

國立臺灣大學生農學院
生物環境系統工程學系暨研究所
碩士論文

Graduate Institute of Bioenvironmental Systems Engineering College of
Bioresources and Agriculture National Taiwan University
Master Thesis

貝氏最大熵法於 Quantum GIS 上之建構 – 應用於台灣
地區對空氣懸浮粒子長期暴露之研究
Development of Bayesian Maximum Entropy Method Toolbox
on Quantum GIS—An Application of Long-term Exposure
Estimation of Particulate Matter in Taiwan

顧尚真

Shang-Chen Ku

指導教授：余化龍 博士

Hwa-Lung Yu, Ph.D.

中華民國九十九年一月

January 2010

摘要

本研究發展 Quantum GIS 上時空統計函式 – 貝氏最大熵法 – 的插件軟體 (Quantum Bayesian Maximum Entropy Toolbox, QtBME) , 可應用於非定常非同質時空過程之推估與繪圖。在時空過程可分成高頻與低頻之假設下，利用核心平滑法將原時空過程分離為定數的低頻趨勢與定常且同質之高頻時空隨機過程，並用粒子群最佳化演算法與 AIC 準則客觀選取適合高頻時空隨機過程之巢狀共變異數，進而利用貝氏最大熵法推估欲推點的資料特性。藉由 Quantum GIS 的圖形運算能力與內建地理資訊系統函式，QtBME 能輕鬆地展現向量式與網格式二種地理資料格式推估的結果。本研究應用 QtBME 推估台灣地區從 2004 至 2008 年的空氣懸浮粒子 PM_{10} 濃度，結果顯示，北部、雲嘉南、高屏等地為 PM_{10} 濃度聚集的區域。花東、宜蘭則有較低的濃度。另外，暴露在高濃度($>50 \mu g/m^3$)懸浮粒子下的機率有週期性。一般來說，在 3 月開始下降到 7 月後，再由 8 月上升至隔年 2 月。交叉驗證的結果顯示，QtBME 預測的相對誤差大部份落在 $\pm 20\%$ 之內，偶有較高的誤差，為該測站特性與附近所提供的資訊較少所造成。

關鍵詞：貝氏最大熵法，空氣懸浮粒子，Quantum GIS，Python 程式語言。

Abstract

This study developed the Quantum Bayesian Maximum Entropy Toolbox (QtBME), which is a spatiotemporal statistics function, can be applied to estimate and map a non-stationary and non-homogeneous spatiotemporal process under the platform of Quantum GIS (QGIS) software. Kernel smoothing method is used to divide the original process into a deterministic trend and a stationary and homogeneous spatiotemporal process, assuming that a spatiotemporal process can be divided into high and low frequency. The covariance model of the process of high frequency is selected objectively by particle swarm optimization (PSO) method and Akaike's information criterion (AIC). Bayesian maximum entropy method is then applied to spatiotemporal mapping of the variable of interest. By means of ability of geoprocessing as well as graphical computing and mapping in QGIS libraries, QtBME can display the results easily with two types of geographical data format, i.e., raster and vector formats. This study evaluated the long-term township-based exposure estimation of particulate matter (PM₁₀) from 2004 to 2008 in Taiwan. Results showed that PM₁₀ concentration are higher in Taipei, Tainan, and Kaohsiung, and lower in Taidon and Ilan. Moreover, the probability of the high PM₁₀ exposure (i.e., higher than 50 $\mu\text{g}/\text{m}^3$) has strong seasonality; in general, it decreases from March to July and then increases from August to February. The results of cross validation show that QtBME provides satisfactory predictions for the PM₁₀ with relative errors less than 20%. High relative error seldom occurred because of the particular characteristic of certain stations and lack of information provided from the stations in the estimation neighborhoods.

Key Words: Bayesian maximum entropy (BME), Particulate matter, Quantum GIS, Python programming language.

目錄

摘要.....	I
目錄.....	III
圖目錄.....	V
第一章 前言.....	- 1 -
1.1 相關背景介紹.....	- 1 -
1.2 研究動機與目的.....	- 3 -
1.3 本文架構.....	- 4 -
第二章 理論介紹	- 5 -
2.1 時空推估理論.....	- 5 -
2.2 核心平滑.....	- 6 -
2.3 巢狀時空共變異數模式推估	- 7 -
2.4 粒子群最佳化.....	- 8 -
2.5 貝氏最大熵法.....	- 10 -
第三章 QtBME 之軟體開發與介紹.....	- 13 -
3.1 軟體開發的程式選擇	- 13 -
3.2 核心程式.....	- 16 -
3.3 軟體使用介紹.....	- 16 -

第四章 QtBME 於台灣地區 PM_{10} 推估應用.....	- 34 -
4.1 資料來源.....	- 34 -
4.2 處理流程.....	- 34 -
第五章 結果與討論	- 38 -
5.1 趨勢與殘差.....	- 38 -
5.2 巢狀共變異數函數.....	- 39 -
5.3 向量式出圖形態.....	- 40 -
5.4 網格式出圖形態.....	- 42 -
5.5 交叉驗證.....	- 44 -
第六章 結論與建議	- 47 -
6.1 結論.....	- 47 -
6.2 建議.....	- 48 -
參考文獻.....	- 50 -

圖目錄

圖一 QtBME-GUI 啟動方法	17
圖二 QtBME-GUI 的主要頁面	17
圖三 資料的輸入	19
圖四 資料輸入格式說明視窗	19
圖五 資料空間套疊參數設定	20
圖六 空間套疊結果	21
圖七 資料的 (圖左) 地理特性 , (圖右) 統計特性	21
圖八 Trend 分頁	22
圖九 Trend 分頁計算結果	23
圖十 Plot & Covariance 分頁	24
圖十一 共變異數圖形	25
圖十二 推估的巢狀共變異數函數圖形(圖上)與結果(圖下)	26
圖十三 Spatiotemporal Estimate 分頁	27
圖十四 程式警告視窗	29
圖十五 向量式出圖結果	30
圖十六 網格式出圖介面	31
圖十七 網格式出圖結果 (左) 平均值 (右) 變異數	32

圖十八 交叉驗證結果	- 33 -
圖十九 研究流程圖	- 35 -
圖二十 空氣懸浮粒子共變異數圖形	- 38 -
圖二十一 空氣懸浮粒子共變異數圖形與擬合(3D).....	- 39 -
圖二十二 空氣懸浮粒子共變異數圖形與擬合(2D).....	- 40 -
圖二十三 空氣懸浮粒子於向量式出圖結果	- 41 -
圖二十四 2004 年 PM_{10} 的顯著性隨月份變化趨勢	- 42 -
圖二十五 空氣懸浮粒子於網格式出圖結果(平均值)	- 43 -
圖二十六 空氣懸浮粒子於網格式出圖結果(變異數)	- 43 -
圖二十七 空氣懸浮粒子交叉驗證結果	- 44 -
圖二十八 相對誤差較大的測站(紅圈與黃圈).....	- 45 -

第一章 前言

1.1 相關背景介紹

空氣中的污染物有許多種，例如、懸浮粒子、二氧化硫、臭氧等。其中懸浮粒子(又稱氣膠)會對大自然環境跟人類造成很大的影響，包括能見度的降低、酸沉降的增加、以及對人類健康危害的風險增加等等[1,2]。粒徑在 10 微米以下的粒子，由於粒徑較小，能深入人體肺部深處造成呼吸系統之危害，其主要來源包括道路揚塵、汽機車排放廢氣、工地營建等[3]。台灣地區的交通與工業逐漸發展至今，空氣品質的維護與空氣污染的防制也越來越受到重視。約民國 70 年，環保署著手於空氣品質監測站的架設，於民國 83 年趨於完整。

一般來說，觀察環境中空氣懸浮粒子的時空分佈，可將其視為隨著空間與時間位置而改變的一種連續隨機過程。相較於以往定數的模擬方式，隨機過程引入了機率的概念，使用機率統計的方式去描述變數的時空分佈。對資料做不確定性的描述也更符合現實生活中我們所碰到的情況。而此隨機過程可以拆解成低頻與高頻的二個部份，低頻的部份，稱為趨勢(Trend)函數。類似於平均值的概念，指的是此隨機過程的分佈趨勢，這部份仍屬於定數的。高頻的部份，稱為時空自相關的殘差項(Residual)，也就是扣除高頻之後的部份[8]。此部份即是

引入機率描述的部份，屬於一種隨機的過程。

地理資訊系統(Geographic Information System, GIS)是屬於一門綜合性的學科，可應用在很多不同的領域。以地理空間資料庫為基礎，採用地理模型分析方法，適時提供多種空間的和動態的地理信息，為提供地理研究和地理決策的電腦技術系統，包含收集、記錄、修正、分析、統計、查詢等等的功能。地理資訊系統從外部來看，它表現為電腦軟硬體系統，而其內涵是由電腦程式和地理數據組織而成的地理空間資訊模型，是一個邏輯縮小、高度資訊化的地理系統[11,15]。

地理資訊的資料通常包含兩個種類，即圖形資料與屬性資料。結合這些資料，便可進一步提供更多有用的資訊，例如：空間分析、拓撲結構等等。有了這些資訊，則更能對我們所關心的細節部份，做出詳細的資料分析與空間展示。

由於現今電腦的發達，提供了許多軟體來幫助實現地理資訊系統的實作。常見的軟體有 ArcGIS、GRASS、MapInfo、Quantum GIS 等等，這些軟體都包含基本的地理資訊系統的圖形處理功能，其之間也有些許的不同。ArcGIS 可以說是目前國內最常使用的地理資訊系統軟體，內含非常豐富的內建函式與插件，但需要付費才能使用 ArcGIS 所提供的功能。GRASS 是一套功能強大且免費的軟體，可處理網格式、向量式資料和影像處理，非常好用。MapInfo 的優點則是在於使

用方便而且價格低廉。支援多種資料輸入與投影系統。Quantum GIS 則是免費且方便的一套軟體，本身即有強大的向量式圖形處理函式，內建結合 GRASS 做網格式圖形的分析。也提供許多應用程式介面可讓使用者自行開發所需要的插件，開發過程因為可以使用 Python 程式語言，使得開發與維護相對的容易許多。

1.2 研究動機與目的

地理資訊系統可以把觀察分析的結果以一種簡潔、清楚的方式呈現，而時空的推估方法可以將觀察分析的結果做更進一步地分析。

一直以來，有關時空分析與推估的方法不斷進步，在分析後的結果呈現方式卻大都相同，且後續的處理需依賴其它的軟體。必須透過多次的資料轉換或傳遞，相當的不方便。

而資料的呈現方式即是地理資訊系統的強項之一。地理資訊系統能提供非常多元的資料呈現方式，在操作上相對簡單。更重要的是，地理資訊系統還可對呈現的資料，進行之後與空間相關的分析。

然而，地理資訊系統仍有它不足的地方。例如：大多僅處理空間相關的分析、相對於很多時空推估、分析的研究來說，其分析的方法簡單許多。若能夠結合強大的時空推估分析理論與多元化的資料呈現方式做更進一步的資料分析是非常方便實用的。

本研究希望能在地理資訊系統方面，加入強而有力的時空分析工

具，配合其原有的強大資料展示與其簡便快速的掌握空間資訊方式，來研究與時空相關的環境特性。

本研究將以行政院環保署所提供之 2004 至 2008 年的懸浮粒子監測資料，運用貝氏最大熵法來觀察台灣地區空氣懸浮粒子時空分佈的趨勢與結構特性。

1.3 本文架構

本研究共分為六個章節，第一章為相關背景介紹、研究的動機與目的以及本研究的架構。第二章是相關理論的介紹，包含貝氏最大熵法、粒子群最佳化演算法等等。第三章是 QtBME 軟體開發與使用的相關介紹，舉出為何選擇 Python 程式語言作為主要開發程式以及 QtBME 詳細的使用說明。第四章是 QtBME 運用的實例，以台灣地區 2004 至 2008 年的空氣懸浮粒子濃度的時空分佈為例子。包含資料處理與處理的流程。第五章是探討整個推估過程的細節成果，探討其中分離低頻與高頻是否成功、巢狀共變異數是否最具代表性，也探討出圖結果的特性與模式的交叉驗證。第六章則是提出結論與建議，作為本研究的總結與提供未來可繼續發展的方向。

第二章 理論介紹

2.1 時空推估理論

一個在時空中連續的隨機過程，我們以 $Z(s, t)$ 表示。其中 s 是空間位置， t 是時間位置。將其拆解成低頻與高頻的二個部份，如下式：

$$Z(s, t) = m(s, t) + \varepsilon(s, t)$$

其中 $m(s, t)$ 屬於低頻部份， $\varepsilon(s, t)$ 屬於高頻部份。通常低頻的部份被視為一種非穩定(Nonstationary)、非同質(Nonhomogeneous)的隨機過程。高頻則是有某種程度自相關的穩定隨機過程[13]。

推估低頻部份的方法可分為兩大類—有母數與無母數的推估方法。有母數的推估方式需要對此隨機過程有一定程度的認識與了解，然而有時依據這些背景知識卻不容易被取得。

反過來說，對於無母數的推估方式，資料的背景知識就不是那麼地重要。也就是說，無母數的推估方式可適用於各種資料的推估。

為了能夠發展通用的推估方式以適應各種不同的資料，本研究選用無母數的推估方式來推估低頻的部份。

高頻的部份則是需要推估其自相關性函數，常用的推估函數有半變異數、共變異數等等。本研究因為貝氏最大熵法函式庫的參數需要，選用共變異數模式來推估高頻的自相關性。

2.2 核心平滑

核心平滑法(Kernel smoothing method)是一種在統計中無母數的推估方法。本研究中將使用核心平滑來推估低頻部份。

基本上，核心平滑的作法是將欲推估點的觀測值視為與其鄰近點資料的距離權重平均。這樣的作法可以將資料平滑化，以推估資料的平均趨勢。

令欲推估點為 P_0 ， P_i 為第 i 筆資料， N 為資料的總數， Z 與 \hat{m} 分別為觀測值與欲推估點的觀測值(趨勢)[14]。則其公式可表示如下：

$$\hat{m}(P_0) = \frac{\sum_{i=1}^N K(P_0, P_i) Z(P_i)}{\sum_{i=1}^N K(P_0, P_i)}$$

