

# 多尺度空間模型及其在空間預測之應用

中央研究院統計研究所 黃信誠

## 一、前言

隨著科技及資料蒐集技術的進步，大量的空間或時空資料在天文、地球科學及農業等領域越來越常見。如何將大量的空間或時空資料轉化為有用的知識則有賴於能處理大量資料的統計方法。例如，監測地球環境的衛星，每日繞行地球南北極數次，產生大量的環境資料。但是分析此大量資料，卻常發生統計運算上的困難。當前較精確的空間預測（平滑、內插）方法，如 kriging，無法有效處理大量的空間資料。另一方面，一些可以處理大量資料的空間預測方法，如移動平均法、最近鄰（nearest neighbor）法及反距離權重（inverse distance weighting）法等，卻無法適切地結合空間的相關性以產生準確的空間預測曲面。如何兼顧統計精度與計算效率兩大目標是空間預測的困難問題。

欲解決此問題，Huang et al. [1] 提出了一種滿足尺度間「質量平衡」概念的多尺度空間模型（multi-resolution spatial model），可以處理大量且在空間不規則分佈的空間資料。此模型不僅可以建構不平穩（nonstationary）的空間關係，亦可以透過 change-of-resolution Kalman filter 演算法快速估計得最佳預測曲面。

多尺度空間模型僅能處理純空間資料，對於隨著時間平緩變化的時空過程，過去的資料可能對時空過程當前的狀態提供極有價值的資訊。例如，衛星監測某環境變量，可能因機械及雲層等因素，導致衛星飛經某區域時，無法獲得可靠的監測資料。欲估計此無資料區域的環境變量，過去的資料可能比此區域週遭當前的資料，更能提供有關的訊息。所以我們進一步推展多尺度空間模型，發展出一種結合時間及空間相關性的動態多尺度空間模型（dynamic multi-resolution spatial model）[2]。

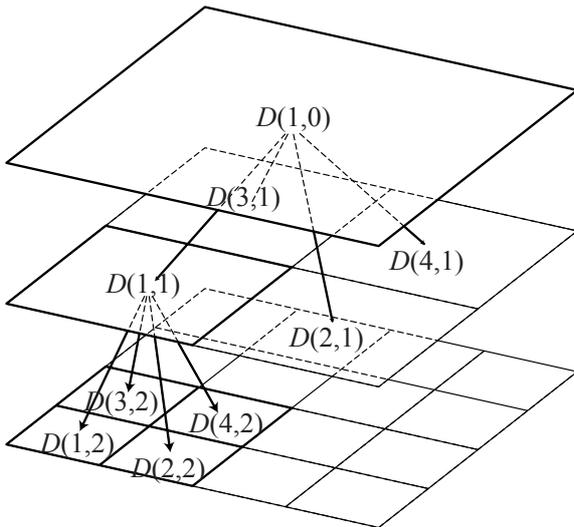
雖然多尺度空間模型相較於傳統的空間預測方法有其統計上及計算上的優勢，然而此模型

亦有一項缺陷，即其空間互變異函數（spatial covariance function）在空間中不連續，因而所產生的空間預測曲面有時可能呈現塊狀的結構。欲改進此缺陷，我們又發展出一種模型平均的方法 [3]，不單採用一個多尺度空間模型，而是採用多個在空間中互為平移關係的多尺度空間模型，並以這些多尺度模型預測曲面的平均，做為最終的預測曲面。因此縱使個別多尺度模型的預測曲面可能呈現塊狀的結構，平均後的曲面依舊是平滑的。我們稱此「multi-resolution spatial」方法為「MURS」。直觀上可能認為 MURS 方法，由於模型平均的計算會衍生嚴重的計算問題，事實上我們探究不同多尺度空間模型間計算上重複的結構，發展出一種演算方法，可以快速地處理大量且在空間不規則分佈的資料。更值得一提的是，我們證明 MURS 方法在空間預測問題具有最佳的統計性質。因此 MURS 方法可說同時達成統計精度與計算效率的兩大目標。

以下各節將詳細說明我們所發展的這些方法。首先於第 2 節中介紹多尺度空間模型，並於第 3 節中簡述動態多尺度空間模型，然後於第 4 節中介紹 MURS 方法。最後於第 5 節中，我們將 MURS 方法應用在分析衛星監測的臭氧資料上。

