



# 應用最小絕對壓縮挑選機制法 改進溫度迴歸模型建置方式

馮智勇 林佑蓉 劉家豪  
陳雲蘭

多采科技有限公司  
交通部中央氣象局預報中心

# 簡報大綱

一、緣起與目的

二、最小絕對壓縮挑選機制法原理

三、以LASSO法建置迴歸模型流程

四、溫度迴歸模型建置實驗

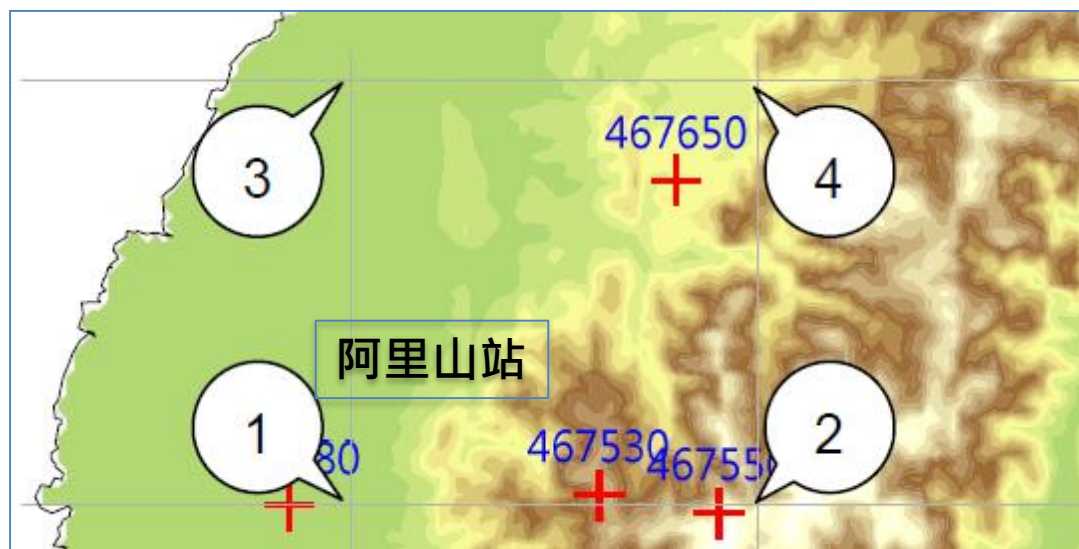
五、研究綜整

# 溫度迴歸模型建置方式

- 以動力模式資料(X)迴歸測站觀測(Y)
  - 參考美國 MDL 作法，以統計方法加值動力模式輸出之3D大氣資料
  - 複線性迴歸方程式(multiple linear regression)
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$
  - 採前進選擇法(Forward selection)挑選預報子
    - 預報子必須增加模型解釋能力
    - 預報子共線性(Colinearity)必須小於寬容值
    - 以 F-test 確認預報子之模型解釋能力顯著性

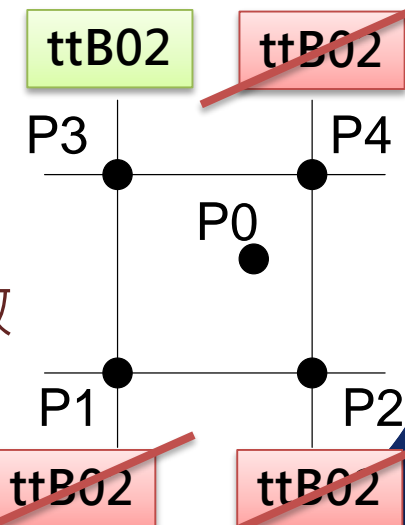
# 動力模式資料前處理

- 各層場需先內插至測站處(P0)
  - 動力模式網格(P1~P4)解析度 50km~5km
  - 以雙線性內插(Bilinear interpolation)為主
  - 台灣地形變化急遽，可能錯誤解讀低層資訊
  - 測站溫度模型主要為低層資訊影響
    - 高度(H)
    - 溫度(T)
    - 水平風分量(U)
    - 垂直風分量(V)
    - 濕度(Q)



