應用高斯過程迴歸於臺中港風速推算

江文山[1] 李俊穎[2] 許義宏[3] 黃國展[4] 賴志峰[5]

[1]國立成功大學水工試驗所研究員
[2]交通部運輸研究所運輸技術研究中心副研究員
[3]交通部運輸研究所運輸技術研究中心副研究員
[4]國立成功大學水工試驗所技術師
[5]國立成功大學水工試驗所技術師

摘要

本文根據臺中港內 4 個風速測站觀測資料，建構測站風速之間高斯過程迴歸模型，並進而推算測站風速應用於測站資料補遺、異常現象判斷。研究成果顯示就推算臺中港南堤紅燈塔附近的風速而言，若僅考慮 6 級風以上的情形，則單以北堤堤頭綠燈塔測站的風速資料，進行高斯過程迴歸模型建構，可以最有效率的方式得到良好的預測結果。然風向基本集中在偏東北與西南向，其他風向幾乎未發生，因此應用上受到限制。當風速降低到 3 級風甚至 1 級風條件，風向涵蓋範圍逐漸擴大到全部領域，但是預測效率顯著降低，同時在西北與東南風向情況預測結果誤差較大，主要是因該風向發生機率較低，資料數量少，導致模型學習不足，其次是因測站所在位置受到結構物影響，導致局部區域出現不穩定的風場。

一、前言

臺中港位於臺灣中部，地理位置優越，靠近臺灣的工業中心以及臺灣海峽，使其成為重要的海上貨物轉運點。臺中港擁有深水港口，能容納大型貨船和貨櫃船隻，這使其成為進口和出口貨物的主要中轉站，以及國際貨運的重要門戶。同時具有現代化的貨物處理設施，包括大型貨櫃碼頭、貨櫃堆場、貨物倉儲和物流中心，能夠高效地處理各種貨物，包括原材料、成品和消費品，因此，對臺灣經濟的多元性和國際貿易的發展都具有關鍵作用。臺中港不僅提供海運服務，還與臺灣內陸地區有優越的聯繫，通過鐵路、公路和管道等多種運輸方式，可將貨物快速送達臺灣各地，促進國內和國際貿易的流通。基於臺中港營運管理與航運安全考量，交通部運輸研究所運輸技術研究中心(以下簡稱運技中心)於臺中港建置包括風力(7 站)、潮流(2 站)、風波流(2 站)與靜穩度(2 站)等監測站，並透過即時傳輸將相關的環境參數提供使用者參考運用。

風力觀測資料因種種因素可能存在錯誤或是偏差，必須經過品管確保數據品質，品管程序通常基於合理性、連續性與關聯性等一系列原則性邏輯，實務上應用時必須定義相關門檻，而門檻的決定並無統一的絕對值，因此，品管結果會因門檻的設定存在灰色地帶，必須借助其他參數或是不同測站的數據進行交叉分析，以期進一步確認數據的品質。本研究嘗試針對強風事件發展風場變動趨勢分析。選定特定目標站，設定一定強度的風速數據為目標資料，選取同時

間同港口其他站的數據，進行高斯過程迴歸(GPR, Gaussian Process Regression)分析，得到目標數據與相關數據的 GRP 模型，並以該模型推算新的觀測數據，若推算結果偏離實測太多，則顯示根據過去經驗所建立的模型與現況不符，因此認定該資料為異常，藉此協助判斷資料的異常。異常發生可能原因為觀測系統特性改變或是環境變遷導致數據偏移，期望藉此可以協助判斷觀測系統的可靠性。

高斯過程迴歸，普遍應用在各領域的科學研究。時間序列預測是指對未來時間點的數據進行預測，它在許多領域中都有應用，例如經濟學、氣象學和交通運輸等。高斯過程迴歸可以用於時間序列預測，其中誤差項的分佈假設為高斯分佈。將高斯過程迴歸應用於時間序列預測可以提高模型的精度和準確性。例如，應用基於高斯過程迴歸的模型於預測貨幣匯率變化，該模型在實證結果中取得了很好的效果。其次，高斯過程迴歸可以用於模式識別，即識別數據中的類別或分組。這種方法被廣泛用於圖像識別、語音識別和自然語言處理等領域。在模式識別中，高斯過程迴歸可以用於分類問題，其中誤差項的分佈假設為高斯分佈。高斯過程迴歸在模式識別中的應用已經被證明是一種有效的方法，例如，在人臉識別方面，高斯過程迴歸被用於人臉識別的特徵提取和識別。再者，醫學影像處理亦是一個非常重要的應用領域，其中高斯過程迴歸可以用於分割、配準和分類等問題。高斯過程迴歸可以用於醫學影像中的不同問題，例如，基於高斯過程迴歸的方法可以用於改進胸部影像的分割。高斯過程迴歸可以用於風速預測，該模型可以根據過去的風速和氣象條件預測未來的風速，Sloughter et al.(2010)基於貝葉斯模型平均法進行風速機率預報，Thorarinsdottir et al.(2012)假設風速機率分佈為高斯分佈，利用高斯過程迴歸法進行強風發生機率預報。胡波等人(2019)應用高斯過程迴歸方法探討浙江沿海冬春季陣風預報。關於臺中港的風場研究，賴等人(2021)建立風速品管與補遺的程序，江等人(2021) 蒐集運技中心的風力資料經品管分析後，針對強風特性進行研究，並結合離散小波轉換與深度學習技術，以探討定點風力預報之基礎模型。藉由對風速資料的特性分析與掌握，進而嘗試建立風速短期預報的技術。本研究則嘗試利用高斯過程迴歸，建構臺中港內各風速測站之間的推算模型，藉以做為判斷各測站出現異常風速狀況之參考，以及資料缺漏時的補遺應用。

二、高斯過程迴歸

高斯過程迴歸是使用高斯過程（Gaussian Process, GP）對數據進行迴歸分析的非參數機器學習模型（non-parameter model）。GPR 機器學習方法是由劍橋大學學者 Carl E. Rasmussen 和愛丁堡大學學者 Christopher K. I. Williams，二者在 1996 年發表的研究，其按函數空間觀點提出 GPR 的求解系統並討論 GPR 超參數的極大似然估計和蒙特卡羅方法求解。GPR 的模型假設包括雜訊（迴歸殘差）和選定核心函數高斯過程先驗兩部分，利用貝葉式推理（Bayesian inference）進行模型參數的求解。若不限制核心函數的形式，GPR 在理論上是緊緻空間（compact space）內任意連續函數的通用近似（universal approximator）。此外，在假設概似函數為正態分佈時，GPR 可提供預測結果的後驗，且該後驗具有解析形式。因此，GPR 是一個具有泛用性和可解析性的

概率模型。基於高斯過程及其核函數所具有的便利性質，GPR 在時間序列分析、圖像處理和自動控制等領域的問題中具實務應用價值。一般而言，GPR 的計算量較大相對效率較低，通常被用於低維和小樣本的迴歸問題。

關於 GPR 的早期研究大致可分為 3 部分，在時間序列分析中， Robert Wiener 在 1940 年代提出了估計 0 均值平穩高斯過程信號的濾波技術 。隨後出現的卡爾曼濾波（Kalman filter）提供了估計高斯隱變量（latent variable）的有效方法。在地理空間統計學領域，1963 年提出的克里金法（Kriging）為 GPR 首次應用在隨機場的非參數迴歸，之後在 1990 年代提出的貝式克里金方法提供了與 GPR 類似的高斯隨機場估計方式。在機器學習方面，包含無限節點的貝氏神經網絡（Bayesian neural network）和各類核學習（kernel learning）方法例如核嶺迴歸（kernel ridge regression）為 GPR 的核函數超參數求解。GPR 假設數據樣本是高斯過程的採樣，因此其估計結果與核函數有密切聯繫。GPR 中核函數的實際意義為利用協方差函數（covariance function），描述數據樣本間的相關性，因此，其不是通過方法簡化計算的手段，而是模型假設的一部分。GPR 實務應用的關鍵為核心函數的選擇，核心函數的選擇要求須滿足 Mercer 定理(Mercer's theorem)，即核心函數在樣本空間內的任意核矩陣為半正定矩陣（semi-positive definite）。假設 GPR 迴歸的數據符合穩態(stationary)高斯過程，可用的核心函數包括徑向基函數核（RBF kernel）、馬頓核（Matérn kernel）、指數函數核（exponential kernel）、二次有理函數核（rational quadratic kernel, RQ kernel）等。另外，GPR 迴歸也可應用在數據為非穩態高斯過程，此時常見的核心函數選擇為週期核（periodic kernel）與多項式核函數（polynominal kernel），二者分別賦予高斯過程週期性和旋轉不變特性（rotation invariant）。

高斯過程迴歸原理係從函數空間角度出發，定義一個高斯過程(GP)來描述函數分布，直接在函數空間進行貝葉式推理。GP 是任意有限個隨機變量均具有聯合高斯分布的集合，其性質完全由均值函數和協方差函數確定，即：

$$\begin{gathered} m(x) = E[f(x)] \\ k(x,x') = \mathrm{E}[(f(x) - m(x))(f(x') - m(x'))] \end{gathered} \tag{1}$$

其中，x，$x' \in R^d$ 為任意隨機變量，因此，GP 可定義為$f(x) \sim \mathrm{GP}(m(x), k(x,x'))$。為了符號上的簡潔，通常針對數據做預處理，使其均值函數等於 0。

對於迴歸問題，考慮如下模型：

$$y = f(x) + \varepsilon \tag{2}$$

其中：x 為輸入向量，f 為函數值，y 為考慮誤差的觀測值。進一步假設誤差$\varepsilon \sim \mathrm{N}(0, \sigma_n^2)$，因此，可以得到觀測值 y 的事前分布為：

$$\mathrm{y} \sim \mathrm{N}(0, \mathrm{K}(\mathrm{X},\mathrm{X}) + \sigma_n^2 I_n) \tag{3}$$

而觀察值 y 和預測值 f*的聯合先驗分布為：

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I_n & K(X,x_*) \\ K(x_*,X) & k(x_*,x_*) \end{bmatrix}\right) \tag{4}$$

