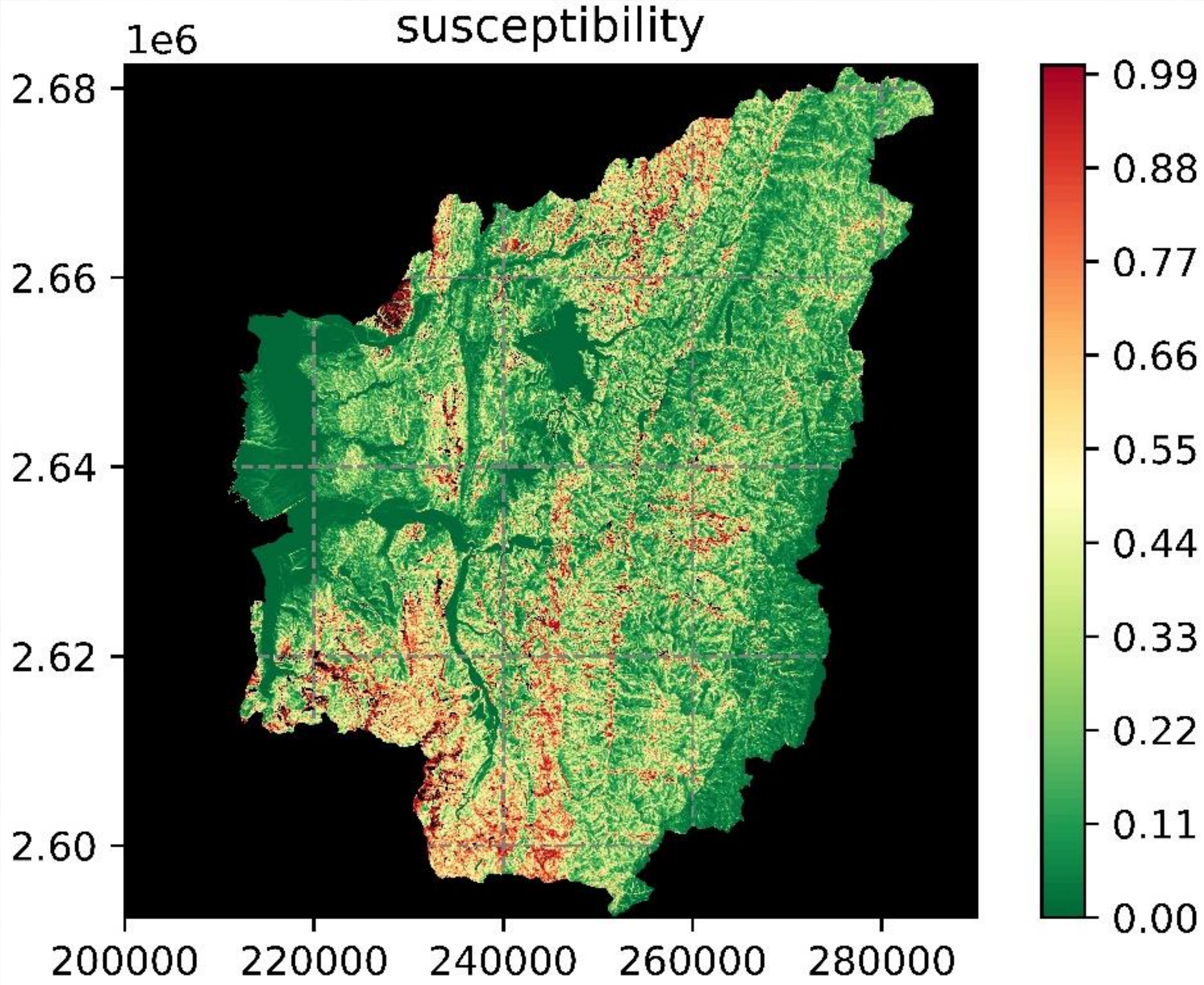
TPR: 0.048
TNR: 0.998
FPR: 0.952
FNR: 0.002
Precision: 0.048
Recall: 0.786

Accuracy: 0.848

不分區



以修正後 Arias Intensity分區



Training data

不分區傳統模型

Confusion matrix **(model 1)**		observation LS	observation NLS
prediction	LS	178916	38266
	NLS	33897	174540

TPR: 0.824
TNR: 0.837
FPR: 0.176
FNR: 0.163
Precision: 0.824
Recall: 0.841

Accuracy: 0.830

Training model AUC: 0.907

increase

全區取樣 分區建模

Confusion matrix **(model 1)**		observation LS	observation NLS
prediction	LS	183416	35734
	NLS	29396	177070

TPR: 0.837
TNR: 0.858
FPR: 0.163
FNR: 0.142
Precision: 0.837
Recall: 0.862

Accuracy: 0.847

Training model AUC: 0.923

Training data

分區取樣 分區建模

Confusion matrix (model 1)		observation LS	observation NLS
prediction	LS	166614	44444
prediction	NLS	46310	167655

TPR: 0.789
TNR: 0.784
FPR: 0.211
FNR: 0.216
Precision: 0.789
Recall: 0.783

Accuracy: 0.868

Training model AUC: 0.868

increase

全區取樣 分區建模

Confusion matrix (model 1)		observation LS	observation NLS
prediction	LS	183416	35734
prediction	NLS	29396	177070

TPR: 0.837
TNR: 0.858
FPR: 0.163
FNR: 0.142
Precision: 0.837
Recall: 0.862

Accuracy: 0.847

Training model AUC: 0.923

Testing data

不分區

Confusion matrix **(model 1)**		observation LS	observation NLS
prediction	LS	169517	3998320
prediction	NLS	24820	16057343

TPR: 0.041
TNR: 0.998
FPR: 0.959
FNR: 0.002
Precision: 0.041
Recall: 0.872

Accuracy: 0.801

Testing model AUC: 0.913

全區取樣 分區建模

Confusion matrix **(model 1)**		observation LS	observation NLS
prediction	LS	178549	4738958
prediction	NLS	15788	15316705

TPR: 0.036
TNR: 0.999
FPR: 0.964
FNR: 0.001
Precision: 0.036
Recall: 0.919

Accuracy: 0.765

Testing model AUC: 0.922

Testing data

分區取樣 分區建模

Confusion matrix **(model 1)**		observation	
		LS	NLS
prediction	LS	152799	3035719
	NLS	41538	17019944

TPR: 0.048
TNR: 0.998
FPR: 0.952
FNR: 0.002
Precision: 0.048
Recall: 0.786

Accuracy: 0.848

Testing model AUC: 0.905

全區取樣 分區建模

Confusion matrix **(model 1)**		observation	
		LS	NLS
prediction	LS	178549	4738958
	NLS	15788	15316705

TPR: 0.036
TNR: 0.999
FPR: 0.964
FNR: 0.001
Precision: 0.036
Recall: 0.919

Accuracy: 0.765

Testing model AUC: 0.922