

中央氣象局 103 年大學生實習成果報告

--紫外線指數資料分析初探

林容伊¹ 施昀汝¹ 蕭朱杏¹ 陳雲蘭²

¹臺灣大學公共衛生學系

²中央氣象局氣象預報中心

摘要

中央氣象局自民國 98 年起與台灣大學公共衛生學系合作，於暑假期間提供升上大四的同學六個星期的實習課程，同學們可藉由豐富的氣象資料及有趣的應用議題來進行實作，磨練在學校所學的統計技術。今(103)年度，我們兩位台大同學接受實習的應用性議題，除了接續去年學長姊在發展紫外線指數迴歸預報模式的工作之外，亦首度嘗試探討台灣地區紫外線指數隨高度變化的統計關係式，以提供氣象局將來對紫外線指數網格化作業需求的參考。首先，我們針對海拔高度以及天氣因子等紫外線相關因子探討他們與紫外線的關係，並從中挑出適合的因子進行建模。過程中使用 2010 至 2012 年的資料建立線性模型，並預測 2013 年的紫外線每日最大值，接續則利用均方根、相關係數以及色塊法等方法進行校驗，比較各模型的預報表現。在紫外線指數隨高度變化關係的探討方面，我們假設紫外線與海拔高度的關係為指數關係或線性關係，完成了初步的分析。

一、紫外線基本資料

2014 年暑假由於台灣大學公共衛生學系的暑期實習，與氣象局預報中心合作，為了協助氣象局日後能夠銜接環保署的紫外線預報工作做準備，首先要先了解紫外線特性以及可以來預測的因子，在去年學長姐的實習工作後已經對於預報模式有特定的了解了，因此我們要了解是否有什麼可以更加改進讓預報更加準確。

1.1 過去紫外線分析

2013 年臺大公衛實習的雅文學姐和崧岳學長在氣象局進行紫外線的特性探討以及建模的驗證。了解到因為太陽角度的關係，紫外線在中午 11 點到下午 1 點會達到最大值，夏天時紫外線會比冬天還要高出許多，相同季節時，南部會比北部高。另外在高山上的紫外線也會因為受大氣層中許多物質影響較少而比平地高，從相同緯度、相同季節、相同時間的紫外線比較中，可以看到高山上的紫外線比平地測站還要高，這項特性我們也會在面化作業時來進行更加詳細的分析。在美國氣象局中也有提過，紫外線會受到空氣的塵埃程度影響，因此我們也會在接下來的分析中特別去看我們所猜測的因子—懸浮微粒。許多文章都有提過，紫外線會受到平流層的臭氧分解而減小，但是在台灣上空平流層的臭氧變化量並不大，因此我們沒有進行太多的分析，有趣的是在對流層中的臭氧和紫外線有高度的正相關，是由於紫外線在對流層所造成的光化學作用，在這邊我們雖然不會對於臭氧與紫外線進行更

多的關係分析，但是對流層的臭氧對於人體的影響很大，或許未來在針對紫外線所造成的健康影響中可以更加深入的探討。

1.2 環保署(EPA)與氣象局(CWB)紫外線資料比較

由於中央氣象局的紫外線測站在台灣本島只有 15 個，2013 年新增了桃園新屋測站為第 16 個測站，測站所測的紫外線與環保署同為紫外線光中的 UVB，因此我們如果可以取得更多資料來當作分析可以使我們的分析結果更加穩定。

我們先進行簡單的分析來看氣象局與環保署測得的紫外線有沒有不同處，我們取相同地區相近的氣象局測站與環保署測站來進行比較。台灣本島可以進行配對的測站中我們先以台北 2010-2013 資料為例，氣象局測站位於台北市，而環保署的測站位於板橋。從日分佈以及年分佈中可以發現，環保署的資料都有略高於氣象局資料 1~2 個單位，猜測可能是因為兩個單位所使用的儀器不同或是測站所在的位子有所差異導致，因此我們也進行其他可以配對測站的分析，以恆春、台東 2010-2013 年資料的年分佈來看都有 EPA 資料高於 CWB 的現象，雖然還是有可能是因為測站地點不同所造成的影響，但是我們認為儀器所帶來的偏差是主要的原因，所以接下來的資料分析我們只使用 CWB 的資料並沒有將兩個單位的紫外線指數資料混合使用，以避免誤差更大。未來若有時間也可以針對 EPA 的資料進行更多的探討，來比較分析結果是否和 CWB 資料的分析結果有所不同。

