Smart Structures Systems, *Vol. 20, No. 2 (2017) 219-229*
DOI: https://doi.org/10.12989/sss.2017.20.2.219

Outlier detection of GPS monitoring data using relational analysis and negative selection algorithm

Ting-Hua Yi[1a], X.W. Ye[*2], Hong-Nan Li[1b] and Qing Guo[1c]

[1]*School of Civil Engineering, Dalian University of Technology, Dalian 116023, China*
[2]*Department of Civil Engineering, Zhejiang University, Hangzhou 310058, China*

(Received December 30, 2016, Revised April 11, 2017, Accepted April 18, 2017)

Abstract. Outlier detection is an imperative task to identify the occurrence of abnormal events before the structures are suffered from sudden failure during their service lives. This paper proposes a two-phase method for the outlier detection of Global Positioning System (GPS) monitoring data. Prompt judgment of the occurrence of abnormal data is firstly carried out by use of the relational analysis as the relationship among the data obtained from the adjacent locations following a certain rule. Then, a negative selection algorithm (NSA) is adopted for further accurate localization of the abnormal data. To reduce the computation cost in the NSA, an improved scheme by integrating the adjustable radius into the training stage is designed and implemented. Numerical simulations and experimental verifications demonstrate that the proposed method is encouraging compared with the original method in the aspects of efficiency and reliability. This method is only based on the monitoring data without the requirement of the engineer expertise on the structural operational characteristics, which can be easily embedded in a software system for the continuous and reliable monitoring of civil infrastructure.

Keywords: structural health monitoring; global positioning system; outlier detection; grey relational analysis; negative selection algorithm

1. Introduction

The importance of structural health monitoring (SHM) has been recognized by civil engineers throughout the world in securing proper operation of infrastructure during the service lifetime and in enhancing the state of knowledge in structural conditions under various environmental loads (Li *et al.* 2015a,b, Ye *et al*. 2013, Ye *et al*. 2016a,b,c). In view of this, the process of elaborate instrumentation, measurement and analysis of response data can be considered as the most challenging task in implementation of an effective SHM system. Advances in the usage of modern geodetic techniques, such as the Global Positioning System (GPS), offer a useful method to recover static, quasi-static and dynamic responses in large-scale civil infrastructure exposed to ambient effects such as temperature differentials, earthquakes or gust-winds (Yi *et al.* 2013). Since its introduction to civilian users in the early 1980's, the GPS has been playing an increasingly important role in deformation monitoring for civil infrastructure because it can directly measure the position coordinates with higher precision and sampling rate. Some typical applications include long-span bridges (Nakamura 2000, Meng *et al.* 2007, Moschas and Stiros 2014), high-rise buildings (Kijewski-Correa *et al.* 2006, Breuer *et al.* 2008, Park *et al.* 2008), and large-span spatial structures (Ogaja 2007, Casciati and Fuggini 2011), can be found in vast amount of references.

Although GPS-based monitoring technology provides an opportunity to identify "anomalies" that may signal unusual loading conditions or abnormal structural behaviors, the accuracy and robustness of this technology, especially the identified outlier level (including false and missed alarms) still not be statistically determined (Psimoulis and Stiros 2013). This is because that the on-line GPS monitoring data often exhibit systematic biases (ionospheric delay, receiver noise), significant variability (GPS satellite geometry and multipath effect), or bad/missing data (instrument failure, poor or uncalibrated instrumentation), etc. (Kijewski-Correa and Kareem 2007, Yi *et al.* 2012, Moschas and Stiros 2013). Without proper detection and pretreatment methods, the necessary interpretation for structural behavior will be difficult. Thus, the sophisticated approaches within the computer-based software process these data using some innovative methods specifically tailored for real-time GPS monitoring are needed. Early in 1997, Mertikas and Rizos investigated cumulative-sum test to monitor and control the quality of GPS measurements for critical real-time and/or deformation applications. Tiberius and Kenselaar (2003) presented reasonably simple variance component estimation formulas for four characteristics of GPS observables' noise. Demonstration on zero baseline GPS data shows that: (1)

*Corresponding author, Associate Professor
E-mail: cexwye@zju.edu.cn
[a] Professor
E-mail: yth@dlut.edu.cn
[b] Professor
E-mail: hnli@dlut.edu.cn
[c] MSc Student
E-mail: guoqing.06@163.com

http://www.techno-press.com/journals/sss&subpage=7
ISSN: 1738-1584 (Print), 1738-1991 (Online)

the variance of a GPS observable generally depends on the elevation of the satellite, (2) covariance between channels seems to be negligible, (3) significant correlation can occur between the observation types, and (4) time correlation over some 10-20 s can be present. Ogaja *et al.* (2003) developed an approach that employs the principal component analysis of wavelet transformed GPS solutions to detect out of control signals. Satirapod *et al.* (2003) proposed a multipath mitigation technique based on the use of wavelet decomposition, which can be used to correct for multipath at permanent GPS stations.

Meanwhile, Satirapod *et al.* (2003) compared the use of wavelets, the semiparametric model, and new stochastic modeling techniques, to mitigate the impact of systematic errors on GPS positioning results in both the theoretical and numerical sense. Following that, Tsai *et al.* (2004) suggested a GPS fault detection and exclusion method based on moving average filters. Kijewski-Correa and Kochly (2007) presented a method to characterize multipath effect using Fourier and wavelet spectra and the GPS distortion signature, which can remove the multipath effects even when a GPS distortion signature is not available on-site. Ragheb *et al.* (2009) proposed a combined method of single epoch, sidereal filtering and multi-antenna array, which was successful in detecting horizontal and vertical displacements in simulated near-real time with reduced errors reaching the few-millimeter level. Zhong *et al.* (2010) developed a sidereal filtering based on single differences for mitigating GPS multipath effects on short baselines. By the characteristic analysis of the GPS signal, Yi *et al.* (2011) established a specific set of generating and monitoring systems for multipath signals and carried out a series of controlled experiments to assess the efficiency of the proposed improved particle filtering. Psimoulis and Stiros (2012) presented a computer-based algorithm computing the mean amplitude of small-scale, high frequency oscillations of civil engineering structures using GPS and Robotic Theodolites (RTS). Yi *et al.* (2013) proposed a novel wavelet based multi-step filtering method by combining the merits of the continuous wavelet transform with the discrete stationary wavelet transform so as to address the GPS deformation monitoring application more efficiently. It can be remarked from these references that rational data interpretations are a biggest challenge for civil engineers who are specialized in GPS based long-term monitoring. Although there have been several important developments, these new technologies have not been widely accepted. Especially, the available approaches can only determine preliminary whether there is abnormal date but not give accurately the interval of the outlier in monitoring data. If so, a new question will arise that is whether these methods are suitable for the real-time GPS displacement monitoring or not.

In this paper, a novel two-phase outlier detection method for on-line GPS monitoring data is proposed. The method combines the advantages of relational analysis and negative selection algorithm (NSA) that can detect and locate the occurrence of abnormal events quickly without requiring prior knowledge of the structural operational characteristics. The paper is organized as follows: Section 2 describes the basic idea, main features and detailed implementation procedures of the proposed method. Sections 3 and 4 explore its detection performance for anomalies by using simulated and field measured data, respectively. Finally, conclusions are drawn in Section 5.

