

# A wavelet-based despiking algorithm for large data of structural health monitoring

Yun-Xia Xia<sup>1</sup> and Yi-Qing Ni<sup>2</sup>

## Abstract

The last two decades have witnessed a rapid increase in the applications of long-term structural health monitoring technologies to the civil structures. A wealth of field data has been collected by the structural health monitoring systems. Nevertheless, the mining of information associated with structural condition from the large database is still a great challenge. In the structural health monitoring signals, spikes are commonly encountered anomalies that have large amplitudes and may show complex spatial and temporal patterns. They can introduce significant errors in data-based condition assessment of structures. Particularly, the long-term structural health monitoring data have an extraordinarily large volume. To remove the spikes in it, the algorithm is highly desired to be both automatic and efficient. An unsupervised and fast despiking method is proposed in this article on the theoretical cornerstone of the wavelet transform. This method is implemented by two steps, namely, spike detection and spike removal. The hypothesis testing and algorithm of searching wavelet modulus maxima chain are incorporated into the spike-detection procedure. Thus, the arrival time of the spikes can be identified fast. And then, the spikes are removed by a cross-scale maxima and minima search algorithm based on the maximum overlap discrete wavelet transform, retaining the unaffected information. The inverse transformation is not required in the spike-detection step, which improves the speed of the algorithm. The spike-removal algorithm removes spikes only from their occurrence frequency bands; thus, the unaffected signal components are intact after despiking. The proposed algorithm is demonstrated using three sets of structural health monitoring data collected from a real bridge, comparing with three other approaches, that is, the time-domain method, frequency filter and traditional wavelet method.

## Keywords

Signal processing, spike detection, spike removal, wavelet transform, structural health monitoring

Date received: 19 April 2018; accepted: 6 November 2018

Handling Editor: Hongyan Ma

## Introduction

A rapid development of structural health monitoring (SHM) technology in civil engineering has been witnessed in the recent two decades. Permanent SHM systems have been implemented on some major civil structures. The SHM systems continuously collect data of loads, environment and responses from the structures. Thus, a wealth of long-term SHM data has been obtained. Though these data are expected to be interpreted into information about the structural condition,

it is still a great challenge at present. Even though the

<sup>1</sup>School of Civil Engineering, Qingdao University of Technology, Qingdao, China

<sup>2</sup>Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

### Corresponding author:

Yun-Xia Xia, School of Civil Engineering, Qingdao University of Technology, 11 Fushun Road, Qingdao 266033, Shandong, China.  
Email: xiayunxia@qut.edu.cn



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<http://www.creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

manufacturers and industries have attempted to improve the long-term stability of the SHM systems, it is unavoidable that there are occasional errors in the sensors, signal conditioning unit, analogue-to-digital converter (ADC), or digital communication network. Consequently, the signals are usually contaminated by noises, spikes and trends, which have hindered us from tapping the potential of the long-term SHM data. The spikes that have freak amplitudes are often non-physical events in the data. They can introduce errors in the data-based damage detection or condition assessment of civil structures.

The spikes in the SHM signals can originate from different sources, such as capacitive or inductive noise in the analogue signal path and communication errors in asynchronous communication protocols.<sup>1</sup> They do not only show average amplitude exceeding a threshold but also have unique durations and shapes. Thus, the spikes may have highly variable shapes and amplitudes. The variations make the identification and elimination of spikes challenging, and the template-based despiking methods are inapplicable sometimes. In particular, the long-term SHM systems collect an extremely huge volume of data to process, and manual and supervised despiking among plentiful channels is highly time-consuming and even impossible to perform. Therefore, robust spike-removal methods that are unsupervised and fast are particularly interested.

The main difficulty in removing spikes from large SHM data is the fast identification of spikes. There have been dozens of spike-detection methods, which can be classified into manual and automated, supervised and unsupervised ones.<sup>2</sup> Among the manual methods, the most common one may be the window discriminator. Signals that exceed an amplitude threshold and pass through a subsequent pair of specific boxes are identified as spikes.<sup>3</sup> However, this method is time-consuming when processing multiple signals due to its supervised and manual nature. Moreover, the fixed window may leave out prolonged, wide, or step-like spikes. Another widely used method is the amplitude thresholding, but this method degrades rapidly when the signal-to-noise ratios are relatively low. Other methods such as the matched filtering<sup>4</sup> and principle components<sup>5</sup> require prior modelling of 'typical' spike, which make these methods supervised.

After a spike is detected, it is often removed by being suppressed to zero or the level of local mean or median or replaced by a fitting-based interpolation. This method is abrupt and the real signal sources may be removed along with the spikes. For those spikes with frequencies sufficiently high for separation, the spike removal can also be achieved by low-pass filters. Nevertheless, the low-pass filters would fail if frequencies of the spikes overlap those of the underlying signals. It has also been proved that even when spike

frequencies are far higher, the low-pass filter may still not be entirely sufficient to isolate the wanted signal because filters may produce ripples at sharp edges.<sup>6</sup>

The wavelet transform (WT)<sup>7,8</sup> is a powerful tool to analyse the complex time series. It provides the time-frequency and multi-resolution information of signals, so it is effective at detecting and removing transient phenomena like spikes. In the SHM of civil structures, the application of the WT has mainly focused on the damage detection by checking anomalies in wavelet coefficients of the vibration signals.<sup>9–15</sup> Some scholars also use the WT in the signal processing such as data reconstruction.<sup>16</sup> However, the application of the WT in processing complex and large SHM data is still in its infancy.

A wavelet method was proposed based on the discrete Haar transform.<sup>17</sup> However, this method used inverse WT (IWT) that may be unnecessarily time-consuming. Other approaches to spike detection using dyadic wavelets.<sup>18,19</sup> These methods mainly detect a single spike in the signal, but cannot estimate parameters such as arrival time of spikes. To remove spikes, the wavelet methods do not simply substitute the spikes with zeros, mean values or interpolations, but remove them in the wavelet domain only in their occurrence frequencies, retaining information of unaffected frequencies. Besides, they are able to robustly eliminate both high- and low-frequency spikes owing to the merit of both time-frequency and multi-resolution. The traditional wavelet-based despiking method removes the spikes by cancelling all wavelet coefficients that exceed a threshold. However, it can yield errors, and this approach significantly underestimates the initial amplitude in the case of spike series.<sup>20</sup>

Fortunately, combining with other techniques, the weaknesses in using the WT alone are expected to be overcome. On the theoretical cornerstone of WT, an unsupervised and fast despiking algorithm is proposed in this article for the long-term SHM data collected from civil structures. Techniques of hypothesis-testing-based detection, parameter estimation and modulus maxima chain are incorporated into this algorithm. It estimates the occurrence time of the spikes fast and automatically and then removes these spikes focusing on their neighbouring time-domains in different wavelet scales, retaining unaffected information at the instant of the spikes. The spike detection is conducted at the level of wavelet coefficients directly in the hypothesis testing framework, and there is no need for an inverse transformation. This makes the spike detection fast and ideal for long-term SHM data. The despiking does not simply interpolate values at the arrival time of spikes, but removes spikes only from their occurrence frequency bands. The information from other frequencies at the spike instant can be retained.

