以愛氏震度分區建立地震誘發山崩潛感羅吉斯回歸模型以及地震後山崩潛感近及時分析

沈楷庭[1]、董家鈞[1]、李錫堤[1]、曾佳漢[2]

(1) 國立中央大學應用地質研究所、(2)中國文化大學地質學系

研究動機及目的

地震造成的損失除了直接導因於地震動外，地震動所引發的山崩也是生命財產損失之重要原因，本研究目標為半自動化近即時繪製山崩潛感圖，盡量減少人工資料處理時間，使用羅吉斯回歸計算潛感值，並撰寫了半自動化Python程式計算羅吉斯回歸模型潛感值，經過測試可以在1分鐘內建立完山崩潛感模型，並在15分鐘左右繪製出山崩潛感圖。

研究方法

本研究使用羅吉斯回歸建立山崩潛感模型，透過山崩與非山崩資料經過計算可得式(1)：

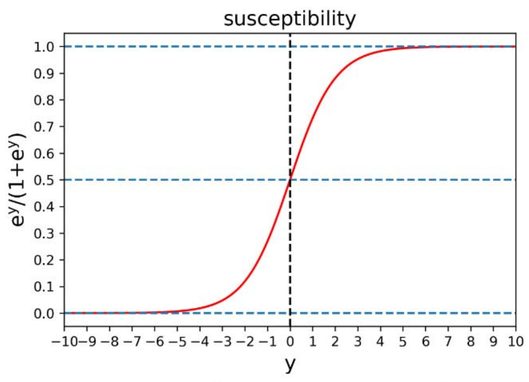
$$y = a_0 + \sum_{k=1}^{n} w_k x_k \qquad 式(1)$$

a_0:截距　w_k:k因子權重

x_k:k因子標準化後之值

接著將資料集中因子值代入式(1)可得y值，再代入式(2)即可得潛感值：

$$susceptibility = \frac{e^y}{1+e^y} \qquad 式(2)$$



圖一 潛感值值域分布曲線

研究模型

本研究使用不同於以往的方法設計三種模型，主要差別在於是否將誘發因子納入模型中：

一、全區建模，全區等比例取樣建模：

此為傳統之建模方法，將訓練資料之地形因子及誘發因子納入羅吉斯回歸模型中，並以山崩與非崩數量接近1:1之方式進行建模。

二、分區建模，全區等比例取樣建模：

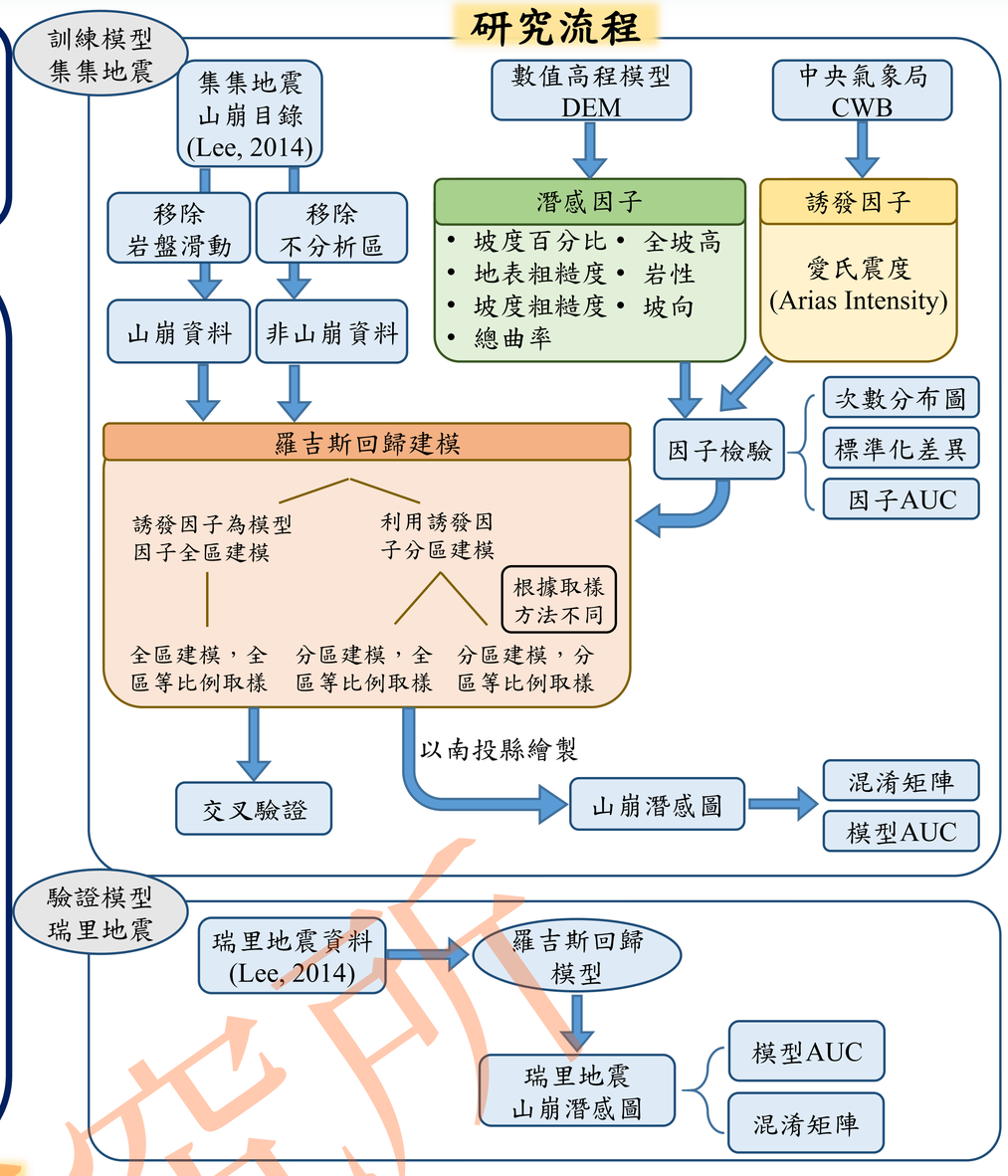
此方法為本研究設計之新建模方法，先進行山崩與非山崩取樣，再以誘發因子為分區標準且不放入模型內，將訓練集切成數等份，最後對每區資料進行建模，因此不同愛氏震度下會對應到不同之羅吉斯回歸模型。

