

# 整合獨立成份分析與分類迴歸樹 在製程干擾辨識上之應用

黃馨瑩・邱志洲\*

(收稿日期：95 年 10 月 13 日；第一次修正：96 年 3 月 9 日；  
第二次修正：96 年 6 月 22 日；接受刊登日期：96 年 6 月 22 日)

## 摘要

爲了在今日全球化的市場上競爭，製造業者不斷的嘗試應用許多的品質管制方法（如管制圖）來監測產品的生產流程以降低成本。但由於監測的產品品質特性，基本上是一種具有高度自我相關的觀測資料，因此常造成製程干擾現象被誤判的結果。爲了解決這樣的問題，在實際執行監測的過程中，品管人員常會採用一般化時間序列模式下的殘差控制圖來進行資料異動形態之辨別。然而，使用殘差控制圖的辨識成功率卻會隨著觀測資料的自我相關程度（當期資料與前期資料的相關係數高低）而變動；而且在多數情況下，資料相關程度的先前資訊往往無法事先獲得。由於相關性製程資料可被視爲一個混合了雜訊及製程干擾訊號的混合資料（mixture data），因此本研究導入近年來被成功應用在訊號分離（signal separation）之獨立成份分析（Independent Component Analysis, ICA）技術於製程管制中，利用 ICA 能將混合訊號分離出潛在來源訊號（latent source signals）之能力，提出一個結合獨立成份分析與分類迴歸樹（Classification and Regression Tree, CART）的製程監控架構。在所提的架構中，本研究先利用 EPC 與 ICA 將製程觀察值（即混合訊號）分離出相互獨立之獨立成份（independent component），之後再使用分類迴歸樹針對獨立成份進行監控，期望能解決傳統管制圖需要對相關性製程的資料進行繁複處理的問題。換言之，本研究的研究重點有二：（一）提出獨立成份分析（Independent component analysis, ICA）的轉換機制，使得原始資料可被分離出更容易被辨識的資訊。（二）嘗試應用分類迴歸樹進行干擾型態的辨識以解決應用傳統管制圖無法清楚解釋製程管制規則的缺點。本研究針對在 IMA(1,1) 下兩種不同類型的製程干擾資料進行製程管制的模擬研究測試，以驗證所提出整合方法之有效性。根據研究結果顯示，藉由 SPC/EPC/ICA 技術的應用，雜訊（noise）確實可以從相關性製程資料中被成功的分離出來。

關鍵詞彙：統計製程管制，工程製程控制，獨立成份分析，分類迴歸樹

## 壹· 導論

在全球市場競爭激烈的今日，持續提昇產品品質是企業獲得競爭優勢的重要方法。爲了要能完整控制生產產品的高品質效能，進行生產線上的監測作

---

\* 作者簡介：黃馨瑩，國立台北科技大學商業自動化與管理研究所研究生；邱志洲，國立台北科技大學商業自動化與管理研究所教授。

本研究部分承國科會（計劃編號：NSC 95-2221-E-027-072-MY3）經費贊助，特此感謝。

業是工業界常採用的策略。而所謂線上的監測作業是指運用各種技術或方法在不改變產品品質特性的前提下記錄所有資料，進而辨識產品在生產過程中是否有任何不良品的一種程序。在品質管制的研究領域裏，統計製程管制 (Statistical Process Control, SPC) 中的管制圖 (control charts) 及工程製程管制 (Engineering Process Control, EPC) 技術是常被應用來進行工程製程監測與控制的工具。其中，SPC 的主要目的在於能迅速地偵測出製程中發生異常現象的可歸屬原因 (assignable cause) 及時點，而所謂的 EPC 技術，則是一種透過回饋控制 (feedback control) 機制，針對製程輸出值來調整輸入變數進而縮小製程輸出值與目標值間差距的一種方法。只是，雖然結合 SPC 與 EPC 的方法可以有效的偵測出製程中發生異常的時點，但若製程數據具有某種程度的自我相關性質時，使用傳統的 SPC 與 EPC 方法對製程進行偵測往往會發生錯誤警訊而增加誤判的結果，因此如何發展出一套能針對相關性製程進行異常偵測與控制的線上系統，便成為當前品質管制領域中相當重要的研究課題之一。

獨立成份分析 (ICA) 是近年來發展極為迅速的一個訊號處理技術，它主要是用來處理未知來源訊號的分離 (Blind Source Separation, BSS) 問題 (Lee, 1998; Hyvärinen and Oja, 2000)，其目的是在沒有任何有關訊號混合機制的情形下，根據觀察到的混合訊號 (mixture signals) 尋找出潛在的來源訊號 (latent source signals) (Hyvärinen et al., 2001)。而基本上，相關性製程的觀測值可以被視為一個包含了雜訊 (noise)、製程干擾項及自我相關性的混合資料 (mixture data)，因此若單純使用 SPC 與 EPC 來進行製程管制，往往無法得到理想的結果。因此在本研究的第一階段中，先嘗試結合 SPC/EPC 與 ICA 的方法來進行資料的轉換，希望透過 ICA 的技術將個別獨立訊號由觀測資料中分離出來。而在本研究的第二階段裏，本研究則將根據分離出來的個別獨立訊號，以分類迴歸樹 (Classification and Regression Tree, CART) 的方法建構一個辨識製程干擾型態的線上模式。事實上截至目前為止，在學術上已有許多的分類辨識工具被成功的發展出來，它們包括了許多傳統統計方法、無母數方法以及人工智慧方法等。而在這眾多的分類方法中，分類迴歸樹是近年來較常被討論、應用的方法之一。所謂的分類迴歸樹 (CART) 是 Breiman et al. (1984) 所發展出來的一種演算法，其基本原理乃是使用二元分割過程來分析龐大的資料集，並透過遞迴 (recursive) 的程序，依照預測變數與相關劃分條件將  $N$  個訓練樣本加以劃分成  $M$  個已知類別，並產生一組劃分好的規則 (rule)，最後再透過此規則對新樣本進行預測的過程。在本研究中，為了解決現存既有方法 (如 SPC

與 EPC) 之缺點，並增加模式的辨識精確度，除了 ICA 方法的應用外，本研究也嘗試提出分類迴歸樹技術來進行製程管制模式的建立。

本研究共分為五部分：第一部份為緒論，說明本文的研究動機、目的及研究的整體架構；第二部分的重點則在回顧統計製程管制、工程製程控制、獨立成份分析與分類迴歸樹的相關文獻；第三部分則分別介紹工程製程控制、獨立成份分析與分類迴歸樹的演算原理與步驟；第四部分則描述結合 SPC/EPC 與 ICA 的資料訊號分離結果，並進一步說明透過分類迴歸樹方法針對相關性製程資料分析所獲得之辨識情形；最後，第五部分則為本研究的結論。

## 貳· 文獻探討

### 一、SPC/EPC於相關性製程之監控

在運用傳統管制圖進行製程資料的監測時，常會先假設觀測資料是來自於一個獨立且符合常態分配的母體才不會對偵測結果產生過大的誤差，因此當製程觀測值具有自我相關性時，傳統管制圖很容易因為低估製程變異使得管制界限比預期來得狹窄，進而造成假警訊的頻率增加。為了解決這樣的問題，目前已有許多方法被提出來 (Tseng & Adams, 1994 ; Faltin, et al., 1997)。而在這許多的方法中，時間序列法是最常被討論與使用的一種技術 (Adams & Tseng, 1998 ; Wieringa, 1999 ; Montgomery, 2001)。

Alwan and Roberts (1988) 曾將原始的製程資料以時間序列模式配適後，再整合一般原因管制圖 (Common-Cause Chart, CCC) 與特殊原因管制圖 (Special-Cause Chart, SCC) 進行實際的監控。Montgomery & Mastrangelo (1991) 則是以時間序列模型下的預測誤差為對象，提出以指數加權移動平均管制圖 (Exponentially Weighted Moving Average Control Chart, EWMA) 進行監控的想法；根據他們的研究結果顯示，當製程資料呈現 IMA(1,1)型式時，結合時間序列及管制圖的方法能有效的發揮其監控的功能。Runger et al. (1995) 則是利用馬可夫鏈方法建構一監控殘差之累和管制圖 (Cumulative Sum Control Chart, CUSUM)，並指出殘差 CUSUM 管制圖可降低傳統 CUSUM 管制圖會因自我相關性而影響管制狀態內平均串長度 (Average run length, ARL) 之缺點。除了以預測誤差為對象的管制方法外，Zhang (1998) 則是以原始資料的變異數為對象，先將其轉換成自我相關函數 (Autocorrelation Function, ACF) 的形式後，再以其所建構之 EWMAST 管制圖進行監控。Lu & Reynolds (1999) 則是針對

