39th IAHR WORLD CONGRESS GRANADA, SPAIN 2022 From Snow to Sea 19-24 June 2022

Proceedings of the 39th IAHR World Congress
19–24 June 2022, Granada, Spain
doi://10.3850/IAHR-39WC2521716X20221215

A New Software Package and Filtering Algorithm Development for Despiking Doppler Velocimeter Data

Farzad Asgari[(1)], Seyed Hossein Mohajeri[(2)], Mojtaba Mehraein[(3)] and Mohammad J. Ostad Mirza Tehrani[(4)]

[(1, 2, 3)] Faculty of Engineering, Kharazmi University, Tehran, Iran,
farzad.asgari@khu.ac.ir, hossein.mohajeri@khu.ac.ir, mehraein@khu.ac.ir

[(4)] Faculty of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran,
mohammad.tehrani@kntu.ac.ir

Abstract

Filtering spike contaminated velocity time series, recorded by Acoustic or Laser Doppler Velocimeter (ADV or LDV) in laboratory open-channel flow studies and field measurements, has always been challenging. Despite numerous conducted studies on the velocity time-series signal filtering methods, the importance of the number of spikes of invalid data on the performance of filtering algorithm has not been determined. Indeed, there is still a lack of comprehensive and updated study on the performance of despiking algorithms of the velocity signals. In the present study, a new developed software package for despiking Doppler Velocimeter Data has been introduced. The package is composed of various detection techniques such as Phase-Space Threshold (PST), Velocity Correlation Filter (VCF), Kernel Density Estimation (KDE). A new filtering technique, so called "*Three Dimensional Fast Kernel Estimation* (3D-fastKDE)" has been developed and employed to detect spikes in highly polluted signals. Implemented replacement algorithms include Last Good Values (LGV), and 12 points cubic polynomial interpolation (12pp). The performance and accuracy of detection and replacement techniques has been explored in this study using different experimental data sets.

Keywords: Doppler Velocimeter; Velocity time-series; Flow measurement; Filtering, Fast Kernel Density Estimation

1. INTRODUCTION

The Acoustic Doppler Velocimeter (ADV) was introduced in the early 1990s and has since been vastly used for experimental and field studies (Lane et al., 1998; Nikora and Goring, 2000; Puertas et al., 2004). Despite the ADV advantages, noises in the measured data, which usually arise in form of spikes in the velocity signal remains as the unsolved main issue (Goring and Nikora, 2002; Lane et al., 1998). A set of researches have been done to compare the results of measured velocity by an ADV with other velocity measuring instruments (Lohrmann et al., 1994; Nikora and Goring, 1998; Voulgaris and Trowbridge, 1998; Hurther and Lemmin, 2008; Khorsandi et al., 2012; Hejazi et al., 2016). The results show the opportune exactitude of ADV in measuring the average velocity and high relatively impact of noises on turbulent characteristics. Consequently, in order to improve the accuracy of the results, comprehending the causes of errors and using post-processing methods to enhancement the veracity of measurements has consummate more attention.

Spikes can transpire in ADV measured data due to flow aeration rate, high turbulence intensities, overstepping of the velocities from the ADV probe velocity range (Goring and Nikora, 2002). The basic flow characteristics such as power spectrum, kinetic turbulent energy, and turbulence characteristics are inconsistent with the actual values due to the presence of these spikes (Lane et al., 1998; Wahl, 2000; García et al., 2005; Cea et al., 2007). Consequently, spikes must be removed or replaced from the time series using appropriate methods.

Variety of methods have been developed to discern spikes, replacing or detaching them from the ADV velocity time series. Goring and Nikora (2002) presented the Acceleration thresholding method, which is claimed to be useful for despiking signals free of complex spikes, and the Phase-Space thresholding method (PST) for complex spikes. The PST method plots each turbulent velocity fluctuation and its derivatives against each other in phase-space using Poincaré map equations to correct inaccurate data resulting from data aliasing. This method was the basis of many subsequent despiking algorithms and was modified to further improve the quality of the data measured by (Wahl, 2003). Utilizing the 3D PST algorithm to omit the spikes from measured velocity temporal signal in bubbly flows, was scrutinized by Mori et al. (2007). Using the PST approach, Cea et al. (2007) presented the Velocity Correlation Filter (VCF) by considering the correlation between velocity components. This technique plots all three velocity components against each other in phase-

space, which is claimed to perform better than PST in highly turbulent flows. Parsheh et al. (2010) proposed the Modified Phase-Space thresholding method (mPST) by considering a pre-filter condition before applying the PST algorithm to prevent the identification of suitable data adjacent to long spikes (due to the high magnitude of velocity fluctuation gradient). Islam and Zhu (2013) proposed a despiking algorithm based on kernel density estimation equations by studying an experimental dataset near a wall jet that was highly contaminated with spikes. They found out that antecedent methods were not utilitarian in these highly polluted datasets, where more than 40% of the data is consists of spikes. Irrespective of significant progression, there is still no comprehensive algorithm that can be used in all flow conditions and with any percentage of data pollution. Considering the spikes number in a velocity signal, some researchers have claimed that their proposed method is suitable for data with low number of spikes and uncomplicated spikes, while some claim that their proposed method is suitable for data with high pollution rates (high number of spikes).

This paper introduces the new developed software package and obtained results. Embedded replacement and detection techniques are addressed, followed by introducing a new developed detection technique. The performance of the developed technique has been explored and discussed, presenting the results obtained by utilizing different experimental data sets.