三、分區建模，分區等比例取樣建模：

此方法與前一模型差別在於取樣的順序不同，此方法會先以誘發因子進行分區，接著再對分區內的資料進行山崩與非山崩的取樣，最後一樣對各分區資料進行建模。

模型二與模型三因為取樣與分區的先後順序不同，影響了山崩與非山崩資料比例的問題。

研究流程

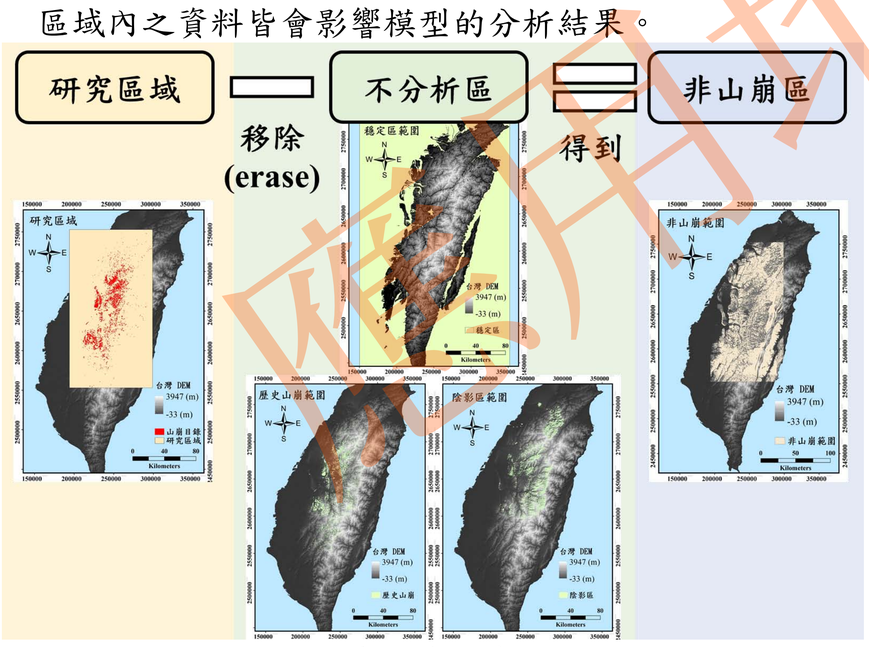


資料處理及統計方法

在進行山崩潛感建模之前，需要先整理訓練資料集，訓練資料內分為山崩資料以及非山崩資料，各資料皆有需要放入模型之因子值，因子又分為連續變數以及類別變數，坡度、曲率、全坡高等等為連續變數，岩性、坡向則為類別變數，以下為山崩與非山崩資料選擇方法：

