Structural Health Monitoring of a Footbridge using Echo State Networks and NARMAX

Adam J Wootton[1,2], John B Butcher*[1], Theocharis Kyriacou[1], Charles R Day[1], Peter W Haycock[2]

[1]*School of Computing and Mathematics, Keele University, Staffordshire, ST5 5BG, UK*

[2]*Foundation Year Centre, Keele University, Staffordshire, ST5 5BG, UK*

Corresponding author: j.b.butcher@keele.ac.uk, +44 (0)1782 733264

Abstract:

Echo State Networks (ESNs) and a Nonlinear Auto-Regressive Moving Average model with eXogenous inputs (NARMAX) have been applied to multi-sensor time-series data arising from a test footbridge which has been subjected to multiple potentially damaging interventions. The aim of the work was to automatically classify known potentially damaging events, while also allowing engineers to observe and localise any long term damage trends. The techniques reported here used data from ten temperature sensors as inputs and were tasked with predicting the output signal from eight tilt sensors embedded at various points over the bridge. Initially, interventions were identified by both ESNs and NARMAX. In addition, training ESNs using data up to the first event, and determining the ESNs' subsequent predictions, allowed inferences to be made not only about when and where the interventions occurred, but also the level of damage caused, without requiring any prior data pre-processing or extrapolation. Finally, ESNs were successfully used as classifiers to characterise various different types of intervention that had taken place.

Keywords:

Echo State Networks, NARMAX, wireless sensor networks, structural health monitoring, bridges, NPL Footbridge

1 INTRODUCTION

As non-destructive sensor networks are increasingly used as a means of monitoring the long term health of structures, computational methods for the analysis of sensor data become increasingly important. With the cost of the repair and upgrading of deficient bridges in the US alone estimated to be $121 billion (Civil Engineers (ASCE), 2013), there is a need for approaches that increase the serviceable life of structures and reduce repair costs. Consequently, a method for providing structural engineers with a picture of how the state of a structure changes over time that is also capable of detecting the onset of damaging events would be of great value.

The contribution of this study is the presentation of two different machine learning methods, Echo State Networks (ESNs) (Jaeger, 2010) and Nonlinear Auto-Regression Moving Average model with eXogenous inputs (NARMAX) (Billings and Chen, 1998), for the detection, localisation and classification of damage to a real-world civil engineering structure. This structure was subjected to a number of deliberate controlled interventions which were likely to undermine the structure's integrity.

These methods used the multi-dimensional, longitudinal, time-series dataset composed of readings that were provided by sensors embedded in the structure. Past work by the authors has seen ESNs successfully applied to two different SHM case studies, one of which was the footbridge discussed here (Wootton et al., 2014, 2015). The current paper builds on this work by introducing two further ESN approaches and comparing all of these to the NARMAX

technique and previous approaches which have been applied to the same structure (see Section 1.1.2).

1.1 The UK National Physical Laboratory Footbridge

The UK National Physical Laboratory (NPL) footbridge project was set up as a UK-wide means of developing new sensor technologies and methods for the processing of large time-series datasets from wireless sensor networks. The datasets that have been obtained from the footbridge project and used in this work have also been made widely available to other research groups. The project was centred on a concrete footbridge that was built in the 1960s and underwent normal use for nearly 50 years prior to the beginning of the project. In 2009 it was taken out of use and embedded with a number of sensors, which took readings over three years at regular five minute intervals, or shorter intervals during interesting time periods. Details of the motivation behind the project and the sensors used have been published extensively elsewhere and so are not discussed here ("Footbridge Monitoring Project (SHM) - Background," n.d.; NakedScientists, n.d.; Barton and Zhang, 2010; Barton, 2011; Barton et al., 2011).

The work in this paper is concerned with the data produced by the ten temperature sensors (embedded in vibrating wire arc weldable strain gauges) and eight electrolevel tilt sensors, which consisted of 365,376 data points collected between January 2009 and May 2012. All of the tilt and temperature sensors were provided by ITMSOIL (ITMSOIL, 2016). Figure 1 shows the spatial arrangement of the temperature and tilt sensors on the bridge and the bridge in-situ. Note that tilt sensors 7 and 8 are attached to the two piers of the bridge and that it is a standalone structure, allowing weights to be suspended from the cantilever where sensor 1 is located.

The temporal relationship between the tilt and temperature sensor data can be seen in Figure 2, which shows the data produced by tilt sensor 1 and temperature sensor 1 between 14th March 2009 and 17th March 2009, before the first significant intervention occurred. It can be seen that under normal circumstances, the bridge undergoes a daily cycle due to changes in temperature. Each day, the temperature increased and peaked at around 2 pm, before then gradually reducing, producing a characteristic daily spike. The tilt sensor data followed this, since the tilt of the bridge at that point increased as the temperature increased and peaked as the temperature peaked. As the bridge cooled, the tilt sensor reading gradually reduced again. There is a clear, observable temporal relationship between the two sets of data.

1.1.1 Deliberate damage to the footbridge

During the course of the study, the bridge was subjected to damage and repair cycles, detailed in full by Livina et al. (2013). Two key types of damage cycle are considered in this paper: static tests and fatigue tests. A typical static test involved suspending a load of several tons of water from the end of the bridge where tilt sensor 1 was located, while fatigue tests simulated damage using an adapted hydraulic system that performed half a million load controlled, 10 kN/s cycles. The purpose of these tests was to displace the bridge to its serviceability limits, induce tensile strain sufficient to cause cracking and to allow assessment of the performance of the deteriorating structure (Worden et al., 2012). A full list of the static tests and fatigue tests is given in Table 1.

1.1.2 Past research on the footbridge

There have been other studies which have analysed the data provided by the project, but as of yet none has been able to detect events accurately, either spatially or temporally, or to characterise long term damage to the bridge. Barton & Esward (2012) investigated errors in

the sensor data, but were not able to assess long term damage. Kromanis & Kripakaran (2014) used a support vector regression technique to model tilt sensor behaviour with some accuracy, but this required significant pre-processing and again did not give any indication about the condition of the footbridge; both the ESN and NARMAX methods described in this paper manage to model the behaviour of the tilt sensors and, hence, show long term damage trends. While later work using support vector regression and moving principal component analysis allowed for anomaly detection (Kromanis, 2015), the ESNs used here were able to identify specific types of anomalous event without any major post-processing procedures. After going through a process of deseasonalising and detrending the temperature data, Livina et al. (2013) managed to detect 'early warning indicators' when significant interventions occurred, but were not able to quantify the damage caused by these events. Worden et al. (2012) used cointegration on the tilt sensor data so as to purge environmental effects and detect damage. However, once the condition of the bridge changed, the data were no longer purged of environmental data and retraining would be required.

Similar problems have been encountered in other past approaches to monitoring the condition of bridges with sensors. A number of papers have been published on a similar real world bridge monitoring project, where the Tamar suspension bridge in Plymouth, UK, was embedded with a range of different sensors and data taken over a four-year period. However, in one of the most recent papers on this project, it was reported that processing the data was difficult, due to problems in finding the 'normal' behaviour of the bridge (Koo et al., 2013). In another paper, it was found that random forest and support vector regression could predict the natural frequency of the bridge well, but potential damage to the bridge was not assessed (Laory et al., 2014). In the case of the Z24 bridge, much of the work dealt with removing environmental effects from the data (Dervilis et al., 2016), in particular temperature effects (Worden et al., 2013).

An advantage of the ESN methods presented in this work is that the data require no significant pre-processing: simply normalising the data between 1 and -1 prior to presenting it to an ESN will suffice. Furthermore, no prior knowledge about the bridge, such as Young's modulus of the concrete, is required. Finally, since ESNs can be used for both regression and classification, it is possible to use two ESNs in parallel, one of which allows the user to observe long term damage trends and the other of which will flag up signals in the data that are suggestive of a damage event.

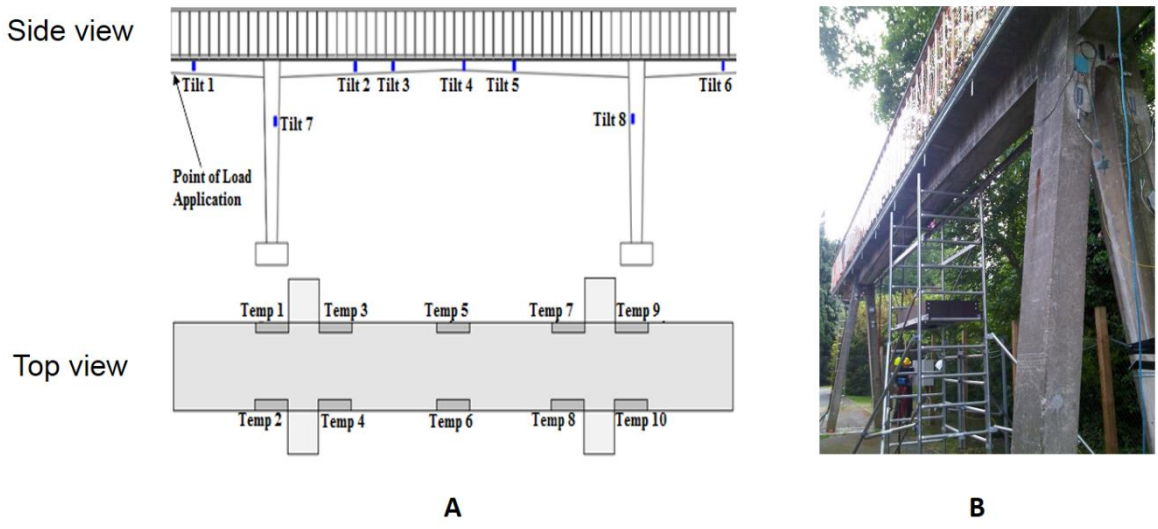

Fig. 1 A: A schematic of the footbridge showing the spatial layout of the eight tilt sensors on the footbridge (top) and the ten temperature sensors (bottom). **B**: A picture of the footbridge in-situ.

