Research Article

TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Transportation Research Record
1–11

DOI: 10.1177/0361198120920631
journals.sagepub.com/home/trr
SAGE

Comparison between a Linear Regression and an Artificial Neural Network Model to Detect and Localize Damage in the Powder Mill Bridge

Kathryn Kaspar[1], Erin Santini-Bell[1], Marek Petrik[2], and Masoud Sanayei[3]

Abstract

This paper evaluates the ability of two different data-driven models to detect and localize simulated structural damage in an in-service bridge for long-term structural health monitoring (SHM). Strain gauge data collected over 4 years is used to characterize the undamaged state of the bridge. The Powder Mill Bridge in Barre, Massachusetts, U.S., which has been instrumented with strain gauges since its opening in 2009, is used as a case study, and the strain gauges used in this study are located at 26 different stations throughout the bridge superstructure. A linear regression (LR) model and an artificial neural network (ANN) model are evaluated based on the following criteria: (a) the ability to accurately predict the strain at each location in the undamaged state of the bridge; (b) the ability to detect simulated structural damage to the bridge superstructure; and (c) the ability to localize simulated structural damage. Both the LR and the ANN models were able to predict the strain at the 26 stations with an average error of less than 5%, indicating that both methodologies were effective in characterizing the undamaged state of the bridge. A calibrated finite element model was then used to simulate damage to the Powder Mill Bridge for three damage scenarios: fascia girder corrosion, girder fracture, and deck delamination. The LR model proved to be just as effective as the ANN model at detecting and localizing damage. A recommended protocol is thus presented for integrating data-driven models into bridge asset management systems.

Structural health monitoring (SHM) has become an increasingly important field in asset management. SHM allows for remote monitoring of infrastructure, which could include instrumentation to measure and store data in relation to temperature, pH levels, humidity, strains, accelerations, or deflections. It is essential, however, for bridge owners to translate SHM data into useable information for making decisions about the service life and maintenance of the asset. In this paper, the Powder Mill Bridge (PMB) in Barre, Massachusetts, U.S., is used as a case study for developing a protocol for using strain gauge data for long-term monitoring of the structural health of a bridge. Since its opening in 2009, the PMB has been instrumented with strain gauges, among other SHM sensors (*1*). The strain readings from single-vehicle truck events collected from 2012 to 2016 are used to characterize the behavior of the undamaged state of the bridge. During each truck event, the strain readings are collected at 26 different stations on the PMB (*2*). Using 25 strain readings as inputs, mathematical models are trained to predict the output strain at the 26th station. Two different methodologies for developing the undamaged bridge model are compared: linear regression (LR) and artificial neural networks (ANN). LR relates the strain readings at all stations to one another using a linear function, while ANNs are capable of interpreting and characterizing nonlinear and complex relationships between the strains at the 26 stations.

The first step was to determine how well the two models could predict the strain at each location on the bridge under undamaged conditions. Both the LR and the ANN models were able to predict the strain at each station with an average error of less than 5%, thus

[1]Department of Civil and Environmental Engineering, University of New Hampshire, Durham, NH
[2]Department of Computer Science, University of New Hampshire, Durham, NH
[3]Department of Civil and Environmental Engineering, Tufts University, Medford, MA

Corresponding Author:
Kathryn Kaspar, kew235@wildcats.unh.edu

establishing confidence in the models' ability to characterize the bridge behavior under undamaged conditions. Following, a calibrated finite element model (FEM) in SAP2000 was used to simulate damage on the PMB. Three damage scenarios were considered: girder fracture, fascia girder corrosion, and deck delamination. Previous research done by Weinstein et al. showed that using ANNs it was possible to detect damage in a bridge, but localization was not possible for all scenarios (*2*). This research thus expanded the work of Weinstein et al. with the main contributions being (a) to compare the damage detection and localization ability of a LR model with an ANN model and (b) to provide a recommended approach for future implementation in bridge management systems (*2*).

Structural Performance Assessment using Physics-Based and Data-Driven Approaches

Background

In the field of SHM, there are two primary types of computer-based models used to analyze the structural health of infrastructure: physics-based models and data-driven models. Physics-based models involve the development of a structural model, such as an FEM, that considers the geometry, material properties, interactions between multiple bodies, and other system variables and uses that information to numerically assess possible behavior outcomes as a result of external forces or stresses (*3*). Data-driven models do not require *a priori* inputs; instead, large amounts of response data are used to learn and characterize the relationships between different components of a system. In this research, data-driven models are proposed as a reliable method for SHM, and a physics-based model was used to simulate various damage scenarios on the bridge. It is thus necessary to understand both approaches and their applications in the field of SHM.

Physics-Based Models in SHM

The research of Toksoy and Aktan represents an early application similar to how SHM is performed today: *in situ* testing results are analyzed in parallel with results from an FEM to understand the changes in the behavior of a structure over time (*4*). During the construction phase of the PMB, Sanayei et al. used strain gauge data from truck load tests on the bridge to calibrate and update an FEM for future SHM (*5*). Santini-Bell et al. and Sanayei et al. further expanded on the idea of using the calibrated FEM to generate objective load ratings of the PMB based on both the observed condition of the bridge from an inspection as well as the actual system response of the bridge under *in situ* load testing (*1*, *6*). Garcia-Palencia et al. also developed a methodology for calibrating the stiffness, mass, and damping parameters of the FEM to create a baseline behavior model of the bridge to be used for future damage detection (*7*). Research indicates that physics-based models are powerful at analyzing global and local structural behavior under various loading conditions as well as indicating the presence of damage, hence their continued use in the field of SHM (*8*, *9*). However, as modeling programs continue to change and improve, the practice of modeling itself has become more complex. Weber and Paultre discussed the intricacies of modeling and model updating of a truss tower to identify damage within the structure (*10*). Despite the power and versatility of physics-based models, these models can be cumbersome to construct, require a deep understanding of the system before implementation, and must be continuously calibrated and updated to reflect changing conditions on the bridge, leading to the rise in popularity of data-driven models (*9*).

