

# Greens 計劃：求解 TBAF 構造分數

廖鎮磐

`<andrew.43@gmail.com>`

November 18, 2015

## 地表型態代碼

代碼	表面型態組	表面型態項目
A	人工鋪面	不透水鋪面
B		透水鋪面
C	園景設施：底土連接	植栽覆蓋不與底土連接
D		植栽覆蓋與底土連接
E	園景設施：植被類型	地被
F		灌木
G		小喬木
H		中喬木
I		大喬木
J		複層植栽
K	生物滯留設施	綠屋頂
L		綠牆
M		雨水花園

# 大綱

現有資料檢查

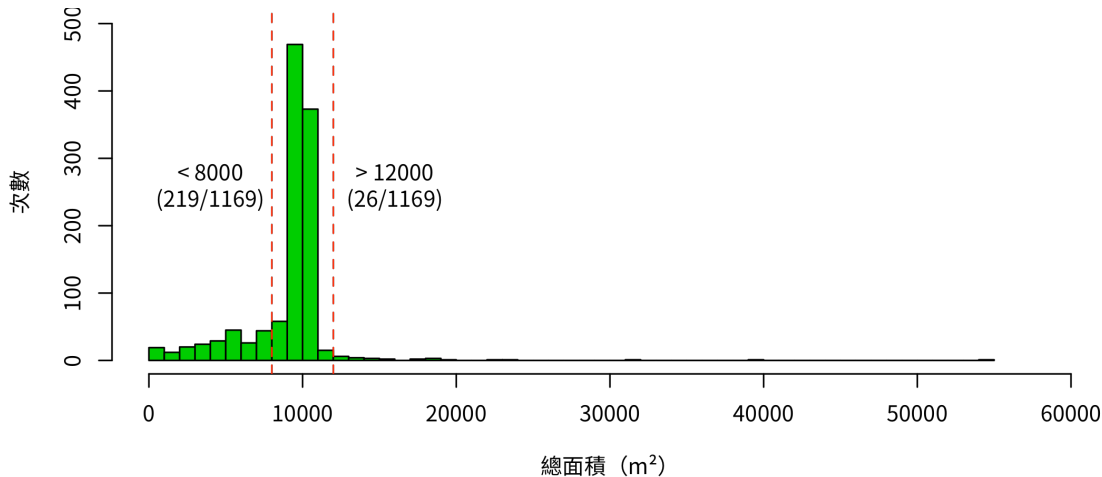
求解侯選之 TBAF 建構分數

評價不同侯選 TBAF 構造分數

預測結果探討

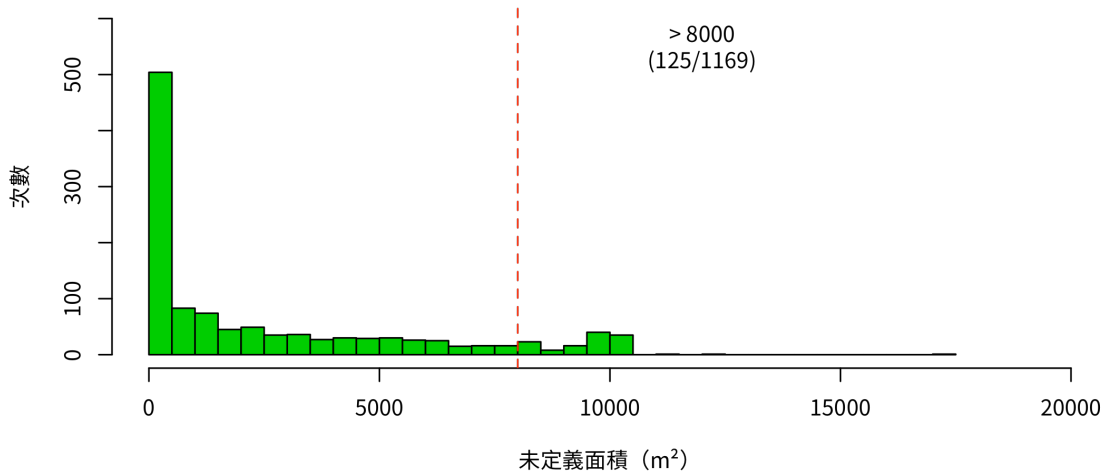
## 樣點總面積偏離 $10,000\text{m}^2$

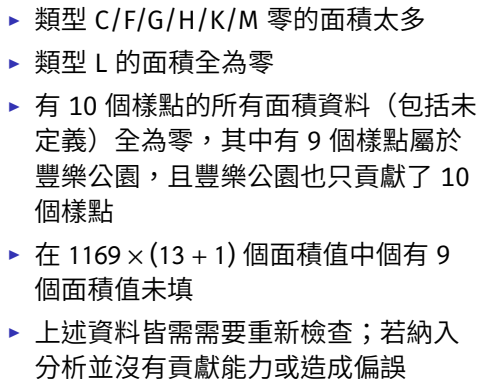
- ▶ 1169 個樣點中有 245 個樣點之總面積（13 已知類型 + 1 未知類型）小於  $8,000\text{ m}^2$  或大於  $12,000\text{ m}^2$
- ▶ 樣點總面積應該都在  $10,000\text{m}^2$  附近才對，需要重新檢查



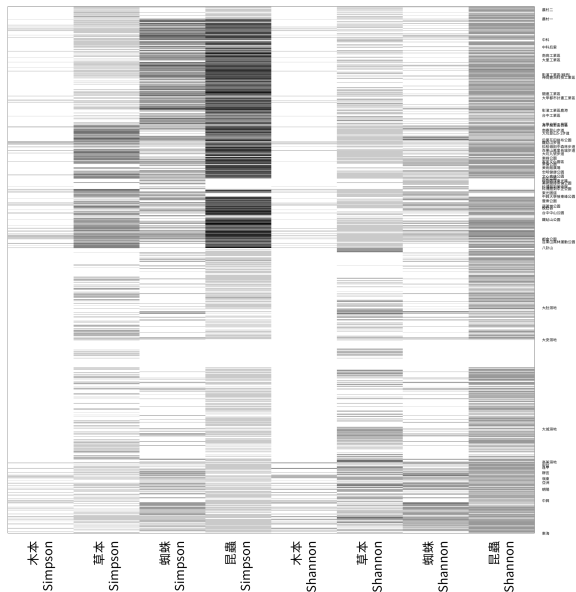
## 過大的未定義面積

- ▶ 1169 個樣點中有 125 個樣點之未定義面積大於  $8,000 \text{ m}^2$
- ▶ 未定義面積過大的樣點對之後的分析可能造成偏誤，需要重新檢查