其中 K 為核心函數，在本研究中，二維(空間與時間軸)的核心函數可表示為：

$$K(P_0, P_i) = f_s\left(\frac{Dis_s(P_0, P_i)}{B_s}\right) f_t\left(\frac{Dis_t(P_0, P_i)}{B_t}\right)$$

f_s 、 f_t 為一權重型函數，具有以下特性：

1. 其函數值皆為非負實數。
2. 通常為對稱圖形。
3. 由函數對稱中心為最大值，往兩旁遞減至 0。

$Dis_s(P_0, P_i)$ 、 $Dis_t(P_0, P_i)$ 分別代表 P_0 與 P_i 的空間與時間的幾何距離。

B_s 、 B_t 為空間與時間的核心頻寬，通常用來決定核心函數在該維度往

兩旁遞減至 0 的距離。

核心函數有很多種類，例如均匀型(Uniform)、三角型(Triangular)、高斯型(Gaussian)、二次方型(Quadratic)、四次方型(Quartic)等等，可依資料的特性做選擇。

2.3 巢狀時空共變異數模式推估

由前述，我們可以利用核心平滑的方式，將資料分為低頻與高頻，高頻部份的自相關性函數便是使用共變異數模式來推估。

巢狀時空共變異數模式(Nested spatiotemporal covariance model)是共變異數模式的一種。令為 $C_{st}(h, \tau)$ ，下標 s 指空間， t 指時間， h 是空間距離， τ 是時間距離。則可由下式描述：

$$C_{st}(h, \tau) = \sum_{l=1}^N b_l C_{sl}(h; A_{sl}) C_{tl}(\tau; A_{tl})$$

N 為巢狀個數， b_l 為門檻值(sill)， $C_{sl}(h; A_{sl})$ 與 $C_{tl}(\tau; A_{tl})$ 分別表示純空間與純時間的共變異數模式，常用的有高斯型、指數型(Exponential)等等。 A_{sl} 與 A_{tl} 則是 C_{sl} 與 C_{tl} 所需使用的參數，一般稱為此共變異數的範圍(range)。式中的下標 l 則表示為第 l 個時空共變異數模式。由上式可看出，巢狀時空共變異數模式是由數個純時間與純空間的共變異數模式相乘，再將其相加組合而成。

本研究使用赤池信息量準則(Akaike information criterion, AIC)判斷選用模式的好壞。如下式：

$$AIC = 2k + n \ln(RSS / n)$$

其中 k 為參數個數， n 為資料筆數， RSS 為殘差的平方和(Residual sum of squares)。愈小的 AIC 值，代表所選出的模式能用最少的參數來合理解釋數據[5]。其中 RSS 的計算方式，因考慮到共變異數之間的相對重要性，本研究加入權重方法來計算誤差平方和(Weighted RSS, WRSS)。強調愈大的共變異數值將有愈大的重要性[10]。如下式：

$$WRSS = \sum_{i=1}^n \left(C'_{st,i} \right)^2 \left(C'_{st,i} - C_{st,i} \right)^2$$

$C'_{st,i}$ 為第 i 筆資料實際的共變異數值。其算法為將所有資料中，屬於第 i 組時空延遲(lag)範圍內的資料挑出，再由此資料計算共變異數做為代表。

2.4 粒子群最佳化

粒子群最佳化(Particle swarm optimization, PSO)是一種模擬一群粒子，根據群體智慧在其搜尋空間搜索出目標函數最佳解的演算法。這樣的方法最早是在 1995 年由 James Kennedy 與 Russell C. Eberhart 所提出[9]。

群體智慧的基礎在於社會心理學的原則，並對社會上的行為提出見解。如人們在解決問題時，會請教別人。而別人的回答，或多或少都會造成人們的認知、態度、甚至是行為的改變。社會中的形形色色對個體的認知與種種產生影響，這就是群體智慧的觀念基礎。

粒子群最佳化即模擬在多維空間中有一群粒子，它們有初始的位置與初始的速度，在空間中搜索目標函數的最佳解。根據群體智慧的觀念，這些粒子會記憶它自己最佳解的位置，與所有粒子中最佳解的位置。前者稱為局部最佳解(Local best)，後者則稱為群體最佳解(Global best)。接著粒子跟粒子間分享自己與群體最佳解的位置，而適當地調整粒子的位置與速度。到整個模擬過程結束，群體最佳解即為此演算法的最佳解。

在 n 維中，第 i 顆粒子的位置表示為 $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ，經過適應後的位置為 $\hat{x}_i = (\hat{x}_{i1}, \hat{x}_{i2}, \dots, \hat{x}_{in})$ 、速度表示為 $v_i = (v_{i1}, v_{i2}, \dots, v_{in})$ ，經過適應後的速度為 $\hat{v}_i = (\hat{v}_{i1}, \hat{v}_{i2}, \dots, \hat{v}_{in})$ 。其局部最佳解表示為 $LB_i = (LB_{i1}, LB_{i2}, \dots, LB_{in})$ 。群體最佳解表示為 $GB = (GB_1, GB_2, \dots, GB_n)$ 。則粒子群最佳化的控制方程式可表示如下：

$$\begin{aligned}\hat{v}_i &= w \cdot v_i + c_1 \cdot Rand() \cdot (LB_i - x_i) + c_2 \cdot rand() \cdot (GB - x_i) \\ \hat{x}_i &= x_i + \hat{v}_i\end{aligned}$$

其中 w 為慣性權重(Inertia weight)， c_1 和 c_2 為加速常數(Acceleration constants)， $Rand()$ 、 $rand()$ 為 $[0,1]$ 範圍內的隨機值。對於參數的不同選擇即對應到不同想法的粒子群最佳化演算法，例如將 w 設為零即是不考慮慣性的作用，相對的加強適應部份的權重。

一般標準的粒子群最佳化演算法流程觀念如下：

1. 初始化一粒子群，其中每顆粒子的速度與位置為在某一給定範圍

內的隨機值。

2. 利用每顆粒子的位置計算目標函數值。
3. 比較每顆粒子的目標函數值與局部最佳解的目標函數值，若目標函數值較佳，則以當前粒子的位置取代局部最佳解。
4. 比較每顆粒子的目標函數值與群體最佳解的目標函數值，若目標函數值較佳，則以當前粒子的位置取代群體最佳解。
5. 根據其演算法的控制方程式，利用局部最佳解、群體最佳解等參數，對每顆粒子的位置與速度做適應性的改變，得到新的位置與速度。
6. 重覆步驟 2 到 5，直到演算法結束，傳回群體最佳解。

本研究將使用粒子群最佳化演算法來搜尋最佳的巢狀時空共變異數模式。

2.5 貝氏最大熵法

貝氏最大熵法(Bayesian maximum entropy, BME)是一種近代時空模式與繪圖(Modern spatiotemporal modeling and mapping)的方法。可自動整合非線性的推估技巧與非高斯的機率密度函數理論，也可以處理不同種類與尺度的資訊與不確定性的資料型態[12]。

在本研究中，空氣中懸浮粒子濃度的時空分佈可視為時空隨機場。場中某個空間時間點 $p = (s, t)$ 的濃度值可令為 $X_p = X_{s,t}$ 。而對於場

中每個點的濃度值 X_p ，我們使用機率密度函數 f_{KB} 來描述。可用下式描述：

$$P_{KB} \left[\chi_1 \leq X_{p1} \leq \chi_1 + d\chi_1, \dots, \chi_k \leq X_{pk} \leq \chi_k + d\chi_k \right] = f_{KB}(p_1, \dots, p_k) dX_{p1} \cdots dX_{pk}$$

下標 KB 代表用來建構的此機率密度函數的“知識根據”(Knowledge base)。

對於知識根據，貝氏最大熵法將其分為兩大類，即 $G - KB$ 與 $S - KB$ 。 $G - KB$ 是指全面性的知識根據(General KB)，記為 G ，包含一些物理、生物的定律跟一些與空間時間相關的時空理論模式(例如：平均值、共變異數、半變異數、高階動差等等)，適用於整個觀察區域。 $S - KB$ 是指特定點的知識根據(Site-specific KB)，記為 S ，包含確定性資料(Hard data)與不確定性資料(Soft data)這兩種資料型態。確定性資料就是觀測點的值為一確切的值，不確定性資料則是使用機率密度函數來描述觀測點的值。而總知識根據 $K - KB$ (Total KB)，記為 K ，即是將全面性與特定點的知識根據結合，也就是 $K = G \cup S$ 。

貝氏最大熵法的基礎理論方程式如下式(相關的細節討論可以參考[6,7])：

$$\left. \begin{aligned} \int d\chi (g - \bar{g}) e^{\mu^T g} &= 0 \\ \int d\chi \xi_S e^{\mu^T g} - A f_K(X_p) &= 0 \end{aligned} \right\}$$

g 為一向量，代表所有由全面性知識根據而來的函式 $g_\alpha (\alpha = 1, 2, \dots)$ 。 \bar{g} 表示 g 的統計期望值。 μ 也為一向量，代表所有與空間時間位置、 g 相關的所有係數 μ_α ，可以將其視為對全面性知識根據的函式之間的相對重要程度。 ζ_S 是特定點知識根據的運算子(Operator)。 A 是標準化參數。 f_K 是每個空間時間中所有每一個點的機率密度函數，下標 K 表示此機率密度函數是基於總知識根據而來的。式中的 g 與 ζ_S 是需輸入的部份， μ 與 f_K 則是未知待定的變數。

因為 f_K 是空間時間中所有每一個點的機率密度函數，所以欲推估的點 p_k 也包含在 f_K 之中。只要得到欲推估點的機率密度函數，便可以依研究上的需要，推估在該點上的某些統計特性，例如：平均值，變異數、高階動差等等。

第三章 QtBME 之軟體開發與介紹

3.1 軟體開發的程式選擇

本研究使用的程式語言是依據以下原則作為考慮條件：

1. 免費的開發軟體設計程式。
2. 軟體易於程式設計者開發與維護。
3. 軟體易於一般使用者使用，提供良好的圖形化使用者介面。
4. 高效能的運算能力與資料圖形展示。

並有以下幾點主要開發目的：

1. 盡可能以方便一般使用者的使用為主。
2. 可將資料分析的結果整合地理資訊系統做展示與分析。
3. 資料分析的結果可繪圖展現。

在與地理資訊系統整合的軟體方面，選擇 Quantum GIS 作為開發主要軟體。

Quantum GIS 是一個實用、程式容量輕巧的免費軟體，提供許多地理資訊系統常用的函式[21]。例如：環域分析(Buffer)、交集(Intersect)、差集(Difference)。也有許多強大的函式。例如：附加元件(Plugins)中的「幾何校正工具」(Georeferencer)可輸入影像檔案並直接進行幾何校正。內插工具(Interpolation)可對向量圖層的某項屬性進行

內插，並有多種內插方法、內插參數可供選擇。

Quantum GIS 提供友善的圖形使用者介面，讓不熟悉指令式程式的使用者也能輕易上手。中央研究院人社中心地理資訊研究專題中心的同仁也為 Quantum GIS 提供中文化的使用者介面 軟體簡介資訊與基本的操作說明，讓使用者更易於上手[24]。

Quantum GIS 也適合程式開發者在其使用介面上開發附加元件，除了可以使用官方提供的應用程式介面(Application programming interface, API)，還可任意執行自行開發的客製化程式。

對於附加元件的開發，Quantum GIS 提供 C++與 Python 二種程式語言作為撰寫工具[20]。由於 Python 程式較易於附加元件的開發與維護，且在官方所提供的附加元件中有「Python 主控台」，可直接使用 Python 語言的指令來即時控制 Quantum GIS 的使用介面，故選用 Python 程式語言作為開發附加元件的撰寫工具[22]。

另外，Quantum GIS 是用 C++程式語言實作而來，運算效能也頗為出色。

而上文中所提到的 Python，是最近幾年開始流行起來的一套功能強大、直譯式、物件導向的程式語言，其優點甚多，包括簡單、易學、易讀、易開發、易與其它程式連結、跨平台、擁有眾多實用的函式庫、被許多知名廠商使用、完全免費等等[17,23]。

在選擇撰寫程式方面，若考慮到軟體執行的效能，則應以 C++ 程式語言作為撰寫程式較好。而考慮到軟體開發與維護的難易，則選擇 Python 為撰寫語言較為適合。此二種語言各有其優缺點，且能互相補足，故本研究將兩者都做為撰寫的程式。其中 Python 適合於界面開發與維護(甚至是開發與 C++ 程式的連結)，C++ 則用來實作大量或重覆簡單的數值計算。

將 Python 與 C++ 語言結合運用是達到軟體運算高效能、介面易於開發與維護的最佳方式。

考慮撰寫 Quantum GIS 軟體中附加元件的使用介面，即圖形化使用者介面(Graphical user interface, GUI)。本研究使用 PyQt4 模組，此為 Qt 在 Python 應用的模組，Qt 也是使用 C++ 所開發，有免費與商業用的二種版本，其提供許多常見的 GUI 元件與易於開發 GUI 的 Qt designer 軟體可供使用，使得附加元件在開發與效能上皆有相當好的表現[18,19]。

本研究在資料特性展現，也就是二維與三維的繪圖部份，使用 matplotlib 模組。matplotlib 是一個 Python 的繪圖模組，提供非常多的繪圖函式與指令，而且可以輕易地修改圖形座標軸、圖形標題等許多繪圖展現所用到的參數[16]。

綜合以上開發程式的選擇，包含 Quantum GIS、Python 與其相關

的模組(PyQt4、matplotlib)、C++，就可以客製化一套在 Quantum GIS 上的附加元件，來實行本研究所希望達到的研究目的。

3.2 核心程式

本研究的核心程式分為以下三大部份：

1. 由 C++ 撰寫 BME 演算法的動態函式庫(BMEprobaMoments.dll)。
2. 由 C++ 撰寫 Python 的動態函式庫(SEKSGUI.pyd)。
3. 由 Python 撰寫的 Quantum GIS 附加元件(QtBME.py)。

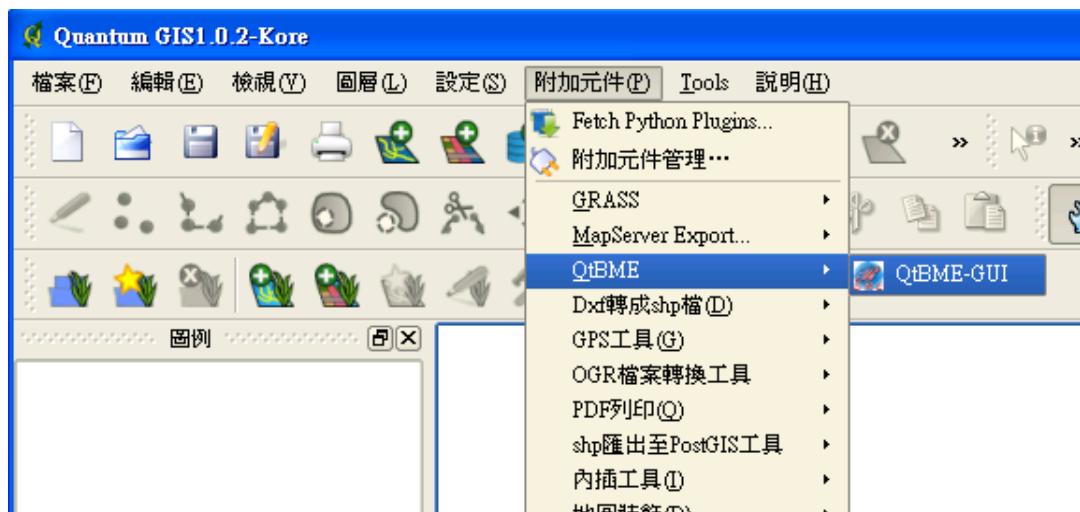
其中 BMEprobaMoments.dll 包含所有 BME 演算法所需要的函式。由余化龍指導教授開發，與本人共同維護及編譯。