2. Proposed method for outlier detection of GPS monitoring data

2.1 Overview of proposed approach

An on-line SHM system installed on a large-scale civil engineering structure can be regarded as a complex nonlinear system which is hardly to be accurately formulated with an explicit mathematical expression. The main reasons are due to the randomness and uncertainty inherent among different structural components as well as the high nonlinearity within various kinds of influential factors. The monitoring data obtained by sensory subsystem are derived from the response outputs of the instrumented structure under the stochastic ambient excitations. For example, the three-dimensional coordinates exported from GPS receivers deployed on the bridge can be easily transformed into the structural displacements in the longitudinal, transversal, and vertical directions by use of the method of coordinate transformation. Therefore, timely and correct interpretation of these collected data is essential to safer system operations and reduction in the number of false alarms.

It is well known that under the normal operational circumstance, the relationships among the structural monitoring data obtained from adjacent measurement locations will follow the certain rules, e.g. positive correlation, negative correlation or no correlation (Fig. 1). Without a doubt, these rules will be changed with the location of the GPS receiver as well as the occurrence of the abnormal monitoring data which are usually induced by the GPS cycle slip, electromagnetic interference, malfunction of the transmission channel, abnormity of the structural condition, etc.

In this connection, outlier detection of the monitoring data can be realized on the basis of the variation of the relational rule within the measurement data. Considering the grey relational theory is able to quickly and quantitatively characterize the correlation among different time series (Deng 1982), thereby it may provide an effective way to identify the abnormal displacement or deflection behavior of structures under operational conditions.

However, this method can only provide the preliminary results of whether outlier exists but not accurately locate the abnormal GPS monitoring data. Therefore, in this study, an improved NSA is proposed subsequently and integrated with the grey relational analysis method for accurate localization of the abnormal GPS monitoring data. Combining these two methods will generate useful insights into the problems of GPS data analysis and interpretation.

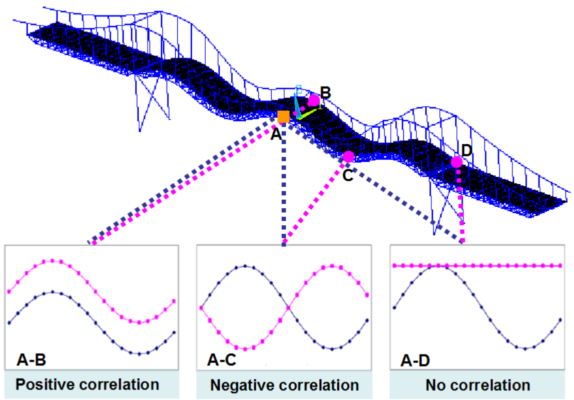


Fig. 1 Relational analysis of monitoring data for outlier detection

2.2 Preliminary outlier detection using grey relational analysis

Gray relational analysis is a kind of system analysis method which has two merits: not requiring a lot of sampling data and not requiring typical distributing rule. The essence of this method is a comparison between two relevant sequences or their curves' geometric shape. The more similar curves will represent the closer developing trends and result in the higher relational degrees. To investigate a SHM system where the number of GPS receivers is m, the displacement time series from two arbitrary receivers are denoted as X_i and X_j (i, j=1, 2, …, m). When regarding X_i is the reference data series and X_j is the analysis data series, two sets of GPS monitoring data in the same time interval n can be expressed as

$$X_i = \{x_i(t)\},\ t = 1,\ 2,\ ...,\ n \tag{1}$$

$$X_j = \{x_j(t)\},\ t = 1,\ 2,\ ...,\ n \tag{2}$$

The basic thought of relational theory is to ascertain the correlation degree according to the similarity among the monitoring data series, that is, the more similarity between two date series, the higher the correlation degree. Assume the moving time window is k, the grey relational grade $R(t')$ will be obtained through continuously computing the relational grades between X_i and X_j (t'=1, 2, …, n/k). The relational grade of the monitoring data should be situated within a reasonable and stable range when the structure is operated under the normal condition and the abnormal situation is not occurred in acquisition and transmission of the GPS monitoring data. Thus, the standard deviation of the relational grade can be set as the alarming threshold by conducting the statistical analysis of the relational grade series of the huge amount of normal monitoring data. In accordance with the theorem of small probability event, the control limit is chosen to be three times of the standard deviation of the relational grade series, and then the normal range can be set as ($\bar{R}-3\sigma/\sqrt{n}, \bar{R}+3\sigma/\sqrt{n}$), where $\bar{R}$ and σ are mean value and standard deviation of the relational grade, respectively. When the relational grade is larger than the alarming threshold, the range of the abnormal data can be roughly determined according to the moving time window k. Considering X_i and X_j belong to the GPS data series of different measurement locations, the normalization process is adopted to eliminate the dimensional effect before the relational analysis is implemented, and then the data sets can be obtained as

$$X_i' = X_i D = \{x_i(1)d, x_i(2)d, ..., x_i(n)d\} = \left\{x_i'(1), x_i'(2), ..., x_i'(n)\right\} \tag{3}$$

$$X_j' = X_j D = \{x_j(1)d, x_j(2)d, ..., x_j(n)d\} = \left\{x_j'(1), x_j'(2), ..., x_j'(n)\right\} \tag{4}$$

where $x_i(t)d = \dfrac{x_i(t) - \min x_i(t)}{\max x_i(t) - \min x_i(t)}$; $t = 1, 2, ..., n$; and D is the normalization factor.

The relational grade of two data series $R_{i,j}(X_i', X_j')$ can be computed after normalization. In this study, the relational grade of B-mode is used to analyze the effect of external influencing factors on the relational degree due to it is able to comprehensively consider the displacement difference, the first-order gradient difference, and the second-order gradient difference (Zheng *et al.* 2014). The relational grade of two data series can be formulated as

$$R_{ij}(X_i', X_j') = \frac{1}{1 + \frac{1}{n} d_{ij}^{(0)} + \frac{1}{n-1} d_{ij}^{(1)} + \frac{1}{n-2} d_{ij}^{(2)}} \tag{5}$$

where $d_{ij}^{(0)} = \sum_{k=1}^{n} \left| x_i'(t) - x_j'(t) \right|$,

$d_{ij}^{(1)} = \sum_{k=1}^{n-1} \left| x_i'(t+1) - x_j'(t+1) - x_i'(t) + x_j'(t) \right|$, and

$d_{ij}^{(2)} = \sum_{k=2}^{n-1} \left| [x_i'(t+1) - x_j'(t+1)] - 2[x_i'(t) - x_j'(t)] + x_i'(t-1) - x_j'(t-1)] \right|$

The distinctive advantage of the relational analysis method is the fast computing speed which means it suitable for online outlier detection. However, the relational grade is devoted to describing the relevance between two data series. When the relational grade is larger than the specified alarming threshold, it can only be concluded that the abnormality in one of the GPS monitoring data series occurs, but which the abnormal receiver cannot be determined. Furthermore, different ranges of abnormal data will be obtained if the moving time window k is set to different values. Thus, the consequence resulting from the above-mentioned factors is that only the preliminary results of outlier detection for GPS monitoring data are achieved. In order to tackle the problem, an improved NSA is proposed for further accurate detection and localization of the abnormal GPS monitoring data.

