

國立交通大學  
工業工程與管理學系  
碩士論文

建構太陽能發電系統之三階段轉換效率  
預測模型

Constructing a Three-Stage Prediction Model for  
Conversion Efficiency of Photovoltaic Systems

研究生：呂哲嘉

指導教授：唐麗英 博士

洪瑞雲 博士

中華民國九十九年六月

建構太陽能發電系統之三階段轉換效率預測模型

Constructing a Three-Stage Prediction Model for Conversion Efficiency  
of Photovoltaic Systems

研究生：呂哲嘉

Student : Je-Jia Lu

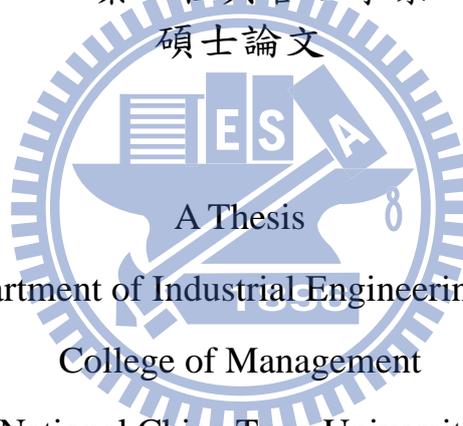
指導教授：唐麗英 博士

Advisor : Lee-Ing Tong

洪瑞雲 博士

Ruey-Yun Horng

國立交通大學  
工業工程與管理學系  
碩士論文



Submitted to Department of Industrial Engineering and Management

College of Management

National Chiao Tung University

In Partial Fulfillment of the Requirements

for the Degree of

Master

in

Industrial Engineering and Management

June 2010

HsinChu, Taiwan, Republic of China

中華民國九十九年六月

# 建構太陽能發電系統之三階段轉換效率預測模型

研究生：呂哲嘉

指導教授：唐麗英、洪瑞雲

國立交通大學工業工程與管理學系碩士班

## 摘要

由於目前發電能源大多為化石燃料或核能，導致這些發電能源正快速消耗，且這些能源在發電的過程中會汙染環境。因此世界各國均非常重視再生能源如太陽能源、風力等的開發；其中太陽能源取之不盡、用之不竭，且太陽能系統發電時對環境產生的汙染很小，因此太陽能成為再生能源中備受矚目的新興能源。然而太陽能發電系統之轉換效率存在技術上的瓶頸，且轉售的電能不及高額的設置成本，成為廠商投資太陽能設備最大的阻礙，因此廠商在進行投資前須準確地預測系統的轉換效率。本研究針對此問題提出一套三階段的太陽能系統轉換效率預測模型。首先利用逐步迴歸法篩選出影響轉換效率的顯著變數；接著使用集群分析法對資料進行分群，以提高預測模型之準確率；然後再建構一套結合自組性演算法與遺傳規劃法的混合預測模型來預測轉換效率，藉由調整殘差來提高預測模型之準確率。本研究最後利用國內某太陽能系統轉換效率之實例來說明本研究所提出之三階段預測模型確實有效可行。

**【關鍵詞】** 太陽能發電系統、系統轉換效率、預測模型、逐步迴歸分析、集群分析法、自組性演算法、遺傳規劃法

# **Constructing a Three-Phase Prediction Model for Conversion Efficiency of Photovoltaic Systems**

Student : Je-Jia Lu

Advisors : Lee-Ing Tong  
Ruey-Yun Horng

Department of Industrial Engineering and Management  
National Chiao Tung University

## **Abstract**

Most of the electricity is generated from resources such as fossil fuels and nuclear power. These resources are facing depletion due to the generation of energy. Moreover, the process of generating electricity will produce pollution. Considering the crisis of energy shortage and environmental protection, many countries emphasize the development of renewable energy such as solar and wind energy. Because solar energy is inexhaustible and the pollution is low when the photovoltaic (PV) system is utilized to generate the electricity, the solar energy is considered as one of the most important renewable energy. However, the conversion efficiency of PV system is restricted to engineering technique. The biggest obstacle of investing the PV system is that the profit of electricity produced by the PV system is less than the high set-up cost of the PV system. Therefore, before investing on the PV system, it is very important to accurately predict the conversion efficiency obtained from the PV system. This study proposes a three-stage prediction model for the conversion efficiency of PV systems. First, the stepwise regression analysis is utilized to select the significant variables for predicting the conversion efficiency of a PV system; then cluster analysis is employed to classify the data into two clusters to improve the prediction accuracy of the prediction model. Finally, a hybrid model which combines the GMDH method and Genetic Programming is adopted to construct the prediction model. The hybrid prediction model is enhanced by adjusting the residuals of the prediction model. One real case is utilized to demonstrate the effectiveness of the proposed model.

**【Key word】** Photovoltaic System, Conversion Efficiency, Prediction Model, Stepwise Regression Analysis, Cluster Analysis, GMDH, GP;

# 目 錄

摘要.....	i
英文摘要.....	ii
目錄.....	iii
表目錄.....	v
圖目錄.....	vi
<b>第一章 緒論</b> .....	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	4
1.3 研究方法.....	4
1.4 研究架構.....	4
<b>第二章 文獻探討</b> .....	6
2.1 太陽能電池簡介.....	6
2.2 太陽能發電系統介紹.....	7
2.4 集群分析法.....	10
2.5 自組性演算法.....	12
2.6 遺傳規劃法.....	16
2.7 總結.....	18
<b>第三章 研究方法</b> .....	19
3.1 研究架構.....	19
3.2 第一階段-篩選顯著變數.....	21
3.3 第二階段-資料分群.....	22
3.4 第三階段-建構轉換效率混合預測模型.....	23
<b>第四章 實例驗證</b> .....	25
4.1 第一階段-篩選顯著變數.....	25
4.1.1 資料蒐集與整理.....	25
4.1.2 利用逐步迴歸法篩選出顯著變數.....	26
4.2 第二階段-資料分群.....	29
4.2.1 利用階層式集群法決定最適分群數.....	29
4.2.2 使用 k-means 法進行分群.....	30
4.3 建構轉換效率混合預測模型.....	31
4.3.1 利用自組性演算法對各群觀測值進行預測.....	31
4.3.2 利用遺傳規劃法調整殘差.....	33
4.4 三階段預測模型之應用.....	36
<b>第五章 結論與建議</b> .....	40
5.1 研究貢獻.....	40
5.2 後續研究.....	41



# 表目錄

表 2.1 矽太陽能電池種類比較表 .....	7
表 2.2 太陽能發電系統之相關設備介紹 .....	8
表 2.3 GA 與 GP 之差異比較表 .....	17
表 3.1 系統資料的解釋變數與反應變數 .....	21
表 4.1 太陽能發電系統原始資料整理 .....	26
表 4.2 逐步迴歸分析結果 .....	26
表 4.3 解釋變數之共線性 .....	27
表 4.4 顯著變數的重要性 .....	28
表 4.5 集群過程的距離報表 .....	29
表 4.6 k-means 法分群結果 .....	30
表 4.7 集群分析之 ANOVA 結果 .....	31
表 4.8 各變數中心點 .....	31
表 4.9 GMDH 預測結果 .....	32
表 4.10 集群一內部驗證結果 .....	33
表 4.11 集群二內部驗證結果 .....	33
表 4.12 預測模型之準確性評估 .....	35
表 4.13 單一階段預測方法結果 .....	35
表 4.14 預測模型之準確性評估 .....	36
表 4.15 邏輯斯迴歸之參數估計表 .....	37
表 4.16 邏輯斯迴歸之分類正確率 .....	38
表 4.17 新資料內容 .....	39

# 圖目錄

圖 1.1 台灣歷年進口能源依存度 .....	2
圖 2.1 太陽能電池發電原理 .....	6
圖 2.2 太陽能電池組裝示意圖 .....	7
圖 2.3 獨立型太陽能蓄電系統 .....	8
圖 2.4 市電併聯型太陽能發電系統 .....	9
圖 2.5 緊急防災型太陽能發電系統 .....	9
圖 2.6 兩階段集群法演算流程圖 .....	11
圖 2.7 自組性演算法之網路架構圖 .....	13
圖 2.8 GMDH 演算法之輸入層 .....	13
圖 2.9 GMDH 演算法各層變數組合過程 .....	14
圖 2.10 語法樹示意圖 .....	16
圖 3.1 建構三階段轉換效率預測模型之流程圖 .....	20
圖 4.1 常態性機率圖 .....	27
圖 4.2 殘差分析圖 .....	28
圖 4.3 分群樹狀圖 .....	30
圖 4.4 實際值與預測值比較折線圖 .....	34
圖 4.5 應用三階段轉換效率預測模型之流程圖 .....	38

# 第一章 緒論

本章介紹本研究之背景與動機，研究目的以及研究方法，最後介紹本篇論文章節之架構。

## 1.1 研究背景與動機

自十九世紀工業革命後，生產技術有了突破性的進步，生產型態也逐漸由機器生產取代傳統人力從事生產工作，不僅可加快生產速度，使全世界經濟快速成長，連帶也改變了人類的生活型態。電已成為人們生產作業與生活中不可缺少的能源，世界各國紛紛投入能源發電的研究行列。近代電能需求大幅上升，各種發電的原料含量快速減少，煤炭、天然氣、石油等自然資源消耗迅速，其中石油價格的波動更是連帶影響了全世界的經濟，根據英國知名石油公司 British Petroleum(BP)[32]指出，全球原油剩餘可開採量為 42 年，天然氣儲量為 60 年，而鈾礦蘊藏量僅剩 53 年，這些有限資源即將耗盡，能源危機的出現讓許多國家砸下龐大資金研發新的發電能源。

人類所面臨問題不單只有能源危機，過度依賴化石能源或是鈾原料進行發電，對於地球生態環境產生莫大的影響。利用化石燃料發電時所排放的大量溫室氣體，如二氧化碳(CO<sub>2</sub>)、甲烷(CH<sub>4</sub>)等使全球氣溫年年升高，形成溫室效應，影響了各地生態環境。利用鈾礦所進行的核能發電，用來冷卻核反應爐的廢水排入大海，對海洋生態圈產生不良影響；此外核能發電的輻射線及其安全性仍有疑慮，核能發電產生的核廢料更是燙手山芋。因此政府在制定能源決策時，除了考慮經濟發展外，也必須盡到保護生態環境的責任。

全球科技發展日新月異，經濟與文明不斷進步，能源短缺與環境生態保護兩大問題日益嚴重。隨著環保聲浪高漲，各國在追求經濟發展的同時

必須正視環境問題，因此紛紛投入大量資金與人力開發風力、水利、太陽能等低污染的再生能源(renewable energy)。太陽能成為眾多再生能源中備受矚目的新能源，相較其他再生能源，太陽能在取得上不易受到地形限制，且太陽能取之不盡，用之不竭，對環境的污染很小，因此近年來歐美、日本等已開發國家大力地發展太陽能發電系統。

台灣地小人稠，天然資源含量不多。為了維持工業生產及經濟活動，除了開採本島的資源外，大部份的能源仍仰賴國外進口以維持供電量。根據日前台灣經濟部[16]所發佈的資料顯示，台灣能源安全持續惡化，2008年進口能源依存度已高達 99.35%，如圖 1.1 所示，如此高的進口能源依存度對台灣的經濟產生一定程度的影響。此外，台灣所需能源中比例最高為石油，佔台灣能源總供給量的一半以上；台灣對於進口石油依賴性高，一旦國際油價市場出現震盪波動，勢必對台灣的經濟造成很大的衝擊，因此政府推動太陽能發電系統的設置，以太陽能替代化石燃料作為供電來源，是一項不錯的選擇。

