

### 實驗目的及探討：

1. 探討感測器對 pm2.5 的精確度
2. 分析感測器資料上的一致性
3. 感測器使用問題與耗電問題

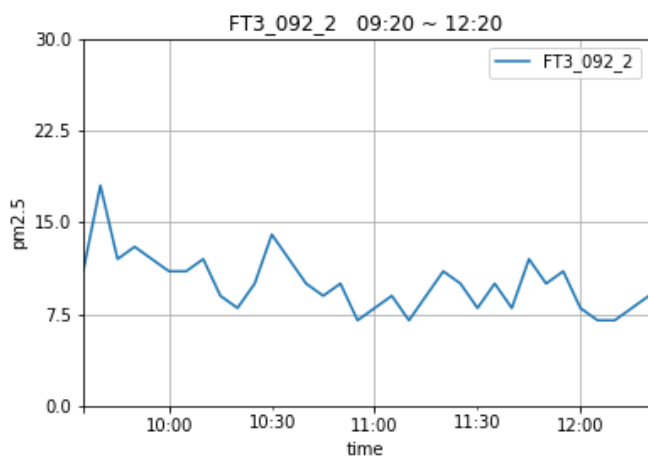
### 實驗（一）：

時間：3/29 01:10 ~ 12:20，共連續約12小時

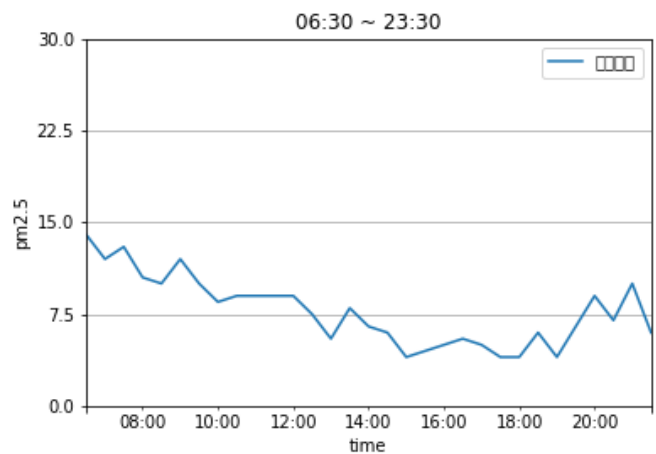
實驗場所：居家陽台（一樓），附近環境多為綠地及公園，方圓四百公尺內皆無廟宇及工廠

實驗當日天氣：降雨量 0 毫米，溫度介於 27 至 24 度，附近無燃燒事件發生或人為污染源

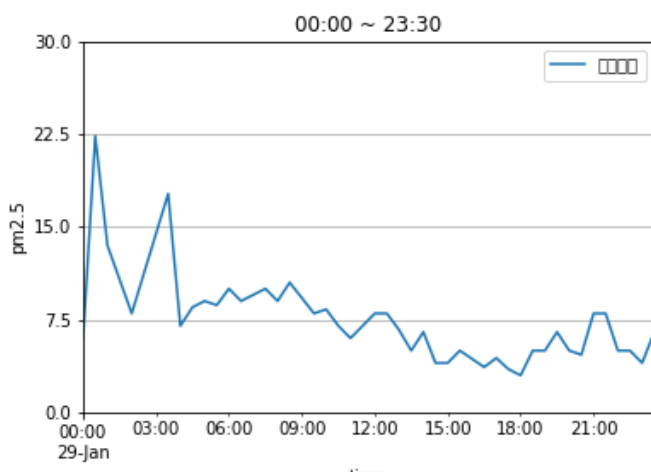
離實驗地點最近的 AirBox 有兩台分別為萬興國小（28C2DDDD423C，距離約八百公尺）及木柵國小（28C2DDDD4354，距離約一公里）內測站還有一台 LASS 感測器 ID：FT3\_092\_2（距離約八百公尺），以下為當日上述三台感測器之變化，藉此表示當日由早至晚附近並無長時間嚴重污染