Data-Driven Models in SHM

Data-driven models rely on statistical and mathematical interpretations of data. One of the simplest models used in statistical analysis is linear regression (LR). In an LR analysis, a model is trained to determine optimal coefficients relating input variables to an output variable that result in the lowest prediction error over the training dataset (*11*). Many relationships can be characterized through an LR analysis. For example, Seo et al. used an LR approach and strain gauge readings as inputs to develop a protocol for assigning objective load ratings to steel bridges (*12*). LR is one of the simplest methods for characterizing the relationship between multiple variables, but it may not be applicable when relationships between variables are nonlinear. During the past few decades, numerous prediction methods capable of characterizing nonlinear variable relationships have been developed in the fields of statistics and machine learning (*13*).

Machine learning is a subfield of artificial intelligence that uses data to find patterns and make predictions. Many machine learning methods build on and extend an LR approach. In this research, ANNs were selected as a methodology for implementing machine learning because of their ability to fit both linear and nonlinear behavior. Hornik et al. showed that ANNs can be used to approximate virtually any function, thus making them an ideal framework to study the behavior of complex infrastructure (*14*). In an SHM application, Smarsly et al. developed a small-scale, four-story test structure in a laboratory setting with one accelerometer at each story level (*15*). Using one ANN at each location, simulated

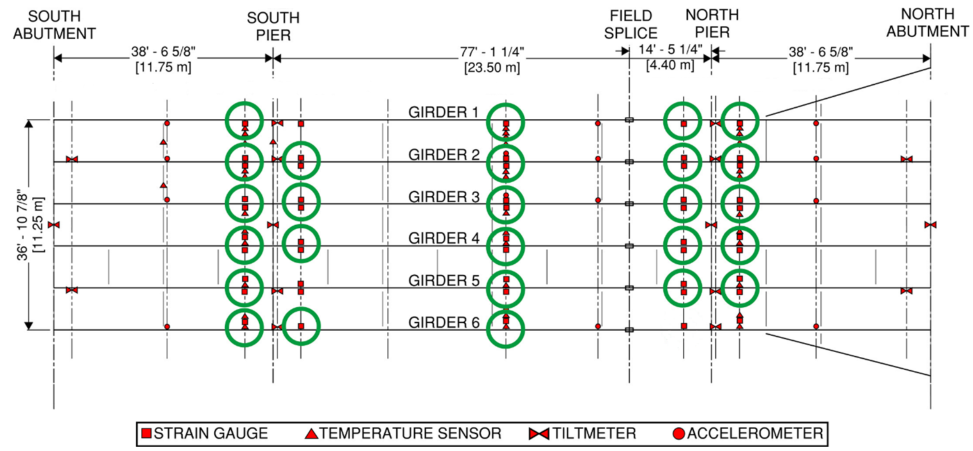

Figure 1. Instrumentation setup on the Powder Mill Bridge (PMB). Adapted from Weinstein et al. (*2*).

bridge damage was detected and localized, using various ANN properties over multiple runs. In lieu of accelerometer data, Kudva et al. used strain readings generated by an FEM and trained ANNs located on a 4x4 bay to detect and localize damage (*16*). The analyses by Smarsly et al. and Kudva et al. are perhaps the most similar to the work done by Weinstein et al. (*2*, *15*, *16*). However, instead of using strain readings generated by an FEM, Weinstein et al. used measured strain data from the PMB to train ANNs at various locations on the bridge and found that damage was detected but not localized for all scenarios (*2*). To address the lack of localization, an alternative to the ANN methodology as a means of localizing damage in all scenarios was explored, thus providing an approach for integrating a data-driven model into a bridge SHM system.

Case Study: Powder Mill Bridge (PMB)

The PMB in Barre, Massachusetts is a three-span steel-girder bridge connecting Vernon Ave to Massachusetts Route 122 over the Ware River. Six steel girders are in composite action with a continuous concrete deck slab spanning 47 m long by a width of 12.7 m at the south and center spans and increasing to a width of 19 m at the north abutment. Since its opening in 2009, PMB has been instrumented with 100 strain gauges, 66 temperature sensors, 16 biaxial tiltmeters, and 16 accelerometers as shown in Figure 1. Details in relation to the placement and types of SHM instrumentation on the PMB are discussed in Sanayei et al. and Santini-Bell et al. (*1*, *5*). The green circles represent the 26 stations where the strain readings were extracted for this research.

PMB is adjacent to a waste management transfer station, thus heavy trucks inducing a live load on the bridge are frequent. The strain readings from single-vehicle, heavy truck loads, here referred to as *truck events*, were used as the inputs to the data-driven models. All truck events were single-vehicle truck events with a maximum positive strain at the midspan greater than 39 microstrain centered on either the southbound or the northbound lanes of the PMB. By using only single-vehicle, heavy truck events in this analysis, the behavior of the structure under similar loading conditions is analyzed and thus changes in the behavior that may indicate damage are more likely to be detected. Weinstein et al. also performed a sensitivity analysis showing that the simulated damage effects are sufficiently representative regardless of the transverse location of the truck within each vehicle lane (*2*). A description of the program developed to process the truck events can be found in Weinstein et al. (*2*). The post-processed strain reading thus represents the change in strain as a result of the load induced by the single truck events.

