Statistical Bridge Damage Detection Using Girder Distribution Factors

Alexandra J Reiff¹, Masoud Sanayei², Richard M. Vogel³

¹ Bridge Engineer, Kleinfelder; email: areiff@kleinfelder.com

Former Graduate Research Assistant, Tufts University, Medford MA, 02155

² Professor, Department of Civil & Environmental Engineering, Tufts University, Medford MA, 02155;

Phone: (617) 627-4116; email: masoud.sanayei@tufts.edu

³ Professor, Department of Civil & Environmental Engineering, Tufts University, Medford MA, 02155;

Phone: (617) 627-4260; email: richard.vogel@tufts.edu

Abstract

A hypothesis testing framework is introduced for bridge damage detection, which enables a rigorous, decision-oriented approach for detection of bridge damage when it exists. A bridge damage detection hypothesis test is developed using girder distribution factors (GDF) under operational, output-only strain monitoring. GDFs are calculated from measured strain data collected during traffic events at the Powder Mill Bridge in Barre, Massachusetts. A sample of GDFs is drawn to establish a baseline over the course of one week, representing the probabilistic behavior of a healthy bridge under normal operating conditions. A new sample can be compared with the baseline at the end of each day, providing a timely and effective operational damage detection method. A calibrated finite element model is used to simulate damaged bridge GDF samples under four damage scenarios. The damaged bridge GDF samples are compared with the healthy baseline sample using the rank-sum test, and the results are employed to develop a damage index capable of alerting bridge owners of potential damage. A simple bootstrap resampling scheme is used to evaluate the probability of failing to issue an alert when the bridge is damaged (Type I error), as well as the likelihood of issuing a false alarm (Type II error). A three-dimensional statistical bridge signature is developed to aid damage localization and assessment. Nonparametric prediction intervals corresponding to a baseline signature are generated using the bootstrap method, creating an envelope of possible baseline bridge signatures. When a bridge signature falls outside the baseline bridge signature envelope, damage is detected. Damage was successfully identified for all four artificial damage cases considered. The overall damage detection method is designed to alert bridge owners when damage is

detected and to provide a probabilistic tool to aid damage assessment and localization while controlling for both Type I and Type II errors.

Keywords: *Bridge signature, structural health monitoring, operational strain monitoring, girder distribution factor, damage index, hypothesis test, damage detection*

1. Introduction

The American Society of Civil Engineers (ASCE) estimated that approximately 210 million trips were taken per day over structurally deficient bridges in the United States in 2013 [1]. In 2010, the Federal Highway Administration (FHWA) reported the cost of improving the nation's aging infrastructure greatly exceeded baseline spending [2]. Visual bridge inspections are required every two years, but these inspections can often be subjective and inconsistent, as shown by Moore et al. [3]. Structural health monitoring systems can be an effective means of supplementing visual inspections with objective measured data. The probabilistic damage detection method presented herein can be implemented to alert a bridge owner when damage is detected and provide a tool to aid damage assessment and localization.

1.1. Literature Review

The live load distribution factor for a bridge is the ratio of the live load applied to each girder when a vehicle crosses the bridge. When a bridge is designed, AASHTO distribution factors are calculated to determine the percentage of the design load to be carried by each girder based on enveloped maximum live loads [4]. These distribution factors are appropriately conservative. The distribution factor can also be calculated using measured strain data. The term Girder Distribution Factor (GDF) is used herein to distinguish the GDF calculated using measured strain data from the AASHTO distribution factor. Ghosn et al. [5] assumed the GDF for identical girders to be the individual girder recorded strain divided by the sum of all girder strains at a transverse location:

$$GDF = \frac{\varepsilon_i}{\sum_{j=1}^N \varepsilon_j} \quad (1)$$

Since the result of (1) represents the percentage of the live load carried by each girder, the sum of the GDFs for a bridge must be equal to 1:

$$\sum_{j=1}^N GDF_j = 1 \quad (2)$$

This method of calculating GDFs using measured strains has been commonly accepted and is referenced throughout the literature [6] [7] [8]. When all girders have the same stiffness, (1) represents the percentage of the live load carried by each girder. When the girders have different stiffnesses, (1) does not represent the true distribution of the live load, but can be thought of as a comparison of girder peak strains relative to other girders. In this form, the GDF is an effective measure of bridge performance and can be used evaluate changes in girder load sharing.

Stallings and Yoo [9] refined the Ghosn et al. [5] method to account for bridges with different interior and exterior girder sizes. This method used the ratio of section moduli to weight the measured strains and calculate the portion of the load carried by each girder. Cardini and DeWolf [8] employed strain data to compute an envelope of acceptable GDFs, noting that a damaged girder would likely produce a GDF below envelope values. Chakraborty and DeWolf [10] used continuous strain monitoring to compute girder stresses during truck events. The cumulative distribution function (CDF) was utilized to show the probability of the measured stress exceeding the design stress. Kim and Nowak [11] measured GDFs under normal truck traffic and used the CDF to comment on trends in traffic patterns. Plude [12] employed GDFs to investigate the observability of various damage cases, using the standard deviation of the GDFs to establish an envelope of acceptable values. Wipf et al. [13] calculated GDFs for ambient and load test traffic on a high-performance steel bridge, finding that measured distribution factors were typically much smaller than AASHTO distribution factors. Kim et al. [14] observed that under very heavy loads, governing distribution factors were reduced, indicating a redistribution of loads to girders further from the most stressed girder. Shenton and Hu [15] used a genetic algorithm to identify the location and severity of damage based on the redistribution of dead load bending moment. Catbas et al. [16] studied the structural response of bridge components under long term monitoring, observing that temperature effects had an impact on overall system reliability.

Data acquisition (DAQ) systems, both long-term and temporary, continue to become more affordable due to advances in technology. Howell and Shenton [17] created an inexpensive and rapidly deployable bridge monitoring system, emphasizing its use in monitoring fatigue life.

Whelan and Janoyan [18] developed and tested a wireless sensor network for real time strain monitoring with remote access capabilities. Teixeira et al. [19] used long-term monitoring for a retrofitted orthotropic bridge deck to observe reduced stresses over one year of monitoring.

Follen et al. [20] defined a bridge signature as the “expected response of a bridge structural system to daily traffic as measured by an instrumentation system”. Peak strains collected for heavy truck events were used by Follen et al. [20] to develop a nonparametric survival distribution function (SDF) representing the probabilistic behavior of a healthy bridge. Nonparametric prediction intervals were then developed using the bootstrap method, with a bridge signature falling outside of these prediction intervals indicating possible bridge damage corresponding to a particular level of confidence. The work described in this article employs the idea of statistical bridge signatures introduced by Follen et al. [20] and extends their ideas within a statistical decision and hypothesis testing framework to design an effective strategy for bridge damage detection.

1.2. Objective and Scope

This research introduces a hypothesis testing framework that enables a rigorous, decision-oriented approach for damage detection on operational bridges. The method targets bridges where single vehicle crossings are common. Rules are presented for extracting data when only one vehicle is crossing the bridge. Two different hypothesis tests for bridge damage detection are developed based on GDFs calculated from measured strain data. A sample of GDFs was drawn to establish a baseline, representing the behavior of a healthy bridge under normal daily traffic. Because the bridge studied is new and is in good condition, a finite element model (FEM) was used to simulate four bridge damage scenarios in order to evaluate the proposed methodology. A FEM is not needed to carry out this damage detection method, and was only used as a substitute for actual data from a damaged bridge. Four levels of damage identification are commonly referenced in structural health monitoring: (1) detection, (2) localization, (3) assessment, and (4) consequence [21]. The proposed two-part probabilistic damage detection method was shown to detect damage, as well as aid damage localization and assessment. In Part I, damage was detected and assessed using a damage index based on the rank-sum hypothesis test statistic. In Part II, a three-dimensional statistical baseline bridge signature envelope was established using a nonparametric probability distribution based on the bootstrap method. Simulated bridge damage

was detected, assessed, and partially localized based on whether or not bridge signatures fell outside of the baseline envelope. The two components of the damage detection method were designed to work together to alert bridge owners of potential damage and aid in damage localization and assessment.

