Journal of Materials and Engineering Structures

Research Paper

Study to process abnormal data for GNSS monitoring system of a long-span cable-stayed bridge in Vietnam

Van Hien Le

University of Transport and Communications, No 3 Cau Giay Street, Dong Da District, Hanoi, Vietnam

ARTICLE INFO

Article history:

Received : 15 December 2022

Revised : 17 December 2022

Accepted : 17 December 2022

Keywords:

GNSS monitoring system

long-span cable-stayed bridge

Structural Health Monitoring – SHM

Abnormal data

ABSTRACT

Global Positioning System (GPS) or currently upgraded to Global Navigation Satellite System (GNSS) has been applied in many SHM systems of the super-structures, especially in the long-span bridges. A GNSS system has the ability in monitoring the global deformation of a long-span cable-stayed bridge at the millimeter level of accuracy in real-time. However, the GNSS monitoring dataset acquired from a SHM system includes various noise data such as abnormal data, missing data, and so on. This paper studies de-noising methods to detect and replace the abnormal data of a GPS monitoring dataset acquired from a real cable-stayed bridge in Vietnam. Firstly, a GPS monitoring dataset of an actual long-span cable-stayed bridge was acquired to study processing abnormal data. A scenario of abnormal data was created in a time-series GPS data, and then the Hampel identifier method was applied to detect and replace the abnormal data. The replacing data were then assessed for precision and reliability by using correlation analysis and RMSE criterion. Finally, a long-term GNSS monitoring dataset processed the abnormal data automatically. The results show, that abnormal data in GPS monitoring data can be detected and replaced with high accuracy and reliability.

F. ASMA & H. HAMMOUM (Eds.) special issue, 4th International Conference on Sustainability in Civil Engineering ICSCE 2022, Hanoi, Vietnam, J. Mater. Eng. Struct. 9(4) (2022)

1 Introduction

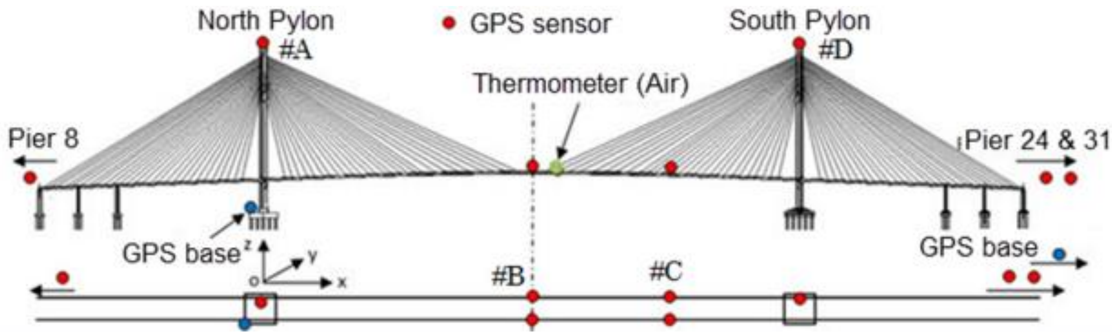
Displacement measurement is a vital work in Structural Health Monitoring (SHM) of a superstructure such as a high-rise building, hydroelectricity dam and long-span bridge, etc. Some traditional methods adopted in displacement measurement of structures are interferometer, electronic distance, or total station. However, these methods have disadvantages in the application; for instance, they need a clear line of orientation in measurement, are affected by climate, and have low accuracy in long distances. Currently, GPS or GNSS technology is a popular technology used for positioning and measuring the distance that has many advantages such as overcoming climate limitations, high accuracy, and measuring in real-time (or near real-time) [1, 2]. Therefore, GPS/GNSS technology has been applied in the monitoring of many structures, especially in long-span bridges [3, 4].