SEKSGUI.pyd 則包含了計算 BME 演算法與前置函式、也包含大部份的數學計算函式。由本人開發與維護。

QtBME.py 則是在 Quantum GIS 上開發客製化的附加元件---QtBME 的主要程式碼。主要是用來架構圖形化的使用者介面，與少部份的計算函式。由本人開發與維護。

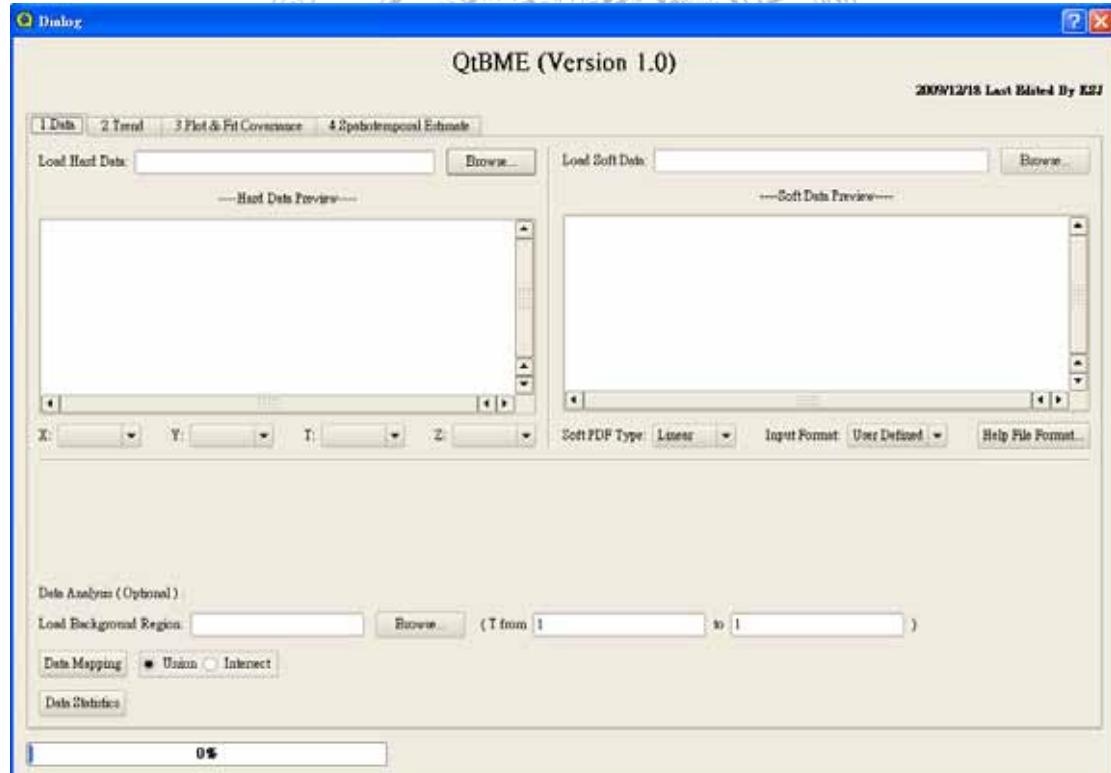
3.3 軟體使用介紹

開啟 Quantum GIS，點選如下圖畫面中，“附加元件(P)”選單下的 QtBME 自製插件，再點選之中的 QtBME-GUI 將其開啟。



圖一 QtBME-GUI 啟動方法

開啟後可以看到下圖的主要頁面。



圖二 QtBME-GUI 的主要頁面

QtBME-GUI 的主要頁面是由“1.Data”、“2.Trend”、“3.Plot & Fit Covariance”、“4.Spatiotemporal Estimate”等四個分頁所組成，各別處理不同分析步驟，以下將對這四個分頁進行介紹與使用說明。

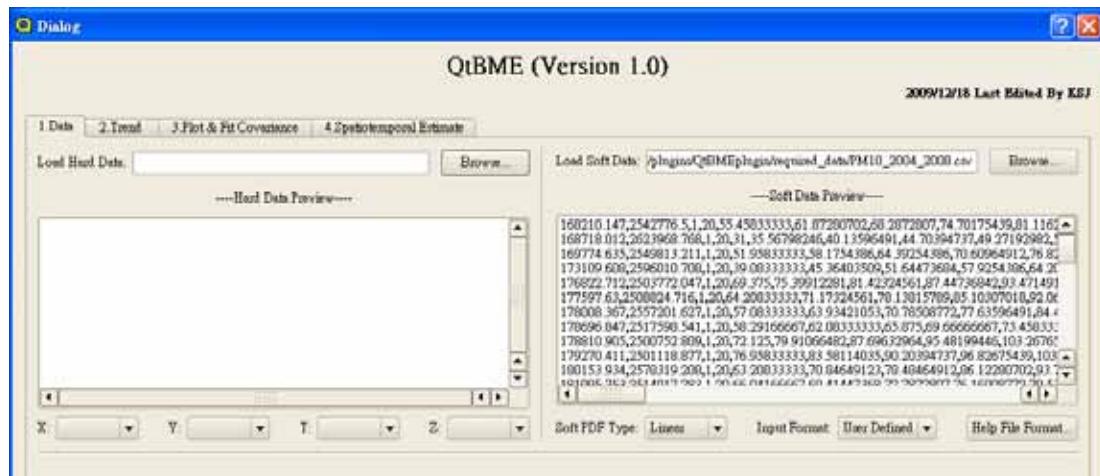
1. Data 分頁

如上頁圖，此分頁的功能是處理資料的輸入與初步分析，包含圖中上方的確定與不確定資料輸入介面、與下方附加的資料的空間座標與資料統計介面(“Data Analysis(Optional)：“)，資料輸入介面的檔案格式皆為逗號分隔型式(.csv)，空間座標與資料統計介面即是一些簡單的鍵入參數。

確定性資料需輸入包含空間與時間座標、觀測數值等四個參數，輸入檔案之後可由下圖下拉式選單選取對應到的資料參數。不確定資料可以由使用者選擇不同的機率密度函數類型(Soft PDF Type)、格式(Input Format)。類型方面有“Linear”與“Histogram”二類可供選擇，表示不同的機率密度函數描繪方式。“Linear”表示機率密度函數圖形為線性連接。而“Histogram”則是用水平連接來表示機率密度圖形。格式方面有“User Defined”“Gaussian”“Uniform”等三種格式可以選擇，分別為“使用者定義型”、“高斯型”、“均勻型”，表示不同的機率密度描述的方法。其中特別的是“使用者定義型”，可讓使用者自行決定機率密度的圖形，非不是常見的高斯圖形。詳細的描述方

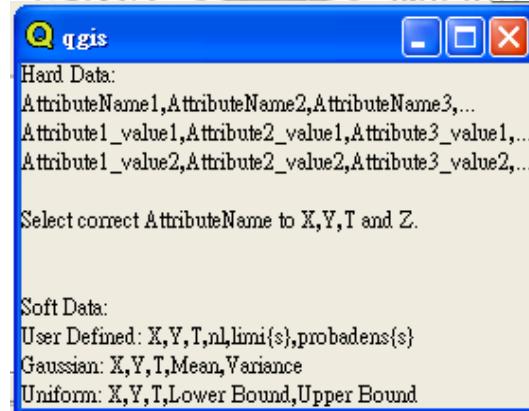
法可由“Help File Format”顯示。

例如使用者按下“Browse...”按鈕尋找要輸入的資料檔案，輸入後將有預覽資料如下圖所示。



圖三 資料的輸入

而點選“Help File Format”將出現如下的視窗可供參考。



圖四 資料輸入格式說明視窗

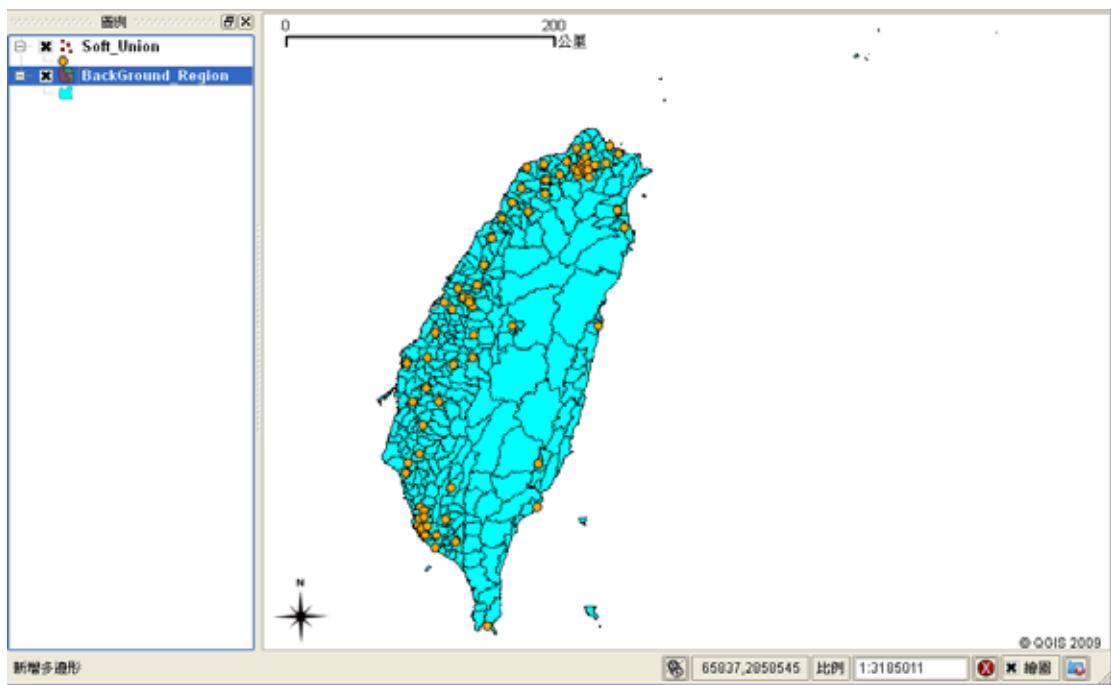
空間座標介面為選擇性的功能。需輸入想套疊的地圖檔案(.shp)、時間範圍、與套疊方式，即會將介面上方輸入的資料座標，依輸入的時間範圍取出，配合選取的套疊方法將資料繪於 Quantum GIS 上。套疊方法中，聯集(“Union”)表示只要在選取的時間範圍中有出現過的測站點皆會標示在圖上，而交集(“Intersect”)則表示測站點必須在所有的時間中皆有出現才會標示。

例如點選介面下方的“Browse...”，選擇台灣地圖(TWN_TOWN.shp)做為想套疊的地圖，時間範圍輸入 1 到 20，套疊方式點選聯集、如下圖。



圖五 資料空間套疊參數設定

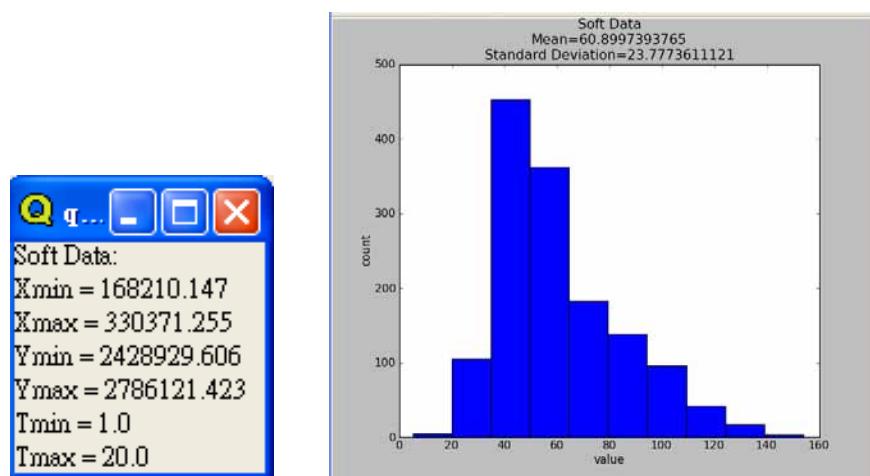
接著點選“Data Mapping”按鈕即在 Quantum GIS 下出現下頁圖結果。



圖六 空間套疊結果

圓點即為範圍內中所有包含的測站點空間座標。

資料統計介面，僅需輸入時間範圍，會計算時間範圍內輸入資料的一些地理與統計特性，地理特性以小視窗顯示、統計特性以組體圖展現。按下“Data Statistics”按鈕後，即出現下圖兩視窗。

