Mechanical Systems and Signal Processing 110 (2018) 139–151

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

# Improvement to the sources selection to identify the low frequency noise induced by flood discharge

Jijian Lian, Xiaoqun Wang *, Bin Ma, Dongming Liu

*State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China*

ARTICLE INFO

*Article history:*
Received 11 October 2017
Received in revised form 9 January 2018
Accepted 15 March 2018
Available online 29 March 2018

*Keywords:*
Low frequency noise
Flood discharge
Single channel
Blind source separation
Source number estimation
Cross-correlation procedure

ABSTRACT

Recent studies indicate that low frequency noise (LFN) is generated by flood discharge of high dam. For prototype observation of the LFN, the observed data usually contain multiple sources induced by different mechanisms. To separate and identify the LFN generated by flood discharge from the multiple sources, an improved version of single-channel blind source separation (SCBSS) is proposed. In this study, the source number estimation is improved by a singular entropy (SE) method based on the eigenvalues calculated by principal components analysis (PCA). Then an SCBSS algorithm with no interruption is proposed. Both traditional method and PCA-SE method may result in extra sources that do not exist actually and are introduced by SCBSS due to the misconduct of human judgment or underestimation of threshold. Therefore, a cross-correlation procedure is proposed to identify and eliminate the extra sources and other sources that we are not really concerned about. The proposed method is first applied to a pre-determined signal to validate its effectiveness. Then the LFN data observed during the flood discharge of the Jin'anqiao hydropower station are analyzed and separated using this improved method. Two components, with dominant frequencies about 0.7 Hz and 0.95 Hz respectively, are successfully recognized as the actual acoustic sources induced by the flood discharge.


## 1. Introduction

Low frequency noise (LFN) with frequency ranging from 0.1 Hz to 20 Hz can travel very far and still be detected thousands of kilometers away from acoustic source. This is because it is difficult for the atmosphere to absorb the energy of LFN [1,2]. Many researchers have reported that houses and workshops near the LFN source have been found vibrating violently and that people who live around the LFN source suffered a variety of clinical symptoms [3,4]. However, past studies mainly concentrate on structure-borne LFN [5], such as infrasound generated by wind turbines [6,7], low-speed fans [8], multi-span bridges [9], and wind farms [10]. Very few literatures have mentioned the LFN induced by flood discharge in hydraulic engineering. Recently, LFN has been detected during flood discharge from many high dams in China, such as Xiang'jiaba [11], Jin'ping and Jin'anqiao hydropower station. Therefore, it is important to reveal the mechanism of the LFN induced by flood discharge. When we conduct prototype observation and model experiment to study the characteristic of the LFN sources, the recorded signal usually contains white noise components, colored noise components and information of LFN induced by flood discharge. These different components in the observed signal should be separated and the LFN induced by flood discharge should be identified and recognized. However, in practical measurement, only one or two sensors are usually used

* Corresponding author.
*E-mail address:* wangxiaoqun@tju.edu.cn (X. Wang).

https://doi.org/10.1016/j.ymssp.2018.03.030

to gather the original acoustic signals because the use of multiple channels is usually uneconomical. Therefore, a technique called single channel blind source separation (SCBSS), which can separate multiple sources from a single channel signal, becomes increasingly important.

Studies on blind source separation (BSS) have been developed intensively over the past decade. It has been applied in many fields, such as machine fault diagnosis [12,13], speech process [14] and damping estimation [15]. One of effective method of BSS is called independent component analysis (ICA), which is based on the assumption that all the sources are non-Gaussian and statistically independent. The ICA method can only solve the problems when the number of channels for recording is larger than the source number [16,17]. As for the SCBSS, the recorded single channel signal should be pre-processed by other techniques, such as Fourier transform, wavelet transform, phase space reconstruction [18] and empirical mode decomposition (EMD). The problem of separability of single channel signal with multiple sources was first studied by Hopgood and Rayner [19]. One idea of SCBSS is based on exploiting the inherent time structure of source signals by learning a priori sets of time-domain basis functions [20]. However, it is difficult to find proper basis functions because little was known about the sources in advance in practical applications. Another idea of SCBSS is to map the single-channel signal into multi-channels by the preprocessing techniques [21]; then sources contained in the multi-channel signals can be separated by ICA. Since the first combination of multi-channel mapping and ICA, many similar SCBSS methods have emerged and been studied. The single channel ICA (SCICA) [21], which combines Fourier transform and ICA, can only deal with stationary signals. The wavelet-ICA (WICA) [22], which decomposes single channel signal by wavelet transform and reconstructs a multi-channel signal as an input of ICA, can deal with no-stationary signal, but it involves the selection of proper wavelet basis function and the number of wavelet decomposition layers. It is difficult to find the most appropriate basis function as it greatly influences the performance of WICA.

