

應用模糊分類與馬可夫概型分析於長期機率預報

吳明進¹、賀介圭²

國立台灣大學大氣科學研究所 中央氣象局預報中心

摘要

本文介紹一種結合模糊分類與馬可夫概型分析作長期機率預報的方法，並將之應用於台北逐月氣溫和降水的長期機率預報。馬可夫概型分析提供進一步抽取長期預報資訊的一個方法，模糊分類解決了傳統分類法在臨界點的不連續問題。由 1961-90 逐月氣溫和降水三分類狀態轉移機率矩陣，顯示氣溫的變化有較明顯的持續性，而降水則否。另外應用此方法於 1991-95 年資料的後預報顯示台北逐月氣溫的預報有相當高的技術，降水則只達邊際技術。

一、前言

旬以上的長期預報之時間尺度已超過大氣確定預報理論上的可預報度範圍，所以必須採用統計的方法來進行。統計預報的方法相當多，主要有兩種類型的方法：一為時間序列法，這裡包括持續法週期法和動態時間序列模式法，其次為多變數統計和物理推斷，這包括類比法、回歸法和綜觀運動法。這些統計方法預報採有的有的是氣候因子的數值有的則是預報類型。不過對於一個長期預報作業產品的使用者來說，由於目前預報技術的能力限制，太精確的數值並沒有什麼意義，所以世界上所有的氣象中心氣溫和降水的預報作業皆以三或五分類的統計級性來表示。即將氣溫降水分成偏高、正常及偏低三者來表示。預報結果評估通常以氣候平均的機率分配當作基準，即若預報之準確率和氣候機率分配相當，作無技術即技術得分為 0。有的時候也以延續性預報當作評估基準，即延續性預報的技術得分為 0。執行預報的結果需較持續性預報準確度大才有預報技術。

其實氣候資料中有預報價值的資訊可能有超過平均統計特性或持續法者。這其中包括本文所要介紹經常用在社會科學或經濟學中的馬可夫概型分析。馬可夫概型分析是以時間序列的內部結構為出發點，應用多元時間序列與馬可夫過程理論從實測的時間序列中抽出隨

機過程的機率規律。馬可夫概型分析的概念在本文第二章第一節中做簡單介紹，詳細的內容請參考羅(1989)第九章。馬可夫概型分析應用到氣象者有 Mo & Ghil (1988)等。

有限狀態的馬可夫過程分析中必須定義狀態，由於氣候因子所表現的狀態判定值為一連續數值，所以若以固定的值做分類標準時則在臨界值產生落差。為解決這些問題本文引進模糊邏輯的概念，模糊數學的概念首先於 1965 年由美國自動控制學家 L.A.Zadeh 所提出。現在這學科以在各種不同領域有廣泛的應用，模糊數學在氣象上的應用有陳(1976)曾與陳(1980)等。模糊分類的概念在本文第二章第二節中做簡單介紹，詳細的內容請參考馮與樓(1991)。

2、方法

2.1 馬可夫概型分析和預報

馬可夫概型分析是以時間序列的內部結構為出發點，應用多時間序列與馬可夫過程理論。從實測的時間序列中抽取隨機過程的機率規律，馬可夫概型分析可以解決與動態系統有關的問題，它主要解決的問題是如果系統現在處於狀態 r (例如乾旱)從現在起經過 n

步轉移後系統處於狀態s(例如洪澇)的概律是多少,或如果某種狀態r現在的觀測值(譬如正距平)的比例為 P_r^0 ,則從現在起經過n個時間步後,處於某種狀態s(譬如負距平)觀測值的比例。而馬可夫過程則定義為一個隨機過程 $x(t)$ 在 τ ($\tau > t$)時刻時系統所處的狀態與時間t時所處的狀態有關而和t時間前的狀態無關。在動態序列模式可表示成爲一階自相關。

在馬可夫過程中系統狀態的轉移可用轉移概率矩陣表示即 $P = P_{ij} = P\{x(t) = j | x(t-1) = i\}$ P_{ij} 表示在時刻t-1時系統處於狀態i的條件下系統在時刻t處於狀態j的概率。

$$P = \begin{pmatrix} P_{11} & \dots & P_{1m} \\ \dots & \dots & \dots \\ P_{m1} & \dots & P_{mm} \end{pmatrix}$$

