

應用空間分布降雨資料於數據 驅動淹水模型之訓練及測試

坡地與洪旱組 江申 陳偉柏 魏曉萍 郭文達 張志新

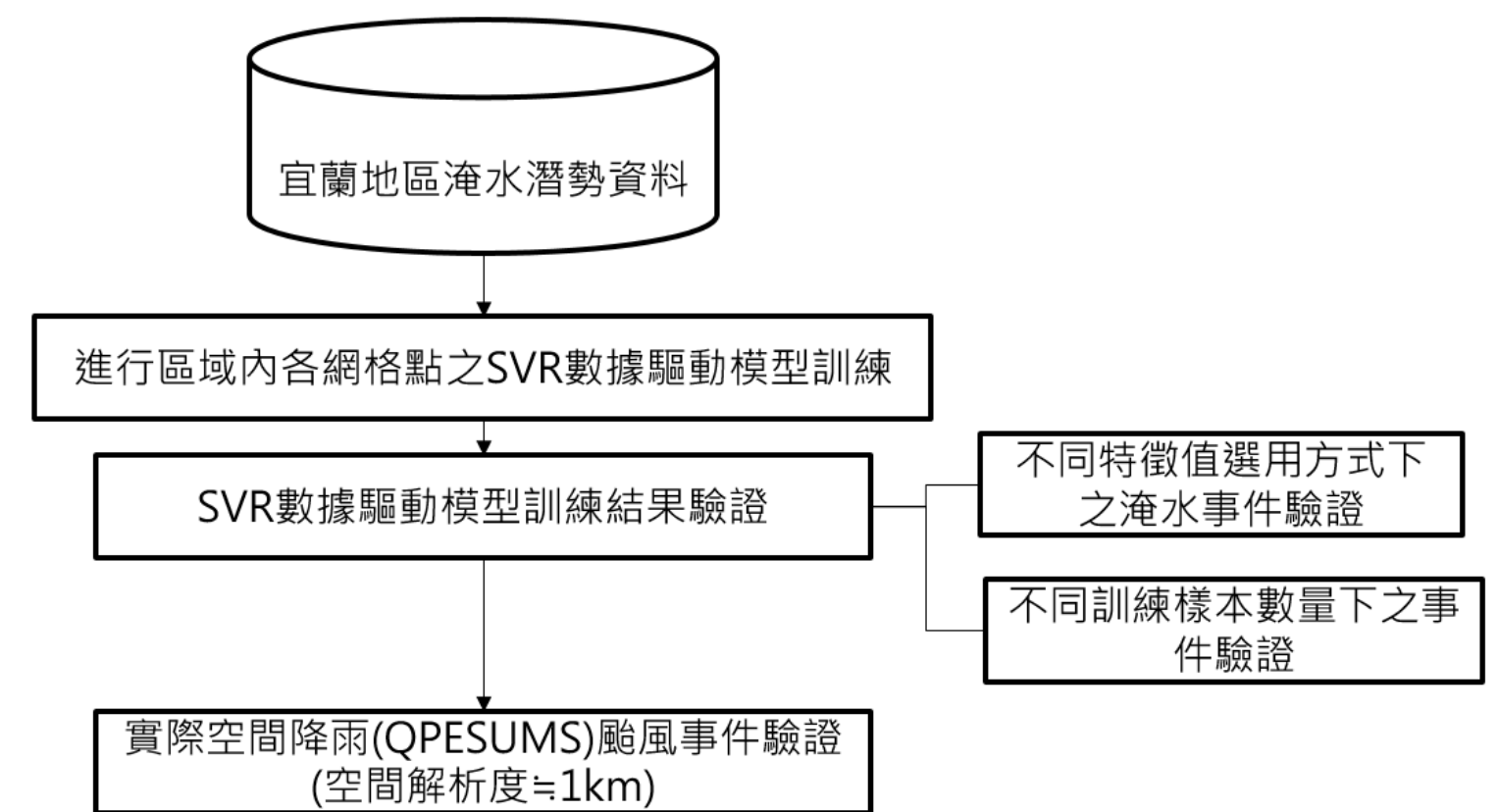


前言

本研究旨在探討數據驅動模式在淹水模擬中之應用。傳統淹水模式的建置需要大量的地文、水文資料以及參數調校，計算資源的耗費也相對較高。相較之下，數據驅動模式的建置無需收集大量的基本資料，且模型的訓練僅需決定演算法及相對應的輸入特徵值與輸出目標值即可。因此，本研究採用數據驅動模式，以「臺灣氣候變遷推估資訊與調適知識平台計畫」所提供的宜蘭地區淹水潛勢圖進行模型建置。其中，本研究所建置之模型為基於支持向量機演算法的支持向量回歸模型，以空間降雨時序列與淹水時序列進行模型訓練。本研究嘗試比較不同特徵值採用策略及模型訓練樣本數之下，數據驅動模式的表現。最後，本研究以歷史颱風事件之QPESUMS資料及淹水區位調查資料進行模式驗證，以評估模型之準確性與可靠性。

研究流程

本研究使用TCCIP產製之宜蘭地區淹水潛勢圖資，此淹水潛勢圖係應用AR5動力降尺度全球大氣模式(High-Resolution Atmospheric Model, HiRAM)之基期(1979-2008)149場颱風事件之時序列空間降雨資料(空間解析度5公里)，透過SOBEK模式得出研究區域內各計算網格點之逐時空間淹水深度資料(空間解析度40公尺)。研究中對此淹水潛勢圖資進行資料清洗及格式轉換，其後對研究區域內之各地表淹水計算格點，以不同特徵值選用策略及不同訓練樣本數量方式建置SVR淹水數據驅動模型，並比較其模式驗證結果。研究流程圖如右圖所示：



研究方法

研究中使用數據驅動模式之演算法為基於支持向量機(Support Vector Machines, SVMs) 演算法之支持向量回歸 (Support Vector Regression, SVR)。作法為於SVM中以核函數[8]將低維度之非線性問題轉換為高維度特徵空間後，再加入損失函數後之預測模型即成為線性回歸問題，為一種監督式學習方法。研究中目標值(target) 為特定時刻之淹水深度；特徵值 (feature) 為特定時刻前雨量值或特定時刻前淹水深度。特徵值之數量多寡將反映於模型訓練花費時間，一般來說特徵值數量愈多，模式訓練所費時間則愈久，訓練完成之模型效果亦可能較好。研究中使用兩組不同之特徵值取用方法，試圖通過比較得到表現較合理之模型訓練結果。第一種特徵值選用策略如式(1)所示：

