ROBUST PARAMETER OPTIMIZATION STRATEGIES IN COMPUTER SIMULATION EXPERIMENTS

by

Renato P. Panis

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctorate of Philosophy

in

Management Science

APPROVED

Ernest C. Houck, Chairman

Raymond H. Myers

Lance A. Matheson

Edward R. Clayton

Ralph D. Badinelli

Terry R. Rakes

August 31, 1994

Blacksburg, Virginia

ROBUST PARAMETER OPTIMIZATION STRATEGIES IN COMPUTER SIMULATION EXPERIMENTS

by

Renato P. Panis

Ernest C. Houck, Chairman

Management Science

(ABSTRACT)

An important consideration in computer simulation studies is the issue of model validity, the level of accuracy with which the simulation model represents the real world system under study. This dissertation addresses a major cause of model validity problems: the dissimilarity between the simulation model and the real system due to the dynamic nature of the real system that results from the presence of nonstationary stochastic processes within the system. This transitory characteristic of the system is typically not addressed in the search for an optimal solution.

In reliability and quality control studies, it is known that optimizing with respect to the variance of the response is as important a concern as optimizing with respect to average performance response. Genichi Taguchi has been instrumental in the advancement of this philosophy. His work has resulted in what is now popularly known as the Taguchi Methods for robust parameter design. Following Taguchi's philosophy, the goal of this research is to devise a framework for finding optimum operating levels for the controllable input factors in a stochastic system that are insensitive to internal sources of variation.

Specifically, the model validity problem of nonstationary system behavior is viewed as a major internal cause of system variation.

In this research the typical application of response surface methodology (RSM) to the problem of simulation optimization is examined. Simplifying assumptions that enable the use of RSM techniques are examined. The relaxation of these assumptions to address model validity leads to a modification of the RSM approach to properly handle the problem of optimization in the presence of nonstationarity. Taguchi's strategy and methods are then adapted and applied to this problem. Finally, dual-response RSM extensions of the Taguchi approach separately modeling the process performance mean and variance are considered and suitably revised to address the same problem.

A second cause of model validity problems is also considered: the random behavior of the supposedly controllable input factors to the system. A resolution to this source of model invalidity is proposed based on the methodology described above.

ACKNOWLEDGMENTS

I am deeply indebted to my chairman and mentor, Dr. Ernest C. Houck, for his guidance, support, enthusiasm, and patience throughout my graduate studies. Dr. Houck's excellent direction of my program of studies and dissertation research is greatly appreciated.

I wish to thank Dr. Raymond H. Myers for the many hours spent imparting his wisdom, expertise, and insightful thoughts. I am grateful for the opportunity to do collaborative work in statistics.

I would like to thank Dr. Lance Matheson for his logistical and computing support. Lance has been a great source of moral support and encouragement.

My sincere thanks to Dr. Edward R. Clayton, Dr. Ralph D. Badinelli, and Dr. Terry R. Rakes for graciously serving in my dissertation committee. The perspectives they added to the research were invaluable.

I would like to thank Dr. Bernard W. Taylor, III, for providing the necessary research as well as monetary support throughout my graduate studies. I am also grateful for the opportunity to teach as a lecturer in the Management Science Department.

My sincere appreciation to my unofficial dissertation committee members, Clark Gaylord and Fernando Siochi. Clark has been an excellent statistical consultant and sounding board. Fern's computing assistance has been invaluable.

Thanks to all my friends at Virginia Tech who made graduate student life truly memorable.

And, finally, many thanks to my brother Roberto, his wife, Norma, and their family for their support and encouragement. I would also like to thank the Gaylords, the Peraltas, and the Siochis, my adopted Blacksburg families, for their friendship and support.

DEDICATION

This research is dedicated to my parents, Celestina and Domingo, for instilling in me a strong appreciation for higher learning.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION 1

1.1 SYNOPSIS..1

1.2 OPTIMIZATION THROUGH COMPUTER SIMULATION.................4

1.2.1 CONCEPTUAL MODEL VALIDATION 7

1.3 PROBLEM DEFINITION AND SCOPE.......................................10

1.3.1 ILLUSTRATIVE PROBLEM 14

1.4 PROPOSED METHODOLOGY...18

1.4.1 GENERAL APPROACH..18

1.4.2 SOLUTION TECHNIQUES 20

Chapter 2

RESPONSE SURFACE METHODS.. 22

2.1 INTRODUCTION...22

2.2 METHODOLOGY...24

2.2.1 FIRST-ORDER ANALYSIS 25

2.2.2 SECOND ORDER ANALYSIS 30

2.3 RSM APPLIED TO THE CONDITIONAL NONSTATIONARITY 34

2.3.1 STATIC MODEL APPROACH.......................................35

2.3.1.1 PHASE 1 38

2.3.1.2 PHASE 2 48

2.3.2 NONSTATIONARY MODEL APPROACH 52

2.3.2.1 NUMERICAL RESULTS.................................56

2.4 CONCLUDING REMARKS 71

Chapter 3

ROBUST PARAMETER DESIGN.. 74

3.1 INTRODUCTION.....74
3.2 TAGUCHI STRATEGY.....77
3.3 TAGUCHI TACTICS.....80
3.3.1 TAGUCHI DESIGNS.....80
3.3.2 TAGUCHI ANALYSIS.....83
3.3.2.1 SIGNAL-TO-NOISE RATIOS.....84
2.3.2.2 OPTIMIZATION PROCEDURE.....87
3.4 CRITICISMS OF ROBUST PARAMETER DESIGN.....89
3.4.1 INADEQUACY OF THE OPTIMIZATION PROCEDURE.....90
3.4.2 INEFFICIENCY OF EXPERIMENTAL DESIGN PLANS.....91
3.4.3 IMPROPRIETY OF S/N AS PERFORMANCE STATISTIC.....93
3.4.3.1 INEFFICIENCY OF S/N IN "TARGET-IS-BEST" CASE.....93
3.4.3.1.1 MULTIPLICATIVE ERROR MODEL CASE.....98
3.4.3.1.2 ADDITIVE ERROR MODEL.....100
3.4.3.2 INEFFICIENCY ON S/N IN MEAN ADJUSTMENT PROCESS.....101
3.5 A MODIFIED ROBUST PARAMETER DESIGN APPROACH.....103
3.5.1 MODIFICATIONS TO THE TAGUCHI METHOD.....104
3.5.2 IMPLEMENTATION AND NUMERICAL RESULTS.....106
3.6 CONCLUDING REMARKS.....128

Chapter 4

DUAL RESPONSE METHODS.....130
4.1 INTRODUCTION.....130
4.2 VARIANCE FITTING APPROACH.....134
4.2.1 NUMERICAL EXAMPLE.....137
4.2.2 CONCLUDING REMARKS.....150
4.3 VARIANCE APPROXIMATION APPROACH.....152
4.3.1 MOTIVATION.....152
4.3.2 ALGORITHM STEPS AND IMPLEMENTATION ISSUES.....157
4.3.3 A NUMERICAL EXAMPLE.....170

Chapter 5
COMPARISONS, CONCLUSIONS, AND EXTENSIONS 194
5.1 CONFIRMATORY EXPERIMENTS 204
5.2 QUESTIONS FOR FURTHER RESEARCH 211
5.3 EXTENSION TO THE RANDOM SIGNAL FACTORS PROBLEM 212
5.3.1 JOB-SHOP EXAMPLE 214
5.3.2 PROBLEM RESOLUTION 217
5.4 CONCLUDING COMMENTS 219

BIBLIOGRAPHY 220

APPENDICES 225
APPENDIX A 225
APPENDIX B 228

VITA 232

LIST OF TABLES

1.1 Model Parameters 16
2.1 Experimental Design and Data for the First Subregion 39
2.2 Regression Analysis Results for the First Suibregion 40
2.3 Results of Exploratory Experiments Along the Gradient 41
2.4 Experimental Results for the Second Subregion 43
2.5 First-order Analysis Results for the Second Subregion 43
2.6 Experimental Results Along the Second Search Direction 44
2.7 Experimental Results for the Third Subregion 44
2.8 First-order Analysis Results for the Third Subregion 45
2.9 Experimental Results for the Third Search Direction 45
2.10 Experimental Results for the Fourth Subregion 46
2.11 First-order Analysis Results for the Fourth Subregion 47
2.12 Augmented Experimental Results for the Fourth Subregion 49
2.13 Second-order Analysis Results for the Fourth Subregion 51
2.14 Experimental Results for the Initial Subregion 57
2.15 First-order Model Results for the Initial Subregion 57
2.16 Experimental Results Along the First Fitted Gradient 59
2.17 Experimental Results for the Second Subregion 60
2.18 Model-fitting Results for the Second Subregion 60
2.19 Second-order Design and Experimental Results 62
2.20 Second-order Analysis Results 63
2.21 Experimental Results for the Second Subregion 65
2.22 Second-order Analysis Results for the Second Subregion 66
2.23 Experimental Results for the Third Subregion 68
2.24 Second-order Analysis Results for the Third Subregion 69
3.1 Experimental Results for the First Experimental Subregion 110

3.2 Fitted First-order S/N Model for the Initial Subregion 111
3.3 Results for the Exploratory Runs for the First Search Direction 112
3.4 Experimental Results for the Second Experimental Subregion 114
3.5 Fitted First-order S/N Model for the Second Subregion 114
3.6 Results of Exploratory Runs for the Second Search Direction 115
3.7 Experimental Results for the Third Experimental Subregion 115
3.8 Fitted First-order S/N Model for the Third Subregion 116
3.9 Results of Exploratory Runs for the Third Search Direction 116
3.10 Experimental Results for the Fourth Experimental Subregion 117
3.11 Fitted First-order S/N Model for the Fourth Subregion 117
3.12 Augmented Results for the Fourth Experimental Subregion 120
3.13 S/N Second-order Model Analysis Results (Fourth Subregion) 121
3.14 Experimental Results for the Fifth Experimental Subregion 123
3.15 S/N Second Order Model Analysis Results (Fifth Subregion) 124
3.16 Experimental Results for the Sixth Experimental Subregion 126
3.17 S/N Second Order Model Analysis Results (Fifth Subregion) 127
4.1 Phase 1 Data - First Experimental Subregion 139
4.2 Fitted First-Order Model - First Subregion 139
4.3 Exploratory Runs Data for First Search Direction 140
4.4 Phase 1 Data - Second Experimental Subregion 141
4.5 Second Fitted First-Order Model 142
4.6 Exploratory Runs Data for Second Search Direction 142
4.7 Phase 1 Data - Third Experimental Subregion 143
4.8 Third Fitted First-Order S/N Model 143
4.9 Phase 2 Data - Third Experimental Subregion 145
4.10 Second Order Model Analysis Results 146
4.11 Mean Response Second Order Model Analysis Results 147
4.12 Experimental Design and Results - First Subregion 169
4.13 First Fitted Composite Model 170
4.14 Experimental Results - Second Subregion 172
4.15 Second Fitted Composite Model 173
4.16 Experimental Results - Third Subregion 175
4.17 Third Fitted Composite Model 176
4.18 Experimental Results - Fourth Subregion 178

4.19 Fourth Fitted Composite Model 179
4.20 Experimental Results - Fifth Subregion 182
4.21 Fifth Fitted Composite Model 183
4.22 Augmented Design and Results - Fifth Subregion 185
4.23 Refitted Fifth Composite Model 186
5.1 Solutions to the Inventory Optimization Problem 196
5.2 Refitted Fifth Composite Model 198
5.3 Results of Confirmatory Trials ($Z=2^3$) 204
5.4 F-test results for H_0: $s^2{}_i=s^2{}_j$ vs H_a: $s^2{}_i<s^2{}_j$ 206
5.5 Service Time Distributions 212

LIST OF FIGURES

3.1 Crossed-array Diagram 81
5.1 JOB-SHOP EXAMPLE 211
SLAM Network - Static Model 223
SLAM Network - Nonstationary Model 227

Chapter 1

INTRODUCTION

1.1 SYNOPSIS

A problem of considerable interest in the analysis of complex stochastic systems is the estimation of operating levels for the controllable input factors that results in some optimum system response or performance level. A popular tool for conducting such studies where the complexity of the system prohibits an analytical derivation of a closed-form solution is computer simulation. In computer simulation the analyst attempts to capture the relevant facets of the system in a mathematical or logical model that is capable of representing the system's behavior under various operating policies. This model is then used to generate an artificial history of the system which, in turn, serves as a basis for drawing inferences about the operating characteristics of the actual system.

An important consideration in computer simulation studies is the issue of model validity, the level of accuracy with which the simulation model represents the real world system under study. Obviously model validity has a direct impact on the relevance, applicability, and optimality of any formulated solution or policy. This dissertation addresses a major cause of model validity problems in simulation optimization: the dissimilarity between the simulation model and the real system due to the dynamic nature of the real system which results from the presence of nonstationary stochastic processes within the system. For example, Hadley and Whitin (1963) note that the stochastic processes that generate demands and leadtimes for inventory systems continually change over time. The same is true even for the costs and items of interest in such systems.

In many simulation optimization studies, this transitory characteristic of the system is not considered during the search for an optimal solution as most contemporary optimum seeking search techniques are not designed to operate under unstable conditions brought about by the system's nonstationary behavior. Law and Kelton (1991) note that the use of classical statistical procedures in the analysis of the output processes of nonstationary systems is usually not accommodated. Instead static simulation models of the real system, i.e., models that do not reflect the transitory nature of the system, are utilized.

In reliability and quality control studies, it is known that optimizing with respect to the variance of the response is as important a concern as optimizing with respect to average performance response. Genichi Taguchi has been instrumental in the advancement of this philosophy in the area of manufacturing. He has focused on the issue of finding input settings robust to variation in the context of designing a consistent product. His work has resulted in what is now popularly known as the Taguchi Methods for robust parameter design. An application frequently cited as a successful implementation of Taguchi Methods in the manufacturing area is the case of the Ina Tile Company of Japan which experienced unacceptable levels of tile size variation attributed to the uneven temperatures within its kilns. The company addressed the problem not by improving or changing its kilns to maintain an even temperature profile (a costly alternative), but rather by finding a tile design via robust parameter design that was insensitive, or least sensitive, to the uneven temperatures (a relatively inexpensive strategy). They found that a change in lime content from 1% to 5% resulted in a considerable reduction (approximately 90%) in tile size variation (see Kackar (1985), and Taguchi and Wu (1980)).

Following Taguchi's philosophy of considering both response mean and variance simultaneously, the goal of this research is to devise a framework for finding optimum operating levels for the controllable input factors in a stochastic system that are insensitive to internal sources of variation. Specifically, the model validity problem in simulation optimization of nonstationary system behavior cited earlier is viewed as a major internal cause of system variation. This perspective clearly suggests that the variance of system response is an appropriate measure of the effect of nonstationarity. Thus a reasonable and desirable objective in establishing system operating conditions becomes one of finding solutions that optimize system response while being minimally affected by this source of variation; that is, we seek a set of robust operating conditions.

In this research the typical application of response surface methodology (RSM) to a simulation optimization problem with nonstationary process components is examined. Simplifying assumptions made on the simulation model to enable the use of RSM techniques are examined. The relaxation of these assumptions to address model validity leads to a modification of the RSM approach to properly handle the problem of optimization in the presence of nonstationarity. Taguchi's strategy and methods for enhancing product or process quality are then adapted and applied to this problem. Finally, dual-response RSM extensions of the Taguchi approach separately modeling the process performance mean and variance are considered and suitably revised to address the same problem. Comparisons of these approaches provide additional insights into the estimation of optimum operating conditions for complex stochastic systems.

A second cause of model validity problems is also considered: the random behavior of the supposedly controllable input factors to the system. Complete control over these inputs is usually assumed whereby the modeler or experimenter attempts to ascertain the quality of performance of the system operating under conditions defined by preselected levels for the inputs. Model validity becomes an issue when the uncontrollable stochastic nature of these inputs in the real world is not reflected in the simulation model. On the other hand, its inclusion in the model impacts the complexity of the analysis of the system. A resolution to this source of model invalidity is presented based on the methods described earlier.

1.2 OPTIMIZATION THROUGH COMPUTER SIMULATION

Computer simulation is the imitation of a real-world system or process using a computer. It is the process of designing a computer-based mathematical model of a real system and conducting experiments on that model. Thus computer simulation allows for the study of the behavior of complex systems with the objective of providing predictive capability regarding future system behavior under conditions defined by different sets of operating parameters and input policies. Subsequent to that goal is the frequent objective of system optimization, identifying the operating policy for the system that results in an overall optimum system performance. One strategy for pursuing this objective is to construct a regression metamodel to relate a set of controllable input variables defining the operating policy to some generated response of interest, and then to use that metamodel to find the levels for the input variables or factors that result in an optimal value for the response.

The role of computer simulation in the aforementioned objectives and analyses is the efficient generation of an artificial history of the system. This artificial history is the result of sampling experiments performed on a computer simulation model. The model is a computer-based abstract representation of the system consisting of a computer program which attempts to capture and emulate the logical, mathematical, and symbolic relationships between the entities and/or components of the system. Based on the assumption that the computer simulation model accurately depicts the system, designed experiments are performed on the model and output statistics are collected. These generated data are utilized to evaluate and explain the performance of the model given the test operating policies specified in the trials via statistical analysis. From these conclusions, inferences regarding true system behavior relative to operating policy are formulated. From these inferential results an optimal system operating policy is determined.

Simulation is an appropriate tool when analysis of the system through the use of a mathematical model instead of a physical model is required. As with any efficient working model of a system, a computer simulation model affords the decision-maker or experimenter valuable surrogate information about the system without the costs that would be incurred if the experimentation was performed on the actual system, or the even greater costs of constructing and experimenting with a previously nonexistent system. Among the mathematical models of the system, the computer simulation model is the model of choice when the complexity of the mathematics of the model renders the derivation of an exact, analytic solution highly intractable if not virtually impossible. Furthermore, simulation remains a viable analysis tool even when the complexity of the original problem prevents the construction of a meaningful mathematical description of the problem in the form of a set of mathematical expressions. This flexibility is due to the empirical nature of the approach whereby the contributions of the system's components and the effects of the policies or decisions are investigated through experimentation.

The simulation process encompasses not only the design and implementation of experiments on the model and the ensuing statistical analysis of the experimental results, but also the process of building the computer simulation model. Banks and Carson[1984] outline the steps involved in the development of computer simulation models. The first step is the gathering of information regarding the system, including data regarding the behavior of and interactions between the different components of the system. Once a sufficient understanding of the system is acquired, a conceptual model of the system is constructed. This conceptual model is a collection of assumptions regarding the components and the logical structure of the system, the hypothesized values for the model parameters defining the operating conditions within the system, and the controllable inputs defining an operating policy. This model serves as a blueprint containing specifications that are deemed integral for the proper emulation of the complex, real-world system . The third step involves the coding of an operational computer model to implement these specifications and the logic of the conceptual model.

It is to be noted that the model building process is not a linear three-step procedure moving progressively from the real system to the conceptual model, then finally to the computer simulation model. Rather it is an iterative process requiring successive refinements at each stage. The iterative step between the computer model and the conceptual model corresponds to model verification. Model verification is the determination of whether the logic and specifications of the conceptual model are properly reflected by the computer model. It consists of checking and debugging the computer program. The other refining iterative step, between the real system and the conceptual model, is model validation. It is intended to assure the veracity of the conceptual model; i.e., to establish that the conceptual model is an accurate as well as an adequate portrayal of the real system. The focus of this research is on this aspect of simulation model building. The significance of model validation relative to the process of identifying an optimal policy is highlighted. A significant problem associated with this procedure related to the presence of nonstationary stochastic processes within the system is discussed. This transitory behavior of stochastic systems is usually not incorporated in simulation models formulated for optimization purposes. Methods for handling the ramifications of directly addressing this model validity problem are presented.

1.2.1 CONCEPTUAL MODEL VALIDATION

As noted in the previous section, a critical issue in computer simulation is the level of accuracy with which the model represents the system under study. Law and Kelton[1991] define model validation as the process concerned with the determination of whether the conceptual simulation model is an accurate or valid representation of the system. They note that since the conceptual model is only an abstraction of the system, it will most likely be different from the system. With this assumption, the matter of model validity then becomes an issue of model adequacy, focusing on the severity of the differences between the original (system) and the imitation (conceptual model). The main concern is whether the severity of the differences affects the relevance or the applicability of the model-based conclusions to the system under scrutiny. As the experimentation and statistical analysis results are dependent on the computer simulation model's generated output, any serious divergence between the system and the conceptual model at the outset of the computer simulation model building process will invalidate or, at the very least, cast serious doubt on the pertinence of the derived inferences and decisions in relation to the system.

Much has been written on model validation. Comprehensive coverages of these works are provided in the following references: Balci (1987), Banks and Carson (1984), Feltner and Weiner (1985), Gass (1983), Gass and Thompson (1980), Law and Kelton (1991), Naylor and Finger (1967), Oren (1981), Sargent (1981,1988), Schellenberger (1974), Shannon (1975), and Van Horn (1971). A comprehensive bibliography of published work on validation is presented by Balci and Sargent (1984).

Naylor and Finger (1967) present a three step validation approach. In the first step, a conceptual model with high face validity is constructed. This model consists of system structure and data assumptions which, on the surface, appear reasonable to the intended model users and others knowledgeable about the system being studied. This is done through the use of all available existing information regarding the system, including general knowledge, existing theory, historical data, intuition, experience and expertise of individuals who have dealt with the real system, and results from similar studies.

The second step is to validate the structural assumptions on how the system operates, as well as the data assumptions for each of the system components. The structural assumptions are verified via actual comparison with the real system. Data assumptions, on the other hand, are evaluated using statistical analysis tools to determine if the generated data are in agreement with observed data from the system's past history.

The third step is to test the simulation model's ability to accurately predict future system behavior. It is made to operate at input settings that have been utilized in the real system in the past, and for which output data has been collected. The simulation model's generated output is statistically compared to the actual system data to determine how accurately the simulation model performed at those specified settings.

This three step approach is a widely accepted, logical, and systematic attempt at validating conceptual models. However, there are inherent problems with the process, particularly in the last two steps. In these steps, statistical comparisons are made between system parameter data and assumed model parameter data , and between system output data and model output data. These comparisons are made with the use of classical statistical analyses tools. Law and Kelton (1991) note that most output data from real world processes and systems are usually correlated. More importantly, the stochastic processes in most real systems are nonstationary. These phenomena tend to create problems with the use of statistical procedures that usually require independent and identically distributed observations. An even more significant obstacle is when the system under study is nonexistent, thereby precluding these data comparisons. These are just some of the more prominent innate problems that stress the fact that not only is model validation a very difficult process - it is also an inexact one. Law and Kelton[1991] conclude:

"There is no completely definitive approach for validating the model of a proposed system. If there were, there might be no need for a simulation model in the first place."

1.3 PROBLEM DEFINITION AND SCOPE

The issue of model validity is a critical consideration in the process of finding an optimal system policy. An invalid model could have serious repercussions on the relevance, applicability, and optimality of any policy relative to the real system. In this study, the dissimilarity between the actual system and the simulation model caused by the transient behavior of most real-world systems is addressed. As noted previously, the output processes for almost all real-world systems are nonstationary (see Law and Kelton (1991)). Given a nonstationary component stochastic process or subsystem within the real-world system, the output process for that subsystem, say $\{O_1, O_2, ...\}$, may not have a steady state distribution, with the distribution of the random variable O_j continually changing over time. Thus, the transitory behavior of the component subsystems impacts on the output process of the entire system.

The ramifications of nonstationarity on simulation modeling are addressed in two ways. In the first scenario, the modeller acknowledges or assumes the existence of nonstationarity in system behavior at the outset of the study and decides to directly confront the problem by incorporating this characteristic into the simulation model. The validity problem in this case is mainly one of estimation of the input parameters of the model. Due to the dynamic nature of the real-world process under consideration, the use of classical statistical methods, which rely on the observations being independent and identically distributed, is not accommodated. Law and Kelton (1991) further observe that there are numerous unresolved output analysis problems of this nature, and that methods available for resolving some of these estimation problems are difficult to apply due to their complexity. As a result, efforts to correctly incorporate nonstationarity in the simulation model result in other problems also impacting model validity.

The second scenario is one wherein the experimenter decides against reflecting the nonstationarity of the real-world processes in the simulation model. Optimal policies have to be derived under static or stationary conditions as methods for finding an optimal solution under transient operating conditions are nonexistent (Safizadeh (1990)). The impact of nonstationarity on such policies is that optimality is guaranteed only for certain static situations. Schmidt and Taylor (1970) note that when a supposedly optimal policy is subjected to varying conditions, its optimality becomes suspect warranting further investigation.

Examples in the literature dealing with optimization in the presence of nonstationarity are in concert with this view. Hadley and Whitin (1963) consider an inventory system with a time-varying rate of demand within a cost minimization context. Their proposed method of solution was a dynamic programming (DP) approach wherein at each stage of the analysis a static subproblem is solved. The main assumption is that the rate of demand can be treated as constant for judiciously defined time intervals. An end result of this sectioning of the planning horizon is that orders can be placed only at certain points in time. Given that this supposition is valid, their immediate problem was estimation and/or prediction of the rate of demand for those time intervals. Since the rates for different intervals will, in all likelihood, be different, the optimal solutions to the corresponding subproblems will likewise be different.

Schmidt and Taylor (1970) also examine a related case of the aforementioned inventory problem with dynamic rate of demand following an increasing linear trend. It is this problem that raised their concern about continued optimality of any solution over time. They propose to gauge the effect of changes in rate of demand as well as changes in mean lead time via regression analysis, with total cost being the dependent variable, and rate of demand and mean lead time being the regressor variables. In this analysis, it should be noted that the regressors are held at fixed values to generate each observation in the dependent variable. This approach therefore implies static system behavior.

In the above studies, the analyses entailed dealing with static versions of the problem. If they were undertaken using simulation models, the problem of model validity remains. Another notable problem, especially in the dynamic programming approach, is that of granularity --- how to choose discrete levels of continuous variables like time or rate of demand.

There are studies in the literature that have dealt with incorporating nonstationarity in the modeling of real-world systems. Hillestad and Carillo (1980) examined queuing models where the demand and service processes are nonstationary. They assumed that the rates for these processes vary according to some function known through a priori considerations. Gaver, Lehoczky, and Perlas (1976) considered queuing models of service systems with nonstationary demand. Miller, Stanton, and Crawford (1984) examine a nonstationary queuing simulation model. In all of these studies, assumptions concerning the transient behavior of the stochastic processes were needed to conduct the analysis of the queuing models. For example, assumptions regarding the functional form of the time dependent demand intensity are essential. Though reasonable, their results are appropriate only to the specialized cases considered. This observation suggests that a more flexible, empirical approach may be more advantageous to reducing the restricting impact of a priori assumptions. More importantly, in all of these cases, the analyses address the adequacy of performance of a given operating policy by observing its sensitivity to variations in operating conditions. The objective of finding an optimal policy, system optimization, is not addressed.

The problem of finding an optimal policy in the presence of nonstationarity remains mostly unresolved. This study addresses primarily one source of nonstationarity and the development of methods for estimating an optimal solution or operating policy for a system that is in a state of flux due to the presence of nonstationarity in its component stochastic processes. A concurrent objective is to formulate techniques that are less seriously impacted by the model validity problem of inaccurate or unreliable model input and output parameter estimation due to transient system behavior. This is especially significant in the case of terminating simulations, simulations whose run lengths are determined by the occurrence of some predetermined event. In such simulation trials, there is the possibility of terminating the simulation before steady-state conditions are attained. Even if the component stochastic processes are stationary, the initial bias conditions realized prior to achieving steady-state could result in nonstationary output data.

In this research the cause of nonstationarity is limited to the transient behavior of the means of the distributions that drive the stochastic processes. The functional forms of the distributions are assumed to remain the same throughout the course of the simulation trials while the mean or location parameter is essentially allowed to vary according to a specified probability distribution. This is a special case of the more general scenario wherein all the parameters of the distributions are allowed to vary. The next subsection details a simulation model of a continuous review inventory problem which will serve as a vehicle for illustrating the characteristics of the nonstationarity problem as well as the mechanisms of the proposed methods of solution presented in later chapters.

Following the development and discussion of proposed methods of solution, a different model validation problem involving the controllability of the input factors is discussed in more detail. A job-shop example is presented to illustrate the issue of model validity for this scenario and a reformulation of the input factors is suggested to enable the usage of the previously proposed methodologies.

1.3.1 ILLUSTRATIVE PROBLEM

Consider a continuous review inventory system where customers enter the system with an interarrival time that is distributed according to the exponential distribution with mean ζ_1. The amount of material required by a customer, called the demand amount, is normally distributed with a mean of ζ_2 and standard deviation of σ_2. If on-hand inventory is greater than this amount, the customer's entire order is filled and on-hand inventory is reduced accordingly. If on-hand inventory is less than the desired amount of material, the customer's order is partially filled by the entire amount available, reducing on-hand inventory to zero level. The difference between the required amount and the available amount is added to lost sales. If the current on-hand inventory level is zero at the arrival of a customer, his entire demand amount is added to lost sales.

In this model, the decision to reorder stock is based on the inventory or stock position which is the sum of on-hand inventory and on-order quantity. On-order quantity is the amount of material already on order or in transit from the supplier. In a continuous review system, the inventory position is monitored after each transaction or customer requirement has been considered. When this quantity drops to or below a certain predetermined level, called the reorder point (r), a prespecified quantity , termed the reorder quantity (Q), is placed on order to replenish the stock. The delivery lead time for reorders is normally distributed with a mean of ζ_3 and standard deviation σ_3. This system is commonly referred to as a (Q, r) system with the two fixed input factors r and Q forming an operating policy (see Hadley and Whitin (1963)).

The main objective of the analysis of this system is to find levels for r and Q that would minimize average total cost, which is the average of the sum of the ordering or set-up cost, carrying or holding cost, and the stockout or shortage cost within a relevant time period. The ordering cost is equal to the number of orders placed within the time period multiplied by a fixed per order cost. Holding cost is computed daily by taking the product of average daily on-hand inventory and a fixed per unit daily holding cost. Each day's average daily on-hand inventory is taken to be the average of the day's beginning and ending on-hand inventories. The holding cost for the relevant time period is computed by summing up the daily holding costs for the entire period. Stockout cost is computed as the product of the total units of lost sales within the relevant period and a fixed opportunity cost per unit of unsatisfied demand. These cost parameters, as well as the other input parameters to the simulation model, are summarized in the following table.

Table 1.1 Model Parameters

VARIABLE	DISTRIBUTION	NOMINAL VALUE
Interarrival Time	exponential(ζ_1)	$\zeta_1 = 0.20$ days
Demand Amount	$N(\zeta_2\ ,\ \sigma_2^2)$	$\zeta_2 = 100$ units $\sigma_2 = 10$ units
Lead Time	$N(\zeta_{3,}\ \sigma_3^2)$	$\zeta_3 = 15$ days $\sigma_3 = 2$ days

Ordering Cost	Holding Cost	Shortage Cost
$1000/order	$0.10/unit/day	$10/unit

The nominal values for ζ_1, ζ_2, and ζ_3 in the above table are values for the most likely static case of the inventory model. Hadley and Whitin (1963) state that for most inventory systems, the demand and lead time processes change over time. For illustrative purposes, we introduce nonstationarity by treating the means of the distributions for interarrival time, demand amount, and lead time as random variables. In particular, we assume that

$$\zeta_1 \sim N(.20,.05^2)$$

$$\zeta_2 \sim N(100,10^2)$$

$$\zeta_3 \sim N(15,2^2).$$

This is a modification of the inventory model considered by Schmidt and Taylor (1970) wherein system nonstationarity is brought about by letting the demand arrival rate be a linear function of time.

The inventory system just described is modeled using the network simulation modeling package SLAM by Pritsker (1986). The initial on-hand inventory level and inventory position are set at 20,000 units. For each experimental run with a specified set of values for r and Q, the length of the simulation is 50,000 days. Assuming that there are five working days per week, and four weeks per month, the costs are aggregated over a 20-day cycle providing monthly total cost data for 2500 months. The results for the first 500 months are discarded to eliminate start-up bias. The remaining 2000 data points comprise the output data stream for use in the analysis. The average of these observations is the usual statistic of interest. In a static model, i.e, ζ_1,ζ_2, and ζ_3 set to their nominal values, these observations would be identically distributed but correlated. In this model, these generated data points are neither independent nor identically distributed.

Though the observations within each experimental run may be correlated, observations between experimental trials or specified combinations of the input factors are independent due to the use of independent random number streams. In this strategy, a different set of seed values is assigned to the random number stream generators for each execution of the simulation model. For a given set of seeds assigned to a particular experimental run, the resulting random number streams are unique to that experimental run. These random number streams, in turn, drive the stochastic process generators.

1.4 PROPOSED METHODOLOGY

1.4.1 GENERAL APPROACH

Model validation problems pertaining to nonstationarity cannot be resolved absolutely since the sources of these problems are for the most part uncontrollable and unpredictable. Nonstationary system behavior results in problems with system parameter estimation which are reflected in the relevance and optimality of any derived solution or operating policy. However, failure to consider nonstationarity in model construction places the validity of the model and the elicited solutions in doubt. With the inclusion of transient system behavior, system instability results obstructing the systematic identification of an optimal solution using conventional optimization procedures. It is imperative, therefore, that these model validation issues be addressed.

Fundamental to developing a general approach to the problem of nonstationarity is understanding its overall effect on system performance. It will be shown in Chapter 2 that the modeling of the nonstationarity in the distributions of system components results in an increase in system response variance. Thus nonstationarity can be viewed as an added internal source of performance variation for a given operating policy. This perspective clearly suggests that variance of system response is an appropriate measure of the effect of nonstationarity.

While the original objective of system optimization is to find an operating policy that will result in optimum system response, the consistency of its performance at the desired optimum level is subject to the variance-generating effect of nonstationarity. Thus another important consideration in establishing the best system operating policy is one of finding solutions that are minimally affected by this source of variation. What we therefore seek are robust as well as optimal operating conditions. This observation coincides with the so-called "Taguchi Methods" for off-line quality improvement, also known as robust parameter design (RPD).

The underlying theme of RPD is the need to simultaneously consider the response variance and the response mean in specifying an operating policy. Its objective is one of finding a solution that optimizes mean system response while being minimally affected by the internal sources of variation. We adopt this philosophy in formulating a general approach to the problem of optimization in the presence of nonstationarity.

In the proposed approach the effects of system instability produced by nonstationarity are minimized by limiting the solution space to the set of robust solutions. These are policies that are least sensitive to the destabilizing effect of nonstationarity. Sensitivity of a given policy to nonstationarity is appropriately measured by the variance of system response for this policy. As the system evolves in the presence of nonstationarity, these solutions will represent the most consistent results in system performance; i.e., they remain "valid" solutions. From this set of "valid" solutions, the policy producing the best results with regard to mean response is chosen as the optimal operating policy.

Thus with this approach, the model validity problems caused by nonstationarity are neither resolved directly nor ignored. Instead of attempting to improve the validity of a model, we focus on improving the quality or validity of the candidate solutions or policies.

1.4.2 SOLUTION TECHNIQUES

The procedures studied are empirical decision-making tools based on the definition and exploration of the response surfaces for the mean and variance responses. In Chapter 2, we consider the techniques of RSM as applied to the usual optimization problem of minimizing mean response. Two scenarios, one relating how the nonstationarity problem is typically handled, and the other showing a possible resolution are discussed. These methods are both applied to the continuous review inventory problem cited earlier. In the process, the problem of increased response variance due to the presence of nonstationarity is examined in detail.

In Chapter 3, we present the strategy and tactics of robust parameter design (RPD) or the Taguchi Methods as they are currently interpreted in this country. The merits as well as the criticisms of RPD are examined, and a modified version of RPD combining the best features of RPD and RSM is presented.

In Chapter 4 we discuss a natural extension of RPD's philosophy of simultaneously considering the response mean and variance in the optimization process to the dual response philosophy of RSM. In this approach, the two performance measures of interest are considered separately. We provide a revised sequential RSM dual response strategy that best embodies this general approach wherein the initial search phase of RSM utilizes the variance response to identify robust policies. The second phase of this strategy involves the mean response in the search for the best robust policy. In formulating this approach, necessary revisions to the original form of the mean-variance dual response methods of RSM are offered, including the addition of some of the sound tactics of RPD.

