Support Vector Regression for Anomaly Detection from Measurement Histories

Rolands Kromanis*, Prakash Kripakaran
College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
*E-mail: rk296@exeter.ac.uk

Abstract. This research focuses on the analysis of measurements from distributed sensing of structures. The premise is that ambient temperature variations, and hence the temperature distribution across the structure, have a strong correlation with structural response and that this relationship could be exploited for anomaly detection. Specifically, this research first investigates whether support vector regression (SVR) models could be trained to capture the relationship between distributed temperature and response measurements and subsequently, if these models could be employed in an approach for anomaly detection. The study develops a methodology to generate SVR models that predict the thermal response of bridges from distributed temperature measurements, and evaluates its performance on measurement histories simulated using numerical models of a bridge girder. The potential use of these SVR models for damage detection is then studied by comparing their strain predictions with measurements collected from simulations of the bridge girder in damaged condition. Results show that SVR models that predict structural response from distributed temperature measurements could form the basis for a reliable anomaly detection methodology.

1. Introduction

Bridges are valuable assets of the national highway infrastructure and their maintenance and management imposes a significant cost on the economy. In the UK, local authorities and Network Rail [1] estimated that they would require over £1.95 billion for the repair and strengthening of their bridge stock. The Federal Highway Administration (FHWA) [2] in the USA noted that almost 24% of the country's bridge stock was classified as structurally deficient or functionally obsolete in 2011. Therefore there is significant interest among the bridge engineering community in innovative technologies and approaches that reduce lifecycle costs of asset management. Current assessment procedures rely primarily on visual inspections, which have the following drawbacks:

- They often fail to detect early-stage damage [3]; Repairs undertaken at an advanced stage of deterioration are generally expensive and cause significant traffic disruption.
- They seldom provide sufficient data for accurately characterizing structural behaviour [3]. Consequently, estimates of structural capacity are typically conservative and impose unnecessary strengthening and replacement costs.

Monitoring systems have the potential to overcome these limitations by enabling early detection of the onset of damage, and accurate evaluation of asset condition and behaviour.

In the last decade, Structural Health Monitoring (SHM) systems have been deployed more frequently on bridges with the objective of tracking their real-time performance [3]. For example, three long-span bridges – Tsing Ma bridge, Kap Shui Mun bridge and Ting Kau bridge, are continuously monitored using over 800 permanently-installed sensors as part of the Wind and Structural Health Monitoring System (WASHMS) by the highways department in Hong Kong [4]. Wireless sensors that take advantage of energy-harvesting technologies are expected to further simplify the installation of future monitoring systems, and the storage and transmission of collected data [5-7]. These developments are envisaged to form the underpinning technologies for smart infrastructures [8] of the future that continuously sense their environment and provide real-time asset condition for effective management. However, this transition is greatly dependent on the development of fundamental methodologies for processing and interpreting the deluge of measurements generated by sensing systems.

The inverse engineering task of defining the state of a system from indirect measurements is often referred to as structural system identification [9]. System identification techniques [10] can be broadly classified into two categories: (i) model-based methods and (ii) data-driven methods. Model-based methods identify one or more behaviour models of the structure that are representative of measured structural behaviour. Since models are directly useful for structural assessment, these methods have been extensively studied by researchers in structural health monitoring (SHM). Many have focused on the evaluation of modal parameters such as mode shapes, frequencies and damping from vibration-based monitoring (VBM) [3, 11]. Model-based methods have also been investigated for interpreting static measurements [12, 13]. In particular, multi-model strategies [14, 15] that explicitly account for modelling and measurement uncertainties have been developed and successfully illustrated for analyzing measurements from static load tests of full-scale bridges [16]. However, challenges still remain, the most difficult being the quantification of the effect of variations in ambient conditions [11] and in particular, temperature variations which are known to greatly affect structural response [11, 17, 18]. Recently, Kulprapha and Warnitchai [19] showed that behavior models could be developed for predicting the thermal response of a multi-span pre-stressed concrete bridge from distributed temperature measurements. However, model development and simulation is often time and resource-intensive and thus not suited for analyzing large volumes of measurements [20].

In contrast to model-based methods, data-driven methods require minimal structural information and hence offer a lot of promise for real-time interpretation of measurements from continuous monitoring. These methods attempt to detect anomalous structural behaviour by evaluating whether new measurements deviate sufficiently from measurements taken when the structure is assumed to be healthy (baseline) state. For example, measurements collected soon after construction could be assumed to represent the normal condition and new measurements could be compared against this data to detect damage. Researchers have investigated many statistical techniques for interpreting quasi-static measurements including wavelet transform [21], pattern recognition [22] and autoregressive moving average models [23]. However, these methods do not incorporate the effects of ambient temperature variations and therefore detect anomalous structural behaviour only at advanced stages of damage since damage-induced changes in structural response are often masked by larger changes due to diurnal temperature variations.

Previous long-term monitoring studies have illustrated that daily and seasonal temperature variations have a great influence on the structural response of bridges [24, 25], and that this influence may even exceed the response to vehicular traffic [26]. Catbas et al. [26] monitored a long-span truss bridge in the USA and observed that the annual peak-to-peak strain differentials for the bridge were ten times higher than the maximum traffic-induced strains. Measurements taken from the Tamar bridge in the UK by Koo et al. [27] also showed that thermal variations were the major driver of deformations in the structure. Therefore there has been considerable interest in the SHM community on quantifying the effect of ambient conditions on structural response [28-30] and in particular, employ it for damage detection. The anomaly detection approach proposed by Posenato et al. [31, 32] relies on correlations between strain measurements and seasonal temperature variations. However, this approach based on moving principal component analysis (MPCA) requires a large set of reference measurements and is also unable to detect anomalous behaviour unless damage is very severe. Laory et al. [33] later illustrated the importance of including temperature effects in the interpretation process by showing that eliminating seasonal temperature variations from the measurement histories could negatively affect the performance of MPCA. However, no previous study has yet attempted to exploit the inherent relationship between distributed temperature and response measurements for anomaly detection.

This research attempts to explicitly capture the relationship between temperature distributions and response using support vector regression (SVR) models, and exploit this relationship for damage detection. SVR essentially employs the same theoretical concepts as support vector machines (SVM), which are a class of supervised learning methods widely used in the computing community for classification tasks. SVRs are chosen in this research due to their many successful applications for anomaly detection in diverse subjects such as computer networks, finance and medicine [34, 35]. In the civil engineering domain, Ray and Teizer [36] used SVR to create blind spot maps based on the construction machinery operator's head pose; the maps could then aid in warning operators of the presence of objects in their blind spots. SVRs have also been previously used in SHM applications. Shengchao et al. [37] proposed a SVR-based fault detection method to detect anomalies in the structure of F-16 fighters without requiring prior measurements for a faulty condition. Other applications in SHM include structural integrity assessment [38] and structural system identification [39]. SVR has also been shown to effectively capture correlations between temperatures and modal frequencies [40]. However, previous studies have not examined the application of SVR for quasi-static measurements, the focus of this research.

This research aims to develop a fast and robust method for anomaly detection by taking advantage of the correlations between temperature distributions across a structure and the measured structural response. The paper first presents an approach for generating SVR models from distributed temperature and response measurements. It then describes a strategy of using such models for anomaly detection. The paper evaluates the feasibility of this methodology on measurements that are obtained from simulations of numerical models representing a bridge girder in healthy and damaged states. It will also assess the performance of the developed methodology in the presence of noise and outliers in measurements.