則若系統在時間0時狀態概率的分佈向量爲 $P^0 = (p_1^0, \dots, p_m^0)$ 則時間1時狀態概率分佈爲 $P^1 = (P)^T \cdot P^0$ 其中 $p_i^0 \geq 0$ 且 $\sum_{i=1}^m p_i^0 = 1$, T 表示矩陣轉置,時間2時狀態概率分佈 $P^2 = (P^*)^T \cdot P^1$ P^* 爲時間1至時間2的轉移概率矩陣。

2.2 模糊分類

目前中央氣象局對進行臺灣地區分區(北、中、南、東)旬、月與季氣溫及降水的長期預報作業。預報的標的爲依觀測值由大至小排列發生概率30%和70%劃分的偏高、正常和偏低的三分類。本研究中依此劃分法對台北的氣候要素的氣溫和降水分別逐月依三類定義其狀態,則各月氣溫狀態的範圍如表一所示。降水三態的範圍如表二所示。對於一個觀測值三態的劃分依古典邏輯概念必爲偏高、正常、偏低三者中之一,即其發生概率分佈必爲(1,0,0),(0,1,0),(0,0,1)三者其中之一。但依此分類介於臨界值的觀測其狀態概率分佈必產生落差。例如累積頻率值爲29.9%的觀測值,其狀態概率爲(1,0,0),但累積頻率值爲30.1%的觀測值其狀態概率則爲(0,1,0),殊爲不合理。爲解決這個問題,我們提出源於模糊邏輯的模糊分類的概念。

模糊邏輯,或稱爲多值邏輯,和古典邏輯不同處爲古典邏輯值定義邏輯真假值爲0或者1,即是或非兩者中必有其一,模糊邏輯則定義邏輯值可爲0至1的任何值。氣候要素的模糊分類是採觀測值依其極大值、極小值、中位值做規格化以定義其模糊狀態概率。以平均的氣溫爲例,首先將1961-90年月平均觀測值由小至大依序排列,每一級次的發生機率各爲 $1/(30-1)=0.034$ 。而後將觀測值換算成累積機率a,最大觀測值a=0,中位數值a=0.5,最小值a=1。若有多次(月)觀測之值相同,則級次和累積概率值以平均數表示。其次定義一個觀測值三個狀態出現頻數爲 (f_1, f_2, f_3) , (f_1, f_2, f_3) 定義爲:
若 $a \leq 0.1$ 則 $f_1 = 1, f_2 = 0, f_3 = 0$
若 $0.1 < a \leq 0.5$,則 $f_1 = 1 - f_2, f_2 = \frac{a-0.1}{0.4}, f_3 = 0$
若 $0.5 \leq a \leq 0.9$,則 $f_1 = 0, f_2 = 1 - f_3, f_3 = \frac{a-0.5}{0.4}$
若 $a \geq 0.9$,則 $f_1 = f_2 = 0, f_3 = 1$

依此定義最大10%觀測值之狀態頻率分佈向量爲(1,0,0),累積機率30%之觀測值爲(1/2,1/2,0),中位數爲(0,1,0),累積機率70%之觀測值爲(0,1/2,1/2),而最小10%觀測值之狀態頻率分佈爲(0,0,1)。而 $f_1 + f_2 + f_3 = 1$,如此將狀態發生頻數模糊化,可以保守原來極端值頻率的特性,且解決古典分類法所造成臨界值不連續的問題。在一次轉移過程,

若時間t-1時之頻率分佈向量爲 $F^0 = (f_1^0, f_2^0, f_3^0)$
而時間t時之頻率分佈向量爲 $F^1 = (f_1^1, f_2^1, f_3^1)$
則一次轉移模糊頻數矩陣爲 $F = (F^0)^T F^1 = P_y$