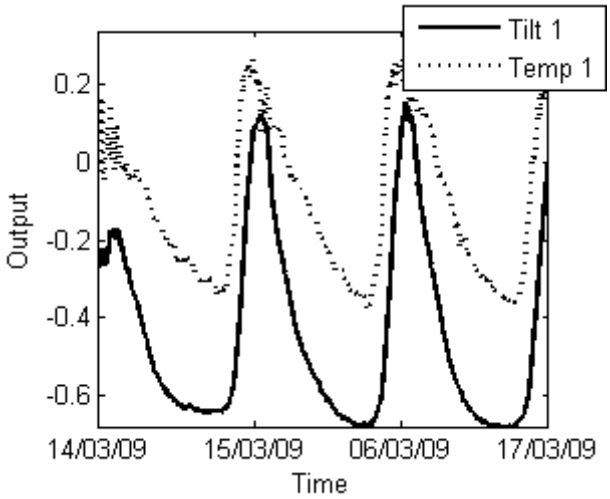

Fig. 2 The normal behaviour of the footbridge over the period 14th March - 17th March 2009. A clear correlation between the tilt sensor 1 reading and the temperature sensor 1 reading can be observed.

1.2 Novelty and Summary of Paper

The chief contribution of this work lies in the use of the ESN to detect long term damage to the structure studied here, in addition to detecting days when damaging events occurred. To the best of the authors' knowledge, no other approach has managed this to date. Furthermore, this was achieved using data that had no prior preprocessing, save for normalisation between +1 and -1. Rather than purging environmental effects, the use of ESNs and NARMAX allowed for the environmental effects to be embraced as a feature of the dataset.
The remainder of the paper is organised as follows. Section 2 gives the background on the ESN and NARMAX approaches. Section 3 presents the four different methodologies used here, while Section 4 discusses the results from these methodologies. Section 5 concludes the paper.

2 THEORETICAL BACKGROUND

2.1 Echo State Networks (ESNs)

ESNs (Jaeger, 2010) are a form of artificial recurrent neural network (RNN) that offer a fast and efficient training procedure, making their application to real-world data very appealing and allowing them to overcome the problems usually associated with RNN training procedures (Lukoševičius and Jaeger, 2009). They have recently been applied in areas such as human activity recognition (Palumbo et al., 2016), hydraulic pump prognostics (Sun et al., 2016) and the recognition of forehands in tennis (Bacic, 2016). In addition, ESNs have previously been used in a structural health monitoring context, having been used for fault diagnosis in a water network (Quevedo et al., 2014) and defect detection in reinforced concrete (Butcher et al., 2014). In all of these studies, ESNs were found to equal, or outperform, alternative state-of-the-art techniques. ESNs were chosen for this study due to their potential ability to capture the temporal relationship between the temperature sensors and the target output tilt sensors.
An ESN architecture involves an input layer, a sparsely and randomly interconnected layer (the dynamical reservoir) and an output layer, each of which is connected in the forward direction to its neighbouring layers. Unlike most artificial neural networks, only the weights

on the connections between the reservoir neurons and the output units are trained. All other weights are randomly generated at the start of the process and are left unchanged after initialisation. The dynamical reservoir has recurrent connections, allowing it to possess a short term memory. This means that at any given time *t*, the output values are affected not just by the inputs at that time step, but at all of the previous time steps as well. Figure 3 shows an example of an ESN architecture.

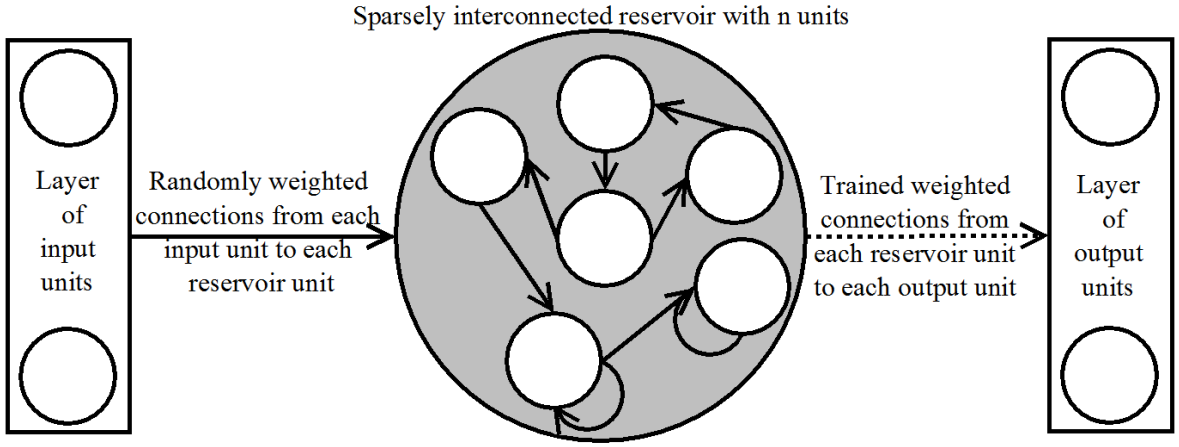

Fig. 3 One of the ESN architectures used, with ten input units feeding into a dynamical reservoir, which in turn feeds into eight output units. Solid arrows indicate weighted connections, while dotted arrows indicate weighted connections that are adjusted during the training process. Note: the number of units in the figure is for illustration purposes.

Although the input and internal weights are randomly generated, there is a number of parameters that can be adjusted in order to tailor an ESN's behaviour to provide improved performance in different applications. The activation of the ESN reservoir neurons is evaluated as follows:

$$\mathbf{x}(t) = f((1-\delta)\mathbf{x}(t-1) + \delta(\mathbf{W}_{res}^{inp}\mathbf{u}(t) + \mathbf{W}_{res}^{res}\mathbf{x}(t-1))) \quad (1),$$

where $\mathbf{W}_{res}^{inp}$ is the input to reservoir weight matrix, $\mathbf{x}(t)$ is the vector of the activations of the reservoir neurons at time *t*, which is the current time step, $t-1$ is the previous time step, $\mathbf{W}_{res}^{res}$ is the reservoir weight matrix (drawn randomly from a Z distribution), $\mathbf{u}(t)$ is the vector of the input data at time *t* and δ is the leak rate, which determines the extent to which ESN reservoir neurons' activations decrease over a period of time. The leak rate is one of the parameters that can be tuned and varies inversely with the extent of the ESN's short term memory, with smaller values increasing the ability of reservoir neurons to recall inputs presented further in the past, but decreasing their ability to recall the most recent inputs (Verstraeten, 2009).

The internal dynamics of the reservoir are largely dependent on the spectral radius of an ESN, α. The value for α is another parameter that can be set at the start of the training process and adjusted to improve ESN performance. Generally speaking, a small value of α will also contribute to a shorter memory and a reservoir that forgets values more quickly, while an ESN with a larger value of α no greater than 1 will be able to recall further into the past (Butcher et al., 2013; Jaeger, 2013). The dependence of the final internal reservoir weights on the spectral radius α can be seen in the following equation (Tong et al., 2007):

$$\mathbf{W}_{res}^{res} = \alpha \times \mathbf{W}'^{res}_{res} / |\lambda_{\max}| \quad (2),$$

where λ_{max} is the maximum eigenvalue of $\mathbf{W}_{res}^{res}$, which represents the initial reservoir weights. The optimal value of α was found using cross validation.
Adjusting the scaling of the input weights allows control over the extent to which the inputs at any given time drive the output of the network, with very high input scaling causing the reservoir neurons to act in an almost binary switching manner. Generally, a high input scaling value leads to highly non-linear network behaviour, while a low input scaling value leads to almost linear behaviour (Butcher et al., 2013). Finding the ideal value for these parameters for any given application is usually achieved by performing a grid search.
Once the activations have been calculated for all of the data in a particular training dataset, the output weights can be trained using a simple linear regression technique, such as ridge regression (Montgomery and Peck, 1982):

$$\mathbf{W}_{out}^{res} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{Y}_{tgt} \quad (3),$$

where λ is a regularisation parameter, **I** the identity matrix, **X** the matrix of all reservoir neurons' activations over the length of the data, $\mathbf{X}^T$ the transpose of the activations matrix and $\mathbf{Y}_{tgt}$ the target output. The output of the network is then calculated:

$$\mathbf{y}(t) = f^{out}(\mathbf{W}_{out}^{res}(\mathbf{x}(t))) \quad (4),$$

where f^{out} is the activation function of the output unit, which is linear. In this work, the ESNs were simulated using the Reservoir Computing Toolbox for MATLAB (Verstraeten et al., 2007).