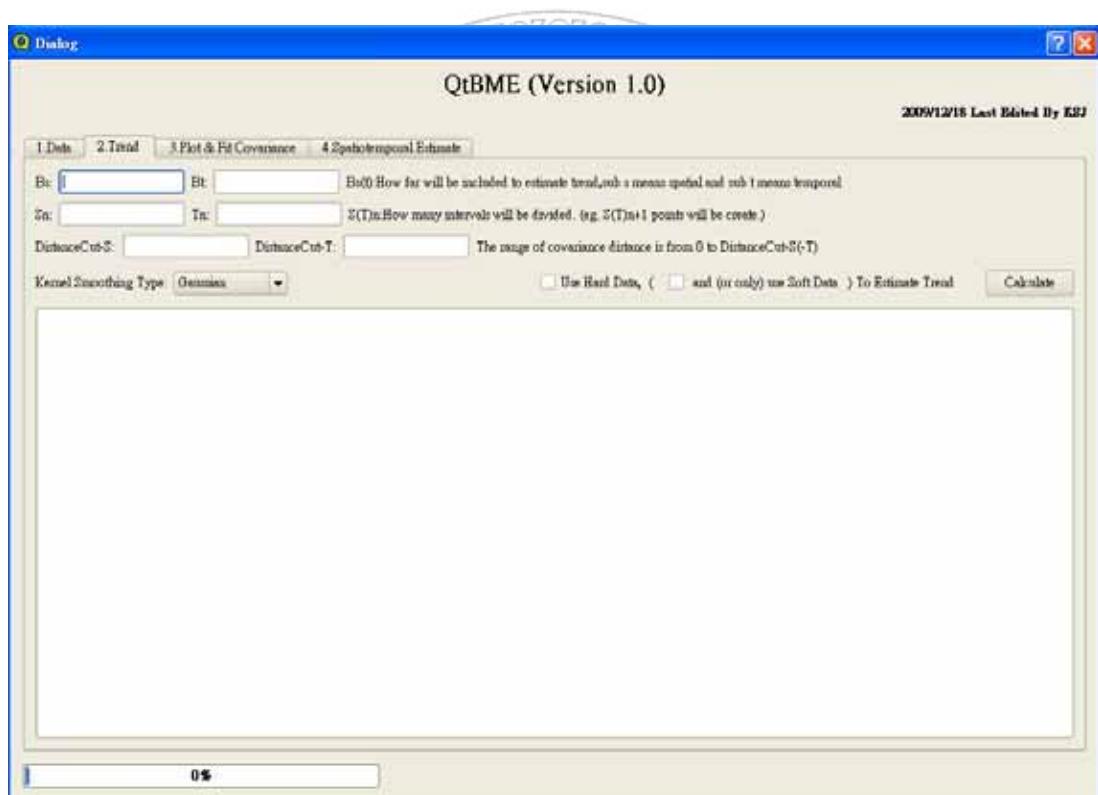


圖七 資料的 (圖左) 地理特性 , (圖右) 統計特性

地理特性視窗中顯示資料的時空位置邊界，統計視窗即展現此時間範圍內的資料值的統計特徵，包含資料分佈情形、平均值、標準差。讓使用者在須對資料進行初步了解時可以選擇使用。

2. Trend 分頁

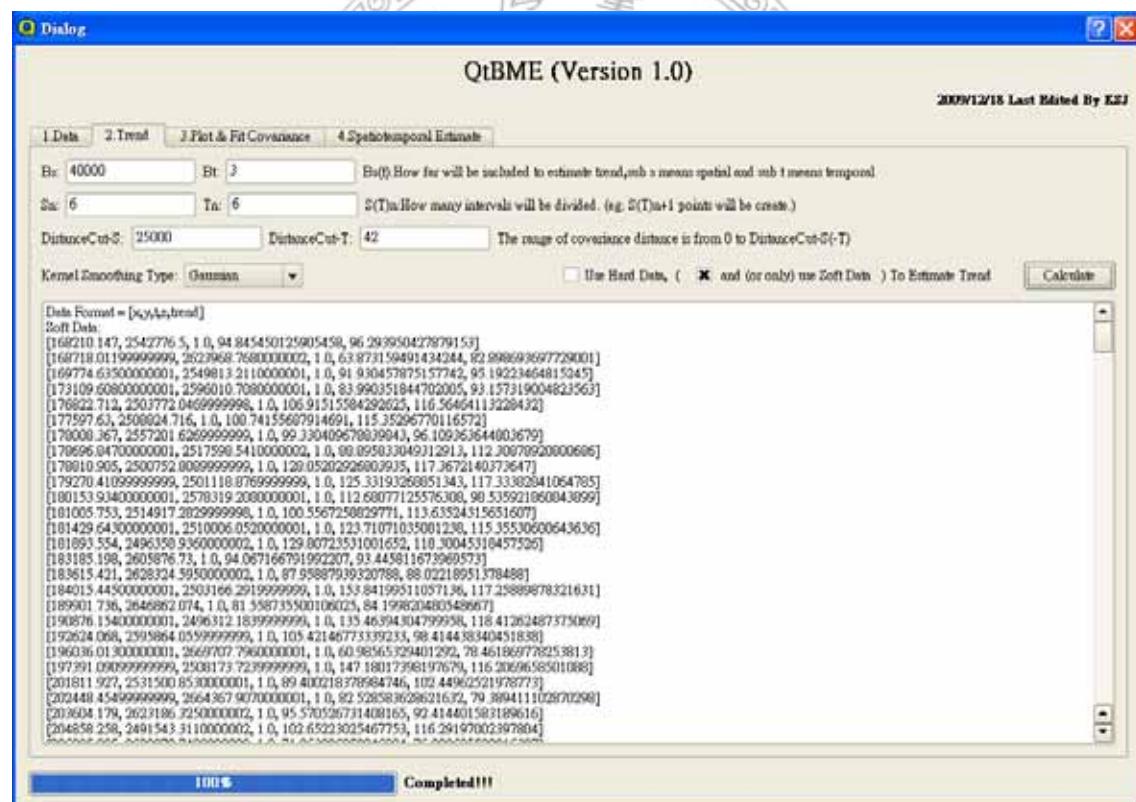
此分頁是將資料觀測值分解為低頻與高頻，以便分別對低頻與高頻的部份做推估。介面如下：



圖八 Trend 分頁

圖中有很多需要輸入的參數，Bs 與 Bt 分別表示核心平滑法中的空間、時間的核心頻寬。Sn 與 Tn 即是將資料的高頻部份做共變異數計算時的統計參數，指將資料在空間與時間上各分成幾個間隔(也就

是會有間隔數+1 個點)來描繪共變異數圖形。DistanceCut-S 與 DistanceCut-T 是選取共變異數在空間與時間想要觀察的延遲範圍。Kernel Smoothing Type 提供高斯型與二次方型的核心函數可供選擇。兩個選取方框“Use Hard Data,”、“use Soft Data”即是考慮到資料的不確定性，讓使用者可以自由選擇適合的資料形態來推估趨勢。例如輸入完所需參數後，按下“Calculate”按鈕即進行趨勢推估，完成後會出現結果如下圖。

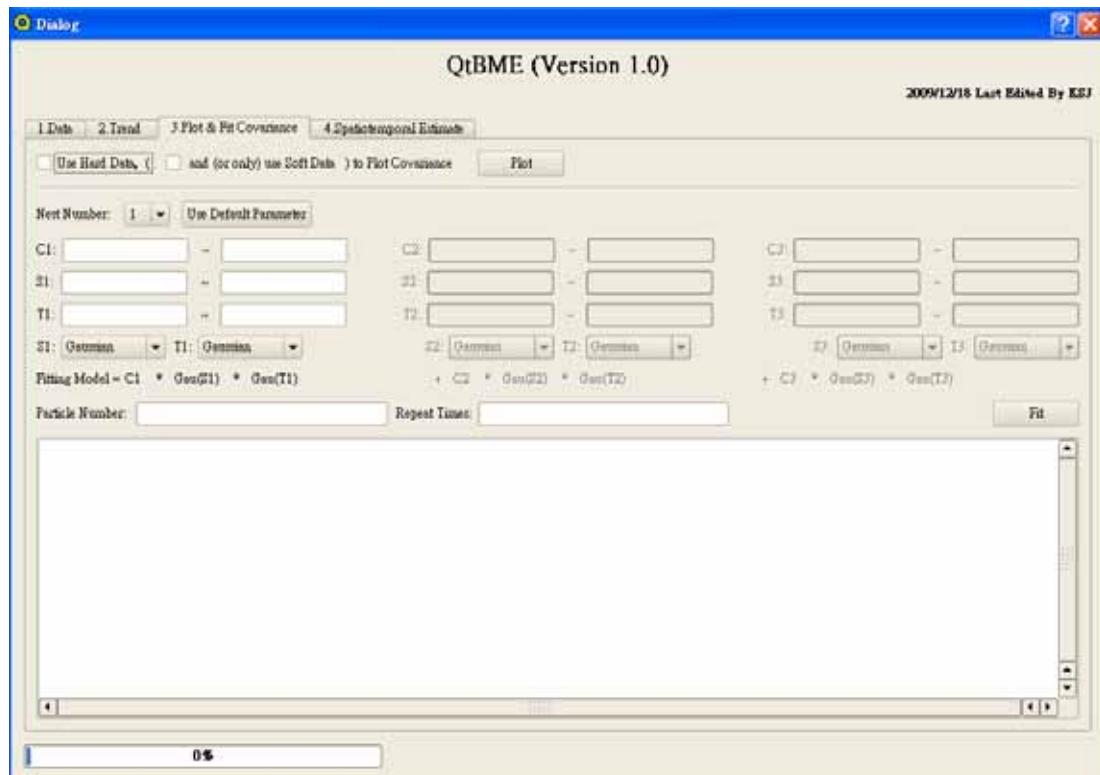


圖九 Trend 分頁計算結果

預覽表格中包含原始的資料值與推估的趨勢。

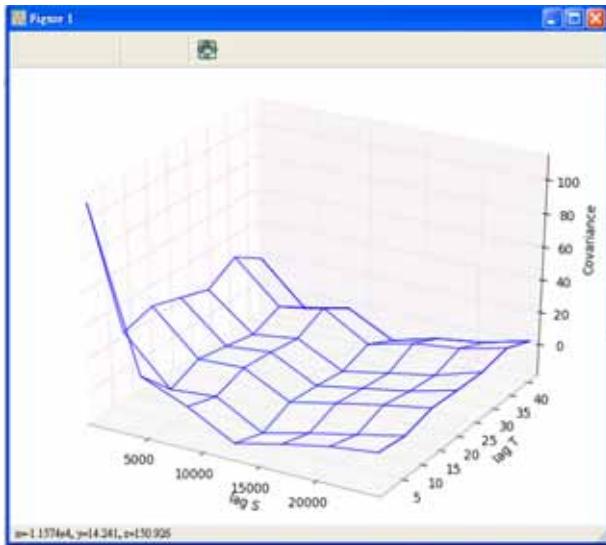
3. Plot & Fit Covariance 分頁

此分頁是將計算好的高頻部份經由之前的統計參數計算共變異數並以 3D 圖形呈現(介面上部)，接著使用粒子群最佳化演算法推估巢狀共變異數函數(介面下部)。其介面如下圖所示。



圖十 Plot & Covariance 分頁

同 Trend 分頁的道理，選擇適當資料型態後即可點選“Plot”按鈕進行共變異數計算與作圖，結果如下視窗：



圖十一 共變異數圖形

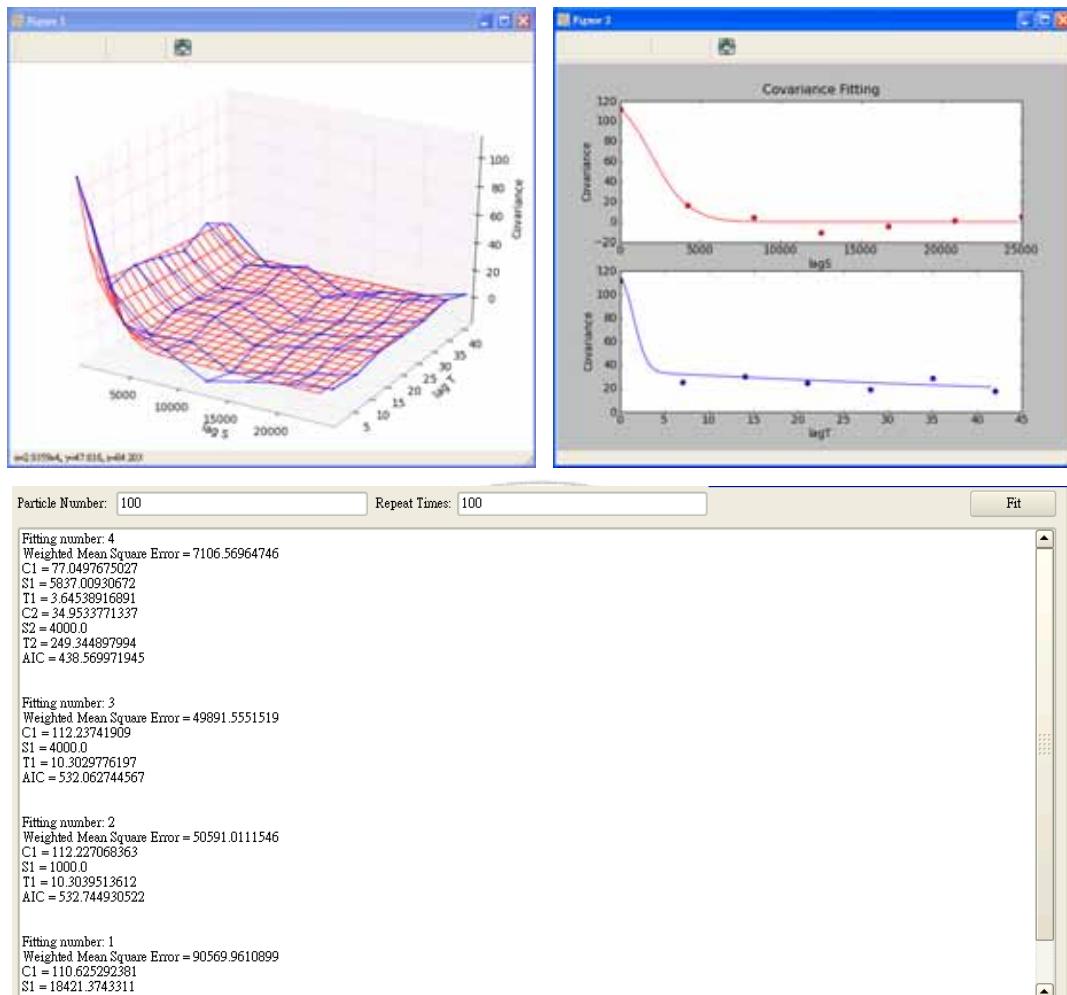
此為可旋轉的座標軸圖形，可任意旋轉圖形至適合範圍觀察。視窗中圖形即為在空間時間延遲平面上的共變異數曲面。

接著即可由粒子群最佳化演算法推估巢狀共變異數函數。此處需輸入包含共變異數的模式型態與粒子群最佳化演算法參數。巢狀個數參數有 1~3 可以選擇，為時空共變異數模式的個數，共變異數模式有高斯、指數、球型、洞正弦(Holesin)、碎塊(Nugget)型等五種模式可以選擇。粒子群最佳化演算法則是設定每個共變異數的變數求解空間與粒子數量、疊代次數等等的參數。

軟體中有提供“Use Default Parameter”按鈕進行設定，會將資料中最大的共變異數值、二點空間時間最大距離的 0.01~100 倍分別與對應的對數設為求解空間範圍，可供參考使用。

選取巢狀個數與相關參數後，按下“Fit”按鈕就會進行推估。推估

後結果如下圖。



圖十二 推估的巢狀共變異數函數圖形(圖上)與結果(圖下)

