

應用基因演算法及基因表示規劃法於區域性 幾何大地起伏值之預測－以台中地區為例

Applying Genetic Algorithm and Gene Expression Programming to Estimate Local Geometric Geoid Model - A Case Study in Taichung Area

高書屏¹ 黃惠琪² 簡子凌³

Szu-Pyng Kao Hui-Chi Huang Zih-Ling Chien

摘要

本研究結合實驗區現有已施測一等水準點正高及水準點上利用 GPS RTK 測得之平面坐標和橢球高，以基因演算法與基因表示規劃法建立實驗區全區，從本研究成果數據分析顯示，利用基因表示規劃法台中全區之區域性幾何大地起伏模型精度可達 ± 5.98 公分，而利用基因演算法台中全區之區域性幾何大地起伏模型精度則可達 ± 1.14 公分，而且推導出之檢核點正高成果均可符合工程測量規範對高程精度之要求。

關鍵詞：基因演算法、基因表示規劃法、區域性幾何大地起伏模型

Abstract

This study combines the existing 1st order leveling data of Taichung and the GPS RTK surveying plane coordinates and ellipsoid heights, adopting the Genetic Algorithm (GA) and Gene Expression Programming (GEP) to fitting local geometric geoid models. According to the experimental results, the obtained accuracy of local geometric geoid models for the Taichung Area is ± 5.98 cm for GEP and ± 1.14 cm for GA method. The derived orthometric heights for check points are all meet the surveying engineering specifications on the accuracy of the height.

Keywords: Genetic Algorithm (GA), Gene Expression Programming (GEP), Local geometric geoid model

¹ 國立中興大學土木工程學系測量資訊組副教授

² 國立中興大學土木工程學系測量資訊組博士候選人

³ 國立中興大學土木工程學系測量資訊組碩士

一、前言

近二、三十年來，隨著 GPS 全球定位系統趨於完善，在世界各地的應用也越來越廣泛，其優點有高精度定位、觀測時間短、可全天候作業、測站間無需通視以及自動化操作簡便等，故以往平面控制測量所使用的傳統測量已逐漸被高效率的 GPS 測量所取代，但在高程控制測量方面，由於 GPS 測量精度不足的關係，一般主要仍是以水準測量來進行施測取得高精度之高程坐標，但是有耗費大量人力、作業時間及成本的缺點，於是如何運用高效率的 GPS 測量取得合乎工程所需之精度，是如今值得探討的一個問題。回顧以往文獻，建立大地起伏模型主要分為兩種方法，第一種是考慮到地球重力等物理因素的重力場模型法，但因台灣地形特殊，其崎嶇的地形及分佈密度問題使得實測重力資料取得不易且需耗費大量成本，導致工程效率不佳，第二種為數學函數擬合法，亦即本研究所使用之方法(鍾智偉，2008、簡子凌，2011、陳佳菱，2011)，此種方法不考慮上述等物理因素，僅使用實測資料做為數據，再透過各種數學函數來擬合建立大地起伏模型，如果精度能夠符合工程規範所需之精度，就能夠有效的運用在實務上。

由於基因演算法具有「結果易於應用」、「處理的資料類型範圍廣」、「適用於最佳化問題」等優點，如：應用於大地起伏模型建立應可有不錯的成果，本研究中將以 GPS 測量數據結合水準測量數據，採用基因演算法與基因表示規劃法建立區域性之大地起伏模型，往後僅需將 GPS 測量的點位平面坐標輸入到大地起伏模型，即可求得該點位的大地起伏值，再經由 GPS 測量所得之橢球高求得工程實務所需之正高，進而提升工程效率以及減少水準測量之作業時間，並期望與以往相關文獻相比，能夠將精度更進一步的提升，或是以較少的點位資料便能夠達到相同水準的精度。

