ORCA – Online Research @ Cardiff

Author's Accepted Manuscript

Denoising and Despiking ADV Velocity and Salinity Concentration Data in Turbulent Stratified Flows

Kourosh Hejazi, Roger A. Falconer, Ehsan Seifi

Cite this article as: Kourosh Hejazi, Roger A. Falconer and Ehsan Seifi, Denoising and Despiking ADV Velocity and Salinity Concentration Data in Turbulent Stratified Flows, *Flow Measurement and Instrumentation,* http://dx.doi.org/10.1016/j.flowmeasinst.2016.09.010

Denoising and Despiking ADV Velocity and Salinity Concentration Data in Turbulent Stratified Flows

Kourosh Hejazi[a*], Roger A. Falconer[b], Ehsan Seifi[c]

[a]Hydro-environmental Research Centre, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK. ,
[b]Hydro-environmental Research Centre, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK.
[c]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada, N2L 3G1.

[*]Corresponding author: Department of Civil Engineering, K. N. Toosi University of Technology, 1346, Valiasr Ave., Mirdamad Intersection, Tehran 19967-15433, Iran

HejaziK@kntu.ac.ir

FalconerRA@cardiff.ac.uk

SMSeifi@uwaterloo.ca

Abstract

Due to the presence of noise and spikes in velocity measurements of turbulent flow fields, understanding the flow pattern may be seriously affected by the spurious values for zones where certain flow behaviour is not necessarily expected. In a series of laboratory experiments the velocity data recorded, using Acoustic Doppler Velocity meters (ADVs), and the salinity measurements were noticeably noisy. In despiking and denoising the velocity data a linear correlation algorithm was established, which successfully lowered the noise levels and removed the spikes. For the assessment of the method, an autoregressive model was used to generate a clean velocity signal. The spikes were generated with a uniformly random time index and a Gaussian distributed value, where White Gaussian noise was added to this simulated signal. Assessment was also undertaken on the signals generated using a three-dimensional numerical model. To enhance the comprehension of the flow field, an interpolation method for producing the missing data has also been developed, which may be deployed to increase the sampling frequency, or to produce data for the spatial domain at locations which are not included in the measurements. For salinity a different strategy was applied where a moving average procedure was carried out, as the data did not suffer from spikes and exhibited almost a constant band of noisy fluctuations.

Keywords: Auto-correlation; Despiking; Denoising; ADV; Stratified Flows; Hydraulic Modeling.

1 Introduction

Horizontal flows driven by buoyancy forces in the presence of a surface or interface are a common occurrence in many environmental flow fields. Stratified flows occur in the atmosphere, oceans, estuaries and lakes. Estuaries inherently include: unsteady, non-uniform, periodically reversal direction, stratified flows, with complex vertical salinity and velocity distributions. Stratified flows are very important in estuarine studies, as salt intrusion is almost always present in harbours and barrages located in these water bodies. Physical hydraulic models of estuaries can be very useful in comparing alternative designs or operational strategies.

High levels of noise and spikes have been reported in measurements using an Acoustic Doppler Velocimeter (ADV) (Nikora and Goring 1998; McLelland and Nicholas 2000). The ADV is a remote-sensing, three-dimensional sensor, originally developed and tested for use in physical hydraulic model facilities (Kraus et al. 1994; Lohrmann et al. 1994), and its operation is based on the Doppler shift effect. Voulgaris and Trowbridge (1998) evaluated the accuracy of ADVs by measurements of open channel flow using an ADV and a laser Doppler velocimeter. They concluded that ADVs are suitable for accurate measurements of mean flow characteristics. The accuracy of ADVs is limited when making measurements close to the bed, or in flows where large spatial gradients are present (Dombroski and Crimaldi 2007). To validate the use of ADVs for the measurement of turbulent flows, Khorsandi et al. (2012) conducted experiments in an axisymmetric turbulent jet and in approximately homogenous isotropic turbulence with a zero mean flow. The experiments showed that the horizontal RMS velocities measured using the ADV were overestimated, in comparison with those velocities obtained using both flying hot-film anemometry measurements and accepted values in the literature. However, the vertical component of the RMS velocity agreed well with comparable measurements obtained for other studies. To correct the data, post-processing filters and a Doppler noise-reduction method were applied to the jet data. Despite decreasing the RMS velocities, they remained erroneously higher than the corresponding accepted values. The results showed no clear relationship between the Doppler noise and the mean flow. In turbulent flows, ADV velocity fluctuations characterise the combined effects of Doppler noise, signal aliasing, velocity fluctuations, installation vibrations and other disturbances (Chanson et al. 2008). Both the spectra and the probability distributions indicate that, as a first approximation, the Doppler noise is Gaussian white noise

(Nikora and Goring 1998). The main source of Doppler noise is the random target distribution in the sampling volume, which induces an instantaneous Lagrangian deviation of the position of the target from the mean

position determined by the spatially averaged velocity. This creates instantaneous random Doppler-phase noise that is added to the spatially averaged Doppler phase (Doroudian et al. 2010). McLelland and Nicholas (2000) explained the physical processes causing spikes by aliasing of the Doppler signal.