Chapter 2

RESPONSE SURFACE METHODS

2.1 INTRODUCTION

Response Surface Methodology (RSM) is a collection of mathematical optimization and statistical experimentation methods used in continuous parameter stochastic optimization problems. For this class of problems, the main objective is the estimation of the operating levels for the real-valued input parameters to some system or process that results in optimal value(s) for that system's or process' response(s) of interest. The mathematical optimization methods are calculus-based search procedures, working mainly with the gradient of some response function that relates the response(s) to the input(s). The statistical component of RSM is the foundation for these optimization techniques since it provides for the estimation of the response function.

RSM was introduced and developed by Box and his co-workers in the 1950s and 1960s as a set of experimental design and analysis tools with primary application in the chemical, food, and textile processing industries (Box and Wilson (1951)). Since that time its application base has expanded to include other areas such as computer simulation. This empirical approach to problem-solving puts a high premium not only on the estimation of optimal operating levels for the input parameters or factors, but also on understanding the mechanisms of the process or system under study. The extensive applications of the methodology are chronicled in an article by Myers, Khuri, and Cornell (1989). Textbooks by Box and Draper (1987), Khuri and Cornell (1987), and Myers (1976) provide a comprehensive coverage of both the traditional experimental design and analysis procedures that comprise RSM. Extensive surveys of the literature on RSM are found in Hill and Hunter (1966), Mead and Pike (1975), and Myers (1988).

The next section presents a general overview of the general RSM strategy. The illustrative example introduced in Chapter 1 is considered in the third section to demonstrate how RSM can be applied to the optimization problem involving transitory operating conditions. The final section summarizes the advantages and disadvantages of this approach.

2.2 METHODOLOGY

The basic assumption in RSM is that for a given system there is a set of k input factors

$$\underline{\xi}' = (\xi_1, \xi_2, \ldots, \xi_k) \tag{2.1}$$

that affect the system response of interest η via some relationship $g(\underline{\xi})$; that is,

$$\eta = g(\xi_1, \xi_2, \ldots, \xi_k). \tag{2.2}$$

It is also assumed that these factors can be controlled or fixed at preselected levels with negligible error and that the functional form of $g(\underline{\xi})$ is unknown and complex. Rather than working directly with the natural variables $\underline{\xi}'$, the study is conducted in terms of the design variables

$$\underline{x}' = (x_1, x_2, \ldots, x_k) \tag{2.3}$$

where $\underline{x}'$ is a linear transformation of $\underline{\xi}'$. The underlying relationship can then be written as

$$\eta = f(\underline{x}). \tag{2.4}$$

Another important assumption in RSM is that no matter how complicated the form of $f(\underline{x})$ may be, it can be closely approximated by some convenient graduating function over restricted subregions of the input factor space. In computer simulations studies, these functions are termed metamodels - simple and explicit algebraic representations of the simulated response. In RSM these graduating functions are low-order polynomials, usually of order one or two. The rationale for the use of polynomial approximations is based on the truncated Taylor series approximation of $f(\underline{x})$. The order of the graduating polynomial used typically depends on the phase of the analysis currently being conducted. First-order metamodels find extensive applications in the initial stages of the investigation where the experimenter seeks to identify the region of optimum response. Once identified, graduating functions of order two generally serve as the basis for the optimization phase of the study. The following two sections present the basic components of the analysis in terms of these two models.

2.2.1 FIRST-ORDER ANALYSIS

In a typical RSM study the subregion in the input variables containing the optimum operating conditions is unknown, and the experimenter's initial objective is to determine its location. The investigation frequently begins in a region chosen on a "best-guess" or heuristic basis. This region is often remote from the region of optimum response, and, through a series of sequential experiments, additional subregions are systematically explored in an effort to identify the optimum subregion. When there exists little curvature in the true surface for the subregion under study, which is frequently the case in the initial stages, the first-order polynomial graduating function

$$\eta = \beta_0 + \sum_{i=1}^{k} \beta_i x_i \tag{2.5}$$

frequently provides an adequate approximation of the surface. The observed responses, y, will, in all likelihood, differ from the true response η due to random or experimental error. Consequently, the response function is written as

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon, \tag{2.6}$$

where the random error, ε, is assumed to have an zero expectation, and is used primarily to account for the difference between the true and observed responses. The true response η is often referred to as the mean response given an input factor combination $\underline{x}$, since the expected value of y is η. Likewise, the model denoted by equation (2.5) is called the mean response (surface) model. Equation (2.6) can be written in matrix form as

$$\underline{y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}. \tag{2.7}$$

These models are used to estimate the direction of steepest ascent or gradient direction. Additional observations are then taken along this gradient to identify the next region for exploration. The use of first-order graduating functions in this manner continues until the model fails to adequately describe the response surface or the optimum region is tentatively identified.

Experimental plans for the estimation of the graduating polynomials used in RSM are an important component of the methodology. Plans that allow for the estimation of the linear and interaction coefficients of first-order response surface models are termed first-order designs. They consist of n systematically chosen settings or combinations for the input variables, with each variable set to at least two levels during the course of the trials. The matrix

$$\mathbf{D} = \begin{bmatrix} x_{11} & x_{12} & . & . & x_{1k} \\ x_{21} & x_{22} & . & . & x_{2k} \\ . & . & . & & . \\ . & . & & . & . \\ x_{n1} & x_{n2} & . & . & x_{nk} \end{bmatrix},$$

where x_{ij} denotes the *ith* level of the *jth* input variable, is termed the design matrix and completely specifies the n>k experiments or factor settings at which observations are to be taken. Note that the *ith* row of **D**,

$$\begin{bmatrix} x_{i1} & x_{i2} & . & . & x_{ik} \end{bmatrix},$$

represents the *ith* experimental run or input factor combination. Also note that the matrix **X** in (2.7) is defined in terms of **D** as

$$\mathbf{X} = [\mathbf{1} \vdots \mathbf{D}].$$

Of particular importance in the estimation of the aforementioned first-order models is the class of orthogonal designs. These designs are characterized by a design matrix whose column vectors are orthogonal or at right angles with each other. This property allows for the independent estimation of the effect of each input factor relative to the other factors resulting in uncorrelated estimates of the model parameters. It also yields parameter estimates with minimum variance. The theoretical development of this design class was first considered by Rao (1947). The most commonly used members of this design class are the 2^k factorial and 2^{k-p} fractional factorial designs.

The 2^k factorial design plans consist of k input factors each set at two levels. It includes all possible combinations for the factor settings thus yielding 2^k design points or factor settings. These points, in turn, permit the estimation of the coefficients of all linear and interaction terms. The two levels for each factor in the natural variables are linearly transformed or coded to -1 and +1, with +1 corresponding to the higher value. A 2^{k-p} fractional factorial design, on the other hand, is a $(1/2)^{p}th$ fraction of the full factorial design. Fractional factorial designs are more economical in terms of the number of experimental runs, but do not provide for the estimation of all first order model coefficients. This general design class also possesses the added property of rotatability; i.e., the variance of predicted response is constant at locations equidistant from the center of the design subregion.

Other commonly used classes of orthogonal first-order designs are simplex designs and Plackett-Burman designs. Comprehensive coverages of these designs are found in Box and Hunter (1957,1961), Box (1952), Plackett and Burman (1946), and in numerous experimental design texts (see, for example, Kempthorne (1979), Myers (1976)).

The ordinary least squares (OLS) estimators of

$$\underline{\beta}' = [\beta_0 \ \beta_1 \ \cdots \ \beta_k],$$

in (2.7) are

$$\underline{b}_{OLS} = [b_0 \vdots \underline{b}']' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underline{y}. \qquad (2.8)$$

Under the assumption that

$$E(\underline{\varepsilon}) = \underline{0} \qquad \text{and} \qquad Var(\underline{\varepsilon}) = \sigma^2\mathbf{I},$$

the variance-covariance matrix for $\underline{b}_{OLS}$ is

$$\mathrm{Var}(\underline{b}_{OLS}) = \sigma^2 (X'X)^{-1},$$

Substitution of (2.8) in (2.5) yields the estimated or fitted graduating polynomial

$$\hat{y} = b_0 + \underline{x}'\underline{b}. \tag{2.9}$$

From this fitted model, the direction of steepest ascent or gradient direction is determined by the vector of the estimated main or linear effects $\underline{b}$. Additional exploratory experiments are conducted along the gradient path until the data exhibits a lack of improvement in the response. Based upon these results a new and more promising subregion is identified, one that is centered around that point on the gradient path that resulted in the current best response. This iterative process then restarts with a new set of experiments utilizing a first order experimental design in this new subregion to identify another search direction or gradient. The first order phase terminates when the regression coefficients become insignificant, at which point the gradient approximates the null vector indicating that minimal improvement in the response will result from any movement away from the current subregion. Thus the subregion containing the optimal solution is tentatively identified, and the second order analysis phase commences.

2.2.2 SECOND ORDER ANALYSIS

This phase of the analysis assumes that the optimal subregion has been reached and identified. The goal at this stage is to locate the stationary point and to characterize the nature of that set of operating conditions. Hence a more elaborate exploration of the optimal subregion is undertaken, with a second-order model being specified to account for possible curvature in the underlying functional relationship between the inputs and the response within that subregion. In this phase the second-order model

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j=i}^{k} \beta_{ij} x_i x_j + \varepsilon \qquad (2.10)$$

is postulated. This model can be rewritten as

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'\mathbf{B}\underline{x} + \varepsilon$$

with

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \cdots & \frac{1}{2}\beta_{1k} \\ & \beta_{22} & \cdots & \frac{1}{2}\beta_{2k} \\ & & \ddots & \vdots \\ \text{sym} & & & \beta_{kk} \end{bmatrix}$$

containing the quadratic and interaction coefficients. The corresponding general linear model formulation is given by

$$\underline{y} = \mathbf{X}\underline{\beta}^* + \underline{\varepsilon}$$

with

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} & x_{11}^2 & \cdots & x_{1k}^2 & x_{11}x_{12} & \cdots & x_{1,k-1}x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} & x_{21}^2 & \cdots & x_{2k}^2 & x_{21}x_{22} & \cdots & x_{2,k-1}x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nk} & x_{N1}^2 & \cdots & x_{Nk}^2 & x_{N1}x_{N2} & \cdots & x_{N,k-1}x_{Nk} \end{bmatrix}$$

and

$$\underline{\beta}^{*\prime} = \begin{bmatrix} \beta_0 & \beta_1 & \cdots & \beta_k & \beta_{11} & \cdots & \beta_{kk} & \beta_{12} & \cdots & \beta_{k-1,k} \end{bmatrix}.$$

An appropriate second order experimental design is then selected for the OLS estimation of the linear, quadratic, and interaction coefficients of (2.10).

Second order experimental designs require at least three levels for each factor to accommodate the estimation of the coefficients of the quadratic terms in the model. The most popular class of second order designs is the class of central composite designs (CCD) developed by Box and Wilson [1951]. These experimental plans are five-level designs consisting of a full factorial design, or a fraction thereof, augmented by axial design points (with two points on each design variable axis located symmetrically about the origin), and center runs as shown in the following design matrix.

$$\begin{array}{cc} & \begin{array}{cccc} x_1 & x_2 & \cdots & x_k \end{array} \\ \mathbf{D} = & \begin{bmatrix} \underline{\pm 1} & \underline{\pm 1} & \cdots & \underline{\pm 1} \\ \alpha & 0 & \cdots & 0 \\ -\alpha & 0 & & 0 \\ 0 & \alpha & \cdots & 0 \\ 0 & -\alpha & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha \\ 0 & 0 & & -\alpha \\ \underline{0} & \underline{0} & \cdots & \underline{0} \end{bmatrix} \end{array} \begin{array}{c} \text{Factorial or fractional factorial design} \\ \text{Axial design point} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \text{Axial design point} \\ \\ \text{Center Runs} \end{array}$$

A CCD involving a fractional factorial is sometimes referred to as a small composite design (SCD). These plans typically substitute less comprehensive first order designs such as fractional factorials and Plackett-Burman designs in place of the full factorial component of the CCD to reduce the number of design points required. Other popular classes of designs are the class of Box-Behnken designs developed by Box and Behnken [1960], and 3^k factorial and 3^{k-p} fractional factorial designs. Both of these experimental plans require just three levels for each input factor, but generally require more design points than the CCD or the SCD.

After a second order experimental design has been chosen, additional experimental runs are conducted. The collected response data are then used for the least squares estimation of (2.10) which leads to the fitted model

$$\hat{y} = \hat{\beta}_0 + \underline{x}'\hat{\underline{\beta}} + \underline{x}'\hat{\mathbf{B}}\underline{x}.$$

From this fitted quadratic function the stationary point is identified through the use of differential calculus. This is accomplished by taking the derivative of the fitted response with respect to $\underline{x}$ and setting it equal to zero,

$$\frac{\partial \hat{y}}{\partial \underline{x}} = \hat{\underline{\beta}} + 2\hat{\mathbf{B}}\underline{x} = 0,$$

and then solving for $\underline{x}$ to obtain the stationary point

$$\underline{x}^* = \frac{-\hat{\mathbf{B}}^{-1}\hat{\underline{\beta}}}{2}.$$

The analysis proceeds with an investigation of the nature of the stationary point. Second-order optimality tests based on second partial derivatives of the fitted model, if applicable, could be used to verify that the stationary point represents the optimal solution. If these fail, other methods to explain the nature of the stationary point are considered. One such procedure is the method of canonical analysis developed by Box and Wilson (1951). This method is used to locate the optimal point when it is contained within the current subregion. The first step in this method involves the translation of the response function from the original design space to a new coordinate system centered around the stationary point. The response function is rewritten in terms of canonical variables which are linear combinations of the original design variables. The axes of the new coordinate system are the principal axes of the contour system about the stationary point. If the new expression for the response function, called the canonical form, reveals that the stationary point is a local maxima or minima within the current experimental region, then an estimate of the true optimal point has been identified and the corresponding solution can be derived via the reverse transformation from the canonical variables to the design variables.

If the stationary point is found to be a saddlepoint or one that lies outside the current experimental region, the method of ridge analysis can be used to identify the best solution within the current region. This method was first studied by Hoerl (1959) and later modified for use in RSM by Draper (1963). With this procedure, the maximum (or minimum) fitted responses on hyperspheres interior to the experimental region of varying radii are determined using Lagrange multipliers, together with the corresponding solutions which give rise to these results. From this set of points the best solution is chosen. This course of action is taken when the experimenter decides to stay within the current experimental region. Should he decide that the true nature of the response surface is accurately depicted by the fitted second order model within the current region and that the optimal solution is not presently contained within the current region, he could then undertake additional exploratory experimental runs either to define a new and more promising experimental region, or to expand the current region in an attempt to contain the true stationary point. Once this is accomplished, the second order analysis phase recommences.

2.3 RSM APPLIED TO THE CONDITIONAL NONSTATIONARITY PROBLEM

In this section the inventory system example introduced in Chapter 1 is analyzed using the techniques of RSM. Two approaches to the problem are discussed. The first approach presents a simplified solution to the conditional nonstationarity problem to allow conventional usage of RSM tools and techniques. This straightforward application of RSM serves to illustrate the methodology as well as to provide a solution that will serve as a benchmark for judging other solutions.

The second approach presents a solution procedure that directly addresses the nonstationarity problem. The motivation behind this approach, as well as issues relating to the feasibility of using RSM in this situation, are discussed and resolved. The continuous review inventory model is analyzed using this approach.

2.3.1 STATIC MODEL APPROACH

Consider the optimization problem described in Chapter 1 of the continuous review inventory model which is complicated by the demand interarrival time, demand amount, and delivery lead time processes being nonstationary. That is, suppose that the mean interarrival time (ζ_1), the mean demand amount (ζ_2), and the mean lead time (ζ_3) are transitory parameters, varying randomly across time. As noted in the previous chapter, we assume that these "parameters" are random variables and are distributed as follows:

$$\zeta_1 \sim \text{Normal}(.20,.05^2)$$

$$\zeta_2 \sim \text{Normal}(100,10^2)$$

$$\zeta_2 \sim \text{Normal}(15,2^2).$$

Let $\underline{\zeta}$ denote the random vector of these mean parameters, and let $\underline{\mu}_\zeta$ be the vector of expected values of these parameters; i.e.,

$$E[\underline{\zeta}] = \underline{\mu}_\zeta' = [.20,\ 100,\ 15].$$

The common simulation optimization scenario for RSM is one dealing with stationary processes. It would seem that the added nonstationary feature of the model would render the direct application of RSM infeasible. This issue together with other ramifications of the nonstationary nature of the system's component processes are considered here.

Consider the output process with the response of interest being the total monthly cost Y. The observations generated by the system will be the time series

$$\{y_{ij}\ ,\ j = 1, ..., q\}$$

where the model is observed for q months given an operating policy defined by the *ith* design setting, $\underline{x}_i$. In the inventory example, this would correspond to the *ith* combination of values for the controllable input factors,

$$\underline{x} = [x_1, x_2]$$

where x_1 is the reorder point and x_2 is the reorder quantity in the coded design variables.

A major concern with this process arises from the nature of the output processes. Specifically, since the internal component processes are nonstationary, the resulting output processes must also be nonstationary. This is evident because the environmental conditions that led to the *jth* observation are unique to that observation. Each observation is a product of a different set of environmental conditions, each of which are, in turn, defined by different combinations for the rates of the component stochastic processes that comprise the model. For example, at some point in time, $t = t_0$, the mean interarrival time would be $\zeta_1(t_0)$, the mean demand amount $\zeta_2(t_0)$, and the mean leadtime equal to $\zeta_3(t_0)$. These values, in all likelihood, would be different for $t = t_k$ due to the fact that they are random variables. Thus each y_{ij} is uniquely distributed relative to the other observations, as it is conditioned on a unique set of environmental conditions that occurred within the relevant time period. In other words, the y_{ij} are not identically distributed random variables. For the *jth* observation,

$$E(y_{ij}) = E(y_{ij} | \{\zeta_1\}_j, \{\zeta_2\}_j, \{\zeta_3\}_j) = E(y_{ij} | \underline{\zeta_1}_j, \underline{\zeta_2}_j, \underline{\zeta_3}_j),$$

where $\{\zeta_l\}_j$ denotes the set of realizations for ζ_l during the *jth* month or time period. The notation is simplified by referring to each $\{\zeta_l\}$ as ζ_l.

Now consider y_{ij} for a specific j, say $j = j_0$. The value of $\underline{x}$ that would result in an optimal (minimum) value for $E(y_{i0})$, which we will denote by $\underline{x}_0$, will be uniquely optimal with respect to y_{i0}. For any other j the optimal solution would most likely be different.

Going back to what was referred to earlier as the typical case or scenario for RSM applications, the y_{ij} in the typical case (also known as the stationary model) would be identically distributed, having common expectation. Thus an optimal solution relative to one y_{ij} will be optimal relative to the rest, not dependent on a particular month or relevant time period.

Therefore, the main concern in the non-identically distributed y_{ij} case (or the nonstationary model) is that we have a different or separate optimization (minimization) problem at each point in the time series. Likewise, we have different optimal solutions, each one optimal only with respect to a particular random variable y_{ij} which is conditionally distributed. Hence, any one of these solutions is not guaranteed to remain optimal as the system evolves across time. Another consideration is that there exists an infinite number of possible combinations for the operating conditions given that the ζ_i are real-valued random variables. Enumeration of all possible optima is thus not practical nor feasible.

To provide a suitable venue for a traditional RSM analysis, it is desirable to have homogeneity among the distributions of the y_{ij} to ensure overall, time-independent optimality of a solution. One approach is to transform the transitory component processes into stationary ones by selecting a suitable combination of values for the time-varying rates, and hold them constant at these values for the duration of the simulation run. The logical choice for this combination of settings would be the vector of expectations for the rates, $\underline{\mu}_{\zeta}$. Under this scheme, the problem is converted to a typical RSM application in simulation optimization. The next subsection illustrates the analysis under this simplifying assumption.

2.3.1.1 PHASE 1

The first phase starts with the choice of an experimental plan for the estimation of the first-order graduating function. For the inventory problem under consideration, the natural variables are

$$\xi_1 \equiv \text{reorder point}$$

and

$$\xi_2 \equiv \text{reorder quantity.}$$

The first order experimental design selected was a 2^2 factorial design augmented with two center runs. Next an arbitrary starting point for the search process was chosen, with ξ_1 and ξ_2 both set equal to 1000 units. This point defines the center of the current experimental subregion. The natural variables are transformed into design variables using the coding convention

$$x_i = 2\left(\frac{\xi_i - \bar{\xi}_i}{d_i}\right), \tag{2.11}$$

where d_i is the spacing between the high and low levels on the natural variable ξ_i defining the width of the experimental subregion along the axis of that variable. The spacing was arbitrarily set to 1000 units for both the reorder point and reorder quantity. Thus d_i=1000 (i=1, 2) and $\bar{\xi}_i$=1000 (i=1, 2). The experimental design specifying the levels for the input variables in both the natural variables and the coded or transformed variables, and the observed responses for each of the experimental runs are given in the following table. The response y_i at each design point represents average monthly total cost.

Table 2.1 Experimental Design and Data for the First Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1	-1	500	500	91110
-1	1	500	1500	84831
1	-1	1500	500	81705
1	1	1500	1500	70922
0	0	1000	1000	80477
0	0	1000	1000	80708

The first order response model was fitted using the statistical analysis software package MINITAB. Results of the regression analysis are summarized below.

Table 2.2 Regression Analysis Results for the First Subregion

```
The regression equation is
y = 81625 - 5828 x1 - 4265 x2

Predictor        Coef        Stdev    t-ratio        p
Constant      81625.5        679.0     120.21    0.000
x1            -5828.5        831.6      -7.01    0.006
x2            -4265.5        831.6      -5.13    0.014

s = 1663          R-sq = 96.2%

Analysis of Variance

SOURCE        DF          SS          MS         F        p
Regression     2   208663600   104331792     37.71    0.007
Error          3     8299451     2766483
Total          5   216963040
```

The fitted first order model is

$$y = 81625.5 - 5828.5x_1 - 4265.5x_2$$

with standardized gradient vector

$$\underline{\hat{\beta}}_{\bullet} = \underline{b}_{\bullet} = \begin{bmatrix} -0.8070 \\ -0.5906 \end{bmatrix}.$$

This vector identifies the search direction or path along which additional observations are to be taken at the experimental points defined by

$$\underline{x}_r = \underline{x}_0 - \underline{b}_0 r. \tag{2.12}$$

Here the variable $\underline{x}_0$ represents the current design center and r is the number of steps of unit length taken along the axis of the gradient vector in the appropriate direction. In this instance, we take the negative gradient direction as we are minimizing the response. Starting with r set equal to one unit, we identify additional points by setting r to successive powers of 2. Additional experimental runs are conducted at these points to identify a more promising region. The results of this search procedure are summarized in the Table 2.3.

Table 2.3 Results of Exploratory Experiments Along the Gradient

r	x_1	x_2	ξ_1	ξ_2	y_i
1	0.8070	0.5906	1403.5	1295.3	74306
2	1.6140	1.1812	1807.0	1590.6	68624
4	3.2279	2.3623	2614.0	2181.2	57987
8	6.4558	4.7246	4227.9	3362.3	39114
16	12.9117	9.4492	7455.8	5724.6	15347
32	25.8234	18.8985	13911.7	10449.2	24092

An examination of these results indicates that the minimum response along the gradient direction occurs near r ≈ 16, which corresponds to the solution

$$\underline{\xi} = \begin{bmatrix} 7500 \\ 5700 \end{bmatrix}.$$

This solution then becomes the center of the new experimental subregion, and another iteration of phase one commences in an effort to find a better subregion. This sequential procedure terminates when the estimated model coefficients from the regression analysis are statistically insignificant, and any attempt to make them significant within the current subregion would cause the model fit to be inadequate.

For our example, three iterations of the search procedure were undertaken with four subregions being identified. The experimental data from the factorial runs, the results of the ensuing first-order regression analyses, and the observations taken along the fitted gradients are given in the following tables.

Table 2.4 Experimental Results for the Second Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1	-1	7000	5200	17415
-1	1	7000	6200	17020
1	-1	8000	5200	13724
1	1	8000	6200	13902
0	0	7500	5700	15346
0	0	7500	5700	15397

Table 2.5 First-order Analysis Results for the Second Subregion

```
The regression equation is
y = 15467 - 1702 x1 - 54.2 x2

Predictor        Coef         Stdev    t-ratio        p
Constant      15467.3          78.5     197.02    0.000
x1           -1702.25         96.15     -17.70    0.000
x2             -54.25         96.15      -0.56    0.612

s = 192.3        R-sq = 99.1%

Analysis of Variance

SOURCE        DF          SS          MS         F        p
Regression     2    11602392     5801196    156.88    0.001
Error          3      110935       36978
Total          5    11713326
```

Table 2.6 Experimental Results Along the Second Search Direction

r	x_1	x_2	ξ_1	ξ_2	y_i
1	1	0	8000	5700	13778
2	2	0	8500	5700	12852
4	4	0	9500	5700	12301
8	8	0	11500	5700	15341

Table 2.7 Experimental Results for the Third Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1	-1	9000	5200	11875
-1	1	9000	6200	12626
1	-1	10000	5200	12589
1	1	10000	6200	13181
0	0	9500	5700	12428
0	0	9500	5700	12372

Table 2.8 First-order Analysis Results for the Third Subregion

```
The regression equation is
y = 12512 + 317 x1 + 336 x2

Predictor        Coef        Stdev     t-ratio        p
Constant      12511.8         50.2      249.11    0.000
x1             317.25        61.51        5.16    0.014
x2             335.75        61.51        5.46    0.012

s = 123.0        R-sq = 94.9%

Analysis of Variance

SOURCE        DF          SS          MS          F        p
Regression     2      853502      426751      28.19    0.011
Error          3       45408       15136
Total          5      898911
```

Table 2.9 Experimental Results for the Third Search Direction

r	x_1	x_2	ξ_1	ξ_2	y_i
0.0	0.00	0.00	9500.0	5700.00	12372
0.5	-0.34	-0.36	9328.3	5518.29	12198
1.0	-0.69	-0.73	9156.6	5336.57	12135
2.0	-1.37	-1.45	8813.2	4973.15	12160
4.0	-2.75	-2.91	8126.4	4246.30	13334

Table 2.10 Experimental Results for the Fourth Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1	-1	8700	4800	12246
-1	1	8700	5800	12440
1	-1	9700	4800	12055
1	1	9700	5800	12598
0	0	9200	5300	12099
0	0	9200	5300	12080

A summary of the results of the regression analysis for this experimental subregion is given in Table 2.11.

Table 2.11 First-order Analysis Results for the Fourth Subregion

```
The regression equation is
y = 12253 - 8.2 x1 + 184 x2

Predictor        Coef        Stdev    t-ratio        p
Constant      12253.0         78.5     156.15    0.000
x1              -8.25        96.10      -0.09    0.937
x2             184.25        96.10       1.92    0.151

s = 192.2       R-sq = 55.1%

Analysis of Variance

SOURCE         DF          SS          MS          F         p
Regression      2      136064       68032       1.84     0.301
Error           3      110827       36942
Total           5      246892
```

The above table shows both estimated model coefficients to be insignificant at a significance level of 0.10. Additionally, model fit is poor as evidenced by the r^2 value of .551 and the F-test for the regression. These results signal the presence of substantial curvature in the response surface. Any expansion of the present subregion size to highlight the linear or first order effects would further adversely affect the quality of fit. On the other hand, reducing the region size to improve the quality of fit would likely result in the relevant estimated first-order model coefficients remaining statistically insignificant. Hence, phase one was terminated at this stage.

2.3.1.2 PHASE 2

This phase of the analysis begins with the estimation of the second-order model in (2.10) to describe the average monthly inventory costs over the last subregion examined in the previous section. This region is defined by

$$\{(\xi_1,\xi_2) \ni \xi_1 \in [8700,9700]\ ,\ \xi_2 \in [4800,5800]\},$$

where ξ_1 and ξ_2 denote the reorder point and reorder quantity, respectively, in the natural variables. The previous experimental design, a 2^2 factorial augmented with two center points (and the observations that resulted by applying that design) is further augmented by additional experimental runs conducted along the axes of the design variables, termed axial runs, to form a central composite design (CCD). For a two-factor experiment, these axial design points are

$$\begin{bmatrix} \alpha & 0 \\ -\alpha & 0 \\ 0 & \alpha \\ 0 & -\alpha \end{bmatrix}$$

with α set equal to $\sqrt{2}\,/\,2$; i.e., the axial points are drawn half the radial distance into the interior of the experimental subregion for bias protection. Observations on this experimental plan are presented in Table 2.12.

Table 2.12 Augmented Experimental Results for the Fourth Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1.000	-1.000	8700	4800	12246
-1.000	1.000	8700	5800	12440
1.000	-1.000	9700	4800	12055
1.000	1.000	9700	5800	12598
0.000	0.000	9200	5300	12099
0.000	0.000	9200	5300	12080
0.707	0.000	9554	5300	12180
-0.707	0.000	8846	5300	12082
0.000	0.707	9200	5654	12259
0.000	-0.707	9200	4946	11971

Table 2.13 provides a portion of the summary statistics for the fitted second-order model generated by the SAS procedure RSREG. The results show the fit to be excellent. The canonical analysis of this fitted model shows that a minimum response exists at the estimated stationary point

$$\underline{\xi} \cong \begin{bmatrix} 9324 \\ 4843 \end{bmatrix}$$

since the eigenvalues positive. It also shows that this is a unique point of minimum response as the eigenvalues are significantly greater than zero. Thus this analysis yields as an estimate of optimum operating conditions a reorder point of 9324 units and a reorder quantity of 4843 units. The estimated average monthly inventory costs corresponding to this solution is $11,985.

Table 2.13 Second-order Analysis Results for the Fourth Subregion

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	9200.000000	500.000000
X2	5300.000000	500.000000

Response Surface for Variable Y

Response Mean	12201
Root MSE	40.972827
R-Square	0.9799
Coef. of Variation	0.3358

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	177235	0.5307	52.787	0.0013
Quadratic	2	119581	0.3580	35.616	0.0028
Crossproduct	1	30450	0.0912	18.138	0.0131
Total Regress	5	327267	0.9799	38.989	0.0017

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	6715.090346	1678.772587

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	89464	12254	7.301	0.0019	12070
X1	1	-12.622633	3.246239	-3.888	0.0177	7.273123
X2	1	-7.695776	2.000998	-3.846	0.0184	188.085704
X1*X1	1	0.000586	0.000175	3.353	0.0285	146.569010
X2*X1	1	0.000349	0.000081946	4.259	0.0131	87.250000
X2*X2	1	0.000459	0.000175	2.623	0.0586	114.649702

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	49594	16531	9.847	0.0256
X2	3	218972	72991	43.479	0.0016

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Critical Value Uncoded
X1	0.247346	9323.672923
X2	-0.914379	4842.810349

Predicted value at stationary point 11985

Eigenvalues	Eigenvectors X1	Eigenvectors X2
177.062033	0.819624	0.572901
84.156679	-0.572901	0.819624

Stationary point is a minimum.

2.3.2 NONSTATIONARY MODEL APPROACH

In the approach described in Section 2.3.1, a substantial restriction was imposed on the stochastic components that drive the simulation model to stabilize the internal as well as the output processes; i.e., the means of the three input distributions were fixed at their nominal values in an effort to approximate stationarity of response. In this section, we allow these means to vary over time. Thus the simulation is conducted in a manner reflecting the true nonstationary nature of these processes; the ensuing analysis therefore reflects the inherent instability of the real world system under study.

The rationale behind the analysis presented in this section is based on the response statistic used in the modeling phase of RSM. Given the output stream

$$\{y_{ij}\,,\ j = 1, \ldots, q\},$$

where the i-subscript represents the *ith* design setting in the signal variables, the j-subscript indexes the observed monthly total costs, and q denotes the simulation run length (which is taken to be equal to 2000 in the inventory example). The response statistic of interest is

$$\bar{y}_i = \frac{\sum_{j=1}^{q} y_{ij}}{q}. \tag{2.13}$$

We exploit certain asymptotic properties of this statistic to justify the application of RSM to the nonstationary problem under consideration.

Consider the expected value of this statistic,

$$E(\bar{y}) = E\left(\frac{\sum_{j=1}^{q} y_j}{q}\right) = \frac{\sum_{j=1}^{q} E\left(y_j | \underline{\zeta}_{1j}, \underline{\zeta}_{2j}, \underline{\zeta}_{3j}\right)}{q}, \tag{2.14}$$

with the i-subscript dropped for convenience. By the Law of Large Numbers (see Hogg and Craig (1970)),

$$E(\bar{y}) = \lim_{q \to \infty} \frac{\sum_{j=1}^{q} E\left(y_j | \underline{\zeta}_{1j}, \underline{\zeta}_{2j}, \underline{\zeta}_{3j}\right)}{q} = E\left(E\left(y_j | \underline{\zeta}_{1j}, \underline{\zeta}_{2j}, \underline{\zeta}_{3j}\right)\right). \tag{2.15}$$

From Rohatgi (1976), we have

$$E(\bar{y}) = E(y_j), \tag{2.16}$$

which is an unconditional expectation. This is the expectation of the unconditional distribution of the y_j ; i.e., independent of the $\underline{\zeta}_k$.

This unconditional distribution is the infinite mixture of all the conditional distributions of the Y_j, with density

$$f(y) = \int_{\mathbf{Z}} f(y|\mathbf{Z}) \cdot g(\mathbf{Z}) \cdot d\mathbf{Z} \tag{2.17}$$

where f(y|**Z**) is the conditional density of y given **Z**, g(**Z**) is the joint distribution of the $\underline{\zeta}_l$, and **Z** is the vector of $\underline{\zeta}_l$. This mixture distribution can be construed as the combined or overall distribution of an infinite number of conditional distributions. Analogously, the mixture distribution is the overall population, while the conditional distributions are the subpopulations.

Considerable research has been done on mixture distributions (for example, see Everitt and Hand (1981), and McLachlan and Basford (1988)). In most of these studies, the focus was on the finite mixture case and primarily concerned with defining the subpopulations and categorizing the observations according to their sub-distributional origin. In contrast, the focus of this second RSM approach is on the mixture distribution and taking advantage of one of its underlying properties. With the y_{ij} all originating from this common or overall distribution, they are indeed identically distributed random variables. The direct application of RSM in the typical manner is thus enabled.

The optimization problem is now one if finding the $\underline{x}$ that will minimize the unconditional expectation of y. Contrasting this problem with the problem considered in the previous section, the difference is in the change of perspective with regard to the output process. The first approach considered the conditional distributions on the response of interest; in particular, the conditional distribution defined by the most likely operating conditions or mean rates. The optimization problem in that scenario is

$$\min_{\underline{x}} E\left(y | \underline{\zeta} = \underline{\mu}_{\zeta} \right).$$

The second approach considers the optimization problem from the perspective of the unconditional or overall distribution of the response,

$$\min_{\underline{x}} E(y).$$

Furthermore, these two optimization problems are not equivalent since

$$E\left(y|\underline{\zeta}=\underline{\mu}_{\zeta}\right) \neq E(y). \tag{2.18}$$

The equality holds only for special cases wherein the conditional expectation of y given $\underline{\zeta}$ can be expressed a linear function of $\underline{\zeta}$. To illustrate that point, consider the following simple contradiction with $\underline{\zeta}'=[\zeta_1, \zeta_2]$ and

$$E\left(y|\underline{\zeta}\right) = \zeta_1^2 + \zeta_2^2.$$

With the above situation we have the left-hand side of (2.18) as

$$E\left(y|\underline{\zeta}=\underline{\mu}_{\zeta}\right) = \mu_{\zeta_1}^2 + \mu_{\zeta_2}^2.$$

The right-hand side of (2.18) for this case is

$$E(y) = E\left[E\left(y|\underline{\zeta}\right)\right] = E(\zeta_1^2) + E(\zeta_2^2)$$

which is obviously not identical to the earlier result.

Another important consideration concerns the variance of y for the cases of conditional and unconditional expectations. This will be discussed in greater detail in Section (2.4).

The following subsections present the results of an RSM analysis using the second approach. These results show that the difference cited above between the two approaches translates to pronounced differences in the derived solutions.

2.3.2.1 NUMERICAL RESULTS

The RSM analysis was conducted in very much the same fashion as in the previous approach, the only difference being in the manner that the data were generated. The same first and second order experimental plans presented previously were also used in this approach. Additionally, the first phase was initiated at the same design location/subregion, namely

$$\{(\xi_1, \xi_2) \ni \xi_1 \in [500,1500],\ \xi_2 \in [500,1500]\},$$

The experimental plans, model analysis, and steepest-descent exploratory results for the first-order phase are summarized in the following tables.

Table 2.14 Experimental Results for the Initial Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1	-1	500	500	90890
-1	1	500	1500	84837
1	-1	1500	500	81700
1	1	1500	1500	71165
0	0	1000	1000	80199
0	0	1000	1000	80372

Table 2.15 First-order Model Results for the Initial Subregion

```
The regression equation is
y = 81527 - 5715 x1 - 4147 x2

Predictor        Coef        Stdev    t-ratio        p
Constant      81527.1        732.7     111.28    0.000
x1            -5715.5        897.3      -6.37    0.008
x2            -4147.0        897.3      -4.62    0.019

s = 1795         R-sq = 95.4%

Analysis of Variance

SOURCE        DF         SS          MS          F        p
Regression     2  199458192    99729088      30.96    0.010
Error          3    9662253     3220751
Total          5  209120432
```

The fitted gradient or vector of estimated first-order coefficients is r'=[-5715.5,-4147] and provides the steepest-ascent direction. We therefore work with the additive inverse of the standardized version of this vector,

$$-\underline{\hat{\beta}}_s = -\underline{b}_s = \begin{bmatrix} 0.8094 \\ 0.5873 \end{bmatrix}.$$

Using this vector as our search direction, additional experiments were conducted at points along this path as defined by (2.12). The results for these exploratory simulation runs along this gradient direction are provided in the next table.

Table 2.16 Experimental Results Along the First Fitted Gradient

r	x_1	x_2	ξ_1	ξ_2	y_i
1	0.8094	0.5873	1404.7	1293.6	74774
2	1.6188	1.1745	1809.4	1587.3	69220
4	3.2376	2.3491	2618.8	2174.5	58722
8	6.4751	4.6982	4237.6	3349.1	40799
16	12.9503	9.3963	7475.1	5698.2	17435
32	25.9005	18.7926	13950.2	10396.3	24148

The steepest-descent results indicate the most promising location to be approximately

$$\underline{\xi} \cong \begin{bmatrix} 9000 \\ 7000 \end{bmatrix},$$

or at r≈20. A new subregion was therefore centered about this point and defined to be of the same size as the previous one. Results of the simulation trials and the ensuing first-order model-fitting procedure are given in the next two tables.

Table 2.17 Experimental Results for the Second Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1	-1	8500	6500	14714
-1	1	8500	7500	15025
1	-1	9500	6500	13956
1	1	9500	7500	14754
0	0	9000	7000	14245
0	0	9000	7000	14330

Table 2.18 Model-fitting Results for the Second Subregion

```
The regression equation is
y = 14504 - 257 x1 + 277 x2

Predictor        Coef        Stdev    t-ratio        p
Constant      14504.0        106.3     136.40    0.000
x1             -257.2        130.2      -1.98    0.143
x2              277.2        130.2       2.13    0.123

s = 260.5        R-sq = 73.8%

Analysis of Variance

SOURCE        DF         SS          MS          F       p
Regression     2     572180      286090       4.22   0.134
Error          3     203521       67840
Total          5     775702
```

Table 2.18 shows that the fitted first-order model provides a poor representation of the response over the present subregion region. Since this result could most likely be attributed to model misspecification, a second-order model was postulated and additional experiments conducted to allow for the estimation of quadratic terms. Specifically, the previous first-order design was augmented with axial design points in the same manner described in Section 2.3.1.2 to form a central composite design. The resulting experimental data and regression analysis are provided in the following tables.

Table 2.19 Second-order Design and Experimental Results

x_1	x_2	ξ_1	ξ_2	y_i
-1.000	-1.000	8500	6500	14714
-1.000	1.000	8500	7500	15025
1.000	-1.000	9500	6500	13956
1.000	1.000	9500	7500	14754
0.000	0.000	9000	7000	14245
0.000	0.000	9000	7000	14330
0.707	0.000	9354	7000	14132
-0.707	0.000	8646	7000	14615
0.000	0.707	9000	7354	14614
0.000	-0.707	9000	6646	14328

Table 2.20 Second-order Analysis Results

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	9000.000000	500.000000
X2	7000.000000	500.000000

Response Surface for Variable Y

Response Mean	14471
Root MSE	68.920945
R-Square	0.9802
Coef. of Variation	0.4763

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	719545	0.7505	75.740	0.0007
Quadratic	2	160953	0.1679	16.942	0.0111
Crossproduct	1	59292	0.0618	12.482	0.0242
Total Regress	5	939790	0.9802	39.569	0.0017

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	19000	4750.096636

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	110333	18337	6.017	0.0038	14324
X1	1	-7.549049	5.381042	-1.403	0.2333	-274.054238
X2	1	-17.544865	4.300428	-4.080	0.0151	262.165049
X1*X1	1	0.000200	0.000294	0.679	0.5346	49.888057
X2*X1	1	0.000487	0.000138	3.533	0.0242	121.750000
X2*X2	1	0.000978	0.000294	3.324	0.0293	244.396340

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	437198	145733	30.680	0.0032
X2	3	455609	151870	31.972	0.0030

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Uncoded
X1	4.886306	11443
X2	-1.753448	6123.275788

Predicted value at stationary point 13425

Eigenvalues	Eigenvectors X1	X2
261.877257	0.276006	0.961156
32.407140	0.961156	-0.276006

Stationary point is a minimum.

The model-fitting results of Table 2.20 indicate that the estimated stationary point is outside the coverage of the current subregion. It also shows that this point is projected to be an estimate of the point of minimum response. Since results outside the current region are inadmissible, another experimental subregion is defined that approximates the location of the projected stationary point. The new design region is defined as

$$\{(\xi_1, \xi_2) \ni \xi_1 \in [9000, 11000], \ \xi_2 \in [5000, 7000]\},$$

with design widths twice their original size to provide greater coverage of the operability space since the projected stationary point was distant from the initial subregion in the second-order analysis phase. The same experimental plan is utilized, with the simulation runs yielding the following results.

Table 2.21 Experimental Results for the Second Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1.000	-1.000	9000	5000	13393
-1.000	1.000	9000	7000	14544
1.000	-1.000	11000	5000	14313
1.000	1.000	11000	7000	15689
0.000	0.000	10000	6000	13702
0.000	0.000	10000	6000	13644
0.707	0.000	10707	6000	14394
-0.707	0.000	9293	6000	13673
0.000	0.707	10000	6707	14353
0.000	-0.707	10000	5293	13544

Table 2.22 Second-order Analysis Results for the Second Subregion

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	10000	1000.000000
X2	6000.000000	1000.000000

Response Surface for Variable Y

Response Mean	14125
Root MSE	84.837333
R-Square	0.9933
Coef. of Variation	0.6006

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	3246775	0.7605	225.6	0.0001
Quadratic	2	981104	0.2298	68.157	0.0008
Crossproduct	1	12656	0.0030	1.758	0.2555
Total Regress	5	4240535	0.9933	117.8	0.0002

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	28789	7197.373069

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	64425	7410.105597	8.694	0.0010	13752
X1	1	-8.982620	1.832525	-4.902	0.0080	514.980505
X2	1	-3.398114	1.168961	-2.907	0.0438	619.830038
X1*X1	1	0.000458	0.000090718	5.049	0.0072	458.005046
X2*X1	1	0.000056250	0.000042419	1.326	0.2555	56.250000
X2*X2	1	0.000288	0.000090718	3.174	0.0337	287.953690

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	1522053	507351	70.491	0.0006
X2	3	2006002	668667	92.904	0.0004

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Uncoded
X1	-0.499102	9500.897874
X2	-1.027519	4972.481404

Predicted value at stationary point 13305

Eigenvalues	Eigenvectors X1	X2
462.535951	0.987271	0.159048
283.422785	-0.159048	0.987271

Stationary point is a minimum.

The second iteration of the second-order analysis phase yielded results similar to those of the first iteration; an estimated stationary point exterior to the current subregion which is also projected to be an estimate of the minimum response point. Some improvement is observed, however, since the location of the estimated stationary point is not as remote relative to the current subregion under investigation. Also that point is exterior only with respect to the second signal variable, x_2. A new subregion is therefore defined moving the design location in the direction of the estimated stationary point, with the design width along the axis of x_2 increased by fifty percent to provide greater coverage in an effort to contain the true stationary point. The new design region is bounded by

$$\{(\xi_1, \xi_2) \ni \xi_1 \in [8500, 10500],\ \xi_2 \in [3500, 6500]\}.$$

The ensuing experimental and second-order analysis results are presented in the next two tables.

Table 2.23 Experimental Results for the Third Subregion

x_1	x_2	ξ_1	ξ_2	y_i
-1.000	-1.000	8500	3500	14481
-1.000	1.000	8500	6500	14778
1.000	-1.000	10500	3500	13185
1.000	1.000	10500	6500	14706
0.000	0.000	9500	5000	13121
0.000	0.000	9500	5000	13276
0.707	0.000	10207	5000	13418
-0.707	0.000	8793	5000	13739
0.000	0.707	9500	6061	13789
0.000	-0.707	9500	3939	12956

Table 2.24 Second-order Analysis Results for the Third Subregion

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	9500.000000	1000.000000
X2	5000.000000	1500.000000

Response Surface for Variable Y

Response Mean	13745
Root MSE	96.270064
R-Square	0.9911
Coef. of Variation	0.7004

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	1667584	0.3983	89.965	0.0005
Quadratic	2	2107465	0.5034	113.7	0.0003
Crossproduct	1	374544	0.0895	40.413	0.0031
Total Regress	5	4149593	0.9911	89.547	0.0003

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	37072	9267.925294

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	95691	8594.981337	11.133	0.0004	13202
X1	1	-15.576873	1.962185	-7.939	0.0014	-319.008668
X2	1	-3.114516	0.550384	-5.659	0.0048	481.380031
X1*X1	1	0.000749	0.000103	7.282	0.0019	749.361291
X2*X1	1	0.000204	0.000032090	6.357	0.0031	306.000000
X2*X2	1	0.000150	0.000045734	3.274	0.0307	336.923151

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	1374838	458279	49.448	0.0013
X2	3	1632683	544228	58.722	0.0009

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Uncoded
X1	0.395368	9895.368227
X2	-0.893917	3659.124401

Predicted value at stationary point 12924

Eigenvalues	Eigenvectors X1	X2
799.920929	0.949500	0.313767
286.363513	-0.313767	0.949500

Stationary point is a minimum.

The analysis presented in Table 2.24 suggests that the second-order metamodel provides an excellent representation of average monthly inventory costs over the current experimental subregion. It is also seen that the canonical analysis of the fitted model yields a stationary point that lies within the region. The estimated minimum response of \$12,924 corresponds to a reorder point of ξ_1 = 9895 units and a reorder quantity of ξ_2 = 3659 units. Examination of the eigenvalues corresponding to this solution shows that the response is approximately two-and-a-half times as sensitive to movement away from the stationary point in the direction of the axis of the first canonical variable as in the second canonical variable axial direction.

2.4 CONCLUDING REMARKS

This Chapter has presented two approaches to the use of RSM for system optimization in the presence of model validation problems caused by transitory system behavior. Both approaches produce estimates of the optimum system operating conditions. Each has its merits as wells as its drawbacks. The first approach is a straightforward implementation of RSM under the assumption that the mean parameters of the distributions describing the process generators for the simulation model are set equal to their expected values. This restriction is not unreasonable nor uncommon as it yields an optimal solution under average conditions. That is, the choice for the constant set of environmental conditions corresponds to be the most likely set. However, the end result is that the derived solution does not address the original problem. It becomes more akin to a simulation optimization situation where the analyst decides to ignore the nonstationary nature of the system's stochastic processes at the outset of the study, and proceeds to construct the simulation model and analyze the results in a manner consistent with his initial assumptions. In other words, it does not address the original model validation problem.

In contrast, the second approach does take into account the transitory nature of the system. It also produces an estimate of the true optimal solution with respect to the mean response that maintains its optimality across time as shown in Section 2.3.2. Here the application of RSM is also straightforward. However, since the model needs to reflect the nonstationarity of the system's stochastic process, the construction of the simulation model entails more programming effort.

A more serious concern stems from conducting the analysis under increased system instability. For a fixed ζ, as in the first approach,

$$Y = E(y|\zeta) + \varepsilon \tag{2.19}$$

where ε is a random error term. The variance of y in this instance depends on just the variance of the error term,

$$Var(y|\zeta) = Var(\varepsilon),$$

as the conditional mean of y is a constant with a fixed ζ. Now if the ζ were random also as is the case with the unconditional distribution of y, the variance of y would be inflated as both components of the sum in (2.19) are random terms; i.e.

$$Var(y) = Var(E(y|\zeta)) + Var(\varepsilon)$$

assuming that the conditional mean and the error term are independent.

This variance inflation result is in agreement with the following formulation from Rao (1973) relating the conditional variance of y given a fixed ζ, $Var(y|\zeta)$, to the variance of the unconditional distribution of y:

$$Var(y) = E(Var(y|\zeta)) + Var(E(y|\zeta)).$$

This implies that

$$Var(y) \geq E(Var(y|\zeta)).$$

Since the conditional variance of y, $Var(y|\zeta)$, is equal to a constant ($Var(\varepsilon)$) once ζ is fixed, the expectation on the right hand side of the equation reduces to the conditional variance of y, yielding the desired result which is

$$Var(y) \geq Var(y|\zeta).$$

Likewise, the response(s) of interest in the second approach will exhibit increased variability. This, in turn, will adversely affect the quality of the optimization results. It is well known in regression analysis that the variance of the response impacts both the variance of the estimated model coefficients and the prediction variance . Thus, in the second approach, as the problem of optimality with respect to the mean response is resolved, a new problem arises in the form of increased response variance. This result will prove to be of particular importance in later developments.

Chapter 3

ROBUST PARAMETER DESIGN

3.1 INTRODUCTION

Robust Parameter Design (RPD), also known as the Taguchi Method, is an empirical decision making approach developed by Genichi Taguchi for the enhancement of product or process quality. The focus of the technique is the reduction in product or process performance variation. Such variation is viewed as the primary cause of quality loss. Articles by Taguchi (1978), Taguchi and Wu (1980), Box (1988), Easterling (1985), and Pignatiello (1988), and Pignatiello and Ramberg (1985) describe and highlight Taguchi's work.

Performance variation is defined as the amount by which a product's performance level deviates from its ideal or target value during its lifespan under different operating conditions and different reproductions. (The terms product and process are used interchangeably whenever reference is made to performance level or variation.) The primary sources of product performance variation are noise factors, which are often environmental variables that define the operating conditions, product utilization or handling, deterioration, and aging. These factors, while directly or indirectly affecting product performance, are deemed to be uncontrollable or quite difficult and expensive to control. The Taguchi Method has as its focus the reduction of this variation by designing the product in such a way as to reduce its performance sensitivity to the noise factors, rather than by controlling or eliminating these sources of performance variation.

An example cited by Kackar (1985) to illustrate this philosophy is the case of the Ina Tile Company, a Japanese ceramic tile manufacturer. In 1953, this company concluded that uneven temperature distribution within the kiln resulted in variation in tile size. Rather than attempting to control this cause of tile size variation at a considerable cost to the firm, they decided instead to find a tile design or formulation whose end product's size uniformity was insensitive to the effect of uneven temperature. Through the application of the Taguchi Method, they found that increasing the level of lime content from 1% to 5% in the tile formulation resulted in a reduction in tile size variation by a factor of 10. Thus, by finding a tile design more resistant to the assignable cause of variation, which in this case is uneven kiln temperature, the company was able to improve product quality without resorting to the more expensive solution of modifying their kilns to control the source of variation.

Taguchi (1978) outlines three stages in the product design development cycle: system design, parameter design, and tolerance design. In the system design phase, scientific and engineering knowledge is utilized to produce a basic prototype of the product, including initial functional design attributes or characteristics. The second stage, parameter design, is intended to determine the optimal parameters of the design; i.e., finding the best values for the controllable inputs or design variables of the product. The third stage is the determination of economic tolerance limits for the design variables. The focus of the Taguchi Method is on the second stage, parameter design, where finding the optimal design parameters is then redefined as identifying settings for the design variables that result in product performance insensitivity to the noise factors. Since the method is utilized during the product design development cycle, robust parameter design is an "off-line" quality control method whereby quality is built into the product by design, not by inspection.

The first step in parameter design is the identification of the system variables. The controllable inputs $\underline{\xi}$ (or in their coded form, $\underline{x}$) are termed the signal or design variables, and the identified uncontrollable factors causing performance variation are classified as noise variables $\underline{\zeta}$. An empirical investigation or parameter design experiment is conducted to determine values for the signal variables that minimize performance variation and still approximate, if not attain, the performance target value. To fully appreciate this approach, we need to separate Taguchi's strategy from his tactics. The next two sections address these two topics and the steps of the method are motivated, developed, and presented as they are currently understood today. The strategy section discusses the motivating philosophy of Taguchi which is based on need for the simultaneous consideration of process mean and variance. The tactics section presents the details of the implementation of the strategy. This section also contains a subsection citing the main criticisms to Taguchi's approach.

Following the discussion of the concepts and techniques of RPD, we again focus on the optimization problem introduced in Chapter 1, that of optimization in the presence of nonstationarity. The increase in system performance variation due to nonstationarity as shown in Chapter 2, and the need to address this additional performance criterion in the optimization process, provide the motivation for the use of the Taguchi approach. The adaptation of RPD to this problem is developed next. Modifications to the tactics of Taguchi are offered to address the major criticisms presented in an earlier section. The revised RPD approach is subsequently applied to the nonstationary inventory optimization problem, and numerical results using this analysis are presented.

We end the chapter with some concluding comments highlighting some of the merits and shortcomings of the original and revised Taguchi Method, especially as they pertain to the optimization problem under study.

3.2 TAGUCHI STRATEGY

The underlying principle of the Taguchi strategy is Taguchi's quality philosophy. Fundamental to this philosophy is the concept of an ideal or target value for the performance characteristic of interest. With this target value, Taguchi defines quality, or the lack of it, as the loss to society incurred when the product or process performance level deviates from a specified target value. This abstract view of the quality improvement problem is quite different from the traditional perspective that hinges on the concept of tolerances on the performance level. With this motivating definition of quality, Taguchi's goal is to minimize societal loss.

The embodiment of Taguchi's departure from the traditional quality philosophy is his use and choice of a *loss function.* This loss function represents a quantification of societal loss resulting from product/process performance deviation from the target value. Taguchi's loss function is the quadratic error loss function

$$l(\underline{x}) = k\ \{y(\underline{x}) - \tau\}^2 \tag{3.1}$$

where τ is the intended or target value for the performance level $y(\underline{x})$ and k is some economic scalar. The use of a loss function transforms the previous abstract problem of minimizing societal loss to a more concrete problem of minimizing expected loss over the space of the noise variables.

The expectation of the loss function, also known as the risk function $R(\underline{x})$, is taken over the noise space to incorporate the random effects of the noise variables. These noise variables are construed as the primary sources of performance deviation from the target value. Pignatiello (1988) characterizes the minimization of the expectation of the loss function as a case of decision making under uncertainty with future realizations of the noise variables defining the unknown future states of nature. In the case of Taguchi's preferred loss function, which is the quadratic loss function, the problem becomes equivalent to minimizing mean squared error. Partitioning mean squared error, the problem translates into the minimization of the sum of variance and squared bias, i.e.,

$$\min R(\underline{x}) = \min E_Z\{l(\underline{x})\} = \min\left\langle\ k\{\mathrm{Var}_Z[y(\underline{x})] + [\tau - E_Z(y(\underline{x}))]^2\}\ \right\rangle. \tag{3.2}$$

The preceding result reflects the soundness of the strategy. The objective of minimization of the expected loss function results in the simultaneous consideration of the variance and the mean, wherein the variance in $y(\underline{x})$ as well as the difference between the mean and the target value τ is minimized.

To accomplish this objective Taguchi proposes the use of empirical optimization techniques. This involves the identification of a performance measure that addresses the expected loss function through the simultaneous consideration of the process mean and variance while reflecting the effects of the noise variables. Planned experiments are conducted and estimates of the performance measure, termed performance statistics, are constructed from the collected data.

These planned experiments entail a systematic exploration of the noise space for each given setting of the input signal factors. They amount to a selective sampling within the noise space, conditioned on a given combination of the controllable input variables, to gauge the reaction or sensitivity of the given input setting's performance when subjected to an array of settings for the noise variables that provide a sufficient spectrum or coverage of the noise variable space. This provides a venue for the later analysis of the input setting's sensitivity to different operating conditions. This systematic exploration of the noise space, which will be referred to as sensitivity sampling, is accomplished by fixing the levels of the noise variables for each experimental run. This is made possible by the experimenter's ability to control or specify the levels of these noise factors in the experimental or prototype stage of the product/process development cycle. In the actual process in the field, these noise factors are, of course, uncontrollable or difficult to control. We note that this type of control over the noise variables is easily accommodated in the simulation context.

Estimation of the levels of the signal variables that optimize system performance is based on estimates of the aforementioned metric, the performance statistic. Thus, the optimality criterion used in specifying the signal factor combinations should be equivalent to that of minimizing the expected loss or risk which is the original objective of the analysis.

3.3 TAGUCHI TACTICS

The Taguchi tactics deal with the implementation of the strategy discussed in the previous section. Whereas the Taguchi strategy provides the motivation, these tactics are concerned with the practical issues regarding the implementation of the strategy. Consideration is given to the methodological details regarding the collection of data, called the Taguchi designs, and the ensuing Taguchi analysis of the experimental results.

3.3.1 TAGUCHI DESIGNS

The data collection procedure or experimental designs advocated by Taguchi illustrate the concept of exploring the noise space $Z = [Z_1 \times Z_2 \times \ldots \times Z_k]$ conditioned on a given combination of the input variables $\underline{x}_i' = [x_{i1}, x_{i2,}, \ldots, x_{ip}]'$. As in the case of the signal variables $\underline{x}$, the study of the noise variables is conducted in terms of their coded or transformed versions $\underline{z}$. The distinguishing feature of a Taguchi design is that it consists of experimental design arrays **D** and **Z** being cross-classified forming a Kronecker product $\mathbf{D}\otimes\mathbf{Z}$. The **D**-array is an orthogonal experimental design in the input factors:

$$\begin{bmatrix} x_{11} & \cdots & x_{1s} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & \cdots & x_{is} & \cdots & x_{ip} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{ns} & \cdots & x_{np} \end{bmatrix},$$

and the **Z**-array is an orthogonal design in the noise variable space:

$$\begin{bmatrix} z_{11} & \cdots & z_{1t} & \cdots & z_{1k} \\ \vdots & \ddots & \vdots & & \vdots \\ z_{j1} & \cdots & z_{jt} & \cdots & z_{jk} \\ \vdots & & \vdots & \ddots & \vdots \\ z_{m1} & \cdots & z_{mt} & \cdots & z_{mk} \end{bmatrix}.$$

An orthogonal design is a design whose columns form multi-dimensional vectors which are at right angles with each other. For later reference, it is convenient to display the crossed-array design and observations in the following form.

					z_{11}	...	z_{j1}	...	z_{m1}	
					:		:		:	
					z_{1t}	...	z_{jt}	...	z_{mt}	
					:		:		:	
					z_{1k}	...	z_{jk}	...	z_{mk}	
x_{11}	...	x_{1s}	...	x_{1p}	y_{11}	...	y_{1j}	...	y_{1m}	$p(\underline{x}_1)$
:		:		:	:		:		:	:
x_{i1}	...	x_{is}	...	x_{ip}	y_{i1}	...	y_{ij}	...	y_{im}	$p(\underline{x}_i)$
:		:		:	:		:		:	:
x_{n1}	...	x_{ns}	...	x_{np}	y_{n1}	...	y_{nj}	...	y_{nm}	$p(\underline{x}_n)$

Figure 3.1 Crossed-array Diagram

Note that the crossed-array structure results in each input factor setting in **D**,

$$\underline{x}_i' = [x_{i1}, x_{i2}, \ldots, x_{ip}]', \quad i = 1, \ldots, n,$$

being combined with each noise factor setting in **Z**,

$$\underline{z}_j' = [z_{j1}, z_{j2}, \ldots, z_{jk}]', \quad j = 1, \ldots, m,$$

producing m replications for each input factor setting. Thus each combination of the p input factors is subjected to a preselected set of m operating conditions defined by the k noise variables. The m observations for a given input factor combination reflect the sensitivity of that input factor combination to variations in the noise variables. Within the input variable space, the m observations generated for each input factor setting are replication data, but these replications are actually systematically controlled replicates from the perspective of the noise variables. In traditional experimental design, these noise variables are neither controlled nor observed, but instead are accommodated through randomization techniques in experimental design. In a setting such as computer simulation, the experimenter has the ability to completely control the signal variables as well as the noise variables. Designs involving crossed inner and outer arrays for sensitivity analysis have a considerable history. Box (in Nair (1992)) cites early examples in agriculture and in the food industry.

As for the structure of the **D** and **Z** matrices, these so-called Taguchi designs are standard orthogonal array experimental designs; i.e., they are of the generalized Graeco-Latin square variety. Some of these designs are due to Plackett and Burman (1946). One notable characteristic of the **D** matrix is that it is a saturated or near saturated design; i.e., the number of runs is sufficient for the estimation of just the first order regression coefficients or the main effects. As with the **D** matrix, the **Z** matrix could be characterized as a near, if not already, saturated first order design. The recommended use of these Taguchi designs in the crossed-array scenario is a major source of criticisms of the Taguchi method and will be discussed in more detail in a later section.

3.3.2 TAGUCHI ANALYSIS

The data analysis phase can be broken down to two sequential stages: the performance statistic construction stage and the marginal means analysis stage. It is in the first phase that the performance statistic P($\underline{x}$) is constructed. Taguchi advocates the blanket use of what he terms signal-to-noise ratios (S/N) for this purpose. The choice of S/N ratio is dependent on the type of problem or goal of the experimenter. The three most discussed goals are :

- *Smaller the better* - The goal is to minimize the performance response,
- *Larger the better* - The goal is to maximize the performance response,
- *Target is best* - The goal is to attain a specific value for the performance response.

A discussion of the corresponding S/N ratios is presented in the next section.

3.3.2.1 SIGNAL-TO-NOISE RATIOS

Taguchi's S/N ratios are the recommended performance statistics for the analysis. They are estimators or functions of estimators of some performance measure that supposedly addresses expected loss. In other words, Taguchi' contends that the optimization of the appropriate S/N ratio equivalently leads to the minimization of expected loss. (This contention will be shown to be inaccurate for certain S/N ratios in a later section.) These ratios are formulated in a manner that requires their maximization to achieve optimality.

For the "smaller-the-better" case, Taguchi considers the fixed target value for the performance response to be zero. The expected loss function then simplifies to

$$R(\underline{x}) = k\, E_Z[y(\underline{x}) - 0]^2 = k\, E_Z[y(\underline{x})]^2 \cdot \tag{3.3}$$

Viewing this as the appropriate performance measure, the naive estimator at the *ith* setting for the input variables becomes

$$k\frac{\sum_{j=1}^{m} y_{ij}^2}{m}$$

where the arithmetic average of the squared terms is computed from the replications described earlier in Figure 3.1: i.e.,

$$y_{ij}, \quad j = 1,\ldots,m.$$

Hence, for each setting of the input signal variables $\underline{x}_i'$, the observations or replications at that point are summarized into one statistic. The final form of the performance statistic, which is a function of the above estimator for the expected loss function, is

$$S/N = -10\log \frac{\sum_{j=1}^{m} y_{ij}^2}{m}. \tag{3.4}$$

The log transformation is applied presumably as a variance stabilizing transformation (see Bartlett and Kendall (1946)). Thus, to minimize $R(\underline{x})$, this performance statistic needs to maximized.

The "larger-the-better' situation is just the opposite of the preceding scenario. Taguchi's fixed target value is some unspecified large number. To cope with such a vague definition of the problem, Taguchi prefers to view this problem in the context of the previous problem. This is accomplished by considering the reciprocal of this performance response, 1/y, whose optimal value in this case would be an estimator of zero. The expected loss function then becomes

$$R(\underline{x}) = k\, E_Z[\{1/y(\underline{x})\} - 0]^2 = k\, E_Z[1/y(\underline{x})]^2 , \tag{3.5}$$

and the naive estimator for this expected loss is

$$k \frac{\sum_{j=1}^{m} \left[\frac{1}{y_{ij}}\right]^2}{m}$$

which leads to

$$S/N = -10 \log \frac{\sum_{j=1}^{m} \left[\frac{1}{y_{ij}} \right]^2}{m}. \tag{3.6}$$

In the case of "target-is-best", the assumption is that there exists some known target value for the response, say τ, whose attainment is desired. The Taguchi risk function is then the basic risk function mentioned earlier in the discussion of the Taguchi strategy, namely

$$R(\underline{x}) = kE_Z[y(\underline{x}) - \tau]^2 . \tag{3.7}$$

This function is equal to mean squared error multiplied by k, some economic factor or constant as stated earlier. Taguchi's approach aims to minimize $R(\underline{x})$. To accomplish this, Taguchi seems to have chosen the coefficient of variation, σ/μ, as the appropriate performance measure. This is, presumably, to account for any linear dependence of the response standard deviation to the mean. The estimator for the above performance measure based on the replications in the input variable space is defined as

$$\frac{s_i^2}{\overline{y}_i^2},$$

where the sample mean and variance for the *ith* input setting are computed as

$$\overline{y}_i = \frac{\sum_{j=1}^{m} y_{ij}}{m}, \text{ and } s_i^2 = \frac{\sum_{j=1}^{m} \left(y_{ij} - \overline{y}_i\right)^2}{m-1}, \tag{3.8}$$

respectively. Taguchi recommends the performance statistic

$$S/N = 10\log\frac{\overline{y}_i^2}{s_i^2} \tag{3.9}$$

which is again to be maximized. Myers, Khuri, and Vining (1992) note that this performance statistic is the only S/N expression involving some semblance of a signal to noise ratio. It is perhaps this ratio that provides an explanation to Taguchi's universal recommendation of these S/N ratios as performance statistics, or perhaps even his usage of the term signal-to-noise. It would seem that his interpretation of the robust parameter design problem is mainly that of an engineer's, such that finding the best design is essentially similar to the problem in communication theory of finding the design that maximizes the signal strength (which he relates to optimizing the mean or reducing bias) while suppressing the obscuring effects of noise (which he relates to minimizing variance).

2.3.2.2 OPTIMIZATION PROCEDURE

The second phase of the analysis involves the selection of the best input factor or signal variable combination. As noted earlier, the optimization criterion is the maximization of S/N. Focus now shifts from the outer array, which is used to generate the replication data needed to estimate S/N statistic values, to the inner array which is used to identify the settings of the signal variables that will yield the most beneficial and robust performance. Traditional analysis of variance (ANOVA) or a marginal means analysis is employed wherein the marginal means of the responses for each level of each of the signal variables is computed. Given an input factor or signal variable, the difference between the marginal means corresponding to its high and low settings is that factor's marginal effect on S/N. Each factor is set to the level that results in the maximum marginal mean value.

With the estimated marginal effects, the signal or design variable set is partitioned

$$\underline{x} = [\,\underline{x}_d \,, \underline{x}_a\,] \,. \tag{3.10}$$

The subset $\underline{x}_d$ of the design variables is the set of significant factors with respect to S/N such that

$$S/N(\underline{x}) \approx S/N(\underline{x}_d). \tag{3.11}$$

The subset $\underline{x}_a$ is identified as the set of adjustment variables; i.e., variables that mainly affect just the mean. The simple "pick-the-winner" exercise, by which levels for the factors producing the maximum value for S/N are chosen and held fixed, involve just the factors in $\underline{x}_d$. Confirmatory experimental runs are undertaken at the selected levels for these inputs, with the adjustment variables initially fixed at arbitrary values. The actual performance response is observed and the adjustment variables are varied sequentially to fine tune the response to the target value.

3.4 CRITICISMS OF ROBUST PARAMETER DESIGN

The Taguchi Method has received considerable attention during the decade of the 1980*s* primarily because of its successful implementations in industry. These successes can be attributed to the soundness of Taguchi's strategy: the simultaneous consideration of the performance mean and variance in the enhancement of quality and the use of the expected quadratic loss function to translate the quality problem to an empirical decision making situation under uncertainty.

Notwithstanding the apparent successes of RPD in industry, statisticians in this country have uncovered flaws in Taguchi's techniques. The controversies surrounding these procedures center mainly on the tactics recommended by Taguchi for the minimization of the expected loss or risk. Criticisms to his tactics are discussed in detail in articles by Leon, Shoemaker, and Kackar (1987), Box (1985, 1988), Welch, Yu, Kang, and Sachs (1990), Easterling (1985), Pignatiello and Ramberg (1985), Hunter (1985), Lucas (1985), Pignatiello (1988), and Myers, Khuri, and Vining (1992). The major criticisms include : (1) inadequacy of the optimization search process, (2) inefficiency of the experimental design plans, and (3) impropriety of S/N ratios as performance statistics. These criticisms are examined in the following subsections.

3.4.1 INADEQUACY OF THE OPTIMIZATION PROCEDURE

We focus first on the criticism regarding the characterization that the Taguchi Method is a single-cycle empirical decision making procedure. Box (in Nair (1992)) laments the inherent lack of understanding of the system due to Taguchi's apparent disregard of the sequential nature of empirical investigations. With RPD, data are collected from one comprehensive set of experimental runs and used in a "pick-the-winner" analysis for the purpose of minimizing expected loss. Aside from the fine tuning exercise that follows the selection of the settings in $\underline{x}_d$, no sequential analysis or investigation is undertaken. This would be appropriate for the case of constrained optimization wherein the experimenter is restricted to a subregion of the operability space corresponding to solutions deemed feasible by the decision-maker. In a case of unconstrained optimization, where prior restrictions or specifications on the solution space have not been determined, this disregard of the sequential nature of experimental investigations could lead to suboptimal decisions, especially when the operability region has not been adequately explored, or if the current experimental subregion does not contain the true optimal solution.

Another problem source is the "pick-the-winner" exercise or marginal means analysis recommended by Taguchi to select the optimal levels for the input signal variables. With this procedure, the only possible choices for the levels of the signal variables would be those specified in the inner array design corresponding to design locations. Constructing a solution which is a combination limited to just these levels could result in a suboptimal policy, especially in the situation where the signal variables are continuous or real-valued variables.

3.4.2 INEFFICIENCY OF EXPERIMENTAL DESIGN PLANS

The use of cross-classified inner and outer design arrays can lead to an inordinate number of experimental runs. Each design setting in the inner array is combined with each of the noise variable combinations specified by the outer array. With n design points for the inner array, and m combinations for the outer array, the total number of experimental runs, mxn, can be excessively large.

To reduce the number of runs required, Taguchi recommends the use of highly saturated design plans for the inner array, plans which are used primarily for assessing just the main effects of the signal factors. The motivation for this is that Taguchi deems interactions between signal variables to be insignificant. He therefore recommends that these effects be ignored in the estimation procedure, and that saturated or near saturated designs be used for the inner array to minimize the number of signal factor combinations (n), and thereby reduce the overall number of experimental runs.

In this attempt to reduce experimental effort, signal factor interactions are sacrificed. This could lead to a potentially critical problem when interactions between signal variables do play a major role in the relation between the input factors and the output response. Furthermore, Sachs and Welch (in Nair (1992)) show that disregard of signal factor interactions could be in serious conflict with the use of the S/N performance statistics. They illustrate this by considering a simple case involving two signal variables, x_1 and x_2, and one noise variable z related to the output variable y through the equation

$$y = x_1 + x_2 + z.$$

Given a "smaller-the-better" optimization scenario, the Taguchi statistic would be

$$S/N = -10\log\frac{\sum_{j=1}^{m} y_{ij}^{2}}{m}.$$

It is obvious from the above expression that S/N is not just a function of the signal main effects, but of the signal interactions as well, since S/N is a function of the squared observations.

3.4.3 IMPROPRIETY OF S/N AS PERFORMANCE STATISTIC

3.4.3.1 INEFFICIENCY OF S/N IN "TARGET-IS-BEST" CASE

The main objective of the Taguchi method is the minimization of the expected loss function $R(\underline{x})$. The appropriateness of the S/N ratios for this objective lies in their ability as estimators of performance measures to address the minimization of $R(\underline{x})$. Taguchi considers the optimization of the expected loss function indirectly, by considering a performance measure that reflects the expected loss function's behavior, then finding optimal design variable settings relative to that performance measure assuming that these settings also result in the minimization of $R(\underline{x})$. The S/N ratios recommended by Taguchi are the estimators of those performance measures, called performance statistics. Taguchi contends that the solutions they generate automatically translate to the minimization of $R(\underline{x})$.

To study Taguchi's contention, Leon, Shoemaker, and Kackar (1987) provide a general optimization procedure that deals with the direct optimization of $R(\underline{x})$. The objective of this procedure is to find $\underline{x}^* = (\underline{x}_d^*, \underline{x}_a^*)$ such that

$$R(\underline{x}^*) = \min_{\underline{x}_a, \underline{x}_d} R(\underline{x}_d, \underline{x}_a) \,. \qquad (3.12)$$

Recall that $\underline{x}_a$ are the adjustment variables and $\underline{x}_d$ are the factors that affect S/N.

The steps of this procedure, which will be referred to as the LSK procedure, are :

(1) Derive $P(\underline{x}_d) = R(\underline{x}_d , \underline{x}_a\{\underline{x}_d\})$ where $\underline{x}_a\{\underline{x}_d\} = \min_{\underline{x}_a} R(\underline{x}_d, \underline{x}_a)$;

(2) Find $\underline{x}_d^*$ that minimizes $P(\underline{x}_d)$; and

(3) Evaluate $\underline{x}_a^* = \underline{x}_a\{\underline{x}_d^*\}$.

Step 1 expresses $\underline{x}_a$ in terms of $\underline{x}_d$ by minimizing R with respect to $\underline{x}_a$. This is done by taking the partial derivative of R with respect to $\underline{x}_a$, equating to zero, and finding an expression for $\underline{x}_a$ in terms of $\underline{x}_d$. Substituting this back into the expression for R results in what Leon, Shoemaker, and Kackar refer to as the performance measure independent of adjustment (PerMIA). This PerMIA is what they consider to be the appropriate performance measure to be optimized with respect to the design variables $\underline{x}_d$. [Note that this is a derived performance measure, and its derivation is made possible only through the a priori knowledge of the true form of the underlying transfer or response function and the partitioning of $\underline{x}$ into $\underline{x}_a$ and $\underline{x}_d$. Hence the utility of this procedure is mainly for the understanding of the (in)efficiency of the S/N ratios which will be discussed in a later section.]

Note that the analog of this PerMIA in the Taguchi method is whatever performance measure the S/N ratio is estimating. If these two could be shown to have the same optimum conditions with respect to $\underline{x}_d$, then utilizing the S/N statistic does result in the minimization of the expected loss and, hence, is an appropriate performance statistic. Otherwise, it is an inefficient performance statistic.

The efficiency of Taguchi's S/N ratio as performance statistic to minimize expected loss will be studied from the standpoint of two approaches to the minimization of expected loss: (1) the LSK procedure where the PerMIA will be derived, and (2) the Taguchi Method. A major difference in these approaches is that the LSK procedure is purely a conceptual framework for the study of robust parameter design problems requiring a priori knowledge of the form of the transfer function relating the output to the inputs, as well as the partitioning of $\underline{x}$ into $(\underline{x}_d, \underline{x}_a)$. The Taguchi approach, on the other hand, does not require prior information about the transfer function, and the identification and partitioning of the design variables into the two subsets is a result of the analysis. Thus RPD will be applied using the prior knowledge of the above information regarding the partitioning of the design variables for the purpose of analyzing and comparing the mechanisms of both procedures.

Two contrived scenarios will be considered affording a priori knowledge of the form of the underlying transfer function relating the response to the controllable input variables. The first scenario involves the multiplicative error model

$$Y = \mu(\underline{x}_d, \underline{x}_a)\cdot\varepsilon(\underline{x}_d) \tag{3.13}$$

where

$$E_Z[\varepsilon(\underline{x}_d)] = 1 \text{ and } Var_Z[\varepsilon(\underline{x}_d)] = \sigma^2(\underline{x}_d). \tag{3.14}$$

The second scenario presents the additive error model case

$$Y = \mu(\underline{x}_d, \underline{x}_a) + \varepsilon(\underline{x}_d) \tag{3.15}$$

such that

$$E_Z[\varepsilon(\underline{x}_d)] = 0 \text{ and } Var_Z[\varepsilon(\underline{x}_d)] = \sigma^2(\underline{x}_d). \tag{3.16}$$

The problem to be analyzed is the "target is best" case, for which Taguchi defines the appropriate S/N ratio as

$$S/N = 10 \log \frac{\bar{y}^2}{s^2} \tag{3.17}$$

for a specified setting of the controllable input variables.

It is assumed that this ratio is intended as an estimator of the performance measure

$$10 \log \frac{E_Z^2(y)}{Var_Z(y)}. \tag{3.18}$$

We note again that this S/N ratio is the only Taguchi statistic that shows any semblance of some ratio involving both "signal" and "noise." It is also the only S/N statistic that does not relate directly to expected loss $R(\underline{x})$. We deviate momentarily to show the connection between these two entities. Phadke (in Nair(1992)) made use of the data rescaling transformation

$$\tilde{y}(\underline{x}) = r \cdot y(\underline{x}) \quad , \text{with} \quad r = \left(\frac{\tau}{E_Z(y(\underline{x}))} \right). \tag{3.19}$$

The data resulting from this transformation has the properties

$$Var(\tilde{y}(\underline{x})) = \left[\frac{\tau}{E_Z(y(\underline{x}))} \right]^2 \cdot Var(y(\underline{x})) \quad \text{and} \quad [\tau - E_Z(\tilde{y}(\underline{x}))] = 0. \tag{3.20}$$

Utilizing the transformed data in constructing the risk function,

$$R(\underline{x}) = \mathrm{Var}[\tilde{y}(\underline{x})] + [\tau - E_Z(\tilde{y}(\underline{x}))]^2,$$

we get

$$R(\underline{x}) = \tau^2 \left[\frac{\mathrm{Var}(y(\underline{x}))}{E_Z^2(y(\underline{x}))} \right]. \tag{3.21}$$

Comparing this to the expression of (3.18), the direct link between the minimization of expected loss and the maximization of S/N is established. It is to be noted, however, that the above result was made possible through the use of the rescaling transformation of (3.19). This transformation is available only when the mean parameter, $E(y(\underline{x}))$, is known or can be accurately estimated.

However, the use of data transformations is not part of the Taguchi approach as noted by Box (1988). In the scenarios of "smaller-the-better" and "larger-the-better", the connection between risk and S/N is obvious, with S/N being the naive estimator of $R(\underline{x})$ in both cases (see (3.3) and (3.4); (3.5) and (3.6)). This is not the case for the "target-is-best" scenario. Thus there is a need to examine the efficiency of S/N statistic in the original setting of the "target-is-best" problem; i.e., the role of S/N without the use of data transformations has to be ascertained.

3.4.3.1.1 MULTIPLICATIVE ERROR MODEL CASE

Given the multiplicative error model and quadratic error loss function, the expected loss function is

$$R(\underline{x}) = R(\underline{x}_d, \underline{x}_a) = \mu^2(\underline{x}_d, \underline{x}_a)\sigma^2(\underline{x}_d) + [\mu(\underline{x}_d, \underline{x}_a) - \tau]^2. \tag{3.22}$$

Utilizing the LSK procedure to find the optimal solution $\underline{x}^* = (\underline{x}_d^*, \underline{x}_a^*)$ to the original problem of minimizing $R(\underline{x})$, we have the following steps:

(1) Taking the partial derivative of R with respect to $\underline{x}_a$ and equating to zero,

$$\frac{\partial R(\underline{x}_d.\underline{x}_a)}{\partial \underline{x}_a} = 2\frac{\partial \mu(\underline{x}_d.\underline{x}_a)}{\partial \underline{x}_a}\cdot[\mu(\underline{x}_d.\underline{x}_a)\cdot\{1+\sigma^2(\underline{x}_d)\} - \tau] = 0, \tag{3.23}$$

yields the normal equation

$$\mu(\underline{x}_d.\underline{x}_a^*) = \frac{\tau}{1+\sigma^2(\underline{x}_d)}. \tag{3.24}$$

which defines $\underline{x}_a^*$ in terms of $\underline{x}_d$. Substituting this result into the expression for R yields

$$R(\underline{x}_d, \underline{x}_a^*\{\underline{x}_d\}) = \frac{\tau^2\sigma^2(\underline{x}_d)}{1+\sigma^2(\underline{x}_d)} = P(\underline{x}_d). \tag{3.25}$$

This is the PerMIA of LSK.

(2) $P(\underline{x}_d)$ can be written as

$$P(\underline{x}_d) = \frac{\tau^2}{\dfrac{1}{\sigma^2(\underline{x}_d)}+1}. \tag{3.26}$$

Therefore, the solution that minimizes P, $\underline{x}_d^*$, is also the solution that minimizes $\sigma^2(\underline{x}_d)$ as P is an increasing function of $\sigma^2(\underline{x}_d)$. Hence they have the same minimizing solution $\underline{x}_d^*$.

(3) To find $\underline{x}_a^*$, substitute $\underline{x}_d^*$ into the normalizing equation defined in step (1),

$$\mu(\underline{x}_d^*, \underline{x}_a^*) = \frac{\tau}{1+\sigma^2(\underline{x}_d^*)}, \tag{3.27}$$

and solve for $\underline{x}_a^*$.

Therefore, given the multiplicative error model, the LSK optimal solution for the design variables is one that minimizes the variance and adjusts the mean according to the previous expression.

To examine the optimal solution generated via the use of Taguchi's S/N ratio as performance statistic, the intended performance measure is considered. This performance measure,

$$10\log\frac{E_Z^2(y)}{Var_Z(y)},$$

combined with the specified multiplicative error transfer function formulation for the response results in the simplification

$$10\log\frac{\mu^2(\underline{x}_d, \underline{x}_a)}{\mu^2(\underline{x}_d, \underline{x}_a)\cdot\sigma^2(\underline{x}_d)} = 10\log\frac{1}{\sigma^2(\underline{x}_d)}. \tag{3.28}$$

To maximize this performance measure, which is what Taguchi aims to accomplish by maximizing its estimator S/N, the required solution $\underline{x}_d^*$ should minimize $\sigma^2(\underline{x}_d)$. Therefore, for the case of a multiplicative error transfer function, maximizing S/N does indeed lead to the minimization of the expected loss with respect to $\underline{x}_d$, hence partially validating Taguchi's claim that S/N is an appropriate performance statistic.

3.4.3.1.2 ADDITIVE ERROR MODEL

Suppose the additive error model given in equation (3.15) defines the form of the underlying transfer function for the response. The expected loss function given the quadratic error loss function is

$$R(\underline{x}) = R(\underline{x}_d, \underline{x}_a) = \sigma^2(\underline{x}_d) + [\mu(\underline{x}_d, \underline{x}_a) - \tau]^2. \tag{3.29}$$

The LSK approach proceeds as follows:

(1) Minimizing with respect to $\underline{x}_a$ defines the condition

$$\mu(\underline{x}_d, \underline{x}_a^*) = \tau \tag{3.30}$$

which when substituted back into the expression for the expected loss function, gives $\sigma^2(\underline{x}_d)$ as the performance measure independent of adjustment.

(2) $\underline{x}_d^*$ is the solution that minimizes $\sigma^2(\underline{x}_d)$.

(3) $\underline{x}_a^*$ is the solution that satisfies the unbiasedness condition derived in step (1) utilizing $\underline{x}_d^*$:

$$[\mu(\underline{x}_d^*, \underline{x}_a) - \tau] = 0. \tag{3.31}$$

To determine if the Taguchi tactic of employing the S/N ratio results in the same optimal solution generated by the previous approach, again consider the intended performance measure which, in combination with the additive error model, simplifies to

$$10\log\frac{\mu^2(\underline{x}_d,\underline{x}_a)}{\sigma^2(\underline{x}_d)}.$$

This expression is maximized when its estimator, S/N, is maximized. Note that the maximization of S/N does not necessarily result in the minimization of $\sigma^2(\underline{x}_d)$. Indeed, a solution to the above maximization exercise could be to specify $\underline{x}_d$ and $\underline{x}_a$ to maximize the numerator $\mu^2(\underline{x}_d, \underline{x}_a)$. This solution does not necessarily imply the minimization of the numerator $\sigma^2(\underline{x}_d)$. By transitivity, the use of S/N as the performance statistic does not necessarily lead to the minimization of the expected loss, hence the inefficiency of S/N as performance statistic.

3.4.3.2 INEFFICIENCY ON S/N IN MEAN ADJUSTMENT PROCESS

Let us assume that the S/N statistic properly addresses the risk $R(\underline{x})$. That is, with the maximization of S/N, minimization of $R(\underline{x})$ is likewise effected. Now consider the two-step optimization procedure outlined earlier, wherein the input factor set $\underline{x}$ is partitioned into $\underline{x}_d$ and $\underline{x}_a$, where $\underline{x}_d$ is the subset of factors affecting S/N and $\underline{x}_a$ is the subset of the signal variables deemed inconsequential to S/N but does exert some influence on the mean. In the procedure, the levels for $\underline{x}_d$ are chosen to maximize S/N, and the settings for $\underline{x}_a$ are selected to adjust the mean to the target value.

Since the adjustment variables $\underline{x}_a$ affect the mean, they must affect the bias component of R($\underline{x}$) given a quadratic error loss function. Therefore, since R($\underline{x}$) is affected by $\underline{x}_a$, S/N must also be impacted by these same variables. This is a consequence of the earlier assumption that S/N is a proper estimator, or is some function of a proper estimator, of R($\underline{x}$). Hence, any tuning or adjustment variable identified through the use of S/N is not a true mean response adjustment variable independent of the performance measure being addressed by S/N. This is especially true when the bias portion of R($\underline{x}$) dominates the variance portion.

It would seem that to properly identify tuning variables, the appropriate performance statistic should address just the variance component of R($\underline{x}$), exclusive of the bias component. Any variable that does not affect this statistic but still influences R($\underline{x}$) must certainly impact just the bias component, and, similarly, the mean. This is true assuming, of course, that adequate variable screening has been completed in the earlier stages of the study wherein the set of relevant signal inputs $\underline{x}$ was determined.

3.5 A MODIFIED ROBUST PARAMETER DESIGN APPROACH

Consider the system optimization problem of the continuous review inventory model presented in Chapter One where the component stochastic processes and the output process are nonstationary. In Chapter Two RSM techniques for optimizing with respect to the mean response of transient systems were developed. It was also shown that with the inclusion of dynamic system behavior within the simulation model to make it more realistic, a new problem surfaced in the form of increased response variance.

In this section we again focus on the same inventory model. In addition to the original objective of optimizing mean response, we also consider the problem of increased variance of the response. We will treat nonstationarity as an internal source of system performance variation in the context of RPD and follow the Taguchi philosophy of directly and simultaneously considering the process mean and variance in specifying an operating policy for the system.

In the process of applying the techniques of RPD, we also take into consideration the major criticisms in the previous section and offer some revisions to the original methodology of Taguchi to address these criticisms. This strategy combines some of the best features of RPD and RSM to formulate a revised RPD approach. This revised approach is applied to the inventory problem and numerical results are presented.

3.5.1 MODIFICATIONS TO THE TAGUCHI METHOD

We focus first on the class of criticisms regarding the inadequacy of the optimization search process. These criticisms involve the lack of sequential experimentation to locate the subregion of the operability space that contains the optimal policy, and the inflexibility of the marginal means analysis in locating the best solution within that subregion. The proposed revisions to RPD to alleviate these problems are based on the incorporation of the sequential search procedures of RSM. We start by postulating the existence of a response surface for the performance measure whose estimator is Taguchi's S/N statistic. We search for the solution that generates the maximum S/N response utilizing response surface methods, sequentially fitting first order S/N response surface models to find the best operating subregion. Applying second order analysis techniques on second order response surface models for S/N, estimates for the optimum operating policy settings are determined.

The second class of criticisms deals mainly with experimental design issues. Under the RPD crossed array scheme, certain saturated or near saturated first order designs are recommended for the inner array experimental design on the signal variables. These designs do not accommodate the estimation of signal variable interactions. We address this problem by proposing the following changes in the construction of the inner design array. For the first order phase of the analysis, 2^k factorial or 2^{k-p} fractional factorial designs with sufficient experimental points to allow the estimation of the main effects and signal factor interactions are considered. The same class of designs is proposed for the outer array. In the second order phase, the use of central composite designs (CCD's) or small composite designs (SCD's) allows for the estimation of the main effects, interactions, and the quadratic effects of the signal variables.

Since the use of such designs will most likely result in an increase in the number of experimental settings for the inner array when compared to those designs recommended by Taguchi, we attempt to offset this potential drawback by fractionating more heavily on the outer array. The motivation for this procedure is twofold: (1) noise by noise interactions are not a major concern since they are not essential in making process performance insensitive to the noise factors (see Phadke in Nair (1992)); and (2) the purpose of the outer array is to provide just a rough measure of sensitivity. Possible alternatives for the outer array are other saturated or near saturated first order designs such as Plackett-Burman designs.

The third set of criticisms involves the use of the S/N statistic. In Section 3.4, it was shown that the maximization of this statistic as formulated by Taguchi does not necessarily lead to the minimization of $R(\underline{x})$ in the "target-is-best" case. We address this problem by formulating a different S/N statistic for this case, one that directly addresses expected loss. Recall that

$$R(\underline{x}) = E_Z(y(\underline{x}) - \tau)^2. \qquad (3.32)$$

If we defined w as

$$w = |y(\underline{x}) - \tau|, \qquad (3.33)$$

$R(\underline{x})$ can be rewritten as

$$R(\underline{x}) = E_Z(|y(\underline{x}) - \tau|)^2 = E_Z(w)^2. \tag{3.34}$$

Using the same procedure considered in formulating the S/N statistic for the "smaller-the-better" case, we get

$$S/N = -10\log\frac{\sum_{j=1}^{m} w_{ij}^2}{m} \tag{3.35}$$

where $w_{ij} = |y_{ij} - \tau|$.

Since τ is assumed to be a known, fixed value, the above data transformation involving a combination of a recentering and the absolute value transformation does not involve any unknown model parameters. The S/N statistic is therefore a naive estimator of $R(\underline{x})$ and maximization of S/N results in minimization of expected loss.

3.5.2 IMPLEMENTATION AND NUMERICAL RESULTS

We now analyze the problem of the continuous review inventory model whose internal process generators are transitory using the revised RPD approach. The optimization problem involves the minimization of average monthly cost which is an example of the "smaller-the-better" case. Implementation issues are considered during this process, and numerical results from the application of the revised methodology are presented.

The first step is the classification of the system input variables. The uncontrollable input factors, reorder point (ξ_1) and reorder quantity (ξ_2), are classified as signal or design variables. The uncontrollable model parameters, the means of the distributions of the internal stochastic process generators, are classified as noise variables due to their variance inflating effect on the system response of interest, monthly total cost, as shown in Chapter Two. Recall that these noise variables are mean interarrival time (ζ_1), mean demand (ζ_2), and mean delivery lead time (ζ_3).

Next we select an arbitrary point in the solution space or operability region of the signal variables. This point will serve as the center of the initial experimental subregion in the search for the subregion containing the optimal solution. In the example problem, we chose the same initial design center used in illustrating the RSM techniques of Chapter Two, namely

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1000 \\ 1000 \end{bmatrix}.$$

The initial experimental subregion centered on the above location is defined by specifying the width of the relevant ranges for the input variables. Arbitrarily setting d_i, the width of the relevant range or spacing between the high and low levels of ξ_i, equal to 1000 units for both signal variables as in Chapter Two, the initial experimental region becomes

$$\{(\xi_1, \xi_2) : \xi_1 \in [500,1500] \,,\ \xi_2 \in [500,1500]\}.$$

We use the transformation of (2.11) to redefine the signal variables in terms of their coded version x_i; i.e.,

$$x_i = 2\left(\frac{\xi_i - \overline{\xi}_i}{d_i}\right), \quad \text{with} \quad d_i = \xi_{i,HIGH} - \xi_{i,LOW}, \quad \overline{\xi}_i = \left(\frac{\xi_{i,HIGH} + \xi_{i,LOW}}{2}\right).$$

The coding convention for the noise variables is different from that of the signal variables as they are random variables distributed as follows:

$$\zeta_1 \sim \text{Normal}(.20, .05^2)$$
$$\zeta_2 \sim \text{Normal}(100, 10^2)$$
$$\zeta_3 \sim \text{Normal}(15, 2^2).$$

The purpose of using different fixed settings for the noise variables is to gauge a signal variable combination's sensitivity to the random behavior of the noise variables. The search process is undertaken in the solution space of the signal variables, not in the space of the noise variables. To define the coding for the noise variables, we make use of the centering and rescaling transformation

$$z_i = \frac{\zeta_i - \mu_{\zeta_i}}{\sigma_{\zeta_i}}, \tag{3.36}$$

where μ_{ζ_i} and σ_{ζ_i} are the mean and standard deviation of the *ith* noise variable in its natural space, respectively.

The high and low settings for the noise variables are specified by

$$\zeta_i = \mu_{\zeta_i} \pm k \cdot \sigma_{\zeta_i}. \tag{3.37}$$

The most common values for k are 1, 2, and $(3)^{.5}/2$. (see Kackar (1985) and Myers et al. (1992)). We choose k = 1 to define the relevant ranges for the noise variables for two reasons: (1) the intervals defined for that value of k contain the most likely values of the means (approximately 68% in the normal case), and (2) using the coding transformation of (3.36), we get

$$z_{i,\mathrm{LOW}} = -1 \,, \quad z_{i,\mathrm{HIGH}} = +1 \,, \quad E(z_i) = 0 \,, \quad \mathrm{Var}(z_i) = 1.$$

Given the above coding schemes, we define our inner and outer array designs, **D** and **Z**. For the inner array design a 2^2 factorial design augmented by 2 center runs defines the levels of the signal variables x_1 and x_2,, and is adequate for estimating the signal main effects. Thus

$$\begin{array}{cc} & \begin{array}{cc} \underline{x_1} & \underline{x_2} \end{array} \\ \mathbf{D} = & \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} . \end{array}$$

The outer array,

$$\begin{array}{cc} & \begin{array}{ccc} \underline{z_1} & \underline{z_2} & \underline{z_3} \end{array} \\ \mathbf{Z} = & \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} , \end{array}$$

is a 2^{3-1} fractional factorial design. The Kronecker product of these two design matrices gives the overall design matrix requiring nxm = 6x4 = 24 experimental points or runs.

For the inventory example, the original optimization objective is to minimize average total monthly cost. The corresponding S/N statistic for the "smaller-the-better" scenario is

$$S/N_i = -10\log\frac{\sum_{j=1}^{4} y_{ij}^2}{4} \quad , \quad i = 1,\ldots,6,$$

where the i-subscript corresponds to the *ith* combination in the signal variables in the inner array, and j denotes the *jth* setting of the noise variables in the outer array. The experimental results for the initial subregion are presented in the following table.

Table 3.1 Experimental Results for the First Experimental Subregion

x_1	x_2	ξ_1	ξ_2	y_1	y_2	y_3	y_4	S/N
-1	-1	500	500	61520.2	137134	112689	79717.2	-100.175
-1	1	500	1500	55643.0	129215	106950	74346.1	-99.633
1	-1	1500	500	51143.0	125337	103997	71098.4	-99.319
1	1	1500	1500	41680.8	112232	93633	62044.0	-98.285
0	0	1000	1000	50593.9	124280	102535	70394.9	-99.226
0	0	1000	1000	50564.1	123892	102606	70459.3	-99.217

The S/N regression model fitting results are summarized in the following table.

Table 3.2 Fitted First-Order S/N Model for the Initial Subregion

```
The regression equation is
s/n = - 99.3 + 0.551 x1 + 0.394 x2

Predictor        Coef        Stdev     t-ratio        p
Constant     -99.3092       0.0682    -1457.02    0.000
x1            0.55096      0.08348        6.60    0.007
x2            0.39402      0.08348        4.72    0.018

s = 0.1670       R-sq = 95.6%

Analysis of Variance

SOURCE        DF         SS          MS          F        p
Regression     2    1.83523     0.91761      32.92    0.009
Error          3    0.08362     0.02787
Total          5    1.91885
```

From the above table, the maximizing or steepest ascent search direction is

$$\underline{b}_l = \begin{bmatrix} .5510 \\ .3940 \end{bmatrix} \Rightarrow \underline{b}_{ls} = \begin{bmatrix} .8134 \\ .5817 \end{bmatrix}$$

where $\underline{b}_{ls}$ is the standardized gradient vector. We define exploratory points $\underline{x}_r$ along this gradient direction again using the relationship

$$\underline{x}_r = \underline{x}_q + \underline{b}_{qs} \cdot r \qquad (3.38)$$

where $\underline{x}_q$ is the current design center, and r is the step size along the steepest ascent direction. The values of r are increasing exponents of 2, i.e., $r = 2^t$, $t = 1,2,3, \ldots$.

The results for the exploratory runs for the first search direction are listed below.

Table 3.3 Results of Exploratory Runs for the First Search Direction

r	x_1	x_2	ξ_1	ξ_2	S/N
1	0.8134	0.5817	1406.7	1290.8	-98.6403
2	1.6269	1.1633	1813.4	1581.7	-98.0560
4	3.2537	2.3266	2626.9	2163.3	-96.8450
8	6.5075	4.6533	4253.7	3326.6	-94.2157
16	13.0149	9.3065	7507.5	5653.3	-88.2744
32	26.0299	18.6130	14014.9	10306.5	-87.4805

The trial runs were terminated after t = 5 as the next step size was deemed excessive, and the incremental improvements in S/N did not warrant further exploratory runs. From the above results, the most promising region seems to be that centered in the neighborhood of

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 14000 \\ 10300 \end{bmatrix}.$$

At this point in the analysis, another iteration of the first order phase begins with a new experimental subregion being defined about the new location, experimental runs being undertaken using the same first order crossed-array design, a first order S/N regression model being fitted, and, if necessary, exploratory runs undertaken along the steepest ascent direction. This sequential procedure identified three additional subregions for exploration. The experimental and model-fitting results and the exploratory search data are provided in the following eight tables.

Table 3.4 Experimental Results for the Second Experimental Subregion

ξ_1	ξ_2	y_1	y_2	y_3	y_4	S/N
13500	9800	28111.6	19345.4	17532.6	22661.2	-86.9579
13500	10800	28951.4	20146.6	18478.2	23646.8	-87.2930
14500	9800	30006.1	21146.8	19436.5	24622.6	-87.6559
14500	10800	31018.0	21916.4	20361.3	25549.4	-87.9758
14000	10300	29556.3	20482.6	18965.1	24025.7	-87.4625
14000	10300	29491.4	20561.5	18930.1	24142.3	-87.4697

Table 3.5 Fitted First-Order S/N Model for the Second Subregion

```
The regression equation is
s/n = - 87.5 - 0.345 x1 - 0.164 x2

Predictor         Coef         Stdev    t-ratio        p
Constant      -87.4691        0.0025  -35049.04    0.000
x1           -0.345196      0.003057    -112.94    0.000
x2           -0.163731      0.003057     -53.57    0.000

s = 0.006113     R-sq = 100.0%

Analysis of Variance

SOURCE        DF          SS           MS          F          p
Regression     2     0.58387      0.29194    7812.29      0.000
Error          3     0.00011      0.00004
Total          5     0.58398
```

Table 3.6 Results of Exploratory Runs for the Second Search Direction

r	ξ_1	ξ_2	S/N
1	13548.2	10085.8	-87.0816
2	13096.4	9871.5	-86.6967
4	12192.9	9443.0	-85.9647
8	10385.8	8586.1	-85.1772

Table 3.7 Experimental Results for the Third Experimental Subregion

ξ_1	ξ_2	y_1	y_2	y_3	y_4	S/N
9500	7500	18010.5	21763.6	20704.0	13178.7	-85.4418
9500	8500	18809.9	21435.7	20269.5	14154.9	-85.5161
10500	7500	19967.0	17887.9	16640.1	14793.7	-84.8230
10500	8500	20880.2	18168.3	17195.9	15510.7	-85.1267
10000	8000	19464.9	19863.1	18323.1	14259.2	-85.1603
10000	8000	19466.8	19460.3	18580.2	14334.8	-85.1467

Table 3.8 Fitted First-Order S/N Model for the Third Subregion

```
The regression equation is
s/n = - 85.2 + 0.252 x1 - 0.0945 x2

Predictor        Coef        Stdev    t-ratio        p
Constant     -85.2024       0.0337   -2529.10    0.000
x1            0.25205      0.04126       6.11    0.009
x2           -0.09447      0.04126      -2.29    0.106

s = 0.08252      R-sq = 93.4%

Analysis of Variance

SOURCE        DF         SS          MS          F        p
Regression     2    0.28982     0.14491      21.28    0.017
Error          3    0.02043     0.00681
Total          5    0.31025
```

Table 3.9 Results of Exploratory Runs for the Third Search Direction

r	x_1	x_2	ξ_1	ξ_2	S/N
1	0.93638	-0.35100	10468.2	7824.5	-84.9483
2	1.87275	-0.70200	10936.4	7649.0	-84.8400
4	3.74550	-1.40400	11872.7	7298.0	-85.0087

Table 3.10 Experimental Results for the Fourth Experimental Subregion

ξ_1	ξ_2	y_1	y_2	y_3	y_4	S/N
10500	7000	19574.6	18193.1	16598.0	14340.1	-84.7541
10500	8000	20462.3	18204.8	16999.2	15164.7	-85.0142
11500	7000	21639.0	16401.0	14523.0	16288.6	-84.8197
11500	8000	22490.9	16492.4	14824.1	17133.4	-85.0892
11000	7500	21047.1	16813.8	15225.7	15723.5	-84.7883
11000	7500	21014.9	16882.0	15472.8	15641.0	-84.8097

Table 3.11 Fitted First-Order S/N Model for the Fourth Subregion

```
The regression equation is
s/n = - 84.9 - 0.0351 x1 - 0.132 x2

Predictor        Coef        Stdev     t-ratio        p
Constant     -84.8792       0.0330    -2575.78    0.000
x1           -0.03513      0.04036       -0.87    0.448
x2           -0.13239      0.04036       -3.28    0.046

s = 0.08072      R-sq = 79.3%

Analysis of Variance

SOURCE        DF          SS          MS          F        p
Regression     2    0.075049    0.037524       5.76    0.094
Error          3    0.019546    0.006515
Total          5    0.094595
```

The model-fitting results for the fourth subregion show that the model fit was marginally adequate to inadequate as given by r^2=.793 and a p-value of 0.094 for the overall F-test. It was surmised that the inadequacy in model fit was due to the presence of considerable curvature in the S/N response surface. The analysis therefore shifted to the second order phase.

In the second order analysis phase, the inner array utilized was a CCD with the axial points drawn in towards the design center half the distance of the design radius in the coded variable space (i.e., α = .707) as a protection against model bias. The **D** inner design array,

$$\mathbf{D} = \begin{array}{c} \begin{array}{cc} \underline{x_1} & \underline{x_2} \end{array} \\ \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ .707 & 0 \\ -.707 & 0 \\ 0 & .707 \\ 0 & -.707 \end{bmatrix} \end{array}$$

consists of the first-order design for the last subregion augmented by axial design points. The coordinates of the axial points in the space of the natural variables are specified by the reverse transformation

$$\xi_i = (x_i \cdot d_i / 2) + \overline{\xi_i}. \tag{3.39}$$

Data from the previous trials in the last iteration of the first order phase, along with observations on the axial points, and are summarized in the Table 3.12. A second order regression model was fitted and an analysis of the fitted S/N response function was conducted using the SAS procedure RSREG. A partial SAS output of the results is shown in Table 3.13.

Table 3.12 Augmented Results for the Fourth Experimental Subregion

x_1	x_2	ξ_1	ξ_2	y_1	y_2	y_3	y_4	S/N
-1.000	-1.000	10500	7000	19574.6	18193.1	16598.0	14340.1	-84.7541
-1.000	1.000	10500	8000	20462.3	18204.8	16999.2	15164.7	-85.0142
1.000	-1.000	11500	7000	21639.0	16401.0	14523.0	16288.6	-84.8197
1.000	1.000	11500	8000	22490.9	16492.4	14824.1	17133.4	-85.0892
0.000	0.000	11000	7500	21047.1	16813.8	15225.7	15723.5	-84.7883
0.000	0.000	11000	7500	21014.9	16882.0	15472.8	15641.0	-84.8097
0.707	0.000	11354	7500	21735.4	16290.6	14965.2	16256.4	-84.8644
-0.707	0.000	10646	7500	20396.2	17430.1	16479.2	14993.9	-84.8294
0.000	0.707	11000	7854	21348.3	17050.7	15467.9	15982.3	-84.9177
0.000	-0.707	11000	7146	20698.9	16858.2	15464.3	15414.3	-84.7326

Table 3.13 S/N Second-order Model Analysis Results (Fourth Subregion)

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	11000	500.000000
X2	7500.000000	500.000000

Response Surface for Variable s/n

Response Mean	-84.861930
Root MSE	0.010062
R-Square	0.9965
Coef. of Variation	-0.0119

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	0.092715	0.7904	457.9	0.0000
Quadratic	2	0.024157	0.2059	119.3	0.0003
Crossproduct	1	0.000022090	0.0002	0.218	0.6647
Total Regress	5	0.116894	0.9965	230.9	0.0001

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	0.000405	0.000101

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	-129.744991	4.096620	-31.671	0.0000	-84.803713
X1	1	0.007034	0.000957	7.354	0.0018	-0.033059
X2	1	0.002029	0.000681	2.979	0.0408	-0.132063
X1*X1	1	-0.000000320	4.2932714E-8	-7.443	0.0017	-0.079883
X2*X1	1	-9.4E-9	2.0123422E-8	-0.467	0.6647	-0.002350
X2*X2	1	-0.000000146	4.2932714E-8	-3.400	0.0273	-0.036492

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	0.011097	0.003699	36.538	0.0023
X2	3	0.088440	0.029480	291.2	0.0000

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Critical Value Uncoded
X1	-0.180394	10910
X2	-1.803655	6598.172468

Predicted value at stationary point -84.681633

Eigenvalues	Eigenvectors X1	Eigenvectors X2
-0.036461	-0.027050	0.999634
-0.079915	0.999634	0.027050

Stationary point is a maximum.

The above analysis reveals the presence of stationary point resulting in a maximum S/N predicted mean response. It also indicates, however, that the estimate of this point is outside the current experimental subregion. Since conclusions regarding the nature of the fitted response surface outside the current subregion are questionable, another iteration of the second order phase was undertaken at a new location.

A new experimental subregion centered close to the previous estimate of the stationary point was chosen. Additionally, the natural variable spacing, d_i, is doubled for greater coverage of the operability space. The same inner and outer design matrices were utilized and a new set of experimental runs undertaken. Responses at these design points are provided in Table 3.14.

Table 3.14 Experimental Results for the Fifth Experimental Subregion

ξ_1	ξ_2	y_1	y_2	y_3	y_4	S/N
10000	5500	17335.4	20430.0	19034.4	12452.2	-84.8970
10000	7500	19042.4	19408.3	18364.8	13988.7	-85.0251
12000	5500	21431.0	15230.8	13605.8	16116.2	-84.5336
12000	7500	23048.7	16085.4	14582.1	17633.7	-85.1639
11000	6500	20268.7	16606.4	15276.5	14888.9	-84.5547
11000	6500	20241.9	17118.9	15351.7	14887.9	-84.6244
11707	6500	21539.0	15718.3	14178.0	16198.4	-84.6775
10293	6500	18839.5	18385.4	17365.2	13621.7	-84.6984
11000	7207	20802.6	17051.7	15513.6	15426.1	-84.7789
11000	5793	19627.8	16422.3	15130.2	14327.6	-84.3503

The results of the RSREG procedure employed on the above data are given in the following partial SAS output.

Table 3.15 S/N Second Order Model Analysis Results (Fifth Subregion)

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	11000	1000.000000
X2	6500.000000	1000.000000

Response Surface for Variable s/n

Response Mean	-84.730380
Root MSE	0.065845
R-Square	0.9675
Coef. of Variation	-0.0777

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	0.236797	0.4434	27.309	0.0047
Quadratic	2	0.216823	0.4060	25.005	0.0055
Crossproduct	1	0.063051	0.1181	14.543	0.0189
Total Regress	5	0.516671	0.9675	23.834	0.0045

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	0.017342	0.004336

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	-130.104758	6.992143	-18.607	0.0000	-84.560689
X1	1	0.007312	0.001564	4.675	0.0095	0.047878
X2	1	0.001771	0.000985	1.798	0.1466	-0.212297
X1*X1	1	-0.000000293	7.040947E-8	-4.163	0.0141	-0.293089
X2*X1	1	-0.000000126	3.2922489E-8	-3.814	0.0189	-0.125550
X2*X2	1	-4.631422E-8	7.040947E-8	-0.658	0.5466	-0.046314

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	0.149637	0.049879	11.505	0.0195
X2	3	0.290263	0.096754	22.316	0.0058

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Critical Value Uncoded
X1	0.806790	11807
X2	-3.385454	3114.546241

Predicted value at stationary point -84.182014

Eigenvalues	Eigenvectors X1	Eigenvectors X2
-0.031263	-0.233151	0.972440
-0.308140	0.972440	0.233151

Stationary point is a maximum.

The above analysis also reveals the presence of stationary point resulting in a maximum S/N predicted mean response. As with the previous iteration of the second-order analysis phase, it also indicates an stationary point estimate exterior to the current subregion under study. Hence another iteration of the second order phase was undertaken at a new location. The sixth experimental subregion is defined near the previous estimate of the stationary point, but not as close as in the previous iteration since that estimated stationary point was quite distant from the previous subregion. To compensate for this, the natural variable spacing d_i was again increased by fifty percent to include the estimated stationary point. Again, the same inner and outer design matrices were utilized, with the new set of experimental runs producing the following results.

Table 3.16 Experimental Results for the Sixth Experimental Subregion

ξ_1	ξ_2	y_1	y_2	y_3	y_4	S/N
10500	2500	16931.7	19072.6	18027.3	12318.2	-84.4993
10500	5500	18382.7	18314.3	16762.0	13255.7	-84.5102
13500	2500	23006.7	16577.4	14351.3	18020.0	-85.2336
13500	5500	24358.5	16314.0	14573.4	19042.7	-85.5464
12000	4000	20393.6	15047.4	13202.2	15184.3	-84.1796
12000	4000	20437.9	15054.2	13423.1	15195.2	-84.2136
13061	4000	22445.4	15450.2	13342.1	17224.5	-84.8330
10939	4000	18293.8	16787.2	15488.3	13168.0	-84.1067
12000	5061	21071.1	15483.2	13519.3	15787.9	-84.4550
12000	2939	20011.4	15232.7	13465.1	14914.6	-84.1340

Table 3.17 S/N Second Order Model Analysis Results (Fifth Subregion)

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	12000	1500.000000
X2	4000.000000	1500.000000

Response Surface for Variable s/r

Response Mean	-84.571140
Root MSE	0.073578
R-Square	0.9900
Coef. of Variation	-0.0870

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	1.104071	0.5091	102.0	0.0004
Quadratic	2	1.020341	0.4705	94.236	0.0004
Crossproduct	1	0.022786	0.0105	4.209	0.1095
Total Regress	5	2.147198	0.9900	79.324	0.0004

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	0.021655	0.005414

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	-118.565973	4.689796	-25.282	0.0000	-84.195065
X1	1	0.005710	0.000841	6.786	0.0025	-0.456789
X2	1	0.001043	0.000342	3.048	0.0381	-0.110137
X1*X1	1	-0.000000245	3.4940111E-8	-7.012	0.0022	-0.551265
X2*X1	1	-3.354444E-8	1.6350716E-8	-2.052	0.1095	-0.075475
X2*X2	1	-8.923974E-8	3.4940111E-8	-2.554	0.0630	-0.200789

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	1.332398	0.444133	82.038	0.0005
X2	3	0.118760	0.039587	7.312	0.0422

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Critical Value Uncoded
X1	-0.400690	11399
X2	-0.198951	3701.573198

Predicted value at stationary point -84.092593

Eigenvalues	Eigenvectors X1	Eigenvectors X2
-0.196772	-0.105857	0.994381
-0.555282	0.994381	0.105857

Stationary point is a maximum.

The results of the analysis of the fitted second order response surface model indicate an excellent model fit (i.e., r^2 = .9900 and p-value = .0004 for the total regression F-statistic). It also shows an estimated stationary point within the current subregion resulting in a maximal S/N value. The corresponding estimate of the true optimal solution is

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \cong \begin{bmatrix} 11399 \\ 3702 \end{bmatrix},$$

with a predicted maximal S/N value of -84.0930. At this point, the second order analysis phase was terminated.

The above optimal solution recommends that when the inventory position drops to or beyond ξ_1 = 11,399 units, an order should be placed for an additional ξ_2 = 3,702 units of inventory.

3.6 CONCLUDING REMARKS

Taguchi's main contribution to quality improvement is his philosophy of minimizing the expected loss to society through the simultaneous consideration of the mean and variance of product or process response. This chapter illustrated how his philosophy can be applied to the problem of system optimization under nonstationary operating conditions in computer simulation studies. In addition to finding a operating policy that would result in an optimum mean performance, the added problem of increased performance variation which is attributed to the dynamic behavior of the system was also jointly addressed.

With the revised RPD approach, some of the best features of RSM and Taguchi Methods are combined to address the major criticisms of the original RPD approach. The resulting methodology based on the analysis of the response surface for S/N provides a systematic, flexible, and efficient determination of an optimal operating policy. With the use of experimental designs on the noise variables to gauge performance sensitivity, a static simulation model is utilized for each experimental run, providing added ease in programming as noted in Chapter 2.

However, there are problems inherent to the simultaneous consideration of the process mean and variance. Some of these problems were highlighted by the criticisms regarding the use of the S/N performance statistic to accomplish the minimization of expected loss. While most of the critical problems were resolved through the proposed revisions in the previous section, a facet of the problem has been neglected. With the simultaneous consideration of mean and variance through the combination of these two responses into one performance measure, understanding of the separate natures of these two responses is not possible . Predictive capability for either response is available only through the use of additional confirmatory experimental runs. While the solution formulated through the use of the S/N model presents some optimal mix in variance and bias of the performance response, the resulting optimal proportions of these components of expected loss are neither explained nor estimable.

Chapter 4

DUAL RESPONSE METHODS

4.1 INTRODUCTION

In the previous chapter, the problem of system optimization in the presence of nonstationary system behavior was analyzed using a modification of Taguchi's tactics for robust parameter design. The methodology of Chapter 3 jointly addressed the dual problems of optimizing mean performance and minimizing the variation in performance brought about by nonstationarity in system behavior. The criterion considered was the minimization of risk or expected loss, given a quadratic error loss function, via the revised robust parameter design (RPD) strategy that combines the best features of response surface methodology (RSM) and the Taguchi philosophy. With this approach, Taguchi's estimator for expected loss, the signal to noise ratio (S/N), was modeled directly and analyzed using RSM techniques.

While this analysis produces an estimated optimal solution with respect to minimizing expected loss or maximizing S/N, it lacks the capability to separately predict the mean and variance response levels at that point. Its predictive ability is limited to estimating S/N at the estimated optimal solution. Although expected loss is minimized through the analysis of S/N, it cannot be partitioned into its variance and bias components to give information regarding the separate contribution of the response mean and variance. Thus this information must be derived empirically, making confirmatory experimental runs essential for predicting the estimated levels for mean and variance corresponding to the estimated optimal solution for the S/N ratio.

Another unresolved issue in S/N modeling pertains to the inefficiency of the mean adjustment process. Specifically, the analysis of S/N is supposed to identify tuning or adjustment variables, signal variables that are inconsequential to S/N but which do influence mean response. It was argued in the previous chapter that S/N is unsuitable for the process of partitioning the signal variables into tuning variables and variables affecting the performance measure independent of adjustment (PerMIA) of Leon, Shoemaker, and Kackar (1987). Furthermore, the possibility of the signal variables conveniently partitioning into these mutually exclusive subsets is questionable.

Leon, Shoemaker, and Kackar (1987) note that Taguchi gives no justification for the use of S/N in identifying tuning variables. They argue that variance is the logical choice for PerMIA for that purpose. This observation is important with respect to Taguchi's two-step optimization procedure for the analysis of S/N since it takes advantage of the natural partitioning of expected loss. The first step would involve minimizing the variance component of expected loss, then minimizing the bias portion by adjusting the mean. In the process, signal variables affecting mainly the variance, termed dispersion variables, are determined, and their complement, tuning variables, that affect mainly the mean and, consequently, the bias component, are also identified.

With this methodology, the partitioning of expected loss into its variance and bias components in the two-step RPD optimization procedure results in the resolution of the mean adjustment process in RPD. A natural extension is to consider the mean and variance separately. Perhaps the resolution of the RPD problem cited above is a precursor to the merits of considering variance and bias (or the mean) separately. The performance measure used in jointly modeling these two components, S/N, does not have a concrete meaning, while the mean and variance responses separately do have clear interpretations that facilitate system analysis. The separate modeling of these two responses is appealing as it provides the predictive capability to assess the performance of any solution with respect to both mean and variance responses.

With the use of these two distinct models, a new strategy based on RPD's sequential two -step procedure using variance as the PerMIA is proposed. Variance is construed as a measure of sensitivity to system instability due to transitory system behavior. In the first step involving the minimization of the PerMIA variance, the solution space is limited to solutions that are least sensitive to nonstationarity. In the next step, adjusting the mean is interpreted as finding the solution within the constrained solution set that results in the best possible mean performance level. This revised strategy is tantamount to finding the most stable solutions, solutions which are least sensitive to the effects of nonstationarity, and thus remain the most "valid" in its presence, and then picking the best mean response solution from that subset.

Given this strategy and the dual response system for the mean and variance responses, we reformulate the optimization problem in the presence of nonstationarity as a constrained optimization problem. The new problem is to optimize (maximize or minimize) the mean response subject to a constraint that the variance be set to its minimum level. In the continuous review inventory system example with nonstationary demand and lead time process, the problem may be reformulated as

$$\underset{\underline{x} \in \mathbf{R}}{\text{minimize}} \; E[y(\underline{x})]$$

$$\text{subject to: } \operatorname{Var}[y(\underline{x})] = \operatorname{Var}[y(\underline{x}^*)]$$

where **R** denotes the current experimental subregion, and $\underline{x}^*$ is the optimal solution to the variance minimization problem

$$\underset{\underline{x} \in \mathbf{R}}{\text{minimize}} \; \operatorname{Var}[y(\underline{x})].$$

Estimates of the optimal solutions to both problems can be solved using standard constrained optimization techniques after models for the mean and variance responses have been constructed. Necessary and sufficient conditions for the equality constrained optimization problem can be found in Luenberger (1984). In irregular cases where these conditions are not satisfied, such as saddle point systems or systems where necessary constraint qualifications do not hold, the optimization methods for dual response systems proposed by Myers and Carter (1973) can be used. These methods are extensions of the ridge analysis approach by Hoerl (1959,1964) and Draper (1963) to the equality-constrained case. As in ridge analysis, these methods involve finding solutions within the current experimental subregion, with each solution derived by adding a constraint limiting the analysis to points on the surface of a hypersphere centered on the design center. By varying the radius of these concentric hyperspheres up to the radius of the current experimental region, solutions which are each optimal relative to their corresponding hyperspheres are identified. In cases where the variance is relatively homogeneous in the current subregion containing the minimum variance response level, ridge analysis would suffice.

The next two sections present two ways of implementing this dual response strategy. Section 4.2 involves the fitting of the variance using the experimental design framework of the Taguchi method. Section 4.3 presents a different variance modeling approach that addresses the identified shortcomings of the first variance modeling approach. Each of these sections contains a subsection illustrating the implementation of the two methods in terms of the nonstationary inventory system example.

4.2 VARIANCE FITTING APPROACH

The first approach to the implementation of the dual response strategy makes use of response surface methods for fitting and analyzing the model for the variance. The overall strategy of searching for and limiting the analysis to the subregion containing the most stable or least sensitive solutions is in harmony with RSM's philosophy of sequentially seeking the best subregion of operability relative to a particular response. The use of response variance as the performance response of interest coincides well with the use of RSM techniques; it can be treated much in the same manner as the usual mean response. The applicability of RSM is justified in part by the presence of variance heterogeneity that is connected to location in the solution space. In the inventory model, for example, variance could be a function of the reorder point and reorder quantity. Variance models have been considered in an RSM context by Bechhofer (1960), Box and Meyer (1986), and Nair and Pregibon (1988).

A major difference between mean modeling and variance modeling procedures is in data requirements. Variance models generally require the collection of independent replications at each design setting to obtain an estimate of the variance through some randomization scheme. In the approach proposed here, we merge the concepts of sensitivity and variance by using the sensitivity sampling method of RPD to generate the replications needed for estimating variance. An experimental design on the noise variables is cross-classified with a design on the signal variables in the manner prescribed by Taguchi thus producing several observations at each signal variable setting. Such observations were originally meant to characterize the design setting's sensitivity to variations in noise settings. Vining and Myers (1990) extend this concept, suggesting that the outer design array on the noise variables is meant not just to characterize design sensitivity, but also to characterize process variance at each design point in the signal variables as well. Myers, Khuri, and Vining (1992) further note that subjecting the system to the noise settings specified in the outer array makes the system as noisy as possible, translating to an extreme case of variability in system performance. They further suggest that these observations can be used to compute the sample variances for each of the design settings in the signal variables, in lieu of randomly generated replicates.

The initial steps in this approach are similar to those of the S/N modeling strategy presented in Chapter 3. Starting at an arbitrary experimental subregion, a first order model relating variance to the signal input variables is postulated. A first order design on the signal variables is cross-classified with a design on the noise variables to form the same inner-outer array structure of RPD. The design considerations for these arrays are equivalent to those discussed in Section 3.5 for S/N modeling. After the experimental runs are undertaken, the sample variance at each of the signal design points is calculated. Before fitting the regression coefficients, a variance stabilizing transformation is applied to the responses. Bartlett and Kendall (1946) recommend the use of the logarithm transformation (the data transformation utilized in calculating the S/N statistic values). For this purpose, Vining and Myers (1990) propose using the square root transformation, thereby having the sample standard deviation as the response on interest. Here we use the logarithm transformation for its added capability of inducing approximate normality for the resultant data (see Kendall and Stuart (1966). With the transformed data assumed to have approximately homogeneous (i.e., stabilized) variance and normal properties, a first order model is fitted using ordinary least squares (OLS). The estimated regression coefficients are then utilized in a steepest-descent search for a better subregion with respect to variance. Once the subregion containing the minimum variance solution is identified or approximated, the corresponding minimum variance level is estimated using second order RSM analysis procedures when applicable. Using the same experimental data for the current subregion, the sample means of the responses are used to fit a second order model for the mean response. Ordinary least squares is again used to estimate the regression coefficients if the variance is relatively constant within the subregion. Otherwise, estimated weighted least squares (EWLS) could be used with the weights being estimated through the use of the fitted variance (or its transformed version) model. This fitted mean model serves as the objective function to be optimized subject to the constraint of setting the second order fitted model on the variance equal to the minimum variance level.

4.2.1 NUMERICAL EXAMPLE

Consider the continuous review inventory model with transient demand and lead time processes. The objective is to minimize monthly cost subject to the monthly cost variance set to its minimum value. Implementation issues are discussed during this process, and numerical results are presented.

We begin by postulating a first order model

$$g(\mathrm{var}(y)) = \gamma_0 + \sum_{i=1}^{k} \gamma_i x_i + \varepsilon \tag{4.1}$$

for some transformed version of the variance response, g(var(y)). Recall that the uncontrollable input factors, reorder point (ξ_1) and reorder quantity (ξ_2), are classified as signal or design variables. The uncontrollable model parameters, the means of the distributions of the internal stochastic process generators, are classified as noise variables due to their variance inflating effect on the system response of interest, monthly total cost. These noise variables are mean interarrival time (ζ_1), mean demand (ζ_2), and mean delivery lead time (ζ_3).

An arbitrary point in the solution space or operability region of the signal variables is selected to serve as the center of the initial experimental subregion. We chose the same initial design center used in illustrating the strategies considered in the previous chapters, namely

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1000 \\ 1000 \end{bmatrix}.$$

The initial experimental subregion centered on the above location is

$$\{(\xi_1, \xi_2) : \xi_1 \in [500,1500],\ \xi_2 \in [500,1500]\},$$

the same initial subregion used in the previous illustrations. The transformation of (2.11) redefines the signal variables in terms of their coded version, namely

$$x_i = 2\left(\frac{\xi_i - \overline{\xi_i}}{d_i}\right), \quad \text{with} \quad d_i = \xi_{i,\text{HIGH}} - \xi_{i,\text{LOW}} \quad \text{and} \quad \overline{\xi_i} = \left(\frac{\xi_{i,\text{HIGH}} + \xi_{i,\text{LOW}}}{2}\right).$$

The coding convention for the noise variables is

$$z_i = \frac{\zeta_i - \mu_{\zeta_i}}{\sigma_{\zeta_i}}$$

where the high and low settings for the noise variables are specified by

$$\zeta_i = \mu_{\zeta_i} \pm k \cdot \sigma_{\zeta_i}$$

with k = 1. Given the above coding scheme, we define our inner and outer array designs, **D** and **Z**, as follows. A 2^2 factorial design augmented by 2 center runs specifies the inner array design. For the **Z** matrix, we use a 2^{3-1} fractional factorial design. The Kronecker product of these two design matrices, $\mathbf{D} \otimes \mathbf{Z}$, gives the overall design matrix which has 24 experimental design points.

Recall that in modeling variance, our statistic is

$$\log\left(s_i^{\,2}\right) = \log\left[\frac{\sum_{j=1}^{4}(y_{ij} - \bar{y}_i)^2}{3}\right] \quad , \quad i = 1, \ldots, 6, \tag{4.2}$$

where the subscript i corresponds to the *ith* combination in the signal variables in the inner array, and the j subscript relates to the *jth* setting of the noise variables in the outer array. The log transformation of the sample variance is used as the response for variance stabilization in accordance with the Bartlett and Kendall (1946) and to induce approximate normality. Experimental results for the initial subregion are presented in the following table. We include the s_i for informational purposes.

Table 4.1 Phase 1 Data - First Experimental Subregion

ξ_1	ξ_2	y_{i1}	y_{i2}	y_{i3}	y_{i4}	$\bar{y}_i$	s_{y_i}	$\log(s^2_{y_i})$
500	500	61432	136413	112643	80041	97632.2	33412.5	9.04782
500	1500	55388	129046	106865	74299	91399.5	32892.2	9.03419
1500	500	51213	125603	103620	71191	87906.7	33134.9	9.04057
1500	1500	41532	112193	93394	61808	77231.7	31601.0	8.99940
1000	1000	50505	124296	102671	70277	86937.2	32904.4	9.03451
1000	1000	50455	123579	102610	70655	86824.7	32579.5	9.02589

Table 4.2 provides summary statistics for the fitted first-order model.

Table 4.2 Fitted First-Order Model - First Subregion

```
The regression equation is
log(s²) = 9.03 - 0.0105 x1 - 0.0137 x2

Predictor        Coef        Stdev     t-ratio        p
Constant      9.03040      0.00355     2543.88    0.000
x1          -0.010508     0.004348       -2.42    0.094
x2          -0.013700     0.004348       -3.15    0.051

s = 0.008695      r² = 84.0%

Analysis of Variance

SOURCE        DF          SS           MS          F        p
Regression     2   0.00119245   0.00059623      7.89    0.064
Error          3   0.00022683   0.00007561
Total          5   0.00141928
```

From the above regression analysis, the steepest descent search direction is

$$-\underline{b}_1 = \begin{bmatrix} 0.0105 \\ 0.0137 \end{bmatrix} \Rightarrow -\underline{b}_{1s} = \begin{bmatrix} 0.6083 \\ 0.7937 \end{bmatrix}$$

where $\underline{b}_{1s}$ is the standardized gradient vector. The results for the exploratory runs for the above search direction are listed in the next table.

Table 4.3 Exploratory Runs Data for First Search Direction

r	x_1	x_2	ξ_1	ξ_1	s_{y_i}	$\log(s^2_{y_i})$
1	0.6083	0.7937	1304.2	1396.8	33649.6	9.05396
2	1.2166	1.5874	1608.3	1793.7	33459.5	9.04904
4	2.4332	3.1748	2216.6	2587.4	32992.5	9.03683
8	4.8665	6.3496	3433.2	4174.8	31394.2	8.99370
16	9.7330	12.6992	5866.5	7349.6	21910.6	8.68131
32	19.4659	25.3984	10733.0	13699.2	2900.1	6.92483

The most promising region is centered approximately on the point

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 10700 \\ 13700 \end{bmatrix}.$$

The first order analysis phase then restarts with a new experimental subregion being defined about the new location. This sequential procedure is repeated until the first order coefficients become insignificant (at which point an improving direction is not available), or when the quality of the fit is poor. In the case where the first order coefficients vanish, the second-order phase of the analysis commences. If the quality of fit is poor, one alternative is to reduce the size of the experimental subregion, and restart the sequential procedure. The other alternative is to proceed to the second-order analysis phase.

This procedure identified two subsequent subregions whose exploration were straightforward. The experimental and model-fitting results and the exploratory search data are provided in the following five tables.

Table 4.4 Phase 1 Data - Second Experimental Subregion

ξ_1	ξ_2	y_{i1}	y_{i2}	y_{i3}	y_{i4}	$\bar{y}_i$	s_{y_i}	$\log(s_{y_i}^2)$
10200	13200	24728.2	21327.1	20774.3	19258.4	21522.0	2309.46	6.72702
10200	14200	25692.6	22157.3	21583.4	20207.6	22410.2	2336.18	6.73701
11200	13200	26712.1	19990.0	18771.1	21101.5	21643.7	3510.41	7.09072
11200	14200	27694.8	20849.7	19629.1	22185.8	22589.8	3559.86	7.10286
10700	13700	26249.7	20699.4	19843.1	20713.4	21876.4	2943.79	6.93781
10700	13700	26156.2	20796.5	19888.2	20708.2	21887.3	2875.22	6.91734

Table 4.5 Second Fitted First-Order Model

```
The regression equation is
log(s²) = 6.92 + 0.182 x1 + 0.00554 x2

Predictor        Coef        Stdev     t-ratio        p
Constant      6.91879      0.00496     1395.98    0.000
x1           0.182386     0.006070       30.05    0.000
x2           0.005536     0.006070        0.91    0.429

s = 0.01214      R-sq = 99.7%

Analysis of Variance

SOURCE         DF          SS           MS          F        p
Regression      2    0.133181     0.066591     451.81    0.000
Error           3    0.000442     0.000147
Total           5    0.133624
```

Table 4.6 Exploratory Runs Data for Second Search Direction

r	x_1	x_2	ξ_1	ξ_2	s_{y_i}	$\log(s_{y_i}^2)$
1	-1	0	10200	13700	2266.1	6.71054
2	-2	0	9700	13700	2268.6	6.71152
4	-4	0	8700	13700	4608.4	7.32710
8	-8	0	6700	13700	11043.5	8.08621

Table 4.7 Phase 1 Data - Third Experimental Subregion

ξ_1	ξ_2	y_{i1}	y_{i2}	y_{i3}	y_{i4}	$\bar{y}_i$	s_{y_i}	$\log(s^2_{y_i})$
9700	13200	23746.5	22915.3	22415.4	18420.8	21874.5	2367.01	6.74840
9700	14200	24678.8	23272.5	22800.1	19339.4	22522.7	2267.24	6.71099
10700	13200	25691.9	20341.4	19516.1	20117.9	21416.9	2871.27	6.91615
10700	14200	26687.5	21132.6	20116.9	21195.1	22283.0	2977.62	6.94774
10200	13700	25262.7	21522.6	21099.9	19749.5	21908.7	2360.39	6.74597
10200	13700	25153.0	21898.9	21063.9	19860.6	21994.1	2266.06	6.71054

Table 4.8 Third Fitted First-Order S/N Model

```
The regression equation is
log(s²) = 6.80 + 0.101 x1 - 0.0015 x2

Predictor        Coef       Stdev    t-ratio        p
Constant      6.79663     0.02967     229.09    0.000
x1            0.10112     0.03634       2.78    0.069
x2           -0.00145     0.03634      -0.04    0.971

s = 0.07267      R-sq = 72.1%

Analysis of Variance

SOURCE        DF          SS          MS          F        p
Regression     2    0.040912    0.020456       3.87    0.147
Error          3    0.015844    0.005281
Total          5    0.056755
```

The analysis for the last subregion indicates that the quality of fit has deteriorated as shown by r^2=0.721 and p-value=0.147 for the F-test on the regression. The results also indicate that the slope coefficients are marginal to insignificant with p-values of 0.069 and 0.971. At this point it was surmised that the stationary point was located within this region which would explain the insignificance of the slope coefficients, and that the inadequacy in model fit was due to the presence of curvature in the response surface. A second-order model of the logarithm of the sample variance response was therefore entertained at this point.

In the second order analysis phase, the inner array utilized was a CCD with the axial points drawn in towards the design center half the distance of the design radius in the coded variable space (i.e., $\alpha = .707$) for bias protection. The data from the previous trials in the last iteration of the first order phase were augmented by additional data from the axial runs and are summarized in the following table.

Table 4.9 Phase 2 Data - Third Experimental Subregion

ξ_1	ξ_2	y_{i1}	y_{i2}	y_{i3}	y_{i4}	$\bar{y}_i$	s_{y_i}	$\log\left(s_{y_i}^2\right)$
9700	13200	23746.5	22915.3	22415.4	18420.8	21874.5	2367.01	6.74840
9700	14200	24678.8	23272.5	22800.1	19339.4	22522.7	2267.24	6.71099
10700	13200	25691.9	20341.4	19516.1	20117.9	21416.9	2871.27	6.91615
10700	14200	26687.5	21132.6	20116.9	21195.1	22283.0	2977.62	6.94774
10200	13700	25262.7	21522.6	21099.9	19749.5	21908.7	2360.39	6.74597
10200	13700	25153.0	21898.9	21063.9	19860.6	21994.1	2266.06	6.71054
10554	13700	25901.8	20874.7	20124.4	20405.7	21826.6	2734.35	6.87371
9846	13700	24450.2	22804.5	22610.2	19263.4	22282.1	2175.14	6.67497
10200	14054	25597.9	21753.4	21322.0	20177.9	22212.8	2352.61	6.74310
10200	13346	24798.9	21301.6	20834.7	19412.7	21586.9	2287.02	6.71854

A second order regression model was fitted using the SAS procedure RSREG. A partial SAS output of the results is given in the following table.

Table 4.10 Second Order Model Analysis Results

Coding Coefficients for the Independent Variables

Factor	Subtracted off	Divided by
X1	10200	500.000000
X2	13700	500.000000

Response Surface for Variable LOGVARY

Response Mean	6.779011
Root MSE	0.023173
R-Square	0.9742
Coef. of Variation	0.3418

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	0.059447	0.7131	55.351	0.0012
Quadratic	2	0.020574	0.2468	19.156	0.0089
Crossproduct	1	0.001190	0.0143	2.216	0.2108
Total Regress	5	0.081211	0.9742	30.246	0.0028

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	0.002148	0.000537

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|	Parameter Estimate from Coded Data
INTERCEPT	1	59.860274	14.361310	4.168	0.0141	6.727308
X1	1	-0.008486	0.002115	-4.013	0.0160	0.108986
X2	1	-0.001605	0.002750	-0.584	0.5908	0.002313
X1*X1	1	0.000000380	9.8879118E-8	3.847	0.0184	0.095087
X2*X1	1	6.9E-8	4.634662E-8	1.489	0.2108	0.017250
X2*X2	1	3.3066076E-8	9.8879118E-8	0.334	0.7549	0.008267

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	0.068556	0.022852	42.555	0.0017
X2	3	0.001277	0.000426	0.793	0.5582

Canonical Analysis of Response Surface
(based on coded data)

Factor	Critical Value Coded	Critical Value Uncoded
X1	-0.618981	9890.509543
X2	0.505950	13953

Predicted value at stationary point 6.694163

Eigenvalues	Eigenvectors X1	Eigenvectors X2
0.095936	0.995195	0.097909
0.007418	-0.097909	0.995195

Stationary point is a minimum.

The above results show that the second order model is quite adequate. They also indicate the presence of an estimated stationary point interior to the subregion corresponding to a minimum response solution for the logarithm of sample variance response. This solution corresponds to a reorder point of 9891 units and a reoder quantity of 13,953 units, with an estimated minimum logarithm of variance response of approximately 6.6942 (or a predicted standard deviation of $2224).

Attention is now directed to the mean response and the estimation of a second-order model based on the experimental runs previously conducted in this region. Results based on ordinary least squares estimation in the space of the coded variables are presented in Table 4.11.

Table 4.11 Mean Response Second Order Model Analysis Results

Response Surface for Variable Y

Response Mean	21991
Root MSE	78.272666
R-Square	0.9765
Coef. of Variation	0.3559

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	973691	0.9343	79.464	0.0006
Quadratic	2	32106	0.0308	2.620	0.1874
Crossproduct	1	11870	0.0114	1.937	0.2363
Total Regress	5	1017667	0.9765	33.221	0.0024

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	4	24506	6126.610258

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEPT	1	21957	38.570842	569.3	0.0000
X1	1	-203.880014	35.005658	-5.824	0.0043
X2	1	391.385900	35.005658	11.181	0.0004
X1*X1	1	188.844898	83.698667	2.256	0.0870
X2*X1	1	54.475000	39.136333	1.392	0.2363
X2*X2	1	-120.248448	83.698667	-1.437	0.2242

Factor	Degrees of Freedom	Sum of Squares	Mean Square	F-Ratio	Prob > F
X1	3	250881	83627	13.650	0.0144
X2	3	790384	263461	43.003	0.0017

The analysis shows a excellent fit for the estimated mean response model. Since the analysis results for the variance indicate a single minimum variance point, the constrained optimization problem on the mean response suggests that an estimate for the optimal mean response be evaluated at that point using the fitted mean response model. This point yields an estimated mean monthly total cost value of approximately $22,352.

4.2.2 CONCLUDING REMARKS

Consider first the manner in which the independent replications at each design point in the signal variable space are generated. For the purpose of estimating the true population variance at a specified signal location, the conventional procedure is to use random sampling methods. The sampling scheme prescribed in RPD of using the outer array design on the noise produces replications that do not constitute a random sample. Rather, it is a controlled sampling technique taking advantage of the experimenter's ability to control the noise variables in the experimental or prototype stage of process or product development. Its purpose is to subject each design combination to carefully specified environmental operating conditions defined by the design settings on the noise variables. These noise combinations in the outer array have traditionally been viewed as a means for systematically exploring the space of the noise variables, not for randomly sampling from it. In this regard, Shoemaker and Tsui (in Nair (1992)) caution that sample variances computed in such a manner do not constitute good variance estimates, only rough measures of sensitivity to the noise factors or, in other words, pseudo-variance estimates. In rebuttal, Myers and Vining (in Nair (1992)) state that the outer array on the noise does put considerable emphasis on randomization. It does so by preventing any possibility of "slippage" or bias tendencies in the sampling of the noise space through carefully chosen central location and separation considerations for the noise factor levels (i.e., setting the high and low level noise settings to $\mu \pm 1\sigma$), making the system operating environment as noisy as possible.

Another concern concerning the legitimacy of the sample standard deviation computed from the crossed-array experimental designs as an estimator of the population standard deviation is the number of replications used in its calculation. In the proposed modification of RPD considered in this chapter, it was recommended that the outer array be fractionated more heavily than the inner array to reduce the potential for an excessive number of experimental runs in the cross-classification of design arrays, . In the previous example, as in the example of Chapter 3, the outer array was reduced to having four settings, corresponding to four replications per signal design point.

Deaton, Reynolds, and Myers (1983) provide a guideline for the use of the sample variances as weights in estimated weighted least squares. In that article, the authors note that unless the number of replications is greater than approximately nine, then the sample variances should not be utilized as weights. It is therefore argued that if the four replications used in the previous experiments could not have provided an adequate variance-based weight for estimated least squares application, they likewise could not have provided an adequate variance response estimate.

4.3 VARIANCE APPROXIMATION APPROACH

In this methodology, a composite model describing the response as a function of both the signal and noise variables is postulated and estimated using RSM. From this composite model, approximating low order polynomial functions for the mean and variance responses are derived by applying the expectation and variance operators on the fitted model. This composite model strategy addresses the problems of the variance model fitting approach associated with the use of crossed-arrays designs. The dual response optimization process exploits the inherent qualities of the variance approximation function to devise a search procedure distinct from that of the approach presented previously in this chapter. The next section provides the motivation behind this strategy and is based mainly on the work of Myers, Khuri, and Vining (1992).

4.3.1 MOTIVATION

Key to the composite model approach for analyzing the dual response system of the mean and variance is understanding the mechanisms behind the generation of the variance response. The previous approach's experimental tactic of cross-classifying the design arrays for the signal and noise variables hints at the importance of interactions between signal and noise factors. The designed crossing pattern is a means of emphasizing these interactions, and it is surmised that these interactions play a vital role in the variance and S/N analyses. The analysis framework to be developed is based on this hypothesis, wherein the presence of these interactions is ascertained, and their effect on variance is exploited.

In Chapter 2, it was shown that shifting from a what was referred to as a static model (where the noise variables are held constant) to a nonstationary one results in an increased variance for the response. We generalized these results through the following model relating the response to the inputs

$$y = E(y|\underline{x},\underline{z}) + \varepsilon. \tag{4.3}$$

In this model, the expected value of the response y is dependent on both input variable types, i.e., signal and noise variables. The term ε represents the random model error. Clearly, if the noise variables are held constant or treated as controllable variables, the variance of the response would be equivalent to the variance of the error term. In such a situation, the variance of the response is assumed homogeneous in the space of the signal variables. If the noise variables, which in the actual process are uncontrollable, behave in their usual random manner, then the variance of the response would depend not just on the variance of the error term, but on the variance of the conditional mean of the response as well. Assuming the error term is independent of the conditional mean, we have

$$\mathrm{Var}(y) = \mathrm{Var}[E(y|\underline{x},\underline{z})] + \mathrm{Var}(\varepsilon). \tag{4.4}$$

Consider $E(y|\underline{x},\underline{z})=f(\underline{x},\underline{z})$; i.e., let the conditional mean of y be some function f of the inputs $\underline{x}$ and $\underline{z}$. In an adequately restricted subregion of the joint design space for both inputs, this function can be approximated by a low order polynomial approximation based on a truncated Taylor series expansion. Let f be a first order polynomial such that

$$f(\underline{x},\underline{z}) = \underline{x}'\underline{\beta} + \underline{z}'\underline{\gamma}. \tag{4.5}$$

where $\underline{\beta}$ and $\underline{\gamma}$ are the regression coefficients of the coded signal and noise variables. In this additive model for the signal and noise variables, the variance of f is

$$\mathrm{Var}[f(\underline{x},\underline{z})] = \underline{\gamma}' \, \mathrm{Var}(\underline{z}) \underline{\gamma} \tag{4.6}$$

which is a function of the variance-covariance matrix of the noise variables. In this scenario, the variance of f and, likewise, the variance of y is not related to the signal variables. Thus in this setting the experimenter or analyst has no control over the source of heterogeneity in the variance of the response. If f, however, is of the form given by

$$f(\underline{x},\underline{z}) = \underline{x}' \underline{\beta} + \underline{z}' \underline{\gamma} + \underline{x}' \Lambda \underline{z} \tag{4.7}$$

where $\Lambda=[\lambda_{ij}]$ represents the matrix of coefficients of the interactions between the signal and noise variables, $x_i z_j$, for $i=1,\dots,p$, and $j=1,\dots,k$, then the variance of f is

$$\mathrm{Var}[f(\underline{x},\underline{z})] = \underline{\gamma}' \, \mathrm{Var}(\underline{z}) \underline{\gamma} + (\underline{x}' \Lambda) \mathrm{Var}(\underline{z}) (\Lambda' \, \underline{x}) \tag{4.8}$$

which is a function of the variance of the noise variables and the settings of the signal variables. Since the experimenter fixes the settings or levels for the signal variables, he is able to exert some influence on the variance of the response y. In this scenario, though the variance of y is heterogeneous, it is controllable and can be minimized (up to the extent encompassed by the variance of $E(y|\underline{x},\underline{z})$) though the appropriate choice of the signal variable settings. That is, we take advantage of the presence of signal by noise factor interactions.

These issues regarding the variance are indeed consistent with the variance-related concepts associated with the Taguchi method. Given variance as the performance measure independent of adjustment (PerMIA), the signal factors $\underline{x}_d$ that influence the variance are exactly those signal factors that interact with the noise variable. That is, their effect on the variance is a byproduct of their interaction with the noise variables and, therefore, their inclusion in the function for the variance. In other words, if $\lambda_{ij} \neq 0$, then $x_i z_j$ is a significant interaction and $x_i \in \underline{x}_d$, the set of dispersion variables.

These results can also be reconciled with the models of Leon, Shoemaker, and Kackar (1987). Given their multiplicative error model

$$y = \mu(\underline{x}_d, \underline{x}_a) \cdot \varepsilon(\underline{x}_d, \underline{z}),$$

it would seem that all the signal variables interact with the noise variables once the multiplication has been carried out. In this case all signal variables are actually dispersion variables. For their additive error model case,

$$y = \mu(\underline{x}_d, \underline{x}_a) + \varepsilon(\underline{x}_d, \underline{z}),$$

the presence of signal by noise interactions depends on the functional form of the error function $\varepsilon(\underline{x}_d, \underline{z})$. Thus variance can only be conditionally controlled or minimized.

The above results suggest a different approach to the modeling of variance. The first step in the analysis is to postulate a linear model for the function f containing main effects for the signal and noise variables, as well as interactions between signal and noise. We replace this model for f into equation (4.3) to get

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{z}'\underline{\gamma} + \underline{x}'\Lambda\underline{z} + \varepsilon. \tag{4.9}$$

This model will be referred to as the composite model. Using this postulated composite model, an appropriate experimental design is selected for fitting this model. The model statement calls for a design that combines both signal and noise variables into a single design array.

At this point it is noted that one of the original motivating factors for this approach is the reduction in the number of experimental runs if a combined design for the noise and signal factors is utilized. Welch, Yu, Kang, and Sacks (1990), Shoemaker, Tsui, and Wu (1991), and Easterling (1985) suggest this combined array format accommodating both types of variables in the direct modeling of the response as a function of the control and noise variables. This approach is only possible if both factor types are included in the model, treating the noise variables as if they were controllable inputs like the signal variables. This is accommodated in a robust parameter design-type analysis where the experimenter is able to control the noise factors during the experiment or prototype stage of process or product development. We exploit the fact that in simulation every aspect of the experiments can be controlled by the analyst.

Once the composite model of (4.9) is fitted, the variance operator is applied to the postulated model whose coefficients are replaced by the estimated coefficients of the fitted composite model to form the variance approximation function. By applying the expectation operator to the same model, a function for the mean can also be derived. The following subsections present a dual response constrained optimization algorithm that utilizes the composite model formulation. Practical implementation considerations are discussed and the numerical results from its application to the illustrative inventory problem are presented.

4.3.2 ALGORITHM STEPS AND IMPLEMENTATION ISSUES

This subsection discusses the steps of the second approach to dual response optimization. As each step is presented, practical issues concerning its implementation are examined. The nonstationary inventory system example is then used to illustrate the implementation of the algorithm.

The first step is to define the composite model. We begin by postulating a first order model for the response involving main effects for the noise and signal variables, as well as noise by signal interactions terms. From equation (4.9),

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{z}'\underline{\gamma} + \underline{x}'\Lambda\underline{z} + \varepsilon,$$

where

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad \underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \underline{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix}, \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix},$$

and Λ is the matrix of signal by noise interaction coefficients,

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pk} \end{bmatrix}.$$

In matrix notation, the general linear model is

$$\underline{y} = C\underline{\beta}^* + \underline{\varepsilon}, \tag{4.10}$$

with

$$C = \left[\begin{array}{c|ccc|ccc|ccc} 1 & x_{11} & \cdots & x_{1p} & z_{11} & \cdots & z_{1k} & x_{11}z_{11} & \cdots & x_{1p}z_{1k} \\ 1 & x_{21} & \cdots & x_{2p} & z_{21} & \cdots & z_{2k} & x_{21}z_{21} & \cdots & x_{2p}z_{2k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Np} & z_{N1} & \cdots & z_{Nk} & x_{N1}z_{N1} & \cdots & x_{Np}z_{Nk} \end{array}\right]$$

and

$$\underline{\beta}^{*\prime} = [\beta_o,\ \beta_1 \cdots \beta_p,\ \gamma_1 \cdots \gamma_k,\ \lambda_{11} \cdots \lambda_{pk}]\ .$$

The reason for choosing a first order model plus interactions terms will become apparent shortly. As for the error term, we assume that

$$E(\underline{\varepsilon}) = \underline{0} \quad \text{and} \quad Var(\underline{\varepsilon}) = \sigma^2 I_N.$$

That is, with the noise variables at fixed levels during the experiments, a potential source of variance heterogeneity has been eliminated, thereby supporting the assumption of homogeneous variance. Nair (1992) cautions that this is possible only if all the noise factors present in the system have been accounted for and treated as design or controllable factors. This condition can be frequently met in the simulation environment since the modeler specifies all facets of the simulation model.

With this constant variance scenario, the coefficients $\underline{\beta}^*$ are estimated by ordinary least squares (OLS), and the OLS estimator is

$$\underline{b}^* = (C'C)^{-1}C'\underline{y} = [\hat{\beta}_o,\ \hat{\beta}_1 \cdots \hat{\beta}_p,\ \hat{\gamma}_1 \cdots \hat{\gamma}_k,\ \hat{\lambda}_{11} \cdots \hat{\lambda}_{pk}]'$$

with variance-covariance matrix

$$\mathrm{Var}(\underline{b}^*) = \sigma^2 (C'C)^{-1}. \tag{4.11}$$

The fitted model is

$$\hat{y} = \hat{\beta}_0 + \underline{x}'\hat{\underline{\beta}} + \underline{z}'\hat{\underline{\gamma}} + \underline{x}'\hat{\Lambda}\underline{z}. \tag{4.12}$$

The suggested designs for fitting the composite model are the usual first order RSM designs like the 2^{p+k} factorial design, or, if the magnitude of p+k leads to an excessive design in terms of experimental effort, fractions of this design may be utilized. For the fractional factorial designs, the defining interactions must be carefully chosen to allow the independent estimation of all relevant model terms (i.e., x_i's, z_j's, and x_iz_j's). This is accomplished by restricting defining interactions that involve just the signal variables to be of at least length 3 (or Resolution III); those defining interactions that involve both signal and noise to be of at least Resolution V; and those involving just the noise to be of at least Resolution III. If the modeler decides to include interactions between signal variables in the model, then the defining interactions consisting of just the signal variables should have at least Resolution V. This two-level design is augmented by at least two center runs (the reason for this will become evident shortly) to produce the design matrix **D**. For example, if p=2 and k=3, then **D** would follow the format

$$\begin{array}{cc} & \begin{array}{ccccc} x_1 & x_2 & z_1 & z_2 & z_3 \end{array} \\ \mathbf{D} = & \left[\begin{array}{ccccc} \pm\underline{1} & \pm\underline{1} & \pm\underline{1} & \pm\underline{1} & \pm\underline{1} \\ \hdashline \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \end{array}\right] \begin{array}{l} \leftarrow \text{Factorial or Fractional Factorial} \\ \leftarrow \text{Center Runs} \end{array} \end{array}$$

The +1 and -1 levels for the coded variables correspond to high and low levels for the natural variables for both signal and noise. The variable transformations from the natural variables to their coded versions were discussed in the previous chapters and are given by

$$x_i = 2\left(\frac{\xi_i - \overline{\xi}_i}{d_i}\right), \quad \text{with} \quad d_i = \xi_{i,\text{HIGH}} - \xi_{i,\text{LOW}} \quad \text{and} \quad \overline{\xi}_i = \left(\frac{\xi_{i,\text{HIGH}} + \xi_{i,\text{LOW}}}{2}\right).$$

The high and low settings for the natural signal variables are functions of the current experimental subregion and correspond to the limits of the subregion in the signal variable space. For the noise variables, the relevant coding transformation is

$$z_i = \frac{\zeta_i - \mu_{\zeta_i}}{\sigma_{\zeta_i}},$$

where the high and low settings for the noise variables are defined as

$$\zeta_i = \mu_{\zeta_i} \pm k\sigma_{\zeta_i}.$$

Note that the noise variable levels utilized do not change in either the natural or coded spaces. Only the levels for the signal variables in their natural space change as their levels depend on the experimental subregions considered within the sequential search for a better operating subregion.

The initial set of experimental runs is undertaken after a choice is made for the initial experimental subregion size and location. Without a priori information, the choice for initial location is made heuristically. A consideration for the initial region size is that it should be sufficiently small so that the true response function can be adequately approximated by the first order graduating function. If the quality of fit for the estimated function is not adequate, the region size is typically reduced. Once an adequate first order model approximation is constructed, the next step is the verification of the applicability of this approach.

The initial utility of the fitted composite model is to ascertain the applicability of this approach when it comes to variance modeling. The feasibility of this approach hinges on the presence of significant signal by noise interactions. It is essential that the presence or absence of these interactions be verified. This verification is accomplished in the following manner. Assume that the fitted first-order fitted model adequately represents the response over some subregion in the signal variable space. If there exists at least one significant noise by signal interaction coefficient, then this variance modeling scheme is applicable. Otherwise, we try to expose the presence of any of these interactions by increasing the experimental subregion size. This strategy is justified by the fact that interactions between variables, if they do exist, typically become more prominent when the range for any one of the variables is expanded. With signal by noise interactions, this range expansion is possible only with respect to the signal variables. Sequentially adjusting the subregion size while monitoring the quality of fit, the procedure terminates in one of two possibilities - an adequate fitted model with significant noise by signal interaction terms is defined, or the quality of fit deteriorates to an unacceptable level before these interactions are exposed thereby indicating that this variance modeling approach is not applicable or the model is not adequate.

To avoid the early deterioration of model fit, a second order model for the response may be postulated with the following limitation. The second order terms must involve just the signal variables, keeping the noise variables terms at first order. Again the reason for this will become apparent shortly. With this relaxation of the model order restriction, our postulated model becomes

$$y = \beta_0 + \underline{x}' \underline{\beta} + \underline{x}' B \underline{x} + \underline{z}' \underline{\gamma} + \underline{x}' \Lambda \underline{z} + \varepsilon, \qquad (4.13)$$

with B being the matrix of coefficients for the second-order and interaction terms for the signal variables given by

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \cdots & \frac{1}{2}\beta_{1p} \\ & \beta_{22} & \cdots & \frac{1}{2}\beta_{2p} \\ & & \ddots & \vdots \\ \text{sym} & & & \beta_{pp} \end{bmatrix}.$$

The increased order in the signal variables is accommodated by augmenting the design matrix **D** with axial runs along the axes of the signal variables. For example, if p=2 and k=3, then the design matrix **D** follows the following format.

$$\begin{array}{cc} & \begin{array}{ccccc} x_1 & x_2 & z_1 & z_2 & z_3 \end{array} \\ \mathbf{D} = & \left[\begin{array}{ccccc} \underline{\pm 1} & \underline{\pm 1} & \underline{\pm 1} & \underline{\pm 1} & \underline{\pm 1} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \alpha & 0 & 0 & 0 & 0 \\ -\alpha & 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 \\ 0 & -\alpha & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{l} \leftarrow \text{Factorial or Fractional Factorial} \\ \leftarrow \text{Center Runs} \\ \leftarrow \text{Axial Runs} \\ \downarrow \\ \downarrow \\ \lrcorner \end{array}$$

The verification or feasibility step affords the experimenter the unique advantage of being able to classify the heterogeneous problem as either being a controllable or uncontrollable situation. If signal by noise interactions do not exist, then the application of any method of solution for the dual response problem will be futile. The heterogeneity in variance is caused by the noise variables which are uncontrollable in the actual system or process. If these interactions do exist, then variance heterogeneity is a function of both the noise variables and those signal variables that play a role in the significant noise by signal interactions. Thus the variance can be controlled and, consequently, minimized via the signal variables.

Assuming the presence of significant signal by noise interactions, the next step is the formulation of the variance approximation function. At this point of the analysis, we shift our perspective from treating the noise variables as controllable variables to treating them as uncontrollable, random factors. Given the model of (4.9), the application of the variance operator results in

$$\mathrm{Var}(y) = \underline{\gamma}' \,\mathrm{Var}(\underline{z})\underline{\gamma} + (\underline{x}' \Lambda)\mathrm{Var}(\underline{z})(\Lambda' \underline{x}) + \sigma^2. \tag{4.14}$$

Since all the terms on the right hand side of the equation are constants except for the signal variables, this expression shows variance as a quadratic form in the signal variables. This observation is precisely the reason why the composite model was limited to first order terms when noise variables and, specially, noise by signal variables are concerned. The variance operator doubles the exponent of the nonrandom multipliers of the noise variables. For example, if a quadratic signal term by noise variable interaction is present in the composite model, the order of the resulting variance function would be at least four in terms of the signal variables. It is obvious that with higher-order models the complexity of the problem increases significantly. For practical considerations, the order of the variance approximating function is limited to two.

In the above model, the model coefficients are generally unknown and estimates are used in their place. We substitute the OLS estimates from the fitted composite model for the true model coefficients. For the variance of the error term, we utilize the sample variance, s^2, computed from the center runs. Thus, the need to estimate s^2 provides the motivation for including replicated runs at the center of the design. With these substitutions the estimated variance function is

$$\hat{\mathrm{V}}\mathrm{ar}(y) = \underline{\hat{\gamma}}' \,\mathrm{Var}(\underline{z})\underline{\hat{\gamma}} + (\underline{x}' \hat{\Lambda})\mathrm{Var}(\underline{z})(\hat{\Lambda}' \underline{x}) + s^2. \tag{4.15}$$

Consider now the properties of the variance approximation function and its estimate. The variance function of (4.14) can be rewritten as

$$\mathrm{Var}(y) = [\underline{\gamma} + \Lambda' \underline{x}]' \mathrm{Var}(\underline{z})[\underline{\gamma} + \Lambda' \underline{x}] + \sigma^2. \tag{4.16}$$

Letting $\underline{w} = \underline{\gamma} + \Lambda' \underline{x}$,

$$\mathrm{Var}(y) = \underline{w}' \mathrm{Var}(\underline{z})\underline{w} + \sigma^2 \tag{4.17}$$

which is the sum of a positive definite quadratic form and a nonnegative constant. This result is due to the positive definiteness of the variance-covariance matrix Var($\underline{z}$). An important result of this property of Var($\underline{z}$) is that the variance function is strictly convex. Hence, its has a unique minimum solution in the subspace of the relevant signal variables, those signal variables that interact with the noise variables. Hence, if all signal variables have significant interactions with the noise variables, the minimum variance solution would be a unique point in the space of the signal variables. Otherwise, it would only be unique in the subspace of the relevant signal variables, the dispersion variables.

The absolute minimum value for Var(y) is Var(ε) which is attained when $\underline{w}=\underline{0}$. With the true model, this is possible only when the number of noise variables is less than or equal to the number of signal variables. Otherwise, the resulting system of equations, $\underline{w}=\underline{0}$, would be overdefined for which a solution may not exist Similarly, for the case of the fitted composite model, this absolute minimum variance value is attainable if the number of significant noise variables is less than or equal to the number of significant signal variables.

Another important characteristic of Var($\underline{z}$) is that it is a diagonal matrix since the noise variables are independent of each other. Furthermore, it is equivalent to an identity matrix $\mathbf{I}_p$, where p is the number of noise variables This is due to the coding scheme used for transforming the natural noise variables to their coded versions which yields the following properties for the coded noise variables $\underline{z}$:

$$z_{i,HIGH} = +1 \quad , \quad z_{i,LOW} = -1 \quad , \quad E(z_i) = 0 \quad , \text{ and } \quad Var(z_i) = 1.$$

With these results, the variance function becomes

$$Var(y) = [\underline{\gamma} + \Lambda' \underline{x}]' [\underline{\gamma} + \Lambda' \underline{x}] + \sigma^2. \tag{4.18}$$

Exploiting the fact that (4.18) is a strictly convex function, its minimum point is a unique stationary point which is determined by taking the derivative of Var(y) with respect to $\underline{x}$, setting it equal to zero, and solving for $\underline{x}^*$, the minimum variance solution, using the resulting normal equation. To derive the normal equation, we rewrite Var(y) as

$$Var(y) = \underline{\gamma}' \underline{\gamma} + 2\underline{\gamma}' \Lambda' \underline{x} + \underline{x}' \Lambda\Lambda' \underline{x} + \sigma^2. \tag{4.19}$$

The derivative of Var(y) with respect to $\underline{x}$ is

$$\frac{\partial Var(y)}{\partial \underline{x}} = 2\Lambda \underline{\gamma} + 2\Lambda\Lambda' \underline{x} \tag{4.20}$$

since $\Lambda\Lambda'$ is a symmetric matrix. Setting this derivative equal to zero yields the normal equation

$$\Lambda(\underline{\gamma} + \Lambda' \underline{x}) = 0. \tag{4.21}$$

This is equivalent to taking the partial derivative of Var(y) with respect to each signal variable x_i (for i=1,...,p), setting the partial derivatives to zero, and solving for the stationary point. Since the analysis is usually based on the fitted model, the estimated coefficients are used in lieu of the true model parameters in (4.18) yielding the estimated normal equation

$$\hat{\Lambda}(\hat{\underline{\gamma}} + \hat{\Lambda}'\underline{x}) = 0. \tag{4.22}$$

An alternative method for finding the stationary point involves the partial derivatives

$$\frac{\partial \hat{V}ar(y)}{\partial x_i} = 0 \qquad \text{for } i = 1,\ldots,p. \tag{4.23}$$

As a verification step, the matrix of second partial derivatives

$$H(\underline{x}) = \begin{bmatrix} \frac{\partial^2 \hat{V}ar(y)}{\partial x_1^2} & \frac{\partial^2 \hat{V}ar(y)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 \hat{V}ar(y)}{\partial x_1 x_p} \\ \frac{\partial^2 \hat{V}ar(y)}{\partial x_2 x_1} & \frac{\partial^2 \hat{V}ar(y)}{\partial x_2^2} & \cdots & \frac{\partial^2 \hat{V}ar(y)}{\partial x_2 x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \hat{V}ar(y)}{\partial x_p x_1} & \frac{\partial^2 \hat{V}ar(y)}{\partial x_p x_2} & \cdots & \frac{\partial^2 \hat{V}ar(y)}{\partial x_p^2} \end{bmatrix},$$

also known as the Hessian matrix, may be computed and its definiteness verified. A sufficient condition for the optimality of the stationary point (in variance minimization) is that the Hessian be positive definite at that point.

If the estimated minimum variance point $\underline{x}^*$ is located within the current experimental region, then the corresponding minimum variance level is estimated. If the minimum variance location is outside the current region, a new experimental subregion containing that location (not necessarily centered on it) is defined, and the entire process is repeated. Once an interior point solution is identified, the corresponding minimum variance level is estimated using the variance model of (4.15).

Some comments regarding this estimate are in order. Myers and Kim (1993) point out that the use of equation (4.15) leads to a biased estimate of the variance. To show this result, consider the variance approximation function as a quadratic form in terms of $\underline{\hat{w}} = \underline{\hat{\gamma}} + \hat{\Lambda}'\underline{x}$. Taking its expectation, we have

$$E[\hat{V}ar(y)] = E[\underline{\hat{w}}' \, Var(\underline{z})\underline{\hat{w}} + s^2]. \tag{4.24}$$

With $E(\underline{\hat{w}})=\underline{w}$ and $Var(\underline{\hat{w}})=\Psi$, this expectation becomes

$$E[\hat{V}ar(y)] = [\underline{w}' \, Var(\underline{z})\underline{w} + tr(Var(\underline{z}) \cdot \Psi)] + \sigma^2, \tag{4.25}$$

using the theorem for taking the expectation of quadratic forms (Graybill (1976)). From (4.17) and noting that $Var(\underline{z})$ is an identity matrix, this result can be rewritten as

$$E[\hat{V}ar(y)] = Var(y) + tr(\Psi). \tag{4.26}$$

This result shows that the estimator for variance is biased on average by an amount equal to the trace of the variance of $\underline{\hat{w}}$. Thus this bias will affect the prediction of the minimum variance level once the approximate minimum variance solution has been identified.

To obtain an unbiased estimator, we use an estimate of the bias as a correction factor, subtracting its value from the original estimate of the variance response. The first step in determining the correction factor is to examine the origins of the variance-covariance matrix Ψ, $\underline{\hat{w}}$. Another way of viewing this vector is as a vector of estimated slopes in the direction of the noise variables. That is,

$$\underline{\hat{w}} = \begin{bmatrix} \frac{\partial \hat{y}}{\partial z_1} \\ \vdots \\ \frac{\partial \hat{y}}{\partial z_p} \end{bmatrix},$$

and, therefore, its variance is

$$\Psi = \mathrm{Var}\begin{bmatrix} \frac{\partial \hat{y}}{\partial z_1} \\ \vdots \\ \frac{\partial \hat{y}}{\partial z_p} \end{bmatrix} = \begin{bmatrix} \mathrm{var}\left(\frac{\partial \hat{y}}{\partial z_1}\right) & \mathrm{cov}\left(\frac{\partial \hat{y}}{\partial z_1}, \frac{\partial \hat{y}}{\partial z_2}\right) & \cdots & \mathrm{cov}\left(\frac{\partial \hat{y}}{\partial z_1}, \frac{\partial \hat{y}}{\partial z_p}\right) \\ & \mathrm{var}\left(\frac{\partial \hat{y}}{\partial z_2}\right) & \cdots & \mathrm{cov}\left(\frac{\partial \hat{y}}{\partial z_2}, \frac{\partial \hat{y}}{\partial z_p}\right) \\ & & \ddots & \vdots \\ \mathrm{sym} & & & \mathrm{var}\left(\frac{\partial \hat{y}}{\partial z_1}\right) \end{bmatrix}.$$

The trace of this variance-covariance matrix is the sum of its diagonal elements,

$$\mathrm{tr}(\Psi) = \sum_{i=1}^{p} \mathrm{var}\left(\frac{\partial \hat{y}}{\partial z_i}\right), \tag{4.27}$$

such that

$$\mathrm{var}\left(\frac{\partial \hat{y}}{\partial z_i}\right) = \mathrm{var}(\hat{\gamma}_i + \sum_{j=1}^{k} \hat{\lambda}_{ij} x_j).$$

Exploiting some results of the orthogonality of factorial designs or their fractions, the variances of the estimated noise main effects and noise by signal interactions are functions of the variance of the random error term ε and the number of design points in the factorial portion of the experimental design; namely,

$$\mathrm{var}(\hat{\gamma}_i) = \mathrm{var}\left(\hat{\lambda}_{ij}\right) = \frac{\sigma^2}{F} \tag{4.28}$$

where F denotes the number of factorial design points. Another significant result of orthogonality in the design is that the covariances of the estimated coefficients are zero-valued. Using s^2 in place of σ^2, the estimator of the bias correction factor (BCF) is

$$\hat{BCF} = \sum_{i=1}^{p}\left[\hat{var}(\hat{\gamma}_i) I_{\hat{\gamma}_i} + \sum_{j=1}^{k} \hat{var}(\hat{\lambda}_{ij}) x_j^2 I_{\hat{\lambda}_{ij}} \right] \tag{4.29}$$

where I_t is an indicator function which assumes the value of 1 if the corresponding estimated model coefficient is nonzero-valued or significant, and 0 otherwise.

Once the minimum variance level has been estimated, the next step is to formulate and solve the constrained mean response optimization problem. The first step is to derive the mean approximation function from the fitted composite model. This function is based on the application of the expectation operator on the composite model for the response y. Since the error term and the coded noise variables $\underline{z}$ having zero expectation,

$$E(y) = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x}, \tag{4.30}$$

assuming that the composite model of (4.13) is appropriate. The estimated mean model or mean approximation function is constructed by substituting the estimated coefficients in place of the true model parameters so that

$$\hat{E}(y) = \hat{\beta}_0 + \underline{x}'\hat{\underline{\beta}} + \underline{x}'\hat{B}\underline{x}. \tag{4.31}$$

The constrained optimization problem is then set up with the mean approximation function as the objective function to be minimized or maximized, and the constraint is formulated by setting the variance approximation function equal to minimum variance level. The variance constraint can also be formulated by using the estimated normal equation of (4.22), or simply by restricting the feasible region in the signal variable space to the projection of the unique stationary point in the subspace of the dispersion signal variables in the solution space of all the signal variables. After the optimal solution to the constrained optimization problem is identified, an estimate of the optimal mean response level can be computed using (4.31).

4.3.3 A NUMERICAL EXAMPLE

For the inventory model with two signal variables and three noise factors, the postulated model is

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{z}'\underline{\gamma} + \underline{x}'\Lambda\underline{z} + \varepsilon,$$

with

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \underline{\gamma} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}, \text{ and } \Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \end{bmatrix}.$$

The design used to estimate the signal and noise factor main effects and interactions is 1/2 fraction of a 2^5 factorial design (Resolution V) with defining contrast $x_1x_2z_1z_2z_3$. The same starting location,

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 1000 \\ 1000 \end{bmatrix},$$

and same initial experimental subregion used in the previous numerical illustrations of the earlier methods,

$$\{(\xi_1, \xi_2) : \xi_1 \in [500,1500] \text{ , } \xi_2 \in [500,1500]\},$$

are utilized. The experimental design in both its natural and coded variable forms, as well as the experimental results for the initial subregion, are given in the following table.

Table 4.12 Experimental Design and Results - First Subregion

x_1	x_2	z_1	z_2	z_3	ξ_1	ξ_2	ζ_1	ζ_2	ζ_3	y_i
-1	-1	-1	-1	1	500	500	0.15	90	17	112428
1	-1	-1	-1	-1	1500	500	0.15	90	13	98795
-1	1	-1	-1	-1	500	1500	0.15	90	13	102933
1	1	-1	-1	1	1500	1500	0.15	90	17	93794
-1	-1	1	-1	-1	500	500	0.25	90	13	61602
1	-1	1	-1	1	1500	500	0.25	90	17	55969
-1	1	1	-1	1	500	1500	0.25	90	17	58755
1	1	1	-1	-1	1500	1500	0.25	90	13	41409
-1	-1	-1	1	-1	500	500	0.15	110	13	136685
1	-1	-1	1	1	1500	500	0.15	110	17	130836
-1	1	-1	1	1	500	1500	0.15	110	17	133155
1	1	-1	1	-1	1500	1500	0.15	110	13	111715
-1	-1	1	1	1	500	500	0.25	110	17	80162
1	-1	1	1	-1	1500	500	0.25	110	13	66747
-1	1	1	1	-1	500	1500	0.25	110	13	70885
1	1	1	1	1	1500	1500	0.25	110	17	62080
0	0	0	0	0	1000	1000	0.20	100	15	80271
0	0	0	0	0	1000	1000	0.20	100	15	80255

The regression analysis results for the fitted composite model are summarized in the Table 4.13.

Table 4.13 First Fitted Composite Model

```
The regression equation is
C11 = 87693 - 5954 x1 - 4281 x2 - 26421 z1 + 10411 z2 + 2275 z3
      - 1137 X1*X2 + 304 X1*Z1 - 235 X1*Z2 + 726 X1*Z3 + 362 X2*Z1
      - 293 X2*Z2  + 330 X2*Z3
```

Predictor	Coef	Stdev	t-ratio	p
Constant	87693	1626	53.93	0.000
x1	-5954	1725	-3.45	0.018
x2	-4281	1725	-2.48	0.056
z1	-26421	1725	-15.32	0.000
z2	10411	1725	6.04	0.002
z3	2275	1725	1.32	0.244
x1*x2	-1137	1725	-0.66	0.539
x1*z1	304	1725	0.18	0.867
x1*z2	-235	1725	-0.14	0.897
x1*z3	726	1725	0.42	0.691
x2*z1	362	1725	0.21	0.842
x2*z2	-293	1725	-0.17	0.872
x2*z3	330	1725	0.19	0.856

s = 6898   R-sq = 98.3%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	12	13883162624	1156930048	24.31	0.001
Error	5	237937600	47587520		
Total	17	14121099264			

From the above results, the quality of fit for the estimated composite model is excellent (r^2=.983 and F-test p-value of .001). The interaction terms between signal and noise variables are, however, not significant. To expose these interactions, the experimental subregion region size for the signal variables was enlarged by increasing the design widths by a factor of 2. The second experimental subregion is

$$\{(\xi_1,\xi_2) : \xi_1 \in [500,2500] \, , \, \xi_2 \in [500,2500]\}.$$

The results from the second set of experimental runs and the ensuing regression analysis are listed in Tables 4.14 and 4.15.

Table 4.14 Experimental Results - Second Subregion

ξ_1	ξ_2	ζ_1	ζ_2	ζ_3	y_i
500	500	0.15	90	17	112428
2500	500	0.15	90	13	87626
500	2500	0.15	90	13	93000
2500	2500	0.15	90	17	78062
500	500	0.25	90	13	61602
2500	500	0.25	90	17	47767
500	2500	0.25	90	17	52093
2500	2500	0.25	90	13	27514
500	500	0.15	110	13	136685
2500	500	0.15	110	17	121981
500	2500	0.15	110	17	124374
2500	2500	0.15	110	13	91447
500	500	0.25	110	17	80162
2500	500	0.25	110	13	56195
500	2500	0.25	110	13	62281
2500	2500	0.25	110	17	48242
1500	1000	0.20	100	15	70893
1500	1000	0.20	100	15	70815

Table 4.15 Second Fitted Composite Model

```
The regression equation is
C11 = 79065 - 10237 x1 - 7965 x2 - 25609 z1 + 10080 z2 + 3047 z3
      - 573 X1X2 + 684 X1Z1 - 468 X1Z2 + 1111 X1Z3 + 1015 X2Z1
      - 620 X2Z2 + 519 X2Z3

Predictor        Coef       Stdev    t-ratio        p
Constant        79065        1789      44.21    0.000
x1             -10237        1897      -5.40    0.003
x2              -7965        1897      -4.20    0.009
z1             -25609        1897     -13.50    0.000
z2              10080        1897       5.31    0.003
z3               3047        1897       1.61    0.169
x1x2             -573        1897      -0.30    0.775
x1z1              684        1897       0.36    0.733
x1z2             -468        1897      -0.25    0.815
x1z3             1111        1897       0.59    0.583
x2z1             1015        1897       0.54    0.616
x2z2             -620        1897      -0.33    0.757
x2z3              519        1897       0.27    0.795

s = 7588        R-sq = 98.1%

Analysis of Variance

SOURCE       DF           SS           MS         F        p
Regression   12  15022108672   1251842304     21.74    0.002
Error         5    287888384     57577664
Total        17  15309996032
```

The estimated composite model still has an excellent quality of fit but the interaction terms remain insignificant. A review of the regression results show an improvement with respect to the actual significance levels (or p-values) of the estimated interaction coefficients. Assuming that an approximately linear improving trend exists relative to the magnitude of the widths of the experimental subregion, the subregion size is further magnified by raising the widths to eight times their original length (i.e., d_i=8000, for i=1,2). The resulting subregion is

$$\{(\xi_1, \xi_2) : \xi_1 \in [500, 8500] \, , \, \xi_2 \in [500, 8500]\}.$$

Another set of experimental runs is undertaken using the original fractional factorial design augmented by 2 center runs. The third set of observations and the corresponding regression analysis results from the construction of the third fitted composite model are presented in Tables 4.16 and 4.17.

Table 4.16 Experimental Results - Third Subregion

ξ_1	ξ_2	ζ_1	ζ_2	ζ_3	y_i
500	500	0.15	90	17	112428
8500	500	0.15	90	13	31181
500	8500	0.15	90	13	61995
8500	8500	0.15	90	17	24816
500	500	0.25	90	13	61602
8500	500	0.25	90	17	20139
500	8500	0.25	90	17	34811
8500	8500	0.25	90	13	16903
500	500	0.15	110	13	136685
8500	500	0.15	110	17	69530
500	8500	0.15	110	17	91572
8500	8500	0.15	110	13	25785
500	500	0.25	110	17	80162
8500	500	0.25	110	13	24086
500	8500	0.25	110	13	40195
8500	8500	0.25	110	17	13552
4500	4500	0.20	100	15	32617
4500	4500	0.20	100	15	32673

Table 4.17 Third Fitted Composite Model

```
The regression equation is
C11 = 50596 - 24591 x1 - 14136 x2 - 16409 z1 + 7356 z2 + 3036 z3
      + 6151 X1X2 + 6830 X1Z1 - 2366 X1Z2 + 724 X1Z3 + 4070 X2Z1
      - 3283 X2Z2 - 552 X2Z3
```

Predictor	Coef	Stdev	t-ratio	p
Constant	50596	3508	14.42	0.000
x1	-24591	3721	-6.61	0.001
x2	-14136	3721	-3.80	0.013
z1	-16409	3721	-4.41	0.007
z2	7356	3721	1.98	0.105
z3	3036	3721	0.82	0.452
x1x2	6151	3721	1.65	0.159
x1z1	6830	3721	1.84	0.126
x1z2	-2366	3721	-0.64	0.553
x1z3	724	3721	0.19	0.853
x2z1	4070	3721	1.09	0.324
x2z2	-3283	3721	-0.88	0.418
x2z3	-552	3721	-0.15	0.888

s = 14883 R-sq = 94.8%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	12	20086517760	1673876480	7.56	0.018
Error	5	1107587072	221517408		
Total	17	21194104832			

The quality of fit for the estimated composite model deteriorated as expected but the fit is still acceptable, with r^2=.948 and F-test p-value at 0.018. The individual p-values for the interaction regression coefficients improved significantly but, at a critical significance level of .10, are not significant. Still under the partially confirmed assumption of the existence of an improving trend in these significance values as region size is magnified, the widths are further widened to d_i=11,000. The fourth subregion corresponding to these specifications is

$$\{(\xi_1, \xi_2) : \xi_1 \in [500,11500]\ ,\ \xi_2 \in [500,11500]\}.$$

A new set of experimental runs in this subregion results in the following data and model fitting results.