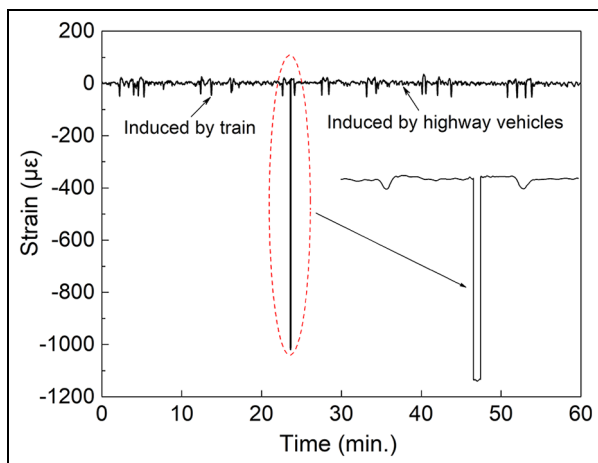
The proposed algorithm is demonstrated by long-term data obtained from the Tsing Ma Bridge (TMB)

in Hong Kong. Three sets of data are used to examine this algorithm: one set with a single spike, another with two adjacent spikes and the last one with three spikes with different shapes and durations. The results are compared with those of a time-domain method, frequency filter and traditional wavelet method. The remainder of this article is organized as follows. The motivation and particularity of the studied problem are illustrated first. The wavelet-based despiking algorithm is proposed subsequently. And then, the proposed algorithm is demonstrated by three sets of real data. Finally, conclusions are given.

## Problem statement

Spikes are those signal components whose amplitudes are significantly larger than the immediate surrounding data. They may be produced by anomaly of any link in the signal path, such as the capacitive and inductive noises and communication errors. Figure 1 shows an example of the spikes embedded in the strain signals collected from the TMB. This signal reflects the structural response of the bridge deck under service loads that mainly include the highway and railway traffic. The components induced by the railway and highway traffic, respectively, have also been indicated in Figure 1. The magnitude of the spike is so large that it would significantly affect the conclusions of the data-based structural analysis.

The spike shown in Figure 1 is easy to handle because it is a single spike with salient amplitude. However, the spikes in the long-term SHM data are various and complex. For example, Figure 2 shows displacement of the bridge deck measured in a full day: Figure 2(a) shows two spikes adjacent to each other, and Figure 2(b) shows three spikes with different



**Figure 1.** A single spike embedded in the strain signal collected from the TMB.

shapes and durations. Due to the diversity of the spikes' durations, the window discriminator method is ineffectual because the fixed window may leave out some spikes, and the frequency filters would also fail. In addition, if the spikes are removed by being replaced with zeros or local means or medians, the detailed information as shown in Figure 2(b) (those on the peak edge of the zoomed spike) would be lost. Therefore, to remove the spikes embedded in SHM signals, a multi-resolution and robust method is required, for which the WT is a promising tool.

The motivation of this study arose from the authors' experience in evaluating the bridge condition based on long-term SHM data. The statistics of the daily maximum structural responses monitored by the SHM system was conducted to obtain the extreme value distributions and then to extrapolate the long-term return levels. However, the spikes embedded in the data have hindered this study. The data volume to tackle is incredibly tremendous because the SHM system is collecting data continuously with a high sampling rate. For example, the sampling rate for the strain monitoring of the TMB is 51.2 Hz, which implies that 4,423,680 data points would be obtained by each strain gauge in a single day. The data to handle were measured by hundreds of sensors in more than 10 years. To process the massive data, a manual technique is nearly impossible. On the contrary, an automatic and efficient algorithm is desired.

