

## Multiscale features and information extraction of online strain for long-span bridges

Baijian Wu<sup>1a</sup>, Zhaoxia Li<sup>\*1</sup>, Tommy H.T. Chan<sup>2b</sup> and Ying Wang<sup>1c</sup>

<sup>1</sup>Department of Engineering Mechanics, Southeast University, Nanjing 210096, China

<sup>2</sup>School of Civil Engineering and Built Environment, Science and Engineering Faculty, Queensland University of Technology, Queensland, Australia

(Received May 14, 2013, Revised April 19, 2014, Accepted April 25, 2014)

**Abstract.** The strain data acquired from structural health monitoring (SHM) systems play an important role in the state monitoring and damage identification of bridges. Due to the environmental complexity of civil structures, a better understanding of the actual strain data will help filling the gap between theoretical/laboratorial results and practical application. In the study, the multi-scale features of strain response are first revealed after abundant investigations on the actual data from two typical long-span bridges. Results show that, strain types at the three typical temporal scales of  $10^5$ ,  $10^2$  and  $10^0$  sec are caused by temperature change, trains and heavy trucks, and have their respective cut-off frequency in the order of  $10^{-2}$ ,  $10^{-1}$  and  $10^0$  Hz. Multi-resolution analysis and wavelet shrinkage are applied for separating and extracting these strain types. During the above process, two methods for determining thresholds are introduced. The excellent ability of wavelet transform on simultaneously time-frequency analysis leads to an effective information extraction. After extraction, the strain data will be compressed at an attractive ratio. This research may contribute to a further understanding of actual strain data of long-span bridges; also, the proposed extracting methodology is applicable on actual SHM systems.

**Keywords:** health monitoring; bridges; strain; information extraction; multi-scale; wavelet

### 1. Introduction

Recently, many technical articles have arisen on the subject of strain-based health monitoring, which extend the original concept of Structural Health Monitoring –the so-called global damage identification based on dynamic responses (Housner *et al.* 1997). Some of these put an emphasis on the function of short/long-term ‘monitoring’ and evaluation on structural state or identification of the environmental events (Cardini and DeWolf 2009, Kamath *et al.* 2010, Li, *et al.* 2012, Li *et al.* 2001, Liu *et al.* 2009, Omenzetter *et al.* 2004, Ye *et al.* 2012), while some of others much concentrate on the function of ‘damage identification’ (Bukkapatnam *et al.* 2005, Hu and Shenton

---

\*Corresponding author, Professor, E-mail: zhxli@seu.edu.cn

<sup>a</sup> Ph.D. Student, E-mail: bjwu2@seu.edu.cn

<sup>b</sup> Professor, Email: tommy.chan@qut.edu.au

<sup>c</sup> Ph.D., E-mail: civil\_wangying@seu.edu.cn

2007, Katsikeros and Labeas 2009, Kesavan *et al.* 2008, Li and Wu 2008, Reynders *et al.* 2007). All the above point to the wide application and important role of strain data in structural health monitoring.

Due to the complexity of structural field environments, there still exists a distance between the laboratorial /theoretical results and practical applications (Farrar *et al.* 2001, Van der Auweraer and Peeters 2003). Consequently, it becomes extremely crucial to obtain clean, reliable and normalized data when we apply the developed techniques, which may have been validated in laboratories, on actual structures. The situation couldn't be an exception for strain-based health monitoring strategies, especially for long-span bridges which are usually under more complex operational environments. A better understanding on these strain data is then required if we want to use them confidently rather than abstractly.

Literature and publications on research concerning with the strain itself are relatively sparse, with the exception of aiming at developing a much robust and low-noise data acquisition system, for example in (Wong *et al.* 2006). Differently from these studies, online strains in this article are all supposed to be 'good' – not highly noisy – data which record the real response of structures, allowing for concentration on the structure of online strain itself. Considering the wide application of SHM systems on long-span bridges, investigations in this article should be helpful for researchers and users of them.

The strain data used in this article is from SHM systems installed on the Tsing-Ma Bridge (TMB) and Run-Yang Bridge (RYB). Tsing-Ma Bridge in Hong Kong with a 2.2 km total span and a main span of 1377 m, is the longest of suspension bridge that carries both railway and highway traffic. Run-Yang Yangtze Bridge, a highway bridge connecting Yang-Zhou and Zhen-Jiang, China, includes a suspension bridge and a cable-stayed bridge of which the suspension bridge with a main span of 1490 m ranks the third longest suspension bridge in the world.

In the article, the characteristic of online strain data is investigated first. Based on its multi-scale feature, multi-resolution analysis and wavelet shrinkage are introduced for separating and extracting the different strain types. A wavelet model for strain data is proposed in Section 4. Results and discussions are presented in Section 5.

## 2. Features of online strain at multiple temporal scales

The structural monitoring system in TMB comprises a total of approximately 900 sensors, including accelerometers, strain gauges, displacement transducers, level sensors, anemometers, temperature sensors and weigh-in-motion sensors, installed permanently on the bridges and the data acquisition and processing system. The strain gauges which were installed to measure strain in bridge-deck sections are shown in Fig. 1. The locations of strain gauges installed in the Tsing Ma Bridge include rail track sections at CH 24662.50, bridge-deck trough section at CH 24664.75 and deck at tower and rocker bearing links at CH 23623.00. Locations of some strain gauges in cross frame at CH 24662.50 are shown in Fig. 1.

Despite of the huge data size collected by hundreds of sensors installed on TMB and RYB, investigation shows that the strain data from sensors at different locations of one bridge exhibit similar features. For this reason, only two strain gauges are chosen as representations to illustrate their common characteristic. They are respectively numbered SSTLN01, a strain gauge installed on the top chord of the longitudinal truss in the cross-section CH24662.50 of TMB, as shown in Fig. 1, and YBH4-13, a gauge locating at the upper deck in the middle-span section of Run-Yang

suspension bridge. They were continuously recording the strain under working conditions at rates of 25.6 Hz and 20 Hz, respectively. A 24-hours strain history was selected for both gauges. Further detailed description on layouts of strain sensors could be seen in (Lau and Wong 1997) and (River 2004).

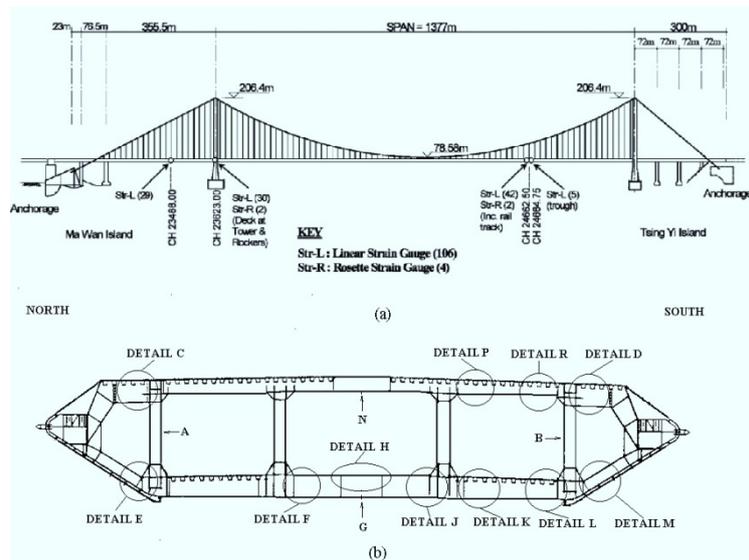
## 2.1 Analysis in the time domain

The characteristic of strain response can be described in brief as “multi-scale”. More specifically, the strain displays distinctive features at three typical temporal scales of  $10^6$  sec,  $10^3$  sec and  $10^0$  sec.