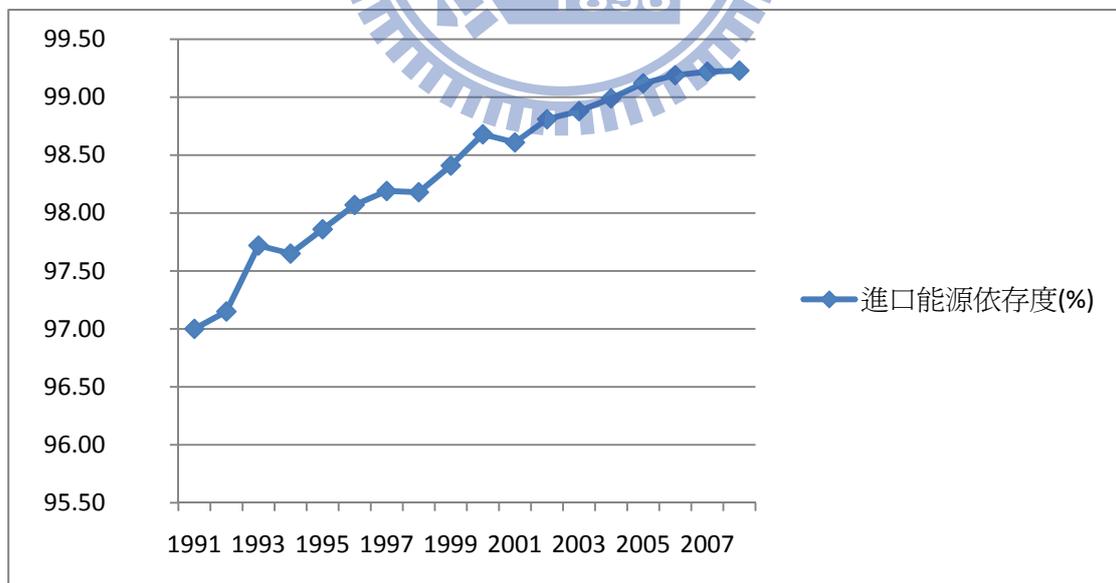


圖 1.1 台灣歷年進口能源依存度[30]

台灣位處太陽輻射能豐富的亞熱帶地區，太陽輻射能每平方公尺超過

3000 千卡，理論上是非常合適架設太陽能發電板的設置地區[33]。根據經濟部能委會的資料顯示，台灣太陽能發電系統的設置容量雖然逐年提升，至 2008 年已設置四百萬瓦，但目前台灣電力供應來源主要依賴核能發電與火力發電，太陽能提供的電力占所有電力供應的比例不到 0.1%，太陽能發電系統技術仍然有許多待克服的難題，不論是外在環境的使用或是技術瓶頸，使太陽能發電系統的轉換效率與供電穩定性欠佳，低於其他發電系統許多。除了技術層面的問題，設置太陽能發電系統需佔用廣大的土地面積，欲設置一座與核能廠供電量相當的太陽能電廠，其設置面積約為核能廠面積的 160 倍，台灣地狹人稠，土地成本高，高昂的設置成本成為廠商投資太陽能產業一大障礙，因此如何提升太陽能發電系統之收入效益絕對是廠商的首要工作。

馮焯生[15]在探討太陽能電池發電原理時，指出若太陽光譜不變，日照強度並不會影響轉換效率，因此轉換效率可作為衡量系統效能優劣的指標。而欲發展太陽能發電系統必須提高轉換效率來平衡成本，如何提高太陽能發電系統的轉換效率相當重要的，目前最先進的太陽能發電系統其轉換效率略高於 19%[34]，而系統的模組零件會影響系統轉換效率。過去探討轉換效率的相關研究，多以實驗室模擬數據為基準，但許多影響系統轉換效率的外在因素無法在實驗室進行模擬，然而這些因素在太陽能系統實際運作時卻相當重要。

本研究利用統計方法分析系統實際運作的資料，並藉由系統模組的變數來預測轉換效率。目前預測轉換效率的文獻並不多，高翊倫[9]使用了遺傳規劃法(Genetic Programming, GP)建構轉換效率預測模型，利用已知的零組件規格去預測轉換效率，而近幾年有文獻提出了很多混合預測模型[23][29](hybrid model)，並驗證使用兩套以上系統的混合預測模型，其預測準確性較單一預測模型來的好。由於高翊倫僅使用單一方法建構預測模型，

其模型準確度仍有改善空間；本研究針對已實際運作的太陽能發電系統轉換效率資料，另提出一套三階段預測模型，此預測模型能有效改善單一預測模型的準確性。

## 1.2 研究目的

本論文主要目的為根據高翊倫[9]所提出之遺傳規劃法進行改善，利用一套結合自組性演算法與遺傳規劃法之混合模型來預測轉換效率。此外，將資料挑選出顯著變數再進行分群，以提升預測準確性，作為投資廠商的參考依據。

## 1.3 研究方法

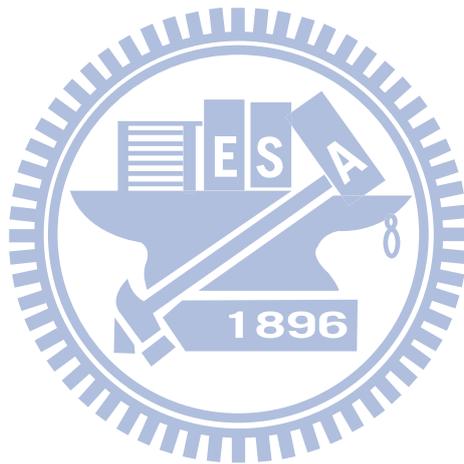
為達以上研究目的，本研究所使用之方法如下：

1. 篩選顯著變數：太陽能發電系統的轉換效率會受不同模組零件影響，將資料整理後，使用逐步迴歸法篩選出影響轉換效率的顯著變數，刪除與轉換效率無關的系統變數，並利用這些顯著變數來建構轉換效率預測模型。
2. 資料分群：由於各個發電系統的轉換效率會受到系統模組零件影響，所表現出來的轉換效率差距很大，若直接將轉換效率進行預測，會產生較大的誤差；利用集群分析法若先將資料進行分群，提高同一群資料的同質性，便能提高預測的準確性
3. 建構混合預測模型：使用自組性演算法建構轉換效率預測模型，並以遺傳規劃法調整殘差項，以提高預測模型的準確性。

## 1.4 研究架構

本論文共分五章，其內容分別如下：第一章為緒論，說明本研究之背

景與動機、研究目的、研究方法與研究架構；第二章為文獻探討，簡介太陽能發電系統、集群分析法、自組性演算法以及遺傳規劃法；第三章為研究方法，詳述本研究所提出的三階段預測模型建構的流程；第四章為實例說明，以太陽能發電系統實際運作資料建構系統轉換效率之三階段預測模型，並與其他文獻的預測結果進行評比，以證實本研究所提出之方法確實具可行性與有效性；第五章為結論與建議，簡述本研究的貢獻並提供未來研究之建議。



## 第二章 文獻探討

本章首先介紹台灣地區太陽能發電系統之相關資訊，接著分別介紹集群分析法、自組性演算法以及遺傳規劃法，並對相關文獻進行回顧。

### 2.1 太陽能電池簡介

太陽能電池(photovoltaic,PV)乃是利用太陽的光能，照射太陽能電池板後轉換成直流電力，再經電力轉換器轉換成需用的直流或交流電壓形式使用。就促進能源種類與來源多元化及永久性能源的觀點而言，太陽光電能是除了石化燃料、水力及核能發電之外，最具潛力值得全力開發的新能源；它不僅取之不盡用之不竭，而且沒有環境污染的問題。

以太陽光為能量來源的光電半導體薄片，在高純度的矽基材中，加入部分 3A 與 5A 族元素，便形成 P 型與 N 型半導體。當受到日光中某些波長的光照射時，會產生伏特效應(Photovoltaic effect)使得電池內部產生電位差，驅使電池內的電子移動而產生電流。

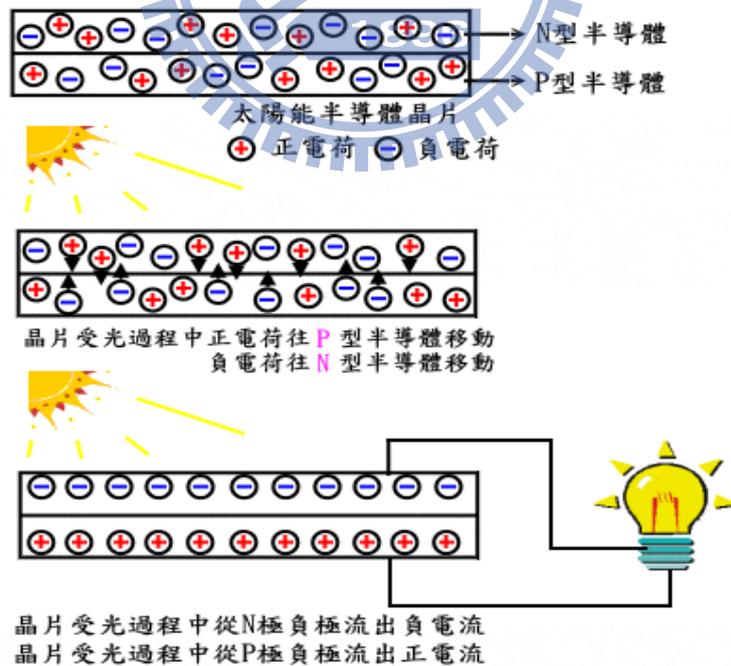


圖 2.1 太陽能電池發電原理[33]

製成太陽能電池可用許多種材料，目前主要使用的材料是矽(silicon)，約佔

96%以上的市場，矽電池分為三種，分別為單晶矽(Single Crystalline)、多晶矽(Poly Crystalline)與非晶矽(Amorphous)，差別在於矽原料內部晶體排列方式的不同，此三種不同矽電池的比較如表 2-1 所示。在本研究使用的資料中，所有的太陽能發電系統皆以矽做為太陽能電池的材料。

表 2.1 矽太陽能電池種類比較表[33]

太陽電池種類	半導體材料	耐用性	市場模組發電轉換效率
晶矽 Crystalline	單晶矽 Single Crystalline	高	12~20%
	多晶矽 Poly Crystalline	中	10~18%
非晶矽 Amorphous	Si、SiC、SiGe、SiH、SiO	低	6~9%

## 2.2 太陽能發電系統介紹

單一太陽能電池所輸出的電量有限，為了提高發電量，將許多太陽能電池經過串並聯的組合後，形成了太陽能電池模組(Solar module)，再將多個太陽能電池模組組合成為更大的方陣(array)，配上蓄電池、變壓器、控制器、與換流器等設備，形成一套太陽能發電系統(Photovoltaic system)[8]。

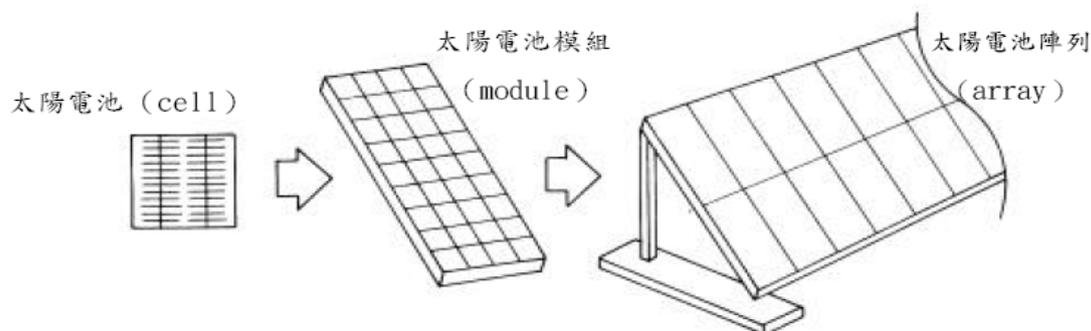


圖 2.2 太陽能電池組裝示意圖[8]

每個太陽能發電系統由許多設備所組成，表 2.2 簡介幾個太陽能發電

系統之相關設備，將名稱與功能整理如下：

表 2.2 太陽能發電系統之相關設備介紹

設備名稱	功用
蓄電池 (battery)	當太陽能發電系統之電力供過於求，蓄電池可儲存多餘電力，以備不時之需。
配電盤 (distribution panel)	通常設置於住宅內部，功能是将接收的電力分流到各個樓層或電器。
換流器 (inverter)	又稱逆變流器，將太陽能電池模組產生之直流電轉換為交流電，供一般家庭使用。
變壓器 (transformer)	在輸送電力過程中，若電壓過低則容易造成能量的損失，因此裝設變壓器可減少電力輸送時的耗損，提高太陽能發電的轉換效率。

目前台灣地區常見的發電系統有以下三種[33]：

1. 獨立型太陽能蓄電系統(圖)

為單一獨立發電系統，使用蓄電池，未與其他電力系統連結，白天太陽光電系統發電，並供負載及充電，夜間由電池供電，可以自給自足。適用地點為高山、離島、基地台等市電無法到達處。