The Type I and Type II error probabilities are of critical importance to any damage detection method. In this research, a Type I error corresponds to issuing a bridge damage alert when no damage is present, often termed a false alarm. The more critical Type II error corresponds to not issuing a damage alert when damage is present. An evaluation of both of these errors is central to the development of the overall methodology and distinguishes this research from previous work.

2. Data Collection and Data Quality Analysis at the PMB

The Powder Mill Bridge (PMB) is a three-span continuous bridge located in Barre, Massachusetts (Figure 1). It was constructed in 2009 and is in good condition. The deck cross section is shown in Figure 2. The bridge is 47 m (154.2 ft) long, with a center span of 23.5 m (77.1 ft) and ends spans 11.75 m (38.6 ft) in length. The bridge is non-skewed and carries two lanes of traffic and a sidewalk. The deck is 200 mm (0.66 ft) thick and is supported by six steel girders, spaced at 2.25 m (7.38 ft) with 732.5 mm (2.4 ft) overhangs. The exterior girders are W920x345 (W36x232) and the interior girders are W920x238 (W36x160). Ten MC460x63.5 (MC18x42.7) transverse diaphragms are provided over the length of the bridge. The deck concrete has a modulus of elasticity of 27,400 MPa (3,974 ksi). Though the bridge instrumentation consists of over 200 sensors, only six strain gauges were required for this damage detection method. A DAQ system mounted underneath the bridge collects data throughout the day. Additional details about the PMB are provided in [23] and [31].

Figure 1: Powder Mill Bridge

Figure 2: PMB Cross Section

GDFs were calculated using strain measurements collected by six strain gauges during traffic events on the PMB. A traffic event was recorded each time a vehicle crossed the bridge. No minimum strain threshold was set for collecting traffic events. The six strain gauges of

interest were located on the girder bottom flange within 1 m (3.28 ft) of midspan on the center span near the location of the maximum positive moment. The DAQ system recorded strain readings each day from 6:00 AM to 6:00 PM at a sampling frequency of 50 Hz. Each evening, a processing program extracted and stored traffic events. Strain gauge readings are known to drift over the long term due to changes in temperature and environmental effects. For the PMB, each traffic event consisted of approximately 20 seconds of recording. This period of time was not long enough for a drastic temperature change, thus the long term strain gauge drift could be removed by zeroing each traffic event. Each event was zeroed by subtracting the average ambient strain prior to the traffic event from each strain reading recorded during the event. This ensured that only strain readings due to the live load were captured. Measurement noise was filtered using a moving average window.

The lever rule can be used to illustrate the variability associated with computing the GDF under various travel paths and vehicle weights. The lever rule calculates the static summation of moments about one point in order to determine the reaction at a second point [4]. It can be used to understand how a vehicle travelling close to the curb will produce different GDFs than a vehicle travelling close to the centerline. While one high GDF could be the result of a vehicle travelling unusually close to a curb, many high GDFs could be indicative of damage. Different vehicle weights also result in different GDFs. Olund and DeWolf [22] computed the GDFs for two different truck configurations on a bridge in Connecticut, noting that the main load carrying girder had a GDF of 0.34 for Truck 1 and 0.38 for Truck 2. For these reasons, a nonparametric statistical approach is well-suited for computing GDFs under operational monitoring.

Strain data was collected at the PMB intermittently over the course of 14 months. During this time, seasonal differences were observed between data collected in summer and winter months. A number of factors could have contributed to these differences, including ground freezing, seasonal traffic pattern changes due to snow embankments, unexpected bearing pad and expansion joint behavior, or DAQ system temperature sensitivities. More long term strain data should be collected to investigate this uncertainty. On bridges where seasonality exists, the authors propose extreme baselines be established for the summer and winter seasons. Future studies correlating a relationship between temperature and strain data could be used to determine a correction for seasonality effects on baseline conditions. The nonparametric bootstrap resampling procedures used throughout this work require data to be resampled in such a way that

the bootstrap samples are independent and identically distributed (iid). This assumes that a vehicle crossing is not influenced by the vehicle that crossed before it (independent), and no overall trend exists in the vehicle crossings (identically distributed). To insure this to be the case, data used in this work comes from a set of three summer months, June, July and August, in which the GDFs were shown to be iid. This method is valid for any set of data that can be shown to be iid, though a more advanced moving blocks or nearest neighbor bootstrap could be applied when this assumption is violated.

Data quality was a critical aspect of GDF collection, thus three rules were created to extract events in the southbound lane and to remove and correct undesirable events: (1) select southbound events, (2) remove events with multiple vehicles, and (3) correct for negative Girder 6 strains. This resulted in a set of 1,177 southbound traffic events available for this research.

2.1. Rule 1: Select southbound events

GDFs produced by northbound and southbound traffic differed due to the location of the vehicle travel path. In order to establish a range of expected GDFs, the events were sorted by lane. Rule 1 was used to select southbound traffic events, the focus of this research. The same damage detection method can be used for northbound traffic. In the initial screening for southbound traffic, all events with Girder 1 peak strains exceeding Girder 5 peak strains were selected. Figure 3 and Figure 4 show recorded strain data for Girders 1 to 6 at a single transverse location during typical northbound and southbound traffic events, respectively. For southbound events, Girders 2 and 3 typically experienced the highest strains, while Girders 5 and 6 recorded the smallest strains.

Figure 3: Typical event in northbound lane

Figure 4: Typical event in southbound lane

2.2. Rule 2: Remove events with vehicle in opposite lane

Events initially sorted into the southbound group were screened a second time for peak strains that indicated multiple vehicles on the bridge. When one vehicle crossed the bridge, it was expected that each girder peak strain would occur within the window of the overall peak strain. When individual girder peak strains occurred outside of this window, it indicated that

multiple vehicles were on the bridge. An example of this is shown in Figure 5, where the peak strain for Girder 5 is outside of the window for the peak overall strain (Girder 2). In practice, these events will be removed from the set when they are identified during the measured strain data quality analysis. During these events, the second vehicle was able to influence the strains recorded for the primary vehicle, making the GDF calculation for the primary vehicle inaccurate. The short span and rural location of the PMB made this occurrence relatively uncommon.

Figure 5: Traffic event with vehicles travelling in opposite directions

2.3. Rule 3: Correct for negative Girder 6 strains

The geometry of the PMB occasionally caused very small negative live load strain readings to be recorded for Girder 6 when a vehicle traveled in the southbound lane near Girder 1. This was caused by the transverse stiffness of the deck and diaphragms, which resulted in very slight uplift and negative bending in Girder 6 when a vehicle was close to the southbound curb. Rule 3 corrected these events by setting ε_6 to 0 since Girder 6 did not contribute to the load carrying of southbound vehicles and the negative bending was minimal. The GDFs for Girders 1 to 5 were calculated using (1).

Strain measurements for approximately 75 usable southbound traffic events were captured each day at the PMB. A sample size of 500 traffic events was needed to establish a baseline probability distribution of GDFs that was both stable and repeatable, meaning the baseline could be collected in approximately one week of monitoring. The simulated potentially damaged bridge sample consisted of 75 events, representing one day of monitoring. Using this method, a baseline can be established for a new or existing bridge in one week of strain monitoring. A potentially damaged bridge sample of traffic events can be collected each day. This would allow the bridge owner, by the end of the day, to know if a change had occurred in the measured GDFs. Whether the bridge is new or existing, any deviation from the baseline is a departure from normal operation and can be indicative of damage.

