Noise correction algorithm for nonlinear turbulent shear velocity time series based on energy spectrum of singular values

Xiuyan Liu\textsuperscript{a}, Xin Luan\textsuperscript{a}, Dalei Song\textsuperscript{b},\textsuperscript{*}

\textsuperscript{a}College of Information Science and Engineering, Ocean University of China, ShanDong 266100, P. R. China.  
\textsuperscript{b}College of Engineering, Ocean University of China, Ocean University of China, ShanDong 266100, P. R. China.

Abstract

Shear velocity time series are essential in characterizing ocean turbulent flows. The moored platform is mounted with two orthogonal shear probes (PNS06) to measure shear data for calculating velocity spectra. However, the shear probes are inevitably contaminated by instrument noise and the complex marine environments. In this paper, a method based on singular spectra decomposition was proposed to attenuate vibration noise by neglecting the higher-order modes in time-series reconstruction. First, this method constructed a Hankel matrix with shear velocity data, then decomposed and reconstructed the shear signals based on the method of conducting inverse singular value decomposition transformation on the values and their corresponding vectors to achieve the purpose of signal de-noising. The corrected spectra match well with the empirical Nasmyth spectrum and dissipation rates calculated from the noise-reduced shear spectra have dropped nearly one order of magnitude. The experimental results show that the proposed method provides an effective and straightforward approach for eliminating the noise signals in shear velocity spectra in ocean dynamics. ©2017 all rights reserved.

Keywords: Turbulent flows, shear spectra, singular values decomposition, noise correction, dissipation rates. 
2010 MSC: 05C12, 05C15, 05C76.

1. Introduction

Ocean turbulence is one of the most challenging fields in physical oceanography. The analysis of ocean turbulence will make contributions to improve our understanding of marine environments and strengthen the forecasting system of marine climate and disaster. Thus, it is important for us to study ocean turbulence mixing processes [4]. The shear velocity spectra measured by shear probes and the rate of turbulent kinetic energy (TKE) dissipation $\varepsilon (\text{Wkg}^{-1})$ are two important parameters to describe the ocean turbulent mixing theory. Generally, the micro-structure turbulence is measured by two orthogonal airfoil shear probes and a fast-response thermistor. However, shear probes usually work in tough marine environments, such as ship-induced movements, the vibrations of profiler and other complex interference factors, the measured turbulence signals contain a large amount of noisy energy which will affect the quality of the measured shear data. Therefore, it is particularly important to effectively eliminate the noise.
signals and obtain uncontaminated turbulence data before correctly analyzing the statistical characteristics of ocean micro-structure turbulence.

Generally, much attention has been paid to separate useful signals from mixed and noised time series during the past decades. Traditional de-noising algorithms, such as the Fourier transform spectral analysis [19], the low-pass Gaussian filter [5], the Wiener filter [11], and the wavelet transform analysis [1, 17] are proposed to minimize the noise signals. These mentioned traditional de-noising methods usually convert the signal from time domain to frequency domain based on fast Fourier transform. For example, a conventional Gaussian low-pass filter can only remove the noise above the cut-off frequency of the filter, and hence the spectra below the cut-off frequency still have lots of noisy signals. However, the noisy energy is distributed in a wide frequency band, not only in a high frequency, the residual noise data still has considerable energy when the shear velocity signals are transformed into the wavenumber domain, thus it cannot be effectively removed by traditional methods. Piera et al. [13] proposed a new time-frequency analysis method based on wavelet transform to identify turbulent patches and reduce the noise. This approach is highly advantageous over either Fourier transform or the classic low-pass filtering, especially when the time series are intermittent in nature. However, the wavelet-based methods are not adapted to be used when the time series are nonlinear or unevenly spaced. These time-variant methods have their limitations to nonlinear and non-stationary time series. An advanced singular spectrum analysis (SSA) method is essentially a principal components analysis in the time domain that extracts information from noisy signals without heavily relying on prior knowledge. Multiscale analysis methods, the proper orthogonal decomposition (POD) [2], singular value decomposition (SVD) [6], and principle component analysis (PCA, or empirical orthogonal functions) [8], are names of the SSA implementations in other disciplines. In this research, we apply an advanced SSA method to decompose the turbulent flow field data for reducing the noise and improving the accuracy of observed turbulence data. Low frequency components can identify the dominant features and represent global characteristics, and the high frequency components retain the local features. Thereby, the noise data is reduced by removing the energy spectrum in higher-order singular values.

