

Elsevier Editorial System(tm) for Engineering Structures
Manuscript Draft

Manuscript Number: ENGSTRUCT-D-11-01111R3

Title: Methodologies for predicting natural frequency variation of a suspension bridge

Article Type: Research Paper

Keywords: Environmental effect; artificial neural network; support vector regression; regression tree; random forest; variable importance; suspension bridge

Corresponding Author: Dr. Irwanda Laory, Ph.D

Corresponding Author's Institution: University of Warwick

First Author: Irwanda Laory, Ph.D

Order of Authors: Irwanda Laory, Ph.D; Thanh Ngoc Trinh, Ph.D.; Ian Smith, Ph.D.; James Brownjohn, Ph.D.

Abstract: In vibration-based structural health monitoring, changes in the natural frequency of a structure are used to identify changes in the structural conditions due to damage and deterioration. However, natural frequency values also vary with changes in environmental factors such as temperature and wind. Therefore, it is important to differentiate between the effects due to environmental variations and those resulting from structural damage. In this paper, this task is accomplished by predicting the natural frequency of a structure using measurements of environmental conditions. Five methodologies - multiple linear regression, artificial neural networks, support vector regression, regression tree and random forest - are implemented to predict the natural frequencies of the Tamar Suspension Bridge (UK) using measurements taken from three years of continuous monitoring. The effects of environmental factors and traffic loading on natural frequencies are also evaluated by measuring the relative importance of input variables in regression analysis. Results show that support vector regression and random forest are the most suitable methods for predicting variations in natural frequencies. In addition, traffic loading and temperature are found to be two important parameters that need to be measured. Results show potential for application to continuously monitored structures that have complex relationships between natural frequencies and parameters such as loading and environmental factors.

Highlights

- We compare five regression methods to predict natural frequencies of a bridge.
- Random forest and support vector regression are the most appropriate methods.
- The relative importance of variables is useful to evaluate environmental effects.
- Traffic loading and temperature are the most influential parameters.
- Obtaining these parameters should be a priority.

Abstract

In vibration-based structural health monitoring, changes in the natural frequency of a structure are used to identify changes in the structural conditions due to damage and deterioration. However, natural frequency values also vary with changes in environmental factors such as temperature and wind. Therefore, it is important to differentiate between the effects due to environmental variations and those resulting from structural damage. In this paper, this task is accomplished by predicting the natural frequency of a structure using measurements of environmental conditions. Five methodologies - multiple linear regression, artificial neural networks, support vector regression, regression tree and random forest - are implemented to predict the natural frequencies of the Tamar Suspension Bridge (UK) using measurements taken from three years of continuous monitoring. The effects of environmental factors and traffic loading on natural frequencies are also evaluated by measuring the relative importance of input variables in regression analysis. Results show that support vector regression and random forest are the most suitable methods for predicting variations in natural frequencies. In addition, traffic loading and temperature are found to be two important parameters that need to be measured. Results show potential for application to continuously monitored structures that have complex relationships between natural frequencies and parameters such as loading and environmental factors.

1 **Methodologies for predicting natural frequency variation of a suspension** 2 **bridge**

3 Irwanda Laory¹, Thanh N. Trinh², Ian F. C. Smith² and James M. W. Brownjohn³

4 ¹*School of Engineering, University of Warwick, Coventry, United Kingdom*

5 *Email: I.Laory@warwick.ac.uk*

6 ²*Applied Applied Computing and Mechanic Laboratory, Swiss Federal Institute of*
7 *Technology Lausanne (EPFL), Lausanne, Switzerland*

8 ³*Vibration Engineering Section, College of Engineering, Mathematics and Physical Sciences,*
9 *The University of Exeter, Exeter, United Kingdom*

12 **Abstract**

13 In vibration-based structural health monitoring, changes in the natural frequency of a structure
14 are used to identify changes in the structural conditions due to damage and deterioration.
15 However, natural frequency values also vary with changes in environmental factors such as
16 temperature and wind. Therefore, it is important to differentiate between the effects due to
17 environmental variations and those resulting from structural damage. In this paper, this task is
18 accomplished by predicting the natural frequency of a structure using measurements of
19 environmental conditions. Five methodologies - multiple linear regression, artificial neural
20 networks, support vector regression, regression tree and random forest - are implemented to
21 predict the natural frequencies of the Tamar Suspension Bridge (UK) using measurements
22 taken from three years of continuous monitoring. The effects of environmental factors and
23 traffic loading on natural frequencies are also evaluated by measuring the relative importance
24 of input variables in regression analysis. Results show that support vector regression and
25 random forest are the most suitable methods for predicting variations in natural frequencies.
26 In addition, traffic loading and temperature are found to be two important parameters that
27 need to be measured. Results show potential for application to continuously monitored
28 structures that have complex relationships between natural frequencies and parameters such as
29 loading and environmental factors.

30 **KEY WORDS:** *Environmental effect, artificial neural network, support vector regression,*
31 *regression tree, random forest, variable importance, suspension bridge.*

32

33 **1. Introduction**

34 Many vibration-based approaches in structural health monitoring have been designed to
35 identify changes in natural frequency values for the purpose of detecting changes in structural
36 conditions that may indicate structural damage and degradation. In reality, however, civil
37 engineering structures are subject to environment and operating effects caused by changes in
38 temperature, traffic, wind, humidity and solar-radiation [1-5]. Such environmental effects also
39 change natural frequency values, hence concealing changes due to structural damage [6-10].
40 Therefore, it is important to distinguish between changes due to structural damage and
41 changes resulting from environmental effects. This task is managed observing then modeling
42 dependencies of natural frequencies on environmental parameters [11]. The prediction of
43 natural frequencies of structures under environmental changes has been studied using methods
44 such as linear regression analysis, artificial neural networks and support vector regression.

45 Multiple linear regression (MLR) was employed to predict changes in natural
46 frequencies of the Alamosa Canyon Bridge (USA) due to environmental temperature variation
47 [9] with natural frequencies formulated as a linear function of temperature data. It was found
48 that the changes in the frequencies were linearly correlated with temperature taken from
49 different locations on the bridge. Peeters et al. [12] conducted a one-year monitoring study for
50 the Z24-Bridge (Switzerland) before it was deliberately damaged, applying a linear regression
51 analysis to distinguish normal frequency changes from abnormal changes due to damage.
52 Also, for this concrete box girder bridge, Peeters and Roeck [13] applied an autoregressive
53 method with exogeneous inputs (ARX) to predict the bridge natural frequencies, where no
54 relationship was found between natural frequencies and wind, rainfall and humidity. Liu and

55 Dewolf [3] simulated the varying natural frequencies under temperature changes using a
56 linear regression analysis, concluding that the long-term variations of natural frequencies are
57 closely related to the variation in in-situ concrete temperature for the three frequencies they
58 measured. The MLR method has also been used to predict natural frequencies of suspension
59 bridges and a footbridge using long-term monitoring data [11, 14].

60 Artificial neural networks (ANNs) have been successfully applied in fields such as
61 pattern recognition [15], artificial intelligence [16] and civil engineering [17-20]. For long-
62 term monitoring of structures, ANNs have been employed to predict time-dependent natural
63 frequencies of a structure in order to eliminate the environmental effects on vibration-based
64 damage detection procedures. For example, Ni et al. [21] applied an ANN to formulate the
65 correlation between the natural frequencies and environmental temperatures taken from the
66 cable-stayed Ting Kau Bridge (Hong Kong). Zhou et al [22] further investigated the
67 performance of the ANNs formulated using the early stopping technique by constructing
68 three different kinds of input, including mean temperatures, effective temperatures and
69 principle components (PCs) of temperatures. The results indicated that when a sufficient
70 number of PCs were taken into account, the ANN using temperature PCs as inputs predicted
71 natural frequencies more accurately than that when using the mean temperatures. More
72 studies on ANNs for the prediction of structural responses are found in references [22-25].

73 Support vector regression (SVR) is an application form of support vector machines that
74 is a learning system using a high dimension feature space [26-27]. An attractive characteristic
75 of SVR is that instead of minimizing the observed training error such as with MLR and
76 ANNs, SVR involves minimizing the generalized error bound in order to achieve good
77 performance. The generalized error bound is the combination of the training error and a
78 regularization term that controls the complexity of prediction functions. A good overview of
79 SVR is given in [28-29]. SVR has been successfully employed in fields such as text

80 categorization and pattern recognition as well as structural health monitoring [27, 30]. Ni et
81 al. [31] applied SVR to predict natural frequencies of the cable-stayed Ting Kau Bridge
82 (Hong Kong) subjected to temperature variations taken from one-year measurement data, the
83 method exhibiting better prediction capability than the MLR method. Also using
84 measurement data of this bridge, Hua et al. [32] combined principle component analysis
85 (PCA) and SVR to simulate temperature-frequency correlations. It was found that the SVR
86 method trained using the PCs of measured temperature data outperformed that trained using
87 measured temperature data directly.

