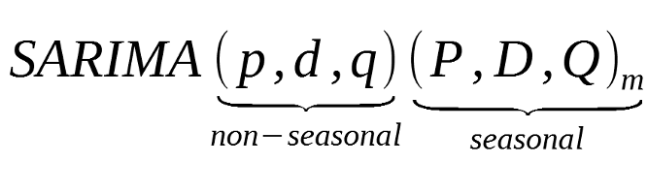
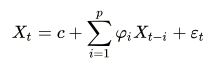
SARIMA: Autoregressive Integrated Moving Average model



1. AR:



函式代表當前的X包含了從 t 開始最多到 t - p的資訊。

p : 為Autoregressive 的階數，代表當前的X和前p個歷史值有關。

1. Integrated :

d : 為使資料平穩而做的差分的次數。

1. MA:



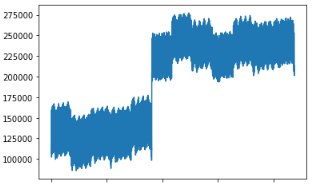
Mu為序列平均，θ1, ... θ*q* 為系統參數，epsilon t ~ t-q 為系統的白雜訊。

q : 為Moving Average的階數，代表當前的y 和前q 個Autoregressive 預測之誤差有關。

(4)對時間序列的季節性部分加入P、D、Q，他們與非季節性部分的用途相同，但他們代表季節性的變化。

m : 為每一段時間的觀察次數，我們的資料為每小時一筆的交通量，將其設置為m = 24。

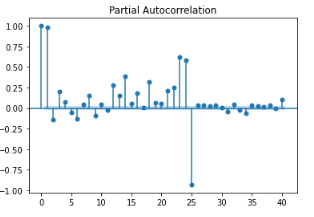
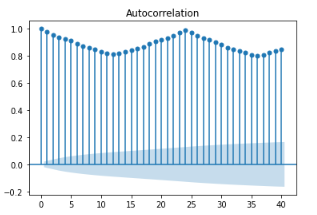
我們的資料為108年某路段(這裡以01F0467N~01F0413N為例)之重型車的交通量或車速(單位為每小時)，所以總共有365\*24 筆資料。為方便舉例，我們將訓練資料和測試資料個別分為108年1-10月((365-61)\*24筆)和108年11、12月(61 \* 24筆)。



01F0467N~01F0413N 108整年的重車交通量(單位:小時)

如果時間序列資料是平穩的，就不需要對資料做差分(d = 0)。而資料是否平穩需透過Autocorrelation Function 和 Partial Autocorrelation Function 觀察。

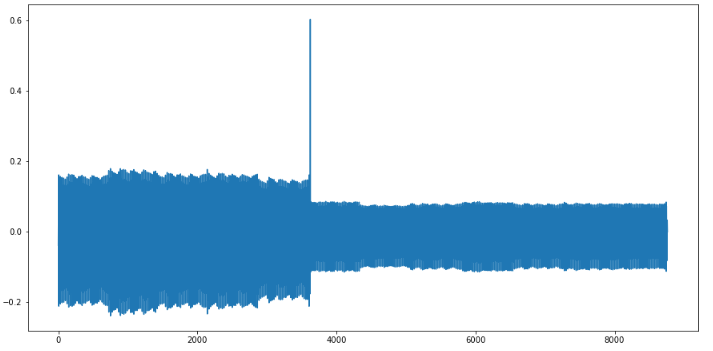
d = 0:

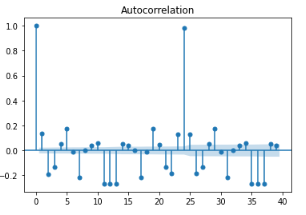
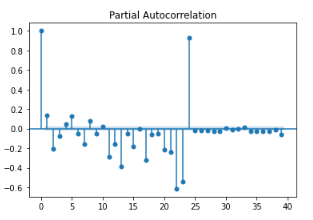


可以看到自相關圖拖尾現象嚴重，必須差分。

因為資料值過大，為使其更加平穩我們先對他取自然對數再做差分。

d = 1:

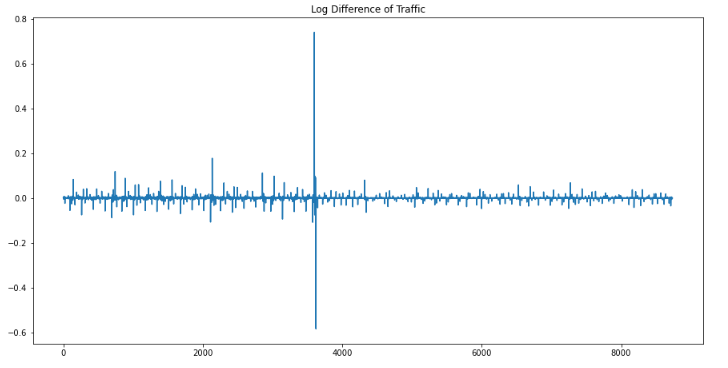


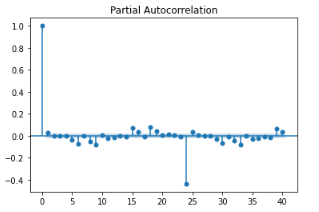
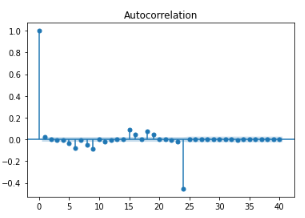
 

