UTCM

"Improving the Quality of Life by Enhancing Mobility"

University Transportation Center for Mobility

A Systems Approach to Risk Reduction of Transportation Infrastructure Networks Subject to Multiple Hazards

Final Report

Mauricio Sanchez-Silva and David V. Rosowsky

Performing Organization
University Transportation Center for Mobility
Texas Transportation Institute
The Texas A&M University System
College Station, TX

Texas Transportation Institute

1. Project No. UTCM 08-01-13	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A systems approach to risk reduction of transportation infrastructure networks subject to multiple hazards		5. Report Date December 31, 2008
		6. Performing Organization Code Texas Transportation Institute
7. Author(s) Mauricio Sanchez-Silva & David V. Rosowsky		8. Performing Organization Report No. UTCM 08-01-13
9. Performing Organization Name and Address University Transportation Center for Mobility Texas Transportation Institute The Texas A&M University System 3135 TAMU College Station, TX 77843-3135		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTRT06-G-0044
12. Sponsoring Agency Name and Address Department of Transportation Research and Innovative Technology Administration 400 7th Street, SW Washington, DC 20590		13. Type of Report and Period Covered Final Report 1/1/08 - 12/31/08
		14. Sponsoring Agency Code
15. Supplementary Notes Supported by a grant from the US Department of Transportation, University Transportation Centers Program		

16. Abstract

Integrity, robustness, reliability, and resiliency of infrastructure networks are vital to the economy, security and well-being of any country. Faced with threats caused by natural and man-made hazards, transportation infrastructure network management must be directed towards: (1) understanding the network performance as a system; (2) modeling the dynamic interaction between the network and the external and internal demands; and (3) defining hazard management strategies to optimize resource allocation. The objective of the project is to develop a model of infrastructure transportation network that can be used to design efficient risk management strategies to ensure an acceptable performance (e.g., in terms of expected damage or recovery times) when subject to the action of individual, simultaneous, or sequential hazards. This study explores the performance of infrastructure networks using a systems approach. This approach is different from most existing modeling techniques in that networks will not be modeled as a collection of separate elements, but rather as a dynamic structured functional unit. This project develops new analytical methods built on a hierarchical structure of the system, which directs the analysis to the interaction and dependencies between components. This are used to characterize and model the emergent properties of the entire system. The performance of the network is integrated with the analysis of individual network components. Time-dependent models will be used for studying the life-cycle performance (mechanical and operational) of network components (e.g., bridges) and to maximize the objective performance function (e.g., cost or efficiency of the response) for different time windows. The case of the transportation network of Texas is used as an illustrative example of some parts of the model.

17. Key Word Infrastructure networks, transportation, hazards, risk management, life-cycle analysis	18. Distribution Statement Public distribution		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 70	22. Price n/a

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

A systems approach to risk reduction of transportation infrastructure networks subject to multiple hazards

Mauricio Sánchez-Silva[1] and David V. Rosowsky[2]
Zachry Department of Civil Engineering

College Station, TX,
December 2008

[1] Associate Professor at Los Andes University, Bogota, Colombia. (msanchez@uniandes.edu.co). Visiting Scholar at the Zachry Department of Civil Engineering, Texas A&M University.

[2] Department head and A.P. Florence Wiley Chair Professor at the Zachry Department of Civil Engineering, Texas A&M University (rosowsky@civil.tamu.edu).

Disclaimer:

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, Transportation Centers Program in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Contents

List of Figures

List of Tables

Executive Summary

As society and technology grow, organizations expand and construct new networks to sustain their development. Within this context, integrity, robustness, reliability, and resiliency of infrastructure networks are vital to the economy, security and well-being of any country. Studying networks' performance and their interaction is important to support decisions about risk mitigation, future development, investments and maintenance policies so that their operation can be made more efficient and reliable. This study focuses on the problem of transportation networks but the conceptual discussions and the models described can be extended to better understand the performance of other infrastructure systems.

Faced with threats caused by natural and man-made hazards, transportation infrastructure network management should look for new ways of modeling the behavior of individual components and their interaction. First, a better understanding of the network performance can be achieved by describing the network as a system i.e., a set of elements interconnected with a purpose (emergent property of the system). Identifying and defining the purpose of a network is key to the solution. In transportation networks the target of the analysis might be connectivity (the ability to travel between two sites) or an overall network performance (e.g., total travel times within the network). Secondly, it is also important to describe and model accurately the dynamic interaction between the system (network) and the imposed demands. Internal and external demands are changes in the normal operation conditions that may cause a negative impact on the system. In transportation networks an internal demand is traffic congestion or pavement deterioration; an external demand is a hurricane or a terrorist attack. Finally, models about the operation and the network's relationship with the environment should provide information that supports decisions for hazard management strategies and for optimizing resource allocation.

The objective of this project is to study new ways of understanding the performance of transportation infrastructure network systems. The study focuses on the performance of both the entire network and the individual network components. The model developed can be used to support and design efficient risk management strategies that can ensure the acceptable performance of the network (e.g., in terms of expected damage or recovery times) when subject to different hazard types. The case of the main transportation network of Texas was chosen to demonstrate the applicability of the model.

This work is divided in two parts. First, the problem of network modeling by using a systems approach and its potential applications are presented in Chapters 2 and 3. This approach is different from most existing modeling techniques in that networks will not be modeled as a collection of separate elements but rather as a dynamic structured functional unit. It uses hierarchical structures in order to better represent the network system allowing a more efficient use of information for decision making when exposed to multiple hazards. In the second part (Chapter 4), a comprehensive probabilistic time-dependent model of the lifetime performance and reliability of individual components in the network is developed and integrated into the analysis. Structural deterioration is modeled as a result of the combined action of progressive degradation (e.g., corrosion, fatigue) and sudden events (e.g., earthquakes). Probabilistic models are used as input for a reliability cost-based optimization model. Results are compared with similar models showing the importance of taking into account the damage history when studying the life-cycle of infrastructure systems.

The results of the study show that a systems approach to understanding the problem of transportation networks provides valuable information that cannot be acquired by traditional "mechanistic" models. By taking into account the internal relationships between network elements, both resource allocation and risk preventive strategies can be improved. In addition, the study of network components shows the importance of incorporating and integrating different deteriorating processes of infrastructure components (e.g., bridges) to planning and operational strategies. In conclusion, it is argued that better decisions about resource allocation for risk mitigation, development and operational planning can be obtained by integrating a systems approach to network modeling with the details of network component performance.

Chapter 1

Description of the project

1.1 The problem

Integrity, robustness, reliability, and resilience of infrastructure networks are vital to the economy, security and well-being of any country. Studying networks' performance and their interaction is important to support decisions about risk mitigation, future development, investments and maintenance policies so that their operation can be made more efficient and reliable. Among all different types of networks, the performance of transportation networks is a major issue in most countries around the world. Transportation networks deteriorate as a result of environmental factors (e.g., chloride ingress) and operation (e.g., pavement deterioration, traffic congestion). They might be also exposed to rare extreme events such as earthquakes, hurricanes, accidental blasts and terrorist attacks. Within this context, and considering the importance and cost of building and maintaining transportation infrastructure systems, resource allocation and operational decisions become paramount (Figure 1.1). It is proposed in this document that for a better understanding and management of transportation infrastructure networks it is important to work towards: (1) understanding the network performance as a system; (2) modeling the dynamic interaction between the network and the external and internal demands; and (3) defining hazard management strategies to optimize resource allocation.

1.2 Objectives

The objective of the project is to develop a model of the performance of infrastructure transportation network systems that can be used to design efficient risk management strategies to ensure acceptable performance when subject to the action of individual, simultaneous, or sequential hazards. The project focuses on the performance of both the entire network as a system and of the individual components.

The main objectives of this study are:

1. provide an overview of the modeling problems and needs of transportation infrastructure systems and components;

2. develop a model that can be used to estimate the impact of multiple hazards on a transportation infrastructure network. The transportation network of Texas is used to demonstrate the potential applicability of the model; and

3. study the infrastructure operational problems and needs of transportation network components (e.g., bridges) through a probabilistic model of the life-cycle performance of components.
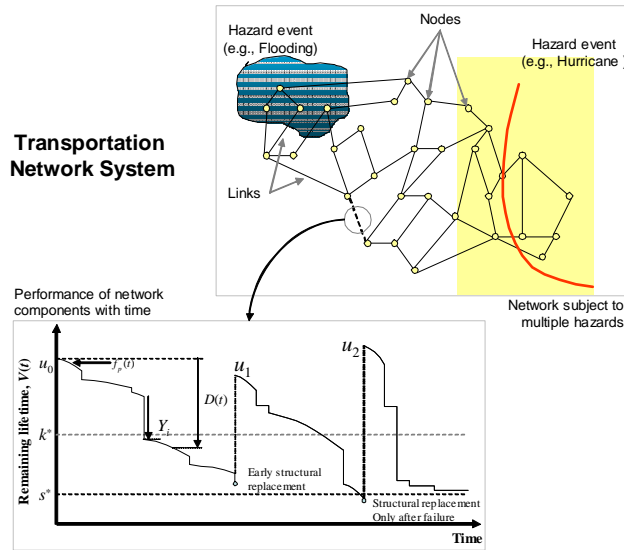
Figure 1.1: Description of the project components.

1.3 Overview

In this project, transportation networks are modeled hierarchically, in order to model different network characteristics. Chapter 2 will describe the systems approach to modeling networks. Studying the performance of infrastructure networks using a systems approach is different from most existing modeling techniques in that networks will not be modeled as a collection of separate elements but rather as a dynamic structured functional unit. In systems thinking, systems are represented hierarchically in order to deal with their complexity. For instance, it is known that with the exception of very small networks with certain geometric characteristics, close form solutions to problems such as connectivity and flow efficiency cannot be found. Given these computational difficulties, new approaches are based on detecting "community" structures in networks. These methods define rules for clustering the network vertices organizing them at different levels within a hierarchical structure allowing a systemic approach to the problem. Chapter 2 focuses on describing how networks can be described hierarchically and the clustering methods that can be used for the particular case of transportation infrastructure networks.

In Chapter 3, the practical application of the model is described with some illustrative and real examples that use the main transportation network of Texas as a case study. A decision support tool developed in Matlab$^{\circledR}$ is also described, and various cases of different hazard events are analyzed. Finally, the impact of Hurricane Ike on Texas is used as an example of the usefulness of the proposed model.

The performance of individual network elements is studied in Chapter 4. The focus will be on the life-cycle performance of infrastructure components (e.g., bridge structures). The model developed considers structural deterioration as a result of the combined action of progressive degradation (e.g., corrosion, fatigue) and sudden events (e.g., earthquakes). Structural reliability is evaluated against prescribed design and operation thresholds that can be used to establish limit states or intervention policies. Finally, probabilistic models are used as input for a reliability cost-based optimization model.

Chapter 2

System modeling of transportation networks

2.1 Introduction

Lifeline systems such as power and energy distribution, transportation, communications, internet and many others are essential for modern life. The world can be described as the multiple interactions of dynamic and diverse networks. Studying networks' behavior and their interaction is important since the results can be used to support decisions about risk mitigation, future developments, investments and maintenance policies so that the network operation can be made more efficient and reliable. Among all different types of networks, transportation networks are critical in most countries around the world. Transportation networks deteriorate as a result of environmental factors and operation. They are also exposed to rare extreme events such as earthquakes, hurricanes and terrorist attacks, which make resource allocation and operational decisions paramount.

This chapter presents a systems approach to model the performance of infrastructure networks. Systems thinking has been used as a modeling strategy in other areas of civil engineering, for instance, to model structural vulnerability and robustness of structures (Agarwal et al., 2003 [1]). The proposed model will assess the form (topology) of the network in order to develop a hierarchical description of the system. The hierarchy is obtained by unraveling the system progressively following state of the art network clustering algorithms. The resulting hierarchy is used to better understand the system's performance and as a tool to support a wide variety of decisions. The proposed approach can be extended to modeling other network types.

2.2 Network classification

Network modeling requires first defining analysis criteria. For transportation networks, the most common criteria are *form* and *flow*. Models that focus on the form of the network concentrate on changes in its structure (e.g., geometry) and address problems such as connectivity failure. On the other hand, models that focus on flow (e.g., traffic) deal with problems of cost and efficiency. Flow, in a general sense, describes whatever travels through a network. Form and flow are related concepts although the correlation may be difficult to assess explicitly.

Networks can be classified in terms of form as (1) centered systems (e.g., communications network); (2) node diverted systems (e.g., transportation network); or (3) as a combination of the above. In the first case, a set of nodes are highly connected in comparison with all other vertices, and most paths between nodes include them. In the second case, connectivity (affinity) is evenly distributed throughout the vertices. In terms of flow, networks can be classified as (1) permanently loaded (e.g., electricity distribution system); (2) unidirectional (e.g., water supply system); (3) multi-directional (e.g., road

transport network); or (4) energy dissipation system (e.g., biological systems). Permanently loaded networks are those where energy is stored within the system and neither the concept of energy flow nor directionality exists. In the second case, energy flows only in one general or local direction, for instance, a water supply system working by gravity. Most transport systems belong to the third case where energy can flow bidirectionally between the system nodes. Finally, there exists the particular case where energy flow may exit or enter into the system permanently through both nodes and links by, for instance, a diffusion processes. Additional network types can be identified in areas outside civil engineering (e.g., social sciences) (Newman, 2003 [45]).

2.3   Modeling networks as graphs

A network system is usually modeled as a graph $G(V, E)$ where V is the set of nodes $V = \{v_1, v_2, \ldots, v_n\}$ and E the set of connecting links (also called arcs), $E = \{e_1, e_2, \ldots e_n\}$. Transportation networks are usually modeled taking V as a set of points of interest (e.g., cities, special facilities like hospitals) and E as the connecting links. The condition of a network is usually modeled in terms of the state of individual links, commonly evaluated as either failure or not-failure. This means that partial failure conditions are not taken into consideration. Then the state vector of the network at time t is $X(t) = \{x_1(t), x_2(t), \ldots, x_n(t)\}$, which denotes the state vector of $G(V, E)$ such that $x_i(t) = 1$ if e_i is operating and $x_i(t) = 0$ otherwise. The purpose of a network can be described by a function $g()$. Then, any functionality analysis should address the performance of $g()$(e.g., reliability assessment). In most studies, $g()$ is described either as a connectivity function or as a travel cost function.

With the exception of very small networks with certain geometric characteristics, close form solutions to problems such as connectivity and flow efficiency cannot be found. Studies that require considering all possible failure scenarios lead to a problem known to be *NP*-hard (Nondeterministic Polynomial time hard) due to the enormous number of possible states that have to be considered (a network with n links has 2^n possible states). In other words, this means that it is most unlikely that there is a general algorithm that can compute all possible cases in polynomial time. Several approaches have been proposed to manage this problem. For example, Konak et al., 2004 [37] divided alternative approaches in three groups: (1) state-based techniques; (2) sample space arc permutations; and (3) bounding techniques. The first approach focuses on efficient methods for state vector generation or on introducing negative correlation between sample state vectors. The second method uses sample space arc permutations instead of the original sample space. The third group includes stratified and important sampling variance reduction techniques.

2.4   Properties and performance indicators

Characterizing networks is essential to building a mathematical model of the network performance. Two broad approaches have been proposed. The first focuses on statistical characterization of the network structure based mainly on topological indicators. The second approach, which has been used extensively in biological and social sciences, focuses on the properties of groups and subgroups within the network (Clausset et al., 2006 [14]).

Statistic-based measures include:

1. degree distribution;

2. maximum degree (Aiello et al., 2000 [2], Cohen et al., 2000 [15]);

3. degree correlations;

4. centrality indices (Wasserman and Faust, 1994 [70], Scott, 2000 [61]);

5. the inverse characteristic path length; and

6. distribution of the average minimum (maximum) path of a graph.

Statistic-based measures are mostly based on topological aspects, within which the degree distribution has become the most popular measure. If the degree $d(v)$ of a vertex $v \in V$ is defined as the number of links (i.e., edges), n, connected to that vertex, it is possible to build a histogram containing the fraction of vertices in the network that have degree k, i.e., p_k. Then, the degree distribution of a network is $P(k) = P(n < k) = \sum_{i=0}^{k} p_i$. Degree correlations describe how vertices with different degrees associate. The inverse characteristic path length is a global indicator of efficiency in network connectivity. It is the mean of the means of the shortest path lengths connecting each vertex to all other vertices (Duenias-Osorio et al., 2007 [18], Holme et. al., 2002 [30] and Newman, 2003 [45]). The average minimum path of a graph is the mean length of the minimum paths.

The second approach for network characterization requires taking into account the network dynamics but also many other aspects that go beyond the topology and flow (Duenias-Osorio et al., 2007 [18]). It is widely known that networks have internal structures, which are difficult to capture by simple indicators and some of them are difficult to identify (Newman and Leicht, 2007 [47]). Finding groups and subgroups is equivalent to building a hierarchical representation of the networks, which is the core of systems thinking.

2.5 Systems approach

A systems approach to any problem is built on the idea that a system cannot be modeled as a collection of separate elements but rather as a dynamic structured functional unit.

A system is defined as a set of interacting components (subsystems) called holons, organized hierarchically. A holon is both a whole and a part at the same time (Blockley and Godfrey, 2000 [8]). Holons are obtained by desegregation of the subsystem in the upper level of the hierarchy. A hierarchy structure of a system is a dynamic form of organizing information. It is not the same as a fault tree nor should it be interpreted as a rigid structure (Blockley and Godfrey, 2000 [10]). A hierarchical representation of a system requires a logical and structured way of identifying subsystems and their relationships at every level. A hierarchy is used to describe different levels of description of the system. In an upper level, holons are more general, have greater scope and are less precisely defined. In the lower layer, they have less scope and are more precisely defined. Holons have emergent properties that result from the interaction (cooperation) of the component holons (i.e., holons in an immediate lower level) but are not properties of the constitutive holons.

A systems approach helps to describe complex systems simply by providing a better representation of the problem. It provides the means to identify processes, properties and characteristics of a system at different scales (levels of precision). This is useful for supporting decisions at different levels and to make a more efficient assignment of resources.

Following a systems approach, network analysis should include four key aspects:

1. defining the network, i.e., links and nodes with their respective attributes;

2. defining the criteria for the analysis i.e., performance function $g()$, e.g., form (connectivity), resistance (strength);

3. describing the network system in a hierarchical manner by successive clustering; and

4. establishing a metric to assess the consequences of a loss in $g()$, e.g., economic consequences caused by excessive travel times.

15

2.6 Building network hierarchies

2.6.1 Clustering

A systems approach to network modeling requires the construction of a hierarchical structure describing the system at different levels of definition. In turn, this requires the development of a strategy for the identification of subsystems (holons). The problem of network separation has been studied for some decades now in many fields that cover a wide variety of networks, which include physical, biological and sociological among many others. Identifying patterns within the network around which communities of elements can be grouped is useful to simplify large and complex problems and essential to construct the hierarchical structure needed in systems thinking.

Clustering is essential for building hierarchical structures. A cluster is defined as a group of same or similar elements gathered or occurring closely together. Thus explicit clustering rules, which depend upon the problem at hand, are required; defining these rules is called cluster analysis. Cluster analysis can be defined as the mathematical study of methods for reorganizing natural groups within classes of entities (van Dongen, 2000 [68]). Cluster analysis is concerned with identifying groups of objects (people, things, events, etc.) so that the degree of association is strong between members of the same group and weak between members of different clusters (see also Lattimore et al., 2005 [38]). Reliable clustering of complex infrastructure networks might provide insights into the functioning of individual elements and the relationships within these groups of elements.

2.6.2 Clustering methods

The purpose of network clustering is to identify k groups of elements (vertices), while minimizing the number of edges that run between vertices in different groups.

Definition 1 *(Alpert et al., 1999 [4]): A graph representation of a network is kpartitioned if there is a set of non-empty clusters $P^k = \{C_1, C_2, ..., C_k\}$ such that each $v_i \in V$ is a member of exactly one C_h $1 \le h \le k$.*

Hierarchical clustering is defined in this study as the process by which a network is divided successively into clusters and subclusters. The process is carried out until the network cannot be subdivided further into smaller components. The hierarchical clustering will lead to a system representation in the form of a hierarchical tree, also called a dendogram (from Greek *dendron* "tree"-*gramma* "drawing"). The concept of hierarchical representation of a system has been widely discussed for hard (e.g., infrastructure) and soft (e.g., sociological) systems (Blockley and Godfrey, 2000 [10] and Scott, 2000 [61]).

There is not a standard approach to construct the hierarchy since it depends on the context, the type of network, the objective of the study and the computational cost. However, two main partitioning methods can be distinguished as: (1) bipartitioning; and (2) multi-partitioning. The first approach focuses on a successive bipartitioning of the network at every level. On the other hand, in the case of multiple partitions there is not a restriction in the number of clusters at every level. Hierarchical algorithms can be constructed "bottom-up" or "top-down." The former, also called agglomerative, starts with individual elements (e.g., vertices) merging new elements sucessively into larger clusters until a stopping criterion is reached (Peng and Xia, 2005 [52]). The latter (i.e., divisive) divides the whole data set sucessively into smaller clusters.