The outline of this paper is as follows. In Section 2, we describe the basic configuration of the measuring platform for turbulence measurements. Section 3 introduces the advanced noise correction method based on singular values decomposition. Subsequently, the trial sea data is used to validate the feasibility and efficiency of the algorithm by analyzing the shear velocity spectra and calculating the dissipation rates. The conclusions remarks are finally presented in Section 5.

2. Experiments and database collection

A moored micro-structure turbulence measuring instrument (MTMI) has been continuously deployed in the South China Sea (SCS) at $21^\circ 9.90'\text{N}, 117^\circ 42.03'\text{E}$ from October 19, 2013 to February 10, 2014. The whole mooring system [10] can be divided into four parts (Figure 1). In part I, groups of glass floats can be used as the upper buoyancy element, providing buoyancy to integrate the system at the desired depth in a mooring line. The CTD system mainly records the conductivity, water depth and real-time water temperature. The MTMI instrument is equipped with two orthogonal shear probes (PNS06) for measuring micro-structure turbulent velocity spectra ($\partial w/\partial x$ and $\partial v/\partial x$), an attitude sensor to measure motion behavior of platform (heading, pitch, and roll) and three-axis accelerations ($A_x, A_y, A_z$). Part II contains the Aanderaa Current Meter (RCM11) which mainly records the water flow state such as averaged current speed, the absolute current speed, and the flow direction, etc. The part III and part IV are acoustic release and anchor block, respectively, the former is utilized to recover the instrument and the latter provides gravity for the whole system to make the instrument keeping a vertical state. The platform is located at approximately 250-m depth and several groups of glass floats are used to provide buoyancy and keep the submerged buoy straight together with the anchor at the bottom.

In this experiment, the MTMI instrument has continuously collected turbulence data for up to 115 days from October 19, 2013 to February 10, 2014. Exclude some incomplete and sporadic data in 18
days, there are totally 97 days which has the complete data samples. Such long time-series turbulence dataset are of interest and valuable for oceanographic researchers to explore the characteristics of ocean turbulence.

![Figure 1: The sketch of the moored platform.](image)

3. Methodology

The singular value decomposition (SVD) method is a mathematical method which has been widely used in many fields, such as meteorology, physics, economics, financial mathematics, and life sciences. It especially provides a special insight into dynamics of the underlying system. Singular spectrum analysis (SSA) is the most remarkable approach to obtain signals to noise separation by decomposing time series into small independent components. The detailed steps of this decomposition algorithm are described as follows [15].

1. First, we assume that a given random signal sequence \( X = \{x_k, k = 1, 2, \ldots, N\} \) is represented as,

\[
X_k = S_k + W_k, \quad k = 1, 2, \ldots, N,
\]

where \( S_k \) is the true signals, \( W_k \) is the noisy signals, and \( N \) is the length of time series.

2. Second, calculate the autocorrelation function \( R_{XX} \) and determine the window length \( M \), then construct the covariance matrix \( C_X \) with \( R_{XX} \),

\[
R_{XX}(\tau) = \sum_{i=1}^{N-M+1} (X_{i+\tau} - \bar{X})(X_i - \bar{X})(\tau = 0, 1, \ldots, M - 1),
\]

where \( \bar{X} \) is the mean of time series and the symbol \( \tau \) refers to the time lag. The Toeplitz matrix \( C_X \) can be calculated directly by using \( C_X = \text{Toeplitz}(R_{XX}) \) in Matlab, where

\[
C_X = \frac{1}{N - |i - j|} \sum_{i=1}^{N-|i-j|} (X_i X_{i+|i-j|}).
\]
The window length is a key to extract noise from the mixed signal. If the parameter M is not properly selected, the source signals could not be extracted from the mixed signal. Thus, we use the method proposed by Wang et al. [18] to select the window length. In general, the window length M is determined by the maximum value of time lag of autocorrelation and the dimension of Toeplitz matrix.

(3) Perform SVD on the Toeplitz matrix $C_X$ to get the eigenvalues and extract the eigenvalues from the diagonal matrix $\Sigma$,

$$C_X V^k = \Sigma V^k,$$

where $C_X \in \mathbb{R}^{M \times M}$, $V \in \mathbb{R}^{N \times N}$ are orthogonal matrices, and $\Sigma$ is a diagonal matrix with size of $M \times N$, which is called singular values of matrix $C_X$. The extracted eigenvalues where $\sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ are non-negative and arranged in a descending order ($\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M \geq 0$). The corresponding eigenvectors of this matrix are called the empirical orthogonal functions (EOFs) matrix in the meteorological literature. The contribution rate $\Delta \lambda$ is used to estimate the energy contribution of every eigenvalues and find number of principal components. That is,

$$\Delta \lambda = \frac{\sigma_k}{\sum_{i=1}^{M} \sigma_i} (k = 1, 2, \ldots, M).$$

To determine the numbers of components of SVD, we set a threshold to select the principal components to reconstruct the original signal. When the cumulative contribution rate $S_\lambda = \sum_{i=1}^{p} \Delta \lambda_i (p < M)$ value exceeds the threshold, the value of $p$ is the numbers of components we want to find. In this paper, the threshold is equal to 0.9, which is enough to select the different contributions for all eigenvalues.