2. Methodology

A typical bridge management framework that employs feedback from monitoring in the decision-making process is shown in Figure 1. The management process is iterative with results from monitoring being used to plan and prioritize interventions (e.g. repair, strengthening) and measurements from the bridge helping with condition assessment. The anomaly detection methodology that is presented in this paper is expected to form part of a suite of data interpretation techniques present within such a framework. These techniques, which may include both model-based and data-driven strategies, will supply information on real-time structural behaviour and condition.

This study will develop data-driven strategies for integrating the thermal response of bridges in the measurement interpretation process (shaded block in the measurement management module in Figure 1). It is, in principle, a first step towards using distributed temperature and response measurements for structural performance monitoring. The objectives are to (i) demonstrate that a data-driven strategy could accurately predict the thermal response of a structure from distributed temperature measurements and (ii) such a strategy could then form the basis of an anomaly detection methodology. While the examples in the paper predominantly focus on the relationship between temperature distributions and the strains they introduce in the structure, the proposed concepts are, however, applicable in general to all types of structural response (e.g. tilt, displacement).

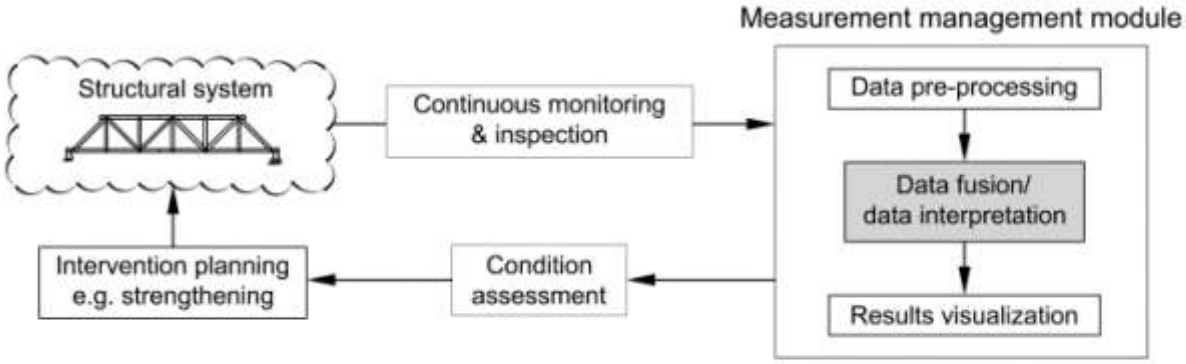

Figure 1: A typical framework for bridge management

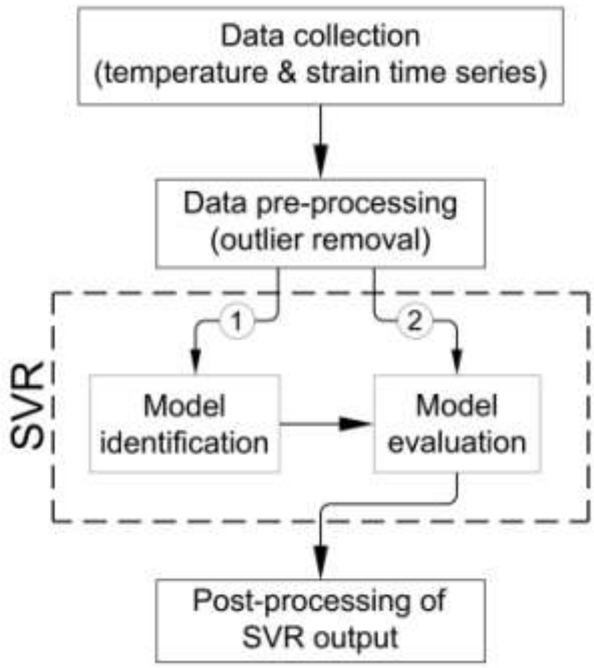

Figure 2: A flow chart of the proposed anomaly detection methodology.

A flowchart of the measurement interpretation strategy presented in this paper is shown in Figure 2. Measurements collected from sensors are first pre-processed to handle noise and remove outliers. These are initially used to train a regression model that captures the relationship between distributed temperature measurements and the measured thermal response. This training phase referred to as model identification in Figure 2 could happen when the structure is known to behave normally such as immediately after construction. The trained regression model is subsequently employed for predicting the structure's thermal response. During this phase, which is noted as model evaluation in Figure 2, the predictions from the regression models are compared with measured thermal response. The prediction errors are later analyzed within a post-processing phase. This study employs signal processing techniques to detect anomalous behaviour from the time series of prediction errors. The emphasis in this paper will be on the concept of employing support vector regression (SVR) for predicting the thermal response of bridges and on the post-processing of SVR output. The following section describes the approach for developing SVR models that form the basis for the proposed anomaly detection methodology.

3. Support Vector Regression (SVR)

3.1 Theory

SVR uses the same features that are central to support vector machines (SVM). In SVMs, datasets are often first transformed to a higher dimensional feature space using a kernel trick. Optimization is then used to find the hyper-plane that best separates datasets in this transformed feature space. The vectors that define the hyper-plane are referred to as support vectors. The process of finding the support vectors can be computation-intensive due to the tuning required as well as the quadratic optimization that is involved. The only addition in SVR is a loss function that determines the degree of complexity and generalization provided by the regression.

There are two main classes of SVR approaches – ε-SVR and ν-SVR. ν–SVR is used in this study since it requires less tuning and fewer number of parameters than ε-SVR. It also automatically minimizes the loss function and has been shown to support more meaningful data interpretation [41, 42]; this premise is validated by results from this research.

As for any machine learning technique, the core task in developing a regression model is to find model parameters that minimize the prediction error. The sensitivity of the SVR model is greatly dependent on the value specified for ν – a parameter which determines the number of support vectors and the number of bias support vectors. In addition to ν, values for two other parameters – a regularization constant (C) and gamma (γ), that also affect the performance of the SVR model have to be specified. Five-fold cross-validation is employed to evaluate the best values for C and γ. In this procedure, a data set is split into five equal parts such that one part constitutes the learning set that is trained on the other four parts. The values for C and γ are chosen such that they maximize the coefficient of determination (or squared correlation coefficient (R^2)), which is derived as follows:

$$R^2 = 1 - \sum_{i=1}^{n}\left(y_{pi} - \bar{y}\right)^2 / \sum_{i=1}^{n}(y_{ri} - \bar{y})^2, \; i = 1, 2, \dots, n \qquad (1)$$

where y_{pi} and y_{ri} represent the prediction and measurement at the i^{th} time-step, $\bar{y}$ is the mean value of the observed data and n is the number of observations. Lastly, several types of kernel functions are examined in this research. However, for reasons of brevity, results are presented only for linear kernels, which were also observed to give the best performance.

3.2 SVR for Anomaly Detection

Temperature and response measurements collected during an initial reference period when the structure has no damage constitute the training set. All measurements are scaled between 0 and 1 to reduce the time required to compute a SVR model. After training, distributed temperature measurements are provided as input to the SVR model and its predictions compared against measured thermal response. The difference (Δy_i) between the predicted and measured response at a given sensor location (Eq. 2), i.e. prediction error, is a measure of the structure's deviation from normal behaviour.

$$\Delta y_i = y_{pi} - y_{ri}, \; i = 1, 2, \dots, n \qquad (2)$$

In this study, strain histories are obtained from simulations of a numerical model that represents a bridge girder in healthy and damaged states. This numerical model is described in detail in the following section.