2.3 Improved negative selection algorithm for accurate outlier detection

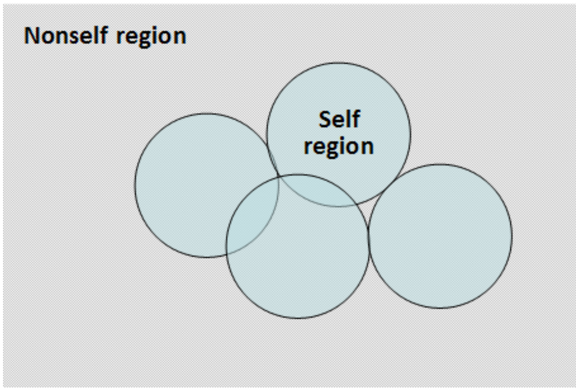
Artificial immune systems are a type of intelligent algorithm inspired by the principles and processes of the human immune system which is a complicated mechanism to protect the human body from the invasion of various types of foreign antigens, such as bacteria and viruses. Among them, the negative selection in binary form is one of the most famous algorithms, which was first proposed by Hofmeyr and Forrest (2000) by use of the immune system's property of distinguishing any foreign cells from the body's own cells. As illustrated in Fig. 2, it includes two data sets, where the "Self" representing the normal data while the data set "Nonself" denotes the abnormal data. This algorithm basically consists of two steps. First, the candidate detector is generated at random and tested as to whether it recognizes the Self samples. Only those that cannot recognize Self samples at all are added to the detector set. Then, the new input data is examined as to whether they are recognized by the detector set. It determines that the recognized data are abnormal and the unrecognized are normal. This method has the distinctive advantages of simplicity and easy to implementation, and has been successfully used for several areas such as fault diagnosis, anomaly detection, and computer security due to its high robustness and parallelism in analyzing of discrete space problem.

Nevertheless, the long-term GPS monitoring data are belonging to the continuous space and it difficult to be expressed by the binary representation. Thus, the NSA in binary form is inappropriate to solve this kind of naturally real valued problem.

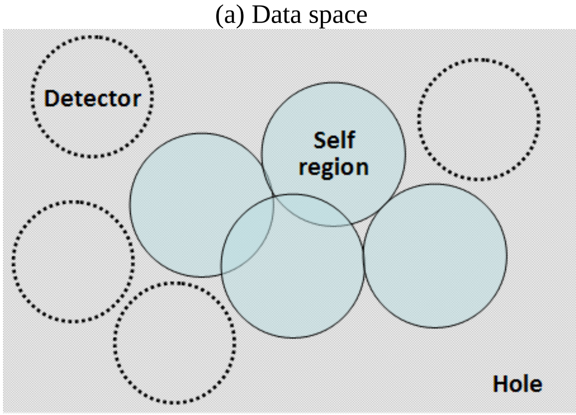
To solve this problem, Gonzalez and Dasgupta (2003) developed a real-valued NSA with a continuous searching space. In this method, the Self and Nonself space corresponds to a unitary hypercube $[0,1]^n$. The Self set and the detector are described by the centers and the fixed radii respectively, that is $s=(c_s, r_s)$ and $d=(c_d, r_d)$, where c_s and c_d are the centers, and r_s and r_d are the fixed radii of them. It should be noted that the distances between the Self samples and the Self center are less than r_s; while those of the Nonself samples are larger than r_s. When using the real-valued matching principle for outlier detection, the abnormal data are recognized if the distances between the elements to be detected and the center of the detector are less than the radius r_d. In general, the distance between two samples in the real-valued space is expressed as Euclidean distance

$$d(x,y) = (\sum_{i=1}^{n} |x_i - y_i|^2)^{1/2} \qquad x, y \in R^n \tag{6}$$

It is known the NSAs no matter based on binary or real-valued encoding will bring forth the well-known issue of "holes", as shown in Fig. 2(b). It means that the perfect detector coverage is realized hardly because a huge amount of detectors are needed to well cover the Nonself space. In addition, the computational efficiency of the algorithm will be reduced to a certain degree if the total number of detectors is not controlled. In this study, an improved algorithm incorporated with the adjustable-sized detector and Self is proposed towards solving the hole problem in the real-valued NSA.



(a) Data space



(b) Detector coverage

Fig. 2 Main concept of negative selection algorithm

(1) Using detector of adjustable radius

A certain area of the Nonself space can be covered by use of the detector with the fixed radius. Usually, a large number of detectors are required to be produced in order to cover more areas of the Nonself space which will unavoidably result in low computational efficiency. Thus, the detectors with adjustable radii are adopted in this study (Ji and Dasgupta, 2009). By so doing, the number of the detector will be dramatically decreased because the radius can be dynamically resized according to the case studied. It includes two basic steps: (i) the detector d is stochastically generated in the Nonself space and the distance between the detector d and the Self sample $x_s \in self$ is calculated by Eq. (6) which is denoted as $d(x_s, d_i)$. It will be chosen as the candidate detector if $d(x_s, d_i)>r_s$, and (ii) the radius of the detector r_d is set as r_d=min($d(x_s, d_i)$)-r_s. Because the number of detectors is a key factor that decides coverage rate, it is desirable to choose as few detectors as possible so as to improve the performance of the algorithm. Therefore, the method of target coverage is selected to overcome the overlapping problem of the detector. The appropriate number of the detector will be automatically created in accordance with the coverage rate which is estimated by Monte Carlo method here. Assume the number of the random samples in the detection space is p, the coverage rate is calculated by c_0=1-1/p if only one sample is not covered.

