Mechanical Systems and Signal Processing 206 (2024) 110934

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Data-driven modeling of long temperature time-series to capture the thermal behavior of bridges for SHM purposes

S. Mariani [a,*], A. Kalantari [a], R. Kromanis [b], A. Marzani [a]

[a] Department of Civil, Chemical, Environmental and Materials Engineering — DICAM, University of Bologna, Viale del Risorgimento 2, Bologna, 40136, Italy

[b] Faculty of Engineering Technology (ET), University of Twente, Drienerlolaan 5, Enschede, 7522, NB, Netherlands

ARTICLE INFO

Communicated by W.-X. Ren

Keywords:
SHM
Thermal inertia
Time-lag
Regression
WaveNet
Bridges

ABSTRACT

Bridges experience complex heat propagation phenomena that are governed by external thermal loads, such as solar radiation and air convection, as well as internal factors, such as thermal inertia and geometrical properties of the various components. This dynamics produces internal temperature distributions which cause changes in some measurable structural responses that often surpass those produced by any other load acting on the structure or by the insurgence or growth of damage. This article advocates the use of regression models that are capable of capturing the dynamics buried within long sequences of temperature measurements and of relating that to some measured structural response, such as strain as in the test structure used in this study. Two such models are proposed, namely the multiple linear regression (MLR) and a deep learning (DL) method based on one-dimensional causal dilated convolutional neural networks, and their ability to predict strain is evaluated in terms of the coefficient of determination R^2. Simple linear regression (LR), which only uses a single temperature reading to predict the structural response, is also tested and used as a benchmark. It is shown that both MLR and the DL method largely outperform LR, with the DL method providing the best results overall, though at a higher computational cost. These findings confirm the need to consider the evolution of temperature if one wishes to setup a temperature-based data-driven strategy for the SHM of large structures such as bridges, an example of which is given and discussed towards the end of the article.

1. Introduction

Bridges are critical components of transportation networks, therefore ensuring their structural integrity is paramount for public safety and economic productivity. Since the 1990s, two complementary approaches have been developed for bridge condition assessment, one based on Bridge Management Systems (BMSs) and the other on Structural Health Monitoring (SHM) [1]. While BMSs are mostly based on periodic visual inspection [2], SHM involves the use of sensors for the collection of structural response and environmental data, which are then analyzed to predict the structural performance under various conditions and to detect the onset of damage [3,4]. Notably, this type of data can also be used to gain a better understanding on the definition of thermal, wind and traffic loads, on how do they mutually interact and on how to separate their effects. Some of the sensors which are typically employed for these tasks are accelerometers, strain gauges, fiber optic sensors, displacement gauges, thermistors and anemometers [5,6].

The SHM techniques used for damage detection in bridges are typically based either on numerical-model-updating or on data-driven algorithms. Model-updating consists in setting a numerical model whose parameters are iteratively fit to experimental

* Corresponding author.
E-mail address: stefano.mariani9@unibo.it (S. Mariani).

https://doi.org/10.1016/j.ymssp.2023.110934
Received 22 August 2023; Received in revised form 21 October 2023; Accepted 6 November 2023
Available online 9 November 2023

data [7]. This is computationally intensive, and there are difficulties in capturing all intricacies of real-life systems such as varying material properties, boundary conditions, and environmental conditions such as temperature, wind and humidity [8,9]. On the contrary, data-driven approaches do not require an exact knowledge of the test structure, and they aim to infer some damage-sensitive information from measured time-series related to structural responses and/or environmental factors [10,11].

Among the various environmental conditions that can affect a structure, it is well recognized that temperature typically has the largest effect on the structural response [12,13]. In fact, temperature effects can also exceed those due to vehicular traffic [14]. For example, Catbas et al. [8] observed that the annual peak-to-peak strain differentials due to seasonal temperature variations on a long-span truss bridge in USA were an order of magnitude higher than those due to traffic. Similarly, Koo et al. [15] noticed that the majority of the observed deformations in the Tamar Bridge in UK was due to thermal variations. Other authors have shown that the modal properties of bridges are highly correlated to ambient temperature [12,16–18].

Large structures such as bridges are exposed to a wide range of thermal loads resulting from the diurnal and seasonal temperature variations experienced in their surroundings. Arguably, the most relevant thermal loads for bridges are direct solar radiation experienced by their sun-facing external surfaces and convective phenomena occurring near the heated surfaces. Within the structure, the rate of temperature changes and the spatial distribution of temperature are governed by thermal inertia, which refers to the property of a material to resist changes in its temperature due to its heat capacity and thermal conductivity, and the geometrical properties of the components. Due to this complex thermal behavior, non-uniform temperature distributions within the structure exist at any instant of time [19–24]. Yet, temperature measurements are only available from a few discrete locations in the external surface of the structure where thermal sensors are installed. This can be improved by using thermal cameras, although they are only able to provide the distribution of temperature across some visible external surfaces. As a result of these thermal phenomena, i.e. of the non-uniform structural temperature distribution that lags behind the environmental temperature variations, time-lags are often observed between the histories of structural responses and temperatures measured at various structural locations [6,25–28]. As a matter of example, Guo et al. [6] estimated an average time-lag of approximately 45 min between displacement and temperature histories measured on a steel box-girder cable-stayed bridge.

Kromanis et al. [29,30] have recognized the importance of measuring temperature in a distributed manner in order to characterize its non-uniform distribution, and they proposed regression-based models that make use of all available thermal sensors to predict structural responses as measured by strain gauges. In order to take into account the time-lags between temperatures and strains measured at different structural locations, they considered the inclusion of past temperature readings in their regression models [30]. However, their regression algorithm, which was based on principal component analysis (PCA) [31] and support vector regression [32], only included a single temperature reading in the past together with temperature and strain measured at the current time [30]. Similarly, Sohn et al. [33] used temperature measured from two sensors at two different temporal instants to model the eigenfrequencies of the Alamosa Canyon Bridge, USA, via multiple linear regression (MLR) [34], while Peeters and De Roeck [12] setup an ARX model [35] based on temperature measured from a single sensor at two different temporal instants to predict the eigenfrequencies of the Z24 bridge, Switzerland. Recently, Wang et al. [36] proposed an approach based on a long short-term memory (LSTM) network [37] fed with six past deflection and temperature measurements to predict the expected deflection of a suspension bridge in Chongqing, China, while Xu et al. [38] suggested a two-step method based on a robust PCA routine followed by classical PCA in order to remove the effects of temperature and other environmental conditions from the eigenfrequencies measured on large structures.

This work proposes the use of regression models that can be fed with a large number of past temperature measurements in order to fully capture the dynamic thermal behavior of the monitored structure and hence to better predict the expected structural response. To this scope, the performance offered by a deep learning method based on WaveNet [39], which has been specifically adapted to the task, is compared to that of MLR. Simple linear regression (LR) [34] is also considered as a benchmark scenario. The algorithms are tested on a subset of the experimental data collected and described in [30] that only includes a pair of strain and temperature sensors exhibiting significant time-lag effects. A scenario where the temperature readings are time-shifted in order to better align strain and temperature variations prior to the regression operation is also considered.

The remainder of the article is structured as follows. Section 2 presents the regression models used in this study to capture the thermal behavior of a given structure and the methodology devised to produce and compare the various results. Section 3 describes the experimental dataset that was used as a test-bed. Section 4 discusses the results derived from the analysis and provides a possible implementation of those into a strategy for SHM purposes. Finally, Section 5 concludes the article by summarizing the key findings.