4. Numerical Model

A numerical model (see Figure 3) representative of a typical reinforced concrete girder found in highway bridges is employed as a test-bed in this study. The model is created using 8-noded plane stress elements in Ansys [43]. Each element has the following dimension: $360\ \text{mm} \times 300\ \text{mm} \times 500\ \text{mm}$ ($length \times width \times thickness$). Fiber Bragg grating (FBG) sensors that measure both strains and temperatures are assumed to be present on top and bottom faces at the quarter-spans of the girder. They have accuracies of ±1με and ±0.1°C. The locations of these sensors are shown in Figure 3 as S-1, S-2, etc.

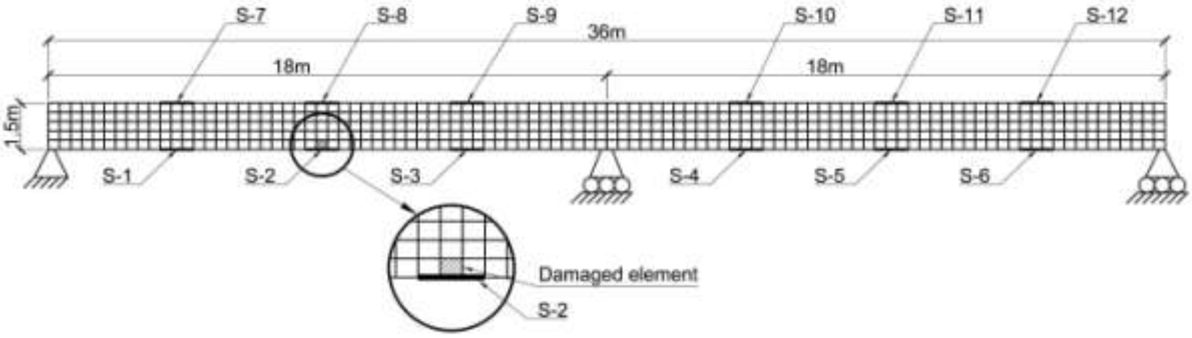

Figure 3: Numerical model of a bridge girder with S-*i* (*i = 1, 2, …, 12*) showing the assumed FBG sensor locations; the damaged element is near S-*2*.

The main purpose of setting up the numerical model is to simulate measurements of strains and temperatures similar to those generated by distributed sensing systems in continuously-monitored bridges under daily and seasonal temperature variations. The temperature distribution in a bridge is dependent on several factors including the ambient temperature, the geographical orientation of the bridge and its exposure to the sun. These effects could lead to complex, nonlinear temperature gradients in the bridge. This study focuses on the computational modelling of the relationship between temperature distributions and thermal response. Since it is the first such investigation into the thermal response of bridges, it evaluates the proposed approach for linear temperature gradients. Specifically, the following temperature distribution (see Figure 4) is considered: TEMP1 – a scenario representing linear temperature gradients across the length and depth of the girder (Figure 4). It is similar to the scenarios used in a previous study by Posenato et al [31]. Other forms of linear temperature gradients and combinations of these distributions have also been evaluated in this research to ensure that the proposed methodology is not sensitive to the nature of temperature distribution. However, results for these cases are not presented in this paper since its focus is on the central theme of anomaly detection.

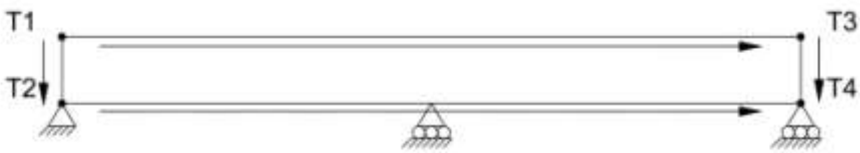

Figure 4: Temperature distribution for model in Figure 3; arrows show the direction of temperature increase.

Temperature histories from the European Climate Assessment & Dataset project (ECAD) project [44] are used to define the temperature distributions outlined in Figure 4. The histories are comprised of minimum, average and maximum daily temperature readings for a specific geographic location. Values for T1-T4 in Figure 4 for each time step are derived from the ECAD temperature histories. This study uses temperature histories recorded in Camborne, Cornwall, UK. Sensor readings are assumed to be taken during the hours when the bridge has minimal vehicular traffic. This is to ensure that the effects of ambient temperature variations dominate the measurements. This study also assumes the frequency of measurement collection to be one reading per day.

The model is used to simulate measurements from a bridge in both normal and damaged states. The structure is assumed to behave normally for the first three years. Damage is introduced after 1100 days (≈ 3 years) as a reduction in the material stiffness in one element. In concrete bridges, damage is often initiated by the corrosion of reinforcing steel due to chemical ingress. This tends to occur closer to mid-spans since the bending moments and the widths of resulting flexural cracks are largest around these locations. In an attempt to generate realistic damage scenarios, damage is modelled close to the middle of the first span of the bridge girder as shown in Figure 3. The following damage scenarios are considered:

(i) D1 – instant stiffness loss of 30%;
(ii) D2 – instant stiffness loss of 10%;
(iii) D3 – instant stiffness loss of 5%;
(iv) D4 – gradual stiffness loss – 1% reduction in stiffness every 15 days for 10 months (until it reaches 10%);
(v) D5 – gradual stiffness loss – 1% reduction in stiffness every 30 days for 10 months.

Measurements from full-scale structures often include outliers and noise. To account for this, randomly distributed outliers are introduced to the data set to represent malfunctioning sensors or external effects that may temporarily affect the sensors. They are introduced in both temperature and response measurements. We consider three outlier scenarios – O1, O2 and O3, equivalent to outlier percentages of 1%, 2% or 4% respectively. Magnitudes of outliers are assumed to be between -100 and +100 units. Measurement noise is added using a uniformly distributed random variable that takes values below 1% (N1), 2.5% (N2) or 5% (N3) of the peak-to-peak range of measurements from the first year.

5. Results

5.1 Performance of SVR model

The efficiency of the SVR strategy proposed in Section 3 is evaluated on data from the numerical model described in Section 4. Strain outputs from the numerical model are taken as the measurement histories in this study. Measurements are simulated for several scenarios, where each corresponds to a combination of a damage scenario and certain levels of outliers and noise. For example, scenario D1O1N1 refers to measurements simulated from the numerical model for damage case D1 taken together with outliers and noise levels corresponding to scenarios O1 and N1 respectively. Figure 5 shows the strain history at sensor S-2 of the girder (Figure 4b) for scenario D1. The figure shows that damage modelled as a 30% loss in stiffness is not visually discernible from the time series. The effects of damage are masked by the larger changes in strains due to daily and seasonal temperature variations.

A SVR model is created for each strain measurement location. Distributed temperature measurements constitute the input to the SVR model. In this study, five-fold cross validation is chosen for the training phase. In this procedure, the dataset is randomly divided into 5 parts; 4 parts are used for training and 1 part for testing the SVR model. Measurements taken during the first year form the training and test sets. The Libsvm package [45] is used for generating SVR models. A linear kernel is selected for the SVR. The SVR model is then evaluated for the task of predicting the structural response, i.e., strains. Figure 6 illustrates predictions from a SVR model trained on the first year of measurements from scenario D1N3. The SVR model is observed to predict strains to a high degree of accuracy.

