

時間序列期末報告

天降甘霖是否能消滅隱形殺手-pm2.5?

指導教授： 林良靖 教授

組員： 陳育婷 H24034019

楊珍妮 H24035049

曾麗娟 H24035015

目錄

壹、研究動機與目的.....	P.3
貳、資料介紹.....	P.4
參、建模流程.....	P.5
肆、模型建立.....	P.6
✓ Pm2.5 模型建立.....	P.6
a.模型判定.....	P.10
b.參數估計.....	P.12
c.殘差診斷.....	P.12
d.其他候選模型的 aic 及殘差診斷結果.....	P.14
✓ 雨水模型建立.....	P.15
a.模型判定.....	P.18
b.參數估計.....	P.19
c.殘差診斷.....	P.19
d.其他候選模型的 aic 及殘差診斷結果.....	P.22
伍、模型預測.....	P.23
陸、相關性檢定.....	P.24
柒、結論.....	P.31
捌、感想.....	P.33
玖、建議.....	P.34
拾、引用.....	P.35

壹、研究動機與目的

在以前，pm2.5 對很多人來說應該是一個相當陌生的名詞。因為大家對於空氣汙染日趨重視，以及政府推廣並廣設偵測站，才讓越來越多人認識這個隱形殺手。身為大學生的我們，也是大概從今年才真正意識到 pm2.5 是甚麼、以及它的危險性。

pm2.5 是漂浮在空氣中，粒徑大小小於或等於 2.5 微米(μm)的粒子，單位以微克/立方公尺($\mu\text{g}/\text{m}^3$)表示。pm2.5 的可怕之處是它非常微細，可穿透肺部氣泡，並直接進入血管中隨著血液循環全身，故對人體及生態所造成之影響是不容忽視的，可能會導致人體器官不同的危害。

今年秋冬，社交軟體或網路新聞上一直出現南台灣紫爆(pm2.5 濃度最高段落)的新聞，戴口罩的人越來越多。身為學生的我們，出入也開始戴上口罩，注意即時 pm2.5 濃度。一時之間，pm2.5 讓大家人心惶惶、小心防範，出門第一件事就是戴上口罩。因此當要選擇時序報告主題時，我們毫不猶豫地選擇了最近影響台南甚遽的 pm2.5。

另外，我們同時也想到，今年九月時三個颱風接連襲及全台，那陣子似乎 pm2.5 濃度警告比較少，且 AQI(空氣汙染指數)沒有那麼高。因此，我們產生了一個假設：下雨是否會洗淨空氣中汙染物，降低 pm2.5 濃度呢？

將此兩議題一結合，就產生了我們此份報告的研究主題：探討降雨量對 pm2.5 的影響。

貳、 資料介紹

資料說明: pm2.5 拿到的資料是以小時為單位，而且包含所有空汙物。因此

我們在 excel 中先將 pm2.5 資料挑出，將檢測值為無效值刪除。刪完後，一天

中若資料<18 筆即刪除該天資料，清完資料後再算日平均，月平均。

目的： 分析台南月均降雨量是否會影響月均 pm2.5 濃度

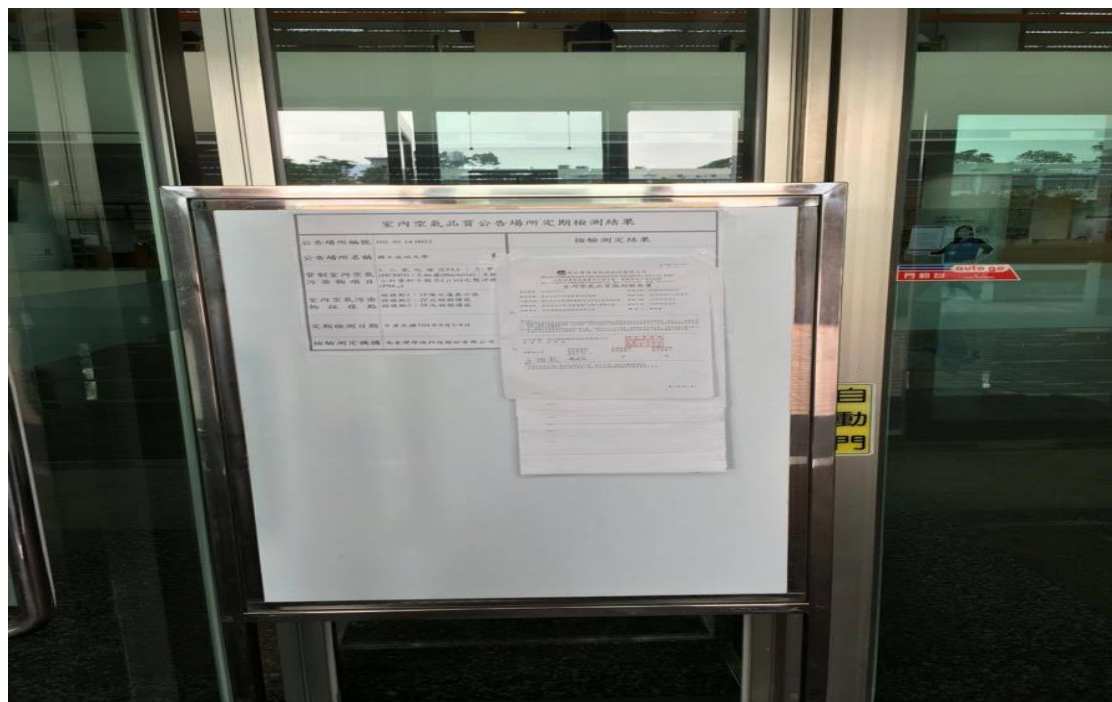
分析軟體： R 軟體、Excel

資料來源： 行政院環境保護署以及交通部中央氣象局資料庫

資料期間： 2007 年 1 月之 2015 年 12 月

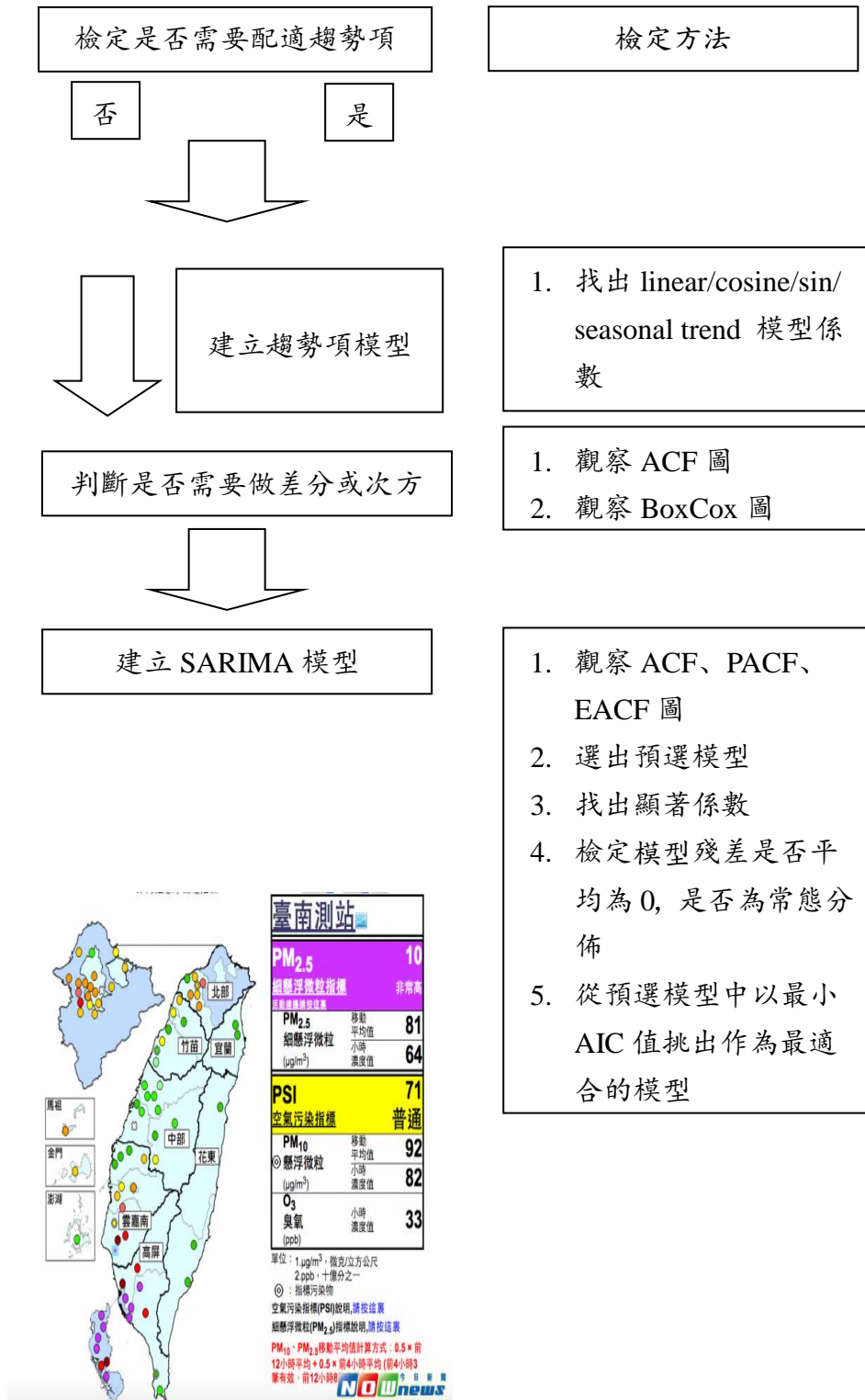
資料間隔： 以月為單位

資料型態： 台南市 pm2.5 月均量(單位: $\mu\text{g}/\text{m}^3$)、台南市降雨量月均量(單位: 毫米)