其中：$K(X,X)=K_n=(K_{ij})$為$n\times n$階對稱正定的斜方差矩陣，矩陣元素$k_y=k(X_i,X_j)$用來度量X_i和X_j之間的相關性；$K(X,x_*)=K(x_*,X)^T$為測試點x_*與訓練集的輸入值 X 之間的$n\times 1$階協方差矩陣；$k(x_*,x_*)$為測試點x_*自身的協方差，I_n為 n 維單位矩陣。

由此可以計算出預測值$f*$的事後分布為：

$$f*\left|X,y,x_*\sim N(\bar{f}*,cov(f*))\right. \tag{5}$$

其中：

$$\bar{f}*=\ K(x_*,X)[K(X,X)+\sigma_n^2 I_n]^{-1}y \tag{6}$$

$$cov(f*)=k(x_*,x_*)-K(x_*,X)\times[K(X,X)+\sigma_n^2 I_n]^{-1}\,K(X,x_*) \tag{7}$$

上述可見高斯過程迴歸模型是一種概率模型，$\hat{\mu}_x=\bar{f}*,\hat{\sigma}_{f_i}^2=cov(f*)$即為測試值$x_*$，對應預測值$f*$的均值和方差。

高斯過程迴歸可以選擇不同協方差函數，普遍使用的有平方指數協方差，即：

$$k(x,x')=\sigma_f^2\exp(-\tfrac{1}{2}(x-x')^{T}M^{-1}(x-x')) \tag{8}$$

其中：$\mathrm{M}=\mathrm{diag}(\tau^2),\tau$為方差尺度，$\sigma_f^2$為信號方差。參數集合$\theta=\{M,\sigma_f^2,\ \sigma_n^2\}$即為超參數，一般通過最大概似法求得。首先建立訓練樣本條件概率的負對數概似函數$\mathrm{L}(\theta)=-\mathrm{logp}(y|X,\theta)$，並令其對超參數$\theta$求偏導數；然後採用牛頓法等方法對偏導數進行最小化以得到超參數的最佳解。其中負對數概似函數$\mathrm{L}(\theta)$及其關於超參數θ的偏導數形式如下所示：

$$\mathrm{L}(\theta)=\tfrac{1}{2}y^TC^{-1}+\tfrac{1}{2}log|C|+\tfrac{n}{2}\log(2\pi) \tag{9}$$

$$\frac{\partial L(\theta)}{\partial\theta_i}=\tfrac{1}{2}tr((\alpha\alpha^T-C^{-1})\frac{\partial C^n}{\partial\theta_i}) \tag{10}$$

其中$\mathrm{C}=K_n+\sigma_n^2 I_n,\alpha=(K+\sigma_n^2 I_n)^{-1}y=C^{-1}y$，獲得最佳參數後，利用式(6)與式(7)即可得到預測點x_*對應的預測值$f*$及其方差$\hat{\sigma}_{f_i}^2$。

三、測站介紹

運輸技術研究中心與臺灣港務股份有限公司在臺中港設置的 5 個風速測站，其背景資料與相關位置如表 1 及圖 1 所示。使用觀測儀器為超音波的三維風速計，每秒測量一筆資料，每 10 分鐘統計輸出平均風速與陣風風速，其中測站 TCWD08M10 因資料時序不規則，未予納入後續分析。

表 1 風速測站背景資料

測站代碼	座標(N)	座標(E)	地點說明	主副站	儀器類別
TCWD02M10	24°17'58.61"	120°29'11.82"	北堤綠燈塔	主站	WindSonic
TCWD03M10	24°18'18.66"	120°31'52.44"	防風林	附屬站	WindSonic
TCWD07M10	24°16'27.81"	120°30'58.82"	31 號碼頭	附屬站	WindSonic
TCWD08M10	24°14'23.28"	120°28'18.45"	工專二	附屬站	SONIC SA-20
TCWD14M10	24°17'24.6"	120°30'02.1"	南堤紅燈塔	附屬站	WindSonic

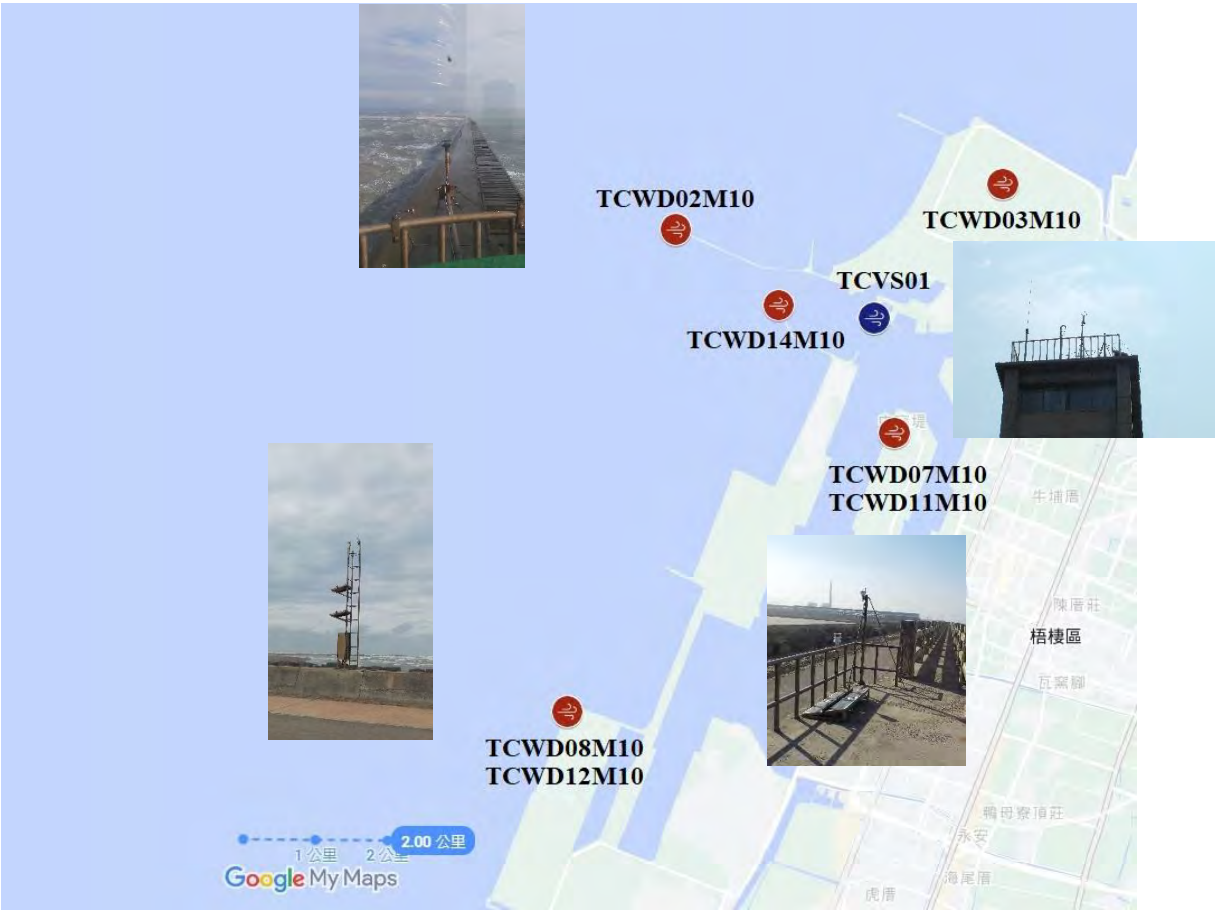

圖 1 運輸技術研究中心與臺灣港務股份有限公司在臺中港設置之風速測站位置

四、資料前處理與分析

觀測資料在進行模式建構前須經過相關的品管程序，資料品管規範有區域性特性，本文主要以往昔臺灣實測資料分析經驗為依據，配合參考國內外相關研究單位的品管藍圖，以合理性、連續性以及關聯性三個原則進行自動化品管檢測，並採用統計理論建立上述原則的檢測標準(如圖 2)。首先將原始風速時序資料轉換為東西向與南北向之風速分量，各別進行資料品管。

(1)合理性品管

合理性檢定為品管中最基本的檢驗工作，一般而言，合理性品管分為兩種限制，第一是指觀測數據不超過量測儀器的可測範圍;第二則是觀測數據不超過量測地點的氣象臨界條件，亦即物理上的限制，一般而言，10m 高風速介於 0-60 m/s。

(2)連續性品管

自然界中大部份物理量的變化是循序漸進的，連續性的品管基於此假設，認為各種觀測物理量在時間及空間上都有連續變化的存在，也就是說其時空的變化量應落在合理的範圍內測，如果在時間上的改變量超過某一限度，或是在空間上相鄰測點的資料無法以物理觀點解釋，都視為可疑資料。連續性品管分為時間連續性與空間連續性兩種品管。時間連續性是指物理量在某段時間內的變化量應小於某一容許值，NDBC 對時間連續性的檢測採 $\sigma_T=0.58\delta\sqrt{\Delta T}$ 公式來推求容許變動值，式中 σ_T 為在時間間距ΔT 時所允許的最大變動量，δ 為一無因次變量，其值隨觀測物理量不同而變。同樣地，在某適當範圍內觀測的資料必定存在空間演變上的連續特性，進行此檢驗時，應選擇空間條件相近的臨近測站，此檢驗可降低隨機取樣造成之隨機誤差(Tucker, 1993)。

(3)關聯性品管

關聯性檢驗是檢驗所量測之資料與其他相關物理量之間關係是否有異常現象，海氣象特性往往受到外在環境的影響而改變(例如波浪與風之關係)，利用不同物理量間之關連性，可互相檢驗資料的品質。此外，同一物理量使用不同測量儀器或是相同儀器同時量測，除了可明顯的比較資料正確性，對於儀器老化或毀損，更可顯示出其偏差量。本文風速資料之品管係根據合理性、連續性與關聯性 3 個原則，其中合理性主要依儀器量測範圍與物理上的限制而訂，連續性則是依相鄰資料應呈平緩變化為原則，依據機率理論而訂立通過(或異常)門檻，近年因氣候變遷而有短時間之極端事件發生，因此，須比對及參考鄰近測站資料建立相關式，依信賴區間訂立通過(或異常)之門檻，避免前述極端資料誤刪並做為相互補遺之依據。另資料一段時間無變化，通常是儀器故障、損壞或遭覆蓋等現象，可於輸出檔中去除(或標註)之，最終輔以人工檢視方式做進一步之確認。詳細步驟說明如下：

