Scalable Data Management and Data-Driven Analytics for Structural Condition Assessment using Structural Monitoring Data

by

Yilan Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Civil Engineering)
in The University of Michigan
2016

Doctoral Committee:

      Professor Jerome P. Lynch, Chair
      Assistant Professor Eunshin Byon
      Professor Vineet R. Kamat
      Professor Victor C. Li
      Professor Atul Prakash

DEDICATION

To my parents, Hong and Ronghua, and my husband, Sean,

for your love and support

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Professor Lynch for the continuous support of my Ph.D. study and related research, for his guidance, motivation, and immense knowledge. He has taught me how to conduct research, contribute to the academic community, and be a scholar. His continuous encouragement and support has driven me to overcome many challenges during the last five years. Professor Lynch has been a teacher, mentor and also friend throughout my time at the University of Michigan for which I will be forever grateful.

I would also like to thank my family. My husband, Sean, continuously provided me love and support which has kept me motivated and encouraged. We first met at the LIST research group and have worked side by side for the last five years. I am grateful to have him to share my life with. My parents, Hong and Ronghua, have taught me the value of hard work and have provided all the support and encouragement needed to come to the University of Michigan and pursue a higher education.

I would also like to thank my dissertation committee, Professor Prakash, Professor Kamat, Professor Li and Professor Byon for their support toward my thesis work. Professor Prakash provided me with guidance in the research direction of data management systems. Professor Kamat guided my research on the application of long-term monitoring data for bridge inspectors.

Professor Li has provided insight regarding the academic contributions of my research and drove me to always consider the novelty of my work. Professor Byon was a mentor on statistical modeling and machine learning. I deeply appreciate the support and the guidance from my committee.

My fellow LIST members and friends in Ann Arbor have also had a great impact on my life during the last five years. My first experiences with the LIST lab were with Dr. Masahiro Kurata. Masa was my daily mentor and a great friend. He taught me the basics about research and provided me with great vision for my future work. My senior LIST members, Dr. Andy Zimmerman, Dr. Junhee Kim, Dr. Sukhoon Pyo, Dr. Michael Kane and Dr. Courtney Peckens, all have been teachers and mentors for me on different subjects, which I deeply appreciate. My fellow graduate students in LIST gave me support and friendship that made the past five years more enjoyable. In particular I'd like to thank Jeff Bergman, Ted Byrne, Nephi Johnson, Andy Burton, Mitsuhito Hirose, Rui Hou, Katherine Flanigan, Hao Zhou and Omid Bahrami. The large group of graduate students in CEE and my friends here made my time in Ann Arbor so much more memorable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Scalable Data Management and Data-Driven Analytics for Structural Condition Assessment using Structural Monitoring Data

By

Yilan Zhang

Chair: Dr. Jerome P. Lynch

Structural monitoring systems are an objective and quantitative-based management tool that have been developed to assist structure owners with their diagnostic and prognostic structural condition monitoring processes. As sensing technologies mature, the deployment of permanent sensing arrays in structures are becoming more popular resulting in increasing volumes of sensing data. This thesis focuses on the advancement of structural condition assessment by developing a scalable data management system for the curation and analysis of large-scale data sets associated with long-term structural monitoring systems. A hybrid database system termed *SenStore* is proposed consisting of a relational database system for the storage of structural information and a hierarchical data format (HDF) repository for the storage of time history sensor measurements. By storing sensor data in an HDF repository, sensor data can be queried at high speeds resulting in greater scalability of analytics to be performed on the data. The thesis also develops novel

approaches to interrogating data associated with long-term structural monitoring systems. Data-driven analytics are built on top of the *SenStore* data management platform to extract information associated with the performance and condition of the monitored structure from the curated data. Automated modal parameter extraction by stochastic subspace identification (SSI) is implemented to obtain the modal properties of structures monitored using accelerometers. Long term trends in modal parameters are modeled by Gaussian process regression (GPR). A load estimation algorithm is also proposed for highway bridges monitored using strain gages. The data-driven analytics developed in the thesis are evaluated using long-term monitoring datasets from the New Carquinez Bridge (Vallejo, CA) and the Telegraph Road Bridge (Monroe, MI).

CHAPTER 1

INTRODUCTION

1.1 Structural Monitoring and the Need for Data Management Solutions

Undetected structural deterioration (*e.g.*, corrosion, fracture) and extreme loading events (*e.g.*, earthquake, hurricane, floods) are two major causes of structural failures in civil infrastructure systems (Brownjohn 2007; Çatbaş et al. 2013; Farrar and Worden 2007; Nagayama and Spencer 2007). In current practice, trained inspectors regularly carry out visual inspections to qualitatively evaluate the condition of structures to ensure their safety for public use. To track long-term conditions and to detect deterioration in a timely manner, visual inspections are typically carried out on a regular basis (*e.g.*, quarterly, yearly, *etc.*). In the United States, highway bridges, railway bridges, and pipelines are just some examples of infrastructure requiring routine visual inspection as mandated by law (DOT 2014; FHWA 2004; US Congress 2006). When extreme load events occur, inspectors inspect the structure immediately following the event to determine if the structure can safely resume operations. For example, recently enacted building occupancy resumption programs allow structure owners to rapidly assess if their structures are safe immediately following a large seismic event; findings from visual inspections are a key input to re-occupancy decisions (BORP 2011; Çelebi et al. 2004). While visual inspections are critical to ensuring safety, they have limitations including: 1) inspectors only have access to the surface of the structure (making subsurface damage difficult if not impossible to detect), 2) inspectors are required to interpret

visual cues to make conclusive statements regarding structural conditions, and 3) visual inspection only provides *qualitative* information in the form of inspector ratings (with no real quantitative basis to those ratings).

Structural monitoring has been proposed as a potential means to help address the limitations of visual inspections by tracking the performance and health of a structure. Structural monitoring systems consisting of sensors and a communication medium (wired or wireless) are installed in a structure to measure structural responses to loads (*e.g.*, dead and live loads) and environmental conditions (*e.g.*, temperature, wind speed), and to communicate the response data to a data repository. The goal of structural monitoring is two-fold. First, structural response data is valuable for validating design assumptions and understanding structural performance under routine and extreme loads. Second, engineers can use data to *quantitatively* assess structural conditions over time. In particular, monitoring systems designed to assess structural health in an automated fashion are differentiated from ordinary structural monitoring systems by being termed structural *health* monitoring (SHM) systems. There are several notable deployments of structural monitoring and SHM systems on operational structures including on buildings (Brownjohn 2003; Çelebi et al. 2004; Kijewski-Correa and Pirnia 2007; Nayeri et al. 2008), dams (de Sortis and Paoliani 2007) and wind turbines (Rolfes et al. 2007; Smarsly et al. 2012). Especially in the field of bridge engineering, wired structural monitoring systems have been widely adopted to monitor critical bridges such as long-span bridges in areas known for seismic events and wind storms (Brownjohn et al. 2010; Çelebi 2006; Fujino et al. 2000; Ni and Wong 2012; Wang et al. 2000; Zhou and Ou 2005). More recently, wireless monitoring systems have emerged as a viable alternative to wired systems. Wireless monitoring systems use wireless communication in lieu of expensive wiring for

the exchange of data between sensors and a data repository (Lynch 2002, 2007; Spencer et al. 2004; Straser and Kiremidjian 1998). Early examples of wireless monitoring systems installed in operational bridges include short-term (Lynch et al. 2006; Pakzad et al. 2008; Straser and Kiremidjian 1998; Whelan and Janoyan 2008), long-term (Jang et al. 2010), and permanent (Feltrin et al. 2010; Kurata et al. 2013; O'Connor et al. 2016) installations. Regardless of the type of communication used (wired versus wireless), structural monitoring systems have the potential to generate large amounts of data with time. For example, the permanent monitoring system installed on the Ting Kau Bridge in Hong Kong has 227 sensing channels recording environmental parameters (*e.g.*, temperature, wind speed) and structural responses (*e.g.*, vibration, displacement, and strain) (Wong 2004). The bridge monitoring system collects 55 MB of data every hour which is equivalent to 470 GB of data every year. As structural monitoring systems reduce in cost and are easier to install (*e.g.,* wireless), more structures instrumented with higher nodal densities are likely. As monitoring grows, the need to store and analyze massive amounts of sensor data grows in tandem. Scalable data management technologies are needed to allow data to be securely stored and analyzed in an autonomous manner.

The need for scalable data management solutions is further motivated by the growing popularity of data-driven approaches (*e.g.,* machine learning) for structural health assessment(Farrar and Worden 2012). Such algorithms typically entail the processing of large sets of data to develop statistical models from the data. One of the earliest approaches to data-driven analytics in structural health monitoring was treating the problem of damage detection as a pattern classification problem. In contrast to traditional approaches to damage detection that are based on physical models of the structure (Doebling et al. 1996), supervised pattern classification methods classify damage based

on statistical models of a feature vector with the statistical models "trained" from previous data. The power of the methodology is the generality of the framework. Different feature vectors and various classification methods can be used to attain a high level of performance (*i.e.,* damage discrimination). With training an integral part of the pattern classification methodology, large data sets are typically needed with data corresponding to the structure in its full range of environmental and operation conditions (EOC). The role of data-driven analytics has rapidly expanded to go well beyond characterizing damage in structures. Data-driven methods have proven valuable for modeling complex and often nonlinear structural behavior under the range of EOC parameters structures are exposed to. For example, modal parameters extracted from experimental data are known to be EOC dependent; advance regression methods can be adopted to fit nonlinear models to modal and EOC parameters(Cross et al. 2012; Hua et al. 2007; Moser and Moaveni 2011; Sohn et al. 2001). With the ability to perform these analyses over large volumes of data, statistical models can be applied to characterize response distributions and model uncertainties. In a similar fashion, the nonlinear behavior of structures can be modeled by nonlinear regressions methods. A common strategy is the use of neural networks to train nonlinear restoring force curves from extensive amounts of experimental data(Nakamura et al. 1998; Pei and Smyth 2006). Another area that benefits from data-driven analysis is the assessment of structural reliability which is based on probabilistic and regression analysis (Ditlevsen and Madsen 2007). To facilitate the continued development of data-driven analytics of structural monitoring data, scalable data management solutions that support high-speed, high-volume querying of large sets of data are needed.

1.2 Existing Data Management Solutions for Scientific and Engineering Applications

The growing need for scalable data management platforms for massive sets of sensor data exists in many engineering and scientific domains including in water management (Goodall et al. 2011; Muste et al. 2012), medicine (Buetow 2005), urban planning (Tomko et al. 2012), and earthquake engineering (Bardet et al. 2004), just to name a few. In all of these domains, researchers typically seek to integrate two types of data: metadata providing information about the object of interest and large-volumes of temporal sensor data. The needs of the structural engineering field are no different.

Comprehensive data management platforms that store information associated with structural management have been developed. In the domain of buildings, building information modeling (BIM) systems have flourished in recent years (Autodesk 2002; Eastman et al. 2011). Similarly, Bridge Information Modeling (Herman et al. 2012) or BrIM is the BIM counterpart for bridges. BIM (and BrIM) models are a digital representation of information associated with the design, construction, and operational phases of a structure's life-cycle unified in a single database system. While effective in their ability to store building information, researchers have only recently began to contemplate the extension of BIM to store sensor data (Hajian and Becerik-Gerber 2009; Liu and Akinci 2009; Nader et al. 2010). These early efforts focus on storing sensor data in a separate database or file system. Sensor locations that are stored in the BIM platform serve as the linkage to the sensor data that is externally stored. This framework largely limits the type and speed of queries that can be done by BIM users on the data. In the domain of bridge engineering, a data management platform termed AASHTOWare Bridge Management (formerly known as PONTIS) was developed to assist bridge owners with storing bridge information (*e.g.*, location, geometry,

materials) along with visual inspection data (Robert et al. 2003). AASHTOWare is designed using an element-based modeling ontology because bridges are rated on an element-by-element basis (*e.g.*, beam, column, deck, bearing). A strength of AASHTOWare is its ubiquitous use by bridge owners in the United States, leading to a widely accepted standard in the representation of bridge information and inspection condition data (*i.e.*, element rating). Unfortunately, AASHTOWare was not designed to store sensor data associated with bridge monitoring.

Several projects are underway in the structural monitoring community exploring scalable data management platforms for structural monitoring. These research efforts have focused on the storage of large sets of sensor data. For example, Koo et al. (2011) proposed a MySQL database system for storing structural response data, visualization tools for viewing sensor data, and interfaces for data processing. In a similar fashion, Smarsly et al. (2012) proposed and implemented an agent-based framework for storing structural monitoring data and performing distributed analytics. A benefit of their system is its use of software agents to automate data processing in a distributed computing environment such as in a wireless sensor network. Other notable efforts in data management for structural monitoring applications includes Kijewski-Correa et al. (2013) who created a powerful SHM database platform for tall buildings and Zárate et al. (2013) who proposed a scalable cyber-environment for storage and analysis of non-destructive evaluation (NDE) data (*e.g.*, acoustic emission data). While these references offer innovative means of storing and exposing structural monitoring data for analysis, these systems often do not store detailed structural information regarding the structure nor link structural information to the collected sensor data. This in effect limits the type of analyses that can be done using the sensor data collected. Ultimately, asset managers will require data management systems

to comprehensively combine: 1) structural information (*e.g.*, geometric details, material properties) herein termed structural metadata, 2) life-cycle information (*e.g.*, inspection and rehabilitation history), and 3) sensor data. One system that seeks to overcome the existing fragmentation between structural metadata platforms (*e.g.*, AASHTOWare) and sensing data platforms is the Wind and Structural Health Monitoring System (WASHMS) implemented by the Highway Department of Hong Kong (Wong 2004). WASHMS stores both bridge metadata (in the form of a finite element model) and sensing data.

1.3 Modal Analysis of Monitored Structures

One of the most common analyses conducted on structural accelerations measured by structural monitoring systems is modal analysis. Modal analysis corresponds to analytical tools used to extract the modal parameters of structures from vibration data: modal frequency, modal damping, and mode shapes (Ewins 2000). Modal parameters are intrinsic global structural properties that have found wide spread use in structural engineering. For example, the spectra-based approach to the design of structures rely on estimates of modal frequencies and modal damping. Modal properties are also used as the basis of structural performance assessment. For example, modal properties are used for the updating of finite element models of the structure (Brownjohn et al. 2003; Friswell and Mottershead 1995; Michael et al. 1997). Finally, modal properties have been proposed for damage detection. While there is some debate surrounding the efficacy of modal properties as damage sensitive features in civil engineering structures, significant research efforts have been devoted to modal-based damage detection (Doebling et al. 1998; Salawu 1997).

Dynamic testing methods for modal analysis of bridges are either based on controlled excitation or ambient excitation. In controlled excitation, the bridge is typically closed to traffic with modal shakers installed to excite the bridge. By controlling the dynamic load, accurate input-output dynamic models of the structure can be developed. However, controlled dynamic testing is expensive and can only be performed for short periods of time. Given these limitations, dynamic measurement of operational structures is often performed using ambient vibration source (*i.e.*, wind and traffic induced vibration). The ambient excitation source is assumed to be broadband, zero mean white noise. This assumption is often necessary to apply classical linear system theory. To process ambient excited vibration response data, output-only system identification algorithms are adopted with operational modal parameters extracted from the models produced. System identification algorithms are divided into two groups: frequency- and time-domain methods. Frequency-domain algorithms identify modes based on Fourier spectra or power spectral density (PSD) functions derived from time history data (Brincker et al. 2001; Heylen et al. 1995; Richardson and Formenti 1982). One common output-only system identification method performed in the frequency-domain is the Frequency Domain Decomposition (FDD) algorithm (Brincker et al. 2001). FDD converts time history data to the frequency-domain through a fast Fourier transfom (FFT). There, PSD matrices are assembled for each discrete frequency with singular value decomposition (SVD) of the PSD matrix revealing modal information including modal frequencies and mode shapes. Time-domain algorithms do not rely explicitly on the frequency-domain and perform their operations in the time-domain using time-series data (Cunha and Caetano 2006). Examples of time-domain system identification algorithms include time series models (*e.g.,* autoregressive models), random decrement technique (Ibrahim 1977), and sub-space methods, just to name a few. A time domain output-only system identification method gaining

popularity in recent years is stochastic subspace identification (SSI) (Peeters and De Roeck 1999) which uses the oblique projection of the time history data to compute the Kalman filter state sequence from which the system matrices of a state-space model can be solved using a least square approach. The aforementioned system identification methods used for modal analysis have been applied widely to a broad array of bridges ranging from short-span multi-girder bridges to long-span suspension bridges.

Modal analysis can be performed on limited data sets due to short-term monitoring campaigns or can be applied to long-term response data collected over prolonged periods. Several projects have observed that the operational modal parameters obtained using vibration-based structural monitoring system are not constant and naturally vary within a 5% range around their mean values (Cross et al. 2012; Li et al. 2010; Sohn et al. 2002). Specifically, EOC parameters affecting bridges, including temperature and wind profiles, have been investigated for their environmental impact on structural modal properties. Cornwell et al. (1999) performed force vibration testing (i.e., hammer excitation) on the Alamosa Canyon Bridge (New Mexico), a steel girder highway bridge. They used a wireless structural monitoring system to collect structural accelerations during excitation while also recording deck temperatures during testing. A peak picking method was applied to extract modal frequencies using two 24-hour test data sets. Regression analysis was performed to model the relationship between modal frequency and temperature. Peeters and De Roeck (2001) performed SSI on one-year worth of vibration data collected from the structural monitoring system instrumented on the Z24-Bridge, a concrete box girder bridge in Switzerland. Their study revealed a bimodal relationship between modal frequencies (extracted from the SSI state space models) and temperature with a clear change in slope near the point of water freezing (0 °C); other researchers

have reported similar observations(Moser and Moaveni 2011). Ni et al. (2005) reported regression modeling on modal frequencies and temperature using a support vector machine (SVM) applied to a 6-month period of monitoring data from the Ting-Kau Bridge which is a pre-stressed concrete cable-stayed bridge in Hong Kong. In a similar fashion, Li et al. (2010) reported on the environmental impact on the modal parameters of the Tianjin Yonghe Bridge, a pre-stressed concrete cable-stayed bridge in Tianjin, China. The relationships between modal parameters (modal frequencies, damping ratio and modal assurance criterion (MAC) values of mode shapes) with temperature and wind velocity were modeled using linear regression and artificial neural networks using 16 days of monitoring data. Cross et al. (2013) performed an environmental impact study on modal frequencies using one-year worth of monitoring data collected from the Tamar Bridge which is a steel truss suspension bridge in the United Kingdom. Their study used accounted for the impact of traffic load, thermal loads and wind loads. Conte et al. (2008) and Hong et al. (2011) individually performed short-term dynamic testing of the New Carquinez Bridge, a large-scale suspension bridge in California. Using the acceleration response of the bridge steel box girder, both studies have identified modal parameters of the bridge using the stochastic subspace identification algorithm. The structural response induced by traffic and aerodynamic loads were also investigated to provide evidence of the EOC impact. The abovementioned bridge monitoring campaigns, modal frequencies and mode shapes were identified; however, bridge modal damping ratios were not reported in all cases due to the high uncertainties that exits in damping estimation by many of the system identification algorithms used. Also lacking in many of these studies are quantitative analysis of the contribution of each EOC variable on changes in modal parameter through regression modeling. Better understanding of the modal properties of different types of

bridges under various EOCs would better inform bridge design methods when considering seismic and aerodynamic loads.

Modal analysis is often conducted in a manual fashion but it can also be automated for execution by computers. Manual analysis is beneficial because intermediate results can be interpreted by the operator during the analysis. However, when contending with large data sets, manual approaches are no longer scalable and automation is required. Automation comes with some challenges including automating an interpretation process that yields more accurate modal results. In the family of frequency-domain methods, Brincker et al. (2007) automated the FDD method by applying a peak picking algorithm that identifies modes based on singular values and distinguishes physical and mathematical modes based on the statistics (kurtosis) of each peak. Peaks with large kurtosis are identified as harmonic solutions. With respect to time-domain algorithms, a series of algorithms have been developed to automate the process of model extraction using the stochastic subspace identification algorithm (Häckell and Rolfes 2013; Magalhães et al. 2009; Zhang et al. 2014). These methods take advantage of the observation that physical modes are stable (invariant) across different model orders of the state space model. Magalhães et al. (2009) applied clustering algorithms to extract real physical modes from the SSI results. In their study, SSI solutions across a range of system model orders are clustered based on similarity in modal frequency and mode shape. Each cluster represents one mode and clusters with little elements are discarded as mathematical modes. While these methods aim to remove the need of an operator to manually determine an appropriate system order, they sometimes lack generality and can add to the computational costs of the algorithm implementation (especially in SSI methods where large number of model sizes are attempted leading to orders of magnitude higher computational costs).

1.4 Traffic Load Identification using Weigh-in-Motion Systems

Weigh-in-motion (WIM) stations are a critical component of highway traffic monitoring systems because federal governments (*e.g.,* Federal Highway Administration in the United States) use traffic data to measure the traffic demands on national highway systems and to allocate highway funding (FHWA 2013). As opposed to traditional weigh station where trucks are diverted from the flow of traffic for weight measuring on scales, WIM stations have the benefit of measuring truck loads continuously at highway speeds. WIM stations consist of load measuring sensors (such as bending plate, piezoelectric sensors and load cells) embedded in the road surface to measure axle weight, axle spacing, and vehicle speed. Installation costs range from approximately $22,600 per lane for piezoelectric-type WIM stations to more than $73,900 per lane for load cell-type WIM stations; maintenance costs also ranges from $7,350 to $10,100 per lane per year (Hallenbeck and Weinblatt 2004). The installation and maintenance of sensors embedded in the road surface requires lane closures resulting in distribution to traffic; this adds cost and complexity to WIM station installation and maintenance.

An alternative to installing WIM stations in the road network is to monitor bridges where bridges serve as a crude WIM station. So called bridge WIM (or BWIM) technologies first emerged as a concept in the late 1970's (Moses 1979). Due to the many advances in sensing technologies over the past two decades, the concept of BWIM is re-emerging as a lower cost and potentially easier to maintain WIM platform. BWIM utilizes an existing bridge structure and uses it as a linear system to compute loads based on structural responses. Compared to bending plates or load cells that must be installed in the road, the cost of installing sensors such as strain gages on a bridge is much lower. Additionally, with the sensors installed underneath the main bridge deck in many

12

instances, installation and maintenance of the BWIM sensors does not require traffic interruptions. An early BWIM system consisting of road surface axle detectors and strain gages on steel girders was proposed in 1979 by Moses (Moses 1979). The road surface axle detectors determined the axle counts while the girder strain measurements were used to estimate gross vehicle weight. Later, Snyder and Moses (1985) expanded the system to extract axle weight by solving a linear regression between theoretical influence lines and the measured strain response of the bridge. While their work advanced BWIM methods, the system still required the use of axle detectors in the road (which can be costly to install) and relied on modeling to derived accurate influence lines. More recently, a contactless sensing method was presented using camera imagery to accurately measure bridge displacements under truck loading (Ojio et al. 2016); gross truck weight and axle weight can then be extracted from span displacements using Moses's algorithm (Moses 1979).

A second generation of BWIM concepts that are extensions of Moses work emerged in Europe in the 1990's. These methods aimed to improve the accuracy of load estimates (Jacob and O'Brien 1998; Jacob et al. 2001) . For example, O'Brien et al. (2006) proposed an algorithm to calculate accurate influence lines from experimental sensing data as opposed to reliance on modeling to derive theoretical influence lines. Their algorithm extracted the single-axle influence lines directly from bridge strain responses that correspond to calibrated multi-axle trucks driven over a bridge. Rowley et al. (2009) later applied Tikhonov regularization (Tikhonov and Arsenin 1977) to improve the traditional method by Moses (Moses 1979). The inversion methods in Moses's approach can be ill-conditioned leading to inaccurate estimates for axle weights; hence, Rowley et al. (2009) introduces a regularization term on the norm of the axle weight vector to improve the least square solution. Other researchers have proposed BWIM methods based on influence

surfaces (Quilligan et al. 2002). In these methods, two dimensional influence surfaces enable multiple vehicles to be identified in a more accurate manner.

Another set of BWIM concepts have been proposed based on the coupled dynamics between bridges and moving vehicles; such methods are often classified as moving load identification methods. The earliest moving load identification algorithms modeled bridge spans as second-order dynamic beam systems. Strain or acceleration response of the bridge are used as the system output and truck axles are modeled as the system input using step functions. One approach in the time-domain is to solve the equations of motion by modal superposition methods in order to extract the axle loads of one or two axle vehicles moving at constant speed over a flexible beam (Law et al. 1997). Another approach is proposed in the time-frequency domain method where axle loads are identified in the modal coordinate space (Law et al. 1999). Gonzalez et al. (2008) expands on prior moving force identification methods to account for two dimensional distribution of moving loads. All of these moving force identification algorithms are reliant on modal parameters; hence, accurate modal parameters are necessary to ensure these methods are accurate in identifying moving loads. This approach has a number of challenges. First, to be accurate, model parameters should be found experimentally. Second, experimentally derived modal parameters have uncertainties associated with them (especially modal damping estimates). Finally, modal properties are well known to exhibit high sensitivity to EOC variations (Brownjohn et al. 2010; Hua et al. 2007; Li et al. 2010; Moaveni and Behmanesh 2012).

A major obstacle in BWIM development is the difficulty associated with algorithm validation. Validation must be conducted under in-service conditions with different types of trucks with varying load configurations required to test the robustness of the algorithms. Some of the aforementioned studies adopted validation procedures based on the use of calibrated trucks; but this controlled truck loading strategy prevents large variations in loads that can be observed on the bridge. Some projects were able to use regular truck traffic with trucks later weighed at a weigh station. Weighing trucks that crossed a bridge can be a major effort that takes time and resources.

Some long-term BWIM studies have been conducted on operational bridges. The WAVE project (Jacob et al. 2001) reported on BWIM system instrumentation on the Autreville Bridge, a three span orthotropic steel plate highway bridge located in France. A total of 21 strain gages were install in the bridge and 44 trucks were randomly selected and pulled over after crossing to measure their static axle weights over a two year period. The 1-dimensional algorithm achieves D+(20) class, meaning 95% of the identified gross weight is within $\pm$20% error, and 2-dimensional algorithm achieves C(15) accuracy. Two short 2-span slab bridges of skewness 7° and 26° were also instrumented and tested using calibrated trucks. The systems achieved class D+ and class C accuracy, respectively, where using COST323 Management Committee accuracy ratings (Jacob and O'Brien 1998). A D+ and C rating are generally good rating with the truck weight estimates within 20 and 15% of their true static gross weights respectively. More testing results on different bridges can be found in the WAVE project report (Jacob et al. 2001). Commercial bridge WIM software termed SiWIM was developed from this project and is distributed by Cestel (http://www.cestel.eu/).

In the United States. The University Transportation Center for Alabama has reported its study of BWIM using SiWIM applied to the I-459 bridge which is a 3 span prestressed concrete highway bridge (Hitchcock et al. 2012). During a 2-day in-service validation test, 24 trucks were pulled out of the traffic with SiWIM identifying15 out of the 24 trucks with class C accuracy for gross weight and class E for axle accuracy (class E characteristic weights to within 30% of static gross weight). Wall et al. (2009) reported a field instrumentation of BWIM on a single-span steel girder bridge with 11.5° skewness and 2.56% slope along I-91 in Connecticut. A nearby (0.8km) weigh station was used to obtain the truck static weight data for comparison. In total, 117 truck events were recorded and the bridge WIM system achieved a 95% confidence interval bounded by -27.3% and 23.4% error on estimated fross weights for 109 trucks in the slow lane and bounded by -39.2% and 21.7% for 8 trucks in the passing lane. Collectively, all of these projects illustrate the potential for BWIM. The studies identify optimal selection criteria for bridges, the algorithm options for different systems, and highlight methods to improve the workability of on-site implementation and data collection. At the same time, several aspects of BWIM still need improvement, such as needs for dense sensor arrays, measurement sensitivity to truck path, the weight identification accuracy and the high costs of system calibration and validation.

1.5 Goals and Objectives

The previous sections highlight some of the advances made in the management and analysis of structural monitoring data. While advances have been made, the previous sections also illuminate areas where additional research is needed. Because data management systems are the foundation upon which analytical tools are implemented, the overarching goal of this thesis to advance the state-of-the-art in monitoring data management to ensure powerful analytical tools can be

implemented in an autonomous fashion on large monitoring data sets. Data management systems that can accommodate the increasing amounts of data are needed, especially as structural monitoring systems gain popularity as a structural asset management tool. An integral part of the thesis goal is to also develop a suite of analytical tools that can query data from the data management system to extract actionable information for system end users (*e.g.,* asset managers and owners).

The specific objectives of the thesis can be summarized as follow:

- *Develop a scalable and secure data management system based on the design of a novel hybrid database architecture:* a database server that supports high speed queries on monitoring data will be designed to store and curate relational data pertinent to the structure in a relational database while storing time history sensor data in a hierarchical data format (HDF) repository.

- *Develop a relational database ontology compatible with existing bridge management data standards*: the relational database ontology will be designed to facilitate a broad array of analyses on data (*e.g.,* finite element modeling) while ensuring representations of data are consistent with element-based representations of the structure.

- *Validate the performance of the data management system using data from long-term wireless bridge monitoring systems*: the performance of the data management system will be validtaed by using it to store and process data associated with two long-term wireless monitoring campaigns on the Telegraph Road Bridge (Monroe, MI) and the New Carquinez Bridge (Vallejo, CA). In addition to validating the data management system,

both monitoring systems provide opportunities to learn many valuable lessons on how to deploy wireless monitoring systems in harsh field settings.

- *Analyze long-term bridge monitoring data to model the influence of EOC parameters on modal parameters*: a fully autonomous implementation of the SSI method for modal analysis will be developed to process long-term monitoring data from the New Carquinez Bridge to discover relationships between modal parameters (*e.g.,* modal frequency, modal damping, and mode shapes) and EOC variations.

- *Develop a data-driven approach to BWIM by estimating truck load information using long-term bridge monitoring data*: a robust data-driven analytical framework will be developed to use the strain response of structural components of steel girder bridges to estimate the parameters associated with vehicular loads (e.g., number of axels, axle weights, and vehicle speed). The framework proposed is validated on the Telegraph Road Bridge using a WIM station in close proximity to the bridge as a baseline for assessing the framework.

An outline of the thesis is presented in Figure 1.1. The thesis can be divided into three major parts. First, two testbed bridges that will be used extensively for validation throughout the thesis are introduced. In Chapter 2 (*Long-term Wireless Monitoring Systems on the New Carquinez Bridge and the Telegraph Road Bridge*) introduces the Telegraph Road Bridge in Monroe, Michigan and the New Carquinez Bridge in Vallejo, California; both bridges have wireless monitoring systems installed. A portion of the chapter will be devoted to a discussion on the lessons learned from these two long-term wireless monitoring campaigns. With wireless monitoring still a relatively new technology that has only been deployed on a small number of structures, a distillation of what does and doesn't work is a valuable contribution of the chapter. With two testbed bridges selected, the

second part of the thesis focuses on data management of sensor data. In Chapter 3 (*SenStore: a Scalable Cyberinfrastructure Platform for Implementation of Data-to-Decision Frameworks for Infrastructure Health Management*), a cyberinfrastructure platform termed *SenStore* is presented. *SenStore* has been developed for the management of the heterogeneous set of data associated with bridge monitoring. A hybrid database architecture is presented consisting of a relational database for the storage of metadata (*e.g.,* structural geometry, materials, coordinates, elements, sensors, sensor locations, sensor parameters) associated with a bridge and its monitoring system and a non-relational data repository (*i.e.,* HDF5) for the storage of large tracts of time history sensor data. The third part of the thesis is devoted to the data analytics that ca nbe implemented on top of *SenStore*, For example, the performance of *SenStore* is validated using long-term monitoring data from the Telegraph Road Bridge in Chapter 4 (*Case Study: Automated Modal Analysis and Finite Element Model Updating using SenStore*). The Telegraph Road Bridge case study demonstrates the capability of *SenStore* in enabling automated modal and finite element analyses. Bayesian model updating using adaptive Markov Chain Monte Carlo (MCMC) simulation is applied to update finite element model parameters using temperature normalized modal parameters. With the capabilities of *SenStore* established and validated, the thesis then turns its attention to the development of two novel analytical frameworks for the processing of long-term bridge response data. In Chapter 5 (*Automated Long-Term Modal Analysis of Suspension Bridges under Varying Environmental Conditions using Wireless Structural Monitoring Systems*) automated modal analysis for the modeling of modal dependencies on EOC parameters for long-span bridges is developed. The SSI algorithm is adopted for the identification of output-only state-space models from which modal frequency, modal damping ratio and mode shapes are extracted. A rigorous framework that standardizes SSI algorithm parameters (*e.g.,* size of Hankel matrix and model order

selection) is proposed to ensure the SSI method is applied in a consistent manner across the large data set. Consistency in the modal results is vital for aggregation of modal properties over long periods of observation for regression against EOC parameters. Ridge regression and Gaussian process regression (GPR) are applied to model relationships between modal parameters and EOCs. Chapter 6 (*Data-driven Regression Analysis for Bridge Weigh-In-Motion*) presents a new approach to BWIM which applies regularized regression analysis on the strain response of a critical bridge element to extract reliable estimates of truck load parameters including truck speed, axle counts and axle weights. The proposed BWIM algorithmic framework is applied to the Telegraph Road Bridge for which load data recorded at a traditional WIM station in close proximity is readily available. Finally, the thesis concludes with Chapter 7 (*Conclusions and Future Work*) which summarizes the key thesis findings and highlights areas where future work is needed to further progress data-driven analytics for structural asset management applications.
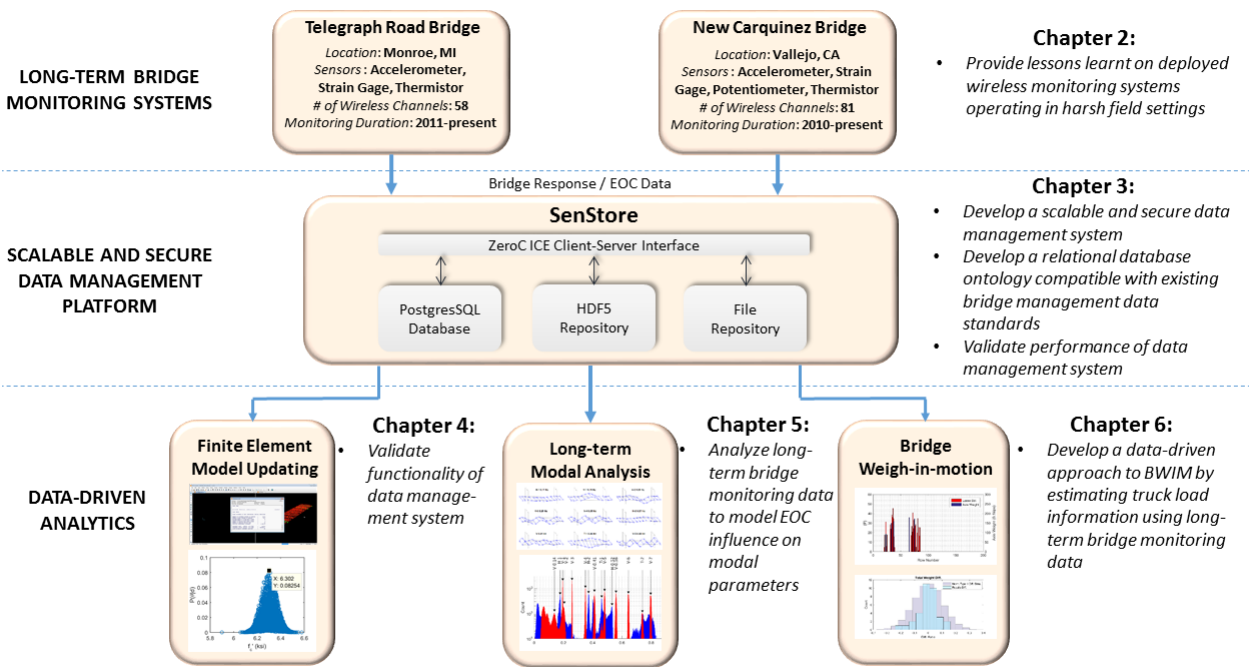
Figure 1.1, Thesis outline: chapter layout and key objectives

CHAPTER 2

LONG-TERM WIRELESS MONITORING SYSTEMS ON THE NEW CARQUINEZ BRIDGE AND THE TELEGRAPH ROAD BRIDGE

2.1 New Carquinez Bridge

The New Carquinez Bridge (Figure 2.1), also named the Alfred Zampa Memorial Bridge, is a long-span suspension bridge that spans over the Carquinez Strait between Crockett, CA and Vallejo, CA. The bridge carries four lanes of westbound traffic on I-80 between Sacramento and San Francisco. The bridge, whose construction was completed in 2003, is the first suspension bridge to be constructed in the United States since 1973. The bridge design consists of a longitudinally continuous closed orthotropic steel box girder 29 m (95 ft) wide and 3 m (9.8 ft) deep with a cross-section designed to be aerodynamically stable (OPAC 2016). The steel box girder and suspension

Figure 2.1, Photo of the New Carquinez Bridge in Vallejo, CA

cables are supported by two concrete towers that are designed to exhibit high ductility under the high seismic demands common for this area of California. The bridge span is divided into three segments that measure 147 m (482 ft), 728 m (2388 ft) and 181 m (594 ft) when going from south to north (Figure 2.2). The main suspension cables consist of 8,584 carbon steel wires (zinc coated) stranded together to form a 51.2 cm (20 in) diameter cable encased in a corrosion protective skin. The vertical load of the deck is transferred to the suspension cables using vertical suspenders that are 12.4 m (40.7 ft) apart along both sides of the bridge span. The bridge is designed for a service life of 150 years with wind and seismic loading considered. The design wind loads were based on wind speeds corresponding to a 100 year return period and drag, lift and flutter coefficients derived from wind tunnel tests of the bridge span and tower cross sections (OPAC 2016). The seismic design was based on nonlinear time history analysis of the bridge using safety evaluation earthquake (SEE) and functional evaluation earthquake (FEE) records with return periods of 1000-2000 and 285-300 years, respectively (OPAC 2016). Soil-structure interaction and spatial variations in the earthquake record were considered during the nonlinear time history analysis.
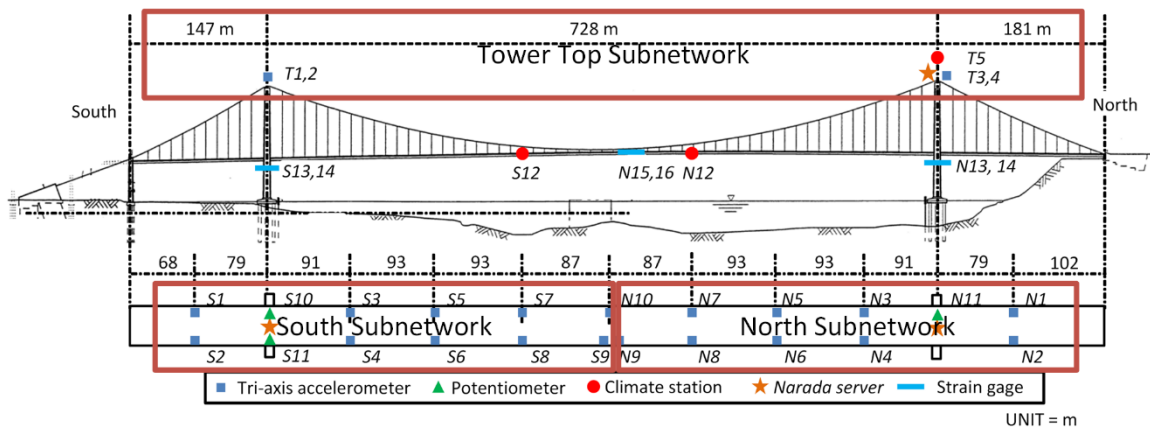
Figure 2.2, Sensor instrumentation layout of the *Narada* wireless monitoring system on the New Carquinez Bridge

A permanent seismic monitoring system was installed on the New Carquinez Bridge by the California Strong Motion Instrumentation Program (http://www.conservation.ca.gov/cgs/smip) shortly after construction. The CSMIP monitoring system consists of 76 force balance accelerometers (installed along the main span, towers, and foundations) and 1 anemometer all wired to a central data acquisition system (Hong et al. 2011). The wired monitoring system is operated based on triggering from ground motions but manual data collection is possible. A separate wireless structural monitoring system was also installed on the New Carquinez Bridge in 2010 with iterations to the monitoring system configuration occurring through 2011 (Kurata et al. 2013). The wireless monitoring system is designed to collect data on a regular basis (*e.g.,* based on a schedule) but is not designed to be triggered based on motion. The bridge response data collected by the wireless monitoring system is the primary focus of this thesis.

