A Design Performance Driven Learning Framework for Conceptual Design Knowledge: Methodology Development and Applications

by

Sakda Chaiworawitkul

M.Eng., Civil and Environmental Engineering
Massachusetts Institute of Technology, 2002

M.Eng., Civil Engineering
The University of Tokyo, 2001

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the
Massachusetts Institute of Technology
September 2007
[February 2008]

Signature of Author:______________________________
Department of Civil and Environmental Engineering
September 21, 2007

Certified by:______________
John R. Williams
Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by:______________________________
Daniele Veneziano
Chairman, Department Committee on Graduate Students

A Design Performance-Driven Learning Framework for Conceptual Design Knowledge: Methodology Development and Applications

by

Sakda Chaiworawitkul

Submitted to the Department of Civil and Environmental Engineering on Sep 21, 2007 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Field of Information Technology

ABSTRACT

This thesis develops a learning framework for automation of acquisition of bridge conceptual design knowledge. The thesis proposes a new learning methodology explicitly aimed at capturing quality design aspects to help engineer gain insight into good design. The research uses the National Bridge Inventory (NBI) data, which contains more than 600,000 bridges. The physical condition ratings are used as proxies for design quality.

In this data the relationships between physical condition ratings and bridge design elements are not well-known. The simultaneous equation model (SEM) technique is employed to model the physical condition ratings. SEM has the advantage over existing methods of state transition probability estimation in that no a-priori subjective conditional grouping is required. The resulting model yields the marginal effects of design variables on condition ratings, which is easy for engineers to interpret. The analysis results reveal that design features available in the NBI database alone do not adequately explain the resulting condition ratings.

Using the identified performance model, COBWEB, an incremental clustering algorithm, is employed to learn mappings from design specification to configuration space. However, the COBWEB branching strategy focuses on probabilistic predictability of feature values. The learned knowledge therefore represents not clusters of good design aspects but rather clusters of local similarity. A modification to the existing strategy is proposed. A set of experiments has been conducted to compare the original and the modified COBWEB. Finally, the thesis provides a detailed discussion of issues related to the quality of the NBI database and proposes strategies for improved analysis of the NBI bridge data.

Thesis Supervisor: John R. Williams

Title: Professor of Civil and Environmental Engineering

To my parents, Pisoot and Krittiya, and my wife Joy.

Acknowledgements

I would like to thank my advisor, Prof. Williams, and other committees, Prof. Connor and Prof. Kocur for their ideas and comments that shape this thesis up and their valuable support throughout my work here at MIT.

Many thanks to William Robert from Cambridge Systematics Inc. and Amjad Waheed from Ohio DOT for knowledgeable discussions on NBI bridge data and insights on Bridge Management System.

Thanks to Paul Lee and Steven Zhu in Credit Analytics group at Banc of America Securities who have dedicatedly committed to authorize employee's educational support fund for my tuition during my employment at Banc of America Securities. Of course, my gratitude is also extended to other folks in Credit Analytics team at Banc of America Securities who have been supportive not only to my career responsibility but also opportunity for finishing up this thesis.

Thanks to Patrick Henaff at Merill Lynch Co. who has spared some of business time from hard and busy work during day time for me to focus and finalize this thesis.

Finally, my love and thanks to my wife Joy, and my beloved parents and family who had to put up with me whilst I was working on this thesis.

A Design Performance-Driven Learning Framework for Conceptual Design Knowledge: Methodology Development and Applications

by

Sakda Chaiworawitkul

Submitted to the Department of Civil and Environmental Engineering on Sep 21, 2007 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Field of Information Technology

ABSTRACT

This thesis develops a learning framework for automation of acquisition of bridge conceptual design knowledge. The thesis proposes a new learning methodology explicitly aimed at capturing quality design aspects to help engineer gain insight into good design. The research uses the National Bridge Inventory (NBI) data, which contains more than 600,000 bridges. The physical condition ratings are used as proxies for design quality.

In this data the relationships between physical condition ratings and bridge design elements are not well-known. The simultaneous equation model (SEM) technique is employed to model the physical condition ratings. SEM has the advantage over existing methods of state transition probability estimation in that no a-priori subjective conditional grouping is required. The resulting model yields the marginal effects of design variables on condition ratings, which is easy for engineers to interpret. The analysis results reveal that design features available in the NBI database alone do not adequately explain the resulting condition ratings.

Using the identified performance model, COBWEB, an incremental clustering algorithm, is employed to learn mappings from design specification to configuration space. However, the COBWEB branching strategy focuses on probabilistic predictability of feature values. The learned knowledge therefore represents not clusters of good design aspects but rather clusters of local similarity. A modification to the existing strategy is proposed. A set of experiments has been conducted to compare the original and the modified COBWEB. Finally, the thesis provides a detailed discussion of issues related to the quality of the NBI database and proposes strategies for improved analysis of the NBI bridge data.

Thesis Supervisor: John R. Williams

Title: Professor of Civil and Environmental Engineering

Contents

LIST OF FIGURES

LIST OF TABLES

It does not take maturity to realize that models are to be used but not to be believed.

- Henri Theil

INTRODUCTION: A METHODOLOGY FOR ENGINEERING CONCEPTUAL DESIGN KNOWLEDGE LEARNING

1. INTRODUCTION

1.1. Scope

The conceptual design phase lies at the heart of the engineering design process during which an initial design is generated. This process, has a great bearing on the quality and success of the produced artifact. Knowledge to perform such a task is only acquired through years of practical experience. The scarcity of synthesis knowledge for conceptual design and the importance – and difficulty – of this design phase has led to a desire to build computer systems which incorporate design knowledge and which can generalize it to new problems. Although the idea was promising, the fact that traditional strategy for acquiring such heuristic knowledge using "knowledge engineers" that tries to capture knowledge directly from experts has been found inefficient and expensive. With the advent of cheap computational power, a number of attempts have been made to acquire such heuristic knowledge automatically using a so-called "Machine Learning" (ML) technique.

This thesis investigates the use of the ML technique for learning of design knowledge from design examples in the light of the resulting performance of the produced artifact. This differs from the majority of studies in the same area whose knowledge is found to be shallow and usually does not provide insight about quality of design to engineers/experts. This involves investigation and development of models capturing the relationships between design choices and the resulting quality of design, in this case measured by performance indices. The knowledge of induction algorithm should then incorporate the identified performance model into the knowledge extraction strategy in an explicit manner rather than using pure statistical measures like learning error.

This thesis uses National Bridge Inventory (NBI) data domain as an example of its application. The NBI database possesses a certain degree of richness in terms of data description for a Bridge Maintenance System (BMS). Part of the current thesis has been devoted to in-depth analysis of the NBI data to gain insights into relationships between various aspects that underscore the NBI bridge specification, design configuration, and performance.

The thesis contribution to current studies on automated design knowledge learning is to extend existing ML techniques whose learning strategy is purely constructed by statistical measures of local similarity. The engineering objective is introduced to accommodate a performance model, thus the resulting methodology tries to simultaneously optimize the two objectives during the learning process. The starting point of the proposed framework is a statistical identification of a performance model estimated from NBI data. Such a model enables engineers to gain insights into the marginal effects of design elements on the resulting performance. Then the model is incorporated into the modified knowledge learning algorithm to produce general design rules for engineers to apply it to new problems. The performance model plays a key and distinctive role here in that it provides insight to engineers as to why, under a given design setting, one would prefer a particular design to another. Thus making the knowledge deep and useful for engineers/experts.

1.2. Motivation

The philosophy espoused in this thesis is that recent improvements in statistical modeling techniques open up new opportunities to view design knowledge learning from a different perspective than the one provided by traditional Artificial Intelligence (AI) or expert systems. This technique enables construction of a design knowledge learning machine with arbitrary learning characteristics and behaviors tailored to fit niches of different learning problems and situations. A learning problem can be viewed as the construction of a hypothesis function that optimizes certain objective functions.

This development perspective is the main motivation of this thesis, in aiming to harness this view to flexibly design a new knowledge learning algorithm that allows one to direct learning to focus on the performance perspective of the artifact in an explicit manner. In particular, this thesis explores the opportunities which are offered by the following ML and econometric methods:

Simultaneous Equation Modeling (SEM): an econometric framework that allows proper treatment of a system of contemporaneous response (or dependent) variables. The framework is particularly useful in our context when modeling multiple performance variables that contemporaneously exist with other exogenous variables to the performance system.

COBWEB: an incremental conceptual clustering algorithm which constructs a concept hierarchy partially ordered by generality through incremental incorporation of new example instances. Each node of the classification tree represents a concept as the probability of the occurrence of each attribute-value pair presenting that concept. The criteria to identify appropriate operators for growing the tree (create a new node, merging or splitting existing nodes) are driven by maximizing a so-called *category utility* measure. The measure is purely probabilistic and represents a tradeoff between accuracy of prediction of feature values in a cluster class and the class size.

The objective of this thesis is therefore to exploit the strength of the SEM model to develop a performance model comprising multivariate performance measures which are hypothesized to be endogenous to each other. The resulting model is then used to augment the weakness of COBWEB to, instead of only focusing on purely balancing node size and predictability of a class by a feature value, simultaneously optimize the local gain of average performance over the global average performance.

1.3. Thesis Overview

There are three main components that form this thesis, namely: statistical analysis of NBI bridge data, NBI bridge performance modeling, and development of learning algorithm for design knowledge learning by performance. The relationship among these three components is illustrated in Figure 1-1.

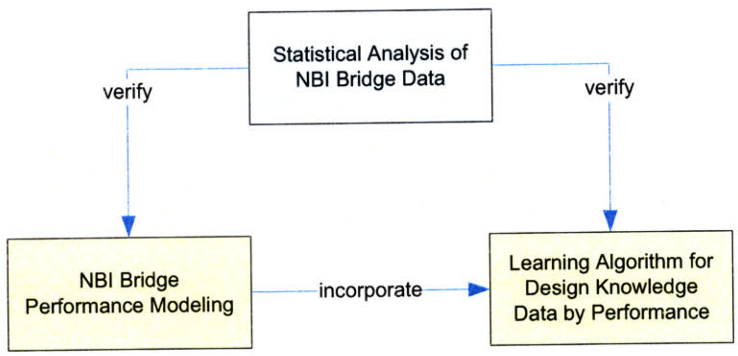


Figure 1-1. Illustration of relationship among three main components of current thesis

The first component focuses on finding trend relationships between three aspects of NBI data, i.e., specifications, design configurations, and performances. This serves as a basis of understanding the underlying relationship structure embedded in the data. These trends are useful insights that can be used to criticize and verify validity of the next two components.

The second component is concerned with modeling of NBI bridge performance whereby performances are measured as a system of multivariate variables. In our application, we select physical condition ratings on deck, superstructure, and substructure elements of the bridge to present design performance measures. They are hypothesized to following common stochastic latent process of deterioration. A proper treatment of such a model is to exploit SEM framework. The model for these performance measures are identified with in-sample data and validated with out-of-sample data.

The third component demonstrates a proposed extension to the current design learning algorithm, particularly COBWEB to incorporate the performance model identified in the previous step into its learning objective. The central idea is to modify the category utility definition to accommodate performance in a proper manner.

1.4. Contributions

The thesis has three contributions:

- In-depth analysis of NBI bridge data: the thesis presents detailed analysis of trend studying in the NBI bridge data particularly on the relationship among specifications, design configurations, and performances. Literature review on accuracy of NBI bridge performance modeling also provides an insight and hint for further improvement for data collection in NBI database to better explain the deterioration process in the physical condition ratings of NBI bridges. A review on accuracy of physical condition rating measurement is also provided to raise an issue on its impact to performance modeling accuracy.

- SEM approach for modeling bridge performance: The thesis presents the necessity to proper treatment of endogeneity possessed by regression of multivariate performance measures such as NBI bridge physical condition ratings. The approach also has advantage over traditional method employed by typical Bridge Management System (BMS) which relies on estimation of transition probability in that it does not require bucketing bridge instances by its characteristic to form conditional probability of transition. This is typically done subjectively due to the difficulty in defining *similarity* on data space formed by NBI data which is a mixture of continuous, nominal, and ordinal-scaled quantities.

- Development of design knowledge learning algorithm with performance objective: Unlike traditional ML approaches employed in similar type of study in which learning is driven by pure statistical objective, the thesis presented a modification of the existing learning algorithm to accommodate performance into its learning objective. The main goal is to construct a learning algorithm that simultaneously optimizes similarity or prediction accuracy in order to measure what was originally employed by the algorithm and the newly introduced performance measure.

- Suggestion for current NBI data enhancement: In the thesis, we have found the inadequacy of the NBI data in its current status to explain the measured performance. Using the experience from studies in the current thesis, a set of suggested enhancement for current NBI data is presented.

1.5. Organization

The thesis is divided into five parts.

The first part of this thesis consists of a brief introduction, followed by extensive review of existing methodology for engineering design knowledge learning. The missing element in the existing studies which is the learning for design performance is pointed out and an outline of the methodological development described in the remainder of the thesis is laid out. The subsequent three parts of the thesis describe the three components of our work: Part I describes the statistical analysis of NBI bridge data, Part II describes the proposed framework for construction of NBI bridge performance and empirical evaluation of the estimated model, and Part III demonstrates to the incorporation of the identified performance model into existing design knowledge algorithm to form a novel algorithm that incorporates performance into its learning objective.

The organization of the chapters is given below:

Introduction:

- Chapter 1 consists of a brief introduction which outlines the scope, motivation, and organization of the thesis as well as summarizing the main contributions which it makes to the realm of automated engineering design knowledge learning study.
- Chapter 2 presents a review of background and recent development of methodology for design knowledge learning. The drawbacks and missing elements of existing automated design knowledge acquisition methodologies are discussed.
- Chapter 3 presents an overview of our proposed methodology in correspondence to the drawback of existing methods identified in Chapter 2. The roadmap to the rest of the thesis is also postulated.

Part I: Overview of NBI Database

- Chapter 4 introduces NBI database and presents a detailed statistical study. Particularly from the perspective to identify major trends and relationships among design specifications, configurations, and the resulting performances.

Part II: Statistical Modeling of Bridge Performance

- Chapter 5 extensively reviews recent developments and methodologies used for bridge performance modeling. It describes transition probability estimation based approach, ordered probit model approach, and simultaneous equation model approach in great detail. These methodologies are qualitatively assessed in the light of their applicability of construction performance model comprising multivariate performance variables.
- Chapter 6 describes empirical evaluation of performance model constructed by the method suggested in Chapter 5. It also provides insights on adverse factors that impair model accuracy observed from the experiment.

Part III: Conceptual Design Knowledge Learning by Design Performance Objective

- Chapter 7 examines existing design knowledge learning framework. The discussion of metric for algorithm selection is introduced and used to identify appropriate methodology for the study. The so-called COBWEB algorithm appears to be the best suitable methodology to use. However, we suggest necessary modification to the existing COBWEB learning strategy to incorporate performance measure obtained from Chapter 6 into its learning objective.

- Chapter 8 describes empirical evaluation of the modified COBWEB and comparison of its performance to the plain vanilla COBWEB.

Conclusions

- Chapter 9 discusses inadequacy of NBI bridge data with respect to the outlook for design for long-term performance scheme which is a topic of interest for Federal Highway Administration (FHWA) and authorities in recent years. Using experience from the studies in this thesis, a set of recommended enhancement to the NBI data is presented.
- Chapter 10 describes the conclusions of the thesis and discusses avenues for future developments.

2. BACKGROUND AND LITERATURE REVIEWS

Lying at the heart of the engineering conceptual design process is the *synthesis of potential solutions* during which an initial solution for the design problem is devised. The synthesis of solutions is usually dependent on the knowledge of the designer. This knowledge cannot easily be taught or captured due to the *heuristic* nature of such knowledge. The advent of intelligent computer systems has set forth attempts to automate acquisition of such knowledge. This chapter investigates existing studies in the area of automatic design synthesis knowledge acquisition. We begin by classifying types of design knowledge to be learned which, in turn, defines different learning schemes in existing literature.

Among all these schemes, the ones relevant to this research interest are the so-called 'learning of design composition' and 'learning for performance evaluation' frameworks. In this chapter, a subsection is devoted to thedetailed discussion and literature review on each of these frameworks. Finally, we point out a gap between the two frameworks and suggest a more suitable learning paradigm for synthesis knowledge.

2.1. Types of Design Knowledge Learned

In this chapter, unless specified, the term learning is used to mean automated learning of design knowledge, in particular learning by the method of Machine Learning (ML) programs. According to Sim and Duffy (1998), the types of design knowledge learned can be classified into the following:

- Product design knowledge
- Design process knowledge.

While the first type of knowledge to be learned is the heuristics behind generation of design artifacts, the second type involves learning the rationale that drives design actions to be taken to progress the design. Among the two types, the first is relevant to the interest of this thesis. Within the paradigm of product design knowledge learning, there exist various branches of research focusing on different types of knowledge specific to the design product/outcome. From all learning types discussed by Sim and Duffy (1998), the schemes relevant to this thesis are:

- composition of the components/subsystems that constitute the product
- performance evaluation knowledge

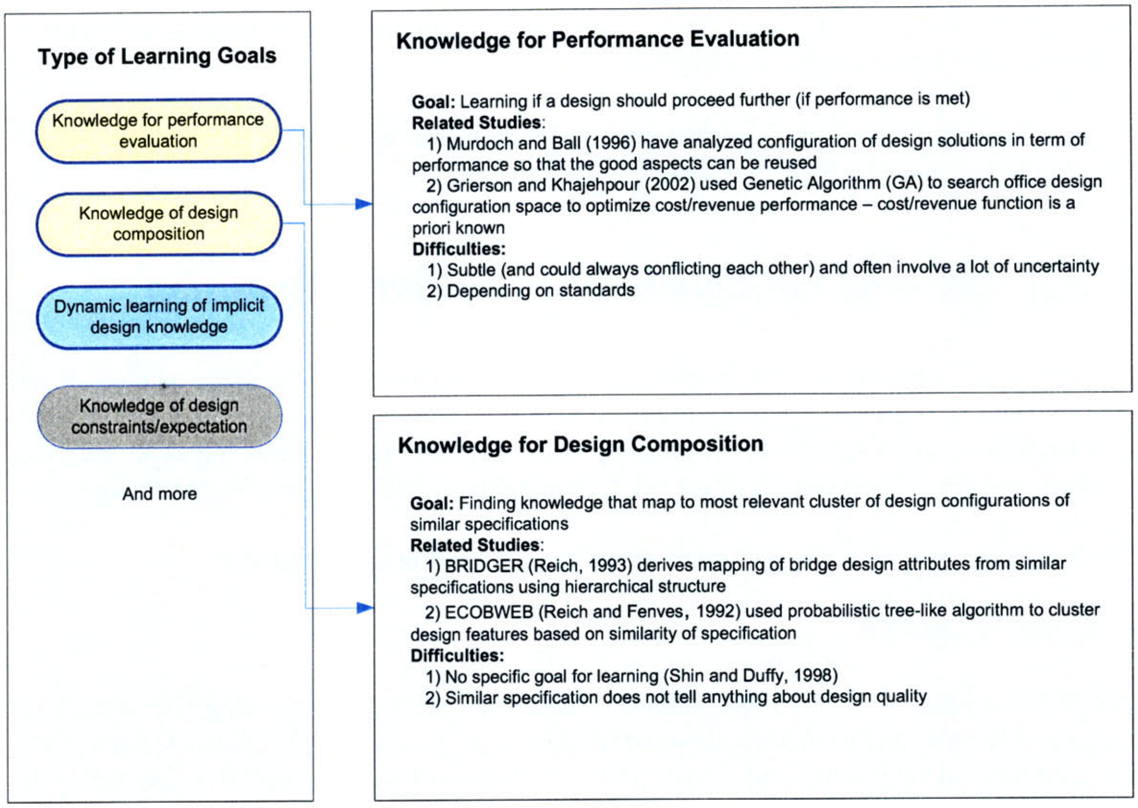


Figure 2-1. Type of knowledge to be learned under product design knowledge learning category

In learning design composition knowledge, past design is used to initiate the synthesis process of design problems that have similar design specification. Examples of this type of learning are BRIDGER (Reich, 1993) and ECOBWEB by Reich and Fenves (1992). The latter employs a hierarchical clustering technique which maximizes a so-called 'utility function' for classification of bridge specification and design attributes into clusters of subspaces based on similarity of bridges specifications. It can be considered as a framework to construct design concept and is employed in BRIDGER primarily for synthesis of the different cable-stayed bridge concepts.

In the paradigm of learning for performance evaluation knowledge, the knowledge to be learned is typically a mapping from the design solution space to the design behavior space. The learned knowledge is useful in various stages of design processes in the sense that it provides a support to the decision for furthering the design or not. Formalization of relationship between decisions about values of design variables and their consequent performances is represented as mapping from feasible design variable space to performance space. Each point on the feasible design solution space represents a particular combination of design decisions regarding the design variables. By identifying the *best or optimal* feasible region on the design solution space, the mapping back from performance space to the original design space yields the design solution which results in such performance. An example of such a learning type is Grierson and Khajehpour (2002). They applied multi-criteria optimization on selected office building design configuration attributes and searched for candidate configuration that optimizes the preset multi-objective using stochastic search.

In the following two sections, philosophy and techniques underlying these two types of knowledge learning are reviewed and discussed in detail.

2.2. Learning for Knowledge of Design Composition

In this type of learning, design decisions are conceptualized and formalized as clusters of designs with resembled specifications. Past designs are used as a starting point for construction of such design conceptualization. Important to this learning type are two components, namely clustering objective functional and clustering technique/strategy.

The followings are reviews of clustering techniques used in the literature.

COBWEB/ECOBWEB

COBWEB (Fisher, 1987) is a hierarchical clustering technique employing five operators to determine how best to incorporate an example (e.g., existing design) into the hierarchy. The category utility (Gluck and Corter, 1985) is used as clustering objective functional. It can be viewed as a function that rewards traditional virtues held in clustering generally – similarity of objects within the same class and dissimilarity of objects in different classes. Classification topology is constructed in such a way as to maximize the average Category Utility (CU) over all clusters on the hierarchy. Precisely, the CU function for the k-th cluster is defined as:

$$CU_k = P(C_k)\left[\sum_i \sum_j P(A_i = V_{ij} \mid C_k)^2 - \sum_i \sum_j P(A_i = V_{ij})^2\right] \qquad (2\text{-}1)$$

where $P(A_i = V_{ij} \mid C_k)$ is the probability of the i-th attribute of the observed data taking the j-th label value given that the observation is classified to the k-th cluster. The first time can be rewritten as follows using the Bayes rule.

$$\sum_i \sum_j P(A_i = V_{ij} \mid C_k) P(C_k) P(C_k \mid A_i = V_{ij}) \qquad (2\text{-}2)$$

The first term in the product of Equation (2-2) can be interpreted as *intra-class* similarity. The larger this probability, the greater the proportion of class members sharing the value and the more predictable the value is of that class member. The last term is the *inter-class* similarity. The higher this probability, the fewer the objects in contrasting classes that share this value and the more predictive the value is of this class.

Thus, the term defined in Equation (2-2) can be viewed as a tradeoff between intra-class similarity and the intra-class dissimilarity. The CU function is therefore defined as the gain of expected number of attribute values that can be correctly guessed given cluster k ($P(C_k)\sum_i\sum_j P(A_i = V_{ij} \mid C_k)^2$) over the expected number of correct guesses with no such prior knowledge ($P(C_k)\sum_i\sum_j P(A_i = V_{ij})^2$).

Examples are permanently incorporated into the hierarchy by sorting through the hierarchy and find the best host node that maximizes average CU over all clusters, i.e.

$$\frac{\sum_{k=1}^{n} P(C_k)\left[\sum_i\sum_j P(A_i = V_{ij} \mid C_k)^2 - \sum_i\sum_j P(A_i = V_{ij})^2\right]}{n} \tag{2-3}$$

Note that COBWEB in its original version can only handle nominal data.

To be able to handle real world data, ECOBWEB (Reich and Fenves, 1992) has been developed as an extension of the original COBWEB to be able to handle real values.

ID3 and C4.5

ID3 was first developed by Quinlan (1986). Its extension C4.5 (Quinlan, 1993) has been a widely used classification/decision tree algorithm. In the context of engineering design knowledge, Kumar, Subramaniam and Teck (2004) employed C4.5 algorithm to inductively learn knowledge of fixtures conceptual design.

At the heart of C4.5 (or ID3) lies the *entropy* function as impurity measure for branching control. Entropy or impurity after splitting at node m is given by

$$I_m = -\sum_{j=1}^{n} \frac{N_{mj}}{N_m} \sum_{i=1}^{K} p_{mj}^i \log p_{mj}^i \tag{2-4}$$

where N_m and N_{mj} denote number of observations in node m and number of observations in node m that takes value j at the branching attribute. And p_{mj}^i the probability of class C_i at node m that takes value j at the branching attribute is defined as

$$P(C_i \mid x, m, j) \equiv p_{mj}^i = \frac{N_{mj}^i}{N_{mj}} \tag{2-5}$$

Although other measures for branching can be used, the most common ones are, for example, *Gini index* (Breiman et al., 1984) and misclassification error.

2.3. Learning for Knowledge for Performance Evaluation

Typically, the problem in this learning scheme involves evaluating multiple conflicting performance criteria. Examples of such learning problems are Grierson and Khajehpur (2002) and Agarwal and Raich (2005). Both employ multi-objective genetic algorithms (Kalyanmoy, 2001) to search feasible design spaces and identify configuration that best optimize the competing strategies. Grierson and Khajehpur (2002) applied such optimization technique totheconceptual office building design problem in which the cost and revenue functions are used as objectives. Several design attributes such as structural system or floor system, etc. are searched to simultaneously optimize capital cost, operating cost and income revenue. On the other hand, Agarwal and Raich (2005) used a similar type of algorithm to locate Pareto optimality for volume, deflections, and stress on the space of truss topologies, geometries, and member size.

It should be noted that the solution obtained from optimization in this research is the actual design configuration itself and it is impossible to extract any heuristics regarding design aspects that separate good from bad design from the resulting Pareto solution set.

2.4. Gap between Methodologies

In this section, we further the discussion to show that although the above two approaches are useful, they do not suffice the purpose of this thesis. Remember from Chapter 1 that the thesis objective is to establish a framework for automated design knowledge acquisition so as to understand which design configurations, as a rule of thumb, lead to good or poor performance.

The learning for design composition knowledge is helpful in this purpose but is insufficient. The knowledge is insufficient because the learning problem which is mostly relying on clustering mechanism is constructed on statistical similarity or probabilistic predictability of each cluster on some common attribute values. To confirm this, Reich and Fenves (1992) reported the capability of adaptability of their COBWEB based algorithm to design periods. In other words, clusters were organized in such a way that they represent design concepts in each period. However, it is unclear from our perspective how performance measures form a context or concept in the design space. Generally speaking, it is unclear how predictability tradeoff between inter-class and intra-class modeled in CU function relates to meaningful design cluster. Yet, the predictability of cluster class with respect to specific attribute values is an indispensable tool. This is because, by virtue of clustering, we would like to be able to determine the obtained clusters that are best represented by which group of attribute values. Nonetheless, to make a conceptual clustering like COBWEB meaningful for our purpose, some adjustments on the control of clustering strategies, in particular, the CU function, should be introduced in such a way that it explicitly directs the clustering strategy to focus on performance measure simultaneously with trading off cluster predictability.

On the other hand, as pointed out earlier, most of the studies in the second paradigm to learn for performance evaluation involve searching for a Pareto solution set that optimizes competing objectives. The result is typically the design configuration itself rather than the structure of heuristics that leads to understanding of what causes good or bad performance. However, studies in this learning type explicitly exploit knowledge of performance (which is a priori known or identified) of which the element that learning for design composition is lacking.

2.5. Conclusions

This chapter first starts with classification of type of engineering knowledge learned in the studies. Among all the types classified by Sim and Duffy (1998), those relevant to the current thesis are the "learning for design composition knowledge" and "learning of performance evaluation knowledge". In the first paradigm, the main trend of studies is to cluster similar designs together and separate dissimilar ones. The clustering strategy can be based on several types of strategies such as predictability of cluster class with respect to specific attribute value group or impurity measures like entropy. In the second paradigm, most studies involve solving for Pareto solution set that simultaneous optimizes competing performance objectives (typically a priori known or identified).

Unfortunately, both paradigms do not suffice the purpose of our thesis. The scheme for learning knowledge of design composition does not explicitly use performance as the design goal for clustering but rather the predictability or impurity. It is unclear how these measures relate to design goal or performance. Nonetheless, they are not completely useless. On the other hand, the learning of knowledge for performance evaluation mostly focuses on solving the Pareto solution set itself rather than extracting constructive knowledge. However, they explicitly exploit the knowledge of performance in deriving optimal solutions.

This observation suggests that these two paradigms cannot be directly applied to our study. In the next chapter, we will devote our discussion to the adjustment to be made on these two paradigms to make it sufficient to our thesis goal.

3. OVERVIEW OF PROPOSED METHODOLOGY

The gap between the existing methodologies described in chapter 2 is due to the lack of clear engineering objective in the learning for design configuration and the type of knowledge learned from design for performance, which is in a functional form and thus opaque to engineers. This makes it difficult for engineers to visualize the underlying structure of learned knowledge.

In this chapter, we propose a methodology to be used in this study to overcome described difficulties. The proposed methodology can be seen as a combination of the two learning paradigms. However, the drawbacks of the two paradigms are addressed with machine learning techniques such as conceptual clustering.

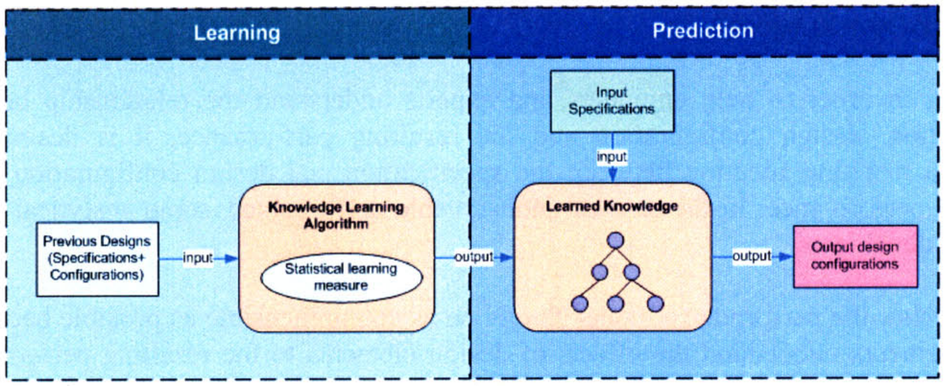
3.1. Proposed Methodology for Engineering Design Knowledge Learning

As discussed in Chapter 2, the learning for design configuration typically assumes knowledge of the design lies within the existing designs. The vast body of research has been using machine learning by induction (mostly clustering technique) to extract design knowledge from the examples. However, in most cases, the knowledge induction strategies are carried out by employing statistical similarity measure. This type of strategy searches for similarity in the underlying specifications to suggest existing design configurations based on similar type of specification. However, it is unclear what the rationale behind the suggested design is and how the performance of the suggested design configuration would be.

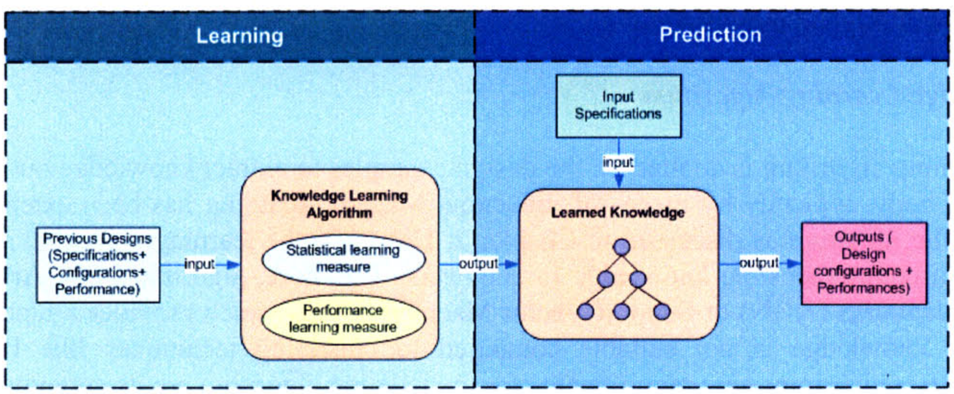
On the other hand, the learning for performance mostly focuses on searches throughout the feasible design configuration spaces constrained by design constraints. Using a priori known as design performance objective function, the search is performed to find the optimal design solution in the feasible space. Unfortunately, the structure of knowledge to help engineers or experts understand the decision is not learned. In the literature, this type of knowledge is called "*incomprehensible*" or "*opaque*" (Reich, 1997) and thus is not suitable for our purpose.

In contrast to this approach, the first paradigm forms clusters of design specifications and configurations and offers less opaque knowledge for engineers to consume. However, what it lacks is the ingredient from the second paradigm which describes the design performance.

Thus, the proposed methodology for this thesis is to incorporate the performance into a learning goal. (see Figure 3-1) This approach can be viewed as design knowledge formation which tries to capture specification and configuration combinations that results in good performance.



(a) Without performance learning measure (most of existing studies)



(b) With performance learning measure (proposed methodology)

Figure 3-1 Learning and prediction of design configuration

3.2. Elements and Requirements of Proposed Methodology

From Figure 3-1, the elements for the proposed learning methodology are:

- Existing design samples (annotated with specifications, configurations and performances)
- Knowledge learning algorithm
- Objective function for knowledge learning algorithm

Existing design samples

The minimum requirements are that the design samples should include specifications and the outcome design configurations. If the associated performance of an artifact is unknown or unobservable, a priori known or estimated performance evaluation function can be used to estimate it.

However, in order to help engineers and experts understand the relationship between specification, design configuration and the resulting performance, it is desirable to identify a mapping function between the specification and design configuration to the observed performance. In the case of unobservable performance, some analytical model could be developed.

Nevertheless, the performance model should be as comprehensible as possible because it helps engineers understand the effects of design elements to the resulting performance. This is a non-trivial task because engineering systems are typically involved with a high degree of complexity and thus require sophisticated models to analyze behavior or characterize performance. However, whenever possible, it is always beneficial to employ a model with high comprehensibility.

Knowledge Learning Algorithm

The learning algorithm is applied on the design examples to extract knowledge out of the example body, typically by means of induction. Machine learning has been extensively used in the literature as discussed in Chapter 2. However, the learning algorithm should provide a comprehensible knowledge to engineers. Therefore, algorithms like Artificial Neural Networks (ANN) or Support Vector Machine (SVM) which provides a functional form of knowledge is not suitable compared to clustering techniques like ID3 or COBWEB. Other important requirements for the learning algorithm are its computational complexity, timing of learning, mode of learning, etc. An in-depth discussion of selection of learning algorithm is deferred to Chapter 8.

Objective function for knowledge learning algorithm

The objective function is a key element to success of meaningful knowledge with respect to our objective, i.e., the capturing of design configurations that leads to good performance. As noted above and in Chapter 2, the existing methodologies typically use statistical measure such as information gain or category utility to define local similarity for knowledge construction. (Other types of statistical measures in learning functional form of knowledge are, for example, L2 norm of point error in typical classification problem, or ε-insensitive error for support vector machine.) As discussed in Chapter 2, these statistical measures result in an unclear structure of the obtained knowledge. More specifically, the statistical similarity of design configurations does not provide an explicit basis for building body of knowledge whose intention is to represent designs that leads to a good performance.

Therefore, as proposed in Section 3.1, a modification of the learning objective function must be made to explicitly prescribe the clustering strategy to focus on designs of good performance. Detailed discussion of the modification of existing learning objective function used in this thesis is deferred to Chapter 8.

3.3. Domain of Application

The domain of application selected in this study is the bridge design synthesis knowledge learning. The main reasons for this selection are:

- There exists a rich and well-defined data bank for study. The Federal Highway Association (FHWA) has been gathering on-site data inspection of the nation-wide highway bridges annually since 1972. This is catalogued in the so-called "National Bridge Inventory (NBI) database". Until the year 2006, there are more than 800,000 entries from which more than 500,000 instances are true bridges with qualified length. Data available in the database covers the specification, design configurations and performances at high-level and thus can be employed for conceptual design knowledge learning problem.
- There has been quite a body of studies in this domain. Detailed lists of studies for statistical analysis of NBI data is reviewed in Chapter 4 while bridge conceptual design knowledge learning studies are reviewed in Chapter 2 and in greater detail in Chapter 8.

3.4. Study Roadmap

Figure 3-2 provides a guide map of this study. This thesis study can be broken down into three main steps:

- **Step 1: Studying and analyzing NBI database**
  This step helps provide insight of design trends and the basic relationship of bridge specification and design configuration to performance. The resulting insight is useful for verifying the results of step two and three. Step 1 is discussed in greater detail in Chapter 4 and 5.

- **Step 2: Studying and analyzing NBI database**
  Using specifications, design configurations and recorded performance, a statistical identification of model performance is conducted in this step. The purpose is to quantitatively model map specification and design configuration to performance. The resulting model is useful for engineers/experts to study marginal effect of design factors to its performance. The resulting model is then used to input into the design configuration/feature learning in step three whose objective is to

inductively learn design configurations which maximizes bridge performance. Step 2 is discussed in greater detail in Chapter 6 and 7.

- **Step 3: Studying and analyzing NBI database**
  In this step, the machine learning algorithm is employed with appropriate modification to incorporate performance measure as a goal of learning in as much as the resulting knowledge captures designs that lead to good performance. Step 3 is discussed in greater detail in Chapter 7 and 8.

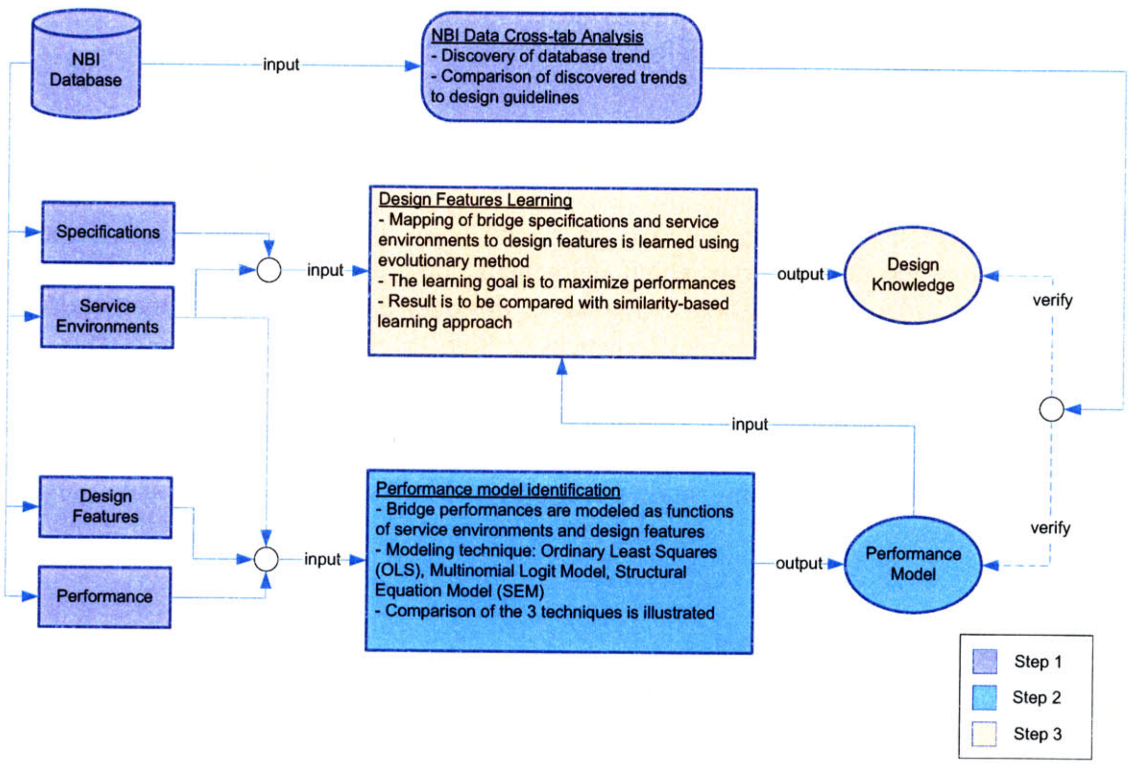


Figure 3-2. Study roadmap

3.5. Conclusions

We presented in this chapter a high-level of proposed study methodology which aims at addressing the capability to explicitly specify performance as learning goal in the existing studies. We proposed that a performance measure is to be incorporated directly into the knowledge induction objective function of the learning algorithm. The concrete presentation of the actual modification is deferred to Chapter 8.

We also have proposed the domain of study to be the NBI bridge data for bridge conceptual design knowledge learning. This is because of the rich data provided by NBI bridge data description and number of existing records. In addition, there have been, significant body of studies on the NBI bridge data.

Finally, we proposed that the study is to be divided into three main steps. These are: statistical analysis of NBI bridge data; performance model identification; and design configuration learning. The first helps provide us insight of typical trends and relationship for design specification, configurations and performances and thus serves as a basis for verification and evaluation of the resulting performance model and learned design knowledge in the latter steps. Step 2 pertains to statistical identification of performance model which aims at capturing a mapping from design specification and configuration space to performance space. It also aims at providing a quantitative insight of design factor effects on its performance. The last step involves inductively learning design knowledge with explicit incorporation of performance measure obtained in step 2 into its objective function.

PART I: OVERVIEW OF NBI DATABASE

4. ANALYSIS OF NBI BRIDGE DATA

THE NBI DATABASE has been widely used for variety of research ranging from understanding of Bridge Management System (BMS) area to prediction of bridge performances. The database provides a descriptive attribute data that characterize highway bridges in different aspects, such as, serviceability, design description, specifications, and performance evaluations. This chapter is devoted to analysis of NBI bridge data and comparison of the results from analysis to other known literatures. The main purpose is to obtain general insights about trends in the existing designs especially in the light of relationship between specifications, design configurations and performance evaluations.

4.1. NBI Database Description

The United States has more than 3.9 million miles of roadway (U.S. Department of Transportation, 1987) and mote than 600,000 bridges (as of 2006). In response to the December 1967 collapse of the Silver bridge over the Ohio River which claimed 46 lives, Congress mandated for the implementation of National bridge inspection standards. The individual bridge inspection records that are based on these standards constitute the National Bridge Inventory (NBI). The purpose of the NBI is to provide a uniform base of bridge information that can be used to identify those bridges that are most in need of repair and to serve as a basis for allocating Federal Highway Administration (FHWA) funding for bridge rehabilitation or replacement. The NBI is administered by the FHWA in Washington, DC, and data are updated continuously based on the latest bridge inspection; most inspections are completed on a 2-year cycle. Culverts that are 20 ft (6 m) or greater in span are included.

Among 116 attributes (called NBI item) (FHWA, 1992), we are interested in the specifications, design configurations and performances aspects of the NBI database. The related fields selected from the NBI data by categories are listed in Table 4-1 to 4-4 below.

Table 4-1. Service environment related attributes

Field Name	Abbreviation	NBI Item #
Age	AGE	N/A
Average Daily Truck Traffic	ADTT	109
Average Daily Traffic	ADT	29
State Code	ADT	1

Table 4-2. Design specifiation related attributes

Field Name	Abbreviation	NBI Item #
Design Load	DL	31
Type of Service On Bridge	TSOB	42
Type of Service Under Bridge	TSUB	42
Lanes On Structure	LOS	28
Lanes Under Structure	LUS	28
Structure Length	SL	50

Table 4-3. Design description related attributes

Field Name	Abbreviation	NBI Item #
Main Structure Material	MSM	43
Main Structure Design	MSD	43
Deck Structure Type	DST	107
Pier Protection	PP	111
Number of Spans in Main Unit	NSMU	45
Number of Approach Spans	NAS	46
Length of Maximum Span	LMS	48
Deck Width	DW	52
Skew	SKW	34

Table 4-4. Performance related attributes

Field Name	Abbreviation	NBI Item #	Brief Description	Possible values
Deck Condition Ratings	DCR	58	Overall condition rating of the deck	0,1,2,3,4,5,6,7,8,9,N
Superstructure Condition Ratings	SPCR	59	Physical condition of all structural members	0,1,2,3,4,5,6,7,8,9,N
Substrucure Condition Ratings	SCR	60	Physical condition of piers, abutments, piles, fenders, footings, or other components	0,1,2,3,4,5,6,7,8,9,N

The bridge age field is inferred from the year built field (item 27). Since our data is as of 2006, bridge age simply equals 2006 minus construction year. Note that most of the attributes in Table 4-2 and 4-3 are nominal type. Meaning, they are discrete label values. On the other hand, although the physical condition ratings listed in Table 4-4 take integer values from 0 to 9, they do not have equidistance between each value. Rather, they are ordinal-value attribute namely the value are discrete labels but possesses order relationship. Table 4-5 tabulates meaning of each possible rating value shown for these physical condition ratings.

Because of the mixture of data type, i.e. continuous, nominal and ordinal, in the NBI database, the conventional continuous-valued analysis technique cannot be employed directly. Rather, approaches like categorical data analysis (Agresti, 1996) is more suitable.

Table 4-5. Meaning of decoded values for physical condition ratings

Code	Description
N	Not applicable
9	Excellent condition
8	Very good condition - no problems noted
7	Good condition - some minor problems
6	Satisfactory condition - structural elements show some minor deterioration
5	Fair condition - all primary structural elements are sound but may have mionr section loss, cracking, spalling or scour
4	Poor condition - advanced section loss, deterioration, spalling or scour
3	Serious condition - loss of section, deterioration, spalling or scour have seriously affected primary structural components. Local failures are possible. Fatigue cracks in steel or shear cracks in concrete may be present
2	Critical condition - advanced deterioration of primary structural elements. Fatigue cracks in steel or shear cracks in concrete may be present or scour may have removed substructure support. Unless closely monitored it may be necessary to close the bridge until corrective action is taken
1	"Imminent" failure condition - major deterioration or section loss present in critical structural ocmponents or obvious vertical or horizontal movement affecting structure stability. Bridge is closed to traffic but corrective action may put back in light service.
0	Failed condition - out ofservice- beyond corrective action

4.2. NBI Database Preprocessing

To ensure the quality of statistical analysis, the input data needs to be preprocessed for appropriate identification and exclusion of anomalies. There are two types of anomalies categorized in our study:

- Data semantic anomalies: This refers to missing data or ambiguous values such as 'Others' or 'Unknown' found in the database. Another type of semantic anomalies pertain data qualification issues such as qualified NBI bridge length or bridge age threshold.
- Statistical anomalies: This refers to outliers in the statistical sense.

4.2.1. Data semantic anomalies

The followings are identified as data semantic anomalies and are excluded from the original data.

- Bridge that does not qualified for NBI bridge length of 20ft. This can be identified by item 112 of code 'N'.
- Bridge constructed earlier than 1930. This threshold is adopted from discretization of bridge period specified by Reich (1992). The discretization is

tabulated in Table 4-6. Our intention is to include only Modern bridges (those constructed from 1941 onward). However, we allow 10 years extra cushion.

- Bridges with reconstruction are excluded because physical condition ratings of these records are spurious. These instances can be identified from 'Year of Reconstructed' field (item 108) with values not equal to 'N'.
- Missing data or ambiguous encoded values such as 'N', 'Unknown', 'Others', etc. are excluded.

Table 4-6. Discretization of bridge period

Construction Date	Period
1818-1870	Craftsman
1871-1890	Emerging
1891-1940	Mature
1941-present	Modern

4.2.2. Statistical anomalies

To identify outlier in the data, we employed *Mahalanobis distance* (Alpaydin, 2004). It is defined as follows:

$$D^2 = (\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \tag{4-1}$$

where $\mathbf{x}$ is and N dimensional multivariate vector (in our case $N = 24$ from Table 4-1 to 4-4). $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are mean vector and covariance matrix of $\mathbf{x}$ respectively. Mahalanobis distance can be viewed as a distance from N dimensional hyperellipsoid centered at $\boldsymbol{\mu}$ whose shape is spanned by covariance matrix $\boldsymbol{\Sigma}$. The Mahalanobis distance approximately follows χ^2 distribution of N degree of freedom.

Note that we simplified data treatment here as continuous value since the proper treatment of outlier finding is beyond the scope of this thesis.

The statistical anomalies or outliers can then be identified as two-tailed extreme values of 10% quantile on the approximated χ^2 distribution.

After application of the above two types of anomalies exclusion, we are left with 260,331 instances.

4.3. Analysis of NBI Data

First, we start with analyzing bridge age in the selected period from 1930-2006. The histogram of bridge age is shown in Figure 4-1. It can be seen that the majority of the bridges are between 0-50 years.