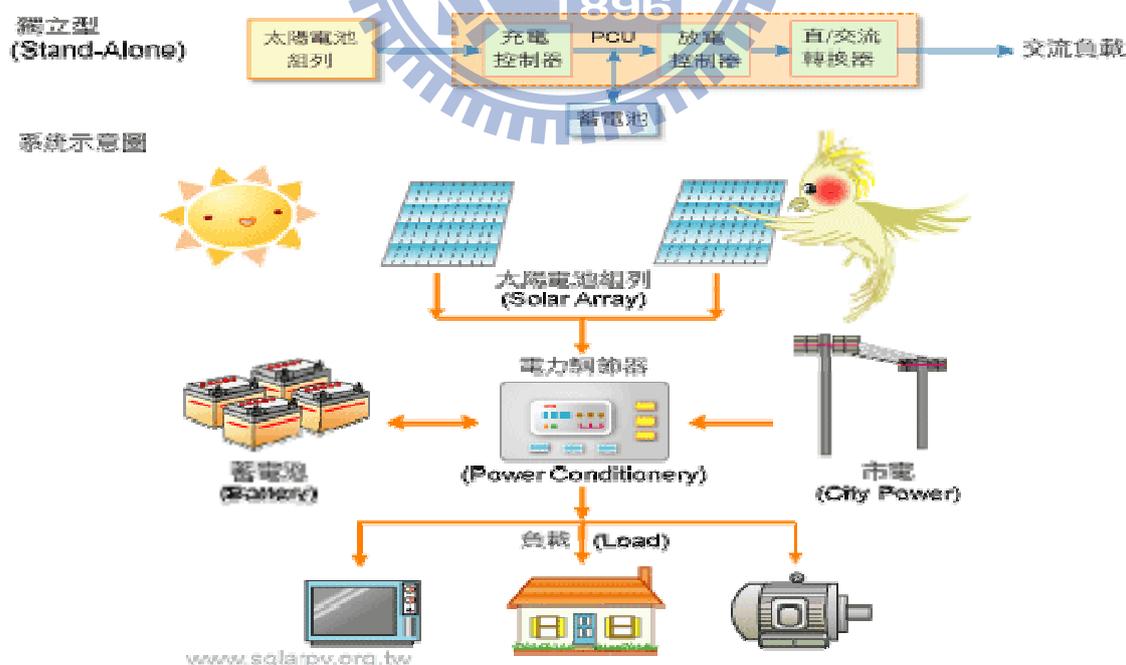


圖 2.3 獨立型太陽能蓄電系統[33]

## 2. 市電併聯型太陽能發電系統(圖 2.4)

此類型的發電系統特色是，市電負載併聯，平時與太陽光電系統併聯發電，並供負載，不夠的電由台電供電。好比將市電電力系統當作一個無限大無窮壽命的免費蓄電池，不會影響一般電器的正常使用。此種發電系統發電效率高且太陽光之發電能量利用率高。

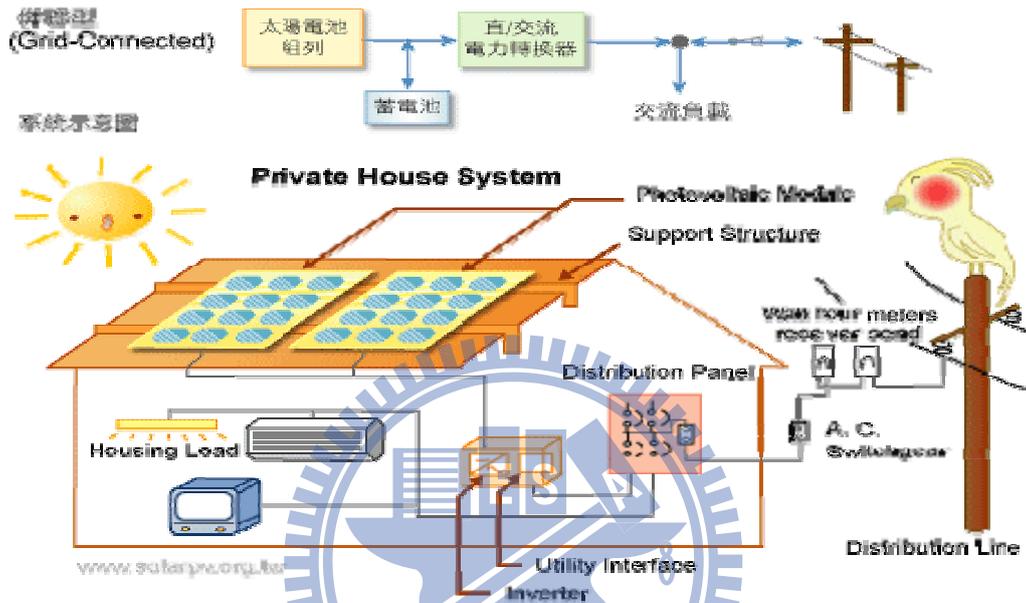


圖 2.4 市電併聯型太陽能發電系統[33]

## 3. 緊急防災型太陽能發電系統(圖 2.5)

此類型的太陽能發電系統其換流器(Inverter)具有逆送電功能，同時裝置蓄電池，可操作於併聯模式或獨立模式之太陽光電發電系統，在有防災需求(照明、汲水、通信…)之公共設施適用，但緊急防災型太陽能包含兩種系統，建置成本較高，系統較複雜

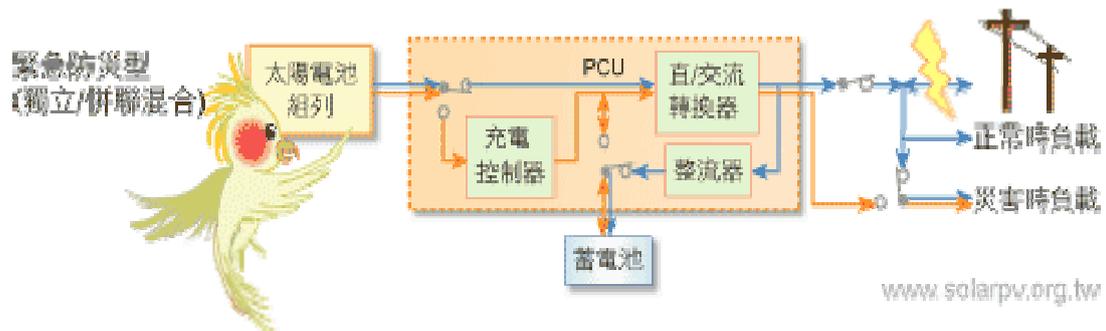


圖 2.5 緊急防災型太陽能發電系統[33]

## 2.3 太陽能系統轉換效率

根據高翊倫[9]提出的系統發電量定義，光能輸入與系統轉換效率為影響發電量的兩大重要因素，其公式如(2-1)：

$$\text{電能輸出(發電量)} = \text{光能輸入} \times \text{系統轉換效率} \quad (2-1)$$

此為系統發電量的定義，由此公式發現系統的發電量是由這兩項變數相乘得出。前項的光能輸入會受到氣候因素所影響，包含地表日照量、日照強度等，這些氣候因素皆無法由人為控制；後項的系統轉換效率則會因為系統內部各零組件規格特性差異而有所不同，例如模組面積、模組單片容量、模組種類等等因素會影響系統轉換效率，這些因素可經由技術上的突破或系統改良設計而提高，屬於人為可控因素，要提升發電量必須改良系統的轉換效率。馮堃生[15]探討太陽能電池的發電原理時，以工程觀點的角度說明，只要太陽光譜不變，日照強度並不會影響轉換效率，因此可將轉換效率是為衡量太陽能系統效能優劣之指標。

綜合上述可知，系統轉換效率的高低決定系統發電量的多寡，轉換效率扮演了舉足輕重的角色，本研究以探討轉換效率為主，將多個系統變數利用一套混合預測模型進行預測，利用已知的系統變數來預測此套系統的轉換效率值。

## 2.4 集群分析法

集群分析法[10][12]發展至今應用於很多領域，如企管、社會、醫學、心理、工程等科學。集群分析是將個體(受測者或變數)分成有意義群組的一種統計分析技巧，依據個體間的相似性，計算出個體的距離，依據距離大小將樣本分成幾群互相沒有交集的群組，使同群組內的個體相似性高，不同群組的個體相似性低。集群分析也可作為資料簡化的工具，目的是將

整個母體分成能提供資訊的幾個小群組，可以了解各小群體的資料特性，且只損失小部分資訊。

集群分析法分成階層式集群法與非階層式分群法，此兩種分群法各有其優缺點。階層式集群法不必事先知道要將資料分成幾群，計算出各資料的距離來進行分群，其缺點為當觀察值分群後，就不會再產生變動，容易受到距離較遠的資料所影響；非階層式集群法中最具代表性為 k-means 法，k-means 法進行分群的過程時，在每階段的分群中，資料會因距離大小而從原本的集群中跳脫出來，重新調整各點所歸屬的集群，避免受到距離太遠的點所影響，其缺點為使用 k-means 法必須先決定集群數目，若未得到較佳的集群數目，k-means 法分群結果未必為最佳分群結果。目前大多使用結合這兩種方法的「兩階段集群分群法」，此方法結合了階層式集群法與非階層式集群法，將這兩種方法的優缺點進行互補，得到最好的分群結果，利用階層式集群法的分群結果當作最佳分群數目，然後再利用 k-means 法調整分群，如此一來結合這兩種分群法的優點，得到最好的分群結果。圖 2.6 為兩階段集群法演算流程圖

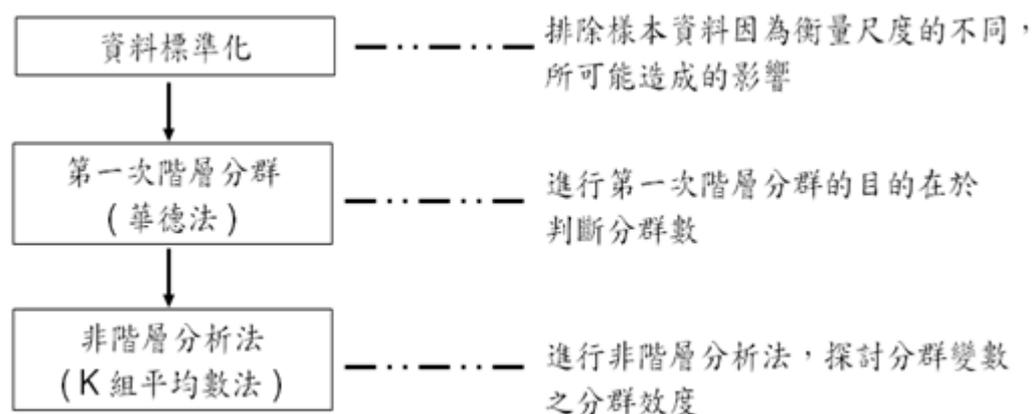


圖 2.6 兩階段集群法演算流程圖[7]

## 2.5 自組性演算法

自組性演算法(Group Method of data handling, GMDH)[4]為俄國學者 A.G.Ivakhenko[21]於 1969 年所發展，他認為在大多數的情況下，透過模式來建立目標與因子之間的關係時，其交互作用往往是最難以得知且不確定的，若在建構數學模式時加入太多主觀的判斷會影響到數學模式之正確性，所以在針對一系統或問題建構數學模式求解時，必須對該系統或問題有相當的了解，才能決定其影響變數間之關係。因此他假設應變數與自變數間呈現高階多項式的關係，且讓資料本身存在的訊息來決定模式之間的關係，也就是資料導向(Data-Oriented)的建模概念。在目標與變數之間關係不明確的系統中，自組性演算法可令變數進行配對、競爭的動作、直到訓練模式的效能達到最佳，或誤差收斂為止。其優點是對於複雜的非線性系統可作出較佳的推測，且不需要大量的訓練資料，並對時間週期較短的預測問題也有較佳的分析能力(Farlow, 1984)

自組性演算法在演算過程上，類似倒傳遞類神經網路，兩者皆透過與目標實際值的誤差比較，訓練出較佳的模式。但最大的不同點在於，倒傳遞類神經網路之隱藏層數目必須一開始就確定，而自組性演算法可以在訓練過程中自動找出最佳的層數，自行衍生出一高階非線性系統，可將其歸類為自組性網路架構(Self-organization)。其模式建立之概念類似動植物演化過程，依據「適者生存」之原則來決定新方程式之取捨，透過競爭後留下的方程式將比原始方程式更適合描述此系統，這些優勝的方程式再次競爭則又可以得出具有更佳解釋能力之方程式。在演化的過程中，對系統影響不顯著的輸入變數會被自行淘汰，以此方式層層向下推演，直到滿足演算法終止條件或是下一代方程式並無明顯優於上一代為止，建構出一個可表示輸入與輸出之關係的高階多項式，圖 2.7 為自組性演算法之網路架

