International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Discussion on the Development of Algorithm for Despiking ADV Data

Huang Yafei

River and Ocean Engineering school, Chongqing Jiaotong University, Chongqing 400074, China

Abstract: ***Normally, ADV data is contaminated by high turbulent intensities, preset velocity range and bed river boundaries, which needs some techniques to solve it. In this paper, a mount of existing methods ,which are Kernel density-based algorithm, wavelet-based algorithm and Phase-Space Thresholding algorithm, are discussed and apply it to two sorts of data from laboratory experiments and field experiments, respectively.***

Keywords: ADV; Spikes; Kernel density-based; Wavelet-based; Phase-Space Thresholding;

1. Introduction

Velocity of measurement using the ADV equipment is significantly important in hydraulic engineering. Accurate velocities is fundamental research of the hydraulic subject. The US Army Engineer Waterways Experiment Station developed the Acoustic Doppler Velocimeters to provide an instrument capable of giving instantaneous velocity measurements in three dimensions[1]. Basically, there is no flawless in ADV data set since it's born due to easily affected by Doppler noise and other reasons. The introduction of spikes of invalid data into the recorded velocity time-series inevitable, with resulting errors in the calculated turbulence characteristics[1].Therefore, many studies are proposed by some experts involved in this. At the beginning of these problems occurring, Phase-Space Thresholding algorithm (PST) has been in the mainstream and quite popular for despiking ADV data. Following that, numerous revised methods based on PST has sprung up in recent years. Besides, Razaz and Kawanisi[2] come up with wavelet-based techniques to detect spikes in combination with linear time series model to replace the gaps after removing spikes. Islam's[3] despiking method is linked to Kernel density-based algorithm. They all work differently. In this paper, we discussed them.

2. Data Resource

We analyzed two sorts of ADV sequences to compare difference between despiking methods applied to do data in this study. One of them was obtained from lab which is uncontaminated data due to no distracting factors. We also have data which was got in 2012 from a field experiment in the Three Gorges. This set of data that contains many large amplitude spikes is considered a consequence of high turbulent flows, bed areas, surf zones and downstream of flow steps[4].Also, we created a third data set by contaminating the clean data obtained in the laboratory experiment through interpolation of 50 large spikes into it at regular intervals. Modern ADV are able to sampling of at a rate of up to 200 Hz. In our research, we set the frequency of data set as 64 Hz in both experiments which include 20000 sampling in the lab mounts and 65300 sampling in the field, respectively. As figure.1., there are different data we analyzed in this paper.