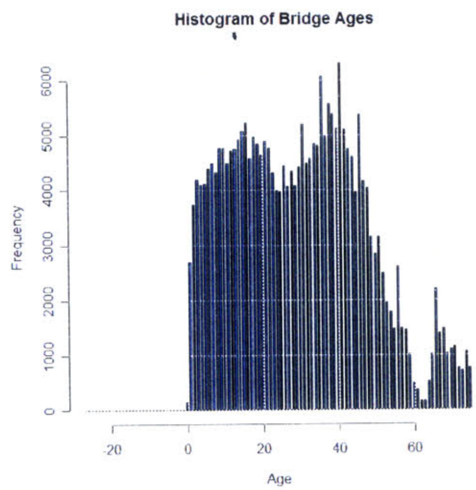
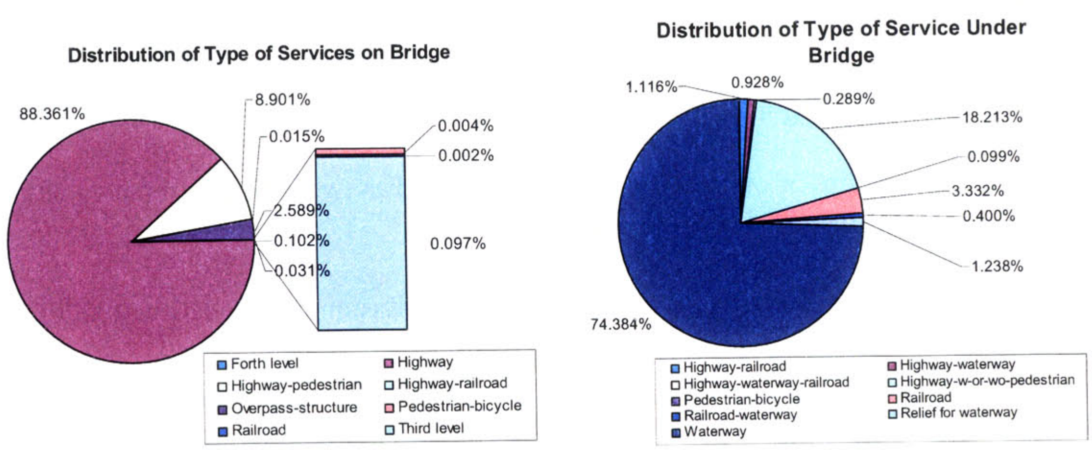


Figure 4-1. Histogram of bridge ages

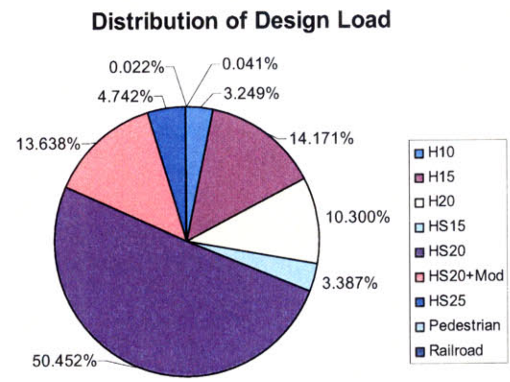
Next the distributions of nominal attributes are shown in Figure 4-2. The dominating populations for type of services on and under bridge are highway and waterway respectively. In term of design load, the newly introduced standard such as H25 or HS25 has not been widely adopted as did H20 and HS20. For construction materials, concrete, prestressed concrete and steel take the majority volume of approximately 30% each. Deck structure of type concrete cast-in-place is predominant. And for physical condition ratings evaluation, ratings of 0-3 only governs less than 5% of the entire population while the majority are ratings of 6-8.



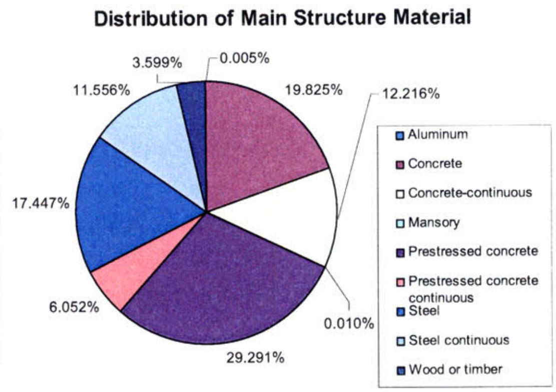
(a) **Type of service on bridge**

(b) **Type of service under bridge**

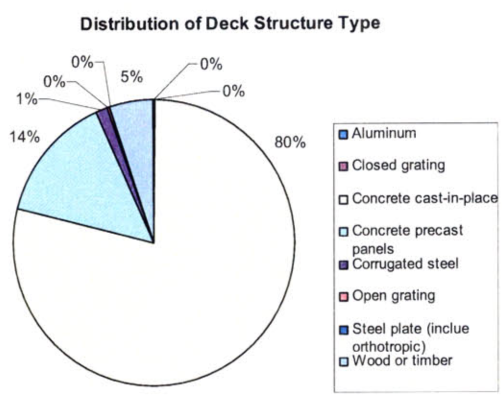
Figure 4-2. Distribution of bridge population by attributes



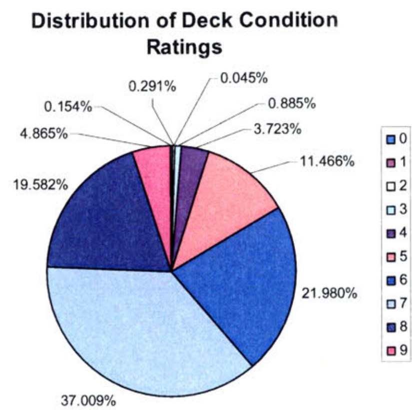
(c) **Design load**



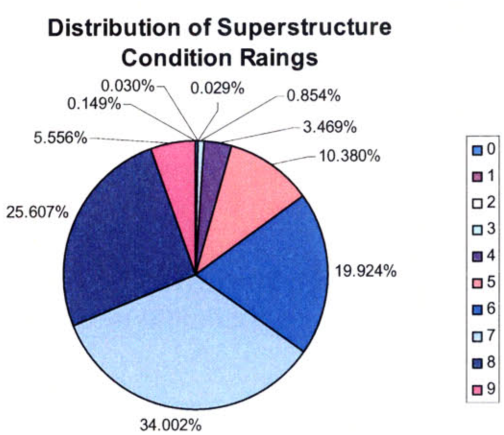
(d) **Main structure material**



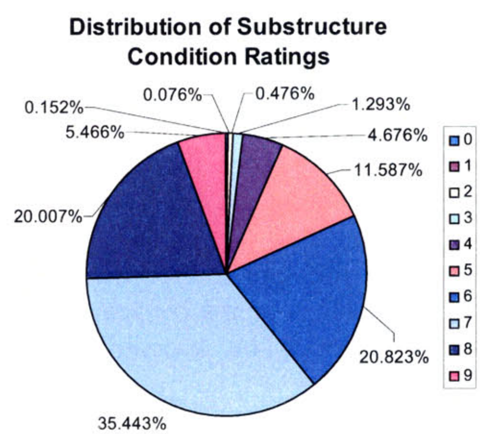
(e) **Deck structure type**



(f) **Deck condition ratings**



(g) **Superstructure condition ratings**



(h) **Substructure condition ratings**

Figure 4-2. Distribution of bridge population by attributes (continued)

4.3.1. Relationships of bridge age to spec/design/performance

Bridge Construction Year vs. Specifications/Service Environments

The highway bridges in NBI data, including overpass at interchange structure, are mostly built in 20th century. Pedestrian and railroad bridges construction period were spanning into 19th century as can be observed in Figure 4-3.

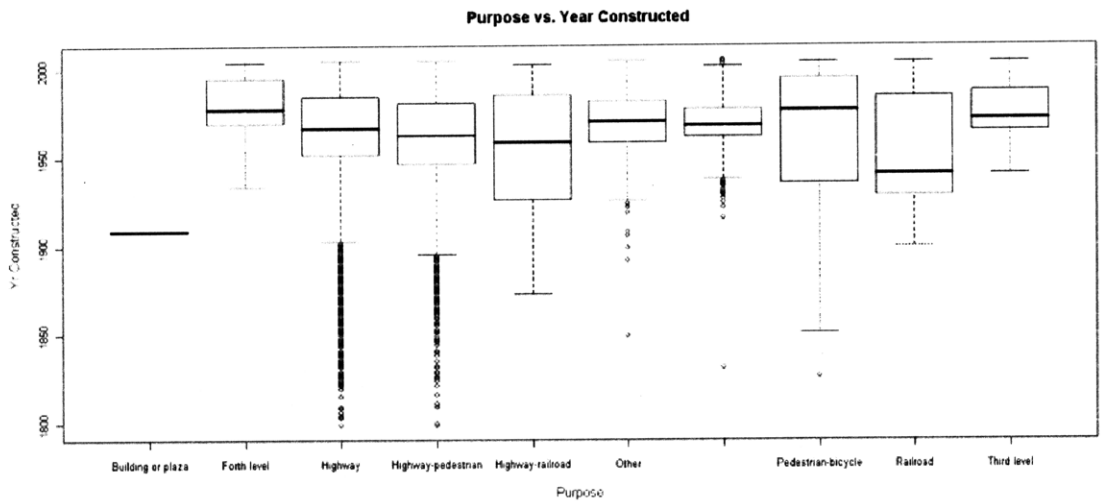


Figure 4-3. Distribution of construction year by bridge purpose (type of service on bridge)

A quantity "Average Daily Traffic per Lane (ADT/Lane)" is defined to reflect normalized traffic volume intensity for comparison purpose among different bridges. Histograms of ADT/Lane by year are plotted to examine distribution characteristics conditioned on different year buckets. The result is shown in Figure 4-4. ADT/Lane quantities of bridges from 1900 to current year have peak frequency at around 0-70 and quickly dissipate after this range. In contrast, bridges prior to 1900 have different distribution patterns for ADT/Lane quantity.

Similarly, a measure for Average Daily Truck Traffic per Lane (ADTT/Lane) is also defined. The histograms by years are plotted in Figure 4-5. Resemble trend to that of ADT/Lane case can be observed in ADTT/Lane.

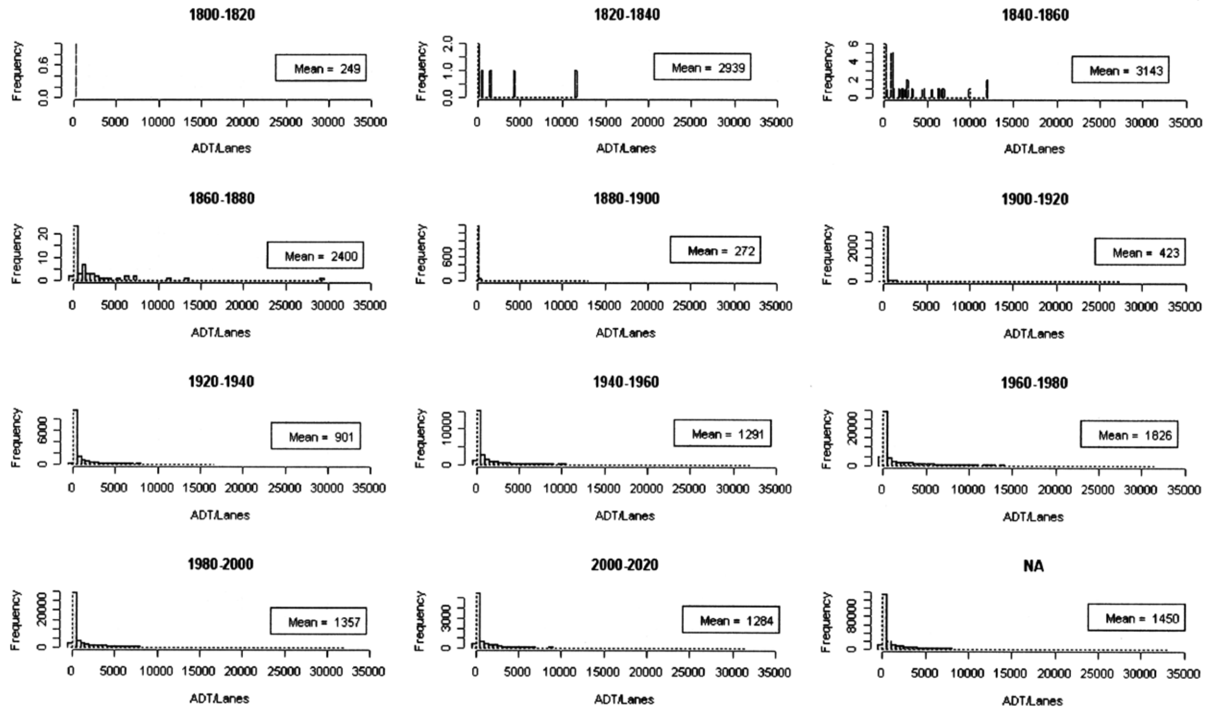


Figure 4-4. Histograms of ADT/Lane by construction years

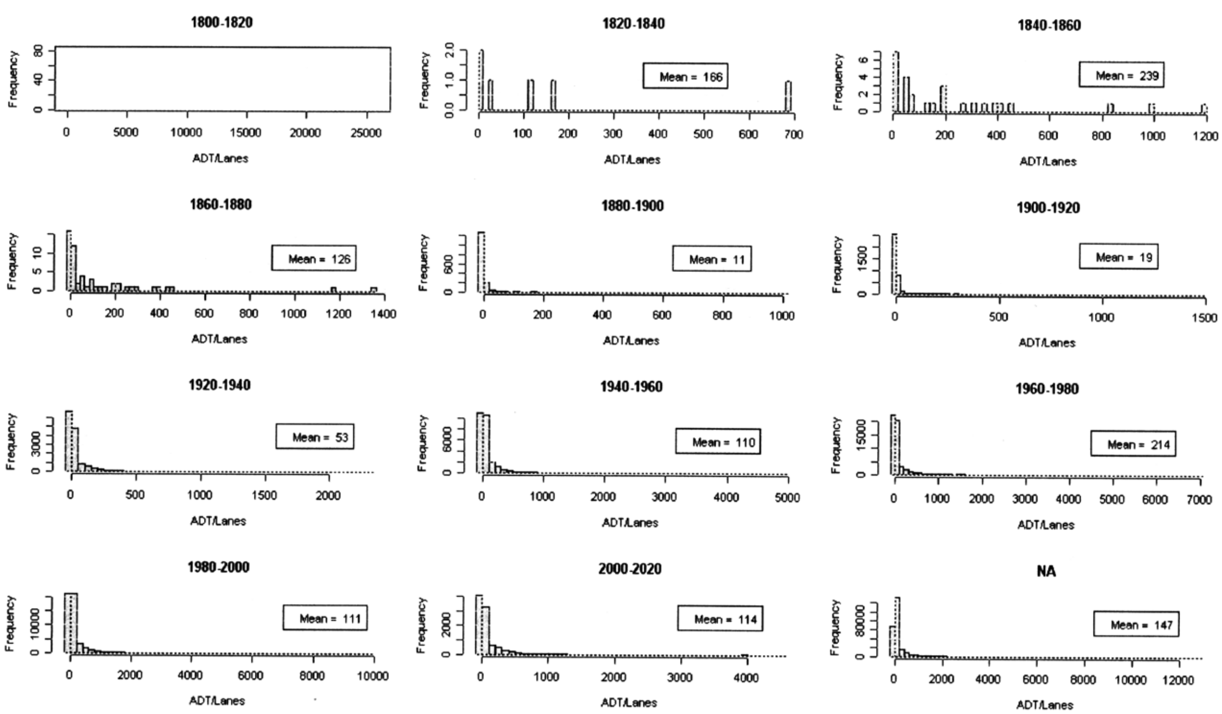


Figure 4-5. Histogram of ADTT/Lan by construction year

To facilitate finding trend of ADT/Lane and ADTT/Lane over constructed year, observed mean value on each year bucket is also noted on the plot in Figure 4-4 and Figure 4-5. Figure 4-6 shows the path of the mean value of ADT/Lane and ADTT/Lane for each year bucket. It is clear from this plot that after 1900, ADT/Lane and ADTT/Lane increase monotonically until 1980 from where the trend starts to move downward. On the other hand, for bridges constructed during 1800s, ADT/Lane and ADTT/Lane mean values have hump shapes in which they peak at 1840-1860 year bucket. Note that the peaks for both ADT/Lane and ADTT/Lane at this year bucket are greater than the peak during 1960-1980.

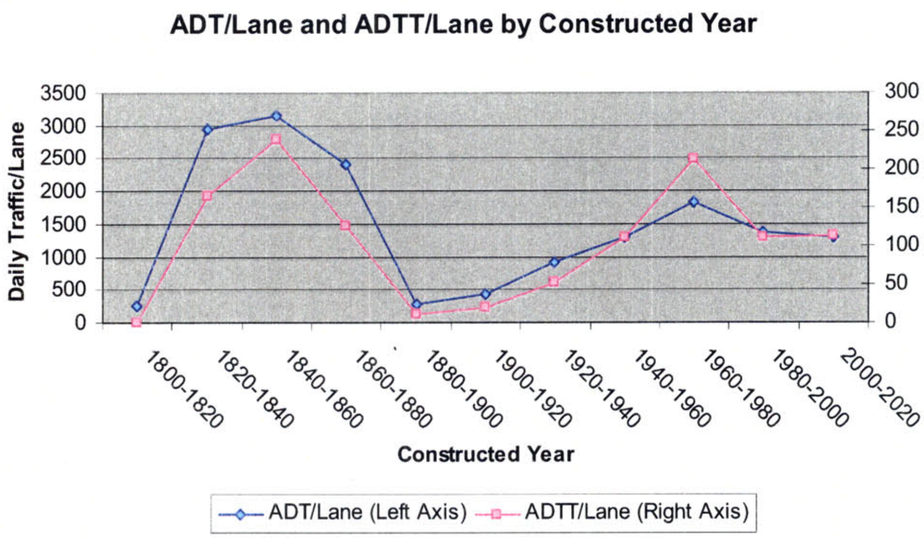


Figure 4-6. Mean value of ADT/Lane and ADTT/Lane by construction years

Another important trend is the relationship between design load assignment and constructed year. Figure 4-7 shows a box plot of constructed year separated by assigned design load. Note that the greater the number after prefix ‘H’ or ‘HS’, the higher load a bridge is assigned to bear for during its service. Figure 4-7 suggests that bigger design load is assigned as year passes by.

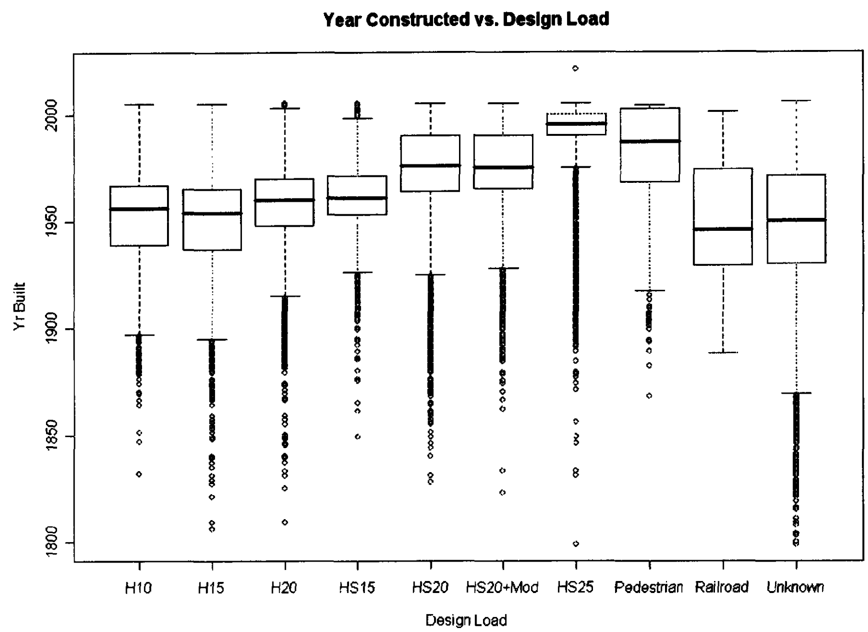


Figure 4-7. Distribution of constructed year by design load

Bridge Construction Year vs. Design Configurations

Figure 4-8 shows a box plot of material of main structure by construction year. It can be seen that mansory or aluminum has been used heavily from 1800s to 1900s. Whereas, modern materials such as steel, concrete and pre-stressed concrete have been used during 1900s, especially after 1950s.

Bridge Construction Year vs. Performance

Figure 4-9 shows box plots of distribution of construction year conditioned by three types of physical condition rating evaluations (deck, superstructure and substructure). From this figure, it is clear that construction years of low rating (0-3) bridges center at around 1950s and gradually increasing to 1990s when reaching highest rating of 9. Also, it is noteworthy that bridges of low rating tend to have widely distributed range of construction year compared to those with high ratings (4-9).

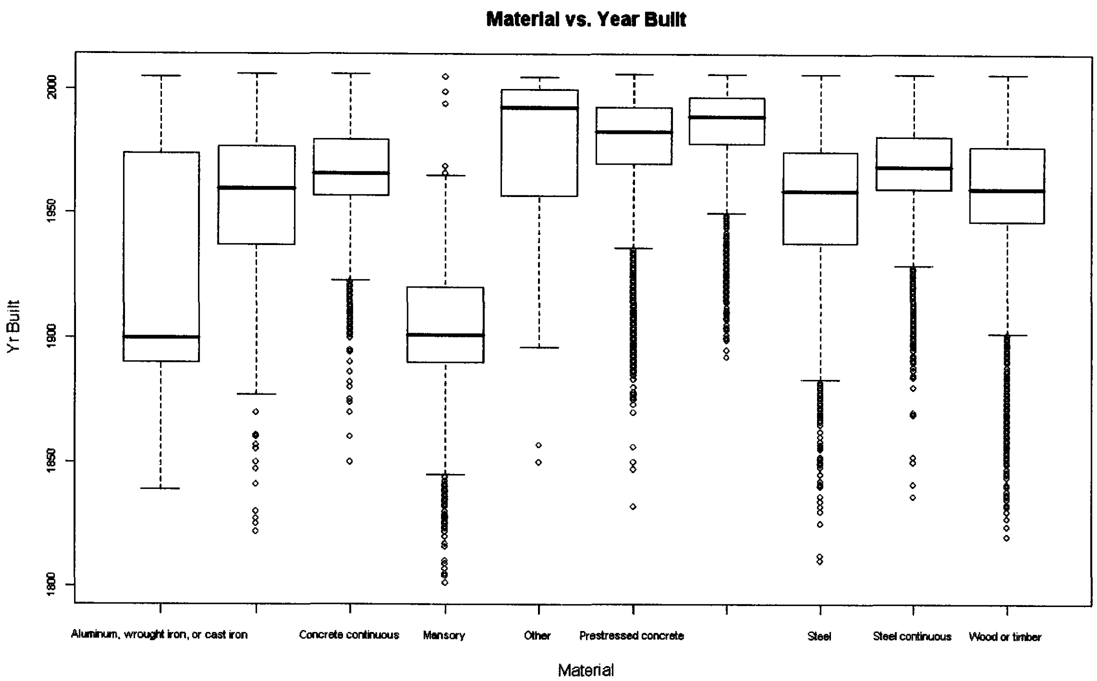
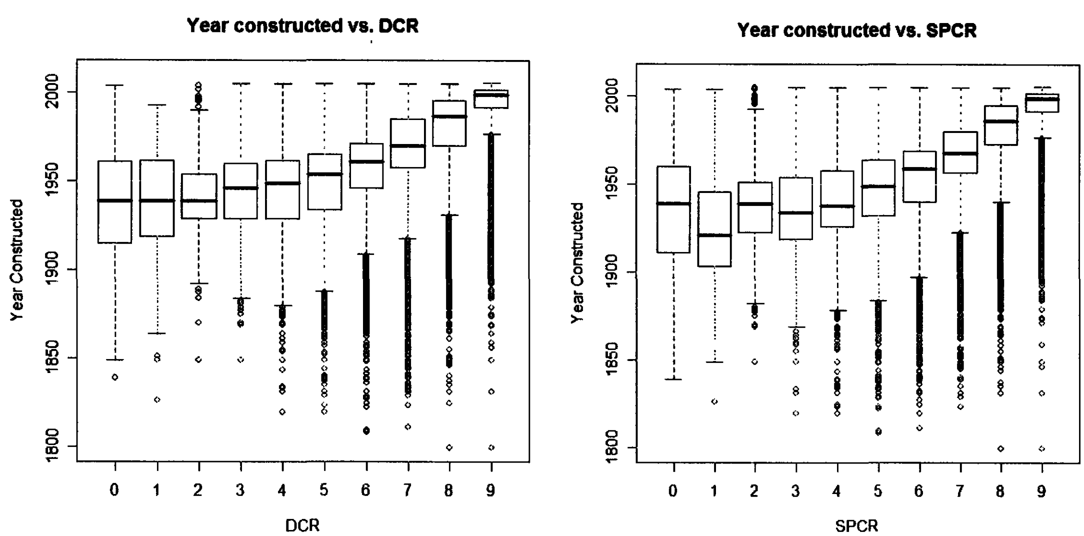
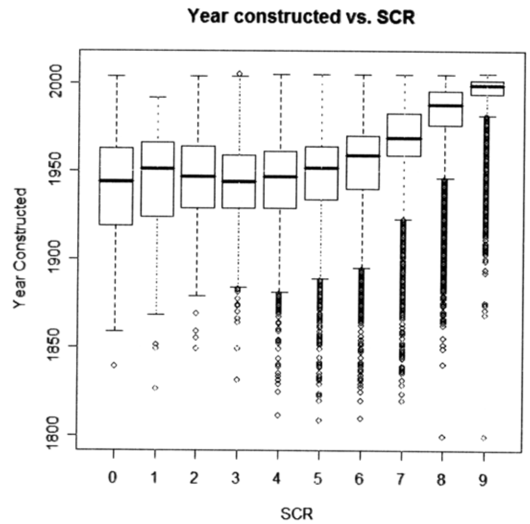


Figure 4-8. Distribution of construction year by material type



(a) DCR **(b) SPCR**

Figure 4-9. Conditional distributions of construction year on physical condition rating



(c) SCR

Figure 4-9. Conditional distributions of construction year on physical condition rating (continued)

Summary of the overall relationship between bridge ages (or construction years) to service environment, specifications, design configurations and performances are summarized in Table 4-7 below.

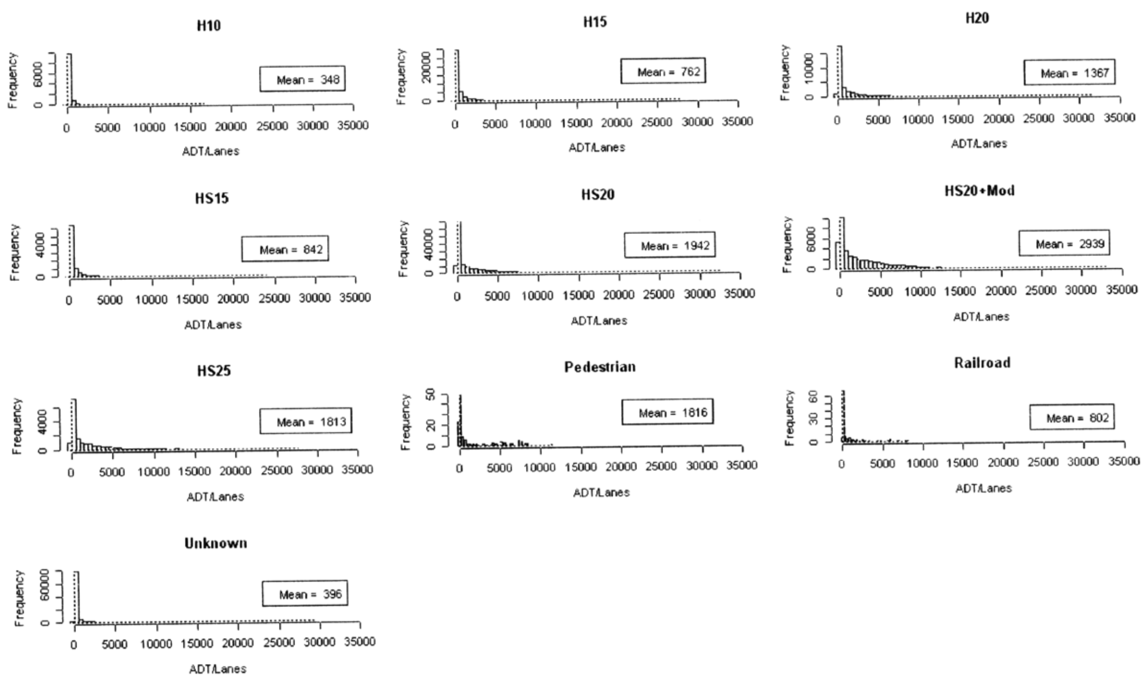
Table 4-7. Summary of relationship of construction year and service environments, specifications, design configurations and performances

	1800	1820	1840	1860	1880	1900	1920	1940	1960	1980	2000
Bridge Purpose	Highway/Pedestrian		Pedestrian/bicycle			Railroad	Interchange				
Design Load	H15 and H20	H10 and HS20	HS15, HS25 and Pedestrian			Railroad					
ADT/Lane	249	2939	3143	2400	272	423	901	1291	1826	1357	1284
ADTT/Lane	0	166	239	126	11	19	53	110	214	111	114
Main Material	Steel and Mansory	Concrete	Aluminum					Prestressed concrete			

4.3.2. Relationships between service environments and specifications

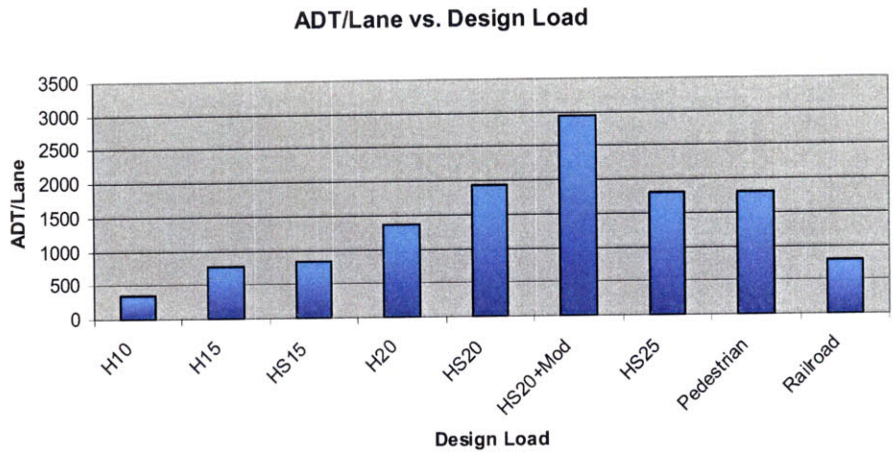
To grasp a basic set of relationships among specification elements of the bridges, firstly, conditioned distributions are plotted on ADT/Lane conditioned by other specifications. The merit of finding relationship between ADT/Lane measure and other specification elements is due to the fact that, intuitively, this quantity is one of the main factors that affect bridge structural condition.

Starting with distribution of ADT/Lane by design load plotted in Figure 4-10, the ADT/Lane's, again, have highest frequency at 0-70 and quickly dissipate away. It can also be seen that majority of the bridges are designed under H20, HS20, and HS20+Mod codes. The numbers after prefixes H and HS indicate weight of the truck used for load testing. The greater the number, the heavier the truck is. As expect, it can be seen that the mean ADT/Lane increases with the design load except for HS25. This is because generally bridges are designed carry a 72,000-lb design load (HS20). The HS25 is recently adopted. According to Chase (2003), bridges after 2001 have been designed under HS25 code for one in every five bridges.



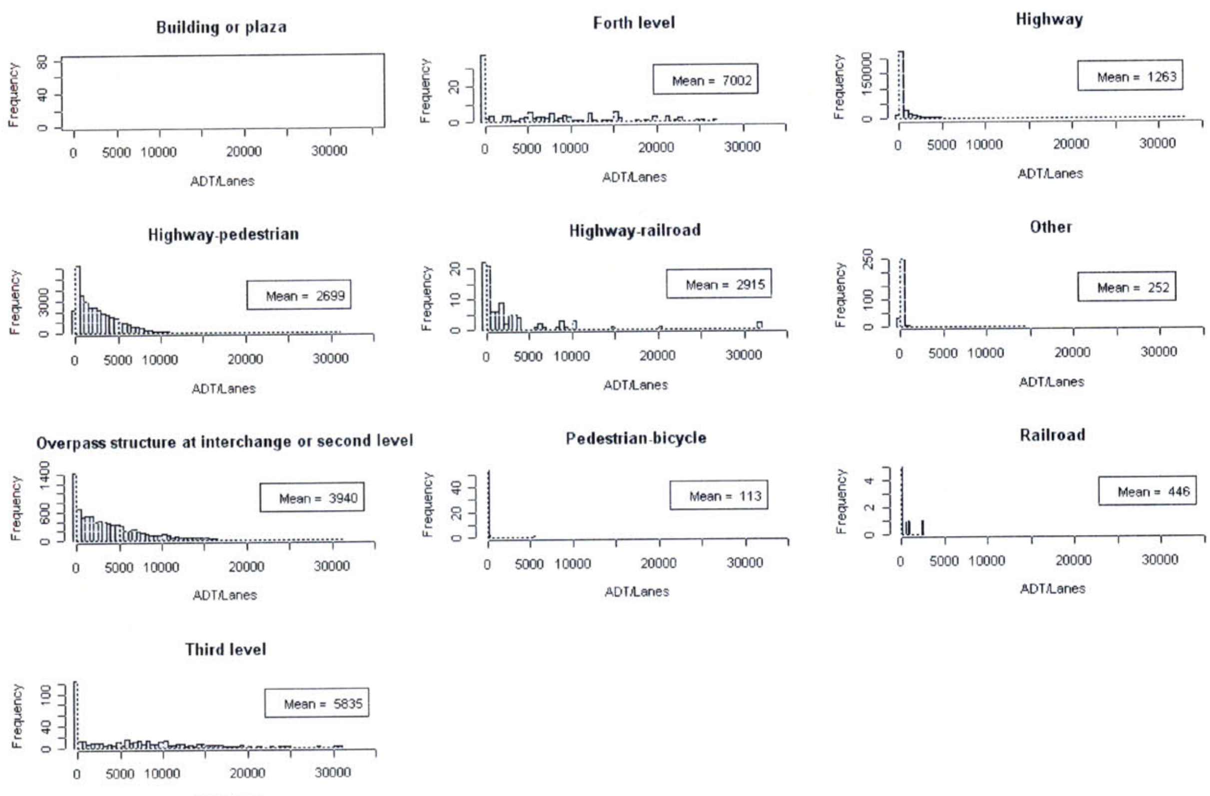
(a) Histograms of ADT/Lane by Design Load

Figure 4-10. Distribution of ADT/Lane conditioned on design load



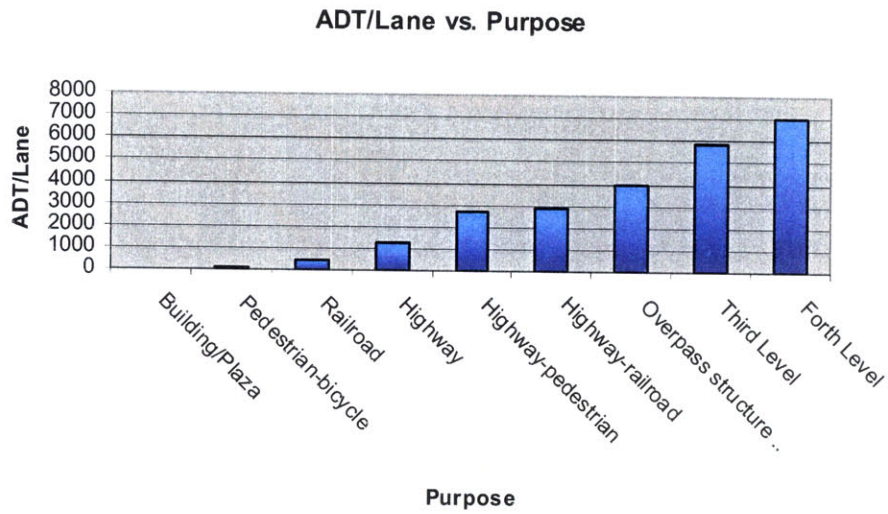
(b) Mean values of ADT/Lane by Design Load

Figure 4-10. Distribution of ADT/Lane conditioned on design load (continued)



(a) Histograms of ADT/Lane by Bridge Purpose

Figure 4-11. Distribution of ADT/Lane conditioned on purpose



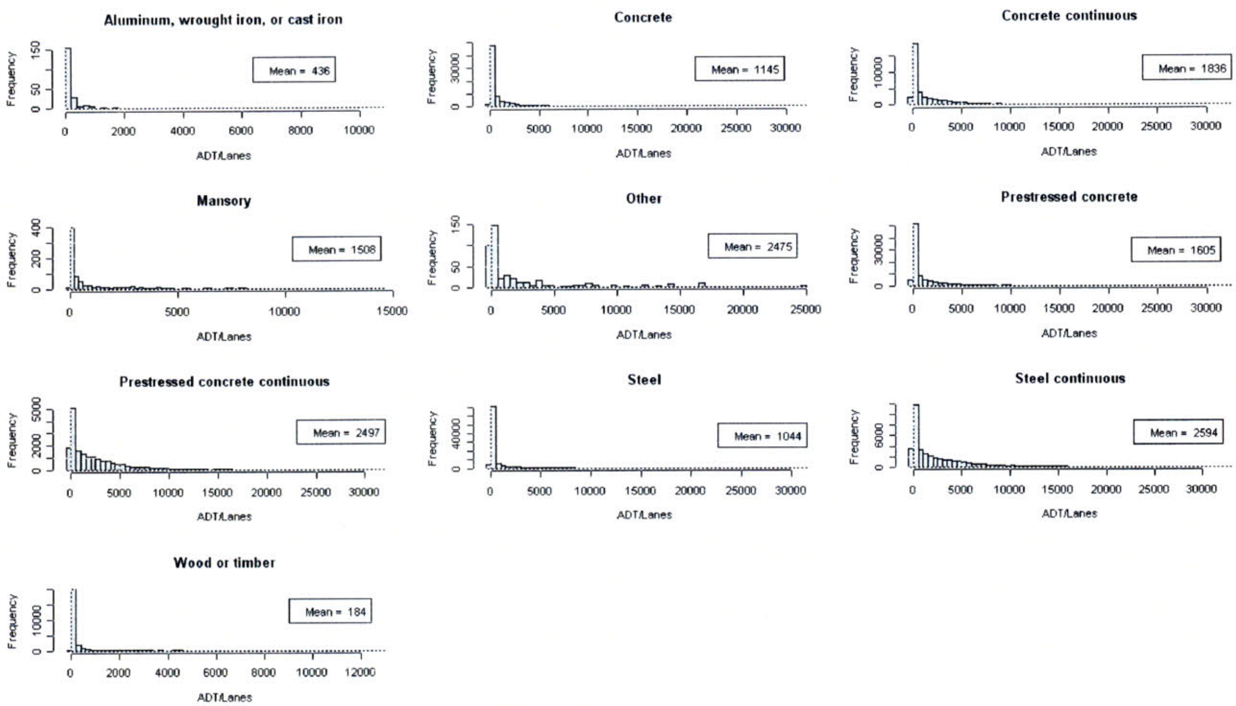
(b) Mean values of ADT/Lane by Bridge Purpose

Figure 4-11. Distribution of ADT/Lane conditioned on purpose (continued)

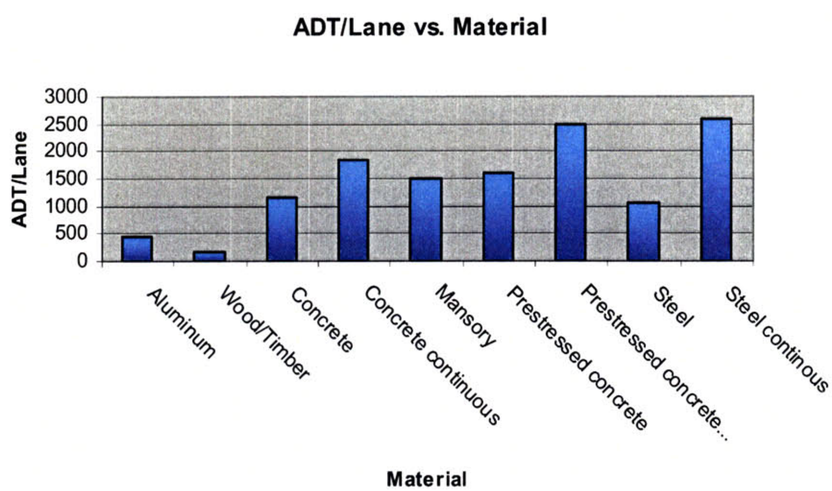
ADT/Lane conditioned on bridge purpose (Figure 4-11) also demonstrates the same trend of distribution. However, the mean values indicate that overpass structure at interchanges tend to have higher volume of ADT/Lane.

4.3.3. Relationships between specifications and designs

Now we examine relationship between bridge specification and design description. The major field for specification to be used here is ADT/Lane measure. We start with visualizing distribution of ADT/Lane by material type shown in Figure 4-12. For all material types, highest frequency appears at very small ADT/Lane, i.e. about 0-70 like in other cases. The majority of the bridges were designed with concrete and steel. Average ADT/Lane observed is greatest in steel continuous and prestressed concrete continuous types. Although concrete is the majority material of bridge design, concrete bridges do not possess high average ADT/Lane.



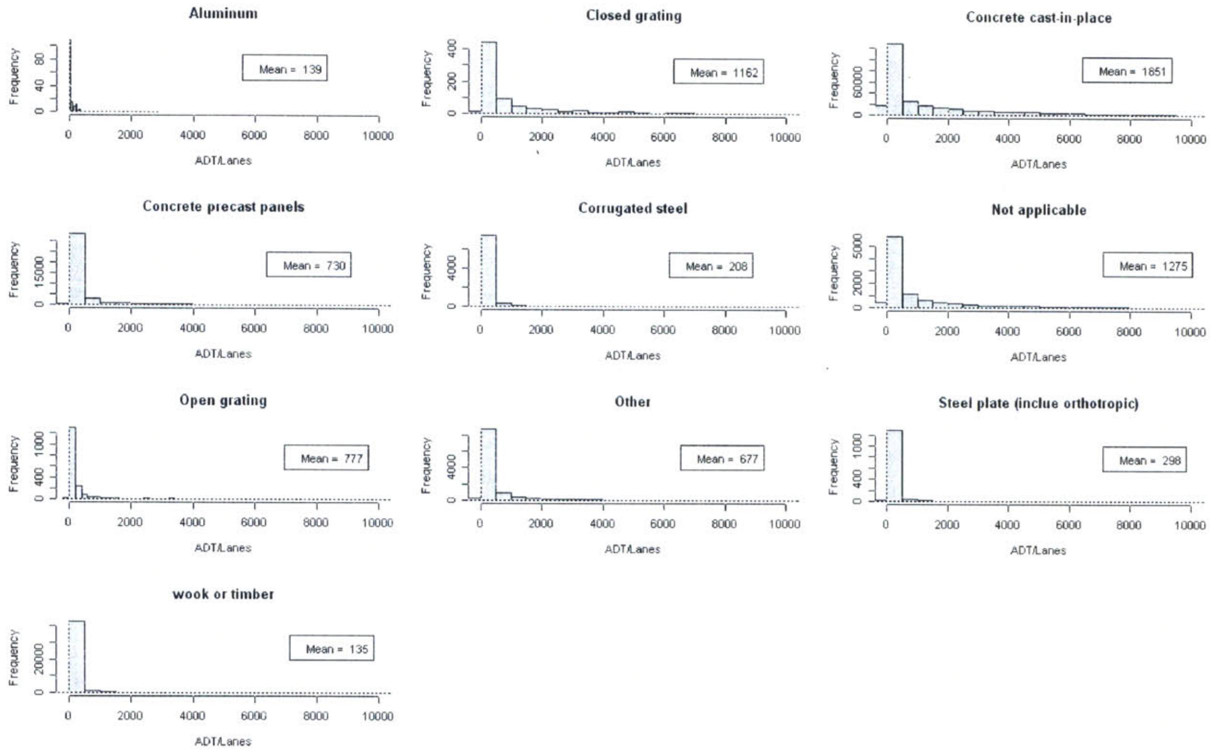
(a) Histograms of ADT/Lane by material



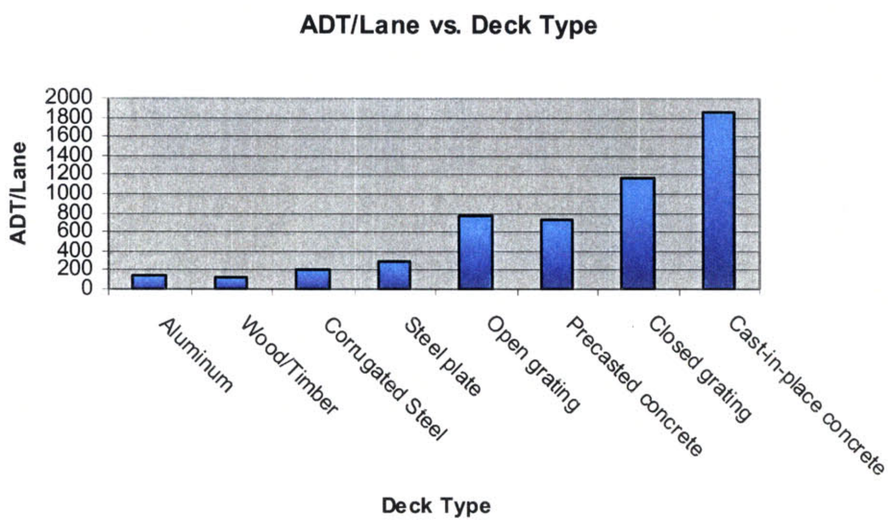
(b) Mean values of ADT/Lane by material

Figure 4-12. Distribution of ADT/Lane conditioned on material

By deck type, precasted concrete and cast-in-place concrete decks have highest ADT/Lane masses (Figure 4-13). However, cast-in-place concrete and closed-grating deck have substantially high ADT/Lane mean values compared to other deck types.



(a) Histograms of ADT/Lane by Deck Type



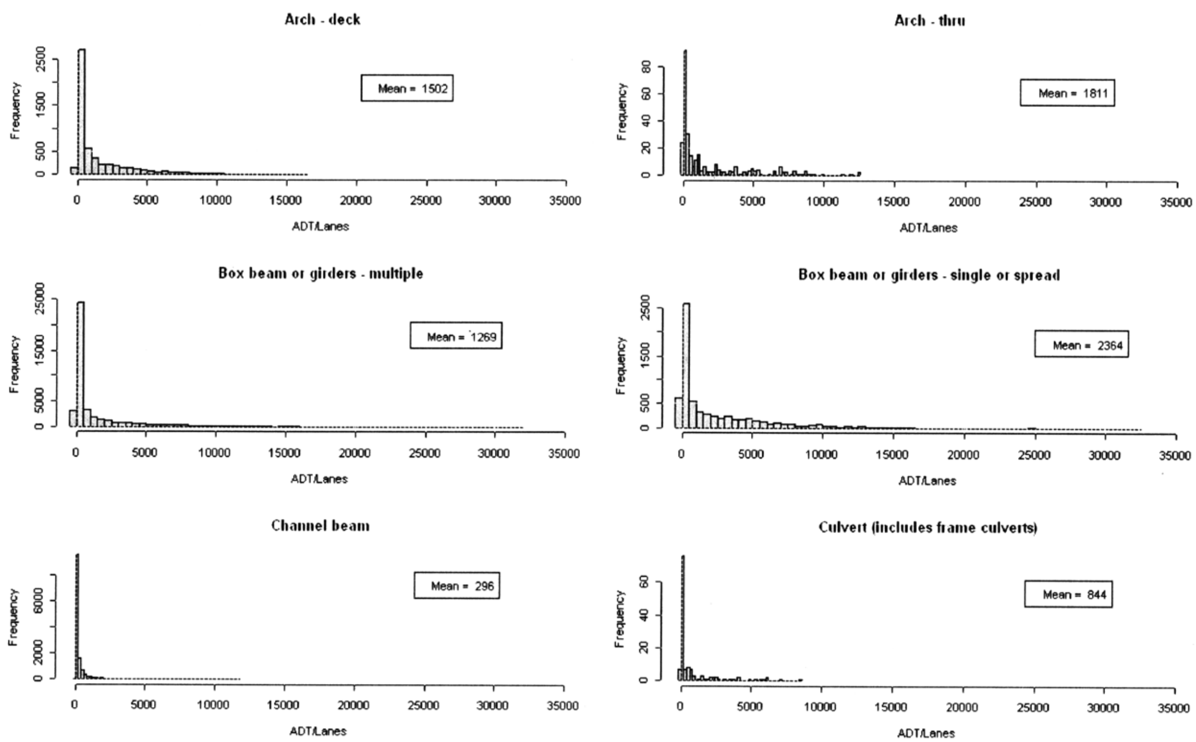
(b) Mean values of ADT/Lane by Deck Type

Figure 4-13. Distribution of ADT/Lane conditioned on type of deck

By main structural design type (Figure 4-14), major design types include

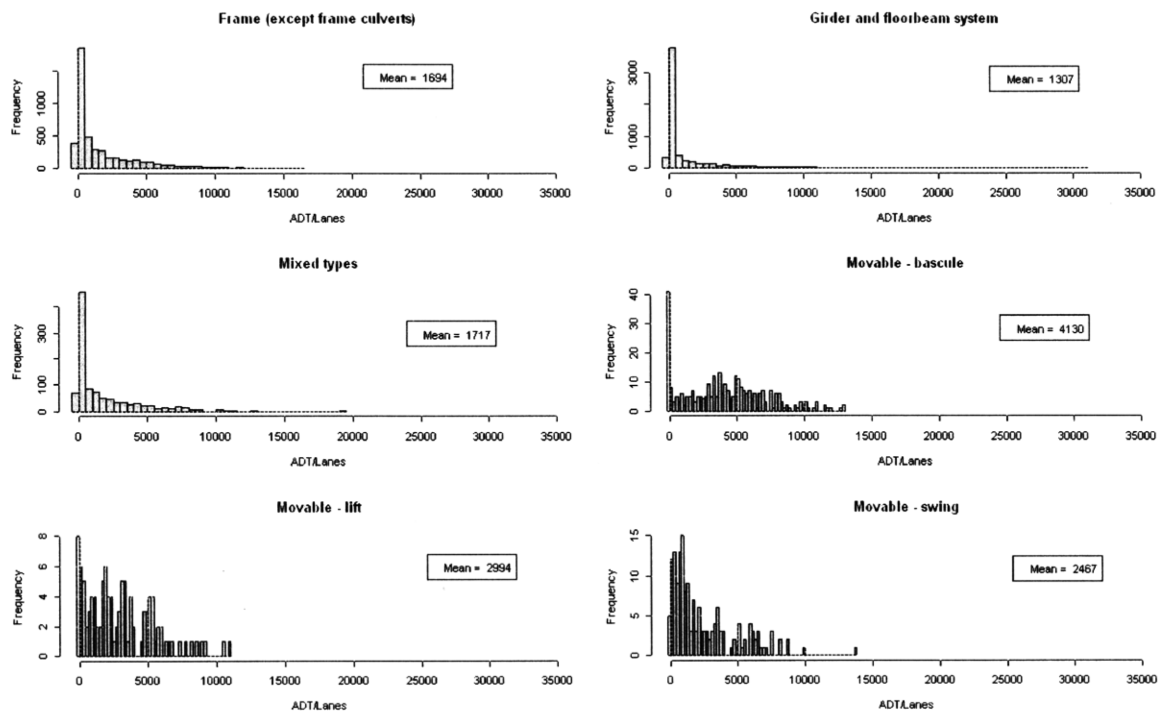
- Box beam or girders (multiple)
- Slab
- Stringer/Multi-beam or girder
- Tee beam
- Truss (thru)

Note the irregular distribution shapes of movable, segmental box girder, stayed girder, suspension and tunnel types. Though peculiar distribution shapes and their limited number of instances, segmental box girder and stayed girder structural designs demonstrate highest observed ADT/Lane mean values.

