

# 基於時空序列 STARMA 之肺結核傳播時空關聯分析

## On the Space-Time Correlation Analysis Based on STARMA Models for Tuberculosis Epidemic in Taiwan 2003-2012

張家豪

亞洲大學生醫資訊學系  
allan8206100@gmail.com

李正宇

亞洲大學生醫資訊學系  
leecheng2005@yahoo.com

### 摘要

傳染病傳播分析常採取純時間或純空間的分析；然而，這將有可能產生誤導性的結果 [13]。本研究以時間序列的時空延伸 -- STARMA (Space-Time Autoregressive Moving-Average) 模型為基礎，對台灣地區的肺結核 (Tuberculosis) 傳播動態進行時空關聯性分析，並對分析結果提出統計的推論，以資未來此類疾病管控的理論參考。我們發現疾病傳播的統計過程是由兩種現象交互作用產生的結果，即：病毒擴散行為與空乏效應 (depletion effect)；這兩種行為的相互抵消與隱藏，讓傳統的純空間分析或純時間分析對疫情易作出誤導性的結論。

**關鍵詞：**時空序列、STARMA、肺結核、傳染性疾病、空乏效應

### Abstract

Traditionally, the methods for epidemic analysis of infectious diseases are solely temporal or solely spatial analyses, and potentially lead to misleading results [13]. In this study, a true space-time model, STARMA (Space-Time Autoregressive Moving-Average) model was used as a framework to study dynamics and space-time correlation of Tuberculosis in Taiwan 2003-2012 to provide statistical inferences. We found that the process of the disease spreading was an interaction of two STARMA processes, i.e. spreading process and depleting process. If only temporal or only spatial analysis is considered, mixture and cancellation of the two processes may hide the dynamics of the disease and lead to a wrong prevention or control policy.

**Keywords:** Space-time series, STARMA, Tuberculosis, Infectious disease, depletion effect

### 1. 介紹

本研究基於時空模型 STARMA (Space-Time Autoregressive Moving-Average) [1][2][6][7]，對台灣地區的肺結核 (Tuberculosis) 疫情進行時空關聯性分析；經由適當的統計推論，我們得到具體結果可供肺結核疾病傳播管控的參考。

由結核桿菌感染所造成的結核病是一種目前仍普遍存在於全世界，尤其是未開發及開發中國家的慢性傳染病。在臺灣一年四季都有病例，男性發生率比女性高，老年人發生率比年輕人高。在初感染時，大約 95% 的人會因自身的免疫力而未發病，但會有終身再活化的潛在危險，只有 5% 的人在初感染後結核菌會經由血行或淋巴液之散播造成肺內或肺外結核（如：結核性腦膜炎）。經過初感染而未發病的人，日後也可能因外在再感染而發病。

結核病可以發生在人體任何器官或組織，如淋巴結、腦膜、胸膜、腎臟、骨骼、皮膚、消化道、泌尿生殖道等；在臺灣比較常見的肺外結核是淋巴結核及骨結核，其次為結核性腦膜炎，肺外結核的發生率遠比肺結核來得低。若給予適當的抗結核藥物治療，結核病幾乎可以百分之百痊癒 (Cure)，但若不予治療，則在 3 年內，約有一半的病人會死亡。

為了瞭解台灣地區肺結核傳播的時空分析中所隱藏的自然機制，我們對台灣行政區定義合理的空間關聯結構 (spatial correlation structure) [5][8]，並使用時空序列分析軟體 IEAST [3] 進行時空關聯性分析。

## 2.方法

### 2.1 時空自我迴歸移動平均 (STARMA)

本研究採取的 STARMA 模型為時間序列 ARMA 模型加入空間的定義[2] (如圖 1) 與關係組合而成[1][2]，它能夠完全表達許多統計現象在空間上和時間上的相關性及交互作用[4][9][10]。

