

經濟部及所屬機關因公出國人員報告書
(出國類別：研習)
出國報告系統識別號：C09205373

數據分析技術研習

服務機關：中國石油公司煉製研究所
職 務：化學工程師
姓 名：蔡坤祥
出國地點：美國
出國期間：92 年 11 月 8 日～92 年 11 月 15 日
報告日期：93 年 2 月 10 日

25/
C09205373

系統識別號: C09205373

公 務 出 國 報 告 提 要

頁數: 14 含附件: 否

報告名稱:

數據技術分析研究

主辦機關:

中國石油股份有限公司

聯絡人/電話:

葉宇容/87258422

出國人員:

蔡坤祥 中國石油股份有限公司 煉製研究所 化學工程師

出國類別: 研究

出國地區: 美國

出國期間: 民國 92 年 11 月 08 日 - 民國 92 年 11 月 15 日

報告日期: 民國 93 年 02 月 10 日

分類號/目: I5/化學與環境科學 /

關鍵詞: 實驗設計法, 混合物實驗設計, 配方優化, 數據分析

內容摘要: 近年來歐美國家及研究機構發展出一種跨領域之新的學門—化學計量學 (Chemometrics)，綜合利用數學、統計、或邏輯方法設計最佳實驗程序，及分析實驗數據，以期獲得最多有用之化學資訊。本次出國研習之主題為化學計量學中之實驗設計法部分—混合物實驗設計及數據分析 (Mixture Experimental Design)。混合物實驗設計 (或配方實驗設計) 是一種特殊的實驗設計，其反應值Y與組成成分間之比例Xi有關而與組成成分之絕對量無關。這類混合物實驗問題在工業、農業和科學試驗中屢見不鮮，例如汽油、潤滑油、混凝土、塑膠、合金、陶瓷、油漆、食品、醫藥、洗滌劑、混紡纖維…等產品都會遇到混合物實驗設計問題。本次研習之主要目的即學習混合物實驗設計技術，希望能以最少的實驗找出最佳的配方。本報告敘述研習心得包括(1)有那些適用之混合物數學模式，(2)如何配置最少之實驗點以配適上述之數學模式，(3)如何檢定配適之數學模式為適當可用之模式，(4)如何找出最佳之配方等。研習所學將可應用於本公司各種配方之開發及優化上，如汽油配方、潤滑油配方、觸媒配方...等。

本文電子檔已上傳至出國報告資訊網

數據分析技術研習

摘 要

近年來歐美大學及研究機構發展出一種跨領域之新的學門—化學計量學 (Chemometrics)，綜合利用數學、統計、或邏輯方法設計最佳實驗程序，及分析實驗數據，以期獲得最多有用之化學資訊。本次出國研習之主題為化學計量學中之實驗設計法部分—混合物實驗設計及數據分析 (Mixture Experimental Design)。混合物實驗設計 (或配方實驗設計) 是一種特殊的實驗設計，其反應值 Y 與組成分間之比例 X_i 有關而與組成分之絕對量無關。這類混合物實驗問題在工業、農業和科學試驗中屢見不鮮，例如汽油、潤滑油、混凝土、塑膠、合金、陶瓷、油漆、食品、醫藥、洗滌劑、混紡纖維...等產品都會遇到混合物實驗設計問題。本次研習之主要目的即學習混合物實驗設計技術，希望能以最少的實驗找出最佳的配方。本報告敘述研習心得包括 (1)有那些適用之混合物數學模式，(2)如何配置最少之實驗點以配適上述之數學模式，(3)如何檢定配適之數學模式為適當可用之模式，(4)如何找出最佳之配方等。研習所學將可應用於本公司各種配方之開發及優化上，如汽油配方、潤滑油配方、觸媒配方...等。