The NCB is an ideal bridge for instrumentation and study. For example, the NCB offers good accessibility for sensor instrumentation. First, there are travelers (*i.e.*, movable platforms) underneath the bridge girder which makes it possible to install wireless sensors along the girder. Second, there are elevators inside the concrete tower which permits equipment to be carried to the top of the towers. Third, there are electrical outlets at certain locations in the bridge which provides substantial convenience during each stage of the installation.

The bridge is subject to live loads such as wind load, heavy traffic load and occasionally seismic load. The different excitation sources and the induced response of the bridge provides valuable data to study the behavior of large-scale suspension bridges. This chapter will describe the design

and installation of the wireless monitoring system installed on the NCB by the thesis author as part of this thesis.

2.1.1 Long-Term Wireless Structural Monitoring System

2.1.1.1 System Overview

The wireless monitoring system was designed based on the *Narada* wireless sensor platform developed at the University of Michigan for structural monitoring applications (Swartz et al. 2005). *Narada* support data acquisition on 4 sensing channels with a 16-bit analog-to-digital conversion (ADC) resolution. The platform communicates data through a power amplified (10 dB gain) IEEE802.15.4 wireless transceiver providing a communication range of 500 m line-of-sight and a communication data rate of 250 kbps. *Narada* has 128 kB of flash memory for program storage and an additional 128 kB of on-board random access memory for temporary data storage. When *Narada* is active including transmitting data at the maximum output power setting on the transceiver the node consumes 375 mW of power; this power consumption can be reduced to 215 mW by placing the radio into a low-power sleep mode. *Narada* has been successfully used to monitor (in both permanent and temporary capacities) a variety of structures including bridges (O'Connor et al. 2016), buildings (Kim and Lynch 2012), ships (Swartz et al. 2012) and wind turbines (Swartz et al. 2010).

The long-term wireless structural monitoring system for the NCB was designed to record bridge responses (*i.e.,* strains, displacements and accelerations) and environmental conditions (*i.e.,* temperature, wind speed and wind direction). In total, 35 *Narada* sensor nodes collecting data

from 76 sensing channels (Figure 2.2) were installed. To measure the vertical and transverse motion of the orthotropic steel deck along its length, 19 tri-axial accelerometers (MEMSIC CXL02TG3) were installed. At the top of each tower, two tri-axial accelerometers (MEMSIC CXL02TG3) measuring one vertical, one transverse and two longitudinal degrees-of-freedom were also installed. To measure strain of the orthotropic steel deck, metal foil strain gages (Tokyo Sokki Kenkyujo FLA-6-11-3LT) were mounted to the interior surface of the box girder at 3 locations: over the north tower, over the south tower, and approximately at mid-span. At each location, three strain gages were installed to measure longitudinal strain on the top and bottom girder surfaces and transverse strain on the top surface. For thermal compensation, three thermistors (National Semiconductor LM35DT) were installed to measure the top and bottom girder surface temperatures and the ambient internal air temperature of the girder. At the girder midspan, one additional strain gage was installed at a bolted plate connection but the bottom girder temperature was not measured. To capture relative longitudinal displacement of the steel orthotropic deck relative to the concrete towers, three string potentiometers (Celesco SP2–25) were installed: two at the south tower and one at the north tower. In addition, three weather stations were installed each consisting of one anemometer (NRG Systems 40H). Also included at each weather station were two additional environmental sensors consisting of either a wind vane (NRG Systems 200P), thermistor (National Semiconductor LM35DT) or humidity sensor (TDK CHS-UPS). In total, 46 channels of acceleration, 10 channels of strain, 3 channels of displacement, and 17 channels of environmental parameters (*i.e.,* wind speed, wind direction, temperature, humidity) have been installed.
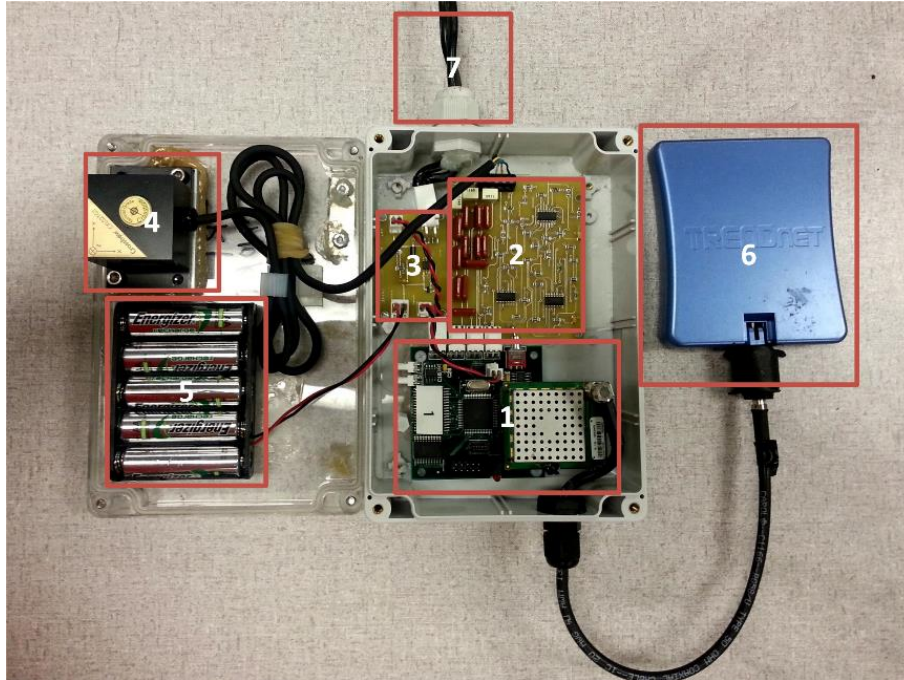
Figure 2.3. Fully packaged *Narada* wireless sensing unit: 1) *Narada* with amplified IEEE802.15.4 radio, 2) anti-aliasing signal conditioning board, 3) solar charge controller, 4) tri-axial accelerometer, 5) rechargeable NiMH battery pack, 6) directional antenna and 7) connection to solar panel

Each *Narada* node (Figure 2.3) has one power amplified long-range IEEE802.15.4 transceiver, one directional antenna, and up to four sensing transducers (*e.g.*, accelerometer, strain gage, etc.) attached. Signal conditioning boards designed for each transducer including anti-aliasing filters, amplifiers, and Wheatstone bridge circuits are adopted to condition sensor outputs before interfacing to the *Narada* sensing interface. All *Narada* nodes (except those measuring strain and displacement) are powered by 5 AA rechargeable nickel-metal hydride (NiMH) batteries (Energizer NH15-2300) with a net nominal voltage of 6 V and 2300 mAh capacity. The node battery pack is continuously charged during daylight hours using a single 5W solar panel (UL Solar 5W 12V) and a custom solar charge circuit. The *Narada* nodes collecting strain and displacement are plugged directly into the bridge electrical system and hence do not have batteries or solar panels. Each node was enclosed within a weatherproof polycarbonate enclosure (Bud

Industries PN-1324-C) but the directional antenna, sensing transducer and solar panel are outside of the enclosure. The nodes installed along the length of the girder are magnetically mounted to the steel box girder while the nodes on the top of the towers are epoxy bonded to the concrete surface.

The wireless monitoring system is divided into three sub-networks (termed the north, south and tower sub-networks) with each sub-network consisting of a collection of *Narada* nodes communicating data to a single base station in a hub-spoke network topology. In Figure 2.2, the nodes assigned to the north, south and tower sub-networks are labeled with the number N*X*, S*X*, and T*X*, respectively (where *X* denotes the node number). Each base station consists of a single-board PC104 computer (WinSystems) that operates the Ubuntu 10.04 LTS operating system. The base stations each have a IEEE802.15.4 transceiver for communication with the *Narada* nodes and a 3G cellular modem for communication with the Internet via a cellular network (Sprint). The north and south sub-network base stations are powered by the bridge electrical system while the tower base station is powered by a solar panel (UL Solar 110W 12V) that charges a 12V sealed lead acid battery (AGM SLA-12V35) using a charge controller (SunSaver 10L). The base stations were packaged in a weatherproof polycarbonate plastic enclosure (BUD Industries NBB-15253) and installed on the bridge. The system is capable of performing scheduled or remote manual data acquisition. Based on a preliminary dynamic study of the bridge (Kurata et al. 2010), data acquisition is scheduled to collect data at 20 Hz for 8 minutes. During favorable solar charging conditions (*i.e.,* daylight hours) the system performs sensing hourly or every other hour; otherwise, sensing is performed every four hours to limit power consumption. The detailed power consumption breakdown is summarized in Table 2.1. The energy demand of the node is 903 mAh

Table 2.1 Power consumption of *Narada* node

	Duration	**Power (mW)**	**Voltage (V)**	**Consumption (per day)**
Active	8 min/hour ×12 hours = 96 min [**]	375	6	96 min*375 mW/6V/60 min/hour = 100 mAh
Sleep	24 hours × 60 min/hour – 96 min = 1344 min	215	6	1344 min*215mW/6V/60 min/hour = 803 mAh
Total				100 mAh + 803 mAh = **903** mAh

[**] Assuming sensing every 2 hours for 8 minutes

per day. The battery pack included with each node has a capacity of 2300 mAh. Hence, the battery

is capable of powering the node for 2.5 days if no solar power was available.

2.1.1.2 Phase One Installation (October 2010)

The first wireless sub-network was installed underneath the girder box along the north side of the

bridge in June 2010 as shown in Figure 2.2. In this system, 11 *Narada* nodes are installed evenly

along the north half of the girder. 10 *Narada* node were connected to tri-axial accelerometers

(MEMSIC CXL02TG3) to measure the vibration response of the bridge. As previously mentioned,

each *Narada* node along with its accelerometer, signal conditioning board, battery pack and solar

charging circuit is packed into a weather proof enclosure. The *Narada* node package is mounted

to the bottom surface of the bridge girder using rare earth magnets (Figure 2.4 (a)). The directional

antenna of the node is then attached to the bottom surface of the enclosure to allow the antenna

direction to be adjusted easily. Another *Narada* node was used to collect the weather profile of the

bridge from anemometer, wind vane and humidity sensor; the detailed installation is shown in

Figure 2.4 (b). Each node is recharged by a 3.3W solar panel. The solar panels are magnetically

mounted on the deck surface with a 60° angle facing south as shown at the bottom of Figure 2.4

(b). About 20 ft wire (gauge 18) is used for each solar panel to convey the power to the *Narada* node located on the bottom surface of the girder.

(a)
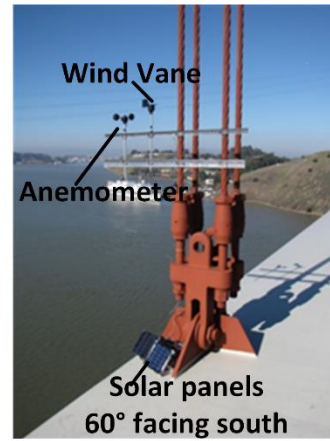
(b)
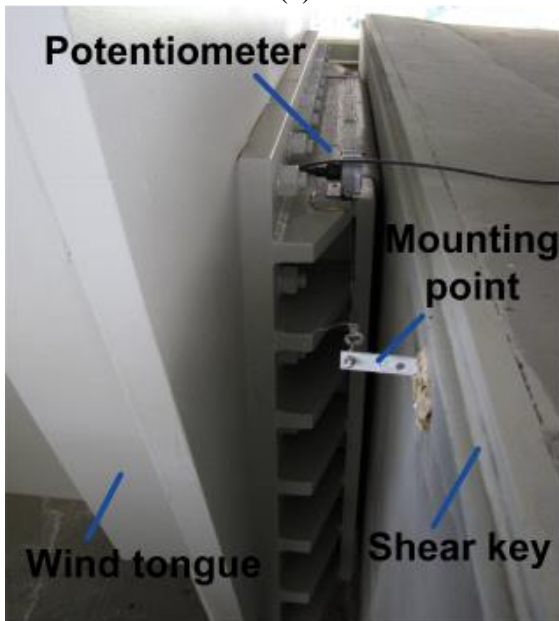
(c)
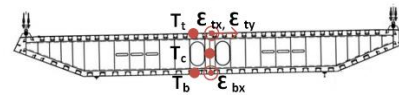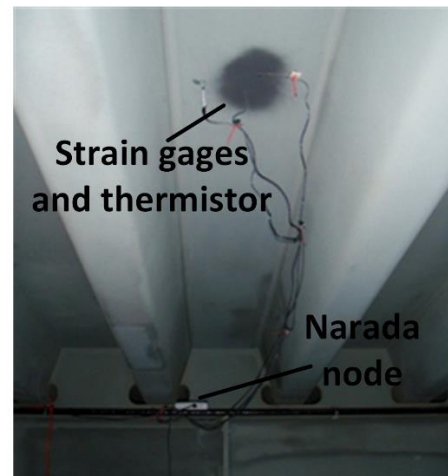
(d)

Figure 2.4 Photos of sensor instrumentation: (a) accelerometers installed underneath the steel box girder, (b) anemometer and wind vane installed on the suspender and the solar panels installed on the deck surface, (c) string potentiometer installed at the wind tongue and (d) strain gages installed inside the girder box

2.1.1.3 Phase Two Installation (January 2011)

The second and third wireless sub-network were installed during January 2011 (Figure 2.2). The second sub-network was installed along the south side of the bridge girder and the third sub-network includes *Narada* nodes installed on the top of the north and south concrete towers to capture the tower dynamic properties. *Narada* nodes and server stations are configured to be consistent with the first sub-network. For example, 9 *Narada* nodes with tri-axial accelerometers are installed along the main deck in a configuration mirroring the north subnetwork (see Figure 2.2). The third subnetwork consisted of the two tower tops with two tri-axial accelerometers installed on each tower top. Another two weather stations are also installed at the south side and on the top of the north tower (and assigned to the south and tower subnetworks, respectively). Therefore, weather profile (wind speed, wind direction, humidity, temperature) of the bridge is measured at three different locations. The south subnetwork is serviced by a base station with cellular service at the south tower; similarly, a third base station is installed on the north tower for the tower subnetwork. To achieve the time synchronization between the three subnetworks, the three *Narada* base station are configured to use the Network Time Protocol (NTP). The three subnetworks are scheduled to take measurement at hour 1, 5, 9, 13, 17, 21 every day for 8 minutes at 20 Hz.

2.1.1.4 Phase Three Installation (May 2012 to March 2013)

More *Narada* nodes with different transducers are added into the long-term monitoring system between May 2012 and March 2013, in order to capture more bridge responses. Three string potentiometers (Celesco Position Transducer SP2-12) are installed at the wind tongue of the bridge

as shown in Figure 2.4 (c); two at the south tower (east and west sides) and one at the north tower (west side) to measure the displacement of the steel box girder in the longitudinal direction (relative to the towers). In addition, 10 metal foil strain gages (Tokyo Sokki Kenkyujo FLA-6-11-3LT) are installed inside the girder box at three locations near the north tower, near the south tower and near the middle of the center span, measuring longitudinal and transverse strains on the top of the girder box (Figure 2.4 (d)) and only longitudinal strain on the bottom of the girder box. The strain gage locations along the girder length are carefully chosen to capture points of maximum flexural moments on the girder. Thermistors are installed on the steel surface near the metal foil strain gages for thermal compensation and to measure the structural temperature. During phase three, the measurement frequency of the monitoring system is increased to acquire more data at different time of the day. The system is configured to collect data every hour from 06:00 to 20:00 and every three hours from 20:00 to 06:00. Each measurement period is for 8 min and is sampled at 20Hz just as before.

After examining the first two years (2010 to 2011) of monitoring system performance, some hardware upgrades were applied during the phase three installation. These upgrades aimed to achieve a more robust system performance and better signal quality. The major weakness of the system during the first two years was the performance of the *Narada* nodes during winters. In the San Francisco Bay Area, solar radiation from sunshine is usually adequate all year round. However, there is a significant rain season in the area during the winter; during this time, the solar panels may not be able to get enough energy. In this case, the *Narada* nodes may still operates, but with very low energy supply. If the energy supply gets too low, operation may malfunction leading to the loss of wireless communication until the node can reset. The node design used in installation

phase one and two only allowed for reset by physically turning the node off and on using the power switch. To avoid this issue, a low voltage cut-off (LVCT) solar charge controller is designed and implemented during phase three to prevent the *Narada* nodes from malfunctioning when there was a low power supply. The LVCT circuit has two major components, an integrated circuit (IC) (Microchip MCP100) and a diode. The MCP100 compares the battery pack voltage and cuts off the load (*i.e.*, *Narada* node) when the voltage is less than 5.7V. The MCP100 also has hysteresis built-in by turning the load back on when the battery voltage increases to 6.2 V. The MCP100 protects *Narada* nodes from a low energy supply by temporarily sacrificing the sensing routine, but this is acceptable because it dramatically improves the system long-term durability. In additional, the solar panels are upgraded during phase three to 5W solar panels; these new panels are prepackaged in aluminum frame for mechanical mounting.

Also during phase three, an active 4-pole butterworth analog signal filter board with adjustable amplification is designed to replace the previous passive filter (Figure 2.3). The new signal conditioning board largely reduces the noise level of measurement signals by rejecting below 0.001 Hz and above 20 Hz. The new signal conditioning board is designed to be physically compact to fit into the *Narada* node enclosure.

2.1.2 Review of System Performance and Observed Vulnerabilities

The system performance of the long-term wireless monitoring system on the NCB improved largely over five years (2011-2015) via the upgrade of both software and hardware of the system over its years of service. Figure 2.5 shows the availability of each node of the bridge monitoring

system from January 1st 2011 to December 31st 2015. Each dot in the figure is when a node logs

data to the base station; absence of a dot implies a node was unsuccessful in reporting its data.

During the phase one and two installations, the system performance was not as desired leading to

many lessons learned and corresponding system changes. However, after the implementation of

various hardware and software improvement in 2012, the system performance was largely

improved, reaching 85% node availability for long periods of time, especially during the non-

winter months. Even still, regular maintenance trips have been needed to repair failed nodes and

to upgrade the base station firmware. The following paragraphs highlight some of the key lessons

learns.

System wide subnetwork outages are usually caused by failure of the base station due to a

connection problem with the wireless radio (*i.e.*, the computer failed to register the USB *Narada*

receiver correctly or when the receiver was damaged due to external disturbances such as birds).

Figure 2.6 (a) shows a damaged receiver radio board. The receiver radio board coaxial connector

for the antenna was soldered on to the radio board top. As external forces (*e.g.*, birds or accidental

human operation) was applied on the coaxial cable, the connector has a tendency to pull off the

radio board leading to the disconnection of the antenna (and the loss of communications). Similar

damage can also happen to coaxial cable extension connector where it enters the antenna base as

shown in Figure 2.6 (b). Again, this type of damage leads to the loss of wireless communications

between the base station and the nodes.
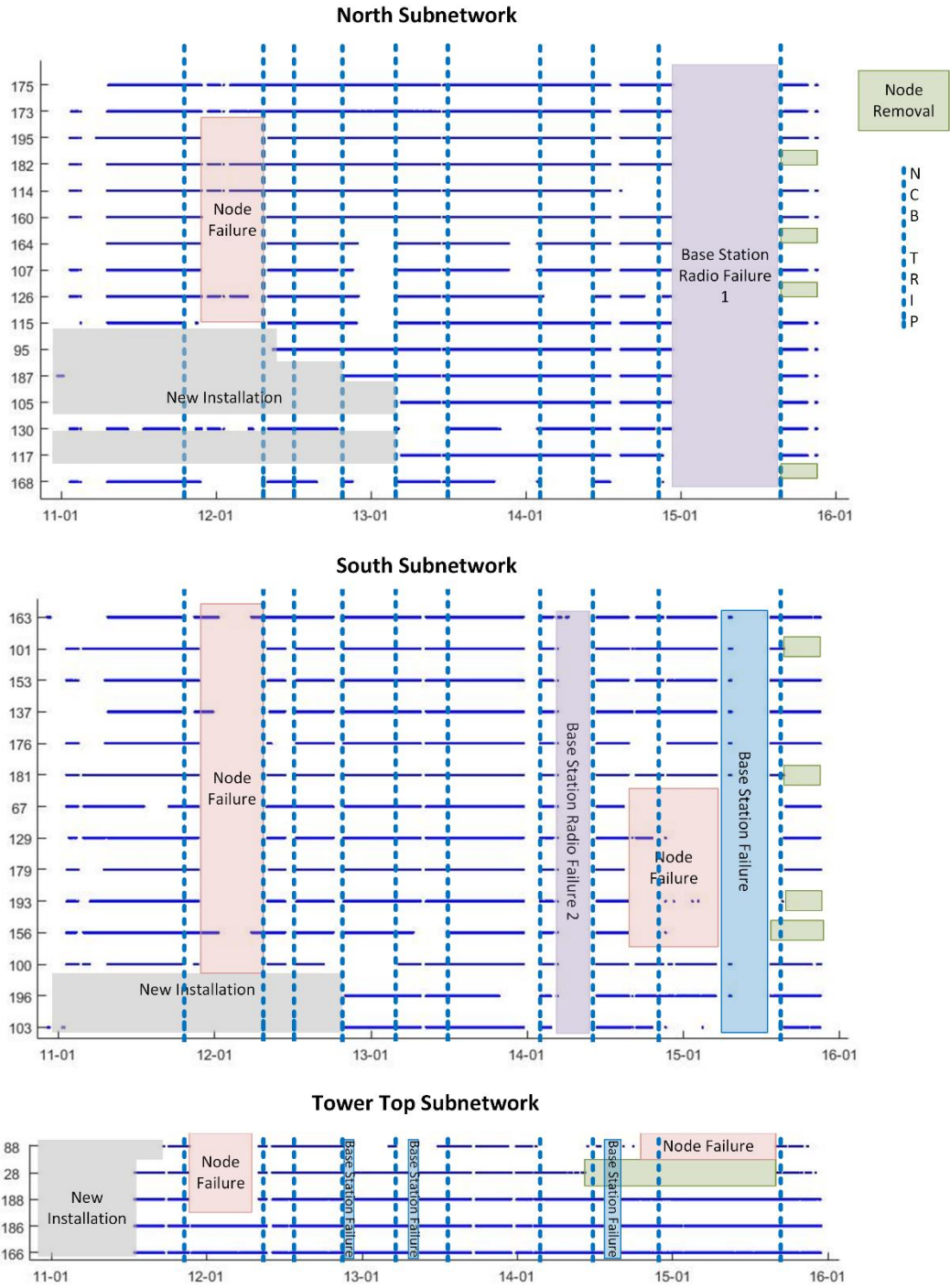
Figure 2.5, *Narada* node and server performance plot
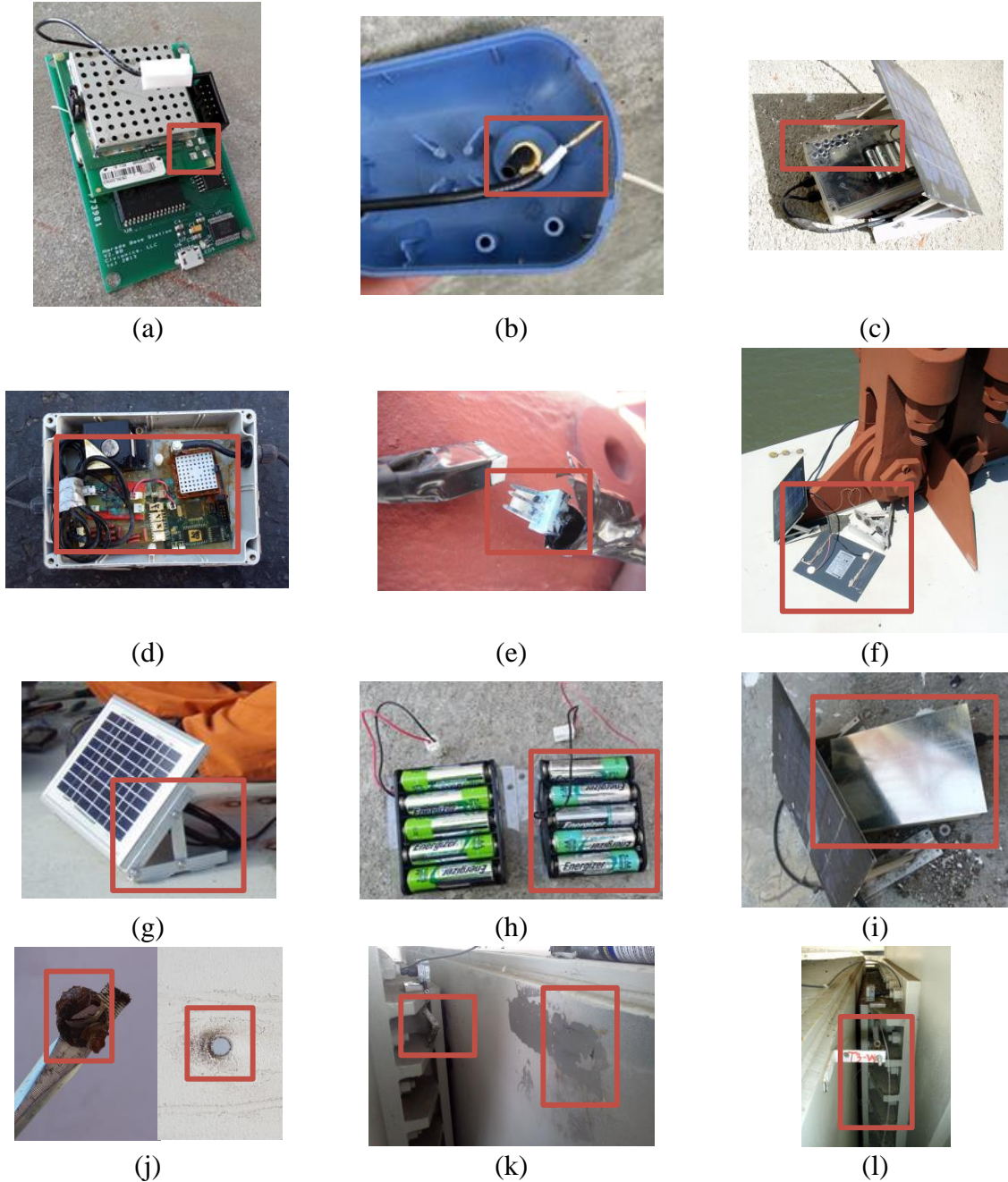
Figure 2.6, Photos of monitoring system failure modes: (a) broaken raido connector, (b) broken antenna extension, (c) water condensation inside *Narada* node enclosere, (d) severe water damage to *Narada* node, (e) corroded connector, (f) solar panel broke off from the brakcet, (g) new solar panel mechanically bonded to brackets, (h) over heated battery pack, (i) reflective shield for *Narada* node, (j) corroded maganet on the tip of a scewdriver and corroded bridge surface, (k) string potentiometer braket broke off from adhesive bonding and (j) broken string from the potentiometer

After the first phase of installation, some nodes failed over time due to two reasons: water damage and low power. While water tight enclose are used to package the nodes, water damage occurred during the first winter. Water was able to penetrate the enclosures through glands installed on the side of the enclosures that allow wires for the solar panel and antenna to enter. To solve this issue, wires entering the enclosure were later placed in a manner to ensure gravity pulls water away from the enclosure. In addition, all glands were also epoxied to enhance their water tightness. A second mode of water intrusion occurred if the water tight enclosures were packaged in a humid environment. Humidity inside the enclosure can lead to condensation inside the enclosure. In Figure 2.6 (c), you can see condensation on the top surface of the enclosure. During later installation silica gel packets were included inside each enclosure to absorb water from humid air. Irrespective of origin, the results of water intrusion can be catastrophic including corrosion of the circuits (Figure 2.6 (d)) and wire connectors (Figure 2.6 (e)).

Some of the nodes also failed after the first phase of installation due to power issues. During the winter when there is less solar power (due to shorten days and inclement weather), the batteries of the nodes are not replenished by the solar panels. This can result in an eventual decrease in battery voltage. When the battery voltage drops below the required voltage of the node's regulator, the microprocessor is affected in a manner that requires a physical turning off and on of the unit to resume normal operations. Even if the battery voltage returns to a higher, acceptable level, without a manual cycling of the power the unit will not operate. To remedy this problem, a low-voltage cut-off circuit was added to the node design which eliminated this issue. The low voltage cut-off circuit detects when the battery voltage is approaching the lower threshold of the regulator and cuts power to the *Narada* board. When voltage rises, the circuit turns the *Narada* board back on.

During the third phase of installation, new units used the low-voltage cut-off circuit and previously deployed units were retrofitted with the circuit.

Another issue encountered in 2015 was the loss of solar panels. The solar panels were attached to the angled brackets initially using glue. Under prolonged exposure to the weather, the glue broke down allowing the solar panels to disengage from the brackets. Under strong winds, the panels would blow off the bridge tearing the wire connecting the panels to the nodes enclosures. Figure 2.6 (f) shows a flipped solar panel that debonded from its bracket. A total of 10 panels were lost in 2015 due to this problem. At that time, all of the panels were retrofitted with a bolted connection to their brackets. A picture of the improved bolted connection of the solar panel to the bracket is shown in Figure 2.6 (g).

It was found that direct exposure of the battery packs to sunlight during hot summer months can lead to catastrophic destruction of the batteries. Figure 2.6 (h) shows on the right a damage battery pack and on the left a new battery pack. To solve this problem, *Narada* enclosures on the top surface of the bridge with direct exposure to sun were fitted with a thin reflective plate. Figure 2.6 (i) shows on such *Narada* node with reflective surface added. This repair was performed during the third installation phase.

In 2014, the rare-earth metal magnets used to bond the *Narada* nodes and solar panels to the steel bridge surface were found corroded. The corroded magnets not only weaken the bonding force, but also cause damage to the bridge painted surface, defeating the original purpose of using

magnets instead of adhesive bonding. Figure 2.6 (j) left shows a corroded magnet and right shows a damage bridge paint surface, bare metal being exposed after removing the magnet. This problem was solved by adding a thin layer of tape (*e.g.*, electric tape) between the magnets and the bridge surface preventing the magnets corrosion spreading to the bridge surface.

Transducers are subject to physical damage as well. Among the various sensors installed on the NCB, string potentiometers have failed several times. The base of the string potentiometer is mounted on the wind tongue and the end of the string is mounted to the shear key. The string potentiometer has failed due to two different failure mechanisms. First the bracket used to hold the string to the shear key has debonded resulting in the string returning to the potentiometer housing. Figure 2.6 (k) shows the failed bracket end. More epoxy has been used to resolve this issue. A second failure mechanism encountered is due to the birds sitting on the potentiometer string. The location of the string potentiometer is a known nesting area for birds as is evident from bird droppings near the shear key. No immediate remedy has been made as this failure mechanism has only occurred once. Figure 2.6 (l) shows the failed string potentiometer due to a bird.

2.2 Telegraph Road Bridge

The second bridge to be extensively used in this thesis is the Telegraph Road Bridge (TRB). The TRB is a highway bridge that carries three lanes of I-275 traffic over Telegraph Road in Monroe, Michigan. The bridge was built in 1973 and has a total length of 68 m (224 ft) as shown in Figure 2.7. The bridge consists of three spans with one main span of 43 m (140 ft) supported by two abutment spans that are each 13 m (42 ft). The bridge has a skewed (57°) reinforced concrete deck

supported by seven steel girders. The main bridge span and the abutment spans are connected through pin and hanger assemblies (as annotated in Figure 2.7) at each of the seven girders. These assemblies consist of steel link plates pinned at one end to the main span and the abutment span at the other; these assemblies fundamentally transfer vertical load from the main span to the abutment spans through axial tension in the plate. The pin and hanger detail was a popular design approach for steel girder bridges up to the 1980s due to their easy design and low cost construction. The bridge sees large loads including heavy trucks associated with the automotive manufacturing industry prevalent in the area. The location of TRB also exposes it to harsh winter weather from December to March each year. Since 2011, the lowest temperature at the bridge site was -26 ℃ and the maximum 24-hour snowfall was 20.3 cm (8 in).

TRB is currently inspected bi-annually by the Michigan Department of Transportation (MDOT). Structural deterioration has been noted by bridge inspectors in the past. For example, material loss on the concrete deck, corrosion of the pin and hanger connections, a severely spalled abutment structure, and web fatigue fracture are just some of the major deteriorations observed during past

Figure 2.7, Side view of the Telegraph Road Bridge located in Monroe, MI (pin and hanger connections are pointed out in the figure)

inspections. These forms of bridge deterioration are typical among similar types of skewed, steel girder bridges. In 2011, the deck condition was sufficiently poor that the top of the existing deck was replaced with high strength concrete with an average compressive strength of 48 MPa (7 ksi). This deck repair increased the deck strength and may have changed the global properties of the bridge (*e.g.*, deck stiffness).

A long-term wireless structural monitoring system was firsly installed on TRB in 2011 (O'Connor et al. 2012). The system was designed to monitor the acceleration and strain response of the bridge for global structural analysis and local damage quantification. The *Narada* wireless sensor node (Swartz et al. 2005, 2007) was installed as the primary data acquisition unit of the TRB wireless monitoring system. The monitoring system was also installed with a single board computer (PC 104) serving as the on-site base station. The base station runs the Ubuntu operating system and has Internet access via a 3G phone card. The wireless *Narada* nodes (and their interfaced sensing transducers) and the base station are all powered by solar energy. Figure 2.8 shows the deployment plan of the TRB monitoring system. The system includes 14 accelerometers installed along the external steel girders (girder 1 and 7) to monitor bridge vibrations from traffic, one tri-axial accelerometer installed in the center of the main span to measure vertical and transverse acceleration, 24 strain gages installed at six locations to monitor the flexural strain response of the main and abutment spans, 16 strain gages are installed at two pin and hanger connections to evaluate the fatigue life of the link plate at each of these connections and 1 strain gage installed one the hanger plate near the bottom pin hole of the middle girder (see Figure 2.9) for bridge weigh-in-motion (BWIM) station application (see Chapter 6). Furthermore, six channels of

thermistors were installed for temperature correction of the strain gages and to measure the structural temperature of the girders.
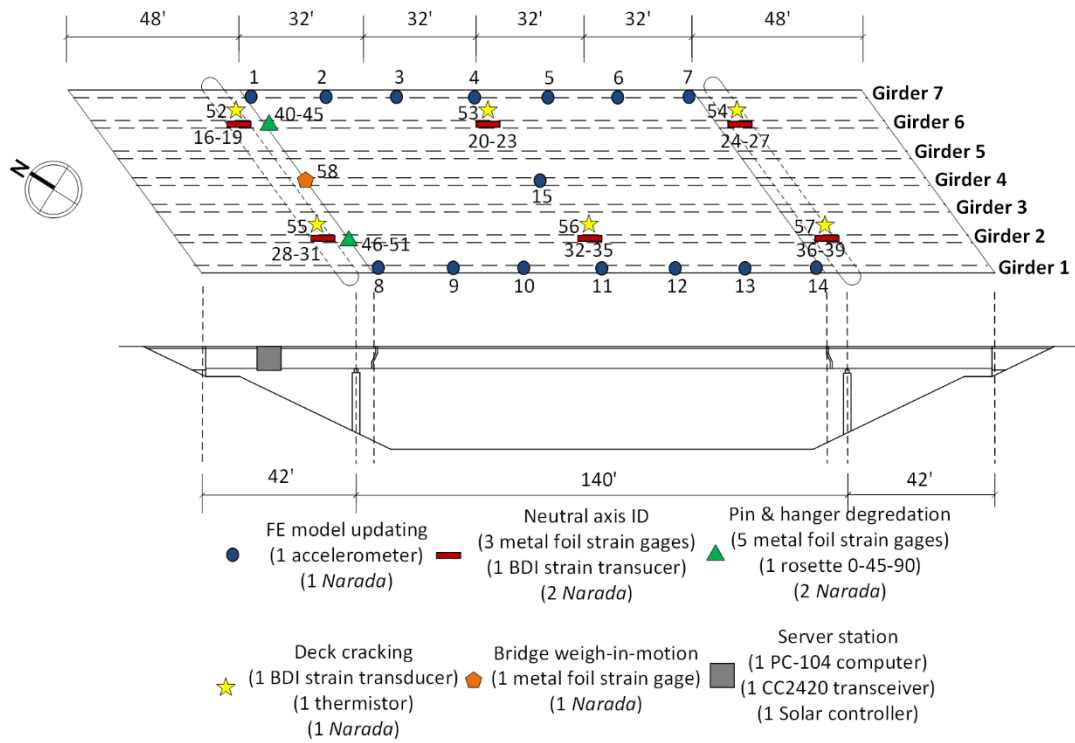
Figure 2.8, Wireless sensor installation plan on the Telegraph Road Bridge
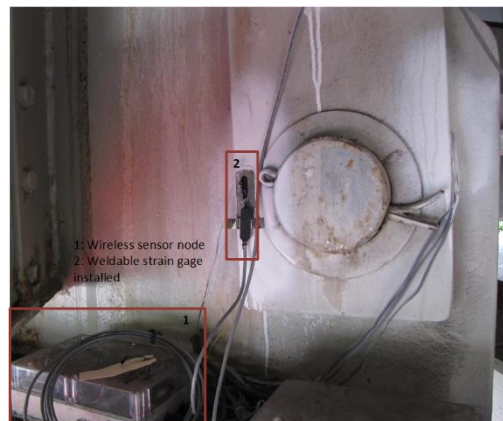
Figure 2.9, Installation details on strain gage near bottom pin hole of middle girder

The long-term wireless sensor network is configured to perform data acquisition at 200 Hz for accelerometers and 100 Hz for strain gages for 1 minute every 2 hours daily. However, during the winter months (December - March) the system is only run when there is sufficient solar energy. This can vary from two to four collections of data per day (as opposed to twelve during the warmer months of April to November). Once wireless sensor data is collected by the TRB base station, it is communicated via a 3G cellular connection to a central data server, *SenStore* (see Chapter 3), implemented on an Ubuntu server (12.04 LTS) at the University of Michigan. The long-term wireless structural monitoring system on TRB alone generates nearly 3 MB of data each measurement cycle, equivalently 7 GB of data each year.