# 逕用動力模式資料建模可行性

- 模型成效是否相當
  - 以動力模式網格P1~P4資料迴歸測站溫度觀測
  - 與使用P0處資料之迴歸模型比較
    - 預報成效、建模成效
- 層場共線性是否有助於迴歸模型預報？
  - 變數挑選方式影響模型預報子共線性
    - 前進選擇法降低共線性以得最佳建模成效
    - LASSO允許存在共線性以得最佳預報成效



# 研究目的

- 探討台灣測站溫度迴歸模型建置方式
  - 逕用動力模式資料(P1~P4)建模的差異
    - 與使用P0資料建置之模型比較，  
評估內插必要性
  - 使用最小絕對壓縮挑選機制法(Least Absolute Shrinkage and Selection Operator, LASSO)建置之迴歸模型預報成效
    - 與使用前進選擇法建置之模型比較，  
評估逕用動力模式資料建模時的預報子挑選方式

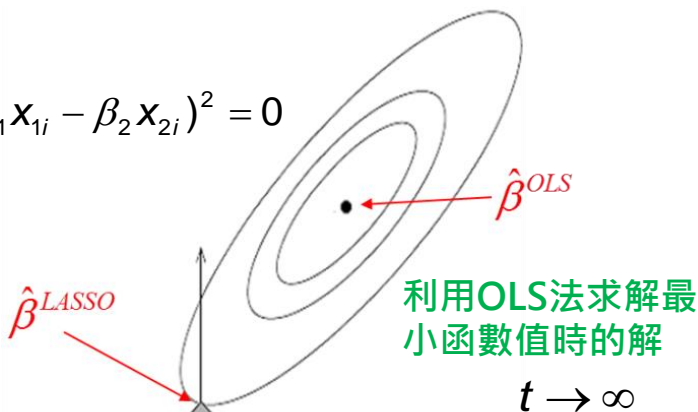
# 最小絕對壓縮挑選機制法原理

$$\sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_k x_{ki})^2$$

估計一組係數  $\beta_j$

限制式  $\sum_{j=1}^k |\beta_j| \leq t$

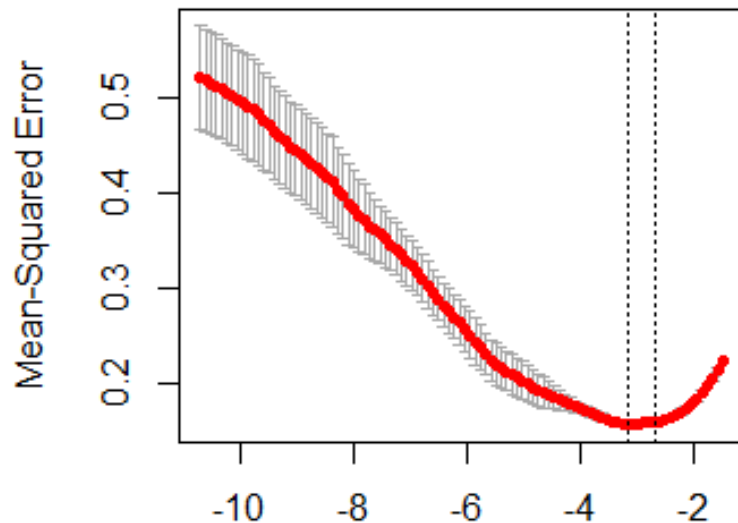
$$\sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i})^2 = 0$$



