

國立臺灣大學工學院環境工程學研究所



碩士論文

Graduate Institute of Environmental Engineering

College of Engineering

National Taiwan University

Master Thesis

應用支援向量回歸於河川洪水預警之研究

——以朴子溪為例

Application of Support Vector Regression in Flood  
Forecast —— Taking the Puzi River as an Example

薛天一

Tian-Yi Xue

指導教授：駱尚廉 教授

Advisor：Prof. Shang-Lien Lo

中華民國 109 年 7 月

July, 2020

# 國立臺灣大學碩士學位論文

## 口試委員會審定書

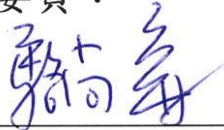
應用支援向量回歸於河川洪水預警之研究—以朴子溪為例

Application of Support Vector Regression in Flood

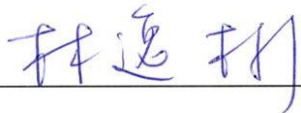
Forecast —— Taking the Puzi River as an Example

本論文係薛天一君(學號 R07541214)在國立臺灣大學環境工程學研究所完成之碩士學位論文，於民國 109 年 7 月 31 日承下列考試委員審查通過及口試及格，特此證明

論文審查委員：



駱尚廉博士  
國立臺灣大學環境工程學研究所特聘教授

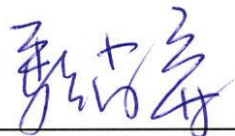


林逸彬博士  
國立臺灣大學環境工程學研究所教授




林進榮博士  
國立宜蘭大學環境工程學系教授

指導教授：



所 長：



## 誌謝



回顧過去兩年的學習生涯，首先要謝謝我的指導教授駱尚廉老師。從平時的課程到碩論研究，老師對於學生的想法特別寬容。無論我提出的問題大小與否，老師一樣會耐心細緻地解答並給予建議，偶爾與同學們分享的人生經驗和待人處事方式也讓我受益良多。論文口試期間內，十分感謝林逸彬教授以及宜蘭大學環工系林進榮教授百忙之中抽空參與，讓這篇碩士論文更臻于完善，在此表示深摯的謝忱。

就讀研究所的兩年時光裡，少不了駱家班、環工所同學、老師們的幫助。確定研究方向時雅順提出了不少建議；曉芸、俊華、一丁、越東在我無法返校的日子裡經常關心近況；口試前夕心潔、風光不斷為我加油打氣，在此一併表示感謝，祝各位未來順心順意。

另外在學校也受到藝寬非常非常多的照顧，不斷為我分擔煩惱、難過和各種糟糕情緒，真的太感謝啦。也謝謝寧宇在論文撰寫時互相勉勵，生活中相互扶持，能碰到這麼好的室友真是非常幸運，以後也要多多聯絡。

當然還有臺北的仲倫、孟庭；花蓮的玟濤；廣州的昀燁和高中同學，意銘和中大電信院學生會網宣部的朋友，天偉、媛媛、古典和大院的好朋友；古來的宗信；吉隆坡的偉恩；福州的俊維；交大的孟勳；西安大略大學的郭洋；上海海事大學的勇志；明尼蘇達大學的澤南；洛杉磯加大的寶文。謝謝各位一直以來的關照和肯定，下次聚會我做東。

最後也是最想感謝的是我的父母與家人，你們的支持讓遠赴異鄉讀書的我後顧無憂，可以全心全意地投身學習，希望我的努力也會給你們帶來喜悅！

## 摘要

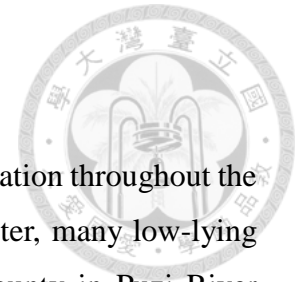


臺灣屬於副熱帶溼潤氣候，夏季雨量充沛，河川逕流量較大。在 2018 年的 823 水災中，部分低窪地區遭遇了洪水及淹水問題，帶來負面影響，其中便包括了朴子溪流域的嘉義縣市。因此，本研究利用嘉義朴子河流域過往之水位及降雨資料，建立支援向量回歸、灰色水位、線性回歸三種水位預測模式，並嘗試在不處理原始資料雜訊的前提下進行河川水位預測，以期判斷各模式是否可用於快速提供當地洪水預警。

本研究使用朴子河流域牛稠溪橋(1)、竹崎、番路三個測站 2017 年一整年之水位、降雨資料率定模式，使用 2018 年三場降雨事件分別作驗證，其中包括了 823 水災。研究先確定洪峰延滯時間為 3 小時，並以此作為預警時間。而後建立支援向量回歸、灰色水位、線性回歸三種模式分別預測三場降雨事件，以事件一為例，支援向量回歸的決定係數、均方根誤差、最大絕對誤差、尖峰水位誤差均為最佳，分別為 0.9902、0.1754、0.9341、0.5713。再根據三場預測得到各模式之總體決定係數，分別為 0.9741、0.9434、0.9560。經由對比三種模式後得到，各模式都可在一定程度上預測河川水位高度，其中以支援向量回歸為最優，其健壯性 (Robustness) 相對另外兩種模式較佳。

關鍵字：洪水預警、水位預測、支援向量回歸、灰色模式、朴子溪

## Abstract



Due to the humid subtropical, Taiwan receives abundant precipitation throughout the summer, which result a high river discharge. In the 823 flood disaster, many low-lying areas experience flooding during heavy rains, including Chiayi county in Puzi River Basin. In this study, three models are utilized for forecasting: Support Vector Regression, Gray Model, Linear Regression. Values of input data was not denoised, in order to provide flood warning as fast as possible. The purpose of this study is to construct a quickly water stage forecasting model at Puzi River.

In this study, water stage and precipitation of the 1550H017, COM700, COM720 station during 2017 is used to establish the models, and three rainfall events from 2018 is used to verification, including the 823 flood disaster. At first, determine lag time of 3 hours as early warning time. Then, establishing Support Vector Regression, Gray Model, Linear Regression model, forecasting three rainfall events. Take Event One as an example, the  $R^2$  (Coefficient of determination), RMSE(Root mean square error), MAE (Maximum absolute error) and EPWL (Error of peak water level) of Support Vector Regression are the best, which are 0.9902, 0.1754, 0.9341 and 0.5713, respectively. In addition, the overall  $R^2$  of each formula are 0.9741, 0.9434 and 0.9560, respectively.

By comparing the three models, the water stage prediction model of Support Vector Regression is more effective in predicting river water level, and its robustness is stronger than the other two models.

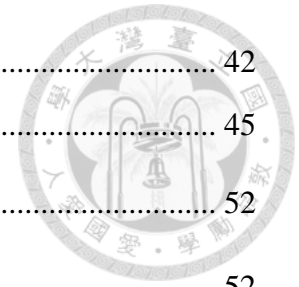
Keywords : Flood Warning 、 Water Stage Forecast 、 Support Vector Regression 、 Gray Model 、 Puzi River

# 目錄



第一章 前言 .....	1
1.1 研究緣起 .....	1
1.2 研究目的 .....	2
第二章 研究背景與文獻回顧 .....	3
2.1 研究區域地文條件概述 .....	3
2.2 研究區域氣象水文資料概述 .....	4
2.3 支援向量回歸 .....	6
2.4 灰色系統理論 .....	7
2.5 線性回歸 .....	11
2.6 模式研究回顧 .....	11
第三章 研究方法與過程 .....	13
3.1 研究區域與暴雨事件之選取 .....	13
3.2 支援向量回歸水位預測模式 .....	16
3.3 灰色水位預測模式 .....	21
3.4 線性回歸預測模式 .....	25
3.5 水位傳播時間 .....	26
3.6 模式評鑒指標 .....	28
第四章 結果與討論 .....	30
4.1 模式運行環境 .....	30
4.2 模式率定資料概述 .....	30
4.3 模式預測資料概述 .....	33
4.4 支援向量回歸預測結果 .....	35
4.5 灰色水位預測結果 .....	39

4.6 線性回歸預測結果 .....	42
4.7 三種模式之比較 .....	45
第五章 結論 .....	52
5.1 結論 .....	52
5.2 建議 .....	52
參考文獻 .....	54



# 表目錄



表 2.2-1 研究區域歷年平均降雨量與平均流量.....	5
表 3.1-1 流域內之水位測站.....	13
表 3.1-2 流域及周邊之雨量測站.....	14
表 3.1-3 研究使用之水位及雨量測站資料.....	15
表 3.1-4 研究使用之預測驗證事件.....	16
表 3.5-1 水位傳播時間相關性分析.....	27
表 4.2-1 原始資料分析.....	33
表 4.3-1 待預測事件之降雨量分析.....	35
表 4.3-2 待預測水位資料分析.....	35
表 4.4-1 支援向量回歸水位預測結果.....	38
表 4.5-1 灰色模式水位預測結果.....	41
表 4.6-1 線性回歸模式水位預測結果.....	44
表 4.7-1 事件一之分析結果.....	46
表 4.7-2 事件二之分析結果.....	47
表 4.7-3 事件三之分析結果.....	49



## 圖目錄



圖 2.1-1 朴子溪流域圖 .....	3
圖 2.2-1 歷年颱風侵襲路徑與概率 .....	5
圖 2.3-1 支援向量預測流程圖 .....	7
圖 2.4-1 灰色預測流程圖 .....	8
圖 3.1-1 朴子溪流域之水位及雨量測站位置圖 .....	13
圖 3.1-2 研究使用之水位及雨量測站位置圖 .....	15
圖 3.2-1 線性回歸與支援向量回歸之比較 .....	18
圖 3.2-2 線性回歸與支援向量回歸之損失函數比較 .....	18
圖 3.2-3 非線性支援向量回歸架構示意圖 .....	21
圖 4.2-1 番路雨量測站 2017 全年之日降雨量 .....	31
圖 4.2-2 牛稠溪橋(1)測站 2017 全年之水位高度 .....	31
圖 4.2-3 竹崎雨量測站 2017 全年之日降雨量 .....	32
圖 4.2-4 流域上游 2017 全年之平均日降雨量 .....	32
圖 4.3-1 事件一之降雨量與水位關係圖 .....	33
圖 4.3-2 事件二之降雨量與水位關係圖 .....	34
圖 4.3-3 事件三之降雨量與水位關係圖 .....	34
圖 4.4-1 事件一支援向量回歸水位預測誤差圖 .....	36
圖 4.4-2 事件一支援向量回歸水位預測圖 .....	36
圖 4.4-3 事件二支援向量回歸水位預測誤差圖 .....	37
圖 4.4-4 事件二支援向量回歸水位預測圖 .....	37

圖 4.4-5 事件三支援向量回歸水位預測誤差圖 .....	38
圖 4.4-6 事件三支援向量回歸水位預測圖 .....	38
圖 4.5-1 事件一灰色模式水位預測圖 .....	39
圖 4.5-2 事件二灰色模式水位預測圖 .....	40
圖 4.5-3 事件二灰色模式水位預測誤差圖 .....	40
圖 4.5-4 事件一灰色模式水位預測誤差圖 .....	40
圖 4.5-5 事件三灰色模式水位預測誤差圖 .....	41
圖 4.5-6 事件三灰色模式水位預測圖 .....	41
圖 4.6-1 事件一線性回歸水位預測圖 .....	42
圖 4.6-2 事件二線性回歸水位預測圖 .....	43
圖 4.6-3 事件一線性回歸水位預測誤差圖 .....	43
圖 4.6-4 事件二線性回歸水位預測誤差圖 .....	43
圖 4.6-5 事件三線性回歸水位預測圖 .....	44
圖 4.6-6 事件三線性回歸水位預測誤差圖 .....	44
圖 4.7-1 事件一之模式結果對比 .....	45
圖 4.7-2 事件一之模式誤差對比 .....	46
圖 4.7-3 事件二之模式誤差對比 .....	47
圖 4.7-4 事件二之模式結果對比 .....	47
圖 4.7-5 事件三之模式結果對比 .....	48
圖 4.7-6 事件三之模式誤差對比 .....	48
圖 4.7-7 支援向量回歸模式預測值與觀測值之散布圖 .....	50
圖 4.7-8 灰色模式預測值與觀測值之散布圖 .....	50

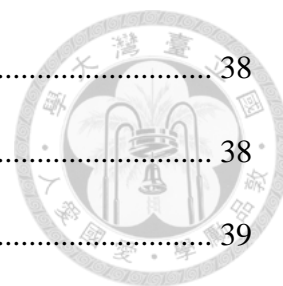


圖 4.7-9 線性回歸預測值與觀測值之散布圖.....51



# 第一章 前言



## 1.1 研究緣起

洪水是常見的自然災害之一，常常會帶來一定經濟損失並有可能導致民眾傷亡，對世界許多地區都產生了一定的負面影響。臺灣地處亞洲東部，屬於亞熱帶海島型氣候，雨量充沛，且降雨大多集中於夏、秋兩季。由於臺灣河川大多流短坡陡，每暴雨侵襲與颱風來臨，河川逕流量大大增加，將可能對流域的低窪地區帶來淹水影響。近年來，全球變暖導致極端降雨事件的頻率和強度增加，又使得洪災發生的頻率變高。此外，城市化過程會改變土地表面特性，並使得人口和個人財產集中在較小區域內，導致洪水所產生的潛在風險增加。尋找一套合理的預測方式，對於水利管理與災害防控具有重要意義（王如意、易任，1979）。

在自然過程中，洪水流量會涉及暴雨降雨量、強度、延時、面積與途徑。傳統水文預報方式是通過水位轉換建立降雨-逕流模式，利用歷史降雨事件之流量率定曲線計算逕流量，再將逕流量轉換為水位。但由於高流量資料較少，需採用外插法補足；同時由於泥沙淤積，河道斷面改變也會增加不確定性。河川水位是指河川某點之水面與一基準點（如海平面）比較之高度差，其尖峰水位可以作為防洪參考，最低水位則決定了灌溉、給水的限制。在災害應變實務中，相關單位常採用水位作為預警參考。

支援向量回歸（Support Vector Regression, SVR）屬於資料驅動模式，可使用較少資料進行即時預報，對於非線性及隨機性等特徵具有良好適應能力（Vapnik et al., 1997）。

灰色系統理論（Grey System Theory）則是水文學中較為傳統的預測方式之一，主要探討在訊息不完整的情況下，如何將問題簡化為數學模式進行分析，而不考慮其物理過程（鄧聚龍、郭洪，1996）。洪水預報所使用的雨量、水位資料十分有限，需要快速計算且可以適當捨棄精確度，符合灰色系統的處理要求。

線性回歸 (Linear Regression) 是最常見的回歸模式之一，原理為通過找出一個可以描述資料集的線性方程式，藉此預測未來的資料。在本研究中，用以代表過度簡化的數學模式，以期觀察簡化是否會帶來誤差。

因此本研究希望藉由分析朴子溪上游之降雨量資料，透過不同時期降雨及水位相關性推估出洪峰延滯時間，再以數學模式分析洪水水位與降雨之關聯性，並評估支援向量回歸與傳統灰色模式以及線性回歸方式間存在的差異。

## 1.2 研究目的

本研究利用朴子溪上游流域之降雨量資料，以灰色模式、支援向量回歸、線性回歸三種方式對嘉義市牛稠溪橋測站之水位進行推估，計算程式模擬數值與實際量測值的差異，評估模式可靠性並與傳統的灰色模式作出對比。

## 第二章 研究背景與文獻回顧



### 2.1 研究區域地文條件概述

本研究之研究區域為朴子河流域。朴子溪位於臺灣南部，流經嘉義縣境，主流長度 75.87 公里，流域面積 426.60 平方公里。分為平原區（面積 305 平方公里，佔全流域面積 72%）、丘陵區（面積 76 平方公里，佔全流域面積 18%）及山地區（面積 46 平方公里，佔全流域面積 11%）三部份（第五河川局，2019）。北臨北港溪流域，南臨八掌溪流域，介於東經 121°90'~120°40'，北緯 23°25'~23°30' 之間。

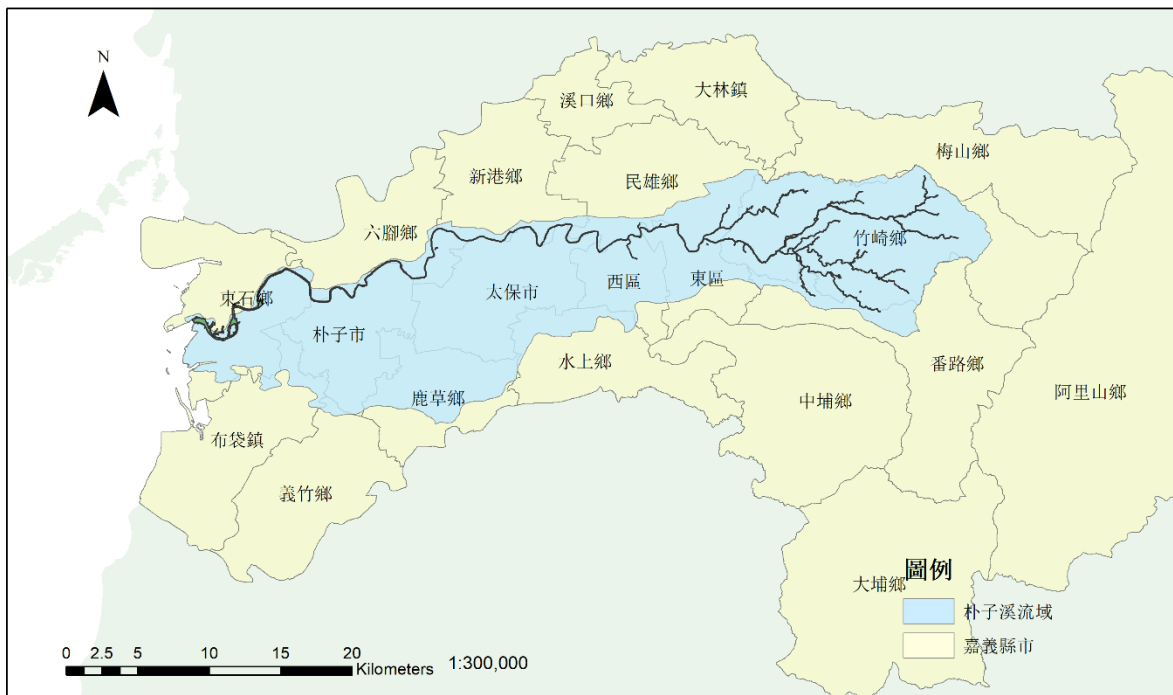


圖 2.1-1 朴子溪流域圖

（資料來源：內政部國土測繪中心、經濟部水利署）

朴子溪主流上游名為牛稠溪，因流經民雄鄉牛稠山而得名，發源於標高 1421 公尺的阿里山山脈四天王山芋菜坑。上游水面平均寬約 20 公尺，流速較快，棲地

型態豐富，分為深流、淺瀨及淺流。河床底質以卵石、礫石為主，灘地多為林地，周邊區域有種植香蕉、火龍果、檳榔及稻田（第五河川局，2017）。

中游河段之區域大約為嘉義市、民雄鄉邊界至太保市、新港鄉交界處，河川在月眉潭（地名）附近始稱朴子溪。此段河段之水面平均寬約 50 公尺，流速緩慢，棲地型態多為淺流及深流。河床底質以沉積泥砂為主，河床中有砂洲淤積，高灘地則為林地。在此河段中，河道兩側已設置防洪堤防，有零星住宅與工廠，種植作物主要為稻米及番茄。（第五河川局，2017）。

河口位於於東石鄉的東石大橋附近，水面寬平均約 165 公尺，因河道寬廣所以河水流速緩慢，棲地型態多為淺流。河口有砂洲淤積，河床底質以沉積泥砂為主，兩側灘地有紅樹林。下游河段的河道兩側已設置堤防，部分堤面目前已採植生綠化工法，並於堤頂設置了涼亭，供周邊民眾休憩使用，左右岸環境多為魚塭，並有零星住宅（第五河川局，2017）。

## 2.2 研究區域氣象水文資料概述

朴子溪流域位於北回歸線附近，地處亞熱帶。受降水時空分佈不均影響，河川流量變化大，夏季為豐水期，冬季為枯水期。

以嘉義站與牛稠溪橋(1)測站為例，流域全年平均降水量約為 148.86 mm，中游河道年平均流量為 9.22 CMS。從歷史記錄來看，一般每年從四月開始，降雨量逐月上升，其中以六至九月降雨量最高，直至十月降水量才又逐漸下降。河川流量的變化類似，六至八月到達最高值，十一月至次年四月份的流量較低，如表 2.2-1 所示。

同時流域也面臨颱風侵襲問題，中央氣象局（2015）依據 1897~2014 年颱風資料進行統計分析，將侵襲臺灣颱風路徑分為九類，如圖 2.2-1 所示。其中第 3 類路徑之颱風將直接侵襲本研究區域，占比約 11%。除此之外，第 2、4、7、8、9 類路徑之颱風占比約 40%，亦會影響朴子溪流域的降雨量。

表 2.2-1 研究區域歷年平均降雨量與平均流量

項目	平均降水量 (mm)		平均流量 (CMS)	
	嘉義	阿里山	牛稠溪橋(1)	灣內橋
一月	23.6	71.7	2.69	4.15
二月	57.4	137.3	1.45	4.99
三月	63.4	166.5	2.22	5.49
四月	103	254.2	3.22	7.22
五月	176.2	493.6	2.52	11.24
六月	314	649.6	27.88	23.03
七月	369.9	668.3	27.62	30.36
八月	380.2	809.3	25.37	44.54
九月	222.6	432.9	10.1	23
十月	27.5	146.7	4.12	10.16
十一月	15.2	46.3	1.81	6.38
十二月	21.3	55.9	1.61	4.96
合計	1774.3	3932.3	110.61	175.52
統計期間	1981-2010	1981-2010	2016-2019	1994-2019

