

## 漫談蒙地卡羅法在農業試驗研究上的應用

鄭文吉<sup>1</sup>

### 摘要

蒙地卡羅法配合電腦模擬技術可以創造理想環境及材料，有效控制變因並減低試驗成本，因此近年來已成為許多領域科學研究的重要工具。然而，在進行模擬試驗前，必須先了解問題本身的特性，以便針對其獨特的機率分布模式產生相對應的亂數和程式，而不能隨意套用其他模擬試驗的模式和程式來進行模擬運算，以免產生錯誤的結果。

本文以兩個關於病蟲害傳播空間分布方面試驗研究的實際案例，說明如何應用蒙地卡羅電腦模擬技術來進行試驗研究，以及操作時應該注意的問題，希望能提供研究人員參考。

關鍵語：蒙地卡羅法、電腦模擬、偽逢機數字

### 前言

前文提過，蒙地卡羅法是一種利用電腦產生大量的亂數(random number, 又稱逢機數字)來進行試驗研究的研究方法，除了可以當作數值分析的計算輔助工具，亦可用以解決各種最佳化問題，因此目前在物理、化學、數學甚至社會科學等領域也都被廣泛應用。隨著電腦科技的進步，現在研究人員可以利用電腦模擬試驗研究過程，除可建立理想的試驗環境外，並可透過亂數模擬試驗產生的結果或外在影響所造成的誤差，如此除了可以大幅減少進行真實試驗所需的時間和成本，也可以避免因試驗材料不夠均勻、或取樣動作不夠逢機等操作上所產生的誤差而影響試驗結果；並且可以任意改變各變因的參數值再重新進行模擬試驗，來觀察在不同參數組合下所得結果的差異情形，而不用實際進行這麼多次試驗。由於電腦模擬技術可以有效控制變因及減低試驗成本，以便在電腦創造出來的理想環境下比較不同理論模式或操作方法的差異，因此近年來已成為許多領域科學研究的重要工具<sup>(4,5)</sup>。

本文將以兩個關於病蟲害傳播空間分布方面試驗研究的實際案例，說明如何應用蒙地卡羅電腦模擬技術來進行試驗研究，以及操作時應該注意的問題，希望能提供研究人員參考。

---

<sup>1</sup> 行政院農業委員會高雄區農業改良場助理研究員

## 亂數的產生

蒙地卡羅電腦模擬試驗是藉由仿照真實試驗的進行流程所設計的電腦程式，加上不斷的擷取亂數來使各變因產生各種變化，使得程式產生不同的結果，以探討各種變因組合下的變化趨勢。因此進行模擬試驗需要大量的亂數，如何正確而迅速的取得亂數，就成為模擬試驗成敗的關鍵。

早期的亂數大多藉由銅板、骰子、抽籤、撲克牌、輪盤等實體(physical)方法產生(如圖 1)，不僅速度慢，而且每次都必須重新取樣。後來也有人製作亂數表(random table)提供研究人員使用，雖然比較方便，但速度仍然很慢；而且如果模擬次數很多，亂數表便不敷使用，因此便開始利用電腦來產生大量隨機數字。



圖 1. 各種不同面體的骰子，可產生不同數量的均勻分布亂數

Fig 1. Use various icosahedral dice to produce various number of uniformly distributed random number.

目前電腦模擬試驗大多使用偽隨機數字(PseudoRandom Numbers, PRNs)，也就是利用數學公式的運算來產生亂數，先給定一個起始值，經過公式運算後可得到下一個值，如此反覆運算即可求得一系列的數字。由於電腦運算速度很快，因此可以在短時間內得到大量亂數，而且範圍和分布狀況也可以配合試驗需要加以設定。

在一般試驗研究的應用上，所需要的大多以均勻分布(Uniform)的亂數，來確保取樣的公平性。因此，一般套裝軟體和程式語言通常都會提供均勻分布  $U(0,1)$  的亂數，然後再配合需要自行調整。以常用的 SAS 統計軟體，其內建的亂數就是用下列公式計算出來的：

$$x_i = 397204094 x_{i-1} \bmod (2^{31} - 1), \quad 0 < x_i < 2^{31} - 1$$

$$u_i = x_i / (2^{31} - 1), \quad 0 < u_i < 1$$

這種運算方式稱為「乘法型同餘公式」，也就是將數值乘以某個固定數字後，再除以另一個數字求取其餘數來當做下一個數值。其中 mod 為餘數運算子，例如  $10 \bmod 3$  即是求 10 除以 3 的餘數，因此得到結果為 1。而上述公式可以產生  $2^{31}$  ~ 約 20 億種隨機數字  $x_i$ ，已足以應付一般需求，因此一般程式語言和許多應用軟體大多是採用這個公式作為內建的隨機數字產生來源。

然而，如果模擬試驗規模龐大，需要更大量的隨機數字，上述公式可能

無法應付。因此 Wichmann and Hill(1982)提出另一個組合型的亂數產生公式：

$$\left. \begin{array}{l} x_i = 171 x_{i-1} \text{ mod } 30269 \\ y_i = 172 y_{i-1} \text{ mod } 30307 \\ z_i = 170 z_{i-1} \text{ mod } 30323 \end{array} \right\} u_i = \left( \frac{x_i}{30269} + \frac{y_i}{30307} + \frac{z_i}{30323} \right) \text{ mod } 1$$