For steady flow situations a few techniques to eliminate the spikes have been developed (Nikora and Goring 1998; Goring and Nikora 2002; Wahl 2003). Goring and Nikora (2002) assumed that good data can be found within a cluster and that points located outside the cluster are spikes. They suggested a method based on iterative phase-space thresholding as the most suitable solution for spike detection. They reported that the method worked extremely well, confirmed by application of the method to numerous ADV data. Spike detection was based on the postulation that under normal flow conditions the instantaneous acceleration in a stream must be of the same order of, or less than, the acceleration of gravity g. To replace the spikes, a polynomial fitted to good data on either side of the spike event, then interpolated across the event, was recommended by the authors. They mentioned that if the data sequence were undergoing a sustained change in velocity (e.g., over a tidal cycle), then the phase-space method would not work unless these long-scale fluctuations were removed by high-pass filtering. They also pointed out that for some records the choice of thresholds is very difficult and subjective. Doroudian et al. (2010) combined a spike-removal procedure on the beam velocities with a noise-reduction method on the flow velocities to improve turbulence measurements, and compared the results with those obtained from ADVP (Acoustic Doppler Velocity Profiler) measurements under the same conditions. They applied the despiking method to the bi-static velocities before the orthogonal velocities were calculated. It was shown that spikes were best removed from ADV beam velocity data before calculating flow velocities, thereby correcting all three flow velocity components at the source.

The noise reduction method is based on the decorrelation of the Doppler noise terms contained in two vertical velocities redundantly sampled in the same volume. Parsheh et al. (2010) proposed a method for reconstructing contaminated time series which integrates two previously developed techniques for detecting and replacing spurious spikes. They reasoned that the phase-space threshold despiking method (Goring and Nikora 2002; Wahl 2003) erroneously also removes some valid data points in the vicinity of the spikes. They modified the phase-space-threshold despiking method so that the data points near the peak of the probability density function (PDF) were not affected by the despiking method, as these points were not spurious. The spikes were first detected using a modified version of the universal phase-space-threshold technique and subsequently replaced by the last valid data points. The accuracy of the approach was evaluated by applying it to identify and remove spikes and reconstruct the spectra of two clean data sets, which were artificially contaminated with random spikes. The results showed that the power spectra of the reconstructed time series contained filtered white noise, caused by the steps in the reconstruction technique using the last valid data point. It was shown that the modified method improved the spectrum over all frequencies, compared to the original despiking method. Spike-removal, however, depends on the flow conditions and caution should be taken when proposing universal guidelines (Doroudian et al. 2010). Chanson et al. (2008) used an ADV for high-frequency velocity measurements in a small estuary. They showed that conventional despiking methods, such as the phase-space thresholding method (Goring and Nikora 2002), were not sufficient. It has been pointed out by a number of authors, however, that spike detection and spike replacement techniques in three receiver ADV data still remain an arbitrary procedure and that none of the currently available methods gives totally satisfactory results (Doroudian et al. 2010). Yin et al. (2001) developed a method to analyse a time series of velocity signals in order to obtain a time series of a moving-averaged velocity. They described the turbulent velocity fluctuations with a Gaussian probability distribution, and the method was developed to calculate the local mean velocity of an unsteady tidal flow from the experimental data using ADVs with noisy signals. de Nijs et al. (2009) used a 10 minute moving average filter to determine the main flow and turbulence statistics. The moving average algorithm is not constructive when the filter passes through peaks that are narrow compared to the filter width. A better procedure is to perform a least squares fit of a small set of consecutive data points to a polynomial and take the calculated central point of the fitted polynomial curve as the new smoothed data point. The smoothing effect of the latter algorithm is not as sharp as in the case of the moving average and the loss and/or distortion of vital information is comparatively limited.

In this paper a time-invariant minimum mean square estimator has been used, and temporal auto-correlation matrices have been developed, to reduce the noise level and remove the spikes for the ADV records. For the noisy fluctuating salinity data a moving average strategy has been utilised. Section 2 outlines the laboratory apparatus, instrumentation and data acquisition, followed by section 3 which describes the sources of the noise and spikes. Section 4 introduces the linear algorithm, which has been developed for minimising the noise and spikes by the establishing temporal correlations for the flow field. For the salinity concentrations recorded in the model, a moving average strategy has been presented in section 5, to eliminate the noise from the conductivity signals.

2 Laboratory Apparatus, Instrumentation and Data Acquisition

A laboratory tidal basin was designed, set up and deployed for velocity, water surface elevation and salinity measurements of an idealised vertically distorted model harbour, for fresh water and salinity stratified conditions. The tidal basin had overall tank dimensions of 7.02 m by 4.01 m and a maximum depth of 0.76 m. A 50 mm thick honeycomb baffle separated the main area of the basin from the manifold system to produce a uniform inflow to the 5.28 m long by 4.01 m wide working plan area of the tank. A further 50 mm thick honeycomb baffle was placed near the other end of the basin to partly absorb the tidal motions before they were reflected by the end wall, and again on the way back to the basin. Two bridges across the tank facilitated the accessibility of every location of the tidal tank for positioning the measuring instruments and providing the possibility of closer observations without any disturbance and interruption during the experiments. Constant amplitude and constant period sinusoidal model tides were reproduced in the tidal tank by means of a vertically oscillating weir, located just inside the tidal tank, and adjacent and parallel to the inlet pipe. Similar laboratory models have been used in a number of studies conducted by Nece and Falconer (1989), Falconer and Mardapitta-Hadjipandeli (1990), Falconer and Yu (1991) and Falconer and Li (1994). The idealised rectangular model harbour had a horizontal base, four vertical 330 mm high sidewalls and a single asymmetric entrance. The model harbour was positioned at one side of the tidal tank and received long-shore model tidal currents. The breakwater was extended along the whole length of the basin and to the honeycomb baffles. The scaled laboratory model harbour had planform dimensions of 1080 mm × 1080 mm, mean depths of 150 mm, 200 mm, and 250 mm for various test arrangements, and an entrance width of 120 mm. Using a Froude law scaling relationship for dynamic similarity, this resulted in a model tidal period of 708 s and 500 s, and ranges of 100 mm and 200 mm for distortion ratios of 10 and 20 respectively. Picture (1a) shows the model harbour location in the basin.