其次將每個模糊轉移頻數矩陣累加,可以得到模糊轉移頻率矩陣,然後將矩陣各元素除以各行元素之和即可得到轉移概率矩陣P。

3、結果

3.1 轉移概率矩陣

表三所示爲台北 1961-90 各月氣溫之狀態的轉移概率矩陣,如以轉移概率向量行中最大概率代表其狀態則知若台北一月高溫則二月也可能是高溫。若一月正常則二月也可能是正常。但若一月低溫則二月較有可能轉爲正常。若二月高溫則三月可能仍是高溫,二月若正常則三月可能也是正常,二月低溫三月可能轉爲正常。三

表一 以三類劃分法所訂的台北月平均氣溫(°c)的範圍

月份	一	二	三	四	五	六	七	八	九	十	十一	十二
偏高	>15.89	16.43	18.02	22.22	25.23	27.59	29.29	29.02	27.63	24.49	21.29	17.86
中值	15.55	15.65	17.35	21.63	24.85	27.00	28.90	28.75	27.27	24.12	20.85	17.53
偏低	<14.93	14.91	17.01	20.94	24.37	26.55	28.58	28.51	26.74	23.75	19.96	16.74

表二 以三類劃分法所訂的台北月降水(mm)的範圍

月份	一	二	三	四	五	六	七	八	九	十	十一	十二
偏高	>122.4	126.9	245.9	162.6	295.0	361.6	289.0	323.2	400.9	129.5	92.8	101.0
正常	74.3	102.7	146.3	129.7	229.7	256.5	185.5	289.8	284.2	96.4	70.9	69.5
偏低	<51.4	81.7	104.5	87.7	166.9	177.9	157.4	166.7	222.2	44.0	59.1	46.3

表三 台北全年逐月氣溫三狀態的一步(月)轉移概率矩陣 (X 1000)

U代表高於正常, N代表正常, B代表低於正常

	U	N	B	U	N	B	U	N	B	U	N	B			
MONTH = 1 TO 2				MONTH = 4 TO 5			MONTH = 7 TO 8			MONTH = 10 TO 11					
U	619	246	134	U	341	389	271	U	471	357	172	U	329	375	296
N	242	416	341	N	301	379	320	N	287	373	341	N	339	432	229
B	76	488	436	B	281	393	326	B	169	433	399	B	239	348	413
MONTH = 2 TO 3				MONTH = 5 TO 6				MONTH = 8 TO 9				MONTH = 11 TO 12			
U	448	324	228	U	337	463	200	U	247	457	296	U	357	433	210
N	281	422	297	N	297	378	325	N	363	393	244	N	287	404	310
B	198	403	399	B	290	318	392	B	296	306	398	B	278	326	395
MONTH = 3 TO 4				MONTH = 6 TO 7				MONTH = 9 TO 10				MONTH = 12 TO 1			
U	440	371	189	U	676	219	105	U	497	281	223	U	322	366	313
N	302	450	248	N	204	464	332	N	228	381	391	N	503	300	197
B	181	321	499	B	62	464	474	B	212	507	281	B	123	522	356

月若高溫, 正常或低溫則四月可能維持高溫, 正常或低溫。四月若高溫五月可能轉成正常。四月正常, 五月可能保持正常。四月若低溫五月可能轉成正常。五月若高溫則六月可能正常, 五月若正常, 六月可能正常。五月若低溫, 則六月可能保持低溫。六月若高溫則七月可能保持高溫, 六月若正常則七月可能保持正常。六月若低溫則七月可能保持低溫。七月若高溫則八月可能仍為高溫, 七月若正常則八月可能仍為正常, 七月若低溫則八月可能轉為正常。八月若高溫則九月可能為正常。八月若正常則九月可能轉為高溫。八月若低溫則九月可能維持低溫。九月若高溫則十月可能仍為高溫。九月若正常則十月可能轉成低溫, 九月若低溫則十月可能正常。十月若高溫則十一月可能轉為正常, 十月若正常, 則十一

月可能保持正常, 十月若低溫則十一月可能保持低溫。十一月若高溫則十二月可能正常, 十一月若正常則十二月可能正常, 十一月若低溫, 則十二月可能保持低溫。十二月若高溫則翌年一月可能正常, 十二月若正常翌年一月可能高溫, 十二月若低溫, 翌年一月可能變成正常。