Existing partitional methods vary widely. The main difficulty of these methods is that it is usually necessary to make strong assumptions about the internal structure of the network, which is unknown. Then there are supervised and unsupervised approaches to the problem. In supervised clustering algorithms the required number of clusters is defined beforehand, and the algorithm looks for the optimal partition based on a predefined object function. Some examples of supervised algorithms are the k-means clustering method (McQueen, 1968 [40]) and the NJW method (Ng et al., 2001 [48]). Most often, it is impossible to give a reasonable estimate for the expected number of clusters. In these cases unsupervised clustering approaches such as the MCL algorithm (see section 2.7) are required.

2.6.3 Hierarchical clustering algorithms

Conceptually, most hierarchical clustering approaches are based on developing a similarity measure m_{ij} between pairs of vertices (v_i, v_j) and a iterative process of vertex grouping up to a point where minimum or maximum similarity value is achieved. For instance, two well-known methods are the *single linkage* and the *complete linkage* method (Scott, 2000 [61]). In sociology (Newman, 2004 [46]), similarity is measured through the so-called structural equivalence. Two points are structural equivalent if they have the same set of neighbors. Similarity is usually a measure of the "distance" between two vertices. Some examples are the Euclidean distance or the Pearson correlation between columns (or rows) of the adjacency matrix (Wasserman and Faust, 1994 [70]). The definition of the similarity function defines the clustering process. Then the same network may have different clusters depending upon the similarity function selected.

Original space separation

Bisection For the case of bipartitioning, two algorithms have dominated the literature: the Kernighan–Lin and the k-means (with $k = 2$) algorithm (Newman, 2004 [46]). The Kernighan-Lin algorithm optimizes the number of within- and between-community edges using a greedy algorithm (Newman, 2004 [46]). Any algorithm is said to be greedy if it recursively finds local optimum, by using a benefit (also called feasible) function, and among them determines the global optimum based on an overall objective function. The Kernighan-Lin algorithm optimizes the similarity function by swapping vertices between two initially defined clusters. This method is highly dependent on the size and selection of initial groups and can be applied only for the bisection case. Its use in complex networks, where the hidden data structure is not known, has to be managed carefully.

On the other hand, the so-called k-means method (McQueen, 1968 [40]) has been used extensively alone and in combination with other methods. The algorithm requires as input the number of clusters k; which for the case of bisection is $k = 2$. The algorithm for any case is as follows:

1. Select the number of clusters, k, for a data set D.

2. Choose the k-cluster centers at random in a domain containing all the points.

3. Compute the distance of each point to every cluster center (e.g., by using Euclidean distance).

4. Assign every point to the closet cluster center (i.e., min. distance).

5. Recompute the cluster centers using the current cluster memberships.

6. If a convergence criterion is met, stop; otherwise go to step 3.

Multiple partitions As mentioned before, the k-means algorithm can be used for multiple partitions by selecting $k > 2$. This method is very reliable and has been used extensively, alone and in conjunction with other methods. Within the category of unsupervised clustering algorithms, an efficient algorithm is the so-called Markov Cluster Algorithm (MCL). It is a fast and scalable algorithm for graphs based on simulation of (stochastic) flow in graphs (van Dongen, 2000 [68]). This algorithm will be described later in section 2.7.1.

Spectral separation

Bisection Spectral clustering has emerged recently as a popular clustering method that uses eigenvectors of a matrix derived from the data. Spectral separation techniques are based on the characteristic or spectral space imbedded within the original network structure. The first and well known case is the spectral bisection method, where separation is not carried out in the real space but in a new space defined by the eigen values of the Laplacian of the connectivity (adjacency or affinity) matrix. The Laplacian of an undirected graph G is defined as: $L_p = D - A$, where D is the diagonal matrix of

vertex degrees ($D_{ii} = \sum_j E_{ij}$) and A is the connectivity matrix with $E_{ii} = 0$. Note that for the Laplacian matrix L, all rows and columns add to zero; therefore, the vector $\mathbf{1} = (1, 1, 1...)$ is always an eigenvector with eigen value zero. It can be proved that by considering the eigen vector corresponding to the second lowest eigen value, i.e., Λ_2, it is possible to separate the vertices in two groups S_1 and S_2 as follows: $v_i \in S_1$ if $\Lambda_2(v_i) > 0$; and $v_i \in S_2$ if $\Lambda_2(v_i) < 0$. The second eigen value λ_2 is usually called the *algebraic connectivity* of the graph; smaller values correspond to better partitions. The spectral bisection method works well where communities can be separated easily, but as networks become very complex the quality of separation is not evident. Newman, 2004 [46] argues that partitioning into larger communities by successive bisection does not always give satisfactory results.

Multiple partitions   Several attempts have been made to extend spectral bipartitioning method to a multiple-eigen vector partitioning (Barnes, 1982 [8], Rendel and Wolkowicz, 1995 [57] and Hadley et al., 1992 [27]). However, methods based on iterative approaches similar to the k-means have proven to be more effective. A major drawback to k-means is that it cannot separate clusters that are non-linearly separable in input space. In order to solve this problem, some algorithms use k-means in a higher-dimensional feature space, which is defined after mapping the elements by using a nonlinear function (kernel). In these cases, data are partitioned by linear separators in the new space. Other approaches use the so-called spectral clustering algorithms, which use the eigen vectors of an affinity matrix to obtain a clustering of the data (Dhillon et al., 2004 [17]). Recently some approaches combine the main characteristics of both approaches resulting in very reliable separation methods (e.g., Ng et al., 2001 [48], Dhillon et al., 2004 [17] and Kanungo et. al., 2002 [33]). One such method, the NJW algorithm (Ng et al., 2001 [48]) will be described and some consideration for its applicability and extension will be presented in section 2.7.2.

2.7   Selected algorithms for modeling transportation networks

Systems modeling of transportation networks requires a multiple partition algorithm. However, the number of clusters at every level is not fixed; on the contrary, it depends on the inherent structure of the network. Therefore two partition methods were selected for this study: (1) the MCL method; and (2) the NJW method. MCL performs an unsupervised search of clusters at every level, while in NJW the number of clusters required has to be provided. In addition, MCL performs the calculation in the original space, while NJW uses the characteristic space defined by the first k-eigen vectors of the affinity matrix. Both algorithms are presented in the next sections.

2.7.1   MCL clustering algorithm

Description

The MCL algorithm finds cluster structure in graphs by a mathematical bootstrapping procedure. This is achieved without any prior knowledge of the graph's cluster structure. The MCL clustering process consists of dividing the node set of a graph into natural groups with respect to the edge relationship. In an analogy to transportation, the MCL method is based on the following assumptions for clustering:

1. for a traveler, there are many different ways of traveling between any two points within the same cluster and only a few ways if they belong to different clusters;

2. if a traveler drives randomly he/she will remain within the same cluster for a long time; and

3. links connecting clusters are likely to be in all shortest paths between nodes located in different clusters.

The MCL algorithm simulates random walks (i.e., flow) within a graph by alternation of two processes: 1) expansion and 2) inflation. The combined iterative process of expansion and inflation

reveals the presence of innate cluster structures in the input graph. The MCL process causes flow to spread out within natural clusters and to evaporate in between different clusters. The mathematics associated with the MCL process shows that there is an intrinsic relationship between the MCL process and cluster structure in graphs. This is very valuable given the many heuristic approaches in cluster analysis.

Expansion models the spreading out of flow; it becomes more homogeneous. In other words, it is used to compute random walks of higher length (i.e., many steps). Since higher length paths are more common within clusters than between different clusters, the probabilities associated with traveling between node pairs within the same cluster are expected to be larger as there are many ways of going from one to the other. Expansion is carried out by squaring a stochastic matrix. On the other hand, inflation models the contraction of flow; it becomes thicker in regions of higher current and thinner in regions of lower current. Inflation boosts the probabilities of walks within the same cluster and will demote inter-cluster walks. Expansion and inflation operators are alternated. Each operator takes the Markov matrix computed by the preceding operator to compute a new Markov matrix, which is passed back to the previous operator (van Dongen, 2000 [68]).

Sequential alternating expansion and inflation results in the separation of the graph into different sets of elements, which are interpreted as clusters. The stopping criterion is reached once the Markov probability matrix stabilizes. An equilibrium state takes the form of a so-called doubly idempotent matrix, i.e., a matrix that does not change with further expansion or inflation steps. The graph associated with such a matrix consists of different connected directed components. Each component is interpreted as a cluster and has a star-like form, with one attractor in the centre and arcs going from all nodes of that component to the attractor. In theory, attractor systems with more than one attractor may occur (these do not change the cluster interpretation). Also, nodes may exist that are connected to different stars, which is interpreted as cluster overlap or as nodes that belong to multiple clusters.

The MCL algorithm is as follows (van Dongen, 2000 [68]):

1. Construct the affinity matrix $A \in \mathbb{R}^{n \times n}$ such that $A_{ij} = s_i - s_j$ if there is a connection between nodes (data points) i and j, and $A_{ij} = 0$ otherwise.

2. Define the diagonal matrix $D \in \mathbb{R}^{n \times n}$ such that $D_{ii} = \sum_{k=1}^{n} A_{ik}$.

3. Construct the *associated Markov graph* as $T_G = AD^{-1}$; and $T_1 = T_G$.

4. $k = 1$.

5. $T_{2k} = f(T_{2k-1})$.

6. $T_{2k+1} = \Gamma_r(T_{2k})$.

7. Check convergence criteria (i.e., $T_{2k-1} \rightarrow$ idempotent). If convergence is not achieved, $k = k + 1$ and go back to step 5.

8. Use T_{2k-1} to create clusters.

In step 1, the affinity (connectivity) matrix of a graph G is created, i.e., matrix A. As mentioned before, matrix A indicates the existence of connectivity between any pair of nodes i and j, i.e., A_{ij}. Connectivity may be described as the existence or not of a link and it may as well include the cost of such a link (e.g., travel time). In step 2, the degree of every vertex is computed and placed in a diagonal matrix D. In step 3, the matrix T_G corresponds to a new graph G' called the *associated Markov graph* of G. The directed weight function of G' is called the *localized interpretation* of the weight function of G. In this transformation the value $(T_G)_{ij}$ represents the level of attraction between vertices i and j only in the context of the other values in the column. This is also called a column stochastic matrix, which is a non-negative matrix within which each column adds to 1.

The process that follows in steps 4 to 7 is the sequential alternation of expansion and inflation until convergence is reached. Expansion can be made in different ways according to a function $f(T_{2k-1})$. Afterwards, inflation is carried out by taking the Hadamard power of a matrix (taking powers entry wise), followed by a scaling step, such that the resulting matrix is stochastic again, i.e., the matrix elements (on each column) correspond to probability values. In practice, it means applying the operator Γ to the expanded matrix, i.e., $T_{2k+1} = \Gamma(T_{2k})$. Then, If $T \in \mathbb{R}^{n \times n}$ and $r > 1$,

$$\Gamma_r : \mathbb{R}^{k \times \ell} \to \mathbb{R}^{k \times \ell}; \Gamma_r(T)_{ij} = \frac{(T_{ij})^r}{\sum_{i=1}^{k}(T_{ij})^r} \tag{2.1}$$

Note that inflation changes the probabilities associated with the random walks departing from one particular node (corresponding with a matrix column) by favoring more probable walks over less probable walks. It is also important to note that a matrix does not have to be square. The parameter r is usually assumed to be 2 but there are no specific restrictions and it is mainly context dependent. Values of $0 \leq r \leq 1$ increase the homogeneity of the argument probability vector or matrix; while Values of $r > 1$ reduces the homogeneity. Negative values of r invert the ordering (van Dongen, 2000 [68]). Convergence is evaluated at every iteration by checking that the matrix is idempotent under matrix multiplication (i.e., it has stabilized); this is also called a doubly idempotent matrix. Idempotence is a property of mathematical operations in which the successive application of the same operation yields the same result.

The identification of clusters is carried out in step 8. The final matrix T_{2k+1}, which is idempotent at the end of the process, can be mapped with the associated graph G. In matrix T_{2k+1} every column corresponds to a vertex and all rows j of T_{2k+1}, with $j \neq 0$, are the cluster attractors and contain the cluster component vertices. Every cluster has at least one element that is unique to the cluster but it is possible that some vertices may belong to several clusters. This problem usually can be overcome by merging clusters with shared elements. Further discussion on this issue can be found in Van Dongen, 2000 [68].

2.7.2 NJW clustering algorithm

The NJW algorithm was proposed by Ng et. al., 2001 [48]. It is an efficient spectral clustering method that uses the first k eigen values of the connectivity matrix. The key to solution is to perform the separation in the $\mathbb{R}^{n \times n}$ space and not in the real space. The NJW is as follows: consider a set of points $S = \{s_1, s_2, ..., s_n\}$ in $\mathbb{R}^l$ that will be clustered into k subsets. Then,

1. Construct the affinity matrix $A \in \mathbb{R}^{n \times n}$ such that $A_{ij} = s_i - s_j$ if there is a connection between nodes (data points) i and j, and $A_{ij} = 0$ otherwise.

2. Recalculate the affinity matrix by using the appropriate nonlinear function (i.e., kernel). For example, for $A \in \mathbb{R}^{n \times n}$, $A_{ij} = \exp\left(-\left\|s_i - s_j\right\|^2 / 2\sigma_{ij}^2\right)$ and $A_{ii} = 0$.

3. Define the diagonal matrix $D \in \mathbb{R}^{n \times n}$ such that $D_{ii} = \sum_{k=1}^{n} A_{ik}$.

4. Calculate the matrix $L = D^{-1/2}AD^{-1/2}$.

5. Define the matrix $X \in \mathbb{R}^{n \times k}$ as the k-largest eigen vectors of L (organized by columns).

6. Compute the matrix $Y \in \mathbb{R}^{n \times k}$ by normalizing the rows of matrix X; i.e., $Y_{ij} = X_{ij} / \left(\sum_{k=1}^{k} X_{ik}^2\right)^{1/2}$.

7. Treating each row of Y as a point in $\mathbb{R}^k$, cluster them into k groups via k-means or any other algorithm.

8. Assign the original points s_i to the cluster of y_i.

For transportation networks, for instance, $A_{ij} = s_i - s_j$ (step 1) corresponds to the cost of travelling between nodes i and j. Step 2 is a way of improving clustering by carrying out a nonlinear mapping of the original data set (input space) into a higher dimensional feature space. This process helps to construct clusters that are not linearly separable in the original space. This non-liner mapping can be performed by selecting the appropriate kernel functions. The most popular kernel function is the Gaussian kernel, i.e., $\kappa(a, b) = \exp(-\|a - b\|^2 / 2\sigma^2)$; where σ^2 is usually called the scaling parameter, which controls the speed at which the affinity (i.e., connectivity) decreases with the distance between a and b. Others kernels such as the polynomial $(\kappa(a, b) = (a \cdot b + c)^2)$ or Sigmoid $(\kappa(a, b) = \tanh(a \cdot b)c + \alpha)$ are also used.

Step 3 places in the diagonal the degree of every vertex once the kernel has been applied. In step 4, the matrix L can also be computed as the Laplacian matrix, i.e., $L_p = D - A$. The difference between L and L_p is in the sign of the eigen values but not in the actual eigen vectors. Step 5 defines the new space where separation will be carried out. This space consists of k-eigen vectors of matrix L, which, in this case, are the largest orthogonal eigen vectors. As the number of eigen vectors considered increases, more noise is introduced in the system and the clustering becomes more difficult (Pentney and Meila, 2005 [53]). In the new space, the coordinates of every vertex, v_i, are defined by $X_{v_i} = \{x_{v_i}^1, x_{v_i}^2, ..., x_{v_i}^k\}$ where $x_{v_i}^j$ corresponds to the j component associated with the j eigen vector that defines the $k$dimensional space. The k-means algorithm used for separation required in step 7 was already explained in section 1.6. Several methods along these lines have been proposed by Alpert and Kahng, 1994 [3], Alpert et al., 1999 [4], Chang et al., 1994 [11], Ng et al., 2001 [48], Kanungo et al., 2002 [33].

Chapter 3

Network modeling implementation: case of the Texas transportation network system

3.1 Introduction

In this chapter, the network clustering algorithms presented in the previous chapter are implemented. The model is applied to the transportation network of Texas as a case study. Texas is the second largest state in the U.S. with an area of $696,621$ km^2 and a population of 22.9 million distributed in large cities such as Houston, Dallas and San Antonio as well as mid-size, but still large, urban centers such as Austin and El Paso. The Gross State Product is the third among all U.S. states (US$ 880.9 billion) with an average annual growth of 4%, a GDP per capita of US$ $38,536$ and an average family income at US$$49,086$ (source: http://www.statemaster.com).

3.2 Hierarchical model

Texas has one the largest transportation networks in the U.S. Figure 3.1 presents a map of Texas in which the main road network can be observed. The main statistics about Texas road infrastructure and some global socioeconomic measures are shown in Table 3.1 (source: http://www.statemaster.com). The case study presented in this chapter focuses on the main road transportation network of Texas, which includes all interstate, primary and secondary roads. The network studied has a total length of 16,065 miles (20% of the total road network) connecting 525 nodes that represent urban centers or major intersections. A hierarchical representation of the system was developed based on the following assumptions: (1) neither the road conditions nor the traffic flow characteristics within the network were taken into account; (2) the cost of traveling was assumed linearly proportional to the distance, i.e., $C_{ij} = \alpha d_{ij}$; and (3) the cost of freight and private vehicles was not differentiated.

In order to construct a hierarchical representation of the network, a modified version of the MCL algorithm (see Chapter 2) for clustering was used. Within this version, an additional step is added to improve clustering separation. The complete algorithm is as follows:

1. Construct the affinity matrix $A \in \mathbb{R}^{n \times n}$ such that $A_{ij} = s_i - s_j$ if there is a connection between nodes (data points) i and j, and $A_{ij} = 0$ otherwise.

2. Recalculate the affinity matrix by using the appropriate nonlinear function (i.e., kernel). For example, for $A \in \mathbb{R}^{n \times n}$, $A_{ij} = \exp\left(-\left\|s_i - s_j\right\|^2 / 2\sigma_{ij}^2\right)$ and $A_{ii} = 0$.

Figure 3.1: Road network of Texas (source: Google Earth$^{\mathrm{TM}}$)

Item	Texas	US Total	US Rank
Road Transportation statistics:			
Total Mileage	83,852	912,024	-
Mileage (Very Good)	3,873	116,520	9
Mileage (Good)	18,853	256,146	1
Mileage (Fair)	41,781	372,002	1
Mileage (Mediocre)	9,254	101,482	2
Mileage (Poor)	10,091	65,874	6
Total Bridges	48,492	590,111	1
Toll Bridges	24	172	3
Bridges (Struct. Deficient)	2,777	79,526	9
Bridges (Funct. Obsolete)	7,543	80,232	1
Socioeconomic statistics:			
Area	696,621	9,829,499	2
Population	22,859,968	300,737,973	2
GDP/capita	38,536	38,014	22

Table 3.1: Basic infrastructure and social statistics of Texas, USA.

Level	Cluster	Modified MCL - parameters				
		r	μ_L	σ_L	σ_L/μ_L	α
2	*Texas*	1.10	20.76	30.62	1.47	0.15
3	S. Antonio	1.09	18.26	21.4	1.17	0.2
	Houston	1.08	12.86	12.17	0.94	0.35
	Dallas	1.06	12.20	11.88	0.97	0.35
	Amarillo	1.08	19.39	21.72	1.12	0.2
	El Paso	1.10	28.2	27.97	0.99	0.15
$\alpha = 1/2\sigma_{ij}^2$						

Table 3.2: Parameters of the model for the clustering process at every level in the hierarchy.

3. Define the diagonal matrix $D \in \mathbb{R}^{n \times n}$ such that $D_{ii} = \sum_{k=1}^{n} A_{ik}$.

4. Construct the *associated Markov graph* as $T_G = AD^{-1}$; and $T_1 = T_G$.

5. $k = 1$.

6. $T_{2k} = f(T_{2k-1})$.

7. $T_{2k+1} = \Gamma_r(T_{2k})$.

8. Check convergence criteria (i.e., $T_{2k-1} \rightarrow$idempotent). If convergence is not achieved $k = k + 1$ and go back to step 6.

9. Use T_{2k-1} to create clusters.

It is important to stress that step 2 in the modified algorithm is included to enhance the clustering process. This modification uses the mean and standard deviation of the distances that a user will have to travel from any vertex to the next in the network to be clustered, and is calculated as:

$$\mu_L = \frac{1}{n} \sum_{i=1}^{n} L_i; \text{ and } \sigma_L = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (L_i - \mu_L)^2} \tag{3.1}$$

where n is the total number of links and L_i the length of link i. Note that this information changes at every level, and therefore, appropriate values have to be chosen accordingly. These parameters for levels 2 and 3 are shown in Table 3.2. It should be stressed that clusters group nodes only, and they do not provide information about physical and spatial boundaries on the region. However, for illustration purposes, the subclusters identified at levels 2 and 3 of the hierarchy are shown in Figures 3.2 and 3.3 by simple and heuristic interpolation. It can be observed that the results are intuitively consistent; for instance, clusters are constructed around the Texas major urban centers. The number of clusters describing the system at every level, from top to bottom were: 1, 5 (Figure 3.2), 28 (Figure 3.3), 95, 199, 284 and 313.