2. Background and methodology

This work builds on the assumption that temperature variations cause changes in some measurable structural responses (e.g. in dynamic properties such as frequencies of vibrational modes or in quasi-static responses such as strains) which, in turns, may or may not also be sensitive to some types of damage. The goal is to define a relation (i.e. a regression model) between temperature measured at some structural location and the structural response measured at some other location. As the distance between these two locations increases, this is complicated by the effects of thermal inertia, which introduce temporal lags between the two readings. It will be shown that in order to take into account the time required for heat propagation phenomena to fully unravel, it is necessary to consider a certain length of history of temperature readings acquired prior to the temporal instant when the structural response is measured. To this scope, regression models will be trained using data taken in a “baseline” period when the structure is at some initial, potentially undamaged state, and will then be used in “monitoring” manner to detect and track deviations from the initial conditions, which can be linked to the insurgence or growth of damage. Three different regression models are considered for the task, namely LR, MLR and a deep-learning architecture based on WaveNet, i.e. on one-dimensional (1D) causal dilated convolutional neural networks (CNNs) [39].

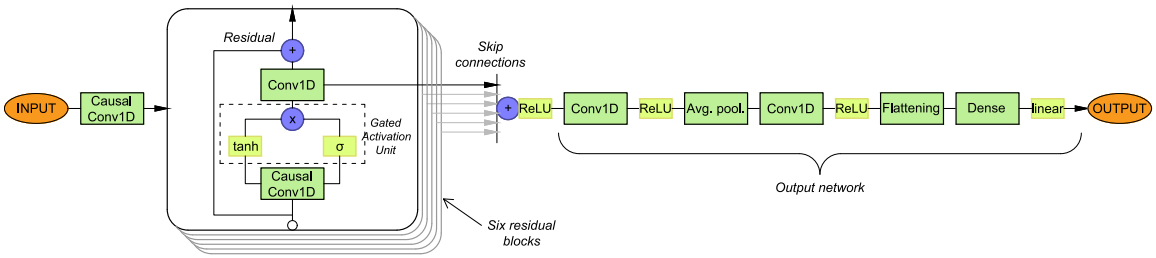

Fig. 1. Implementation of a WaveNet-based architecture for the regression task. The operations within the "output network" produce an unbounded scalar output. Refer to [39,43] for an explanation of symbols and terms.

2.1. Linear regression and multiple linear regression

LR is a statistical technique used to model the relation between a dependent variable and an independent variable, by assuming a linear relationship between the two. The model can be expressed mathematically as:

$$y = \beta_0 + \beta_1 x + \epsilon \tag{1}$$

where y is the dependent variable, x is the independent variable, β_0 is the y-intercept, β_1 is the slope of the line and ϵ is the error term that captures the random variation in the data. In this study, y is a structural response reading, namely strain, and x is a temperature measurement. Note that these two readings are not necessarily acquired at the same temporal instant, as discussed in Section 2.3. Since LR is often used to model the relation between temperature and structural response (see e.g. [20,27]), it is here used as a benchmark to assess the performance of the other methods.

Multiple linear regression (MLR) extends the concept of simple linear regression to include more than one independent variable [40]. In this case, the model can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \tag{2}$$

where y is still the dependent variable, $x_1, x_2, \ldots, x_n$ are n multiple independent variables, $\beta_0, \beta_1, \ldots, \beta_n$ are the corresponding coefficients that measure the impact of each independent variable on the dependent variable, and again the error term ϵ captures the unexplained variation in the model. In this work, $x_1, x_2, \ldots, x_n$ represent a temporal sequence of temperature measurements recorded by an individual sensor, rather than an array of environmental and/or structural measurements acquired by different sensors at the same temporal instant, as it is typically done in literature (see e.g. [24]). For example, assuming that x_1 is the temperature measured at the same temporal instant as the dependent variable y (i.e. strain), x_2 would be the temperature measured at the previous sample in time, and so for up to x_n being the temperature measured $n-1$ samples in the past. Note that considerations on how to choose the value of n are provided in Section 4.2. For both LR and MLR models, the method used to estimate the coefficients is the least squares fit, which minimizes the sum of squared differences between the observed and predicted values of a vector of dependent variables.

2.2. Deep learning method based on 1D causal dilated CNNs (WaveNet)

WaveNet was presented in 2016 as a deep generative model capable of operating directly on raw audio waveforms of thousands data samples [39]. The core of WaveNet's architecture is a stack of 1D causal dilated CNNs. By using dilated convolutions, WaveNet is able to expand the receptive field exponentially while keeping the computational cost manageable, and this allows the model to capture long-range dependencies in the input time-series. Since its first release, the core scheme of WaveNet has been adapted and applied to solve a variety of different tasks, including classification and regression, in numerous fields [41–44].

Fig. 1 depicts the WaveNet-based deep learning architecture that was developed to perform the regression task required in this work. Analogously to the MLR model, the input time-series fed to the algorithm is a temporal sequence of temperature measurements recorded by an individual sensor. This is passed to a stack of 1D causal dilated CNNs that is followed by the "output network" shown in Fig. 1, whose final output is an unbounded scalar term being the predicted structural response (i.e. strain). The hyperparameters of the model were determined by gradually increasing its complexity, and the chosen final architecture of Fig. 1 employs six residual blocks, each with kernel size of 3 and with increasing dilation rates of 3, 9, 27, 81, 243 and 729 (see [39,43] for an explanation of the terms), giving a total of 8225 parameters to learn. In the remainder of the article, this algorithm will be simply called WaveNet.

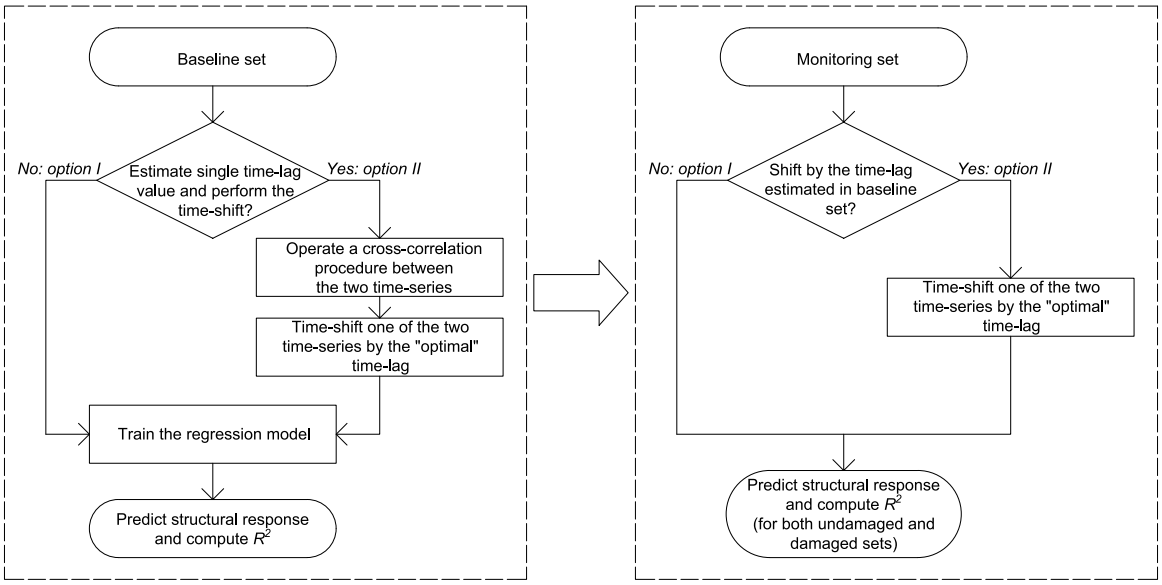

Fig. 2. Workflow used to train each regression model on the baseline set (left box), and to then apply the trained model on the monitoring set of data (right box). For each regression model the whole procedure is repeated twice, i.e. first under "Option I" then under "Option II".