Strain gauges were used for measuring the strain from each truck event at a sampling rate of 50 Hz (0.2-s intervals) to a data acquisition box located underneath the bridge (*2*). From this dataset, only the maximum strains were extracted for each location as well as the corresponding strains at the other 25 stations to ensure a high signal-to-noise ratio. For example, Figure 2 shows the corresponding strains at 25 stations throughout the bridge when the station Girder 1 Station 2 (G1-S2) reaches its maximum strain of approximately -20 microstrain for a sample truck event. The blue star represents the target output strain while the other 25 strain readings marked by circles represent the input strains for the model. The goal of the model is to predict the output strain at a single location given only the input strains at the 25 other locations. As a result, an individual model

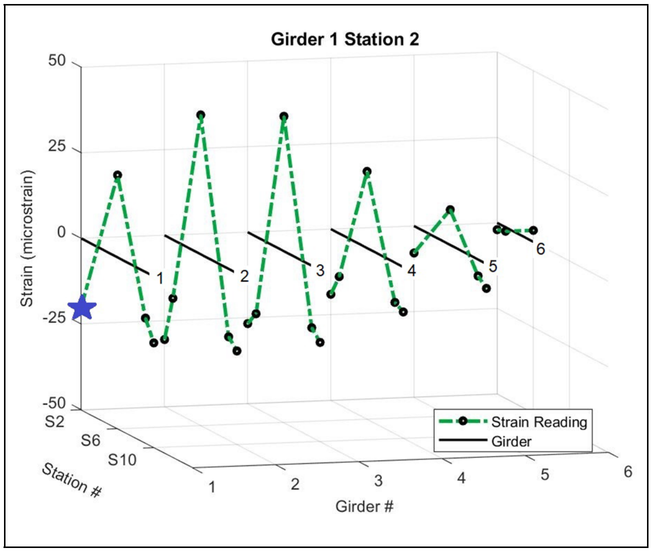

Figure 2. Extraction of the maximum output strain at location G1-S2 and the corresponding input strains at the other 25 locations.

was developed for each location on the bridge, thus totaling 26 different models to characterize the entire bridge. Throughout this paper, the term *individual model* refers to the model trained at one specific location whereas the *final model* refers to the cumulative results of the models at all 26 stations. The final LR and ANN models were evaluated in their ability to (a) accurately predict the strain at each station under undamaged conditions, (b) detect simulated bridge damage, and (c) localize simulated bridge damage.

Preliminary Model Testing

A total of 1,929 truck events collected over 4 years were used in this research and randomly divided to be used for two purposes: 1,509 truck events were used in the training phase and 420 truck events were set aside to run the preliminary model and damage tests, which was the same breakdown used by Weinstein et al. for comparison (*2*). MATLAB was first used to determine the coefficients and parameters for both the LR and ANN models that resulted in the lowest prediction errors over the 1,509 truck events. Each individual model was then tested against the 420 truck events to determine the prediction errors of the trained model.

Damage Detection

The models were trained to predict the outcome at each location under undamaged conditions. However, this research focuses on the final model's ability to detect and localize bridge damage in the structure. The hypothesis is that, under damaged conditions, the prediction errors of the model will be larger than in the undamaged state. To test this hypothesis, several damage scenarios were developed and simulated using an FEM in SAP2000. Details in relation to the development and calibration of the FEM can be found in Sanayei et al. (*5*). Once the model was calibrated and validated, various simulated bridge damage scenarios were applied to the structure.

Reiff et al. identified three damage scenarios that typical bridges are susceptible to and provided references to bridges displaying these damage conditions (*17*). Case 1 represents varying degrees of fascia girder corrosion, whereby the exposed exterior girders experience corrosion from drainage, rain, and other environmental factors. This was modeled by reducing the elastic modulus of the web of the girder by 5%, 10%, and 15%. Case 2 is the most critical structurally, with a girder fracture simulated at the midspan of Girder 2. To simulate the girder fracture, the elastic modulus was reduced to close to zero at the midspan. Case 3 represents deck delamination in the southbound lane, which was modeled by reducing the elastic modulus of the deck concrete by 35% in the southbound lane. It was estimated that the maintenance cost of deck repairs was equivalent to the combined total of all other maintenance costs on a typical bridge, emphasizing the prevalence and persistence of deck delamination (*18*). Early detection of damage can thus reduce maintenance costs and increase safety of bridges.

For each bridge damage scenario, as well as in the undamaged state, a single simulated HS20 truck was run over the FEM centered on both the northbound and the southbound directions (*2*). The peak strain at all 26 different locations was identified for each damage scenario and the 25 strains at the corresponding locations were also extracted. To determine the effects of the simulated bridge damage on the FEM, the percent difference between the strains in the undamaged state and the strains in each damage scenario was compiled for both the southbound and northbound directions. The percent difference was then applied to the truck event with undamaged conditions to create a new truck event with strain readings reflecting the bridge simulated damage.

The simulation process was repeated to simulate damage on multiple truck events. A truck event that has been modified through this process will here be referred to as a *truck event with simulated damage conditions*, while a truck event that has been unchanged will be referred to as a *truck event with undamaged conditions*. For each truck event with undamaged or simulated damaged conditions, the final LR and ANN models are used to predict the strain at all stations on the bridge and the prediction errors of the model are compiled. It is anticipated that the model will yield higher prediction errors when predicting the strains at each station for the truck events with simulated damage conditions.

