Environmental Effects on the Identified Natural Frequencies of the Dowling Hall Footbridge

Peter Moser¹ and Babak Moaveni²

Abstract

Continuous monitoring of structural vibrations is becoming increasingly common as sensors and data acquisition systems become more affordable, and as system and damage identification methods develop. In vibration-based structural health monitoring, the dynamic modal parameters of a structure are usually used as damage-sensitive features. The modal parameters are often sensitive to changing environmental conditions such as temperature, humidity, or excitation amplitude. Environmental conditions can have as large an effect on the modal parameters as significant structural damage, so these effects should be accounted for before applying damage identification methods. This paper presents results from a continuous monitoring system installed on the Dowling Hall Footbridge on the campus of Tufts University. Significant variability in the identified natural frequencies is observed; these changes in natural frequency are strongly correlated with temperature. Several nonlinear models are proposed to represent the relationship between the identified natural frequencies and measured temperatures. The final model is then validated using independent sets of measured data. Finally, confidence intervals are estimated for the identified natural frequencies as a function of temperature. The ratio of observed outliers to the expected rate of outliers based on the confidence level can be used as a damage detection index.

¹ Assistant Engineer, Fay, Spofford & Thorndike, Burlington, MA; Former Graduate Student, Tufts University.
² Assistant Professor, Dept. of Civil and Environmental Engineering, Tufts University, Medford, MA. E-mail: babak.moaveni@tufts.edu, tel.: 617-627-5642, fax: 617-627-3994 (corresponding author).
Keywords

Structural health monitoring; vibration monitoring; temperature effects on natural frequencies; system identification; automatic modal analysis; damage detection

1. Introduction

Major structural failures in recent years have brought the need for improved infrastructure monitoring and maintenance to public attention [1]. The nation’s roadways include more than 600,000 bridges [2] which are normally monitored by visual inspections. In many cases, accurate assessment of a bridge’s performance is not really possible by these visual inspections. The American Society of Civil Engineers issued its Report Card for America’s Infrastructure in 2009 [3], giving bridges an overall grade of “C.” A key statistic determining this grade is the number of bridges classified as either “structurally deficient” or “functionally obsolete.” Nationwide, more than 26% of bridges fall into at least one of these classifications. Structural health monitoring (SHM) can help maintain and improve the transportation system by providing accurate, timely, and objective information about the condition of bridges [4]. Continuous vibration monitoring is one strategy for SHM. In such a monitoring system, the vibration responses of a bridge due to traffic, wind, and other sources of ambient excitation are measured. From the measured response, the dynamic modal parameters (natural frequencies, damping ratios, and mode shapes) of the structure can be determined using output-only modal analysis techniques. The identified modal parameters can then be used for condition assessment of the structure. A typical damage scenario involves loss of stiffness in some portion of the structural system and a corresponding change in the modal parameters.

Vibration-based SHM methods often rely on changes in the modal parameters for damage identification. Outside the laboratory, however, modal parameters are usually sensitive to other factors
as well as damage. Temperature affects the Young’s modulus of steel [5] and concrete [6]. The stiffness of asphalt has been shown to vary dramatically near the freezing point [7]. Chen and Virgin [8] present an analysis of significant changes in dynamic behavior as thermal loads approach buckling loads. Sohn [9] reviews several studies which indicate that boundary conditions change with temperature and that the boundary conditions in turn affect modal parameters. Large volumes of traffic can change a structure’s mass [10, 11] and thus its modal parameters. Interaction of the structure with surrounding air currents can produce changes in structural dynamics as wind speeds change [12]. In a nonlinear structure, dynamic properties change with excitation amplitude [13]. All these factors can influence the modal parameters of a bridge but temperature is the most commonly considered environmental variable [14]. Changes in natural frequency in the order of 10% from environmental or operational sources are not unusual [15, 16]. This presents a challenge for SHM methods because changes from environmental effects can be larger than changes due to significant damage [17]. If environmental effects are neglected, changes due to damage may be completely masked [9]. To accurately and reliably identify damage, these environmental effects must be accounted for. This can be done through development of a mathematical model correlating the modal parameters (e.g., natural frequencies) to environmental effects (e.g., temperature). Other approaches to account for environmental effects have also been studied and reported, namely using features sensitive to damage but insensitive to environmental variability [18] or eliminating the environmental effects without measuring the underlying environmental variables [19, 20, 21, 22, 23].

Temperature effects on natural frequencies have been accurately modeled using AutoRegressive models with eXogenous inputs (ARX models) [24, 16]. The ARX family of models can account for dynamics in a relationship by relating present outputs to past inputs and outputs [25]. This is ideal for representing dynamic properties of a structure when they depend (linearly) on the rate of change or trend
in temperatures as well as the present temperature [26]. One complication in the modeling of environmental effects is that the relationship between frequency and temperature may be nonlinear. For example, Peeters and De Roeck [24] observed a bilinear relationship between temperature and the natural frequencies of the Z24 Bridge, with a dramatic change in behavior near the freezing point. In climates where temperatures below freezing are common, the environmental model must be accurate both above and below freezing. Another class of methods to model the temperature-natural frequency correlations is based on statistical learning algorithms such as support vector regression technique [27].

This paper presents results from a vibration-based continuous monitoring system deployed on the Dowling Hall Footbridge at Tufts University in Medford, MA. The system includes eight accelerometers, ten thermocouples, data acquisition equipment, and fully automated modal identification programs developed by the authors. The nonlinear relationship between natural frequencies and temperatures is modeled using measured data taken from this system during the sixteen-week period beginning on January 5, 2010 and ending on April 25, 2010. The process of choosing temperature variables and model structure is highlighted and an accurate model is selected. This model is then used to establish confidence intervals with which future data may be screened for damage. The model is validated using independent sets of measured data selected by two different data-splitting strategies. Finally, the mathematical form of the model is examined to draw conclusions about the underlying process by which temperatures affect the natural frequencies of the bridge.

2. Continuous Monitoring of the Dowling Hall Footbridge

2.1. Bridge Structure

The Dowling Hall Footbridge, shown in Figure 1, is a pedestrian bridge located on the Medford, MA campus of Tufts University. The bridge is 44 m (144 ft) long and 3.7 m (12 ft) wide. This two-span
continuous steel frame bridge connects the main campus with the student services offices on the seventh floor of Dowling Hall. The concrete walkway is equipped with a snowflake sensor and a heating system to prevent ice buildup on the deck during winter months. More details about the Dowling Hall Footbridge can be found in [28]. A continuous structural monitoring system is installed on this bridge as described in [29]. The system is briefly reviewed here.