2. METHODOLOGY

2.1. Filtering Algorithms

The developed program includes distinct detection algorithms composed of Phase-Space Thresholding Method (PST), Velocity Correlation Filter (VCF) and classical Kernel Density Estimation (KDE). Also, the 3D-fastKDE has been developed and added to this package. The PST method is one of the most common and popular detection techniques, which is the basis of many subsequent algorithms. This loading method was introduced by Goring and Nicor (2002) and then other researchers conducted further studies on it. The basis of this method is the involvement of phase-space equations of velocity and its first and second derivatives in the despiking operation ($u - \Delta u - \Delta^2 u$). Following this technique, Cea et al. (2007) proposed VCF by conceptualizing the PST method in order to consider the correlation between velocities in different directions and to improve the algorithm in turbulent flows with very high concentrations of air bubbles. Studies have shown that in this type of flows, there is a possibility of producing additional spikes if replacement algorithms are used. In this method, instead of examining the instantaneous velocity in each direction against the first-order and second-order derivatives used in PST, the velocities in all three directions are plotted against each other ($u - v - w$). Similar to the PST algorithm, the equations used to calculate the criteria ellipse diameters are derived from the Universal threshold and the standard deviation of each velocity component.

The KDE is also based on the kernel bivariate distribution function (Islam and Zhu, 2013). One of the most important advantages of this method is its non-iterative process. The kernel equations act as a 2D histogram that can specify a 2D density estimate in the dataset measured by ADV. In this method, after normalizing and transferring the data to the center of the coordinate, the dataset density is calculated using kernel equations.

2.1.1. 3D-fastKDE *detection*

The principal idea of this algorithm is derived from the KDE algorithm proposed by Islam and Zhu (2013) and the velocity correlation Filter (VCF) algorithm proffered by Cea et. al (2007). In algorithms such as phase-space and kernel density, which are calculated based on the instantaneous velocity of each direction and their derivatives, the correlation between velocities in other directions is not considered, and the maximum correlation participation is the clearance of spikes detected in one direction from all three velocity components. Accordingly, the logic of the Velocity Correlation Filter algorithm, which operates by considering the correlation and covariance between the velocity components, has been utilized. To consider the correlation between velocity components, the longitudinal, transverse, and vertical velocity components are placed against each other, and calculations are performed given pursuant thereto. Thus, by reckoning the three-dimensional kernel density estimation, the peak location and maximum density are selected and the appraisals are performed in a three-dimensional space. Nevertheless, the main objection to this idea was program runtime. By selecting 64 grid points, the average time to estimate 3D density was about 50 minutes. By selecting 256 grid points, the approximate time for estimating 3D density was about 12 hours for Python which is an interpreted programming language. To solve this enormous hitch, many occasions were investigated. Including converting one three-dimensional kernel to three two-dimensional kernels or using conditional equations to estimate 3D density. Notwithstanding the fact that converting the 3D kernel to 2D kernels greatly increased the execution speed of the program and, despite selecting 256 mesh networks, reduced the execution time of the algorithm to about 30 seconds, however the logic of the algorithm is in 3D space and sphere coordinates

should be utilized (Wahl, 2003). Furthermore, the use of conditional equations did not significantly succor the runtime of the algorithm and the filtering operation. Especially when the total number of data was high, much more time was spent estimating 3D density. To untangle the knot of filtration runtime by estimating 3D density, the method of Fast Kernel Density Estimation (fastKDE) was used. O'Brien et al. (2016) proposed equations for estimating fast univariate and multivariate density estimation in terms of vector correlations. The fastKDE method demonstrates statistical accuracy and generates kernel density estimates several times faster. For a succinct revision, for a generalized, multivariate kernel density estimation $P(X)$ for d-dimensional multivariate data X_i (for 1 = 1, 2, …, n) using a haphazard shape of kernel function $K(X)$:

$$P(X) = N^{-1}\sum_{i=1}^{n} K(X - X_i) = N^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty} K(s).\,\delta\big(X - X_j - s\big).\,ds \qquad [1]$$

where $\delta(x)$ denotes the Dirac delta function and X denotes the data coordinates. The inverse Fourier transform of KDE can be demonstrated as:

$$\phi(t) = f^{-1}\big(P(X)\big) = \kappa(t).\,l(t) \qquad [2]$$

where t stands for frequency spaced coordinates and f^{-1} denotes the multi-dimensional inverse Fourier transform from X to t, κ denotes the inverse Fourier transform of kernel and l represent the ECF which is defined as:

$$l(t) = \frac{1}{N}\sum_{j=1}^{N} e^{i\vec{x}_j.\vec{t}} \qquad [3]$$

and an optimal transform kernel is defined as Bernacchia and Pigolotti (2011):

$$\widehat{K}(t) = \frac{N}{2(N-1)}\left[1 + \sqrt{1 - \frac{4(N-1)}{N^2|\varphi(t)|^2} I_A(t)}\right] \qquad [4]$$

where $I_A(t)$ denotes a frequency filter which is equal to 1 for the frequencies A used in kernel density estimation and 0 otherwise. Utterly particulars have been provided in O'Brien et al. (2016). Using this method to estimate density, diminished the execution time of the 256, 256, 256 three-dimensional mesh grids to 13 seconds. To use the new 3D fast kernel density estimation algorithm to despike the data measured by the ADV, the rotation of the three-dimensional matrix is performed by the vectors of longitudinal velocity u, transverse velocity v and vertical velocity w. The 3D density of the transmitted and standardized data must then be calculated. After 3D estimation of the density, which is selected as a three-dimensional matrix $n \times n \times n$ where n denotes the number of grid points, the location of the density peak is distinguished amidst all three dimensions of the matrix. Figure 1 shows a 3D density estimation of data measured by ADV. Subsequently, we can do the same as the simple kernel density estimation algorithm and calculate the criteria ellipse diameters. Instead of the fixed criterion of 0.4, the dynamic criterion introduced in Universal threshold equation is used. The number of 257 grid points was chosen.