二、理論基礎

2.1 基因演算法

在 60 年代，基因演算法是由密西根大學的約翰·霍蘭德和他的同事對細胞自動機 (cellular automata) 進行研究時所率先提出。在 80 年代中期之前，科學界對於基因演算法的研究僅止於理論方面，隨著計算機計算能力的發展和實際應用需求的增多，基因演算法逐漸進入實際應用階段，大多數都利用基因演算法進行時間表安排、數據分析、未來趨勢預測、預算及解決很多其他組合優化問題。

基因演算法為計算數學中用於解決最優化的搜索演算法，是進化演算法的一種，進化演算法最初是借鑒進化生物學中的一些現象而發展，這些現象包括遺傳、突變、自然選擇以及雜交等，通常是以計算機模擬方式實行。

對於一個最優化問題，一定數量的候選解（稱為個體）的抽象表示（稱為染色體）的種群向更好的解進化。傳統上，是用二進制表示（即 0 和 1 的字串），但也可以用其他表示方法。進化從完全隨機個體的種群開始，之後一代一代發生。在每一代中，整個種群的適應度被評價，從當前種群中隨機地選擇多個個體（基於它們的適應度），通過自然選擇和突變產生新的生命種群，該種群在演算法的下一代迭代中成為當前種群（基

因演算法架構如圖 1)。

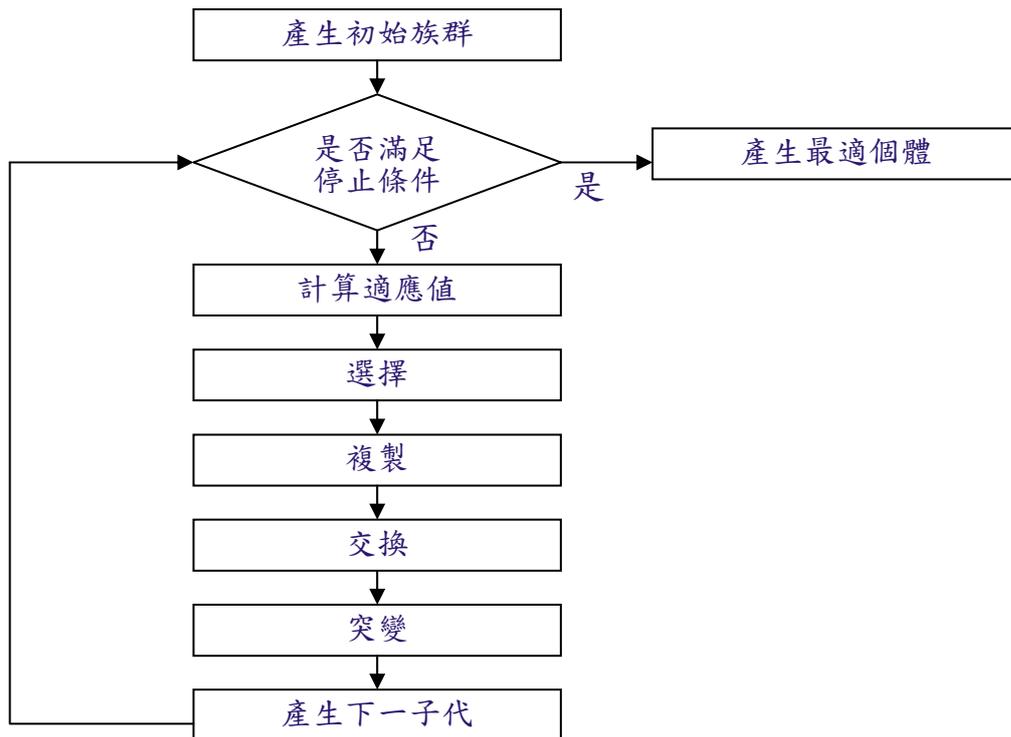


圖 1. 基因演算法架構圖

2.1.1 基因演算法方法流程

由圖 1. 基因演算法架構圖可知其流程，將該流程簡述說明如下：

Step1：隨機產生 N 個染色體。

Step2：利用適應函數算出每個染色體的適應值。

Step3：選擇與複製：選擇適應值高之染色體複製成多個染色體。

Step4：利用交換及突變的動作產生新的染色體。

Step5：測試新染色體的適應值是否達到門檻值。

Y：則停止

N：則 Go to Step3

持續迭代至直到終止條件滿足為止，一般終止條件有以下幾種：

- (1) 進化次數限制。
- (2) 計算耗費的資源限制（例如計算時間、計算佔用的內存等）。
- (3) 一個個體已經滿足最優值的條件，即最優值已經找到。
- (4) 適應度已經達到飽和，繼續進化不會產生適應度更好的個體。
- (5) 人為干預。
- (6) 以上兩種或更多種的組合。

2.1.2 基因演算法參數設定

(1) 種群規模(P,population size)：即種群中染色體個體的數目。

(2) 字串長度(l,string length):字串編碼的長度越長則精準度越高，但所需的編碼、解碼運

算也相對增加。

- (3) 交叉機率(pc, probability of performing crossover)：控制著交叉算子的使用頻率。交叉操作可以加快收斂，使解達到最有希望的最優解區域，因此一般取較大的交叉機率，但交叉機率太高也可能導致提早收斂。
