非線性類似法預測旬平均溫度

陳圭宏
中央氣象局　氣象預報中心

摘　要

不規則（非線性）的時間序列裏，假設兩小線段在過去短時間的演化軌跡類似，未來也會有類似的演化軌跡。所以在時間序列的資料庫裏，找尋和目前線段類似的演化軌跡；以其未來的演化做爲目前時刻對未來演化的預測。因爲不規則時間序列的變動很不定，所以不同方式的類似會有不同的結果。爲了掌握各個不規則時間序列的特性，只有做大量的測試，選出較好的類似方式。

用非線性類似法測試台北、台中、高雄和花蓮四個氣象站的旬平均溫度。結果選擇絕對值誤差平均有比尤拉最短距離或均方根的類似法好。而季節因素的考慮，雖然減少了可供比較的線段個數，但它卻有掌握季節的特性，反而有比較好的成績。雖然可以選到比較持續法好的類似法組合，可是其技術得分不高，命中率不到 50 %，在實際應用上仍需小心。

一、前言

（ Sugihara and May ， 1990 ）認爲造成自然界動力系統預報的不確定，主要是由於觀測的誤差和動力系統本身的複雜性，容易導出混亂的軌跡。他們認爲用非線性類似方法可以對此種不規則軌跡作短時間的預測。（ Drosdowsky ， 1994 ）用同樣的方法預測南方振盪指數（ SOI ）。他認爲有比持續法更好的預報能力。本文採用同樣的方法測試和校驗台北、台中、高雄和花蓮的旬平均溫度時間序列。

二、資料和方法

旬平均溫度的定義是每個月分爲三旬，上、中、下旬的平均溫度分別是 1 日到 10 日、 11 日到 20 日和 21 日到月底的平均溫度。則一年計有 36 旬。各個旬再做一次常態化，去除各旬季節的變化。選定四個旬平均溫度的時間序列，包括有台北（ 46692 ）、台中（ 46749 ）、高雄（ 46744 ）和花蓮（ 46699 ）。時間序列長度都是從 1950 到 1995 年， 56 年共有 1656 個時間序列點。圖一是上述四個氣象站 1950 年第一旬到 1959 年第 36 旬的時間序列，可見每個氣象站的時間序列都是很不規則的。

定義時間序列　X_t ， t=1,2,3,・・・,1656

時間序列類似的選取，首先必須選定 E －空間（ embedding dimension ），它是由時間序列的部份線段（ X_t ， X_{t-1} ， X_{t-2} ，・・・・， $X_{t-(e-1)}$ ）構成。 X_t 是時間序列中的某一旬， X_{t-1} 是 X_t 的前一旬， $X_{t-(e-1)}$ 是 X_t 的前（ e-1 ）旬。其中的元素（ X_t ， X_{t-1} ， X_{t-2} ，・・・， $X_{t-(e-1)}$ ）也構成 e 維 E －空間的座標。例如：當 e ＝ 2 ，組成二維 E －空間座標爲（ X_t ， X_{t-1} ），當 e ＝ 3 ，組成三維 E －空間座標爲（ X_t ， X_{t-1} ， X_{t-2} ），其中 t ＝ 1,2,3,........,1656 。所以在三維的 E －空間可有 1656 － 2 ＝ 1654 個部份線段。假設時間序列的演化在過去 e － 1 旬到現在的時間內類似，未來也會有類似的演化。因此在三維 E －空間內可有 1653 個部份線段做比較（本身除外）。找出最類似的線段，其外推數旬，即爲對未來數旬的預測。兩線段的類似程度，可有多種方法估計。一般是用最小尤拉距離（ Euclidean distance ）、均方根（ root mean square ）和絕對值誤差的平均。而選取類似的線段可爲最類似的一個，也可以有較類似的數個線段，再乘以權重函數決定。權重函數有 $\exp(-d^2)$ ， d 是距離。它的權重曲線如圖二所示。當距離增加時，權重比例會急劇下降，當距離大於 2 時，權重趨近於零。另外有等權重函數和距離 d 無關。至於要以何種方式的組合較佳，只有大量測試和校驗去選取。

三、測試和校驗

不規則時間序列的演化非常不定，要重覆的測試校驗選取較適合 e 維 E －空間、類似方法和類似程度的權重函數。本文中所測試的組合方式有 E －空間從 1 維到 10 維，類似方法有尤拉距離、均方根和絕對值誤差的平均，並且取最好的 1 、 3 、 5 、 10 個線段，而權重函數取 $\exp(-d^2)$ 和等權重函數，再加上季節因

素的考慮與否。至於何種組合方式比較好，本文用命中率和技術得分（表一）（陳，1994）的好壞，判斷那一種組合較佳。因爲常態化後的時間序列，根據常態曲線分布，在小於-0.52 和大於 0.52 的部分，各約占 30％，而介於-0.52 到 0.52 之間約占 40％。這種分布和中央氣象局劃分低溫、正常溫度、高溫的頻率分布 30％、40％、30％很相近。所以只要預測的溫度和實際溫度是在同一溫度區間就算命中。命中率就是命中的次數除以預測總數。而技術得分主要是和純機率預測相比較。如果得分爲正，表示預報有優於純機率預測的技術。