FT3\_092\_2 上午九點二十至下午十二點半之變化

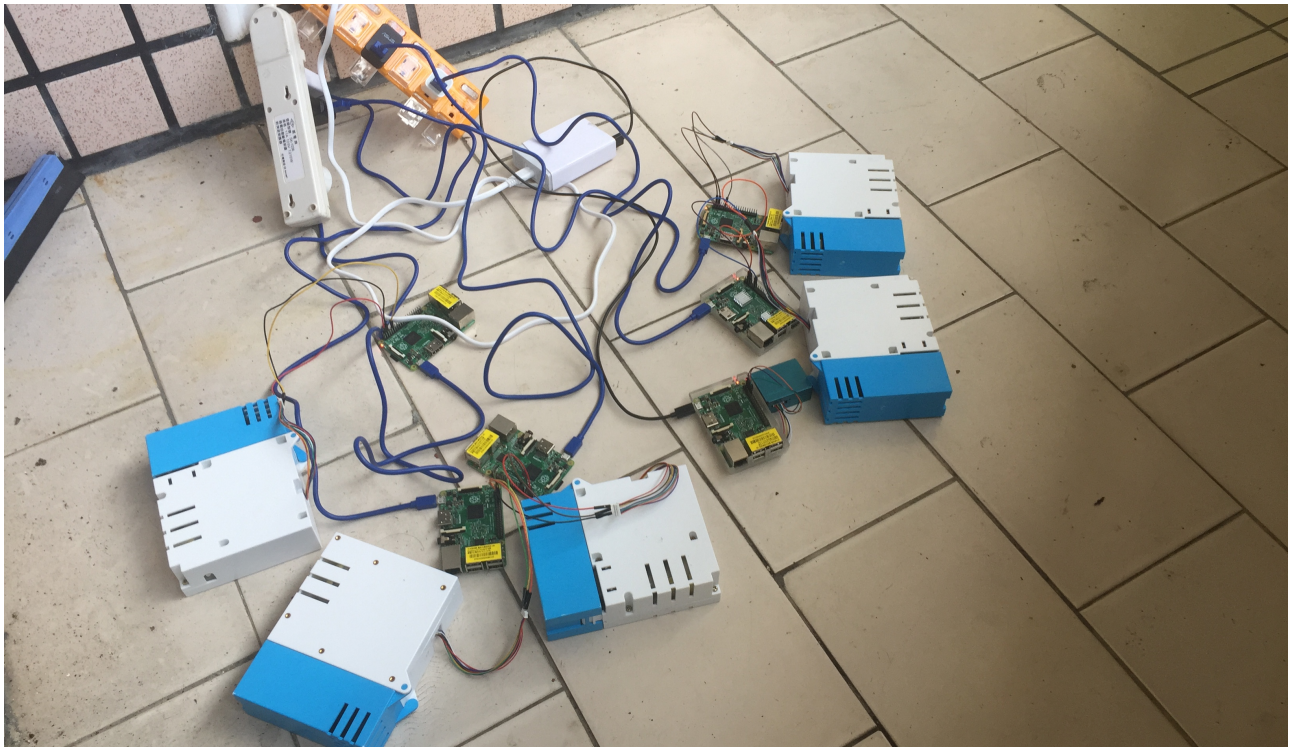


木柵國小 上午六點半到晚上十一點之變化



萬興國小 凌晨到晚上十一點半之變化

實驗進行時之拍攝：



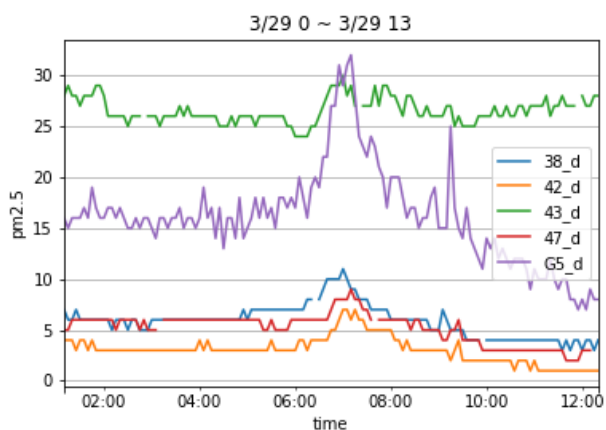
位於一樓陽台並靠近馬路

實驗方式：

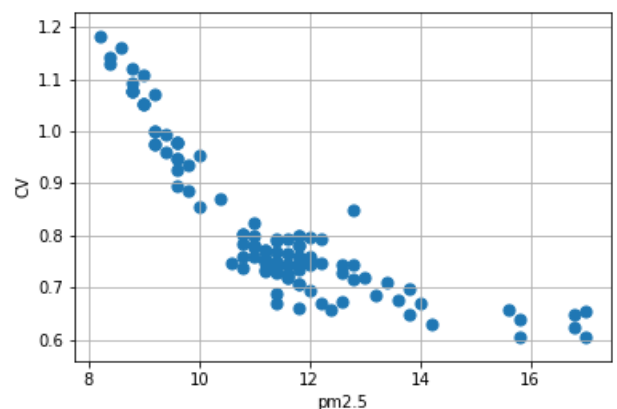
隨機取出四台感測器（編號分別為：38、43、42、47）與一台 G5 感測器作實驗，並以 G5 感測器作為標準對照組，透過樹莓派連接並讀取資料，每台感測器以每五分鐘讀取一筆資料做實驗，並試分析結果。資料的讀取方式是先由 sensor 接上 GPIO 所得到的 byte array 解碼成 32 byte，而字串固定開頭為 "4d4200000000"，固定結尾為 "009X"，pm2.5 的資料位置則是開頭後四碼 "0004"，最後再轉為十進位制。