成功大學圖書館正門口的室內空氣品質檢驗報告，大家越來越重視空氣汙染

參、建模流程

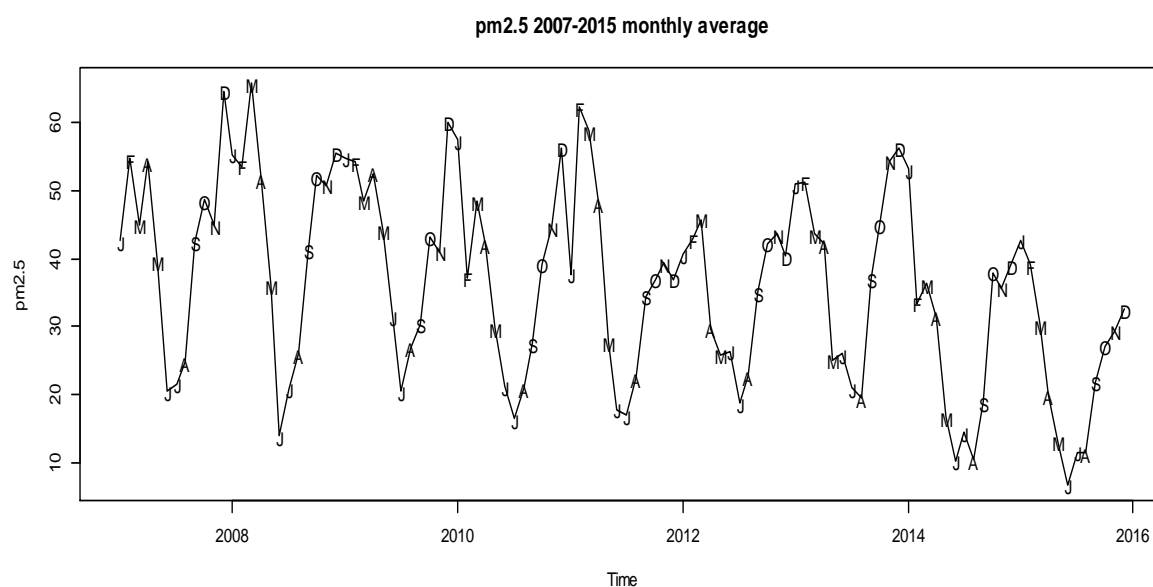


透過 pm2.5 即時值讓我們知道空汙程度

肆、模型建立

(一) pm2.5 模型建立

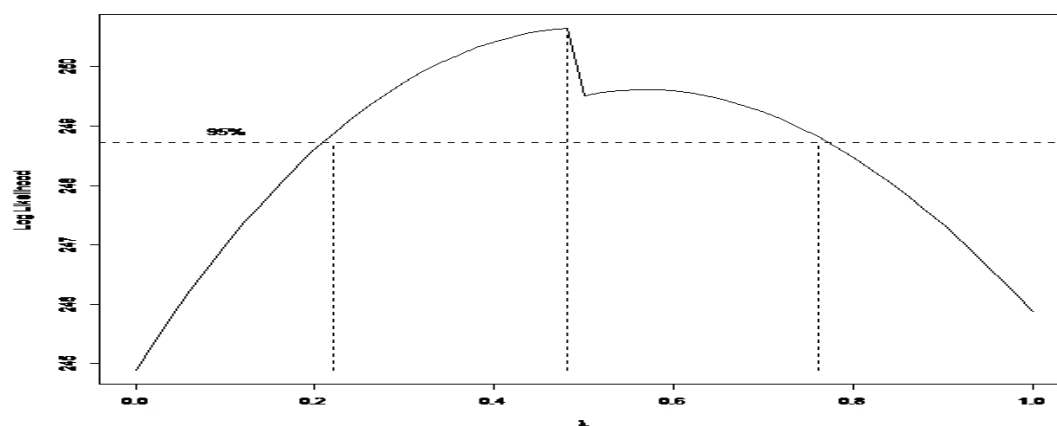
首先，觀察台南市 pm2.5 從 2007 年到 2015 年月平均濃度的時間序列圖。



圖(一)

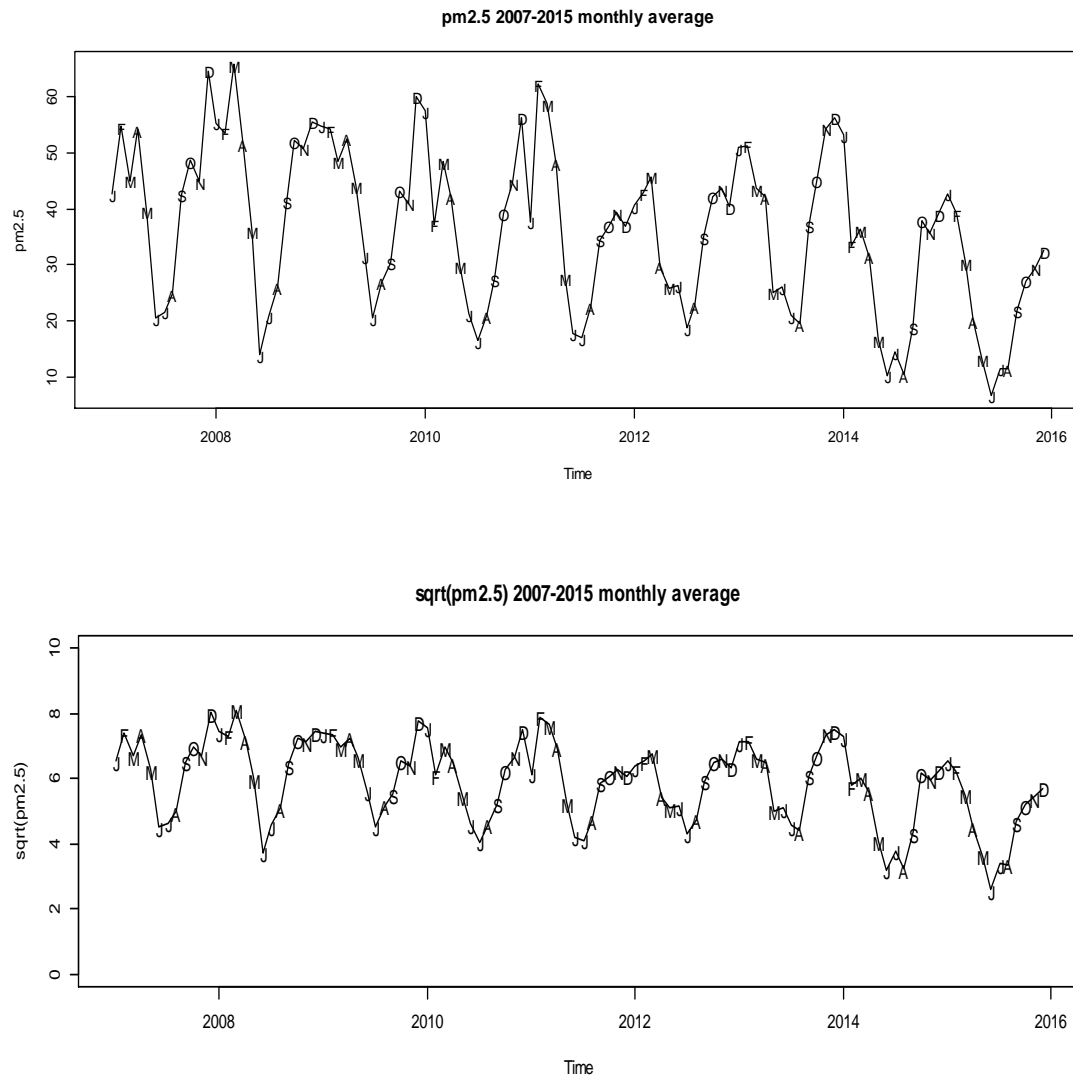
可以看到此筆時間序列似乎有往下降的趨勢，並且有季節效應，十二月到三月濃度似乎比較高，五到八月濃度比較低。

我們先將資料做 Box-Cox 轉換，並且繪製 λ 與對數概似估計值圖。



(圖二)

可以看到，最大概似估計量 λ 的 95%信賴區間介於[0.2,0.8]，最大值發生在 $\lambda=0.5$ 時。因此我們決定將資料做根號轉換。



(圖三)

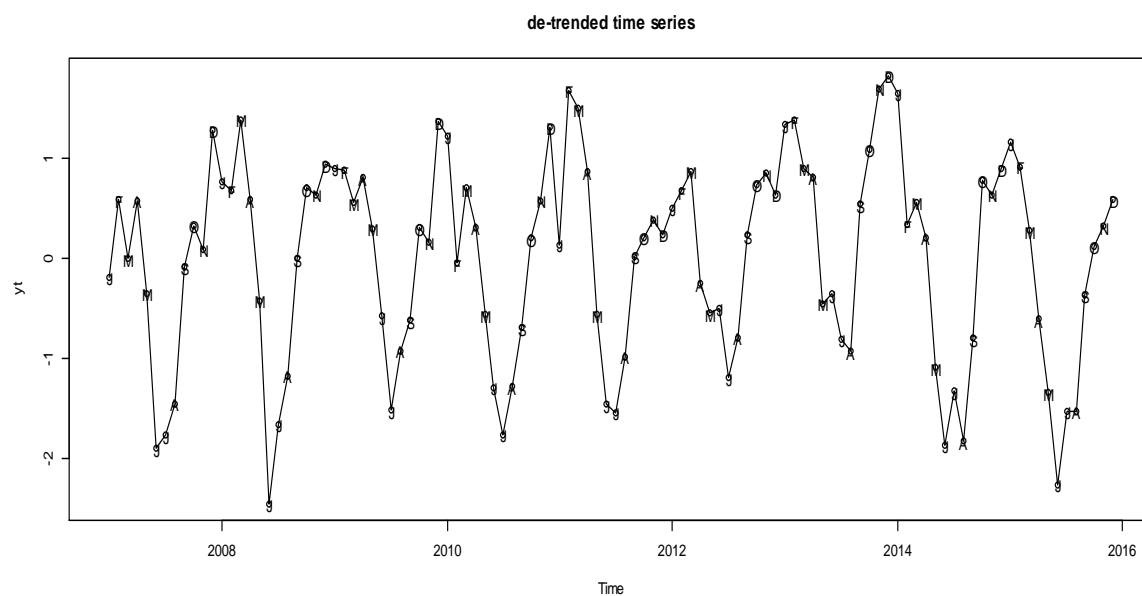
上圖為原始時間序列，月跟月之間的變動程度較大。下圖為經過根號轉換後的時間序列，月跟月之間的變異減少且趨於穩定。但還是可以看到序列有往下降的趨勢，因此我們決定先配適一次項的時間趨勢。

因此我們的模型為: $\sqrt{Y_t} = \mu_t + Zt$ where $\mu_t = \beta_0 + \beta_1 t$

	Estimated	S.E.	t value	P value
β_0	391.7253	86.4842	4.529	<0.0001
β_1	-0.1918	0.0430	-4.461	<0.0001

常數項和一次項的係數皆顯著，代表此模型有一個顯著向下時間趨勢。

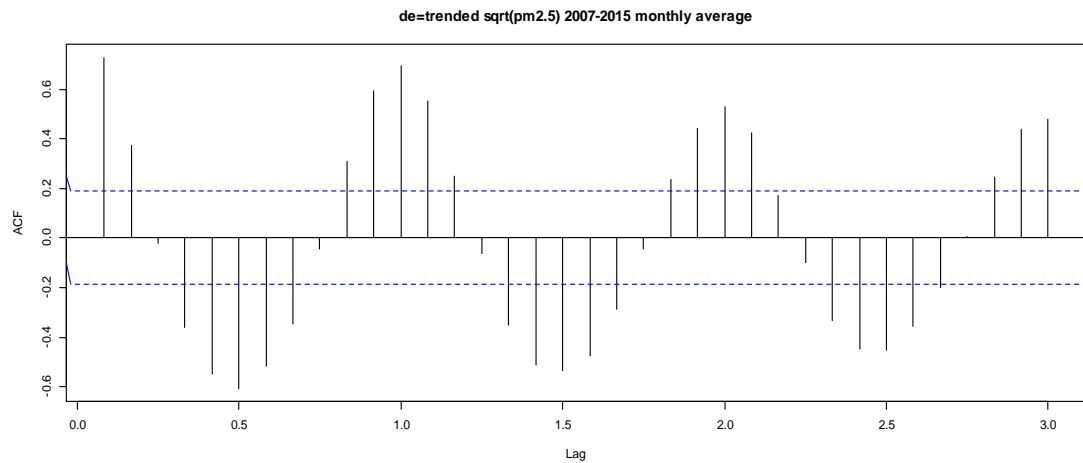
下圖為去除掉一次項時間趨勢後的时间序列圖。



(圖四)