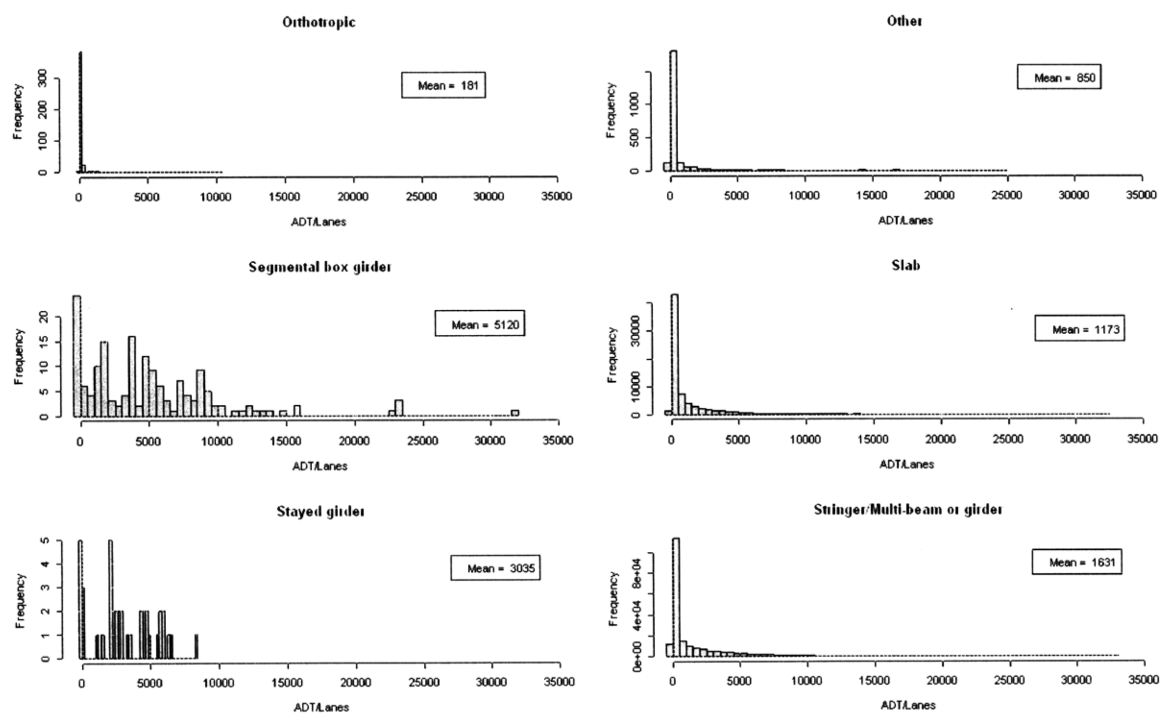


(a) Histograms of ADT/Lane by main structure type

Figure 4-14. Distribution of ADT/Lane conditioned on structure type of the deck

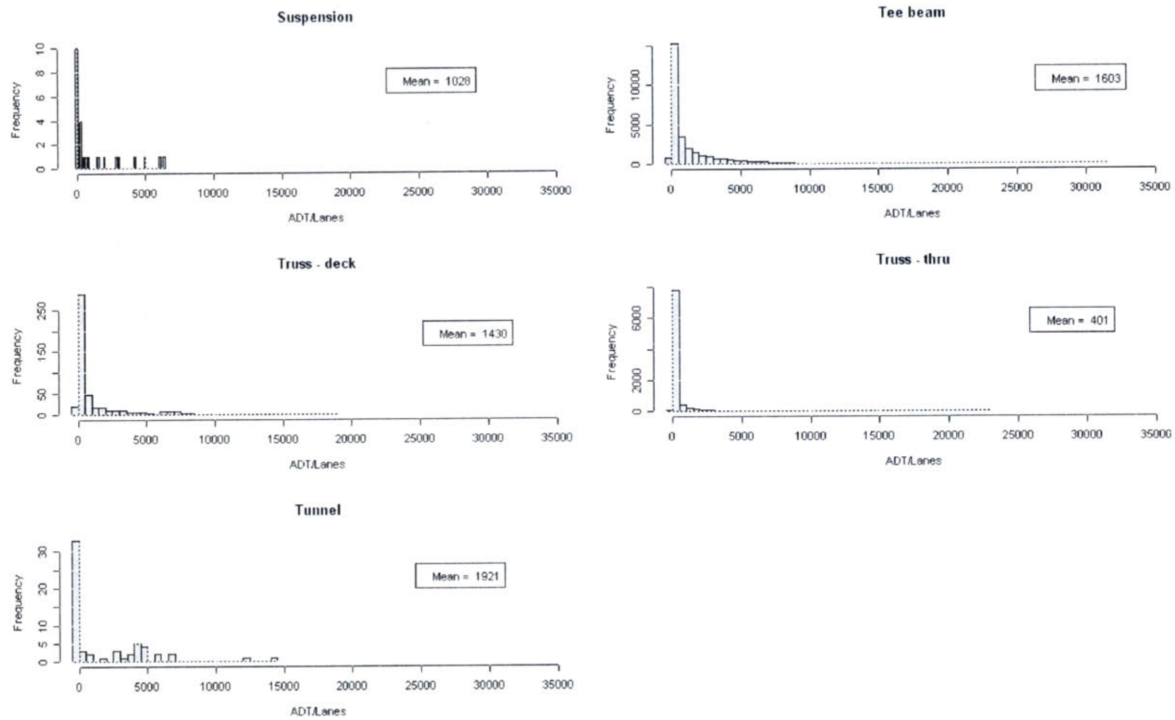


(b) Histograms of ADT/Lane by Main Structure Type

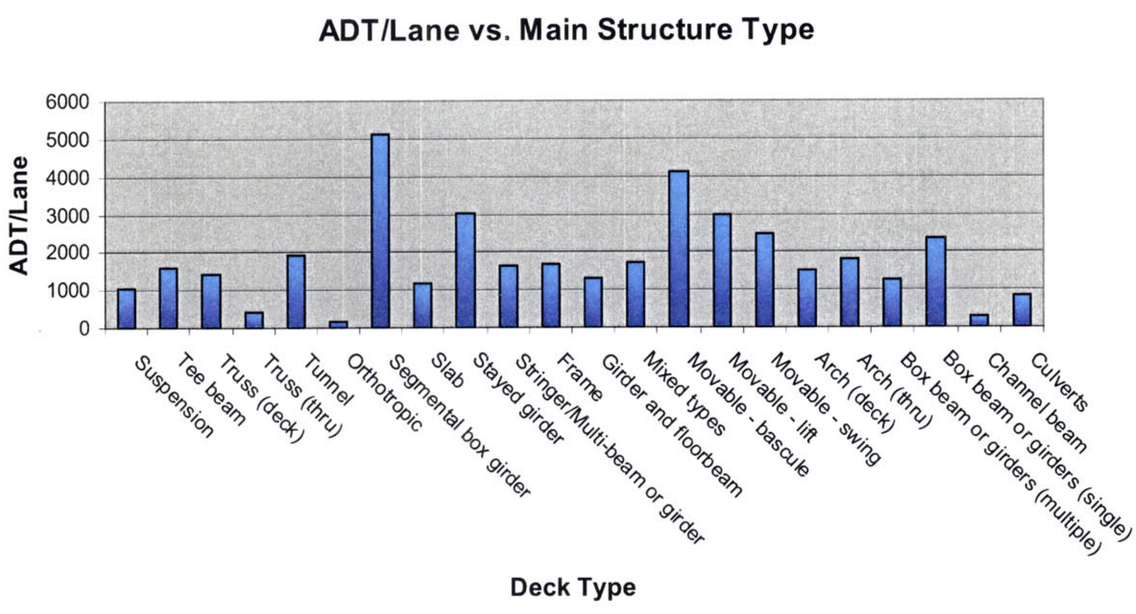


(c) Histograms of ADT/Lane by Main Structure Type (continued)

Figure 4-14. Distribution of ADT/Lane conditioned on structure type of the deck (continued)



(d) Histograms of of ADT/Lane by Main Structure Type (continued)



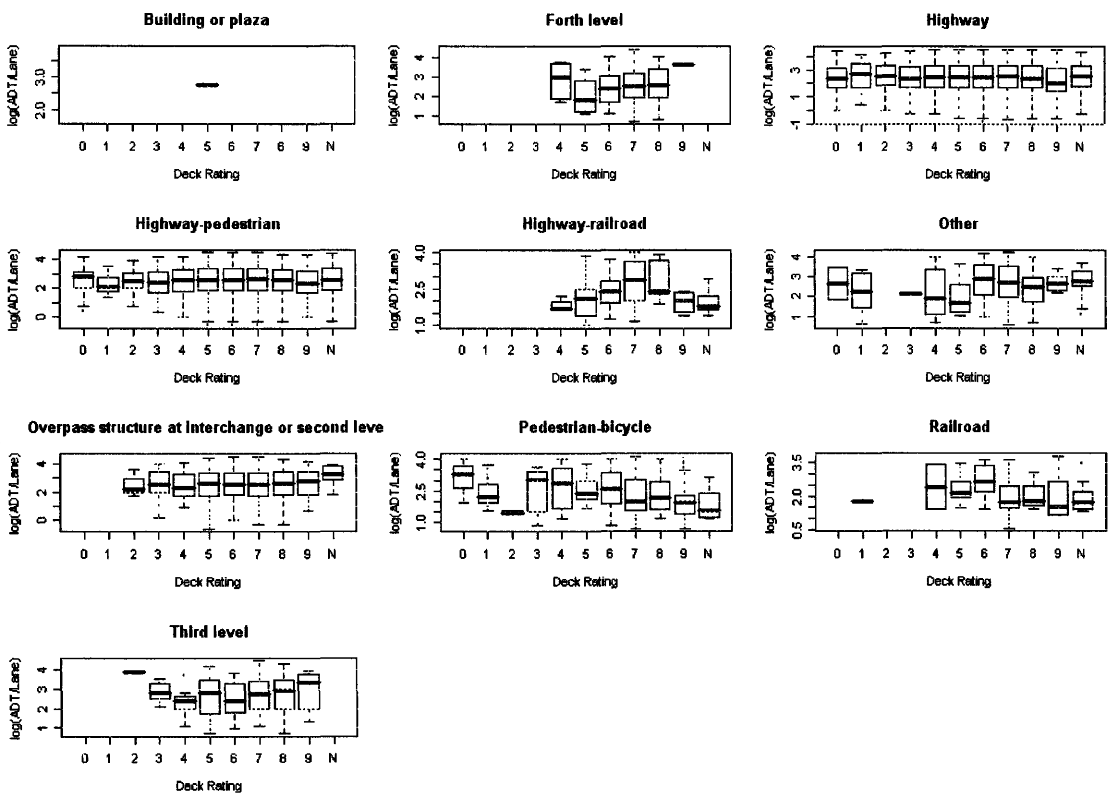
(d) Mean values of ADT/Lane by Main Structure Type

Figure 4-14. Distribution of ADT/Lane conditioned on structure type of the deck (continued)

4.3.4. Effects of specifications and designs on performances

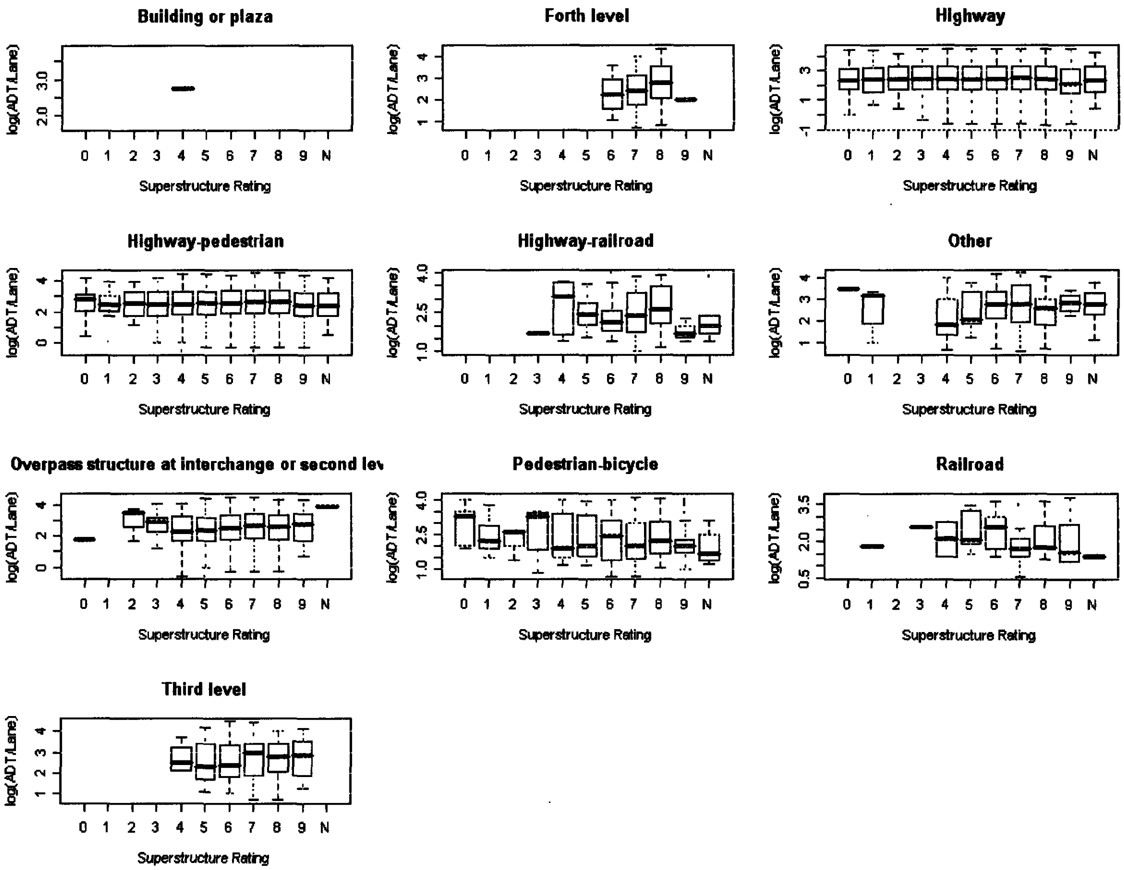
NBI data contains item 58 to 60 that pertain condition ratings of deck, superstructure and substructure of bridges. These items are defined by 10 level scales, i.e. 0 to 9. 0 indicates failed condition whereas 9 means excellent condition. Structural deficiency can be defined as bridges with values of these items less than or equal to 4.

Denoting DCR, SPCR, SCR for deck, superstructure and substructure condition rating respectively, Figure 4-15 plots distribution of log ADT/Lane (base 10) versus the three condition ratings by type of services.



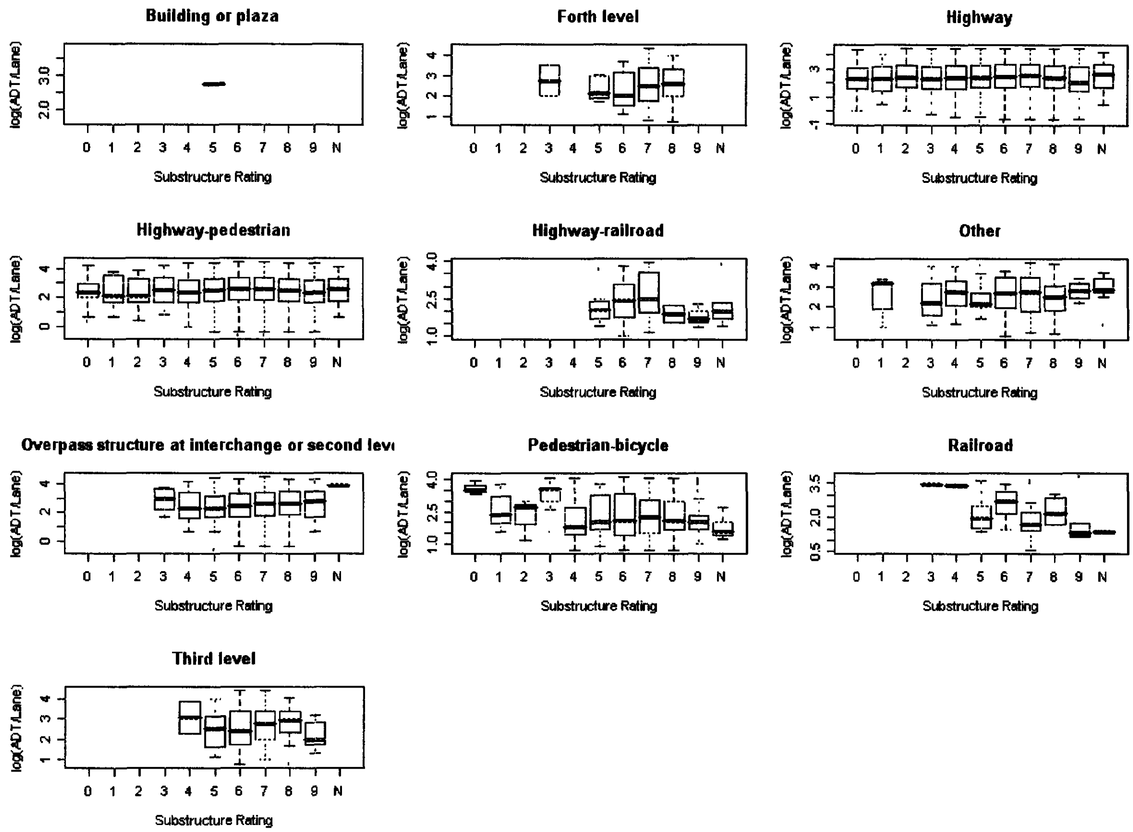
(a) DCR

Figure 4-15. Log(ADT/Lane) vs. condition rating by purpose



(b) SPCR

Figure 4-15. Log(ADT/Lane) vs. condition rating by purpose (continued)

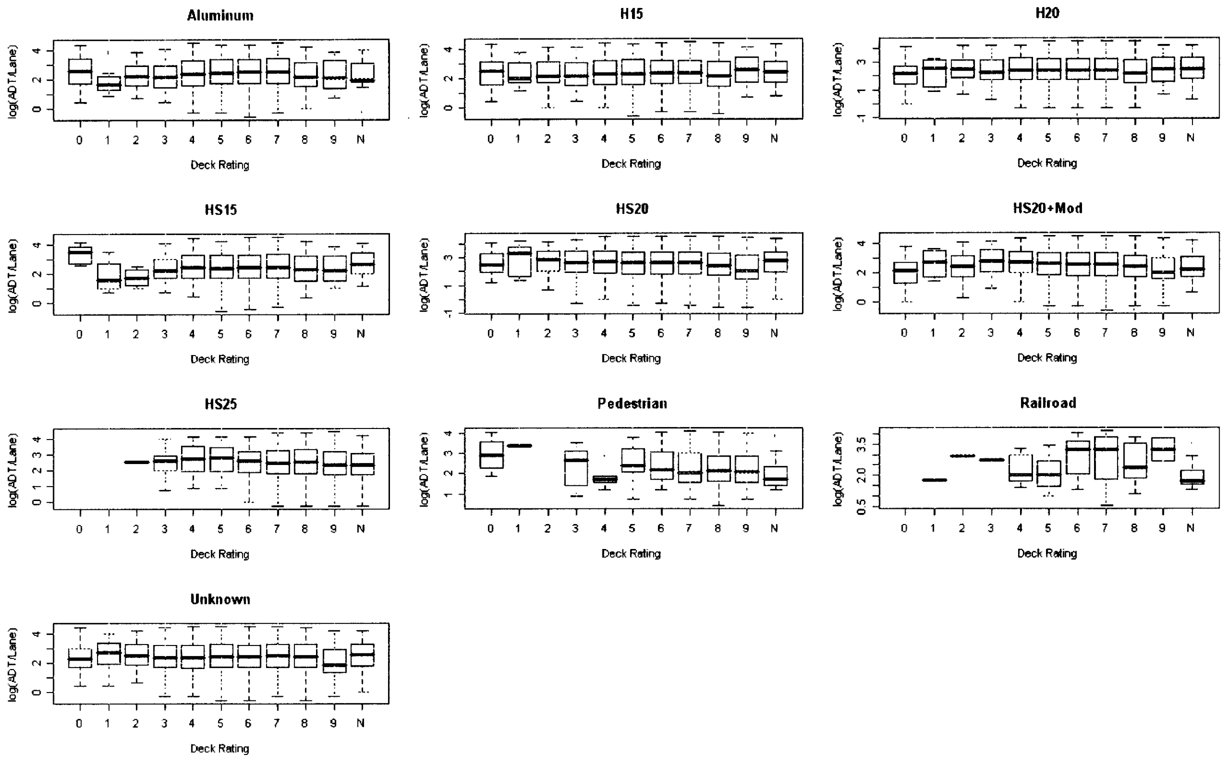


(c) SCR

Figure 4-15. Log(ADT/Lane) vs. condition rating by purpose (continued)

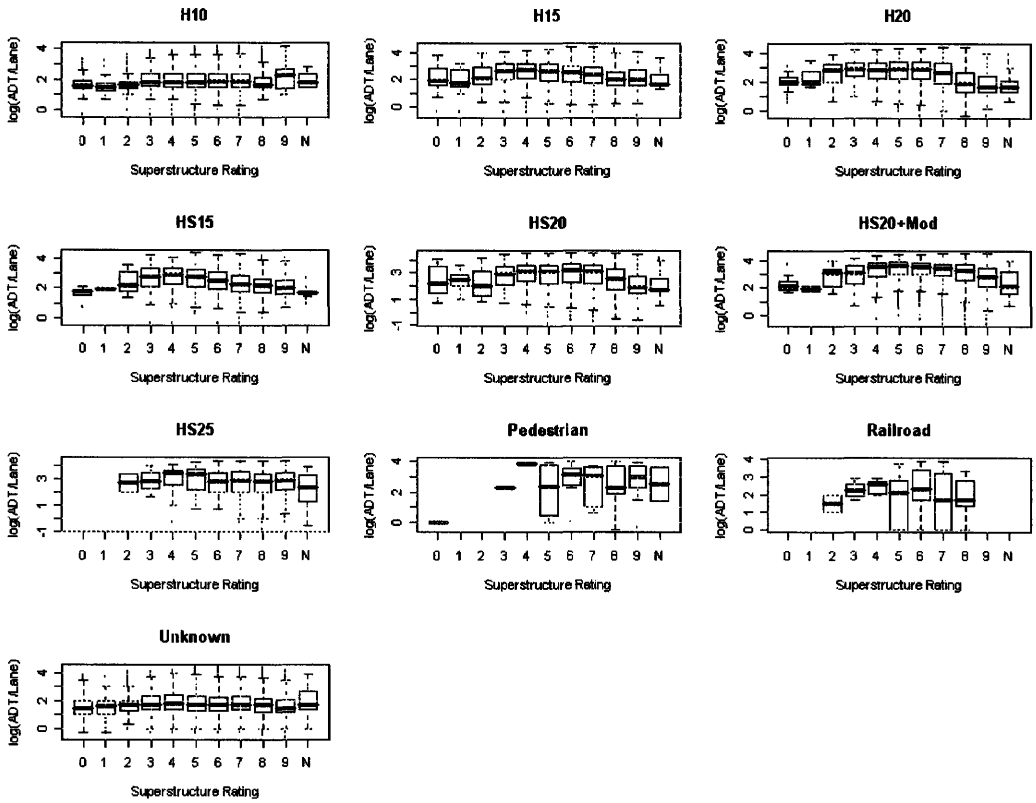
The overall trends from these three plots suggest that structural deficiency concentrates in highway and pedestrian type of bridges. For highway bridges, the ADT/Lanes amount does not vary much among ratings. On the other hand, the pedestrian bridges have relatively high ADT/Lane volume for those found structurally deficient.

Next is the trend for structural deficiency categorized by design load codes. Box plots for DCR, SPCR, and SCR are provided in Figure 4-16. As expected, it can be observed from these plots that the carry load a bridge is designed for, the higher ADT/Lane volume that would cause the deficiency (rating 0-4). Note the lack of structurally deficient instances of bridges designed by HS25 code. This is because the code is relatively new and not many bridges have been designed under this load.



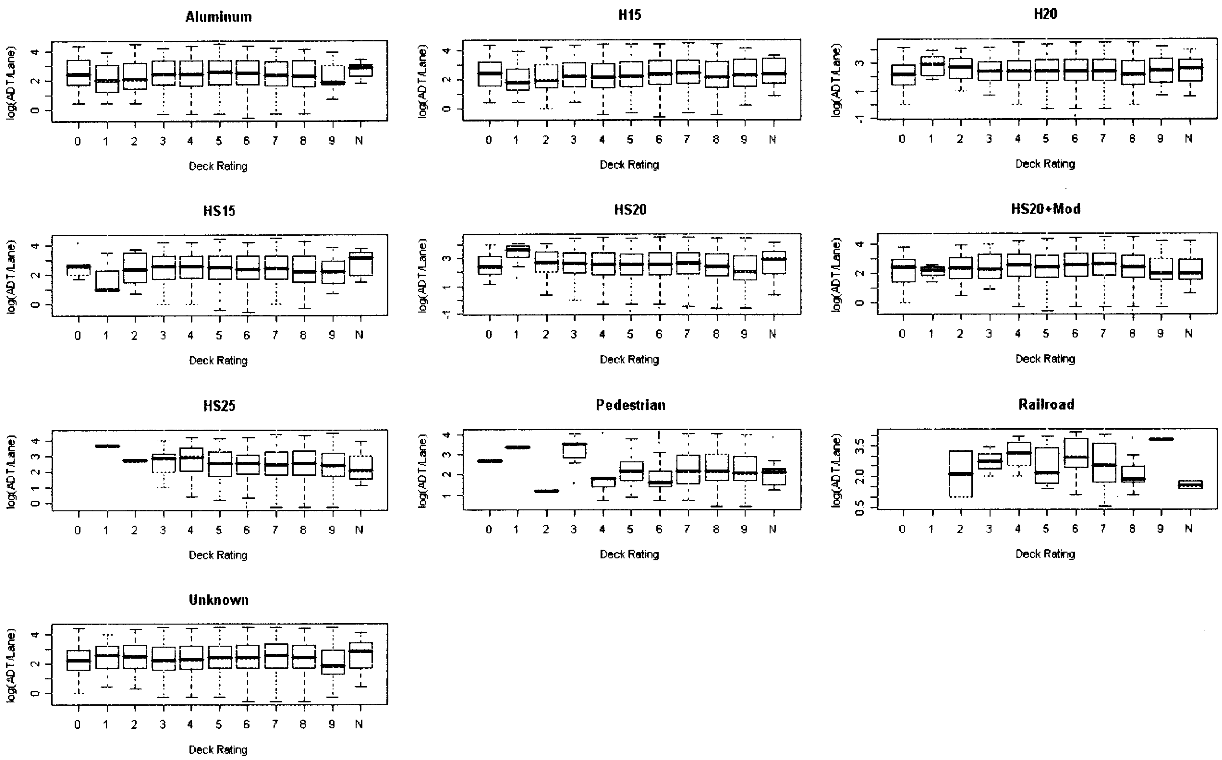
(a) DCR

Figure 4-16. Log(ADT/Lane) vs. condition rating by design load



(b) SPCR

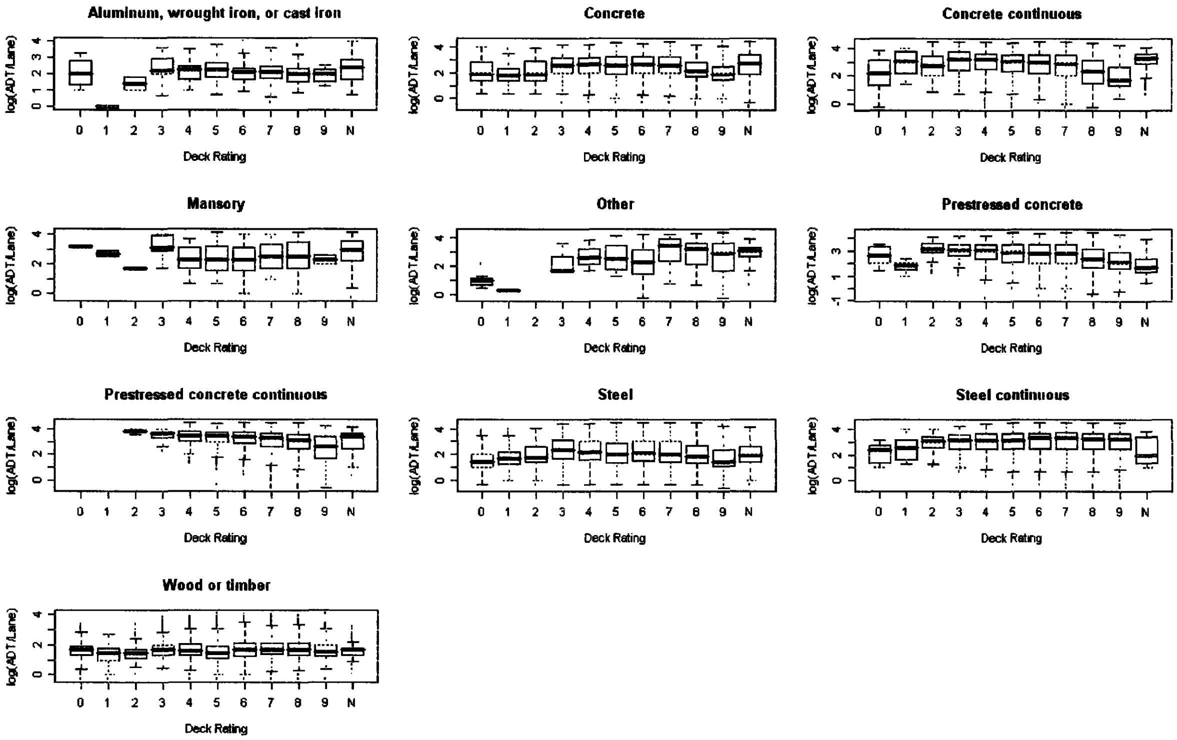
Figure 4-16. Log(ADT/Lane) vs. condition rating by design load (continued)



(c) SCR

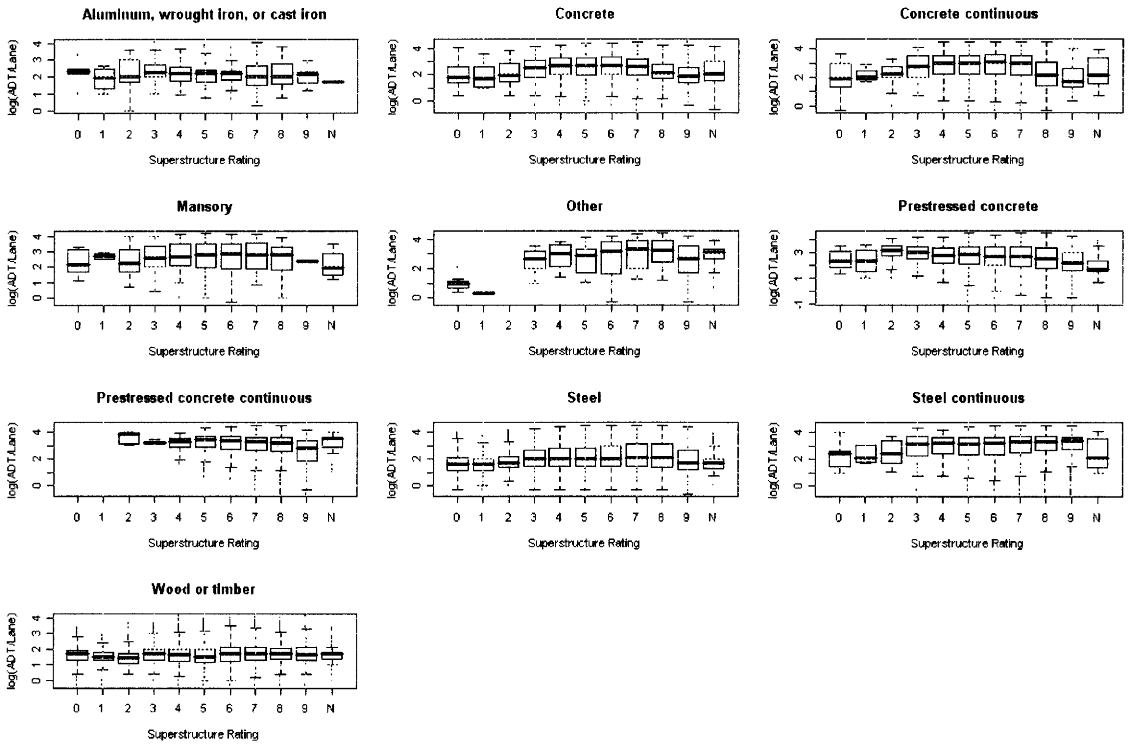
Figure 4-16. Log(ADT/Lane) vs. condition rating by design load (continued)

In term of effects of traffic volume, ADT/Lane and design choice to bridge structural condition, similar box plots on bridge material has been created in Figure 4-17. These plots suggest high traffic volume on prestressed and steel continuous bridges. Note also that the traffic volume that causes deficiency in prestressed concrete is higher than other types of materials.



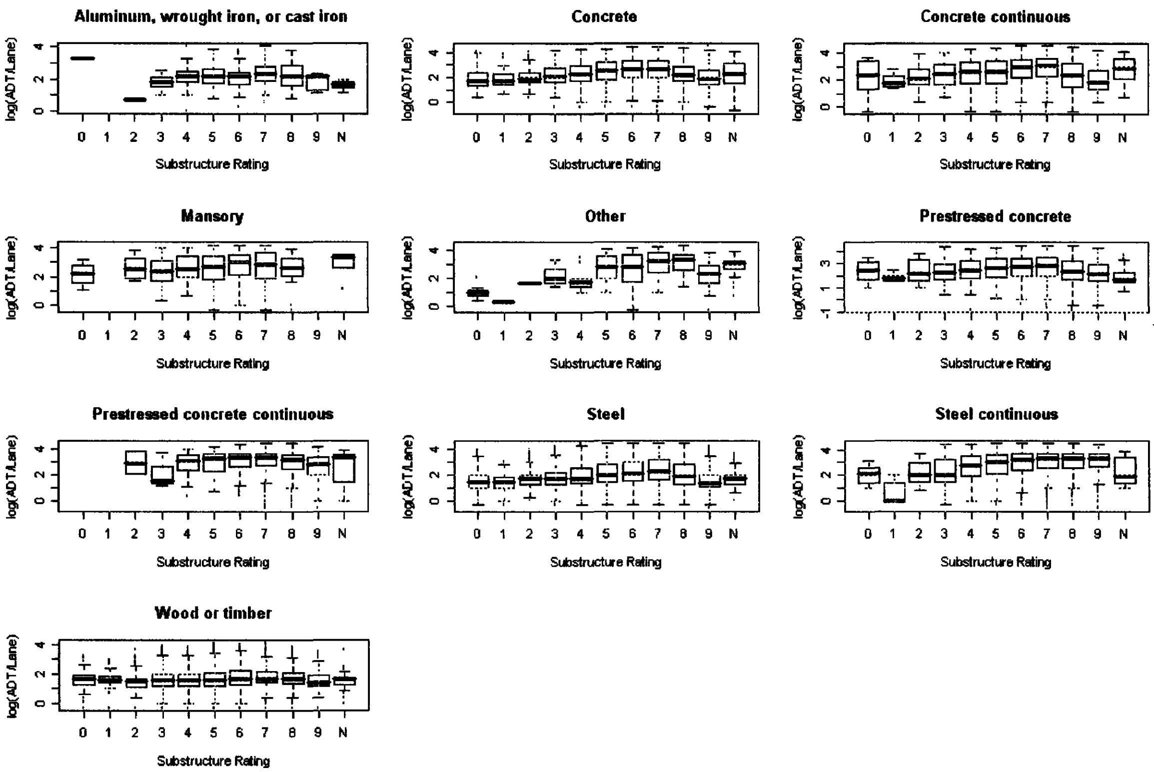
(a) DCR

Figure 4-17. Log(ADT/Lane) vs. condition rating by material



(b) SPCR

Figure 4-17. Log(ADT/Lane) vs. condition rating by material (continued)

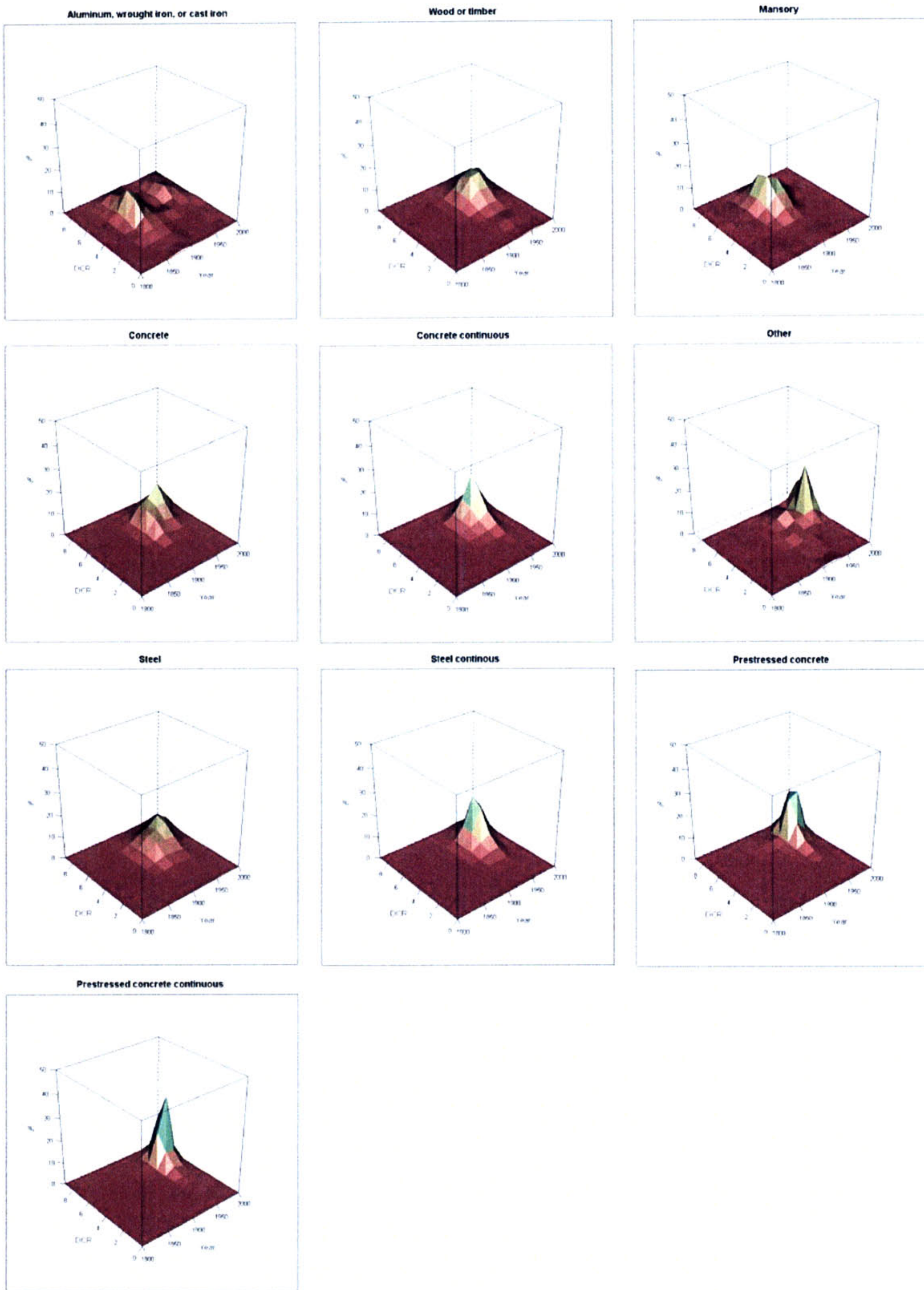


(c) SCR

Figure 4-17. Log(ADT/Lane) vs. condition rating by material (continued)

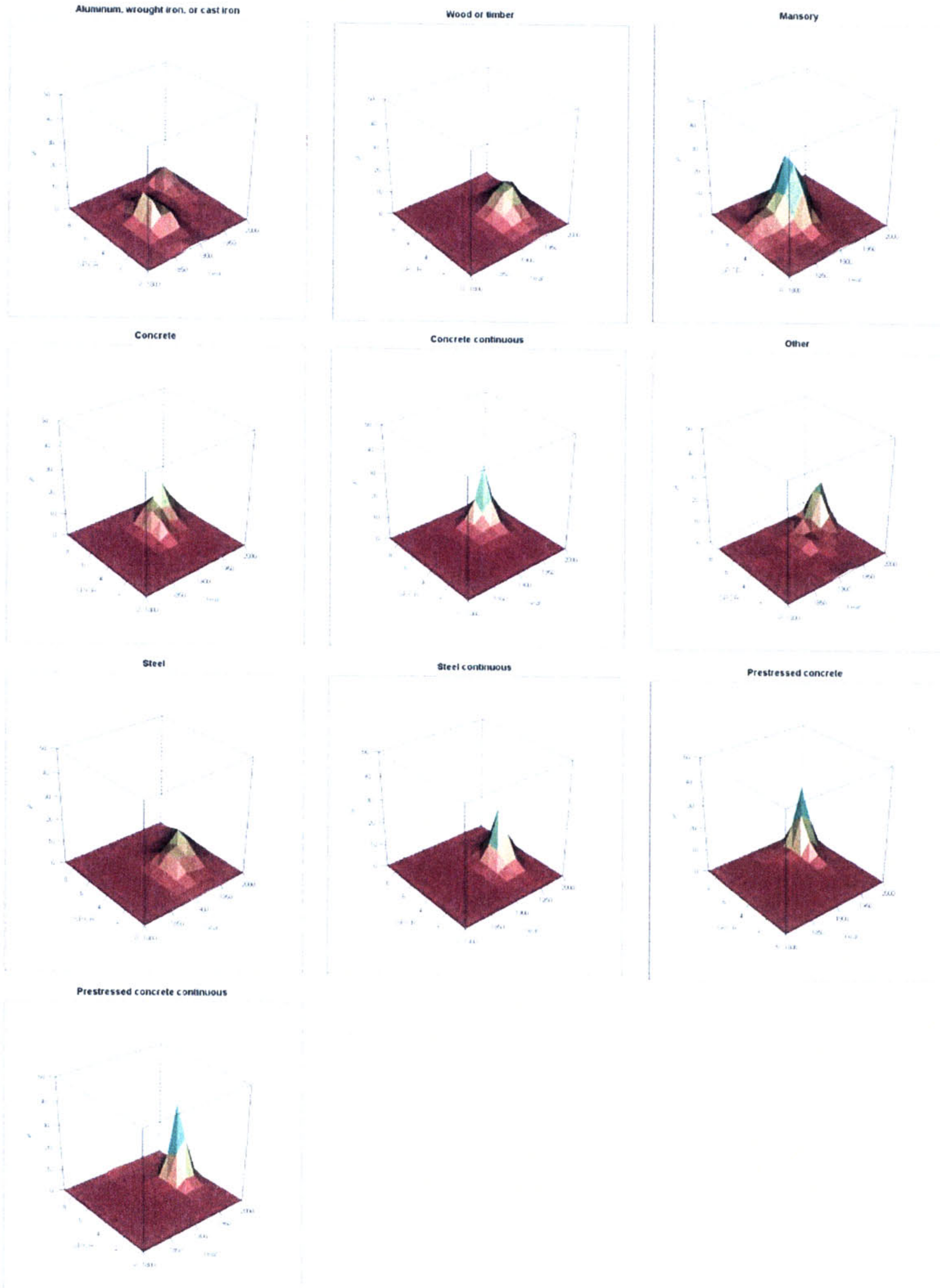
To help understand effect of design material on structural condition, percentage of bridges in each bucket by DCR, SPCR and SCR are plotted in Figure 4-18. For attribution to deck condition rating, it can be seen that aged bridges with aluminum, mansory and steel bridges have distribution widely spread to structural deficiency region. The degree of spreading is smaller in case of concrete and steel. For modern material such as prestressed concrete, the distribution concentrates around good condition.

For superstructure condition, it can be seen that the width of the distributions are typically greater than those for deck condition ratings. Specifically, aluminum, wood, mansory, steel and prestressed concrete bridges now have higher proportion of population that enter into structurally deficient realm. Note the shift in the centers of the distributions of the above materials to the lower rating side. This pattern also holds for substructure condition rating.



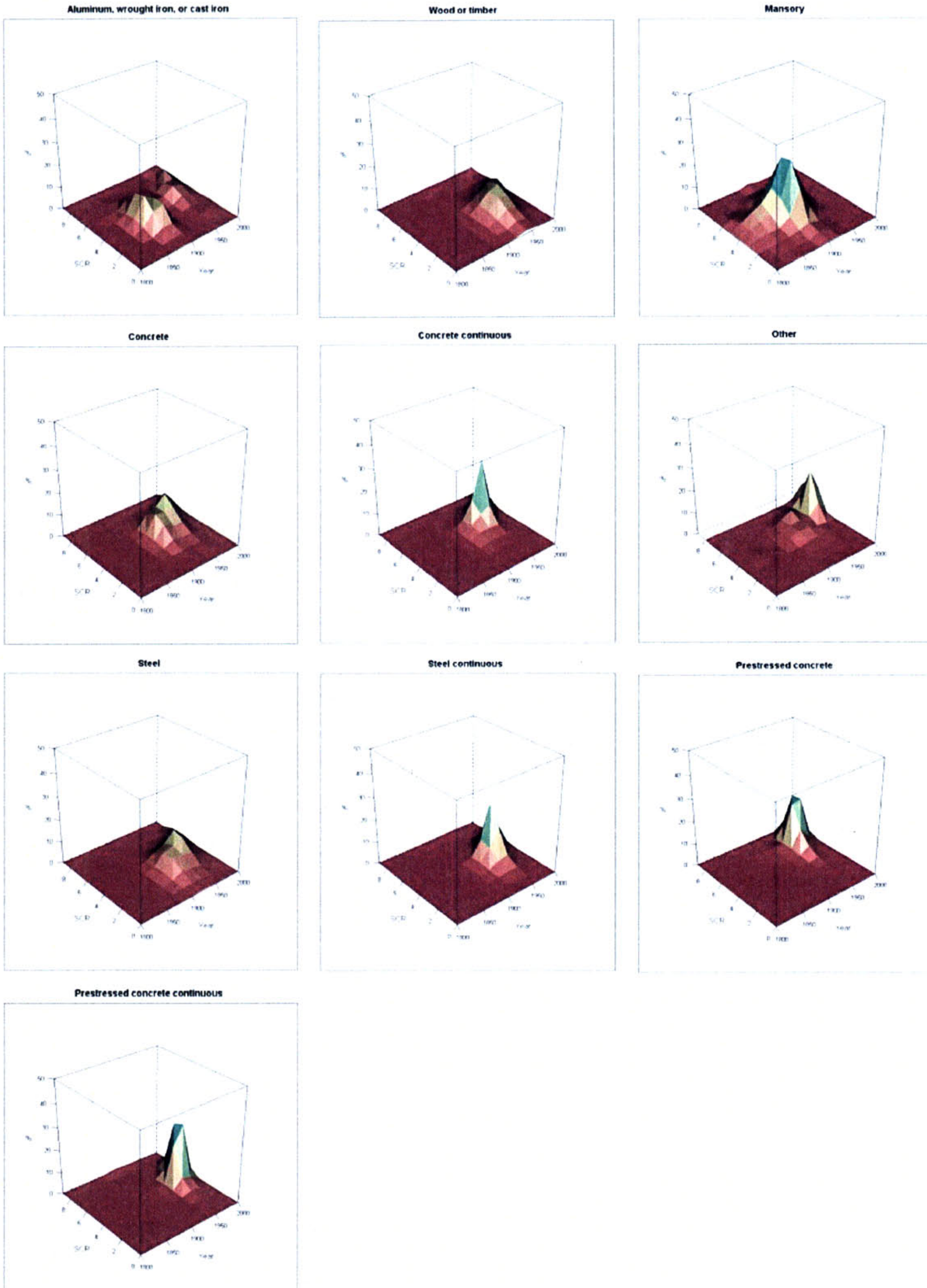
(a) DCR

Figure 4-18. Joint distribution of condition rating and construction year



(b) SPCR

Figure 4-18. Joint distribution of condition rating and construction year (continued)



(c) SCR

Figure 4-18. Joint distribution of condition rating and construction year (continued)

4.4. Conclusions

In this chapter, we have presented empirical relationships between service environments, specifications, design configurations and performances using visualization tools. Starting from basic understanding of distribution of bridge populations by different nominal NBI items to 2- or 3-way conditional analysis which helps us understand relationships between different aspects of the NBI database. Especially, the focus of our interests is mutual relationships/effects of how service environments, specifications, design configurations could affect bridge performances which in we mainly use physical condition ratings as proxies for performances.