3.3 Impact index-based maps

3.3.1 Hazard impact index

In most hazard analyses the impact of any event is usually restricted to computing the direct losses. Direct losses are estimated based on the structural response of the system components within the event path or the event area of influence. However, for the particular case of infrastructure systems, the impact of an event on a set of components (arcs or vertices) might have consequences that propagate beyond the event's area of influence. In the proposed model the impact of any event on a given element
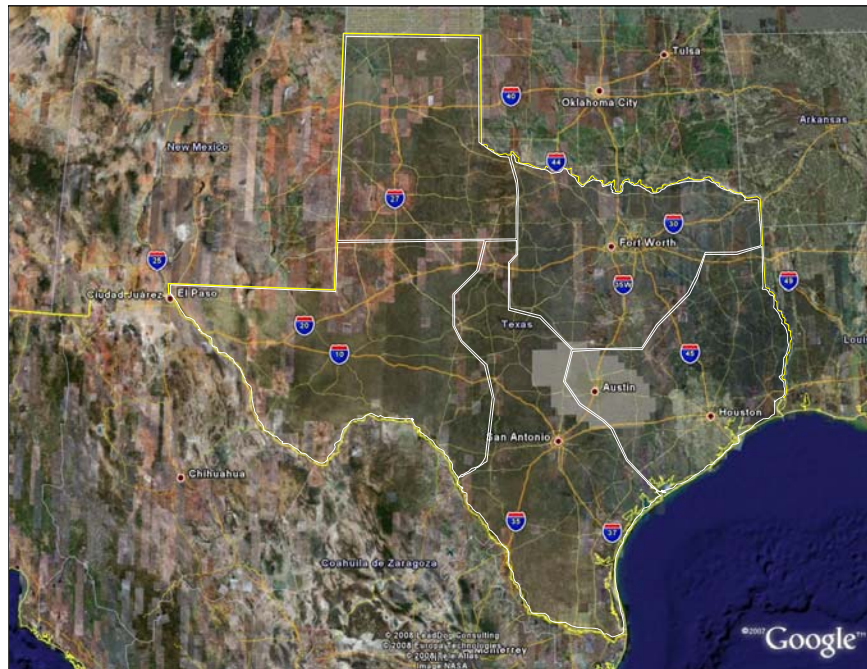
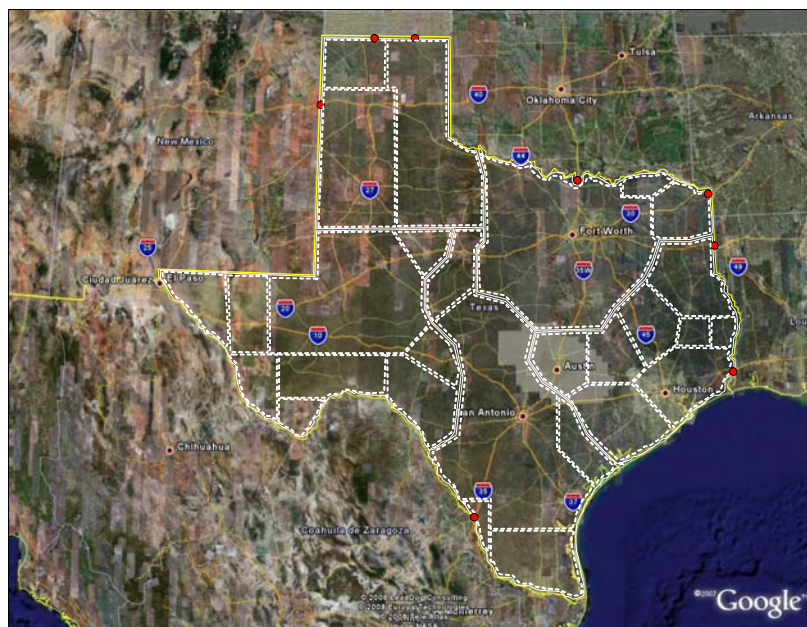Figure 3.2: Subclusters identified by the modified MCL-algorithmin in the second level of the hierarchy.

Figure 3.3: Subclusters identified by the modified MCL-algorithm in the third level of the hierarchy.
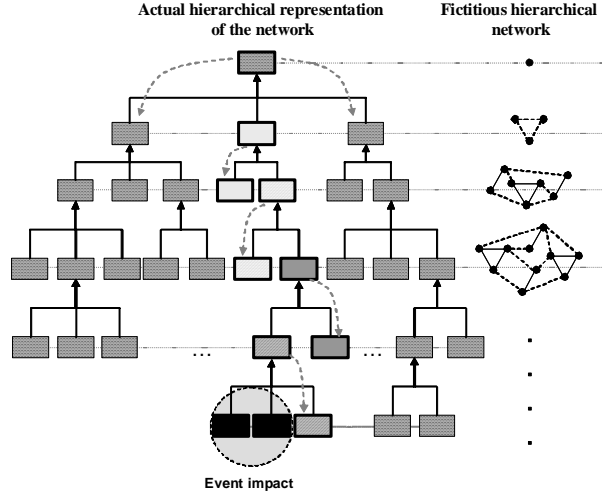
Figure 3.4: Network hierarchy representation, fictitious network and impact index assignment.

(e.g., node) of the network is a measure of the "closeness" of that element to the hazard event. This closeness is twofold: in terms of the direct physical closeness and indirectly through the relationship with other affected nodes. Then nodes that are not directly (physically) affected by the event will be affected indirectly if they share the cluster with the nodes directly affected. While the direct impact is fixed for a given event, the indirect impact is defined by the level in the hierarchy. As the analysis moves up in the hierarchy, clusters are larger and more nodes share the clusters.

For a system described hierarchically, a parallel fictitious network can be constructed (Figure 3.4 and Figure 3.5). In the fictitious network, every element (node) corresponds to a cluster. Links between nodes are parallel arrangements of the actual links connecting original nodes between clusters. At the top of the fictitious hierarchy there is one single element that represents the whole network while, at the bottom, clusters and original nodes are the same. Figure 3.5a and 3.5b show the complete network (for simplicity only the nodes are shown) and the fictitious networks at levels 2 and 3 (nodes shown as filled circles and links in dotted lines).

3.3.2   Construction of the impact index

In order to construct the impact index map, it is initially required to define the physical characteristics of the event (extent and intensity) and the decision level L at which the analysis will be carried out.

Define $\mathbf{C}_L = \{C_1, C_2, \ldots, C_q\}$ as the set of all clusters at level L in the hierarchy. $\mathbf{C}_A^L = \{C_1, C_2, \ldots, C_p\}$, with $p < q$, is the subset of clusters containing nodes directly affected by the event at level L ($\mathbf{C}_A^L \subseteq \mathbf{C}_L$). Then the index value assigned to every node e_i with $i = 1, 2, \ldots k$, in level L is given by:

$$I_L(e_i) = \left\{ \begin{array}{ll} f(L, \Phi) \cdot E_I & e_i \in \mathbf{C}_A^L \\ 0 & e_i \notin \mathbf{C}_A^L \end{array} \right\} \tag{3.2}$$

where L is the level at which the element is evaluated and Φ is a vector of parameters and attributes that characterize the network and its components, e.g., the number of levels considered in the analysis (N_L), the number of clusters at every level or any other relevant criteria for the decision. The function $f(L, \Phi)$ can take values within the range $[0, 1]$; it provides the range of evaluation for the index. The function f is maximum at the lowest level in the hierarchy and decays as the analysis moves up to higher levels. Although decay rules are case-specific, polynomial and linear decay shapes seem to be reasonable to use in practice. E_I is a value within the range $[0, 1]$ describing the event intensity; the

27

Figure 3.5: Real and fictitious networks: (a) level 2; and (b) level 3.

lower the value, the lower the event intensity and $E_I = 1$ the maximum intensity. At any level, the same value will be assigned to all the elements that share the cluster with the nodes affected.

After the assessment level by level, the final index value assigned to an element e_i of the network can be computed as:

$$h(e_i) = \max_{\omega} \left[I_{\omega}(e_i) \right]; \ \omega = 1, 2, ..., N_L \tag{3.3}$$

This index provides information on the relationship between local and global effects of the system. If the assessment is taken to the upper levels, the interconnection of components is more redundant, and the emergent properties are more complex. Then the impact of the event on the system is defused leading to smaller index values and adding more elements to the set of affected nodes. It is stressed that this model challenges certain aspects of existing network failure mechanisms such as cascade failure since damage propagation is associated to the cluster structure within the hierarchical organization and not only to its topology. Note also that by establishing N_L, the map of index values will be automatically defined. Then maps do not indicate the level of expected damage nor the size of the threat. They indicate the relative degree in which every part of the system is affected as a result of the network interactions.

3.3.3 Quantification of the impact

In addition to the index, decisions are based on actual data that can measure the consequences. If Φ is a vector parameter describing a set of loss units (e.g., casualties, GDP, revenue, etc.), the impact of the event on the entire network system can be estimated as:

$$m\left(\Phi\right) = \sum_{i=1}^{k} r\left(\Phi_{i,\omega}\right) h(e_i) \tag{3.4}$$

where ω represents a particular attribute loss to be measured and $i = 1, 2, ..., k$ the number of nodes. The function r can be seen as a vulnerability function relating the event intensity with the losses to be measured. For instance, it might describe the population assigned to a node. The assessment of losses, as a result of link failures, is beyond the scope of this paper.

3.3.4 Overall algorithm

The overall algorithm of the proposed methodology to estimate the losses caused by a hazardous event on the transportation network system is as follows:

1. define the network (nodes, links, attributes);

2. select criteria for clustering (e.g., cost of traveling);

3. construct the hierarchy (up to a prespecified level of detail, N_L) by carrying out a successive non-supervised clustering process;

4. define and characterize the event or set of potential damaging events to be considered (spatial distribution, intensity etc.);

5. identify the set of elements (nodes or links) affected by the event;

6. compute the indicator $h(e_i)$ for every node based on the hierarchical structure; and

7. when applicable, make the appropriate interpolation to take into account the spatial consequences of the event.

3.4 Model implementation

A network management tool in the form of a mathematical algorithm was developed in Matlab$^{\circledR}$. This tool can handle different hazard event types and provides information about their impact on the network. The consequences of a hazard are evaluated in terms of the impact index and by an estimation of the direct and indirect consequences. The software also provides information about the properties of any network. The interface of the tool is shown in Figure 3.6.

3.4.1 Type of events

In the algorithm developed, four event types can be used to describe the area of influence of a hazard: (1) individual node disconnection; (2) circular region; (3) irregular region (e.g., flooding); and (4) event path (e.g., hurricane). The model also allows combining different hazard types. The extent of the hazard and its intensity are usually defined by the physical characteristics of the event and are an input to the model. Figure 3.7 shows an illustrative example of the first two cases and Figure 3.8 an example of cases 3 and 4. For illustrative purposes, the results obtained for the events shown in Figures 3.7 and 3.8 were computed using the following basic information:

- the analysis was carried out at level 3 in the hierarchy for all cases;

- the model used to compute the hazard index (i.e., f) was assumed to be linear;

- the limits for defining the qualitative measure of the impact index were: "High" = $[0.75 - 1]$, "Moderate" = $[0.40 - 0.75]$, "Low" = $[0.15 - 0.40]$, and "Minor" = $[0 - 0.15]$;

- the limits (minimum and maximum) of the population affected were computed from a uniform distribution with mean 50% and COV 25%;

- the limits (minimum and maximum) of the impact on productivity were computed from a uniform distribution with mean 25% and COV 60%; and

- the hazard event intensity was taken as $E_I = 1$.
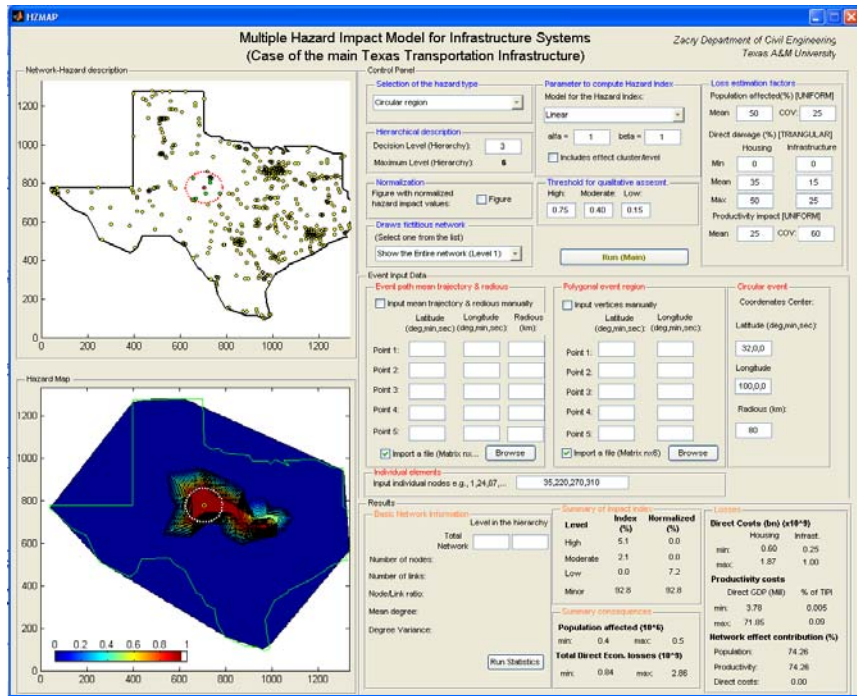
Figure 3.6: Interface of the software developed in Matlab® to model the impact of various hazard events on the network.
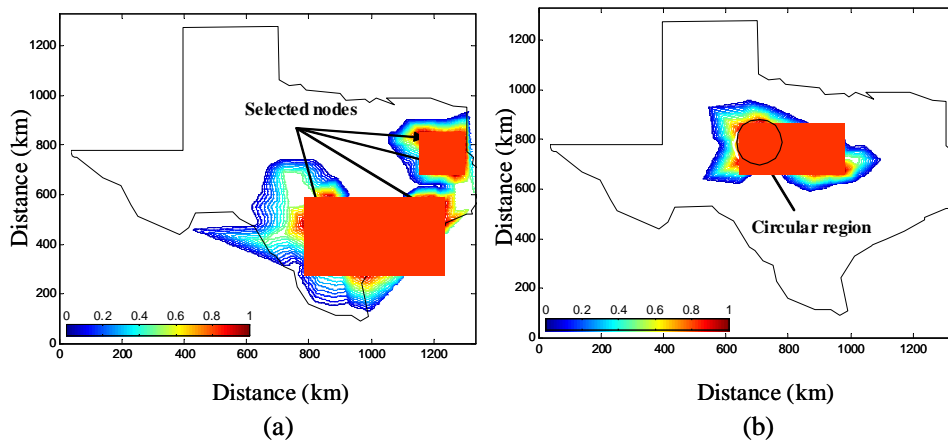

Figure 3.7: Impact index scenarios for: (a) selected nodes (35, 220, 270, 310); and (b) for a circular region.
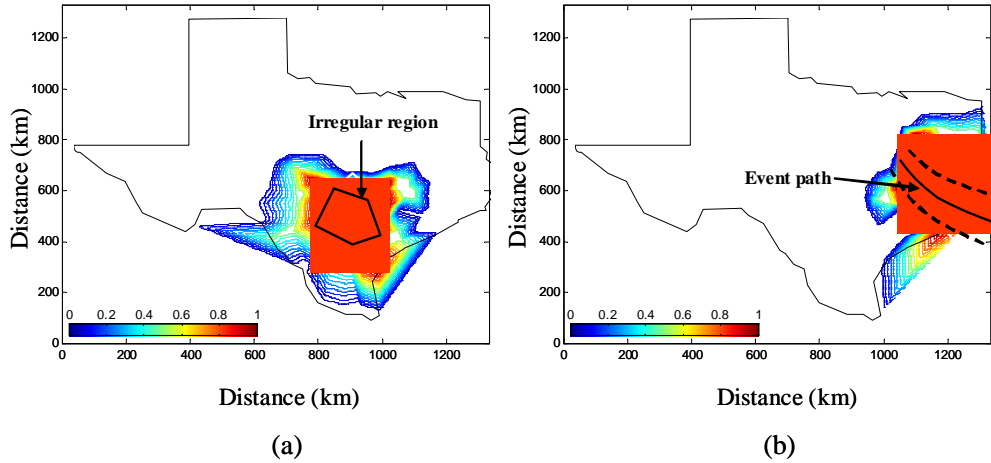
Figure 3.8: Impact scenaros for: (a) irregular region; and (b) event path.

Table 3.3 presents a summary of the impact of these events, the consequences and the contribution of the elements of the network. It can be observed from the figures that the impact of the hazard events is not restricted to their areas of influence. On the contrary, the impact includes other nodes that are beyond the event physical limits. Losses are estimated based on the attributes of every node. The properties of every node are defined by assigning to it a population size, housing and infrastructure value and productivity according to the actual population spatial distribution. The contribution to damage of nodes inside and outside the physical area of influence of the hazard event were separated. Then the percent contribution of nodes outside the physical area of the event is shown in the last row of Table 3.3. It can be observed that it is higher when the event consists of a set of individual selected nodes (97.56%). As for the other cases, the result depends highly on the nodes affected and the connectivity of the network. The model can also be adapted to carry out simulation studies. In this case, events can be generated randomly on a specific area, and the consequences can be characterized probabilistically to support intervention decisions.

3.4.2   Evaluation at different levels

The main advantage of using a hierarchical structure to support the decision making process is that information and evidence can be obtained at different levels of precision, i.e., different levels within the hierarchical description of the system. Upper levels provide a general overview of the problem and contribute with information with large uncertainty; lower levels provide more detailed descriptions of the impact and smaller uncertainty. If the event is known and can be completely characterized (i.e., spatial distribution and intensity), lower levels in the hierarchy become very important to make accurate decisions. Nevertheless, if there is uncertainty about the event location or the event intensity, middle and upper hierarchical levels become more important.

Figure 3.9 presents the case of a hazard event defined by a circular region. For this case, the impact index is evaluated at levels 2, 4 and 6. It can be observed that for the upper levels in the hierarchy the most critical area is larger than for the lower levels, where it is reduced along with the uncertainty. The results of the impact of this event for each level are shown in Table 3.4, where only indirect impact is reported (population affected and productivity losses). It can be observed that as the analysis is taken to the lower levels of the hierarchy, the estimation of losses and population affected decrease.

31

Measure	Selected Nodes	Circular Region	Irregular Region	Event Path
Impact Index (% of total network)				
High	25.1	5.1	14.1	17
Moderate	8.2	2.1	7.6	9.3
Low	0	0	0	0
Minor	66.7	92.8	78.3	73.7
Population Affected (million)				
min	4.2	0.4	2	3.2
max	5.3	0.5	2.5	3.9
Direct Losses (US$ million)				
Housing + Infrastructure				
min	1.27	0.85	14.08	31.13
max	4.33	2.87	47.76	105.65
Productivity Losses (US$ million)				
min	42.34	3.78	19.88	31.55
max	804.51	71.85	377.63	599.41
Network Effect Contribution(%)	97.56	74.26	57.45	25.33

Table 3.3: Results obtained for all four illustrative hazard events considered.
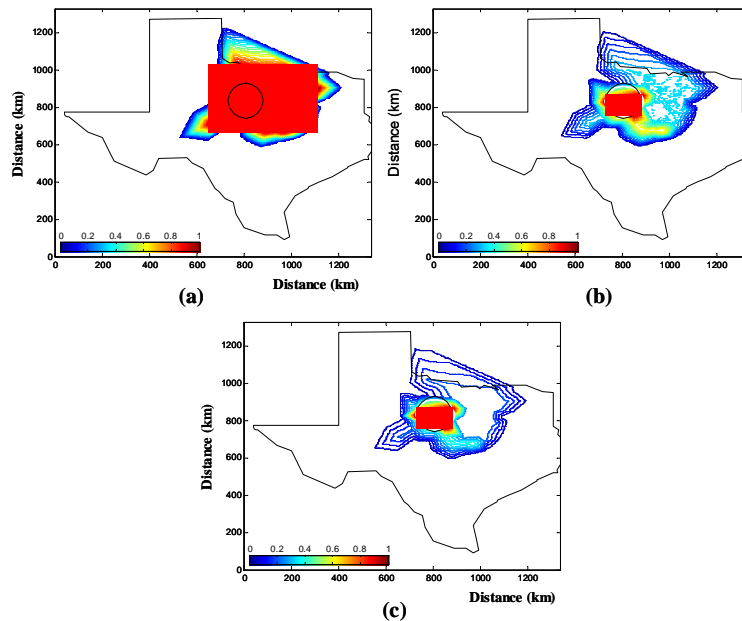

Figure 3.9: Impact index-based map for a circular hazard and three levels in the hierarchy: 2, 4 and 6.

Measure	Hierarchical Level		
	2	**4**	**6**
Impact Index (% of total network)			
High	26.9	1.3	1.3
Moderate	0	2.9	0
Low	0	22.7	25.5
Minor	73.1	73.1	73.1
Population Affected (million)			
minimum	3.2	1.2	0.7
maximum	3.9	1.5	0.9
Productivity Losses (US$ million)			
minimum	31.75	11.87	7.35
maximum	603.24	225.47	139.58
Network Effect Contribution (%)	98.22	95.24	92.31

Table 3.4: Impact index maps computed at level (a) 2; (b) 4; and (c) 6.

3.4.3 Impact of Hurricane Ike on the Texas transportation network

Description of Hurricane Ike

Hurricane Ike was the third major hurricane of the 2008 season and made landfall at 2:10 a.m. CDT on September 13 over the east end of Galveston Island (Figure 3.10). At landfall, Ike was classified as a category 3 hurricane with wind speeds of 185 km/hr. The consequences of the hurricane are still being evaluated but rough estimates from different sources predict about 5 million people left without water and power, insured economic losses between $8bn and $18bn and extensive environmental damage in Galveston and large areas along the coast.