原始資料可能因現場量測與通訊或後端資料庫異常等因素，而使資料呈時間非連續狀態，對於一些分析(如頻譜)與預測所需，勢必無法直接使用，因此，在進行品管作業前先對資料做等時距處理，由於自動化程式之開發係以 Matlab 作為開發環境，因此無資料者以 NaN(Not a Number)填入。

其次進行合理性檢定，通常合理性品管分為兩種限制，第一是指觀測數據不超過量測儀器的可測範圍;第二則是觀測數據不超過量測地點的氣象臨界條件，亦即物理上的限制，一般而言，

10m 高風速介於 0-60 m/s，而風向資料以 0~360 度為有效數據。因此在自動化程式中，將非介於上述範圍之數據輸出為 NaN。

扣除上述明顯錯誤值後，進行資料之連續性檢定，其原理是前後時間資料不應有突然之變化。本案利用時間非連續之移動平均進行此異常值檢定，採用的步驟為：計算原始值與 2 小時移動平均值之標準偏差，大於 3 倍處則視為離群異常值，即假定差值呈常態分布兩側機率各佔 0.15%者。由於異常值可能是短時間之劇烈天候變化，附近測站有可能同時測得，因此留待關聯性檢定時，做進一步之處理。

關聯性檢定之原理為測站間資料如有高度相關性，可做為修正短期天候變化而未通過連續性品管之資料改列有效數據。步驟首先為進行待品管測站資料與同一港口所有風速站同時期資料之迴歸分析，由最高相關係數且呈高度正相關(相關係數大於 0.7)者才進行關聯性檢定，由線性迴歸所得曲線並計算其信賴區間，假定於 95%信賴區間範圍內者代表兩站同時測得有高度可靠性，因此，可將上述未通過連續性品管之資料改列有效數據。至於超出 95%信賴區間代表皆未通過上述兩個檢定，以 NaN 取代原數據。

完成上述自動化品管後，以人工檢視做進一步品管，方法是將原始資料與自動品管後資料進行時序比對，觀察是否有誤刪或異常仍保留之資料，查詢當時天候與作業紀錄，以決定是否修正。

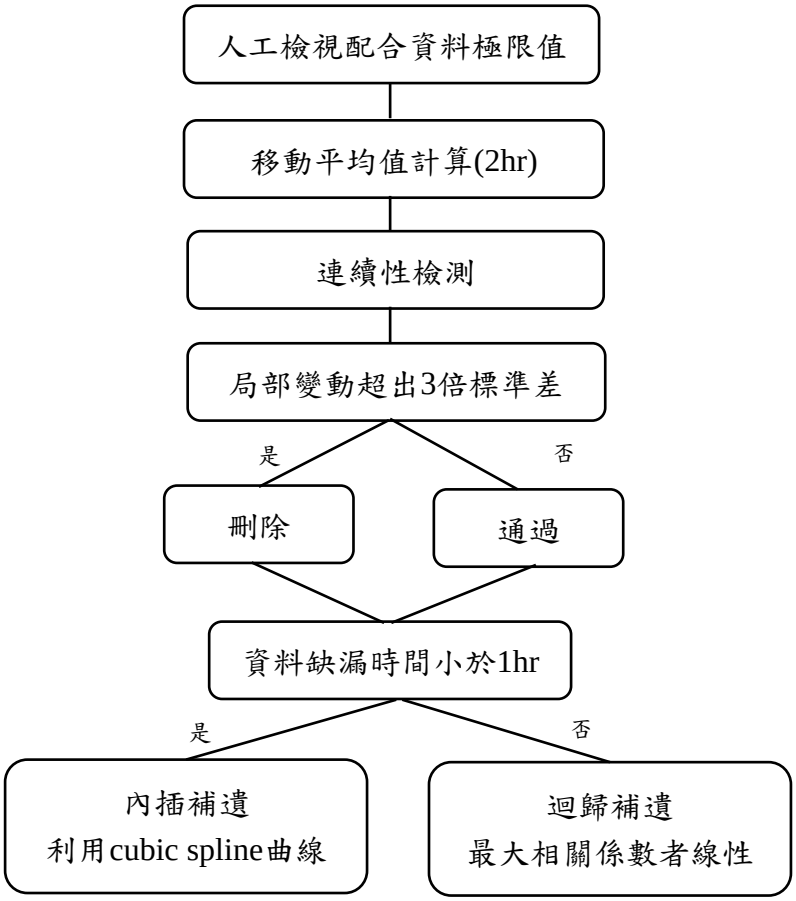

圖 2 風力資料品管及補遺作業標準作業程序之流程圖

五、風速基本特性

根據臺中港 4 個風速測站 2020 年 1 月至 2023 年 08 月的資料篩選風速超過 6 級風的資料進行分析，就北堤堤頭 TCWD02M10 測站，可見強風主要以偏北風時期發生機率最高，同時風速亦較強，部分時間亦會發生西南向強風，但是風速不若北風時強，詳如圖 3。位於港域北側與中心區域的內陸測站 TCWD03M10、TCWD07M10 與南側靠近外海的測站 TCWD08M10，風速超過 6 級風的情況都是發生在東北風時期。分析各測站強風發生機率在一天當中的分佈情況如圖 4，可見靠近內陸的測站 TCWD03M10 與 TCWD07M10 強風發生機率較集中，以及強風主要發生在下午時間，而相對靠近外海的測站 TCWD02M10 與 TCWD08M10 強風發生機率較均勻。另統計每場超過 6 級風的強風事件的延時與事件的平均風速繪如圖 5，顯示每場強風事件隨著發生期間事件平均風速的增強，延時亦增加，然增加的趨勢隨平均風速的增強呈現逐漸緩和。靠近內陸的測站 TCWD03M10 與 TCWD07M10 強風相對較弱，而相對靠近外海的測站 TCWD02M10 與 TCWD08M10 強風則是較強。測站之間的強風速差異最大達 10m/s 以上。為了解各測站出現強風事件時的空間尺度規模，設定同港口全部測站都發生超過 6 級風的門檻進行強風事件分析，發現事件數極其有限，顯示異常強風的時空尺度通常不大，只出現在短時間而且局部區域，並未涵蓋整個港區。以臺中港為例檢視測站 TCWD02、03、07 與 08 為例，只在 2021 年 10 月 11 日曾經發生涵蓋全港區的強風事件。當日強烈熱帶風暴(日本氣象廳)襲侵菲律賓造成嚴重損失。

統計各測站發生超過 6 級風的次數與總時數列，結果如表 2 所示，可見測站 TCWD02M10 最常發生強風事件，持續的時間最久，其次為 TCWD07M10，呈現越往內陸發生次數越少的趨勢。

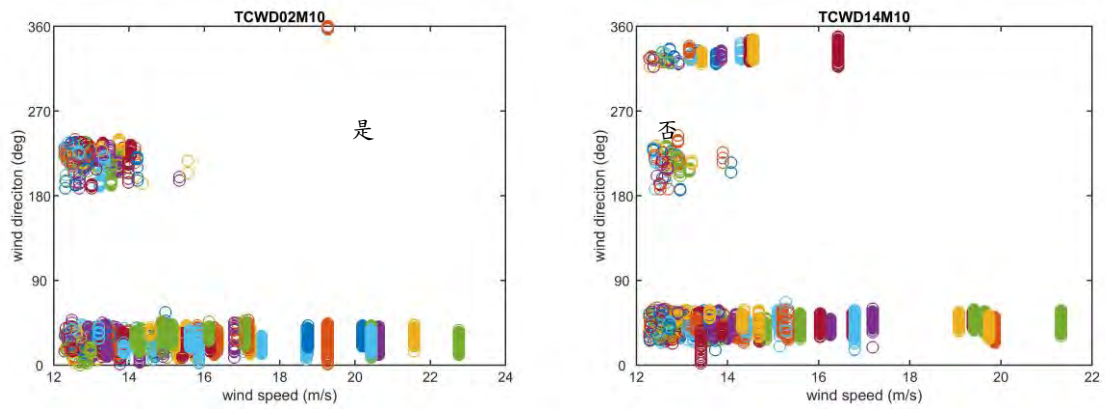

圖 3 臺中港測站 TCWD02M10 與 TCWD14M10 超過 6 級風的事件其風速與風向分布

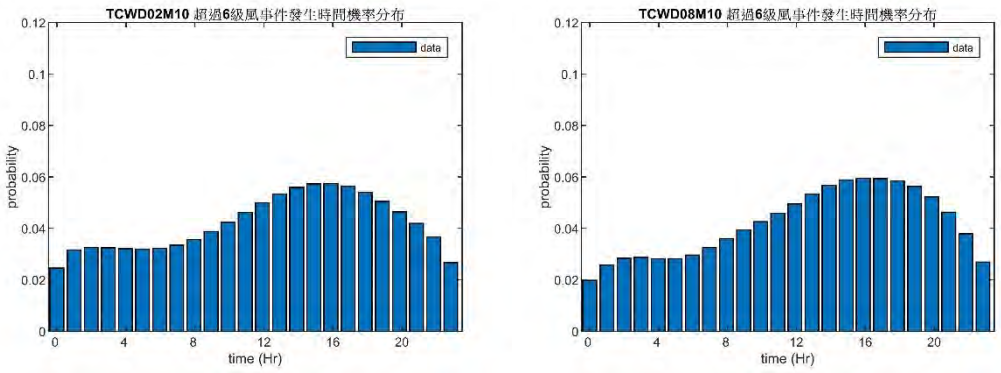

圖 4 臺中港測站 TCWD02M10 與 TCWD14M10 強風發生機率在一天當中的分佈

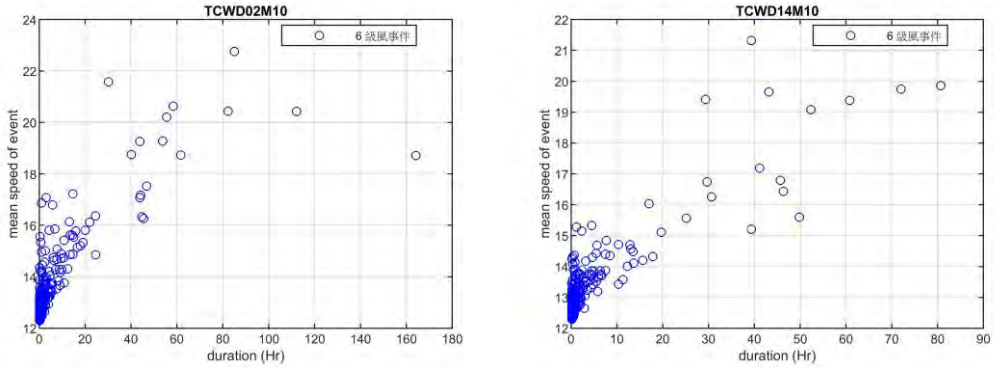

圖 5 臺中港測站 TCWD02M10 與 TCWD14M10 超過 6 級風事件延時與事件平均風速

表 2 臺中港各風速測站超過 6 級風事件的統計

測站	期間	事件定義	發生事件總數	發生總時數(小時)
臺中港_TCWD02M10	202301_202308	超過6級風事件	306	1688
臺中港_TCWD03M10	202301_202308		138	192
臺中港_TCWD07M10	202301_202308		267	408
臺中港_TCWD14M10	202301_202308		340	1160