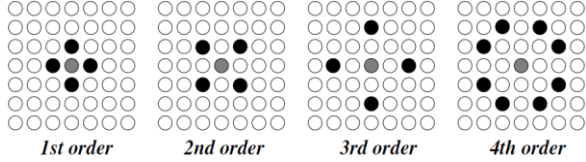


圖 1. 空間階次定義的範例：相對於中心灰色的空間點，與第一圖中的四個黑點的關係為空間第一階；第二圖則為空間第二階；依此類推

若  $Z_{i,t}$  是時空變數，代表在點  $i$  ( $i=1, 2, \dots, N$ )、時間  $t$  時的時空隨機變量，基於空間階次的定義 (可轉換為下列方程式中的  $W$  矩陣)，STARMA 模型一般可表達成 [4][5][12]：

$$Z_t = \sum_{k=1}^p \sum_{l=0}^r \phi_{kl} W^{(l)} Z_{t-k} - \sum_{k=1}^q \sum_{l=0}^s \theta_{kl} W^{(l)} \varepsilon_{t-k} + \varepsilon_t \quad (2.1)$$

其中，

$p, r$  是自我迴歸項的最大時間階次、空間階次； $q, s$  是移動平均項的最大時間階次、空間階次； $\phi_{kl}, \theta_{kl}$ ：在時間延遲  $k$  和空間延遲  $l$  時自我迴歸、移動平均的模型參數；

$W^{(l)}$ ：空間  $l$  階時  $N \times N$  權值矩陣

$\varepsilon_t$ ：時間  $t$  時的隨機雜訊向量

上述的 STARMA model 當中，假如  $q=0$ ，則形成時空自我迴歸模型，可以表達成以下公式：

$$Z_t = \sum_{k=1}^p \sum_{l=0}^r \phi_{kl} W^{(l)} Z_{t-k} + \varepsilon_t \quad (2.2)$$

而時空自我迴歸模型為一可以同時考慮時間與空間並為最適合描述疾病傳播的統計模型 [3]。

### 2.2 權值矩陣

在模型中有一方陣  $W$ ，稱為權值矩陣，是描述空間關係的關鍵[11][12]，說明如下。如果空間的總位點是  $N$ ，相對應的權值矩陣尺寸就是  $N \times N$ ；所有權值矩陣中元素的值都必須大於 0，每一列或每一行的加總必須等於 1，如下式。

$$\sum_{j=1}^N \omega_{ij}^{(l)} = 1 \text{ 且 } \omega_{ij}^{(l)} \geq 0 \quad (2.3)$$

空間上權值矩陣的定義依序包含了三步驟：

(1) 空間階次的定義 (SOD, Spatial Order Definition)，(2) 空間關聯矩陣 (SRM, Spatial Relation Matrix) 的定義 (3) 計算空間權值矩陣 (SWM, Spatial Weighting Matrix)。對於大多數的應用，一旦 SOD 定義完成，SRM 與 SWM 就可以由 IEAST 系統自動產生，而 SWM 就是 STARMA 模型中的  $W^{(l)}$ ，可供模型描述空間關係 [7, 10]。以下我們舉個簡單的例子來說明上述過程：

首先定義空間關聯結構 SOD，它相對地描述了每一個空間點與其相鄰點的空間階次關係：

$$\text{SOD} = \begin{bmatrix} - & 4 & 3 & 4 & - \\ 4 & 2 & 1 & 2 & 4 \\ 3 & 1 & 0 & 1 & 3 \\ 4 & 2 & 1 & 2 & 4 \\ - & 4 & 3 & 4 & - \end{bmatrix} \quad (2.4)$$

接著在各空間階次中，按照實際空間關聯情形適度分配權重，使得各階次中的所有權重總和為 1 且能反應真實的空間關係，如下：

$$\text{SRM} = \begin{bmatrix} - & .125 & .25 & .125 & - \\ .125 & .35 & .40 & .35 & .125 \\ .25 & .25 & 0 & .25 & .25 \\ .125 & .15 & .10 & .15 & .125 \\ - & .125 & .25 & .125 & - \end{bmatrix} \quad (2.5)$$

