



## 應用機器學習建立霧之診斷模型

劉正欽、洪景山

2018.09.11

### 凌霧行

秋城海霧重，職事凌晨出。

浩浩合元天，溶溶迷朗日。

霧色漸侵衣，稍視沾衣密。

道狹至不分，郊樹都如失。

霏微誤噓吸，膚腠生寒栗。

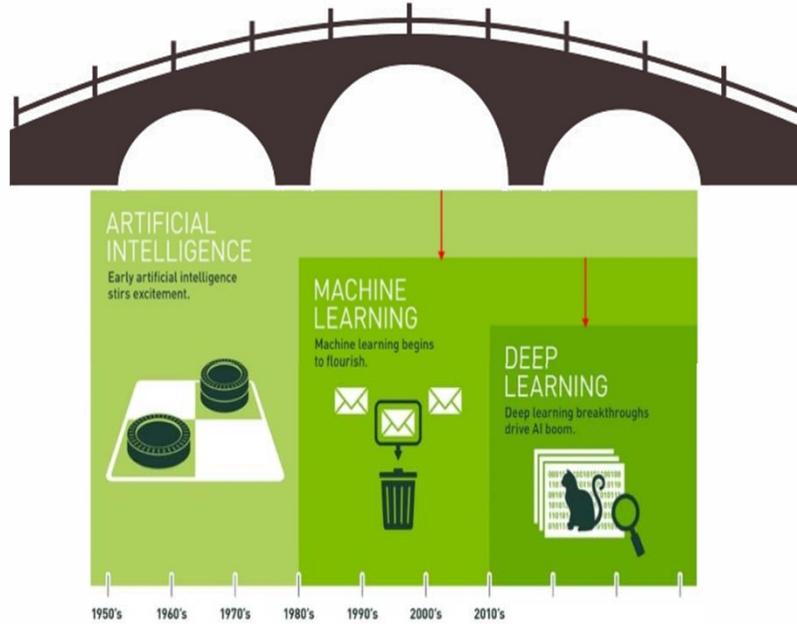
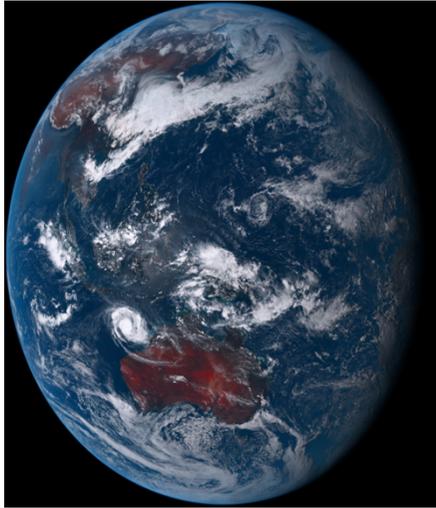
歸當飲一杯，庶用蠲斯疾。