A traditional WIM station is installed 12 miles downstream of the TRB on the I-275 corridor; the WIM station is managed by the Michigan Department of Transportation. It is a 2-lane type II quartz station that measures vehicle direction, class, gross weight, speed, number of axles, axle spacing and axle weight. The station has $\pm 30\%$, $\pm 20\%$, $\pm 15\%$, $\pm 2 km/h$ ($1 mph$) and $\pm 150 mm$ ($0.5 ft$)tolerance for 95% probability of conformity for axle load, axle group load, gross vehicle weight, speed and axle spacing, respectively. There are four highway entrances/exits along the I-275 highway between the TRB and the WIM station. The WIM station is one mile before the next exit, Eureka Road, which leads to the Detroit Metropolitan Airport. The WIM station provides insight to the vehicular loads TRB sees. It will also be a valuable tool for validation a BWIM solution presented in Chapter 6.

CHAPTER 3

SENSTORE: A SCALABLE CYBERINFRASTRUCTURE PLATFORM FOR IMPLEMENTATION OF DATA-TO-DECISION FRAMEWORKS FOR INFRASTRUCTURE HEALTH MANAGEMENT

3.1 Introduction

Cost-effective and dense low-power wireless sensor networks have been proven effective in offering the ability to serve as long-term structural monitoring systems for the evaluation of structural performance and health (Lynch et al. 2009). While such systems offer tremendous promise, they also have the potential to produce large volumes of sensing data. Comparatively less work has been conducted on data management and processing architectures that can harness the full potential of the data collected. Furthermore, to extract meaningful information from the data, decision makers still urgently require computational tools that can automate the processing of data to make structural assessments. A secure cyberenvironment, customized for structural condition monitoring applications, is required to manage not only large volumes of sensing data, but also structural metadata (*e.g.*, the structure's geometric details, material properties, inspection history). Previously, several tools have been developed by researchers that begin to address these needs including sensor data storage using single relational database systems (Allen et al. 2003; Cross et al. 2012; Koo et al. 2011; Magalhães et al. 2012). The objective of this chapter is to

propose a cyberinfrastructure platform for data management using a top-down design strategy where user needs dictates the design of the database ontology.

This chapter presents a novel data management platform that aims to unify the storage and curation of structural metadata, life-cycle information and sensor data collected from monitored structures. Cyberinfrastructure technologies now allow scalable and secure data management platforms to be developed that facilitate analysis of data in the cloud (*i.e.*, Internet-based computing). First, the chapter describes *SenStore*, a data server architecture that is designed to serve as a core cyberinfrastructure component upon which data-to-decision frameworks for infrastructure health management can be built. The design of the *SenStore* ontology is next described. The ontology provides a rich representation of the monitored structure that enables a broad class of performance and health assessment analyses to be applied to the collected sensor data. The chapter concludes with the performance of *SenStore* evaluated and quantified.

3.2 *SenStore*: A Platform for Data-Driven Structural Asset Management

Figure 3.1 provides an architectural overview of the *SenStore* cyberinfrastructure framework. The proposed framework can be delineated into three major components: 1) sensor network (*i.e.*, structural monitoring system), 2) data management system, and 3) system end users. The sensor network consists of sensors (wired or wireless) installed in a structure for structural monitoring. This monitoring system connects to the Internet via a wireless link such as through cellular communication. The structural monitoring system utilizes this link to push data to a remotely located database server that is designed to securely store sensor data and to expose the data to

authorized clients for interrogation and information extraction. While the database server will serve as the heart of the data-to-decision framework, it may be complemented by both third-party databases (housing relevant information such as metrological data) and data processing services distributed on the Internet. The automated analytics performed using the data stored in the database server (as well as data from the third party databases) is intended to provide actionable information that empowers the decision making process of the system end users.

The database server is the central component of the proposed data-to-decision framework. Thus, its design must explicitly account for the following data management and decision making needs of the end-users. First, the database server must integrate sensor data with other data (*e.g.*, structural metadata including geometric details, materials, sensor types and sensor locations) to provide context for the sensor data. This is especially important for data processing clients which
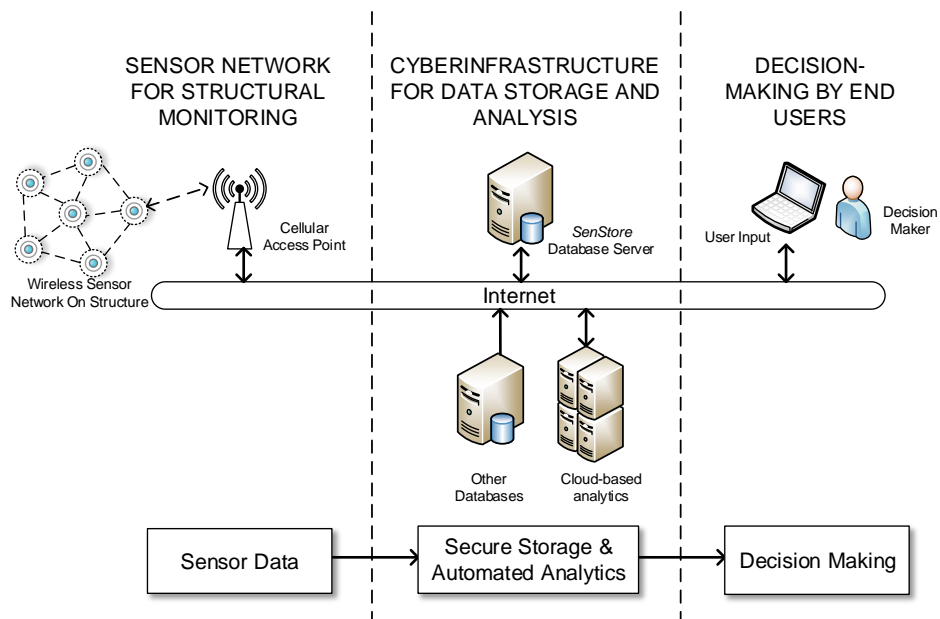
Figure 3.1, Overview of the *SenStore* cyberinfrastructure architecture for structural asset management

require this context to process the sensor data. In this chapter, this set of other data will be generically called *system metadata*. Second, the database server must be compatible with other structural management systems. For example, in the case of bridges, the representation of data in the database must be compatible with AASHTOWare (formerly called PONTIS). Third, the database server must scale well to manage massive amounts of sensor data. Specifically, the query speed of the database must not degrade significantly to ever increasing volumes of sensor data stored in the database server. This is critical for ensuring timely execution of data processing algorithms accessing the data for analysis. Fourth, the database server must balance security of the data with access. Especially after September 11th, 2001, infrastructure owners require data regarding their assets to be stored in a secure manner to ensure individuals with malicious intent cannot exploit data to identify structural vulnerabilities. Finally, the database server must allow authors of processing clients the flexibility to use programing languages of their choice. These desired functionalities are summarized in Table 3.1.

SenStore, jointly developed by SC Solutions (Sunnyvale, CA) and the University of Michigan, is a state-of-the-art data management platform whose design aims to include the functionality summarized in Table 3.1. The implementation details that specifically address each of the five design objectives (*i.e.*, desired functionality) are summarized in Table 3.1 and described in detail in the chapter sub-sections that follow.

Table 3.1, Design requirements and implementation strategies for *SenStore*

Design Objective	Implementation Strategy
Unification of structural metadata and temporal sensor data for data analytics	Combination of a relational database, flat file system and HDF5 as a hybrid database system; the relational database also maintains the relationship between the data stored in the HDF5 and flat file systems
Compatibility with existing structural management systems (*e.g.*, AASHTOWare)	SenStore database ontology that includes industry established structural element classes
Scalability to handle growing volumes of data from dense sensor array	Hybrid database; store data in their most efficient format (storing temporal data in HDF5, multi-media file in flat file system)
Secure access to the sensor data and structural metadata	IP address restriction, web access certification, and the logging of all data access
Support for multiple programming language	Selection of ZeroC Ice middleware

3.2.1 Hybrid Databases System

Data sources relevant to structural asset management can be categorized into five major source types: 1) structural metadata; 2) sensor metadata; 3) functional element data; 4) inspection data; and 5) sensor data. Structural metadata includes structural dimensions, geometric configuration, material properties, etc. Structural metadata could be represented in various forms. However, in the design of *SenStore*, the representation commonly used by finite element method (FEM) modeling tools will be used. For example, structural geometries, boundary conditions and material properties will be defined using nodes and elements defined on a Cartesian coordinate system. Sensor metadata corresponds to information regarding the monitoring system configuration such as sensor type (*e.g.*, accelerometer, strain gage, etc.), sensor location, and sensor properties (*e.g.*, sensitivity, accuracy, range, etc.). Functional element data refers to the representation of the

structure on an element-by-element basis (*i.e.*, beam, column, joint, bearing, wall, etc.). This is especially valuable when *SenStore* is used for bridge management because bridge inspection data stored in AASHTOWare is based on bridge elements. Inspection data obtained from visual inspections can be numerical (*e.g.*, element ratings) or graphical (*e.g.*, photos). Finally, sensor data is numerical and consists of time histories of sensor measurements. The full integration of these five data sources within *SenStore*'s database design will facilitate advanced data analyses (*e.g.*, finite element model updating, reliability and risk analyses, life-cycle analyses, etc.) in an automated fashion. Furthermore, by relating sensor data to the other four data sources, the context of the sensor measurements is established, thereby allowing any conceivable analysis on the data to be implemented and automated in the future.

To store and establish relationships within the data associated with the five data sources, a relational database is a viable option. Many mature options exist for storing relational data including relational databases based on structured query language (SQL). However, the database architecture must ensure that the speed of data input/output is adequate for the quick processing of data by data processing agents. Some of the data types previously defined could present a performance bottleneck for the database in terms of query speed. Specifically, large vectorized sensor time history data and file-like data objects such as inspector reports, inspector pictures, and bridge drawings might slow query speeds. As a result, these data sources are stored in separate repositories better suited for the fast querying of these data types. Therefore, three separate data repositories each optimized for a specific data type are integrated in *SenStore*. A relational database is used to store the majority of the data. PostgreSQL (http://www.postgresql.org/) is chosen for *SenStore* because it is well developed and open source licensed. PostgreSQL is used in *SenStore*

to store data in a relational manner for data from four of the data sources: structural metadata, sensor metadata, functional element data, and some of the inspector data such as element ratings. Three basic data types are used in PostgreSQL including integers (*int*), floating-point numbers (*float*) and strings (*string*); pre-defined enumeration is also used for fields with limited values (*e.g.*, material type, boundary type) in order to minimize database redundancy.

Sensor data is not stored in the PostgreSQL database. Rather, a hierarchical numerical data repository is used because it will be faster in querying numerical and vectorized data (van der Linden et al. 2013). HDF5 (http://www.hdfgroup.org/HDF5/), netCDF (http://www.unidata.ucar.edu/software/netcdf/) and CDF (http://cdf.gsfc.nasa.gov/) are three popular hierarchical data repositories for scientific data. For *SenStore*, HDF5 (Hierarchical Data Format) is chosen to store the large volume of time history sensor data. HDF5 stores this time history data in its natural format (*i.e.*, 2D arrays of time stamp and measurement) thereby improving input/output speed. HDF5 also supports random access to its multi-dimensional data objects using PyTable.

File-like data sources are ill-suited for storage in a relational database or in a hierarchical numerical data repository. Hence, these file-like sources (*e.g.*, inspector reports and images) are stored in their original format in a file repository integrated into *SenStore*. Links are established between PostgreSQL, HDF5 and the file repository to maintain the relationships between the data stored across the three different databases. Specifically, object IDs from HDF5 and the file locations (*i.e.*,

HDF5
REPOSITORY

Sensor Temporal Data:
-Signals/DaqUnitChannelData
Analytical Results Mapping:
-Arrays/StaticLoadToSensorMapping
-Arrays/SensorModalAnalysisResult
-Arrays/SensorProbabilityAnalysisResults

Object
IDs

Object
IDs

POSTGRESQL
DATABASE

Structure General Information:
-Structure
-StructureOwner
Structural Geometries:
-FEMNode
-FEMBeam
-FEMShell
-FEMMaterial
Structural Elements:
-PontisElement
-StructureComponent
Sensing System Information:
-DaqUnit
-DaqUnitChannel
-Sensor
-SensorChannelConnection
-MeasurementCycle
Inspection Results and Utility:
-InspectionAgency
-Inspector
-ComponentInspElement
-StructureComponentPoint
-StructureComponentRating
-FixedCamera
Analytical Results:
-SensorAnalysisGroup
-SensorAnalysisMember
-StaticLoadToSensorMapping
-SensorModalAnalysisResult
-SensorProbabilityAnlaysisResult

File Location

File Location

FILE
REPOSITORY

Structure Drawings
Inspection Reports
Inspection Multimedia

Figure 3.2, An overview of the *SenStore* hybrid database architecture

paths) are stored in corresponding PostgreSQL tables as links. The integrated hybrid *SenStore* database is shown in Figure 3.2.

The ontology of the PostgreSQL relational database is designed to express the relationships between the different data types contained in the database. The data contained in the database is generally dived into 6 groups of tables as shown in Figure 3.3. The first group of tables (10 tables), contains structural inventory information such as structure location, year of construction, etc. The second group of tables (86 tables) pertains to the geometric design of the structure including definition of structure nodes and elements that define the structure in a manner similar to how geometry is represented in a FEM program. The third group of tables (3 tables) define the structure

50

as elements such as beams, columns, piers, decks, etc. The fourth group of tables (8 tables) relates to the sensing system including sensor location, type, specifications, etc. The fifth group (31 tables) corresponds to data associated with visual inspections including dates of inspection, inspection ratings, inspectors' names, etc. Finally, the sixth group of tables (7 tables) corresponds to results derived from analysis of sensor data including modal parameters, load characteristics, etc. In total, the *SenStore* database schema consists of 145 tables. For simplification, only a smaller select number of tables are shown in Figure 3.3 to illustrate the database ontology.

Figure 3.3 depicts how the tables build relationships across the data. These relationships are vital to providing context for data processing clients. Consider a simple steel girder-concrete deck highway bridge. The table `Structure` contains the structure name, links to structural drawings (stored in the file repository), owner name, date built, inventory ID, and location (*e.g.*, longitude and latitude coordinates). The relationship between a structure and its structural elements is maintained by a one-to-many relationship of table `Structure` to table `StructureComponent`. The geometry of the structure is contained in the group of tables associated with the structure geometry.

For example, `FEMNode` is a table containing the X, Y, Z coordinates of nodes defining the bridge in 3-D space. The geometry of the structure adds spatial information to the structural elements by maintaining relationship of `FEMNode` and `StructureComponent`. The element-based structure inspection results (*e.g.*, `StructureComponentRating`) is stored with its associated `StructureComponent`. The sensor system information contains metadata including information of data acquisition (DAQ) channels, sensor types, the connection between each DAQ
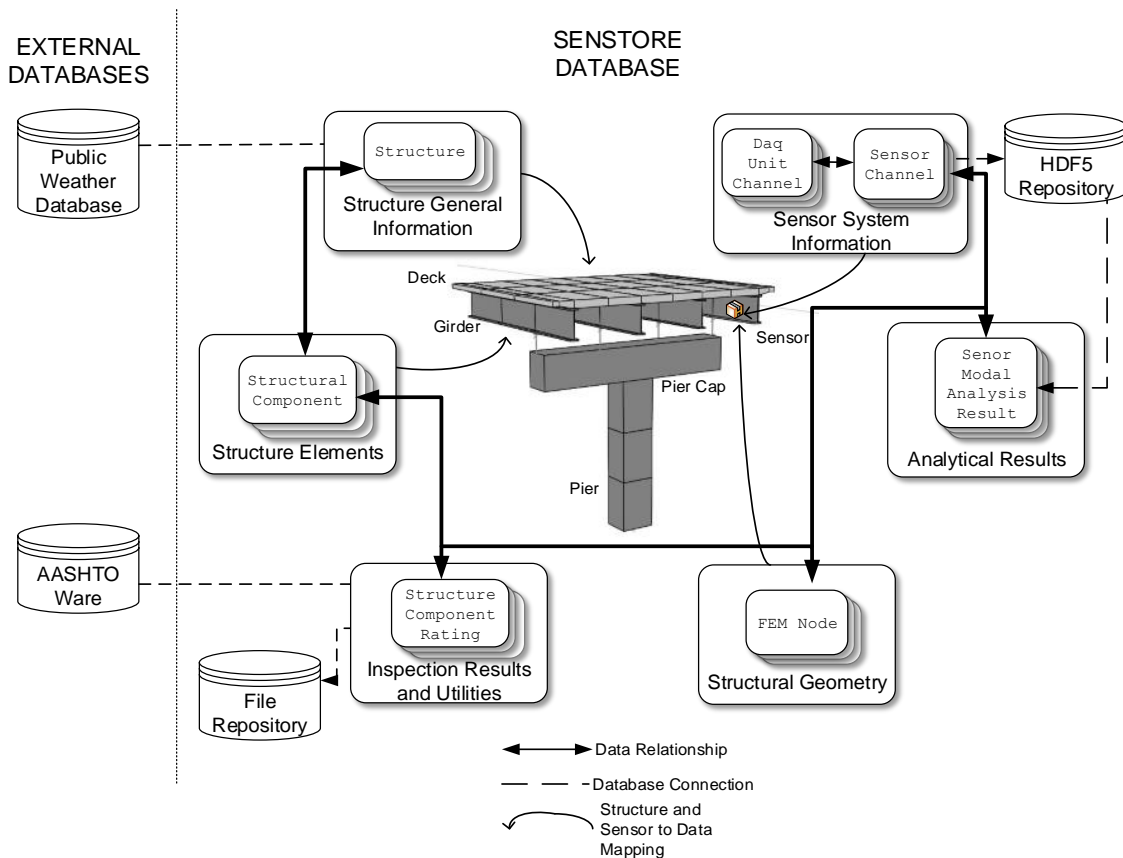
Figure 3.3, Illustration of the relationship in the *SenStore* database ontology

sensor channel and data collection parameters. The sensor system information also maintains the spatial definition of each sensor by associating a `FEMNode` to `DAQUnitChannel`. Analytical results can also be stored in HDF5 with analysis parameters stored in PostgreSQL. For example, if modal analysis is performed, links (*i.e.*, object IDs) to extracted modal parameters stored in HDF5 are stored in the table `SensorModalAnalysisResults`. In addition to the seamless connections among PostgreSQL, HDF5 and the file system, the design of the *SenStore* ontology also offers the compatibility between *SenStore* and external databases. Two illustrative examples are shown in Figure 3.3: 1) additional environment data can be extracted from online weather databases (*e.g.*, Weather Underground); 2) in the case of bridges, structural component rating data in AASHTOWare database can be directly pushed into `StructureComponentRating` table.

3.2.2 Ice Middleware for Remote Access

Middleware is needed that allows remote users (mostly, professional engineers and researchers) easy access to data contained in *SenStore*. At the same time, the middleware must have a sufficient level of security to protect confidential data contained in the database. An additional requirement is an expressive application programming interface (API) that supports multiple programming languages. The Internet Communication Engine (Ice) that is distributed by ZeroC (http://www.zeroc.com/) is chosen for *SenStore* over other existing middleware options (*e.g.*, SOAP, RMI, CORBA) due to its strong technical support, easy implementation, proven robust performance (van der Linden et al. 2013). The three major functions of the Ice middleware layer are: 1) to direct the flow of data internal to *SenStore* among its three integrated data repositories (*i.e.*, PostgreSQL, HDF5 and a file repository system); 2) to provide a defined API for remote users that streamlines the development of data processing clients; 3) to provide a moderate level of security.

Python is chosen as the primary programming language for the *SenStore* implementation because of the Python libraries available which speeds the *SenStore* development time. For example, PyTables in Python is one such supporting library which interface with the HDF5 repository to provide high HDF5 input/output data rates. Numpy is another library in Python which supports advance math operations; it is used in *SenStore* to offer array operations. In addition, intrinsic Python features such as compatibility with the object-oriented programming paradigm, readable syntax and easy integration with the server operating system all enhance the advantage of choosing Python over other programming languages for *SenStore*'s implementation.

After the ZeroC Ice middleware layer is implemented, wrappers are created to perform the read and write functions for each data repository (*e.g.*, PostgreSQL, HDF5, and flat file system) of *SenStore* (Figure 3.4). Table access is supported with a series of Python methods/functions (*e.g.*, add new objects, modify fields of a certain object, equality search with multiple conditions) which provide sufficient and powerful data access options. The data traffic among the three different data repositories is managed by the Ice server layer with database architectural details hidden from the users (van der Linden et al. 2013). A server-client model is also established through the Ice implementation to create a distributed development environment. The majority of *SenStore* users are structural engineers, so the API required must be easy to understand and implement. Furthermore, the supporting language for the API should not be limited to a single language. ZeroC Ice offers both features by standardizing and simplifying the functions/syntax provided by the API to perform complex operations (*e.g.*, data query across different data repository). Several popular programming languages are supported by ZeroC Ice, such as Python, Java and C#, thus offering system and users the freedom to develop in their preferred programming environment.
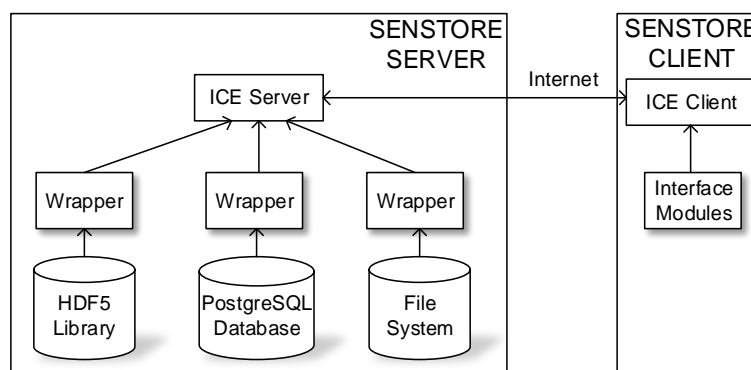
Figure 3.4, *SenStore* implementation with ZeroC Ice

3.3 Performance Validation of *SenStore*

Infrastructure health management systems are established with different purposes, such as to evaluate structure safety after extreme load events and to monitor progressive structural deterioration. These systems face the same set of challenges: rapidly growing data volume (e.g., continuous sensing data, updated inspection inputs), increasing data complexity (e.g., data relationships, external data linkage), and high demand computing (e.g., automated data analysis on a continuous basis). This section showcases how *SenStore* meets the demands of *SenStore* end users such as structural owners. The Telegraph Road Bridge of Chapter 2 will be used as the bestbed. The performance of *SenStore* is validated by quantifying data query speed and illustrating scalability and controlled access.

SenStore has been implemented to manage the metadata and sensing data of the Telegraph Road Bridge since 2011. The metadata of TRB consists of bridge inventory information from the bridge owner, detailed geometry information, inspection data (*e.g.*, element ratings) and sensor system configuration information. This set of metadata stored in PostgreSQL is about 1 MB in size. Multimedia files, including bridge drawings, inspection photos and scanned reports are also stored in the *SenStore* flat file system and are about 7.5 MB in size. The size of the metadata and multimedia files are not very large and more importantly, they generally do not grow with time. The sensing data consists of data and its time stamp from each TRB sensing channel of the wireless monitoring system. Unlike metadata, the size of the sensing data grows continuously with time as shown in Figure 3.5 (a). *SenStore* gracefully handles the growth of sensor data due to the hybrid database system. Figure 3.5 (a) illustrates the cumulative amount of TRB data stored in *SenStore* from 2011 to 2014. As more sensing channels are added into the system, the data size grows faster

(*e.g.*, the slope of data accumulation in late 2012 is steeper than the slope of early 2012). During the winter season, limited solar energy results in less sensing activities and therefore the growth of the data size is slower (*e.g.*, no data in February 2013) and slow accumulation during the winter months of early 2014. As sensing data is stored in HDF5 in *SenStore*, the growth in the amount of sensor data stored will have no negative effect on the performance of *SenStore* queries.

Query speed of a database directly reflects the fitness of the selected hybrid database architecture and the design of the *SenStore* middleware. In order to examine the performance of *SenStore*, two

(a)

(b)
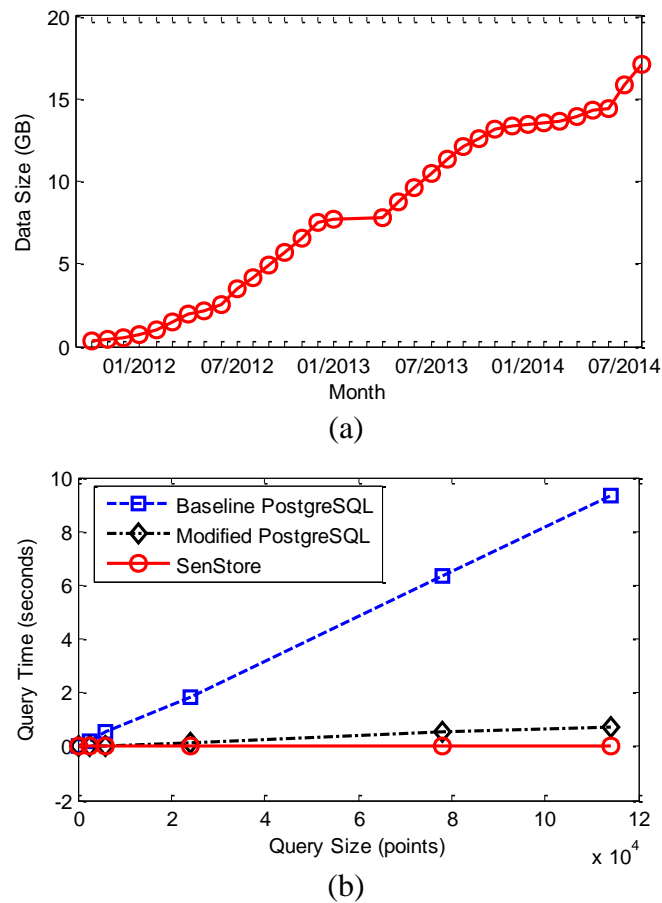
Figure 3.5, *SenStore* performance validation: (a) growth of data stored in *SenStore*, (b) query time comparison between *SenStore* and baseline PostgreSQL database

PostgreSQL databases, termed as the "baseline PostgreSQL" and "modified baseline PostgreSQL",

are built to serve as comparative baselines. Both comparison databases mimic the schema of

SenStore and use a single PostgreSQL database to store both metadata and sensing data (schema

shown in Figure 3.6). For sensing data, the baseline PostgreSQL database uses one big table to

hold the all sensing data (`value`) together with time stamp (`time_stamp`) and data acquisition

channel information (`channel_id`); most traditional relational database intended for storing

monitoring data adopt this design strategy. The modified baseline PostgreSQL database creates

separate tables for each data acquisition channels in order to achieve the best performance of

queries on sensing data. In this modified baseline PostgreSQL database, each table holds sensing

data (`value`) and time stamp (`time_stamp`). *SenStore* and the comparison PostgreSQL

databases (and their data query clients) are implemented on the same server to eliminate unfairness

caused by different machine performance or remote communications. The server has dual core

Intel 64 bits 3.40GHz CPU, 4GB RAM and 500GB SSD, it operates on Ubuntu 12.04 LTS and

```
CREATE TABLE data (
    unit_num integer,
    channel_num integer,
    time_stamp double precision,
    value double precision
);
```

```
...
CREATE TABLE data101 (
    time_stamp double precision,
    value double precision
);

CREATE TABLE data102 (
    time_stamp double precision,
    value double precision
);
...
```

(a)   (b)

Figure 3.6, Baseline and modified baseline PostgreSQL schema: (a) baseline PostgreSQL uses one big table to store sensor data from all channels, (b) modified baseline PostgreSQL creates separate tables for each channel to optimize performance

runs PostgreSQL 9.1. B-tree indexes are created on the `time_stamp` data to improve the performance of range search on time information; this is the most common search method used when processing time history sensor data. A hash index is created on the `channel_id` on the data table of the baseline PostgreSQL database to reduce the data acquisition channel search time to constant. Four years' worth (2011 to 2014) of sensing data is stored in *SenStore*, while for test purposes only 1 year worth (Year 2013) of sensing data is stored in the baseline PostgreSQL database and modified baseline PostgreSQL database. Test queries are designed to examine query speed for big queries on the stored sensing data. The overhead is near 1 second of *SenStore* queries and 0.04 second of PostgreSQL database queries. Figure 3.5 (b) illustrates the time needed to process queries of different sizes after removing the overhead. When the size of the query is small (0 to ~10000 data points), the baseline PostgreSQL databases slightly outperform *SenStore* if considering the overhead introduced by the *SenStore*. When the query size increases, the baseline PostgreSQL database performs poorly as the query time grows linearly. The query time of the modified baseline PostgreSQL and *SenStore* system both grow sub-linearly, but the growth of the *SenStore* query time is slower than the modified baseline PostgreSQL. *SenStore* adopts a binary search algorithm and the PostgreSQL database adopts B-tree indexing to reduce search costs. As the time complexity of both searching algorithm is $O(\log n)$, the input/output speed of PostgreSQL is the bottleneck with HDF5 input/output speed significantly outperforming PostgreSQL. If these trends are extrapolated out to large data sets performing large queries, the time advantages of using *SenStore* are significant.

Data security is one of the basic requirements for any infrastructure health management system. *SenStore* offers multiple tiers of access control to provide a moderate level of security. At the

SenStore client level, users are required to obtain the *SenStore* client software package, the server address of the *SenStore* server and the open port number from the server administrator. Client access is also controlled by IP address with clients categorized by the server into groups with different access permissions (*e.g.*, only administrator can delete data). On the *SenStore* server side, data access through client APIs and the web viewer are logged with the name and IP address of the client with the content accessed and any system errors raised. The server administrator can use the log file for routine server maintenance or to detect malicious attacks.

3.4 Chapter Summary

With technological advancement of data acquisition systems and structural monitoring systems now proceeding rapidly, the lack of scalable data management systems that can effectively aggregate data, curate data and process data in an autonomous manner is evident. As a result, the value of structural monitoring is greatly hindered as structural owners wrestle with how to extract valuable information from data collected that beneficially impacts their decision making. This chapter described *SenStore* which offers an efficient hybrid data management system for structural monitoring applications. *SenStore* provides a cloud-based server for multiple, distributed clients that processes sensor data in an automated fashion. The system scales well with the growing volume of sensing data with current *SenStore* instantiations now hosting 18GB of metadata and sensor data from the wireless structural monitoring system installed permanently on the Telegraph Road Bridge since 2011. Query speed tests of *SenStore* show that *SenStore* outperforms baseline relational database systems based on strictly relational database architectural designs. The query overhead is nearly 1 second using *SenStore*; however, the data search and input/output cost grows much slower with the size of the query compared to the "baseline PostgreSQL" and "modified

baseline PostgreSQL" databases selected for comparison in this study. *SenStore* not only serves as the central data repository for sensor data and bridge metadata, but also provides users with a platform for data analytics such as data mining, data visualization, cloud computing, and data integration.

CHAPTER 4

AUTOMATED MODAL ANALYSIS AND FINITE ELEMENT

MODEL UPDATING USING *SENSTORE*

4.1 Introduction

SenStore has been successfully implemented as part of two long-term structural monitoring systems operating on the New Carquinez Suspension Bridge (Kurata et al. 2013) and the Telegraph Road Bridge (O'Connor et al. 2012) since 2010 and 2011, respectively. In this chapter, a set of automated data analysis tools are presented to demonstrate how a data-to-decision framework can be created on top of *SenStore*. Automated finite element model updating using data from the long-term wireless structural monitoring system installed on the Telegraph Road Bridge is used to showcase and validate the functionality of *SenStore*.

Finite element method (FEM) analysis is a standard tool for advanced structural analysis. Structure owners often maintain FEM models to assist with structural design (Friswell and Mottershead 1995) and to estimate structural capacity during states of deterioration (Michael et al. 1997). More recently, FEM models have been expanded via model updating procedures to assess structural damage using sensor data (Brownjohn et al. 2003; Ni et al. 2009). Regardless of the role of the FEM model, standard practice is to perform model updating in order to ensure the model accurately

reflects the physical system as is represented by the monitoring data collected. Model updating minimizes the difference between the FEM model and the actual structure due to either structural uncertainty (Farhat and Hemez 1993) or due to progressive structural deterioration (Brownjohn and Xia 2000). To showcase the functionality of *SenStore*, an automated data analysis framework is created in this study. The framework seeks to automate the updating of FEM models that are part of an owner decision making process (*e.g.*, capacity estimation, etc.). The example used in this chapter is intentionally generic so that the process of setting up an automated analytic framework based on the *SenStore* architecture could be highlighted. With model updating serving as an integral part of many data-to-decision frameworks (*e.g.*, damage detection, capacity estimation), it is considered a well suited analytical example.

A set of distributed *SenStore* data analysis clients are implemented to concurrently meet four analytical objectives: 1) automate FEM model generation for multiple FEM platforms using *SenStore* metadata; 2) automate modal parameter extraction from vibration-based sensor data; 3) use of machine learning techniques to develop and remove the non-linear regression relationships between modal and environmental parameters; 4) automate the updating of FEM models to develop an accurate analytical representation of the structure. The computing algorithms used to achieve each objective are well established in the literature. These four data analysis clients will be combined in a single work flow (Figure 4.1) to demonstrate the ease and flexibility of developing powerful data-to-decision frameworks based on the *SenStore* platform. The automated finite element model generation client generates a model input file for multiple FEM platforms (*e.g.*, SAP2000, ABAQUS) using the structure geometry information stored in *SenStore*. The client interfaces and runs with the FEM software using the input file generated. The automated modal
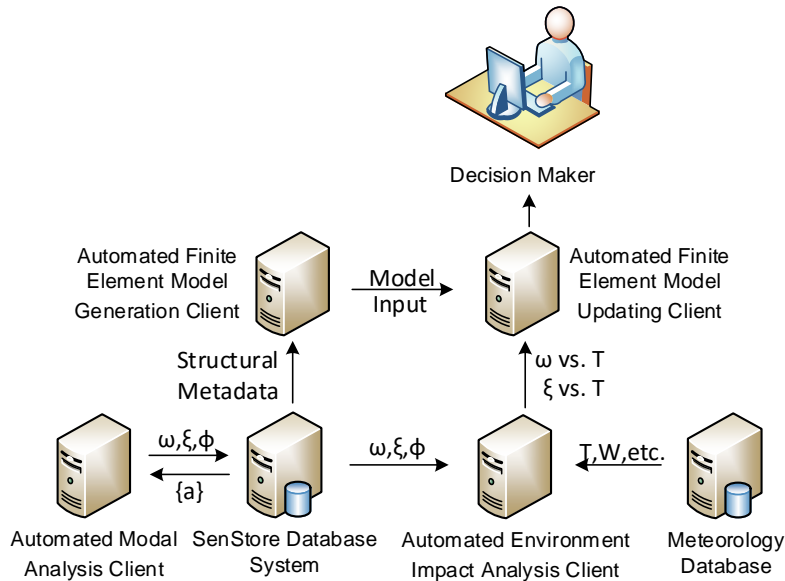
Figure 4.1, Flowchart of the prototype model updating framework based on the *SenStore* platform

analysis client extracts structural modal parameters autonomously from large sets of response data stored in *SenStore* and writes its results (*e.g.*, modal frequency (ω), damping (ξ), and mode shape (φ)) back to *SenStore*. The environmental impact analysis client then connects with a public weather database to query environmental data such as temperature time history data of the structure site, connects with *SenStore* to read modal data, and then uses machine learning techniques (*e.g.*, support vector regression) to model a non-linear relationship between the modal parameters and environmental variables and to exploit the modal properties uncertainties decoupled from environmental dependency. The automated model updating client interacts with the automated finite element model generation client and the environmental impact analysis client to develop an accurate and representative finite element model of the structure that explicitly accounts for parameter uncertainty.

The following chapter subsections will describe in detail each of the clients proposed. The clients will be used to update a FEM model of the Telegraph Road Bridge.

