



Improving forecasting performance by employing the Taguchi method

出處:European Journal of Operational
Research 176(2007)

作者: Tai-Yue Wang, Chien-Yu Huang

指導老師:童超塵 老師

報告人:卓益如

- 
- Abstract
 - Introduction
 - Methodology
 - An illustrated example
 - Conclusion

Abstract

- 為因應需求多變性的市場，決策者必須有一套精準的預測模型以降低開發前置時間。
- 此論文採取田口方法校準預測模型的可控制因子，找出影響較大的因子。

Introduction (I)

- 預測執行視 horizon length(水平時間範圍)、開發成本、data period(數據週期)、校正次數(頻率)、應用種類、模型辨識能力以及觀測值大小(Delurgio, 1998)而定。但是必須更深層研究如何找出關鍵可控制因子和計算彼此交互作用並且去證實其最佳水準設置(Radaev, 1998; Russell et al.,2000)才是目前的課題。

Introduction (II)

- 田口方法用來降低成本，改善品質 (Taguchi, 1986) 並且提供穩健設計之解決方法 (Khoei et al., 2002)。即使因子有交互作用 (interactions) 存在，利用田口方法仍能夠建立最佳設計架構，譬如 Casab (2003) 的酵素結合免疫吸附法 (ELISA) 的最佳化；或者 McMillan (1998) 運用田口方法設計兩階段篩選實驗決定顯著因子等。
- 使用雜訊比 (S/N) 和變異數分析 (ANOVA) 做主效應和交互作用的分析。

Introduction (III)

- 此論文的研究目標如下：
 1. 探討可控因子主效應和特定時間序列裡的交互作用。
 2. 將參數套用在預測模型上，選擇影響大的作為雜訊因素。
 3. 透過內外部直交表討論可控制因子和水準設置的最佳組合。
 4. 讓決策者能明確指出適當的數據週期、水平時間、需要的觀測值大小做資料收集，然後建立有效的預測模型。