因此，對於所有  $Z$  中的點，空間  $l$  階相鄰點所產生的影響可寫為：

$$W^{(l)}Z = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots \\ .125 & .35 & .40 & \dots & \dots & .125 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \\ \vdots \\ \vdots \\ p \end{bmatrix} = \begin{bmatrix} a^* \\ b^* \\ c^* \\ \vdots \\ \vdots \\ p^* \end{bmatrix} \quad (2.6)$$

其中，為了數學計算上的便利，二維的空間 (矩陣) 將會改寫為行向量 (column vector)，也就是以一維向量表達二維空間狀態，如下：

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} \rightarrow \begin{bmatrix} a \\ b \\ c \\ \vdots \\ \vdots \\ p \end{bmatrix} = Z \quad (2.7)$$

至於本研究中所使用的空間階次的定義 (SOD)，在此說明如下。將台灣各個縣市依序排列，有些縣市並存，例如：台北市、新北市、基隆市，我們將這三個歸類再同一個區。接著用一百五十萬分之一的台灣地圖，將每個縣市依序等比例(0.7cm)的放大 1~5 倍，再比較放大過 1 倍的區域與其他縣市原來倍率有無交集，若交集超過一半，我們將視為有關聯並定義此區為第一階，以此類推放大至 5 倍；若超過 5 倍都未有交集，我們將視為無關聯；此定義有例外，因為台灣中間有著中央山脈，所以山脈的東西邊並不太會互相影響，例如：台中與花蓮，雖然地圖上是相連的，但隔著中央山脈病毒不易傳遞，因此無較大的關聯。

下表即是台灣各區空間階次定義：

	台北	桃園	新竹	苗栗	台中	彰化	雲林	嘉義	台南	高雄	屏東	台東	花蓮	宜蘭	南投
台北	0	1	2	3	5	0	0	0	0	0	0	0	0	2	0
桃園	1	0	1	2	4	5	0	0	0	0	0	0	0	2	0
新竹	2	1	0	1	2	4	5	0	0	0	0	0	0	3	0
苗栗	3	2	1	0	1	2	4	5	0	0	0	0	0	4	0
台中	5	4	2	1	0	1	3	4	5	0	0	0	0	0	2
彰化	0	5	4	2	1	0	1	2	3	5	0	0	0	0	3
雲林	0	0	5	4	3	1	0	1	3	4	0	0	0	0	3
嘉義	0	0	0	5	4	2	1	0	1	2	4	5	0	0	3
台南	0	0	0	0	5	3	3	1	0	1	3	4	0	0	5
高雄	0	0	0	0	0	5	4	2	1	0	3	3	0	0	0
屏東	0	0	0	0	0	0	0	4	3	3	0	3	0	0	0
台東	0	0	0	0	0	0	0	5	4	3	3	0	5	0	0
花蓮	0	0	0	0	0	0	0	0	0	0	0	5	0	3	0
宜蘭	2	2	3	4	0	0	0	0	0	0	0	0	3	0	0
南投	0	0	0	0	2	3	3	3	5	0	0	0	0	0	0

表 2-1 台灣地區肺結核傳播空間階次定義

### 2.3 時空自我相關函數定義

基於 STARMA 模型，時空自我相關函數可定義如下[3]：在時間延遲  $s$  時在  $l^{th}$  和  $k^{th}$  相鄰的空間階次的時空共變異定義成

$$\gamma_{lk}(s) = E_t \left[ \frac{[W^{(l)} Z_t][W^{(k)} Z_{t+s}]}{N} \right] \quad (2.8)$$

而在時間延遲  $s$  時在  $l^{th}$  和  $k^{th}$  相鄰的空間階次的時空的自我相關函數定義成：

$$\rho_{lk}(s) = \frac{\gamma_{lk}(s)}{\sqrt{\gamma_{ll}(0)\gamma_{kk}(0)}} \quad (2.9)$$

### 2.4 資料來源

由衛生署疾病管制局取得資料來源，疾病的資料是取 2003~2012。在時間上我們選擇用年資料；空間上，由於疾病管制局將所有行政區域劃分的太細（分區太小將使空間關聯範圍過大），因此我們將所有的行政區域適當的合併，如：台北市、新北市、基隆市，這三個行政區我們將他歸類在同一個行政區中，至於，外島部分由於病例數相對而言過少，在統計上不具意義，因此我們不納入考慮範圍。

