農業智慧化之後的挑戰 – 數據分析

台大農藝系 劉力瑜

lyliu@ntu.edu.tw

Google: 利用先進科技 (包含大數據、雲、物聯網) 追蹤、監測、自動化及分析於農產業的管理。

2022/06/16

Google

what is "smart agriculture"? 何謂智慧農業

全部 新聞 圖片 影片 地圖 更多 工具

約有 6,950,000 項結果 (搜尋時間：0.55 秒)

Smart farming is a management concept focused on providing the agricultural industry with the infrastructure to leverage advanced technology – including big data, the cloud and the internet of things (IoT) – for tracking, monitoring, automating and analyzing operations.

https://www.techtarget.com › iotagenda › smart-farming

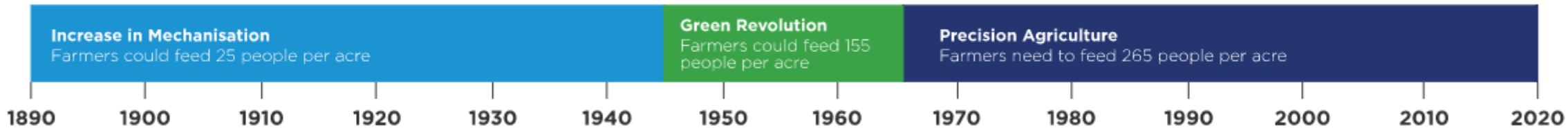
What is smart farming? Definition from WhatIs.com - TechTarget

https://smartertechnologies.com/the-complete-guide-to-smart-farming-agriculture/

所以智慧農業要在下一世代
餵養 1000+ people per acre?

Timeline of agricultural revolutions

Since humans have been cultivating food and rearing livestock, technological innovations have been improving agricultural outputs.



https://smartertechnologies.com/the-complete-guide-to-smart-farming-agriculture/

WHAT IS SMART AGRICULTURE AND SMART FARMING?

Also known as Farming 4.0 or digital farming, smart farming is the application of information and data technologies to optimise complex farming systems. It involves individual machines and all farm operations.

How does digital farming work?

Smart farming incorporates information and communication technologies into machinery, equipment and sensors used in agricultural production systems. Technologies such as the IoT and cloud computing are advancing this development even further by introducing more robots and artificial intelligence into farming.

For example, farmers can use smartphones and tablets to access real-time data about the condition of almost anything involved in their day-to-day operations:

Soil

Plants

Terrain

Weather

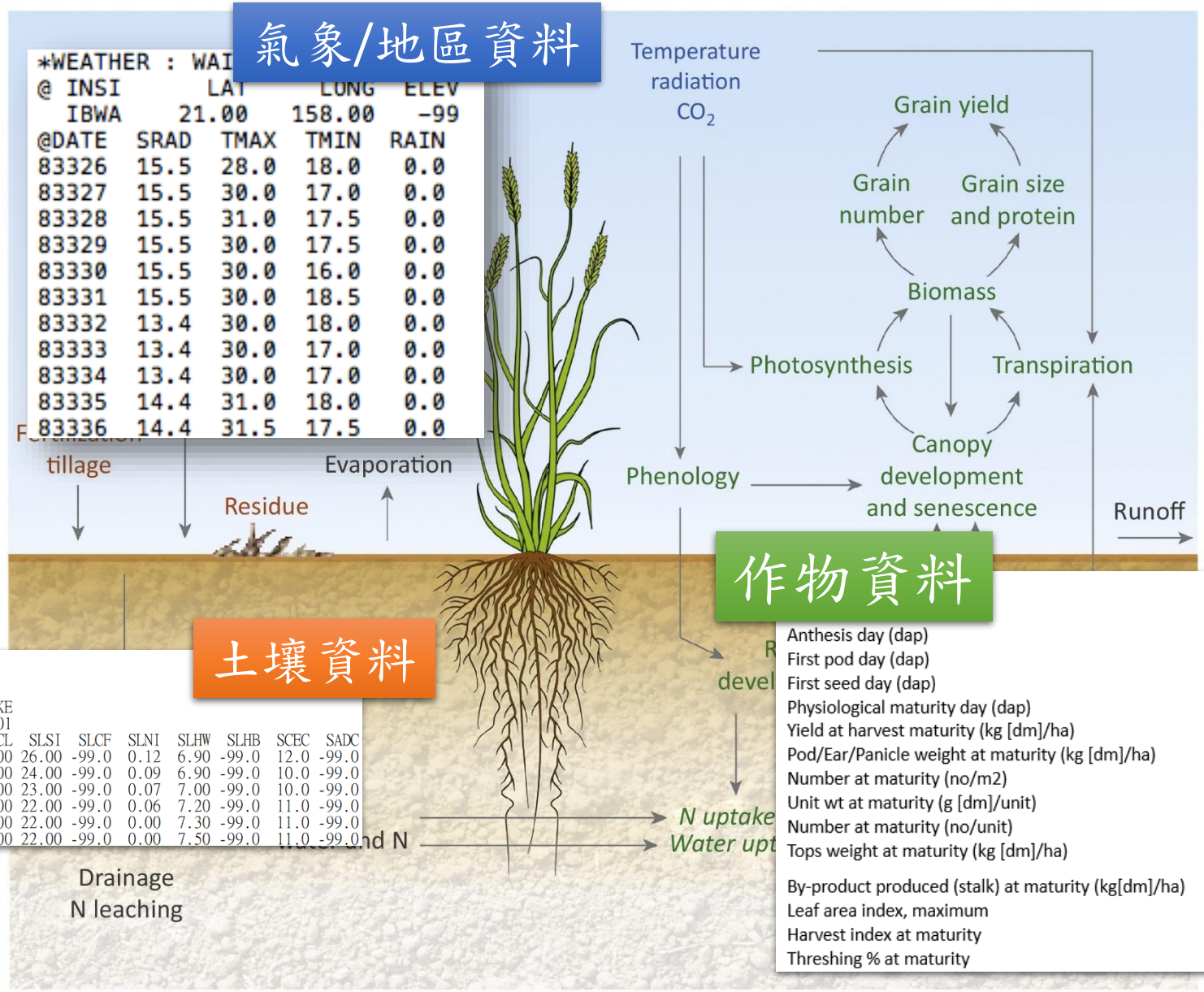
Location of assets

Condition of assets

Smart agricultural practices allow for the generation of a large volume of data and information. Farmers can use this information to make data-based decisions and take action for improved productivity and profitability.

“Make data-based decision and take action”

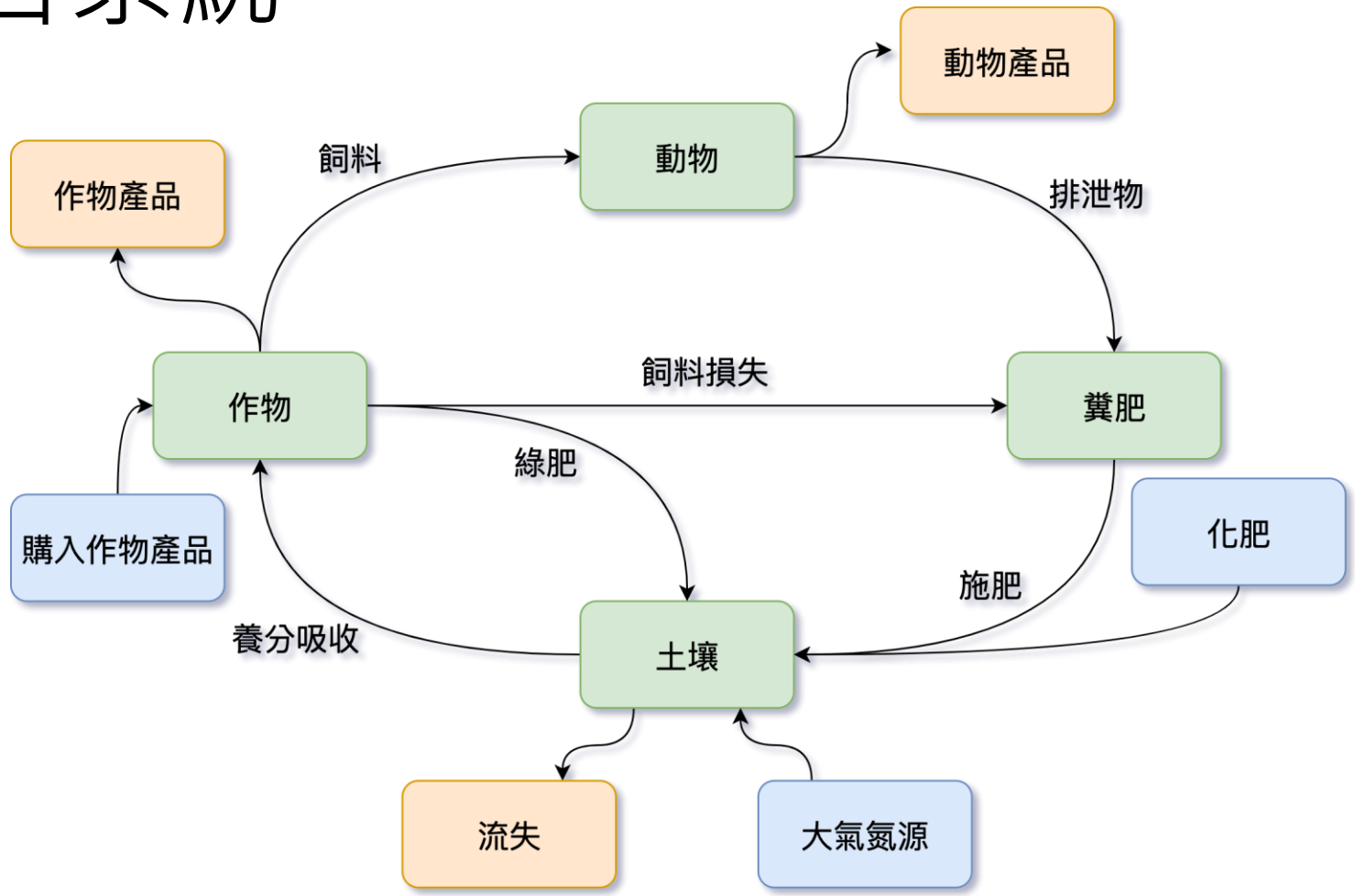
數據/資訊哪裡來?



Trends in Plant Science

Trends Plant Sci. 2017. 22:472-490.

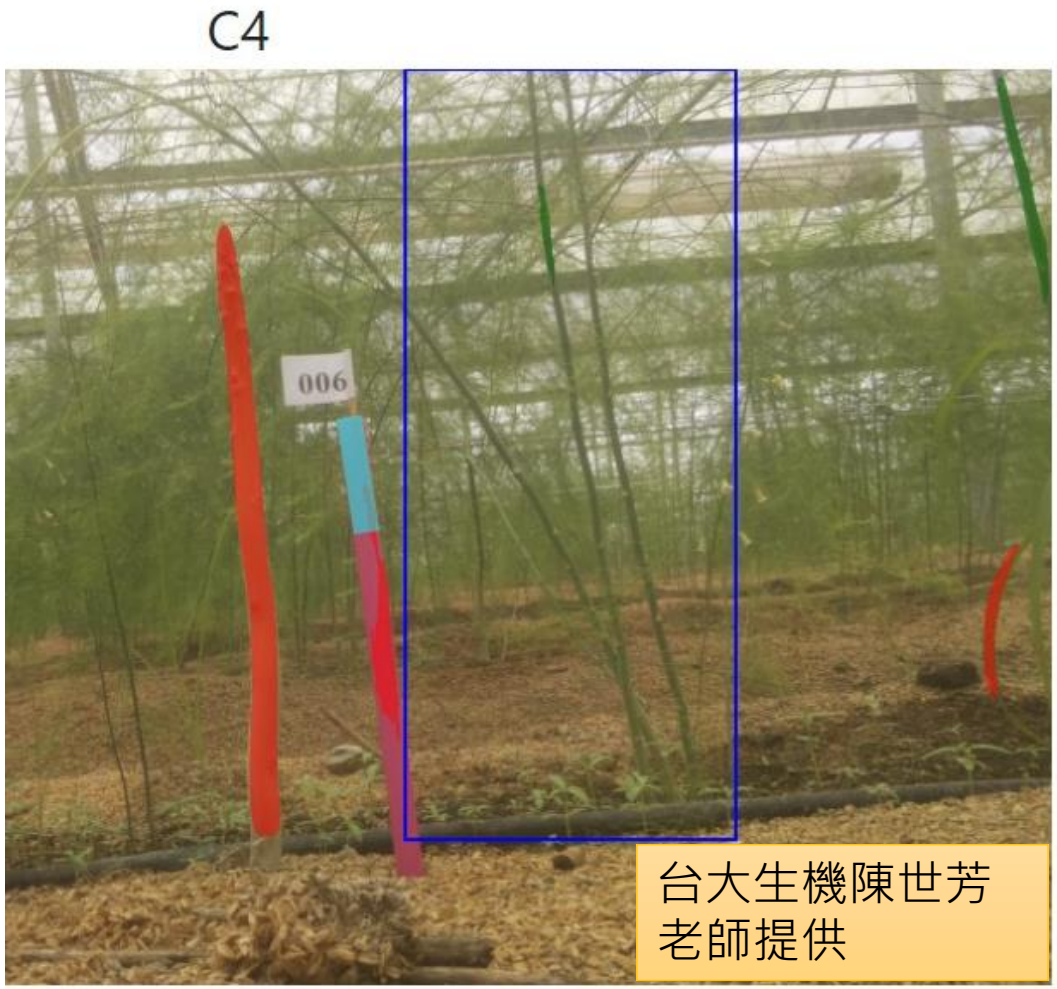
資料來源:
農牧整合系統



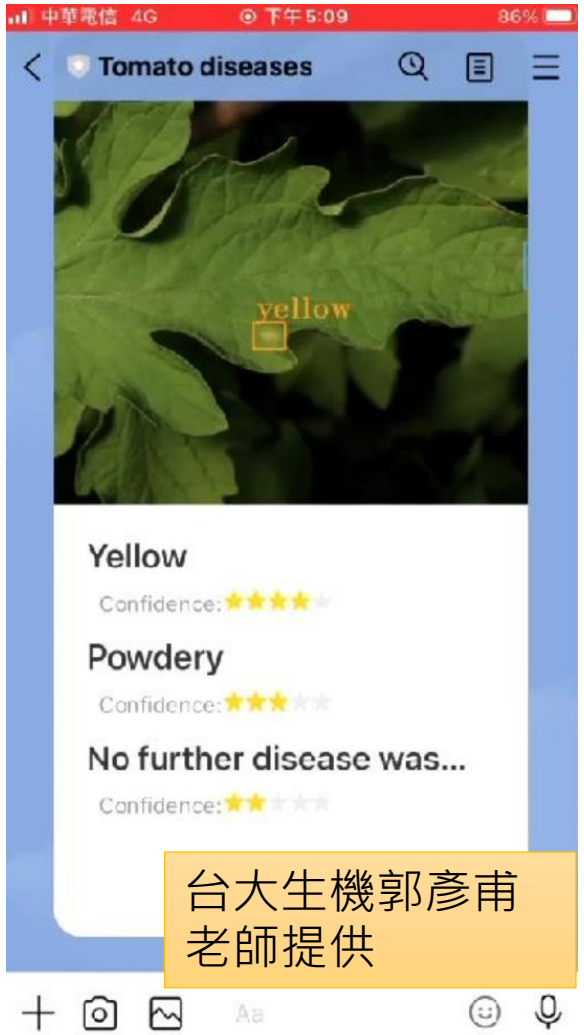
Groot et al. (2012) Agricultural Systems 110: 63–77 (根據 Fig. 1 重製)