ARMA(1,1)製程資料發展出一新穎的殘差 EWMA 管制方法；根據他們的研究結果指出，當製程資料具有中度或低度的自我相關性時，以觀測值為基礎的 EWMA 管制方法與以殘差值為基礎的 EWMA 管制方法表現相當，但在高相關製程時，殘差 EWMA 管制圖在製程偏移時的偵測速度較快。Jiang et al. (2000) 則是根據 ARMA 統計量提出了 ARMA 管制圖的想法，該研究指出 SCC 管制圖與 EWMAS 管制圖均為 ARMA 管制圖的特例，並且 ARMA 管制圖的偵測效果較前述兩種管制圖為佳。Lu & Reynolds (2001) 曾針對觀測資料先以一時間序列模型予以配適後，再以其殘差值為基礎建構一雙邊 CUSUM 管制圖；根據其研究結果顯示，傳統 CUSUM 管制圖與以殘差值為基礎之 CUSUM 管制圖，在製程資料呈現中、低度相關性時的偵測能力十分相近，但殘差管制圖在處理高度相關性製程資料時的表現較佳。

除了上述結合時間序列模式與管制圖的方法外，結合 SPC 與工程製程管制 (Engineering Process Control, EPC) 的技術也曾被學者提出來，針對具有自我相關性製程的資料進行監控。而所謂的 EPC 技術，是一種透過回饋控制 (feedback control) 機制，針對製程輸出值來調整輸入變數進而縮小製程輸出值與目標值間差距的一種方法。藉著結合 SPC 可以成功辨識製程發生干擾原因的優點，EPC 已經成功的被應用在許多實際的製程監控問題上 (Box and Kramer, 1992 ; Faltin et al., 1993 ; Montgomery et al., 1994 ; Luceño, 1995 ; Box et al., 1997 ; Shao, 1998)。只是儘管結合 SPC 與 EPC 的方法能夠成功的偵測並修正製程的干擾因素，但對於干擾因素過小或不明顯的情況下，結合 SPC 與 EPC 的方法卻也常發生低估或高估干擾因素的錯誤狀態 (Shao and Chiu, 1999)。

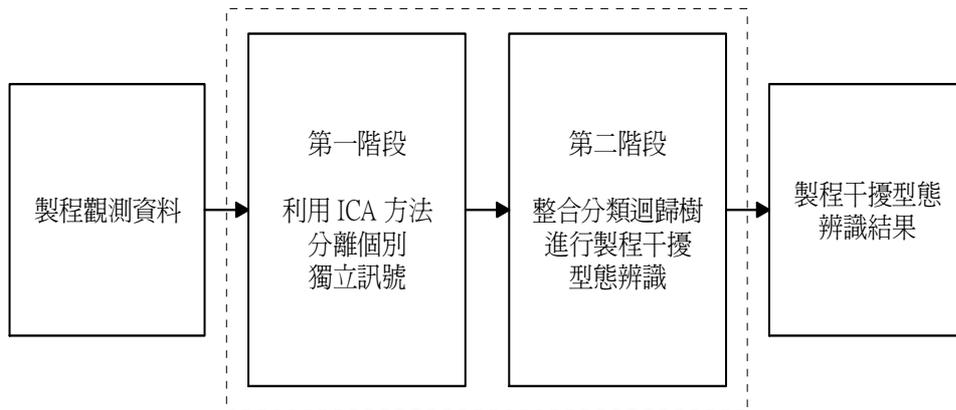
## 二、ICA獨立成分分析於統計製程監控

相對於在醫學訊號處理、語音訊號處理、特徵萃取及人臉辨識上的應用，ICA 模式較少被運用在處理統計製程管制的問題中。Xia and Howell (2003) 及 Xia (2003) 提出了一個稱為光譜 ICA (Spectral ICA) 的方法來進行製程的監控，他們首先利用間斷傅立葉轉換 (Discrete Fourier Transform) 將時間域 (time domain) 下的製程資料轉換至頻率域 (frequency domain) 資料，希望藉此消除觀測值中因為製程變化而產生的時間延遲問題，接著再針對頻率域資料運用 ICA 技術產生獨立成份 (ICs)，最後根據所對應出的較大震盪區域獨立成份進行製程之監控。Kano et al. (2003) 則是以獨立成份取代原始製程觀測值進

行製程的管制；該研究首先針對正常狀態下的製程觀測值進行 ICA 技術的應用並產生獨立成份，之後再利用蕭華特 (Shewhart) 管制圖計算每一個獨立成份的管制上下限，藉此監控製程。Lee et al. (2003, 2004) 則是以 ICA 方法針對化學製程及廢水處理過程進行監控。綜合上述研究，本研究發現由於在使用傳統 ICA 模式時，常會假設須有  $M$  個觀察訊號的存在，以滿足  $M$  個獨立成份的萃取 (Hyvärinen et al., 2001)，因此大多數的研究皆是以多品質特性的製程來進行監控。換言之，現有的 ICA 方法在 SPC 問題上的應用，大多是適用於多變數的製程監控，而較少應用在單變數的製程監控中。

## 參. 研究方法

呈如前言，本研究將先利用 ICA 的方法來進行資料的轉換，希望透過 ICA 的技術將個別獨立訊號由觀測資料中分離出來；再根據分離出來的個別獨立訊號，整合至第二階段的分類迴歸樹模式中，以建構一個辨識製程干擾型態的線上模式。圖一為本研究之研究架構。有關架構圖中所應用到之獨立成份分析方法及分類迴歸樹技術則分別描述如下。



圖一 觀念性研究架構

### 一、獨立成份分析 (Independent Component Analysis, ICA)

基本上，獨立成份分析是一種用來尋找隨機變數中隱藏因子的統計方法 (Comon, 1994 ; Lee, 1998 ; Hyvärinen & Oja, 2000)。通常使用者會先假設觀察到的隨機變數是由數個潛在變數 (latent variables)，以未知的線性組合方式混合而成。這些潛在的變數在互為獨立的假設情況下，被稱為觀察資料的潛在來源 (latent sources)，而所謂的 ICA 就是在只有觀察資料，卻無任何混合機制資訊與來源訊號的情況下，進行潛在來源訊號估計的一種方法。它的整個估計過程描述如下。

假設  $\mathbf{X}=[x_1, x_2, \dots, x_M]^T$  是一個由  $x_i$  所組成且維度為  $M \times N$  的混合訊號矩陣，而  $x_i$  代表了維度為  $1 \times N$  的時間序列訊號，則一個基本的 ICA 模式可以表示為 (Hyvärinen & Oja, 2000)

$$\mathbf{X} = \mathbf{A}\mathbf{S} = \sum_{i=1}^M \mathbf{a}_i \mathbf{s}_i \quad (1)$$

在(1)式中， $\mathbf{a}_i$  代表了在未知混合矩陣 (mixing matrix)  $\mathbf{A}$  中的第  $i$  個行向量 (column vector)，而  $\mathbf{s}_i$  除了是未知來源矩陣 (source matrix)  $\mathbf{S}$  中的第  $i$  個列向量 (row vector) 外，它也是在混合訊號矩陣  $\mathbf{X}$  中無法直接觀測到的潛在來源訊號。在 ICA 方法假設來源訊號  $\mathbf{s}_i$  互為統計獨立的條件下，一個維度  $M \times M$  的解混合矩陣 (de-mixing matrix)  $\mathbf{W}$  會被先估計出來，之後藉由  $\mathbf{W}$  可以將所觀察到的混合訊號矩陣  $\mathbf{X}$  進行轉換，以產生維度為  $M \times N$  的矩陣  $\mathbf{Y}$ ， $\mathbf{y}_i$  代表了在矩陣  $\mathbf{Y}$  中的第  $i$  個列向量，它同時也代表了期望由混合訊號中萃取出來的潛在來源訊號 (即所謂的獨立成份 (Independent Component, IC)，如(2)式。

$$\mathbf{Y} = [\mathbf{y}_i] = \mathbf{W}\mathbf{X} \quad (2)$$

最後，在比較方程式(1)與(2)之餘，可以發現當  $\mathbf{W}$  為混合矩陣  $\mathbf{A}$  的反矩陣時，即  $\mathbf{W}=\mathbf{A}^{-1}$ ，獨立成份  $\mathbf{y}_i$  將可用來估計潛在來源訊號  $\mathbf{s}_i$ 。換言之，第  $i$  個獨立成分  $\mathbf{y}_i$  可以用來代替  $\mathbf{s}_i$ ，並可以經由  $\mathbf{W}$  矩陣中相對應的第  $i$  個列向量  $\mathbf{w}_i$  乘上矩陣  $\mathbf{A}$  得到，亦即  $\mathbf{y}_i = \mathbf{w}_i \mathbf{X}$ ， $i = 1, 2, \dots, M$

