Failure Amplification Method: An Information Maximization Approach to Categorical Response Optimization

V. Roshan Joseph and C. F. Jeff Wu

School of Industrial and Systems Engineering

Georgia Institute of Technology

Atlanta, GA 30332-0205, USA

Abstract

Categorical data arise quite often in industrial experiments because of an expensive or inadequate measurement system for obtaining continuous data. When the failure probability/defect rate is small, experiments with categorical data provide little information regarding the effect of factors of interests and are generally not useful for product/process optimization. We propose an engineering-statistical framework for categorical response optimization that overcomes the inherent problems associated with categorical data. The basic idea is to select a factor that has a known effect on the response and use it to amplify the failure probability so as to maximize the information in the experiment. New modeling and optimization methods are developed. It is illustrated with two real experiments.

KEY WORDS: Accelerated Life Testing, Operating Window Method, Process Capability, Quality Engineering, Robust Parameter Design.

1. INTRODUCTION

It is a well known fact that continuous data contain more information than categorical data for assessing the performance of a system. Yet in many situations, experimenters have to be content with categorical data due to an inadequate or expensive measurement system for obtaining continuous data. Categorical data possess several advantages. First, they are easy to measure and record. Second, they can be easily converted into monetary units and therefore are useful for managerial decision making and for communications among non-statisticians. But they also introduce new challenges and complications for data analysis that are not present in continuous data. The accumulation analysis introduced by Taguchi for categorical data was quite popular in industries for sometime until it was proved to be faulty by Nair (1986) and Hamada and Wu (1990). Generalized linear modeling is the most accepted and well received approach for categorical data analysis. See, for example, McCullagh and Nelder (1989) and Wu and Hamada (2000).

Categorical data possess serious problems in terms of effect estimation and system optimization compared to using continuous data. To illustrate this point, consider a 2^{3-1}_{III} experiment on an electro-plating process with 3 control factors: temperature (x_1), pH (x_2), and current (x_3). The plated thickness inside 5 holes is measured and given in Table 1. Data analysis gives the fitted model for the thickness $\hat{y} = 71.8 + 2.2x_1 - 4.3x_2 + 8.7x_3$, which can be used for optimization. Now suppose the experimenter instead used a go/no-go measurement system designed for a specification 75 ± 25 microns. We see that all the holes conform to the specification, thus resulting in zero defectives for the 4 runs. Based on these categorical data, we cannot estimate the effect of the three factors and therefore the outcome of this experiment is inconclusive. Needless to say, no sophisticated statistical analysis can rescue this experiment. This situation is a consequence of low failure probability/defect rate for which the data provide little information. This is an inherent problem associated with categorical data. We propose in this paper an engineering-statistical approach to overcome it.

The basic idea is to select a factor with known effect on the failures based on physical knowledge of the product/process. This factor can then be used to amplify (or excite) the failure probability so as to *maximize the information* in the experiment. We call this factor an *amplification factor* and this approach the *Failure Amplification Method* (FAMe).

For example, in the electro-plating process the plating time can be used as the amplification

Table 1: Data of the electro-plating example

run	x_1	x_2	x_3	thickness	under-plating	good	over-plating
1	$-$	$-$	$+$	86 79 86 84 78	0	5	0
2	$-$	$+$	$-$	53 51 57 64 55	0	5	0
3	$+$	$-$	$-$	66 70 69 67 76	0	5	0
4	$+$	$+$	$+$	79 87 75 74 77	0	5	0

factor. From Faraday's law we know that the plating thickness is proportional to the plating time. Thus, decreasing time from its current value will produce defective holes due to under-plating (thickness $<$ 50 microns) and increasing time will produce defective holes due to over-plating (thickness $>$ 100 microns). With this amplification of the defective rate, the experiment will produce defective holes and the experimenter can hope to draw some conclusions regarding the effects of factors of interest. This method differs from the traditional approach because of the introduction of the amplification factor. It is a factor of no direct interest to the experimenter as its effect on the thickness is already known. In the traditional approach this factor is not studied and the experiment is performed by keeping it at its current value. In FAMe such a factor is identified and varied in the experiment so as to maximize the information. The amplification factor should have a large effect on the failures, so that other factors in the system will not be able to reverse its effect. This will ensure that the estimated model at the amplified conditions remains the same at the user conditions.

Failure amplification can be achieved in several ways. The amplification factor can be chosen from the set of control factors, the set of noise factors, or other types of factors. The choice is often determined by the ease in conducting the experiments. Depending on the type of amplification factor, we classify FAMe into three types: (i) control factor method, (ii) complexity factor method, and (iii) noise factor method (see Figure 1). As the name suggests, in a control factor method we use a control factor as the amplification factor. The plating process described above is a good example because the plating time is a control factor as it can be set at any value the manufacturer wishes. A *complexity factor* is a factor that determines how complex a product is to manufacture. Extreme settings can lead to high rejections/reworks, making the product difficult to manufacture.

For example, hole diameter in a printed circuit board (PCB) is a complexity factor. PCBs with small hole diameters lead to high rejections in drilling as well as in hot air leveling due to blocked holes. The complexity factor differs from a control factor as it is specified by the customer and hence not in the control of the manufacturer. Noise factors are factors that are not in the control of the manufacturer/user but some of which can be systematically varied in the experiment. For example, in drilling holes the life of the drill bit (new to worn out) is a noise factor and can be used to amplify the failures. In the next two sections we explain the different modeling and optimization strategies that can be adopted for the control factor and complexity factor methods. We will not discuss the noise factor method as we have not encountered it in any real experiments, but we note that this is a feasible option for failure amplification.

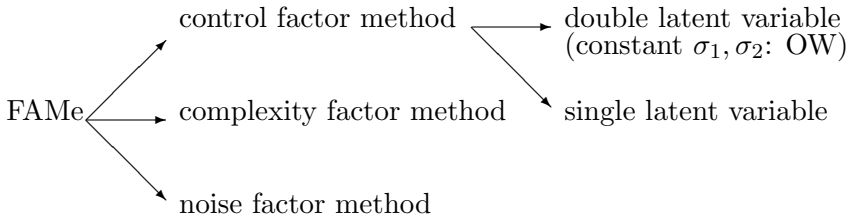

Figure 1: Classification of FAMe

The article is organized as follows. In Section 2 modeling and optimization methods are developed for the control factor method. Connections of this approach with the operating window method (Clausing, 1994) are established. A paper feeder experiment from Fuji-Xerox is used to illustrate the approach. In Section 3 an experiment from PCB manufacturing is used to explain the complexity factor method. Experimental strategies for failure amplification are discussed in Section 4. In Section 5 the efficiency of FAMe is compared with that of traditional experiments . Further discussions and concluding remarks are given in Section 6.

2. CONTROL FACTOR METHOD

Consider the paper feeder in a copier machine. It has two failure modes: misfeed (the feeder fails to feed a sheet) and multifeed (the feeder feeds more than one sheet of paper). Here the stack force can be used as an amplification factor. Since the designer can set the stack force at any

desired value, it is a control factor. Its effect on the two failure modes is known. A low stack force leads to misfeed, whereas a high stack force leads to multifeed.

Let M be the amplification factor, p_1 the probability of failure due to failure type 1 (say misfeed), and p_2 the probability of failure due to failure type 2 (say multifeed). Assume p_1 to be a strictly decreasing function in M and p_2 a strictly increasing function in M. A typical behavior is shown in Figure 2. Note that if M does not have a conflicting effect on p_1 and p_2, then M could be adjusted to a low value or a high value to eliminate both failures. Thus, to apply the control factor method, M must have conflicting effects on at least two failure modes.

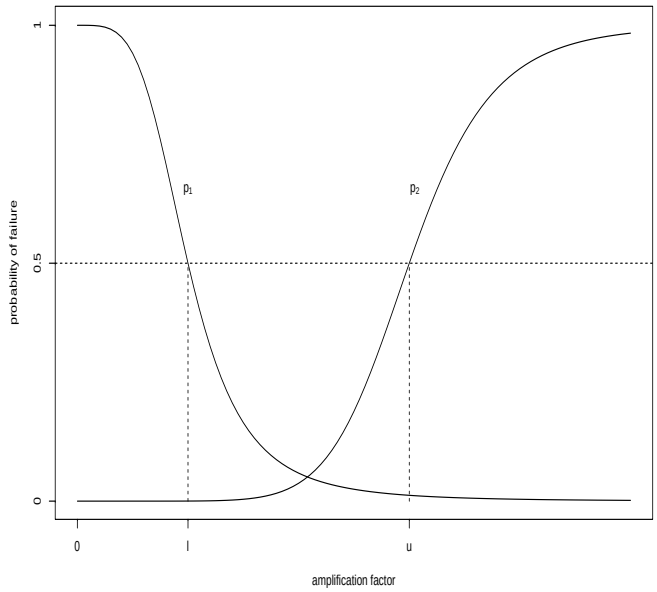

Figure 2: Probability of failure curves against amplification factor

We will now describe an experiment carried out at Fuji-Xerox by Y. Norio and O. Akira. These data were also analyzed by Miyakawa (1993). There were 8 control factors and one noise factor. Their levels are shown in Table 2. The control array (i.e., array for the control factors) is obtained by modifying the $OA(18, 2^1 \times 3^7)$ in Table 5, by coding level 3 as level 1 in column x_4 (so that x_4 can be used for the two-level factor "center roll"). The data on misfeed and multifeed are given in Table 3. The data give the number of misfeeds and multifeeds out of 5 sheets at different stack force levels. For example, in run 1 at the first noise level, there are 5 misfeeds at $M = 20$, 5 misfeeds at $M = 40$, 1 misfeed at $M = 42.5$, and so on. The stack force, which is the amplification factor, is

Table 2: Factors and levels for the paper feeder experiment

Control factors	Notation	Levels		
		1	2	3
Feed belt material	x_1	Type A	Type B	-
Speed	x_2	288 mm/s	240 mm/s	192 mm/s
Drop height	x_3	3 mm	2 mm	1 mm
Center roll	x_4	Absent	Present	-
Belt width	x_5	10 mm	20 mm	30 mm
Tray guidance angle	x_6	0	14	28
Tip angle	x_7	0	3.5	7
Turf	x_8	None	1 sheet	2 sheets
Noise factor				
Stack quantity	N	High	Low	-

varied sequentially to produce about 50% of failures, which is about 2 or 3 failures. The objective of this experiment is to find control factor settings to minimize both misfeeds and multifeeds. We will come back to the analysis of this experiment in Section 2.4 after developing the necessary modeling and optimization methods.