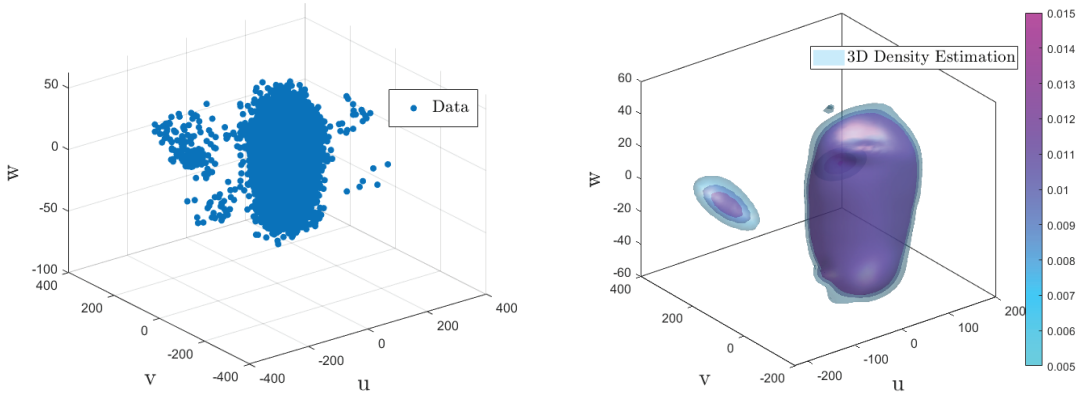

Figure1. 3D density estimation of data measured by Acoustic Doppler Velocimeter.

Ultimately, using the 3D matrix of the data and criteria ellipse, points laying outside the ellipse are detected as spikes in all three directions. This algorithm does not have repetition in the implementation of the algorithm and it is non-iterative, the replacement operation is optional.

2.2. Replacement Techniques

In this research, two important replacement algorithms that are widely used have been investigated which are Last Good (Valid) Value (LGV) and 12 Points Polynomial (12PP). In the LGV method, the detected spike is simply replaced with the last valid data point. The advantage of this method is that it does not depend on subsequent spikes, especially when several consecutive data have been identified as spikes (Jesson et al., 2013). In such cases, one of the disadvantages of this method is that it creates a flat area in the signal. If this happens to a large number of identified spikes, the effect on calculating the mean and standard deviation will be significant (Goring and Nikora, 2002). The use of this alternative method is recommended when using the PST detection algorithm modified by Parsheh et al. (2010). In the 12PP method, a cubic interpolation is made between 24 points around the spike (12 data on each side), and the replacement value is calculated. The vector consisting of 24 side points must be free of spikes. Otherwise, the spikes that have not yet been removed will affect the replacement and the next loop of the detection algorithm, fully described in Goring and Nikora (2002).

2.3. ADV Data

Some algorithms (such as Acceleration thresholding) have been suggested to be suitable only for low-polluted datasets that do not have a complex spike (Goring and Nikora, 2002). Also, some algorithms (such as KDE) have been suggested to be suitable only for highly-polluted datasets (Islam and Zhu, 2013). Nevertheless, none of them give a precise definition of pollution. To evaluate the performance of different algorithms concerning data contamination, 3 types of flow have been collected. In this study, datasets with pollution rates less than 4~5% are called low-polluted, datasets with pollution rates more than 4~5% and less than 10~15% are called polluted and datasets with pollution rates above 15% are highly polluted. The rate of pollution is assumed to be equal to the number of unique spikes detected and replaced, divided by the total number of data.

In the present study, 5 datasets were examined. For the first datasets used in this experiment, a channel with a 90-degree arc was used. The central radius of this arch is equal to 1.2 meters and the width of this channel is equal to 0.6 meters. The floor material of the canal is sand with a diameter of 2.4 mm. The upstream flow velocity is 35.6 cm/s and the water depth is 11.7 cm. A straight spur dike is located at a 45-degree angle to the beginning of the arc. An ADV made by Nortek was used to measure the velocity field. At a sampling frequency of 100 Hz, about 18,000 data were collected upstream in 3 minutes (Referred to as LP-M in this paper).

The second series of data is obtained in a flume with a cross-section of about 0.25 square meters and an approximate length of 6 meters. An axial symmetrical jet discharged into the flume by a 1 cm diameter brass tube was fed by a fixed-head tank 2.8 m above the ground. The Reynolds Jet was set at around 10,000, with an output velocity of about 1 meter per second. The perimeter of the jet was static fluid. Velocity field measurements were performed by an ADV manufactured by Nortek. A sampling frequency of 200 Hz was selected. The data were taken at a distance of 75 cm from the jet outlet and across the channel and on a plane parallel to the channel floor (Referred to as LP-KH in this paper).

