

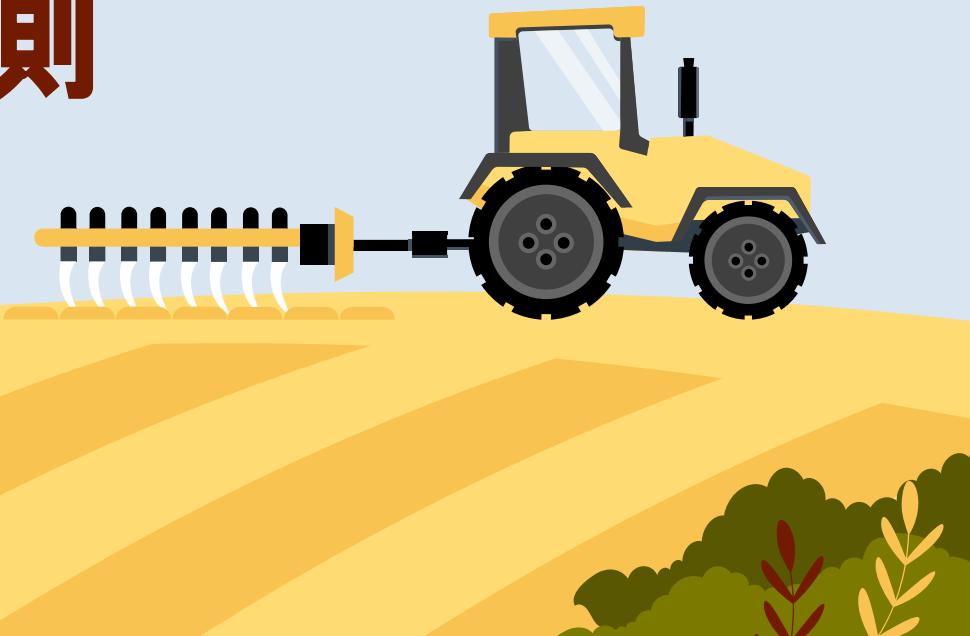
Group 6

中部市場蔬菜 作物價格預測

114753209 資訊碩一 涂冠瑛 108104049 哲學四 宋岷叡

110703057 資訊四 陳芎月 110304032 統計四 宋庭萱

110405098 統計三 呂欣蓉 110207317 金融四 吳彥勳



目錄

01 簡介

02 資料蒐集與整理

- 農產品價格資料
- 天氣觀測資料

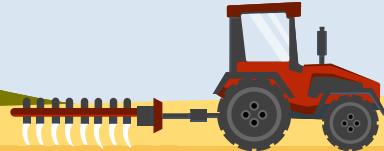
03 預測模型

- Arima
- XGBoost
- RandomForest

04 結論

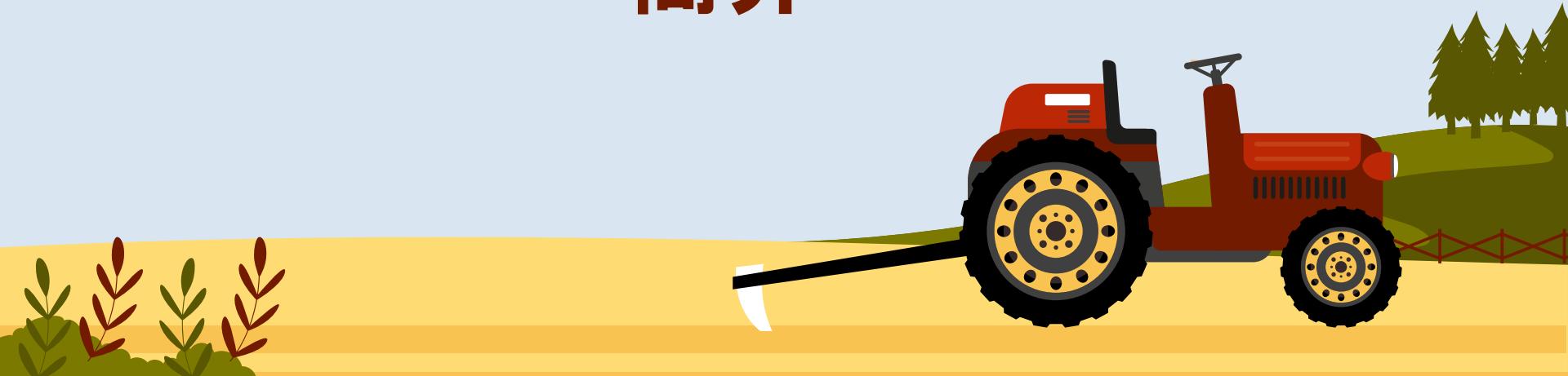
05 Demo

06 附錄



01

簡介



研究動機

生活經驗啟發

從日常觀察中發現蔬果價格與天氣變化密切相關，如冬季高麗菜價格下跌，颱風過後特定蔬菜價格暴漲

氣候因素影響

氣溫、降雨量、日照時數等氣象條件直接影響農作物生長狀況，進而影響產量與價格

市場價格波動

蔬果價格波動幅度大，影響民生消費與農民收益，準確預測有助穩定市場

預測模型需求

建立科學模型量化天氣與蔬果價格關係，提供市場參考依據

研究架構

天氣是否能預測蔬菜價格？

蔬果價格資料

主要抓取中部果菜市場的
交易價格

農業天氣資料

抓取蔬果交易日前三個月
的天氣資料作為變數

RandomForest

除了天氣之外納入歷史交
易價格進行預測

XGBoost

以天氣作為變數進行預測

Arima

時間序列模型納入天氣外
生變數進行預測

工作分配與規劃

04/18

資料搜集

冠瑛、欣蓉 負責交易資料搜集
彥勳負責天氣資料蒐集

05/04

模型建置

芎月用 XGBoost 建模
庭萱以 Arima 預測
珉叡用 Random Forest 分析

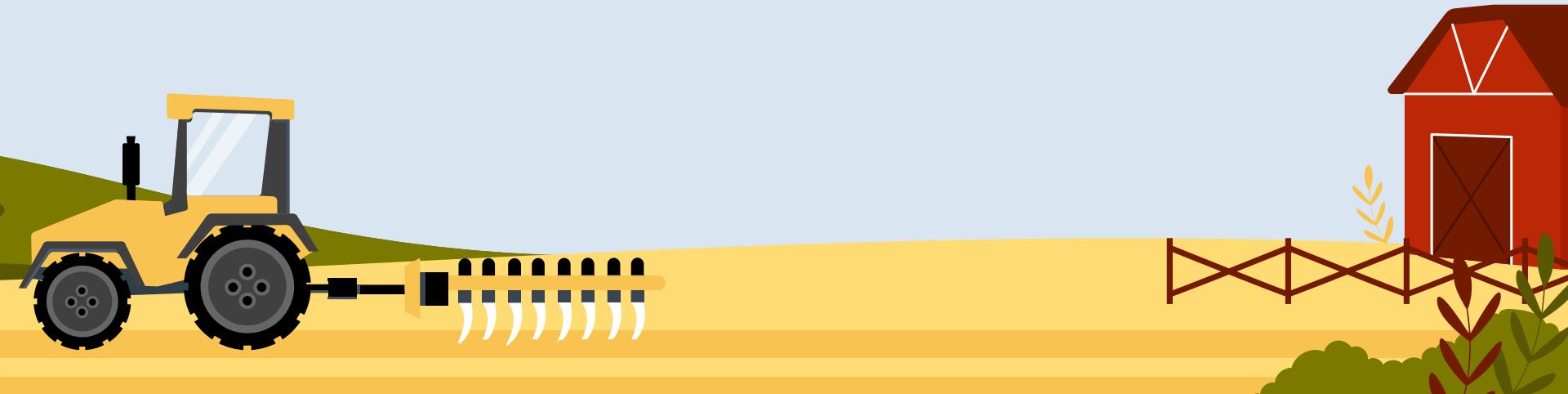
05/25

海報、簡報製作

簡報根據上述分配進行分工
芎月負責海報設計

02

資料蒐集與整理





農產品價格資料 - 資料集下載

1. 選擇三種為短期作物且產地集中於中部的蔬菜：
 - 花椰菜(青梗)
 - 茼蒿
 - 甜豌豆
2. 使用網站：農產品批發市場交易行情站
3. 資料集期間：2015/01/01 ~ 2025/04/30
4. 中部市場名稱：台中市，豐原區，永靖鄉，溪湖鎮，南投市，西螺鎮



農產品價格資料 - 資料集下載

資料集欄位

欄位名稱	資料型態	單位
日期	時間	每天/每週
市場	類別型	台中市, 豐原區, 永靖鄉, 溪湖鎮, 南投市, 西螺鎮
上價	數值型	元 / 公斤
中價	數值型	元 / 公斤
下價	數值型	元 / 公斤
平均價	數值型(目標)	元 / 公斤
增減%	數值型	元 / 公斤
交易量	數值型	公斤
增減%	數值型	公斤



農產品價格資料 - 資料集整理

1. 移除增減%之欄位
2. 確認是否有缺失值：
 - 缺失值僅1-2筆，故直接刪除
3. 將日資料整理為周資料
 - 日期整理為兩個欄位：年分、周次
 - 平均值：上價，中價，下價，平均價，交易量
 - 加總值：交易量



氣象資料合併方法

01

資料選用方法

選擇中部地區，資料完整之氣象站

02

氣象資料整理方法

去除不合理的極值、多種補值方法

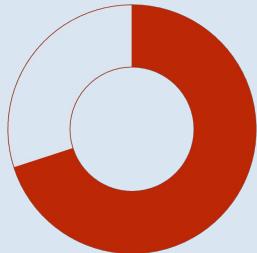
03

與蔬果資料合併方法

將資料集從每日資料轉換為每週資料，並以週為單位合併三個月的天氣資料



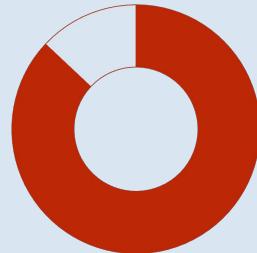
氣象資料選用—近產地的氣象站(年產量)