1.3 CWB測站資料

我們將各測站的紫外線資料起始時間到目前為止做一個簡單的統計，可以看到全台灣的 20 個測站有 19 個測站的資料至少長達 4 年，最新加入的新測站只有一年的時間。缺測率的部分在日月潭以及玉山站有比其他地區大許多的現象，原始資料中我們發現，日月潭在 921 大地震之後有長達數月的缺測資料，因為地震時的受損導致測站受損而無法正常運行，若將此連續缺測資料去除後，日月潭的缺測率會和其他測站差不多；玉山測站方面，缺測的月份並沒有特別的分佈，但是從原始資料中可以看見，每一年缺測的資料時間多集中在一個連續的時間點上，我們猜測可能是因為高山站如果遇到儀器的故障維修不易，容易導致出現連續缺測的現象。

1.4 紫外線危險級探討

由於紫外線到達危險級時容易對人體造成健康上的影響，因此在這邊我們也進行了關於紫外線危險級的探討，希望可以更加了解危險級出現的時間。已開始我們先將各個測站的歷史資料中每日 UVI max 的第一名拿出來看，可以看見台灣不論是本島或是外島，不論北部或南部都有出過危險級。為了瞭解各測站歷史資料的前幾名是否都會是危險級，我們將各測站所有歷史資料 top10 列出來後，可以從各個測站的 top10 中發現最大值常出現在七月與八月，少數會出現在五月，因此我們也猜測會不會紫外線危險級容易出現在某些月份，為此我們將紫外線每日最大值達危險級所出現的最多次月份進行比較，台灣的測站的都在七月時最容易出現危險級，六月與八月次之。特別是玉山站在每個月都有可能出現危險級，即使是在冬天，由此可見高度對於紫外線影響之甚。

一般想像中六月應該是台灣的梅雨季節，因此紫外線達危險級的機會應該比七月以及八月低，但是從其他測站的比較中我們發現每個測站都是七月最多，而有些測站六月平均出現過危險級的次數比八月高。因此我們也針對究竟是八月或是六月出現出較多危險級進行比較。初步看出在西半部與花蓮，六月平均出現過危險及次數比八月多，另外東北部以及東南部則是八月會比六月多。我們猜測可能跟地形所造成的降雨時間不同所影響，以及地區的降雨情況不同而有所影響。

二、紫外線相關因子探討

從學長姐分析的文獻中發現，每日最高溫與雲量都和紫外線呈現非常高的正相關，在建模時也是很好的因子，但是在建模時會有些微的誤差，另外在雲量預報上也難預報到很準確，因此若能以相對濕度代替高空中的水氣(雲)以及低空

中的水氣，可以讓預報上更加順利。此外，在環保署與美國氣象局的資料中都有提到空氣乾淨度會影響紫外線的強度，因此我們也探討了懸浮微粒與紫外線的關係。

2.1 相對溼度

2.1.1 相對溼度基本探討

相對溼度為空氣中的水氣程度除以在此溫度下可容納的總水氣。因此可以猜測日變化應該在中午時隨著溫度提高而下降。在這邊我們想知道除了日變化以外，相對溼度會不會隨著季節而有所變化，因此我們畫出台北和台南 2010-2013 中午 12 點相對溼度的月變化 boxplot，來進行比較與探討。從圖中可以看到四面環海的台灣，不論是在北部或南部相對溼度都會大於 40，北部可以看到較明顯的季節變化，夏天明顯低於冬天，因為北部冬天會有東北季風將潮濕的空氣推過來，又溫度低才會使得相對溼度較其他月份大許多也比同時間的南部大很多，而東北季風對於南部的影響不大，因此很難看出特定的季節變化的趨勢。

2.1.2 相對溼度與紫外線

取台北 2010-2013 一天中同一個時間點的相對溼度和紫外線來進行分析，可以發現相關性都呈現很強的負相關，雖然會隨著月份有些微的變化，但是大致上都呈現相同的趨勢。台北測站的相關係數介於(-0.51,-0.77)，因此在接下來的建模工作中我們也將考慮相對溼度這個因子。如果我們可以讓高溫加相對溼度比單純只有高溫時好便可以將此模型納入預報時的參考。