構圖，並說明其演算步驟。

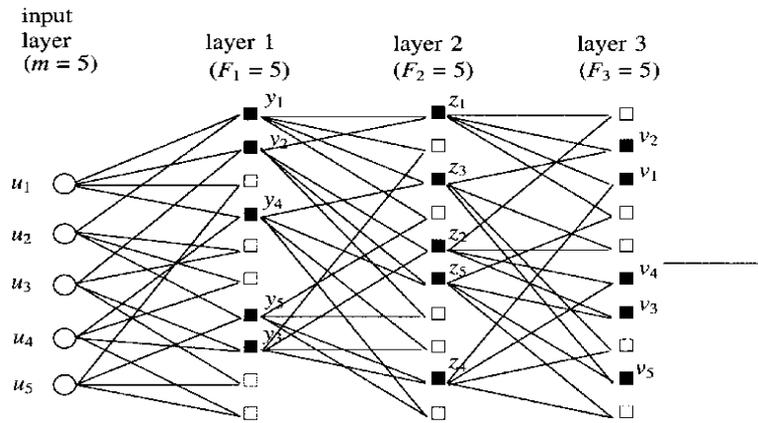


圖 2.7 自組性演算法之網路架構圖[24]

步驟一：假設建模資料共有  $n+m$  筆，分為訓練資料(M)及內部驗證資料(n)兩部份，目標值為  $y$ ，變數為  $x_1, x_2, x_3, \dots, x_p$ ，共有  $p$  個變數，作為 GMDH 之輸入層，如圖 2.8 所示。

Y		X <sub>1</sub>		X <sub>2</sub>		.....		X <sub>p</sub>
Y <sub>m1</sub>		X <sub>m11</sub>	X <sub>m12</sub>	1896				X <sub>m1p</sub>
Y <sub>m2</sub>		X <sub>m21</sub>	X <sub>m22</sub>					X <sub>m2p</sub>
Y <sub>m3</sub>		X <sub>m31</sub>	X <sub>m32</sub>					X <sub>m3p</sub>
.		.	.					.
.		.	.					.
Y <sub>nm</sub>		X <sub>nm1</sub>	X <sub>nm2</sub>					X <sub>nm p</sub>
Y <sub>n1</sub>		X <sub>n11</sub>	X <sub>n12</sub>					X <sub>n1p</sub>
Y <sub>n2</sub>		X <sub>n12</sub>	X <sub>n12</sub>					X <sub>n1 p</sub>
.		.	.					.
.		.	.					.
Y <sub>nn</sub>		X <sub>n1n</sub>	X <sub>n12</sub>					X <sub>n1p</sub>

圖 2.8 GMDH 演算法之輸入層[4]

步驟二：建構各層間輸入與輸出之關係。將  $p$  個變數兩兩組合，假設所選取的變數為  $u, v$ ，再代入內建方程式：

$$f(u, v) = Y = A + Bu + Cv + Du^2 + Ev^2 + Fuv$$

以最小平方法(Ordinary Least Squares, OLS)求取  $A, B, C, D, E, F$  之值，因此在每一層中將會延伸出  $C_2^p$  個變數，如圖 2.9 所示

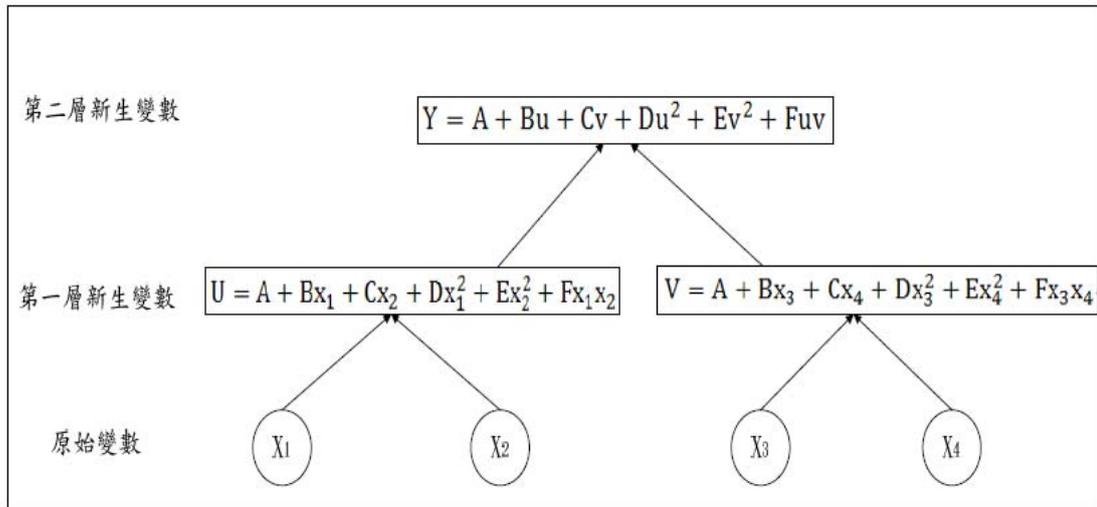


圖 2. 9GMDH 演算法各層變數組合過程

步驟三：求出  $u$ 、 $v$  等新生變數的方程式後，再以內部驗證資料作驗證，並以均方根(Root Mean Square, RMS)或誤差均方根(Root Mean Square Error, RMSE)等指標來評估總誤差大小，並依照誤差由小到大對各新生變數進行排序。均方根與誤差均方根之說明如下：

(1) 均方根

$r_i$  為第  $n$  層中第  $i$  條方程式之均方根值， $y(t)$  為第  $t$  個時間點之實際觀測值， $Z_i^{(n)}(t)$  為第  $n$  層第  $i$  條方程式於第  $t$  個時點之預測值，其中  $r_i$  公式如下：

$$r_i = \sqrt{\frac{\sum_{t=1}^T (y(t) - Z_i^{(n)}(t))^2}{\sum_{t=1}^T (y(t))^2}}$$

(2) 誤差均方根

$\varepsilon_i$  為第  $n$  層中第  $i$  條方程式之均方根值，其公式如下：

$$\varepsilon_i = \sqrt{\frac{\sum_{t=1}^T (y(t) - Z_i^{(n)}(t))^2}{n}}$$

步驟四：假設每層之最小誤差為 $R_{\min}$ ，若某一層之 $R_{\min}$ 小於上一層之 $R_{\min}$ 則回到步驟二。

步驟五：若某一層之 $R_{\min}$ 大於上一層之 $R_{\min}$ ，即以上一層之結果作為模式之最終輸入。

步驟六：依據模式輸出結果層層往上回推，即可找出模式選定的影響因子。

若最初輸入層之輸入變數為 $x_1$ 、 $x_2$ 、 $x_3 \cdots x_p$ ，則第一層所產生之新生變數標示為 $Z_1^1 = f(x_1, x_2)$ ， $Z_2^1 = f(x_1, x_3)$ ， $\dots$ ， $Z_{r(1)}^1 = f(x_{p-1}, x_p)$ ，依此類推第 $k+1$ 層新變數為 $Z_1^{k+1} = f(Z_1^k, Z_2^k)$ ， $Z_2^{k+1} = f(Z_1^k, Z_3^k)$ ， $\dots$ ， $Z_{p(k+1)}^{k+1} = f(Z_{p(k)-1}^k, Z_p^k)$ ，其中 $p(k)$ 表示第 $k$ 層中變數的數量。

一般的非線性系統來說，當模式發展到第三層獲第四層(方程式最高階為8階次或16階次)時，即可以得到一個最佳模式[1]，否則常會發生模式對於訓練資料的趨勢模擬相當接近，但對於訓練資料以外的驗證資料會產生較大的誤差。

自組性演算法發展至今，被各種領域的學者廣泛運用，當作預測分析的工具[11]。Dipti[19]進行需求電量的預測時，利用移動平均法、時間數列分析法、倒傳遞類神經網路以及自組性演算法進行預測，其研究結果指出自組性演算法在這些方法中表現最好。謝易霖[17]針對洪水暴雨的現象，對河水水位的變化使用自組性演算法結合馬斯金更公式進行預測，成功地使誤差縮小至30公分以內；余靜芳[3]運用自組性演算法結合基因演算法，發展出一套最佳化單品質及多品質特性混合實驗之演算法，以求出一組可使產品品質特性最大化的成分比例。林敬凱[5]對於半導體公司之蝕刻製程的資料，利用自組性演算法結合資料包絡法，提出一套多品質特性同時最佳化演算法。游翔百[14]先使用分類迴歸樹(Classification and Regression

Trees)進行變數篩選，再分別使用判別分析、羅吉斯迴歸、機率類神經網路、倒傳遞類神經、一般迴歸神經網路(General Regression Neural Network, GRNN)、自組性演算法、k-means及學習向量量化網路(Learning Vector Quantization, LVQ)等八種方法進行分類，發現複合式模型之準確率較單一方法好，另一個研究結果為：在違約資料之判別率上自組性演算法又比其他單一方法來得高。

自組性演算法不需繁複的計算，且其使用限制很少，成為各領域相關研究的熱門分析工具。

## 2.6 遺傳規劃法

遺傳規劃法[9]是由美國史丹福大學 Koza[22]教授於 1992 年提出，主要目的是希望在不提供詳細程式指令的情況下，電腦能透過自發性學習去解決問題。遺傳規劃法的基本概念是沿襲遺傳演算法(或稱基因演算法)而來，因此在演算過程中也具有染色體、適合度函數、複製交配與突變等運作機制，但有別於遺傳演算法利用二進位 0 與 1 的序列來表示染色體排序方式，遺傳規劃法則將染色體排列方式利用語法樹(tree)來表示，如圖 2.10 所示。每一個語法樹表示一組方程式，以語法樹對應一個將進行演化的個體，在演化過程中會不斷改變樹狀結構(染色體的排列方式)，直到產生最佳的語法樹，即為最佳的規劃方程式。

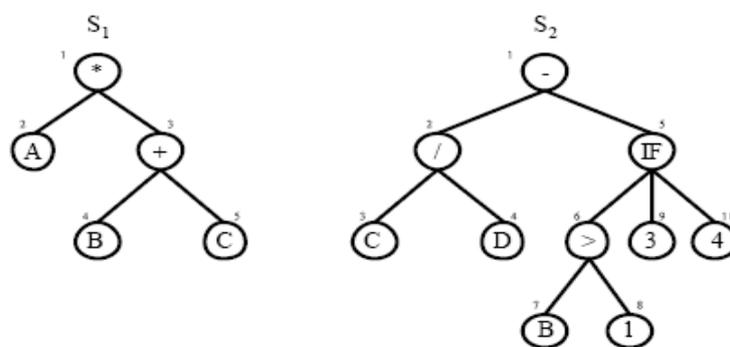


圖 2.10 語法樹示意圖[9]

遺傳規劃法結合機器學習(machine learning)與啟發式樹狀結構(evolved tree structure)的演算法則，特色是可透過自發性學習，設計一個存在有最佳解的演算法，因此使用者不需設計複雜的演算法，節省撰寫程式的時間，適用領域相當廣，應用彈性也相當大，特別是應用在不容易以數學或統計方法求解的問題。

遺傳規劃法與遺傳演算法在演化的機制上有許多相似的行為，但兩者在概念上之差異如表 2.3 所示：

表 2.3 GA 與 GP 之差異比較表[9]

概念差異	遺傳演算法	遺傳規劃法
染色體表示法	二進位字串	樹狀結構，並對應方程式
模式之結構與最佳參數	固定結構下尋找最佳參數	同時尋找模式之最佳結構與最佳參數
解的呈現方式	編碼方式呈現	數學方程式

遺傳規劃法運算流程為以下四個步驟：

1. 針對問題定義的函數隨機地組成產生初始族群(Population)。
2. 評估每個族群的個體(Individual)對於問題的適合度(Fitness)，適合度大小表示個體與目標值的差異距離。
3. 將最好的個體複製，並隨機的選擇個體進行演化操作，包括了個體之交配(Crossover)與個體之突變(Mutation)，進而演生出下一個世代的族群。
4. 紀錄每一個世代中最好的個體，直到達到目標設定值為止，最好之個體為該問題之最佳解。