70%

花椰菜

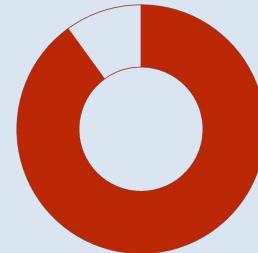
近七成位於彰化縣



87%

豌豆

近九成來自彰化線



90%

萵苣

九成來自雲林縣

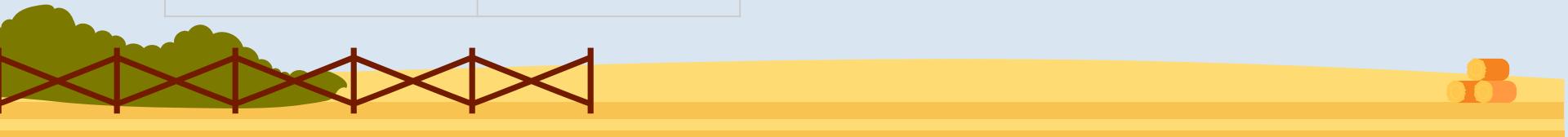




選擇 2014 至今資料完整的觀測站

雲林線農業氣象觀測站	
測站中文名稱	測站資料起始時間
口湖工作站	2018/12/5
南改斗南分場	1987/1/1
水試臺西試驗場	1990/1/2
臺大雲林校區	2019/12/12
海口故事園區	2021/12/29
草嶺	2024/1/8
四湖植物園	2018/11/21
麥寮合作社	2019/7/22

彰化縣農業氣象觀測站	
測站中文名稱	測站資料起始時間
臺中農改	1987/1/1
王功漁港	2021/12/29
北斗	2024/1/8
芳苑	2024/1/8
鹿港	2024/1/8





通霄鎮
公館
苗栗縣

測站、產地相對位置



氣象資料概覽

資料來源

農業氣象觀網往監測系統

選擇日資料

資料期間

2014/10/1 - 2025/4/30

抓取農業資料期間 + 前 3 個月

類別	欄位名稱
氣壓	平均氣壓(hPa)、日最高氣壓(hPa)、日最低氣壓(hPa)
氣溫	平均氣溫(°C)、最高氣溫(°C)、最低氣溫(°C)
露點/濕度	平均露點溫度(°C)、平均相對溼度(%)、最低相對溼度(%)
風速	平均風速(m/s)、最大陣風風速(m/s)
降雨	累計雨量(mm)、最大 10 分鐘降水量(mm)、最大 60 分鐘降水量(mm)
日照	累積日照時數(hr)、累積日射量(MJ/m ²)
地溫	平均地溫：5cm、10cm、20cm、50cm、100cm(單位：°C)

極值處理—將顯著不合理 值設為 NA

地區	欄位名稱	最小值	最大值	問題說明
雲林	最低氣溫(°C)	-39.6	28.5	台灣平地不可能出現此低溫，應為錯誤值
雲林	平均相對溼度(%)	0	99	0% 幾乎不可能，可能為缺值填零或記錄錯誤
雲林	最低相對溼度(%)	0	99	同上，0% 為不合理值
彰化	最低氣溫(°C)	-4.9	29.9	雖非完全不可能，但已屬台灣極罕見值，應確認是否為誤植
彰化	最低相對溼度(%)	0	100	同上，0% 為不合理值

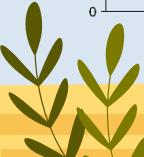
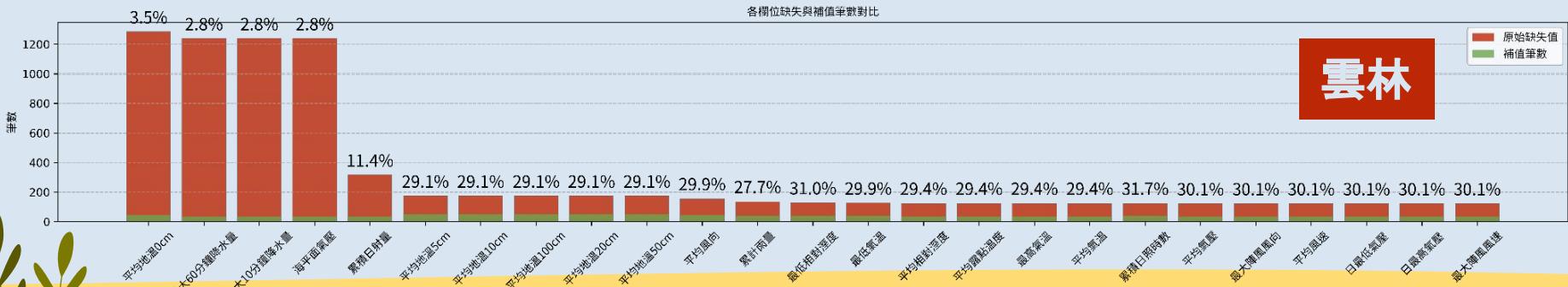
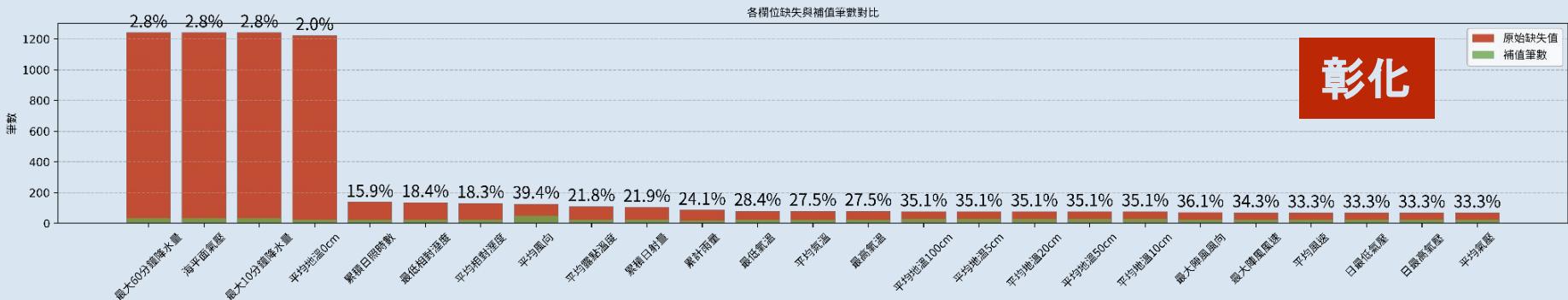


交叉比較—將差異顯著不合理者設為 NA

欄位名稱	索引	雲林數值	彰化數值	異常原因
平均氣溫(°C)	324	0	29.8	雲林氣溫為 0°C 明顯錯誤
最高氣溫(°C)	324	0	35	雲林氣溫為 0°C 明顯錯誤
最低氣溫(°C)	324	0	25.7	雲林氣溫為 0°C 明顯錯誤
平均露點溫度(°C)	324	0	26	雲林露點為 0°C 明顯錯誤
平均相對溼度(%)	324	0	81	雲林濕度為 0% 不合理
最低相對溼度(%)	324	0	61	雲林濕度為 0% 不合理
平均地溫0cm(°C)	2821	20.4	30.1	溫度差異超過 10°C, 可能誤植
平均地溫5cm(°C)	2821	20.1	30.2	溫度差異超過 10°C, 可能誤植
平均地溫10cm(°C)	2821	20.2	30.1	溫度差異超過 10°C, 可能誤植
平均地溫20cm(°C)	2821	19.8	30.1	溫度差異超過 10°C, 可能誤植
平均地溫50cm(°C)	2821	19.4	29.4	溫度差異超過 10°C, 可能誤植
平均地溫100cm(°C)	2821	19	28.3	溫度差異超過 10°C, 可能誤植

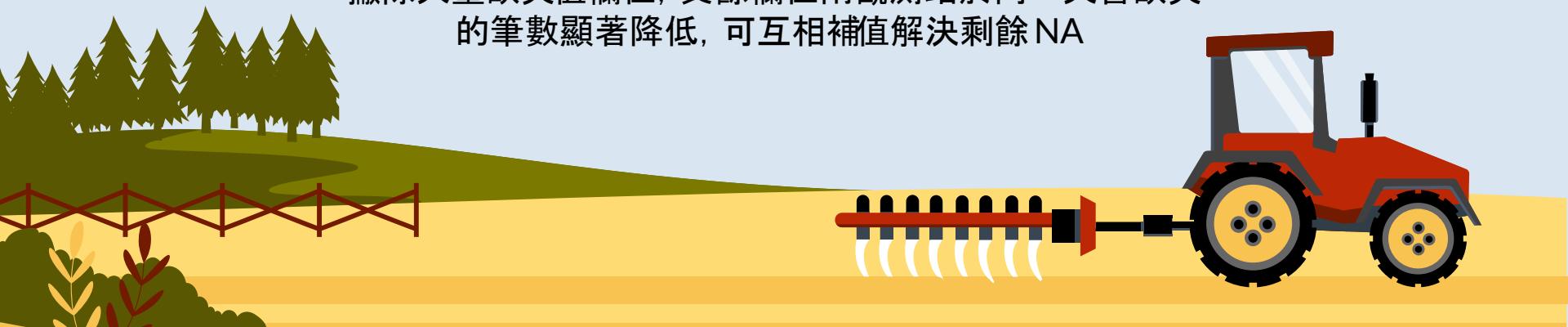


補值—取前一天與後一天 值取平均





16 → 6



撤除大量缺失值欄位，其餘欄位兩觀測站於同一天皆缺失的筆數顯著降低，可互相補值解決剩餘 NA

合併兩觀測站資料

01

如果彰化、雲林皆有 值

如果欄位為平均欄位，則取兩者平均，如果是極 值欄位，則選擇極值(Max/min)