At the scale 1 (1 day or 86400 sec): As shown in Figs. 2(a), 3(a), 4(a), and 4(b), both strains demonstrate a trend line (Type 1) in 24 hours, and the varying magnitude is about  $100 \mu\epsilon$ . Moreover, this strain type has a periodic character with a cycle time of 24 hours because the daily curve is similar.

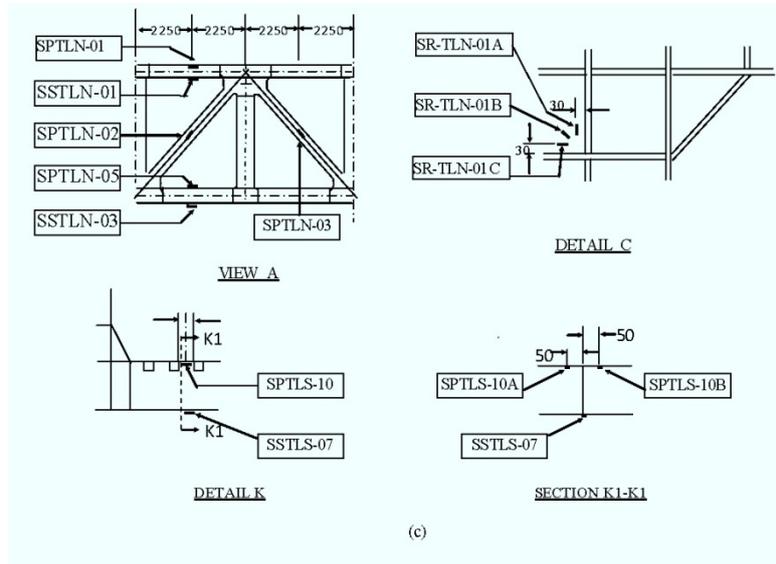
At the scale 2 (1000 sec): In SSTLN01 on Tsing-Ma Bridge, a new strain type (Type 2) could be recognized at this scale in Figs. 2(c) and 3(d). It could be described as a “local impulse” or a “local strain cycle” with the duration of approximately 20sec and order of variation magnitude  $10 \mu\epsilon$ . However, this strain type couldn't be found in Fig. 4(c) in Run-Yang Bridge.

At the scale 3 (1 sec): As seen in Figs. 3(h) and 4(e) for Tsing-Ma and Run-Yang Bridge, respectively, another strain type (Type 3) can be recognized at this time scale. Type 3 presents the feature of ‘local oscillating impulse’ or ‘consecutive local impulses’. The key point is that they all last about 1sec from the beginning to the end, displaying two, occasionally three, consecutive peaks or valleys. The variation range differs between the two bridges. For TMB, it is less than  $10 \mu\epsilon$ , with an average of several strain units, while for RYB, it could be large than  $50 \mu\epsilon$ , with an average of tens of strain units.



(a) Layout of strain gauges on Tsing-Ma Bridge, (b) Strain gauge location in the cross frame at CH24662.50

Continued-



(c) Some figures for details and views in Fig. 1(b)

Fig. 1 Layout and locations of strain gauges on Tsing-Ma Bridge

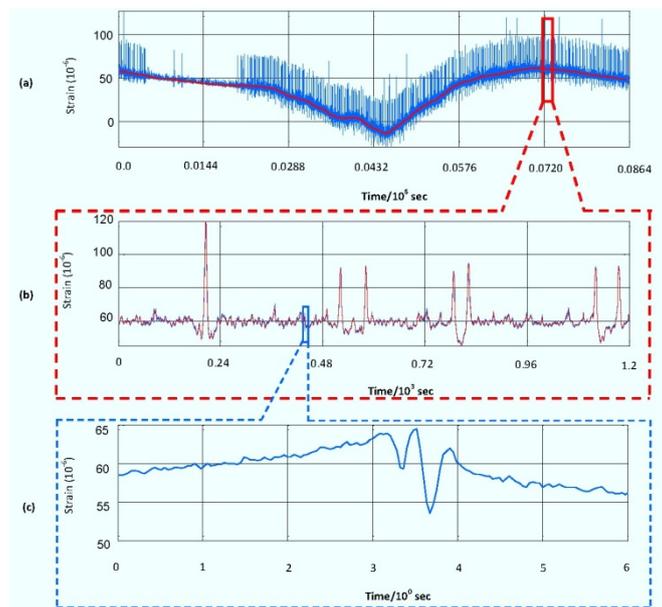


Fig. 2 The original strain history from SPTLN01 on Tsing-Ma Bridge at different scales: (a)  $10^6$  sec, at this scale only the strain caused by temperature variation could be recognized, (b)  $10^3$  sec, only strain under railway load could be recognized, (c)  $10^0$  sec, at this scale, the details of truck induced strain cycles could be clearly displayed