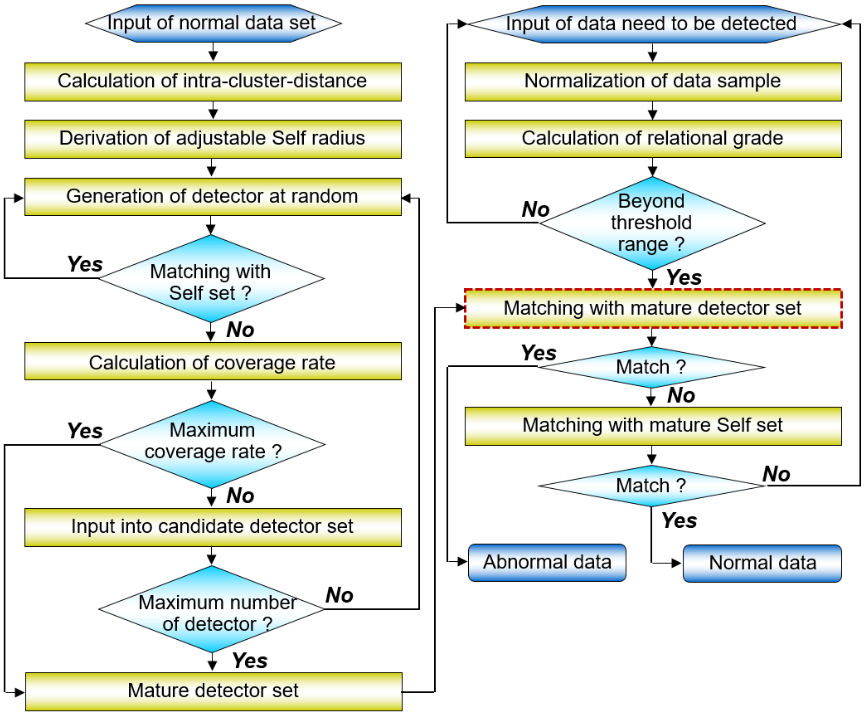


Fig. 3 Flowchart of proposed approach for outlier detection

(2) Using Self of adjustable radius

In the real-world SHM practice, the normal long-term monitoring data are continuously varied, and only part of the structural response data will be acquired as the number of the sensors is limited. Thus, the improvement of the representation pattern of the Self is urgently needed to effectively cover the normal monitoring data and avoid the false alarm in checking the abnormal data due to the excessive coverage of the Nonself space. As we know that, the normal data of the structural responses obtained by the GPS receiver are fluctuated within certain amplitude and aggregated in the scatter diagram (Yi *et al.* 2013), while the abnormal data usually are away from the data aggregation center. Inspired by this phenomenon, the radius of the Self is designed to adjust automatically according to the distance between the monitoring data and the data aggregation center, making that the radius is enlarged for the Self sample close to the data aggregation center and reduced for the Self sample far away from the center.

In the Self space, the intra-cluster-distance is defined as

$$ds_i = \sum_{j=1}^{n} d(s_i, s_j) \tag{7}$$

where $d(s_i, s_j)$ is the Euclidean distance of the Self sample.

Then, the adjustable radius of the Self sample can be express as

$$rs_i = \frac{(ds_i - \min(ds))(c_1 - 1)}{(\max(ds) - \min(ds))c_1} rs \tag{8}$$

where c_1 is the influencing parameter related to the radius; and $ds=(ds_1, ds_2, \ldots, ds_n)$ is the intra-cluster-distance of the whole Self samples.

In Eq. (8), the effect of the distance between the monitoring data and the data aggregation center on the radius of the Self sample is considered by the parameter c_1. It makes sure that the radius of the Self sample is also able to be adaptively adjusted in accordance with the distribution of the normal monitoring data, which further improves the detection rate of the abnormal data. Because the parameters adopted in this algorithm are not known in advance, hence, it is necessary to determine them with a trial and error method. What need to be mentioned is that a large number of the normal monitoring data are required for the all NSAs to generate the mature detector. In addition, the efficiency of the outlier detection will be greatly reduced if all the GPS monitoring data are participated in the calculation. Therefore, the relational model can be firstly employed to quickly judge the occurrence of the abnormal data, and then an accurate analysis for the localization of the abnormal data is carried out by use of the improved NSA with the adjustable radius.

2.4 Implementation procedure of the proposed method

As illustrated in Fig. 3, the implementation procedure of the proposed method for outlier detection of the GPS monitoring data can be divided into the following steps.

Step (1): The GPS monitoring data under the normal vibration condition are captured and processed by the mean removal method. The time window k is set up and the Self is obtained through windowing the monitoring data series.

Step (2): The Self radius r_s and its influencing factor c_1 are settled and the Self set data are normalized. The adjustable Self radius is generated by Eq. (8).

Step (3): The maximum number of the detector T_{max} and the ideal coverage rate of the detector are given and the mature detector is created by use of the generation method of the detector.

Step (4): The relational model is established for the data samples studied. The relational grade series R_{12}={R_{12}(1), R_{12}(2), …, R_{12}(n/k)} are calculated by Eq. (5). The mean value $\bar{R}$ and the standard deviation σ of R_{12} are also derived.

Step (5): The threshold range $(\bar{R}-3\sigma/\sqrt{n},\bar{R}+3\sigma/\sqrt{n})$ is set up, which is employed for preliminary outlier detection of the GPS monitoring data.

Step (6): The preliminarily detected GPS data are then put into the Self set and the mature detector set. These data should be estimated as the normal data if they are matched with the Self set; while if they are matched with the mature detector set, the range of the abnormal data will be final determined.

3. Numerical simulation study

3.1 Preparation of simulation data

To verify the effectiveness of the proposed method in outlier detection of the GPS monitoring data, a numerical study is conducted by used of the sinusoidal periodic signal. A sinusoidal time history with a sampling frequency of 10 Hz is expressed by

$$y_1=\begin{cases}\sin(\pi t) & t\in(0,6.4)\cup(8.5,10)\\ \mathrm{rand}(t)+\sin(\pi t) & t\in(6.5,8.4)\end{cases} \tag{9}$$

where t is the sampling time; and rand(t) means the random values in the range of [-1, 1].

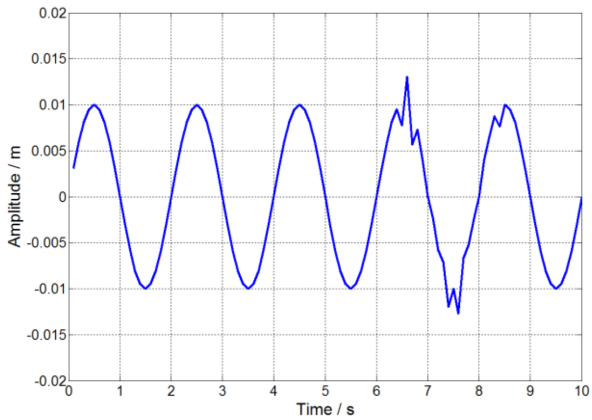
The abnormality is simulated by inputting into the random variation between 6.5 s and 8.4 s, as illustrated in Fig. 4(a). As shown in Fig. 4(b), the simulated GPS reference data series can be expressed by

$$y_2=\sin\left[\pi(t-2)\right],\quad t\in(1,10) \tag{10}$$

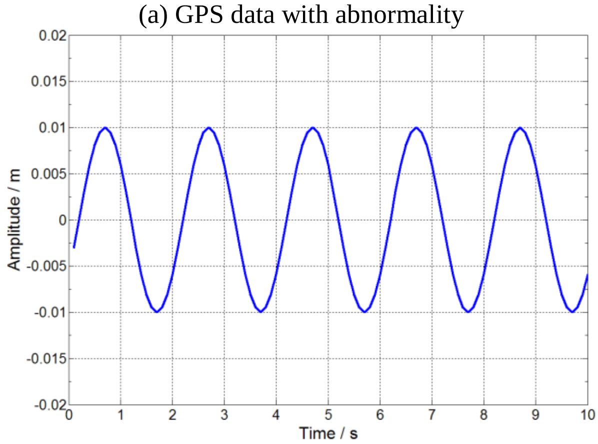
3.2 Discussion of numerical results

The moving time window is set as k=10 as the sampling frequency of the GPS receiver is 10 Hz here. The relational grade series $R(t)$ are calculated by Eq. (5), and the threshold range is obtained by the statistical analysis. Fig. 5 shows the alarming model by the grey relational analysis for preliminary outlier detection. It is seen from Fig. 5 that the relational grade series between 6 s and 9 s exceeded the alarming threshold. From this observation, a preliminary decision can be made that the abnormal data should be existed in the above-mentioned two GPS data set between 6 s and 9 s. The NSA is employed to identify the abnormal data set and time range. In this study, both the original (with the fixed radius) and the improved NSA (with the adjustable radius) are applied to carry out a comparative investigation on their effectiveness in outlier detection during the detector generation and abnormality identification stages. The GPS data with abnormality is processed by windowing with a length of 2 and a two-dimensional vector y_1=($y_{1.1}$, $y_{1.2}$, …,$y_{1.50}$) with a length of 50 are obtained. The first 30 samples are taken as the Self data set. For the original NSA, the detector radius is 0.05, the Self radius is 0.02, and the number of the detector is 65. For the improved NSA, the coverage rate of the detector is 0.95, the maximum number of the detector is 200, and the influencing factor of the Self radius is 2.