舉例：4d42000000000000400093 -> 解碼 pm2.5 = 4

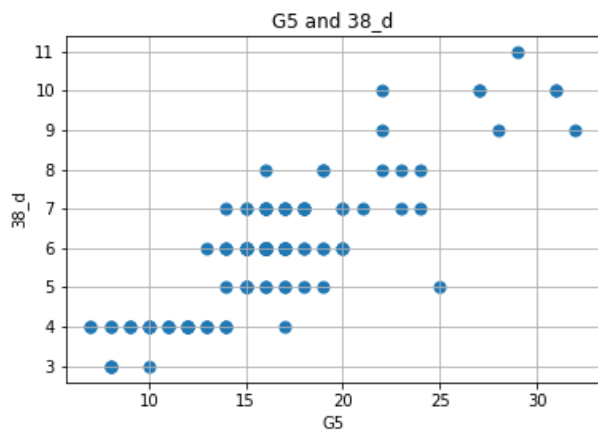
實驗結果：



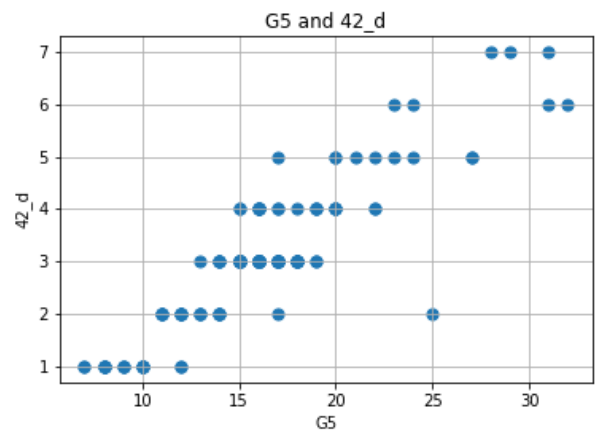
圖一，編號38、42、43、47、G5 連續十二小時之結果



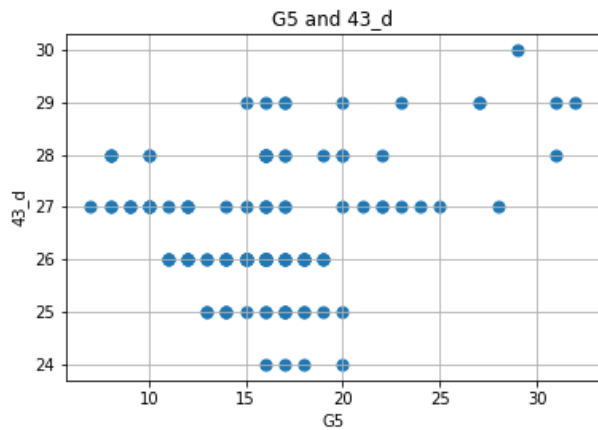
圖二，每五分鐘的感測器數值平均，與共變異數分析



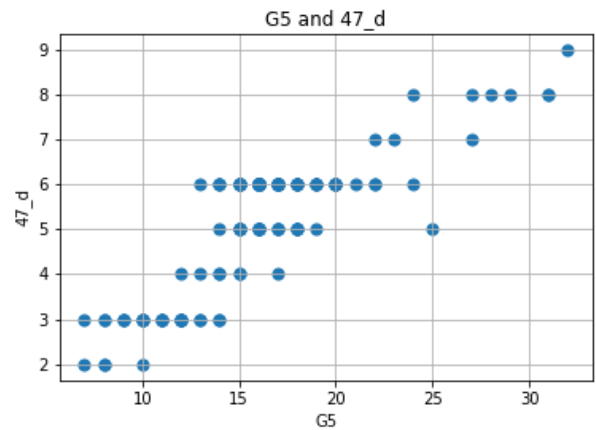
圖三，G5 與 38 號分佈圖，兩者相關係數0.85



圖四，G5 與 42 號分佈圖，兩者相關係數0.89



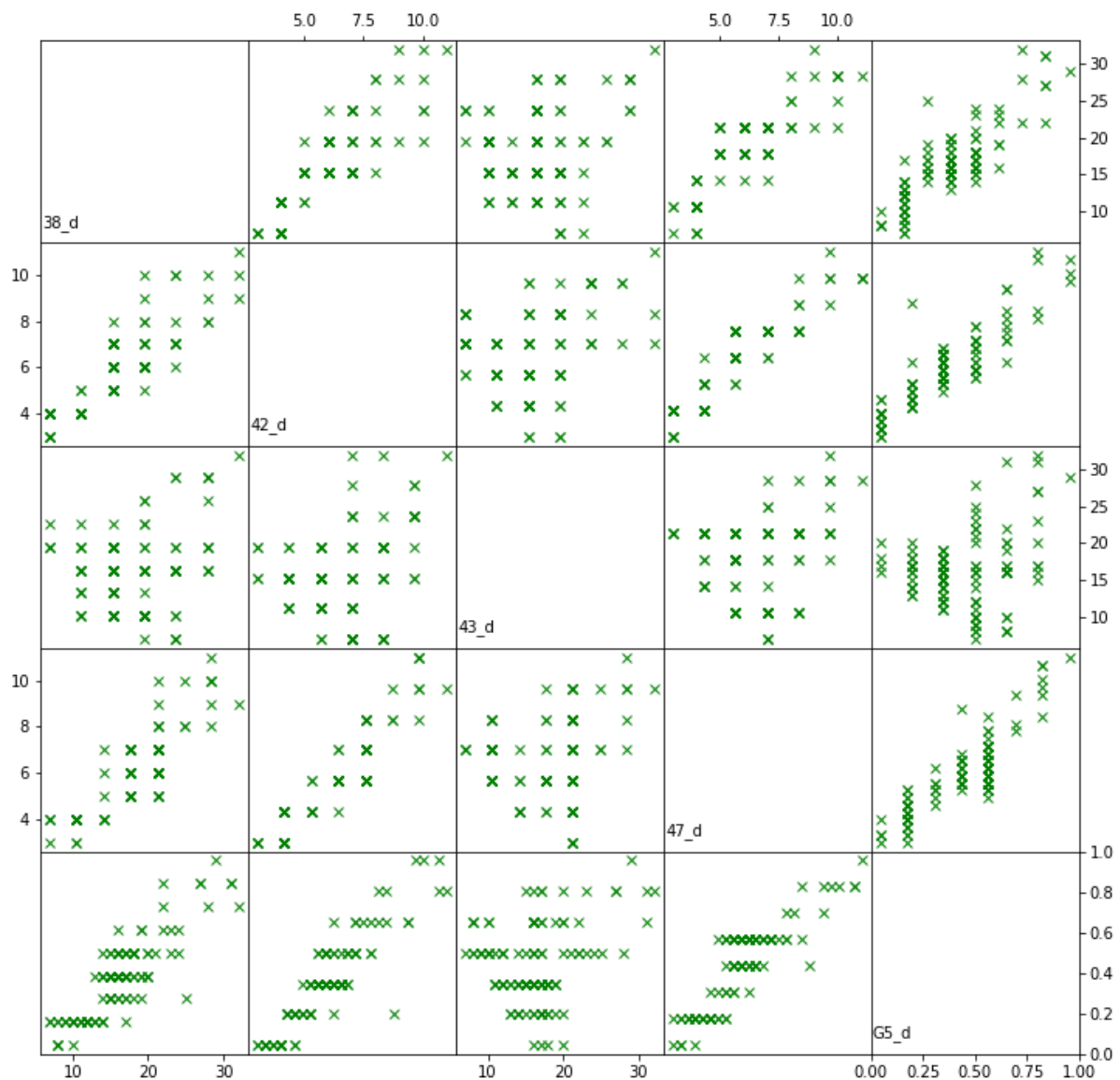
圖五，G5 與 43 號分佈圖，兩者相關係數0.24



圖六，G5 與 47 號分佈圖，兩者相關係數0.85

	38_d	42_d	43_d	47_d	G5_d
count	133.000000	135.000000	132.000000	132.000000	135.000000
mean	5.902256	3.059259	26.568182	5.181818	16.162963
std	1.608993	1.308770	1.224391	1.476628	4.755527
min	3.000000	1.000000	24.000000	2.000000	7.000000
25%	5.000000	2.000000	26.000000	4.000000	14.000000
50%	6.000000	3.000000	26.000000	6.000000	16.000000
75%	7.000000	4.000000	27.000000	6.000000	18.000000
max	11.000000	7.000000	30.000000	9.000000	32.000000

圖七，各個感測器實驗中數據分析，本次實驗共135筆資料



圖八，各個感測器之間的SCATTER PLOT MATRIX

分析結果：

首先，由圖一觀察得知編號 43 資料與其他工研院感測器明顯結果不同，又與G5資料之兩者相關係數（0.24）得知兩者缺乏關係性，初步判定編號 43 是比較有問題的裝置，其資料與其他工研院感測器之間正確性將在實驗二說明。