代表  $|\beta_1| + |\beta_2| \leq 1$  區域

最小估計誤差法(OLS)與  
LASSO法之最佳解示意

97 93 90 80 55 32 9 3



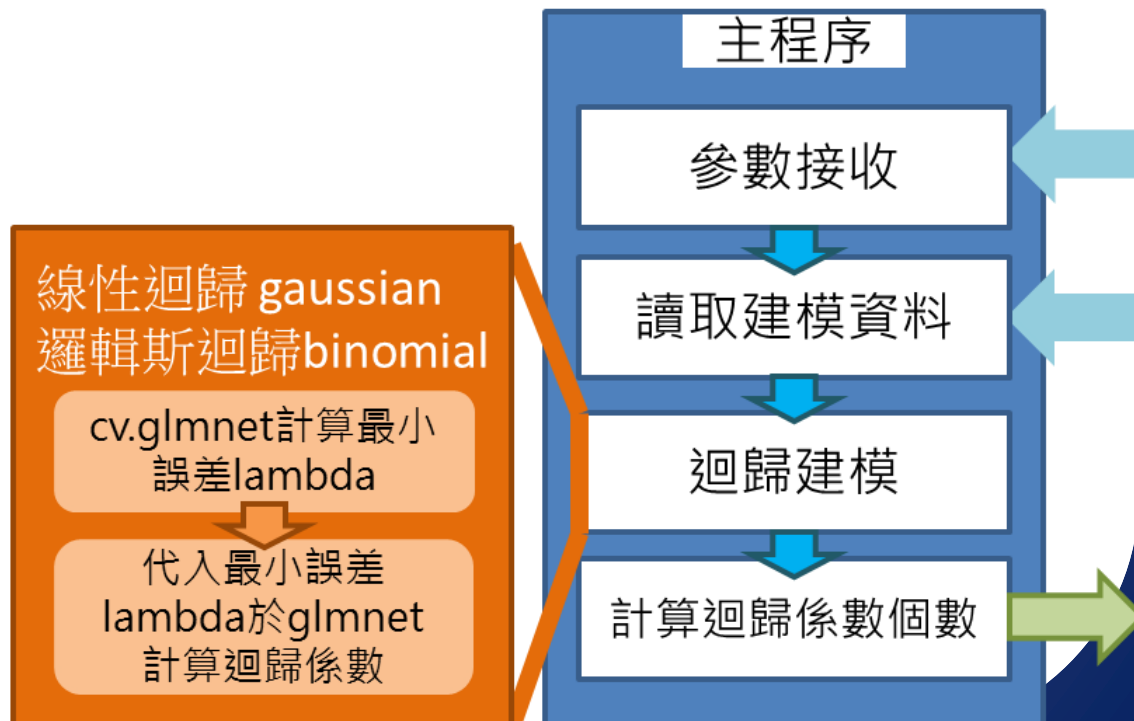
模型預報均方差(縱軸)隨  
預報子數(上方橫軸)變化示意

# 以LASSO法建置迴歸模型

- R統計軟體LASSO套件 glmnet
  - cv.glmnet 函式與 glmnet 函式

- cv.glmnet 函式自動隨機分割資料成n等份，進行交叉檢驗 (cross validation)
- cv.glmnet 函式由模型預報均方差隨預報子數變化關係，獲得最佳預報子個數資訊。
- glmnet 函式以完整資料與最佳預報子個數資訊建置模型