2.2 懸浮微粒

2.2.1 懸浮微粒基本探討

PM10 為空氣中粒徑在 10 微米 (μm) 以下之粒狀污染物，在空氣污染的地區濃度較高，台灣地區還有可能是受到沙塵暴的影響，濃度會有提高的現象。PM2.5 則是空氣中粒徑在 2.5 微米 (μm) 以下之粒狀污染物。

我們以相近緯度但都市化程度不同的城市來看，冬天時 PM10 和 PM2.5 都比夏天高，且東部地區不論是 PM10 還是 PM2.5 都比西部低。美國氣象中心也表示，在空氣乾淨的地方懸浮微粒少會讓紫外線比懸浮微粒多的地區強，同樣的現象在台灣也能看見，在夏天同緯度的地區，懸浮微粒較多的西部紫外線普遍都高於懸浮微粒較少的東部。

2.2.2 懸浮微粒與紫外線

宏觀的來看我們看見了:同緯度,同海拔高度,空氣乾淨的地方紫外線會比懸浮微粒多的地方大。為了瞭解懸浮微粒和紫外線真正的關係,我們先從簡單的 Scatter plot 來看。以台東、台中、台南、花蓮 EPA 的空氣測站中,所測得 2010~2013 年七月的中午十一點的 PM10、PM2.5 以及 CWB 在約略相同地區的紫外線指數測量當做例子。我們發現很難直接看出懸浮微粒和紫外線有什麼樣的相關性,但是在前面我們也有發現雲量是一個影響紫外線非常重要的因子,所以我們猜測會不會在不同雲量時懸浮微粒會和紫外線有不同的相關分布型態,因此在這邊先將雲量分為三個等級:多雲(深灰色):雲量>7、少雲(淺灰色):2<雲量≤7、極少雲(藍色):雲量≤2。將雲量分級後,我們發現只有在雲極少的時候紫外線會和懸浮微粒 PM10,PM2.5 呈現負相關,但是不同地區又會不太一樣。

三、統計預報建模較驗

之前學長姐的報告中建議的模型為”日最高溫”與”日最高溫加雲量”此二種模型;我們在探討相關因子時也發現相對濕度與紫外線有高度相關,我們也發現懸浮微粒雖然與紫外線呈現負相關,但關係不明顯,所以我們不會將懸浮微粒當作變項放入模型,但可以當作預報時的參考指標。故我們將會建立以下三種模型並且進行比較:

迴歸模型一:

每日紫外線最大值(UVI Max)~ Tx(日最高溫)

迴歸模型二:

每日紫外線最大值(UVI Max)~ Tx(日最高溫)+CDmin(11點與14點雲量最小值)

迴歸模型三:

每日紫外線最大值(UVI Max)~ Tx(日最高溫)+RHmin(11點至13點相對濕度最小值)

3.1 校驗方法

3.1.1 均方根、相關係數

我們將會使用各測站 2010~2012 年各月的資料進行建模,每個測站、每個月分都有自己的模型,並用這些模型預測 2013 年的每日紫外線最大值,求出預測值與觀察值的均方根以及相關係數進行校驗。均方根代表預測值與觀察值的差,我們希望均方根能小於 2010~2012 觀測值的標準差,並且越小越好。

3.1.2 色塊法

前面所使用的均方根與相關係數的校驗方法是把紫外線當作連續變項進行校驗,我們覺得如果紫外線預測值與觀測值落在同樣等級內應該也是可以接受的,因此我們也將紫外線當作類別變相進行校驗,我們將這樣子的方法稱為色塊法。

A. 畫出預測值與觀察值關係圖,並在圖上依紫外線等級畫出色塊。

B. 算出色塊內黑點數。

C. 將色塊內黑點數除以該月天數,若此值愈高,模型預測能力愈佳。

3.2 建模校驗—恆春

以過去建模時經驗較不好的恆春站來當作例子,進行兩種模式較驗。發現觀測值標準差介於 1~3 個單位,預測值和觀測值的均方根約介於 0.5~2.5 個單位,在 2 月時大於標準差,其他都小於標準差,相關係數大都小於 0.7,因此我們認為這三個模型預測能力並不算太好。恆春的各月均方根比較可知夏天(6~8 月)各模型預測能力比其他月份好,但若是從色塊法來看,恆春夏天(6~8)月各模型預測能力較其他月份差。