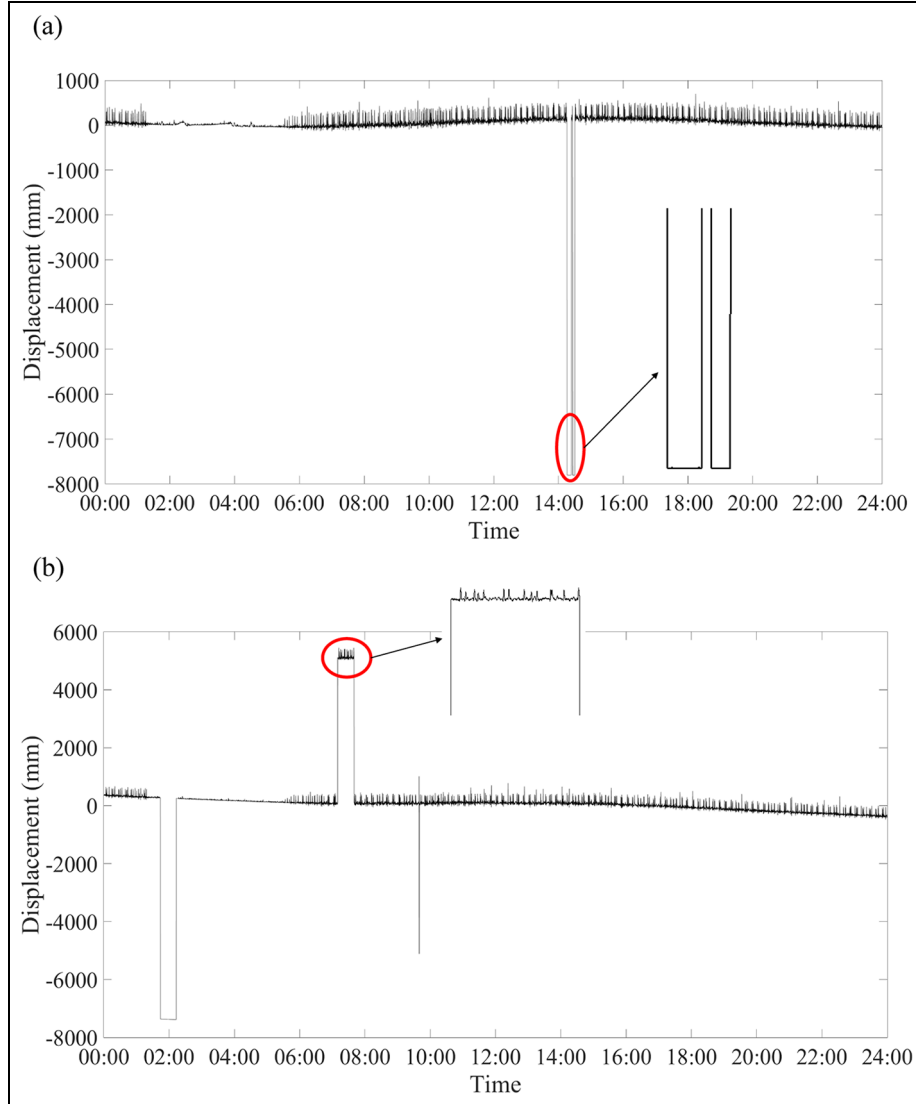
## Wavelet-based despiking algorithm

The wavelet-based despiking algorithm for large SHM data involves two successive steps: (1) spike detection (time estimation included) and (2) spike removal. Techniques of statistics, hypothesis testing and wavelet modulus maxima chain are incorporated into this algorithm to complement the WT.

### Spike detection

A spike-detection method<sup>2</sup> developed to identify spikes in neural recordings is referred to by modifying some parameters and procedures according to the characteristics of long-term SHM data. The procedure is stated briefly as follows. The multi-scale decomposition of the signal is performed first. And then, the Bayesian hypothesis testing is conducted at each scale to detect the presence of spikes. Finally, the arrival time of individual spikes is estimated by combining the decisions of different scales.

The choice of a suitable wavelet function is important for spike detection. In principle, wavelet coefficients measure the similarity between the signal and the wavelet basis. Thus, to select a spike-like mother wavelet is reasonable. Consequently, the signal will have a



**Figure 2.** Variety of spikes: (a) two spikes adjacent to each other and (b) three spikes with different shapes and durations.

sparse representation. To reduce the computational cost, a limit set of scales

$$S = \{s_0, s_1, \dots, s_j, \dots, s_J\} \quad (1)$$

are used, where  $s_0$  and  $s_J$  depend on the sampling rate of the signal and the minimum and maximum durations of the spikes. The intermediate scales  $\{s_0, s_1, \dots, s_{J-1}\}$  are uniformly sampled between  $s_0$  and  $s_J$  with an arbitrary step.

Wavelet coefficients in multi-scales are obtained by the WT to represent the signal. The coefficients corresponding to the ‘spike plus signal’ are random variables with non-zero means. For unsupervised spike detection, the wavelet coefficients should be separated by estimating the noise level  $\sigma$ . The ideas from Donoho and Johnstone<sup>21</sup> are borrowed to obtain estimates of the noise level. In their method, only those wavelet

coefficients that exceed a threshold are accepted, and an IWT is performed subsequently. The thresholding policy at different scales follows the hard thresholding rule

$$\rho_T(x) = \begin{cases} X, & \text{if } |X| > T \\ 0, & \text{if } |X| \leq T \end{cases} \quad (2)$$

The threshold  $T$  is usually chosen as

$$T = \sigma \sqrt{2 \log_e N} \quad (3)$$

where  $N$  is the number of samples and  $\sigma$  is the variance of the noise coefficients. The variance  $\sigma$  must be estimated from the wavelet coefficients because the noise coefficients are unknown. Assuming that the wavelet coefficients are Gaussian random variables independent on each other, the variance of the noise coefficients at scale  $s_j$  is

$$\hat{\sigma}_j = M\{|W(j, 0) - \bar{w}_j|, \dots, |W(j, N-1) - \bar{w}_j|\} / 0.6745 \quad (4)$$

where  $\bar{w}_j$  is the sample mean of  $w_j$  and  $M\{\cdot\}$  means the sample median. However, the wavelet coefficients considerably different from noise coefficients will contain the concerned signal, that is, such an estimation will be biased. In this case, the threshold will depend on the wavelet basis.<sup>8</sup>

In this study, the threshold is estimated by hypothesis testing on the wavelet coefficients. Unlike the method of Donoho and Johnstone,<sup>21</sup> the IWT is not necessary and the computation efficiency can be improved significantly as a result. The spike detection is conducted in the time-scale domain through sequential binary hypothesis tests on the wavelet coefficients. The spike is not present under the hypothesis  $H_0$ , while both spike and signal are present under  $H_1$ . The hypothesis testing rule for each wavelet coefficient  $W(j, k)$  is<sup>2</sup>

$$|W(j, k)| \underset{H_1}{\overset{H_0}{<}} \frac{\hat{\mu}_j}{2} + \frac{\hat{\sigma}_j^2}{\hat{\mu}_j} \log_e \gamma_j \triangleq \Theta_j \quad \forall k \in B \quad (5)$$

where  $\hat{\sigma}_j$  is determined by equation (4),  $\hat{\mu}_j$  is the sample mean of the absolute value of the wavelet coefficients at scale  $s_j$  under hypothesis  $H_1$ , and  $\gamma_j$  is a parameter determined by the acceptable costs of false alarms and omissions, denoted by  $\lambda_{FA}$  and  $\lambda_{OM}$  and the prior probabilities of the two hypotheses. The parameter  $\Theta_j$  is an acceptance threshold for the hypothesis  $H_1$  at scale  $s_j$ . More details about the rules of the hypothesis testing can be found in Appendix II of Nenadic and Burdick's paper.<sup>2</sup>

Because most of the wavelet coefficients do not contain information of spikes,  $\hat{\mu}_j$  cannot be estimated from  $w_j$ . Nevertheless, a rough estimate can be obtained by the thresholding rule in equation (2). It splits  $w_j$  into two disjoint subsets: a signal subset  $w_j^{sg}$  and a spike subset  $w_j^{sp}$ , where<sup>2</sup>

$$w_j^{sg} \triangleq \{W(j, k) \in w_j : |W(j, k)| \leq T_j\}$$

and

$$w_j^{sp} \triangleq \frac{w_j}{w_j^{sg}}$$

Therefore,  $\hat{\mu}_j$  is the sample mean of  $w_j^{sp}$ .  $\gamma_j$  is determined by<sup>2</sup>

$$\log_e \gamma_j = \log_e \left[ \frac{\lambda_{FA} P(H_0)}{1 - \lambda_{FA} P(H_1)} \right] \triangleq LL_M + \log_e \frac{P(H_0)}{P(H_1)} \quad (6)$$

where  $L$  is in effect a new parametrization of  $\lambda_{FA}/\lambda_{OM}$  and  $L_M = 36.7368$  is a chosen scaling factor.