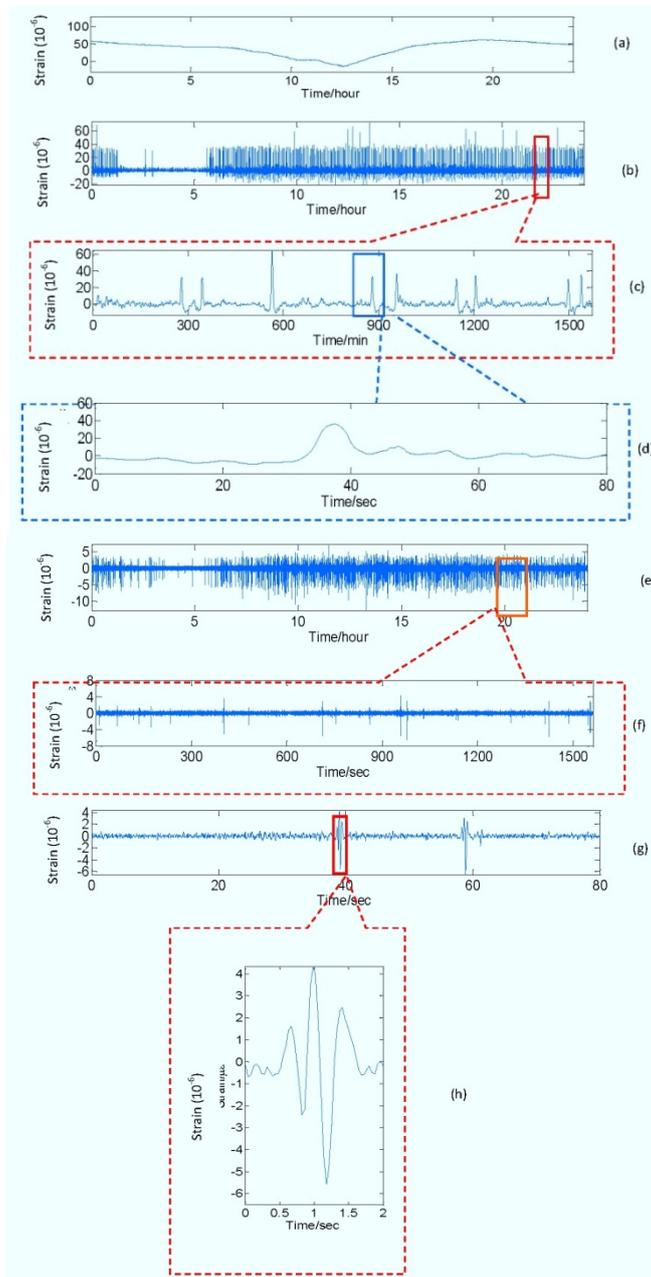


Fig. 3 The separated strain components from SSSLN01: (a) temperature strain  $\hat{\varepsilon}_t$  at scale 1(24hours), (b) railway strain  $\hat{\varepsilon}_r$  at scale 1, (c)  $\hat{\varepsilon}_r$  at intermediate scale (25min), (d)  $\hat{\varepsilon}_r$  at scale 2 (80sec), (e) highway strain  $\hat{\varepsilon}_h$  at scale 1 (24hours), this strain type could not be recognized, but globally, the range of strain cycles is much smaller than railway strain, (f)  $\hat{\varepsilon}_h$  at intermediate scale (25min), this type of strain still could not be recognized, (g)  $\hat{\varepsilon}_h$  at scale 2, the details are not so clear (80sec), (h)  $\hat{\varepsilon}_h$  at scale 3 (2sec)

It should be pointed out that other time scales were also chosen during the study, without noticeable strain types discovered. A further investigation indicates that the above three strain types are caused by events of temperature variation, train (railway loads) and heavy truck (highway loads), respectively. The first strain type reflects the daytime temperature change (Chan *et al.* 2001). The second strain type is the structural response to trains passing through the bridges, evidenced by the numerical simulation implemented by Chan *et al.* (2003). The third type is caused by heavy trucks of those the gross weight is above 14,000 Kg, where the consecutive two or three peaks/valleys are corresponding to the two or three lines of front and rear wheels. Additionally, the absence of the second strain type in RYB could be smoothly explained for TMB is a double-duty bridge for both railway and highway loads while RYB is singly a highway bridge.

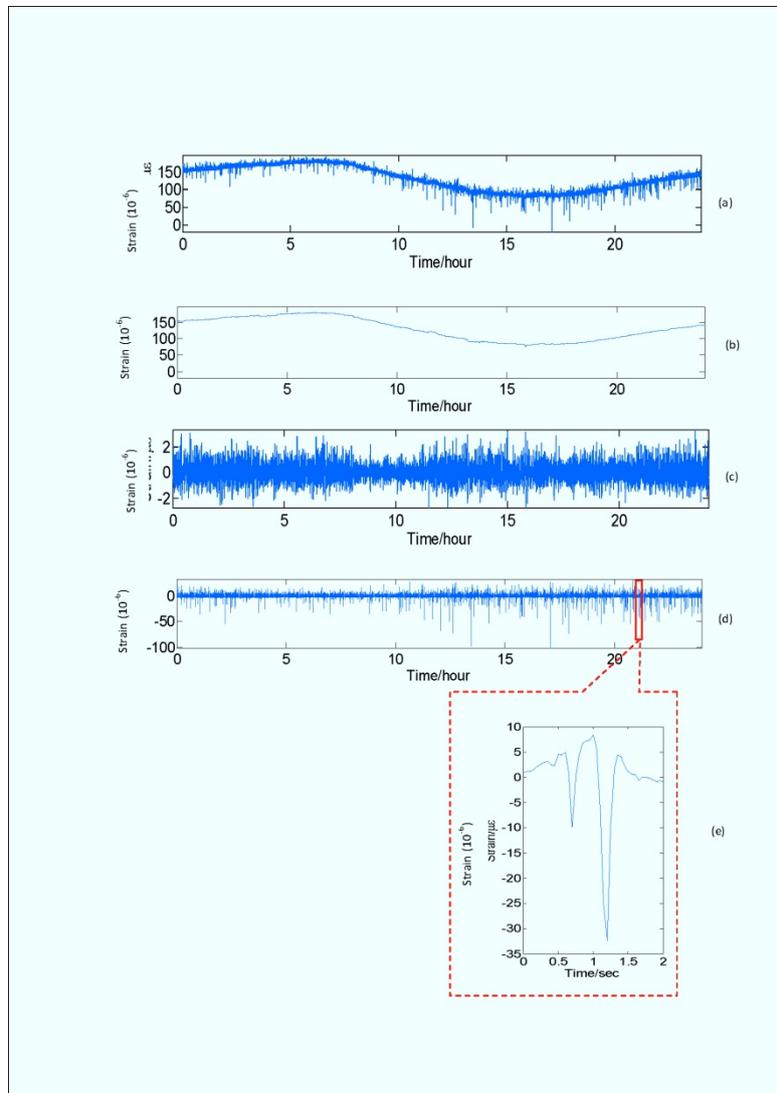


Fig. 4 actually could be treated as noise signal considering its in SSTLN01 (e)  $\hat{\varepsilon}_h$  at scale 3 (2sec)

## 2.2 Analysis in the frequency domain

In order to discriminate the above three types of strains, their characteristics in frequency domain are discussed in this section.

### Type 1:

The 1<sup>st</sup> type of strains resembles to a periodical function with a cycle time of 24 hours. According to the Fourier series theory of periodical functions, the fundamental frequency can be approximated as  $f_b = 1 / (3600 \times 24) = 1.2 \times 10^{-5}$  Hz.

Consequently, a sufficiently large multiple of the fundamental frequency is chosen as the cut-off frequency of its energy distribution. Considering the temperature varies slowly, it is reasonable to assume that the temperature remains smooth within certain minutes. So the cut-off frequency could be expected qualitatively as about 1000 times of the fundamental frequency.

$$f_{c,t} = 1 \times 10^{-2} \text{ Hz} = O(10^{-2}) \quad (1)$$

### Type 2:

As described above, the 2<sup>nd</sup> type of strains is similar to an impulse with the duration of approximately 20 sec. On the basis of Fourier Transform theory, the great majority energy of a  $\delta$ -duration impulse concentrates in a frequency interval named Main Lobe from 0 to Main Lobe Width  $\Delta$ . Table 1 presents several typical impulse types with their Main Lobe Widths (Mitra and Kuo 2006).

It can be inferred that the cutoff frequency  $f_c$  of impulses with duration  $\delta$  should be in the order of magnitude of  $1/\delta$  although their curve shapes may be different. Thus by choosing a representative duration  $\delta_2=20\text{sec}$ , the cutoff frequency of the 2<sup>nd</sup> strain type could be estimated as

$$f_{c,r} = \frac{2}{\delta_2} = 1 \times 10^{-1} \text{ Hz} = O(10^{-1}) \quad (2)$$

### Type 3:

The third type of strains is similar to an oscillating impulse or consecutive impulses with two valleys (peaks) generally lasting less than 1sec. Likewise, by choosing a reasonable value  $\delta_3$  as 0.5sec, the representative cutoff frequency of this type could be calculated below as

$$f_{c,h} = \frac{2}{\delta_3} = 4 \times 10^0 \text{ Hz} = O(10^0) \quad (3)$$

Table 1 Main-Lobe Widths of typical impulses

Types of Impulse	Rectangular	Triangular	Cosine
Main Lobe Width $\Delta$	$1/\delta$	$2/\delta$	$2/\delta$

### 3. Multi-resolution analysis and wavelet shrinkage

Corresponding to the multi-scale characteristic revealed above, wavelet analysis method is applied to deal with the strain data. In this section, we will briefly introduce the theory of multi-resolution analysis (MRA) and wavelet shrinkage.