多年來國內外不少的學者應用遺傳規劃法於各種領域的研究，應用最多的是金融、經濟等商學領域。以下將介紹工程學相關之文獻。

蕭德勇[18]應用遺傳規劃法在核能電廠爐心設計，由於填換爐心的時間點必須考慮許多電廠的重要參數，且爐心更換時產生事故的機率始終存在，因此以遺傳規劃法分析爐心設計的各種參數，並利用電廠過去實際資

料，驗證以遺傳規劃法進行爐心設計的分析是可行的。

林嘉宏[6]在研究建材的表面紋理對於摩擦力的影響時，以遺傳規劃法建立數個表面紋理的參數對於摩擦力的預測模型，並將預測結果與線性的多元迴歸模型進行比較，結果顯示遺傳規劃法的預測準確性較高。

## 2.7 總結

此小節將上述所提及的文獻，重點整理如下：

1. 複合式或二階段預測模型，能有效提升預測準確性，較單一方法預測結果佳。
2. 自組性演算法不論在線性或非線性系統都有不錯的預測能力，廣泛使用於能源預測領域。
3. 進行預測前，若能先將資料進行分群後再建構預測模型，能夠有效提升模型準確性。



## 第三章 研究方法

各國科技與經濟發展至今，能源危機與環境保護的問題日趨嚴重，歐美、日本等先進國家積極開發再生能源，以減少其他能源的消耗。而太陽能取之不盡，且太陽能發電時造成的環境汙染甚小，成為新興再生能源中最受期待的能源。然而建造一套太陽能發電系統所費不貲，礙於技術的侷限，目前系統的轉換效率仍然偏低，發電所轉售的利潤無法平衡昂貴的設置成本，成為許多廠商在投資太陽能產業時裹足不前的主要原因。因此探討一套發電系統在實際運作後，其轉換效率會有多大，將是所有投資廠商有興趣的議題。

如第二章所敘述，過去國內外討論太陽能發電系統轉換效率的相關研究中，建構轉換效率的預測模型較為罕見。本研究便利用太陽能發電系統的內部規格，並改善預測不準確的原因，建構轉換效率之預測模型。以下說明此三階段預測模型的流程與步驟。

### 3.1 研究架構

本研究之主要目的為利用太陽能發電系統中各零件特性與規格，建構出準確性高之太陽能發電系統轉換效率預測模型。當資料太複雜時如果直接進行預測，其預測的準確度不會太高；倘若我們先對資料做前置處理再進行預測，將會大大改善其預測準確性。本篇是採用三階段的預測方式，第一階段先篩選出顯著變數，避免不顯著變數影響模型準確性；第二階段將資料進行分群，使集群內的資料同質性高，提高模型的預測準確性；第三階段將處理好的資料利用一套混合預測模型建構轉換效率的預測模型，透過此三階段流程的預測，希望能有效提升模型準確性，並將預測結果與未經過資料處理，直接進行預測的其他方法作比較，如逐步迴歸法分析法(Stepwise Regression)、遺傳規劃法以及自組性演算法結合遺傳規劃法等預

測模型，作為評估準確性之用。圖 3.1 為本研究之流程圖。



圖 3.1 建構三階段轉換效率預測模型之流程圖

### 3.2 第一階段-篩選顯著變數

#### 步驟一：蒐集整理太陽能發電系統的轉換效率資料

本研究以台灣某太陽能光電研究單位提供的「太陽能光電系統轉換效率」資料進行分析。根據文獻[9]，此 12 項系統變數分析上較有意義，於是搜集此 12 項系統變數資料，而各系統變數即為預測模型的解釋變數，以變數的屬性可分為類別型變數(含離散計數型與二元型)與連續型變數兩類，反應變數則為系統轉換效率，如表所示。

表 3.1 系統資料的解釋變數與反應變數

種類	解釋變數	反應變數
類別型	模組廠牌	系統轉換效率
	模組種類	
	換流器廠牌	
	換流器組數	
	有無變壓器	
	變壓器輸出規格	
連續型	模組單片面積	
	模組裝設片數	
	模組單片容量	
	系統設置總容量	
	換流器輸出容量	
	變壓器容量	

由於轉換效率之數值必須考慮其合理性，若轉換效率超出現有技術所能達到的水準，則判定此筆資料為異常。就目前太陽能產業技術而言，最

先進開發技術其模組的實際轉換效率僅略高於 20%[34]，因此將資料中轉換效率超過 20%的資料視為異常，予以刪除。刪除缺漏與異常的資料後，每一筆資料均記錄其系統裝置內容，接著進行後續分析流程。

### **步驟二：利用逐步迴歸法篩選出顯著變數**

在本研究中，並非所有的解釋變數與轉換效率存在顯著關係，因此必須挑選出與轉換效率有顯著關係的解釋變數，以利進行轉換效率的預測。由於逐步迴歸分析法在篩選顯著變數時，可以同時處理類別型變數與連續型變數，因此本研究選擇使用逐步迴歸分析法篩選出對轉換效率有影響之解釋變數。

## **3.3 第二階段-資料分群**

此階段是將轉換效率進行分群處理，使每群體內的個體同質性高，提高轉換效率預測模型的準確性。本階段所採用的分群法為集群分析法，因為集群分析法進行分群時，計算出個體間的距離，根據距離遠近作為分群的依據，為一套客觀的分類程序；且集群分析並沒有嚴格的統計假設必須遵守，便能將資料進行分群，達到群體內個體同質性高的目的。而集群分析法中又以「二階段集群分析法」使用最普遍。

### **步驟一：階層式集群法取得最佳分群數目**

執行 k-means 法進行分群前必須決定集群數目，所以先利用階層式集群法得出最適集群數。決定最適集群數的準則參考陳順宇的多變量分析 [12]：先找出集群過程中距離最遠的階段當成切割處，因為此階段群內同質性最高，群間異質性最高；接著計算切割處和樹狀圖的接點數目當作最適集群數目。

### **步驟二：k-means 法進行分群**

決定最適集群數之後，以 k-means 法進行資料分群，參數 k 則為階層

式集群法所決定的最適集群數。

### 3.4 第三階段-建構轉換效率混合預測模型

建構轉換效率之混合預測模型之流程說明如下：

#### 步驟一：利用自組性演算法對各群觀測值進行預測

自組性演算法的優點在於：1.不受限於樣本筆數多寡；2.資料若為複雜的非線性資料也有不錯的預測能力。本研究鑒於自組性演算法有良好的預測能力，針對各群觀測值利用自組性演算法來分別進行預測，得出各系統之轉換效率預測值。

本研究執行自組性演算法所使用的軟體為「NeuroShell2.0 軟體」，可計算出類神經網路法演算架構，此軟體為套裝形式，不需要自行編寫繁複的程式語言，只需設定少量參數，軟體便會依照自組性演算法之運算機制自動進行。

#### 步驟二：利用遺傳規劃法調整殘差項

利用自組性演算法求得轉換效率預測值後，接著調整殘差項提升模型準確性。建構殘差預測模型則是利用遺傳規劃法，得出每筆資料的殘差預測值，作為殘差調整項。

#### 步驟三：得出最終轉換效率預測值

將自組性演算法預測的轉換效率預測值與遺傳規劃法得到的殘差預測值進行疊加，得出最終轉換效率預測值。

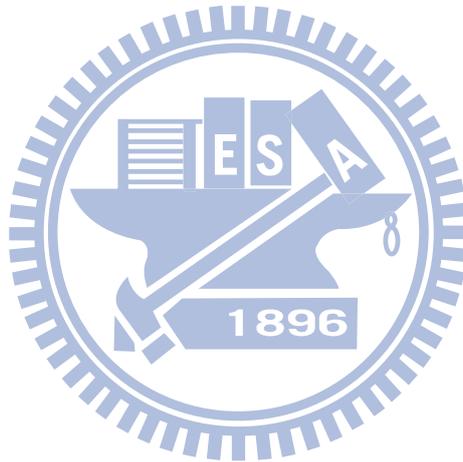
為了評估此套三階段轉換效率預測模型之準確性，將預測的結果利用準確性績效指標進行評估。本研究所採用的績效指標為平均絕對差(mean absolute deviation, MAD)以及平均絕對誤差百分比(Mean Absolute Percentage Error, MAPE)，利用這兩個指標來判別預測模型之結果優劣，Lewis[26]於 1982 年提出 MAPE 為預測模型準確性的評估指標中最好的，若 MAPE 小於 10%，代表此預測模型擁有高準確性。公式如下：

$$\text{MAD} = \frac{\sum |\text{實際值}_t - \text{預測值}_t|}{n}$$

$$\text{MAPE} = \frac{\sum \frac{|\text{實際值}_t - \text{預測值}_t|}{\text{實際值}_t} \times 100}{n}$$

其中  $n$  代表資料筆數。

此兩個指標通常用來評估預測模型的準確性，屬於望小的特性，指標愈小表示誤差愈小，預測模型越準確。接著本研究將利用實際案例，驗證此套三階段預測模型建構預測模型時，較單一階段預測方法來的準確。



## 第四章 實例驗證

本章利用實際案例來驗證本篇所提出之轉換效率三階段預測模型為可行且有效，並說明整個預測模型的流程。4.1 節為篩選顯著變數：敘述資料蒐集與整理的結果，並篩選出顯著變數；4.2 節將資料進行分群：將系統轉換效率進行分群以提高預測的準確性；4.3 節為建構轉換效率混合預測模型：建構自組性演算法結合遺傳規劃法之混合預測模型，並將結果與其他預測方法進行比較，評比預測模型的準確性；4.4 節介紹如何應用本研究所提出的三階段混合預測模型。

### 4.1 第一階段-篩選顯著變數

#### 4.1.1 資料蒐集與整理

此節說明太陽能發電系統之系統轉換效率資料，進行搜集與整理的過程和結果。

本研究以台灣某太陽能光電研究單位所提供 2006 年之太陽能發電系統資料進行資料分析，共計 49 筆。由於目前實際運作的太陽能發電系統，轉換效率也僅略高於 20% [34]，考慮轉換效率的合理性，將收集到的 49 筆資料中，刪除 5 筆轉換效率大於 20% 的觀測值，最後可進行分析的資料共有 44 筆。

在這 5 筆異常值中，進一步觀察其解釋變數是否異常，其中 1 筆資料的單片模組尺寸為 166mm×100mm，此尺寸明顯小於其他觀測值的單片模組尺寸，可能是進行量測時發生判斷錯誤，導致系統轉換效率出現異常。目前多數系統之發電量資料仍依賴人工方式抄寫，出現錯誤的機率比電子記錄來的高。其他 4 筆轉換效率異常的資料，其解釋變數並無太多差異，判定可能是人為疏失所造成。表 4.1 列出太陽能發電系統轉換效率資料整

理後的結果。

表 4.1 太陽能發電系統原始資料整理

所有資料總計	轉換效率異常(>20%)	有效資料
49	5	44

刪除缺漏的資料與異常的資料後，進行後續預測流程。

#### 4.1.2 利用逐步迴歸法篩選出顯著變數

由於資料中的解釋變數多達 12 個，因此考慮先行篩選對於轉換效率有顯著的解釋變數，減少不相關的解釋變數所造成的預測誤差，資料的解釋變數分為類別型變數與連續型變數，而逐步迴歸法能同時處理這兩種類型的變數，篩選出對轉換效率有顯著影響的解釋變數。表 4.2 為逐步迴歸法篩選變數結果之統計報表，此 12 個變數中「模組種類」、「模組單片面積」以及「模組單片容量」，而「有無變壓器」的 p-value 為 0.050336，略大於顯著水準 0.05，根據文獻[9]此變數也會影響顯著變數，因此本研究將「有無變壓器」納入顯著變數，其餘的變數由於皆未達顯著水準，故均予以刪除。

