Journal of Nonlinear Science and Applications

Applications of discriminant analysis on shear turbulence data in wavenumber domain

Yongfang Wang[a], Xin Luan[b], Tongxing Li[a,*]

[a] School of Informatics, Linyi University, Linyi, Shandong 276005, P. R. China.
[b] College of Information Science & Engineering, Ocean University of China, Qingdao, Shandong 266100, P. R. China.

Communicated by Martin Bohner

Dedicated to Professor Ravi P. Agarwal

Abstract

This paper proposed a discriminant analysis method to realize the auto matching of shear spectra and improve the precision of the shear turbulence data. The discriminant analysis method includes two parts, firstly, in order to eliminate noise data, cross validation method is used to data preprocessing, and secondly, maximum likelihood method is used to get discriminant function to realize the auto matching of the spectra. South China Sea experiment is used to verify the validity of the method. ©2015 All rights reserved.

Keywords: Discriminant analysis, cross validation, data preprocessing, maximum likelihood.
2010 MSC: 62H30, 68T10.

1. Introduction

Discriminant analysis is a multivariate statistical analysis method, and according to different criteria, the commonly used discriminant analysis methods [6] include distance discriminant [5], Fisher discriminant [7], and Bayes discriminant [1], etc. In essence, it uses sample data to get discriminant function based on a certain criterion, and especially for large amounts of sample data, the discriminant analysis method plays an important role in research.

*Corresponding author

Email addresses: greenworld6@163.com (Yongfang Wang), luanxin@qingdao.gov.cn (Xin Luan), litongx2007@163.com (Tongxing Li)

In the research of ocean turbulence, the observation data is the basis to analyze its characteristics [10]. In recent years, the analysis of turbulence data has been one of hot issues. Dong and Dosso [2] applied Bayesian inversion to estimate seabed shear-wave speed profiles and their uncertainties from interface-wave dispersion data. Piera et al. [11] proposed that using wavelet to identify the turbulence and calculate the Thorpe scale. Xie et al. [15] found the anomalous change of the spectra in SCSMEX, and through analyzing the shear spectra, they developed new view, new thinking, and new approaches to study the PSI mechanism. All of the results are based on the observation turbulence data, which contains abundant information resources.

When the observed turbulence data is processed on computer, noise data will influence its accuracy, and in wavenumber domain, the realistic shear spectra which are calculated from the observed turbulence data will be compared with the Nasmyth empirical spectra [8] in dissipation subrange. Generally artificial recognition method is used to evaluate the spectral quality, which is subjectivity and imprecise. Therefore, the paper proposed a discriminant analysis method to improve the precision of the observed turbulence data and realize the spectra auto matching. Section 2 introduces the wavenumber spectra in dissipation subrange. In Section 3, the discriminant analysis method is introduced. Firstly, cross validation [12] is used to get the valid data, and then maximum likelihood method [3] is used to get the discriminant function. Section 4 introduces the result which is used to illustrate the validity of the method, and Section 5 makes a conclusion about the method.

2. Wavenumber Spectra

Ocean turbulence data is acquired by sensor, and through electronic circuit system, it is converted into digital signal, then through data processing technology [9], digital signal is changed into corresponding physical parameters in wavenumber domain. Wavenumber spectra is one of the physical parameter, which plays an important role to evaluate the characteristics of turbulence [13].