Figure 1. Dowling Hall Footbridge (left photo from [28] and right photo from <www.maps.bing.com>)

2.2. Continuous Monitoring System

In order to properly design the continuous monitoring system, several dynamic tests were performed during the spring of 2009. The objective of these tests was to determine the natural frequencies and mode shapes of the bridge and assess the level of bridge response amplitudes due to ambient excitation. More details about the preliminary dynamic tests performed on the Dowling Hall Footbridge can be found in [29]. Figure 2 shows the first six vertical mode shapes with their natural frequencies as identified from one test. Mode 1 (4.66 Hz in this example) and Mode 2 (6.21 Hz) showed vertical motion with anti-symmetric and symmetric shapes, respectively. Mode 3 (7.08 Hz) and Mode 4 (8.88 Hz) showed vertical-torsional motion. Mode 5 (13.23 Hz) and Mode 6 (13.62 Hz) showed vertical motion with one node in each span of the bridge. A vibration-based continuous monitoring system was
designed for the Dowling Hall Footbridge based on results of initial testing and installed on the bridge in November 2009.

Eight piezoelectric accelerometers were installed on the bridge to monitor vibrations. Figure 3a shows the layout of these accelerometers. Each accelerometer was mounted on the underside of the bridge using aluminum L-brackets. Ten type T thermocouples were installed on the bridge to measure temperatures. These sensors measure the temperature of the masonry piers at two locations, the temperature of the concrete deck at two locations, the temperature of the steel frame at four locations, and the temperature of the air at two locations. Figure 3b shows the layout of the thermocouples.
Thermocouples are designated “S” for steel, “C” for concrete, and “A” for air, and their temperature readings are referred to as “Ts”, “Tc”, and “Ta”, respectively.

![Image of sensor layout](image)

Figure 3. Sensor layout: (a) accelerometers, (b) thermocouples

On-site data acquisition is performed by a National Instruments cRIO-9074 integrated chassis/controller. The monitoring program on the cRIO continuously samples the acceleration channels at a 2048 Hz sampling rate while temperatures are recorded at a rate of one sample per second. A 5-minute data sample is recorded every hour or when triggered by large vibration amplitudes. The program also performs file and memory management, automatic error recovery, and system status messaging. Communication with the data acquisition system occurs using a wireless bridge connected to the campus network. New data is retrieved each hour by a computer in the Civil and Environmental Engineering Department at Tufts University via File Transport Protocol (FTP). Data processing is performed after data is transferred from the cRIO to the processing computer. Data processing includes filtering (bandpass between 2 and 55 Hz), down-sampling (from 2048 to 128 Hz), and other data cleaning tasks as described in [29].

2.3. Automatic System Identification
The modal parameters of the structure are extracted from the measured ambient vibration data using the data-driven Stochastic Subspace Identification (SSI-Data) algorithm. This output-only system identification method was developed by Van Overschee and De Moor [30]. A reference-channel-based version of the algorithm [31] is implemented here. SSI-Data begins by creating “past” and “future” Hankel matrices from the data record. The reference channels (1, 2, 3, 5, 6, and 7) are loaded into a “past” Hankel matrix of 96 block rows. Both reference and non-reference channels are loaded into a “future” Hankel matrix of 96 block rows. Note that the “future” data is part of the measured data; the word future signifies that it comes 96 samples after the “past” data sequence. SSI-Data computes the optimal prediction of the future data using the past data by projecting the row space of the future data into the row space of the past data. A linear system is realized in state-space based on the singular value decomposition of this projection. The modal parameters of the structure are determined from the realized state-space matrices.

In the application of SSI-Data, a system order must be specified for the linear model to be realized. If the system order specified is too low, some observable modes will go undetected. If the system order specified is too high, non-physical modes will be identified along with the physical ones [32]. One means of selecting a system order and eliminating the identified non-physical modes is the stabilization diagram [33]. In this strategy the modal analysis is performed at sequentially increasing system orders. Modes which correspond to the physical system generally have similar modal parameters at different orders. A mode is judged to be “stable” between different system orders if its estimated characteristics agree within set limits. The system order can be chosen to maximize the number of stable modes.

In this work, the identification is performed sequentially for state-space system orders of 2 to 96 (numbers of modes from 1 to 48). At each step, modes identified at the current system order are
compared with modes identified at the previous system order. If the frequency matches within 1%, the damping ratio matches within 30% (relative), and the Modal Assurance Criterion (MAC) [34] between the current mode shape and the mode shape identified at previous iteration is higher than 95%, the mode is judged to be “stable” between the two system orders. A mode which remains stable for seven successive system orders is considered to be a physical mode of the system. Since the damping of the bridge is known to be very low, modes with identified damping ratios higher than 2% or less than zero are also excluded. The best system order is then determined as the smallest order with the maximum number of physical modes of interest.

2.4. Environmental Effects on Natural Frequencies

Automatic modal analysis was completed for each of 2700 records from the sixteen-week period beginning on January 5, 2010 and ending April 25, 2010. The identification results were screened for outliers using the procedure described in [29]. The identified natural frequencies during this period are shown in Figure 4. A clear pattern of daily variation in the natural frequencies can be observed. Measured temperatures at selected channels during the same period are shown in Figure 5a while Figure 5b shows temperature data from three channels (C1 = pier, C2 = deck, S3 = steel) for the period between January 25 and February 8. Correlation between natural frequencies and temperatures is evident: higher natural frequencies generally occur at lower temperatures. To illustrate this correlation, the identified natural frequencies of the six identified modes are plotted versus the temperature of the northern abutment (C1) in Figure 6. From this figure it can be observed that: (1) the identified natural frequencies increase as temperatures decrease, (2) this effect is more significant when the temperatures are below the freezing point resulting in a nonlinear relationship, and (3) Modes 1, 3, and 4 show more clear correlation with temperature while Modes 2, 5, and 6 show more scatter. One explanation for the third observation is that Modes 2, 5, and 6 are identified with larger estimation uncertainties as seen in Figure
4. Statistics of the identified modal parameters during the considered 16 week period are reported in Table 1. The MAC values are computed between each identified mode shape and the average mode shape for that mode. The percent rate at which a mode is successfully identified is reported as “ID Rate”. During this period, natural frequencies of the six identified modes varied by 4 to 8 percent while temperatures ranged from -14 to 39 °C. Because of this inherent variability in the natural frequencies, a model of the relationship between temperature and natural frequency must be developed before the dynamic characteristics of the bridge can be used for condition assessment. It is worth noting that there are different sources of variability in the identified modal parameters such as relative humidity, amplitude of excitation due to wind velocity and/or pedestrian traffic, weight of pedestrians on the footbridge, and estimation uncertainty in the SSI-DATA system identification method. However, the number of pedestrians on the Dowling Hall Footbridge at one time is generally small; therefore the weight of pedestrians will have little contribution to the total variability of the identified modal parameters. In a study by He [16], it was found that changes in relative humidity of the air and the wind velocity have much smaller influence on the modal parameters of a highway overpass bridge in San Diego, CA than changes in the air temperature. The estimation uncertainty for different modes is different. Based on the observed modal contributions of the bridge response and the observed variability in the stabilization diagrams, vibration Modes 1, 3, and 4 are expected to be identified with smaller estimation uncertainties than Modes 2, 5, and 6.