這個公式先使用 3 個乘法型同餘公式分別產生 3 個整數遞迴數字  $x_i$ 、 $y_i$ 、 $z_i$ ，然後進行合併，再除以 1 得到剩餘的小數部分，最後得到服從均勻分布  $U(0,1)$  的亂數  $u_i$ 。由於個別公式所用的除數並不大(相對於前一個公式所用的除數  $2^{31}-1$  而言)，因此這個方法雖然分為多個步驟，但計算時速度並不會減緩太多，而循環週期卻可擴大到  $6.95 \times 10^{12}$ 。以每秒需要使用 10,000 個亂數的模擬試驗程式來說，要循環一週期大約需要 22 年，而前面所提到的慣用公式則只要 2.48 天就會發生循環現象。因此，對於需要非常大量亂數才能進行的模擬研究，建議可以採用這個亂數產生公式來自行產生所需的亂數<sup>(9,10)</sup>。

有時候模擬試驗需要的不一定是均勻分布的隨機數字，而是符合某種機率分布的隨機數字，以便更加貼近試驗所需的狀況。因此一般常用軟體除了均勻分布亂數外，通常也提供其他類型分布亂數。例如 EXCEL 軟體內建的「亂數產生器」，除均勻分布外，還可以產生常態分布、伯努利分布、二項分布、卜瓦松分布等分布類型的亂數，其他各種常用套裝軟體所提供的亂數種類整理如表 1<sup>(1)</sup>。

表 1. 各種常用套裝軟體所提供的亂數種類一覽表

Table 1. Random number provided by the various software packages.

(余, 2007)

隨機變數的分配函數	SAS 6.12	S-PLUS 2000	MINITAB 12	EXCEL 97	SPSS 8.0
均勻(uniform)分配	*	*	*	*	*
常態(normal)分配	*	*	*	*	*
指數常態(lognormal)分配		*	*		*
指數(exponential)分配	*	*	*		*
貝他(beta)分配		*	*		*
伽嗎(gamma)分配	*	*	*		*
科西(cauchy)分配	*	*	*		*
羅吉斯(logistic)分配		*	*		*
韋伯(weibull)分配		*	*		*
F 分配		*	*		*
T 分配		*	*		*
卡方(chi-square)分配		*	*		*
Laplace 分配			*		*
Pareto 分配					*
伯努利(bemoulli)分配			*	*	*
二項(binomial)分配	*	*	*	*	*
幾何(geometric)分配		*	*		*
負二項(negbinomial)分配		*			*
卜瓦松(poisson)分配	*	*	*	*	*
超幾何(hypergeometric)分配		*	*		*

如果軟體或程式語言本身沒有提供我們所需要的分布類型的亂數，就需要先產生均勻分布亂數後，再進行轉換。目前常見分布類型的亂數大多已經有學者研究出轉換方法，其轉換公式可自行查閱相關書籍，在此不多贅述。

不論是使用套裝軟體或自己設計的亂數產生公式，由於這些公式都是從上一個數字經過運算產生下一個數字，因此都必須先設定初始值  $x_0$ ，這個數值將決定亂數會從數列的哪個位置開始出現，又稱為亂數的種子(seed)。若  $x_0$  相同，也就是提供相同的種子，那麼不管程式重新執行幾次，都會得到完全相同的亂數數列。因此在使用套裝軟體或程式語言內建的亂數產生指令時，通常必須另外給予會隨時變動的亂數種子，才能確保每次執行都會重新取樣，而不會得到完全相同的結果。

### 蒙地卡羅法模擬試驗的設計

電腦模擬試驗目的在模擬真實試驗的進行，因此為了能呈現真實試驗的進行狀況，研究人員必須先了解試驗材料的分布情形，及各種處理變因的影響效果模式。這些背景知識可經由過去相關文獻探討，了解類似研究的機率分布情形，然後給予合理的假設作為理論基礎，如此才能藉由電腦程式的執行流程，來正確的重現出實際試驗可能發生的情形<sup>(4,5)</sup>。

由於模擬試驗的材料和處理效果實際上只是一堆數值變化，並非真的有些那些材料和處理存在。因此要進行模擬試驗，就必須設法將試驗材料和處理效果轉變成數學模式，以便讓電腦程式掌控它們的變化，達到我們想要探討的目標。

以下整理出幾點進行模擬試驗需要考慮的變因：

#### 一、試驗材料：

由於模擬試驗並沒有真正的試驗材料，因此是先依照試驗規模的需要設定資料陣列的大小，作為取樣空間。例如有  $m$  種試驗處理，進行  $n$  重複，則需設定  $m \times n$  大小的陣列；若為多重參試因子的試驗，則只要相對設定多維資料陣列即可。陣列設定完成後，再將裡面所含資料的初始值依照試驗需要加以設定，若假定材料是均質狀態，則將起始值設定為完全相同(通常設為 0 或某個固定值)；若試驗前提假定材料並非均值，亦可依需要設定為若干種初始值，或是服從某種機率分布變化的亂數值。

#### 二、試驗處理：

「試驗材料」準備完成後，即可開始進行「處理」，也就是將初始值依照處理作用的效果更改為新數值，作為處理後的結果。若為多重參試因子的試驗，則依照試驗的設定分別加入各種處理的效果。至於處理效果大小可視需