A spike's wavelet coefficients are neighbours in both time and scale. Combining the sets of acceptance of  $H_1$ , denoted by  $B^{H_1}$

$$B^{H_1} = \bigcup_{s_j \in S} B_j^{H_1} \quad (7)$$

where  $B_j^{H_1}$  is a subset of the translation set  $B$  that corresponds to the acceptance of  $H_1$  at scale  $s_j$ , the continuities across different scales are combined.<sup>2</sup>

To estimate the arrival time of spikes, the acceptance set of  $H_1$  is organized into its contiguous constituents<sup>2</sup>

$$B^{H_1} = \bigcup_{i=1}^{N_c} C_i^{H_1} \quad (8)$$

where  $C_i^{H_1}$  are the contiguous regions of  $B^{H_1}$  and  $N_c$  is the number of contiguous regions. The estimated location of the  $i$ th spike at scale  $s_j$  is

$$T_i^j \triangleq \arg \max_{k \in C_i^{H_1}} \{|W(j, k)| : |W(j, k)| > \Theta_j\} \quad (9)$$

$$\forall i = 1, 2, \dots, N_c \quad \forall s_j \in S$$

And then, the arrival time candidate of the  $i$ th spike is estimated by averaging  $T_i^j$  over the scales which accept  $H_1$  on the  $i$ th contiguous region

$$T_i = \frac{1}{\|S_i^{H_1}\|} \sum_{s_j \in S_i^{H_1}} T_i^j \quad \forall i = 1, 2, \dots, N_c \quad (10)$$

where

$$S_i^{H_1} \triangleq \{s_j \in S : |W(j, k)| > \Theta_j, k \in C_i^{H_1}\}$$

and  $\|\cdot\|$  is the set size. Post-processing of the spike candidates is needed for those contiguous transients.

### Spike removal

To robustly remove the diverse spikes that have a wide range of shapes and durations from the long-term SHM data, a method proposed in the neuroscience to remove spikes in the functional magnetic resonance imaging signals<sup>22</sup> is referred to. For the WT, the maximum overlap discrete wavelet transform (MODWT)<sup>23</sup> is employed. The MODWT has the following advantages over the traditional discrete wavelet transform that uses dyadic scales.<sup>22</sup> First, it is applicable for all sample sizes, so the length of signal to be analysed has no restrictions of 'power of two'. Moreover, transient detection from coarser scales is easier due to the MODWT. And then, the wavelet coefficients at each scale are temporarily aligned based on the phase delay properties of the wavelet basis. For example, for the db4 basis, at each scale, the circular shift is defined by  $T_s$

$$T_s = 2^{s-1}(L-1) - 1 \quad (11)$$

Therefore, all wavelet coefficients  $W_{s,t}$  ( $s$  represents the scale and  $t$  represents time) are redefined as

$$W_{s,t} = W_{s,t-T_s} \quad (12)$$

The spikes are represented as chains of maximal and minimal wavelet coefficients at the same time point in multiple scales. After the temporal alignment, the  $2 \times 2$  neighbourhood of each coefficient is searched for maximal or minimal wavelet coefficients.<sup>22</sup> The maximal or minimal coefficients correspond to those with values at least half the size of the local maximum or minimum. That is, for  $s = \{1, \dots, J\}$  and  $t = \{2, \dots, N-3\}$ <sup>22</sup>

$$\begin{aligned} W_{\max} &\equiv \{W_{s,t} \geq 0.5 \cdot \max(\{W_{s,t-2}, \dots, W_{s,t+2}\})\} \\ W_{\min} &\equiv \{W_{s,t} \leq 0.5 \cdot \min(\{W_{s,t-2}, \dots, W_{s,t+2}\})\} \end{aligned} \quad (13)$$

For  $t = \{0, 1, N-2, N-1\}$  and  $s = \{1, \dots, J\}$ , the coefficient  $W_{s,N-1}$  is considered as maximal if

$$W_{s,N-1} \geq 0.5 \cdot \max(\{W_{s,N-3}, \dots, W_{s,N-1}, W_{s,1}, W_{s,0}\}) \quad (14)$$

and minimal if

$$W_{s,N-1} \leq 0.5 \cdot \min(\{W_{s,N-3}, \dots, W_{s,N-1}, W_{s,1}, W_{s,0}\}) \quad (15)$$

To identify those maximal and minimal wavelet coefficients originated from spikes, a lenient threshold is employed. The survivals are denoted as  $W'_{\max}$  and  $W'_{\min}$ , respectively. A search algorithm looking at the position and directionality of the maximal or minimal coefficients in the same or adjacent scales is implemented to identify the spikes in the wavelet domain.<sup>13</sup> The maxima or minima chains representing the spikes are set to zero. Subsequently, all wavelet coefficients are re-shifted back out of temporal alignment as follows

$$W_{s,t} = W_{s,t+T_s} \quad (16)$$

After the coefficients corresponding to the spikes have been removed, the signal can be reconstructed by the inverse MODWT (iMODWT).

The flowchart to implement the proposed algorithm is shown in Figure 3, and nine sub-steps that are indicated by the numbers in red are included. Sub-steps 1–3 correspond to the spike detection, and the following steps are used to implement the spike removal. It can be found that the IWT is not performed in the procedure of spike detection, and the spike removal only focuses on the time range corresponding to the spike.