山崩資料：使用李錫堤教授研究室所繪製之集集地震山崩目錄，先將研究範圍的網格資料轉換成資料點，而所有位在山崩目錄內的資料點則為山崩資料。

非山崩資料：將在非山崩區域隨機選擇與山崩資料數量相似之資料點為非山崩資料，位在山崩目錄外的範圍即為非山崩區域，根據李錫堤教授之建議，另需刪除穩定區、陰影區及歷史山崩區，因為位在這些區域內之資料皆會影響模型的分析結果。



圖二 非山崩區資料處理方法

在製作完山崩潛感模型後，將使用混淆矩陣以及AUC說明模型的好壞：

混淆矩陣

從混淆矩陣中可以看出資料有多少資料被正確預測或是誤判，而其可衍伸出兩個統計參數為召回率與準確率：

召回率：預測到山崩與實際山崩的比例

準確率：預測正確與所有事件的比例

$$召回率(Recall) = \frac{TP}{TP+FN}$$

$$準確率(Accuracy) = \frac{TP+TN}{TP+TN+FP+FN}$$

混淆矩陣		觀測值	
		LS	NLS
預測值	LS	TP	FP
	NLS	FN	TN

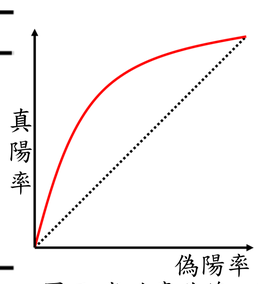
LS : Landslide
NLS : Non-Landslide
TP : True Positive
FP : False Positive
TN : True Negative
FN : False Negative

Area Under Curve (AUC)

AUC代表的是成功率曲線下之面積，而成功率為真陽率與偽陽率之關係，根據設定不同門檻值計算真陽率及偽陽率，並繪製成曲線，若曲線越接近左上代表模型表現越佳，因此AUC值則會越大，

AUC之值	模型結果
AUC > 0.9	非常準確
0.9 > AUC > 0.8	準確
0.8 > AUC > 0.7	普通
0.7 > AUC > 0.6	不良
0.6 > AUC > 0.5	非常不良

表一 AUC對應之模型結果



圖三 成功率曲線

研究結果

經過羅吉斯回歸計算後可得模型中各因子之權重，以傳統羅吉斯回歸模型為例，如表一，其中以坡度及愛氏震度權重最高，代表此兩因子較能區分山崩以及非山崩資料，而全區取樣分區建模以及分區取樣分區建模模型由於每個分區皆有一個模型，因此回歸結果則不在此一一列出。

接著使用南投縣所有網格資料點繪製山崩潛感圖，當成驗證資料並利用統計方法驗證模型，根據召回率來看，全區取樣分區建模為最佳，代表此模型預測山崩的能力比以往傳統模型更佳，但整體的準確率卻下降。

表二 傳統模型因子權重

因子	權重
坡度	0.958
地表粗糙度	0.567
坡度粗糙度	0.221
曲率	0.270
全坡高	0.208
愛氏震度	0.987

※坡向、岩性為類別變數，未一一列出

表三 傳統模型混淆矩陣

混淆矩陣(傳統模型)		觀測值	
		山崩	非山崩
預測值	山崩	164934	4285959
	非山崩	29403	15769704

召回率:0.849　準確率:0.786

表四 全區取樣混淆矩陣

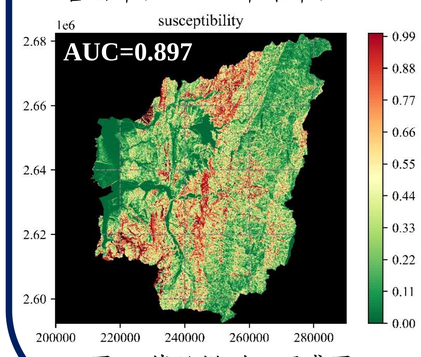
混淆矩陣(全區取樣)		觀測值	
		山崩	非山崩
預測值	山崩	177419	4904383
	非山崩	16918	15151280