3. Finite element modeling to simulate bridge damage

In order to test the proposed method of damage detection, and since the PMB is a new bridge with no known damage, data for the damaged bridge was simulated using a FEM to

examine the capability of the proposed method. A calibrated FEM for the PMB was developed by Sanayei et al. [23]. The initial model was created based on design drawings using eight-node solid elements to model the deck and four-node shell elements for the girders. Steel reinforced elastomeric neoprene bearing pads were modeled using springs with axial, shear, and rotational stiffnesses calculated to represent the support behavior, as described in [23]. Data was collected during a diagnostic load test and was used to calibrate the FEM. Three steps were taken to calibrate the model: (1) concrete strength updated from the design value of 30 MPa (4.35 ksi) to 33.6 MPa (4.87 ksi) based on cylinder break data, (2) parapet stiffness added, and (3) deck stiffness reduced in the negative bending region from 27,400 MPa (3,974 ksi) to 18,000 MPa (2,611 ksi) due to concrete in tension. The undamaged bridge was defined as Case U. Four different damage cases were analyzed: (A) interior girder fracture, (B) fascia girder corrosion, (C) diaphragm fractures, and (D) deck delamination.

GDFs were extracted from the FEM for a simulated HS-20 truck centered in the southbound lane. The HS-20 truck weighed 72 kips with three axles spaced at 4.3 m (14 ft). The simulated strains were calculated at the same midspan bottom flange strain gauge locations used to establish the measured baseline bridge signature for the PMB. Figure 6 shows a summary of the GDFs calculated for Cases A to D and U. Girder 6 did not produce large enough strains under southbound traffic to justify reporting a change in the GDF.

Figure 6: Summary of GDFs for Cases A-D and U

For Cases A to D, ΔGDF was defined as the difference between the GDF for the damaged bridge and the GDF for the undamaged bridge, Case U. This represented the anticipated change in GDF for a damaged bridge under the given damage scenario. Similar to (2), the sum of the six ΔGDF values must always be equal to 0:

$$\sum_{j=1}^N \Delta GDF_j = 0 \quad (3)$$

It is assumed that a damaged bridge will produce:

$$GDF_{\text{Damaged}} = GDF_{\text{Undamaged}} + \Delta GDF. \quad (4)$$

Though the position and configuration of traffic under operational monitoring vary, ΔGDF remains relatively constant. In order to simulate a sample of damaged bridge GDFs, a subset of 75 traffic events was selected from the set of 1,177 southbound traffic events. To create damaged

bridge samples of 75 events for Cases A to D, ΔGDF was added to the GDF for each event in the subset.

3.1. Case A: Interior girder fracture

Case A was the sudden brittle fracture of Girder 2 at the midspan. This damage case was based on brittle fractures causing sudden, localized damage to girders on the Hoan Bridge [24], Delaware's I-95 Bridge over the Brandywine River [25], and US 422 bridge over Schuylkill River in Pennsylvania [26]. In all cases, the fracture occurred close to the midspan near a steel detail location. Farrar and Jauregui [27] created a fatigue crack in a plate girder using a 9.5mm (0.38 in) wide cut that was initiated in the web of the girder and progressed downward through the bottom flange. To simulate a realistic girder fracture on the PMB, a reduced stiffness was assigned to a 2 mm (0.08 in) wide section by setting the elastic modulus to a near zero value. This simulated crack extended from the middle of the web through the bottom flange and was located on Girder 2 at a diaphragm connection near the midspan.

Timely detection is critical for sudden brittle fracture of steel girders. In two of the three case studies mentioned, cracks were not identified until a regularly scheduled repair or inspection. The damage detection method developed in this research would allow only one day's worth of traffic to cross the damaged bridge before an alert would be issued. The GDFs computed for Case A are shown in Figure 6. The crack reduced the stiffness of Girder 2, resulting in lower Girder 2 GDFs. To compensate for this, Girders 1, 3, 4, and 5 carried larger portions of the load, producing higher GDFs.

3.2. Case B: Fascia girder corrosion

Case B was section loss on the fascia girder due to corrosion. This is common in steel girders, and can be accelerated by chloride in de-icing salts that can mix into runoff water and pour over the edge of a bridge deck onto exterior girders [28]. This causes corrosion and results in section loss. Miller et al. [29] tested corroded fascia girders removed from a deteriorated bridge and estimated the global stiffness loss to range from 13-32%, with the majority of the section loss occurring in the tension flange and web. To model this damage case for the PMB, the web and bottom flange sections of Girder 1 were reduced to produce a stiffness loss of 30%. The GDFs for Case B are shown in Figure 6. Similar to Case A, the reduced stiffness of Girder 1

produced a lower Girder 1 GDF. The neighboring girders carried higher loads to counterbalance this, yielding higher GDFs for Girders 2 and 3. The damage in Case B was less severe than in Case A, resulting in less significant ΔGDF values.

3.3. Case C: Diaphragm fractures

Case C was fatigue cracking in diaphragms at diaphragm to girder connections. Zwerneman et al. [30] investigated the cause of fatigue cracking in diaphragm to girder connections on Oklahoma's I-40 bridge near Weatherford. It was determined that a high level of restraint at these connections caused nearly 1/3 of the diaphragms to fracture. The cracks initiated in the coping of the steel channel diaphragms and propagated upwards through the web. Though this did not threaten the structural integrity directly, it affected the distribution of the loads. Over time, this modified load distribution could cause members to become overstressed. Similar to the midspan cracking pattern observed by Zwerneman et al., four of the ten diaphragm to girder connections near the midspan of the PMB were modeled with cracks. For details of the selected cracking pattern, see [31]. The GDFs for Case C are shown in Figure 6. Diaphragm cracking resulted in disruption of the load distribution. Since the southbound lane was centered on Girders 2 and 3, these girders were forced to carry more of the load, producing higher GDFs. Girders 1 and 4 were not able to carry as much of the load, resulting in lower GDFs.

3.4. Case D: Deck delamination

Case D was deck delamination. Deck damage is one of the most costly repairs faced by bridge owners [32]. Delamination is caused by corrosion of reinforcement due to long term exposure to chloride ions or moisture. Corrosion usually occurs in the top layer of deck reinforcement, decoupling the concrete from the rebar and reducing the strength of the structure [33]. Deck delamination for the PMB was modeled as a reduction in deck stiffness of 35%, roughly the stiffness lost if the top layer of rebar were to decouple from the concrete and become damaged. This damage was applied in a patch measuring 10 m (32.8 ft) in the transverse direction by 16 m (52.5 ft) in the longitudinal direction, centered in both directions on the middle span of the PMB. The GDFs for Case D are shown in Figure 6. Deck delamination reduced the deck's ability to distribute the load, similar to Case C. As a result, Girders 2 and 3, directly

underneath the southbound lane, carried higher loads. This resulted in behavior similar to Case C: GDFs for Girders 2 and 3 increased, while GDFs for Girders 1 and 4 decreased.

The following sections describe two damage detection methods evaluated using the damaged bridge samples simulated in this section. Part I introduces a damage index designed to detect damage and issue alerts to bridge owners. In Part II, a 3D statistical bridge signature is used to aid bridge owners in the assessment and localization of potential damage.

4. Part I: A Damage Index for Damage Detection

In Part I of this damage detection method, a damage index (*DI*) was developed as a hypothesis test statistic. A hypothesis test provides a useful framework for making decisions based on available data. The goal was to develop a hypothesis test for determining whether a sample from a potentially damaged bridge differed enough from a sample from an undamaged bridge to conclude that the bridge was damaged. Hypothesis tests are advantageous because they present a rigorous and standardized approach for making decisions with limited data, and enable evaluation of the likelihood of making Type I and Type II errors. A rigorous hypothesis testing framework can identify the likelihood of both underdesign and overdesign, key concepts in infrastructure planning and management problems [34].

