

Comparison on outlier detection methods using measured data from a long span tied-arch bridge

Peng Ni

Department of Bridge Engineering, Tongji University, Shanghai 200092, China

Haili Jiang, Wurong Fu

Shanghai Road and Bridge (Group) Co., Ltd., Shanghai 200433, China

Ye Xia, Limin Sun

State Key Laboratory for Disaster Reduction in Civil Engineering, Tongji University, Shanghai 200092, China

Contact: yxia@tongji.edu.cn

Abstract

As the demand for the detections of outliers in the structural health monitoring data-set increases, numerous approaches are presented for it. However, the characteristics of the existing methods dealing with different kinds of measured data are not yet clear enough for practical use. Therefore, this paper conducts a comparative study of several popular rule-based methods based on monitoring data of an arch-tied bridge in China. For measured data, outliers are not known in advance. In this way, this study evaluates and compares the detection performances rely on two indicators: the quantity of the detected outliers and the extreme value of the outliers deviating from the mean of the data. Conclusions on the features and applicable situations of involved methods are given. Additionally, combining the results of different methods proves to be beneficial. Finally, a software incorporating the research results is developed for outlier detection.

Keywords: outlier detection, data faults, structural health monitoring

1 Introduction

For structural health monitoring (SHM), the quality of the measured data is indispensable for accurate results. However, there often are many kinds of data faults due to various factors in the collected data-set. Among them, outlier is a more common one, meaning the data that dissimilar or irrelevant from the other observations present in the data set.

Numerous outliers detection approaches have been presented. Sharma et al. [1] divided these

methods into four types and conducted a comparative study. In [2], popular outlier detection methods are categorized into statistical-based ones, depth-based ones, deviation based ones, distance-based ones, density-based ones, and high dimensional ones. Studies on the performances of the various detection methods were conducted a lot. However, most of them were applied to the artificial data instead of the real measured data. Thus, the features of these methods in practical application are not yet clear enough.