PART II: STATISTICAL MODELING OF BRIDGE PERFORMANCE

5. METHODOLOGY FOR CONSTRUCTING BRIDGE PERFORMANCE MODEL

BRIDGE PERFORMANCE with respect to design choice is a special interested topic for Bridge Management System (BMS). It obviously serves as a basis of understanding and quantifying effects of design configurations choice on performances. Unfortunately, the studies of bridge performance are typically a difficult problem because of complex uncertainties that governs over the long lifespan of bridge infrastructure.

In this chapter, we review existing methodologies for bridge performance modeling. Especially, with regards to NBI bridge data in which several performance measurements coexist and typically change in tandem as shown in chapter 4. A system approach that allows one to study effects of specifications and design configurations on performances is proposed and presented.

5.1. The Use of Discrete Scale Performance in BMS

From the perspective of Bridge Management System (BMS), performance definitions could be formed using either continuous or ordinal quantities. In the NBI database, several of performance measures exist, such as physical condition ratings, operating ratings, overall status appraisal, etc. However, there are benefits of measuring performances in ordinal scales over the continuous scale. Discrete ratings are used instead of continuous condition indices primarily for reducing the computational complexity of the Maintenance and Rehabilitation (M&R) decision-making process (Madanet, Mishalani, and Ibrahim, 1995). This is mainly because, from this level of management perspective, detail of deterioration or physical condition is not necessary.

The choice of this choice of value scale has causes conventional technique for continuous-scale values in applicable for modeling the performance. In the following sections, we reviewed the main trends of techniques in detail.

5.2. Approach for Modeling Probability Transition

This approach is based on estimation of Markov transition probabilities from time series of physical condition ratings. This is possible using NBI database because typical inspection cycle of NBI bridges is about 1 to 2 years. This type of approach mainly consists of the following steps:

1. Bridges are classified into buckets where each group presents bridges with similar attributes. This is done based on experts' classification. From each bucket, one

can obtain a set of condition ratings, Y, and age, t. The main purpose of this grouping is to conditioned transition probabilities on these common explanatory variables.

2. For each group, one can use linear regression to identify linear relationship between condition ratings Y and t. Precisely,

$$Y_i = \beta_1 + \beta_2 t_i + \varepsilon_i \tag{5-1}$$

3. On the other hand, the theoretical expected value of condition rating can be derived from Markov chain structure. The transition matrix, $\mathbf{P}$, has the following general structure:

$$\mathbf{P} = \begin{bmatrix} p_{kk} & p_{k(k-1)} & \cdot & \cdots & \cdot & p_{k1} \\ 0 & p_{(k-1)(k-1)} & p_{(k-1)(k-2)} & \cdots & \cdot & p_{(k-1)1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & p_{22} & p_{21} \\ 0 & 0 & 0 & 0 & 0 & p_{11} \end{bmatrix} \tag{5-2}$$

where k is the highest condition rating and 1 represents the lowest rating. Because of the upper triangular structure of this matrix, it represents non-reconstruction or rehabilitation assumption. Therefore, it is important to exclude bridges with reconstruction out of the study samples as described in chapter 4. Furthermore, it can be seen from the matrix that bridges are permitted only to either stay in its current state or deteriorate to some lower state. And not the other way round. Note that the elements of this matrix which represent the target transition probabilities are to be identified.

4. Given this matrix structure and Markov chain assumption, one can model the theoretical expected value of condition rating at time t, denoted by $E(t,P)$ as:

$$E(t,P) = \mathbf{M}_t\mathbf{R} \tag{5-3}$$

where $\mathbf{M}_t$ is a row vector denoting the probability mass function of the state of the bridge at age t where the first entry is the probability that the bridge is in state k and the last entry is the probability that the facility is in state 1; $\mathbf{R}$ denotes a column vector of condition ratings.

Here, the probability mass function $\mathbf{M}_t$ can be written as:

$$\mathbf{M}_t = \mathbf{q}_\tau^{'}\mathbf{P}^{(t-\tau)} \tag{5-4}$$

where $\mathbf{q}_\tau$ is a vector of condition state probabilities at age τ whose entries are the frequencies of the bridges in the various stages. $\mathbf{P}^{(t-\tau)}$ reflects the fact that probabilities transition from time τ to t consists of $(t-\tau)$ steps of independent transitions. This is the property of Markov chain.

5. The unknown transition probabilities in matrix **P** in equation (5-2) can be identified such that it minimizes summation of point-wise distance between the expected condition rating estimated by (5-3) and that from equation (5-1). Precisely,

$$\min W = \sum_{i=\tau}^{\tau+\Delta T-1} \left|\hat{Y}_i - E(t,P)\right| \tag{5-5}$$

$$\text{Subject to } 0 \le p_{ij} \le 1; i,j = 1,2,\ldots,k$$

$$\sum_{i=1}^{k} p_{ij} = 1; i,j = 1,2,\ldots,k$$

The probability of state transition estimation has been widely used and is advantageous because it can reflect uncertainty from different sources such as uncertainty in initial condition, uncertainty in applied stresses, presence of condition assessment errors, and inherent uncertainty of the deterioration process (Lounis, 2000). The state-of-the-art BMS system Pontis and BRIDGIT have adopted such methodology for predicting performances of bridge components (Golabi and Shepard, 1997; Hawk, 1995). However, linear regression is typically used to model facility condition rating described by other observable indicators. This is not appropriate because the condition rating itself is not continuous but rather discrete and ordinal. Moreover, the approach does not recognize the latent nature of infrastructure deterioration. Deterioration is unobservable. Nevertheless, it is the surface and subsurface distress which are manifestation results of deterioration that is observable (Ben-Akiva and Ramaswamy, 1993). The expected value used for transition probability estimation utilizes such ratings observations without linking them to the underlying true deterioration thus has a major drawback. The limitation comprises an unrealistic representation of infrastructure condition and its deterioration.

5.3. The Ordered Probit Model

Madanet, Mishalani, and Ibrahim (1995) used ordered probit model to overcome the limitation of Markov chain model by incorporating latent variable presenting deterioration process and bind this to the realized rating scores by threshold parameters. The probability of transition is then equal to the area under the density curve between these threshold values. In the case of probit model, this is normal probability density. (see Figure 5-1) This approach turns out to be more realistic in capturing the unobservable process as latent variable. Contribution of each explanatory variable in the regression model has quantitative interpretation through its estimated coefficients.

The detail of the ordered probit is briefly narrated below. Let y be an ordered response taking discrete values of $\{0,1,\ldots,J\}$. The ordered probit model for y (of course, conditional on explanatory variables $\mathbf{x}$ ($\in \mathbf{R}^K$)) can be derived from a latent variable model. Let y^* be a latent variable and be determined by:

$$y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \mid \mathbf{x} \sim N(0,1) \tag{5-6}$$

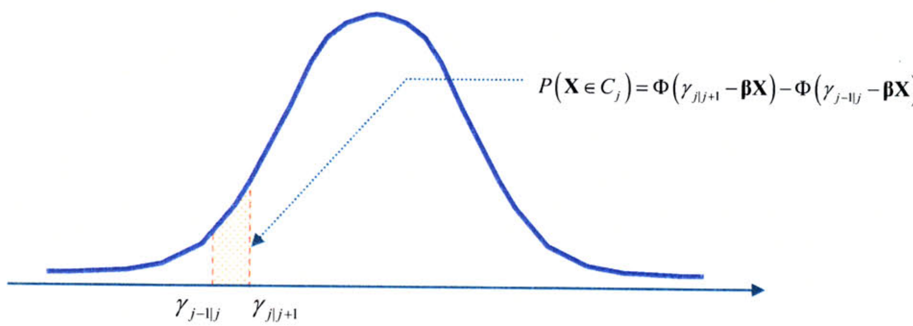


Figure 5-1. Illustration of ordered probit model key idea

where $\boldsymbol{\beta}$ is $K \times 1$. Note that by construction, the explanatory variable matrix $\mathbf{x}$ does not contain a constant. Let $\gamma_1 < \gamma_2 < \ldots < \gamma_J$ be unknown cut points such that

$$\begin{aligned} &y = 0 \quad \text{if } y^* \le \gamma_1 \\ &y = 1 \quad \text{if } \gamma_1 \le y^* \le \gamma_2 \\ &\vdots \\ &y = J \quad \text{if } y^* > \gamma_J \end{aligned} \tag{5-7}$$

From normality assumption of ε, one can derive response probability for each discrete value of y as follow:

$$\begin{aligned} &P(y=0 \mid \mathbf{x}) = P(y^* \le \gamma_1 \mid \mathbf{x}) = P(\mathbf{x}\boldsymbol{\beta} + \varepsilon \le \gamma_1 \mid \mathbf{x}) = \Phi(\gamma_1 - \mathbf{x}\boldsymbol{\beta}) \\ &P(y=1 \mid \mathbf{x}) = P(\gamma_1 \le y^* \le \gamma_2 \mid \mathbf{x}) = \Phi(\gamma_2 - \mathbf{x}\boldsymbol{\beta}) - \Phi(\gamma_1 - \mathbf{x}\boldsymbol{\beta}) \\ &\vdots \\ &P(y=J-1 \mid \mathbf{x}) = P(\gamma_{J-1} \le y^* \le \gamma_J \mid \mathbf{x}) = \Phi(\gamma_J - \mathbf{x}\boldsymbol{\beta}) - \Phi(\gamma_{J-1} - \mathbf{x}\boldsymbol{\beta}) \\ &P(y=J \mid \mathbf{x}) = P(y^* > \gamma_J \mid \mathbf{x}) = 1 - \Phi(\gamma_J - \mathbf{x}\boldsymbol{\beta}) \end{aligned} \tag{5-8}$$

Thus, one can form a likelihood function for estimation of the unknown $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ using Maximum Likelihood Estimation (MLE). Precisely, for each i, the log-likelihood function is:

$$\begin{aligned} L_i(\boldsymbol{\gamma},\boldsymbol{\beta}) = I(y_i=0)\log\Phi(\gamma_1-\mathbf{x}\boldsymbol{\beta}')+I(y_i=1)\log\left[\Phi(\gamma_2-\mathbf{x}\boldsymbol{\beta})-\Phi(\gamma_1-\mathbf{x}\boldsymbol{\beta})\right]+ \\ \ldots+I(y_i=J)\log\left[1-\Phi(\gamma_J-\mathbf{x}\boldsymbol{\beta})\right] \end{aligned} \tag{5-9}$$

5.4. Simultaneous Equation Model (SEM) Approach

Although the ordered probit model is suitable for treatment of ordinal-scale response variable, it does not suffice our application purpose on NBI bridge data. This is because we hypothesized that the performance measures, DCR, SPCR and SCR are highly correlated due to some common latent factors of deterioration. They should have strong explanatory power on one another and therefore form a simultaneous equation.

However, the difficulty in having these response variables contemporaneously appear on the RHS of equation for estimation of another response variable lies in endogeneity that is formed by such a construction. The Simultaneous Equation Model (SEM) approach is developed to deal with such nature of the problem.

Endogeneity can be described as a situation where disturbance terms of these equations are likely to be contemporaneously correlated (Henningsen and Hamann, 2006). This is because unconsidered factors that influence the disturbance term in an equation probably influence the disturbance in other equations simultaneously. Therefore, ignoring this contemporaneous correlation and estimating these equations separately leads to inefficient and bias parameter estimates.

However, another difficulty emerges from the fact that the response variables are ordinal and not continuous. Traditional methods like 2-Step Least Squares (2SLS) estimation (Judge, et. al, 1980) cannot be applied directly. Such models are considered in Lee (1981), and are postulated as follows. Consider the following simultaneous equation model:

$$\mathbf{Y\Gamma}+\mathbf{XB}+\mathbf{E}=\mathbf{0} \tag{5-10}$$

where $\mathbf{Y}$ $(T\times M)$ is the M-variate discrete response matrix of T observations. $\mathbf{X}$ is a $T\times K$ matrix of K-variate exogenous variables. $\mathbf{E}$ is a error term matrix of $T\times M$. $\mathbf{\Gamma}$ is a $M\times M$ coefficient matrix of endogenous variables. And $\mathbf{B}$ is a exogenous variable coefficient matrix of $K\times M$.

Lee (1981) suggested an alternative method to direct MLE to estimate (5-10) which employs a two-stage technique. It relies on consistent estimation of the reduced form in the first stage. Consider the reduced form of (5-10)

$$\mathbf{Y} = \mathbf{X\Pi} + \mathbf{V} \tag{5-11}$$

From (5-10), one can confirm that $\mathbf{\Pi} = -\mathbf{B\Gamma}^{-1}$ and $\mathbf{V} = -\mathbf{E\Gamma}^{-1}$. The reduced form equations for specific endogenous variables can be written as:

$$y_i = \mathbf{X}\pi_i + v_i \tag{5-12}$$

The parameters π_i can be consistently estimated using probit model.

In the second stage, the structural parameters are estimated. Let the *i*-th structural equation be:

$$y_i = Y_i\gamma_i + X_i\beta_i + \varepsilon_i \tag{5-13}$$

And partition of the reduced form in (5-11) as

$$\left[y_i \; Y_i \; Y_i^*\right] = X\left[\pi_i \; \Pi_i \; \Pi_i^*\right] + \left[v_i \; V_i \; V_i^*\right] \tag{5-14}$$

The partition represents LHS endogenous variable (y_i), the RHS endogenous variable (Y_i) and the excluded endogenous variables (Y_i^*) in the *i*-th equation. From (5-11), $Y_i = X\Pi_i + V_i$. Substituting for Y_i in (5-13) yields

$$\begin{aligned} y_i &= \left(X\Pi_i + V_i\right)\gamma_i + X_i\beta_i + \varepsilon_i \\ &= X\Pi_i\gamma_i + X_i\beta_i + V_i\gamma_i + \varepsilon_i \\ &= X\Pi_i\gamma_i + X_i\beta_i + v_i \end{aligned} \tag{5-15}$$

If $X\hat{\Pi}_i\gamma_i$ is added to and subtracted from (5-15), one obtains

$$\begin{aligned} y_i &= X\hat{\Pi}_i\gamma_i + X_i\beta_i + X\left(\Pi_i - \hat{\Pi}_i\right)\gamma_i + v_i \\ &= X\hat{\Pi}_i\gamma_i + X_i\beta_i + \omega_i \end{aligned} \tag{5-16}$$

This is the second stage of estimation where one can use probit model to estimate the structural coefficients.

5.5. Conclusions

We have reviewed two main types of modeling approach for performance on NBI bridge data. The first type of methods which is broadly employed by BMS system is based on estimation of transition probabilities between ratings. This approach has a major drawback in that it depends on subjective bucketing of bridge group for conditioning the probability. Another approach with less assumption is the ordered probit model which attempts to capture the latent process underlying the observation response. We have discussed that this type of model is more preferable for our application yet it requires further modification when dealing with system of performance models which contemporaneously involve each other such as one that is formed by DCR, SPCR and SCR. The effect of endogeneity causes the estimation of coefficients to be inefficient and bias. The framework that provides special treatment of such a model is the SEM. However, unlike the SEM on continuous response variable, a direct application of 2SLS is not possible due to ordinal nature of the response variables. We thus follow Lee (1981) suggestion to employ two-stage approach for estimation of SEM model.

6. EMPIRICAL EVALUATION OF PERFORMANCE MODEL

FOLLOWING THE CONSTRUCTION of Simultaneous Equation Model (SEM) postulated in the previous chapter, we prescribe a model for physical condition ratings comprising DCR, SPCR and SCR from NBI database. We show the estimation results and performances both for in- and out-of sample data. Finally, we include some remarks to justify the performance of the estimated model.

6.1. Model Specification (Model: PERF1)

Using notation of NBI data in chapter 4 and SEM model in chapter 5, we describe our SEM model for NBI bridge physical condition rating as follows:

$$\mathbf{Y\Gamma} + \mathbf{XB} + \mathbf{E} = \mathbf{0} \tag{6-1a}$$

$$\mathbf{Y} = [DCR, SPCR, SCR] \tag{6-1b}$$

$$\mathbf{X} = [ADT, AGE, MSM, MSD, ASM, ASD, NSMU, NAS, LMS, DST] \tag{6-1c}$$

$$\mathbf{\Gamma} = \begin{bmatrix} 0 & -\gamma_{12} & -\gamma_{13} \\ -\gamma_{21} & 0 & -\gamma_{23} \\ -\gamma_{31} & -\gamma_{32} & 0 \end{bmatrix} \tag{6-1d}$$

$$\mathbf{B} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & 0 & 0 & 0 & 0 & 0 & \beta_{1,8} & \beta_{1,9} & \beta_{1,10} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} & \beta_{2,4} & \beta_{2,5} & \beta_{2,6} & \beta_{2,7} & 0 & 0 & 0 \\ \beta_{3,1} & \beta_{3,2} & 0 & 0 & 0 & \beta_{3,6} & \beta_{3,7} & \beta_{3,8} & 0 & 0 \end{bmatrix} \tag{6-1e}$$

In this model, we try to capture effects of design configuration choice on the physical condition ratings.

6.2. Estimation Results of PERF1 Model

Two batches of 130,000 records of NBI data were I.I.D. sampled by keeping the probability of SPCR rating constants to the original data. The first batch is used for estimation while the second is used for testing purposes. Estimation of the model in equation (6-1) based on the two-stage estimation using ordered probit model is tabulated in Table 6-1.

The R^2 definition in the context of discrete response model is not the same as that used in continuous response regression model. One of the possibilities to compute R^2 is to calculate the square of correlation between the fitted and actual response value. (Agresti, 1996) However, here we report the Cragg & Uhler's pseudo R^2 (Long, 1997) measure which is a variant of pseudo R^2 used for Generalize Linear Model (GLM) literature. The Cragg & Uhler's pseudo R^2 is defined as follows:

$$R^2 = \frac{1-\left\{L\left(M_{Intercept}\right)/L\left(M_{Full}\right)\right\}^{2/N}}{1-L\left(M_{Intercept}\right)^{2/N}} \tag{6-2}$$

where M_{Full} and $M_{Intercept}$ are models with predictors and without predictors (intercept only) respectively. (Note that the latter is usually called "null" model in the context of generalized linear model) L is the estimated likelihood value. The numerator is indeed the Cox & Snell's pseudo R^2. The intuition behind this measure is that the ratio of the likelihood yielded from the two models suggests the level of improvement over the intercept model offered by the full model. Furthermore, since the definition of $L(M)$ is the conditional probability of the dependent variable given the independent variables. If there are N observations in the dataset, $L(M)$ is the product of N such probabilities. Taking the N-th root provides an estimate of the likelihood of each estimated dependent value. The denumerator here is served as a normalization factor to ensure that the pseudo R^2 defined in (10) is bounded between 0 and 1.

Table 6-1. PERF1 model step 1 estimation result of model (6-1)

	DCR*		SPCR*		SCR*	
	Value	t-value	Value	t-value	Value	t-value
AGE	(0.0463)	(248.4478)	(0.0475)	(212.1396)	(0.0492)	(260.5324)
MSM	(0.0319)	(38.2316)	(0.0511)	(50.9812)	(0.0108)	(12.9246)
MSD	(0.0006)	(0.9492)	(0.0049)	(6.2589)	(0.0172)	(26.0962)
ASM	(0.0295)	(8.6387)	0.0107	2.6801	0.0104	3.0505
ASD	(0.0018)	(0.6386)	(0.0162)	(4.7797)	(0.0213)	(7.4285)
NSMU	(0.0030)	(6.1839)	(0.0027)	(4.1976)	(0.0037)	(7.5126)
NAS	0.0003	0.6674	(0.0021)	(4.3734)	(0.0009)	(1.9033)
LMS	(0.0036)	(19.0642)	0.0021	9.4644	0.0022	11.5932
DST	(0.0493)	(43.9736)	(0.0916)	(68.2227)	(0.0953)	(84.4790)
ADT	(0.0155)	(39.6772)	(0.0088)	(18.9224)	(0.0066)	(16.8563)
0|1	(5.4695)	(132.5260)	(5.5446)	(111.7711)	(5.5768)	(133.0554)
1|2	(5.3956)	(133.6180)	(5.4925)	(112.5739)	(5.4019)	(135.3987)
2|3	(5.1929)	(134.3048)	(5.2520)	(113.9767)	(5.0610)	(134.1081)
3|4	(4.7770)	(129.2523)	(4.8778)	(110.8214)	(4.6523)	(126.9232)
4|5	(4.1416)	(114.7853)	(4.2734)	(99.8301)	(4.0575)	(112.6151)
5|6	(3.3835)	(94.6845)	(3.5042)	(82.8379)	(3.3579)	(93.9471)
6|7	(2.5181)	(70.8839)	(2.6614)	(63.3028)	(2.5398)	(71.4393)
7|8	(1.1893)	(33.7039)	(1.4004)	(33.5380)	(1.2018)	(34.0356)
8|9	0.1023	2.8951	0.1475	3.5316	0.1096	3.1023
R-Squared		**0.3241**		**0.3439**		**0.3510**

Here, some of the distinctive trends can be observed, age negatively affects physical condition ratings. Structure length on the other hand has positive effect on DCR but not for SPCR and SCR. Note the very dense cutting points between ratings from 0-3 for DCR, SPCR and SCR. The R^2 values obtained for three cases are very low suggesting poor quality of fit of the specified SEM.

Comparison of the fraction of estimated instances to the total number of observations (i.e. probability) of each condition between the actual observation and the estimated values yielded from fitted models are shown in Table 6-2 and plotted in Figure 6-1.

One can observe the failure of the model to capture ratings 0-4. On the other hand, the rating 7 for DCR, SPCR and SCR are overly predicted by almost double of the actual population.

Table 6-2. Comparison of fraction of population for fitted and actual observed rating for each rating level (yielded from PERF1 model step 1 estimation)

	DCR*		SPCR*		SCR*	
	Observed	Fitted	Observed	Fitted	Observed	Fitted
0	0.0011	0.0000	0.0010	0.0000	0.0011	0.0000
1	0.0003	0.0000	0.0002	0.0000	0.0006	0.0000
2	0.0012	0.0000	0.0012	0.0000	0.0029	0.0000
3	0.0053	0.0000	0.0042	0.0000	0.0078	0.0000
4	0.0245	0.0000	0.0194	0.0000	0.0294	0.0002
5	0.0848	0.0016	0.0718	0.0063	0.0850	0.0129
6	0.2016	0.1268	0.1739	0.0822	0.1874	0.1019
7	0.3988	0.6647	0.3633	0.5431	0.3882	0.6441
8	0.2255	0.2068	0.2998	0.3683	0.2334	0.2409
9	0.0569	0.0000	0.0652	0.0000	0.0643	0.0000

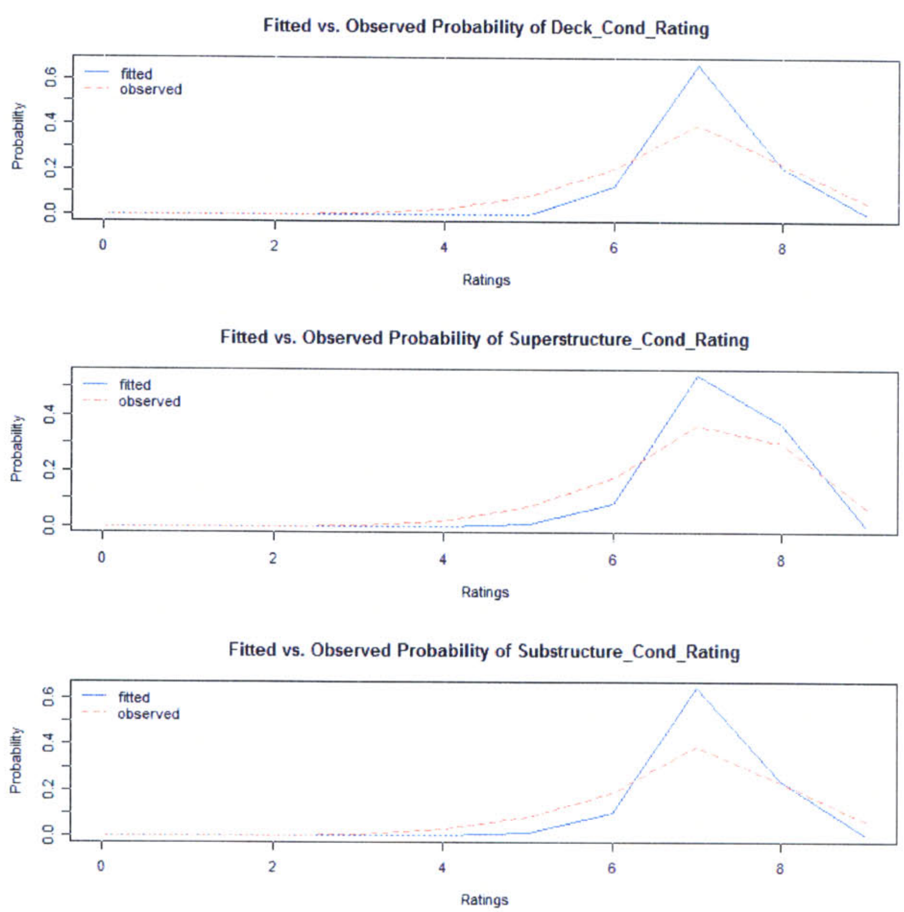


Figure 6-1. Fitted vs. observed population fraction of each condition rating (yieled from PERF1 model step 1 estimation)

Table 6-3 shows the result of step 2 estimation obtained by using the fitted DCR, SPCR and SCR obtained from model estimated in step 1. Interestingly, one can observe that SPCR has positive impact on DCR while SCR's is negative. For SPCR, DCR has negative effect while SCR has, by far, greater positive effect. The same applies for SCR, meaning SPCR's positive impact is well above the level of negative impact yielded by DCR.

Unfortunately, the R^2 values obtained in this step are also in the level of 30% suggesting low quality of fit. Nonetheless, for completeness purpose, comparison of fraction of population predicted and actually observed for each rating is tabulated in Table 6-4 and plotted in Figure 6-2.

Table 6-3. PERF1 model step 2 estimation result of model (6-1)

	DCR		SPCR		SCR	
	Value	**t-value**	**Value**	**t-value**	**Value**	**t-value**
DCR*	N/A	N/A	(0.1181)	(2.4300)	(0.5502)	(15.0229)
SPCR*	0.8158	39.70	N/A	N/A	1.1976	49.7103
SCR*	(0.1411)	(5.12)	1.0486	34.9819	N/A	N/A
AGE	(0.0146)	(15.44)	(0.0020)	(2.0392)	(0.0166)	(20.2631)
MSM	N/A	N/A	(0.0410)	(24.1893)	N/A	N/A
MSD	N/A	N/A	0.0139	16.4468	N/A	N/A
ASM	N/A	N/A	(0.0051)	(1.2589)	N/A	N/A
ASD	N/A	N/A	0.0096	3.1691	N/A	N/A
NSMU	N/A	N/A	0.0011	2.2082	(0.0015)	(2.9914)
NAS	N/A	N/A	(0.0013)	(2.8078)	0.0022	4.5508
LMS	(0.0046)	(24.88)	(0.0005)	(1.6010)	(0.0019)	(6.8855)
DST	0.0033	1.49	N/A	N/A	N/A	N/A
DW	(0.0106)	(18.78)	N/A	N/A	N/A	N/A
ADT	(0.0049)	(10.84)	(0.0035)	(4.9755)	(0.0037)	(6.7391)
0|1	(5.3494)	(225.33)	(5.4690)	(138.0396)	(5.1779)	(224.4102)
1|2	(5.2660)	(233.84)	(5.4183)	(138.5188)	(5.0260)	(242.0067)
2|3	(5.0504)	(249.61)	(5.1874)	(138.3743)	(4.7086)	(265.5967)
3|4	(4.6104)	(258.73)	(4.8108)	(132.9900)	(4.3125)	(271.3054)
4|5	(3.9409)	(239.00)	(4.2045)	(118.8468)	(3.7074)	(251.8697)
5|6	(3.0919)	(194.89)	(3.4167)	(97.6038)	(3.0004)	(211.7139)
6|7	(2.3793)	(152.88)	(2.6218)	(75.3547)	(2.1992)	(159.1558)
7|8	(1.1077)	(73.38)	(1.4057)	(40.7151)	(0.8762)	(66.1696)
8|9	0.1705	11.25	0.1217	3.5225	0.4108	30.8431
R-Squared	**0.3198**		**0.3242**		**0.3573**	

Table 6-4. Comparison of fraction of population for fitted and actual observed rating for each rating level (yielded from PERF1 model step 1 estimation)

	DCR		SPCR		SCR	
	Observed	Fitted	Observed	Fitted	Observed	Fitted
0	0.0016	0.0000	0.0015	0.0000	0.0015	0.0000
1	0.0005	0.0000	0.0002	0.0000	0.0008	0.0000
2	0.0018	0.0000	0.0017	0.0000	0.0034	0.0000
3	0.0082	0.0000	0.0056	0.0000	0.0092	0.0000
4	0.0367	0.0001	0.0247	0.0000	0.0346	0.0002
5	0.1274	0.0964	0.0886	0.0237	0.0961	0.0176
6	0.1876	0.0170	0.1799	0.0920	0.1942	0.1268
7	0.3752	0.7005	0.3548	0.5485	0.3829	0.6425
8	0.2089	0.1860	0.2823	0.3357	0.2176	0.2127
9	0.0521	0.0000	0.0606	0.0000	0.0597	0.0001

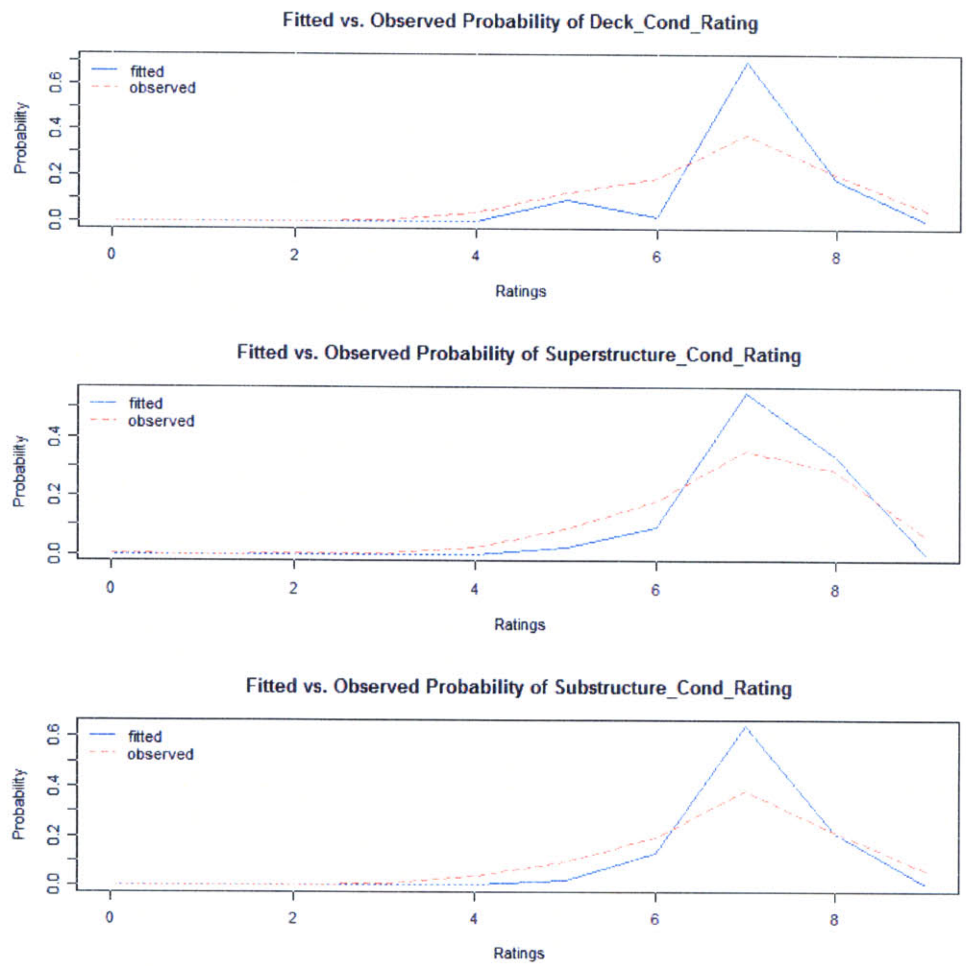


Figure 6-2. Fitted vs. observed population framction of each condition rating (yielded from PERF1 model step 2 estimation)

6.3. Investigation of Model PERF1's Performance

In correspondence to poor quality of fitness obtained from earlier attempt to fit a SEM of bridge physical condition rating (DCR, SPCR and SCR) with design attributes, the following analyses to analyze the cause of the problem has been conducted. The analyses carried out in this note are aimed to pinpoint the cause of poor quality of fitness by examining degree of association of these ratings to design attributes. The tool used to study such relationship is the so-called categorical data analysis. Detection of no strong association (or in other words independence) between a certain rating to a given design attribute evidences the undermined predictability of that rating by that design attribute.

6.3.1. Analyses of degree of associations

In this section, the so-called categorical data analysis framework is applied to help understand predictability of NBI bridge physical condition ratings by design attributes as explanatory variables. The general idea here is to measure degree of association between each condition rating score to each design attribute. Association is defined as a negation of independence between a combination of rating and design attribute under consideration. In other words, if the independence test fails for a pair of rating value and design attribute with high observed probability, one can say that the pair of rating value and design attribute possesses association among each other. Hence, supports predictability of that rating using the design attribute under consideration.

Using two-way table analysis technique, one can apply Chi-square test to test independency between response and explanatory variables in the contingency table. The adjusted residual from the Chi-square test is defined as:

$$residual = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}\left(1-p_{i+}\right)\left(1-p_{+j}\right)}} \tag{6-3}$$

The n_{ij} and the $\hat{\mu}_{ij}$ are the observed count in the table and the expected frequencies of cell (i,j) respectively. And p_{i+} and p_{+j} are the probability of occurrence in for the i-th row and the j-th column respectively.

As a rule of thumb, adjusted residual from the test at each cell whose magnitude are greater than 3 in absolute value is a strong evidence against independence among response and a given explanatory variable level (Agresti, 1997). For example, given the following two-way table (see Table 6-6(a)) whose response is Superstructure Condition Rating (SPCR) and explanatory variable is Main Structure Material (MSM), the result of Chi-squared independence test yields p-value of approximately 0.0003, a clear evidence against the null hypothesis of independence. This means that MSM has association with SPCR in a certain way. However, if one examines the adjusted residuals at each cell (tabulated in Table 6-6(c)), those in red have absolute magnitude of less than 3 and thus shows independence. In other words, MSM of those types cannot help distinguishing SPCR in red in the corresponding row from other values. This is expectable because in most cases, the observation frequencies for those cells are 0.

(The same type of tables for DCR and SPCR are shown in Table 6-5 and 6-7 respectively.)

Table 6-5. Main Structure Material vs. Deck Condition Rating

(a) Observed frequencies

Main Structure Materials	Deck Condition Rating 0	1	2	3	4	5	6	7	8	9
Prestressed concrete	1	3	54	202	1013	3951	12607	38246	29103	6938
Mansory	0	0	0	0	0	0	5	3	2	1
Wood or timber	89	20	101	177	834	2599	4957	5452	2711	383
Prestressed concrete continuous	0	0	3	32	164	514	1958	6908	4390	1085
Concrete	12	3	14	152	782	3780	10611	21501	12029	2893
Other	0	1	0	1	0	6	23	85	68	36
Steel	33	42	124	589	2492	9194	15568	21399	11316	3695
Steel continuous	2	3	43	296	1179	3260	8131	12445	4535	1143
Concrete continuous	3	16	32	236	1081	3161	7613	13788	5251	1682
Aluminum	0	0	1	0	3	9	25	18	18	3

(b) Expected frequencies (as if the two attributes are independent)

Main Structure Materials	Deck Condition Rating 0	1	2	3	4	5	6	7	8	9
Prestressed concrete	42.293	26.584	112.379	509.028	2280.202	7997.625	18578.151	36204.405	20972.243	5395.089
Mansory	0.005	0.003	0.013	0.061	0.272	0.955	2.218	4.323	2.504	0.644
Wood or timber	7.953	4.999	21.133	95.724	428.797	1503.972	3493.664	6808.321	3943.878	1014.559
Prestressed concrete continuous	6.912	4.344	18.365	83.186	372.633	1306.979	3036.057	5916.554	3427.301	881.670
Concrete	23.772	14.942	63.165	286.110	1281.639	4495.246	10442.269	20349.503	11787.922	3032.432
Other	0.101	0.063	0.268	1.216	5.446	19.100	44.369	86.465	50.087	12.885
Steel	29.591	18.600	78.628	356.150	1595.384	5595.681	12998.534	25331.057	14673.603	3774.770
Steel continuous	14.250	8.957	37.863	171.505	768.261	2694.612	6259.472	12198.225	7066.105	1817.749
Concrete continuous	15.088	9.484	40.091	181.595	813.460	2853.145	6627.736	12915.884	7481.826	1924.692
Aluminum	0.035	0.022	0.094	0.425	1.906	6.685	15.529	30.263	17.530	4.510

(c) Adjusted Residuals

Main Structure Materials	Deck Condition Rating 0	1	2	3	4	5	6	7	8	9
Prestressed concrete	-6.350	-4.574	-5.507	-13.608	-26.537	-45.249	-43.808	10.730	56.145	21.006
Mansory	-0.071	-0.056	-0.116	-0.247	-0.522	-0.977	1.868	-0.636	-0.319	0.443
Wood or timber	28.738	6.709	17.373	8.307	19.568	28.236	24.757	-16.438	-19.632	-19.828
Prestressed concrete continuous	-2.629	-2.084	-3.585	-5.612	-10.808	-21.934	-19.565	12.889	16.444	6.848
Concrete	-2.414	-3.089	-6.186	-7.929	-13.956	-10.668	1.651	8.072	2.220	-2.532
Other	-0.318	3.717	-0.518	-0.196	-2.334	-2.998	-3.208	-0.158	2.531	6.440
Steel	0.627	5.426	5.117	12.338	22.448	48.103	22.537	-24.705	-27.718	-1.298
Steel continuous	-3.245	-1.990	0.835	9.506	14.819	10.892	23.655	2.234	-30.111	-15.826
Concrete continuous	-3.112	2.116	-1.278	4.037	9.380	5.763	12.102	7.674	-25.791	-5.532
Aluminum	-0.188	-0.149	2.956	-0.652	0.792	0.895	2.403	-2.229	0.112	-0.711

Table 6-6. Main Structure Material vs. Superstructure Condition Rating

(a) Observed frequencies

Main Structure Materials	Superstructure Condition Rating									
	0	1	2	3	4	5	6	7	8	9
Prestressed concrete	1	0	83	0	11	0	28	2	1	0
Mansory	1	0	21	0	3	1	7	1	0	0
Wood or timber	66	0	133	2	19	0	128	10	4	0
Prestressed concrete continuous	198	0	249	3	171	1	527	74	68	0
Concrete	881	0	862	23	857	2	2451	397	439	3
Other	3193	0	2449	185	3779	14	8665	2128	1881	9
Steel	9109	5	4602	900	9584	19	16616	6559	5951	20
Steel continuous	28596	3	5473	4420	21483	56	20109	12620	15739	21
Concrete continuous	41829	3	3044	8027	12779	87	12359	7480	6890	22
Aluminum	8244	0	407	1494	3091	40	3562	1766	1890	2

(b) Expected frequencies (as if the two attributes are independent)

Main Structure Materials	Superstructure Condition Rating									
	0	1	2	3	4	5	6	7	8	9
Prestressed concrete	38.064	0.005	7.158	6.220	21.395	0.091	26.632	12.825	13.579	0.032
Mansory	10.271	0.001	1.932	1.679	5.773	0.025	7.186	3.461	3.664	0.009
Wood or timber	109.358	0.013	20.565	17.871	61.467	0.261	76.514	36.846	39.013	0.091
Prestressed concrete continuous	390.003	0.047	73.341	63.735	219.210	0.931	272.872	131.402	139.133	0.326
Concrete	1786.884	0.213	336.028	292.014	1004.358	4.268	1250.225	602.049	637.469	1.494
Other	6737.593	0.805	1267.020	1101.063	3787.016	16.091	4714.077	2270.074	2403.629	5.632
Steel	16121.224	1.925	3031.633	2634.544	9061.298	38.501	11279.502	5431.668	5751.230	13.475
Steel continuous	32783.195	3.915	6164.955	5357.457	18426.535	78.294	22937.347	11045.529	11695.371	27.403
Concrete continuous	27949.698	3.338	5256.004	4567.563	15709.758	66.751	19555.504	9416.995	9971.026	23.363
Aluminum	6191.710	0.739	1164.365	1011.854	3480.190	14.787	4332.140	2086.152	2208.886	5.176

(c) Adjusted Residuals

Main Structure Materials	Superstructure Condition Rating									
	0	1	2	3	4	5	6	7	8	9
Prestressed concrete	-6.008	-0.067	28.347	-2.494	-2.247	-0.302	0.265	-3.023	-3.414	-0.178
Mansory	-2.893	-0.035	13.720	-1.296	-1.154	6.228	-0.070	-1.323	-1.914	-0.093
Wood or timber	-4.146	-0.114	24.793	-3.754	-5.417	-0.511	5.886	-4.423	-5.606	-0.302
Prestressed concrete continuous	-9.722	-0.216	20.512	-7.608	-3.256	0.071	15.384	-5.008	-6.031	-0.571
Concrete	-21.430	-0.462	28.693	-15.742	-4.650	-1.098	33.960	-8.357	-7.861	1.233
Other	-43.183	-0.897	33.206	-27.607	-0.130	-0.521	57.544	-2.982	-10.660	1.419
Steel	-55.228	2.216	28.521	-33.793	5.491	-3.143	50.247	15.296	2.634	1.777
Steel continuous	-23.126	-0.462	-8.813	-12.808	22.516	-2.520	-18.675	14.981	37.391	-1.223
Concrete continuous	83.019	-0.185	-30.511	51.187	-23.383	2.478	-51.462	-19.961	-30.855	-0.282
Aluminum	26.082	-0.860	-22.195	15.157	-6.597	6.557	-11.701	-7.009	-6.785	-1.396

Table 6-7. Main Structure Material vs. Substructure Condition Rating

(a) Observed frequencies

		Substructure Condition Rating									
		0	1	2	3	4	5	6	7	8	9
Main Structure Materials	Prestressed concrete	0	4	34	202	1202	4484	12060	36910	29376	7846
	Mansory	0	0	0	0	0	0	7	2	2	0
	Wood or timber	92	79	301	703	1801	3134	4242	4388	2205	378
	Prestressed concrete continuous	0	0	2	1	30	234	1107	6464	5462	1754
	Concrete	13	70	267	536	2001	5377	11583	17619	11197	3114
	Other	0	1	2	1	4	7	30	79	62	34
	Steel	29	33	253	808	3249	9737	16201	20333	10129	3680
	Steel continuous	4	0	19	84	544	2180	6823	13791	6140	1452
	Concrete continuous	3	0	10	51	335	1370	5294	16306	7529	1965
	Aluminum	0	0	0	2	3	10	27	19	14	2

(b) Expected frequencies (as if the two attributes are independent)

		Substructure Condition Rating									
		0	1	2	3	4	5	6	7	8	9
Main Structure Materials	Prestressed concrete	42.595	56.491	268.259	721.399	2769.896	8015.449	17332.317	35015.969	21785.781	6109.843
	Mansory	0.005	0.007	0.032	0.086	0.331	0.957	2.070	4.181	2.601	0.730
	Wood or timber	8.010	10.623	50.447	135.661	520.885	1507.323	3259.382	6584.833	4096.866	1148.970
	Prestressed concrete continuous	6.961	9.232	43.839	117.892	452.659	1309.891	2832.462	5722.339	3560.250	998.476
	Concrete	23.942	31.752	150.781	405.479	1556.883	4505.264	9742.020	19681.515	12245.190	3434.175
	Other	0.102	0.135	0.641	1.723	6.615	19.143	41.394	83.627	52.030	14.592
	Steel	29.802	39.525	187.692	504.740	1938.007	5608.152	12126.865	24499.547	15242.810	4274.860
	Steel continuous	14.351	19.033	90.384	243.059	933.252	2700.618	5839.718	11797.810	7340.208	2058.568
	Concrete continuous	15.196	20.153	95.701	257.359	988.158	2859.503	6183.286	12491.910	7772.054	2179.680
	Aluminum	0.036	0.047	0.224	0.603	2.315	6.700	14.488	29.269	18.210	5.107

(c) Adjusted Residuals

		Substructure Condition Rating									
		0	1	2	3	4	5	6	7	8	9
Main Structure Materials	Prestressed concrete	-6.526	-6.984	-14.303	-19.338	-29.791	-39.445	-40.047	10.122	51.424	22.211
	Mansory	-0.071	-0.082	-0.179	-0.294	-0.575	-0.978	3.427	-1.067	-0.373	-0.854
	Wood or timber	29.676	20.979	35.276	48.710	56.089	41.898	17.211	-27.072	-29.557	-22.745
	Prestressed concrete continuous	-2.638	-3.038	-6.319	-10.766	-19.866	-29.727	-32.421	9.804	31.872	23.910
	Concrete	-2.236	6.788	9.465	6.482	11.256	12.987	18.652	-14.702	-9.472	-5.464
	Other	-0.319	2.355	1.698	-0.551	-1.017	-2.775	-1.771	-0.506	1.382	5.081
	Steel	-0.147	-1.038	4.767	13.498	29.780	55.134	36.997	-26.619	-41.420	-9.098
	Steel continuous	-2.732	-4.363	-7.509	-10.202	-12.742	-10.018	12.867	18.351	-14.009	-13.369
	Concrete continuous	-3.129	-4.489	-8.760	-12.863	-20.778	-27.855	-11.309	34.125	-2.757	-4.598
	Aluminum	-0.189	-0.217	-0.474	1.799	0.450	1.275	3.287	-1.898	-0.987	-1.375

By iterating combination of condition ratings (deck, superstructure and substructure) and design attributes, one can accumulate number of design attribute value which has no association (and hence predictability) over a given rating score. This total score is then normalized by total number of design attribute value (76 for this analysis). The final results for deck, superstructure and substructure condition ratings are tabulated in Table 6-8.

Table 6-8. Summary result of independence analysis for physical condition ratings

	Physical Condition Rating		
Rating	Deck	Superstructure	Substructure
0	85.53%	85.53%	84.21%
1	86.84%	96.05%	77.63%
2	75.00%	73.68%	67.11%
3	63.16%	61.84%	57.89%
4	51.32%	46.05%	48.68%
5	51.32%	48.68%	51.32%
6	48.68%	48.68%	44.74%
7	57.89%	55.26%	47.37%
8	47.37%	44.74%	44.74%
9	43.42%	42.11%	40.79%

Table 6-8 coincides with the result of poor fitness quality obtained from the SEM analysis conducted earlier. It is clear from Table 6-8 that ratings of 0-3 have higher rate of independency over design attributes. This mean that they tend to be indistinguishable given most of the design attributes. Especially, the superstructure condition rating 1 which possesses approximately 96% of independency detected over all design attributes. This means that given a design attribute, says Main Structure Material, rating 1 is independent of the type of material used. In other words, one observes similar probability of obtaining rating 1 regardless of whatever material type selected. This can also be confirmed by Table 6-6(c).

As one could observe from Table 6-5 to 6-6, the independency of design configuration attributes to low ratings are typically due to low proportion of population in the low rating regions (0-3) compared to higher ratings. This can be confirmed by the distributions of data by ratings shown in chapter 4 in which we observed that ratings 0-3 forms less than 5% of the entire population. This imbalance of population thus cause the tight distances between cutting points for these low ratings in the estimated model and consequently impairs the quality of fit.

In the next subsection, we perform another set of examination on bridges with low ratings to confirm peculiarity of this population that cannot be explained by the model.