- (4) 變異機率(pm, probability of mutation)：控制著變異因子的使用頻率。
- (5) 中止條件(termination criteria):適應度已經達到飽和，繼續進化不會產生適應度更好的個體。

2.2 基因表示規劃法

本研究第二種所使用之方法為基因表示規劃法，是以基因演算法(Genetic Algorithm, GA)及基因規劃法(Genetic Programming, GP)為基礎概念所提出，以下將個別簡述其概念。基因演算法是在1960年由John Von提出的理論為概念，直到1975年才由John Holland正式定義其數學基礎與名稱，其流程是將問題編碼成染色體之型式，根據所定義之適應值函數作為篩選標準，計算出族群中不同染色體的適應值，其中適應值好的留下、差的淘汰，並透過基因操作產生新一代族群，以此方式不斷改進當前解，直到找出最適解為止(Holland, 1975)。基因規劃法則是於1989年由Koza所提出，並在1992年發表為著作，與基因演算法的最大相異之處在於基因演算法演化單位是以二元字串的編碼方式呈現，而基因規劃法則是表示成不定大小及形狀的樹狀結構，能直接將演算結果以結合理論基礎與經驗公式的程式化方式呈現(Koza, 1992)。

本研究所採用的基因表示規劃法是由Ferreira在2001年所提出，與上述兩者之主要差異在於基因表示規劃法可將變異過程與評估過程分開，先編碼成固定長度的染色體，直到評估適應值時再表示成不同大小、形狀的樹狀結構(Ferreira, 2001)，既具有基因演算法的簡單性，又具有基因歸劃法的功能，在解決非線性擬合問題方面有優異的數據處理能力及搜索能力(鐘義嘯，2006)。

2.2.1 基因表示規劃法方法流程

基因表示規劃法與以往的基因演算法及基因規劃法一樣，一開始先隨機產生一組初始族群，其中族群中染色體的數量依不同問題而自行設定，再將染色體表示成樹狀結構，透過所設定之適應值函數計算各個染色體的適應值，在此過程中，適應值越好的代表越符合我們對問題解答的要求，適應值較差則否，根據此標準將優良的染色體保留，並進行突變、轉換及重組等基因操作產生下一代的新族群(何亞威，2010)以求得更佳解，如此重複上述過程不停迭代，直到達到所設定之終止條件即停止運算，而經過實際測試後，在本研究中最優適應值通常於迭代次數第 2000 代之前就會出現，僅少數會在第 2500 代前後出現，故依測試經驗最後將終止條件設定為迭代次數第 3000 代，使其 2500 與 3000 之間有一定之容許誤差量，避免例外情況發生，最後求得問題之最適解。而有關基因表示規劃法中基因的長度、如何表示成樹狀結構等原理請參照以往文獻(向勇等，2010)，在此便不再贅述。