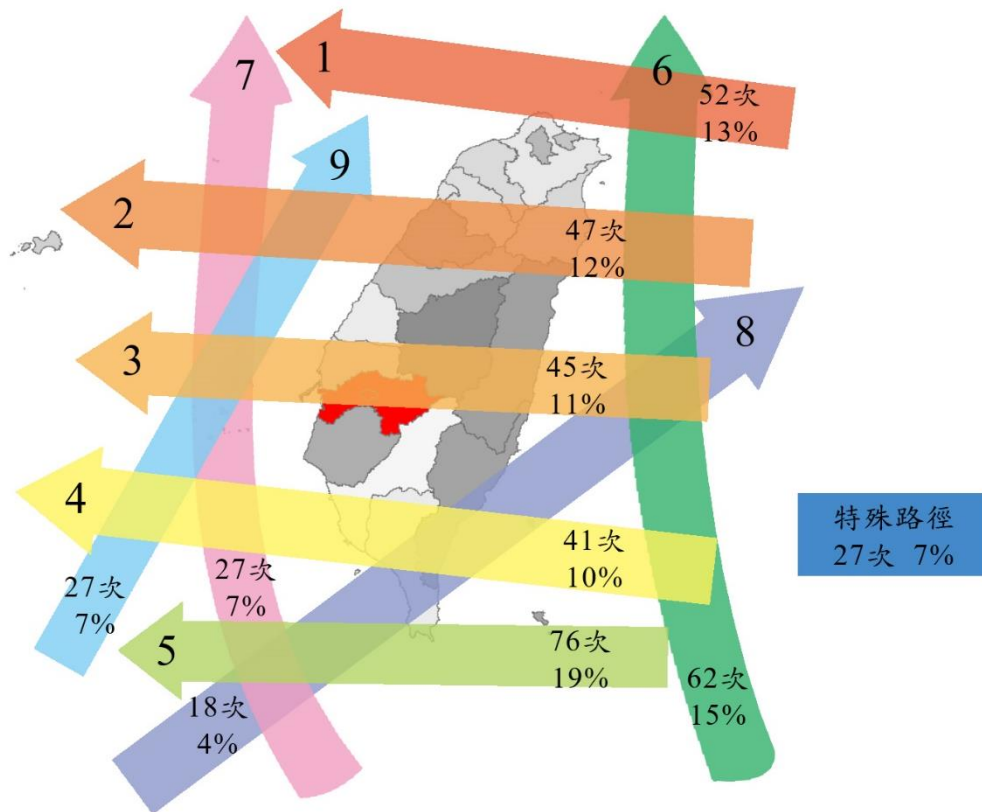


圖 2.2-1 歷年颱風侵襲路徑與概率

(資料來源：內政部國土測繪中心、中央氣象局)





## 2.3 支援向量回歸

支援向量回歸源自於支援向量機 (Support Vector Machine, SVM) 理論，支援向量機屬於一種基於機器學習 (machine learning) 模式，由 Vapnik 於 1963 年提出並不斷改進發展。Vapnik and Corinna Cortes (1995) 利用風險最小化原理 (Structured risk minimize)，提出軟間隔 (Soft margin)，此方法可以有效處理標記無關聯性的錯誤的樣本。

與眾多基於計算機模擬的模式一樣，支援向量回歸可視為一黑色系統，即不採用傳統物理、現實邏輯機制建立模式，由計算機負責求解其中的關係 (盧瑞山，2000)。近年來關於支援向量機的研究已趨於成熟，並應用於許多領域，如：生物、交通、影像辨識、文本分類、資料探勘等。

支援向量回歸屬於支援向量機的延伸應用，回歸是指使用一系列現有的值預測另一個連續的值，透過不同的不敏感損失函數 (Insensitive Loss Function)，其求得的回歸函數也會有所差別。相較於傳統線性回歸使用的最小平方法，支援向量回歸存在二次損失函數機制，僅需要少量且有用的資料就可以建立決策函數，而傳統線性回歸則需要利用所有的資料。

支援向量的求解過程引入了拉格朗日乘子 (Lagrange multipliers)  $\alpha_i$ 、 $\alpha_i^*$ ，每一組乘子分別對應一筆訓練資料，若  $\alpha_i - \alpha_i^*$  非零，則表示該筆資料可用來建立決策函數 (Decision function)，稱其為「支援向量」(support vector)。簡單來講，支援向量是指這些可被用來協助、支援、定義決策函數的資料 (C. Cortes et al., 1995)，其預測流程如圖 2.3-1 所示。

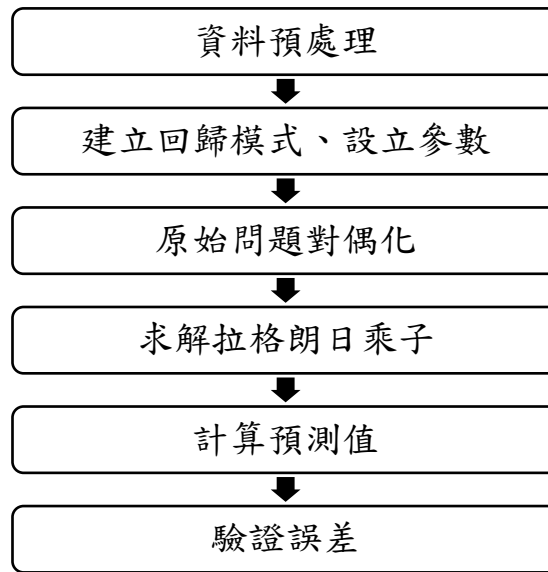


圖 2.3-1 支援向量預測流程圖

## 2.4 灰色系統理論

灰色系統理論最初是由華中理工大學鄧聚龍教授在 1979 年創立，應用於掌握不完整訊息的通訊系統控制問題。鄧聚龍教授將控制系統分為三類：第一類是參數確知的，第二類是參數完全不確知的，第三類是參數部分確知的(鄧聚龍、郭洪, 1996)。其中第三類控制系統，用盡量少的參數、資料控制不完整的系統，建立了最少訊息模型，此即是「灰色系統」(鄧聚龍, 1982)。「灰色」，相對於「黑」與「白」，是指系統之訊息不完全、不充分。

灰色系統的定義為：部份訊息已知，內涵架構亦部份已知之系統。主要探討訊息不完整、關係不明確以及操作機制不明確的系統，著重於參數及結果明確而內涵不明確問題之研究(鄧聚龍, 1999)。水文系統涉及到無數的物理過程，變數頗多，難以完全依照物理機制來描述，因此在描述一水文系統時，常將其簡化，僅留下最主要的影響因數，依其物理現象描述。此時研究者並未完全採用水文系統的物理機制、參數，屬於灰色系統。灰色預測流程可概括為圖 2.4-1，建立灰色預測模式的步驟為(吳漢雄等人, 1996)：

1. 將原始數據作累加生成處理，建立灰微分方程；



2. 求其數據矩陣和係數向量；
3. 經由灰微分方程求得預測值的累加數列；
4. 經由後項差分還原成預測數據。

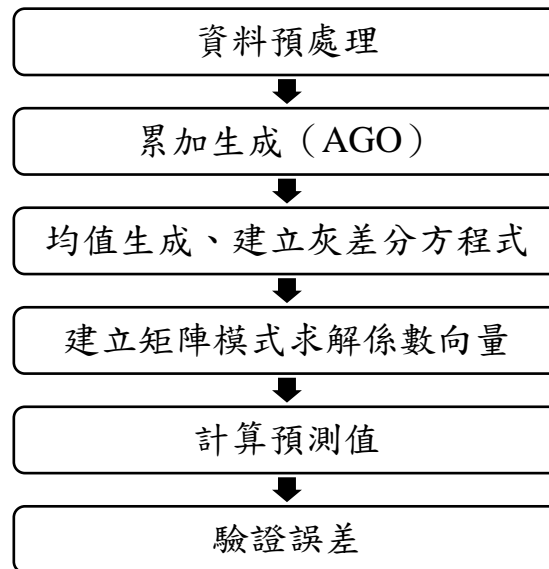


圖 2.4-1 灰色預測流程圖

以最簡單的 GM (1,1) 模式建立為例 (鄧聚龍, 1996; 樊新海等人, 2003), GM (1,1) 代表一階微分一個變數之灰模型, GM 為 Gray Model (灰模式) 縮寫, 第一個 1 表示一階微分, 第二個 1 表示一個變數。模式建立流程為:

1. 累加生成 (Accumulated Generating Operation, AGO)

假設原始序列表示為  $x^{(0)}$

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3) \dots x^{(0)}(t)\} \quad (2-1)$$

一次累加生成的序列  $x^{(1)}$  可表示為

$$x^{(1)} = AGO(x^{(0)}) = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3) \dots x^{(1)}(t)\} \quad (2-2)$$

$$x^{(1)}(t) = \sum_{k=1}^t x^{(0)}(k) \quad (2-3)$$



2. 建立灰模式，GM (1,1) 之微分方程可表示為

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (2-4)$$

其中 a、b 為模式參數（常數）

3. 轉換為灰差分方程式

$$(a) \quad \frac{dx^{(1)}(k)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{x^{(1)}(k) - x^{(1)}(k - \Delta t)}{\Delta t} \quad (2-5)$$

對於一離散序列，取  $\Delta t = 1$ ，則

$$\frac{dx^{(1)}(k)}{dt} = x^{(1)}(k) - x^{(1)}(k - 1) = x^{(0)}(k) \quad (2-6)$$

(b) 假設  $z^{(1)}(k)$  為  $x^{(1)}(k)$  之背景值（鄰近均值），取

$$z^{(1)}(k) = 0.5 \cdot (x^{(1)}(k) + x^{(1)}(k - 1)) \quad (2-7)$$

此時可將 (2-4) 式轉化為灰差分方程式

$$x^{(0)}(k) + a \cdot z^{(1)}(k) = b \quad (2-8)$$

由此可知，微分方程式與差分方程式存在對應關係：

$$\begin{aligned} \frac{dx^{(1)}(k)}{dt} &\Rightarrow x^{(0)}(k) \\ x^{(1)}(k) &\Rightarrow z^{(1)}(k) \end{aligned} \quad (2-9)$$

4. 將 (2-8) 式轉化為矩陣形式

建立資料矩陣與求解係數向量：



$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \\ \vdots \\ x^{(0)}(t) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ -z^{(1)}(4) & 1 \\ \vdots & \vdots \\ -z^{(1)}(t) & 1 \end{bmatrix}, \hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} \quad (2-10)$$

其中  $Y$  與  $B$  為數據矩陣， $\hat{a}$  為係數向量。

此時 (2-8) 式可表示為： $Y = B \cdot \hat{a}$

利用最小二乘法求解，則

$$\hat{a} = (B^T B)^{-1} B^T Y \quad (2-11)$$

參數  $a$ 、 $b$  即可求得。

#### 5. 求解預測資料（白化響應）

$a$ 、 $b$  求得後，代入 (2-4) 式中，起始條件為

$$x^{(1)}(1) = x^{(0)}(1) \quad (2-12)$$

可得其解：

$$x^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-ak} + \frac{b}{a} \quad (2-13)$$

再將  $x^{(1)}$  做後項差分還原成  $x^{(0)}$ ，可得預測數列

$$x^{(0)}(k+1) = x^{(1)}(k+1) - x^{(1)}(k) \quad (2-14)$$

解得

$$x^{(0)}(k+1) = (1 - e^a) \left(x^{(0)}(1) - \frac{b}{a}\right) e^{-ak} \quad (2-15)$$



## 2.5 線性回歸

線性回歸是最簡單也是最常見的回歸模式之一，原理為通過找出一個可以描述資料集的線性方程式，藉此預測未來的資料。每一筆資料與方程式的距離平方稱之為誤差，線性回歸的目標是將此種誤差降到最低，並以此作為線性方程權重的參考 (Triola, M. F., 2008 與 Torres, R. A., 2013)。

在本研究中，水位與雨量之關聯顯然非線性關係，因此使用簡單線性回歸僅用以模擬過度簡化的數學模式，以期觀察簡化模型是否會帶來誤差。

## 2.6 模式研究回顧

K. W. Chau et al. (1999) 對長江上游羅山測站水位進行分析，預測下游漢口站值變化。研究者使用了基於遺傳演算法 (Genetic algorithm-based) 的人工神經網路 (Artificial Neural Networks, ANNs) 和基於自我調整神經網路 (Adaptive-network-based) 的模糊理論系統 (ANFIS) 的模式，並以傳統之線性回歸模式作為對比參照基準。其結論之一為，基於機器學習演算法之模式相對於傳統數學模式，在離散度較大的資料集中具有更強之健壯性。限制因數主要為需要大量的輸入參數，和較長的計算時間。

See, L., and Openshaw, S. (2000) 使用了混合神經網路、模糊理論、時間序列模式 (Autoregressive-moving-average model) 以及天真預測 (Naïve model 直接用當前值作為預測值) 建立洪水預報模型，並將其綜合後評估模型性能。結果顯示，加入模糊理論的貝葉斯模型的總體結果較為優秀。

CH Wu et al. (2004) 利用支援向量回歸 (Support Vector Regression, SVR) 建立基於不同交通工具的行程時間預測模式，同樣也為基於灰色系統之研究。

Pao-Shan Yu et al. (2006) 在前人基礎上定量了洪水之水力停留時間，並採用支援向量回歸，建立提前 1-6 小時之洪水水位預報模式。結果顯示，支援向量回歸

模式具有與其他模式相同甚至更優秀的性能，並且對輸入變數的滯後不敏感。但作者認為支援向量回歸模式不容易理解，特徵空間之非線性映射過程較為複雜。

Han, D et al. (2007) 使用支援向量機進行洪水預報，並發現過度學習/過度擬合 (Over-fitting) 產生的負面影響大於學習不足/低度擬合 (Under-fitting)。在同一流域中，線性與非線性之核函數 (Kernel functions) 都具有優良性能。

戴凌全等人(2010)利用最小二乘法原理建立水位及流量關係之多項式方程(與灰色系統相似)，分別進行二至四次擬合，模式之平均誤差隨擬合次數增加而減少。

Hosseini,S.M.,& Mahjouri,N.(2016) 將支援向量回歸模式與基於地理環境的人工神經網路 (geomorphologic-based ANN model) 結合，建立了新型降雨逕流模式：SVR-GANN。研究者使用了人工神經網路作為支援向量回歸之核函數，提高了模式精度，較以往的多種單一模式 (ANN-BP、traditional SVR、ANN-GA、ANFIS、GANN) 之預測準確性更高。

E Sharifi Garmdareh et al. (2018) 對伊朗納馬克湖流域 55 個水文站之 1992-2012 年資料分析，預測洪峰流量。研究者採用 gamma test (GT)、GT plus 分類與專家意見選擇變數，對比支援向量回歸 (SVR)、自我調整神經模糊理論系統 (ANFIS)、人工神經網路 (ANN) 和非線性回歸 (NLR) 四種模式之性能。結果顯示，GT+ANFIS 和 GT+SVR 模型的性能最好，由此可以看出，gamma test 決定之輸入變數較適合作為水文預報。

閔英男 (2018) 使用灰色系統、BP 神經網路、RBF 神經網路、支援向量回歸等多種模式進行黑河地下水水位預測，結果表明，以上之模式都可以很好地預測水位變化趨勢及波動性，若使用混合模式，效果將比起使用單一模式還能得到進一步提升。

霍文博 (2019) 分別對中國浙江省、陝西省的 4 個流域使用新安江模型和支援向量機模型進行即時水位預報，並使用 k-NearestNeighbor 法對新安江模型結果即時校正。結果表明，新安江模型和支援向量機模型在不同流域洪水預報中各有優勢，支援向量機模型預報精度受降雨精度影響較大。

### 第三章 研究方法與過程



#### 3.1 研究區域與暴雨事件之選取

朴子河流域內共有 6 個水位測站，9 個雨量測站，周邊共有 16 個雨量測站。水位測站上下游各三個，雨量測站分為局屬測站和自動雨量站，前者的資料監測頻率更高，分別如圖 3.1-1 及表 3.1-1、3.1-2 所示。

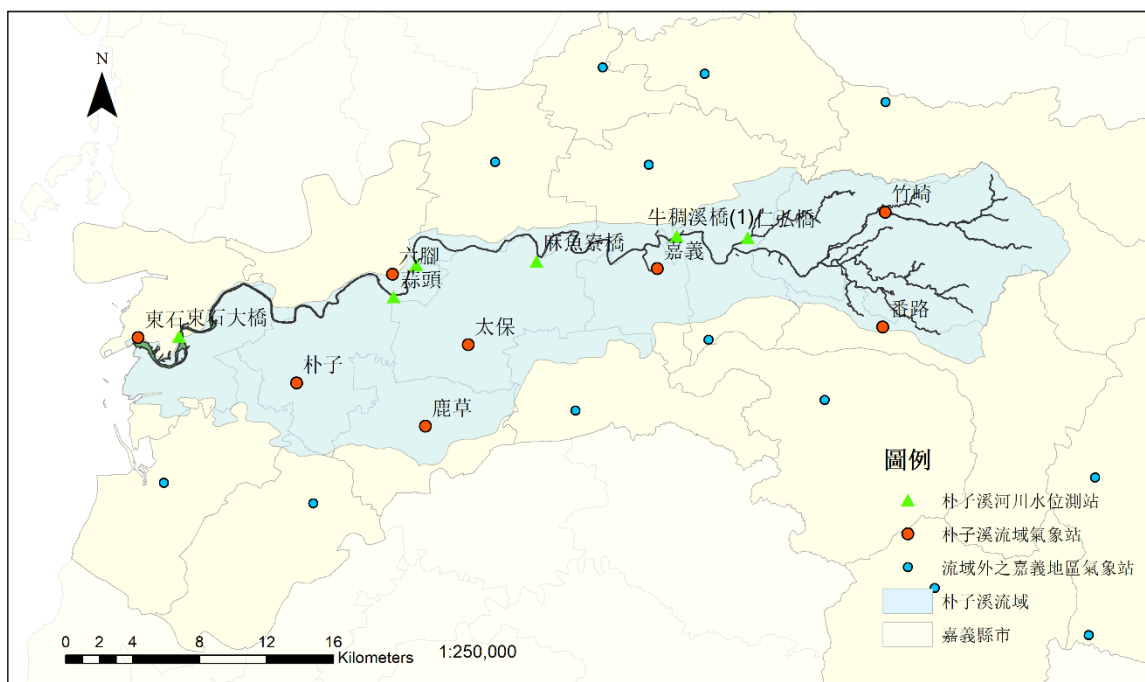


圖 3.1-1 朴子溪流域之水位及雨量測站位置圖  
(資料來源：內政部國土測繪中心、經濟部水利署)

表 3.1-1 流域內之水位測站

測站名稱	測站代號	經度	緯度	所屬縣市	所屬鄉鎮
牛稠溪橋(1)	1550H017	120.4432831	23.5135498	嘉義市	東區
東石大橋	1550H018	120.175829	23.45945	嘉義縣	東石鄉
麻魚寮橋	1550H016	120.367981	23.49979019	嘉義縣	太保市
蒜頭	1550H011	120.2912292	23.48080444	嘉義縣	六腳鄉
灣內橋	1550H012	120.3033218	23.49786949	嘉義縣	六腳鄉
仁弘橋	1550H014	120.4815826	23.51259804	嘉義縣	民雄鄉



表 3.1-2 流域及周邊之雨量測站

測站名稱	測站代號	種類	經度	緯度	所屬縣市	所屬鄉鎮
嘉義市東區	C0M730	自動	23.4577	120.4605	嘉義市	東區
朴子	C0M650	自動	23.4346	120.239	嘉義縣	朴子市
太保	C0M680	自動	23.455	120.3314	嘉義縣	太保市
大林	C0M670	自動	23.6006	120.4584	嘉義縣	大林鎮
布袋	C0M750	自動	23.3809	120.1678	嘉義縣	布袋鎮
馬頭山	C0M410	自動	23.3244	120.582	嘉義縣	大埔鄉
表湖	C0M850	自動	23.2419	120.6417	嘉義縣	大埔鄉
中埔	C0M640	自動	23.4254	120.5229	嘉義縣	中埔鄉
六腳	C0M740	自動	23.4929	120.2906	嘉義縣	六腳鄉
民雄	C0M760	自動	23.5517	120.4284	嘉義縣	民雄鄉
奮起湖	C0M530	自動	23.4939	120.6993	嘉義縣	竹崎鄉
竹崎	C0M700	自動	23.5262	120.5555	嘉義縣	竹崎鄉
東石	C0M710	自動	23.4589	120.1538	嘉義縣	東石鄉
茶山	C0M800	自動	23.2991	120.6648	嘉義縣	阿里山鄉
裡佳	C0M810	自動	23.3891	120.7166	嘉義縣	阿里山鄉
山美	C0M830	自動	23.3838	120.6682	嘉義縣	阿里山鄉
達邦	C0M820	自動	23.454	120.7495	嘉義縣	阿里山鄉
水上	C0M690	自動	23.4197	120.3889	嘉義縣	水上鄉
嘉義梅山	C0M770	自動	23.5854	120.5556	嘉義縣	梅山鄉
鹿草	C0M780	自動	23.4113	120.3083	嘉義縣	鹿草鄉
番路	C0M720	自動	23.4646	120.5541	嘉義縣	番路鄉
新港	C0M790	自動	23.5533	120.3457	嘉義縣	新港鄉
溪口	C0M660	自動	23.6041	120.4036	嘉義縣	溪口鄉
東後寮	C0M520	自動	23.3699	120.248	嘉義縣	義竹鄉
嘉義	467480	局屬	23.4959	120.4329	嘉義市	西區
阿里山	467530	局屬	23.5082	120.8132	嘉義縣	阿里山鄉