2.3. Methodology

The three regression models are tested on the experimental dataset described in the next section, where the measured structural response y is strain. Fig. 2 shows a flowchart that summarizes the workflow used to train each regression model on the baseline set of data. The trained model is then used in the monitoring period to predict strain readings based on the measured temperature histories. The flowchart also indicates that for each of the three models the whole procedure is repeated twice, i.e. either along the branch termed "Option I" or that called "Option II". In the former scenario, the temporal sequences of temperature readings fed to MLR or WaveNet models include temperature measured at the same instant as strain together with a chosen value $n-1$ of past temperature readings, as described in Section 2.1 (note that when a value $n=1$ is chosen, only a single temperature reading is used as the independent variable and this corresponds to the LR model). "Option II", instead, is designed to test whether a time-shifting of one of the two time-series (temperature or strain) in order to better align their temporal variations would enhance the performance of the regression models. In this scenario, the temporal sequences of temperature readings fed to the models include temperature measured at an "optimal" average time-lag from the time when strain is measured together with $n-1$ past temperature readings from that one. Such "optimal" time-lag is assumed as that giving the highest lagged cross-correlation value computed on the baseline set of data.

The primary metrics that will be used to characterize the performance of the various models is the coefficient of determination R^2 [45], which is defined as:

$$R^2 = \left(1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}\right) \cdot 100 \tag{3}$$

where y_i is the ith value of measured strain, $\hat{y}_i$ is its prediction by the regression model, and $\bar{y}_i$ is the mean of the measured strains, i.e. $\bar{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i$. Furthermore, moving averages of the squared differences computed between predicted and measured strains will be plotted as a function of time and discussed.

3. Experimental case study

Experimental data was collected in 2013 on an aluminum truss setup in a laboratory environment at the University of Exeter, UK. The experimental setup is already described in [30], thus here is only briefly outlined. Fig. 3 shows the truss dimensions and the location of sensors, namely 10 strain gauges and 31 thermocouples. Three infrared heating lamps were placed at vertical and horizontal distances of 0.5 and 0.2 m from the truss, respectively. The lamps were automatically turned on for 45 min every 90 min, therefore simulating 16 diurnal cycles every day. All strain gauges and thermocouples were set to sample data at a 10-s interval. Additionally, a thermal imaging camera was used to capture the external temperature of the truss, though the camera's data was not used in this article. However, a close inspection of the thermal image shown in Fig. 4, which was published in [30], shows an important aspect of the test-setup, namely that the lateral corners of the truss are essentially the only portions of the structure that are not directly illuminated by the lamps. In particular, the portion at the right corner lies at the furthest distance from direct

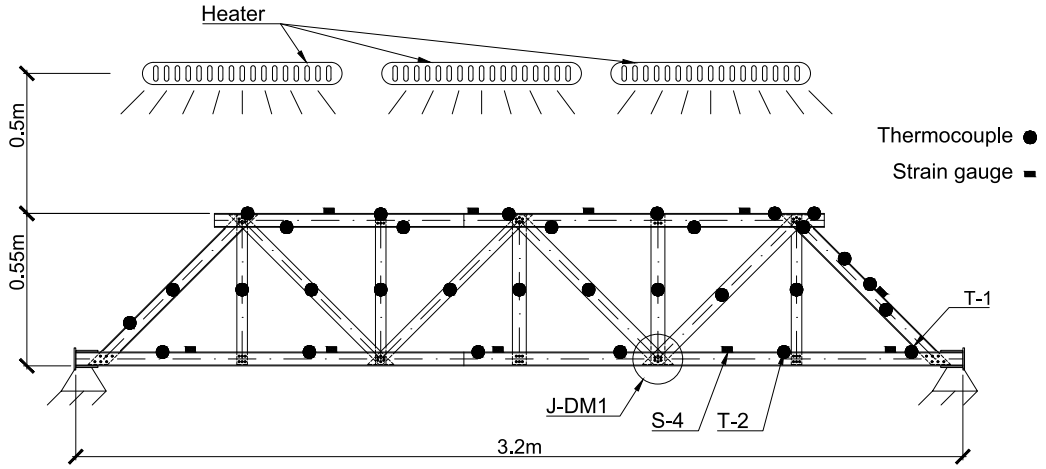

Fig. 3. Schematic of the truss setup in the laboratory of University of Exeter, UK. Both supports are fixed. Only the strain gauge labeled S-4 and the thermocouples labeled T-1 and T-2 are used in this article. J-DM1 indicates the joint where three bolts were removed on 2 September 2013 to simulate damage. The approximate location of the three heating lamps is also shown.

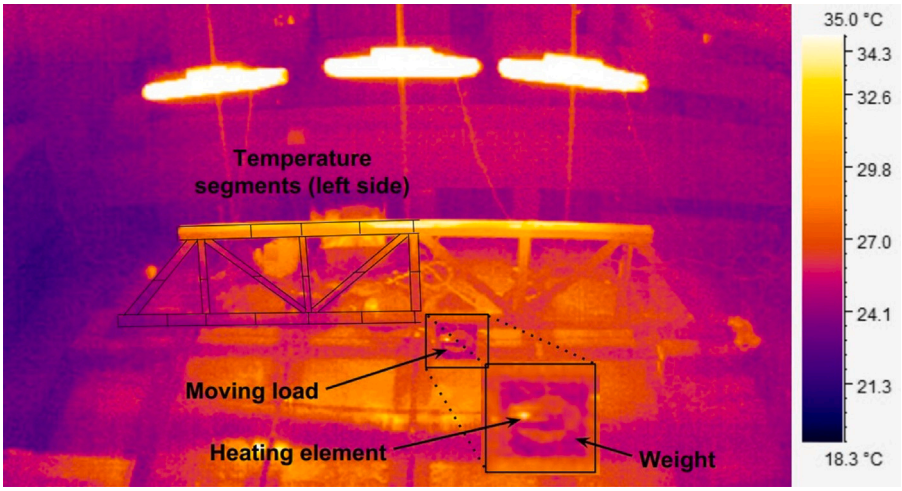

Fig. 4. Thermal image of the experimental setup, courtesy of [30].

illumination of any lamp, since the ensemble of the three lamps is slightly skewed towards the left of the truss. Therefore, the largest observable time-lags due to thermal inertia in this setup are expected to exist between pairs of strain and temperature sensors where one of the two is located in the illuminated area of the truss and the other towards one of the shaded lateral corners, particularly the right one.

As discussed in [30], the bottom chord of the truss was also instrumented with a mobile platform driven by a motor with the aim of simulating the effects of vehicular traffic (such moving load is indicated in Fig. 4). The platform was set to move on average once per day, each time for a total duration of anything between 1 and 3 h, and before each movement it was loaded with one of five possible weights, namely 0, 40, 100, 140 or 180 N. At all other times the platform was kept stationary and without added weight.

Data was acquired between 15 August and 16 September 2013, and over this period three artificial damages were introduced, though only the first damage (denoted DM1) is relevant for this article. Such damage was introduced on 2 September by removing three bolts (out of six) from the joint labeled J-DM1 in Fig. 3. This article only uses data collected up to the evening of 5 September, which include about 18 days of undamaged truss and the full period of 3 days in which DM1 was present. Data after 5 September was excluded for the reasons expressed in the next section.

3.1. Pair of sensors chosen for the study

The pair of strain and temperature sensors that was selected to carry out this study consists of strain gauge S-4 and thermocouple T-1 (Fig. 3). S-4 was chosen as it was expected to show some sensitivity to the simulated damage DM1 due to its proximity to J-DM1.

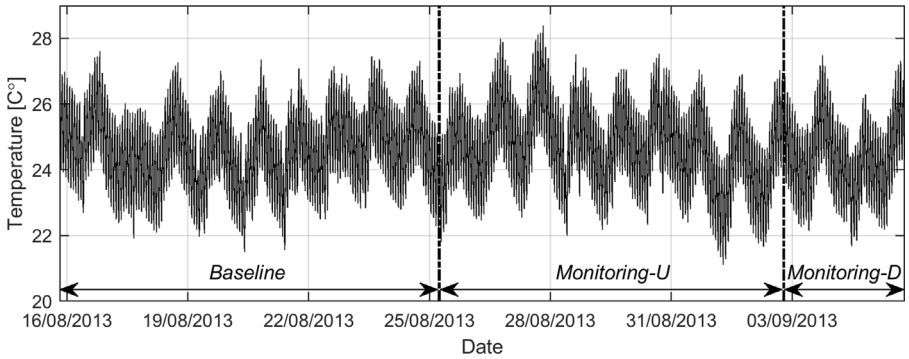

Fig. 5. Temperature measured by thermocouple T-1 at 10-s intervals between 15 August and 5 September 2013. The time frame is split into a baseline ("Baseline") and two monitoring sets ("Monitoring-U" and "Monitoring-D") by the two vertical lines. The rightmost line indicates the time of insertion of damage DM1 on the evening of 2 September.