02

如果彰化、雲林一方為 NA

取單一值作為該時間點的欄位 值

03

如果兩者皆為 NA

保留為 NA，但在下一階段轉換為週資料時有可能進行插 值



轉換為週資料並與蔬果資料合併

01

極值取極值，平均取平均

如果該週每週皆有值，則平均欄位取平均，極值欄位取極值

02

累積型變數強制要求一週至少六天有值

累積型變數(如累積雨量)因為其差一天值會差異極大，因此如果有超過 2天缺值則記 NA

03

與交易資料合併

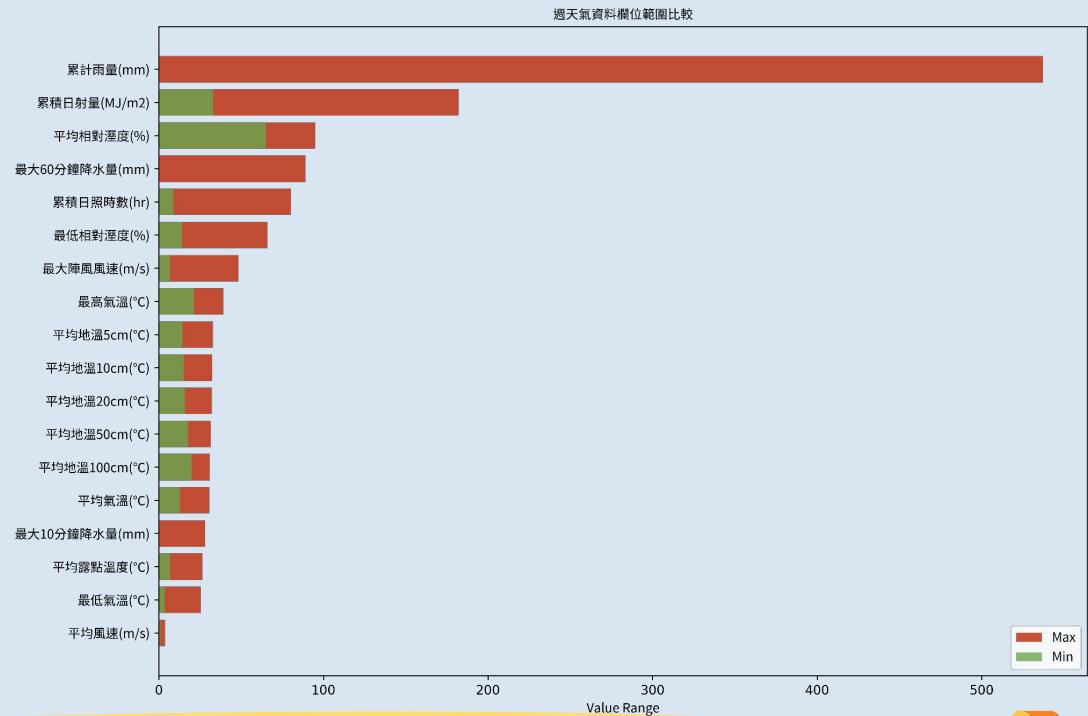
以交易日該週為 w，合併 w ~ w-12 週的所有天氣欄位資料(共 283 個欄位)





Brief Summary

Variable	Missing_Values
累計雨量(mm)	5
最大10分鐘降水量(mm)	170
最大60分鐘降水量(mm)	170
累積日照時數(hr)	4
累積日射量(MJ/m2)	10





03

預測模型

三種模型

01

ARIMA

使用週資料集，以花椰菜為例。進行永靖 鄉、溪湖鎮、西螺鎮三個市場進行建模

02

XGBoost

使用週資料集，以花椰菜為例。進行永靖 鄉、溪湖鎮、西螺鎮三個市場進行建模

03

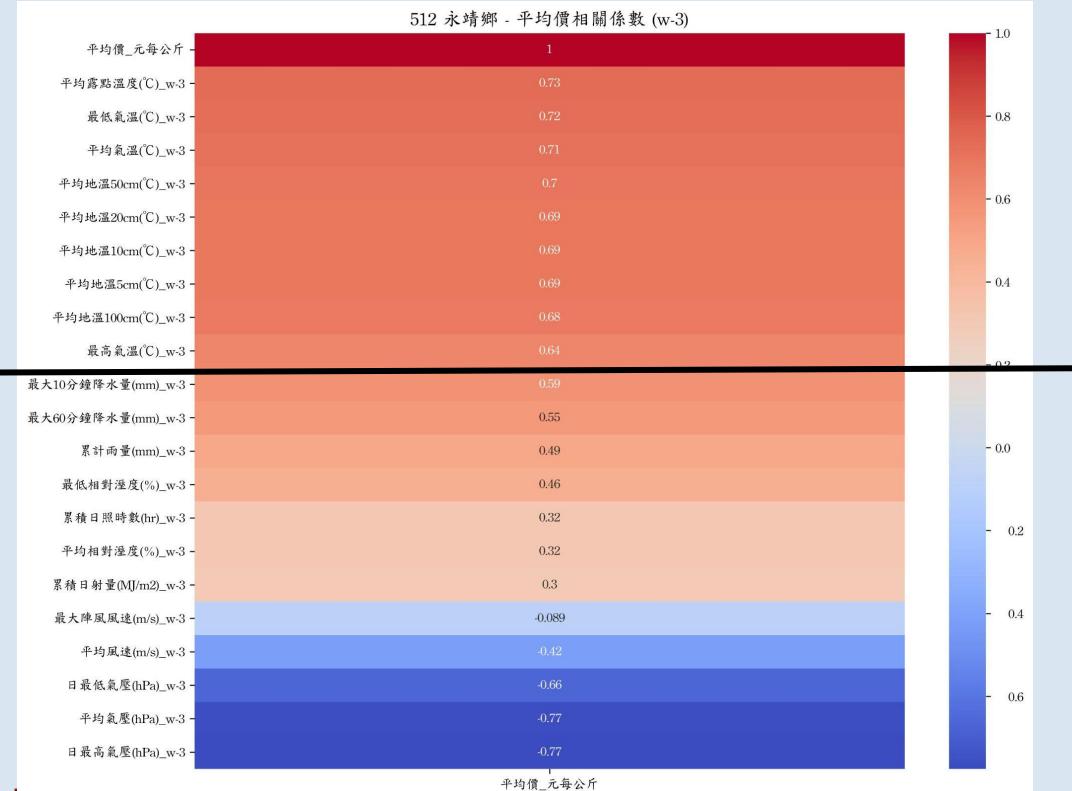
RandomForest

使用日資料集，以花椰菜為例。進行永靖 鄉、溪湖鎮、西螺鎮三個市場進行建模

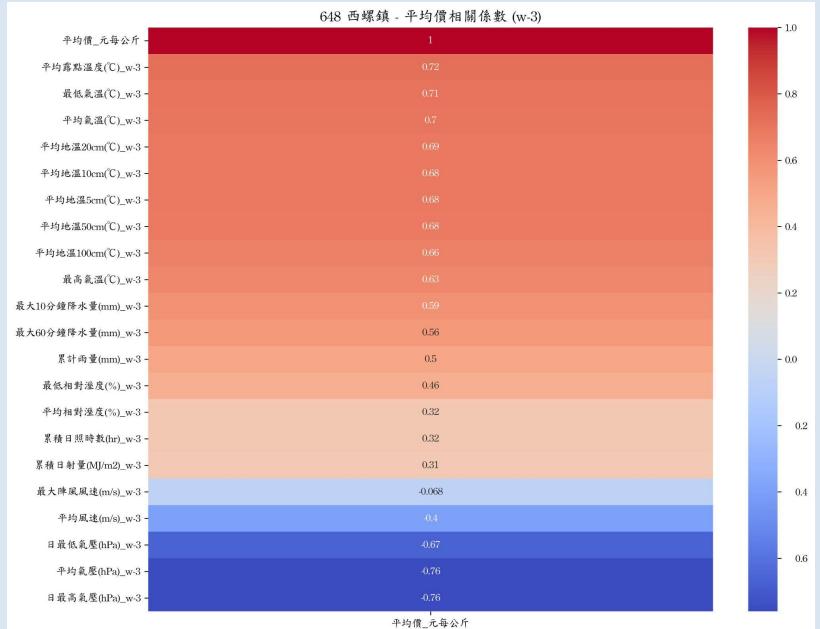
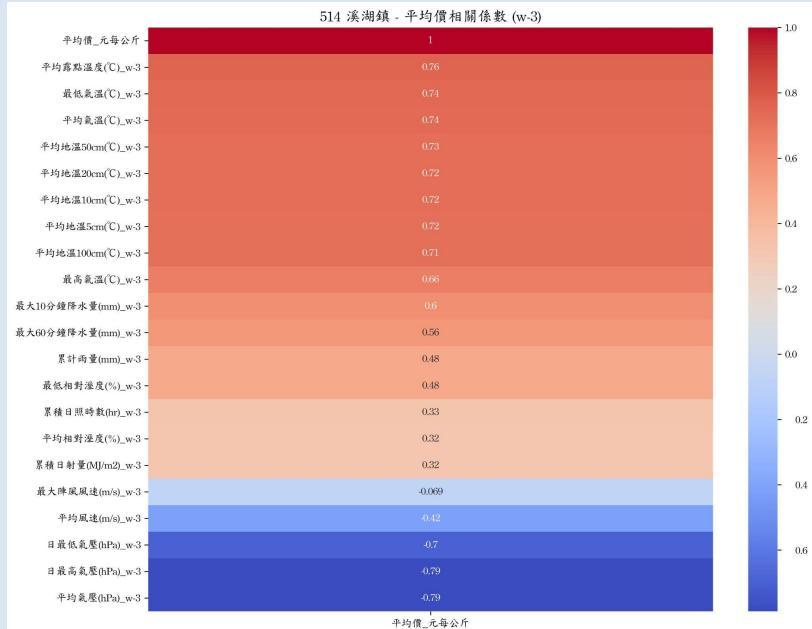