本研究所選取的水位測站是「牛稠溪橋(1)」測站，位於嘉義市的北側邊界處，較容易體現朴子溪水位變化對市區的影響。為了盡可能測試各模式之健壯性，雨量站僅選取了「番路 C0M700」、「竹崎 C0M720」兩個距離較近且處於河道上游的測站，以便分析可能存在的缺陷。水位測站及雨量測站位置、監測頻率等如圖 3.1-2 與表 3.1-3 所示，所使用的資料為 2017 年全年資料。

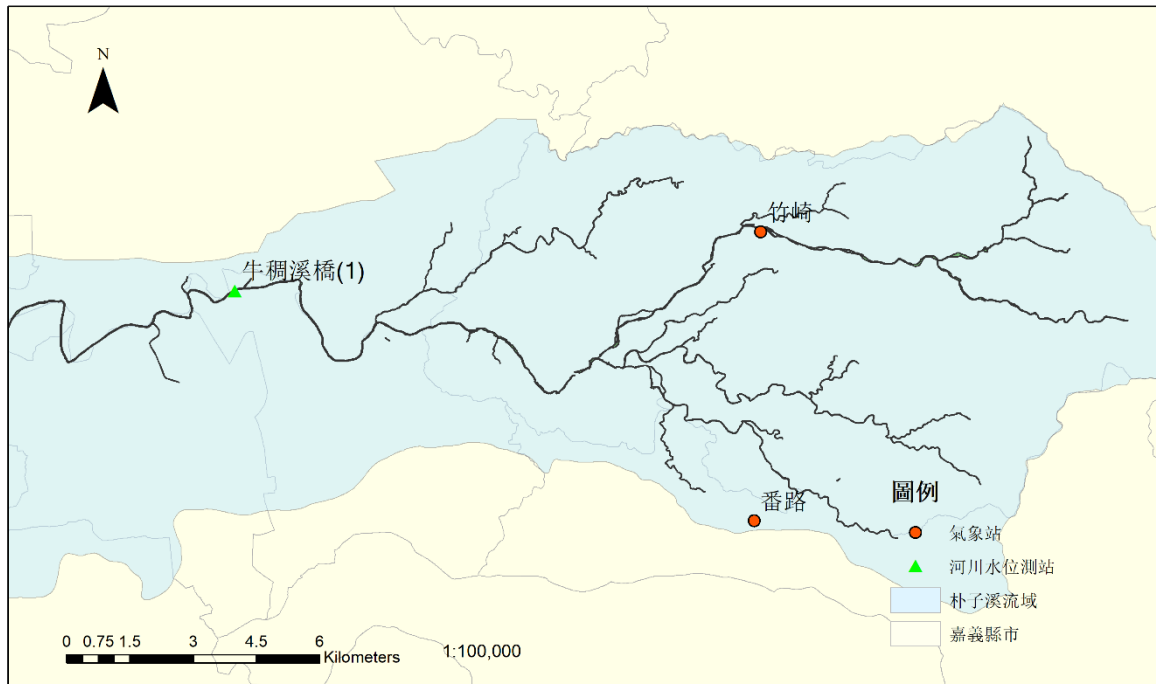


圖 3.1-2 研究使用之水位及雨量測站位置圖  
(資料來源：內政部國土測繪中心、經濟部水利署)

表 3.1-3 研究使用之水位及雨量測站資料

測站種類	名稱	編號	經度	緯度	資料種類	監測頻率	所屬單位
水位測站	牛稠溪橋(1)	1550H 017	120. 4433	23.5 135	水位、 流量	每 10 分鐘一筆	第五河川局
自動雨量站	番路	COM7 00	120. 5541	23.4 646	雨量	每小時一筆 (1998-2017) 或 即時記錄	中央氣象局
自動雨量站	竹崎	COM7 20	120. 5555	23.5 262	雨量	(2018-)	中央氣象局

其中河川水位測站之資料為每 10 分鐘一筆，本研究因受自動雨量站之資料頻率限制，採用平均法計算每小時之水位供模式使用。

自動雨量站一般之資料間隔為每小時一次，並且從 2018 年往後的資料中，雨量站在降雨瞬間會開始記錄數值，時間間隔不固定，本研究將間隔不固定之降雨量累加，得到每小時降雨量。

局屬雨量站之資料間隔為每分鐘一次，其監測頻率非常高，具有更強之參考意義，但由於本研究選用之測站未包括局屬雨量站，故僅將預測精度精確到小時，在後續研究中可以考慮藉此提高模式精確度。

用以驗證預測之降雨則選用了三次 2018 年之降雨事件，事件一為 823 暴雨，曾引發嘉義地區接近一個月的淹水問題（陳昆懋，2019），後兩次則是隨機選取的一般降雨，如表 3.1-4 所示：

表 3.1-4 研究使用之預測驗證事件

事件	事件說明	資料時間（年-月-日-小時）
事件一	823 暴雨	2018-08-22-00:00 ~ 2018-09-03-23:00
事件二	一般降雨	2018-07-01-00:00 ~ 2018-07-14-23:00
事件三	一般降雨	2018-06-10-00:00 ~ 2018-06-22-23:00

### 3.2 支援向量回歸水位預測模式

假設有一組資料序列  $\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\}$ ，線性支援向量回歸之函數如 (3-1) 式所示：

$$f(x) = (\omega \cdot x) + b \quad (3-1)$$

其中  $x_i$ ：輸入向量

$y_i$ ：實際輸出向量

$f(x)$ ：回歸函數之輸出向量

$\omega$ ： $f(x)$  之複雜度 (complexity)

$b$ ：常數

$\omega$  之數值愈小代表模式愈不複雜， $b$  為常數，通常最後才決定。



依據結構風險最小化法則（陳磊等人，2006；張逸凡，2005），比照前人之研究（張逸凡，2005）可以將線性支援向量回歸問題表示為（3-2）式：

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{Subject to} \quad & y_i - (\omega \cdot x_i + b) \leq \varepsilon + \xi_i \\ & (\omega \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{aligned} \quad (3-2)$$

其中  $\frac{1}{2} \|\omega\|^2$  代表模式複雜度， $\varepsilon$  為容許寬度（ $\varepsilon$ -tube）， $C \sum_{i=1}^l (\xi_i + \xi_i^*)$  代表控制經驗風險（Empirical Risk）， $\xi_i$  與  $\xi_i^*$  為寬鬆變數（Slack Variable），分別代表位於容忍區間上方及下方出現誤差的訓練資料。

若  $C \sum_{i=1}^l (\xi_i + \xi_i^*)$  較大，則表示模式需要對資料做大的調整，因此計算時會將寬鬆變數之和乘上為懲罰參數（Penalty parameter） $C$ ， $C$  值越大代表誤差發生時對目標函數影響越大，不敏感損失函數為：

$$|y - f(x)|_\varepsilon = \begin{cases} 0, & \text{if } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases} \quad (3-3)$$

上式中  $y$  為實際值， $f(x)$  為估計值。我們定義一個可容許的誤差容忍區間（ $\varepsilon$ -tube），若實際值落於此區間內則損失函數（Loss function）值為零；否則損失函數值則不為零，計算懲罰，此時點落於誤差容忍區間外，由此損失函數即可定義出實際值與估計值之誤差，如圖 3.2-1 與圖 3.2-2 所示。

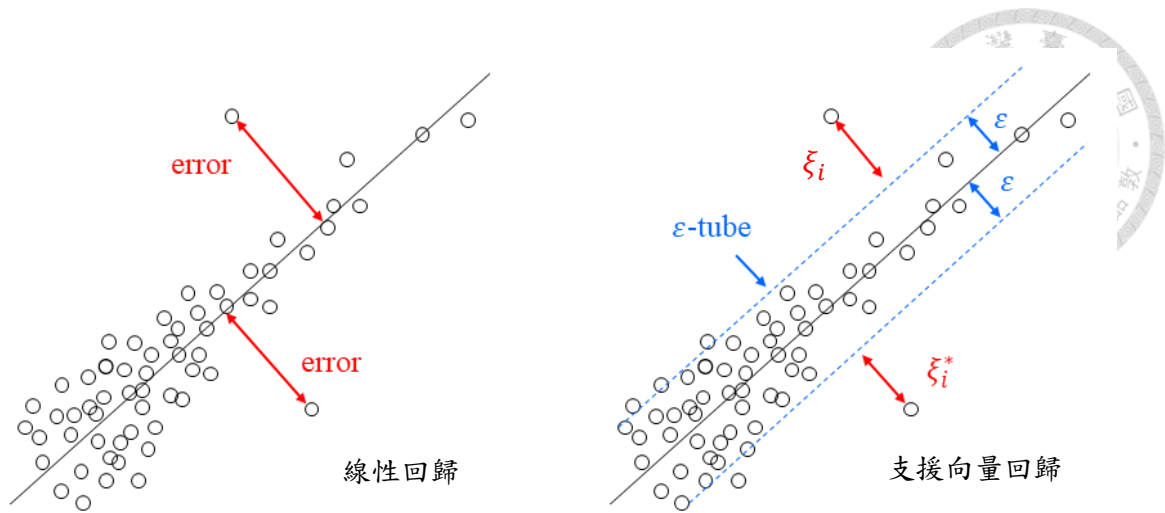


圖 3.2-1 線性回歸與支援向量回歸之比較 (改繪自 Vapnik, 1998)

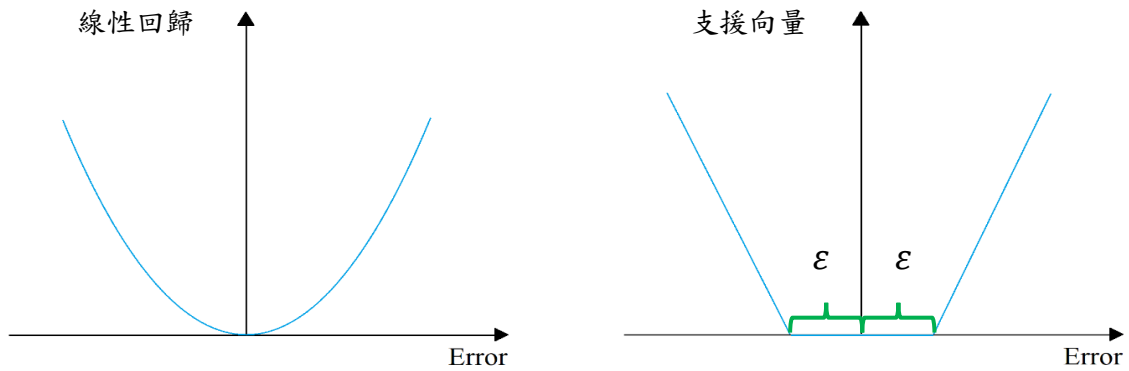


圖 3.2-2 線性回歸與支援向量回歸之損失函數比較 (改繪自 Vapnik, 1998)

引入拉格朗日乘子 (Lagrange multipliers)  $\alpha, \alpha^*, \eta, \eta^*$ ，得到下面的拉格朗日函數，求解：

$$\begin{aligned}
 L(\omega, b, \xi, \xi^*; \alpha, \alpha^*, \eta, \eta^*) = & \frac{1}{2} \|\omega\|^2 + C \left( \sum_{i=1}^l (\xi_i + \xi_i^*) \right) \\
 & - \sum_{i=1}^l \alpha_i [\varepsilon + \xi_i - y_i + (\omega \cdot x_i) + b] \\
 & - \sum_{i=1}^l \alpha_i^* [\varepsilon + \xi_i^* + y_i - (\omega \cdot x_i) - b] - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*)
 \end{aligned} \tag{3-4}$$

對上式分別求偏導，即  $\max_{\alpha, \alpha^*, \eta, \eta^*} \min_{\omega, b, \xi, \xi^*} L(\omega, b, \xi, \xi^*; \alpha, \alpha^*, \eta, \eta^*)$  可得：

$$\begin{aligned}
\frac{\partial L}{\partial \omega} = 0 &\rightarrow \omega - \sum_{i=1}^l (\alpha_i^* - \alpha_i) x_i = 0 \\
\frac{\partial L}{\partial b} = 0 &\rightarrow \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \\
\frac{\partial L}{\partial \xi_i} = 0 &\rightarrow C - \alpha_i - \eta_i = 0 \\
\frac{\partial L}{\partial \xi_i^*} = 0 &\rightarrow C - \alpha_i^* - \eta_i^* = 0
\end{aligned}
\tag{3-5}$$



$$\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0, i = 1, \dots, l$$

將上式代入 (3-4)，可將  $\omega$ 、 $b$ 、 $\xi$  消去，只留下  $\alpha$ 、 $\eta$ ，得到 SVR 對偶問題 (dual problem)：

$$\begin{aligned}
&\text{Maximize}_{\alpha, \eta} \\
&-\frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i \cdot x_j) - \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i)
\end{aligned}
\tag{3-6}$$

Subject to

$$\begin{aligned}
\sum_{i=1}^l (\alpha_i^* - \alpha_i) &= 0 \\
0 &\leq \alpha_i \leq C \\
0 &\leq \alpha_i^* \leq C \\
i &= 1, \dots, l
\end{aligned}$$

上式即為二次規劃 (quadratic programming) 問題，求解上式即可得到  $\alpha_i$ 、 $\alpha_i^*$ ，再將非零的  $\alpha_i$ 、 $\alpha_i^*$  代入  $\omega = \sum_{i=1}^l (\alpha_i^* - \alpha_i) x_i$  可得  $\omega$ 、 $b$  值由 Karush-Kuhn-Tucker (KKT) conditions (Fletcher, 1987) 求得，分別如 (3-7)、(3-8) 式所示：

$$\omega = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \cdot x_i \tag{3-7}$$

$$\alpha_i (\omega \cdot x_i + b - y_i + \varepsilon + \xi_i) = 0 \tag{3-8}$$

綜上所述，線性支援向量回歸模式可表示為 (3-9) 式， $\alpha_i^* - \alpha_i \neq 0$  所對應的資料即為支援向量：

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) (x \cdot x_i) + b \quad (3-9)$$

對於非線性支援向量回歸模式，則是在 (3-1) 式引入非線性映射函數 (nonlinear mapping function)  $\phi(x)$ ，決策函數變為：

$$f(x) = \omega \cdot \phi(x) + b \quad (3-10)$$

所以 (3-6) 式將變為下式，其限制條件不變：

$$\begin{aligned} & \underset{\alpha, \eta}{\text{Maximize}} \\ & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) [\phi(x_i) \cdot \phi(x_j)] - \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \\ & \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \end{aligned} \quad (3-11)$$

求解式中  $\phi(x_i)$ 、 $\phi(x_j)$  需要引入核函數 (Kernel Function)  $K$ ，函數滿足  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ ，本研究參考前人之研究，採用多項式 (Polynomial) 之核函數。非線性支援向量回歸模式可表示為：

$$f(x) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x, x_i) + b \quad (3-12)$$

$$K(x_i, x_j) = [\gamma(x_i, x_j) + c]^d \quad (3-13)$$

其中  $\gamma$ 、 $c$ 、 $d$  為核函數之參數。

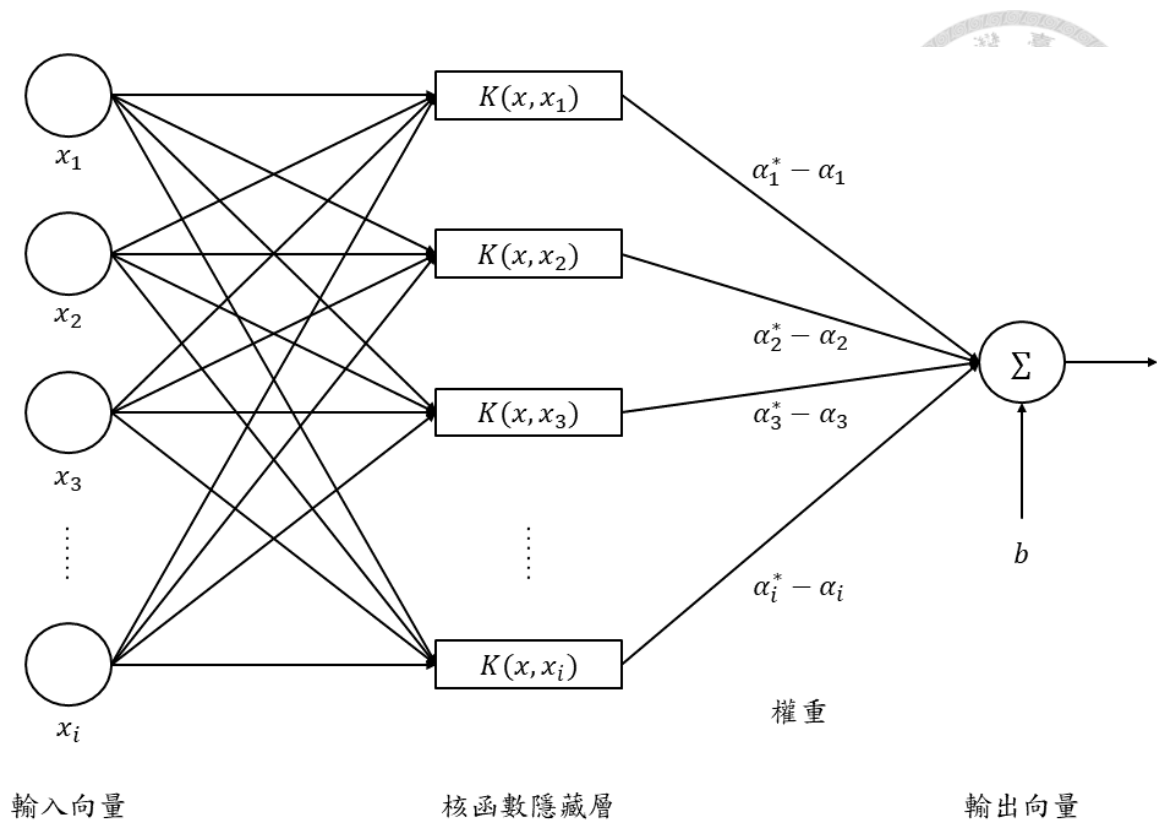


圖 3.2-3 非線性支援向量回歸架構示意圖 (改繪自 Vapnik, 1998)

### 3.3 灰色水位預測模式

降雨——逕流過程具有因果關係，符合灰色系統理論，故可將其視為一組具有不確定特性的灰色過程。利用雨量與流量之累加生成序列，建立降雨——逕流預報模式，可應用於集水區之即時逕流預報 (張簡鳳蓮，2002)。

微分水文灰色模式 (Differential Hydrological Grey Model) (夏軍, 1989; 莊文南, 1993) 採用時間序列變換方式建立，弱化原始數列之隨機性，簡稱 DHGM (n+1, r+1) 模式。由於雨量與流量為非線性的對應關係，所以儘管原始序列有漲落起伏變化，但經累加生成處理後，變換序列則呈有規律增長趨勢，可能弱化為近似線性關係，易於分析與模擬 (莊文南, 1993)。本研究參考包玉海 (1995)，僅考慮一次累加生成之 DHGM (2,2) 模式：



$$\frac{dQ^{(1)}(t)}{dt} + a_1 Q^{(1)}(t) = b_0 \frac{dR^{(1)}(t-1)}{dt} + b_1 R^{(1)}(t-1) \quad (3-14)$$

其中  $Q^{(1)}(t)$ ：流量之一次累加生成，即累積流量（cms）

$R^{(1)}(t)$ ：雨量之一次累加生成，即累積雨量（mm）

$a_1$ 、 $b_0$ 、 $b_1$ ：模式參數

(3-14) 式為微分方程，依照前述 GM (1,1) 模式之原理 ((2-4) 式至 (2-9) 式)，將其轉化成離散型之差分式：

$$Q^{(0)}(t) + a_1 z^{(1)}(t) = b_0 R^{(0)}(t-1) + b_1 R^{(1)}(t-1) \quad (3-15)$$

$$z^{(1)}(t) = 0.5 \cdot [Q^{(1)}(t) + Q^{(1)}(t-1)] \quad (3-16)$$

(3-16) 式為  $Q^{(1)}(t)$  的白化值， $Q^{(0)}(t)$  為原始流量，是  $Q^{(1)}(t)$  與  $Q^{(1)}(t-1)$  的差值， $R^{(0)}(t-1)$  同理。將其代入推導化簡 (3-15) 式：

$$Q^{(0)}(t) + \frac{a_1}{2} [Q^{(1)}(t) + Q^{(1)}(t-1)] = b_0 R^{(0)}(t-1) + b_1 R^{(1)}(t-1) \quad (3-17)$$

$$\begin{aligned} Q^{(1)}(t) - Q^{(1)}(t-1) + \frac{a_1}{2} Q^{(1)}(t) + \frac{a_1}{2} Q^{(1)}(t-1) \\ = b_0 [R^{(1)}(t-1) - R^{(1)}(t-2)] + b_1 R^{(1)}(t-1) \end{aligned} \quad (3-18)$$

$$\begin{aligned} \left(1 + \frac{a_1}{2}\right) Q^{(1)}(t) + \left(-1 + \frac{a_1}{2}\right) Q^{(1)}(t-1) \\ = (b_0 + b_1) R^{(1)}(t-1) - b_0 R^{(1)}(t-2) \end{aligned} \quad (3-19)$$



$$\begin{aligned}
(2 + a_1)Q^{(1)}(t) &= (2 - a_1)Q^{(1)}(t - 1) + (2b_0 + 2b_1)R^{(1)}(t - 1) \\
&\quad - 2b_0R^{(1)}(t - 2)
\end{aligned}
\tag{3-20}$$

$$Q^{(1)}(t) = \frac{2 - a_1}{2 + a_1}Q^{(1)}(t - 1) + \frac{2b_0 + 2b_1}{2 + a_1}R^{(1)}(t - 1) + \frac{-2b_0}{2 + a_1}R^{(1)}(t - 2)
\tag{3-21}$$