六、模擬參數與條件

針對強風事件規劃發展風場變動趨勢分析，選定南內堤紅燈塔測站 TCWD14M10 為目標站，設定超過 6 級風的數據為目標資料，根據選定風速資料之間相關性的分析結果如表 3，顯示南內堤紅燈塔測站與其他 3 站的相關性以北堤堤頭測站最高為 0.79，其次為防風林與 31 號碼頭之相關係數 0.68，因此，納入此 3 站的風場資料作為後續分析的輸入因子，選取同時間這些站的數據，就如前面基本資料分析所顯示的北堤堤頭測站的風速相對較高，而各測站之間風速有近似的相關的變動趨勢。然仔細檢視會發現在某一些時段特定測站的變動趨勢會與其他測站有

局部性的差異。顯示雖然測站的位置都是固定，但是仍然可能因為環境參數的改變如風向，而使得測站之間的風速變動趨勢改變。

選定測站 TCWD14M10 超過 12m/s 的資料與對應之其他站的風速與風向資料(如圖 6)，超過設定門檻的強風條件主要發生在冬季期間，隨季節進到春季夏季發生次數逐漸降低，同時風速也降低。對應的風向主要都是在偏東北風情況，少數為偏西南風情形。

臺中港風速測站基於測站 TCWD14M10 超過 9m/s 的資料與對應之其他站的風速之發生機率分布與各測站彼此之間的散佈關係顯示各測站的風速分布存在顯著差異，資料的散佈分布趨勢亦不同，各測站之間的風速相關係數，就測站 TCWD14M10 而言，與之相關係數最高的為測站 TCWD02M10 達到 0.79 以上，至於與其他 3 個測站之相關系數都小於 0.15，除了是測站位置較為接近的因素以外，各測站周遭環境的特性亦有直接的影響。

分析臺中港全部測站各風向之風速發生機率如圖 7，可見北北東~東北風是發生機率最高的風向，其次為南南西~西北西向，再者為北北西~北向，其餘的風向發生的機率都非常低。進一步分析各測站平均風速與風向之關係如圖 8，可見平均風速的分布各站之間的強弱隨著風向改變而有不同的趨勢，普遍是以北防波堤堤頭的測站為最高，其次為測站 TCWD14M10，再者為測站 TCWD03M10 與測站 TCWD107M10。風向介於北到東北之間，以測站 TCWD02M10 風速最強，風向轉為東北至東之間時則以測站 TCWD14M10 風速最大；介於東至南的風向則又以 TCWD02M10 測站風速最強，同時測站 TCWD14M10 轉為最弱；南至西風的情況，測站 TCWD02M10 與 TCWD14M10 風速接近，大於其它測站；西至西北風情形，以測站 TCWD02M10 風速最弱。顯示各測站的風速除了受到所在地理位置偏向外海或者是往內陸的影響之外，也受到各測站局部周遭環境的影響，例如，防波堤或是建築結構物。至於各測站的極端高風速主要是來自偏北方向，其次則是來自南向。北防波堤堤頭測站的最大風速明顯高於其他 3 個測站，其相對位於最外海的位置，周遭空曠沒有遮蔽物，應是主要的原因。

設定測站 TCWD14M10 的風速為推算的目標，為了解不同預測因子對預測結果的影響，分析數種不同的情境如表 4，第一個情境是單純只考慮與測站 TCWD14M10 相關系數最高的測站 TCWD02M10 的風速為預測因子；第二個情境是在情境一之外預測因子再加上測站 TCWD02M10 的風向；之後情境則是逐漸將其他測站的風速與風向加入預測因子之中。第五個情境則是考慮將風速門檻降低為 4 級風條件約為 9m/s，以探討風速變動的影響。各情境都是選取資料中的 80%進行模型訓練，其餘 20%則是做為驗證，高斯過程迴歸為一機率模型，所以預測結果中，除平均值以外，另亦考慮信賴區間做為評估結果的參考。

表 3 測站相關係數表

測站代號	TCWD02M10	TCWD03M10	TCWD07M10	TCWD14M10
TCWD02M10	1.0000	0.0552	0.1640	0.7905
TCWD03M10	0.0552	1.0000	0.6794	0.1364
TCWD07M10	0.1640	0.6794	1.0000	0.0794
TCWD14M10	0.7905	0.1364	0.0794	1.0000

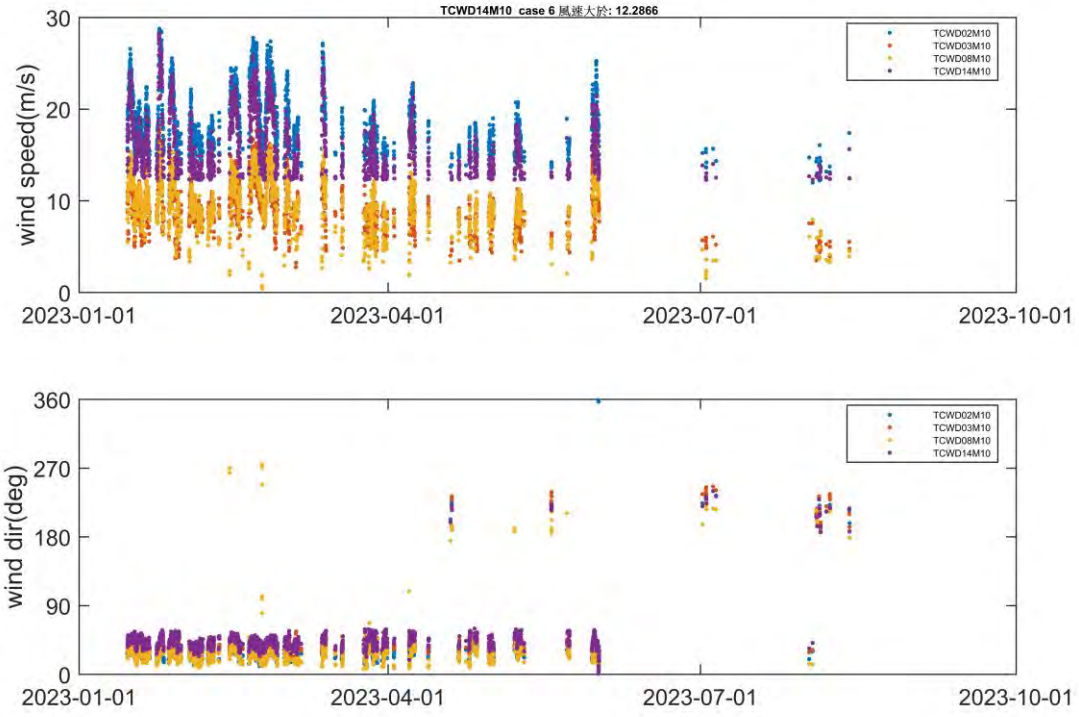

圖 6 TCWD14M10 超過 12m/s 的資料與對應之其他站的風速與風向資料

表 4 分析推算情境表

條件	輸入因子說明	資料期間	資料篩選	訓練資料
Case 6-1	測站 TCWD02M10 風速	1/1~8/31, 2023	風速大於 12m/s	資料總數的 80%
Case 6-2	測站 TCWD02M10 風速風向			
Case 6-4	測站 TCWD02M10 、TCWD03M10 風速風向			
Case 6-6	測站 TCWD02M10 、TCWD03M10、TCWD07M10 風速風向			
Case 4-6	測站 TCWD02M10 、TCWD03M10、TCWD07M10 風速風向		風速大於 6 m/s	