(4) Transform the time series into a Hankel (trajectory) matrix $H_{M \times N}$, and divide the matrix into two subsets, principal components subset ($S$) and noise subset ($W$),

$$H_{M \times N} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ x_2 & x_3 & \cdots & x_{N+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+1} & \cdots & x_{M+N-1} \end{bmatrix} = S + W,$$

where $S$ and $W$ are the trajectory matrices in reconstruction phase space corresponding to the true and noisy signal respectively.

(5) Project $H$ onto the rearranged matrices of eigenvalues $V$, where $W = HV$ and yield the corresponding principal components (PCs) $P_k$:

$$P_k(t) = \sum_{j=1}^{p} x(t+j-1)V(j)^k.$$

(6) Finally, apply the inverse Hankelization process by anti-diagonal averaging and reconstruct the original series $X$ with the selected principal components.

In short, a trajectory matrix is constructed according to the given time series embedded into the delay coordinate phase space, and principal signals are extracted from the original mixed time series by decomposing and reconstructing this matrix, then extract global features and analyze characteristics of the signals, such as the long-term trend signal, the periodic signal, and noisy signal. This algorithm is suitable for analyzing and separating partial information from mixed signals.

4. Results

4.1. Shear velocity series

Shear velocity spectra measured by shear probes is one of the most important parameters to study characteristic of ocean turbulence. The sampling frequency of the shear probes is set to 1024Hz. As the
probes move horizontally through the water at flow speed $U$, the vertical component of the cross-stream velocity generates a lift force on the probe that causes the piezo-ceramic beam to bend microscopically, the shear force $F_p$ is sensed by the shear probe and the piezoelectric ceramics makes the force into charge signal $E_p$ [9]. Then, $E_p$ is converted into digital signals by an analog-digital converter (A/D) after 11 times amplification by preamplifier and then filtered by a low-pass filter (130Hz). The digital signal ($E_p$) is transported to the pre-amplifier by lead wires and is converted into space sequence to obtain vertical shear velocity ($\partial w/\partial x$) by applying Taylor’s frozen field hypothesis theorem ($\partial / \partial x = U^{-1} \partial / \partial t$) [14].

\[
E_p = 2\sqrt{2}SUw, \\
E_s = \frac{\partial E_p}{\partial t} = 2\sqrt{2}SU^2 \frac{\partial w}{\partial x},
\]

where $S$ is the sensitivity of shear probes, $U$ is the horizontal axial velocity of probe, $w$ is the cross-stream velocity, $E_p$ is the output voltage of shear probes, $E_s$ is the output of an analog differentiator, and the numeric factor $2\sqrt{2}$ is an artifact calibration method based on the laboratory results.

Time series of turbulence recorded by two orthogonal probes are presented in Figure 2. The data are sampled in 180s and the velocity $U$ is 0.25 m s$^{-1}$. The voltages are changed into shear signals using the Taylor frozen theorem. The two calculated shear velocity spectra ($\partial w/\partial x$ and $\partial v/\partial x$) present intermittent nature and intermittent nature in component of $\partial v/\partial x$ is relatively less energetic (Figure 2). Shear probe signal from probe #2 ($\partial v/\partial x$) is very identical with the shear data recorded by probe #1 ($\partial w/\partial x$), indicating strong consistency for measuring isotropic turbulence. The unprocessed shear signals contain the contamination from platform motions which obscures the environmental signal in temporal space, thus shear data are sharply low-pass filtered at 130 Hz to remove high frequency signals.

![Figure 2: Shear variance signals measured by two orthogonal shear probes. The $\partial w/\partial x$ is sensed by probe #1 and $\partial v/\partial x$ is measured by probe #2. Shear signals are low-pass filtered at 130Hz to remove the contamination signals at high frequency.](image)

4.2. Singular spectrum

The advanced singular spectrum analysis method (SSA) is applied to attenuate and identify the vibration noise and improve the quality of the measured turbulence shear data. The turbulence shear signals are embedded into a time series, and then this method constructs a Hankel matrix of shear data and
decomposes this matrix into \( M \) dimension, where \( M \) is the window length. Finally a low-order reconstruction is implemented. The relevant results of detailed implementation are present here for \( w \) shear velocity component.