Another version of SCBSS called EMD-ICA is a method which can deal with the nonlinear, nonstationary, even chaotic signals [23]. The empirical mode decomposition (EMD) can map the single channel signal into numbers of intrinsic mode functions (IMFs), and then several independent components are calculated by performing the ICA algorithm to the IMFs. The sources of interest have to be selected in this method. However, the characteristics of important sources are uncertain, which makes the selection difficult. Furthermore, the number of IMFs is usually larger than that of actual sources, which may deteriorate the result of separation. Another algorithm, EEMD-PCA-ICA (ensemble empirical mode decomposition-principal component analysis-independent component analysis) [24,25], which employs a more robust version of EMD called ensemble empirical mode decomposition (EEMD), makes use of the principal component analysis (PCA) to estimate the number of sources. However, the EEMD-based method usually have a problem called edge effect when using the EEMD algorithm, so an edge effect elimination method called extreme point symmetry extension (EPSE) has been introduced by Guo et al. [26] Furthermore, one has to interrupt the algorithm to determine the number of sources to be separated. It is inconvenient especially when a number of single channel signals are needed to deal with. In addition, the source number determined is not always appropriate. If the source number selected is larger than the actual number, the extra sources are illusive because they do not exist actually but are calculated by the algorithm. Also, for the LFN observation, there may be some colored sources that occur casually, which are not the sources we are concerned about. The existing algorithm has not provided a specific guidance about how to identify and eliminate the extra sources. And researchers in the past usually select the sources of interest directly. However, we have no priori-knowledge about the characteristics of the LFN sources induced by flood discharge, which makes the selection more difficult.

To overcome the problems of EEMD-PCA-ICA, we introduce a cross-correlation analysis procedure to eliminate the extra sources called ghost source (GS) and casually-occurred sources (CS). In the improved separation algorithm, the number of sources is determined directly by a threshold which is selected by pre-experiment. Then the separation process with no additional intervention is ready for further analysis to eliminate the GS and CS.

This study is organized as follows. In Section 2, the method of EEMD-PCA-ICA is briefly reviewed. To overcome the shortcomings of previous approaches, the source number estimation procedure in EEMD-PCA-ICA is simplified by the singular entropy (SE) method along with a threshold. In addition, a cross-correlation procedure to eliminate unimportant sources is introduced into the separation algorithm. In Section 3, a pre-determined experiment is conducted to validate the proposed method. In Section 4, the proposed and validated method is applied to analyze the LFN data observed during the flood discharge. Two source components are identified and recognized as the actual sources. It is also demonstrated that the proposed approach can be also used to determine the exact location of actual sources of the LFN induced by flood discharge. Conclusions are drawn in Section 5.

## 2. Methodology

### 2.1. Description of the problem and review of EEMD-PCA-ICA

Multiple acoustic sources are usually mixed in the single channel signal during the observation of the LFN induce by flood discharge. In this paper, it is assumed that the observed signals are the linear combination of all the acoustic sources:

$$\mathbf{x}(n) = \alpha_1\mathbf{s}_1(\mathrm{n}) + \alpha_2\mathbf{s}_2(\mathrm{n}) + \ldots + \alpha_i\mathbf{s}_i(n) + \ldots + \alpha_N\mathbf{s}_N(n) + \sigma\mathbf{n}(n) \quad (1)$$