The decision matrix for the damage detection hypothesis test is formulated in Figure 7. The null hypothesis, H_0 , corresponds to an undamaged bridge. The alternate hypothesis, H_A , corresponds to a damaged bridge. The goal is to ascertain which of these hypotheses represents the truth based on a limited sample of bridge data. At the outset, the decision maker does not know which of these two hypotheses are correct, and the purpose of the test is to make this determination. As shown in Figure 7, two different errors are possible. A Type I error implies that H_0 is rejected when it is actually true. Thus, a Type I error results from concluding that an undamaged bridge is damaged, leading to over-preparedness and unnecessary costs. Normally, the probability of a Type I error is set prior to the test and is represented by α , termed the *significance level*. A significance level of 5% is common and was the α -level assumed in this research. A Type II error results from concluding that a damaged bridge is not damaged, which can lead to critical conditions. The probability of a Type II error indicates the likelihood of missing damage when it is present, which corresponds to under-preparedness and is represented by β . The complementary probability, $1-\beta$, is termed the *power of the test*. A test with high power

results in a low probability of a Type II error, thus a powerful damage detection hypothesis test will lead to a low probability of under-preparedness. It is of critical importance to consider both the Type I and II errors, as shown in Figure 7, yet as Vogel et al. [34] lament, it is remarkably uncommon to report both errors. Rosner et al. [35] further illustrate how both errors can be integrated into a rigorous decision making framework for infrastructure planning by considering the expected regret associated with each of the possible errors.

Figure 7: Damage detection hypothesis test decision matrix

4.1. Rank-sum test

The two-sided Wilcoxon Rank-Sum hypothesis test statistic is employed as the initial test statistic. It was originally developed to investigate the use of ranking methods to determine the difference between samples of data [36]. This nonparametric test statistic determines the probability that two independent samples come from continuous distributions with equal medians. Previous studies have demonstrated the power ($1-\beta$ in Figure 7) of the Wilcoxon rank-sum test statistic for various distributions [37] [38]. When it is expected that one of the samples has either a larger or smaller median than the other, a one-sided test is performed; otherwise a two-sided test is performed [39]. A two-sided test was performed in this study because it was not known *a-priori* in which direction the shift would occur. There are two methods for performing the rank-sum test depending on the sample size. For sample sizes less than 10, the exact rank-sum test is necessary [39]. When both sample sizes under consideration exceed 10, as is the case in this research, an approximate method is suitable.

In the approximate test, the damaged bridge sample, x_i , has a sample size of n . The undamaged bridge sample, y_i , has a sample size of m . The combined sample is of size M , where $M = n + m$. The null hypothesis assumes that the two samples come from the same distribution so that:

$$H_0: \text{Prob}(x_i \geq y_i) = 0.5 \quad (5)$$

The alternate hypothesis for a two-sided test, when it is not known whether x_i is expected to be larger or smaller than y_i , assumes that:

$$H_A: \text{Prob}(x_i \geq y_i) \neq 0.5 \quad (6)$$

The test statistic used is the sum of the ranks, W , given in (7). To find W , x_i and y_i are combined and ranked in ascending order. Each measurement receives a rank, R , ranging from 1 to M based on the rank of the measurement within the combined sample. The test statistic is the sum of the ranks for the sample having the smaller sample size:

$$W = \sum_{i=1}^{n'} R_i \quad (7)$$

where $n' = \min(n, m)$. When both sample sizes n and m , exceed 10, the distribution of W can be approximated by a normal distribution [39]. Under the null hypothesis of no damage, the mean μ_w and standard deviation σ_w of W are:

$$\mu_w = \frac{n(M+1)}{2} \quad (8)$$

$$\sigma_w = \sqrt{\frac{n \cdot m(M+1)}{12}} \quad (9)$$

where n denotes the size of the potentially damaged bridge sample, m is the size of the undamaged bridge sample, and $M = n+m$. The cumulative probability associated with a particular value of W can be calculated using a standard normal variable Z , defined as:

$$Z = \frac{W - \mu_w}{\sigma_w} \quad (10)$$

For a two-sided test, the p -value is doubled to reflect the probability of the test statistic being on either side of the mean:

$$p = 2[1 - \Phi(|Z|)] \quad (11)$$

where $\Phi(Z)$ represents the CDF of the standard normal distribution. The decision to reject the null hypothesis is made by comparing the p -value with the assumed significance level α . The decision of the damage detection rank-sum test is the Boolean decision variable h . When $p \geq \alpha$, the decision h is 0; when $p < \alpha$, the decision h is 1.

4.2. Results of hypothesis test

For the hypothesis test presented in Figure 7, there are two possible outcomes that result in no error. If the null hypothesis is known to be true (bridge is undamaged) and the rank-sum p -value is greater than or equal to the assumed significance level, the null hypothesis is accepted

and it is correctly determined that the bridge is not damaged. If the alternate hypothesis is true (bridge is damaged) and the p -value is less than the significance level, the null hypothesis is rejected by the test and it is correctly determined that the bridge is damaged.

Similarly, there are two situations that produce errors and lead to incorrect decisions. If the bridge is not damaged and the p -value is less than α , it is incorrectly concluded that the bridge is damaged and a Type I error occurs. This results in over-preparedness, and could lead to unnecessary resources being allocated to what is actually a healthy bridge. If the bridge is damaged and the p -value is greater than or equal to α , it is incorrectly concluded that the bridge is not damaged and a Type II error occurs, resulting in under-preparedness. In bridge damage detection, a Type II error may have considerable consequences.

It is important to understand the conclusions and limitations of a hypothesis test. The test determines whether the available data indicate that the null hypothesis should be accepted or rejected based on an assumed significance level. The Type I error probability α is set *a-priori*, whereas the Type II error β is unknown. When the null hypothesis is rejected, it cannot be concluded with certainty that the alternate hypothesis is true because of the possibility of a Type I error. The significance level directly affects the Type II error. When the acceptable level of the Type I error is increased, the null hypothesis is rejected more often. As a result, the Type II error decreases. Similarly, if a lower significance level is set, the Type II error increases because the null hypothesis is rejected less frequently. For bridge monitoring, it is tempting to increase the significance level in order to reduce the probability of the Type II error. This increases the Type I error probability, however, which reduces the ability of the test to discern damage from false alarms. Thus, acceptable values of α and β should be determined based on the needs of the bridge owner. In this instance, the only way to decrease both α and β would be to collect more data (see [34] for further discussion).

4.3. Minimum Detectable Damage Levels

Simulations, and in some instances analytical probabilistic calculations, can be used to study the likelihood of a Type II error associated with a hypothesis test. Most studies which employ hypothesis tests tend to focus only on the likelihood of rejecting the null hypothesis without considering the power of the test [34]. The probability of accepting the null hypothesis when the alternate hypothesis is true is often critical in infrastructure applications because the

Type II error typically has potential for disaster. The classic elevator cable problem is often used as a pedagogic tool for understanding the consequences of Type I and Type II errors. In this problem, the Type I and Type II errors correspond to making the elevator cable either too strong or too weak, respectively. Surely one would want to minimize the uncontrollable Type II error corresponding to the consequences of making an elevator cable too weak.