2.5. Environmental Effects on Damping Ratios and Mode Shapes

The identified damping ratios from the sixteen-week period between January 5 and April 25 were examined for variations due to environmental effects. Figure 7 shows the damping ratios versus the temperature of the western abutment C1. Significant scatter is seen in the identified damping ratios and no obvious pattern of variation is observed. It is concluded that the damping ratios are not identified
with enough precision to detect environmental effects. In a similar way, the identified mode shapes were examined for variations due to environmental effects; no obvious variation in mode shapes was observed. Similar observation has been found in other studies as well [16].

Figure 4. Variation of the natural frequencies of the first six modes between January 5 and April 25, 2010

Figure 5. Temperature variations: (a) January 5 to April 25, (b) January 25 to February 8, 2010
Table 1. Statistics of modal parameters identified between January 5 and April 25, 2010

<table>
<thead>
<tr>
<th>Mode</th>
<th>Natural Frequencies</th>
<th>Damping Ratios</th>
<th>Mode Shapes (MAC)</th>
<th>ID Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean [Hz] STD [Hz]</td>
<td>Mean [%] STD [%]</td>
<td>Mean [%] STD [%]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.68 0.03 4.61 4.81</td>
<td>0.29 0.23 0.01 1.76</td>
<td>99.4 1.6 85.1 100.0</td>
<td>84</td>
</tr>
<tr>
<td>2</td>
<td>5.99 0.08 5.76 6.21</td>
<td>0.42 0.29 0.03 1.95</td>
<td>99.4 1.4 86.1 100.0</td>
<td>92</td>
</tr>
<tr>
<td>3</td>
<td>7.16 0.10 6.97 7.51</td>
<td>0.34 0.22 0.02 1.92</td>
<td>99.4 1.4 85.4 100.0</td>
<td>98</td>
</tr>
<tr>
<td>4</td>
<td>8.93 0.06 8.81 9.18</td>
<td>0.16 0.13 0.01 1.80</td>
<td>99.8 0.9 86.7 100.0</td>
<td>99</td>
</tr>
<tr>
<td>5</td>
<td>13.18 0.08 12.94 13.58</td>
<td>0.49 0.22 0.05 1.57</td>
<td>95.9 3.2 85.2 99.9</td>
<td>71</td>
</tr>
<tr>
<td>6</td>
<td>13.71 0.15 13.29 14.34</td>
<td>0.65 0.27 0.09 1.99</td>
<td>97.3 2.8 85.1 100.0</td>
<td>88</td>
</tr>
</tbody>
</table>

3. Data Assessment

3.1. Correlation between Temperature Measurements

The continuous monitoring system includes ten temperature sensors (Fig. 3b). An environmental model should quantify the relationship between the identified natural frequencies of the bridge and these temperatures. However, since some temperature channels provide nearly identical information, a model which uses all ten temperature measurements as inputs is unnecessarily complicated. A subset of these measurements should be chosen which produces an accurate and simple model. For example,
Thermocouples S3 and S4 both measure the steel frame temperature on the north side of the bridge. Figure 8 shows these two channels between January 25 and February 8; the records were nearly indistinguishable. Using both of these measurements in the environmental model would provide redundant information and could lead to ill-conditioning in regression analysis. Therefore, only one of the two should be used.

![Figure 7. Damping ratios of the first six identified modes versus temperature at C1](image)

![Figure 8. Steel temperatures on the north side of the bridge (S3 and S4)](image)
To quantify the similarities between measured temperatures, the correlation coefficient of the time-history of each pair of temperature sensors was calculated. Thermocouples C1 and C4 both measured the temperature of a masonry pier and had a correlation coefficient of 99.2%. Thermocouples C2 and C3 both measured the temperature of the bridge deck and had a correlation coefficient of 99.5%. Thermocouples S1 and S2 measured the steel frame temperature at different locations on the south side of the bridge and had a correlation coefficient of 99.6%. Finally, all four of Thermocouples S3, S4, A1, and A2 had mutual correlation coefficients of at least 98.9%. Four sets of temperature measurements (\{C1 C4\}, \{C2 C3\}, \{S1 S2\}, and \{S3 S4 A1 A2\}) were therefore found having correlation coefficients of 98.9% or greater within each set. Furthermore, the identified sets \{S1 S2\} and \{S3 S4 A1 A2\} displayed significant dependence: all measurements between the two sets had a correlation coefficient of at least 92%. These observations suggested that either three or four variables should be used in the environmental model with no more than one variable from each set.

It is interesting to observe that the steel temperatures on the north side of the bridge and the air temperatures underneath the bridge were nearly equivalent. The steel frame on the north side of the bridge receives little direct sunlight and so the steel temperature varied in a pattern similar to the air temperature. In contrast, the south side of the bridge experiences direct sunlight for much of the day. Figure 9 shows the difference between the temperatures at S1 and S3. During the daytime hours the sensor on the south side (S1) was significantly warmer than the sensor on the north side (S3). During the night hours the sensors produced much closer measurements. Overall these two measurements displayed a high correlation coefficient (94.7% during the full sixteen-week period), but differences between the channels were noticeable at times when the south side of the bridge was exposed to direct sunlight.
3.2. **Principal Component Analysis of Temperatures**

The number of temperature measurements useful for modeling can also be estimated using Principal Component Analysis (PCA) [35]. First, the singular value decomposition (SVD) of the matrix of all temperature measurements ($T$) is formed as $T = U \Sigma V^T$. PCA can be used to generate a transformed set of variables (equal to $U \Sigma$) in which all columns are mutually orthogonal. In the present case the analysis was used solely to judge the appropriate number of variables. The matrix $\Sigma$ is diagonal and contains the singular values. The number of nonzero (or relatively large) singular values indicates the number of independent measurements.

Figure 10 shows the singular values of the temperature measurements. The first three singular values were seen to be significantly above zero; the fourth was also nonzero but smaller than the first three. The remaining six singular values were close to zero. Dividing the sum of the first three singular values by the sum of all ten singular values showed that the first three singular values represented 90.1% of the total sum; adding a fourth brought the ratio to 94.5%. These results agreed with the observations.
from correlation analysis that there were four non-dependent sets of temperature measurements and that three or four temperature variables should be used in the environmental model.

![Singular values from principal component analysis of all temperature measurements](image)

Figure 10. Singular values from principal component analysis of all temperature measurements

### 3.3. Natural Frequency-Temperature Relationship

Correlation coefficients and principal component analysis indicated that three or four temperature measurements should be used in the environmental model. Figure 11 shows the identified natural frequency of Mode 3 (f ≈ 7.2 Hz) versus a temperature from each of the four non-dependent sets. It was observed the identified natural frequencies of the bridge increased at colder temperatures with a very sharp increase near the freezing point. The effect was most pronounced for temperature C2 (the temperature of the concrete deck), suggesting that the concrete deck plays a major role in the stiffening of the structure at freezing temperatures. For all temperature sensors the relationship between identified natural frequencies and measured temperatures was nonlinear.
4. Numerical Modeling

4.1. Model Quality Metrics

Metrics are needed to quantify the goodness of fit between the data and various models and to select the optimal model structure. The easiest measure to consider is the estimation error variance $\sigma_e^2$, in which $e$ is the estimation error or residual of the model. A smaller error variance is an indicator of a closer fit. Note that least-squares regression generates unbiased estimators so the error variance and mean-square error are equal. Another common related statistic is the coefficient of determination $R^2$. Kvalseth [36] gives a definition of $R^2$ as:

$$R^2 = 1 - \frac{\sigma_e^2}{\sigma_d^2}$$  \hspace{1cm} (1)

in which $\sigma_d^2$ is the variance of the measured data. This statistic can be understood as the fraction of the total variance that is explained by the model. A higher $R^2$ is an indicator of a closer fit. The problem
with $\sigma^2_e$ and $R^2$ as measures of model quality is that they always improve with increasing model complexity and cannot distinguish between a good model and an over-fit model.