改寫為：

$$Q^{(1)}(t) = C_1 \cdot Q^{(1)}(t - 1) + C_2 \cdot R^{(1)}(t - 1) + C_3 \cdot R^{(1)}(t - 2)
\tag{3-22}$$

其中  $C_1 = \frac{2-a_1}{2+a_1}$ ,  $C_2 = \frac{2b_0+2b_1}{2+a_1}$ ,  $C_3 = \frac{-2b_0}{2+a_1}$ , 為模式參數。

由於流量與雨量之累積值相差過大，因此需先將雨量與流量做無因次化處理後再去率定參數。

$$\tilde{Q}^{(1)}(t) = C_1 \cdot \tilde{Q}^{(1)}(t - 1) + C_2 \cdot \tilde{R}^{(1)}(t - 1) + C_3 \cdot \tilde{R}^{(1)}(t - 2)
\tag{3-23}$$

其中  $\tilde{Q}^{(1)}(t) = \frac{Q^{(1)}(t)}{\bar{Q}}$ ：無因次化後之累積流量

$\tilde{R}^{(1)}(t) = \frac{R^{(1)}(t)}{\bar{R}}$ ：無因次化後之累積雨量

$\bar{Q}$ ：平均流量 (cms)

$\bar{R}$ ：平均雨量 (mm)

$C_1$ 、 $C_2$ 、 $C_3$ ：模式參數

張逸凡 (2005) 假定河流集水區為一個整體系統，此時水位變化可視為流量變化之另一形式輸出，建立了類似上述灰色降雨-逕流模式的降雨-水位高度灰色模式。本研究效仿過往研究 (莊文南, 1993; 陳憲宗, 1998) 之無因次化處理方法，採取均值化處理，即是將 2017 年一整年之雨量與水位高度值各除以年度平均雨量與平均水位，與支援向量回歸所使用之資料相同。本研究採用此種灰色水位預報模式，並與支援向量回歸、線性回歸之水位預報模式做對比。

令  $S_L$  為牛稠溪橋水位高度； $R_L$  為朴子溪流域之雨量，由流域內所選用的雨量站資料經徐昇網法 (Thiesson network method) 求得。上述變量之「水位」即為「水位漲差」。因此 (3-23) 式可以改寫為：

$$\tilde{S}_L^{(1)}(t) = C_1 \cdot \tilde{S}_L^{(1)}(t-1) + C_2 \cdot \tilde{R}_L^{(1)}(t-1) + C_3 \cdot \tilde{R}_L^{(1)}(t-2) \quad (3-24)$$

其中  $\tilde{S}_L^{(1)}(t) = \frac{S_L^{(1)}(t)}{\bar{S}_L}$ ：無因次化後之累積水位高度

$\tilde{R}^{(1)}(t) = \frac{R^{(1)}(t)}{\bar{R}}$ ：無因次化後之累積雨量

$\bar{S}_L$ ：平均水位高度 (m)

$\bar{R}$ ：平均雨量 (mm)

對 (3-24) 式之無因次化後的累積雨量與累積水位高度值，以最小二乘法求解灰色降雨——水位高度預報模式之參數  $C_1$ 、 $C_2$  與  $C_3$ ：

$$Y = \begin{bmatrix} \tilde{S}_L^{(1)}(3) \\ \tilde{S}_L^{(1)}(4) \\ \tilde{S}_L^{(1)}(5) \\ \vdots \\ \tilde{S}_L^{(1)}(t) \end{bmatrix}, B = \begin{bmatrix} \tilde{S}_L^{(1)}(2) & \tilde{R}^{(1)}(2) & \tilde{R}^{(1)}(1) \\ \tilde{S}_L^{(1)}(3) & \tilde{R}^{(1)}(3) & \tilde{R}^{(1)}(2) \\ \tilde{S}_L^{(1)}(4) & \tilde{R}^{(1)}(4) & \tilde{R}^{(1)}(3) \\ \vdots & \vdots & \vdots \\ \tilde{S}_L^{(1)}(t-1) & \tilde{R}^{(1)}(t-1) & \tilde{R}^{(1)}(t-2) \end{bmatrix}, \hat{a} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} \quad (3-25)$$



$$\hat{a} = (B^T B)^{-1} B^T Y \quad (3-26)$$

模式率定結果為：

$$C_1 = 0.9996$$

$$C_2 = 0.003462$$

$$C_3 = -0.003254$$

$$\bar{S}_L = 15.83m$$

$$\bar{R} = 0.2704mm$$

此時代入上述率定結果參數，即可建立牛稠溪橋水位站  $t$  時刻的灰色無因次累積水位方程式為：

$$\tilde{S}_L^{(1)}(t) = 0.9996 \cdot \tilde{S}_L^{(1)}(t-1) + 0.003462 \cdot \tilde{R}_L^{(1)}(t-1) - 0.003254 \cdot \tilde{R}_L^{(1)}(t-2) \quad (3-27)$$

牛稠溪橋水位站  $t$  時刻的灰色水位方程式為：

$$S_L^{(1)}(t) = \tilde{S}_L^{(1)}(t) \times \bar{S}_L \quad (3-28)$$

$$S_L^{(0)}(t) = S_L^{(1)}(t) - S_L^{(1)}(t-1) \quad (3-29)$$

超前  $l$  時刻的灰色水位預測方程式為：

$$S_L^{(0)}(t+l) = S_L^{(1)}(t+l) - S_L^{(1)}(t-1+l) \quad (3-30)$$

### 3.4 線性回歸預測模式

本研究之資料序列為  $\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\}$ ，簡單線性回歸方程可寫為：

$$f(x) = a + bx \quad (3-31)$$

$$a = \bar{y} - b\bar{x} \quad (3-32)$$

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \quad (3-33)$$



其中  $x$ ：雨量觀測值

$\bar{x}$ ：平均雨量 (mm)

$y$ ：水位高度觀測值 (m)

$\bar{y}$ ：平均水位 (m)

$f(x)$ ：水位高度預測值 (m)

$a, b$ ：線性回歸方程常數

$n$ ：樣本數

### 3.5 水位傳播時間

洪水之傳播時間為洪水事件之重要參數之一，這是由於河川之上、下游水位變化存在遲滯現象。假設有一河川水位測站位於下游，當降雨空間分佈偏上游，則傳播時間較長，偏下游則傳播時間較短（王如意、易任，1979）。

因為洪水預測中需要儘快得出結論，同時可適當捨棄精確度。故本研究嘗試忽略多峰形之水位曆線，只將尖峰水位（主峰段）抵達時間作為參考依據。事實上多峰形之水位曆線複雜多變，在最高水位抵達之前可能已經有較低峰值經過，實務工程中需進行進一步之定量分析。

尖峰水位之抵達時間則經由降雨、水位變化之相關性分析得到，如表 3.5-1 所示。



表 3.5-1 水位傳播時間相關性分析

參數代號	測站	延滯時間	相關性
700rf	竹崎	0	0.227769
700rf1	竹崎	1	0.27871
700rf2	竹崎	2	0.366778
700rf3	竹崎	3	0.390836
700rf4	竹崎	4	0.363266
700rf5	竹崎	5	0.330714
700rf6	竹崎	6	0.30221
700rf7	竹崎	7	0.280705
700rf8	竹崎	8	0.266516
700rf9	竹崎	9	0.255547
700rf10	竹崎	10	0.242862
720rf	番路	0	0.209793
720rf1	番路	1	0.260448
720rf2	番路	2	0.344293
720rf3	番路	3	0.364185
720rf4	番路	4	0.337326
720rf5	番路	5	0.306748
720rf6	番路	6	0.281169
720rf7	番路	7	0.261228
720rf8	番路	8	0.249325
720rf9	番路	9	0.239534
720rf10	番路	10	0.228113

本研究列舉降雨開始後 0-10 小時之降雨及水位相關性。對於「番路 COM700」測站，三小時延滯時間之相關性最高，其次為二、四；對於「竹崎 COM720」測站，其結果與前者相仿，但相關性較前者低。為了簡化模式架構，本研究選用之參數為：700rf2、700rf3、700rf4、720rf2、720rf3、720rf4，將兩個雨量測站的資料，分別計算 2、3、4 小時之延滯期，即使用過去 2、3、4 小時之資料，輸入支援向量回歸與線性回歸模式。對於灰色水位模式，本研究僅選擇 3 小時作為兩個雨量站的水位延遲時間。



### 3.6 模式評鑒指標

本研究採用了四個評鑒指標，分別為：決定係數、均方根誤差、尖峰水位時間誤差、絕對最大誤差。

決定係數 (Coefficient of determination,  $R^2$ ) 是一個評價兩個變數線性相關度的指標，為回歸平方和與總平方和之比 (Glantz et al., 1990)。對於簡單的線性回歸而言，決定係數為樣本相關係數 (Coefficient of Correlation) 的平方。 $R^2$  取值在 0 到 1 之間，且無單位， $R^2$  越高回歸方程式越優。

$$R^2 = 1 - \frac{SSE}{SST} \quad (3-34)$$

$$SSE = \sum_{i=1}^n (y^{(i)} - \bar{y})^2 \quad (3-35)$$

$$SST = \sum_{i=1}^n (y^{(i)} - \bar{y})^2 \quad (3-36)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y^{(i)} \quad (3-37)$$

其中 SSE：殘差平方和，Sum of Squares for Error

SST：總平方和，Sum of Squares for total

$y^{(i)}$ ：觀測值

$\bar{y}$ ：平均觀測值

$\mu_y$ ：觀測值對應之預測值

均方根誤差 (Root-mean-square error, RMSE) 是預測值與真實值偏差的平方與觀測次數比值的平方根，即預測值與真實值之差的樣本標準差，可視為預測誤差 (Hyndman et al., 2006)。因其單位與資料集相同，所以在資料集相同的情況下，

不僅可反映不同模式中的誤差高低，更可用於定量分析。均方根誤差越接近 0，代表誤差越小。



$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^i (y^{(i)} - \hat{y}^{(i)})^2}{n}}$$

(3-38)

其中  $y^{(i)}$ ：真實值

$\hat{y}^{(i)}$ ：對應之預測值

$n$ ：樣本數

絕對最大誤差 (Maximum absolute error, MAE) 代表絕對誤差之範圍，又稱絕對誤差界 (張逸凡，2005)。

$$\text{MAE} = \max \{ |y^{(i)} - \hat{y}^{(i)}| \}$$

(3-39)

尖峰水位誤差 (Error of peak water level, EPWL) 為實際尖峰流量到達時刻，預測水位與真實值之差值。差值為正，則預測值過高，反之則為預測值過低。在洪水預測中，尖峰水位誤差越接近 0 越好，同時負值的誤差相較於正值更不能被接受。

$$\text{EPWL} = \hat{y}_{Peak} - y_{Peak}$$

(3-40)

其中  $y_{Peak}$ ：真實洪峰高度

$\hat{y}_{Peak}$ ：對應之預測洪峰高度



## 第四章 結果與討論



研究首先需要將訓練資料格式化處理，並將其中缺失之數值予以插值，而後分別對資料進行支援向量回歸、灰色水位、線性回歸模式訓練，並預測兩場降雨之水位高度、評估模式可靠性。

### 4.1 模式運行環境

研究使用了基於 Python 3.x (Python Software Foundation, 2001) 的機器學習套件 SKlearn (David Cournapeau, 2007)，匯入 LinearRegression 以及 SVM 模組進行實作。

因研究訓練集較少(52560 筆資料)，使用純 CPU 運算之速度已經十分可觀。在本人的 AMD FX-9590 上(稍弱於現代的 Intel i5-9400)，除去資料前處理工作、洪峰延滯期決定、繪圖所需之時間，使用 SVM 模組學習並預測的過程僅需 1~2 分鐘即可完成，非常符合快速預測之標準。

在資料量不大之情況下，任何一臺較新的個人電腦都可以勝任預測工作，相較於 TensorFlow 等基於 GPU 加速的資料流(Dataflow programming)學習架構(段生月等人, 2019、Nevo et.al., 2019)，對硬體要求更低。但如果資料集過於龐大，學習所需的算力將會以指數形式增長，此時則需使用高性能電腦乃至工作站才可以勝任。

### 4.2 模式率定資料概述

本研究選取了「牛稠溪橋(1)」水位測站作為水位參考，測站每 10 分鐘記錄一筆資料，因其資料記錄間隔與雨量站記錄間隔(每小時一筆)不同，故需要經過前處理，以每小時為單位進行平均，最後得到 8760 筆資料，如圖 4.2-1 所示。

研究選用之雨量站為「番路 COM700」、「竹崎 COM720」，位於河道上游集水區。資料間隔在 1998-2017 年間為每小時一次，共 8760 筆，分別如圖 4.2-2 及圖 4.2-3 所示。率定使用的雨量為兩個測站之平均雨量，資料整理後可得到圖 4.2-4。值得注意的是，中央氣象局提供之雨量資料從 2018 年 1 月 1 日 0 時開始變更為即時記錄，間隔不固定，故需要將即時之降雨量資料進行累加，得到以小時為單位的每小時降雨量。

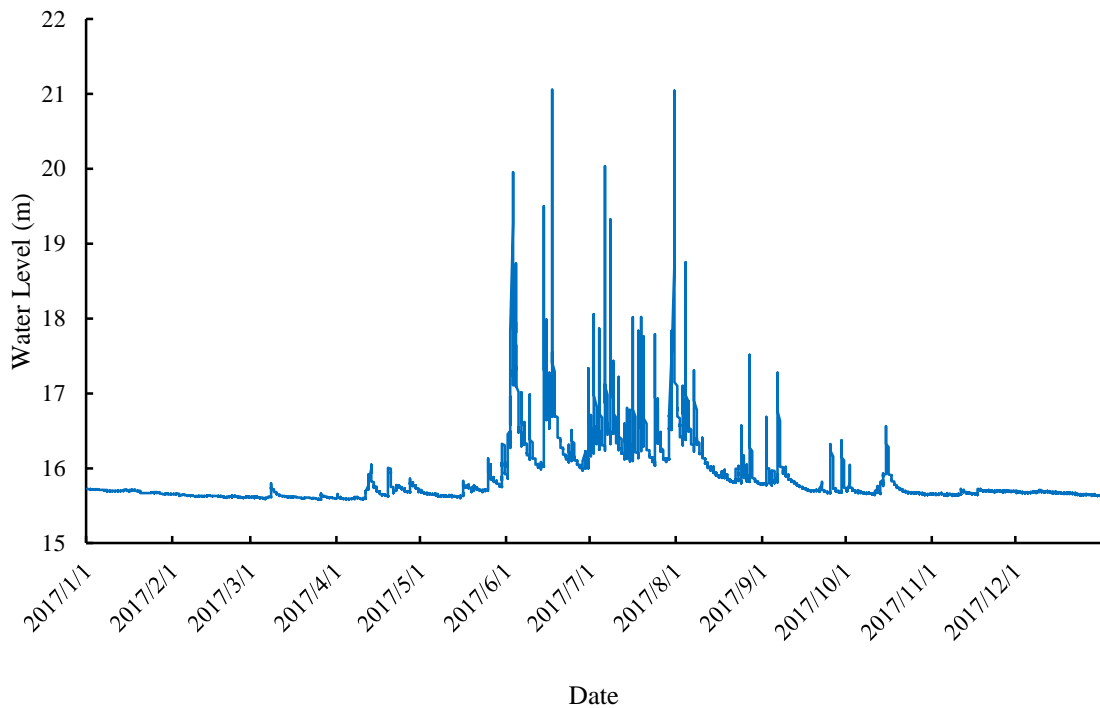


圖 4.2-2 牛稠溪橋(1)測站 2017 全年之水位高度

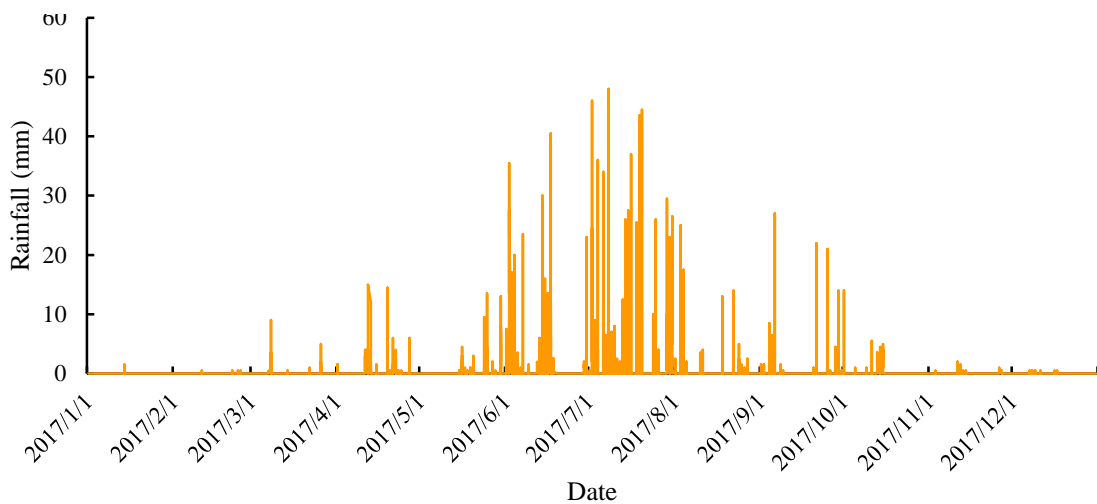


圖 4.2-1 番路雨量測站 2017 全年之日降雨量

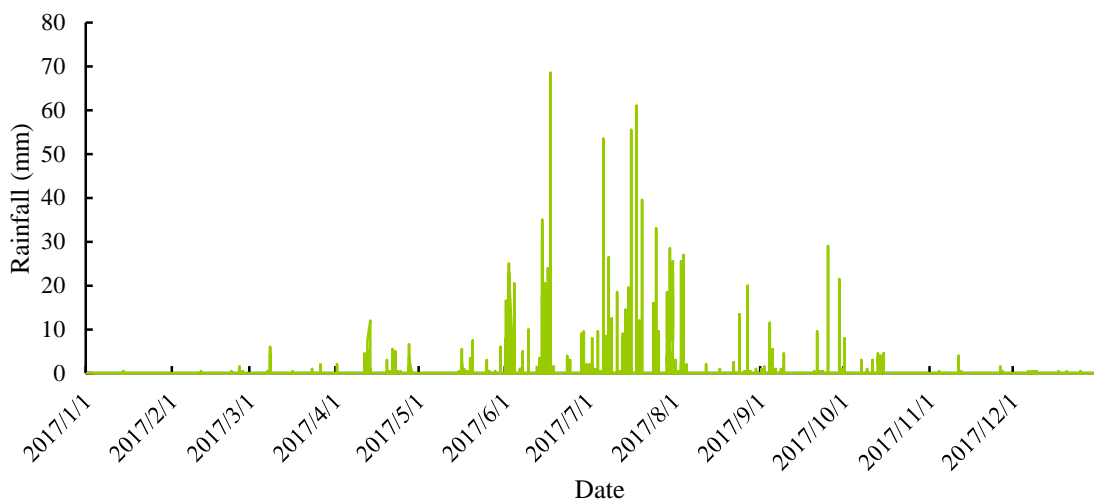


圖 4.2-3 竹崎雨量測站 2017 全年之日降雨量

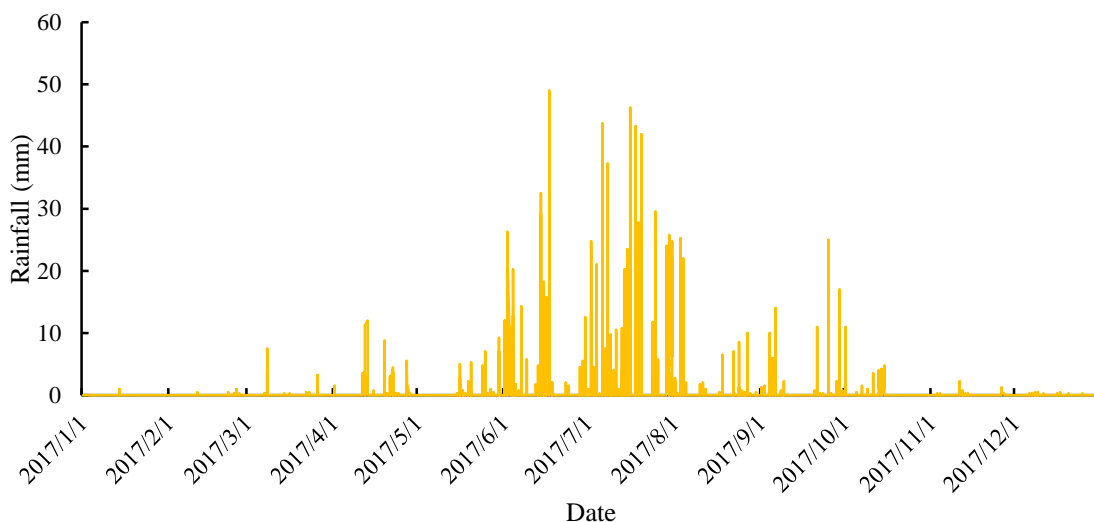


圖 4.2-4 流域上游 2017 全年之平均日降雨量

如圖 4.2-1 至圖 4.2-4 可明顯觀察到，朴子溪流域的雨季集中在全年七月前後，河川水位高度在此期間也有一定抬升。原始資料經過分析后，可得到如表 4.2-1 所示之初步資料。牛稠溪橋(1)水位測站在 2017 一年間，水位平均高度為 15.8632m，最高水位為 21.0567m，最低水位為 15.58m，水位標準差為 0.4338m，水位均方根為 15.8692m。