Network impact index

The impact of Hurricane Ike was divided into two parts: flooding and strong winds. The network impact indices of both events (evaluated at level 2 in the hierarchical model) are shown in Figure 3.11. The physical extent of the hazard is represented by the dotted lines. It can be observed in Figure 3.11 that the impact of Ike on the network goes beyond the area affected directly by the event. This implies that there is an important impact on connectivity, i.e., cost of traveling and transporting goods. This assessment is an improvement over traditional models that focus only on estimating direct costs reported in the affected area.

The idea behind constructing the hierarchy is to provide a tool that can be used at different levels of precision. Then any assessment about the performance of the network, g, can be made with more precision as more information about the event and the nodes is provided. Figure 3.12 shows the impact of high wind speeds on the network when evaluated at level 4 in the hierarchy. In this case, given that more detailed information about nodes and links is provided, the impact index is computed again. Then some nodes that had a high impact index at level 2 were given a smaller value. The analysis in the upper levels of the hierarchy are useful to make rough estimates of the event's impact, but more precise forecasts of the consequences can be obtained from analysis in lower hierarchical levels.

Social and economic impact

In order to make an estimation of direct costs (damage to housing and infrastructure) the model uses a strategy similarly to the ones used in traditional analysis. Every node in the event path (influence area) has as one of its attributes the value of housing and infrastructure. Then based on an approximate triangular distribution relating flooding and damage and wind-speed and damage, the total direct costs were estimated within the range [$10.92, $37.92] billions. These values are independent of the level at

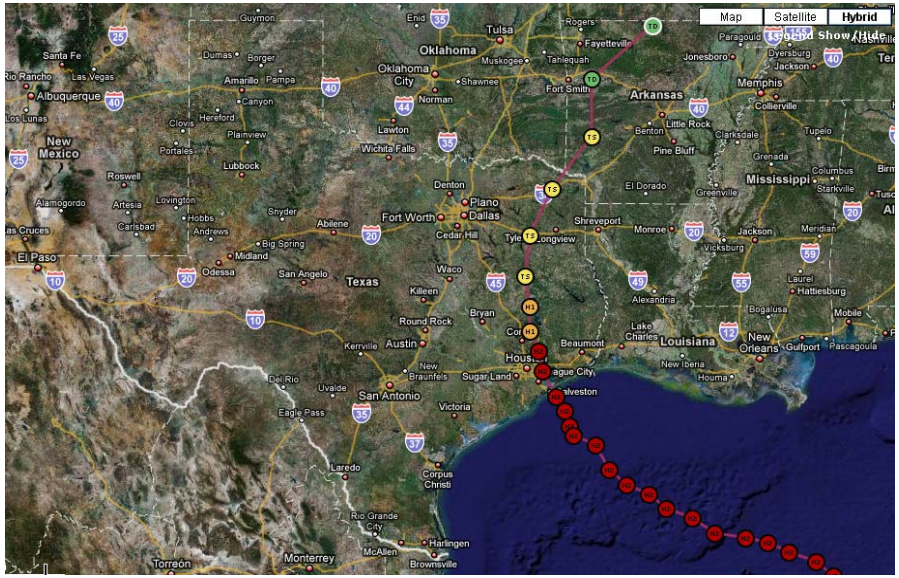Figure 3.10: Path of Hurricane Ike over Texas (source: http://stormadvisory.org/map/atlantic).
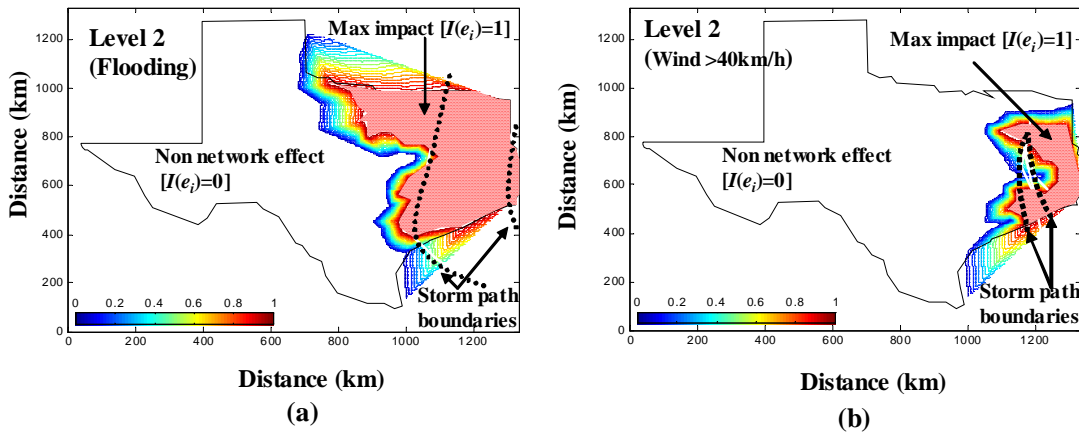

Figure 3.11: Network impact index at level 3. Spatial distribution of Ike considering: (a) flooding; and (b) wind speed.
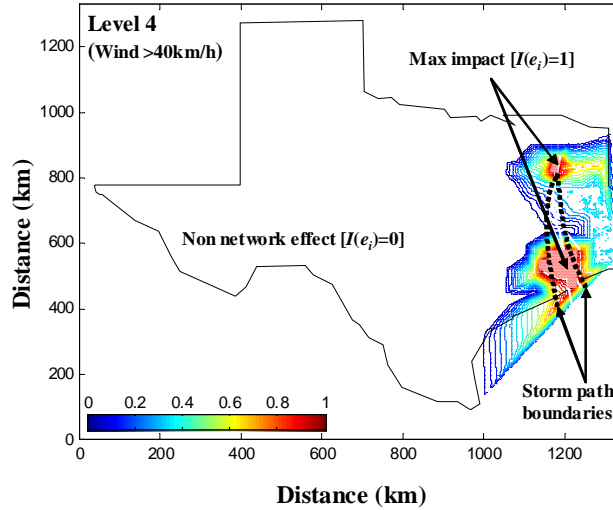
Figure 3.12: Impact index for winds at level 4.

which the analysis is carried out because they result from a direct evaluation of losses in nodes located in the event path. The evaluation of indirect losses is, however, the most important aspect of the model since it captures the network effect. As indirect losses two values were measured: 1) population affected; and 2) loss of productivity.

The estimates of people affected by the flooding and wind speed are shown in Figure 3.13a. The expected number of people affected by flooding (level 2: $[6.2, 12.8]$ million; and level 6: $[4.4, 9.1]$ million) is larger than the number of people affected by wind (level 2: $[4.6, 5.8]$ million; and level 6: $[3.7, 4.6]$ million). It can be observed that if the estimation is made in a lower level, the uncertainty (range of possible values) decreases from 6.6 at level 6 to 4.7 million at level 2 for flooding and from 1.2 to 0.9 million for wind. This is explained by the fact that there is a better knowledge of the network interrelations in the lower level of the hierarchy (smaller clusters). In the assessment, the contribution of nodes outside the event path decreases from 46.8% in level 2 to 25% in level 6 for the case of high speed winds and from 79% to 73% for flooding (Figure 3.13b). Then, estimates made only from nodes directly affected will significantly underestimate actual values.

The estimation of productivity loss included the assessment of two measures: the GDP and the percent loss in the Texas Industrial Productivity Index (TIPI), which measures the output of the manufacturing, mining, and utility sectors of the Texas economy (FRBD, 1989 [20]). The loss in productivity was roughly estimated by a function describing the number of days off work or days with no productivity. According to the Federal Reserve Bank of Dallas, manufacturing of nondurable goods and mining indicators of the TIPI fell 0.23% and 0.47%, respectively, in September. Since they have shown a steady growing trend during last months, it is therefore reasonable to assume that most of this fall might have been caused by the impact of Ike. Figure 3.13b shows the contribution of the network components located outside the event's area of influence, at level 2, on the impact index factor. This means that in the case of wind damage, 21% of the loss in TIPI comes from network nodes located within the storm path and 79% from locations outside the storm path. For the case of flooding, these values are 47% and 53%, respectively. On the other hand, the impact over productivity in terms of percentage of the GDP lost, assuming proportionality to population, is $[0.14, 0.29]$ for level 2 and $[0.10, 0.21]$ for level 6 in the case of flooding and $[0.07, 0.14]$ for level 2 and $[0.05, 0.11]$ for level 6 in the case of wind.
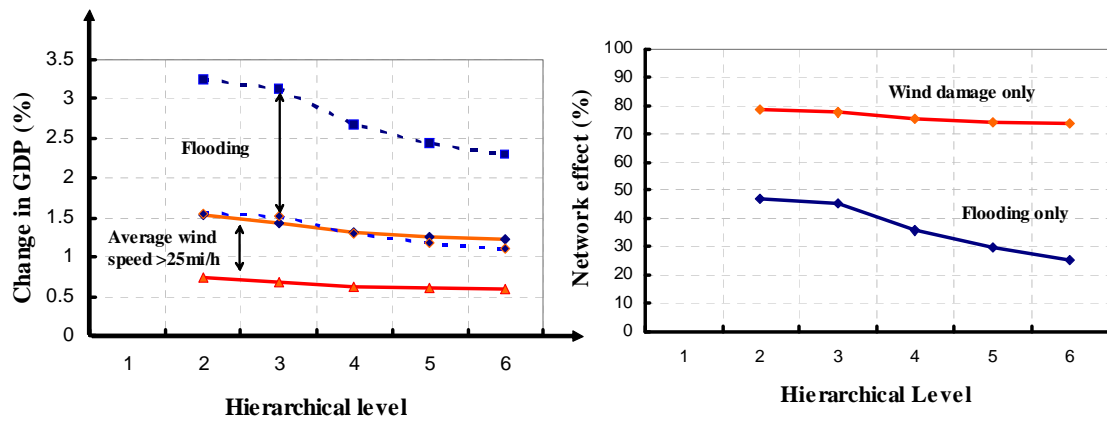
Figure 3.13: (a) Boundaries defining people affected for both flooding and wind speed; and (b) Contribution of the network elements outside the storm path to the estimation of people affected and loss of productivity.

Chapter 4

Probabilistic modeling of deteriorating systems

4.1 Introduction

The main objective of this chapter is to provide a comprehensive probabilistic model to study the time dependence performance of infrastructure components (Figure 1.1). It integrates the uncertainty about the occurrence time and damage extent caused by extreme loads and the damage accumulation as a result of progressive deterioration. Probabilistic models and stochastic processes are used to improve on traditional life-cycle cost analysis and reliability based design optimization. Although the proposed model has applicability in many areas, the emphasis of this paper is primarily on civil engineering infrastructure systems.

The specific objectives of this paper are:

1. develop a stochastic model of the life-cycle performance of infrastructure systems;

2. model the time-dependent performance of structures subject to multiple deterioration processes; and

3. improve existing reliability-based design and operation (maintenance) models of infrastructure components.

A review of the state of the art of deteriorating systems in various engineering areas is presented in section 4.2. Section 4.3 presents an overall description of the problem, the basic considerations and goals of the proposed model. Section 4.4 focuses on developing a probabilistic model of extreme events only. Progressive deterioration is included in the model in section 4.5. In section 4.6 the model for total system failure is presented and discussed. Finally the problem of life-cycle cost analysis and reliability-based design optimization is presented in section 4.7.

4.2 Structural deterioration

In practice, most infrastructure systems deteriorate as a result of the action of both sudden extreme events (i.e., shocks) and continuous progressive degradation caused mainly by aging and environmental factors. They both cause damage to accumulate with time stripping life units from the system. Earthquakes and hurricanes are examples of extreme events, while corrosion and fatigue are typical cases of progressive degradation. In civil engineering infrastructure systems, life units are described by a system performance measure (e.g., displacement or drift). Structural deterioration can be managed with a variety of intervention measures, which can be grouped into preventive maintenance (i.e.,

intervention to avoid total failure) and full reconstruction. The need for an intervention is defined by performance thresholds, which are prescribed in design codes or operation manuals. Then the main goal in infrastructure management is to maintain the system operating in acceptable conditions at a minimum cost. This objective is achieved by a cost-based optimization that requires a balance among the economic investment, the benefits derived from the existence of the project and the consequences in case of failure. In other words, designing, constructing and maintaining infrastructure may be viewed as a decision problem in which the maximum economic benefit of the life-cycle of the project is achieved while the reliability requirements are fulfilled simultaneously at the decision point (Rackwitz, 2000 [55]).

Progressive deterioration is a reduction of the structural capacity caused mainly by chloride ingress, which usually leads to steel corrosion, loss of effective cross-section of steel reinforcement in RC structures, concrete cracking, loss of bond and spalling. The details of these processes are beyond the scope of this paper but are well described by Val and Stewart, 2005 [65] and Liu and Weyers, 1998 [39]. Progressive deterioration has been managed traditionally by using the reliability index profile, which describes the change of the reliability index with time. Simplified degradation models have been proposed by Ellingwood and Mori, 1993 [19], Mori and Ellingwood, 1994 [43], Frangopol et al., 2004 [24], Petcherdchoo et. al., 2004 [54] and Cinlar et. al., 1977 [13]. Many papers have been published proposing models to keep track of structural deterioration with time and to define optimal intervention policies. Commonly, these models are based on Markov Decision Processes (MDP) (Harper et al., 1990 [28], Gopal and Majidzadeh, 1991 [25], Kleiner, 2001 [34], Mishalani and Madanat, 2002 [42], Guillaumot et al., 2003 [26], and Kubler and Fabber, 2003 [35]), Bayesian probability (Pandey, 1998 [50], Jian and Xila, 2005 [31]), renewal theory (Rackwitz, 2002 [55]) and approximate functions obtained from experimental data (Mori and Ellingwood, 1994 [43]). A review of common probabilistic models for life-cycle performance of deteriorating structures can be found in Frangopol et. al., 2004 [24].

Deterioration caused by extreme events is usually associated to earthquakes, hurricanes or blasts (terrorists attacks). In these cases, the structure is subject to randomly occurring shocks, and each shock causes a random amount of damage. Most shock maintenance and failure models are based on a control-limit policy, in which limit state violations (requirement for an intervention) are carried out once the accumulated damage exceeds a critical value or at failure, whichever occurs first (Valdez-Flores and Feldman, 1989 [66]). Extensive research has been carried out on mathematical models for shock degradation (Barlow and Proschan, 1965 [6], Sherif and Smith, 1981 [63], Aven and Jensen, 1999 [5], Taylor, 1975 [64], Nakagawa, 1976 [44], Feldman, 1976 [21], Feldman, 1977 [22], Feldman, 1977 [23], and Zukerman, 1977 [72]). Although this problem has been discussed extensively in civil engineering-related problems, only a few analytical solutions have been proposed within the context of structural optimization and life-cycle cost analysis. The first works on this topic were published by Resemblueth and Mendoza, 1971 [58], Hasofer, 1974 [29] and Rosemblueth, 1976 [59] in the context of earthquake resistance design optimization. Their ideas were reconsidered by Rackwitz, 2000 [55], to propose a general framework for optimal design and reliability verification. Rackwitz considered two main possible structural operation policies: structures abandoned after first failure and structures reconstructed systematically. Furthermore, for each case three loading conditions were considered: 1) time invariant loads; 2) extreme overloads (performance under Poisson perturbances); and 3) failures described by Poisson perturbances. Rackwitz uses renewal theory and takes advantage of the form of the discount function to achieve a rather practical (time-independent) solution to the expected life-cycle cost of a structure. The merit of the solution for random failures in time with systematic reconstruction (e.g., seismic design case) is that it does not depend on a specific lifetime of the structure, which is a random variable very difficult to quantify and usually underestimated by codes of practice. The solution is based on failure intensities and not on time dependent failure probabilities. It is not necessary to define arbitrary reference times of intended use or to compute first passage time distributions. Similarly, Wen and Kang, 2001 [71] developed a model to minimize the life-cycle cost with respect to the design load and resistance. The random occurrence and intensity variation of the hazards in time is described by a simple random process. The model also includes several possible damage states after an event with the corresponding probability of reaching every damage state.

After a thorough review of the models briefly outlined, it can be concluded that a comprehensive life-cycle model should account for the following aspects:

1. structural deterioration as a combination of both progressive and sudden damaging events;

2. multiple damage states (not only the failure and not-failure states);

3. the role of deterioration history in estimating the structural condition at a given time;

4. after failure, structural reconstruction does not necessarily take the component to its initial condition; and

5. the selection of acceptable operational and failure thresholds that is critical in life-cycle analysis.

4.3   General deterioration model

Consider a structural component with an initial remaining life u_0 (Figure 4.1). As deterioration increases, the remaining life of the component decreases. Then if $D(t)$ describes the deterioration of the component at time t; the *remaining life* of the component can be expressed as:

$$V(t) = u_0 - D(t) \tag{4.1}$$

An intervention (maintenance or reconstruction) is carried out once the remaining life of the system reaches a threshold value. Two possible threshold values (limit states) can be identified. The first one corresponds to the minimum level of performance of the system a level below which the system cannot be in service under any circumstance; this does not necessarily implies collapse. This is a fixed and predefined value represented by s^* in Figure 4.1. The second threshold is an operation threshold. This value (k^* in Figure 4.1) describes the minimum acceptable operation level defined, for instance, by a regulatory agency. In other words, the system can still operate below this level but the operation will not be considered satisfactory, efficient or acceptably safe. In summary, the system will be in acceptable operation ("on") as long as the remaining life is larger than a given value k^* in Figure 4.1 and the system will be out of service ("off") once the remaining life falls below k^* and until it is repaired and taken to a new "initial" remaining life value $u > k^*$. The length of time in which the system remains "off" is the time required to maintain and/or retrofit the structure. Finally, after the intervention, the condition of the structure will reach a new value corresponding to the remaining life that might differ from the value at the beginning of the previous cycle.

Consider a structural component with a sample path as the one shown in Figure 4.1. Assuming that both continuous and sudden damaging events are independent, the deterioration at time t, for the first cycle can be computed as:

$$D(t) = \int_0^t r_p(\tau)d\tau + \sum_{i=1}^{N_t} Y_i \tag{4.2}$$

where Y_i is the loss of remaining life caused by shock i and $r_p(t) > 0$ describes the rate of progressive deterioration. Note that the term *remaining life* used here does not necessarily describe time, but more realistically, a loss of capacity expressed in physical units (e.g., structural drift, stiffness).

A system fails completely when the accumulated damage reaches a predefined damage threshold s^*. In most structures the threshold s^* corresponds to the ultimate structural capacity and therefore this failure type can be assimilated to structural collapse. Nevertheless, maintenance programs are based on early repairs, that is, preventive repairs carried out before the structure collapses. Preventive maintenance is conditioned on a level of damage k^* with $k^* > s^*$. Note that failure and preventive maintenance are mutually exclusive since for any instantaneous event that takes the accumulated damage below the threshold k^*, only one of these two cases is possible. Information about the structural
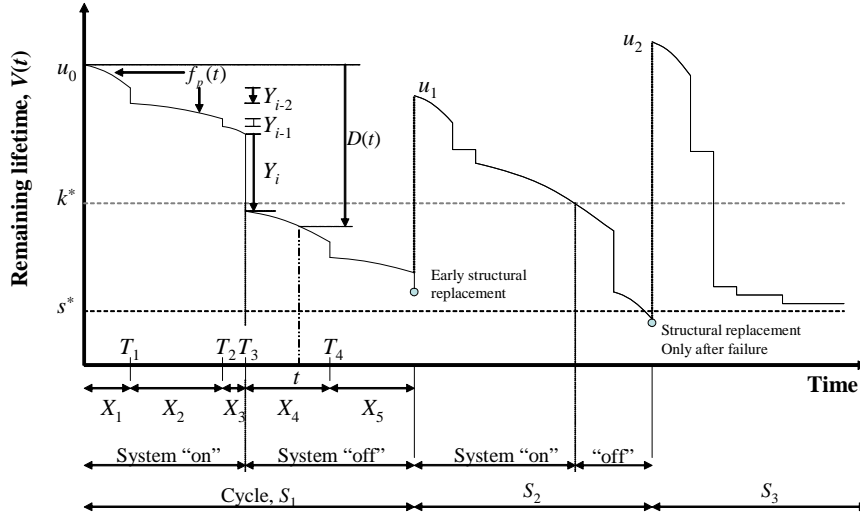
Figure 4.1: General description of the system remaining lifetime.

remaining life is important in decision-making. For example, after a shock (e.g., earthquake) the operator could make decisions based on the following criteria:

1. if the remaining life level is below threshold s^*, the structure is fully replaced immediately;

2. if the remaining life is between k^* and s^*, preventive maintenance is carried out; and

3. if the remaining life is larger than k^* the structure is not modified (Figure 4.1).

The derivations that will be presented in this document can be used to respond to questions such as: what is the probability that the remaining life of the system falls below a given level z before time t? What is the average availability of the system? What are the optimum design and operation policies? And what is the remaining life of the system at time t? Information from the model can also be used to develop fragility curves and sensitivity analysis of various design and operation parameters.

4.4   Effect of shocks

The proposed model for structural deterioration as a result of only successive shocks is based on the following assumptions:

1. damage accumulates as a consequence of consecutive shocks;

2. the loss of remaining life describes "structural damage" as a result of the shock, not the shock's "intensity";

3. the size of shocks (damage) are independent and identically distributed;

4. all replacements and repairs return the system to *good as new*, i.e., statistically identical cycles; and

5. repairs and replacements are instantaneous; therefore, the system does not remain "off" at any time.