The shear spectra in wavenumber domain is shown in Figure 1, x axis represents wavenumber, y axis represents power density spectrum (PSD). The black line is the realistic shear spectra, and the blue line is the Nasmyth empirical spectra, the red dotted line is the Kolmogorov cut-off wavenumber [4], it is defined as the reciprocal of Kolmogorov micro-scale, and the data after the red dotted line is not considered in the method.
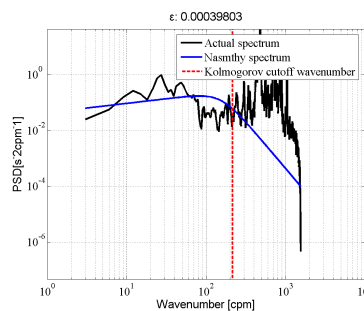

Figure 1: the wavenumber spectra

Figure 1 shows that noise energy will affect matching precision of the spectra, such as wavenumber of 30, the PSD reaches 100, which is obviously due to the influence of noise data. Ideally, the realistic shear spectra should fluctuate in a small range around Nasmyth empirical spectra, nevertheless in the process of observation, noise is inevitable, so 10-fold cross-validation method is use to data preprocessing to eliminate the noise data. In general, 10-fold cross-validation is a common test method to test the algorithm accuracy, but in this paper it used to test the observed turbulence data category to get valid data. After data pretreatment, based on the valid data in line with normal distribution [14], the maximum likelihood method is used to estimate the characteristic parameters of the turbulence data, and then the quadratic discriminant function is got according to the characteristic parameters. Finally, the effective fluctuation

range is obtained by the discriminant function to realize the auto matching of the shear spectra and the noise signal is eliminated to improve the data accuracy.

3. Discriminant Analysis Method

Since noise signal will change the distribution characteristics of turbulence, cross validation method is firstly used to eliminate the noise data, its purpose is to do data preprocessing for the discriminant function in the next step.

3.1. Data Preprocessing

Cross validation method is a statistical analysis method, it divides the data set into two complementary subsets: training set and testing set, the training set is used to modeling, and the testing set is used to prediction. The commonly used cross validation methods are holdout, k-fold and leave one out cross validation. The flow chat of k-fold is shown in Figure 2.
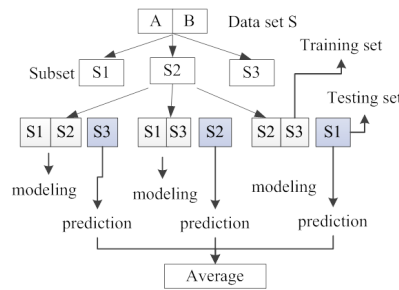

Figure 2: k-fold cross validation (S={S1, S2, S3})

In the paper, 10-fold cross validation is used to eliminate the noise data. The shear turbulence data S is divided into 10 sets, 9 sets are used as training set: Sm, and the least one is used as testing set: Sn. The class mid-value V_{mid} and its corresponding probability value P_{mid} of the training set are used to verify the testing set. If the data in the testing set contains in the two values, then it will be marked as 1, and after 10-fold cross validation, if the mark numbers of the data are less than 5, it will be regarded as noise data and then eliminated. The flow chart of 10-fold is shown in Figure 3 (Assume $A = \{1, 2, 3, \ldots, 10\} = \{x_1, x_2, \ldots, x_{10}\}$. If $n = x_i$, $i \in A$, then $m \in A - \{x_i\}$).
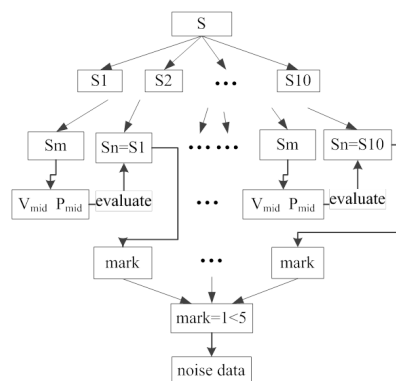

Figure 3: 10-fold cross validation (S={Sm, Sn})

3.2. Getting Discriminant Function

Since each observation of ocean turbulence, the observed data is large, we are looking for an effective method to realize the shear spectra auto matching and improve the efficiency of the data processing. After data preprocessing, according to the characteristic that turbulence data conforms to normal distribution, the maximum likelihood method is used to get the parameters of the probability density function, the parameters makes the probability that sample data appears maximum, and finally the discriminant function is got by the parameters. Generally maximum likelihood is analyzed under two different conditions: discrete random variables and continuous random variables. Here we use the discrete random variables condition.