We will use a latent variable approach to model the failure probabilities p_1 and p_2. In this approach the failures can be thought of as extremities of certain (unobserved) functional characteristics. We only observe categorical data on failures because of the inability to measure these functional characteristics. Sometimes it is also difficult to identify the right functional characteristics corresponding to the failure modes under study. This suggests that we can treat the *functional characteristics as latent variables* and infer about their properties indirectly from the categorical data. Two situations are considered in the following subsections: two latent variables or a single latent variable generating the two failures.

2.1 Double Latent Variable Modeling

Let Y_1 and Y_2 be the two latent variables corresponding to the two failure modes. Suppose the

Table 3: Data from the paper feeder experiment

	misfeed												multifeed											
Run	N_1						N_2						N_1						N_2					
1	20	40	42.5	45	50	60	20	30	40	50	60	70	80	82.5	85	90	120	160	60	62.5	65	70	80	90
	5	5	1	0	0	0	5	0	0	0	0	0	0	2	2	2	2	2	0	1	1	3	2	3
2	0	10	15	20	30	40	0	10	15	20	40	60	30	35	40	50	60		30	40	60	70	75	80
	5	3	0	0	0	0	5	3	0	0	0	0	0	1	3	3	3		0	1	1	1	2	2
3	0	10	15	20	25		0	10	15	20	30	40	20	25	30	40			20	30	35	40	0	
	5	5	1	1	0		5	3	2	0	0	0	0	2	2	2			0	1	1	3	3	
4	20	25	30	40	60		0	20	25	30	40		50	60	65	70	80		40	50	55	60		
	5	3	1	0	0		5	5	1	0	0		0	1	2	2	2		0	0	2	2		
5	20	25	30	40	50		20	25	30	40	50		30	40	45	50	60		40	50	55	60		
	4	1	0	0	0		4	1	0	0	0		0	1	3	3	3		0	0	2	2		
6	10	15	20	30	40		10	15	20	30	40		30	40	45	50			30	40	50	55	60	
	4	2	1	0	0		3	0	0	0	0		0	1	2	3			0	1	2	2	3	
7	10	20	30	35	40		10	20	25	30	40		20	30	35	40	50		20	30	40	60	70	80
	5	4	2	1	0		5	3	0	0	0		0	1	2	2	3		0	1	1	1	2	2
8	15	20	30	35	40		20	30	35	40	60	70	70	80	100	110	120		60	70	75	80	100	
	3	2	2	3	0		5	2	4	1	1	0	0	1	1	2	2		0	1	2	2	2	
9	10	15	20	30	40		10	15	20	25	30	40	40	60	65	70	80		40	50	55	60	70	
	5	4	1	0	0		5	5	5	4	0	0	0	1	1	2	3		0	0	1	2	3	
10	0	5	10	15	20		0	5	10	15	20		5	10	15	20	30		0	5	15	20	30	
	5	1	0	0	0		5	0	0	0	0		0	1	1	3	3		0	1	0	2	2	
11	0	5	10	15	20		0	5	10	15	20		5	10	15	20	30		0	5	10	15	20	30
	5	2	0	0	0		5	1	0	0	0		0	1	1	4	3		0	1	1	0	2	2
12	0	10	15	20			0	10	15	20			30	40	50	55	60		40	45	50			
	5	4	0	0			5	4	0	0			0	1	1	2	5		0	0	1			
13	0	10	15	20			0	10	15	20			30	40	80	85	90	100	55	60				
	5	5	1	0			5	4	2	0			0	1	1	4	3	2	0	2				
14	10	20	25	30	35	40	10	20	25	30	35	40	20	30	35	40	45		20	25	30	35	40	50
	5	3	2	2	0	0	5	4	0	0	0	0	0	0	0	2	2		0	0	1	1	4	3
15	0	5	10	15			0	5	10	15	20		20	30	35	40	50		10	15	20	30		
	1	0	0	0			4	0	0	0	0		0	1	3	2	4		0	1	4	4		
16	5	10	20	30	35		5	10	20	30	40		20	30	35	40	50	60	30	40	50	55	60	
	5	1	0	0	0		5	1	0	0	0		0	1	0	2	3	5	0	1	1	2	2	
17	10	20	30	40	45	50	10	20	25	30	40		80	90	95	100			100	105	110			
	5	4	5	2	0	0	5	3	0	0	0		0	1	1	1			0	1	1			
18	10	15	20	30			10	20	30	35	40		60	65	70	80	90	120	60	70	75	80	90	
	5	5	1	0			5	5	5	0	0		0	1	2	2	2	3	0	1	2	2	3	

failures are generated by the following mechanism:

$$\text{if } Y_1 < T_1 \text{ then failure type 1; if } Y_2 > T_2 \text{ then failure type 2.} \tag{1}$$

Assume Y_1, Y_2, and M to be nonnegative variables taking values in $[0, \infty)$. Let $\mathbf{X}$ be the set of control factors and $\mathbf{Z}$ the set of noise factors. Consider the following models:

$$Y_1 = \beta_1(\mathbf{X}, \mathbf{Z})M \text{ and } Y_2 = \beta_2(\mathbf{X}, \mathbf{Z})M. \tag{2}$$

We will see that the above models together with the failure generating mechanism in (1) produce the probability of failure curves as shown in Figure 2.

Partition the set of noise factors as $\mathbf{Z} = \{\mathbf{N}, \mathbf{U}\}$, where $\mathbf{N}$ is the set of observed (or known) noise factors which can be systematically varied in the experiment, and $\mathbf{U}$ the set of unobserved (or unknown) noise factors. Let $\mu_i(\mathbf{X}, \mathbf{N}) = E_U[\log \beta_i(\mathbf{X}, \mathbf{Z})|\mathbf{N}]$ and $\sigma_i^2(\mathbf{X}, \mathbf{N}) = Var_U[\log \beta_i(\mathbf{X}, \mathbf{Z})|\mathbf{N}]$. Assuming a normal distribution for $\log Y_i$ for given $\mathbf{N}$, we have

$$\log Y_i|\mathbf{N} \sim N(\mu_i(\mathbf{X}, \mathbf{N}) + \log M,\ \sigma_i^2(\mathbf{X}, \mathbf{N})), \tag{3}$$

for $i = 1, 2$. We also assume Y_1 and Y_2 are independent for given $\mathbf{N}$.

From (1) we see that Y_1 is a larger-the-better characteristic and Y_2 a smaller-the-better characteristic. A loss function that can be used to evaluate the system based on the values of Y_1 and Y_2 is

$$L = \frac{c_1}{Y_1} + c_2 Y_2. \tag{4}$$

For a discussion and justification of this class of quality loss functions, see Joseph (2003). The expected loss is

$$\begin{aligned} E(L|\mathbf{N}) &= c_1 E(\frac{1}{Y_1}|\mathbf{N}) + c_2 E(Y_2|\mathbf{N}) \\ &= c_1 E(e^{-\log Y_1}|\mathbf{N}) + c_2 E(e^{\log Y_2}|\mathbf{N}) \\ &= c_1 e^{-\mu_1(\mathbf{X},\mathbf{N}) - \log M + \sigma_1^2(\mathbf{X},\mathbf{N})/2} + c_2 e^{\mu_2(\mathbf{X},\mathbf{N}) + \log M + \sigma_2^2(\mathbf{X},\mathbf{N})/2}. \end{aligned}$$

The above result is exact under the normal distribution assumption and can be shown to be a second order approximation with other distributions. By taking expectations with respect to $\mathbf{N}$,

$$\begin{aligned} EL &= E_N[E(L|\mathbf{N})] \\ &= \frac{c_1}{M} E_N(e^{-\mu_1(\mathbf{X},\mathbf{N}) + \sigma_1^2(\mathbf{X},\mathbf{N})/2}) + c_2 M E_N(e^{\mu_2(\mathbf{X},\mathbf{N}) + \sigma_2^2(\mathbf{X},\mathbf{N})/2}). \end{aligned}$$

Because the amplification factor M is a control factor, we can set it at any value we wish. Obviously we should choose a value that minimizes the expected loss. Equating the first derivative of EL with respect to M to zero, we get

$$M^* = \left\{ \frac{c_1 E_N(e^{-\mu_1(\mathbf{X},\mathbf{N})+\sigma_1^2(\mathbf{X},\mathbf{N})/2})}{c_2 E_N(e^{\mu_2(\mathbf{X},\mathbf{N})+\sigma_2^2(\mathbf{X},\mathbf{N})/2})} \right\}^{1/2}. \tag{5}$$

Because EL is strictly convex in $M > 0$, M^* minimizes EL. At this optimal setting for M, the expected loss is

$$EL^* = 2\left\{ c_1 c_2 E_N(e^{-\mu_1(\mathbf{X},\mathbf{N})+\sigma_1^2(\mathbf{X},\mathbf{N})/2}) E_N(e^{\mu_2(\mathbf{X},\mathbf{N})+\sigma_2^2(\mathbf{X},\mathbf{N})/2}) \right\}^{1/2}. \tag{6}$$

Letting D be the feasible region for $\mathbf{X}$, our objective is to find an $\mathbf{X} \in D$ to minimize the expected loss. Because EL^* is evaluated at the optimum setting M^* of M, it can be viewed as a performance measure independent of adjustment (PerMIA) with M being the adjustment parameter. For details on PerMIA, see Leon, Shoemaker, and Kacker (1987) and Leon and Wu (1992).