88 The methodologies used above are based on parametric functions that specify the form
89 of the relationship between inputs and a response (output) but in many cases, the form of the
90 relationship is unknown. Regression tree (R_Tree) methods offer a non-parametric
91 alternative [33] that has been used extensively in a variety of fields. The method has been
92 found to be especially useful in biomedical and genetic research, speech recognition and other
93 applied sciences [34]. Recent studies in the machine-learning field found that significant
94 improvements in prediction accuracy have resulted from growing an ensemble of trees in a
95 random way, a methodology called *random forest* (RF) [35]. It has been demonstrated that RF
96 has improved prediction accuracy in comparison to other regression methods [36] but
97 additionally provides measures of variable importance for each input variable [37-38]. This
98 method has not been evaluated for its applicability to structural health monitoring, so this
99 paper investigates the performance of RF on predicting natural frequencies through a case
100 study of a suspension bridge.

101 The studies mentioned above have proposed methodologies for predicting the dynamic
102 responses of bridges, but none has compared methodologies for prediction accuracy. This
103 paper compares five methodologies – multiple linear regression, artificial neural networks,
104 support vector regression, regression tree and random forest – in terms of their ability to

105 predict natural frequencies of a suspension bridge. Confidence intervals are then defined for
106 the best method to differentiate the effects due to environmental changes from those caused
107 by structural damage. Furthermore, the individual effects of temperature, wind and traffic
108 loading on the natural frequency responses of the bridge are evaluated using the variable
109 importance metric in regression analysis.

110 2. Methodologies for predicting natural frequencies of the bridge

111 2.1. Multiple linear regression (MLR)

112 Assuming that a response variable y (for example natural frequency) is linearly related
113 to the p input variables (for example temperature, wind and traffic loading) x_1, \dots, x_p so that

$$114 \quad y = \beta_0 + \sum_{i=1}^p \beta_i x_i + e. \quad (1)$$

115 This relationship is known as a linear regression analysis, where β_i is the regression
116 coefficient associated with the i^{th} input variable x_i and e the random error with mean zero
117 and variance σ^2 . Using the dataset of n observations in measurement time series, the
118 unknown coefficients β_i are determined using the least-squares method.

119 2.2. Artificial neural networks (ANNs)

120 Artificial neural networks can be used as a nonlinear regression method to predict the
121 natural frequency of a bridge. ANN is a two-stage regression in which the first stage is to
122 create derived features Z_m , represented by hidden layer, from linear combinations of the
123 inputs and the second stage is to model the output Y_m as a function of linear combinations of
124 the Z_m . Z_m could be considered as a basis expansion of the original input X .

$$\begin{aligned}
Z_m &= \phi(\alpha_{0m} + \alpha_m^T X), m = 1, \dots, M, \\
T_k &= \beta_{0k} + \beta_k^T Z, k = 1, \dots, K, \\
f_k(X) &= T_k + e, k = 1, \dots, K,
\end{aligned} \tag{2}$$

126 where $Z = (Z_1, Z_2, \dots, Z_M)$, $\phi(v)$ is the activation function which is usually chosen to be the
127 sigmoid $\phi(v) = 1/(1 + e^{-v})$, e the random error, α_i and β_i are unknown parameters. Given a
128 training set $\{x_i, y_i\}$ ($i = 1, \dots, N$), the ANN regression model is formulated by searching these
129 unknowns so that the sum-of-squared errors as a measure of fit reaches a minimum value.

$$R(\alpha, \beta) = \sum_{k=1}^K \sum_{i=1}^N (y_{ik} - f_k(x_i))^2 \tag{3}$$

131 The generic approach to minimizing, $R(\alpha, \beta)$, is by gradient descent, called back-
132 propagation. A two-layer back-propagation neural network (BPNN) is employed to predict
133 the natural frequencies of a structure. BPNN is first trained using the training set in order to
134 formulate the relationship between the natural frequencies and environmental factors
135 including direct loading such as traffic. BPNN is composed of one hidden layer and one
136 output layer with a tan-sigmoid transfer function in the hidden layer and a linear transfer
137 function in the output layer. The tan-sigmoid transfer function is capable of capturing the
138 nonlinear relationship between input variables (in our example three of them) and output
139 variables (in our example individual natural frequencies).

140 An important parameter to be determined when using BPNN for prediction tasks is the
141 optimal number of hidden nodes in the hidden layer. A network with too few hidden nodes
142 might not have enough flexibility to capture the nonlinearities in the relationship while a
143 network with too many hidden nodes may have a tendency to overfit the training data.

144 2.3. *Support vector regression (SVR)*

145 The strategy of SVR is to transform nonlinear relationships from the original space into
146 linear relationships in a new space (or feature space) defined using a kernel function so as to
147 discover relationships more easily [27, 36]. The linear function in the new space is given by

$$148 \quad y(x) = w^T \varphi(x) + b + e \quad (4)$$

149 where w is the weight vector; b is the bias constant and $\varphi(x)$ is the mapping function that
150 transfers the input vector x into the new space. Given a training set $\{x_i, y_i\}$ ($i = 1, \dots, N$), a
151 SVR model is obtained by minimizing the following objective function [39]

$$152 \quad \min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (5)$$

subject to $y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, \dots, N.$

153 where γ is the regularization parameter and e_i is the error. Such optimization that is subject
154 to a condition is solved using the Lagrangian function

$$155 \quad L(w, b, e, \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i \{w^T \varphi(x_i) + b + e_i - y_i\} \quad (6)$$

156 where α_i is a Lagrange multiplier. The conditions for optimality are given by

$$\begin{cases}
\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i \varphi(x_i) \\
\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\
\frac{\partial L}{\partial e} = 0 \rightarrow \alpha_i = \gamma e_i, & i = 1, \dots, N. \\
\frac{\partial L}{\partial \alpha} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0, & i = 1, \dots, N.
\end{cases} \quad (7)$$

158 Elimination of w and e yields a set of linear equations that are written in the matrix form

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & \Omega + \gamma^{-1} I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (8)$$

160 where $Y = [y_1, \dots, y_N]^T$, $1_N = [1, \dots, 1]^T$ and $\alpha = [\alpha_1, \dots, \alpha_N]^T$. I_N is an $N \times N$ identity matrix

161 and Ω is a $N \times N$ kernel matrix defined by a kernel function as

$$\Omega_{ij} = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j). \quad (9)$$

163 The kernel function is designed to compute inner-products in the new space using only the

164 original input data. The choice of K implicitly determines φ and the new space. Thus, the

165 advantage of kernel functions is that if a kernel function K is given, it is not necessary to

166 know the explicit form of the mapping function $\varphi(x)$. The selection of the kernel function

167 generally depends on the application domain. It has been shown that Gaussian radial-basis

168 function (RBF) is a reasonable first choice of kernel functions since it has only a single

169 parameter (standard deviation, σ) to be determined [27, 40]. The Gaussian RBF is expressed

170 as

171
$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} . \quad (10)$$

172 Solving Equation (8) identifies the values of α and b . Then, substituting these values into
 173 Equation (4) leads to the prediction

174
$$y(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b . \quad (11)$$

175 There are only two tuning parameters, γ and σ , that need to be determined when using the
 176 RBF kernel function and their optimal values are determined using the grid search method .
 177 Possible intervals for the two parameters are first defined. Then all grid points are tried to find
 178 the one giving the best accuracy. For each combination of the two parameters, SVR is trained
 179 using the training data and their performance is evaluated by a ten-fold cross-validation
 180 scheme.

181 *2.4. Regression tree (R_Tree)*

182 Regression tree is a nonparametric statistical method [33] that offers an alternative to
 183 parametric regression methods which usually require assumptions and simplifications to form
 184 the relationship. A regression tree is built by recursively partitioning the entire dataset,
 185 represented by a *root node*, into more homogeneous groups with each to be represented by a
 186 node. When the splitting process terminates, each resulting group is referred to as a terminal
 187 node. Splitting at each node is based on one value of an input variable that leads to the most
 188 homogeneous resulting nodes. Assuming that we have a partition into M regions $R_1, R_2, \dots,$
 189 R_M the system model is identified as

190
$$y(x) = ave(y_j | x_j \in R_m) + e \quad (11)$$

191 Where y_j and x_j represent the response and input variables at j^{th} observation respectively.
 192 Equation 10 shows that the predicted response is the average of y_j in region R_m with the
 193 error e .

194 A simple regression tree is built with two input variables x_1 and x_2 and a response y
 195 by considering a recursive partition as shown in Figure 1(a). First, we select the *splitting*
 196 *variable* (for example, x_1) and the *split point* (for example s_1) in order to achieve the most
 197 homogeneous splitting groups and split the space of the dataset into two groups. The selected
 198 variable and point solve

$$199 \quad \min_{j,s} \left[\sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 + \right] \quad (11)$$

200 c_1 and c_2 are the mean value of all the responses in the corresponding groups. Then, each of
 201 these groups is further split into two more groups. As shown in Figure 1(a), the group $x_1 \leq s_1$
 202 is split at $x_2 \leq s_2$ and finally the group $x_1 > s_1$ is split at $x_1 = s_3$. The process results in four
 203 groups R_1, \dots, R_4 . This process can be represented by the binary tree (Figure 1(b)). The
 204 entire dataset sits at the top of the tree, as a so-called root node. Observations (data points)
 205 satisfying the condition at each node are assigned to the left branch, and the others to the right
 206 branch. The terminal nodes of the tree correspond to the groups, R_1, \dots, R_4 . Once a tree has
 207 been built, the response for any new observation can be predicted by following the path from
 208 the root node down to the appropriate terminal node of the tree, based on the observed values
 209 of the splitting variables.