表 4.2 逐步迴歸分析結果

Regression Summary for Dependent Variable: 平均轉註						
R= .87301908 R2= .82839621 Adjusted R2= .8190979						
F(12,31)=2.6716 p<.01373 Std. Error of estimate: 2.296						
N=44	Beta	Std.Err. of Beta	B	Std.Err. of B	t(31)	p-level
Intercept			12.0929	4.936335	2.44977	0.020142
模組廠牌	-0.19743	0.167318	-0.1541	0.130579	-1.17994	0.246999
模組種類	-0.52176	0.242975	-2.9816	1.388481	-2.14738	0.039694
模組單片面積	-1.46861	0.631909	-16.0750	6.916674	-2.32409	0.026847
模組裝設片數	-0.08740	0.972145	-0.0036	0.040075	-0.08991	0.928940
模組單片容量	2.19721	0.719790	0.1469	0.048121	3.05256	0.004627
系統設置總容量	-0.23637	1.111742	-0.0719	0.338199	-0.21262	0.833019
inverter組數	0.26576	0.367162	0.9361	1.293357	0.72381	0.474607
inverter廠牌	0.01360	0.191885	0.0119	0.167975	0.07089	0.943942
inverter輸出容量	0.12949	0.473989	0.0489	0.178841	0.27318	0.786525
有無變壓器	-0.51487	0.252840	-2.9422	1.444855	-2.03634	0.050336
變壓器容量	-0.08481	0.525611	-0.0198	0.122482	-0.16136	0.872857
變壓器輸出規格	0.25207	0.256783	0.4867	0.495814	0.98166	0.333869

執行逐步迴歸分析後，必須檢查顯著變數之間是否存在共線性，表 4.3 得知顯著變數間未具有共線性問題。

表 4.3 解釋變數之共線性

Variable	Redundancy of Independent Variables; DV: 轉換效率 (Spreadsheet)			
	Toleran.	R-square	Partial Cor.	Semipart Cor.
模組廠牌	0.582460	0.223541	-0.177319	-0.151236
模組種類	0.378616	0.821384	-0.345770	-0.265242
模組單片面積	0.397125	0.940286	-0.360595	-0.285178
模組裝設片數	0.012780	0.983220	-0.143345	-0.029109
模組單片容量	0.306085	0.958252	0.569862	0.407121
系統設置總容量	0.027733	0.968394	-0.043710	-0.027440
inverter組數	0.128335	0.872425	0.191776	0.119493
inverter廠牌	0.396724	0.582436	0.028085	0.009512
inverter輸出容量	0.097725	0.919414	0.059038	0.037645
有無變壓器	0.312743	0.732246	-0.339229	-0.264887
變壓器容量	0.049302	0.969398	-0.027927	-0.022420
變壓器輸出規格	0.271804	0.729996	0.194232	0.133181

使用逐步迴歸法必須服從三大假設，圖 4.1 為逐步迴歸模式之常態機率圖，圖形上殘差並沒有嚴重偏離直線，因此判定常態性假設成立

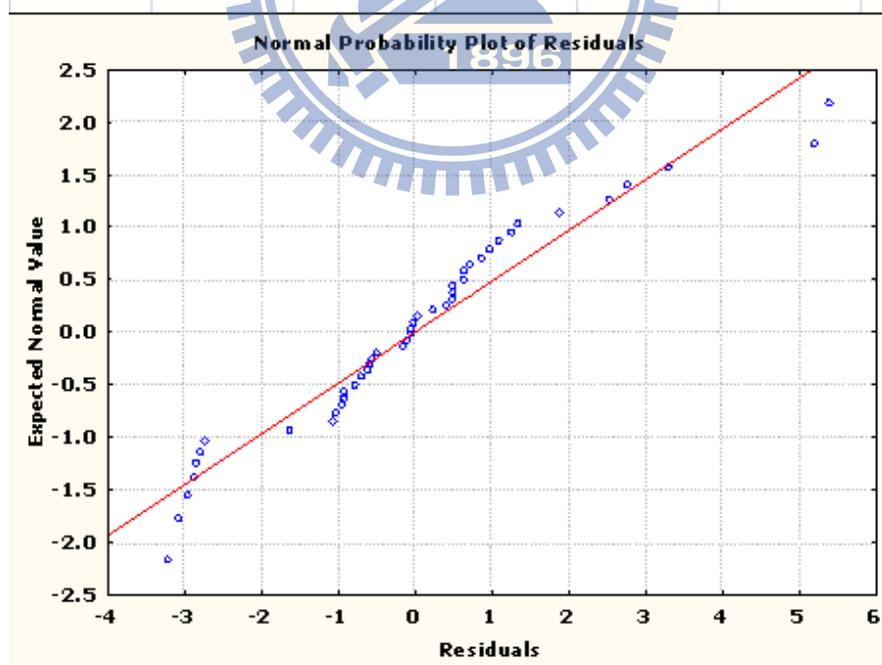


圖 4.1 常態性機率圖

圖 4.2 為殘差之散佈圖，觀察後判定圖型無明顯趨勢，因此獨立性和

同質性之假設皆成立。

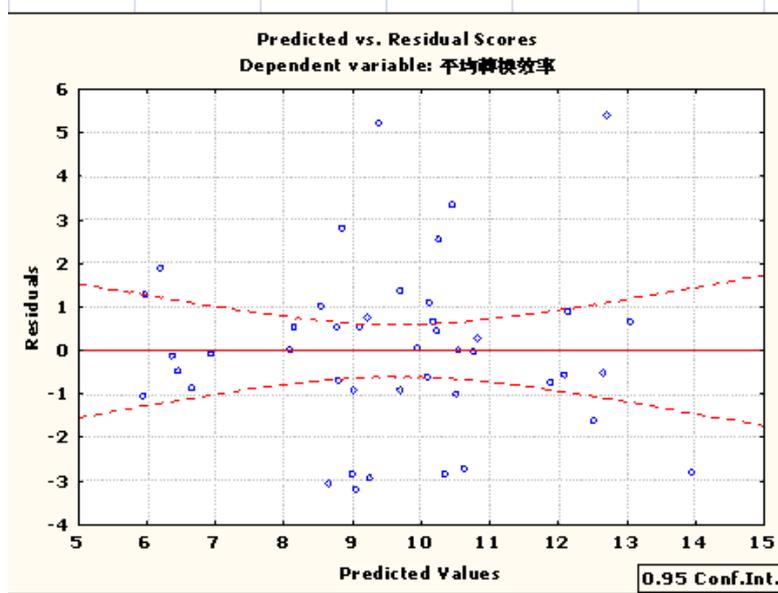


圖 4.2 殘差分析圖

篩選出顯著變數後，將未顯著的解釋變數予以刪除，避免產生預測誤差，表 4.4 為各顯著變數的重要性排序。

表 4.4 顯著變數的重要性

Variable	Summary of Stepwise Regression; DV: 轉換效率 (篩選變數原始資料)						
	Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variables included
模組單片容量	1	0.432327	0.186906	0.186906	9.654574	0.003382	1
模組種類	2	0.523687	0.274248	0.087342	4.934225	0.031912	2
inverter組數	3	0.563325	0.317335	0.043086	2.524598	0.119957	3
有無變壓器	4	0.587183	0.344783	0.027449	1.633803	0.208733	4
模組單片面積	5	0.648426	0.420456	0.075673	4.961794	0.031912	5
模組廠牌	6	0.682315	0.465553	0.045097	3.122071	0.085487	6
系統設置總容量	7	0.697883	0.487041	0.021488	1.508017	0.227411	7
變壓器輸出規格	8	0.711558	0.506314	0.019274	1.366413	0.250329	8

由表 4.4 得知，顯著變數中「模組單片容量」最為重要，其次為「模組種類」，接著為「有無變壓器」以及「模組單片面積」。而顯著變數的定義為：

模組種類：1：單晶矽，2：多晶矽

模組單片面積：0.6324~2.0625(平方公尺)

模組單片容量：75~210(瓦)

有無變壓器：1：無變壓器，2：有變壓器

## 4.2 第二階段-資料分群

### 4.2.1 利用階層式集群法決定最適分群數

使用 k-means 法分群前，必須先決定分群數目，使用階層式集群法來決定分群數是個客觀且合理的方法，將資料利用 SPSS 執行階層式集群法，連結方法設定為 Ward's 法。表 4.5 為集群過程的距離報表(節錄)，由報表得知在第 43 階段時，當資料 1 與資料 22 歸屬於同一群時，集群間的距離最大，所以此階段為分群的最佳切割線。再從圖 4.3 的分群樹狀圖中找尋資料 1 與資料 22 的集群處，便為分群切割線所在；計算此切割線與樹狀圖的接點數目，即為最適分群數的數目。在此實例中，從樹狀圖可以發現最適集群數目為 2 群。

表 4.5 集群過程的距離報表

29	1	12	.143	13	0	38
30	2	21	.145	0	0	36
31	25	44	.210	18	0	32
32	22	25	.224	26	31	39
33	13	19	.247	0	10	40
34	26	38	.365	27	28	39
35	3	5	.475	0	22	38
36	2	11	.826	30	0	40
37	4	8	1.282	0	0	41
38	1	3	1.891	29	35	41
39	22	26	4.381	32	34	43
40	2	13	4.616	36	33	42
41	1	4	5.021	38	37	42
42	1	2	7.385	41	40	43
43	1	22	12.550	42	39	0

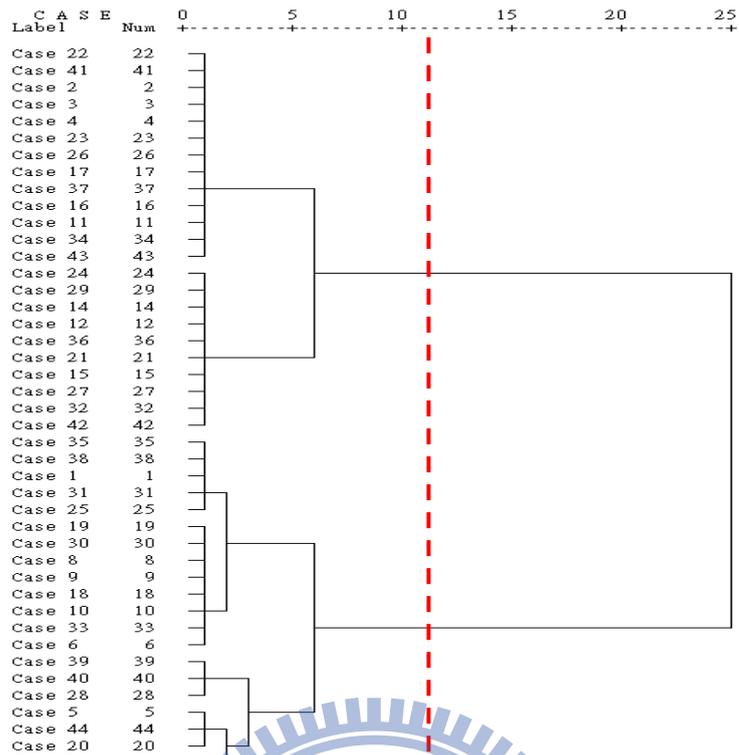


圖 4.3 分群樹狀圖

#### 4.2.2 使用 k-means 法進行分群

利用階層式集群法決定最適分群數目後，使用 k-means 法將資料分成 2 群，表 4.6 為分群結果。

表 4.6 k-means 法分群結果

Case No.	Members of and Distance Cluster con. Distance	Case No.	Members of and Distance Cluster con. Distance
C 2	2.29294	C 1	10.83512
C 3	2.29294	C 5	10.84610
C 4	2.29294	C 6	9.16856
C 11	4.22321	C 7	9.17608
C 12	0.35686	C 8	9.16888
C 14	4.79098	C 9	9.16888
C 15	2.73205	C 10	9.16884
C 16	2.72607	C 13	5.84757
C 17	2.72607	C 18	9.16888
C 21	2.30004	C 19	9.16888
C 22	2.29294	C 20	9.18153
C 23	4.78758	C 25	5.83636
C 24	4.79098	C 28	0.91794
C 26	4.78758	C 30	9.16888
C 27	2.73205	C 31	10.83512
C 29	4.79098	C 33	11.66845
C 32	7.72257	C 35	10.83512
C 34	4.22322	C 38	10.83512
C 36	0.35695	C 39	10.84006
C 37	2.72607	C 40	10.84006
C 41	2.29294	C 44	5.85672
C 42	12.72056		
C 43	5.28730		

分群結果為集群一有 21 筆資料，集群二有 23 筆資料；表 4.7 為 k-means 法的 ANOVA 分析結果，從此表得知此兩群的模組種類、模組單片面積、模組單片容量均有顯著差異，而有無變壓器在這兩群間沒有顯著差異。