由於各獨立成份間必須統計獨立是在運用 ICA 方法時的一個基本假設，因此各獨立成份間的獨立性高低常會先被定義成一個目標函數。然後，藉由某一特定的最佳化技術或非監督式的演算法，解混合矩陣  $\mathbf{W}$  將可以成功的被估計出來 (Hyvärinen & Oja, 2000)。此外，由於所謂的獨立成份必須具有非高斯分配 (non-Gaussian distribution) 的限制，因此在 ICA 方法中獨立成份的非高斯特性 (non-Gaussianity)，常會被定義成目標函數來衡量各獨立成份間的獨

立性 (Hyvärinen & Oja, 2000 ; Hyvärinen, et al., 2001)。事實上，現今已有許多成熟的技術被用來量測 ICA 中所謂獨立成份的非高斯性，例如高階統計量法 (high order cumulants)、共同資訊法 (mutual information) 及負熵法 (negentropy) 等。而在這一些方法中，以負熵法最常被討論與使用 (Hyvärinen, et al., 2001 ; David & Sanchez, 2002)。

負熵法是延伸資訊理論中熵 (entropy) 的定義所提出來的方法，在資訊理論中熵被視為一資訊量的測量單位，且被用來表示隨機觀察值的資訊程度，當隨機觀察值的結構性越差越亂 (即越無法預測) 時，其熵值會越大。若令隨機向量  $\mathbf{y}$  之機率密度函數為  $p_{\mathbf{y}}(\eta)$ ，則  $\mathbf{y}$  的熵  $H$  可定義為  $H(\mathbf{y}) = -\int p_{\mathbf{y}}(\eta) \log p_{\mathbf{y}}(\eta) d\eta$  (Cover & Thomas, 1991)。根據資訊理論，在所有具有相同變異數的隨機分佈中，呈高斯分佈的變數會具有最大的熵。因此若變數具非高斯分佈，則熵值會較其具有高斯分佈時來的小 (Hyvärinen, et al., 2001)。由此延伸，可以對熵  $H$  的定義做一修正當作非高斯的測量值。熵  $H$  修改後的定義稱為負熵  $J$  (negentropy)，定義如下 (Hyvärinen, et al., 2001)：

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (3)$$

其中  $\mathbf{y}_{gauss}$  表示和  $\mathbf{y}$  具有相同變異數的高斯變數向量。負熵的值永不為負值，即  $J(\mathbf{y}) \geq 0$ ，只有  $\mathbf{y}$  為高斯分佈時  $J(\mathbf{y})$  才會為零，所以 ICA 的目標函數就變成要最大化隨機向量  $\mathbf{y}$  的負熵，即 *Maximize*  $J(\mathbf{y})$ 。

然而，負熵最大的問題就是計算太過複雜，因為根據定義，要估計出負熵就必須先估計  $\mathbf{y}$  的機率密度函數  $f(\mathbf{y})$ 。為了解決這個問題，Hyvärinen (1999) 發展出負熵的近似函數

$$J(\mathbf{y}) \propto [E\{G(\mathbf{y})\} - E\{G(\nu)\}]^2 \quad (4)$$

其中  $\nu$  是平均值為零且變異數為 1 的高斯分佈隨機變數。 $G$  可為任何非二次方函數 (non-quadratic function)，因為  $G$  若為二次方函數， $J(\mathbf{y})$  必為零。

## 二、分類迴歸樹

基本上，CART 在建構模式時，必須先提供其兩組預先切割好的資料 (訓練樣本與測試樣本)，之後它會針對訓練樣本運用劃分條件建構出具有最多結點數目的決策樹。在進行決策樹建構的過程中，CART 會根據資料的「分散

度」來決定分隔變數的先後順序。而所謂的「分散度」是用來評估分隔變數重要程度的一個衡量指標：當分散度很高時，表示決策樹終結點的資料組合包含許多個不同類別的資料，而分散度很低時則表示終結點的資料類別較單純。有關 CART 的運算流程可大略整理如下，詳細之步驟則可參考 Breiman (1984)。

### (一)步驟一：建構最大樹狀結構

CART 乃使用遞迴二元劃分規則來進行資料的切割動作，並且使用每個結點的分散度高低來進行分隔變數的挑選，Breiman (1984) 等人稱之為結點複雜度 (node impurity)。其中最常用來評估結點異質性的指標為 Gini Index。

$$Gini(t) = 1 - \sum_i p_i^2 \quad (5)$$

其中  $t$  為已知之結點； $P_i$  則為第  $t$  個結點中第  $i$  個類別的機率。當 Gini 值越小時代表在該切割點所產生的子結點中，其組成份子越單純。而建構最大樹狀結構的流程會一直由上往下 (top-down) 進行，直到所有的資料均被劃分到同一類別，或是已經細分到某一類別具備有單一個樣本時，或是滿足了某些條件的情況下才停止。

### (二)步驟二：修剪樹狀結構，產生許多子樹群

當最大樹狀結構建構完成後，CART 會開始修剪樹狀結構以降低誤判率，並找出大小適當的決策樹。通常決策樹修剪過程所使用的準則為重代估計值 (re-substitution estimate)。若令  $R(T)$  為整個樹狀結構的重代估計值，當  $R(T)$  越小時代表末端結點個數越多，意謂著樹狀結構中的切割點越多，因此樹狀結構也越大。有關修剪樹狀結構的方程式可整理如下：

$$R(T) = \sum_{t \in T} R(t) \quad (6)$$

$$R(t) = \gamma(t)p(t) \quad (7)$$

$$\gamma(t) = 1 - \max_j p(j|t) \quad (8)$$

$$p(j|t) = \frac{p(j,t)}{P(t)} \quad (9)$$

$$p(t) = \sum_j p(j, t) \tag{10}$$

其中  $R(T)$  為第一步驟所產生的最大樹狀結構  $T$ ； $T$  為結點  $t$ ； $R(t)$  為第  $t$  個末端結點的重代估計值； $\gamma(t)$  為第  $t$  個末端結點的誤判率； $p(j|t)$  為第  $t$  個結點中，第  $j$  個類別所佔的比率； $p(t)$  為任一樣本落入結點  $t$  的機率； $p(j, t)$  為任一樣本落入結點  $t$  中，歸屬於第  $j$  個類別的機率。若令  $\alpha(\alpha \geq 0)$  為該樹的複雜度係數，則  $R_\alpha(T)$  可視為樹狀結構成本與複雜度的線性組合，並具備了誤判成本的意義。如此便可以針對每個  $\alpha$  值，找出其個別子樹使得其  $R_\alpha(T)$  為最小，之後再根據  $R_\alpha(T)$  值來進行樹的修剪。

### (三)步驟三：挑選最佳樹狀結構

根據步驟二中每個  $\alpha$  值所求得的樹，便可以獲得巢狀樹群  $T_1 > T_2 > \dots > \{t_k\}$  和其所對應的複雜度係數數列  $0 = \alpha_1 < \alpha_2 < \alpha_3 < \dots < \alpha_k (T_k = T(\alpha_k), \alpha_1 = 0)$ ，其中  $t_k$  為第  $k$  個使得  $R_\alpha(T_t) = R_\alpha(\{t\})$  的結點，最後再將測試樣本代入並計算巢狀樹群中的每個樹狀結構  $R(T_k)$  值，並選取具有最小  $R(T_k)$  值的樹狀結構為最佳的樹狀結構。