2.2 The NARMAX model estimation methodology

NARMAX is a polynomial function that can be used to represent the input/output relationship between the inputs, as represented by **u**, and the output, as represented by y, of a Multiple-Input/Single-Output (MISO) non-linear system. The aim of the NARMAX model estimation methodology is to determine the terms (the structure) and the coefficients of these terms (the parameters) for the polynomial using input/output examples of the system under investigation. The purpose of the estimated model is twofold: to predict the output for new or unseen values of the input and to characterise the system by studying the model structure and parameter values. This model estimation methodology has been used in several domains (Billings and Chen, 1998; Iglesias et al., 2007; Kyriacou et al., 2008).
For a noisy MISO system, the output *y* at discrete time *t* can be generally represented by:

$$y(t) = f(\mathbf{u}_i(t-j)^p, \mathbf{e}_i(t-k)^p, \mathbf{y}(t-m)^p) \quad (5),$$

where:

$$i \in (\{\mathbb{N}_+\} \mid i \leq d)$$
$$j \in (\{\mathbb{N}_0\} \mid j \leq N_u)$$
$$p \in (\{\mathbb{N}_+\} \mid p \leq l)$$
$$k \in (\{\mathbb{N}_0\} \mid k \leq N_e)$$
$$m \in (\{\mathbb{N}_0\} \mid m \leq N_y)$$

In equation 5, $\mathbf{u}(t)$ and $\mathbf{e}(t)$ are the sampled input and error signals at time *t*, respectively, N_u, N_e and N_y are the regression orders of the input, error and output respectively, and *d* is the dimension of the input. *f*() is a polynomial expansion of the arguments over all possible

combinations of *i, j, p, k* and *m*, subject to the degree, *l*, of the polynomial being the highest sum of powers over all of its terms. Hence, the degree provides a limit to the polynomial expansion.
The NARMAX methodology is an iterative process that tries to minimize the difference between the actual and model predicted output by changing the structure and parameters of the model (Korenberg et al., 1988). Two sets of data are used: one for estimating the model structure and its parameters and the other for validating the model. During each iteration of the algorithm the structure of the model polynomial is changed by removing non-contributing terms and the remaining model parameters are re-calculated so that the best possible fit is achieved to the estimation set output. This iterative process stops when all remaining polynomial terms are considered to be significant contributors to the calculation of the output. In other words, the removal of any of the remaining terms would cause the prediction error of the model on the estimation data set to increase beyond a preset acceptable limit.
The significance of a model term is expressed using its corresponding Error Reduction Ratio (ERR). This value is calculated for every term as part of the NARMAX estimation process. The ERR is an indication of the reduction in the model's prediction error that occurs when the model term considered is introduced into the model. This reduction is expressed in proportion to the maximum error (a constant) that results from removing all the terms from the model. The value of the ERR is therefore proportional to the significance of the term to which it corresponds.
The initial structure of the NARMAX polynomial is determined by d, N_u, N_y, N_e, and l. The number of terms in the initial model can be very large depending on these variables; however, only a few of them will be left in the final model as most of them will have negligible or no effect on the error (very small ERR).
After the iterative estimation process ends, the resulting NARMAX model is tested using a validation data set. The true performance of the model is assessed in this test.

3 Materials and Methods

3.1 NARMAX and ESN_α: detecting manual interventions

Both the NARMAX and the ESN approaches were applied to the NPL bridge dataset, using the ten temperature sensors as their input data (independent variables), and the eight tilt sensor readings as their output (dependent variables). The ESN used at this stage of the study is henceforth referred to as ESN_α. ESN_α and NARMAX were, therefore, given the task of learning the relationship between the temperature of different parts of the bridge and the tilt at eight different points on the bridge. The hypothesis was that any significant deviation between the output of the two models and the actual tilt sensors from the bridge would be an indication of the presence of an anomalous structural change within the bridge. This deviation could then be used to alert an engineer to inspect the bridge for potential defects.
The first 70% of the data was used to train both of the models, with the remaining 30% used for validation. In order to reduce the size of the dataset, one sample per hour was extracted from the dataset spanning the years 2009, 2010 and 2011. The best NARMAX model was found to be a linear model (degree 1) and used a lag of 12 (i.e. a maximum delay between input and output data of 12 hours). The data presented to the ESN were the same as used by the NARMAX model but were also normalised over the range -1 to +1: a commonly used preprocessing technique that delivers improved neural network performance (Lukoševičius , 2012). As the weights of the ESN are random at network creation, 100 ESNs were simulated using the parameters that enabled best performance and the outputs for each tilt sensor were

averaged. The trained models were presented with the whole dataset as their input data and were required to give the tilt sensor readings across the entirety of the data as their output.

3.2 ESN_β: detecting lasting damage

It was expected that the approach laid out in Section 3.1 would work well for the detection of significant interventions, but would not show whether or how the bridge had been damaged as a consequence of the intervention. This was partly due to the fact that a number of significant interventions took place in the portion of data that was used for training. This meant that the ESNs were potentially being trained to treat the intervention-modified behaviour of the tilt sensors due to damage as part of the typical behaviour of the bridge. Therefore, the approach in this second stage of the work was to use only the data prior to the first significant intervention for training and then to apply the trained ESNs to the remainder of the data. This approach means that the ESNs used, which are referred to as ESN_β for the remainder of the paper, were trained on the behaviour of the tilt sensors only under normal conditions. As the portion of the data available for training was now limited to that taken prior to the first significant intervention (only 12.79% of the full dataset), no sampling was performed and the full dataset was instead used. The data were again normalised between -1 and +1, with no other pre-processing performed.

An ESN_β topology that was found to deliver best performance was used and 100 ESNs using this topology were simulated and their output for each tilt sensor averaged. The error between the output predicted by ESN_β and the actual tilt sensor data at each point was calculated and then a moving average of these values was taken. This average error method was employed to address the noise that might otherwise lead to useful information being obscured. Due to the high resolution of the data used, the moving average was taken over 10,000 points, equivalent to approximately a month's worth of data. 10,000 points was found to be an optimal window size for removing the noise without introducing a significant lag into the results.

The revised training regime outlined here meant that ESN_β would recognise normal patterns of data, but would also be better able to detect unusual events. Specifically, a significant divergence between the data recorded by each tilt sensor and the normal behaviour of the bridge as predicted by ESN_β was taken as a change in the error of at least 0.01; this is used here to indicate that something of note had occurred. It should be noted that when there was a discrepancy of at least 0.01, the full duration of this discrepancy was noted, since a brief discrepancy may have been simply an initial response to an intervention, while a discrepancy maintained over several months may indicate permanent damage. In this analysis, it was hypothesised that the greater the magnitude of the discrepancy between real and predicted sensor data, the greater the level of damage. Some ground truth about the date and type of events was known (see Table 1), making it possible to localise the onset and end of these events in time and to verify the findings. However, the extent of any long lasting damage caused by an intervention was not known.

3.3 ESN_γ and ESN_δ: classification of interventions

In addition to determining if damage had occurred, an attempt was also made to identify specific types of intervention. Figure 2 shows the daily cycle of the bridge under normal conditions. Figure 4 shows how this daily cycle could be perturbed by a static test. In the period shown in Figure 4, two static tests occurred, one on 24th June 2011 and one on 27th June 2011. The signal from tilt sensor 1 on these two days is different from the normal tilt sensor daily cycle in the bridge: there is a second spike seen shortly after the usual peak in the temperature data, due to the weight of the load, applied after 2 pm, causing the bridge to tilt

significantly before returning to its normal tilt once the load had been released. This characteristic ‘double spike’ shape can be seen for all of the 22 static tests that were performed, and was a basis for expecting an ESN technique to characterise the intervention. Similarly, Figure 5 shows how the normal cycle was perturbed by a fatigue test. During the period shown in Figure 5, fatigue tests occurred on 28th September, 29th September, 2nd October and 3rd October 2011. On these days, the normal daily cycle of the bridge was perturbed as the bridge was kept in a stressed state for several hours, prompting a sharp rise in the value of the tilt sensor readings at the onset of the event and a sharp drop in the value of the tilt sensor readings at the end of the event. This characteristic shape can be seen for all eight fatigue tests that were performed. Again, this was a basis for expecting an ESN technique to characterise the intervention.

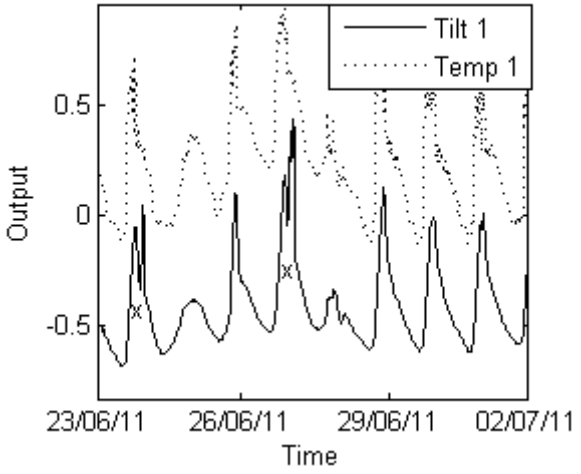

Fig. 4 Tilt sensor 1 and temperature sensor 1 over the period 23rd June to 2nd July 2011. Static tests with a characteristic 'double spike' can be observed on 24th June and 27th June and are marked with an 'x'.

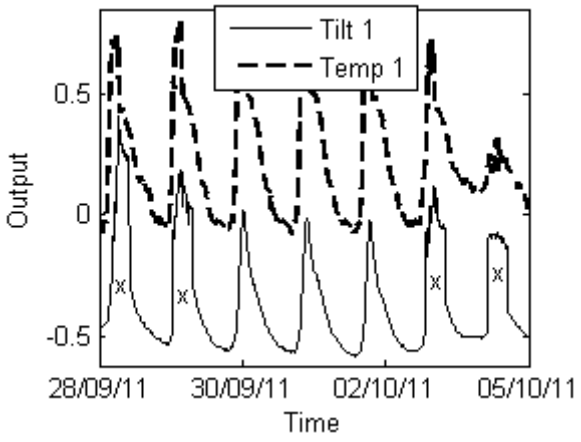

Fig. 5 Tilt sensor 1 and temperature sensor 1 over the period 26th September to 5th October 2011. The characteristic shape of the fatigue tests can be observed on 28th September, 29th September, 2nd October and 3rd October and are labelled with an 'x'.