The decomposed eigenvalues are given in decreasing order (Figure 3, left panel). There is a distinct grouping of the first two eigenvalues where their cumulative contribution rates distribution are larger than 0.8 (Figure 3, right panel), and the accumulation of the other three eigenvalues is less than 0.15. It is apparent that the first five eigenvalues are a distinct grouping which followed by a steep slope and the residual eigenvalues form a mildly sloping and flattening out ‘tail’ of the singular value spectrum. The cumulative contribution rates of the first five eigenvalues are up to 0.95 which indicates that the five singular values will contain most of energy and represent the global characteristics. Therefore, we will reconstruct shear velocity spectra by using lower order singular values. A similar behavior is also observed for the \( v \) component of shear velocity (not shown here).

![Figure 3: Singular values spectrum of the shear variance (\( \partial w/\partial x \)) time series where the embedding dimension \( M \) is 97 and the shear component \( \partial w/\partial x \) is similar with \( \partial v/\partial x \) (not shown). The eigenvalues are plotted, as usual, in decreasing order. The symbol \( \Delta \lambda \) is the proportion distribution for all eigenvalues.](image)

According to the eigenvalues calculated by applying the proposed algorithm to the Toeplitz matrix, the first five empirical orthogonal functions (EOFs) corresponding to the first five singular eigenvalues are shown in Figure 4. The first two leading pairs of EOFs (EOFs1 and EOF2) are in quadrature which corresponds to the first two eigenvalues that are approximately equal and whose error bars are overlapped (Figure 4 (a)). The other three EOFs are almost periodic and orthogonal thus it corresponds to the oscillation behavior of the signals (Figure 4 (b)). The EOFs analysis of the data suggested that the flows exhibited a high vertical mode structure, creating strong vertical shear that could easily trigger turbulence. The small scale shear will improve our understanding of the ocean evolution processes of multi-scale turbulence.

The SSA technique provides a multi-scale analysis for turbulence signals with several functions which is reasonable, because there are not many possible structures at scales that approach the sampling timescale. This method also provides the simultaneous analysis by many different wavelet analysis functions at large scales turbulence which reflects the large complexity of the structures. SSA can thus provide sharper regime transitions in the evolution of a nonlinear system than traditional method of wavelet analysis which uses a fixed set of basic functions.
4.3. Corrected shear spectra

In the constructed process, we perform a low-order reconstruction by using the first five components, and the higher order components may be associated with noise, thus we filter and correct shear spectra by neglecting the higher-order components which captured noise energy. The first five leading principal components (PCs) corresponding to the EOFs are also in quadrature and periodic, and they strongly suggest periodic variability at different periods. The partial reconstruction of long-term time series is done by summing the variability of first five PCs associated with the leading pairs of singular values. The reconstructed data contains 43% of the total variance which is not including the large vibrations energy.

As is shown in Figure 5, the decontaminated time series also preserve the intermittent and cascade nature of the turbulence series in scale. We eliminate the vibrations noise by reconstructing the signals based on the first five leading principal components, the original shear (Figure 5, black line) fluctuate apparently larger than the reconstructed signals (Figure 5, red line). Importantly, the reconstructed time series present the properties of the original time series over every scale. The agreement between the original shear data and the reconstructed shear spectra is excellent, most of the vibration noise is removed by neglecting the higher components and the accuracy of shear turbulence data is surely improved a lot by reconstructing the original time series based on the PCs.

The normal distribution and the probability density function (PDF) are given to further validate the effectiveness of the corrected shear spectra after using the proposed de-noising algorithm. Townsend’s results [16] showed that the structure of the turbulent shear flow was in normal distribution (or Gaussian distribution). The Gaussian distribution and probability density distribution of the valid data is shown in Figure 6. It is obvious that both the cleaned data conform to the Gaussian distribution which is in accordance with the Townsend’s results. In short, the corrected shear data is effective and can be used to
analyze the characteristics of turbulence.

![Figure 5: Raw turbulence shear time series (black) and the corrected shear signals (red) based on singular values construction method.](image)

Figure 5: Raw turbulence shear time series (black) and the corrected shear signals (red) based on singular values construction method.