若以最大轉移機率代表轉移狀態, 全年逐月氣溫之狀態的一步(月)最大可能轉移過程示意圖如圖 1 所示, 由於氣溫正常區間取觀測值累積機率 30%-70%之間佔所有發生機率的 0.4, 則根據馬可夫過程適應性原理, 可以發現多次轉換後轉移機率逐漸向(0.3, 0.4, 0.3)趨近, 不管初始值為何, 預報正常所佔比例最大。

圖 1. 台北逐月氣溫三狀態的一步最大可能轉移過程示意圖。

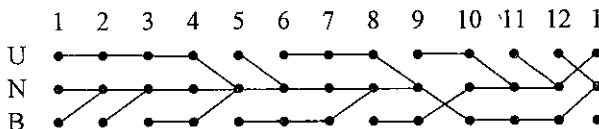
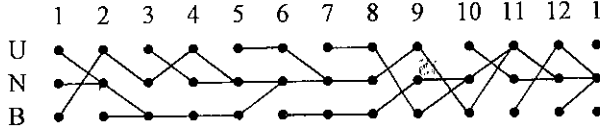


圖 2. 台北逐月降水三狀態的一步最大可能轉移過程示意圖。



表四所示為台北各月降水之狀態的一步轉移機率矩陣。

若台北降水一月偏多則二月可能偏多，一月若正常則二月可能正常，一月若偏少則二月可能也偏少。

二月若偏多則三月可能也偏多，二月若正常則三月可能也正常，二月若偏少三月可能也偏少。

三月若偏多則四月可偏轉為正常，三月若正常則四月可能變成偏多，三月若偏少四月可能正常。

四月若偏多五月可能正常，四月若正常五月可能保持正常，四月若偏少五月可能變成偏多。

五月若偏多六月可能也偏多，五月若正常六月可能也正常，五月若偏少六月可望正常。

六月若偏多七月可能正常，六月若正常七月可能偏少，六月若偏少七月可能正常。

七月若偏多八月可能正常，七月若正常八月可能偏

多，七月若偏少八月可能正常。

八月若偏多九月可能也偏多，八月若正常九月可能偏少，八月若偏少九月可能正常。

九月若偏多十月可能偏少，九月若正常十月可能正常，九月若偏少十月可能正常。

十月若偏多十一月可能偏少，十月若正常十一月可能也正常，十月若偏少十一月可能偏多。

十一月若偏多十二月可能正常，十一月若正常十二月可能偏多，十一月若偏少十二月可能也偏少。

十二月若偏多次年一月可能仍偏多，十二月若正常一月可能偏少，十二月若偏少一月可能也偏少。

全年逐月降水三狀態的一步最大可能轉移過程示意圖則如圖 2 所示。和氣溫相同，降水正常區間取為 30%-70% 的 40%，所以多次轉換後轉移矩陣輻合至一定值。

表四 台北全年逐月降水三狀態的一步(月)轉移概率矩陣 (X 1000)

U代表高於正常，N代表正常，B代表低於正常

	U	N	B		U	N	B		U	N	B		U	N	B
MONTH = 1 TO 2				MONTH = 4 TO 5				MONTH = 7 TO 8				MONTH = 10 TO 11			
U	306	380	314	U	315	364	322	U	362	305	332	U	180	548	272
N	239	452	309	N	304	500	196	N	320	455	225	N	355	312	332
B	393	309	298	B	303	266	431	B	235	380	385	B	373	318	310
MONTH = 2 TO 3				MONTH = 5 TO 6				MONTH = 8 TO 9				MONTH = 11 TO 12			
U	284	519	197	U	535	231	233	U	297	307	397	U	182	558	260
N	300	348	352	N	168	421	411	N	360	325	315	N	301	380	319
B	339	301	361	B	253	497	250	B	250	543	207	B	439	222	339
MONTH = 3 TO 4				MONTH = 6 TO 7				MONTH = 9 TO 10				MONTH = 12 TO 1			
U	161	577	261	U	299	404	298	U	333	277	391	U	344	396	260
N	395	322	283	N	242	521	237	N	345	412	244	N	325	352	323
B	342	276	382	B	397	199	404	B	233	464	303	B	247	390	363