〈唐·韋應物〉

## 大綱

- 動機與目的
- 流程
- 資料分析
- 結果與討論
- 未來工作

# 動機與目的



## 資料科學

利用機器學習技術連結大氣科學與資料科學

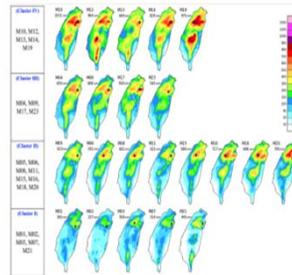


分類、回歸



監督式學習

分群



非監督式學習

分群後分類



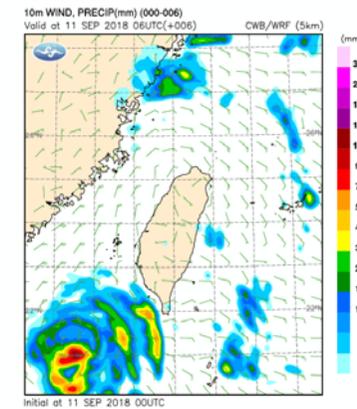
半監督式學習

分群、分類、回歸



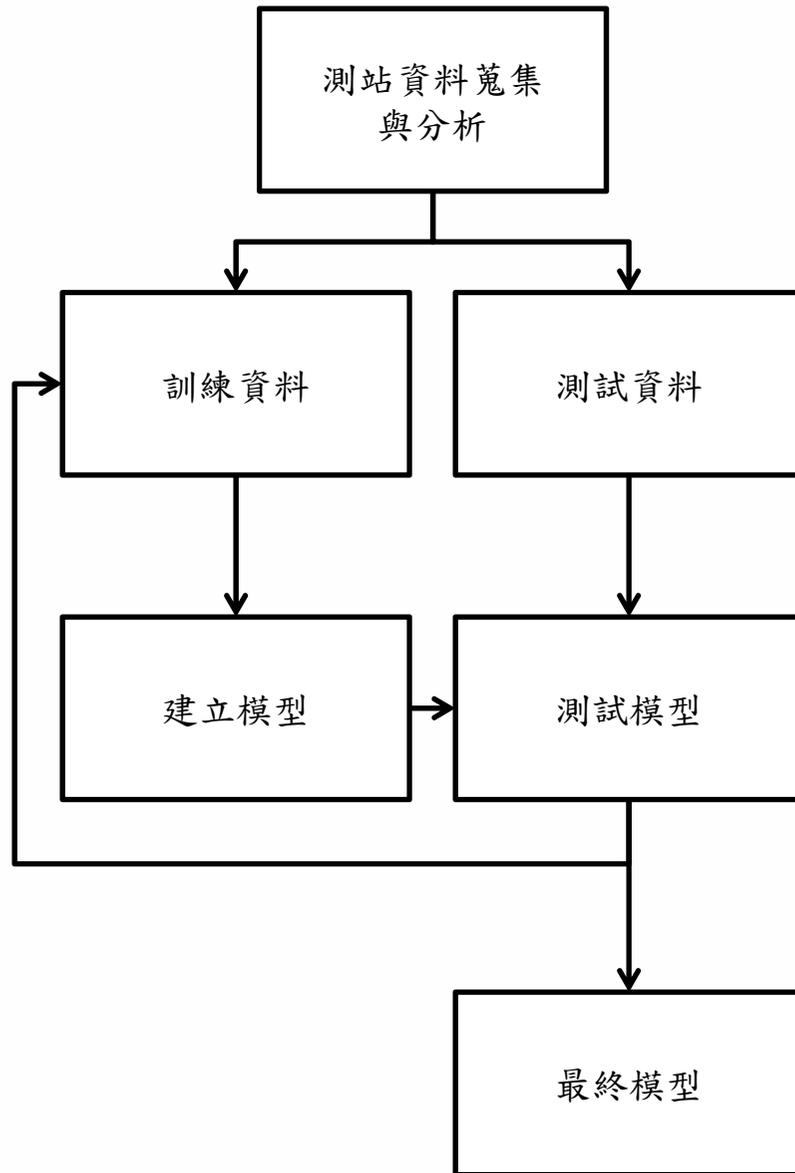
增強式學習

# 動機與目的



1. 數值預報是利用數值方法模擬大氣的物理過程。而針對霧的天氣現象，因其**非線性**過程，在霧是否會生成的診斷上有諸多挑戰。
2. 霧，一種肉眼不易分辨，極細微而密集的水滴所組成，懸浮於近地面的空氣中。根據世界氣象組織的定義，有霧時水平方向之能見度必須不足一公里。一旦生成會嚴重影響陸、海及空之交通安全及運輸規劃。
3. 因大氣資料具有非線的特性，期望透過機器學習**有非線性演算法**的優勢，嘗試發展霧的診斷模型。