210 When determining tree size, note that a small tree may not capture a nonlinear
 211 relationship that may exist while a very large tree may over-fit the data. Therefore, tree size is
 212 a tuning parameter and the optimal tree size should be adaptively chosen from the data. The

213 preferred strategy is to gradually increase the tree size and evaluate the accuracy of each tree
214 size until each node contains fewer than a given number of observations (for example, 5).
215 Then this large tree is pruned by sequentially cutting off branches that add the smallest
216 capability to predictive performance of the tree according to a specified pruning criterion.

217 2.5. *Random forest (RF)*

218 A random forest is a combination of regression trees that are grown in random ways
219 [35]. The idea behind the random forest method is to generate an ensemble of low-correlated
220 regression trees and average results in order to reduce variance. The low-correlated trees are
221 generated by adding randomization in two steps: (i) each tree is grown using a random sub-
222 dataset of observations and (ii) each node of a tree is split using a random subset of input
223 variables. Figure 2 shows the layout of the random forest method.

224 The first step is to generate B sub-datasets of observations by randomly copying
225 observations from the original training set L until each sub-dataset has the same number of
226 observations N as the original training set. Some observations can be chosen several times
227 for each sub-dataset, whereas others are not chosen at all. It has been proved that about 37%
228 of the observations in the original training set are not chosen for each sub-dataset [38, 41].
229 The collection of non-chosen observations corresponding to each sub-dataset functions as a
230 validation set. Each sub-dataset is denoted L_b where $b = 1, 2, \dots, B$.

231 The second step involves growing a regression tree (T_b) using a sub-dataset (L_b). This
232 step is to reduce further the correlation between the regression trees that enter into the
233 averaging step later. This is achieved during the tree-growing process by randomly selecting
234 a subset of m input variables from all p input variables ($m \leq p$) before splitting each node.
235 A regression tree is grown by recursively repeating the following three sub-steps for each
236 node until the specified number of observations within each node is reached.

237 - Randomly select a subset of m variables from all p variables.

238 - Find the best split among the m variables.

239 - Split the selected node into two resulting nodes.

240 After B regression trees are grown from B sub-datasets, an ensemble of these B trees
241 is called a random forest. The random forest makes a prediction for a new observation x by
242 using each regression tree T_b in the forest to obtain a prediction $y_b(x)$ and then averaging B
243 prediction values from the B trees:

$$244 \quad y(x) = \frac{1}{B} \sum_{b=1}^B y_b(x) + e \quad (12)$$

245 3. Case study subject: The Tamar suspension bridge

246 The Tamar Suspension Bridge, as shown in Figure 3, is a road bridge connecting
247 Saltash to Plymouth in southwest England. The original bridge was designed as a
248 conventional suspension bridge with symmetrical geometry and was first opened in 1961. The
249 total length is 642 m with a main span of 335 m and side spans of 114 m and the tower height
250 is 73 m. Trusses are 4.9 m deep with chords of welded hollow box structures. To meet the
251 requirement that bridges should be capable of carrying lorries up to 40 tons, the Tamar Bridge
252 was strengthened and widened in March 1999 and the upgrading was completed in December
253 2001[42-43]. The upgrading included replacing the original composite main deck by a three-
254 lane orthotropic steel deck, adding single lane cantilevers at each side of the truss and
255 installing sixteen new cables acting as additional stays to carry the additional dead load of
256 new cantilever lanes and associated temporary works. Figure 2 shows the layout of one of the
257 truss sections with the main orthotropic deck and two cantilever lanes.

258 Many types of sensors were installed during and subsequent to the strengthening and
259 widening to monitor the behavior of the bridge [44]. They included anemometers,

260 displacement sensors, thermometers, load cells and accelerometers. Most recently, a robotic
261 total station was added to monitor the deflection of the bridge deck and a pair of
262 extensometers installed to track relative movement across the single expansion joint located
263 around the Saltash Tower [45]. Measurement data have been collected continuously since
264 February 2007.

265 These data used in this study include air temperature, wind velocity, the measured
266 natural frequencies of the bridge and the number of vehicles crossing the bridge every hour.
267 Vehicle crossing data were available from the bridge toll reports, temperature and wind values
268 are 30-minute averages of data sampled at either 1Hz from four thermistors on the cable and
269 deck and an anemometer close to midspan, while frequencies are derived from modal analysis
270 of 64-Hz sampled acceleration signals from a pair of accelerometers located near mid span.
271 Locations of these sensors are shown in Figure 3. The covariance-driven stochastic subspace
272 identification (SSI-COV) procedure operated automatically on the acceleration data after 8-
273 fold decimation, reporting frequency and damping estimates at 30 minute intervals.

274 Figure 6 shows the time history of air temperature for three years, including daily and
275 seasonal temperature variations. The temperature ranges from -5 °C to 25 °C between winter
276 and summer. The first five natural frequencies of the bridge are summarized in Table 1.

277 **4. Results**

278 The five regression methodologies presented in the previous section are applied to
279 predict the natural frequency variation of the Tamar Bridge based on environmental factors as
280 well as traffic loading. The prediction is performed for each natural frequency separately. The
281 measurement data taken from July 2007 to January 2010 on the Tamar Bridge are divided into
282 two non-overlapping and independent data sets: a training set of 70% and a test set of 30%.
283 While the training set (from July 2007 to May 2009) is used for regression analysis to predict

284 natural frequencies, the test set is used for assessing prediction accuracy (from June 2009 to
285 January 2010).

286 4.1. *Multiple linear regression*

287 The relationship between natural frequency responses and air temperature, wind and
288 traffic loading are first formulated for each frequency using the least square method. Figure 7
289 shows the prediction of the 10-day time histories (from 20 to 30 July 2009) of the third
290 frequency that is compared with the measured frequency in the test phase. This figure
291 indicates that the predicted frequency is unable to capture the high variation in the natural
292 frequency. Table 2 presents the mean square error (MSE) values in the training and test sets,
293 where the error is the difference between the measured frequency value and its corresponding
294 predicted value. For frequency 1 and 3, MSE values in the training set are somewhat larger
295 than those in the test set. This is because there are more outliers in the training data than in
296 the test set.

297 4.2. *Artificial neural networks*

298 The optimal number of hidden nodes in the hidden layer is determined so that the
299 validation error reaches the minimum value. To do this, a set of neural networks with respect
300 to the increasing number of hidden nodes from 1 to 50 are trained using training data. The
301 number of hidden nodes of the neural network that gives the minimum error is taken as the
302 optimal number.

303 Table 3 presents the optimal numbers of hidden nodes for five natural frequencies,
304 together with MSE values of the training set and the test set. The optimal number of hidden
305 nodes for five frequencies ranges from 10 to 33 nodes. These values are close to the optimal
306 value (19 hidden nodes) for the first natural frequency of the Ting Kau cable-stayed bridge
307 [46]. Figure 8 shows the predicted natural frequency along with the measured frequency.
308 Comparing Figure 7 and Figure 8 indicates that an artificial neural network achieves a better

309 prediction value than multiple linear regression. Comparing the prediction capability of ANN
310 with other methods is further discussed in Section 4.6.

311 4.3. Support vector regression

312 Table 4 presents the optimal values of γ and α that give the best performance (lowest
313 MSE) of SVR for five natural frequencies. The corresponding MSEs are also listed in this
314 table. Comparing the MSE values in Table 4 with Table 2 and Table 3 indicates that the SVR
315 method has a better performance than the MLR and ANN methods in both the training set and
316 the test set. For example, the prediction error (in the test set) for frequency 5 using SVR is
317 reduced by 20% when compared with the prediction error using MLR. Figure 9 shows the
318 predicted and measured time histories of frequency 3 from July 20 to 30, 2009. It is shown
319 that the predicted frequencies closely match the measured ones.

320 4.4. Regression tree

321 Figure 10 shows the mean squared error with respect to the increasing number of
322 terminal nodes (i.e. tree size) of the pruned tree for the first natural frequency. The optimal
323 tree size (i.e. the point where increasing tree size only leads to minor decrease of MSE) for
324 this frequency is composed of 31 terminal nodes. The optimal tree sizes for the second and
325 third frequencies are 44 and 35 terminal nodes respectively (Table 5). It is observed that the
326 higher frequency requires more terminal nodes, leading to a larger tree size, i.e. higher tree
327 complexity. Table 5 also presents the mean squared errors in the training and test sets. The
328 prediction of the R_Tree method is better than that of the ANN and MLR methods but it is not
329 as good as that of the SVR method.

330 Figure 11 shows the 10-day time histories of measured and predicted frequencies from
331 July 20 to July 30, 2009. For both sets, the predicted frequency time history reasonably
332 matches the measured one. Flatness exists at some peaks of the predicted time history. This

333 is because the observations at the peaks fall into the same groups where the predicted
334 responses are equal to the mean of measured responses within the corresponding group.