6.3.2. Examination of low physical condition rating (0-3) bridges

Example of bridges of low physical condition ratings (from 0-3) in each type of NBI condition appraisals (i.e. deck, superstructure and substructure) are selected from the entire NBI database for examination. In addition, bridges with higher ratings but have similar characteristics are also selected and compared side-by-side aid the analysis.

Deck Condition Rating (DCR)

Table 6-9 shows the comparison of bridges with low (highlighted in yellow) and high DCRs of more or less similar characteristics. It can be seen from the table that while most of other design aspects and age remain more or less similar, the most distinctive feature that makes those with low DCR differ from those of high DCR is the Average Daily Traffic (ADT). For low rating group (highlighted in light yellow), the ADT are over 5,000, that of the group with high rating is bounded within 2,000 except the one in PA. Location-wise, it can be seen that those with low-rating (and probably caused by high volume of daily traffic) are mostly in CA.

A noteworthy point is the negative correlation tendency between DCR and Superstructure Condition Rating (SPCR) and Substructure Condition Rating (SCR) for those with low ratings. This trend is also reflected in the previous regression analysis in which negative coefficient of SCR appear to be negative in the Reduced-Form (RF) equation of DCR. A probable reason could be that the bridges in CA are typically built to tolerate seismic motion and thus have a good design consideration for superstructure and substructure and deck condition are simply a function of ADT.

Table 6-9. Sample bridges with low and high DCR of similar characteristics

State	Year Built	Lanes on Structure	Lanes under Structure	Design Load	Type of Service on Bridge	Type of Service under Bridge	Average Daily Traffic	Year of Average Daily Traffic	Main Structure Material	Main Structure Design	Number of Spans in Main Unit	Length of Maximum Span	Structure Length	Deck Width	Deck Structure Type	Operating Rating	Inventory Rating	DCR	SPCR	SCR
CA	1960	2	0	HS20	Highway	Waterway	6850	1997	Steel	Stringer/Multi-beam or girder	3	19.8	56.7	12.9	Concrete cast-in-place	47.1	25.4	0	6	5
CA	1963	2	2	HS20+Mod	Highway	Highway-railroad	11150	1998	Steel	Stringer/Multi-beam or girder	4	43	137.2	14.2	Concrete cast-in-place	53.5	32.6	1	7	6
MI	1932	3	0	HS20+Mod	Highway-pedestrian	Waterway	18086	2004	Steel	Stringer/Multi-beam or girder	1	21.9	22.8	16	Concrete cast-in-place	8.7	5.2	2	3	4
CA	1931	2	0	HS15	Highway	Waterway	11000	1998	Concrete	Tee beam	4	11.9	48.8	11.6	Concrete cast-in-place	47.1	29	3	7	7
IA	1938	2	0	H15	Highway	Waterway	1940	2004	Steel	Stringer/Multi-beam or girder	1	15.2	16.5	8.5	Concrete cast-in-place	40.2	20.9	4	5	5
PA	1934	2	0	HS20	Highway	Waterway	4781	2006	Concrete	Tee beam	1	9.8	10.1	7.9	Concrete cast-in-place	47.7	28.8	5	4	4
AL	1937	2	0	H15	Highway	Waterway	1410	2003	Steel continuous	Stringer/Multi-beam or girder	5	6.1	30.9	8.4	Concrete cast-in-place	60.8	39	6	6	6
AR	1931	2	0	H15	Highway	Waterway	1400	2004	Concrete	Tee beam	5	11	55.2	8	Concrete cast-in-place	28.1	17.2	7	7	7
VA	1932	2	0	HS20+Mod	Highway	Waterway	110	2004	Steel	Stringer/Multi-beam or girder	1	10.7	11	7.8	wook or timber	58.5	35.1	8	8	8
VA	1932	1	0	HS20+Mod	Highway	Waterway	120	2004	Concrete	Slab	1	7.6	7.6	5.7	Concrete precast panels	66.6	40.5	9	9	8

Superstructure Condition Rating (SPCR)

Table 6-10 shows the comparison of bridges with low (highlighted in yellow) and high SPCRs of more or less similar characteristics. Unlike the case of DCR in which ADT is a major cause for the deterioration of deck structure, the pattern of SPCR deterioration cannot distinctively be seen in Table 6-10. The result in Table 6-10 suggests that the majority of bridges with low rating for superstructure members are steel bridges. To gain more insight in regard to this, Table 6-11 shows bridges with low SPCR ratings by state and by main structure material. One can see that the deficient bridges are clustered within steel, wood and concrete. Moreover, Oklahoma turns out to possess highest fractions of deficient steel and wood bridges whereas Pennsylvania has highest fraction of defective concrete bridges. This suggests that to obtain better statistical explanation of what causes such a pattern, one would probably need geographical information which is unfortunately not available in this study.

Table 6-10. Sample bridges with low and high SPCR of more or less similar characteristics

State_ Code	Year Built	Lanes on Structure	Lanes under Structure	Design Load	Type of Service on Bridge	Type of Service under Bridge	Average Daily Traffic	Year of Average Daily Traffic	Main Structure Material	Main Structure Design	Number of Spans in Main Unit	Number of Approach Spans	Length of Maximum Span	Structure Length	Deck Width	Deck Structure Type	Approach Spans Material	Approach Spans Design	Operating Rating	Inventory Rating	DCR	SPCR	SCR
PA	1937	2	0	H15	Highway	Waterway	566	2006	Steel	Stringer/Multi-beam or girder	1	0	7.3	8.2	7.6	wood or timber	Other	Other	9	5.4	3	0	5
CA	1946	2	0	HS20	Highway	Waterway	625	1980	Steel	Stringer/Multi-beam or girder	4	0	8.2	32.6	8.8	Concrete cast in-place	Other	Other	27.4	16.4	7	1	5
OH	1932	2	0	H15	Highway	Waterway	835	1991	Concrete continuous	Tee beam	2	0	9.8	21.3	7.3	Concrete cast in-place	Other	Other	24.3	16.2	3	2	4
IA	1931	2	0	H20	Highway-pedestrian	Waterway	740	2004	Steel	Girder and floorbeam system	2	2	30.5	87.5	9.9	Concrete cast in-place	Steel continuous	Stringer/Multi-beam or girder	37.6	21.9	4	3	4
IL	1931	2	0	H20	Highway	Waterway	1300	2005	Concrete	Slab	1	0	9.4	9.8	12.9	Concrete cast in-place	Other	Other	62.1	32.4	4	4	5
KY	1931	2	0	HS20	Highway	Waterway	3832	2005	Concrete	Tee beam	3	0	43.6	43.6	7.4	Concrete cast in-place	Other	Other	70.8	39	5	5	6
IL	1931	2	0	H20	Highway	Waterway	2600	2005	Concrete	Slab	1	0	9.4	9.8	12.9	Concrete cast in-place	Other	Other	61.2	32.4	6	6	5
CA	1931	2	0	HS20+ Mod	Highway	Waterway	397	1998	Concrete continuous	Slab	3	0	3.7	11.3	9.2	Concrete cast in-place	Other	Other	22.7	13.6	7	7	6
ME	1931	2	0	HS20	Highway	Waterway	7055	2005	Steel	Stringer/Multi-beam or girder	1	0	7.6	17.7	14.9	Concrete cast in-place	Other	Other	69.9	39	8	8	6
VA	1932	2	0	HS20+ Mod	Highway	Waterway	1136	2004	Concrete	Slab	1	0	8.2	8.2	7.8	Concrete precast panels	Other	Other	44.1	32.4	9	9	8

Table 6-11. Fraction of bridges with low ratings by state and material (from total sample bridges)

State	Aluminum	Concrete	Concrete continuous	Mansory	Other	Prestressed concrete	Prestressed concrete continuous	Steel	Steel continuous	Wood or timber	Total
AL	0.0000%	0.0038%	0.0003%	0.0000%	0.0000%	0.0000%	0.0000%	0.0098%	0.0003%	0.0160%	**0.0301%**
AK	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0003%	0.0000%	0.0014%	0.0000%	0.0019%	**0.0035%**
AZ	0.0000%	0.0008%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0005%	0.0000%	0.0005%	**0.0019%**
AR	0.0000%	0.0030%	0.0005%	0.0000%	0.0000%	0.0000%	0.0000%	0.0051%	0.0008%	0.0005%	**0.0100%**
CA	0.0000%	0.0016%	0.0024%	0.0000%	0.0000%	0.0003%	0.0000%	0.0244%	0.0011%	0.0005%	**0.0304%**
CO	0.0000%	0.0003%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0016%	0.0027%	0.0022%	**0.0068%**
CT	0.0000%	0.0019%	0.0003%	0.0000%	0.0000%	0.0014%	0.0000%	0.0022%	0.0003%	0.0000%	**0.0060%**
DE	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0003%	0.0000%	0.0000%	0.0000%	0.0000%	**0.0003%**
DC	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0005%	0.0000%	0.0000%	0.0000%	0.0000%	**0.0005%**
FL	0.0000%	0.0005%	0.0000%	0.0000%	0.0000%	0.0014%	0.0000%	0.0014%	0.0005%	0.0043%	**0.0081%**
GA	0.0000%	0.0022%	0.0000%	0.0000%	0.0000%	0.0003%	0.0000%	0.0076%	0.0008%	0.0041%	**0.0149%**
HI	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0003%	0.0000%	0.0005%	0.0000%	0.0003%	**0.0011%**
ID	0.0000%	0.0014%	0.0000%	0.0000%	0.0000%	0.0014%	0.0000%	0.0022%	0.0000%	0.0005%	**0.0054%**
IL	0.0000%	0.0070%	0.0005%	0.0000%	0.0000%	0.0176%	0.0000%	0.0089%	0.0014%	0.0011%	**0.0366%**
IN	0.0000%	0.0054%	0.0022%	0.0000%	0.0000%	0.0027%	0.0014%	0.0092%	0.0005%	0.0005%	**0.0220%**
IA	0.0000%	0.0057%	0.0008%	0.0000%	0.0000%	0.0016%	0.0000%	0.0295%	0.0035%	0.0068%	**0.0480%**
KS	0.0000%	0.0089%	0.0024%	0.0000%	0.0005%	0.0005%	0.0000%	0.0138%	0.0005%	0.0141%	**0.0409%**
KY	0.0000%	0.0041%	0.0005%	0.0000%	0.0000%	0.0024%	0.0000%	0.0198%	0.0005%	0.0000%	**0.0274%**
LA	0.0000%	0.0160%	0.0000%	0.0000%	0.0003%	0.0024%	0.0000%	0.0182%	0.0003%	0.0577%	**0.0948%**
ME	0.0000%	0.0008%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0046%	0.0000%	0.0005%	**0.0060%**
MD	0.0000%	0.0005%	0.0003%	0.0000%	0.0000%	0.0003%	0.0000%	0.0049%	0.0003%	0.0011%	**0.0073%**
MA	0.0000%	0.0014%	0.0011%	0.0000%	0.0000%	0.0008%	0.0000%	0.0068%	0.0003%	0.0003%	**0.0106%**
MI	0.0000%	0.0070%	0.0005%	0.0000%	0.0000%	0.0122%	0.0000%	0.0469%	0.0008%	0.0008%	**0.0683%**
MN	0.0000%	0.0024%	0.0008%	0.0000%	0.0000%	0.0024%	0.0000%	0.0114%	0.0000%	0.0033%	**0.0203%**
MS	0.0000%	0.0114%	0.0000%	0.0000%	0.0000%	0.0005%	0.0000%	0.0168%	0.0003%	0.0572%	**0.0862%**
MO	0.0000%	0.0222%	0.0051%	0.0000%	0.0000%	0.0000%	0.0000%	0.0122%	0.0014%	0.0014%	**0.0423%**
MT	0.0000%	0.0005%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0014%	0.0000%	0.0027%	**0.0046%**
NE	0.0000%	0.0005%	0.0008%	0.0000%	0.0000%	0.0003%	0.0000%	0.0184%	0.0000%	0.0133%	**0.0333%**
NV	0.0000%	0.0000%	0.0000%	0.0000%	0.0003%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	**0.0003%**
NH	0.0000%	0.0024%	0.0003%	0.0003%	0.0000%	0.0000%	0.0000%	0.0073%	0.0005%	0.0008%	**0.0117%**
NJ	0.0000%	0.0011%	0.0003%	0.0000%	0.0000%	0.0005%	0.0000%	0.0041%	0.0003%	0.0000%	**0.0062%**
NM	0.0000%	0.0019%	0.0008%	0.0000%	0.0000%	0.0000%	0.0000%	0.0014%	0.0008%	0.0019%	**0.0068%**
NY	0.0000%	0.0011%	0.0003%	0.0000%	0.0000%	0.0008%	0.0000%	0.0130%	0.0005%	0.0003%	**0.0160%**
NC	0.0003%	0.0008%	0.0000%	0.0000%	0.0000%	0.0008%	0.0000%	0.0087%	0.0014%	0.0016%	**0.0135%**
ND	0.0000%	0.0005%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0054%	0.0000%	0.0019%	**0.0079%**
OH	0.0000%	0.0146%	0.0043%	0.0000%	0.0000%	0.0016%	0.0000%	0.0347%	0.0024%	0.0000%	**0.0577%**
OK	0.0000%	0.0095%	0.0016%	0.0005%	0.0003%	0.0008%	0.0000%	0.1599%	0.0081%	0.1005%	**0.2813%**
OR	0.0000%	0.0000%	0.0038%	0.0000%	0.0000%	0.0008%	0.0000%	0.0024%	0.0008%	0.0038%	**0.0117%**
PA	0.0000%	0.0241%	0.0081%	0.0000%	0.0000%	0.0192%	0.0000%	0.0428%	0.0041%	0.0003%	**0.0986%**
RI	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0005%	0.0000%	0.0016%	0.0000%	0.0005%	**0.0027%**
SC	0.0000%	0.0016%	0.0000%	0.0000%	0.0000%	0.0008%	0.0000%	0.0027%	0.0000%	0.0022%	**0.0073%**
SD	0.0000%	0.0035%	0.0011%	0.0000%	0.0000%	0.0000%	0.0000%	0.0089%	0.0005%	0.0030%	**0.0171%**
TN	0.0000%	0.0033%	0.0000%	0.0000%	0.0000%	0.0003%	0.0003%	0.0111%	0.0014%	0.0041%	**0.0203%**
TX	0.0000%	0.0019%	0.0011%	0.0000%	0.0000%	0.0003%	0.0000%	0.0130%	0.0019%	0.0098%	**0.0279%**
UT	0.0000%	0.0005%	0.0005%	0.0000%	0.0000%	0.0005%	0.0000%	0.0008%	0.0000%	0.0000%	**0.0024%**
VT	0.0000%	0.0003%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0068%	0.0000%	0.0005%	**0.0076%**
VA	0.0000%	0.0046%	0.0000%	0.0000%	0.0000%	0.0003%	0.0000%	0.0068%	0.0014%	0.0005%	**0.0135%**
WA	0.0000%	0.0005%	0.0008%	0.0000%	0.0000%	0.0003%	0.0003%	0.0005%	0.0000%	0.0024%	**0.0049%**
WV	0.0000%	0.0081%	0.0016%	0.0000%	0.0000%	0.0019%	0.0000%	0.0089%	0.0033%	0.0000%	**0.0238%**
WI	0.0000%	0.0068%	0.0024%	0.0000%	0.0003%	0.0027%	0.0003%	0.0070%	0.0011%	0.0019%	**0.0225%**
WY	0.0000%	0.0003%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0030%	0.0003%	0.0005%	**0.0041%**
PR	0.0000%	0.0003%	0.0016%	0.0000%	0.0000%	0.0008%	0.0000%	0.0019%	0.0000%	0.0000%	**0.0046%**
Total	**0.0003%**	**0.1973%**	**0.0474%**	**0.0008%**	**0.0016%**	**0.0829%**	**0.0022%**	**0.6322%**	**0.0450%**	**0.3282%**	1.3379%

Substructure Condition Rating (SCR)

Table 6-12 shows the comparison of bridges with low (highlighted in yellow) and high SCRs of more or less similar characteristics. There is no clear distinguishable trend that separates those with low and high ratings. Like in Table 6-11, a breakdown of SCR in each rating by state is tabulated in Table 6-13. One can see that the top three states with high fractions substructure-deficient bridges are Mississippi, Louisiana and Kansas. The first two are in the gulf area suggesting that there could be geographical environmental factors that the regression analysis cannot capture for these bridges. Unfortunately, this data is not available in the current study.

Table 6-12. Sample bridges with low and high SCR of similar characteristics

State Code	Year Built	Lanes on Structure	Lanes under Structure	Design Load	Type of Service on Bridge	Type of Service Under Bridge	Average Daily Traffic	Year of Average Daily Traffic	Main Structure Material	Main Structure Design	Number of Spans in Main Unit	Number of Approach Spans	Length of Maximum Span	Structure Length	Deck Width	Deck Structure Type	Operating Rating	InventoryRating	DCR	SPCR	SCR
MI	1932	2	0	HS20+ Mod	Highway-pedestrian	Waterway	5546	1989	Concrete continuous	Tee beam	1	0	15.8	29.8	16.1	Not applicable	50.9	37.3	5	5	0
GA	1954	2	0	H15	Highway	Waterway	640	2004	Concrete	Slab	19	0	4.5	86.8	7.4	Concrete precast panels	3.6	1.8	5	5	1
PA	1932	2	0	HS20	Highway	Waterway	10021	2006	Concrete continuous	Stringer/Multi-beam or girder	1	0	7.9	9.1	10.4	Concrete cast-in-place	48.6	32.4	4	4	2
PA	1932	2	0	H20	Highway	Waterway	4600	2006	Concrete	Tee beam	2	0	8.2	19.5	7.9	Concrete cast-in-place	54.9	32.4	4	3	3
PA	1931	2	0	H20	Highway-pedestrian	Waterway	854	2003	Concrete	Tee beam	2	0	20.1	42.7	10.7	Not applicable	51.3	29.7	5	4	4
ME	1931	2	0	HS20	Highway	Waterway	1932	2005	Concrete	Tee beam	1	0	15.2	16.2	8.3	Concrete cast-in-place	57.1	31.8	5	6	5
CA	1931	2	0	HS20+ Mod	Highway	Waterway	397	1998	Concrete continuous	Slab	3	0	3.7	11.3	9.2	Concrete cast-in-place	22.7	13.6	7	7	6
CA	1931	6	6	H20	Overpass structure at interchange or second level	Highway, w. or w/o pedestrian	29000	2000	Concrete continuous	Tee beam	6	0	19.8	73.2	41.9	Concrete cast-in-place	63.5	37.2	5	7	7
VA	1932	2	0	HS20+ Mod	Highway	Waterway	1534	2004	Concrete	Slab	1	0	7.3	7.3	9.8	Concrete cast-in-place	69.3	46.8	7	8	8
TX	1940	2	0	HS20	Highway	Waterway	2783	2000	Concrete	Box beam or girders - multiple	3	0	12.2	30.5	12.8	Concrete cast-in-place	44.4	32.7	9	9	9

Table 6-13. Fraction of bridges with SCR ratings by state (from total sample bridges)

State	0	1	2	3	4	5	6	7	8	9	Total
AL	0.0035%	0.0014%	0.0070%	0.0436%	0.1119%	0.3032%	0.5336%	0.8016%	0.3978%	0.1518%	**0.0049%**
AK	0.0008%	0.0003%	0.0027%	0.0065%	0.0114%	0.0531%	0.0442%	0.0585%	0.0648%	0.0106%	**0.0011%**
AZ	0.0003%	0.0000%	0.0003%	0.0011%	0.0054%	0.0160%	0.0799%	0.3266%	0.2320%	0.0019%	**0.0003%**
AR	0.0003%	0.0005%	0.0030%	0.0163%	0.0672%	0.1748%	0.4959%	0.7851%	0.6482%	0.0862%	**0.0008%**
CA	0.0003%	0.0003%	0.0038%	0.0111%	0.0344%	0.1127%	0.4932%	3.0363%	0.0778%	0.0014%	**0.0005%**
CO	0.0000%	0.0000%	0.0003%	0.0068%	0.0290%	0.1301%	0.3051%	0.7144%	0.3390%	0.0057%	**0.0000%**
CT	0.0008%	0.0000%	0.0005%	0.0019%	0.0127%	0.0328%	0.1374%	0.3051%	0.0341%	0.0000%	**0.0008%**
DE	0.0000%	0.0000%	0.0000%	0.0000%	0.0014%	0.0060%	0.0206%	0.0645%	0.0195%	0.0016%	**0.0000%**
DC	0.0000%	0.0000%	0.0003%	0.0000%	0.0003%	0.0019%	0.0079%	0.0057%	0.0016%	0.0000%	**0.0000%**
FL	0.0008%	0.0003%	0.0016%	0.0103%	0.0225%	0.0745%	0.2163%	1.0488%	0.4057%	0.0667%	**0.0011%**
GA	0.0084%	0.0051%	0.0133%	0.0268%	0.1046%	0.2499%	0.4404%	0.8393%	0.2547%	0.0051%	**0.0135%**
HI	0.0000%	0.0000%	0.0000%	0.0008%	0.0022%	0.0081%	0.0369%	0.1081%	0.0084%	0.0011%	**0.0000%**
ID	0.0008%	0.0000%	0.0005%	0.0073%	0.0257%	0.1005%	0.3111%	0.3469%	0.1098%	0.0165%	**0.0008%**
IL	0.0016%	0.0003%	0.0038%	0.0252%	0.1138%	0.2992%	0.6141%	1.2648%	1.7677%	0.4650%	**0.0019%**
IN	0.0011%	0.0003%	0.0057%	0.0306%	0.1217%	0.2707%	0.5157%	1.0984%	0.8948%	0.0686%	**0.0014%**
IA	0.0038%	0.0011%	0.0173%	0.1564%	0.4328%	0.7447%	0.8209%	1.0439%	0.8607%	0.5325%	**0.0049%**
KS	0.0168%	0.0035%	0.0117%	0.0496%	0.1827%	0.4794%	0.8607%	1.2718%	1.0680%	0.1499%	**0.0203%**
KY	0.0005%	0.0011%	0.0125%	0.0111%	0.1024%	0.2610%	0.6469%	1.0515%	0.5927%	0.0168%	**0.0016%**
LA	0.0474%	0.0033%	0.0111%	0.0897%	0.1691%	0.2141%	0.7482%	0.6677%	0.7883%	0.0444%	**0.0507%**
ME	0.0003%	0.0000%	0.0019%	0.0098%	0.0225%	0.0602%	0.1328%	0.1388%	0.0642%	0.0108%	**0.0003%**
MD	0.0008%	0.0003%	0.0005%	0.0030%	0.0127%	0.0696%	0.1778%	0.2439%	0.0612%	0.0049%	**0.0011%**
MA	0.0005%	0.0003%	0.0003%	0.0051%	0.0206%	0.0900%	0.1691%	0.2710%	0.0504%	0.0228%	**0.0008%**
MI	0.0070%	0.0011%	0.0060%	0.0352%	0.0797%	0.1911%	0.4133%	0.6867%	0.4772%	0.0488%	**0.0081%**
MN	0.0016%	0.0008%	0.0030%	0.0442%	0.1225%	0.2198%	0.3423%	0.6940%	0.4715%	0.1715%	**0.0024%**
MS	0.0179%	0.0377%	0.1203%	0.0962%	0.2978%	0.4222%	0.3669%	0.6271%	1.3320%	0.3157%	**0.0556%**
MO	0.0022%	0.0005%	0.0027%	0.0209%	0.1309%	0.4312%	0.7889%	0.9891%	0.7146%	0.9070%	**0.0027%**
MT	0.0005%	0.0000%	0.0030%	0.0138%	0.0344%	0.0924%	0.1789%	0.4331%	0.3203%	0.0203%	**0.0005%**
NE	0.0041%	0.0008%	0.0035%	0.0472%	0.2135%	0.3485%	0.3883%	0.5577%	0.6144%	0.6165%	**0.0049%**
NV	0.0000%	0.0000%	0.0003%	0.0008%	0.0024%	0.0041%	0.0220%	0.1306%	0.0282%	0.0003%	**0.0000%**
NH	0.0011%	0.0054%	0.0011%	0.0038%	0.0146%	0.0512%	0.0946%	0.1948%	0.0940%	0.0195%	**0.0065%**
NJ	0.0014%	0.0000%	0.0000%	0.0051%	0.0287%	0.1257%	0.2282%	0.3648%	0.0881%	0.0187%	**0.0014%**
NM	0.0005%	0.0005%	0.0030%	0.0073%	0.0347%	0.1111%	0.1344%	0.1753%	0.0214%	0.0016%	**0.0011%**
NY	0.0003%	0.0003%	0.0035%	0.0179%	0.1165%	0.4358%	0.6057%	0.6721%	0.4851%	0.3271%	**0.0005%**
NC	0.0000%	0.0000%	0.0041%	0.0314%	0.1696%	0.5276%	0.9672%	0.7341%	0.7070%	0.1344%	**0.0000%**
ND	0.0000%	0.0003%	0.0027%	0.0168%	0.0528%	0.0908%	0.1144%	0.1450%	0.1360%	0.0444%	**0.0003%**
OH	0.0014%	0.0008%	0.0041%	0.0363%	0.1385%	0.3566%	1.0330%	1.4301%	1.2035%	0.6515%	**0.0022%**
OK	0.0089%	0.0041%	0.1705%	0.1415%	0.3854%	0.9274%	0.8797%	1.2124%	0.2214%	0.0314%	**0.0130%**
OR	0.0005%	0.0000%	0.0014%	0.0146%	0.0648%	0.1406%	0.3230%	0.6434%	0.3986%	0.0168%	**0.0005%**
PA	0.0033%	0.0014%	0.0157%	0.0577%	0.3515%	0.8187%	0.8875%	1.0206%	0.2770%	0.0474%	**0.0046%**
RI	0.0005%	0.0000%	0.0000%	0.0011%	0.0130%	0.0285%	0.0398%	0.0244%	0.0022%	0.0003%	**0.0005%**
SC	0.0014%	0.0014%	0.0008%	0.0160%	0.1556%	0.2832%	0.3374%	0.6629%	0.4401%	0.0225%	**0.0027%**
SD	0.0000%	0.0005%	0.0089%	0.0328%	0.0745%	0.1444%	0.3363%	0.4672%	0.0287%	0.0011%	**0.0005%**
TN	0.0016%	0.0014%	0.0117%	0.0244%	0.1192%	0.4604%	0.5791%	0.8721%	0.3650%	0.0390%	**0.0030%**
TX	0.0103%	0.0035%	0.0076%	0.0488%	0.1531%	0.5135%	1.7143%	3.4441%	0.7382%	0.0382%	**0.0138%**
UT	0.0000%	0.0000%	0.0003%	0.0038%	0.0152%	0.0344%	0.0753%	0.2347%	0.1146%	0.0022%	**0.0000%**
VT	0.0003%	0.0003%	0.0011%	0.0038%	0.0260%	0.0534%	0.1209%	0.1664%	0.1057%	0.0000%	**0.0005%**
VA	0.0000%	0.0000%	0.0000%	0.0027%	0.0732%	0.4447%	0.6892%	0.8975%	0.2995%	0.0366%	**0.0000%**
WA	0.0003%	0.0000%	0.0008%	0.0076%	0.0317%	0.0986%	0.2553%	0.8555%	0.2829%	0.0041%	**0.0003%**
WV	0.0003%	0.0000%	0.0003%	0.0192%	0.0650%	0.1406%	0.2734%	0.5352%	0.3374%	0.0683%	**0.0003%**
WI	0.0011%	0.0008%	0.0019%	0.0238%	0.0762%	0.2144%	0.4634%	0.6764%	0.9282%	0.2385%	**0.0019%**
WY	0.0000%	0.0003%	0.0003%	0.0035%	0.0100%	0.0493%	0.1756%	0.3363%	0.0076%	0.0003%	**0.0003%**
PR	0.0008%	0.0000%	0.0005%	0.0027%	0.0098%	0.0488%	0.1607%	0.0927%	0.0260%	0.0008%	**0.0008%**
Total	**0.1561%**	**0.0797%**	**0.4770%**	**1.2940%**	**4.6680%**	**11.5616%**	**20.7976%**	**35.4689%**	**20.0081%**	**5.4891%**	**0.2358%**

From these analyses, we conclude that there seem to be geographic regional information that cannot capture by the model.we therefore propose the following two remedies:

- Excluding rating 0-3 out of the further analysis: they dominate only marginal proportion of the entire population. They can be counted as anomalies that model variables are not sufficient to explain it. This is consistent with comments from Chase, Small and Nutakor (1999).
- Add additional exogenous variables such as design specifications and state information to help increase explanatory ability.

6.4. Model Specification (Model: PERF2)

Unlike the PERF1 model specified in section 6.1, PERF2 includes design specifications and state as exogenous variables to the SEM model. At the same time, Average Daily Truck Traffic (ADTT) is also added to emphasize the adverse effect of truck traffic to physical rating conditions. The exact model now becomes

$$\mathbf{Y\Gamma} + \mathbf{XB} + \mathbf{E} = \mathbf{0} \tag{6-4a}$$

$$\mathbf{Y} = [DCR,\ SPCR,\ SCR] \tag{6-4b}$$

$$\mathbf{X} = \begin{bmatrix} STATE, LOS, LUS, DL, TSOB, TSUB, \\ ADT, ADTT, AGE, MSM, MSD, NSMU, NAS, LMS, SL, DW, DST \end{bmatrix} \tag{6-4c}$$

$$\mathbf{\Gamma} = \begin{bmatrix} 0 & -\gamma_{12} & -\gamma_{13} \\ -\gamma_{21} & 0 & -\gamma_{23} \\ -\gamma_{31} & -\gamma_{32} & 0 \end{bmatrix} \tag{6-4d}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_{1,9} & \beta_{1,10} & \beta_{1,11} & \beta_{1,12} & \beta_{1,13} & \beta_{1,14} & \beta_{1,15} & \beta_{1,16} & \beta_{1,17} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_{2,9} & \beta_{2,10} & \beta_{2,11} & \beta_{2,12} & \beta_{2,13} & \beta_{2,14} & \beta_{2,15} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_{3,9} & \beta_{3,10} & \beta_{3,11} & \beta_{3,12} & \beta_{3,13} & 0 & \beta_{3,15} & 0 & 0 \end{bmatrix}$$

(6-4e)

It is important to note that the response variables DCR, SPCR and SCR included in the fitting now only include ratings from 4-9. The same sampled in- and out-sample data were used as is but with exclusion of observations with ratings 0-3.

6.5. Estimation Results of PERF2 Model

Estimation results of step 1 and 2 are tabulated in Table 6-14. Similar to PERF1 model result, the observed and predicted proportion of each rating for both steps are summarized in Table 6-15 and plotted in Figure 6-3.

Table 6-14. PERF2 model estimation result of model (6-4)

(a) Step 1

	DCR*		SPCR*		SCR*	
	Value	t-value	Value	t-value	Value	t-value
Lanes_On_Structure	(0.0174)	(2.3840)	(0.0020)	(0.154)	(0.0301)	(8.3049)
Lanes_Under_Structure	(0.0030)	(4.2201)	(0.0061)	(8.394)	(0.0843)	(8.9879)
State	(0.0066)	(2.1092)	(0.0049)	(2.043)	(0.0055)	(2.5940)
Design_Load	0.0256	6.2048	0.0389	10.394	0.0783	3.8635
Type_of_Service_On_Bridge	(0.0496)	(7.2942)	(0.0171)	(3.422)	(0.0114)	(0.2850)
Type_of_Service_Under_Bridge	0.0150	0.9485	0.0150	1.985	0.0109	1.2064
ADT	(0.1331)	(18.2042)	(0.1066)	(20.495)	(0.1024)	(15.6959)
ADTT	(0.2764)	(12.3404)	(0.3371)	(11.394)	(0.3118)	(13.4436)
Main_Structure_Material	(0.0565)	(29.3484)	(0.0167)	(34.384)	(0.0248)	(36.0659)
Main_Structure_Design	(0.0154)	(6.2404)	0.0028	4.959	0.0031	2.9685
Number_of_Spans_In_Main_Unit	(0.0004)	(10.2049)	(0.0002)	(8.068)	(0.0026)	(17.5969)
Number_of_Approach_Spans	0.0011	4.1304	(0.0056)	(2.495)	(0.0038)	(4.4054)
Length_of_Maximum_Span	(0.0003)	(4.3420)	0.0015	2.395	0.0019	23.7218
Structure_Length	(0.0002)	(1.3494)	(0.0002)	(3.696)	(0.0002)	(4.6976)
Deck_Width	(0.0089)	(20.4455)	(0.0009)	(10.496)	(0.0012)	(10.6950)
Deck_Structure_Type	(0.0328)	(15.2837)	(0.1104)	(53.395)	(0.1077)	(44.3857)
Age	(0.0428)	(298.3487)	(0.0465)	(267.050)	(0.0426)	(248.7834)
4|5	(4.4284)	(89.3238)	(4.4672)	(86.496)	(4.0269)	(89.3875)
5|6	(3.5583)	(83.4855)	(3.5554)	(88.687)	(3.1945)	(87.3725)
6|7	(2.5583)	(80.3948)	(2.6429)	(81.055)	(2.3088)	(83.9687)
7|8	(1.2373)	(74.4943)	(1.3356)	(77.961)	(0.9138)	(74.2843)
8|9	0.0292	60.4943	0.2495	60.506	0.4217	60.3859
R-Squared	0.3743		0.4092		0.3852	

Table 6-14. PERF2 model estimation result of model (6-4) (continued)

(b) Step 2

	Value	t-value	Value	t-value	Value	t-value
DCR*	N/A	N/A	(0.0408)	(4.0593)	(0.0771)	(8.3836)
SPCR*	0.086	6.783	N/A	N/A	0.0767	9.2632
SCR*	(0.033)	(2.456)	0.0669	7.0212	N/A	N/A
Lanes_On_Structure	N/A	N/A	N/A	N/A	N/A	N/A
Lanes_Under_Structure	N/A	N/A	N/A	N/A	N/A	N/A
State	(0.0066)	(31.2995)	N/A	N/A	N/A	N/A
Design_Load	N/A	N/A	N/A	N/A	N/A	N/A
Type_of_Service_On_Bridge	N/A	N/A	N/A	N/A	N/A	N/A
Type_of_Service_Under_Bridge	N/A	N/A	N/A	N/A	N/A	N/A
ADT	(0.011)	(20.833)	(0.0078)	(14.1487)	(0.0058)	(10.9442)
ADTT	(0.028)	(13.702)	(0.0707)	(1.9789)	0.0179	0.5016
Main_Structure_Material	(0.053)	(33.339)	(0.0360)	(25.3486)	(0.0406)	(28.8145)
Main_Structure_Design	(0.0003)	(0.5635)	0.0033	5.5119	0.0028	4.7554
Number_of_Spans_In_Main_Unit	(0.0045)	(8.1564)	(0.0041)	(7.4474)	(0.0058)	(10.6393)
Number_of_Approach_Spans	0.0001	0.1359	(0.0075)	(8.6625)	N/A	N/A
Length_of_Maximum_Span	(0.0013)	(5.8007)	N/A	N/A	N/A	N/A
Deck_Width	(0.0358)	(12.1839)	N/A	N/A	0.0020	3.1418
Deck_Structure_Type	(0.0069)	(10.6882)	(5.4784)	(60.5797)	(4.8451)	(49.0280)
Age	(0.0424)	(200.0899)	(0.0472)	(216.7022)	(0.0456)	(213.5423)
4|5	(6.5241)	(72.5597)	(4.3103)	(48.0553)	(4.0567)	(201.2973)
5|6	(5.7049)	(63.5095)	(3.4960)	(38.9652)	(3.2418)	(170.8179)
6|7	(4.8495)	(54.0584)	(2.6436)	(29.4744)	(2.3713)	(129.1735)
7|8	(3.6126)	(40.4258)	(1.4781)	(16.5322)	(0.9929)	(56.4096)
8|9	(2.4119)	(27.0541)	(0.0724)	(0.8112)	0.3327	18.7081
R-Squared	**0.3799**		**0.4131**		**0.3901**	

Table 6-15. Comparison of fraction of population for fitted and actual observed rating for each rating level obtained from PERF2 model

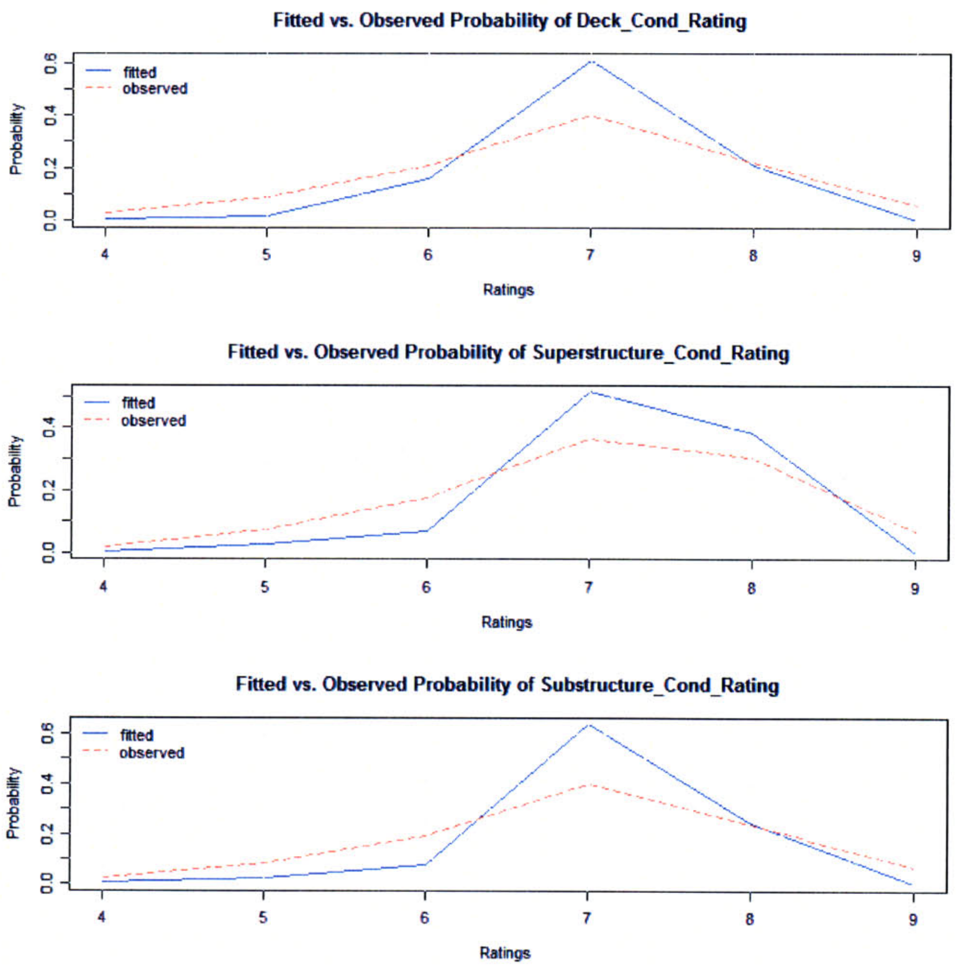
(a) Step 1

	DCR*		SPCR*		SCR*	
	Observed	Fitted	Observed	Fitted	Observed	Fitted
4	0.02730	0.00414	0.01818	0.00313	0.02811	0.01085
5	0.08778	0.01823	0.07133	0.02714	0.08356	0.02697
6	0.20832	0.16026	0.17491	0.07043	0.19044	0.07951
7	0.40173	0.61100	0.36519	0.51530	0.39711	0.63835
8	0.21902	0.20636	0.30410	0.38399	0.23636	0.24431
9	0.05585	0.00001	0.06629	0.00001	0.06442	0.00001

Table 6-15. Comparison of fraction of population for fitted and actual observed rating for each rating level obtained from PERF2 model (continued)

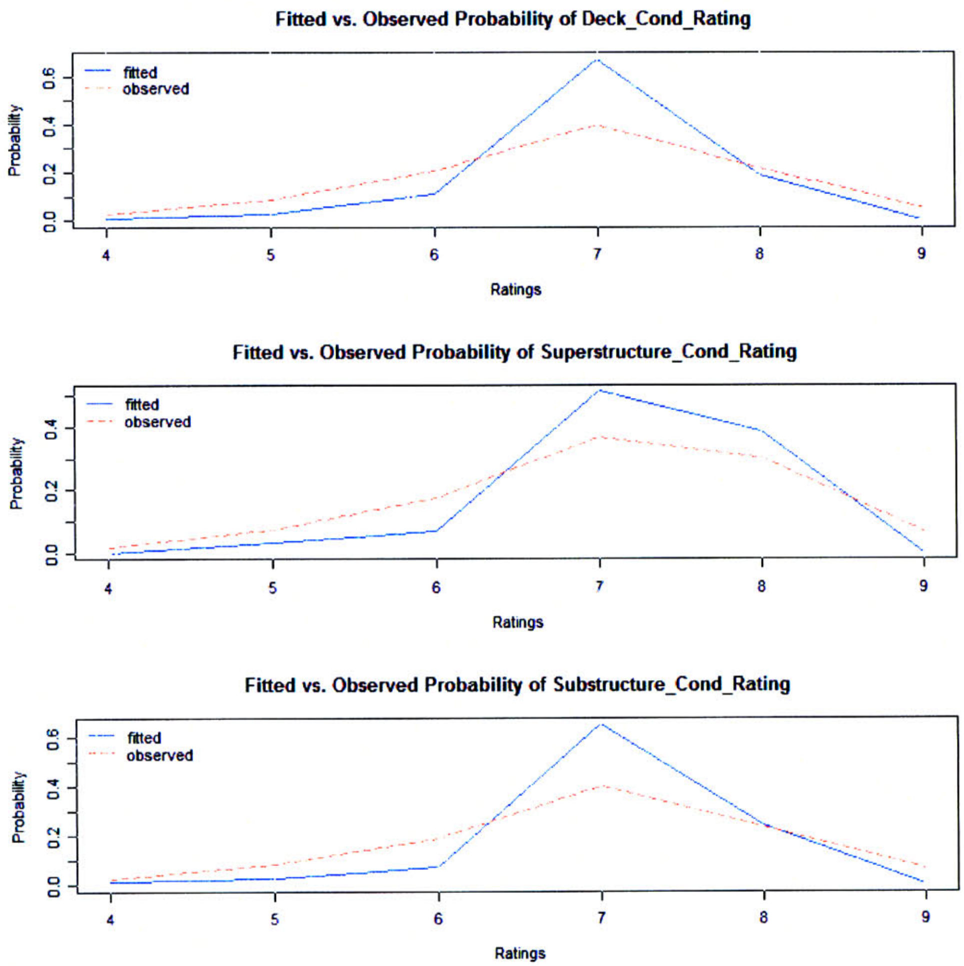
(b) Step 2

	DCR		SPCR		SCR	
	Observed	Fitted	Observed	Fitted	Observed	Fitted
4	0.02730	0.00326	0.01818	0.00052	0.02811	0.01320
5	0.08778	0.01869	0.07133	0.02862	0.08356	0.02488
6	0.20832	0.08954	0.17491	0.06271	0.19044	0.07013
7	0.40173	0.64378	0.36519	0.50348	0.39711	0.64828
8	0.21902	0.24473	0.30410	0.40463	0.23636	0.24351
9	0.05585	0.00000	0.06629	0.00003	0.06442	0.00001



(a) **Step 1**

Figure 6-3. Fitted vs. observed population fraction of each condition rating from PERF2 model (step 1 and 2 estimation)



(b) Step 2

Figure 6-3. Fitted vs. observed population fraction of each condition rating from PERF2 model (continued)

The estimation results show slight improvement over the PERF1 model by 3-5% in R^2 value. This suggests that exclusion of low rating observations and inclusion of more explanatory variables yields little improvement of the explanatory ability of the model.

For the sake of completeness, we have computed out-of-sample predictions and compared to the in-sample ones. The observed and fitted proportion of each ratings for out-of-sample data are tabulated in Table 6-16 and plotted in Figure 6-4. It can be seen that while keeping the distribution of rating population fixed when sampling, the estimated model behaves consistently on the out-of-sample data although low fitting quality still can be observed in this case.

Table 6-16. Comparison of fraction of population for fitted and actual observed rating for each rating level obtained from PERF2 model on the out-of-sample data

	DCR		SPCR		SCR	
	Observed	**Fitted**	**Observed**	**Fitted**	**Observed**	**Fitted**
4	0.02730	0.00326	0.01818	0.00052	0.02811	0.01320
5	0.08778	0.01869	0.07133	0.02862	0.08356	0.02488
6	0.20832	0.08954	0.17491	0.06271	0.19044	0.07013
7	0.40173	0.64378	0.36519	0.50348	0.39711	0.64828
8	0.21902	0.24473	0.30410	0.40463	0.23636	0.24351
9	0.05585	0.00000	0.06629	0.00003	0.06442	0.00001
Total	1.00	1.00	1.00	1.00	1.00	1.00

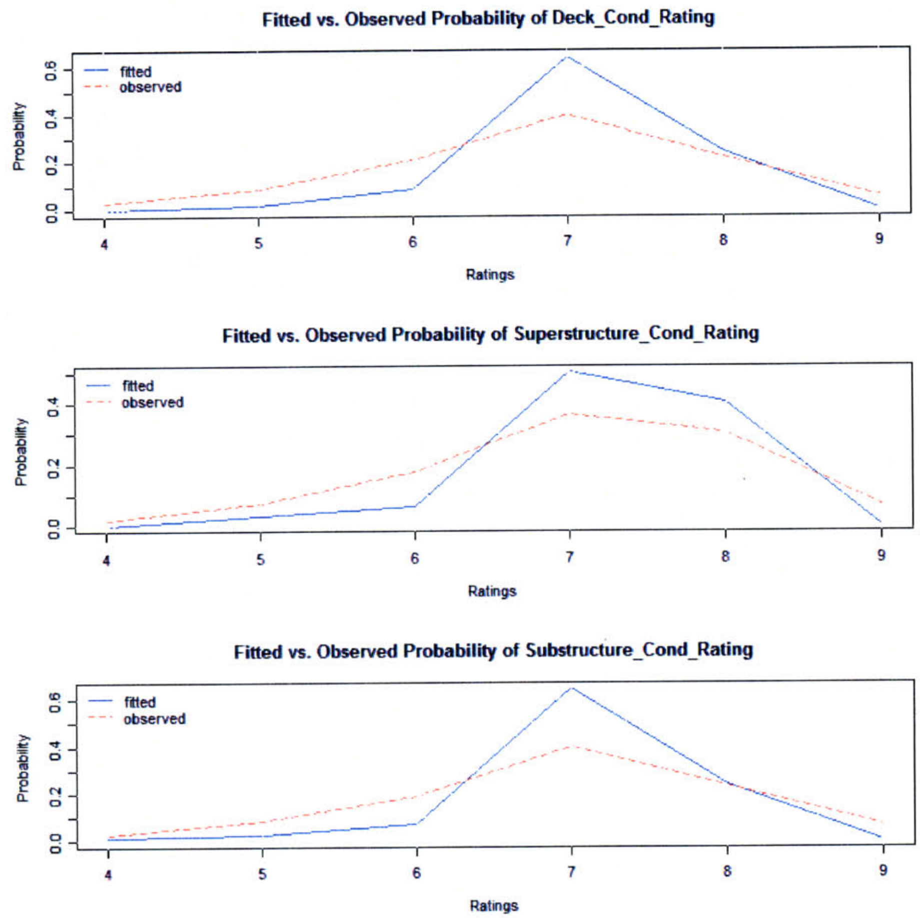


Figure 6-4. Fitted vs. observed population fraction of each condition rating from PERF2 model on the out-of-sample data

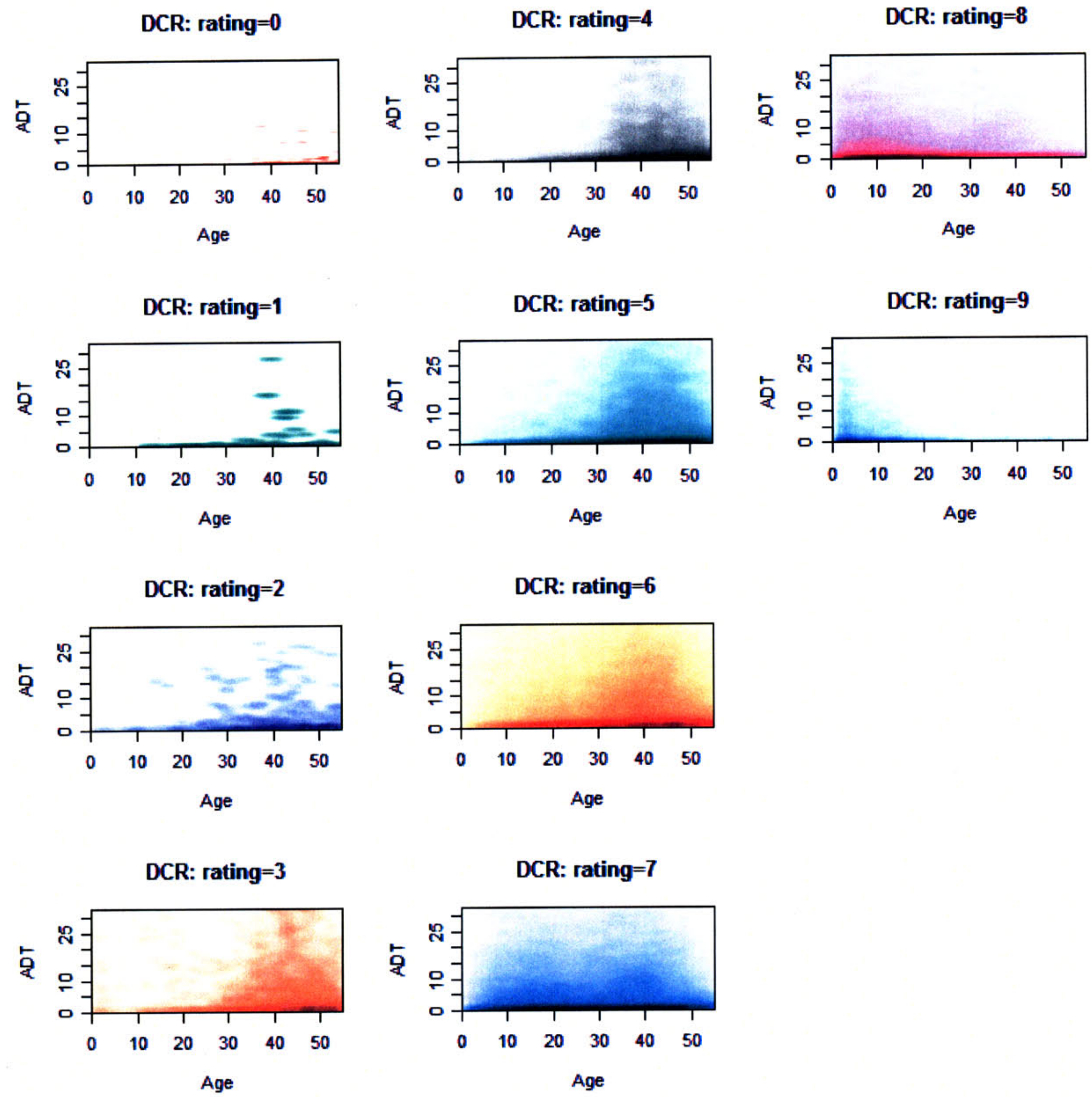
6.6. Discussions

6.6.1. Visualizations of data conditional distributions

To help gain insight of low quality of fitting from the model perspective, visualization of different joint distribution of data on selected attributes is performed.