田間大數據: 影像

台大生機陳世芳
老師提供



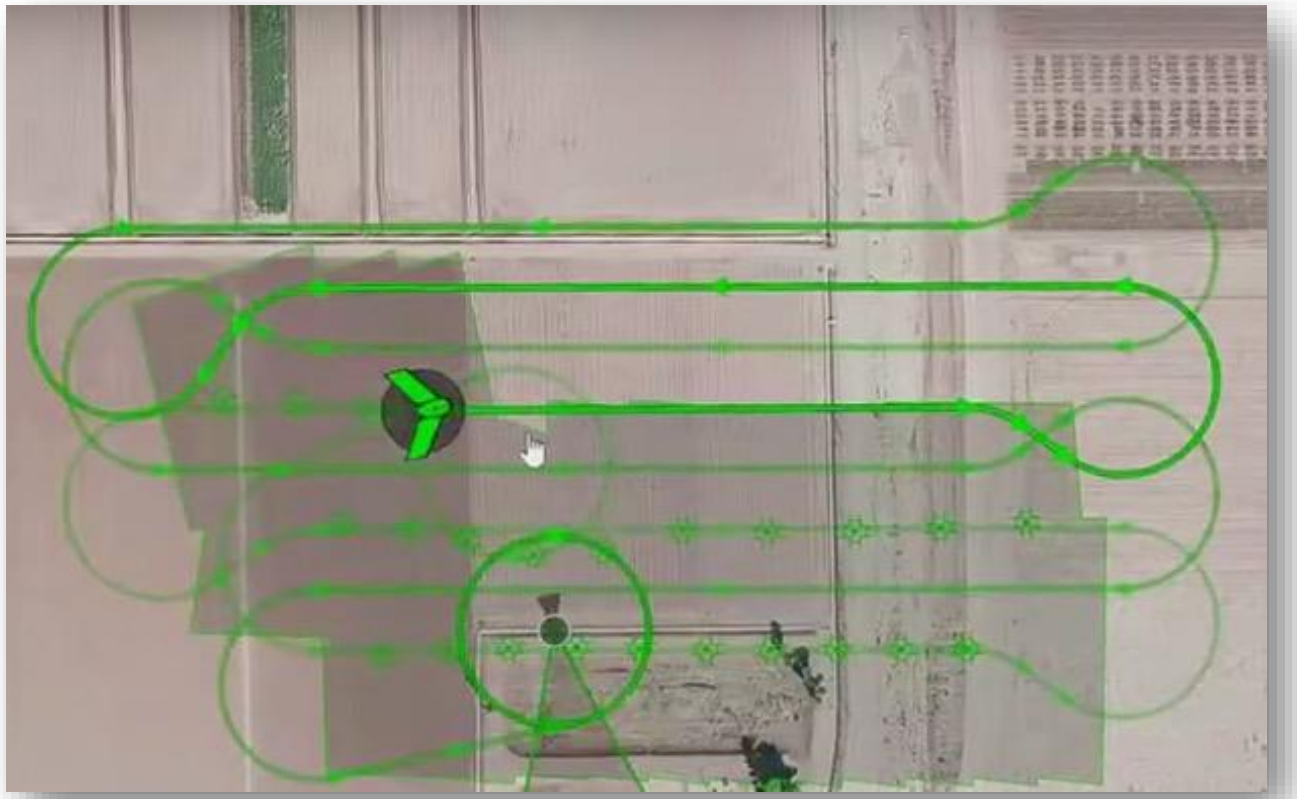
台大生機陳世芳
老師提供



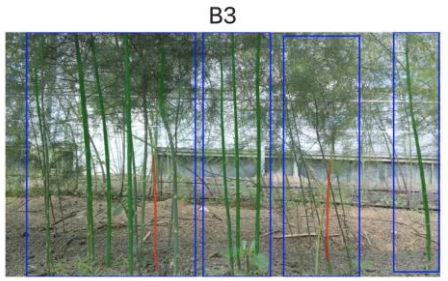
台大生機郭彥甫
老師提供

田間大數據: 影像

Images from Unmanned Aerial Vehicle



數據整合



spear

Score: 100.00 %

Length: 311.33 mm

Length in 24 hrs: 315.85 mm

Width: 19.96 mm

Area: 8351.68 mm^2

現在（ 出土10小時）

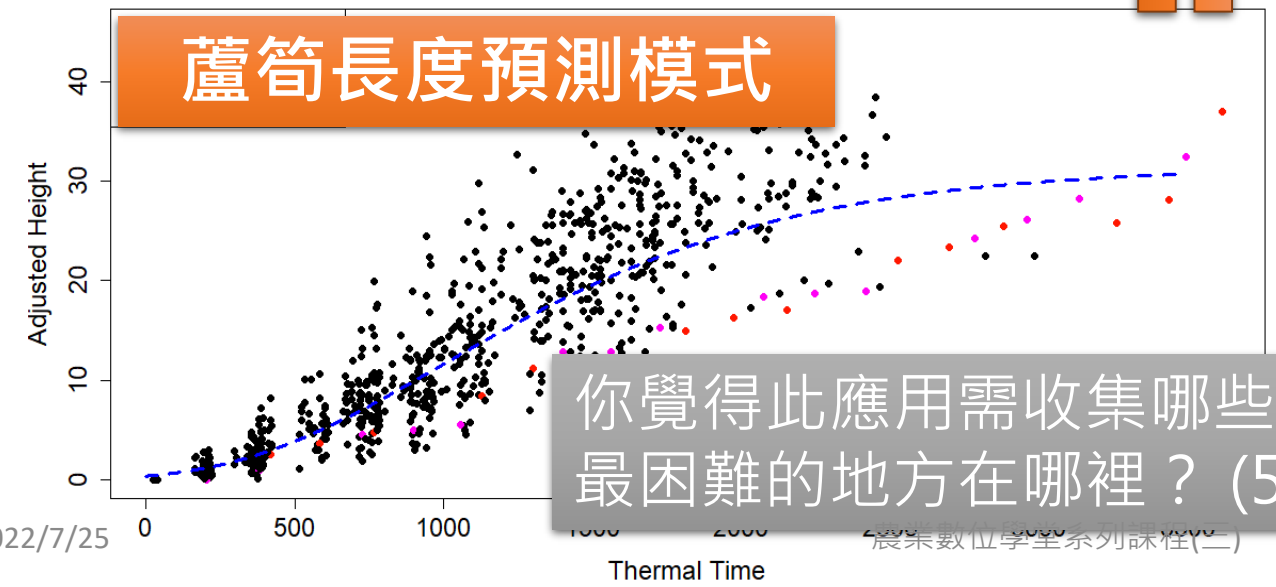
5公分

預測（ 24小時後）

10公分

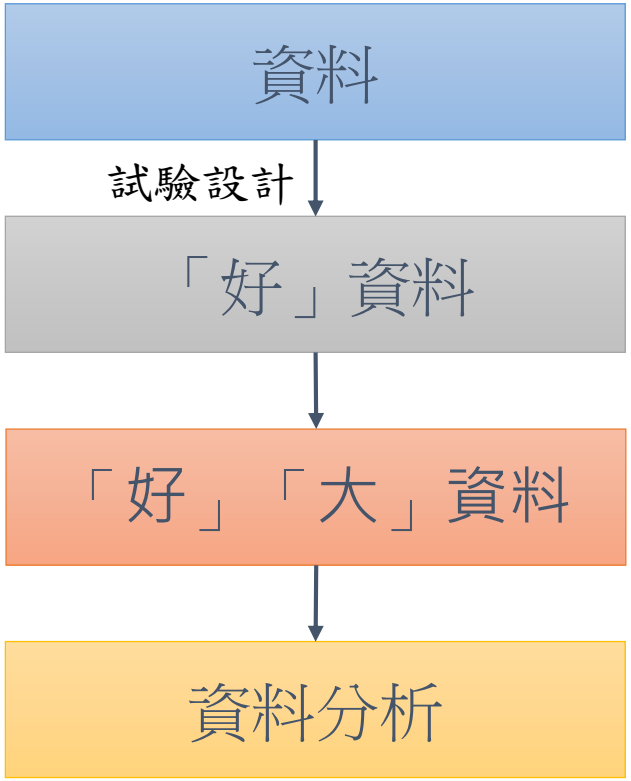
預測三天後早晨 可採收

蘆筍長度預測模式



你覺得此應用需收集哪些資料？
最困難的地方在哪裡？ (5 + 5 min)

從資料收集出發



蘆筍生長？

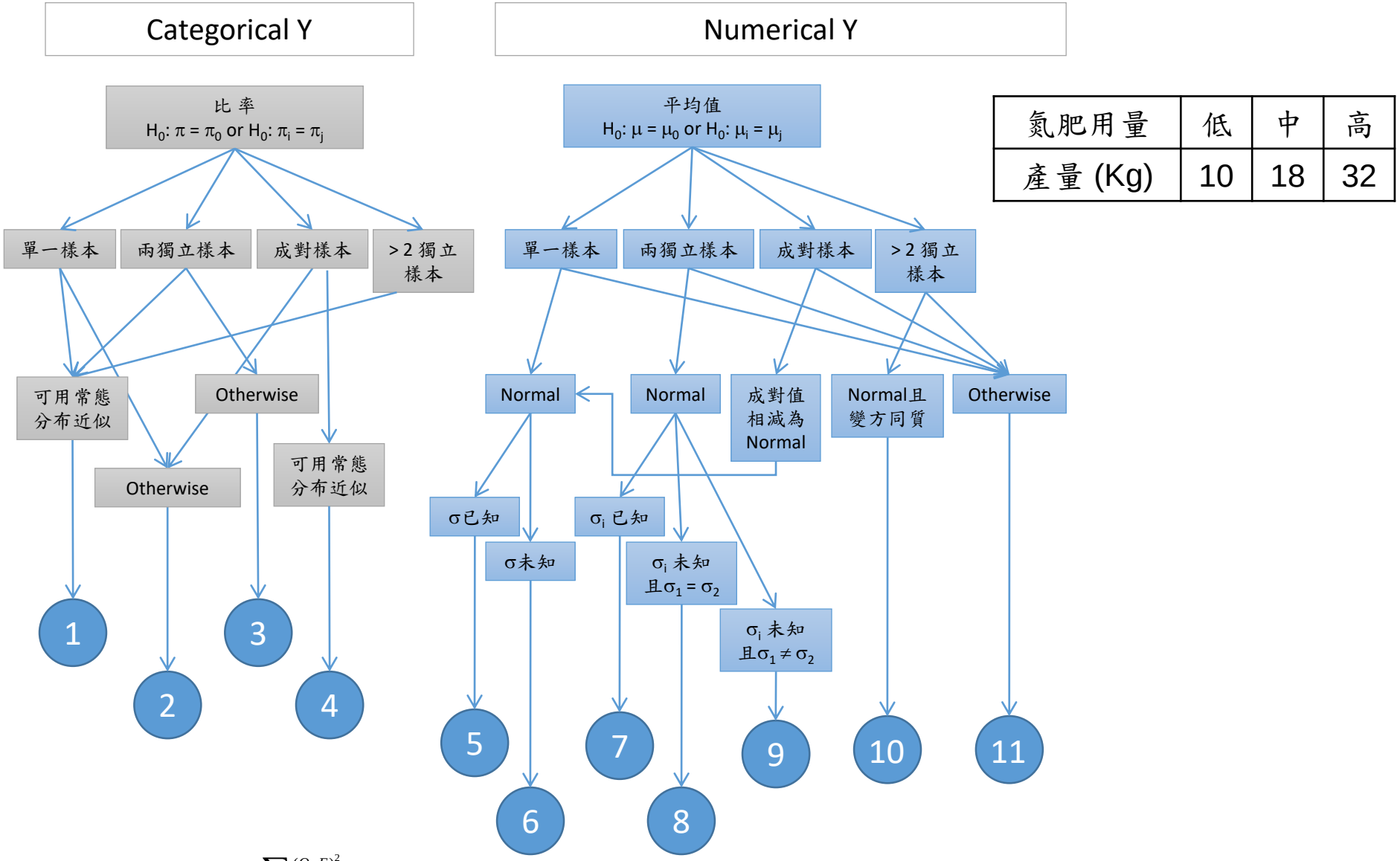
氮肥用量與產量：

氮肥 (Kg) (X)	0	0.5	1.0	1.5	2.0	2.5
產量 (Kg) (Y)	10	18	32	48	55	62

氮肥用量 (X)	低	中	高
產量 (Kg) (Y)	10	18	32

能統計分析的資料，才是好資料!
(統計：排除人為因素、抽樣誤差等干擾)

Y (response)	X (predictor)	
	類別變數	連續變數
類別變數	列聯表 / 比率資料分析	廣義線性模式 (logistic regression, etc.)
連續變數	平均數檢定	線性迴歸分析

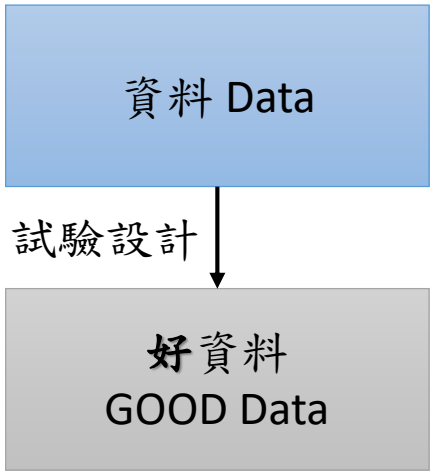


氮肥用量	低	中	高
產量 (Kg)	10	18	32

1: **chi-squared test** $\sum \frac{(O-E)^2}{E}$
2: Binomial test
3: Fisher's exact test
4: McNemar test

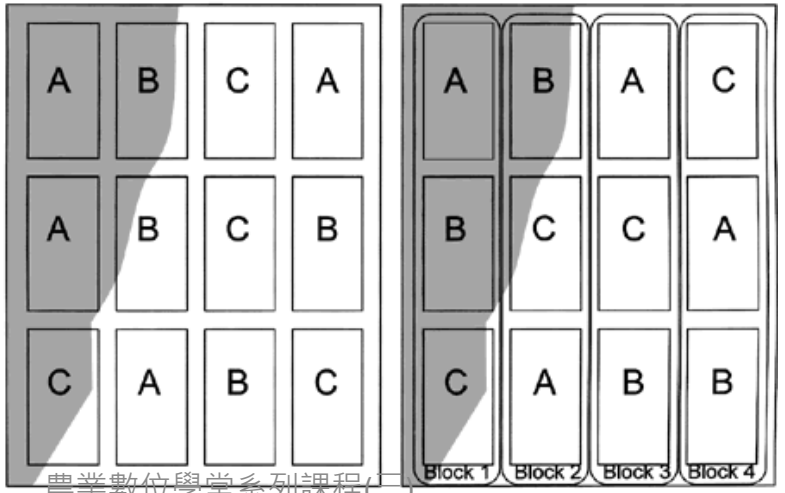
5: one-sample z test
6: one-sample t test
7: two-sample z test
8: two-sample t-test
9: Welch two-sample t-test

10: 變方分析 (ANOVA)
11: 變數轉換或無母數法



一個好的研究必須要有嚴謹的設計，客觀的試驗過程及合理的推論。因此試驗時必須遵守下列三個原則:

- 設置重複: 每個處理組合需獨立施於兩個以上的試驗單位。
- 隨機排列: 隨機分配接受各處理之試驗單位控制未知來源干擾。
- 設置區集: 控制已知來源的干擾。



資料分析

FWT Penning de Vries, DM Jansen, HFM ten Berge, A Bakema (1989)
Simulation of ecophysiological processes of growth in several annual crops.

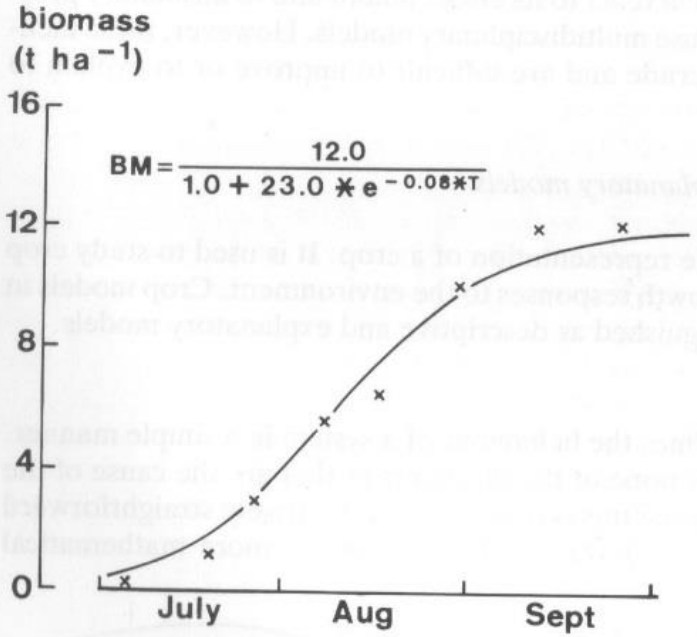


Figure 2. The course of the dry weight of a maize crop in the Netherlands in 1972. Crosses represent observations, the line the regression equation $BM = 12.0 / (1.0 + 23.0 \cdot e^{-0.08 \cdot T})$, where BM is the biomass in t ha^{-1}, 12.0 is the maximum value of BM, T is the time in days since emergence and 1.0, 23.0 and 0.08 are constants.

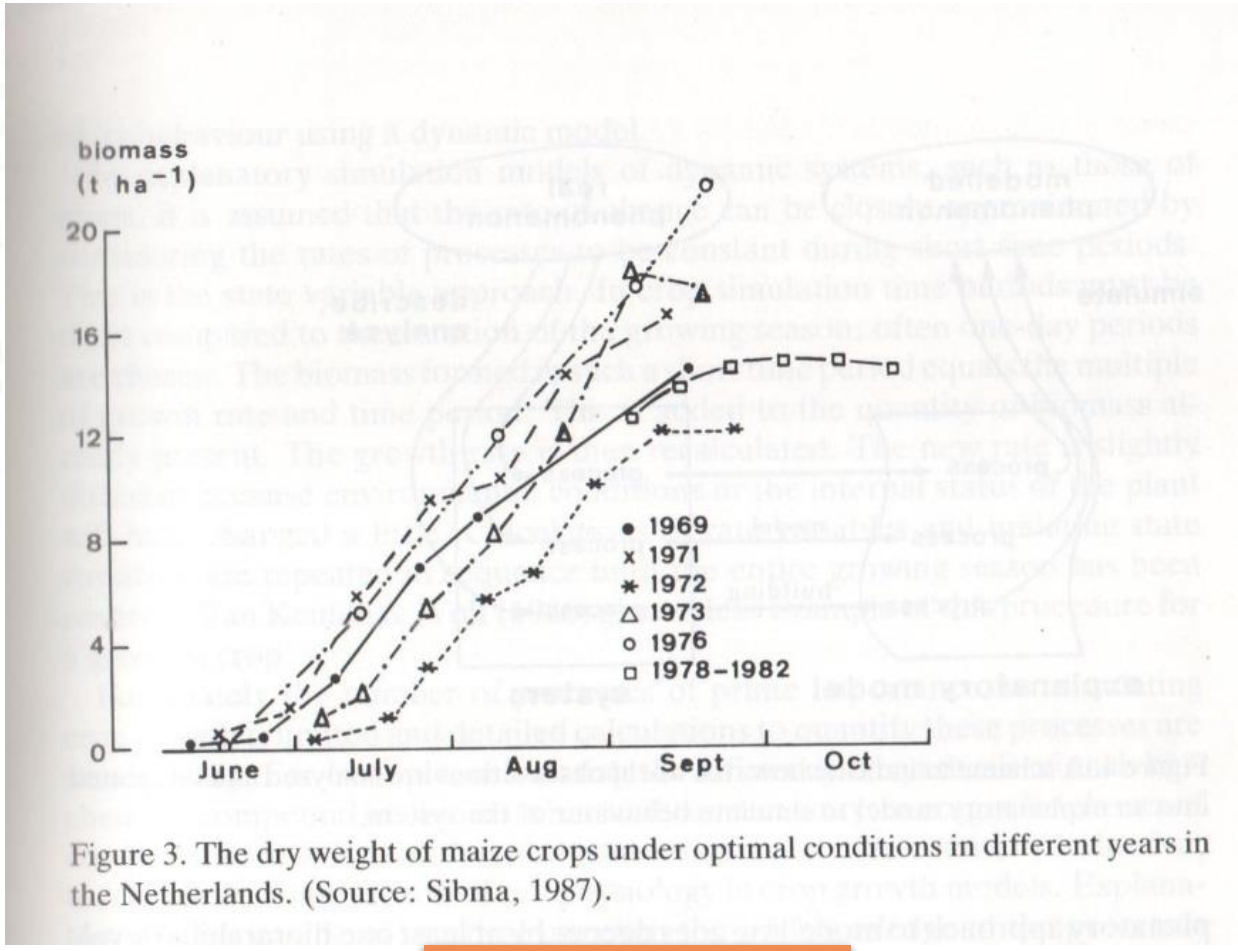
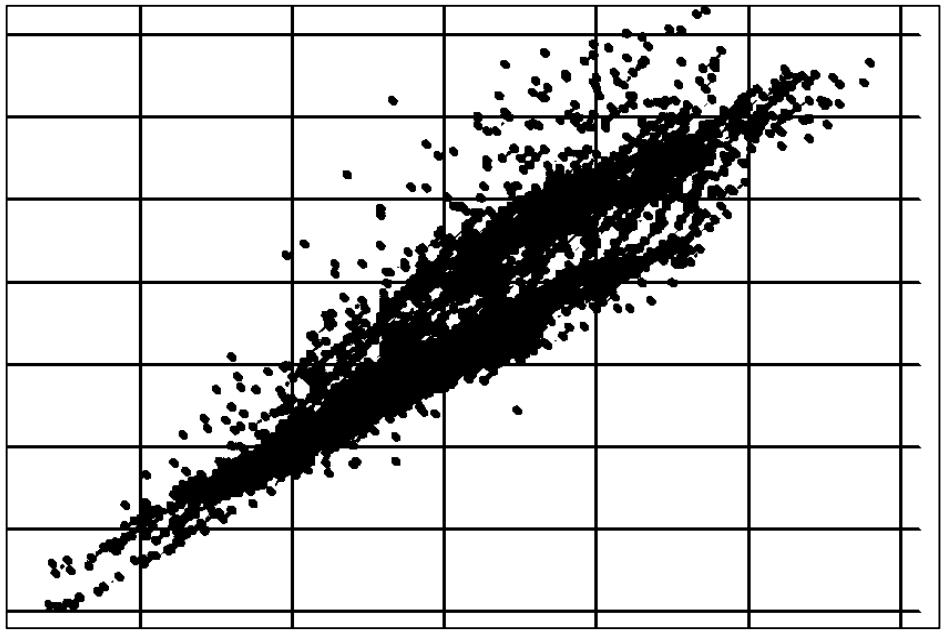


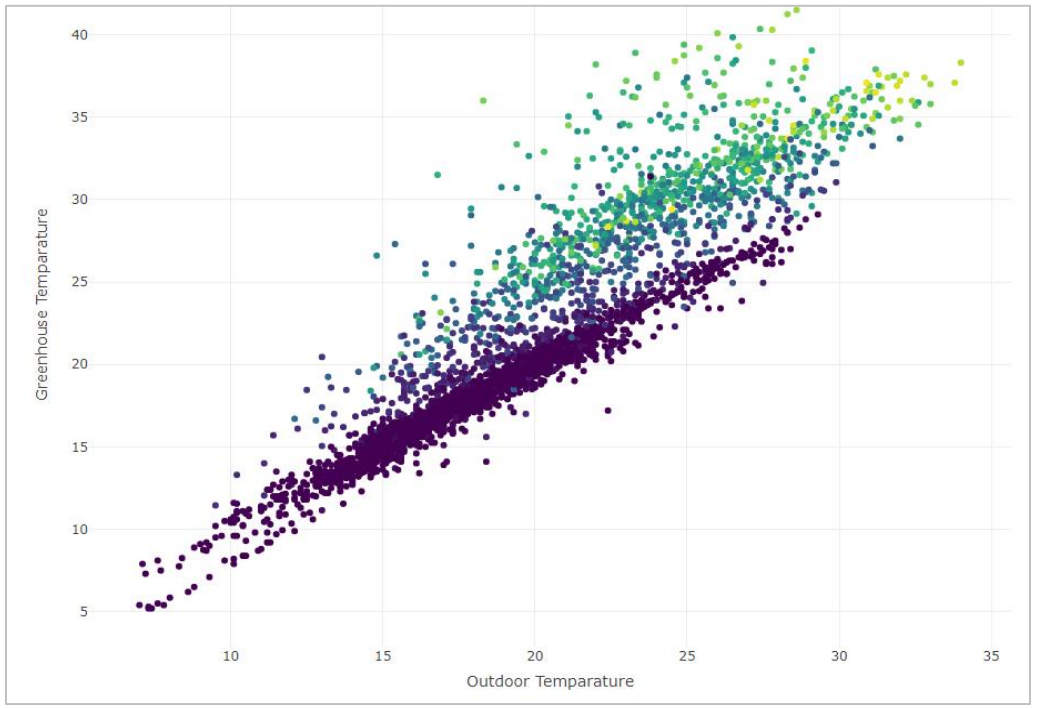
Figure 3. The dry weight of maize crops under optimal conditions in different years in the Netherlands. (Source: Sibma, 1987).