Fig. 6 shows the generated results of the detector. The dashed circles are the detector set, the solid circles represent the Self set, and the triangles are denoted as the sample set to be detected. The generated detectors with the fixed radius are illustrated in Fig. 6(a). It is seen from this figure that less Nonself space is covered by the detectors. The abnormal data cannot be detected if they are located in the holes. While it can be obviously observed from Fig. 6(b) that the Nonself space is effectively covered by use of the adjustable radius detectors. In addition, the radius is also adaptively adjusted in accordance with the Self samples to cover more Self space. With regard to the number of the detector, the number of the fixed radius detectors is 60 and the number of the adjustable radius detectors is just 51 when the coverage rate is 0.95.



(a) GPS data with abnormality



(b) GPS reference data

Fig. 4 Simulated GPS sinusoidal periodic signals

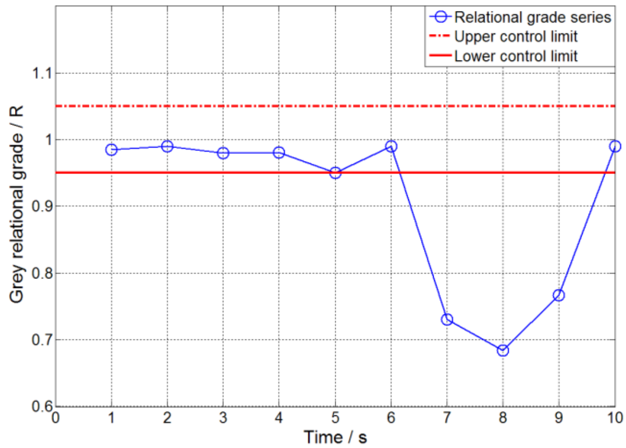
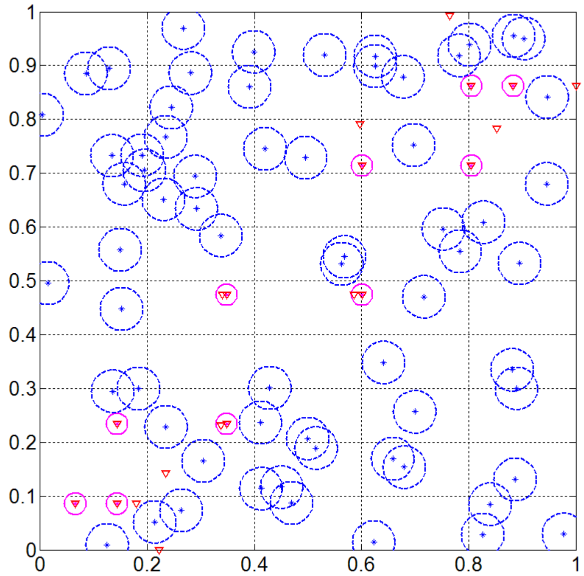
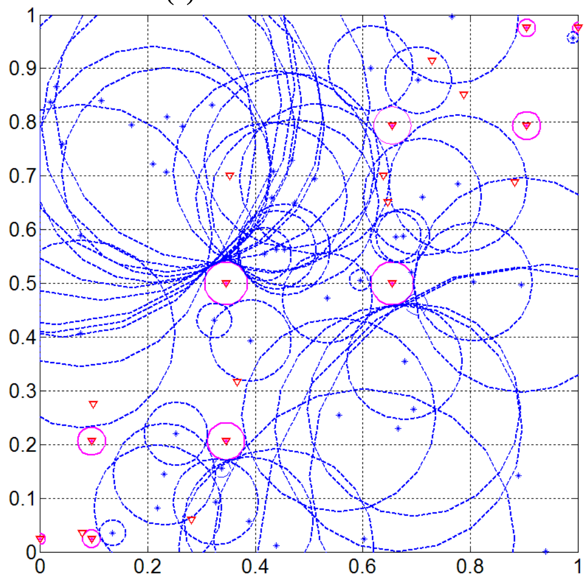


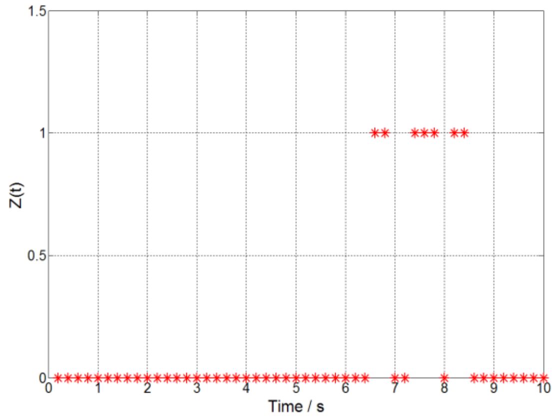
Fig. 5 Grey relational analysis for outlier detection



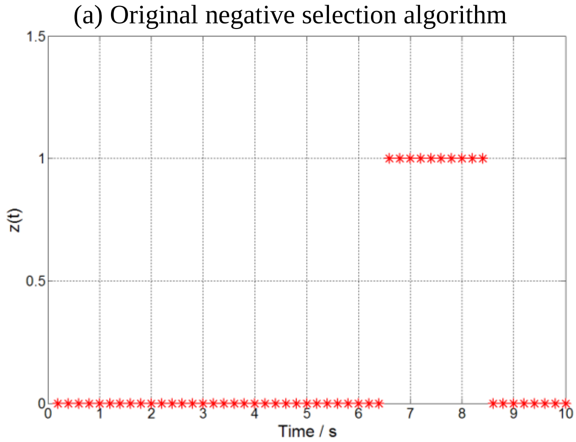
(a) Fixed radius detector



(b) Adjustable radius detector
Fig. 6 Generated results of detectors



(a) Original negative selection algorithm



(b) Improved negative selection algorithm
Fig. 7 Results of outlier detection

Thus, it can be concluded that more Nonself space can be covered by the adjustable radius detectors with less number which yield a better algorithmic performance. Fig. 7 shows the results of the outlier detection by use of the two algorithms. The samples are normal data if $z(t)$=0; while they are abnormal data if $z(t)$=1. It can be seen from Fig. 7(a) that the false alarming is made by the fixed radius detector at t=7 s, 7.2 s, and 8 s. Fig. 7(b) illustrates the obtained outlier detection results by use of the improved algorithm. It is seen from Fig. 7(b) that the range of the abnormal data can be accurately determined as (6.5 s, 8.4 s).