Since S-4 is well within the illuminated area, thermocouple T-1, which is located at the far end of the right corner, was then chosen with the aim of maximizing the effects of thermal inertia and, therefore, of the time-lag existing between strain and temperature traces.

Fig. 5 shows the temperature history measured by T-1 in the time frame of interest, which is split into three periods by the two black vertical lines, the rightmost of which indicates the time of insertion of damage DM1 on the evening of 2 September. The period running from the start to the early morning of 25 August is used as baseline to train the regression models (as in the left box of Fig. 2), while the rest constitutes two monitoring sets, one undamaged ("Monitoring-U") and one damaged ("Monitoring-D"). Note that shortly after the introduction of a second damage on the evening of 5 September, the overall ambient temperature in the laboratory cooled down rather abruptly below the temperature range available in the baseline set (namely, the T-1 reading started to cycle in the ~19 to ~25 °C range rather than the ~22 to ~28 °C encompassed by the baseline set). Previous works performed by the first author of this article in the field of guided wave-based SHM have shown that any thermo-mechanical data-driven model constructed on a given set of baseline data can only characterize the structural behavior within the temperature range encompassed by the baseline set, and the use of the trained model outside the baseline range typically leads to very inaccurate predictions [46,47]. For this reason, this study only considers data collected up and including the first damaged scenario, i.e. the whole period shown in Fig. 5. Note that the practical implications of such requirement for the monitoring of actual structures are discussed in the Conclusions section.

4. Results and discussion

4.1. Estimate of an "optimal" time-shift via cross-correlation

As discussed in Section 2.3, a procedure was performed to estimate an "optimal" time-shift between temperature and strain time-series. For the recordings of strain gauge S-4 and thermocouple T-1 in the baseline set, the lagged cross-correlation gives the maximum value when S-4 is delayed by 35 samples (~6 min), or, conversely, T-1 is anticipated by the same amount, meaning that S-4 variations typically precede those of T-1. This is expected, since the structural location of S-4 and its surrounding are directly illuminated by the lamps, hence this area is rather quickly affected by the thermal loads of the heating cycles. Conversely, T-1 is located in the shaded area at the far right corner of the truss, hence its measures are more largely affected by thermal inertia effects that slow down the heat transfer from the heated area.

Fig. 6(a–b) show the recordings of S-4 and T-1 zoomed in the early morning of 24 August 2013, both (a) before and (b) after the imposed shift of 35 samples. Note that in this case it was arbitrarily chosen to fix the strain readings at their original temporal locations and to back-shift temperature. The same data plotted in Fig. 6(a–b) is also used to produce the "temperature vs strain" scatter plots of Fig. 6(c–d), respectively, which clearly show the enhanced correlation between temperature and strain after the shift. The unperturbed pair of signals of plot (a) were used when testing "Option I", while those of plot (b) were used for "Option II". For comparison, Fig. 6(e–f) plot the readings of strain gauge S-4 together with those of thermocouple T-2, which is closer to S-4 than T-1, as seen in Fig. 3. For this pair of signals, the "optimal" time-shift was estimated at only 3 samples (30 s), in this case with temperature T-2 preceding strain S-4. Again, in Fig. 6(f) strain is fixed at its original position and temperature is forward-shifted by 3 samples, though the differences with plot (e) are almost indiscernible. Note that, as expected, the temperatures measured by T-2 reach higher values than those of T-1.

4.2. Predictions from the regression models

The readings of T-1 and S-4 in the baseline period were used to train LR, MLR and WaveNet models under both options I and II. For both MLR and WaveNet one needs to choose the number n of independent variables $x_1, x_2, \ldots, x_n$ as indicated in

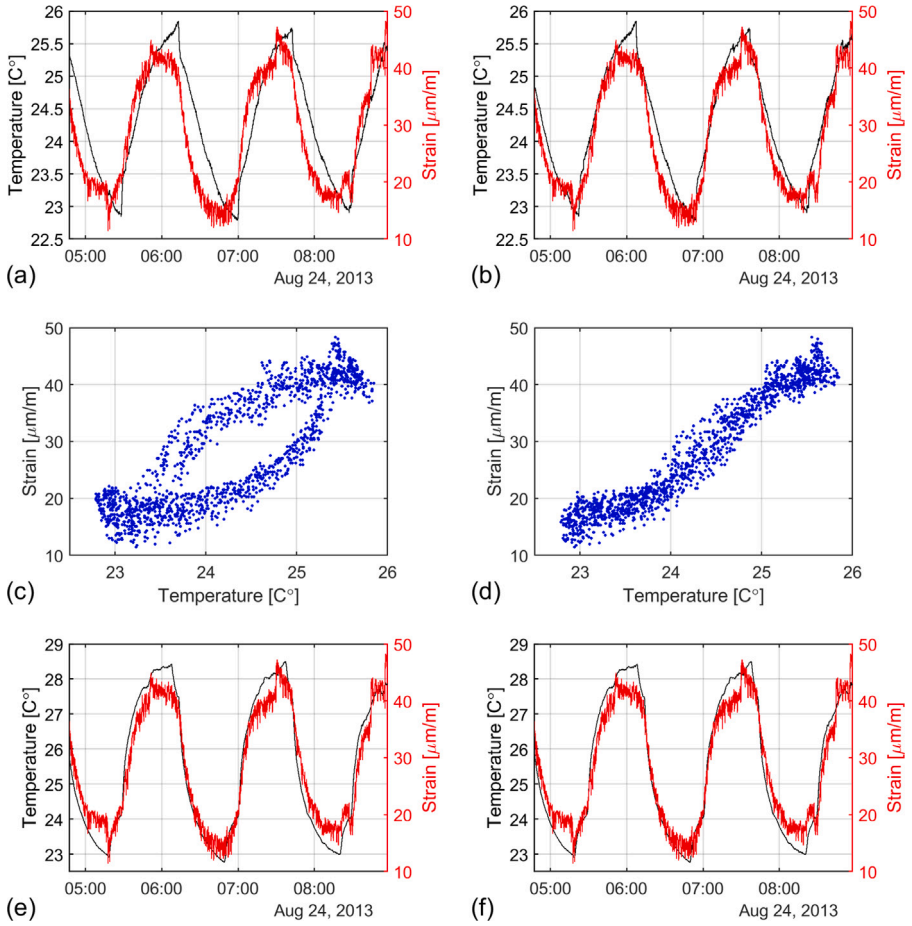

Fig. 6. (a-b) Readings of strain gauge S-4 (in red) and thermocouple T-1 (in black); (a) shows both readings at their original temporal locations, while in (b) T-1 is back-shifted by 35 samples (~6 min). (c–d) "Temperature vs strain" scatter plots using the same data shown in (a–b), respectively. (e–f) Readings of strain gauge S-4 (in red) and thermocouple T-2 (in black); in (c) they are both at their original locations, while in (d) T-2 is forward-shifted by 3 samples (30 s). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Eq. (2), i.e. the length of temperature history input to the model. Clearly, the choice on the value n will be highly dependent on the employed sampling interval and on the typical rates of thermal variations expected in the monitored structure. For the case at hand, the influence of varying n on the performance of MLR was studied by increasing it from 1 (i.e. the LR scenario) to a rather large value of 1801, which corresponds to a 5 hours-difference between the times of measurement of x_1 and x_n, i.e. slightly more than the duration of three heating cycles. Fig. 7 summarizes the results of the study on both "Option I" (black lines) and "Option II" (red lines) scenarios by plotting three values for the coefficient of determination R^2, respectively computed on the baseline (solid lines), the undamaged monitoring (dashed lines) and the damaged monitoring (dash-dotted lines) sets, as n is increased. Note that the horizontal axis in Fig. 7 actually refers to number of past temperature values (where the "current" time would be the time when strain is measured or the "optimal" time-shift from the latter under options I or II, respectively), i.e. to $n-1$, hence the results of LR are found on the y-axis for the abscissa equal to 0. The symbols used for LR are crosses, empty circles and solid circles for baseline, undamaged monitoring and damaged monitoring sets, respectively. Also note that plot (b) is zoomed from plot (a), and that plot (a) also shows the time required for the training of MLR as $n-1$ increases on a PC mounting an Intel(R) Core(TM) i7-6700 CPU at 3.40 GHz.