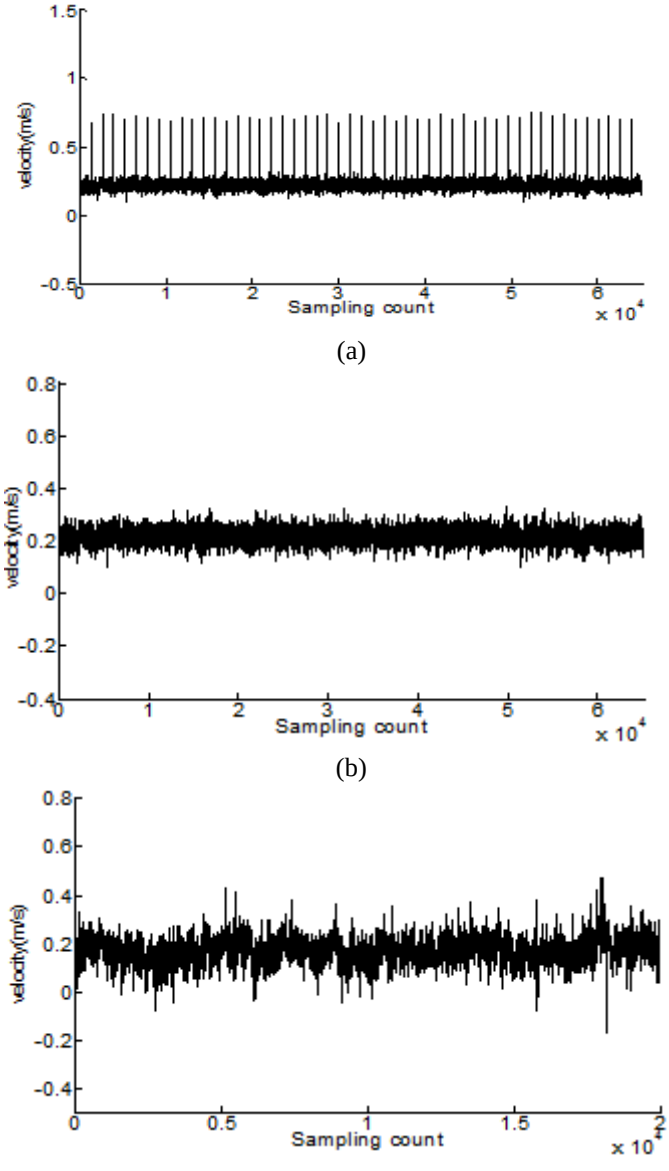


Figure 1: (a)contaminated data set from the lab (b)noncontaminated data set (c)origin data set from field

experiment

3. Despiking Methods

All techniques include two parts that are identification then removal and replacement. There are 3 ways discussed below in this paper.

1. The Modified version of PST

Goring and Nikora[5] firstly created PST algorithm and many revised methods have come along it in the past decade. A effective modified version proposed by Parsheh[6]are discussed in the paper. The basic principle are different from it linked to three aspects in the process of identification and a totally different replacing technique. Firstly, Parsheh applied median absolute deviation, analogous to the standard deviation, which is the fruit of Wal's research[7],instead of standard deviation. Secondly, A formula is used to control some valid points and flag it as unchangeable points as result of the original PST algorithm which is the defect of this technique. Removal of spike of a large magnitude can help reduce the number of iteration in the last step After all procedures run, the techniques use the PST ellipsoid to identify spikes and the last valid points to replace it

2. Wavelet-based algorithm

Application of Wavelet algorithm has existed decades ago. Here, we discussed the technique by Razaz[2].Differently and more importantly, nearly all despiking techniques focus on the identification ,but Razaz's techniques pay more attention on replacement of spikes which significantly improve this method. This technique developed a consistent threshold by thresholding the wavelet basis in wavelet space and transforming ADV data into wavelet space. Besides, a more robust scale estimator ϑ_n [8]is introduced into the technique in combination of linear time series modeling or a Kalman predictor which totally different from interpolation and extrapolation and so on ,which many experts used. Linear are typically applied to auto correlated time series data that are fitted to time series data rather to better understand the data or perhaps to predict future points in the series which is the need of this technique. The principle of Kalman filter doesn't differ from it greatly.

3. Kernel density-based algorithm

The Kernel density estimate amplifies the special points, which is the very spikes, to identify the spikes in the Kernel density plot and use interpolation to replace those eliminated by identification which is a statistical method.