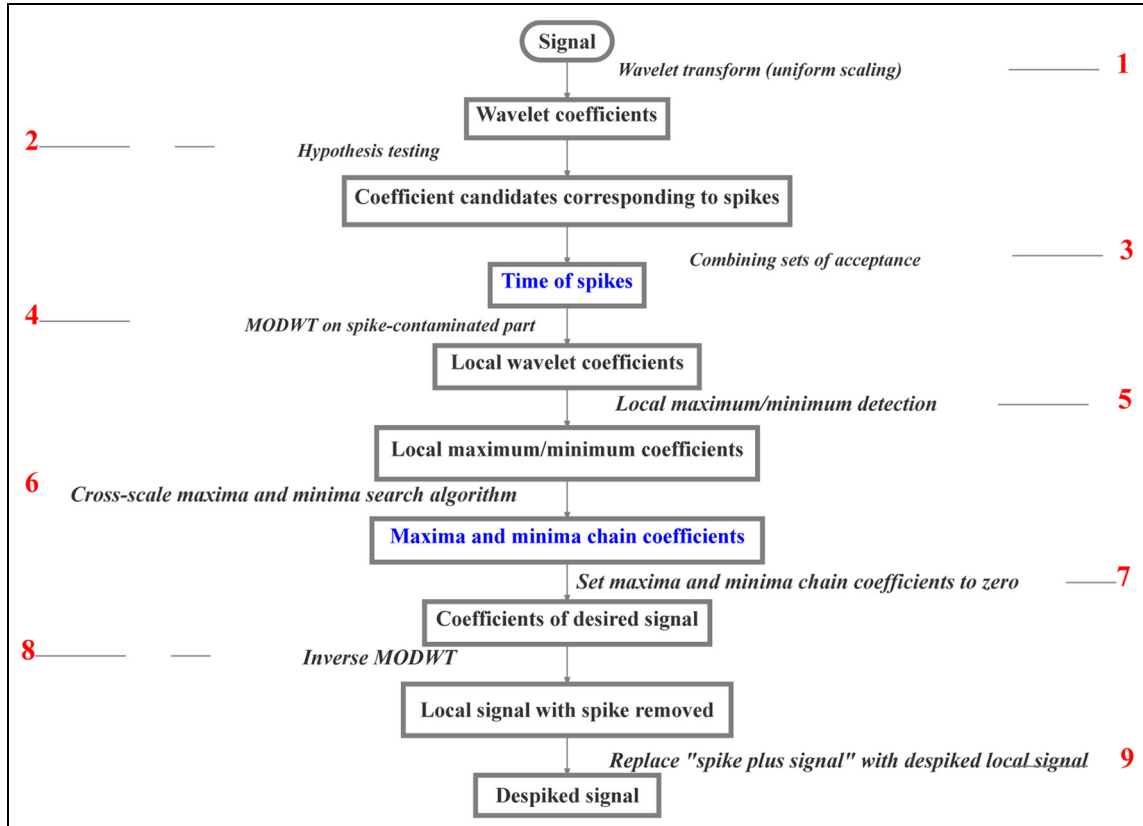
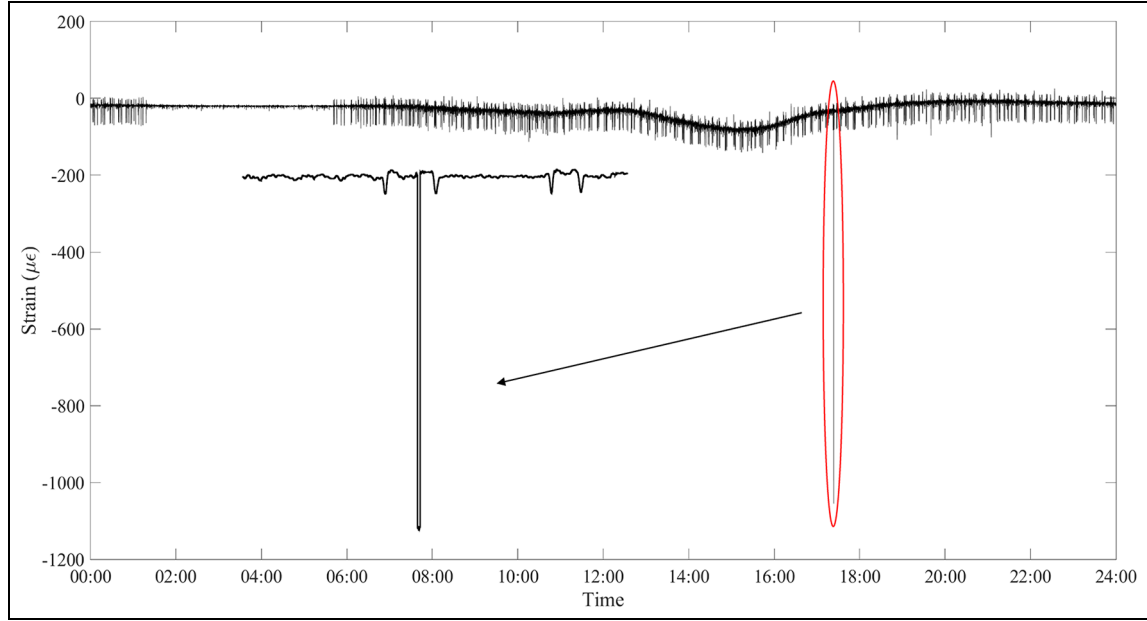


Figure 3. Flowchart of proposed despiking algorithm.



**Figure 4.** Time series of data set 1.

As a result, the efficiency of the algorithm has been improved significantly, which is especially important for the processing of large data. In spike removal, only those wavelet coefficients associated with the spikes are eliminated, retaining those signal components at the time of the spikes. It is more reasonable than those methods that compress the spikes to the level of local median.