你以為的資料分析

實際上的資料分析

資料分析 – 讓資料說話




Greenhouse Temparature
Outdoor Temparature
40
35
30
25
20
15
10
5
10
15
20
25
30
35

Statistics → AI; Machine Learning → AI

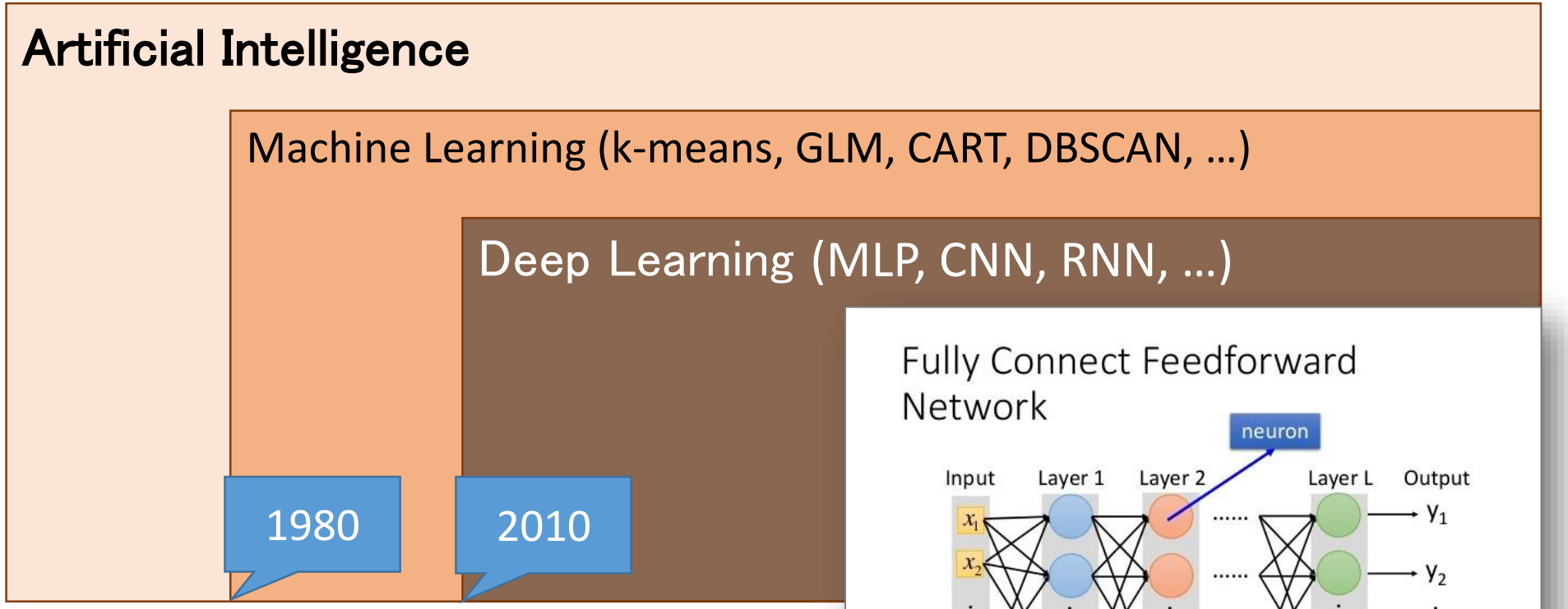
"隨著農業科技的發展，各種相關應用也開始整合人工智慧(AI)，智慧農業(smart farming)技術將善用各種資料促進供應鏈透明度(transparency)，運用先進分析與預測性推論(predictive inference)協助農人進行更好的農務決策，以提升生產力與獲利。"

https://www.digitimes.com.tw/iot/article.asp?cat=158&id=0000555478_DXP5ALNO6DR1WO9YBVP9E

- 統計分析：通常會想要利用數學模型去學習資料，找出一組參數來「描述」資料，目標是找出資料背後的分佈，以解釋資料間的關係。
- 機器學習：透過抽象模型學習擬合資料，會著重在學習模型的最佳化過程，目標是達到最好「預測」效果。

曾婉菁 (2018)。機器學習探究。印刷科技 34:1-32

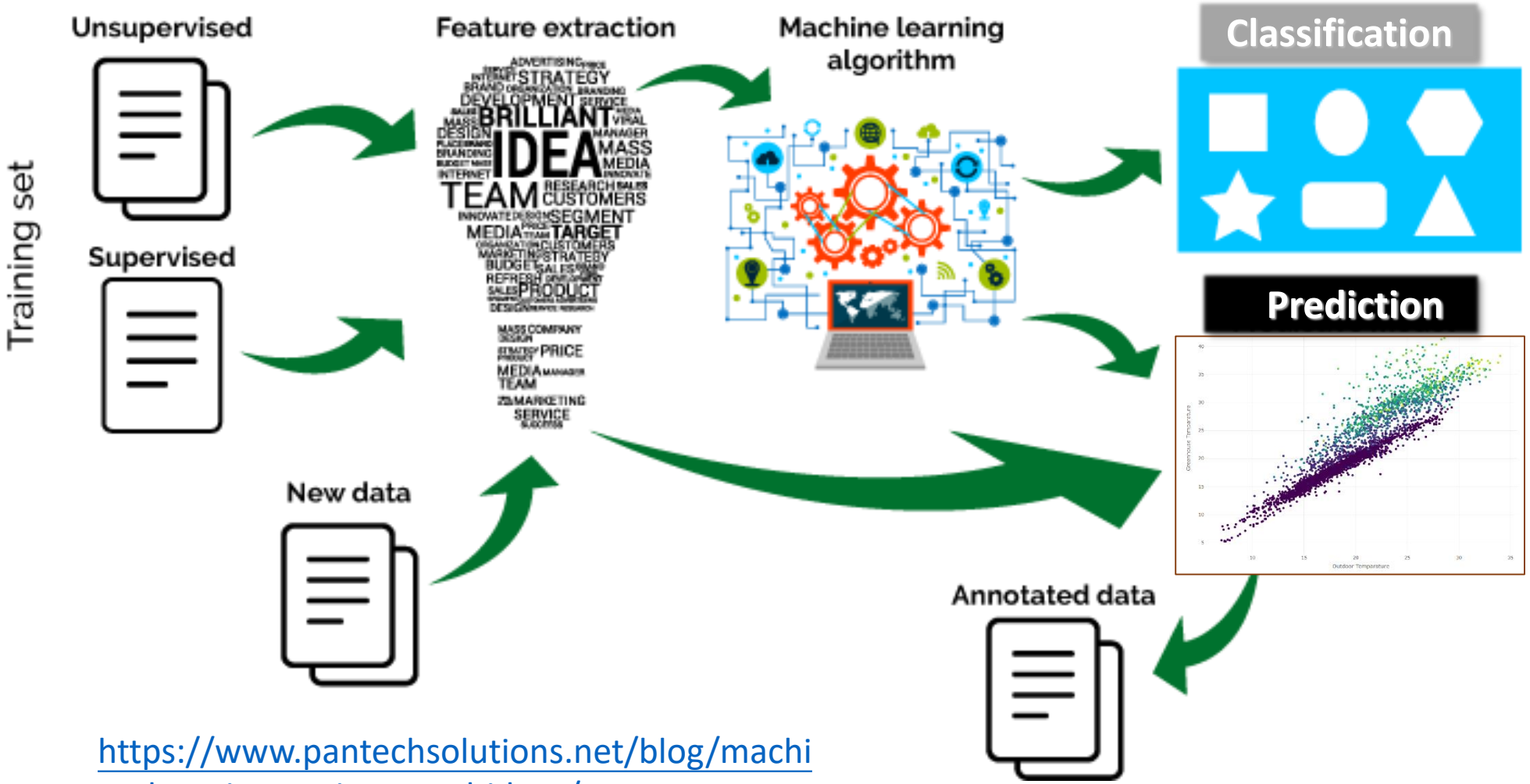
AI、機器學習、深度學習



1950: Turing Test

https://panx.asia/archives/53209

Machine Learning



https://www.pantechsolutions.net/blog/machine-learning-projects-and-ideas/

AI的農業應用

- 人工智慧在農業上的應用卻相當受限
  - 作物生長周期長
  - 資料不易大量累積
  - 資料標註高度仰賴具有專業知識的農業專家
  - 作為機器學習的訓練資料集嚴重不足

農業大數據實例應用

田間大數據: 影像 Drone Image Data

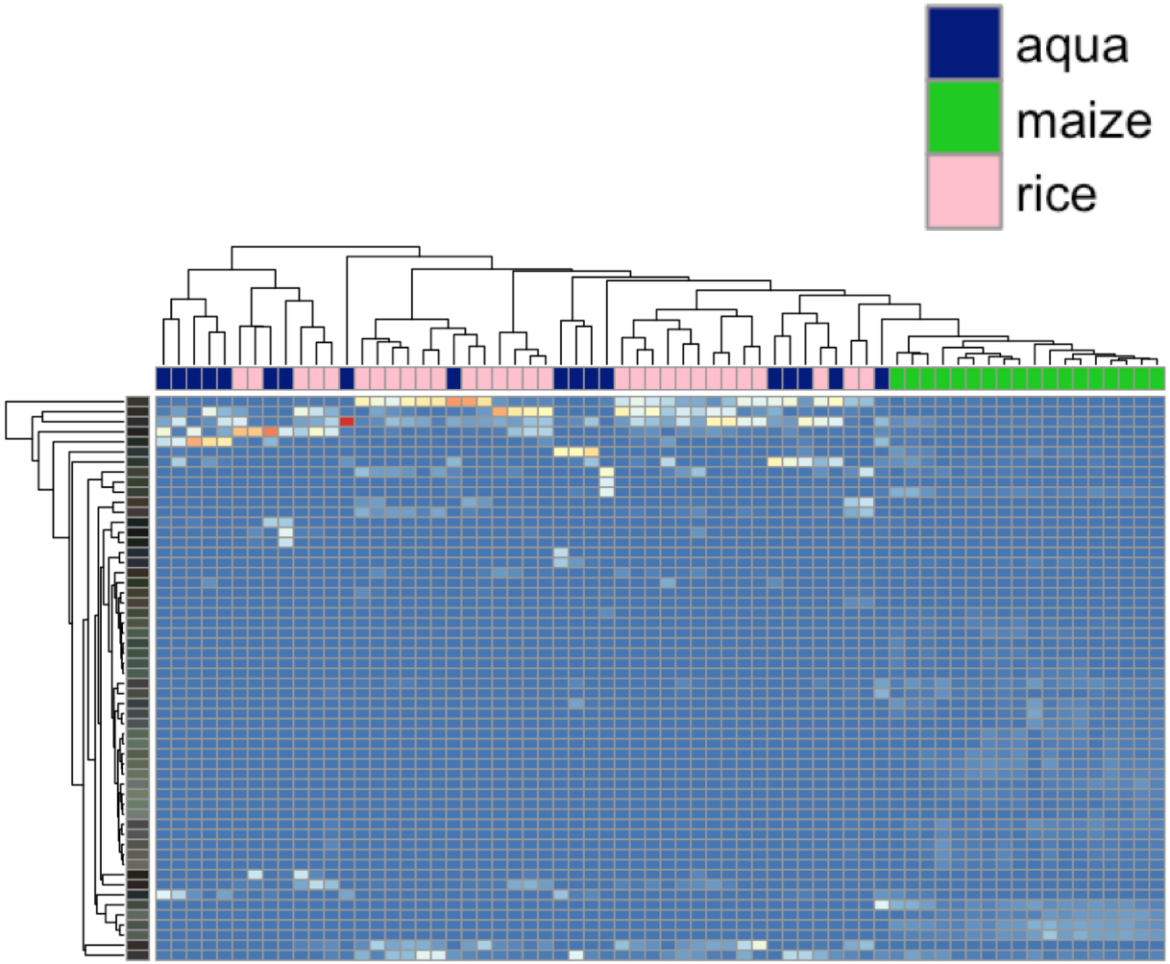
~ 1TB per route

玉米

甘蔗

Rice (2019 Dec)

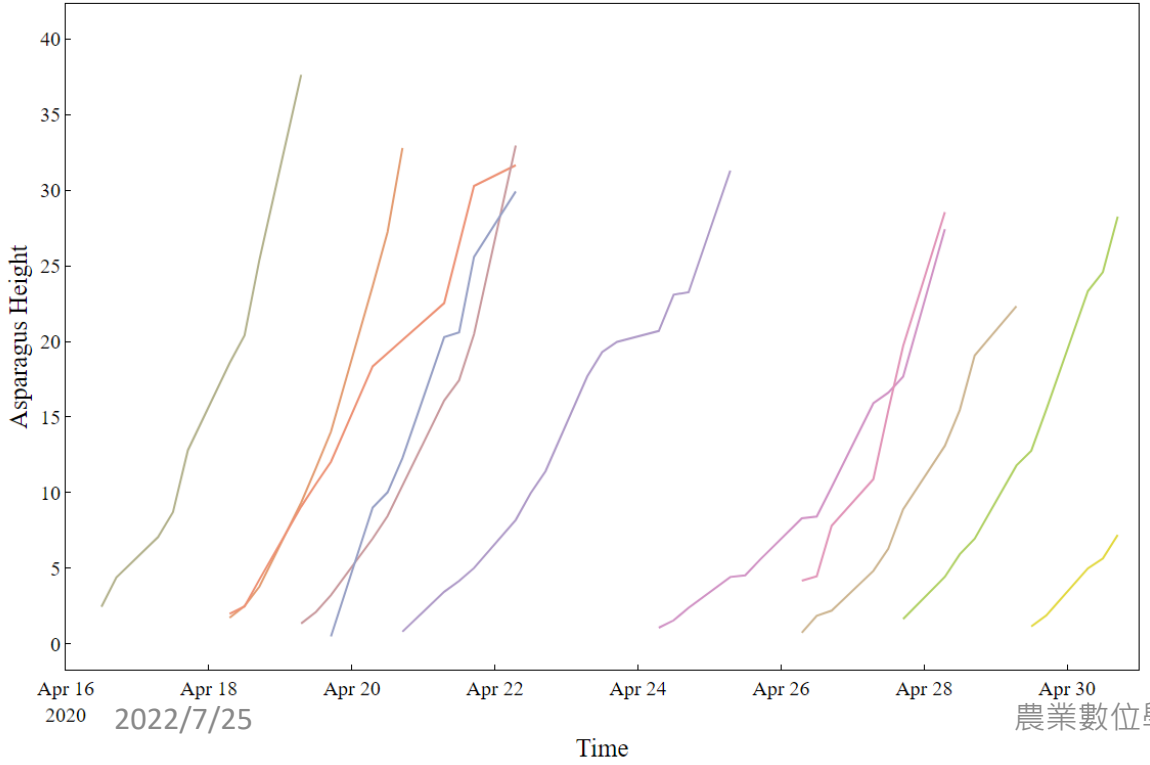
Rice (2020 March)



採收紀錄與IoT資料整合

影像資料(應變數)

- 每五分鐘回傳一張影像
- 紀錄**每日7:00、12:00、17:00**三個時間點
- IC Measure量測、邊長16公分四格板子
- 以**調整後高度**做紀錄



環境感測資料(自變數)

- 每十分鐘回傳一次感測資料
- 溫室內IoT感測 + 室外測站

環境變數	單位	義竹研究場域	將軍研究場域	義竹室外氣象測站
大氣溫度	°C	✓	✓	✓
大氣相對溼度	%	✓	✓	✓
土壤水分張力	kPa	✓	✓	
土壤溫度	°C	✓	✓	
土壤濕度	%	✓	✓	
土壤電導度	mS/cm	✓	✓	
日照量	W/m^2	✓		✓