## 應用R統計軟體LASSO套件glmnet 建置複迴歸模型流程示意



**隨機分割建模資料，模型不具重製性，不利校驗!**



# 以 glmnet 函式改良交叉檢驗作法

以所有資料估計50個lambda

$i = 1, 10$

切割資料並取第  $i$  分為  
校驗資料，剩下資料為  
訓練資料

以訓練資料估計50個  
lambda <sub>$i$</sub>

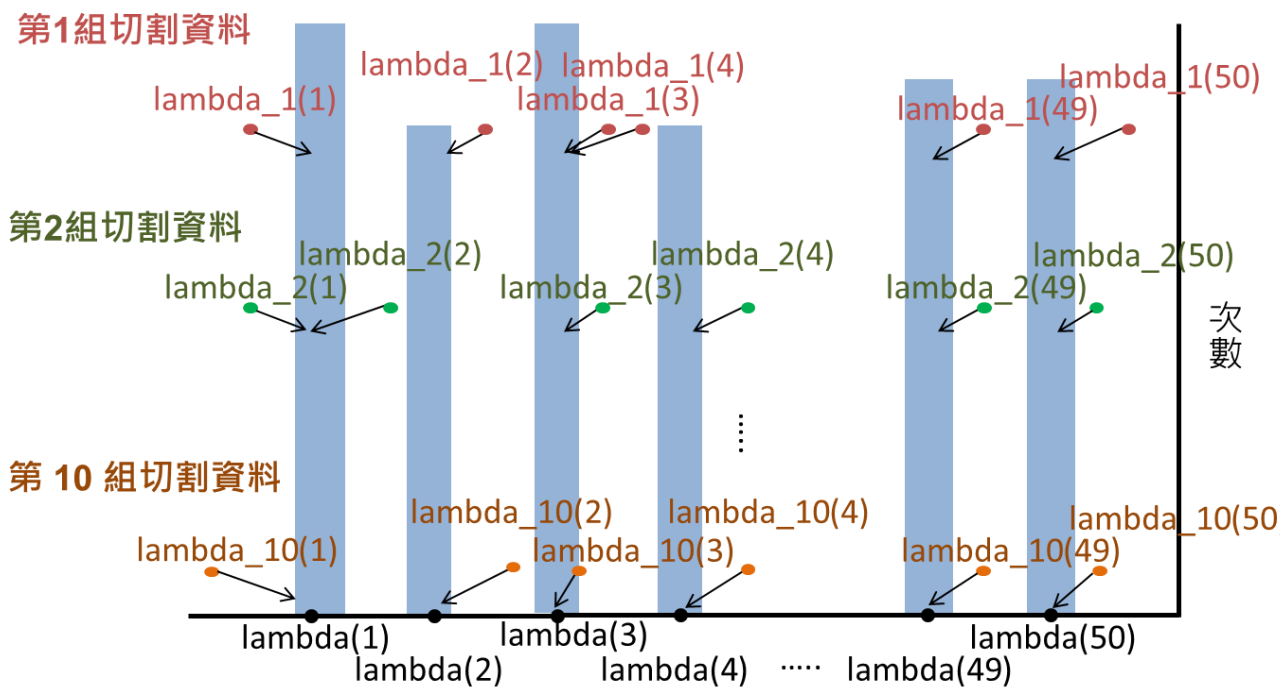
$j = 1, 50$

以校驗資料代入係數  
計算估計誤差

lambda <sub>$i$</sub> 落在lambda-1至  
lambda之間，則此估計誤  
差計為lambda之估計誤差

計算10組分割資料的50個  
lambda平均誤差

取最小誤差之lambda做為最  
佳參數計算迴歸係數



方案	cv.glmnet 自動隨機切割	glmnet + 手動固定切割
建模時間	17分5秒	17分55秒

# 溫度迴歸模型建置實驗

## 建模：07~09年ECTIGGE模式

### □ 最大調整判定係數(Adjusted $R^2$ · AdjR2)

用以瞭解建模成效。

- 925/850/700/500/300mb
- 高度(H)/溫度(T)/水平風分量(U)/  
垂直風分量(V)/濕度(Q)

## 事後預報：11年JMA模式

### □ 平均誤差(Mean Error, ME)

用於量化的事後預報值與觀測值誤差平均值，越接近0越佳

### □ 相關係數(Correlation Coefficient, CORR)

用於掌握事後預報值與觀測值時間序列變化趨勢，越高越佳

### □ 均方根誤差與標準差比值(RMSE/YSD)

掌握事後預報值變異度與不同氣候特性觀測值/預報值的比較，越低越佳

## PP策略00Z溫度迴歸模型

FS+P0 前進選擇法+測站點內插(P0)

FS+4P 前進選擇法+測站點四周格點

LA+4P LASSO法+測站點四周格點

前進選擇法建模參數：

預報子個數為10

共線性寬容值為0.005

# 2月份測站溫度模型建置成效

使用P1~P4資料建置之迴歸模型成效略優於使用P0資料建置者：

1. 採前進選擇法時，幾乎均略優
2. 採最小絕對壓縮挑選機制法時，僅部分測站略不如「FS+P0」
  - 甚至略優於「FS+4P」
  - 不如「FS+P0」者，差異均屬些微

2月份00Z溫度迴歸模型擬合度比較

站碼 \ 建置方式	FS+P0	FS+4P	LA+4P
466910	0.964	0.964	0.962
466920	0.890	0.894	0.914
466930	0.952	0.954	0.955
466940	0.821	0.840	0.850
466990	0.834	0.838	0.830
467080	0.906	0.917	0.924
467410	0.857	0.867	0.889
467440	0.826	0.851	0.864
467480	0.859	0.880	0.882
467490	0.873	0.880	0.896
467530	0.843	0.857	0.853
467550	0.813	0.841	0.848
467610	0.885	0.901	0.900
467650	0.860	0.876	0.889
467660	0.876	0.894	0.908