#### 3.1 Multi-resolution analysis

The multi-resolution theory was developed by Mallat (Mallat 1989). Through defining the multi-resolution structure  $\{V_j\}_{j=-\infty}^{j=+\infty}$  of  $L^2(R)$  space that  $V_j \subset V_{j-1}$  and  $L^2(R) = \bigcup V_j$ , he proved there existed a scaling function  $\varphi(x)$ , the set of the translations and dilations to which,  $\{\varphi_{j,n}(x) = 2^{-j/2}\varphi(2^{-j} - n), n \in Z\}_{j \in Z}$ , is an orthogonal basis of the multi-resolution approximation spaces  $V_j$ . Subsequently, the theory constructs the orthogonal complement of  $V_j$  in  $V_{j-1}$ , named the detail spaces  $W_j$ . Finally the wavelet function  $\psi(x)$  could be constructed that the translations and dilations of which  $\{\psi_{j,n}(x) = 2^{-j/2}\psi(2^{-j} - n), n \in Z\}_{j \in Z}$  form the basis of  $W_j$ .

For a recorded time history of strain  $\varepsilon(t)$  at a finite sampling rate – assume  $\varepsilon(t) \in V_0$ , the coefficients at the level of  $j$  are defined as

$$a_{j,n} = \int_{-\infty}^{+\infty} \varepsilon(t) \varphi_{j,n}(t) dt, d_{j,n} = \int_{-\infty}^{+\infty} \varepsilon(t) \psi_{j,n}(t) dt \quad (4)$$

Define  $a_j(t) = \sum_n a_{j,n} \varphi_{j,n}(t)$  and  $d_j(t) = \sum_n d_{j,n} \psi_{j,n}(t)$  as the  $j$ -th approximation and the  $j$ -th detail respectively, the full expanding of  $\varepsilon(t)$  at the  $J$ -th level can be expressed as

$$\varepsilon(t) = a_0(t) = a_J(t) + \sum_{j=1}^J d_j(t) \quad (5)$$

For practical implementation, Mallat also provided a pyramidal algorithm for computing coefficients in signal decomposition and reconstruction based on convolutions with quadrature mirror filter derived from  $\varphi(x)$ . Giving the coefficients  $a_{j,k}$  of the  $j$ -th approximation  $a_j(t)$ , the decomposition operation is necessary for determining the coefficients of the coarse-resolution approximation  $a_{j+1}$  and the detailed signal  $d_{j+1}$ . The reconstruction should be conducted when computing  $a_j$  from  $a_{j+1}$  and  $d_{j+1}$ .

From the point of digital signal process, the decomposition is equal to a filtering process. After a level of decomposing on  $a_j$ , the separated coarse-resolution approximation  $a_{j+1}$  keeps the information mainly on the lower 1/2 band of  $a_j$ , while the detail  $d_{j+1}$  contains the main information on the higher 1/2 band. The detailed interpretation of MRA theory could be found in articles by Mallat (1989) and Daubechies (1992).

#### 3.2 Wavelet shrinkage

Wavelet Shrinkage was proposed by David and Johnstone (1994), and David (1995) for

problems of recovering an unknown function from noisy data. Sometimes this process was also named wavelet de-noising (Mallat and Hwang 1992) or wavelet filtering.

A straightforward strategy for shrinkage or de-noising is thresholding wavelet coefficients. The general procedure follows the next three steps:

Step 1: (Decomposition) decompose the original signal  $f$  at a selected level  $N$ .

Step 2: (Thresholding) select a threshold and then apply soft or hard thresholding to detail coefficients computed in Step1. The approximation coefficients will not be changed.

Step 3: (Reconstruction) reconstruct signal using the approximation coefficients at the  $N$ -th level and the modified detail coefficients at levels from 1 to  $N$ .

The key issue for wavelet shrinkage is to determine thresholds on which an overview was provided in (Nason, 1995) and (Vidakovic, 2009). Here two methodologies are introduced.

The first is Mallat's model and percentile threshold (S. G. Mallat, 1989). He found that the probability density of image detail coefficients can be experimentally modelled as the family of exponential function:

$$f(d) = Ce^{-(d/\alpha)^\beta} \tag{6}$$

Where the constant  $C$  is normalized as  $C = \frac{\beta}{2\alpha\Gamma(1/\beta)}$ , and the coefficients  $\alpha$  and  $\beta$  can be calculated through estimating the first and second moment of the probability density.

The above model could be used to design the percentile threshold. The induced threshold is

$$T = q_{1-\alpha/2} \tag{7}$$

Where  $q_{1-\alpha/2}$  is the  $(1-\alpha/2)$  quantile of the estimated distribution function or the empirical distribution of the sampling data.

The second is so-called universal threshold by Donoho and Johnstone (1994). The universal threshold is proposed as

$$T = \hat{\sigma}\sqrt{2\log n} \tag{8}$$

Where  $n$  is the number of data points,  $\hat{\sigma}$  is an estimation of the noise level and could be obtained by computing the standard deviation of the detail coefficients (at the  $j$ -th level)

$$\hat{\sigma} = s = \left( \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (d_{j,i} - \bar{d}_j)^2 \right)^{1/2} \tag{9}$$

When SNR (signal-to-noise ratio) is small, the MAD (median absolute deviation from the median) estimation is much robust:

$$\hat{\sigma} = \frac{1}{0.6745} \text{MAD}(d_j) = \frac{1}{0.6745} \text{median}(|d_j - \text{median}(d_j)|) \tag{10}$$

Where median ( $\cdot$ ) is the middle value (50% percentile) of empirical distribution.

In summary, we could see that the percentile threshold much concentrates on the wanted information of which a prior knowledge has usually to some extent been achieved, while the

universal threshold is as much as in the viewpoint of controlling the risk of noise.

#### 4. Wavelet model of strain data

As it has been revealed, online strain of bridges consists of three types of strain, with their own but separable time scales. The subsequent question becomes that how to separate and extract them one by one. In this section, the wavelet model of strain with an emphasis on information extraction is proposed.

##### 4.1 The strain model

As a general form, the strain response under service environment could be deconstructed that

$$\varepsilon(t) = \varepsilon_t(t_d) + \varepsilon_r(t_r) + \varepsilon_h(t_h) + \varepsilon_w(t) + r(t) \quad (11a)$$

Where  $\varepsilon_t$ ,  $\varepsilon_r$ ,  $\varepsilon_h$  and  $\varepsilon_w$  represent the strain components due to temperature variation, railway loads (trains), significant highway loads (heavy trucks) and wind load (especially typhoons), respectively.  $t_d$ ,  $t_r$  and  $t_h$  represent their characteristic time scales. As estimated in Eqs. (1)-(3), they should be in the order of hours, minutes and seconds, respectively.  $r(t)$  refers to the strain caused by random noises or insignificant random loads. Strain signals without any loading were investigated in our study. It was found to usually have a small vibration magnitude and a wide frequency band. Moreover, its empirical distribution is found not to reject the normal distribution at a significant level of 5%. Since strains evolved in this study were not under a significant wind environment, the character of wind-induced strain and its separation will not be included. Thus in this article, Eq. (11(a)) could be simplified as

$$\varepsilon(t) = \varepsilon_t(t_d) + \varepsilon_r(t_r) + \varepsilon_h(t_h) + r(t) \quad (11b)$$