本研究為了快速得到結果，未對率定資料做出篩選。若希望使得模式準確度進一步提升，可以考量排除旱季之資料，甚至排除所有未降雨日期之資料。

表 4.2-1 原始資料分析

項目	水位測站每小時高度 (m)	番路測站每小時降雨量 (mm)	竹崎測站每小時降雨量 (mm)	平均每小時降雨量 (mm)
平均值	15.8632	0.2833	0.2550	0.2691
最大值	21.0567	48	68.5	49
最小值	15.58	0	0	0
標準差	0.4338	2.1880	2.1399	2.0247
均方根	15.8692	2.2062	2.1551	2.0425

### 4.3 模式預測資料概述

如第三章所述，研究選取了三次 2018 年之降雨事件，分別為 823 暴雨及兩次隨機選取的一般性降雨，用以驗證水位預測模式。預測資料使用之水位測站資料為每小時水位平均值，處理方式與訓練資料相同。降雨資料需要進行前處理，將自動雨量站即時記錄之降雨量資料按小時進行累加，得到以小時為單位的降雨量，缺失項資料取前後項平均值，最終得到之總資料數量分別為 936、1008、936 筆，如圖 4.3-1 至 4.3-5 所示。

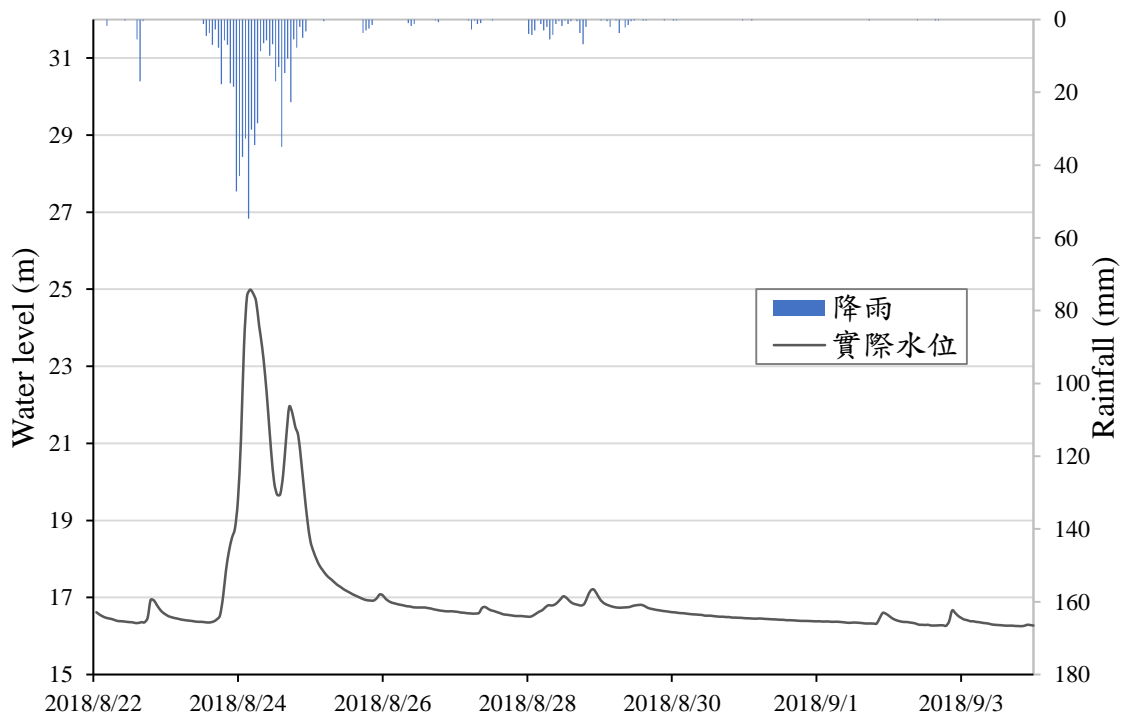


圖 4.3-1 事件一之降雨量與水位關係圖

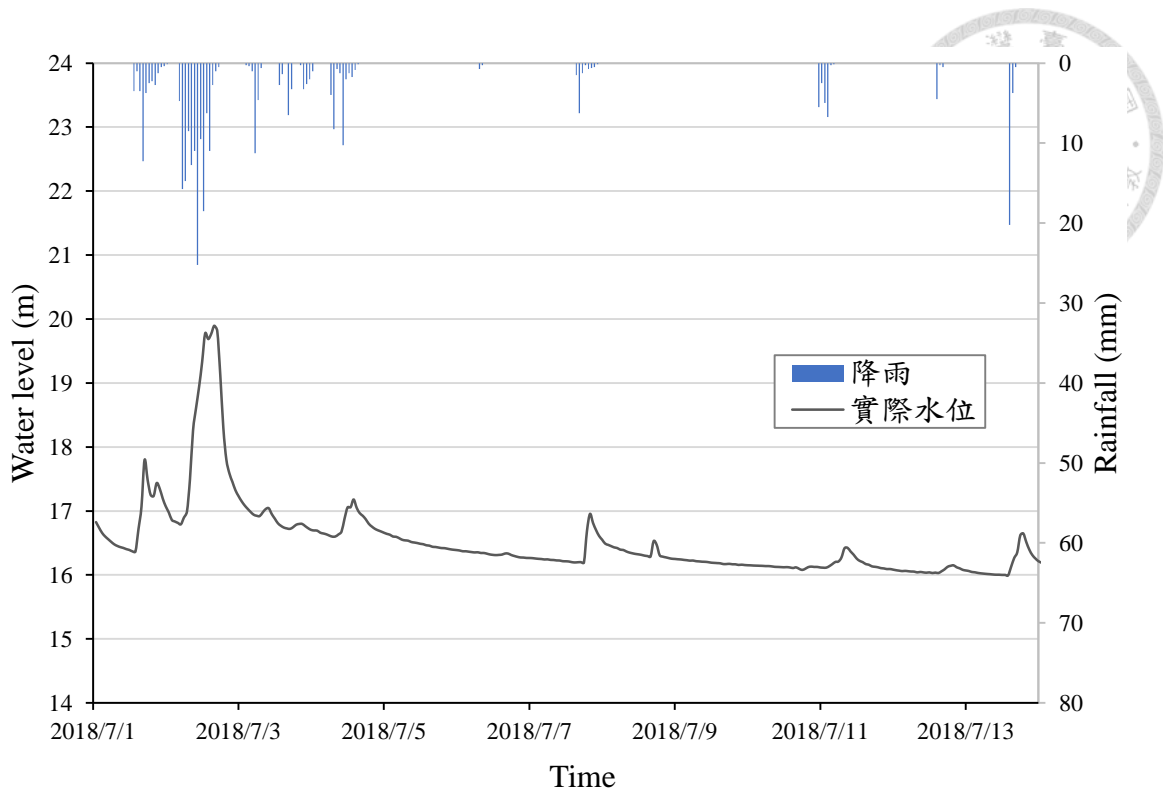


圖 4.3-3 事件二之降雨量與水位關係圖

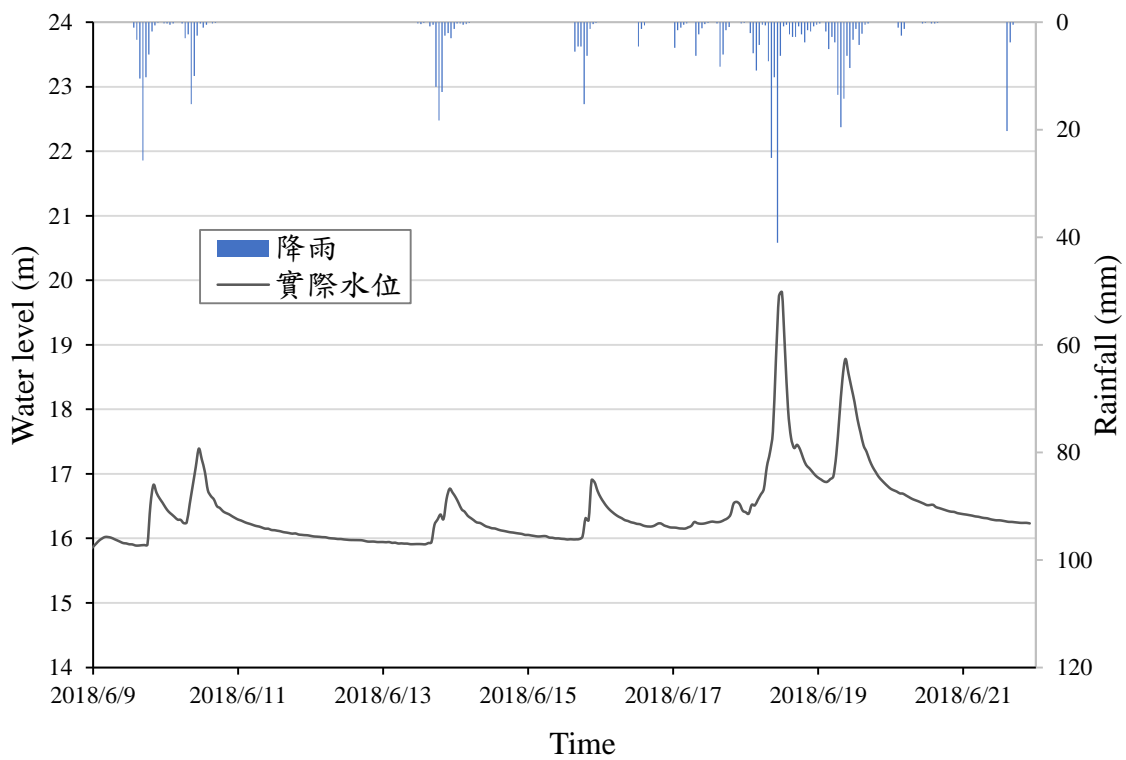


圖 4.3-2 事件三之降雨量與水位關係圖

表 4.3-1 待預測事件之降雨量分析

項目	事件一上游流域 平均降雨量 (mm)	事件二上游流域 平均降雨量 (mm)	事件三上游流域 平均降雨量 (mm)
平均值	2.1915	0.9899	1.5405
一小時最大值	54.75	20.25	41
六小時最大值	245.75	88	90.75
最小值	0	0	0
標準差	7.1003	2.9969	4.3306
均方根	7.4308	3.1459	4.5964

表 4.3-2 待預測水位資料分析

項目	事件一水位高度 (m)	事件二水位高度 (m)	事件三水位高度 (m)
平均值	17.0293	16.5034	16.4208
最大值	24.9883	19.895	19.8133
最小值	16.2567	15.994	15.82
標準差	1.5143	0.6112	0.5845
均方根	17.0965	15.9379	16.4312

表 4.3-1 與表 4.3-2 分別整理了待預測事件的雨量及水位資料，在三場降雨事件中，事件一降雨量最高，水位變化最大。以圖 4.3-1 所示的事件一為例，流域上游平均降雨量為 2.1915mm，一小時最高累積降雨量為 54.75mm，六小時最高累積降雨量為 245.75mm，雨量標準差為 7.1003mm，雨量均方根為 7.4308mm。

#### 4.4 支援向量回歸預測結果

本研究之支援向量回歸模式使用了基於多項式之核函數，並隨機將資料分為 90% 的訓練集與 10% 的驗證集，以 2017 年一整年資料進行訓練，缺失部分使用前後之資料進行插值處理。