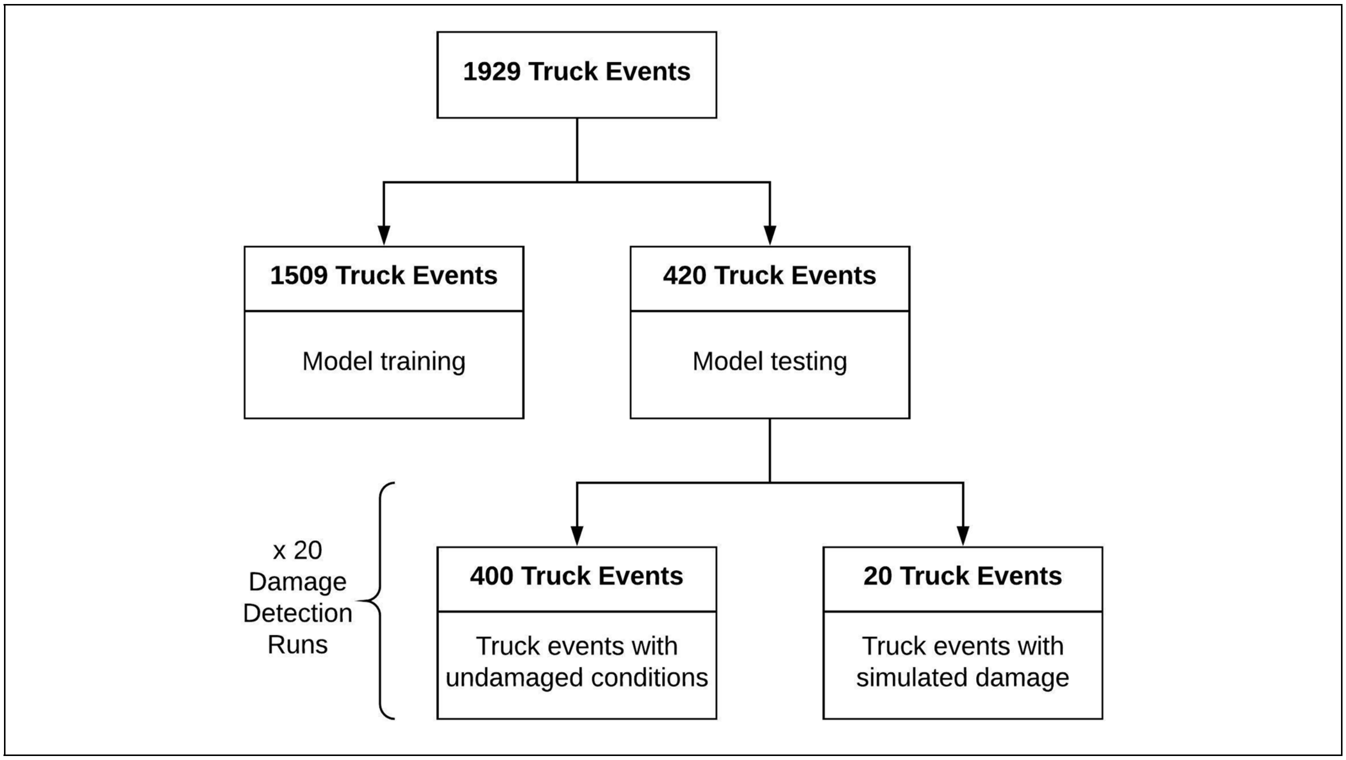

Figure 3. Data breakdown used for both the linear regression and artificial neural network models.

Damage is simulated on the truck events in the testing dataset. A total of 20 bridge damage detection computer runs were performed in this research, each containing a unique breakdown of 20 truck events with simulated damage conditions and 400 truck events with undamaged conditions as shown in Figure 3. For each damage detection run, the models were tested to determine if the model could detect a difference between the prediction errors of the 400 truck events with undamaged conditions compared with the prediction errors from 20 truck events with simulated damage conditions. Only 20 truck events with simulated damage conditions were used, as this represents the approximate number of heavy, single-vehicle truck events that occur on the PMB in one week. To ensure that each bridge damage detection run contained a unique set of 20 truck events with simulated damage conditions, a 20-fold cross-validation pattern was used to allocate which truck events remained undamaged and which truck events were used for simulating damage. Cross validation involves selecting the truck events with simulated damage conditions in a rotating pattern for each damage detection run rather than randomly allocating events into the damaged and undamaged categories, so as to test 400 unique truck events with simulated damage conditions over the course of the 20 damage detection runs (*11*).

Damage detection is here defined as the ability of the trained models to detect a difference between the prediction errors of the truck events with undamaged bridge conditions and the prediction errors of the truck events with simulated damage conditions using a Wilcoxon rank-sum test. A Wilcoxon rank-sum test is a method used to determine if two sets of data are statistically different. This test was selected because of the ability to compare two datasets of different sizes (*19*). There are two types of prediction errors in this analysis: a Type I error would be false damage detection in the undamaged case, while a Type II error would be a failure to identify damage. The null hypothesis of the Wilcoxon rank-sum test is that the prediction errors from the 400 truck events with undamaged conditions and the prediction errors from the 20 truck events with simulated damage conditions have the same median and variance at all locations while assuming a normal distribution (*20*). The test generates a p-value that represents a value between 0 and 1 that evaluates the null hypothesis, with a low p-value indicating that the null hypothesis should be rejected. Using a significance level of 0.1%, Weinstein et al. noted that no Type I or Type II errors occurred during the damage detection analysis, thus the same significance level was implemented in this research (*2*). A p-value less than the significance level of 0.001 thus indicated with 99.9% certainty that the prediction errors of the two datasets were different, and thus damage was detected by the model at that station. The p-values from each damage detection run were evaluated and the

Table 1. Description of Bridge Damage Scenarios and Corresponding Criteria to Determine If Damage was Effectively Localized

Damage scenario	Description	Localization criteria
Case 1a	Fascia Girder Corrosion: 5% reduction of the elastic modulus of the web of Girder 1 across the entire length of the span	Largest prediction error located on Girder 1
Case 1b	Fascia Girder Corrosion: 15% reduction of the elastic modulus of the web of Girder 1 across the entire length of the span	Largest prediction error located on Girder 1
Case 1c	Fascia Girder Corrosion: 25% reduction of the elastic modulus of the web of Girder 1 across the entire length of the span	Largest prediction error located on Girder 1
Case 2	Girder 2 Fracture: A section of Girder 2 at the midspan is altered to have an elastic modulus close to zero	Largest prediction error located at the midspan of Girder 2
Case 3	Half Deck Delamination: 35% reduction in the elastic modulus of the deck concrete across the southbound deck lane	Largest prediction error located on Girder 1 or Girder 2

percentage of damage detection runs with detected damage was evaluated for each station.

Damage Localization

Damage localization here refers to the models' ability to identify the location on the bridge at which the damage scenario was simulated, both longitudinally and transversely. For example, Case 1 consists of corrosion applied to Girder 1. Damage is considered to be effectively localized for this damage scenario if the largest prediction errors generated by the final model are located at any of the stations on Girder 1. Table 1 shows a summary of the damage scenarios as well as the localization criteria used.