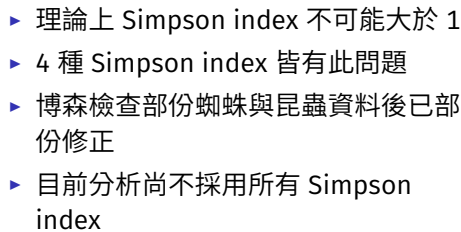




# 綜觀生態資料



- ▶ 三個濕地的多樣性資料有太多零，不存在足夠變異
- ▶ 昆蟲或蜘蛛的 Simpson 與 Shannon 指數之大小趨勢明顯相反
- ▶ Shannon 指數的變異比較均勻，而 Simpson 的變數在昆蟲明顯較大；之後分析若採用昆蟲 Simpson 指數的原始數字，可能會將昆蟲資料過度放大
- ▶ 木本資料太多零，排除之





# 大綱

現有資料檢查

求解侯選之 TBAF 建構分數

評價不同侯選 TBAF 構造分數

預測結果探討

## 面積資料的挑選與標準化規則

1. 去除存在任何缺失資料的樣點
2. 去除總面積（包括未定義面積）小於 5,000 m<sup>2</sup> 或大於 15,000 m<sup>2</sup> 的樣點
3. 去除未定義面積大於 5,000 m<sup>2</sup> 的樣點（剩 793 個樣點）
4. 以每樣點為單位，將已定義類型面積和調整為 1（即不考慮未定義面積的資訊），再減去平均以中心化

$$\frac{g_{jk}}{\text{sum}(g_{j\cdot})} - \text{mean}\left(\frac{g_{jk}}{\text{sum}(g_{j\cdot})}\right)$$

5. 以一次項，二次項及一次項間乘積為解釋變數；太多面積為 0 的變數直接手動移除
6. 最後剩下 793 × 27 個面積資料，稱為 G

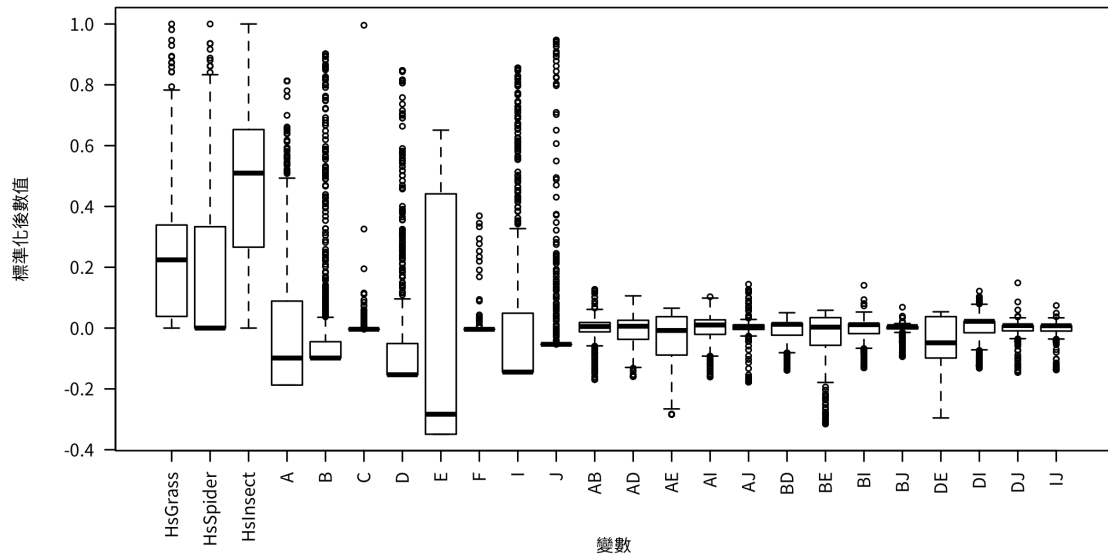
## 生態資料的挑選與標準化規則

1. 我選用 3 種 Shannon 指數（木本資料不考慮），以每種指數為單位，將 3 項指數最大值為 1 且最小值為 0 進行標準化

$$\frac{b_{jk}}{\max(b_{.k})}$$

2. 因為木本 Shannon 指數的零太多且與其它變數相關低，不再考慮

# 挑選與調整後的所有變數



## 以複迴歸求解侯選之 TBAF 建構分數

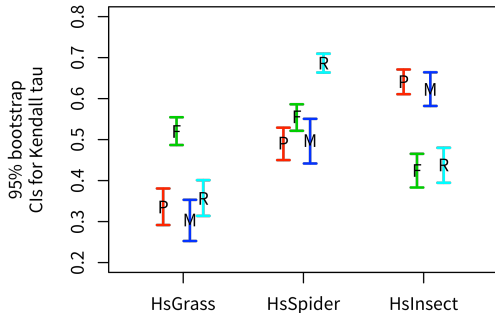
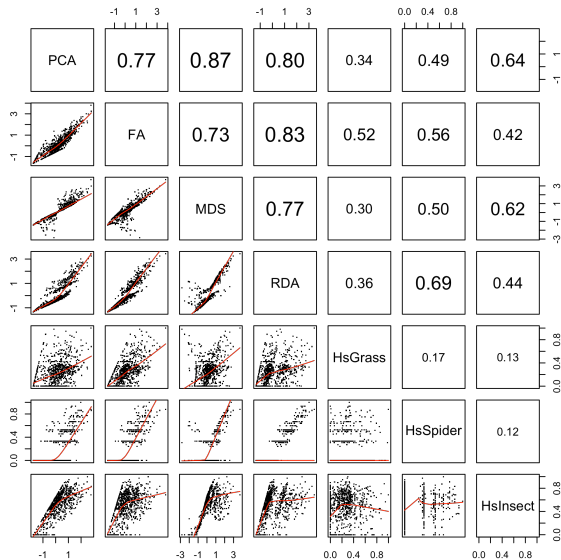