縣市名稱	0~4			5~9			10~14			15~19		
	男	女	不詳	男	女	不詳	男	女	不詳	男	女	不詳
台北市	0	0	0	2	1	0	1	1	0	15	7	0
基隆市	0	2	0	0	0	0	1	2	0	2	5	0
新北市	2	2	0	2	2	0	6	7	0	34	25	0

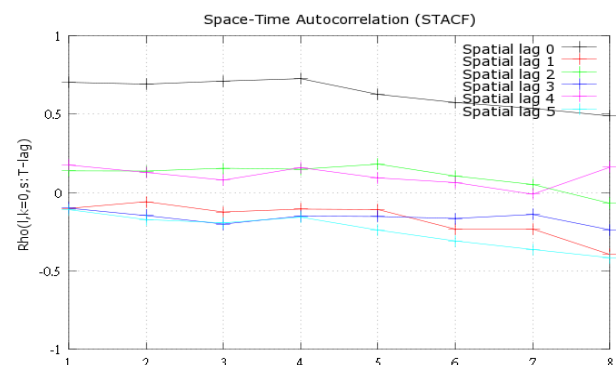
### 2.5 IEAST

我們使用的時空分析平台是 IEAST v1.3，即專為時空自我迴歸移動平均模型為架構所設計之時空分析軟體（由美國密西根州立大學新興傳染性疾病研究中心於 2002 年開發）[3]。IEAST 是分析時空自我迴歸移動平均模型的整合環境 (Integrated Environment for Analyzing STARMA models) 的縮寫，目前需在 Unix 下安裝操作，並使用 GNU Octave 執行，使用方法可參考英文手冊。  
(<http://www.masys.url.tw/AU/doc/ieastman.pdf>)

### 3. 結果

本研究使用 IEAST，並在 Windows 平台下的 GNU Octave 3.6.4 環境執行。我們將台灣地區在空間上分為 15 區，空間關聯性定義為 0~5 階次，因此我們的時空變數  $Z_t$  為一個 15x1 的向量，而權值矩陣  $W^{(l)}$  則為 15x15 的方陣，其中  $l=0\sim5$ 。

我們使用 IEAST 計算台灣在 2003~2012 的肺結核傳播時空變數  $Z_t$  時空自我關聯係數 (如 (2.9) 式)，時空自我相關係數圖則說明如下。下圖即是本研究的時空自我相關係數圖：



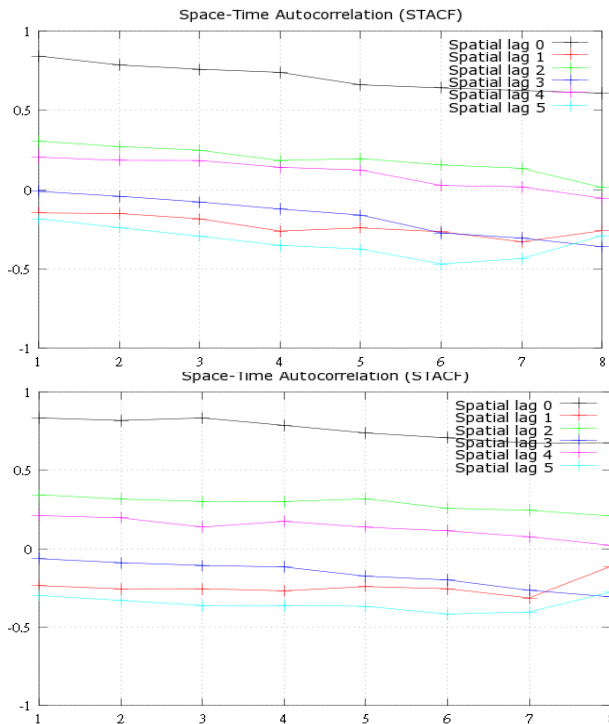


圖 2. 時空自我相關圖：16~20 歲(上)；21~25 歲(中)；26~30 歲(下)