To give a nice interpretation of the performance measure to be discussed later, we need the following re-parameterization. From (1) and (3) we have

$$p_1(\mathbf{X},\mathbf{N},M) = P(Y_1 < T_1|\mathbf{N}) = \Phi\left(\frac{\log T_1 - \mu_1(\mathbf{X},\mathbf{N}) - \log M}{\sigma_1(\mathbf{X},\mathbf{N})} \right) \tag{7}$$

and

$$p_2(\mathbf{X},\mathbf{N},M) = P(Y_2 > T_2|\mathbf{N}) = \Phi\left(\frac{\mu_2(\mathbf{X},\mathbf{N}) + \log M - \log T_2}{\sigma_2(\mathbf{X},\mathbf{N})} \right), \tag{8}$$

where Φ is the standard normal distribution function. Note that p_1 is strictly decreasing in M and p_2 is strictly increasing in M.

Let l and u be the *lower* and *upper threshold* values of M that produce 50% failures for failure type 1 and failure type 2 respectively (see Figure 2). These are known as the median lethal dose (LD50) in bioassays. Thus,

$$p_1(\mathbf{X},\mathbf{N},l) = 0.5 \text{ and } p_2(\mathbf{X},\mathbf{N},u) = 0.5. \tag{9}$$

The threshold could be defined with respect to other failure probabilities, but in this case 50% is the most convenient one. From (7) and (8) we get

$$\mu_1(\mathbf{X},\mathbf{N}) = \log T_1 - \log l(\mathbf{X},\mathbf{N}) \text{ and } \mu_2(\mathbf{X},\mathbf{N}) = \log T_2 - \log u(\mathbf{X},\mathbf{N}).$$

Then (7) and (8) become

$$p_1(\mathbf{X},\mathbf{N},M)=\Phi\left(\frac{\log\{l(\mathbf{X},\mathbf{N})/M\}}{\sigma_1(\mathbf{X},\mathbf{N})}\right) \text{ and } p_2(\mathbf{X},\mathbf{N},M)=\Phi\left(\frac{\log\{M/u(\mathbf{X},\mathbf{N})\}}{\sigma_2(\mathbf{X},\mathbf{N})}\right), \tag{10}$$

and (6) becomes

$$EL^*=2\left\{\frac{c_1c_2T_2}{T_1}E_N(l(\mathbf{X},\mathbf{N})e^{\sigma_1^2(\mathbf{X},\mathbf{N})/2})E_N(\frac{e^{\sigma_2^2(\mathbf{X},\mathbf{N})/2})}{u(\mathbf{X},\mathbf{N})})\right\}^{1/2}.$$

Because c_1, c_2, T_1, and T_2 are given constants, an equivalent performance measure to minimize is

$$PM(\mathbf{X})=E_N(l(\mathbf{X},\mathbf{N})e^{\sigma_1^2(\mathbf{X},\mathbf{N})/2})E_N(\frac{e^{\sigma_2^2(\mathbf{X},\mathbf{N})/2}}{u(\mathbf{X},\mathbf{N})}). \tag{11}$$

Based on the theory of PerMIA (see references given above), we can employ the following *two-step optimization* procedure.

1. Find $\mathbf{X}^*\in D$ to minimize $PM(\mathbf{X})$ in (11).
2. Set $M=\left\{\frac{c_1E_N(l(\mathbf{X}^*,\mathbf{N})e^{\sigma_1^2(\mathbf{X}^*,\mathbf{N})/2})}{c_2T_1T_2E_N(e^{\sigma_2^2(\mathbf{X}^*,\mathbf{N})/2}/u(\mathbf{X}^*,\mathbf{N}))}\right\}^{1/2}$.

(12)

Note that we do not need c_1, c_2, T_1, and T_2 to find $\mathbf{X}^*$. But for the adjustment step we need to know the ratio $(c_1/T_1)/(c_2T_2)$. This ratio can be taken as 1 if we assume the loss due to failure type 1 is equal to the loss due to failure type 2.

The two-step procedure can only be implemented with the input of the experimental data. This brings us to the issue of estimation. Because Y_1 and Y_2 are unobserved, the parameters cannot be estimated from (3). They must be estimated indirectly from the failure data such as that given in Table 3. The sole purpose of the failure amplification was to improve the estimation by increasing the information in the data. Let y_1 and y_2 be the number of failures observed in a sample of size r for failure type 1 and 2, respectively. We have $y_i\sim binomial(r,p_i), i=1,2$. The equations in (10) can be written as

$$\Phi^{-1}(p_1)=a_1(\mathbf{X},\mathbf{N})+b_1(\mathbf{X},\mathbf{N})\log M \text{ and } \Phi^{-1}(p_2)=a_2(\mathbf{X},\mathbf{N})+b_2(\mathbf{X},\mathbf{N})\log M, \tag{13}$$

where

$$\begin{aligned} l(\mathbf{X},\mathbf{N}) &= \exp(-\frac{a_1(\mathbf{X},\mathbf{N})}{b_1(\mathbf{X},\mathbf{N})}),\ \sigma_1(\mathbf{X},\mathbf{N})=-\frac{1}{b_1(\mathbf{X},\mathbf{N})},\\ u(\mathbf{X},\mathbf{N}) &= \exp(-\frac{a_2(\mathbf{X},\mathbf{N})}{b_2(\mathbf{X},\mathbf{N})}),\ \sigma_2(\mathbf{X},\mathbf{N})=\frac{1}{b_2(\mathbf{X},\mathbf{N})}. \end{aligned} \tag{14}$$

Then express the functions a_1, a_2, b_1, and b_2 as linear models in $\mathbf{X}$ and $\mathbf{N}$. They can be estimated by doing two separate probit regression as defined by (13) using the complete data in the experiment. After obtaining the fitted values for a_1, a_2, b_1, and b_2, the functions l, u, σ_1, and σ_2 can be obtained from (14). The two-step procedure in (12) can now be implemented to find the optimal settings of the factors. This approach is known as *response modeling* in the robust parameter design literature. Another approach is the *performance measure modeling*. In this approach the performance measure is estimated for each control run and then modeled with respect to the control factors. This can be done as follows. Assume that the experimental design is a cross array. Consider a run defined by the control setting $\mathbf{X}_i$ and noise setting $\mathbf{N}_j$. Now fit two separate probit regressions to estimate a_1, a_2, b_1, and b_2 by using only the data in this run. To emphasize the difference from response modeling, we will denote $\hat{a}_1(\mathbf{X}_i, \mathbf{N}_j)$ by $\hat{a}_{1ij}$, $\hat{b}_1(\mathbf{X}_i, \mathbf{N}_j)$ by $\hat{b}_{1ij}$ and so on. Then the PM in (11) can be estimated as

$$\hat{PM}_i = \frac{1}{J}\sum_{j=1}^{J} \hat{l}_{ij} e^{\hat{\sigma}_{1ij}^2/2} \frac{1}{J}\sum_{j=1}^{J} \frac{e^{\hat{\sigma}_{2ij}^2/2}}{\hat{u}_{ij}},$$

where $\hat{l}_{ij}, \hat{u}_{ij}, \hat{\sigma}_{1ij}$, $\hat{\sigma}_{2ij}$ are obtained from (14). Now we can fit $\log \hat{PM}$ in terms of $\mathbf{X}$ using least squares and perform the optimization. Although response modeling has a stronger statistical justification, performance measure modeling is popular among practitioners because of its simplicity. For details on both approaches, See Wu and Hamada (2000, Chapter 10).

2.2 A Special Case: Operating Window Method

We can consider a more general loss function than (4), say

$$L = \frac{c_1}{Y_1^{\gamma_1}} + c_2 Y_2^{\gamma_2}, \tag{15}$$

where γ_1 and γ_2 are some positive constants. Then derivations similar to those for (11) give the PerMIA

$$PM_{\gamma_1,\gamma_2}(\mathbf{X}) = E_N^{1/\gamma_1}(l^{\gamma_1}(\mathbf{X},\mathbf{N}) e^{\gamma_1^2\sigma_1^2(\mathbf{X},\mathbf{N})/2}) E_N^{1/\gamma_2}(\frac{e^{\gamma_2^2\sigma_2^2(\mathbf{X},\mathbf{N})/2}}{u^{\gamma_2}(\mathbf{X},\mathbf{N})}). \tag{16}$$

Consider now a special case with $\sigma_1(\mathbf{X},\mathbf{N}) = \sigma_2(\mathbf{X},\mathbf{N}) = 0$. As σ_1 and σ_2 approach zero, p_1 and p_2 become step functions given by $p_1(\mathbf{X},\mathbf{N},M) = \Phi_0(l(\mathbf{X},\mathbf{N}) - M)$ and $p_2(\mathbf{X},\mathbf{N},M) = \Phi_0(M - u(\mathbf{X},\mathbf{N}))$, where

$$\Phi_0(x) = \begin{cases} 0 & , x < 0 \\ 1/2 & , x = 0 \\ 1 & , x > 0 \end{cases} \tag{17}$$