要設為固定，或是服從某種機率分布變化而產生些微的差異。

一般來說，進行模擬試驗的目的就是要探討在各種參試因子變化組合下的結果差異情形，因此在處理部份都會依照各參試因子可能的變化範圍加以組合，然後逐一進行模擬試驗。例如有三個參試因子，每個都有 10 種變化，總共就有  $10 \times 10 \times 10 = 1000$  種參試處理組合，這些處理組合的試驗結果都必須加以彙整，以探討其結果變化的趨勢。

### 三、調查取樣：

當「試驗處理」結束後，原本的試驗材料(初始值)已因各種處理作用的效果而產生變化，這時可依照試驗的規劃進行調查取樣。先依照調查方式另行產生亂數，代表「選中」資料陣列裡的哪些位置，其所含的資料即為被選中的樣本觀測值，可用以進行後續的統計分析，或只單純就其變化情形加以探討。

### 四、重複試驗：

由於模擬試驗的材料、處理效果及取樣調查方式都是依照不同的機率分布產生的亂數，若只進行一次試驗並無法代表真正的結果，因此模擬試驗通常會在同樣的參數條件下重複進行很多次，來探討試驗結果的變化情形，或可能產生某種結果的機率。如此一來，真正執行的模擬試驗次數可能非常多，例如前面例子所提的三因子試驗，假設每個都有 10 種變化，就有 1000 種參試處理組合，若每種組合都重複執行 100 次，就需要進行 10 萬次的模擬試驗運算才能得到結果。因此，電腦模擬試驗通常都需要進行長時間的大量運算，如何提昇運算速度以縮短時間，也成為研究人員必須面對的問題。

## 電腦模擬試驗範例

以下利用兩個關於病蟲害傳播空間分布方面試驗研究的實際案例，說明如何應用電腦模擬技術來進行試驗研究，以及各種變因的設計方法與操作時應該注意的問題，希望能提供大家作為研究時的參考。

### 一、海關檢疫抽樣方法效果的比較

依照 2004 年紐西蘭農林部(MAF)的資料，該國進口香蕉平均每次運送量為 22,500 箱，每箱裝 16 串(hands)，總規模達 36 萬串。依照紐西蘭檢疫標準，進口新鮮農產品的病蟲害感染率需在 0.005 以下，故每次需抽樣 600 個單位(香蕉的抽樣單位為一串)，以確保在 95%的信心水準下，只要有 1 個單位有問題即可被檢出。

由於現行的簡單隨機取樣法(Simple Random Sampling, SRS)操作上較為

費時，為簡化檢疫程序，海關檢疫人員希望能使用操作較簡便的叢聚取樣法(Cluster Sampling, CS)進行抽樣，但卻不知道這樣做是否妥當。因此 Barron(2006)進行模擬試驗，比較在不同群聚水準與感染率下，使用簡單隨機取樣法(SRS)與叢聚取樣法(CS)對檢疫抽樣效果的差異<sup>(2)</sup>。

此次模擬試驗包含的變因如下：

### 1. 試驗材料：

依照紐西蘭平常實際進口規模，先在電腦中設定一個  $22,500 \times 16$  的資料陣列作為樣本空間以供後續處理，預設值設為 0，表示尚未受到感染。

### 2. 試驗處理：

此研究的處理為群聚水準的效果，也就是被感染的樣本是均勻分散或者集中在樣本空間的某些範圍內，其參數為群聚指標(Index of Aggregation, IA)，代表被感染的樣本聚集在樣本空間的  $1/IA$  範圍內。若  $IA = 1$ ，表示均勻散布沒有群聚現象，反之若  $IA = N$ ，則表示全部聚集在某 1 箱裡。

以圖 2 為例，假設某次進口量有 16 箱、每箱有 16 單位，若  $IA = 2$ ，表示被感染的樣本是以均勻分布的方式散布聚集在其中  $1/2$ ~也就是 8 箱的範圍內(圖 2 上黑色部分)。這時分別以 SRS 法與 CS 法取樣 16 單位，可能就會造成左右兩種不同結果：以 CS 法只取一箱的 16 單位進行檢驗，可能會漏掉受感染的箱子而未檢出(圖 2 下右)；而以 SRS 法抽檢 16 個箱子，每箱都抽一個單位進行檢驗，就比較有機會抽中受感染的樣本(圖 2 下左)，然而因為要檢驗 16 個箱子，耗費的人力和成本也相對提高。

為探討不同群聚性及感染率下這兩種抽檢方法的效果，本試驗的處理變因設定 4 種群聚指標( $IA = 1、2、10$  及  $20$ )，及 4 種感染率( $p = 0.005、0.01、0.03$  及  $0.05$ )，共 16 種組合。由於全部有  $22,500 \text{ 箱} \times 16 \text{ 單位/箱} = 36 \text{ 萬單位}$ ，故先分別逢機選取其中的  $22,500 / IA$  箱作為被感染範圍，再逢機選取其中  $360,000 \times p$  單位的樣本，將其資料值設為 1，表示該樣本已被感染。

### 3. 調查取樣：

本試驗比較兩種不同取樣方法的效果，同樣選取 600 個單位，SRS 法是以一串為單位，直接在全體樣本空間中逢機取樣其中的 600 單位作為樣本；CS 法則先以箱為取樣單位，先逢機選取其中的 38 箱，共得到  $16 \times 38 = 592$  個樣本，另外再選取 1 箱並抽取其中的 8 個樣本，以湊足 600 個樣本。由於試驗一開始即已設定，在樣本空間中確實存在一定數量的被感染者，若樣本中出現被感染者，即表示成功檢出，否則即視為未檢出。