Coefficient

溫度相關的特徵，
相關係數都比較高



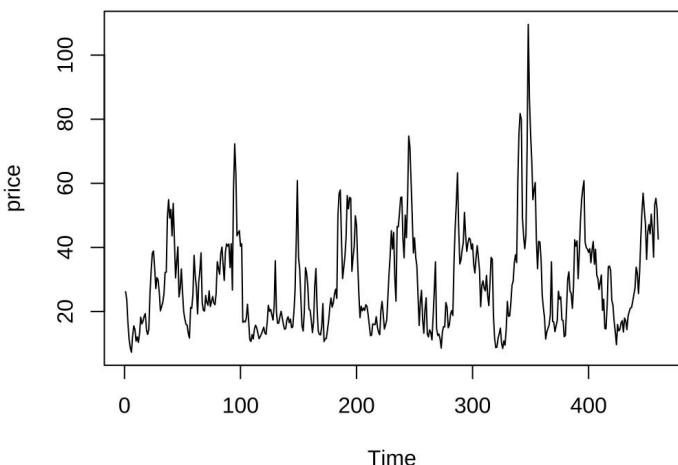
Coefficient



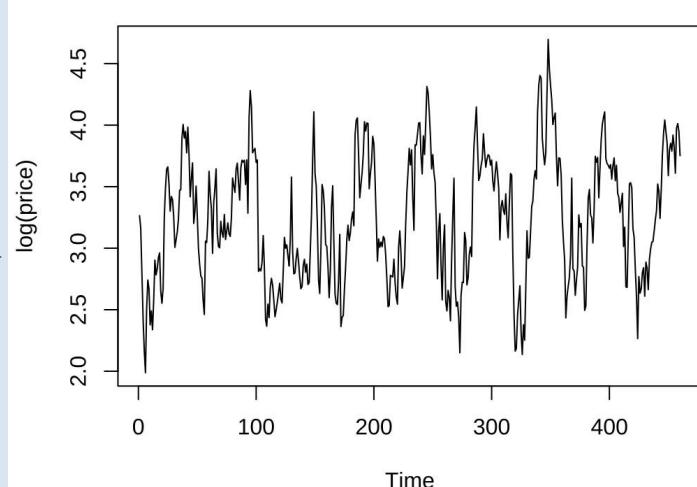
ARIMA model

以512永靖鄉資料為例

以平均地溫100cm-w3當作外生變數來做迴歸detrend



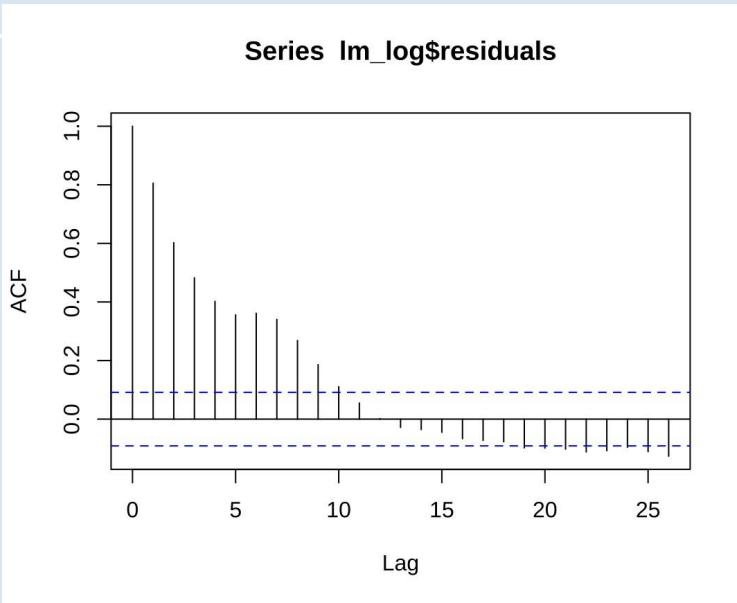
原time series plot



進行detrend後的time series plot

ARIMA model

以512永靖鄉資料為例



```
> adf.test(lm_log$residuals)

Augmented Dickey-Fuller Test

data: lm_log$residuals
Dickey-Fuller = -5.0212, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

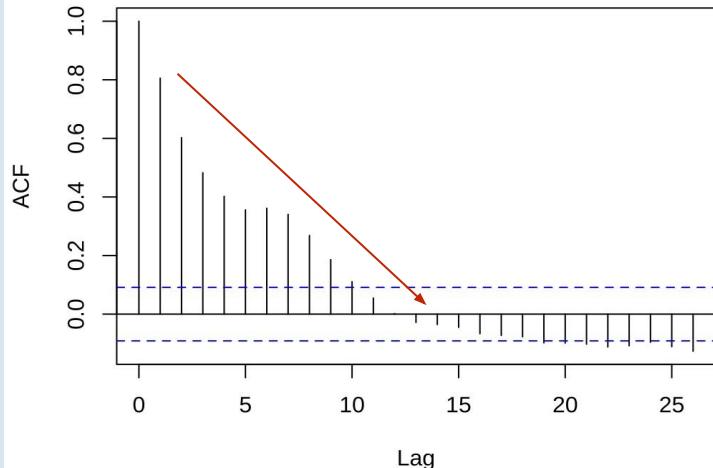
Warning message:
In adf.test(lm_log$residuals) : p-value smaller than printed p-value
```

- **adf.test**
 - H_0 : the time series is not stationary
 - H_1 : the time series is stationary
- **p-value < 0.05 , reject**
- **stationary**

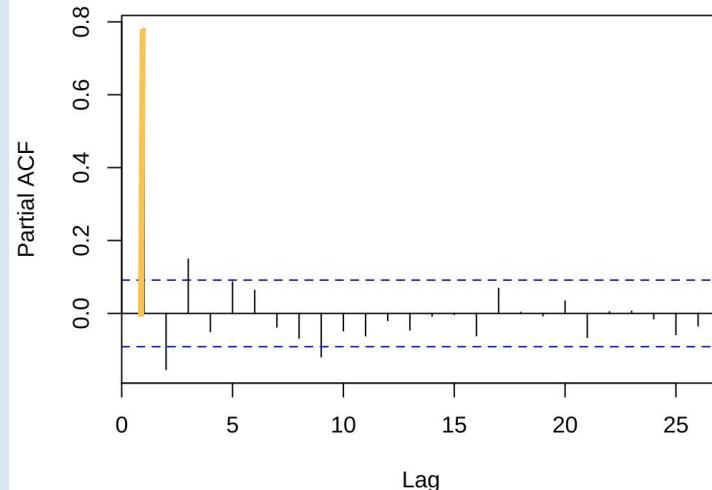
ARIMA model

以512永靖鄉資料為例

Series lm_log\$residuals



Series lm_log\$residuals



- nonseasonal: tail off

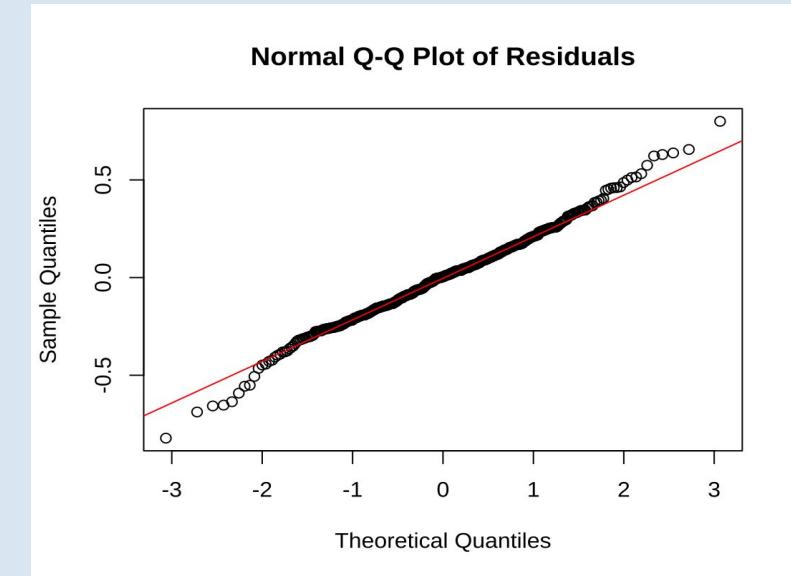
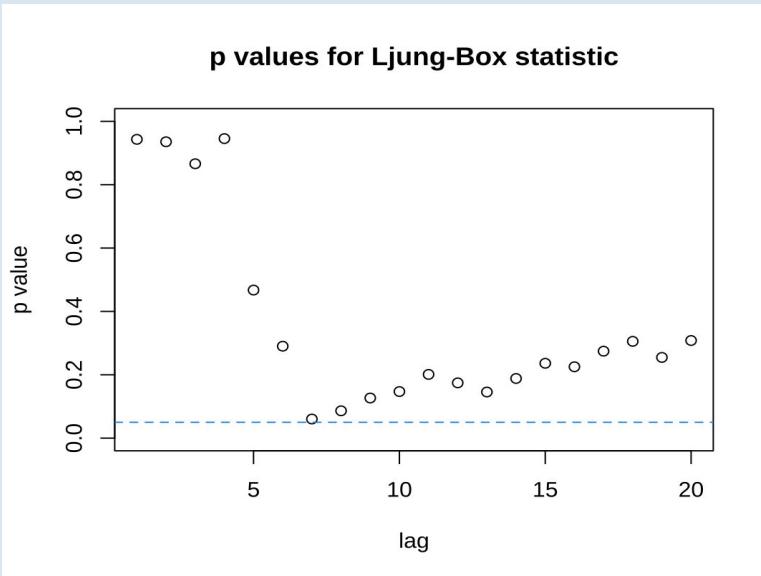
- nonseasonal: cut off at lag =1

典型AR(1)

ARIMA model

以512永靖鄉資料為例

ARIMA(3, 0, 1)



H_0 : 殘差無自相關(為白噪音)