Decompose the strain signal to the  $N_h$ -th,  $N_r$ -th and  $N_t$ -th level in sequence, and then reconstruct the details between consecutive levels, the strain could be divided into four parts,

$$\varepsilon(t) = a_{N_t}(t) + \sum_{j=N_r+1}^{N_t} d_j(t) + \sum_{j=N_h+1}^{N_r} d_j(t) + \sum_{j=1}^{N_h} d_j(t) \quad (12)$$

At an given sampling frequency, we could determine the appropriate numbers of  $N_h$ ,  $N_r$  and  $N_t$  after estimating cut-off frequencies through Eqs. (1)-(3), yielding

$$\begin{aligned} a_{N_t}(t) &\approx \hat{\varepsilon}_t \equiv \varepsilon_t(t_d) + r_3^d(t) \\ \sum_{j=N_r+1}^{N_t} d_j(t) &\approx \hat{\varepsilon}_r \equiv \varepsilon_r(t_r) + r_3^d(t) \\ \sum_{j=N_h+1}^{N_r} d_j(t) &\approx \hat{\varepsilon}_h \equiv \varepsilon_h(t_h) + r_2^d(t) \end{aligned}$$

$$\sum_{j=1}^{N_h} d_j(t) \approx r_1^d(t) \tag{13}$$

Here  $\hat{\varepsilon}_t$ ,  $\hat{\varepsilon}_r$  and  $\hat{\varepsilon}_h$  are actual strain components containing noises corresponding to  $\varepsilon_t$ ,  $\varepsilon_r$  and  $\varepsilon_h$  defined in Eq. (11). Because  $r(t)$  is a wide-band noise, its multi-level components will still be superposed on the actual strain caused by significant loads. Therefore,  $r_3^a(t)$ ,  $r_3^d(t)$  and  $r_2^d(t)$  could be understood as noises on temperature strain, railway strain and highway strain, respectively.

#### 4.2 Level determination

Assuming that the sampling rate of  $\varepsilon(t)$  is  $f_s$ , then the Nyquist frequency will be  $f_m=f_s/2$ . By estimating the three cut-off frequencies through Eqs. (1)-(3), the value of  $N_h$ ,  $N_r$  and  $N_t$  can be determined as follows

$$N_h \approx \log_2\left(\frac{f_m}{f_{c,h}}\right), N_r \approx \log_2\left(\frac{f_m}{f_{c,r}}\right) \text{ and } N_t \approx \log_2\left(\frac{f_m}{f_{c,t}}\right) \tag{14}$$

Where  $f_{c,h}$ ,  $f_{c,r}$  and  $f_{c,t}$  are cut-off frequencies for the strain types induced by highway load, railway load and temperature, respectively.

Take strain data on TMB with  $f_s=25.6$  Hz and  $f_m=12.8$  Hz as an example:

(1)  $N_h$  can be chosen as 1 since  $f_{c,h}$  is the estimated as 4 Hz. For convenience,  $N_h$  is directly set as zero because  $N_h$  is a minor value in this case.

(2) Considering  $f_{c,r}$  is at the order of  $O(10^{-1})$  Hz, it is reasonable to choose  $f_{c,r}$  as 0.2 Hz or 0.4 Hz. Thus the  $N_r$  could be calculated as 5 or 6.

(3) Similarly, let  $f_{c,t}$  be 0.01 Hz, the value of  $N_t$  should be taken as 10 or 11.

Due to the environmental and operational complexity, the cut-off frequencies given in Expression (1)-(3) are more qualitative rather than quantitative. Therefore, the composition level could be slightly adjusted according to the practical situation and specific requirement. For instance, if we care more about  $\varepsilon_r(t)$  and require the separated  $\hat{\varepsilon}_r$  be precious enough, then a relatively large value of  $f_{c,r}$  such as 0.8 Hz and a relatively small  $f_{c,t}$  such as 0.005 Hz can be taken, which leads  $N_r$  to be 4 and  $N_t$  to be 11 even 12.

#### 4.3 Information extraction of different strain types

Information extraction of bridge strain has two levels. The first level could be named as *information separation* that is separating the three types of strain into  $\hat{\varepsilon}_t$ ,  $\hat{\varepsilon}_r$  and  $\hat{\varepsilon}_h$  in Eq. (13). The second level could be called the real *information extraction* which identifies and picks up the local strain cycles or impulses of the 2<sup>nd</sup> and 3<sup>rd</sup> types one by one, that is to say, to extract  $\varepsilon_t(t_d)$ ,  $\varepsilon_r(t_r)$  and  $\varepsilon_h(t_h)$  from the noisy data. In this article, wavelet shrinkage methods are adopted to realize the extraction operation above.

##### 4.3.1 The procedure of information extraction

Based on above analysis, both of the separation and extraction should conduct decomposition and reconstruction operation three times (or twice when  $N_h=0$ ) with one type of strain being separated each time. Distinctively, the extraction has a shrinkage (thresholding) operation on wavelet coefficients before reconstruction. Fig. 5 illustrates the flowchart of information extraction.

We make a summation of all required parameters from the practical level of implementing wavelet shrinkage, they are 1) the levels of decomposition and reconstruction and 2) thresholds for de-noising. The determination of decomposition levels could be got through Eq. (14). Next in Section 4.3.2, we will give two strategies for calculating thresholds expressed in Eqs. (16) and (17), respectively suitable for extracting strain types due to trains and heavy trucks.

#### 4.3.2 Threshold determination

In this section the determination of threshold for wavelet shrinkage is discussed. Based on the prior knowledge of each strain type, this article proposes a specific threshold computing method by referring the essential strategies mentioned in Section 3.2 instead of directly applying them on the practical data. Namely, when the strain caused by train loads is picked up, the percentile threshold is utilized by controlling its occurrence probability. On the other hand, the strategy of controlling noise is applied for extracting the strain caused by heavy trucks. These two methods are simply named as P-method and N-method, respectively.