第二，由圖二中共變異數與每五分鐘所有感測器平均數據分析可說明，當pm2.5 值越小，感測器之間相關程度越大，變動幅度更相近，並可呈現線性正關係。

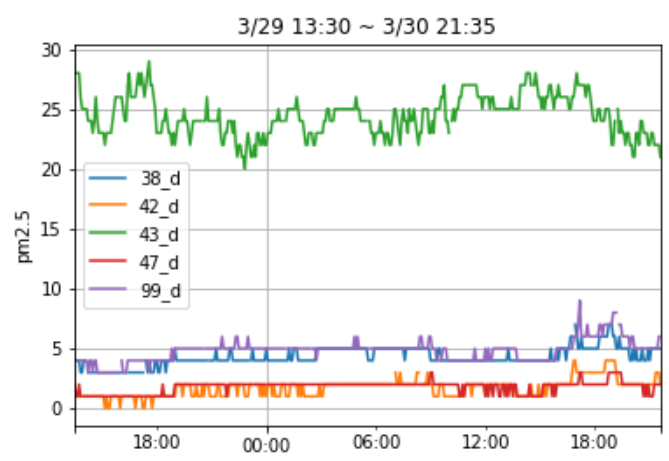
第三，透過相關係數分析，可得知 G5 感測器與編號 38、42、47 三台有極高的相關程度，其資料雖然不夠精準，但是變化趨勢大致相同，但是由各組內標準差可看出 G5 資料的離散程度高於各個工研院感測器，再由圖一推判，除了 G5 以外的感測器，資料的變化幅度不大，資料間離散程度低。

實驗（二）：

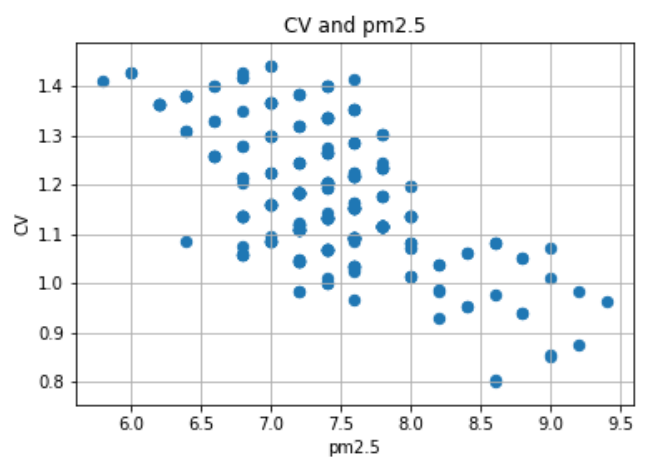
時間：3/29 13:30 ~ 3/30 21:35，共連續約32小時

實驗方式：繼續以編號 38、43、42、47、G5 做連續實驗，並加入一台編號 99，觀察工研院感測器之間資料的一致性。（最後，由於 G5 感測器讀取資料時發生供電不足的問題，因此無法順利取得資料只好做工研院感測器組內之分析）

實驗結果：



圖一，編號 38、42、43、47、99 連續32小時之結果



圖二，每五分鐘的感測器數值平均，與共變異數分析

	38_d	42_d	43_d	47_d	99_d
count	377.000000	373.000000	379.000000	378.000000	376.000000
mean	4.302387	1.697051	24.432718	1.777778	4.816489
std	0.764101	0.741922	1.686418	0.492203	0.885946
min	3.000000	0.000000	20.000000	1.000000	3.000000
25%	4.000000	1.000000	23.000000	1.000000	4.000000
50%	4.000000	2.000000	24.000000	2.000000	5.000000
75%	5.000000	2.000000	26.000000	2.000000	5.000000
max	7.000000	4.000000	29.000000	3.000000	9.000000