4.2 Automated Finite Element Model Generation Client

The *SenStore* PostgreSQL database ontology is designed to store details on the structural properties in a manner consistent with finite element model standards. *SenStore* stores structural geometric (*e.g.*, nodes and elements) and material (*e.g.*, constitutive law) information in its relational database to ensure relationships between element, materials, and boundary conditions are maintained. The automated finite element model generator automatically extracts structural metadata from *SenStore*, interprets the relationships between information, constructs an input file for a user-defined finite element analysis engine (*e.g.*, SAP2000, ABAQUS, and ADINA) and executes the finite element analysis using the input file. The generator presented here is based on the finite element representation of the Telegraph Road Bridge and uses the finite element software package SAP2000 (Computers and Structures, Inc.).

The generator classifies the finite element model and analysis data into five groups: *FEMmaterial, FEMnode, FEMbeam, FEMshell* and *FEMcontroldata*. Class is constructed under each group and each class is dedicated to one input field of the finite element engine and is designed to map the relevant data stored in *SenStore* to a field in the finite element engine input file. For example, as shown in Figure 4.2, under group *FEMNode*, six classes are constructed. First, *JointCoor* class defines coordinates of each finite element node. Within this defined class are *columns* and *SenTable* which define the involved tables and columns of *SenStore* storing the structural node coordinates. Also *match* maps data from *SenStore* to the specific format of the input file for SAP

```
class FEMNode:
    def __init__(self,mngr):

    class JointCoor:
        def __init__(self):
        # set up table properties
            self.columns = ['Joint','CoordSys','CoordType','XorR','Y','Z','SpecialJt','GlobalX','GlobalY','GlobalZ']
            self.format = ['Text','Text','Text','ft','ft','ft','Yes/No','ft','ft','ft']
            self.defaultvalues = ['','','',0,0,0,'Yes','',0,0,0]
            self.SenTable = ['FEMNode','FEMNode'] # first coorsystem is called GLOBAL
            self.match =[['%LocalID%System > 1','#FEMCoordSystem.mLocalID#id,System','#FEMCoordSystem.mType#id,System','X','Y','Z','NONE','X','Y','Z'],
                        ['%LocalID%System == 1','+GLOBAL','#FEMCoordSystem.mType#id,System','X','Y','Z','NONE','X','Y','Z']]

    class JointAddedMassAss:

    class JointRestAss:

    class ConsDefiEqual:

    class ConsDefiBody:

    class JointConsAss:

    def run(self):

class FEMShell:
    def __init__(self,mngr):

    class ConnArea:
        def __init__(self):
        # set up table properties
            self.columns = ['Area','Joint1','Joint2','Joint3','Joint4']
            self.format = ['Text','Text','Text','Text']
            self.defaultvalues = ['','','','']
            self.SenTable = ['FEMShell']
            self.match = ['*Material*ElementID','#FEMNode.mLocalID#id,N1','#FEMNode.mLocalID#id,N2','#FEMNode.mLocalID#id,N3','#FEMNode.mLocalID#id,N4']

    class AreaSecPropDesPara:

    class AreaSecProp:

    class AreaSecAss:

    def run(self):
```

Figure 4.2, Screenshot of automated finite element analysis driver code

2000. The syntax is supported by a set of utility functions that extracts data from *SenStore* according to the mapping definition and computes the data according to the syntax. The *run* method processes each class and generates the corresponding portion of the input file. The analysis output is also defined. Once the input file is complete, the generator interfaces with SAP2000 using its defined Open Application Programming Interface (OAPI) provided that triggers the analysis.

4.3 Automated Modal Analysis Client

Modal parameters, including modal frequency (ω) and mode shape (φ), are important structural properties. Most model updating algorithms aim to minimize the error between the modal parameters of a finite element method model and those extracted from vibration-based sensor data. The automated modal analysis client takes advantage of the *SenStore* platform in order to automate

the following processes: sensor data extraction, system identification and analysis results management. *SenStore* maintains sensor type and location information of data acquisition nodes that allows clients to distinguish sensors by functionalities. Additionally, *SenStore* also allows clients to create a new analysis, define the associated data acquisition nodes and later store analysis results back to *SenStore* in matrix format. These built-in features of *SenStore* enable the automation of sensor data extraction and the results management processes. The automated modal analysis process is implemented using existing output-only system identification algorithms such as frequency domain decomposition (Brincker et al. 2001) and stochastic subspace identification (SSI) (Peeters and De Roeck 1999). Here, SSI is implemented for demonstration.

A linear dynamic system in continuous-time state space can be modeled as:

$$\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \tag{4.1}$$

$$y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \tag{4.2}$$

where $\boldsymbol{A} \in \mathbb{R}^{2n \times 2n}$ is the system matrix, $\boldsymbol{B} \in \mathbb{R}^{2n \times m}$ is the input matrix, $\boldsymbol{C} \in \mathbb{R}^{l \times n}$ is the output matrix and $\boldsymbol{D} \in \mathbb{R}^{l \times m}$ is the direct transmission matrix. Also l is the number of elements in the system output (y), m is the number of inputs (u) and $2n$ is the number of elements in the state matrix, X. Therefore, n is also a representation of the model order. In a stochastic system, the input, u, is assumed to be white noise defined by a zero-mean Gaussian process.

The SSI algorithm estimates the system matrix, $\boldsymbol{A}$, and the output matrix, $\boldsymbol{C}$ via regression. The detailed algorithm derivation and explanation can be found in Peeters and De Roeck (1999 and 2001a); the method will also be described in Chapter 5. Modal parameters, including the structure modal frequencies, damping ratios and mode shapes, are extracted from the system matrix via a

standard eigenvalue decomposition. To discard modal parameters associated with mathematical modes, a knowledge-based automated modal parameter extraction method (Zhang et al. 2014). The method follows four steps: 1) set up expected ranges of modal frequencies and damping ratios; 2) loop through each defined frequency range; 3) exclude modes outside of the allowable modal frequency ranges; 4) exclude modes outside of the allowable damping ratios ranges; 5) exclude modes based on the modal assurance criteria (MAC) (Allemang 2003) value.

For automated modal analysis of the TRB, the SSI algorithm and extraction algorithm are executed using the 14 vertical accelerations recorded on the main span as shown in Figure 2.8. The SSI analysis is performed on a model order $2n = 28$ and 35; these model orders are established based on a stability plot analysis. Based on the finite element model of the TRB (Mosavi et al. 2014), the range of expected modal parameters are known; based on this *a priori* knowledge higher and lower bounds can be set as shown in Table 4.1 to aid the knowledge-based modal parameter extraction method. The MAC value is established on mode shapes for the two models ($2n = 28$ and 35). For a mode to be a true physical mode, the MAC value is ensuring the mode is relatively unchanged as a function of model order.

Table 4.1, Modal frequencies of TRB and parameter setup for automated modal parameter extraction method

Modes	Expected Modal Frequency (Hz)	Lower Bound (Hz)	Higher Bound (Hz)	Damping Ratio Limit	MAC Value Limit
1	2.4	2.35	2.7	0.03	0.95
2	2.8	2.7	3.0	0.1	0.95
3	4.8	4.6	5.2	0.1	0.95
4	8.0	7.5	8.5	0.1	0.95
5	9.0	8.6	9.4	0.1	0.95
6	9.5	9.4	10.1	0.1	0.9

Figure 4.3 shows a histogram of the first four modal frequencies and damping ratios. The Automated Modal Analysis client autonomously extracted from TRB data collected from September 2012 to February 2014. The distributions of modal frequency are hard to identify using common distributions (*e.g.*, normal, log normal) due to environmental influence (as will be discussed shortly) but the distributions are clustered around the expected frequencies. For the first four modes, the counts are negligible near the pre-set boundaries (Table 4.1) verifying that the knowledge-based automated modal extraction method successfully identifies and separates these modes. The distribution of mode 1 appears to consist of two individual distributions with one centered near 2.49 Hz and the other near 2.56 Hz. A similar phenomenon is observed in the distribution of mode 3. As will be later discussed, the lack of a single Gaussian distribution will be due to the environmental influence on the modal frequencies. Modal damping ratios for the first four modes have a log normal distribution with modal damping of 1% or lower.
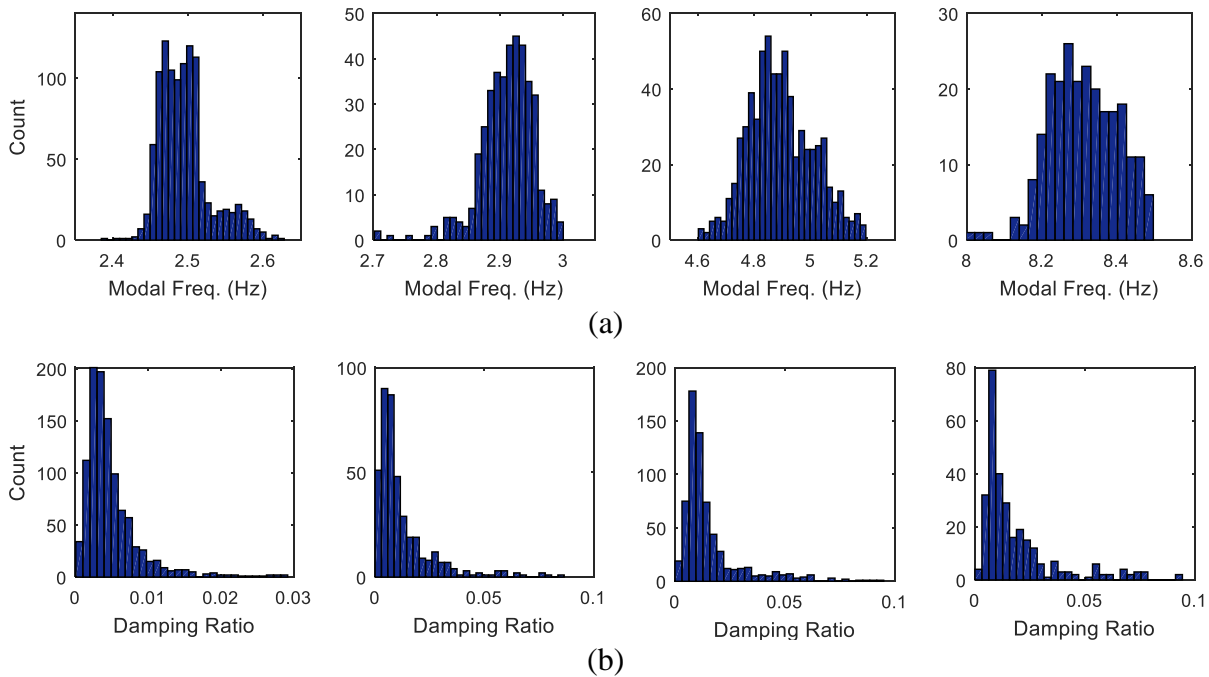
Figure 4.3, Distributions of modal parameters for first four modes: (a) modal frequency distributions, (b) modal damping ratio distributions

During the sensor data extraction process, the automated modal analysis client repeatedly queries sensor data from *SenStore* therefore resulting in unneglectable cost in time. Unlike the query example demonstrated in Chapter 3, the query size of each channel in this client depends on the sampling frequency and duration and is usually fixed, but data from different channels are queried which requires access to different tables of the database. To process long-term monitoring data for modal analysis, this type of query is repeated over thousands of times. For the case of the TRB testbed, accelerometers sense at 200 Hz for 1 min, resulting 12000 points from each channel per dataset, and 15 accelerometers (15 channels) are installed on TRB. From September 2012 to February 2014, about 1000 quality datasets were collected for modal analysis. Therefore, the query extracts 12000 points from 15 different tables repeatedly 1000 times. To validate *SenStore* performance, the query time of *SenStore* is again compared with the baseline PostgreSQL and modified baseline PostgreSQL databases using the same hardware. The query time (including overhead) is 1.01s, 7.50s and 1.49s for *SenStore*, baseline PostgreSQL and modified baseline PostgreSQL, respectively. Figure 4.4 projects the time cost when large amounts of data are processed; hence, the benefit of using *SenStore* is reaped when large amounts of data are queried. Users can further optimize *SenStore* performance by selecting a batch query size in order to reduce overhead and avoid disk input/output at the same time.
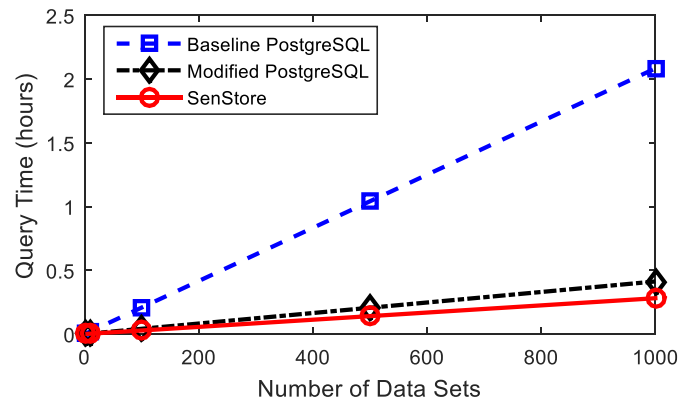
Figure 4.4, Query time comparison using an identical dataset (15 sensor channels collected at 200Hz for 1 min)

4.4 Automated Environmental Impact Analysis Client

Environmental and operational conditions (EOC) (*e.g.*, temperature, wind speed, wind direction, precipitation, and traffic) are known to impact modal properties (Brownjohn et al. 2010; Cross et al. 2012; Hua et al. 2007; Moaveni and Behmanesh 2012; Moser and Moaveni 2011; Peeters and De Roeck 2001a; Zhang and Lynch 2013). The wireless structural monitoring system on TRB is equipped with thermistors to measure structure surface temperature. Other environmental parameters of interest are also available from online weather databases while traffic information is obtained from the Michigan Department of Transportation (MDOT) weigh-in-motion station (WIMS) databases. Taking advantage of the open development platform provided by *SenStore*, the environmental impact analysis client is able to seamlessly query data from the sensing network and public weather database to develop nonlinear regression relationships between modal parameters and EOCs.

The automated environmental impact analysis client interacts with the automated modal analysis client to extract the identified modal parameters from different data sets, extracts temperature information from the thermistor data stored in *SenStore*, and interfaces with an online weather database (www.wunderground.com) and the MDOT WIMS database to obtain additional EOC data during the same time period as the TRB sensor data set. Figure 4.5 (a) illustrates the bi-linear dependency between the 1st modal frequency (identified by automated modal analysis client) and structural temperature with a distinguishable pivot near the freezing point (0° C). The data shown in Figure 4.5 is consistent with the data used in Section 4.2. To characterize the relationship between modal frequency and environmental factors, a non-linear regression algorithm termed support vector regression (SVR) (Chang and Lin 2011; Smola and Schölkopf 2004) is
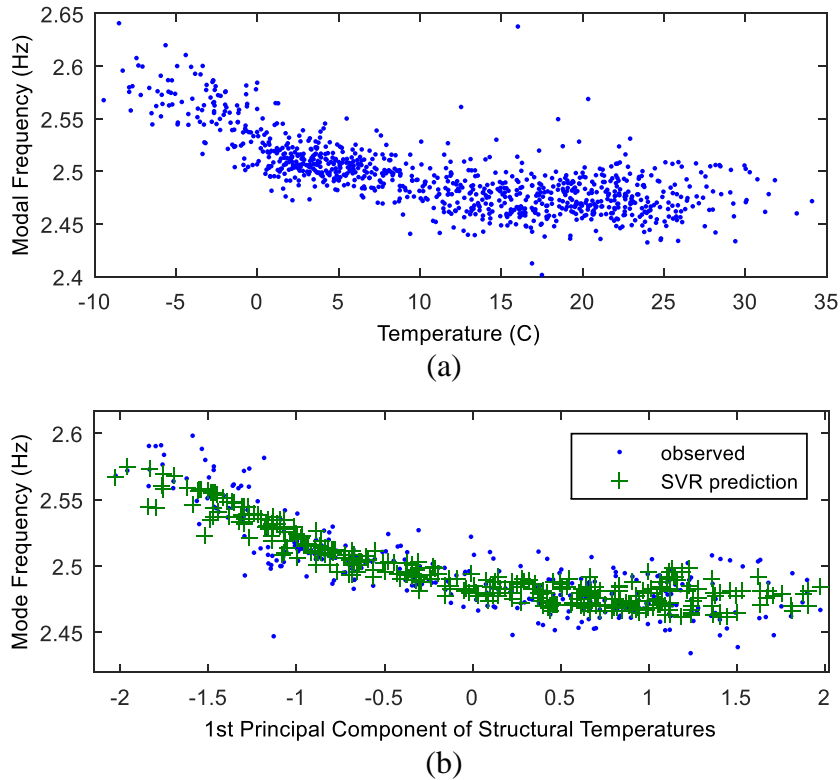
Figure 4.5, Impact of temperature on the 1st mode frequency: (a) 1st mode frequency versus structural temperature, (b) prediction of the 1st mode frequency using support vector regression

implemented. The SVR algorithm projects low dimensional data into high dimension space via a kernel function (*e.g.*, Gaussian kernel) and searches for a hyperplane that minimizes the absolute model error. To discover the relationship of TRB modal frequencies and EOC, seven environmental parameters are chosen as features including: the first principal component (Jolliffe 2005) of structural temperatures, wind speed, wind direction, air pressure, dew point, visibility and estimated traffic load from data collected by the nearest WIMS. The data is divided into training and testing groups for support vector regression analysis. The first principle component of structural temperatures, wind speed and direction are identified as the features having the strongest correlation with modal frequency. The SVR prediction is shown in Figure 4.5 (b), that the prediction and original testing data have an excellent match both numerically and geometrically.

The mean square error of the prediction from SVR and the testing data (normalized to zero mean and unit variance) is as low as 0.3, thereby validating SVR as an effective tool for removing the impact of EOC on the 1st modal frequency.

The error, ε, between observed modal frequencies and predicted frequencies from SVR reveals that the environmental dependency is effectively removed with only a small error process remaining. The sum of the mean, μ, of modal frequencies and error, ε, is considered as the *corrected* model frequencies which are independent of environmental impact.

$$f_{corrected} = \varepsilon + \mu \tag{4.3}$$

$$\varepsilon = f_{observed} - f_{SVR_{prediction}} \tag{4.4}$$

$$\mu = mean(f_{observed}) \tag{4.5}$$

The *corrected* modal frequencies are further used in the automated model updating client.

4.5 Automated Model Updating Client

Model updating calibrates the existing finite element model so that the model has similar modal properties as those estimated from the vibration-based sensor data. A wide collection of model updating algorithms have been proposed over the past 50 years (Friswell and Mottershead 1995; Levin and Lieven 1998; Mottershead and Friswell 1993; Perera and Ruiz 2008). Due to the rehabitation of the TRB deck surface in 2011, the stiffness, thickness and material density of the deck have been changed and require updating in the finite element model. Therefore, in this work, the model updating scheme focuses on estimating the distribution of the deck stiffness (*i.e.*, compressive strength, f_c') for given *corrected* modal parameters inherited from the aforementioned environmental impact analysis client. To demonstrate the capability to distribute and automated

the finite element analysis using *SenStore*, Bayesian model updating using adaptive Markov Chain Monte Carlo (MCMC) simulations (Andrieu and Thoms 2008; Beck and Katafygiotis 1998; Behmanesh and Moaveni 2015) is used to update the finite element model parameters for given observed bridge modal parameters.

Under the Bayesian model updating framework, the probability distribution function (PDF) of the finite element model parameters, $\boldsymbol{\theta}$, given the *corrected* model parameters, $\mathbf{d}$, and the probabilistic model class, M, is defined as:

$$P(\boldsymbol{\theta}|\mathbf{d}, \mathrm{M}) = cP(\mathbf{d}|\boldsymbol{\theta}, \mathrm{M})P(\boldsymbol{\theta}|\mathrm{M}) \tag{4.6}$$

$P(\boldsymbol{\theta}|\mathrm{M})$ is the *prior* PDF, $P(\boldsymbol{\theta}|\mathbf{d}, \mathrm{M})$ is the *posterior* PDF, $P(\mathbf{d}|\boldsymbol{\theta}, \mathrm{M})$ is the *likelihood* function and c is the normalization constant. In this study, *corrected* modal frequencies are used as the evidence and only one probabilistic model class is considered (hence, the variable, M, is dropped). Therefore,

$$P(d|\theta) = \prod_{m=1}^{N_m} P(\tilde{f}_m |f_{\boldsymbol{m}}(\theta)) \tag{4.7}$$

where, N_m is the total DOF considered in the model updating, $\tilde{f}_{\boldsymbol{m}}$ is the mth *corrected* modal frequency and $f_{\boldsymbol{m}}(\theta)$ is the mth eigen-frequency of the finite element model given model parameter, θ. The error between the *corrected* model parameters and the eigen-frequencies of the finite element model is modeled as a random variable of a zero-mean Gaussian distribution:

$$e_m = \tilde{f}_{\boldsymbol{m}} - f_{\boldsymbol{m}}(\theta) \tag{4.8}$$

The *likelihood* function can therefore be modeled as:

$$P(\tilde{f}_m |f_{\boldsymbol{m}}(\theta)) \propto \exp(-\frac{1}{2}\frac{(\tilde{f}_m - f_{\boldsymbol{m}}(\theta))^2}{\sigma_{\tilde{f}_m}^2}) \tag{4.9}$$

$$P(\boldsymbol{d}|\theta) = P(d_1, d_2 \ldots d_{N_k}|\theta) \propto \hat{c} \prod_{k=1}^{N_k} \exp(-\frac{1}{2}\sum_{m=1}^{N_m}\frac{(\tilde{f}_{k,m} - f_{\boldsymbol{m}}(\theta))^2}{\sigma_{\tilde{f}_{k,m}}^2}) \qquad (4.10)$$

and, where N_k is the number of evidence. $\sigma_{\tilde{f}m}$ is formulated as $COV_{\tilde{f}m}\tilde{f}_m$ and $COV_{\tilde{f}m}$ is the average

coefficient of variance of $\tilde{f}_m$.

In this study, the first 4 *corrected* modal frequencies from 66 data sets are used as the evidence, **d**,

and compressive strength of the concrete, f'_c, is used as the hypothesis, **θ**. The algorithm 4: AM

algorithm with global adaptive scaling as presented in Andrieu and Thoms (2008) is used to

updated the finite element model parameters due to its efficiency in model parameter convergence

(Behmanesh and Moaveni 2015 and Andrieu and Thoms 2008). The sample acceptance rate is set

at 40% and to ensure the updating parameters are meaningful in the physical domain, boundary

conditions are applied when drawing samples as shown in Table 4.2; samples selected from outside

of the boundaries are discarded. The histogram of the total 21278 accepted samples and the

posteriori distribution are shown in Figure 4.6 (a) and (b), respectively. Taking the maximum *a*

posteriori estimation of the results, the compressive strength of the new concrete deck is estimated
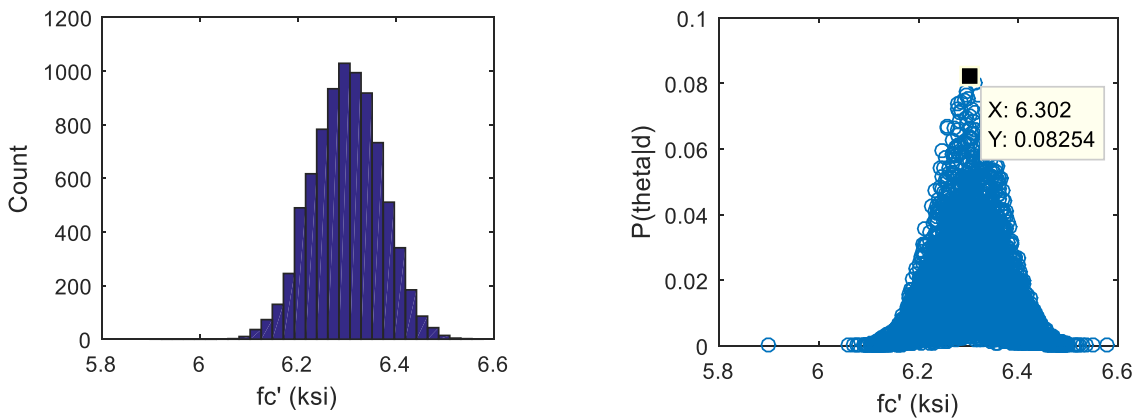
Figure 4.6, Bayesian model updating results: a) histogram of accepted samples, (b) posterior distribution

Table 4.2, Modal parameters and physical parameters of original and updated finite element model

	Original	Updated	Observed Mean	Error (%)	Lower Bound	Upper Bound
1st Modal Freq. (Hz)	2.46	2.498	2.488	0.40	N/A	N/A
2nd Modal Freq. (Hz)	2.86	2.906	2.934	-0.95	N/A	N/A
3rd Modal Freq. (Hz)	4.75	4.822	4.890	-1.39	N/A	N/A
4th Model Freq. (Hz)	8.30	8.429	8.295	1.62	N/A	N/A
Concrete Compressive Strength (MPa)	34.5	43.4	N/A	N/A	13.8	62.1

at 43.4 MPa (6.3 ksi). The updated modal frequencies and the model parameters results are listed in Table 4.2.

4.6 Chapter Summary

The case study presented in this chapter demonstrated a data-to-decision framework performing autonomous modal analysis and finite element model updating using structural metadata and sensor data from the Telegraph Road Bridge. This test case consisted of four data clients executing in an autonomous manner: automated finite element model generation client, automated modal analysis client, automated environmental impact analysis client and automated finite element model updating clients. The ease of integrating these distributed clients into a cohesive data processing framework was illustrated in this study. Future work will focus on advancing the *SenStore* cyberinfrastructure platform including utilization of a distributed database system to provide a cloud-based computing platform with flexible APIs and user-friendly data visualization tools. Another focus for future work is the creation of more meaningful data-to-decision frameworks that can be upon the *SenStore* foundation that are more specific to the health management of structures.

CHAPTER 5

AUTOMATED LONG-TERM MODAL ANALYSIS OF SUSPENSION BRIDGES UNDER VARYING ENVIRONMENTAL CONDITIONS USING WIRELESS STRUCTURAL MONITORING SYSTEMS

5.1 Introduction

Modal characteristics (*e.g*., modal frequency, modal damping ratio and mode shape) are considered one of the most important properties of a civil structure because they control the free and forced (*e.g.,* ambient vibrations, earthquakes, and wind load) vibration response of a structure. When designing seismically excited structures using design spectra, modal frequencies and damping ratios are carefully considered so that the maximum response of the structure in the modal domain can be determined (ASCE 2013). Alternatively, if a time history analysis approach is taken to analyze the maximum response of the structure, the accuracy of the approach is dependent on the accuracy of the structural model. Modal parameters are often extracted from structural monitoring data using system identification techniques so that accurate structural models can be obtained through model updating methods (Friswell and Mottershead 1995).

Due to the difficulties inherent to measuring the load imposed on a large operational civil engineering structure such as a bridge, output-only system identification methods are often used

to estimate global modal parameters. System identification methods in both the frequency (*e.g.*, frequency domain decomposition (Brincker et al. 2001), peak picking) and time (*e.g.,* auto-regression model (Ljung 1998), stochastic subspace identification (Peeters and De Roeck 1999)) domains are widely used to obtain the modal frequencies, damping ratios and mode shapes of a structure. Theoretically, modal frequencies and damping ratios are inherent properties of a structure and are assumed to be invariant. However, several studies (including Chapter 4 of this dissertation) have shown that operational modal parameters obtained using vibration-based structural monitoring systems are not fixed and can vary within 5% of their range from their mean value. In the case of bridges, environmental parameters such as temperature, wind profile and traffic are all known to influence modal frequency (Cross et al. 2012; Hua et al. 2007; Moser and Moaveni 2011; Peeters and De Roeck 2001a; Zhang and Lynch 2013). By observing modal parameters over long periods of time, the variations in modal properties can be modeled over large EOC profiles. Better models of modal parameter estimation can improve the understanding of large structures leading to improvement of future design methods.

Research presented in this chapter aims to: 1) develop data-driven techniques to expand the stochastic subspace identification algorithm to efficiently extract structural physical modes for long-term vibration monitoring data, 2) quantitatively identify environmental and operational impact on large-scale suspension bridge structures using linear and non-linear models, and 3) provide statistical models for modal frequency, damping and mode shapes for the bridge monitored. In this study, modal analysis as well as an operational and environmental impact study are performed using vibration-based monitoring data collected over 2 years by the long-term wireless structural monitoring system installed on the New Carquinez Bridge, a large-scale suspension

bridge in Vallejo, California. A novel data-driven method for modal parameter estimation is presented which utilizes statistics from long-term data to improve detection robustness for operational modes that are low in energy or close to one another. Modal analysis is performed in an autonomous manner on over 4,000 datasets using SSI optimized for automated mode selection and computational speedup. In total, 27 of the bridge are identified under 2 Hz. The statistics of the identified modal frequencies and damping ratios of each mode are tabulated. Environmental and operational impact on the bridge acceleration, strain and displacement response are also investigated. Specifically, the relationships between modal frequencies, operational and environmental parameters, including time of the day, girder acceleration, structural temperature and wind speed, are modeled using Ridge regression and Gaussian Process regression for performance comparison.

5.2 Bridge Response to Environmental and Operational Conditions

The New Carquinez Bridge is subject to a variety of dynamic excitation sources including traffic, wind, temperature and earthquakes. Bridge responses to traffic, wind and temperature loads are evident from the response data collected. A typical response of the bridge to traffic load on Thursday, June 27, 2013 at 3:00 pm is shown in Figure 5.1. The top plots illustrate the vertical acceleration response of the bridge measured at sensor N1, N5, N10, S5 and S1 (see Figure 2.2). The vertical acceleration response, which is attributed to traffic loads is typically less than 50 mg and in this figure well below 30 mg. While individual vehicle responses cannot be easily deciphered from the time history acceleration plots, there is an ordered delayed acceleration response with higher energy content as vehicle groups travel from north (N1) to south (S1). To further characterize the traffic impact on the bridge accelerations, spectrum analysis is performed
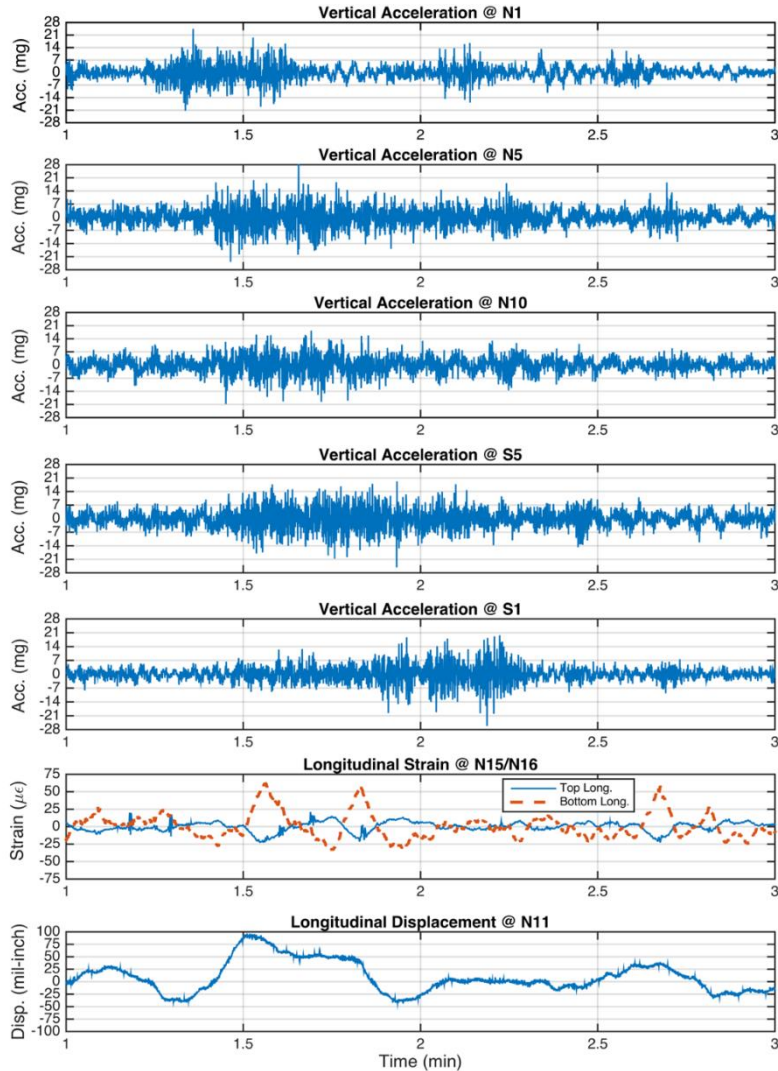
Figure 5.1, Typical time history response of the New Carquinez Bridge for select sensing channels (Recorded: June 27, 2013, 3:00 PM)

(Figure 5.2) using vertical acceleration responses collected from sensors located at N10 (middle point of the main span) and N5 (quarter points of the main span) for the same time period as shown in Figure 5.1, albeit for a longer time history. Figure 5.2 includes sample time history plots of vertical acceleration (top), corresponding spectrum amplitudes (middle) and time frequency spectrogram (bottom) plots for each sensing locations. The spectrum amplitude plots based on Fast Fourier Transform (FFT) of the entire time history record (9600 data points) reveal the modal

frequencies of the bridge. Several dominant modes are observed at both locations including 0.20, 0.26, 0.48, 0.79, 1.03, and 1.16 Hz.

The spectrum plot at N10 generally provides larger modal peaks than the N5 spectrum plot but some modes have larger peaks at N5 such as modes near 0.35 and 0.65 Hz. The spectrogram plots of Figure 5.2 are based on the Fast Fourier Transform (FFT) of 10 seconds of data (200 data points) with 50% overlap resulting in a resolution of 0.1 Hz in the frequency-domain. They reveal that the first few modes of the bridge dominate the frequency spectrum and are always present in the data. However, higher frequency content in the bridge is evident at times when the acceleration amplitudes are higher; this is attributed to traffic on the bridge.

(a)   (b)
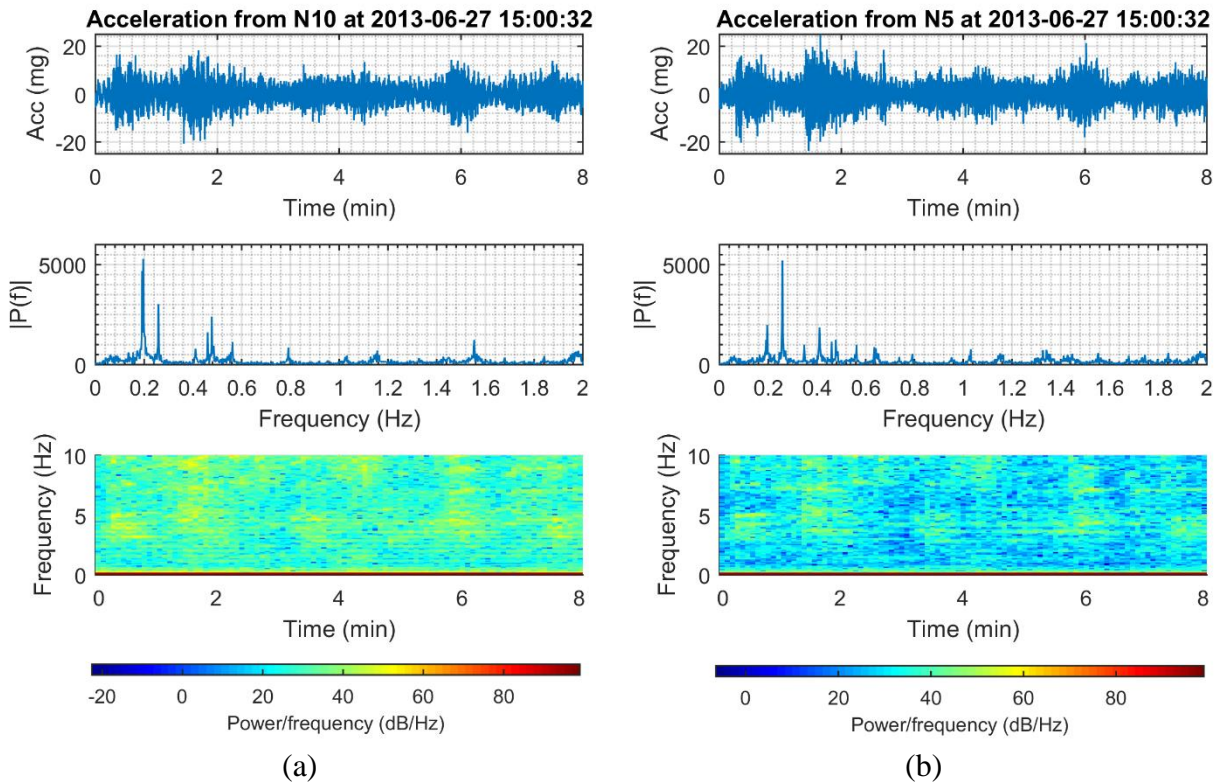
Figure 5.2, Acceleration time history plot, spectrum amplitude and spectrogram (top to bottom) of the vertical acceleration responses from sensors located at: (a) N10 and (b) N5 (Recorded: June 27, 2013, 3:00 PM)
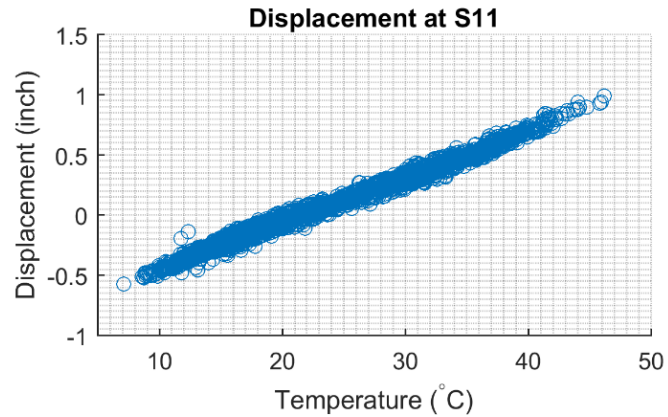
Figure 5.3, Structural temperature (thermistor at N16) impact on bridge girder longitudinal displacement (string potentiometer at S11) measured 2013-2014

Time synchronized plots of the longitudinal strain measured at the center of the main span (*i.e.,* by sensor N15 and N16) is shown. The bottom strain ranges from 30 (compressive) to 65 (tensile) microstrain while the top strain ranges from 30 (compressive) to 20 (tensile) microstrain. The strain response is based on global deflections of the girder as would be expected for the orthotropic steel box girder. The induced flexural moment on the span ensures that when the bottom is in tension, the top is in compression (and vice versa). Furthermore, the bottom strain is larger in magnitude due in part to the composite action of the road surface that is on the top of the steel girder. Finally, the top strain gage experiences local excitation (roughly at 1.2, 1.3, 1.6, 1.7 and 1.85 min in Figure 5.1) which is hypothesized to be due to traffic wheel loads immediately over the gage location resulting in a localized strain response in the wall of the box girder. The girder also displaces longitudinally relative to the tower as measured by the linear potentiometer at N11 as shown in Figure 5.1. This displacement is relatively small (less than 0.25 cm) and is likely due to the slow variations in the weight of traffic on the bridge superimposed on the global modal response of the girder excited by traffic and wind. However, more dramatic displacements are observed over longer periods of monitoring. For instance, Figure 5.3 plots the longitudinal displacement of the

girder relative to the tower collected at S11 over two years of monitoring (from January 1st, 2013 to December 31st, 2014). The displacement is perfectly linear as a function of the top girder structural temperature (as measured at N16) revealing the thermal expansion of the deck (estimated to be 0.095 cm/°C).