Results

Preliminary Model Testing

The first criterion evaluated was the ability of the LR and ANN models to predict the strain at each station under undamaged conditions. The average prediction error of all 26 stations compiled from the 420 testing events was 3.9% for the final LR model and 2.7% for the final ANN model. The results indicate that both models were able to reliably predict the strain at each station under the undamaged scenario. Having low prediction errors in the training phase provides the user with more certainty in the model if significant prediction errors occur when using the model for damage detection purposes.

Damage Detection Results

A Wilcoxon rank-sum test was performed for each damage detection run at each station to determine the final LR and ANN models' ability to detect damage. If the *p*-value was less than the significance level, then damage was considered to be detected for that damage detection run. The percentage of the 20 damage detection runs for which damage was detected is shown in Table 2. Damage detection was evaluated for three bridge damage scenarios as well as an undamaged scenario, Case U, where all 420 events represent the undamaged bridge.

The results indicate that no stations detected bridge damage in any of the damage detection runs when tested against undamaged truck events, thus the final LR model did not produce Type I errors. For comparison, the final ANN model also did not produce any Type I errors. The results also indicate that most sensors detected damage in at least one of the 20 damage detection runs with multiple stations detecting damage in all 20 damage detection runs. For most of the damage scenarios, damage was more easily detected at Stations 4, 6, and 8. This is likely because of the higher strain readings at these stations compared with Stations 2 and 10, so the changes because of the simulated damage are more easily distinguished from the noise. Similarly, external Girders 1 and 6 did not detect damage as readily as some of the other girders despite Girder 1's proximity to the simulated damage. This could be because the traffic lanes were centered above Girders 2–5, so the strain readings were lower along the exterior girders. With the simulated 5%, 15%, and 25% fascia girder corrosion, 100% detection rates were found at 4, 8, and 11 different stations respectively using the final LR model. Using the final ANN model, damage was detected for all damage detection runs at 3, 12, and 11 different stations respectively for the three stages of fascia girder corrosion, indicating similar results. For the Case 2 girder fracture, 12 stations detected damage in all damage detection runs using the final LR model compared with 10 stations using the final

Table 2. Percentage of Damage Detection Runs with Detected Damage at Each Station for the Final Linear Regression Model

	Percentage of runs with damage detected					
	Undamaged	Girder 1 corrosion			Girder 2 fracture	Deck delamination
Sensor location	Case U	Case 1a 5%	Case 1b 15%	Case 1c 25%	Case 2	Case 3
G1-S2	0	25	90	100	0	60
G1-S6	0	10	20	70	10	30
G1-S8	0	15	55	95	100	0
G1-S10	0	100	100	100	0	0
G2-S2	0	20	95	100	95	10
G2-S4	0	100	100	100	100	100
G2-S6	0	100	100	100	100	0
G2-S8	0	40	45	50	100	0
G2-S10	0	80	100	100	0	10
G3-S2	0	25	60	95	0	0
G3-S4	0	60	100	100	100	95
G3-S6	0	100	100	100	100	5
G3-S8	0	40	70	100	100	0
G3-S10	0	90	100	100	20	0
G4-S2	0	65	95	95	95	95
G4-S4	0	75	100	100	95	100
G4-S6	0	0	0	0	100	0
G4-S8	0	0	25	70	100	95
G4-S10	0	0	0	10	30	40
G5-S2	0	15	60	70	100	85
G5-S6	0	30	50	60	100	0
G5-S8	0	10	75	90	10	100
G5-S10	0	0	90	95	5	75
G6-S2	0	0	5	60	75	0
G6-S4	0	0	0	0	100	55
G6-S6	0	5	20	40	90	5
Number of stations with 100% detection	0	4	8	11	12	3

ANN model. Lastly, both the final LR model and the final ANN model were able to detect the Case 3 deck delamination in all damage detection runs at three different stations. It can thus be said with 99.9% confidence that both final models were able to detect damage at more than one station on the bridge for all damage scenarios, with the two final models generating similar detection capabilities.

Damage Localization Results

Damage localization was assessed individually for the 20 truck events with simulated damage conditions in all 20 damage detection runs, thus totaling 400 truck events with simulated damage conditions. It was hypothesized that the model would generate the largest prediction errors at the location of simulated damage as a result of the change in structural behavior. For example, Figure 4 shows the prediction errors at each location along the bridge when the final LR model was tested using a northbound truck event with simulated Case 2 damage

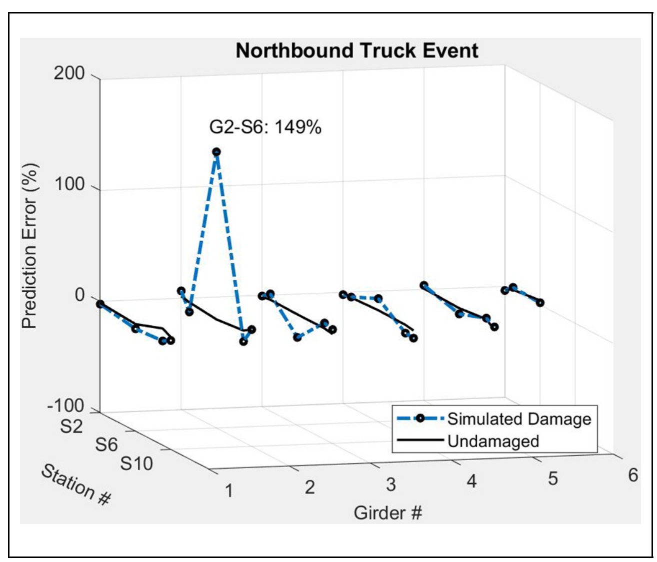

Figure 4. Prediction error of the final linear regression model for a Case 2 northbound truck event with simulated damage conditions.