Since the two primary types of intervention both had recognisable signatures in the data, an experiment was set up using ESNs to determine automatically the type of intervention using the data from all eight tilt sensors as inputs. A new ESN was set up with three output units to classify each data point as either part of the bridge’s normal cycle, a static test or a fatigue test, respectively. As with ESN_α and ESN_β, two separate training regimes were employed, producing ESN_γ and ESN_δ. The two separate training regimes are detailed in Figure 6. ESN_γ

was presented with the full data for the 22 days when static tests were performed, the eight days when the fatigue tests were performed and 32 randomly selected days when no intervention occurred. During testing, the ESNs were presented with the entire three-year dataset in a continuous stream. Although training the ESNs to recognise all of the available static and fatigue tests is unconventional, the approach has value in demonstrating that the ESNs are still able to detect static and fatigue tests in a dataset where the majority of the data pertains to ordinary days, without misclassifying these ordinary days as days of interest. The approach of ESN_{δ} took this a stage further, as only six static test days and 15 fatigue test days were included as part of the training dataset, meaning that there were 9 genuinely unseen test patterns contained within the full dataset that was presented to the ESNs. In both of the ESN_{γ} and ESN_{δ} training regimes, each output unit was set up to give a value of +1 whenever either a static test, fatigue test or normal behaviour was detected respectively, and a value of -1 at all other times. A moving window of one day was used for assessing the performance of $ESN\gamma$ and ESN_{δ}. This meant that if at any point during a single day the static test classification node produced a value of '+1', that day would be considered as a day on which the ESN suggested a static test had occurred. This was done because of the large ratio of ordinary days to significant damage events in the dataset. The motivation behind this approach was to demonstrate the ability of ESNs not only to distinguish between anomalous behaviour and ordinary behaviour, but also between different types of anomalous behaviour. The results of the classification task were then analysed using the classification accuracy measures described by Baldi et al (2000), each of which has its own advantages and disadvantages. The sensitivity of a classifier gives the probability of correctly predicting a positive sample, while the specificity gives the probability of correctly predicting a negative sample. These are calculated according to equations 6 and 7.

$$Sensitivity = TP/(TP + FN) \quad (6),$$

$$Specificity = TN/(TN + FP) \quad (7),$$

In equations 7 and 8, TP refers to the number of true positives, FN refers to the number of false negatives, FP refers to the number of false positives and TN to the number of true negatives. The positive predictive value (PPV), which is the proportion of true positives in data points classified as a positive, and the negative predictive value (NPV), the proportion of true negatives in the total number of data points classified as a negative, were also used. These were calculated using equations 8 and 9.

$$PPV = TP/(TP + FP) \quad (8),$$

$$NPV = TN/(TN + FN) \quad (9),$$

Fawcett suggests that Receiver Operator Characteristic (ROC) curve analysis provides a richer measure of classification accuracy due to de-coupling of classifier performance from class skew and error costs (Fawcett, 2006). For this reason, a graph of sensitivity against false positive rate (FPR) was plotted for each output node classification and the area under the curve (AUC) calculated. A value of close to unity for the AUC is indicative of very good classification, while any value less than 0.5 suggests that the results from the classifier are little better than guessing. It should be noted that when the AUC was calculated, the moving window of one day was not used and the output of each node for each individual point of data was instead used. This was because the ROC curve was used to determine the most appropriate threshold for each output node and the moving window was used only after the threshold had been applied. The FPR was calculated in accordance with equation 10.

$$FPR = FP/(FP + TN) \quad (10),$$

This approach to arriving at a threshold for each node was necessary due to the nature of ESNs. The chief advantage of using ESNs for SHM is that rather than having a simple threshold, they possess a recurrent short term memory that provides a kind of context sensitive model of the underlying system.
Since the Matthews Correlation Coefficient (MCC) (Matthews, 1975) takes into account TP, FP, TN and FN, rather than just considering either the negatives or the positives, it is considered to be a fair summary of the performance of a classifier when compared to measures such as PPV or NPV. It is also good for assessing the performance of a classifier on an unbalanced dataset, where there are significantly more instances of one class than another, as is the case here. The MCC was calculated using:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \quad (11),$$

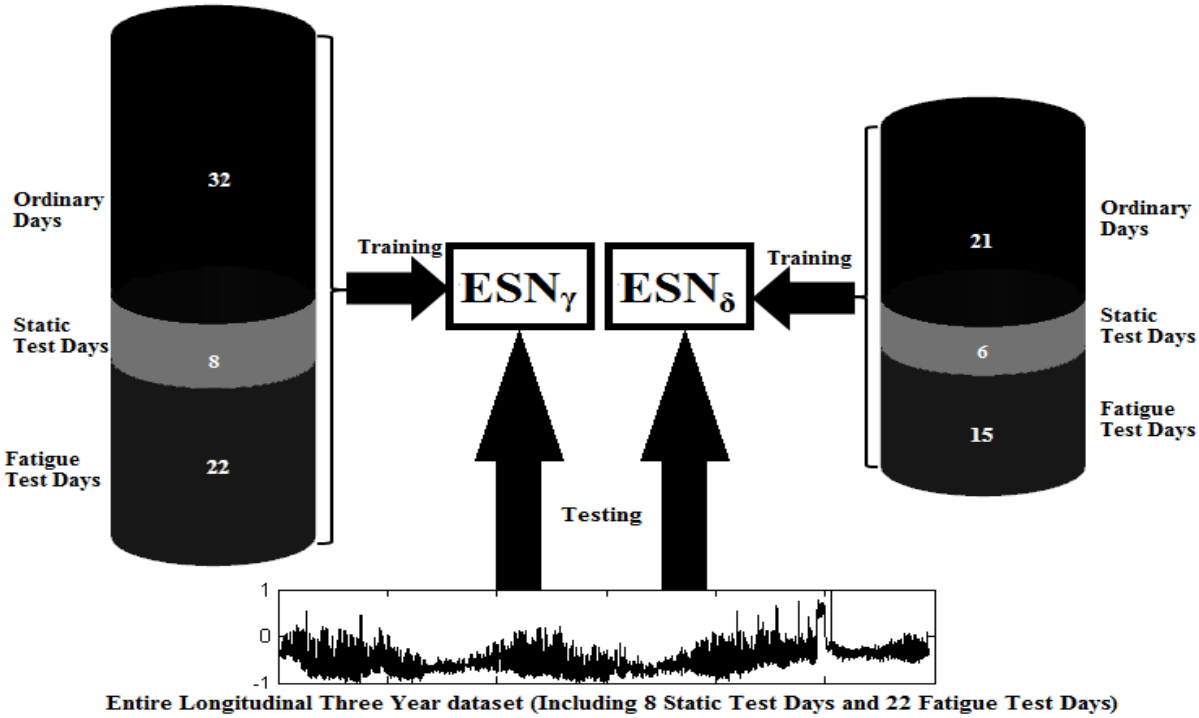

Fig. 6 The different training and test regimes used for ESN_{γ} and ESN_{δ}. ESN_{γ} was trained on the data from 32 ordinary days, 8 static test days and 22 fatigue test days, whereas ESN_{δ} was trained on a reduced training set of 21 ordinary days, 6 static test days and 15 fatigue test days. Each ESN was then presented continuously with the entire longitudinal three-year dataset.

3.3 Summary of Methods

In summary, this section presented four different ESN approaches. The first of these, ESN_{α}, would be directly compared with the NARMAX methodology, and would be trained to replicate the behaviour of the tilt sensors based on the temperature sensors. ESN_{α} and NARMAX used 70% of the dataset for training, sampled at one data point per hour. The second ESN, ESN_{β}, extends this work further, performing the same task using just the first 12.79% of the dataset with no sampling, since this would cover the period prior to the first significant intervention. The final two approaches, ESN_{γ} and ESN_{δ}, aim to classify every single individual day in the dataset as either an ordinary day or featuring either a static test or a fatigue test using the tilt sensor data. The optimal topologies that were found for these four approaches are given in Table 2.

4 Results and Discussion

4.1 Results for detecting manual interventions with ESN_α and NARMAX

The absolute error for both models was plotted against every main test event that the bridge had undergone for each sensor, with a selection of these plots shown in Figure 7. It was found that both approaches captured the relationship between the temperature sensors and the tilt sensors well, with similarly small error rates for the majority of the dataset. For ESN_α and NARMAX, agreement between each model's predicted and actual sensor readings for each time step can be assessed by calculating the Pearson correlation coefficient, which is given in Table 3 along with the averaged values. The closer the Pearson correlation coefficient is to 1, the more accurately the technique modelled the sensor data. These results for the Pearson correlation coefficient were found to be significant ($p < 0.01$). Where larger errors do occur for some of the tilt sensors, such as tilt sensor 2, these correlate well with the majority of events that the bridge experienced over the test period. The comparability of the two techniques is confirmed by inspection of the correlation coefficients for both ESN_α and NARMAX (Table 3), as the two models have similar values.