The mean values of the velocity components in three directions, water surface elevation and salinity were the elements of data acquisition. Measurement of the turbulent, fluctuating components, or their mean values, was not part of the scope of these studies. The values of the velocity components in three directions were measured using Nortek acoustic Doppler velocimeters (NORTEK AS 1997a). HR Wallingford wave probe monitors (HR Wallingford 1994) were used for measuring water levels. For salinity data, the conductivity was measured by a Jenway conductivity meter, model 4320 (Jenway Ltd. 1996). Picture (1) shows the three ADVs with down-looking, side-looking and up-looking probes, with the conductivity meter and wave probe monitor being in operation. In clean, quiescent water, seeding materials were added to ensure sufficient scattering strength. Non-soluble seeding material from Nortek was used for the experiments.

The software for ADVs provided by Nortek (NORTEK AS 1997b) was used for data acquisition and controlling the procedure of measuring the velocity components. For measuring the water surface levels and conductivity, GENIE data acquisition and control software were adapted and used. GENIE (ADVANTECH 1995) enables simulating and running real-time data acquisition and process control strategies. The measurements were taken across a mesh of 12×12 cm grids. Tidal conditions were initiated from low water level in order to minimise the effects of transient surface disturbances within the saline water column.

3 Sources of the Noise and Spikes

Voulgaris and Trowbridge (1998) divided the source of the noise from the ADV velocity records into three categories: i) sampling errors related to the electronic circuitry of the sensor, and the accuracy of the ADV's analogue to digital board in resolving the changes in phase, which are independent of the flow and depend on the pulse length, and are as set by the velocity range of the ADV; ii) Doppler noise, which is an intrinsic feature of all Doppler backscatter systems and is flow-related, and dominates at rapid flows and is caused by turbulence and particle scattering, beam divergence, and the finite residence time of the particles in the sampling volume, and iii) the error attributable to mean velocity gradients in the sampling volume, which becomes important in flows with sharp velocity gradients. The most significant source of noise in ADV measurements is the Doppler noise, which is inherent to the technique (Garbini et al. 1982; Lohrmann et al. 1994). Doppler noise is white noise, which does not influence the mean velocity. Hurther and Lemmin (2001) characterised the Doppler noise as having a flat spectrum over the frequency domain, unbiased and therefore not affecting the mean velocity, statistically independent of the true velocity fluctuations and true Doppler frequency, and having statistically independent noise from one receiver to the next. Nikora and Goring (1998) concluded that: i) both the spectra and the probability distributions indicate that as a first approximation, the Doppler noise is Gaussian white noise; ii) the horizontal noise components are both approximately of the same energy level and have significantly more energy than the vertical component; iii) the velocity range has a significant effect on the noise and iv) the cross-correlation functions between noise components indicate that the cross-correlations are absolute maximum at zero time-lag. If the orthogonal noise components were independent, one would expect the cross-correlations to be zero. In fact, they are not independent but are linear functions of the three direct, along-beam, noise components.
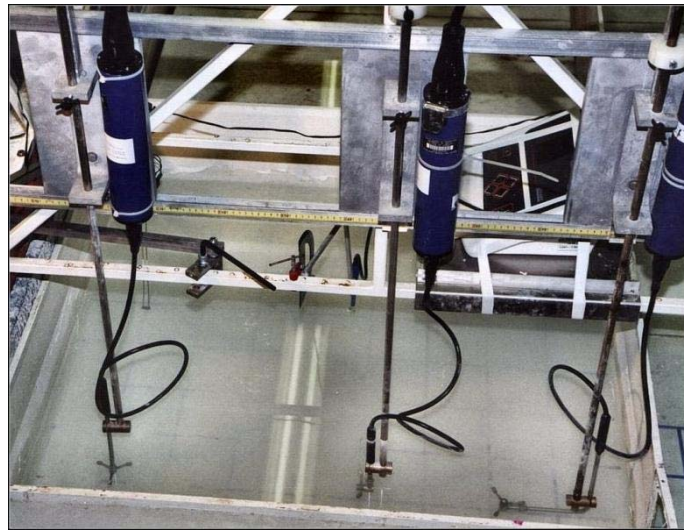
(a) Harbour location in the basin   (b) ADVs, wave probe and conductivity meter in operation

(c) From left: 3D down-looking, side-looking and up-looking ADVs

(d) Wave probe   (e) Conductivity meter probe   (f) ADV probes

Picture 1 Harbour and 3D ADV velocimeters, wave probe monitor and conductivity-meter in operation

Assuming the correlation between the velocity and the noise to be zero, Nikora and Goring (1998) subtracted the measured noise from the measured velocity to estimate the true velocity. Lemmin and Lhermitte (1999) questioned this method, remarking that the Doppler noise should be an increasing function of the mean velocity of the flow. To study the effect of the sampling frequency of the ADV on the turbulence statistics, McLelland and Nicholas (2000) estimated the contribution of noise to the velocity variance at various sampling frequencies, at different mean velocities and depths in a channel flow. They showed that although higher sampling frequencies can characterise a larger range of turbulence frequencies, the total noise in the velocity variance increased at higher sampling rates.