表 4.7 集群分析之 ANOVA 結果

Variable	Analysis of Variance (分群原始資料)					
	Between SS	df	Within SS	df	F	signif. p
模組種類	6.37	1	3.810	42	70.2545	0.000000
模組單片面積	2.39	1	0.385	42	261.0713	0.000000
模組單片容量	65255.86	1	9136.318	42	299.9837	0.000000
有無變壓器	0.24	1	9.938	42	1.0309	0.315759

從表 4.8 可以看出此兩群的變數中心點位置，集群一的特色為：模組種類為單晶矽、模組單片面積較小、模組單片容量較小；集群二的特色為：模組種類為多晶矽、模組單片面積較大、模組單片容量較大；至於有無變壓器則無特別差異。

表 4.8 各變數中心點

Variable	Cluster No. 1	Cluster No. 2
模組種類	1.23809	2.0000
模組單片面積	10.80052	1.2672
模組單片容量	98.33334	175.4348
有無變壓器	0.71429	0.5652

### 4.3 建構轉換效率混合預測模型

本小節提出一套轉換效率混合預測模型：先利用自組性演算法預測轉換效率，得出轉換效率預測值與殘差；為了提高模型準確性，殘差部分使用遺傳規劃法進行調整，以下各小節說明使用混合預測模型之步驟。

#### 4.3.1 利用自組性演算法對各群觀測值進行預測

篩選顯著變數並將資料進行分群後，利用混合預測模型對轉換效率進行預測，首先使用自組性演算法對各群觀測值分別建構轉換效率預測模型，

利用軟體 NeuroShell2.0 軟體執行自組性演算法，各群執行自組性演算法時設定值如下：

- (1) 檔案定義(Field Delimiter)為 Comma。
- (2) 利用「80-20 法則」，將資料分為訓練樣本與測試樣本兩部份，集群一訓練樣本為 17 筆，測試資料為 4 筆；集群二訓練樣本為 18 筆，測試資料為 5 筆。
- (3) 變數尺度不進行轉換。
- (4) Model Non-linearity、Model Complexity、Model Diversity 三者皆設定成 high 值。
- (5) 其餘的參數則使用軟體內建的預設值。

透過 NeuroShell2.0 建構出系統轉換效率預測模型後，將預測結果整理成表 4.9。此兩集群的 GMDH 預測模型之 R 值都很大，解釋的變異量很高。

表 4.9 GMDH 預測結果

群	預測模型	R 值	
集群一	$Y = 0.22 * X_1 - 0.29 * X_2 - 0.9 * X_3 - 0.75 * X_3^2 + 1.9 * X_3^3 - 0.8 * X_1 * X_2 * X_3$	MSE	0.017331
		R squared	0.907023
集群二	$Y = -0.57 + 6 * X_2 - 2.2 * X_3 + 1.5 * X_4 - 7.4 * X_2^2 - 4.8 * X_3^2 - 2.7 * X_2^3 + 13 * X_3^3 - 4.9 * X_2 * X_3 - 9 * X_2 * X_4 - 1.1 * X_3 * X_4 + 8.8 * X_2 * X_3 * X_4$	MSE	0.019984
		R squared	0.892789

由於自組性演算法建構預測模型時，會將變數進行轉換來提高預測準確性，變數轉換內容如下：

集群一之預測模型變數轉換：

$$\begin{aligned}
 x_1 &= 2 * (\text{模組種類} - 1) - 1. \\
 x_2 &= 2 * (\text{模組單片面積} - 63) / 33 - 1. \\
 x_3 &= 2 * (\text{模組單片容量} - 75) / 45 - 1. \\
 x_4 &= 2 * (\text{有無變壓器} - 1) - 1. \\
 Y &= 2 * (\text{轉換效率} - 4.89) / 8.88 - 1.
 \end{aligned}$$

集群二之預測模型變數轉換：

$$x_1=2.*(模組種類-2.)-1.$$

$$x_2=2.*(模組單片面積-1.17)/.15-1.$$

$$x_3=2.*(模組單片容量-150.)/36.-1.$$

$$x_4=2.*(有無變壓器-1.)-1.$$

$$Y=2.*(轉換效率-5.58)/12.54-1.$$

至於 GMDH 的模型內部驗證部分，為了驗證 GMDH 測試模式在實際運作時具有再現性，表 4.10 與 4.11 為兩集群測試資料之驗證結果。

表 4.10 集群一內部驗證結果

單位	預測值	實際值	誤差百分比
3	4.88	4.89	7.28%
13	6.98	6.86	1.81%
14	9.18	8.65	6.13%
21	7.34	8.09	9.37%

表 4.11 集群二內部驗證結果

單位	預測值	實際值	誤差百分比
4	9.43	8.77	7.52%
11	5.88	5.58	5.41%
16	10.97	11.13	1.45%
24	12.13	12.81	5.31%
35	14.38	13.71	4.96%

由表 4.10 及 4.11 得知，此兩集群的測試樣本之誤差百分比皆小於 10%，表示這兩個預測模型準確性高，具有再現性。

#### 4.3.2 利用遺傳規劃法調整殘差

誠如 3.4 節所述，利用遺傳規劃法來調整轉換效率預測模型之殘差部分，提升預測模型的準確性。使用模擬軟體執行遺傳規劃法，其參數設定如下：

- (1) 解釋變數個數為 4 個
- (2) 解釋變數之間由於單位不同，故先將資料進行標準化。
- (3) 演化次數設為一極大值，確保演算結果達到收斂。
- (4) 繁殖比例設定為 0.8，選取 80% 的親代進行染色體之交配。
- (5) 突變比例設定為 0.2，選取 20% 之子代進行染色體突變。

模擬結果所得出殘差預測方程式為：

$$\text{集群一：}\hat{\epsilon} = 0.254 - \frac{x_1^2 x_2^2}{x_3}$$

$$\text{集群二：}\hat{\epsilon} = \frac{x_2^2 \times x_4^2}{x_4 - x_2} + 0.632$$

其中  $\hat{\epsilon}$  表示轉換效率之殘差估計值。

### 4.3.3 求出最終轉換效率預測值

將自組性演算法預測的轉換效率預測值與遺傳規劃法得到的殘差預測值進行疊加，得出最終轉換效率預測值，圖 4.2 為實際值與預測值的折線圖。

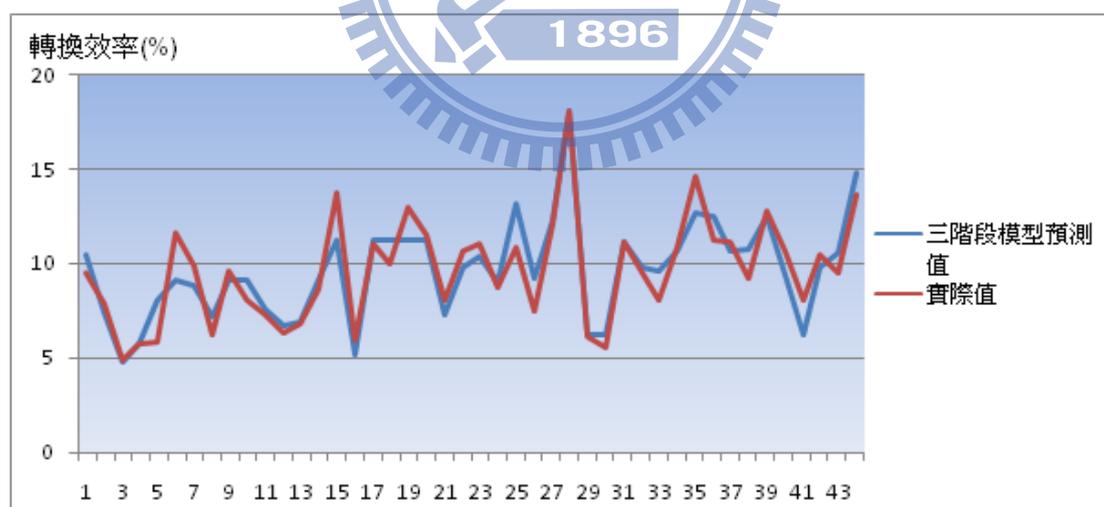


圖 4.4 實際值與預測值比較折線圖

接著探討預測模型準確性，將此混合預測模型的預測結果，計算出 MAD 與 MAPE，如表 4.12 所示。根據 Lewis[26]提出的 MAPE 評估準則，顯示此兩集群的預測模型皆屬於高準確性。

表 4.12 預測模型之準確性評估

	MAPE(%)
集群一之混合預測模型	10.0985%
集群二之混合預測模型	9.1672%

為了驗證本研究所提出的預測模型準確性優於其他預測方法，將資料利用逐步迴歸法、遺傳規劃法以及自組性演算法結合遺傳規劃法進行預測，此三種預測模型皆使用單一階段預測方法建構模型，以比較本研究的三階段預測模型較單一階段來的準確。

參數設定的部分，逐步迴歸分析法之參數設定為軟體預設值，遺傳規劃法參數設定與前一小節相同，至於混合預測模型所有參數與本研究第三階段的混合預測模型相同，利用此三種預測模型對轉換效率進行預測，收集預測結果並整理之。

表 4.13 單一階段預測方法結果

預測方法	預測模型
逐步迴歸法	$Y = -0.52176 \text{ 模組種類} + 1.46861 \text{ 模組單片面積} + 2.19721 \text{ 模組單片容量}$
遺傳規劃法	$Y = 0.7258 - 0.742(\text{模組單片面積} - \text{模組單片容量}) - 0.347(\text{模組種類} + 1.374)X(\text{模組單片面積} - \text{模組單片容量} + 0.674)$
自主性演算法+遺傳規劃法	$Y = -0.31 * \text{換流器廠牌} - 0.11 * \text{系統設置總容量} + 7.8E-002 * \text{變壓器輸出規格} + 0.31 * \text{模組廠牌} - 2 + 21 * \text{模組單片面積} - 20 * \text{模組單片容量} - 17 * \text{模組單片面積}^2 - 17 * \text{模組單片容量}^2 + 6.4 * \text{換流器組數}^2 - 6.2 * \text{模組單片面積}^3 + 4.8 * \text{模組單片容量}^3 + 4.9 * \text{換流器組數}^3 + 36 * \text{模組單片面積} * \text{模組單片容量} + 10 * \text{模組單片面積} * \text{換流器組數} - 10 * \text{模組單片容量} * \text{換流器組數} + 2.1 * \text{模組單片面積} * \text{模組單片容量} * \text{換流器組數} + 0.39 * \text{變壓器容量}^2 + 0.38 * \text{換流器廠牌}^3 - 0.19 * \text{模組廠牌}^2 - 0.51 * \text{模組廠牌}^3$ $e = 2 * \text{模組裝設片數} / (2 * \text{模組廠牌} - \text{模組裝設片數} + 2 * \text{模組單片容量} + \text{系統設置總容量} + \text{模組廠牌})$

為了評估模型準確性，利用 MAD 與 MAPE 來評估預測模型的準確性，將三階段預測模型與其他三個單一預測模型的預測結果求算出 MAD 與 MAPE，整理如表 4.14 所示。

表 4.14 預測模型之準確性評估

預測模型	MAD	MAPE(%)
逐步迴歸分析	1.714682	18.4378
GP(高翊倫)	1.368373	15.4534
GMDH+GP	1.265656	14.8766
三階段預測模型 (集群一)	0.8951	10.0985
三階段預測模型 (集群二)	0.8924	9.1672

由表 4.14 得知，在單一階段預測方法中，遺傳規劃法的預測準確性優於傳統統計的逐步迴歸法，可能的原因是：系統轉換效率的資料中，系統變數與轉換效率之間的關係，可能存在非線性或複雜的交互作用，逐步迴歸法屬於線性預測方法，因此所建立的預測模型準確性不如遺傳規劃法來的好；而混合預測模型(自組性演算法+遺傳規劃法)又比遺傳規劃法準確性高，因混合預測模型進行殘差調整，可以提高模型準確性；而本研究所提出的三階段預測模型，集群一與集群二預測結果的 MAD 與 MAPE 皆為這些方法中最低的，且 MAPE 皆小於 10%，因此可證實此套三階段預測模型準確性最高，先篩選顯著變數後將資料分群後，再利用混合預測模型進行預測，結合這些分析工具的優點有效提高模型準確性。