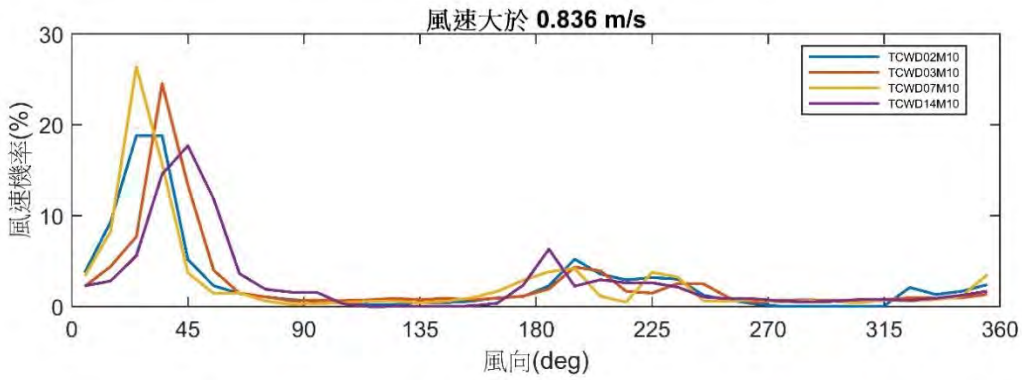

圖 7 臺中港內不同風速測站的各種風向的發生機率

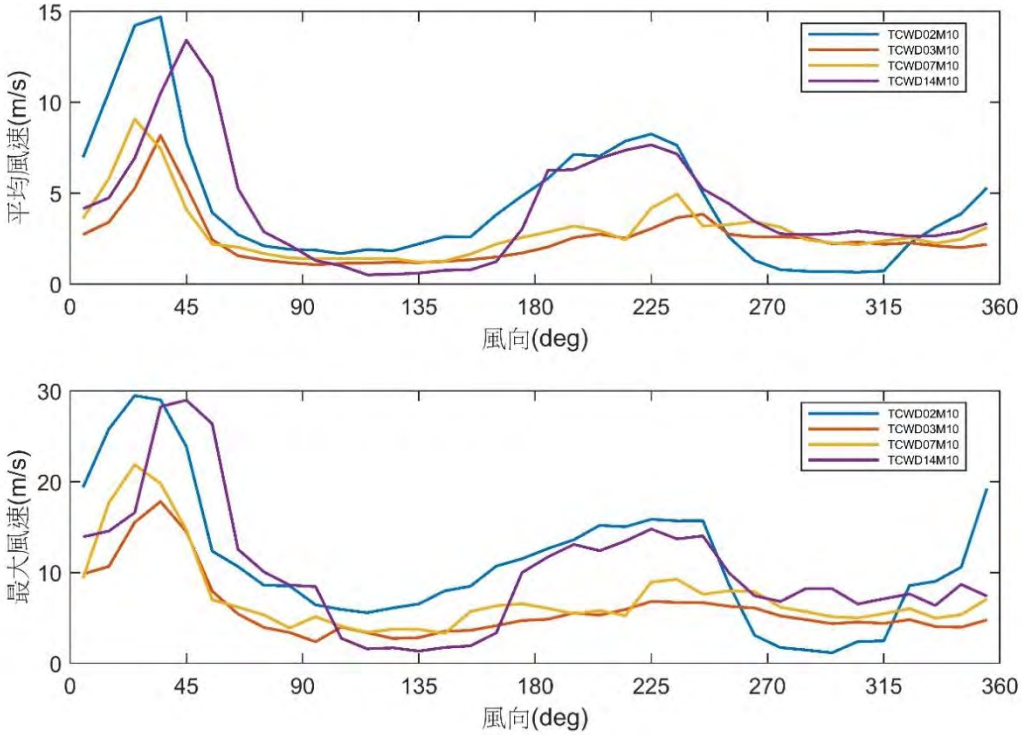

圖 8 臺中港內不同風速測站的平均風速隨風向的變化

七、模擬結果與討論

7.1 情境一: case 6-1

圖 9 為情境一預測與實測風速得比較，圖中灰色帶狀區域為 99%信賴區間的範圍，普遍的預測結果都在合理的區間範圍內，少數預測超出容許範圍。圖 10 情境一模擬結果誤差與風向統計，可見當風向方位角小於約 25 度的情況普遍預測的結果都是偏高，隨著風向轉變，當方位角介於約 25 度至 60 度時，預測的結果偏高與偏低的情況都有相當的數量，但是約略是以低估的情況較大，當方位角大於 180 度時，則預測的結果偏低佔絕大多數。亦即當大區域的風向改變，各測站局部範圍因環境條件不同，對風向改變的反應不同，進一步測站之間的相對關係隨之改

變，所以若未考慮風向的影響，會出現因風向改變導致的系統性誤差。至於絕對誤差在方位角小於約 25 度的情況呈現最大平均達到 2m/s，其他的方位角條件下平均誤差都在 1m/s 以下。

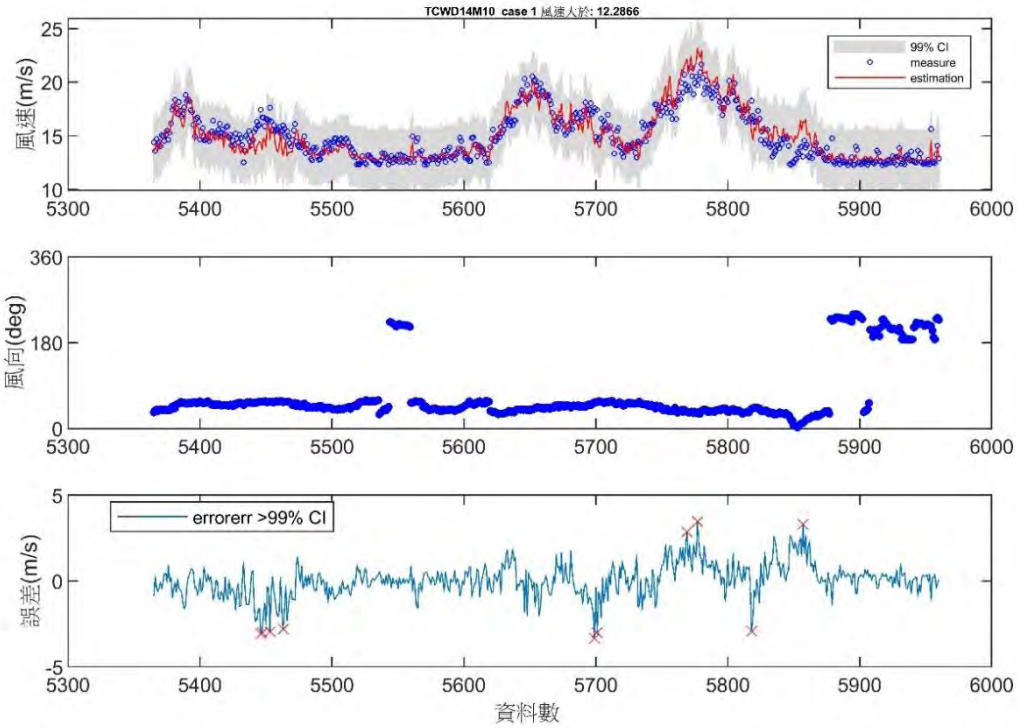

圖 9 情境一模擬與實測之比較

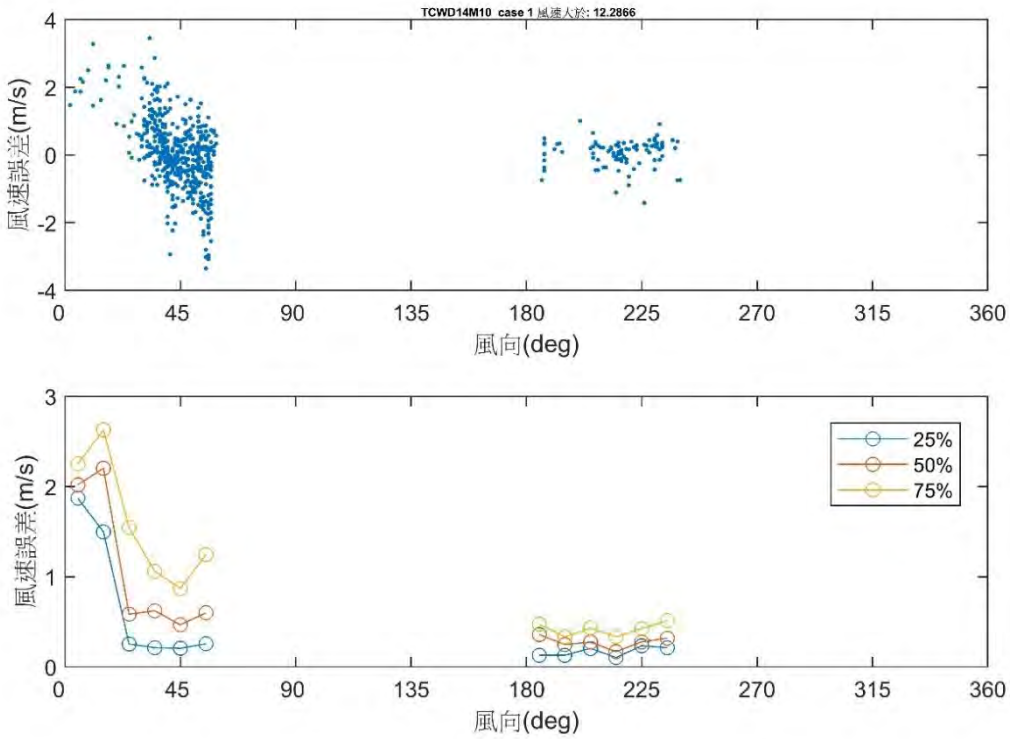

圖 10 情境一模擬結果誤差與風向統計

7.2 情境二: case 6-2

預測結果的信賴區間普遍相對較小，亦即不確定性降低，但是在少數幾個時間預測的信賴區間極度放大，誤差比較大，主要都是發生在風向為非東北風情況見圖 11。圖 12 情境二模擬結果誤差與風向統計，可見相對情境一預測結果存在普遍性地偏高或偏低的情況，當將風向納為預測因子之後獲得改善，而且誤差普遍都呈現比較小。但是在風向方位為北風與西南風的情況，因為發生機率低，選取資料的前 80%做為訓練資料，幾乎都未納入此種條件，導致其預測的結果呈現顯著偏高的現象，其平均誤差達 3m/s 比情境一都大。由此可見訓練資料必須廣泛的涵蓋各種風向條件，否則將資料進行越細部的分類分析時，可能在部分發生機率較低的條件造成更大的誤差。

7.3 情境三: case 6-4

納入測站 TCWD03M10 的風速與風向為預測因子後，依然發生情境二中出現的預測信賴區間極度放大的現象，以及風向方位為北風與西南風的情況，預測結果呈現顯著偏高的現象。

7.4 情境四: case 6-6

情境四與情境三的結果整體趨勢差異不明顯，惟預測誤差略為下降。顯示在風速設定為 6 級風以上為前提的情況下，引入測站 TCWD02M10 之外的鄰近測站資料加入為預測因子進行模型迴歸分析得到的結果，對整體助益不明顯。

7.5 情境 5: case 4-6

設定超過 4 級風資料為模擬目標的條件下，通過篩選的資料數超過 6 級風條件下的 2 倍以上，納入訓練資料的涵蓋條件範圍就相對大許多，迴歸模型亦得到相對較完整的訓練，因此，預測結果的誤差有顯著的改善，除了在風向為北向、南向與西向等情況，即使在一整年當中其發生機率依舊是非常的低，如圖 13 中所顯示的現象，此種狀況其預測結果相對其他風向還是偏大，整體而言，平均誤差幾乎都落在 1m/s 以下。