In order to compare a large number of models, it was desirable to use objective criteria which judge both the goodness of fit and the complexity of the model without performing pairwise hypothesis testing. Two criteria used for this purpose are Akaike’s Information Criterion (AIC) [37] and the Bayesian Information Criterion (BIC) [38]. Both of these metrics compare the magnitude of errors while including penalties for the number of parameters that must be estimated. In the case that the errors are normally distributed with zero mean and constant variance, the AIC and BIC have the following forms:

\begin{align}
AIC &= N \ln \left( \sigma^2_e \right) + 2p \\
BIC &= \ln \left( \sigma^2_e \right) + \frac{p}{N} \ln(N)
\end{align}

where $p$ represents the number of model parameters that must be estimated and $N$ represents the number of data points used in the regression. For both AIC and BIC, a lower value is an indicator of a better fit. For all the models evaluated, the estimated error standard deviation ($\sigma_e$), the $R^2$ value, the AIC, and the BIC are reported.

4.2. Model Structure Selection

The nonlinear shape of the natural frequency vs. temperature graphs (Fig. 11) suggested use of nonlinear polynomial models. Also, polynomial models have been used to accurately describe the stiffness of construction materials near freezing temperatures [7]. To aid in the selection of the input variables and the polynomial order, regression analysis was performed systematically for all possible combinations of: (1) polynomial model orders from first through fourth (without cross terms), (2) numbers of input temperatures varying from one to ten, and (3) all possible subsets of temperature
measurements from the available ten channels. Then at each model order and number of inputs, the subset of inputs which give the best $R^2$ and AIC were identified. Figures 12(a) and 12(b) show these best $R^2$ and AIC values (averaged across three modes as described below) as a function of the number of input variables and model order. Since Modes 1, 3, and 4 displayed the least scatter in identification (Figs. 4, 6), the quality of each fit was assessed by averaging the $R^2$ and AIC values of each regression for these three modes. Note that all the polynomials evaluated by this method were static models and that both natural frequency and temperature data were used in the model without normalization.

Several observations were made from Figure 12. First, adding more input variables in general improved the fit but the benefits diminished as the number of variables increased. After three or four variables were included, the improvement was quite small. Second, the fit improved as the model order increased. Specifically, the first-order model fit the worst and the fourth-order model fit the best. It should also be mentioned that extending the model to fifth-order often produced ill-conditioned solutions. These observations suggested the fourth-order polynomial model with three temperature inputs as a good candidate model.

![Figure 12](image)

Figure 12. Model quality metrics (a: $R^2$ and b: AIC) versus number of temperature measurements considered for models of order one to four.
4.3. **Input Variable Selection**

The procedure described above suggested a fourth-order polynomial model with three temperature inputs. After systematically performing regression on all possible combinations of temperature inputs for this model structure, the best results were obtained using the measurements S3 (steel temperature), C2 (deck temperature), and C1 (pier temperature). Notice that this method led to the choice of one variable from each of the independent sets identified earlier. It should be mentioned that replacing one of these temperatures with another from the same set resulted in very minor reduction in quality.

It was decided that temperatures from the set \{S1 S2\} should not be included in the environmental model. The basis for this decision was that: (1) the procedure above gave the best three variables from the other sets, (2) the inclusion of a fourth variable resulted in only marginal improvement in quality (Fig. 12), (3) the correlation between the sets \{S1 S2\} and \{S3 S4 A1 A2\} was in all cases at least 92%, and (4) natural frequency vs. temperature plots for this set exhibited more scatter than for the other sets (Fig. 11). Physically this is explained as follows: the difference between the sets \{S1 S2\} and \{S3 S4 A1 A2\} was mostly due to direct sunlight near the sensors on the south side of the bridge. This local effect appears to have had little impact on the global dynamic behavior of the structure. Based on all these considerations, a fourth-order polynomial model with three input variables (S3, C2, and C1) was tentatively selected. The selection of these three variables admits a clear physical interpretation: the stiffness of the frame, the stiffness of the deck, and the boundary conditions all contribute significantly to the dynamic behavior of this structure.

After this initial assessment, a large number of alternative models were evaluated in detail. These included numbers of temperature inputs from two through five, model orders one through five, models
including and excluding cross terms, a bilinear model, and an autoregressive exogenous (ARX) model. Not all these can be presented, but the most significant ones are shown here.

4.4. Estimated Models

4.4.1. Linear Polynomial Model

In the linear polynomial model each natural frequency of the bridge was assumed to be a linear combination of the temperatures plus a constant term and an error term $e$:

$$y_i' = \beta_0' + \sum_{k=1}^{n_{var}} x_{ik} \beta_k' + e_i' \quad \text{or} \quad Y = X\beta + e$$  \hspace{1cm} (4)

In which:

$$Y = \begin{bmatrix} y_1^1 & \cdots & y_{n_{mode}}^1 \\ \vdots & \ddots & \vdots \\ y_N^1 & \cdots & y_{n_{mode}}^N \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,n_{var}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \cdots & x_{1,n_{var}} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0^1 & \cdots & \beta_{n_{mode}}^1 \\ \vdots & \ddots & \vdots \\ \beta_0^{n_{var}} & \cdots & \beta_{n_{var}}^{n_{mode}} \end{bmatrix} \quad e = \begin{bmatrix} e_1^1 & \cdots & e_{n_{mode}}^1 \\ \vdots & \ddots & \vdots \\ e_N^1 & \cdots & e_{n_{var}}^{n_{mode}} \end{bmatrix}$$  \hspace{1cm} (5)

In these equations, $N$ represents the number of data points used in the regression, $n_{var}$ represents the number of temperature inputs used, and $n_{mode}$ represents the number of modes for which regression is performed. The subscript $i$ denotes the time index ($i=1,2,\ldots,2700$), the superscript $j$ represents the mode number ($j=1,2,\ldots,6$), and the subscript $k$ represents the temperature sensor ($k=1,2,3$). Each $x$ represents a temperature measurement and each $y$ represents an identified frequency; the values of $\beta$ are the coefficients of the model and are estimated from the data by a linear least-squares solution. It is worth
noting that the coefficients of models represented in the following sections are all estimated using linear least square solutions.