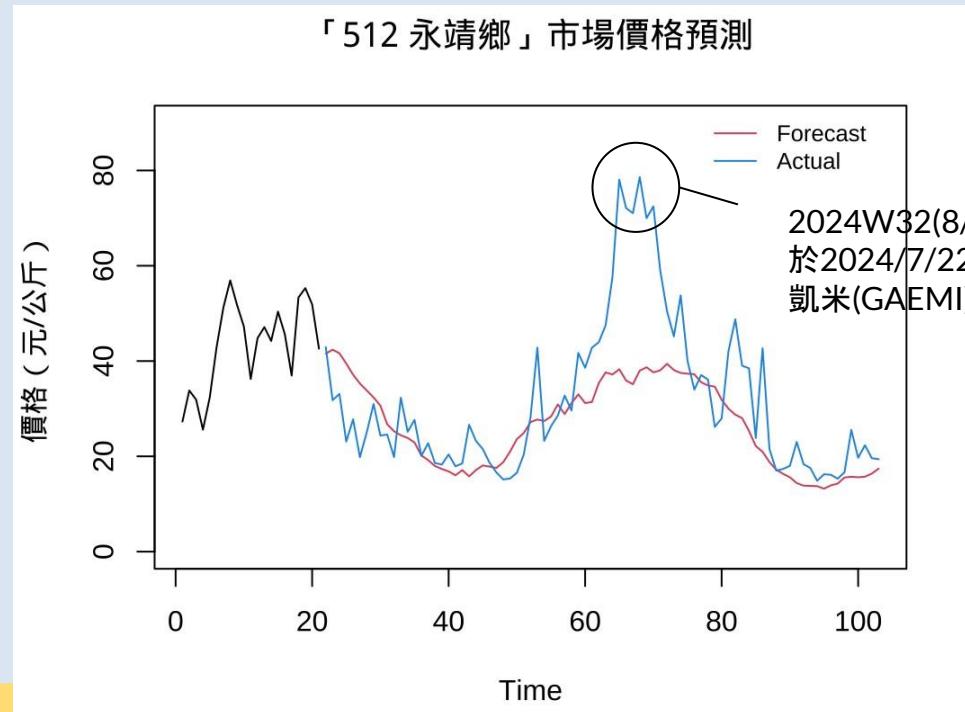
H_1 : 殘差存在自相關

ARIMA model

以512永靖鄉資料為例

ARIMA(3, 0, 1)

$R^2 = 0.3552$
 $RMSE = 12.8866$
 $MAE = 8.4518$

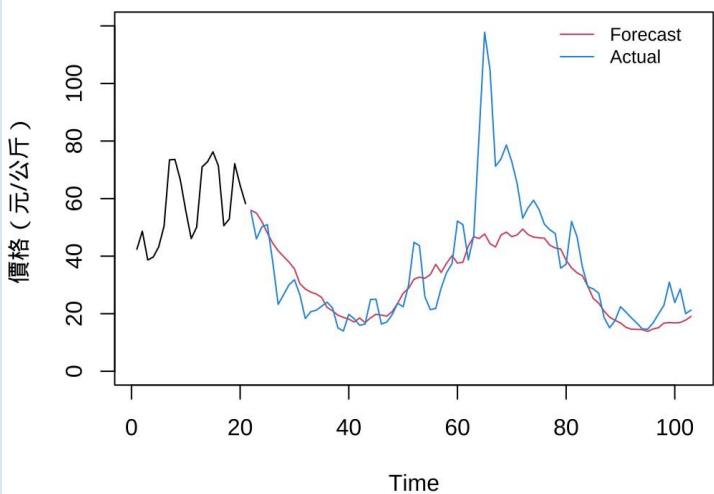


ARIMA model

514 溪湖鎮、648 西螺鎮

ARIMA(3, 0, 1)

「514 溪湖鎮」市場價格預測

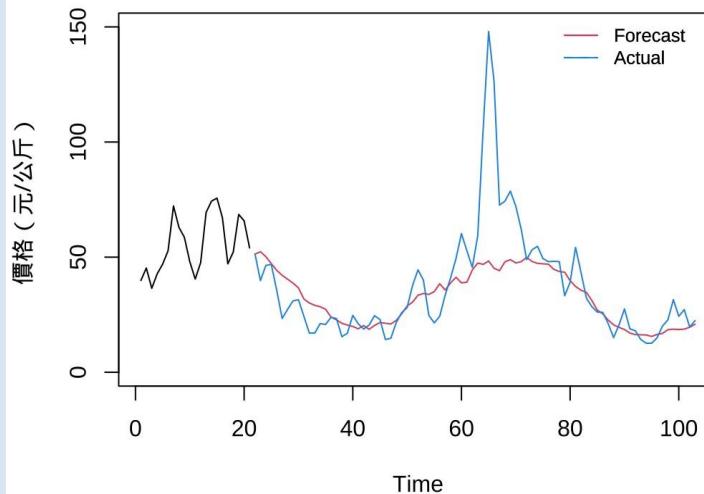


$$R^2 = 0.5191$$

$$RMSE = 14.4107$$

$$MAE = 8.5927$$

「648 西螺鎮」市場價格預測



$$R^2 = 0.4184$$

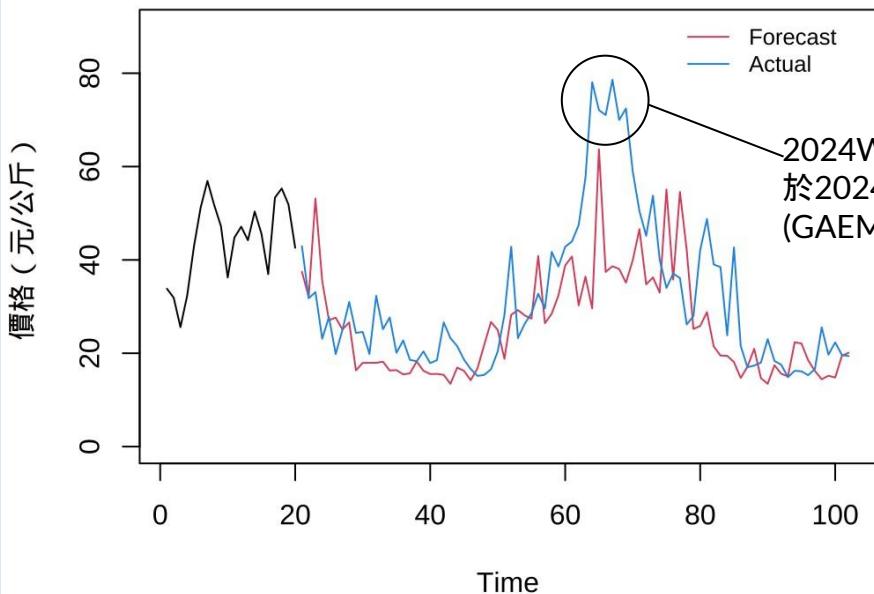
$$RMSE = 18.2386$$

$$MAE = 9.4673$$



Xgboost model

「512 永靖鄉」市場價格預測



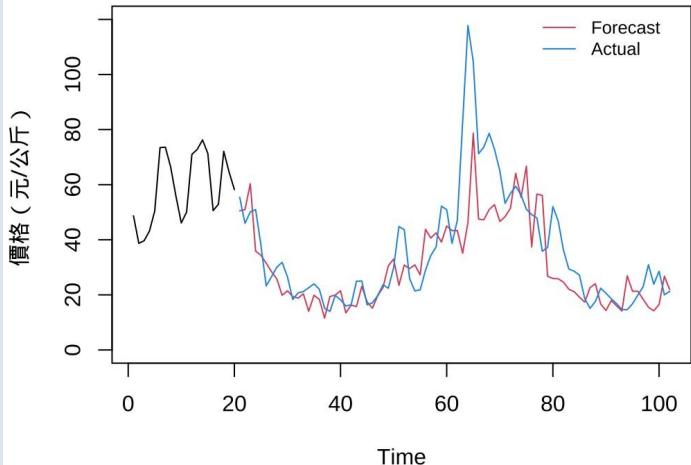
```
> cat("RMSE : ", rmse, "\n")
RMSE : 12.88985
```

```
> cat("MAE : ", mae, "\n")
MAE : 8.874705
```

```
> cat("R-squared : ", r_squared, "\n")
R-squared : 0.3549066
```

Xgboost model

「514 溪湖鎮」市場價格預測

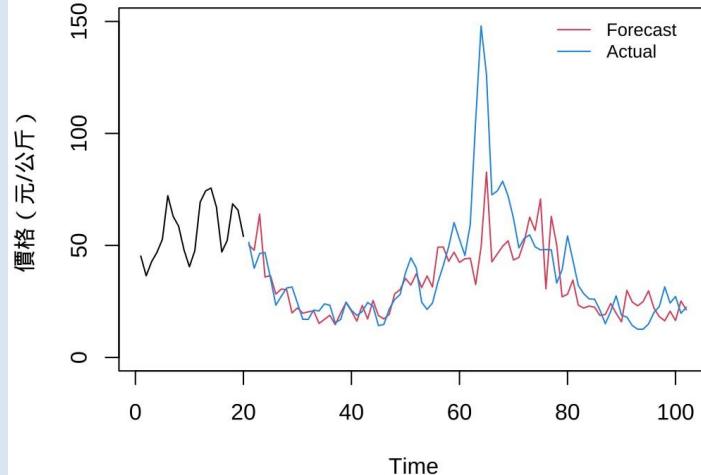


```
> cat("RMSE : ", rmse, "\n")
RMSE : 15.13406

> cat("MAE : ", mae, "\n")
MAE : 9.796615

> cat("R-squared : ", r_squared, "\n")
R-squared : 0.4696208
```

「648 西螺鎮」市場價格預測



```
> cat("RMSE : ", rmse, "\n")
RMSE : 17.87373

> cat("MAE : ", mae, "\n")
MAE : 10.76841

> cat("R-squared : ", r_squared, "\n")
R-squared : 0.4413238
```