4.2 Discussion of detection of manual interventions with ESN_α and NARMAX

The tilt sensors 1, 2 and 6, the data from which is plotted in Figure 7, were found to produce the largest point discrepancy between actual and predicted values. An example of this can be seen on the day of the first intervention, where the error for tilt sensor 1 reached 2.348 (NARMAX) and 1.103 (ESN_α). On the same day, the error for tilt sensor 2 reached 0.661 (NARMAX) and 0.307 (ESN_α), while the error for tilt sensor 6 reached 0.738 (NARMAX) and 0.354 (ESN_α). For comparison, no other tilt sensor on that day produced an error value greater than 0.330 using NARMAX, or 0.120 using ESN_α. This could be used to give an indication of the point at which the interventions had the greatest impact. Although the error from the remaining tilt sensors was relatively small, there were increases in the error from particular additional sensors that correspond to some events. For example, the error for tilt sensor 4 was generally less than for tilt sensors 1, 2 and 6, but increased to 0.330 (NARMAX) and 0.161 (ESN_α) on the day of the first intervention. This shows that some sensors reveal information about certain events experienced by the bridge, but not all. Nevertheless, this information is important when assessing the extent and location of changes in the behaviour of the bridge. These results suggest that some areas of the bridge (in particular the areas around tilt sensors 1, 2 and 6) were affected more than others.
The transparency of the NARMAX model, given by the important polynomial terms and their values, helped to provide insight into the characteristics of the bridge studied. The use of a degree of one for the best NARMAX model means that important aspects of the relationship between the temperature and tilt sensor data can be characterised as a linear system. However, the resulting terms and coefficients used by the NARMAX model showed that some inputs from as far back as 12 hours were important. This is demonstrated by the presence of significantly strong terms up to $t - 12$. The suitability of a lag term of 12 for the best performing NARMAX suggests that the bridge is in part a very slow system, with large immediate responses to temperature change (as seen in Figure 2), but the tilt sensor readings being further influenced by changes in temperature up to 12 hours earlier. In contrast, input 4 and input 9 – i.e. temperature sensors 4 and 9 – were eliminated from the final NARMAX

model, indicating that those input units were not important in characterising the relationship between temperature and tilt sensor data.
These insights into the type of system that best captures the relationship between the temperature and tilt data from the bridge is one advantage of NARMAX over RNNs and other types of 'black box' neural networks, which are often complex and difficult to analyse.

4.3 Results for the detection of lasting damage with ESN_{β}

It was expected that the ESNs would be able to accurately estimate the behaviour of the bridge prior to manual interventions, but that the data could potentially deviate thereafter. The Pearson correlation coefficient was again used as a measure of the agreement between the predicted and actual tilt sensor readings, but this time was applied to the data in the training period. The Pearson correlation coefficient for each tilt sensor and ESN_{β} can be seen in Table 4.
It can be seen from Tables 3 and 4 that the change in training regime from ESN_{α} to ESN_{β} led to improved agreement between the two sets of values for the real and predicted tilt sensor data. For six out of the eight tilt sensors, ESN_{β} produced a greater correlation coefficient than ESN_{α} and the overall average increased by 0.18. This is consistent with any discontinuities in the data caused by the interventions not being fed into the ESN training regime, meaning that the ESNs predicted only the normal behaviour of the bridge. An example of the good correlation between real and predicted tilt sensor data in the training portion can be seen in Figure 8, which shows the close correlation between the two sets of data for tilt sensor 4 over the period of a month in the spring of 2009. This is significant, since it shows that ESN_{β} was able to predict accurately how the bridge should normally behave, strongly implying that any difference between the real and predicted values in the testing portion of the dataset would be due to a change in the state of the bridge, rather than a fault in the predictive capability of ESN_{β}.

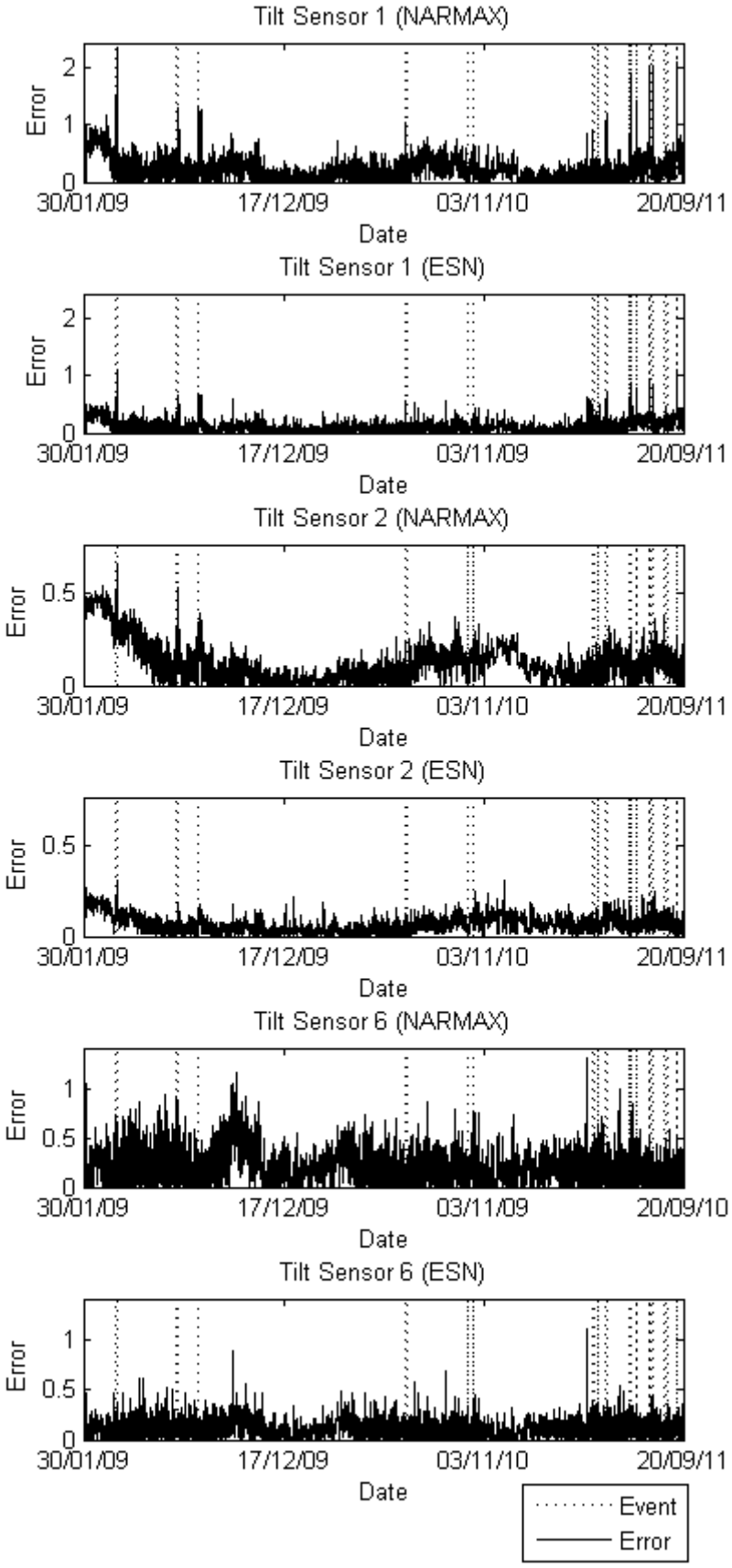

Fig. 7 The absolute error between the output of the NARMAX and ESN models and the target tilt sensor data for three selected tilt sensors. Note the differing Y-axis scales.

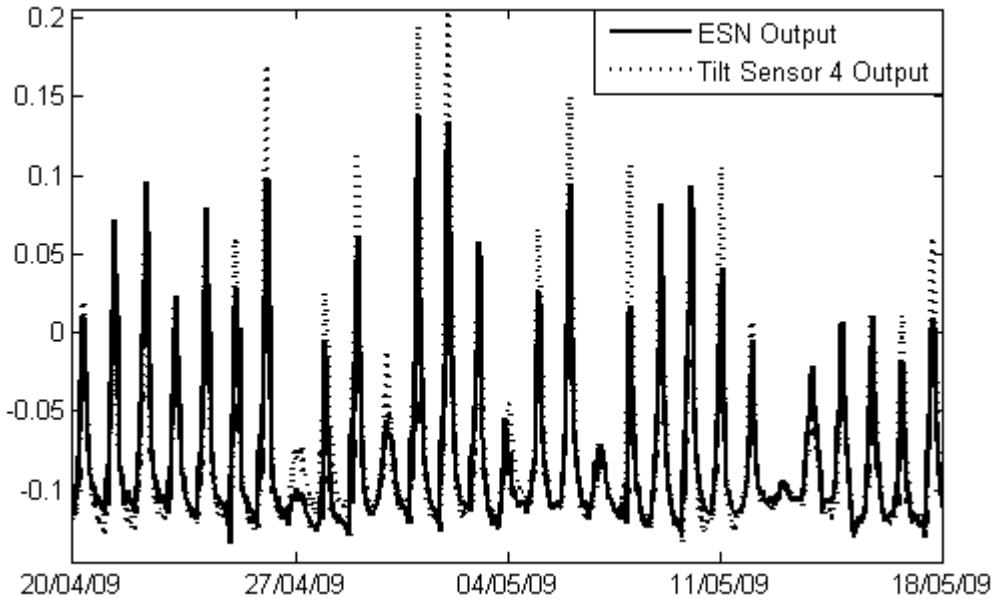

Fig. 8 The output of ESN_{β} and the actual output for tilt sensor 4. The good correlation between the real data and predicted data during this part of the training period can be observed.

While the correlation between tilt sensor 2 and ESN output improved when using ESN_{β}, the value of 0.59 for the Pearson correlation coefficient was still notably lower than that for the other tilt sensors. Further analysis of the output of ESN_{β}, which is detailed in the following paragraphs, showed that tilt sensor 2 was particularly sensitive to the deliberate interventions. It is possible that the region around tilt sensor 2 was prone to damage and that there was an underlying structural condition in the bridge at this point that had developed during the 50 years of ordinary use that the bridge underwent prior to the beginning of the sensor monitoring. This could have caused the bridge to start behaving erratically in this region, making it especially difficult for any regression technique to model its behaviour. Two key damage events were identified by inspection of the long term trends in the data. Figure 9 shows a moving average of the error between the predicted data and the actual data for selected tilt sensors at these two key points in time. Each solid vertical line represents a significant event that occurred at that point. A rise in the error level due to an event that is then maintained over a long time period indicates that the event caused permanent damage or at least medium term modification to the bridge. A rise in the error level following an event but which is followed by the error returning quickly to its prior level is indicative of the event affecting the bridge at that point, but not causing any lasting damage. If there is no change in the error level due to an event, then it probably did not affect the bridge at that tilt sensor position.