5.3 Modal Analysis using Stochastic Subspace Identification

As previously presented, a primary study objective is to assess the relationship between modal properties (*i.e.,* modal frequencies and damping) of long-span bridges exposed to common EOCs including thermal, wind and traffic loads. Since the initiation of the monitoring system deployment on the New Carquinez Bridge in 2010, a total of 35 GB of sensor data has been collected from the wireless monitoring system. The extensive amount of data collected opens the possibility of assessing bridge properties as a function of EOCs but manual data processing is not a feasible option. In this study, an automated approach to modal analysis is created to extract the modal properties of the New Carquinez Bridge. Modal analysis of operational structures often adopts output-only system identification algorithms because the forcing input of the structure is difficult to measure and quantify. In this study, stochastic subspace identification (SSI) is adopted to extract the modal parameters of the New Carquinez Bridge from the long-term data acquired. Specifically, the deck and tower accelerations acquired from January 1st, 2013 to December 31st, 2014 are selected for analysis.

5.3.1 Stochastic Subspace Identification

The SSI algorithm used in this study is based on the works of Van Overschee and De Moor (1996). A brief description will be provided but interested readers are referred to Van Overschee and De

Moor (1996) for a complete theoretical treatment of the SSI algorithm. The dynamics of a structural system can be modeled as a linear, time invariant input-output state space model (Equation 5.1) in continuous time space:

$$\dot{x}(t) = A_c x(t) + B_c u(t)$$
$$y(t) = Cx(t) + Du(t)$$
(5.1)

where $A_c \in \mathbb{R}^{n \times n}$ is the system matrix, $B_c \in \mathbb{R}^{m \times n}$ is the input matrix, $C \in \mathbb{R}^{l \times n}$ is the output matrix, and $D \in \mathbb{R}^{l \times m}$ is the direct transmission matrix; $x(t) \in \mathbb{R}^{n \times 1}$ is the state vector at time t and $y(t) \in \mathbb{R}^{l \times 1}$ is the output vector at time t. Dimensions n, m and l represent the state-space model order (also the dimension of the state vector), the number of inputs and the number of outputs/sensor locations, respectively. However, the structural system is observed using a digital, discrete time data acquisition system; hence, the system output, y, and input, u, vectors can only be obtained at discrete times defined by the sampling rate of the data acquisition system ($f_s = 1/\Delta t$). The state space model of Equation 5.1 must be transformed into discrete time state space:

$$x_{k+1} = Ax_k + Bu_k$$
$$y_k = Cx_k + Du_k$$
(5.2)

where $A = \exp(A_c \Delta t)$, $B = [A - I]A_c^{-1}B_c$ and $x_k = x(k\Delta t)$ are the state matrix, input matrix and state in discrete time, respectively.

As previously mentioned, the system input, u, is hard to measure and is replaced in the model with a broadband stochastic disturbances, $\omega_k \in \mathbb{R}^{n \times 1}$ and $v_k \in \mathbb{R}^{l \times 1}$ defined in discrete time:

$$x_{k+1} = Ax_k + w_k$$
$$y_k = Cx_k + v_k$$
(5.3)

Under normal ambient excitation, these noise vectors are assumed to be zero mean white noise processes with covariance matrices as defined:

$$E\left[\binom{w_p}{v_p}\begin{pmatrix} w_q^T & v_q^T \end{pmatrix}\right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}\delta_{pq} \tag{5.4}$$

where δ_{pq} is the Kronecker delta function. SSI is a time-domain system identification method that will estimate an optimal system (A) and output (C) matrices corresponding to the linear time invariant model of Equation 5.3.

The SSI method begins with the construction of the Hankel matrix of the measured system output

$$H \overset{\text{def}}{=} \frac{1}{\sqrt{j}}\begin{pmatrix} y_0 & y_1 & \cdots & y_{j-1} \\ \cdots & \cdots & \cdots & \cdots \\ y_{i-2} & y_{i-1} & \cdots & y_{i+j-3} \\ y_{i-1} & y_i & \cdots & y_{i+j-2} \\ \hline y_i & y_{i+1} & \cdots & y_{j-1} \\ y_{i+1} & y_{i+2} & \cdots & y_{i+j} \\ \cdots & \cdots & \cdots & \cdots \\ y_{2i-1} & y_{2i} & \cdots & y_{2i+j-2} \end{pmatrix} \overset{\text{def}}{=} \begin{pmatrix} Y_{0|i-1} \\ Y_{i|2i-1} \end{pmatrix} \overset{\text{def}}{=} \begin{pmatrix} Y_p \\ Y_f \end{pmatrix} \in \mathbb{R}^{2li \times j} \tag{5.5}$$

where y_i is the output vector at $t = i$, Y_p denotes the upper half of the Hankel matrix and is referred to the "past" portion of the data matrix, and Y_f denotes the lower half of the Hankel matrix and is referred to the "future" portion of the data matrix. Based on the theory of SSI presented by Van Overschee and De Moor (2012), the geometric projection of the row space of the "future" output on the row space of the "past" output can be calculated directly and is equal to the product of the extended observability matrix, O, and Kalman state sequence, $\hat{X}_i$:

$$P_i = Y_f/Y_p = Y_f Y_p^T \left(Y_p Y_p^T\right)^\dagger Y_p = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{pmatrix}\left(\hat{x}_i \ \hat{x}_{i+1} \ \dots \ \hat{x}_{i+j-1}\right) = O_i \hat{X}_i \in \mathbb{R}^{li \times j} \tag{5.6}$$

Hence, by performing the projection the resulting projection matrix, P_i, can be factored by singular value decomposition (SVD) to expose the extended observability matrix, O, and Kalman state sequence, $\hat{X}_i$. The projection matrix is rewritten as:

$$P_i = U_1 S_1 V_1{}^T \tag{5.7}$$

where $S_1 \in \mathbb{R}^{n \times n}$ is the singular value matrix, $U_1 \in \mathbb{R}^{li \times n}$ is the left singular vector matrix, $V_1 \in \mathbb{R}^{j \times n}$ is the right singular vector matrix, and n is the effective rank of P_i. The extended observability matrix can be calculated as:

$$O_i = U_1 S_1{}^{1/2} \in \mathbb{R}^{li \times n} \tag{5.8}$$

Once the extended observability matrix is found, the Kalman state sequence, $\hat{X}_i$ is determined:

$$\hat{X}_i = O_i^{\dagger} P_i \in \mathbb{R}^{n \times j} \tag{5.9}$$

This Kalman state sequence is the key to the estimation of the state space model matrices A and C. Specifically, a second state sequence is required which is derived from repartitioning the Hankel matrix as follows:

$$H \overset{\text{def}}{=} \frac{1}{\sqrt{j}} \begin{pmatrix} y_0 & y_1 & \cdots & y_{j-1} \\ \cdots & \cdots & \cdots & \cdots \\ y_{i-2} & y_{i-1} & \cdots & y_{i+j-3} \\ y_{i-1} & y_i & \cdots & y_{i+j-2} \\ y_i & y_{i+1} & \cdots & y_{j-1} \\ y_{i+1} & y_{i+2} & \cdots & y_{i+j} \\ \cdots & \cdots & \cdots & \cdots \\ y_{2i-1} & y_{2i} & \cdots & y_{2i+j-2} \end{pmatrix} \overset{\text{def}}{=} \left(\frac{Y_{0|i}}{Y_{i+1|2i-1}} \right) \overset{\text{def}}{=} \left(\frac{Y_p{}^+}{Y_f{}^-} \right) \tag{5.10}$$

The projection of the repartitioned "future" output row space on the "past" output row space is:

$$P_{i-1} = Y_f{}^- / Y_p{}^+ = O_{i-1}\hat{X}_{i+1} \in \mathbb{R}^{l(i-1) \times j} \tag{5.11}$$

The extended observability, O_{i-1}, is derived from an SVD of P_{i-1} as described in Equation 5.8. The Kalman state sequence can be found as:

$$\hat{X}_{i+1} = O_{i-1}^{\dagger} P_{i-1} \in \mathbb{R}^{n \times j} \tag{5.12}$$

Equation 5.3 can be modified in a more condense form using the Kalman state estimate matrix, $\hat{X}_i$, and a reduced data matrix of the system output, $Y_{i|i} \in \mathbb{R}^{l \times j}$:

$$\begin{pmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} = \begin{pmatrix} A \\ C \end{pmatrix} (\hat{X}_i) + \begin{pmatrix} \rho_w \\ \rho_v \end{pmatrix} \tag{5.13}$$

The unknown system matrix and output matrix can be determined using the least square method:

$$\begin{pmatrix} A \\ C \end{pmatrix} = \begin{pmatrix} \hat{X}_{i+1} \\ Y_{i|i} \end{pmatrix} \hat{X}_i^{\dagger} \tag{5.14}$$

The modal properties of the monitored system can be extracted from the discrete time linear state space model. An eigenvalue decomposition of the state matrix, A:

$$A = \Psi \Lambda \Psi^{-1} \tag{5.15}$$

yields the diagonal eigenvalue matrix $\Lambda \in \mathbb{C}^{n \times n}$ and the eigenvector matrix $\Psi \in \mathbb{C}^{n \times n}$. Diagonal terms of the eigenvalue matrix correspond to pole locations on the discrete time complex plane. For lightly damped systems, such as bridges, the eigenvalues will be complex valued and will be within the unit circle. To extract the frequency and damping ratio corresponding to each mode, the conjugate pair eigenvalues, λ_q and λ_q^*, are mapped to the continuous time complex plane:

$$\lambda_{cq} = \frac{\ln(\lambda_q)}{\Delta t} \tag{5.16}$$

where λ_{cq} is the eigenvalue of the qth mode in the continuous time complex plane and Δt is the time step. Modal frequency, ω_q, and damping ratio, ξ_q, for the qth mode is then obtained:

$$\omega_q = |\lambda_{cq}| \tag{5.17}$$

$$\xi_q = \frac{Re(\lambda_{cq})}{\omega_q} \tag{5.18}$$

The operational mode shapes can be obtained as:

$$\Phi = C\Psi \tag{5.19}$$

"Operational" mode shapes is a more accurate term due to possible violations in the broadband white noise input assumptions stated at the outset of the SSI derivation.

5.3.2 Automated Implementation of Stochastic Subspace Identification

The critical component of the SSI algorithm is determination of the order (n) of the state-space model. This is often done using stabilization diagram analyses. Stabilization diagrams plot modal frequency or modal damping ratio (abscissas) for a range of model orders (ordinate) varied sequentially (Ewins 2000). Physical modes tend to stabilize (*i.e.,* modal properties don't vary) as the model order is increased. Hence, the stabilization diagram serves two purposes: it helps identify physical modes from non-physical (*i.e.*, mathematical) modes and it provides a tool for picking an appropriate model order, n. However, the SSI algorithm has two other user-defined tuning knobs to consider: the dimensions of the Hankel matrix (namely, i and j). Typically, the end user of the system identification algorithm would manually adjust these parameters to ensure the SSI method yields physically reasonable modal parameters. However, when seeking to automate the SSI algorithm, the system end-user is no longer in the loop. In addition, when the SSI method is applied to large datasets with results aggregated in a comparative manner, a consistent approach to the application to SSI must be used to ensure modal results from one data set can be more fairly compared to modal results from another. In this section, the method for establishing an appropriate Hankel matrix size that can be used in an automated implementation of the SSI algorithm will be described. In addition, the method for the determination of the model order, n, is also presented.

A Hankel matrix is constructed with synchronized acceleration response data corresponding to the accelerometer channels ($l = 46$) installed along the bridge girder and towers. Each data set is collected at 20 Hz for 8 minutes resulting in $N_s = 9600$ points per channel per data set. The size of the Hankel block is constrained by the amount of data by the relationship:

$$2i + j - 1 \leq N_s \tag{5.20}$$

For a given i, it is convenient to set the Hankel matrix width $j = N_s - 2i + 1$ in order to maximize the number of data points used for estimation. Maximizing the amount of data used in the Hankel matrix is known to lead to improved estimates for the modal damping ratio (Meng et al. 2011; Pridham and Wilson 2003). The Hankel matrix width, j, will bias the determination of the model order, n, during SVD of the projection matrix (Peeters et al. 1997). To ensure the bias is consistent, this study will fix j while varying the number of block rows, $2i$, in the Hankel matrix. Therefore, to compare the estimated modal parameters using different values of $2i$, the width of the Hankel block is set as:

$$j = N_s - \max(2i) + 1 \tag{5.21}$$

where $\max(2i)$ corresponds to the largest $2i$ considered through a stabilization diagram analysis where i will be varied.

SVD of the projection matrices P_i and P_{i-1} are critical steps in the data-driven SSI approach which establish the model order, n. Given the size of $P_i \in \mathbb{R}^{li \times j}$ where $li < j$, a total of li singular values will be obtained during SVD of P_i. Within the set of li singular values are n physical (true) modes of the system where $n < li$. For some structures, it can be difficult to determine an appropriate model order. Over the years, many rule-based methods have been developed to distinguish physical modes from mathematical modes (Häckell and Rolfes 2013; Zhang et al. 2014). In this

study, the method for choosing n is to examine the difference between two neighboring singular

values, $\Delta\lambda$, as a means of establishing the effective rank of P (van Overschee and de Moor 2012).

When analyzing large amounts of data sets in an automated fashion as done herein, rules are

established to automate the determination of n. All singular values obeying the following two

conditions (*i.e.,* rules) are deemed as corresponding to physical modes counting towards the model

order:

$$\lambda \geq k_1 \times max(\lambda) \tag{5.22}$$

$$\lambda \geq \lambda_p \tag{5.23}$$

where λ_p is the last singular value that is true for $\Delta\lambda_j \geq k_2 \times ma\,x(\Delta\lambda)$, $\Delta\lambda_j = \Delta\lambda_j - \Delta\lambda_{j-1}$.

Here, k_1 and k_2 are determined empirically to ensure most of the modes extracted are physical

modes with mathematical modes eliminated. The first condition (Equation 5.22) ensures the most

dominant singular values count towards the model order. The second constraint (Equation 5.23)

ensures singular values are contributing significantly towards the model with the smaller singular

values in the tail excluded. The number of singular values satisfying the two constraints

correspond to the model order, n. In this study, manual evaluation of a large number of

acceleration response data sets from the New Carquinez Bridge is conducted to the determine an

appropriate value of k_1 and k_2; they are set at $k_1 = 0.05$ and $k_2 = 0.0025\sqrt[2]{2i}$.

Computational optimization of the data driven SSI method is important to ensure scalability when

implemented on the extensive data sets acquired. SVD of the projection matrix, $P_i \in \mathbb{R}^{li \times j}$,

taking $\mathcal{O}(l^2 i^2 j)$ floating point operations is the most expensive step in the automated

implementation of SSI (Trefethen and Bau 1997). Since only the left singular vectors and singular

values are required for determination of the state space model, a normalization trick can be applied
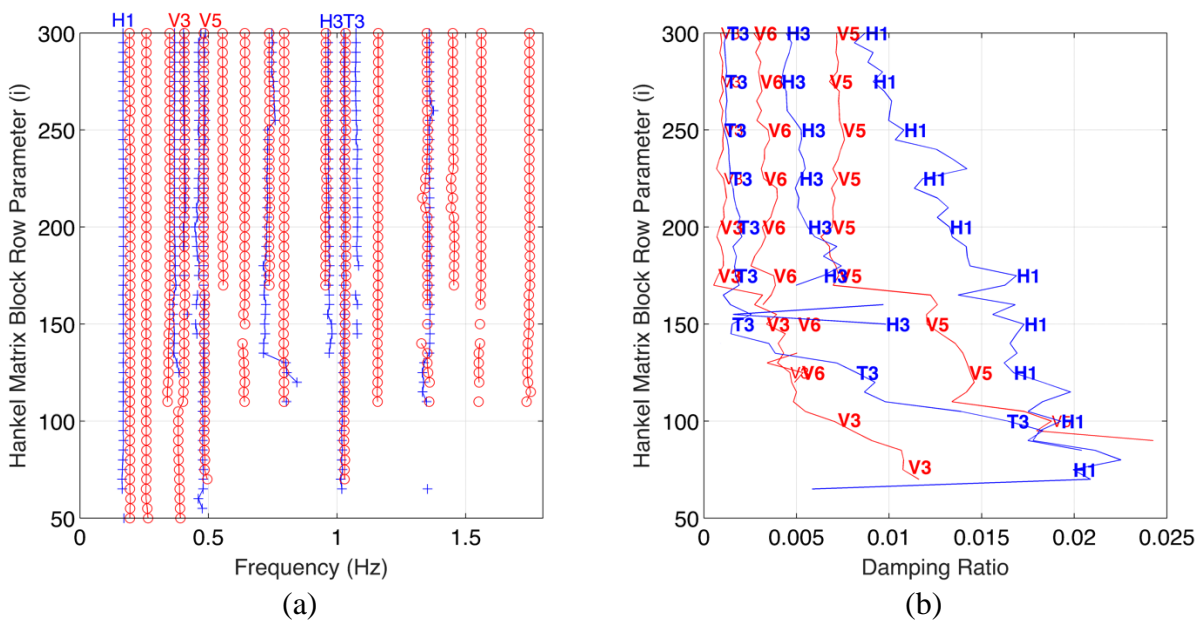
Figure 5.4, Stabilization diagram as a function of the Hankel matrix block row size i: (a) modal frequencies separated as vertical and transverse (horizontal); (b) modal damping ratios for select modes

to reduce the computation time of Equation 5.7. Instead of performing SVD of the projection, $P_i \in$

$\mathbb{R}^{li \times j}$, the SVD is performed on $P_i P_i^T \in \mathbb{R}^{li \times li}$ which for nonsingular P_i reduces to:

$$P_i P_i^T = U_1 \, S_1^2 {U_1}^T \tag{5.24}$$

SVD on $P_i P_i^T$ reduces the computation costs to $\mathcal{O}(l^3 i^3)$ floating point operations.

For illustration purposes, consider one instance of the New Carquinez Bridge acceleration data set

collected at 6 PM on June 26th, 2013. First, the accelerations are digitally filtered using a 4-pole

Butterworth low-pass filter with a cutoff frequency at 2.5 Hz. Second, the size of the Hankel

matrix must be determined. The choice of i has a direct effect on the modal results due to under-

and overfitting issues of the SSI model. The choice of i is determined by examining the

stabilization diagrams of modal frequencies and damping ratios over a range of i. Figure 5.4 is

the stabilization diagram which shows modal frequencies stabilize when $i > 110$; similarly, modal

90
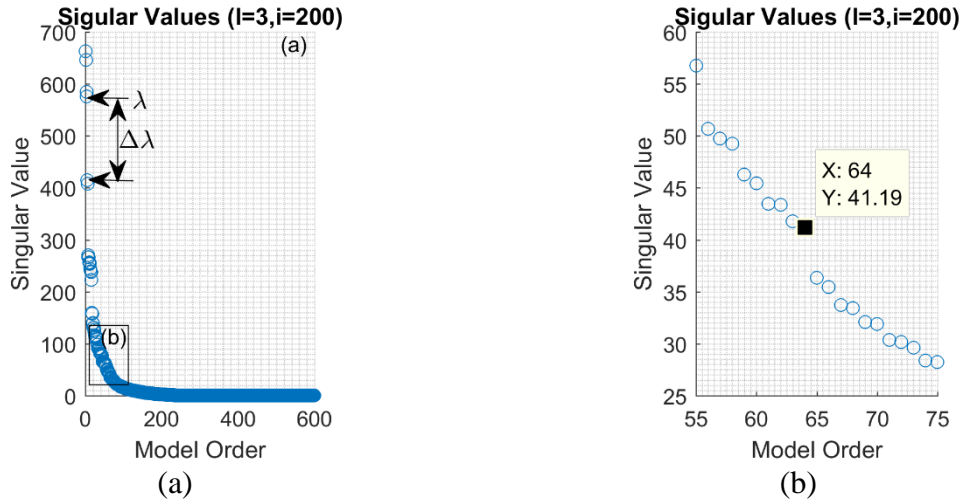
Figure 5.5, Singular values of the projection when $l = 3$ and $i = 200$: (a) all singular values; (b) zoom on the first 30 singular values

damping ratios stabilizing when $i > 200$. In order to visualize the damping ratio of individual modes more clearly, only a subset of the modes is shown in Figure 5.4b. Therefore $i \in [200,250]$ is determined to be the range to consider for the Hankel matrix block row size in this study; with $max(2i) = 500$, the width of the Hankel matrix is fixed at $j = 9101$ based on Equation 5.21. Based on the stabilization diagrams, the modal parameters are extracted including modal frequency and damping ratio. The results corresponding to the transverse (horizontal) modes appear slightly less stable when viewing both modal frequencies and damping ratios. Furthermore, there appears to be some overlap between the transverse and vertical modal frequencies (*i.e.*, near 0.41, 0.48, 0.74, 0.97 and 1.03 Hz) in this data set. This will be investigated in greater depth in the later phases of this study.

A key component of assembling the stabilization diagrams is the automated determination of the model order, n, based on the rules established in Equation 5.22 and 5.23. To highlight this facet of the implementation, using three outputs (*i.e.,* vertical accelerations at N1, N5 and N10 in Figure 2.2) where $i = 200$ and $j = 9101$ are used. The singular values from this set are plotted in Figure
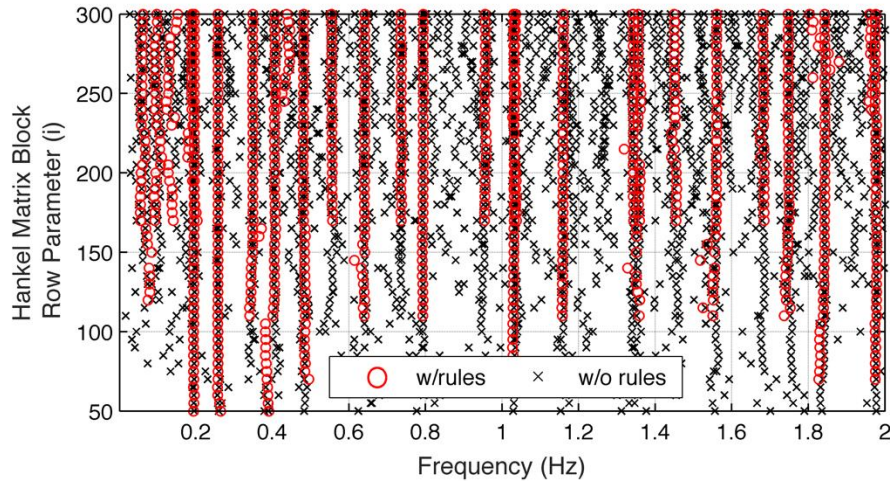
Figure 5.6, Stabilization diagram of modal frequency with and without rules on singular values

5.5 (a); applying the rules to the singular values of this data set results in $n = 64$ autonomously determined as shown in Figure 5.5 (b). The 10 modes extracted between 0.15 and 1 Hz are at 0.196, 0.258, 0.349, 0.406, 0.481, 0.555, 0.641, 0.737, 0.794, 0.956 Hz, which match well with the spectrum amplitude peaks in Figure 5.2; this underscores the effectiveness of the rules on the singular values in effectively discarding mathematical modes. Figure 5.6 shows a comparison of modal frequencies obtained with and without rules on the singular values. All of the consistent modes extracted without rules were also extracted with rules proving that the rules retain the physical modes while removing mathematical modes. As previously mentioned, SVD is performed on $P_i P_i^T$ in lieu of P_i. When $l = 3$, $i = 200$ and $j = 9101$ the time to compute the SVD on $P_i P_i^T$ using MATLAB R2015b on a standard personal computer (i5-2400 CPU at 3.10 GHz, 6 GB RAM) is approximately 0.2 seconds while the corresponding time to perform SVD on P_i is approximately 4.7 seconds; this is a speedup of 23 times.

Considering all acceleration channels ($l = 46$) in this data set, a total of 15 operational mode shapes are extracted under 2 Hz, including 5 transverse (horizontal), 7 vertical and 3 torsional modes. The
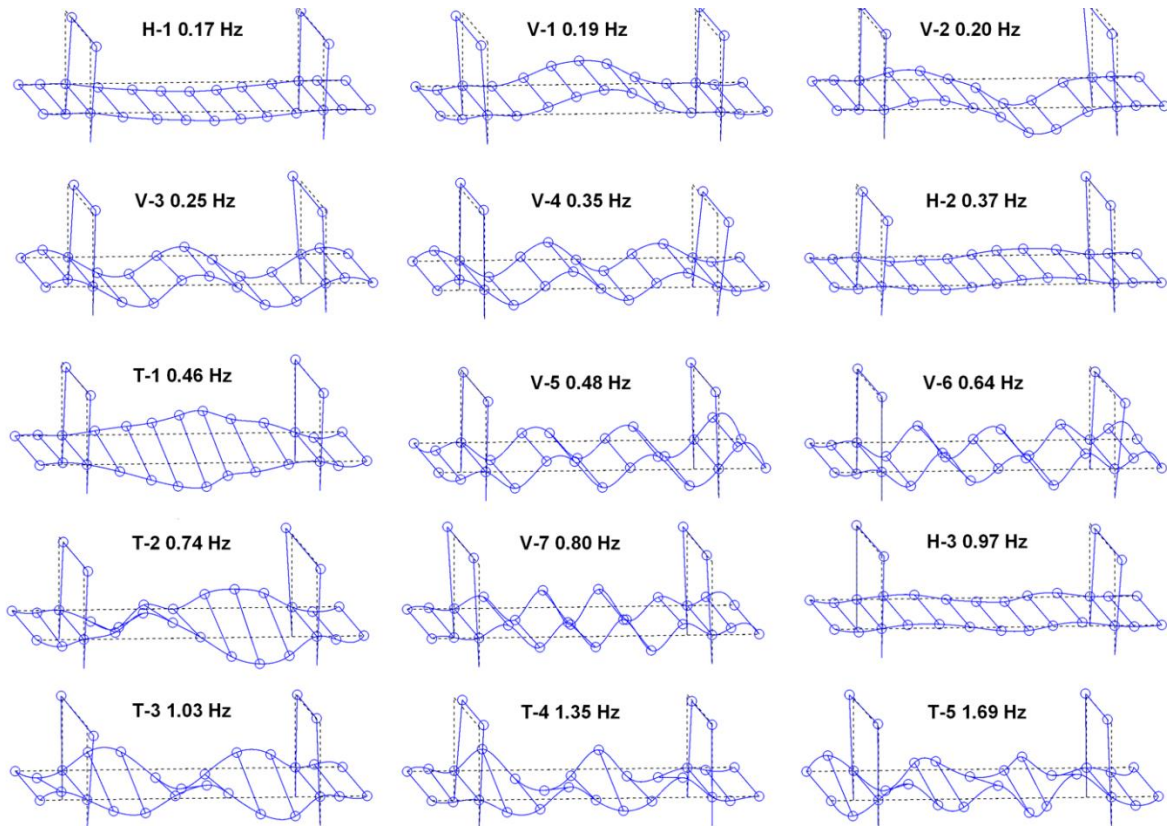
Figure 5.7, Mode shapes of the bridge extracted using SSI analysis

operational mode shapes are shown in Figure 5.7. A cubic spline is used to interpolate the shape between nodes. The torsional modes (T-1, T-2, T-3, T-4 and T-5) show displacement in both vertical and transverse directions which explains the frequency overlapping near 0.74 and 1.03 Hz in Figure 5.7. The primary vertical mode, V-1 at 0.19 Hz, is closely located with mode V-2 at 0.20 Hz; therefore, it is difficult to extract mode V-2 for some data sets as has been pointed out by other studies (Kurata et al. 2013). Modes near 0.41 and 0.56 Hz have similar mode shapes as modes V-4 and V-5, respectively, with the exception that the displacements on the end spans are much larger. It is also worth noting that the number of plotted mode shapes is limited by the number of accelerometers installed along the main span of the girder, which is 7 on both sides of the bridge main span, meaning more vertical modes may exist under 2 Hz. The operational mode shapes

determined are in close agreement with those previously reported for the bridge (Conte et al. 2008; He et al. 2009).

5.3.3 Identification of Modal Frequencies

Acceleration time history data is first filtered using a 4-pole Butterworth low-pass digital filter with a cutoff frequency at 2.5 Hz to remove high frequency signal components outside the spectrum of interest (< 2 Hz). The SSI algorithm is conducted separately on the data in the vertical and transverse directions in order to avoid model coupling. A histogram of all frequency solutions obtained using five variations in the Hankel matrix row parameter $i = [210, 220, 230, 240, 250]$ over all of the datasets is shown in Figure 5.8. This histogram is intended to determine the primary modal frequencies of the New Carquinez Bridge based on histogram peaks. Each labeled peak is considered a mode (structural or operational) due to being consistently identified over the set of 4426 datasets and variations in the Hankel matrix row parameter, i. Based on analysis of operational mode shapes at these modal frequencies, modes are labeled as vertical (V), horizontal (H), or torsional (T). Operational modes considered the true structural modes based on comparison to previous reported modal analyses (Conte et al. 2008; He et al. 2009) are labeled according to mode order (*e.g.,* V-1 for the first vertical mode, V-2 for the second vertical mode, *etc*.). The other modes are considered as operational deflection shapes influenced by the bridge loading; these modes are then labeled according to mode frequency (*e.g.,* V-0.14 at 0.14 Hz).
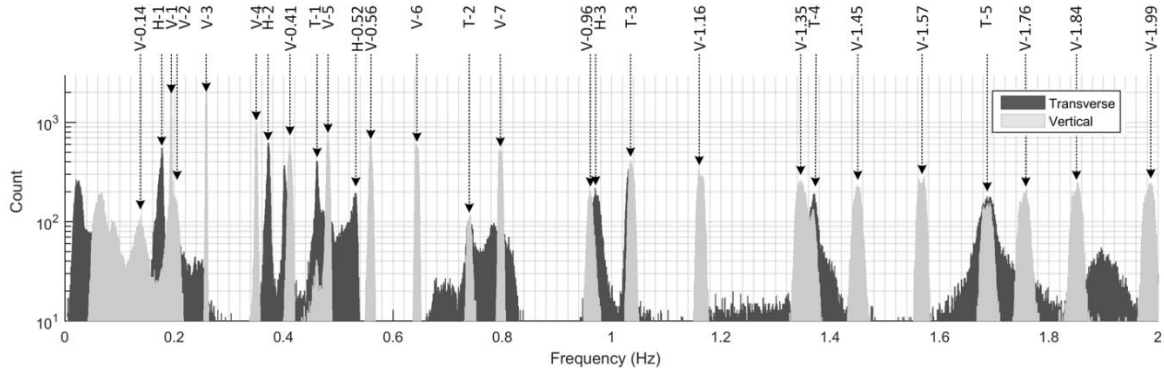
Figure 5.8, Histogram of operational modal frequencies extracted from SSI solutions using 4426 data sets from N1, N5 and N10 and five SSI models of varying Hankel matrix size

Using a histogram to aggregate all modal frequencies extracted from a large set of bridge response data under a consistent set of SSI algorithm parameters (*e.g.,* five Hankel matrix row parameters) is a novel data-driven approach that more accurately determines the operational modal frequencies of the bridge. First, aggregating modal frequencies over such a large volume of data sets makes structural modes easier to distinguish from mathematical modes. Specifically, structural modes are more consistent over all the data sets resulting in high counts in the frequency bins; in contrast, mathematical modes will vary more resulting in lower counts in the frequency bins. Second, the method provides a dramatic improvement in the ability to identify closely spaced modes over traditional implementations using a limited set of bridge response data. For example, the mode V-2 peak (at 0.20 Hz) is "buried" in the mode V-1 peak (at 0.19 Hz) in the spectrum amplitude plots of Figure 5.2. In the data-driven method, the two modes are easy to distinguish with the modal frequencies showing as overlapped distributions in Figure 5.8.

According to Figure 5.8, 27 peaks under 2 Hz are selected as potential modal frequencies. Based on an analysis of the mode shapes (Figure 5.7), a total of 15 true structural modes are finally identified from the 27 peaks as denoted in Figure 5.8. The 15 peaks identified in this study are in

agreement with the modes previously identified by He *et al.* (2009) in an experimental modal

analysis of the bridge prior to its opening in 2003.   The other peaks are deemed to correspond to

operational deflection shapes that have dependency on the loads. For example, the peaks below

H-1 (less than 0.19 Hz) in Figure 5.8 are hypothesized to be induced by traffic and wind loads.

Mode V-1 and V-2 are closely located in the frequency-domain; the peaks overlap but the V-1

peak is more dominant. Torsional modes (T-1, T-2, T-3, T-4 and T-5) are identified in both the

vertical and transverse signals. He et al. (2009) identified structural modes at 0.96 Hz, 1.16 Hz,

1.35 Hz, and 1.57 Hz (denoted in Figure 5.8 as V-0.96, V-1.16, V-1.35, V-1.57). However, the

insufficient number of accelerometers along the main span installed in the wireless monitoring

system did not allow the mode shapes to be accurately identified. In a similar fashion, the three

peaks above 1.7 Hz (V-1.76, V-1.85, and V-1.99) in Figure 5.8 are likely structural modes, but

they cannot be confirmed as well. For example, He et al. (2009) was also not able to identify these

modes due to the limited number of sensors those authors used in their study.

5.3.4 Statistical Analysis of Modal Frequency and Modal Damping Ratio

The 27 peaks below 2 Hz identified in Figure 5.8 are analyzed further. For each dataset, all

frequencies obtained are discretized into 27 frequency bins (ranges). For each bin, the median

frequency and damping ratio from the five SSI models (corresponding to $i =$

$[210,220,230,240,250]$) are used as the modal frequency and damping ratio for the dataset.

Histograms of the median modal frequency and median damping ratio of mode H-1, V-1, V-2, V-

3, V-4, H-2, T-1 and V-5 are plotted in Figure 5.9 (a) and (b), respectively. The histograms of

modal frequency are similar to Figure 5.8 with symmetric distributions seemingly consistent with

a normal distribution. However, when fitting a normal distribution to modal frequencies it is

discovered that high kurtosis in the data renders a poor fit. After exploring alternative distributions, the logistic distribution is found to fit the modal frequency data best. The modal damping ratios have asymmetric distributions, heavily skewed to the left; it is reasonable to categorize such as log logistic. The probability density functions of logistic and log logistic distribution are plotted over the histograms of the modal frequency and damping ratio, respectively, to illustrate the fitness of distribution. The damping ratio for modes in the vertical directions are low with most falling well below 0.8%. The transverse and torsional modes show slightly higher damping ratios most of which are under 1.5%. The much higher damping ratios of mode V-2 are likely a result of high

(a)

(b)
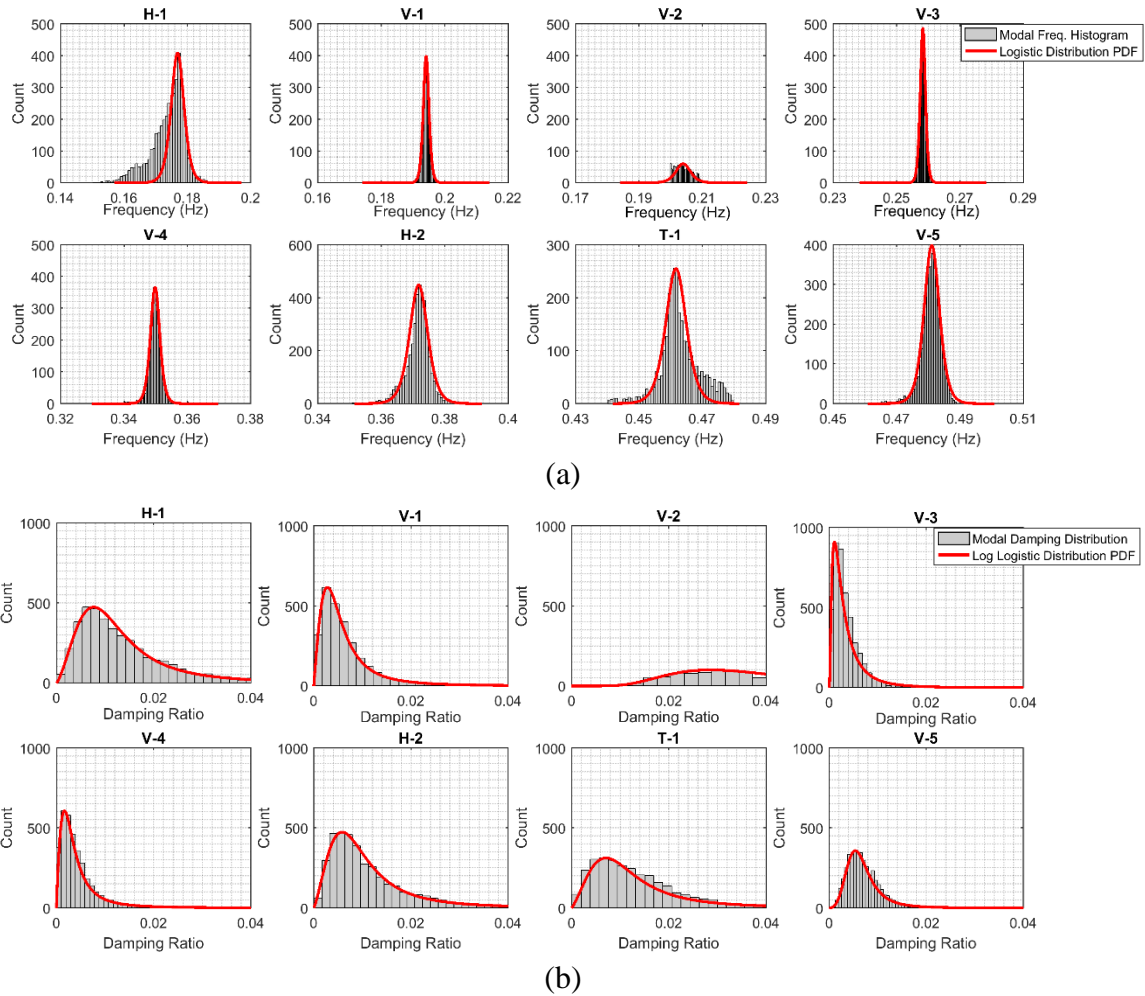
Figure 5.9, Histograms and logistic statistical distributions of the modal properties of the New Carquinez Bridge: (a) modal frequency; (b) modal damping ratio

uncertainty in the damping estimation due to the low participation factor of this mode resulting in a small number of data sets for which V-2 was identifiable. The relatively low damping ratios are similar to what the literature has reported for structures experiencing low level vibrations (Fujino et al. 2012). Due to the origin of damping (*e.g.,* friction) it should be kept in mind that higher damping ratios are likely when response amplitudes are higher, such as during earthquakes and strong winds (Spence, S. and Kareem 2014).