#### 4.4 三階段預測模型之應用

上一節已證實使用此套三階段預測模型的預測能力相當不錯，此節將

說明如何使用此套預測模型來預測太陽能系統之轉換效率。

本研究所提出的三階段預測模型，是將資料依太陽能系統之特色，分為兩個集群來預測轉換效率，集群一資料的特色為：模組單片面積較小且模組單片容量較小；集群二資料的特色為：模組單片面積較大且模組單片容量較大。在應用本研究方法預測某太陽能系統之轉換效率前，必須先知道此筆資料屬於哪個集群，才能利用該群的混合預測模型預測轉換效率，因此本研究必須先建立一個判定模式。在判別資料之工具中，當解釋變數有連續型與類別型變數時，邏輯斯迴歸法之判別能力較佳，因此本研究將44筆建模資料利用邏輯斯迴歸法建立邏輯連接函數(logic link function)，用以判別某太陽能系統應屬何集群。

建立邏輯連接函數必須設定一個切割點，Sharma[28]指出，如果集群數目為兩群時，切割點應設定成0.5，則正確分類率最好，故本研究執行邏輯斯迴歸時，將切割點設定成0.5。表4.15為本研究應用SPSS執行邏輯斯迴歸法之結果。

表 4.15 邏輯斯迴歸之參數估計表

		B	S.E.	Wald	自由度	顯著性 <sup>a</sup>	Exp(B)
步驟 1 <sup>a</sup>	模組種類(1)	8.756	24.954	3.783	1	.000	6348.534
	模組單片面積	101.144	43.425	8.038	1	.000	8.435E+43
	模組單片容量	.090	4.995	1.585	1	.000	1.094
	有無變壓器(1)	7.569	9.318	3.171	1	.056	1936.787
	常數	-131.824	83.642	1.086	1	.008	.000

由上表可得到邏輯連接函數為：

$$Z = -131.824 - 8.756 * \text{模組種類} + 101.144 * \text{模組單片面積} + 0.09 * \text{模組單片容量} - 7.569 * \text{有無變壓器} \quad (\text{公式 4-1})$$

為了驗證此邏輯斯迴歸分類結果的準確程度，於是進行分類正確率檢驗，

執行結果為表 4.16。

表 4.16 邏輯斯迴歸之分類正確率

觀察			預測		
			集群		百分比修正
			1.00	2.00	
步驟1	集群	1.00	21	0	100.0
		2.00	0	23	100.0
概要百分比					100.0

a. 分割值為 .500

當切割點為 0.5 時，此邏輯斯迴歸分類正確率為 100%，代表此分類結果相當好。將資料帶入邏輯連接函數，得出此資料的 Z 值。而分類機率值  $p$  須由 Z 值推算，公式如下：

$$p = \frac{1}{1 + e^{-z}}$$

分類機率值  $p$  若小於 0.5 則將太陽能系統資料歸屬於集群一，若大於 0.5 則歸屬於集群二。圖 4.5 為應用三階段轉換效率預測模型之流程圖，本研究以一個例子說明如何應用此套三階段轉換效率之預測模型。

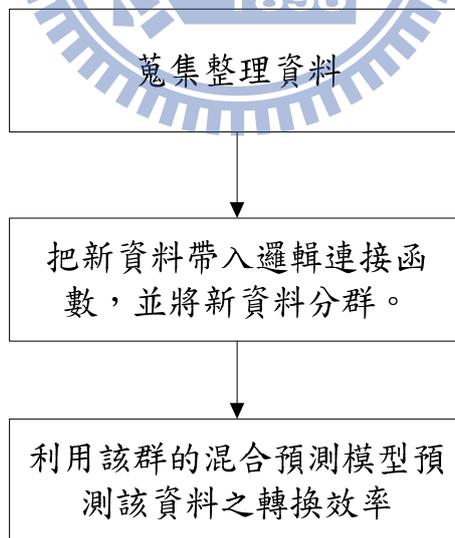


圖 4.5 應用三階段轉換效率預測模型之流程圖

若某廠商欲預測某太陽能系統之轉換效率，該系統資料包含了很多系統變數。將這些資料帶入本研究之三階段轉換效率預測模型，所使用之預

測變數值如表 4.17 所示。

表 4.17 新資料內容

模組種類	模組單片面積	模組單片容量	有無變壓器
1	0.9291	120	2

將表 4.17 之資料代入邏輯連接函數(公式 4-1)求算出分類機率值約為 0.0001，小於 0.5，因此判定此筆太陽能系統資料歸屬於集群一，為模組單片面積小且模組單片容量小的資料，於是利用集群一的混合預測模型來預測轉換效率。將此筆資料代入集群一的轉換效率方程式，得出此筆資料的  $\hat{Y}=11.3035$ ；接著代入集群一的殘差預測方程式，求出此筆殘差預測值為  $\hat{\epsilon}=0.2468$ ；將這兩個值進行疊加得出最終轉換效率預測值為 11.5503。當有太陽能系統之新資料須預測轉換效率時，只需依照此流程便能得出轉換效率預測值，以作為工程人員進行系統改良的參考。



## 第五章 結論與建議

在環保意識抬頭的今日，太陽能成為備受矚目的再生能源之一，利用太陽能系統發電來解決電力供應不足的問題不失為一個好作法。然而欲發展太陽能產業的資方最關切的問題，便是太陽能發電系統能有多大的轉換效率，使廠商能從轉售電能中平衡系統設置成本，從中賺取利潤。因此本研究以太陽能發電系統之轉換效率資料進行分析，提出了一套整合三項系統的三階段預測模型，此預測模型對系統轉換效率擁有良好的預測能力，提供太陽能相關產業進行後續研究的參考，協助工程人員能更精確地了解太陽能發電系統的運作情況，且了解系統各部分組件規格與實際運作之間的差異。5.1 節說明本研究的貢獻，5.2 節說明後續研究。

### 5.1 研究貢獻

本研究之貢獻有以下三點：

1. 過去一些太陽能轉換效率相關研究中，較少使用資料分析的手法。本研究嘗試利用實際運作的資料，藉由已知的零組件特性規格去預測轉換效率，此預測結果可作為工程師改良系統時的參考依據。
2. 本研究提出的三階段預測模型，為一套結合數種分析工具的預測方法，目的為提高傳統單一預測模型的準確性，並透過實際案例的說明，驗證此三階段預測模型確實可行且有效。
3. 本研究提出的三階段預測模型，在建構預測模型的階段，使用了自組性演算法結合遺傳規劃法的混合預測模型進行預測，這兩種方法不需遵守統計假設，且在非線性資料預測能力也很好，可避免資料不符合假設所造成的預測誤差，得到更貼近真實情況的轉換效率預測值。

## 5.2 後續研究

1. 本研究以資料分析的方式來建構系統轉換效率的預測模型，在太陽能系統相關研究中算是新的嘗試，不過礙於資料上的侷限，沒辦法取得其他系統參數資料或是氣候資料，因此無法對這些因素深入探討。日後研究如能取得這些資料，便能以類似的手法建構出更精確的預測模型，使結果能更貼近實際運作的情況。
2. 本研究在建構預測模型的部分使用的方法為自組性演算法以及遺傳規劃法，但仍有其他演算法在建構預測模型時有相當不錯的預測能力，如灰預測模型(Grey Forecasting Model)、支援向量機(Support Vector Machine, SVM)等，因此後續研究可利用其他方法進行預測，發展出更準確的預測模型。



## 參考文獻

- [1] 王英銘(2005),「自組非線性系統應用於濁度預測」,國立成功大學水利及海洋工程研究所,碩士論文。
- [2] 邵承矩(1993),「能源應用」,東華書局出版。
- [3] 余靜芳(2008),「利用自組性演算法與基因演算法於混合實驗最佳化之研究」,國立交通大學工業工程與管理學系,碩士論文。
- [4] 林愷為(2008),「應用資料分析技術於國家公園棲地環境評估」,國立中山大學海洋環境工程與管理學系,碩士論文。
- [5] 林敬凱(2008),「具多反應實驗設計最佳化演算法之研究」,國立交通大學工業工程與管理學系,碩士論文。
- [6] 林嘉宏(2007),「鋪面紋理與摩擦力關聯性之初步研究」,逢甲大學交通工程研究所,碩士論文。
- [7] 胡建勳、陳宗亮(2008),「應用資料群聚與關聯法則於網路拍賣使用者之特性與偏好研究」,玄奘大學資訊科學研究所,知識社群與系統發展研討會。
- [8] 張子文(2000),「太陽電池應用於建築上之研究」,國立成功大學建築研究所碩士論文。
- [9] 高翊倫(2009),「建構台灣地區太陽能發電系統之發電量預測模型」,國立交通大學工業工程研究所,碩士論文。
- [10] 陳同孝、陳雨霖、劉明山、許文綬、林志強、邱永興(2006),「結合 k-means 及階層式分群髮之二階段分群演算法」,國立台中技術學院資訊科技與應用研究所。
- [11] 陳英豪(2005),「應用自組性演算法建構企業信用評等模型」,國立交通大學工業工程與管理學系,碩士論文。

- [12] 陳順宇(2005),「多變量分析」四版,華泰書局出版。
- [13] 黃美玲、施啟翔(2009),「以模糊分群法結合邏輯斯迴歸於醫療輔助診斷」,國立勤益科技大學工業工程與管理系。
- [14] 游翔百(2004),「建構複合式信用評等模型」,交通大學工業工程研究所,碩士論文。
- [15] 馮焯生(2008),「太陽電發電原理與應用」,五南圖書出版。
- [16] 經濟部能源委員會(2007),「中華民國九十六年台灣能源統計年報」,經濟部能源委員會出版。
- [17] 謝易霖(1993),「暴雨時期 GMDH 模式結合距離—水位模式應用於未設站河段即時水位之預測」,國立成功大學水利及海洋工程研究所,碩士論文。
- [18] 蕭德勇(2004),「以遺傳規劃技術應用於核能電廠之爐心燃料設計研究」,元智大學資訊管理所,碩士論文。
- [19] Dipti Srinivasan(2008), “Energy demand prediction using GMDH networks,” *Neurocomputing*, Vol72, pp.625-629.
- [20] Gregg, A., Perker, T. and Swenson, R.(2005), “A real world examination of PV system design and performances,” *Photovoltaic Specialists Conference*, IEEE, Vol.31
- [21] Ivakhnenko, A. G. (1971), “Polynomial Theory of Complex Systems,” *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 1, No. 4, pp. 364-378
- [22] Koza, J. R.(1992), Genetic Programming, On the Programming of Computer by means of Natural Selection, *MIT Press*, Cambridge, MA.
- [23] Kuan-Yu Chen, Cheng-Hua Wang(2007), “A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan,” *Expert System with Applications*, Vol(32), pp.254-264.

- [24] Muller, J. A., Ivachnenko, A. G. and Lemke, F. (1998), “GMDH Algorithms for Complex Systems Modeling,” *Mathematical and Computer Modeling of Dynamical Systems*, Vol.4, No.4.
- [25] Nahi Kandel, Rene Wamkene, Maarouf Saad and Semaan Georges(2006)“ An efficient approach for short term load forecasting using artificial neural networks.”*Power and Energy System*,pp 525-530,Vol28
- [26] Lewis, C. D.(1982), *Industrial and Business Forecasting Methods*, Butterworth, London.
- [27] R. E Abdel-Aal(2006), “ Modeling and forecasting electric daily peak loads using abductive networks. ” *Electrical Power and Energy Systems*, pp 133-141, 2006.Vol28.
- [28] Subhash Sharma(1995),“Applied Multivariate Techniques.” *John Wiley & Sons Inc.*
- [29] Zhang G. Peter(2003), “Time Series forecasting using a hybrid model ARIMA and neural network model,” *Nerocomputing*, Vol.50,pp.159-175.

#### 網路資料

- [30] 台灣經濟部能源局：<http://www.moeaboe.gov.tw/>
- [31] 明道大學產業創新及經營學系：<http://tinyurl.com/2437dfc>
- [32] British Petroleum：  
<http://www.bp.com/bodycopyarticle.do?categoryId=1&contentId=7052055>
- [33] KUAS E.E. PEAL：  
[http://hpds.ee.kuas.edu.tw/index\\_link/intro/download\\_file/20060505.ppt](http://hpds.ee.kuas.edu.tw/index_link/intro/download_file/20060505.ppt)

[34] PR Newswire, “SunPower Announces the Planet’s Most Powerful Solar Panel,”

<http://news.prnewswire.com/ViewContent.aspx?ACCT=109&STORY=/www/story/05-20-2009/0005029765@EDATE>