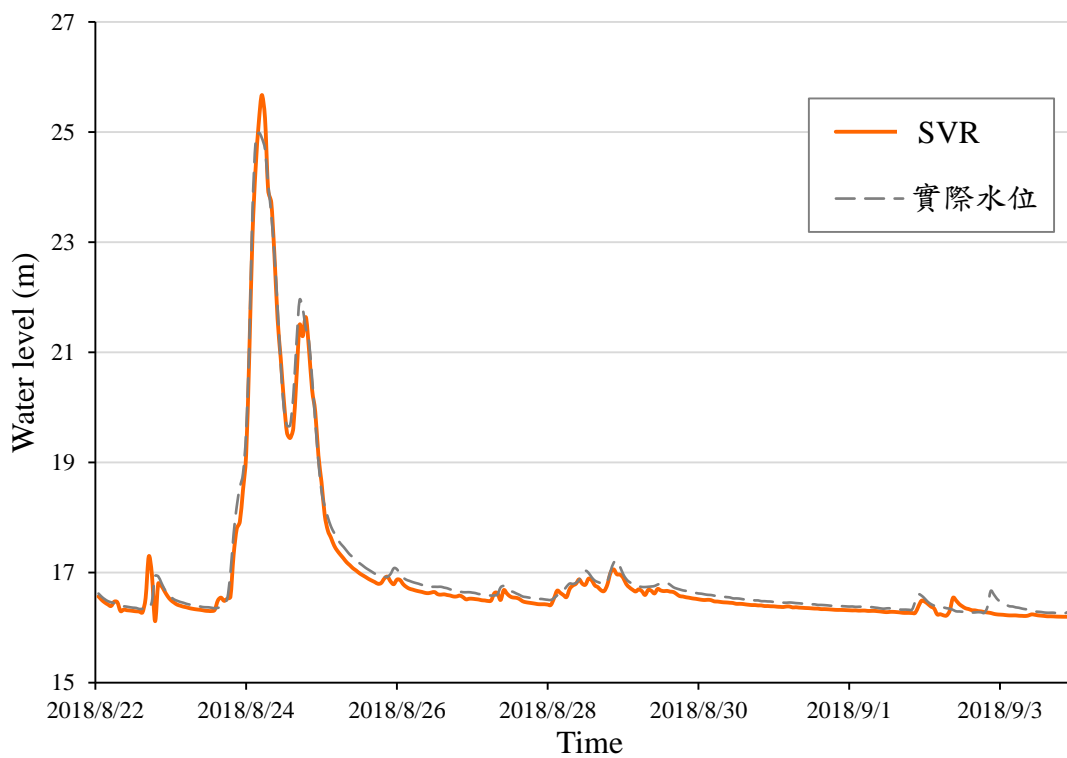


圖 4.4-2 事件一支援向量回歸水位預測圖

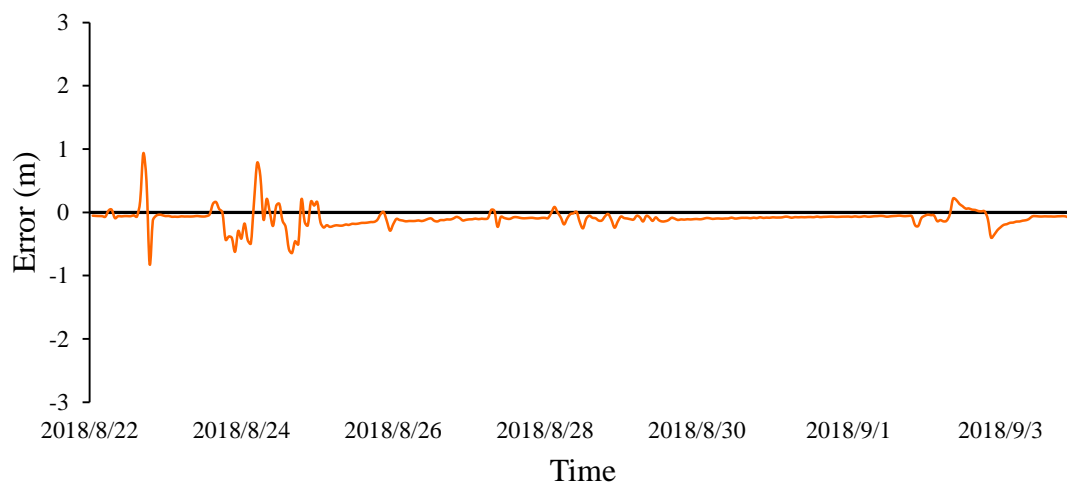


圖 4.4-1 事件一支援向量回歸水位預測誤差圖

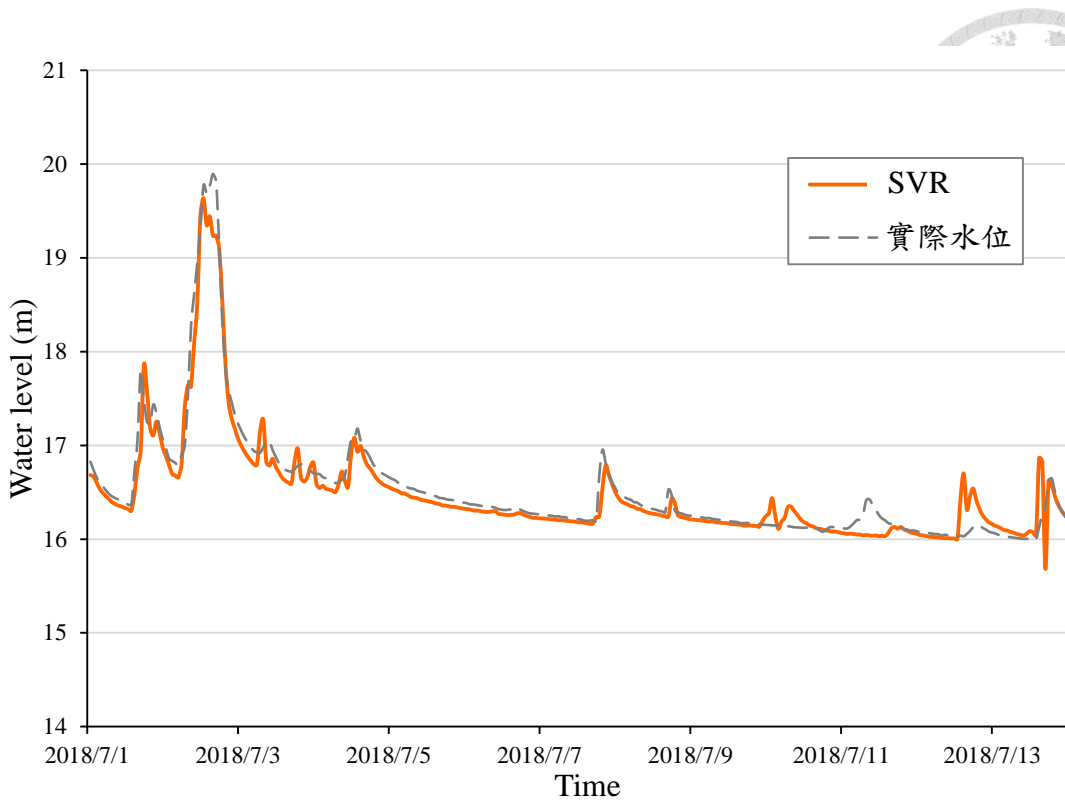


圖 4.4-4 事件二支援向量回歸水位預測圖

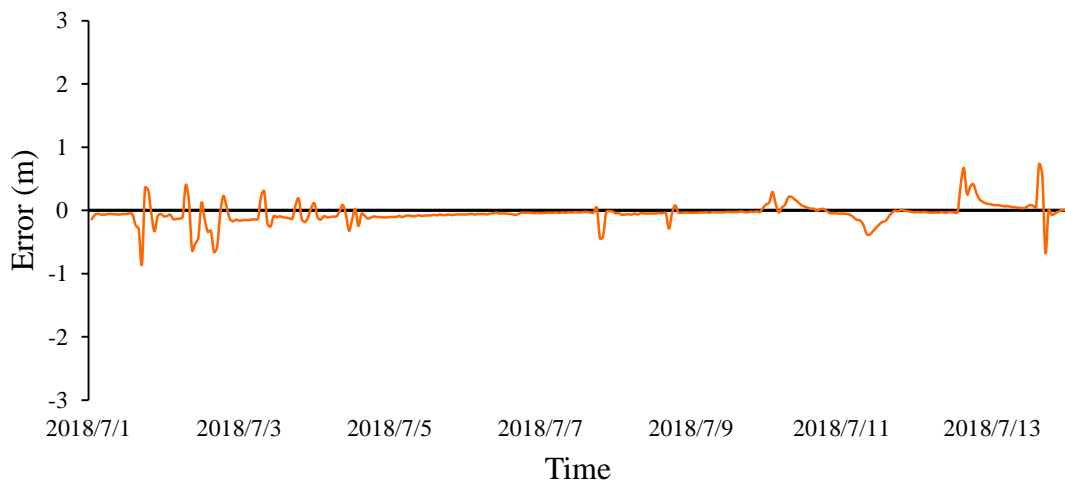


圖 4.4-3 事件二支援向量回歸水位預測誤差圖



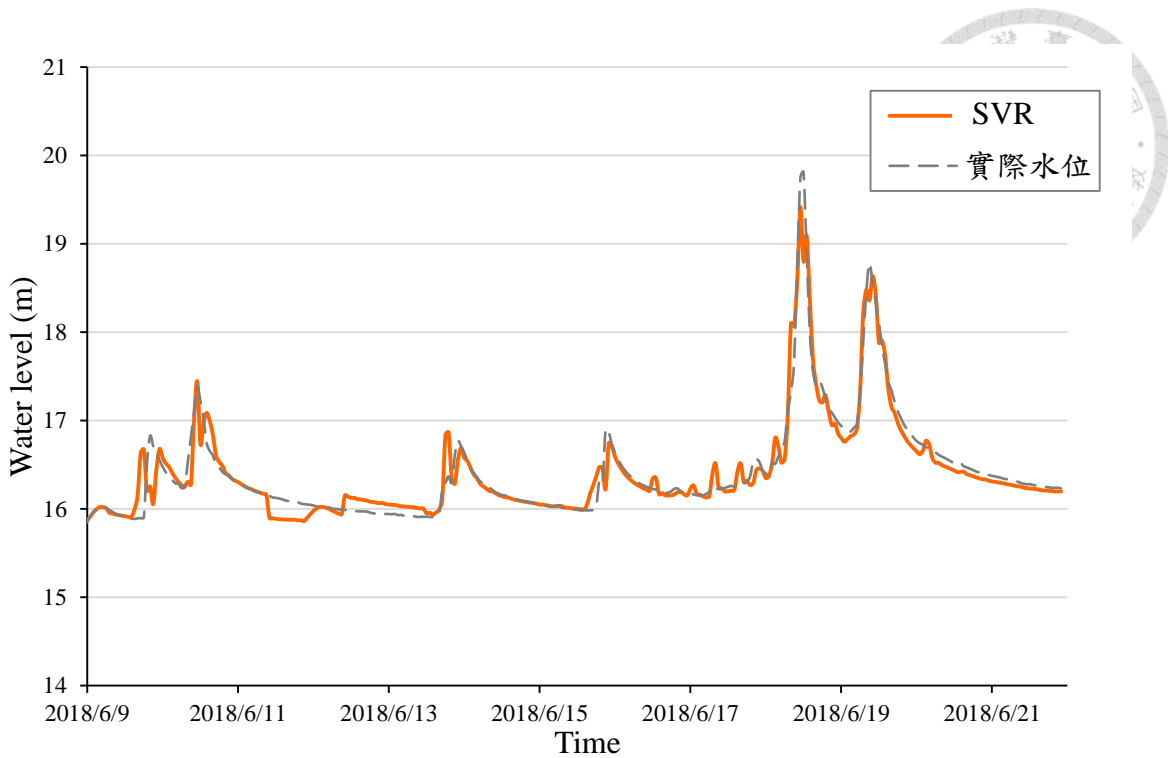


圖 4.4-6 事件三支援向量回歸水位預測圖

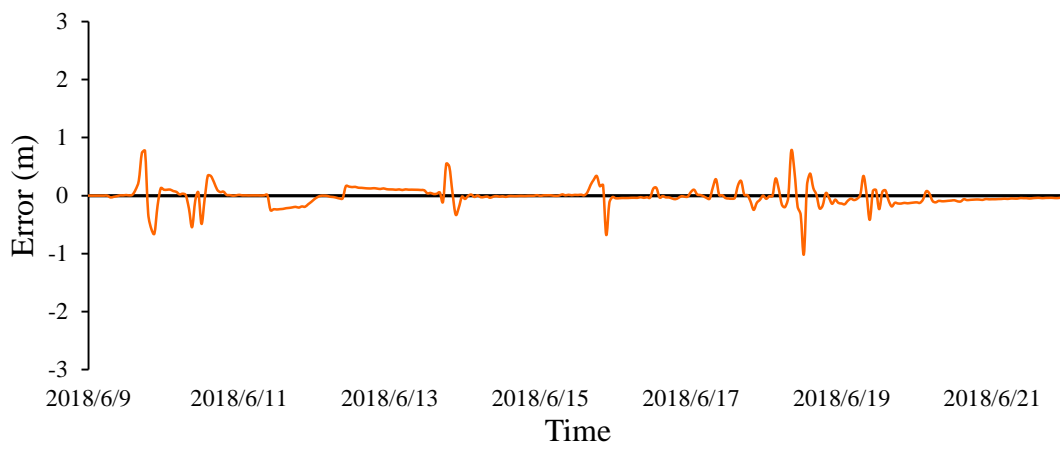


圖 4.4-5 事件三支援向量回歸水位預測誤差圖

表 4.4-1 支援向量回歸水位預測結果

事件	$R^2$	RMSE	MAE	EPWL
事件一	0.9902	0.1754	0.9341	0.5713
事件二	0.9295	0.1728	0.8525	-0.6597
事件三	0.9147	0.2061	1.0129	0.1773

使用支援向量回歸模式對三場降雨事件進行水位預測，得到之結果如圖 4.4-1 至圖 4.4-6 所示。表 4.4-1 整理了模式之評鑒指標，在三場事件中，事件一的決定係數最高，達到了 0.9902，事件三最低，為 0.9147。事件一與事件二的均方根誤差相仿，約為 0.17，最大均方根誤差與最大絕對誤差出現在事件三，分別為 0.2061 與 1.0129，事件二尖峰水位誤差最高，低估了 0.6597 米。

## 4.5 灰色水位預測結果

本研究採用之灰色模式為一次累加生成之 DHGM (2,2) 模式，假定河川上游集水區為一整體系統，將水位高度變化視為流量變化之另一形式輸出，效仿灰色降雨——逕流模式建立了一個降雨-水位高度關係之灰色模式，具體如第三章(3-27)式所述。

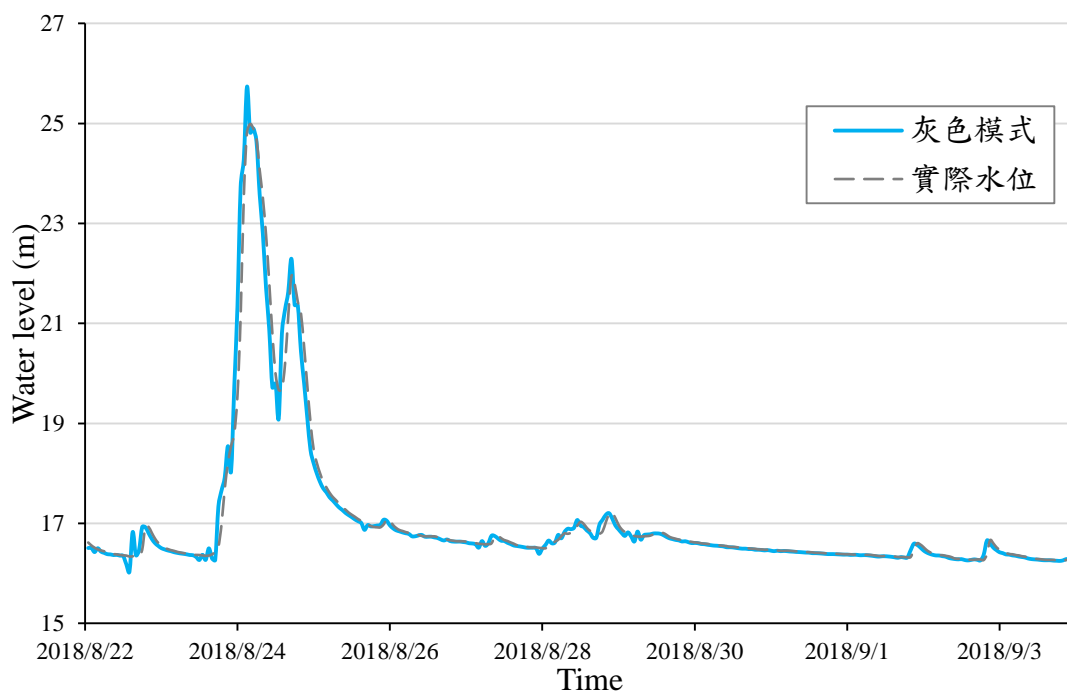


圖 4.5-1 事件一灰色模式水位預測圖

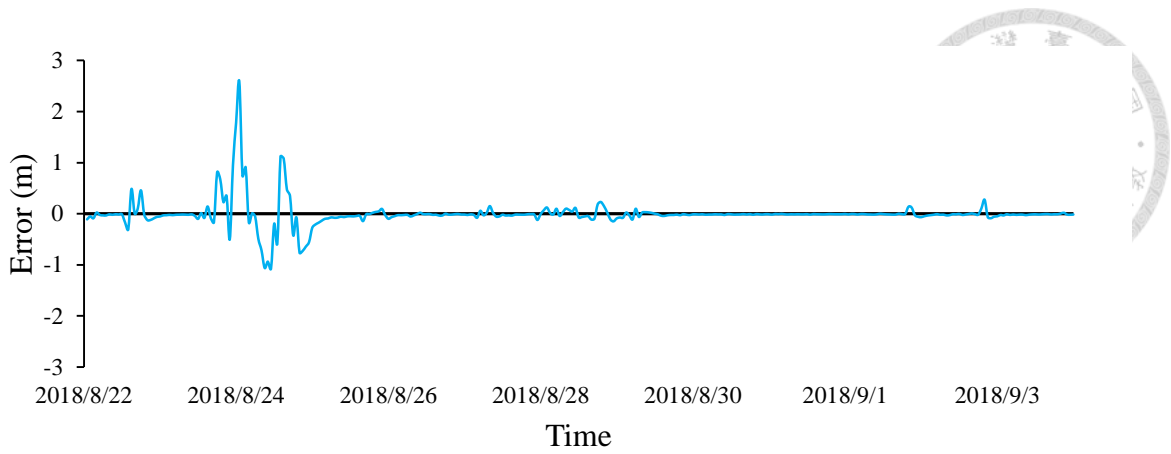


圖 4.5-4 事件一灰色模式水位預測誤差圖

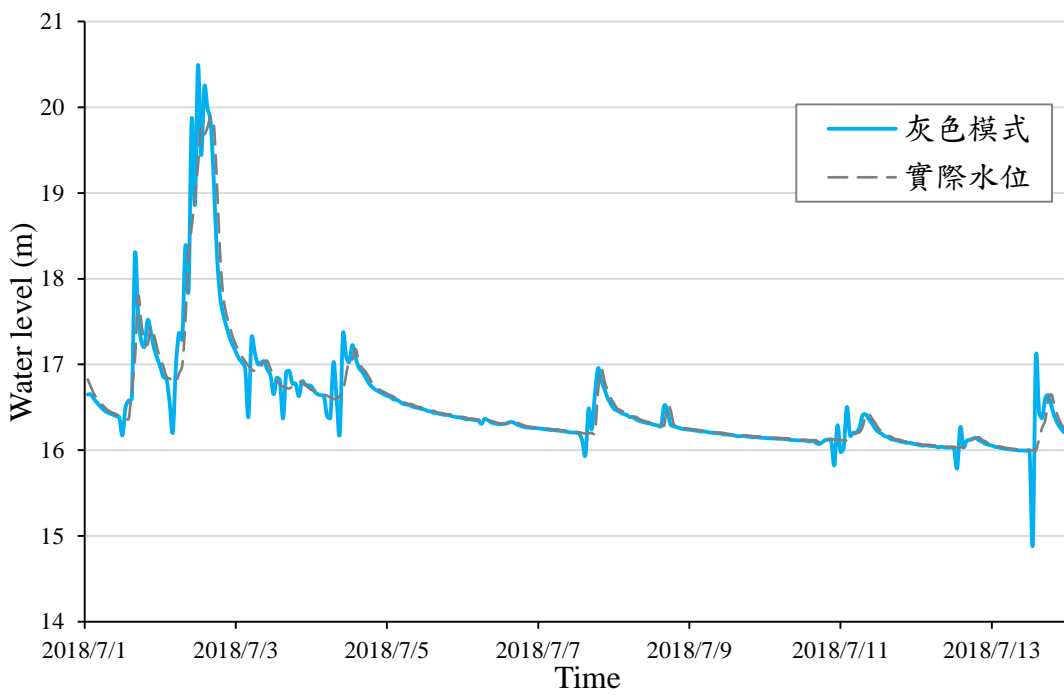


圖 4.5-2 事件二灰色模式水位預測圖

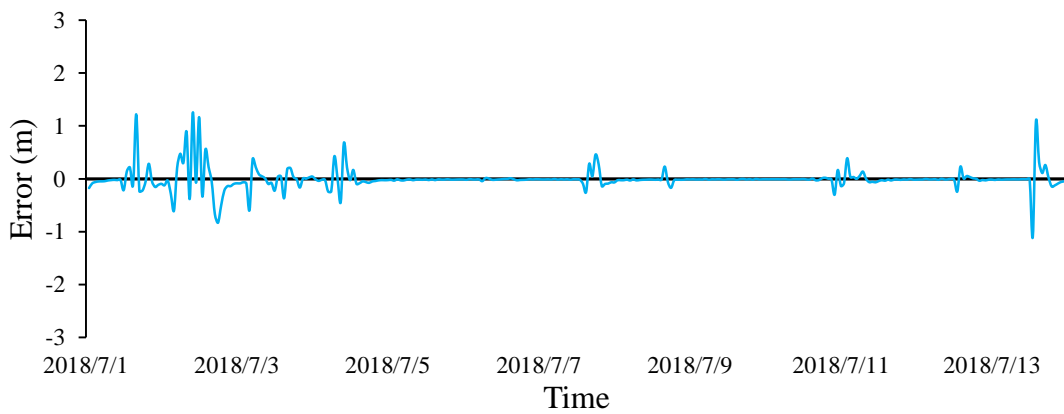


圖 4.5-3 事件二灰色模式水位預測誤差圖

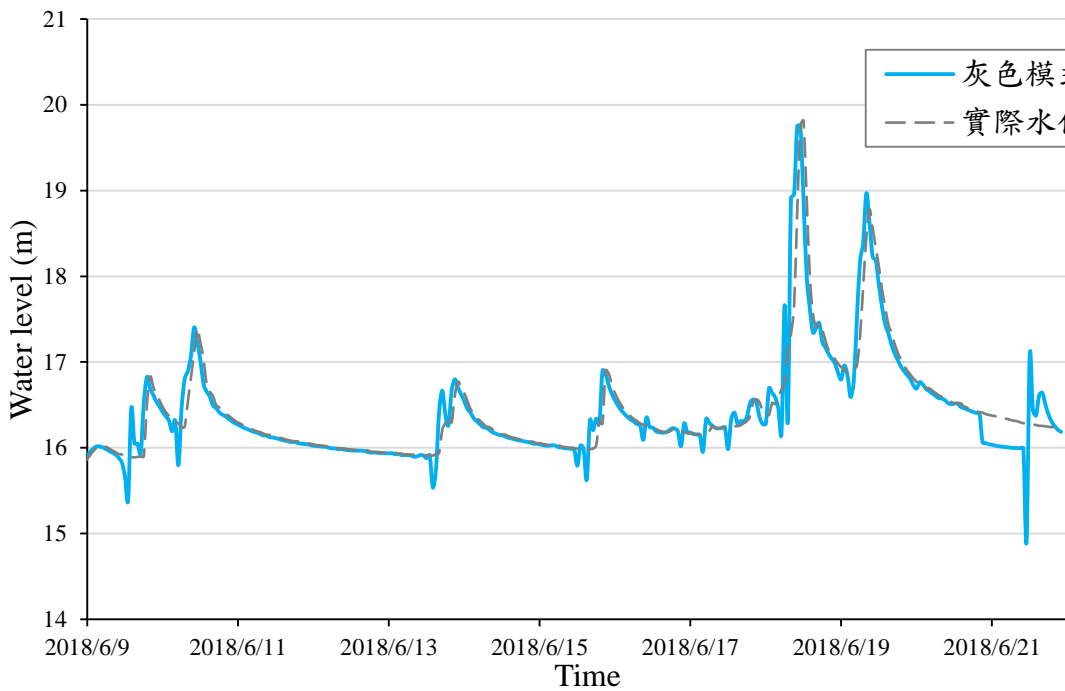


圖 4.5-6 事件三灰色模式水位預測圖

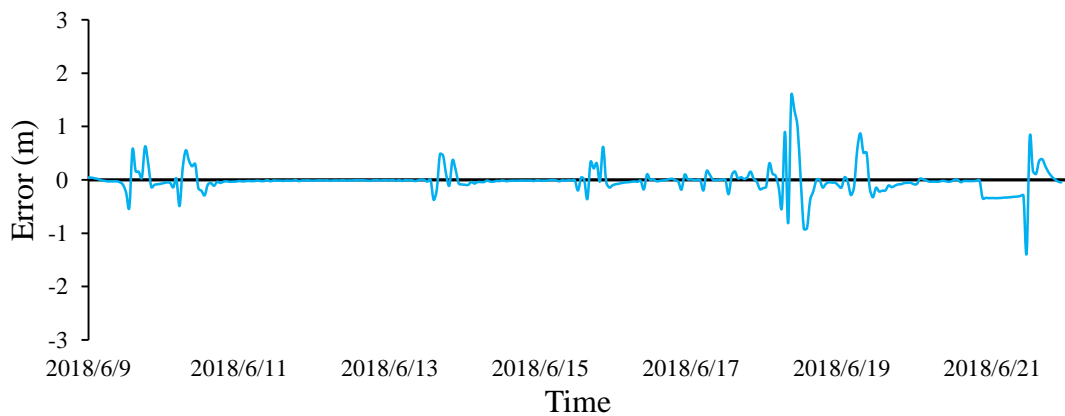


圖 4.5-5 事件三灰色模式水位預測誤差圖

表 4.5-1 灰色模式水位預測結果

事件	$R^2$	RMSE	MAE	EPWL
事件一	0.9678	0.2707	2.5886	0.9104
事件二	0.8853	0.2251	1.2568	-0.3297
事件三	0.8671	0.2329	1.5830	-0.9162

灰色模式的水位預測結果如圖 4.5-1 至圖 4.5-6 所示，表 4.5-1 整理了模式之評鑒指標。在三場降雨事件中，事件一的決定係數最高，為 0.9678，但最大均



方根誤差與最大絕對誤差亦為最大，分別為 0.2707 與 2.5886。事件二、三之決定係數相對較低，分別為 0.8853 與 0.8671。尖峰水位誤差以事件一、三為最大，分別高估 0.9104 米與低估 0.9162 米。

## 4.6 線性回歸預測結果

本研究採用簡單線性回歸方程，如第三章 (3-31) 至 (3-33) 式所述，目的為探討模式過度簡化后之誤差大小。同灰色模式一樣，研究假定河川上游集水區為一整體系統，將水位高度變化視為流量變化之另一形式，建立降雨-水位高度關係之線性方程。