黃底：優於左方欄位者

綠底：優於FS+P0者

# 2月份測站00Z溫度模型預報成效

## 2011年 JMA 模式事後預報統計指標比較

使用P1~P4之模型各項  
指標均具優於使用P0者

- 「LA+4P」幾乎全面優於「FS+P0」
- 台北,台南,玉山,成功略有差異

站碼	建置方式	站名	ME										
			FS			LA+4P	FS			LA+4P	FS		
			P0	4P	P0		4P	P0	4P		LA+4P		
466910	鞍部	-0.698	-0.712	-0.547	0.270	0.270	0.220	0.309	0.306	0.262			
466920	臺北	0.439	0.495	0.599	0.448	0.425	0.516	0.239	0.276	0.237			
466930	竹子湖	0.205	0.427	0.218	0.849	0.852	0.884	0.523	0.525	0.460			
466940	基隆	0.052	0.234	0.044	0.880	0.840	0.887	0.561	0.705	0.494			
466990	花蓮	-0.607	-0.842	0.383	0.943	0.915	0.933	0.329	0.408	0.359			
467080	宜蘭	0.055	0.034	0.083	0.839	0.837	0.837	0.587	0.640	0.651			
467410	臺南	0.547	0.843	0.795	0.21	0.511	0.455	0.61	0.663	0.647			
467440	高雄	0.107	0.019	0.052	0.41	0.587	0.499	0.708	0.600	0.565			
467480	嘉義	0.064	0.277	0.381	0.875	0.858	0.858	0.440	0.443	0.457			
467490	臺中	0.335	0.030	0.037	0.897	0.897	0.888	0.444	0.458	0.507			
467530	阿里山	-1.015	-0.616	-0.493	0.945	0.947	0.943	0.444	0.458	0.507			
467550	玉山	0.322	0.256	0.234	0.911	0.890	0.940	0.498	0.469	0.345			
467610	成功	0.649	0.699	0.808	0.931	0.903	0.921	0.414	0.738	0.398			
467650	日月潭	-0.651	-0.310	-0.130									
467660	臺東	0.304	1.243	0.224									