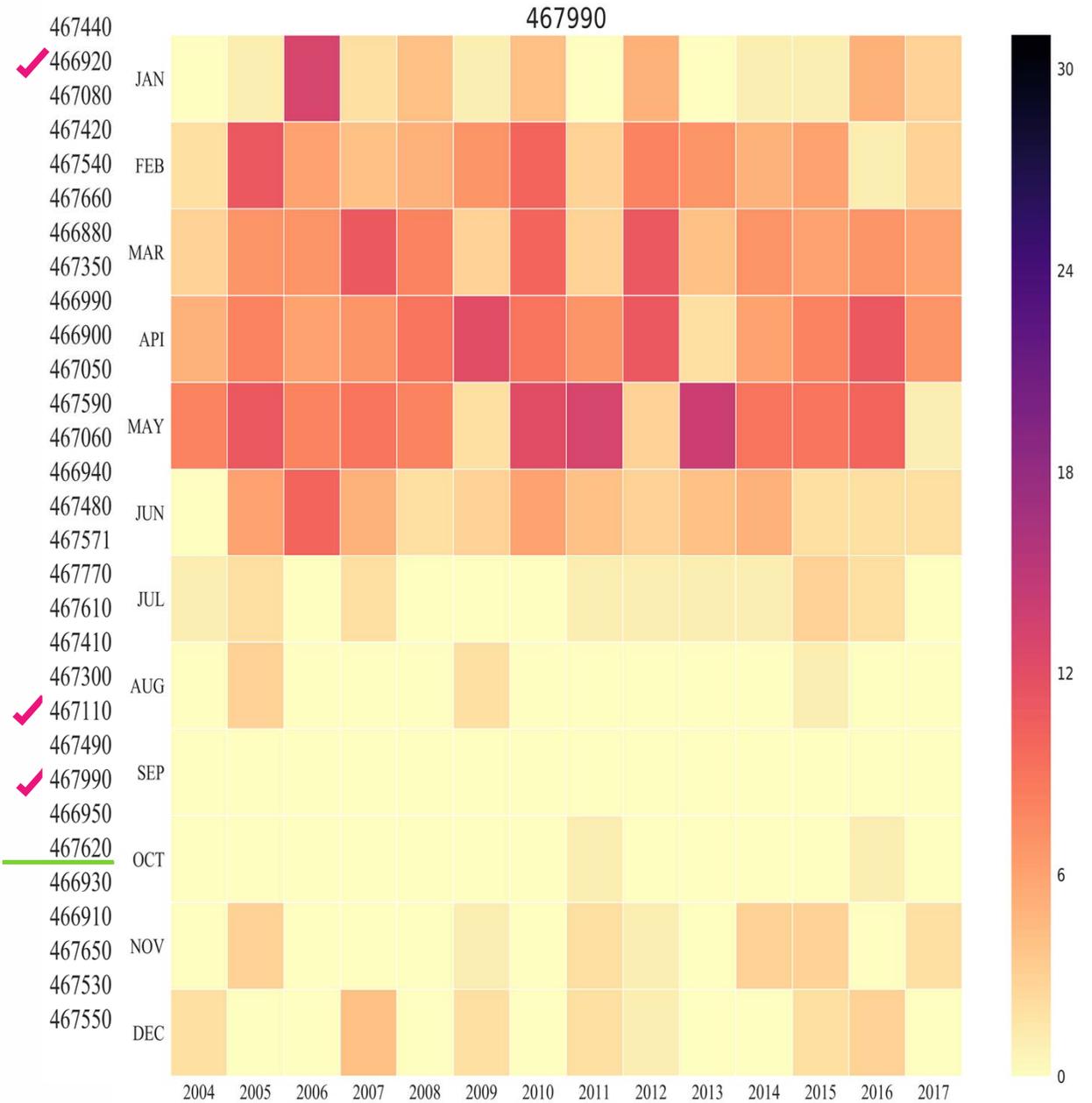
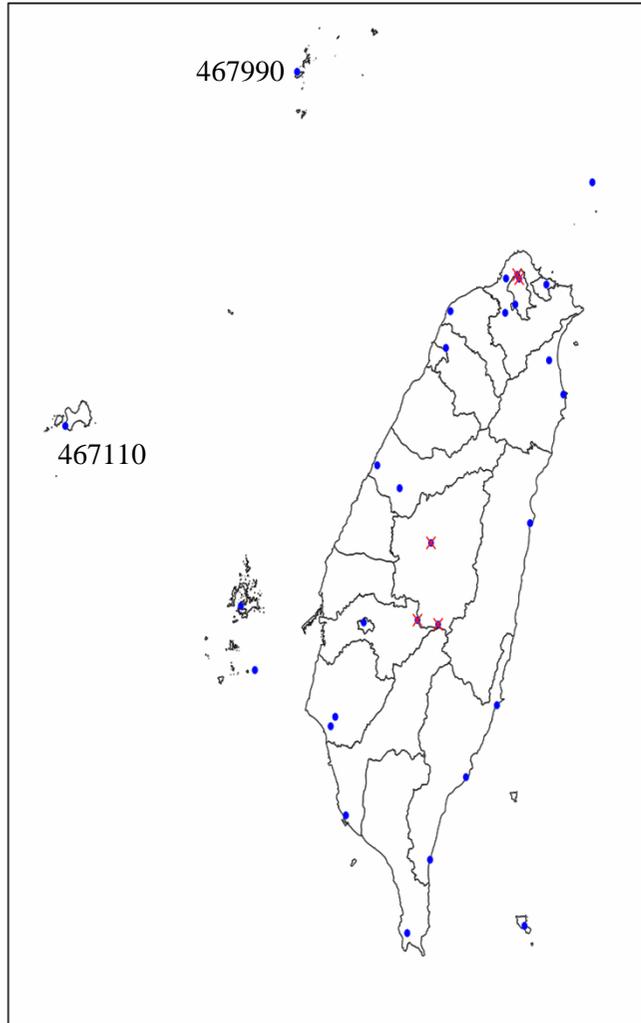
# 流程