To better understand the likelihood of Type II errors associated with this test, a power study was performed to estimate β for increments of ΔGDF . The goal was to determine at what damage level the test could identify damage for each girder corresponding to both a Type I and Type II error probability of 5%. Two-hundred damage increments were evaluated from $\Delta GDF = 0.001$ - 0.02 . At each increment, a random sample of 500 events was drawn, with replacement, to establish a baseline. Drawing samples with replacement is known as the bootstrap and ensures that the generated samples have the same probability distribution as the original sample [40]. A random sample of 75 events was drawn, with replacement, to represent the potentially damaged bridge sample. Damage was added to this sample based on ΔGDF increments. A rank-sum test was performed with a significance level α of 5% and the h -value was recorded for each test. This was performed 1,000 times for each increment of bridge damage. The Type II error probability β was the percentage of simulations that incorrectly concluded that the bridge was not damaged. As the damage level increased, the null hypothesis was correctly rejected more frequently, decreasing the Type II error. Figure 8 shows a summary of these simulations for each girder, with a horizontal line denoting β of 5%. From Figure 8, the likelihood of concluding that a damaged bridge is not damaged (probability of a Type II error) can be found for each girder at any damage level. Table 1 presents the minimum change required in the GDF for damage detection on each girder. The first row of Table 1 summarizes the minimum detectable ΔGDF increments for each girder based on the chosen α and β levels. The second row of Table 1 provides the minimum percent change the GDFs must undergo before damage is detected. For southbound traffic, the GDF for Girder 6 was not significant enough to justify reporting.

Figure 8: Detectable damage level for $\alpha = 5\%$ and $\beta = 5\%$

Table 1: Minimum detectable damage levels for $\alpha = 5\%$ and $\beta = 5\%$

The transverse position of the vehicle on the bridge impacts the range of the GDFs collected, and consequently the ΔGDF threshold required for damage identification. A wider distribution of GDFs was collected for the girders not directly underneath the southbound lane. As a result, a larger change in GDF was required to identify damage in Girders 1, 4, and 5. Girders 2 and 3, directly below the southbound lane, were less affected by variations in the transverse position of the vehicle. As a result, a smaller percent change in GDF was needed to identify damage for Girders 2 and 3.

4.4. Bridge Damage Index

A bridge damage index was developed to quantify the overall damage based on the decision variable, h , obtained from the rank-sum test. Here, simulated damaged bridge data obtained from the FEM was used to test the proposed damage index. A single baseline was established for the undamaged bridge by selecting a sample of 500 consecutive events. This was termed *Sample A*, and is referenced again in Part II. A potentially damaged sample of 75 consecutive events, termed *Sample B*, was drawn outside of the baseline. These samples were drawn to mimic the collection methods that would be used to gather an actual baseline and potentially damaged sample. Varying levels of damage were applied to Sample B based on the ΔGDF values calculated in Section 3. The damage index, DI , was calculated on a scale from 0.0 to 1.0 using N as the number of girders and j as the damage case in (12). Table 2 shows a summary of h -values and DI values for the five cases. Case U, the undamaged bridge, was included to evaluate the Type I error.

$$DI_j = \frac{\sum_{i=1}^N h_i}{N} \quad (12)$$

Table 2: Summary of damage indices for sample damage cases

Figure 6 and Figure 8 can be used to better understand the results of Table 2. For example, in Case D, damage was not identified on Girder 4 for $\Delta GDF = -0.0029$. From Figure 8, ΔGDF of 0.0029 corresponds to a Type II error probability of 73% for Girder 4. This Type II error probability is too high to expect damage to be identified, thus it was not surprising that the rank-sum test did not detect damage on Girder 4 for Damage Case D. Damage Cases A to D produced

nonzero *DI* values, correctly indicating damage was present. As expected, the undamaged Case U produced a *DI* of 0, signifying no damage was identified.

A simulation experiment was run to investigate at which damage index levels alerts should be issued. The goal was to determine *DI* values where damage was correctly identified without large Type I error probabilities. To calculate the *DI* for one trial, a baseline sample of 500 events was drawn randomly, with replacement. A sample of 75 events was independently drawn randomly with replacement. Five potentially damaged samples were created from this sample of 75 events, representing Cases A to D and Case U. The *DI* was calculated for each of the five potentially damaged samples. This was performed for 1000 trials.

For each *DI* level from 0.0 to 1.0, the percentage of trials producing *DI* values at or above the given *DI* level was calculated. Figure 9 shows *DI* plotted against the conditional probability of detecting damage. This represents the probability of detecting damage for each of the five cases. In other words, for a given *DI* level, Figure 9 shows the probability of computing at least that *DI* level for each damage case.

Figure 9: Simulation to determine *DI* values that indicate damage

All trials with $DI > 0$ for Case U were the result of the Type I error. The conditional probability of detecting damage computed for Case U represents the Type I error probability corresponding to each *DI* level. For example, Figure 9 shows that when the bridge was undamaged, it was expected that a *DI* of at least $1/6$ would be computed for 12% of trials.

Since no damage was added to Girder 6 for Cases A to D, the highest anticipated *DI* for these four damage cases was $5/6$. All trials with $DI < 5/6$ for Cases A to D were the result of the Type II error.

Figure 9 can be further explained by comparing Cases A and B. From Table 2, the damage associated with Case A is expected to be more severe than Case B, resulting in smaller ΔGDF values applied to each girder in Case B. Figure 9 shows that 100% of trials calculated a *DI* of at least $3/6$ for Case A, and that 90% of trials calculated a *DI* of at least $3/6$ for Case B. Since Case B was less severe, it makes sense that fewer trials calculated a *DI* of at least $3/6$. Damage was less likely to be identified on girders when smaller ΔGDF values were applied. For

this reason, the probability of calculating a given DI level was expected decrease as the DI level increased.

The first damage index threshold was set at $DI \geq 3/6$. This DI would be the result of at least half of the bridge girders detecting a change in measured GDFs. When the bridge was undamaged, only 3% of trials computed a DI of at least $3/6$. When the bridge was damaged, 90 to 100% of trials, depending on the damage case, computed a DI of at least $3/6$. Put another way, the Type II error probability associated with $DI = 3/6$ was 0-10% depending on the damage case. Based on these α and β values, $DI \geq 3/6$ appeared to be a very effective indicator of damage.

A second DI threshold was established at $DI \geq 1/6$ to issue an alert indicating that damage was possible and the bridge should be monitored closely. For a six girder bridge such as the PMB, one girder indicating damage results in $DI = 1/6$. When a single girder is damaged, the load is redistributed to the undamaged girders, resulting in multiple girders indicating a change in GDF rather than just one. It is unlikely that a serious damage case would occur that would result in only one girder showing a change in GDF. For this reason, a moderate warning was appropriate for a DI greater than $1/6$ but less than $3/6$. From Figure 9, when the bridge was not damaged, a damage index of at least $1/6$ was calculated for 12% of trials. When the bridge was damaged, a damage index of at least $1/6$ was computed for 100% of trials. The Type I error probability of 12% was too high for a severe warning to be issued. However, the Type II error probability of 0% indicates that if damage were actually present, a moderate alert would most likely be issued.

DI ranges and recommended responses are summarized in Table 3. A color based system, similar to a traffic light, was used to categorize the probability of damage as low (green), medium (yellow), and high (red). When DI was 0, the probability of damage was low, and the bridge could be assumed healthy. A DI value greater than $1/6$ but less than $3/6$ indicated that damage was possible and an inspection needed to be scheduled. When $DI \geq 3/6$, damage was probable and an inspection was needed as soon as possible. The DI values computed in Table 2 for Cases A to D would all result in the most severe warning possible. Though a FEM was used to establish this set of DI ranges and recommended responses, future work should include parametric studies to develop these ranges for various bridge geometries, which would preclude the need for a FEM. This would result in a damage detection method requiring only measured data.

Table 3: Summary of Damage Index

5. Part II: Damage detection using statistical bridge signatures

The bridge damage index developed in Part I provides a warning of the likelihood of potential damage, but does not provide information on the magnitude or location of the damage. In Part II, bridge signatures are used to evaluate bridge damage based on the observability of damage in bridge signatures when compared with a baseline signature envelope. The distribution of GDF data exhibited multiple modes, making it challenging to select a suitable theoretical probability distribution function (PDF) for this analysis. Instead, a nonparametric approach was used to establish a bridge signature, making it unnecessary to select and fit a theoretical PDF to the observations. For a more detailed discussion and justification for the use of nonparametric methods in developing statistical bridge signatures, see [20].