圖中的橫軸代表時間延遲(以年為單位)；縱軸代表關聯性；lag0~lag5 分別代表各個空間階次

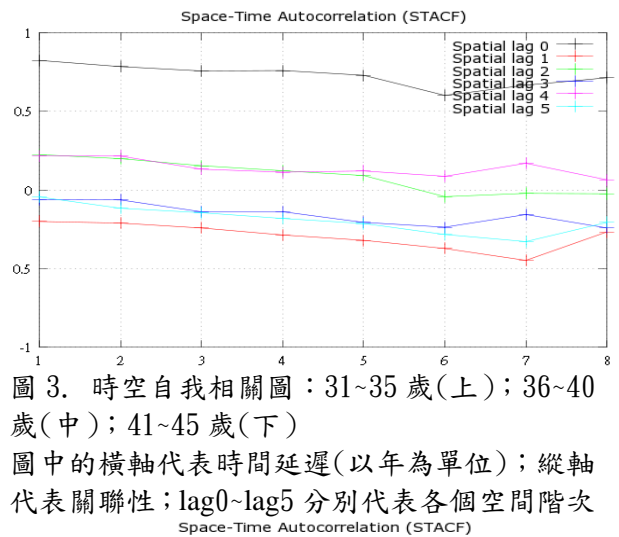
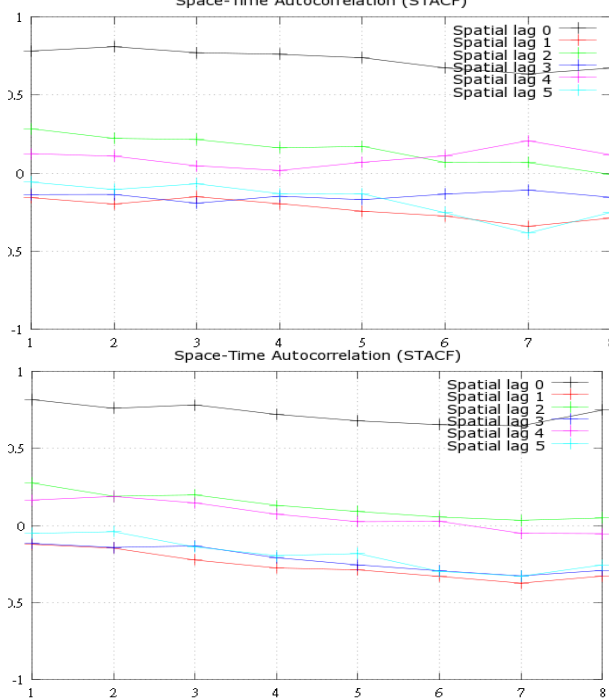


圖 3. 時空自我相關圖：31~35 歲(上)；36~40 歲(中)；41~45 歲(下)