## 二、多尺度空間模型

首先，我們將有興趣的一個長方形區域  $D \subset \mathbb{R}^2$  分割成規則的長方格  $\{D(i, 0) : i = 1, \dots, n_0\}$ ，做為最粗的尺度（尺度 0）。然後以遞迴的方式將尺度  $j$ （ $j = 0, 1, \dots, J-1$ ）的長方格  $D(i, j)$  分割成  $m_j$  個在尺度  $j+1$  上規則的小長方格  $\{D(ch_k(i, j)) : k = 1, \dots, m_j\}$ ，形成樹狀分枝結構，其中  $ch_k(i, j)$  代表節點  $(i, j)$  的第  $k$  個子節點（child node）。因而在尺度  $j+1$ ，共有  $n_{j+1} = n_0 m_0 m_1 \cdots m_j$  個長方格。如此得到區域  $D$  在  $J+1$  個尺度的一組巢狀分割；見圖一。



圖一 多尺度空間分割之一例，圖中所顯示為在三個尺度下的四分樹

依據區域  $D$  在  $J+1$  個尺度的一組巢狀分割  $\{D(i, j) : i=1, \dots, n_j, j=0, \dots, J\}$ ，我們定義具自迴歸結構的多尺度空間模型。考慮一個二維常態 (Gaussian) 空間隨機過程  $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ ，並假設我們對最細尺度  $J$  的空間集成 (aggregated) 變數有興趣：

$$Y(i, J) \equiv \frac{1}{|D(i, J)|} \int_{D(i, J)} Y(\mathbf{s}) d\mathbf{s}; \quad i=1, \dots, n_j, \quad (1)$$

其中  $|D(i, J)|$  代表  $D(i, J)$  的面積。多尺度空間模型則透過從較粗尺度到較細尺度的自迴歸關係建立：

$$Y(i, j) = Y(pa(i, j)) + w(i, j); \quad i=1, \dots, n_j, j=1, \dots, J, \quad (2)$$

其中  $pa(i, j)$  代表樹狀結構中節點  $(i, j)$  的父代節點 (parent node)， $\{w(i, j)\}$  為平均值為 0 的常態隨機變數，而  $w(i, j)$  與  $Y(pa(i, j))$  呈獨立關係。另外，假設資料  $\mathbf{Z} \equiv \{Z(i, j) : i \in I\}$  在位置  $I \subset \{1, \dots, n_j\}$  觀測，且滿足以下的量測方程式：

$$Z(i, j) = Y(i, j) + \varepsilon(i, j); \quad i=1, \dots, n_j, j=0, \dots, J, \quad (3)$$

其中  $\{\varepsilon(i, j)\}$  為平均值為 0 的獨立常態隨機變數，代表測量誤差。

模型(2)和(3)的最大優點在於  $Y(i, j)$  的最佳空間預測值  $E(Y(i, j) | \mathbf{Z})$  及預測誤差變異數，可藉由快速的演算法 [4] 獲得。此演算法為時間序列 Kalman filter 演算法的推廣，可應用於任何具

樹狀 Markov 結構的常態模型。Huang et al. [1] 發現，運用貝氏 (Bayes) 定理，可以從更簡易的概念推導此演算法，並稱其為 change-of-resolution Kalman filter。此外，[1] 中亦研究滿足尺度間「質量平衡」關係的多尺度空間模型：

$$|D(i, j) | Y(i, j) = \sum_{k=1}^{m_j} |D(ch_k(i, j)) | Y(ch_k(i, j));$$

$$i=1, \dots, n_j, j=0, 1, \dots, J-1$$

附帶一提的是，Huang 和 Cressie [5] 更將 change-of-resolution Kalman filter 演算法推廣到可以處理具有更複雜 Markov 關係的多尺度圖形模型 (multi-resolution graphical model)。