Table 5.1 summarizes the statistical properties of the modal frequency and modal damping ratio for the operational modes identified. For each operational mode, the number of data sets for which the mode is identified is presented as a percentage of the 4426 data sets analyzed. The mean (μ) and scale parameter (s) of a best fit logistic distribution to the histograms of median modal frequency are documented. Treating the damping ratios, ξ, as log logistic distributed, Table 5.1 also reports the mean and scale parameters of $\ln(\xi)$ as well as the mode of the distribution based on a best fit log logistic distribution. Table 5.1 provides a statistical basis for modal frequencies and damping ratios for future designs of long-span suspension bridges such as the New Carquinez Bridges. For damping ratio in particular, the assumed 0.3% damping ratio assumed during the design phase of the bridge (OPAC 2016) was an accurate estimate that is now supported by the empirically derived modal damping ratios in this study.

5.4 Environmental and Operational Condition Impact on Modal Parameters

Modal parameters are well known to exhibit sensitivities to EOCs (Çatbaş et al. 2008; Cross et al. 2012; Fraser et al. 2010; Hua et al. 2007; Loh and Chen 2013; Moser and Moaveni 2011; Sohn et

al. 1999; Zhang et al. 2014). In this study, the sensitivity of modal parameters of the New Carquinez Bridge to the bridge environmental and operation conditions is investigated.

Specifically, three environmental and operation conditions are considered: temperature, wind, and traffic loads. While the temperature of the steel box girder and ambient air are both measured, it

Table 5.1, Modal frequency and damping ratio statistical properties of identified operational modes

Operational Mode	ID Rate (%)[*]	Modal Frequency, ω (Hz)		Modal Damping Ratio, ξ		
		μ_ω	s_ω	$\mu_{\ln(\xi)}$	$mode(\xi)$	$s_{\ln(\xi)}$
V-0.14	9.38	0.138	0.0028	-2.723	0.0066	0.1963
H-1	95.77	0.175	0.0029	-4.483	0.0113	0.4268
V-1	62.45	0.194	0.0006	-5.320	0.0049	0.5014
V-2	26.19	0.204	0.0015	-3.358	0.0348	0.2485
V-3	95.93	0.258	0.0005	-5.907	0.0027	0.4798
V-4	78.69	0.350	0.0010	-5.757	0.0032	0.5368
H-2	88.88	0.372	0.0020	-4.709	0.00912	0.4490
V-0.41	99.23	0.412	0.0019	-4.721	0.0089	0.2635
T-1	67.40	0.463	0.0035	-4.527	0.0108	0.4432
V-5	95.12	0.481	0.0016	-5.060	0.0064	0.2783
H-0.51	73.72	0.516	0.0092	-4.878	0.0076	0.3600
V-0.56	98.64	0.559	0.0017	-5.352	0.0047	0.2740
V-6	95.73	0.644	0.0015	-5.490	0.0041	0.3200
T-2	23.14	0.739	0.0020	-5.609	0.0037	0.4070
V-7	96.57	0.797	0.0018	-5.799	0.0030	0.3144
V-0.96	54.90	0.961	0.0030	-5.250	0.0053	0.3178
H-3	83.15	0.972	0.0047	-4.380	0.0125	0.3000
T-3	85.99	1.033	0.0028	-5.335	0.0048	0.3880
V-1.16	97.58	1.163	0.0031	-5.089	0.0062	0.2086
V-1.35	46.43	1.347	0.0043	-4.946	0.0071	0.2453
T-4	68.50	1.372	0.0083	-5.005	0.0067	0.4220
V-1.45	88.43	1.450	0.0041	-5.053	0.0064	0.2283
V-1.57	97.74	1.568	0.0036	-5.362	0.0047	0.2438
T-5	71.22	1.687	0.0049	-5.432	0.0044	0.2838
V-1.76	86.49	1.755	0.0048	-5.293	0.0050	0.2661
V-1.85	87.14	1.851	0.0043	-5.280	0.0051	0.2269
V-1.99	34.07	1.985	0.0042	-5.397	0.0045	0.2563

[*] Identification rate from 4426 data sets

is the temperature of the box girder that is used because it reflects the thermal condition of the structure. In addition, the wind speed and wind direction are measured by the monitoring system anemometers and wind vanes. Traffic load is not directly measured in this study so the study adopts two quantifiable parameters that correlate to traffic: the peak vertical acceleration of a data set and the time of the day. More vehicles moving faster increase the bridge vertical accelerations while time of the day (on a 24 hour scale) roughly captures traffic volume (*e.g.,* 17:00 experiences high traffic volume due to rush hour while 03:00 experiences very low traffic volumes).
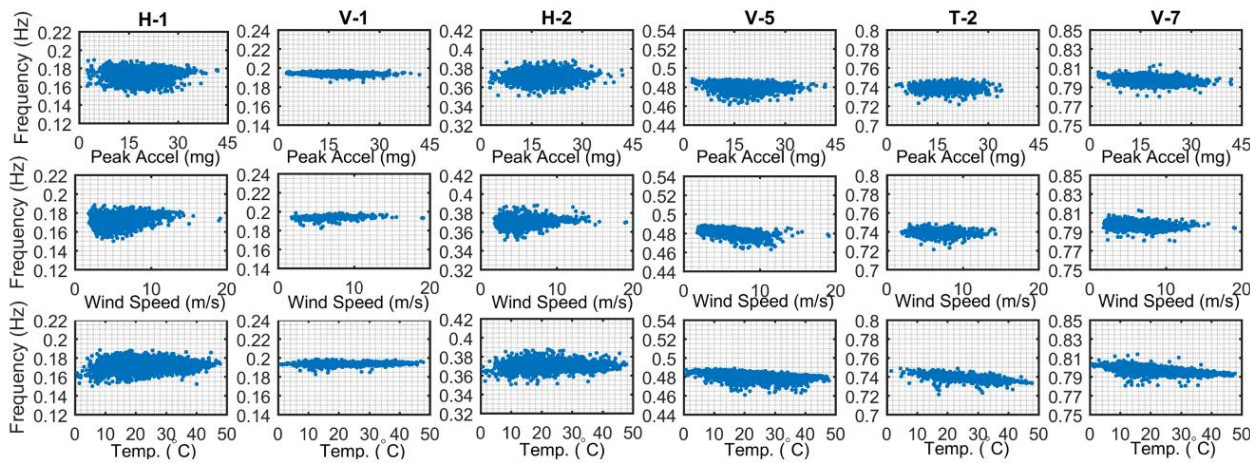
(a)
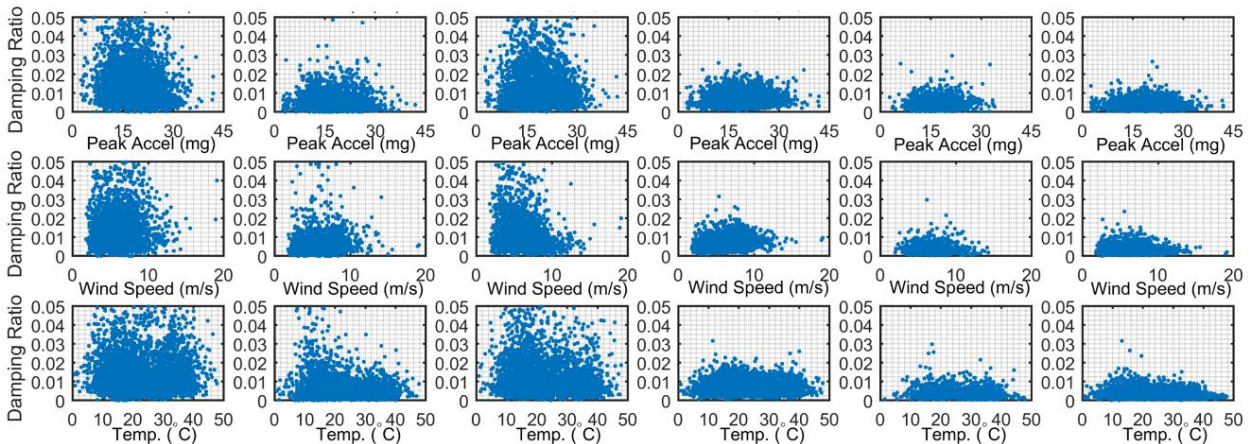
(b)

Figure 5.10, Relationship between modal, environmental (wind speed, temperature) and operational parameters (peak acceleration): (a) modal frequency, (b) modal damping ratio

Figure 5.10 shows the relationship between three EOC parameters (peak acceleration, wind speed, and temperature) and the model parameters for six representative modes (H-1, V-1, H-2, V-5, T-2, V-7). These six modes span frequency range and mode type (*i.e.*, horizontal, vertical and torsional). Close inspection of the modal frequencies reveal that many of the modal frequencies exhibit sensitivity to all three EOC parameters. For example, modal frequency trends downward with increasing temperature and increasing wind speed as is visually evident in modes V-5, T-2 and V-7. While some trends are evident in modal frequency, the modal damping ratios are too scattered for trends to be observed as a function of EOC parameters. As a result, regression analysis is performed on the modal frequencies to develop models between modal frequency and bridge EOC parameters. To quantify the relationship between EOC parameters and modal frequencies, two forms of regression analysis were performed: ridge regression and Gaussian Process Regression (GPR).

If the relationship between EOC parameters (system input) and modal frequency (system output) is assumed linear but there is a belief that the input parameters have a high degree of correlation, ridge regression would be an appropriate choice for regression due to its benefits in dealing with collinearity. Ridge regression represents a modification on traditional linear regression by penalizing the size of the regression coefficients through the introduction of regularization in the least square regression solution (Hoerl and Kennard 1970). Ridge regression begins with a linear model as:

$$y = X\beta \tag{5.25}$$

where $y \in \mathbb{R}^{n \times 1}$ is the modal frequency vector in this regression analysis, $X \in \mathbb{R}^{n \times m}$ is the feature vector (consisting of environmental and operation parameters), $\beta \in \mathbb{R}^{m \times 1}$ is the vector of

regression coefficients corresponding to each feature, n is the number of observations and m is the number of features. In ridge regression, the optimal set of regression coefficients, β, are those that minimize the following cost function:

$$minimize \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{m} \beta_j^2 \qquad (5.26)$$

where λ is the penalty term associated with large regression coefficients. The minimization is solved by least squares to yield the optimal regression coefficients, $\hat{\beta}$:

$$\hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda \boldsymbol{I}_p)^{-1}\boldsymbol{X}^T y \qquad (5.27)$$

where $\boldsymbol{I}_p \in \mathbb{R}^{m \times m}$ is the identity matrix. Ridge regression is dependent on the selection of the penalty term. To select an appropriate penalty term, λ, cross validation should be conducted on the data.

For this study, five input features are chosen: structural temperature (°C), wind speed (m/s), cosine of wind direction, maximum peak vertical acceleration (g) and hour of the day (hour of day on 24 hour clock); the output of the regression is the modal frequency of each mode. The first four input features and the output are continuous real variables while the hour of the day is a discrete integer parameter. Vectorization is used to handle the discrete parameter, which converts each hour into a 1 by 24 vector with 1 on the column of the hour and 0 on the rest. For example $[0,0,0,1,0 \dots ,0] \in \mathbb{R}^{1 \times 24}$ represents 04:00. Data normalization (*i.e.*, data is shifted and scaled to have a zero mean and unit variance) is also applied to both the input and output vectors. Half of the data is chosen randomly and used as the training data set to estimate β; the remainder of the data is then used to evaluate the regression performance. To select λ, the data randomly sampled once to create the training data set. Then, λ is varied from 0 to 1 in 0.1 increments to produce regression models for

each λ. The regression model with the minimum mean square error (MSE) on the predicted modal frequency using the remaining (testing) data corresponds to the best λ. After λ has been established for each mode, training and testing sets are again randomly chosen with ridge regression repeated 30 times; the mean value of the MSE is used to evaluate the regression performance (as summarized in Table 5.2). The MSE is normalized using the variance of the testing data. If the normalized MSE ≥ 1, the regression model cannot capture any relationships between the inputs and the output. If the normalized MSE is close to 0, the regression model perfectly captures a linear relationship. According to the results shown in Table 5.2, ridge regression performs well on most of the modes with the normalized MSE falling below 1 for all structural modes, and well below 0.5 for many of the modes (*e.g.,* V-6, V-7, T-3, T-5). An intrinsic benefit of ridge regression is the interpretation of the regression coefficients β. The larger the absolute value of the regression coefficient, the more that input parameter influences the modal frequency. Table 5.2 summarizes the average regression coefficient β associated with each operational mode. As can be seen, temperature is the most dominant regression input (with a mostly negative regression coefficient) for half of the modes with wind speed (with a mostly positive regression coefficient) and maximum acceleration (with a uniformly negative regression coefficient) the next two dominant inputs.

To explore if nonlinearity exists in the relationship between the EOC parameters and modal frequency, Gaussian process regression (GPR) is conducted using the same input and output features as those used in the ridge regression analysis. Gaussian process regression has received considerable attention in recent years due to its strong statistical foundation. A brief description of GPR is provided based on the description of the method in Rasmussen and Williams (2006). The regression model is stated as a stochastic (Gaussian) process:

$$y = f(\boldsymbol{X}) + \varepsilon \tag{5.28}$$

$$\varepsilon \sim N(0, \sigma^2) \tag{5.29}$$

$$f(\boldsymbol{X}) \sim \mathbb{N}(0, K) \tag{5.30}$$

where $y \in \mathbb{R}^{n \times 1}$ is the output vector of modal frequencies, $\boldsymbol{X} \in \mathbb{R}^{n \times m}$ is the feature set of environmental and operational parameters, and $\varepsilon \in \mathbb{R}^{n \times 1}$ is a zero mean Gaussian noise process of variance σ^2. The observation is considered from a zero mean Gaussian process with the covariance function K. In GPR, the joint distribution of observations, y, and model predictions, y^*, is assumed as:

$$\begin{bmatrix} y \\ y^* \end{bmatrix} |_{X,X^*} \sim \mathbb{N}\left(0, \begin{bmatrix} K(\boldsymbol{X},\boldsymbol{X}) + \sigma^2 \boldsymbol{I} & K(\boldsymbol{X},\boldsymbol{X}^*) \\ K(\boldsymbol{X}^*,\boldsymbol{X}) & K(\boldsymbol{X}^*,\boldsymbol{X}^*) + \sigma^2 \boldsymbol{I} \end{bmatrix}\right) \tag{31}$$

Where $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ is the unity matrix. Based on Bayesian inference, the posterior probability for y^* is defined as:

$$P(y^*|\boldsymbol{X}^*,\boldsymbol{X},y) \sim \mathbb{N}(\mu, \Sigma) \tag{32}$$

$$\mu = K(\boldsymbol{X},\boldsymbol{X}^*)[K(\boldsymbol{X},\boldsymbol{X}) + \sigma^2 \boldsymbol{I}]^{-1} y \tag{33}$$

$$\Sigma = K(\boldsymbol{X}^*,\boldsymbol{X}^*) - K(\boldsymbol{X},\boldsymbol{X}^*)[K(\boldsymbol{X},\boldsymbol{X}) + \sigma^2 \boldsymbol{I}]^{-1} K(\boldsymbol{X}^*,\boldsymbol{X}) \tag{34}$$

Many different covariance functions, K, can be considered. Here a kernel trick is applied with the square exponential (SE) kernel selected:

$$k(x_i, x_j) = \exp\left(\frac{\|x_i - x_j\|^2}{-2l^2}\right) \tag{35}$$

where x_i and x_j are the ith and j^{th} rows of $\boldsymbol{X}$ and l is a hyperparameter that is selected through maximization of the marginal log-likelihood function:

$$\log P(y|\boldsymbol{X}, l) = -\frac{1}{2}\log|K(\boldsymbol{X},\boldsymbol{X}) + \sigma^2 \boldsymbol{I}| - \frac{1}{2}y^T(K(\boldsymbol{X},\boldsymbol{X}) + \sigma^2 \boldsymbol{I})^{-1}y - \frac{n}{2}\log 2\pi \tag{36}$$

An advantage of GPR is that the regression model provides a statistical basis with the mean and variance of the regression offered by the GPR model.

Table 5.2, Mean Square Error of Regression Analysis of Environmental Impact on Modal Frequencies

Oper. Mode	Mean Square Error/Variance		Ridge Regression Average $\widehat{\beta}$				
	Gaussian Process Regression	Ridge Regression	Temper-ature	Wind Speed	Cos (Wind Dir.)	Max. Acc.	Hour[*]
V-0.14	1.00	1.17	0.11	0.06	-0.02	**-0.29**[***]	0.15
H-1	0.51[**]	0.66	**0.37**	0.24	0.17	-0.01	0.28
V-1	0.82	0.82	0.15	**0.26**	0.02	-0.15	0.19
V-2	0.85	0.86	0.24	0.08	0.05	**-0.27**	0.22
V-3	0.85	0.83	-0.20	**0.27**	0.02	-0.14	0.13
V-4	0.79	0.80	**-0.34**	0.16	0.01	-0.12	0.15
H-2	0.85	0.87	0.16	**0.19**	0.14	0.00	0.17
V-0.41	0.51	0.53	0.00	0.00	0.03	**-0.45**	0.25
T-1	0.99	0.99	0.07	0.05	0.00	0.08	**-0.10**
V-5	0.60	0.61	-0.24	**-0.32**	0.04	-0.19	0.13
H-0.51	0.45[**]	0.65	**0.31**	0.24	0.12	0.04	0.28
V-0.56	0.64	0.65	**-0.37**	-0.12	0.00	-0.25	0.14
V-6	0.43	0.43	**-0.53**	0.03	0.01	-0.20	0.17
T-2	0.55	0.55	**-0.64**	-0.05	-0.05	0.01	0.12
V-7	0.33	0.32	**-0.62**	-0.01	0.01	-0.16	0.18
V-0.96	0.67	0.67	**-0.36**	-0.04	0.04	-0.29	0.15
H-3	0.80	0.80	0.22	0.05	0.08	0.00	**0.23**
T-3	0.31	0.32	**-0.56**	0.02	0.02	-0.23	0.20
V-1.16	0.29	0.30	**-0.58**	-0.01	0.01	-0.19	0.20
V-1.35	0.66	0.66	**-0.46**	0.06	0.06	-0.15	-0.17
T-4	0.46[**]	0.57	-0.08	0.19	0.14	0.00	**0.28**
V-1.45	0.57	0.57	**-0.50**	-0.02	-0.02	-0.07	0.16
V-1.57	0.25	0.25	0.11	0.06	-0.02	**-0.29**	0.15
T-5	0.41	0.41	**0.37**	0.24	0.17	-0.01	0.28
V-1.76	0.55	0.55	0.15	**0.26**	0.02	-0.15	0.19
V-1.85	0.38	0.38	0.24	0.08	0.05	**-0.27**	0.22
V-1.99	0.34	0.33	-0.20	**0.27**	0.02	-0.14	0.13

[*] The maximum $\widehat{\beta}$ of hour 0 to 23 is used to represent the impact of time of the day
[**] Significant improvement using Gaussian Process Regression
[***] Highlighted columns labels the highest impact variable

In this study, GPR is applied to the same modal frequency (output, *y*) and environmental and operational data (input, *X*) as used for the ridge regression analysis. Normalized MSE is again used as the basis of evaluation of the regression model yielded. To handle the randomness in the selection of training and testing datasets, the regression is conducted 10 times and the mean MSE is used for comparison. Table 5.2 summarizes the mean normalized MSE for GPR applied to each operational modal frequency. The performance of GPR is shown to be similar to ridge regression. Operational modes involving modal participation in the transverse directions, H-1, H-2, H-0.51, T-2, H-3, T-3, T-4 and T-5, all have better results from GPR. For mode H-1, H-0.51 and T-4, GPR substantially improves the regression with mean normalized MSE reducing 0.15, 0.20 and 0.11, respectively. The fit of the GPR model for each modal frequency can be best seen when plotting
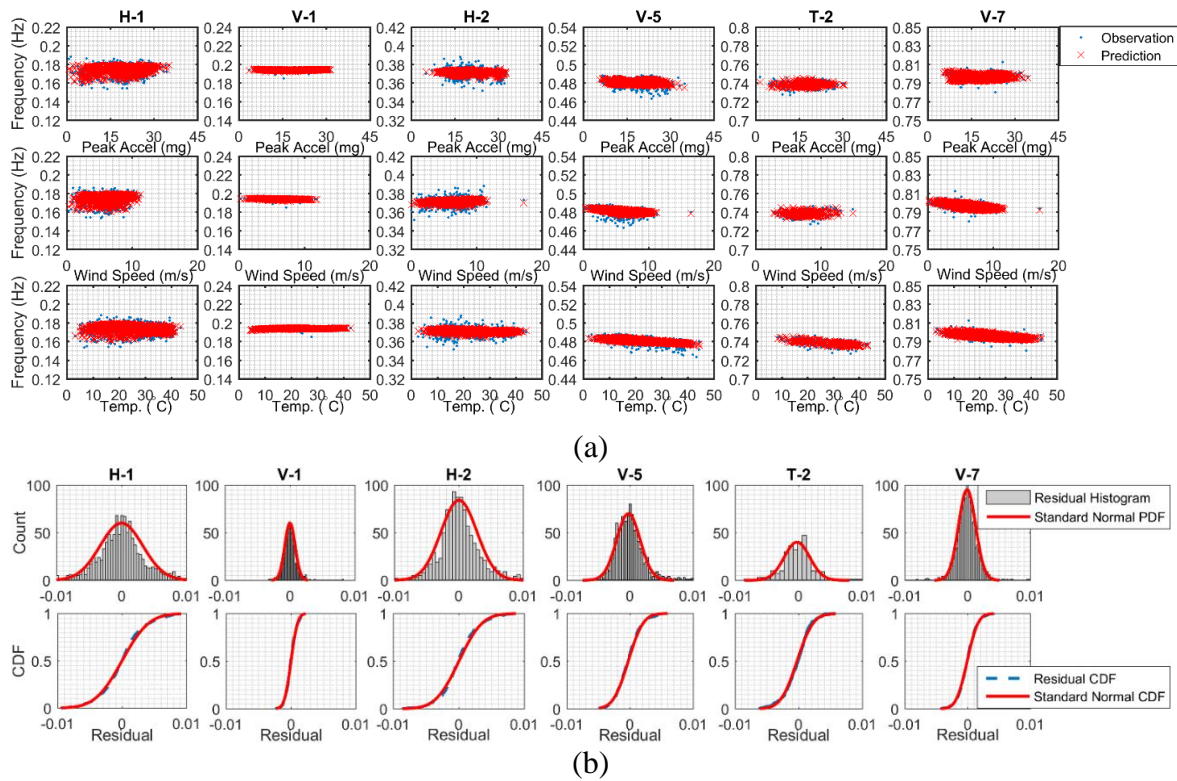

(a)

(b)

Figure 5.11, Gaussian process regression (GPR) model of modal frequency versus EOC parameters for selected modes: (a) GPR prediction versus observed data, (b) distribution and cumulative distribution function plot of GPR model residuals

106

predicted modal frequency as a function of the complete EOC vector. Projections of the GPR results on the modal frequency (Hz)-peak acceleration (mg), -wind speed (m/s), and –temperature (°C) spaces are presented in Figure 5.11 (a) for the same six selected modes of Figure 5.10. The residuals of the GPR results (*i.e.*, difference between the prediction and observation) of the six modes are plotted as histogram and cumulative distribution functions (CDF) in Figure 5.11 (b).

The probability density functions (PDF) as well as the CDF of standard normal distributions are plotted over the data histograms and the data CDF, respectively. The fitness between the GPR residual distribution and the standard normal distribution validates the Gaussian noise assumption in the GPR model (equation 5.28). In all of the cases, the GPR model accurately captures the modal frequency dependency on peak acceleration, wind speed and temperature. In all of the cases, the GPR model accurately captures the modal frequency dependency on peak acceleration, wind speed and temperature.

5.5 Chapter Summary

A long-term wireless structural monitoring system was first installed on the New Carquinez Bridge in 2010. Since that time, the wireless monitoring system has collected bridge responses (accelerations, strain and displacements) and bridge EOCs (temperature, wind speed, wind direction) daily for 8 minutes at 20 Hz every hour or every four hours, depending on energy availability. This has produced an extensive set of bridge response data that can be used for analysis of the bridge behavior under its normal loading environment. In this study, 4426 data sets collected between January 1, 2013 to December 31, 2014 were utilized to evaluate the long-term modal properties of the bridge. Motivation for the study was the importance of modal parameters

in both the design of the bridge and in updating bridge finite element models that are central to rapid assessment of the bridge health after an earthquake. The analysis of such a vast repository of response data necessitated fully automated processing strategies. The study developed a consistent approach to the application of SSI for the estimation of an output-only state space model from which modal properties were extracted. A key innovation of the study was the use of multiple projection sizes (as specified by the size of the Hankel matrix row space) for each data set with the median modal results utilized. By standardizing the implementation of SSI, modal parameters extracted could be fairly compared. To ensure scalability of the modal parameter extraction strategy, a limited number of sensor outputs (6 deck acceleration measurements at node N1, N5 and N10) were considered in the SSI implementation. A novel approach to the creation of a modal frequency histograms was introduced that aggregated the result of five SSI models with different Hankel matrix row spaces ($i = [210,220,230,240,250]$) to ensure modal frequencies are easily identified. Using the histogram, a total of 27 operational modes were identified from the results of which 15 were confirmed to correspond to previously reported structural modes. With accurate estimation of modal frequencies, the median modal frequency and damping ratio of each mode was reported for each data set analyzed allowing extensive histograms of modal frequency and modal damping ratio to be created. Investigation of the histograms revealed that the normal distribution was not an appropriate statistical model for describing the distributions due to high kurtosis levels. The logistic function was determined to be the best statistical model that accurately fit the modal frequency distributions. Similarly, log-logistic distributions proved very accurate in modeling the distribution of modal damping ratio. The empirically derived distributions for modal damping revealed that the mode of the distributions for most structural modes were well below 0.8% with many low order modes less than 0.3% such as the first, third and fourth vertical

structural modes. These low modal damping ratios confirm that the 0.3% damping ratio used in the design of the New Carquinez Bridge was a reasonable value. Finally, the study developed two regression frameworks for mapping the relationship between the bridge EOCs and modal frequency. A GPR model proved to provide relatively accurate estimates of modal frequency given the bridge temperature, wind speed, wind direction, peak vertical acceleration and time of day.

As time progresses, future efforts should update the statistical models developed using newly collected bridge response and EOC data. Given the high uncertainty associated with estimations of modal damping, future work should explore approaches that aim to reduce the uncertainty in modal damping ratios. Due to the origin of damping, more data is needed for the bridge under larger levels of excitation (e.g., earthquake and strong wind loads) from which modal damping can be explored. It is anticipated that higher responses in the bridge will lead to increases in modal damping as has been reported in the literature for other bridges. Finally, the significance of the variations reported in modal parameters should be studies using a finite element model of the bridges. Specifically, the sensitivity of critical structural responses investigated after earthquakes (such as stress levels in key load carrying components) should be evaluated under the normal variation in modal parameters.

CHAPTER 6

DATA-DRIVEN REGRESSION ANALYSIS FOR BRIDGE WEIGH-IN-MOTION

6.1 Introduction

Identification of moving loads on highway bridges is critical for assessing structural performance. In the domain of bridge structural health monitoring (SHM), direct estimation of load on the bridge system can enhance understanding of structural response data. Weigh-in-motion (WIM) stations are routinely used to record load information on vehicles passing a specific location within a highway network. WIM stations are typically installed in the highway pavement system but unfortunately not on bridge decks. Furthermore, existing WIM stations may be located far from bridges of interest. WIM stations, regardless of their design (*e.g.*, bending plates, piezoelectric sensors), are designed to record the number of axles and the *static* weight imposed by the axle to the road. Translating measured static load of a vehicle to true dynamic load imposed on a bridge requires extensive analysis which must account for the dynamic properties of a vehicle's suspension system, the dynamic properties of the bridge, and the roughness of the bridge deck. An alternative strategy to identifying and characterizing the dynamic loads imposed on bridges is to instrument the bridge in a manner that allows it to be used as a WIM station of sorts. A clear advantage of this strategy is that the response of the bridge can also be used for other purposes such as for structural health identification.

Several bridge WIM (BWIM) systems developed in academia and industry exist and typically consist of sensing and data processing units. A typical sensing unit, the nothing-on-road (NOR) system, uses a group of strain gages as an axle detector which take advantage of local stress peaks caused by the wheel passing the instrumented area. Axle information is extracted by identifying the sharp peaks in the strain response data and is later used as critical input to the weight identification process. This type of axle detector is highly sensitive to the wheel path, thus reducing the success rate of the truck identification. In terms of data processing, weight identification is constructed as a regression problem constructed from the axle influence line and bridge strain response. Having accurate axle influence line, speed and axle configuration is critical to the accuracy of the weight estimation. Most BWIM system obtain the influence line via structural modeling or finite element modeling, which can introduce errors related to differences between the model and structure.

Research presented in this chapter focuses on the extension of the bridge condition monitoring systems for bridge weigh-in-motion measures in order to obtain explicit traffic load information. The algorithm developed aims to remove the need for axle detectors, reduce identification error from physical modeling, add robustness against highway traffic and automate the identification process for long-term monitoring. In this chapter, a data-driven regression method is proposed to identify truck gross weight, axle weight, axle spacing and speed with a single strain gage sensor. The algorithm uses experimental influence lines and an expansion of the Moses's algorithm to a ℓ_0-norm regression problem. To solve the ℓ_0 regression, a ℓ_1-norm algorithm least absolute shrinkage and selection operator (LASSO) is applied to estimate the solution; the best subset selection method is used to extract the exact axle spacing and weight. Truck speed is determined

based on best fitting of the resampled influence line during regression analysis. The data-driven regression method removes the needs for dense sensor arrays, making the system a good candidate for quick installation short-term traffic evaluation and ultra-low cost long-term traffic monitoring. The system was installed and tested on the Telegraph Road Bridge, a highway bridge in Monroe Michigan, and achieved a type II (ASTM 2009) WIM station accuracy. The regression algorithm was also applied to track truck weight using long-term data from May 2015 to April 2016. The truck weight distribution from the proposed approach is in high agreement with weight records from the downstream MDOT owned WIM station, demonstrating the long-term reliability of the bridge WIM system and success of the proposed algorithms.

6.2 Data-driven Bridge Weigh-In-Motion Algorithm Overview

The data-driven BWIM algorithm discussed in this chapter is based on best fitting of structural strain response data with experimentally obtained influence lines. The algorithm can extract truck total weight, speed, number of axles (tandem), axle weight and axle spacing. It can also be automated, a significant advantage for long-term operations. Similar to traditional methods, strain response is used to estimate axle weight. However, this algorithm also extracts the number of axles, axle spacing and speed information from the same strain response data, thereby removing the need for an axle detector. The algorithm has three major steps as shown in Figure 6.1.

- Step 1: An experimental static influence line is extracted using calibration vehicle information and measured strain data.
- Step 2: A ℓ_0 normalization problem is constructed using measured strain (as the dependent variable) and a Toeplitz matrix of the time-shifted experimental influence line (as the

112

independent variables). Least absolute shrinkage and selection operator (LASSO) regression, which is a ℓ_1 solver, is used to estimate the ℓ_0 regression due to the computational demand of the ℓ_0 solution. It is repeated for many Toeplitz matrices over a range of speeds (each matrix is constructed with influence lines of a single speed). The best fit solution thereby provides a best fit estimation of vehicle speed and potential axle/tandem locations.

- Step 3: Based on potential axle or axle group location information, a list of combinations from all potential axle occurrence time stamps is constructed. For each combination in the list, least square linear regression is applied using the strain from the last step (as the dependent variable) and a subset of the Toeplitz matrix (as the independent variable). The best fit solution provides number of axles, axle weight, axle spacing and total weight of the vehicle.

1. Extract experimental
influence line
using least square

Influence
Line

2. Extract vehicle speed and
potential axle locations
using L0 normalization ⟶ Speed

Axle/Tandem
Locations

3. Extract axle weight
and spacing
using least square ⟶ # of axle
Axle weight
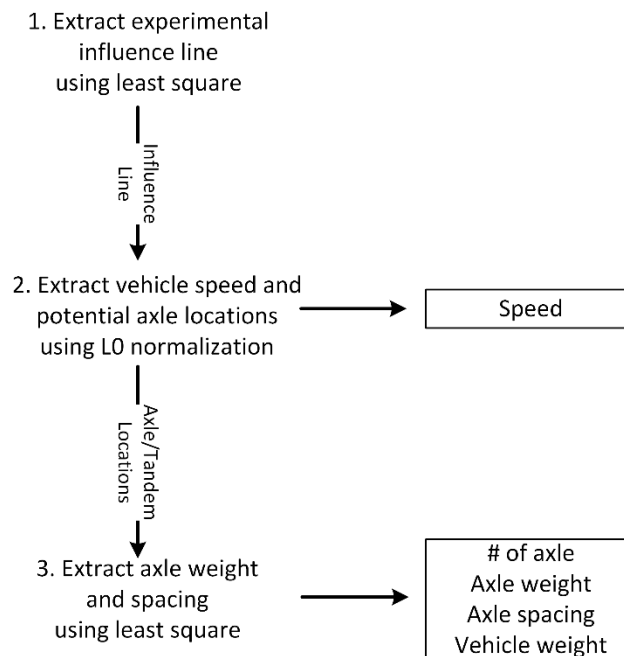Axle spacing
Vehicle weight

Figure 6.1, Flow chart of data-driven truck identification algorithm

The strain gage used for the bridge WIM regression analysis is installed near the bottom pin hole on a hanger plate (as shown in Figure 2.9) of the 4th girder (middle lane) on the leaving span. The stress concentration effect around the pin hole magnifies the strain response thereby improving the signal-to-noise ratio. The weldable axial water-proof strain gage (HBWF-35-125-6-10GP-TR) from Hitec Products Inc. (http://www.hitecprod.com/) measures both static and dynamic responses; an external signal conditioning circuit board is used to magnify the strain gage output signal by a factor of 500. The installation required a lift truck near the bridge pier and did not disrupt traffic on or below the bridge. Since the strain gage is installed as an add-on component to an existing permanent structural monitoring system, the cost was minimal, including $100 for the weldable stain gage, $100 for the wireless *Narada* node, $133 for the solar power supply and roughly $60 for miscellaneous circuits and packing materials. At a total additional cost of $400, the existing bridge monitoring system was equipped with a WIM system as well, proving the exceptional value addition by the proposed BWIM approach.

There are four entrances/exits between the TRB site and the WIM station located 12 miles downstream from the TRB. In order to synchronize the bridge strain response collected at the TRB and the corresponding truck data recorded at the WIM station, two cameras were setup at the TRB site and the WIM station site to record continuous video images of the traffic for truck synchronization. Based on the preliminary video test, nearly 70% trucks passing the TRB will pass the WIM station, therefore providing a basis to that similar traffic characteristics are expected between the TRB and the WIM station. Additionally, the video at the TRB site is used to synchronize a particular truck event with the corresponding strain signal and the video at the WIM

station is used to extract the corresponding vehicle records based on time stamp and consecutive vehicle events.

6.3 Influence Line Extraction

The algorithm used to extract the influence line using experimental sensing data is a modified version of the algorithm discussed in O'Brien et al. (2006). At time step k, the load in the hanger plate, S_k is given by:

$$S_k = E\varepsilon_k^t \tag{6.1}$$

where E is the Young's modules of hanger plate and ε_k^t is the theoretical tensile strain of the hanger plate at time step k.