### Application to real data

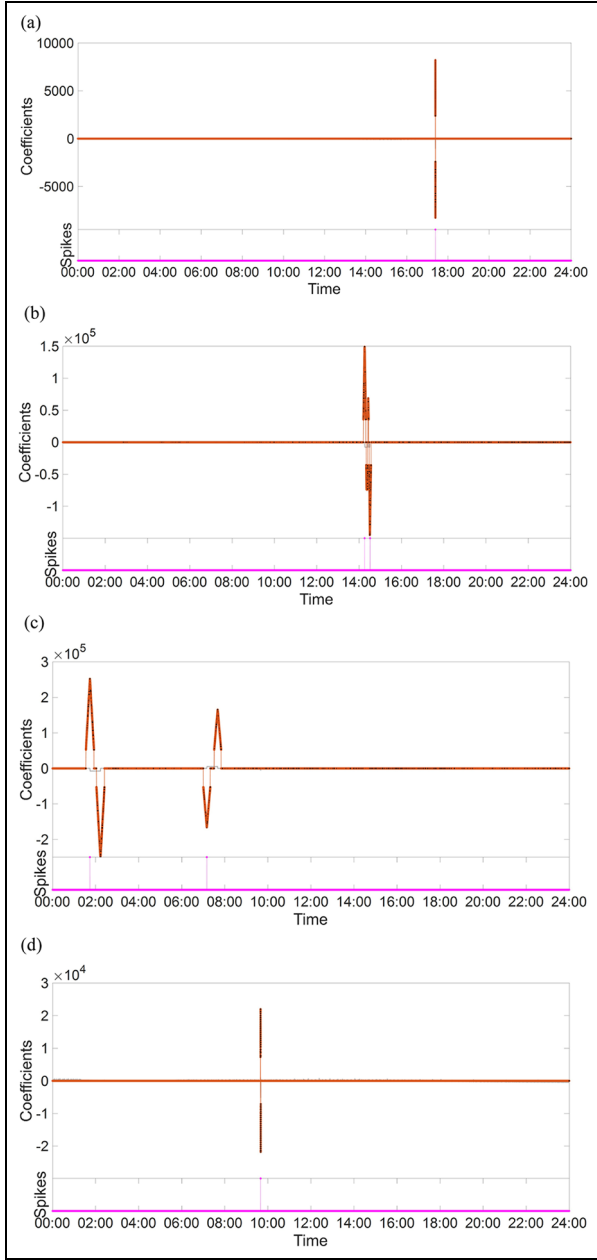
Three sets of data collected from the TMB in Hong Kong are used to examine the capability of the proposed algorithm. Set 1 is the strain data that were collected at a sampling rate of 51.2 Hz. A single spike is embedded in the signal, as shown in Figure 4. Set 2 and Set 3 are the vertical displacement of the deck. They were collected at a sampling rate of 2.56 Hz. Set 2 is shown in Figure 2(a), which has two spikes adjacent to each other. And Set 3 is shown in Figure 2(b), which is contaminated by three spikes with different shapes and durations.

The arrival time of the spikes in the three data sets is identified based on the sparse representation of the signals and hypothesis tests on the wavelet coefficients, and the results are shown in Figure 5. The detection of spikes in Set 3 was executed in two steps because the duration of the third spike is much shorter than those of the two previous ones. The magenta lines in Figure 5 indicate the arrival time of the spikes. Focusing on the time domain surrounding the transients, the spikes are eliminated using the spike removal algorithm. The

despiked signals are shown in Figure 6. It can be observed that the spikes have been well removed. The computational time is 11.24, 2.15 and 4.47 s for Sets 1 to 3, respectively. Set 1 costs much more time because its sampling rate is 51.2 Hz, much higher than those of the other two sets (2.56 Hz). Set 3 consumes more time than set 2 because it has two more spikes.

The proposed method is compared against three other methods: (1) a time-domain method that identifies spikes by a threshold and compresses them to the level of the local median; (2) frequency filter; and (3) a traditional wavelet method without the fast spike-detection algorithm. The time it takes to process the three sets of data using the time-domain method is much less, which is 5.09, 1.97, and 2.34 s, respectively. However, useful signal components in the red circle of Figure 2(b) are eliminated together with the spikes, as shown in Figure 7(a). In contrast, the proposed algorithm retains those signal components well (Figure 7(b)). The long-term SHM data collected from the TMB are stimulated by various loads, and the interactions between the loads and the bridge structure are intricate. Consequently, the signals are composed of multiple components. It is unavoidable that in the frequency domain, the spikes overlap with the useful signal components. Thus, it is found that the frequency filter is not able to extract the spikes as expected. Even when the frequencies of the spikes can be discriminated, a serious Gibbs phenomenon is observed. Figure 8 shows the first two spikes extracted from data set 3 by the frequency filter. The Gibbs phenomenon can be easily observed from Figure 8. The traditional wavelet

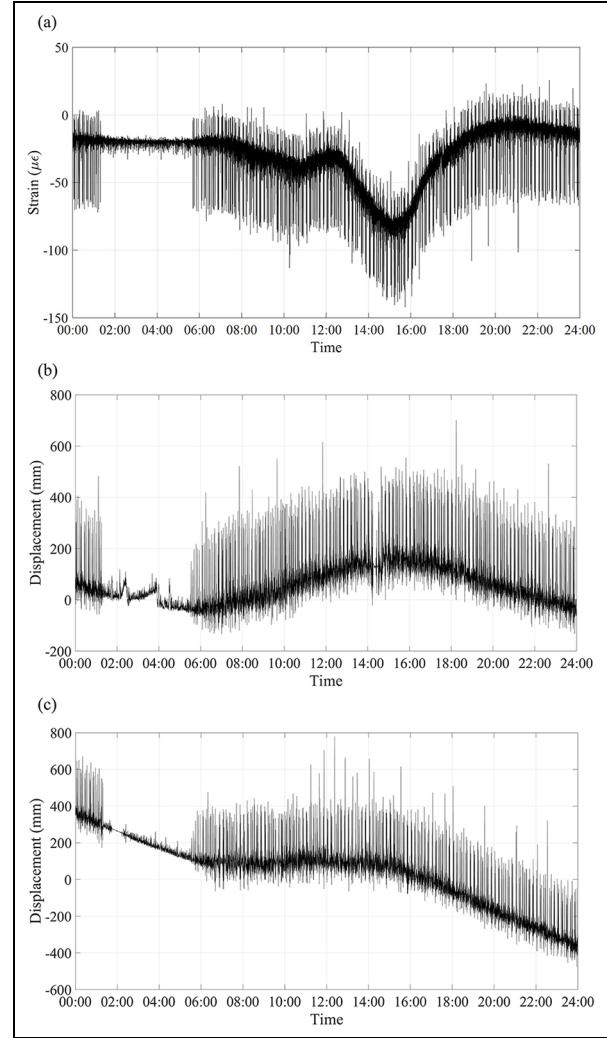




**Figure 5.** Arrival time of spikes: (a) Set 1 (17:23:38), (b) Set 2 (14:15:46 and 14:31:22, respectively), (c) Set 3 (1:44:5 and 7:9:48 for the first and the second spikes, respectively) and (d) Set 3 (9:39:48 for the third spike).

method is found to be too computationally expensive because it needs to identify the chains of maximum or minimum coefficients representing the spikes from the beginning to the end of the data first. Particularly, when this method is used to process data set 1 that has a large volume, the computer stopped due to out of memory.