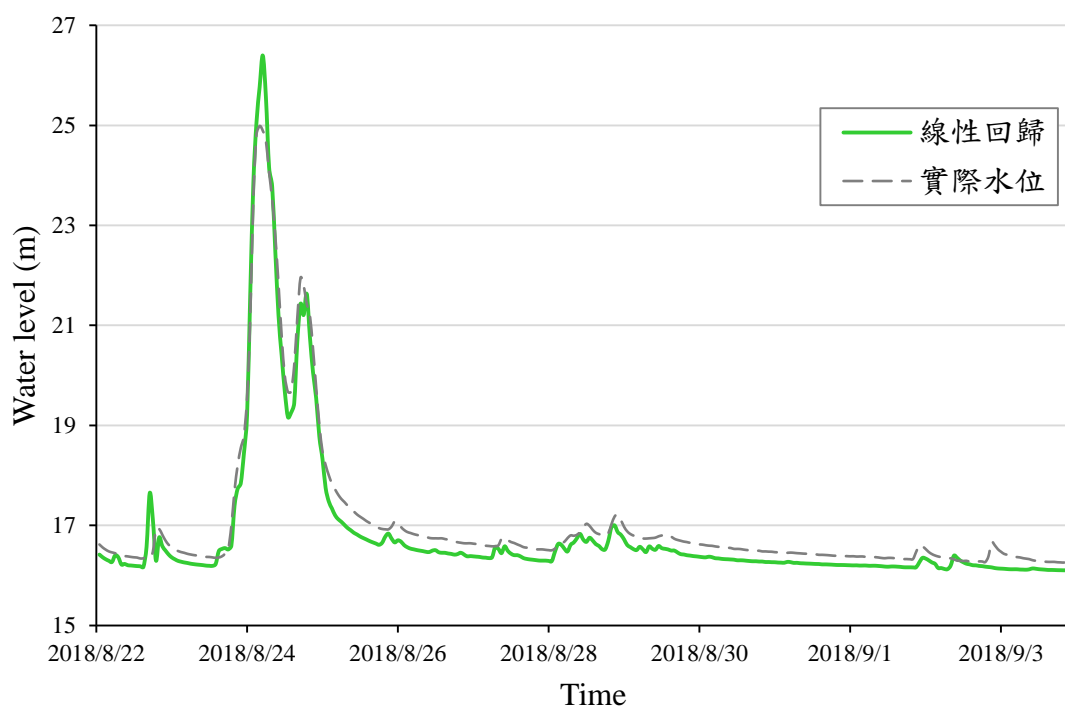


圖 4.6-1 事件一線性回歸水位預測圖

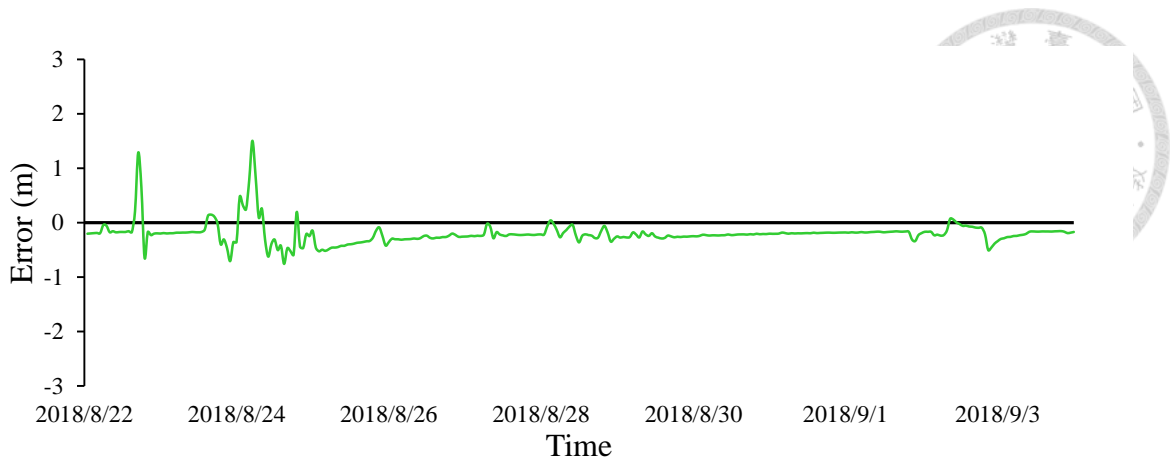


圖 4.6-3 事件一線性回歸水位預測誤差圖

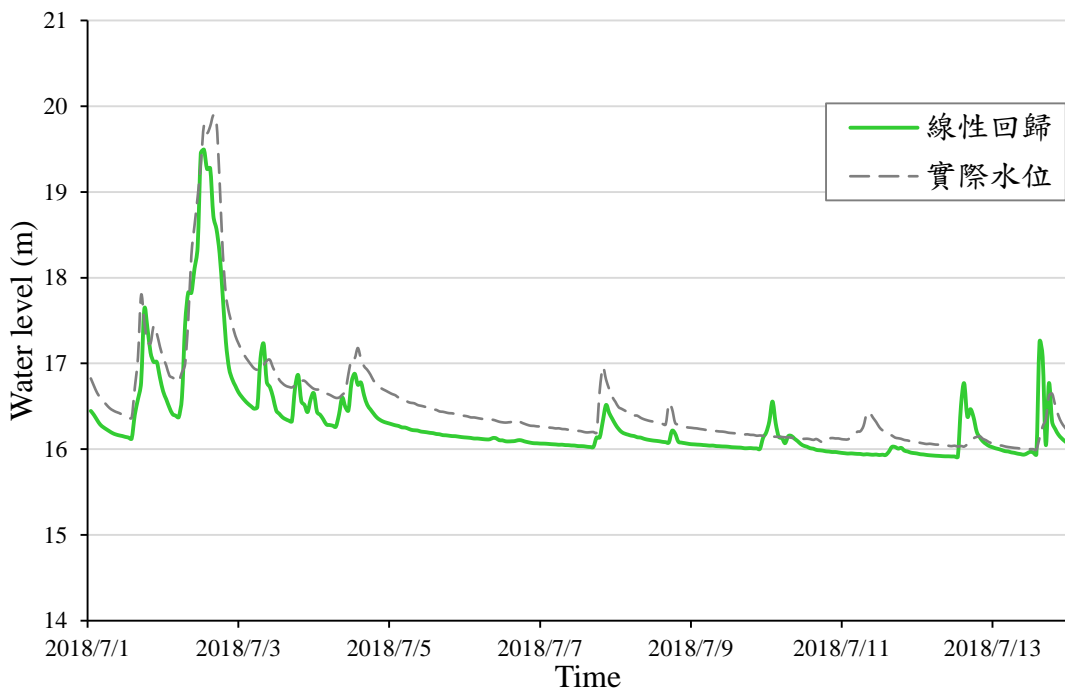


圖 4.6-2 事件二線性回歸水位預測圖

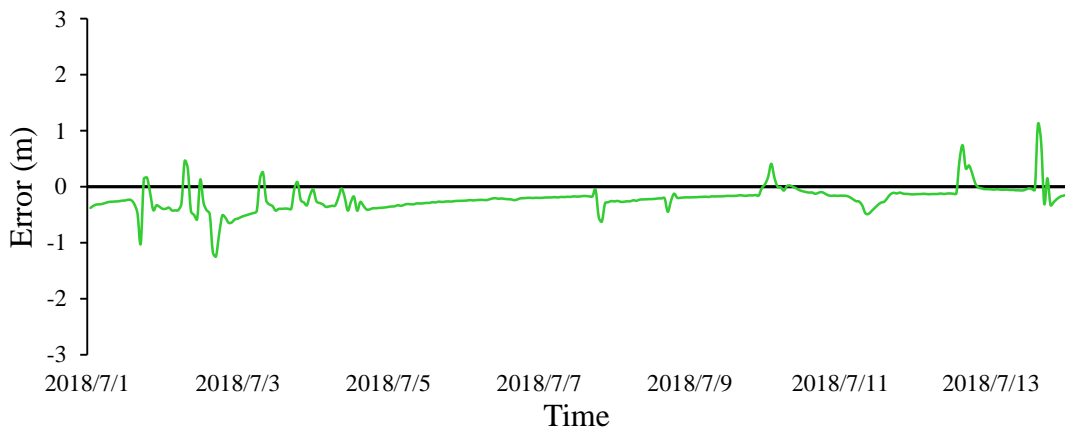


圖 4.6-4 事件二線性回歸水位預測誤差圖

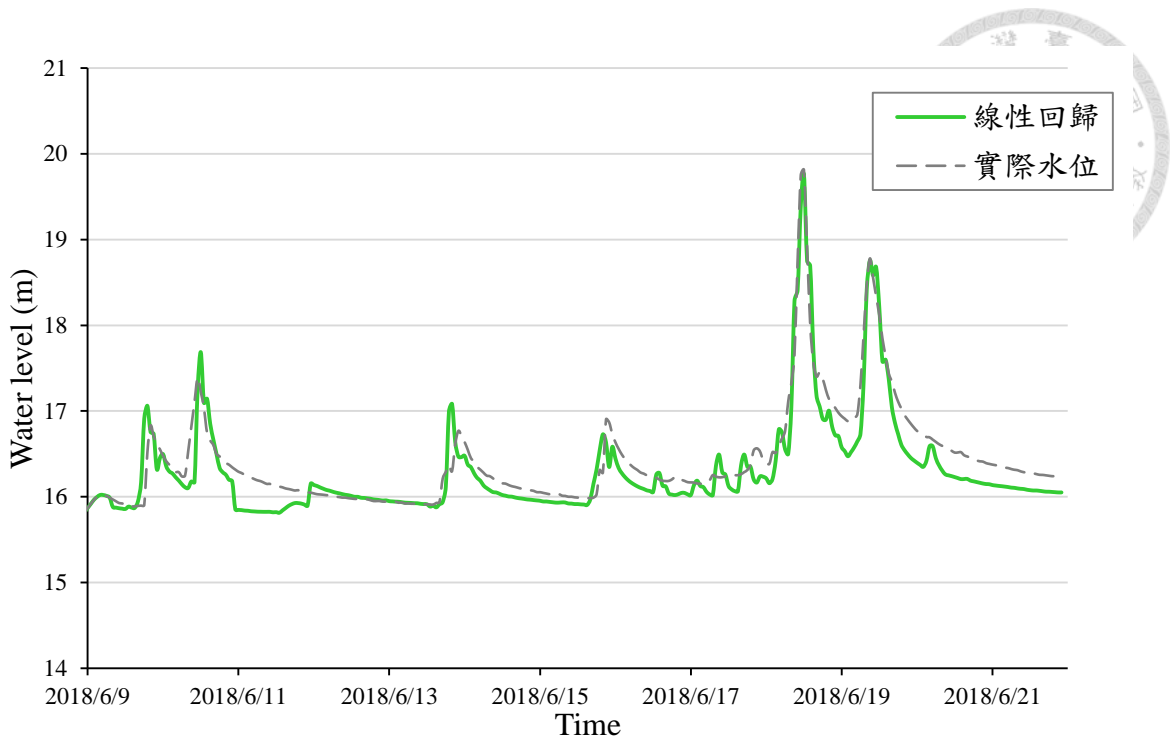


圖 4.6-5 事件三線性回歸水位預測圖

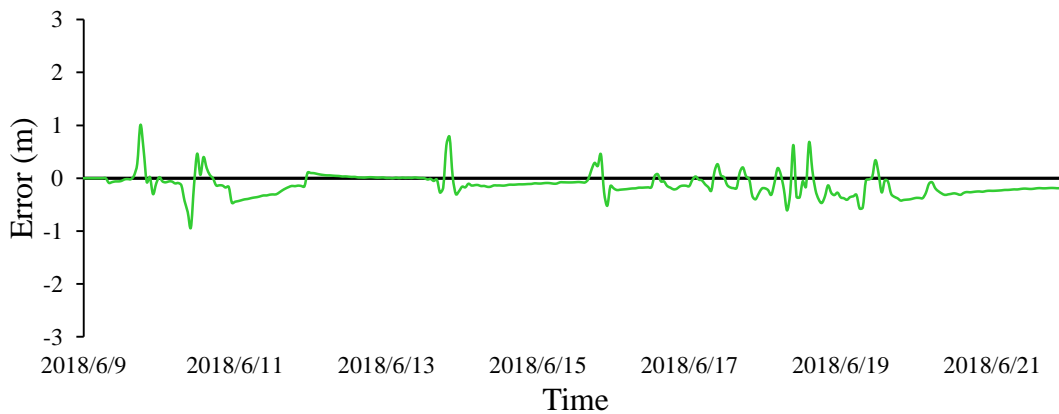


圖 4.6-6 事件三線性回歸水位預測誤差圖

表 4.6-1 線性回歸模式水位預測結果

事件	$R^2$	RMSE	MAE	EPWL
事件一	0.9832	0.2911	1.5039	1.5039
事件二	0.8735	0.3079	1.2471	-1.1714
事件三	0.8751	0.2428	1.0085	0.6828

線性回歸模式水位預測結果如圖 4.6-1 至圖 4.6-6 所示，表 4.6-1 整理了模式之評鑒指標。在三場降雨事件中，事件一的決定係數最高，為 0.9832，事件二、三之決定係數相對較低，都在 0.87 左右。最大均方根誤差以事件一、二為最高，分別為 0.2911 與 0.3079。在三場事件中，最大絕對誤差與最高尖峰水位誤差都出現在了事件一，均為 1.5039 米，可能是線性回歸對高水位時的峰值推估不佳所致。

## 4.7 三種模式之比較

以三種模式在三場降雨事件中的預測結果與實際水位進行比較，得到其預測值與誤差值如圖 4.7-1 至圖 4.7-6 所示，資料散佈如圖 4.7-7 至圖 4.7-9 所示。圖中紅藍綠三種彩色線條分別為支援向量回歸 (SVR)、灰色模式、線性回歸模式之預測值，黑色虛線為實際值。

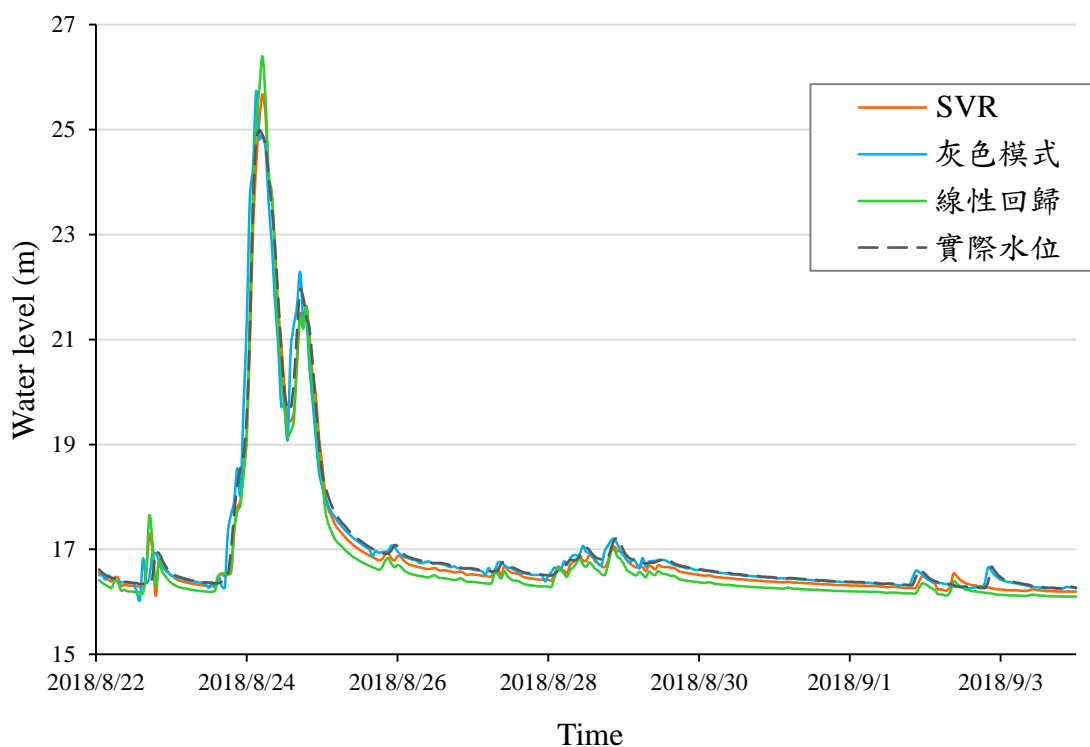


圖 4.7-1 事件一之模式結果對比



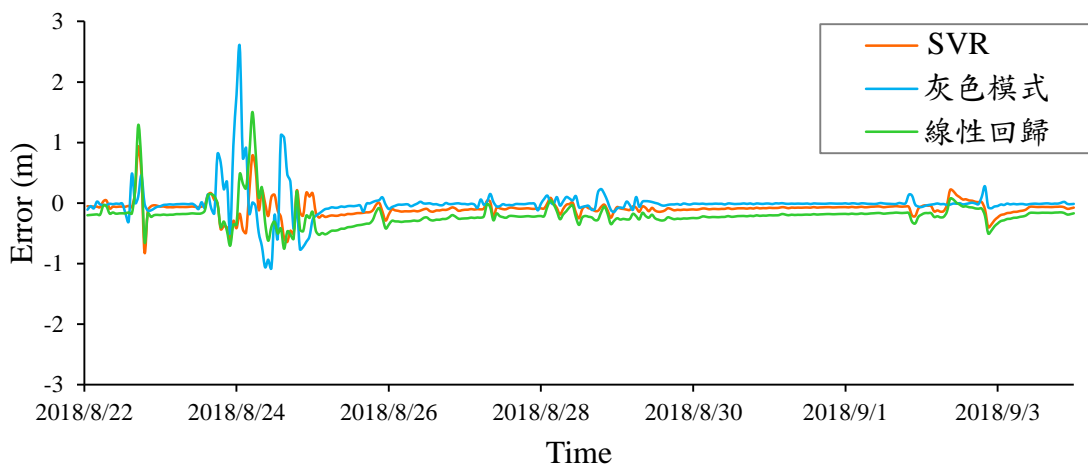


圖 4.7-2 事件一之模式誤差對比

表 4.7-1 事件一之分析結果

模式	R <sup>2</sup>	RMSE	MAE	EPWL
支援向量回歸	0.9902	0.1754	0.9341	0.5713
灰色模式	0.9678	0.2707	2.5886	0.9104
線性回歸	0.9832	0.2911	1.5039	1.5039

對比事件一之預測結果可得到圖 4.7-1 與圖 4.7-2，其評鑒指標如表 4.7-1 所示，三種模式之決定係數呈現支援向量回歸 > 線性回歸 > 灰色模式，RMSE 則為支援向量回歸 > 灰色模式 > 線性回歸，且支援向量回歸之優勢較大。灰色水位之 MAE 誤差最大，出現在峰值流量前之 4 小時，達到 2.5886 米，並且第二高誤差出現在前 5 小時，此時河川之水位已開始上漲，但未達到高點。推測與當時之雨量起伏不定，集水區降雨量不均，或模式對高水位抬升（約 8.7m）之適應性相對較差有關。在 EPWL 誤差中，支援向量回歸表現最佳，為 0.5713 米，線性回歸之預測誤差最高，已達到 1.5 米，其參考價值有待考量。

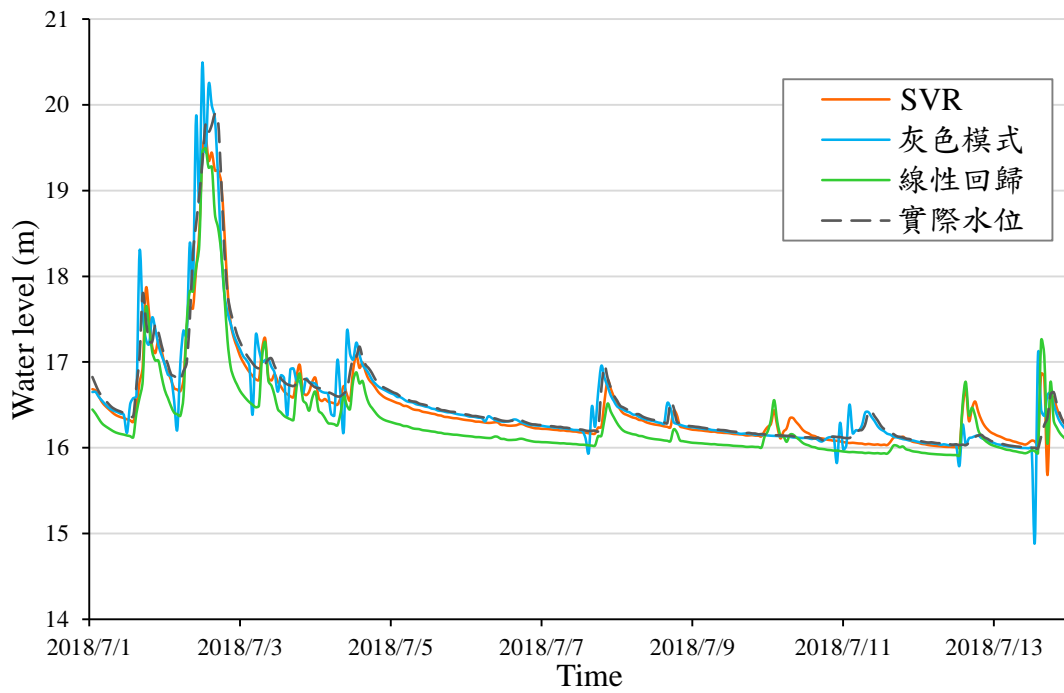


圖 4.7-4 事件二之模式結果對比

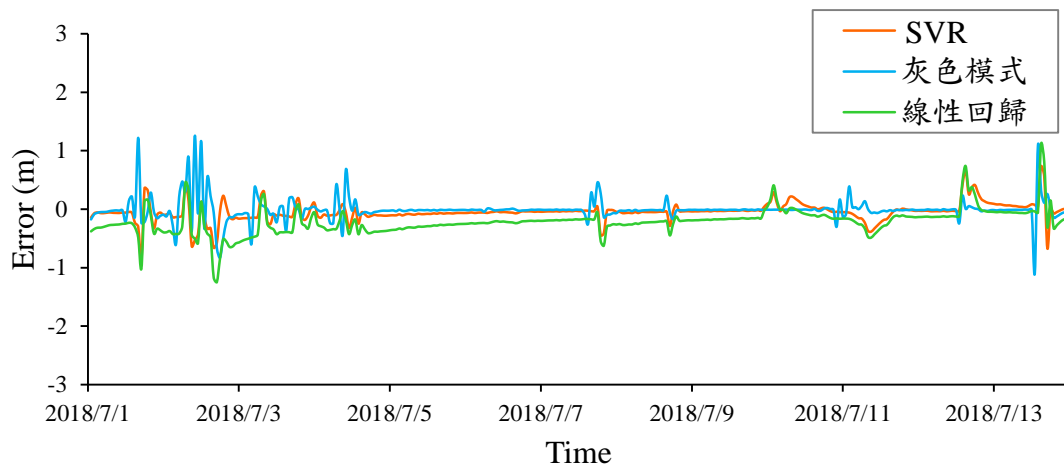


圖 4.7-3 事件二之模式誤差對比

表 4.7-2 事件二之分析結果

模式	$R^2$	RMSE	MAE	EPWL
支援向量回歸	0.9295	0.1728	0.8525	-0.6597
灰色模式	0.8853	0.2251	1.2568	-0.3297
線性回歸	0.8735	0.3079	1.2471	-1.1714

對比事件二之預測結果可得到圖 4.7-3 與圖 4.7-4，其評鑒指標如表 4.7-2 所示。事件二之決定係數與 RMSE 均呈現支援向量回歸 > 灰色模式 > 線性回歸，且支援向量回歸之優勢較大，決定係數與均方根誤差分別為 0.9295 與 0.1728。MAE 表現以支援向量回歸為最佳，灰色模式、線性回歸相近，均優於事件一，這是由於降雨強度減弱所致。EPWL 誤差中，灰色模式表現最佳，支援向量回歸與之有一定差距，同樣也是線性回歸之預測誤差最高。值得注意的是，三種模式在此事件中均低估了洪水水位，在實務應用中需要謹慎考量。

