Research of Despiking ADV Data

Huang Yafei
River and Ocean Engineering School
Chongqing Jiaotong University
Chongqing, China
1195727827@qq.com

Abstract— **Data recorded by Acoustic Doppler Velocimeter(ADV) are usually affected by several factors, including signal aliasing and Doppler noise, etc,which causes many spikes occurring in the data demanding us to despike it. This paper includes two steps to despike ADV data which was measured in the Three Gorges and the the whole process includes the first one is using the modified version of phase-space thresholding(PST) method which proposed by Parsheh to detect the spike and extrapolation is applied in the second step.**

Index Terms— **PST;ADV;extrapolation;despike;spike;**

I. INTRODUCTION

This is an open sesearch problem. Whether in China or for global ranges,there has been no accurate methods to fix this problem which many experts have studied for many years. Due to therer is no flawless contion in field measurement,many amplitude spikes obviously occur in the ADV data as figure.1. Those data which is hugely bigger than normal ones are considered as spikes which could effect its statistical properties and power spectral density of the data[1].There can be ambiguity for the root cause of these spikes[2]. However,the majority of studies showed that spikes occurance is relevant to the preset velocity range of the equipment,high turbulenceintensities and complex flow boundaries and so on[1].Still,ADV is a popular equipment of measuring velocity in the recent past years.Many advantages exist in this equipment.Above all,we come up with modified method[1] based on previous research which mostly use PST[2] to detect spikes and then use a new extrapolation to rerplace the spikes taken out due to there is a gap if not replaced.
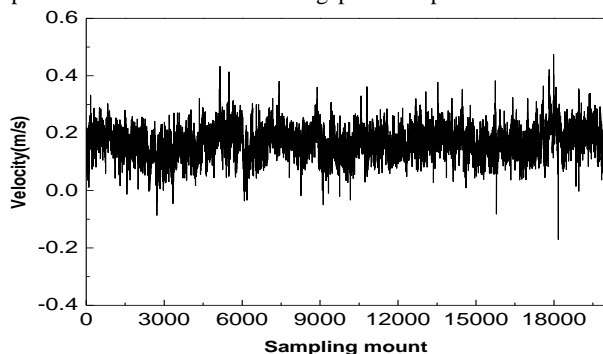

Figure.1. Example of ADV data including spikes from measurements of The Three Gorges

II. DETECTING SPIKE TECHNIQUES

Goring and Nikora[2] proposed a new means called phase-space thresholding to detect spikes which is a milestone for despiking ADV data.From that time onwards,millions of ways linked to PST have sprung up globally and there are also many methods which has nothing to do with such as wavelet[3] and Kernel Density-based way[4].etc. The PST which uses the very concept of a 3D poincare point where the variable and its derivatives are ploted against each other. The points which lie outside the ellipsoid defined by the criterion are regarded as spikes and the points inside are valid. The technique iterates until no spike is detected again[2].

Parsheh[1] and Wahl[5]improved the effectiveness of the PST.Wahl suggests that the median absolute deviation (MAD)is more effective than the standard deviation which can be distracted by the extreme values in the distribution. Parsheh proposed that comfirming some valid data which might be detected as spike and this points cant be changed in the whole iteration process.Following that,they remove the obvious spikes characterized by a large magnitude compared to the rest of the set through the way mentioned below.

The whole process are below:

1.calculate the MAD as scaling quantity:

$$\delta_u = median|u - median(u)| \tag{1}$$

2.velocity fluctuations in the range are flagged as valid and permanent data points:

$$-c_1\delta_u \le u' \le c_1\delta_u \tag{2}$$

3.removal of the large magnitude of u' outside the following scope:

$$|u'| \ge c_2\delta_u\sqrt{2\ln(n)} \tag{3}$$

Above parameters, Parsheh get and after massive numerical studies which is no longer valid in our data messured in The Three Gogres that more complicated than Parsheh's.we regard those 1.6 and 1.5 respectively in our research.

4.after all that,we just need do the basic PST procedure to detect spikes.

After taking out the spikes, there are some gaps in the data set. There are numerous choices we can choose:

III. FILLING THE GAPS WHERE THE SPIKES LIED

1. the last valid data points to fill the gap which is recommended by Parsheh;

2. extrapolation from the previous data point:
$$u_i = u_{i-1};$$

3. extrapolation from the two proceeding points: $u_i = 2u_{i-1} - u_{i-2};$

4. extrapolation from 5 previous points on the one side in combination with smooth estimate;

IV. APPLICATION RESULTS

Tests were performed on all five described replacing techinques in combination with the modified PST techniquet to identify the spikes events in the data set. Comparing this five methods,as Figure 2,we found extrapolation from 5 previous points on the one side combining with smooth is the best method in all of them.From the Figure 2, we can see they all are able to effectively detect spikes except extrapolation from the two preceeding points. The reason is that ADV is a three-dimensional equipment to measure velocity,which means there could be negative velocities and positive velicities in the data set when the flow is complicated.For example,if the previous point is a positive one and the one before the previous is a negative one,using $u_i = 2u_{i-1} - u_{i-2}$ to calculate the u would get a larger velocity than original one.

With the iteration going on,the u would be larger and larger as shows in Figure.2(c). Spikes become larger than before after application of this method. So this replacing method is not practical. The ways of the last valid points and extrapolation from the preceding one are almost same. Both are replaced with data points. However,the latter is less effective than the former due the provious data might be a spike the PST isnt able to find.what's more,neither arent able to identify all spikes which we can see in the Figure.2(a)and (b).The combination way is valid in all methods,which we can see all spikes are nearly detectedin the Figure.2(d) and it is still keep the tendency what it is.
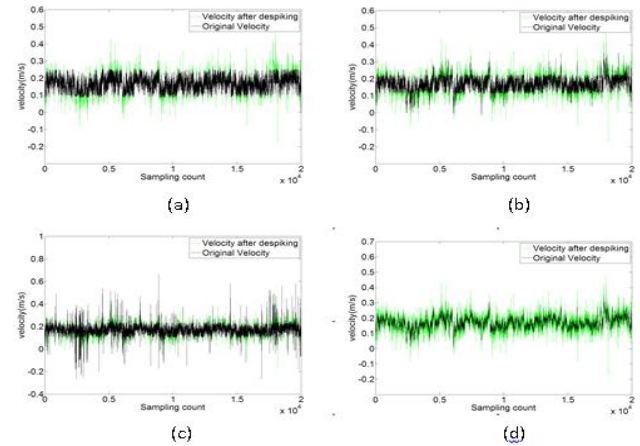

Figure.2. Velocities before and after tackling using differetn way mentioned above(a) the last valid data points to fill the gap (b) extrapolation from the previous data point(c) extrapolation from the two preceding points (d) extrapolation from 5 previous points on the one side in combination with the smoothed estimate;

V. CONCLUSION

In this paper,one detection of spikes and different replacing ways have been used in the ADV data set. We drew conclusions as followed:

1.The methods of the last valid data points to fill the gap and extrapolation from the previous data point are similar, which both are effective but aren't able to find all spikes. At the same time,there is no tendency after application of them.

2.The way $u_i = 2u_{i-1} - u_{i-2}$ is impractical.

3.The extrapolation from 5 previous points on the one side in combination with smooth estimate that we reccommended is the best valid.It can identify all spikes and keep the tendency of original data set.

REFERENCES

[1] 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

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

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

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

[5]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.