The probability distribution of shear turbulence data S is $P\{S = s\} = p(s; \theta_1, \theta_2, \ldots, \theta_k)$, and the probability distribution of $S_1, S_2, \ldots, S_n$ is

$$P\{S_1 = s_1, \ldots, S_n = s_n\} = \sum_{i=1}^{n} p(s_i; \theta_1, \theta_2, \ldots, \theta_k).$$

Then the likelihood function is

$$L(\theta_1, \theta_2, \ldots, \theta_k) = \prod_{i=1}^{n} p(s_i; \theta_1, \theta_2, \ldots, \theta_k).$$

It indicates that the possibility of the data appears, and θ is the parameter of the probability density function.

Since parameter θ is unknown, maximum likelihood is used to find an optimal parameter θ to maximizing the probability, and then the quadratic discriminant function of shear turbulence data is constructed by parameter θ. The process of the maximum likelihood is shown as follows:

(1) Firstly, we get the likelihood function of valid data through mean value μ and variance value Σ, the likelihood function is

$$L(s_1, s_2, \ldots, s_n | \theta) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (s_i - \mu)^{\mathrm{T}} \Sigma^{-1} (s_i - \mu) \right\}.$$

(2) Secondly, through

$$\frac{\partial |A|}{\partial A} = [\mathrm{adj}(A)]^{\mathrm{T}} = |A| (A^{-1})^{\mathrm{T}}, \quad \frac{\partial \log(L)}{\partial \mu} = 0, \quad \text{and} \quad \frac{\partial \log(L)}{\partial \Sigma} = 0,$$

the maximum likelihood estimation of the parameters m and $\hat{\Sigma}$ are got.

(3) Thirdly, the discriminant function is

$$p(s_{\mathrm{valid}}) = \frac{1}{(2\pi)^{1/2} |\hat{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (s - m)^{\mathrm{T}} \hat{\Sigma}^{-1} (s - m) \right\}.$$

Finally, the discriminant function is got to realize the auto matching of the spectra, and then the South China Sea observation experiment is applied to evaluate the effectiveness of the progress.

4. Results

The ocean turbulence observation instrument is Vertical Microstructure Profiler (VMP), the falling depth in the ocean is 146m, and the average sinking velocity is 0.73m/s. Through computer graphics processing, these signals are change into graphics, which are shown in Figures 4 and 5.

Sampling time is 240s and sampling frequency is 1024Hz. After used the cross validation method, the normal distribution fitting of the valid data is shown in Figure 6, it indicates that the valid data is in line with Gaussian distribution, which coincides with Townsend's result.

Then discriminant function is used to realize the shear spectra auto matching, which is shown in Figure 7, we can see that at the frequencies of the fluctuation (such as 30Hz, 40Hz), the peaks are apparently reduced.
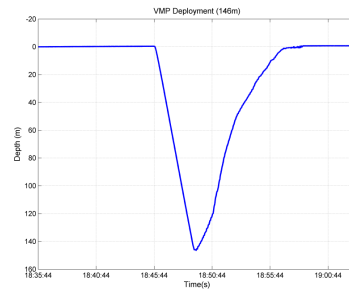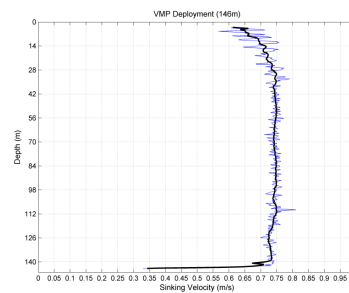
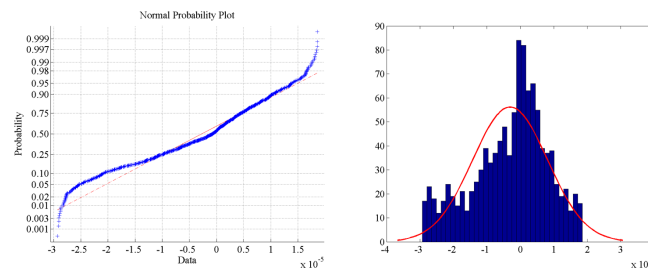Figure 4: Falling depth