3.2 預報評估

表五所示則為台北 1991-95 年各月氣溫的觀測和一次轉移預報機率分佈矩陣。以最大狀態機率代表狀態，圖 3 則為觀測及一步轉移預測之狀態的示意圖，由圖可見氣溫一步轉移預報有相當的預報技術，二步以上的轉移由上述馬可夫過程的適歷性原理和正常狀態佔 40% 的定義，所以預報多偏正常。

同樣由台北 1991-95 年各月降水觀測及一步轉移預測之機率分配矩陣，可知降水觀測和預報的狀態。圖 4 則為降水觀測及一步轉移觀測之狀態的示意圖。由圖可見，降水一步轉移預報只有邊際的預報技術。

Preisendorfer and Mobley(1984)曾經使用 Mobley Score(M Score) 評估在美國使用幾種不同方法所做的長期預報之預報技術。Mobley Score 之定義為：首先

表五 1991-95台北逐月氣溫觀測和一次轉移預報之三分類機率(X 1000)

U代表高於正常，N代表正常，B代表低於正常

		1991	1991	1992	1992	1993	1993	1994	1994	1995	1995
		OBS	PR1	OBS	PR1	OBS	PR1	OBS	PR1	OBS	PR1
1 月	U	733	323	517	322	000	322	690	480	043	322
	N	267	365	483	366	418	366	310	308	957	366
	B	000	312	000	313	582	313	000	212	000	313
2 月	U	776	519	000	437	819	145	474	502	000	259
	N	224	292	612	328	181	458	526	299	009	409
	B	000	190	388	234	000	396	000	199	991	333
3 月	U	1000	411	560	249	474	418	000	360	000	199
	N	000	346	440	415	526	341	483	375	785	403
	B	000	244	000	337	000	241	517	265	216	398
4 月	U	767	440	000	379	000	367	1000	239	445	276
	N	233	371	828	406	483	413	000	383	555	422
	B	000	189	172	215	517	220	000	378	000	302
5 月	U	1000	331	000	297	014	291	991	341	000	318
	N	000	386	224	382	986	386	009	389	526	383
	B	000	282	776	321	000	323	000	271	474	298
6 月	U	1000	337	043	292	474	297	819	337	819	294
	N	000	463	957	332	526	380	181	463	181	350
	B	000	200	000	377	000	323	000	201	000	357
7 月	U	1000	676	129	224	819	428	819	591	129	591
	N	000	219	871	454	181	348	181	264	871	264
	B	000	105	000	322	000	224	000	146	000	146
8 月	U	948	471	000	311	302	437	474	437	474	311
	N	052	357	677	371	698	360	526	360	526	371
	B	000	172	323	319	000	203	000	203	000	319
9 月	U	690	253	000	342	259	328	000	308	000	308
	N	310	454	806	365	741	412	138	423	957	423
	B	000	293	194	294	000	260	862	269	043	269
10 月	U	000	413	000	225	000	297	000	214	776	227
	N	285	312	000	406	336	355	336	490	224	387
	B	716	275	1000	370	664	347	664	296	000	386
11 月	U	259	267	000	239	1000	273	1000	273	000	331
	N	741	372	634	348	000	376	000	376	289	388
	B	000	361	366	413	000	351	000	351	711	281
12 月	U	1000	305	1000	284	129	357	1000	357	000	281
	N	000	411	000	375	871	433	000	433	310	349
	B	000	284	000	341	000	210	000	210	690	371

由觀測值和預測值分類列連表，找出一階誤差數 V，和 2 階誤差數 W，Mobley Score = V + 2W，一個預報的 Mobley Score 越高表示技術越差。以下引用 Mobley Score 來評估，利用模糊分類和馬可夫概型分析所做台北 1991-1995 平均氣溫及降水長期預報之結果如表六所示，對氣溫一共做了 60 個月的預報，所以一個隨機預報一階誤差數為 7.2×4 ，二階誤差數為 5.4×2 ，