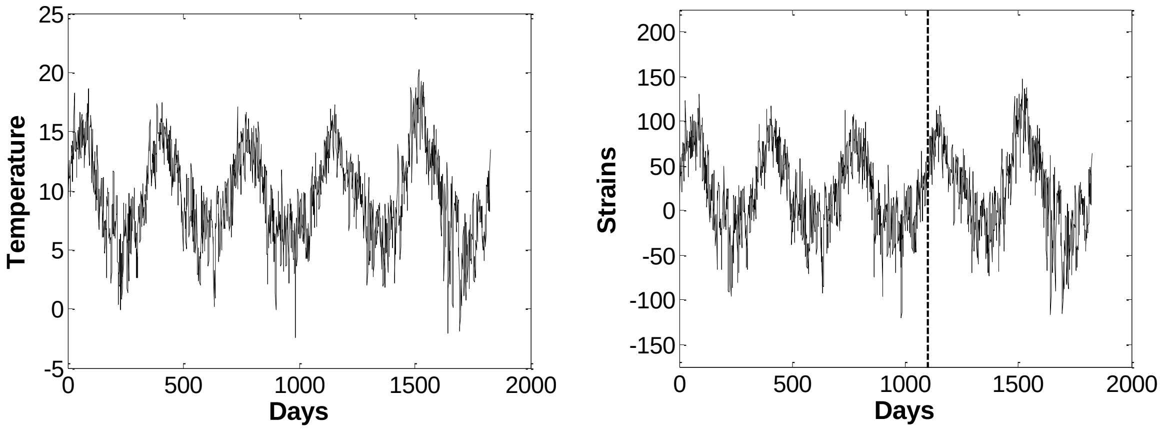

Figure 5: Temperature (left) and strain (right) readings from sensor S-2; dashed line indicates the introduction of damage.

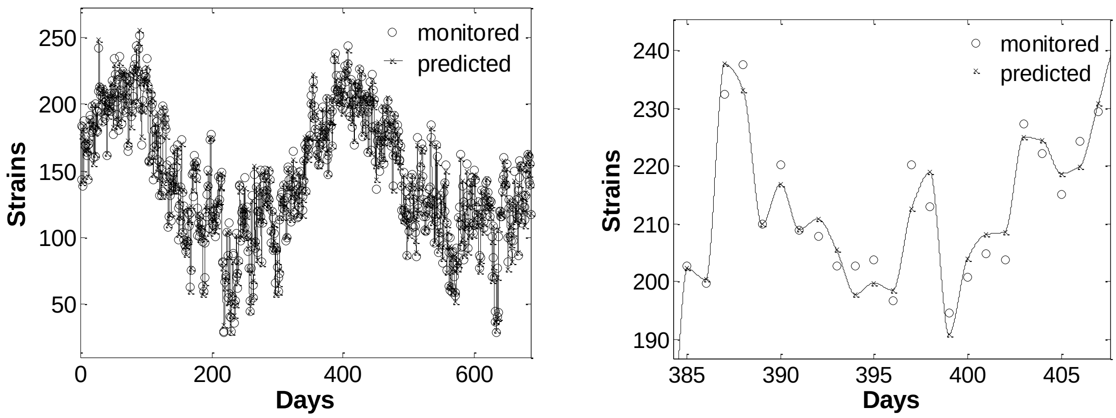

Figure 6: Comparison of measured and predicted strains for scenario D1N3 for two years (left) and a zoomed-in view for two weeks (right).

The prediction error (Δy), which is the difference between the measured strain and the prediction from the SVR model, could be an indicator of damage. This difference is plotted in Figures 7, 8 and 9 for sensor S-2 for damage scenarios D1, D3 and D5. There is a noticeable drop in the prediction error Δy after the damage is introduced; this illustrates that there is a deviation from normal behaviour. The time series could also be indicative of a transition to a new stable state upon collection of sufficient measurements after damage occurrence. This could help in monitoring progress of damage or deterioration. In the next step, time histories of predicted errors are analyzed with signal processing methods for automated detection of onset of anomalous structural behaviour.

5.2 Post-processing of SVR predictions

This research applies moving fast Fourier transform (MFFT) [46] to find statistical evidence of anomalous behaviour from the time series of prediction errors. MFFT is the fast Fourier transform of a moving window of data points from a time series, which in this case is on a sequence of Δy values. An anomaly is said to be detected when the indicator, which is the amplitude of the lowest frequency from MFFT, deviates significantly from its baseline value. The baseline value is defined as the mean value (μ) of the indicator during the reference period, i.e., the first year. The maximum permissible deviation from the baseline value beyond which a measurement is classified as an anomaly is defined as a constant *n* times the standard deviation (σ) of the indicator values during the reference period [33]. The assumption is that indicator values follow a Gaussian distribution with mean μ and standard deviation σ, and therefore, measurements that lead to indicator values outside the interval of $[\mu - n\sigma, \mu + n\sigma]$ have a high probability of representing anomalies. While increasing *n* reduces the sensitivity of the anomaly detection technique, it also minimizes the likelihood of false alarms. In this study, n=6 is chosen since it is observed to provide consistent and accurate results as shown below. The influence of this parameter on the performance of this methodology will be the focus of future research.

The time to damage detection is measured as the number of days between the introduction of damage and the detection of an anomaly. Results are illustrated for three damage scenarios D1, D3 and D5 in Figures 7, 8 and 9 respectively. In all three scenarios, the MFFT indicator shows a visible jump after damage occurrence and clearly detects anomalous structural behaviour.

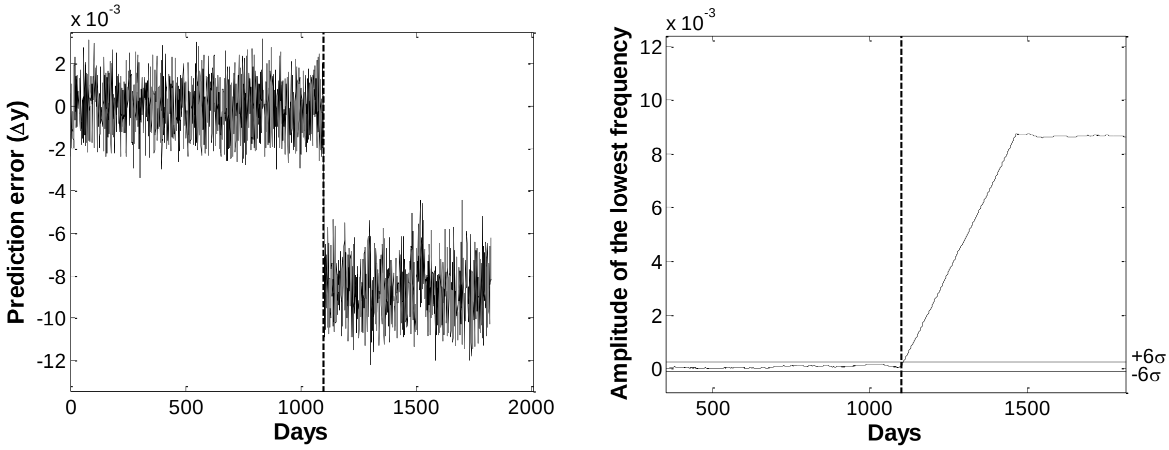

Figure 7: Time series of prediction errors (Δy) at sensor S-2 for scenario D1 (left) and results from MFFT (right); dashed line indicates the introduction of damage