以色塊去校驗方式出來的結果和預期不同,為了探討原因,我們將 7 月為例進行討論。從恆春 7 月各模型的色塊圖上可看出:如同先前算出的各模型預測值與觀測值之相關係數預測值與觀察值相關性不低,但迴歸線卻都偏離 $Y=X$ 這條線,因此落入色塊的比例才會比其他月份低,我們猜想或許是因為每年紫外線相關因子與紫外線的迴歸線都會有些微的不同,而這件事我們也在每一年的相關性比較中驗證了。每年最高溫與紫外線的迴歸線都會有些微的變動:若溫度一樣,各年的紫外線可能差到 3 個單位。

四、紫外線指數隨高度變化關係探討

建模只能針對單個測站進行預報,若要對其他地方進行預報,則需要透過面化,將這幾個測站的預測值利用內插法推出各地的紫外線預測值。之前在進行建模時並未考慮高度這個因素,因此在面化前,我們會先探討紫外線和高度的關係。根據 WHO 文獻估計,高度每升高 300 公尺,UVI 值約增加 4%,為了解這樣的通則是否可以用在台灣我們將透過緯度相差不多,高度不同的四個中部測站(台中、嘉義、阿里山、玉山)來進行探討。

4.1 高度對紫外線的影響

4.1.1 高度對於紫外線呈現指數成長

假設紫外線與海拔高度的關係為指數關係(高度每上升 300 公尺,UVI 上升 X 倍)。

分析方法

$$UVI_{山地} = UVI_{平地} \times (1 + X)^{\text{高度}/300}$$

$$\rightarrow \ln UVI_{山地} = \ln UVI_{平地} + \ln(1 + X)^{\text{高度}/300}$$

$$\rightarrow \ln UVI_{山地} = \ln UVI_{平地} + (\text{高度} \div 300) \times \ln(1 + X)$$

我們先假設此四站海拔高度為零時 UVI 相同，

($\ln UVI_{山地}$) 為應變項， $(\text{高度} \div 300)$ 為自變項進行線性迴歸 ($\ln UVI_{山地} = a_0 + a_1 \times (\text{高度} \div 300)$)，再將得到的斜率項取 \exp 後再減 1 即為所求 ($X = \exp(a_1) - 1$)。

透過前述方法所算出的係數，可發現月份不同，UVI 隨高度增加的倍數不同：夏天係數通常比冬天低。此外，全部個案求得的係數之 R-squared 都小於 0.3。由於 R-squared 並沒有絕對的好壞，我們之後係數的 R-squared 將會與 0.3 進行比較。我們發現雲少和雲極少的 R-squared 多於 0.3，相較於沒分雲量，可信度較高。然而，雲量多的 R-squared 卻比未分雲量時低，我們猜測可能因為紫外線進入少，隨高度變化不明顯。

4.1.2 高度對於紫外線呈現線性成長

同樣地，我們假設此四站海拔高度為零時 UVI 相同，並將 $UVI_{山地}$ 為應變項， $(\text{高度} \div 300)$ 為自變項進行線性迴歸 ($UVI_{山地} = a_0 + a_1 \times (\text{高度} \div 300)$)。

為透過前述方法所算出的係數。從圖表中可發現月份不同，UVI 隨高度增加的倍數不同：夏天係數通常比冬天低。此外，全部個案求得的係數之 R-squared 都比指數關係大，雲少和雲極少係數的 R-squared 也都比指數關係大。然而，雲量多的 R-squared 和指數關係差異不大，我們猜測可能因為紫外線進入少，隨高度變化不明顯。我們認為，普遍而言，紫外線與高度的關係偏向線性關係。

五、討論與結語

在建模時為了控制不同地區不同月份的差異我們在每一個測站的十二個月都建模，如果我們可以使用太陽頂角在加上其他氣象因子與地形或許可以將台灣的紫外線預報整合成一個模型。另外在探討模型校驗時我們常常發現在二月以