Table 2 shows goodness of fit metrics for this multiple linear regression using the full sixteen-week data record. The model had four parameters for each mode \( \beta_0, \beta_1, \beta_2, \beta_3 \). Figure 13 provides a visual evaluation of the fit for Mode 3 \((f \approx 7.2 \text{ Hz})\). Figure 13a compares the time-variation of the identified and simulated natural frequency of Mode 3 between January 25 and February 8, while Figure 13b shows the correlation between the identified and simulated natural frequency of Mode 3 for the entire sixteen-week period along with the straight line representing a perfect fit. Although the model produced an \( R^2 = 0.77 \) for Mode 3, a distinct S-shape can be seen in Figure 13b which indicates that the model does not accurately represent the relationship at extreme temperatures. This is not surprising since Figures 6 and 11 show that the natural frequency-temperature relationship was nonlinear.

![Figure 13](image_url)

Figure 13. (a) Time-variation of identified and linear model simulated natural frequency of Mode 3 between January 25 and February 8, 2010; (b) correlation between identified and simulated natural frequency for the entire sixteen-week period.
<table>
<thead>
<tr>
<th>Mode</th>
<th>Model</th>
<th>Linear Static</th>
<th>ARX</th>
<th>Quadratic</th>
<th>Bi-linear</th>
<th>4th Order with cross terms</th>
<th>4th Order no cross terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\sigma_e$ [Hz]</td>
<td>0.020</td>
<td>0.019</td>
<td>0.017</td>
<td>0.017</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.642</td>
<td>0.671</td>
<td>0.720</td>
<td>0.742</td>
<td>0.763</td>
<td>0.749</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>$10^3 x$</td>
<td>-17.8</td>
<td>-18.0</td>
<td>-18.3</td>
<td>-18.5</td>
<td>-18.6</td>
</tr>
<tr>
<td>2</td>
<td>$\sigma_e$ [Hz]</td>
<td>0.059</td>
<td>0.057</td>
<td>0.055</td>
<td>0.054</td>
<td>0.053</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.426</td>
<td>0.457</td>
<td>0.494</td>
<td>0.519</td>
<td>0.543</td>
<td>0.526</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>$10^3 x$</td>
<td>-14.0</td>
<td>-14.2</td>
<td>-14.3</td>
<td>-14.5</td>
<td>-14.5</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td></td>
<td>-5.65</td>
<td>-5.70</td>
<td>-5.77</td>
<td>-5.82</td>
<td>-5.78</td>
</tr>
<tr>
<td>3</td>
<td>$\sigma_e$ [Hz]</td>
<td>0.048</td>
<td>0.048</td>
<td>0.040</td>
<td>0.038</td>
<td>0.033</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.766</td>
<td>0.765</td>
<td>0.841</td>
<td>0.851</td>
<td>0.891</td>
<td>0.875</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>$10^3 x$</td>
<td>-16.0</td>
<td>-16.0</td>
<td>-17.1</td>
<td>-17.2</td>
<td>-18.0</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td></td>
<td>-6.06</td>
<td>-6.04</td>
<td>-6.43</td>
<td>-6.49</td>
<td>-6.73</td>
</tr>
<tr>
<td>4</td>
<td>$\sigma_e$ [Hz]</td>
<td>0.031</td>
<td>0.031</td>
<td>0.024</td>
<td>0.023</td>
<td>0.022</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.703</td>
<td>0.695</td>
<td>0.819</td>
<td>0.831</td>
<td>0.852</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>$10^3 x$</td>
<td>-18.6</td>
<td>-18.6</td>
<td>-20.0</td>
<td>-20.1</td>
<td>-20.4</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td></td>
<td>-6.95</td>
<td>-6.92</td>
<td>-7.44</td>
<td>-7.50</td>
<td>-7.56</td>
</tr>
<tr>
<td>5</td>
<td>$\sigma_e$ [Hz]</td>
<td>0.061</td>
<td>0.061</td>
<td>0.059</td>
<td>0.057</td>
<td>0.056</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.470</td>
<td>0.480</td>
<td>0.505</td>
<td>0.543</td>
<td>0.563</td>
<td>0.542</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>$10^3 x$</td>
<td>-10.7</td>
<td>-10.7</td>
<td>-10.8</td>
<td>-11.0</td>
<td>-11.0</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td></td>
<td>-5.57</td>
<td>-5.58</td>
<td>-5.63</td>
<td>-5.70</td>
<td>-5.64</td>
</tr>
<tr>
<td>6</td>
<td>$\sigma_e$ [Hz]</td>
<td>0.127</td>
<td>0.121</td>
<td>0.117</td>
<td>0.113</td>
<td>0.105</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.290</td>
<td>0.360</td>
<td>0.395</td>
<td>0.435</td>
<td>0.517</td>
<td>0.453</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>$10^3 x$</td>
<td>-9.8</td>
<td>-10.1</td>
<td>-10.2</td>
<td>-10.4</td>
<td>-10.7</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td></td>
<td>-4.11</td>
<td>-4.21</td>
<td>-4.26</td>
<td>-4.33</td>
<td>-4.40</td>
</tr>
</tbody>
</table>

4.4.2. **Linear ARX Model**

AutoRegressive models with eXogenous inputs (ARX models) have been used for environmental modeling of natural frequencies of structures with good results [24, 16]. ARX models have the form:

$$y^I_k = a_1^I y^I_{k-1} + \cdots + a_{n_y}^I y^I_{k-n_y} = b_1^I u^I_{k-n_y} + b_2^I u^I_{k-n_y-1} + \cdots + b_{n_u}^I u^I_{k-n_u-n_y-1} + e^I_k$$

(6)
in which \( \mathbf{u} \) is a vector of the inputs (temperatures in this case), the model coefficients are defined by the values of \( a \) and \( b \), and the integers \( n_u, n_h, \) and \( n_k \) describe the model order. A model of this form can account for dynamics in the data by relating the current output to previous inputs and outputs. Note that the equations are still strictly linear. An ARX model can be fit to data using a linear least-squares solution [25].

An ARX model with \( n_u = 4, n_h = 2, \) and \( n_k = 0 \) was fit to the data. The model order was selected to give the smallest value of the AIC using the “selstruc” command from the Matlab “System Identification” toolbox [39]. The model had ten parameters for each mode after data normalization. Data normalization was performed to give each variable zero mean and unit standard deviation, as in the work of Peeters and De Roeck [24]. The goodness of fit metrics are listed in Table 2. The linear ARX model performed only marginally better than the (static) linear regression even though it was more complex. A linear model could not accurately describe the nonlinear relationship seen here.

Linear ARX models have been used with good results in several recent studies. However, these studies have involved either structures which rarely experienced freezing (in which case the below-freezing data was simply not used for modeling) or structures which never experienced freezing at all. For the climate in which the Dowling Hall Footbridge is located, freezing temperatures occur for a large part of the year and could not be omitted from the model.

### 4.4.3. Quadratic Model

A quadratic model of the form:

\[
y_i^j = \beta_0^j + \sum_{k=1}^{n_u} x_{ik} \beta_k^j + \sum_{k=1}^{n_h} x_{ik}^2 \beta_{n_h+k}^j + e_i^j \tag{7}
\]
was fit to the natural frequency-temperature data. Notice that this second-order equation did not include any cross terms. The model had seven parameters for each mode. The goodness of fit metrics are listed in Table 2. The results were improved over the linear models but still did not accurately capture the relationship at extreme temperatures.