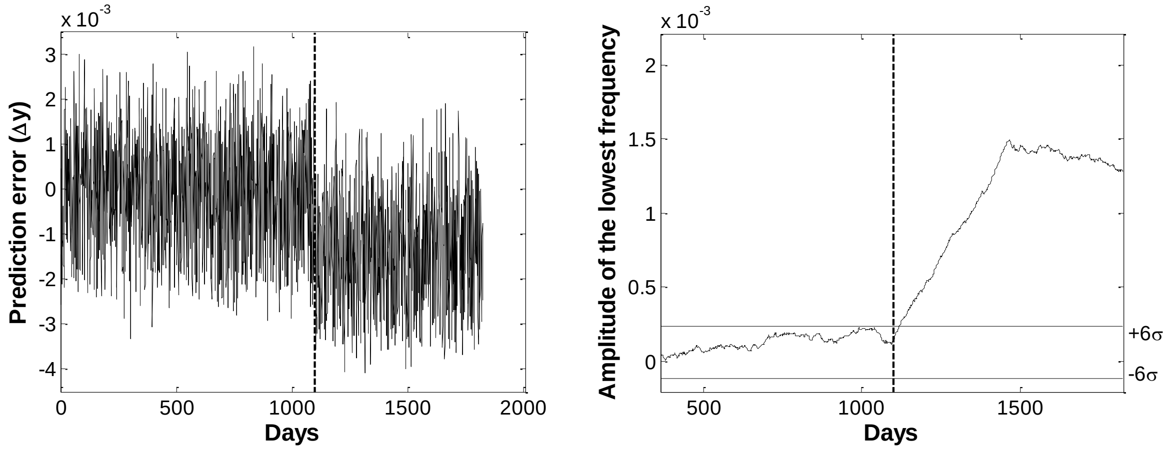

Figure 8: Time series of prediction errors (Δy) at sensor S-2 for scenario D3 (left) and results from MFFT (right); dashed line indicates the introduction of damage

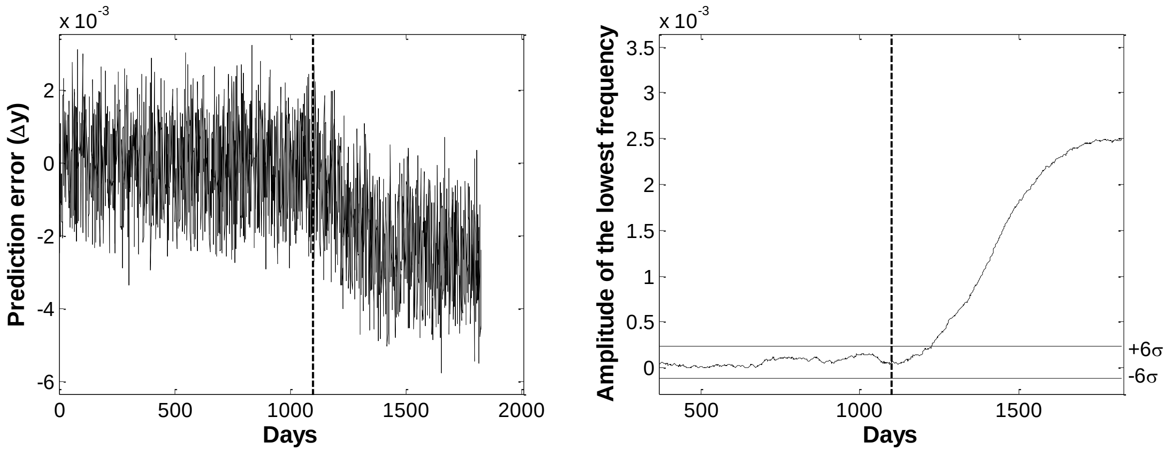

Figure 9: Time series of prediction errors (Δy) at sensor S-2 for scenario D5 (left) and results from MFFT (right); dashed line indicates the introduction of damage.

5.3 Performance under noise and outliers

The performance of the methodology in the presence of noise and outliers in the measurements is studied. The time series of strains and temperatures are first pre-processed to handle outliers. There are two fundamental approaches to managing outliers – (a) exclude measurements classified as outliers from analysis and (b) replace outliers with appropriate values. The former requires that valid measurements collected at the same time-step at which an outlier is detected are also excluded and hence may lead to loss of useful data. For this reason, the latter approach of outlier replacement is employed in this study. Following a preliminary evaluation of outlier replacement techniques [31] such as three-σ analysis and interquartile range (IQR), IQR is chosen to manage outliers in this study. IQR was also shown to outperform other outlier detection techniques in an earlier study by Posenato et al [31], which compared a number of outlier replacement techniques to pre-process measurement time histories for analysis using moving principal component analysis (MPCA). IQR technique uses the statistics of data within a moving window to determine the outliers. A moving window of size equivalent to two months of measurements is employed in this study. The value located in the middle of a moving window is evaluated against statistical thresholds defined for that window and then classified either as an outlier or a valid measurement. A value classified as an outlier is replaced by the median value for the moving window. The application of IQR to temperature and strain time series from sensor S-2 for scenario D5O3 are shown in Figures 10 and 11 respectively.

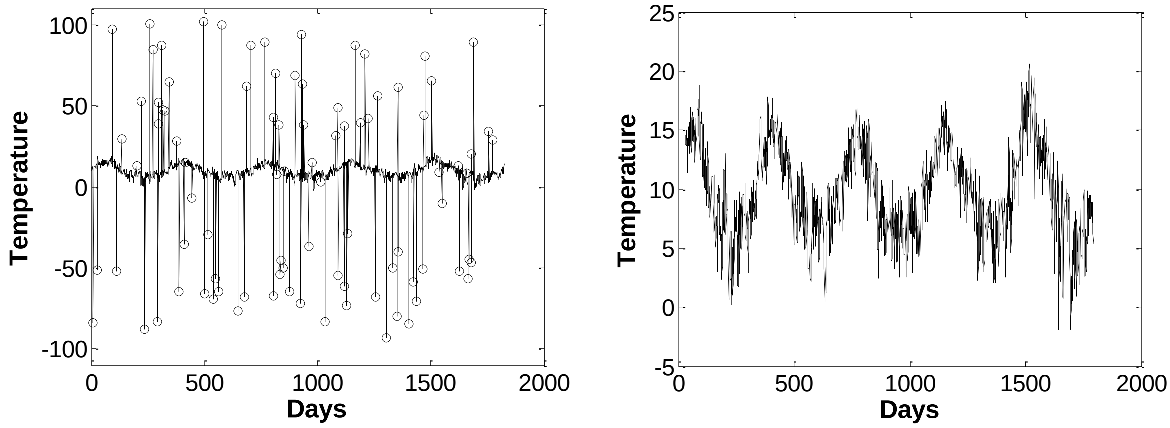

Figure 10: Time series of temperatures collected at S-2 for scenario D5O3, before (left) and after (right) outlier removal.