1. 利用 PCA、FA(1)、MDS(1)、RDA 萃取出生態資料的第一空間分數之標準常態分數  $S_i$
2. 以  $S_i$  為反應變數，面積分數  $G$  為解釋變數，建立 4 個線性複迴歸模型

$$M_i : S_i = \beta_i^T G + E_i$$

3. 採用二種自動化變數挑選方法
  - 3.1 以僅包括一次項為起點，每次加入或刪去一項變數，直到找到最小 AIC（以下標 1 表示）
  - 3.2 以包括一次項及交互作用項為起點，每次加入或刪去一項變數，直到找到最小 BIC（以下標 F 表示）
4. 強迫保留 A D E F I 五類面積，使各表面型態組皆有資料，不然會「變數沒選到 D 結果東海都第一名」
5. 得到  $4 \times 2 = 8$  個迴歸式

# 生態資料降維



- ▶ 4 種降維方法間呈高度相關
- ▶ 唯有 FA 與草本多樣性的相關較高
- ▶ RDA 強調蜘蛛多樣性，其它方法也不太差
- ▶ PCA 和 MDS 與昆蟲多樣性的相關較高

# 複迴歸及變數選擇後的原始迴歸係數（侯選之 TBAF 建構分數）

	PCA <sub>1</sub>	PCA <sub>F</sub>	FA <sub>1</sub>	FA <sub>F</sub>	MDS <sub>1</sub>	MDS <sub>F</sub>	RDA <sub>1</sub>	RDA <sub>F</sub>
Int.	0.045	0.045	0.050	0.050	-0.010	-0.040	0.045	0.079
A	0.336	0.336	0.218	0.218	-0.171	-0.202	0.350	1.333
A × B								1.404
A × D	-2.167	-2.167			-2.835	-2.833		
A × E	1.753	1.753	1.503	1.503	1.107	1.082	1.216	1.326
A × I	2.671	2.671	2.457	2.457			3.014	3.117
A × J			-2.161	-2.161	-1.799	-1.897	-2.117	-1.933
B								1.329
B × E								1.142
C								1.640
D	0.900	0.900	1.036	1.036	0.893	0.860	1.325	2.190
D × I	2.551	2.551			1.365	1.369		
E	-0.016	-0.016	-0.090	-0.090	-0.044	-0.085	-0.084	0.890
F	-0.324	-0.324	-0.187	-0.187	-0.413	-0.499	0.077	1.328
I	0.666	0.666	0.311	0.311	0.282	0.007	0.493	1.378
I × J						-4.881		
J	0.963	0.963	0.745	0.745	0.752		0.743	1.647

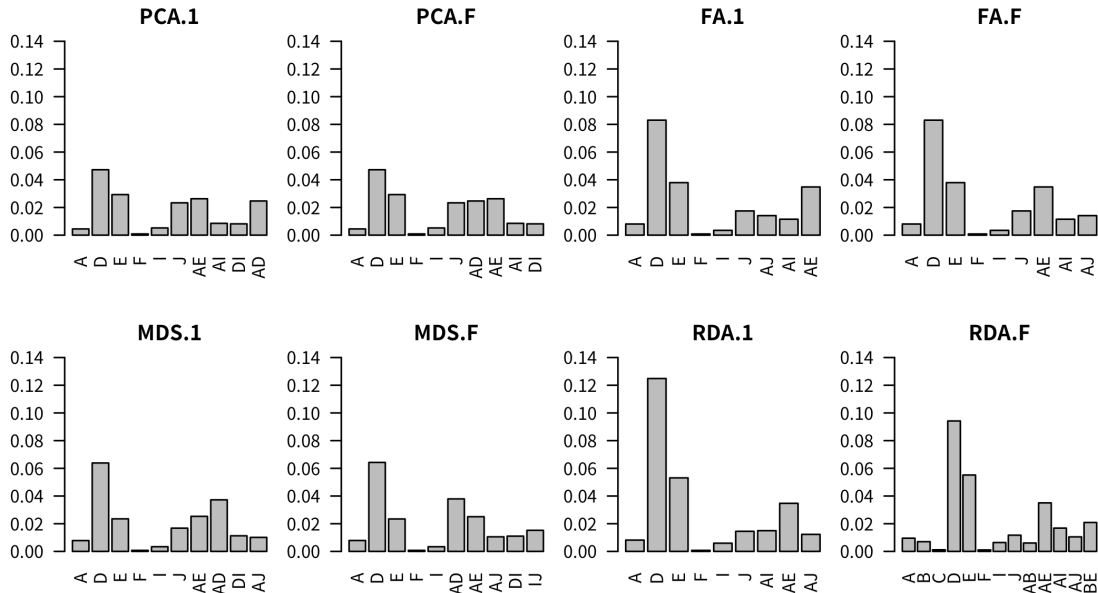
## 解釋變數之 VIF

	PCA <sub>1</sub>	PCA <sub>F</sub>	FA <sub>1</sub>	FA <sub>F</sub>	MDS <sub>1</sub>	MDS <sub>F</sub>	RDA <sub>1</sub>	RDA <sub>F</sub>
<i>A</i>	4.4	4.4	3.4	3.4	4.2	4.1	3.4	28.3
<i>A</i> × <i>B</i>								1.7
<i>A</i> × <i>D</i>	4.1	4.1			4.1	4.2		
<i>A</i> × <i>E</i>	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.3
<i>A</i> × <i>I</i>	1.8	1.8	1.8	1.8			1.8	1.8
<i>A</i> × <i>J</i>			1.8	1.8	1.8	1.8	1.8	1.9
<i>B</i>								27.0
<i>B</i> × <i>E</i>								3.5
<i>C</i>								1.7
<i>D</i>	4.8	4.8	2.4	2.4	4.7	4.6	2.4	44.7
<i>D</i> × <i>I</i>	2.4	2.4			2.4	2.4		
<i>E</i>	4.1	4.1	4.1	4.1	4.2	3.9	4.1	78.0
<i>F</i>	1.1	1.1	1.1	1.1	1.1	1.1	1.1	2.0
<i>I</i>	3.2	3.2	2.5	2.5	2.7	2.5	2.5	34.2
<i>I</i> × <i>J</i>						2.4		
<i>J</i>	1.5	1.5	2.3	2.3	2.3		2.3	18.2