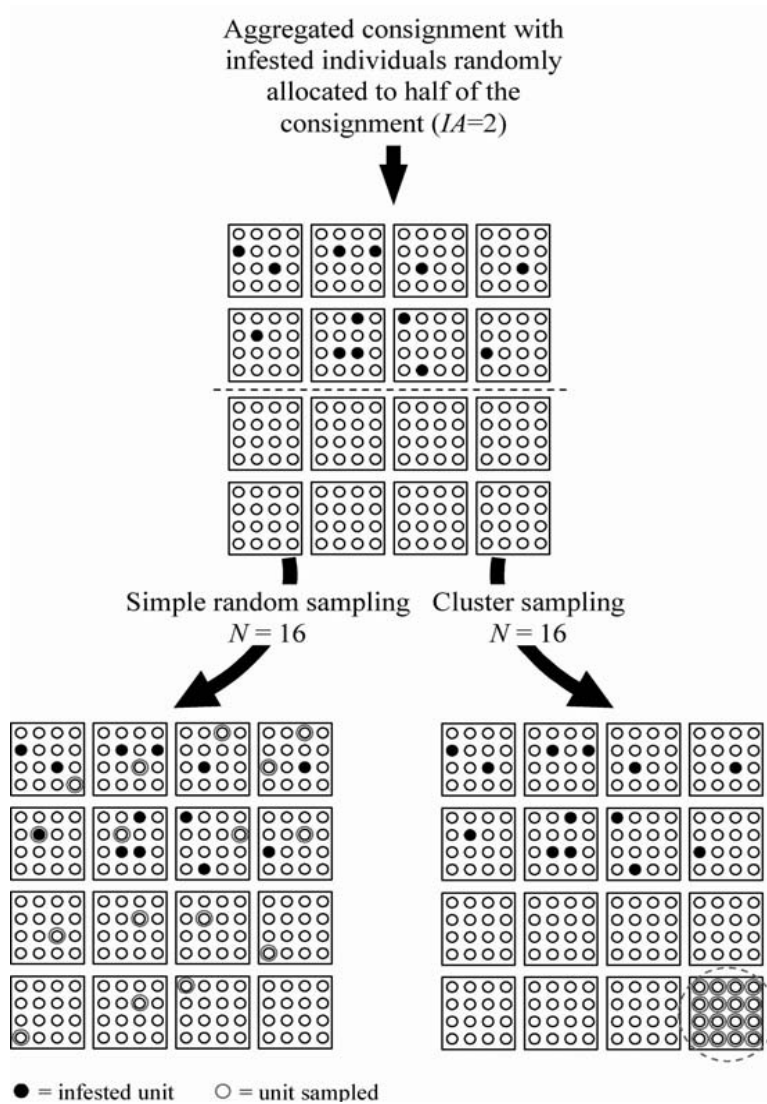


圖 2. 群聚指標  $IA=2$  的分布情形及兩種不同取樣方式結果差異的示意圖  
 Fig. 2. Diagram showing model allocation of infested units in an aggregated consignment and the difference between simple random and cluster sampling.

(Barron, 2006)

4. 重複試驗：

在本試驗中，共有 4 種感染率( $p$ ) × 4 種群聚指標( $IA$ ) × 2 種取樣方法 = 32 種組合，每種組合分別進行 100,000 次模擬試驗，分別計算實際上存在有被感染樣本時，使用這兩種取樣方法的未檢出率。由於檢疫的要求是當進口物品含有被感染個體時，便必須能加以檢出，因此若未能成功檢出，即視為失敗一次，同時累計失敗次數。而本試驗的目的，即在比較這兩種取樣方法的檢出效率。



### 5. 試驗結果：

彙整兩種取樣方法在不同感染率及群聚程度下的未檢出率(如圖 3)，結果顯示，SRS 法不論在何種群聚水準下未檢出率均保持一致，除在  $p = 0.005$  的感染率下未檢出率約 0.05 外，其餘三種感染率下的未檢出率均接近 0(圖 3a)。但 CS 法除群聚情形很低( $IA = 1$  或 2)時的未檢出率與 SRS 法結果類似外，隨群聚程度增加( $IA = 10$  或 20)，未檢出率即迅速提升至 0.1 或 0.2(圖 3b)。顯示隨樣本群聚情形提高，使用 CS 法進行抽樣可能有其風險。

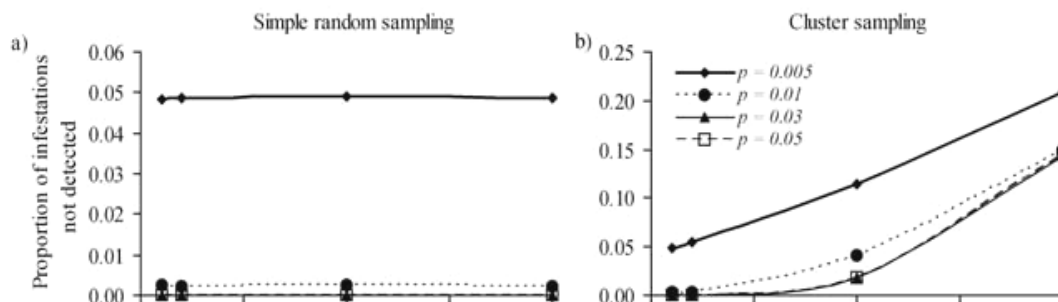


圖 3. 不同感染率及群聚水準下，兩種取樣方法 600 個樣本的未檢出率之比較

Fig. 3: The proportion of infestations not detected by a 600 unit sample (a & b) from simulations of simple random and cluster sampling under different levels of aggregation. (Barron, 2006)