自相關圖拖尾現象還是存在，還不是很穩定。

於是我們將時間分成每一天為一個週期做差分。(m = 24)

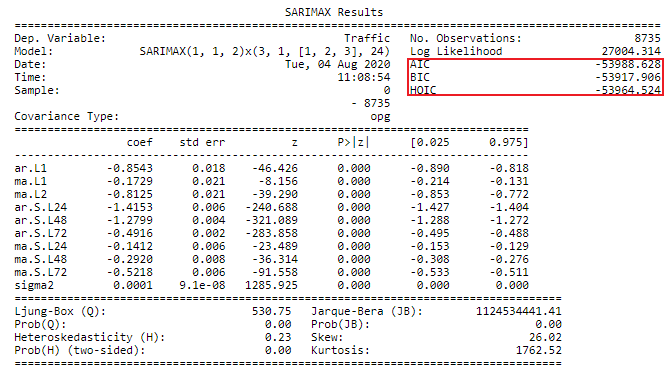
d = 1 ， m = 24:

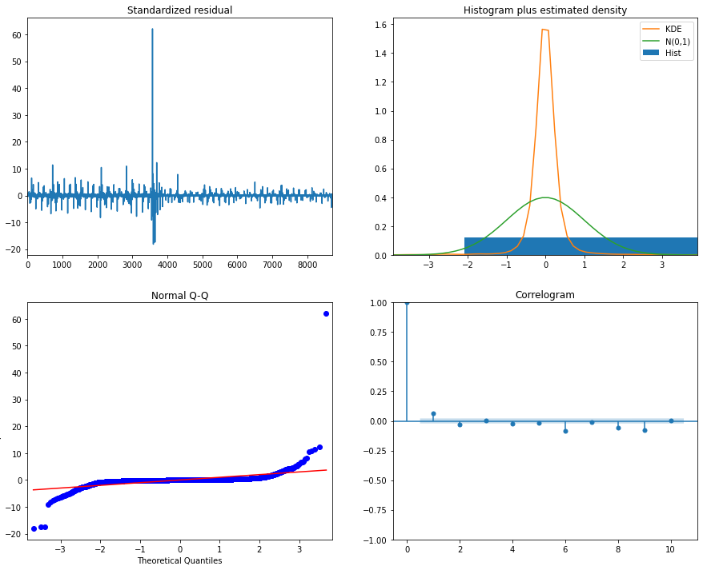


  
可以看自相關圖托尾拖了幾個階數後再lag = 26後結尾，是相當好的結果。而偏相關係數再 lag = 3後落入藍色背景區間內，所以MA階數(q)可以選擇3或者是2。

利用這種方法判斷模型好壞需要花費太多時間，且結果不精確，我們可以透過auto\_arima 設定p、q、d階數上限，且用赤池資訊準(Akaike Information Criterion)、貝葉斯資訊準則(Bayesian Information Criterion)和QAIC方法判斷模型好壞。

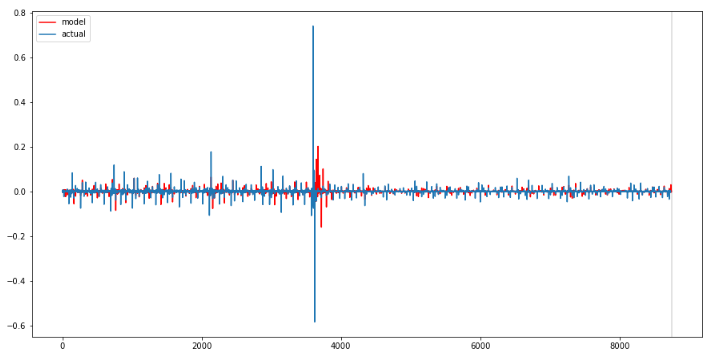
最後我們選擇的模型為SARIMA(p，q，d)(P，Q，D)m = SARIMA(1,1,2)(3,1,3)24



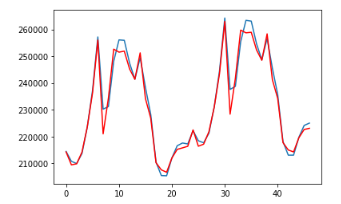


從右上的密度圖和左下方的Q-Q 圖可以看出我們得到的結果幾乎為直線，表示系統符合正態分布右下方的相關圖表明殘差中不存在自相關，因此它們實際上是白噪聲。

預測結果:



將108 11/1~2資料還原:



108 11、12月預測結果:

