1

INTRODUCTION

This introductory chapter defines the goal of the book. It discusses the nature of the problem and the proposed approach to its solution and briefly describes the related methods in existing literature. The chapter also serves as a guide to the book by providing a brief description of the various chapters and their relation with each other.

1.1 THE GOAL

Swami Vivekananda, a great Indian saint, emphasized the importance of a goal with the following quote:

Awake, arise and stop not till the goal is reached.

This aptly applies to writing a book, where one has to define a suitable goal and work hard to accomplish it irrespective of the difficulties faced. Therefore, we begin by clearly stating the goal of this book:

The goal of this book is to help decision makers who use multidimensional systems to make robust decisions by recognizing *various patterns with an appropriate measurement scale and to use simple procedures to optimize the system.*

The decision makers could be doctors, managers, academicians, executives from service organizations, such as banks and insurance companies, or any other type of pattern recognizers. The robust decisions made with the measurement scale help lower the cost of diagnosis and minimize incidents of false diagnosis and diagnosis time.

A *multidimensional system* may be defined as a decision-making (diagnostic or pattern-recognition) system based on the numerous measurements. A multidimensional system could be an inspection system, a medical diagnosis system, a sensor system, a face recognition system, a voice recognition system, or a university admission system. Because multidimensional systems are used in day-to-day life, the proposed methods can find applications in several areas. In this book, case studies and examples related to different types of multidimensional systems are presented.

The main goal can be divided into the following subgoals, which are based on the problems encountered by decision makers in different fields:

- To introduce a measurement scale based on the input characteristics to measure the degree of unhealthiness or abnormality of different conditions
- To quantify functionality of a system with a suitable measure
- To minimize the number of the variables required (in terms of original variables) for an effective diagnosis
- To predict the performance of a multidimensional system under various conditions
- To establish different zones of treatment of a product or patient based on severity and cost so that the decision maker can take appropriate actions
- To identify the direction of abnormal conditions
- To overcome commonly encountered multivariate problems, such as multicollinearity (high correlations) and low correlations

- To demonstrate applicability of proposed methods with actual case studies

This book also compares the proposed methods with classical multivariate methods and artificial neural networks.

1.2 THE NATURE OF A MULTIDIMENSIONAL SYSTEM

To develop a suitable measurement scale for the purpose of diagnosis, it is important to understand the nature of the system in terms of the variables controlling the system. It is also necessary to know the noise conditions under which the diagnosis process is to be performed. Since all the variables may not be necessary for the diagnosis process, it is important to identify the useful set of variables, which is a subset of the original variables. The future diagnosis is performed with the variables in the useful set. While conducting the diagnosis, the decision maker may also be interested in defining a set of actions to be taken for different abnormal conditions.

1.2.1 Description of Multidimensional Systems

A typical multidimensional system used in this research is shown in Figure 1.1. In this figure, X_1, X_2, ..., X_k correspond to k variables (dimensions), which provide information that can be used to make a decision. A correct decision has to be made about the state of the system regardless of noise conditions.

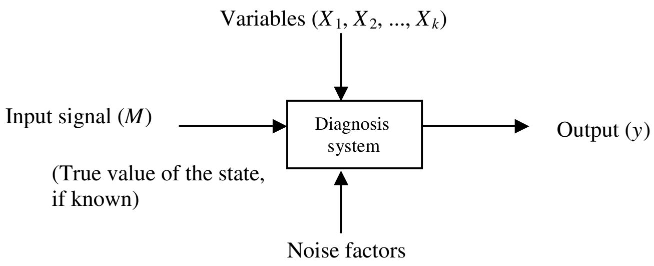

Figure 1.1 Multidimensional diagnosis system.

In Figure 1.1, the input signal M is the true value of the state of the system, if known (for example, a rainfall prediction system where rainfall can be measured and recorded). The noise conditions are the changes in the usage environment, such as conditions in different places and manufacturing variability of the system. The output y should match closely with the input. In future based on the output, a decision has to be made about the state of the system, and accordingly corrective actions have to be taken. In multivariate systems, a correct decision cannot be made by looking at all the variables independently because of the presence of correlations. It is important to take the correlations between the variables into account.

1.2.2 Correlations between the Variables

A measure of linear association between two variables is usually provided by the covariance between them. Covariance between two variables depends on the units of measurement. The correlation coefficient is a standardized version of the covariance and it does not depend on the units of measurements. The patterns of observations in multidimensional systems highly depend on the correlation structure of the variables in the system. One can make wrong decisions about the patterns if each variable is looked at separately without considering the correlation structure.

To construct a multidimensional measurement scale, it is important to have a *distance measure,* which is based on correlations between the variables and by which different patterns could be identified and analyzed with respect to base or reference point. Fortunately, there exists one such measure, called *Mahalanobis distance.* The Mahalanobis distance was introduced in 1936 by P. C. Mahalanobis, a famous Indian statistician and the founder of the Indian Statistical Institute.

1.2.3 Mahalanobis Distance

The Mahalanobis distance (MD) is a generalized distance, which can be considered a single measure of the degree of divergence in the mean values of different characteristics of a population by considering the correlations between the variables. The Mahalanobis distance (Mahalanobis 1936) is a very useful way of determining the similarity of a set of values from an unknown sample to a set of values measured from a collection of known samples. This method has been applied successfully for spectral discrimination in a number of cases. One of the main reasons for using MD is that it is very sensitive to intervariable changes in the reference data. MD is superior to other multidimensional distances, such as Euclidean distance, because it takes distribution of the points (correlations) into account. Traditionally, the Mahalanobis distance is used to classify observations into different groups.

In our approach, the Mahalanobis distance is modified by suitable scaling and is used first to define a base or reference point of the scale with a set of observations from a reference group. The average of the scaled distance in the reference group converges to unity, because of the properties of the scaled distance. Since the reference group has average unit distance, the reference group is also known as a *unit group.* Because the reference group contains scaled Mahalanobis distance, the group is sometimes referred to as *Mahalanobis space* (MS). Mahalanobis space is a database containing the means, standard deviations, and correlation structure of the variables in the reference group. The scaled Mahalanobis distance is also used to measure the distances of unknown observations from the reference point. Defining a reference group or Mahalanobis space and selection of variables for constructing such group depend entirely on the decision maker's discretion.

Since the definition of Mahalanobis space (unit group) is very important in our approach and is based on the information about variables, Figure 1.1 can be modified as shown in Figure 1.2. This figure forms a basis for all the discussions in this book.

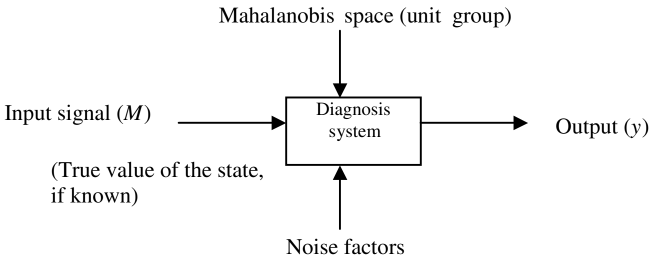

Figure 1.2 Modified multidimensional diagnosis system.

The Mahalanobis distance (and hence the scaled distance) can be computed in two ways: (1) using the inverse of the correlation matrix, and (2) using the Gram–Schmidt orthogonalization process. The advantages of the Gram–Schmidt process are clearly spelled out in this book. A detailed discussion on the Mahalanobis distance and scaled distance is presented in Chapter 2.

1.2.4 Robust Engineering/Taguchi Methods

The concepts of *robust engineering* (RE) are based on the philosophy of Genichi Taguchi, who introduced the concepts after several years of research. Robust engineering systematically evolved starting in the 1950s and aims at providing industries with a cost-effective methodology for enhancing their competitive position in the global market. These concepts are also referred to as *Taguchi Methods.*

In Taguchi Methods, there are two types of quality: (1) customer-driven quality and (2) engineered quality. Customer quality leads to the size of the market segment and includes product features such as color, size, and appearance. The market size becomes bigger as customer quality gets better. Customer quality is addressed during the product planning stage and is extremely important in creating a new market. *Engineered quality* includes the defects, failures, noise, vibrations, pollution, etc. Engineered quality can be measured in terms of deviations from ideal performance (function). While the customer quality defines the market size, the engineered quality helps in winning the market share within the segment. Robust engineering aims at improving the engineered quality. In multivariate applications, *scaled Mahalanobis distance* is similar to engineered quality, because it measures the degree of abnormality (function) of observations from the known reference group (Mahalanobis space).

Since, in this book, the Taguchi Methods are combined with the scaled MD, a brief description of the Taguchi Methods is provided here. Taguchi Methods are based on five principles.

1. Measurement of function using energy transformation
2. Taking advantage of interactions between control and noise factors
3. Use of orthogonal arrays and signal-to-noise ratios

4. Two-step optimization
5. Tolerance design using quality loss function and on-line quality engineering

Taguchi (1987), Phadke (1989), and Park (1996) provide a detailed discussion on the principles of TM. These principles are extremely useful and cost effective and have been successfully applied in many engineering applications to improve the performance of a product/process. A brief illustration of these principles is given below:

1. *Measurement of Function Using Energy Transformation*
   The most important aspect of Taguchi Methods (TM) is to find a suitable function (called an ideal function) that governs the system. Taguchi Methods state that "to improve quality, don't measure quality." It is important, instead, to measure the functionality of the system to improve the product performance (quality).
2. *Take Advantage of Interactions between Control and Noise Factors*
   In TM, we are not interested in measuring the interaction between the control factors. We are interested in the interaction between the control and noise factors, since the objective is to make design robust against the noise factors.
3. *Use of Orthogonal Arrays* (*OAs*) *and Signal-to-Noise* (*S/N*) *Ratios*
   OAs are used to minimize the number of runs (or combinations) needed for the experiment. Many people are of the opinion that the application of OA is TM, but the application of OAs is only a part of TM. *S/N* ratios are used as a measure of the functionality of the system. *S/N* ratios capture the magnitude of real effects (signals) after making some adjustment to uncontrollable variation (noise).
4. *Two-Step Optimization*
   After conducting the experiment, the factor–level combination for the optimal design is selected with the help of two-step optimization. The first step is to minimize the variability (maximize *S/N* ratios). In the second step, the sensitivity (mean) is adjusted to the desired level since it is easier to adjust the mean after minimizing the variability.

5. *Tolerance Design Using Quality Loss Function and On-Line Quality Engineering*
   While the first four principles are related to parameter design, the fifth principle is related to the tolerance design and on-line quality engineering (QE). Having determined the best settings using parameter design, the tolerancing is done with the help of quality loss function. If the performance deviates from the target, a loss is associated with the deviation, known as a loss to the society. This loss is proportional to the square of the deviation. It is recommended that safety factors be designed using this approach. On-line QE is used to monitor the process performance and detect the changes in the process.

1.3 MULTIVARIATE DIAGNOSIS—THE STATE OF THE ART

A significant body of literature exists on the concepts of multivariate diagnosis that are being used in various multidimensional systems. So, where is the need for another book on multivariate diagnosis? This section addresses this question by briefly reviewing some of the major tools used to analyze multidimensional systems and identifying the limitations of the available tools.

1.3.1 Principal Component Analysis

Principal component analysis (PCA) is used for explaining the variance–covariance structure through a fewer linear combinations of original variables. The objectives of PCA are (1) data reduction and (2) data interpretation. Although p components are required to reproduce the total system variability, often much of this variability can be explained by k ($k < p$) principal components. The principal components are particular linear combinations of the p random variables, X_1, X_2, X_3, ..., X_p. These linear combinations represent the selection of a new coordinate system obtained by rotating the original system with maximum variability to provide a simpler description of the covariance structure. Johnson and Wichern (1992) provide a clear discussion on PCA. To calculate the principal components, we need all original variables. Hence, PCA is not helpful in reducing the dimensionality in terms of original variables.

1.3.2 Discrimination and Classification Method

The discrimination and classification method is a multivariate technique concerned with separating the distinct sets of objects or observations and allocating new objects or observations to previously defined groups. Discriminant analysis is carried out using a discriminant function or Mahalanobis distance. The discriminant function is a part of Mahalanobis distance, which provides another rule for classification.

When there are k populations, in the first stage discriminant functions are developed for all the groups. It is assumed that all the groups have same covariance matrix. In the second stage, the classification of a new observation is done based on the following rule:

- Assign a new observation X to the group whose mean is closest to this observation (minimum Mahalanobis distance), or
- Assign a new observation X to the group that has the largest discriminant function.

The classification is done in such a way that the expected cost of misclassification (ECM) is minimized. ECM computations are based on the cost of misclassification and prior probabilities. Therefore, the method of discrimination and classification is probabilistic in nature. A good discussion on this method is provided in Johnson and Wichern (1992). Discrimination and classification methods have certain limitations. In these methods, the emphasis is given to classification of an observation in a group. These methods are less helpful for accurately measuring the level of severity of an abnormality in order to take appropriate corrective actions.

1.3.3 Stepwise Regression

Stepwise regression is widely used for selection of a useful subset of variables in multivariate applications. The procedure iteratively constructs a sequence of regression models by adding and removing variables at each step. Based on a specified value of the F-random variable, addition and deletion of the variables in the model is carried out. This method requires several iterations, if the number of variables is high. The method of stepwise regression has been criticized because it does not guarantee the best subset regression model. A discussion on stepwise regression is given in Montgomery and Peck (1982).

1.3.4 Test of Additional Information (Rao's Test)

The test of additional information, also known as Rao's test, is used to identify a set of useful variables. In Rao's test, which uses Fischer's linear discrimination function, subsets are tested for significance by computing an F-statistic. A high F-ratio indicates that the subset of q variables provides additional information on the discriminant analysis. If the F-ratio is not high, then we can discard the subset of variables. For this test procedure please refer to Rao (1973). In this method, selection of the subset of variables is carried out based on prior knowledge about variables or expert opinion. Moreover, testing the significance of variables using an F-test may not be adequate to decide the important variables. This fact is illustrated in Section 9.4.

1.3.5 Multiple Regression

In multiple regression (MR), the characteristic y (dependent variable) is estimated based on the p independent variables $X_1,X_2,\ldots,X_p$. Based on the value of y, a decision can be made regarding the classification of an observation X, which consists of $X_1,X_2,\ldots,X_p$. A discussion of multiple regression is provided in Montgomery and Peck (1982). MR models are developed using least-squares estimates and are based on certain assumptions about the error term. MR models may become complex if the number of variables is high.

1.3.6 Multivariate Process Control Charts

Multivariate process control charts are an extension of univariate control charts, where more than one variable is monitored and controlled over a period of time. Several types of these charts, such as multivariate Shewhart charts and multivariate Cusum charts, are being used. The purpose of these charts is to monitor and control the multivariate conditions. These charts operate just like univariate charts, in which corrective actions are taken whenever the process is out of the control limits. An extensive literature is available on multivariate control charts. In multivariate charts, generally, the variables $X_1,X_2,...,X_p$ are assumed to follow a p-dimensional normal distribution and therefore the control limits are probabilistic.

1.3.7 Artificial Neural Networks

Artificial neural networks (ANN) are used for pattern recognition, learning, classification, generalization, and interpretation of noisy inputs. A structure (network) is composed of interconnected units (artificial neurons). Each unit has an input/output (I/O) characteristic and implements a local computation or function. The output of any unit is determined by its I/O characteristic and its interconnection to other units and (possibly) external inputs. ANN constitutes not one network, but a diverse family of networks. ANN and the proposed methods in this book do not require any probabilistic assumptions, but ANN have certain limitations:

- The data (patterns) are to be randomized for training the network.
- ANN will not provide a relationship between input and output.
- The dimensionality reduction cannot be easily done.
- The degree of abnormality cannot be measured on a scale.

An important goal of this book is to overcome the limitations of existing multivariate/pattern recognition methods. In the proposed methods, the problem of multivariate diagnosis is viewed with an entirely different perspective. In Chapter 10, a detailed discussion of classical multivariate methods and neural networks is provided. These methods are also compared with proposed methods by using suitable examples.

1.4 APPROACH

In this section, the overall approach to meeting the goal of the book is discussed. The approach is developed by considering the nature of problems faced by decision makers while dealing with multidimensional systems. In proposed methods the scaled Mahalanobis distance is used to develop a measurement scale for the multidimensional systems, and the principles of Taguchi Methods (TM) are used to optimize the system and predict its performance. Therefore, this methodology is referred to as the *Mahalanobis–Taguchi System* (*MTS*). In MTS, scaled MDs are computed using the inverse of the correlation matrices. If the scaled MDs are computed using the Gram–Schmidt orthonormalization process, then such a method is referred to as the *Mahalanobis–Taguchi–Gram–Schmidt* (*MTGS*) *method.* MTS and MTGS methods are discussed in Chapters 2 and 3.

1.4.1 Classification versus Measurement

In classical methods, such as discrimination and classification methods (and sometimes in multiple regression), the objective is to classify the observations into different groups (populations). On the contrary, the main objective of the methods proposed in this book is to provide a measurement scale to measure the degree of abnormalities on a continuous scale. This will help determine the appropriate actions to take based on the degree of abnormalities.

1.4.2 Normals versus Abnormals

In classical methods, both the normal group and abnormal group are considered separate populations. The classification is based on the distances of an observation from the means of these populations. In the MTS/MTGS methods, there are no populations. We need only a group of observations called a “normal” or “healthy” group to obtain correlation structure and to define the reference point to the measurement scale. Selection of this group is entirely at the discretion of decision maker. In the MTS/MTGS methods, every abnormal condition (or a condition outside “healthy” group) is considered unique, since the occurrence of such a condition is different. The degree of abnormality is measured in reference to the normal group. In this context, it is worthwhile to note Tolstoy's quote in *Anna Karenina:*

> *All happy families look alike. Every unhappy family is unhappy after its own fashion.*

1.4.3 Probabilistic versus Data Analytic

In classical multivariate methods, probability-based inference is used for analyzing multivariate systems. For example, in discrimination and classification methods, the cost of misclassification and probabilities are used for classification and the stepwise regression models are probabilistic. On the contrary, in the MTS/MTGS methods, the measures and procedures used are data analytic. The quadratic loss function concept is used for determining the value of a threshold. The loss function approach minimizes the total cost by considering various cost elements. This method of finding threshold is not probabilistic—it uses the measures of descriptive statistics. Based on the threshold, the multivariate systems can be monitored and appropriate actions can be taken accordingly.

1.4.4 Dimensionality Reduction

In multivariate systems, dimensionality reduction is still a challenge. Principal component analysis is used to reduce the dimensionality of the systems by computing the principal components. However, to calculate one component, we need the entire original set of variables. In other words, this technique does not provide a methodology for dimensionality reduction in terms of original variables. Methods like test of additional information, though intended for dimensionality reduction in terms of original variables, require prior knowledge about the variables. They also depend on the F-ratio, which may not be sufficient to identify the variables of importance. Stepwise regression also depends on an F-ratio. This technique requires several iterations if the number of variables is large.

We propose the use of S/N ratios to identify the variables of importance. Based on S/N ratios, we can directly obtain the useful set of original variables.

1.5 REFINING THE SOLUTION STRATEGY

Given the preference to a data analytic and simple methods for multivariate diagnosis, the top-level solution strategy integrates Mahalanobis distance and the principles of Taguchi Methods. Having chosen a top-level solution strategy, the subrequirements can be defined as shown in the Figure 1.3. In addition to having an effective diagnostic procedure, it may be necessary to make robust decisions under the influence of noise factors, to identify the direction of abnormals and to overcome effects of multicollinearity and small correlations. The solution strategy developed takes these aspects into account.

Much of this book is concerned with expanding the solution strategy shown in Figure 1.3. This book will provide a set of equations, concepts, and procedures that help decision makers make effective decisions while diagnosing multidimensional systems. The book aims at continually measuring multidimensional systems with the help of simple measures and procedures.

1.6 GUIDE TO THIS BOOK

Most of the chapters in this book are developed based on Chapter 2, which provides an introduction to MTS/MTGS methods. To get an idea about these methods and how they differ from existing methods and for actual cases, readers can browse through Chapters 10 and 11 using Chapter 2 as a basis. The following are brief chapter summaries of the book:

Chapter 1 is an introductory chapter that defines the goal of the book, discusses the nature of the problem, and proposes an approach to its solution. The chapter briefly discusses the related methods in existing literature.

Chapter 2 introduces the MTS and MTGS methods. A comparison is made between these two methods with the help of medical diagnosis data.

Chapter 3 discusses the advantages and limitations of MTS and MTGS methods. The examples provided in this chapter include a medical diagnosis system, a graduate admission sys-