RDA<sub>F</sub> 有嚴重共線性，應不考慮



# 解釋變數的相對 $R^2$ 貢獻量



# 大綱

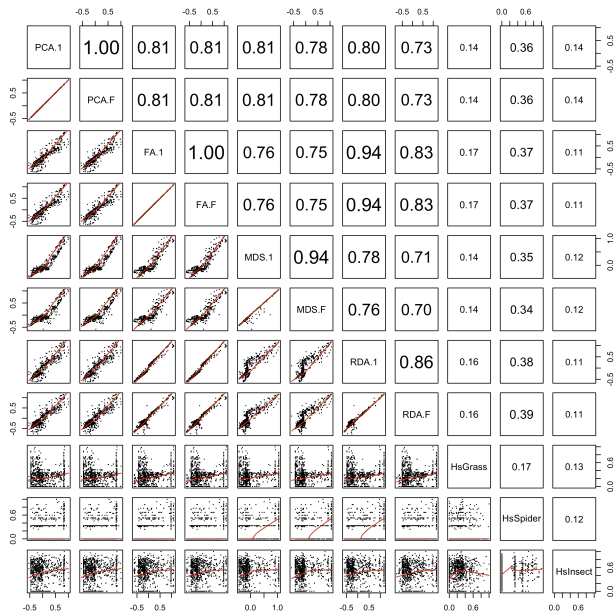
現有資料檢查

求解侯選之 TBAF 建構分數

評價不同侯選 TBAF 構造分數

預測結果探討

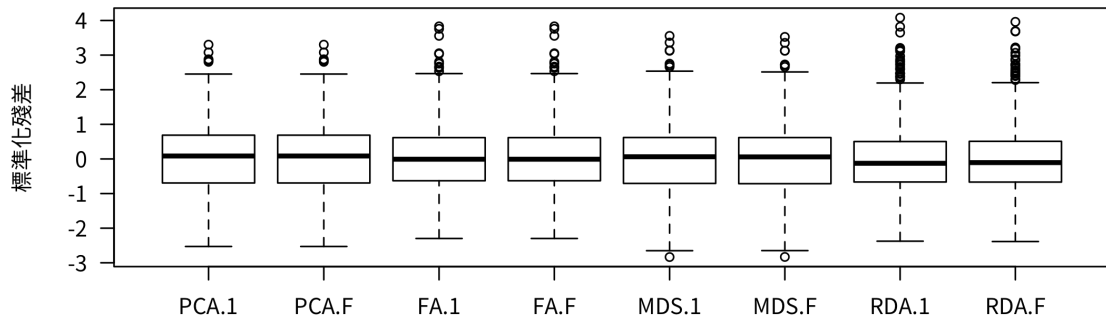
# 生態分數的預測值與實際值的相關性



- ▶ 8 種模型對蜘蛛多樣性指數的預測能力最好，但也在 Kendall rank correlation coefficient = 0.37 左右
- ▶ 8 種模型的效果似乎十分相似

# 候選模型之預測能力

	PCA <sub>1</sub>	PCA <sub>F</sub>	FA <sub>1</sub>	FA <sub>F</sub>	MDS <sub>1</sub>	MDS <sub>F</sub>	RDA <sub>1</sub>	RDA <sub>F</sub>
# IV	10.000	10.000	9.000	9.000	10.000	10.000	9.000	13.000
NRMSE	0.906	0.906	0.888	0.888	0.894	0.894	0.854	0.851
$R^2$	0.178	0.178	0.211	0.211	0.200	0.199	0.269	0.276
Adjusted $R^2$	0.168	0.168	0.202	0.202	0.190	0.189	0.261	0.264
Kendall tau	0.278	0.278	0.303	0.303	0.272	0.270	0.335	0.339
AIC/1000	2.118	2.118	2.083	2.083	2.096	2.097	2.023	2.024
BIC/1000	2.174	2.174	2.135	2.135	2.152	2.153	2.074	2.094



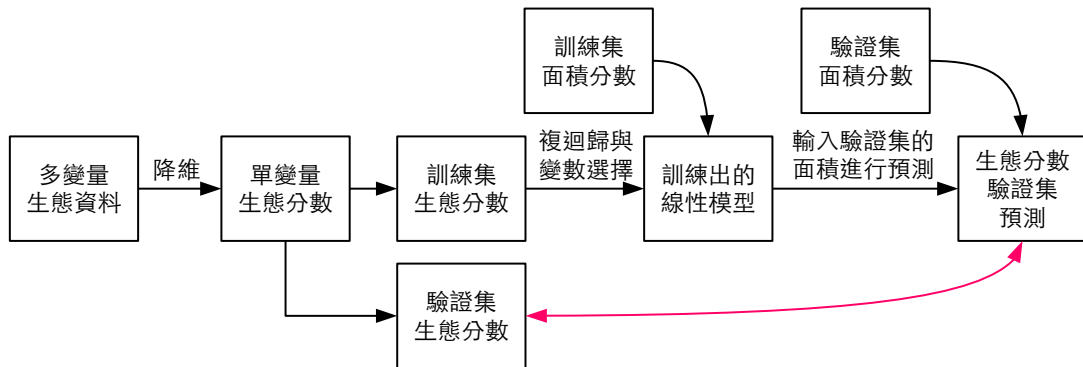
## 十折交插驗證

1. 將資料分 10 等份，其中 9 份為訓練資料集，另 1 份為驗證資料集
2. 以訓練資料集 DV 與 IV 建模  $M$
3. 以驗證資料集 IV 丟入模型  $M$  取得訓練資料集的預測結果
4. 重覆 2-3 步驟 10 次，使所有資料都被驗證 1 次，取得所有預測結果  $\hat{y}$
5. 驗證資料集 DV 為  $y$ ，求 normalized root-mean-square error (NRMSE)

$$\sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y)^2}{n}} / \text{SD}(y)$$