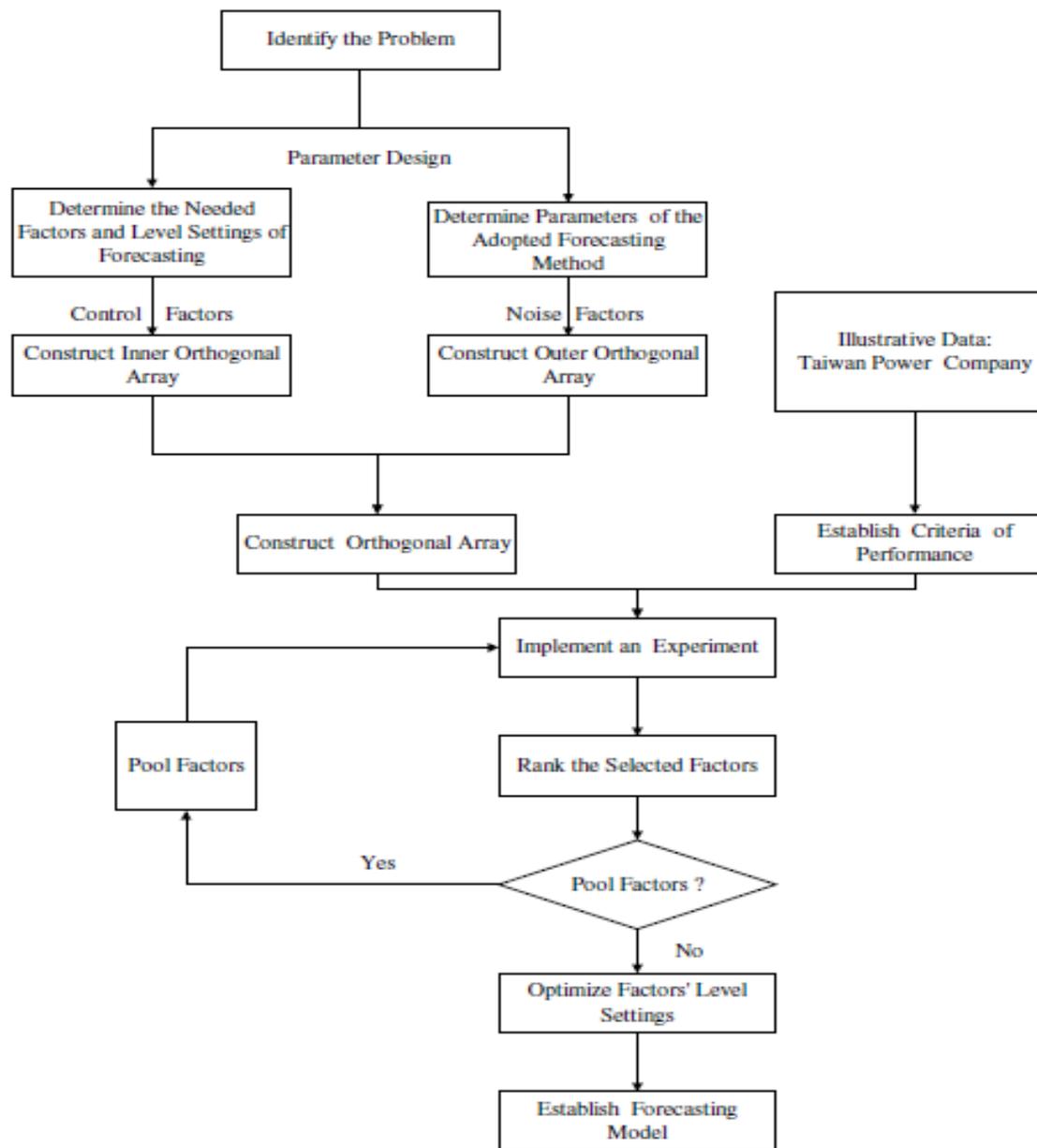


Fig. 1. Outline of present research.

Methodology (I)

1. Taguchi method

田口玄一於1960年發展出一套改善產品品質的有效方法。以品質損失函數所發展的信號噪訊比(signal to noise ratio，簡稱S/N比)作衡量品質的指標，並用直交表做實驗設計，可以用較少之實驗次數達到穩健參數設計的目的。使得可以較低的成本而獲得較高的產品品質。

Methodology (II)

2. 決定可控制因子和水準設置以建立內部直交表
3. 決定外部直交表的參數

田口將控制因素和雜訊因子以內部陣列和外部陣列分開

4. 利用田口方法評估可控因子和水準

信號雜訊比(S/N)是將許多重覆之品質特性值整合成一個反應變異存在大小值；其值越大越好。

$$S/N = -10 \log_{10}(MSD)$$

$$MSD = (y_1^2 + y_2^2 + y_3^2 + \dots) / n,$$

y_1, y_2, \dots 為試驗結果

n 為重覆次數

5. Ranking of selected factors

利用ANOM (analysis of mean)篩選因子和排序因子。

6. Pooling of factors

將不顯著的因子合併至誤差項

7. Validation of experiment

利用指標「平均絕對誤差率」做最後確認

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \bar{x}_i}{x_i} \right| \times 100 \%$$

An illustrated example

- 選擇台灣電力公司為分析對象，蒐集1987年1月至2001年12月的資料，並且以BPN(倒傳遞網路模型)預測模型為基礎。
- Data collection

Table 2

Status of Taiwan Power Company at year-ends 1998–2002

Description	1998	1999	2000	2001	2002	Average growth rate (%)
Installed nameplate capacity (MW)	26,680	28,480	29,634	30,136	31,915	4.4
Energy production (million KWH)	1430	1458	1565	1581	1659	4.6
Sales (million KWH)	1281	1317	1424	1436	1512	5.0

■ Determine level settings of controllable factor

Table 3

Parameter level settings for forecasting model (month)

Name of parameter	Level settings (month)
Data period	1, 3, 6
Horizon length	12, 24, 36
Number of observations required	96, 120, 144

因為單位為月，
可以將資料劃分為以月、季、
半年來作數據期間的水準

■ Determine BPN's topology and parameters for outer orthogonal array

* 考慮BPN拓樸和參數為雜訊因子(外部直交表)

Table 4
Noise factor (BPN's topology and parameter) and level

Variables	Level 1	Level 2
A: number of hidden layers	1 or 2	3 or 4
B: number of neurons in hidden layer	Number of input neuron \times 1.0	Number of input neuron \times 1.5
C: learning rate	\square 0.5	$>$ 0.5

Table 5
 $L_4(2^3)$ outer orthogonal arrays for BPN noise factors

Experiment/column	L_4		
	A	B	C
	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

A: number of hidden layers; B: number of neurons in hidden layer; C: learning rate.

■ 符號說明

f	degree of freedom
S	sum of squares
V	mean squares (variance)
F	variance ratio
P	percent contribution

■ ANOVA table

Table 6

Analysis of the time series related factors experiment (ANOVA)

Column	Factors	f	S	V	P
1	Factor A	2	6.8718	3.4359	3.37
2	Factor B	2	185.8200	92.9110	91.18
3	Interaction A × B	2	0.9073	0.4537	0.45
4	Factor C	2	10.1890	5.0945	5.00
All other/error		0	0	0	0
Total		8	203.7881		100%

顯著

不顯著，列入誤差項

Table 7
 Analysis of the time series related factors experiment (pooled ANOVA)