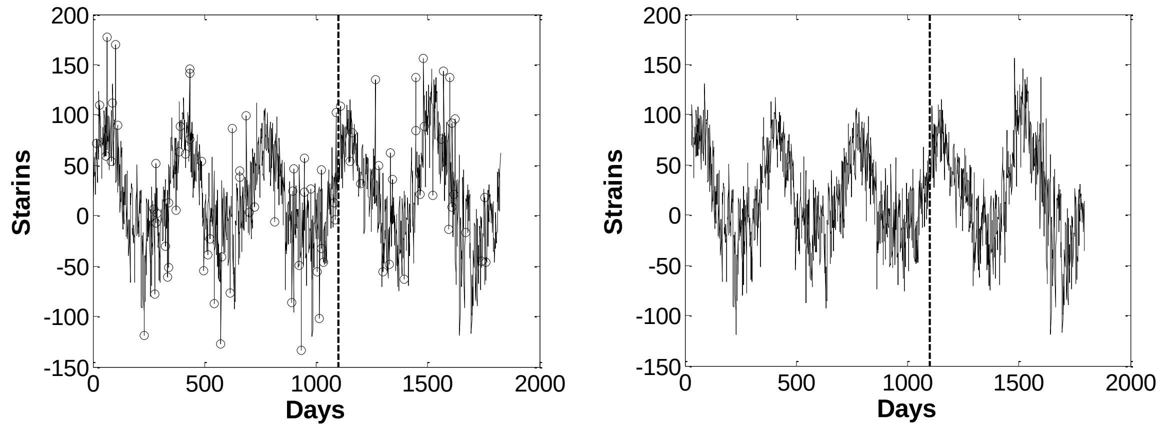

Figure 11: Time series of strains collected at S-2 for scenario D5O3, before (left) and after (right) outlier removal; dashed line indicates the introduction of damage.

IQR analysis does not fully eliminate the problems posed by outliers. First, they seldom identify all outliers in the data. Second, the median values that replace the outliers may still have significant errors. Therefore, even after pre-processing, outliers could still detrimentally affect the training of regression models and the accuracy of predictions. The use of SVR helps address these issues. The generalization ability of SVR is useful in producing robust models. Also, the outliers in the input strain and temperature measurements magnify the prediction errors (Δy) and therefore, produce equivalent outliers in the Δy time series. These outliers that are missed during pre-processing could be eliminated by cleansing the Δy time series using the same outlier removal technique (IQR analysis). A moving window of a length of one month is chosen for this task. This procedure is illustrated in Figures 12 and 13. The plots on the left in these two figures show the time series of Δy values before and after outlier removal respectively for sensor S-2 under scenario D5O3. The plots on the right in Figures 12 and 13 provide the results from MFFT. It is clear that the removal of outliers reveals a drop in the prediction error which could then be identified as an anomaly using MFFT (see Figure 13).

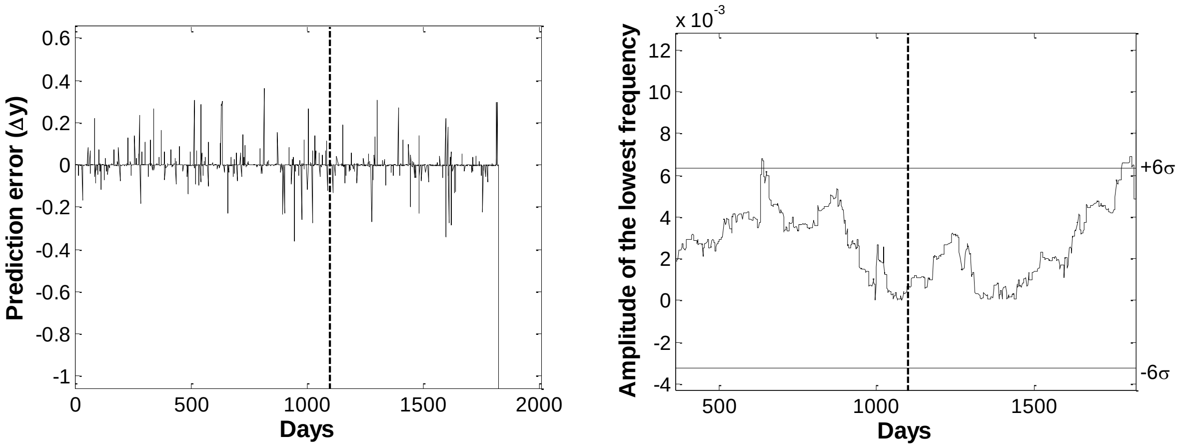

Figure 12: Time series of prediction errors (Δy) at sensor S-2 for scenario D5O3 after pre-processing strain/temperature measurements for outliers (left) and results from MFFT (right); dashed line indicates the introduction of damage.

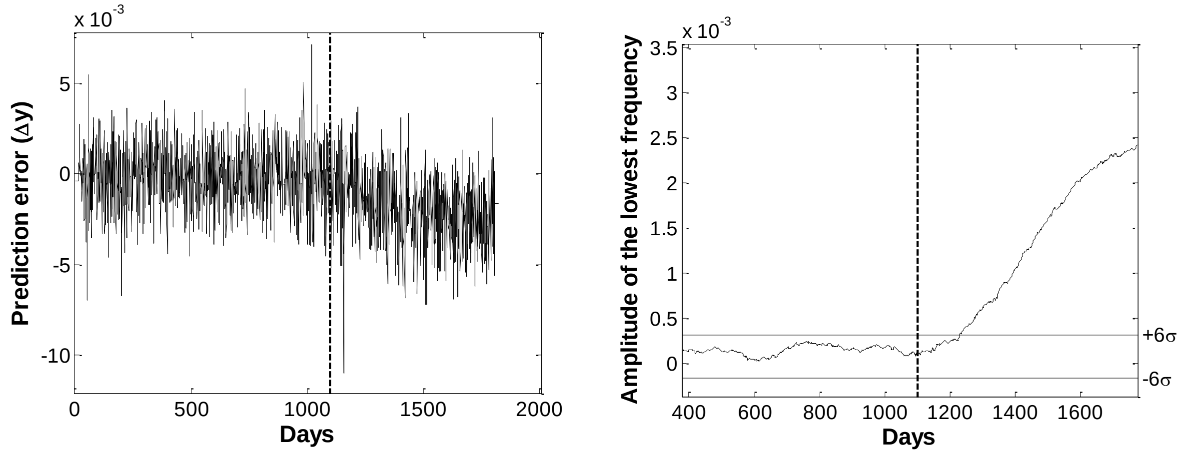

Figure 13: Time series of prediction errors (Δy) (left) produced after applying IQR analysis to data in Figure 12 and corresponding results from MFFT (right); dashed line indicates the introduction of damage.

Next the robustness of the methodology is evaluated for increasing levels of noise. The magnitude of noise is derived from peak-to-peak values of sensor readings from the first year (365 days). The time series of prediction errors (Δy) has increased distortion in the presence of noise. This will increase the variability in the baseline data and hence delay the detection of damage. The prediction error and corresponding results from MFFT for the scenario D5N2 is represented in Figure 14.