Figure 4.2: Relationship between the accumulated effect of shocks and the system cycles.

4.4.1 Remaining lifetime

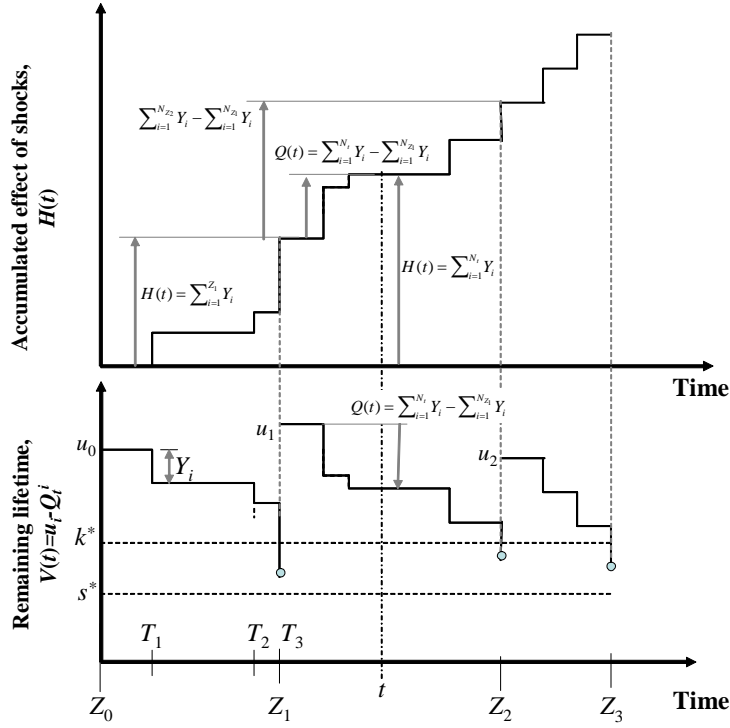For extreme events, the damage at shocks is described by the point process $Y = \{Y(t), t > 0\}$ in which the interarrival times of successive sudden events (i.e., shocks Y_i) is a sequence of nonnegative, independent, random variables $X_1, X_2, \ldots$ with common distribution $F(t) = p(X \leq t)$ (Figure 4.1). If the system is not intervened, the accumulated damage at a given time t is

$$H_t = \sum_{i=1}^{N_t} Y_i \tag{4.3}$$

where $H(t) = H_t$, N_t is a random variable describing the total number of shocks by time t (Figure 4.2); Y_i is the amount of damage produced by the ith shock. Y_i, $i = 1, 2, \ldots$ are positive, independent and identically distributed random variables. Further, it is assumed that the distribution of Y_i does not depend upon the previous states of the system.

The lifetime of the system consists of successive independent and stochastically identical cycles. One cycle starts with the system in a condition *as good as new* and it deteriorates as a result of shocks; once the system fails, immediate reconstruction (or maintenance) is carried out and a new cycle starts. If Z_i describes the ith replacement time (cycle), the total amount of shocks by time t in the ith cycle can be computed as:

$$Q^i(t) = \begin{bmatrix} \sum_{j=1}^{N_t} Y_j & i = 1 \\ \left[\sum_{j=1}^{N_t} Y_j - \sum_{j=1}^{N_{Z_{i-1}}} Y_j \right] \cdot \mathbf{1}_{\{Z_i < t < Z_{i+1}\}} & i > 1 \end{bmatrix} \tag{4.4}$$

where the second term on the right hand side of equation 4.4, for $i > 1$, is the history of demands up to the last replacement (Figure 4.2). Note that the accumulation of damage is defined by statistically identical cycles. The remaining lifetime of the system $V(t)$ defined through $\{H_t, Z_{\ell(t)}\}$ is a regenerative process, implying that the condition of the system depends on the its history only through the state
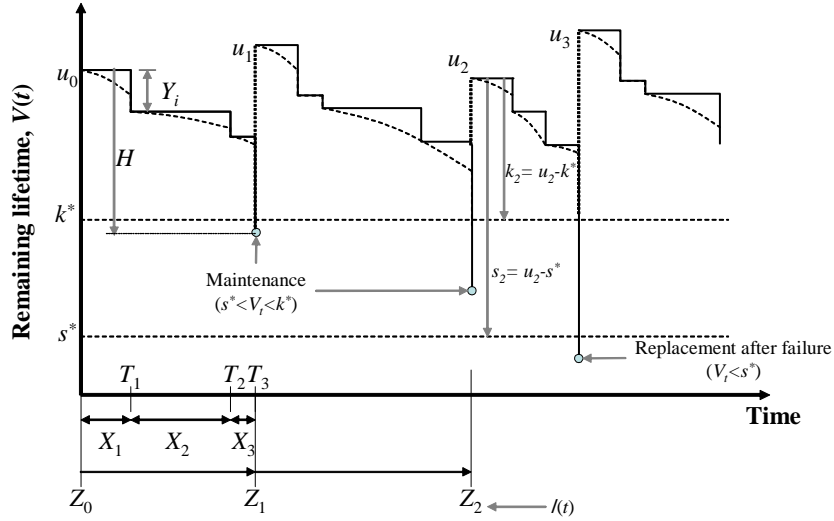
41

Figure 4.3: Structural maintenance and structural replacement after collapse.

at the end of the previous cycle. At the beginning of every cycle i the remaining lifetime is reset to a random value u_{i-1}. Therefore, the loss of remaining lifetime at a given cycle i is computed by substracting the accumulated damage caused by shocks from u_{i-1}; this is (Figure 4.2),

$$V_i(t) = u_{i-1} - Q_t^i; \quad i = 1, 2, 3, ... \tag{4.5}$$

where $Q_t^i = Q^i(t)$.

4.4.2 Probabilistic model of interventions

Define a probability space $(\Omega, \Im, P)$ with respect to which all random variables that will be used in this model are measurable. Further, define the counting process $\{\ell(t), t \geq 0\}$ representing the number of interventions (i.e., early or full replacement) by time t, and a cycle as the time between any two interventions. The performance of the model throughout its lifetime is defined by two thresholds, k^* and s^* (Figure 4.3). Then, the structural life ranges that define failure or preventive maintenance in a cycle (Figure 4.3) can be generalized as: $s_{\ell(t)} = u_{\ell(t)-1} - s^*$ and $k_{\ell(t)} = u_{\ell(t)-1} - k^*$; where $u_{\ell(t)-1}$ is the remaining lifetime of the system at the beginning of the $\ell(t)^{th}$ cycle. In the $\ell(t)^{th}$ cycle, any intervention takes place every time the remaining life falls below k^*, i.e., $V_{\ell(t)}(t) < k^*$; if the structure is not abandoned, it is put back in service immediately.

Consider the event *the system is intervened*, which means that either *the system is reconstructed after failure* or *preventive maintenance is carried out*. Then interventions are modeled as a first passage problem defined in terms of the *instantaneous intervention rate*. This is the probability that if the system has survived for a time t (i.e., has not been intervened), it will not survive an additional time dt; this is also commonly referred to as the *hazard function*. The hazard function, expressed as the rate $\lambda(t)$, is a renewal counting process (Bremaud, 1981 [7], Aven and Jansen, 1999 [5]):

$$\lambda(t) = \sum_{n \geq 0} \frac{f(t - T_n)}{1 - \int_0^{t-T_n} f(x)dx} \cdot \mathbf{1}_{\{T_n < t \leq T_{n+1}\}} \tag{4.6}$$

where $T_n = \sum_{i=1}^n X_i$ is the time to the latest shock before t. At this point, it is assumed that interventions occur only at shock times. The rate defined in equation 4.6 does not provide information

about the remaining life of the system. Since the magnitude of damage and the occurrence times are independent, the probability that an intervention is required at time t will be:

$$dP(\text{intervention at time } t) = \lambda(t)p(Q^{\ell(t)}(t) > u_{\ell(t)-1} - a^*) \tag{4.7}$$

where a^* takes the value of k^* for preventive maintenance and of s^* for failure. The expression $dP(.)$ describes the derivative of the probability. Further, if $dG(y)$ is the probability of having a shock size (i.e., damage size) between y and $y + dy$, then,

$$dP(\text{intervention at time } t) = \int_{V^*_{\ell(t)}(t,a^*)}^{\infty} \lambda(t)dG(y) \tag{4.8}$$

where $V^*_{\ell(t)}(t, a^*) = V_{\ell(t)}(t) - a^*$.

Let's define $d\ell(t) = \ell(t + dt) - \ell(t)$; with $\ell(t) = 1, 2, ...$ being the cycle in which the system is at time t. Then, $d\ell(t) = 1$ if there is an intervention and $d\ell(t) = 0$ otherwise. Furthermore, if $\mathcal{L}_t$ is the process history up to t, the *instantaneous intervention rate* can be written for all cases considered in this paper as:

$$P(d\ell(t) = 1 \mid \mathcal{L}_{t-}) = \begin{bmatrix} \int_{V^*_{\ell(t)}(t,s^*)}^{\infty} \lambda(t)dG(y) \cdot \mathbf{1}_{\{Z_{\ell(t)-1} < t \le Z_{\ell(t)}\}} & \text{Failure} \\ \int_{V^*_{\ell(t)}(t,k^*)}^{V^*_{\ell(t)}(t,s^*)} \lambda(t)dG(y) \cdot \mathbf{1}_{\{Z_{\ell(t)-1} < t \le Z_{\ell(t)}\}} & \text{Prev. Maintenance} \\ \int_{V^*_{\ell(t)}(t,k^*)}^{\infty} \lambda(t)dG(y) \cdot \mathbf{1}_{\{Z_{\ell(t)-1} < t \le Z_{\ell(t)}\}} & \text{Any intervention} \end{bmatrix} \tag{4.9}$$

where t^- indicates that the regenerative process is right continuous. The limits of the integral in every case define the range of shock sizes that may cause the intervention (failure or maintenance). The complexity of the solution for $P(d\ell(t) = 1 \mid \mathcal{L}_{t-})$ lies on computing the limits of the integral which depend on the number of shocks in the cycle and the distribution of the sum of shock sizes (i.e., accumulated damage). The limits of the integral can be defined, for any cycle, as:

$$V^*_{\ell(t)}(t, a^*) = V_{\ell(t)}(t) - a^* = u_{\ell(t)-1} - a^* - Q^{\ell(t)}(t) \tag{4.10}$$

Rewriting equation 4.10,

$$V^*_{\ell(t)}(t, a^*) = \begin{bmatrix} u_0 - a^* - \sum_{j=1}^{N_t} Y_j & \ell(t) = 1 \\ u_{\ell(t)-1} - a^* - \left[\sum_{j=1}^{N_t} Y_j - \sum_{j=1}^{N_{Z_{i-1}}} Y_j \right] & \ell(t) > 1 \end{bmatrix} \tag{4.11}$$

If $G(y) = p(Y \le y)$ is the distribution function of the damage caused by any shock and $G^{(j)}(y)$ the j-fold Stieltjes convolution of $G(y)$ with itself, then by conditioning on the number of shocks:

$$V^*_{\ell(t)}(t, a^*, n) = \begin{bmatrix} \int_0^{u_0-a^*} (u_0 - a^* - y)dG^{(n)}(y) & \ell(t) = 1, G^{(n)}(y) \le (u_0 - a^*) \\ \int_0^{u_{\ell(t)-1}-a^*} (u_{\ell(t)-1} - a^* - y)dG^{(n)}(y) & \ell(t) > 1, G^{(n)}(y) \le (u_{\ell(t)-1} - a^*) \end{bmatrix} \tag{4.12}$$

Note that $V^*_{\ell(t)}(t, a^*, n)$ has been described not only as a function of time but also as a function of the number of shocks since it has to account for the accumulated damage.

4.4.3 System abandoned after first failure

This section considers the case of a structure exposed to successive shocks until the remaining life falls below a prescribed threshold value. Once this happens, it is abandoned; the system is not reconstructed or intervened in any way. Note that for this particular case $a^* = s^* = k^*$. Under these assumptions, it is necessary to only model the first cycle and equation 4.9 specializes to:

$$P(d\ell(t) = 1 \mid \mathcal{L}_{t-}) = \int_{V^*_1(t,a^*,n)}^{\infty} \lambda(t)dG(y) \tag{4.13}$$
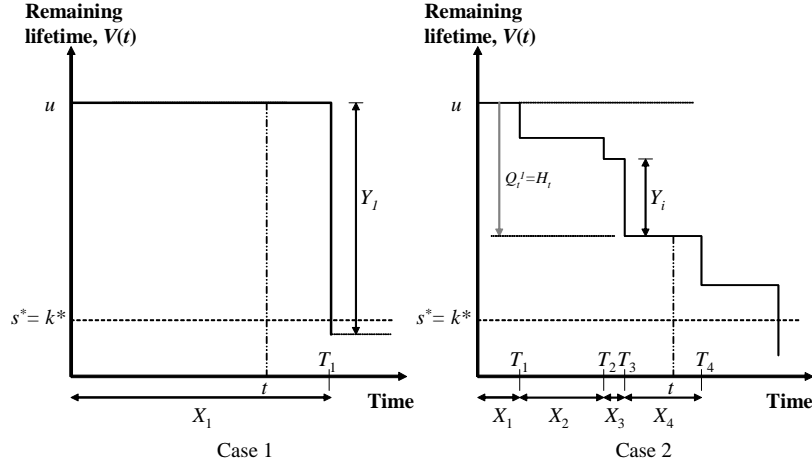
Figure 4.4: Failure cases studied: 1) single shock failure; and 2) failure after a sequence of random shocks.

The lower limit of the integral is a function of the initial remaining life and the shock history; then,

$$V_1^*(t, a^*, n) = \int_0^{u_0 - a^*} (u_0 - a^* - y) dG^{(n)}(y); \text{ for } G^{(n)}(y) \le (u_0 - a^*) \tag{4.14}$$

By removing the condition on the number of cycles, the failure rate can be rewritten as:

$$P(d\ell(t) = 1 \mid \mathcal{L}_{t^-}) = \sum_{n=0}^{\infty} \left(\int_{V_1^*(t, a^*, n)}^{\infty} \lambda(t) dG(y) \right) P(N = n, t) \tag{4.15}$$

The probability that an intervention is required before or at time t is given by:

$$\Lambda(t) = \int_0^t \left(\sum_{n=0}^{\infty} \left(\int_{V_1^*(\tau, a^*, n)}^{\infty} \lambda(\tau) dG(y) \right) P(N = n, \tau) \right) d\tau \tag{4.16}$$

where $p(N = n, t)$ is the probability of having n shocks by time t, and N is the number of shocks.

Illustrative example 1 Consider the case of a structure exposed to earthquakes with magnitude $M > 4$ (those which can actually cause some damage) that follows a Poisson process with rate $\mu = 1/year$. In addition, consider a structure designed and built to meet a target performance defined by a vector parameter $\mathbf{u}$. For the purpose of this paper, the remaining life of the structure will be described as a percentage of $\mathbf{u}$. The structural remaining life will be distributed lognormally with $\mu = 100$ (percentage) and $COV = 0.25$. Assume further that the damage caused by an earthquake is governed by an exponential distribution $G(y, \theta)$ with $\theta = 0.05$ and that the intensity of damage as a result of a shock is not conditioned on the accumulated damage prior to the event. The cases that will be studied are shown in Figure 4.4.

The Poisson process used to model earthquake occurrences implies that interarrival times between earthquakes are exponential, therefore,

$$\lambda(t) = \frac{f(t - T_n)}{1 - \int_0^{t - T_n} f(x) dx} = \frac{\mu e^{-\mu(t - T_n)}}{1 - (1 - e^{-\mu(t - T_n)})} = \frac{\mu e^{-\mu(t - T_n)}}{e^{-\mu(t - T_n)}} = \mu \tag{4.17}$$

44

which reflects the memoryless property of the exponential distribution. The instant failure probability is:

$$P(V(t+dt) \leq s^* \mid V(t) > s^*) = \int_{V_1^*(t,s^*,n)}^{\infty} \lambda(t) dG(y) = \int_{V_1^*(t,s^*,n)}^{\infty} \mu dG(y) \tag{4.18}$$

If it is assumed that the initial remaining life is deterministic, i.e., $u = u_0$, for case 1 (Figure 4.4):

$$P(V(t+dt,u) \leq s^* \mid V(t) > s^*) = \mu \left[1 - \int_0^{u_0 - s^*} g(y) dy \right] = \mu[1 - G(u_0 - s^*)] \tag{4.19}$$

For the case when the initial remaining life is defined as a random variable it is necessary to uncondition equation 4.19 as follows:

$$P(V(t+dt) \leq s^* \mid V(t) > s^*) = \int_0^{\infty} \mu \left[1 - G(u_0 - s^*) \right] dF_U(u); \ s < u \tag{4.20}$$

where $dF_U(u)$ is the differential distribution function of the initial remaining life.

In case 2 (Figure 4.4) damage accumulates with time, and the lower limit of the integral in equation 4.18 becomes random and depends on the number of shocks, n, and their size (i.e., damage). The accumulated damage is described by the sum of the damage caused by every shock. For the particular case where the damage caused by every shock is exponentially distributed, the accumulated damage after n shocks follows an Erlang distribution,

$$dG^{(n)}(y,\theta) = \frac{1}{(n-1)!} \theta^n y^{n-1} \exp(-\theta y) \tag{4.21}$$

where x is the amount of damage and θ is the average damage size observed at every shock. Then:

$$V_1^*(t,s^*,n) = \int_0^{u-s^*} (u - s^* - y) dG^{(n)}(y) dy \tag{4.22}$$

Note that the term $u - s^*$ is the total remaining life available to the system, and y is the remaining life that has been taken from the system after n shocks (damage). In other words, $V_1^*(t,s^*,n)$ defines the minimum damage size required to cause the system's remaining life to fall below s^*. In case 2 (Figure 4.4), the solution is conditioned on the number of shocks and on the initial remaining life u. Then, by conditioning out these variables the final expression for the instant probability of failure becomes:

$$P(V(t+dt) \leq s^* \mid V(t) > s^*) \tag{4.23}$$

$$= \int_0^{\infty} \sum_{n=1}^{\infty} \left[\mu \left[1 - G(V_1^*(t,s^*,n)) \right] p(N=n,t) \right] f_U(u) du; \ s < u \tag{4.24}$$

where the probability of having n shocks by time t with exponential interarrival times is:

$$P(N=n,t) = \frac{(\theta t)^n \exp(-\theta t)}{n!} \tag{4.25}$$

Comparing cases 1 and 2 (Figure 4.5) it can be observed that case 2 (multiple shocks) leads to larger failure probabilities. This is expected since damage accumulates with time, and the probability of exceeding the threshold increases with time. In addition, equation 4.23 can be used to define time-dependent fragility curves by defining appropriate threshold values (i.e., s^*) that characterize damage states (Figure 4.6). Since thresholds (i.e., s^*) define damage states, instant and accumulated probability of intervention are larger for minor damage states and larger for severe damage states.
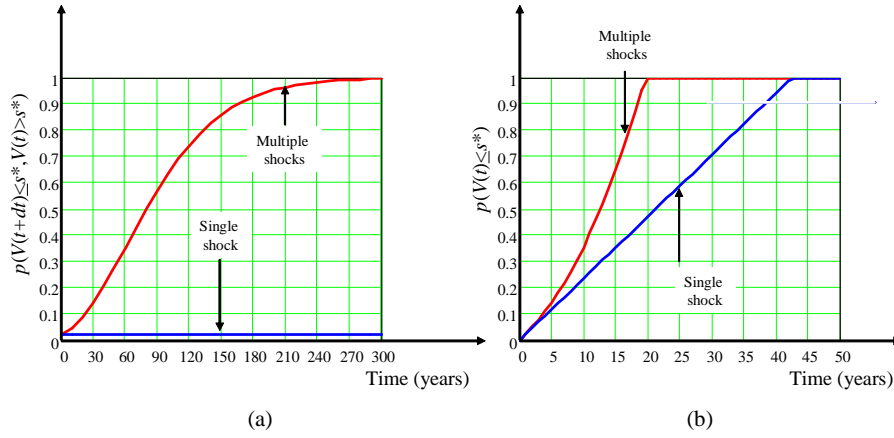
Figure 4.5: Comprison between (a) the instantaneous probability of failure caused by one and multiple shocks; (b) accumulated probability of failure caused by one and multiple shocks. The limiting condition is s* = 0.25 and $dF_U(u)$ is distributed lognormally.
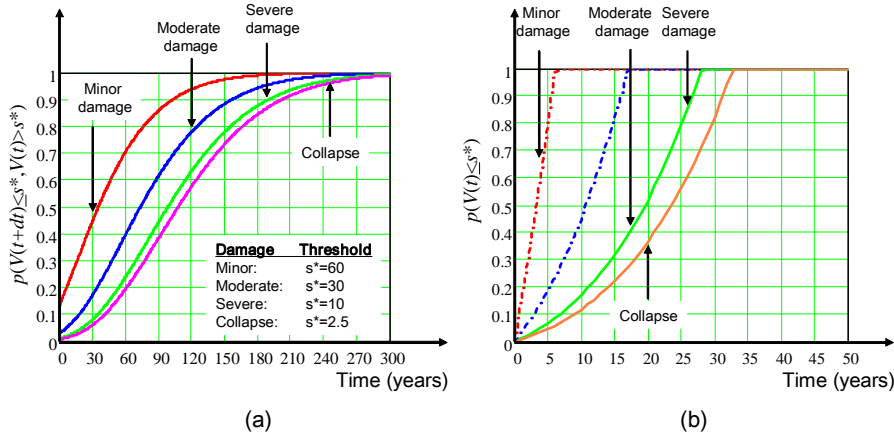
Figure 4.6: A structure subject to multiple shocks: (a) instantaneous probability of reaching a given damage state by time t; (b) accumulated probability of reaching a given damage state before or at time t.