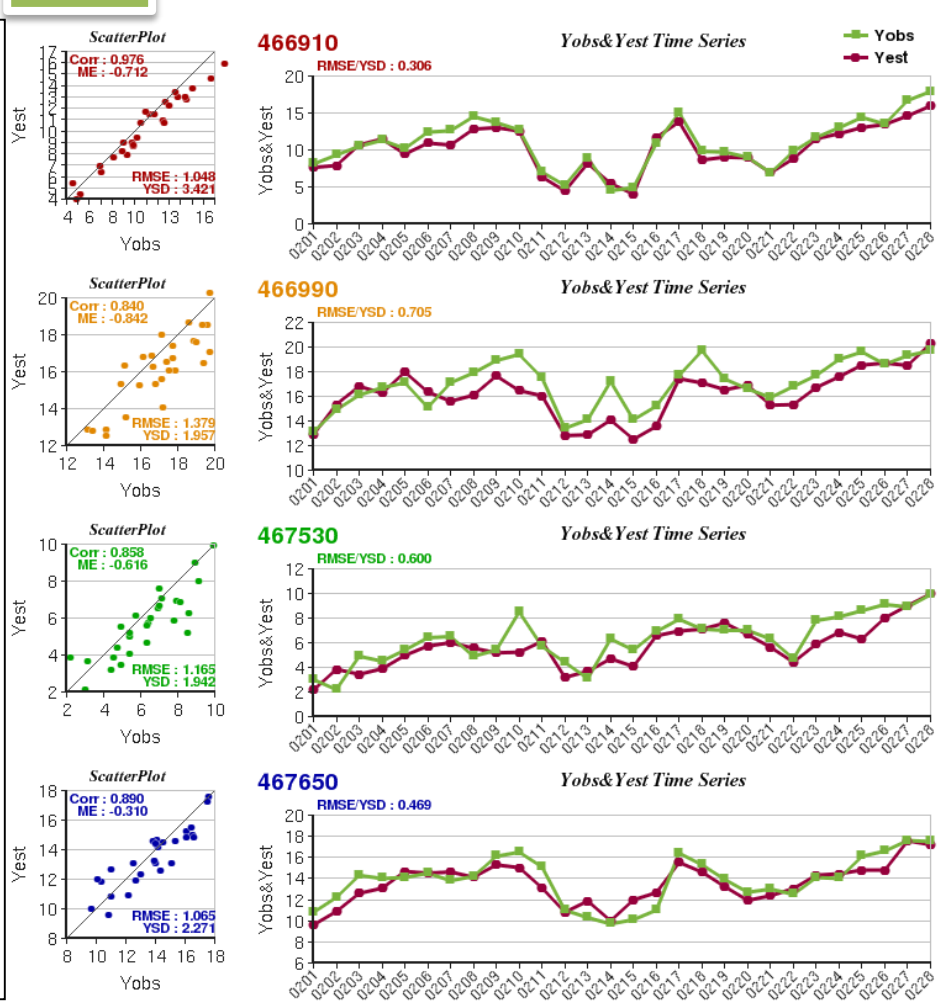
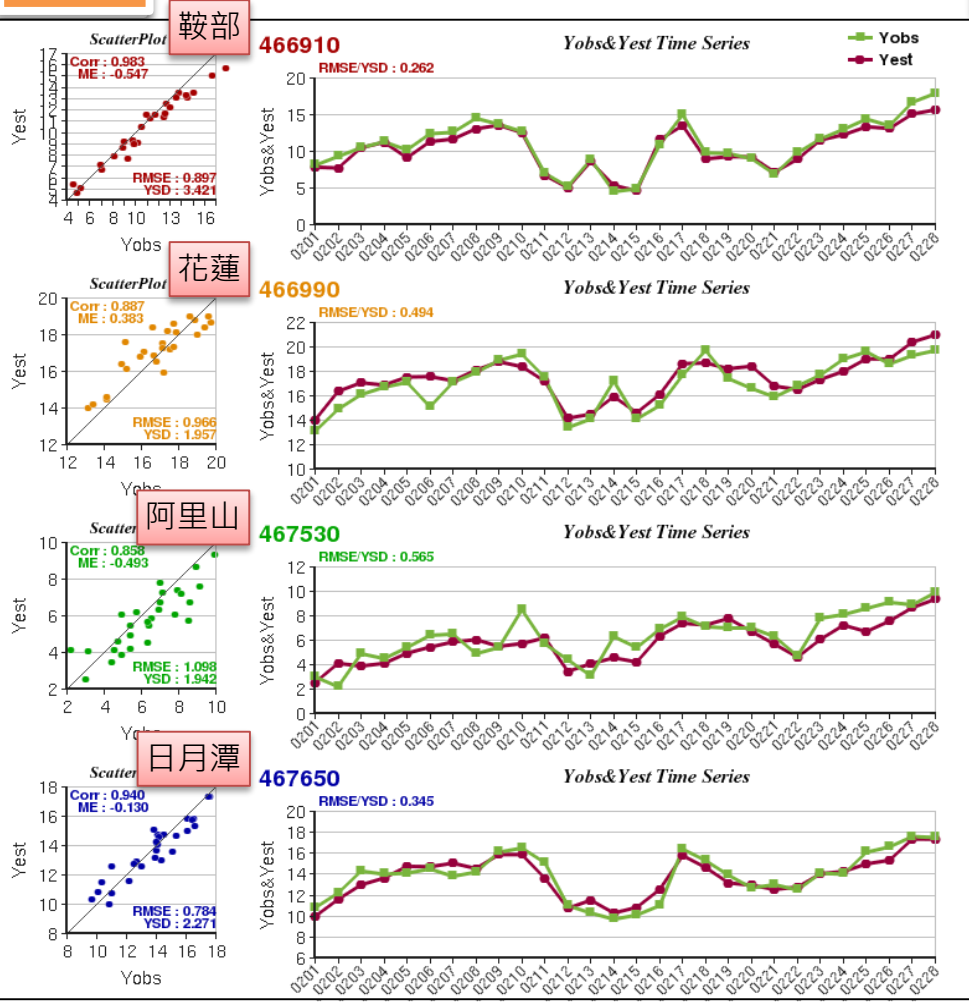
僅部分測站「LA+4P」  
略不如「FS+4P」