召回率:0.913　準確率:0.756

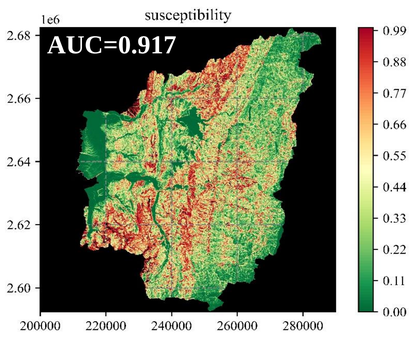
表五 分區取樣混淆矩陣

混淆矩陣(分區取樣)		觀測值	
		山崩	非山崩
預測值	山崩	157923	4009565
	非山崩	36414	16046098

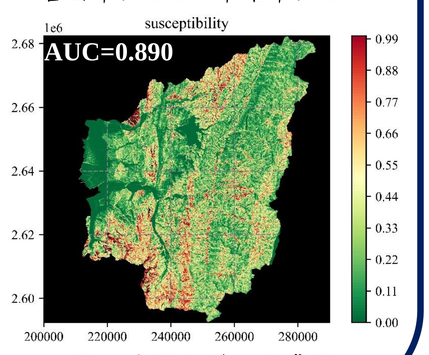
召回率:0.813　準確率:0.800



圖四 傳統模型之潛感圖



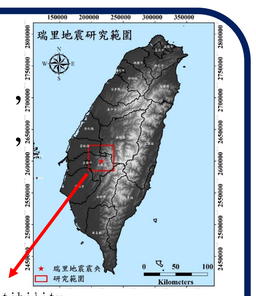
圖五 全區取樣之潛感圖

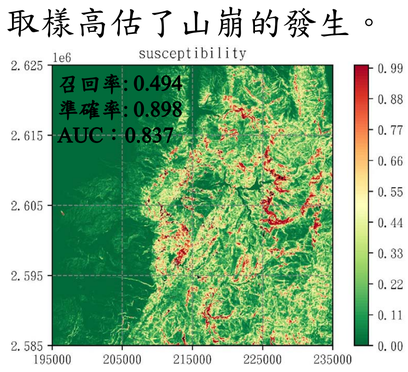


圖六 分區取樣之潛感圖

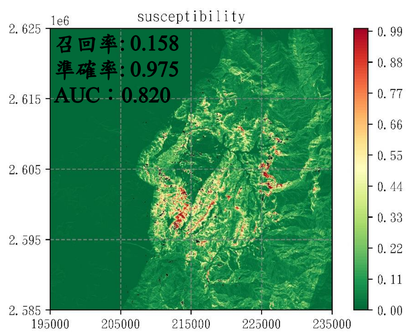
驗證資料

在製作完模型後，本研究選擇了另一次的地震事件來驗證，目的是要了解若是未來要拿此模型對地震事件做預測結果會如何，因此選擇了1998年的瑞里地震作為測試事件。由此結果可以發現，召回率以分區取樣最高但準確率最低；準確率以全區取樣最高，但召回率最低，說明了全區取樣稍微低估了山崩的發生，而分區取樣高估了山崩的發生。

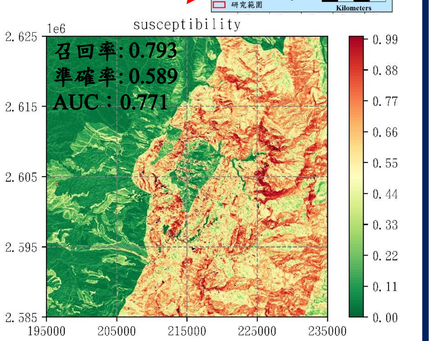




圖七 傳統模型之測試資料潛感圖



圖八 全區取樣之測試資料潛感圖



圖九 分區取樣之測試資料潛感圖

結論

1. 以驗證資料來看，若是要達到防災等目的，可以選擇分區建模，全區等比例取樣，因其召回率最高，代表有最多的山崩被準確預測；而若是要兼顧山崩與非山崩，則可選擇全區建模，全區等比例取樣，雖其準確率不是最高，跟分區建模，分區等比例取樣比起來雖犧牲了一點準確率，但提升了較多的召回率。
2. 以測試資料來看，分區建模，全區等比例取樣有些低估山崩，而分區建模，分區等比例取樣有些高估山崩，全區建模，全區等比例取樣則有較合理的分布，但值得注意的是，全區建模，全區等比例取樣的研究區域邊緣持續有偏中高潛感值存在，意味著若是將繪圖區域擴大，或許會持續有高潛感出現造成準確率下降，而分區建模，全區等比例到外圍則逐漸收斂，對於大範圍的山崩潛感圖或許較準確。

參考資料：

1. M.A. (Nowicki) Jessee, M.W. Hamburger, K. Allstadt, D.J. Wald, S.M. Robeson, H. Tanyas, M. Hearne, and E.M. Thompson (2017). A Global Empirical Model for Near Real-time Assessment of Seismically Induced Landslides. ***Journal of Geophysical Research: Earth Surface.*** doi: 10.1029/2017JF004494
2. Chyi-Tyi Lee(2014). Statistical seismic landslide hazard analysis: An example from Taiwan. ***Engineering Geology***, 182, 201–212.

應用地質研究所