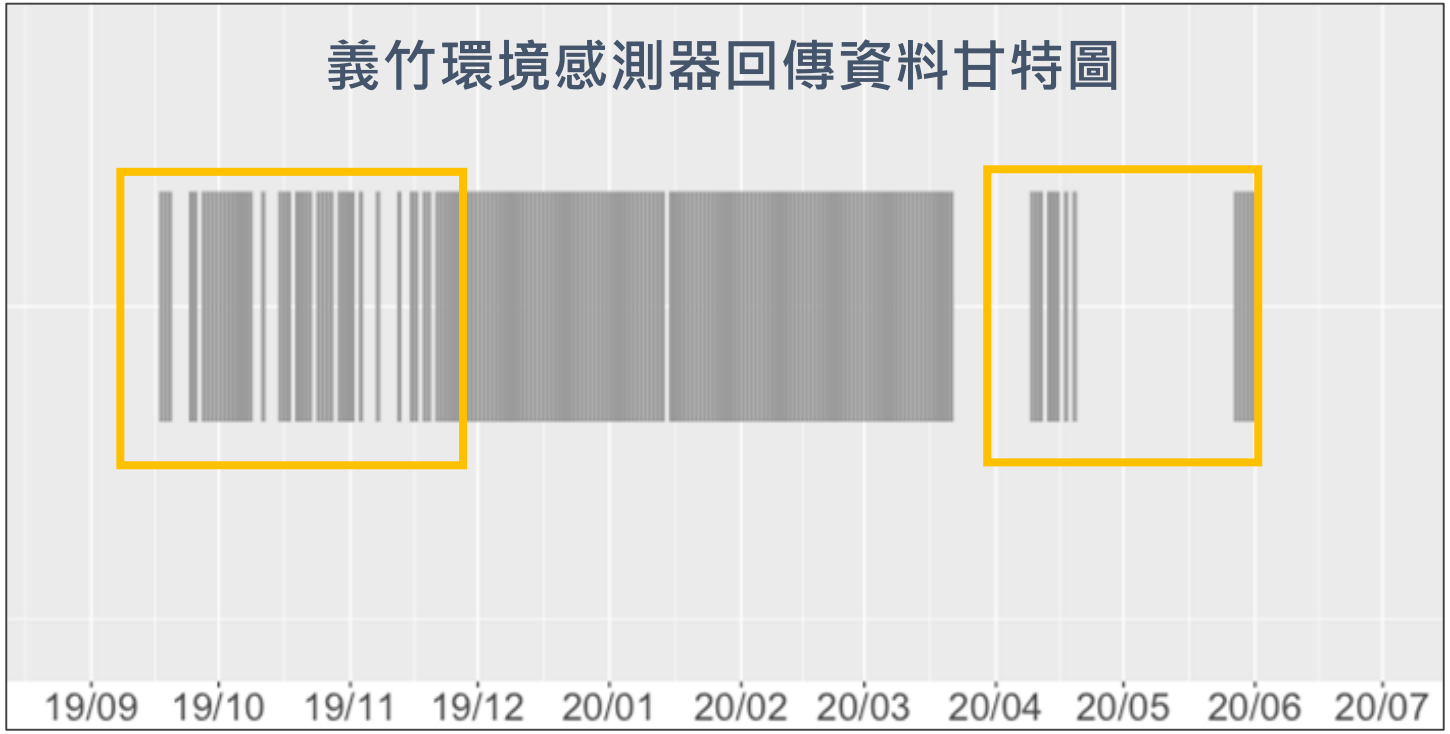
採收紀錄與IoT資料整合

卡爾曼濾波器 Kalman Filter

- 遞迴的估計方法
- 短時間中時間序列的估計

室外測站+多元線性迴歸模型

- 掌握室外/室內環境變數關聯
- 模擬溫室內溫度
- 適用於無IoT感測資料/資源之情況



採收紀錄與IoT資料整合

採收前 k 天日射量 (k = 1, …, 7)

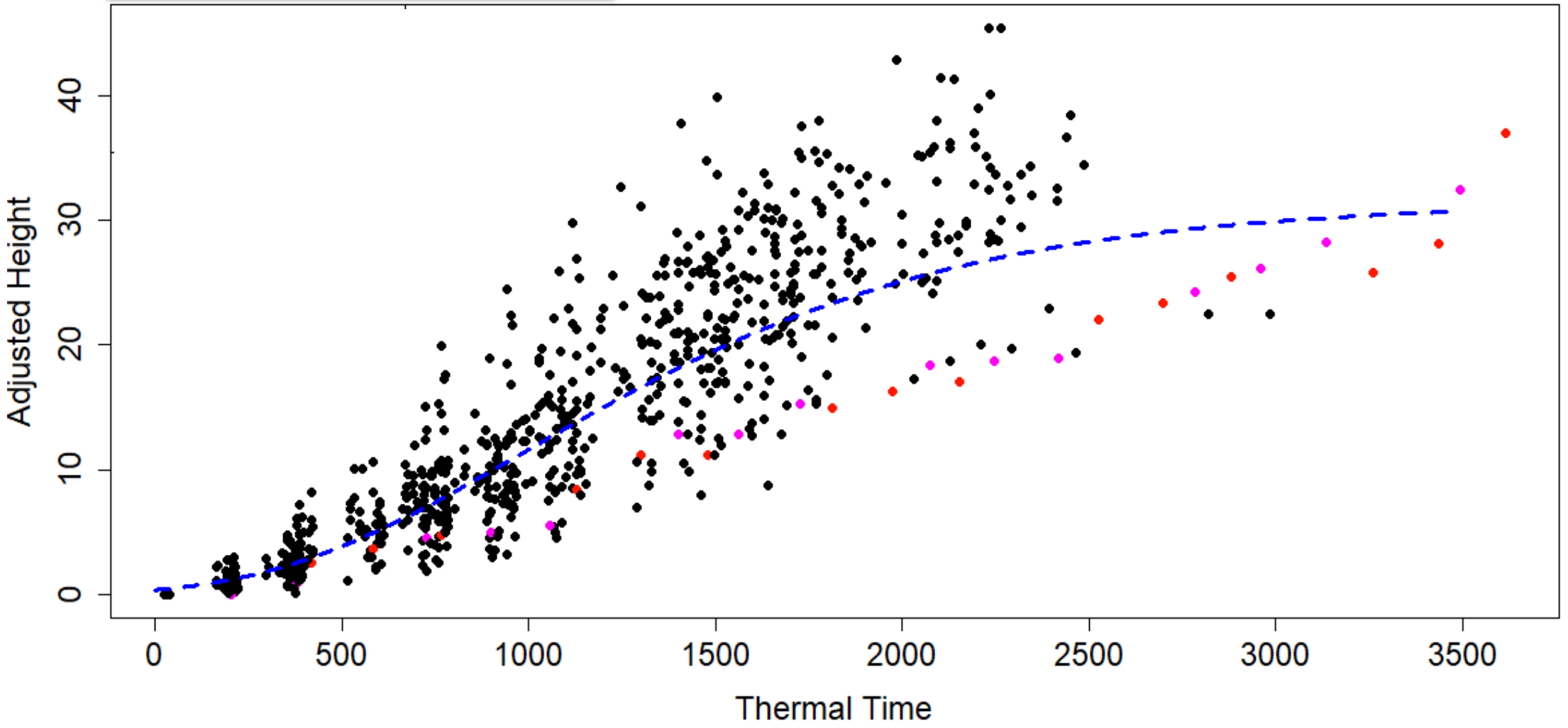
採收日期 採收重量

採收前 k 天溫度 (k = 1, …, 7)

採收前 k 天濕度 (k = 1, …, 7)

	Time	Weight	AirTemp1	AirTemp2	AirTemp3	AirTemp4	AirTemp5	AirTemp6	AirTemp7	AirHum1	AirHum2	AirHum3	AirHum4	AirHum5	AirHum6	AirHum7	Irrad
1	2021/11/25	8.6	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
2	2021/11/25	11.4	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
3	2021/11/25	15.3	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
4	2021/11/25	8.3	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
5	2021/11/25	9.1	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
6	2021/11/25	14.1	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
7	2021/11/25	4.9	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
8	2021/11/25	8.3	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
9	2021/11/25	6.4	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
10	2021/11/25	8.1	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
11	2021/11/26	0.4	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
12	2021/11/26	13.3	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
13	2021/11/26	5.0	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
14	2021/11/26	15.8	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
15	2021/11/26	13.0	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
16	2021/11/26	13.7	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
17	2021/11/26	18.9	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	

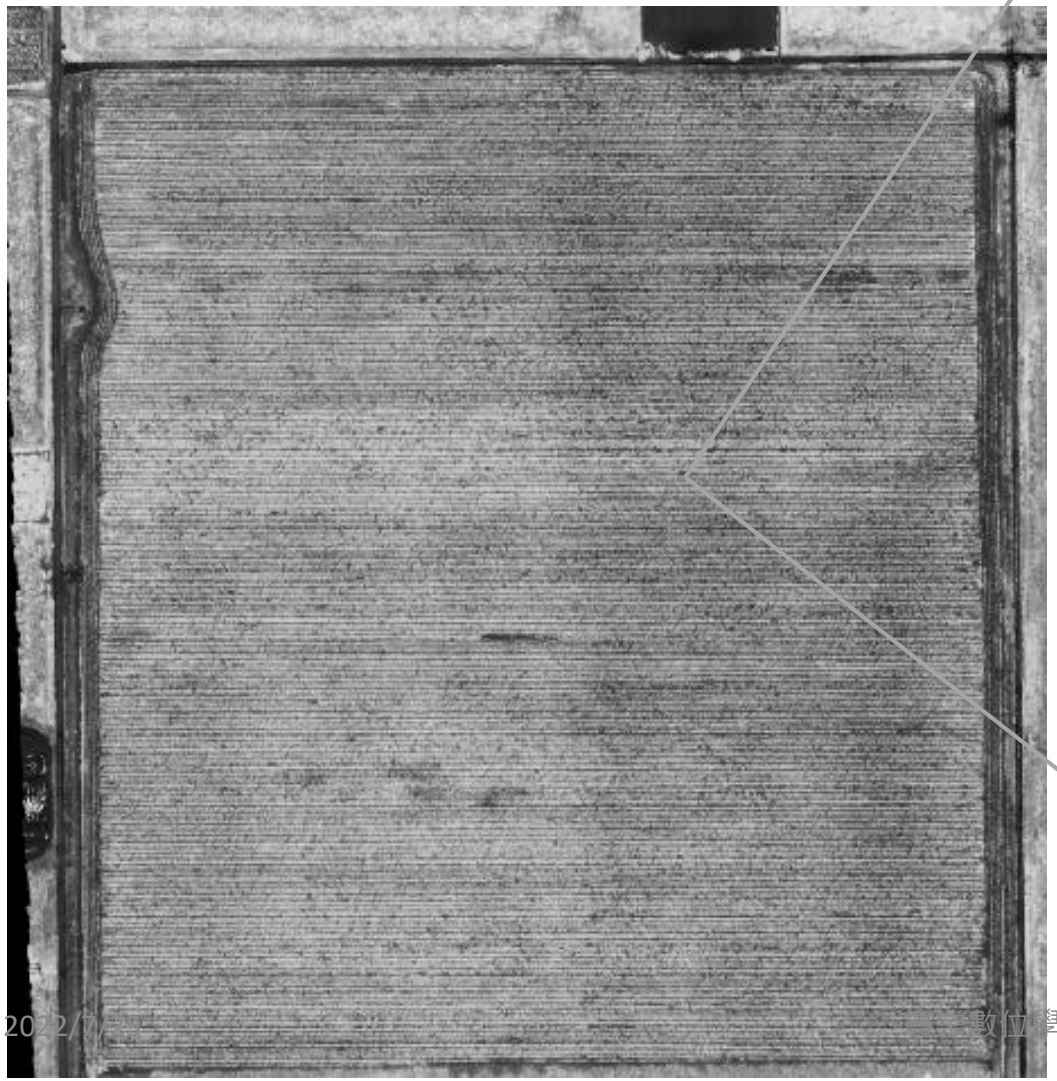
蘆筍長度預測模式

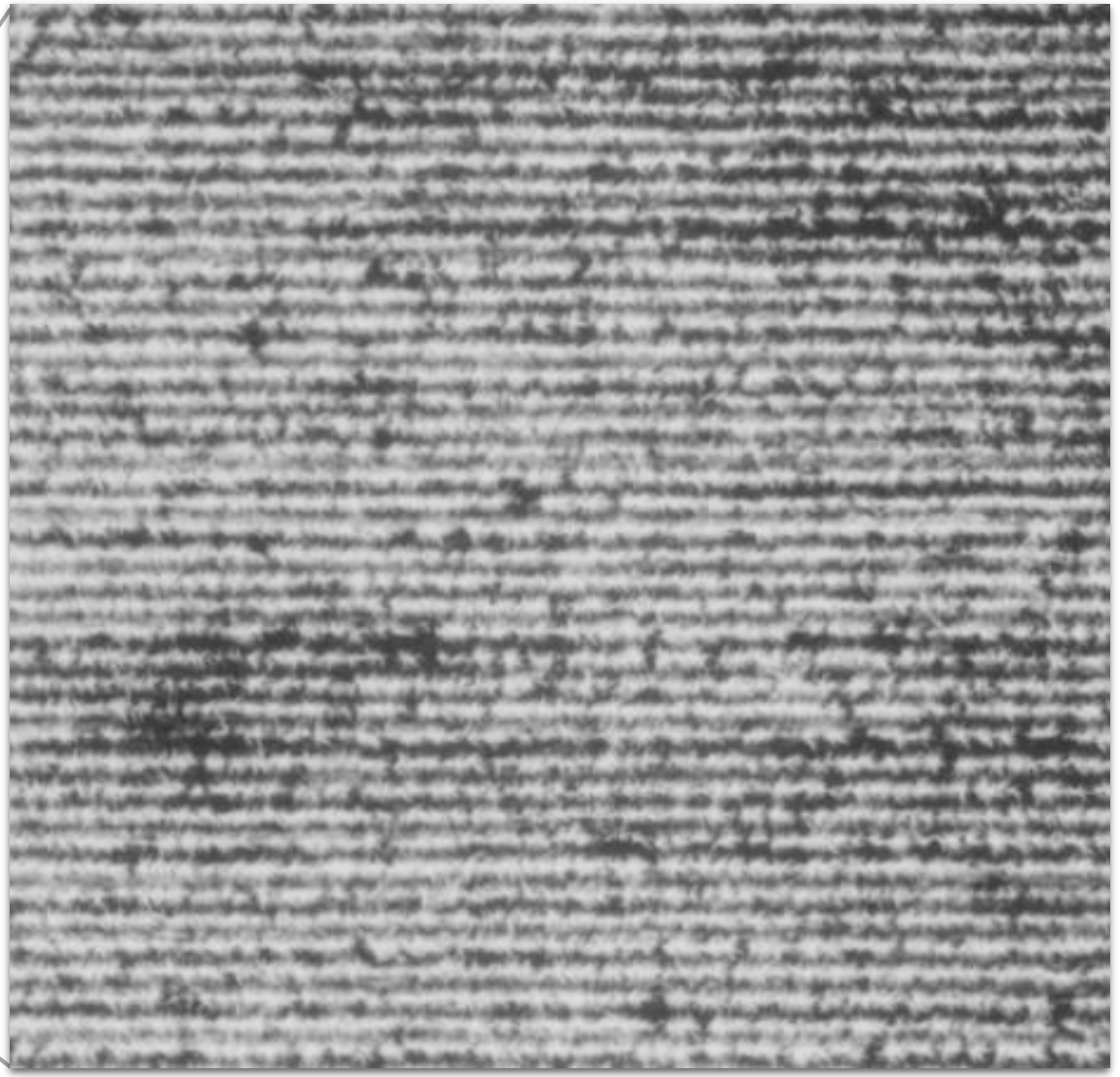

Adjusted Height
0
10
20
30
40
0
500
1000
1500
2000
2500
3000
3500
Thermal Time

- **挑戰:**
  - 標註所需的專業知識
  - 圖資檔案龐大
  - 田間數據多雜訊
- **機會: 透過統計與機器學習的整合**
  - More informative
  - Computationally efficient / less time consuming
  - Applicable on small sample