5

In contrast to Doppler noise, random spikes are not inherent to the measurement technique and occur mostly in poor measurement environments (Doroudian et al. 2010). For many velocimeter measurements, velocity spikes are the dominant source of error. Spikes can be disproportionate sources of error because they can have large values compared with the mean velocity. Large spikes can be removed from the data, but it appears that even despiked data have higher noise levels than comparable data without spikes. The source of spikes is not well understood. Spikes may occur as a result of the environment itself, but it is clear that the velocimeter hardware and algorithms are a primary source of spikes. Wahl (2003) considered that spikes in ADV time series may be caused by many factors, including high turbulence intensities, aerated flows that have undesirable acoustic properties, and phase difference ambiguities that occur when velocities exceed the upper limits of ADV probe velocity ranges. Spikes may also be introduced into the data when obstacles, such as solid particles or air bubbles, block the sound path between the emitter and the receiver (Doroudian et al. 2010). Parsheh et al. (2010) showed that the power spectral density is significantly susceptible to the presence of spikes, and an improvement in the quality of the contaminated time series is directly reflected in the spectra. By removing the spikes they recovered the low-frequency region of the spectra, which is the region with the highest energy content. The high-frequency region cannot be accurately recovered, due to the steps introduced in the time series by the application of the last valid data point for signal reconstruction.

4 The Autocorrelation Algorithm for the ADV Velocity Data

Velocimeter uncertainty is a random white noise. This means that each velocity estimate is independent of the next, and it means that the noise has no preferred frequency. The noise level varies widely according to the characteristics of the velocimeter, its set up, the characteristics of the flow, and the acoustic scattering environment. The velocity noise level limits the frequency range over which useful data can be obtained. However, the velocity spikes are a major source of error as they can have large values compared with the mean velocity and are more difficult to be detected, removed and replaced by a clean signal. Due to the presence of intermittent spikes contaminating the time series measured by an acoustic Doppler velocimetry, statistical properties, and the power spectral density of such data sets can have unrealistic values (Parsheh et al. 2010). These spikes appear when the flow velocity exceeds the pre-set velocity range of the equipment, the turbulence intensities are high, or where there are contaminations from previous pulses reflected from the flow boundaries.

The smoothing filter developed herein is a linear minimum mean square error (LMMSE) estimator. To estimate a parameter φ based on the data vector $\boldsymbol{x}$, defined as equation (1), LMMSE is the best estimator in the sense that among the class of all linear estimators of the form of the equation (2), it minimises the Bayesian mean square error (MSE), defined in equation (3), and only relies on the correlation between the random variables (Kay 1993), with the key equations being given as:.

$$\boldsymbol{x} = \left[x[0], x[1], \ldots, x[N-1] \right]^T \tag{1}$$

$$\hat{\varphi} = \sum_{n=0}^{N-1} a_n \, x[n] \tag{2}$$

$$\mathrm{Bmse}\left(\hat{\varphi}\right) = \mathrm{E}\left[\left(\varphi - \hat{\varphi}\right)^2 \right] \tag{3}$$

where N is the number of data points, $\hat{\varphi}$ is an estimation for φ and a_ns are coefficients. For noise reduction of the velocity signals, the variable $\boldsymbol{v}$, which is the true velocity vector, should be estimated based on the noisy recorded signal $\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{n}$, where $\boldsymbol{n}$ is the Doppler noise, which is believed to be a Gaussian white noise (Nikora and Goring 1998).

4.1 Exploiting Correlations in Time

The smoothing filter finds coefficients $w_{i,j}$ to best estimate the variable:

$$\hat{v}_i = \sum_{j=0}^{N-1} w_{i,j} \, v_j \tag{4}$$

such that for each $i \in \{0, \ldots, N-1\}$ the Bayesian mean square error is minimised to give:

$$\text{Bmse}\left(\hat{v}_i\right)=\text{E}\left[\left(v_i-\hat{v}_i\right)^2\right] \tag{5}$$

In the vector form $\hat{\boldsymbol{v}}=\overline{\boldsymbol{W}}\,\boldsymbol{x}$, where $\overline{\boldsymbol{W}}$ is an $N\times N$ matrix, where N is the number of data points. Assuming the Doppler noise vector $\boldsymbol{n}$ and the true velocity signal $\boldsymbol{v}$ are uncorrelated, and then solving for $\overline{\boldsymbol{W}}$ to minimize equation (5) reduces to:

$$\overline{\boldsymbol{W}}=\overline{\boldsymbol{R}}_{vv}\left(\overline{\boldsymbol{R}}_{vv}+\overline{\boldsymbol{R}}_{nn}\right)^{-1} \tag{6}$$

where $\overline{\boldsymbol{R}}_{vv}$ and $\overline{\boldsymbol{R}}_{nn}$ denote the autocorrelation matrices for the true velocity and the Doppler noise, respectively (Kay 1993). Equation (6) can then be rewritten as follows:

$$\overline{\boldsymbol{W}}=\left(\overline{\boldsymbol{R}}_{xx}-\overline{\boldsymbol{R}}_{nn}\right)\overline{\boldsymbol{R}}_{xx}^{-1} \tag{7}$$