An inspection of Fig. 7 shows that all lines sharply increase as $n-1$ is incremented from 0 up to about 300 (corresponding to a time-interval of 50 min), after which they all tend to plateau. This indicates that the performance of MLR enhances significantly as the considered length of temperature history is increased to about half a heating cycle, though some additional improvements are obtained as such length is risen further. The fact that R^2 for the damaged monitoring sets remain quite lower than for the other sets indicates that, as expected, the strain measured by S-4 is sensitive to damage DM1, therefore the strain values predicted by the regression models trained on the baseline data tend to deviate from the actual measured values in the damaged monitoring set. A comparison of red and black lines show that the time-shift operated in "Option II" is generally beneficial, although significant advantages are only seen for very low $n-1$ values, e.g. the red symbols corresponding to LR in "Option II" are quite higher than the

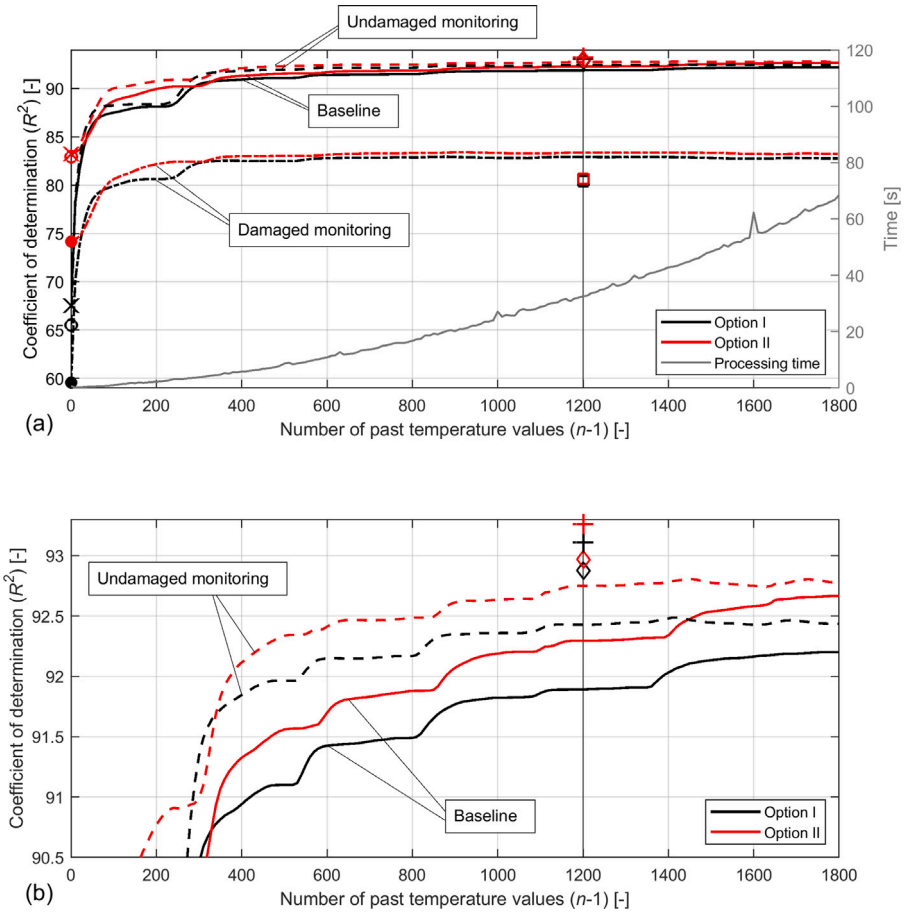

Fig. 7. R^2 obtained in the baseline (B), undamaged monitoring (M-U) and damaged monitoring (M-D) sets as a function of past temperature values ($n-1$) used in the time-series input to the regression models. (b) is zoomed from (a). The results obtained in options I and II are plotted in black and red, respectively. LR corresponds to $n-1=0$ and uses crosses, empty circles and solid circles for R^2 in B, M-U and M-D, respectively. MLR uses solid, dashed and dash-dotted lines for R^2 in B, M-U and M-D, respectively. WaveNet uses diamonds, plus signs and squares for R^2 in B, M-U and M-D, respectively. (a) also plots in a gray line the time required for the training of MLR. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

black ones of LR in "Option I". Another interesting feature of the plot is that generally the dashed lines corresponding to R^2 in the undamaged monitoring period are slightly higher than the solid lines corresponding to the baseline set. This is likely a result of the effects of the moving loads that were activated on the structure at different time-intervals, running duration and weight intensities over the course of the data collection time frame. At all cases, the evolution of the plots that essentially grow in tandem suggests that overfitting of the models to the baseline set is not occurring at least until $n-1$ increases over ~1200, when the solid lines keep increasing, though only slightly, while the dashed lines remain essentially flat. For this reason, $n-1$ equal to 1200 (corresponding to 3 h and 20 min of temperature recording) is selected as an approximately optimal value for MLR in this test setup, and this value is also used for WaveNet.

In order to both reduce computational times and to combat overfitting on the training data, the baseline set given to WaveNet for training was downsampled by a factor of 22, i.e. only 1 every 22 strain values (corresponding to a time-interval of 220 s) and corresponding history of temperatures was retained. This downsampled baseline data was then split into training and validation sets at a 90/10% ratio, while both undamaged and damaged monitoring data were only used for testing, as indicated by the flowchart in Fig. 2, and the retained trained model was that giving the lowest validation loss. The training was set to last for a total of 50 epochs, a duration which proved to be sufficient to reach low plateaus on both training and validation losses. On average, training such WaveNet model at $n-1=1200$ in either option I or II required ~14 min on the same machine used for MLR training. This compares to ~30 s needed for MLR at the same value of $n-1=1200$, as seen in Fig. 7(a).

Since the parameters of any machine learning model are always trained to at least slightly different values each time the training is repeated, WaveNet was trained five times for both options I and II. Then, the models that were retained for the two options were those producing the highest R^2 in the baseline set. Table 1 gives the R^2 values obtained by the ten WaveNet models (five per each option) in the full baseline set (rather than in the downsampled set used for training) and in the two monitoring sets, with the

Table 1
Coefficient of determination R^2 obtained in the baseline (B), undamaged monitoring (M-U) and damaged monitoring (M-D) by LR ($n-1=0$), MLR (for $n-1=1200$) and WaveNet (for $n-1=1200$) under both options I and II. WaveNet was trained five times for each option giving the results labeled WNT(1) to WNT(5), which are ordered in descending order for R^2 in the baseline set. WNT(1) were retained as the final models, while WNT(2) to WNT(5) were discarded.