For the third data, the experimental data of Righetti (2008) was used, which was kindly provided to the authors. A laboratory flume with 150 m length, 2 m width, 2 m depth, and the bed was used. The walls are made of concrete and the final part of the flume is finished with an adjustable rectangular overflow. Water flows through a closed circuit using a group of four pumps, up to a maximum discharge of 1.57 m^3/s. The slope of the channel bed is 0.18%. The downstream part of the flume was uniformly covered with 45 m long Salix bushes. Data from this channel were collected using a Nortek ADV device and a sampling frequency of 25 was selected (Referred to as P-R in this paper). Full details are provided by Righetti (2008).

The fourth series of data used in this study is related to Devi and Kumar (2016) and was kindly presented to the authors. This data was collected in a sloping flume with 20 m of length, 1m of width, and 72 cm of depth in a plant sand bed. A tank with a length of 2.8 meters, a width of 1.5 meters, and a depth of 1.5 meters are built upstream of the flume to regulate the flow to enter the flume Two data from this channel was collected (Referred to as Dataset4 in this paper) using a Nortek ADV. A sampling frequency of 200 was selected and about 24,000 data were collected in 2 minutes (Referred to as HP-K1 and HP-K2 in this paper). Full details are provided by Devi and Kumar (2016).

In this study, the Cartesian right-handed coordinate system was used. According to this system, the *x*-direction is considered to be in the channel path from the channel input, the y-direction from the right wall to the left wall, and the z-direction in the vertical direction from the channel bed upwards. The four introduced datasets had different flow conditions and different specifications. Moreover, the rate of their pollution is different from each other. Table 1 shows the datasets and the estimated pollution by the base method (SSA).

Table1. Estimated pollution of datasets by base method (SSA).

Dataset	Pollution	Category
LP - M	~2.80%	Low Polluted
LP - KH	~1.48%	Low Polluted
P –R1	~7.47%	Polluted
P – R2	~8.88%	Polluted
HP – K1	~16.80%	Highly Polluted
HP – K2	~23.28%	Highly Polluted

3. RESULTS

The accuracy of each algorithm is evaluated by the detected spike percentage, power spectrum, and data normality test. The number of replaced data by each despiking algorithm is compared with the number of spikes detected by the base filtering algorithm (Singular Spectrum Analysis). In order to study the statistical properties of turbulence of filtered signals, the power spectrum of turbulence for the whole laboratory data set was compared with the common properties of the power spectrum of turbulence, which was examined using the approved law of -5/3. This law has received much attention in the fluid and hydraulic mechanics community in recent decades and is the basis of many turbulent flow models (Lewandowski and Pinier, 2016), also known as Kolmogorov's law. Kolmogorov's law states that in some limits of inertia, the energy density of the flow E(k) behaves like $c^{te}k^{-5/3}$, where *k* represents the number of flow waves (Lewandowski and Pinier, 2016). According to the central limit theorem, a dataset with a large number of members must have a normal distribution (Ross, 2017). For a dataset with a perfect normal distribution, the amount of skewness is 0. If Pearson's law is used, the value of kurtosis is ideally 3. The skewness is equal to the third momentum of the normalized data ($\overline{u'^3}/\overline{u'^2}^{1.5}$) and is, in fact, a measure of the degree of symmetry of the distribution function. kurtosis is equal to the fourth momentum of the normalized data ($\overline{u'^4}/\overline{u'^2}^{2}$), in other words, kurtosis is a measure of the sharpness of the curve at the maximum point (Tennekes and Lumley,1972). Also, for a more detailed study of the importance of despiking, higher moments have been calculated and examined.

In low-polluted datasets (pollution less than 4~5%), no significant change is expected in the power spectra. In fact, unfiltered data must also follow Kolmogorov's -5/3 rule. As may be seen in Figure 2, the power spectrum of each low-pollution dataset follows Kolmogorov's -5/3 rule before despiking, and as expected, there is no noticeable difference between the spectrum of filtered and unfiltered data.

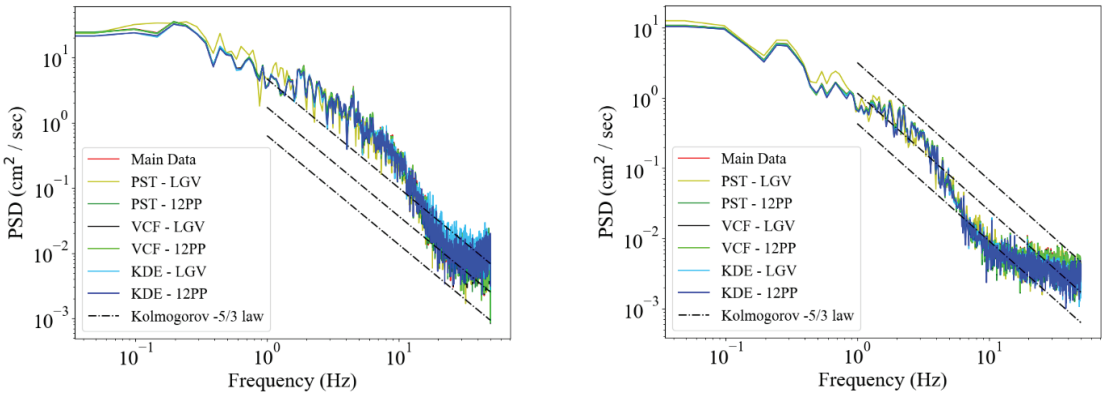

Figure 2. a) Power spectra for dataset LP-M, b) Power spectra for dataset LP-KH.