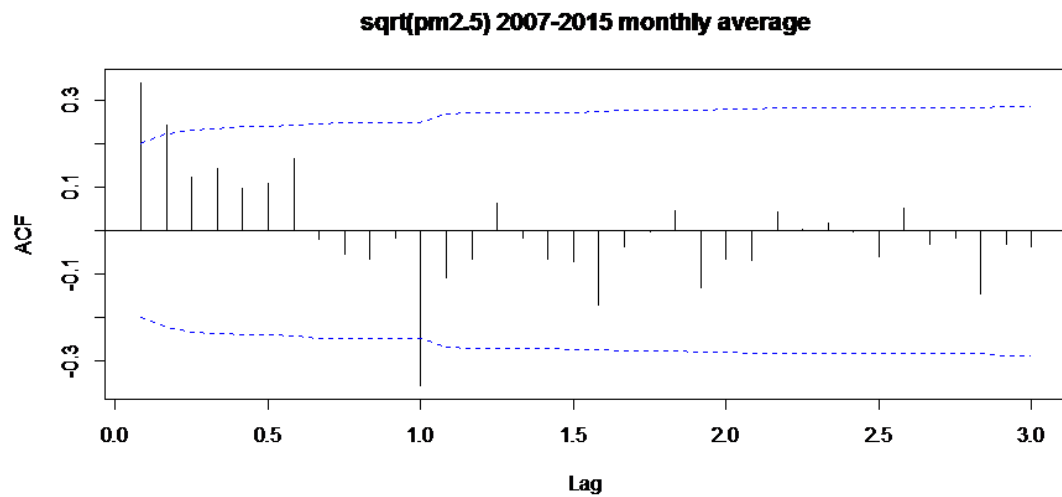
可以發現時間序列已沒有往下的趨勢，在零的上下穩定跳動。

接著我們觀察去除一次項時間趨勢後時間序列的 acf。



(圖五)

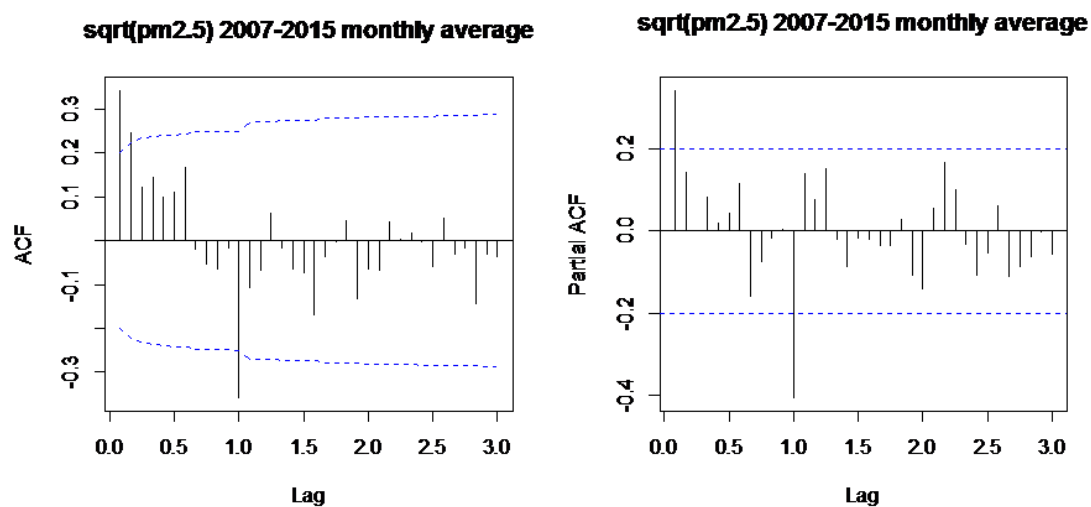
acf 圖可以看出，此序列尚未平穩，acf 呈現波浪狀，且有明顯的季節效應存在，大約是以十二個月為一周期，因此我們決定先對去除一次項後的時間序列做一階季節差分。



(圖六)

上圖為經過一階季節差分後的 acf 圖，可以發現季節效應已去除，序列已經平穩，所以可以進行模型判定。

a. 模型判定



(圖七)左圖為 acf 圖，右圖為 pacf 圖

1. 從 acf 圖中，在 lag=12 時 acf 明顯突出信賴線，lag1、lag2 也有突出信賴線。因此先判斷為 $\text{SARIMA}(0,0,2)\text{X}(0,1,1)_{12}$
2. 從 pacf 圖中，看到有可能是 $\text{SARIMA}(1,0,0)\text{X}(1,1,0)_{12}$ 或 $\text{SARIMA}(12,0,0)\text{X}(1,1,0)_{12}$ ，但合理推測後者步數太大，係數一定會有許多項不顯著。

eacf

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	x	o	o	o	o	o	o	o	o	o	x	o	o
1	x	o	o	o	o	o	o	o	o	o	o	x	o	o
2	o	o	o	o	o	o	o	o	o	o	o	x	o	o
3	o	x	o	o	o	o	o	o	o	o	o	x	o	o
4	o	x	o	o	o	o	o	o	o	o	o	x	o	o
5	x	o	x	o	o	o	o	o	o	o	o	x	o	o

圖(八)

eacf 顯示可能的模型為 $\text{SARIMA}(0,0,2)\text{X}(0,1,0)_{s=12}$ 。

候選模型整理成表格如下圖。

根據圖表	可能模型
acf	$\text{SARIMA}(0,0,2)\text{X}(0,1,1)_{s=12}$
pacf	$\text{SARIMA}(1,0,0)\text{X}(0,1,0)_{12}$ $\text{SARIMA}(12,0,0)\text{X}(0,1,0)_{12}$
eacf	$\text{SARIMA}(0,0,2)\text{X}(0,1,0)_{s=12}$

綜合以下結果，我們決定先建模 $\text{SARIMA}(0,0,2)\text{X}(0,1,1)_{s=12}$ 。

b. 參數估計

模型: $\text{SARIMA}(0,0,2)\text{X}(0,1,1)_{s=12}$

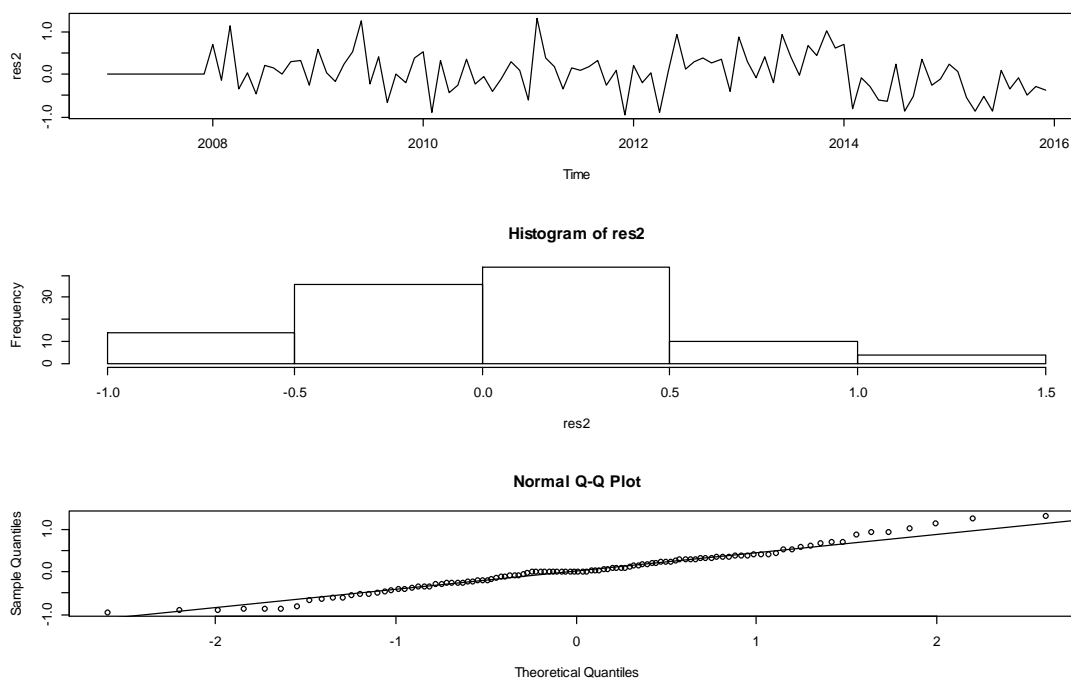
$$(1 - B^{12})Z_t = (1 - \theta_1 B^{12})(1 - \theta_1 B - \theta_2 B^2) e_t \quad e_t \sim \text{wn}(0, 0.2443) \text{ for all } t$$

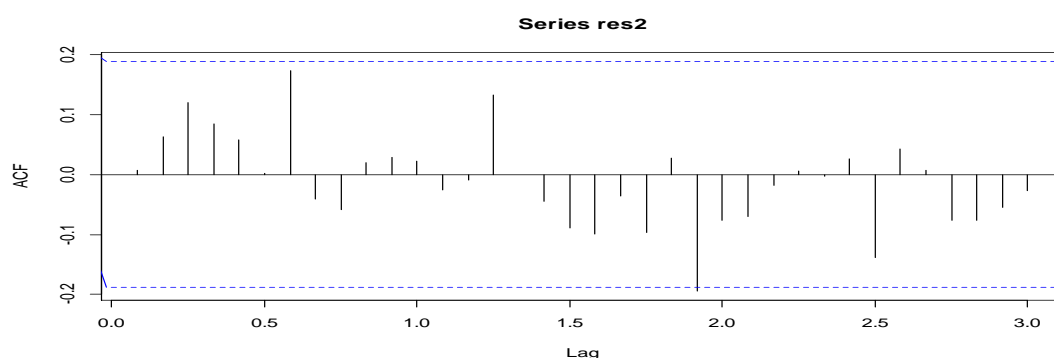
	ma1	ma2	sma1
估計值	0.3362	0.1872	-0.7481
標準差	0.1021	0.0914	0.1343
是否顯著?	是	是	是

σ^2	log likelihood	aic
0.2443	-73.57	153.14

係數皆顯著，因此可以進行殘差診斷。

c. 殘差診斷





(圖九)