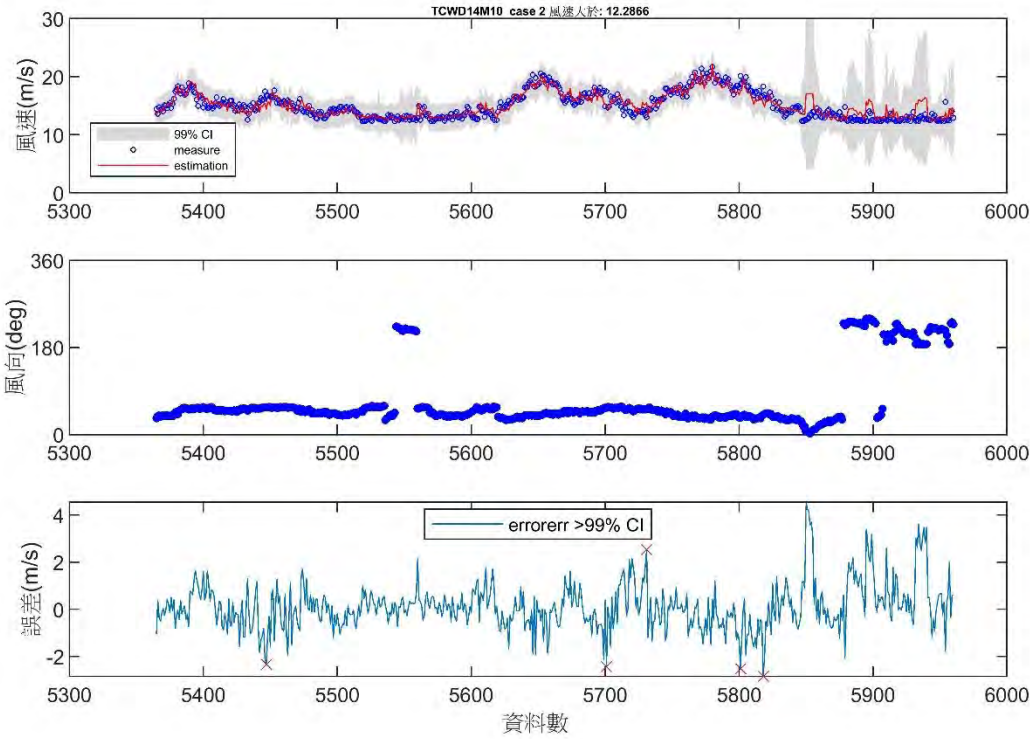

圖 11 情境二模擬與實測之比較

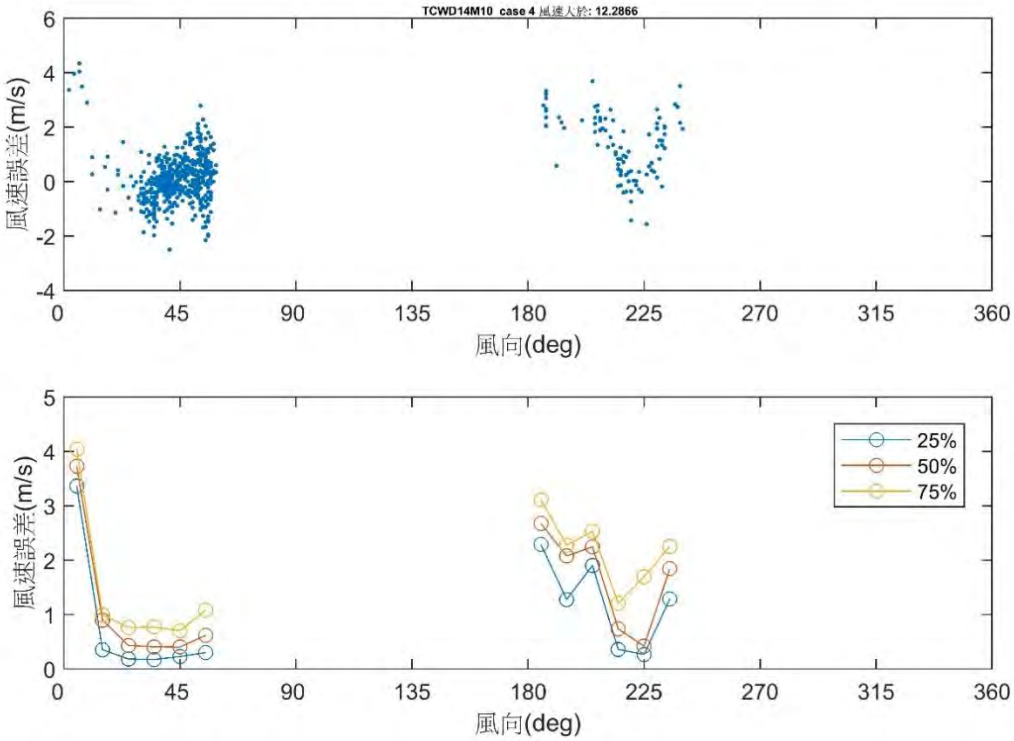

圖 12 情境二模擬結果誤差與風向統計

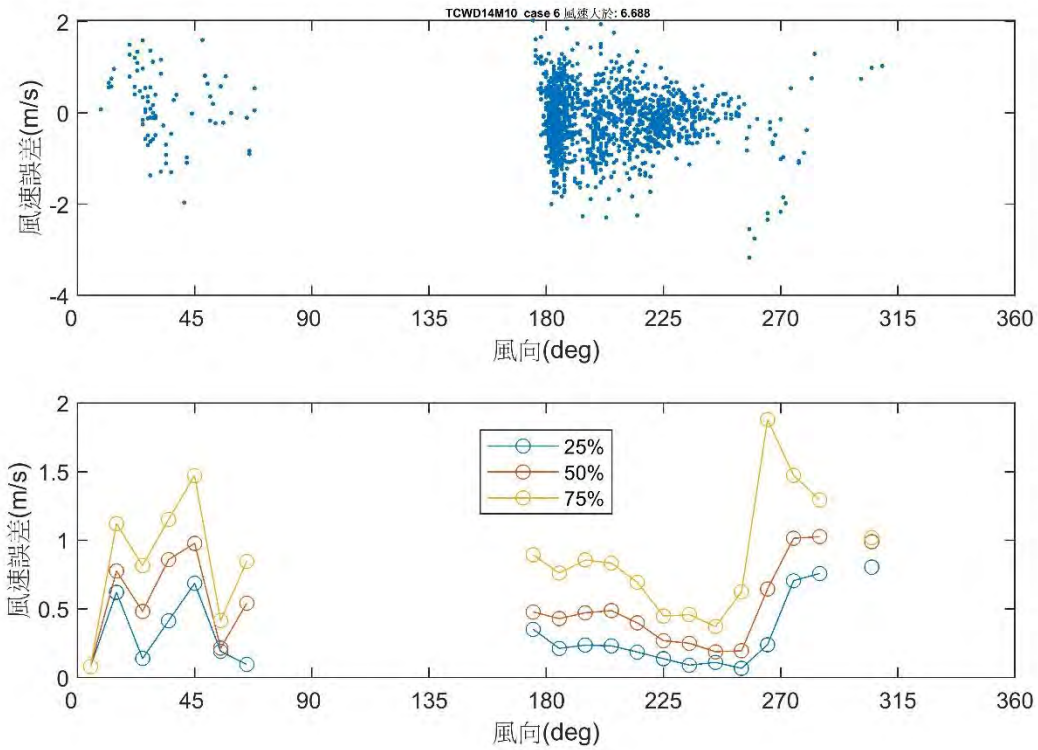

圖 13 情境五模擬結果誤差與風向統計

7.6 模擬結果整合評估

上述分析初始設定是針對 6 級強風以上的狀況，為了解不同風速強度下的現象，亦進行設定不同風速強度為門檻條件的模擬分析，並比較說明其差異。

綜整設定超過 6 級風資料為模擬目標的條件下，各情境模擬結果的誤差隨風向的變化如圖 14，考慮最快效率，僅以相關性最高的測站 TCWD02M10 並且不納入風向的影響進行模擬，整體而言可以得到比考慮其他因子的模擬結果更佳的預測，一方面是因為測站 TCWD02M10 與目標測站 TCWD14M10 相關性最高，尤其是在偏北風的情況，另一方面亦是因為越強風速的條件，其他風向發生次數非常低，模型學習的經驗不足，導致納入風向做為預測因子其預測結果反而較差。

設定超過 5 級風資料為模擬目標的條件下，若考慮最快效率，僅以相關性最高的測站 2 並且不納入風向的影響進行模擬，則除了北風情況平均誤差達到 2m/s 以外，其餘情況平均誤差都在 1m/s 以下。若同時再考慮測站 TCWD02M10 與測站 TCWD03M10 的風速風向影響進行模擬，可以得到最佳的效果，平均誤差在各風向都維持在 1m/s 以下。考慮全部 4 個測站的風速風向影響進行模擬，未必可以得到更好的結果，反而可能引入不同特性的因子，形成干擾導致結果未進一步改善。

設定超過 4 級風資料為模擬目標的條件下各情境模擬結果的誤差隨風向的變化如圖 15。通過篩選的資料數超過 6 級風條件下的 2 倍以上，若同時考慮納入超過兩個測站以上的影響，就各種不同的情況，模型普遍得到更充分的訓練與學習，普遍可以得到平均誤差小 1m/s 以下的結果。但是就發生的風向而言，在 4 級風以上的情況，風向主要還是集中在偏東北風與偏西南風相當的範圍內。

設定超過 3 級風資料為模擬目標的條件下，亦得到與以 4 級風為條件的結果類似。隨著設定風速條件越低，納入分析的資料越多，發生的情況亦趨複雜，同時一些很少發生的風向情況亦被納入分析，因此，其結果在整體平均誤差未必上升，但是部分發生機率少的現象其預測誤差偏大，尤其在特殊風向的情況更是明顯。不過在此條件下，風向的涵蓋範圍擴及西到北向，唯獨偏東南風向還是未被納入，亦即該風向並未發生 3 級風以上的條件。當設定 1 級風為條件時，各情境模擬結果的誤差隨風向的變化如圖 16，則幾乎所有的資料都納入高斯過程迴歸分析，此時所有的風向都涵蓋，但是在西北與東南向的誤差相對最大，尤其是單考慮以測站 TCWD02M10 資料為預測因子時，誤差是最大的。綜上，考慮 3 級風以上的情況時，以單一測站 TCWD02M10 的資料為預測因子，與考慮所有測站資料為預測因子的預測結果差異不顯著。若考慮所有各種風速的條件，則單以測站 TCWD02M10 的資料為預測因子，在偏西北與東南方向的預測誤差會明顯較考慮所有測站資料為預測因子的預測結果誤差明顯偏大。