花蓮、台東「FS+4P」  
明顯不如「FS+P0」  
「LA+4P」顯著改善

# 2月份測站00Z溫度時序比較

LA+4P

FS+4P



# 3月份測站00Z溫度模型預報成效

## 2011年 JMA 模式事後預報統計指標比較

使用P1~P4之模型各項  
指標均具優於使用P0者

- 「LA+4P」幾乎全面優於「FS+P0」
- 高雄略有差異

站碼	建置方式	站名	統計值								
			ME			FS			FS		
			FS	LA+4P	FS	LA+4P	FS	LA+4P			
P0	4P	P0	4P	P0	4P	P0	4P	LA+4P			
466910		鞍部	-0.753	-0.647	-0.683	0.986	0.990	0.987	0.258	0.220	0.244
466920		臺北	-0.407	-0.489	-0.463	0.883	0.903	0.896	0.490	0.458	0.470
466930		竹子湖	-0.403	-0.533	-0.387	0.977	0.978	0.977	0.240	0.256	0.240
466940		基隆	-0.138	-0.225	-0.356	0.830	0.856	0.848	0.567	0.523	0.551
466990		花蓮	-0.862	-0.766	0.122	0.782	0.810	0.799	0.675	0.643	0.462
467080		宜蘭	-0.112	-0.059	-0.045	0.782	0.812	0.825	0.301	0.319	0.290
467410		臺南	-0.780	-0.307	-0.396	0.782	0.810	0.799	0.682	0.589	0.612
467440		高雄	-1.085	-1.111	-1.342	0.782	0.812	0.825	0.767	0.745	0.788
467480		嘉義	-0.668	-0.331	-0.045	0.782	0.812	0.825	0.645	0.595	0.586
467490		臺中	-0.795	-0.642	-0.045	0.782	0.812	0.825	0.568	0.460	0.560
467530		阿里山	-0.025	-0.303	-0.045	0.782	0.812	0.825	0.419	0.468	0.401
467550		玉山	-0.444	-0.139	-0.045	0.782	0.812	0.825	0.618	0.601	0.611
467610		成功	-0.367	-0.732	-0.493	0.943	0.943	0.931	0.438	0.639	0.540
467650		日月潭	-0.366	-0.146	-0.124	0.879	0.879	0.903	0.492	0.473	0.426
467660		臺東	-0.443	-0.661	-0.490	0.957	0.947	0.959	0.377	0.564	0.422

僅高雄站「LA+4P」  
略不如「FS+4P」

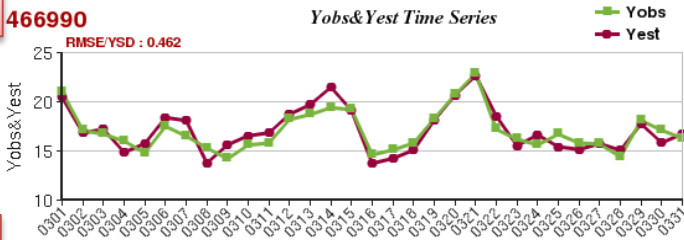
成功、台東「FS+4P」  
明顯不如「FS+P0」  
「LA+4P」顯著改善，  
但仍略不如「FS+P0」