## 肆· 實證研究

為了驗證在本研究中所提整合方法的有效性，本研究針對兩種不同類型的常見干擾：平移式干擾 (step-change disturbance) 與線性式干擾 (linear disturbance) 進行驗證與測試。基本上，所謂的平移式干擾可以表示成：

$$D_t = \begin{cases} L, & t \geq t_d \\ 0, & t < t_d \end{cases} \tag{11}$$

其中  $D_t$  代表在時點  $t$  所產生的干擾項， $L$  則是代表了平移式干擾的幅度大小； $t_d$  則是干擾產生的特定時間點。另外，有關線性式干擾則可以表示成：

$$D_t = \begin{cases} S(t - t_s), & t \geq t_d \\ 0, & t < t_d \end{cases} \tag{12}$$

其中  $S$  是線性式干擾的趨勢大小。之後，在假設  $B$  為一個後退式操作元 (即  $B2Dt=Dt-2$ ) 及  $\beta_t$  是一個當干擾發生時才會有數值 (其餘狀態皆為零) 的隨機變數下，可以將方程式(11)與(12)轉換成下式：

$$D_t = \frac{\beta_t}{1-B} \quad \beta_t = \begin{cases} L, & t = t_d \\ 0, & t \neq t_d \end{cases} \quad \text{for 平移式干擾} \quad (13)$$

$$D_t = \frac{\beta_t}{(1-B)^2} \quad \beta_t = \begin{cases} S, & t = t_d \\ 0, & t \neq t_d \end{cases} \quad \text{for 線性式干擾} \quad (14)$$

由於在文獻中 (MacGregor, 1988) 曾指出，一階模式可以成功有效地配適製造程序中的製程資料，因此本研究將擬研究的製程資料定義成一階模式，此外也採用了具有一階性質的 IMA(1, 1) (Integrated Moving Average) 模型來表達製程資料中的噪音項。

綜合上述，本研究中所採用的驗證資料可以使用下列方程式來表達：

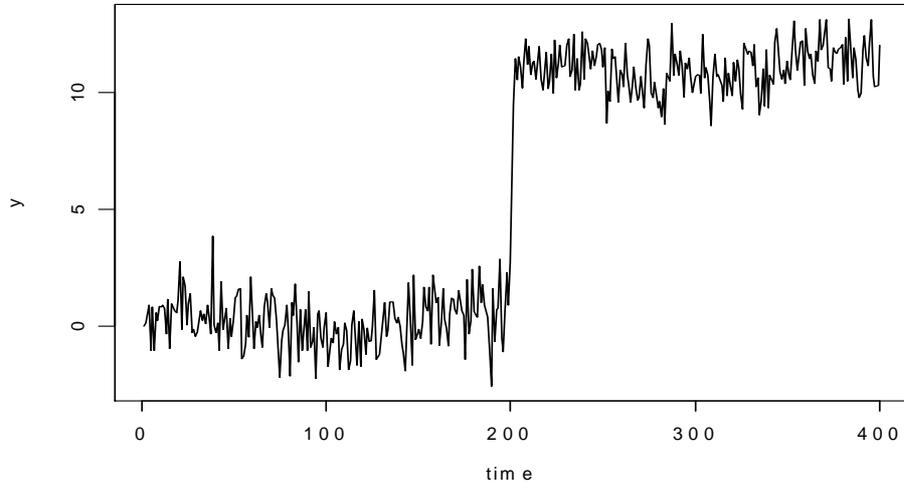
$$y_{t+1} = \frac{q}{1-pB} X_t + d_{t+1} + D_{t+1} \quad (15)$$

$$d_{t+1} = \frac{1-\theta B}{1-B} \varepsilon_{t+1} \quad (16)$$

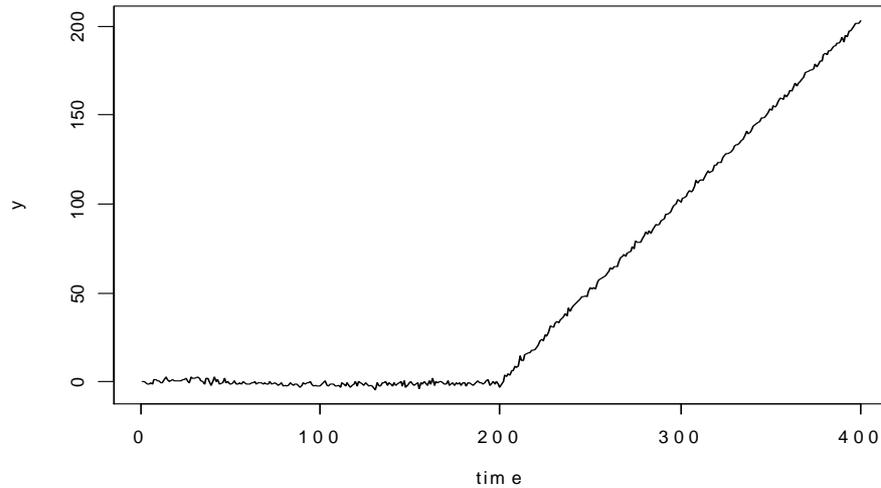
其中  $y_{t+1}$  表示在  $t+1$  時間點時的製程資料，而  $p$  與  $q$  則是一階模式中的固定參數( $p+q=1$ )； $X_t$  是在時間點  $t$  下，針對製程的控制變數； $d_{t+1}$  則是在  $t+1$  時點下，符合 IMA(1, 1)性質的噪音項； $\varepsilon_{t+1}$  則是一符合平均值為零、變異數( $\sigma^2$ )為 1 常態隨機變數，而  $\theta$  為  $\varepsilon_t$  的自我相關係數。

在設定  $p$ ,  $q$  與  $\theta$  參數分別為 0.8, 0.2 與 0.9 的情況下，本研究根據方程式 (15)~(16) 模擬產生 A、B 兩組總共包含 16 小組的訓練資料集 (即 A1-A8 及 B1-B8 的訓練資料)。其中 A 組內的 A1-A4 訓練資料屬於「平移式干擾」資料型態；A5-A8 資料則是屬於「線性式干擾」資料型態。每一小組資料都包括了 400 個無干擾資料向量及 400 個平移式或線性式干擾的資料向量。而 A1 小組的 400 個資料向量是根據當  $L=1$  的方程式(13)及方程式(15)~(16)所產生出來的。同理，A2-A4 小組的資料向量則是以  $L=5, 10$  與  $15$  的設定來產生。至於 A5-A8 小組的資料向量，則是分別在  $S=0.5, 1.0, 1.5$  及  $2.0$  的情況下，以方程式(14)及方程式(15)~(16)模擬產生而得。至於 B 組中所包含的 8 小組訓練資

料，其模擬機制與 A 組中資料的產生方式相同。圖二與 3 分別代表在設定  $t_d=201$  (干擾發生時點) 的情況下，A 組中當  $L=5$  與  $S=1.0$  的製程資料。



圖二 模擬「平移式干擾」資料 ( $L=5$ )



圖三 模擬「線性式干擾」資料 ( $S=1.0$ )

## 一、階段一：結合SPC、EPC與ICA的資料轉換

爲了配合實際製程的監控作業，本研究假設當製程資料符合一階模式且其噪音項具有  $IMA(1, 1)$  性質時，Proportional Integral (PI) 控制器 (如式 (17))，會被用來扮演 EPC 的角色：進行製程的調整。而在這裡本研究所謂的 PI 控制器，通常以最小化偏離值的均方誤差 (Minimum Mean Squared Error, MMSE) 爲目標所建立的控制器。

$$X_t = -\frac{(1-\theta)p}{q} \left( y_t + \frac{1-p}{p} \sum_{i=-\infty}^t y_i \right) \quad (17)$$

$$d_t + D_t = y_t + (1-\theta) \sum_{i=1}^{t-1} y_i \quad (18)$$

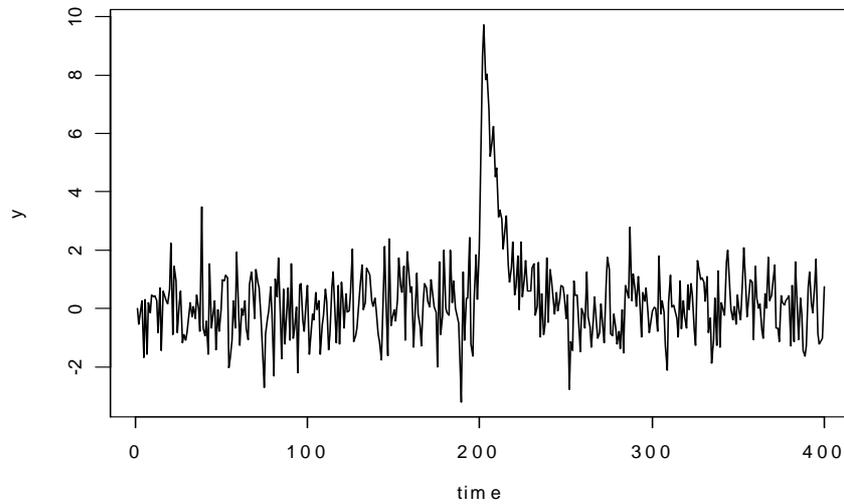
若將方程式(17)代入方程式(15) (亦即應用 PI 或 MMSE 控制器針對製程資料進行調整) 後，原始的製程資料將會被轉換成現階段時間點下的製程資料  $y_t$  與加權後歷史製程資料  $((1-\theta) \sum_{i=1}^{t-1} y_i)$  的總和 (如式(18))。