A bridge signature was developed using a nonparametric survival distribution for each girder based on GDFs calculated for traffic events. A survival distribution is the complement of the CDF and represents the relationship between a random variable and its exceedance probability. The statistical bridge signature method developed by Follen et al. [20] characterizes the probabilistic behavior of a bridge using a two-dimensional signature based on peak strains, whereas the method proposed in this article develops a three-dimensional signature based on girder load distribution.

5.1. Baseline Bridge Signature Envelope

Sample A, the same baseline sample of 500 events associated with an undamaged bridge used in Part I, was used again in this section. The potentially damaged bridge signature is based on 75 events, the number of events collected during one day of monitoring. To match the potentially damaged signatures, baseline signatures were developed using a suite of bootstrap samples of 75 events from the full baseline sample, chosen randomly with replacement. The following approach was used to plot the empirical survival function, which is a plot of the ranked GDF values (ranked in descending order) versus their associated exceedance probabilities. The Weibull plotting position, $i/(I+n)$, was used to estimate the exceedance probability associated with each ranked value, where i was the rank and $n=75$. The Weibull plotting position is an

attractive choice because it provides an unbiased estimate of the exceedance probability of any ranked random variable, regardless of its underlying distribution [39].

The bootstrap is a generalized resampling approach which can be used to replace nearly every theoretical statistical approach that exists [40]. The bootstrap is used here to establish an envelope of all possible SDF curves using nonparametric prediction intervals. In hydrology, SDFs are referred to as flow duration curves, which are used widely in the probabilistic analysis of daily streamflow. Vogel and Fennessey [41] introduced a nonparametric approach for developing confidence intervals for flow duration curves or empirical survival functions. An analogous approach was used to establish prediction intervals for SDF curves.

Each bootstrap sample was chosen by selecting 75 events randomly, with replacement, from the baseline set of 500. This was performed 1000 times, creating 1000 samples of 75 events. At each rank, 1000 bootstrapped GDFs were available; these were ranked in ascending order. A 99% prediction interval was created by eliminating the highest and lowest 0.5% of GDFs. To establish this prediction interval, the 5th and 995th GDFs were selected at each rank, resulting in two SDF curves containing 75 readings each. It was expected that 99% of all healthy bridge signatures would fall within this envelope. The resulting baseline bridge signature envelope for Girder 1 is shown in Figure 10, where the probability of exceedance is plotted against the GDF.

Figure 10: Girder 1 baseline bridge signature envelope

5.2. Evaluating an Undamaged Bridge Signature

The baseline bridge signature envelope shown in Figure 10 can be thought of as prediction intervals associated with the GDF for each girder. The prediction intervals for the GDF of the six girders are plotted in three dimensions in Figure 11, representing an envelope of healthy bridge signatures. To ensure that this envelope captured all possible healthy bridge signatures, Sample B, a set of 75 healthy events drawn consecutively outside of the baseline, was plotted against the baseline envelope. Potential damage to the bridge was analyzed by comparing the potentially damaged bridge GDF with the healthy envelope GDFs at each quantile. A few potentially damaged bridge GDFs falling outside of the envelope could be the result of a Type I error. Since the GDF prediction intervals enclosed 99% of all healthy bridge signatures, the Type

I error was expected to occur with a probability of 1% for healthy bridges. When many potentially damaged bridge GDFs plotted outside of the envelope, damage was likely. The dark (blue) surfaces represent the healthy bridge signature envelope. The light (green) surface represents the potentially damaged bridge signature. Sample B was found to plot entirely inside of the baseline healthy bridge signature envelope, providing further evidence that the envelope shown in Figure 11 was a valid representation of healthy bridge behavior.

Figure 11: Baseline bridge signature envelope

5.3. *Evaluating Damaged Bridge Signatures*

The damage index established in Part I correctly identified damage for all four simulated damage cases. This would result in an alert instructing the bridge owner to inspect the bridge immediately. Bridge signatures can be used as a tool to give the owner a better understanding of the damage prior to an inspection. The damaged bridge samples for Cases A to D in Part I were used again in this section. The damaged bridge signature for each case was plotted as a surface against the baseline envelope to determine whether damage could be identified, assessed, and localized (Figure 12). Again, the dark (blue) surfaces represented the baseline bridge signature envelope for an undamaged bridge, while the light (green) surface represented the damaged bridge signature. From the visibility of the damaged bridge signatures through the baseline envelope, damage could be assessed and partially localized. Though the simulated damage cases were known in this research, it is important to note that the same results would have been obtained with no prior knowledge of the damage locations or severities.

In Case A, Girder 2 was damaged due to the brittle fracture of the girder near the middle of the center span. Figure 12(a) shows the damaged bridge signature well outside of the baseline signature envelope for Girders 1 to 5, indicating damage was correctly identified. This was the result of Girder 2 carrying less of the load, producing lower GDFs. To compensate for this, Girders 1, 3, 4, and 5 carried larger loads, resulting in higher GDFs for these girders. The high magnitudes of these departures from the baseline, particularly for Girder 2, indicated a serious damage condition requiring immediate attention. Damage was not added to Girder 6, thus it correctly indicated no damage. An owner evaluating this bridge signature would be able to

identify Girder 2 as the likely cause of the damage alert and could direct the inspection team to focus on this girder.

In Case B, the stiffness of Girder 1 was reduced due to corrosion along the entire girder, (Figure 12(b)). The neighboring girders compensated for this, resulting in higher GDFs for Girders 2 and 3. This damage was correctly identified in Figure 12(b). The magnitudes of the departures from the baseline envelope for Girders 1 to 3 were smaller than those seen in Figure 12(a), correctly indicating a less severe damage condition. The damage to Girders 4 and 5 was not significant enough to show departures from the baseline healthy bridge signature. Figure 12(b) correctly indicated no damage had occurred to Girder 6, which was in fact undamaged.

In Case C, diaphragm cracking prevented the load from being properly distributed across the girders (Figure 12(c)). The southbound lane was centered on Girders 2 and 3; as a result, these girders produced higher GDFs. The bridge signature in Figure 12(c) correctly identified this damage. The magnitudes of the damage for Girders 2 and 3 were relatively low. Since these girders correspond to the main southbound load carrying girders, the owner may suspect this damage to be related to higher loads on these girders. This could, in turn, be identified as a load distribution issue, and load distribution components, such as diaphragms and the deck, would be checked by an inspection team.

In Case D (Figure 12(d)), concrete deck delamination in the positive bending region of the center span reduced the deck's ability to distribute the load. As a result, Girders 2 and 3, directly underneath the southbound lane, carried more of the load. The bridge signature correctly identified damage, and looked similar to the signature produced by Case C. Both cases involved damage due to load distribution issues, thus their signatures showed higher GDFs on Girders 2 and 3, directly below the southbound lane. Similar to Case C, the bridge signature produced by Case D would lead an inspection team to focus their inspection on load distribution components, such as the deck and diaphragms.

Figure 12: Signatures for four damage cases: (a) interior girder fracture, (b) fascia girder corrosion, (c) diaphragm fractures, and (d) deck delamination.

The results of the damage index established in Part I and the statistical bridge signatures developed in Part II can be compared to illustrate how the two methods are most powerful when used together. As an example, Cases A to D could represent four different bridges managed by a bridge owner. The owner would receive severe (red) alerts for all four bridges based on the *DI* values computed in Part I. The owner could use statistical bridge signatures to help determine which damaged bridge was a priority. For example, Cases A and C both reported a *DI* of 5/6. From the statistical bridge signatures, however, it would be clear that Case A was a more severe damage case because the Case A signature plotted well outside of the baseline bridge signature envelope. Another example can be drawn from Case B, which returned a *DI* of 3/6, the lowest of the four damage cases. From the bridge signature, however, it was clear that damage in Case B was more severe than in Cases C and D. These examples show that using the damage index in combination with statistical bridge signatures can be a powerful tool for detecting damage as well as a useful aid in damage localization and assessment.