Table 4.18 Experimental Results - Fourth Subregion

ξ_1	ξ_2	ζ_1	ζ_2	ζ_3	y_i
500	500	0.15	90	17	112428
11500	500	0.15	90	13	32391
500	11500	0.15	90	13	55306
11500	11500	0.15	90	17	17430
500	500	0.25	90	13	61602
11500	500	0.25	90	17	25714
500	11500	0.25	90	17	32091
11500	11500	0.25	90	13	25680
500	500	0.15	110	13	136685
11500	500	0.15	110	17	47055
500 0	1150	0.15	110	17	82703
11500	11500	0.15	110	13	18784
500	500	0.25	110	17	80162
11500	500	0.25	110	13	29722
500	11500	0.25	110	13	36593
11500	11500	0.25	110	17	20341
6000	6000	0.20	100	15	21815
6000	6000	0.20	100	15	21677

Table 4.19 Fourth Fitted Composite Model

```
The regression equation is
C11 = 47677 - 23778 x1 - 14802 x2 - 11930 z1 + 5588 z2 + 1323 z3
      + 8221 X1X2 + 10154 X1Z1 - 3752 X1Z2 - 827 X1Z3 + 4490 X2Z1
      - 2098 X2Z2 + 703 X2Z3

Predictor        Coef       Stdev    t-ratio        p
Constant        47677        4288      11.12    0.000
x1             -23778        4548      -5.23    0.003
x2             -14802        4548      -3.25    0.023
z1             -11930        4548      -2.62    0.047
z2               5588        4548       1.23    0.274
z3               1323        4548       0.29    0.783
x1x2             8221        4548       1.81    0.130
x1z1            10154        4548       2.23    0.076
x1z2            -3752        4548      -0.82    0.447
x1z3             -827        4548      -0.18    0.863
x2z1             4490        4548       0.99    0.369
x2z2            -2098        4548      -0.46    0.664
x2z3              703        4548       0.15    0.883

s = 18191       R-sq = 91.9%

Analysis of Variance

SOURCE        DF          SS          MS         F        p
Regression    12 18725023744  1560418560      4.72    0.049
Error          5  1654531584   330906112
Total         17 20379553792
```

The preceding results indicate that with the enlargement of the subregion size, the quality of fit becomes marginal as r^2=.919 and the F-test p-value is at 0.049, results which are still acceptable. What is gained from this exercise is the identification of a significant signal by noise interaction, x_1z_1 (the interaction between reorder point and the demand arrival rate), with a significant p-value of 0.076 when a conservative significance level of 0.10 is utilized. From these results, it is concluded that the next phase of the analysis be commenced.

In this phase of the analysis, the same critical p-value is used for determining which signal factor main effects and interactions are included in the fitted composite model. A more liberal criterion is utilized for the inclusion of model terms involving the noise variables, including the noise main effects and noise by signal interactions. A critical p-value of 0.50 is prescribed so as to accommodate the weaker model terms involving the noise variables. The reasoning behind this is twofold: the application of the variance operator on the fitted composite model will result in the squaring of the corresponding model coefficients thus amplifying their effect; and secondly, this approach will result in the inclusion of the weaker noise by signal interactions. Shin Taguchi (in Nair (1992)) notes that the consideration of even the weaker noise by signal interactions could result in substantial variability reduction.

Adopting the aforementioned criteria, the fitted composite model at the current subregion is

$$\begin{aligned}\hat{y} = 47677 - 23778x_1 - 14802x_2 - 11930z_1 + 5588z_2 \\ + 10154x_1z_1 - 3752x_1z_2 + 4490x_2z_1.\end{aligned}$$

Utilizing the estimated regression coefficients in this fitted composite model, the variance approximation function is

$$\begin{aligned}\hat{V}\mathrm{ar}(y) = s^2 &+ (-11930)^2\sigma_{z_1}^2 + (5588)^2\sigma_{z_2}^2 + (10154x_1)^2\sigma_{z_1}^2 \\ &+ (-3752x_1)^2\sigma_{z_2}^2 + (4490x_2)^2\sigma_{z_1}^2 + 2(-11930)(10154x_1)\sigma_{z_1}^2 \\ &+ 2(-11930)(4490x_2)\sigma_{z_1}^2 + 2(5588)(-3552x_1)\sigma_{z_2}^2 \\ &+ 2(10154x_1)(4490x_2)\sigma_{z_1}^2.\end{aligned}$$

With s^2=9,522 and the variances of the coded noise variables all equal to unity, the above variance approximation function reduces to

$$\begin{aligned}\hat{V}\mathrm{ar}(y) = 9522 &+ (-11930)^2 + (5588)^2 + [(10154)^2 + (3752)^2]x_1{}^2 + (4490)^2 x_2{}^2 \\ &+ 2[(-11930)(10154) + (5588)(-3552)]x_1 + 2(-11930)(4490)x_2 \\ &+ 2(10154)(4490)x_1x_2.\end{aligned}$$

Taking the first partial derivatives and setting them equal to zero, we get the stationary point

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cong \begin{bmatrix} 1.4893 \\ -0.7110 \end{bmatrix}$$

which corresponds to

$$\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \cong \begin{bmatrix} 14{,}191 \\ 2{,}089 \end{bmatrix}.$$

It is noted that this stationary point is outside the current subregion of operability and, thus, is not an acceptable solution. A new subregion is thus defined with the same dimensions as the most recent subregion, but with a different location so as to include this stationary point. The new experimental region is

$$\{(\xi_1, \xi_2) : \xi_1 \in [7000, 18000],\ \xi_2 \in [500, 11500]\},$$

with the new design center in closer proximity to the stationary point location. A new set of experimental trials undertaken at this new subregion yielded the following data.

Table 4.20 Experimental Results - Fifth Subregion

ξ_1	ξ_2	ζ_1	ζ_2	ζ_3	y_i
7000	500	0.15	90	17	56063
18000	500	0.15	90	13	45025
7000	11500	0.15	90	13	22852
18000	11500	0.15	90	17	27767
7000	500	0.25	90	13	19977
18000	500	0.25	90	17	38694
7000	11500	0.25	90	17	14803
18000	11500	0.25	90	13	38748
7000	500	0.15	110	13	63891
18000	500	0.15	110	17	41055
7000	11500	0.15	110	17	57470
18000	11500	0.15	110	13	29565
7000	500	0.25	110	17	28795
18000	500	0.25	110	13	42715
7000	11500	0.25	110	13	15909
18000	11500	0.25	110	17	33282
12500	6000	0.20	100	15	17524
12500	6000	0.20	100	15	17575

Table 4.21 Fifth Fitted Composite Model

```
The regression equation is
C11 = 33984 + 1068 x1 - 5989 x2 - 6923 z1 + 3047 z2 + 1203 z3
      + 1223 X1X2 + 8176 X1Z1 - 3499 X1Z2 - 3110 X1Z3 + 2559 X2Z1
      + 960 X2Z2  + 2078 X2Z3
```

Predictor	Coef	Stdev	t-ratio	p
Constant	33984	2796	12.15	0.000
x1	1068	2966	0.36	0.733
x2	-5989	2966	-2.02	0.099
z1	-6923	2966	-2.33	0.067
z2	3047	2966	1.03	0.351
z3	1203	2966	0.41	0.702
x1x2	1223	2966	0.41	0.697
x1z1	8176	2966	2.76	0.040
x1z2	-3499	2966	-1.18	0.291
x1z3	-3110	2966	-1.05	0.342
x2z1	2559	2966	0.86	0.428
x2z2	960	2966	0.32	0.759
x2z3	2078	2966	0.70	0.515

s = 11863 R-sq = 81.8%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	12	3163364608	263613712	1.87	0.253
Error	5	703669248	140733840		
Total	17	3867033856			

The resulting estimated model's fit characteristics are marginal to unacceptable at r^2=.818 and regression significance level at p-value=0.253. To improve the quality of fit without sacrificing the presence of a significant signal by noise interaction coefficient through the reduction of the subregion size, the order of the model was increased to second order in the signal variables. Thus the hypothesized model is

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Lambda\underline{z} + \varepsilon .$$

Here the matrix of coefficients B of the second order terms in the signal variables is given by

$$B = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} \\ \text{sym} & \beta_{22} \end{bmatrix}.$$

To facilitate the estimation of the additional model terms, the original design was augmented with axial runs, with the axial points being located a distance half the radius of the spheroidal subregion in the coded signal space to provide protection against bias. The results of the axial runs are augmented to the already available factorial and center runs' data for the current subregion. The complete data set and the results of the ensuing model fitting exercise are summarized in the following two tables.

Table 4.22 Augmented Design and Results - Fifth Subregion

ξ_1	ξ_2	ζ_1	ζ_2	ζ_3	y_i
7000	500	0.15	90	17	56063
18000	500	0.15	90	13	45025
7000	11500	0.15	90	13	22852
18000	11500	0.15	90	17	27767
7000	500	0.25	90	13	19977
18000	500	0.25	90	17	38694
7000	11500	0.25	90	17	14803
18000	11500	0.25	90	13	38748
7000	500	0.15	110	13	63891
18000	500	0.15	110	17	41055
7000	11500	0.15	110	17	57470
18000	11500	0.15	110	13	29565
7000	500	0.25	110	17	28795
18000	500	0.25	110	13	42715
7000	11500	0.25	110	13	15909
18000	11500	0.25	110	17	33282
12500	6000	0.20	100	15	17524
12500	6000	0.20	100	15	17575
12500	6000	0.20	100	15	17499
12500	6000	0.20	100	15	17522
16389	6000	0.20	100	15	25258
8611	6000	0.20	100	15	12539
12500	9889	0.20	100	15	20753
12500	2111	0.20	100	15	16783

Table 4.23 Refitted Fifth Composite Model

```
The regression equation is
y  = 15951 + 1534 x1 - 5471 x2 - 6923 z1 + 3047 z2 + 1203 z3
      + 1223 X1X2  + 8176 X1Z1 - 3499 X1Z2 - 3110 X1Z3 + 2559 X2Z1
      + 960 X2Z2   + 2078 X2Z3 + 10108 X1**2 + 9847 X2**2

Predictor        Coef        Stdev    t-ratio        p
Constant        15951         2139       7.46    0.000
x1               1534         1293       1.19    0.266
x2              -5471         1293      -4.23    0.002
z1              -6923         1333      -5.19    0.000
z2               3047         1333       2.29    0.048
z3               1203         1333       0.90    0.390
x1x2             1223         1333       0.92    0.383
x1z1             8176         1333       6.13    0.000
x1z2            -3499         1333      -2.63    0.028
x1z3            -3110         1333      -2.33    0.045
x2z1             2559         1333       1.92    0.087
x2z2              960         1333       0.72    0.490
x2z3             2078         1333       1.56    0.153
x1**2           10108         5490       1.84    0.099
x2**2            9847         5490       1.79    0.106

s = 5332        R-sq = 94.9%


Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression   14  4796211200   342586368     12.05    0.000
Error         9   255872272    28430240
Total        23  5052080128
```

Using a critical significance level value of 0.10 for the estimated signal variable coefficients' p-values and employing the more liberal approach described earlier for the consideration of the coefficients of the model terms involving the noise variables , the fitted composite function is

$$\begin{aligned}\hat{y} = 15951 - 5471x_2 + 10108x_1^2 + 9847x_2^2 - 6923z_1 + 3047z_2 + 1203z_3 \\ + 8176x_1z_1 - 3499x_1z_2 - 3110x_1z_3 + 2559x_2z_1 + 960x_2z_2 + 2078x_2z_3.\end{aligned}$$

The estimated coefficients from the fitted model, together with s^2=1028.6667, produce the following variance approximation function

$$\begin{aligned}\hat{V}ar(y) = 1028.6667 + (6923)^2 + (3047)^2 + (1203)^2 + [(8176)^2 + (3499)^2 + (3110)^2]x_1^2 \\ + [(2559)^2 + (960)^2 + (2078)^2]x_2^2 + 2[(-6923)(8176) + (3047)(-3499) + (1203)(3110)]x_1 \\ + 2[(-6923)(2559) + (3047)(960) + (1203)(2078)]x_2 \\ + 2[(8176)(2559) + (-3499)(960) + (-3110)(2078)]x_1x_2.\end{aligned}$$

The stationary point for this quadratic function in both signal variables is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.7589 \\ 0.3280 \end{bmatrix} \Leftrightarrow \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 16,674 \\ 7,804 \end{bmatrix}$$

which is interior to the current subregion. To verify that it is the unique minimum estimated variance solution, the Hessian matrix is formulated and its definiteness is ascertained. The Hessian matrix is

$$\mathbf{H}(\underline{x}) = \begin{bmatrix} 177,524,160 & 22,201,528 \\ 22,201,528 & 23,576,330 \end{bmatrix}$$

which is positive definite. The stationary point solution is therefore the minimum estimated variance solution. Utilizing the variance approximation function, the predicted variance at that point is 741,009 dollars^2, or a predicted standard deviation of \$860.819.

Next we solve for the estimate for the bias correction factor to get the unbiased variance and standard deviation estimates. We start with the vector of estimated slopes with respect to the noise variables

$$\hat{\underline{w}} = \begin{bmatrix} -6923 + 8176x_1 + 2559x_2 \\ 3047 - 3499x_1 + \ \ 960x_2 \\ 1203 - 3110x_1 + 2078x_2 \end{bmatrix}.$$

The trace of its estimated variance using equations (4.27) and (4.29) is

$$B\hat{C}F = 3s^2\left(\frac{1}{16} + \frac{1}{16}x_1^2 + \frac{1}{16}x_2^2\right)$$

which, when evaluated at the stationary point, has a value of approximately 324.708 dollars^2. Subtracting this from the earlier biased estimate yields an unbiased variance estimate of 740,685 dollars^2, or a corrected estimated standard deviation of \$860.630.

Since the minimum variance solution is a singleton in the entire signal variable subspace, the constrained optimization problem has a single point feasible region which is that stationary point. Thus all that remains is to estimate the mean response at that location. To do this, the mean response approximation function is derived from the fifth fitted composite model. Applying the expectation operator on this fitted model and taking advantage of the noise variables having zero expectation, the resulting mean response approximation function is

$$\hat{E}(y) = 15951 - 5471x_2 + 10108x_1^{\ 2} + 9847x_2^{\ 2}.$$

Substituting the optimal variance solution values for the signal variables gives an expected monthly cost value of approximately \$21,037.

Chapter 5

COMPARISONS, CONCLUSIONS, AND EXTENSIONS

In the previous chapters, three distinct approaches for the estimation of the optimum levels for the controllable input factors to a system in the presence of system nonstationarity were discussed. In addition to their common objective of system optimization, these approaches share other features as reflected in overlaps in their respective implementation methodologies as well as optimization criteria. The advantages and disadvantages of each approach correspond to differences in optimization philosophy and strategy.

The first approach focuses primarily on the optimization of mean system performance. It makes use of response surface methodology (RSM) in modeling this traditional response of interest. To enable the direct application of RSM tools and techniques, the initial implementation strategy necessitates the use of a simplified version of the simulation model, referred to as the static simulation model. The analysis of mean response is based on the conditional distribution of the output response, that distribution being one characterized by static operating or environmental conditions. The choice for these conditions corresponds to the most likely environmental conditions.

The second strategy for the analysis of the mean response likewise makes use of RSM modeling techniques. It differs in the specification of the mean parameters of the process generators of the simulation model utilized in the experimental trials. This specification requires that the transient system behavior be emulated by the simulation model. The use of RSM on the non-steady-state output process is made possible by invoking the law of large numbers and basing the analysis of mean response on the unconditional distribution of the output response. This distribution is the infinite mixture of all the conditional distributions for the output response which, in turn, cause the nonstationary nature of the output process. Each conditional distribution is defined by a unique set of environmental conditions.

A comparative study of the two mean modeling strategies reveals the following differences. The first approach, due to its simplicity, is the more straightforward approach, requiring less effort in the construction of the simulation model. The drawback to this approach is that it is not very realistic; i.e., model validity is doubtful. Its corresponding solution is an estimate of just the conditionally optimal solution The second strategy uses a more realistic simulation model. However, due to the added transient feature to the simulation model, and the ensuing analysis being based on the unconditional (or the mixture) distribution of the system response, substantially more effort is required in the programming of the simulation model as well as in the analysis of the output results due to the variance inflating effect of the nonstationary system behavior incorporated into the simulation model. This added instability impacts the precision of the regression model parameter estimates, hampering both prediction and optimization. For example, as was shown in Chapter 3 and especially in Chapter 4, not only do the noise variables have an inflationary effect on variance - this effect is also location-dependent. With the presence of significant noise by signal factor interactions, variance is heterogeneous from location to location in the signal factor space. Under such a scenario, weighted least square (WLS) should be utilized to get minimum variance estimators (see Myers (1976) and Graybill (1976)).

To handle the problem of increased as well as heterogeneous response variance resulting from these added sources of variation, i.e., conditional nonstationarity in system behavior, a robust parameter design (RPD) approach was proposed. This approach embodies a quality philosophy that minimizes risk or expected loss from product or process inconsistency. Using Taguchi's preferred loss function, the quadratic error loss function, this approach attempts to model risk which is expressed in the form of a mean squared error (MSE). This allows for the simultaneous consideration of the response mean and variance. A function of an estimator of MSE, the S/N ratio, is constructed using data generated from the simulation model via a sensitivity sampling plan. This sampling technique is a nonrandom sampling technique that uses cross-classified experimental designs on the controllable (signal) and uncontrollable (noise) input factors. The noise factors are, of course, assumed to be uncontrollable in the actual real-world scenario, but are completely controllable during the experimental (simulation) stage. The estimate for the optimal solution is that combination for the signal variables which maximizes S/N, which, in turn, minimizes risk or MSE.

Certain major criticisms regarding the appropriateness and efficiency of the original RPD approach, also known as the Taguchi Method (TM), were discussed and a majority of these were addressed in Chapter 3. Their resolution brought about the revised RPD approach combining some of the best features of TM and RSM in modeling MSE. An important unresolved issue in that chapter is the criticism regarding the lack of predictive capability with respect to the mean and variance responses. Their joint consideration via the modeling and analysis of the S/N response involves summarizing the data into a single statistic and finding the solution that optimizes the S/N response. Thus estimates of the optimal S/N response level and its corresponding optimal MSE level can be obtained. However, the partitioning of that optimal MSE level into its variance and bias components is not possible. This observation precludes the separate estimation of the response mean and variance at the minimum MSE (maximum S/N) solution making additional confirmatory or validation experimental runs essential.

A logical and natural extension to the MSE modeling approach of RPD with respect to the joint consideration of the mean and variance responses is the dual response approach, i.e., the separate modeling of those two responses. This affords the experimenter the ability to estimate the corresponding levels for these two responses of interest given a selected solution in the signal factors.

Two implementation strategies were presented in Chapter 4. The first strategy starts with the sensitivity sampling procedure of RPD, employing the crossed-arrays experimental design structure to generate a set of independent replications for each design setting or combination in the signal factors. Instead of summarizing the replication results in the form of a S/N statistic value, the observations are used to estimate the expectation and variance at each design setting for the signal factors. Estimates of the response surface functions of these two responses are fitted via regression analysis using the appropriate statistics thereby forming a dual response surface system.

The RPD philosophy of limiting the solution space to solutions that are least sensitive to the destabilizing effects the noise variables is adapted in the resolution of this dual response system. To identify the least affected solutions, the analysis first focuses on the minimization of the variance response. Once the subregion containing the minimum variance solution(s) is determined or approximated, the analysis then addresses mean response. The solution that results in the best possible mean response level subject to the minimum variance constraint is identified as an estimate of the optimal solution.

In the preceding analysis, the concept of sensitivity to noise variables was equated to variation in response. The optimization procedure involved a constrained optimization exercise on the mean response function, with the variance response function serving to locate and restrict the solution space to the set least affected by the noise variables which defines the minimum variance solution(s).

The second dual response strategy, the variance function approximation method, constitutes an attempt to address the potential for an inordinate amount of experimental effort when the crossed-arrays experimental plans are used in the variance fitting approach. Instead of specifying two separate experimental plans for the signal and noise factors, a common experimental design containing both variable types was examined. This approach is made possible by postulating a composite model combining both factor types into a single metamodel, treating the noise variables in the same light (i.e., as fixed or controllable factors) as the signal factors. As with the other approaches, this strategy hinges on the experimenter's ability to completely control both signal and noise variables during the course of the experiment --- a characteristic feature of simulation experiments. The mean and variance function approximations are constructed by applying the expectation and variance operators on the postulated composite function and substituting the estimated coefficients of the fitted composite model in place of the unknown model parameters. The same dual response constrained optimization tactic employed in the earlier dual response strategy is used to find the estimated optimal solution.

In both dual response modeling scenarios, the experimenter is afforded the ability to estimate the corresponding mean and variance values for the estimates of the optimal solution. The second method has the added capability of providing a preliminary test for the applicability of the dual response approach. It does so by detecting the presence (or otherwise) of significant noise by signal interactions which are crucial to the experimenter's ability to control product or process variation. Without these interactions, the variance approximation function would be devoid of terms involving signal factors, indicating that the optimization of the mean approximation function is the only feasible objective.

Another advantage of the second strategy is that the search procedure is made simpler by the analysis of a strictly convex quadratic variance approximation function. A major consideration in the successful implementation of this strategy is the representations provided by both the postulated and fitted composite models. Concerns regarding the composite model involve a plethora of possibilities for model misspecification due to the limitation of the order of the approximating polynomial to second order, the use of heuristic screening procedures for the inclusion (or exclusion) of model terms in the fitted model, and the use of estimated noise variable mean and variance values in the coding transformation for the noise variables.

The five strategies presented and their implementation tactics were applied to the illustrative problem of a continuous review inventory system with nonstationary demand and lead time process components. The generated solutions and their corresponding predicted optimal performance levels are summarized in the following table.

Table 5.1 Solutions to the Inventory Optimization Problem

METHOD	ξ_1 Reorder Point	ξ_2 Reorder Quantity	$\bar{y}$ (predicted)	s_y (predicted)
RSM (Static) Mean Model Approach	9,324	4,843	$11,985	*
RSM (Nonstationary) Mean Model Approach	9,895	3,659	$12,924	*
S/N (MSE) Model Approach	11,399	3,702	*	*
Dual Response (Crossed Array) Model Approach	9,891	13,953	$22,352	$22,24.00
Dual Response (Composite) Model Approach	16,674	7,804	$21,307	$860.63

A review of the above results shows a large variation in predicted mean response values. There are, however, logical groupings that can be accounted for by the similarities in optimization criteria. The clustering of the predicted means resulting from the solutions derived by the two mean modeling approaches form a group, while those for the dual response methods' solutions form another class. The main difference in the two groupings is that variance minimization was paramount with respect to the dual response methods, with consideration for the mean response being of secondary importance.

Inspection of the predicted standard deviations for these two methods shows a large discrepancy between the two estimates. Since the predicted standard deviation value for the crossed array dual response method was derived from fitting the variance response directly to the data, we focus our attempt to reconcile or explain this disparity on the composite model dual response method. One possibility is that the predicted standard deviation for this approach underestimates the true process standard deviation. If this is the case, the variance function would be biased by the exclusion of certain variance function "root terms" in the composite model. Here we refer to model terms in the composite function that involve the noise variables. When the variance operator is applied to the composite model, these root terms essentially are the foundations for the variance function. With their exclusion from the composite model, the resulting variance function may be inadequate in explaining true process variation.

In the inventory example, variance root terms involving the noise by noise factor interactions ($z_i z_j$) were not included in the postulated composite model. We therefore begin our reexamination of the results for the composite model approach by including these terms in the analysis of the experimental results for the last subregion considered. The postulated composite model is then

$$y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{i=1}^{p} \sum_{j=i}^{p} \beta_{ij} x_i x_j + \sum_{i=1}^{p} \sum_{j=1}^{m} \lambda_{ij} x_i z_j + \sum_{i=1}^{m} \sum_{j=i+1}^{m} \delta_{ij} z_i z_j + \varepsilon. \tag{5.1}$$

Using the experimental results from Table 4.22, the regression analysis results from fitting this composite model to the aforementioned data is presented in Table 5.2.

Table 5.2 Refitted Fifth Composite Model

The regression equation is

```
y = 15951 + 1534 x1 - 5471 x2 - 6923 z1 + 3047 z2 + 1203 z3 + 1223 x1x2
          + 8176 X1Z1 - 3499 X1Z2 - 3110 X1Z3 + 2559 X2Z1 + 960 X2Z2
          + 2078 x2z3 - 1987 z1z2 - 1425 z1z3 - 138 z2z3 + 10108 x1**2
          + 9847 x2**2
```

Predictor	Coef	Stdev	t-ratio	p
Constant	15951	2071	7.70	0.000
x1	1534	1252	1.23	0.266
x2	-5471	1252	-4.37	0.005
z1	-6923	1291	-5.36	0.002
z2	3047	1291	2.36	0.056
z3	1203	1291	0.93	0.387
x1x2	1223	1291	0.95	0.380
x1z1	8176	1291	6.34	0.001
x1z2	-3499	1291	-2.71	0.035
x1z3	-3110	1291	-2.41	0.053
x2z1	2559	1291	1.98	0.095
x2z2	960	1291	0.74	0.485
x2z3	2078	1291	1.61	0.158
z1z2	-1987	1291	-1.54	0.175
z1z3	-1425	1291	-1.10	0.312
z2z3	-138	1291	-0.11	0.919
x1**2	10108	5315	1.90	0.106
x2**2	9847	5315	1.85	0.113

$s_y = 5162$ $r^2 = .968$

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	17	4892180480	287775232	10.80	0.004
Error	6	159904864	26650800		
Total	23	5052084224			

Using the same model term screening procedures of Chapter 4, the fitted composite model becomes

$$\hat{y} = 15951 - 5471x_2 + 10108x_1^{\,2} - 6923z_1 + 3047z_2 + 1203z_3$$
$$+ 8176x_1z_1 - 3499x_1z_2 - 3110x_1z_3 + 2559x_2z_1$$
$$+ 960x_2z_2 + 2078x_2z_3 - 1987z_1z_2 - 1425z_1z_3,$$

from which the new variance approximation function is constructed, namely

$$\begin{aligned}\hat{V}\mathrm{ar}(y) = {} & 1028.6667 + (6923)^2 + (3047)^2 + (1203)^2 + [(8176)^2 + (3499)^2 + (3110)^2]x_1^2 \\ & + [(2559)^2 + (960)^2 + (2078)^2]x_2^2 + 2[(-6923)(8176) + (3047)(-3499) + (1203)(3110)]x_1 \\ & + 2[(-6923)(2559) + (3047)(960) + (1203)(2078)]x_2 \\ & + 2[(8176)(2559) + (-3499)(960) + (-3110)(2078)]x_1x_2 \\ & + (-1987)^2 + (-1425)^2.\end{aligned}$$

Note that this variance approximation function differs from the previous one by just a constant (1987^2+1425^2). This is due to $\mathrm{cov}(z_i,z_iz_j)=0$ which is a result of the z_i being independent and having zero expectation. With the inclusion of these additional variance approximation function terms, the $\hat{\delta}_{ij}^2$, the new bias correction factor is

$$\mathrm{BCF}_{new} = \mathrm{BCF}_{old} + \sum_{i=1}^{m}\sum_{j=i+1}^{m} \mathrm{var}(\hat{\delta}_{ij}). \qquad (5.2)$$

Since the new terms in the variance approximation function do not involve the signal variables, the optimal solution in these variables remains unchanged, i.e,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.7589 \\ 0.3280 \end{bmatrix} \Leftrightarrow \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 16{,}674 \\ 7{,}804 \end{bmatrix}.$$

With this solution, the predicted mean response value remains \$21,037.38, and the new estimate for the process standard deviation is \$2,592.17. This predicted standard deviation value is closer to that of the crossed array approach (\$2,224).

In the next section the five solutions are examined and their predicted performance levels validated using independent confirmatory runs. The results of these runs are used to further highlight the merits and shortcomings of the different approaches.

5.1 CONFIRMATORY EXPERIMENTS

Confirmatory trials or validation runs are used to gauge and examine the relative performances of the different operating policies derived through the use of the different optimization strategies. The confirmatory trials are undertaken using a simulation model instead of the real world system if and when this model sufficiently emulates transient system behavior. A primary issue to be resolved prior to conducting these trials concerns the choice for the simulation modeling approach. One approach to conducting these experiments makes use of the simulation model introduced in Chapter 2 that incorporates the transitory nature of the demand and lead time processes. A second approach is fashioned after the sensitivity sampling technique. The five different solutions are cross-classified with a design matrix on the noise variables to observe their performance under the "extreme" operating conditions defined by the design settings in the noise factors. A 2^3 fractional factorial design on the noise variables in a sensitivity sampling scenario results in eight independent replications for each solution combination in the signal variables.

Key to the selection of the appropriate simulation modeling approach is the utility of the generated output data in the estimation of system performance with respect to the process variance. Consider the first approach described above and the manner in which the performance statistic is computed from the output data. Given the *ith* trial operating policy or design setting in the signal variables, the output process can be denoted

$$\{y_{ij}\ ,\ j = 1, \ldots, q\}$$

where y_{ij} represents the total monthly inventory cost for the *jth* month. The response statistic of interest is

$$\bar{y}_i = \frac{\sum_{j=1}^{q} y_{ij}}{q},$$

which we denote simply by y_i. For the nonstationary simulation model of Chapter 2, the monthly cost output is

$$y_{ij} = E(y_{ij} | \underline{\zeta_1}_j, \underline{\zeta_2}_j, \underline{\zeta_3}_j) + \varepsilon_{ij}, \tag{5.3}$$

with $\underline{\zeta_k}_j$ denoting the vector of realizations for the *kth* noise variable within the *jth* month. The statistic y_i in terms of the monthly cost output is

$$y_i = \frac{\sum_{j=1}^{q} E(y_{ij} | \underline{\zeta_1}_j, \underline{\zeta_2}_j, \underline{\zeta_3}_j)}{q} + \bar{\varepsilon}_i. \tag{5.4}$$

As $q \rightarrow \infty$, we again invoke the law of large numbers and the theorem on the expectation of a conditional expectation cited in Chapter 2 to yield

$$y_i \approx E(y_{ij}) + \bar{\varepsilon}_i. \tag{5.5}$$

Thus the variance of the performance statistic is

$$\mathrm{var}(y_i) = \mathrm{var}(\bar{\varepsilon}_i), \tag{5.6}$$

which is a function of just the pure experimental error variation.

Consider now the experimental setting if the crossed-arrays approach is used; i.e., the *ith* solution in the signal factors is cross-classified with a design array on the noise variables such that the monthly cost variable for a given noise variable combination is

$$y_{ijt} = E(y_{ijt} | \zeta_{1t}, \zeta_{2t}, \zeta_{3t}) + \varepsilon_{ij} \tag{5.7}$$

where the subscript t denotes the *tth* design setting in the noise variables. Note that here the noise variable levels are fixed by design; i.e., they are not random. The response statistic is therefore

$$\begin{aligned} y_{it} &= \frac{\sum_{j=1}^{q} E(y_{ijt} | \zeta_{1t}, \zeta_{2t}, \zeta_{3t})}{q} + \bar{\varepsilon}_i \\ &= E(y_{ijt} | \zeta_{1t}, \zeta_{2t}, \zeta_{3t}) + \bar{\varepsilon}_i. \end{aligned} \tag{5.8}$$

After the different trial combinations in the noise variable are considered, k independent realizations of the response statistic are observed. We now treat the noise variables as random variables, with the k observations being viewed as independent replication data given a signal factor setting. Dropping the t-subscript, we have

$$y_i = E(y_{ij} | \zeta_1, \zeta_2, \zeta_3) + \bar{\varepsilon}_i. \tag{5.9}$$

The variance of this statistic is

$$\mathrm{var}(y_i) = \mathrm{var}[E(y_{ij} | \zeta_1, \zeta_2, \zeta_3)] + \mathrm{var}(\bar{\varepsilon}_i), \tag{5.10}$$

assuming independence between the conditional mean and the error term.

A comparison of the variances of the output response statistics of the two simulation modeling approaches reveals that the variance of the nonstationary model approach does not reflect the effect of the noise variables. That is, the averaging process used to obtain the response statistic which is the sample mean of the output stream suppresses the variance-inflating effect of the noise variables. The second approach (crossed-arrays) does reflect the effect of the noise variables on the process variance. Thus the estimate of response variance calculated from this data is more representative of the true process variation discussed in the previous chapters. This simulation approach is therefore the appropriate approach for conducting confirmatory experimental runs.

Sensitivity sampling confirmatory trials were conducted with each of the five solutions in the signal variables cross-classified with a full (2^3) factorial design on the noise variables. The results of these experiments are presented in Table 5.3.

Table 5.3 Results of Confirmatory Trials ($Z=2^3$)

METHOD	ξ_1 Reorder Point	ξ_2 Reorder Quantity	$\bar{y}$ (predicted)	s_y (predicted)	$\bar{y}$ (actual)	s_y (actual)
RSM (Static) Mean Model Approach	9,324	4,843	11,985	*	20,190.3	11,788.7
RSM (Nonstationary) Mean Model Approach	9,895	3,659	12,924	*	18,918.4	9,733.8
S/N (MSE) Model Approach	11,399	3,702	*	*	17,389.4	4,667.4
Dual Response (Crossed Array) Model Approach	9,891	13,953	22,352	2,224.00	23,775.4	5,731.9
Dual Response (Composite) Model Approach	16,674	7,804	21,037	2,592.17	26,445.3	4,802.4

From the above results, the best solution with respect to the mean response appears to be the MSE or S/N modeling strategy's, with the performances of the two mean modeling approaches' solutions yielding similar results. This observation, however, has to be tempered by the fact that the sample standard deviations are quite large. Turning now to the other response of interest, the variance response, the MSE modeling solution seems to be the best performer, with comparable results for the dual response methods' solutions. The performances (with respect to the variance response) of those solutions derived by methods giving consideration to variance are further compared to those of the mean modeling approaches to verify the earlier supposition (from the assessment of the predicted variance values in Section 5.1) regarding which solutions would produce the more consistent results. One-sided hypothesis tests of the form

$$H_o: \sigma_i^2 = \sigma_j^2$$
$$H_a: \sigma_i^2 < \sigma_j^2$$

based on the statistic

$$F = s_i^2 / s_j^2$$

were conducted. The i-subscript represents a variance-oriented solution method, while the subscript j denotes a mean modeling approach. With independent samples from normal populations, this statistic follows an F-distribution with numerator degrees of freedom (ν_i) and denominator degrees of freedom (ν_j) both set equal to 7. The actual significance level of each test is given by a p-value defined as

$$\text{p} - \text{value} = \Pr\text{ob}\{F(\nu_i, \nu_j) < s_i^2 / s_j^2\}.$$

The results for all the tests are presented in Table 5.4.

Table 5.4 F-test results for H_0: $\sigma^2_i=\sigma^2_j$ vs H_a: $\sigma^2_i<\sigma^2_j$

$F = {s_1}^2 / {s_2}^2$ $v_1 = v_2 = 7$	σ^2_j RSM(Static)Mean Model Approach	σ^2_j RSM (Nonstationary) Mean Model Approach
σ^2_i S/N (MSE) Model Approach	F=0.1568 p-value=.0.0129	F=0.2299 p-value=0.0356
σ^2_i Dual Response (Crossed-array) Model Approach	F=0.2364 p-value=0.0382	F=0.3468 p-value=0.0929
σ^2_i Dual Response (Composite) Model Approach	F=0.1660 p-value=0.0151	F=0.2434 p-value=0.0410

The above test results indicate that the variance-oriented approaches (MSE (S/N), dual response (crossed-array), and dual response (composite) modeling approaches) yielded solutions that are less sensitive to the noise variables. Again, this interpretation of the results must be viewed with caution as the tests are not very powerful due to the small number of degrees of freedom involved, and the relatively large variances of the sample process variances as given by

$$\sigma^2_{s^2_y} = \frac{2\sigma^4_y}{n-1}.$$

5.2 QUESTIONS FOR FURTHER RESEARCH

Some ramifications of composite model misspecification were illustrated in Section 5.1. The difficulties encountered clearly indicate the need for a more definitive composite model building procedure. For example, the inclusion (or exclusion) of model terms in the fitted composite model relied heavily on the p-values of the significance tests for each of the fitted regression coefficients. A conservative critical p-value of 0.10 was used in screening model terms involving just the signal variables, while a more liberal one (0.50) was adopted for assessing the variance root terms (i.e., model terms that involve the noise variables). The motivation behind this is that while the effect of these variance root terms may be insignificant in the composite model, they may play a more significant role in the variance model since their coefficients get squared when the variance operator is applied to the composite model. Another equally important consideration is the need for a more precise method for assessing the quality of the fitted composite model.