## 二、病原體飄散傳播的空間分布模擬

為探討病原體(如病毒或孢子)在自然界傳播的情形，Xu 與 Madden(2004)以不同起始空間型態及不同程度的病原飄散傳播的距離進行模擬試驗，並以空間分析距離指標(Spatial analysis by Distance Indices, SADIE)來計算經過一段時間後病害傳播的空間分布情形<sup>(11)</sup>。說明如下：

### 1. 試驗材料：

假設田區有 150 行、每行 150 株的植株，其行株距均為 1 單位( $dc = dr = 1$ )，故先在電腦中設定  $150 \times 150$  的資料陣列。實際調查時則只取中央的  $120 \times 120$  區域進行取樣，周圍 15 行為保護行，故忽略不計。由於田區為二維空間，因此理論上病原可由罹病株向任意方向飄散進行傳播，但為簡化計算，將此二維空間簡化為網格(grid)型態，也就是只考慮水平與垂直方向的傳播。

首先設定起始罹病植株( $y_0$ )的數量，分別設為 4 株及 9 株兩種，再依以下 3 種空間分布型態( $y_s$ )分別進行設定，得到 6 種初始罹病狀態( $y$ )：

- (1) 群聚型(clump)：罹病株集中在田區最中央的  $2 \times 2$  或  $3 \times 3$  區域的植株。
- (2) 規則型(regular)：將田區分割為  $2 \times 2$  或  $3 \times 3$  的區塊(block)，再將罹病株放在各區塊的最中央。
- (3) 逢機型(random)：將 4 或 9 株罹病株逢機放在田區的任一位置。



## 2. 試驗處理：

本試驗共有以下幾種影響因子：

### (1) 病原傳播能力：

假設植株受感染後，會先經過5天的潛伏期，發病後會有10天的傳染能力，在此期間病原是否傳播，則服從平均每天0.7次的卜瓦松分布。因此可先從 $\lambda=0.7$ 的卜瓦松分布進行逢機取樣，得到當天的傳播次數，若次數大於0就進行下一個傳播的模擬步驟，否則就忽略不計，每個罹病植株都需連續進行10天的傳染過程。

### (2) 病原傳播距離：

影響微小粒子(如花粉、細菌、孢子等)傳播距離的因子很多，包括粒子本身的大小、重量、形狀，氣候條件如風力、風向、溫溼度等。一般來說，其傳播距離的機率密度大致呈現尖峰分布的型態，也就是越靠近傳播源的密度越高，距離越遠則密度越低。至於應採用何種分布則各家學者說法不一，例如 Dijk(1987)曾使用常態分布函數來描述風媒植物花粉的傳播密度<sup>(3)</sup>，Shaw(1995)則認為可能呈現指數(exponential)分布、雙指數(dual exponential)分布或科西(Cauchy)分布等不同狀況<sup>(8)</sup>。在本試驗中，Xu 與 Madden 採用 Shaw(1995)所建議的半科西(half-Cauchy)分布，其機率密度函數(pdf)及累計機率密度函數(CDF)分別為<sup>(7,8)</sup>：

$$f(y) = \frac{2}{\pi\sigma[1 + (\frac{y-\mu}{\sigma})^2]} \quad F(y) = \frac{2}{\pi} \arctan(\frac{y-\mu}{\sigma}) \quad (\text{Olive, 2008})$$

式中的平均傳播距離 $\mu$ 在本試驗設定為 $\mu=0.25、0.50、1.0、2.0$ 和 $4.0$ 等5種傳播距離。由於病原可向東西南北四個方向傳播，故先選取1-4的逢機數字以決定傳播方向後，再選取服從半科西分布的亂數來決定病原傳播距離。若距離超過行距或株距的一半( $dc/2$ 或 $dr/2$ )，表示病原已傳播至鄰接的植株，便將之設定已罹病，否則就忽略不管。新罹病株在5天潛伏期後也會有10天的傳染期，繼續傳染其他植株。

### (3) 試驗停止期：

為探討病害傳播的空間分布情形，作者設定罹病株比例需達到全體的65%以上才停止模擬試驗。依照起始分布和傳播距離等參數的不同，最後經過48-145天後才停止。停止期的設定對電腦模擬試驗十分重要，若沒有設定停止條件，電腦可能會不斷計算下去而永遠沒有結果。

## 3. 調查取樣：

將 $120 \times 120$ 的田區分別劃分為 $25(5 \times 5)$ 、 $36(6 \times 6)$ 、 $100(10 \times 10)$ 及 $144(12 \times 12)$