上圖上部左方為推估後的巢狀共變異數函數與實際數值的擬合情形，藍色曲面為實際資料，紅色曲面為推估結果。

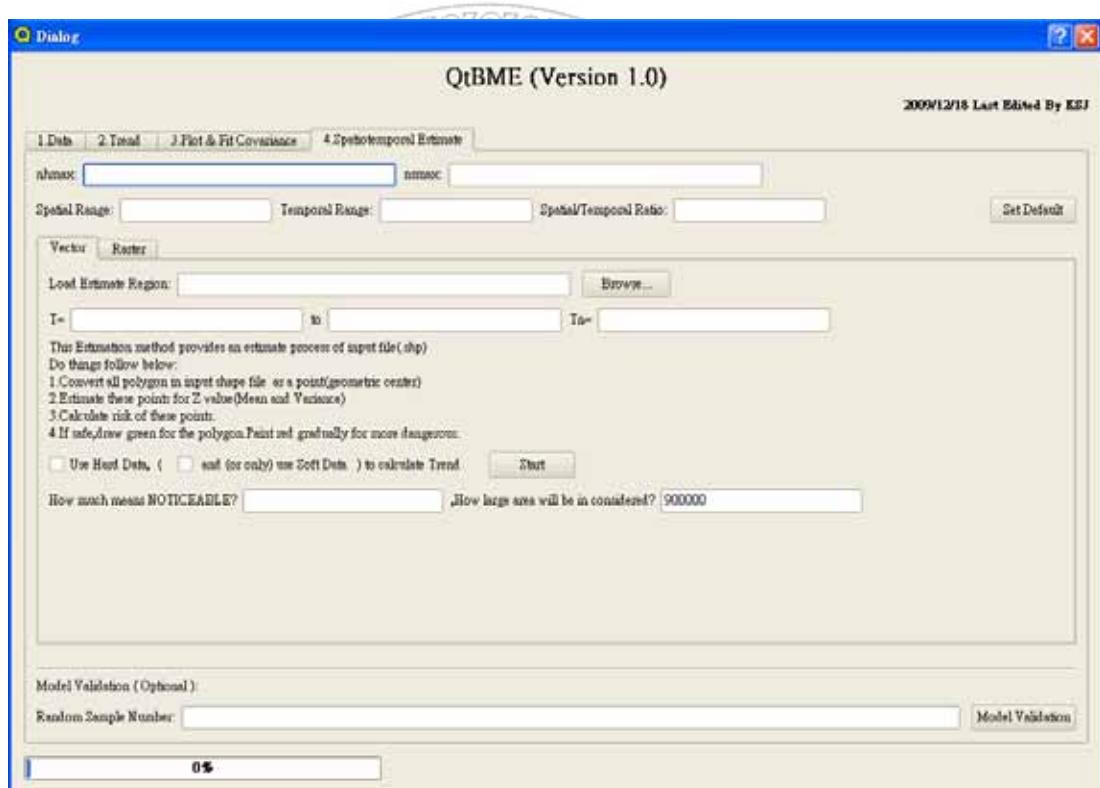
上圖上部右方為先只考慮空間延遲(圖片上方)或只考慮時間延遲(圖片下方)的 2D 巢狀共變異數函數。實心圓點為實際資料值，曲線為推估結果。

上圖下部為推估的結果記錄，包含推估編號(Fitting number:)、權

重後的方均差、各個共變異數模型的參數、與 AIC 值，可作為巢狀共變異數函數推估好壞的準則。

4. Spatiotemporal Estimate 分頁

此分頁即運用貝氏最大熵法計算欲推估點的特性與 Quantum GIS 不同的出圖形態結合、展現，最後還有提供模式驗證的功能。其介面如下圖。



圖十三 Spatiotemporal Estimate 分頁

上方須輸入貝氏最大熵法所需的參數，包含推估時所納入的最大資料筆數(nhmax、nsmax)、最大搜索半徑(Spatial Range、Temporal Range)與時空距離的比值(Spatial/Temporal Ratio)。右上方有提供參考

值的設定，利用“Set Default”按鈕，程式會將 nhmax 設為 15，nsmax 設為 5，Spatial Range 設為資料空間中最大二點距離的 1/3，Temporal Range 設為巢狀共變異函數中最重要的共變異模式裡的時間範圍，而 Spatial/Temporal Ratio 則設為巢狀共變異函數中最重要的共變異模式裡的空間範圍除以時間範圍。

接著是選取不同的 Quantum GIS 出圖型態，不同的出圖形態配合其原有特性，有不同的推估過程，敘述如下：

向量式(Vector 分頁)

為程式預設選項，需輸入預推估的區域檔案(.shp 檔)，程式會取得輸入檔案中每個地區圖形的形心作為欲推估點的空間座標。接著輸入時間範圍，例如：T= 10 to 13 Tn=3，即從時間點 10 到 13 之間切為三等間隔，也就是將在 10, 11, 12, 13 四(Tn+1)個時間點上各自進行資料推估。

推估的結果會得到時空座標上欲推估點的高頻特性(平均值與變異數)。配合低頻部份的推估(可選擇不同資料型態)，將高頻低頻結合，即可得知欲推估點的推測值(也為平均值與變異數)。再配合使用者輸入的資料顯著值(“How much means NOTICEABLE?”)與面積過濾門檻(“How large area will be in considered?”，預設值 900000 為台灣鄉鎮市圖中的面積門檻，可依需求改變)，按下“Start”按鈕後，程式將

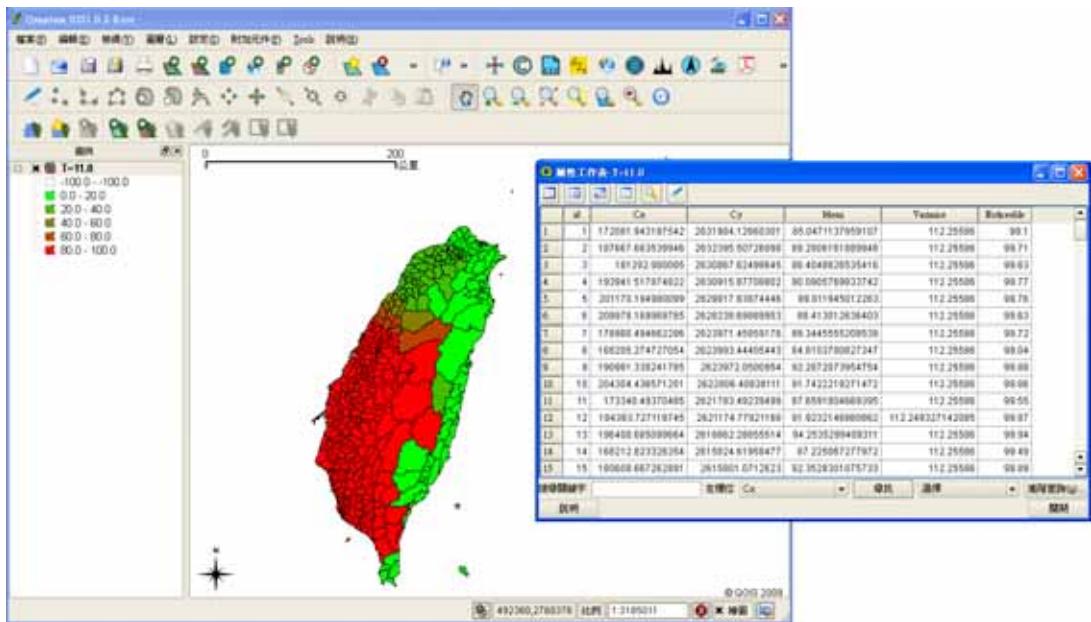
正式運行。

由於計算量可能較大，程式會出現注意視窗提醒使用者耐心等待，請勿按下其它按鈕，否則程式會“沒有回應”(雖然仍在執行)。



圖十四 程式警告視窗

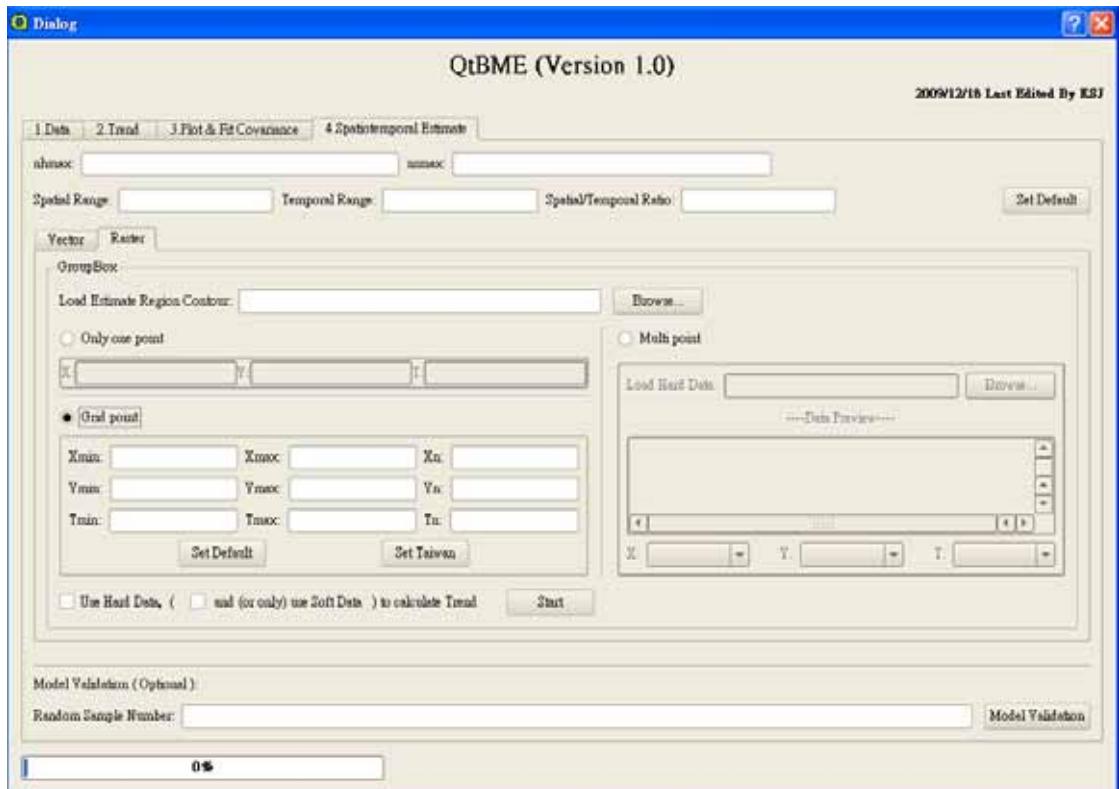
程式將考慮面積大於面積過濾門檻值的地區圖形，假設其推估點值的機率密度函數為常態分佈，利用平均值與變異數計算超過資料顯著值的機率。在 0~100% 之間以 20% 為間隔分成五等分，將地區圖形依顯著機率由低到高填滿由綠到紅的顏色。如果推估點資料有 NaN 值出現，即無法計算，區塊將以白色填滿且將顯著值設為 -100，最後在 Quantum GIS 下以向量式圖形與屬性表(需額外開啟)展現(如下頁圖)。



圖十五 向量式出圖結果

網格式(Raster 分頁)

首先必須按下“Browse...”按鈕輸入欲推估區域的輪廓，為.shp 檔案(之後做遮罩用)。接著是建立欲推估點，欲推估點的時空座標必須是網格式的(下頁圖中的“Grid point”選項)



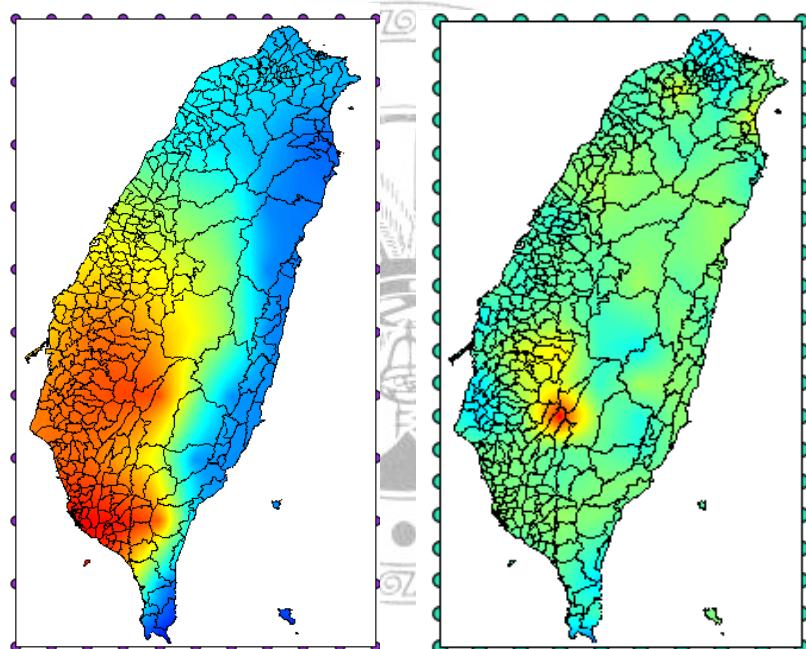
圖十六 網格式出圖介面

輸入空間時間的邊界(X/Y/Tmin, X/Y/Tmax)與間隔數(X/Y/Tn)即可創造出網格式的欲推估點。“Set Default”按鈕會自動搜尋輪廓檔案中的圖形邊界設為空間座標邊界，間隔為 10 等份(11 個點)。時間則是預設為 1，單一時間點。“Set Taiwan”與“Set Default”按鈕類似，僅空間座標邊界不同，顧名思義，即是台灣本島的空間座標。