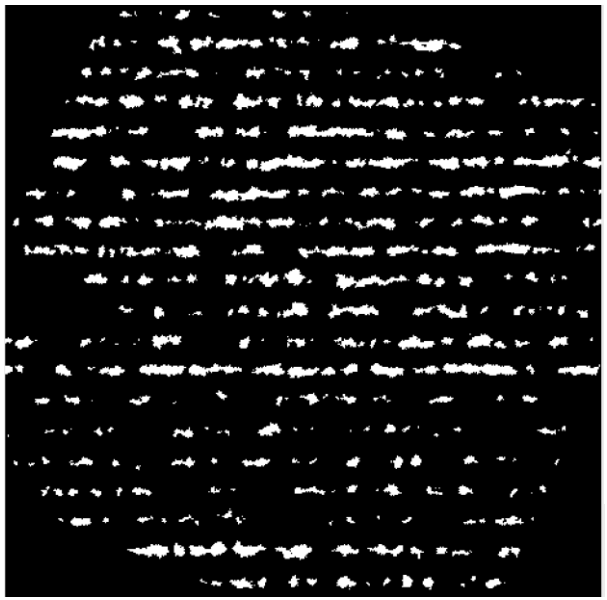
資料會說話 …

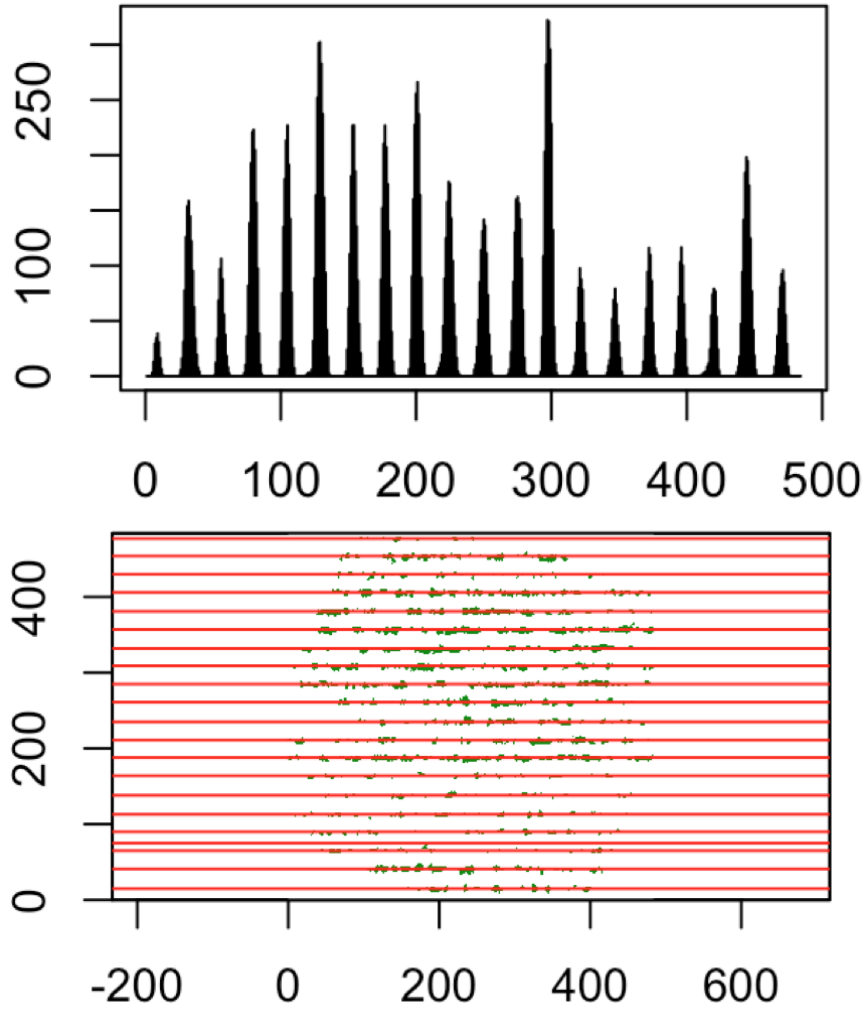
甘蔗田缺株辨識

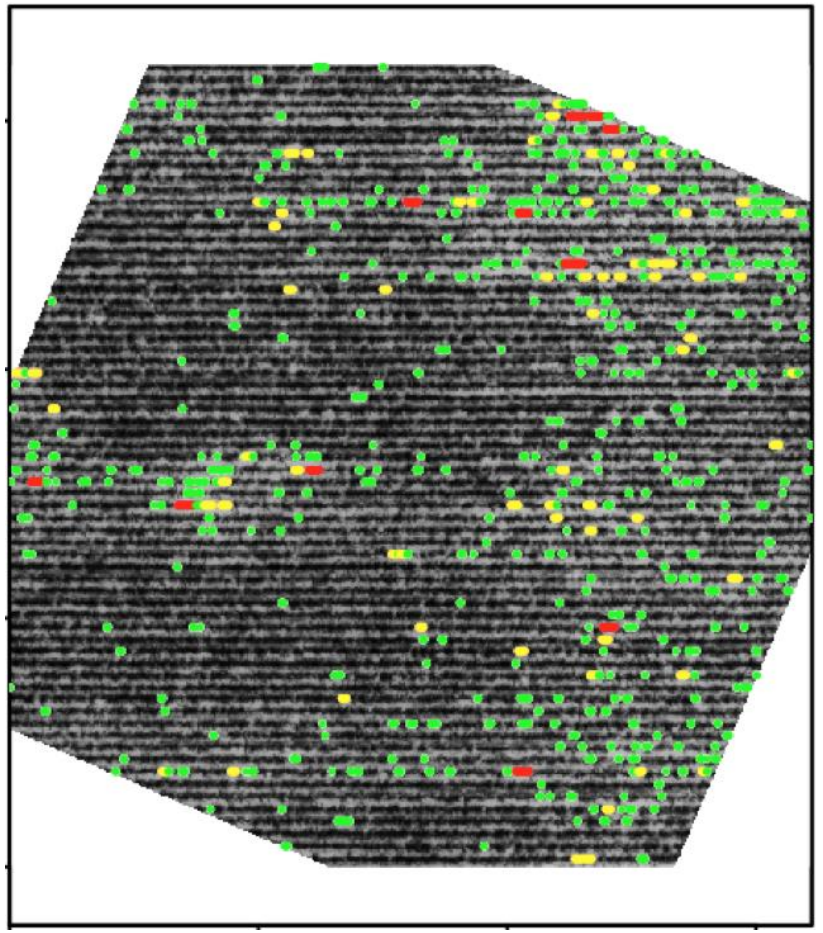




甘蔗田缺株辨識



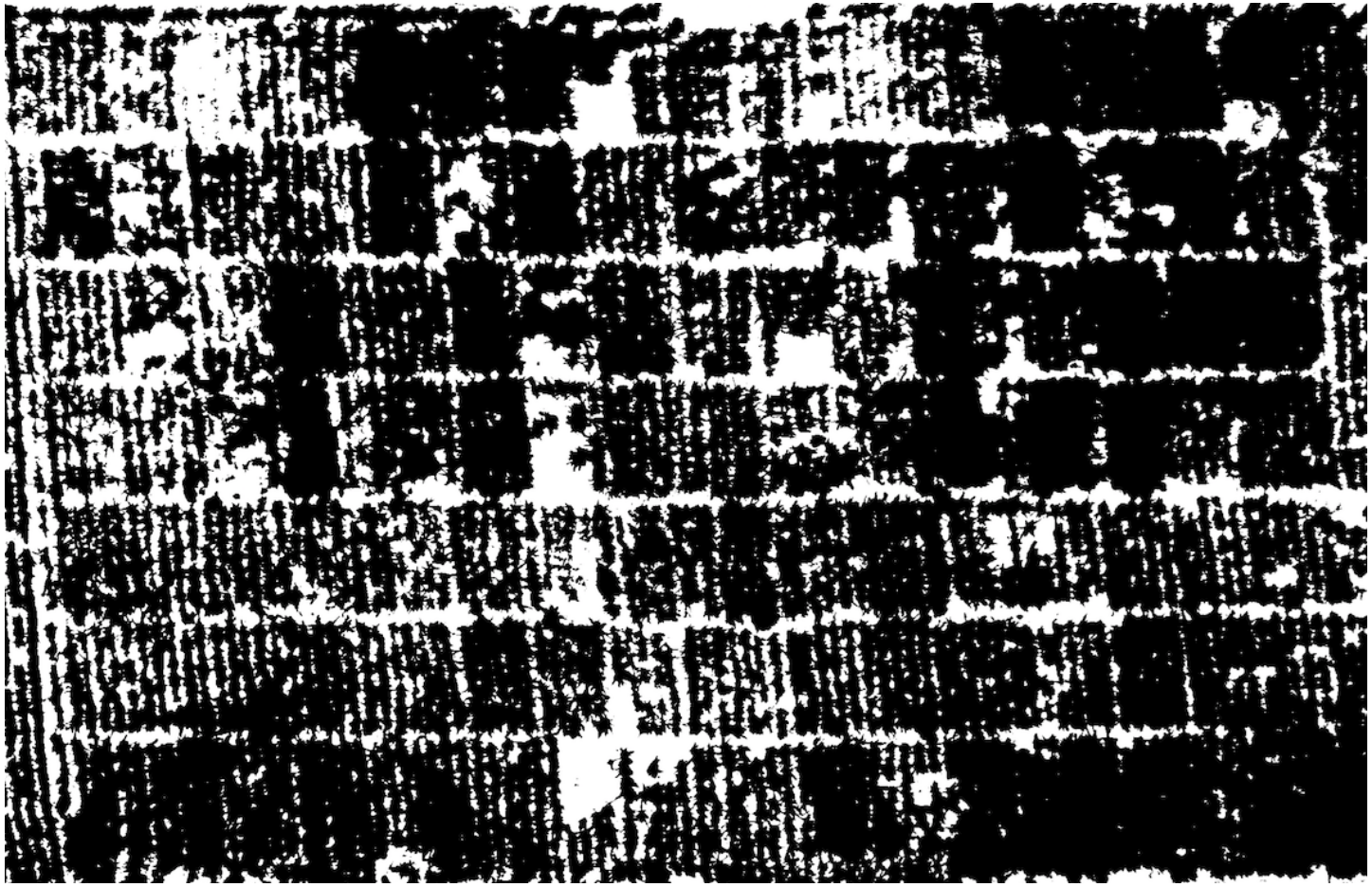

0
100
200
300
400
500
0
100
250
0
200
400
-200
0
200
400
600


0-10
11-20
>20

稻熱病影像辨識

37%	57%	53%	56%	33%	21%	18%	14%
56%	58%	34%	37%	34%	46%	20%	12%
21%	0%	52%	43%	0%	0%	4%	0%
8%	28%	9%	2%	4%	0%	8%	0%
8%	47%	0%	16%	66%	63%	46%	0%
28%	56%	50%	30%	47%	21%	0%	39%
42%	14%		29%	49%	19%		

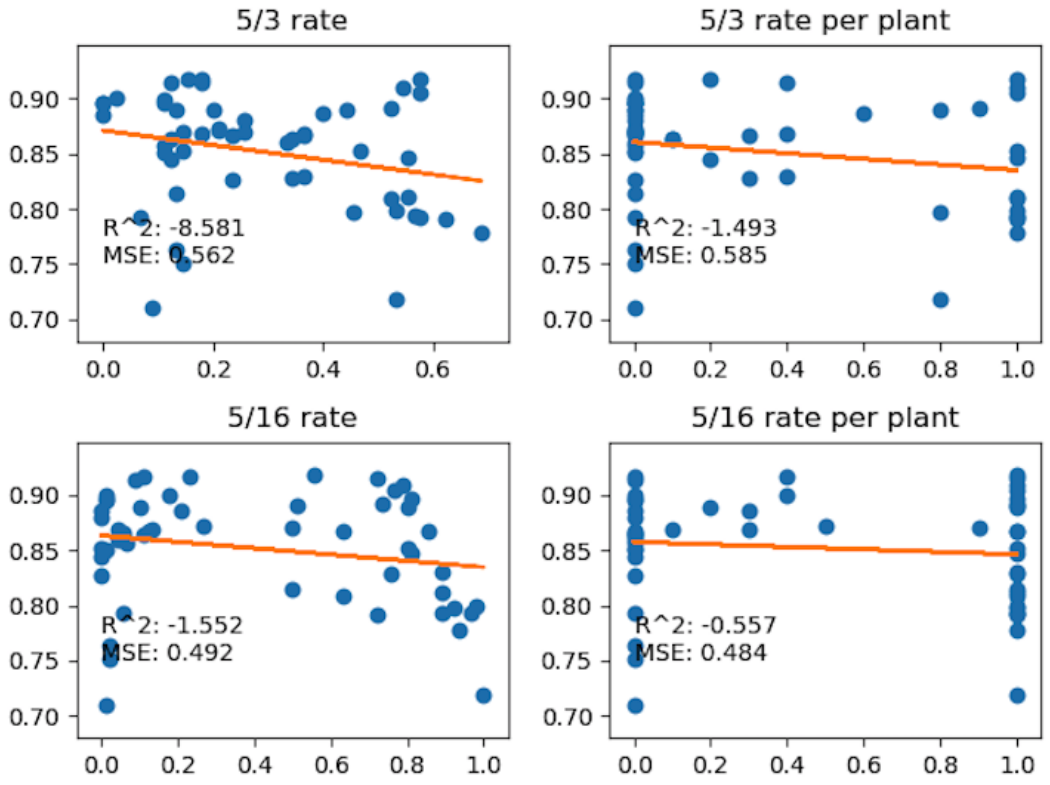
49%	58%	71%	70%	22%	14%	7%	3%
26%	49%	14%	16%	22%	72%	0%	11%
21%	0%	52%	43%	0%	0%	4%	0%
8%	28%	9%	2%	4%	0%	8%	0%
8%	47%	0%	16%	66%	63%	46%	0%
28%	56%	50%	30%	47%	21%	0%	39%
42%	14%		29%	49%	19%		