Figure 5: Sinking velocity
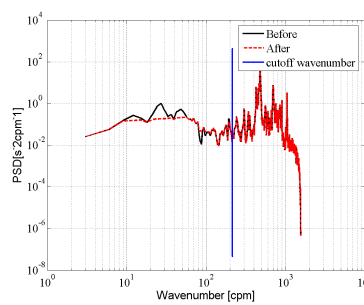

Figure 6: Normal distribution fitting

Figure 7: the change of the wavenumber spectra before and after

5. Conclusion

High accuracy of the observation data plays an important role in ocean turbulence theory research. In the paper, we find an effectively discriminant analysis method to study the characteristic of the ocean turbulence and provides some new way to process the shear turbulence data, and finally the results indicate that the method is feasible and effective. However, due to the limitations of the observed turbulence data, the method still needs to be further optimized to fit the complexity of the marine environment, and then,

most importantly, the information contained in the observed turbulence data needs to be further mining.

References

[1] George E. P. Box, George C. Tiao, *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, Inc., (1992). 1

[2] H. Dong, S. E. Dosso, *Bayesian inversion of interface-wave dispersion for seabed shear-wave speed profiles*, IEEE Journal of Oceanic Engineering, **36** (2011), 1–11. 1

[3] F. Ince, *Maximum likelihood classification, optimal or problematic? A comparison with the nearest neighbour classification*, International Journal of Remote Sensing, **12** (1987), 1829–1838. 1

[4] A. N. Kolmogorov, *Dissipation of energy in the locally isotropic turbulence*, Proceedings of the USSR Academy of Sciences, **32** (1941), 16–18. (Russian). 2

[5] J. Liang, S. Yang, A. Winstanley, *Invariant optimal feature selection: A distance discriminant and feature ranking based solution*, Pattern Recognition, **41** (2008), 1429–1439. 1

[6] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, Inc., (1992). 1

[7] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, K.-R. Müller, *Fisher discriminant analysis with kernels*, Neural networks for signal processing IX, Processing of the 1999 IEEE Signal Processing Society Workshop, (1999), 41–48. 1

[8] P. W. Nasmyth, *Ocean turbulence*, Ph. D. Thesis, University of British Columbia, Vancouver, (1970). 1

[9] N. S. Oakey, *Determination of the rate of dissipation of turbulent energy from simultaneous temperature and velocity shear microstructure measurements*, Journal of Physical Oceanography, **12** (1982), 256–271. 2

[10] J. Piera, *Signal processing of microstructure profiles: integrating turbulent spatial scales in aquatic ecological modeling*, University of Girona, (2001). 1

[11] J. Piera, E. Roget, J. Catalan, *Turbulent patch identification in microstructure profiles: A method based on wavelet denoising and Thorpe displacement analysis*, Journal of Atmospheric and Oceanic Technology, **19** (2002), 1390–1402. 1

[12] J. Shao, *Linear model selection by cross-validation*, J. Amer. Statist. Assoc., **88** (1993), 486–494. 1

[13] G. I. Taylor, *The spectrum of turbulence*, Proceedings of the Royal Society of London. Series A–Mathematical and Physical Sciences, **164** (1938), 476–490. 2

[14] A. A. Townsend, *The Structure of Turbulent Shear Flow*, Cambridge University Press, Cambridge, (1980). 2

[15] X.-H. Xie, X.-D. Shang, G.-Y. Shang, L. Sun, *Variations of diurnal and inertial spectral peaks near the bi-diurnal critical latitude*, Geophysical Research Letters, **36** (2009), L02606. 1