4.1.1 Autocorrelation Matrices

For the case that the velocity signal $\boldsymbol{v}$ and the additive noise $\boldsymbol{n}$ are considered to be wide sense stationary signals, all the matrices would be symmetric Toeplitz matrices. Hence, $\overline{\boldsymbol{R}}_{xx}$ can be approximated using samples of the recorded data as follows:

$$\overline{\boldsymbol{R}}_{xx}\left(i,j\right)=\overline{\boldsymbol{R}}_{xx}\left(\left|i-j\right|\right)=\text{E}\left[\left(x_i-\overline{\text{E}}_x\right)\cdot\left(x_j-\overline{\text{E}}_x\right)\right]=\frac{1}{N-1-\left|i-j\right|}\sum_{k=0}^{N-1-\left|i-j\right|}\left(x_k-\overline{\text{E}}_x\right)\cdot\left(x_{k+\left|i-j\right|}-\overline{\text{E}}_x\right) \tag{8}$$

The assumption of Doppler noise being white Gaussian noise, imposes $\boldsymbol{n}$ to be a wide sense stationary signal, and results in $\overline{\boldsymbol{R}}_{nn}$ being a diagonal matrix with diagonal elements of noise variance σ_n^2. To approximate the noise variance, the DFT of the signal $\boldsymbol{x}$ was calculated. Based on these calculations, the higher frequencies only contain white noise, as the velocity signal is wrapped around lower frequencies, and using the fact that white noise is spread almost equally over the spectrum, an approximate value for $\hat{\sigma}_n^2$ may be calculated.

To compute the noise power for each point, a high-pass filter was designed to reject the main signal and only let part of the white noise go through. Figure (1) shows the frequency spectrum of velocity sequence of a point, for which it is expected that for frequencies more than 0.2 Hz the signal only contains additive white Gaussian noise. From this frequency spectrum, which is computed by using a Fast Fourier Transform (FFT), it is evident that the velocity signal exhibits a low pass signal behaviour. This means that the frequency content of the velocity signal is wrapped around lower frequencies, while the frequency content of the white noise is equally spread all over the frequency spectrum. Therefore the power of the signal for frequencies more than 0.2 Hz is calculated and is extended for the whole frequency spectrum.

4.1.2 Fourier Transform Interpretation

For the smoothing problem, in which the sequence v_n is to be estimated based on $\left\{\ldots,x[-1],x[0],x[1],\ldots\right\}$ or $x[k]$ for all k, corresponding to infinite impulse response non-causal filtering, the smoothing estimator takes the form of equation (9):

$$\hat{v}_n=\sum_{k=-\infty}^{\infty}w_{k,n}\,x_k=\sum_{k=-\infty}^{\infty}w\left(\left|n-k\right|\right)x_k=w_n*x_n \tag{9}$$

where * denotes convolution. Equation (7) represents a set of linear equations to be solved for the filter impulse response values. These values correspond to a Finite Impulse Response (FIR) filter of length N. In the case of an Infinite Impulse Response (IIR) filter, equation (4) translates to equation (9). However, coefficients $w_{k,n}$ cannot simply be solved using the method of linear equations. Therefore, to find the FIR filter which minimises the mean squared error, a Fourier transform method shall be used. The coefficients w_n in equation (9) are then solved as follows:

$$\sum_{k=-\infty}^{\infty}w\left(\left|n-k\right|\right)\boldsymbol{R}_{xx}\left(k\right)=w_n*\boldsymbol{R}_{xx}\left(n\right)=\boldsymbol{R}_{vv}\left(n\right) \tag{10}$$
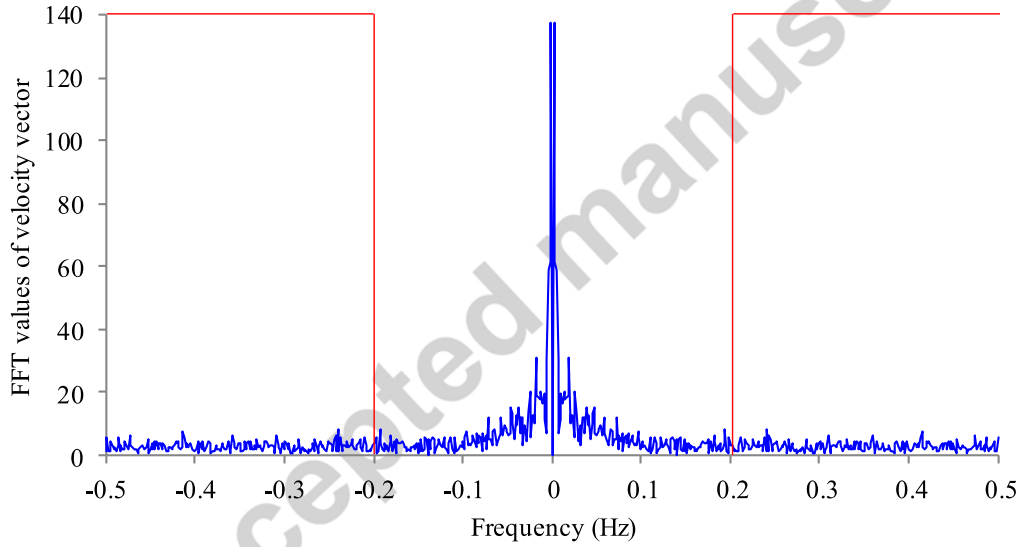
7