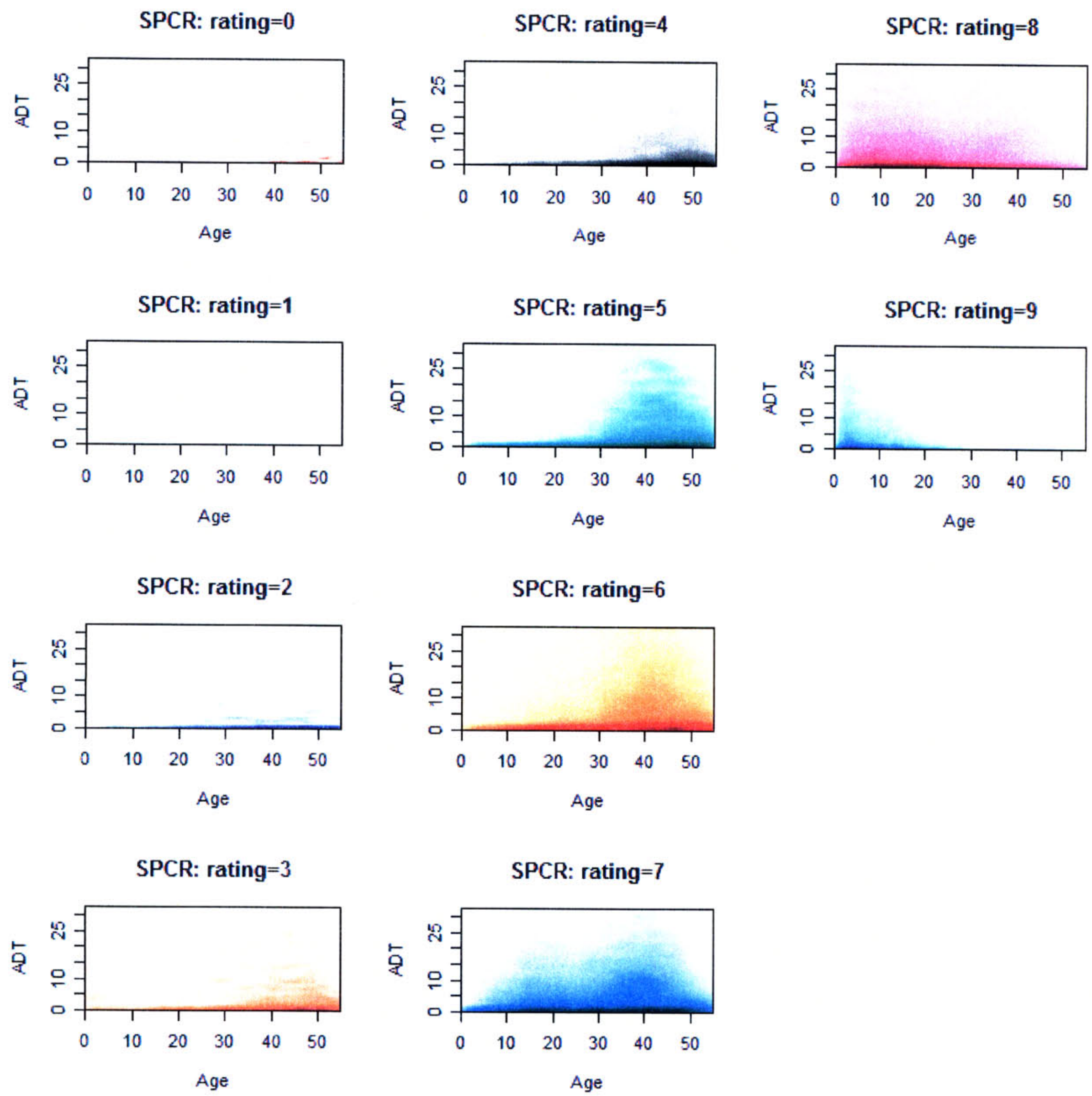
1) *Data Distribution of DCR, SPCR, and SCR conditioned on AGE and ADT (Note that ADT is in x1000 scale)*

   Comments: Throughout the plots (Figure 6-4), one can observe that the mass of ratings 0-5 mainly centers around old ages (more than 20 years) and relatively low volume of traffic. Where ratings of 8-9 tend to center around ages of 1-20 years with, again, low ADT. Exception is in ratings 6-7 in which population mass spans across all ranges of ages. Order probit model discriminates data by identifying a set of parallel linear planes that best separate data. Plus, in term of population number, rating 7 holds the majority of the entire population. From these evidences, it is quite intuitive that discriminant planes would be mainly influenced by rating 7 which will likely to fold in ratings 0-6 and some of ratings 8-9.



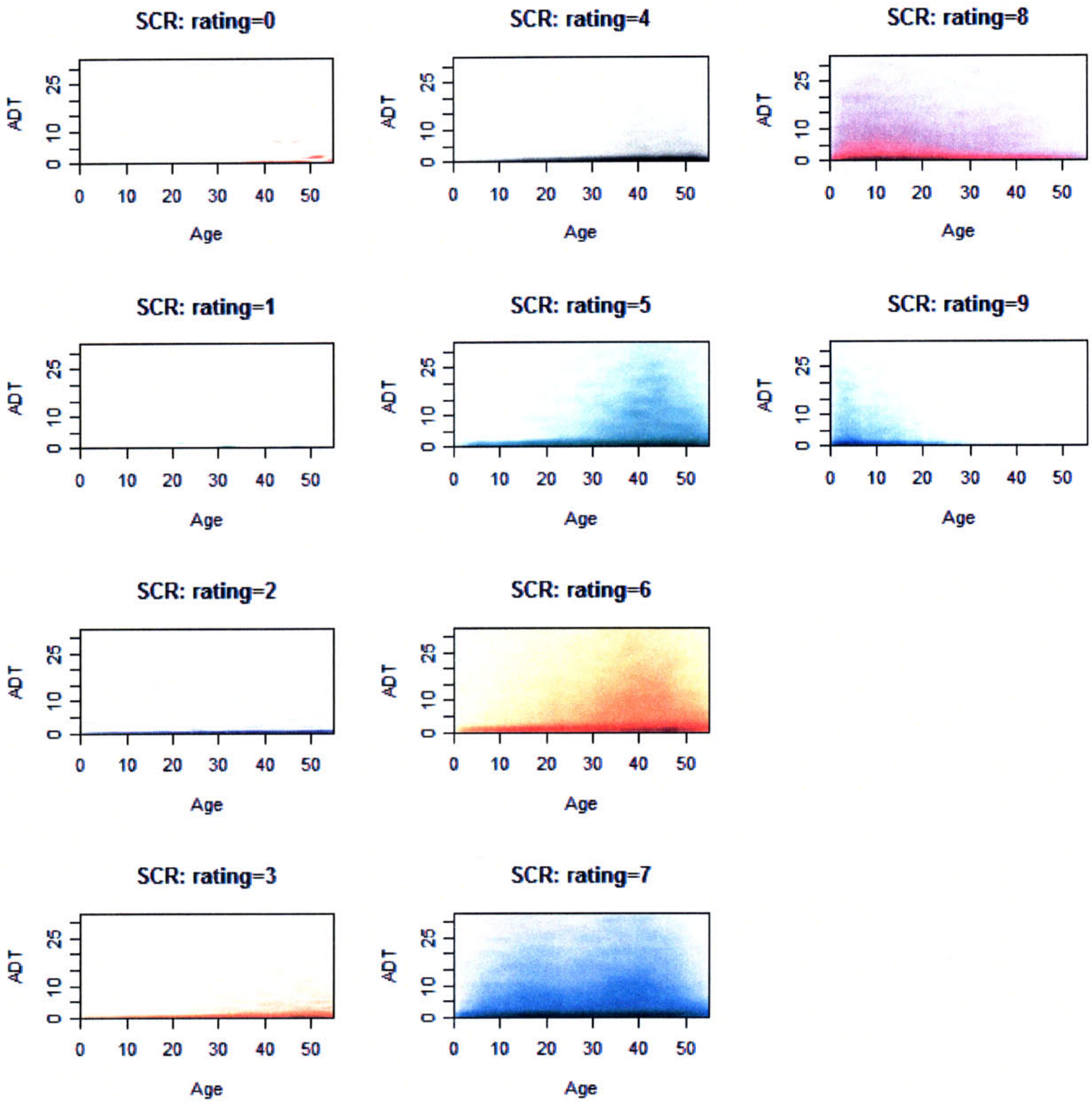
(a) DCR

Figure 6-5. Distributiion of data for each condition rating on Age and Average Daily Traffic space



(b) SPCR

Figure 6-4. Distribution of data for each condition rating on Age and Average Daily Traffic space (continued)

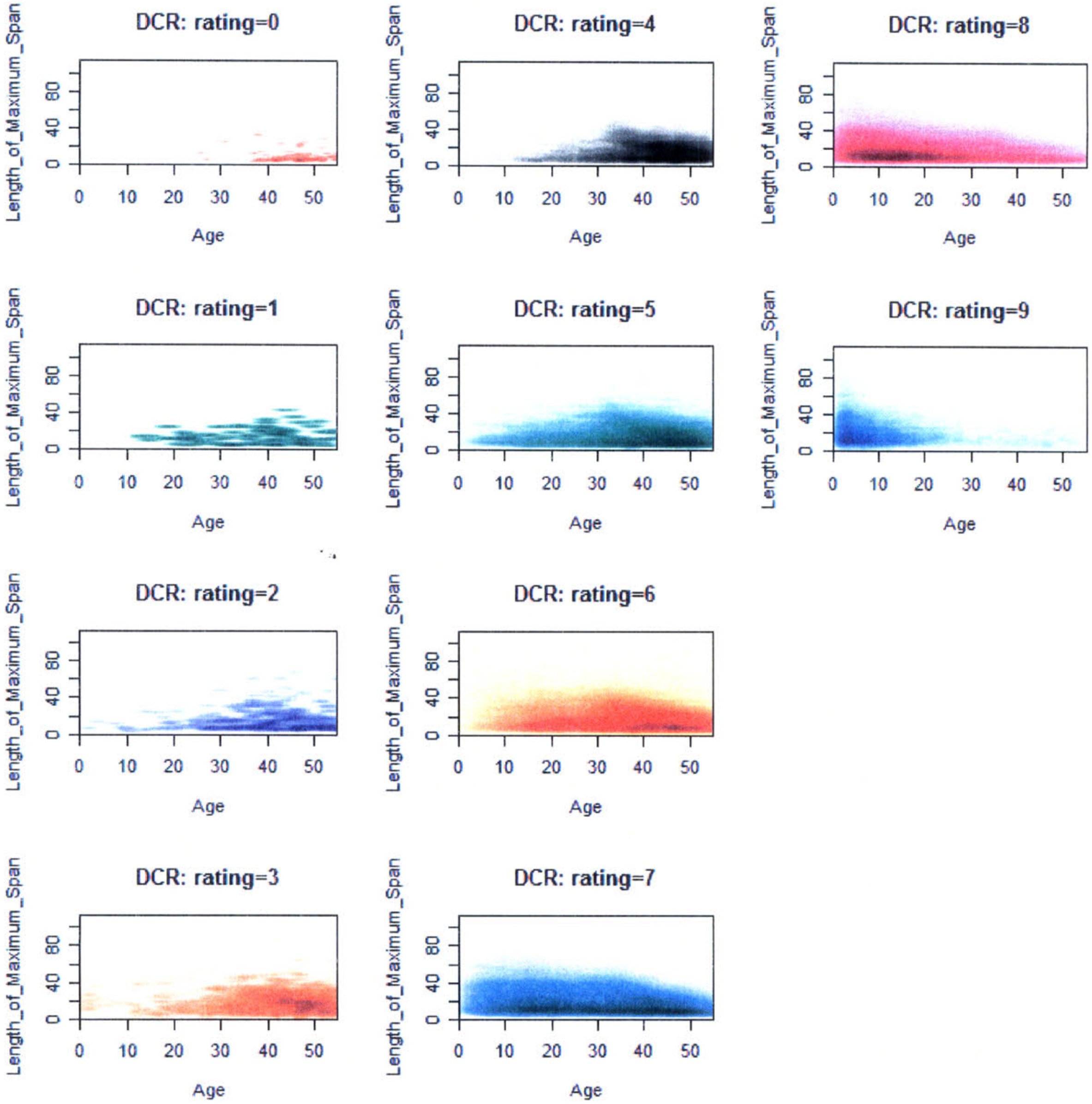


(c) SCR

Figure 6-4. Distribution of data for each condition rating on Age and Average Daily Traffic space (continued)

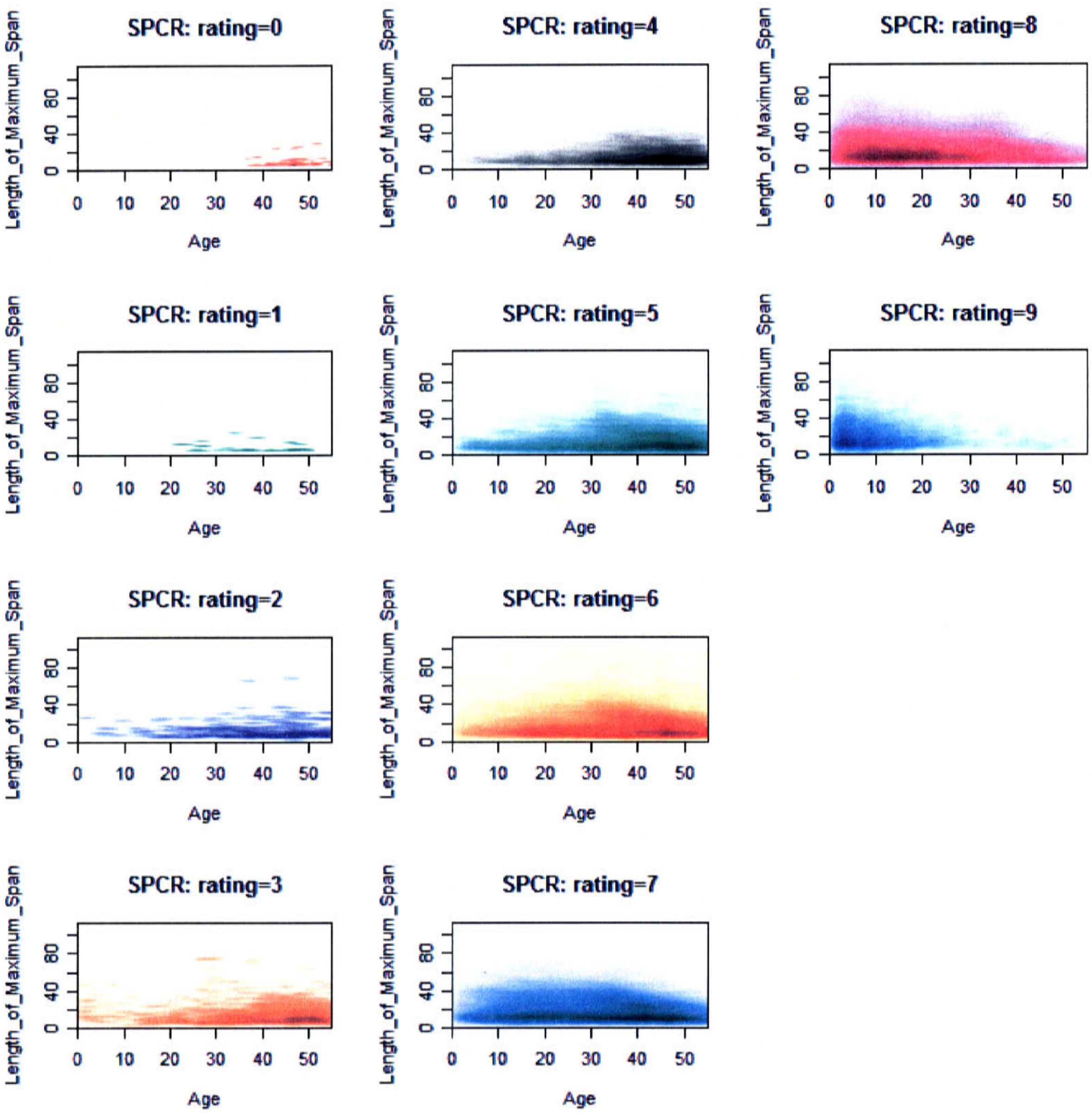
2) *Data distribution of DCR, SPCR, and SCR conditioned on AGE and Length of Maximum Span*

<u>Comments</u>: Similar trend but more emphasized can be found in these plots (Figure 6-5).



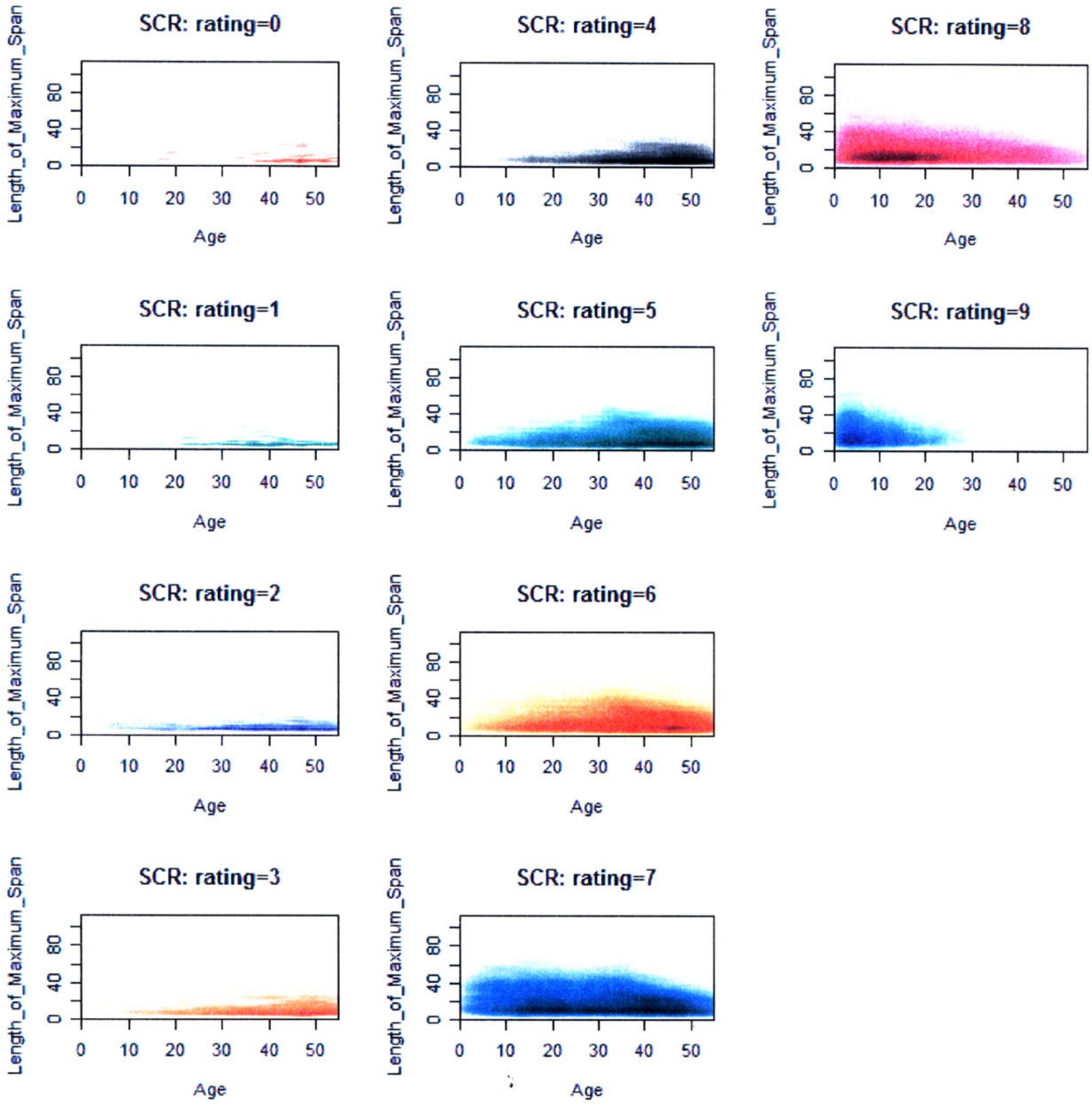
(a) DCR

Figure 6-6. Distribution of data for each condition rating on Age and Length of Maximum Span space



(b) SPCR

Figure 6-5. Distribution of data for each condition rating on Age and Length of Maximum Span space (continued)

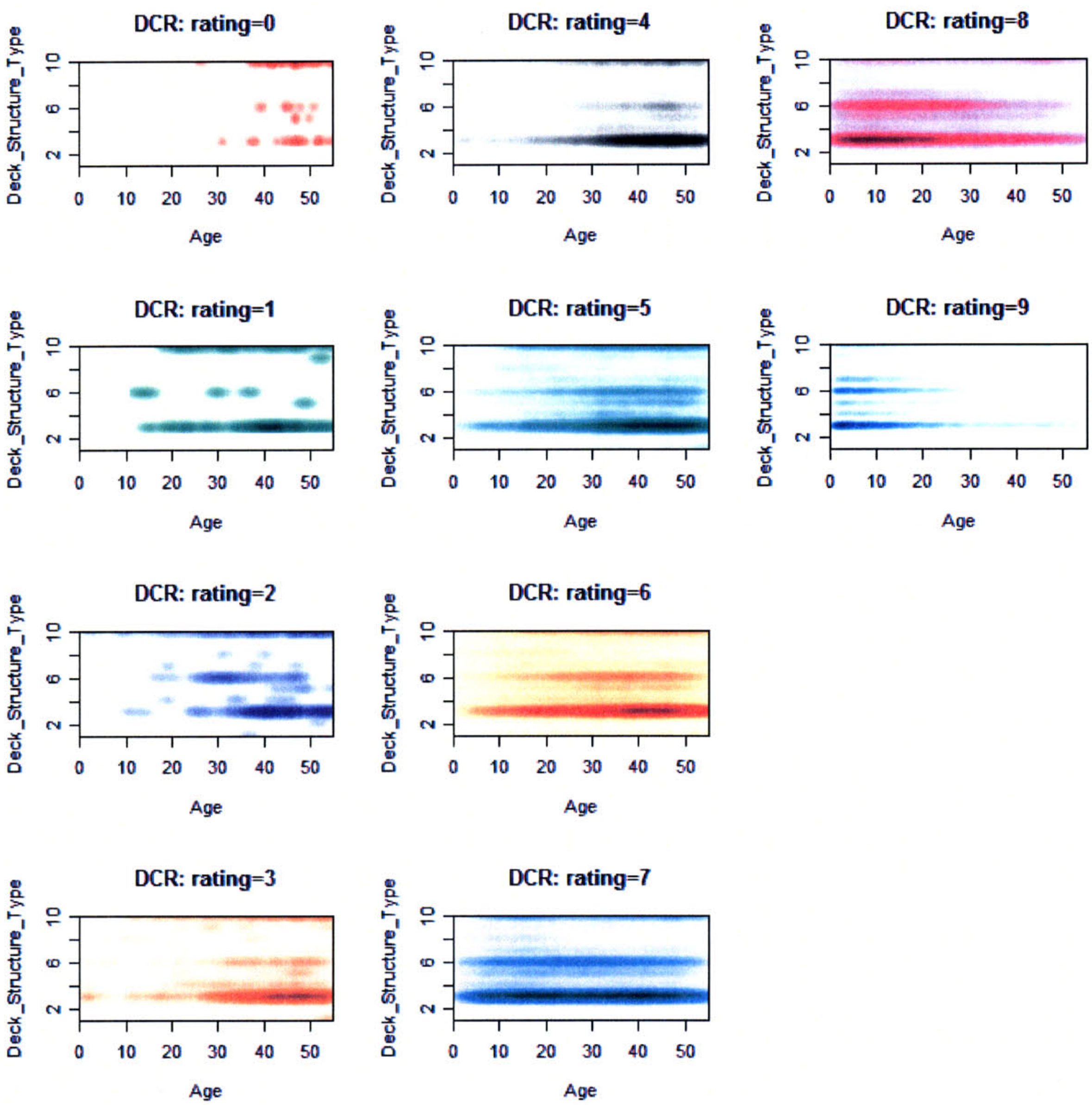


(c) SCR

Figure 6-5. Distribution of data for each condition rating on Age and Length of Maximum Span space (continued)

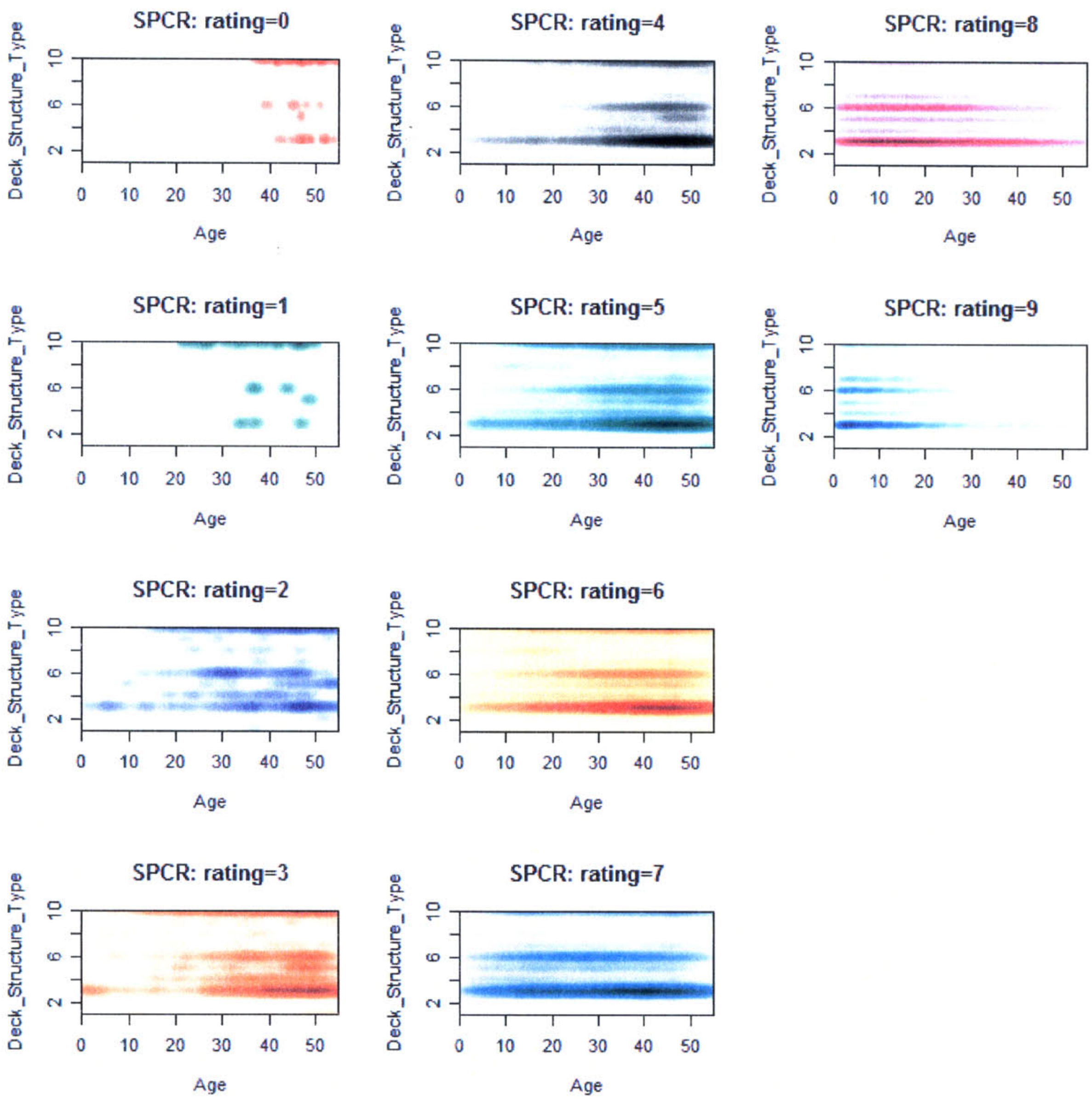
3) Data distribution of DCR, SPCR, and SCR conditioned on AGE and Deck Structure Type

Comments: The majority of bridges are designed with Deck Structure Type of 2, 6 and 10 (i.e. Closed Granting, Not Applicable, and Wood or Timber). While those with conditions 0-6 tends to have heavier mass on high ages range, ratings from 8-9 centered around 1-40 years. Exception is rating 7 group in which all ranges of age seem to appear. See Figure 6-6 for more details



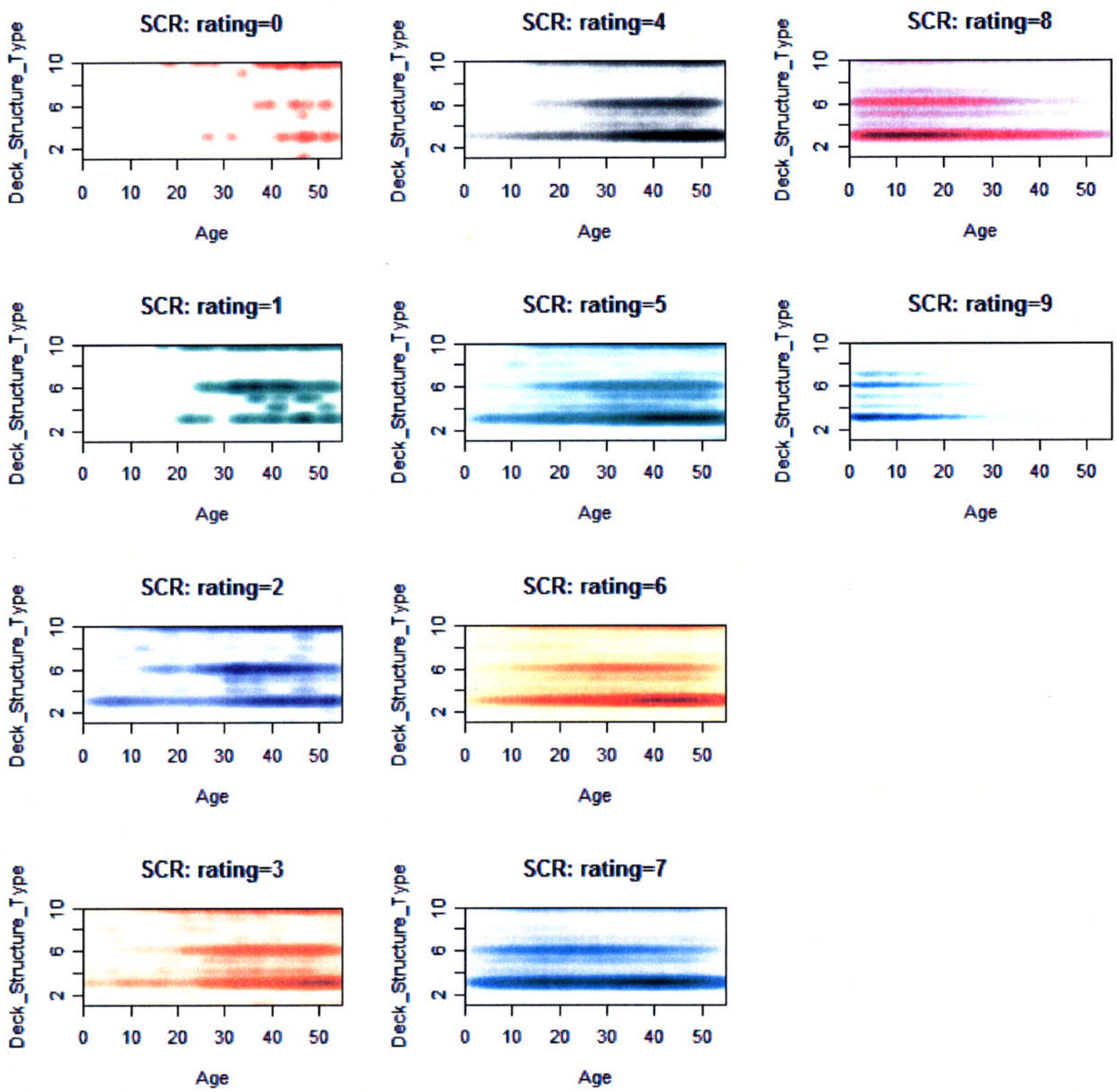
(a) DCR

Figure 6-6. Distribution of data for each condition rating on Age and Deck Structure Type space



(b) SPCR

Figure 6-6. Distribution of data for each condition rating on Age and Deck Structure Type space (continued)



(c) SCR

Figure 6-6. Distribution of data for each condition rating on Age and Deck Structure Type space (continued)

From these visualizations of data distributions, a hint as to why the linear model such as ordered probit model would exhibit poor fitting can be easily drawn. Recall that the ordered probit model separates the probability space into different regions for classification purposes by straight lines (in case of 2D space). However, using analogy of these joint distribution plots illustrated in Figure 6-7, where pink, green and orange dots represent rating 0, 7 and 9 respectively. It is almost impossible to draw lines such that data in low rating ranges can be distinguished from those with high ratings. Moreover rating-7 population clouds almost over the regions of the entire data space. Hence, it is not surprising that ordered probit model will try to accommodate this relatively large-sized population with wide window of cutting points. As a result, we see that the model tend to classify most of the data into rating 7.

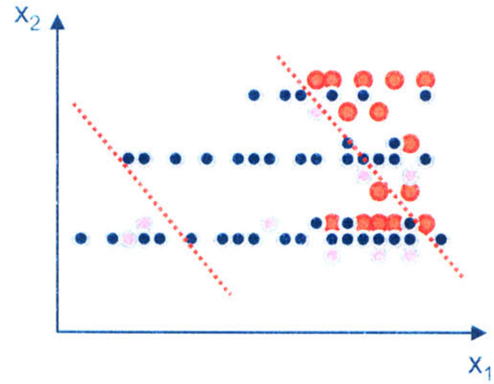


Figure 6-7. Illustration of failure of ordered probit model

6.6.2. Effects of regional environmental information

We compared our results with similar type of analysis conducted by Chase, Small and Nutakor (1999). They use NBI data along with GIS regional environmental data such as precipitation rate, temperature range, etc. to fit the physical condition ratings. Note that they also exclude ratings 0-2 out of the analyses due to the same issue that we have found.

For comparison purpose, we have computed Residual Square Error (RSE) on our in- and out-of-sample data and compared with them. Everything similar except their richer set of explanatory variables and smaller set of sample population of size 30,000, the RSE comparison is charted in Figure 6-8.

A big difference in the level of RSE can be observed between PERF2 and Chase's result. PERF2 exhibits RSE of about 1.0 whereas the Chase's are about 0.15-0.18. Intuitively, incorporation of geographical environment information shall improve the result of the fit. However, given that the quality of physical condition rating measurement which typically errs by +/-1 (to be discussed in the next subsection), it is unclear to what extent additional information on geographical environment could add more value to the accuracy of the model. Another important point to note is the detail of sampling which is not clearly stated by Chase's study. Since the estimation is done via MLE which is indeed a fitting of observed distribution of the regression error tem, the underlying density must be preserved when sampling to reflect the actual observation. In an extreme case, one can sample the data so that populations from each rating are of equal proportion. In this case, the problem of tight cutting points which reflects the problem of underlying data population would have not been realized in the model. By this way the result from the model can be spurious.

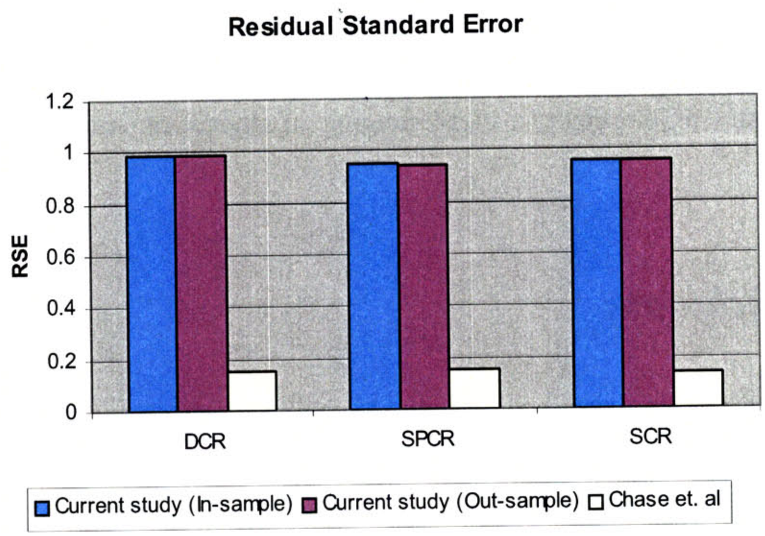


Figure 6-8. Comparison of RSE from in- and out-of-sample data of PERF2 model and that of Chase, Small and Nutakor (1999) study

6.6.3. Accuracy of physical condition rating measurement

Another source (and probably the main source) that contributes most to the poverty of fitness quality in our model is due to low accuracy of the physical condition rating measurement reported in NBI database.

Washer (2003) has conducted a field study to empirically measure the accuracy of visual inspection by inspectors and the nondestructive evaluation (NDE). The study was conducted by asking 49 practicing bridge inspectors from across the country to examine the test bridges in Virginia and Pennsylvania. Each inspector performed 10 separate tasks, including routine and indepth inspections. They used common hand tools such as a masonry hammer, plumb bob, carpenter's level, binoculars, and other nonintrusive tools.

During the routine inspections, the inspectors were asked to provide a condition rating for the superstructure, substructure, and deck. The study revealed a wide distribution of condition ratings reported by inspectors evaluating the same bridge sections. On average, they assigned between four and five condition ratings for each separate component. For some components, inspectors provided as few as three different condition ratings; for others, inspectors provided as many as six. The average was between four and five. Reportedly, the results indicated that only 68% of the reported condition rating for these elements would vary between +/-1 from the average rating for a particular element.

This obviously will have a direct impact on quality of fitness of the model. Now the source of uncertainty in the observations consists of two sources following the context of latent variable models such as the ordered probit model. The first source is, of course, the deterioration process itself which we tried to capture it by ordered probit model. The second source is from the measurement error. Without knowing the structure of probability density of the error, a clear gauging of impact on model accuracy is not possible.

6.7. Conclusions

Presented in this chapter is an empirical evaluation of SEM model on NBI bridge physical condition ratings modeling. Using model with only design description and the one with additional design specifications, we have found that the quality of fit is generally poor with only information available in NBI. By comparing to study done by Chase (1999) in which regional environment information is included, the result suggested that with additional environmental information does improve model accuracy. However, the degree of improvement is dubious because the accuracy of the measured ratings is reportedly low. This also serves as a cause of poor fitness exhibited in our model for both PERF1 and PERF2. Finally, from the model perspective, we found from the visualization analysis that the simple ordered probit model which draws lines (or planes) on the normal density space can perform poorly because the overall data is shadowed by rating 7.

PART III: LEARNING OF CONCEPTUAL DESIGN KNOWLEDGE IN THE PRESENCE OF DESIGN OBJECTIVES

7. METHODOLOGY FOR LEARNING BRIDGE DESIGN KNOWLDGE

THE ADVENT OF POWERFUL computational powers has given rise to the desire of intelligent systems which incorporate design knowledge which can be generalized to solving new design problems. Quite a volume of studies have been devoted to development of a so-called *inductive* Machine Learning (ML) algorithms whereby knowledge of the domain is constructed by induction over a given set of examples. In design contexts, the idea of acquiring design knowledge from previous design examples is an appealing one (Potter, et al., 2001). In this chapter, we briefly reviewed the methodology employed in literatures on engineering design knowledge learning and conclude that the approach of conceptual clustering is more suitable for our learning goal. Finally, we proposed a new clustering measure which aims toward building cluster structure such that it pays more attention to capturing characteristics of designs that leads to high performance.

7.1. Review of Inductive ML Techniques for Engineering Design Knowledge Learning

There are typically two types of learning depending on the form of obtained knowledge (Potter, et al., 2001). The first paradigm is to learn and obtain structure of the knowledge such as case base or hierarchical tree. One can view this type of learning scheme as acquisition of classification knowledge. On the other hand, the second type of obtained knowledge is in a functional form. Examples of resulting functions are neural networks, support vector machines, etc. This type of knowledge may or may not be comprehensible for human or engineers. In the below subsections, we devote our discussion in more detail to review each of these approaches.

7.1.1. Conceptual Clustering Approach

The most popular technique used in the first type of learning is a so-called *conceptual clustering*. Based on the defined measure of similarity used for clustering strategy, design examples are grouped together and usually in a hierarchical structure. Each cluster represents a "concept". The more upstream a concept lies on the structure, the more abstract or general design concept a cluster represents. The solutions under each cluster are used as basis for derivation of solution to new design problem. The followings are brief reviews of representative techniques in the class of conceptual clustering.

Decision Tree (ID3 or C4.5)

One of the techniques for conceptual clustering that has been widely used is a so-called "decision tree" algorithm. It is also known as ID3 or C4.5 developed by Quinlan (1986 and 1993). Decision trees normally consist of a number of nodes, each representing an attribute with different instances that are used to classify the test cases and build the tree. When decision trees are built, they are usually pruned and converted to a set of rules that can be easily understood and applied by users. Examples of study for acquisition of knowledge using this algorithm are conceptual fixtures design by Kumar, Subramaniam and Teck (2004), and construction of critics based on labeled living design space to aid criticism of new design configuration by Williams (2003).

Most of the decision trees employed in various research are variations of the core algorithm exemplified by the ID3 (Quinlan, 1986) and its successor C4.5 (Quinlan, 1993). It utilizes a top-down, greedy search through the space of possible decision trees. For the sake of further discussion, we briefly review the basic algorithm ID3 below.

Given a set of examples in form of attribute-value, the ID3 is grown by recursively branching its leaf with the attribute which is *most useful* for classification of supplied examples. The usefulness measure here is measured by a statistical quantity, called *information gain*. The information gain can be viewed as expected reduction of entropy caused by partitioning examples by an attribute. Here, entropy can be considered as a measure of impurity at a particular node with respect to class labels of the node local data. Precisely, an entropy function for K-class problem at node m can be defined as:

$$I_m = -\sum_{i=1}^{K} p_m^i \log p_m^i \tag{7-1}$$

where p_m^i is the probability of class C_i at node m, i.e.

$$P(C_i \mid x, m) \equiv p_m^i = \frac{N_m^i}{N_m} \tag{7-2}$$

The information gain from using attribute A to branch node m is simply a reduction of entropy and thus can be defined as:

$$Gain(m, A) = I_m - I_{m,a} = -\sum_{i=1}^{K} p_m^i \log p_m^i - \left(-\sum_{j=1}^{n} \frac{N_{mj}}{N_m} \sum_{i=1}^{K} p_{mj}^i \log p_{mj}^i \right) \tag{7-3}$$

where N_{mj} denotes number of observations in node m that takes value j at attribute A. And p_{mj}^i the probability of class C_i at node m that takes value j at the branching attribute is defined as

$$P(C_i \mid x, m, j) \equiv p_{mj}^i = \frac{N_{mj}^i}{N_{mj}} \tag{7-4}$$

Thus, the ID tree can be considered as an algorithm that branches at the node which yields greatest expected reduction of entropy. Pseudo code of ID3 algorithm (for binary classification on categorical data attributes) is provided in Figure 7-1.

ID3 (Examples, Target_Attribute, Attributes)

- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root, with label = +.
- If all examples are negative, Return the single-node tree Root, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Otherwise Begin
    - A = The Attribute that best classifies examples.
    - Decision Tree attribute for Root = A.
    - For each possible value, v_i, of A,
        - Add a new tree branch below Root, corresponding to the test A = v_i.
        - Let Examples(v_i), be the subset of examples that have the value v_i for A
        - If Examples(v_i) is empty
            - Then below this new branch add a leaf node with label = most common target value in the examples
        - Else below this new branch add the subtree ID3 (Examples(v_i), Target_Attribute, Attributes – {A})
- End
- Return Root

Figure 7-1. Pseudo-code for ID3 algorithm (Mitchell, 1997)

It should be noted that the major drawbacks of ID3 include the followings: missing values, continuous attribute and tendency to overfit data. C4.5 is the successor of ID3 that accounts for these shortcomings.

Cobweb

COBWEB (Fisher, 1987) is a hierarchical clustering technique employing five operators to determine how best to incorporate an example (e.g. existing design) into the hierarchy. The category utility (Gluck and Corter, 1985) is use as clustering objective functional. It can be viewed as a function that rewards traditional virtues held in clustering generally – similarity of objects within the same class and dissimilarity of objects in different classes. Classification topology is constructed in such a way to maximize average Category Utility (CU) over all clusters on the hierarchy. Precisely, the CU function for the k-th cluster is defined as:

$$CU_k = P(C_k)\left[\sum_i \sum_j P(A_i = V_{ij} \mid C_k)^2 - \sum_i \sum_j P(A_i = V_{ij})^2\right] \qquad (7\text{-}5)$$

where $P(A_i = V_{ij} \mid C_k)$ is the probability of the i-th attribute of the observed data taking the j-th label value given that the observation is classified to the k-th cluster. The first time can be rewritten as follow using Bayes rule.

$$\sum_i \sum_j P(A_i = V_{ij} \mid C_k) P(C_k) P(C_k \mid A_i = V_{ij}) \tag{7-6}$$

The first term in the product of (7-6) can be interpreted as *intra-class* similarity. The larger this probability, the greater the proportion of class members sharing the value and the more *predictable* the value is of class member. The last term is the *inter-class* similarity. And the higher this probability is, the fewer the objects in contrasting classes that share this value and the more *predictive* the value is of this class.

Thus, the term defined in (7-6) can be viewed as a tradeoff between intra-class similarity and the intra-class dissimilarity. The CU function is therefore defined as the gain of expected number of attribute values that can be correctly guessed given cluster k ($P(C_k)\sum_i \sum_j P(A_i = V_{ij} \mid C_k)^2$) over the expected number of correct guesses with no such prior knowledge ($P(C_k)\sum_i \sum_j P(A_i = V_{ij})^2$).

Examples are permanently incorporated into the hierarchy by sorting through the hierarchy and find the best host node that maximizes average CU over all clusters, i.e.

$$\frac{\sum_{k=1}^{n} P(C_k)\left[\sum_i \sum_j P(A_i = V_{ij} \mid C_k)^2 - \sum_i \sum_j P(A_i = V_{ij})^2\right]}{n} \tag{7-7}$$

Note that COBWEB in its original version can only handle nominal data. To account for this, the idea is to replace the discrete prior and conditional probabilities above with integral over probability density functions. Example calculation of term $\sum_i \sum_j P(A_i = V_{ij} \mid C_k)^2 - \sum_i \sum_j P(A_i = V_{ij})^2$ in equation (7-5) assuming normal distribution with an observed mean μ and standard deviation σ is (Haglin, 2007):

$$\sum_i \sum_j P(A_i = V_{ij} \mid C_k)^2 - \sum_i \sum_j P(A_i = V_{ij})^2 \Leftrightarrow \sum_i \left(\int f(A_i \mid C_k)^2 \, dA_i - \int f(A_i)^2 \, dA_i\right) \tag{7-8}$$

$$\sum_i \left(\int f(A_i \mid C_k)^2 \, dA_i - \int f(A_i)^2 \, dA_i\right) = \frac{1}{2\sqrt{\pi}} \sum_i \left(\frac{1}{\sigma_{ik}^2} - \frac{1}{\sigma_i^2}\right) \tag{7-9}$$

Note that unlike ID3, the COBWEB is an incremental learning algorithm. That is, the tree is updated every time a new example is presented to it. COBWEB tries to accommodate the new example into an existing hierarchy starting from root then recursively traversed through the tree by the following operators (see Fisher, 1987).