圖七，各個感測器實驗中數據分析，本次實驗共379筆資料



分析結果：

圖一中，可得知實驗一中對於編號 43 感測器的懷疑及不確定性是正確的，其平均值及全距皆與其他同時同地之分析結果相距甚遠，判定編號 43 是有問題的裝置。

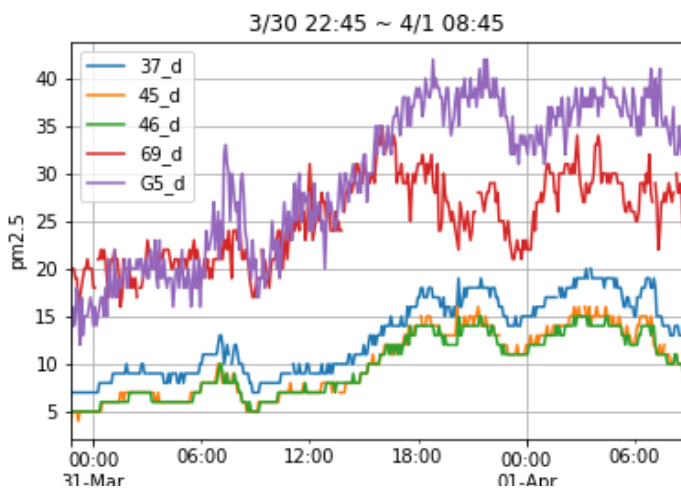
第二，圖二中共變異數與平均值之散佈圖，雖然沒有實驗一密集且趨近線性，但是在  $pm_{2.5}$  上卻有著更高的相關性，更可證明若不考慮有問題的編號 43，工研院製感測器具有相當高的一致性與相等的變化幅度及趨勢。