4. Results and Discussion

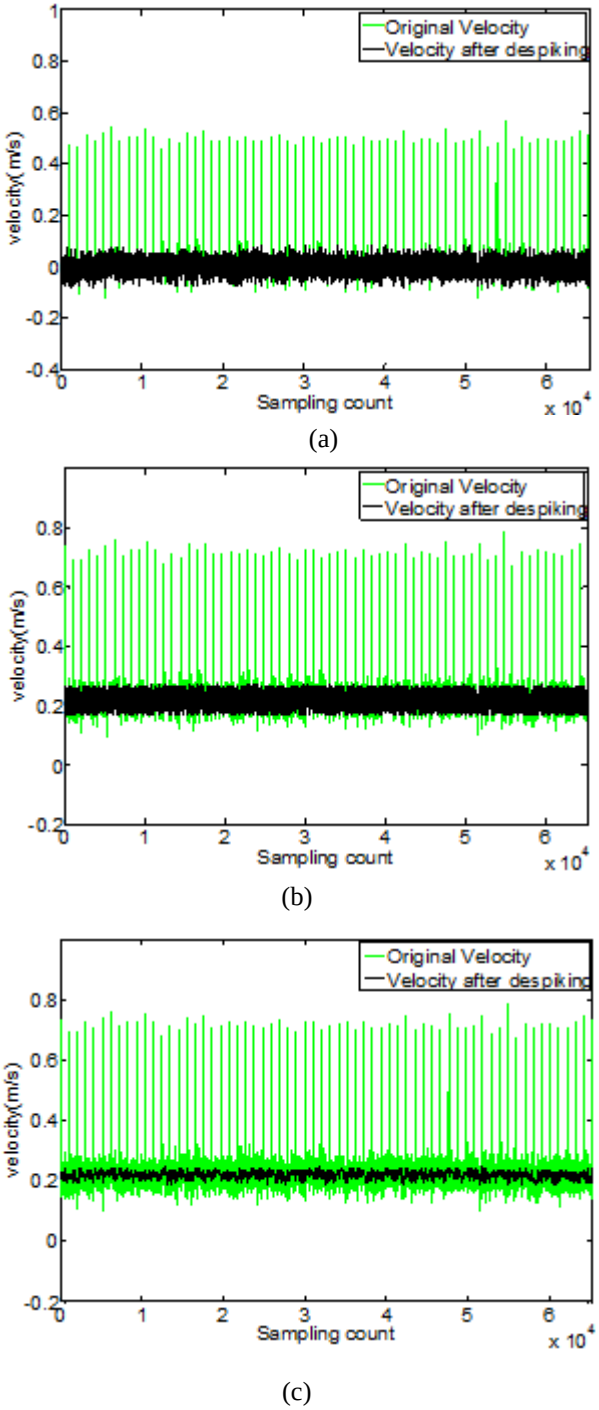
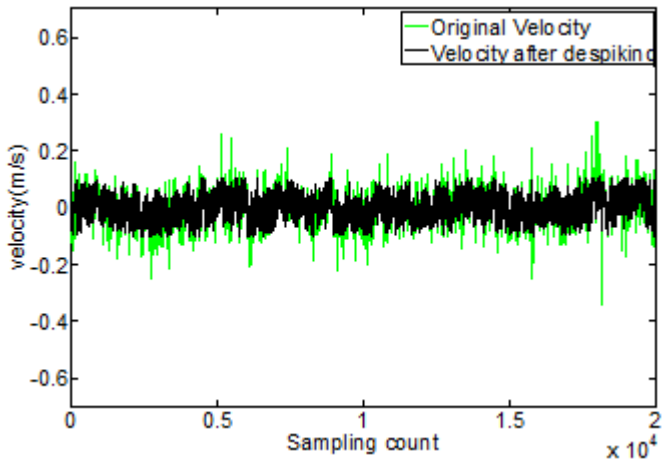
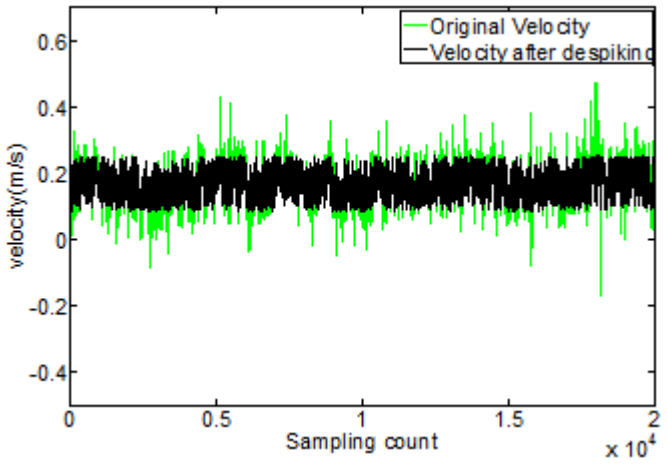