圖中的橫軸代表時間延遲(以年為單位)；縱軸代表關聯性；lag0~lag5 分別代表各個空間階次

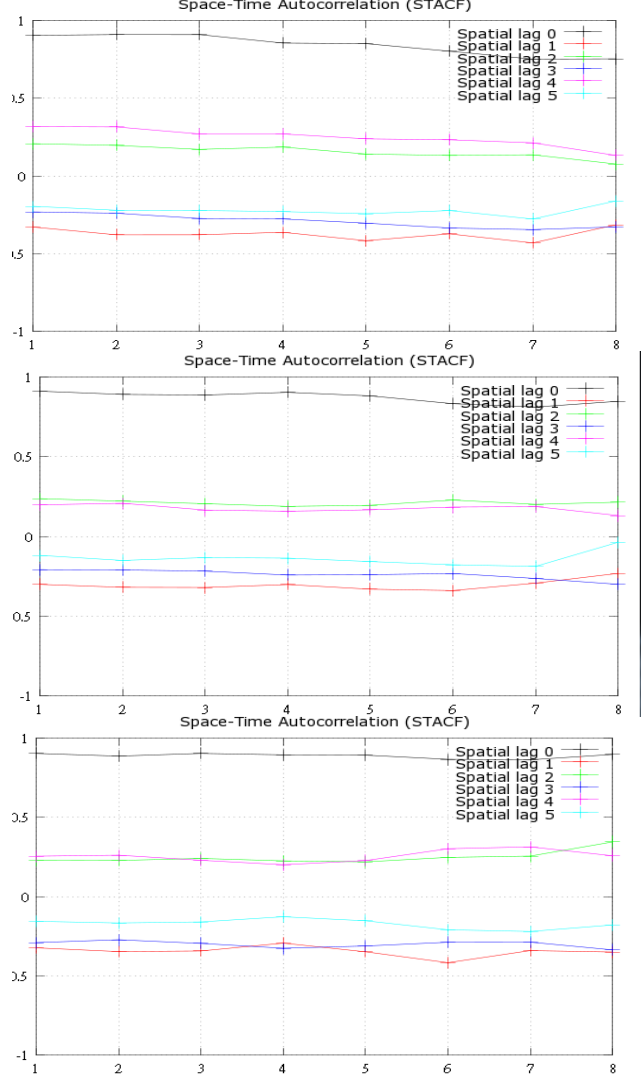


圖 4. 時空自我相關圖：46~50 歲(上)；51~55 歲(中)；56~60 歲(下)

圖中的橫軸代表時間延遲(以年為單位)；縱軸代表關聯性；lag0~lag5 分別代表各個空間階次



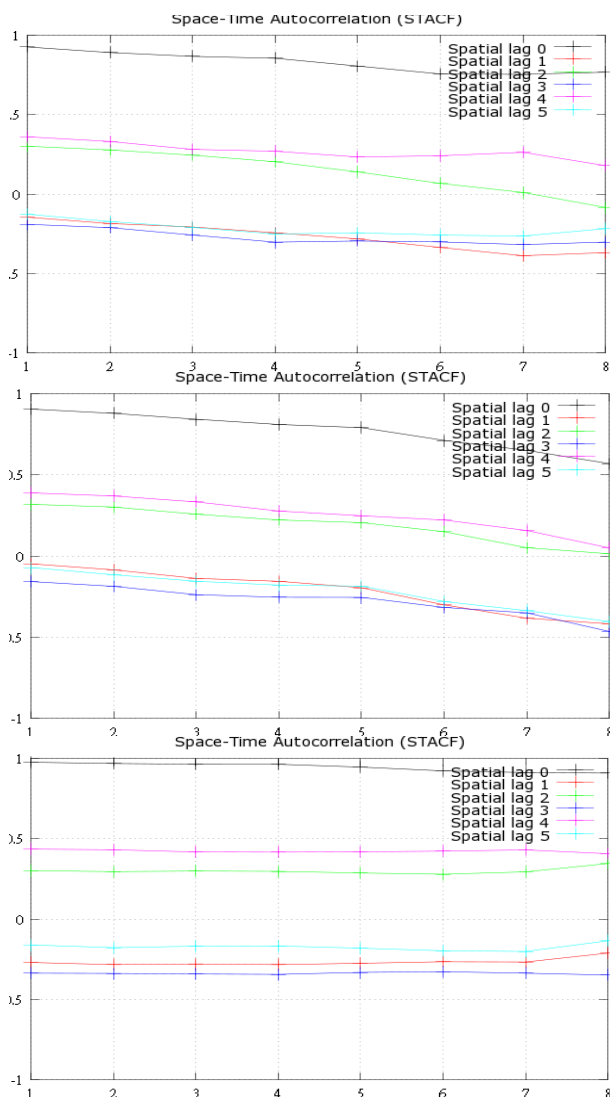


圖 5. 時空自我相關圖：61~65 歲(上)；66~70 歲(中)；71~75 歲以上(下)

圖中的橫軸代表時間延遲(以年為單位)；縱軸代表關聯性；lag0~lag5 分別代表各個空間階次

#### 4. 討論

在前節的結果中，我們發現一些共通的行為及現象值得討論，分述如下：

1. 隨著年齡增加，整體的時空關聯性是漸增的，而且無論任何空間階次或時間階次皆如此。此現象應為真實的反應，我們推論其成因為：a. 年輕人有較短的康復時間與較長的免疫維持時間；b. 老年人相較於其它年齡層病患的行動範圍極為有限。