46
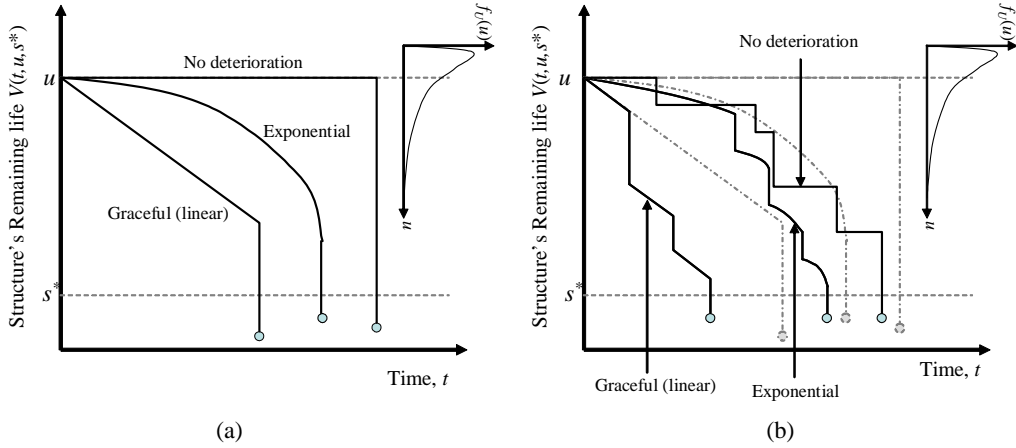
Figure 4.7: (a) Sample path for a single shock failure case and various deterministic degradation models; (b) sample path for the case of multiple shocks and various deterministic degradation models.

4.5 Progressive deterioration

Progressive degradation is usually a slow continuous time dependent phenomenon, which is caused for instance, by chloride ingress, corrosion, fatigue or biodeterioraton. Since there is little evidence of a strong correlation between progressive deterioration and damage as a result of shocks, independence will be assumed. A deterministic and a random model for deterioration are presented in this section.

4.5.1 Deterministic deterioration model

Figure 4.7 shows two deterioration models: linear and exponential and the combined effect of these deterioration models with the case of intervention after a single shock (Figure 4.7a) or after multiple shocks (Figure 4.7b). For the deterministic case, it is assumed that progressive degradation has a continuous positive rate $r_p(t) > 0$. Therefore, the remaining life of the structure at time t can be computed as:

$$V_i(t) \quad = \quad u_{i-1} - Q_t^i - \int_{Z_{i-1}}^t r_p(\tau)d\tau \tag{4.26}$$

$$= \quad u_{i-1} - Q_t^i - A_p(t - Z_{i-1}) \tag{4.27}$$

where the sub-index i describes the cycle the system is in. By including a progressive deteriorating function, the integration limits in equation 4.9 are modified. Progressive deterioration reduces the shock damage required to exceed the threshold value by reducing the remaining life of the system. The integral limits in equation 4.9 will change to:

$$V_{\ell(t)}^*(t, a^*) \quad = \quad V_{\ell(t)}(t) - a^* - A_p(t - Z_{\ell(t)-1})$$

$$= \quad u_{\ell(t)-1} - a^* - Q^{\ell(t)}(t) - A_p(t - Z_{\ell(t)-1}) \tag{4.28}$$

By making $A_p(t, Z) = A_p(t - Z_{\ell(t)-1})$ and introducing the Stieltjes convolution of $G(y)$ with itself,

$$V_{\ell(t)}^*(t, a^*, n) = \left[\begin{array}{ll} \int_0^{T_R} (T_R - y)dG^{(n)}(y) & \ell(t) = 1, G^{(n)}(y) \le T_R \\ \int_0^{T_{R_{\ell(t)}}} (T_{R_{\ell(t)}} - y)dG^{(n)}(y) & \ell(t) > 1, G^{(n)}(y) \le T_{R_{\ell(t)}} \end{array} \right] \tag{4.29}$$

where $T_R = u_0 - a^* - Q^{\ell(t)}(t) - A_p(t, Z)$ and $T_{R_{\ell(t)}} = u_{\ell(t)-1} - a^* - Q^{\ell(t)}(t) - A_p(t, Z)$.
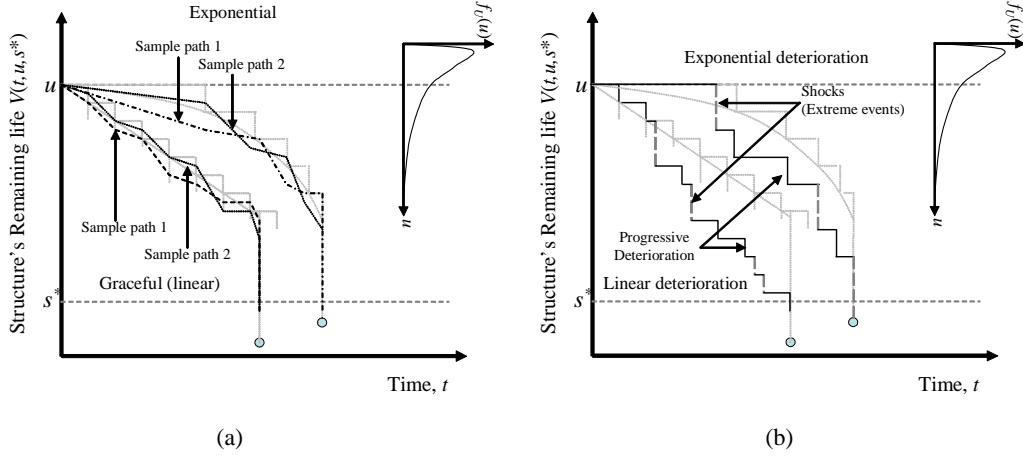
47

Figure 4.8: (a) Sample path of random progressive deterioration for linear and exponentially time-spaced shocks; (b) Shocks models used to simulate both extreme events and progressive deterioration.

4.5.2   Random progressive deterioration

The assumption that deterioration is a deterministic function works fine in many practical cases. However, if the uncertainty of the process needs to be included, a good approximation can be obtained by assuming that progressive deterioration is also a jump process in which the size of every jump is random and jumps occur at fixed time intervals. Figures 4.8a and 4.8b show in light grey lines the deterministic distribution of jumps in time. In Figure 4.8a, various sample paths of deterioration are presented. The combined effect of the jump model for progressive deterioration and the model of shocks is shown in Figure 4.8b.

Define D_i as the loss of remaining life caused by shock i, $G_D(x) = p(D \leq x)$ and the random accumulated damage caused by progressive deterioration as: $S_{\ell(t)} = \sum_{i=0}^{e} D_i$. The remaining life by time t becomes:

$$
\begin{aligned}
V_{\ell(t)}^*(t, a^*) &= V_{\ell(t)}(t) - a^* - S_{\ell(t)} & (4.30) \\
&= u_{\ell(t)-1} - a^* - Q^{\ell(t)}(t) - \sum_{i=0}^{e} D_i & (4.31)
\end{aligned}
$$

The integral limit for the first cycle is

$$
V_1^*(t, a^*, n, e) = \int_0^{u_0 - a^*} \left[\int_0^{u_0 - a^* - d} (u_0 - a^* - d - y) dG^{(n)}(y) \right] dS^{(e)}(d) \tag{4.32}
$$

where $dS^{(e)}$ is the stochastic differential of the distribution of the sum of e shocks describing the progressive deterioration process. Note that e is deterministic and known at every time t. For all cycles after the second,

$$
V_{u_{\ell(t)-1}}^*(t, a^*, e) = \int_0^{u_{\ell(t)-1} - a^*} \left[\int_0^{u_{\ell(t)-1} - a^* - d} (u_{\ell(t)-1} - a^* - d - y) dG^{(n)}(y) \right] dS^{(e)}(d) \tag{4.33}
$$

Note that for the case where progressive deterioration jumps are not random but deterministic with $A_p(t, Z) = D = t$, equation 4.32 becomes,

$$
V_1^*(t, a^*) = \int_0^{u_0 - a^* - t} (u_0 - a^* - t - y) dG^{(n)}(y) \tag{4.34}
$$

which is the same solution obtained for the deterministic case. If the occurrence times are governed by any deterministic function $g_d(t)$,

$$V^*_{\ell(t)}(t, a^*) = u_{\ell(t)-1} - a^* - Q^{\ell(t)}(t) - \sum_{i=0}^{N_{g_d(t)}} D_i \tag{4.35}$$

where $N_{g_d(t)}$ defines the number of jumps by time t. Note that the process rate (equation 4.6) will not be affected since the jumps associated to progressive deterioration occur at fixed times.

Illustrative example 2 Consider the same structure described in illustrative example 1 and assume that the system deteriorates with time. Initially consider that deterioration is a deterministic function $g_{d_1}(t) = t$ and $g_{d_2}(t) = \exp(\alpha t)$, with $\alpha > 0$. The limit state was defined by $s^* = 0.25$. Also consider only the structural performance during the first cycle. Under these assumptions, the integral limits defined by equation 4.29 become:

$$
\begin{aligned}
V^*_1(t, s^*, n) &= \int_0^{u_0 - s^* - g_{d_1}(t)} (u_0 - s^* - g_{d_1}(t) - y) dG^{(n)}(y) \\
&= \int_0^{u_0 - s^* - t} (u_0 - s^* - t - y) dG^{(n)}(y)
\end{aligned}
\tag{4.36}
$$

and

$$V^*_1(t, s^*, n) = \int_0^{u_0 - s^* - \exp(\alpha t)} (u_0 - s^* - \exp(\alpha t) - y) dG^{(n)}(y) \tag{4.37}$$

If randomness is introduced to the deterioration models by describing them as a jump process with fixed interarrival times, the integral limits defined in equation 4.32 become

$$V^*_1(t, s^*, n, e)_{g_{d_1}(t)} = \int_0^{u_0 - s^*} \left[\int_0^{u_0 - s^* - d_1} (u_0 - s^* - d_1 - y) dG^{(n)}(y) \right] dS^{(e_1)}(d_1) \tag{4.38}$$

$$V^*_1(t, s^*, n, e)_{g_{d_2}(t)} = \int_0^{u_0 - s^*} \left[\int_0^{u_0 - s^* - d_2} (u_0 - s^* - d_2 - y) dG^{(n)}(y) \right] dS^{(e_2)}(d_2) \tag{4.39}$$

For the random models and in both cases, i.e., failure after a single shock or multiple shocks, it was assumed that the damage at every jump is distributed exponentially and jumps are *iid*. The rate of jump size was assumed to be $\vartheta = 0.75$. Since jumps are exponentially distributed, $dS^{(e_2)}(d)$ will follow an Earlang distribution. For the case in which jumps are exponentially distributed in time, the distances between progressive deterioration evaluations (jumps) will get closer as time becomes larger. At the beginning the damage will be closer to the condition of no damage, and over time it will grow faster.

The integral limits for all deterministic and random cases are shown in Figure 4.9 (equations 4.36, 4.37, 4.38 and 4.39). It can be observed that by introducing randomness, the integral limits change reducing the remaining life available with respect to the deterministic case. For the conditions of this example this difference is particularly important for the case of linear deterioration.

Figure 4.10 presents the instant intervention rate for various performance models. It can be observed that when multiple shocks are taken into account, the instantaneous failure rate increases and grows smoother than when a single shock model is used. The accumulated probabilities of intervention by time t are shown in Figure 4.11 for the same cases shown in Figure 4.10. As expected, the accumulated probability of intervention increases when progressive deterioration is included. Also random deterioration models lead to larger intervention probabilities that the deterministic models.
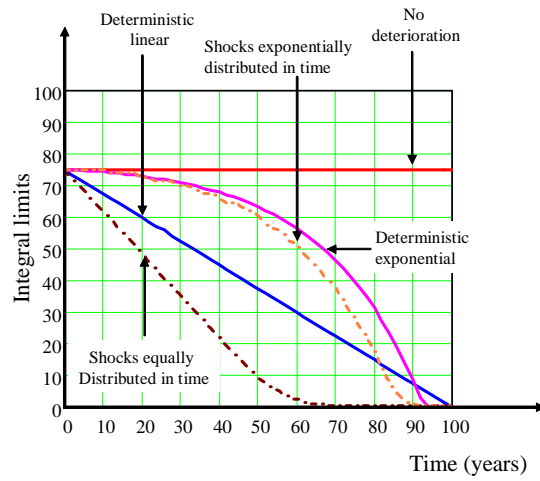
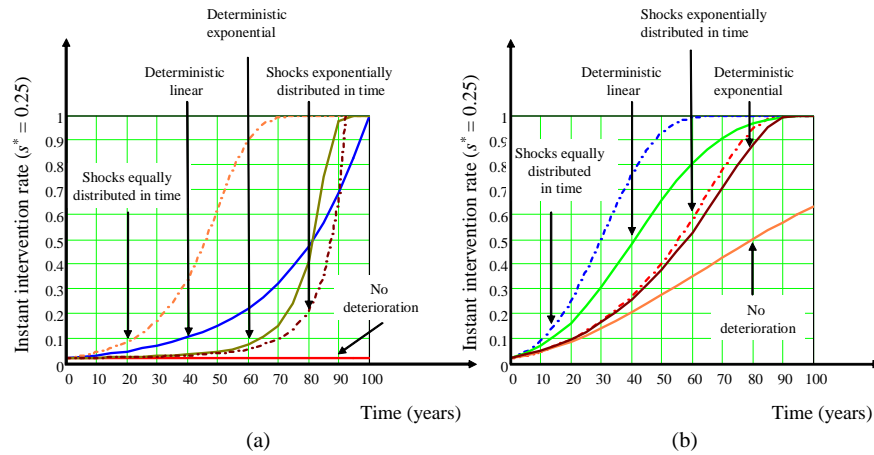Figure 4.9: Integral limits for all deterioration models considered.

Figure 4.10: Instant intervention rates for all models considered. (a) The structure is intervened after a single shock; and (b) the structure is exposed to multiple shocks.
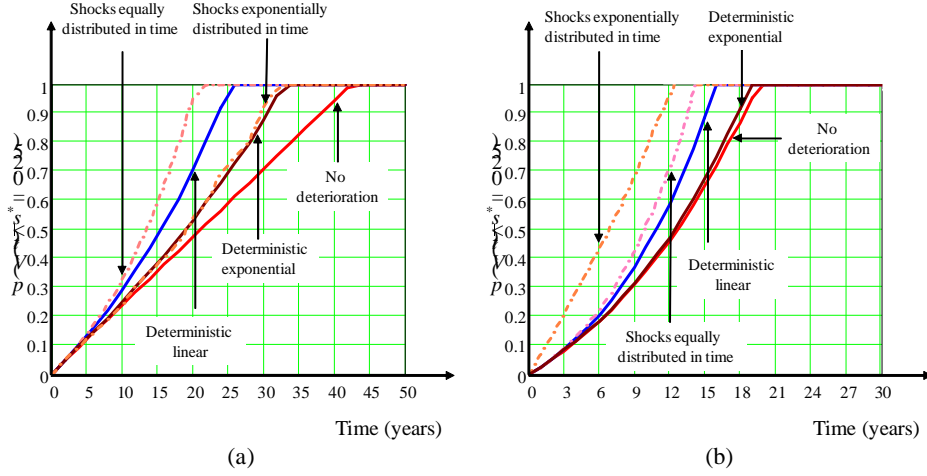
Figure 4.11: Probability of intervention by time t of a structure subject to (a) various deterioration models and (b) multiple shocks.

4.6 Total system failure

As discussed in previous sections, the system may fail as a result of shocks only, progressive deterioration or a combination of both. Although the loss of remaining life is a consequence of both processes, it is assumed that at the time of failure they will not occur simultaneously and that the actual failure is caused by only one of them. For simplicity, only the case for the first cycle will be described, but it can be easily extended to all other cycles.

The probability of having an intervention, i.e., that the loss of remaining life exceeds $u_0 - a^*$, was defined for all three intervention cases in equation 4.9 for the case of shocks only. If the random progressive deterioration model is included, the probability of having an intervention (equation 4.9) has to be modified to:

$$
\begin{aligned}
P(d\ell(t) \quad &= \quad 1 \mid \pounds_{t^-}) = \\
&\left[
\begin{array}{ll}
\left[\int_{V^*_{\ell(t)}(t,s^*)}^{\infty} \lambda(t)dG(y) + \int_{V^*_{\ell(t)}(t,s^*)}^{\infty} \delta(t)dS(y)\right] \cdot \mathbf{1}_{\left\{Z_{\ell(t)-1} < t \le Z_{\ell(t)}\right\}} & \text{(a)} \\
\left[\int_{V^*_{\ell(t)}(t,k^*)}^{V^*_{\ell(t)}(t,s^*)} \lambda(t)dG(y) + \int_{V^*_{\ell(t)}(t,s^*)}^{\infty} \delta(t)dS(y)\right] \cdot \mathbf{1}_{\left\{Z_{\ell(t)-1} < t \le Z_{\ell(t)}\right\}} & \text{(b)} \\
\left[\int_{V^*_{\ell(t)}(t,k^*)}^{\infty} \lambda(t)dG(y) + \int_{V^*_{\ell(t)}(t,s^*)}^{\infty} \delta(t)dS(y)\right] \cdot \mathbf{1}_{\left\{Z_{\ell(t)-1} < t \le Z_{\ell(t)}\right\}} & \text{(c)}
\end{array}
\right] \quad (4.40)
\end{aligned}
$$

where case (a) is failure, case (b) is preventive maintenance; and case (c) any intervention (failure of maintenance). Note that in equation 4.40 the second term in brackets accounts for the progressive deterioration. The parameter $\delta(t)$ can only be 0 or 1. It is 1 at times of "scheduled" fictitious shocks (describing the accumulated deterioration during a fixed period of time) and 0 otherwise.

If the structure is reconstructed after failure or intervened as a result of preventive maintenance, the process regenerates (Figure 4.3). This means that the cycle in which the structure is found at the time of evaluation, i.e., $\ell(t)$, becomes important in the assessment. Regenerative processes are only of interest in the case of damage accumulated with time until the system fails; otherwise, the problem can be managed using basic renewal theory.

The instantaneous intervention rate, conditioned on the cycle in which the system is found, can be

written as:

$$P(d\ell(t) = 1 \mid \pounds_{t-}) = \sum_{n=0}^{\infty} \left(\int_{V_{\ell(t)}^*(t,a^*,n)}^{\infty} \lambda(t)dG(y) \right) P(N = n, t) \tag{4.41}$$

In order to condition out the cycle from this solution two approaches can be taken: 1) find the actual analytical distribution of the length of a cycle and the number of cycles by time t; and 2) find a limiting solution ($t \to \infty$).

4.6.1 Analytical solution

Define the length of a cycle as $L_i = Z_i - Z_{i-1}$ and $\mathbb{Z}(z) = P(L_i \leq z)$. Therefore, $Z_m = \sum_{j=1}^m L_j$ and its distribution is defined by the mth convolution with itself: $Z_m \sim \mathbb{Z}^{(m)}(z)$. By assuming that the length of cycles is independent and identically distributed,

$$
\begin{aligned}
P(d\ell(t) &= 1 \mid \pounds_{t-}) \\
&= \sum_{m=0}^{\infty} \left[\int_0^t \sum_{n=0}^{\infty} \left(\frac{\int_{V_m^*(t,a^*,n)}^{\infty} \lambda(z)dG(y)}{P(N = n, t)} \right) d\mathbb{Z}^{(m)}(z) \right] P(M = m, t) \tag{4.42}
\end{aligned}
$$

where $M(t)$ is a counting process describing the number of cycles (interventions) and $P(M = m, t)$ is the probability that there have been m cycles by time t. The limit of the integral in equation 4.42 is:

$$V_m^*(t, a^*, n) = \int_0^{u_{m-1}-a^*} (u_{m-1} - a^* - y)dG^{(n)}(y); \text{ for } G^{(n)}(y) \leq (u_{m-1} - a^*) \tag{4.43}$$

It is important to stress that if the origin of the process does not coincide with the origin of a cycle, the distribution of length of the first cycle will be different from the distribution of the following cycles; therefore, some special considerations are needed in equation 4.42. The solution for equation 4.42 can become extremely complex and consequently, a better approach might be to take advantage of the limiting results for regenerative renewal processes.

4.6.2 Asymptotic solution

If in a renewal process the origin of the analysis coincides with the initiation of a cycle, the distribution function to the first renewal and the distribution of any other cycle are equal. These types of processes are called *ordinary* renewal processes. On the other hand, in a *delayed* renewal process, the origin does not coincide with the initiation of a cycle; therefore, the first renewal has a different distribution than the subsequent cycles. However, if the process has been running for a long time, the effects of the origin vanish as $t \to \infty$, and the delayed and the ordinary process converge asymptotically although the transient behavior is different.