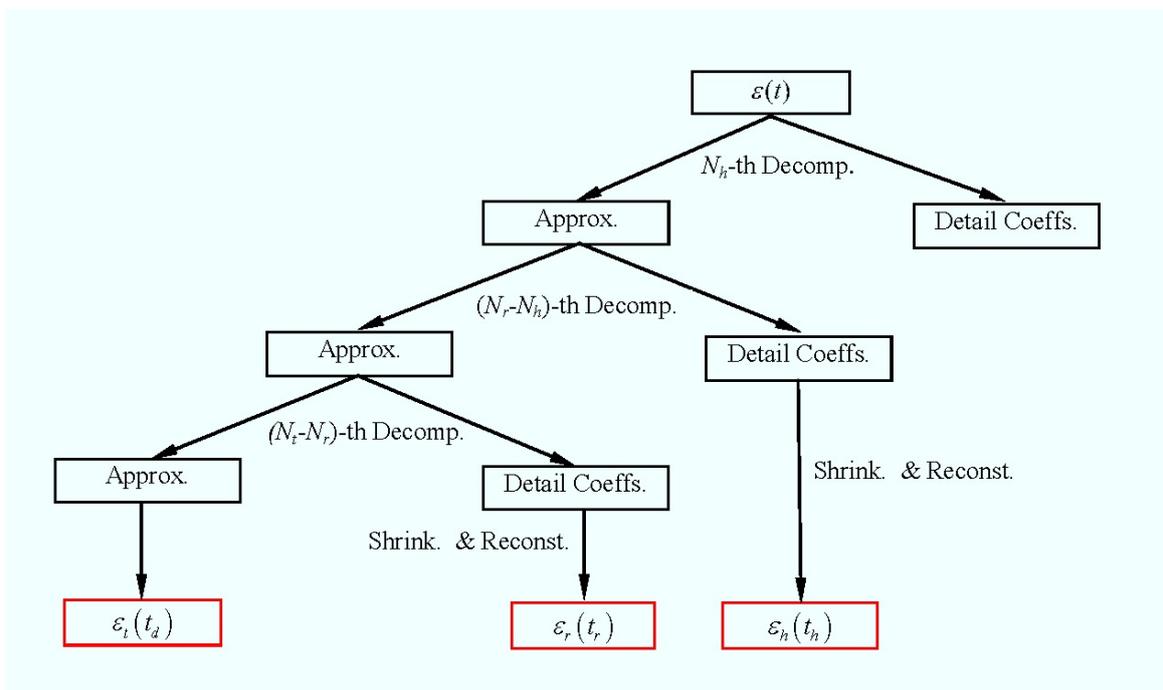


Fig. 5 Flow chart of information extraction

P-method for extraction of the strain due to trains

The investigation shows that the passing train traffic is highly regular in Tsing-Ma Bridge. Firstly, the number of trains crossing the bridge every day is strictly stipulated with little deviation. Secondly, the duration time of passing the bridge is steady because of the limited speed. These characters induce the idea of P-method that is to control the happening of the train strain  $\varepsilon_r(t_r)$ .

The procedure is listed as follows:

1) Through counting the average number of the 2<sup>nd</sup> type in a unit of a single day, assuming as  $N_{train}$ , the expected occurrence rate could be calculated as

$$f_{train} = N_{train} / \text{day} = 1.1574 \times 10^{-5} N_{train} \text{ Hz}$$

2) The occurring probability of the 2<sup>nd</sup> type could be obtained as

$$p_{train} = f_{train} \times \Delta t_{train} \tag{15}$$

Here  $\Delta t_{train}$  is the average duration of the 2<sup>nd</sup> strain types.

3) The threshold of detail coefficients at level  $j$  could thus be determined at the percentile of  $p_{train}$ .

$$T_j = (|d_j|)_{1-p_{train}} \tag{16}$$

Where  $(\cdot)_{1-p}$  denotes the 1- $p$  percentile of the data vector.

Investigations on TMB show that  $N_{train}$  is about 460 times which happen during about 19 hours every day since trains are not scheduled to pass the bridge between 1:00 and 6:00 A.M. The probability  $p_{train}$  could thus be achieved as 10.6% by estimating  $\Delta t_{train}$  as 20s.

N-method for extraction of the strain due to heavy trucks

The strain responding to the truck load has a different character; its occurring rate and duration time have a relatively high variability. Therefore, it is more practical to control the noise rather than its happening probability.

According to Eq. (8) suggested by Donoho, the threshold at level  $j$  could be obtained by  $T_j = \sqrt{2 \log n_j} \hat{\sigma}_j$ , Here, the length of strain collected by health monitoring system is enormous, taking the sampling rate 20 Hz for example, the size of data point from a single strain gauge could nearly reaches up to 2 million after 24 hours. Although the data size of the detail coefficients will reduce to nearly a half after every level of decomposition according to the Mallat’s algorithm, the length could still be about tens of thousands after 6 levels decomposition. So the value of  $\sqrt{2 \log n_j}$  seems too strict for recovering the actual strain response caused by trucks. Thus, we preserve the form of the above equation but make an adjustment of the multiplier factor  $\beta_j$  on the estimated noise level  $\hat{\sigma}_j$  that is

$$T_j = \beta_j \hat{\sigma}_j \tag{17}$$

Assume the noise in strain data follows the norm distribution because the empirical distribution doesn’t reject the norm distribution at a significance level of 5%. Therefore, the cumulative

function for standard normal distribution  $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$  could be referred to determine the threshold value. It can be found that, when  $x$  is larger than 3, the occurrence probability on the two tails  $2(1-\Phi)$  will descend to a magnitude order of  $10^{-3}$ , a reasonable value to suppress noise. Therefore, the value of  $\beta$  is chosen as 3.0 in this paper.

For wavelet selection, Daubechies wavelets are used in our study. Daubechies wavelets are a family of orthogonal and compact supported wavelets. There is balance in choosing a specific Daubechies wavelet. A larger support length will lead to a larger vanishing moment, which makes the frequency separation of the signal more effective. On another side, a larger support length may result in a bad time-domain resolution, that is to say, may not be effective for strain type extraction. It should be noted that although the above consideration is important, our practice shows that by choosing DB2 to DB10, results only have slight differences. We suggest DB8 (the number 8 denoting the support length) for strain separation, and DB4 for information extraction. These two are also the wavelets we used in the study.

## 5. Results and discussion

Results of strain separation are shown in Figs. 3 and 4. In fact, they have been explained in detail when revealing the multi-scale characteristic of bridge strain. We could find that the wavelet method could successfully separate these different strain types and thus makes it possible to conduct a concentrated research on a specific kind of strain type.

### 5.1 Extraction of train-induced and heavy-track-induced strains

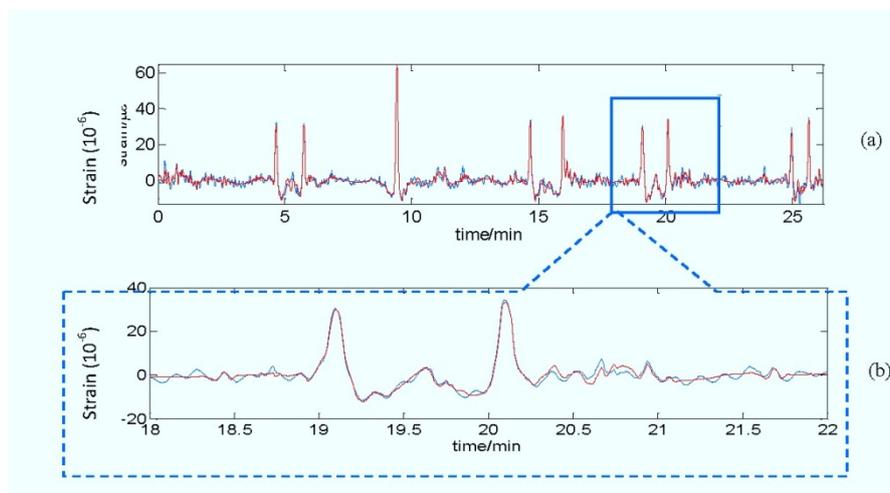


Fig. 6 Extraction of  $\varepsilon_r(t_r)$  (red) from  $\hat{\varepsilon}_r(t_r)$  (blue)

As revealed above, strain data from SSSLN01 includes both train-induced strains and heavy-truck induced strains, while the strain data from YBH4-13 on Run-Yang Bridge only contains the heavy-truck induced strains. Moreover, it was found that the YBH4-13 extraction result on the truck-induced strain is very similar to that of SSSLN01. With these in mind, SSSLN01 is taken for an example to display the results of information extraction. In Fig. 6, the extracted train-induced strain  $\varepsilon_r(t_r)$  (in red) and the strain before extraction  $\hat{\varepsilon}_r(t_r)$  (in blue) are plotted together for comparison. It is shown that noise has been reduced to a lower level, while on another hand, strain curves and peak values induced by trains could be exactly recovered.