Table 2 contains the calculated higher momentum of the dataset LP-M and LP-KH normality test. According to the results, the PST-LGV algorithm has unacceptable results in both datasets. Also, the PST-LGV despiking algorithm was still finding spikes after many iterations and has identified a lot of valid data as spikes. The skewness of the filtered data using the PST-LGV despiking algorithm is very close to the ideal value, especially in the LP-M dataset. However, due to the large number of suitable data which are identified as spikes by this algorithm, these results cannot be reliable. 12pp replacement method leaded better results for the PST algorithm. The number of spikes detected by this algorithm is close to the pollution estimated by the base algorithm. Also, the skewness of the filtered data with the PST-12PP algorithm provides better and more reliable values than PST-LGV. The results of the VCF algorithm showed that the replacement algorithm has a low effect on the despiking operation and that the skewness and kurtosis values are close to the ideal values. According to the results of the normality test of filtered data using the VCF detection algorithm, the LGV replacement method has a slight advantage over the 12PP method. Also, the filtration speed is higher. Since the KDE algorithm is non-iterative, it is obvious that the amount of pollution is the same with each replacement algorithm. Also, the amount of skewness and kurtosis of the replaced data are very close to each other in both algorithms. As mentioned by Islam and Zhu (2013), this algorithm has not performed well in low-polluted datasets and has identified a lot of valid data as spikes. The best skewness test results are related to the KDE detection algorithm. However, these results are not reliable due to the identification of a large number of suitable data as spikes.

Table 2. Mean longitudinal flow velocity, standard deviation, skewness, kurtosis, and pollution of dataset LP-M and LP-KH.

		Mean (cm/s)	Standard Dev.	Skewness	Kurtosis	Pollution
LP-M	Main	7.46	5.60	-0.20	3.49	-
	PST-LGV	7.49	5.79	-0.01	3.64	10.77 %
	PST-12PP	7.47	5.55	-0.20	3.25	2.09 %
	VCF-LGV	7.46	5.57	-0.16	3.26	0.26 %
	VCF-12PP	7.47	5.57	-0.17	3.29	0.26 %
	KDE-LGV	7.46	5.46	-0.13	2.89	8.42 %
	KDE-12PP	7.43	5.42	-0.18	2.89	8.42 %
LP-KH	Main	6.96	2.27	0.06	3.03	-
	PST-LGV	6.97	2.35	0.07	3.15	6.9 %
	PST-12PP	6.96	2.26	0.09	2.85	1.49 %
	VCF-LGV	6.96	2.27	0.07	2.95	0.07 %
	VCF-12PP	6.96	2.27	0.07	2.93	0.07 %
	KDE-LGV	6.95	2.22	0.03	2.62	7.64 %
	KDE-12PP	6.96	2.22	0.03	2.59	7.64 %

Figure 3 shows the detected spikes by PST-LGV and KDE-LGV despiking algorithms for the LP-M dataset's longitudinal flow velocities. It may be seen that these two methods have identified many suitable data as spikes. The results also show that the mean longitudinal flow velocities in all filtered and unfiltered data are almost the same. In general, it can be said that the best performance in the low-polluted datasets was related to the PST-12PP algorithm. Although the PST detection algorithm yielded the best filtration result, studies show that the final results of this algorithm are strongly influenced by the selective replacement

method. The results also showed that by selecting a suitable replacement method for this algorithm, the correct spikes can be identified.

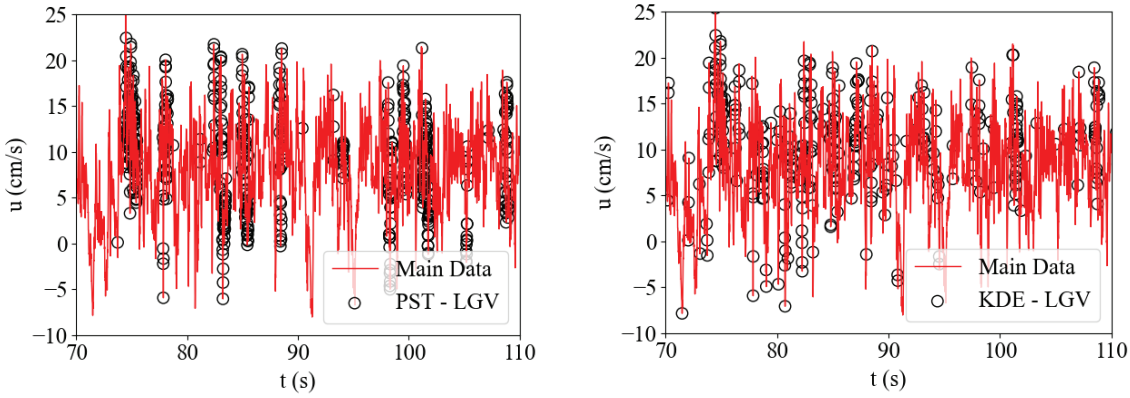

Figure 3. a) Longitudinal flow velocities and detected spikes by PST-LGV in dataset LP-M, b) Longitudinal flow velocities and detected spikes by KDE-LGV in dataset LP-M

Table 3 contains the second and third-order moments of the raw and filtered data by the PST-12PP method (superior algorithm) for low-polluted datasets. The second-order moment of dataset LP-M has decreased by 0.54%. Also, the third-order moment of this dataset has decreased by about 0.95%. The second and third-order moments for dataset LP-KH decreased by only 0.07% and 0.13%, respectively. Considering that the percentage of moments changes in these two datasets is very small, in general, it can be said that the issue of filtering in these datasets is not very important.