1. Expanding the root, if it does not have any sub-classes, by creating a new class and attaching the root and the new example as its sub-classes;

2. adding the new example as a new sub-class of the root;
3. adding the new example to one of the sub-classes of the root;
4. merging the two best sub-classes and putting the new example into the merged sub-class; or
5. splitting the best sub-class and again considering all the alternatives.

If the example has been assimilated into an existing sub-class, the process recurses with this class as the top of a new hierarchy. COBWEB again uses category utility to determine the next operator to apply. Figure 7-2 summarizes the sketch for COBWEB algorithm.

```
COBWEB (Object, Root)

  • Update counts of the Root
  • IF Root is a leaf
    THEN Return the expanded leaf to accommodate the new object
    ELSE Find that child of Root that best hosts Object and
            perform one of the following
                  a) Consider creating a new class and do so if appropriate
                  b) Consider node merging and do so if appropriate and
                        call COBWEB (Object, Merged node)
                  c) Consider node splitting and do so if appropriate and
                        call COBWEB (Object, Root)
                  d) IF none of the above (a, b, or e) were performed
                    THEN call COBWEB (Object, Best child of Root)
```

Figure 7-2. Pseudo-code for COBWEB algorithm (Fisher, 1987)

One can view node merging and splitting as inverse operators (Fisher, 1987). They allow COBWEB to move bidirectionally through a space of possible hierarchies. In general, merging is invoked when initial observations suggest that the environment is a space of highly similar objects, relative to the actual structure of the environment suggested by subsequent observations. In contrast, splitting is invoked when the environment is more 'compressed' than suggested by initial input. One can view merging and splitting as a mechanism to decrease the sensitivity of COBWEB to input ordering due to their inverse relation.

7.1.2. Artificial Neural Networks (ANN)

An artificial neural network - ANN comprises a network of a minimal unit called perceptron. Each perceptron takes in input signals, an internal processing function called "threshold function", and an output. The input may come from environments or from the outputs from other perceptrons. Each input signals are *voted* with weights attached to the arch connected to the destined perceptron. The result of voting from inputs is then thresholded by the threshold function and is output. While the individual perceptrons are relatively simple, when connected into a network they can display complex behavior.

The archs connecting between perceptrons are associated by numerical weighting which serves to "amplify" or "diminish" the strength of the numerical signals passed along it. And the core of the learning by ANN is to search through feasible space formed by real vector represented by these weights and locate the optimal solution with respect to a given objective function. Examples of these objective functions could be binary error (for classification problem), L2 norm error (for regression problem), etc. (see Alpaydin, 2004)

The widely-used network is a feed-forward network and can be "trained" to respond to certain input patterns by producing some associated output. The learning or adjustment of network weights could be performed by a so-called "backpropagation" algorithm (Rumelhart et al., 1986) Using this, the network is trained by repeatedly presenting examples of the correct combinations of inputs and outputs to the network, and gradually altering the connection weightings so that, for every example, the input produces the desired network output.

Unlike ID3 or COBWEB, ANNs are subsymbolic learners: they expect data to be represented numerically (and typically as normalized values between 0 (or -1) and 1). To learn more complex associations, ANNs tend to need greater numbers of intermediate perceptron units—and more layers of these units.

Examples of application of ANNs in the area of engineering design knowledge learning consist of:

- Batill and Swift (1993) applied ANN technique to learn structural design configurations that optimize structural performance. The problems concerned are: the configurational design of a 10 bar truss for minimum weight, and the configurational design of a four spar light aircraft wing-box with weight, displacement, and natural frequency as constraints/objective functions.
- Srinivas and Ramanjaneyulu (2007) used Genetic Algorithms (GA) in combination with ANN for cost optimization of bridge deck configuration. In this study, ANN is used for prediction of bridge response given design configuration. GA is then used to search through feasible design configuration space that results in optimal cost.

7.2. Selection of Methodology for Bridge Conceptual Design Knowledge Learning with Performance Constraints

In this section, we shall proceed by first comparing the aforementioned learning paradigms and techniques with respect to characteristics of knowledge that we seek to obtain for our study. By investigation of these algorithms on several dimensions required by our objective knowledge, we can then derive a preferable choice of methodology to be used for bridge conceptual design task.

According to Reich (1997), Machine Learning (ML) techniques can be characterized along many dimensions that influence the applicability of the techniques. Some selected dimensions are:

1) Complexity of input data representation and learned knowledge;
2) Mechanisms for learning knowledge and the functional form of these resulting models/functions;
3) Mode of learning (i.e. batch or incremental);
4) Computational complexity of the algorithm;
5) Comprehensibility (clear to opaque), and
6) Timing in which the knowledge is learned (early or late).

Examples of comprehensibility of the learned knowledge here include extremists like clear for the case of rules and opaque for the case of black-box systems such as neural networks.

Here *early* learning means that learning pro-actively occurs when receiving new data. The algorithm learns from data and stores the knowledge that can be subsequently used for new problem solving. On the other hand, *late* learning refers to reactive learning by storing data and subsequently retrieves it, learn from it (locally), and adapt to the problem.

The last two characteristics were specially paid attention by Reich as he postulated that they form three clusters of machine learning approach into which most of the studies on machine learning in civil engineering can be classified. The clusters are shown in Figure 7-3.

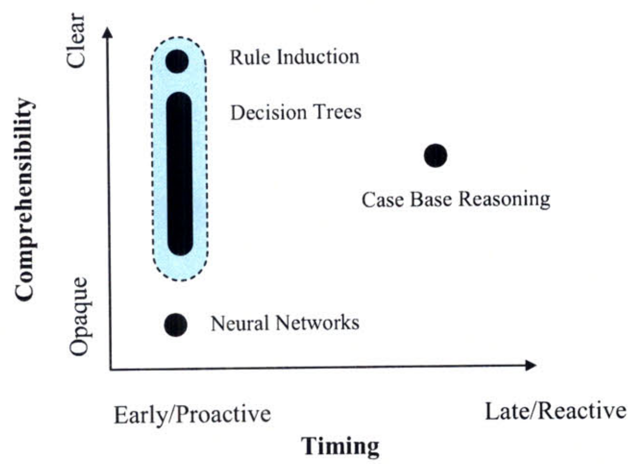


Figure 7-3. Key dimensions of learning systems (Reich, 1997)

From our perspective, we are specially interested in classify best groups of existing designs that yield high performance (in our case, we use physical condition ratings as target) so that the resulting knowledge provides insight on which design configuration conditioned on given sets of requirements leads to high performance and vice versa. Therefore, clear comprehensibility of learned knowledge is indispensable. It follows that a functional form type of knowledge obtained from ANN or similar type of algorithms would not suffice this requirement. Rather, a well-structured knowledge such as those obtained from rule induction or hierarchical structures is more preferable.

On the other hand, the timing aspect of learning algorithm for our application is not so important. However, late or reactive learning such as K-nearest neighbor algorithm requires search for local similarity based on certain distance measure. This could pose difficulty to our application because it is unclear how to model distance measure on a mix of ordinal and nominal data exhibited by NBI data. Thus, methods such as decision tree or COBWEB which is less sensitive to this type of distance measure are more suitable.

Last but not least, in term of complexity of the knowledge and computational effort for hierarchical clustering algorithms, algorithm like COBWEB could be far more efficient than ID3. This is because COBWEB is an incremental learning algorithm and thus does not require re-training like does ID3. A new example can be presented and added to the existing structure. COBWEB is also constructed such that it confines itself into small and useful number of clusters/classes compared to ID3. Reich (1992) reported that the COWEB in the application of bridge conceptual design domain possesses branching factor of about 2.8 and tree size normalized by number of samples consistently equal to 1.5. This can be seen by the fact that COBWEB involves bidirection operators (i.e. merging and splitting) that helps further abstraction/generalization of knowledge by smaller size of cluster numbers.

With these reason, we shall prefer COBWEB to ID3 and other algorithms for application to this thesis. In the next subsection, we will discuss some shortcomings of the COBWEB that needs further adjustment to suffice our thesis purpose. We will also propose appropriate adjustments in detail.

7.3. Proposed Modification for COBWEB

Careful examination of Category Utility (CU) function in equation (7-7) reveals that COBWEB trades off between predictive accuracy of feature values of a class, i.e. $\left[\sum_i \sum_j P\left(A_i = V_{ij} \mid C_k\right)^2 - \sum_i \sum_j P\left(A_i = V_{ij}\right)^2\right]$ term, and the class size $P\left(C_k\right)$. The class size term could lead to bias toward larger category size (Choi and Kim, 2005). Therefore, the concept hierarchy COBWEB constructs may not reflect the underlying class structure of instances.

From the thesis perspective, this hinders the algorithm to focus on design solutions that lead to high performance. Basili, Pazienza and Velardi (1993) proposed a straight line weighted sum between the original CU and another bias term. The similar idea can be utilized to force COBWEB to focus on capturing well-performed design solutions. For this thesis, we propose the following modified CU functions:

$$\begin{aligned} CU_k^{DCR} &= \nu\left[P(C_k)\left\{\sum_i \sum_j P\left(A_i = V_{ij} \mid C_k\right)^2 - \sum_i \sum_j P\left(A_i = V_{ij}\right)^2\right\}\right] \\ &\quad + (1-\nu)P(C_k)\left\{\frac{\overline{DCR_k}}{\max\left(DCR_k\right)} - \frac{\overline{DCR}}{\max\left(DCR\right)}\right\}, \nu \in [0,1] \end{aligned} \tag{7-10a}$$

$$\begin{aligned} CU_k^{SPCR} &= \nu\left[P(C_k)\left\{\sum_i \sum_j P\left(A_i = V_{ij} \mid C_k\right)^2 - \sum_i \sum_j P\left(A_i = V_{ij}\right)^2\right\}\right] \\ &\quad + (1-\nu)P(C_k)\left\{\frac{\overline{SPCR_k}}{\max\left(SPCR_k\right)} - \frac{\overline{SPCR}}{\max\left(SPCR\right)}\right\}, \nu \in [0,1] \end{aligned} \tag{7-10b}$$

$$\begin{aligned} CU_k^{SCR} &= \nu\left[P(C_k)\left\{\sum_i \sum_j P\left(A_i = V_{ij} \mid C_k\right)^2 - \sum_i \sum_j P\left(A_i = V_{ij}\right)^2\right\}\right] \\ &\quad + (1-\nu)P(C_k)\left\{\frac{\overline{SCR_k}}{\max\left(SCR_k\right)} - \frac{\overline{SCR}}{\max\left(SCR\right)}\right\}, \nu \in [0,1] \end{aligned} \tag{7-10c}$$

$$CU_k^{Mix} = \nu\left[P(C_k)\left\{\sum_i\sum_j P(A_i = V_{ij} \mid C_k)^2 - \sum_i\sum_j P(A_i = V_{ij})^2\right\}\right]$$
$$+\frac{(1-\nu)}{3}P(C_k)\left\{\frac{\overline{DCR_k}}{\max(DCR_k)} - \frac{\overline{DCR}}{\max(DCR)}\right\}$$
$$+\frac{(1-\nu)}{3}P(C_k)\left\{\frac{\overline{SPCR_k}}{\max(SPCR_k)} - \frac{\overline{SPCR}}{\max(SPCR)}\right\} \quad (7\text{-}10d)$$
$$+\frac{(1-\nu)}{3}P(C_k)\left\{\frac{\overline{SCR_k}}{\max(SCR_k)} - \frac{\overline{SCR}}{\max(SCR)}\right\}, \nu \in [0,1]$$

where the weighting term ν can be identified by cross validation process. $\overline{DCR_k}$, $\overline{SPCR_k}$ and $\overline{SCR_k}$ denote the average of DCR, SPCR and SCR measures within the k-th cluster.

Analogous to the original CU definition, the modifications introduced in equation 7-10 can be considered as an incremental in local average condition ratings (be they DCR, SPCR, SCR or average of the three) over overall average condition ratings. Therefore, another tradeoff that the modified version of COBWEB needs to perform is to balance between cluster size and this gain concurrently with the original predictability of a feature values in a cluster class from the original CU. The importance of trade off is expressed by the weighting term ν.

7.4. Conclusions

In this chapter, we have reviewed the two main machine learning algorithms being used for engineering design knowledge learning namely the conceptual clustering and ANN techniques. While, the latter is capable of adapt to highly complex association in the underlying data, its critical drawback lies in its opaqueness of the resulting knowledge. In other words, the model works like a black-box and thus hinders engineers/experts to benefit from understanding the knowledge structure. The former is more suitable for the thesis purpose from this perspective. With regards to complexity and adaptability of learning algorithm, COBWEB is more preferable because it learns in an incremental mode. Thus does not require retraining once a new example is presented to the COBWEB.

However, the original COBWEB can lead to arbitrary clustering that could results in spurious cluster structures that do not reflect the underlying groups of design configurations that lead to high performance. To circumvent this, we proposed a straight-line weighting scheme between the original CU and average performance measure. The latter term is added to introduce bias toward high performance designs. The weighting coefficient can be empirically identified during the process of cross validation.

8. EMPIRICAL EVALUATION OF LEARNING MODEL

EXPERIMENTS PERFORMED in this chapter are targeted at evaluating and measuring the degree in which COBWEB is capable of capturing high performance design examples for induction of design knowledge. First we apply the original COBWEB described in the previous chapter directly to NBI data. Results obtained are examined in detail. Then we move on to the modified COBWEB models. Comparison of the obtained knowledge between the original and the modified models is discussed.

8.1. Experiment Setups with Original COBWEB

Input into COBWEB for learning is a set of attribute-value design specification and configurations. Design specification, configurations and service environments are summarized in Table 8-1. For prediction (or testing) purpose, only the service environment and specification are input into the tree. The COBWEB tree is traversed from root given the incomplete set of information and the cluster with best matched environment and specification is located based on the following error measure at the k-th node.

$$\varepsilon_k = \frac{1}{N}\left(\sum_{i \in \mathcal{N}} I\left(x_i = y_i^k\right) + \sum_{i \in \mathcal{C}} \frac{\left|x_i - \bar{y}_i^k\right|}{\left(\max y_i^k - \min y_i^k\right)} \right) \tag{8-1}$$

where x_i and y_i^k denote the i-th attribute value of input vector and node value vector respectively. $\mathcal{N}$ and $\mathcal{C}$ are sets of nominal- (or ordinal-) and continuous-valued attribute respectively. The $I(\bullet)$ is an indicator operator whose value is 1 when argument expression is true and 0 otherwise. $\bar{y}_i^k$ denotes average value of the i-th attribute (whose attribute is a continuous one) at the k-th cluster. N is the dimension of input vector.

Equation (8-1) represents cluster representative value by using the followings:

- Value with maximal cardinality if the attribute is nominal, or
- Mean value if the attribute is continuous.

Then the error measure in equation (8-1) is simply a combination of error measure for nominal- and continuous-valued attributes normalized by the dimension size of the query vector. For continuous attribute, the difference between input vector value and the node mean is normalized by the attribute range local to cluster k.

Table 8-1. Design specification and configurations for training data

Service Environment		Design Specification		Design Configuration	
Attribute	**Acronym**	**Attribute**	**Acronym**	**Attribute**	**Acronym**
Average Daily Traffic	ADT	Lanes On Structure	LOS	Main Structure Material	MSM
Average Daily Truck Traffic	ADTT	Design Load	DL	Main Structure Design	MSD
State Code	SC	Type of Service On Bridge	TSOB	Number of Spans In Main Unit	NSMU
		Type of Service Under Bridge	TSUB	Length of Maximum Span	LMS
		Structure Length	SL	Deck Width	DW
				Deck Structure Type	DST

After the best matched cluster is located, the design description of the query vector is simply the design description of that particular cluster. The query example is not added to the tree in case of prediction.

The input query vector combined with the predicted design configuration is then input to the performance model identified in chapter 7 to determine the predicted performance either DCR, SPCR and SCR (see Figure 8-1).

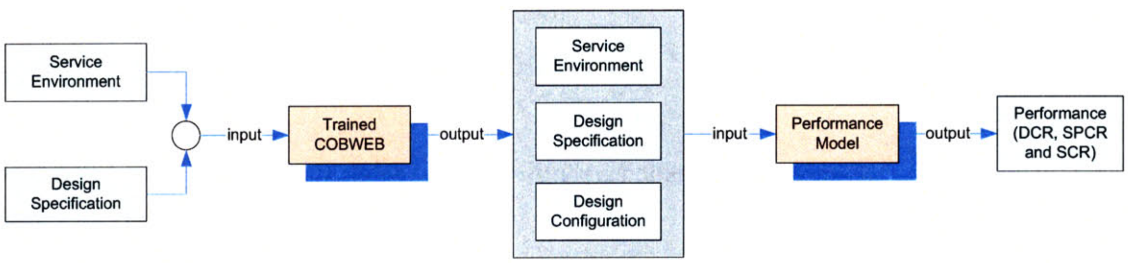


Figure 8-1. Flow of experiment

Finally, the performances and the design configurations in the clusters at the first few top levels of the COBWEB trees are examined to gain insight of the resulting knowledge in term of how design configurations attributes to performances.

The training process of COBWEB involves selection of number of training examples that optimizes validation set error (prescribed in equation 8-1). The error is calculated against the actual attribute value found in the validation set. The range used for error calculation of continuous attributes is computed from the range of the entire validation set on that attribute. Finally, the validation error over the entire validation set is normalized by its size. Because of limitation in computational resource, we constrain our experiment by using 15,000 examples as number for training set and 30,000 examples for validation set. These examples are I.I.D. and drawn from the entire set of NBI database. Finally, trainings are performed on sample set of sizes {100, 200, 300, 500, 1000, 2000, 3000, 4000, 5000, 10000, 15000} to determine the optimal COBWEB tree.

8.1.1. Identification of Optimal COBWEB

Figure 8-2 shows average prediction error over the validation set as defined in equation (8-1).

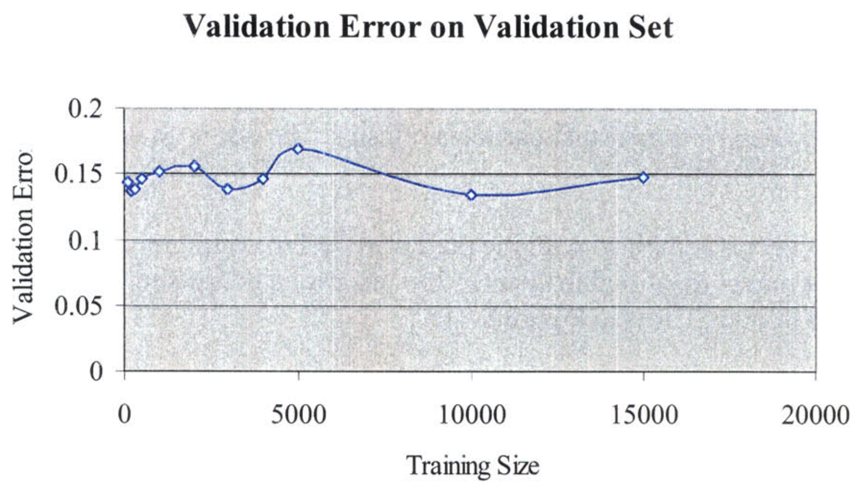


Figure 8-2. Validation error vs. training sample size

While validation error at 1,000 and 10,000 are more or less the same, the COBWEB with 1,000 has a potential to lack observations with some design attribute groups and therefore should not be selected. Therefore, we select 10,000 point as our optimal training size and define the optimal COBWEB as that trained with 1,000 observations.

An observation that confirms Reich (1992) and Fisher (1987) on the behavior of COBWEB is that the normalized tree size (i.e. number of clusters over the number of training samples) is kept consistently low for COBWEB. In our case, we observe that COBWEB demonstrates average normalized tree size of 1.35 (although the overall trend is still sloping downward). This is shown in Figure 8-3. This behavior is desirable because the relative complexity of the learned hypothesis to the problem size does not grow with the problem size.

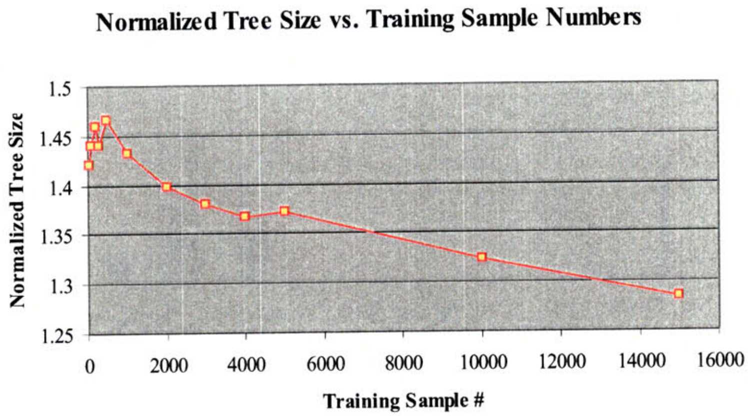
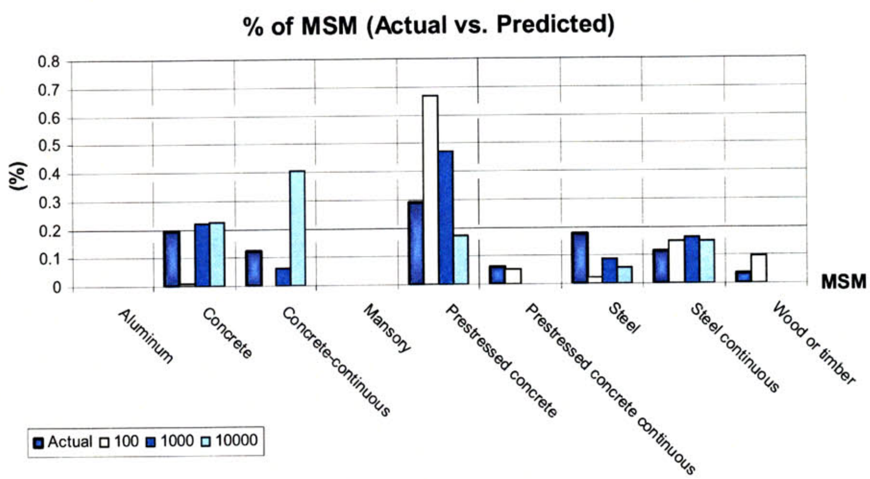


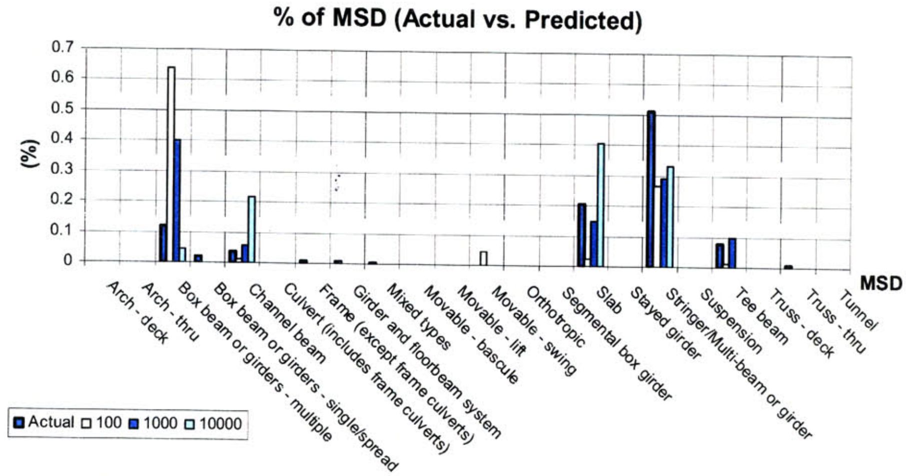
Figure 8-3. Normalized tree size vs. number of training samples

Next, let us examine effects of training sample size on predicted values. Figure 8-4 shows histograms of actual value of each design configurations (shown in Table 8-1) versus the predicted values from COBWEB at different training sizes. Densities are plotted instead of histogram in case of continuous-valued attributes (i.e. NSMU, LMS and DW)

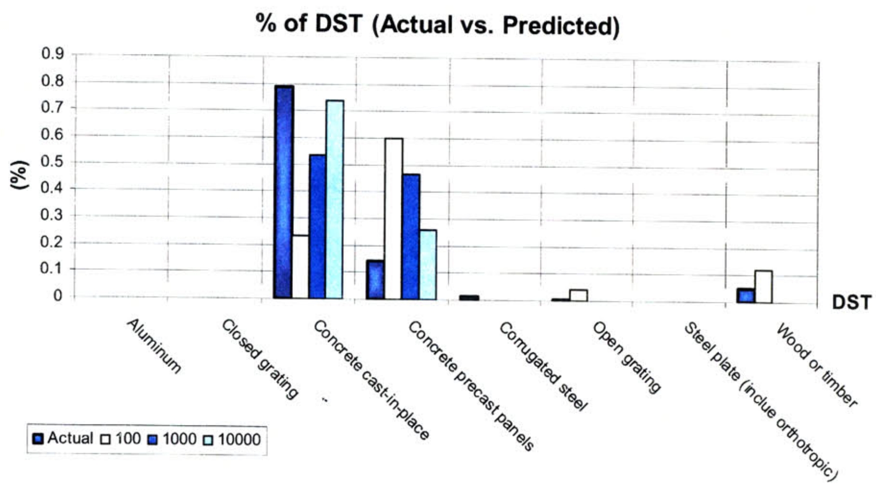


(a) Main Structure Material (MSM)

Figure 8-4. Histograms of actual and predicted design configurations by training size

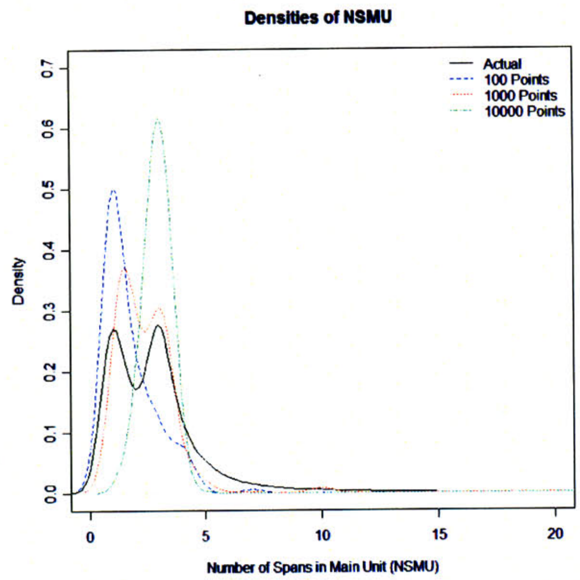


(b) Main Structure Design (MSD)

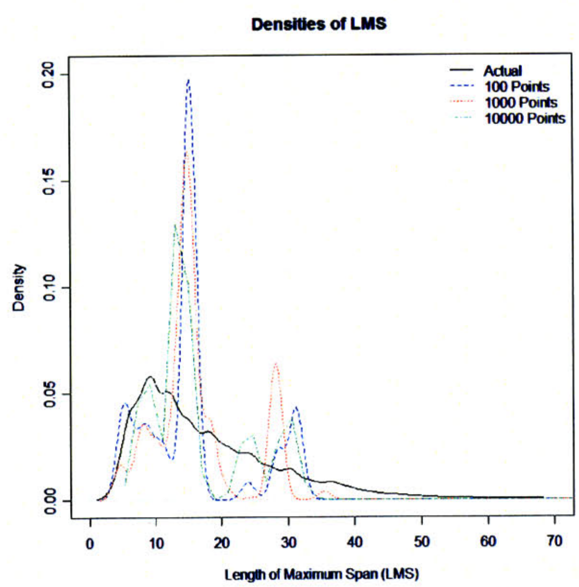


(c) Deck Structure Type (DST)

Figure 8-4. Histograms of actual and predicted design configurations by training size (continued)

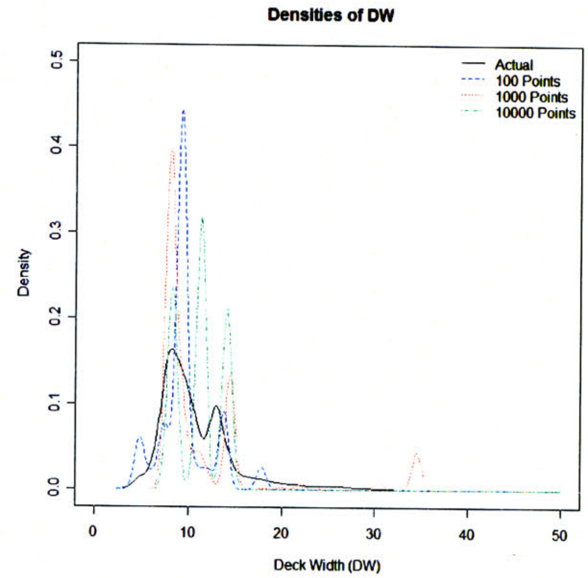


(d) Number of Spans in Main Unit (NSMU)



(e) Length of Maximum Span (LMS)

Figure 8-5. Histograms of actual and predicted design configurations by training size (continued)



(f) Deck Width (DW)

Figure 8-5. Histograms of actual and predicted design configurations by training size (continued)

From the comparison of the distributions between the predicted and actual attribute values, it can be seen that the 10,000 point sample COBWEB provides a reasonable matches for all attributes compared to others.

As a result, we deduce that the optimal COBWEB tree should be formed by training size of 10,000 points. For the remaining of this subsection, we will use COBWEB trained from 10,000 point samples for our discussion.

8.1.2. Analyzing COBWEB Knowledge Tree

Extraction of the first 3 levels of the COBWEB tree using 10,000 points for training is illustrated in Figure 8-5. In each node (or cluster), the top three probabilities are listed. The children automatically subsume these probabilities in most cases.

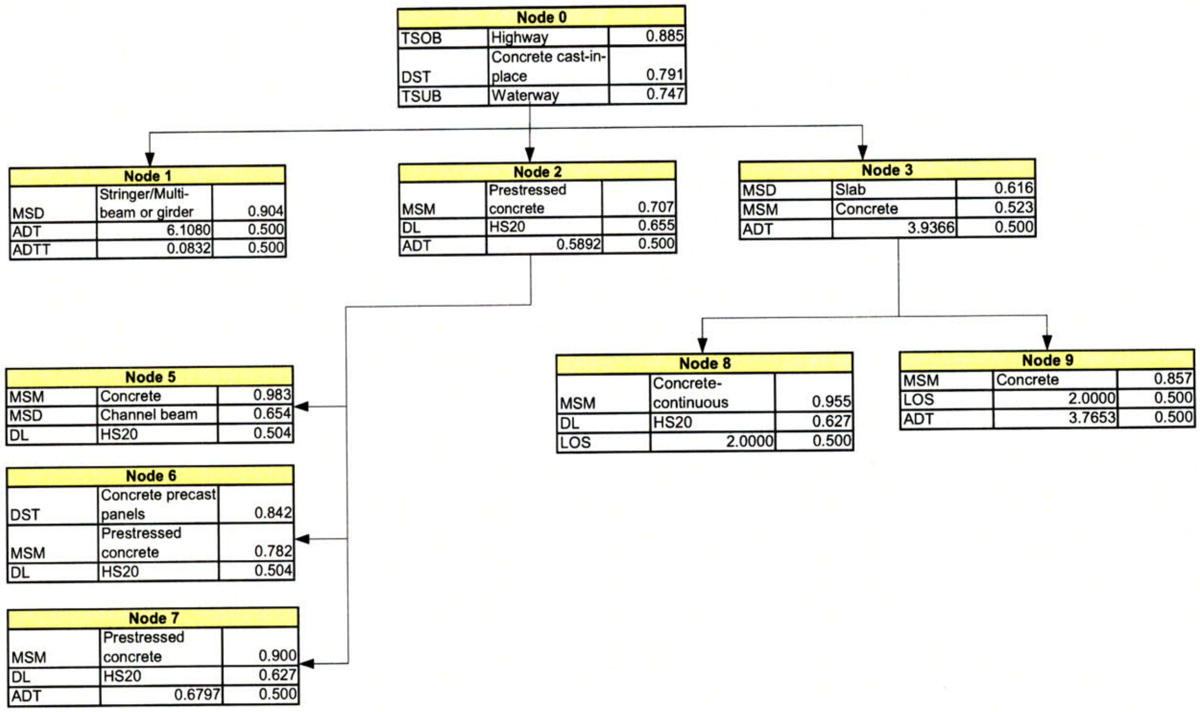


Figure 8-6. Optimal COBWEB tree obtained in section 8.1.1

At the root where all training instances are added, one can see that the majority are highway bridges with concrete-cast-in-place type of deck structure. The service underneath highway is mostly waterway.

In the next level, main structure design or material starts to divide samples into different clusters. In this case, we have three. The first one (node 1) has stringer/multi-beam structure design whereas node 3 has its majority instances of type slab. Node 2's majority consists of bridges of prestressed concrete material, and so on.

It can also be seen that variety of design materials could be observed in the deeper nodes. For example, node 5 to 7 contains different type of concrete material although they subsume their parent, node 2 whose material is governed by prestressed concrete.

Unfortunately, it is hard to obtain a clear trend of designs that lead to high performances from clusters formed by original COBWEB. As one can see from example in Figure 8-5, the clusters are formed in such a way that it gather statistically similar trend into big nodes to avoid high complexity of the resulting tree function. Therefore, the clusters could be formed quite arbitrarily with respect to performances.

8.1.3. Out-of-sample Performance of Simple COBWEB

The out-of-sample data is prepared by mutual exclusively, to the previous training and validation set, I.I.D. sampling data from the NBI database. The sampled out-of-sample data is of size 30,000 points.

Using the trained COBWEB with 10,000 points data, we obtain the prediction results for out-of-sample data set on DCR, SPCR, SCR and average score. The results are summarized in Table 8-2. From the tabulated results, it could be seen that COBWEB tends to overestimate the physical condition ratings. The main trend is that the COBWEB predicts the ratings for out-of-sample data to only rating 7 and 8 for all categories. This is not surprising because the training has been done without inclusion of these performance measures. The COBWEB is free to cluster around design specification and configurations that yield the best trade-off between inter- and intra-cluster accuracy as defined by CU. In addition, the error from performance model could worsen this performance as mentioned in Chapter 7.

Another view of prediction accuracy is the Residual Squared Error (RSE) which is simply the sum of squared residuals on all observations normalized by sample size. RSE for all performance indices are calculated and tabulated in Table 8-3. Averagely, we found RSE of approximately 2 using this simple COBWEB.

Table 8-2. Out-of-sample prediction of performance measures by COBWEB

(a) DCR

DCR		Predicted						
		4	5	6	7	8	9	Total
Actual	4	0.00%	0.00%	0.00%	0.46%	2.26%	0.00%	2.72%
	5	0.00%	0.00%	0.00%	1.55%	7.39%	0.00%	8.94%
	6	0.00%	0.00%	0.00%	3.39%	17.41%	0.00%	20.80%
	7	0.00%	0.00%	0.00%	6.70%	33.06%	0.00%	39.75%
	8	0.00%	0.00%	0.00%	3.57%	18.55%	0.00%	22.12%
	9	0.00%	0.00%	0.00%	0.90%	4.76%	0.00%	5.66%
Total		0.00%	0.00%	0.00%	16.56%	83.44%	0.00%	100.00%

(b) SPCR

SPCR		Predicted						
		4	5	6	7	8	9	Total
Actual	4	0.00%	0.00%	0.00%	0.28%	1.57%	0.00%	1.85%
	5	0.00%	0.00%	0.00%	1.34%	5.91%	0.00%	7.25%
	6	0.00%	0.00%	0.00%	2.91%	14.66%	0.00%	17.57%
	7	0.00%	0.00%	0.00%	5.94%	30.42%	0.00%	36.36%
	8	0.00%	0.00%	0.00%	5.00%	25.29%	0.00%	30.29%
	9	0.00%	0.00%	0.00%	1.09%	5.58%	0.00%	6.67%
Total		0.00%	0.00%	0.00%	16.56%	83.44%	0.00%	100.00%

Table 8-2. Out-of-sample prediction of performance measures by COBWEB (continued)

(c) SCR

SCR								
		Predicted						
		4	5	6	7	8	9	Total
Actual	4	0.00%	0.00%	0.00%	0.49%	2.50%	0.00%	2.99%
	5	0.00%	0.00%	0.00%	1.42%	6.86%	0.00%	8.28%
	6	0.00%	0.00%	0.00%	3.22%	15.71%	0.00%	18.94%
	7	0.00%	0.00%	0.00%	6.57%	32.56%	0.00%	39.13%
	8	0.00%	0.00%	0.00%	3.80%	20.39%	0.00%	24.19%
	9	0.00%	0.00%	0.00%	1.06%	5.41%	0.00%	6.48%
Total		0.00%	0.00%	0.00%	16.56%	83.44%	0.00%	100.00%

(d) Average

Average								
		Predicted						
		4	5	6	7	8	9	Total
Actual	4	0.00%	0.00%	0.00%	0.15%	0.67%	0.00%	0.82%
	5	0.00%	0.00%	0.00%	1.20%	6.00%	0.00%	7.20%
	6	0.00%	0.00%	0.00%	3.85%	18.61%	0.00%	22.46%
	7	0.00%	0.00%	0.00%	6.64%	33.59%	0.00%	40.22%
	8	0.00%	0.00%	0.00%	3.76%	19.65%	0.00%	23.41%
	9	0.00%	0.00%	0.00%	0.97%	4.92%	0.00%	5.89%
Total		0.00%	0.00%	0.00%	16.56%	83.44%	0.00%	100.00%

Table 8-3. Out-of-sample prediction RSE performance measures by COBWEB

	RSE
DCR	2.311
SPCR	1.931
SCR	2.239
Average	1.953

8.2. Experiment Setups with Modified COBWEB

While inputs to the modified COBWEB remains the same as tabulated in Table 8-1, the search strategy has been changed to allow bias on high performance examples as described in section 7.3. The performance model obtained in chapter 7 is not only used for evaluation purpose, but also in the branching decision. As a proof of concept, we only illustrate the last modification, i.e. the one in equation 7-10d which represents the case of average performance. Similar to experiments in section 8.1, the same set of numbers of training examples is used to identify optimal COBWEB on the validation sample set of size 30,000.

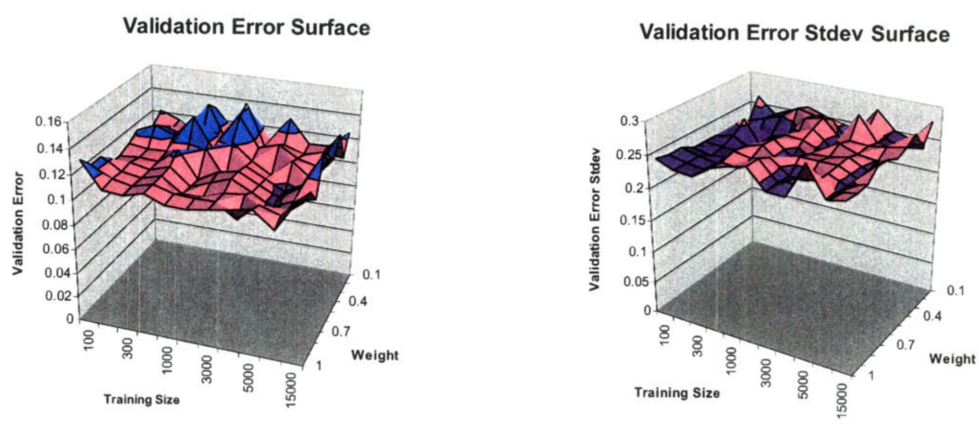
8.2.1. Identification of Optimal Modified COBWEB

Unlike the case of simple COBWEB, the cross-validation stage for modified COBWEB also needs to identify optimal value of straight line weight ν. Thus, we empirically vary the value of ν from 0.1 to 1 using 0.1 interval and find the optimal ν and training size over the validation set. The criteria for optimality cannot employ the validation error directly because it does not include any information about the gain in term of performance from the bias that we introduced into equation 7-10. For this purpose, we propose that the criterion for validation is:

$$\kappa = \frac{\overline{PERF}}{\sigma_{PERF}} \frac{1}{\overline{\varepsilon}} \tag{8-2}$$

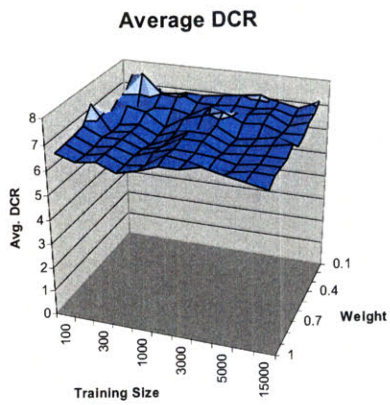
where $PERF$ is defined as the predicted performance index, i.e. DCR, SPCR, SCR or average of the three. σ_{PERF} is the standard deviation of the predicted performance index $PERF$. $\overline{\varepsilon}$ is the mean of validation error. The intuition for this objective measure is to direct the search toward the direction that simultaneously maximizes the average performance and minimizes the standard deviation of the predicted performances and average error of the prediction.

As a result, the empirical search for optimal COBWEB is performed on the 2-dimensional space formed by sample size vector and vector of weighting coefficient ν. Below are results of the validation process.

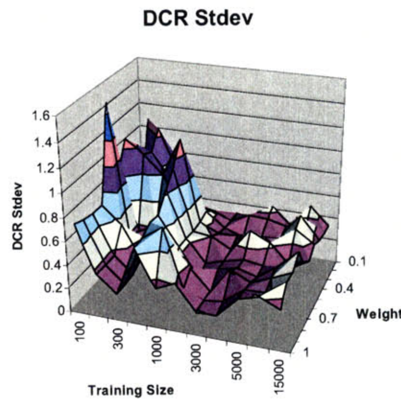


(a) Mean (b) Standard Deviation

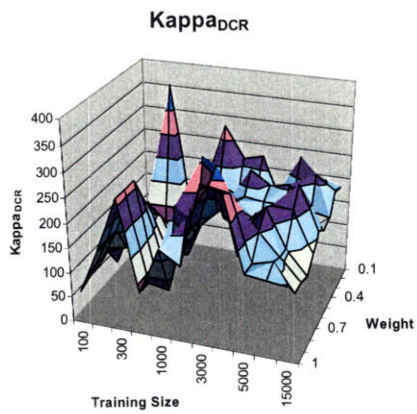
Figure 8-7. Mean and standard deviation surface of validation error



(a) Mean

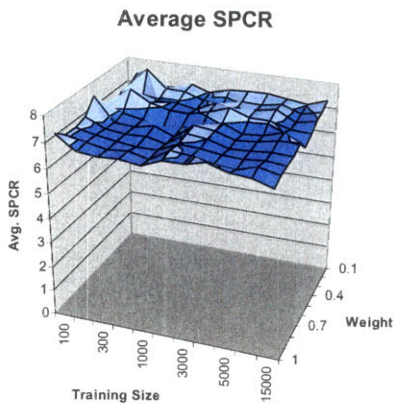


(b) Stadard deviation

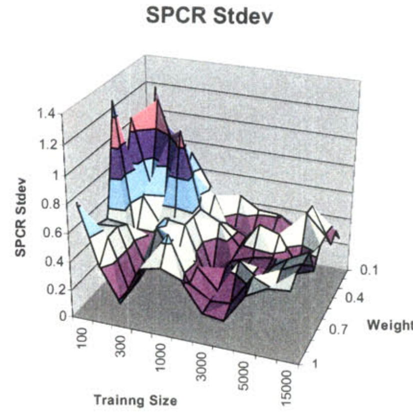


(c) κ_{DCR}

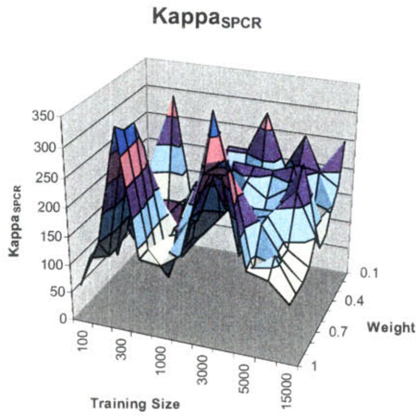
Figure 8-8. Average, standard deviation and κ_{DCR} on validation set



(a) Mean

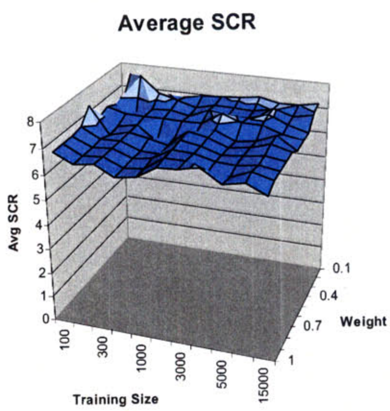


(b) Stadard deviation

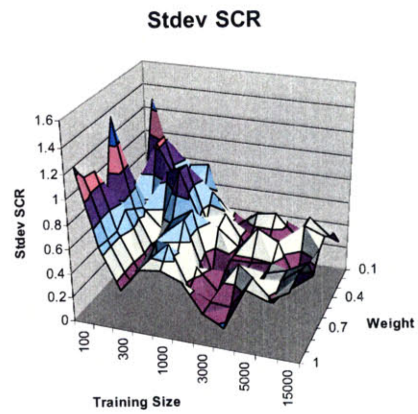


(c) κ_{SPCR}

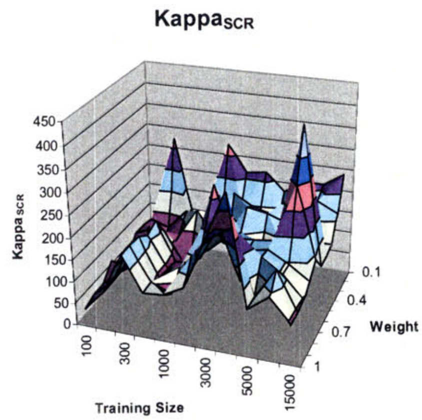
Figure 8-9. Average, standard deviation and κ_{SPCR} on validation set



(a) Mean

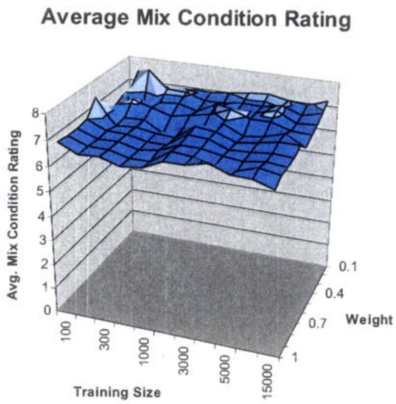


(b) Stadard deviation

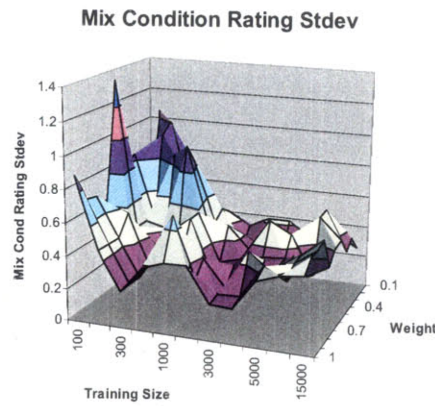


(c) κ_{SCR}

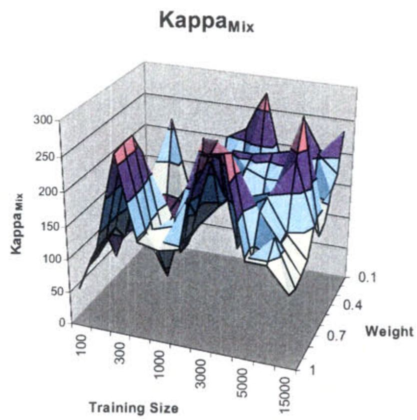
Figure 8-10. Average, standard deviation and κ_{SPCR} on validation set



(a) Mean



(b) Stadard deviation



(c) κ_{Mix}

Figure 8-11. Average, standard deviaion and κ_{Mix} on validation set

As one could see from Figures 8-6 to 8-10, the trend is that with low size of training set, we observed high variation of the predicted condition ratings. On the other hand, for validation error, the trend is opposite. That is, one can observe that the higher the weights the cluster-local condition ratings gain (relative to that of validation error), one see poorer quality of fitting. Thus, we shall seek a good combination of training size and weight that yield maximal value of κ. For the case of average performance model modification (equation 7-10d), we observe a peak of κ_{Mix} at training size of 10,000 and ν of 0.5. Therefore, we will use this configuration for our discussion from this point onward.

8.2.2. Analyzing Knowledge from Modified COBWEB Tree

The modified COBWEB for average condition ratings is drawn until node at level 3 in Figure 8-11.