For $l(\mathbf{X},\mathbf{N}) < u(\mathbf{X},\mathbf{N})$, both $p_1(\mathbf{X},\mathbf{N},M)$ and $p_2(\mathbf{X},\mathbf{N},M)$ are 0 for $M \in (l,u)$. Thus, (l,u) represents a window with no failures, which is known as the *operating window* (Clausing, 1994). The performance measure in this case becomes

$$PM_{\gamma_1,\gamma_2}(\mathbf{X}) = E_N^{1/\gamma_1}(l^{\gamma_1}(\mathbf{X},\mathbf{N}))E_N^{1/\gamma_2}(\frac{1}{u^{\gamma_2}(\mathbf{X},\mathbf{N})}). \tag{18}$$

A small value of l and a large value of u lead to a small value of $PM_{\gamma_1,\gamma_2}(\mathbf{X})$. Thus, minimizing $PM_{\gamma_1,\gamma_2}(\mathbf{X})$ is roughly equivalent to maximizing the size of the operating window. The *operating window signal-to-noise ratio* (Taguchi, 1993 and Joseph and Wu, 2002) is given by

$$SN = -\log\left(\frac{1}{J}\sum_{j=1}^{J} l_j^2 \frac{1}{J}\sum_{j=1}^{J}\frac{1}{u_j^2}\right), \tag{19}$$

where the summation is taken over different noise levels ($j = 1, 2, \cdots, J$). It is the sample analog of $-2\log PM_{2,2}$. Thus, the operating window signal-to-noise ratio is a special case of the performance measure in (16). Therefore, the operating window (OW) method can be viewed as a special case of the failure amplification method (see Figure 1).

The signal-to-noise ratio can be used for optimization in more general cases, when $\sigma_1(\mathbf{X},\mathbf{N})$ and $\sigma_2(\mathbf{X},\mathbf{N})$ are independent of $\mathbf{X}$ and $\mathbf{N}$. This generalization will allow some failures in (l,u). The interpretation of the operating window can still be retained by modifying the definition of the operating window by defining it with respect to a threshold failure rate. See Clausing (1994) and Joseph and Wu (2002) for details. If the dependency of the σ's on $\mathbf{X}$ is ignored, the best settings can be missed, thus resulting in a large loss. In the performance measure (11), in addition to minimizing l and maximizing u, as is done using the SN ratio, one also minimizes σ_1 and σ_2. The proposed data analysis is also different from the existing practice. Because the Clausing-Taguchi approach to analysis only uses the values of l_j and u_j in (19), most of the case studies reported in the literature give only the values of l and u and not the complete data as in Table 3. The complete data are usually discarded after obtaining l and u in some manner. *This practice should be discouraged* because it throws away valuable information in the original data and makes it impossible to estimate σ_1 and σ_2. Furthermore, if the investigator decides to use a different method of modeling and analysis , he/she will need to have access to the complete data.

Although in theory σ_1 and σ_2 can be modeled as functions of $\mathbf{X}$ and $\mathbf{N}$, in practice it is very difficult to obtain good estimates unless there is a huge amount of data. See Nair (1986) and

Hamada and Wu (1990) for discussions about the difficulties associated with estimating dispersion effects from categorical data. With the paucity of data we may assume σ_1 and σ_2 to be constants and therefore in practice the operating window method becomes an important special case of FAMe.

2.3 Single Latent Variable Modeling

The plating process described in the introduction is a good example for single latent variable modeling. We may view the plating thickness as the underlying latent variable for the two failures: under-plating and over-plating. A major distinguishing feature of the single latent variable modeling is that the *two failure modes cannot happen together*. Such is allowed in the double latent variable modeling. Let Y be the single latent variable having a failure generating mechanism:

$$\text{if } Y < T_1 \text{ then failure type 1; if } Y > T_2 \text{ then failure type 2.}$$

Suppose Y and M are related by $Y = \beta(\mathbf{X}, \mathbf{Z})M$. Using arguments similar to those made in Section 2.1, assume

$$\log Y|\mathbf{N} \sim N(\mu(\mathbf{X}, \mathbf{N}) + \log M,\ \sigma^2(\mathbf{X}, \mathbf{N})).$$

Reparameterize T_1 and T_2 in terms of T and Δ by using $\log T_1 = \log T - \Delta$ and $\log T_2 = \log T + \Delta$, where T can be interpreted as the target for Y. Also let $\theta(\mathbf{X}, \mathbf{N}) = \mu(\mathbf{X}, \mathbf{N}) - \log T$. Then the failure probabilities can be obtained as

$$p_1(\mathbf{X}, \mathbf{N}, M) = P(Y < T_1|\mathbf{N}) = \Phi(-\frac{\Delta + \theta(\mathbf{X}, \mathbf{N}) + \log M}{\sigma(\mathbf{X}, \mathbf{N})}) \tag{20}$$

and

$$p_2(\mathbf{X}, \mathbf{N}, M) = P(Y > T_2|\mathbf{N}) = \Phi(\frac{-\Delta + \theta(\mathbf{X}, \mathbf{N}) + \log M}{\sigma(\mathbf{X}, \mathbf{N})}). \tag{21}$$

The loss function can be taken as

$$L = c\left(\frac{T}{Y} + \frac{Y}{T}\right),$$

which is a special case of (4) with $Y_1 = Y_2$ (see Joseph (2003) for details). Using derivations similar to those in Section 2.1, the optimum setting for M is

$$M^* = \left\{ \frac{E_N(e^{-\theta(\mathbf{X},\mathbf{N})+\sigma^2(\mathbf{X},\mathbf{N})/2})}{E_N(e^{\theta(\mathbf{X},\mathbf{N})+\sigma^2(\mathbf{X},\mathbf{N})/2})} \right\}^{1/2}, \tag{22}$$

and the PerMIA is

$$PM(\mathbf{X}) = E_N(e^{-\theta(\mathbf{X},\mathbf{N})+\sigma^2(\mathbf{X},\mathbf{N})/2}) E_N(e^{\theta(\mathbf{X},\mathbf{N})+\sigma^2(\mathbf{X},\mathbf{N})/2}). \tag{23}$$

We can also use a quadratic loss function in $\log Y$, i.e. $L = c(\log Y - \log T)^2$. Then we obtain

$$M^* = e^{-E_N(\theta(\mathbf{X},\mathbf{N}))} \tag{24}$$

and the PerMIA is

$$PM(\mathbf{X}) = Var_N(\theta(\mathbf{X},\mathbf{N})) + E_N(\sigma^2(\mathbf{X},\mathbf{N})), \tag{25}$$

which are much simpler to work with than (22) and (23).

Here $\log u(\mathbf{X},\mathbf{N}) - \log l(\mathbf{X},\mathbf{N}) = 2\Delta$ is a constant, where l and u are defined in (9) and are solved from (20) and (21). Therefore, in the single latent variable modeling we do not have the interpretation of an operating window as in the double latent variable approach.

Under the single latent variable setting, the data on failures follow a multinomial distribution. The estimation of the performance measure can be done by using the response modeling or the performance measure modeling, similar to that in Section 2.1 except that the parameters are to be estimated by maximizing the multinomial log-likelihood:

$$\begin{aligned}\sum \quad \{ \quad & y_1 \log \Phi(-\frac{\Delta + \theta(\mathbf{X},\mathbf{N}) + \log M}{\sigma(\mathbf{X},\mathbf{N})}) + y_2 \log \Phi(\frac{-\Delta + \theta(\mathbf{X},\mathbf{N}) + \log M}{\sigma(\mathbf{X},\mathbf{N})}) \\ + \quad & (r - y_1 - y_2) \log(1 - \Phi(-\frac{\Delta + \theta(\mathbf{X},\mathbf{N}) + \log M}{\sigma(\mathbf{X},\mathbf{N})}) - \Phi(\frac{-\Delta + \theta(\mathbf{X},\mathbf{N}) + \log M}{\sigma(\mathbf{X},\mathbf{N})}))\},\end{aligned}$$

where the summation is over the complete data in the experiment.

2.4 Probability of Failure Modeling

Although the latent variables are unobserved, some knowledge about them is essential to choosing a meaningful loss function that evaluates the performance of the system. An alternative approach is to define the loss in terms of the probability of failures. In this approach (Joseph and Wu, 2002), the loss function is defined as

$$L = c_1 \frac{p_1}{1 - p_1} + c_2 \frac{p_2}{1 - p_2}. \tag{26}$$

Underlying this choice is the double latent variable assumption. If a single latent variable is more appropriate, then a loss function such as $L = (c_1 p_1 + c_2 p_2)/(1 - p_1 - p_2)$ may be considered. If the experimenter is uncertain about the underlying latent variable structure, then the loss function can be taken as $L = c_1 p_1 + c_2 p_2$. Joseph and Wu (2002) postulated models for p_1 and p_2 as

$$p_1 = \frac{1}{1 + \left(\frac{M}{l}\right)^{\alpha_1}} \text{ and } p_2 = \frac{1}{1 + \left(\frac{u}{M}\right)^{\alpha_2}}, \tag{27}$$

where α_1 and α_2 are some positive constants. This can be obtained based on the latent variable modeling described in Section 2.1, but with a logistic distribution instead of a normal distribution. Assume

$$\log Y_i|\mathbf{N} \sim logistic(\mu_i(\mathbf{X},\mathbf{N}) + \log M,\ \sigma_i^2) \tag{28}$$

for $i = 1, 2$. Here σ_1 and σ_2 are assumed to be independent of $\mathbf{X}$ and $\mathbf{N}$. Then under the failure generating mechanism given in (1), we get the models in (27), where

$$\alpha_1 = \frac{\sqrt{3}}{\pi\sigma_1},\ l(\mathbf{X},\mathbf{N}) = T_1 e^{-\mu_1(\mathbf{X},\mathbf{N})},\ \alpha_2 = \frac{\sqrt{3}}{\pi\sigma_2},\ \text{ and }\ u(\mathbf{X},\mathbf{N}) = T_2 e^{-\mu_2(\mathbf{X},\mathbf{N})}.$$

Using the loss function in (26), we can derive the PerMIA as

$$PM = E_N^{1/\alpha_1}(l^{\alpha_1}(\mathbf{X},\mathbf{N}))E_N^{1/\alpha_2}(1/u^{\alpha_2}(\mathbf{X},\mathbf{N})). \tag{29}$$

Minimizing the PM in (11) is equivalent to minimizing $E_N(l(\mathbf{X},\mathbf{N}))E_N(1/u(\mathbf{X},\mathbf{N}))$, which does not depend on σ_1 and σ_2, whereas the PM in (29) depend on them. This difference is mainly due to the difference in the loss functions. Fortunately the PM in (29) is not very sensitive to σ_1 and σ_2 and therefore the conclusions from the two approaches are expected to be close.