	Option I			Option II		
	B	M-U	M-D	B	M-U	M-D
LR	67.51	65.52	59.57	83.23	82.94	74.15
MLR	91.89	92.43	82.91	92.30	92.75	83.37
WNT(1)	92.87	93.11	80.45	92.97	93.26	80.67
WNT(2)	92.82	93.06	81.72	92.83	92.98	81.56
WNT(3)	92.81	93.12	81.95	92.75	93.18	81.49
WNT(4)	92.68	93.20	81.52	92.72	93.10	80.91
WNT(5)	92.53	92.98	80.66	92.61	93.02	80.27

models listed in descending order for R^2 in the baseline set. The retained WaveNet models for options I and II are denoted WNT(1), while WNT(2) to WNT(5) were discarded. Table 1 also shows the R^2 produced by LR (corresponding to the symbols on the y-axis at $n-1=0$ in Fig. 7) and by MLR for $n-1=1200$.

The R^2 values produced by WNT(1) are also plotted in Fig. 7 to facilitate the comparison with MLR. Again, black and red symbols correspond to options I and II, respectively, and in this case diamonds, plus signs and squares indicate baseline, undamaged monitoring and damaged monitoring sets, respectively. The trend for which "Option II" is beneficial is confirmed also when using WaveNet, however in this case the increases in R^2 are rather marginal. For both options, WaveNet returns higher R^2 than MLR in the two undamaged sets, and lower R^2 for the damaged monitoring set. This behavior indicates that WaveNet models have built closer relations between histories of temperature and strain in the undamaged scenarios, such that when the structure is damaged the predictions deviate more significantly from the actual measured strains. An example of this is seen in Fig. 8, which superposes the measured strains to those predicted by WNT(1) in "Option I" over two periods of 12 h within the undamaged (plot(a)) and damaged (plot(b)) monitoring sets. The predictions in the undamaged period closely follow the actual values, while those in the damaged set are significantly distinct from the measured ones. In particular, the measured strain values shown in Fig. 8(b) are generally higher than the predicted ones, indicating that the truss member where the strain gauge S-4 is installed is subjected to higher tensile strains than those that would be expected under the given history of temperature. Note that the increased discrepancy between measured and predicted strains seen roughly between 14:30 and 16:00 in Fig. 8(a) is due to the activation of the moving load passing on the structure over the course of those 90 min.

The R^2 computed on the various undamaged and damaged sets of data are useful to characterize and compare the performance offered by the tested regression models. However, practically, the time of occurrence of damage is not known in advance and, on the contrary, the onset of damage is what needs to be detected. Therefore, one needs to set a monitoring strategy potentially able to automatically inspect incoming data and flag suspicious behaviors. A possible such strategy is discussed in the next section, and the results are used to further compare the performances offered by the tested regression models.

4.3. Moving average of the squared difference between measured and predicted strains

After a regression model has been trained on the baseline set, the squared difference (SD) between measured and predicted strains can be computed and tracked over time, and an alarm can be flagged when this exceeds a pre-determined threshold value. Clearly, isolated spikes of SD can cross the threshold even in the absence of damage when a faulty measurement occurs e.g. due to some sensor or data acquisition issue, though various methods exist to effectively detect and remove such outliers from the analysis [10]. It is more challenging, instead, to distinguish a series of large SD values due e.g. to some transitory load (such would be the case of the moving platform causing the discrepancies seen in Fig. 8(a)) from permanent changes induced by damage. Both types of issues can be attenuated by computing moving averages of the SD data obtained over time, a quantity that will be denoted MA-SD in the remainder of the article. To this scope, care should be taken when setting the temporal length of the sliding window, which should be long enough to effectively reduce isolated or short-termed large SD values, but not too long as that the delay in calling a true alarm would become unacceptable. Another trade-off characterizes the choice on the call-threshold level, which should be large enough to limit the number of false calls, but not too large as to miss the signatures of actual damage. As a matter of example, Fig. 9(a–f) plot trailing moving averages obtained by using a sliding window of 7200 samples (corresponding to 20 h of measurements) on the SD produced by the regression models discussed in this article over the full ~21 days-period and for both options I and II. In each plot the red horizontal line is the threshold level, which was conservatively set as twice the largest MA-SD value obtained in the baseline set (as similarly done in [48]), and whose exact value is given in the legend. As in Fig. 5, the black vertical lines indicate the times of transition between baseline and monitoring sets. Fig. 9(a–b) show the MA-SD given by LR in options I and II, respectively. As already indicated by the low R^2 values seen in Fig. 7 and Table 1, the "standard" LR based on temperature and strain pairs acquired at the same temporal instants (i.e. "Option I" that is plotted in (a)) cannot accurately predict strain, hence large MA-SD are obtained even in the baseline set, and the corresponding threshold is out of the plot at this zoom-level. Regardless of the threshold, the MA-SD values in the damaged monitoring period remain at comparable levels to those seen in the previous undamaged periods, therefore it would be impossible to consistently call this or analogous damages without also calling

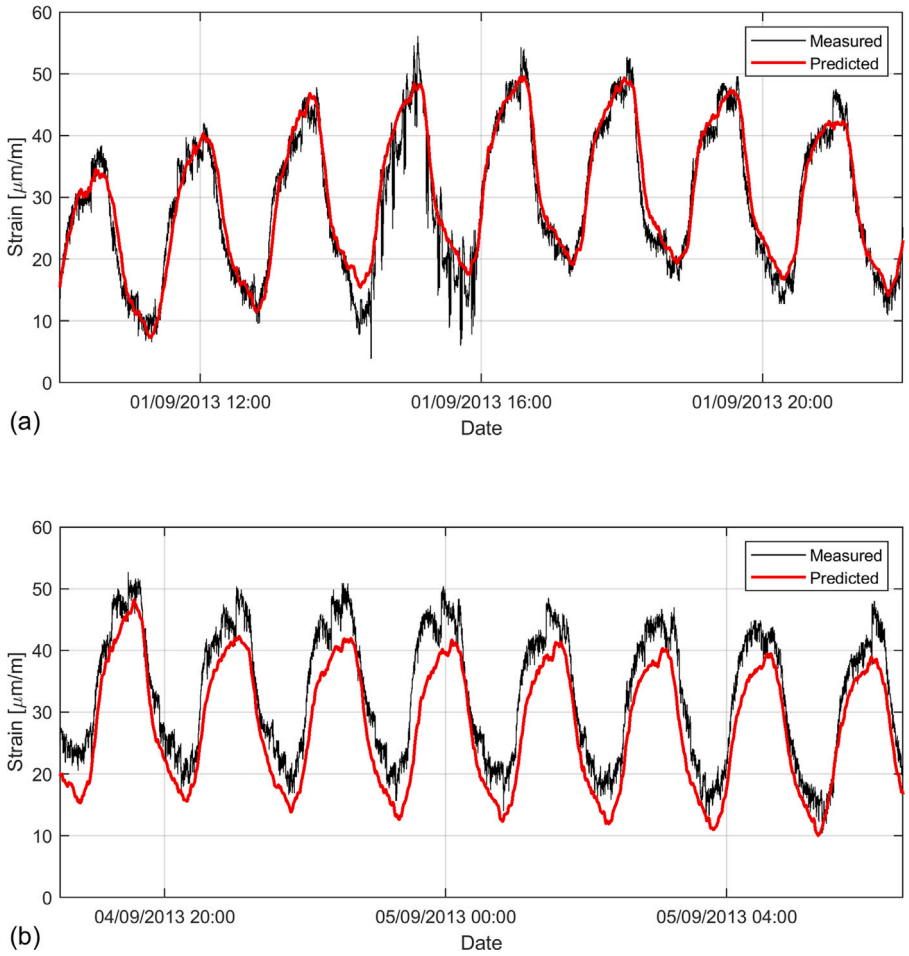

Fig. 8. Measured strains (black) vs those predicted by WNT(1) (red) in "Option I" over 12 h periods within the (a) undamaged and (b) damaged monitoring sets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

an unacceptable large number of false alarms. The "optimal" time-shifting operated in "Option II", which has already been shown to significantly improve the performance of LR in terms of R^2, yields overall lower values for MA-SD, as seen in Fig. 9(b), though without effectively enabling the detection of the considered damage. Significantly better results are obtained when using either MLR or WaveNet with temperature histories comprising 1200 past samples. Note that Fig. 9(c–f) are at a different zoom-level than (a–b). Fig. 9(c–d) consider MLR under options I and II, respectively. In both scenarios the MA-SD values in the damaged region are considerably larger than those in the two undamaged sets. Despite the slight improvement in terms of R^2 seen in Fig. 7 and Table 1, there appears to be very limited advantage in using "Option II" over the standard "Option I", though the threshold in plot (d) is indeed slightly lower than that in (c). In either case, this damage would not be called at this specific threshold and within the available time frame, but clearly a slightly different choice for the threshold would enable the detection. Finally, Fig. 9(e–f) show a further improvement when using WaveNet over MLR. The thresholds for both options I and II are lower than the corresponding ones in MLR, and the MA-SD values reached in the damaged set are larger than those produced by MLR. As a result, in both options I and II this damage would be called under this particular choice for the threshold and within the available time frame. As for MLR, the improvement given by "Option II" over "Option I" remains marginal.