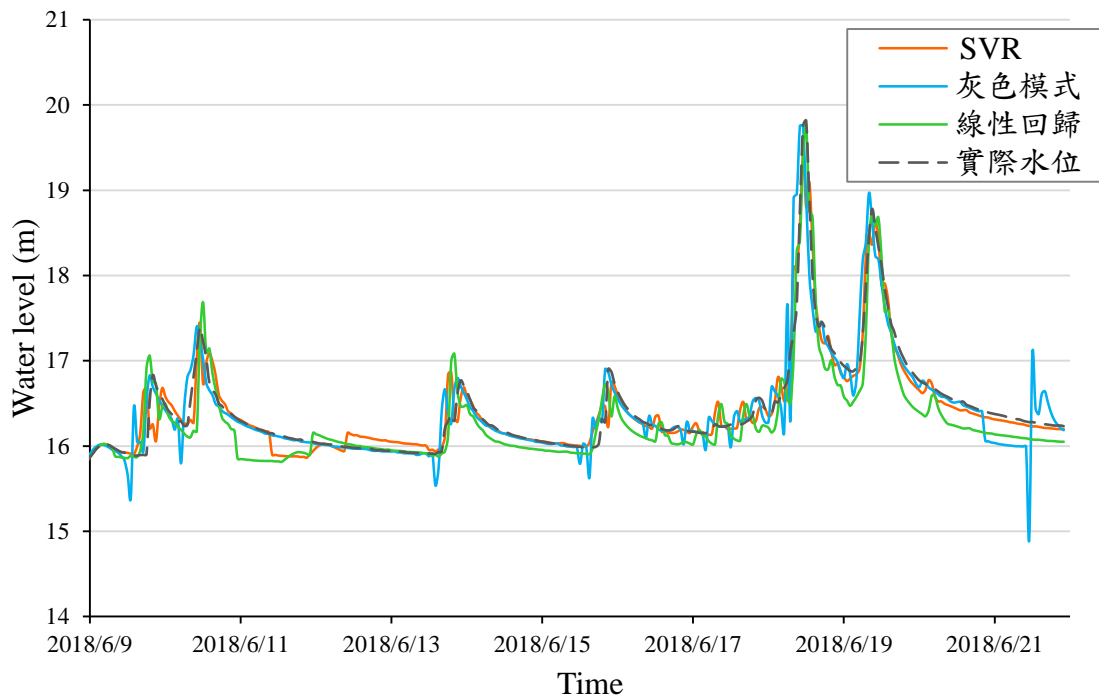


圖 4.7-5 事件三之模式結果對比

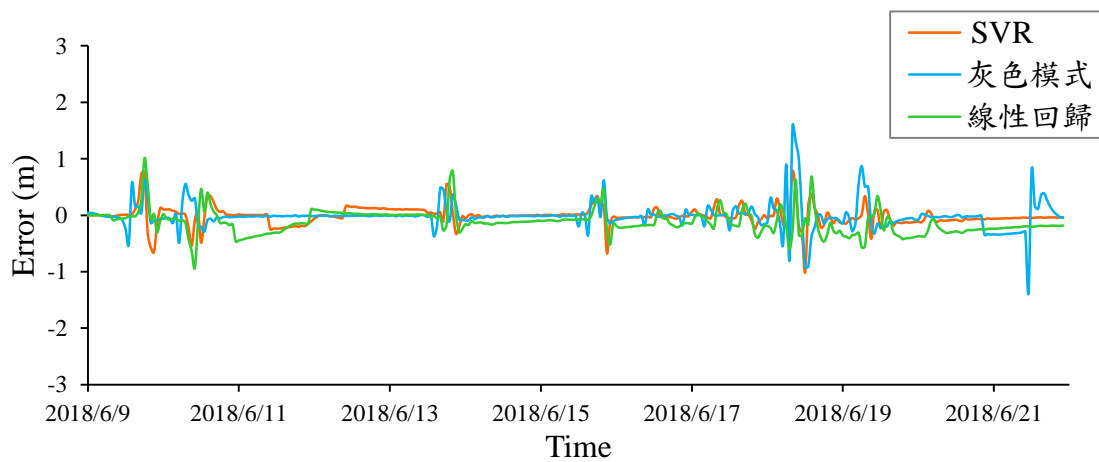


圖 4.7-6 事件三之模式誤差對比



表 4.7-3 事件三之分析結果

模式	R <sup>2</sup>	RMSE	MAE	EPWL
支援向量回歸	0.9147	0.2061	1.0129	0.1773
灰色模式	0.8671	0.2329	1.5830	-0.9162
線性回歸	0.8751	0.2428	1.0085	0.6828

對比事件三之預測結果可得到圖 4.7-5 與圖 4.7-6，其評鑒指標如表 4.7-3 所示。事件三之決定係數與 RMSE 特性類似事件一，均為支援向量回歸最佳，為 0.9147，灰色模式與線性回歸較為相近。MAE 表現以支援向量回歸、線性回歸為最佳，約為 1.0 米，灰色模式較差。在 EPWL 誤差中，支援向量回歸最佳，灰色模式表現最差，低估了 0.9162 米。線性回歸模式在此事件中可相關指標表現與灰色模式較為類似，是由於降雨次數較多且雨量較小，同時灰色模式的雨量資料輸入模式有一定限制所致。

可以觀察到在三場事件剛開始降雨的前幾個小時，灰色水位模式出現了較大波動，表現為先低估水位后高估水位。本人推測是由於灰色模式之微分階數或者項次不夠，導致其在降雨轉折點之預測能力不足。

在預測的總共 960 筆水位資料中，支援向量回歸、灰色模式、線性回歸的相關係數分別為 0.9741、0.9434、0.9560，其預測值與觀測值之散布圖如圖 4.7-7 至 4.7-9 所示。

在所有模式中，支援向量回歸決定係數最高，均方根、最大絕對誤差、尖峰水位誤差幾乎均為最小，證明其健壯性相對較強，在絕大多數情況下具有良好的表現，特別是針對高強度降雨之情形。灰色模式受原始資料影響，預測準確率次之。線性回歸模式某些預測值偏差值較高，在事件一、二中表現糟糕，可參考性相對較低。

因此，本研究以支援向量回歸為最佳洪水預測模式。

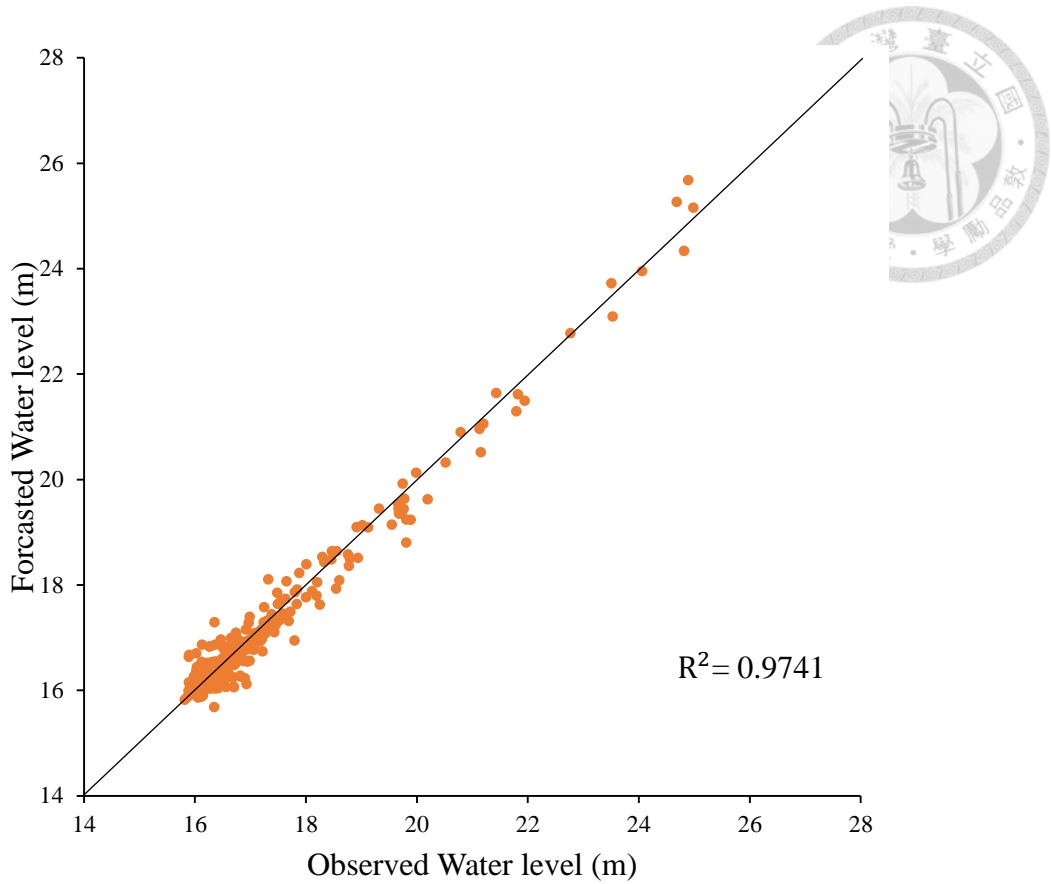


圖 4.7-8 支援向量回歸模式預測值與觀測值之散布圖

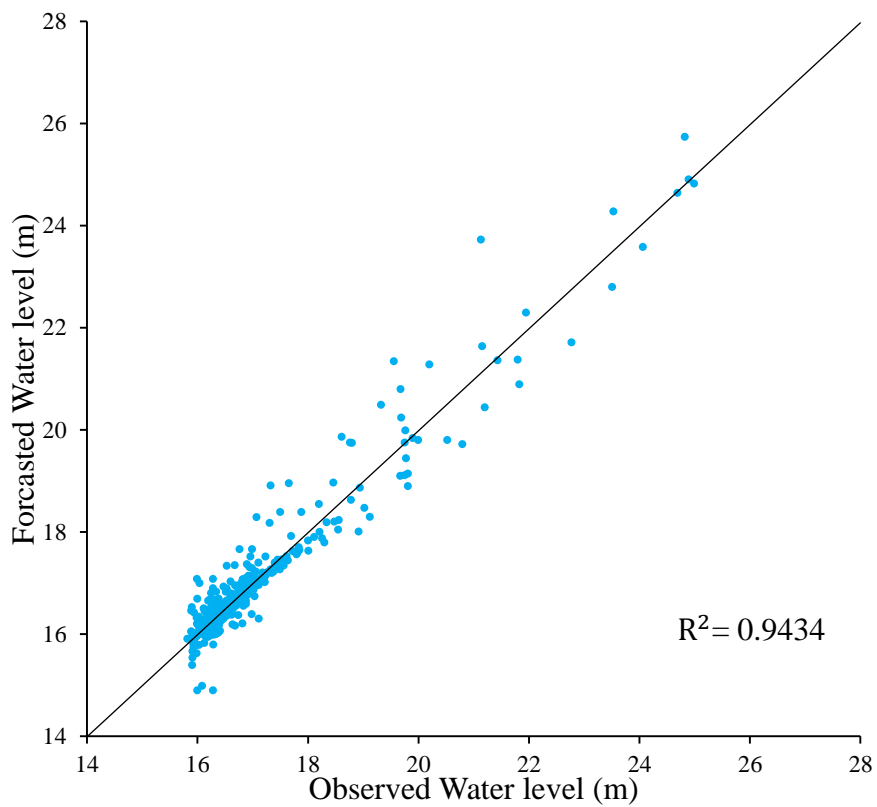


圖 4.7-7 灰色模式預測值與觀測值之散布圖

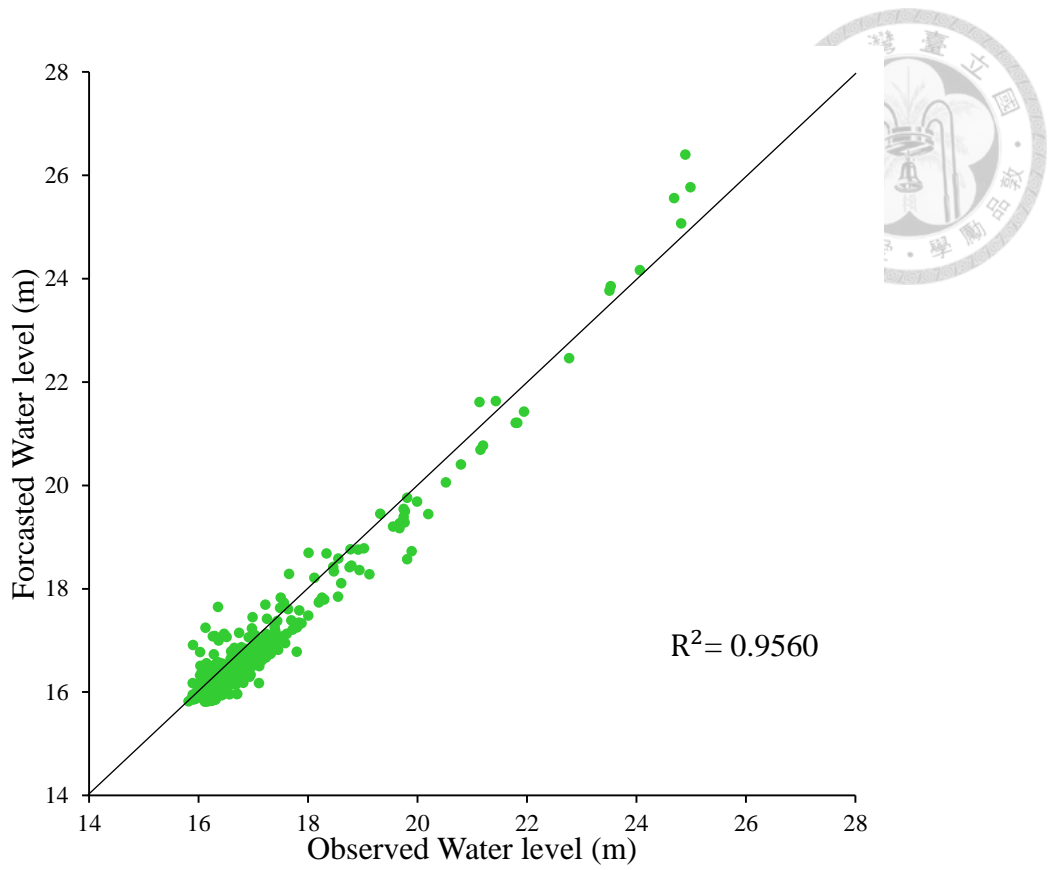


圖 4.7-9 線性回歸預測值與觀測值之散布圖

## 第五章 結論



### 5.1 結論

本研究利用朴子溪流域牛稠溪橋(1)水位測站，竹崎、番路雨量測站 2017 年一整年之水位、降雨資料，分別建立支援向量回歸、灰色水位、線性回歸三種降雨—水位模式，以集水區降雨量預測河川水位。建立模式后，使用 2018 年三場降雨事件分別作驗證，獲得以下結論：

1. 支援向量回歸模式在三場降雨事件中，決定係數分別為 0.9902、0.9295、0.9147，證明其在朴子溪中上游河段具有一定之預測能力，利用歷史資料訓練得到的模式架構，可快速預測未來之水位。
2. 在輸入資料未經篩選與降噪之情況下，支援向量回歸模式仍具有相對較好的預測能力，且優於其它兩種模式，證明其健壯性較強。
3. 然則受限於率定使用之降雨量推估方式、時間精度、資料雜訊，各模式在尖峰水位之預測中均具有一定偏差。
4. 在朴子溪上游的三次降雨事件中，各預測模式均具有一定之洪水預測能力，綜合精確度為支援向量回歸 > 灰色模式 > 線性回歸，推測支援向量回歸更擅長處理資料量有限之情形。
5. 本研究使用三種預測模式中，線性回歸在某些評鑒指標中均具有較大劣勢。這是由於水文過程中，降雨量與水位並非線性關係，若模式過度簡化，將引入較大誤差。

### 5.2 建議

1. 本研究使用之資料種類及數量較為苛刻，比如將整年之資料不加篩選直接輸入

作為變數。後續研究可以考慮僅使用有降雨之時段作為訓練集，以減少資料中的雜訊（Noise，即偏離期望值）。

2. 為達到快速預測之目的，研究僅使用上游兩個雨量站作為降雨參考，並未統計區域內其餘地區雨量，預測精確度一般。在實際應用中若條件允許，應利用更多雨量測站配合插值法，得到更加精確的上游集水區降雨量。同時，降雨產生之徑流量也與地形、集水區邊界距離有關，如能對集水區構建 3D 地圖或者相應的模式，可以更好地決定雨量站資料權重。
3. 在預測精度有限的情況下，研究未考量前置時間（Lead time）之模擬，故預測時間停留在了三小時，未來研究可考慮增加此項。
4. 由於本研究之預測事件均為短時、高强度降雨，模式不一定適用於所有降雨情形，未來研究應驗證盡可能多的降雨模式。
5. 朴子溪中上游的水性較為單純、坡短流急，降雨量與河川流量之相關度較高。未來的研究若針對大型河川預測水位，應考量更多變數。



## 參考文獻



- 王如意、易任 (1979)。應用水文學。茂昌圖書有限公司，臺北。
- 鄧聚龍 (1982)。灰色控制系統。华中工学院学报，10(3)，9-18。
- 夏軍 (1989)。水文系統 DHGM 模型識別與實時灰色預測，灰色系統論文集。
- 莊文南 (1993)。微分水文灰色模式之研究與應用。國立臺灣大學農業工程研究所碩士論文。
- 包玉海、寶力格 (1995)。改進 DHGM(2,2) 模型。內蒙古師範大學學報: 自然科學版，(2)，54-57。
- 駱尚廉 (1995)。環境數學。茂昌圖書有限公司，臺北。
- 吳漢雄、鄧聚龍、溫坤禮、夏郭賢 (1996)。灰色分析入門。高立圖書公司，臺北。
- 鄧聚龍、郭洪 (1996)。灰預測原理與應用。全華科技出版社，臺北。
- 長江防洪系統即時調度研究編輯委員會 (1997)。長江防洪系統即時調度研究。中國水利電力出版社，北京。
- 鄧聚龍 (1999)。灰色系統理論與應用。高立圖書有限公司，臺北。
- 盧瑞山、駱尚廉 (2000)。類神經網路於土壤復育工程之應用。財團法人中興工程顧問社。
- 國立臺灣大學生物環境系統工程學系 (2001)。水文設計應用手冊。經濟部水資源局。
- 張簡鳳蓮 (2002)。結合模糊理論於濁水溪流域逕流預報模式之研究。國立成功大學水利及海洋工程學系碩士論文。
- 樊新海、苗卿敏、王華民 (2003)。灰色預測 GM(1,1) 模型及其改進與應用。裝



甲兵工程學院學報，(2)，21-23。

張逸凡 (2005)。支撐向量機在即時河川水位預報之應用。成功大學水利及海洋工程學系碩士論文。

陳磊、張土喬 (2006)。基於貝葉斯最小二乘支持向量機的時用水量預測模型。天津大學學報(09)，1037-1042。

李明軒 (2008)。支撐向量機與模糊推論於流量預報即時誤差修正之研究。成功大學水利及海洋工程學系碩士論文。

劉鑽塗 (2009)。利用機器學習修正 QPESUMS 雷達估計降雨，成功大學水利及海洋工程學系碩士論文。

劉新立譯(2008)。初級統計學。清華大學出版社，北京。Triola, M. F.(1986).Elementary statistics, Benjamin-Cummings Publishing Company.

國家標準 GB/T 22482-2008 (2008)。水文情報預報規範。中華人民共和國水利局。

戴凌全、戴會超、蔣定國、李華、陳小燕 (2010)。基於最小二乘法的河流水位流量關係曲線推算。人民黃河，32(9)，37-39。

鄭又瑜、張良正 (2011)。曾文水庫即時最佳防洪操作之研究。國立交通大學土木工程研究所碩士論文。

中央氣象局 (2015)。1897-2014 年侵臺颱風路徑分類統計。

(<http://farmer.iyard.org/cwb/cwb.htm>)

第五河川局 (2017)。朴子溪環境生態簡介。水利署電子報，0238。

閔英男 (2018)。基於灰色理論與機器學習的黑河地下水水位預測研究。蘭州大學。

霍文博、朱躍龍、李致家、馮鈞、周莉、孔俊 (2018)。新安江模型和支持向量機模型即時洪水預報應用比較。河海大學學報自然科學版，v.46(04)，283-289。



陳昆懋(2019)。國軍救災傳播策略之研究—以 2018 年花蓮 0206 震災與南臺灣 823 水災為例。國防大學政治作戰學院新聞碩士班，碩士論文。

第五河川局 (2019)。局務介紹：朴子溪水系。

(<https://www.wra05.gov.tw/13615/13616/13618/13624/16544>)

ApacheCN (2019)，scikit-learn (sklearn) 官方文檔中文版 0.21.3。

(<https://github.com/apache/sklearn-doc-zh>)

段生月、王長坤、張柳豔 (2019)。基於正則化 GRU 模型的洪水預測。電腦系統應用，28(5)，196-201。

鄭文吉 (2020)。中央氣象局自動氣象站觀測資料彙整。

(<http://farmer.iyard.org/cwb/cwb.htm>)

感測資料-水利署河川水位資料。民生公共物聯網資料服務平臺。

([https://ci.taiwan.gov.tw/dsp/environmental\\_water\\_wra.aspx](https://ci.taiwan.gov.tw/dsp/environmental_water_wra.aspx))

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, 17(1), 113-126.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

David Cournapeau, (2007). scikit-learn 0.23. (<https://scikit-learn.org/stable/>)

Glantz, S. A., Slinker, B. K., & Neilands, T. B. (1990). *Primer of applied regression and analysis of variance* (Vol. 309). New York: McGraw-Hill.

Han, D., Chan, L., & Zhu, N. (2007). Flood forecasting using support vector machines. *Journal of hydroinformatics*, 9(4), 267-276.

Hosseini, S. M., & Mahjouri, N. (2016). Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling. *Applied Soft Computing*, 38, 329-345.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4), 679-688.

Nevo, S., Anisimov, V., Elidan, G., El-Yaniv, R., Giencke, P., Gigi, Y., ... & Tirumali, A. (2019). ML for flood forecasting at scale. arXiv preprint arXiv:1901.09583.

Python Software Foundation,(2001). Python 3.7.7 (<https://www.python.org/>)

Sharifi Garmdareh, E., Vafakhah, M., & Eslamian, S. S. (2018). Regional flood frequency analysis using support vector regression in arid and semi-arid regions of Iran. *Hydrological sciences journal*, 63(3), 426-440.

Torres, R. A. (2013). Prediction of NBA games based on Machine Learning Methods. University of Wisconsin, Madison.

Vapnik, V., Golowich, S. E., & Smola, A. J. (1997). Support vector method for function approximation, regression estimation and signal processing. In *Advances in neural information processing systems* (pp. 281-287).

Vapnik, V. N. (1998). *Statistical Learning Theory*, Wiley, New York.

Wu, C. H., Ho, J. M., & Lee, D. T. (2004). Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems*, 5(4), 276-281.