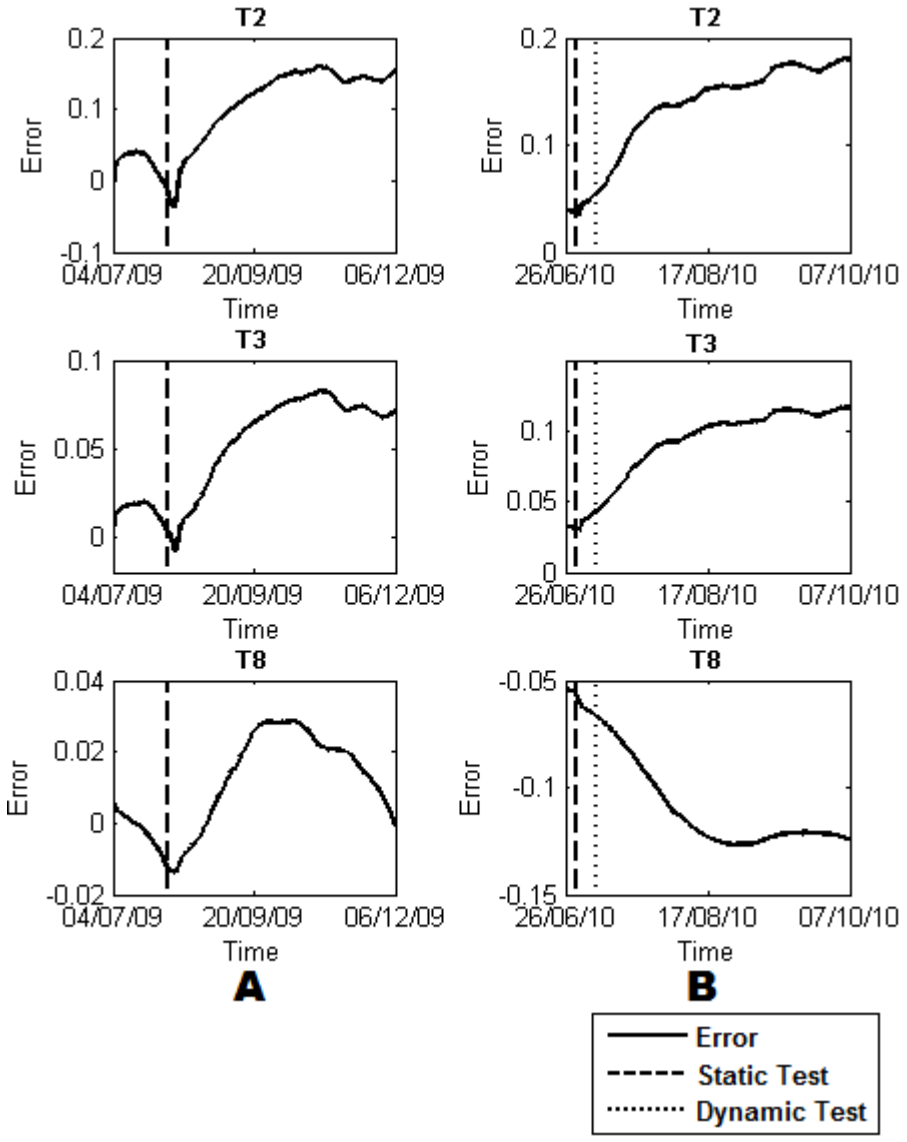

Fig. 9 The modulus of the 10,000 moving average error between ESN_β predicted and actual output for selected sensors and events. Figure 6A depicts the error for tilt sensors 2 (top), 3 (middle) and 8 (bottom) for the first intervention, while Figure 6B depicts the error for tilt sensors 2 (top), 3 (middle) and 8 (bottom) for the second intervention.

4.4 Discussion of the detection of lasting damage with ESN_β

The first of the two key interventions found using ESN_β occurred in August 2009, when two water tanks suspended from one end of the bridge were filled with water and then emptied. The data for the onset of this event is shown in Figure 9a (vertical line) and can be seen to have had a large effect on tilt sensors 2, 3 and 8. The effect on tilt sensors 2 and 3 was maintained not just immediately after the event onset, but over a sustained period. However, the small increase in error for tilt sensor 8 can be seen to return back to its original level; a timescale for reaching equilibrium of several months is noteworthy. This can all be compared to the data in Table 3, where both NARMAX and ESN_α gave low correlation coefficients for tilt sensors 2, 3 and 8.

When comparing Figures 1 and 5, a picture of both the initial effect of the intervention and the slower long term response emerges. The water tanks were loaded onto the cantilever closest to tilt sensor 1. The application of this weight caused the bridge to pivot on the first pier, where tilt sensor 7 is located. The impact of the resultant force was most strongly felt by

tilt sensor 1, as shown in Figure 4, since the cantilever is relatively free to move. However, that freedom of movement allowed it to return to equilibrium almost immediately. Longer term effects were observed most strongly between tilt sensors 2 and 3, as shown by the increase in error, but the pivoting also applied a tensile force to the second pier, where tilt sensor 8 is located. The error continued to increase over a period of months as the structural state of the bridge continued to change, even after the cessation of the intervention. The subsequent levelling out of the error in tilt sensors 2 and 3 suggests that the bridge then reached a state of equilibrium. The fact that this equilibrium was then sustained suggests that the bridge had been permanently changed at this point, with a strained state existing between tilt sensors 2 and 3, although the pier had returned to its initial state. The second of the key events came about due to the loading and unloading of two half-full tanks on the bridge between the 30th of June and the 2nd of July 2010. By this time, the region around tilt sensors 2 and 3 had reached equilibrium, following the event marked in Figure 9a. The effects of this second intervention can be seen in tilt sensors 2, 3, and 8, as shown in Figure 9b, and are very similar to those produced by the first major event. In addition to this, a change in the error signal can be seen in the data from every tilt sensor, which suggests that the tests had an impact that could be felt across the whole bridge, although only errors in the areas around tilt sensors 2, 3 and 8 were suggestive of permanent damage. It is probable that the loading resulted in the bridge again pivoting on the first pier, exacerbating the damage already caused in August 2009. In this case, the effect on the other pier was more long lasting.

It is interesting to note that some of the events that might have been expected to have caused long term problems do not appear to have had any significant effect. For example, in July 2011, static tests involving the water tanks were performed and some of the steel reinforcing bars in the bridge were cut. Although a major immediate response was seen in the tilt sensors using this technique, no further long-term damage was indicated. The same can be said for the other two events that involved the cutting of reinforcing bars, which occurred in October 2010 and April 2011. Similarly, a creep test in late October 2011, wherein a heavy weight was loaded on the bridge over 17 days, was found to have some effect on the data for tilt sensors 1, 2, 4 and 5 while the event was on-going, but once the test had finished the data suggest that the bridge reverted back to its previous state. However, it was also observed that after subsequent interventions, tilt sensor 1 never returned to its initial baseline, suggesting that cutting the reinforcing bars destroyed the cantilever's ability to return to its initial state after further interventions.

4.5 Results for the classification of interventions with ESN_{γ} and ESN_{δ}

A confusion matrix for the performance of ESNγ in the intervention classification task is given in Table 5. Table 6 gives the MCC, sensitivity, specificity, PPV, NPV, FPR and AUC for each output node. The values for the MCC were all found to be significant ($p <= 0.001$). Note that the AUC value given considers the output in terms of the classification of individual points of data, rather than the classification of days. Of the three output nodes, the classification of the fatigue test output node was the most successful, having correctly identified all eight fatigue tests without a single false positive. It therefore achieved the highest score in every single performance measure.

The data in Table 5 show that although they did not perform quite as well as the fatigue test classification node, the ESN_{γ} static test and normal behaviour classification nodes were still successful. The static test classification node managed to detect all but one of the static tests, while the normal behaviour classification node misclassified only one damage event as normal behaviour. The high AUC, sensitivity and specificity values for all three sensors

highlight how successful ESN_γ was. The FPR value was close zero for all three nodes, which also suggests that they performed well. There are relatively modest values for PPV for the static test classification node and NPV for the normal behaviour classification node. However, this can be put down to the fact that there is a far greater number of ordinary days in the full dataset than there are days on which a significant event occurred. Although the number of false positives and true positives are quite close for the static test classification node, it correctly classified 96.82% of the ordinary days and 95.45% of the static tests. Similarly, the normal behaviour classifier successfully classified 96.56% of the significant events and 96.61% of the normal days, despite a low NPV value.

There are further details that come to light when looking at the days when the static test classification node incorrectly gave a positive response, which are shown in Table 5. Of the 30 days seen in Table 7, the one that stands out as being particularly anomalous is the false positive on the 15th March 2010. However, this can be accounted for, as the sensors were switched off between 10th March 2010 and 15th March 2010. It is probably the case that the sudden discontinuity in the data led ESN_γ to believe that a significant event had occurred, resulting in the false positive. Similarly, a creep test was performed on the bridge over the period 11th October 2011 to 28th October 2011. This test significantly altered the normal cycle seen in the tilt sensor data and would explain the nine positives during this period. Although there were no static tests on these occasions, it is safe to say that the bridge was behaving in a particularly abnormal pattern over this period. It is also very interesting to note that 18 of the 30 days came within 31 days of events that were determined by earlier ESN_β experiments to have caused permanent damage to the bridge. It is possible that when the bridge underwent significant structural changes due to permanent damage, a series of 'aftershocks' were recorded in the data that were then automatically detected by ESN_γ. Consequently, it is possible that a single positive reading would indicate an external factor causing the bridge to behave atypically, but a succession of positive ESN_γ readings over the course of a month or so suggests that the initial atypical behaviour has resulted in lasting damage to the bridge.