由上而下分別為殘差時間序列圖、殘差的直方圖、殘差的常態機率圖及殘差的 acf 圖。

One sample t test	Shapiro-wilk	Box-Ljung test
P-value = 0.4264	P-value = 0.1187	P-value = 0.6741
符合 mean=0	符合 normality	第 23 步沒有相關

(1)殘差的變異數是否為常數

從殘差時間序列圖，可以發現除了因為經過一階季節差分，第一年的預測值無法求得只好設為 0 之外，殘差大致上在 0 上下跳動，沒有特定趨勢。

(2)殘差的平均是否為 0

從殘差直方圖可以發現其呈現以 0 為中心的對稱分佈， <0 那半邊次數似乎比較多一點，但經過 t-test 檢定後仍接受殘差平均為 0。

(3)殘差是否與時間序列無關

殘差的 ACF 圖，可以發現第 23 步很靠近信賴線，但經過 Box-Ljung test 檢定後顯示我們仍可以接受 lag23 的 acf 為 0。

(4)殘差是否為常態分配

殘差的 QQ-plot，可以發現點大多數都落在直線上，且經過 Shapiro-Wilk test，其 $p\text{-value}=0.1187 > 0.05$ ，顯示殘差符合常態分布。

由於此模型殘差的四個性質都符合，因此判斷模型通過殘差診斷。最後選定的完整模型如下：

$$\sqrt{Y_t} = 391.7253 - 0.1918t + Z_t$$

$$(1 - B^{12})Z_t = (1 - 0.7481B^{12})(1 - 0.3362B - 0.1872B^2) e_t$$

$$e_t \sim N(0, 0.2443)$$

d. 其他候選模型的 aic 及殘差診斷結果

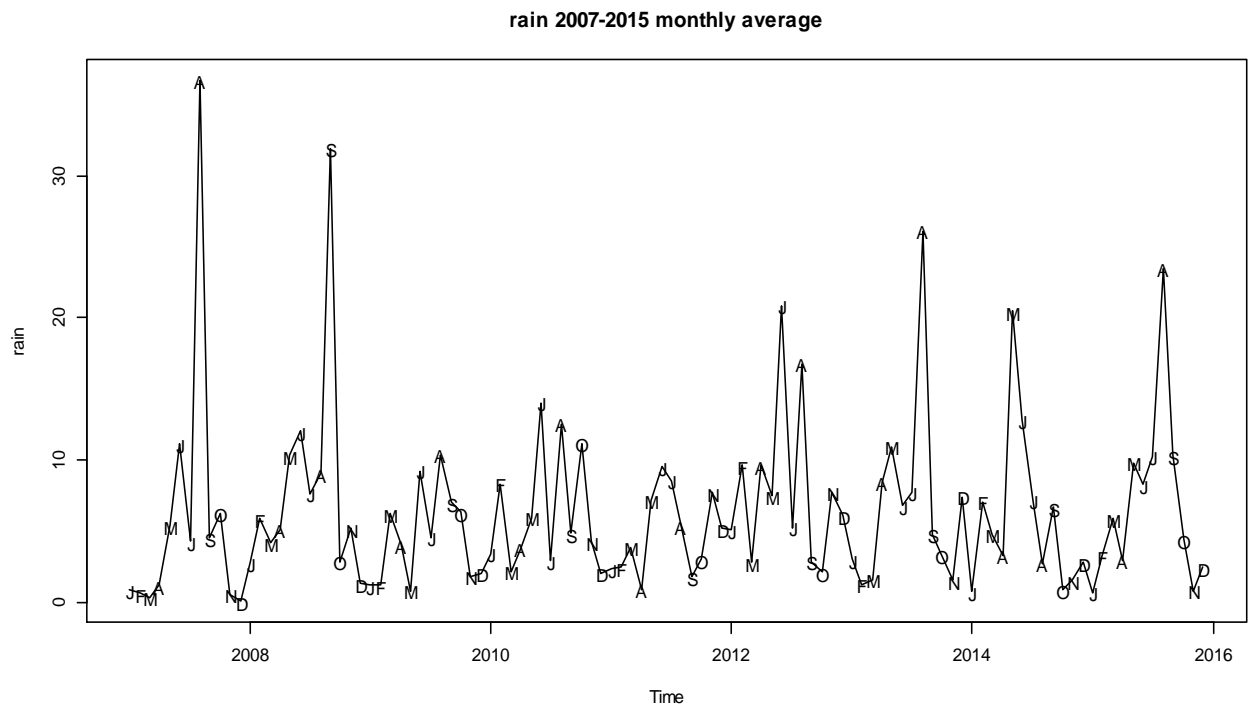
上面過程是呈現 AIC 最小的模型的建模流程，但其實過程中我們也有試了其他模型，下面呈現其他配適模型的 aic 及殘差診斷結果。

$$X_t = \sqrt{Y_t}$$

模型	aic	殘差診斷
$X_t \sim \text{SARIMA}(0,1,1)^*(0,1,1)$	178	有通過
$X_t = Bt + \mu_t + Z_t,$ $\mu_t = B_{1...}B_{12} Z_t = (1 - \theta_1 B) e_t$	296	有通過
$X_t = Bt + \mu_t + Z_t,$ $\mu_t = B_{1...}B_{12} (1 - \phi B) Z_t = e_t$	292	有通過

(二)月平均雨量模型建立

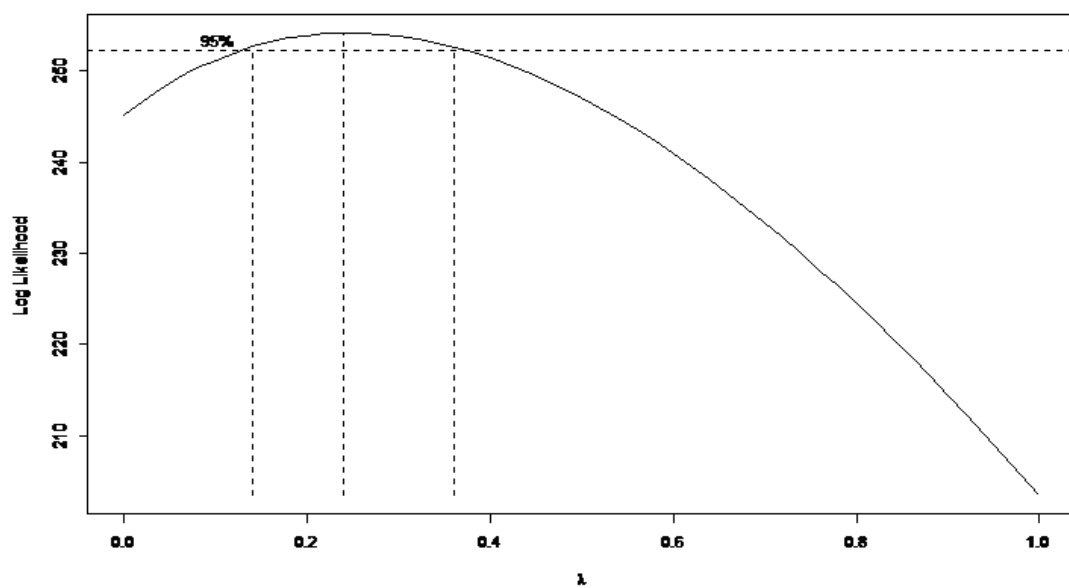
首先，觀察台南市從 2007 年到 2015 年月平均降雨量的時間序列圖。



(圖十)

可以發現此筆時間序列月跟月之間的變異很大，但平均似乎維持在某一常數，並沒有往上升或往下降的趨勢。此序列有明顯的季節趨勢，在夏天時雨量較高，尤其是在 7-9 月雨量會達到高峰，冬天雨量最低。

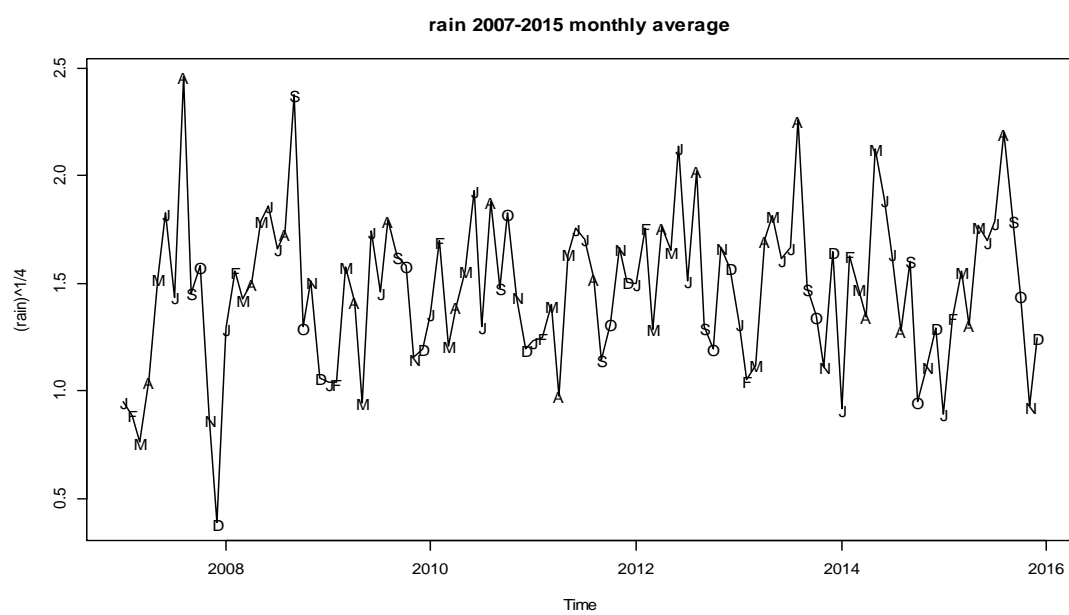
我們先將資料做 Box-Cox 轉換，並且繪製 λ 與對數概似估計值圖。



(圖十一)

可以看到，最大概似估計量 λ 的 95%信賴區間介於約 $[0.1, 0.36]$ ，最大值發生在 $\lambda = 0.25$ 時。因此我們決定將資料做雙重根號轉換。

轉換後的時間序列圖如下



(圖十二)

可以看到月跟月之間變異減少了，沒有往上升或往下降的趨勢，但是仍然有明顯的季節效應。所以我們決定配一次項季節趨勢。

$$Y_t = \mu_t + X_t \quad \mu_t \equiv \begin{cases} B_1 & \text{For } t=1,13,25,\dots \\ B_2 & \text{For } t=2,14,26,\dots \\ \vdots & \\ B_{12} & \text{For } t=12,24,36,\dots \end{cases}$$