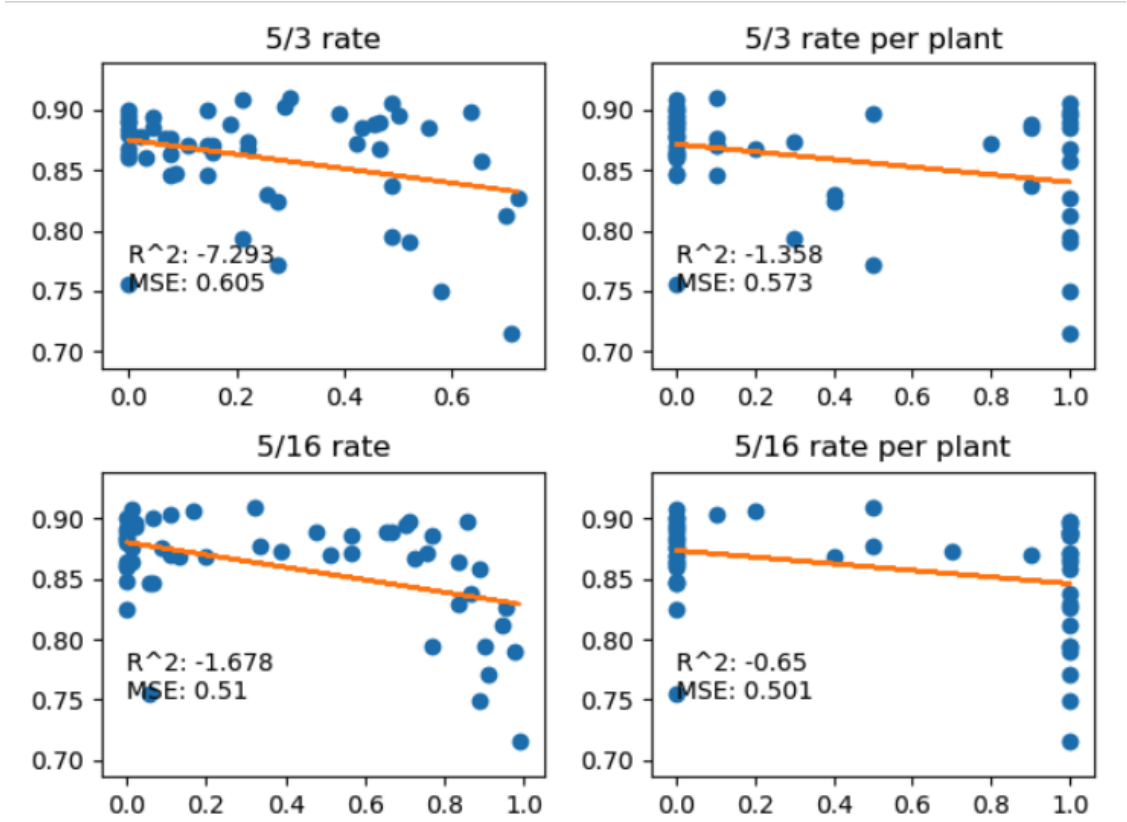
稻熱病影像辨識

Correlation between **NDVI** and **Ground Truth**

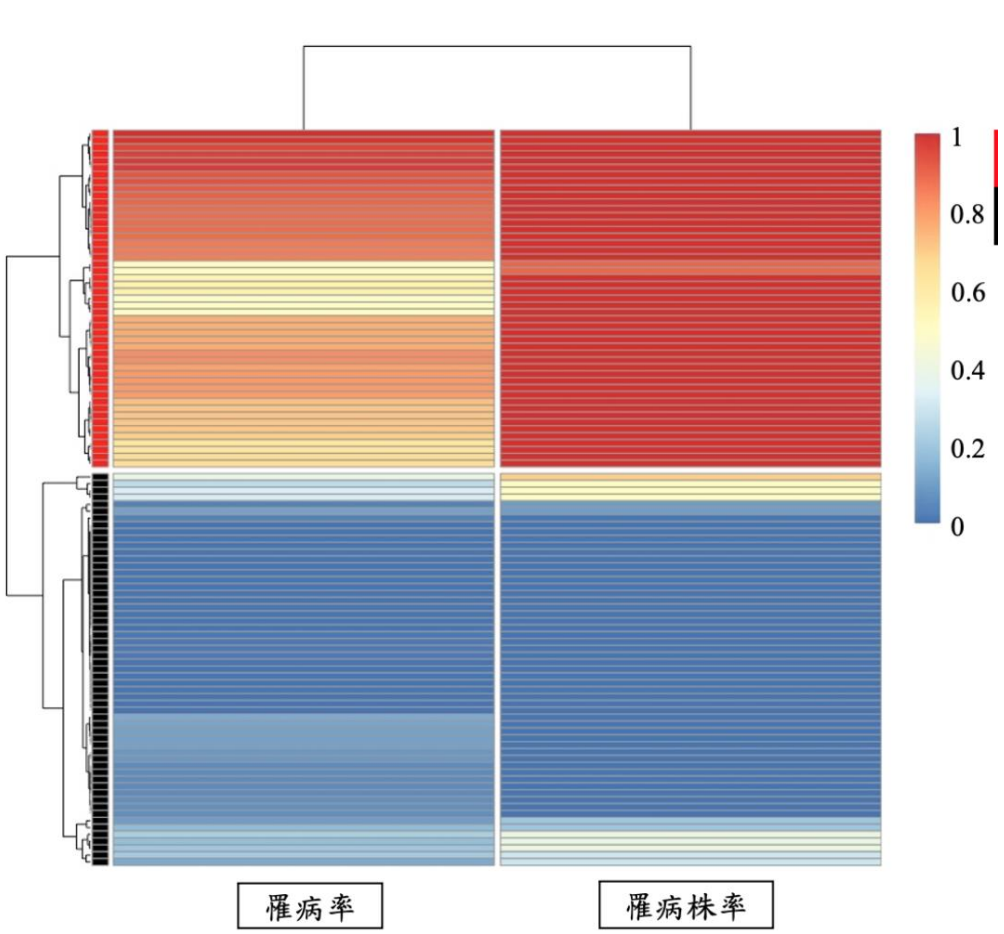
(i) 左半病圃

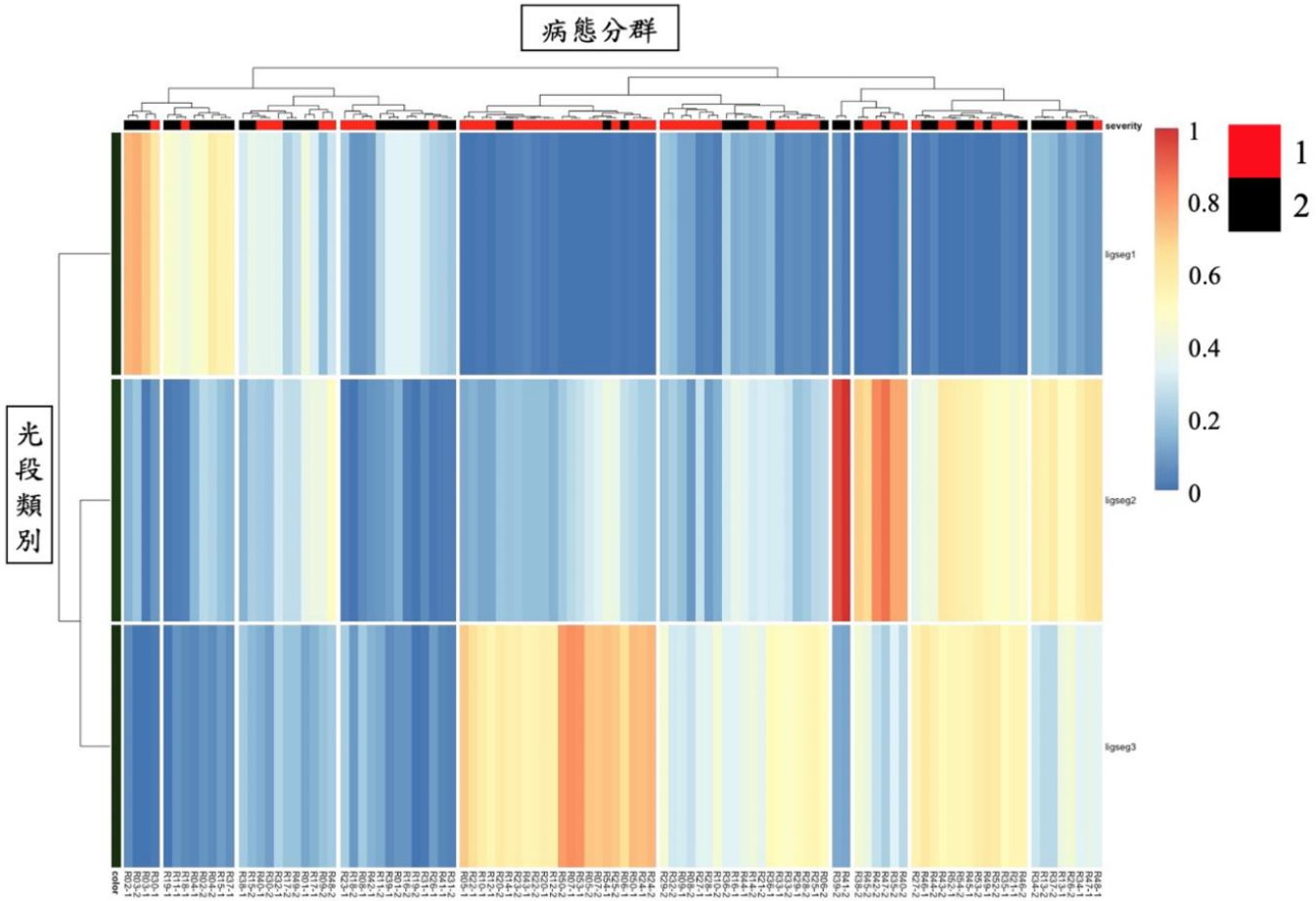


(ii) 右半病圃

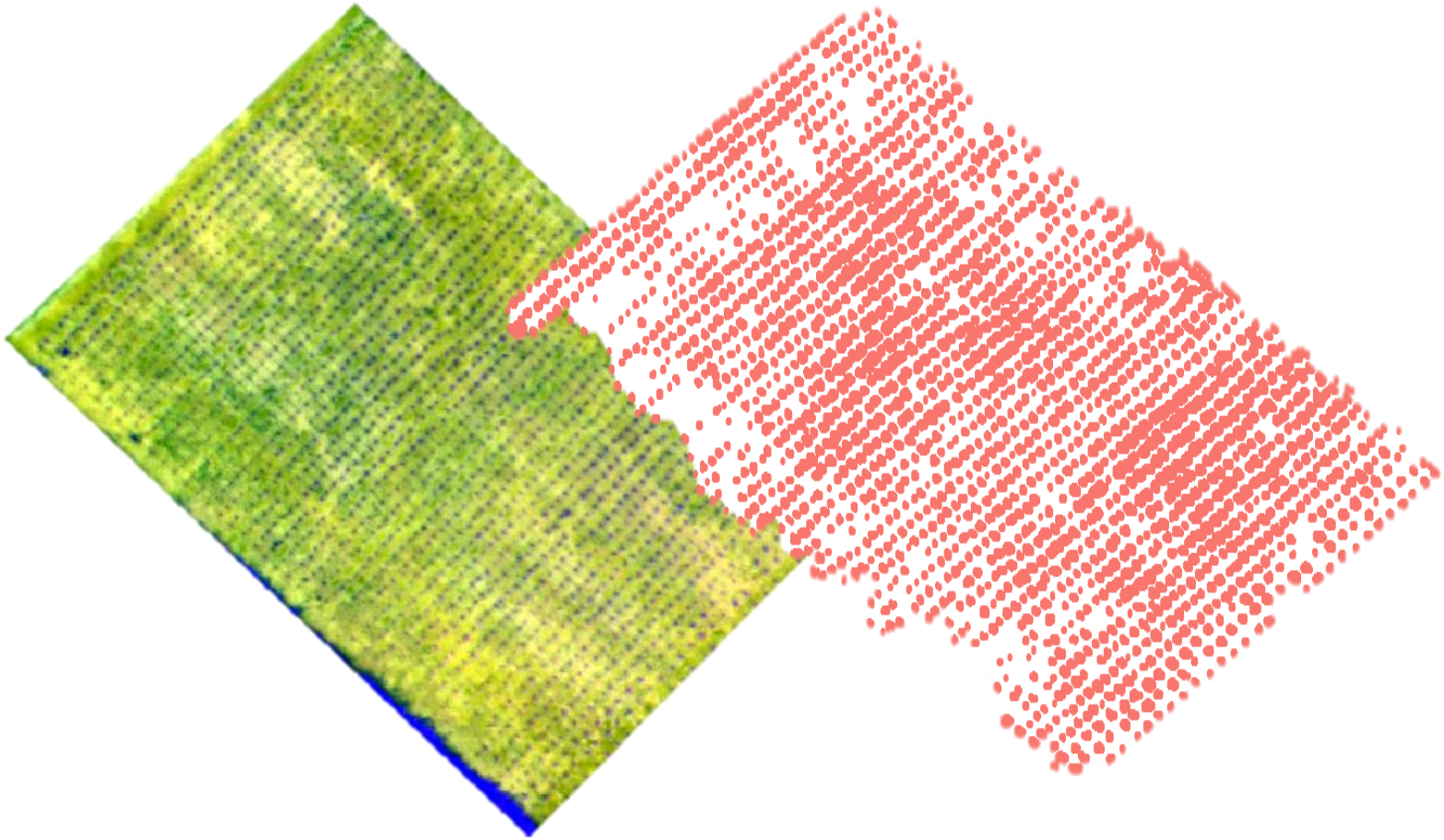


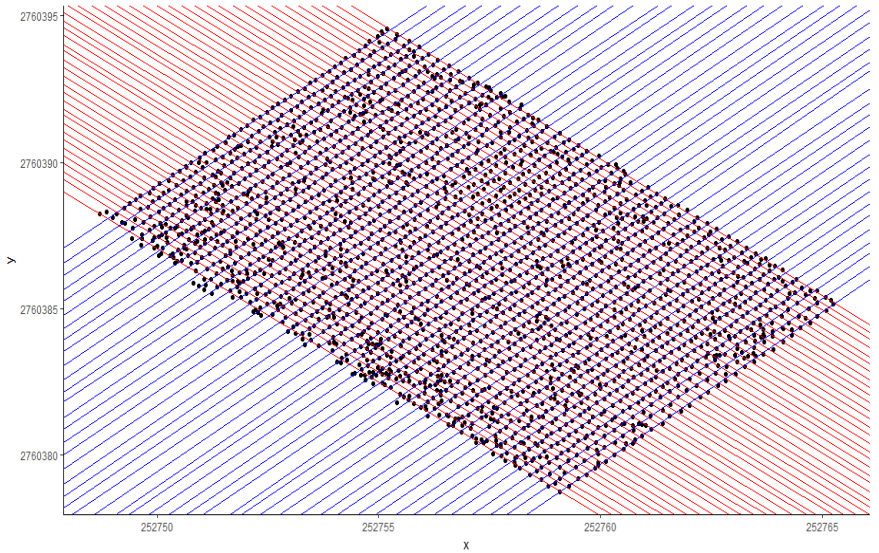
稻熱病影像辨識

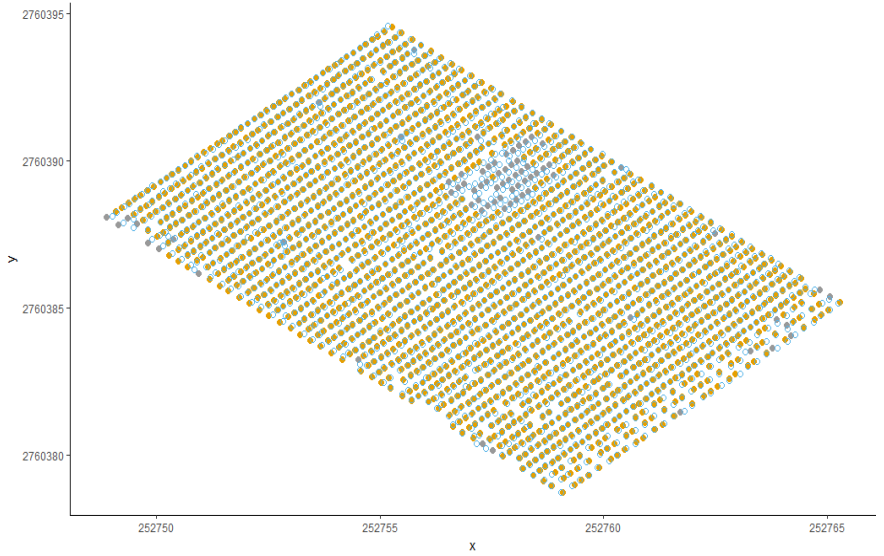

罹病率
罹病株率
1
0.8
0.6
0.4
0.2
0


病態分群
光段類別
severity
ligseg1
ligseg2
ligseg3
color
1
0.8
0.6
0.4
0.2
0
1
2

尋找秧苗中心位置

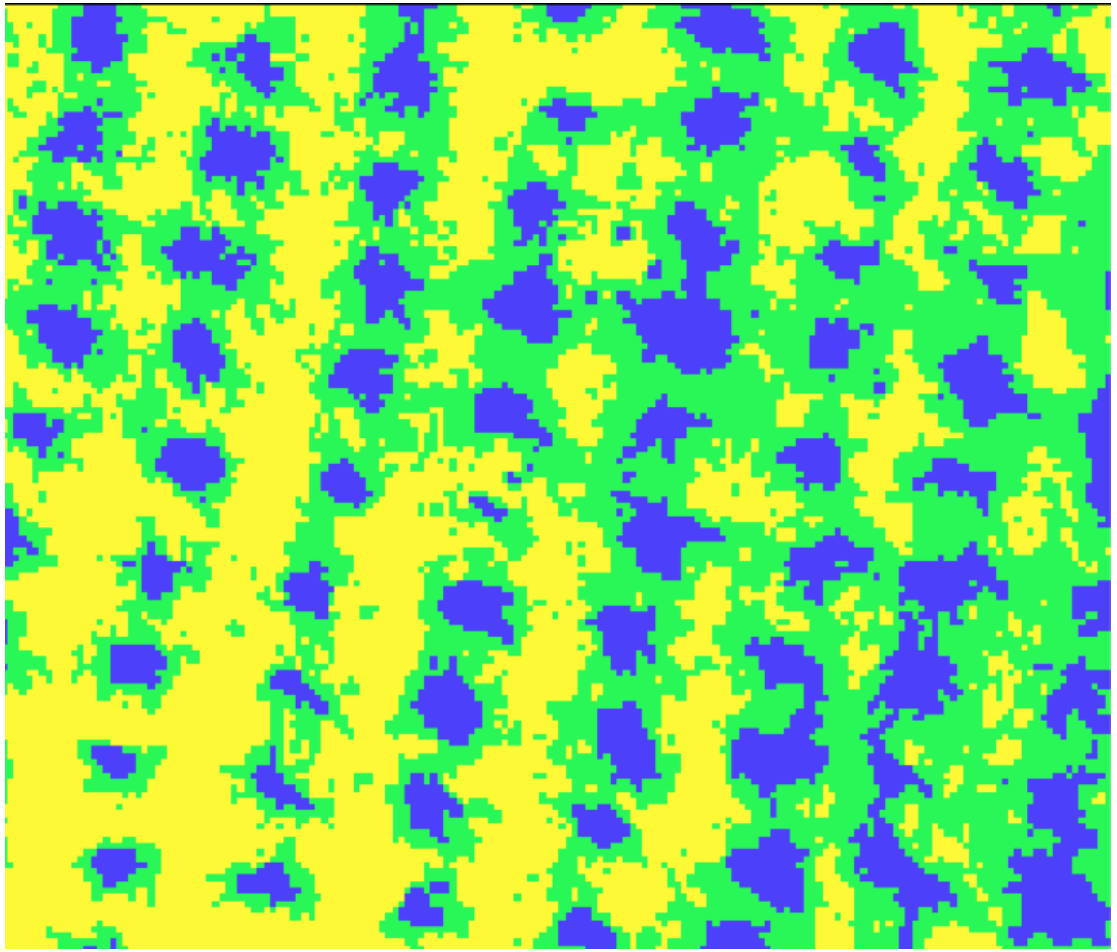



2760395
2760390
2760385
2760380
y
252750
252755
252760
252765
x

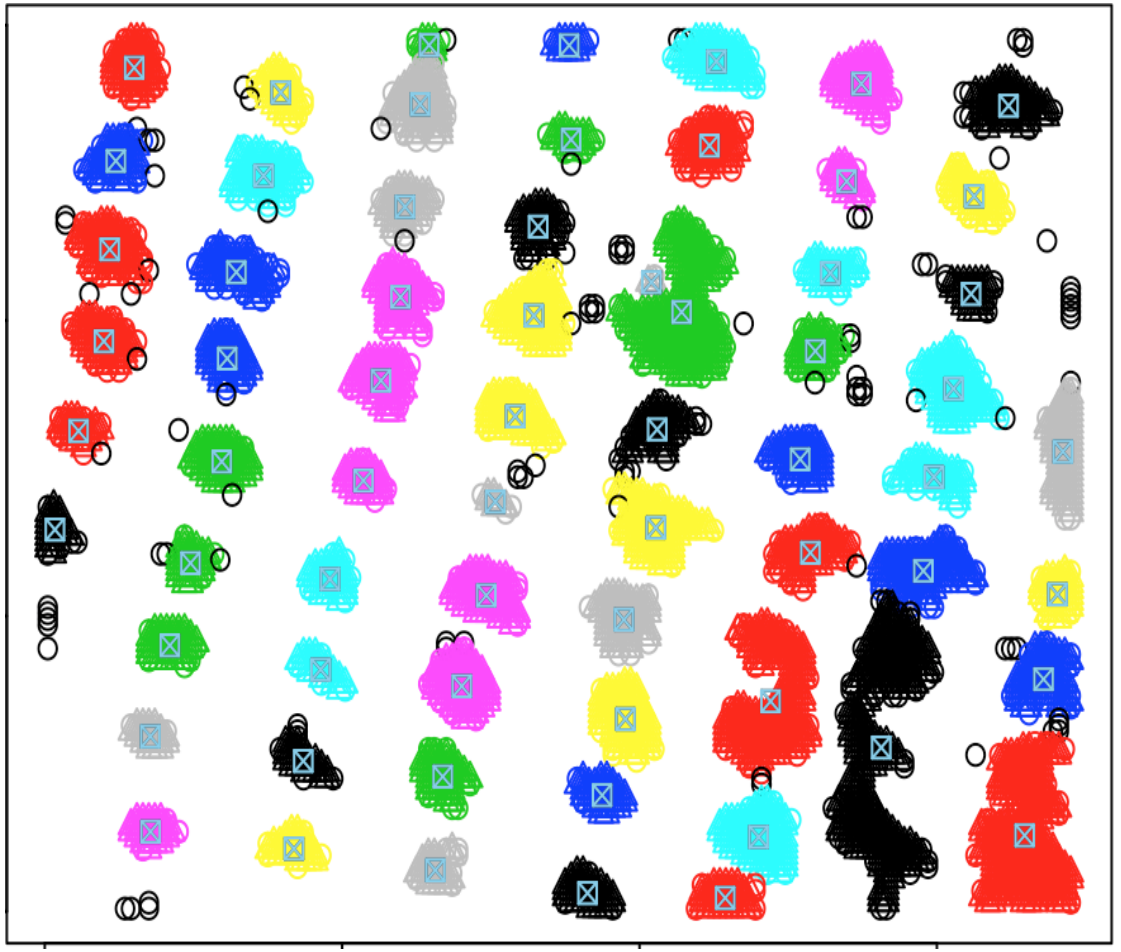

2760395
2760390
2760385
2760380
y
252750
252755
252760
252765
x

尋找秧苗中心位置

DBSCAN: Density-Based Spatial Clustering of Application with Noise

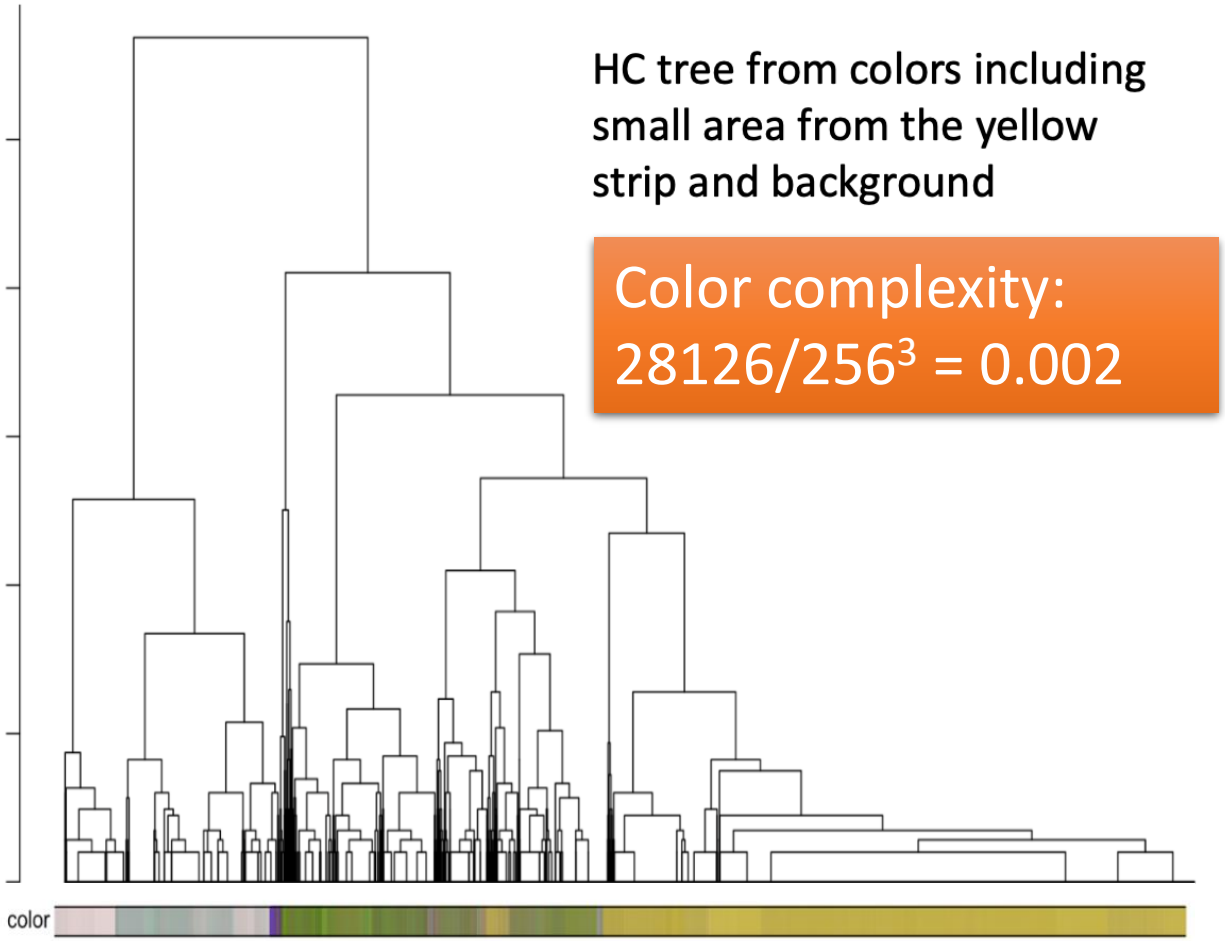


False color image



Grouping by DBSCAN

水試紙分析



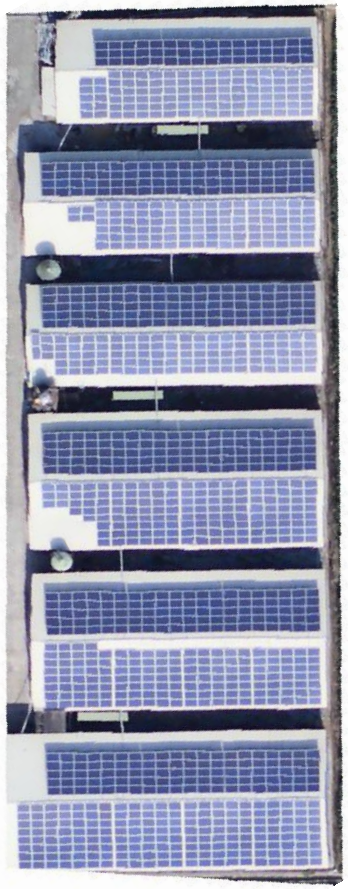
HC tree from colors including small area from the yellow strip and background