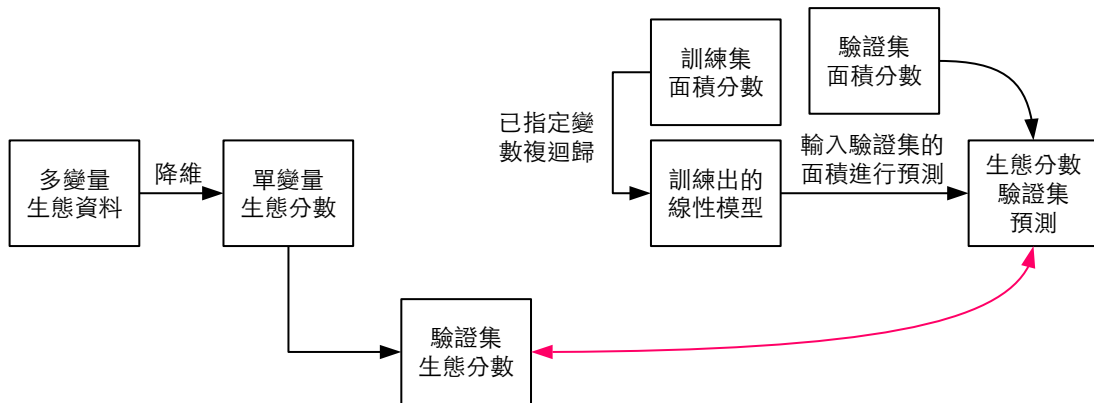
若  $\text{NRMSE} = 1$  表示預測值與驗證值的差之幾何平均佔 1 個  $y$  的標準偏差

# 驗證迴歸及自動化變數挑選方法



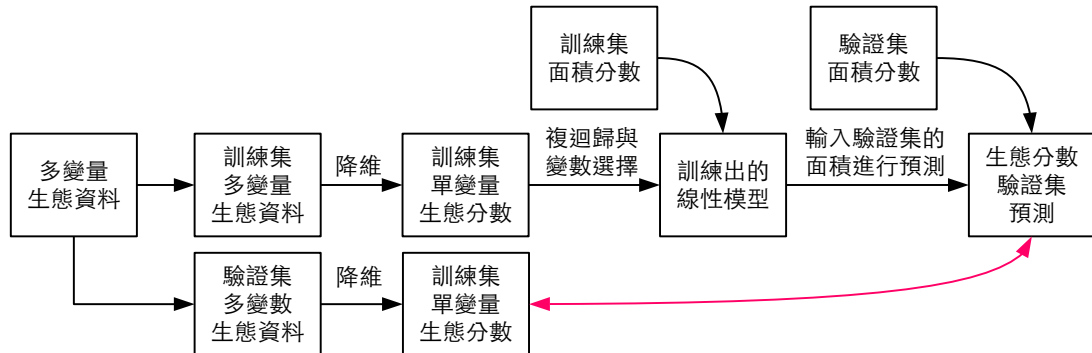
	PCA <sub>1</sub>	PCA <sub>F</sub>	FA <sub>1</sub>	FA <sub>F</sub>	MDS <sub>1</sub>	MDS <sub>F</sub>	RDA <sub>1</sub>	RDA <sub>F</sub>
NRMSE	0.929	0.929	0.902	0.904	0.918	0.922	0.873	0.874
Pearson r	0.375	0.374	0.431	0.428	0.400	0.393	0.489	0.488
Kendall tau	0.236	0.235	0.267	0.262	0.225	0.216	0.297	0.297

# 驗證面積資料變數組合



	PCA <sub>1</sub>	PCA <sub>F</sub>	FA <sub>1</sub>	FA <sub>F</sub>	MDS <sub>1</sub>	MDS <sub>F</sub>	RDA <sub>1</sub>	RDA <sub>F</sub>
NRMSE	0.924	0.923	0.902	0.902	0.911	0.908	0.867	0.868
Pearson r	0.382	0.385	0.432	0.431	0.412	0.420	0.498	0.497
Kendall tau	0.240	0.242	0.269	0.256	0.231	0.235	0.309	0.306

# 綜合比較不同求解方式



	$PCA_1$	$PCA_F$	$FA_1$	$FA_F$	$MDS_1$	$MDS_F$	$RDA_1$	$RDA_F$
NRMSE	0.935	0.938	0.928	0.931	1.138	1.137	0.937	0.945
Pearson r	0.361	0.356	0.380	0.378	-0.094	-0.088	0.378	0.368
Kendall tau	0.224	0.222	0.233	0.232	-0.048	-0.046	0.227	0.218



## 綜合比較結論

- ▶ PCA 與 RDA 降維後，無論以 AIC 或 BIC 法，其預測值與實際生態資料的相關性較高且交叉驗證的誤差較低
- ▶ BIC 的結果比 AIC 還精簡（留下的解釋變數少了約一半），且交叉驗證結果指出二者都沒有明顯過度配適（overfitting）問題
- ▶ RDA 在降維時一併考慮了面積資料，但目前資料並不是非常全面性的；若以 RDA 結果為 TBAF 構造分數要更小心「以外插法求解往往是危險的」的情況
- ▶ MDS 的特性是著重不相似性（概念上類似 4 種多樣性指數的組成比例在樣點間的差距）
- ▶ FA 太過偏好昆蟲多樣性指數，且在每次訓練過程過度偏好訓練子集資料，驗證能力較差
- ▶ 不同的迴歸解釋變數組合以 PCA 與 RDA 驗證能力較佳，但其實二者並不是非常相似；原因可能是資料特性（共線性）或需配合專業知識辨別何者較合理

## 目前我認為最好的 TBAF 建構係數組

按目前結果，我認為  $RDA_{AIC}$  的係數可以發展為 TBAF 建構分數。A-J 的單位為比例 (0-1)