2.5 The Paper Feeder Experiment

As the roller in the paper feeder rotates, it develops a friction force over the surface of the top paper. At the same time a friction force is created between the top paper and the one below it, opposing the movement of the top paper. Thus, the driving force on the top paper is the difference between these two friction forces. Clearly the driving force (Y_1) can be taken as the latent variable behind the misfeed. If the driving force is small, the feeder cannot feed the paper and therefore results in misfeed. The friction force between the two papers (Y_2) can be taken as the latent variable underlying the multifeed. When this friction force is large, the feeder tends to feed the second paper as well, thus causing multifeed. Thus, a simple model for the paper feeder can be written as

$$\text{if } Y_1 < T_1 \text{ then misfeed; if } Y_1 > T_1 \text{ and } Y_2 > T_2 \text{ then multifeed.} \tag{30}$$

Note that if the driving force is not large enough to feed the top paper, then even if Y_2 is large, multifeed cannot happen. Thus, in this case, although a double latent variable approach seems plausible, both failure modes cannot happen together.

With the above latent variables, the models in (2) can be justified using Coloumb's law of friction, because the friction forces are proportional to the normal force. A quick look at the data

in Table 3 will reveal that the probability of both misfeed and multifeed happening at the same level of M is very small. Thus,

$$p_2(\mathbf{X}, \mathbf{N}, M) = P(Y_1 > T_1, Y_2 > T_2|\mathbf{N}) = P(Y_2 > T_2|\mathbf{N}) - P(Y_1 \leq T_1, Y_2 > T_2|\mathbf{N}) \approx P(Y_2 > T_2|\mathbf{N}).$$

This implies that the failures in the paper feeder can be approximated by the models used in Section 2.1.

We only consider the response modeling approach. As recommended in Wu and Hamada (2000), the two degrees of freedom for 3-level factors are split into linear and quadratic components with contrasts $x_l = (-1, 0, 1)$ and $x_q = (1, -2, 1)$ respectively. The 2-level factors x_1, x_4 and the noise factor N are coded as $x_l, N_l = (-1, 1)$. Fitting two separate probit regressions in (13) and using the AIC (Akaike Information Criterion) for variable selection, we get

$$\begin{aligned}
\hat{a}_1(\mathbf{X}, \mathbf{N}) &= 7.84 + 1.07x_{2l} - 1.62x_{6l} - .93x_{1l} + .72x_{7q} - 1.17x_{5l} - .56x_{2l}x_{6l}, \\
&\quad -.90x_{1l}x_{6l} + 1.13x_{4l}x_{6l} - .18x_{2q}x_{7q}, \\
\hat{b}_1(\mathbf{X}, \mathbf{N}) &= -3.08, \\
\hat{a}_2(\mathbf{X}, \mathbf{N}) &= -6.25 - 1.73x_{2l} - 1.36x_{3l} + .72x_{6l} - .14x_{2l}x_{8q} - .19x_{2q}x_{3l} - .48x_{4l}x_{6l}, \\
\hat{b}_2(\mathbf{X}, \mathbf{N}) &= 1.46 + .33x_{2l} + .32x_{3l}.
\end{aligned}$$

In the model selection only those models that satisfy the effect heredity principles (Wu and Hamada, 2000, Section 3.5) are considered. In the misfeed data, a few of the stack force levels are 0, which are increased to one fourth of the next lowest level of M, to avoid $-\infty$ values. Note that the noise factor does not appear in the models, implying this particular noise factor was not a good choice for the experiment. The PM in (11) simplifies to

$$\begin{aligned}
\log \hat{PM} &= -(\log \hat{u} - \log \hat{l}) + \frac{\sigma_1^2 + \sigma_2^2}{2} \\
&= \hat{a}_2/\hat{b}_2 - \hat{a}_1/\hat{b}_1 + .5/\hat{b}_1^2 + .5/\hat{b}_2^2.
\end{aligned}$$

Noting that x_1, x_4, and x_8 are qualitative and the others are quantitative, we choose the experimental region $D = \{\mathbf{X}|x_1, x_4 = 1, 2; x_8 = 1, 2, 3; 1 \leq x_2, x_3, x_5, x_6, x_7 \leq 3\}$. Minimizing $\log \hat{PM}$ with the control factors restricted to be in D, we get $\mathbf{X}^* = (2, 2.22, 1, 2, 3, 3, 2, 1)$. At this optimal setting M should be adjusted to $\sqrt{lu}\exp((\sigma_1^2 - \sigma_2^2)/4) = 12$.

Table 4: Factors and levels for the PCB experiment

Control factors	Notation	Levels		
		1	2	3
Preheat	x_1	No	Yes	-
Surface preparation	x_2	Scrub	Pumice	Chemical
Lamination speed	x_3	1.2 mpm	1.5 mpm	1.8 mpm
Lamination pressure	x_4	20 psi	40 psi	60 psi
Lamination temperature	x_5	95 ^{0}C	105 ^{0}C	115 ^{0}C
Exposure energy	$x_6(m)$	14	17	20
Developer speed	x_7	3 fpm	4 fpm	5 fpm
ORP	x_8	500	530	560

3. COMPLEXITY FACTOR METHOD

We will describe the complexity factor method by using an experiment reported by Maruthi and Joseph (1999) on the inner layer manufacturing process of PCBs. Two types of defects are encountered in the circuits: shorts and opens. The defect rate is very small, less than 3%. But the defect rate can shoot up due to some special causes in the process, particularly during the production of fine line circuits. Therefore it is important to improve the process capability so that even under the influence of special causes the defect rate will not go very high. Eight control factors were selected from different subprocesses of the process and are given in Table 4.

Because the defect rate is small, a very large sample size will be needed to find some shorts and opens unless some failure amplification is applied. To use the control factor method, the exposure energy (x_6) would probably the best choice for the amplification factor. A low energy during exposure means that the dry film photo-resist over the circuitry portion will be under-polymerized, which will be washed off during developing. This will lead to open circuits in the etching process. A high energy during exposure will polymerize the non-circuitry portions of the dry film photoresist and will not be removed during developing. This will then act as etch-resistant, leading to shorts in the circuits. But using exposure energy as an amplification factor is not very easy. The measurements of shorts and opens take more than one hour per panel. Thus a sequential

experiment will not be feasible because it will greatly prolong the experimentation time and will severely disrupt production. Also, the material cost in the experiment will be very high, because one panel need to be processed for each level of the exposure energy.

The complexity factor method is an ideal choice for this experiment. A small line width for the conductor will increase the chance of opens and a small spacing between the conductors will increase the chance of shorts. The line width and spacing can be taken as the complexity factors as they determine the complexity of PCB manufacturing. The company was engaged in producing PCBs with line width/spacing greater than or equal to 5 mil. For the purpose of amplification, it was decided to add 3 mil and 4 mil line width/spacing to the circuits. A special test pattern was designed for the experiment with line width and spacing varying from 3 to 7 mils with an increment of 1 mil. The test pattern has 80 pairs of conductors. Thus, data for all line widths and spacings can be obtained by using only one panel. See Maruthi and Joseph (1999) for more details of this experiment. Note that the objective of this experiment was not to estimate the defect rate as a function of line width and spacing. They are varied in the experiment only to amplify (i.e., generate) failures so that the effects of the factors can be estimated more precisely. Therefore the purpose of amplification is to increase the information in the experiment for studying the factor effects. The data based on an $OA(18, 2^1 \times 3^7)$ are given in Table 5. A quick glance at the data conveys the importance of failure amplification for this experiment. Very few shorts and opens were observed under 5-7 mils. If the experiment were run with only 5-7 mils, it would have been impossible to get meaningful estimates of the effects of interest. Because of failure amplification there is enough variation in the data, so that we can get good estimates of the effects and draw conclusions about their optimal settings.