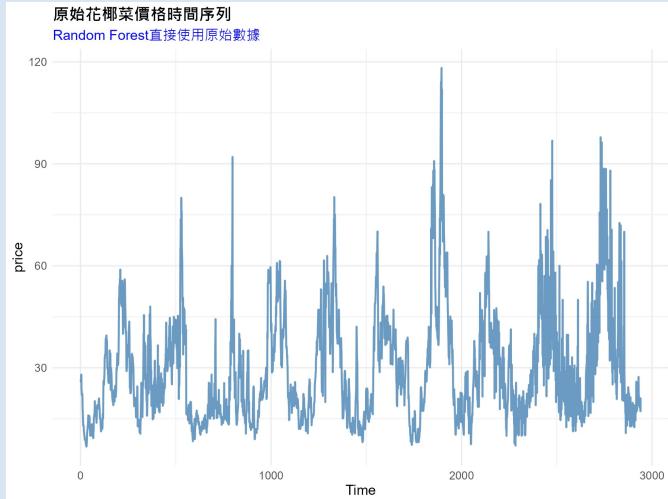




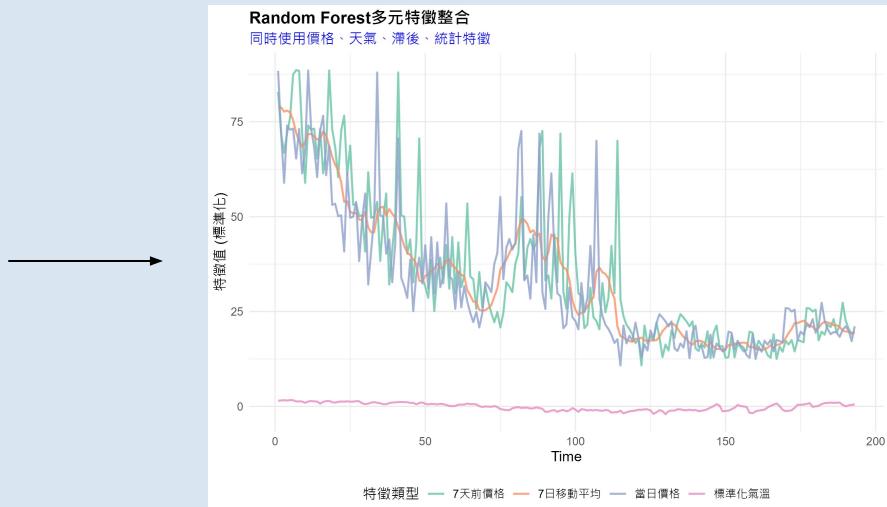
Ranger model

以512永靖鄉資料為例

Random Forest直接處理原始非平穩數據，無需detrend預處理



原始花椰菜價格時間序列

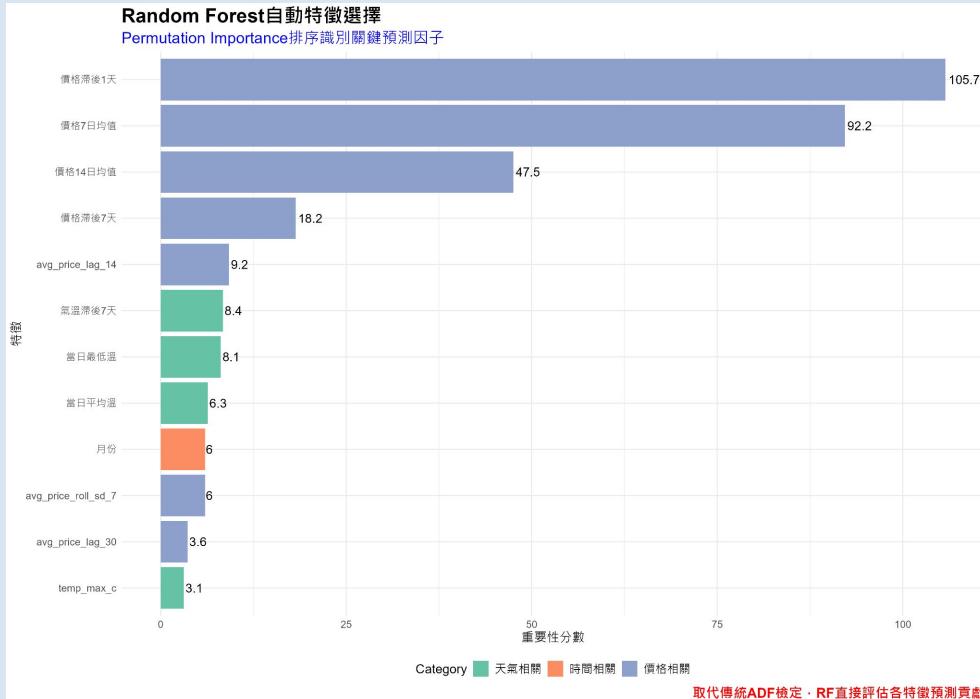


Random Forest多元特徵整合



Ranger model

以512永靖鄉資料為例



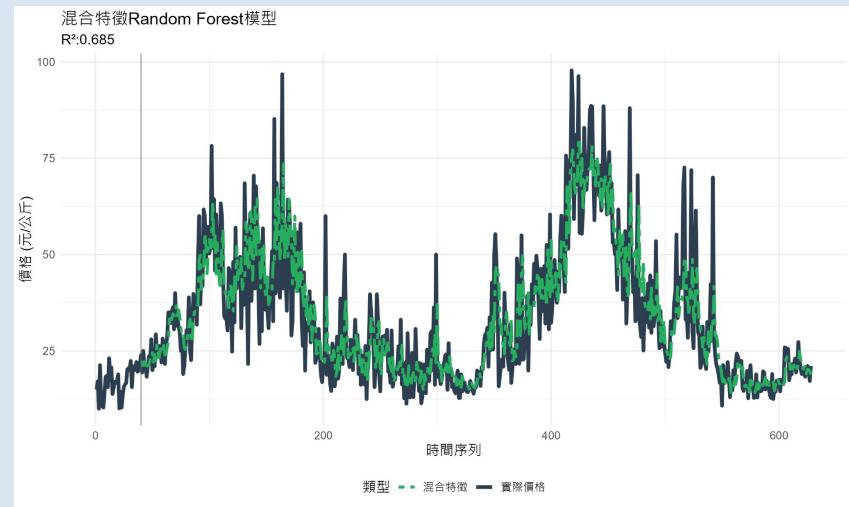
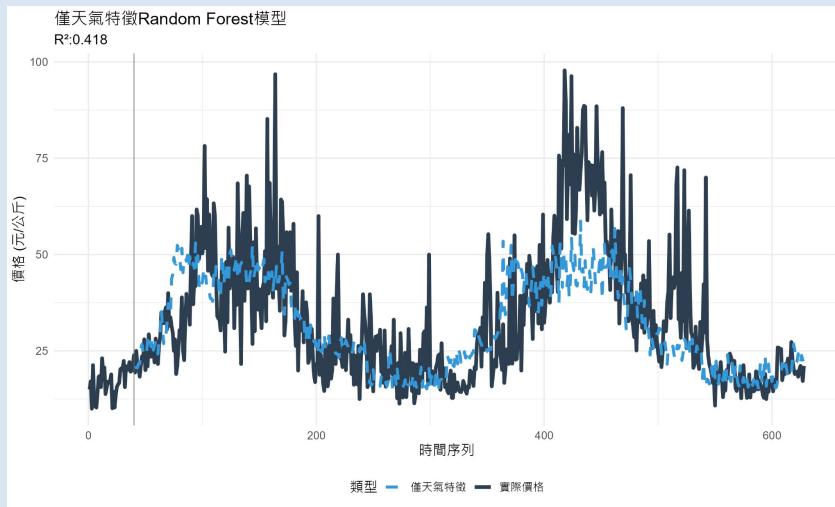
- Random Forest 自動特徵選擇，
無需手動平穩性檢定
- Permutation Importance 取代
統計假設檢驗

Ranger model

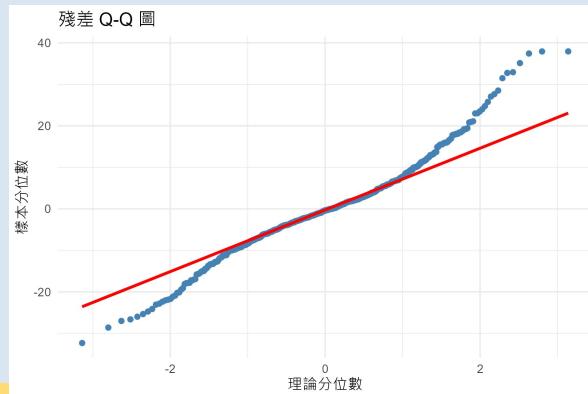
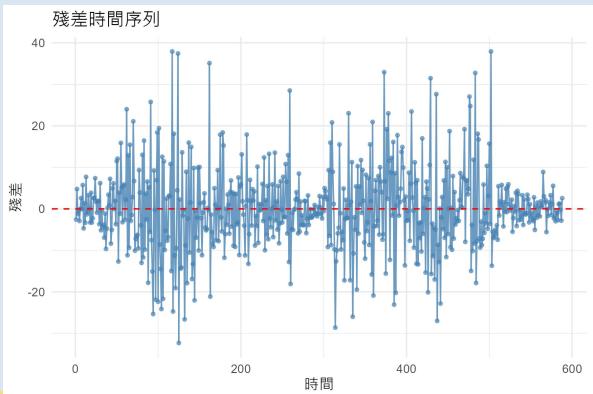
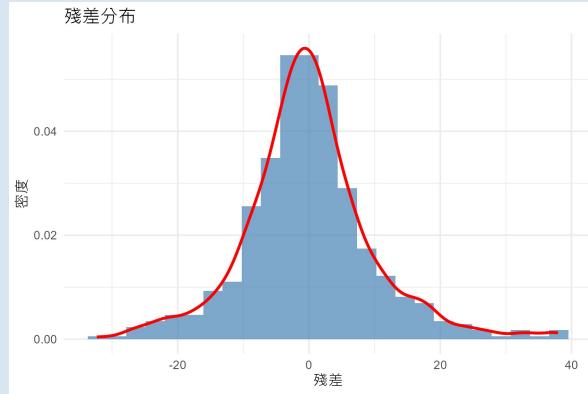
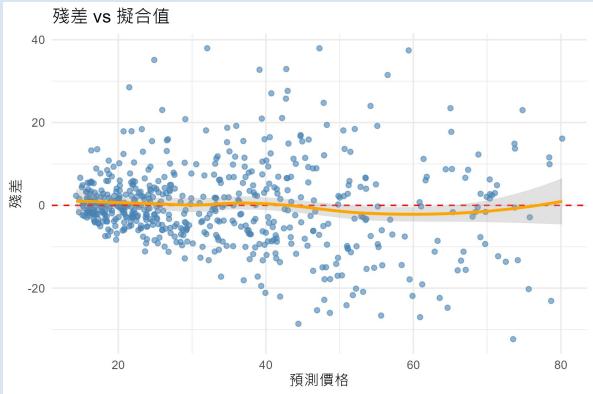
以512永靖鄉資料為例