關鍵詞：實驗設計法、混合物實驗設計、配方優化、數據分析

數據分析技術研習報告

目 錄

一、前言	4
二、行程	4
三、Stat -Ease 公司簡介	5
四、研習課程內容	6
五、研習心得	7
六、感想.....	15

數據分析技術研習

一、前言：

在研發或改善過程中，通常需要做些實驗，如何安排實驗，使實驗次數儘量少，而又能達到好的實驗效果，這是經常會碰到的問題；另外由於分析量測儀器之進步，研究及現場操作均產生非常大量之數據，如何從這些數據中分析得到資訊，並進而轉化為有用的知識，亦為提昇研發效率及現場改善效率之重要課題。近年來歐美大學及研究機構發展出一種跨領域之新的學門—化學計量學 (Chemometrics)，綜合利用數學、統計、或邏輯方法設計最佳實驗程序，及分析實驗數據，以期獲得最多有用之化學資訊。此學門經約 30 年之發展，成效非常卓著。本次出國研習之主題為化學計量學中之實驗設計法部分—混合物實驗設計及數據分析 (Mixture Experimental Design)。

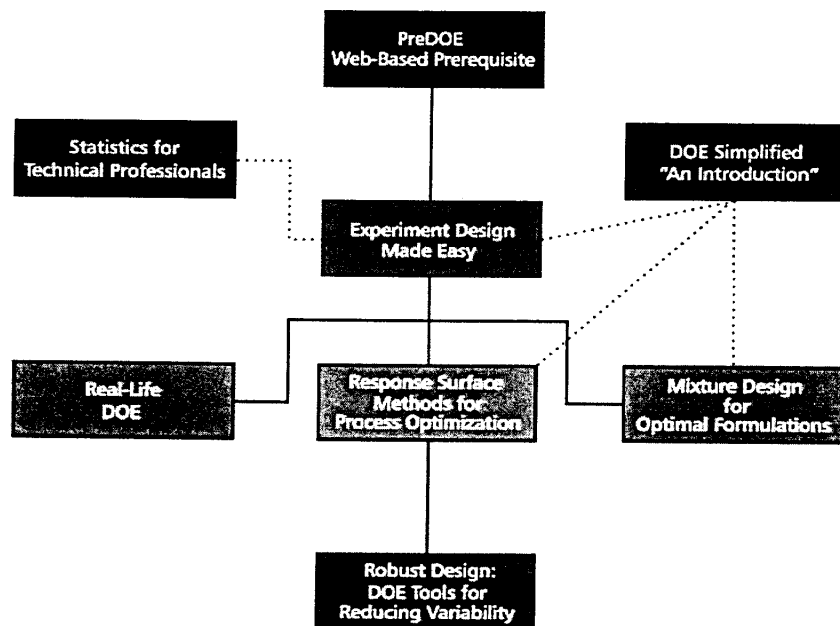
混合物實驗設計（配方實驗設計）是一種特殊的實驗設計，其反應值 Y 與組成成分間之比例 X_i 有關而與組成成分之絕對量無關。這類混合物實驗問題在工業、農業和科學試驗中屢見不鮮，例如汽油、潤滑油、混凝土、塑膠、合金、陶瓷、油漆、食品、醫藥、洗滌劑、混紡纖維...等產品都會遇到混合物實驗設計問題。本次研習之主要目的即學習混合物實驗設計技術，希望能以最少的實驗找出最佳的配方。

二、行程

日 期	工 作 紀 要	地 點
11/8	起程	台北—洛杉磯
11/9	起程	洛杉磯—明尼亞波里斯
11/10	參訪 Stat-Ease 公司	明尼亞波里斯
11/11~11/13	參加 “Mixture Designs for Optimal Formulations” 課程	明尼亞波里斯
11/14~11/15	返程	明尼亞波里斯—台北