Column	Factors	f	S	V	F	P
1	Factor A	(2)	(6.8718)	Pooled		
2	Factor B	2	185.8200	92.9110	31.0250	88.24
3	Interaction A \times B	(2)	(0.9073)	Pooled		
4	Factor C	(2)	(10.1890)	Pooled		
All other/error		6	17.9681	2.9947		11.76
Total		8	203.7881			100%

- 使用ANOM分析，分別計算三因子各水準之平均S/N。

$$\bar{A}_1 = \frac{(SN_1 + SN_2 + SN_3)}{3} = \frac{(26.821 + 20.063 + 14.869)}{3} = 20.584$$

Table 8

Analysis of the time series related factors experiment (main effect)

Factor/level	Level 1	Level 2	Level 3
Data period (A)	20.584	21.830	22.715
Horizon length (B)	27.838	20.320	16.971
Interaction (AB)	21.533	21.440	22.155
Number of observation required (C)	21.994	22.847	20.288

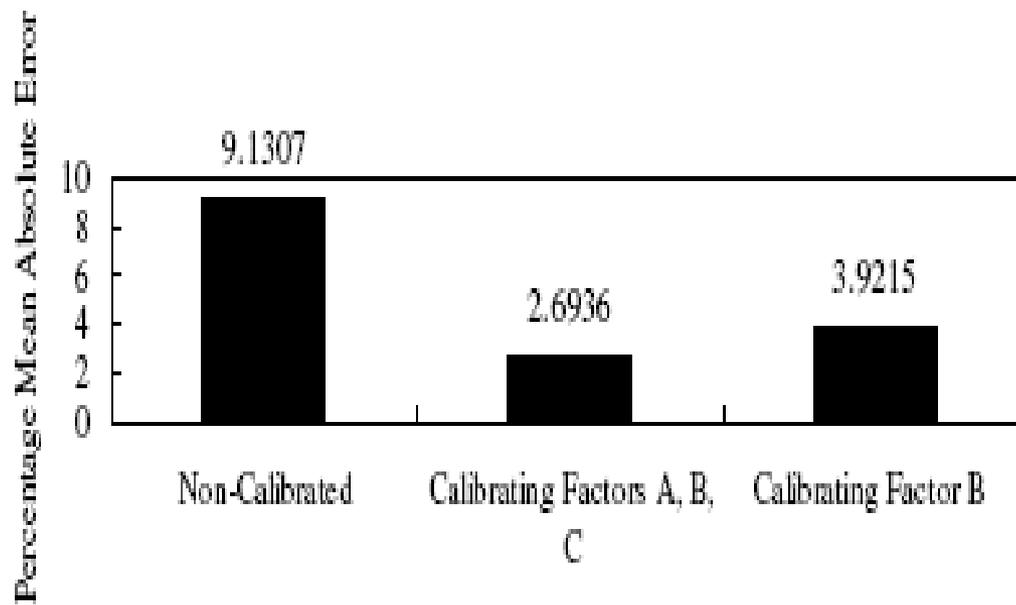
Factor A—Level 3

Factor B—Level 1

Factor C—Level 2

- 故最佳水準設置即為:data period 為6個月；horizon length 為12個月；需要觀測值大小為120個月。

■ Validation of experiment



[table1](#)

Fig. 2. Comparison on forecasting performance of different controllable factors.

Conclusion

- 台電的例子說明，即使大量資料裡有很多不顯著的可控制因子，但是田口方法還是能提供預測的正確性。
- 因為同時考慮內部與外部直交表的結果，能使決策者快速處理及有效地蒐集資料並且做出競爭性策略。
- 可將此研究擴展到其他生產應用的議題上。

$L_9(3^4)$ inner orthogonal array for forecasting model (BPN)

Experiment/column	L_9				Results					S/N _i
	A	B	AB	C	Y ₁	Y ₂	Y ₃	Y ₄	\bar{Y}	
	1	2	3	4						
1	1	1	1	1	4.9129	4.7203	4.1752	4.3961	4.5511	26.821
2	1	2	2	2	10.6529	9.3897	9.5462	10.073	9.9154	20.063
3	1	3	3	3	13.5263	27.397	15.3083	11.6487	16.9701	14.869
4	2	1	2	3	5.0397	2.2619	7.1054	3.6721	4.5198	26.267
5	2	2	3	1	7.5895	7.4801	10.1366	9.4406	8.6617	21.171
6	2	3	1	2	10.4738	12.2062	13.5069	13.6117	12.4497	18.052
7	3	1	3	2	1.8902	0.9949	4.4845	3.4048	2.6936	30.426
8	3	2	1	3	7.5483	8.074	14.7624	9.2666	9.9128	19.727
9	3	3	2	1	10.3526	14.4084	13.5689	11.6791	12.5022	17.991

A: Data period used, B: Horizon length, C: Number of observations required.

$$4.5511 + 4.5198 + 2.6936 / 3 = 3.9215$$