* *Corresponding author. Tel.: +84 981110910.*
E-mail address: hienlv@utc.edu.vn

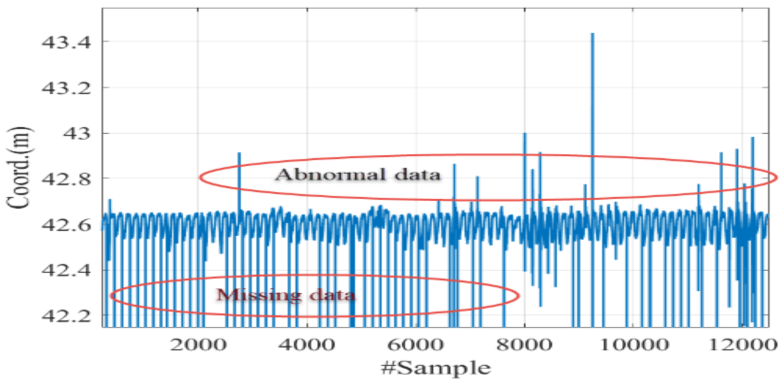
e-ISSN: 2170-127X,

RESEARCH REVIEW of Sciences and Technologies

Additionally, there are some studies on the application of GPS in SHM of long-span bridges showing that the acquired monitoring data are complicated and have many abnormal data such as missing data. For instance, the study of GPS monitoring data of the Can Tho cable-stayed bridge in Vietnam showed that there were so many missing data and abnormal data acquired from the SHM system [1]. In the case of structural health assessment through the monitoring data, abnormal data processing is so important to work to gain an accurate dataset for assessment. There were some studies that verified methods for handling missing parts in time-series data. Those studies showed that the polynomial function estimation could be used to interpolate a few missing data based on the principle of least-square approximation [5, 6]. The polynomial function was also applied automatically to interpolate the GPS missing data of the Can Tho cable-bridge, however, this study just interpolated for less than 4 continuous missing data in a time-series [1].



(a) Dimension of target bridge



(b) Example of a month's actual data

Fig. 1. Description of the target bridge and actual data

Essentially, abnormal data processing is to detect the outliers of time-series and replace them with accurate values. One of the most effective robust methods in detecting and replacing outliers is the Hampel identifier which has been used in many fields and achieved good results[7]-[8]-[9]-[10]. This paper studies applying the Hampel identifier to detect and replace the abnormal data of a GPS monitoring dataset acquired from a real cable-stayed bridge in Vietnam. Firstly, a GPS monitoring dataset of an actual long-span cable-stayed bridge was acquired to study processing abnormal data. A scenario of abnormal data was created in a time-series GPS data, then the Hampel identifier method was applied to detect and replace the abnormal data. The replacing data were then assessed for precision and reliability by using correlation analysis and RMSE criterion. Finally, a long-term GNSS monitoring dataset was processed the abnormal data automatically.

2 GPS monitoring system of Can Tho cable-stayed bridge

The target bridge in this study is the Can Tho cable-stayed bridge located in Southend Vietnam which has a 550m length of main span. This is a symmetric bridge with 2 diamond shape towers. The SHM system of the target bridge was operated in 2010 that combines many kinds of monitoring sensors such as accelerometer, temperature, wind speed, strain, and so on.

A system of GPS sensors was also set up at some specific locations of the bridge for observing displacements at 1Hz data acquisition. The previous study on this bridge [1] showed that there are four GPS locations emphasizing the global deformation of the bridge which are two top tower points, the center main span, and the quarter main span. These four points are selected to study in this paper. Fig. 1a shows the dimension of the target bridge with locations of GPS sensors and air-temperature sensors, while Fig. 1b shows an example of 10-minute average data in a month with a lot of abnormal data and missing data.

3 Application of Hampel identifier for processing abnormal data

In this section, the principle of the Hampel identifier method is described, and then the application strategy for GPS monitoring data is adopted.

3.1 Description of the Hampel identifier method

The Principle of the Hample identifier is to use median value and median absolute deviation for estimating data position, then detecting outliers from the standard deviation of data.