三、Stat-Ease 公司簡介

Stat-Ease 是一家 15 人之小公司，成立約 10 年，提供實驗設計軟體、書籍、訓練及諮詢服務...等。該公司之核心技術為應用統計技術於改善產品品質、發展有效製程、迅速解決生產問題、進行突破性之研發等。他們發展之實驗設計軟體 Design-Expert 功能齊全，使用簡單，廣受好評，為許多近代實驗設計書籍列為參考軟體。此公司每年定期舉辦 20 多場次之實驗設計訓練班如下圖，本次參加之班次為圖中之”Mixture Design for Optimal Formulation”。該公司之刊物 Stateaser、DOE FAQ Alert 以及網站 www.statease.com 提供許多非常實用之資訊。



四、研習課程內容

Section 1 -- Introduction to Mixtures (Day 1)

- What makes a mixture?
- Mixture (Scheffe) polynomials
- Mixture tutorial
- Gold simulation

Section 2 -- Simplex Design (Day 1)

- Simplex-Lattice designs
- Augmenting simplex designs
- Interpreting model coefficients
- Design evaluation
- Simplex-Centroid designs
- Model reduction

Section 3 -- Constrained Mixtures, Simplex (Day 2)

- Mixture tutorial revisited: Actual - Real - Pseudo
- Optimization of multiple responses
- Propagation of error (POE)
- Cement optimization
- ABS Pipe: Model reduction & optimization
- Homework assignment

Section 4 -- Constrained Mixtures, Extreme Vertices (Day 2)

- Constrained mixtures, extreme vertices: Shampoo
- Algorithmic point selection: Flare
- Simulation of extreme vertices: Fruit punch
- Ratio constraints: Stability
- An additional equality constraint

Section 5 -- Combining Mixture and Process variables (Day3)

- User defined cross designs: Fish
- D-optimal crossed designs: Fish and Cure time
- Mixture amount experiments; Ibuprofen
- Mixtures with categorical factors: Paint

Section 6 -- Screening Components (Day3)

- Simplex: Additive package
- Extreme vertices: Snee eight ingredients
- Blue Haze

Section 7-- Application & Review (Day3)

- Homework
- Optional review exercises

Section 8 -- Appendix (Day3)

- Using ratios
- Slack variable polynomials
- Explanation of optimization techniques
- Piepel's vs. Cox's direction

五、研習心得

混合物實驗設計法就是依據統計原理設計最佳之實驗配置，進行實驗，然後由實驗結果應用迴歸分析配適(fit)一適當之實驗模式 $Y=f(X_i)$ ；由此模式即可瞭解組成變數 X_i 對反應變數 Y 之影響情形；由此模式可進一步找尋最佳配方。因此混合物實驗設計法之主要內容為(1)有哪些適用之數學模式，(2)如何配置最少之實驗點以配適上述之數學模式，(3)如何檢定配適之數學模式為適當可用之模式，(4)如何找出最佳之配方等。茲將研習之心得整理如下：

(一). 混合物實驗設計之數學模式：

在混合物試驗中，每個組成的貢獻都以成分比例表示。每個成分的比例必須是正的，而且它們的總和必須是 1，這就決定了混合物之迴歸設計是一種受特殊約束的迴歸設計問題。用 Y 表示反應值， X_1 、 X_2 、……、 X_n 表示混合物中 n 種成分各占的百分比，混合物迴歸設計就是要在混合物條件：

$$X_i \geq 0 \quad (i=1, 2, \dots, n) \quad \text{及}$$

$$X_1 + X_2 + \dots + X_n = 1 \quad \dots\dots\dots(1)$$

或者除混合物條件(1)之外，再加上一些其他條件的約束下，合理地安排試驗，求出關於 Y 的回歸方程，使得分析容易、準確，便於計算和推測最佳混合物比率。

由於混合物條件(1)的限制，混合物回歸設計不能採用一般的多項式模式，否則會引起矩陣退化而無解。通常依混合物組成變數 X_i 與反應變數 $Y(\eta)$ 關係之複雜程度，可選擇一次、二次、或三次 Scheffe 多項式模式。Scheffe 多項式如下(其中 q 代表成份數， η 反應變數， β_i 係數， X_i 組成變數)：