Color complexity: $28126/256^3 = 0.002$

Shuting Liao, Li-Yu Liu, Ting-An Chen, Kuang-Yu Chen and Fushing Hsieh. (2021)

田區分類

greenhouse



maize

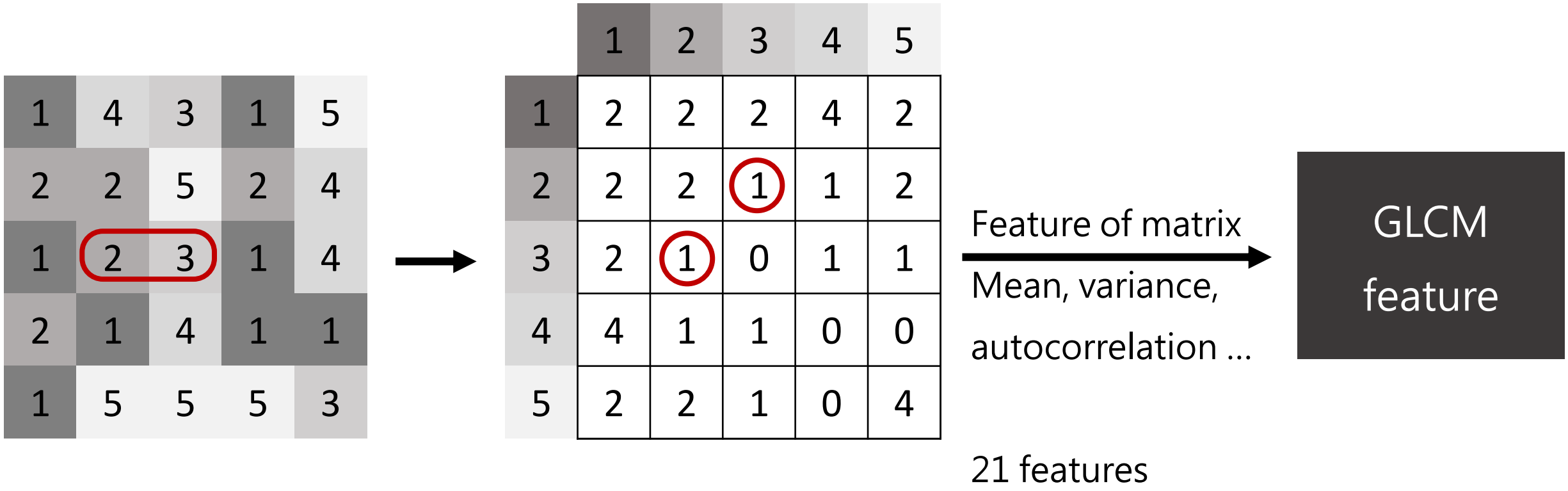
cruciferae

fruitTree

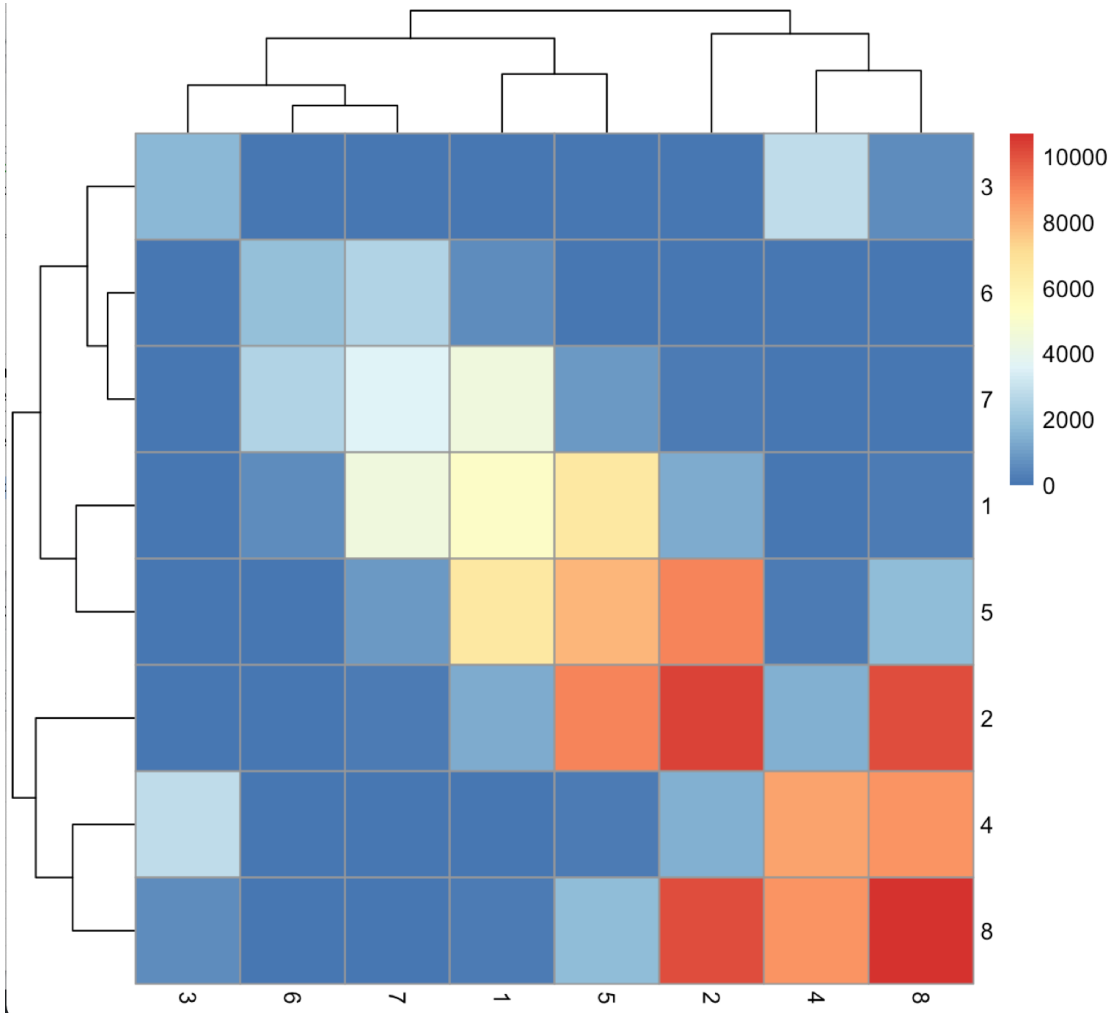
sugarcane

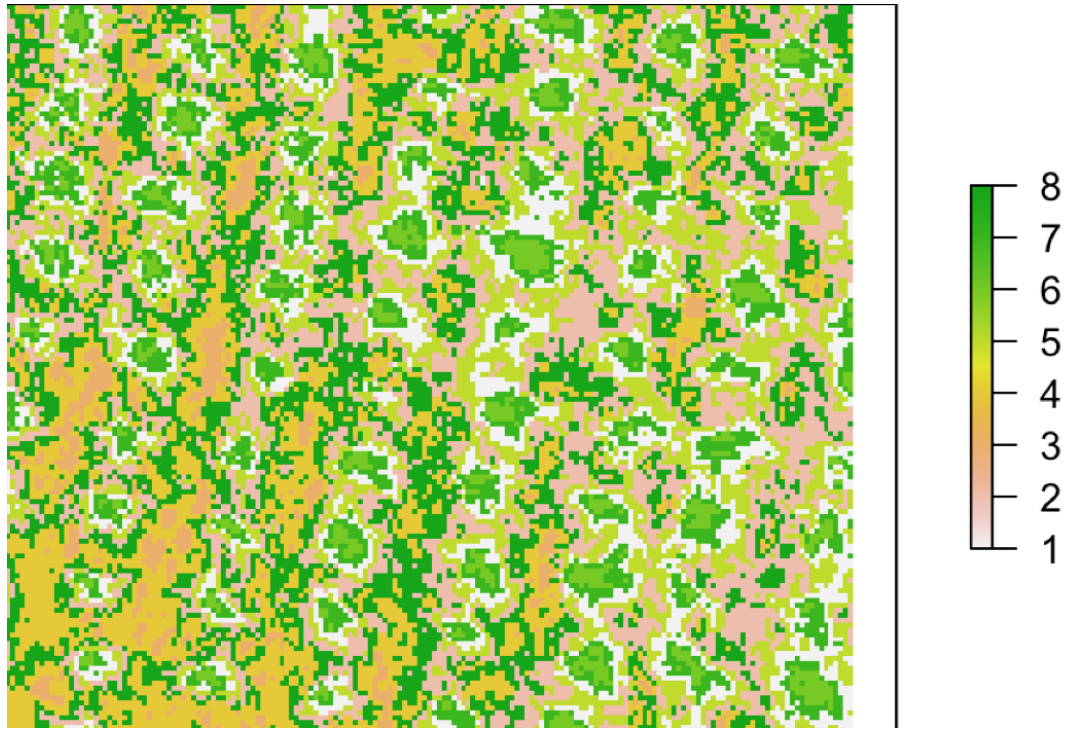
rice

田區分類 - GLCM


1 4 3 1 5
2 2 5 2 4
1 2 3 1 4
2 1 4 1 1
1 5 5 5 3
1 2 3 4 5
1 2 2 2 4 2
2 2 2 1 1 2
3 2 1 0 1 1
4 4 1 1 0 0
5 2 2 1 0 4
Feature of matrix
Mean, variance,
autocorrelation …
21 features
GLCM
feature

田區分類 - GLCM





- 找同一時間不同品種病圃比較 transition matrix 差異
- 找不同時間同一塊病圃比較transition matrix 差異
- Transition matrix 的變化會不會跟罹病率有關？

採收紀錄與IoT資料整合

採收前 k 天日射量 (k = 1, …, 7)

採收日期 採收重量

採收前 k 天溫度 (k = 1, …, 7)

採收前 k 天濕度 (k = 1, …, 7)

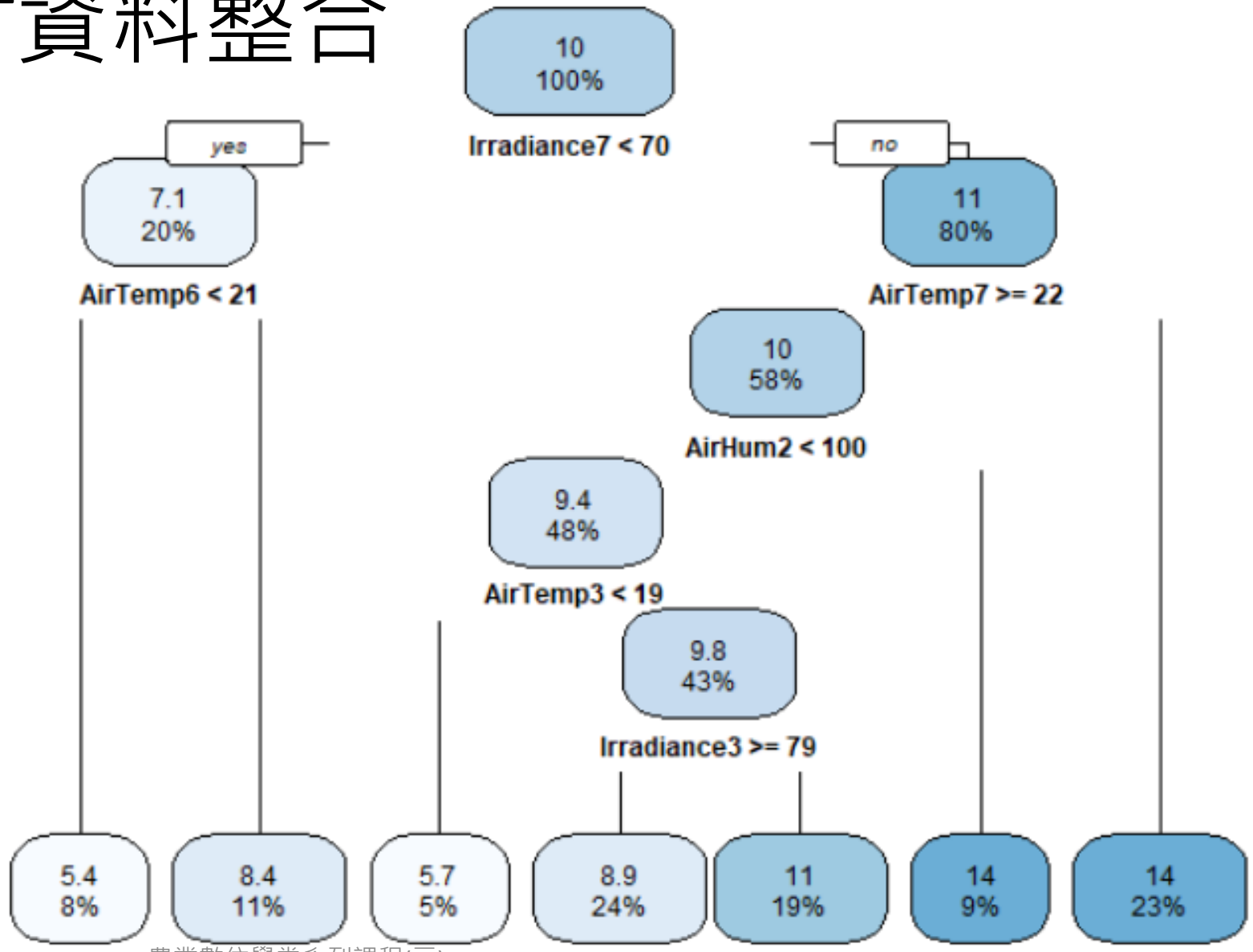
	Time	Weight	AirTemp1	AirTemp2	AirTemp3	AirTemp4	AirTemp5	AirTemp6	AirTemp7	AirHum1	AirHum2	AirHum3	AirHum4	AirHum5	AirHum6	AirHum7	Irrad
1	2021/11/25	8.6	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
2	2021/11/25	11.4	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
3	2021/11/25	15.3	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
4	2021/11/25	8.3	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
5	2021/11/25	9.1	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
6	2021/11/25	14.1	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
7	2021/11/25	4.9	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
8	2021/11/25	8.3	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
9	2021/11/25	6.4	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
10	2021/11/25	8.1	22.150	20.212	22.650	25.670	24.904	25.048	24.710	98.136	99.110	98.842	98.306	98.856	98.868	98.816	
11	2021/11/26	0.4	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
12	2021/11/26	13.3	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
13	2021/11/26	5.0	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
14	2021/11/26	15.8	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
15	2021/11/26	13.0	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
16	2021/11/26	13.7	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	
17	2021/11/26	18.9	21.780	22.150	20.212	22.650	25.670	24.904	25.048	99.758	98.136	99.110	98.842	98.306	98.856	98.868	