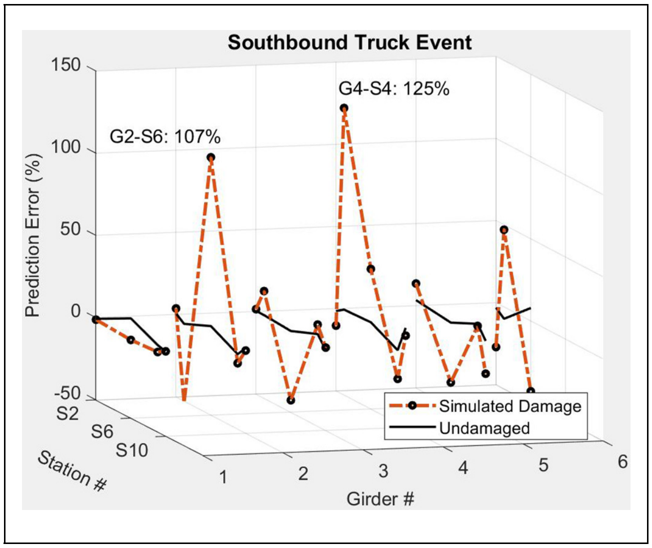

Figure 5. Prediction error of the final linear regression model for a Case 2 southbound truck event with simulated damage conditions.

conditions. The prediction errors of a sample northbound truck event under undamaged conditions have also been included for comparison. Because the simulated damage was induced at the midspan of Girder 2, damage was considered to be effectively localized for this truck event with simulated damage conditions. Conversely, Figure 5 shows the largest prediction error occurring on Girder 4 for the same damage scenario using a southbound truck event, thus damage was not effectively localized.

The final LR and ANN models were thus evaluated to determine the percentage of the 186 northbound events and 214 southbound events for which the model was able to localize each damage scenario, as shown in Table 3. The results from the southbound and northbound truck events were analyzed separately to note the effects of the direction of the vehicle on the models' ability to localize damage. The results indicate that, regardless of the mathematical model used, the northbound truck events were effectively localized while the southbound truck events were not. The following section investigates further into the possible reasons behind this distinction. Ultimately, these results stress the importance of separating the truck events based on direction of traffic.

Discussion

The results indicate success in simulated damage detection for all scenarios and damage localization for northbound truck events. Because both the LR and the ANN models were not able to localize damage effectively for the southbound truck events with simulated damage conditions, the localization criteria were further scrutinized.

Table 3. Percentage of Northbound and Southbound Truck Events Where Simulated Bridge Damage Was Effectively Localized

Damage scenarios	Linear regression		Artificial neural network	
	Northbound	Southbound	Northbound	Southbound
Case 1a	98.9%	11.7%	87.1%	2.8%
Case 1b	98.9%	0.0%	84.9%	0.0%
Case 1c	98.9%	0.0%	82.3%	0.0%
Case 2	98.4%	52.8%	96.8%	0.5%
Case 3	97.3%	0.0%	98.9%	0.5%

The process for simulating damage involved extracting the strains at all stations on the bridge during a truck event for both the undamaged and damaged conditions. The percent difference in strain between the undamaged and damaged conditions was determined, and this percentage was applied to actual truck event strains to simulate truck events with damage conditions. It was initially hypothesized that the stations closest to the simulated damage experienced the largest changes in strain as a result of the damage. However, further inspection showed this only to be true for northbound truck events, thus resulting in better localization results. For the southbound events, the location of the largest change in strain generated by the FEM varied for each damage scenario and did not always align with the localization criteria selected. This is reflected in the results as shown in Figure 5, where large prediction errors can be seen at the location of simulated damage (G2-S6) as well as at other stations on the bridge where a change in structural behavior is noted by the final LR model. Further inspection showed that the FEM also noted significant changes in strain at these locations in addition to at the midspan of Girder 2. This could be as a result of load redistribution on the bridge: as the southbound lane undergoes simulated damage, the adjacent girders are also subjected to increased loading. Furthermore, the indeterminacy of the bridge could also contribute to increased strains at additional stations as well as at the location of simulated damage when testing the damage scenarios in the southbound lane. While damage was not effectively localized using the localization criteria used, the final LR and ANN models were able to indicate the critical locations at which a change in structural behavior occurred. It was thus possible to localize damage using engineering judgment to interpret changes in the bridge structural behavior.

Proposed Management Strategy

Applicability

PMB, a three-span steel girder bridge, served as the case study and thus the application of this research is targeted