6. Implementation and discussion

A permanent or temporary data acquisition system can be used to establish an initial baseline healthy bridge signature using measured strains. For the Powder Mill Bridge, this required one week of monitoring. Ideally, the healthy baseline would be established when the bridge is first opened. Since this methodology evaluates departures from the baseline, however, it is also effective for bridges already in operation. Once the healthy baseline is established, a new sample can be collected in one day of monitoring for the potentially damaged bridge. For bridges with permanent DAQ systems, such as the PMB, the potentially damaged bridge sample can be updated each time a traffic event is recorded. When a new event is added to the potentially damaged bridge sample, the oldest event in the sample is removed. This would allow the damage index and bridge signature to update in real time, increasing the timeliness of the alerts. Permanent DAQ systems are ideal, but can be expensive to install and maintain. Rapidly deployable strain monitoring systems can also be used to monitor departures from the baseline. An ideal use of the rapidly deployable system would be as a supplement to a visual bridge inspection, when a crew would already be on the bridge.

The two methods outlined in this research are best used together. The damage index provides an objective comparison of the baseline and potentially damaged signatures. *DI* values

can easily be interpreted for decision making, but do not report the magnitude of damage to individual girders. Using bridge signatures, trends can be identified based on the observability of the damaged bridge GDF magnitudes through the healthy bridge signature envelope. The best method of implementing this damage detection method is by monitoring the bridge for *DI* values outside of the healthy (green) range. When a warning is received due to a *DI* value exceeding the damage threshold, the bridge signature should be generated to aid in assessment and localization of damage.

This method of damage detection is developed using a six-girder bridge. Future work should investigate its application for wider bridges with a higher number of girders. Wider bridges will produce lower GDFs, which may make damage less observable. This could be improved by computing separate sets of GDFs for each travel lane based on the girders closest to each lane. Additional research should study environmental and operational effects on this damage detection method. Future work should also be performed to investigate the impact of different damage cases, damage magnitudes, sensor locations, and sensor types. A parametric study should be performed to establish sets of *DI* ranges and recommended responses for various bridge geometries.

7. Conclusions

A nonparametric statistical decision approach for damage detection under operational strain monitoring was established using girder distribution factors. A baseline sample of GDFs was drawn to represent a healthy bridge under normal operating conditions. A FEM was used to simulate a damaged bridge response to demonstrate the ability of GDFs to depict bridge damage. The change in GDF observed from data obtained from a FEM in four damage scenarios was used to create response samples for the damaged bridge. The rank-sum test was used to establish a damage index based on a comparison of undamaged and damaged bridge sample medians. Using a simple bootstrap, studies of the Type I and Type II errors associated with this test were used to establish damage alert thresholds. The damage index was shown to successfully identify damage in all four simulated damage cases with acceptable Type I and Type II errors. The approach was also used to document detectability of damage at different damage threshold levels. Three-dimensional statistical bridge signatures proved to be a useful tool to aid damage localization and assessment, providing a comparison of measured bridge GDFs with a range of expected GDFs at

each quantile of the probability distribution. The two nonparametric damage detection methods can be used together to alert a bridge owner when a bridge becomes damaged and to aid the owner in assessing damage severity prior to inspection.

The following conclusions are drawn from this research:

1. Girder distribution factors collected under operational strain monitoring during single vehicle crossings are a robust indicator of bridge performance and potential damage.
2. A damage index based on a nonparametric comparison of sample medians can provide an effective bridge damage alert system while controlling for both the Type I and Type II errors.
3. Three-dimensional statistical bridge signatures based on GDFs at each quantile of a probability distribution provide a successful nonparametric tool for identifying damage and a useful aid for damage localization and assessment.

The hypothesis testing framework introduced here for bridge damage detection enables a rigorous, decision-oriented approach for detection of bridge damage when it exists. Importantly, this can be accomplished while simultaneously assuring that false alarms are not issued, preventing expensive bridge repairs when they are unnecessary.

8. Acknowledgments

The authors are grateful for the funding of this research by the National Science Foundation Partnerships for Innovation (PFI) Program under Grant No. 0650258. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Additionally, the authors would like to thank MassDOT and the town of Barre for access to the Powder Mill Bridge, and Fay Spofford and Thorndike Inc. for access to design calculations and drawings. The authors would also like to thank Geocomp Corporation for extensive help with implementation and upkeep of the DAQ system. Finally, the authors are grateful for the use of the calibrated finite element model created by previous Tufts graduate student John Phelps.

9. References

- [1] American Society of Civil Engineers. (2013). "2013 Report Card for America's Infrastructure.", accessed 7/03/2013, <http://www.infrastructurereportcard.org/a/#p/bridges/conditions-and-capacity>.
- [2] Federal Highway Administration. (2010). 2010 Status of the Nation's Highways, Bridges, and Transit: Conditions and Performance.
- [3] Moore, M., B. Phares, B. Graybeal, D. Rolander, and G. Washer. (2001). Reliability of Visual Inspection for Highway Bridges: Federal Highway Administration.
- [4] American Association of State Highway and Transportation Officials. (2010). "AASHTO LRFD Bridge Design Specifications." 5th ed.
- [5] Ghosn, M., F. Moses, and J. Gobieski. (1986). "Evaluation of Steel Bridges using in-Service Testing." Transp. Res. Rec. 1072: 71-78.
- [6] Eom, J. and A. Nowak. (2001). "Live Load Distribution for Steel Girder Bridges." Journal of Bridge Engineering 6 (6): 489-497.
- [7] Harris, Devin K. (2010). "Assessment of Flexural Lateral Load Distribution Methodologies for Stringer Bridges." Engineering Structures 32 (11): 3443-3451.
- [8] Cardini, A. J. and J. T. DeWolf. (2009). "Long-Term Structural Health Monitoring of a Multi-Girder Steel Composite Bridge using Strain Data." Structural Health Monitoring 8 (1): 47-58.
- [9] Stallings, J. and C. Yoo. (1993). "Tests and Ratings of Short-Span Steel Bridges." Journal of Structural Engineering 119 (7): 2150-2168.
- [10] Chakraborty, S. and J. DeWolf. (2006). "Development and Implementation of a Continuous Strain Monitoring System on a Multi-Girder Composite Steel Bridge." Journal of Bridge Engineering 11 (6): 753-762.
- [11] Kim, S. and A. Nowak. (1997). "Load Distribution and Impact Factors for I-Girder Bridges." Journal of Bridge Engineering 2 (3): 97-104.
- [12] Plude, S. (2011). "Implementing a Long-Term Bridge Monitoring Strategy for a Composite Steel Girder Bridge." Master's Thesis, University of Connecticut.
- [13] Wipf, T., B. Phares, L. Greiman, D. Hemphill, J. Doornink, and P. Lu. (2006). Remote Continuous Evaluation of a Bridge Constructed using High-Performance Steel: Iowa Department of Transportation.
- [14] Kim, Y., R. Tanovic, and R. Wight. (2009). "Recent Advances in Performance Evaluation and Flexural Response of Existing Bridges." Journal of Performance of Constructed Facilities 23 (3): 190-200.
- [15] Shenton, H. and X. Hu. (2006). "Damage Identification Based on Dead Load Redistribution: Methodology." Journal of Structural Engineering 132 (8): 1254-1263.