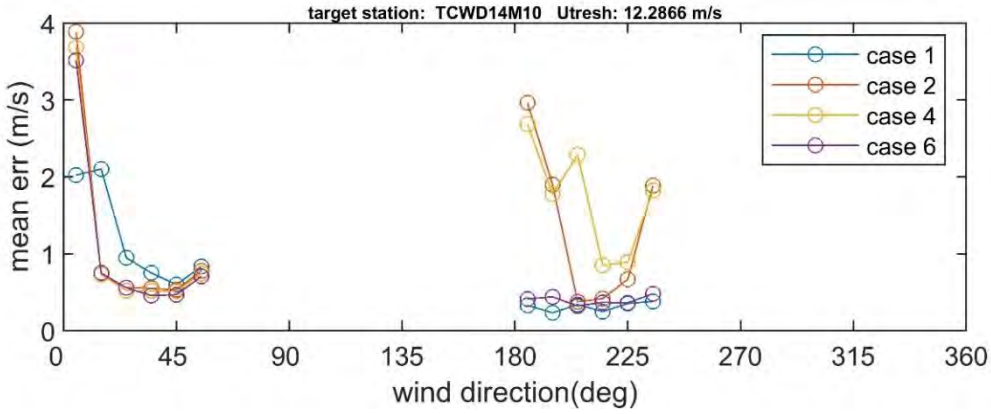

圖 14 設定超過 6 級風條件考慮不同預測因子進行模擬結果之誤差比較

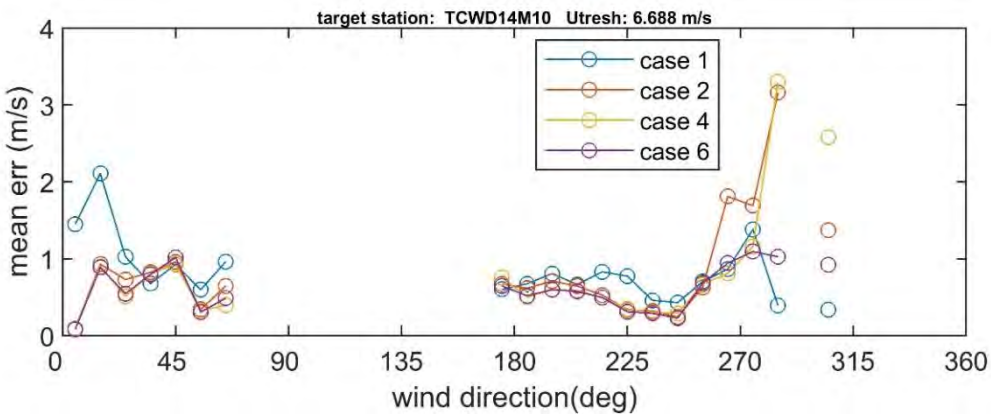

圖 15 設定超過 4 級風條件考慮不同預測因子進行模擬結果之誤差比較

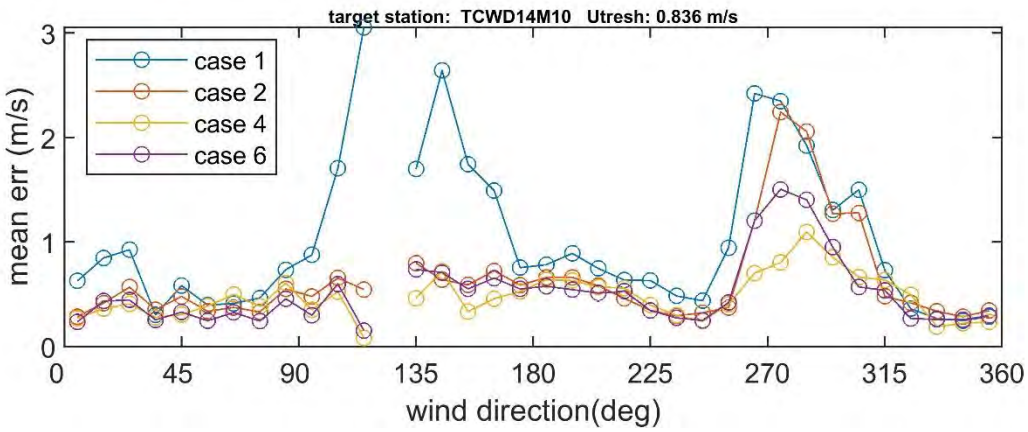

圖 16 設定超過 1 級風條件考慮不同預測因子進行模擬結果之誤差比較

7.7 模擬結果在異常資料判斷應用

圖 17 為設定超過 6 級風條件進行模擬預測與實測的誤差超出信賴區間的現象之一，發生於 112 年 4 月 8 日 5 點 40 分，該期間 4 個測站中有 3 個風速呈現微幅上升的現象，唯有測站 TCWD14M10 風速逐漸下降，前述之相關性分析顯示大趨勢上測站之間是成正相關，因此，當大部分測站呈現上升現象，模型推測結果測站 TCWD14M10 亦是往上趨勢，而實測卻呈現往下，導致預測結果與實測落差較大。雖然如此，但是各測站風速時序列的變動，基本上呈現緩和變化顯示並非儀器錯誤，而是局部環境的變化。圖 18 預測誤差超出信賴區間的另一個例子，發生於 5 月 31 日 0 點 10 分，該期間 4 個測站風速大趨勢呈現微幅下降的現象，但是測站 TCWD14M10 風速在 10 分鐘之間呈現大幅度的起伏，然後又回復緩變的現象，模型推測結果並未見短時間劇烈變動現象，導致預測結果與實測落差較大。這些誤差的原因有可能是短時間在測站局部區域出現小尺度的風場變動現象，只有單一測站偵測到其變動，其他測站則是同時受到大尺度風場現象的影響而呈現一致的變化趨勢，因此利用其他測站推算的結果與局部量測有較大差異。另外設定風速 1 級以上的資料進行預測，出現較大的誤差的情況如圖 19，同港口的 3 個測站普遍是偏西風的情況，然測站 TCWD02M10 的風向明顯與其他站不同，呈現大幅度搖擺狀態，而且風速顯著偏低與一般情況下該測站為最大的現象明顯不同，判斷與該測站位於燈塔東側，當吹西風狀況位於燈塔下游，風場受到燈塔阻擋在下游側出現不穩定尾流的流況影響。圖 20 則是在偏東南風情況測站 TCWD14M10 風速呈現顯著的降低，偏離一般情況下其風速與測站 TCWD02M10 接近的現象，而且風向與其他站不一致，明顯受到局部環境影響。統計全部模擬結果出現誤差超出 99%信賴區間的資料繪製發生月份、時間與風向的分佈如圖 21，可見大部分情況是風向偏向東南東風與西北風，影響的因素主要就是發生機率低與局部環境受到結構物影響，偏西北風時測站 TCWD02M10 受到燈塔阻擋，偏東南東風時測站 TCWD14M10 受到南堤紅燈塔鄰近構造物阻擋。實務應用上，在上述這些情況下，若根據特定單一受到局部影響的測站資料進行風場的解讀，則可能會出現誤判的情形。

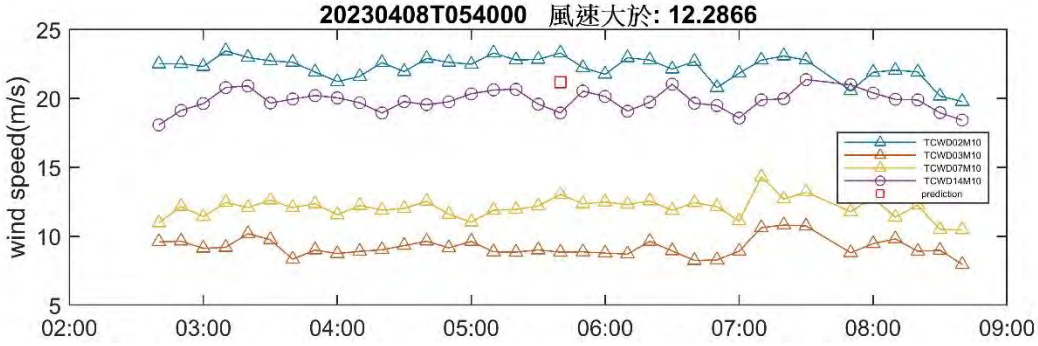

圖 17 預測結果與實測誤差超出信賴區間的現象 (設定風速 6 級以上的資料進行預測)

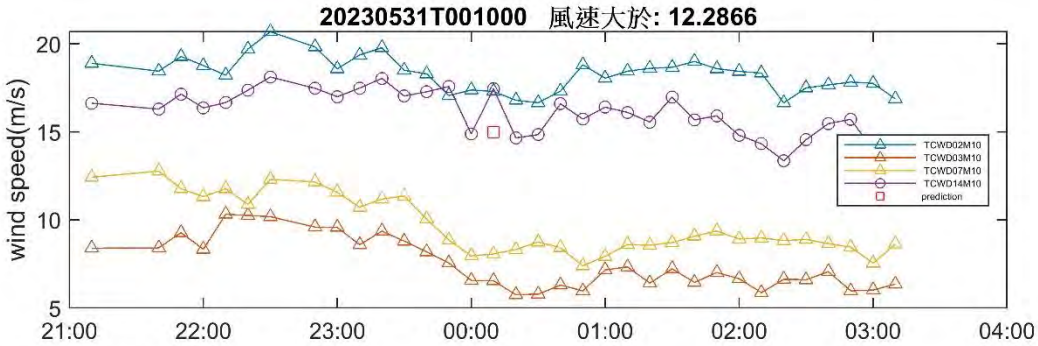

圖 18 預測結果與實測誤差超出信賴區間的現象 (設定風速 6 級以上的資料進行預測)