### 三、動態多尺度空間模型

因為過去的資料可能對當前研究的變量提供訊息，我們進一步將時間相關性與多尺度空間模型結合，提出動態多尺度空間模型 [2]。

在此我們考慮一個常態時空隨機過程  $\{Y(\mathbf{s}, t) : \mathbf{s} \in D \subset \mathbf{R}^2, t \in \mathbf{Z}\}$ ，並考慮區域  $D$  在  $J+1$  個尺度的一組巢狀分割  $\{D(i, j) : i=1, \dots, n_j, j=0, \dots, J\}$ 。如同多尺度空間模型，我們想要預測的是在  $t$  時間最細尺度  $J$  的時空集成變數：

$$Y(i, J, t) \equiv \frac{1}{|D(i, J)|} \int_{D(i, J)} Y(\mathbf{s}, t) d\mathbf{s};$$

$$i=1, \dots, n_j, t \in \mathbf{Z}$$

動態多尺度空間模型假設在任意固定時間  $t \in \mathbf{Z}$ ， $\{Y(i, j, t) : i=1, \dots, n_j, j=0, \dots, J\}$  構成多尺度空間模型：

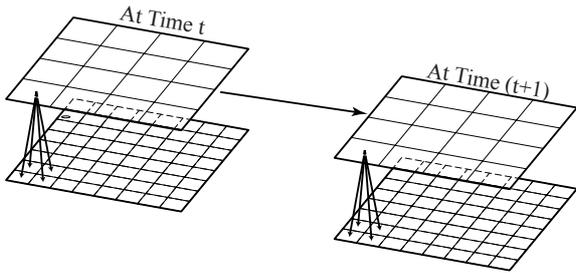
$$Y(i, j, t) = Y(pa(i, j), t) + w(i, j, t);$$

$$i=1, \dots, n_j, j=1, \dots, J, \quad (4)$$

而其時間相關結構則在尺度 0 建構。因為尺度 0 的變量可視為維度  $n_0$  的多變量時間序列  $\mathbf{Y}(0, t) \equiv (Y(1, 0, t), \dots, Y(n_0, 0, t))^T$ ； $t \in \mathbf{Z}$ ，其相關結構可以透過適當的多變量時間序列模型建立。在此我們以一個多變量 AR(1) 模型為例說明：

$$\mathbf{Y}(0, t) = \mathbf{H}(t) \mathbf{Y}(0, t-1) + \boldsymbol{\eta}(t); \quad t \in \mathbf{Z}, \quad (5)$$

其中  $\mathbf{H}(t)$  為  $n_0 \times n_0$  的自迴歸方陣。因為模型(4)和(5)所對應的時空變數關係仍呈現樹狀馬可夫結構 (如圖二所示)，其中  $\mathbf{Y}(0, t+1)$  為此樹狀結構中  $\mathbf{Y}(0, t)$  的子變數(child variable)，所以最佳時



圖二 動態多尺度空間模型之馬可夫關係圖

空預測值及預測誤差變異數仍可透過 change-of-resolution Kalman filter 演算法計算。

#### 四、多尺度空間模型平均

這節中我們將介紹 MURS 方法 [3]。提出此方法的動機在於解決多尺度空間模型在資料較少區域，可能產生不平滑預測曲面的問題。

假設  $\{D(i, j) : i = 1, 2, \dots, j = 0, \dots, J\}$  為  $\mathbf{R}^2$  在  $J+1$  個尺度的一組巢狀分割，在此考慮  $D(i, J)$  為  $1 \times 1$  之方格。定義  $D_m(i, j) = D(i, j) + \mathbf{m}$ ;  $\mathbf{m} \in \mathbf{Z}^2$ ，則對任意  $\mathbf{m} \in \mathbf{Z}^2$ ， $T_m \equiv \{D_m(i, j) : i = 1, 2, \dots, j = 0, \dots, J\}$  亦構成  $\mathbf{R}^2$  在  $J+1$  個尺度的一組巢狀分割。我們將所有不同巢狀分割所形成的集合表示為  $\{T_m : \mathbf{m} \in \mathbf{M} \subset \mathbf{Z}^2\}$ 。值得注意的是，集合  $\mathbf{M}$  中的元素個數通常相當大。