$$\text{Linear}(1^{\text{st}} \text{ order}): \eta = \sum_{i=1}^q \beta_i x_i$$

$$\text{Quadratic}(2^{\text{nd}} \text{ order}): \eta = \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \sum_{j=2}^q \beta_{ij} x_i x_j$$

$$\text{Special Cubic(reduced } 3^{\text{rd}} \text{ order}):$$

$$\eta = \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \sum_{j=2}^q \beta_{ij} x_i x_j + \sum_{i < j} \sum_{j < k} \sum_{k=3}^q \beta_{ijk} x_i x_j x_k$$

$$\text{Full Cubic}(3^{\text{rd}} \text{ order}):$$

$$\eta = \sum_{i=1}^q \beta_i x_i + \sum_{i < j} \sum_{j=2}^q \beta_{ij} x_i x_j + \sum_{i < j} \sum_{j=2}^q \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i < j} \sum_{j < k} \sum_{k=3}^q \beta_{ijk} x_i x_j x_k$$

Sheffe Model 之項次數（係數之數目）G 隨成分數 q 之增加而增加，其關係可整理如下表：

Number of Components(q)	Model(G)			
	Linear	Quadratic	Special cubic	Cubic
2	2	3		
3	3	6	7	10
4	4	10	14	20
5	5	15	25	35
6	6	21	41	56
7	7	28	63	84
8	8	36	92	120
9	9	45	129	165
10	10	55	175	220

混合物實驗之不同實驗配置之數目必須大於上表之數目才可解出 Model 之係數，另外為了檢定 Model 之適用性，應額外加上最少 3 個不同實驗點數(Min. 3 lack-of-fit df)及 4 個重複實驗點 (Min. 4 pure error df)。故適當之實驗數目為上表之數字加 6~7。

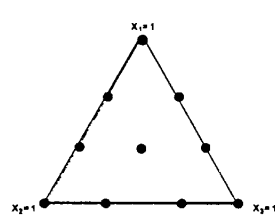
(二). 混合物實驗設計之種類：

依實驗狀況之不同，為了配適(fit)上述之 Scheffe 模式，有許多實驗設計可供選擇：

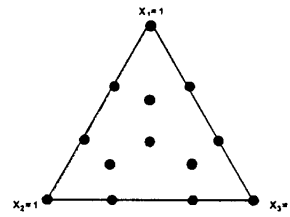
1. Simplex Design
 - Simplex-Lattice Design
 - Augmenting Simplex Design
 - Simplex-Centroid Design
2. Constrained Mixture Design
 - Simplex
 - Extreme Vertices
3. Combining Mixture and Process Variables
 - Mixture-Process Designs

- Mixture-Amount Designs
 - Mixture with Categorical Factors
4. Screening Components
- Simplex Screening Designs
 - Extreme Vertices Screening Design