採收紀錄與IoT資料整合

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 354.37825  375.73955   0.943   0.3468
AirTemp1      1.37090    1.37467   0.997   0.3199
AirTemp2      2.23799    1.31114   1.707   0.0894 .
AirTemp3      1.30636    2.62679   0.497   0.6195
AirTemp4     -4.42639    2.80824  -1.576   0.1166
AirTemp5      0.25631    1.06760   0.240   0.8105
AirTemp6     -1.05308    2.10163  -0.501   0.6169
AirTemp7     -1.23025    1.54830  -0.795   0.4278
AirHum1      -5.35468    4.69598  -1.140   0.2556
AirHum2       1.08650    1.80163   0.603   0.5472
AirHum3      -0.73495    1.19370  -0.616   0.5388
AirHum4       1.33373    1.42752   0.934   0.3513
AirHum5      -1.87958    0.91146  -2.062   0.0405 *
AirHum6      -0.07772    1.01347  -0.077   0.9390
AirHum7       2.58723    1.79002   1.445   0.1499
Irradiance1  -0.11774    0.06390  -1.843   0.0669 .
Irradiance2  -0.15368    0.07014  -2.191   0.0296 *
Irradiance3   0.07826    0.07986   0.980   0.3283
Irradiance4   0.19199    0.10564   1.817   0.0707 .
Irradiance5   0.24461    0.10822   2.260   0.0249 *
Irradiance6  -0.01046    0.15624  -0.067   0.9467
Irradiance7  -0.27508    0.21166  -1.300   0.1953
```

$R^2 = 0.2986$

採收紀錄與IoT資料整合

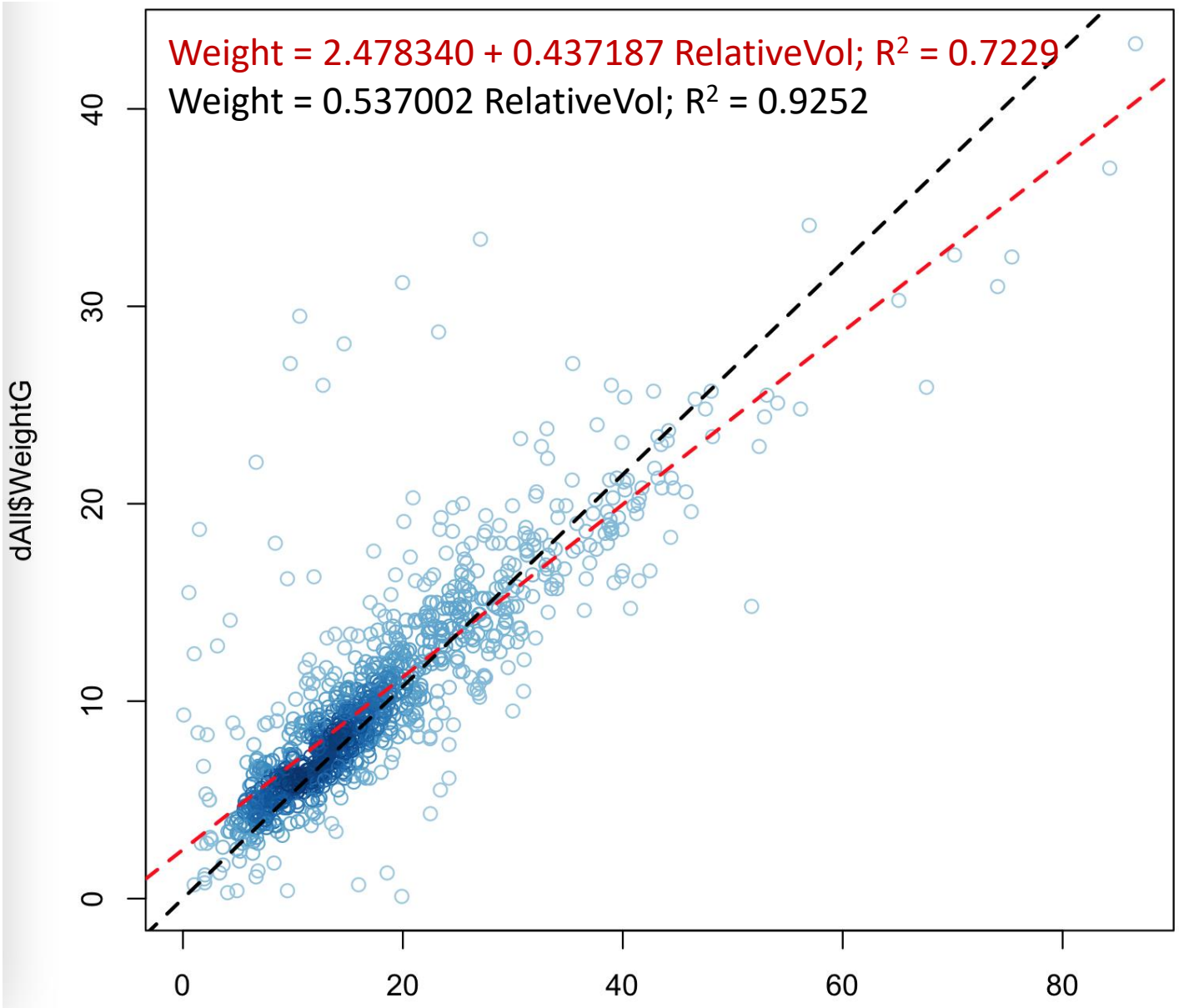


分析蘆筍收穫紀錄，以體積估算重量！

體積 (Volume) = $(2\pi \times 半徑^2) \times 長度$

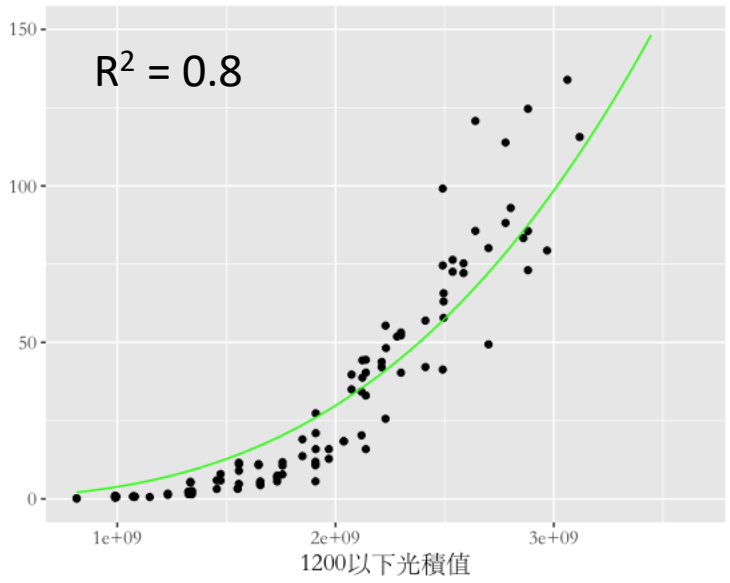
$= (2\pi \times (寬度/2)^2) \times 長度$

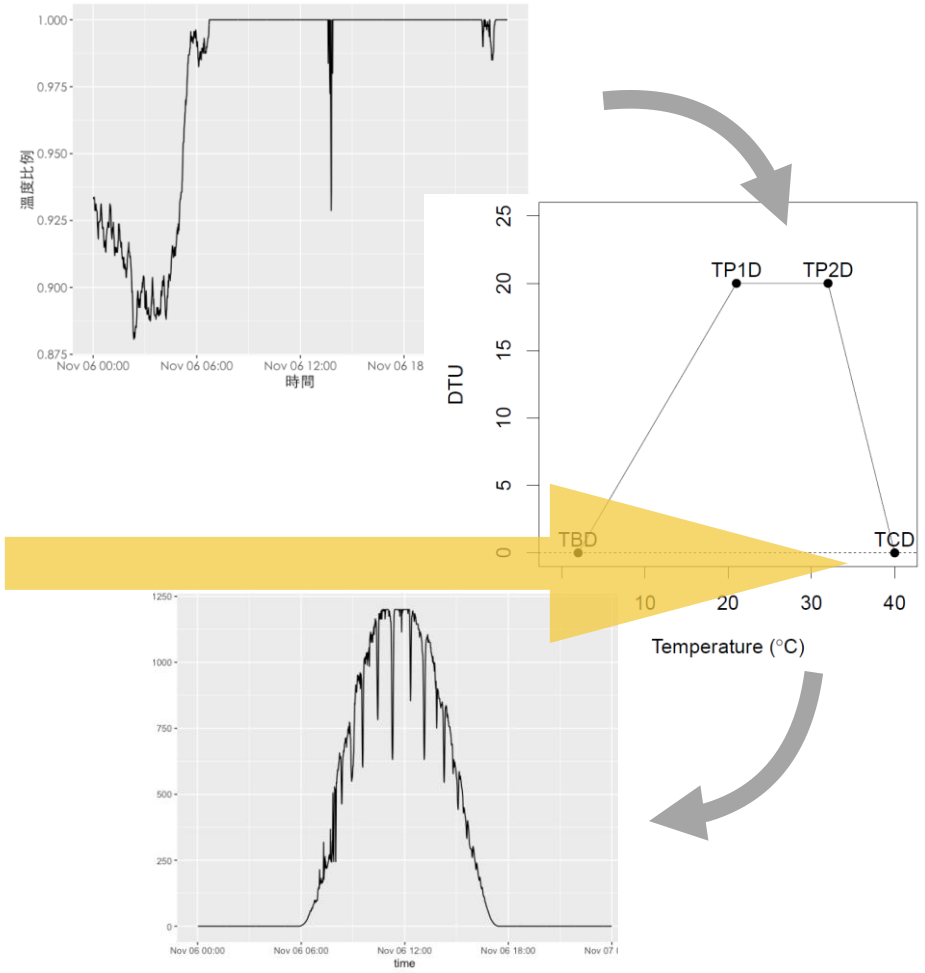
RelativeVol = $寬度^2 \times 長度$



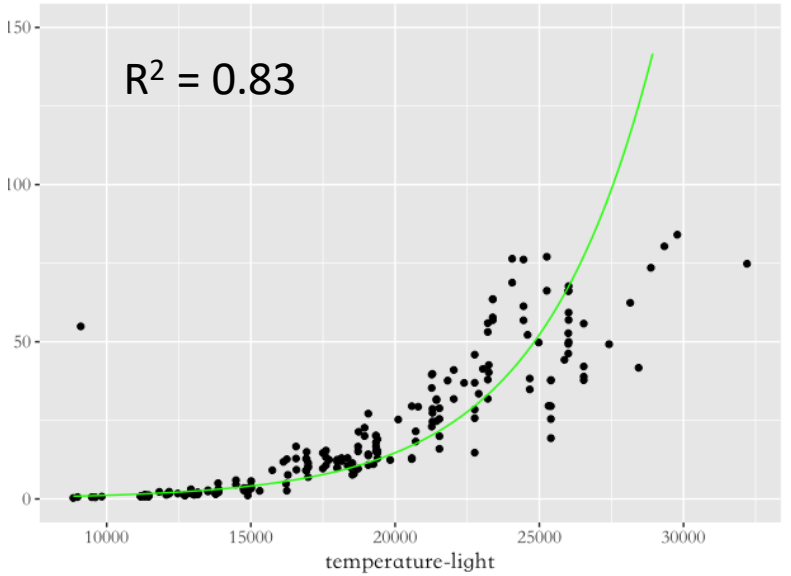
葉菜產量預測模型

2020 光積值產量預測模型





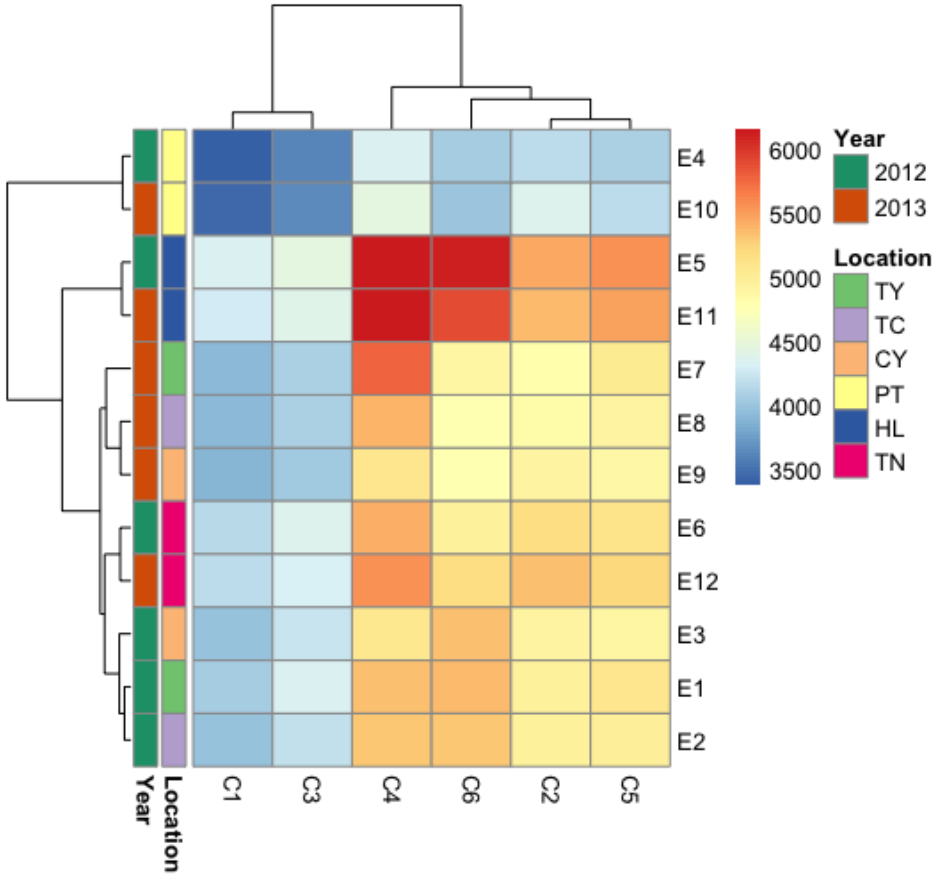
2020 校正光積值
產量預測模型



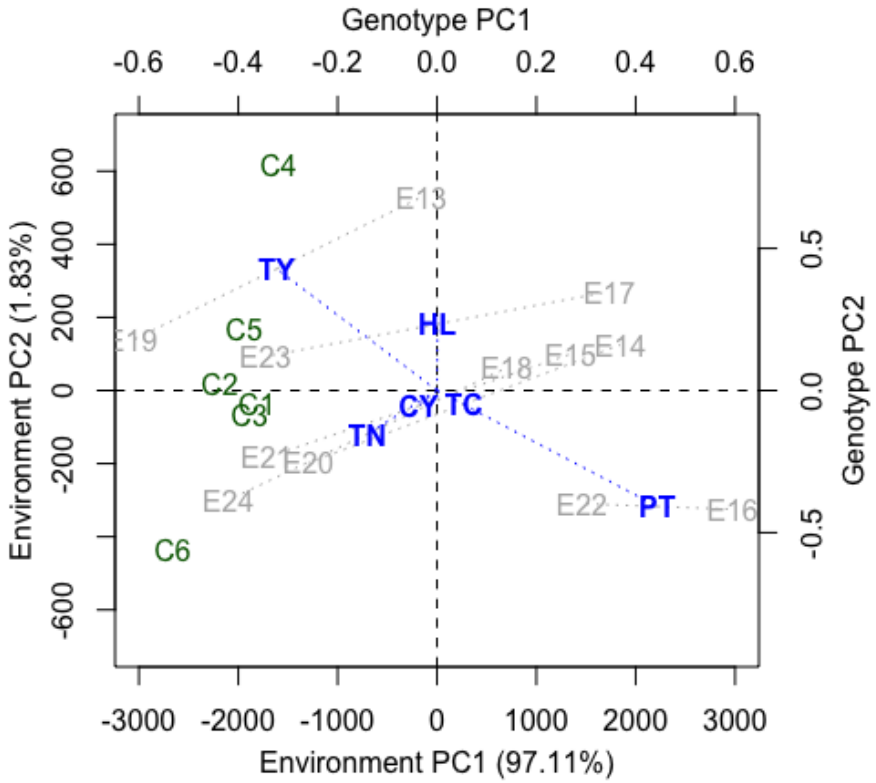
產量預測準確度提升：
青梗白菜：30.84%
小白菜：25.90%
蕹菜：27.52%

Application: 推薦適種品種

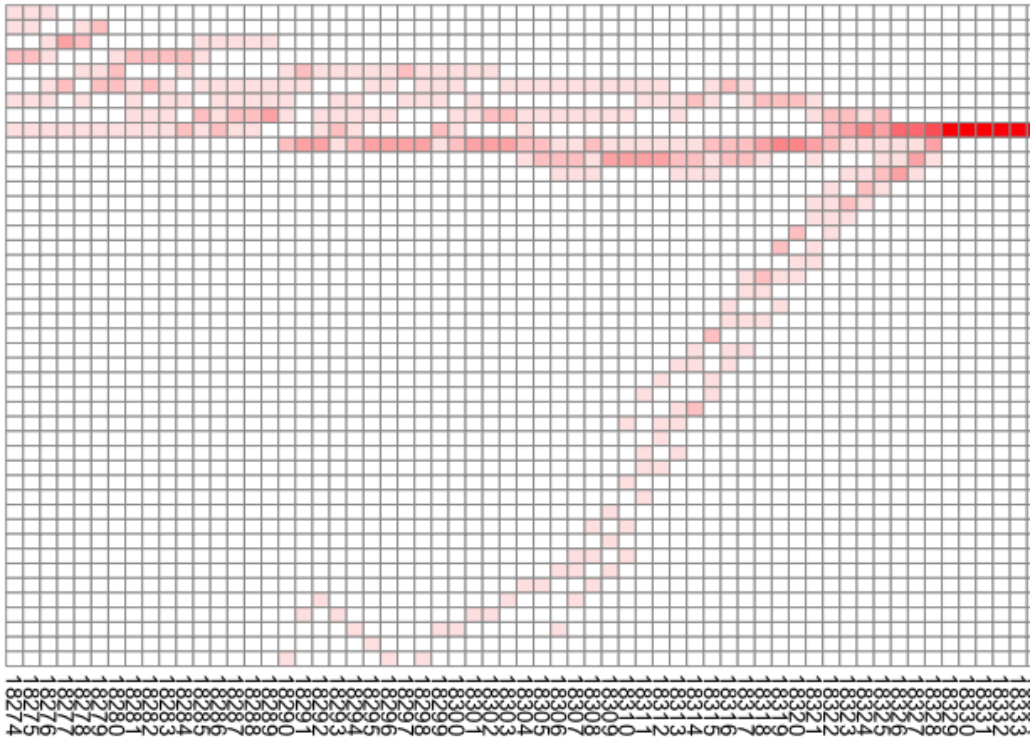
Heatmap
熱圖

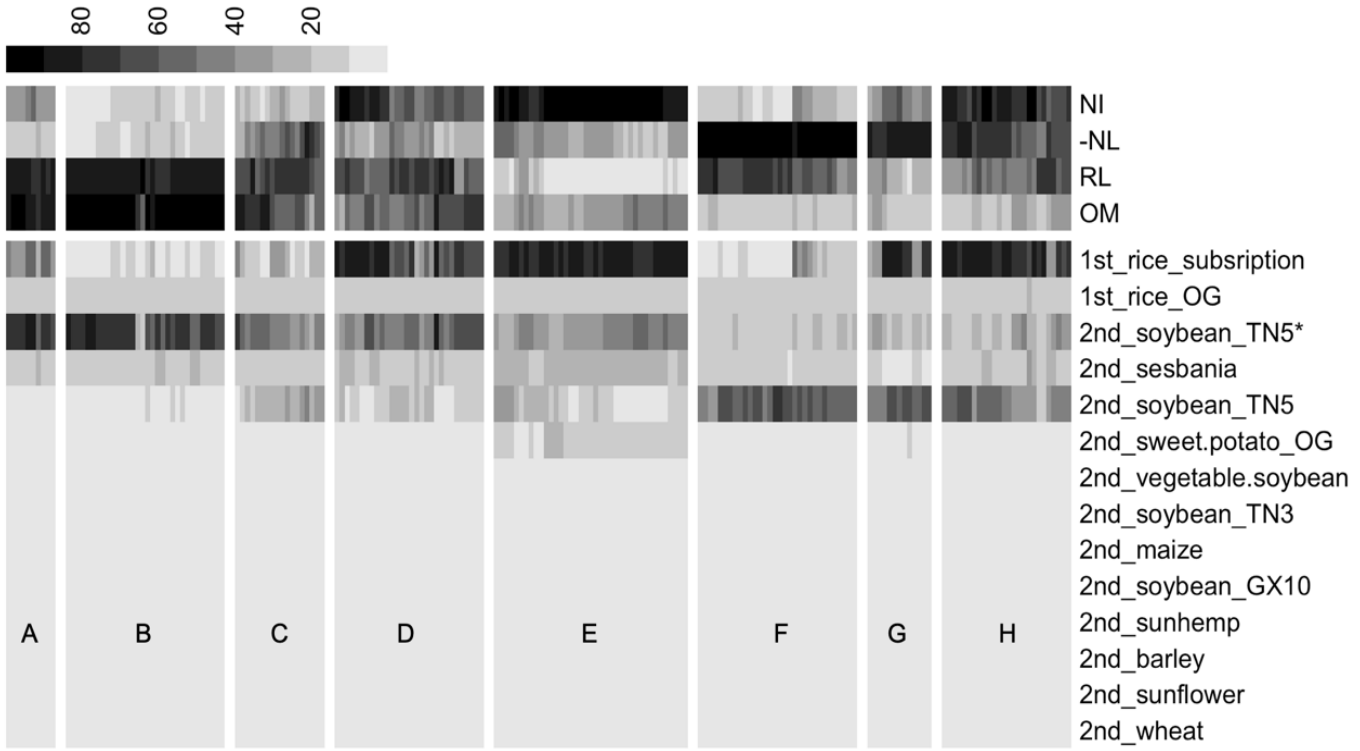


Biplot
雙軸圖



其他系統整合應用




80
60
40
20
NI
-NL
RL
OM
1st_rice_subsription
1st_rice_OG
2nd_soybean_TN5*
2nd_sesbania
2nd_soybean_TN5
2nd_sweet.potato_OG
2nd_vegetable.soybean
2nd_soybean_TN3
2nd_maize
2nd_soybean_GX10
2nd_sunhemp
2nd_barley
2nd_sunflower
2nd_wheat
A
B
C
D
E
F
G
H

訊息傳遞

- 對象：機器？農民？決策者？官員？
- 訊息接收方式？
- Data Visualization

Suggestions:

機器學習、大數據分析與資料視覺化

- An appropriate experimental design is essential to obtain information-rich data that can convey scientific insights. (Get **GOOD** data!)
- No matter the sample size, you MUST to know your data by proper visualization before analyses. (Data intelligence (**DI**) instead of AI!)
- A deep learning algorithm is not necessarily better than a simple machine learning algorithm / statistical modeling method. (Do not overfitting!!)
- **Language?** Pick one and master it!

Acknowledgement

教育部 Ministry of Education

教育部人工智慧競賽與標註資料蒐集辦公室

國立臺灣大學生物資源暨農學院 College of Bioresources and Agriculture, National Taiwan University

經緯航太科技股份有限公司 GEOSAT AEROSPACE & TECHNOLOGY INC.

先端智農實驗室 Apex Agri-Intelligence Lab

臺大工科海洋系-知識計算實驗室

農委會 臺南區農改場 TAINAN D.A.R.E.S. 1902

桃園區農業改良場 TAOYUAN DARES 1982

桃改場賴信忠研究員