在同一年同一空間區域內，隨著年齡群的年齡區間上升，關聯性也跟著提升，到 50 歲以上的族群，已升至高達 0.9。因此，為了降低時間關聯性，破壞肺結核(TB)傳播的時空連結，我們應提升 50 歲以上人口的免疫力，並縮短其康復時間。

2. 時空關聯性在空間上的群集現象(空間第 0 階為第一群，第 2/4 階為第二群，第 1/3/5 階為第三群)是真實存在的，並有以下特性：a. 相鄰空間階次間的時空關聯性並非連續變化，而是跳躍式的關聯性降低，或甚至產生負效應；b. 隨著年齡愈大，群集愈趨明顯；c. 每隔一個空間階次就出現負的效應，空間階次 1/3/5 時空關聯性小於 0，為負效應。此現象應屬空乏效應(depletion effect)；空乏效應的大小決定於疫情管制政策與各年齡層的平均免疫力的表現。因此我們可以推論空乏效應的產生應與患病康復者具有免疫力、媒體的疫情公布、民眾刻意在空間上迴避已知病源地及疫苗的使用有關。

#### 5. 結論

以 2003~2012 年台灣肺結核病毒疫情為例，以時空分析法較為恰當，因為在整個傳播的過程中有兩種時空自我迴歸現象混合在裡面，一種是主要的病毒傳播，另一種則是附在其中的空乏效應，若只單純的看數據是不可能看出有無空乏現象，更無法作出適當的預測與管控。

但目前國內研究仍大多著重在使用時間序列的方法，若加入空間要素的時空自我迴歸移動公式，再建立出模型，時空分析便可在生物及醫學的應用廣泛；從大尺度的生態學、環境學、傳染病學，到小尺度的個人生理(如癲癇病患腦電波的模擬與預測)，其潛力無窮[4]。

#### 參考文獻

- [1] P. E. Pfeifer. *Spatial Dynamic Modeling (Unpublished Ph.D.dissertation)*. Georgia Institute of Technology, Atlanta, Georgia, 1979.
- [2] P. E. Pfeifer and S. J. Deutsch. A three-stage iterative procedure for space-time modeling. *Technometrics*, 22 (1):35-47, 1980.
- [3] Cheng-Yu Lee. *IEAST user manual. IEAST - an integrated Environment for Analyzing STARMA models*, 2005.
- [4] Cheng-Yu Lee. *Space-time modeling and application to emerging infectious diseases*, Ph.D dissertation. Michigan State University. East Lansing, Michigan, 2005.
- [5] J. S. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B*, 36:197-242, 1974.

- [6] L. A. Aroian. Time series in m-dimensions: autoregressive models. *Communications in Statistics - Simulation and Computation*, B9 (5):491-513, 1980.
- [7] L. A. Aroian. Time series in m-dimensions definition, problems and prospects. *Communications in Statistics - Simulation and Computation*, B9 (5):453-465, 1980.
- [8] R. J. Bennett. The representation and identification of spatio-temporal systems: an example of population diffusion in north-west england. *Transactions of the Institute of British Geographers*, 66:73-94, 1975.
- [9] A. D. Cliff and J. K. Ord. Space-time modelling with an application to regional forecasting. *Transactions of the Institute of British Geographers*, 64:119-128, 1975.
- [10] B. K. Epperson. Spatial and space-time correlations in ecological models. *Ecological Modelling*, 132:63-76, 2000.
- [11] P. M. Hooper and G. J. D. Hewing. Some properties of space-time processes. *Geographical Analysis*, 13:203-223, 1981.
- [12] R. L. Martin and J. E. Oeppen. The identification of regional forecasting models using space-time correlation functions. *Transactions of the Institute of British Geographers*, 66:95-118, 1975.
- [13] 林惠君. 傳染性疾病傳播之空間分析與時空分析適切性比較：以 2002 年底特律之西尼羅病毒疫情為例, 應用統計學術研討會, 台北大學, May 30, 2008.