設定完成並選定推估低頻的資料形態後，按下“Start”按鈕程式就會正式運作。

程式會使用與向量式相同的推估方法，推估網格上所有的資料點，取得該點的平均值與變異數。接著取得各個時間點上所有點(扣

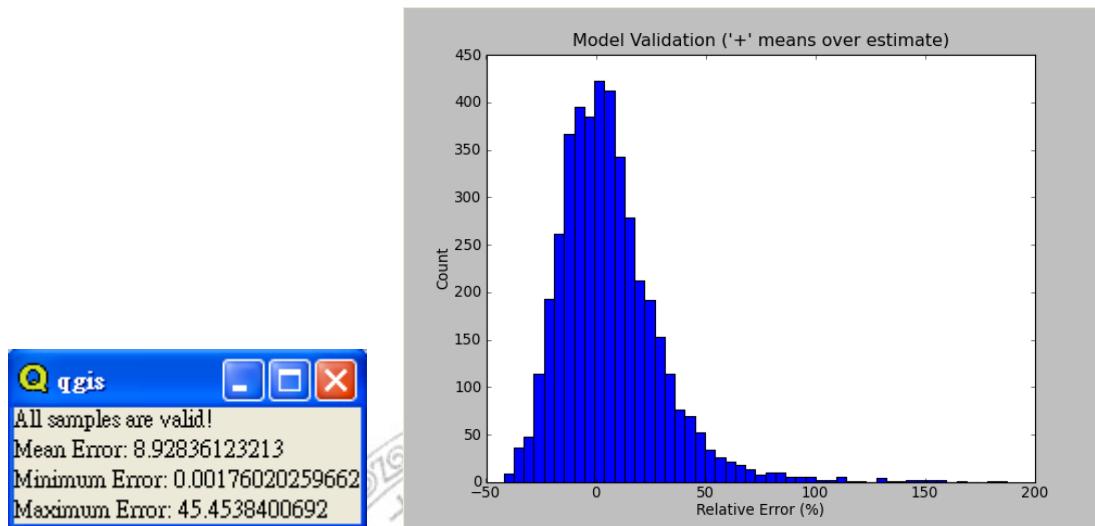
除 NaN 值)的各個空間邊界，依此邊界，以時間點的值為命名依據建立平均值與變異數的 GTiff 的灰階圖檔(較長的邊為 512 像素，像素為正方形，1 波段)。將每個像素依自然鄰近值內插法(Natural neighbor interpolation)填入其值(將推估值做線性轉換至 0~255 以符合圖檔需要)，以擬彩色(Pseudo color)的方式呈現(較低的值為藍色、中間的值為綠色、較高的值為紅色。如下圖。



圖十七 網格式出圖結果 (左) 平均值 (右) 變異數

當推估完成，可以選擇性地輸入採樣數(Random sample number)，是交叉驗證所取的資料筆數。按下“Model Validation”按鈕可進行交叉驗證(Cross validation)。此處使用的交叉驗證為留一驗證法(Leave-one out)，即將需驗證的資料取出一個當成未知點，再由其

它資料點推估該未知點，以驗證資料模型的準確性。結果會輸出大小二個視窗如下圖：



圖十八 交叉驗證結果

圖左交叉驗證的資料特性，有最大、最小誤差、資料可用性等等。第一行表示是否有推估點無法推估(圖中為所有推估點皆可推估)，在推估點時空附近沒有其他資料點時(無資訊納入)會造成無法推估的情形。平均誤差即是將實際值與推估值相減後取絕對值再取平均。誤差最大值與最小值即是在實際值與推估值相減後的絕對值中的最大值與最小值(所以都是正數)。圖右則是顯示相對誤差可作為模式準確與否的參考，橫軸是相對誤差，算法為推估值減去實際值後，再除以實際值，圖上單位為百分比。縱軸為累積的資料筆數。

第四章 QtBME 於台灣地區 PM_{10} 推估應用

4.1 資料來源

本研究使用的空氣懸浮粒子(PM_{10})資料係由行政院環保署的空氣品質監測網(<http://taqm.epa.gov.tw/taqm/zh-tw/default.aspx>)所下載。

本研究使用包含全台灣所有的測站共 73 站, 截取從 2004 至 2008 年共 5 年的長期小時資料做為原始資料。將原始資料依每天至少有超過 18 筆(75% 資料可用率)的小時資料平均, 轉為日資料。接著將日資料轉為每月數值的不確定性資料(將日資料轉為月資料的機率密度函數)。扣除不可用的資料與未測量的測站資料後, 剩下 4327 筆可用的月資料做為研究使用。

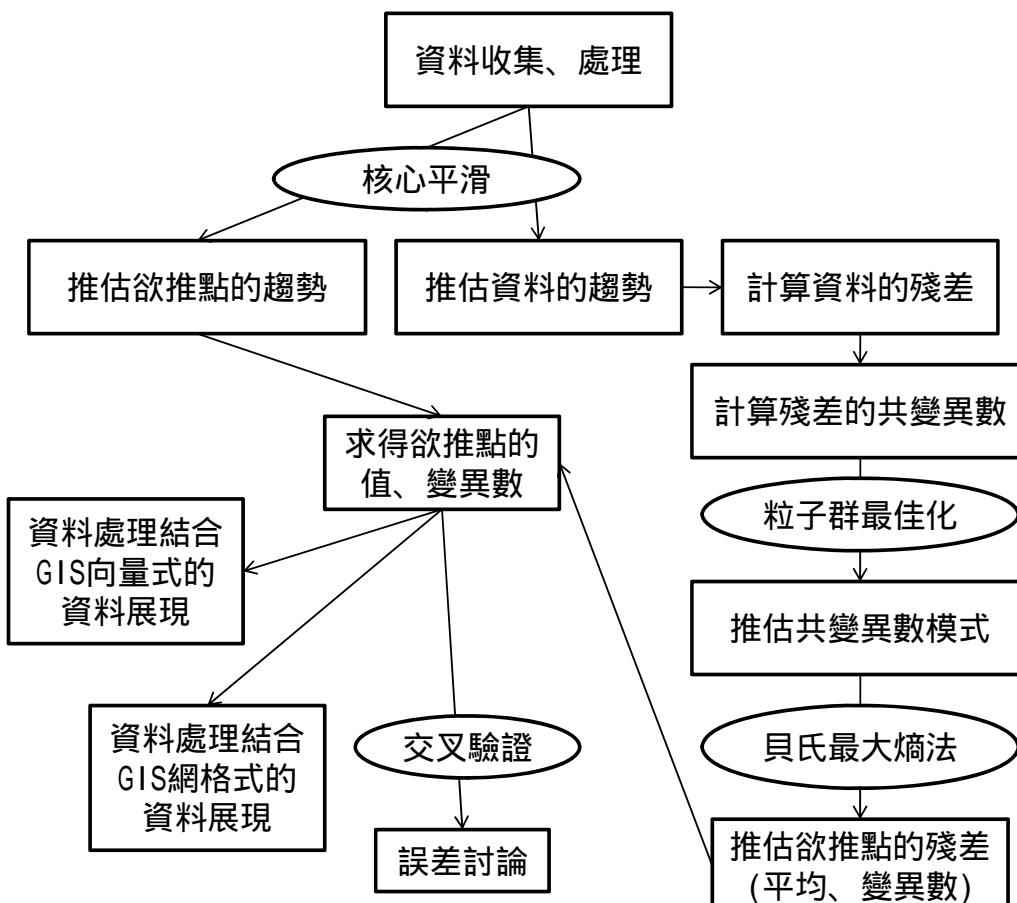
4.2 處理流程

本研究的研究流程圖如下頁：

首先將收集到的原始資料進行處理。由於是每小時的資料, 必須先將其轉為天資料、再轉為月資料以供利用。

一般來說轉換資料的方式大部份是使用平均值的方式, 將小時資料轉為天資料, 然後再轉換一次得到月資料。原因是因為在一個資料點上只能有“一”個確切的值, 若要考慮到機率, 也是假設為高斯常態分佈。但貝氏最大熵法卻能接受在該點上的值為一個非高斯的機率

密度函數。所以本研究希望在小時資料轉天資料時使用平均值的方法，而在天資料轉為月資料時，使用天資料做為月資料的機率密度函數。如此便能使資料更符合現實的情況。



圖十九 研究流程圖

接著在核心平滑部份，核心頻寬為決定低頻部份的主要因素，取的太大會造成資料不平滑，太小則會造成過度的擬合。本研究中選取頻寬是依據推估後的共變異數圖形是否達到穩定性與同質性來決定。通常取值過小會讓共變異數圖形很快地隨著延遲增加而遞減至

0，無法有效的決定共變異數函數的擬合。取值過大會造成圖形二軸末端沒有趨於穩定與同質，不符合共變異數的假設。所以必須先做第一次的猜測，再由結果做修正。而在計算共變異數圖形之前，共變異數統計的參數也有一定的重要程度， S_n 與 T_n 取的太大，會造成共變異數圖形不平滑，取太小又看不出圖形走向。這部份通常是選取 5~10 個等份，在這些結果中選取較適合的計算結果。於是便可得到資料的趨勢與相對應的共變異數圖形。

在擬合共變異數圖形時，會先直接用預設的參數進行推估，已便先看出共變異數圖形在各別軸上的分佈情形，利於選擇共變異數模式。都常圖形下降較快的是使用高斯型的來擬合、較慢的是使用指數型。球型因其在頻寬後就降至為零，不像高斯、指數型在無限遠處仍有值存在，故適合於共變異數模式的疊加。而洞正弦型(Holesin)則可以擬合共變異數圖形裡有週期性的部份，碎塊型則處理碎塊效應(Nugget effect)。

有了共變異數的擬合函數，再配上先前得到核心平滑參數跟原始資料，就可以使用貝氏最大熵法來做計算了。貝氏最大熵法所需要的參數是使用預設功能來給定，而推估點則是台灣區域，視使用向量式或網格式的推估來決定確切範圍，對於向量式，欲推點就是各鄉鎮的形心中心，顯著參數以環保署空氣污指標值(PSI)為“良好”至“普

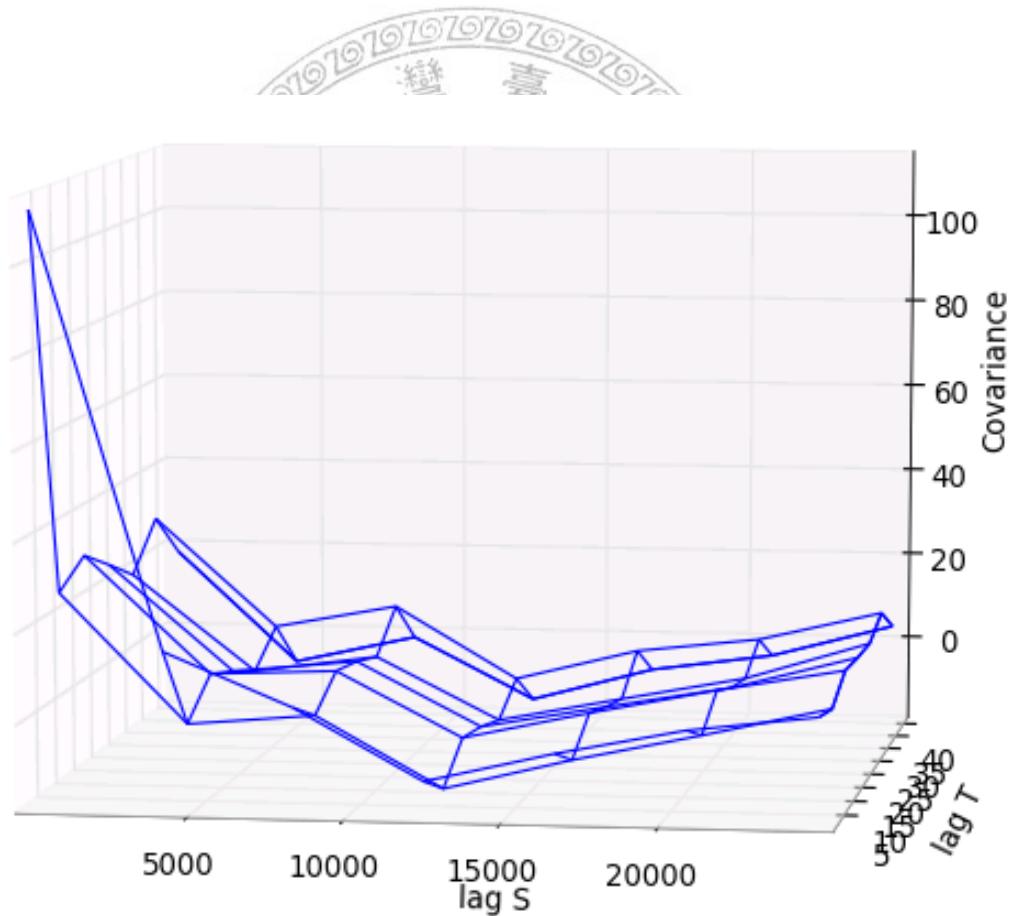
通”的中間值 50 所對應到的 PM_{10} 濃度值 $50 \mu g / m^3$ 。面積過濾參數則是設為台灣鄉鎮中最小的面積以下，設定面積過濾參數可以讓一些較小、較不重要的面積區域不做推估，加速程式的運行速度。對於向量式，欲推估點由於台灣南北長、東西短，本研究將台灣東西邊界為 10 個點、南北方向為 20 個點，讓點與點之間的距離在方向上不要差異太大。推估時間的選取則是用最靠近時間中間的部份的二個時間點，也就是研究資料時間的中值，這樣較不易發生在二側時間點上推估點資訊較少的情形。交叉驗證部份，由於資料點總數過多，較理想的方式是隨機選取幾個點來做驗證，再對誤差做分析討論。



第五章 結果與討論

5.1 趨勢與殘差

本研究皆使用不確定性的空氣懸浮粒子濃度資料形態。在設定核心平滑相關參數 $B_s = 40000$ 、 $B_t = 3$ 、 $S_n = 6$ 、 $T_n = 6$ 、 $DistanceCut-S = 25000$ 、 $DistanceCut-T = 42$ 、Kernel Smoothing Type = Gaussian 下，將資料分成趨勢與殘差，利用殘差繪出共變異數圖形如下圖。