### 實驗（三）

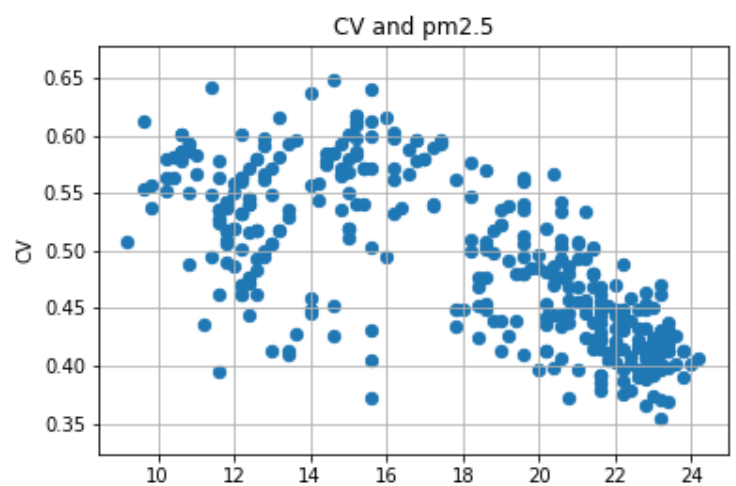
時間：3/30 22:45 ~ 4/1 08:45，共連續約34小時

實驗方式：以編號 37、45、46、69、G5 做連續實驗，觀察工研院感測器之間資料的一致性以及和對照組之間的精準度。

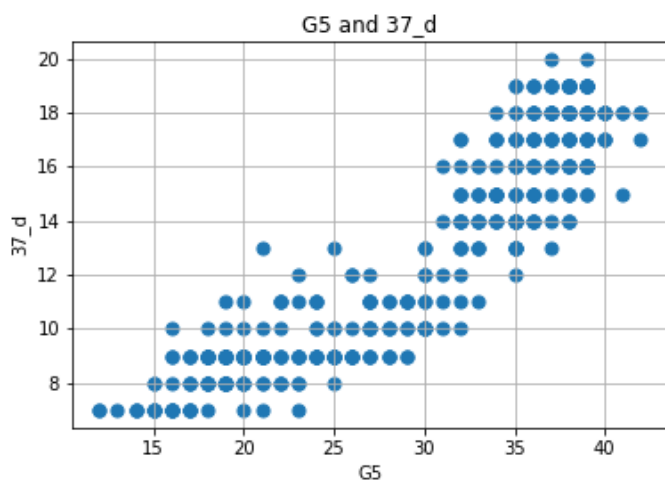
實驗結果：



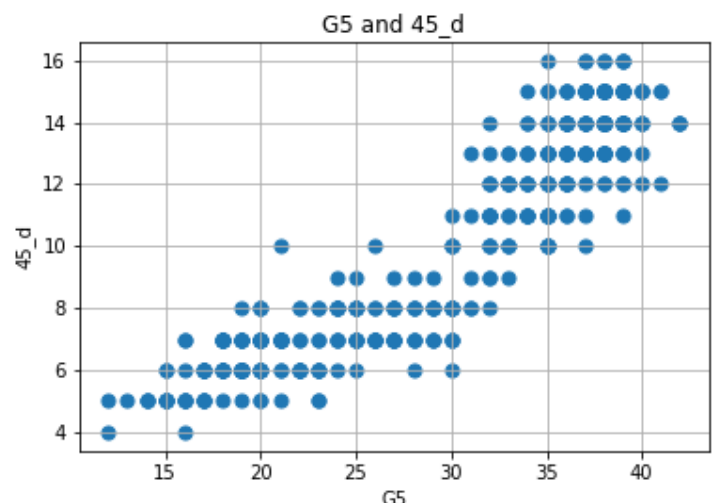
圖一，編號 37、45、46、69 連續34小時之結果



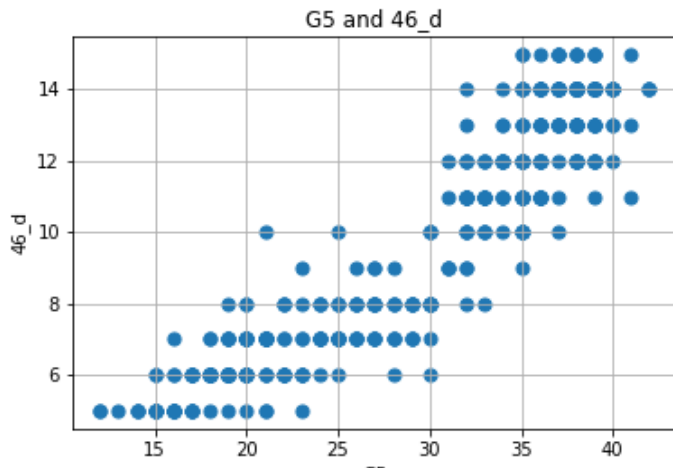
圖二，每五分鐘的感測器數值平均，與共變異數分析



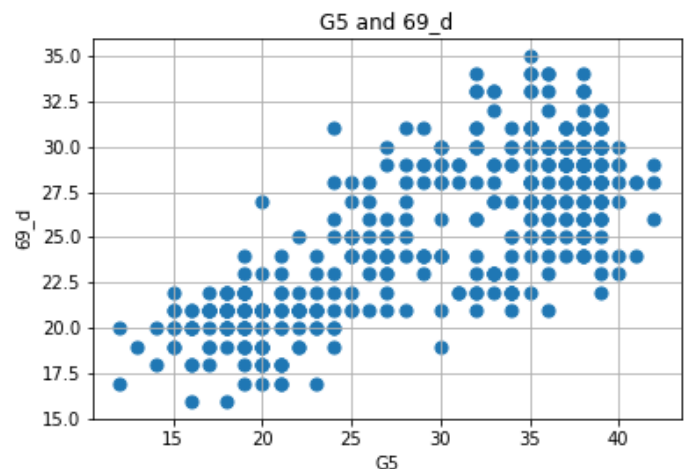
圖三，G5 與 37 號分佈圖，兩者相關係數0.91



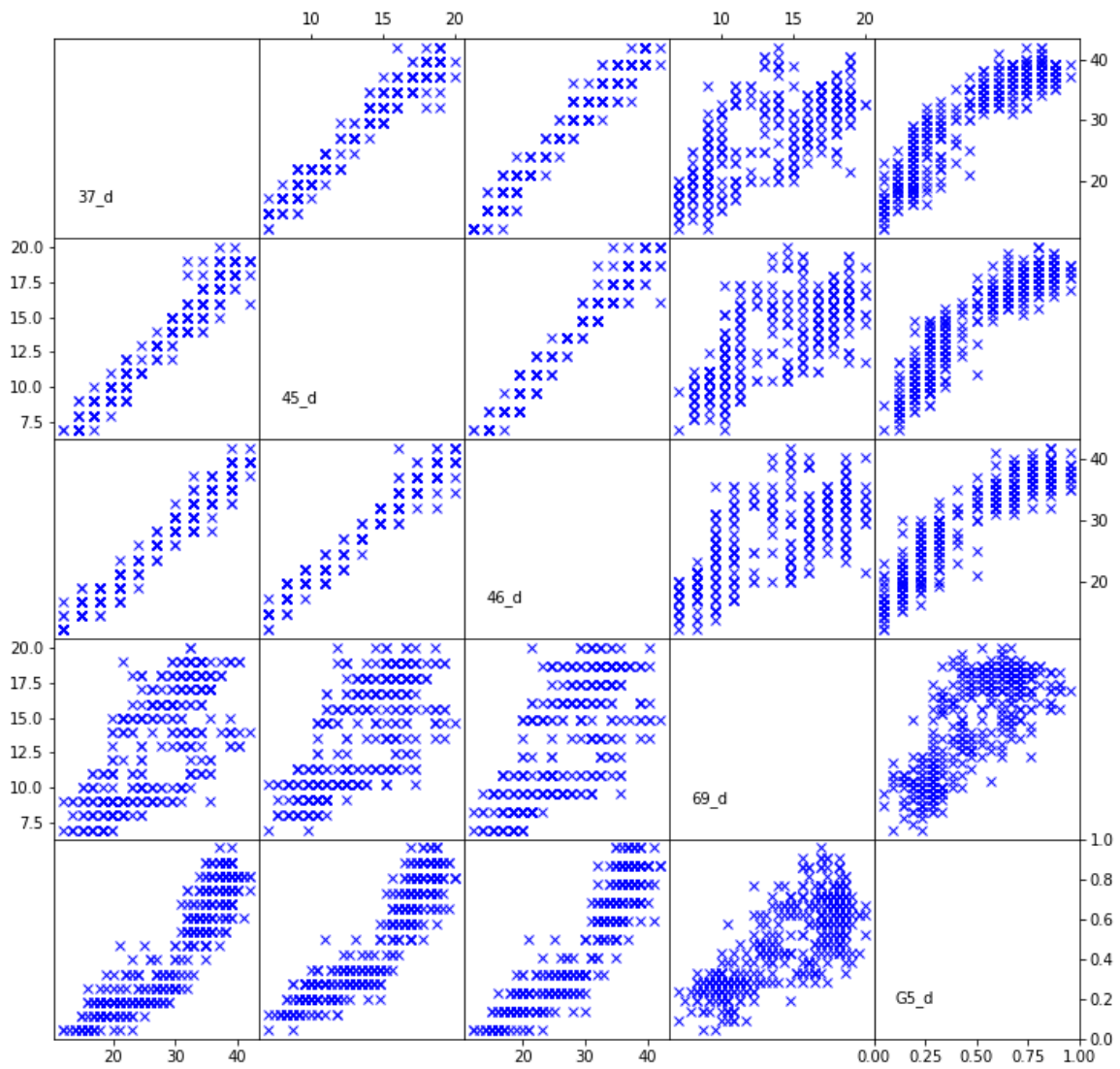
圖四，G5 與 45 號分佈圖，兩者相關係數0.92



圖五，G5 與 46 號分佈圖，兩者相關係數0.92



圖六，G5 與 69 號分佈圖，兩者相關係數0.75



圖八，各個感測器之間的SCATTER PLOT MATRIX

	37_d	45_d	46_d	69_d	G5_d
count	398.000000	399.000000	402.000000	401.000000	409.000000
mean	12.839196	10.072682	9.803483	25.169576	29.457213
std	3.988557	3.485394	3.255333	4.338914	8.069171