所以隨機預報的 M Score = $7.2 \times 4 + 5.4 \times 2 \times 2 = 50.4$ 。對於氣溫的預報一階誤差為 $17+2+6+10=35$ ，二階誤差為 2，所以 M Score = $35+2 \times 2=39<50.4$ 可見氣溫的長期預報有相當的技術得分。對於降水的預報一階誤差數為 $10+7+4+13=34$ ，二階誤差為 $4+1=5$ ，所以 M Score = $34 + 5 \times 2 = 44<50.4$ 。所以降水的預報仍有邊際預報技術。

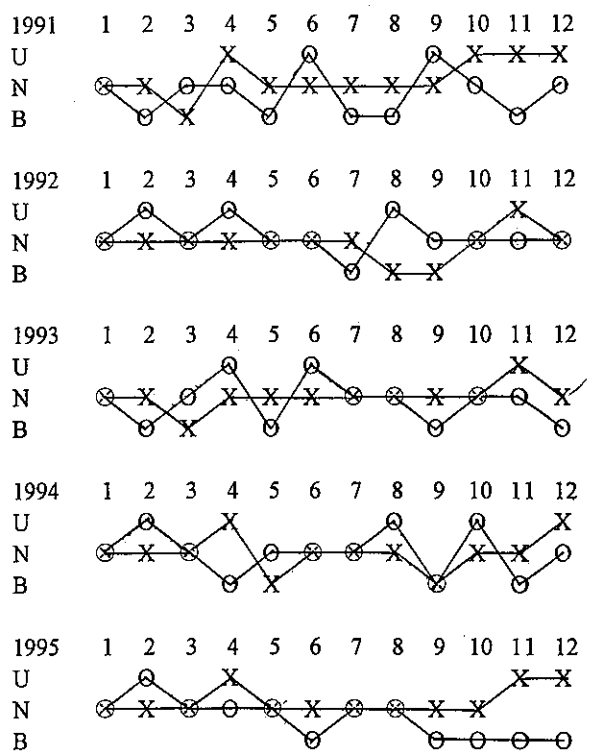
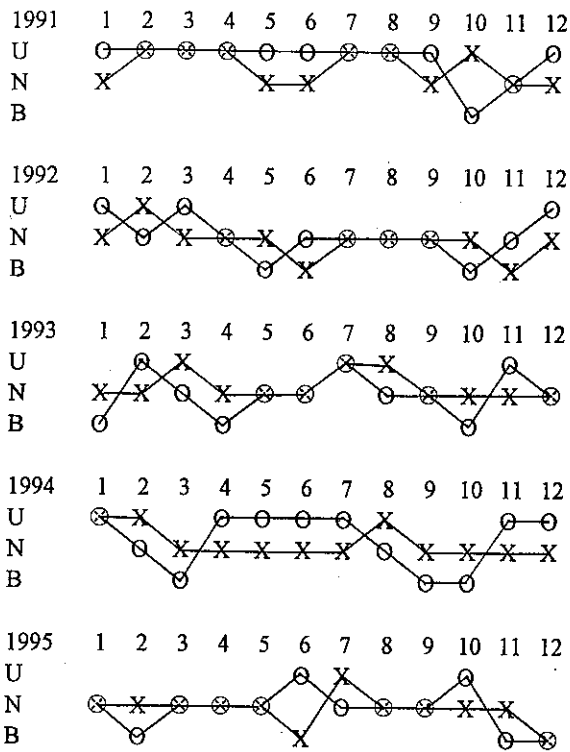


圖 3 以最大狀態機率代表狀態之台北 1991-95 年逐月氣溫觀測與一次轉移預報結果，O 代表觀測，X 代表預報。U 為高於正常，N 為正常，B 代表低於正常

圖 4 同圖 3，但為台北逐月降水

表六 高於正常 U(30.1%)，<N 正常，N(40%)，低於正常(30%)