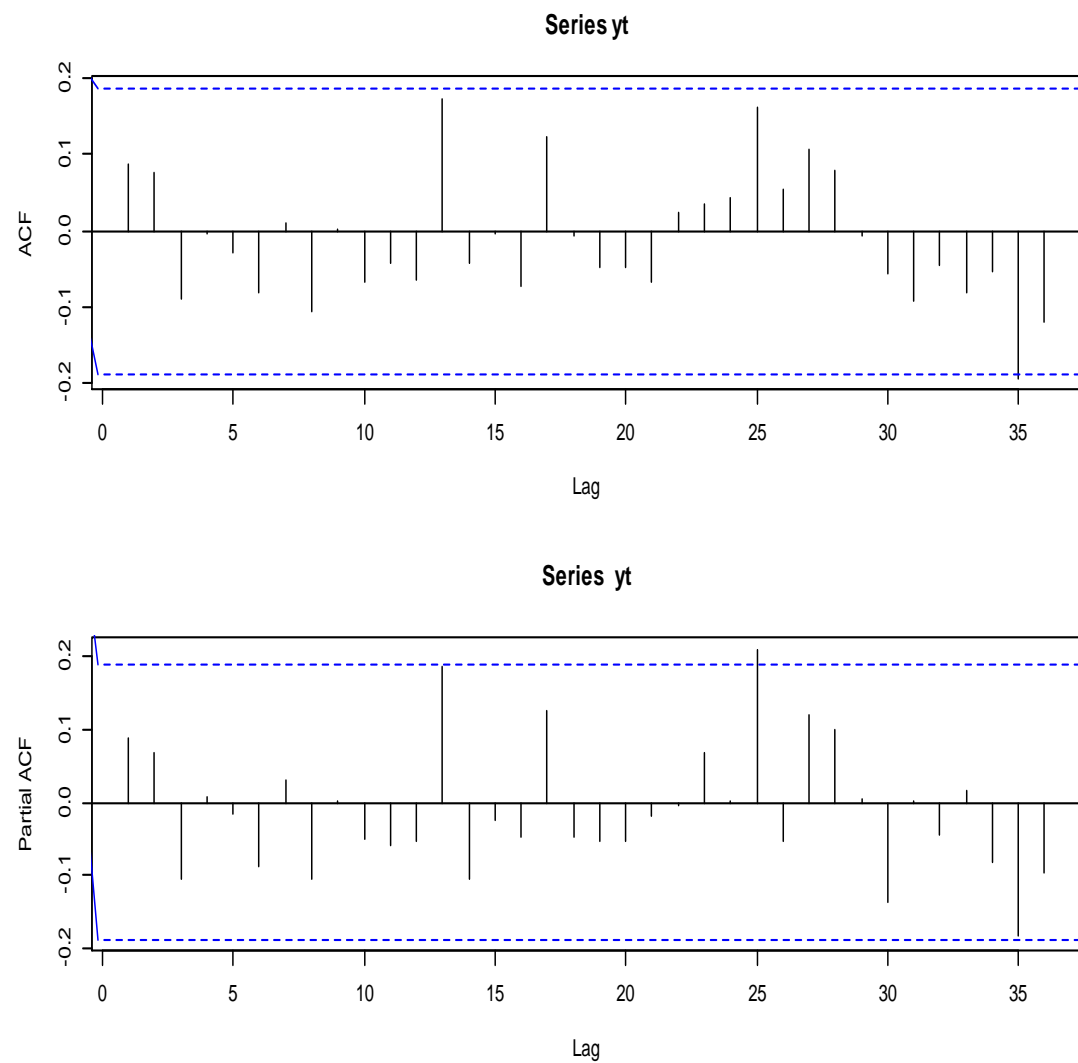
month	estimate	Standard error	P value
January	1.16697	0.09584	<0.0001
February	1.35703	0.09584	<0.0001
March	1.31315	0.09584	<0.0001
April	1.38365	0.09584	<0.0001
May	1.64691	0.09584	<0.0001
June	1.82869	0.09584	<0.0001
July	1.57500	0.09584	<0.0001
August	1.90696	0.09584	<0.0001
September	1.58435	0.09584	<0.0001
October	1.39482	0.09584	<0.0001
November	1.27351	0.09584	<0.0001
December	1.23512	0.09584	<0.0001

σ^2 estimated	R square	Adjusted R square
0.2875	96.79%	96.39%

可以看到季節趨勢項每一項皆顯著，且季節趨勢模型可以解釋序列約 97% 的變異。接下來我們檢查配適完季節趨勢項後的时间序列 $\{X_t\}$ ，進入模型選擇。

先看 $\{X_t\}$ 的 acf 、 pacf 、 eacf 圖。

a. 模型判定



(圖十三)

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	o	o	o	o	o	o	o	o	o	o	o	o	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o
2	x	x	o	o	o	o	o	o	o	o	o	o	o	o
3	o	o	x	o	o	o	o	o	o	o	o	o	o	o
4	x	o	x	o	o	o	o	o	o	o	o	o	o	o

(圖十四)

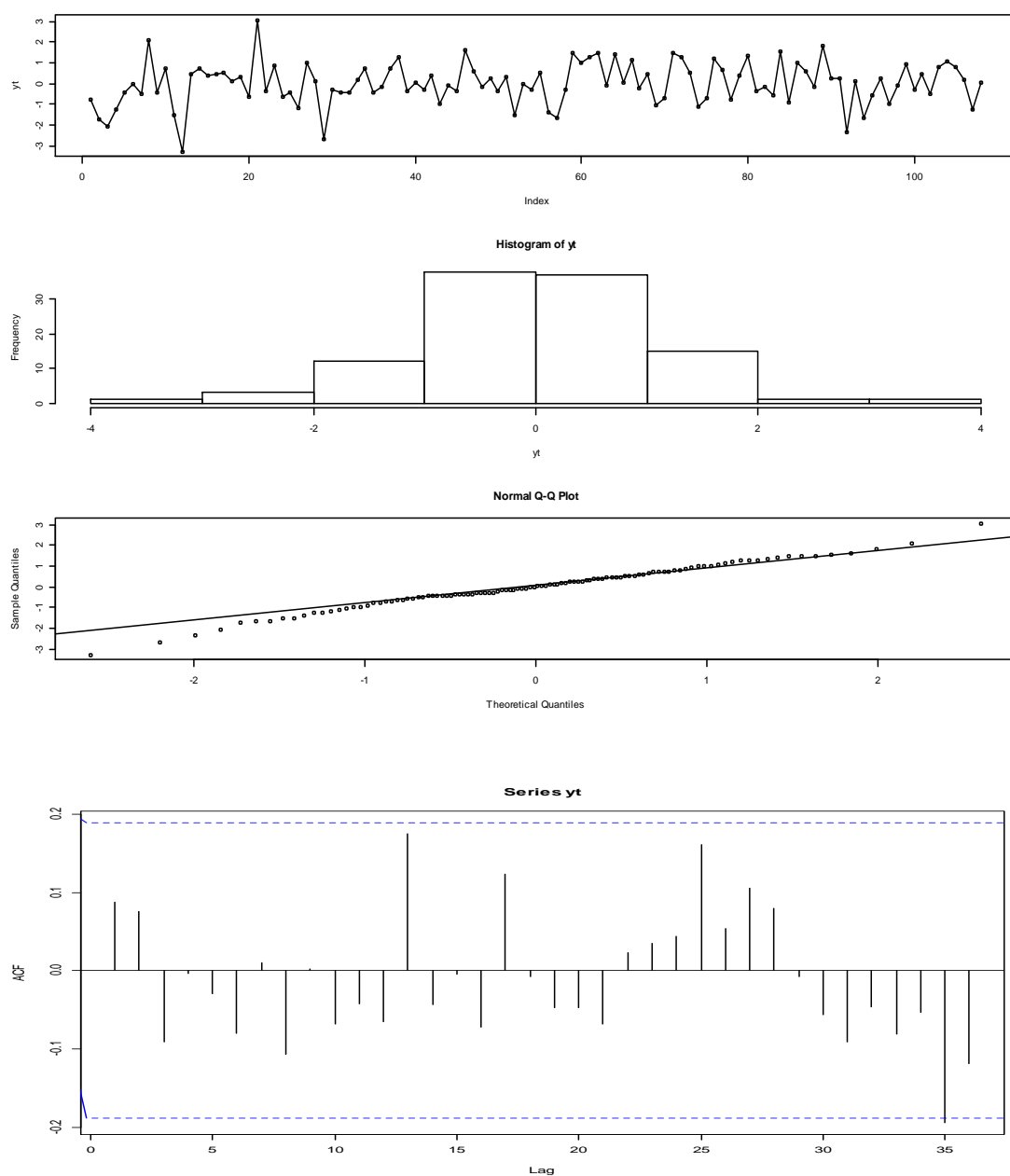
從 acf 跟 pacf 圖可以看到，除了 acf 圖中的 lag35 和 pacf 圖中的 lag25，acf 跟 pacf 值都在信賴線，但 MA(35)及 AR(25)步數太大，不考慮。另外檢查 eacf，也同樣得到 ARMA(0,0)的結論。因此從 acf、pacf、eacf 圖，我們得出共同結論，時間序列 $\{X_t\}$ 為白噪音。

b. 參數估計

$$X_t \sim wn(0, 0.2875)$$

c. 殘差診斷

進一步 我們檢查 $\{X_t\}$ 是否通過符合下面幾個必要條件。



(圖十五)

由上而下分別為 $\{x_t\}$ 時間序列圖、 $\{x_t\}$ 的直方圖、 $\{x_t\}$ 的常態機率圖及 $\{x_t\}$ 的 acf 圖。

One sample t test	Shapiro-wilk normality	Box-Ljung test
P-value = 0.9887	P-value = 0.4485	P-value = 0.2511
符合 mean=0	符合 normality	第 35 步沒有相關

(1)殘差的變異數是否為常數

從殘差時間序列圖，可以發現殘差大致上在 0 上下跳動，沒有特定趨勢。

(2)殘差的平均是否為 0

從殘差直方圖可以發現其呈現以 0 為中心的對稱分佈，且經過 t-test 檢定後接受殘差平均為 0。

(3)殘差是否與時間序列無關

殘差的 ACF 圖，可以發現第 35 步很靠近信賴線，但經過 Box-Ljung test 檢定後顯示我們仍可以接受 lag35 的 acf 為 0。

(4)殘差是否為常態分配

殘差的 QQ-plot，可以發現點大多數都落在直線上，且經過 Shapiro-Wilk test，其 p-value=0.4485 > 0.05，顯示殘差符合常態分布。

由於此模型殘差的四個性質都符合，因此判斷模型通過殘差診斷。最後選定

的完整模型如下：

$$\sqrt[4]{Y_t} = \mu_t + X_t \quad \mu_t = \begin{cases} B_1 & \text{For } t=1,13,25,\dots \\ B_2 & \text{For } t=2,14,26,\dots \\ \vdots & \\ B_{12} & \text{For } t=12,24,36,\dots \end{cases} \quad X_t \sim N(0,0.2875)$$