$$D(t) \ 1:r(t-1) \ 2:r(t-2) \ 3:r(t-3) \quad (1)$$

其中D(t)表計算網格點於時刻t之淹水深度、r(t-1)、r(t-2)及r(t-3)則表示t之前1、2及3時刻之降雨量。第二種特徵值選用策略如式(2)所示：

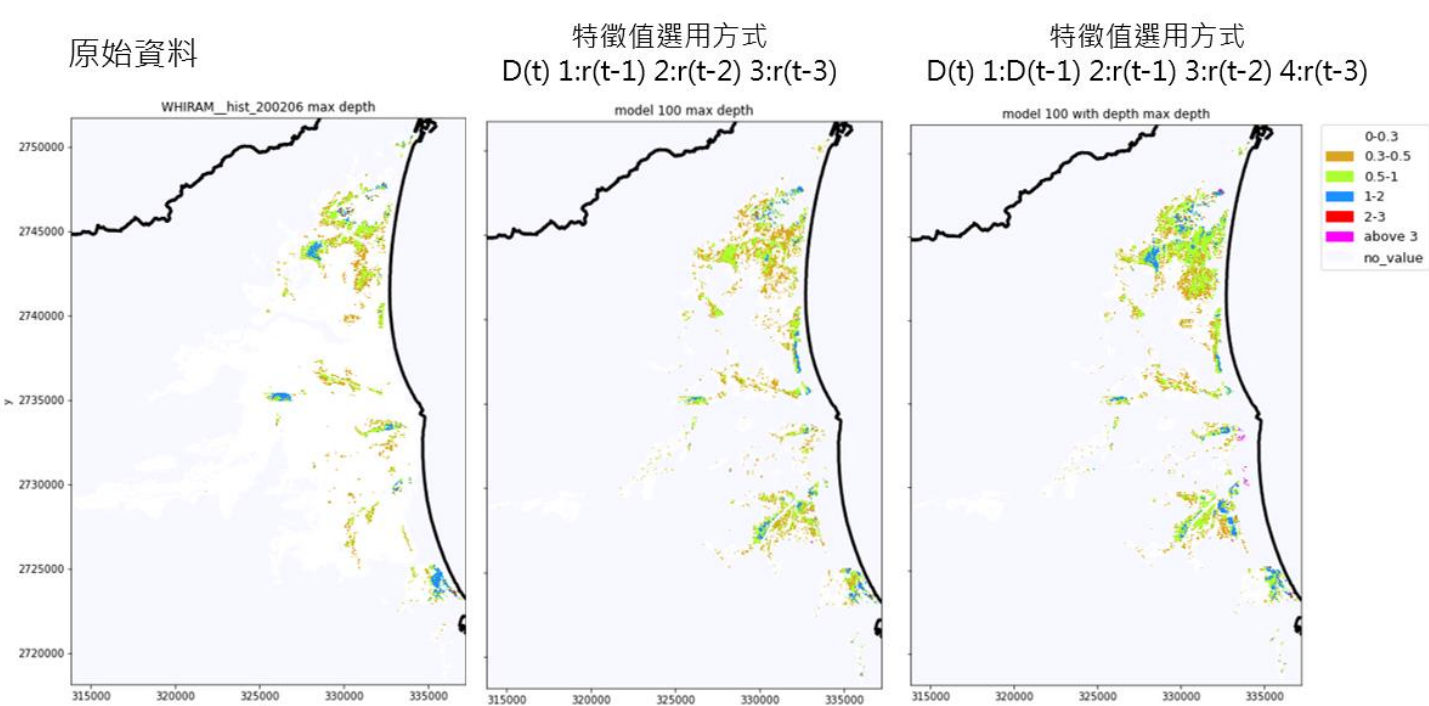
$$D(t) \ 1:D(t-1) \ 2:r(t-1) \ 3:r(t-2) \ 4:r(t-3) \quad (2)$$