For most existing structures, which have been in operation for a long time, the asymptotic behavior is (Cinlar, 1975 [12]):

$$\lim_{t \to \infty} P(V(t + dt) \leq s^* \mid V(t) > s^*) = \frac{1}{E[L]} \left[\left(\sum_{n=0}^{\infty} \left(\int_{V_1^*(t,a^*,n)}^{\infty} \lambda(y)dG(y) \right) P(N = n, t) \right) \right] \tag{4.44}$$

where $E[L]$ is the expected value of the length of a cycle. The length of one cycle is the expected time between interventions given that repair and reconstruction times are not significant with respect to the total life-cycle.

4.6.3 Renewal approach to expected length of a cycle

In order to compute $E[L]$, remember that the total accumulated damage in a cycle was defined in equation 4.3 as $H_t = \sum_{i=1}^{N_t} Y_i$ where N_t is the nth demand intensity that causes the system performance to fall below a^*; this is:

$$N_t = \min\left\{ n : \sum_{i=1}^{n} Y_i > u_{\ell(t)-1} - a^* \right\} \qquad (4.45)$$

In addition, if $X_i, i > 1$ are the interarrival times of the shocks, the expected operating time will be obtained by using Wald's lemma: $E\left[\sum_{i=1}^{N_t} X_i\right] = E[X]E[N_t]$; this is:

$$E\left[\text{Amount of time for which } V_{\ell(t)}(t) > a^*\right] = E\left[\sum_{i=1}^{N_t} X_i\right] \qquad (4.46)$$
$$= E[X]E[N_t] \qquad (4.47)$$

Furthermore, it can be proved that $E[T_{N(t)+1}] = \mu[m_G(t) + 1]$; where $T_{N(t)+1}$ is the time until the $N(t) + 1$ shock, $\mu = E[X]$ and $m_G(t) = E[N_t]$ is the renewal function (for more details see [60], [12] and [5]), which can be computed as follows:

$$m_G(t) = \sum_{n=1}^{\infty} G^{(n)}(t) \qquad (4.48)$$

If all cycles are independent and identically distributed: $L = L_1 = L_{\ell(t)}$, and

$$E[L \mid u] = E[X]\left(m_G(u - a^*) + 1\right) \qquad (4.49)$$

where $E[L]$ is conditioned on the remaining life at the beginning of every cycle, u.

4.6.4 Distribution function of the time to failure in a cycle

A direct way of computing the expected length of a cycle is to find the distribution function of the time between regenerations. Assuming independence between failures and disturbances, the density of the time to n^{th} shock can be computed in terms of the time required for the remaining life to fall below s^*. For the general case (i.e., shocks and deterministic progressive deterioration), this is:

$$P(V(t) \le s^*) = \sum_{n=0}^{\infty} \left[P\left(u - A_p(t) - s^* \le \sum_{i=1}^{n} Y_i \right) \right] P(N = n, t)$$

which can be written as:

$$F_V(s^*, t) = P(V(t) \le s^*) \qquad (4.50)$$
$$= \sum_{n=0}^{\infty} \left(\int_{V(n,t,s^*)}^{\infty} dG^{(n)}(y) \right) P(N = n, t) \qquad (4.51)$$

where $V(u, t, s^*)$ was defined in equation 4.12 for the case of shocks only and in equations 4.29 and 4.32 for progressive deterioration. Then,

$$E[L] = \int_0^{\infty} (1 - F_V(s^*, t))dt \qquad (4.52)$$

Illustrative example 3 In example 1, earthquake damage was modeled by a compound Poisson process with a damage rate $\theta = 0.05$ and earthquake interarrival times exponentially distributed, $E[X] = 1/\mu$, (μ is the earthquake rate). For the case of failure after the first shock (Figure 4.4a), $E[L] = E[X] = 1$, and the limiting results are exactly the same as the solution presented in equation 4.19 (Figure 4.12). If the structure is subject to multiple shocks (damage accumulation process) the
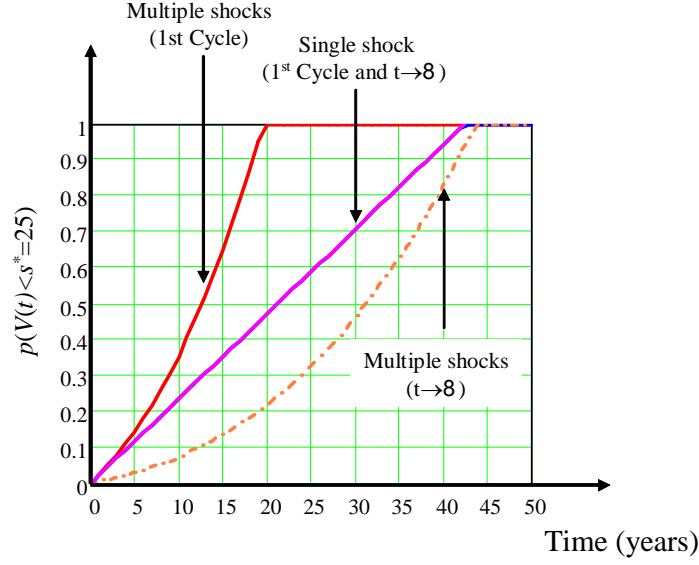
Figure 4.12: Comparison of accumulated structural probability of intervention of one cycle and the limiting result; both for the case of a structure subject to multiple shocks and for a structure that fails after the first shock.

solution is obtained by evaluating equation 4.44. It is then necessary to compute the expected length of the cycle (equation 4.49). Given that damage is exponentially distributed, the renewal function can be defined as:

$$M_G(T) = \sum_{n=1}^{\infty} \int_0^T \left[\frac{1}{(n-1)!} \theta^n t^{n-1} \exp(-\theta t) \right] dt \tag{4.53}$$

For $a^* = s^* = 25$ and $u = 100$ (deterministic),

$$E[L] = E[X](M_G(u - a^*) + 1) \tag{4.54}$$

$$= 1 \cdot (3.75 + 1) = 4.75 \text{ years.} \tag{4.55}$$

Therefore,

$$\lim_{t \to \infty} P(V(t) \leq s^*) = \frac{1}{4.75} \left[\int_0^t \left(\sum_{n=0}^{\infty} \left(\int_{V_1^*(t,a^*,n)}^{\infty} \lambda(z) dG(y) \right) P(N = n, t) \right) dt \right] \tag{4.56}$$

This limiting probability can be observed in Figure 4.12. It can be observed that the limiting solution leads to smaller intervention probabilities at any given time t.

4.7 Cost-based optimization

The life-cycle cost (LCC) of a project is defined as the distribution of total cost that is incurred, or may be incurred, in all stages of the project life-cycle. Within this definition, the life-cycle is defined by the time window required to achieve the functional or economic objectives for which the project was intended. LCC analysis provides a framework to support decisions about resource allocation related to the design, construction and operation of infrastructure systems at minimum life-cycle cost. The probabilistic description of the system performance with time also has implications on the cost effectiveness of the investment.

54

The cost structure of any project (objective function) can be described as:

$$Z(\mathbf{p}) = B(\mathbf{p}) - C_0(\mathbf{p}) - L(\mathbf{p}) \tag{4.57}$$

where $B(\mathbf{p})$ is the benefit for the existence of the structure, $C_0(\mathbf{p})$ the construction cost and $L(\mathbf{p})$ the value of losses if there is failure, which is usually written in terms of the the cost of losses, $C_L(\mathbf{p})$, and the failure probability, i.e., $L(\mathbf{p}) = C_L(\mathbf{p})P_f(\mathbf{p})$. The vector parameter $\mathbf{p}$ accounts for all design parameters.

The decision about the project has to be made analyzing the probability distribution of $Z(\mathbf{p})$ at time $t = 0$. For the project to be feasible, $Z(\mathbf{p})$ has to be positive. Therefore, the optimum design, construction and operation requirements are associated with the vector parameter $\mathbf{p}$ for which $Z(\mathbf{p})$ is positive and maximum. Equation 4.57 can be rewritten in many ways by, for instance, discretizing costs further, including multiple limit states, or extending the problem to multiple hazards (Wen and Kang, 2001 [71], Onoufriou and Frangolpol, 2002 [49], Val and Stewart, 2005 [65], and Mander et al., 2007 [41]). Optimal investment decisions require both maximizing the economic benefit of the life-cycle of the project and meeting the reliability requirements simultaneously at the decision point.

4.7.1 Derivation of terms for the objective function

Benefit function

The benefit $B(\mathbf{p})$ in equation (4.57) is a value that has to be discounted to the decision point. According to Streicher et al., 2008 [62], for a benefit rate that is not affected by short reconstruction times and constant, i.e., $b(t) = b$,

$$B = \int_0^\infty b\exp(-\gamma t)dt = \frac{b}{\gamma} \tag{4.58}$$

A more detailed derivation for computing the total benefit taking into account the time-dependence of $b(t)$ is given in Streicher et al., 2008 [62], and the results are shown here for completeness. Define $\theta_i = t_i - t_{i-1}$ to be the time between renewals with density $f(t, \mathbf{p})$. The density of the times between renewals was derived for the case of shocks and progressive deterioration in section 4.6.4.

The total benefit is (see the Appendix for full derivation):

$$
\begin{aligned}
B(\mathbf{p}) &= E\left[\sum_{i=1}^{\infty}\left(\exp\left[-\gamma\sum_{k=1}^{i}\theta_k\right]\int_0^{\theta_i}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)\right] \tag{4.59}\\
&= E\left[\int_0^{\theta_1}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right] + \frac{E[\exp\left(-\gamma\theta_1\right)]E\left[\int_0^{\theta}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right]}{1 - E[\exp\left(-\gamma\theta\right)]} \tag{4.60}
\end{aligned}
$$

For the case where the structure is abandoned after first failure or at a finite service time t_s (Streicher et al., 2008 [62]),

$$
\begin{aligned}
B(\mathbf{p}) &= E\left[\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right]\\
&= \int_0^{t_s}\left(\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)f_1(t, \mathbf{p})dt \tag{4.61}
\end{aligned}
$$

and for the case of systematic reconstruction, equation 4.60 becomes (Streicher et al., 2008 [62]):

$$
\begin{aligned}
B(\mathbf{p}) &= \int_0^{\infty}\left(\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)f_1(t, \mathbf{p})dt\\
&\quad + \frac{f_1^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})}\int_0^{\infty}\left(\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)f(t, \mathbf{p})dt \tag{4.62}
\end{aligned}
$$

55

Sometimes it might be more accurate to use the CDF instead of the density function. In these cases the following relationship can be used: $f^*(\kappa, \mathbf{p}) = \kappa F^*(\kappa, \mathbf{p})$. In summary, equations 4.58, 4.61 and 4.62 describe the three most common forms of modeling the benefit.

Loss function

The optimization of the objective function has to be made based on the expected values. Define θ_i as before (i.e., the time between renewals) and C_n the discrete cost associated with an intervention (early replacement or reconstruction). Then based on Streicher et al., 2008 [62], for systematic reconstruction, the discounted expected damage cost can be computed as (see the Appendix for full derivation):

$$
\begin{aligned}
L(\mathbf{p}) &= E\left[\sum_{n=1}^{\infty} C_n \exp\left[-\gamma \sum_{k=1}^{n} \theta_k\right]\right] \\
&= \frac{C_1 E[\exp(-\gamma\theta_1)]}{1 - E[\exp(-\gamma\theta)]} + \frac{-C_1 E[\exp(-\gamma\theta_1)]E[\exp(-\gamma\theta)]}{1 - E[\exp(-\gamma\theta)]} \\
&\quad + \frac{E[\exp(-\gamma\theta_1)]C_n E[\exp(-\gamma\theta)]}{1 - E[\exp(-\gamma\theta)]}
\end{aligned}
\tag{4.63}
$$

where θ is the time between any two interventions and C_n is the cost associated with the intervention in cycle n, which may include reconstruction and repair. The expected values in equation 4.63 can be computed as:

$$
E[\exp(-\gamma\theta_1)] = \int_0^{\infty} \exp(-\gamma t) f_1(t, \mathbf{u}) dt = f_1^*(t, \mathbf{u})
\tag{4.64}
$$

$$
E[\exp(-\gamma\theta)] = \int_0^{\infty} \exp(-\gamma t) f(t, \mathbf{u}) dt = f^*(t, \mathbf{u})
\tag{4.65}
$$

which describe the Laplace transform of the densities of the time to the first intervention and between any two interventions respectively. For the modified renewal and the case of systematic reconstruction, the cost of losses can be written as:

$$
\begin{aligned}
L(\mathbf{p}) &= \frac{C_1 f_1^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})} + \frac{-C_1 f_1^*(t, \mathbf{p}) f^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})} + \frac{f_1^*(t, \mathbf{p}) C_n f^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})} \\
&= \frac{f_1^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})} \left[C_1 + -C_1 f^*(t, \mathbf{p}) + C_n f^*(t, \mathbf{p})\right]
\end{aligned}
\tag{4.66}
$$

When convenient, the cost C_n can be extended to include loss of business and compensation cost as a result of risk to human life.

4.7.2 Objective function for optimization

Based on the discussion presented in previous sections, the objective function for optimization will be defined for the following cases: 1) structures abandoned after failure or at a fixed time t_s; and 2) structures systematically reconstructed after any intervention. For the first case, the objective function for the modified renewal process is:

$$
Z(\mathbf{p}) = B(\mathbf{p}) - C(\mathbf{p}) - [C_1 + (-C_1 f^*(t, \mathbf{p}) + C_n f^*(t, \mathbf{p}))] \frac{f_1^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})}
\tag{4.67}
$$

Note note that for systems that have been in operation for a long time or for systematically renewed structures $C_1 = C_n = C_L$ and, therefore, the last two terms in equation 4.66 cancel leading to: $L(\mathbf{p}) = C_1 f_1^*(t, \mathbf{p})/1 - f^*(t, \mathbf{p})$ and

$$
Z(\mathbf{p}) = B(\mathbf{p}) - C(\mathbf{p}) - C_L \frac{f_1^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})}
\tag{4.68}
$$

The function of $B(\mathbf{p})$ can be selected from equations 4.58, 4.61 and 4.62 based on the modeling requirements.

4.7.3 Discounting aspects

Any investment decision about the project has to be made at time $t = 0$; in other words, all future investments have to be discounted. Discounting costs throughout the project's life-cycle is used to compute the Net Present Value (NPV). In order to compute the NPV, all costs have to be discounted by using a continuous discount function: $\delta(t) = \exp(-\gamma t)$; where γ is the interest rate and t the time in suitable units. Note that for small values of γ this expression is similar to $\delta(t) = (1 + \gamma')^{-t}$, which is widely used, and where γ' is the yearly discount rate.

Selection of discount rates for infrastructure systems is a matter of ongoing debate. A broader approach suggests that they should be chosen to reflect how society becomes wealthier; therefore, they should be calculated as the long term average of the economic growth per capita. In this case, values vary between 0.9% (Africa) and 2.5% (USA and Canada) (Rackwitz et. al., 2005 [56]). In industrialized countries interest rates vary between 2% and 8%, while in moderate and low developed countries interest rates may vary from 8% to 18% and from 15% to 30% respectively. Wen and Kang, 2001 [71], states that for the public sector these rates are between 4% and 6%, while for the private sector, they vary between 6% and 10%. Other related issues that are beyond the scope of this paper but that must be addressed are the variation of discount rates with time (see, Rackwitz et. al., 2005 [56]) and the discount rates for the loss of human lives (Pate-Cornell, 1984 [51], Johansson, 2001 [32] and Bayer, 2003 [9]). An interesting discussion on public and private discounting for life-cycle cost analysis can be found in Corotis, 2005 [16].

4.7.4 Illustrative example

Consider the information and the data discussed in the illustrative examples throughout this document. The cost optimization data used in the example are as follows: the benefit obtained from the system operation is assumed to be constant $b = 0.15 \times C_0$; the part of the construction cost (independent of the target parameters) is $C_0 = 3 \times 10^6$; and the construction cost as a function of the design parameter is $C(p) = C_0 + C_1 \cdot p^\beta$, with $\beta = 1.1$. In case of failure, it was initially assumed that for low threshold operational conditions, i.e., $s^* < 25$, the cost of failure (structural reconstruction) is $C(p)$; however, in addition to reconstruction costs, an additional constant cost associated to indirect losses (loss of business opportunity, life loss compensation, etc.) was taken as $L = 15 \cdot C_0$. Finally, a time invariant annual discount rate $\gamma = 0.05$ was assumed. All costs are assumed to be in appropriate currency units. The structure is subject only to multiple shocks (progressive deterioration is not considered) with the stochastic characteristics described in previous examples.

The optimization will focus on determining the value of the mean initial remaining life for which the structure has to be designed to maximize the profit. This can be expressed as:

$$
\begin{aligned}
\max \quad & Z(p) = B(p) - C(p) - C_L \frac{f_1^*(t,p)}{1 - f^*(t,p)} \\
\text{Subject to:} \quad & Z(\mathbf{p}) \geq 0; \\
& p \geq s^* = 25
\end{aligned}
\tag{4.69}
$$

where $C_L = C(p) + 15 \cdot C_0$. It is assumed also that $f_1^*(t,p) = f^*(t,p)$. The components of the objective function are shown in Figure 4.13a. The optimization yields to a $p_{opt} = 116.01$. As expected the optimum value varies depending upon the selection of the threshold that defines the need for an intervention, i.e., s^*; Figure 4.13b presents the variation of the optimum solution, i.e., $p_{opt}(s^*)$. Large threshold values imply higher operational standards. In every case the relationship between the optimum and the threshold changes and varies between $2 \leq p_{opt}/s^* \leq 10$ for threshold values between 10 and 90.

Consider now, for comparison purposes, the case of a system exposed to multiple shocks where damage does not accumulate with time. Then shocks occur randomly in time but at every shock the system might fail or not. It it does not fail, it is assumed that the remaining life remains as in its initial
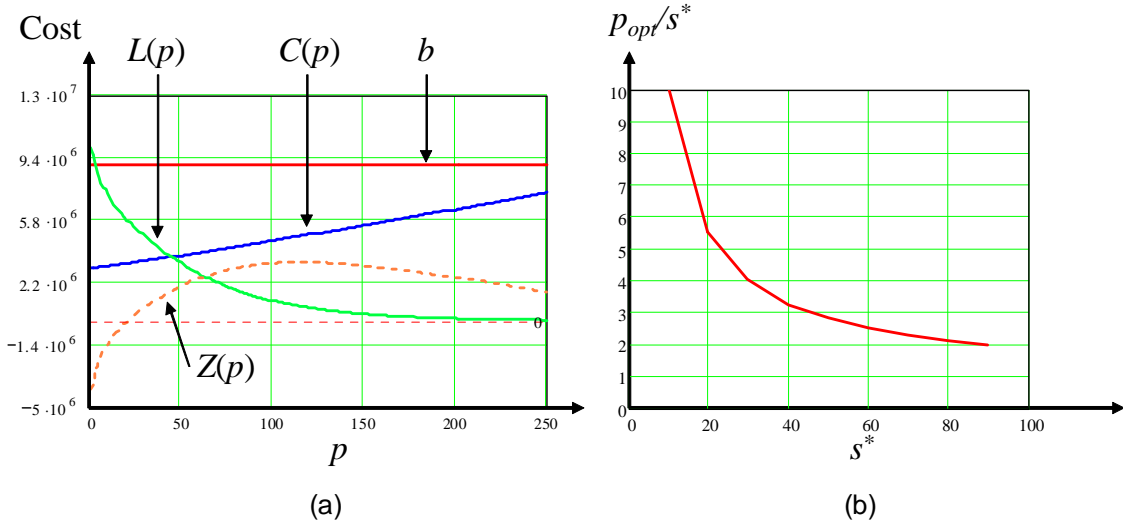
Figure 4.13: a) Basic optimization components; and b) optimum design values for different intervention thresholds.

condition. In this case, the cumulative distribution function of the time to failure can be computed as:

$$P(V(t) \leq s^*, t) = \sum_{n=0}^{\infty} \sum_{i=1}^{n} \left[P_f(p, s^*)(1 - P_f(p, s^*))^{i-1} \right] P(N = n, t) \qquad (4.70)$$

where $P_f(p, s)$ is the probability that the remaining life falls below a predefined threshold s^*. The condition on time is removed in the objective function by computing the Laplace transform. Comparing the cases of damage accumulation and not accumulation, it is observed that for the former case larger optimum values are obtained (Figure 4.14). It is justified by the increase in probability of failure with time as damage accumulates. This indicates that a renewal model that considers the case of failure or not-failure at every shock will always underestimate the optimum.