四個氣象站旬平均溫度時間序列，較佳的 E 一空間維數、類似方法和權重函數等組合方式不一樣。也很難找到有某一特定的組合，可以適用於各種不同的時間序列，且大部分組合的技術得分比持續法的技術得分低，甚至爲負的得分，沒有實用價值。不過一般在考慮了季節因素後，會有較好的技術得分。所謂考慮季節因素是在時間序列上選取部分線段（X_t，X_{t-1}，X_{t-2}，．．．．，$X_{t-(e-1)}$）時，只取和 t 時間在日曆上是同一旬或左右兩旬比較，其他的旬不考慮。如此可供選取比較的線段就會減少很多。但它卻表現了各個季節變化的特性。反而有較好的預報技術。例如：圖三的 3 × 3，9 格小圖中是測試在某些組合方式下，花蓮旬平均溫度預測的技術得分。計有 1985 — 1994，10 年 360 個預報個案。類似方法：第一行（column）用尤拉距離；第二行用均方根；第三行是絕對值平均誤差。權重函數：第一列（array）、第二列是用 $\exp(-d^2)$；第三列是用等權重函數。第二列和第三列有考慮季節因素；第一列沒有考慮。其他的參數：最好的三組線段和 5 維 E 一空間都相同。實線是實際預測的技術得分，虛線是持續法預測的技術得分。實線和虛線上的數字，表示對第 1、2、3、．．旬的預測。由圖三 a、d、g 的比較，在考慮了季節因素後的 d、g 小圖，在第 1、2、3、4 旬預測的技術得分，有顯著的增加，甚至優於持續法的技術得分。同樣圖三的第二行和第三行在考慮了季節因素後，也有類似明顯改善技術得分的表現。

至於何種組合方式的類似較具有實用價值，除了要有優於持續法的技術得分外，穩定性也很重要。所以四個時間序列除了測試 1985 — 1994，10 年外，同時也測試 1985 — 1989 和 1990 — 1994 兩個 5 年的技術得分。確定是否有穩定優於持續法的技術得分。而且主要的重點在第 1、2、3、4 旬的預測。圖四是根據上述的原則，對台北、台中、高雄、花蓮四個旬平均溫度的時間序列挑出較好類似法的組合方式。台北：絕對值誤差平均、最佳的一組線段和 5 維 E 一空間。台中：絕對值誤差平均、最佳的一組線段 3 維 E 一空間。高雄：絕對值誤差平均、最佳的一組線段和 6 維 E 一空間。花蓮：絕對值誤差平均、最佳的三組線段、$\exp(-d^2)$ 的權重函數和 7 維 E 一空間。其中只有高雄不含季節因素外，其他三個時間序列都考慮了季節因素。四個時間序列都共同選用絕對值誤差平均。除了花蓮是選用最好的三組線段外，其他三個時間序列都是取用最好的一組線段而已。至於 E 一空間的維數，四個時間序列都不一樣。

四、討論與結論

（Drosdowsky，1994）在建立好 E 一空間後，建議用尤拉最短距離或均方根來判斷類似。但是本文對四個氣象站的旬平均溫度測試的結果，絕對值誤差平均有更好的成績，且大部分都只取最好的一組線段而已。至於季節因素的考慮，雖然減少了可供比較的線段個數，但卻含有季節的特性，反而有較好的技術得分。

雖然對台北、台中、高雄、花蓮四個旬平均溫度的時間序列有挑出較好類似法的組合方式，但是只有花蓮第 1、2 旬預測的技術得分可以達 15％左右，其他都在 10％左右，相對於命中率只有 45％左右。尚不及 50％的命中率。即預測兩次，答對不到一次。在實際預報的應用上，仍要小心使用。

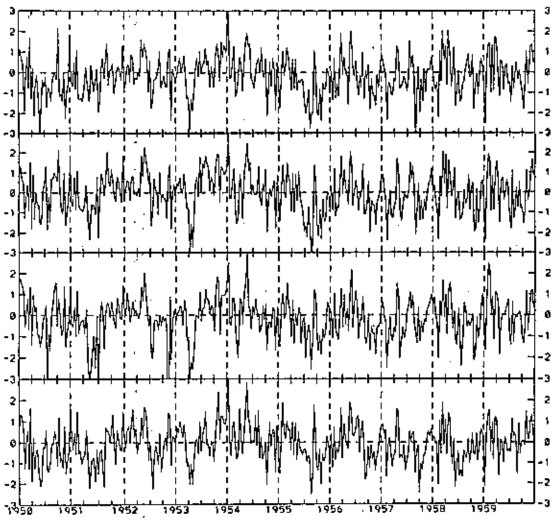
考文獻參：

曾振發，陳圭宏，林燕璋，1994：赴日本研習長期天氣預報技術。行政院所屬各機關因公出國報告書。

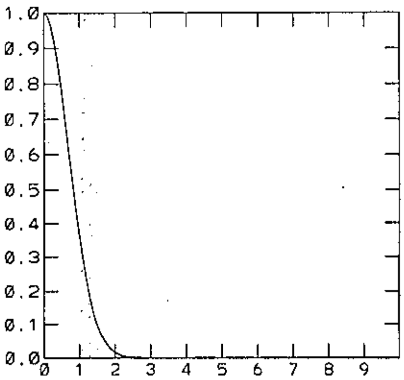
戚啓勳，嚴夢輝，1978：氣象統計學，復興書局。

Drosdowsky，W.，1994：Analog（nonliear）forecasts of the southern oscillation index time series. Weather and Forecasting , 9, 78-84.