4.4.4. Bilinear Model

Before attempting higher-order polynomial models, a bilinear model was fit. In this model, the data was split into two sets. Data was assigned to one set if $T_{C2}$ was below 5 °C and to the other set if not. This approach was based on examination of Figure 11 and similar data for other modes: the sharpest division between below-freezing behavior and above-freezing behavior was seen for $T_{C2}$ and occurred near 5 °C. The relationship was approximately linear on either side of this division.

Each of the two data sets was fit to a separate linear model of the form given in Equation (4). This model had eight parameters for each mode. The goodness of fit metrics are listed in Table 2 and a comparison between the identified and simulated natural frequency of Mode 3 is displayed in Figure 14. Figure 14a compares the time-variation of the identified and simulated natural frequency of Mode 3 between January 25 and February 8, while Figure 14b shows the correlation between the identified and simulated natural frequency of Mode 3 for the entire sixteen-week period along with the straight line representing a perfect fit. This model performed better than the linear and quadratic models but better results were still desired. One problem with a simple bilinear model is that discontinuities can occur at the boundary between the two data sets unless continuity-enforcing constraints are placed on the least-squares solution. Indeed, several large jumps can be seen in Figure 14a.
4.4.5. 4th Order (with Cross Terms)

A fourth-order model of the form:

\[ y_i^j = \beta_0^j + \sum_{k=1}^{n_{uv}} x_{ik} \beta_k^j + \sum_{k=1}^{n_{uv}} \sum_{l-k} x_{il} x_{il} \beta_{kl}^j + \sum_{k=1}^{n_{uv}} \sum_{l-k} \sum_{m-l} x_{ik} x_{il} x_{lm} \beta_{klm}^j + \sum_{k=1}^{n_{uv}} \sum_{l-k} \sum_{m-l} \sum_{n-m} x_{ik} x_{il} x_{lm} x_{ln} \beta_{klmn}^j + e_i^j \]  

was fit to the natural frequency-temperature data. Notice that this model included all possible cross terms. The goodness of fit metrics are listed in Table 2. This model fit the data better than any of the previous models. However, this model had thirty-five parameters for each mode. Because of the large number of parameters used in the fourth-order cross-term model, it was considered too complicated to be practical.

4.4.6. 4th Order Model (without cross terms)

A fourth-order model of the form:
\[ y_i^j = \beta_0^j + \sum_{k=1}^{n_{\text{var}}} x_k \beta_k^j + \sum_{k=1}^{n_{\text{var}}} x_k^2 \beta_k^j + \sum_{k=1}^{n_{\text{var}}} x_k^3 \beta_k^j + \sum_{k=1}^{n_{\text{var}}} x_k^4 \beta_k^j + \epsilon_i^j \]  \hspace{1cm} (9)

was fit to the natural frequency-temperature data. Notice that this model did not include any cross terms.

This model had thirteen parameters for each mode. The goodness of fit metrics are listed in Table 2 and comparisons between the identified and simulated natural frequencies of the six identified modes are displayed in Figure 15. Figure 15a compares the time-variation of the identified and simulated natural frequency of all six modes between January 25 and February 8, while Figure 15b shows the correlation between the identified and simulated natural frequency of these modes for the entire sixteen-week period along with the straight line representing a perfect fit. For all modes the error standard deviation was significantly less than the standard deviation of the raw data (Table 1). For Mode 3, the standard deviation was reduced from 0.10 Hz to 0.035 Hz, a factor of nearly three. This is important because it allows a narrow confidence interval to be established for this mode. Mode 6 was modeled with the least accuracy with a reduction in standard deviation of only 26%. Other than the full fourth-order model (with cross terms), this model provided the best fit yet examined.

The fourth-order model with no cross terms was selected as the best practical model for the relationship between natural frequencies and temperatures. This final model has the structure of Equation (9) where \( n_{\text{var}} = 3 \) and \( n_{\text{mode}} = 6 \). This model was selected as the final model structure for all vibration modes. Separate evaluations for Modes 1, 3, and 4 that were the most accurately identified vibration modes resulted in the same model structure. However, different model structures provide a slightly better fit for natural frequencies of Modes 2, 5, and 6 that were identified with higher estimation uncertainty. In this study, the same model structure was selected for all six modes due to the fact that (1) the difference in the quality of fit of the alternative models for Modes 2, 5 and 6 is negligible, and (2) the same model structure provides the best fit to the natural frequencies of the more accurately identified
natural frequencies of Modes 1, 3, and 4. The coefficients $\beta$ are given in Table 3 for each mode and input variable. Notice that the higher-order terms have smaller coefficients as indicated by the multiplier in the right-most column of the table. For the coefficients of each polynomial order, several patterns can be seen. Coefficients are of a similar magnitude for all three temperatures inputs, with higher-frequency modes generally having larger values. For each input term, coefficients across all six identified modes also usually have the same sign. This is an expression of the similar shape of the natural frequency-temperature surfaces (see Fig. 6) for the six identified modes.

![Figure 15](image.png)

Figure 15. (a) Time-variation of identified and fourth-order model simulated natural frequencies of all six modes between January 25 and February 8, 2010; (b) correlation between identified and simulated natural frequencies for the entire sixteen-week period.

It is worth noting that obtained models in this study consider temperature as the sole source of variability in the identified natural frequencies while there are different contributing sources such as estimation error in system identification. Therefore, the predicted models will better correlate the natural frequency-temperature relationship when the variability in the identified natural frequencies due to other sources of variability/uncertainty is small. This is verified by having the best fit for Modes 1, 3, and 4 which are the same modes identified with the least scatter in identified natural frequency (Fig. 4). The
best fit is seen for Modes 1, 3, and 4, which are the same modes identified with the least scatter in identified natural frequency (Fig. 4). The highest value of $R^2$ is 87% for Mode 3. The plot of identified natural frequency versus simulated natural frequency for Mode 2 (Fig. 15) shows a cluster in which the data is above the simulated natural frequency. This corresponds to the scattered frequencies which are identified higher than the main trend observed for this mode in Figure 4. Modes 5 and 6 display more scatter in general with no obvious pattern. Overall, Figure 15 indicates that the model fits the data well.