Let M_1 be the line width, M_2 the spacing, and $\mathbf{X}$ the set of control factors. We will treat the exposure energy (x_6) as an *adjustment factor* and will denote it by m. So $\mathbf{X} = \{x_1, x_2, x_3, x_4, x_5, x_7, x_8\}$. Note that the exposure energy could have been used as an amplification factor, but in this experiment it was not because it would then result in increased experimentation cost as described in the previous paragraphs. Also note that there are no observed noise factors in this experiment. As in the paper feeder problem, we may use a latent variable approach to model shorts and opens. But because many subprocesses from inner layer manufacturing are involved in this experiment, it is difficult to find a meaningful latent variable. Therefore we opt to use the probability of

Table 5: $OA(18, 2^1 \times 3^7)$ and data from the PCB experiment

									Opens					Shorts				
Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	3	4	5	6	7	3	4	5	6	7
1	1	1	1	1	1	1	1	1	33	7	4	0	1	1	0	0	0	0
2	1	1	2	2	2	2	2	2	7	9	1	0	0	4	1	0	0	0
3	1	1	3	3	3	3	3	3	14	3	1	0	0	19	2	0	0	0
4	1	2	1	1	2	2	3	3	2	0	2	0	0	9	0	0	0	0
5	1	2	2	2	3	3	1	1	7	1	2	1	0	22	1	1	1	0
6	1	2	3	3	1	1	2	2	78	30	7	1	1	8	0	0	0	0
7	1	3	1	2	1	3	2	3	9	1	3	0	0	19	1	0	0	0
8	1	3	2	3	2	1	3	1	7	0	1	0	1	4	0	1	0	0
9	1	3	3	1	3	2	1	2	4	3	0	0	0	7	0	0	0	0
10	2	1	1	3	3	2	2	1	6	0	0	0	0	22	1	0	0	1
11	2	1	2	1	1	3	3	2	13	2	0	0	0	34	2	2	0	0
12	2	1	3	2	2	1	1	3	34	5	0	1	3	13	4	1	0	0
13	2	2	1	2	3	1	3	2	8	3	0	0	0	7	0	1	0	0
14	2	2	2	3	1	2	1	3	25	8	0	2	1	25	1	0	0	0
15	2	2	3	1	2	3	2	1	7	0	0	0	0	41	1	0	0	1
16	2	3	1	3	2	3	1	2	10	6	0	0	0	45	9	5	0	1
17	2	3	2	1	3	1	2	3	8	0	0	0	0	3	0	0	0	0
18	2	3	3	2	1	2	3	1	12	2	0	0	1	7	2	0	0	0

failure modeling approach for this experiment. The shorts/opens are measured by passing current between/through each pair of conductors. This gives rise to 80 opportunities for shorts and 160 opportunities for opens under each line width/spacing per run. (Recall there are in total 80 pairs of conductors.) Thus, the data on opens follow a $binomial(160, p_1(\mathbf{X}, m, M_1))$ and the data on shorts follow a $binomial(80, p_2(\mathbf{X}, m, M_2))$. In one particular conductor there could be many opens. Because of the nature of measurement, we do not get to observe the number of opens in a conductor. The data would have been more informative had we counted the number of opens/shorts. We can assume that the number of opens in a conductor and the number of shorts between a pair of conductors to follow Poisson distributions with means $\lambda_1(\mathbf{X}, m, M_1)$ and $\lambda_2(\mathbf{X}, m, M_2)$, respectively. Then we have the relations $p_1(\mathbf{X}, m, M_1) = 1 - exp(-\lambda_1(\mathbf{X}, m, M_1))$ and $p_2(\mathbf{X}, m, M_2) = 1 - exp(-\lambda_2(\mathbf{X}, m, M_2))$. We assume the following models,

$$\lambda_1(\mathbf{X}, m, M_1) = \frac{\lambda_1(\mathbf{X})}{M_1^{\alpha_1} m^{\gamma_1}} \text{ and } \lambda_2(\mathbf{X}, m, M_2) = \frac{\lambda_2(\mathbf{X}) m^{\gamma_2}}{M_2^{\alpha_2}},$$

where $\alpha_1, \alpha_2, \gamma_1$, and γ_2 are some positive constants. Then

$$\log\log \frac{1}{1-p_1} = \log \lambda_1(\mathbf{X}) - \gamma_1 \log m - \alpha_1 \log M_1 \tag{31}$$

and

$$\log\log \frac{1}{1-p_2} = \log \lambda_2(\mathbf{X}) + \gamma_2 \log m - \alpha_2 \log M_2. \tag{32}$$

This suggests that we can estimate the parameters by fitting two separate binomial GLMs (Generalized Linear Models) with a complementary log-log link function (McCullagh and Nelder, 1989).

The loss is proportional to the number of defects. So we can take the loss function to be

$$L = c_1 \lambda_1(\mathbf{X}, m, M_1) + c_2 \lambda_2(\mathbf{X}, m, M_2).$$

Because M_1 and M_2 are not in the control of the manufacturer, we will try to find an $\mathbf{X}$ to minimize the expected loss taken over the distributions of M_1 and M_2. Thus, the expected loss is

$$\begin{aligned} EL &= c_1 E\lambda_1(\mathbf{X}, m, M_1) + c_2 E\lambda_2(\mathbf{X}, m, M_2) \\ &= c_1 \frac{\lambda_1(\mathbf{X})}{m^{\gamma_1}} E\left(\frac{1}{M_1^{\alpha_1}}\right) + c_2 \lambda_2(\mathbf{X}) m^{\gamma_2} E\left(\frac{1}{M_2^{\alpha_2}}\right). \end{aligned}$$

We can set m to minimize EL. This gives

$$m^* = \left\{ \frac{\gamma_1 c_1 E(1/M_1^{\alpha_1}) \lambda_1(\mathbf{X})}{\gamma_2 c_2 E(1/M_2^{\alpha_2}) \lambda_2(\mathbf{X})} \right\}^{1/(\gamma_1+\gamma_2)}. \tag{33}$$

As in the previous sections, the PerMIA can be obtained by evaluating the expected loss at $m = m^*$. Minimizing this PerMIA is equivalent to minimizing

$$PM(\mathbf{X}) = \frac{1}{\gamma_1} \log \lambda_1(\mathbf{X}) + \frac{1}{\gamma_2} \log \lambda_2(\mathbf{X}). \tag{34}$$

By fitting a GLM with a complementary log-log link function and using forward variable selection based on AIC, from (31) and (32), we get the following models

$$\begin{aligned} \log \hat{\lambda_1} &= 10.72 - .73x_{5l} - .33x_{2l} - .27x_{1l}x_{5q}, \\ \log \hat{\lambda_2} &= -6.66 + .48x_{1l} + .20x_{4l} - .15x_{1l}x_{5q}, \end{aligned}$$

$$\hat{\gamma_1} = 2.768, \hat{\alpha}_1 = 5.06, \hat{\gamma}_2 = 4.70, \text{ and } \hat{\alpha}_2 = 7.664.$$

Substituting in (34) and minimizing we get $x_1^* = 1, x_2^* = 3, x_4^* = 1$, and $x_5^* = 2.34$. The other variables can be retained at the current levels. With $c_1 = c_2$ and a uniform distribution for line width and spacing over 5, 6, and 7, we get $m^* = 15.7$. Note that the optimization is carried out at the user conditions and not at the amplified conditions. The amplification is applied only for the purpose of estimation and not for optimization. Also $\mathbf{X}^*$ does not depend on the distribution of line width and spacing. Therefore only the setting of m needs to be changed if there is a change in the distribution of line width and spacing. This is one of the advantage of using PerMIA for optimization.

The major difference between the complexity factor method and the control factor method is that we need to take expectations with respect to the complexity factor in the optimization. The optimization in the noise factor method will be similar to that of the complexity factor method, because we should average the loss over the noise factor. Yet the two approaches are different because the variations in the complexity factor are specified by the customer whereas the variations in the noise factor are not. There can be other methods of failure amplification. For example, if the measurements are based on a go/no-go gauge, then the specifications can be tightened to produce more failures. We can treat the tolerance as a complexity factor and classify this method under the complexity factor method.

4. STRATEGIES FOR AMPLIFICATION

In applying FAMe, a practitioner will be faced with two important questions. How much failure amplification should be applied? And how to achieve the specified amplification? We address these

questions in this section. First we will present some results from optimal design theory that will give some guidelines about optimal amplification. Then we will discuss two design strategies to achieve it.

4.1 Optimal Amplification

Consider the following model

$$p = \Phi\left(\frac{\log\{M/u\}}{\sigma}\right) \tag{35}$$

as in (10). Suppose σ is known and we want to estimate u. At the current value of M in the process, the failure probability is very small. Therefore we will change M so as to improve the estimation of u. The Fisher information of $\eta = (\log M - \log u)/\sigma$ is

$$I(\eta) = \frac{\phi^2(\eta)}{\Phi(\eta)[1-\Phi(\eta)]}, \tag{36}$$

where ϕ is the standard normal density function. From maximum likelihood theory we know that $Var(\hat{u}) \approx \sigma^2 u^2 / I(\eta)$. Thus maximizing $I(\eta)$ will minimize the variance of the estimate $\hat{u}$ of u. The value of M that maximizes $I(\eta)$ is $M^* = u$. This means that we should adjust M to the LD50 point to maximize the information in the experiment. Unfortunately we do not know u (if we knew u, there would be no need to collect data to estimate it!) and therefore this result is not practical. But it suggests the experimenter that the failure rate should be amplified to about 50%. The optimality of the LD50 point also holds true for the logit link. For the complementary log-log link, because of the skewness, the optimal amplification is about 80%. See Wu (1988) for details on optimal designs with binary data.

If the scale parameter σ is unknown in (35), then we need to use more than one point for M to estimate both u and σ. We can use D- or A-optimality or some other optimal design criterion to select the points. Sitter and Wu (1993) show that a 2-point design is optimal for both probit and logit links. According to D-optimality the optimal amplifications are 13% and 87% for the probit link and 18% and 82% for the logit link.