Assume a time-series $x_1, x_2, x_3, \ldots, x_n$ and a sliding window of length k. To define point-to-point median and estimate standard deviation, Eqs. (1), (2) are calculated [11].

$$m_i = median(x_{i-k}, x_{i-k+1}, x_{i-k+2}, \ldots, x_i, x_{i+k-2}, x_{i+k-1}, x_{i+k}) \tag{1}$$

$$\sigma_i = \kappa.median(|x_{i-k} - m_i|, \ldots, |x_{i+k} - m_i|) \tag{2}$$

where: m_i is the local median; $\sigma_i$is standard deviation; $\kappa = \frac{1}{\sqrt{2}erfc^{-1}(1/2)} \approx 1.4826$ is the unbiased estimation of Gaussian distribution. If a sample xi allows condition $|x_i - m_i| > t.\sigma_i$ that means x_i is an outlier, and it's replaced by the local median m_i. The given threshold t equals 3 normally, but for stricter detection, the threshold can be 2.

3.2 Processing GPS abnormal data of the target bridge

The GPS system of the Can Tho cable-stayed bridge was set to acquire data at 1Hz, the raw data were then stored in some types of average values such as 1 min-average, 10 min-average, 1 hour-average, and 1 day-average. The study data here is the 10 min-average GPS monitoring data of four global deformation points which are two points of the top tower, the center main span point, and the quarter main span point. The previous study [1] showed that the air-temperature data has a significant correlation with GPS data of four global deformation points, especially with the deck points. Therefore, the air-temperature data are also used here to verify the interpolated data by assessing the correlated coefficients. The correlated coefficients between GPS data and air-temperature data calculate by Eq. (3).

$$r = \frac{Cov(x,y)}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3}$$

Processing GPS abnormal data of the target bridge is done through 2 main steps:

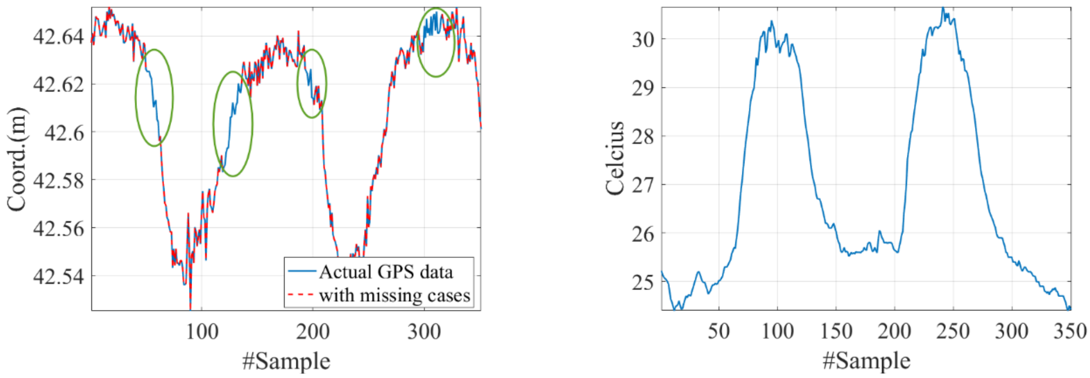
Step 1: clean long-term data in around 2 days of the center main span along to vertical direction and air-temperature are extracted. Then, some specific locations of GPS data are set into missing data. The Hampel identifier is applied to interpolate the specific missing data. The output data are then verified by the correlated coefficients with air-temperature and Root Mean Square Error (RMSE) of the difference between interpolated data and the actual data.

Step 2: The Hampel identifier is then applied to interpolate the abnormal data of a month's GPS monitoring data of four global deformation points. The change in correlation coefficients is also adopted.