335 4.5. *Random forest*

336 When applying the random forest method for regression analysis, three parameters need
337 to be determined: (i) the sufficient number B of trees, (ii) the optimal number of observations
338 in each terminal node and (iii) the number m of input variables randomly chosen as
339 candidates for splitting at each node. For the Tamar Bridge, there are three input variables (i.e.
340 $p = 3$) including temperature, wind and traffic. For this case study, to reduce the correlation
341 between regression trees, the number of input variables chosen for splitting at each node is 2
342 (i.e. $m = 2$).

343 Figure 12 shows that when the number of trees increases, the mean squared error
344 computed from the validation set decreases. The prediction is stable at about 100 trees for
345 both cases of 1 and 50 observations in each terminal node. It is seen that the tree with 50
346 observations in terminal nodes performs better than that with only one observation in terminal
347 nodes. This is attributable to the over-fitting situation when growing a tree to its maximum
348 size (i.e. one observation in terminal nodes).

349 Figure 13 shows the change in the normalised mean squared error with respect to the
350 increase in the number of observations in terminal nodes for 5 frequencies. For each
351 frequency, the normalised MSE is calculated by dividing the MSE by the difference between
352 maximum and minimum MSE. When the number of observations in terminal nodes starts
353 increasing, initially the normalised MSE of all three frequencies drops dramatically to a
354 minimum and then it increases gradually. The optimal number of observations in terminal
355 nodes ranges from 10 to 50 observations. Table 6 presents the optimal number of
356 observations for each frequency together with its mean squared errors computed from the
357 training set and the test set. The results show that the random forest method has the smallest

358 errors as compared with those from the previous four methods. Figure 14 compares the
359 predicted natural frequency with the measured frequency. The predicted frequency closely
360 matches the measured one.

361 4.6. *Performance comparisons and discussions*

362 In order to find a suitable method to predict the natural frequency responses from
363 environmental measurement data for a suspension bridge, the prediction capability of five
364 regression methods are compared. The result of a regression method can have a very good fit
365 to the training data; however, it may poorly predict the response for a new observation. Thus,
366 the prediction capability of these methods is evaluated through prediction error that is defined
367 as the mean squared error from the test set, with a smaller prediction error indicating a better
368 prediction capability. When comparing the prediction error of the five regression methods
369 from Table 2 to Table 6, it can be seen that the four nonlinear regression methods (ANNs,
370 SVR, R_Tree and RF) predict frequencies more accurately than the MLR method. Table 7
371 presents the reduction in the prediction error for these methods when using the prediction
372 error of the MLR method as a basis. For frequency 5, SVR and RF can reduce the prediction
373 error up to 20% when compared with MLR. The good performances of SVR and RF indicate
374 the possibility of existence of non-linear correlations between natural frequency responses and
375 environmental factors as well as traffic loading for the Tamar Suspension Bridge. In addition,
376 comparing Figure 11 and Figure 14 demonstrates that RF employing multiple trees, which are
377 grown in a random way, can lead to better predictions than the R_Tree method that employs a
378 single tree. RF is able to capture the high variations at peaks of frequency time histories.

379 The performance of SVR and RF are further assessed through a normality test [47].
380 From a statistical point of view the error, which is the difference between the predicted value
381 and the corresponding measured frequency value, complies with a normal distribution with
382 zero mean [32]. Figure 15 compares the observed probability density functions of the error

383 with the corresponding theoretical curves of the normal distribution obtained using the mean
384 and standard deviation values computed from error values. The figure shows the observed
385 probability distribution of the error for SVR and RF methods is in good agreement with a
386 normal distribution with zero mean.

387 SVR and RF are used to define the confidence intervals around the predicted natural
388 frequencies for a new observation. It is found that the error in the training data for SVR and
389 RF also have a normal distribution with zero mean. Thus, the confidence interval is defined
390 based on the error variance of the training data. Figure 16 shows the identified and predicted
391 natural frequencies for RF between July 20 and July 30 (2009), together with the 95%
392 confidence interval for the second natural frequency. For the test set, the ratio of the data that
393 falls within the 95% confidence level to the full set of the data is referred to as the success
394 rate. For SVR, the success rates for frequencies 1 to 5 are 98%, 91%, 98%, 94% and 91%,
395 respectively. The corresponding success rates for RF are 98%, 91%, 98%, 94% and 89%.
396 These high success rates indicate that the variations in bridge natural frequencies can be
397 accounted for by measuring temperature, wind and traffic loading. These rates also
398 demonstrate the consistency of continuously monitored data from 2007 to 2010, thereby
399 establishing a baseline data for continuous health monitoring of the bridge. In addition, the
400 success rate can be used as a damage-detection index. If the success rates for future natural
401 frequencies change, it is likely that the bridge has experienced some kind of structural change.

402 **5. Effects of environmental factors and traffic loading on natural frequencies of the** 403 **bridge**

404 The changes in bridge natural frequencies are adequately accounted for by three factors:
405 temperature, wind and traffic loading. This study identifies the degree to which each factor
406 has an effect on the frequency change. Simultaneous effects of these factors on the first five
407 natural frequency responses are evaluated. This is carried out by using the measure of relative

408 importance of input variables in regression analysis. The measure of relative importance
409 indicates the variables that are highly related to the response for interpretation purposes.

410 *5.1. Evaluation of effects using relative importance metrics of the multiple linear regression* 411 *method*

412 Multiple linear regression can be used to evaluate the contribution of an individual input
413 variable $x_j (j = 1, \dots, p)$ to the prediction of a response y . The contribution of each variable is
414 compared with that of other variables using a metric of so-called relative importance. Several
415 relative importance metrics have been proposed to assess the amount of variation in the
416 response that is explained by each individual variable [48-49]. In this study, since the
417 correlation between input variables is negligible, the relative importance of each individual
418 variable is defined as the squared correlation coefficient of an input variable x_j with the
419 response y .

420 Figure 17 shows the relative importance of temperature, wind and traffic loading using
421 MLR for the first five natural frequencies of the bridge. The effects of temperature, wind and
422 traffic loading on the first frequency are 8%, 18% and 74%. Such effects correspond to 34%,
423 10% and 56% for the second frequency. They are 28%, 10% and 62% for the third frequency
424 and 22%, 10% and 68% for the fifth frequency. Except for the fourth frequency (i.e. 70%,
425 21% and 9%), based on relative importance metrics defined using MLR, traffic loading is the
426 main factor that affects the natural frequencies.

427 *5.2. Evaluation of effects using relative importance metrics of the random forest method*

428 The random forest method has improved the prediction accuracy in comparison to other
429 prediction methods. Besides, RF also evaluates the relative importance of variables in a
430 dataset in order to measure the prediction strength of each variable.

431 As mentioned in Section 2.5, approximately 63% of the observations in the original
432 training set are used for each sub-dataset on which to grow each individual tree. The non-

433 chosen observations (about 37%) are utilized as validation observations for that tree. The
 434 computation of the importance of an input variable x_j is carried out one tree at a time. First,
 435 when the b^{th} tree T_b is grown, the validation observations are then used to determine the mean
 436 squared error from the validation data MSE_b . Next, the values of variable x_j in the validation
 437 data are randomly permuted while leaving the values of all other variables unchanged. Then,
 438 the permuted observations are used in the tree T_b and the mean squared error from the
 439 permuted validation data $MSE_b(x_j)$ is computed. If x_j is important, permuting its observed
 440 values will reduce the prediction accuracy of each observed value in the validation data. Thus,
 441 $MSE_b(x_j)$ from the permuted validation data is larger than MSE_b from the un-permuted
 442 data.

443 Finally, a measure of the importance of the j^{th} variable x_j is obtained by averaging the
 444 mean squared errors from the permuted validation data over all of the trees:

$$445 \quad \text{imp}(x_j) = \frac{1}{B} \sum_{b=1}^B (MSE_b(x_j) - MSE_b). \quad (13)$$

446 The relative importance of each variable is computed by normalizing its importance to
 447 the summation of the importance of all variables. The relative importance metrics are
 448 expressed in percentage. Figure 18 shows the relative importance of temperature, wind and
 449 traffic loading on the natural frequency responses of the bridge. There is a significant effect
 450 of traffic loading on the frequency. Figure 18 also indicates that while the effect of traffic
 451 loading decreases from frequencies 1 to 5, the effect of temperature increases respectively.
 452 Both effects are almost similar for frequency 5 and the effect of temperature is more dominant
 453 than that of traffic loading for frequency 4.

454 5.3. Discussion

455 Comparing Figure 17 and Figure 18 shows that although variable importance metrics
456 are defined in two different ways using multiple linear regression and random forest, the
457 importance rankings for temperature, wind and traffic are identical. For the first frequencies,
458 the averaged percentages of the effects taken from both variable importance metrics are about
459 8%, 17% and 75% respectively. Such percentages correspond to 26%, 15% and 59% for the
460 second frequency, 24%, 9% and 66% for the third frequency, 60%, 22% and 18% for the
461 fourth frequency and 31%, 11% and 58% for the fifth frequency. A possible reason for the
462 effect of traffic loading is that there is a significant contribution of the traffic mass to the total
463 mass of the truss-span suspension bridge. Despite the strong influence on other frequencies,
464 the relative effect of the traffic on the fourth frequency is quite small. This could be due to the
465 fact that the fourth frequency refers to a torsional vibration mode while other frequencies refer
466 to vertical and lateral modes.