4.2 Fixed Design

Let $\mathcal{D}(a)$ denote the design for factors a. In a fixed design approach we first select a design for the control and noise factors, $\mathcal{D}(\mathbf{X}, \mathbf{N})$. See Wu and Hamada (2000) for the choice of $\mathcal{D}(\mathbf{X}, \mathbf{N})$. This design will be crossed with the design for the amplification factor. This can be denoted as $\mathcal{D}(\mathbf{X}, \mathbf{N}) \otimes \mathcal{D}(M)$, which means all the runs in $\mathcal{D}(M)$ will be repeated at every run of $\mathcal{D}(\mathbf{X}, \mathbf{N})$.

The $\mathcal{D}(M)$ is a set of points $\{M_1, M_2, \cdots, M_n\}$. Optimal design theory tells us that M should be set at the LD50 point (for the probit/logit link with known scale parameter). But the LD50 point is a function of both $\mathbf{X}$ and $\mathbf{N}$ and is unknown.

We now describe the construction of a fixed design that can increase the information in the experiment compared to conducting all the experiments at the current value of $M = M_0$. Consider the model $p(\mathbf{X}, \mathbf{N}, M) = \Phi(\frac{\log\{M/u(\mathbf{X},\mathbf{N})\}}{\sigma})$ with $\sigma > 0$ known. Suppose $p(\mathbf{X}, \mathbf{N}, M_0) \leq \alpha < 0.5$ for all the runs in $\mathcal{D}(\mathbf{X}, \mathbf{N})$. Let $M^0 = M_0 \exp(-2\sigma\Phi^{-1}(\alpha))$. Then choose n points from (M_0, M^0) as $\mathcal{D}(M)$. If there is no prior knowledge about σ, then the experimenter should try to guess an M^0 satisfying $p(\mathbf{X}, \mathbf{N}, M^0) \leq 1 - \alpha$ for all the runs in $\mathcal{D}(\mathbf{X}, \mathbf{N})$.

The PCB experiment is an example of a fixed design. For the opens, $\mathcal{D}(M_1) = \{3, 4, 5, 6, 7\}$. Only 3 and 4 mil line widths were used for amplification. From the data in Table 5 it is clear that using 6 and 7 mils did not produce many failures and therefore they were a complete waste. More samples of 3 and 4 mils could have been used instead, which would have increased the information in the experiment.

4.3 Sequential Design

In a sequential design $\mathcal{D}(M)$ is different for each run in $\mathcal{D}(\mathbf{X}, \mathbf{N})$ and the points in $\mathcal{D}(M)$ are chosen sequentially. At each run, the value of M_{k+1} is determined based on the data from the previous k design points. There are many sequential methods available in the literature for optimally choosing M_{k+1}. The best method for our problem depends on the modeling assumptions. For example, consider the model in (35). If we are not interested in modeling the scale parameter σ as a function of $\mathbf{X}$ and $\mathbf{N}$, then the stochastic approximation method for estimating LD50 given in Wu (1985) will be a good choice. This method produces a sequence of design points for M that will converge to the LD50. It is well known that stochastic approximation methods do not provide enough information about σ for precise estimation. Therefore if we are interested in modeling σ also as a function of $\mathbf{X}$ and $\mathbf{N}$, then sequential methods (e.g., McLeish and Tosh (1990) and Neyer (1994)) that spread out the design points will be more appropriate.

We now describe Wu's sequential procedure as applied to our problem. Suppose we are at the n^{th} run of $\mathcal{D}(\mathbf{X}, \mathbf{N})$ and k^{th} run of $\mathcal{D}(M)$. Assume that σ is a constant. Let y_{ij} be the observed number of failures from a sample of size r when $M = M_{ij}$. Then the MLE's for $u_1, \cdots, u_n$, and σ

can be obtained by maximizing the log-likelihood

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\left\{y_{ij}\log\Phi(\frac{\log\{M_{ij}/u_i\}}{\sigma})+(r-y_{ij})\log[1-\Phi(\frac{\log\{M_{ij}/u_i\}}{\sigma})]\right\}.$$

This can be easily done using probit regression. The sequential design is then to set $M_{n,k+1}$ at the MLE of u_n.

The paper feeder experiment discussed in Section 2 is an example of a sequential design. The experimenter actually used a trial and error approach to find the LD50 point for each run. This search could have been improved by using the sequential procedures discussed in this section.

5. COMPARISON WITH TRADITIONAL EXPERIMENTS

In this section we will compare the efficiency of FAMe with the traditional way of conducting experiments with categorical data. Because the amplification factor is not of primary interest to the experimenter, the amplification factor is kept at its current setting (say M_0) in the traditional approach. In FAMe this factor is varied to amplify the failures. Fixed design and sequential design can be adopted for amplification. Intuitively, sequential designs are better than fixed designs. In sequential designs the optimal amplification can be achieved for every run, whereas in fixed designs the amount of amplification varies from run to run and can result in a loss of information. As we have seen, about 50% amplification will maximize the information in the experiment. So if we test r samples at the LD50 point, the variance of the estimate will be smaller than that of testing r samples at M_0. But in practice we do not know the LD50 point. We need to spend some resources to search and estimate the LD50 point. Is this still an efficient approach compared to spending all the resources at M_0 ? The following result provides some insight on this question.

PROPOSITION 1 *Let* $p(\mathbf{X},\mathbf{N},M)=\Phi(\frac{\log\{M/u(\mathbf{X},\mathbf{N})\}}{\sigma})$, *where* σ *is known. Suppose there exists an* $\alpha\in(0,.5)$ *and an* $a\in(0,\infty)$ *such that* $p(\mathbf{X},\mathbf{N},M_0)\le\alpha$ *and* $u(\mathbf{X},\mathbf{N})\le a$ *for all the runs in* $\mathcal{D}(\mathbf{X},\mathbf{N})$. *Then we can find a fixed design that performs better than the traditional design. Moreover there exists a sequential design that performs better than the fixed design.*

The proof of this proposition is given in the Appendix. Note that, for $\sigma>0$ the proof holds only asymptotically. The result for $\sigma=0$ is stronger. In this case the traditional design does not provide any information about u. For both fixed designs and sequential designs the posterior

variance of u goes to 0 as $n \to \infty$. Moreover the efficiency of the sequential design compared to the fixed design goes to ∞ as $n \to \infty$, implying the superior performance of sequential designs. Thus, based on the above proposition we can conclude that FAMe is more efficient than the traditional approach.

6. FURTHER DISCUSSIONS AND CONCLUDING REMARKS

The failure amplification method (FAMe) was originally motivated by our attempt to understand, justify and extend the operating window (OW) method. The OW method was proposed by Don Clausing at Xerox and developed by Taguchi and his Japanese colleagues at Fuji-Xerox. It has been used in some industrial sectors and case studies can be found in the applied literature. Joseph and Wu (2002) appear to be the first to study its properties analytically and develop a systematic approach to the choice of a performance measure, modeling, analysis, and system optimization. We realized that the core idea behind the OW method is failure amplification which led to this work. Because failure amplification can be achieved in many different ways, the application areas are considerably expanded and the OW method becomes a special case of FAMe. As noted near the end of Section 2.2, the proposed approach to modeling and analysis utilizes the complete data. It fully exploits the information in the data and thus is an improvement over the existing OW approach.

FAMe was developed in this paper as a general approach to experimentation and information extraction. In experiments with an imprecise measurement system (or due to the very nature of the system) categorical data are often collected. They are cheaper or easier to measure but contain less information than continuous data. In extreme situations they provide little or no information for the investigators to estimate the factor effects. To ameliorate this shortcoming, an amplification factor can be used to extract more information by generating different numbers of failures with experiments run at various settings of the amplification factor. A major assumption in FAMe is that the optimum setting arrived at in the amplified conditions will remain the same in the current conditions. It is encapsulated in the models employed in this paper. Note that in many applications this is a reasonable assumption. Two versions of the FAMe are studied in the paper: the control factor method and the complexity factor method. Distinctions between the two versions are discussed and illustrated with real examples.

The approach of failure amplification is akin to the accelerated life testing (ALT) in reliability

studies (Meeker and Escobar, 1998) but there are some differences. An obvious one is that ALT is used to study reliabilty and FAMe to improve quality. The response in ALT is life time, whereas in FAMe it is categorical (type of failure or no failure). FAMe deals with only sudden failures and not the degradation failures as in ALT. In ALT the acceleration is applied to reduce the testing time by accelerating the degradation of the system, whereas in FAMe it is used to reduce the sample size by amplifying the probability of sudden failure. The literature on ALT focuses on estimating and demonstrating reliability at user conditions through extrapolation of the acceleration factor, whereas the FAMe is developed for system optimization and therefore the details differ completely from the ALT. Further research can benefit from recognizing the connections with the ALT approach and using its vast literature on modeling and extrapolation.

Some simplifying assumptions are made in the models in order to derive tractable results and procedures. For example, the models in (2) and a similar one in Section 2.3 assume linearity through the origin. More general models can be considered in future research, but the identifiability and estimability of parameters are a major concern. The loss functions are chosen primarily to facilitate the derivations of tractable results. Other loss functions can be used. Comparisons between traditional design, fixed design, and sequential design are made under a set of rather restrictive conditions. More realistic conditions should be considered. The design strategies discussed in Section 4 are to be applied for each failure mode separately, which may not be optimal when we consider all the failure modes together. Optimal design strategies with multiple failure modes is an interesting topic for future research.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and an Associate Editor for valuable comments and to Mr. K. Tatebayashi of Fuji-Xerox for giving the details on the paper feeder experiment. This research was supported by the NSF grant DMI-0217395.

APPENDIX: PROOF OF PROPOSITION 1

We consider two cases. The notation $\mathbf{X}, \mathbf{N}$ is omitted in the proof for simplicity.

Case (i): $\sigma > 0$.