等 4 種不同的取樣樣方大小(sampling quadrat size)，用以計算其群聚指標(Index of Aggregation,  $I_a$ )，計算公式為： $I_a = D/E_a$ 。其中  $D$  為從現有空間分布狀態移到成為規則型空間型態所需的最短移動距離，又稱為規則性距離； $E_a$  則是先將現有空間分布狀態進行  $r$  次隨機排列，分別計算其最短移動距離  $D_i, i = 1 \sim r$ ，最後計算此  $r$  次排列的平均  $D_i$  值，即為  $E_a$  值。因此群聚指標  $I_a$  的大小，即可用以判定此空間分布型態與完全隨機排列形成的分布型態的群聚程度，當  $I_a > 1$  時，表示群聚程度較高，稱為群聚型(clump)；當  $I_a < 1$  時，表示分散程度較高，稱為規則型(regular)；當  $I_a$  接近 1 時，表示分散程度與隨機分布差不多，稱為隨機型(random)。

#### 4. 重複試驗：

本試驗共有 2 種起始罹病植株數( $y_0$ ) $\times$ 3 種空間分布型態( $y_s$ ) $\times$ 5 種平均傳播距離( $\mu$ ) = 30 種組合，每種組合分別進行 100 次模擬試驗，隨後分別以 4 種不同的取樣樣方大小計算其平均群聚指數  $I_a$ ，以探討其群聚情形。

#### 5. 試驗結果：

各種參數組合下的平均群聚指數  $I_a$  如表 2 所示。另外以  $10 \times 10$  的取樣樣方大小為例，在  $\mu=0.25$  及  $\mu=4.0$  等兩種平均傳播距離的條件下，3 種空間分布型態的最終傳播結果的罹病株數及等高線分布情形如圖 4 所示。

表 2. 不同取樣樣方大小、平均傳播距離及罹病株起始分布型態下進行 SADIE 分析所得的平均群聚指標  $I_a$  大小

Table 2. Average  $I_a$  index of SADIE for each combination of sampling quadrat size, median spore dispersal distance and initial pattern of diseased plants.

(Xu and Madden, 2004)

Initial pattern	Median spore dispersal distance <sup>a</sup>	Quadrat size			
		5 × 5	6 × 6	10 × 10	12 × 12
Clumped	0.25	3.24 <sup>b</sup>	2.82	1.91	1.67
	0.5	3.30	2.87	1.94	1.69
	1	3.33	2.89	1.94	1.70
	2	3.40	2.95	1.99	1.73
	4	3.53	3.06	2.04	1.77
Regular	0.25	2.24	1.97	1.41	1.24
	0.5	2.15	1.89	1.34	1.16
	1	2.54	2.23	1.57	1.38
	2	2.42	2.12	1.49	1.29
	4	2.79	2.45	1.71	1.49
Random	0.25	4.29	3.73	2.53	2.22
	0.5	4.19	3.64	2.47	2.16
	1	4.25	3.69	2.49	2.18
	2	4.50	3.91	2.64	2.30
	4	4.66	4.04	2.72	2.37

Clumped initial pattern, $\mu = 0.25$										Clumped initial pattern, $\mu = 4.0$													
5	7	5	7	0	8	12	3	19	11	0	2	0	2	4	11	9	15	10	8	10	12	2	2
2	9	8	7	4	2	10	9	50	57	1	0	6	6	2	8	36	27	17	5	7	7	8	7
15	17	23	12	10	34	27	35	6	12	8	6	1	11	8	19	27	35	18	38	23	11	8	3
13	88	40	32	69	43	49	39	20	8	1	2	6	8	14	31	53	62	72	65	62	32	18	16
16	37	37	39	73	84	80	49	44	6	10	24	4	26	28	51	85	82	92	97	78	52	15	20
5	74	97	88	97	100	100	88	29	14	12	38	5	15	41	74	98	99	99	91	60	41	37	23
3	41	58	88	99	99	99	98	68	9	1	3	4	16	28	65	85	99	100	64	52	22	21	16
7	8	8	38	68	94	93	76	29	8	12	0	5	22	17	47	73	84	78	56	38	19	14	6
5	7	22	35	66	41	46	24	24	33	16	5	2	6	11	12	36	45	42	34	24	20	5	14
1	33	32	7	22	21	13	5	7	38	4	1	2	4	5	7	12	11	12	16	6	7	3	0
4	3	2	1	3	2	6	2	1	2	9	1	5	6	2	7	9	8	7	14	6	9	4	0
3	2	3	1	3	5	8	8	0	0	0	1	24	54	5	5	2	0	3	1	1	4	2	1

Regular initial pattern, $\mu = 0.25$										Regular initial pattern, $\mu = 4.0$													
2	5	11	1	2	3	18	2	10	5	10	6	5	13	23	14	13	12	30	36	32	8	4	2
1	5	23	13	5	2	8	10	62	42	21	17	20	56	59	35	19	20	15	52	47	18	3	3
7	20	75	82	36	20	3	25	86	87	26	2	24	79	96	73	41	22	11	29	55	62	8	5
4	24	69	74	16	4	3	23	88	86	27	3	30	58	80	86	40	10	14	27	43	75	19	6
3	37	35	14	6	5	9	14	33	35	20	9	18	29	37	41	40	37	21	19	13	25	27	5
3	9	15	24	4	4	30	13	14	43	7	9	15	14	19	23	17	9	8	12	9	26	21	4
1	2	34	48	12	10	38	10	14	17	9	15	9	18	24	18	16	9	8	7	15	24	41	9
12	24	34	48	21	9	12	26	41	49	9	14	8	18	45	42	29	31	15	18	15	36	50	24
32	72	95	96	43	19	10	60	100	96	29	58	24	45	88	75	51	18	13	19	61	72	53	32
35	61	97	97	47	27	15	38	98	89	29	23	25	34	66	69	25	11	11	31	80	70	34	47
6	22	60	92	33	10	7	61	47	31	49	19	7	12	23	28	28	17	6	23	29	23	15	17
6	3	20	28	7	6	4	10	25	13	10	4	5	1	19	9	26	8	7	6	16	6	4	3