B_1	B_2	B_3	B_4	B_5	B_6
1.16697	1.35703	1.31315	1.38365	1.64691	1.82869
B_7	B_8	B_9	B_{10}	B_{11}	B_{12}
1.57500	1.90696	1.58435	1.39482	1.27351	1.23512

d. 其他候選模型的 *aic* 及殘差診斷結果

除了配適上面的季節趨勢項模型外，其實我們也有配 SARIMA(0,1,1)-用季節差分後的時間序列去配 SMA(1)，此模型預測表現跟我們最後選定的季節趨勢模型差不多，通過殘差檢定，也有顧到季節性。原本應該可用 *aic* 去比較兩模型孰優孰劣，但因為季節趨勢項模型殘差為白噪音，沒有 *aic*，所以我們用預測均方差(MSPE)來做判斷。它的原理是先去掉一年資料，用剩下資料分別配適兩模型，分別預測一年的資料，再計算真實值跟預測值的均方差，哪個小我們就選定哪一個，最後得到的結果如下：

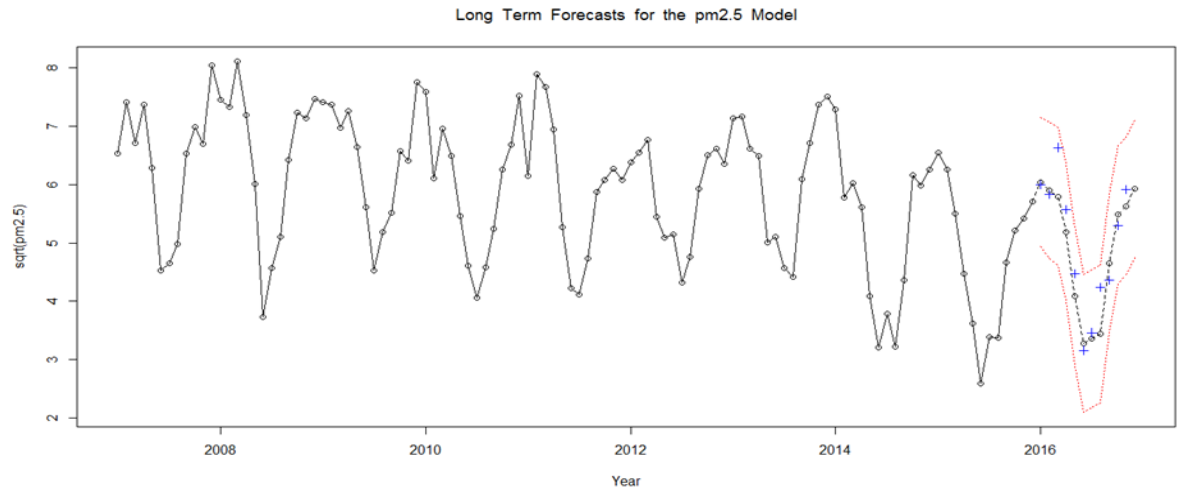
模型	MSE	殘差診斷
$\sqrt[4]{Y_t} \sim \text{SARIMA}(0,1,1)$	2.011366	有通過
我們最終選定的季節趨勢項模型	1.863035	有通過

因此最後我們選定季節趨勢項的模型。

伍、模型預測

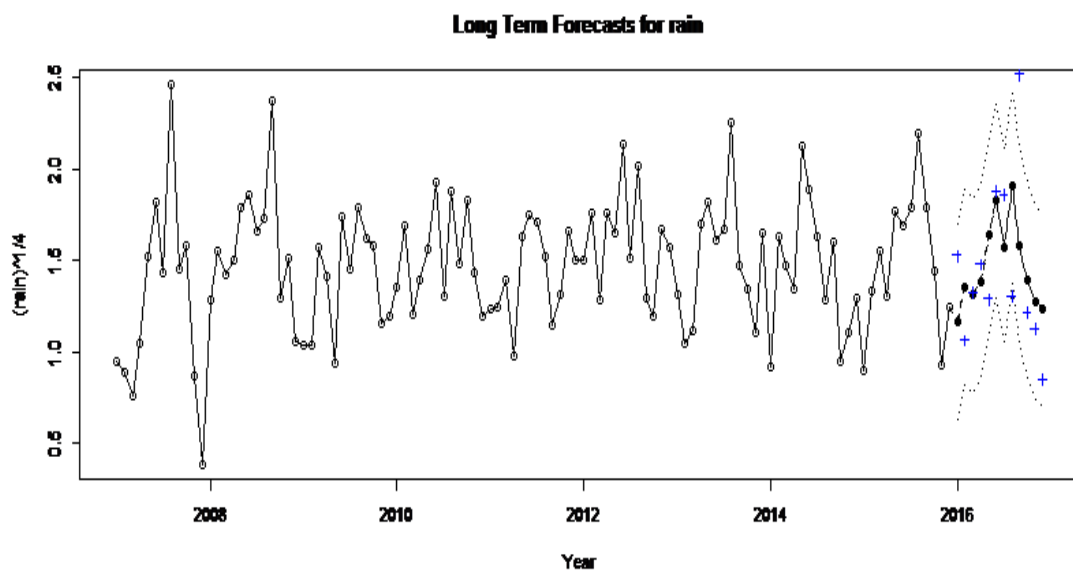
以下是經過轉換後的 pm2.5 與雨量 2016 年每月之預測值，因為此份報告

主題是探討相關性，預測不是最重要的，故不須轉換回原始值。



(圖十六) 2016 年台南 pm2.5 濃度月平均預測圖

由預測圖可知，2016 年的實際觀測值和配適出來的預測值差異不大，甚至有幾個預測值落在觀察值點上。大部分預測值皆落於模型 2 倍信賴區間內，故此模型的預測能力佳。

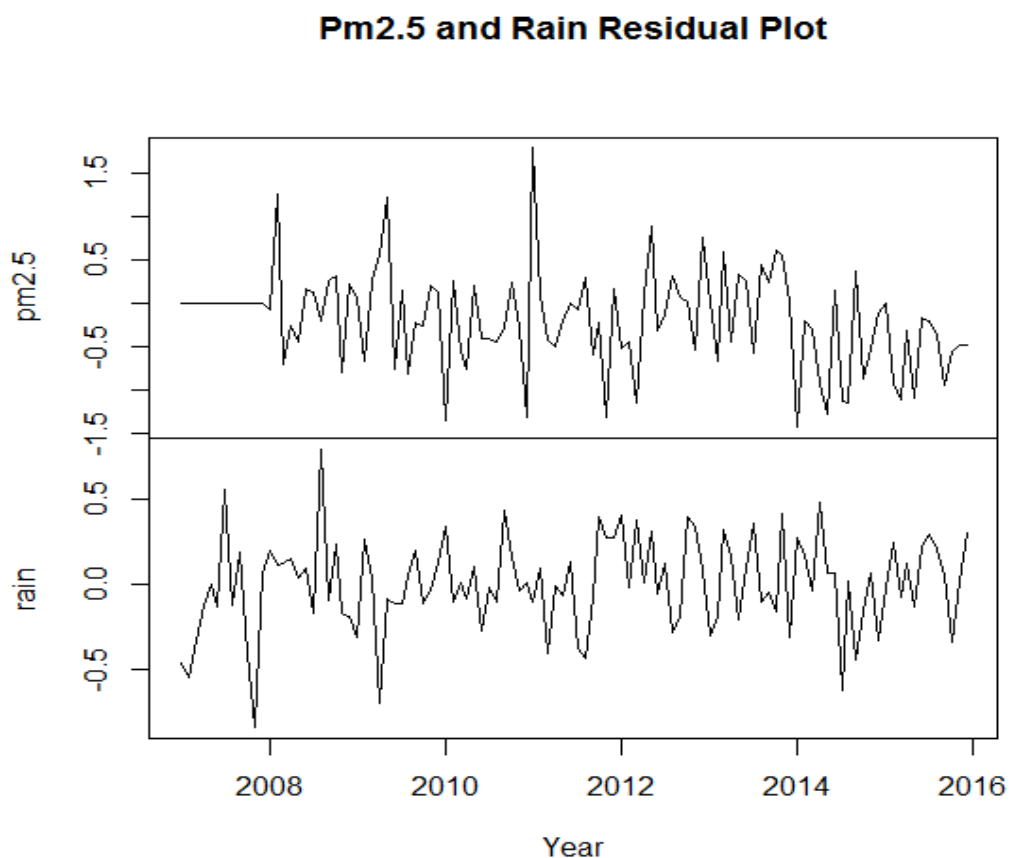


(圖十七) 2016 年台南雨量月平均預測圖

根據預測圖可知，2016 年的實際觀測值和模型配適出來的預測值差異不大，僅有 9 月可能因為接連來三個颱風，觀測值沒有落於預測值的 2 倍信賴區間內之外，其他預測值都落於 2 倍信賴區間內。故此模型的預測能力佳。

陸、相關性分析

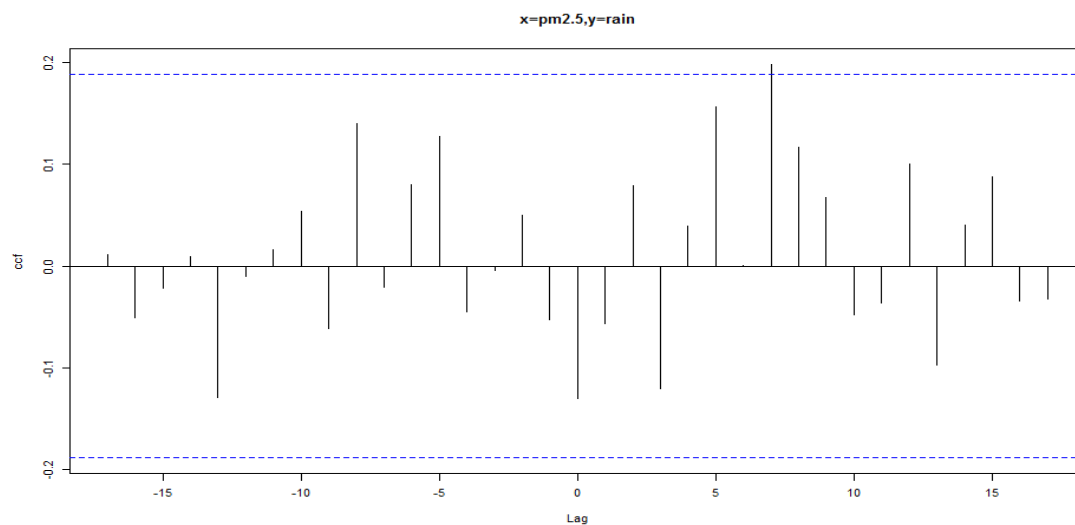
(一)比較殘差時間序列圖



(圖十八)pm2.5 與雨量的殘差時間序列圖

由此圖可以看出 pm2.5 與雨量的殘差跳動的趨勢似乎不同，相關性不大，且兩序列都沒有極端值出現。

(二) 分析兩個模型殘差的 CCF



(圖十九)