Table 3. Fourth-order model coefficients

<table>
<thead>
<tr>
<th>Term</th>
<th>Mode</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.7 Hz</td>
<td>6.0 Hz</td>
</tr>
<tr>
<td>$T_{C2}$</td>
<td>-3.689</td>
<td>-2.454</td>
</tr>
<tr>
<td>$T_{C1}$</td>
<td>-1.604</td>
<td>-11.121</td>
</tr>
<tr>
<td>$T_{S3}^2$</td>
<td>0.853</td>
<td>4.128</td>
</tr>
<tr>
<td>$T_{C1}^2$</td>
<td>-1.199</td>
<td>-5.031</td>
</tr>
<tr>
<td>$T_{S3}^3$</td>
<td>0.875</td>
<td>0.857</td>
</tr>
<tr>
<td>$T_{C1}^3$</td>
<td>1.848</td>
<td>10.334</td>
</tr>
<tr>
<td>$T_{S3}^4$</td>
<td>-2.854</td>
<td>-4.138</td>
</tr>
<tr>
<td>$T_{C1}^4$</td>
<td>-5.050</td>
<td>-31.833</td>
</tr>
</tbody>
</table>

4.4.7. Confidence Intervals

The fourth-order model presented can be used to generate confidence intervals around simulated natural frequencies based on the error variance and the covariance of the temperature inputs. Assuming a
zero mean stationary Gaussian random process for the model error, the confidence interval for a single future observation [40] is computed as:

\[ \hat{y} \pm t_{\alpha/2,N-p} \sigma_e \sqrt{1 + \frac{1}{N-p} \left( \mathbf{x}' \mathbf{x} \right)^{-1} \left( \mathbf{x}' \mathbf{x} \right)^{1/2}} \text{ with } \hat{y} = \mathbf{x}' \hat{\beta} \]  

(10)

in which \( t \) represents the critical value of the student \( T \)-distribution, \( \alpha \) represents the desired significance level, \( N \) represents the number of samples used in the regression, \( p \) represents the number of parameters (including the constant term) in the model, \( \sigma_e \) represents the estimated error standard deviation, \( \mathbf{x} \) represents the input data in a row vector for the future observation, and \( \mathbf{X} \) represents the matrix of all data used in the regression. Note that the confidence interval is a function of \( \mathbf{x} \).

Figure 16 shows the identified and simulated natural frequencies simulated using the fourth-order model during the period between March 15 and March 29, together with the 95% confidence intervals for the natural frequencies. The confidence intervals are indicated by the gray shaded region, with the simulated natural frequency indicated by a darker gray line. Data points which fall outside the confidence interval are shown with a filled circle. Note that the width of the confidence interval compared to the total variability in natural frequency is an indicator of the accuracy of the model for each mode. Across the data shown, the simulated natural frequency and confidence intervals generally match the data, and outliers are scattered rather than concentrated. Another indicator of the model quality is the number of data points which fall outside the generated 95% confidence interval for them. Table 4 shows this fraction; for all modes it is close to the expected 5%. Deviations of this fraction from exactly 5% could be caused by non-Gaussian distribution of the residuals.

The three temperature measurements used in the environmental model are shown in Figure 17 for the same time period. For most of the period shown, the fluctuations in natural frequencies were small. On the nights of March 27 and 28 the temperature dropped below freezing, producing a sharp
increase in the natural frequencies. It is also interesting to observe that in late March the temperature of the concrete deck (C2) moved below freezing with the temperature of the steel frame (S3) whereas in the earlier portion of Figure 5a it remained significantly warmer. This is most likely because the heating system in the deck was turned off for the year sometime in March, allowing the concrete to freeze more easily at this time.

![Figure 16. Identified and simulated natural frequencies between March 15 and March 29 using the fourth order model without cross terms, together with the 95% confidence interval (CI) and the outliers.](image)

Table 4. Outliers in the fourth-order model

<table>
<thead>
<tr>
<th>Mode</th>
<th>Outlier Rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (4.7 Hz)</td>
<td>5.2%</td>
</tr>
<tr>
<td>2 (6.0 Hz)</td>
<td>6.5%</td>
</tr>
<tr>
<td>3 (7.2 Hz)</td>
<td>4.4%</td>
</tr>
<tr>
<td>4 (8.9 Hz)</td>
<td>5.9%</td>
</tr>
<tr>
<td>5 (13.2 Hz)</td>
<td>5.0%</td>
</tr>
<tr>
<td>6 (13.7 Hz)</td>
<td>6.0%</td>
</tr>
</tbody>
</table>
4.4.8. Residuals

Examination of the sequence of residuals (or errors, $e$) provides another way to evaluate the quality of the model’s fit. If the residuals are reasonably interpreted as white noise as assumed for Equation (10), this indicates that the model is accurately representing all the information in the data. If trends are observed in the residuals this indicates that more useful information can still be extracted from the data. The autocorrelation functions of the residuals can be useful in detecting trends which might be missed by visual inspection [24]. Since the autocorrelation of a Gaussian white noise process is a delta function, inspection of this function allows assessment of the “whiteness” of the residuals.

The residuals of the 4th order model without cross terms are shown in Figure 18a. The residuals are smallest for Modes 1, 3, and 4, confirming that the model fits these modes best. Mode 2 shows a set/cluster of residuals well above zero representing the frequencies which are identified above the main trend for this mode (similar to Fig. 4). The residuals of Mode 6 are the largest and the least like white noise. The autocorrelation functions of the residuals are shown in Figure 18b. Notice that each function displays a spike at the zero delay point with all other points closer to zero. However, other smaller peaks

Figure 17. Variation of the three temperature channels considered in the model between March 15 and March 29
are seen regularly in the autocorrelation functions indicating that there is some pattern of oscillation in the residuals. A nonlinear autoregressive model might account for the nonlinear dynamic relationship while eliminating some of the short-period patterns in the residuals [24], but it would also be much more complex. The current model captures most of the variation seen in natural frequencies, including nonlinearities, and is accurate enough to be useful. The long-period patterns observed in the residuals could be caused by unmeasured environmental variables such as variations in air humidity or ambient excitation amplitude.

Figure 18. Residuals between the identified and simulated natural frequencies of the six identified modes (a) and their autocorrelation functions (b)

4.5. Validation

A numerical model whose performance is only evaluated by the data to which it was fit can give overly optimistic estimates of its accuracy as noise in the input data is force-fit to noise in the output data [41]. A better test of a model’s validity is its ability to simulate data points which were not used in the fitting process. This type of model validation can be accomplished by splitting the available data
into a “modeling” set and a “validation” set. The model is fit using the modeling data and then its performance is evaluated using the validation set. Typically the first and larger portion of the data is used for modeling while a later and smaller portion is used for validation.

In the present study the bulk of the data was obtained in the winter or early spring. Using the last portion of the data for validation would involve temperatures that are warmer than most used to form the model and therefore will yield to biased validation results. To avoid this problem, two alternative methods of splitting the data are used for validation. Since each validation method discussed involves removing a different portion of the data before fitting the model, each method will produce a model that differs slightly from that given in Table 3. The values given in Table 3 were calculated using all the available data without splitting.

4.5.1. Validation Based on Random 20% Data Subset

One method for validation is removing data at random from the modeling set [42]. In this work, a random 20% of the data is removed from the modeling data and used as validation data. The model is fit to the modeling data (80% of the total data) and then used to simulate the validation data. The simulated and observed validation data are compared to assess the suitability of the model. Results from this validation method are given in Table 5 and shown in Figure 19.

4.5.2. Leave-One-Out Cross Validation

The leave-one-out method is a form of cross-validation which allows a more complete assessment of model quality than the previously-discussed technique [42]. The first data point is removed from the modeling data and the model is fit without using the first point. The removed point is then used to validate the model. The process is repeated for each data point. This method is more
computationally expensive than the previous validation method since it requires the regression to be calculated repeatedly. The advantage is that it allows validation across all data points.