The performance of ESN_δ is shown in Tables 8 and 9. Neither the static test output node nor the fatigue test output node struggled to detect significant damage events – they both managed similar performance of the equivalent nodes for ESN_γ in this regard – but they were both slightly poorer at discerning between ordinary days and damage events. The most likely cause of this is the fact that ESN_δ used fewer training samples than ESN_γ and thus had less opportunity to learn to distinguish between different significant events and regular days. However, given the small pool of data available to it, ESN_δ performed extremely well and demonstrated the capability of ESNs to learn to detect characteristic signals in large time-series datasets given a limited amount of training data. The data in Table 9 reflects this, as all three nodes for ESN_δ have high values of sensitivity, specificity and AUC, while also having a very low FPR.

Table 10 gives the days when the static test classification node for ESN_δ incorrectly gave a positive response. It can be seen that, like the false positives produced by ESN_γ, the majority of these days fall within 31 days of a static test occurring, which is suggestive of the presence of 'aftershocks' in the data. There are only 11 false positives that do not fall within 31 days of a static test, with one of these again being due to the switching off of the bridge sensors on the 15th March 2010. Ten of these 11 false positives all fell within 60 days of a static test. There are twelve days on which ESN_δ produced a false positive that fell in the period 11th October – 28th October 2011. As with ESN_γ, the response of the static test output node at this point is probably due the creep test that took place during this period, causing the bridge to behave abnormally for the duration of the test.

4.6 Discussion of the intervention classification

The results for ESN_γ and ESN_δ demonstrate the suitability of ESNs for attempting to classify real-world time-series data. By using ESN_γ it was shown that it is possible to train an ESN on certain patterns and then to have it recognise them in a larger test dataset with very few false positives. ESN_δ showed that an ESN is capable of learning to recognise a particular type of damage event and then detecting previously unseen damage events in a large test dataset. The work done here suggests that to obtain a fuller picture of the state of a structure, the optimal approach would be to use a classifying ESN to detect and characterise significant deviations from normal behaviour, but to also task another ESN with predicting the typical behaviour of the tilt sensors so that the real and predicted values could be compared for the observation of long term trends and changes, and the assessment of damage. By doing this, both abnormal events and long term permanent damage could be detected.

5 Conclusion

It has been shown that ESNs could have multiple applications on a structural health monitoring dataset, as they are versatile enough to be used for both classification and regression. This meant that they were able to provide comparable performance of a NARMAX model when trained using the same data. Additionally, ESNs were also shown to be capable not only of detecting specific types of intervention, but of allowing a user to assess and locate damage to the bridge over a long term period, a key requirement in SHM. This offers a clear improvement on the alternative techniques that have been applied to the same data, which were only able to model the tilt sensor data and detect that anomalous events had occurred, without considering the specific type of event or potential long term effects. Significantly, this was done using only the first six months of data, without introducing lengthy pre- or post-processing procedures, offering greater 'real-world' applicability.

Four separate ESN approaches were successfully used to analyse the data analysing from the NPL footbridge. The first approach compared favourably to the NARMAX model used, as both were given the task of using the temperature sensor data to predict the tilt sensor data as their output and achieved average Pearson correlation coefficients of 0.63 (NARMAX) and 0.64 (ESN_α). Adjusting the ESN training regime to create ESN_β made it possible to detect both the initial and long term responses of the bridge to the manual interventions, with two events in particular being found to affect the bridge permanently. The final two ESN approaches (ESN_γ and ESN_δ) were used to characterise the behaviour of the bridge and any interventions. The fatigue test classification output was found to accurately detect the occasions on which fatigue tests occurred, while the static test and normal behaviour classification outputs still performed very well, detecting 21 out of 22 static tests with very few false positives. When a portion of the training data was removed so that there were some unseen examples of damage events in the testing dataset, ESN_δ was able to offer similar performance to the event detection capability of ESN_γ with only a small loss of ability to discern between ordinary days and significant events. Like ESN_γ, ESN_δ managed to detect 21 out of 22 static tests and all eight fatigue tests, while producing false positives on only eight more days than ESN_γ in the testing dataset.

Whilst the NARMAX model is not capable of classifying events, it can be seen to be useful for more in depth investigations of the behaviour of structures, determining the magnitude and timescale of the response of one part of the bridge to stimuli across the whole of the structure.

The successes reported in this work shows that ESNs are a very useful technique for the online SHM of real world structures such as bridge. This work has shown that ESNs could plausibly be incorporated into a centralised data processing scheme for online SHM in sensor networks. In a real-world scenario, one could envisage using the ESN_{γ}/ESN_{δ} approach to detect when a potentially damaging event had occurred, with ESN_{β} allowing any effects from this event to be localised and monitored over the long term. The damage events in this study occurred over a relatively short period of time. It would therefore be of future interest to apply the approaches detailed above to structures whose gradual degradation occurs over a much longer time-scale. The testing and use of ESNs for this purpose, along with the incorporation of other sensor modalities, is the focus of future work.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr Elena Barton and the UK National Physical Laboratory for conducting the experiments on the footbridge and making the data available for analysis.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of Interest: The authors declare that they have no conflict of interest.

REFERENCES

American Society of Civil Engineers (ASCE), 2013. Report Card for America's Infrastructure. Available from: http://www.infrastructurereportcard.org/a/#p/bridges/investment-and-funding. Accessed 2/1/16.

Bacic, B., 2016. Echo State Network for 3D Motion Pattern Indexing: A Case Study on Tennis Forehands, in: Braunl, T., McCane, B., Rivera, M., Yu, X. (Eds.), Image Video Technology: 7th Pacific-Rim Symposium, PSIVT 2015, Auckland, New Zealand, November 25-27, 2015, Revised Selected Papers. Springer International Publishing, Cham, pp. 295–306.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., Nielsen, H., 2000. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinform 16, 412–424.

Barton, E., 2011. SHM Demonstrator at NPL: Two Years of Monitoring Experience and Future Challenges.

Barton, E., Esward, T., 2012. The Origins of Measurement Uncertainty in SHM - NPL Footbridge Case Study, in: 6th European Workshop on Structural Health Monitoring.

Barton, E., Middleton, C., Koo, K., Crocker, L., Brownjohn, J., 2011. Structural Finite Element Model Updating Using Vibration Tests and Modal Analysis for NPL footbridge - SHM demonstrator. Journal of Physics: Conference Series 305, 012105.

Barton, E.N., Zhang, B., 2010. Details of temperature compensation for strain measurements on NPL bridge - demonstrator for SHM. Appl Mech Mater 24 - 25, 173–178.

Billings, S.A., Chen, S., 1998. The determination of multivarmulti nonlinear models for dynamical systems. Neural Network Systems, Techniques and Applications 82, 231–278.

Butcher, J.B., Day, C.R., Haycock, P.W., Verstraeten, D., Schrauwen, B., 2014. Defect Detection in Reinforced Concrete using Random Neural Architectures. Comput-aided Civ Inf 29, 191–207.

Butcher, J.B., Verstraeten, D., Schrauwen, B., Day, C.R., Haycock, P.W., 2013. Reservoir computing and extreme learning machines for non-linear time-series data analysis. Neural Networks 38, 76–89.

Dervilis, N., Shi, H., Worden, K., Cross, E.J., 2016. Exploring Environmental and Operational Variations in SHM Data Using Heteroscedastic Gaussian Processes, in: Pakzad, S., Juan, C. (Eds.), Dynamics Civil Structures, Volume 2: Proceedings 34th IMAC,A Conference Exposition Structural Dynamics 2016. Springer International Publishing, Cham, pp. 145–153.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn Lett 27, 861–874.

Footbridge Monitoring Project (SHM) - Background [Internet]. National Physical Laboratory; Available from: http://www.npl.co.uk/content/ConWebDoc/2879