Table 3. Second and third order of moments of main and filtered data, for low-polluted datasets.

Dataset	Moment	Second-order	Third-order
LP-M	Main	87.00	1081.49
	Filtered	86.53	1071.26
LP-KH	Main	53.62	445.83
	Filtered	53.58	445.25

In polluted datasets (pollution more than 4~5% and less than 15%), the main data spectrum was expected not to follow Kolmogorov's -5/3 rule. Figure 4 shows the power spectrum of unfiltered and filtered data for both polluted datasets (dataset 3-1 and dataset 3-2). It is clear that the power spectrum of the original data does not follow Kolmogorov's rule as expected. However, the power spectrum of filtered data using all algorithms more or less follows Kolmogorov's law.

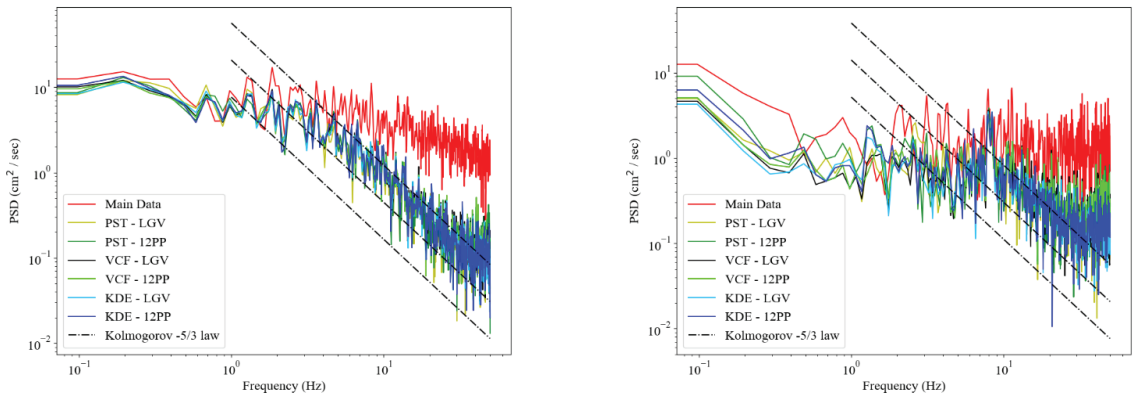

Figure 4. a) Power spectra for dataset P-R1, b) Power spectra for dataset P-R2.

Figure 5 shows the longitudinal velocity time-series signal for the dataset P-R2 before despiking. It may be seen that this dataset can challenge the intelligence of any algorithm. In the trend of this dataset, the data fluctuates in a certain range. This part of the data was examined separately, which can be said to have very few spikes. After about 120 seconds, sharp fluctuations appear in the speed time-series signal.

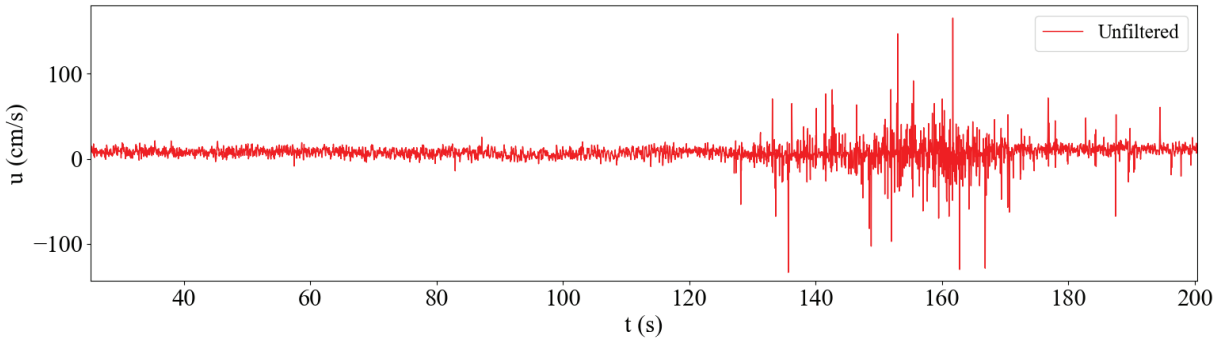

Figure 5. Longitudinal velocity time-series signal for dataset P-R2.