Denoting the frequency response of the infinite smoother by $H(f)$, where f is the frequency, the Fourier transform of equation (10) results in the following representation:

$$H(f) = \frac{P_{vv}(f)}{P_{xx}(f)} = \frac{P_{vv}(f)}{P_{vv}(f) + P_{nn}(f)} = \frac{\dfrac{P_{vv}(f)}{P_{nn}(f)}}{\dfrac{P_{vv}(f)}{P_{nn}(f)} + 1} \tag{11}$$

$P_{vv}(f)$, $P_{xx}(f)$ and $P_{nn}(f)$ correspond to the spectral density of discrete time random processes v, x and n respectively, which are the Fourier transform of the autocorrelation series of each of the processes, namely $R_{vv}(n)$, $R_{xx}(n)$ and $R_{nn}(n)$. The Fourier transform of discrete series $R_{vv}(n)$, for example, is as follows:

$$P_{vv}(f) = \sum_{n=-\infty}^{\infty} R_{vv}(n) e^{-i2\pi fn} \tag{12}$$

Hence, the smoother filter emphasises portions of the frequency spectrum of the data where signal to noise ratio (SNR) is high and attenuates those where it is low. The proposed filter, among all possible low pass filters will select the optimum low pass filter in the sense of minimising the mean squared error. To clarify this, the smoothing frequency impulse response is $H(f) \approx 0$ for a low local SNR, defined as $\eta(f) = P_{vv}(f)/P_{nn}(f)$, and $H(f) \approx 1$ for a high local SNR.
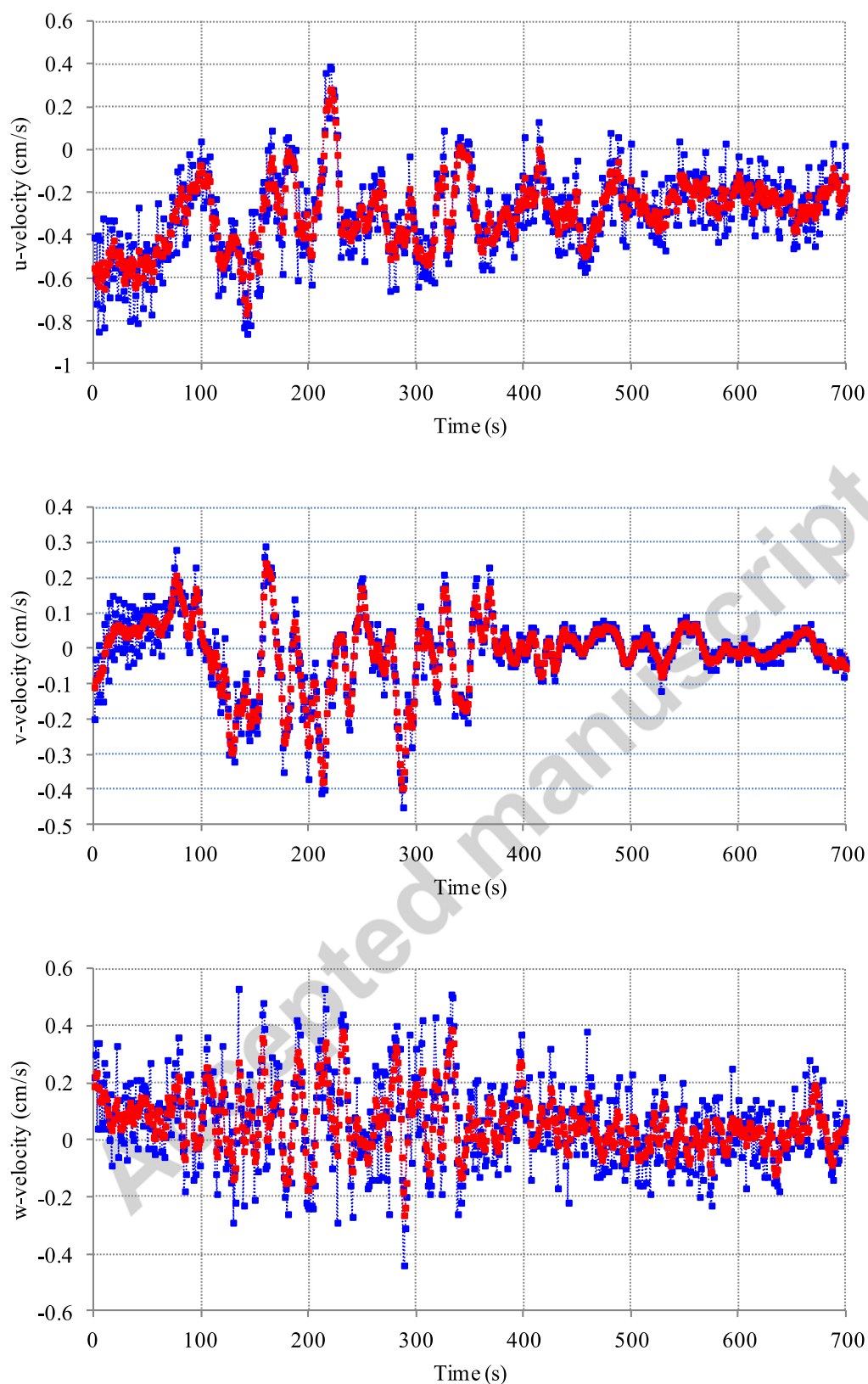