有第 $t$ 時刻的資料就有一定能力診斷第 $t$ 時刻的天氣現象是否為霧

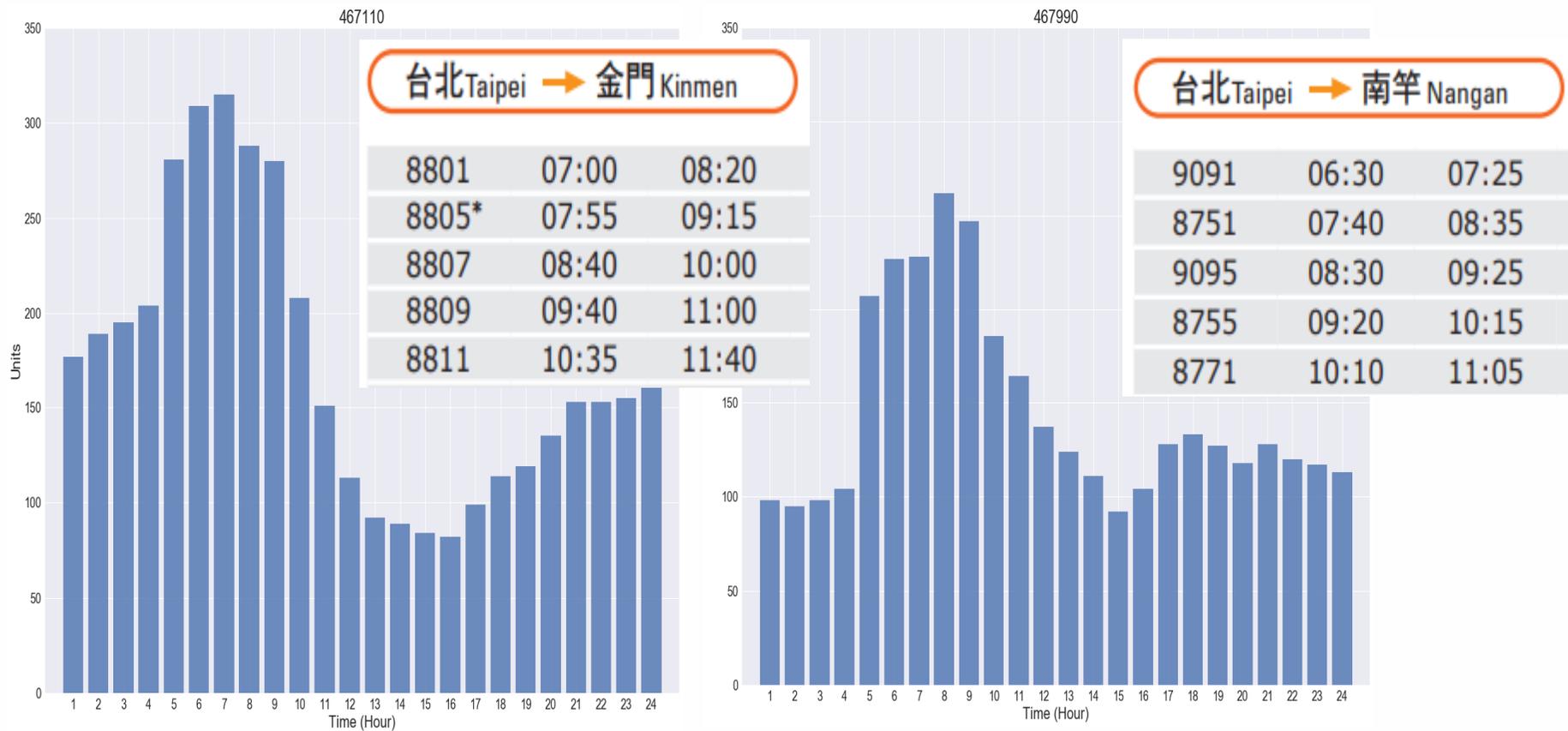
# 資料分析

## 針對氣象局局屬氣象站



# 資料分析

霧的種類很多樣化，會因不同的天氣形態，氣候條件不同甚至於地形不同而形成不同種類的霧。台灣平地地區常見的為輻射霧及平流霧。



# 資料分析

## A組

測站	資料長度	資料比數	原始資料	資料處理
金門	2004~ 2016	22,734	測站氣壓、溫度、露點、 水氣壓、相對濕度、風速、	測站氣壓、溫度、露 點、 <b>溫度露點差值</b> 、
馬祖	2004~ 2016	75,601	風向、降雨量、天氣型態	水氣壓、相對濕度、 風速、風向、降雨量、 天氣型態

## B組

測站	資料長度	資料比數	原始資料	資料處理
金門 馬祖浮標	2011~ 2016	6,383	測站氣壓、溫度、露點、 水氣壓、相對濕度、風速、	測站氣壓、溫度、露 點、 <b>溫度露點差值</b> 、
馬祖 馬祖浮標	2011~ 2016	7,354	風向、降雨量、天氣型態、 海溫、浮標風速、浮標風 向	水氣壓、相對濕度、 風速、風向、降雨量、 天氣型態、海溫、浮 標風速、浮標風向、 <b>溫度海溫差</b>

## 結果與討論

### ◆ 演算法

支撐向量機(Support vector machine, SVM)

隨機森林(Random forest, RF)

極限梯度提升(Extreme gradient boosting, XGBoost)

### ◆ 校驗準則

		觀測	
		霧	非霧
診斷	霧	a	b
	非霧	c	d

$$\text{偵測率}(POD) = \frac{a}{a + c}$$