- 809 [16] Catbas, F. Necati, Melih Susoy, and Dan M. Frangopol. 2008. "Structural Health
810 Monitoring and Reliability Estimation: Long Span Truss Bridge Application with
811 Environmental Monitoring Data." *Engineering Structures* 30 (9): 2347-2359.
- 812 [17] Howell, D. and H. Shenton III. (2006). "A System for in-Service Strain Monitoring of
813 Ordinary Bridges." *Structures Congress* 2005: 1-7.
- 814 [18] Whelan, M. and K. Janoyan. (2009). "Design of a Robust, High-Rate Wireless Sensor
815 Network for Static and Dynamic Structural Monitoring." *Journal of Intelligent Material*
816 *Systems and Structures*. 20(7), 849-864.
- 817 [19] Teixeira de Freitas, S., H. Kolstein, and F. Bijlaard. (2012). "Structural Monitoring of a
818 Strengthened Orthotropic Steel Bridge Deck using Strain Data." *Structural Health*
819 *Monitoring* 11 (5): 558-576.
- 820 [20] Follen, C., M. Sanayei, B. Brenner, and R. Vogel. (2014). "Statistical Bridge
821 Signatures." *Journal of Bridge Engineering* 19(7), 04014022, January 2014. DOI:
822 10.1061/(ASCE)BE.1943-5592.0000596
- 823 [21] Rytter, A. (1993). "Vibrational Based Inspection of Civil Engineering Structures."
824 Department of Building Technology and Structural Engineering, Aalborg University,
825 Denmark.
- 826 [22] Olund, J. and J. DeWolf. 2007. "Passive Structural Health Monitoring of Connecticut's
827 Bridge Infrastructure." *Journal of Infrastructure Systems* 13 (4): 330-339.
- 828 [23] Sanayei, M., J. Phelps, J. Sipple, E. Bell, and B. Brenner. (2012). "Instrumentation,
829 Nondestructive Testing, and Finite-Element Model Updating for Bridge Evaluation using
830 Strain Measurements." *Journal of Bridge Engineering* 17 (1): 130-138.
- 831 [24] Wisconsin Department of Transportation and Federal Highway Administration. (2001).
832 Hoan Bridge Forensic Investigation Failure Analysis Report.
- 833 [25] Chajes, M., D. Mertz, S. Quiel, H. Roecker, and J. Milius. (2005). "Steel Girder Fracture on
834 Delaware's I-95 Bridge Over the Brandywine River." In , 1-10: American Society of Civil
835 Engineers.
- 836 [26] Kaufmann, E., R. Connor, and J. Fisher. (2004). Failure Analysis of the US 422 Girder
837 Fracture: ATLSS Reports.
- 838 [27] Farrar, C. and D. Jauregui. (1998). "Comparative Study of Damage Identification
839 Algorithms Applied to a Bridge: I. Experiment." *Smart Materials and Structures* (7): 701.
- 840 [28] Enright, M. and D. Frangopol. (2000). "Survey and Evaluation of Damaged Concrete
841 Bridges." *Journal of Bridge Engineering* 5 (1): 31-38.
- 842 [29] Miller, T., M. Chajes, D. Mertz, and J. Hastings. (2001). "Strengthening of a Steel Bridge
843 Girder using CFRP Plates." *Journal of Bridge Engineering* 6 (6): 514-522.
- 844 [30] Zwerneman, F., A. West, and K. Lim. (1993). "Fatigue Damage to Steel Bridge
845 Diaphragms." *Journal of Performance of Constructed Facilities* 7 (4): 207-224.
- 846 [31] Reiff, A. (2014). "Bridge Structural Health Monitoring using Nonparametric Statistical
847 Damage Detection and Advanced Load Rating Methods." Master's Thesis, Tufts University.

- [32] Federal Highway Administration. (2002). Corrosion Costs and Preventive Strategies in the United States.
- [33] Warhus, J., J. Mast, and S. Nelson. (1995). "Imaging Radar for Bridge Deck Inspection." SPIE 2456: 296-305.
- [34] Vogel, R. M., A. Rosner, and P. H. Kirshen. (2013). "Brief Communication: Likelihood of Societal Preparedness for Global Change: Trend Detection." *Nat.Hazards Earth Syst.Sci.* 13 (7): 1773-1778.
- [35] Rosner, A., R.M. Vogel and P.H. Kirshen, A Risk-Based Approach to Flood Management Decisions in a Nonstationary World, (2014) *Water Resources Research*, 50, doi:10.1002/2013WR014561.
- [36] Wilcoxon, Frank. (1945). "Individual Comparisons by Ranking Methods." *Biometrics Bulletin* 1 (6): 80-83.
- [37] Blair, R. Clifford and James J. Higgins. (1980). "A Comparison of the Power of Wilcoxon's Rank-Sum Statistic to that of Student's t Statistic Under various Nonnormal Distributions." *Journal of Educational Statistics* 5 (4): 309-335.
- [38] Neave, H. R. and C. W. J. Granger. (1968). "A Monte Carlo Study Comparing various Two-Sample Tests for Differences in Mean." *Technometrics* 10 (3): 509-522.
- [39] Helsel, D. R. and R. M. Hirsch. (2002). In *Statistical Methods in Water Resources Techniques of Water Resources Investigations*, Book 4, Chapter a3, 522 pages.
- [40] Efron, B. and R. Tibshirani. (1993). *An Introduction to the Bootstrap*.
- [41] Vogel, R. and N. Fennessey. (1994). "Flow-Duration Curves. I: New Interpretation and Confidence Intervals." *Journal of Water Resources Planning and Management* (120)4, 485–504.

872 List of Figures

873 Figure 1: Powder Mill Bridge

874 Figure 2: PMB Cross Section

875 Figure 3: Typical event in northbound lane

876 Figure 4: Typical event in southbound lane

877 Figure 5: Traffic event with vehicles travelling in opposite directions

878 Figure 6: Summary of GDFs for Cases A-D and U

879 Figure 7: Damage detection hypothesis test decision matrix

880 Figure 8: Detectable damage level for $\alpha = 5\%$ and $\beta = 5\%$

881 Figure 9: Simulation to determine DI values that indicate damage

882 Figure 10: Girder 1 baseline bridge signature envelope

883 Figure 11: Baseline bridge signature envelope

884 Figure 12: Signatures for four damage cases: (a) interior girder fracture, (b) fascia girder
885 corrosion, (c) diaphragm fractures, and (d) deck delamination.

886

887

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

		<u>Truth</u>	
		H_0 Bridge is Not Damaged	H_A Bridge is Damaged
<u>Decision</u>	Conclude Bridge is Not Damaged	No Error $1 - \alpha$ (No Damage)	Type II Error β (Under-preparedness)
	Conclude Bridge is Damaged	Type I Error α (Over-preparedness)	No Error $1 - \beta$ (Damage)

Figure 7

Figure 8

Figure 9

Figure 10

Figure 11

Figure 12

Tables

Table 2: Minimum detectable damage levels for $\alpha = 5\%$ and $\beta = 5\%$

	Girder 1	Girder 2	Girder 3	Girder 4	Girder 5	Girder 6
Damage Level Detected (Δ GDF)	0.013	0.011	0.008	0.008	0.005	0.004
Minimum percent change (Δ GDF/GDF)	6.4 %	3.5 %	3.1 %	5.2 %	8.8 %	N/A

Table 3: Summary of damage indices for sample damage cases

	G1	G2	G3	G4	G5	G6	Damage Index
Case U: Undamaged	0	0	0	0	0	0	0.00
Case A: Interior girder fracture	1	1	1	1	1	0	0.83
Case B: Fascia girder corrosion	1	1	1	0	0	0	0.50
Case C: Diaphragm fractures	1	1	1	1	1	0	0.83
Case D: Deck delamination	1	1	1	0	1	0	0.67

Table 4: Summary of Damage Index

		Damage Index Range	Probability of Damage	Type I Error Probability	Type II Error Probability	Action
●	GREEN	0 - 0.16	Low	-	-	No action required
●	YELLOW	0.17 - 0.49	Medium	12%	0%	Monitor closely and inspect when possible
●	RED	0.5 - 1.0	High	3%	0-10%	Inspect bridge