如同第二節，我們考慮一個二維常態空間隨機過程  $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbf{R}^2\}$ ，我們的目的是於預測(1)中的集成變數  $\{Y(1, J), \dots, Y(n_j, J)\}$ 。對於任一個  $\mathbf{m} \in \mathbf{M}$ ，我們依其巢狀分割  $T_m$  定義多尺度空間模型：

$$Y_m(i, j) = Y_m(pa(i, j)) + w_m(i, j);$$

$$i = 1, 2, \dots, j = 1, \dots, J$$

其中  $Y_m(i, J) = Y(i, J)$ ，且  $Y_m(i, j)$  的變異數與  $\mathbf{m}$  無關。假設資料  $\mathbf{Z} \equiv \{Z(i, J) : i \in \mathbf{I}\}$  滿足量測方程式(3)，因此對以  $\mathbf{m} \in \mathbf{M}$  所定義的模型， $Y(i, j)$  的最佳空間預測值為  $\hat{Y}_m(i, j) \equiv E_m(Y(i, j) | \mathbf{Z})$ 。

有別於個別多尺度空間模型的預測曲面，MURS 方法所預測的曲面為  $\mathbf{M}$  中所有多尺度空間模型預測曲面的平均，即  $Y(i, j)$  的預測值可寫為

$$Y^*(i, j) = \frac{1}{|\mathbf{M}|} \sum_{\mathbf{m} \in \mathbf{M}} \hat{Y}_m(i, j),$$

$$i = 1, \dots, n_j, j = 0, \dots, J$$

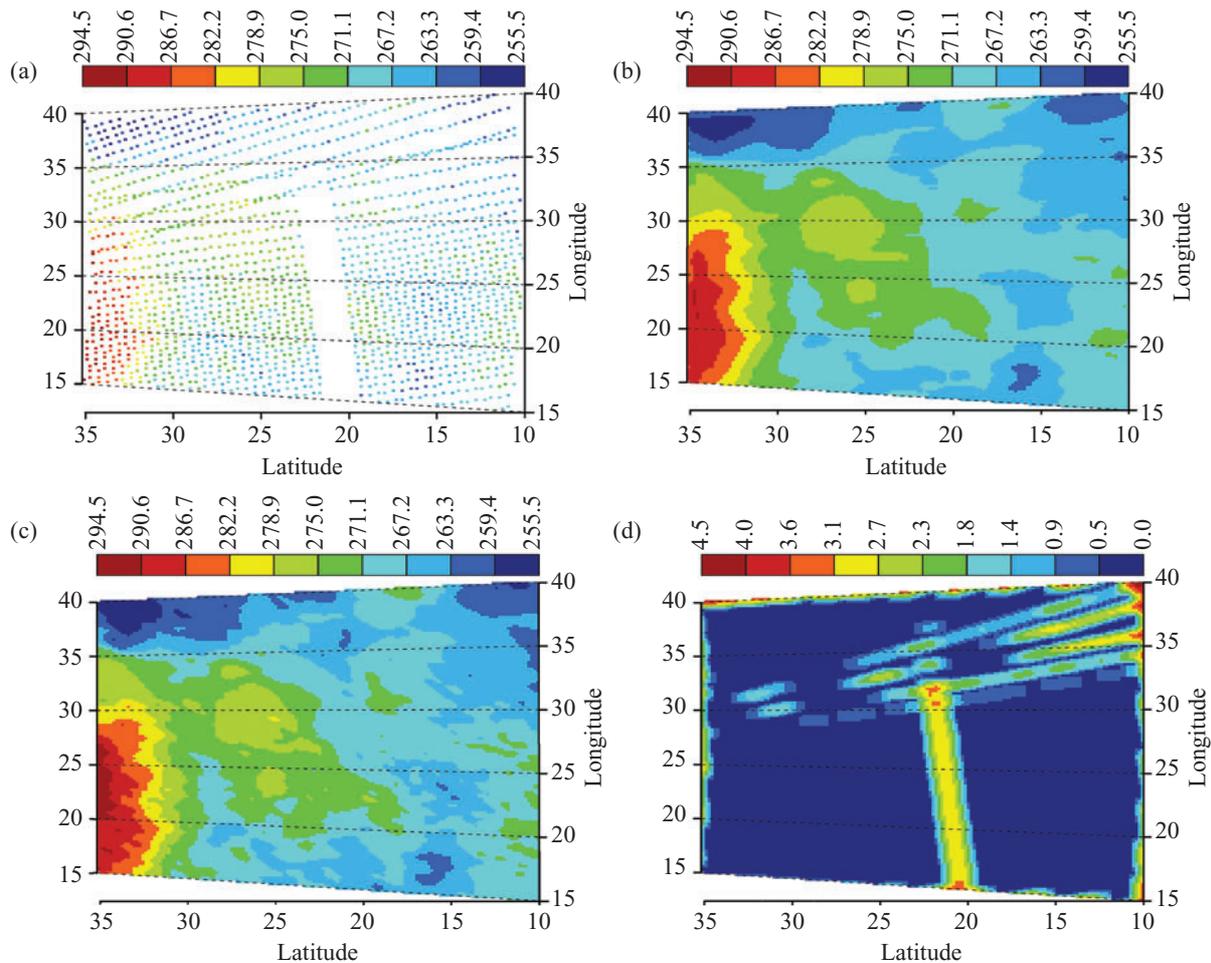
其中  $|\mathbf{M}|$  表示集合  $\mathbf{M}$  之元素個數。乍看之下 MURS 的計算量為多尺度空間模型計算量的  $|\mathbf{M}|$  倍，所以當  $|\mathbf{M}|$  大時可能產生計算上的困難。事實上我們藉由探究  $|\mathbf{M}|$  個 change-of-resolution Kalman filter 間的關係，成功地發展一種快速計算 MURS 預測值及預測誤差變異數的演算法。此外我們證明 MURS 方法在相當廣泛的空間互變異函數結構下，具有最小均方預測誤差的性質。因此 MURS 方法除了產生平滑的曲面外，亦兼具快速及最佳兩項優點，因而同時達到統計精度與計算效率兩大目標。