467 As for temperature effects, the influence on the variation of the fourth and fifth
468 frequencies is larger than that of the other frequencies. This could be caused by the non-linear
469 temperature distribution due to solar radiation. In general, for successful data interpretation
470 when monitoring natural frequency responses of suspension bridges, the effects of both traffic
471 loading and temperature need to be taken into account.

472 6. Conclusions

473 This paper compares five methodologies to predict the natural frequency responses of a
474 suspension bridge using measurements of temperature, wind and traffic loading. The
475 following conclusions are drawn

- 476 • Random forest and support vector regression are the most appropriate methods for
477 predicting the natural frequencies of a suspension bridge using measurement data of
478 temperature, wind and traffic loading. This may be due to non-linear behavior.

- 479 • The relative importance of input variables of regression analysis is a useful metric to
480 evaluate the simultaneous effects of environmental factors and traffic loading on the
481 long-term natural frequency responses of a bridge.
- 482 • Traffic loading and temperature are the most influential parameters on natural frequencies
483 of the suspension bridge studied. Obtaining these parameters should be a priority when
484 using natural frequency changes to detect damage.

485 **Acknowledgements**

486 The authors are grateful to Dr. Minh-Hai Pham and Dr. Ki-Young Koo for their contributions
487 to this research.

488 **References**

- 489 [1] Wahab MA, De Roeck G. Effect of temperature on dynamic system parameters of a highway
490 bridge. *Structural Engineering International*. 1997;7:266-70.
- 491 [2] Yuen K-V, Kuok S-C. Ambient interference in long-term monitoring of buildings. *Engineering*
492 *Structures*. 2010.
- 493 [3] Liu C, DeWolf JT. Effect of temperature on modal variability of a curved concrete bridge under
494 ambient loads. *Journal of Structural Engineering*. 2007;133:1742-51.
- 495 [4] Xu Z-D, Zhishen Wu. Simulation of the effect of temperature variation on damage detection in a
496 long-span cable-stayed bridge. *Structural Health Monitoring*. 2007;6:177-89.
- 497 [5] Brownjohn JM. Thermal effects on performance on tamar bridge. in *4th International*
498 *Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII-4)*. Zurich,
499 Switzerland. 2009.
- 500 [6] Laory I, Trinh TN, Smith IFC. Evaluating two model-free data interpretation methods for
501 measurements that are influenced by temperature. *Adv Eng Inform*. 2011;25:495-506.
- 502 [7] Posenato D, Lanata F, Inaudi D, Smith IFC. Model-free data interpretation for continuous
503 monitoring of complex structures. *Advanced Engineering Informatics*. 2008;22:135-44.
- 504 [8] Posenato D, Kripakaran P, Inaudi D, Smith IFC. Methodologies for model-free data
505 interpretation of civil engineering structures. *Computers & Structures*. 2010;88:467-82.
- 506 [9] Sohn H, Dzwonczyk M, Straser EG, Kiremidjian AS, Law K, Meng T. An experimental study
507 of temperature effect on modal parameters of the alamosa canyon bridge. *Earthquake*
508 *Engineering and Structural Dynamics*. 1999;28:879-97.
- 509 [10] Trinh TN, Koh CG. An improved substructural identification strategy for large structural
510 systems. *Structural Control and Health Monitoring*. 2011:doi: 10.1002/stc.463.
- 511 [11] Moser P, Moaveni B. Environmental effects on the identified natural frequencies of the dowling
512 hall footbridge. *Mechanical Systems and Signal Processing*. 2011;25:2336-57.
- 513 [12] Peeters B, Maeck J, Roeck GD. Vibration-based damage detection in civil engineering:
514 Excitation sources and temperature effects. *Smart Materials and Structures*. 2001;10:518.

- 515 [13] Peeters B, De Roeck G. One-year monitoring of the Z24-bridge: Environmental effects versus
516 damage events. *Earthquake Engineering & Structural Dynamics*. 2001;30:149-71.
- 517 [14] Yang D, Youliang D, Aiqun L. Structural condition assessment of long-span suspension bridges
518 using long-term monitoring data. *Earthquake Engineering and Engineering Vibration*.
519 2010;9:123-31.
- 520 [15] Bishop C. *Neural networks for pattern recognition*: Oxford University Press, USA; 1996.
- 521 [16] Raphael B, Smith IFC. *Fundamentals of computer-aided engineering*: West Sussex, England :
522 Wiley; 2003.
- 523 [17] Chan THT, Ni YQ, Ko JM. Neural network novelty filtering for anomaly detection of tsing ma
524 bridge cables in *The 2nd International Workshop on Structural Health Monitoring*. Stanford
525 University, Stanford, CA, USA. 1999.
- 526 [18] Fernandez B, Parlos AG, Tsai WK. Nonlinear dynamic system identification using artificial
527 neural networks (anns). New York, NY, USA: IEEE. 1990.
- 528 [19] Huang C-C, Loh C-H. Nonlinear identification of dynamic systems using neural networks.
529 *Computer-Aided Civil and Infrastructure Engineering*. 2001;16:28-41.
- 530 [20] Adeli H. *Neural networks in civil engineering: 1989–2000*. *Computer-Aided Civil and*
531 *Infrastructure Engineering*. 2001;16:126-42.
- 532 [21] Ni YQ, Zhou HF, Ko JM. Generalization capability of neural network models for temperature-
533 frequency correlation using monitoring data. *J Structural Engineering*. 2009;135:1290.
- 534 [22] Zhou HF, Ni YQ, Ko JM. Constructing input to neural networks for modeling temperature-
535 caused modal variability: Mean temperatures, effective temperatures, and principal components
536 of temperatures. *Engineering Structures*. 2010;32:1747-59.
- 537 [23] Zhou HF, Ni YQ, Ko JM. Performance of neural networks for simulation and prediction of
538 temperature-induced modal variability. in *Smart Structures and Materials 2005: Sensors and*
539 *Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, 7 March 2005*.
540 USA: SPIE-Int. Soc. Opt. Eng. 2005.
- 541 [24] Freitag S, Graf W, Kaliske M, Sickert JU. Prediction of time-dependent structural behaviour
542 with recurrent neural networks for fuzzy data. *Computers & Structures*. In Press, Corrected
543 Proof.
- 544 [25] Graf W, Freitag S, Sickert J-U, Kaliske M. Prediction of time-dependent structural behavior
545 with recurrent neural networks. in *The Sixth International Structural Engineering and*
546 *Construction Conference*. Zürich, Switzerland. 2011. 789-94.
- 547 [26] Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression
548 estimation, and signal processing. in *Advances in Neural Information Processing Systems 9*.
549 1996.
- 550 [27] Saitta S, Kripakaran P, Raphael B, Smith IFC. Feature selection using stochastic search: An
551 application to system identification. *Journal of Computing in Civil Engineering*. 2010;24:3-10.
- 552 [28] Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and Computing*.
553 2004;14:199-222.
- 554 [29] Basak D, Pal S, Patranabis D. Support vector regression. *Neural Information Processing –*
555 *Letters and Reviews*. 2007;11.
- 556 [30] Zhang J, Sato T. Experimental verification of the support vector regression based structural
557 identification method by using shaking table test data. *Structural Control and Health*
558 *Monitoring*. 2008;15:505-17.

- 559 [31] Ni YQ, Hua XG, Fan KQ, Ko JM. Correlating modal properties with temperature using long-
560 term monitoring data and support vector machine technique. *Engineering Structures*.
561 2005;27:1762-73.
- 562 [32] Hua XG, Ni YQ, Ko JM, Wong KY. Modeling of temperature-frequency correlation using
563 combined principal component analysis and support vector regression technique: ASCE; 2007.
- 564 [33] Breiman L, Friedman J, Stone C, Olshen RA. *Classification and regression trees*: Chapman &
565 Hall/CRC; 1984.
- 566 [34] Izenman A. *Modern multivariate statistical techniques : Regression, classification, and manifold
567 learning*: Springer New York; 2008.
- 568 [35] Breiman L. Random forests. *Machine Learning*. 2001;45:5-32.
- 569 [36] Zheng L, Watson DG, Johnston BF, Clark RL, Edrada-Ebel R, Elseheri W. A chemometric
570 study of chromatograms of tea extracts by correlation optimization warping in conjunction with
571 pca, support vector machines and random forest data modeling. *Analytica Chimica Acta*.
572 2009;642:257-65.
- 573 [37] Archer KJ, Kimes RV. Empirical characterization of random forest variable importance
574 measures. *Computational Statistics & Data Analysis*. 2008;52:2249-60.
- 575 [38] Grömping U. Variable importance assessment in regression: Linear regression versus random
576 forest. *The American Statistician*. 2009;63:308-19.
- 577 [39] Suykens JAK, Gestel TV, Brabanter JD, Moor BD, Vandewalle J. *Least squares support vector
578 machines*. Singapore: World Scientific; 2002.
- 579 [40] Hsu C, Chang C, Lin C. *A practical guide to support vector classification*. 2010.
- 580 [41] Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: Data mining, inference
581 and prediction*. 2 ed: Springer; 2009.
- 582 [42] Koo KY, Brownjohn JMW, List D, Cole R. Innovative structural health monitoring for
583 suspension bridges by total positioning system. in *IABMAS10*. Philadelphia, USA. 2009.
- 584 [43] Westgate R, Brownjohn JMW. Development of a tamar bridge finite element model. in *IMAC
585 XXVIII*. Florida, USA. 2010.
- 586 [44] Koo KY, Brownjohn JMW, Cole R, List DI. Structural health monitoring of tamar suspension
587 bridge. Submitted to *Journal of Structural Control and Health Monitoring*. 2011.
- 588 [45] Koo KY, Brownjohn JMW. Robotic total station for long-term structural health monitoring of
589 the tamar bridge. Submitted to *Engineering Structures*. 2011.
- 590 [46] Ni YQ, Zhou HF, Ko JM. Generalization capability of neural network models for temperature-
591 frequency correlation using monitoring data: ASCE; 2009.
- 592 [47] Freund RJ, Wilson WJ, Mohr D. *Statistical methods*. London: Academic Press; 2010.
- 593 [48] Chao Y-CE, Zhao Y, Kupper LL, Nylander-French LA. Quantifying the relative importance of
594 predictors in multiple linear regression analyses for public health studies. *Journal of
595 Occupational and Environmental Hygiene*. 2008;5:519 - 29.
- 596 [49] Groemping U. Relative importance for linear regression in r: The package relaimpo. *Journal of
597 Statistical Software*. 2006;17.
- 598
- 599
- 600
- 601