4. Field experimental verification

To verify the effectiveness of the proposed method for outlier detection of on-site GPS monitoring data, field experimental investigations on the Dalian BeiDa Bridge are carried out. As illustrated in Fig. 8, the bridge is located at the Binhai Road, Dalian, China, which was constructed in 1987 in order to commemorate the establishment of the friendly relationship between Kitakyushu, Japan and Dalian, China. It is a three span simply-supported stiffening truss suspension bridge with a length of 230 m. The length of the mid-span is 132 m and the length of two side-spans is 48 m. The width of the bridge is 12 m, the length of the roadway is 8 m, and the length of each footway is 2 m. The height of the tower is 35 m.

4.1 Outline of field measurement

Based on the results of the finite element analysis, the first several order frequencies of the bridge are less than 2 Hz. The sampling frequency of the GPS receiver is thus can be set as 10 Hz according to Nyquist sampling theorem. The dual-frequency carrier-phase NovAtel GPS receiver is applied in the field experiments, which is able to communicate and transmit the wired or wireless signals. The advantages of this device include measurement of the code and phase with high accuracy, low signal-to-noise ratio, and less power consumption. A total of three GPS receivers are installed on the bridge with one GPS receiver located at the coast near the bridge working as the reference station and two GPS receivers located at the mid-span working as the rover station. Fig. 8 gives the site deployment of the GPS receives with the detailed specifications. The raw data measured by the GPS receiver are processed by the commercial software GrafNav which is a Windows-based post-processing package for GPS data developed by Waypoint Products Group, NovAtel Inc., Canada (2012). This software has a powerful post-processing capability for GPS data by use of the carrier phase dynamic Kalman filtering technology. The modulated data can be transformed into the structural response data in the longitudinal, transverse, and vertical directions using the coordinate transformation.

4.2 Analysis of experimental results

In this study, the measurement data in the vertical direction obtained from the two GPS rover stations are acquired for further investigation. As illustrated in Figs. 9(a) and 9(c), the two set of measurement GPS data are denoted as X_1 and X_2 with the time duration of 500 s. A random abnormal data set is incorporated into X_1 from t= 476 s to t= 478 s to form X_3, as shown in Fig. 9(b).

By use of the data series X_2 and X_3, the relational grade series R_{23} are calculated according to Eq. (5) by setting the moving time window k as 10. As illustrated in Fig. 10, the relational grade from t= 474 s to t= 479 s is larger than the alarming threshold which means that the abnormality exits in the data sample number from 4740 to 4790.

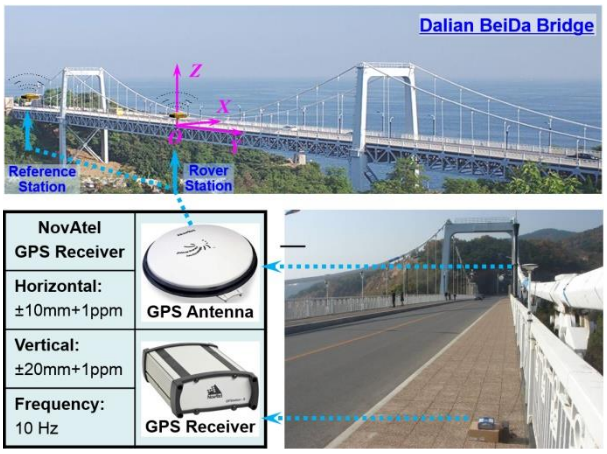
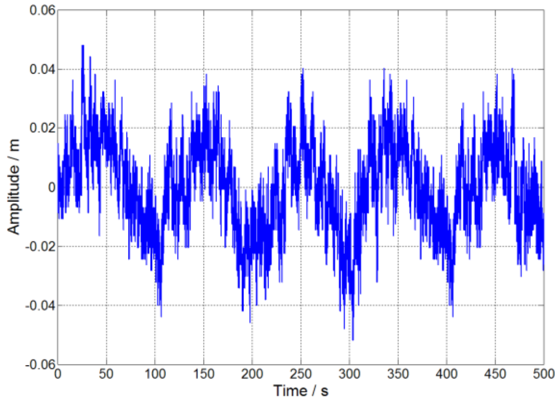
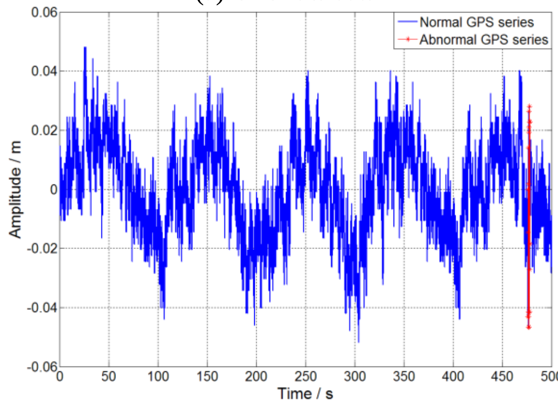


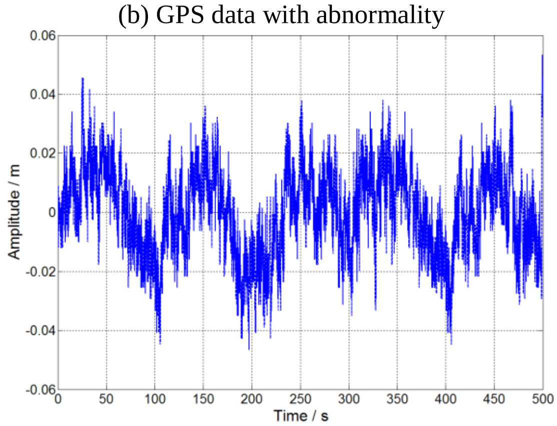
Fig. 8 Deployment of GPS receivers on Dalian BeiDa Bridge