	FA <sub>1</sub> 已中心化	RDA <sub>1</sub> 已中心化	FA <sub>1</sub> 未中心化	RDA <sub>1</sub> 未中心化
Int.	0.04972	0.04469	-0.0594563053	-0.1660450534
<i>A</i>	0.21781	0.34981	-0.5471588221	-0.3980290910
<i>A</i> × <i>E</i>	1.50333	1.21649	1.5033315634	1.2164858262
<i>A</i> × <i>I</i>	2.45668	3.01409	2.4566805629	3.0140897708
<i>A</i> × <i>J</i>	-2.16068	-2.11685	-2.1606755410	-2.1168453546
<i>D</i>	1.03635	1.32543	1.0363492651	1.3254339384
<i>E</i>	-0.09021	-0.08447	-0.3719367565	-0.3124394208
<i>F</i>	-0.18668	0.07723	-0.1866814866	0.0772270472
<i>I</i>	0.31146	0.49309	-0.1489308375	-0.0717603507
<i>J</i>	0.74532	0.74315	1.1502336883	1.1398493291

# 大綱

現有資料檢查

求解侯選之 TBAF 建構分數

評價不同侯選 TBAF 構造分數

預測結果探討

# FA<sub>1</sub> 預測 1-20 樣點 (對草本敏感)

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
朝陽	5	0	0	0	0	0	0	0	0	5658	0	10	4	10	2	1
朝陽	9	1003	0	0	0	0	0	0	0	6358	0	7	1	2	4	1
朝陽	14	0	0	0	0	0	0	0	0	5046	0	4	10	5	9	1
中台	7	0	0	0	0	0	0	0	0	5361	0	10	5	6	3	1
中台	8	0	0	0	0	0	0	0	0	5942	0	3	6	5	4	1
鐵砧山公園	14	0	0	0	0	0	0	0	0	6273	0	10	7	7	3	1
客家文化園區	8	0	0	0	0	0	0	0	0	5001	0	10	8	10	10	1
大坑登山 5-1 步道	2	0	0	0	0	0	0	0	0	7777	0	10	7	10	4	1
朝陽	8	673	0	0	0	42	0	0	0	6850	0	10	5	3	4	9
大坑登山 5-1 步道	3	0	0	16	0	0	0	0	0	8321	0	10	7	10	3	10
大坑登山 5-1 步道	4	0	0	33	0	0	0	0	0	9839	0	10	7	10	5	11
朝陽	6	565	0	0	0	337	0	0	0	7413	0	10	5	4	5	12
中台	5	953	0	0	0	342	0	0	0	5944	0	1	6	2	4	13
朝陽	11	1154	0	0	0	559	0	0	0	7634	0	10	4	5	8	14
中台	6	408	0	0	0	654	0	0	0	7042	0	10	9	2	7	15
朝陽	13	4597	0	0	41	244	0	0	0	5114	0	10	6	6	3	16
朝陽	12	317	0	0	2	1167	0	0	0	8255	0	10	1	10	2	17
大坑登山 5-1 步道	1	0	0	109	0	0	0	0	0	8356	0	10	7	10	6	18
朝陽	15	3939	0	0	76	144	0	0	0	5644	0	6	6	4	4	19
大坑登山 5-1 步道	5	31	0	192	0	0	0	0	0	8898	0	10	7	7	3	20

FA<sub>1</sub> 預測 21-40 樣點

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
朝陽	7	1670	0	0	0	1972	0	0	0	6335	0	10	10	5	9	21
大坑登山 5-1 步道	6	0	0	320	0	0	0	0	0	7063	0	10	7	10	6	22
中台	1	4096	0	0	0	485	0	0	0	574	0	10	7	10	5	23
東海	3	237	0	0	0	8471	0	0	0	0	0	10	7	10	4	24
東海	5	4113	0	0	0	6032	0	0	0	0	0	10	10	10	2	24
東海	8	4782	0	0	0	5231	0	0	0	0	0	7	8	5	4	24
東海	9,11	3172	0	0	0	6838	0	0	0	0	0	10	4	6	6	24
東海	12	4907	0	0	0	5100	0	0	0	0	0	10	4	6	8	24
東海	17	3732	0	0	0	6259	0	0	0	0	0	10	9	8	4	24
東海	24	3799	0	0	0	6146	0	0	0	0	0	10	8	7	6	24
東海	26,27	3259	0	0	0	6757	0	0	0	0	0	10	10	3	4	24
東海	30	1988	0	0	0	8063	0	0	0	0	0	7	10	2	4	24
東海	31	1653	0	0	0	8354	0	0	0	0	0	7	10	4	5	24
東海	32	2454	0	0	0	7535	0	0	0	0	0	8	3	4	4	24
東海	33,34	2225	0	0	0	7818	0	0	0	0	0	10	5	7	4	24
東海	35	3323	0	0	0	6651	0	0	0	0	0	7	9	10	5	24
東海	36,44	3581	0	0	0	6364	0	0	0	0	0	10	5	3	5	24
東海	37	2543	0	0	0	7508	0	0	0	0	0	5	7	5	6	24
東海	38	3292	0	0	0	6645	0	0	0	0	0	10	10	6	3	24
東海	39,40	34	0	0	0	10025	0	0	0	0	0	10	5	2	5	24