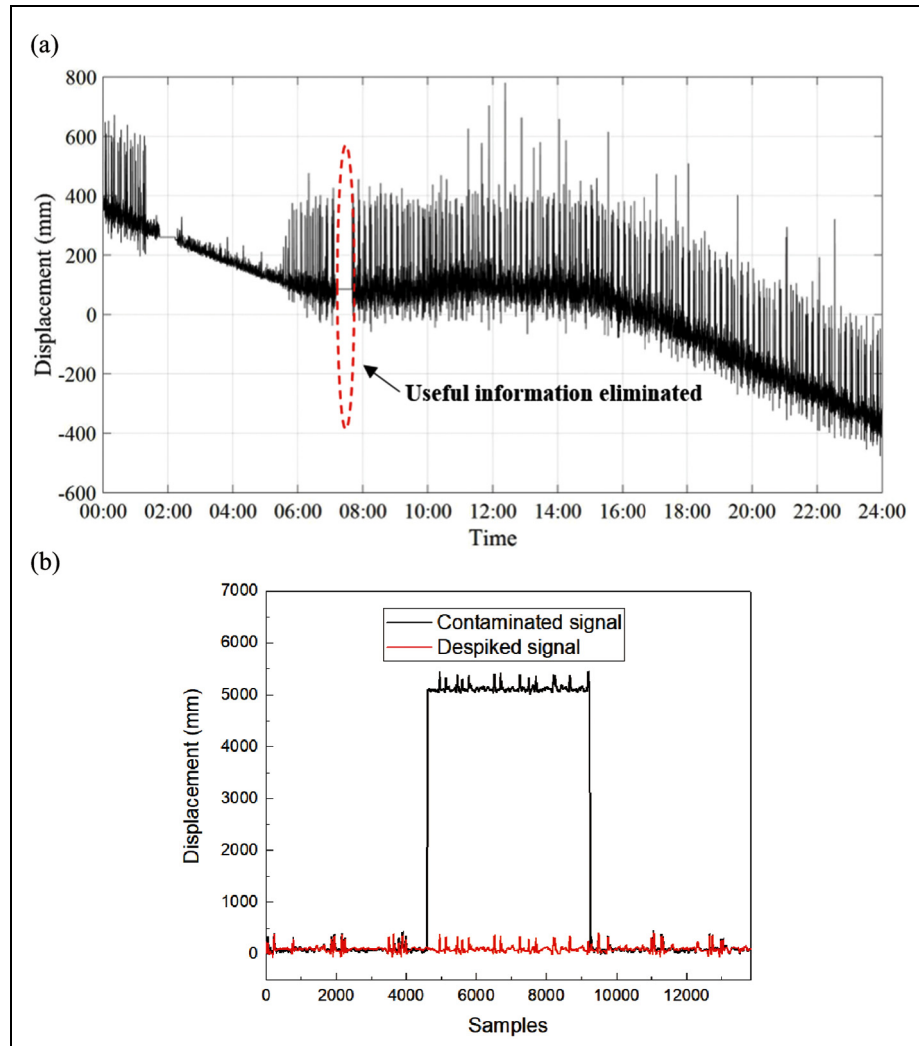
The result comparison of the proposed algorithm with other three methods is shown in Table 1. Both the computational time and the effectiveness are evaluated.



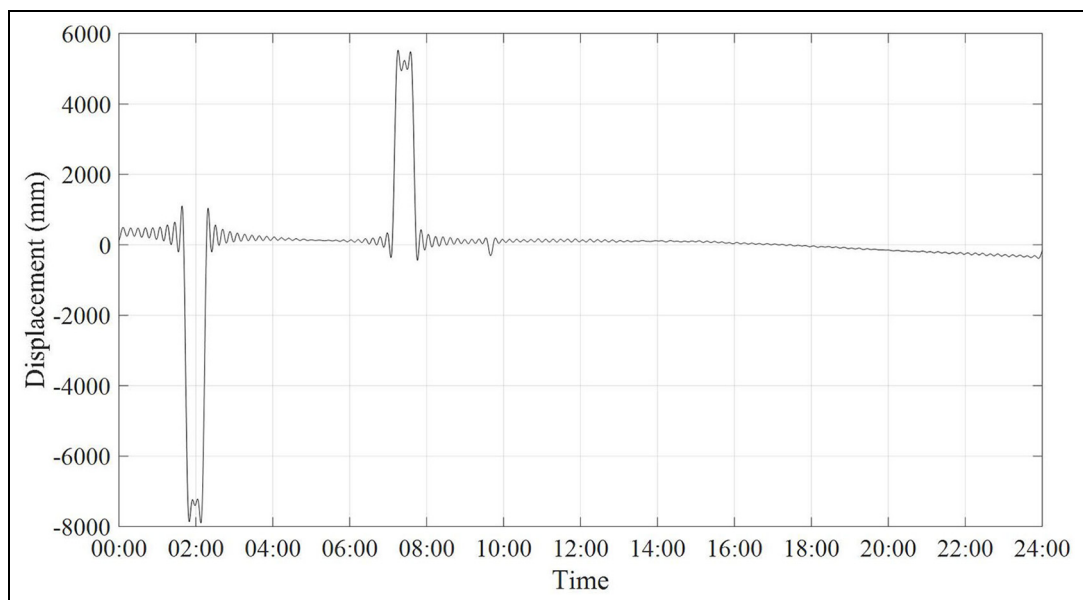
**Figure 6.** Despiked signals for: (a) Set 1, (b) Set 2 and (c) Set 3.

The consumed time may fluctuate when using different computers or the same computer at different times. It can be found that when there is a single spike that has a salient magnitude and does not contain much useful information, the time-domain method is enough. However, in general, the proposed method outperforms the other three methods. Even though in computation cost it is shaded by the time-domain method, it wins in the effectiveness. Actually, its computational cost is also acceptable, because even when it is used to process the data collected at the highest sampling rate for the bridge, the time to process the test data is only around 10 s. In particular, after thresholds are set for the algorithm based on the knowledge of a certain data type, the despiking can be performed automatically. This is of great significance to process data of large volume. Therefore, to remove the spikes embedded in large SHM data, which have complex spatial and temporal patterns, the proposed wavelet-based despiking algorithm is promising.