其中D(t)表計算網格點於時刻t之淹水深度、D(t-1)表時刻t-1之淹水深度、r(t-1)、r(t-2)及r(t-3)則表示t之前1、2及3時刻之降雨量。

研究成果

SVR數據驅動模式於不同特徵值選用方式下之淹水事件驗證結果

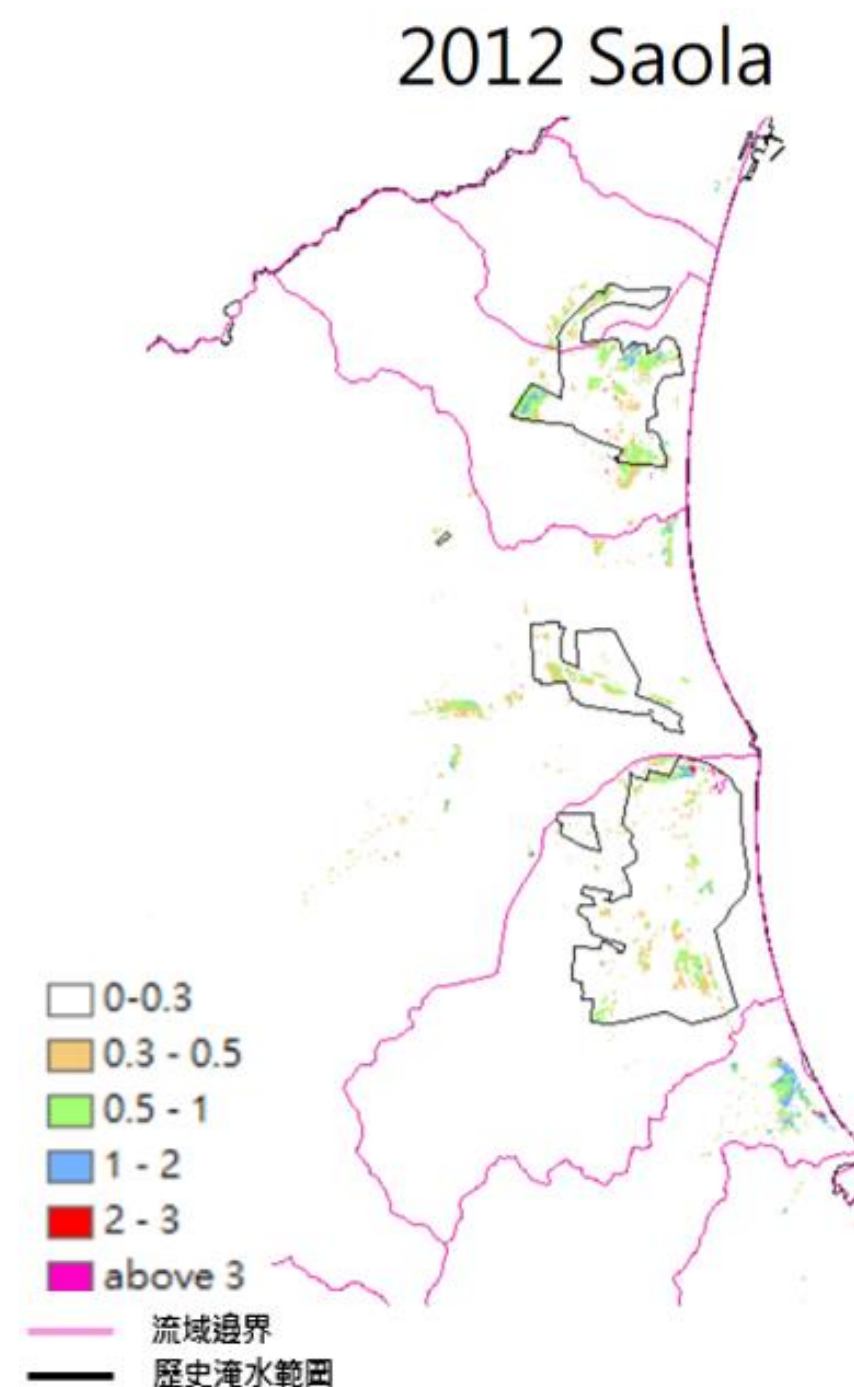
下圖為透過兩種不同特徵值取用策略進行模型訓練後，以指定颱風事件(未使用於模型訓練資料集)進行驗證後與原始資料之最大淹水深度比較圖。