Fig. 7 shows the result of heavy-truck-induced strain extraction. It could be seen in Fig. 7(a) that after extraction  $\varepsilon_h(t_h)$  will become zero for most of the time period. In contrast, when a heavy truck passes through, the induced strain cycle still could be exactly extracted as show in Fig. 7(b) --strain curves and peak values both agree well with the strain before extraction.

To verify the effectiveness of information extraction, we define an index named average peak error (APE)  $e_p$  to quantitatively test the global accuracy of exacted strain types. This is because peak values are usually of importance for bridge health monitoring.

$$e_p = \frac{\sum_{j=1}^N |p_j - \hat{p}_j|}{N} \tag{18}$$

Where  $N$  means the counting number of strain Type 2 or Type 3 and  $p_j$  are their peak (valley) values. For a kind of strain type extraction, a smaller APE value implies a more accurate recovery.

As listed in Table 2, it could be seen that APE values, for both extracted strains caused by trains (Type 2) and heavy trucks (Type 3), are very small by considering that the accuracy of strain gauges is usually of  $1 \mu\varepsilon$ .

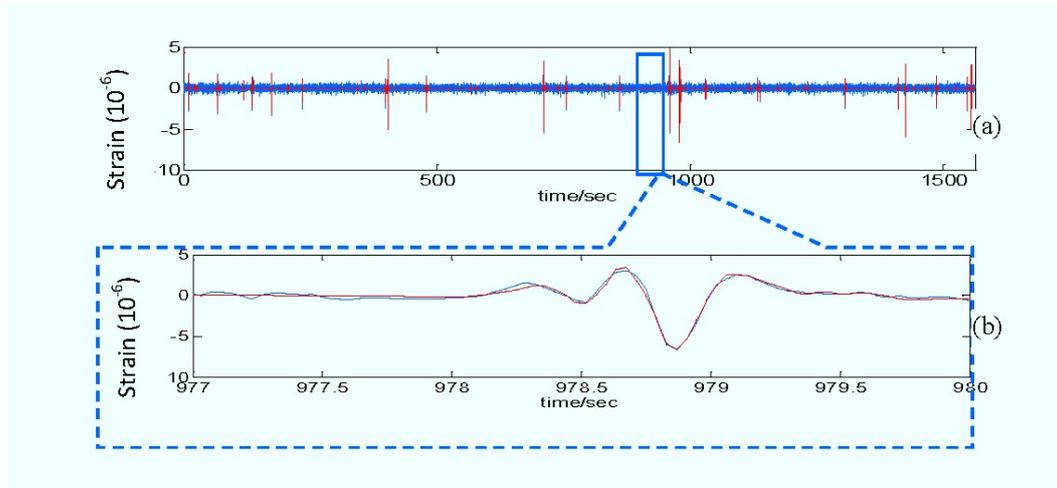


Fig. 7 Extraction of  $\varepsilon_h(t_h)$  (red) from  $\hat{\varepsilon}_h(t_h)$  (blue)

Compression ratio and energy ratio are used to investigate extraction results. Compression ratio is defined in the way that

$$\rho_c = \frac{\text{MemSize}(\varepsilon)}{\text{MemSize}(\hat{\varepsilon})} \quad (19)$$

where MemSize (·) represents the required size for digital storage. After wavelet shrinkage it is roughly equal to the number of nonzero coefficients.

Energy ratio is defined as

$$\rho_e = \frac{\|\varepsilon\|_2^2}{\|\hat{\varepsilon}\|_2^2}, \quad (20)$$

where  $\|\cdot\|_2$  denotes the  $L^2$  norm of a signal, thus  $\|\cdot\|_2^2$  is the energy.

Values of above two indexes are also shown in Table 2. It is found that noise in  $\hat{\varepsilon}_h$  is relatively higher than in  $\hat{\varepsilon}_r$ . This is because the variation range of strain cycles due to trains and that due to the heavy trucks are different—usually  $10^1 \mu\varepsilon$  versus  $10^0 \mu\varepsilon$ , while noise are usually universal at every level.

### 5.2 The capability of data compression

As we known, the data size collected by health monitoring system would be extremely huge during its persistently cumulative process. Therefore, investigations on the extent to which the data could be compressed without compromising its useful information will be necessary for both system designers and users.

Assuming the length of original data is  $l$ , after information separation,  $\hat{\varepsilon}_h$ ,  $\hat{\varepsilon}_r$  and  $\hat{\varepsilon}_t$  will be stored in the form of wavelet coefficients with the length of  $\frac{l}{2^{N_h}} \left(1 - \frac{1}{2^{N_r - N_h}}\right) \approx \frac{l}{2^{N_h}}$ ,  $\frac{l}{2^{N_r}} \left(1 - \frac{1}{2^{N_t - N_r}}\right) \approx \frac{l}{2^{N_r}}$  and  $\frac{l}{2^{N_t}}$ , respectively. Furthermore, after extraction, the memory size for storing  $\varepsilon_h$  and  $\varepsilon_r$  will be about  $\frac{\rho_{c,h}}{2^{N_h}} l$  and  $\frac{\rho_{c,r}}{2^{N_r}} l$ . The total ratio for compression could be estimated as

$$\begin{aligned} \rho_c &\approx \frac{1}{l} \left[ \frac{\rho_{c,h}}{2^{N_h}} l + \frac{\rho_{c,r}}{2^{N_r}} l + \frac{l}{2^{N_t}} \right] \\ &= \frac{\rho_{c,h}}{2^{N_h}} + \frac{\rho_{c,r}}{2^{N_r}} + \frac{1}{2^{N_t}} \end{aligned} \quad (21)$$

As representative values, we choose  $N_h=0$ ,  $N_r=4$ ,  $N_t=12$ ,  $\rho_{c,r} = p_{\text{train}} = 10\%$  and  $\rho_{c,h} = 0.5\%$ , which lead the total ratio  $\rho_c$  to be 1.4%. This is a considerable value and seems to make information extraction very attractive for bridge health monitoring systems.

Table 2 Comparisons on average peak error, compression ratio and energy ratio

	$e_p/10^{-6}$	$\rho_c$	$\rho_e$
$\varepsilon_r / \hat{\varepsilon}_r$	0.40	10.7% (= $p_{\text{train}}$ )	96.3%
$\varepsilon_h / \hat{\varepsilon}_h$	0.56	0.45%	39.1%

## 6. Conclusions

This article reveals the multi-scale features of online strains recorded by health monitoring systems of long-span bridges. It is shown that despite of the diversity of bridge types and the complexity of their respective working environments, the actually measured strain could still be understood and deconstructed in a common model. The strain within three typical temporal scales of  $10^5$ ,  $10^2$  and  $10^0$  sec have different characteristics, which have their individual cut-off frequencies of different magnitude orders of  $10^{-2}$ ,  $10^{-1}$  and  $10^0$  Hz, and are caused by temperature change, trains and heavy trucks, respectively.

Multi-resolution analysis and wavelet shrinkage methods can successfully separate and extract the different strain types, which make the research on a specific strain type more convenient. To make things better, the extraction process compresses the collected data at the same time to an attractive ratio.

The methodology developed during the information extraction, especially the choice of compression level and the determination of threshold, is applicable to the actual digital strain processing for long-span bridges, as the load conditions are similar.

The findings and the developed method for the monitoring data analyses will help bridge engineers and managers to obtain a reliable and clear description of the bridge response under different type of loading, therefore have better understanding to their bridge's behavior from a multi-scale viewpoint and subsequently can evaluate the different influences on the structure induced by different load types.