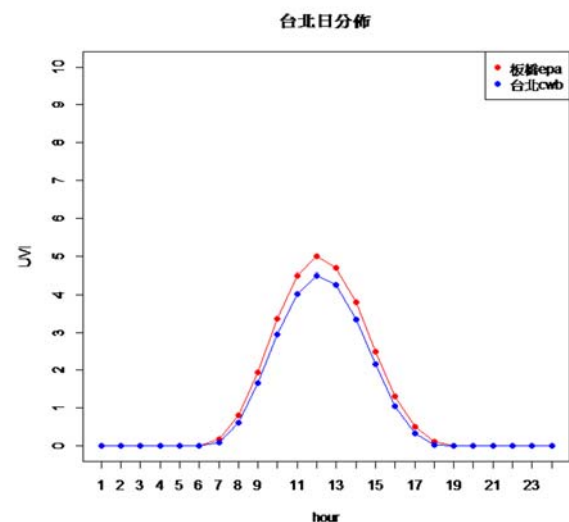
及九月的模式校驗上都發現模式沒有預測得很好，猜測可能是因為這兩個月份是台灣季節轉換之際，每年天氣的變化都不太一樣，所以使得預報上容易受到每一年氣候變化不同影響無法像其他月份精準，或許可以加入其他氣象因子讓這兩個月更加準確，便可以提升整體預報能力。

另外我們在建模式也看見最高溫加雲量的模式是預報的最好的模式，但是雲量是十分位變項，無法反映出雲層厚度，若在預報紫外線時可以參考衛星雲圖的預報，加入雲量以及雲種，或許可以讓預報員在雲量加高溫的模式上更有信心。

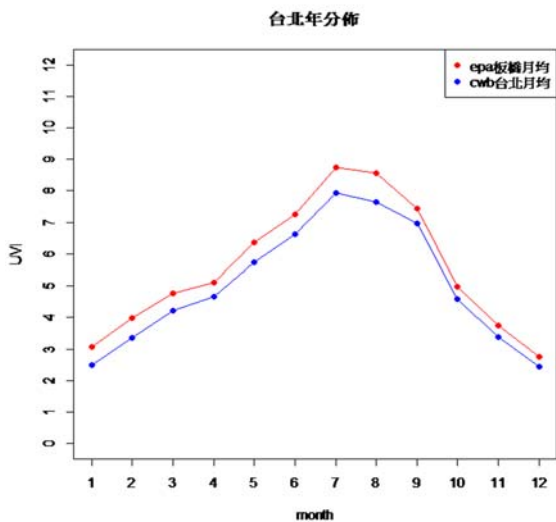
六週的實習讓我們收穫良多。最後，我們想要向提供及安排此實習機會的中央氣象局長官及台大公共衛生學系的師長們表達謝意。

五、參考文獻

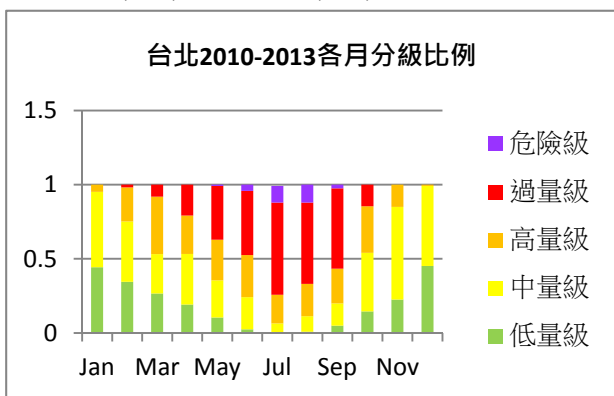
陳雅文、黃崧岳，2013：2013 年中央氣象局實習報告。
陳圭宏，2010：紫外線指數分析及探討，中央氣象局。



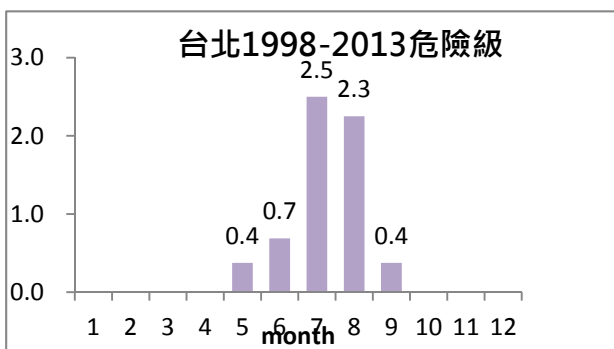
圖一，台北紫外線 2010-2013 年紫外線日分佈，每個小時的平均紫外線指數比較，紅色為 EPA 資料，藍色為 CWB 資料。兩者有些許差異。