The load can also be expressed as the superposition of axle loads:

$$S_k = \sum_{i=1}^{N} P_i I_{k-C_i} \tag{6.2}$$

$$C_i = \frac{D_i f}{v} \tag{6.3}$$

where P_i is the axle load of the ith axle, N is the total number of axles, , D_i is the distance between axle i and $i-1$ (D_1 is set to be 0), f is the sampling frequency, v is the speed of the vehicle, C_i is the number of data points offset between axle i and $i-1$ and I_{k-C_i} is the influence line ordinates of the ith axle at time stamp k (as shown in Figure 6.2). Combing equations 6.1 and 6.2, ε_k^t is expressed as:

$$\varepsilon_k^t = \frac{1}{E}\sum_{i=1}^{N} P_i I_{k-C_i} \tag{6.4}$$
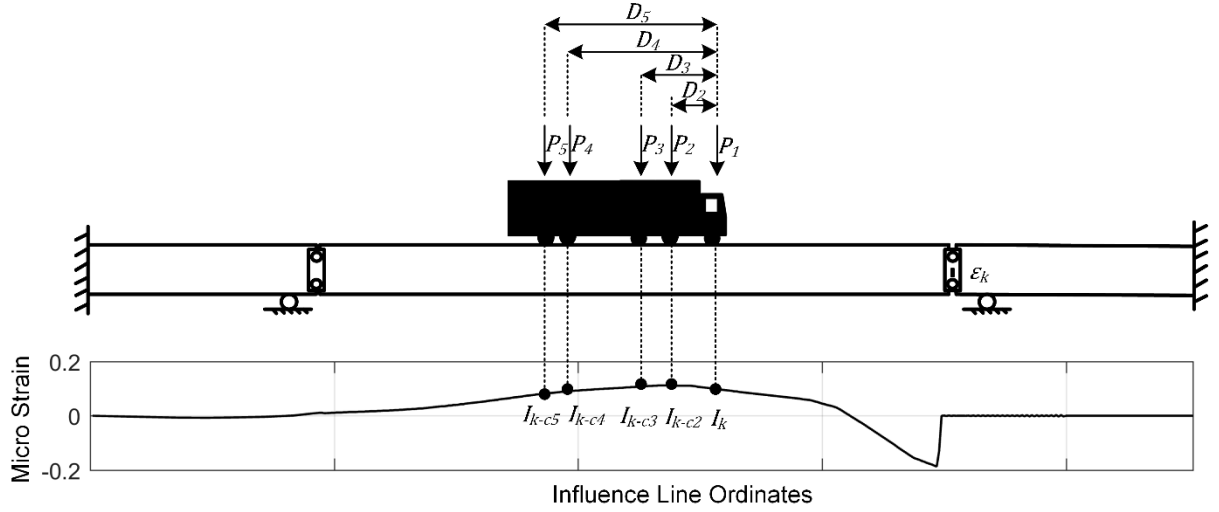
115

Figure 6.2, Influence line ordinates of a 5 axle truck at time stamp k

Define the error between the measured strain response and the theoretical values as:

$$Error = \sum_{k=1}^{K}(\varepsilon_k^m - \varepsilon_k^t)^2 = (\varepsilon_1^m - \varepsilon_1^t)^2 + (\varepsilon_2^m - \varepsilon_2^t)^2 + \cdots + (\varepsilon_K^m - \varepsilon_K^t)^2 \qquad (6.5)$$

Take equation 6.4 into equation 6.5, the error is expressed as:

$$Error = \frac{1}{E^2}[(\varepsilon_1^m - P_1 I_1)^2 + \cdots + \left(\varepsilon_{1+C_2}^m - P_1 I_{1+C_2} - P_2 I_1\right)^2 + \cdots$$

$$+ \left(\varepsilon_K^m - \sum_{i=1}^{N} P_i I_{k-C_i}\right)^2] \qquad (6.6)$$

To minimize the error, partial derivatives of the error with respect to a set of influence ordinates are set to be zero:

$$\frac{\partial Error}{\partial I_R} = 0, R = 1,2,\dots,K - C_N \qquad (6.7)$$

The $K - C_N$ equations from the partial derivations can be re-organized into matrix format:

$$[W]_{(K-C_N)\times(K-C_N)} \times \{I\}_{(K-C_N)\times 1} = \{\varepsilon\}_{(K-C_N)\times 1} \qquad (6.8)$$

Where the I is the influence line ordinates vector. The rows of $\{\varepsilon\}$ are defined as:

$$\varepsilon_R = E\left(P_1\varepsilon_R^m + P_2\varepsilon_{R+C_2}^m + \cdots + P_N\varepsilon_{R+C_N}^m\right), R = 1,2,\ldots,K - C_N \tag{6.9}$$

The elements of $[W]$ are defined as:

$$W_{i,i} = \sum_{i=1}^{N} P_i^2$$

$$W_{i,i+(C_j-C_k)} = P_j P_k, i + \left(C_j - C_k\right) < K - C_N \tag{6.10}$$

$$i = 1,2,\ldots,K - C_N, j > k, j = 1,\ldots,N, k = 1,\ldots,N$$

$\{I\}$ can be solved by taking the least square solution of equation 6.8.

Figure 6.3 shows the influence line extracted from experimental strain response data from a calibration vehicle load. The calibration vehicle was chosen to be a two axle truck with axle weights of 7.8 kips and 11.2 kips and axle spacing of 22.8 ft. The super positioned influence line was computed according to equation 6.4. As the experiment was conducted without traffic interruption, ideal calibration truck conditions (bridge at rest before the truck arrives, truck driving on the middle lane and no other truck on the bridge at the same time) existed. The specifications of the calibration truck (*i.e.*, axle weight and spacing) were obtained from the WIM station. The speed of the truck, which was used when computing C_i, was unknown but could be determined using a grid search. A range of speeds were used to compute the C_i individually and the speed resulting in the minimum squared sum error, $\sum_{k=1}^{K-C_N}(\varepsilon_k^t - \varepsilon_k^m)^2$, was used as the calibration truck speed, which was 96 ft/s (65 mph). The experimental influence line follows the same trend as the one from the CSiBridge model, but more accurately represents the influence line of the structure. Therefore, the experimental influence line shall be used for future analyses.
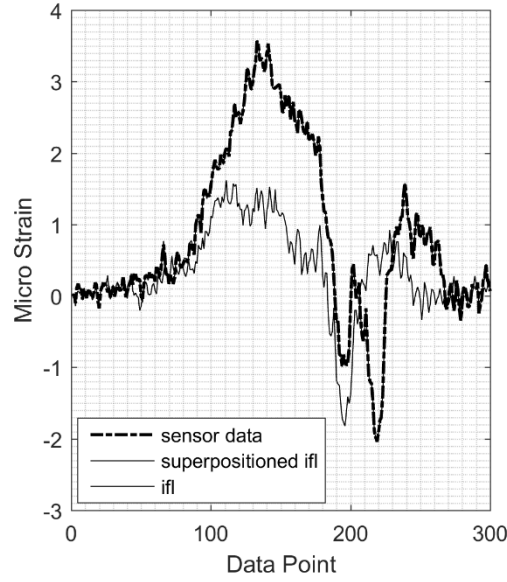
Figure 6.3, Influence line extracted from experimental sensing data

6.4 Speed and Potential Axle/Axle Group Location Extraction

6.4.1 Axle/Axle Group Estimation

Estimation of the axle weight and spacing of a truck can be treated as a regression problem between the measured strain response vector, $\{\varepsilon\}_{m \times 1}$, and the superposition of the influence line vector. The regression problem is constructed as:

$$[I]_{m \times t} \times \{P\}_{t \times 1} = \{\varepsilon\}_{m \times 1} \tag{6.11}$$

where m is the number of data points in the strain response, ε, and t is the number of time-shifted influence lines, $\{I\}$ from equation 6.8, comprising the Toeplitz matrix $[I]$,

$$[I]_{m \times t} = \begin{bmatrix} I_1 & 0 & \cdots & 0 & 0 \\ I_2 & I_1 & \cdots & \vdots & \vdots \\ \vdots & I_2 & \cdots & 0 & 0 \\ I_K & \vdots & \cdots & I_1 & 0 \\ 0 & I_K & \cdots & I_2 & I_1 \\ 0 & 0 & \cdots & \vdots & I_2 \\ \vdots & \vdots & \cdots & I_K & \vdots \\ 0 & 0 & \cdots & 0 & I_K \end{bmatrix} \tag{6.12}$$

118

and $\{P\}$ is the axle weight sparse vector, where the non-zero elements represent axle weights and the number of zeros between the neighboring non-zero elements can be translated to axle spacing:

$$\{P\}_{t\times 1} = [0 \quad ... \quad 0 \quad P_1 \quad 0\times C_2 \quad P_2 \quad 0\times C_3 \quad P_3 \quad ... \quad P_N \quad 0 \quad ... \quad 0]^T \tag{6.13}$$

To solve for the axle weight vector, $\{P\}$, the regression problem is subject to:

$$\{P\} = argmin\left\{\sum_{i=1}^{m}(\varepsilon_i - \alpha - \sum_{j=1}^{t}P_j I_{ij})^2\right\} \; subject \; to \; \sum_{j=1}^{t}\|P_j\|_0 = N \tag{6.14}$$

where α is the intercept, which is a small constant correcting the offset of the strain signal. This is therefore a ℓ_0-norm regularized regression problem.

However, ℓ_0-norm regularization problem is computationally hard. It is often solved using the best subset selection (BSS) algorithm which evaluates all $\binom{t}{N}$ models that have exactly N non-zero elements of $\{P\}$, and then selects the best model. One obvious problem of BSS is that if t is non-trivial, the number of models becomes too large to implement. Therefore, ℓ_1-norm regularization is often used to estimate the ℓ_0-norm solution. The least absolute shrinkage and selection operator (LASSO) algorithm (Tibshirani 1996) is applied to solve:

$$\{P\} = argmin\left\{\sum_{i=1}^{m}(\varepsilon_i - \alpha - \sum_{j=1}^{t}P_j I_{ij})^2\right\} \; subject \; to \; \sum_{j=1}^{t}\|P_j\|_1 \leq w \tag{6.15}$$

where w is a positive tuning parameter. The solver for Lasso solutions is an iterative algorithm that uses the Kuhn-Tucker conditions (Kuhn and Tucker 1951) and soft shrinkage function. The detailed algorithm can be found in (Tibshirani, 1995). The solver returns a set of $\{P\}$ for a range of w, and the best fitting $\{P\}$ and w are determined using cross-validation of the squared sum error (SSE) defined as:

$$SSE = \sum_{i=1}^{m} \left(\varepsilon_i - \alpha - \sum_{j=1}^{t} f(P_j) I_{ij} \right)^2$$

(6.16)

$$f(P_j) = \begin{cases} P_j, if \ P_j \geq l \\ 0, if \ P_j < l \end{cases}$$

As P_j essentially means axle weight of the truck, which is usually larger than 2 kips, the threshold function $f(P_j)$ eliminates small P_j, so the SSE focuses on the fitness of the axle influence line instead of noises. The threshold, l, can be set to a constant number or can be determined dynamically based on the peak of the strain response (*i.e.*, the higher the strain, the larger l is).

6.4.2 Truck Speed Extraction

To determine the truck speed, a range of possible truck highway speeds are tested and the best model is selected. As the calibration truck had a speed of 96 ft/s (65 mph), several influence lines corresponding to a reasonable speed range of [88,110] ft/s ([60,75] mph) were obtained by resampling the calibration influence line, discretized by 1 ft/s, resulting in a set of 23 influence line vectors. Therefore, the Lasso regression is repeated 23 times using 23 different Toeplitz matrices (equation 6.11). The model with the smallest SSE is selected and the corresponding speed is used to represent the truck speed.

6.4.3 Validation and Discussion

To partially validate the proposed algorithm, the calibration truck sensor data shown in Figure 6.3 is used as the measured strain response data, $\{\varepsilon\}$, in equation 6.11, and the Lasso solution is shown in Figure 6.4. Figure 6.4 (a) shows that the Lasso solution fits the sensor data very well and is
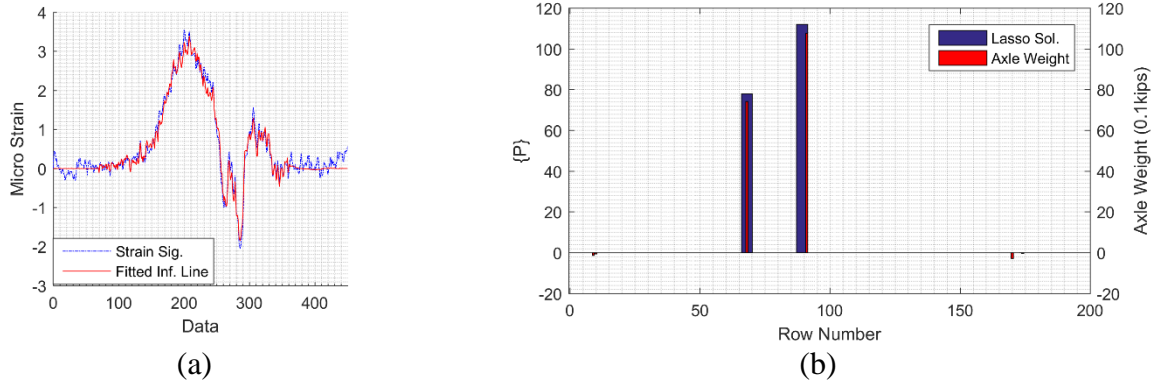
Figure 6.4, Lasso solution of axle weight and axle spacing for the calibration truck: (a) measured strain and fitted strain and (b) Lasso solution of $\{P\}$ and axle weight and spacing from WIM station

robust to the noisy data at the ends caused by consecutive vehicles and structure dynamics. Figure 6.4 (b) is a bar plot showing the Lasso solution of $\{P\}$ and truck axle information from the WIM station (axle weights of 7.8 kips and 11.2 kips, axle spacing of 22.8 ft and speed of 96 ft/s). The Lasso solution has eight non-zero elements, but only two of which have significant positive values (row 68 and 91 in Figure 6.4 (b)), representing the two axles of weight 7.4 kips and 10.8 kips, respectively. Axle spacing is computed as $\frac{91-68}{100Hz} \times 96ft/s = 22.08ft$. The validation test accurately identified axle weight and spacing, thereby showing the effectiveness of the LASSO estimate to the ℓ_0 regression solution.

The results of a 5-axle truck (axle weight of 9.3, 15, 15.9, 17.9 and 17.3 kips and axle spacing of 11, 4.4, 30.8 and 10.2 ft., according to WIM station records) is also shown to demonstrate the proposed algorithms on regular traffic data (non-calibration vehicle data). When using regular traffic data, due to signal noise, structure dynamics and slight differences in truck path (with respect to the influence line path, *e.g.,* truck shifted in transverse direction within the lane), higher

121

rank Lasso solutions always produce better fits (lowest squared sum errors), as shown in Figure 6.5 (a1). However, the amplitude of these non-zero elements are often not accurate as shown in Figure 6.5 (a2). These are referred to as "over-fitted" solutions, and can be eliminated by limiting the maximum number of non-zero elements in $\{P\}$. Therefore, an additional filter was applied to $\{P\}$:

$$\sum_{j=1}^{t} \|P_j\|_0 < n \tag{6.17}$$

This is similar to the ℓ_0-norm condition with the upper boundary labelled as n. A smaller n can eliminate "over-fitted" solutions. However, too few n (*e.g.*, $n = N$) can also lead to an "under-fitted" solution, where the Lasso solution only concentrates near the dominant axles, or takes a position between axles and omits the remaining axles as shown in Figure 6.5 (b1) . This is also the result of signal noise, structure dynamics and a difference in truck path from the influence line path. The Lasso solution "spreads" the axle weight among the neighboring elements, so a small n will force the algorithm to shrinkage the light weight axles and pick a compromised position between axles. For trucks with less than five axles, $n = 30$ is a good choice as shown in Figure 6.5 (c1) and Figure 6.5 (c2). As shown in Figure 6.5 (c1), the "well-fitted" Lasso solution of $\{P\}$ is a good fit in the time history strain response, especially the latter half of the data where axle loads cause sharp peaks. The non-zero elements of $\{P\}$ shown in Figure 6.5 (c2) are spread into three groups (near row 23, 35 and 77, with more elements concentrated near row 35 and 77) which can be treated as a single axle or a tandem of two axles. In conclusion, the ℓ_0-norm regularized regression modelling estimated using Lasso has been shown to provide accurate information on truck axle/tandem weights and spacing.
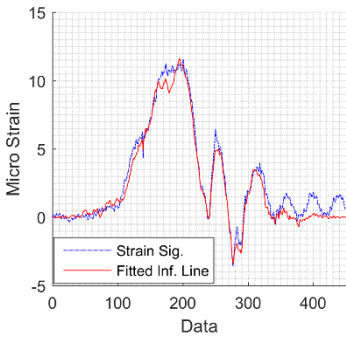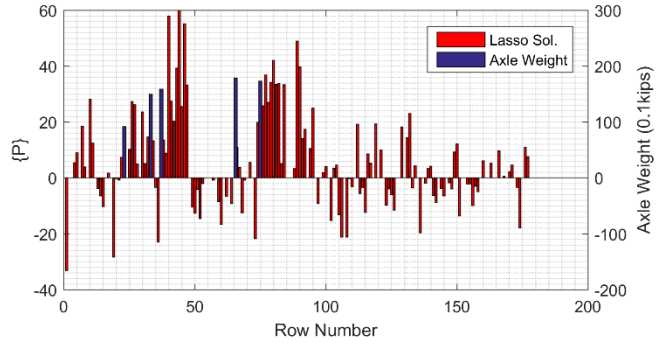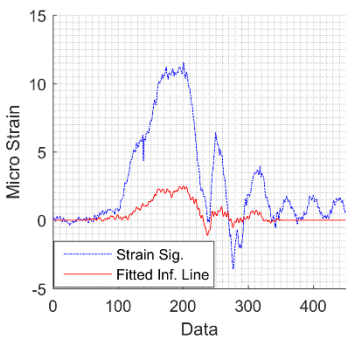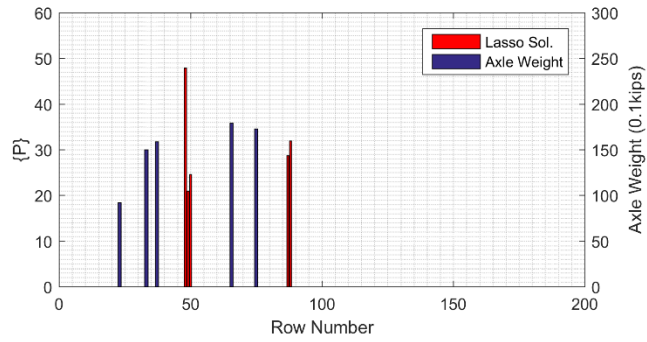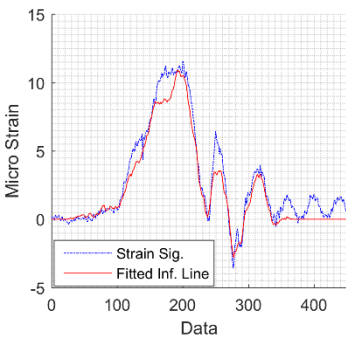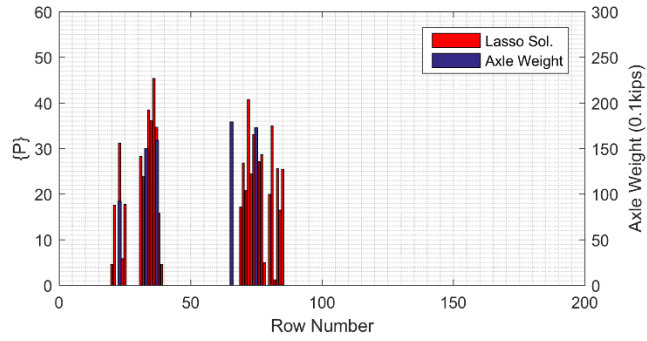
Figure 6.5, Lasso solution of axle weight and axle spacing for: (a) "over-fitted" condition, $n = 180$, (b) "under-fitted" condition, $n = 5$ and (c) "well-fitted" condition, $n = 30$; (1) measured strain and fitted strain and (2) Lasso solution of $\{P\}$ and axle weight and spacing from WIM station

6.5 Number of Axles, Axle Weight and Spacing Extraction

By revealing estimations of truck speed and axle locations (within a range) using ℓ_1 methods, the originally posed regression problem can be re-posed but with greatly reduced dimensions (focused only on the estimated speed and axle location dimensions). With the reduced problem setup, ℓ_0 methods become tractable using the BSS algorithm. Reduced models are constructed by:

1) Apply a threshold function (from equation 6.16) on Lasso solutions to zero out insignificant elements.

2) Form N_g groups of row numbers from the Lasso solution, based on neighboring positive elements. Each group contains N_{gi} elements (i is the group number) and represents an axle or a tandem.

3) For each group, construct the possible combinations (2-combination for tandem and 1-combination for single axle) of the row elements, resulting in a total of $\binom{N_{gi}}{2} + \binom{N_{gi}}{1}$ combinations.

4) Construct the possible combinations of axles for all N_g groups, with only one single/tandem axle selected from each group. In total, there are $\prod_{i=1}^{N_g}\left[\binom{N_{gi}}{2} + \binom{N_{gi}}{1}\right]$ models; each model has N elements.

For each model, re-construct the Toeplitz matrix from equation 6.11 by extracting the columns according to the row number elements in the N_g-combination. The regression problem is re-structured as:

$$[I]_{m \times N} \times \{P\}_{N \times 1} = \{\varepsilon\}_{m \times 1} \tag{6.18}$$

which can be solved using least square. Models with non-physical solutions (*e.g.*, negative and insignificant values) should be discarded. The best fitting model is selected by evaluating the squared sum error.

6.5.1 Validation and Discussion

After applying the threshold function on the Lasso solution from last step, the solution could be divided into 3 groups of row numbers: $[23], [31:37], [70:85]$ having significant values. Therefore, totally 3808 models are evaluated. The proposed algorithms successfully picked a best fitting model with $N = 5$ axles, having row numbers of $[23, 33, 37, 72, 84]$. At a speed of 92 ft/s and considering 100 Hz sampling frequency, the row numbers are converted into axle spacings of $[9.2, 3.68, 32.2, 11]$ ft. The least square solution for this model is $\{P\} = [10.8\ 11.2\ 14.2\ 23.1\ 15.9]^T$ kips. The difference between the proposed solution and the truck data received from the WIM station is listed in Table 6.1. The total weight difference is 0.3%, individual axle weight ranges from 8.1% to 29%, and axle spacing ranges from 4.5% to 16%.

Table 6.1, Truck information comparison between regression results and WIM station data

	Total Weight	Weight 1	Spacing 1 to 2	Weight 2	Spacing 2 to 3	Weight 3	Spacing 3 to 4	Weight 4	Spacing 4 to 5	Weight 5
WIMS	75.4	9.3	11	15	4.4	15.9	30.8	17.9	10.2	17.3
Regression	75.2	10.8	9.2	11.2	3.68	14.2	32.2	23.1	11	15.9
Difference	-0.3	16	-16	-25	-16	-11	4.5	29	7.8	-8.1

6.6 Results and Discussion

Of all 189 truck events considered in this experiment which crossed the TRB, 150 of them also crossed WIM station, but only 42 of them satisfied identification conditions: 1) driving on the middle lane and, 2) being the only vehicle on the bridge when passing the TRB. Therefore, the data-driven truck identification algorithm was applied only to these 42 truck events. Among the 42 trucks, 4 were 2-axle, 3 were 3-axle, 3 were 4-axle and 32 were 5-axle. In total, 34 of these truck events resulted in a successful number of axles detected (according to WIM results). Of the 8 unsuccessful cases, 1 was mis-identified by the WIM station (a 5-axle truck as a 3-axle and 2-axle), 4 occurred on trucks with multiple axles where there was difficulty separating tandem from single axle, and 3 occurred on trucks with low axle loads (low values in $\{P\}$).

Figures 6.6 (a) and 6.6 (b) show a distribution of truck weights from the WIM station records and identified using the data-driven regression algorithm. Figure 6.6 (c) shows the distribution of total weight difference between the regression analysis and the WIM station records. A difference of 15% or less is shown for 91% of trucks. As WIM station records are not necessarily a true representation of the truck total weight, further analysis was performed to evaluate the quality of the regression analysis. Given the WIM station has an accuracy of 95% trucks with error of 15% for truck total weight, the error of the WIM station record is simulated as a random variable, S_1, from normal distribution of $N(0,0.075)$. Assuming the regression analysis also carries the accuracy of type II WIMS, a random variable, S_2, is draw from the same distribution. The histogram of the difference between S_1 and S_2 is normalized and plotted as reference to evaluation of the quality of the regression
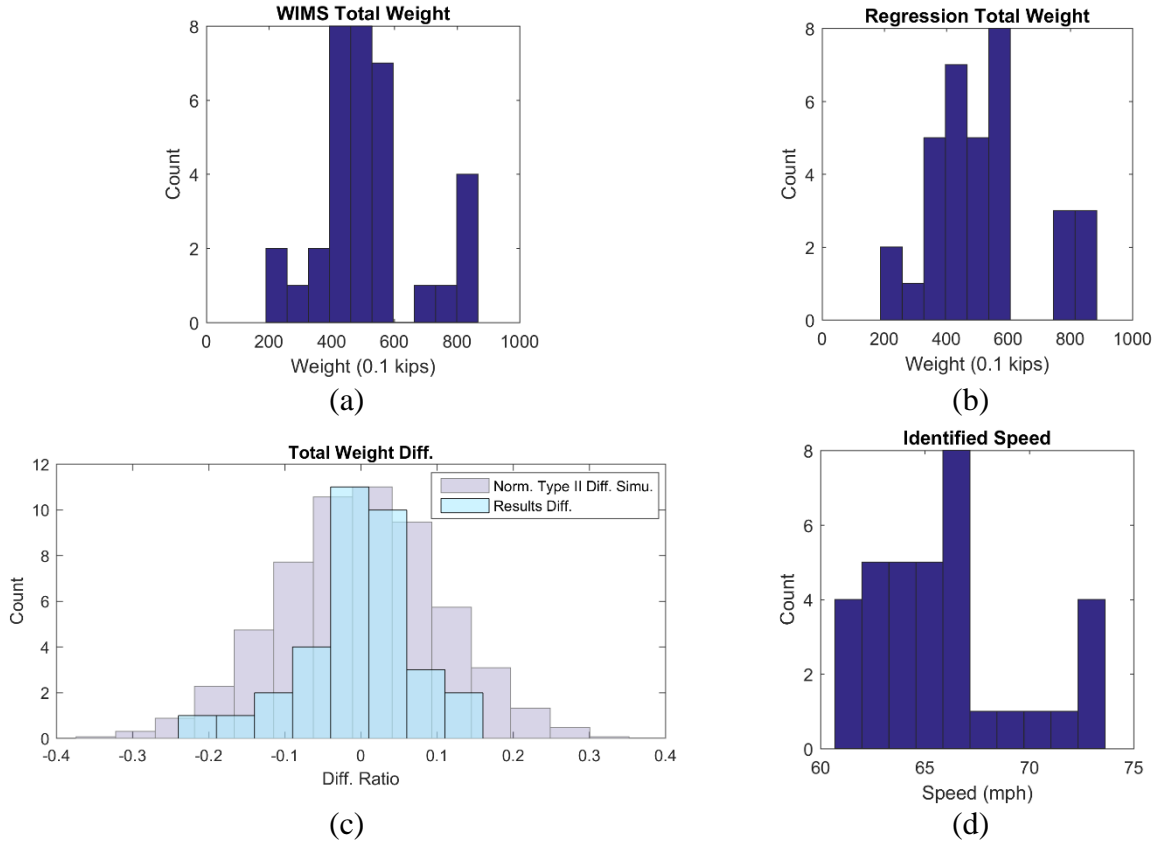
Figure 6.6, Distribution of identified truck weight and speed: (a) truck weight from WIMS record, (b) truck weight identified using data-driven regression algorithm, (c) truck weight difference and (d) speed identified using data-driven regression algorithm

analysis. As shown in Figure 6.6 (c), the difference between the regression analysis and the WIM station records is under the area of the difference between two simulations, meaning the regression analysis is likely to have a smaller error variance than the type II WIM station and performing at least as good as a type II WIM station in total weight identification. Figure 6.6 (d) shows a distribution of identified truck speed with mean speed of 66 mph.

Among the 34 trucks, 45 single axles and 59 tandem/axle groups were identified. The identified axle weight information is shown in Figure 6.7. The type II WIM station has $\pm30\%$ and $\pm20\%$

tolerance for 95% probability of conformity for single axle and axle group weight. Therefore, the reference difference distributions are plotted in Figure 6.7 (a) and (b) using $N(0,0.15)$ and $N(0,0.10)$, respectively. As shown in Figure 6.7, the performance of the axle weight identification using the data-driven regression analysis is as good as a type II WIM station for the both single axle and axle group. The one outlier in the axle group weight identification is from a light load axle group of 13.7 kips, well below the mean axle group weight of 21.0 kips. Low axle weight results in lower signal-to-noise ratio and dynamic-to-static ratio in the strain response, therefore reducing the accuracy of the regression analysis.
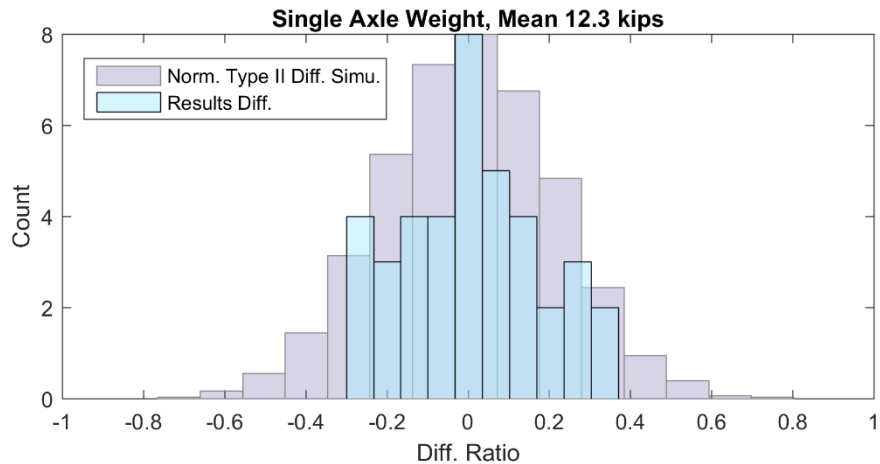
(a)
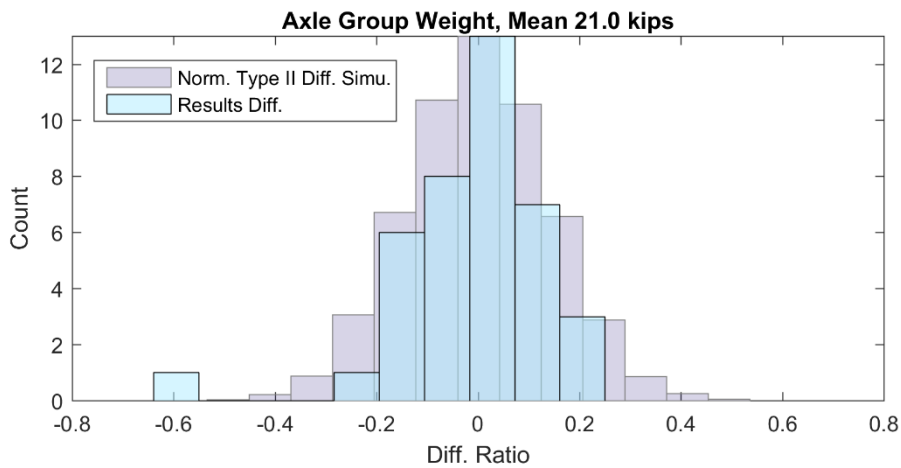
(b)

Figure 6.7, Axle weight difference between regression analysis and WIM station records: (a) single axle weight and (b) axle group weight

Axle spacing results are shown by histogram in Figure 6.8 as the difference between the identified spacing and the WIM station records. The accuracy of the axle spacing is reasonable but not as good as a WIM station, given the WIM station has an accuracy of 0.5 ft. Signal noise, dynamic impact and speed change are the three main causes for the estimation error. The axle spacing detection is based on the assumption that the vehicle is driving at a constant speed across the bridge. However, based on observation, some of the truck drivers tend to slow down when approaching the road side experiment setup, including a video camera and two researchers, and the drivers tend to speedup once pass the cameras. This change of speed could introduce errors in axle spacing estimation.

(a)

(b)

(c)

(d)
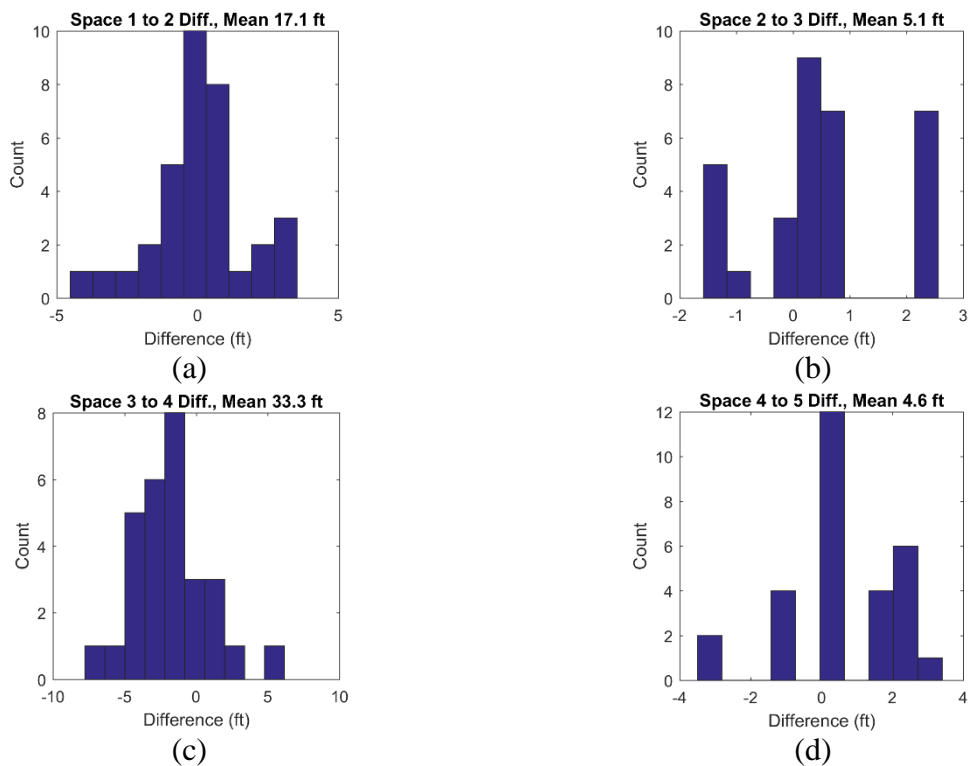
Figure 6.8, Axle spacing difference between regression analysis and WIM station records

6.7 Truck Weight Identification using Long-term Data

To evaluate the long-term traffic conditions of the TRB, the proposed data-driven bridge WIM approach is expanded to identify the gross vehicle weight for all truck events captured from May 2015 to April 2016. Data acquisition is performed every two hours for one minute due to power constraints; as a result, the data is only a sample representation of bridge traffic. Peak picking and threshold algorithms are used to first identify truck events in the strain response, and the bridge WIM algorithm is applied to extract the gross truck weight. In total, 1,968 trucks were identified using the 1-year data. Figure 6.9 (a) shows the distribution of gross vehicle weight. 92% of the trucks are under the legal weight 80 kips. Among these trucks, three groups are identified which corresponds to the three types of typical Michigan trucks: light weight medium, medium and standard semi-trailer, according to an MDOT report (MDOT 2012). Over 1,000,000 trucks traveling north were recorded at the WIM station from May 2015 to April 2016. Figure 6.9 (b) shows the distribution of gross vehicle weight. Without permanent video data at each site, it is difficult to produce a one-to-one mapping between the trucks passing TRB and the trucks passing

Figure 6.9, Histogram of gross vehicle weight: (a) truck identified at the TRB bridge WIM and (b) truck records at the WIM station

the WIM station. However, by observing histograms of large data sets from each location, very similar statistics can be seen in the distribution shap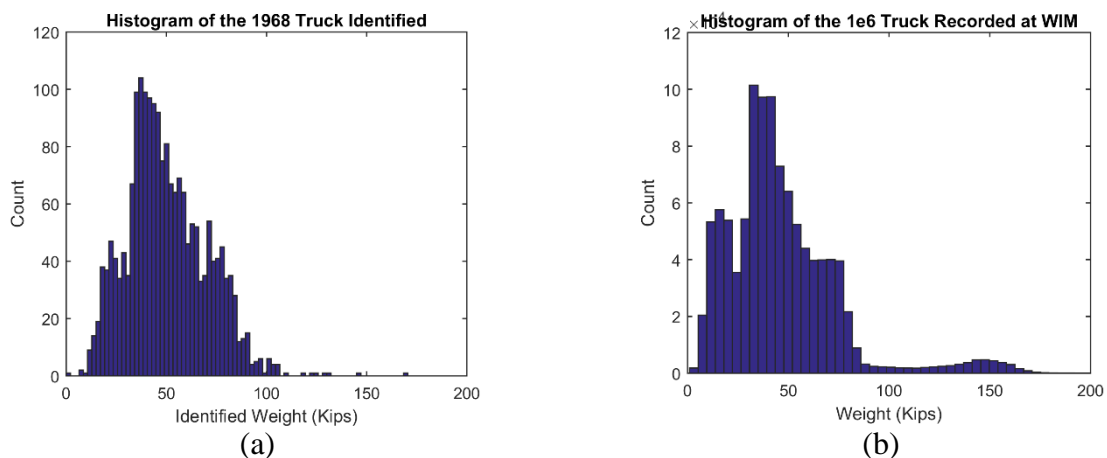es, providing evidence of the reliability of the proposed data-driven bridge WIM system for long-term implementation and analysis. The representation of the traffic (*i.e.*, gross vehicle weight) collected using the ultra-low cost bridge WIM system can be further used for long-term load estimation and traffic analysis.

6.8 Chapter Summary

This chapter described a data-driven regression algorithm for bridge weigh-in-motion system using a single strain gage sensor. The algorithm constructs a ℓ_0 linear regression problem of the influence line Toeplitz matrix and the strain response to extract gross weight, number of axles, speed, axle weight and spacing. The axle influence line used in the analysis was extracted using experimental sensing data from a calibration truck. The Least Absolute Shrinkage and Selection Operator (LASSO) solver was first applied to estimate the solution of the ℓ_0 regression problem. The estimated solution provides information on possible axle locations and truck speed. A list of all reasonable truck models are constructed based on the LASSO solution and the best subset selection is applied to evaluate all models using Moses's theory. The best fitting model and its solution provide details on the truck gross weight, number of axles, axle weight and spacing. The testbed used to validate the algorithm and evaluate the accuracy is the Telegraph Road Bridge, a pin-and-hanger highway bridge in Monroe, Michigan. The bridge carries northbound traffic and has three traffic lanes. A strain gage installed on the hanger plate of the middle girder is used as the bridge WIM sensor. Truck data recorded at an existing MDOT owned WIM station 12 miles downstream on the I-275 highway was used for reference data. Video data captured at both the TRB site and the WIM station was used to generate a one-to-one mapping between trucks passing

both sites. Based on the results obtained from 34 truck events satisfying the identification conditions, the bridge WIM system using the data-driven regression analysis algorithm achieves a type II WIM accuracy. The algorithm has also been applied to long-term data from May 2015 to April 2016 where nearly 2000 trucks were identified. Long-term data histograms of the bridge WIM and the MDOT owned WIM have very similar shapes, proving that the ultra-low cost bridge WIM system is capable of providing an accurate representation of long-term traffic conditions for local traffic analysis.