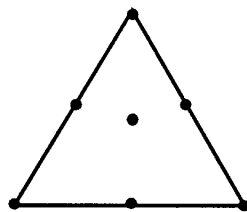
以三成分之混合物實驗為例，各種實驗配置如下圖：



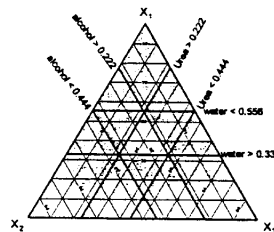
Simplex Lattice



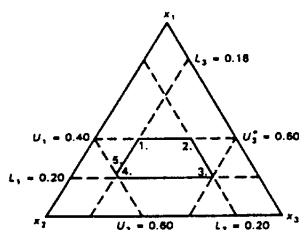
Simplex Lattice Augmented



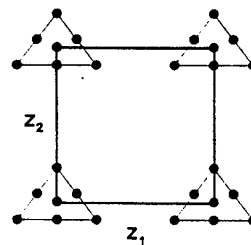
Simplex Centroid



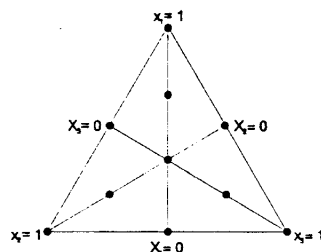
Constrained: Simplex



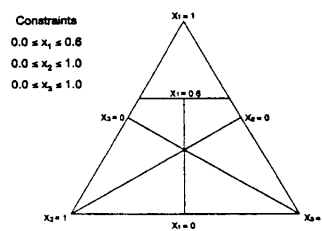
Constrained: Extreme Vertices



Mixture-Process Design



Simplex Screening Design



Extreme Vertices Screening Design

(三). 如何選擇適當之混合物實驗設計？

1. 單純型設計（格子設計或中心點設計）[Simplex designs (lattice or centroid)]
 - 適用於沒有限制條件或有限制條件但實驗範圍仍形成單純型(simplex)
 - 所有組成分之範圍均相同
 - ◆ 增加查核實驗點及中心實驗點
 - ◆ 增加重複實驗點以便計算實驗誤差(pure error)
2. 極端頂點設計(Extreme vertices designs)
 - D-最適設計
 - ◆ 組成分之範圍不相同，或複雜之實驗範圍
 - ◆ 增加查核實驗點及重複實驗點
 - 基於距離之設計
 - ◆ 增加重複實驗點
 - 自行設計
 - ◆ 自己選擇實驗點
3. 篩選設計
 - 選擇能配適一次線性模式之設計以找出最大線性效果之成分
 - 組成分超過 6 個時應先將重要成分篩選出來
 - 如果實驗範圍形成一單純型時使用單純型篩選設計，如否則應使用極端頂點篩選設計。
4. 含混合物組成變數及製程變數之實驗設計
先分別選定混合物組成變數及製程變數之實驗模式，然後應用 D-Optimal 選擇最佳實驗點，並另增加增加查核實驗點及重複實驗點。

(四). 混合物實驗設計數據分析

1. 選擇適當之數學模式
 - 避免選擇變數交混在一起(aliasing)之模式

- 依序模式平方和 (Sequential Model Sum of Squares)：當加入之項為顯著($P < 0.05$)，且非交混模式則應選擇最高次之多項模式
- 缺適度檢定 (Lack of Fit Test)：選擇之模式其缺適度檢定結果應不顯著($P > 0.10$)
- 缺適度測試需
 - ◆ 額外的實驗點 (多於估計模式係數之實驗點)，以計算這些實驗點之結果與模式之差異
 - ◆ 重複實驗點以便計算實驗誤差

2. 查驗選擇之模式

- 檢視 ANOVA 結果，模式之 F-檢定結果應為顯著($P < 0.05$)，而缺適度檢定結果應不顯著($P > 0.10$)
- 檢視迴歸係數之 F-檢定結果， $P > 0.10$ 之係數可考慮去除，以簡化模式，但應維持變數之完整階層性(hierarchy)。
- 檢查信號雜訊比是否足夠，即“Adeq Precision”是否大於 4

3. 應用殘差分析及診斷圖以檢視 ANOVA 之假設是否合理

4. 應用 Trace Plot 及 Contour plot 察看在實驗範圍內反應值隨組成變數改變之情形

(五). 找尋最佳配方

在此次研習中主要是以模糊邏輯方法求解多目標之最優化配方，此法先將所要探討之反應值 Y_i 先轉換為 0~1 間之歸屬函數 (Membership Function)，例如目標為極大值 (越大越好) 之轉換式如下：