此外，若將方程式(13)與(14)代入方程式(18)之中，便可獲得在不同干擾型態下，應用 SPC 與 EPC 作業所轉換而成的製程資料 (如式(19)與(20))。

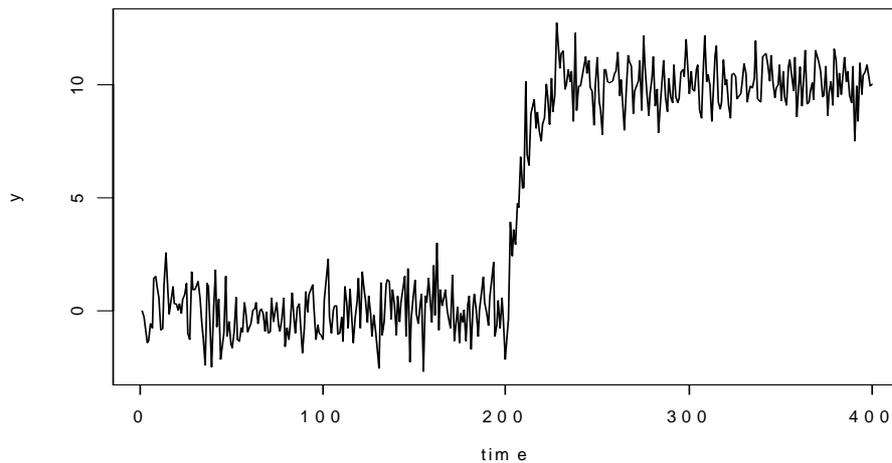
$$y_{t_d+i} = L\theta^{i-1} + \varepsilon_{t_d+i} \quad \text{for 平移式干擾} \quad (19)$$

$$y_{t_d+i} = S\left(\frac{\theta^i - 1}{\theta - 1}\right) + \varepsilon_{t_d+i} \quad \text{for 線性式干擾} \quad (20)$$

圖四與圖五分別代表利用方程式(19)與(20)所產生之製程資料。



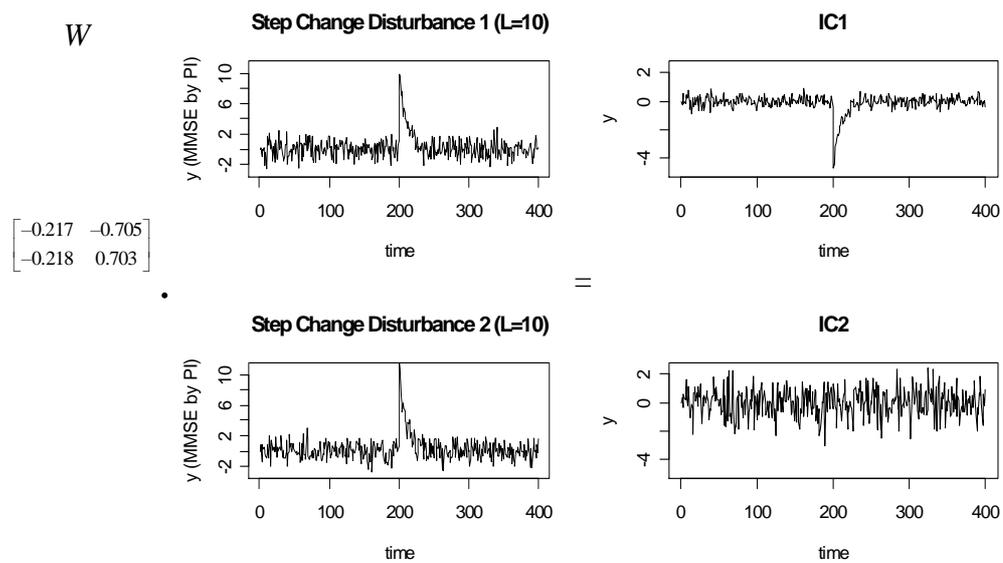
圖四 應用 PI 或 MMSE 控制器進行調整後的「平移式干擾」資料 (L=10)



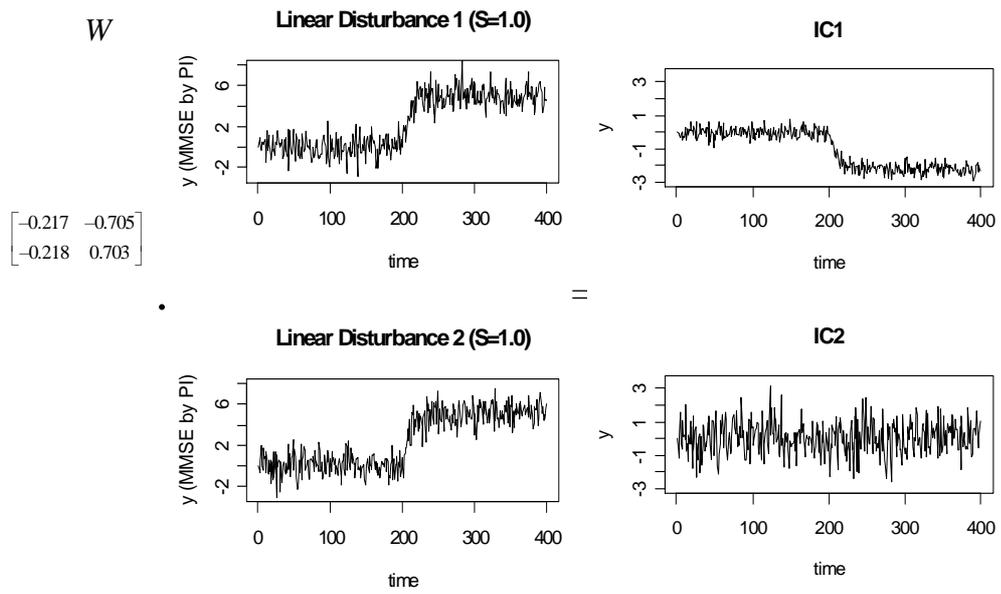
圖五 應用 PI 或 MMSE 控制器進行調整後的「線性式干擾」資料 (S=1.0)