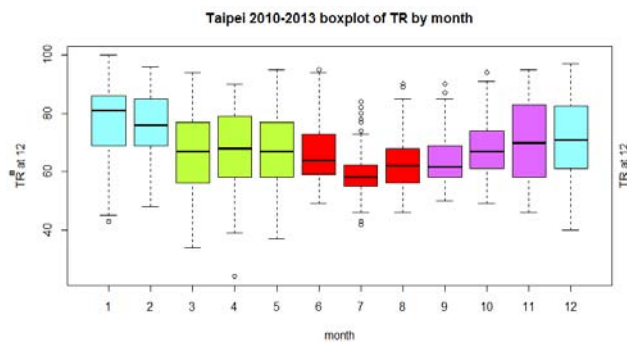
圖二，台北紫外線指數 EPA 和 CWB 每月平均資料比較。EPA(紅色)會高於 CWB(藍色)的資料 1~2 單位。



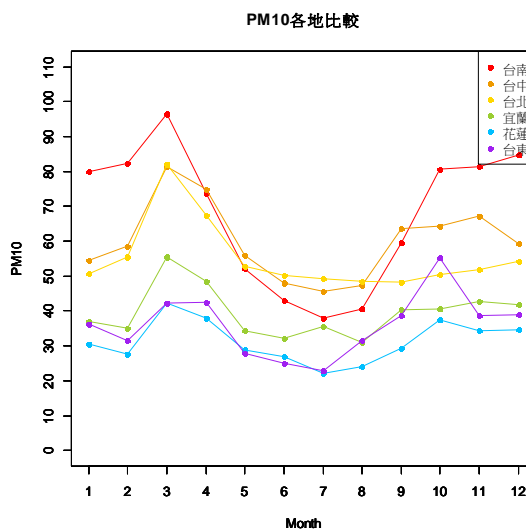
圖三，台北 2010-2013 紫外線每日最大值分級比例，可以看見在夏天時有大於 50%的天數紫外線會高出過量級。



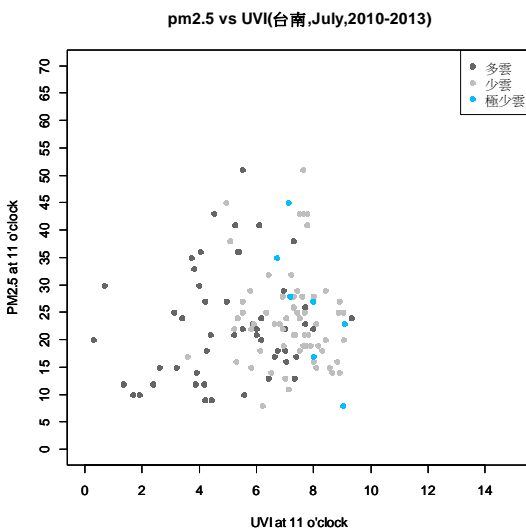
圖四，台北歷史資料上每日紫外線最大值達到危險級出現的平均天數。在七月最多，八月次之。