由圖中可以看出，兩種特徵值選用策略均可大致重現指定颱風事件之淹水範圍，惟第二種特徵值選用策略之驗證結果在淹水深度上較第一種特徵值選用策略驗證結果為接近。

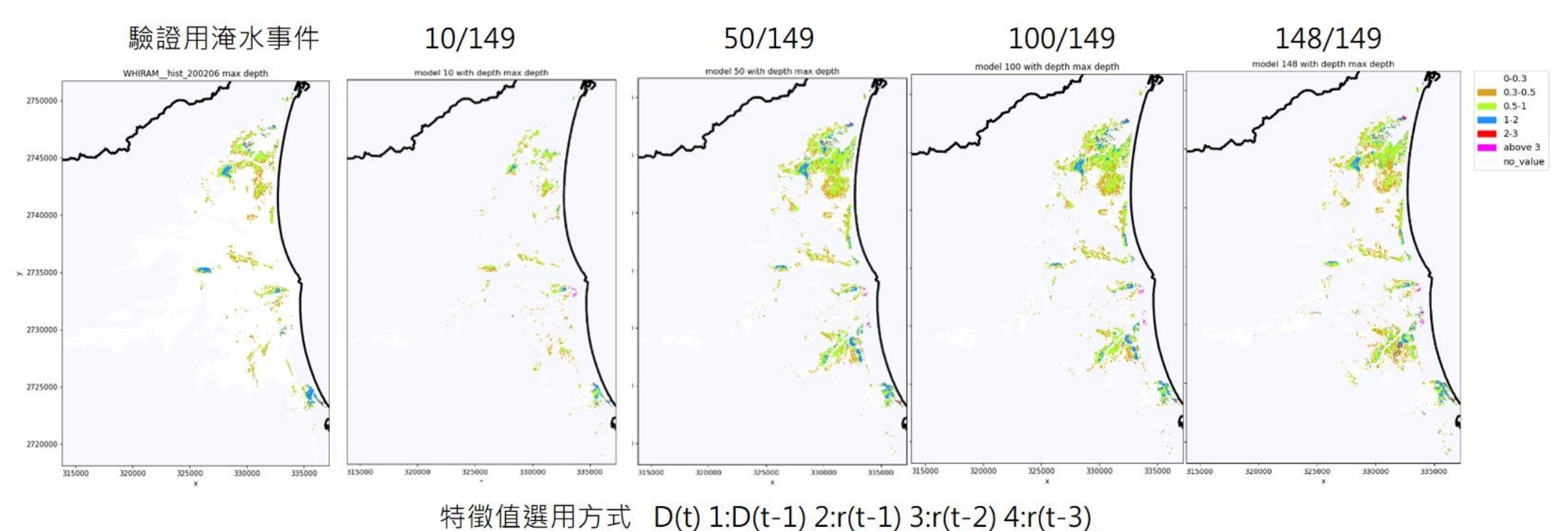
真實颱風事件驗證結果

2012 Saola蘇拉颱風夾帶豐沛水氣，影響臺灣期間 (2012/06/28 06:00~2012/06/29 09:00)，雨量主要分布在宜蘭陽平原南側 (太平山區)、雪山山脈西麓包括桃竹苗山區及台中、南投山區。造成多處地區發生土石流、淹水、道路中斷等災情，共有7人死亡，農損逾12億元。右圖為蘇拉颱風之淹水範圍及數據驅動模式之驗證成果。由圖中可看出模式之驗證結果淹水範圍大致吻合淹水範圍調查結果，但蘇澳地區不在淹水調查範圍之內，但數據驅動模式之驗證結果在此區有相當深度之淹水，需待進一步資料收整及研判。



SVR數據驅動模型之不同訓練樣本數量下之事件驗證結果

除數據驅動模型之特徵值選取會影響到模型訓練結果之外，訓練之樣本數也可能會影響訓練結果。研究中使用149場颱風事件中之10、50、100及148場事件進行模型訓練，並將驗證結果與未使用於模型訓練之颱風淹水事件進行比較，下圖為其最大淹水深度之比較。



特徵值選用方式 D(t) 1:D(t-1) 2:r(t-1) 3:r(t-2) 4:r(t-3)

由圖中可看出以50、100及148場事件進行模型訓練所得之驗證結果在淹水之深度及範圍上差距不大，但與以10場事件進行模型訓練所得之驗證結果相差較大。經與驗證用淹水事件比較，50、100及148場事件進行模型訓練所得之驗證結果在深度及範圍上均較驗證用淹水事件為深及廣；以10場事件進行模型訓練所得之驗證結果則較驗證用淹水事件在淹水深度及範圍上為小。但若以時序列進行比較，則以10場事件進行模型訓練所得之驗證結果較為合理。

結論、建議及未來展望

1. 淹水數據驅動模式之建置不需要大量之地文、水文及水工結構物等資料之收整工作，運算時間亦較傳統水理淹水模式之速度為快。以本研究為例，研究中所使用產製淹水訓練用資料集之SOBEK水理模式，演算24小時之淹水需時約90分鐘，而本研究之數據驅動模式僅需約10分鐘。
2. 數據模型訓練時之特徵值選用策略將影響數據驅動模式之良窳，本研究使用兩種特徵值選用策略進行模型訓練並驗證其訓練結果，比較結果顯示增加前一時刻淹水深度作為特徵值進行模型訓練之模式驗證結果優於僅使用淹水之前三時刻降雨作為特徵值之模型訓練結果。
3. 雖然加入前一時刻之淹水深度作為特徵值已可大致重現淹水之空間範圍，但各淹水點位之時序列變化與驗證事件尚有差距，未來可嘗試增加淹水時刻前二時刻或前三時刻之淹水深度作為特徵值，或將鄰近網格點之淹水深度亦加入特徵值進行模型訓練。
4. 研究中於149場次颱風淹水事件中取用10、50、100、150及148場次進行模型訓練，並比較其模式驗證結果。結果發現50、100、150及148場次之模式驗證結果相差不大。
5. 數據驅動模式之建置用模型訓練樣本之代表性(時空間變異性)應較樣本數量重要，如何確認訓練樣本之代表性可為下階段研究重點。