圖二十 空氣懸浮粒子共變異數圖形

可以看到在空間與時間都沒有延遲的共變異數為最高，隨著空間

延遲增加，其值漸漸減少至 0 附近，而隨著時間延遲增加，其值漸減至 20 左右。代表殘差部份為穩定的隨機過程，在時間方向上為一直都有相關程度。

5.2 巢狀共變異數函數

本研究利用粒子群最佳化演算法推估出的巢狀共變異數函數，其巢狀個數為 3，可以下式代表：

$$C_{st}(h, \tau) = C_1 Sph(h; S_1) Sph(\tau; T_1) + C_2 Sph(h; S_2) Exp(\tau; T_2) + C_3 Hos(h; S_3) Hos(\tau; T_3)$$

$$C_1 = 56.66 \quad S_1 = 4159.21 \quad T_1 = 8.46$$

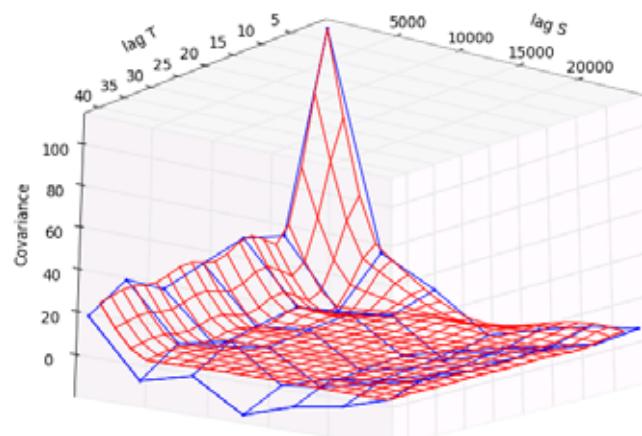
$$C_2 = 29.38 \quad S_2 = 4028.96 \quad T_2 = 503.13$$

$$C_3 = 26.20 \quad S_3 = 13246.12 \quad T_3 = 8.11$$

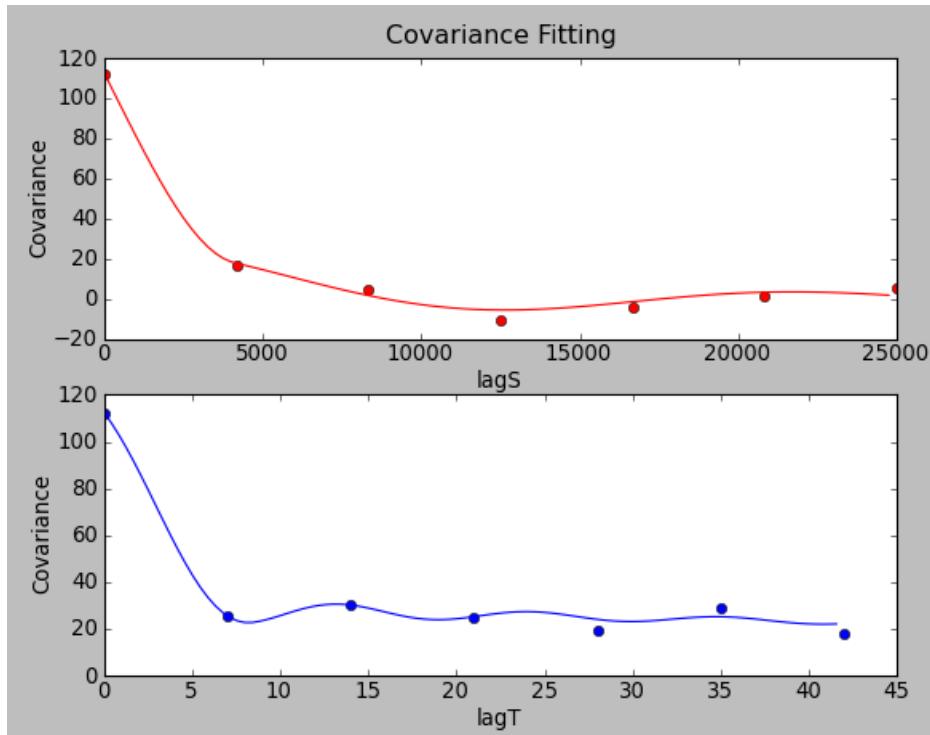
$$AIC = 432.58$$

其中 Sph 為球形、 Exp 為指數型、 Hos 為洞正弦型。

可以繪出下圖的函數圖形：



圖二十一 空氣懸浮粒子共變異數圖形與擬合(3D)

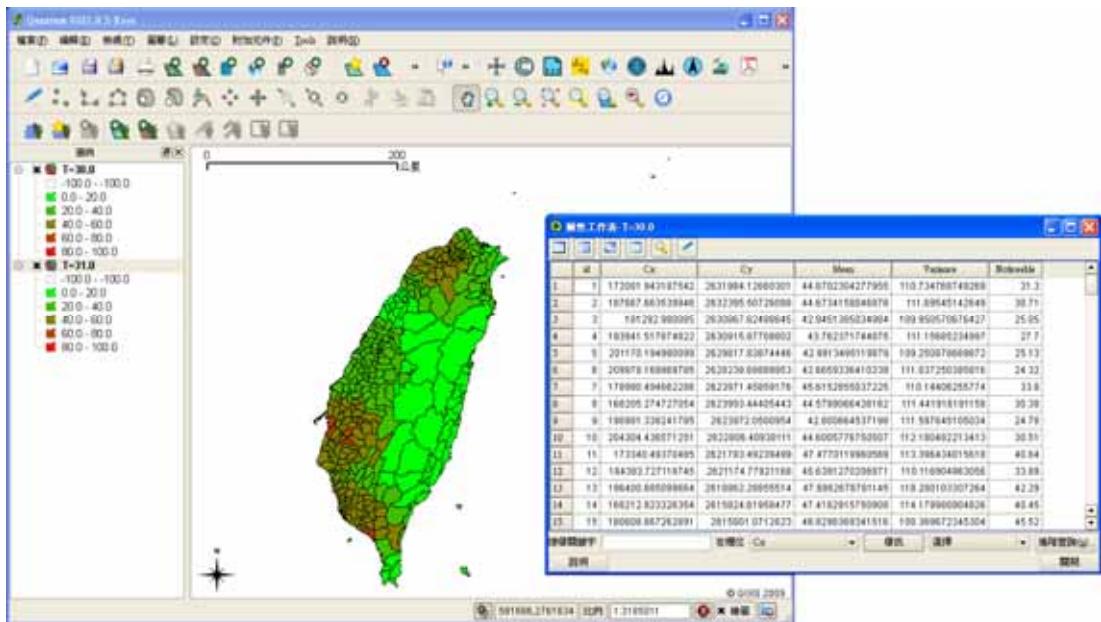


圖二十二 空氣懸浮粒子共變異數圖形與擬合(2D)

特別的是，空間與時間的延遲方面，尾端都有類似週期性的微小振盪，適合用洞正弦型來描述。

5.3 向量式出圖形態

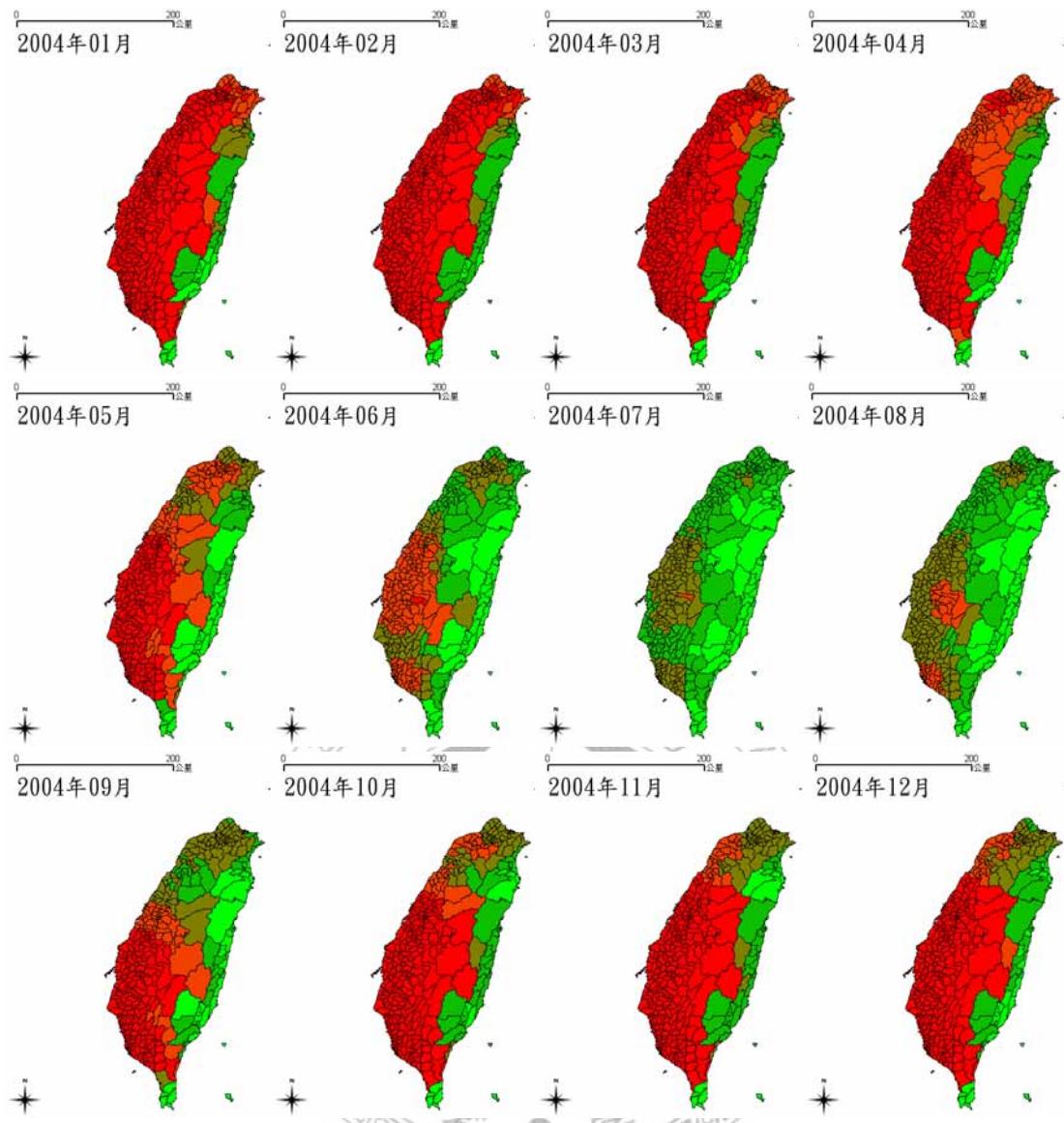
貝氏最大熵法參數使用預設值($nhmax = 15$ 、 $nsmax = 5$ 、 $Spatial Range = 122809.16$ 、 $Temporal Range = 25.38$ 、 $Spatial/Temporal Ratio = 491.72$)，選取台灣的鄉鎮地圖做為輸入的.shp 檔案。空間部份，觀察時間點 30(2006 年 6 月)，資料顯著值為 50，面積過濾參數為 900000，結果如下頁圖：



圖二十三 空氣懸浮粒子於向量式出圖結果

可以發現到，時間點 30 的空氣懸浮粒子濃度超過資料顯著值的機率在北部台北、西部台南、與南部高雄各有較大的值(約 60~70)，表示有較大的機會暴露在較高的濃度下。但參考值僅為 50(屬於環保署空氣污染指標的“普通”，也就是安全值範圍內)，所以並不會危害到健康。

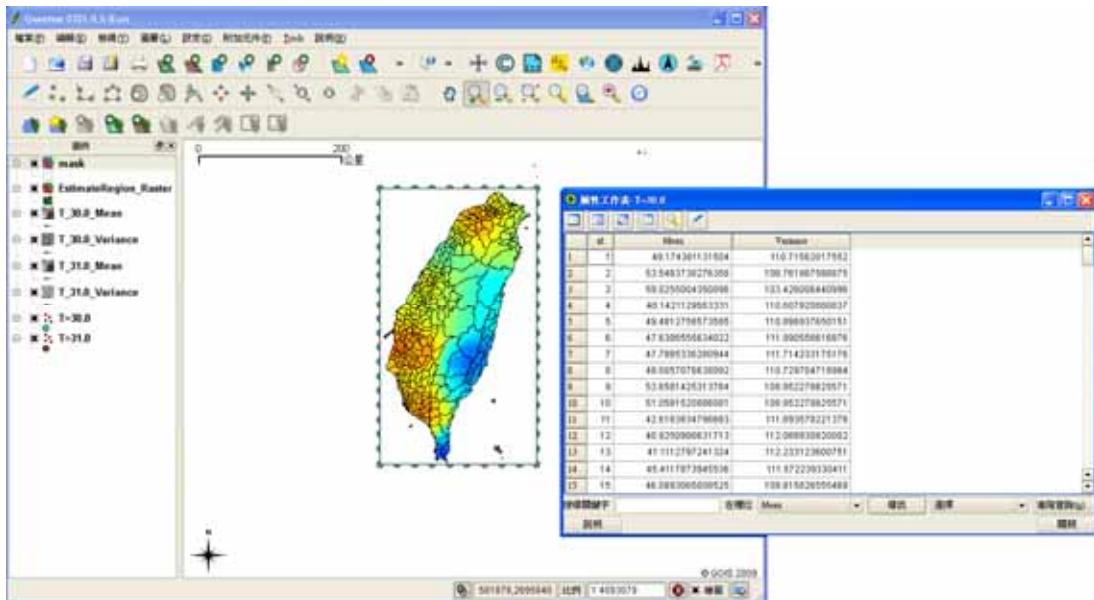
另由時間部份來看，由 PM_{10} 的顯著性隨月份變化的趨勢(如下頁圖，僅 2004 年代表，2005~2008 有相似的結果)可看出 PM_{10} 濃度具有規律，約在 3 月開始下降到 7 月後，再由 8 月上升至隔年 2 月。與一般認知相符[4]。



圖二十四 2004 年 PM_{10} 的顯著性隨月份變化趨勢

5.4 網格式出圖形態

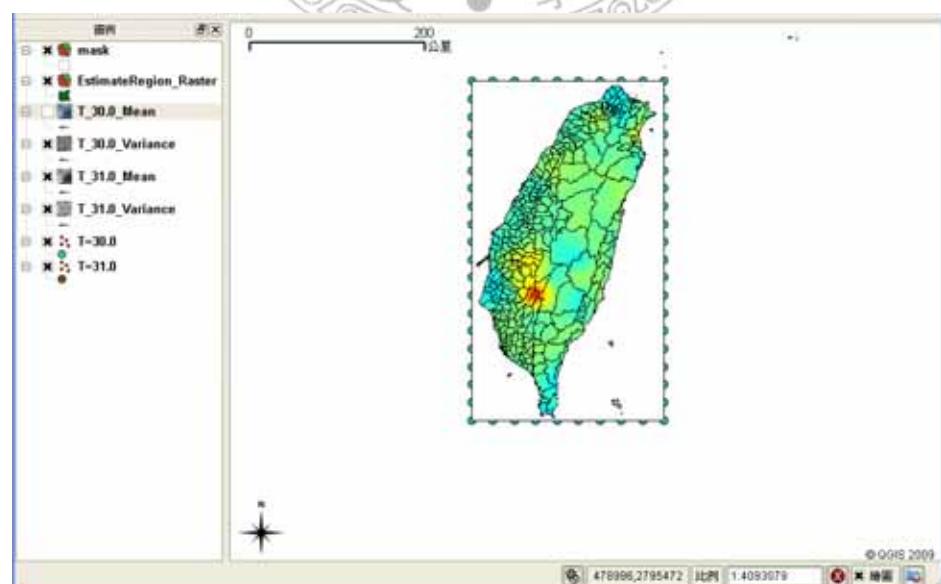
與向量式出圖形相同的貝氏最大熵法參數，選取台灣鄉鎮圖層作為遮罩，用台灣邊界當做欲推點邊界(“Set Taiwan”按鈕)，將 x 方向切成 9 等間隔(10 個點)、y 方向切成 19 等間隔(20 個點)，再配上時間點 30、31，共有 $10*20*2=400$ 個欲推估點。結果如下頁圖：