Random initial pattern, $\mu = 0.25$										Random initial pattern, $\mu = 4.0$													
2	2	3	33	4	0	0	0	5	11	7	7	2	4	5	9	3	7	31	50	70	55	59	24
2	0	1	3	1	2	7	2	10	24	29	22	2	21	21	8	6	13	38	59	98	90	66	24
0	0	0	0	3	1	7	11	35	78	79	31	12	31	16	15	9	17	34	75	99	88	57	28
3	3	1	2	2	4	9	19	56	100	97	50	15	36	43	23	24	23	25	64	97	80	38	25
1	0	0	0	6	4	4	12	62	99	92	56	25	71	87	42	26	13	14	43	65	83	75	38
1	2	9	20	23	5	12	32	45	68	95	79	34	61	80	57	14	9	16	25	35	56	87	58
1	7	6	25	14	4	7	11	21	52	100	100	10	27	27	15	16	9	14	21	19	55	41	27
5	54	62	79	64	9	7	9	18	45	89	61	24	37	18	12	10	5	5	27	25	23	13	6
22	57	94	95	31	10	12	31	43	74	53	53	13	47	16	8	5	1	3	12	8	11	6	1
13	49	75	32	9	19	13	67	94	98	46	36	1	12	21	2	4	14	11	22	8	17	3	4
14	42	7	12	13	6	11	27	77	99	50	33	1	2	3	2	2	12	8	9	8	1	1	4
6	2	2	2	16	1	2	30	33	37	16	2	0	0	4	1	2	13	4	5	1	0	2	1

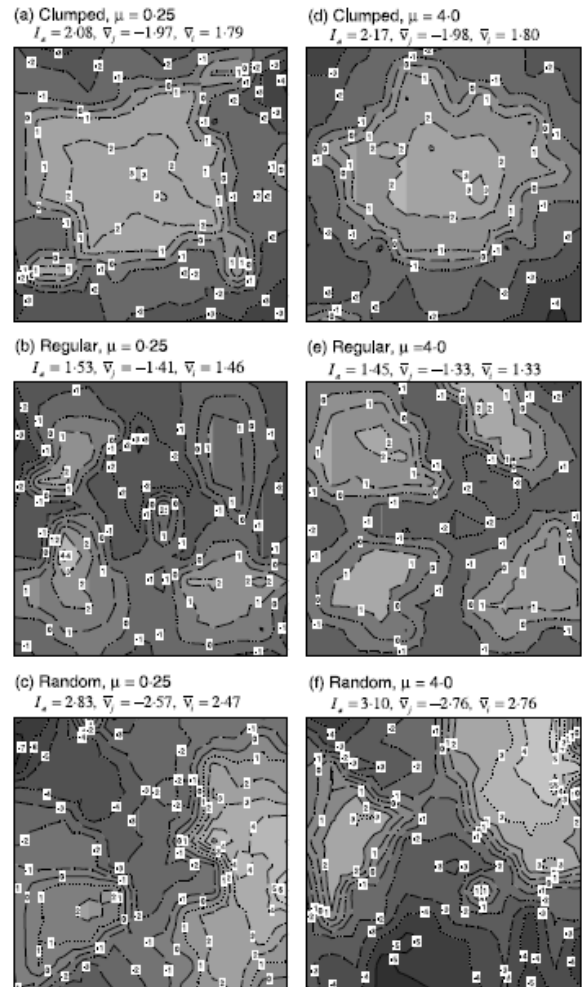


圖 4. 在 10×10 的取樣方大小下，兩種平均傳播距離及 3 種起始空間分布型態的最終傳播結果的罹病株數及等高線分布情形

Fig 4. Six examples of mapped quadrat (10×10) data of number of diseased plants from simulated epidemics, and contour maps of individual cluster indices ( $V_i$ , positive and  $V_j$ , negative) for the six selected simulation treatments. Selected spatial statistics are given above each contour map. Mean disease incidence was  $\approx 0.25-0.27$  for each data set. (Xu and Madden, 2004)

### 三、兩種範例試驗的比較

前面所介紹的兩個範例中都用到群聚指標(Index of Aggregation, IA)這個參數，然而其意義卻完全不同。至於二者的試驗進行和亂數的產生方式更是相差甚遠，前者的目的在探討兩種抽樣方式的差異，因此假定被感染的樣本是均勻分散或者集中在樣本空間的某些範圍內(依其群聚指標大小而定)，隨後再以不同抽樣方式觀察是否能抽中被感染的樣本；而後者的目的則是觀察