三分類六十次預報結果之檢定

- (a) 隨機預報頻次預報 (b) 誤差階次預報
(c) 氣溫一次轉移預報 (d) 降水一次轉移預報頻次

(a)

		預報		
		U	N	B
觀測	U	5.4	7.2	5.4
	N	7.2	9.6	7.2
	B	5.4	7.2	5.4

(b)

		預報		
		U	N	B
觀測	U	0	1	2
	N	1	0	1
	B	2	1	0

(c)

		預報		
		U	N	B
觀測	U	7	17	1
	N	6	15	2
	B	1	10	1

(d)

		預報		
		U	N	B
觀測	U	0	10	1
	N	7	20	4
	B	4	13	1

四、討論與結論

長期預報的時間尺度因超過天氣的可預報度範圍，所以必須使用統計方法。而統計預報結果的校驗評估則需依賴氣候資料的統計。傳統隨機概率的觀點可能尚不足以完全說明氣候資料可應用在長期預報上的資訊。馬可夫概型分析提供了進一步統計資料應用在長期預報資訊抽取的一個方法。但傳統的分類方法應用於馬可夫概型分析時在分類的臨界點產生巨大落差，減低了馬可夫概型分析的應用價值。為解決這個問題，本文引入了模糊分類的概念--由邏輯值的連續化去除分類臨界點所造成的不連續問題，因而提高了馬可夫概型分析的應用價值。

本研究即結合模糊分類與馬可夫概型分析方法，並將之應用於台北月平均氣溫和降水三分類的長期預報。首先以 1961-90 三十年資料來建立轉移概率矩陣，由逐月氣溫的轉移概率矩陣可見氣溫有明顯的持續性，由逐月降水的轉移矩陣可見降水的類型轉移較為複雜。

其次利用 1991-95 年的資料作後預報來檢定此一預報方法的預報技術，結果顯示氣溫的長期預報有相當的預報技術。降水的長期預報也有些微的預報技術。顯然我們對於一般的預報時評估其技術時需採較高水平。又 1991-95 年的月平均氣溫，顯示高於正常的情形比氣候多（氣候值應為 60 個月的 30% 即 18 個月，但實際發生者為 25 個月），低於正常值（12 個月）較氣候值少，溫度的上升異常顯著，所以做長期預報時，特別是夏季氣溫可以多報高於正常。

參考文獻

- 陳國範, 1976: 模糊數學與天氣預報. 氣象, 1976 第 6 期, 北京.
- 羅積玉, : 多元統計分析方法與應用. 科技圖書公司, 臺北.
- 曹鴻真, 陳國範; 1980: 天氣過程的 Fuzzy 劃分. 科學通報, 1980 第十期, 北京.
- 馮德益, 樓世博, 1991: 模糊數學方法與應用.
- MO, K. and M. Ghil; 1988: Cluster Analysis of multiple planetary flow regimes, J. of Geophy Res. vol. 93. No. D9. 10927-10952.
- Preisendorfer, R. and C. Mobley, 1984: climate forecast verification, U.S. mainland, 1974-1983. Mon. Wea. Rev., 112, 809-825.

Applying Fuzzy Classification and Markov Process Analysis to the Long-range Probability Forecasting

Ming-Chin Wu¹ Jieh-Guei Ho²

1. Department of Atmospheric Sciences
National Taiwan University

2. Forecasting Center, the Central Weather Bureau

Abstract

In this paper, a long-range probability forecasting method is introduced. This method is the combination of fuzzy classification and Markov process analysis. This method is applied to the long-range probability forecasting of the monthly mean temperature and precipitation in Taipei. Markov process analysis provided a further scheme to abstract the forecasting information contained in the climate data. The fuzzy classification resolves the problem caused by the discontinuity appeared at the critical value of conventional classification.

The transfer probability matrix of the 3 class of the monthly temperature and precipitation of 1961-90 indicate that the temperature is quite persistent. On the other hand, the precipitation is not. The hindcasts of the 1991-95 data indicate that applying this method to forecast the temperature shows ratherly skillful. However, the skill for the precipitation forecasting is ratherly marginal.