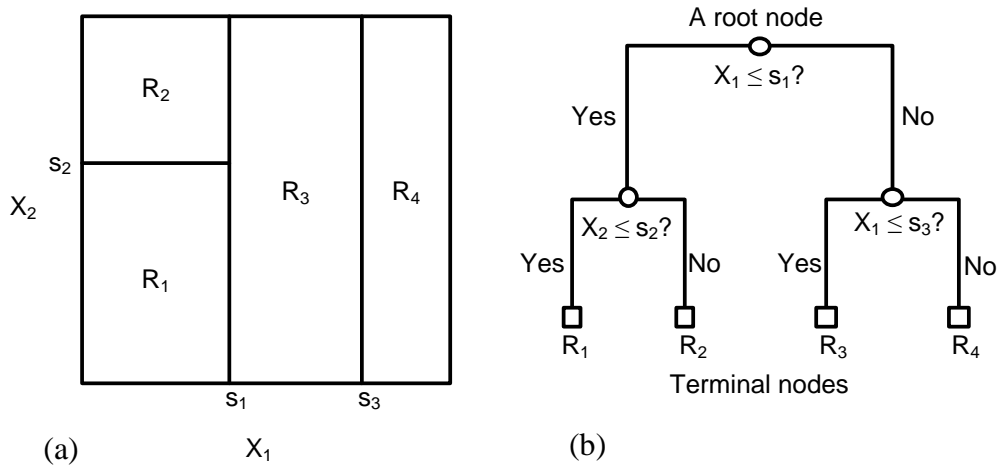


Figure 1. (a) The partitioning of a two-dimensional feature space into four regions, R_1 - R_4 ; (b) a decision tree with three splits and four terminal nodes corresponding to the four partitions.

602
603
604

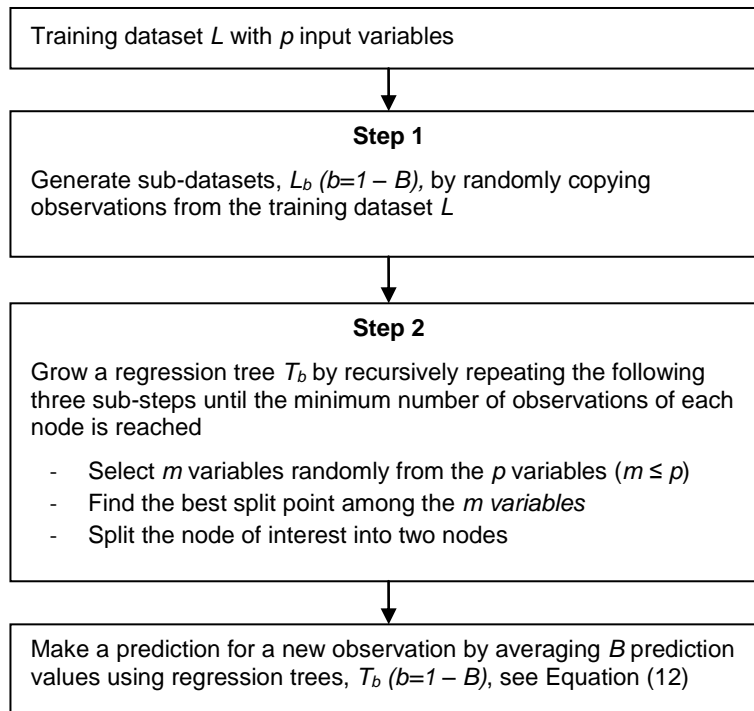


Figure 2. A layout of the random forest analysis method.

605

606



Figure 3. The Tamar Suspension Bridge

607

608

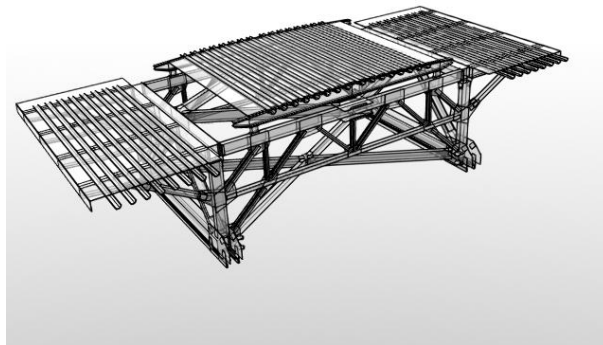


Figure 4 The truss section with the main orthotropic deck and two cantilever lanes

609

610

611

612

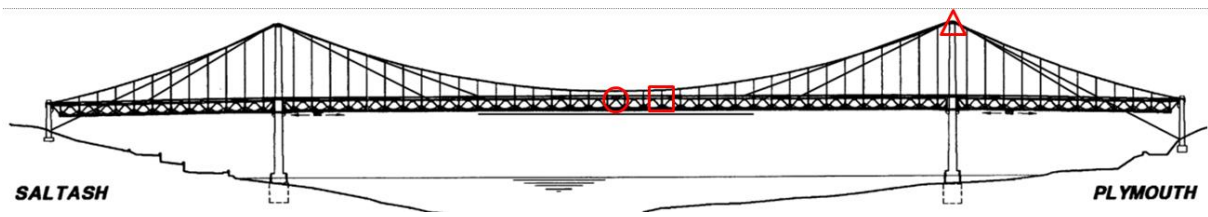


Figure 5 Sensor locations (circle for accelerometers, square for thermistors and triangle for anemometer)

613
614
615

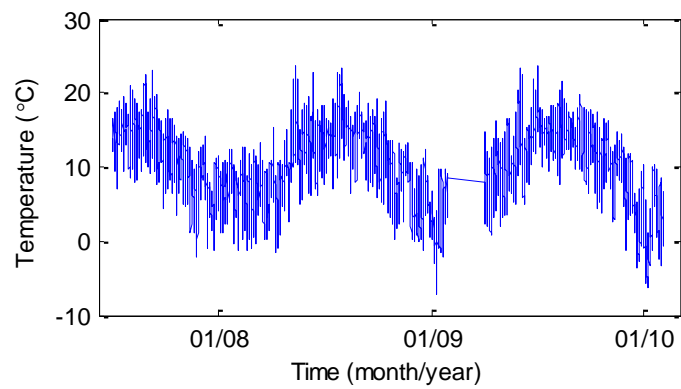


Figure 6. Time history of temperature measured from 2007 to 2010

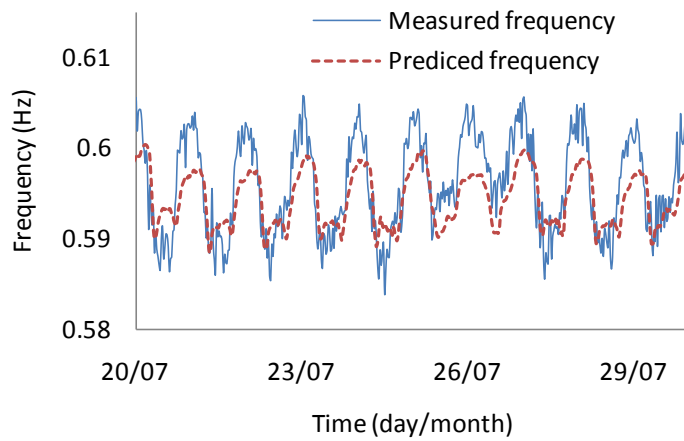


Figure 7. Measured and predicted natural frequencies between July 20 and July 30, 2009 using the multiple linear regression method.