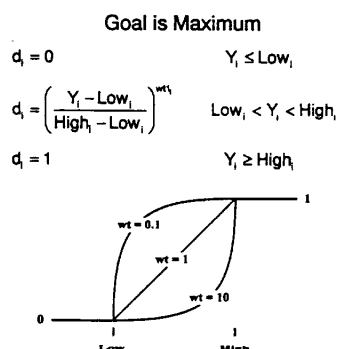
$$m(Y_i) = \begin{cases} 0, & Y_i \leq Y_{\min} \\ \left[\frac{Y_i - Y_{\min}}{Y_{\max} - Y_{\min}} \right]^R, & Y_{\min} < Y_i < Y_{\max} \\ 1, & Y_i \geq Y_{\max} \end{cases}$$

上式中 $Y_{\min}$ 為可接受之最小值，

$Y_{\max}$ 為當 Y_i 大於此值時並無好處，

R 為一常數，通常為 0.1 至 10 間之數值。

如果某一特性值 Y_i ，愈大愈好，則選定 $Y_{\min}$ 及 $Y_{\max}$ 後，即可依上式轉換為 0~1 間之數值，如下圖。目標值為極小值，或特定值時亦有相對應之轉換公式。



在多目標最優化之問題中，每一特性值 Y_i 可得一 $m(Y_i)$ ，所有特性質之總合歸屬值 $M(Y)$ 可以下式計算：

$$M(Y) = \frac{\sum_{i=1}^I \beta_i \times m(Y_i)}{\sum_{i=1}^I \beta_i} \quad , \text{ if all } m(Y_i) > 0$$

$$M(Y) = 0 \quad , \text{ if any } m(Y_i) = 0$$

上式中： $M(Y)$ 為所有特性值之歸屬值

$m(Y)$ 為個別特性值之歸屬值

β_i 為特性值 i 之權重

i 為特性值之個數。

模糊邏輯最適方法，即改變自變數使 $M(Y)$ 值最大。

求解最佳配方之程序如下：

1. 針對每一不同之反應值建立一適當之實驗模式，此實驗模式應能適當地反應實際情形。檢查下列各項：
 - (1). 模式之 F 檢定顯著，即大的 F 值其 $p < 0.05$
 - (2). 缺適度檢定結果不顯著，即 F 值接近 1 其 $p > 0.10$
 - (3). 殘差正常
 - (4). 適當之精密度，即大於 4
2. 設定最適化之準則
 - (1). 針對每一反應變數設定一目標
(最大、最小、目標值、範圍)
 - (2). 對每一目標設定高、低限值
 - (3). 必要時對不同的目標可依其相對重要性給予不同之權重
3. 進行求解，由數個可能的解中選出最優解
4. 觀察最優結果圖

六、感想

- (一). 本次到實驗設計法軟體公司參加"Mixture Design for Optimal Formulations Workshop"，經由專家有系統及深入的介紹，及許多實例的探討，受益良多，也解決過去的一些疑問。
- (二). 研習所學可應用於各種配方之開發及優化上，如汽油配方、潤滑油配方、觸媒配方...等。
- (三). 現在的企業要有競爭力必須(1) 和別人做同樣的事但比別人做得更好（與人競爭），或(2)和別人做不同的事，提供獨佔性的價值或整合性價值（不和別人競爭）。Stat-Ease 是一家小公司，能在競爭劇烈之市場存活下來所依賴的是統計應用之核心技術。與他們接觸，感覺上他們似乎知道所有技術細節，這或許是專家和一般人不同之處吧。
- (四). 最近非常流行之 6 Sigma 管理，應用非常多之統計技術（包括實驗設計法），這些統計技術正是 Stat-Ease 公司之專長，因此他們不但不受不景氣影響，反而業務大量增加。本公司及本所如要進一步降低成本及改善品質應多使用事半功倍之統計工具，如實驗設計法。
- (五). 研習資料：(1)Mixture Design for Optimal Formulations 講義及(2)Handbook for Experimenters 將複製一份存於本所圖書館供參考。
- (六). 本次能出國研習增長見聞，非常感謝公司及煉研所何組長、副所長、所長。