CHAPTER 7

CONCLUSIONS

7.1 Summary of Results and Contributions

The main focus of this thesis is to advance structural condition monitoring by taking advantage of big data opportunities created by long-term structural monitoring systems. The low cost of sensing technology encourages deployment of dense sensor arrays in structural monitoring systems, which generate a large volume of sensing data as well as information (metadata) of sensing activities. Data size can grow into the GB and TB scale, especially for long-term monitoring systems with high-frequency sampling (*e.g.*, collecting acceleration data for modal analysis). Large volume sensing data creates opportunities for researchers to apply data-driven analytics, extracting physical system (structure) information based on statistical modeling. Additionally, long-term monitoring adds a level of confidence in modeling results since a full range of structural observations and environmental conditions are experienced over an extended period of time. With big data opportunities also comes challenges of data management such as date curation, storage, distribution processing of data in an effective and efficient manner.

This thesis proposes and validates solutions for managing the challenges associated with long-term structural condition monitoring. The solution presented is two-fold: 1) a customized data

management system for structural condition monitoring purposes, and 2) a series of data-driven analytic tools to improve and automate existing algorithms in order to extract new information, about structural performance while improving modeling accuracy. The research results and contributions of each chapter are presented as follows:

- In Chapter 2, two long-term wireless monitoring systems, on the New Carquinez Bridge (NCB) in Vallejo, CA and the Telegraph Road Bridge (TRB) in Monroe, MI, were presented to showcase the current state-of-the-art in long-term wireless monitoring systems for operational structures. Evaluation of the systems' performance over six-years of monitoring activities validated the long-term reliability of the wireless monitoring systems and their readiness for application to large-scale civil infrastructure. A description of lessons learned over a three phase systme installation on the NCB contributes valuable information to the wireless monitoring community in terms of harsh environment field deployment. First, the system design of the wireless monitoring system on the NCB was customized for the owner's decision making needs: assistance in post-earthquake decision making. As the current decision making process employed by Caltrans on the NCB relies on non-linear time history finite element analysis, the majority of sensing channels were assigned to collect bridge acceleration responses for modal analysis and finite element model updating purposes. To further study the environmental impact on the bridge properties, weather stations were implemented to collect ambient environment conditions. Sensor locations were carefully chosen to offer easy accessibility for installation and maintenance without compromising sensing accuracy.

- In Chapter 3, a cyberinfrastructure *SenStore* was presented as a solution to data management of long-term structural monitoring data. The top-down database ontology design for customized structural condition monitoring applications is a key feature. Specifically, the needs of system end-users (*e.g.,* structural owner, stake holders) dictate the design of the database ontology. *SenStore* is designed to host structural metadata, inspection data, sensing metadata as well as sensing data itself; *SenStore* also offers compatibility to other existing databases such as AASHTOWare, a bridge inventory and inspection software used by the Departments of Transportation in multiple states. To curate and store the relational metadata and large volume sensing data, *SenStore* adopts a hybrid databases architecture using PostgreSQL (for relational data) and HDF5 (for sensor data). Segregating vectorized sensor data from the relational database maintain the data relationships without losing query speed when integrating data. The system scalability and efficiency was evaluated using metadata and 18GB of sensing data from the long-term wireless monitoring system on the TRB. Despite the constant overhead (about 1 second) introduced by the database system, the sublinear query speed of *SenStore* outperformed the relational database-only solutions, highlighting the efficiency of the hybrid database design. The cloud-based *SenStore* server provides distributed clients with secure access to data via an application programming interface (API) in multiple programming languages. Therefore, analytical clients can use *SenStore* as a cloud platform to perform and data mining, data visualization.

- In Chapter 4, a case study of an automated finite element model updating tool was presented to demonstrate the capability of *SenStore* for complex structural analysis.

135

SenStore served as the data management platform upon which four distributed clients, (serving different functionalities of the total analysis) were implemented in an autonomous manner. The automated FEM model generation client created a FE model using the structural metadata stored in *SenStore*; the automated modal parameter extraction client performed modal analysis from the long-term sensing data also stored in *SenStore*; the automated environment impact modeling client used machine learning techniques to develop and remove the non-linear relationships between modal and environmental parameters; and, the automated FE model updating tool incorporated the temperature impact on the modal parameters and updated a finite element model of the structure based on Bayesian statistical modeling. This is just one example that demonstrates the capability and efficiency of *SenStore* to serve as a data center and analytics platform for advanced structural condition management. This implementation utilized the TRB and its long-term data to validate the *SenStore*-based tool chain.

- Chapter 5 presented two novel data-driven analytical tools which automate the modal analysis of structures and models the environmental and operational conditions (EOC) impact on modal characteristics using long-term monitoring data. Improvements made on the output-only time series algorithm, stochastic subspace identification (SSI), contributes to the modal analysis community in the following four areas. First, it automates the SSI algorithm to extract only physical modes (from mathematical modes). Mathematical modes are specifically rejected by applying rules on the singular values from the output projections. Second, the singular value decomposition process was performed on the product of the projection matrix and its transpose (as apposed on the projection

matrix itself) in order to improve the computational time by a factor of 20. Third, the implementation removes bias in terms of model order selection by controlling the data size for modal estimation. Finally, the implementation provides confidence information of each mode by investigating the distribution and identification success rate of the SSI results using two years' worth of data. The second part of the chapter focuses on the results and contributions from linear and non-linear modeling of EOC impact on modal frequencies of the NCB given wind, temperature and time of the day information. The selected regression frameworks quantitatively identified the EOC impact on the bridge model frequencies. The extracted model parameters (*i.e.*, modal frequencies, damping ratios and mode shapes) contribute to the bridge design community as valuable experimental reference data for future suspension bridge designs. The quantitative EOC modeling of modal frequencies contributes a rare and complete EOC impact study for large-scale suspension bridges. In addition to the direct contributions to modal analysis and the bridge design communities, these two data-driven analytic tools also showcase the opportunities to incorporate big data techniques on long-term monitoring data in order to extract new information and behavioral insight about bridges.

- Chapter 6 presented a novel bridge weigh-in-motion (BWIM) system using a low cost wireless monitoring system and a series of data-driven algorithms. Based on the existing TRB long-term wireless monitoring system, the proposed BWIM system only required one additional strain gage sensor installed at a strategically selected location to capture the structural strain response due to a moving truck load. The algorithm applies a series of regression analyses on strain data to extract the truck configuration by solving a ℓ_0

regression problem. The 3-step algorithm first extracts the experimental influence line using a least square regression. Second, a ℓ_1 regression solver, Least Absolut Shrinkage and Selection Operator (Lasso), is applied to estimate the solution. Finally, the best subset method is used to solve a ℓ_0 regression problem, estimating vehicle gross weight, axle weight, axle spacing and speed. The proposed BWIM system removes the need for standard-practice axle detectors, which are fragile when installed on the road surface or too sensitive to truck path when installed underneath the bridge. Moreover, the algorithm does not require physical modeling of the bridge (*e.g.*, finite element model), which is one of the major sources of computational error in existing BWIM systems. Based on the results of 34 testing trucks, the BWIM strain gage sensor together with the algorithm developed for truck identification added full weigh-in-motion functionality to the existing bridge wireless monitoring. The accuracy of the system was validated to be a type II weigh-in-motion on gross vehicle weight, axle weight and axle group weight identification. The long-term application of such a system is also proven promising. This novel BWIM introduced a new direction to the BWIM community for truck identification algorithms based on data-driven analytics.

7.2 Future Research Recommendations

The potential of data-driven analytics using long-term monitoring data for structural condition monitoring is growing. The increasing amount of commodity sensing units available on the market ease the installation and maintenance of the sensing system. Therefore, the research direction will be focused on the advancement of cloud-based data management systems and improved data-driven analytics. Several future research directions are identified:

138

- Fully utilize cloud services to improve the system availability, scalability and computational power of the data management system, as well as reduce the initial investment and operational cost. If using public cloud services, data security must be a primary concern.

- Expand the bridge weigh-in-motion system to comprehend multi-truck events and trucks on different lanes using more sensors. One sensor each lane should suffice, and by solving the load distribution and using superposition, an optimization algorithm could extract the response data for each truck event.

- Moving load identification algorithms using the dynamics of the bridge may be adopted by the bridge weigh-in-motion system to improve the accuracy. System matrices can be determined using the free vibration data following a truck event. The input matrix can be estimated using sampling and optimization techniques.

- Apply machine learning and statistical modeling techniques to structural damage detection applications. With the implementation of dense sensing arrays, the characteristics of each sensing channel as well as the correlations between sensing channels can be translated into feature vectors or an index for damage detection to improve the success rate and accuracy.

REFERENCES

Allemang, R. J. (2003). "The modal assurance criterion –twenty years of use and abuse." *Journal of Sound and Vibration*, 1(August), 14–21.

Allen, D. W., Clough, J. A., Sohn, H., and Farrar, C. R. (2003). "A software tool for graphically assembling damage identification algorithms." *Proceedings of SPIE - The International Society for Optical Engineering*, San Diego, CA, 138–144.

Andrieu, C., and Thoms, J. (2008). "A tutorial on adaptive MCMC." *Statistics and Computing*, 18(4), 343–373.

ASCE. (2013). "Minimum design loads for buildings and other structures." American Society of Civil Engineers, Reston, VA.

ASTM. (2009). "Standard specification for highway weigh-in-motion (WIM) systems with user requirements and test methods." American Society for Testing and Materials, West Conshohocken, PA.

Autodesk. (2002). "White paper: building information modeling." Autodesk Inc., San Rafael, CA.

Bardet, J.-P., Law, K. H., Peng, J., and Swift, J. (2004). "Overview of NEES data and metadata models." *13th World Conference in Earthquake Engineering*, Vancouver, B.C., Canada, paper 4001.

Beck, J. L., and Katafygiotis, L. S. (1998). "Updating models and their uncertainties. I: Bayesian statistical framework." *Journal of Engineering Mechanics*, 124(4), 463–467.

Behmanesh, I., and Moaveni, B. (2015). "Probabilistic identification of simulated damage on the Dowling Hall footbridge through Bayesian finite element model updating." *Structural Control and Health Monitoring*, 22, 463–483.

BORP. (2011). "Building occupancy resumption program." Department of Building Inspection Emergency Operations Plan, San Francisco, CA.

Brincker, R., Andersen, P., and Jacobsen, N.-J. (2007). "Automated frequency domain decomposition for operational modal analysis." *Conference: 2007 IMAC-XXV: Conference & Exposition on Structural Dynamics*, Orlando, FL.

Brincker, R., Zhang, L., and Andersen, P. (2001). "Modal identification of output-only systems using frequency domain." *Smart Materials and Structures*, 10, 441–445.

Brownjohn, J. M. W. (2003). "Ambient vibration studies for system identification of tall buildings." *Earthquake Engineering and Structural Dynamicsngineering & Structural Dynamics*, 32(1), 71–95.

Brownjohn, J. M. W. (2007). "Structural health monitoring of civil infrastructure." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 589–622.

Brownjohn, J. M. W., Magalhaes, F., Caetano, E., and Cunha, A. (2010). "Ambient vibration re-testing and operational modal analysis of the Humber Bridge." *Engineering Structures*, Elsevier Ltd, 32(8), 2003–2018.

Brownjohn, J. M. W., Moyo, P., Omenzetter, P., and Lu, Y. (2003). "Assessment of highway bridge upgrading by dynamic testing and finite-element model updating." *Journal of Bridge Engineering*, 8(3), 162–172.

Brownjohn, J. M. W., and Xia, P.-Q. (2000). "Dynamic assessment of curved cable-stayed bridge by model updating." *Journal of Structural Engineering*, 126(2), 252–260.

Buetow, K. H. (2005). "Cyberinfrastructure: empowering a' third way' in biomedical research." *Science*, 308(5723), 821–4.

Çatbaş, F. N., Kijewski-Correa, T., and Aktan, A. E. (2013). *Structural identification of constructed systems.* American Society of Civil Engineers, Reston, VA.

Çatbaş, F. N., Susoy, M., and Frangopol, D. M. (2008). "Structural health monitoring and reliability estimation: long span truss bridge application with environmental monitoring data." *Engineering Structures*, 30(9), 2347–2359.

Çelebi, M. (2006). "Real-time seismic monitoring of the new Cape Girardeau Bridge and preliminary analyses of recorded data: an overview." *Earthquake Spectra*, 22(3), 609–630.

Çelebi, M., Sanli, A., Sinclair, M., Gallant, S., and Radulescu, D. (2004). "Real-time seismic monitoring needs of a building owner-and the solution: a cooperative effort." *Earthquake Spectra*, 20(2), 333–346.

Chang, C., and Lin, C. (2011). "LIBSVM: a library for support vector machines." *ACM Transactions on Intelligent Systems and Technologies*, 2(3), 1–27.

Conte, J. P., He, X., Moaveni, B., Masri, S. F., Caffrey, J. P., Wahbeh, M., Tasbihgoo, F., Whang, D. H., and Elgamal, A. (2008). "Dynamic testing of Alfred Zampa Memorial Bridge." *Journal of Structural Engineering*, 124(6), 1006–1015.

Cornwell, P., Farrar, C. R., Doebling, S. W., and Sohn, H. (1999). "Environmental variability of modal properties." *Experimental Techniques*, 23(6), 45–48.

Cross, E. J., Koo, K. Y., Brownjohn, J. M. W., and Worden, K. (2012). "Long-term monitoring and data analysis of the Tamar Bridge." *Mechanical Systems and Signal Processing*, Elsevier, 35(1-2), 16–34.

Cunha, Á., and Caetano, E. (2006). "Experimental modal analysis of civil engineering structures." *Sound and Vibration*, 40(6), 12–20.

Ditlevsen, O., and Madsen, H. O. (2007). *Structural reliability methods*. Department of Mechanical Engineering Technical University of Denmark.

Doebling, S. W., Farrar, C. R., and Prime, M. B. (1998). "A summary review of vibration-based

damage identification methods." *The Shock and Vibration Digest*, 30(2), 91–105.

Doebling, S. W., Farrar, C. R., Prime, M. B., and Shevitz, D. W. (1996). *Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review*. No. LA-13070-MS, Los Alamos National Lab, NM.

DOT. (2014). "Track safety standards; improving rail integrity." Department of Transportation, Washington, DC.

Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2011). *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers, and contractors*. John Wiley & Sons, Hoboken, New Jersey.

Ewins, D. J. (2000). *Modal testing, theory, practice, and application (mechanical engineering research studies: engineering dynamics series)*. Research Studies Press Ltd, Baldock, Hertfordshire, England.

Farhat, C., and Hemez, F. M. (1993). "Updating finite element dynamic models using an element-by-element sensitivity methodology." *AIAA Journal*, 31(9), 1702–1711.

Farrar, C. R., and Worden, K. (2007). "An introduction to structural health monitoring." *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 365, 303–315.

Farrar, C. R., and Worden, K. (2012). *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, Chichester, UK.

Feltrin, G., Meyer, J., Bischoff, R., and Motavalli, M. (2010). "Long-term monitoring of cable stays with a wireless sensor network." *Structure and Infrastructure Engineering*, 6(5), 535–548.

FHWA. (2004). *National bridge inspection standards*. Federal Highway Administration, Washington, DC.

FHWA. (2013). "Traffic Monitoring Guide." Federal Highway Administration, Washington, DC.

Fraser, M., Elgamal, A., He, X., and Conte, J. P. (2010). "Sensor network for structural health monitoring of a highway bridge." *Journal of Computing in Civil Engineering*, 24(1), 11–24.

Friswell, M., and Mottershead, J. (1995). *Finite element model updating in structural dynamics*. Springer Netherlands.

Fujino, Y., Kimura, K., and Tanaka, H. (2012). "Wind resistant design codes for bridges in Japan." *Wind Resistant Design of Bridges in Japan*, Springer Japan, 1–7.

Fujino, Y., Murata, M., Okano, S., and Takeguchi, M. (2000). "Monitoring system of the Akashi Kaikyo Bridge and displacement measurement using GPS." *the International Society for Optical Engineering (SPIE)*, San Diego, CA, 229–236.

Gonzalez, A., Rowley, C., and O'Brien, E. J. (2008). "A general solution to the identification of moving vehicle forces on a bridge." *International Journal for Numerical Methods in Engineering*, 75(3), 335–354.

Goodall, J., Robinson, B., and Castronova, A. (2011). "Modeling water resource systems using a service-oriented computing paradigm." *Environmental Modelling & Software*, Elsevier Ltd, 26, 573–582.

Häckell, M. W., and Rolfes, R. (2013). "Monitoring a 5MW offshore wind energy converter— condition parameters and triangulation based extraction of modal parameters." *Mechanical Systems and Signal Processing*, 40(1), 322–343.

Hajian, H., and Becerik-Gerber, B. (2009). "A research outlook for real-time project information management by integrating advanced field data acquisition systems and building information modeling." *International Workshop on Computing in Civil Engineering*, Austin, Texas, 83–94.

Hallenbeck, M., and Weinblatt, H. (2004). *Equipment for collecting traffic load data*. National Cooperative Highway Research Program - Report 509, Transportation Research Board, Washington, DC.

He, X., Moaveni, B., Conte, J. P., Elgamal, A., and Masri, S. F. (2009). "System identification of Alfred Zampa Memorial Bridge ysing dynamic field test data." *Journal of Structural Engineering*, 135(1), 54–66.

Herman, G. A., Peterson, J. C., and Trotta, B. W. (2012). "Bridge information modeling." United States Patents, US8160841 B2, United States.

Heylen, W., Lammens, S., and Sas, P. (1995). *Modal analysis theory and testing*. Department of Mechanical Engineering, Katholieke Universiteit Leuven, Belgium.

Hitchcock, W. A., Uddin, N., Sisiopiku, V., Salama, T., Kirby, J., Zhao, H., Toutanji, H., and Richardson, J. (2012). *Bridge weigh-in-motion (B-WIM) system testing and evaluation*. University Transportation Center for Alabama, Tuscaloosa, Alabama.

Hoerl, A. E., and Kennard, R. W. (1970). "Ridge regression: biased estimation for nonorthogonal problems." *Technometrics*, 12(1), 55–67.

Hong, A. L., Ubertini, F., and Betti, R. (2011). "Wind analysis of a suspension bridge: identification and finite-element model simulation." *Journal of Structural Engineering*, 137(1), 133–142.

Hua, X. G., Ni, Y. Q., Ko, J. M., and Wong, K. Y. (2007). "Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique." *Journal of Computing in Civil Engineering*, 21(2), 122–135.

Ibrahim, S. R. (1977). "Random decrement technique for modal identification of structures." *Journal of Spacecraft and Rockets*, 14(11), 696–700.

Jacob, B., Cebon, D., O'Brien, E. J., Henny, R., Grandpre, D., Hallstrom, B., Jehaes, S., Baumgartner, W., Huhtala, M., Znidaric, A., and Caprez, M. (2001). *Weigh-in-motion of axles and vehicles for Europe (WAVE)*. Laboratoire Central des Ponts et Chaussees, Paris, France.

Jacob, B., and O'Brien, E. J. (1998). "European specification on weigh-in-motion of road vehicles (COST323)." *Second European Conference on Weigh-in-mition of Road Vehicles*,

Held Lisbon, Portugal, 171–183.

Jang, S., Jo, H., Cho, S., Mechitov, K., Rice, J. A., Sim, S., Jung, H., Yun, C., Spencer, B. F., and Agha, G. (2010). "Structural health monitoring of a cable-stayed bridge using smart sensor technology: deployment and evaluation." *Smart Structures and Systems*, 6(5-6), 439–459.

Jolliffe, I. (2005). *Principal component analysis*. Wiley StatsRef: Statistics Reference Online.

Kijewski-Correa, T., Kwon, D. K., Kareem, A., Bentz, A., Guo, Y., Bobby, S., and Abdelrazaq, A. (2013). "Smartsync: An integrated real-time structural health monitoring and structural identification system for tall buildings." *Journal of Structural Engineering*, 139(10), 1675–1687.

Kijewski-Correa, T., and Pirnia, J. D. (2007). "Dynamic behavior of tall buildings under wind: Insights from full-scale monitoring." *Structural Design of Tall and Special Buildings*, 16(4), 471–486.

Kim, J., and Lynch, J. P. (2012). "Autonomous decentralized system identification by Markov parameter estimation using distributed smart wireless sensor networks." *Journal of Engineering Mechanics*, 138(5), 478–490.

Koo, K. Y., de Battista, N., and Brownjohn, J. M. W. (2011). "SHM data management system using MySQL database with Matlab and web Interfaces." *5th International Conference on Structural Health Monitoring of Intelligent Infrastructure*, Cancun, Mexico, Mexico.

Kuhn, H. W., and Tucker, A. W. (1951). "Nonlinear programming." *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California, 481–492.

Kurata, M., Kim, J., Lynch, J. P., van der Linden, G. W., Sedarat, H., Thometz, E., Hipley, P., and Sheng, L.-H. (2013). "Internet-enabled wireless structural monitoring systems: development and permanent deployment at the New Carquinez Suspension Bridge." *Journal of Structural Engineering*, 139(10), 1688–1702.

Kurata, M., Lynch, J. P., van der Linden, G., Jacob, V., and Hipley, P. (2010). "Preliminary study of a wireless structural monitoring system for the New Carquinez Suspension Bridge." *5th World Conference of Structural Control and Monitoring*, Tokyo, Japan, 1–14.

Law, S. S., Chan, T. H. T., and Zeng, Q. H. (1997). "Moving force identification: a time domain method." *Journal of Sound and Vibration*, 201(1), 1–22.

Law, S. S., Chan, T. H. T., and Zeng, Q. H. (1999). "Moving force identification - a frequency and time domains analysis." *Journal of Dynamic Systems, Measurement, and Control*, 121(3), 394–401.

Levin, R. I., and Lieven, N. A. J. (1998). "Dynamic finite element model updating using simulated annealing and genetic algorithms." *Mechanical Systems and Signal Processing*, 12(1), 91–120.

Li, H., Li, S., Ou, J., and Li, H. (2010). "Modal identification of bridges under varying environmental conditions: Temperature and wind effects." *Structural Control and Health*

Monitoring, 17, 495–512.

van der Linden, G., Emami-Naeini, A., Zhang, Y., and Lynch, J. P. (2013). "Cyber-infrastructure design and implementation for structural health monitoring." *Proceedings of SPIE - The International Society for Optical Engineering*, T. Y. Yu, A. L. Gyekenyesi, P. J. Shull, A. A. Diaz, H. F. Wu, and A. E. Aktan, eds., San Diego, CA, 1–9.

Liu, X., and Akinci, B. (2009). "Requirements and evaluation of standards for integration of sensor data with building information models." *International Workshop on Computing in Civil Engineering*, Austin, Texas, 95–104.

Ljung, L. (1998). "System identification." *Signal Analysis and Prediction*, Birkhäuser Boston, 163–173.

Loh, C.-H., and Chen, M. (2013). "Modeling of environmental effects for vibration-based SHM using recursive stochastic subspace identification analysis." *Key Engineering Materials*, 558, 52–64.

Lynch, J., Kamat, V., Li, V. C., Flynn, M., Sylvester, D., Najafi, K., Gorden, T., Lepech, M., Emami-Naeini, A., Krimotat, A., Ettouney, M., Alampalli, S., and Ozdemir, T. (2009). "Overview of a cyber-enabled wireless monitoring system for the protection and management of critical infrastructure systems." *Proceedings of SPIE - The International Society for Optical Engineering*, H. F. Wu, A. A. Diaz, P. J. Shull, and D. W. Vogel, eds., San Diego, CA, 72940L–72940L.

Lynch, J. P. (2002). "Decentralization of wireless monitoring and control technologies for smart civil structures." Ph.D. Dissertation, Stanford University.

Lynch, J. P. (2007). "An overview of wireless structural health monitoring for civil structures." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 345–372.

Lynch, J. P., Wang, Y., Loh, K. J., Yi, J.-H., and Yun, C.-B. (2006). "Performance monitoring of the Geumdang Bridge using a dense network of high-resolution wireless sensors." *Smart Materials and Structures*, 15(6), 1561–1575.

Magalhães, F., Amador, S., Cunha, Á., and Caetano, E. (2012). "Dynamo–software for vibration based structural health monitoring." *International Association for Bridge Maintenance and Safety (IABMAS)*, Stresa, Lake Maggiore, Italy.

Magalhães, F., Cunha, Á., and Caetano, E. (2009). "Online automatic identification of the modal parameters of a long span arch bridge." *Mechanical Systems and Signal Processing*, 23(2), 316–329.

MDOT. (2012). *Michigan's truck-weight law and truck-user fees*. Michigan Department of Transportation, Lansing, MI.

Meng, F., Ozbek, M., Rixen, D. J., and van Tooren, M. J. L. (2011). "Comparison of system identaification techniques for predicting dynamic properties of large scale wind turbine by using the simulated time response." *Structural Dynamics and Renewable Energy, Volume 1*, Springer New York, 339–349.

Michael, B., Mertz, D. R., and Commander, B. (1997). "Experimental load rating of a posted bridge." *Journal of Bridge Engineering*, 2(1), 1–10.

Moaveni, B., and Behmanesh, I. (2012). "Effects of changing ambient temperature on finite element model updating of the Dowling Hall Footbridge." *Engineering Structures*, 43, 58–68.

Mosavi, A. A., Sedarat, H., O'Connor, S. M., Emami-Naeini, A., and Lynch, J. P. (2014). "Calibrating a high-fidelity finite element model of a highway bridge using a multi-variable sensitivity-based optimisation approach." *Structure and Infrastructure Engineering*, 10(5), 627–642.

Moser, P., and Moaveni, B. (2011). "Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge." *Mechanical Systems and Signal Processing*, 25(7), 2336–2357.

Moses, F. (1979). "Weigh-in-motion system using instrumented bridges." *Transportation Engineering Journal of ASCE*, 105(TE3), 233–249.

Mottershead, J. E., and Friswell, M. (1993). "Model updating in structural dynamics: a survey." *Journal of Sound and Vibration*, 167(2), 347–375.

Muste, M., Bennett, D., Secchi, S., Schnoor, J. L., Kusiak, A., Arnold, N. J., Mishra, S. K., Ding, D., and Rapolu, U. (2012). "End-to-end cyberinfrastructure for decision-making support in watershed management." *Journal of Water Resources Planning and Management*, 139(5), 565–573.

Nader, W., Neal, J., Shi, J., Berryman, C., Cho, Y., Siu-Kit, L., Li, H., Schwer, A., Shen, Z., Stansbury, J., and Zhang, T. (2010). "Real time power monitoring & integration with BIM." *2010-36th Annual Conference on IEEE Industrial Electronics Society (IECON)*, Glendale, AZ, 2454–2458.

Nagayama, T., and Spencer, B. F. (2007). "Structural health monitoring using smart sensors." Newmark Structural Engineering Laboratory. University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.

Nakamura, M., Masri, S. F., Chassiakos, A. G., and Caughey, T. K. (1998). "A neural network approach to damage detection in a building from ambient vibration measurements." *SPIE 3321, 1996 Symposium on Smart Materials, Structures, and MEMS, 126*, Bangalore, India.

Nayeri, R., Masri, S., Ghanem, R. G., and Nigbor, R. L. (2008). "A novel approach for the structural identification and monitoring of a full-scale 17-story building based on ambient vibration measurements." *Smart Materials and Structures*, 17(2), 025006.

Ni, Y. Q., Hua, X. G., Fan, K. Q., and Ko, J. M. (2005). "Correlating modal properties with temperature using long-term monitoring data and support vector machine technique." *Engineering Structures*, 27(12), 1762–1773.

Ni, Y. Q., Xia, Y., Liao, W. Y., and Ko, J. M. (2009). "Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower." *Structural Control and Health Monitoring*, 16(1), 73–98.

Ni, Y., and Wong, K. (2012). "Integrating bridge structural health monitoring and condition-based maintenance management." *Civil Structural Health Monitoring Workshop*, Berlin, Germany.

O'Brien, E. J., Quilligan, M. J., and Karoumi, R. (2006). "Calculating an IL from direct measurements." *Proceedings of the Institution of Civil Engineers, Bridge Engineering*, Institution of Civil Engineers, 159, 31–34.

O'Connor, S. M., Lynch, J. P., Ettouney, M., Linden, G. Van Der, and Alampalli, S. (2012). "Cyber-enabled decision making system for bridge management using wireless monitoring systems : Telegraph Road Bridge demonstration project." *Proceedings of American Society for Nondestructive Testing- NDE/NDT for Highways and Bridges*, New York, NY.

O'Connor, S. M., Zhang, Y., Lynch, J. P., Ettouney, M., and Jansson, P. (2016). "Long-term performance assessment of the Telegraph Road Bridge using a permanent wireless monitoring system and automated statistical process control analytics." *Structure and Infrastructure Engineering*, in press.

Ojio, T., Carey, C. H., O'Brien, E. J., Doherty, C., and Taylor, S. E. (2016). "Contactless bridge weigh-in-motion." *Journal of Bridge Engineering*, in press.

OPAC. (2016). "Third Carquinez Strait Bridge." <http://www.opacengineers.com/features/carquinez> (May 27, 2016).

van Overschee, P., and de Moor, B. (2012). *Subspace identification for linear systems: Theory—Implementation—Applications*. Springer US.

Pakzad, S. N., Fenves, G. L., Kim, S., and Culler, D. E. (2008). "Design and implementation of scalable wireless Sensor network for structural monitoring." *Journal of Infrastructure Systems*, 14(1), 89–101.

Peeters, B., and De Roeck, G. (1999). "Reference-based stochastic subspace identification for output-only modal analysis." *Mechanical Systems and Signal Processing*, 13(6), 855–878.

Peeters, B., and De Roeck, G. (2001a). "One year monitoring of the Z24 bridge: environmental influences versus damage events." *Earthquake Engineering and Structural Dynamics*, 30(2), 149–171.

Peeters, B., and De Roeck, G. (2001b). "Stochastic system identification for operational modal analysis: a review." *Journal of Dynamic Systems, Measurement, and Control*, 123(4), 659–667.

Peeters, B., De Roeck, G., Pollet, T., and Schueremans, L. (1997). "Stochastic subspace techniques applied to parameter identification of civil engineering structures." *New Advances in Modal Synthesis of Large Structures*, A. A. Balkema, Rotterda, Netherlands, 145–156.

Pei, J., and Smyth, A. W. (2006). "New approach to designing multilayer feedforward neural network architecture for modeling nonlinear restoring forces. I: formulation." *Journal of Engineering Mechanics*, 132(12), 1290–1300.

Perera, R., and Ruiz, A. (2008). "A multistage FE updating procedure for damage identification

in large-scale structures based on multiobjective evolutionary optimization." *Mechanical Systems and Signal Processing*, 22(4), 970–991.

Pridham, B. A., and Wilson, J. C. (2003). "A study of damping errors in correlation-driven stochastic realizations using short data sets." *Probabilistic Engineering Mechanics*, 18(1), 61–77.

Quilligan, M., Karoumi, R., and O'Brien, E. J. (2002). "Development and testing of a 2-dimensional multi-vehicle bridge-WIM algorithm." *3rd International Conference on Weigh-in-Motion (ICWIM3)*, Orlando, FL, FL, 199–208.

Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press, Boston, MA.

Richardson, M. H., and Formenti, D. L. (1982). "Parameter estimation from frequency response measurements using rational fraction polynomials." *Proceedings of the International Modal Analysis Conference*, Orlando, FL, 167–182.

Robert, W., Marshall, A., Shepard, R., and Aldayuz, J. (2003). "The PONTIS bridge management system: state-of-the-practice in implementation and development." *9th International Bridge Management Conference*, Orlando, FL, 29–46.

Rolfes, R., Zerbst, S., Haake, G., Reetz, J., and Lynch, J. P. (2007). "Integral SHM-system for offshore wind turbines using smart wireless sensors." *the 6th International Workshop on Structural Health Monitoring (IWSHM)*, Stanford, CA, 1–8.

Rowley, C., Gonzalez, A., O'Brien, E. J., and Znidaric, A. (2009). "Comparison of conventional and regularized bridge weigh-in-motion algorithms." *International Conference on Heavy Vehicles HVParis 2008*, John Wiley & Sons, Inc, Hoboken, NJ.

Salawu, O. S. (1997). "Detection of structural damage through changes in frequency: a review." *Engineering Structures*, 19(9), 718–723.

Smarsly, K., Law, K., and Hartmann, D. (2012). "Multiagent-based collaborative framework for a self-managing structural health monitoring system." *Journal of Computing in Civil Engineering*, 26, 76–89.

Smola, A., and Schölkopf, B. (2004). "A tutorial on support vector regression." *Statistics and Computing*, Kluwer Academic Publishers, 14(3), 199–222.

Snyder, R. E., and Moses, F. (1985). "Application of in-motion weighing using instrumented bridges." *Transportation Research Record*, 1048, 83–88.

Sohn, H., Dzwonczyk, M., Straser, E. G., Kiremidjian, A. S., Law, K. H., and Meng, T. (1999). "An experimental study of temperature effect on modal parameters of the Alamosa Canyon Bridge." *Earthquake Engineering and Structural Dynamics*, 28(8), 879–897.

Sohn, H., Worden, K., and Farrar, C. R. (2002). "Statistical damage classification under changing environmental and operational conditions." *Journal of Intelligent Material Systems and Structures*, 13(9), 561–574.

Sohn, H., Worden, K., and Farrar, R. C. (2001). "Novelty detection using auto-associative neural

network." *Symposium on Identification of Mechanical Systems: international mechanical engineering congress and exposition*, New York, NY.

de Sortis, A., and Paoliani, P. (2007). "Statistical analysis and structural identification in concrete dam monitoring." *Engineering Structures*, 29(1), 110–120.

Spence, S. and Kareem, A. (2014). "Tall buildings and damping: a concept-based data-driven model." *Journal of Structural Engineering*, 140(5).

Spencer, B. F., Ruiz-Sandoval, M. E., and Kurata, N. (2004). "Smart sensing technology: opportunities and challenges." *Structural Control and Health Monitoring*, 11(4), 349–368.

Straser, E. G., and Kiremidjian, A. S. (1998). "A modular, wireless damage moitoring system for structures." The John A. Blume Earthquake Engineering Center, Stanford University, Stanford, CA.

Swartz, R., Jung, D., Lynch, J. P., Wang, Y., Shi, D., and Flynn, M. P. (2005). "Design of a wireless sensor for scalable distributed in-network computation in a structural health monitoring system." *the 5th International Workshop on Structural Health Monitoring (IWSHM)*, Stanford, CA.

Swartz, R., Lynch, J., Sweetman, B., and Rolfes, R. (2010). "Structural monitoring of wind turbines using wireless sensor networks." *Smart Structures and Systems*, 6(3), 183–196.

Swartz, R., Zimmerman, A., and Lynch, J. P. (2007). "Structural health monitoring system with the latest information technologies." *5th Infrastructure & Environmental Management Symposium*, Yamaguchi, Japan.

Swartz, R., Zimmerman, A., Lynch, J. P., Rosario, J., Brady, T., Salvino, L., and Law, K. H. (2012). "Hybrid wireless hull monitoring system for naval combat vessels." *Structure and Infrastructure Engineering, Taylor & Francis Group*, 8(7), 621–638.

Tibshirani, R. (1996). "Regression shrinkage and selection via the LASSO." *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288.

Tikhonov, A. N., and Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Vh Winston.

Tomko, M., Greenwood, P., Sarwar, M., Morandini, L., Stimson, R., Bayliss, C., Galang, G., Nino-Ruiz, M., Voorsluys, W., Widjaja, I., Koetsier, G., Mannix, D., Pettit, C., and Sinnott, R. (2012). "The design of a flexible web-based analytical platform for urban research." *20th International Conference on Advances in Geographic Information Systems (ACM)*, New York, NY, 369–375.

Trefethen, L. N., and Bau, D. I. (1997). "Computing the SVD." *Numerical linear algebra*, Society for Industrial and Applied Mathematics, Philadelphia, PA.

US Congress. (2006). *Pipeline inspection, protection, enforcement, and safety act of 2006*. Public Law: 109–468, 109th Congress.

Wall, C. J., Christenson, R. E., McDonnel, A. H., and Jamalipour, A. (2009). *A non-intrusive bridge weigh-in-motion system for a single span steel girder bridge using only strain measurements*. Report No. CT-2251- 3-09-5, University of Connecticut, Connecticut

Transportation Institute, Storrs, CT.

Wang, J. Y., Ko, J. M., and Ni, Y. Q. (2000). "Modal sensitivity analysis of Tsing Ma Bridge for structural damage detection." *Proceedings of SPIE - The International Society for Optical Engineering*, San Diego, CA, 1–12.

Whelan, M. J., and Janoyan, K. D. (2008). "Design of a robust, high-rate wireless sensor network for static and dynamic structural monitoring." *Journal of Intelligent Material Systems and Structures*, 20(7), 849–863.

Wong, K. (2004). "Instrumentation and health monitoring of cable-supported bridges." *Structural Control and Health Monitoring*, 11(2), 91–124.

Zárate, B. A., Caicedo, J. M., and Ziehl, P. (2013). "Development and implementation of a cyberinfrastructure framework for research in nondestructive evaluation using acoustic emission data." *Journal of Computing in Civil Engineering*, 28(3), case study.

Zhang, Y., Häckell, M., Lynch, J. P., and Rolfes, R. (2014). "Automated modal parameter extraction and statistical analysis of the New Carquinez Bridge response to ambient excitations." *Model Validation and Uncertainty Quantification*, Springer International Publishing, 161–170.

Zhang, Y., and Lynch, J. P. (2013). "Long-term modal analysis of the New Carquinez long-span suspension bridge." *Topics in Dynamics of Bridges*, Springer New York, 73–82.

Zhou, Z., and Ou, J. (2005). "Development of FBG sensors for structural health monitoring." *Sensing Issues in Civil Structural Health Monitoring*, Springer Netherlands, 197–206.