Random Forest模型比較: 僅評估天氣因素 vs 納入價格連續性

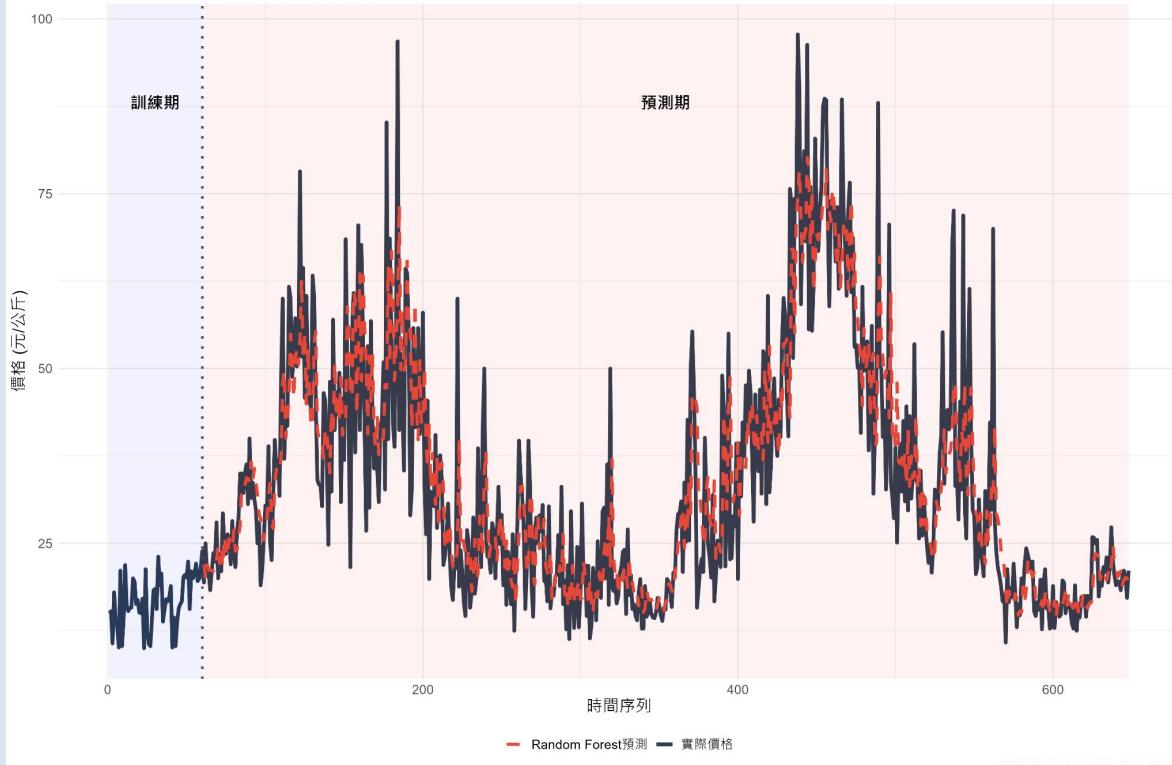


Random Forest模型診斷: 殘差分析與預測穩健性檢驗



永靖鄉 Random Forest花椰菜價格預測

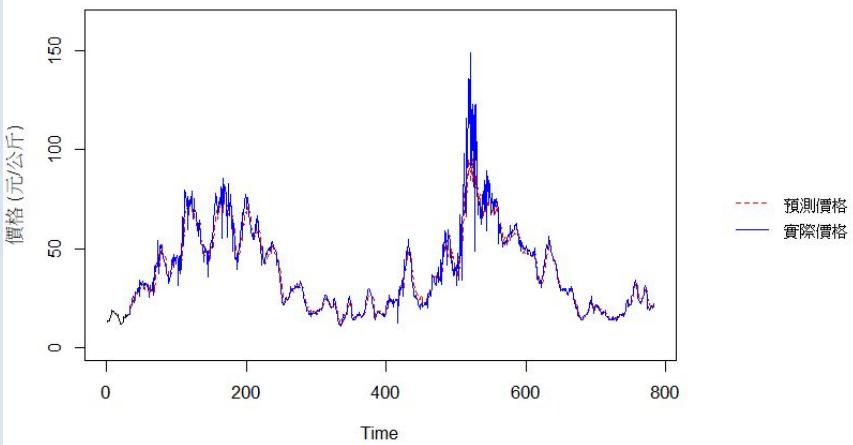
RF極端值處理能力與波動捕捉效果展示



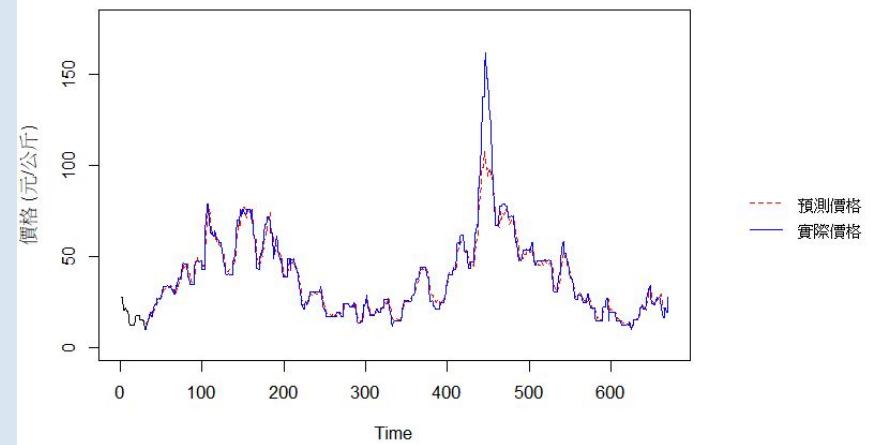
Ranger model

514 溪湖鎮、648 西螺鎮

「514 溪湖鎮」市場花椰菜價格預測

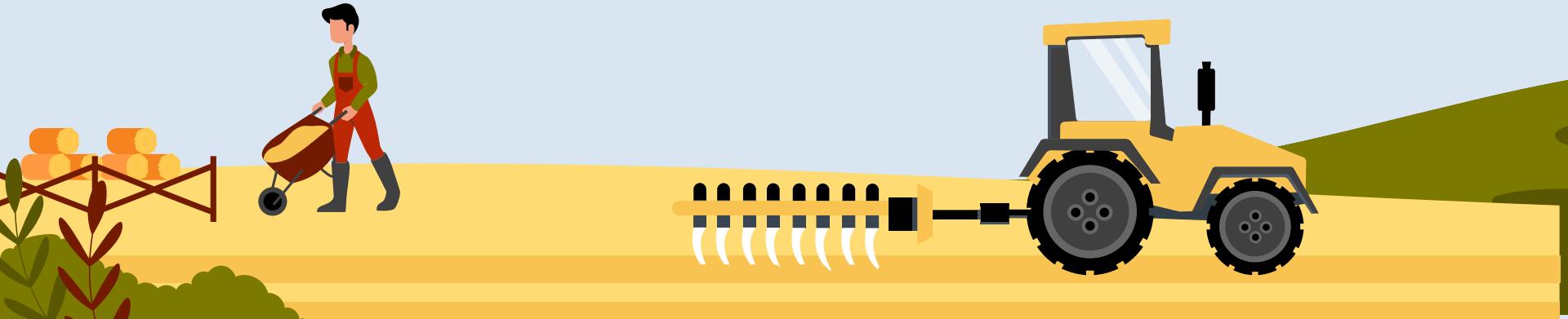


「648 西螺鎮」市場花椰菜價格預測



04

結論



研究限制與未來改進方向

01

聚焦於花椰菜價格預測

資料完整，無季節性缺乏資料之問題，然而專注於單一蔬菜限制了模型的泛化能力

02

ARIMA 與 XGBoost 模型表現差

較難準確預測價格的短期極端變動，尤其是高點的出現時機

03

特徵選擇考量

聚焦於歷史價格、天氣條件（氣溫、降雨）、與市場別等主要變數，但可能排除部分對高點預測具關鍵影響的變數



05

Demo



Shinyapp

<https://hsinjunglu.shinyapps.io/code/>

中部市場蔬菜價格預測

組別: Group 6
組員:
涂冠英, 宋蕙韻, 陳筠月
宋庭萱, 吳欣蓉, 吳彥勳

Home
ARIMA
XGBoost
RandomForest
相關係數

主題簡介

本組探討氣候因素對蔬果價格的影響，透過分析氣溫、降雨量、日照時數等資料，建立價格預測模型，希望找出影響價格波動的關鍵天氣因素，協助消費者、農民與產業做出更有效的決策。

預測模型

此平臺提供三種預測模型之價格預測圖與相關結果：

- ARIMA - 時間序列分析模型
- XGBoost - eXtreme Gradient Boosting模型
- Random Forest - 隨機森林模型

使用說明

- 點擊左側選單選擇您想要查看的模型
- 各模型頁面中可以選擇不同的市場
- 查看模型結果圖表和相關資訊

QRcode 連結



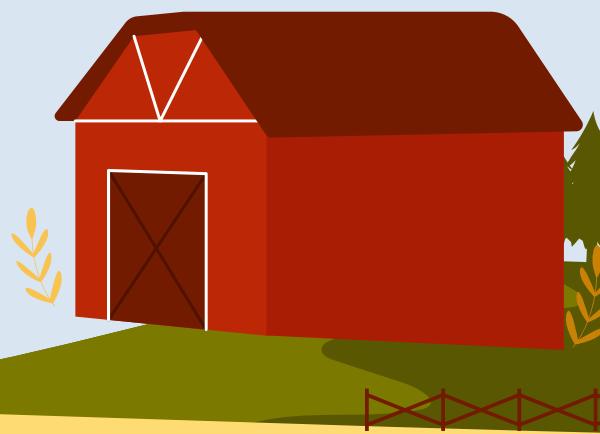


Thanks!