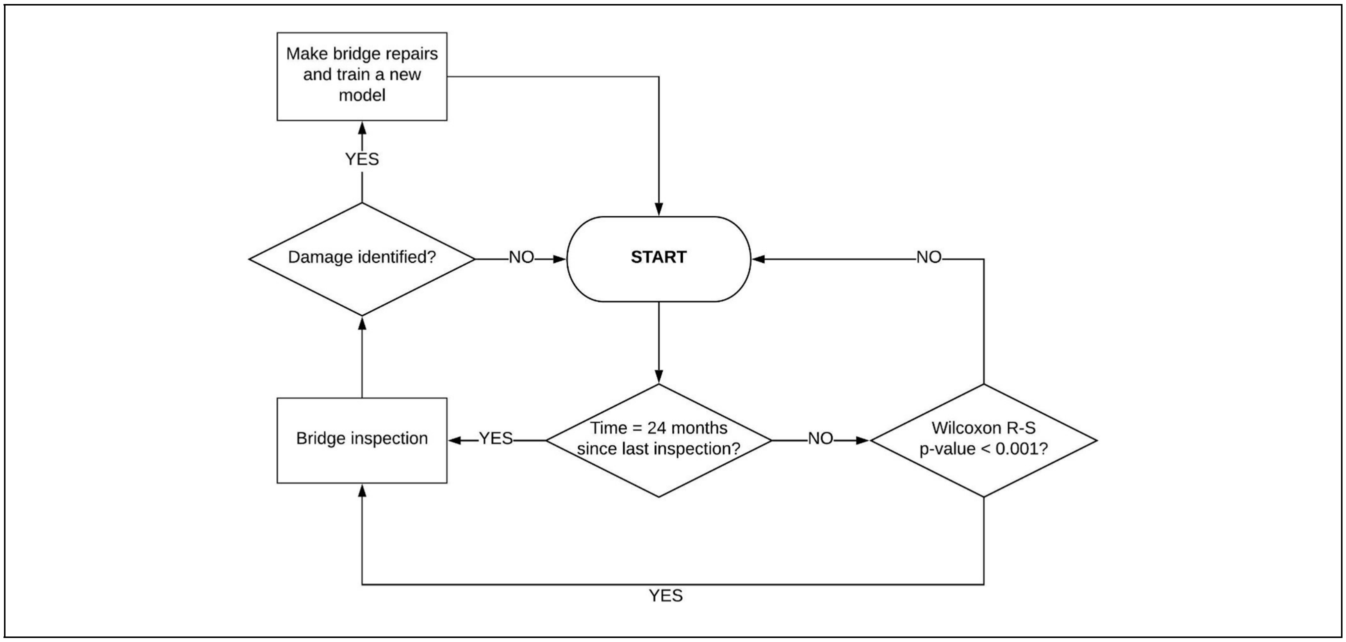

Figure 6. Sample flow chart for integrating a data-driven model into bridge inspection protocol.

at similar bridges. Girder bridges are the oldest and most common bridge type in the world and currently make up more than half of all state bridges in New Hampshire (*21*). PMB is located adjacent to a waste management transfer station in a low-traffic area, and thus a typical loading event consists of a single-vehicle, heavy truck event. This type of loading condition is not uncommon: examples include isolated bridges servicing coal mines in West Virginia, timber harvesting operations in Maine, or any other kind of manufacturing plant or distribution center with heavy shipments.

Both mathematical models implemented in this research were trained using 1,509 truck events compiled over 4 years to characterize the undamaged state of the bridge. Strain readings were captured at four or five different locations per girder at the critical locations for a three-span bridge, and the models were tested using approximately one week's worth of truck events. It is thus necessary to have an instrumentation system installed at the inception of the bridge with enough strain gauges and truck events to capture the bridge's behavior. The critical assumption that is made when training a data-driven model for this application is that the trained model represents the undamaged state of the bridge. If a bridge with existing damage were instrumented with strain gauges and a data-driven model were trained under the guidelines provided by this paper, the model would not be able to detect the existing damage in the bridge. It would, however, likely be able to detect any future changes in the structural behavior. The process for how a model could be trained and continuously updated over the life of the bridge is subsequently discussed.

Proposed Monitoring Strategy

Once a trained LR or ANN model of the bridge has been developed, the model can then be tested at periodic intervals to determine if damage is detected. Bridge inspection routinely occurs every 24 months; depending on how much data is available, the trained model could be tested every week, for example, which was the time frame used in this study. Using one week's worth of truck events, a Wilcoxon rank-sum test can be performed to test for a statistical difference between the prediction errors from the new events compared with the prediction errors from the original testing dataset. If damage is detected and repairs are made to the bridge, a new model will need to be trained to represent the new, assumed healthy state of the bridge. Older versions of the trained model could be used to analyze small, long-term changes in structural behavior such as creep and bridge bearing deterioration or to retest the bridge after repairs have been made; more research is needed to confirm these hypotheses. Figure 6 represents a sample decision-making flow chart illustrating how a data-driven model can be integrated within existing inspection protocol for long-term bridge monitoring.

Conclusion

When comparing the LR and ANN methodologies, the final LR model proved to be just as reliable as the final

ANN model at predicting the strain at each station on the bridge under undamaged conditions as well as detecting and localizing damage under simulated damage conditions. This is remarkable given the simplicity and ease of use of LR. Approximately 1,500 truck events were sufficient to characterize the behavior of the bridge under undamaged conditions with average prediction errors of approximately 5% at each station for both final models. Damage was detected in all 20 damage detection runs when testing the final models with one week of truck events with simulated damage conditions. Both final models were also able to identify the locations at which changes in structural behavior occurred, though these locations did not always correspond to the localization criteria used.

Compared with the ANN, the trained LR model is perhaps of more value to the engineer. Once the model has been trained, each variable is given a coefficient that relates all stations on the bridge to one another. This information could be useful in interpreting current and future behavior of the bridge. Deciphering the relationship between the different variables in an ANN is more complicated and requires a deep understanding of the mathematical foundations of ANN. Because of its simplicity and power, LR is recommended as an initial approach for model training. The results in this research suggest that a trained LR model with low prediction errors of approximately 5% should be able to detect and localize changes in structural behavior of the bridge caused from fascia girder corrosion, girder fracture, and deck delamination. Should the LR model generate large prediction errors in the preliminary testing phase, it is possible that the ANN methodology could be used to train an appropriate model.

Finally, engineering judgment is critical for interpreting information about the current and future states of the system. However, this research has shown the power that data-driven models can offer as a remote monitoring tool for infrastructure management.

Future Work

The results from this research are promising for both the future of PMB as well as other bridges nationwide. Further analysis should be done to identify the number of strain gauges needed to detect damage in a structure. For example, a full sensitivity analysis could be performed to better understand how many strain gauges are needed and the location of the gauges to fully characterize the bridge behavior. A sensitivity analysis could also be performed to determine the minimum number of truck events used in the testing dataset to detect damage without generating many false positives. When selecting truck events for the training and testing of the models, the data collected over 4 years was purposefully randomized. However, an analysis could be done to determine the impact of seasonal and diurnal temperature changes on the strain readings. Long-term monitoring of the PMB using this model would allow for confirmation of the recommended protocol and generate information in relation to the type and severity of damage that the model is able to detect and localize.