Figure 2: Comparison before and after despiking ADV data from the lab by different ways (a) Kernel Density-based algorithm (b) PST algorithm (c) Wavelet-based algorithm

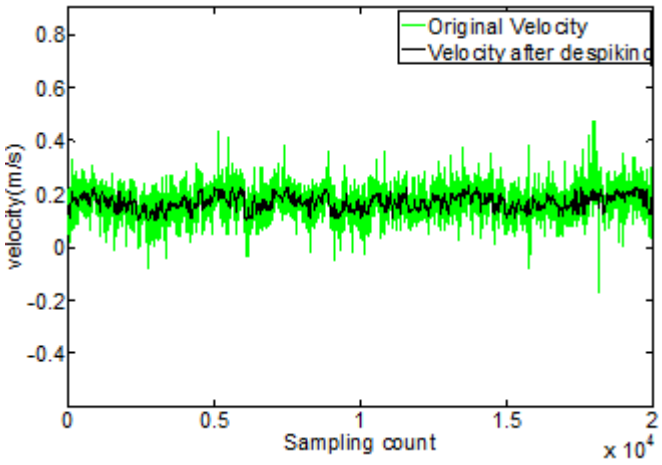
As we can see it from Figure.2., the three ways mentioned above are all effective to despike laboratory data. Besides those 50 artificial points, the despiking ways also eliminated some points belonging to the data themselves. It means either there are some spikes itself in the data on account of no perfect environment or no distraction in this experiment, or else it is defect of these three ways that = identify wrong some valid points. The root cause need us to do further research. The reason why the trend of data after despiking by Wavelt



(a)



(b)



(c)

Figure 3: comparison before and after despiking ADV data from the field experiments by different ways (a) comparison before and after despiking ADV data from the lab by different ways (a) Kernel Density-based algorithm (b) PST algorithm (c) Wavelet-based algorithm

The aim of our research is to detect spike and replace it with valid ways to make it what it is supposed to be. It really helps improve issues of the Three Gorges which is a very significant to Chinese economic, even to the globe, and Chinese environment. From the Figure.3., those three techniques all can effectively reconstruct original velocity in the flow. However, it isn't able to identity all spikes, which need specialized people who interested in this subject to do more further research.

References

[1] Michael Jesson;Mark Sterling;John Bridgeman. Despiking velocity time-series—Optimisation through the combination of spike detection and replacement methods. Flow Measurement and Instrumentation. 2013. Vol.30.45-51.

[2] Razaz M, Kawanisi K. Signal post-processing for acoustic velocimeters: detecting and replacing spikes[J]. Measurement Science and Technology, 2011, 22(12): 125404.

[3] Islam M R, Zhu D Z. Kernel Density–Based Algorithm for Despiking ADV Data[J]. Journal of Hydraulic Engineering, 2013, 139(7): 785-793.

[4] Amir Hossein Birjandi;Eric Louis Bibeau . Improvement of Acoustic Doppler Velocimetry in bubbly flow measurements as applied to river characterization for kinetic turbines. International Journal of Multiphase Flow. 2011. Vol.37. No.8. 919-929.

[5] Goring, D. G., and Nikora, V. I. _2002_. "Despiking acoustic Dopplervelocimeter data." J. Hydraul. Eng., 128_1_, 117–126.

[6] Parsheh, M.,Sotiropoulos, F., and Porté-Agel, F. (2010)."Estimation of Power Spectra of Acoustic-Doppler Velocimetry Data Contaminated with Intermittent Spikes." J. Hydraul. Eng., 136(6), 368

[7] Wahl T L. Discussion of "Despiking acoustic doppler velocimeter data" by Derek G. Goring and Vladimir I. Nikora[J]. Journal of Hydraulic Engineering, 2003, 129(6): 484-487.

[8] Rousseeuw P J and Croux C 1993 Alternatives to median absolute deviation J. Am. Stat. Assoc. 88 1273–83.