2.2.2 基因表示規劃法參數設定

本研究所使用之基因表示規劃法程式是利用網路上所提供之資源，再自行修改程式碼使其可應用於擬合大地起伏模型，而由於基因表示規劃法具有許多參數，故特於此小節說明本研究所設定之各參數值，分別如表1. 所列：

表1. 參數設定表

參數名稱	參數值
基因頭部長度	10, 11, 12
染色體之基因個數	2
族群大小	2000, 2500, 3000
常數個數	1
最大迭代次數	3000
函數集	'+', '-', 'x', '÷', 'Q', 'S', 'E', 'O'
突變率	0.05
IS轉換率	0.1
RIS轉換率	0.1
單點重組率	0.3
雙點重組率	0.3
基因重組率	0.1
常數突變率	0.1
常數互換率	0.1
常數範圍	0~10
連結基因之函數	'+'

函數集中 Q 為平方根、 S 為正弦函數(sin)、 E 為指數函數(exp)、 O 則代表自然對數(log)，而在參數修改部分，上述參數中有關基因操作之部分皆延用程式內部之設定值，包含突變率、IS轉換率、RIS轉換率、單點重組率、雙點重組率、基因重組率、常數突變率及常數互換率等，其餘參數之設定均為參考以往相關文獻(岳崇倫，2011)，並透過實際測試將參數值進行調整，其中基因頭部長度在10、11、12及族群大小在2000、2500、3000時均能於測試中得到良好之成果，故在本研究中是將各種組合實際測試後，再將其中精度最佳者作為最後成果。

2.2.3 基因表示規劃法適應值函數

評估演算結果之優劣取決於適應值，其值是透過適應值函數所求得，由此可知，適應值函數是整個演化過程的方向指標，對演算的結果及收斂速度有很大的影響，故根據不同問題之特性而選擇合適的適應值函數是基因表示規劃法中相當關鍵且重要的部份，回顧以往相關文獻中(王艷春，2010)，常用的適應值函數包括以絕對誤差作為演化方向、相對誤差作為演化方向、利用統計學上常用的相關係數修改等，而在本研究中，因所使用之精度評估指標為均方根誤差，故採用以均方根誤差作為演化方向的式(1)做為

- (1) 演化代數gen_no=1，由程式讀入點位坐標，擬合出大地起伏之二次多項式函數，每個函數各有六個係數，其中 a_0 屬性須宣告為'normalized'(正規化)。
- (2) 將所有大地起伏之二次多項式函數演算適應函數RMSE。
- (3) 經由適應函數的比較，決定下個迭代GA的較佳值。
- (4) 更新方程式係數。
- (5) $gen_no = gen_no + 1$ 。
- (6) 產生新的大地起伏之二次多項式函數及適應函數值RMSE。
- (7) 重複(3)~(6)步驟。
- (8) 當演算代數gen_no=500(程式所設定的代數上限)，得到最佳解RMSE及大地起伏模型。
- (9) 程式中止。

3.3 基因表示規劃法研究步驟

首先取得台中市實驗區資料，數據包括GPS所測得之縱坐標、橫坐標和橢球高，以及水準測量所測得之正高數據，再將實驗區數據分為用於擬合大地起伏模型之擬合點，與用於檢核模型優劣之檢核點兩部份，將擬合點數據中的縱坐標、橫坐標等平面坐標作為輸入，在忽略垂線偏差之影響下，橢球高減去正高所求得之大地起伏值作為輸出，透過C++環境下的GEP程式求得兩者之間的轉換函數公式，即大地起伏模型公式，最後以MATLAB作為運算平台，將檢核點輸入至擬合點所擬合出的大地起伏模型公式中，再將求得之大地起伏估計值與原始數據中之實測值之間進行比較成果分析。