Iglesias, R., Kyriacou, T., Nehmzow, U., Billings, S.A., 2007. Task identification and characterisation in mobile robotics through non-linear modelling. Robot Auton Syst 55, 267–275.
ITMSOIL, 2016. Available from: http://www.itmsoil.com/.
Jaeger, H., 2010. The "echo state" approach to analysing and training recurrent neural networks - with an Erratum note. Fraunhofer Institute for Autonomous Intelligent Systems.
Jaeger, H., 2013. A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach (No. 4). Fraunhofer Institute for Autonomous Intelligent Systems (AIS).
Koo, K.Y., Brownjohn, J.M.W., List, D.I., Cole, R., 2013. Structural health monitoring of the Tamar suspension bridge. Struct Control Hlth 20, 609–625.
Korenberg, M., Billings, S.A., Liu, Y.P., McIlroy, P.J., 1988. Orthogonal parameter estimation algorithm for non-linear stochastic systems. Int J Control 48, 193–210.
Kromanis, R., 2015. Structural Performance Evaluation of Bridges: Characterizing and Integrating Thermal Response.
Kromanis, R., Kripakaran, P., 2014. Predicting thermal response of bridges using regression models derived from measurement histories. Comput Struct 136, 64–77.
Kyriacou, T., Nehmzow, U., Iglesias, R., Billings, S.A., 2008. Accurate robot simulation through system identification. Robot Auton Syst 56, 1082–1093.
Laory, I., Trinh, T.N., Smith, I.F.C., Brownjohn, J.M.W., 2014. Methodologies for predicting natural frequency variation of a suspension bridge. Eng Struct 80, 211–221.
Livina, V., Barton, E., Forbes, A., 2013. Tipping point analysis of the NPL footbridge. Journal of Civil Structural Health Monitoring 3, 1–8.
Lukoševičius, M., 2012. A practical guide to applying echo state networks. In *Neural networks: Tricks of the trade* (pp. 659-686). Springer Berlin Heidelberg.
Lukoševičius, M., Jaeger, H., 2009. Reservoir computing approach to recurrent neural network training. Computer Science Review 3, 127–149.
Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. BBA-Protein Struct M 405, 442–451.
Montgomery, D.C., Peck, E.A., 1982. Introduction to Linear Regression Analysis. Wiley.
NakedScientists. Structural Health Monitoring [Internet]. Available from:
https://www.youtube.com/watch?v=oO7E2G2WfL4
Palumbo, F., Gallicchio, C., Pucci, R., Micheli, A., 2016. Human activity recognition using multisensor data fusion based on Reservoir Computing. J Ambient Intell Smart Environ 8, 87–107.
Quevedo, J., Chen, H., Cuguero, M.A., Tino, P., Puig, V., Garcia, D., Sarrate, R., Yao, X., 2014. Combining learning in model space fault diagnosis with data validation/reconstruction: Application to the Barcelona water network. Eng Appl Artif Intell 30, 18–29.
Sun, J., Li, H., Xu, B., 2016. Prognostic for hydraulic pump based upon DCT-composite spectrum and the modified echo state network. SpringerPlus 5, 1–17.
Tong, M.H., Bickett, A.D., Christiansen, E.M., Cottrell, G.W., 2007. Learning grammatical structure with Echo State Networks. Neural Networks 20, 424–432.
Verstraeten, D., 2009. Reservoir Computing: computation with dynamical systems. Ghent University.
Verstraeten, D., Schrauwen, B., D'Haene, M., Stroobandt, D., 2007. An experimental unification of reservoir computing methods. Neural Networks 20, 391–403.
Wootton, A.J., Day, C.R., Haycock, P.W., 2014. Echo State Network Applications in Structrual Health Monitoring, in: Proceedings of the 53rd Annual Conference of The British Institute Non-Destructive Testing (NDT 2014). pp. 289–300.
Wootton, A.J., Day, C.R., Haycock, P.W., 2015. An Echo State Network Approach to Structural Health Monitoring, IEEE IJCNN. pp. 1–7.
Worden, K., Cross, E., Barton, E., 2012. Damage Detection on the NPL Footbridge Under Changing Environmental Conditions, in: 6th European Workshop on Structural Health Monitoring.
Worden, K., Cross, E.J. and Brownjohn, J.M., 2013. Switching response surface models for structural health monitoring of bridges. In Surrogate-Based Modeling and Optimization (pp. 337-358). Springer New York.

Table 1: Significant interventions performed as part of the NPL footbridge project, along with the date when these were performed.

Date	Intervention Type
24th March 2009	Static test
29-30th June 2009	Static test
3rd August 2009	Static test
30th June – 2nd July 2010	Static test
8th October 2010	Static test
18th October 2010	Static test
26th April 2011	Static test
18 – 19th May 2011	Static test
24th June 2011	Static test
27th June 2011	Static test
6 – 7th July 2011	Static test
28th July 2011	Static test
21st August 2011	Static test
24th August 2011	Static test
9th September 2011	Static test
28th September 2011	Fatigue test
29th September 2011	Fatigue test
3rd October 2011	Fatigue test
4th October 2011	Fatigue test
6th October 2011	Fatigue test
7th October 2011	Fatigue test
10th October 2011	Fatigue test
11th October 2011	Fatigue test
11th October 2011	Static test
11th – 28th October 2011	Creep test
28th October 2011	Static test
8th November 2011	Static test

Table 2: Optimal Parameters of the four ESN architectures used here.

Parameter	ESN_{α}	ESN_{β}	ESN_{γ}	ESN_{δ}
Input units	10	10	8	8
Output units	8	8	3	3
Reservoir Units	200	250	500	500
Spectral Radius	0.4	0.9	0.9	0.9
Input Scaling	0.25	0.25	1	1
Leak Rate	1	1	1	1

Table 3 The Pearson correlation coefficients between the output of both models and the target tilt sensor data for all eight tilt sensors with the average also shown, together with the standard deviation in parentheses.

Model	Tilt 1	Tilt 2	Tilt 3	Tilt 4	Tilt 5	Tilt 6	Tilt 7	Tilt 8	Ave
NARMAX	0.73	0.33	0.48	0.75	0.83	0.92	0.74	0.23	0.63 (0.25)
ESN_{α}	0.72	0.39	0.52	0.73	0.84	0.91	0.75	0.22	0.64 (0.24)

Table 4 The Pearson correlation coefficients between the output of ESN_{β} and the target tilt sensor data for all eight tilt sensors during the training period, with the average and standard deviation also shown.

Model	T1	T2	T3	T4	T5	T6	T7	T8	Ave
ESN_{β}	0.72	0.59	0.75	0.92	0.91	0.88	0.92	0.86	0.82 (0.12)

Table 5 A confusion matrix for ESN_{γ}.

		ESN_{γ} Prediction		
		Static Test	Fatigue Test	Normal Day
Ground Truth	Static Test	21	0	1
	Fatigue Test	0	8	0
	Normal Day	30	0	914

Table 6 Measures of the performance of each of the three classification nodes for ESN_{γ}. Note that the AUC value was calculated for the correct classification of individual data points, rather than windows of days.

Classification node	MCC	Sens.	Spec.	PPV	NPV	FPR	AUC
Static Test	0.616	0.955	0.968	0.412	0.999	0.032	0.977
Fatigue Test	1.000	1.000	1.000	1.000	1.000	0.000	1.000
Normal Behaviour	0.659	0.999	0.966	0.999	0.467	0.034	0.998

Table 7 Days when the static test classification node for ESN_{γ} produced a false positive, along with the closest static test prior to the date and the difference between the two in days.

False Positive Date	Nearest Event Date	Difference
22-Jul-09	29-Jun-09	23
27-Jul-09	29-Jun-09	28
09-Aug-09	03-Aug-09	6
11-Aug-09	03-Aug-09	8
13-Aug-09	03-Aug-09	10
16-Aug-09	03-Aug-09	13
19-Aug-09	03-Aug-09	16
23-Aug-09	03-Aug-09	20
27-Aug-09	03-Aug-09	24
12-Sep-09	03-Aug-09	40
21-Sep-09	03-Aug-09	49
25-Sep-09	03-Aug-09	53
15-Mar-10	03-Aug-09	224
10-Jul-10	02-Jul-10	8
11-Jul-10	02-Jul-10	9
12-Jul-10	02-Jul-10	10
19-Jul-10	02-Jul-10	17
25-Jul-10	02-Jul-10	23
08-Aug-10	02-Jul-10	37
09-Aug-10	02-Jul-10	38
16-Aug-10	02-Jul-10	45
14-Oct-11	11-Oct-11	3
15-Oct-11	11-Oct-11	4
18-Oct-11	11-Oct-11	7
19-Oct-11	11-Oct-11	8
20-Oct-11	11-Oct-11	9
22-Oct-11	11-Oct-11	11
23-Oct-11	11-Oct-11	12
24-Oct-11	11-Oct-11	13
25-Oct-11	11-Oct-11	14

Table 8 A confusion matrix for ESN_{δ}.

		ESN_{δ} Prediction		
		Static Test	Fatigue Test	Normal Day
Ground Truth	Static Test	21	0	1
	Fatigue Test	0	8	0
	Normal Day	38	1	905

Table 9 Measures of the performance of each of the three classification nodes for ESN_{δ}. Note that the AUC value was calculated for the correct classification of individual data points, rather than windows of days.

Classification node	MCC	Sens.	Spec.	PPV	NPV	FPR	AUC
Static Test	0.570	0.955	0.960	0.356	0.999	0.040	0.967
Fatigue Test	0.942	1.000	0.999	0.889	1.000	0.001	1.000
Normal Behaviour	0.611	0.957	0.966	0.999	0.406	0.034	0.998

Table 10 Days when the static test classification node for ESN_{δ} produced a false positive, along with the closest static test prior to the date and the difference between the two in days.

False Positive Date	Nearest Event Date	Difference
22-Jul-09	29-Jun-09	23
23-Jul-09	29-Jun-09	24
27-Jul-09	29-Jun-09	28
09-Aug-09	03-Aug-09	6
11-Aug-09	03-Aug-09	8
13-Aug-09	03-Aug-09	10
16-Aug-09	03-Aug-09	13
19-Aug-09	03-Aug-09	16
23-Aug-09	03-Aug-09	20
26-Aug-09	03-Aug-09	23
27-Aug-09	03-Aug-09	24
06-Sep-09	03-Aug-09	34
12-Sep-09	03-Aug-09	40
18-Sep-09	03-Aug-09	46
21-Sep-09	03-Aug-09	49
24-Sep-09	03-Aug-09	52
25-Sep-09	03-Aug-09	53
26-Sep-09	03-Aug-09	54
27-Sep-09	03-Aug-09	55
15-Mar-10	03-Aug-09	224
10-Jul-10	02-Jul-10	8
11-Jul-10	02-Jul-10	9
19-Jul-10	02-Jul-10	17
25-Jul-10	02-Jul-10	23
09-Aug-10	02-Jul-10	38
16-Aug-10	02-Jul-10	45
13-Oct-11	11-Oct-11	2
14-Oct-11	11-Oct-11	3
15-Oct-11	11-Oct-11	4
17-Oct-11	11-Oct-11	6
18-Oct-11	11-Oct-11	7
19-Oct-11	11-Oct-11	8
20-Oct-11	11-Oct-11	9
21-Oct-11	11-Oct-11	10
22-Oct-11	11-Oct-11	11
23-Oct-11	11-Oct-11	12
24-Oct-11	11-Oct-11	13
25-Oct-11	11-Oct-11	14