min	7.000000	4.000000	5.000000	16.000000	12.000000
25%	9.000000	7.000000	7.000000	21.000000	21.000000
50%	13.000000	10.000000	10.000000	25.000000	32.000000
75%	17.000000	13.000000	13.000000	29.000000	37.000000
max	20.000000	16.000000	15.000000	35.000000	42.000000

圖七，各個感測器實驗中數據分析，本次實驗共409筆資料

分析結果：

從圖一中，長時間來看編號 69 感測器明顯高過其他工研院製感測器，其資料離散程度及當日平均可從圖七判定與標號 37、45、46 超出許多，因此參考實驗二對編號 43 結論，判定其準確度是有待商榷的。

第二，由相關係數角度發現編號 69 雖然與 G5 依舊有著 0.75 的高相關程度，但是比起其他工研院製感測器，更可推判其差異。

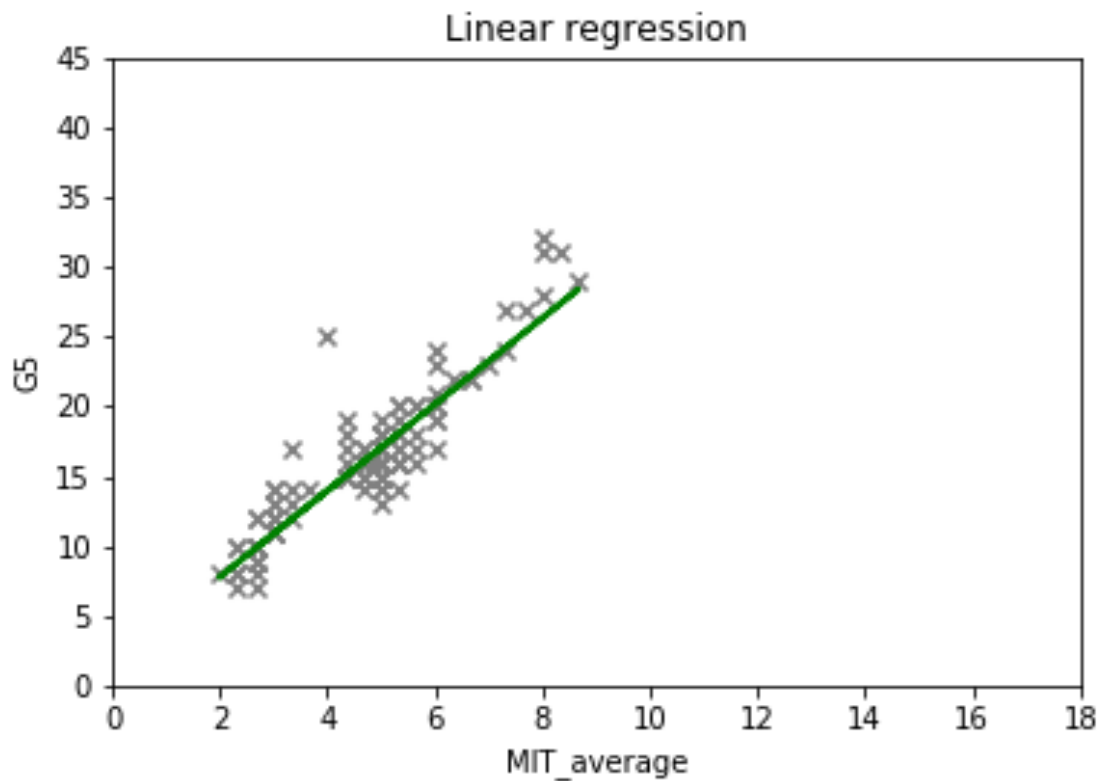
第三，每台感測器之間的相關程度，我們可從圖二觀察，雖然沒有實驗一和實驗二所得到的高相關性，但是卻能得到更大尺度的空氣濃度相關性，尤其是在空氣濃度高於18 ( $\mu\text{g}/\text{m}^3$ ) 之後，空氣濃度越大，感測器間的相關性越低，由實驗一的圖二中，空氣濃度超過16 ( $\mu\text{g}/\text{m}^3$ ) 之後，感測器之間共變異數約為 0.6，佐證出實驗三對於共變異數分析的正確性。