基本上，針對製程資料進行干擾型態的辨識工作，主要是根據資料所呈現的 **pattern** 來進行判別，如果能將混合在製程資料中的雜訊予以萃取並過濾將有助於干擾 **pattern** 的明顯化，因此本研究將應用獨立成份分析的技術來進行資訊萃取的工作。在應用獨立成份分析方法的部分，可以針對經由 **SPC** 與 **EPC** 方法所轉換求得的資料 **Y**，運用解混合矩陣 (**de-mixing matrix**) **W** 進行線性轉換以求得一新變數矩陣 **Z** (即  $Z=WY$ )。而在 **Z** 矩陣當中，列向量 (表示成

z) 間除了彼此互不相關外 (uncorrelated)，本研究也假設其變異數必須為 1；亦即  $E(\mathbf{z}\mathbf{z}^T)=1$ 。而為了順利的估計出本研究中的獨立成份  $\mathbf{Z}$  值，本研究首先根據 Hyvarinen et al. (2001) 所提出的 gradient algorithm，利用經過 SPC 與 EPC 方法所轉換後的 A、B 組訓練資料集，進行解混合矩陣  $\mathbf{W}$  的估計。此處使用 A、B 組訓練資料集來估計  $\mathbf{W}$  矩陣的原因，在於本研究觀察到無論是平移式干擾製程資料或線性式干擾製程資料都可能包含兩種不同的訊號來源：一個是干擾製程資料本身，另一個則是噪音訊號。因此本研究建構了一個  $2 \times 2$  的  $\mathbf{W}$  矩陣來進行獨立成份的估計；而利用方程式  $\mathbf{Z}=\mathbf{W}\mathbf{Y}$  及 A、B 兩組經過轉換後資料所估計出來的  $\mathbf{W}$  矩陣及獨立成份部分結果，可以整理如圖六。之後再利用上述所估計出來的  $\mathbf{W}$  矩陣，以一組新模擬資料搭配一組噪音訊號的方式，進行獨立成份的計算（在此階段，只使用一組模擬資料的原因在於本研究假設只針對具有一個品質特性的製程資料進行監控）。相關的結果可以整理如圖七所示。由圖六與圖七中可以發現，經由 ICA 方法所萃取出來的獨立成份 (IC1 與 IC2)，確實能將噪音訊號從混合訊號中成功地分離出來。

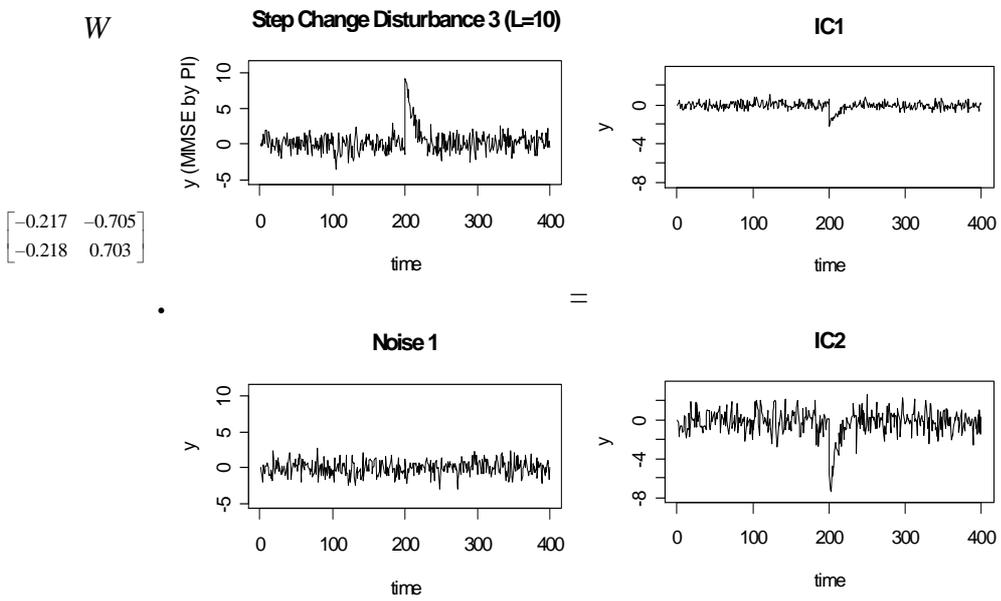


(a) 「平移式干擾」資料 ( $L=10$ ) 的獨立成份估計結果

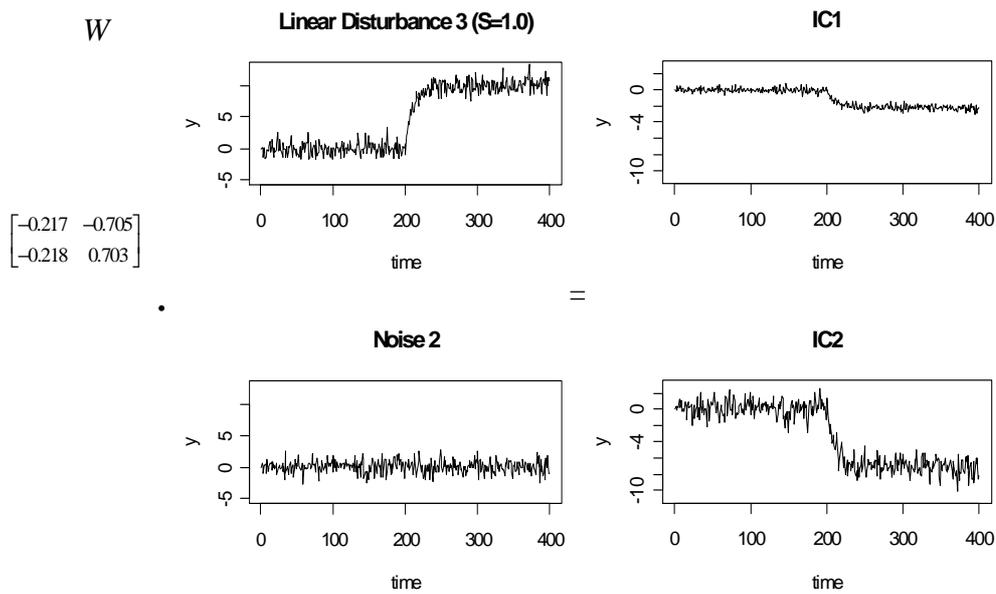


(b) 「線性式干擾」資料 (S=1.0) 的獨立成份估計結果

圖六 根據轉換後訓練資料集所估計出之  $W$  矩陣與獨立成份部分結果



(a) 新一組「平移式干擾」資料 (L=10) 的獨立成份計算結果



(b) 新一組「線性式干擾」資料 (S=1.0) 的獨立成份計算結果

圖七 應用一組新模擬資料與一組噪音訊號所求算之獨立成份部分結果

## 二、階段二：建立製程干擾辨識與分類模式

在經過 EPC 與 ICA 的資料轉換後，噪音訊號可以從製程資料中成功的被分離出來。此外，由於每一筆轉換後的資料均可依照原始模擬過程中所使用的參數值 (L 與 S) 而被歸屬到特定的類別中，因此可以進一步針對轉換後的所有資料建立製程干擾辨識與分類預測的模式。在建構干擾辨識與分類模式的技術上，本研究採用了分類迴歸樹的方法來進行，主要目的乃在於希望能利用決策樹在建樹時所產生的法則，更清楚地讓品管人員了解到監控對象與產生干擾因素間的關係，以作為後續調整製程的一個重要參考。

在輸入變數的選擇上，由於在第一階段中經由 ICA 所分離出來的非噪音項獨立成份只有一個，因此本研究根據該獨立成份的趨勢進行可能資訊的篩選。在觀察過該獨立成份的樣式 (pattern) 後，本研究採取 X1：前 5 期平均數、X2：前 10 期變異數、X3：t-1 期斜率、X4：t-2 期斜率、X5：單點搭配 5 期平均數、X6：前 3 期平均數等六個變數來進行不同製程干擾類型與大小的分類與辨識。而為了測試以分類迴歸樹方法建構干擾辨識與分類模式的準確性，

本研究以隨機選取之方法，挑選轉換後獨立成份資料中的 2880 筆 (90%) 資料作為訓練模式之用，而剩下之 320 筆 (10%) 則保留為測試資料。

在經過分類迴歸樹方法的運算後可以得到如表一的變數重要程度、表二的推論規則 (部分) 及相關的分析結果 (彙總如表三及表四)，其中表三及表四分別為不同干擾型態下，在不同相關程度( $\theta$ )時 CART 方法所建構製程干擾辨識模式的辨識成功率。根據表一的數據可以發現六個變數在分類規則裡的相對重要程度依序為 X5：單點搭配 5 期平均數、X1：前 5 期平均數、X2：前 10 期變異數、X6：前 3 期平均數 X3：t-1 期斜率、X4：t-2 期斜率，故可以發現 X5：單點搭配 5 期平均數與 X1：5 期平均數變數的重要性最高，這分別代表單點觀察值搭配五期平均數能成功辨識出資料的變化趨勢；此外，X2：前 10 期變異數與 X6：