In a fixed design (F.D.) $\mathcal{D}(M) = \{M_1, \cdots, M_n\}$ with r samples tested at each level of M. In the traditional design (T.D.) all the nr samples are tested at the current value of the amplification

factor, M_0. Then, under the T.D.,

$$\sqrt{nrI(\eta_0)}(\hat{u}_n - u) \Rightarrow N(0, \sigma^2 u^2),$$

where $\eta_0 = \log\{M_0/u\}/\sigma$ and $I(\eta)$ is given in (36).

Let $M^0 = M_0 \exp(-2\sigma\Phi^{-1}(\alpha))$, so that $p(M^0) \leq 1-\alpha$. For a fixed design and $M_0 < M_i < M^0$ for all i, we have under the F.D.,

$$\sqrt{r\sum_{i=1}^{n} I(\eta_i)}(\hat{u}_n - u) \Rightarrow N(0, \sigma^2 u^2),$$

where $\eta_i = \log\{M_i/u\}/\sigma$. Because $p(M)$ is strictly increasing in M, we have $p(M_0) < p(M_i) < 1-p(M_0)$ for all $i = 1, \cdots, n$. Because $I(\eta)$ is a unimodal and symmetric function of η, $I(\eta_i) > I(\eta_0)$ for all $i = 1, \cdots, n$. Thus, the asymptotic relative efficiency (ARE) of the fixed design with respect to the traditional design is

$$ARE(F.D., T.D.) = \lim_{n\to\infty} \frac{Var(\hat{u}_n)_{T.D.}}{Var(\hat{u}_n)_{F.D.}} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} I(\eta_i)}{nI(\eta_0)} > 1.$$

Now consider a sequential design (S.D.). Let M_1 be the midpoint of (M_0, M^0) and $M_{k+1} = \hat{u}_k$, where $\hat{u}_k$ is the MLE of u after k experiments. Then, based on the asymptotic results of Ying and Wu (1997),

$$\sqrt{nr2/\pi}(\hat{u}_n - u) \Rightarrow N(0, \sigma^2 u^2).$$

The maximum value of $I(\eta)$ is $2/\pi$. Thus

$$ARE(S.D., F.D.) = \lim_{n\to\infty} \frac{Var(\hat{u}_n)_{F.D.}}{Var(\hat{u}_n)_{S.D.}} = \lim_{n\to\infty} \frac{n2/\pi}{\sum_{i=1}^{n} I(\eta_i)} > 1.$$

Case (ii) $\sigma = 0$.

Here $p(M) = \Phi_0(M-u)$, with Φ_0 defined in (17). In this case a unique MLE does not exist. Therefore we will use a Bayesian formulation of the problem. Assume u follows a uniform distribution $U(0, a)$. $p(M_0) \leq \alpha < 0.5$ implies $M_0 < u$ and $p(M_0) = 0$. So we could as well take the prior distribution of u to be $U(M_0, a)$. Without loss of generality, assume $r = 1$. Then the number of failures $y_i|u \sim Bernoulli(p(M))$. Then at $M = M_0$, $\mathbf{y} = (y_1, \cdots, y_n) = (0, \cdots, 0)$. Thus,

$$f(u|\mathbf{y}) \propto f(\mathbf{y}|u)f(u) = (\Phi_0(u - M_0))^n \, 1_{(M_0,a)}(u) = 1_{(M_0,a)}(u),$$

which is same as the prior distribution. In other words no information is gained by doing experiments at M_0. The best estimate of u is $E(u|\mathbf{y}) = (M_0 + a)/2$ and the posterior variance of u in the traditional design is

$$Var(u|\mathbf{y})_{T.D.} = \frac{(a - M_0)^2}{12}.$$

For a fixed design, one sample is tested at n levels of M. Let $M_i = M_0 + (a - M_0)i/(n+1)$ for $i = 1, \cdots, n$. Let $M_{n+1} = a$. Then $\mathbf{y}$ can take $n+1$ possible values, $\mathbf{y}^k = (y_1, \cdots, y_k, y_{k+1}, \cdots, y_n) = (0, \cdots, 0, 1, \cdots, 1), k = 0, 1, 2, \cdots, n$. Then

$$\begin{aligned} f(u|\mathbf{y} = \mathbf{y}^k) &\propto f(\mathbf{y}^k|u)f(u) \\ &= \Phi_0(u - M_1) \cdots \Phi_0(u - M_k)\Phi_0(M_{k+1} - u) \cdots \Phi_0(M_n - u)1_{(M_0,a)}(u) \\ &= \Phi_0(u - M_k)\Phi_0(M_{k+1} - u)1_{(M_0,a)}(u), \end{aligned}$$

which is 1 if $u \in (M_k, M_{k+1})$, 1/2 if $u = M_k$ or M_{k+1}, and 0 otherwise. Thus, the posterior distribution of u differs from $U(M_k, M_{k+1})$ with measure 0. Denote this distribution by $\tilde{U}(M_k, M_{k+1})$. Because $M_{k+1} - M_k = (a - M_0)/(n + 1)$ for all k, the posterior variance of u does not depend on $\mathbf{y}$ and is given by

$$Var(u|\mathbf{y})_{F.D.} = \frac{(a - M_0)^2}{12(n+1)^2}.$$

Because equal spacing minimizes the maximum length of the intervals, the above fixed design minimizes the maximum posterior variance over all possible fixed designs.

For a sequential design, let $M_1 = (M_0 + a)/2$ and $M_{k+1} = \hat{u}_k$, where $\hat{u}_k$ is the posterior mean of u after k experiments. After observing y_1,

$$f(u|M_1, y_1) \propto f(y_1|u, M_1)f(u|M_1) = (\Phi_0(M_1 - u))^{y_1} (\Phi_0(u - M_1))^{1-y_1} 1_{(M_0,a)}(u).$$

Thus, the posterior distribution of u is $\tilde{U}(M_1, a)$ if $y_1 = 0$ and $\tilde{U}(M_0, M_1)$ if $y_1 = 1$. So $M_2 = (M_1 + a)/2$ if $y_1 = 0$ and $M_2 = (M_0 + M_1)/2$ if $y_1 = 1$. Note that the interval (M_0, a) is halved after observing y_1. Now observe y_2 at $M = M_2$. Proceeding similarly, after n steps the posterior distribution of u is $\tilde{U}(M_k, M_n)$ for $k = 0, 1, \cdots, n-1, n+1$, depending on $\mathbf{y}$. Each interval (M_k, M_n) or (M_n, M_k) has a length of $(a - M_0)/2^n$, which does not depend on $\mathbf{y}$. Thus the posterior variance of u is

$$Var(u|M_1, \cdots, M_n, \mathbf{y})_{S.D.} = \frac{(a - M_0)^2}{12 \times 2^{2n}}.$$

Thus, we have

$$Var(u|M_1, \cdots, M_n, \mathbf{y})_{S.D.} < Var(u|\mathbf{y})_{F.D.} < Var(u|\mathbf{y})_{T.D.}, \text{ for all } n > 1.$$

$\diamondsuit$

REFERENCES

Clausing, D. (1994), *Total Quality Development: A Step-by-Step Guide to World Class Concurrent Engineering*, ASME Press, New York, NY.

Hamada, M. and Wu, C. F. J. (1990), "A Critical Look at Accumulation Analysis and Related Methods," (with discussion), *Technometrics*, 32, 119-162.

Joseph, V. R. (2003), "Quality Loss Functions for Nonnegative Variables and Their Applications," *Journal of Quality Technology*, to appear.

Joseph, V. R. and Wu, C. F. J. (2002), "Operating Window Experiments: A Novel Approach to Quality Improvement", *Journal of Quality Technology*, 34, 345-354.

Leon, R. V., Shoemaker, A. C., and Kacker, R. N. (1987), "Performance Measures Independent of Adjustment: An Explanation and Extension of Taguchi's Signal-to-Noise Ratios," (with discussion), *Technometrics*, 29, 253-285.

Leon, R. V. and Wu, C. F. J. (1992), "A Theory of Performance Measures in Parameter Design," *Statistica Sinica*, 2, 335-358.

Maruthi, B. N. and Joseph, V. R. (1999), "Improving the Yield of Printed Circuit Boards Using Design of Experiments," *Quality Engineering*, 12, 259-265.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall.

McLeish, D. L. and Tosh, D. (1990). "Sequential Designs in Bioassay," *Biometrics* 46, 103-116.

Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: John Wiley & Sons.

Miyakawa, M. (1993), "Logistic Regression Analysis for Observations from the Operating Window Method," *Technical Report* IA-TR-93-4, Science University of Tokyo, Japan.

Nair, V. N. (1986), "Testing in Industrial Experiments with Ordered Categorical Data," *Technometrics*, 28, 283-291.

Neyer, B. T. (1994). "D-Optimality-Based Sensitivity Test," *Technometrics* 36, 61-70.

Sitter, R. R. and Wu, C. F. J. (1993), "Optimal Designs for Binary Response Experiments: Fieller, D, and A criteria," *Scandinavian Journal of Statistics*, 20, 329-341.

Taguchi, G. (1993), *Taguchi on Robust Technology Development*, ASME Press, New York.

Wu, C. F. J. (1985), "Efficient Sequential Designs with Binary Data," *Journal of American Statistical Association*, 80, 974-984.

Wu, C. F. J. (1988), "Optimal Design for Percentile Estimation of a Quantal Response Curve," *Optimal and Analysis of Experiments* (Y. Dodge et al., Ed.), 213-223, Amsterdam: Elsevier.

Wu, C. F. J., and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: John Wiley & Sons.

Ying, Z. and Wu, C. F. J. (1997), "An Asymptotic Theory of Sequential Designs Based on Maximum Likelihood Recursions," *Statistica Sinica*, 7, 75-91.