#### 五、衛星資料之空間預測

接下來，我們將 MURS 方法應用在大氣中臭氧的空間預測問題。我們採用 1988 年 10 月 1 日由 Nimbus-7 衛星上的 total ozone mapping spectrometer (TOMS) 儀器在東經 15-40 度間及北緯 10-35 度間（即撒哈拉沙漠區域）所監測的臭氧資料，共有 1,744 個臭氧觀測值（圖三(a)）。必須注意的是，這筆資料的量並不算大，但是已經瀕臨常用的 kriging 方法在個人電腦上所能處理的限度，我們採用這筆資料的主要原因在於比較 MURS 與 ordinary kriging 兩方法。

首先，我們將此區域分割成  $142 \times 150$  個網格，做為最細的尺度，並以一種交互驗證 (cross validation) 的模型選擇法，得出適當的樹狀結構及模型的參數值。圖三(b)為 MURS 方法所預測的臭氧濃度，其中所選用的樹狀結構，對應 12,800 個多尺度空間模型，因此圖三(b)之預測濃度相當於 12,800 個多尺度空間模型預測平均的結果。

圖三(c)為 ordinary kriging 方法所預測的臭氧濃度。比較圖三(b)與三(c)可看出兩方法的預測濃度極為相近。事實上，從其他的資料分析或模擬實驗結果，亦皆顯示 MURS 與 ordinary kriging 的表現難分軒輊。但是在計算上 MURS 方法完全凌駕於 ordinary kriging 方法；以此資料分析為例，MURS 的計算速度約比 ordinary kriging 快 60 倍。最後，圖三(d)為 MURS 預測誤差的標準差，如所預期的，在資料較稀疏的地方有較大的標準差，而其計算亦可藉由 [3] 中所發展的演算法獲得。



■三 (a) 1988 年 10 月 1 日的臭氧監測值；(b) MURS 方法所預測的臭氧濃度；(c) Ordinary kriging 方法所預測的臭氧濃度；(d) MURS 方法預測誤差的標準差

#### 參考文獻

- [1] H.-C. Huang, N. Cressie and J. Gabrosek, *Journal of Computational and Graphical Statistics*, **11**, 63 (2002).
- [2] G. Johannesson, N. Cressie and H.-C. Huang, Dynamic multi-resolution, *Environmental and Ecological Statistics*, to appear (2005).
- [3] S. Tzeng, H.-C. Huang and N. Cressie, A fast, optimal spatial-prediction, *Journal of the American Statistical Association*, to appear (2005).
- [4] K. C. Chou, A. S. Willsky and R. Nikoukhah, *IEEE Transactions on Automatic Control*, **39**, 479 (1994).
- [5] H.-C. Huang, and N. Cressie, Multiscale graphical modeling in space: applications to command and control. In *Spatial Statistics: Methodological Aspects and some Applications*, (Ed.) M. Moore, Springer-Verlag, New York, 83-113 (2001).