# FA<sub>1</sub> 預測倒數 1-20 樣點

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
彰濱工業區 (線西)	34	1415	7615	0	0	0	0	0	0	970	0	10	10	10	5	793
台中工業區	2	4439	4657	0	0	0	0	0	0	885	0	10	10	10	10	792
大里工業區	2	3986	4885	0	0	0	0	0	0	263	0	10	10	10	5	791
台中工業區	8	4774	4194	0	0	0	0	0	0	1046	0	10	8	10	3	790
大甲幼獅工業區	1	4578	1936	0	0	0	0	0	0	0	0	10	9	10	6	779
彰濱工業區鹿港	2	2605	7395	0	0	0	0	0	0	0	0	10	10	10	6	779
關連工業區	17	4573	5427	0	0	0	0	0	0	0	0	10	10	10	8	779
關連工業區	20	4869	5131	0	0	0	0	0	0	0	0	10	10	10	5	779
關連工業區	23	4594	5406	0	0	0	0	0	0	0	0	10	10	10	5	779
關連工業區	33	2469	7531	0	0	0	0	0	0	0	0	10	10	10	6	779
關連工業區	34	595	9405	0	0	0	0	0	0	0	0	10	10	10	5	779
神岡豐洲科技工業區	4	4585	10000	0	0	0	0	0	0	0	0	10	8	10	3	779
彰濱工業區 (線西)	27	4747	5253	0	0	0	0	0	0	0	0	10	9	10	5	779
農村一	3	3726	6275	0	0	0	0	0	0	0	0	10	10	10	5	779
農村一	4	865	9101	0	0	0	0	0	0	0	0	10	8	10	3	779
彰濱工業區 (線西)	37	3955	3848	0	0	0	0	0	0	1394	0	10	10	10	8	778
中科	4	0	7270	0	0	0	0	0	0	2730	0	10	10	10	7	777
台中工業區	10	3805	4647	0	0	0	0	295	0	1298	0	10	10	10	9	776
神岡豐洲科技工業區	6	2150	3199	0	0	0	0	0	0	1224	0	10	9	10	5	775
中科	23	4429	5413	0	0	0	180	0	0	0	0	10	8	10	3	774

# FA<sub>1</sub> 預測倒數 21-40 樣點

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
台中工業區	3	3277	4184	0	0	0	0	0	0	1785	0	10	10	10	2	773
中科	20	4785	3852	0	0	0	0	1111	0	0	0	8	9	10	10	772
彰濱工業區鹿港	27	846	7569	1585	0	0	0	0	0	0	0	10	9	10	4	771
台中工業區	11	4053	4786	0	0	0	500	0	0	657	0	10	6	10	9	770
中科	42	3326	5609	0	0	0	412	655	0	0	0	10	7	7	7	769
中科	21	3331	4735	0	0	0	223	1716	0	0	0	8	8	7	5	768
大里工業區	10	3623	3946	294	0	0	0	0	0	2137	0	10	9	10	6	767
農村一	7	0	7295	0	0	0	0	0	0	0	0	10	8	10	4	766
彰濱工業區(線西)	19	0	6075	995	0	0	0	0	0	2930	0	10	9	10	2	765
高美濕地	3	0	0	0	0	0	9997	0	0	0	0	10	2	10	3	697
高美濕地	17	0	0	0	0	0	10001	0	0	0	0	10	5	10	10	697
高美濕地	20	0	0	0	0	0	9995	0	0	0	0	10	10	10	10	697
高美濕地	21	0	0	0	0	0	9995	0	0	0	0	10	8	10	10	697
高美濕地	22	1966	0	0	0	0	8029	0	0	0	0	10	10	10	10	697
高美濕地	28	0	0	0	0	0	9999	0	0	0	0	10	8	10	9	697
高美濕地	31	4318	0	0	0	0	5680	0	0	0	0	10	7	10	6	697
高美濕地	38	0	0	0	0	0	9988	0	0	0	0	10	9	10	4	697
高美濕地	41	0	0	0	0	0	9995	0	0	0	0	10	2	10	10	697
高美濕地	51	548	0	0	0	0	4721	0	0	0	0	10	9	10	9	697
大城濕地	7	0	0	0	0	0	10000	0	0	0	0	10	7	10	10	697

# RDA<sub>1</sub> 預測 1-20 樣點 (對蜘蛛敏感)

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
東海	3	237	0	0	0	8471	0	0	0	0	0	10	7	10	4	1
東海	5	4113	0	0	0	6032	0	0	0	0	0	10	10	10	2	1
東海	8	4782	0	0	0	5231	0	0	0	0	0	7	8	5	4	1
東海	9,11	3172	0	0	0	6838	0	0	0	0	0	10	4	6	6	1
東海	12	4907	0	0	0	5100	0	0	0	0	0	10	4	6	8	1
東海	17	3732	0	0	0	6259	0	0	0	0	0	10	9	8	4	1
東海	24	3799	0	0	0	6146	0	0	0	0	0	10	8	7	6	1
東海	26,27	3259	0	0	0	6757	0	0	0	0	0	10	10	3	4	1
東海	30	1988	0	0	0	8063	0	0	0	0	0	7	10	2	4	1
東海	31	1653	0	0	0	8354	0	0	0	0	0	7	10	4	5	1
東海	32	2454	0	0	0	7535	0	0	0	0	0	8	3	4	4	1
東海	33,34	2225	0	0	0	7818	0	0	0	0	0	10	5	7	4	1
東海	35	3323	0	0	0	6651	0	0	0	0	0	7	9	10	5	1
東海	36,44	3581	0	0	0	6364	0	0	0	0	0	10	5	3	5	1
東海	37	2543	0	0	0	7508	0	0	0	0	0	5	7	5	6	1
東海	38	3292	0	0	0	6645	0	0	0	0	0	10	10	6	3	1
東海	39,40	34	0	0	0	10025	0	0	0	0	0	10	5	2	5	1
東海	42	2145	0	0	0	7873	0	0	0	0	0	8	3	5	6	1
東海	43	2728	0	0	0	7265	0	0	0	0	0	10	6	6	8	1
東海	45	2685	0	0	0	7366	0	0	0	0	0	6	5	6	7	1