5. Conclusions

Temperature variations cause structural responses in large structures such as bridges that are typically on the same order or higher than those induced by any other load or by the insurgence or growth of damage. The complex relation between temperature and structural changes is governed by external thermal loads and internal factors such as thermal inertia, which is responsible for the delays typically seen between responses measured at different locations. This article has investigated the capability of two regression models, namely MLR and a WaveNet-based algorithm, to predict strain measured on a test structure based on long temporal sequences of temperature readings acquired on a different location of the same structure. The performances of the two

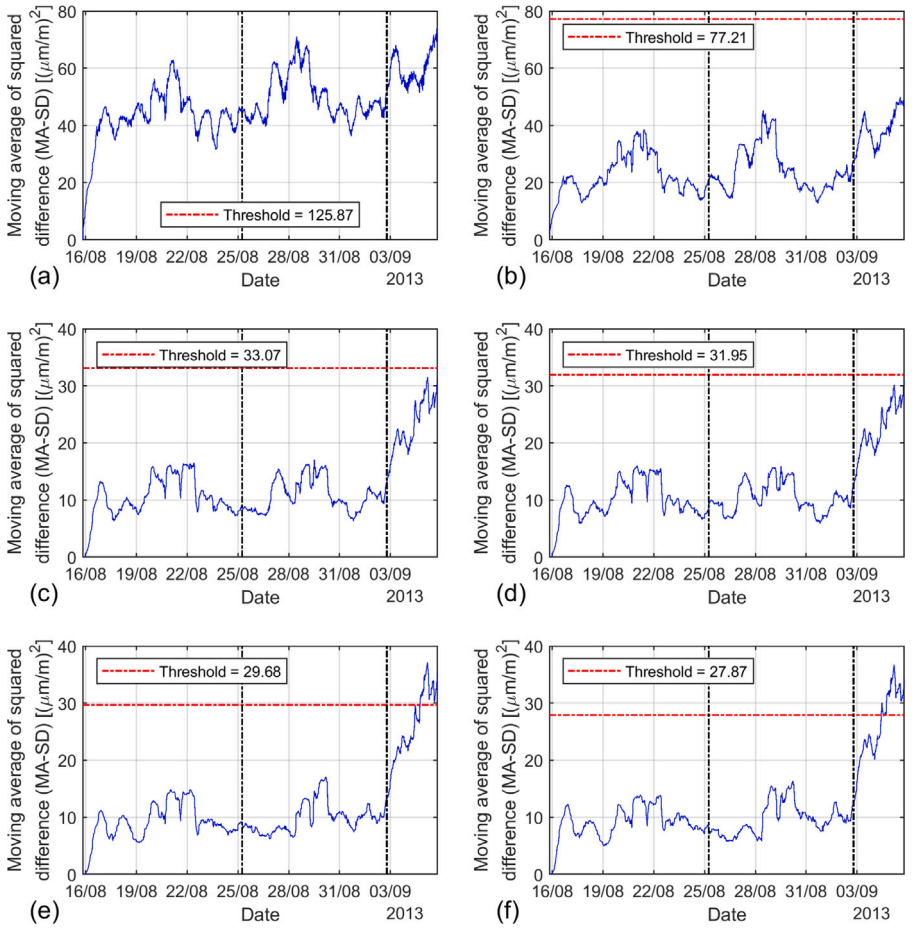

Fig. 9. Trailing moving averages with sliding window of 7200 samples (corresponding to 20 h of measurements) computed on the squared differences between measured strains and those predicted by (a-b) LR, (c–d) MLR at $n-1=1200$ and (e–f) WaveNet at $n-1=1200$, under options (a,c,e) I and (b,d,f) II. In each plot the threshold was set as twice the largest MA-SD value found in the baseline set. As in Fig. 5, the black vertical lines indicate the times of transition between baseline and monitoring sets. Note that the zoom-level in the y-axis of (a–b) is different than that of (c–f).

models were evaluated in terms of R^2 computed between predicted and measured strain values, and were also compared to that offered by simple LR that uses single temperature readings to predict strain. It has been shown that the two former models largely outperform LR, with the WaveNet-based model achieving the best results overall, though at the cost of being computationally more expensive. A further scenario was considered in which temperature and strain sequences were better aligned prior to the regression operation. This was accomplished by time-shifting temperature by the time-lag giving the highest lagged cross-correlation between the two time-series. This pre-processing operation was shown to be highly beneficial for LR, though its performance remained significantly poorer than that of the other two models, and only marginally beneficial for MLR and WaveNet.

Finally, a data-driven SHM strategy was presented and applied to the considered experimental dataset, which also included measurements taken after an artificial damage was introduced. The strategy requires an initial phase where baseline data is collected and used to train one of the considered regression models in order to characterize the structure in its initial, potentially undamaged condition. Such training period is then followed by the active monitoring phase where damage can be automatically detected. The application of this SHM strategy offered a second test-bed to compare the performances offered by LR, MLR and the WaveNet-based model, for both scenarios of unperturbed and time-shifted temperature prior to regression. All findings based on R^2 as discussed above were confirmed by the analysis of the results, with the WaveNet-based model being the only one successfully enabling the detection of the considered damage within the available time-frame and under the chosen length of moving average window and method for setting the threshold.

It is worth noting that all analyses discussed in this article have been performed on data acquired over the course of ~21 days in a relatively small aluminum truss setup in a laboratory environment, where the only imposed external loads were thermal cycles exerted by heating lamps and a moving platform engineered to simulate the effects of vehicular traffic, which was set in motion for about 10% of the total measurement time. Therefore, the proposed procedures will have to be validated on more challenging structures where vehicular traffic can be heavier and more frequent, where other significant loads may exist, such as those due

to wind, and where the employed measurement systems can be affected by higher levels of noise. Similarly, structures composed of other materials, such as concrete or masonry, can be considered. Lastly, with regards to the proposed SHM strategy, it must be noted that for the trained regression model to be able to fully characterize the given monitored structure, the baseline data should encompass the full temperature range expected during its service life. Practically, this translates into a minimum duration for the baseline period of at least six months, as required for temperature to cycle between minimum and maximum seasonal values.