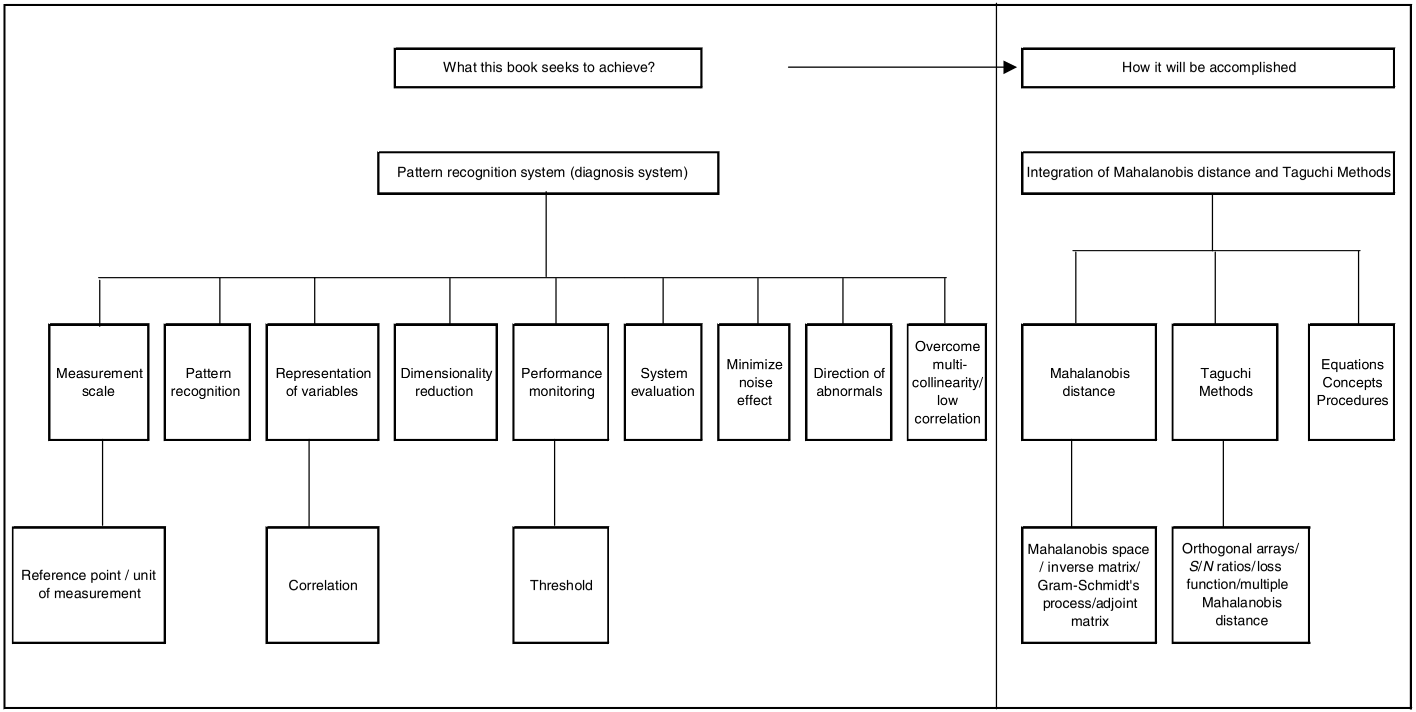

Figure 1.3 Solution strategy.

tem, and the data set provided by American Supplier Institute (ASI).

Chapter 4 introduces the role of orthogonal arrays (OAs) and signal-to-noise (*S/N*) ratios in multivariate diagnosis. The chapter shows how dimensionality reduction can be performed with the help of OAs and *S/N* ratios. The chapter also discusses the advantages of using *S/N* ratios in multivariate diagnosis.

Chapter 5 explains the applicability of MTS/MTGS methods for categorical data. This feature is explained using a sales promotion case study.

Chapter 6 describes different ways of treating noise factors in MTS/MTGS methods. Noise conditions are changes in usage environment, such as conditions in different places and manufacturing variability of the systems.

Chapter 7 describes the role of quality loss function in determining thresholds for MTS/MTGS methods. These methods cannot use the prior probabilities and expected cost of misclassification because they are data analytic.

Chapter 8 discusses the standard error of the measurement scale developed using MTS/MTGS methods.

Chapter 9 focuses on advanced topics related to MTS/MTGS methods. These topics are included to provide suitable procedures to overcome problems due to multicollinearity and small correlations, to select subsets for complex cases, and to select a good Mahalanobis space from historical data.

Chapter 10 compares MTS/MTGS with classical multivariate statistical methods and artificial neural networks, using a medical diagnosis example.

Chapter 11 describes the applicability of this methodology to the case studies in different areas, representing both American and Japanese industry.

Chapter 12 concludes by highlighting the important points of proposed methods, their scientific contributions, limitations, and directions for future research.