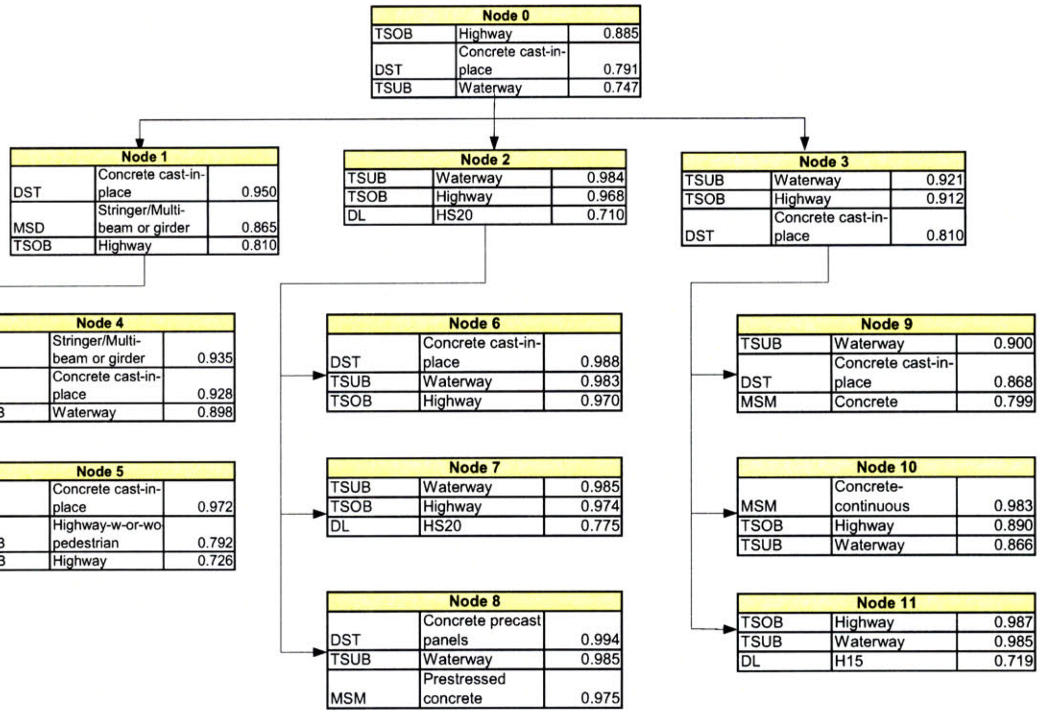


Figure 8-12. Optimal modified COBWEB tree (by 7-10d) obtained in section 8.2.1

Similar root node as in Section 8.1 is found here because the underlying validation is the same. However, a clear distinction from the original COBWEB tree is that, the modified COBWB tree branches mostly from design specifications, i.e. TSUB, TSOB and DL and cascade the design configuration detail down along the path of the tree. In contrast, the original COBWEB tree (Figure 8-5) starts its hierarchy from basic design configurations on the top and the detail of design specification will become more specific along the path of the tree.

Thus constraining the branching strategy to explicitly focus on performance results in tree clusters that is built from specifications, and then is propagated to specific design details. In other words, the modified COBWEB provides a hierarchy of knowledge that one can start from locating concepts of specification and then locate specific design detail along the deeper node of the path.

8.2.3. Comparison of Original and Modified COBWEB

To better understand characteristics of the modified COBWEB, we compare distribution of each attribute value predicted by the original COBWEB and the modified one. Comparisons of attribute value distribution are plotted in Figure 8-12.

First, in term of main structure material (Figure 8-12a), it is revealed from the plot that unlike the original COBWEB, the predicted values from the modified COBWEB are limited among concrete, steel and pre-stressed concrete. Especially, it is almost as is the missing populations in other type of materials apart from these three were allocated to pre-stressed concrete.

Second, for main structure design (Figure 8-12b), the modified COBWEB constrains the prediction between slab and stringer/multiple beam type. Some occurrences of box type structure are also predicted. However, in contrast to the original COBWEB, approximately 20% of the entire population predicted as channel beam type does not appear in prediction of the modified model.

Third, for deck structure type (Figure 8-12c), the results from the two models are quite consistent. The majority is predicted as concrete cast-in-place whereas some (about 5-20%) are assigned as concrete precast panel.

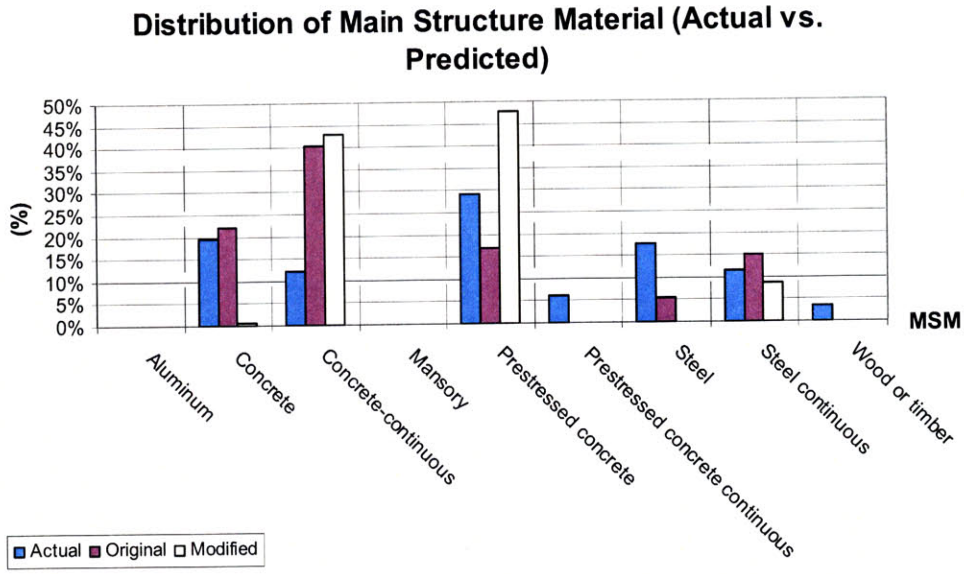
Forth, for numbers of span in main unit (Figure 8-12d), comparison of the predicted and actual density suggests that the extreme value range for both low and high sides are not captured by both models. Both models produced very closely similar results.

However, for length of maximum span attribute Figure (8-12e), it is clear that both models tend to pick up values in certain ranges that are separated from each other. Examples of ranges are 5-18, 20-25, 25-30. Comparing the original to the modified model, the original model produces wider range of distribution which covers most of the range that the actual observation does whereas the modified model produces prediction in the range of 10-30.

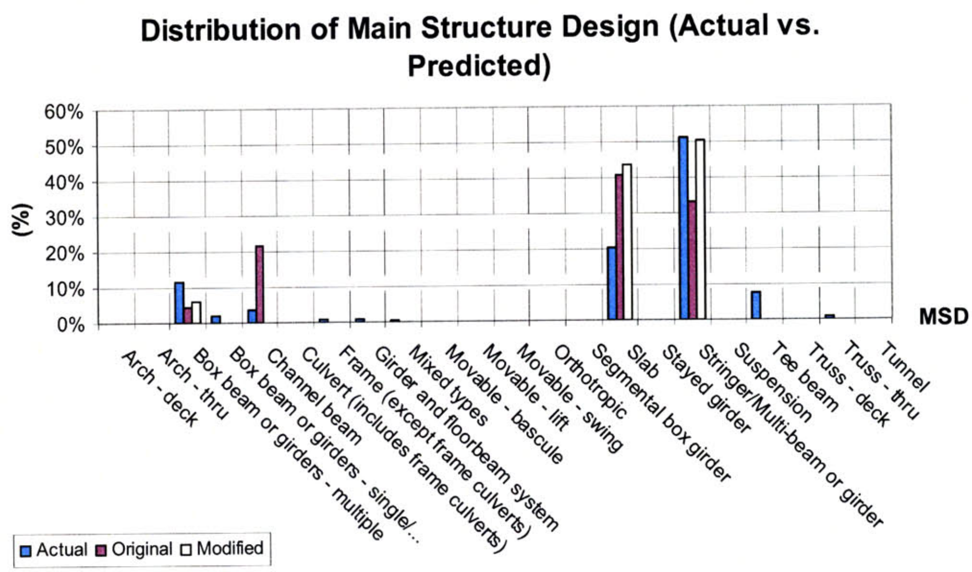
For deck width (Figure 8-12f), the two models predictions are quite close to the actual observation especially the original one. The modified model generates narrower distribution.

On the condition rating side (Figure 8-12g-8-12j), a typical trend can be observed that both models limits their prediction to range of 6-8. For the modified model, the distribution tends to skew toward higher rate, i.e. 8 more than does the original one. The important result is in Figure 8-12j which compares distribution of prediction on average ratings from both model and references to the actual observation. The density of the modified model clearly skews toward rating 7 and 8 whereas the original model produces quite symmetric density around rating 7.

Therefore, it is clear that the modified model leads to clustering algorithm which results in a more limited focus on high rating side. Thus satisfies our thesis hypothesis.

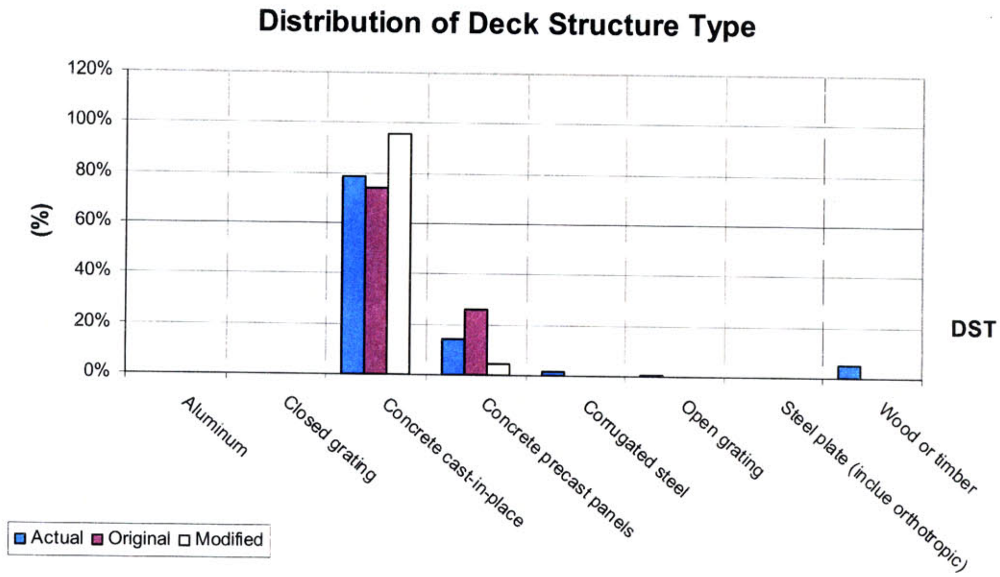


(a) Main structure material

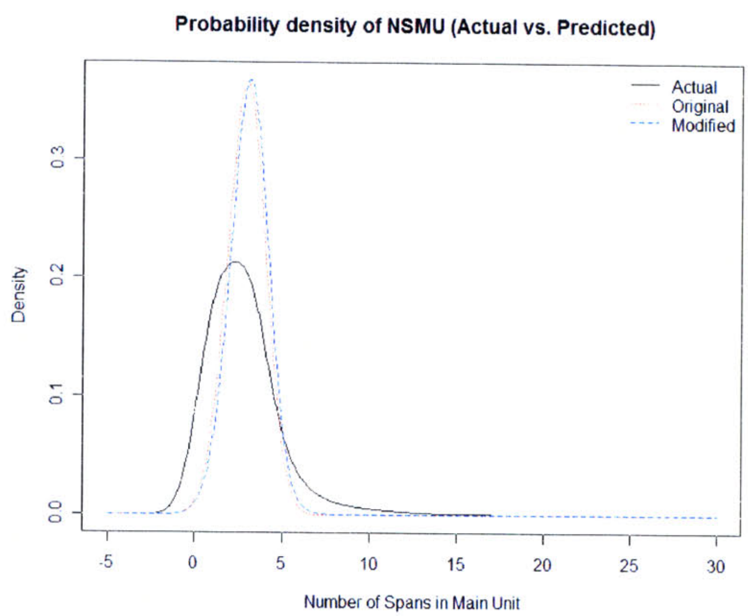


(b) Main structure design

Figure 8-13. Comparison of fraction of population by attribute value (actual, original COBWEB, and modified COBWEB)

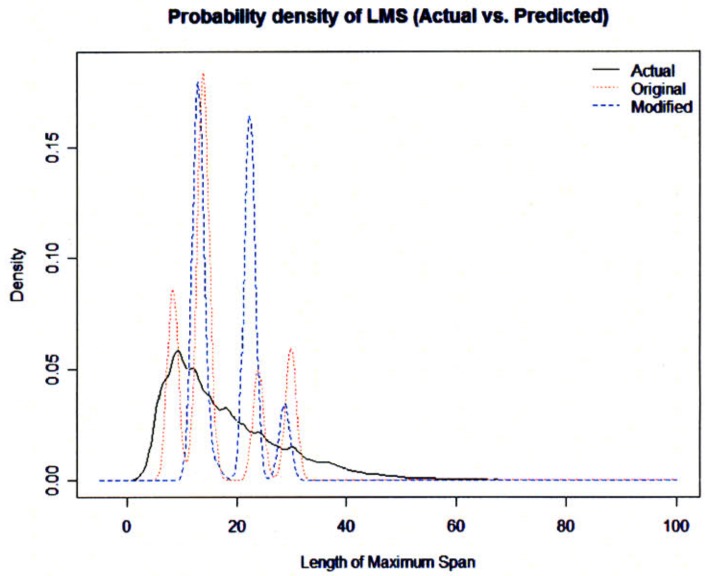


(c) Deck structure type

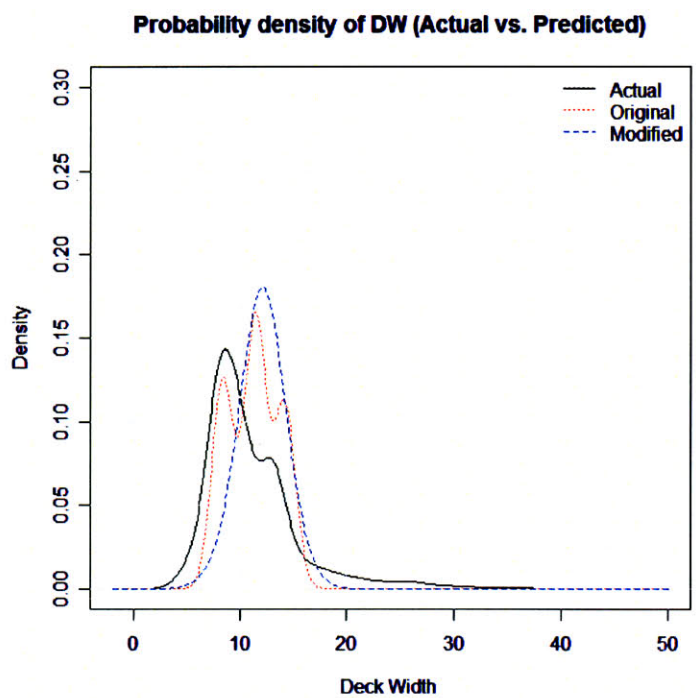


(d) Number of spans in main unit

Figure 8 13. Comparison of fraction of population by attribute value (actual, original COBWEB, and modified COBWEB) (continued)

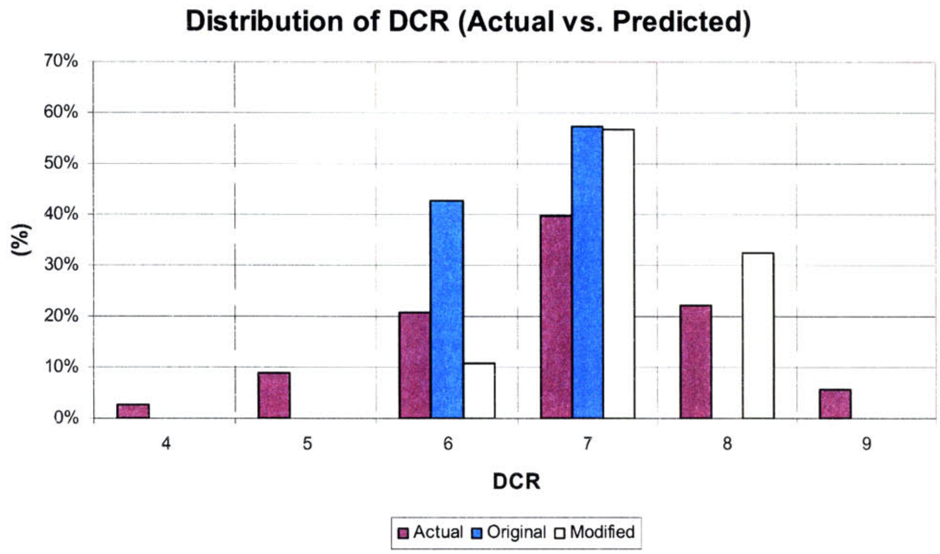


(e) Length of maximum span

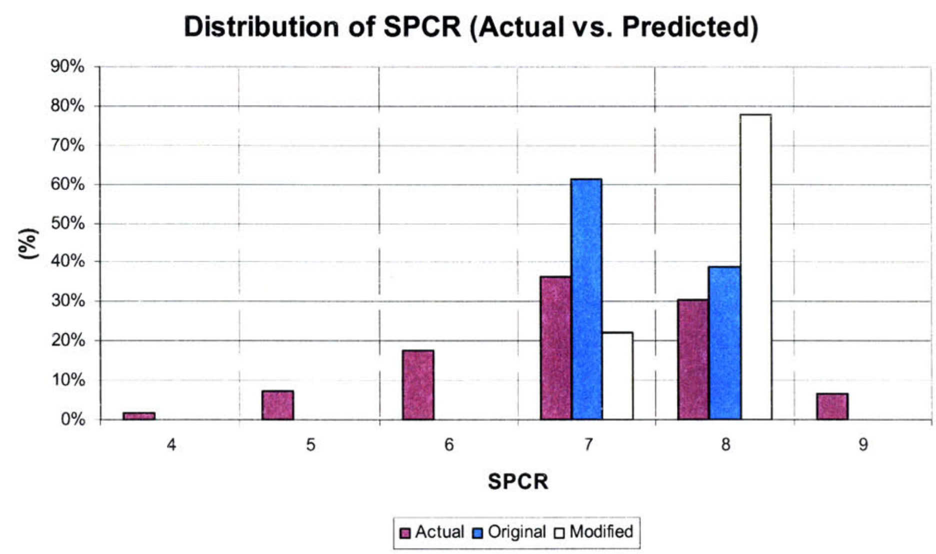


(f) Deck width

Figure 8 13. Comparison of fraction of population by attribute value (actual, original COBWEB, and modified COBWEB) (continued)

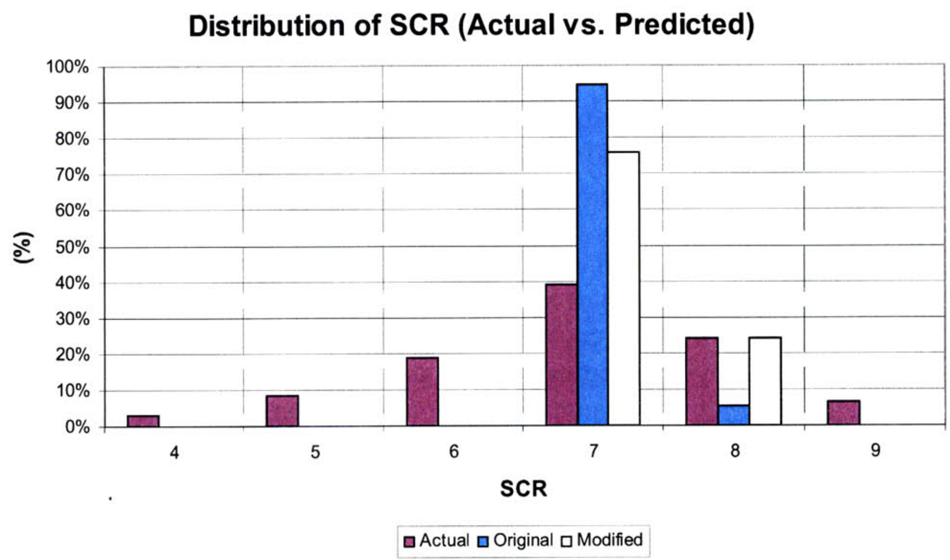


(g) Deck condition rating

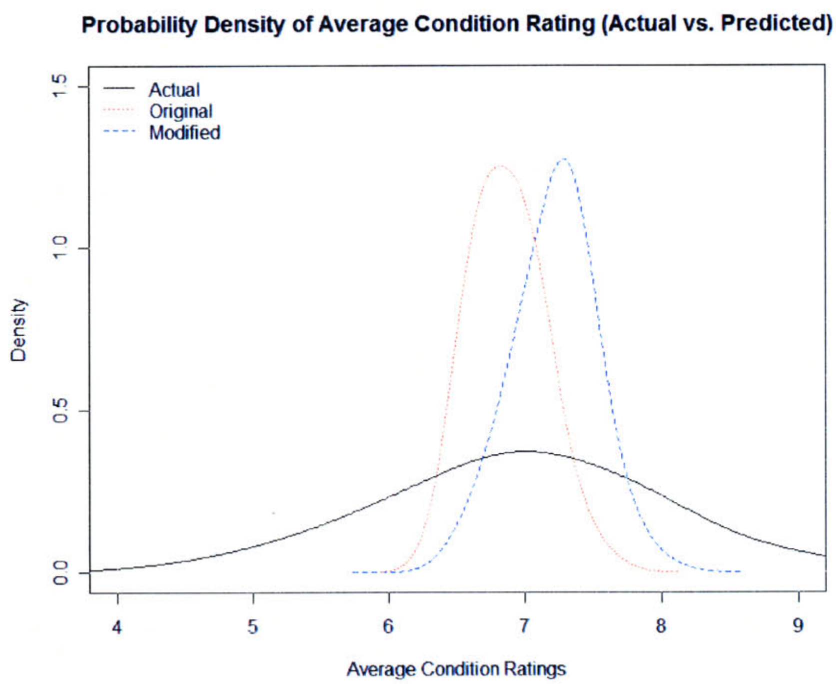


(h) Superstructure condition rating

Figure 8 13. Comparison of fraction of population by attribute value (actual, original COBWEB, and modified COBWEB) (continued)



(i) Substructure condition rating



(j) Average condition rating

Figure 8 13. Comparison of fraction of population by attribute value (actual, original COBWEB, and modified COBWEB) (continued)

8.3. Conclusions

In this chapter, we have presented empirical experiments that aim to illustrate characteristics of the proposed modified COBWEB model which explicitly direct the branching toward high performance designs. The optimal simple COBWEB (or original model), in the sense of training size, is empirically identified by searching for model that yields minimal validation error on the validation test. The optimal model is the one built from 10,000 points of training set. The knowledge obtained from examination of the resulting hierarchy reveals that design materials or design structure types of type concrete and stringer/multi-beam or girder represents the main trends in design. The original COBWEB clusters observation instances by their contribution to the local node predictability on its attribute value rather than performance. Therefore, the predicted design configurations cause high error of condition ratings of approximately 2 in term of RSE.

In contrast, the modified model exhibits a clear tendency to form such a hierarchy that focuses around observations with high performances. This can be observed from the predicted design configurations that tend to centre around modern materials. As a result, the predicted condition ratings have low variability and limits themselves to only rating 7 or 8. This is a clear evidence to support the success of our model development using this type of branching strategy.

9. DISCUSSIONS

BRIDGE DESIGN IS TYPICALLY conducted to meet budget and short-term performance measures. Rarely has it been focused on long-term performance due to the high complexity and uncertainty evolved during the long life-span of the bridge service life. As a result, the NBI data reveals that more than 150,000 bridges are structurally deficient or functional obsolete. This number is likely to increase due to increased traffic volume, aging and the continuing deterioration process. Additionally there are limited funds for rehabilitation and maintenance.

National Cooperative Highway Research Program (NCHRP) and other organizations such as FHWA and AASHTO have started initiatives to address the issue of incorporation into the design phase. NCHRP (2006) drafted detailed suggestions to accommodate long-term performance into design as a component for asset management and budget allocations. FHWA in cooperation with AASHTO (2004) initiated work to provide strategic roadmaps toward performance specification which is meant for long-dated performance.

Both attempts involve defining meaningful long-term performance measures and a higher accuracy modeling approach that yields better prediction of future performance of the design artifact. In contrast to the current scheme in which design is set to meet initial cost and short-term performance, the long-term performance measure is used to estimate cost over the service life cycle and is incorporated into the design phase. The objective of design becomes more emphasized on estimating future conditions with respect to the level of funding.

The following sections will be devoted to discussion on this new design aspect especially from the NBI data and discoveries of this thesis work perspective. Firstly, the new concept of design for long-term performance is introduced. The elements necessary to enable such a scheme are described. Retrospectively, the current NBI bridge data is used to reflect the inadequacy of the current data and plausible development to achievement of design for performance is suggested.

9.1. Overview of Design for Long-term Performance

With the limited budget for maintenance and deterioration in the bridges nation-wide, several initiatives have been set forth to develop a new design framework that allows engineers or designers to prescribe target long-term performance and use such information to tradeoff between future conditions and investment budget especially from the asset management perspective.

FHWA in cooperation with AASHTO (2004) initiated works around performance specifications. The goal is to establish a framework for prescription of design performance during order process. The main objective is to ensure that the targeted performance is satisfied over the planned life cycle. The emphasis is to determine meaningful measures of performance that can be either estimated or tested using key construction test or modeling approach, or measured after some predetermined time. The former is known as 'performance-related measure' (PRS) and the latter as 'warranty'. Among all fundamental requirements for PRS, the key items are measurability of quality characteristics – which partly implies accuracy of the measurement itself - and predictability of performance.

On the other hand, National Cooperative Highway Research Program (NCHRP) started an initiative to establish a framework for life-cycle performance measures as a target for asset management. The main idea is to identify meaningful long-term performance measures and to provide guidelines on how these should be incorporated into asset management plan. Especially, for the design stage, how the long-term performance shall be reflected as investment cost and how it could affect life-cycle budgeting at an individual and network level are thoroughly studied. Key components are more or less similar to what PRS embraces and contain sound analytic constructions of performance measures and solid quality foundation data. Without being able to accurately estimate future performance, the estimation quality of budgeting for asset management level is degraded.

Recently, FHWA has formed a synergy with bridge owners, bridge industry and academia to form a so-called 'Long-Term Bridge Performance (LTBP) Program' with similar objectives to the two initiatives mentioned earlier (FHWA, 2007). The program focus is on data quality and collection, data mining and analysis, etc. In addition, it clearly specifies periodical inspection of bridges using Non-Destructive Evaluation (NDE) method with visual inspection to document deterioration.

9.2. Solid Foundation Data

In the report 551 from NCHRP, the criticality of solid foundation data to attribution of successful asset management for performance has been emphasized. The interviews in the study reveal emphasis on data quality and evaluation of data usefulness to support performance measures meaningfully, accurately and reliably. Below are sample quotes from the report that reelects the importance of data on performance evaluation.

- There is a direct relationship between specific performance measures and the data needed to support these measures. "The most common data problems are in ascertaining the quality of the data and in acquiring it in the exact form desired".
- Data that are highly uncertain translate into performance measure values that are likewise highly uncertain, reducing their management value. "Investments in accurate, high-quality data collection systems are essential to successful

performance measurement and, by extension, to achieving the overall strategic goals of the agency".

- Some factors that are important, however, either cannot be measured at all or cannot be measured accurately at an acceptable cost. "Transportation agencies need to consider the uncertainty introduced by inaccurate data when taking action based on their system of performance measures".

From this thesis study, the NBI data in its current stage fails to meet the above desired characteristics. Firstly, the accuracy of measured performance is poor and thus rules out the accountability of the resulting models to predict or analyze performance. For example, over 68% of the measurement that varies between 1 and -1 unit for physical condition ratings reported by Washer's study (2003) on comparison of routine visual inspection and NDE approach.

Secondly, the NBI data collects data based on availability rather than the necessity driven by business process of asset management. The NBI data contains more information on results and manifestation of current deterioration or serviceability rather than detail of factors that explains deterioration process. A clear example is impressive prediction accuracy of physical condition ratings reported by Chase, Small and Nutakor (2000) with geographic environmental explanatory variables. They achieved approximately 15-20% in residual squared error (RSE) using ordered response model type of regression. This is a significant improvement over this thesis achievement using data available in NBI data alone. From an interview, Chase has emphasized that these factors indeed significantly enhance explanatory power to condition rating. This supportively suggests that the criticality of augmenting NBI data with necessary information for better describing deterioration or future performance. This information provides more reliable foundation for evaluation of design solutions and more accountable decision for investment allocation at the bridge network level.

Thirdly, to address the cost of data collection and quality issue, the new technique such as NDE should be used. Not only that the NDE can provide a more economical means for data collection and inspection, it can also yield higher accuracy and less subjective data. In the next section, we provide an example of such NDE measurement.

9.3. Accountable Performance Measures

One way to distinguish between overall health and critical deficiencies is through the type of performance measures used and the establishment of critical threshold values.
For individual facilities, overall health can be gauged through indexes based on a set of conditions (e.g., sufficient score in NBI). Critical deficiencies can be identified by establishing a threshold for these indexes, the value of which experience shows is serious enough to threaten the structural integrity of the facility, dramatically increase user costs, or result in many customer complaints.

Another approach is to focus on particular conditions that are critical to facility performance and to define detailed measures and thresholds (e.g., physical condition ratings).

Nonetheless, the performance measures should be defined in the most objective or quantitative way as much as possible to avoid subjectivity or bias during evaluation or measurement. A good example is found from Washer (2003) study that NBI bridge condition rating (scaling from 0 to 9) has averagely +/-1 range of error. Also, among about 50 inspectors, 4-5 different rating score can be assigned to the same bridge element which well reflects subjectivity during the routine visual inspection process.

In contrast, a new methodology such as Non-Destructive Evaluation (NDE) technique can be used to quantitatively and accurately evaluate performance or condition of the bridges. For example, Figure 1 compares vibration mode measured from ambient vibrometer of a healthy and structurally deficient bridge. A clear difference can be observed in that the healthy one vibrates in its natural modes and have its spectrum concentrated only on a few frequencies whereas the damaged bridges possess vibration in almost every frequency, i.e. resonance. With this particular example, this type of spectrum can be taken on the newly constructed bridge as a baseline performance and each year vibration spectrum can be measured and compared with the baseline. The approximate location and severity of an unknown defect is determined using a damage correlation index calculated from the frequency response function (FRF) of the structure. The damage index is a relative measure whose value depends on the differences in the dynamical properties of the undamaged (baseline) and damaged structures (Mal, A., et. al., 2005). This also emphasizes superiority of NDE technique over the conventional visual inspection in that the NDE can identify damage which is hidden inside the structure where visual inspection cannot be reached or assessed. Thus NDE can provide a more quantitative and more accountable performance measure for use of design for long-term performance.

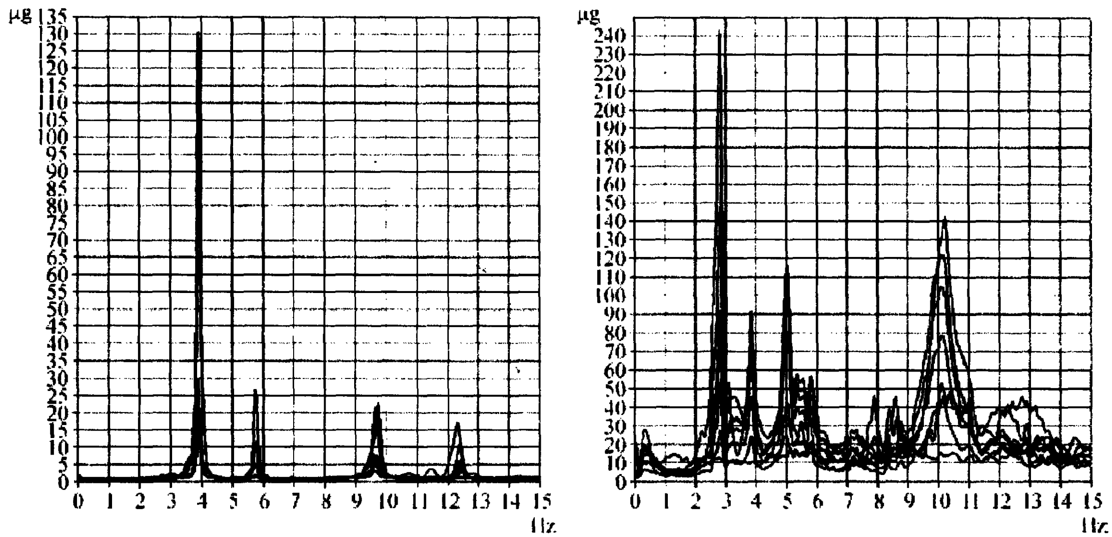


Figure 9-1 Spectrum of a sound bridge (left) and spectrum of a damaged bridge (right) (source: SAMCO, 2007)

9.4. Suggestions for Improvement of NBI Data

In this chapter, we have used the design for long-term performance scheme to emphasize shortfalls of the NBI data. Firstly, it does not serve as a complete set of data containing information that supports understanding or explains the observed bridge performances. Second, the data quality in the NBI data is low. In particular, the measurement of the observed performance is subjective and erroneous. Third, the performance of the bridge is defined on such a subjective and erroneous measure and thus causes the defined performance to lack of accountability. As a result, it deters accountability of the resulting model when analyzing such a performance measure.

In this subsection, we suggest some of the enhancement to the existing NBI bridge data in the context of design for long-term performance.

- Provide supportive data that explains the observed performance: an example from this study is geographic information which Chase, Small and Nutakor (2000) have incorporated. The database should be self-contained in such a way that it provides sufficient and meaningful data that is necessary to describe or support the observed performance.
- Reliable data accuracy: the NBI data especially the performance measures tend to include error. This is due to a subjective and inaccurate data collection process. To avoid this problem, new techniques such as NDE can be introduced to yield reliable level of accuracy. Additionally, the NDE technique can be used to assess performance of bridges or structures in the unreachable area and hence provide better coverage of performance measure.

- Encourage usage of quantitative measures for performance: the problem with performance assessment observed in the NBI data is that they are subjectively measured. In particular, the physical condition ratings. As described in an earlier subsection, the ratings are subjective and tend to be erroneous. Moreover, an indirect implication from Washer (2003) experiment, that could explain why a wide range of ratings could be assigned to the same bridge element by different inspectors, is that the scale distance between each rating is fuzzy. In his report, Washer claimed that it could be that the scale is too close to each other, for example, there is very little difference between rating 7 to 9. To address this problem, the NBI should instead adopt quantitative driven measures. For example, change in baseline spectrum can be quantitative derived as a measure to indicate structural health condition.
- Suggested standard model for performance evaluation: the baseline model for performance evaluation especially for long-term investment purposes should be standardized among authorities. The NBI should suggest the types of model for performance evaluation and should ensure that necessary data for the model is included in the database. Not only would it serve as a clear evidence for supporting the observed or collected performance, but it also provides a baseline for researchers to be able to suggest improvements of analysis on performance model which, in turn, is beneficial for FHWA and authorities.

9.5. Conclusions

We have presented examples of initiatives that steers design philosophy or objectives from initial cost or construction performance to long-term performance and life-cycle budget allocation planning. The value of incorporation of long-term performance into design stage lies in the fact that it allows assessment of required budget level to sustain long-term performance during bridge life cycle in the hope of obtaining control over maintenance cost.

At the heart of these initiatives, solid data quality and accountable performance measures are the crucial factors. Desirable data quality must be collected in an inexpensive and accurate manner. Data must also well support and explain the observed or evaluated performances. Performance measures should be a reliable measure that subjects to less subjectivity and well capture health or status of the infrastructure.

Retrospectively, from this thesis study, it can be deduced that the current NBI data is lacking in both areas. The performance data in NBI is erroneous with expensive visual inspection. The available attributes do not possess explanatory power to the observed deterioration condition. The use of subjective rating score is subjective and thus provides a very crude and inaccurate means to assess performance. This also suggests a critical need to introduce a new way of data measurement and performance defining achievable by techniques like NDE to address the two factors for long-term performance design. Finally, a set of recommendations for NBI to address these shortfalls are presented.

10. CONCLUSIONS AND SUGGESTED FUTURE DEVELOPMENTS

This thesis proposes a developed a methodology for conceptual design knowledge learning which enables incorporation of design quality, particularly design performance, as learning objectives.

Using the NBI bridge as application domain, we have presented empirical relationships between service environments, specifications, design configurations and performances using visualization tools. Starting from basic understanding of distribution of bridge populations by different nominal NBI items to 2- or 3-way conditional analysis which helps us understand relationships between different aspects of the NBI database. The focus of our interests is mutual relationships/effects of how service environments, specifications, design configurations could affect bridge performances which in we mainly use physical condition ratings as proxies for performances.

In particular, we have found that bridges with low physical condition ratings (0-3) constitute less than 5% of the entire population. We found that bridges with low ratings tend to have more variety of materials used for construction and a wider range of ages thus suggesting that aging and some of the classic materials used contributes to deterioration. We also discovered that the normalized Average Daily Traffic per lane quantity tends to be lower for lower rating bridges compared to those with higher ratings.

For bridge performance modeling purposes, we have employed the Simultaneous Equation Modeling (SEM) approach to model the selected performances, i.e., the physical condition ratings on deck, superstructure and substructure elements. Using the estimated SEM models with only design description and the one with additional design specifications, we have found that the quality of fit is generally poor.

To justify this observation, a set of literature has been reviewed. Compared to a study done by Chase (1999) in which regional environment information is included, the result suggested additional environmental information does improve model accuracy. However, the degree of improvement is dubious because we have found that the accuracy of the measured ratings is reportedly low. Over 67% probabilistically, the measurement includes error of +/-1 scale. This also serves as a cause of poor fitness exhibited in our models. Finally, from the model perspective, visualization of the underlying joint distributions between data attributes helps to understand the cause of poor fitness quality. This is because the majority of the data space is clouded by observations with rating 7 and thus the model will pose tendency to predict most of the observation to rating 7.

For selection of design knowledge learning algorithm, we have reviewed the two main machine learning algorithms being used for engineering design knowledge learning namely the conceptual clustering and ANN techniques. The latter suffers from the opaqueness of the resulting knowledge from which engineers can interpret rationale of the design and thus provide only shallow knowledge. In contrast, the clustering type of algorithm, particularly COBWEB which is an incremental hierarchical clustering method is found more suitable for our thesis purpose.

However direct adoption of COBWEB would not satisfy our objective because its hierarchical is built based on pure statistical measure called category utility which measures a tradeoff between cluster size and predictability of a cluster class by a feature value. We proposed a modification to existing category utility function to incorporate a linearly weighted sum between the original predictability of cluster class by an attribute value and the local increment in average performance over the global average performance which is measured by the model constructed in the previous section.

A set of experiments to evaluate the proposed modified COBWEB is conducted. The first one uses the original COBWEB to build hierarchical clusters which represents design knowledge. However, no explicit incorporation of performance measure is included in learning strategy. This forms a benchmark to which the modified COBWEB is to be compared. An example of mixed performance COBWEB is used for illustration. The optimal COBWEB is the one with the linear weighting term such that it maximizes the average predicted rating normalized by prediction error. The rationale is to identify the COBWEB that best balances capturing of good design features and consistency of predicted design configuration with the actual observation. Experimental result has shown that the modified COBWEB tends to suggest design with limited ranges of choices, e.g. prestressed concrete, concrete deck type, etc. which tends to provide higher performance compared to the original COBWEB. Also, the average out-of-sample predicted rating on all three categories, i.e., DCR, SPCR and SCR are higher than those obtained from original COBWEB. Thus, the result confirms that the modified COBWEB has fulfilled our thesis purpose.

Finally, we discussed the insufficiency of the NBI in present to explain the performance measure. While the design for long-term performance has been a topic of interest for FHWA and authorities in the recent years, the quality of data in the current NBI is not adequate to support such a vision. Not only that its data quality is not reliable and erroneous, NBI's bridge performance measure such as physical condition ratings are measured subjectively and cannot be explained by the data available in the NBI database. Our recommended enhancement included: accurate data collection with the use of new technique such as NDE, less subjective and more accountable performance measures, supportive data that explains the observed performance evaluation, and standardization of analytical model for performance evaluation.

In terms of suggested future developments, potential extension in the following three main areas could be pursued.

- <u>Measurement noise as another source of disturbance to performance model</u>: as noted in chapter 6 that the physical condition rating is typically erred by +/-1. This has a great impact on performance model accuracy because the true rating now is also latent to the model. A general structural equation model such as recticular action model (RAM) (McArdle and McDonald, 1984) can be utilized. However, the difficulty lies in the fact that the measurement error probability density form is unknown and thus would falsify the model estimation if using MLE on normality assumption. Another challenge is to capture the structure of correlation between the measurement errors and the actual disturbance from latent deterioration process.

- <u>Explanatory power analysis of exogenous factors to NBI bridge condition ratings</u>: The physical condition rating model by Chase, Nutakor, and Small (1999) has included several regional environmental variables as explanatory variables. These results have, by far, exceeded our model accuracy which suggests the additive value of such information. Especially from BMS perspective, this additional information is useful for understanding the deterioration process. However, explanatory ability of each exogenous variable for modeling of performance is still worthwhile to investigate. A general framework such as PCA can be employed to estimate relevancy and the degree of variation in performance explained by each explanatory variable. However, the PCA type of analysis is limited to stationary process and continuous-scaled variable. The difficulty thus lies in development of PCA-like methodology that does not resort on stationarity and can handle nominal and ordinal variables. This is especially useful for improvement of data collection in NBI data. As far, it should be clear from chapter 6 that the available attributes in NBI are not sufficient to capture the deterioration.
- <u>Filtering of relevant features during learning process</u>: As data contains more attributes in practice, it becomes difficult for engineers to interpret knowledge at each node level, particularly in the case of COBWEB. Also this could deteriorate effectiveness of learning algorithms. A natural extension is to perform local feature selection before deploying the COBWEB. However, this type of method oversimplified the problem and could lose global information that is useful for describing local node. Rather, a construction of algorithm in which feature selection at the node level is performed while learning is more preferable. This would enable the clustering system to use all features at the global level while limiting itself to a smaller set of relevant features at the local level.

BIBLIOGRAPHY

Agarwal, P. and Raich, A.M. (2005). "Optimal Design of Bridge and Roof Truss Systems Using Multi-Objective Genetic Algorithms", Journal of Computing in Civil Engineering, 179(76).

Alpaydin, E. (2004). Introduction to Machine Learning, Cambridge, MA. MIT Press.

Agresti, A. (1996). An Introduction to Categorical Data Analysis, John Wiley and Sons Inc.

Basili, R., Pazienza, M.T. and Velardi, P. (1993). Hierarchical clustering of verbs. ACL-SIGLEX Workshop on Lexical Acquisition, Columbus, OH.

Batill, S.M. and Swift, R.A. (1993). "Structural Design Space Definition Using Neural Networks and a Reduced Knowledge Base", AIAA, AHS, and ASEE, Aerospace Design Conference, Irvine, CA.

Ben-Akiva, M. and Ramaswamy, R. (1993). "An approach for predicting latent infrastructure facility deterioration.", Transpration Science, 27(2), 174-193.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). Classification and Regression Trees. Belmont, CA. Wadsworth International Group.

Chase, S. (2003). "The bridge of the future: FHWA Bridge Plan Part II", *Roads & Bridges*, 41(7), < http://209.85.165.104/search?q=cache:bXPm2_LMTxAJ:www.roadsbridges.com/The-bridge-of-the-future-FHWA-Bridge-Plan-Part-II-article4287+design+load+HS25+adopt&hl=en&ct=clnk&cd=2&gl=us>, last accessed 08/30/2007.

Chase, S.B., Nutakor, C. and Small, E.P. (1999). "An In-depth Analysis of the National Bridge Inventory Database Utilizing Data Mining, GIS and Advanced Statistical Methods", 8-th Transportation Research Board Conference on Bridge Management, Denver, CO., C-6, 1-17.

Choi, J.Y. and Kim, P.J. (2005). "Incremental Conceptual Clustering Using a Modified Category Utility", International Conference on Circuites/Systems, Computers and Communications, 23-24.

Federal Highway Administration (FHWA). (2007), Long-Term Bridge Performance Program, < http://www.tfhrc.gov/structur/ltbp.htm>, last accessed 10/2/2007.

Federal Highway Administration (FHWA). (2004), Performance Specifications Strategic Roadmap: A Vision for the Future, < http://www.fhwa.dot.gov/construction/pssr04tc.cfm>, last accessed 10/2/2007.

Federal Highway Administration (FHWA) (1992). Recording and Coding Guide for the Structure Inventory Appraisal of the Nation's Bridges, Bridge Division, Office of Engineering R&D.

Fisher, D.H. (1987). "Knowledge acquisition via incremental conceptual clustering" Machine Learning, 2(2), 139-172.

Gluck, M., and Corter, J. (1985). "Information uncertainty and the utility of categories." Proc. Seventh Ann. Conf. Cognitive Science Society, 283-287. Irvine, CA: Lawrence Erlbaum Associates.

Golabi, K., and Shepard, R. (1997). "Pontis: A System for Maintenance Optimization and Improvement of U.S. Bridge Networks." Interfaces, 27, 71-88.

Grierson, D.E. and Khajehpour, S. (2002). "Method for Conceptual Design Applied to Office Buildings", Journal of Computation in Civil Engineering, 16: 83-103.

Haglin, D. (2007). GRB Tool Shred, <http://grb.mnsu.edu/grbts/>, last accessed 08/30/2007.

Hawk, H. (1995). "BRIDGIT Deterioration Models." Transportation Research Record 1490, Transportation Research Board, Wasihngton, D.C., 19-22.

Henningsen, A. and Hamann, J. (2006). "systemfit: A Package to Estimate Simultaneous Equation Systems in R", <http://mpra.ub.uni-muenchen.de/1421/>, last accessed 08/30/2007.

Judge, G.G. et. al. (1985). The Theory and Practice of Econometrics, New York, USA. John Wiley & Sons.

Kalyanmoy D. (2001). Multi-Objective Optimization Using Evolutionary Algorithms, West Sussex, England. John Wiley & Sons.

Kumar A.S., Subramaniam, V. and Teck T.B. (2004). "Conceptual Design of Fixtures Using Machine Learning Techniques.", The International Journal of Advanced Manufacturing Technology, 16(3): 176-181.

Lee, L. (1981). "Simultaneous Equations Models with Discrete and Censored Dependent Variables.", in C. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, Mass., 346-364.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications.

Lounis, Z. (2000), "Reliability-based life prediction of aging concrete bridge decks.", Life prediction and aging management of concrete structures, D. Naus, ed., RILEM, Paris, 229–238.

Madanet, S. Mishalani, R. and Ibrahim W.H.W. (1995). "Estimation of Infrastructure Transition Probabilities from Condition Rating Data.", Journal of Infrastructure System, 1(2), 120-125.

Mal, A., et. al. (2005), A Conceptual Structural Health Monitoring System Based on Vibration and Wave Propagation, Journal of Structural Health Monitoring, 4(3), 283-293

McArdle, J. J., and McDonald, R. P. (1984). "Some Algebraic Properties of the Reticular Action Model.", British Journal of Mathematical and Statistical Psychology, 37, 234–251.

Mitchell, T. (1997). Machine Learning, McGraw-Hill Higher Education, 56.

Morcous, G. (2006). "Performance Prediction of Bridge Deck Systems Using Markov Chain", Journal of Performance of Constructed Facilities, 20(2), 146-155.

National Cooperative Highway Research Program (NCHRP). (2006), Report 551: Performance Measures and Targets for Transportation Asset Management, Transportation Research Board of the National Academies, < http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_551.pdf>, last accessed 10/2/2007.

Potter, S., et al. (2001), "Design Synthesis Knowledge and Inductive Machine Learning.", Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 15(3), 233-249.

Quinlan, J.R. (1986) "Induction of Decision Trees." Machine Learning, 1: 81-106.

Quinlan, J.R. (1993) C4.5 Programs for Machine Learning. San Mateo, CA. Morgan Kaufmann.

Reich, Y. (1997). "Machine Learning Techniques for Civil Engineering Problems." Microcomputing in Civil Engineering, 12(4), 307-322.

Reich,Y. (1993). "The development of BRIDGER—A methodological study of research in the use of machine learning in design." Artificial Intelligence in Engineering, 8(3), 165–181.

Reich, Y. and Fenves, S.J. (1992). "Inductive learning of synthesis knowledge." International Journal of Expert Systems: Research and Applications, 5(4), 275-297.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning internal representations by error propagation. In Parallel Distributed Processing: Explorations in the Microstructures of Cognition, (Rumelhart, D.E. and McClelland, J.L., Eds.), 1, 318–362, Cambridge, Massachusetts, MIT Press.

SAMCO (2007). Feedback from Monitoring to Bridge Design, < http://www.samco.org/network/download_area/ambvibmon/03_feedback.pdf>, last accessed 10/2/2007.

Sim, K.K. and Duffy, A.H.B. (1998). "A foundation for machine learning in design." Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 12, 192-209.

Srinivas, V. and Ramanjaneyulu, K. (2007). "An Integrated Approach for Optimum Design of Bridge Decks Using Genetic Algorithms and Artificial Neural Networks", Advances in Engineering Software, 38(7), 475-487.

U.S. Department of Transportation. (1987). "Our Nation's Highways: Selected Facts and Figures.", U.S. Department of Transportation, Federal Highway Administration, Office of Highway Information Management HPM-10/4/87 (35M)E.

Washer, G.A. (2003). "Improving Bridge Inspections: Researchers at FHWA Are Developing Innovative Nondestructive Evaluation Technologies to Assess the Condition of Bridges", <http://findarticles.com/p/articles/mi_m3724/is_3_67/ai_113646119 >, last accessed 08/30/2007.

Williams, R.E. (2003). "Training Architectural Computational Critics by Example.", Master of Engineering Thesis, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2003.

Winston, P.H. (1993). Artificial Intelligence, 3rd Edition, Addison-Wesley, Massachusetts.