616

617

618

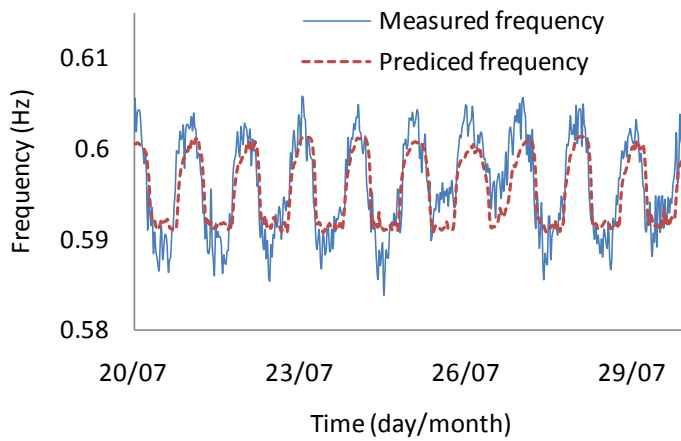


Figure 8. Measured and predicted natural frequencies between July 20 and July 30, 2009 using the artificial neural network method.

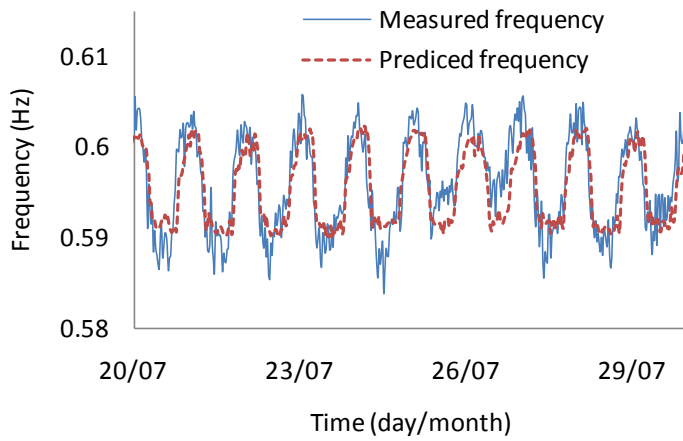


Figure 9. Measured and predicted natural frequencies between July 20 and July 30, 2009 using the support vector regression method.

619

620

621

622

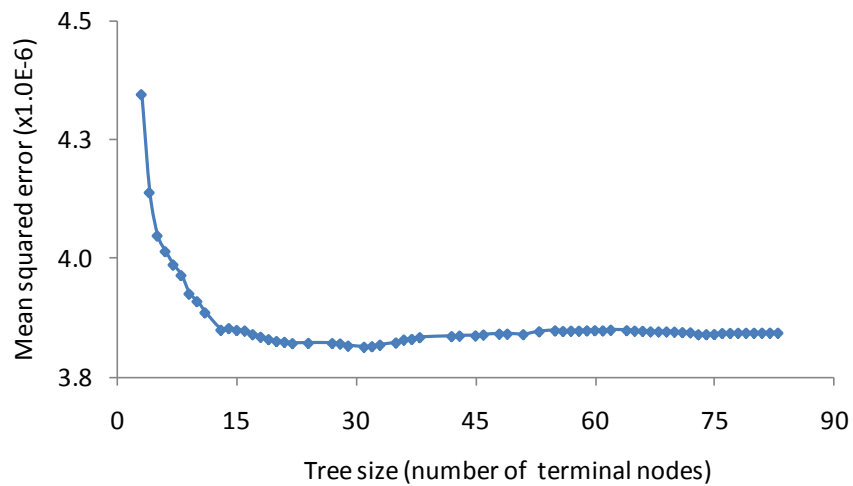


Figure 10. Mean squared errors versus the number of terminal nodes of a tree.

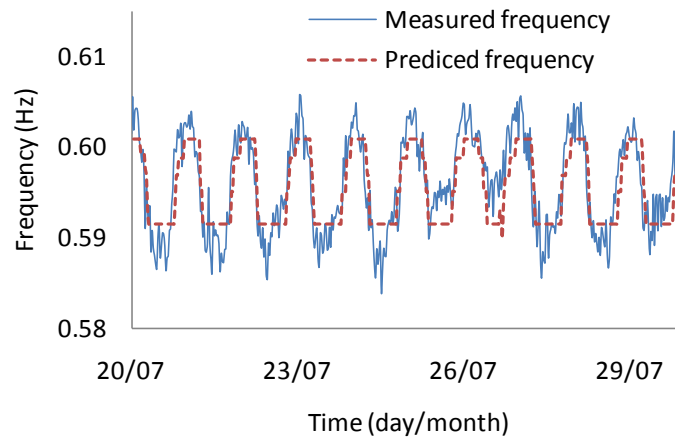


Figure 11. Measured and predicted natural frequencies between July 20 and July 30, 2009 using the regression tree method.

623

624

625

626

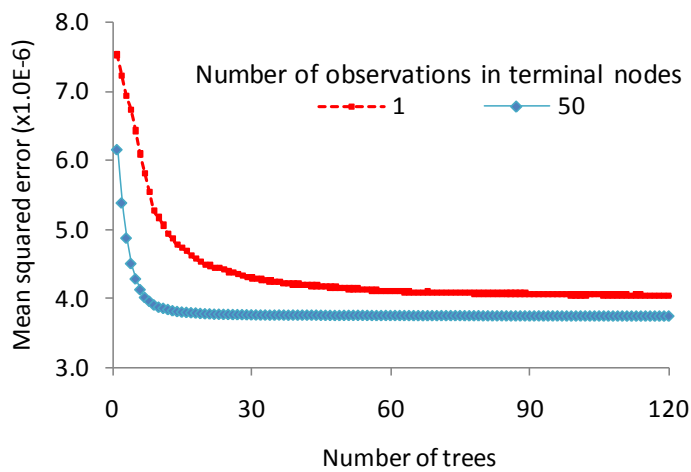


Figure 12. Mean squared errors versus number of trees for two cases of 1 and 50 observations in terminal nodes.

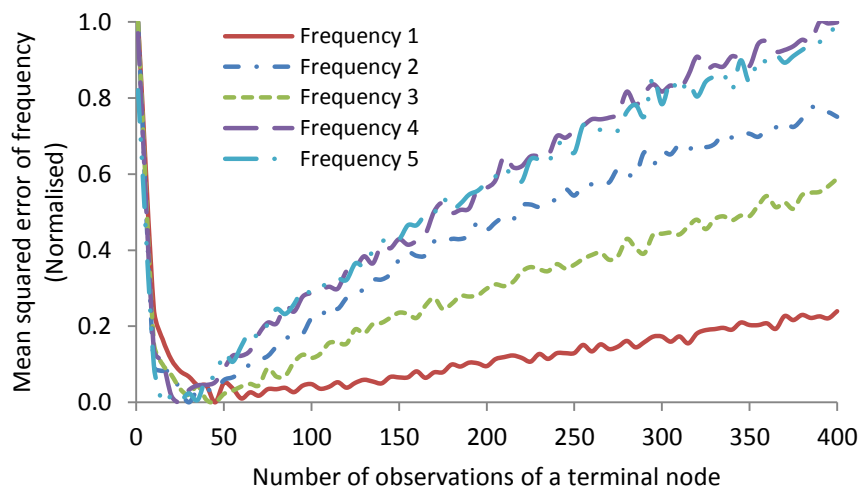


Figure 13. Mean squared errors versus the number of observations in a terminal node

627

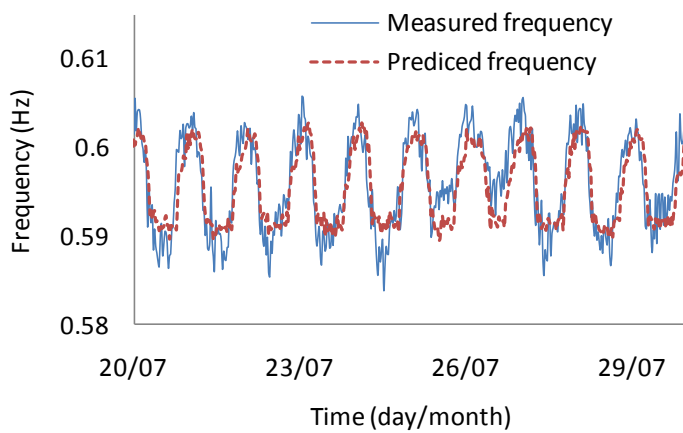


Figure 14. Measured and predicted natural frequencies between July 20 and July 30, 2009 using the random forest method