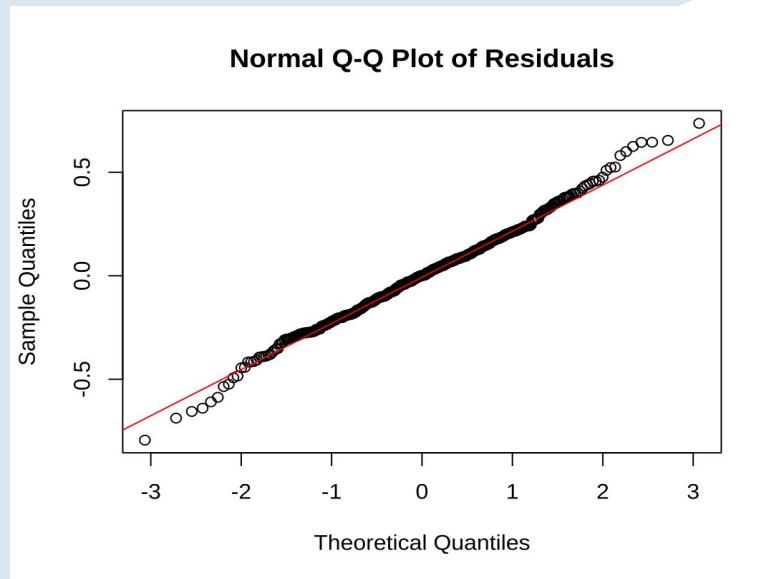
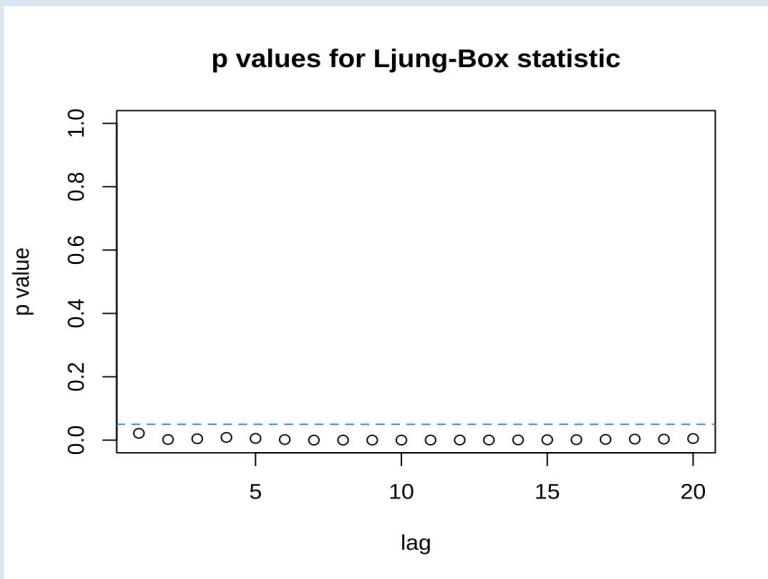


06

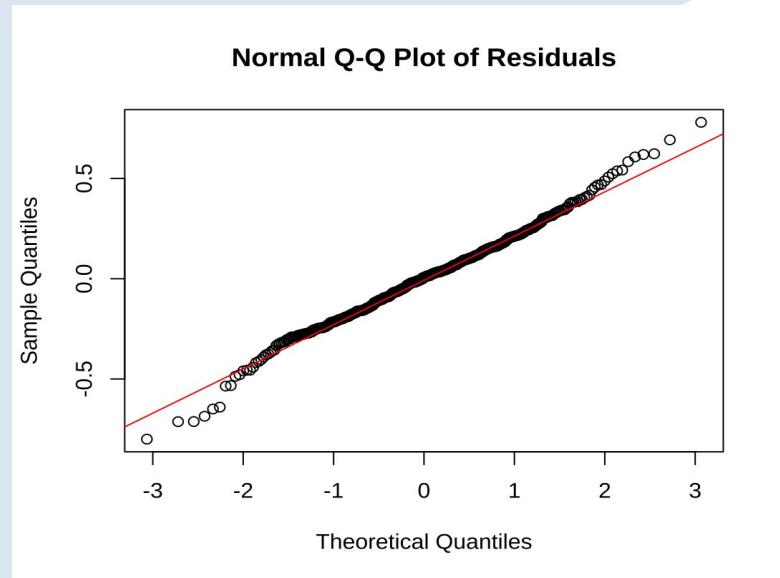
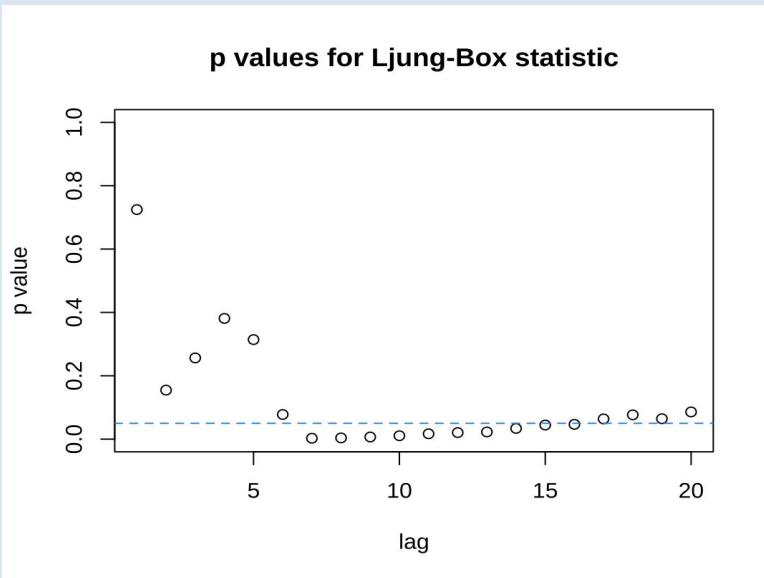
附錄



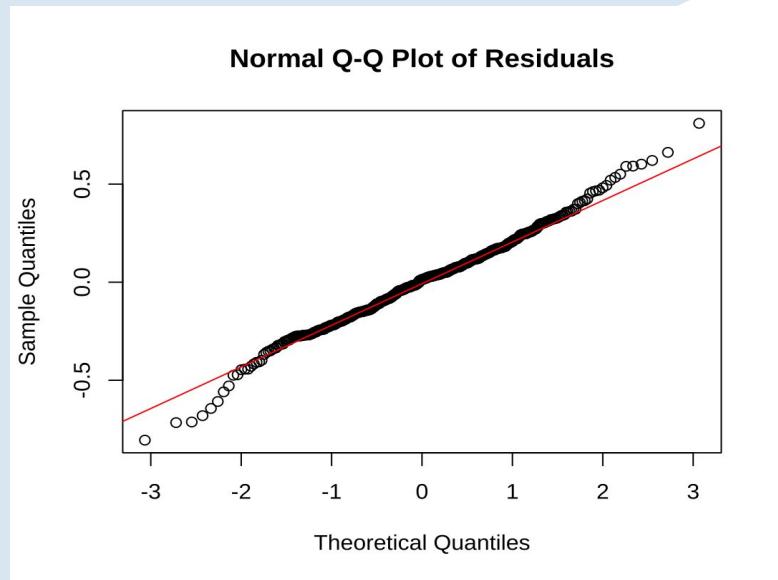
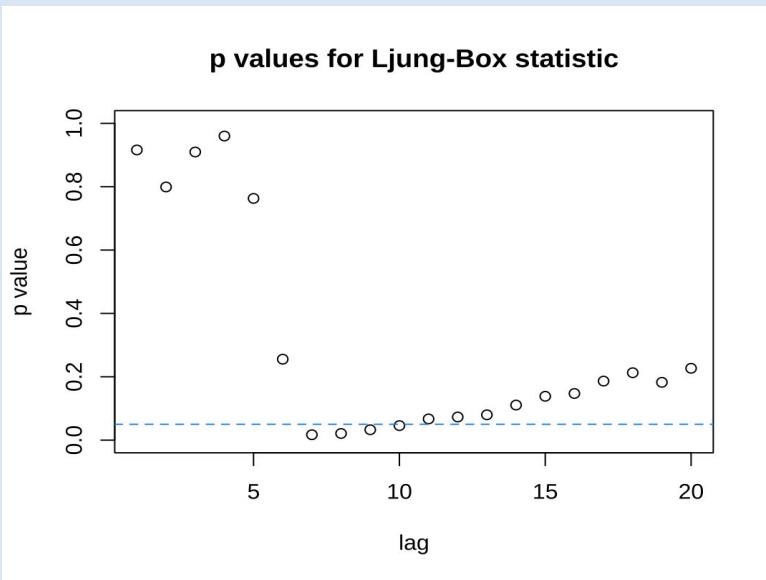
ARIMA (1, 0, 0)



ARIMA (2, 0, 0)



ARIMA (2, 0, 1)





xgboost_512永靖鄉_模型參數

```
> # 設定XGBoost參數
> params <- list(
+   objective = "reg:squarederror",
+   eval_metric = "rmse",
+   eta = 0.1,
+   max_depth = 6,
+   min_child_weight .... [TRUNCATED]

> # 使用訓練集訓練模型，並監控驗證集表現
> xgb_model <- xgb.train(
+   params = params,
+   data = dtrain,
+   nrounds = 1000,
+   watchlist = list(train = dtrain, vali .... [TRUNCATED]
[1]     train-rmse:29.143583    validation-rmse:31.370753
```

```
> cat("\n最佳迭代次數:", best_iteration, "\n")
```

最佳迭代次數: 42



xgboost_514 溪湖鎮_模型參數

```
> # 設定XGBoost參數  
> params <- list(  
+   objective = "reg:squarederror",  
+   eval_metric = "rmse",  
+   eta = 0.1,  
+   max_depth = 6,  
+   min_child_weight .... [TRUNCATED]  
  
> # 使用訓練集訓練模型，並監控驗證集表現  
> xgb_model <- xgb.train(  
+   params = params,  
+   data = dtrain,  
+   nrounds = 1000,  
+   watchlist = list(train = dtrain, vali .... [TRUNCATED]  
[1]     train-rmse:33.614218      validation-rmse:41.792935
```

```
> cat("\n最佳迭代次數:", best_iteration, "\n")
```

最佳迭代次數: 78



xgboost_648 西螺鎮_模型參數

```
> # 設定XGBoost參數  
> params <- list(  
+   objective = "reg:squarederror",  
+   eval_metric = "rmse",  
+   eta = 0.1,  
+   max_depth = 6,  
+   min_child_weight .... [TRUNCATED]  
  
> # 使用訓練集訓練模型，並監控驗證集表現  
> xgb_model <- xgb.train(  
+   params = params,  
+   data = dtrain,  
+   nrounds = 1000,  
+   watchlist = list(train = dtrain, vali .... [TRUNCATED]  
[1]      train-rmse:35.323519      validation-rmse:41.976228
```

```
> cat("\n最佳迭代次數:", best_iteration, "\n")
```

最佳迭代次數: 65