圖二十五 空氣懸浮粒子於網格式出圖結果(平均值)

濃度推估的平均值看起來與向量式的結果一致，北部、西部、與南部各有較大的值(上圖暖色處)，實際的值可以開啟工作表來得知。

另外變異數的推估結果如下頁圖

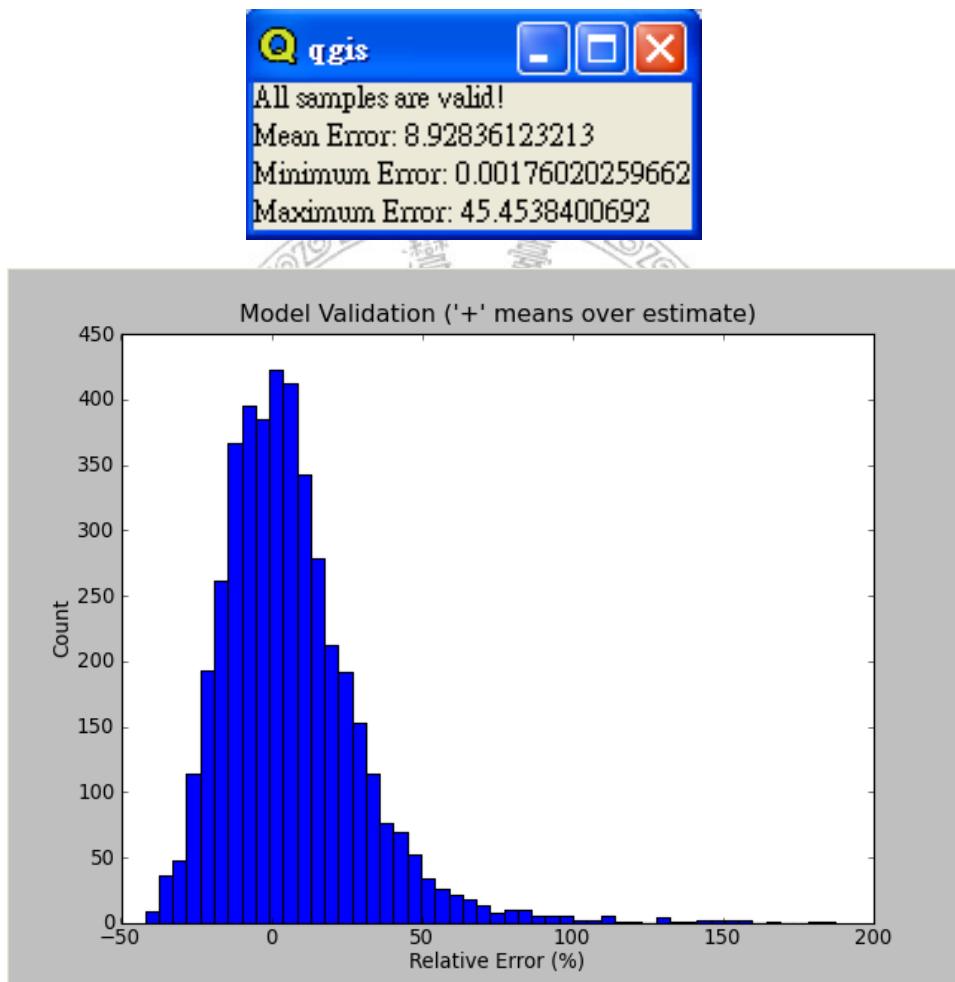


圖二十六 空氣懸浮粒子於網格式出圖結果(變異數)

雖然圖片中有明顯的顏色變化，但查看工作表中的變異數值皆在 100~110 附近，差異並不大。也就是大部份的點標準差約在 10 左右。

5.5 交叉驗證

將所有的點進行交叉驗證，結果如下圖。

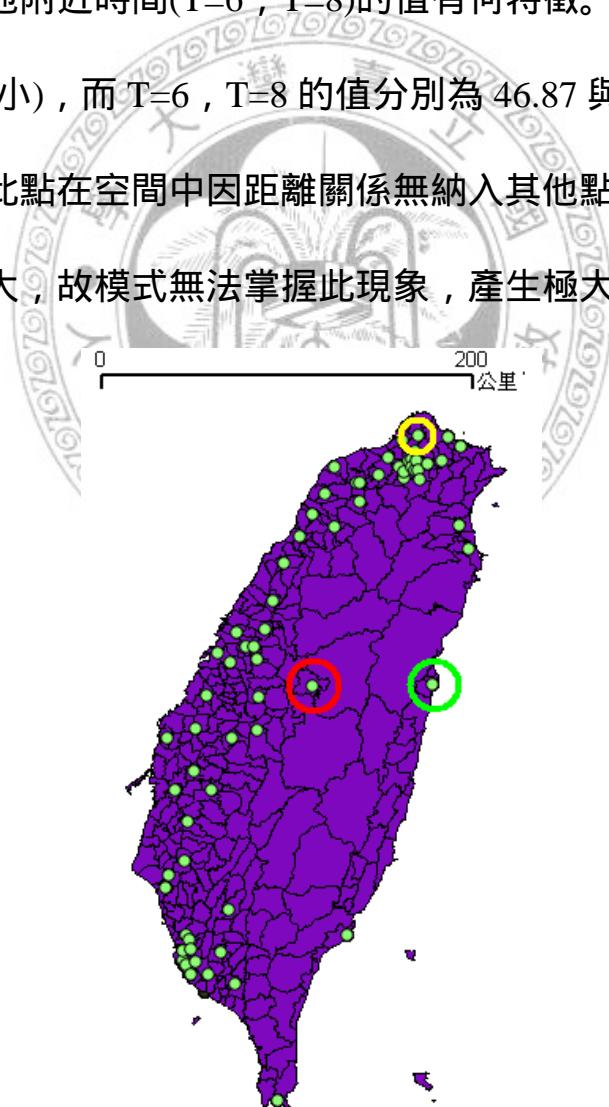


圖二十七 空氣懸浮粒子交叉驗證結果

由小視窗得知，平均誤差為 8.93，誤差最小值為 0.0018、最大值為 45.45。另外經過計算與查詢後發現，平均相對誤差為 17.51%，相

對誤差的標準差為 22.75。有 70%的資料在 $\pm 20\%$ 的誤差範圍內，也有少數 4% 資料相對誤差高達 50% 以上 (其中有一筆為 850%)，且皆為高估的情形。

查看 850% 最高誤差點的座標位置(空間如下圖紅圈處，時間為 $T=7$)，其附近的點相較其他點為少(空間部份)，但其右方的點(圖中綠圈處)應該有更高的誤差，所以此現象應該為誤差來源的一部份。再探討此點與其他附近時間($T=6$ ， $T=8$)的值有何特徵。發現 $T=7$ 時的值為 5.03(非常小)，而 $T=6$ ， $T=8$ 的值分別為 46.87 與 45.80。判斷誤差原因是由於此點在空間中因距離關係無納入其他點資訊，且在時間上其值跳動較大，故模式無法掌握此現象，產生極大的誤差。



圖二十八 相對誤差較大的測站(紅圈與黃圈)

另外如上頁圖黃色圈圈部份，其相對誤差在任何時間點上皆超過平均的相對誤差值，且全部為高估，該點在所有時間上的平均誤差為110.47%。造成誤差的原因，是此測站為台北市陽明測站，因測站位於山上，其測值較附近其他點為低，且其附近的點在空間部份較時間部份給予的權重較大，故該點的推估值總是較實際值為大。



第六章 結論與建議

6.1 結論

本研究成功建構基於貝氏最大熵法的插件軟體(QtBME) , 於 Quantum GIS 上 , 加強了地理資訊系統軟體時空推估上先天的不足 , 以圖形化使用者介面方便使用者使用。並以台灣地區從 2004 至 2008 的空氣懸浮粒子 PM_{10} 的濃度為例 , 進行軟體實作與懸浮粒子長期暴露之研究。

在推估趨勢與殘差部份 , 當殘差的共變異數圖形呈現穩定與同質性 , 即表示成功地把資料分為高頻與低頻二部份。

在推估巢狀共變異數函數方面 , 應用粒子群最佳化演算法與 AIC 準則 , 提供啟發式的求解方法與客觀的判定共變異數函數好壞的依據 , 並減少人為試誤的時間 , 找出最適合的巢狀共變異數函數。

在資料展現方面 , 與 Quantum GIS 軟體結合展現出強力的出圖形態 , 向量式出圖形態適合於推估確切點的數值與其代表面積的顯著性。網格式出圖形態適合於觀察區域中整體的數值表現與推估的變異數分佈。

研究發現以時間點 30(2006 年 6 月) , 空間平面中於台北、台南、高雄等地有較高的濃度暴露值(PSI 普通) , 而台東、屏東偏南等地有較低的濃度暴露值(PSI 良好) , 與一般認知相符。

以時間角度，由濃度值的顯著性可看出，PM₁₀ 濃度具有週期性，約在每年 3 月開始下降到 7 月後，再由 8 月上升至隔年 2 月。

對於推估模式準確度，交叉驗證的相對誤差大部份落在 $\pm 20\%$ 內。偶有較大的相對誤差，皆屬高估情形。相對誤差高達 200% 以上的資料，為該測站的特性與其附近測站較少，所能得到的資訊較少所造成。

6.2 建議

本研究提出以下幾點建議：

1. 程式部份，核心平滑的頻寬選擇、或是巢狀共變異數模式的選擇。可以應用最佳化演算法自動或半自動的求出最佳解，會減少許多人工求解的時間。
2. 程式部份，貝氏最大熵法除了依出圖的形態分為二種，每一種出圖形態下可加以細分為不同需求使用的推估方法或出圖形態。
3. 程式部份，網格式出圖形態僅能看出大略或相對的濃度高低，以後可在出圖部份加入實際值、加上等高線，或可以用滑鼠點選之類的方法加以辨別並幫助說明濃度值的範圍與分佈。
4. 程式部份，核心平滑法中的核心函數可討論非對稱性，例如考慮風速時將屬於不對稱的情形。
5. 實作部份，進一步可探討例如土地利用、人口密度等可與地理資

訊系統結合計算的參數與空氣懸浮粒子相關性的研究。

6. 實作部份，可研究短時間或小地區的空氣懸浮粒子暴露分析。例如區域性失火或焚化爐附近的空氣懸浮粒子濃度分析。
7. 可探討貝氏最大熵法與以往的克利金法二種不同的推估方式所得的結果有何差異。



參考文獻

1. 詹俊南，1996，台灣地區 PM10 污染特性分析，國立臺灣大學環境工程研究所碩士學位論文。
2. 楊忠盛，1998，台北都會區懸浮微粒特性及來源之探討，國立臺灣大學環境工程研究所碩士學位論文。
3. 蔣本基，1992，北桃地區空氣污染受體模式應用之研究(三)，行政院環境保護署。
4. 顏有利，2002，空氣品質長期趨勢分析與年報編撰，行政院環境保護署。
5. Akaike H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19 (6), 716–723.
6. Christakos G. and Serre M.L. (2000). BME analysis of spatiotemporal particulate matter distributions in North Carolina, *Atmosph. Environ.* 34 (20), 3393–3406.
7. Christakos, G., & Olea, R. A. (2005). New space-time perspectives on the propagation characteristics of the Black Death epidemic and its relation to bubonic plague. *Stochastic Environmental Research and Risk Assessment*, 19, 307–314.
8. Cressie, N. (1989). The American Statistician. *Geostatistics*. 43(4), 197-202.

9. Eberhart R. and Kennedy J. (1995). A new optimizer using particle swarm theory. Presented at *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Nagoya, Japan.
10. Eulogio Pardo-Igúzquiza, (1999). VARFIT: a fortran-77 program for fitting variogram models by weighted least squares. *Computers & Geosciences*, 25(3), 251-261.
11. Goodchild, M., (1992) Geographical information science. *Int. J. Geographical information system*, Vol. 6, No. 1, 31-45
12. H-L, Yu, Christakos G. and Chen J-C, (2007). Spatiotemporal air pollution modeling and prediction in epidemiologic research. *In Air Pollution Research Trends*, Columbus, F. (ed.), Nova Science Publishers, Inc., Hauppauge, NY. 57-75.
13. Marcella N. and Cira P. (2003). Kernel smoothing for the analysis of climatic data. *Quaderni di Statistica* Vol. 5.
14. Wand M. P. and Jones M. C. (1995). Bandwidth selection, Monographs on statistics and applied probability 60, *Kernel smoothing*, CRC Press, 63-64.
15. *Geographic information system – Wikipedia, the free encyclopedia*, Retrieved December 24, 2009, from the World Wide Web:
http://en.wikipedia.org/wiki/Geographic_information_system
16. *matplotlib: python plotting – Matplotlib v0.99.1.1 documentation*, Retrieved December 24, 2009, from the World Wide Web:

<http://matplotlib.sourceforge.net/>

17. *Python Programming Language -- Official Website*, Retrieved December 24, 2009, from the World Wide Web:
<http://www.python.org/>
18. *Riverbank / Software / PyQt / What is PyQt?*, Retrieved December 24, 2009, from the World Wide Web:
<http://www.riverbankcomputing.co.uk/software/pyqt/intro>
19. *The PyQt4 tutorial*, Retrieved December 24, 2009, from the World Wide Web: <http://zetcode.com/tutorials/pyqt4/>
20. *QGIS Coding and Compilation Guide*, Retrieved December 24, 2009, from the World Wide Web:
http://download.osgeo.org/qgis/doc/manual/qgis-1.3.0_coding-compilation_guide_en.pdf
21. *Welcome to the Quantum GIS Project*, Retrieved December 24, 2009, from the World Wide Web: <http://qgis.org/>
22. *Writing Python Plugins –Quantum GIS Wiki*, Retrieved December 24, 2009, from the World Wide Web:
http://www.qgis.org/wiki/Writing_Python_Plugins
23. 什麼是 Python? Why Python? , 民 98 年 12 月 24 日 , 取自 :
<http://ez2learn.com/index.php/python-tutorials>
24. *Quantum GIS 資源網@sinica* , 民 98 年 12 月 24 日 , 取自 :
<http://gis.rchss.sinica.edu.tw/qgis/>