![Figure 6: Probability density distribution for normal distribution (Gaussian distribution) calculated from the corrected shear data for two shear probes.](image)

Figure 6: Probability density distribution for normal distribution (Gaussian distribution) calculated from the corrected shear data for two shear probes.

4.4. Dissipation rates

The dissipation rates ($\varepsilon$) of turbulent kinetic energy (TKE) is another primary parameter to describe characteristics of turbulence. The MTMI uses two shear probes to obtain high-resolution of turbulent quantities, especially the dissipation rates of TKE which provides critical information about the magnitude
of the fluctuations associated with the turbulent flow. The dissipation rates are calculated from cleaned shear spectra by using [3, 7],

\[ \varepsilon = 7.5 \nu (\frac{\partial w}{\partial x})^2 = 7.5 \nu \int_{k_0}^{k_c} \Phi(k) dk, \]

where dissipation rate of TKE is marked as \( \varepsilon \), and \( \nu = 1.64 \times 10^{-6} \text{m}^2\text{s}^{-1} \) is the kinematic viscosity coefficient. The \( k_0 \) is a lower integration limit and \( k_c \) is the Kolmogorov wavenumber where \( k_c = (\varepsilon/\nu^3)^{1/4} \). Generally, the Nasmyth (1970) [12] theoretical spectrum is considered as a criterion to evaluate the other spectrum, and the model spectrum \( \Phi_{\text{Nas}} \) is defined as,

\[ \Phi_{\text{Nas}}(k) = \frac{8.05(k/k_c)}{1 + (20 \times k/k_c)^{3.5}}. \]

The original and cleaned spectra in comparison with their corresponding Nasmyth spectra are shown in Figure 7. It is noted that the original shear spectrum (the black solid line) has two apparent peaks induced by the vibration contamination of the platform and these two vibration noise are obviously removed in the cleaned spectrum (the red solid line) using the proposed method. In comparison with the standard Nasmyth model before the Kolmogorov wavenumber \( k_c \) (Figure 7, two vertical dashed lines), the corrected power spectrum matches well with the Nasmyth spectrum and the dissipation rates \( \varepsilon \) are decreased from \( 3.83 \times 10^{-5} \text{Wkg}^{-1} \) to \( 2.39 \times 10^{-6} \text{Wkg}^{-1} \). The dissipation rates of TKE are computed from the raw shear velocity spectra where the 130Hz low pass filter is applied to remove high-frequency noise and the variation is approximately ranged from \( 10^{-6} \text{Wkg}^{-1} \) to \( 10^{-7} \text{Wkg}^{-1} \) (Figure 8, the red solid circles). Note that the corrected dissipation rates calculated from the reconstructed shear data have significantly decreased more than one order of magnitude (Figure 8, the gray solid circles), indicating that the proposed method is effective to remove noisy energy and the shear spectra are fully resolved.

![Figure 7: Original (black solid line) and corrected spectra (red solid line) in comparison with their corresponding empirical Nasmyth spectrum (dashed lines). Two vertical dashed lines denote the cutoff wavenumber.](image-url)
Figure 8: Turbulent kinetic energy dissipation rates calculated from the raw shear spectra measured by probe #1 (red solid circles) and the corrected shear spectra (gray solid circles). Identification holds on shear probe #2 (not shown).

5. Conclusions

The moored turbulence measuring instrument (MTMI) was deployed at 21°09.90′N, 117°42.03′E in the South China Sea (SCS). The long time series of shear velocity data has collected for up to 115 days during from October 19, 2013 to February 10, 2014. To improve the accuracy of the measured shear velocity data, an advanced noise reduction method based on energy spectrum of singular values is proposed. The experimental data show that the vibration noise can be effectively removed by decomposing and reconstructing the shear data and improve the quality of shear velocity data. Turbulent dissipation rates calculated from corrected shear spectra will make contributions to improve our understanding of ocean turbulence evolution processes in the South China Sea. However, this noise correction method still needs to be optimized owing to the tough marine experiments and instrument noise.

Acknowledgment

This work is financially supported by the National High Technology Research and Development Program of China (“863 Program”) under Grant No.2012AA090901 and No. 2014AA093404, the National Nature Science Foundation under Grant No.61203070. We would also like to thank the Ocean University of China for the platform of the vessel “Dong Fang Hong 2”, and thank the whole teams of the research and development for their help of this experiment.

References

[14] H. R. Rahai, J. C. LaRue, Assessment of Taylor hypothesis and local isotropy due to strain applied to a nearly homogeneous and isotropic flow, Exp. Fluids, 26 (1999), 136–144. 4.1