Table 5. Validation results based on random 20% of data

<table>
<thead>
<tr>
<th>Mode</th>
<th>$\sigma$ [Hz]</th>
<th>$R^2$</th>
<th>Outlier Rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (4.7 Hz)</td>
<td>0.016</td>
<td>0.75</td>
<td>4.6%</td>
</tr>
<tr>
<td>2 (6.0 Hz)</td>
<td>0.050</td>
<td>0.57</td>
<td>5.2%</td>
</tr>
<tr>
<td>3 (7.2 Hz)</td>
<td>0.037</td>
<td>0.87</td>
<td>4.9%</td>
</tr>
<tr>
<td>4 (8.9 Hz)</td>
<td>0.022</td>
<td>0.86</td>
<td>5.8%</td>
</tr>
<tr>
<td>5 (13.2 Hz)</td>
<td>0.058</td>
<td>0.54</td>
<td>5.3%</td>
</tr>
<tr>
<td>6 (13.7 Hz)</td>
<td>0.115</td>
<td>0.45</td>
<td>7.2%</td>
</tr>
</tbody>
</table>

Figure 19. Random 20% validation of fourth-order model for all six identified modes

In many applications of cross-validation, not only the calculation of the model parameters but also the variable and model structure selection process is repeated each time the data is divided [43]. In the present work the input variables and model structure are fixed at the selected values and the leave-one-out strategy is simply used to provide different divisions of the data. Results from this validation method are given in Table 6 and shown in Figure 20.
Table 6. Leave-one-out cross validation results

<table>
<thead>
<tr>
<th>Identified f [Hz]</th>
<th>$\sigma_e [\text{Hz}]$</th>
<th>$R^2$</th>
<th>Outlier Rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7 Hz</td>
<td>0.015</td>
<td>0.79</td>
<td>3.2%</td>
</tr>
<tr>
<td>6.0 Hz</td>
<td>0.057</td>
<td>0.51</td>
<td>8.6%</td>
</tr>
<tr>
<td>7.2 Hz</td>
<td>0.035</td>
<td>0.88</td>
<td>6.1%</td>
</tr>
<tr>
<td>8.9 Hz</td>
<td>0.020</td>
<td>0.89</td>
<td>2.6%</td>
</tr>
<tr>
<td>13.2 Hz</td>
<td>0.049</td>
<td>0.58</td>
<td>2.9%</td>
</tr>
<tr>
<td>13.7 Hz</td>
<td>0.124</td>
<td>0.34</td>
<td>7.4%</td>
</tr>
</tbody>
</table>

Figure 20. Leave-one-out cross validation of fourth-order model for all six identified modes

4.5.3. Validation Comments

The leave-one-out method allows for the validation to be performed across all data points and therefore provide a more complete assessment of model performance than a single random division of the data. In the random 20% validation, only one fifth of the data are used for validation and the results can vary depending on exactly which data is selected. However, in both validation methods, the error standard deviation $\sigma_e$, the $R^2$ statistic, and number of outliers compare favorably with the values estimated from the modeling step (Tables 3 and 4) and with each other. After considering these validation methods it is concluded that the model fits the data well without being over-fit.
5. Conclusions

The continuous monitoring system on the Dowling Hall Footbridge consists of eight accelerometers to monitor pedestrian traffic/ambient vibrations and ten thermocouples to monitor the air temperature, the steel frame temperature, the temperature of the heated concrete deck, and the temperature of the piers. Six vibration modes of the footbridge could be reliably identified from purely ambient vibration sources such as wind and ground vibration using a fully automated system identification algorithm. For these modes the natural frequencies and mode shapes are accurately identified; the modal damping ratios are extracted with higher estimation uncertainty. Pedestrian traffic during daytime hours adds significant amplitude to the excitation and allows for more accurate system identification results. It is worth noting that the modal analysis was possible at low excitation amplitudes due to the flexibility of the bridge.

During the sixteen-week period from January 5, 2010 to April 25, 2010, the identified natural frequencies of the Dowling Hall Footbridge were observed to vary by 4 to 8 percent. During the same time period, measured temperatures ranged from -14 to 39 °C. Examination of the natural frequency and temperature data from the continuous monitoring system revealed that natural frequency and temperature were strongly correlated and that the relationship was nonlinear. No such temperature dependence was observed for the identified damping ratios or mode shapes.

Due to the high correlation between some temperature records, only three records were used to model the relationship between the identified six natural frequencies and the measured temperatures. The three temperature measurements which gave the best results for modeling were S3 (temperature of the steel frame), C2 (temperature of the concrete deck) and C1 (temperature of the western abutment). The fact that these three variables gave the best model suggests the structural frame, the concrete deck, and the boundary conditions all significantly affect the dynamics of this bridge. Furthermore, the
observed increase in natural frequency at below-freezing temperatures was by far the most sharp for C2 which suggests that freezing of the concrete deck is an important mechanism for stiffening of the structure at freezing temperatures.

After evaluating many candidate models (including a static linear model, an ARX model, a bilinear model, and polynomials of various orders) a fourth-order polynomial static model was selected. While the fourth-order polynomial without cross terms was selected as the final model, the bilinear model also performed well and shows promise for environmental modeling near the freezing point.

The validity of the selected 4th order model was assessed based on separate sets of validation data from two different data-splitting methods. For both validation sets, the error standard deviation and percentage of outliers were comparable with statistics estimated directly from the modeling data. This indicates that the model can be applied to the future data with almost the same accuracy that it fits the modeling data as long as future temperatures are not outside the range of modeling temperatures.

It is worth noting that the estimation uncertainty/variability of most vibration-based damage identification methods such as finite element model updating is a function of the uncertainty/variability of the modal parameters (especially the natural frequencies) used in these methods. Therefore, reduction in the variability of the natural frequencies will yield more accurate damage identification results. In this study, the standard deviation of the identified modal parameters was reduced significantly after removing the temperature effects using the selected 4th order model. For Mode 3 the standard deviation decreased from 0.10 Hz (before modeling) to 0.035 Hz (after modeling) with an $R^2$ of 87%. This large reduction in standard deviation will allow more accuracy in damage identification results. Along with the estimated model, confidence intervals were established for future data, giving a simple means of identifying damage: if a future natural frequency develops a trend outside the confidence intervals
established for the corresponding temperatures it is likely that the bridge has experienced structural change.

6. Acknowledgements

This work was partially funded by the Tufts University Faculty Research Award Committee (FRAC) whose support is gratefully acknowledged. Mr. Stefano Sensoli is recognized for his contribution to the project during the preliminary tests in April and June 2009. The authors wish to thank the Industrial Enclosure Corporation for donating a weatherproof enclosure to the project. The authors are grateful to Mr. James Pearson and Mr. Robert Reppucci at Tufts Facilities for flexibility in coordinating the project, and to Mr. Bidia Amana and Tufts University Information Technology for indispensable network support.
7. References