(a) GPS raw data



(b) GPS data with abnormality



(c) GPS reference data

Fig. 9 Field measured GPS data

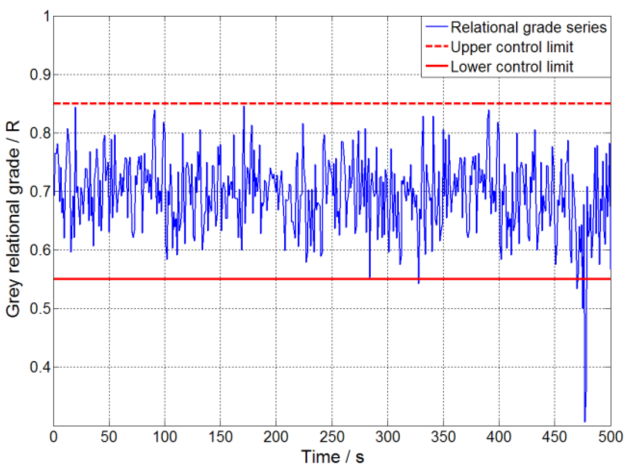
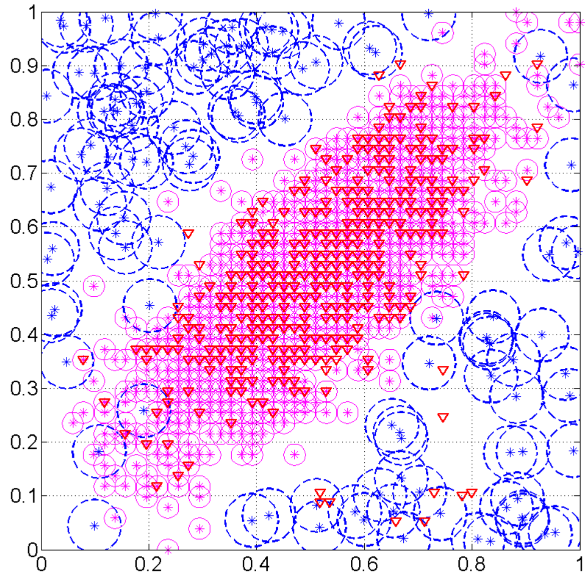
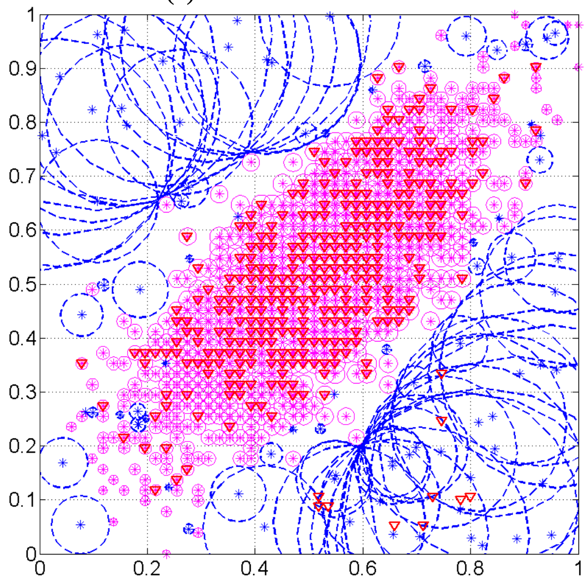


Fig. 10 Grey relational analysis for outlier detection

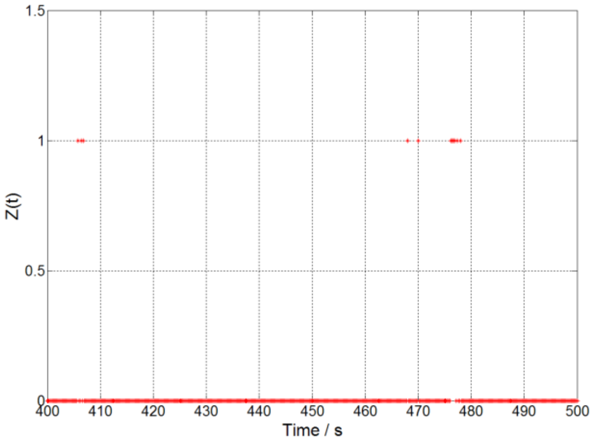


(a) Fixed radius detector

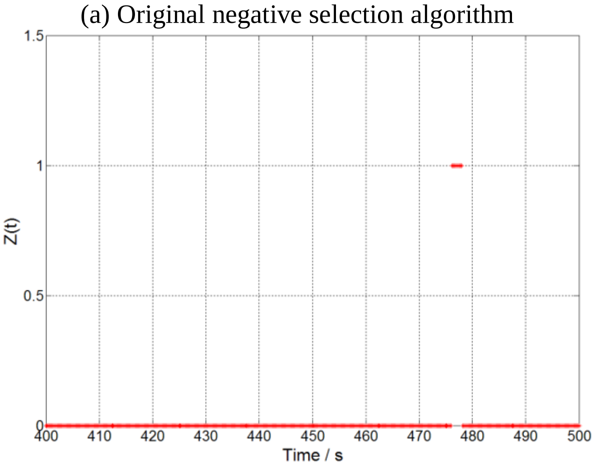


(b) Adjustable radius detector

Fig. 11 Generation of detectors



(a) Original negative selection algorithm



(b) Improved negative selection algorithm

Fig. 12 Results of outlier detection

The two algorithms are used to conduct the outlier detection of the data series as before. The first 4000 data samples are regarded as the Self set and the data series from 4000 to 5000 are the samples to be detected. The number of the detector with the fixed radius is 110, the Self radius is 0.01, and the radius of the detector is 0.05. For the improved algorithm with the adjustable radius, the coverage rate is set as 0.95, the influencing factor of the ontology radius is 2, and the maximum number of the detector is 200.

The generated detectors with the fixed radius and the adjustable radius are illustrated in Fig. 11. The dashed circles are the detector set, the solid circles represent the Self set, and the triangles are denoted as the samples to be detected. It is seen from Fig. 11(a) that there are many holes in the Nonself space covered by the fixed radius detectors.

It is shown in Fig. 11(b) that the number of the hole is effectively reduced by use of the adjustable radius detectors and Self samples. The number of the detector with the fixed radius is 110 and the number of the detector with the adjustable radius is 99 with the coverage rate being as 0.95. It is therefore concluded that the relatively large Nonself space can be well covered by the improved algorithm with less number of the detector, and the hole areas are greatly decreased.

The results of the outlier detection for the samples to be detected are illustrated in Fig. 12. In the same way, the samples are normal data if $z(t)=0$; while they are abnormal data if $z(t)=1$. It is revealed from Fig. 12(a) that the false alarming is made by the fixed radius detector for the sample number from 4055 to 4070 and samples 4680 and 4700. As demonstrated in Fig. 12(b), the range of the abnormal data from 4680 to 4780 can be accurately identified by the improved NSA.

5. Conclusions

This paper put forward a practicable method for detecting abnormal data in a GPS-based monitoring system. Using the simulated and field measured data, the performance of the method was investigated in detail. Some conclusions are drawn as follows:

(1) The proposed method merges the merits of relational analysis and NSA in two aspects: the fast computing speed of the former means it suitable for online operation, while the accurate identifying ability of the latter can ensure it has a higher detection rate.

(2) To improve the coverage of Self region, an improved NSA by integrating Self of adjustable radius into the training stage is presented. The modified Self samples with variable radii can cover the Self region effectively by fewer samples, thus reduces the computation cost well.

(3) Numerical and experimental results demonstrated that the proposed method is encouraging compared with the original method in efficiency and reliability without significant increase in complexity. This method is only based on the system's monitoring data without the requirement of the engineer expertise on the structural operational characteristics, which can be easily embedded in a software system for the continuous and reliable monitoring of civil infrastructure.

Acknowledgments

This research work was jointly supported by the National Natural Science Foundation of China (Grant Nos. 51625802, 51478081), the 973 Program (Grant No. 2015CB060000), the Fundamental Research Funds for the Central Universities of China (Grant No. 2017QNA4024), and the Key Lab of Structures Dynamic Behavior and Control (Harbin Institute of Technology), Ministry of Education of the PRC.

References

Breuer, P., Chmielewski, T., Gorski, P., Konopka, E. and Tarczynski, L. (2008), "The Stuttgart TV Tower-displacement of the top caused by the effects of sun and wind", *Eng. Struct*, **30**(10), 2771-2781.

Casciati, F. and Fuggini, C. (2011), "Monitoring a steel building using GPS sensors", *Smart Struct. Syst.*, **7**(5), 349-363.

Deng, J.L. (1982), "Control problems of grey systems", *Syst. Control Lett.*, **1**(5), 288-294.

Gonzalez, F.A. and Dasgupt, D. (2003), "Anomaly detection using real valued negative selection", *Genet. Program. Evol. M.*, **4**(4), 383-403.