The next step in this research is to test this same methodology on a different bridge. This paper focuses on bridges with heavy, single-vehicle truck events, but other SHM instrumentation such as accelerometers or tiltmeters could be used as the input data to the model detecting damage over time. This methodology could also be used for localized damage detection. The Memorial Bridge constructed in Portsmouth, New Hampshire in 2009 became one of just a few bridges in the world to implement gusset-less bridge connections (*22*). Because of the lack of inspection protocol and data in relation to these connections, training a mathematical model to detect damage could be especially advantageous for long-term monitoring.

Acknowledgments

The research team is grateful to Tufts University, in particular Masoud Sanayei and Jordan Weinstein for sharing information in relation to the data collected at the Powder Mill Bridge.

Author Contributions

The authors confirm contributions to the paper as follows: study conception and design: K. Kaspar, E. Santini-Bell, M. Petrik, M. Sanayei; data collection: M. Sanayei; analysis and interpretation of results: K. Kaspar, M. Petrik, E. Santini-Bell, M. Sanayei; draft manuscript preparation: K. Kaspar, M. Petrik, E. Santini-Bell. All Authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This material is based on work partially supported by the National Science Foundation under Grant No. 0620258 and funding from the NHDOT Research Advisory Council.

References

1. Santini-Bell, E., P. J. Lefebvre, M. Sanayei, B. R. Brenner, J. D. Sipple, and J. Peddle. Objective Load Rating of a Steel-Girder Bridge using Structural Modeling and Health

Monitoring. *Journal of Structural Engineering*, Vol. 139, No. 10, 2013, pp. 1771–1779. http://doi.org/10.1061/(ASCE)ST.1943-541X.0000599.

2. Weinstein, J. C., M. Sanayei, and B. R. Brenner. Bridge Damage Identification using Artificial Neural Networks. *Journal of Bridge Engineering*, Vol. 23, No. 11, 2018, p. 04018084. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001302.
3. Kakadiaris, I. A. *Physics-Based Modeling, Analysis and Animation*. Technical Report No. MS-CIS-93-45. University of Pennsylvania Department of Computer and Information Science, 1993.
4. Toksoy, T., and A. E. Aktan. Bridge-Condition Assessment by Modal Flexibility. *Experimental Mechanics*, Vol. 34, No. 3, 1994, pp. 271–278. https://doi.org/10.1007/BF02319765.
5. Sanayei, M., J. E. Phelps, J. D. Sipple, E. Santini-Bell, and B. R. Brenner. Instrumentation, Nondestructive Testing, and Finite-Element Model Updating for Bridge Evaluation using Strain Measurements. *Journal of Bridge Engineering*, Vol. 17, No. 1, 2012, pp. 130–138. http://doi.org/10.1061/(ASCE)BE.1943-5592.0000228.
6. Sanayei, M., A. J. Reiff, B. R. Brenner, and G. R. Imbaro. Load Rating of a Fully Instrumented Bridge: Comparison of LRFR Approaches. *Journal of Performance of Constructed Facilities*, Vol. 30, No. 2, 2016, p. 04015019. http://doi.org/10.1061/(ASCE)CF.1943-5509.0000752.
7. Garcia-Palencia, A. J., E. Santini-Bell, J. D. Sipple, and M. Sanayei. Structural Model Updating of an In-Service Bridge using Dynamic Data. *Structural Control and Health Monitoring*, Vol. 22, No. 10, 2015, pp. 1265–1281. http://doi.org/10.1002/stc.1742.
8. Kim, C., and M. Kawatani. Pseudo-Static Approach for Damage Identification of Bridges Based on Coupling Vibration with a Moving Vehicle. *Structure and Infrastructure Engineering – Maintenance, Management, Life-Cycle Design and Performance*, Vol. 4, No. 5, 2008, pp. 371–379. https://doi.org/10.1080/15732470701270082.
9. Shahidi, G. *Physics-Based and Data-Driven Methods with Compact Computing Emphasis for Structural Health Monitoring*. Doctoral dissertation. Lehigh University, 2016.
10. Weber, B., and P. Paultre. Damage Identification in a Truss Tower by Regularized Model Updating. *Journal of Structural Engineering*, Vol. 136, No. 3, 2010, pp. 307–316.
11. James, G., D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer, New York, N.Y., 2013.
12. Seo, J., T. M. Czaplewski, J. Kimn, and G. Hatfield. Integrated Structural Health Monitoring System and Multi-Regression Models for Determining Load Ratings for Complex Steel Bridges. *Measurement*, Vol. 75, 2015, pp. 308–319. http://dx.doi.org/10.1016/j.measurement.2015.07.043.
13. Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, Cambridge, 2006.
14. Hornik, K. M., M. Stinchcombe, and H. White. Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks*, Vol. 2, 1989, pp. 359–366.
15. Smarsly, K., K. Dragos, and J. Wiggenbrock. Machine Learning Techniques for Structural Health Monitoring. *Proc., 8th European Workshop on Structural Health Monitoring (EWSHM)*, Bilbao, Spain, 2016.
16. Kudva, J. N., N. Munir, and P. W. Tan. Damage Detection in Smart Structures Using Neural Networks and Finite-Element Analyses. *Smart Materials and Structures*, Vol. 1, 1992, pp. 106–112.
17. Reiff, A. J., M. Sanayei, and R. M. Vogel. Statistical Bridge Damage Detection using Girder Distribution Factors. *Engineering Structures*, Vol. 109, 2016, pp. 139–151.
18. Lee, S. K. Current State of Bridge Deterioration in the U.S. – Part 2. *Materials Performance*, Vol. 51, 2012, pp. 2–7.
19. Mann, H. B., and D. R. Whitney. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, Vol. 18, 1947, pp. 50–60.
20. Sugiyama, M. *Introduction to Statistical Machine Learning*. Morgan Kaufmann, Cambridge, Mass., 2016.
21. New Hampshire Department of Transportation. *Bridge Program Recommended Investment Strategy*. New Hampshire Department of Transportation, 2018.
22. University of New Hampshire. *The Living Bridge Portsmouth Memorial Bridge: Project Overview*. https://livingbridge.unh.edu/. Accessed June 21, 2018.

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.