前 3 期平均數則代表了特定時段內的資料變化與平移現象能協助辨識有無干擾的發生。由表二的規則中，可以很清楚地了解到 L=10 與 L=15 的兩種平移式干擾製程，能藉由 X1：前 5 期平均數與 X6：前 3 期平均數兩種變數的特定上下限成功的被辨識出來。其中，單點搭配 5 期平均數=-0.556513 及前 3 期平均數=-5.29759 兩個特定數值，將可以被參考來進行製造程序的改善。

表一 輸入變數的相對重要程度 (%) ( $\theta=0.9$ )

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
重要程度	83.03	65.37	35.74	29.72	100.0	63.29

表二 CART 模式所建構之製程干擾辨識規則 (部分)

終結點	規則	類別
1	If 是否 X <sub>5</sub> ≤ -0.556513 & 是否 X <sub>6</sub> ≤ -5.29759	Class=4 平移式干擾 (L=15)
2	If 是否 X <sub>5</sub> ≤ -0.556513 & 是否 X <sub>6</sub> > -5.29759 & 是否 X <sub>6</sub> ≤ -3.1004	Class=3 平移式干擾 (L=10)

表三 整合 SPC/EPC/ICA 與 CART 監控法之辨識正確率 (平移式干擾)

	L				
	0	1	5	10	15
0.9	99.4%	59.6%	99.8%	100.0%	99.6%
0.8	98.1%	45.3%	80.1%	100.0%	99.9%
0.7	98.8%	20.5%	100.0%	99.7%	99.7%
0.6	98.1%	45.1%	84.7%	100.0%	100.0%
0.5	99.4%	40.0%	80.0%	100.0%	99.8%
0.4	98.1%	30.1%	69.6%	100.0%	100.0%
0.3	99.4%	35.0%	74.6%	95.0%	99.6%
0.2	96.3%	34.5%	69.9%	90.3%	99.6%
0.1	98.8%	39.9%	70.5%	84.9%	100.0%
0.0	97.5%	29.9%	70.4%	95.4%	94.5%
-0.1	97.5%	40.3%	85.5%	95.0%	100.0%
-0.2	98.1%	9.8%	44.5%	90.4%	89.8%
-0.3	96.3%	45.1%	95.1%	85.2%	99.8%
-0.4	97.5%	45.3%	44.6%	85.3%	95.4%
-0.5	95.6%	15.1%	70.0%	85.1%	95.3%
-0.6	99.4%	30.3%	59.6%	95.4%	94.8%
-0.7	96.3%	44.8%	90.3%	95.3%	100.0%
-0.8	98.8%	14.7%	89.9%	95.2%	100.0%
-0.9	91.3%	45.1%	95.2%	99.7%	99.5%

表四 整合 SPC/EPC/ICA 與 CART 監控法之辨識正確率 (線性式干擾)

	S				
	0	0.5	1.0	1.5	2.0
0.9	98.8%	25.4%	74.9%	50.1%	95.4%
0.8	98.7%	39.6%	29.9%	69.8%	79.6%
0.7	99.2%	30.3%	54.9%	55.2%	79.8%
0.6	98.2%	44.5%	34.7%	75.4%	74.9%
0.5	99.0%	30.0%	20.0%	39.9%	60.2%
0.4	97.5%	29.9%	35.2%	29.6%	75.5%
0.3	99.1%	40.4%	20.2%	35.0%	64.5%
0.2	96.4%	20.1%	19.6%	19.9%	60.0%
0.1	99.1%	10.3%	20.2%	29.9%	74.6%
0.0	97.2%	14.7%	14.9%	54.6%	65.0%
-0.1	97.4%	10.5%	24.9%	25.4%	59.9%
-0.2	97.3%	15.4%	10.1%	30.2%	44.8%
-0.3	95.9%	30.5%	39.8%	15.2%	40.3%
-0.4	98.0%	5.2%	35.5%	15.4%	49.8%
-0.5	95.0%	15.5%	15.1%	14.6%	65.1%
-0.6	98.8%	14.6%	14.9%	14.8%	35.4%
-0.7	95.3%	19.5%	15.4%	30.2%	34.5%
-0.8	99.7%	19.7%	10.5%	25.1%	40.4%
-0.9	92.2%	9.5%	5.3%	10.3%	35.1%

根據表三與表四的數據可以發現當資料中相關性( $\rho$ )愈高時，使用 CART 模式所建構的製程干擾辨識模式在進行干擾型態與類別辨識時的辨識正確率也會隨之增高；此外，不論是平移式干擾資料或線性式干擾資料，辨識模式的辨識成功率會普遍會隨著 L 值或 S 值的增加而升高；最後，從表三與表四的數據中可觀察到，辨識模式對平移式干擾資料的辨識成功率高於對線性式干擾資料的辨識成功率；究其因本研究發現是與轉換後的獨立成份 (IC) 資料有關：平移式干擾資料在經過獨立成份分析方法轉換後的資料波動性遠比線性式干擾資料者明顯許多，因此當平移式干擾發生時，CART 方法所建構的製程干擾辨識模式可以較正確地偵測出該干擾型態及類別。

## 伍· 結論

本研究將具有自我相關性製程的資料視為一個包含了雜訊與製程資料本身的混合訊號，並導入了近年來成功應用於訊號分離之獨立成分分析 (ICA) 技術與分類迴歸樹方法 (CART)，提出一個以結合 SPC、EPC、ICA 與 CART 的製程監控架構。不同於傳統的製程監控方法，整合 SPC/EPC/ICA 與 CART 的方式除了避免繁複的數學函數配適外，對於欲監測的觀察變數亦能快速且正確的予以修正，促使製程監控效率得以提升。

在整合的過程中，本研究先使用結合 SPC/EPC 與 ICA 的技術來進行資料的轉換，目的是希望能先利用 EPC 將製程中的干擾或偏誤做即時且適當地修正，之後再透過獨立成份分析 (Independent Component Analysis, ICA) 方法的應用，將噪音訊號從製程資料中分離出來。最後，本研究針對萃取出的重要訊號，透過分類迴歸樹 (CART) 的方法來建構一個製程干擾的辨識模式。選擇決策樹技術做為建構辨識模式方法的主要原因在於，其所提供的推論準則能夠相當透明化的解析辨識的結果；而品管人員或製程分析師也將可以根據相關的推論規則，來針對製造程序進行進一步的改善。為了驗證所提出整合方法的有效性，本研究同時針對兩種不同類型的常見干擾：平移式干擾 (step-change disturbance) 與線性式干擾 (linear disturbance) 運用 Shewhart 與 CUSUM 管制圖進行驗證與測試。而根據實驗結果顯示在辨識不同型態及程度的干擾資料時，ICA 可以成功的將噪音訊號從製程資料中分離出來。

## 參考文獻

- Adams, B. M. and Tseng, L. T. "Robustness of Forecast-based Monitoring Schemes", *Journal of Quality Technology*, (30), 1998, pp.328-329.
- Alwan, L.C. and Roberts, H.V., "Time Series Modeling for Statistical Process Control", *Journal of Business and Economic Statistics*, (6), 1988, pp.87 -95.
- Bartlett, M. S., Movellan, J. R. and Sejnowski, T. J., "Face Recognition by Independent Component Analysis", *IEEE Transactions on Neural Networks*, (13), 2002, pp.1450-1464.
- Beckmann, C. F. and Smith, S. M., "Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging", *IEEE Transactions on Medical Imaging*, (23), 2004, pp.137-152.
- Box, G. E. P. and Kramer, T., "Statistical Process Monitoring and Feedback Adjustment - A Discussion", *Technometrics*, (34), 1992, pp.251-285.