GrafNav (2012), NovAtel, Inc., Canada, http://www.novatel.com.

Hofmeyr, S.A. and Forrest, S. (2000), "Architecture for an artificial immune system", *Evol. Comput.*, **8**(4), 443-473.

Ji, Z. and Dasgupta, D. (2009), "V-detector: An efficient negative selection algorithm with 'probably adequate' detector coverage", *Inform. Sciences*, **179**(10), 1390-1406.

Kijewski-Correa, T., Kareem, A. and Kochly, M. (2006), "Experimental verification and full-scale deployment of Global Positioning Systems to monitor the dynamic response of tall buildings", *J. Struct. Eng. - ASCE*, **132**(8), 1242-1253.

Kijewski-Correa, T. and Kareem, A. (2007), "Performance of wavelet transform and empirical mode decomposition in extracting signals embedded in noise", *J. Eng. Mech. - ASCE*, **133**(7), 849-852.

Kijewski-Correa, T. and Kochly, M. (2007), "Monitoring the wind-induced response of tall buildings: GPS performance and the issue of multipath effects", *J. Wind Eng. Ind. Aerod.*, **95**(9), 1176-1198.

Li, J. and Hao, H. (2015a), "Damage detection of shear connectors under moving loads with relative displacement measurements", *Mech. Syst. Signal Pr.*, **60-61**, 124-150.

Li, J. and Hao, H. (2015b), "Structural damage identification with power spectral density transmissibility: Numerical and experimental studies", *Smart Struct. Syst.*, **15**(1), 15-40.

Meng, X., Dodson, A.H. and Roberts, G.W. (2007), "Detecting bridge dynamics with GPS and triaxial accelerometers", *Eng. Struct.*, **29**(11), 3178-3184.

Mertikas, S.P. and Rizos, C. (1997), "On-line detection of abrupt changes in the carrier-phase measurements of GPS", *J. Geodesy*, **71**(8), 469-482.

Moschas, F. and Stiros, S.C. (2014), "Three-dimensional dynamic deflections and natural frequencies of a stiff footbridge based on measurements of collocated sensors", *Struct. Control. Health Monit.*, **21**(1), 23-42.

Moschas, F. and Stiros, S. (2013), "Noise characteristics of high-frequency, short-duration GPS records from analysis of identical, collocated instruments", *Measurement*, **46**(4), 1488-1506.

Nakamura, S. (2000), "GPS measurement of wind-induced suspension bridge girder displacements", *J. Struct. Eng. - ASCE*, **126**(12), 1413-1419.

Ogaja, C., Li, X. and Rizos, C. (2007), "Advances in structural monitoring with global positioning system technology: 1997-2006", *J. Applied Geodesy*, **1**(3), 171-179.

Ogaja, C., Wang, J. and Rizos, C. (2003), "Detection of wind-induced response by wavelet transformed GPS solutions", *J. Surv. Eng. - ASCE*, **129**(3), 99-104.

Park H.S., Sohn, H.G., Kim, I.S. and Park, J.H. (2008), "Application of GPS to monitoring of wind-induced responses of high-rise buildings", *Struct. Des. Tall. Spec.*, **17**(1), 117-132.

Psimoulis, P.A. and Stiros, S.C. (2012), "A supervised learning computer-based algorithm to derive the amplitude of oscillations of structures using noisy GPS and Robotic Theodolites (RTS) records", *Comput. Struct.*, **92**, 337-348.

Psimoulis, P.A. and Stiros, S.C. (2013), "Measuring deflections of a short-span railway bridge using a robotic total station", *J. Bridge Eng. - ASCE*, **18**(2), 182-185.

Ragheb, A.E., Edwards, S.J. and Clarke, P.J. (2009), "Using filtered and semicontinuous high rate GPS for monitoring deformations", *J. Surv. Eng. - ASCE*, **136**(2), 72-79.

Satirapod, C., Khoonphool, R. and Rizos, C. (2003), "Multipath mitigation of permanent GPS stations using wavelets", *In Proceedings of 2003 International Symposium on GPS/GNSS*, Tokyo, Japan (pp. 15-18).

Satirapod, C., Wang, J. and Rizos, C. (2003), "Comparing different global positioning system data processing techniques for modeling residual systematic errors", *J. Surv. Eng. - ASCE*, **129**(4), 129-135.

Tiberius, C. and Kenselaar, F. (2003), "Variance component estimation and precise GPS positioning: case study", *J. Surv. Eng. - ASCE*, **129**(1), 11-18.

Tsai, Y.H., Chang, F.R. and Yang, W.C. (2004), "GPS fault detection and exclusion using moving average filters", *IEE Proceedings-Radar, Sonar and Navigation*, **151**(4), 240-247.

Ye, X.W., Ni, Y.Q., Wai, T.T., Wong, K.Y., Zhang, X.M. and Xu, F. (2013), "A vision-based system for dynamic displacement measurement of long-span bridges: algorithm and verification", *Smart Struct. Syst.*, **12**(3-4), 363-379.

Ye, X.W., Dong, C.Z. and Liu, T. (2016a), "Image-based structural dynamic displacement measurement using different multi-object tracking algorithms", *Smart Struct. Syst.*, **17**(6), 935-956.

Ye, X.W., Su, Y.H., Xi, P.S., Chen, B. and Han, J.P. (2016b), "Statistical analysis and probabilistic modeling of WIM monitoring data of an instrumented arch bridge", *Smart Struct.*

Syst., **17**(6), 1087-1105.

Ye, X.W., Dong, C.Z. and Liu, T. (2016c), “Force monitoring of steel cables using vision-based sensing technology: methodology and experimental verification”, *Smart Struct. Syst.*, **18**(3), 585-599.

Yi, T.H., Li, H.N. and Gu, M. (2011), “Characterization and extraction of global positioning system multipath signals using an improved particle-filtering algorithm”, *Meas. Sci. Tech.*, **22**(7), 1-11.

Yi, T.H., Li, H.N. and Gu, M. (2012), “Effect of different construction materials on propagation of GPS monitoring signals”, *Measurement*, **45**(5), 1126-1139.

Yi, T.H., Li, H.N. and Gu, M. (2013), “Recent research and applications of GPS-based monitoring technology for high-rise structures”, *Struct. Control Health Monit.*, **20**(5), 649-670.

Yi, T.H., Li, H.N. and Gu, M. (2013), “Wavelet based multi-step filtering method for bridge health monitoring using GPS and accelerometer”, *Smart Struct. Syst.*, **11**(4), 331-348.

Yi, T.H., Li, H.N. and Gu, M. (2013), “Experimental assessment of high-rate GPS receivers for deformation monitoring of bridge”, *Measurement*, **46**(1), 420-432.

Zheng, C., Li, W., Wang, S., Sun, D., Ma, L., Zhang, M. and Zhang, B. (2014), “Hepatoma cells recognition based on matrix absolute gray relational degree of B-mode”, *Optik*, **125**(4), 1579-1584.

Zhong, P., Ding, X., Yuan, L., Xu, Y., Kwok, K. and Chen, Y. (2010), “Sidereal filtering based on single differences for mitigating GPS multipath effects on short baselines”, *J. Geodesy*, **84**(2), 145-158.