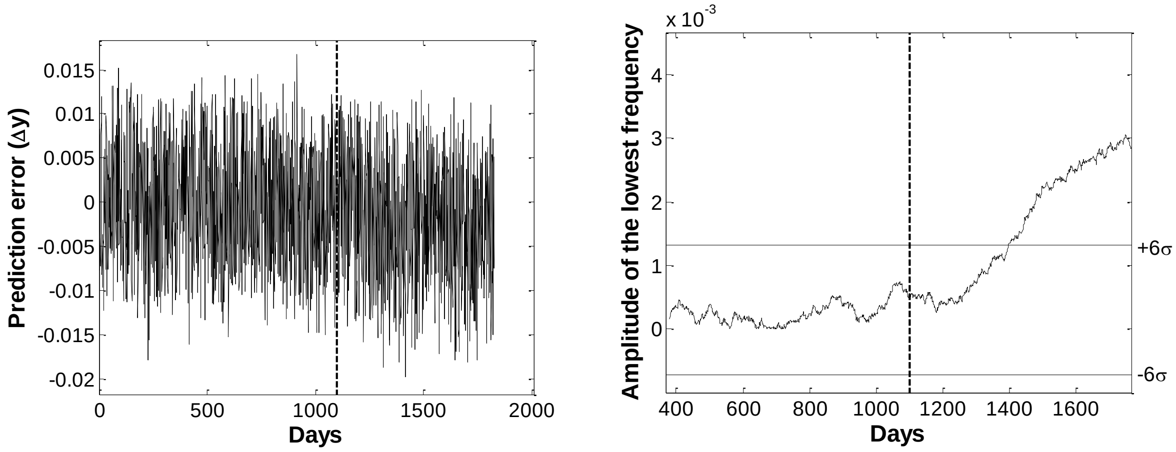

Figure 14: Time series of prediction errors (Δy) at sensor S-2 for scenario D5N2 (left); results from MFFT of the Δy time series (right); dashed line indicates the introduction of damage.

5.4 Discussion

The previous section presented notable results for only a few scenarios. This research, however, has investigated the proposed methodology that combines SVR and MFFT for a much larger set of scenarios. These results are summarized in Table 1. As expected, time to detect damage varies depending upon the chosen scenario. The introduction of outliers and noise has a significant impact on the performance of the methodology. The presence of noise and outliers in the measurements increase the time to detect damage and for large levels of noise, the methodology completely fails to detect anomalies as shown in Table 1.

This study has also compared the performance of the proposed methodology with moving principal component analysis (MPCA) of the response time histories as previously proposed by Posenato et al. [31]. These are presented in Table 1. Results illustrate the superior performance of the proposed methodology over the MPCA-based approach. The MPCA approach fails to detect damage in all scenarios except for the ones where the intensity of damage is the strongest i.e. a reduction of 30% of material stiffness. Moreover, the evidence for occurrence of an anomaly may also be weak, i.e., the threshold is exceeded only briefly and the eigenvectors do not clearly indicate anomalous behaviour by transitioning to a new stable state as would be expected. An example of such behaviour is illustrated for scenario D1O1 in Figure 15.

Table 1: Time (days) to anomaly detection of the proposed methodology and MPCA [31] for a range of scenarios

Algorithm	Noise and outlier scenario	Damage scenario				
		D1	D2	D3	D4	D5
Proposed approach/ MPCA [31]	-	7/4	25/x	19/x	81/x	126/x
	O1	5/70*	22/x	21/x	79/x	116/x
	O2	8/x	17/x	42/x	105/x	139*/x
	O3	25/x	9/x	75/x	80/x	129/x
	N1	3/62	106/x	153/x	140/x	105/x
	N2	105/52	276/x	225/x	102/x	297/x
	N3	71/x	151/x	x/x	x/x	436*/x
	O1N1	43/56	129/x	x/x	294/x	265/x
	O1N2	159/89	x/x	x/x	x/x	x/x
	O1N3	242/x	x/x	x/x	x/x	x/x

* – weak evidence of anomalous behaviour
x – failure of algorithm to detect anomaly

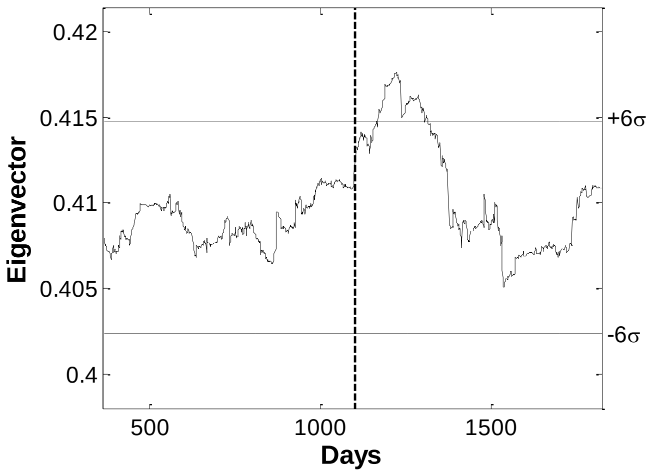

Figure 15: Plot of the component corresponding to sensor S-2 in the first principal component from MPCA of strain measurements for scenario D1O1

6. Conclusions

Conclusions from this study are as follows:

- The relationship between distributed temperature and response measurements can form the basis for anomaly detection techniques that are faster and more accurate than the interpretation of the response time histories using MPCA.
- SVR models can be trained to accurately predict the thermal response of a structure from distributed temperature measurements.

- The prediction error, which is the difference between a prediction from a SVR model and a corresponding measurement, is a reliable indicator of damage. The time series of prediction errors can be analyzed by MFFT for anomaly detection.
- The proposed methodology that combines SVR and MFFT is shown to reliably detect anomalous structural behaviour from distributed response and temperature measurement in the presence of outliers and measurement noise.

Future research will evaluate the developed methods on measurements from laboratory and full-scale structures. Work is also underway on extending these approaches to find the location of damage. Further investigation is required on the sensitivity of the SVR-based approach for anomaly detection to tuning parameters such as *v*. A long-term research goal is to combine the developed methods with strategies that identify traffic loads on the structure.

7. References

[1] G. Cole, The role of assessments in highway bridge management, Proceedings of the ICE - Bridge Engineering, 161 (2008) 133-139.

[2] FHWA Bridge Programs, Tables of Frequently Requested NBI Information, http://www.fhwa.dot.gov/bridge/deficient.cfm, (Accessed: September 2012).

[3] J. Brownjohn, Structural health monitoring of civil infrastructure, Philos Transact A Math Phys Eng Sci. , 365 (2007) 589–622.

[4] Y.Q. Ni, H.F. Zhou, K.C. Chan, J.M. Ko, Modal Flexibility Analysis of Cable-Stayed Ting Kau Bridge for Damage Identification, Computer-Aided Civil and Infrastructure Engineering, 23 (2008) 223-236.

[5] J.P. Lynch, An overview of wireless structural health monitoring for civil structures, Philosophical Transactions of the Royal Society A, 365 (2007).

[6] E. Sazonov, H. Li, D. Curry, P. Pillay, Self-Powered Sensors for Monitoring of Highway Bridges Sensors Journal, IEEE 9(2009) 1422 - 1429

[7] S.D. Glaser, H. Li, M.L. Wang, J. Ou, J. Lynch, Sensor technology innovation for the advancement of structural health monitoring: a strategic program of US-China research for the next decade., Smart Structures and Systems, 3 (2007) 221-244.

[8] RAE, Smart infrastructure: the future, in, The Royal Academy of Engineering, 2012.

[9] L. Ljung, System Identification - Theory for the User, 2nd edition ed., Prentice Hall, 1999.

[10] A.E. Aktan, F.L. Moon, F.N. Catbas, Long-Term Vision for the ASCE Technical Committee: Structural Identification of Constructed Systems, in: Proceedings of the 3rd International Conference on Structural Health Monitoring of Intelligent Infrastructure, Vancouver, British Columbia, Canada, 2007.

[11] Y. Xia, B. Chen, S. Weng, Y.-Q. Ni, Y.-L. Xu, Temperature effect on vibration properties of civil structures: a literature review and case studies, Journal of Civil Structrual Health Monitoring, 2 (2012) 29-46.