As we known, the actual environment of civil structure is quite complicated. For example, besides temperature variation, railway loads and heavy trucks, the wind loads, especially typhoons (Chen and Wu 2008, Li *et al.* 2002, Xu *et al.* 2007), may also induce significant strain responses. It should be pointed out that the strain history in this article is the response under service environment; therefore, the features revealed here should be an elementary model. Strains caused by occasional environmental events which make the strain more sophisticated will have their own characters and possibly could also be identified and separated by wavelet method. Further investigations could be conducted on these subjects.

## Acknowledgements

The work described in this article was supported by the grants from the National Natural Science Foundation of China (No. 11072060 and 51008069) and Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions, which are gratefully

acknowledged. The authors also wish to thank the Highways Department of the Hong Kong SAR Government for their support throughout the study.

## References

- Bukkapatnam, S.T.S., Nichols, J.M., Seaver, M., Trickey, S.T. and Hunter, M.. (2005), “A wavelet -based, distortion energy approach to structural health monitoring”, *Structural Health Monit.*, **4**(3), 247-258.
- Cardini, A.J. and DeWolf, J.T. (2009), “Long-term structural health monitoring of a multi-girder steel composite bridge using strain data”, *Struct. Health Monit.*, **8**(1), 47-58.
- Chan, H.T., Guo, L. and Li, Z.X. (2003), “Finite element modelling for fatigue stress analysis of large suspension bridges”, *J. Sound Vib.*, **261**(3), 443-464.
- Chan, H.T., Li, Z.X. and Ko, J.M. (2001), “Fatigue analysis and life prediction of bridges with structural health monitoring data—Part II: application”, *Int. J. Fatigue*, **23**(1), 55-64.
- Chen, S.R. and Wu, J. (2008), “Performance enhancement of bridge infrastructure systems: Long-span bridge, moving trucks and wind with tuned mass dampers”, *Eng. Struct.*, **30**(11), 3316-3324.
- Daubechies, I. (1992), *Ten lectures on wavelets* (Vol. 61): SIAM.
- Donoho, D.L. and Johnstone, J.M. (1994), “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika*, **81**(3), 425-455.
- Donoho, D.L.. (1995). “De-noising by soft-thresholding”. *Information Theory, IEEE Transactions on*, **41**(3), 613-627.
- Farrar, C.R., Sohn, H., Fugate, M.L. and Czarnecki, J.J. (2001), *Integrated structural health monitoring*, Paper presented at the 6th Annual International Symposium on NDE for Health Monitoring and Diagnostics.
- Housner, G.W., Bergman, L.A., Caughey, T.K., Chassiakos, A.G., Claus, R.O., Masri, S.F. and Yao, T.P. (1997), “Structural control: past, present, and future”, *J. Eng. Mech. - ASCE*, **123**(9), 897-971.
- Hu, X.F. and Shenton H.W. (2007), “Dead load based damage identification method for long-term structural health monitoring”, *J. Intelligent Mat. Syst. Str.*, **18**(9), 923-938.
- Kamath, G.M., Sundaram, R., Gupta, N. and Rao, M.S. (2010), “Damage studies in composite structures for structural health monitoring using strain sensors”, *Struct. Health Monit.*, **9**(6), 497-512.
- Katsikeros, C.E. and Labeas, G.N. (2009), “Development and validation of a strain-based structural health monitoring system”, *Mech. Syst. Signal Pr.*, **23**(2), 372-383.
- Kesavan, A., John, S. and Herszberg, I. (2008), “Strain-based structural health monitoring of complex composite structures”, *Struct. Health Monit.*, **7**(3), 203-213.
- Lau, C.K. and Wong, KY. (1997), “Design, construction and monitoring of the three key cable-supported bridges in Hong Kong”, *Proceedings of the 4th International Kerensky Conference on Structures in the new millennium*, Hong Kong.
- Li, S.Z. and Wu, Z.S. (2008), “Modal analysis on macro-strain measurements from distributed long-gage fiber optic sensors”, *J. Intel. Mat. Syst. Str.*, **19**(8), 937-946.
- Li, Y. H., Tang, L.Q., Liu, Z.J. and Liu, Y.P. (2012), “Statistics and probability analysis of vehicle overloads on a rigid frame bridge from long-term monitored strains”, *Smart Struct. Syst.*, **9**(3), 287-301.
- Li, Z.X., Chan, H.T. and Ko, J.M. (2001), “Fatigue analysis and life prediction of bridges with structural health monitoring data—Part I: methodology and strategy”, *Int. J. Fatigue*, **23**(1), 45-53.
- Li, Z.X., Chan H.T. and Ko, J.M. (2002), “Evaluation of typhoon induced fatigue damage for Tsing Ma Bridge”, *Eng. Struct.*, **24**(8), 1035-1047.
- Liu, M., Frangopol, D.M. and Kim, S. (2009), “Bridge system performance assessment from structural health monitoring: a case study”, *J. Struct. Eng. - ASCE*, **135**(6), 733-742.
- Mallat, S.G. (1989), “A theory for multiresolution signal decomposition: the wavelet representation”, *IEEE T. Pattern Anal.*, **11**(7), 674-693.
- Mallat, S. and Hwang, W.L. (1992), “Singularity detection and processing with wavelets”, *IEEE T. Inform.*

- Theory*, **38**(2), 617-643.
- Mitra, S.K. and Kuo, Y.H. (2006), *Digital signal processing: a computer-based approach* (Vol. 2), McGraw-Hill New York.
- Nason, G.P. (1995), "Choice of the threshold parameter in wavelet function estimation", *Wavelets Statistics*, 261-280.
- Omenzetter, P., Brownjohn, J.M.W. and Moyo, P. (2004), "Identification of unusual events in multi-channel bridge monitoring data", *Mech. Syst. Signal Pr.*, **18**(2), 409-430.
- Reynders, E., Roeck, G.D., Gunders B.P. and Sauvage, C. (2007), "Damage identification on the Tilff Bridge by vibration monitoring using optical fiber strain sensors", *J. Eng. Mech. - ASCE*, **133**(2), 185-193.
- River, The Governmental Headquarter of Bridge Construction on Yangtzi (2004), *The Design of structural health monitoring system on Run-Yang Yangtzi Bridge (in Chinese)*, Jiangsu Province, China.
- Van der Auweraer, H. and Peeters, B. (2003), "International research projects on structural health monitoring: an overview", *Struct. Health Monit.*, **2**(4), 341-358.
- Vidakovic, B. (2009), *Statistical modeling by wavelets*, Wiley-interscience.
- Wong, C.L., Childs, P.A. and Peng, G.D. (2006), "Multiplexed fibre Fizeau interferometer and fibre Bragg grating sensor system for simultaneous measurement of quasi-static strain and temperature using discrete wavelet transform", *Measurement Sci. Technol.*, **17**(2), doi:10.1088/0957-0233/17/2/021.
- Xu, Y.L., Guo, W.W., Chen, J., Shum, K.M. and Xia, H. (2007), "Dynamic response of suspension bridge to typhoon and trains. I: Field measurement results", *J. Struct. Eng. - ASCE*, **133**(1), 3-11.
- Ye, X.W., Ni, Y.Q., Wong, K.Y. and Ko, J.M. (2012), "Statistical analysis of stress spectra for fatigue life assessment of steel bridges with structural health monitoring data", *Eng. Struct.*, **45**, 166-176.