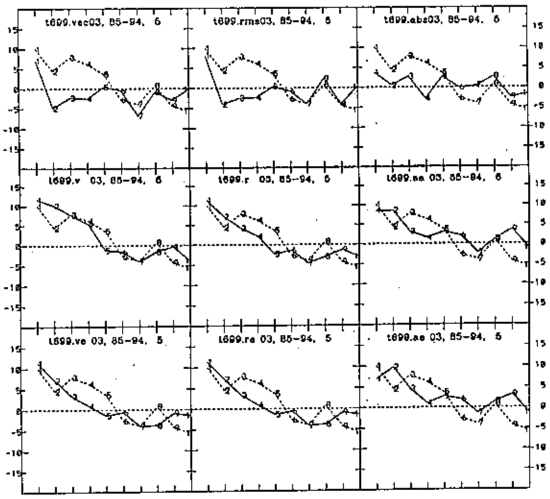
Sugihara，G.，and R. M. May，1990：Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, 344, 734-741.



圖一、（ａ）台北（ｂ）台中（ｃ）高雄（ｄ）花蓮四個氣象站 1950 年第一旬到 1959 年第 36 旬的旬平均溫度時間序列。



圖二、 $\exp(-d^2)$ 的權重曲線，橫軸是距離 d ，縱軸是權重比例。



圖三、花蓮旬平均溫度時間序列測試的技術得分，第一行（ column ）用尤拉距離；第二行用均方根；第三行是絕對值平均誤差。權重函數：第一列（ array ）、第二列是用 $\exp(-d^2)$ ；第三列是用等權重函數。第二列和第三列有考慮季節因素；第一列沒有考慮。其他的參數：最好的三組線段和 5 維 E 一空間都相同。實線是實際預測的技術得分，虛線是持續法預測的技術得分。實線和虛線上的數字，表示對第 1 、 2 、 3 、‧‧‧旬的預測。

表一、技術得分的算法和列聯表

		預測			
		−	o	+	
觀測	−	n_{11}	n_{12}	n_{13}	$n_{1.}$
	o	n_{21}	n_{22}	n_{23}	$n_{2.}$
	+	n_{31}	n_{32}	n_{33}	$n_{3.}$
		$n_{.1}$	$n_{.2}$	$n_{.3}$	n

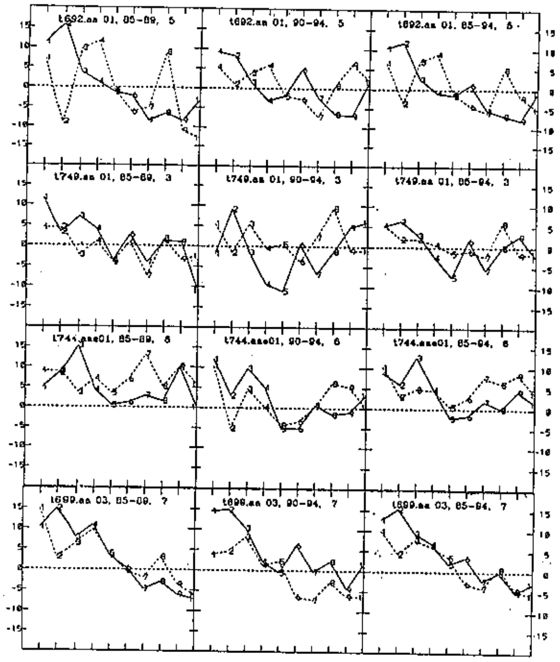
總數 $n = n_{11} + n_{12} + \cdots + n_{33}$

命中數 $C = n_{11} + n_{22} + n_{33}$

命中率 rate $= C/n$

$$S = \frac{1}{n}\left(n_{.1} \times n_{1.} + n_{.2} \times n_{2.} + n_{.3} \times n_{3.}\right)$$

技術得分 $Sc = [\,(C-S)/(n-S)\,] \times 100\%$



圖四、台北、台中、高雄、花蓮四個氣象站旬平均溫度時間序列，較好的類似方式組合。第一行是 1985-1989 ；第二行是 1990-1994 ；第三行是 1985-1994 測試的技術得分。第一列台北：絕對值誤差平均；最好一組線段； 5 維 E 一空間，第二列台中：絕對值誤差平均;最好一組線段； 3 維 E 一空間,第三列高雄：絕對值誤差平均；最好一組線段； 6 維 E 一空間，第四列花蓮：絕對值誤差平均；最好三組線段； exp(-d^2)權重函數； 7 維 E 一空間。只有高雄含有季節因素；其他三個氣象站沒有考慮。