本研究之步驟，簡述如下：

- (1) 將實驗區數據分為擬合點及檢核點兩部份。
- (2) 在 GEP 程式中輸入擬合點(訓練樣本)數據，包括輸入值 a(縱坐標)、b(橫坐標)，輸出值 N(大地起伏值)。
- (3) 運行 GEP 程式求得輸入與輸出之間的轉換公式，即大地起伏模型之函數公式。
- (4) 在 MATLAB 環境下，將檢核點(測試樣本)之 a(縱坐標)、b(橫坐標)數據代入所求得之大地起伏模型公式，求得檢核點的 N' (大地起伏估計值)。
- (5) 計算檢核點的 N' (大地起伏估計值)、N(大地起伏值)之間的差值，即兩者之殘差，並對其進行成果分析。

四、研究成果

為判別擬合成果精度之優劣，本研究以目前所蒐集之參考文獻中，區域擬合精度最佳的最小二乘支持向量機(沈昱廷，2011)內具較高精度的多項式核函數與徑向基(RBF)核函數作為比較對象，使用相同實驗區資料且挑選相同之擬合點，未參與模型訓練之點位為檢核點，計算其均方根誤差並進行分析比較，最後以成果中的大地起伏殘差值作為依據計算均方根誤差，再對其進行比較分析並探討成果數據所代表之含義。

4.1 全區分析

先從最小二乘支持向量機與基因表示規劃法兩者比較可知，此兩種方法均有因隨擬

合點數增加，RMSE 值隨之變小的趨勢，其中擬合點 30 點為最小二乘支持向量機文獻中，全區最佳精度之點位數量，故為使比較更具完整性，在此一併納入探討，而由成果表中可看出在同樣擬合點數 10 點及 30 點時，基因表示規劃法之精度均較差於最小二乘支持向量機，且最小二乘支持向量機在陸續增加擬合點數時精度還能夠持續提升，直至點數增加至 30 點為止，多項式核函數精度可達到 ± 1.469 公分，徑向基核函數則是 ± 1.453 公分，分別提昇了約 4 公分以及 2 公分的精度，而基因表示規劃法則僅能提升約 1 公分之精度。

而從基因演算法與基因表示規劃法兩者比較則可知，在擬合點數 10 點、檢核點數 68 時，基因演算法之均方根誤差精度為 ± 1.483 公分，則基因表示規劃法之均方根誤差精度為 ± 6.662 公分；而在擬合點數 30 點、檢核點數 48 時，基因演算法之均方根誤差精度為 ± 3.339 公分，則基因表示規劃法之均方根誤差精度為 ± 5.976 公分，由此可知，故由表 2. 可得採用基因演算法是優於基因表示規劃法。

而前述最佳之方法基因演算法與目前所蒐集之參考文獻中，區域擬合精度最佳的最小二乘支持向量機(沈昱廷, 2011)內具較高精度的多項式核函數與徑向基(RBF)核函數作為比較對象，可以得知最小二乘支持向量機之徑向基核函數是與基因演算法均方根誤差值最為接近，惟最小二乘支持向量機之徑向基核函數之均方根誤差是隨擬合點增加而提高；而基因演算法之均方根誤差是隨擬合點增加而降低，由於最小二乘支持向量機之徑向基核函數所得之均方根誤差是已經採取最佳擬合點數後，所得之結果，因此，為分析基因演算法用於求取區域性大地起伏模型應採取之多少擬合點數為最佳，則於下節進行後續分析。

表 2. 台中全區檢核點 RMSE 成果表(單位：cm)

擬合方法	最小二乘支持向量機(沈昱廷, 2011)				GEP		GA	
	多項式核函數		徑向基核函數					
擬合點數	10	30	10	30	10	30	10	30
檢核點數	68	48	68	48	68	48	68	48
RMSE 值	± 5.448	± 1.469	± 3.445	± 1.453	± 6.662	± 5.976	± 1.483	± 3.339