4 Experiment analysis

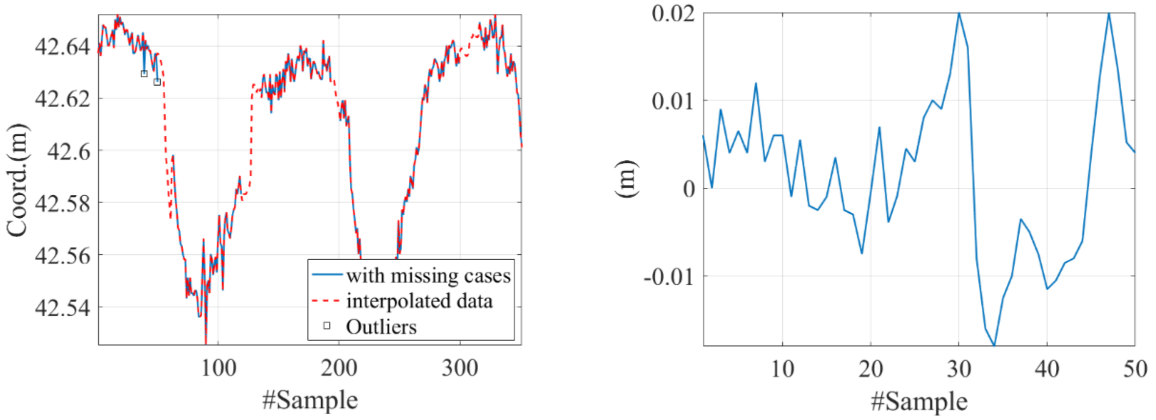
4.1 Assessment of interpolating abnormal data in step 1

Fig. 2a shows overlapped plot of actual GPS data and the scenario of the missing case, while Fig. 2b plots the air-temperature monitoring data at the same time. The correlated coefficient between actual GPS data and air-temperature data is -0.896 which is so high inverse correlation. There are four scenarios of missing cases that are randomly selected at 51 to 61, 120 to 135, 194 to 200, and 300 to 315 (see Fig. 2a). The missing value at each location is set to not a number (NaN). The correlated coefficient between GPS data with missing case and air-temperature is NaN which is a large change compared to the actual GPS data.



(a) Scenario of missing case (b) Air-temperature data

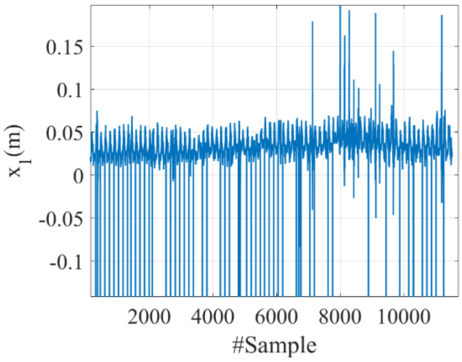
Fig. 2. Missing case scenario and air-temperature data



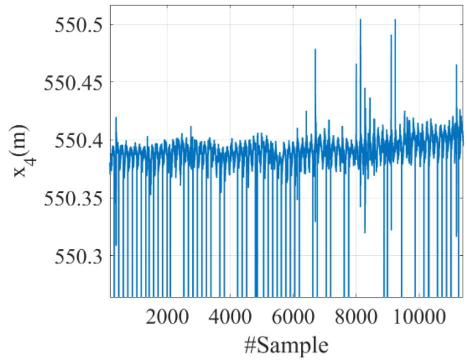
(a) Overlapped input and output data (b) Difference between actual and interpolated data (RMSE = 0.008m)

Fig. 3. Results of interpolating data

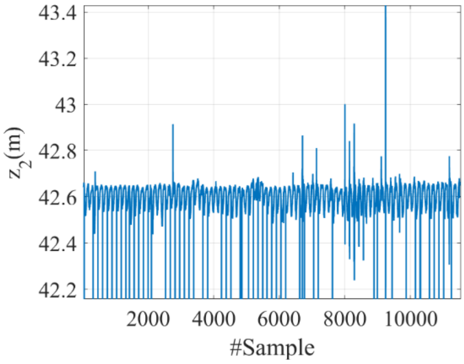
The Hampel identifier method is applied with the sliding window of length k = 10. Fig. 3a shows overlapped plots of GPS data with missing cases and output interpolated results. It can be seen that the output interpolated data matches the input GPS data both in values and data trends. The correlation coefficient between interpolated data and the air-temperature data is -0.888 which has a small change compared with the actual GPS data (-0.896). Furthermore, Fig. 3b shows the small difference between actual data and interpolated data with RMSE = 0.008m which is less than 10% of the amplitude of actual GPS data. It can be concluded that the interpolated data by using the Hample identifier method has high accuracy and reliability.