**Figure 7.** Drawback of time-domain method: (a) signal despiked by time-domain method and (b) information-retaining capability of proposed despiking algorithm (the signal segment is the one marked by a red circle in Figure 2(b)).



**Figure 8.** Gibbs phenomenon due to frequency filter.

**Table 1.** Result comparison.

Data set	Time-domain method		Frequency filter		Traditional wavelet method		Proposed algorithm	
	Time	Result	Time	Result	Time	Result	Time	Result
1	5.09	Satisfactory	63.80	Fail	–	–	11.24	Satisfactory
2	1.97	Unsatisfactory	4.06	Unsatisfactory	11.04	Satisfactory	2.15	Satisfactory
3	2.34	Unsatisfactory	3.94	Unsatisfactory	11.13	Satisfactory	4.47	Satisfactory

The unit for *Time* is second.

## Conclusion

To remove the diverse and complex spikes embedded in the large SHM data fast and robustly, a wavelet-based algorithm is proposed. Other techniques of hypothesis testing and search of wavelet modulus maxima chain are incorporated. This algorithm detects the spikes first by conducting hypothesis tests on the wavelet coefficients and combining the sets of acceptance in different scales. The arrival time of the spikes is identified in this procedure. And then, the MODWT is performed on the signals surrounding the spikes and removes the transients fast, with unaffected information intact. The performance of this proposed algorithm is demonstrated by three sets of data measured from a real bridge. These data sets are characterized by different spatial and temporal patterns. Through comparing with three other methods, the time-domain method, frequency filter and traditional wavelet method, the performance of this algorithm is demonstrated. The conclusions are as follows:

- For signals with a single spike that has salient magnitude and contains little useful information, the time-domain method is enough.
- Frequency filters may be incompetent to remove the spikes in long-term SHM data because of the complexity of the signals and the Gibbs phenomenon.
- The traditional wavelet method loses in computational cost when processing large data, because it costs too much time in searching the spikes from the beginning of the signals to the end.
- The proposed wavelet-based method combining other techniques is promising to realize fast and unsupervised removal of spikes from long-term SHM data.

## Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this

article: This work was supported by the grant from the National Natural Science Foundation of China (Grant No. 51708315) and the Provincial Key Research and Development Program of Shandong (2018GSF120017).

## ORCID iD

Yun-Xia Xia  <https://orcid.org/0000-0001-8101-3636>

## References

1. Halfpenny A. Data preprocessing for damage detection. In: C Boller, FK Chang, Y Fujino, et al. (eds) *Encyclopedia of structural health monitoring*. Hoboken, NJ: Wiley, 2009, pp.557–575.
2. Nenadic Z and Burdick JW. Spike detection using the continuous wavelet transform. *IEEE T Biomed Eng* 2005; 52(1): 74–87.
3. Welsh JP and Schwarz C. Multielectrode recording from the cerebellum. In: MAL Nicolelis (ed.) *Methods for neural ensemble recordings*. 1st ed. Boca Raton, FL: CRC Press, 1998, pp.79–100.
4. Bankman IN, Johnson KO and Schneider W. Optimal detection, classification, and superposition resolution in neural waveform recordings. *IEEE T Biomed Eng* 1993; 40: 836–841.
5. Abeles M and Goldstein MH. Multispike train analysis. *P IEEE* 1977; 65: 762–773.
6. Zanos TP, Mineault PJ and Pack CC. Removal of spurious correlations between spikes and local field potentials. *J Neurophysiol* 2011; 105: 474–486.
7. Daubechies I. *Ten lectures on wavelets*. Philadelphia, PA: SIAM, 1992.
8. Mallat S. *A wavelet tour of signal processing*. 2nd ed. San Diego, CA: Academic Press, 1999.
9. Cantero D and Basu B. Railway infrastructure damage detection using wavelet transformed acceleration response of traversing vehicle. *Struct Control Hlth* 2015; 22: 62–70.
10. Hera A and Hou Z. Application of wavelet approach for ASCE structural health monitoring benchmark studies. *J Eng Mech* 2004; 130: 96–104.
11. Hester D and González A. A wavelet-based damage detection algorithm based on bridge acceleration response to a vehicle. *Mech Syst Signal Pr* 2012; 28: 145–166.
12. Kim H and Melhem H. Damage detection of structures by wavelet analysis. *Eng Struct* 2004; 26: 347–362.
13. Omenzetter P, Brownjohn JMW and Moyo P. Identification of unusual events in multi-channel bridge monitoring data. *Mech Syst Signal Pr* 2004; 18: 409–430.

14. He WY, Zhu S and Ren WX. A wavelet finite element-based adaptive-scale damage detection strategy. *Smart Struct Syst* 2014; 14: 285–305.
15. He WY and Zhu S. Adaptive-scale damage detection strategy for plate structures based on wavelet finite element model. *Struct Eng Mech* 2015; 45: 239–256.
16. Ye XW, Su Y, Xi P, et al. Structural health monitoring data reconstruction of a concrete cable-stayed bridge based on wavelet multi-resolution analysis and support vector machine. *Comput Concrete* 2017; 20: 555–562.
17. Yang X and Shamma SA. A totally automated system for the detection and classification of neural spikes. *IEEE T Biomed Eng* 1988; 35: 806–816.
18. Frisch M and Messer H. The use of the wavelet transform in the detection of an unknown transient signal. *IEEE T Inform Theory* 1992; 38: 892–897.
19. Liu TT and Fraser-Smith AC. Detection of transients in 1/f noise with the undecimated discrete wavelet transform. *IEEE T Signal Proces* 2000; 48: 1458–1462.
20. Costabel S and Müller-Petke M. Despiking of magnetic resonance signals in time and wavelet domains. *Near Surf Geophys* 2014; 12: 185–197.
21. Donoho DL and Johnstone JM. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 1994; 81: 425–455.
22. Patel AX, Kundu P, Rubinov M, et al. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *Neuroimage* 2014; 95: 287–304.
23. Percival DB and Walden AT. *Wavelet methods for time series analysis*. London: Cambridge University Press, 2006.