圖五，台北相對濕度各月分佈，在四面環海的台灣，相對濕度大多都會大於 40。台北季節變化明顯。

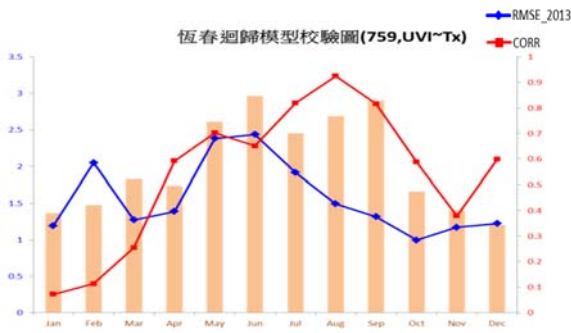


圖六，台灣東西部懸浮微粒之比較，可以看出冬天懸浮微粒較多，且東部懸浮微粒普遍比西部低。

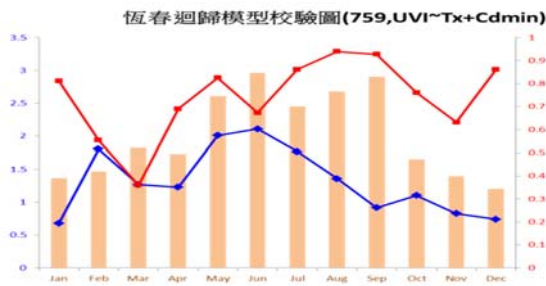


圖七，台南 2010~2013 年七月中午十一點懸浮微粒與紫外線關係圖，綜觀而言懸浮微粒與紫外線

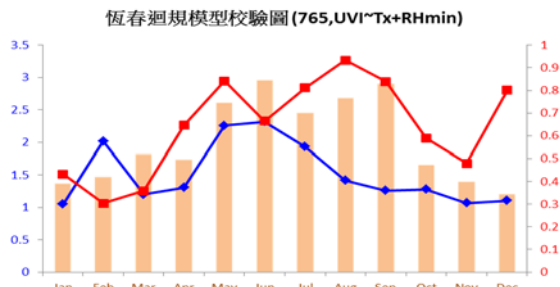
關係不明顯。將雲量分類後，在極少雲(淺藍色點)可稍微看出紫外線與懸浮微粒之關係。



圖八，恆春模型(UVI max~Tx)校驗圖，二月的RMSE 大於標準差，模型對於預報能力較差。夏天 CORR 高於 0.6，RMSE 都很好模型預報能力很好。



圖九，恆春模型(UVI max~Tx+Cdmin)校驗圖

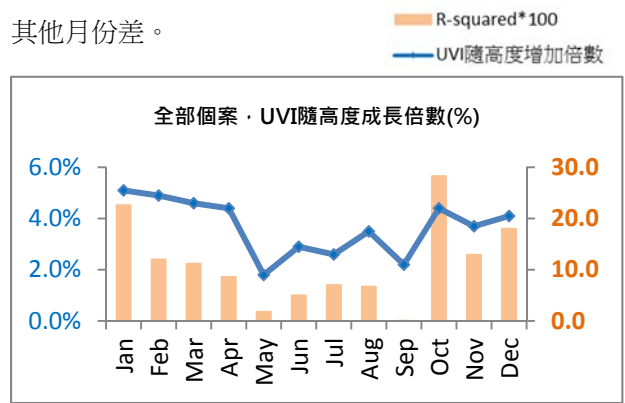


圖十，恆春模型(UVI max~Tx+RHmin)校驗圖

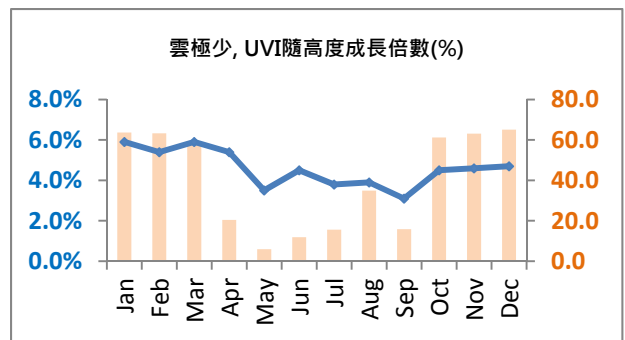


圖十一，恆春的色塊法校驗比較圖，圖八-十中可知夏天(6~8月)各模型預測能力比其他月份好，但從圖十一的色塊法卻發現夏天(6~8)月各模型預測能力較

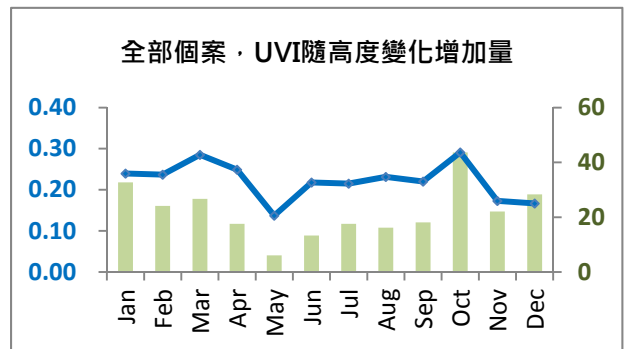
其他月份差。



圖十二，紫外線指數隨高度呈現指數上升，各月隨高度上升 300m，紫外線指數上升的倍數。



圖十三，在雲極少時，紫外線指數隨高度呈現指數上升，各月隨高度上升 300m，紫外線指數上升的倍數。



圖十四，紫外線指數隨高度呈現線性上升。各月隨高度上升 300m，紫外線增加量。



圖十五，雲極少，紫外線指數隨高度呈現線性上升。各月隨高度上升 300m，紫外線增加量。