Consider now the case where the cost of a loss is conditioned on the threshold defining the need for an intervention and, if an intervention is required, on the level of damage. Under this assumption, the cost of an intervention is redefined as:

$$C_L(p, x) = \left\{ \begin{array}{cc} C_o + C_1 p^\beta & x \leq S_{\min} \\ \varphi C_o + C_1 x^\beta & s^* \geq x > S_{\min} \\ 0 & x > s^* \end{array} \right\} \qquad (4.71)$$

where $\varphi > 0$ and $\beta > 0$ are constants, and x is the remaining life after the event that causes the intervention. In this case, a new threshold value S_{min} has been defined as the level of damage (loss of remaining life) below which the intervention requires total reconstruction. For this intervention model the optimum values are reduced significantly since not all interventions will cost the same nor have the same probability of occurrence. For instance for a value of $S_{min} = 10$ and $s^* = 40$ the optimum for the three cases discussed are: 130.71 (multiple shocks and degradation), 113.04 (multiple shocks and no degradation) and 68.6 (multiple shocks and degradation).
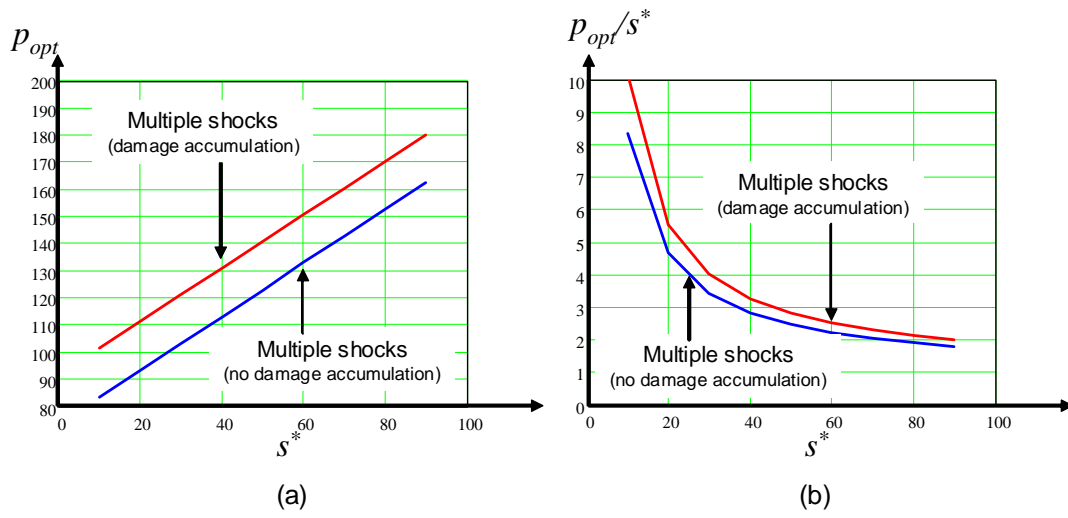
Figure 4.14: Comparison of the models (a) with and (b) without damage accumulation.

Chapter 5

Summary and conclusions

This research is directed to model the performance of infrastructure network systems when they are exposed to the presence of both internal and external demands. The first part of this reports deals the problem of network modeling by assessing the performance of the entire system. The second part is related to the loss of operational capacity of individual components with time.

A systems approach to modeling infrastructure networks exposed to different hazards is presented in the first part of this document (Chapters 2 and 3). The model combines systems thinking with strategies for network clustering. Networks are described by a hierarchical structure that is obtained after successive clustering. The hierarchical structure is used to make estimations of the impact of any event on the system, allowing loss estimations and measurements of the hazard impact that can go beyond direct damage evaluation. The practical application of the model is illustrated by using the transportation network of Texas as a case study. Several fictitious hazard events illustrate the model and the potential application in practice. In addition, the impact of Hurricane Ike on the network was also studied. Estimations of population affected and productivity losses were made based on the network structure. The proposed model is a tool that can be used to make estimations of both direct and indirect losses. However, the main advantage of the model is its potential to estimate indirect losses which are strongly related with the emergent properties of the network (connectivity, flow, etc.) and not to the direct physical response to the event.

In the second part of the document (Chapter 4), a probabilistic model of structural deterioration was developed to model the performance of network components subject to several demand types. The model can be used to characterize the performance of structures that deteriorate as a consequence of both sudden events, i.e., shocks (e.g., earthquakes, terrorist attacks, accidents) and progressive degradation (biodegradation, sulfur attack, corrosion, fatigue). The model describes the system's performance in terms of its remaining life, which in practice would describe structural properties such as a structure inter-story drift or maximum displacement, resistance or safety availability. It takes into account both the accumulation of damage as a consequence of successive shocks and progressive deterioration. Shocks are modeled as independent events that occur randomly in time whose size are also random, and which are described by a probability distribution. Progressive deterioration was first modeled as a deterministic continuous function. Afterwards, randomness was included by modeling it as successive small shocks distributed according to a deterministic function. In addition to providing the means to compute the system's failure probability, the model provides the basis for modeling and defining preventive maintenance strategies. For instance, it provides flexibility to define different damage limit states that define intervention levels. The results of the illustrative examples presented and discussed stress the importance of taking into account the accumulation of damage caused by both shocks and progressive deterioration on the probability of failure. Results also show that the probability of failure can be significantly high if these progressive failure mechanisms are considered.

Bibliography

[1] Agarwal J., Blockley D.I. and Woodman N. (2003), Vulnerability of structural systems. Structural Safety 25, 263-268.

[2] Aiello W. Chung F. and Lu L.(2000), A random graph model for massive graphs. In Proc of the 32nd Annual ACM Symposium on Theory of Computing, Association of Computing Machinery, New York, 171-180.

[3] Alpert C.J. and Kahng A.B. (1994), Multiway partitioning via spacefilling curves and dynamic programming. In Proc of the 31st ACM/IEEE Design Automation Conference, ACM New York, NY, USA, 652-657.

[4] Alpert C.J., Kahng A.B. and Yao S-Z. (1999), Spectral partitioning with multiple eigen vectors. Discrete Applied Mathematics, 90, 3-26.

[5] Aven T.J. and Jansen U. (1999), Stochastic Models in Reliability. Applications of Mathematics: Stochastic Modeling and Applied Probability Series 41, Springer.

[6] Barlow R.E. and Poschan F. (1965), Mathematical Theory of Reliability. Wiley, New York.

[7] Bremaud P. (1981), Point Processes and Queues. Spring Verlag, New York.

[8] Barnes E.R. (1982), An algorithm for partitioning the nodes of a graph. SIAM J. Alg. Disc. Math., 3, 541-549.

[9] Bayer S. (2003), Generation-adjusted discounting in long-term decision-making. Intl. J. Sustainability Development, 6(1), 133-49.

[10] Blockley D.I. and Godfrey P. (2000), Do it differently: systems for rethinking construction. Thomas Telford, London.

[11] Chang P.K., Schlang M.D.F. and Zien J. (1994), Spectral k-way ratio cut partitioning and clustering. IEEE Transactions CAD, 1088-1096.

[12] Cinlar E. (1975), Introduction to stochastic processes. Prentice Hall Inc, New Jersey.

[13] Cinlar E., Bazant Z.P. and Osman E. (1977), Stochastic process for extrapolating concrete creep. Journal of the Engineering Mechanics Division, 103 (EM6), 1069–1088.

[14] Clauset A., Moore C., and Newman M. (2006), Structural inference of hierarchies in networks. Proc. Workshop on Statistical Network Analysis, 23rd International Conference on Machine Learning (ICML '06). Lecture Notes in Computer Science 4503, 1–13.

[15] Cohen R., Erez K., Ben-Avraham D. and Havlin S. (2000), Resilience of the internet to random breakdowns. Phys. Rev. Lett., 85, 4626-4628.

[16] Corotis R.B. (2005), Public versus private discounting for life-cycle cost. In Safety and Reliability of Engineering Systems and Structures: Proceedings of the 9th International Conference on Structural Safety and Reliability (ICOSSAR'05). Millpress, Rotterdam, The Netherlands, 249.

[17] Dhillon I.S., Guan Y. and Kulis B. (2004), Kernel k-means spectral clustering and normalized cuts. In Proc of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'04, Seattle Washington, USA, 551-556.

[18] Dueñas-Osorio L, Craig J.I. and Goodno B.J. (2007), Seismic response of critical interdependent networks. Earthquake Eng. Struct. Dyn., 36, 285–306.

[19] Ellingwood B.R. and Mori Y. (1993), Probabilistic methods for condition assessment, life prediction of concrete structures in nuclear power plants. Nuclear Engineering and Design, 142, 155–166.

[20] Federal Reserve Bank of Dallas (FRBD) (2008), http://www.dallasfed.org.

[21] Feldman R.M. (1976), Optimal replacement with semi-Markov shock models. Journal of Applied Probability, 13, 108-117.

[22] Feldman R.M. (1977), Optimal replacement for systems governed by Markov additive shock processes. Annals of Probability, 5, 413-429.

[23] Feldman R.M. (1977), Optimal replacement with semi-Markov shock models using discounted costs. Mathematical Operations Research, 2, 78-90.

[24] Frangopol D.M., Kallen M-J. and Noortwijk M. (2004), Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. Steel Construction Prog. Struct. Engng Materials, 6, 197-212.

[25] Gopal S. and Majidzadeh K. (1991), Application of Markov decision process to level of service based maintenance systems. Transportation Research Record, 1304.

[26] Guillaumot V.M., Durango P.L. and Madanat S. (2003), Adaptive optimization of infrastructure maintenance and inspection decisions under performance model uncertainty. J. Infrastruct. Syst., ASCE, 9(4), 133-139.

[27] Hadley S.W., Mark B.L. and Vannelli A. (1992), An efficient eigen vector approach to finding netlist partitions. IEEE Transactions CAD 11, 885-892.

[28] Harper W., Lam J., Al-Salloum A., Al-Sayyari S., Al-Theneyan S., Ilves G. and Majidzadeh K. (1990), Stochastic optimization subsystem of a network-level bridge management system. Transportation Research Record, 1268, 68-74.

[29] Hasofer A.M. (1974), Design for infrequent overloads. Earthquake Engineering and Structural Dynamics, 2(4), 387-8.

[30] Holme P., Kim B.J., Yoon C.N. and Han S.K. (2002), Attack vulnerability of complex networks. Physical Review E., 65, 056109.

[31] Jiang W. and Xilia L. (2005), Prediction of structural performance and life-cycle analysis based on Bayesian dynamic models. In Safety and Reliability of Engineering Systems and Structures: Proceedings of the 9th International Conference on Structural Safety and Reliability (ICOSSAR'05), Millpress, Rotterdam, The Netherlands, 1679-1686.

[32] Johansson P.O. (2001), Is there a meaningful definition of the value of statistical life? Health Economics, 20, 131-139.

[33] Kanungo T., Mount D.M., Netanyahu N.S., Piatko C.D., Silverman R. and Wu A.Y. (2002), An efficient k-means clustering algorithm: analysis and implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (7), 881-892.

[34] Kleiner Y. (2001), Scheduling inspection, renewal of large infrastructure assets. Journal of Infrastructure Systems, 7(4), 136–143.

[35] Kubler O. and Faber M.H. (2003), Optimal design of infrastructure facilities subject to deterioration. In Proc of International Conference on Applied Statistic and Probability in Civil Engineering, ICASP'03, Millpress Science Publishers, San Francisco, 1031-1039.

[36] Klutke G-A. and Yang Y. (2002), The availability of inspected systems subject to shocks and graceful degradation. IEEE Transactions on Reliability, 51(3), 371-374.

[37] Konak, A., Smith A.E., and Kulturel-Konak, S. (2004), New Event-driven Sampling Techniques for Network Reliability Estimation. In Proc of Winter Simulation Conference 2004, Washington, D.C., 224-231.

[38] Lattimore B.S.,Van Dongen S. and Crabbe M.J. (2005), GeneMCL in microarray analysis. Computational Biology and Chemistry, 29(5), 354-359.

[39] Liu Y. and Weyers R.E. (1998), Modeling the time-to-corrosion cracking of the cover concrete in chloride contaminated reinforced concrete structures. ACI Materials Journal 95, 675-81.

[40] McQueen J. (1968), Some methods for classification and analysis of multivariate observations. Computer and Chemistry, 4, 257-272.

[41] Mander J.B., Dhakal R.P., Mashiko N. and Solberg K.M. (2007), Incremental dynamic analysis applied to seismic financial risk assessment of bridges. Engineering Structures, 29(10), 2662-2672.

[42] Mishalani R.G. and Madanat S. M. (2002), Computation of infrastructure transition probabilities using stochastic duration models. Journal of Infrastructure Systems, 8(4), 139–148.

[43] Mori Y. and Ellingwood B. (1994), Maintaining reliability of concrete structures. I: role of inspection and repair. ASCE Journal of Structural Engineering, 120(3), 824-825.

[44] Nakagawa T. (1976), On a replacement problem of cumulative damage model. Operational Research Quarterly, 27(4), 895-900.

[45] Newman M.E.J. (2003), The structure and function of complex networks. Society for Industrial and Applied Mathematics (SIAM) Review, 45(2), 167-256.

[46] Newman M.E.J. (2004), Detecting community structure in networks. Eur. Phys. J., 38, 321–330.

[47] Newman M.E.J. and Leicht E.A. (2007), Mixture models and exploratory analysis in networks. PNAS, 104 (23), 9564-9569.

[48] Ng, A., Jordan, M., Weiss, Y. (2002). On spectral clustering: analysis and an algorithm. In Advances in Neural Information Processing Systems 14, MIT Press, Cambridge, 849- 856.

[49] Onoufriou T. and Frangopol D.M. (2002), Reliability-based inspection of complex structures: a brief retrospective. Computers and Structures, 80, 1133-1144.

[50] Pandey, M.D. (1998), Probabilistic models for condition assessment of oil and gas pipelines. Int. J. Non-Destructive Testing and Evaluation, 31(5), 349-358.

[51] Paté-Cornell M.E. (1984), Discounting risk analysis: capital versus human safety. In Proc Symp on Structural Technology and Risk. University of Waterloo Press, Waterloo, ON, Canada, 17-20.

[52] Peng J. and Xia Y. (2005), A new theoretical framework for k-means-type clustering, StudFuzz, 180, 79–96.

[53] Pentney W. and Meila M. (2005), Spectral clustering of biological sequence data. In Proc of the 25th Annual Conference of the American Association for Artificial Intelligence (AAAI), 1, 845–850.

[54] Petcherdchoo A., Kong J.S., Frangopol D.M. and Neves L.C. (2004), NLCADS (new life-cycle analysis of deteriorating structures) user's manual. Structural Engineering and Structural Mechanics Research Series No. CU/SR-04/3, Department of Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder 04/3, 63.

[55] Rackwitz R. (2000), Optimization - the basis for code making and reliability verification. Structural Safety, 22, 27-60.

[56] Rackwitz R., Lentz A. and Faber M. (2005), Socioeconomically sustainable civil engineering infrastructures by optimization. Structural Safety 27:187-229.

[57] Rendel F. and Wolkovicz H. (1995), A projection technique for partitioning nodes of a graph. Annual Operations Research, 58, 155-179.

[58] Rosemblueth E. and Mendoza E. (1971), Optimization in isostatic structures. J. Eng. Mech. Div., ASCE, (EM6) 1625-42.

[59] Rosemblueth E. (1976), Optimum design for infrequent disturbances. J. Struct. Div., ASCE, 102, (ST9) 1807-25.

[60] Ross S.M.(1996), Stochastic processes. 2nd Edition, Wiley.

[61] Scott J. (2000), Social network analysis: a handbook. 2nd Ed. Sage, London,

[62] Streicher H., Joanni A. and Rackwitz R. (2008), Cost-benefit optimization and risk acceptability for existing, aging but maintained structures. Structural Safety, 30, 375-393.

[63] Sherif Y. and Smith M. (1981), Optimal maintenance models for systems subject to failure - a review. Naval Research Logistics Quarterly, 28, 47-74.

[64] Taylor H.M. (1975), Optimal replacement under additive damage and other failure models. Naval Research Logistics Quarterly, 22, 1-18.

[65] Val D. and Stewart M. (2005), Decision analysis for deteriorating structures. Reliability Engineering & System Safety, 87, 377-385.

[66] Valdez-Flores C. and Feldman R.M. (1989), A survey of preventive maintenance models for stochastically deteriorating single unit systems. Naval Research Logistics Quarterly, 419-446.

[67] Van Dongen S. (2000), Graph clustering by flow simulation. PhD thesis, University of Utrecht.

[68] Van Dongen S. (2000), A cluster algorithm for graphs. Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, May.

[69] Van Dongen S. (2000), A stochastic uncoupling process for graphs. Technical Report INS-R0011, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam.

[70] Wasserman S. and Faust K. (1994), Social network analysis. Cambridge University Press, Cambridge.

[71] Wen Y.K. and Kang Y.J. (2001), Minimum building life-cycle cost design criteria. I: methodology. ASCE J. of Structural Engineering, 127(3), 330-337.

[72] Zuckerman D. (1977), Replacement models under additive damage. Naval Research Logistics Quarterly, 24, 549-558.

Chapter 6

Appendix

6.1 Benefit function

Derivation of the equation to estimate losses after Streicher et al., 2008 [62]. Let's define $\theta_i = t_i - t_{i-1}$ to be the time between renewals with density $f(t, \mathbf{p})$. Then, the total benefit is

$$
\begin{aligned}
B(\mathbf{p}) &= E\left[\sum_{i=1}^{\infty}\left(\exp\left[-\gamma\sum_{k=1}^{i}\theta_k\right]\int_0^{\theta i}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)\right] \\
&= E\left[\int_0^{\theta_1}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right] + E\left[\sum_{i=2}^{\infty}\prod_{k=2}^{i-1}\left(\exp\left[-\gamma\theta_k\right]\int_0^{\theta_i}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)\right] \\
&= E\left[\int_0^{\theta_1}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right] + E[\exp\left(-\gamma\theta_1\right)]\sum_{i=2}^{\infty}E[\exp\left[-\gamma\theta\right]]^{i-2}E\left[\int_0^{\theta_1}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right] \\
&= E\left[\int_0^{\theta_1}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right] + \frac{E[\exp\left(-\gamma\theta_1\right)]E\left[\int_0^{\theta}\exp\left(-\gamma\tau\right)b(\tau)d\tau\right]}{1 - E[\exp\left(-\gamma\theta\right)]}
\end{aligned}
\tag{6.1}
$$

Note that the derivation uses the following known relationship: $\sum_{n=k}^{\infty}aq^{n-k} = a/(1-q)$ for $k < \infty$. Since it is sometimes more accurate to use the CDF, i.e., $F(t, \mathbf{p}) = \int_0^t f(\tau, \mathbf{p})d\tau$ instead of the density function, the integration theorem for Laplace transforms can be used when necessary; this is, $F^*(\kappa, \mathbf{p}) = \int_0^{\infty}\exp(-\kappa t)f(t, \mathbf{p})dt = \frac{f^*(\kappa, \mathbf{p})}{\kappa}$, which leads to $f^*(\kappa, \mathbf{p}) = \kappa F^*(\kappa, \mathbf{p})$.

For the case of systematic reconstruction, equation 4.60 can be simplified to (Streicher et al., 2008 [62]):

$$
\begin{aligned}
B(\mathbf{p}) &= \int_0^{\infty}\left(\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)f_1(t, \mathbf{p})dt \\
&\quad + \frac{f_1^*(t, \mathbf{p})}{1 - f^*(t, \mathbf{p})}\int_0^{\infty}\left(\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)f(t, \mathbf{p})dt
\end{aligned}
\tag{6.2}
$$

and for the case where the structure is abandoned after first failure or at a finite service time t_s (Streicher et al., 2008 [62]),

$$
\begin{aligned}
B(\mathbf{p}) &= E\left[\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right] \\
&= \int_0^{t_s}\left(\int_0^t\exp\left(-\gamma\tau\right)b(\tau)d\tau\right)f_1(t, \mathbf{p})dt
\end{aligned}
\tag{6.3}
$$

6.2 Loss function

Let's define θ_i as before (i.e., the time between renewals) and C_n the discrete cost associated to an intervention (early replacement or reconstruction). Then, based on Streicher et al., 2008 [62], for systematic reconstruction the discounted expected damage cost can be computed as:

$$
\begin{aligned}
L(\mathbf{p}) &= E\left[\sum_{n=1}^{\infty} C_n \exp\left[-\gamma \sum_{k=1}^{n} \theta_k\right]\right] \\
&= E[C_1 \exp(-\gamma\theta_1)] + E\left[\sum_{n=2}^{\infty} C_n \prod_{k=1}^{n} \exp[-\gamma\theta_k]\right] \\
&= E[C_1 \exp(-\gamma\theta_1)] + E\left[\sum_{n=2}^{\infty} C_n E[\exp(-\gamma\theta_1)]\left(\prod_{k=2}^{n-1} \exp[-\gamma\theta_k]\right)\exp[-\gamma\theta_n]\right] \\
&= E[C_1 \exp(-\gamma\theta_1)] + \sum_{n=2}^{\infty} E[\exp(-\gamma\theta_1)] E[\exp(-\gamma\theta)]^{n-2} E[C_n \exp(-\gamma\theta_n)] \\
&= E[C_1 \exp(-\gamma\theta_1)] + E[\exp(-\gamma\theta_1)]\frac{E[C\exp(-\gamma\theta)]}{1 - E[\exp(-\gamma\theta)]} \\
&= \frac{C_1 E[\exp(-\gamma\theta_1)]}{1 - E[\exp(-\gamma\theta)]} + \frac{-C_1 E[\exp(-\gamma\theta_1)]E[\exp(-\gamma\theta)]}{1 - E[\exp(-\gamma\theta)]} \\
&\quad + \frac{E[\exp(-\gamma\theta_1)]CE[\exp(-\gamma\theta)]}{1 - E[\exp(-\gamma\theta)]}
\end{aligned}
$$

where θ and C_n are the time and cost between any two interventions.

University Transportation Center for Mobility

Texas Transportation Institute

The Texas A&M University System

College Station, TX 77843-3135

Tel: 979.845.2538   Fax: 979.845.9761

utcm.tamu.edu

**Texas
Transportation
Institute**