[12] M. Sanayei, S. Scampoli, Structural Element Stiffness Identification from Static Test Data, Journal of Engineering Mechanics, 117 (1991) 1021-1036.

[13] Sanayei, G. Imbaro, J. McClain, Brown, Structural Model Updating Using Experimental Static Measurements, Journal of Structural Engineering, 123 (1997) 792-798.

[14] Y. Robert-Nicoud, B. Raphael, O. Burdet, I.F.C. Smith, Model Identification of Bridges Using Measurement Data, Computer-Aided Civil and Infrastructure Engineering, 20 (2005) 118-131.

[15] J.L. Beck, K.-V. Yuen, Model Selection Using Response Measurements: Bayesian Probabilistic Approach, Journal of Engineering Mechanics, 130 (2004) 192-203.

[16] J.A. Goulet, P. Kripakaran, I.F.C. Smith, Multimodel Structural Performance Monitoring, Journal of Structural Engineering-Asce, 136 (2010) 1309-1318.

[17] B. Moaveni, I. Behmanesh, Effects of changing ambient temperature on finite element model updating of the Dowling Hall Footbridge, Engineering Structures, 43 (2012) 58-68.

[18] O.S. Salawu, Detection of structural damage through changes in frequency: a review, Engineering Structures, 19 (1997) 718-723.

[19] N. Kulprapha, P. Warnitchai, Structural health monitoring of continuous prestressed concrete bridges using ambient thermal responses, Engineering Structures, 40 (2012) 20-38.

[20] J.M.W. Brownjohn, A.D. Stefano, Y.-L. Xu, H. Wenzel, A.E. Aktan, Vibration-based monitoring of civil infrastructure: challenges and successes, Journal of Civil Structural Health Monitoring, 1 (2011) 79-95.

[21] P. Moyo, J. Brownjohn, Detection of anomalous structural behaviour using wavelet analysis, Mechanical Systems and Signal Processing, (2002) 429-445.

[22] C.R. Farrar, K. Worden, An introduction to structural health monitoring, Philosophical Transactions of the Royal Society A, 365 (2007) 303-315.

[23] E. Carden, J. Brownjohn, ARMA modelled time-series classification for structural health monitoring of civil infrastructure, Mechanical Systems and Signal Processing, (2008) 295-314.

[24] K.V. Yuen, S.C. Kuok, Ambient interference in long-term monitoring of buildings, Engineering Structures, 32 (2010) 2379-2386.

[25] X.G. Hua, Y.Q. Ni, J.M. Ko, K.Y. Wong, Modeling of Temperature--Frequency Correlation Using Combined Principal Component Analysis and Support Vector Regression Technique, Journal of Computing in Civil Engineering, 21 (2007) 122-135.

[26] F.N. Catbas, M. Susoy, D.M. Frangopol, Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data, Engineering Structures, 30 (2008) 2347-2359.

[27] K.Y. Koo, J.M.W. Brownjohn, D.I. List, R. Cole, Structural health monitoring of the Tamar suspension bridge, Structural Control and Health Monitoring, (2012) n/a-n/a.

[28] E. Mirambell, A. Aguado, Temperature and Stress Distributions in Concrete Box Girder Bridges Journal of Structural Engineering 116 (1990) 2388-2409.

[29] S. Moorty, C.W. Roeder, Temperature Dependent Bridge Movements, Journal of Structural Engineering, 118 (1992) 1090.

[30] K.-V. Yuen, S.-C. Kuok, Ambient interference in long-term monitoring of buildings, Engineering Structures, 32 (2010) 2379–2386.

[31] D. Posenato, P. Kripakaran, D. Inaudi, I.F.C. Smith, Methodologies for model-free data interpretation of civil engineering structures, Computers and Structures, 88 (2010) 467-482.

[32] D. Posenato, F. Lanata, D. Inaudi, I.F.C. Smith, Model-free data interpretation for continuous monitoring of complex structures, Advanced Engineering Informatics, 22 (2008) 135–144.

[33] I. Laory, T.N. Trinh, I.F.C. Smith, Evaluating two model-free data interpretation methods for measurements that are influenced by temperature, Advanced Engineering Informatics, 25 (2011) 495-506.

[34] V. Chandola, A. Banerjee, V. Kumar, Anomaly Detection : A Survey, ACM Computing Surveys, 41 (2009) 1-58.

[35] K.A. Heller, K.M. Svore, A.D. Keromytis, S.J. Stolfo, One Class Support Vector Machines for Detecting Anomalous Windows Registry Accesses, In Proc. of the workshop on Data Mining for Computer Security, (2003).

[36] S.J. Ray, J. Teizer, Coarse head pose estimation of construction equipment operators to formulate dynamic blind spots, Advanced Engineering Informatics, 26 (2012) 117-130.

[37] S. Shengchao, D. Na, H.Y. Gang, An SVR-Based Online Fault Detection Method in: 2011 Third International Conference on Measuring Technology and Mechatronics Automation, Shangshai, 2011, pp. 447-450.

[38] M.N. Noori, Y. Cao, Z. Hou, S. Sharma, Application of Support Vector Machine for Reliability Assessment and Sturctural Health Monitoring, International Journal of Engineering Under Uncertainty: Hazards, Assessment and Mitigation, 2 (2010) 89-98.

[39] S. Saitta, P. Kripakaran, B. Raphael, I.F.C. Smith, Feature Selection Using Stochastic Search: An Application to System Identification, Journal of Computing in Civil Engineering 24 (2010) 3-10.

[40] Y.Q. Ni, X.G. Hu, K.Q. Fan, J.M. and Ko, Correlating modal properties with temperature using long-term monitoring data and support vector machine technique, Engineering Structures, 27 (2005) P. 1762-1773.

[41] B. Schölkopf, J. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, Neural Computation, 13 (2001) 1443-1471.

[42] C.-C. Chang, C.-J. Lin, Training v-Support Vector Regression: Theory and Algorithms, Neural Computation, 14 (2002) 1959-1977.

[43] ANSYS, ® Academic Research, Release 13.0, Help System, Mechanical APDL, Element Reference, in: ANSYS, Inc., 2011.

[44] A.M.G. Klein Tank, J.B. Wijngaard, G.P. Können, R. Böhm, G. Demarée, A. Gocheva, M. Mileta, S. Pashiardis, L. Hejkrlik, C. Kern-Hansen, R. Heino, P. Bessemoulin, G. Müller-Westermeier, M. Tzanakou, S. Szalai, T. Pálsdóttir, D. Fitzgerald, S. Rubin, M. Capaldo, M. Maugeri, A. Leitass, A. Bukantis, R. Aberfeld, A.F.V. van Engelen, E. Forland, M. Mietus, F. Coelho, C. Mares, V. Razuvaev, E. Nieplova, T. Cegnar, J.A. López, B. Dahlström, A. Moberg, W. Kirchhofer, A. Ceylan, O. Pachaliuk, L.V. Alexander, P. Petrovic, Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment, International Journal of Climatology, 22 (2002) 1441-1453.

[45] C.-C. Chang, C.-J. Lin, LIBSVM : a library for support vector machines, ACM Transactions on Intelligent Systems and Technology, 2 (2011) 27:21--27:27.

[46] B.G. Sherlock, D.M. Monro, Moving discrete Fourier transform, Radar and Signal Processing, IEE Proceedings F 139 (1992).