結論：

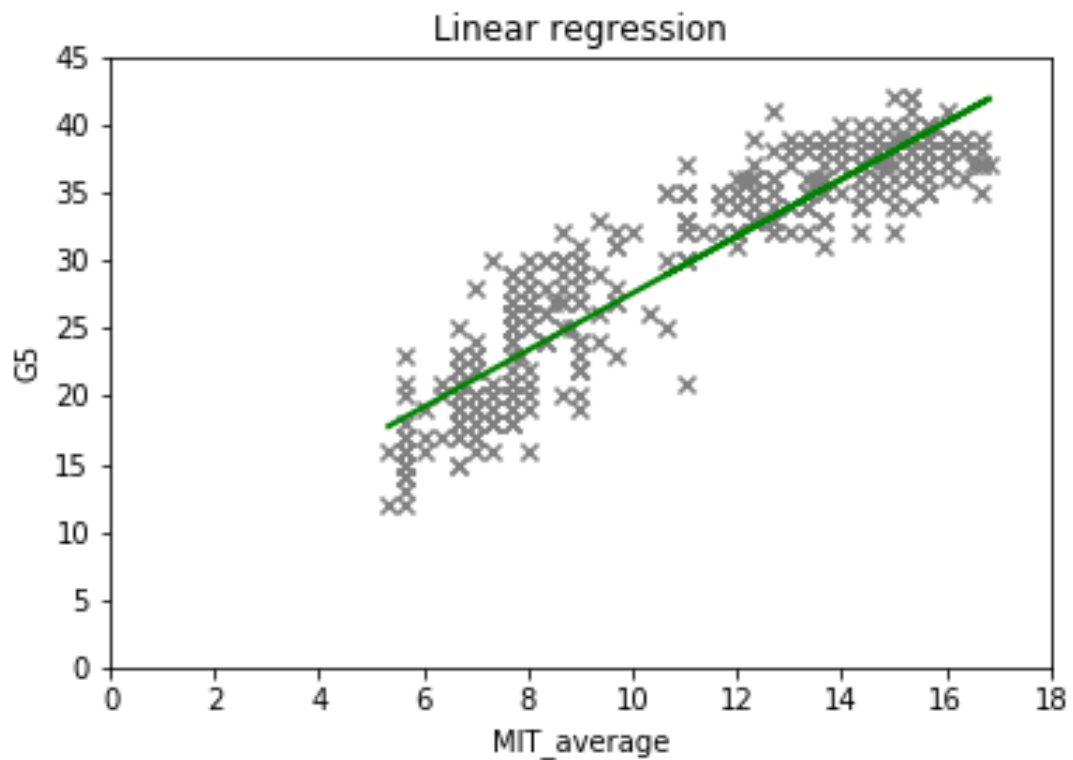
由實驗一、二、三對於工研院製感測器結果分析得知，排除有問題的編號 47、69 號，其他感測器相關性與 G5 有著極高的相關性及高相似度的變化趨勢，資料上也有著完整的一致性（由實驗二得知），但是與對照組的 G5 相比，資料缺乏準確度，關於 G5 和 工研院感測器相關的回歸分析在下方有更完整說明。最後，透過實驗一和實驗三的圖八比較可以清楚看出有問題的裝置，與其他明顯有著不同的分佈現象。整個實驗過程中，工研院製感測器噪音實在惱人，不宜放置在一般民眾寧靜的居家環境中。另外，感測器本身需要的電流量高，若是供電不足，將造成資料讀取不完整以及資料格式不清楚的情況，一般樹莓派在接上第三個感測器就會出現斷線或是資料不清楚的案例發生，關於耗電問題在最後篇幅將有說明。



實驗一與實驗三之迴歸分析：



圖一，實驗一中 G5 與 感測器之相關線性迴歸分析 ( $Y = 3.08 * X + 1.61$ , R-SQUARE = 0.816)



圖二，實驗三中 G5 與 感測器之相關線性迴歸分析 ( $Y = 2.09 * X + 6.58$ , R-SQUARE = 0.855)

分析結果：

實驗一迴歸分析中，X 變量為扣除有問題的編號 47 及其餘感測器之平均，Y 變量為對照組 G5 資料，放進 sklearn library 進行迴歸分析，得到的預測將稱為 f 方程式。而迴歸方程式  $f: Y = 3.08 * X + 1.61$ 。

實驗三迴歸分析中，X 變量為扣除有問題的編號 69 及其餘感測器之平均，Y 變量為對照組 G5 資料，同樣去做現性回歸，方程式  $f: Y = 2.09 * X + 6.58$ ，為了解釋兩次實驗迴歸分析的配合度，我們以決定係數（R - square）來描述。R - square 值域介於 0 與 1 之間，大小表示 Y 之變異是受 X 影響程度。若 R-square 越大，我們越能信賴迴歸分析知結果。在做完兩次迴歸分析後，實驗一的 R - square 為 0.816，實驗二的 R - square 為 0.855，因此我們有可以有強力的證據去相信分析完後的線性迴歸方程式。

**耗電問題：**

樹莓派沒有接上任何裝置前電流量是 0.2 安培，在接上感測器後，電流量增加到了 0.55 ~ 0.58 安培，推估工研院製國產感測器運轉電流量為 0.35 ~ 0.38 安培，耗電量為 1.75~1.9（瓦特），相較 G5 感測器電流量 0.01~0.02 安培，耗電量為 0.05~0.02（瓦特）高出數三十五倍。