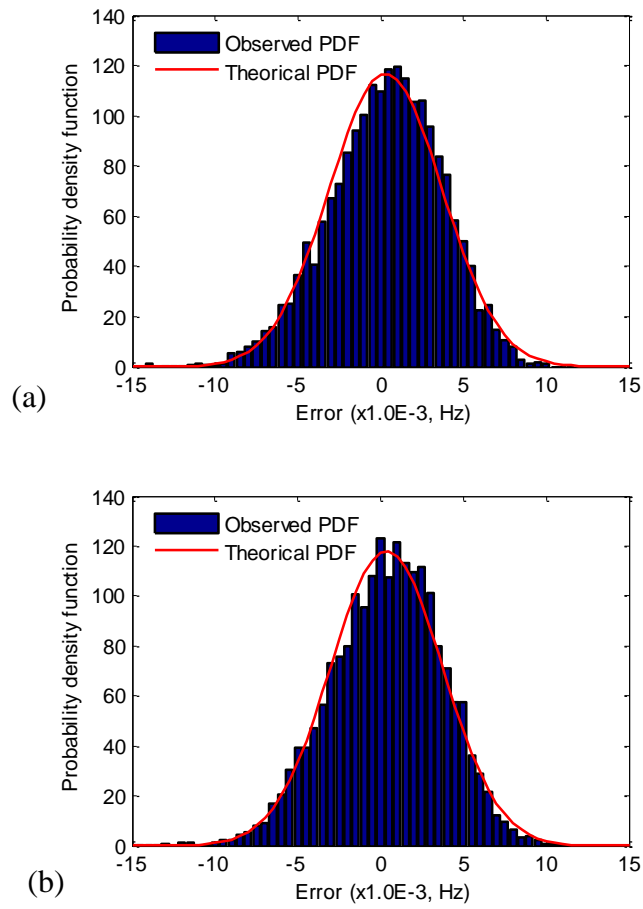


Figure 15. Probability distribution of errors for (a) support vector regression and (b) random forest

628

629

630

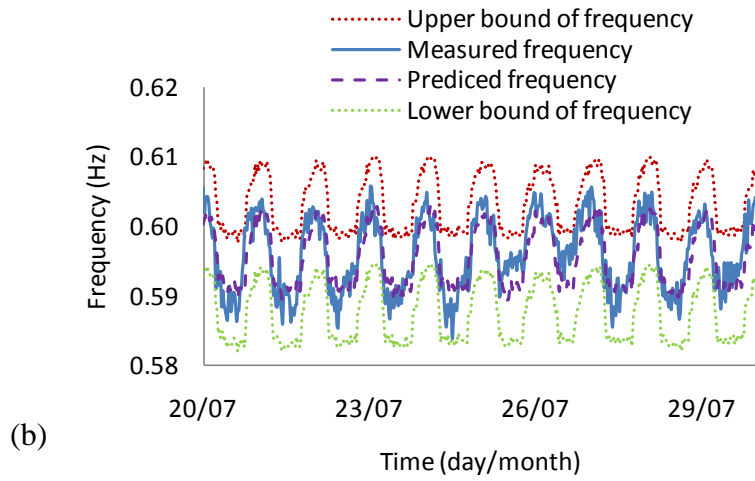
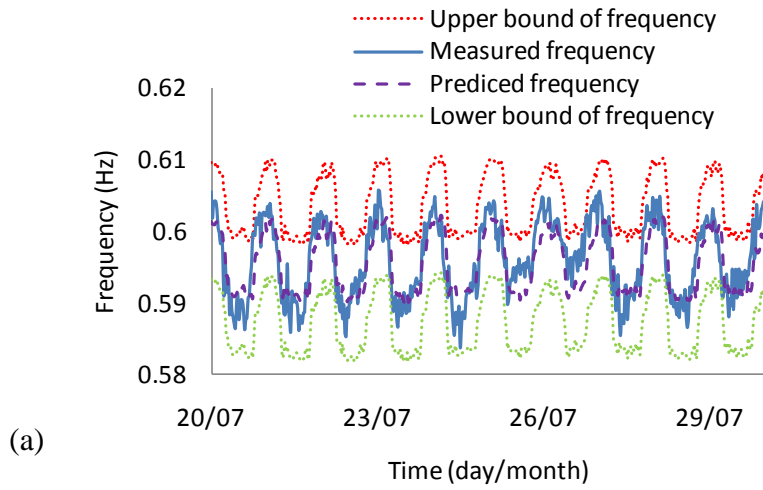


Figure 16. Measured and predicted natural frequencies (20 – 30 July, 2009) together with the 95% confidence interval using (a) support vector regression and (b) random forest

631

632

633

634

635

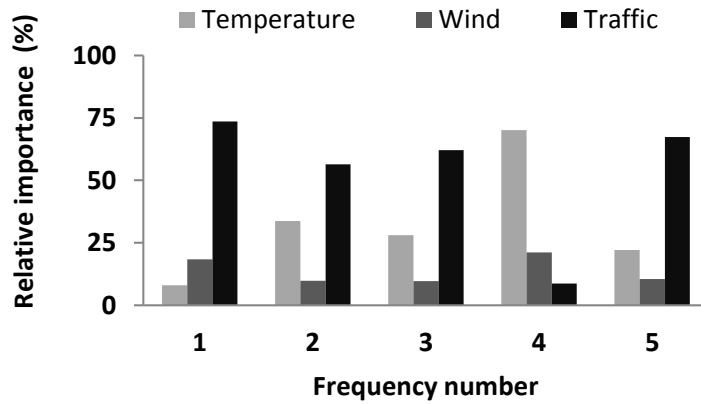


Figure 17. Evaluating simultaneous effects of temperature, wind, and traffic loading on the natural frequency responses through the relative importance of variables using the MLR method.

636
637
638
639

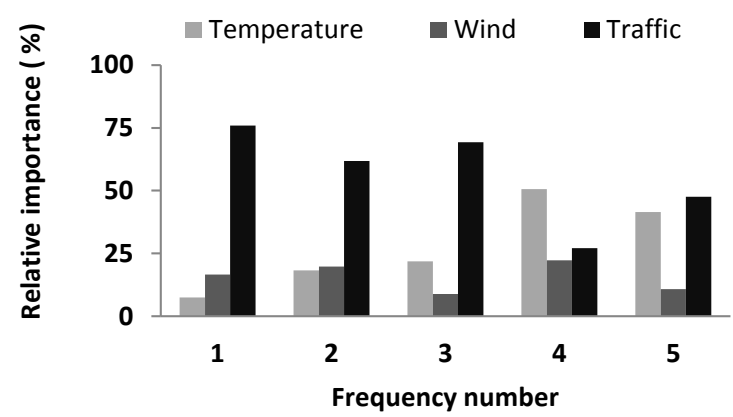


Figure 18. Evaluating simultaneous effects of temperature, wind, and traffic loading on the modal frequency responses through the relative importance of variables using the RF method.

640

641

Table 1. Parameters of measured natural frequencies of the bridge.

Mode number	Average frequency (Hz)	Frequency range (Hz)		Maximum difference (%)	Standard deviation (Hz)
		minimum	maximum		
1	0.39	0.38	0.41	8	0.00
2	0.47	0.41	0.57	34	0.01
3	0.60	0.58	0.61	5	0.01
4	0.69	0.67	0.70	4	0.00
5	0.73	0.71	0.75	6	0.01

642

643

644

645

Table 2. Results of the MLR method for the first five modes of the bridge

Frequency number	Mean squared error ($\times 10^{-6}$)	
	Training set	Test set
1	4.0	2.8
2	84.4	99.1
3	20.5	13.7
4	11.7	11.6
5	18.5	23.1

646

647

648

Table 3. Results of the ANN method for the first five modes of the bridge

Frequency number	Number of hidden nodes	Mean squared error ($\times 10^{-6}$)	
		Training set	Test set
1	11	3.9	2.7
2	10	87.6	95.7
3	21	22.8	12.3
4	33	15.7	18.6
5	21	19.5	20.0

649

650

651

Table 4. Results of the SVR method for the first five modes of the bridge

Frequency number	γ	α	Mean squared error ($\times 10^{-6}$)	
			Training set	Test set
1	20	0.8	3.7	2.5
2	9	0.17	69.5	96.7
3	20	0.56	16.6	11.8
4	12	0.28	9.4	10.6
5	14	0.36	15.7	18.6

652

653

654

655

656

Table 5. Results of the R_Tree method for the first five modes of the bridge

Frequency number	Number of terminal nodes	Mean squared error ($\times 10^{-6}$)	
		Training set	Test set
1	31	3.7	2.6
2	44	74.1	98.1
3	35	17.1	11.7
4	54	9.7	10.8
5	78	16.4	19.4

657

658

659

660

Table 6. Results of the RF method for the first five modes of the bridge

Frequency number	The optimal number of observations in terminal nodes	Mean squared error ($\times 10^{-6}$)	
		Training set	Test set
1	45	3.5	2.5
2	30	68.8	96.7
3	45	15.2	11.4
4	25	9.1	10.1
5	15	13.6	18.4

661

662

663

664

Table 7. Reduction in prediction errors of the ANN, SVR, R_Tree and RF methods when using the prediction error of the MLR method as a basis.

665

666

Frequency number	Error reduction (%)			
	ANN*	SVR	R_Tree	RF
1	6.5	10.8	9.8	12.9
2	3	2.4	1.1	2.4
3	10.4	14.2	14.9	17.0
4	8.6	8.5	6.9	12.9
5	13.5	19.4	15.8	20.1

667

(*) The error reduction for ANN when compared with the error of MLR, MSE_{MLR} , is equal to

668

$(MSE_{ANN} - MSE_{MLR}) \times 100 / MSE_{MLR}$; the same formulation is also applied for SVR, R_Tree and

669

RF.