# 3月份測站00Z溫度時序比較

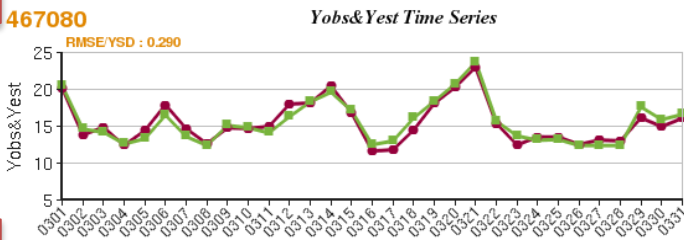
LA+4P

FS+4P

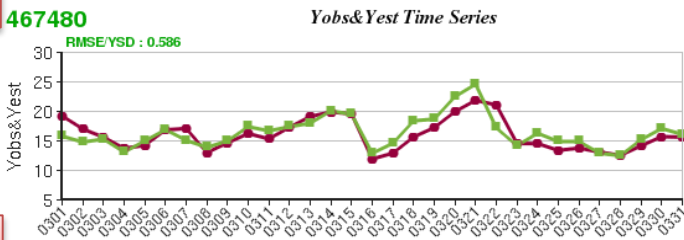
花蓮



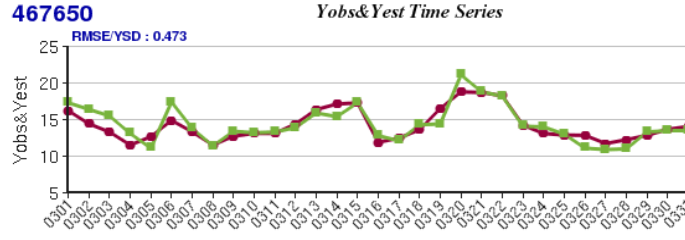
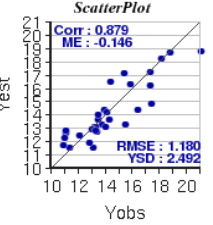
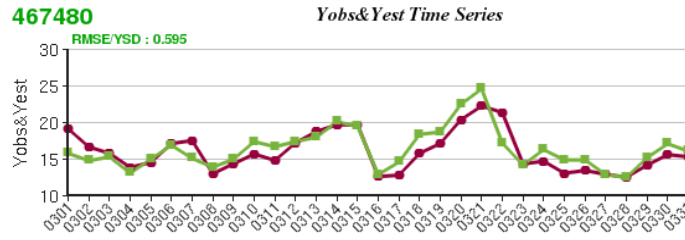
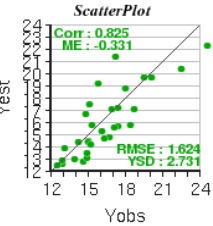
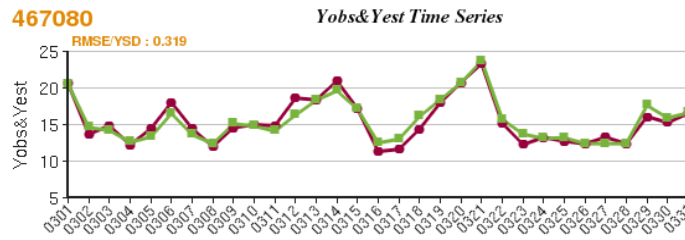
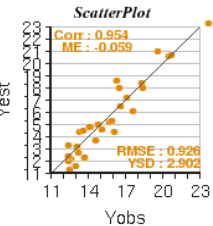
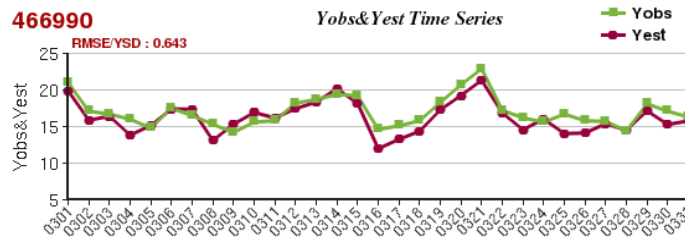
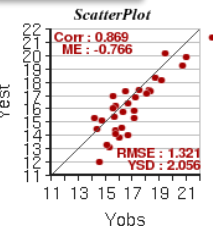
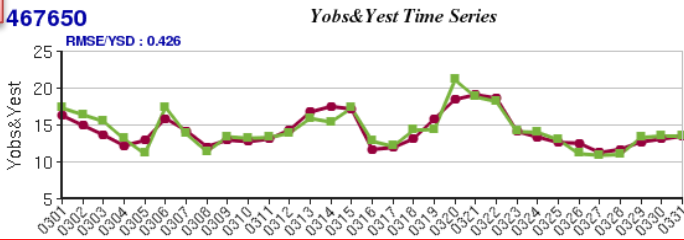
宜蘭



嘉義



日月潭



# 研究綜整

- 利用R統計軟體glmnet套件改良交叉檢驗作法
  - 使採LASSO法建置之模型利於後續校驗
- 溫度模型建置建議使用「LA+4P」為佳
  - 直接使用P1~P4建置之模型預報成效幾乎均優於使用P0者，並可得到符合物理現象之預報子
  - FS追求建模成效佳，LASSO追求預報成效佳  
→原理皆正確，應以追求預報成效佳為最終目標。
  - 「LA+4P」較「FS+4P」具大幅改善預報成效的能力
- 建議針對各測站特性分別建模滿足作業需求
  - 建議統計分析連續三年建模交叉檢驗與逐月建置
  - 研究選用區域模式建模&預報



簡報結束  
敬請指教