4.2 最佳化測試

測試區域性大地起伏模型之最佳擬合點數，首先選取研究區外圍六個點位作為已知擬合點，分別是 AB655、GBC055、HT136、2035、2128、3091，再每次加入兩點中間點位進行計算，同時點位盡量均勻分布，選取的點位稱之為參考點，其餘點位當成檢核點，利用基因演算法計算不同點數擬合之均方根誤差及其執行時間，並以檢核點之均方根誤差作為精度評鑑標準，以找出最佳之擬合點數，由於這些點未參加擬合，所以這種檢核具有外部檢核的性質，其結果也較具說服力，具體成果如表 3. 所列。

表 3. 基因演算法擬合點數成果表(cm)

擬合點數	基因演算法 RMSE	執行時間
6	±0.148	3.8s
8	±1.228	3.8s
10	±1.483	3.9s
12	±1.696	4.0s
14	±1.782	4.1s
16	±2.119	4.0s
18	±2.300	3.9s
20	±2.666	4.1s
22	±2.828	4.1s
24	±3.156	1.7s
26	±3.030	1.5s
28	±3.020	1.7s
30	±3.339	1.6s
32	±3.049	1.9s
34	±3.144	1.6s
36	±2.964	1.9s
38	±3.009	1.9s
40	±2.952	1.8s
42	±2.882	2.2s
44	±2.775	2.0s
46	±2.713	2.0s
48	±2.837	2.5s
50	±2.747	2.3s
52	±2.692	2.0s
54	±2.734	2.3s
56	±2.548	2.2s
58	±2.550	2.2s
60	±2.619	2.3s

由表 3 之結果顯示，擬合點數在 6 至 30 點之間，均方根誤差隨著點數增加而變大，在 30 點時誤差最大，亦即精度最低；而在 30 點至 52 點間，均方根誤差有隨著點數增加而逐漸減小的情況，在 52 點時更達到誤差最小，52 至 60 點間誤差則有不規則變化情況，故以 52 點進行實驗區全區的擬合測試。本研究實驗區擬合點與檢核點之分佈圖如圖 3。

參考文獻

1. 王艷春 (2010) 基因表達編程算法及其應用綜述，計算機應用與軟件，第 27 卷，第 6 期。
2. 向勇、唐常杰、朱明放、陳瑜、代述成 (2010) 內嵌基因表達式編程及其在函數發現中的應用，電子科技大學學報，40(1)。
3. 何亞威 (2010) 基因表達程式規劃於全球定位系統中幾何精度稀釋因子之逼近，國立高雄大學電機工程學系碩士論文。
4. 沈昱廷 (2011) 以最小二乘支持向量機擬合區域性大地起伏值之研究—以台中地區為例，國立中興大學土木工程學系碩士論文。
5. 岳崇倫、曾苑 (2011) 基因表達式編程方法在 GPS 高程擬合中的應用，地礦測繪，1。
6. 陳佳菱 (2011) 以粒子群演算法改善傳統二次曲面擬合區域性大地起伏精度之研究，國立中興大學土木工程學系碩士論文。
7. 鍾義嘯、唐常杰、陳宇、段磊、魏大剛 (2006) 提高基因表達式編程發現知識效率的回溯策略，四川大學學報自然科學版，1。
8. 鍾智偉 (2008) 不同模式多面函數法改進推求區域性大地起伏值方法之研究—以台中市為例，國立中興大學土木工程學系碩士論文。
9. 簡子凌 (2011) 以基因表示規劃法建立區域性大地起伏模型之研究—以台中地區為例，國立中興大學土木工程學系碩士論文。
10. Ferreira, C. (2001) Gene expression programming : A new adaptive algorithm for solving problems. *Complex Systems*, 13(2): 87-129.
11. Holland, J.H. (1975) *Adaption in natural and artificial systems*. Ann Arbor, MI: The University of Michigan Press.
12. Koza, J.R. (1992) *Genetic programming - on the programming of computers by means of natural selection*. Cambridge, MA. MIT Press.