With different overlapping optimization criteria, there exists a plethora of possible hybrid approaches using combinations of the different strategies. One such approach is the use of the variance model to locate the best operability subregion. With information regarding the location of least sensitivity, MSE modeling could then be undertaken with the experimenter having some idea regarding the range of the variance within the present experimental subregion of interest.

Another possible hybrid approach is to make use of the mean modeling approach with the nonstationary simulation model to find an estimate of the best possible mean response level. Using this as the target value τ, the "target-is-best" MSE modeling approach discussed in Chapter 3 could then be employed. This approach could quite possibly put more emphasis on the mean response.

5.3 EXTENSION TO THE RANDOM SIGNAL FACTORS PROBLEM

In this section we examine a second possible source of model validity problems, the random behavior of the input signal factors. This characterization of the signal factors manifests itself in terms of a lack of complete control over the setting of the levels of the input factors in the real-world system. Box (1963) considers the problem of errors in setting factor levels in designed experiments noting that random differences between actual signal variable settings and their design-specified levels occur when the signal factors are random, and hence not completely controllable. He further notes that the resulting random discrepancies account for a considerable portion of total system performance variation.

Kackar (1985) and Leon, Shoemaker, and Kackar (1987) cite possible causes for the randomness in the signal variables. These include human operator error and job irregularities, instrument calibration problems due to manufacturing imperfections, and deviations in manufacturing component characteristics from their nominal values. Box

(1963, and in Nair (1992)) termed this third cause an "error transmission problem" wherein component imperfections are transmitted as system response variation. Myers, Khuri, and Vining (1992) note that similar problems, often referred to as "errors in control", occur in the mechanical and chemical processing fields.

The issue of model validity arises when a simulation model of the actual system whose signal factors are not completely controllable is formulated. If the discrepant levels of the signal variables in the real-world system are known or can be accurately observed, this information can be incorporated into the formulation of the simulation model and exact analytical tools can be utilized for the input-output analysis of simulation results. However, due to the frequent randomness of all or some of the signal factors, this is often not possible.

Consider the situation where the experimenter decides to include the stochastic behavior of the signal factors into the simulation model thereby providing a more realistic representation of the real-world system. This scenario is similar to the situation described by Box (1963) regarding errors in control of the design variable settings in designed experiments. He notes that serious complications in the input-output analysis of the experimental results arise when the random characterization applies not only to the output response but to the input design factors as well. Box(1963), Madansky (1959), and Berkson (1950) present and analyze some of the complications that could result for the analysis of simple linear metamodels of such situations.

In most real-world experiments the input-output analysis is usually performed under the assumption that the designed or intended signal input levels were actually utilized. This creates natural incompatibilities between the input and output data. The assumption of complete controllability of the signal input factors is easily accommodated in simulation modeling and experimentation, since the modeler has complete control over all aspects of the simulation model, including the input factors. With the experimenter exercising complete control over these variables, the input-output analysis follows a more conventional path without the complications brought about by the incompatibilities between the inputs and the output. The issue of model validity, however, does exist in this setting because of the variability in the input factors for the real system.

A job-shop simulation scenario is presented in the next section to illustrate some possible ramifications on model validity in these situations.

5.3.1 JOB-SHOP EXAMPLE

Consider the following job-shop example originally studied by Nozari, Arnold, and Pegden (1987) shown in Figure 1. In this example jobs arrive into the system according to a Poisson process with an arrival rate of 10 jobs per hour. Arriving jobs initially undergo an automated assembly operation at station 1. Upon completion of that operation, 80% of the jobs are routed to station 2 for minor adjustments, 15% to station 3 for major adjustments, and the remaining 5% exit the system as they do not require any rework. After the machine-made adjustments in either stations 1 or 2 are finished, the completed jobs likewise exit the system.

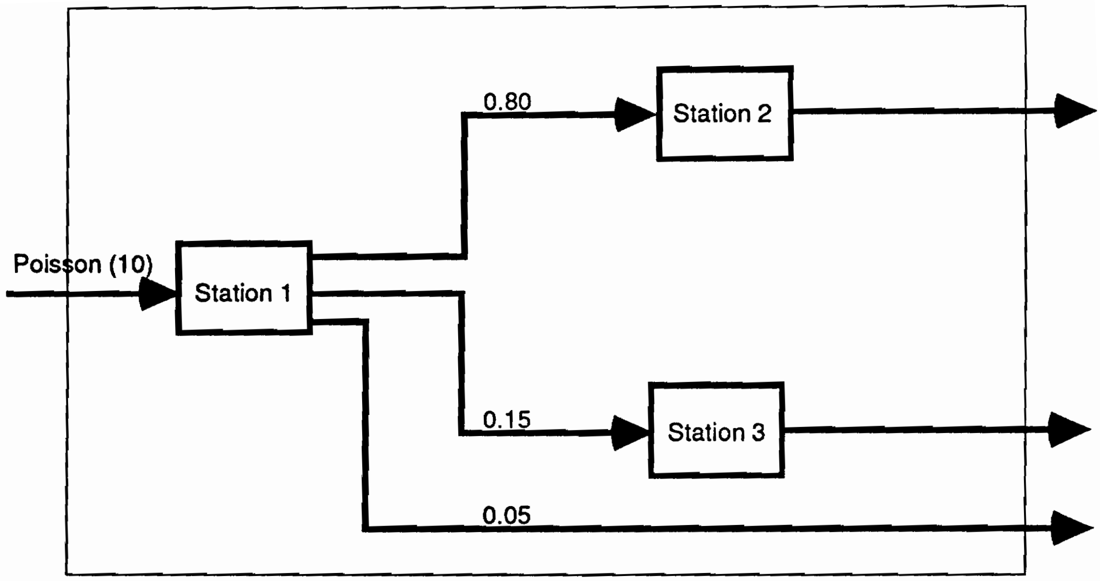


Figure 5.1 JOB-SHOP EXAMPLE

The shop starts admitting and processing jobs at 8:00 a.m. and closes its doors at 4:00 p.m. each day, continuing its processing of all admitted jobs prior to 4:00 p.m. until all are completed. Service time at station 1 is assumed to be constant, while the service times at stations 2 and 3, which depend on the type of equipment in use at those stations, are assumed to be randomly distributed according to different uniform distributions. The service time distributions given the current existing set of machines at the three stations is given in Table 5.10.

Table 5.5 Service Time Distributions

Station	Service Time Distribution (in Minutes)
1	Constant (3.6)
2	Uniform (2.4, 3.6)
3	Uniform (18, 30)

It is also assumed that the equipment in each of the stations can be independently updated or replaced to change the corresponding average service times. The system performance measure of interest is the daily average job velocity, defined as the reciprocal of average sojourn time for all jobs entering the system in a working day. The purpose of the study by Nozari et al. (1987) was to study the effect on daily average job velocity of proposed changes in equipment from the existing ones in the three stations. To do this, they considered analyzing the effects of three qualitative variables (equipment in use at the 3 stations) at two levels (existing and proposed) on average daily job velocity.

Certain comments are in order at this juncture. Another way of defining the problem and its input factors is by considering service times at the three stations to be the input factors. This would result in the use of three quantitative input factors. Treatment of service time distributions as qualitative factors in the earlier analysis stresses the consequences of the random behavior of the service time variables --- they preclude a modeling strategy using quantitative input factors. That is, levels for random quantitative variables can not be controlled and thus can not be fixed at specified levels in a designed experimental scenario. The inability to use quantitative factors, in turn, has serious implications on the optimization process, especially when there are numerous (if not an infinite number of) alternative system configurations.

5.3.2 PROBLEM RESOLUTION

As stated earlier, the use of categorical/qualitative variables as input factors enables the modeling of random input (signal) factors but removes these signal factors from consideration during the optimization process. We resolve this dilemma by considering the effect of the random or uncontrollable behavior of the signal input variables on the actual factor settings in a simulation experiment wherein the signal factors are allowed to vary randomly. The random deviations from the selected or intended design levels are viewed as "errors in control" of the signal factors. Following the recommendations of Myers, Khuri, and Vining (1992), noise variables can be used to reflect such errors. We do so by partitioning an originally random signal input variable into a sum of a nonrandom variable set equal to the intended design level (our new signal variable), and a random noise variable representing the error in control or random deviation from the intended level. In the job-shop example, if we let ξ_i be the average service time at station i (i=1,2,3) and let ζ_i be equal to the actual deviation from the average service time, then the sum $\xi_i+\zeta_i$ is the actual service time. The optimization problem can then be redefined in terms of the new signal factors (the ξ_i), which can be completely controlled, and the noise variables (the ζ_i) which are random. In the job-shop example, the average service times at the three stations can be independently and accurately set by updating or replacing the equipment at the stations. The optimization process would then be undertaken using the methods dealing with noise variables discussed in the earlier chapters. For example, using the dual response (composite) model approach, we can define the composite model in terms of the coded variables as

$$
\begin{aligned}
y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \\
& + \gamma_2 z_2 + \gamma_3 z_3 + \lambda_{12} x_1 z_2 + \lambda_{13} x_1 z_3 + \lambda_{22} x_2 z_2 + \lambda_{23} x_2 z_{23} \\
& + \lambda_{32} x_3 z_2 + \lambda_{33} x_3 z_3 + \varepsilon.
\end{aligned}
$$

The uncoded noise variables and their distributions would then be

$$\zeta_2 \sim \text{Uniform } (-0.6, 0.6) \text{ and } \zeta_3 \sim \text{Uniform } (-6, 6).$$

The natural variables corresponding to the signal variables are assumed to be able to take on a range of values. Essentially, these values can be fixed or controlled with negligible error. The same coding transformations used in the previous chapters relate the natural variables, the ξ_i and ζ_i, to their coded versions, the x_i and z_i, respectively.

5.4 CONCLUDING COMMENTS

The discussions in the previous chapters highlight the benefits of having variance as an active player in the optimization process. Variance, or sensitivity, should not be considered as just an ex post facto consideration in system simulation optimization. There are, however, limits to the successful application of these optimization methods that feature robustness to the effects of the noise variables. As Kackar (in Nair (1992)) correctly points out, it is essential that certain noise by signal interactions exist, and that these interactions are identifiable and estimable.

Furthermore, these methods are empirical in nature. Closed-form solutions, if available, will always be preferable. But until such solutions are available, these empirical techniques should provide adequate alternatives.

BIBLIOGRAPHY

1 Balci, O.(1987),"Credibility Assessment of Simulation Results: The State of the Art," *Proc, Conference on Methodology and Validation*, Orlando, Fla., pp.19-25.

2 Balci, O. and R. G. Sargent (1984),"A Bibliography on the Credibility Assessment and Validation of Simulation and Mathemetical Models, *Simuletter*, 15: 15-27.

3. Banks, J. and J. S. Carson (1984), *Discrete-Event System Simulation*, New Jersey: Prentice-Hall.

4. Bartlett, M. S. and D. G. Kendall (1946), "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society,* Ser. B, 8, 128-150.

5. Bechhofer, R.E. (1960),"A multiplicative Model for Analyzing Variances Which are Affected by Several Factors," *Journal of the American Statistical Association* 55, pp.245-264.

6. Berkson, J. (1950), "Are There Two Regressions ?," *Journal of the American Statistical Association*, 45, 164-180.

7. Box,G. E. P. (1963), "The Effects of Errors in the Factor Levels and Experimental Design," *Technometrics*, Vol.5, No.2, pp.247-262.

8. Box, G. E. P. (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology*, 17, 189-190.

9. Box, G. E. P. (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations," *Technometrics*, 30, 1-17.

10. Box, G.E.P. and R.D. Meyer (1986),"Dispersion Effects from Fractional Designs," *Technometrics* 28, pp.19-27.

11. Davies, O. L.(1956), *Design and Analysis of Industrial Experiments*, 2nd ed., New York:Hafner Publishing Company, Inc.

12. Draper, N. R. (1963),"Ridge Analysis of Response Surfaces," *Technometrics*, *5*(4), 469.

13. Easterling, R. B. (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology,* 17, 191-192.

14. Easterling, R. B. and J. S. Ramberg (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology,* 17, 198-206.

15. Feltner, C. E. and S.A. Weiner (1985)."Models, Myths and Mysteries in Manufacturing," *Ind. Eng.*, 17:66-76.

16. Gass, S. I. (1983),"Decision-Aiding Models: Validation, Assessment, and Related Issues in Policy Analysis," *Operation Res.*, 31:603-631.

17. Gass, S.I and B.W. Thompson (1980),"Guidelines for Model Evaluation: An Abridged Version of the U.S. General Accounting Office Exposure Draft," *Operations Res.*, 28:431-439.

18. Gaver, D. P., J. P. Lehoczky, and M P. Perlas (1975), "Service Systems with Transitory Demands," Chapter in *Logistics*, Vol. 1 of North Holland/TIMS Studies in Management Science, North Holland/American Elsevier, New York and Amsterdam.

19. Graybill, F. A. (1976), *Theory and Application of the Linear Model*, California:Wadsworth & Brooks/Cole.

20. Hadley, G. and T. M.Whitin (1963), *Analysis of Inventory Systems*, New Jersey: Prentice-Hall, Inc.

21. Hillestad, R. J. and Carillo, M. J. (1980),"Models and Techniques for Trcoverable Item Stockage When Demands and the Repair Process are Nonstationary -- Part I: Performance Measurement," The Rand Corporation, N-1482-AF.

22. Hoerl, A. E.(1959),"Optimum Solution of Many Variables," *Chem. Eng. Prog.* 55(11),69.

23. Hoerl, A. E. (1964).Ridge Analysis.*Chemical Engineering progres*, Symposium Series, *60*, 76-77.

24. Hunter, J. S. (1985), "Statistical Design Applied to Product Design," *Journal of Quality Technology,* 17, 210-221.

25. Kackar, R. (1985), "Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology*, 17, 176-188.

26. Kendall, M.A. and A. Stuart (1966), *The Advanced Theory of Statistics.* Vol.3, New York: Hafner Publishing Co.

27. Khuri, A. I. and J.A. Cornell (1987), *Response Surfaces: Designs and Analyses.* New York: Marcel-Dekker.

28. Law, A. M.,and W. D. Kelton (1991), *Simulation Modeling and Analysis,* New York: McGraw-Hill, Inc.

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

30. Lucas, J. M. (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology,* 17, 195-197.

31. Luenberger, D. G. (1984)*Linear and Nonlinear Programming*,2nd ed.,Massachusetts:Addison-Wesley.

32. Madansky, A. (1959), "The Fitting of Straight Lines When Both Variables Are Subject to Error," *Journal of the American Statistical Association*, 54, 173-206.

33. Miller, L. W., R. E.Stanton, and G. B. Crawford (1984),"DYNA-SIM: A Nonstationary Queuing Simulation With Application to the Automated Test Equipment Problem," The Rand Corporation, N-2087-AF.

34. Myers, R. H. (1976), *Response Surface Methodology*, Blacksburg, VA: Author.

35. Myers, R. H (1986), *Classical and Modern Regression with Applications*, Boston: Duxbury Press.

36. Myers, R. H. and W. H. Carter (1973),"Response Surface Techniques for Dual Response Systems," *Technometrics*, 15:2, 301-317.

37. Myers, R. H., A. I, Khuri, and G. Vining (1992), "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," *The American Statistician,* 46, 131-139.

38. Myers, R. H. and Y. Kim (1993), "Response Surface Methods and the Use of Noise Variables," Technical Report, Department of Statistics, Virgia Polytechnic Institute and State University, Blacksburg, Va. 24061.

39. Nair, V.N. and Pregibon, D. (1988),"Analyzing Dispersion Effects from Replicated Factorial Experiments," *Technometrics* 30, pp.247-257.

40. Nair, V. N. et al.(1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 127-161.

41. Naylor, T. H. and J. M. Finger (1967),"Verification of Computer Simulation Models,"*Management Sci.*, 14: 92-101.

42. Nozari, A., S. F. Arnold, and C. D. Pegden (1987), "Statistical Analysis for Use with the Schruben and Margolin Correlation Induction Strategy," *Operations Research*, 35, 127-139.

43. Oren, T. I. (1981),"Concepts and Criteria to Assess Acceptability of Simulation Studies: A FRame of Reference,"*Communications of the Association of Computing Machinery*, 24: 180-189.

44. Pignatiello, J. J. (1988), "An Overview of the Strategy and Tactics of Taguchi," *IIE Transactions,* 20(3), 247-254.

45. Pignatiello, J. J. and J. S.Ramberg (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology,* 17, 198-206.

46. Plackett, R. L. and J. P. Burman (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 305-325.

47. Pritsker, A. A. B. (1986), *Introduction to Simulation and SLAM II*, 3rd ed., New York: John Wiley & Sons.

48. Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, New York: Wiley.

49. Safizadeh, M. H. (1990), "Optimization in Simulation: Current Issues and the Future Outlook," *Naval Research Logistics,* 37, 807-825.

50. Safizadeh, M. H. and B. M. Thornton (1984),"Optimization in Simulation Experiments Using Response Surface Methodology," *Comput. & Indus. Engng.*, Vol.8, No.1, pp.11-27.

51. Sargent, R. G. (1980),"Verification and Validation of Simulation Models,"*Progress in Modelling and Simulation*," ed. by F.E.Cellier, London: Academic Press.

52. Sargent, R.G. (1988), "A Tutorial on Validation and Verification of Simulation Models,"*Proc. 1988 Winter Simulation Conference*, San Diego, Calif., pp. 33-39.

53. Schellenberger, R. E.(1974),"Criteria for Assessing Model Validity for Managerial Purposes,"*Decision Sci.*, 5:644-653.

54. Schmidt, J. W. and R. E. Taylor (1970), *Simulation and Analysis of Industrial Systems,* Illinois: Richard R. Irwin, Inc.

55 Schruben, L. W. (1980), "Establishing the Credibility of Simulations,"*Simulation*, 34: 101-105.

56. Shannon, R. E. (1975), *System Simulation: The Art and Science*, New Jersey: Prentice-Hall.

57. Shoemaker, A. C., Tsui, K. L., and C. F. J. Wu (1991),"Economical Experimentation Methods for Robust Design," *Technometrics*, 33, 415-427.

58. Taguchi, G. (1978), "Off-Line and On-Line Quality Control Systems," *Proceedings of International Conference on Quality Control*, Tokyo, Japan.

59. Taguchi, G. and Y.Wu (1980), "Introduction to Off-line Quality Control," Central Japan Quality Control Association (available from American Supplier Institute, 32100 Detroit Industrial Expressway, Romulus, MI 48174).

60. Van Horn, R. L. (1971),"Validation of Simulation Results," *Management Sci.*, 17: 247-258.

61. Vining, G. G. and R.H. Myers (1990), "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach," *Journal of Quality Technology*, 22, 38-45.

62. Welch, W. J., T. K.Yu, S. M. Kang, and J. Sacks (1990), "Computer Experiments for Quality Control by Parameter design," *Journal of Quality Technology*, 22, 15-22.

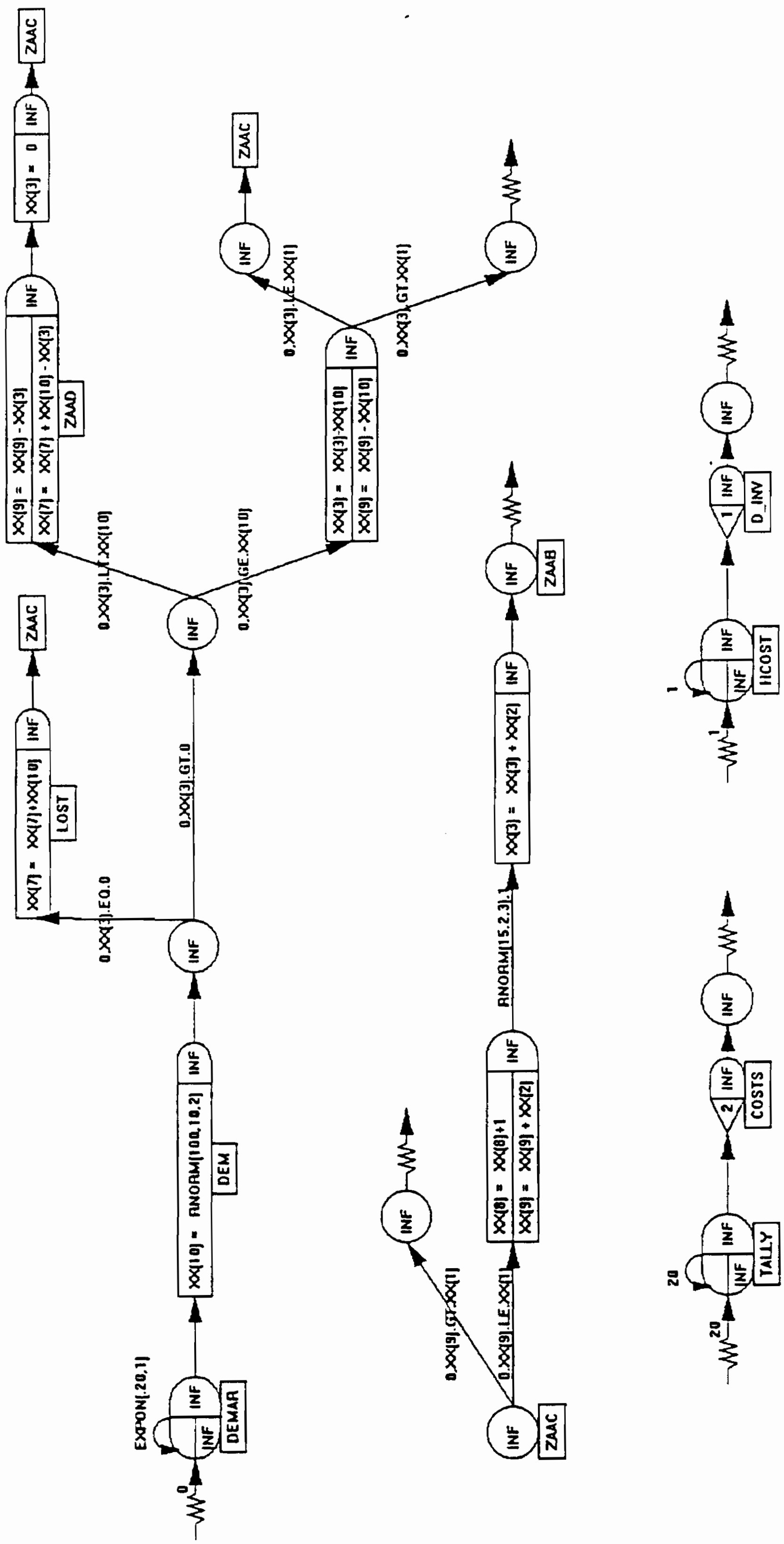
APPENDICES

APPENDIX A

SLAM CODE AND FORTRAN USER FUNCTIONS FOR THE STATIONARY SIMULATION MODEL

```
//Rxx JOB 445B8,RENEP,MSGLEVEL=(2,0),TIME=30,REGION=3M
/*PRIORITY IDLE
/*ROUTE PRINT VTVM1.RPANIS
/*JOBPARM LINES=5
//STEP1 EXEC SLAMCG
//FORT.SYSIN DD *
       SUBROUTINE EVENT(I)
       INCLUDE (SLMSCOM1)
       GO TO (1,2) I
    1  XX(6) = XX(6) + ( XX(4) + XX(3) ) /2
       XX(4) = XX(3)
       RETURN
    2  HOLD_COST  = XX(6) * .10
       STOCK_OUT  = XX(7) * 10.00
       ORDER_CST  = XX(8) * 1000.00
       TOTAL_CST  = HOLD_COST + STOCK_OUT + ORDER_CST
       IF (TNOW.LE.10001.) GO TO 6
       WRITE(10,5) TOTAL_CST,XX(1),XX(2)
    5  FORMAT(3F12.2)
    6  XX(6)=0.
       XX(7)=0.
       XX(8)=0.
       RETURN
       END
//GO.SYSIN DD *
GEN,RENE,TEST CASE,02/30/1994,1,Y,Y,Y/Y,Y,Y/1,132;
LIMITS,,1,100;
SEEDS,4957461(1),4918271(2),4819861(3);
INTLC,XX(1)=** X1 **,XX(2)=** X2 **,XX(3)=20000,XX(4)=20000,XX(5)=0;
INTLC,XX(6)=0,XX(7)=0,XX(8)=0,XX(9)=20000,XX(10)=0;
NETWORK;
;
DEMAR CREATE,EXPON(** Z1 **,1);
      ACTIVITY;
DEM    ASSIGN,XX(10)=RNORM(** Z2 **,10,2);
```

```
       ACTIVITY;
       GOON;
       ACTIVITY,,XX(3).GT.0;
       ACTIVITY,,XX(3).EQ.0,LOST;
       GOON;
       ACTIVITY,,XX(3).GE.XX(10);
       ACTIVITY,,XX(3).LT.XX(10),ZAAD;
       ASSIGN,XX(3)=XX(3)-XX(10),XX(9)=XX(9) - XX(10);
       ACTIVITY,,XX(3).GT.XX(1);
       ACTIVITY,,XX(3).LE.XX(1),ZAAC;
       TERMINATE;
ZAAC   GOON;
       ACTIVITY,,XX(9).LE.XX(1);
       ACTIVITY,,XX(9).GT.XX(1),ZAAB;
       ASSIGN,XX(8)=XX(8)+1,XX(9)=XX(9) + XX(2);
        ACTIVITY,RNORM(** Z3 **,2,3);
       ASSIGN,XX(3)=XX(3) + XX(2);
       ACTIVITY;
       TERMINATE;
ZAAB   TERMINATE;
ZAAD   ASSIGN,XX(9)=XX(9) - XX(3),XX(7)=XX(7) + XX(10) - XX(3);
       ACTIVITY;
       ASSIGN,XX(3)=0;
       ACTIVITY,,,ZAAC;
LOST   ASSIGN,XX(7)=XX(7)+XX(10);
       ACTIVITY,,,ZAAC;
;
TALLY CREATE,20,20;
       ACTIVITY;
COSTS EVENT,2;
       ACTIVITY;
       TERMINATE;
;
HCOST CREATE,1,1;
       ACTIVITY;
D_INV EVENT,1;
       ACTIVITY;
       TERMINATE;
       ENDNETWORK;
INITIALIZE,,50000,Y;
SIMULATE;
FIN;
/*
//FT10F001 DD SYSOUT=C,DCB=LRECL=132
/*
//
```



SLAM NETWORK - STATIC MODEL

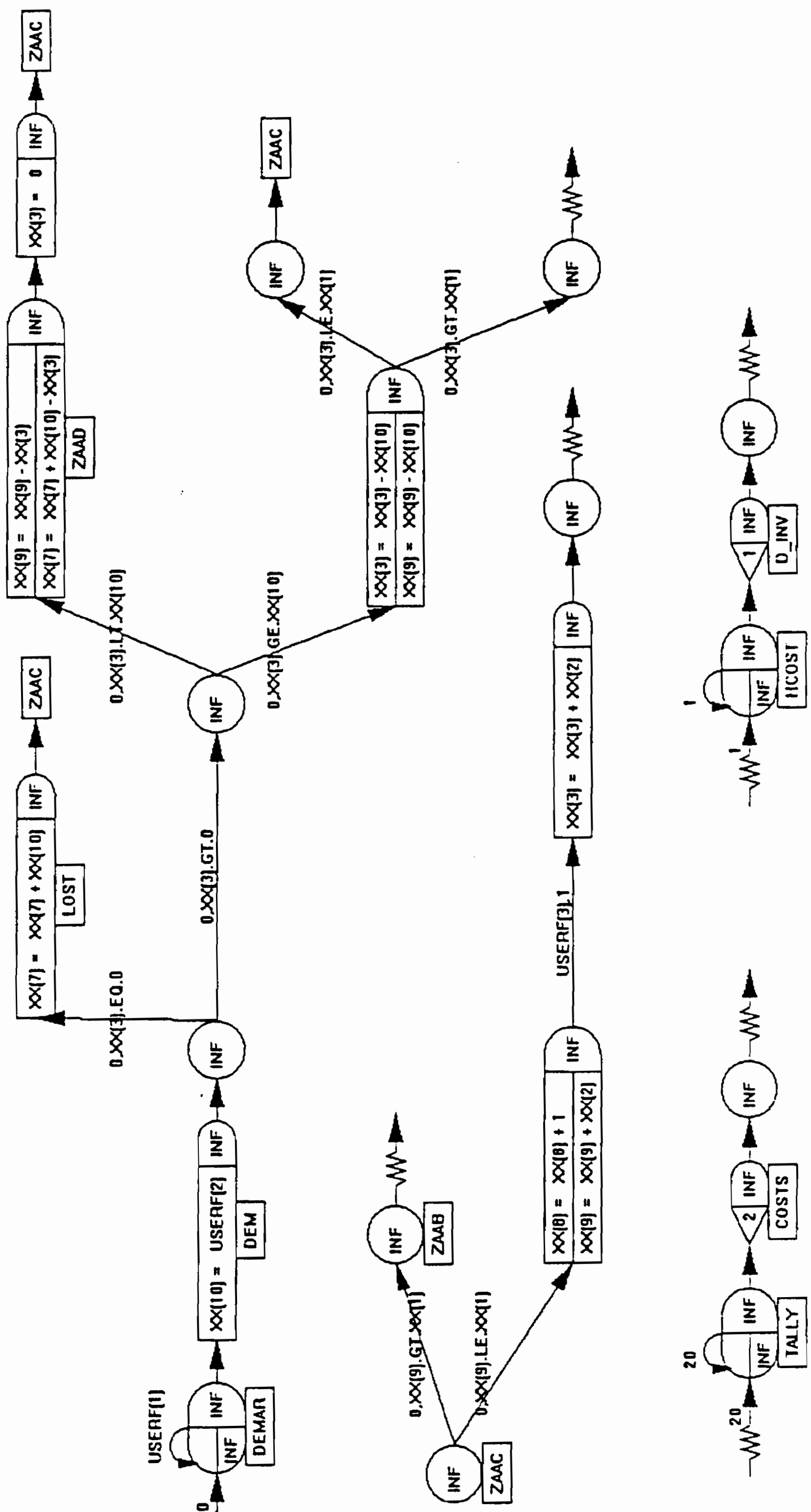
APPENDIX B

SLAM CODE AND FORTRAN USER FUNCTIONS FOR THE NONSTATIONARY SIMULATION MODEL

```
//Rxx JOB 445B8,RENEP,MSGLEVEL=(2,0),TIME=30,REGION=3M
/*PRIORITY URGENT
/*ROUTE PRINT VTVM1.RPANIS
/*JOBPARM LINES=5
//STEP1 EXEC SLAMCG
//FORT.SYSIN DD *
       SUBROUTINE EVENT(I)
       INCLUDE (SLMSCOM1)
       GO TO (1,2) I
    1  XX(6) = XX(6) + ( XX(4) + XX(3) ) /2
       XX(4) = XX(3)
       RETURN
    2  HOLD_COST  = XX(6) * .10
       STOCK_OUT  = XX(7) * 10.00
       ORDER_CST  = XX(8) * 1000.00
       TOTAL_CST  = HOLD_COST + STOCK_OUT + ORDER_CST
       IF (TNOW.LE.10001.) GO TO 6
       WRITE(10,5) TOTAL_CST,XX(1),XX(2)
    5  FORMAT(3F12.2)
    6  XX(6)=0.
       XX(7)=0.
       XX(8)=0.
       RETURN
       END
       FUNCTION USERF(IFN)
       INCLUDE (SLMINIT)
       GO TO (1,2,3) IFN
    1  Z1=RNORM(.20,.05,1)
       IF (Z1.LT.0) Z1=0
       USERF=EXPON(Z1,2)
       RETURN
    2  Z2=RNORM(100.0, 10.0, 3)
       IF (Z2.LT.0) Z2=0
       DEMAND=RNORM(Z2,10.0,4)
       IF (DEMAND.LT.0.) DEMAND=0.0
       USERF=DEMAND
       RETURN
    3  Z3=RNORM(15.0, 2.0, 5)
       IF (Z3.LT.0) Z3=0
       LTIME=RNORM(Z3,2.0,6)
       IF (LTIME.LT.0.0) LTIME=0.0
       USERF=LTIME
       RETURN
       END
```

```
//GO.SYSIN DD *
GEN,RENE,TEST CASE,02/31/1994,1,Y,Y,Y/Y,Y,Y/1,132;
LIMITS,,1,100;
SEEDS,651741(1),817091(2),573731(3);
SEEDS,913761(4),752151(5),645361(6);
INTLC,XX(1)=** X1 **,XX(2)=** X2 **,XX(3)=20000,XX(4)=20000,XX(5)=0;
INTLC,XX(6)=0,XX(7)=0,XX(8)=0,XX(9)=20000,XX(10)=0;
NETWORK;
;
DEMAR CREATE,USERF(1);
      ACTIVITY;
DEM   ASSIGN,XX(10)=USERF(2);
      ACTIVITY;
      GOON;
      ACTIVITY,,XX(3).GT.0;
      ACTIVITY,,XX(3).EQ.0,LOST;
      GOON;
      ACTIVITY,,XX(3).GE.XX(10);
      ACTIVITY,,XX(3).LT.XX(10),ZAAD;
      ASSIGN,XX(3)=XX(3) - XX(10),XX(9)=XX(9) - XX(10);
      ACTIVITY,,XX(3).GT.XX(1);
      ACTIVITY,,XX(3).LE.XX(1),ZAAC;
      TERMINATE;
ZAAC  GOON;
      ACTIVITY,,XX(9).LE.XX(1);
      ACTIVITY,,XX(9).GT.XX(1),ZAAB;
      ASSIGN,XX(8)=XX(8) + 1,XX(9)=XX(9) + XX(2);
      ACTIVITY,USERF(3);
      ASSIGN,XX(3)=XX(3) + XX(2);
      ACTIVITY;
      TERMINATE;
ZAAB  TERMINATE;
ZAAD  ASSIGN,XX(9)=XX(9) - XX(3),XX(7)=XX(7) + XX(10) - XX(3);
      ACTIVITY;
      ASSIGN,XX(3)=0;
      ACTIVITY,,,ZAAC;
LOST  ASSIGN,XX(7)=XX(7) + XX(10);
      ACTIVITY,,,ZAAC;
;
      TERMINATE;
;
TALLY CREATE,20,20;
      ACTIVITY;
COSTS EVENT,2;
      ACTIVITY;
      TERMINATE;
;
HCOST CREATE,1,1;
      ACTIVITY;
D_INV EVENT,1;
      ACTIVITY;
      TERMINATE;
      ENDNETWORK;
INITIALIZE,,50000,Y;
SIMULATE;
```

```
FIN;
/*
//FT10F001 DD SYSOUT=C,DCB=LRECL=132
/*
//
```



SLAM NETWORK - NONSTATIONARY MODEL

VITA

The author was born in San Fernando, the Philippines on May 9, 1960. He graduated from the University of the Philippines (UP) with a Bachelor of Science degree in Mathematics in 1981. After working for a year as a programmer/systems analyst with the Atlantic Gulf and Pacific company, the author entered the M.S. program in Mathematics in UP in 1982, working concurrently as a lecturer in the UP Mathematics department. He graduated with a Master's degree in Mathematics in 1984.

The author started employment as a systems consultant and partner of S.I. Consultants, Inc. in 1984. Two years later, he entered the graduate program in Management Science at Virginia Tech, majoring in Operations Research and Statistics. He worked as a graduate teaching assistant and lecturer in the Management Science department at Virginia Tech from September 1986 to May 1994. He graduated with a Doctor of Philosophy degree in Management Science in August 1994.

The author is presently employed as a consultant at Sabre (American Airlines) Decision Technologies.