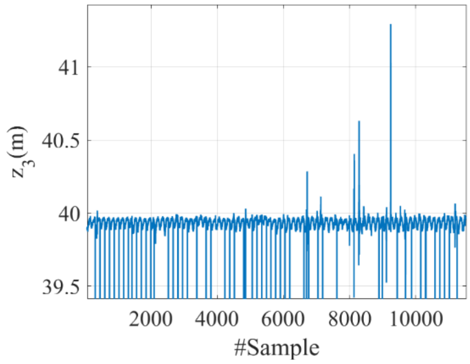
(a) *x*-coordinate of Noth tower (**r = 0.006**)



(b) *x*-coordinate of South tower (**r = 0.007**)

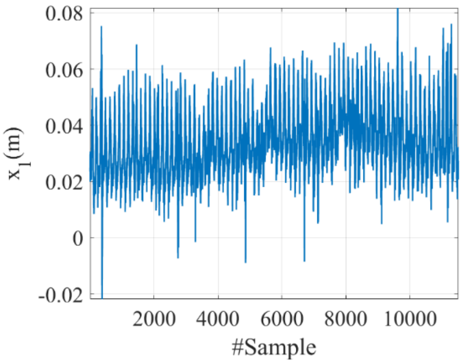


(c) *z*-coordinate of center span (**r = 0.008**)

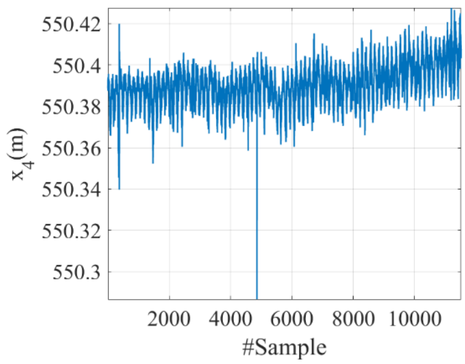


(d) *z*-coordinate of quarter span (**r = 0.008**)

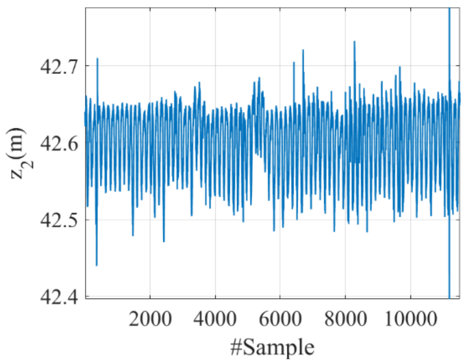
Fig. 4. Actual GPS long-term data and correlation coefficient with air-temperature



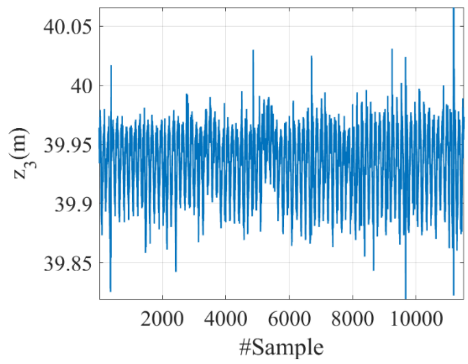
(a) *x*-coordinate of Noth tower (**r = 0.53**)



(b) *x*-coordinate of South tower (**r = 0.44**)



(c) *z*-coordinate of center span (**r = -0.63**)



(d) *z*-coordinate of quarter span (**r = -0.57**)

Fig. 5. Actual GPS long-term data and correlation coefficient with air-temperature