The results of the normality test of the P-R1 and P-R2 datasets are shown in Table 4. In both datasets, it was observed that the mean longitudinal velocities for the raw and filtered data are almost close to each other. In these types of datasets, the PST-LGV algorithm performed better than in the previous type of datasets (low-polluted datasets). The skewness and kurtosis test of the data filtered using this method is close to the ideal values. Also, the number of spikes identified by this method is close to the number of spikes identified by the base method. In these datasets, the PST-12PP algorithm detected fewer spikes than the PST-LGV algorithm, similar to low-pollution datasets. According to the results, it can be said that the replacement method had a great effect on the filtering operation with the PST algorithm. However, the replacement method had less effect on the VCF and KDE detection algorithms. The VCF detects fewer spikes than other algorithms. This algorithm also did not show acceptable results in the kurtosis test, especially in the second dataset. In the second dataset, the kurtosis value of the filtered data with the VCF-12PP algorithm was 4.34. However, the skewness value of this data was 0.07, which is a very acceptable number. The performance of the KDE algorithm has improved due to the increase in data pollution. Both replacement methods had relatively acceptable results with the KDE detection algorithm. The normality test of the filtered data by the KDE method also had acceptable results. It can be said that the KDE-LGV algorithm had the best results in the kurtosis test with PST-LGV. According to the results, PST-LGV, PST-12PP, and KDE-LGV algorithms have acceptable performance in polluted datasets. However, the PST algorithm can be strictly selected as the superior algorithm with a slight superiority. The PST-12PP algorithm was also selected as the superior algorithm for this series of datasets due to the more accurate detection of spikes by the PST-12PP algorithm and the slight difference in the normality test.

Table 4. Mean longitudinal flow velocity, standard deviation, skewness, kurtosis, and pollution of dataset P-R1 and P-R2.

		Mean (cm/s)	Standard Dev.	Skewness	Kurtosis	Pollution
P-R1	Main	7.70	10.39	-1.08	54.11	-
	PST-LGV	7.99	4.71	-0.09	3.32	11.57 %
	PST-12PP	8.06	5.17	-0.12	4.11	10.34 %
	VCF-LGV	8.11	5.02	0.14	4.05	6.35 %
	VCF-12PP	8.10	5.24	0.07	4.34	5.95 %
	KDE-LGV	8.16	4.74	0.15	3.51	11.41 %
	KDE-12PP	8.16	4.91	0.15	3.80	11.41 %
P-R2	Main	53.10	12.29	-4.74	37.56	-
	PST-LGV	54.80	6.48	-0.24	3.04	9.1 %
	PST-12PP	54.65	6.61	-0.23	3.17	8.19 %
	VCF-LGV	54.65	6.74	-0.3	3.27	2.66 %
	VCF-12PP	54.63	6.68	-0.28	3.19	2.66 %
	KDE-LGV	54.73	6.56	-0.21	3.01	10.41 %
	KDE-12PP	54.62	6.62	-0.3	3.28	10.41 %

The performance of the new algorithm in this dataset was also examined in detail. This algorithm was also able to clear the entire signal from the spikes as well as the KDE method. The filtered signal had a normal distribution and the values of skewness and elongation were very close to the ideal values. Figure 6 shows the main and filtered signal of dataset obtained by Righetti (2008) (Figure 6-a) and Devi and Kumar (2016) (Figure 6-b), using 3dKDE despiking method. It is quite obvious that the introduced method was able to pinpoint the true signal consummately and annihilate the spikes from it.

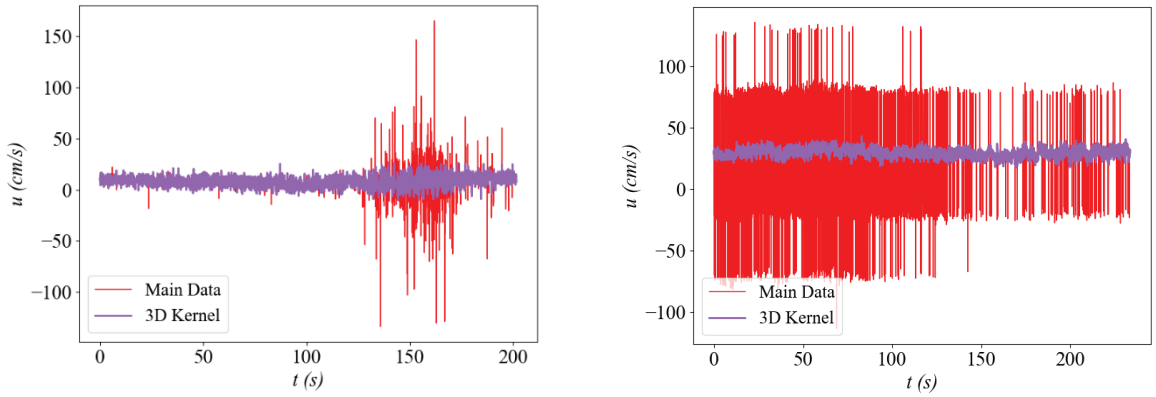

Figure 6. Time-series longitudinal velocity of unfiltered and filtered signal using 3dKDE despiking method, a) Righetti (2008), b) Devi and Kumar (2016).

4. CONCLUSIONS

In the present paper, a new velocity data despiking software package has been introduced and the accuracy of different techniques has been discussed. The decoction technique is composed of PST, VCF, KDE and 3D-fastKDE, while the replacement techniques are LGV and 12PP. The main observations in this study are:

- Generally, it is found that an appropriate filtering method depends on the level of spikes presence.
- For the low polluted signals (less than 4~5% of the data are spikes), the PST algorithm detects the spikes better than other spike detections methods. The results also show that in this type of signals, the replacement method does make a drastic impact. The LGV replacement method had too bad results and the 12PP method had too good results.
- For the polluted signals (more than 4~5% and less than 10~15% of the data are spikes), the best filtering results have been obtained using PST as detection and 12PP as a replacement method.
- For highly polluted signals (more than 15% of the data are spikes), our results show that most of the detection available algorithms do not properly work. However, the best results were obtained using KDE as a detection method and 12PP as a replacement method.
- The 3D-fastKDE also shows high detection accuracy in low moderate and highly polluted data. Considering this fact, application of this algorithm in laboratory and field measurements has been suggested.
- The observation in this study is essential as a guideline for most of the laboratory and river velocity and turbulence measurement using ADV in the data preparation step.

5. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Babak Khorsandi, Amirkabir University of Technology (Tehran Polytechnic), for his constructive and fruitful discussion. The last author acknowledges the support of the Iranian National Elites Foundation and the International Cooperation Center of the Scientific and Technological Department of Presidential Office.

6. REFERENCES

Bernacchia, A., and Pigolotti, S. (2011). Self-consistent method for density estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73: 407-22.

Cea, L., Puertas, J., and Pena, L. (2007). Velocity measurements on highly turbulent free surface flow using ADV. *Experiments in Fluids*, 42: 333-48.

Devi, T.B., and Kumar B. (2016). Channel hydrodynamics of submerged, flexible vegetation with seepage. *Journal of Hydraulic Engineering*, 142: 04016053.

García, C. M., Mariano I. C., Yarko N., and Marcelo, H.G., (2005). Turbulence measurements with acoustic doppler velocimeters. *Journal of Hydraulic Engineering*, 131: 1062-73.

Goring, D.G., and Nikora, V.I. (2002). Despiking acoustic doppler velocimeter data. *Journal of Hydraulic Engineering*, 128: 117-26.

Hejazi, K., Falconer, R.A., Seifi, E. (2016). Denoising and despiking ADV velocity and salinity concentration data in turbulent stratified flows. *Flow Measurement and Instrumentation*, 52 :83-9.

Hurther, D., Lemmin, U. (2008). Improved turbulence profiling with field-adapted acoustic doppler velocimeters using a bifrequency doppler noise suppression method. *Journal of Atmospheric and Oceanic Technology*, 25: 452-63.

Islam, M.R., and Zhu, D.Z. (2013). Kernel density–based algorithm for despiking ADV data. *Journal of Hydraulic Engineering*, 139: 785-93.

Jesson, M., Sterling, M. and Bridgeman, J. (2013). Despiking velocity time series optimisation through the combination of spike detection and replacement methods. *Flow Measurement and Instrumentation*, 30, pp.45-51.

Khorsandi, B., Mydlarski, L., and Gaskin, S. (2012). Noise in turbulence measurements using acoustic doppler velocimetry. *Journal of Hydraulic Engineering*, 138: 829-38.

Lane, S.N., Biron P.M., Bradbrook, K.F., Butler, J.B., Chandler, J.H., Crowell, M.D., McLelland, S.J., Richards, K.S., and Roy, A.G. (1998). Three-dimensional measurement of river channel flow processes using acoustic Doppler velocimetry. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Group*, 23: 1247-67.

Lewandowski, R. and Pinier, B. (2016). The Kolmogorov law of turbulence what Can rigorously be proved? Part II. *In The Foundations of chaos revisited: from Poincaré to recent advancements* (pp. 71-89). Springer, Cham.

Lohrmann, A., Cabrera, R. and Kraus, N.C. (1994). Acoustic-Doppler velocimeter (ADV) for laboratory use. *Fundamentals and advancements in hydraulic measurements and experimentation*, 351-65. ASCE.

Mori, N., Suzuki, T., and Kakuno, S. (2007). Noise of acoustic doppler velocimeter data in bubbly flows. *Journal of Engineering Mechanics*, 133(1), pp.122-125.

Nikora, V., and Goring, D. (2000). Flow turbulence over fixed and weakly mobile gravel beds. *Journal of Hydraulic Engineering*, 126: 679-90.

O'Brien, T.A., Kashinath, K., Cavanaugh, N.R., Collins, W.D., and O'Brien, J.P. (2016). A fast and objective multidimensional kernel density estimation method: fastKDE. *Computational Statistics & Data Analysis*, 101: 148-60.

Parsheh, M., Sotiropoulos, F., Porté-Agel, F. (2010). Estimation of power spectra of acoustic-doppler velocimetry data contaminated with intermittent spikes. *Journal of Hydraulic Engineering*, 136: 368-78.

Puertas, J., Pena, L., and Teijeiro, T. (2004). Experimental approach to the hydraulics of vertical slot fishways. *Journal of Hydraulic Engineering*, 130: 10-23.

Righetti, M. (2008). Flow analysis in a channel with flexible vegetation using double-averaging method. *Acta Geophysica*, 56: 801-23.

Ross, S.M. (2017). Introductory statistics. *Academic Press*.

Tennekes, H., Lumley, J. L. (1972). A first course in turbulence. *The MIT Press, Cambridge*. Massachusetts and London, England.

Voulgaris, G., and Trowbridge, J.H. (1998). Evaluation of the acoustic doppler velocimeter (ADV) for turbulence measurements. *Journal of Atmospheric and Oceanic Technology*, 15: 272-89.

Wahl, T.L. (2000). Analyzing ADV data using Win-ADV. *In: Hotchkiss, R.H., Glade, M. (eds) Building partnerships*. Reston, VA: ASCE, pp.1–10.

Wahl, T.L. (2003). Discussion of "Despiking acoustic doppler velocimeter data" by Derek G. Goring and Vladimir I. Nikora, *Journal of Hydraulic Engineering*, 129: 484-87.