$$\text{誤報率}(FAR) = \frac{b}{a + b}$$

# 結果與討論

A組

氣象站	演算法	校驗			
		偵測率		誤報率	
		訓練	驗證	訓練	驗證
金門	支撐向量積	0.74	0.52	0.14	0.53
	隨機森林	0.92	0.51	0.24	0.52
	極限梯度提升	0.53	0.43	0.32	0.53

B組

金門	支撐向量積	0.70	0.53	0.16	0.5
	隨機森林	0.87	0.6	0.28	0.47
	極限梯度提升	0.53	0.38	0.23	0.45

# 結果與討論

A組

氣象站	演算法	校驗			
		偵測率		誤報率	
		訓練	驗證	訓練	驗證
馬祖	升				
	支撐向量積	0.57	0.31	0.13	0.67
	隨機森林	0.74	0.34	0.35	0.75
馬祖	升				
	支撐向量積	0.7	0.5	0.12	0.52
	隨機森林	0.82	0.52	0.25	0.55
	極限梯度提	0.28	0.3	0.37	0.42

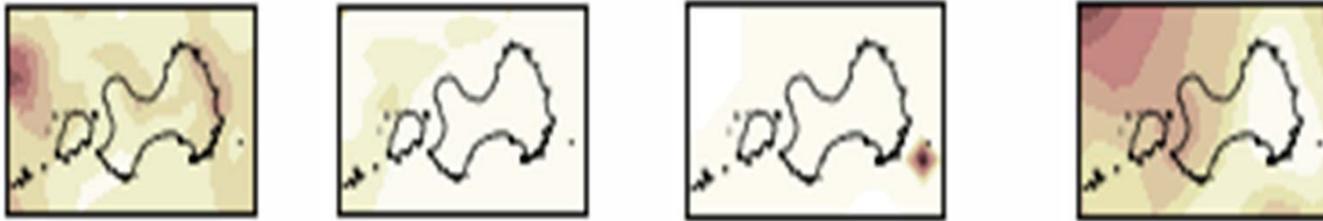
B組

## 結果與討論

- 使用單站且單一時刻的觀測資訊對於機器學習技術診斷非線性現象有其難度。
- 依照霧的分類而選取測站資訊對於診斷結果有幫助，但礙於資料相對不足，信任度會較低。
- 建議可以先將資料經過方程式(能見度推估方程式)的計算後，再作為機器學習之輸入(增加有效的影響因子)，進而發展診斷模式。

## 未來工作

- 利用卷積類神經網路技術搭配數值模式分析場(00, 06, 12, 18 UTC)建立霧之診斷模型。



共246筆  
186訓練、60測試

	偵測率	誤報率
訓練	0.56	0.42
測試	0.29	0.57

謝謝您的聆聽