以 pm2.5 為 X，降雨量為 Y，差距為 k 期的相關係數。從 CCF 圖發現在正負第 7 期值稍微超出信賴線一點，代表現在的雨量與 7 個月後的 pm2.5 有正相關。雖然感覺不太合理，因為照理說雨量多，pm2.5 濃度會減少，但我們還是按照 ccf 圖結果檢驗看看。

經過 d=7 移動，pm2.5 月均濃度和月均雨量的殘差時間序列

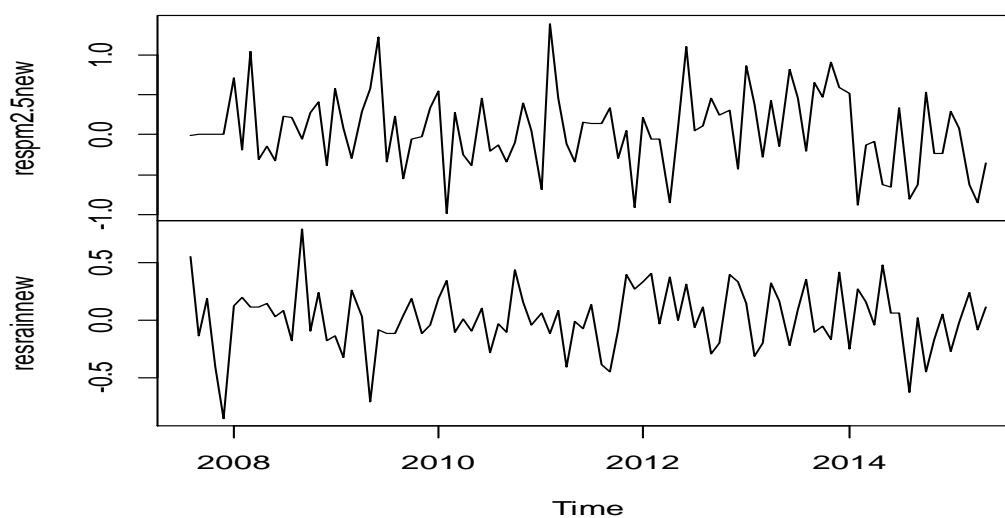


圖 (二十)

雨量殘差維持不變，pm2.5 殘差往後平移 7 個月的兩模型殘差時間序列圖中，

pm2.5 與雨量的殘差跳動的趨勢看起來沒有明顯的一致性。

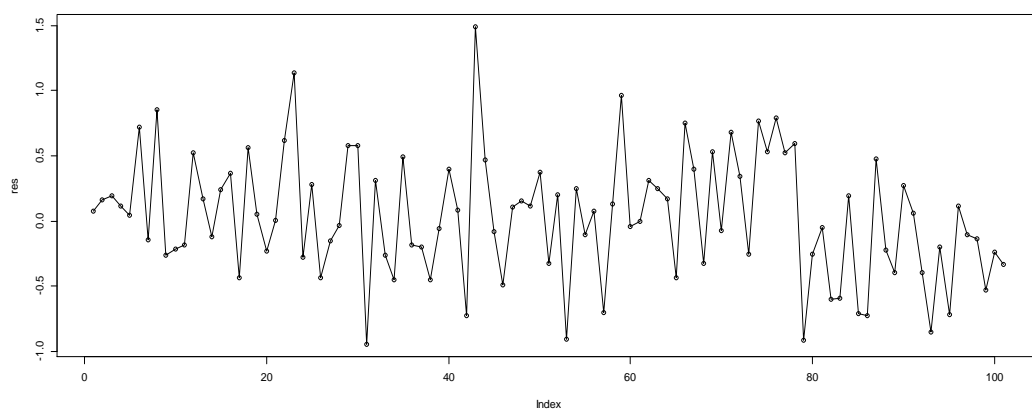
(三)對移動後的殘差配適迴歸模型

	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	0.03183	0.04735	0.672	0.5030
降雨量	0.35459	0.17236	2.057	0.0423
Adjusted R-squared: 0.041		Standard error: 0.4759		

	Estimate	Std. Error	t-value	Pr(> t)
降雨量	0.3533	0.1719	2.056	0.0424
Adjusted R-squared: 0.04054		Standard error: 0.4745		

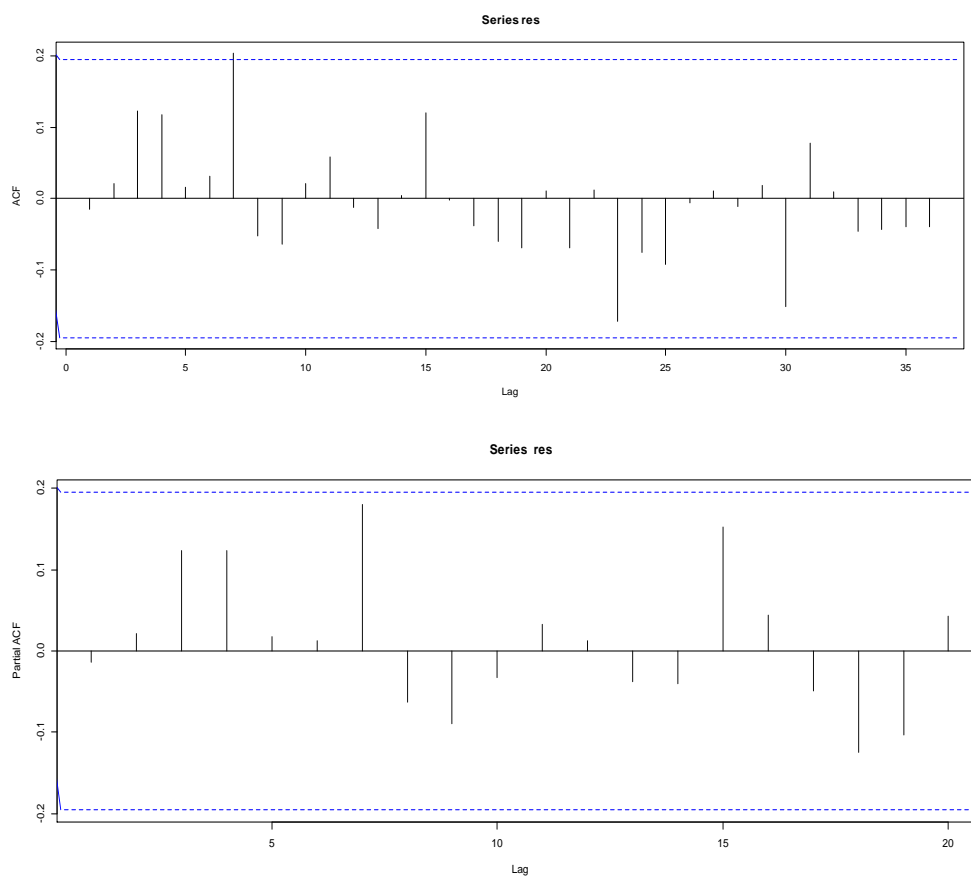
迴歸模型截距項係數不顯著，一次項係數稍顯著(type I error rate=0.05 時顯著)，可建立迴歸模型。係數為 0.3533，符合 ccf 圖顯示此兩序列為正相關的結果。照理來說，降雨量應該不會影響 7 個月後的 pm2.5 濃度，也不會呈現正相關，但我們還是接受模型配適結果。但要注意，兩者之間相關性不高，係數也並沒有很顯著。且此模型雨量只能解釋 pm2.5 濃度不到 5%的變異。

(四)模型配適

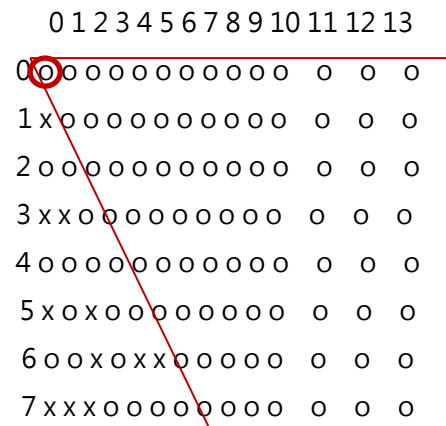


(圖二十一)殘差的時間序列圖

此序列模型並沒有特定趨勢，隨機跳動，彼此之間看起來沒有關係。



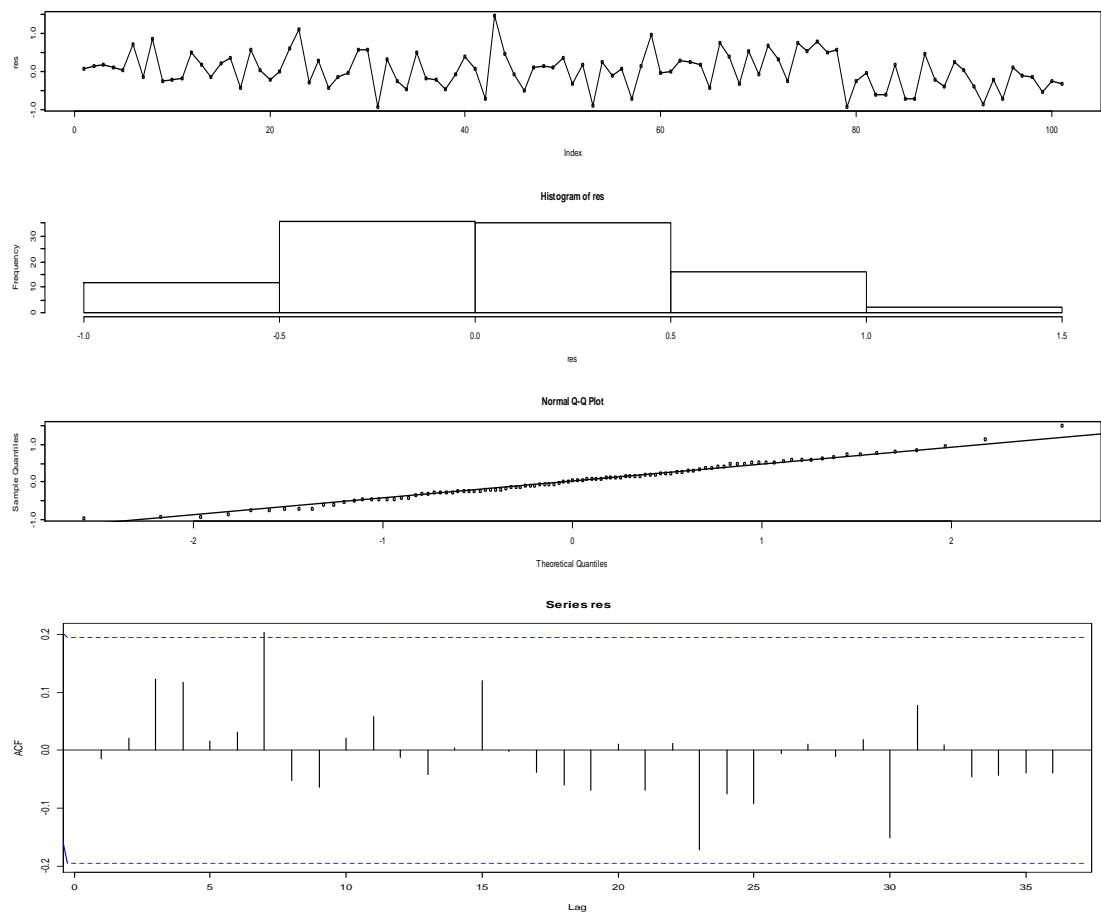
(圖二十二)殘差的 acf 及 pacf 圖



(圖二十三)殘差的 eacf

殘差的 acf、pacf、eacf 圖皆顯示此模型的殘差為白噪音。

(五)殘差診斷



由上而下分別為殘差時間序列圖、殘差的直方圖、殘差的常態機率圖及

殘差的 acf 圖。

One sample t test	Shapiro-wilk	Box-Ljung test
P-value = 0.5009	P-value = 0.7714	P-value = 0.3408
符合 mean=0	符合 normality	第 7 步沒有相關