# RDA<sub>1</sub> 預測 21-40 樣點

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
東海	46	2196	0	0	0	7846	0	0	0	0	0	10	2	6	6	1
東海	47	2664	0	0	0	7378	0	0	0	0	0	10	2	5	6	1
東海	48	3861	0	0	0	6160	0	0	0	0	0	6	4	7	4	1
東海	50	2903	0	0	0	7077	0	0	0	0	0	10	3	7	6	1
東海	51	277	0	0	0	9706	0	0	0	0	0	10	7	10	5	1
東海	52	1265	0	0	0	8724	0	0	0	0	0	10	8	3	6	1
東海	53,54	2861	0	0	0	7180	0	0	0	0	0	10	9	5	4	1
東海	56	3133	0	0	0	6848	0	0	0	0	0	10	7	5	7	1
東海	57	2476	0	0	0	7513	0	0	0	0	0	6	9	4	9	1
東海	58,59	396	0	0	0	9607	0	0	0	0	0	10	8	6	5	1
東海	60,61	3094	0	0	0	6921	0	0	0	0	0	10	8	6	7	1
東海	67	2602	0	0	0	7382	0	0	0	0	0	3	7	7	3	1
東海	69	0	0	0	0	9989	0	0	0	0	0	10	8	8	4	1
東海	70	62	0	0	0	9980	0	0	0	0	0	10	10	4	6	1
東海	71,74	0	0	0	0	9972	0	0	0	0	0	10	8	5	7	1
東海	72	1179	0	0	0	8799	0	0	0	0	0	6	8	5	4	1
東海	73	955	0	0	0	9087	0	0	0	0	0	4	10	5	5	1
東海	75,77	395	0	0	0	9650	0	0	0	0	0	10	8	5	5	1
東海	76	706	0	0	0	6875	0	0	0	0	0	4	10	3	4	1
東海	78	1173	0	0	0	8875	0	0	0	0	0	10	9	10	7	1

# RDA<sub>1</sub> 預測倒數 1-20 樣點

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
台中工業區	2	4439	4657	0	0	0	0	0	0	885	0	10	10	10	10	793
彰濱工業區 (線西)	34	1415	7615	0	0	0	0	0	0	970	0	10	10	10	5	792
台中工業區	8	4774	4194	0	0	0	0	0	0	1046	0	10	8	10	3	791
大里工業區	2	3986	4885	0	0	0	0	0	0	263	0	10	10	10	5	790
彰濱工業區 (線西)	37	3955	3848	0	0	0	0	0	0	1394	0	10	10	10	8	789
中科	4	0	7270	0	0	0	0	0	0	2730	0	10	10	10	7	788
大甲幼獅工業區	1	4578	1936	0	0	0	0	0	0	0	0	10	9	10	6	777
彰濱工業區鹿港	2	2605	7395	0	0	0	0	0	0	0	0	10	10	10	6	777
關連工業區	17	4573	5427	0	0	0	0	0	0	0	0	10	10	10	8	777
關連工業區	20	4869	5131	0	0	0	0	0	0	0	0	10	10	10	5	777
關連工業區	23	4594	5406	0	0	0	0	0	0	0	0	10	10	10	5	777
關連工業區	33	2469	7531	0	0	0	0	0	0	0	0	10	10	10	6	777
關連工業區	34	595	9405	0	0	0	0	0	0	0	0	10	10	10	5	777
神岡豐洲科技工業區	4	4585	10000	0	0	0	0	0	0	0	0	10	8	10	3	777
彰濱工業區 (線西)	27	4747	5253	0	0	0	0	0	0	0	0	10	9	10	5	777
農村一	3	3726	6275	0	0	0	0	0	0	0	0	10	10	10	5	777
農村一	4	865	9101	0	0	0	0	0	0	0	0	10	8	10	3	777
神岡豐洲科技工業區	6	2150	3199	0	0	0	0	0	0	1224	0	10	9	10	5	776
台中工業區	10	3805	4647	0	0	0	0	295	0	1298	0	10	10	10	9	775
台中工業區	3	3277	4184	0	0	0	0	0	0	1785	0	10	10	10	2	774

# RDA<sub>1</sub> 預測倒數 21-40 樣點

Location	Spot	X	A	B	C	D	E	F	I	J	M	木	草	蛛	蟲	預
中科	23	4429	5413	0	0	0	180	0	0	0	0	10	8	10	3	773
彰濱工業區鹿港	27	846	7569	1585	0	0	0	0	0	0	0	10	9	10	4	772
台中工業區	11	4053	4786	0	0	0	500	0	0	657	0	10	6	10	9	771
高美濕地	3	0	0	0	0	0	9997	0	0	0	0	10	2	10	3	703
高美濕地	17	0	0	0	0	0	10001	0	0	0	0	10	5	10	10	703
高美濕地	20	0	0	0	0	0	9995	0	0	0	0	10	10	10	10	703
高美濕地	21	0	0	0	0	0	9995	0	0	0	0	10	8	10	10	703
高美濕地	22	1966	0	0	0	0	8029	0	0	0	0	10	10	10	10	703
高美濕地	28	0	0	0	0	0	9999	0	0	0	0	10	8	10	9	703
高美濕地	31	4318	0	0	0	0	5680	0	0	0	0	10	7	10	6	703
高美濕地	38	0	0	0	0	0	9988	0	0	0	0	10	9	10	4	703
高美濕地	41	0	0	0	0	0	9995	0	0	0	0	10	2	10	10	703
高美濕地	51	548	0	0	0	0	4721	0	0	0	0	10	9	10	9	703
大城濕地	7	0	0	0	0	0	10000	0	0	0	0	10	7	10	10	703
大城濕地	13	1306	0	0	0	0	8687	0	0	0	0	10	10	10	8	703
大城濕地	19	1390	0	0	0	0	8610	0	0	0	0	10	9	10	3	703
大城濕地	22	0	0	0	0	0	10006	0	0	0	0	10	10	7	6	703
大城濕地	32	4081	0	0	0	0	5925	0	0	0	0	10	7	10	10	703
大城濕地	36	0	0	0	0	0	10000	0	0	0	0	10	5	10	6	703
大城濕地	45	0	0	0	0	0	9994	0	0	0	0	10	10	10	5	703

## 筆記、其它解決方案及未來工作

- ▶ 不考慮未定義面積（目前辦法）及考慮未定義面積的預測能力差不多
- ▶ 嘗試過合併（加總）太多零的面積類別，但沒有效果
- ▶ 交叉驗證的 NRMSE 之間不能直接比較
- ▶ 需要完全檢查生態資料
- ▶ 需要完全檢查怪異的面積資料
- ▶ 自動化變數選擇應該必須避免，但已有部份改善
- ▶ 模型預測效果不佳，可以再找新的建模方法（如二階式迴歸）、數據（如大樣區的總面積）、以別的「角度」採用面積資料（如採用  $B - A$ 、 $\frac{B}{A}$  之類的面積變數）