4.2 Application of interpolation method for GPS long-term data

The GPS long-term monitoring data of the target bridge were extracted from January 1st, 2016 to March 30th, 2016. According to the GPS monitoring system, the missing value is replaced by -9999. The previous study [1] showed that the global deformation of the target bridge effecting by air-temperature was realized by the longitudinal direction of tower points (x-direction) and the vertical direction of the deck points (z-direction). Thus, the x- and z- directions of those four points are used to analyze this study. Fig. 4 shows the actual GPS long-term data of four points where the correlation coefficients with the air-temperature are attached separately. It can be seen that the correlation coefficients are so small by the effects of abnormal data, especially missing data. This does not reflect the actual situation of a cable-stayed bridge. Fig. 5 shows the interpolation data by using the Hampel identifier method along each point with attaching correlation coefficient.

It can be concluded that the Hampel identifier method gives good results in processing the abnormal GPS data of the target bridge, especially for missing data. The correlation coefficients with air-temperature also verify the correctness of the actual situation of the structure.

5 Conclusion

In this paper, the GPS abnormal monitoring data acquired from an actual cable-stayed bridge were processed by applying the Hampel identifier method. The output interpolation results were also assessed by some criterions such as correlation coefficients with the air-temperature and RMSE values. Firstly, it can be concluded that there are a lot of abnormal data in an actual GPS monitoring system, especially with missing data. The abnormal data need to be interpolated before using for structural health conditions assessment. Secondly, using the Hampel identifier method to interpolate the GPS abnormal data achieved good results, especially for long-term monitoring data. This method can be applied as a robust estimation for dealing with abnormal data both online and post-processing data.

REFERENCES

[1]- V.H. Le and M. Nishio, Time-series analysis of GPS monitoring data from a long-span bridge considering the global deformation due to air temperature changes. J. Civ. Str. Hea. Monit., 5(4) (2015) 415-425. doi:org/10.1007/s13349-015-0124-9.

[2]- P. Cheng, W. John, and W. Zeng. Large structure heath dynamic monitoring using GPS technology. in FIG XXII International Congress 2002. Washington, DC USA. (2002).

[3]- M. Kaloop and H. Li, Monitoring of bridge deformation using GPS technique. KSCE J. Civ. Eng., 13(6) (2009) 423-431. doi:org/10.1007/s12205-009-0423-y.

[4]- M. Celebi, GPS in dynamic monitoring of long-period structures. Soil. Dyn. Earthquake Eng., 20(5) (2000) 277-483. doi:org/10.1016/S0267-7261(00)00094-4.

[5]- T.D. Pigott, A review of methods for missing data. Educ. Res. Eval., 7(4) (2001) 353-383. doi:org/10.1076/edre.7.4.353.8937.

[6]- C.F. Gerald and P.O. Wheatley, Applied Numerical Analysis. Seventh Edition. Pearson, 2004.

[7]- J. Astola and P. Kuosmanen, Fundamental of Nonlinear Digital Filtering. CRC Press, (1997). doi:org/10.1201/9781003067832.

[8]- R.K. Pearson, Outlier in process modeling and identification. IEEE Trans. Cont. Sys. Technol., 10(1) (2002) 55-63. doi:org/10.1109/87.974338.

[9]- H. Liu, S. hah, and W.Jiang, On-line outlier detection and data cleaning. Comput. Chemic. Eng., 28(9) (2004) 1635-1647. doi:org/10.1016/S0098-1354(04)00024-9.

[10]- R.D. Martin and D.J. Thomson. Robust-resistant spectrum estimation. in Proceedings of the IEEE. (1982), 1097-1115. doi:org/10.1109/PROC.1982.12434.

[11]- Z. Yao, J. Xie, Y. Tian, and Q. Huang, Using Hampel Identifier to Eliminate Profile-Isolated Outliers in Laser Vision Measurement. J. Sens., (2019). doi:dx.doi.org/10.1155/2019/3823691.