- Box, G. E. P. and Luceño, A., "Statistical Control by Monitoring and Feedback Adjustment", John Wiley and Sons, New York, NY. , 1997.
- Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., "Classification and Regression Trees", Pacific Grove, CA:Wadsworth and Brooks.Gee, 1984.
- Chai, B. B., Huang, T., Zhuang, X. H., Zhao, Y. and Sklansky, J., "Piecewise Linear Classifiers Using Binary Tree Structure and Genetic Algorithm", *Pattern Recognition*, 1996, pp.1905-1917.
- Cichocki, A. and Amari, S. I., "Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications", John Wiley & Sons, New York, NY. , 2002.
- Comon, P., "Independent Component Analysis: A New Concept", *Signal Processing*, (36), 1994, pp.287-314.
- Cover, T. M. and Thomas, J. A., "Elements of Information Theory", John Wiley and Sons, New York, N.Y., 1991.
- David, V. and Sanchez , A., "Frontiers of Research in BSS/ICA", *Neurocomputing*, (49), 2002, pp.7-23.
- Déniz, O., Castrillón, M. and Hernández, M., "Face Recognition Using Independent Component Analysis and Support Vector Machines", *Pattern Recognition Letters*, (24), 2003, pp.2153-2157.
- Faltin, F. W., Hahn, G. J., Tucker, W. T. and Vander Wiel, S. A., "Algorithmic Statistical Process Control: Some Practical Observations", *International Statistic Reviews*, (61), 1993, pp 67-80.
- Faltin, F. W., Mastrangelo, C. M., Runger, G. C. and Ryan, T. P., "Considerations in the Monitoring of Autocorrelated and Independent Data", *Journal of Quality Technology*, (29), pp. 131-133 (1997).
- Hyvärinen, A., "Fast and Robust Fixed-point Algorithms for Independent Component Analysis", *IEEE Transactions on Neural Networks*, (10), 1999, pp.626-634.
- Hyvärinen, A., Karhunen, J. and Oja, E., "Independent Component Analysis", John Wiley & Sons, New York, N.Y., 2001.
- Hyvärinen, A. and Oja, E., "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, (13), 2000, pp.411-430.
- Jiang, W., Tsui, K. L. and Woodall, W. H., "A New SPC Monitoring Method: the ARMA Chart", *Technometrics*, (42), 2000, pp.399-410.
- Kano, M., S. Tanaka, S., Hasebe, I., Hashimoto and Ohno, H., "Monitoring Independent Components for Fault Detection", *AIChE Journal*, (49), 2003, pp.969-976.
- Kass, G. V., "An Exploratory Technique for Investigating Large Quantities of Categorical Data", *Applied Statistics*, (29), 1980, pp.119-127.
- Kim, T. K., Kim, H., Hwang, W. and Kittler, J., "Independent Component Analysis in a Local Facial Residue Space for Face Recognition", *Pattern Recognition*, (37), 2004, pp.1873-1885.

- Lee, J. M., Yoo, C., and Lee, I. B., "Statistical Process Monitoring with Multivariate Exponentially Weighted Moving Average and Independent Component Analysis", *Journal of Chemical Engineering of Japan*, (36), 2003a, pp.563-577.
- Lee, J. M., Yoo, C., and Lee, I. B., "On-line Batch Process Monitoring Using Different Unfolding Method and Independent Component Analysis", *Journal of Chemical Engineering of Japan*, (36), 2003b, pp.1384-1396.
- Lee, J. M., Yoo, C., and Lee, I. B., "Statistical Process Monitoring with Independent Component Analysis", *Journal of Process Control*, (14), 2004, pp.467-485.
- Lee, T. W., "Independent Component Analysis: Theory and Application", Kluwer Academic Publishers, Boston, MA., 1998.
- Lin, Q. H., Zheng, Y. R., Yin, F. and Liang, H. L., "Speech Segregation Using Constrained ICA", *Lecture Notes in Computer Science*, (3173), 2004, pp.755-760
- Loh, W. and Vanichsetakul, N., "Tree-structured Classification via Generalized Discriminant Analysis", *Journal of the American Statistical Association*, (83), 1998, pp.715-728.
- Lu, C. W. and Reynolds, M. R. JR., "EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes", *Journal of Quality Technology*, (31), 1999, pp.166-188.
- Lu, C. W. and Reynolds, M. R. JR., "CUSUM Charts for Monitoring an Autocorrelated Processes", *Journal of Quality Technology*, (33), 2001, pp.316-334.
- Luceño, A., "Choosing the EWMA Parameter in Engineering Process Control", *Journal of Quality Technology*, (27), 1995, pp.162-168.
- Macgregor, J. F., "On-line Statistical Process Control", *Chemical Engineering Progress*, (84), 1988, pp.21-31.
- Montgomery, D. C. and Mastrangelo, C. M., "Some Statistical Process Control for Autocorrelation Data", *Journal of Quality Technology*, (23), 1991, pp.179-193.
- Montgomery, D. C., "Introduction to Statistical Quality Control", Fourth Edition, John Wiley & Sons, New York, NY, 2001.
- Montgomery, D. C., Keats, J. B., Runger, G. C., and Messina, W. S., "Integrating Statistical Process Control and Engineering Process Control", *Journal of Quality Technology*, (26), 1994, pp.79-87.
- Morgan, J. N. and Sonquest, J. A., "Problem in the Analysis of Survey Data and a Proposal", *Journal of the American Statistical Association*, (58), 1963.
- Ohmann, C., Moustakis, V., Yang, Q. and Lang, K., "Evaluation of Automatic Knowledge Acquisition Techniques in the Diagnosis of Acute Abdominal Pain", *Artificial Intelligence in Medicine*, 1996, pp.23-36.
- Quinlan, J. R. "Induction of Decision Trees", *Machine Learning*, 1986, pp.81-106.
- Runger, G. C., Willemain, T. R. and Prabhu, S., "Average Run Lengths for CUSUM Control Charts Applied to Residuals", *Communication in Statistics-theory and Methods*, (24), 1995, pp.273-282.

- Shao, Y. E. and Chiu, C. C., "Developing Identification Techniques with the Integrated Use of SPC/EPC and neural networks", *Quality and Reliability Engineering International*, (15), 1999, pp.287-294.
- Shao, Y. E., "Integrated Application of the Cumulative Score Control Chart and Engineering Process Control", *Statistic Sinica*, (8), 1998, pp.239-252.
- Shao, Y. E., RUNGER, G. C., HADDOCK, J. and WALLACE, W. A., "Adaptive Controllers to Integrate SPC and EPC", *Communications in Statistics- Simulation and Computation*, (28), 1999, pp.13-36.
- Shao, Y. E. and Chiu, C. C., "Developing Identification Techniques with the Integrated Use of SPC/EPC and Neural Networks", *Quality and Reliability Engineering International*, (15), 1999, pp.1-8.
- Tseng, S. and Adams, B. M., "Monitoring Autocorrelated Processes with an Exponentially Weighted Moving Average Forecast", *Journal of Statistical Computation and Simulation*, (50), 1994, pp.187-195.
- Vigario, R., J. Sarela, V. Jousmaki, M. Hamalainen and Oja, E., "Independent Component Approach to the Analysis of EEG and MEG Recordings", *IEEE Transactions on Biomedical Engineering*, (47), 2000, pp.589-593.
- Wieringa, J. E., "Statistical Process Control for Serially Correlated Data", Ph. D. Dissertation, University of Groningen, Netherlands, 1999.
- Xia, C., "Control Loop Measurement Based Isolation of Faults and Disturbances in Process Plants", Ph.D. these, University of Glasgow, UK, 2003.
- Xia, C. and Howell, J., "Isolating Multiple Sources of Plant-wide Oscillations via Independent Component Analysis", *Proceedings of 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Washinton, D.C., USA, 2003.
- Zhang, N. F., "A Statistical Control Chart for Stationary Process Data", *Technometrics*, (35), 1998, pp.37-52.

# Disturbance Identification Using Independent Component Analysis with CART Approach

SHIN-YING HUANG, CHIH-CHOU CHIU \*

## ABSTRACT

It is well known for many years that statistical process control (SPC) and engineering process control (EPC) are both effective in the monitoring and adjusting of the manufacturing process. However, both schemes assume that process data are identically and independently distributed. Moreover, the real process data are actually serially correlated. The presence of autocorrelation has an adverse effect on the performance of traditional SPC or EPC approach. Because the correlated process data could be a mixture of noise and process characteristics such as process disturbances and/or autocorrelation, a disturbance identification scheme based on independent component analysis (ICA) is proposed in this paper. Basically, ICA is a novel statistical signal processing technique that was originally applied to find the latent source signals from observed mixture signal. In this paper, the ICA technique is first integrated with SPC and EPC approach to extract the independent components that contain different characteristics of the process. Then, the CART methodology is used to monitor the independent components and identify the process disturbance.

Keywords: statistical process control, engineering process control, independent component analysis, CART

---

\* Shin-Ying HUANG, Graduate Student, Institute of Commerce Automation and Management, National Taipei University of Technology. Chih-Chou CHIU, Professor, Institute of Commerce Automation and Management, National Taipei University of Technology.