(1)殘差的變異數是否為常數

從殘差時間序列圖，殘差大致上在 0 上下跳動，沒有特定趨勢。

(2)殘差的平均是否為 0

從殘差直方圖可以發現其呈現以 0 為中心的對稱分佈，<0 那半邊次數似乎比較多一點，但經過 t-test 檢定後仍接受殘差平均為 0。

(3)殘差是否與時間序列無關

殘差的 ACF 圖，可以發現第 7 步很靠近信賴線，但經過 Box-Ljung test 檢定後顯示我們仍可以接受 lag7 的 acf 為 0。

(4)殘差是否為常態分配

殘差的 QQ-plot，可以發現點大多數都落在直線上，且經過 Shapiro-Wilk test，其 p-value=0.5009 > 0.05，顯示殘差符合常態分布。

由於此模型殘差的四個性質都符合，因此判斷模型通過殘差診斷。最後選定的完整模型如下：

$$e_t = 0.3553X_{t-7} + \varepsilon_t \quad \varepsilon_t \sim N(0, 0.4857)$$

e_t : pm2.5 模型的殘差 X_t : 雨量模型的殘差

ε_t : 回歸模型的殘差

(六) 比較原始資料及配適資料相關性圖

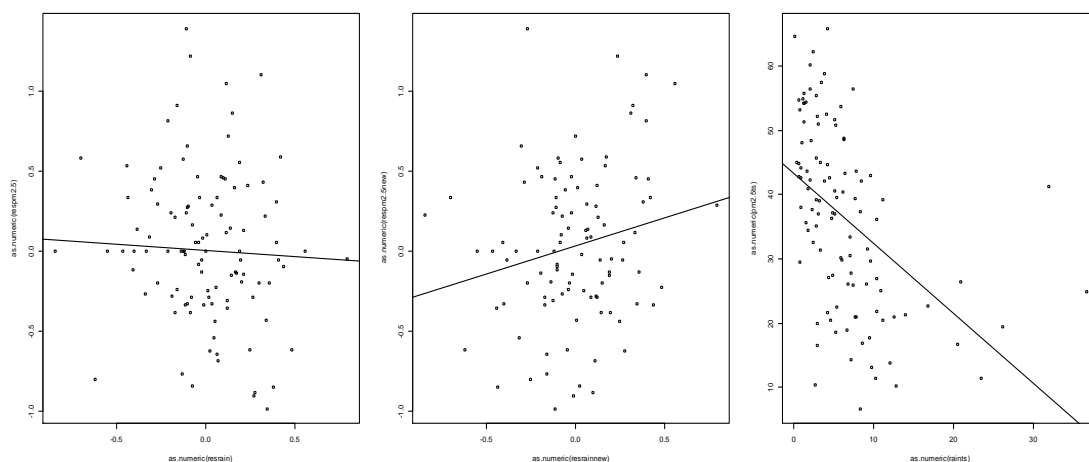


圖 (二十四)

(右)原始資料 pm2.5 和雨量散布圖 (中)pm2.5 往後移動 7 步後和雨量殘差散布圖

(左)pm2.5 和雨量殘差散布圖

原始資料表示 Pm2.5 和降雨量呈現明顯的負相關，相關係數=-0.47。但考慮時間效應後，透過殘差相關性檢驗及建立迴歸模型過程，發現當月的雨量其實並不會影響當月的 pm2.5 濃度；且用殘差繪製的雨量、pm2.5 散布圖看起來沒有特定的趨勢，相當分散，相關係數只有約-0.15。而將 pm2.5 往後移動 7 個月後的兩序列相關係數也只有約 0.2，為正相關，並非負相關。由此可知，兩筆資料皆有時間效應存在，若用原始資料看相關性，會得到虛偽相關，就如同很多中小學科展的做法一樣，會得到錯誤的結果。

柒、結論

根據 2007 年到 2015 年 pm2.5 及降雨量的資料，分析後得到以下結果：

- Pm2.5 模型：

$$\sqrt{Y_t} = 391.7253 - 0.1918t + Z_t \quad e_t \sim N(0, 0.2443)$$

$$(1 - B^{12})Z_t = (1 - 0.7481B^{12})(1 - 0.3362B - 0.1872B^2)e_t$$

- 降雨量模型：

$$\sqrt[4]{Y_t} = \mu_t + X_t, \quad \mu_t = \begin{cases} \beta_1, t = 1, 13, 25, \dots \\ \beta_2, t = 2, 14, 26, \dots \\ \vdots \\ \beta_{12}, t = 12, 24, 36, \dots \end{cases}, \quad X_t \sim N(0, 0.2875)$$

B_1	B_2	B_3	B_4	B_5	B_6
1.16697	1.35703	1.31315	1.38365	1.64691	1.82869
B_7	B_8	B_9	B_{10}	B_{11}	B_{12}
1.57500	1.90696	1.58435	1.39482	1.27351	1.23512

- Pm2.5 及降雨量的預測

大部分原始資料皆落在預測模型的 2 倍信賴區間內，但雨量預測圖中，

2016/9 月值超出信賴線，可能因那月有三個颱風，導致實際數值超出 2 倍信賴區間。

- Pm2.5 及降雨量相關性模型：

$$e_t = 0.3553X_{t-7} + X_t$$

$$X_t \sim N(0, 0.4857)$$

e_t : pm2.5 模型的殘差 X_t : 雨量模型的殘差

ε_t : 回歸模型的殘差

雖然很多媒體報導，還有很多中小學科展都認為降雨量及 pm2.5 是負相關、當月降雨量會使當月 pm2.5 濃度變低等等，但是根據 2007 年到 2015 年資料分析的得知：其實當月降雨量不會影響當月 pm2.5 濃度，反而是當月雨量會影響七個月後 pm2.5 濃度，而且還是正相關。雖然最後結果是正相關，但其實回歸模型係數只有稍微顯著，相關係數也只有約 0.2，且回歸模型只能解釋不到 5% 的變異，所以其實 pm2.5 跟雨量之間的相關性還是很低的。

此份分析結果是用 2007~2015 年資料分析得到的結果，其中中央氣象局提供的觀測值可能有誤差，或是清理資料的時候把一些重要的關鍵點刪掉(但因為儀器檢測無效值過多，也沒有辦法)。總之，有時間效應的模型，如果直接看原始資料的相關性，很容易得到虛偽相關，進而導出錯誤結論，所以一定要建立時間序列模型後，才能用殘差進行相關性分析。

捌、感想

(1)育婷:

一開始選修時間序列分析這堂課，只是想到迴歸分析有提到：有時間效應的資料分析法會在大三此堂課教，所以想說多學一種資料分析法，沒想到變成我這學期最充實的課。老師一開始介紹理論，教導我們模型背後的原因及計算若干統計量的方法等等；看似無聊，但其實最重要。理論打底可讓我們不是只會表面功夫，在遇到非傳統時間序列或是變化多端的問題時，靠著這些基礎，我們可以想出應變方式，而不是只會依樣畫葫蘆。後面的運用，讓我們知道了吸收大量理論的意義，並且能將此技能運用在實際生活、改善問題，對我的幫助超乎我想像。

此份報告，讓我實際做了一份完整研究，從一開始想題目、找尋資料、篩選及清理資料、寫程式、建立模型、預測、找尋相關性、解釋程式跑出的結果及目的、最後寫成報告。此份報告讓我訓練做一份研究所需要的所有能力，每個環節都在訓練我不同方面的技術，並且是做我覺得很有意義的主題。謝謝老師的教導及訓練，此份機會不可多得，難能可貴！

(2)珍妮:

本來修這堂課沒有任何目的及期待。但是接近期末考回顧這學期所修的課才真正發現這堂課比起其他課收穫最多。透過每個禮拜的作業及這次報告讓我更了解如何在實務上運用所學的東西，並真的去了解我做出來的東西是甚麼，不只是單純照課本或課堂上所教的東西去套用就結束。

其實我很想做匯率方面的報告，但是因為之前學長姊都有做了所以無法再做。

做報告過程最難的部份是在選擇模型的時候。因為模型有很多可以用，但選擇最適合的模型很難。有的模型預測的比較好，有的模型 AIC 比較小。建立模型才是這次報告的挑戰，因為應用了這學期所學到的東西。

(3)麗娟:

整個上課時間都需要不斷地抄筆記，可能每個人的讀書方式不一樣，我自己吸收比較慢，所以無法真正領悟時間序列分析其中的奧妙。但是，整體來說，此課程對統計系的我們是非常重要的一項課程，可以訓練我們的邏輯思考還有分析能力，是我少數覺得很有用的課程，有點慶幸自己有選到該課程。

覺得老師的配分方式不錯合理，包含 1 次考試成績，作業分數，期末報告成績。作業的分標準與其他老師不一樣，從 12 份作業中抽取幾份分數最高的作業成績，還有團隊作業分數，讓我們互相配合，大家可減少功課題數，更快更有效完成作業，因為自己寫作業影響整個小組的分數，所以自己寫時也會謹慎一些。

玖、 建議

- 建議每個禮拜還是有作業，因為有作業我們才會每個禮拜複習當週上的課。
- 期中考後的作業還是自己做比較好。這樣大家才會學到一樣的份量。如果作業分工的話有的東西沒有學到很可惜。如果要減輕作業量，建議減少作業題數。每個禮拜的作業如果，分組的話，平均一個人作業只寫一題，分組做感覺大家學的東西不均等，有的會漏掉。

拾、引用

- 行政院環保署-空氣品質監測網

<http://taqm.epa.gov.tw/taqm/tw/YearlyDataDownload.aspx>

- 交通部中央氣象局-每日雨量

<http://www.cwb.gov.tw/V7/climate/dailyPrecipitation/dP.htm>

- 原始資料整理方法的報告依據

<http://www.marcopolo-panda.eu/wp/wp-content/uploads/2015/06/>

[AirINFORM_D2.3_API_Evaluation_Report_part3_cn.pdf](#)