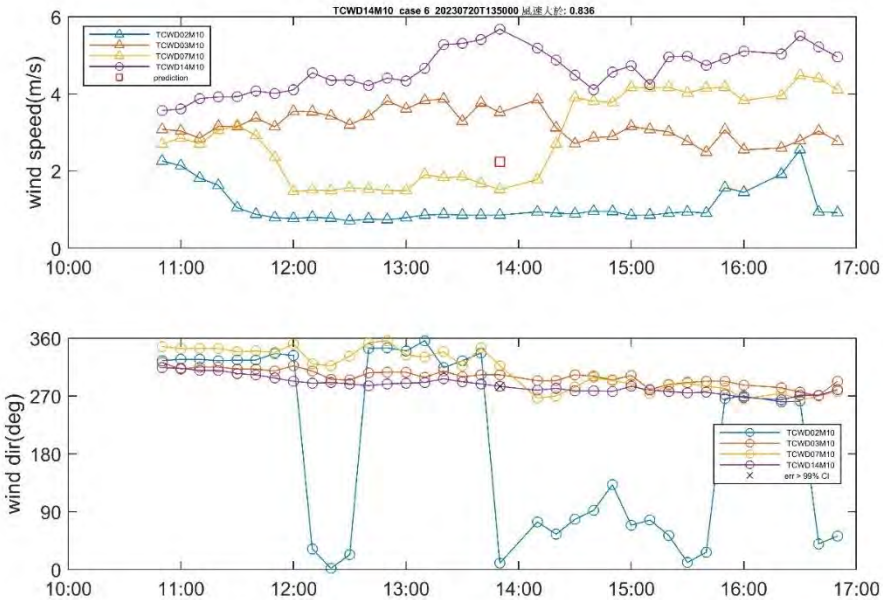

圖 19 預測結果與實測誤差超出信賴區間的現象(設定風速 1 級以上的資料進行預測)

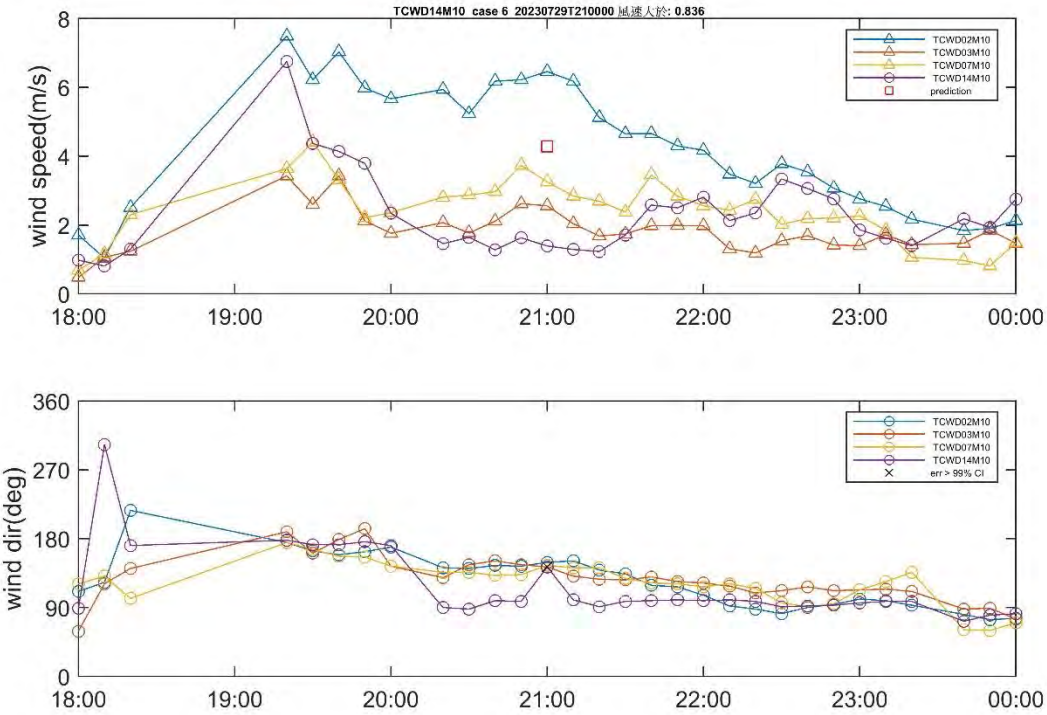

圖 20 預測結果與實測誤差超出信賴區間的現象(設定風速 1 級以上的資料進行預測)

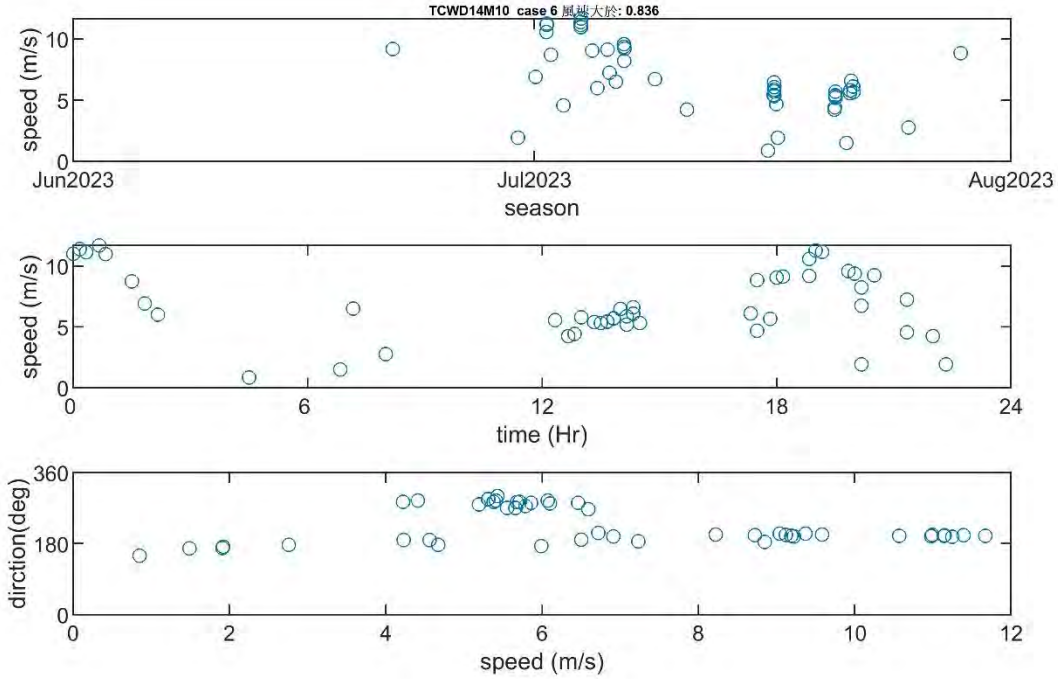

圖 21 預測結果與實測誤差較大的發生日期與時間(設定風速 1 級以上的資料進行預測)

八、結論

本文根據 2023 年 1 月至 8 月臺中港內 4 個風速測站的資料，建構測站風速之間高斯過程迴歸模型，並進而推算測站風速。研究成果顯示就推算臺中港南堤紅燈塔測站 TCWD14M10 附近的風速而言，若僅考慮 6 級風以上的情形，則單以北堤堤頭綠燈塔測站 TCWD02M10 的風速資料，進行高斯過程迴歸的模型建構，可以最有效率的方式得到良好的預測結果。考慮 5 級風以上的情形，則引入測站 TCWD02M10 與 TCWD03M10 進行模擬預測，可兼顧效率同時得到良好成果。考慮 4 級風以上的情形都納入推算，通過篩選的資料數超過 6 級風條件下的 2 倍以上，則引入臺中港現有全部 4 個測站資料進行模擬預測誤差普遍可以維持在 1m/s 以下，然因資料量大模擬效率相對較低。在 4 級風以上的情況，風向基本集中在偏東北與西南向，其他風向幾乎未發生，因此應用上受到限制。當風速降低到 3 級風甚至 1 級風條件，風向涵蓋範圍逐漸擴大到全部領域，且預測誤差維持在相似的範圍內，但是預測效率顯著降低，同時在西北與東南風向情況預測結果誤差較大。在異常風速分析，顯示當風向來自上游方向時，測站因鄰近結構物的阻擋，位於下游處，導致偶爾出現不穩定風場，其風速變動偏離一般趨勢，因而被偵測為異常現象。目前的分析資料尚未涵蓋完整一年的季節性變動現象，因此，在風向發生機率較低的部分可能因為資料數量少，導致模型學習不足，因而影響預測結果，尤其是在偏東南風與西至西北風情況更是顯著，後續應納入更完整資料後再持續探討。

參考文獻

1. 胡波，俞燎霓，藤代高(2019)，「高斯過程迴歸方法在浙江沿海海島冬春季陣風預報中到應用試驗」，熱帶氣象學報，35(6)，767-779 頁。
2. 111 年港區風力及能見度觀測系統建置維護與觀測資料品管統計成果報告，交通部運輸研究所港灣技術研究中心(2022)。
3. 江文山、賴志峰、黃國展、蔡宜秦、衛紀淮(2021)，「結合離散小波轉換與深度學習技術預測短期風速研究」，港灣季刊，第 118 期，第 41-51 頁。
4. 賴志峰、江文山、黃國展、蔡宜秦、衛紀淮(2021)，「港區風力資料品管與補遺」，港灣季刊，第 118 期，第 52-58 頁。
5. Bushnell, M., & Worthington, H. (2022). Manual for Real-Time Quality Control of High Frequency Radar Surface Current Data: a Guide to Quality Control and Quality Assurance for High Frequency Radar Surface Current Observations. Version 2.0.
6. Rasmussen, Carl Edward, and Christopher KI Williams. Gaussian processes for machine learning. Vol. 1. Cambridge, MA: MIT press, 2006.
7. Sloughter J. M., Geniting, T., Raftery A. E. (2010). Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. J Amer Stat. Assoc. 105(489): 25:35

8. Thorarinsdottir T. L., Johnson M. S.(2012) Probabilistic wind gust forecasting using nonhomogeneous Gaussian regression. Mon. Wea. Rev. 140(3): 889-897