CRediT authorship contribution statement

S. Mariani: Conceptualization, Data curation, Formal analysis, Methodology, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **A. Kalantari:** Formal analysis, Methodology. **R. Kromanis:** Investigation, Writing – review & editing. **A. Marzani:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Eloi Figueiredo, James Brownjohn, Three decades of statistical pattern recognition paradigm for SHM of bridges, Struct. Health Monit. 21 (6) (2022) 3018–3054.
[2] Eloi Figueiredo, Ionut Moldovan, Manuel Barata Marques, Condition assessment of bridges: Past, present, and future. A complementary approach, Univ. Católica Ed. (2013) 199.
[3] Hoon Sohn, et al., A review of structural health monitoring literature: 1996–2001, Los Alamos Natl. Lab. USA 1 (2003) 16.
[4] Charles R. Farrar, Keith Worden, Structural Health Monitoring: A Machine Learning Perspective, John Wiley & Sons, 2012.
[5] Jerome P. Lynch, Kenneth J. Loh, A summary review of wireless sensors and sensor networks for structural health monitoring, Shock Vib. Dig. 38 (2) (2006) 91–130.
[6] Tong Guo, et al., Displacement monitoring and analysis of expansion joints of long-span steel bridges with viscous dampers, J. Bridge Eng. 20 (9) (2015) 04014099.
[7] A. Alvandi, C. Cremona, Assessment of vibration-based damage identification techniques, J. Sound Vib. 292 (1-2) (2006) 179–202.
[8] F. Necati Catbas, Melih Susoy, Dan M. Frangopol, Structural health monitoring and reliability estimation: Long span truss bridge application with environmental monitoring data, Eng. Struct. 30 (9) (2008) 2347–2359.
[9] James-A. Goulet, Prakash Kripakaran, Ian F.C. Smith, Multimodel structural performance monitoring, J. Struct. Eng. 136 (10) (2010) 1309–1318.
[10] Daniele Posenato, et al., Methodologies for model-free data interpretation of civil engineering structures, Comput. Struct. 88 (7–8) (2010) 467–482.
[11] João Pedro Santos, et al., Multivariate statistical analysis for early damage detection, Eng. Struct. 56 (2013) 273–285.
[12] Bart Peeters, Guido De Roeck, One-year monitoring of the Z24-Bridge: environmental effects versus damage events, Earthq. Eng. Struct. Dyn. 30 (2) (2001) 149–171.
[13] Guang-Dong Zhou, Ting-Hua Yi, et al., A summary review of correlations between temperatures and vibration properties of long-span bridges, Math. Probl. Eng. 2014 (2014).
[14] Yong Xia, et al., Temperature effect on vibration properties of civil structures: a literature review and case studies, J. Civ. Struct. Health Monit. 2 (2012) 29–46.
[15] Ki-Young Koo, et al., Structural health monitoring of the Tamar suspension bridge, Struct. Control Health Monit. 20 (4) (2013) 609–625.
[16] E.J. Cross, et al., Long-term monitoring and data analysis of the Tamar bridge, Mech. Syst. Signal Process. 35 (1–2) (2013) 16–34.
[17] X.G. Hua, et al., Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique, J. Comput. Civ. Eng. 21 (2) (2007) 122–135.
[18] Y.Q. Ni, et al., Correlating modal properties with temperature using long-term monitoring data and support vector machine technique, Eng. Struct. 27 (12) (2005) 1762–1773.
[19] Brock D. Hedegaard, Catherine E.W. French, Carol K. Shield, Effects of cyclic temperature on the time-dependent behavior of posttensioned concrete bridges, J. Struct. Eng. 142 (10) (2016) 04016062.
[20] Brock D. Hedegaard, Catherine E.W. French, Carol K. Shield, Time-dependent monitoring and modeling of I-35W St. Anthony Falls Bridge. I: Analysis of monitoring data, J. Bridge Eng. 22 (7) (2017) 04017025.
[21] Youliang Ding, et al., Thermal field characteristic analysis of steel box girder based on long-term measurement data, Int. J. Steel Struct. 12 (2012) 219–232.
[22] Hanbing Liu, Xianqiang Wang, Yubo Jiao, et al., Effect of temperature variation on modal frequency of reinforced concrete slab and beam in cold regions, Shock Vib. 2016 (2016).
[23] Fabio Casciati, et al., Estimating thermal inertia and temperature distribution consistent with monitored data from a railway bridge, in: Proceedings of the 4th ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, 2013, pp. 12–14.
[24] Gabriele Comanducci, et al., On vibration-based damage detection by multivariate statistical techniques: Application to a long-span arch bridge, Struct. Health Monit. 15 (5) (2016) 505–524.
[25] James M.W. Brownjohn, et al., Structural health monitoring of short to medium span bridges in the United Kingdom, Struct. Monit. Maint. 3 (3) (2016) 259.
[26] Kang Yang, et al., Modeling of temperature time-lag effect for concrete box-girder bridges, Appl. Sci. 9 (16) (2019) 3255.
[27] Chuang Chen, et al., Reliability assessment for PSC box-girder bridges based on SHM strain measurements, J. Sens. 2017 (2017).
[28] Xiaoyu Gong, et al., A temperature-driven approach for quantitative assessment of strengthening effect of continuous bridges using structural health monitoring data, Struct. Health Monit. (2023) 14759217231181882.
[29] Rolands Kromanis, Prakash Kripakaran, Predicting thermal response of bridges using regression models derived from measurement histories, Comput. Struct. 136 (2014) 64–77.

[30] Rolands Kromanis, Prakash Kripakaran, Data-driven approaches for measurement interpretation: analysing integrated thermal and vehicular response in bridge structural health monitoring, Adv. Eng. Inform. 34 (2017) 46–59.
[31] Ian T. Jolliffe, Principal Component Analysis for Special Types of Data, Springer, 2002.
[32] Vladimir Vapnik, The Nature of Statistical Learning Theory, Springer science & business media, 1999.
[33] Hoon Sohn, et al., An experimental study of temperature effect on modal parameters of the Alamosa Canyon Bridge, Earthq. Eng. Struct. Dyn. 28 (8) (1999) 879–897.
[34] David A. Freedman, Statistical Models: Theory and Practice, cambridge University Press, 2009.
[35] Lennart Ljung, System identification, in: Signal Analysis and Prediction, Springer, 1998, pp. 163–173.
[36] Chengwei Wang, et al., LSTM approach for condition assessment of suspension bridges based on time-series deflection and temperature data, Adv. Struct. Eng. 25 (16) (2022) 3450–3463.
[37] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[38] Mingqiang Xu, et al., Structural damage detection by integrating robust PCA and classical PCA for handling environmental variations and imperfect measurement data, Adv. Struct. Eng. 25 (8) (2022) 1815–1828.
[39] Aaron van den Oord, et al., Wavenet: A generative model for raw audio, 2016, arXiv preprint arXiv:1609.03499.
[40] Leona S. Aiken, Stephen G. West, Steven C. Pitts, Multiple linear regression, Handb. Psychol. (2003) 481–507.
[41] Hendrik Purwins, et al., Deep learning for audio signal processing, IEEE J. Sel. Top. Sign. Proces. 13 (2) (2019) 206–219.
[42] Shu Lih Oh, et al., Classification of heart sound signals using a novel deep WaveNet model, Comput. Methods Programs Biomed. 196 (2020) 105604.
[43] Stefano Mariani, et al., Causal dilated convolutional neural networks for automatic inspection of ultrasonic signals in non-destructive evaluation and structural health monitoring, Mech. Syst. Signal Process. 157 (2021) 107748.
[44] Suubi Racheal, Yongkuo Liu, Abiodun Ayodeji, Improved WaveNet for pressurized water reactor accident prediction, Ann. Nucl. Energy 181 (2023) 109519.
[45] George Casella, Roger L. Berger, Statistical Inference, Cengage Learning, 2021.
[46] Stefano Mariani, Sebastian Heinlein, Peter Cawley, Location specific temperature compensation of guided wave signals in structural health monitoring, IEEE Trans. Ultrason. Ferroelectr. Freq. Control 67 (1) (2019) 146–157.
[47] Stefano Mariani, Peter Cawley, Change detection using the generalized likelihood ratio method to improve the sensitivity of guided wave structural health monitoring systems, Struct. Health Monit. 20 (6) (2021) 3201–3226.
[48] Stefano Mariani, et al., Performance of a guided wave pipe monitoring system over extended periods of field operation, Mater. Eval. 81 (8) (2023) 36–46.