病害傳播的情形，因此設定許多的參數和傳播模式(如每天傳播次數為卜瓦松(Poisson)分布，而平均傳播距離為半科西分布等)，試圖模擬病原體在真實田間環境下的傳播情形。然而，關於病原傳播距離的機率分布模式各家學者說法不一，亦有學者指出可能呈現指數(exponential)分布、雙指數(dual exponential)分布或科西(Cauchy)分布等不同狀況<sup>(8)</sup>。如果改採這些機率分布進行模擬，或許也會得到不同的結果<sup>(11)</sup>。

### 討論

從前面的介紹看來，蒙地卡羅模擬法可以有效控制各種變因，使試驗能在理想狀態中進行，而不受試驗材料及參試處理能否保持均質的影響；並且可以任意改變各變因的參數值再重新進行模擬試驗，來觀察在不同參數組合下所得結果的差異情形，而不用實際進行這麼多次試驗，因此成為一項重要的研究工具。

然而，在使用蒙地卡羅法進行模擬試驗前，必須對問題本身進行了解，這樣才能知道問題的特性，以便針對其獨特的機率分布模式產生相對應的亂數，進而得到最終的「觀測值」，因此不能隨意套用其他模擬試驗的模式和作法來進行模擬運算。因為各種試驗研究，都各有其不同的背景條件，也各有不同的模式來解釋這些問題發生原因及特性。若不先了解問題本身的特性以配合正確的模式，而任意取用其他試驗所使用的解法，這樣不但無法解決問題，可能還會得到錯誤的答案，造成更大的問題。

對問題先作了解的另一個目的是可以減少問題的複雜度，雖說現在的電腦運算速度已經大幅提昇，但若不考慮真正的問題所在，即使使用目前最先進的高速電腦設備，還是會浪費大量不必要的運算時間。在目前研究計畫經費和人力普遍不足的情況下，如果能先簡化問題，減少其中不必要的變項，就可以大幅減少模擬計算次數及所需要的時間，讓一般研究人員也能利用手邊的個人電腦來進行模擬運算，達到最佳的效率。

除了對問題本身的了解外，另一個常被忽略的是模擬運算本身可能產生的問題，包括所取得的逢機數字是否真的是隨機產生，以及所使用的模擬運算程式在語法甚至邏輯上是否有問題等等。這是因為蒙地卡羅法本身大量使用逢機數字，如果這些數字本身並非隨機產生，就可能會有分布不夠均勻，或者有奇怪的規律性出現，這時不但無法隨著取樣數的增加而使得估算精密度提升，甚至可能產生錯誤的估算結果。至於程式本身語法和邏輯上的問題所產生的影響就更不在話下，使用錯誤的程式自然會產生錯誤的估算結果。因此在正式的大規模運算進行前，最好能先進行一些簡單的測試運算，來檢

測程式本身所得到的結果是否正常。此外，由於電腦模擬試驗可能使用許多假設狀況，先假定某些環境條件均質而對試驗不會產生干擾，以簡化計算的複雜度。因此對於試驗結果也要小心評估，以免不慎產生誇大不實的結論<sup>(4,5)</sup>。

對農業試驗研究來說，由於研究對象是生物，其複雜性較高，難以使用簡單的數學模式描述作物生長發育等現象，因此也難以使用電腦模擬整個生長發育過程。目前所發展出來的許多作物生長模式，往往因為對於真正的生長發育機制仍無法完全掌控，因而無法完全適用於所有環境和品種。如果想要有效的模擬出可以適應各種生長環境(包括在許多開發中國家常見的間作和輪作制度)和品種組合的生長模式，就必須了解更多關於植物生理和回饋機制等方面的基礎知識，並配合更為嚴格的建模技術(像是已知流程的標準化模組)，並且需要與其他技術例如統計和電腦模擬技術等加以結合才行。否則，一個只能適用於特定狀況下的生長模式，仍然只能當作參考用<sup>(6)</sup>。

#### 參考文獻

1. 余清祥. 2007. 3.模擬、10.模擬檢定、11.Bootstrap. 政治大學統計系「統計計算與模擬」課程上課講義.
2. Barron, M. C. 2006. Effects of aggregation on the probability of detecting infestations in fresh produce consignments. *New Zealand Plant Protection*. 59:103-108.
3. Dijk, H. V. 1987. A method for the estimation of gene flow parameters from a population structure caused by restricted gene flow and genetic drift. *Theor. Appl. Genet.* 73:724-736.
4. Gentle, James E. 2003. Random number generation and Monte Carlo method. Springer-Verlag, New York.
5. Liu, J. S. 2001. Monte Carlo strategies in scientific computing. Springer-Verlag, New York.
6. Mohammadi, S.A. and B.M. Prasanna. 2003. Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci.* 43:1235- 1248.
7. Olive, D. J. 2008. The Half Cauchy Distribution. *A Course in Statistical Theory*, Chapter 10. Some Useful Distributions. p.300.
8. Shaw, M.W. 1995. Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. *Biol. Sci.* 259(1356):243-248.

9. Wichmann, B. A. and I. D.Hill. 1982. Algorithm as 183: An efficient and portable pseudo-random number generator. Applied Statistics 31(2):188-190.
10. Wichmann, B. A. and I. D.Hill. 1984. Correction: Algorithm as 183: An efficient and portable pseudo-random number generator. Applied Statistics 33(1):123.
11. Xu, Xiangming and L. V. Madden. 2004. Use of SADIE statistics to study spaial dynamics of plant disease epidemics. Plant Pathology 53:38-49.