Fig. 1 Frequency spectrum for u-component of velocity data, recorded at a point 24 cm from the centre of the harbour, at 100 mm above bed for saline water, with a tidal range of 100 mm

4.1.3 Velocity Filtered Results

A filter with a window size of 5 has been applied to the u, v and w components of the noisy velocity vector, collected in the model harbour at a point 24 cm from the centre of the harbour and at 10 cm above the bed, for a tidal range of 100 mm and for saline water. For the u-velocity the estimated value of σ_n^2 equates to 0.0094, which corresponds to an estimated SNR = 5.28 dB. The filter coefficients are calculated as follows:

$$\mathbf{h} = \big[h[-2], h[-1], h[0], h[1], h[2]\big] = [0.0616, 0.1410, 0.4804, 0.1410, 0.0616] \tag{13}$$

Fig. 2 Noisy (blue) and filtered (red) velocity time series for a point 24 cm from the centre of the harbour, at a layer 100 mm above the bed for saline water status, with a tidal range of 100 mm

which are symmetric with respect to the central point and emphasise lower frequencies correlations as expected. The filtering then proceeds as follows:

$$\hat{v}_i = \sum_{j=i-2}^{j=i+2} h_{j-i}\, v_j \tag{14}$$

The filtered data for the u, v and w velocity components are plotted in figure (2). The spikes are more influenced than the noisy signal, as was to be expected.

4.1.4   Assessment of the Method Using an Autoregressive Model

For the assessment of the method, an autoregressive model has been used to generate a clean velocity signal. The autoregressive (AR) model is a representation of a type of random process; as such, it describes many time-varying processes in nature. The rationale behind using the autoregressive model is that the random process generated using this model will be stationary for a suitable choice of parameters. Therefore, the covariance matrix for the process will be well-defined. The clean signal has been generated using the following equation:

$$X_t = \sum_{i=1}^{p} \rho_i\, X_{t-i} + \varepsilon_t \tag{15}$$

where X_t is the clean signal, the velocity, for example, $\rho_i, i \in \{1, \cdots, p\}$ are the parameters of the model, and ε_t is white noise. Parameters used to generate the velocity signal are $p = 2, \rho_1 = 0.7, \rho_2 = 0.2$ and ε_t with unit variance. The spikes are generated with a uniformly random time index and a Gaussian distributed value, with a variance equal to 20. White Gaussian noise is added to this simulated signal, such that the SNR before applying the correlation algorithm was equal to 5.28 dB. After filtering with a window size equal to 5, the spikes were smoothed and the SNR for the filtered signal was computed to be 7.78 dB, which indicates an improvement of about 2.50 dB in the SNR. Using the method of Goring and Nikora (2002), as modified by Wahl (2002), the samples were removed in large portions of the signal. The performance of the linear correlation algorithm for a velocity signal generated by using an autoregressive model is illustrated in figure 3. The clean signal was generated with 10,000 samples, whereas the figure is plotted for 200 signals to produce higher resolution graphs. However, the computations included all samples.
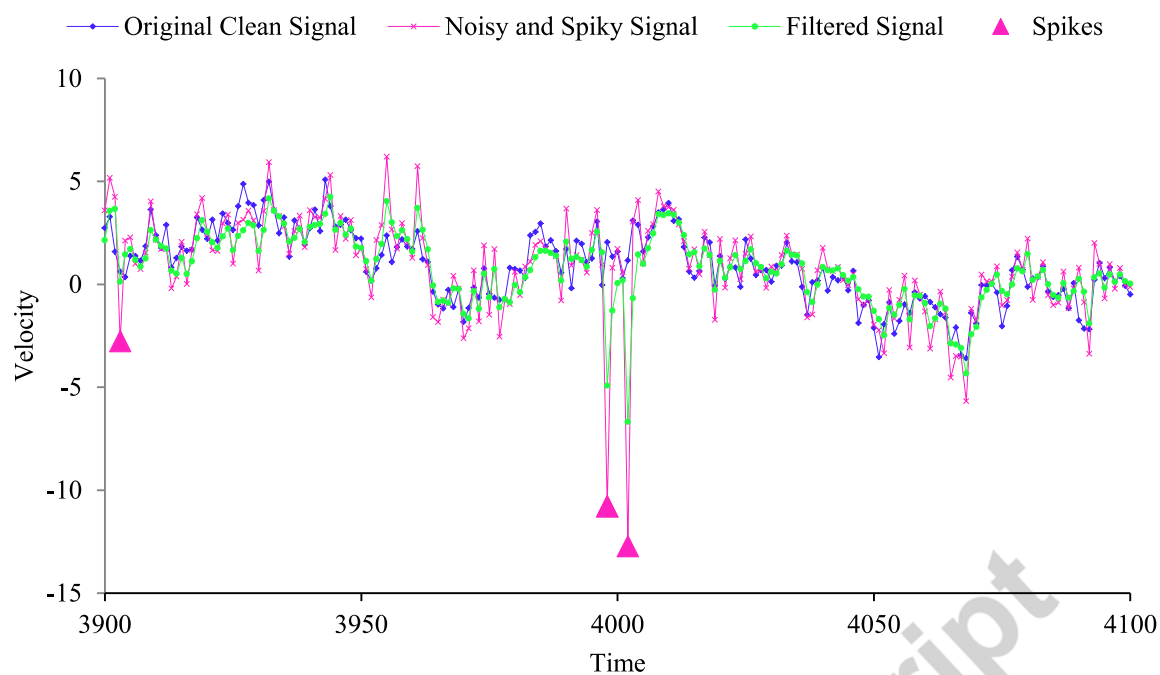
To better observe how the filter reduces the squared error, the empirical cumulative distribution function (ECDF) for the noisy and filtered signals are plotted in figure 4 using a vertical logarithmic scale. The squared error at any time index t for a given signal Y with respect to a clean signal X is the squared difference between the given signal and the clean signal in that time index, which is presented in equation (16) as follows:
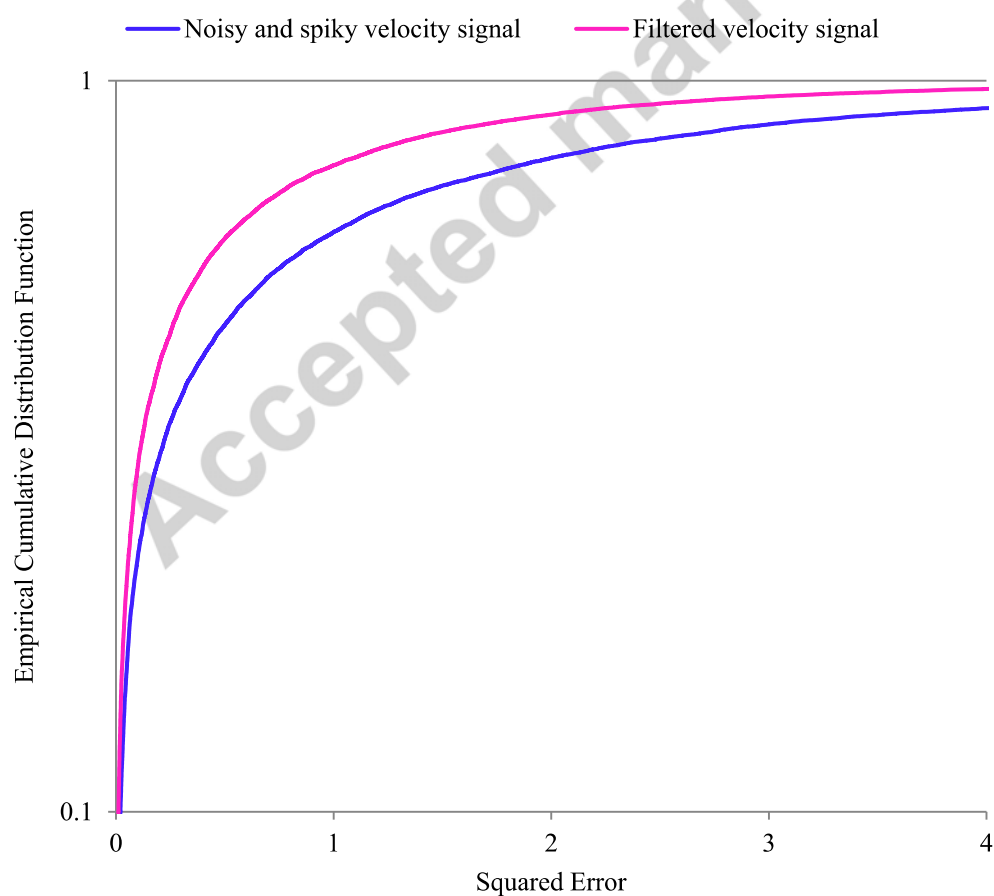
$$e_t = (X_t - Y_t)^2 \tag{16}$$

When X_t and Y_t are both stochastic processes, the sequence e_t will in turn be a stochastic process itself. As explained in Section (4.1.2) the proposed filter aims to minimise the expected value of e_t at any given time index t when Y is replaced with the filtered signal. This is due to the fact that the less the value of e_t becomes, the better signal Y represents the clean signal X. As for any other random variable, the squared error possesses a cumulative distribution function (CDF). The cumulative distribution function (CDF) of a random variable X, evaluated at x, is the probability that X takes a value less than or equal to x. The distribution function for a discrete variable is related to a discrete probability $P(x)$ and is represented as follows:

$$D(x) = P(X \le x) = \int_{y \le x} p_X(y)\, dy \tag{17}$$

where $p_X(y)$ denotes the probability density function for random variable X. The empirical cumulative distribution function estimates the cumulative distribution function of the squared error of the contaminated and filtered samples. Smaller values for random variable e_t translates to faster convergence of CDF to unity. Therefore, the faster CDF approaches unity the better signal Y is representing the clean signal X. It is noticeable that the ECDF for the filtered signal approaches 1 much faster than the ECDF for noisy signal, which shows that a larger percentage of the filtered signals possess less of a square error when compared to the noisy signal.

10

Fig. 3 Performance of the linear correlation algorithm for a velocity signal generated using an autoregressive model.

Fig. 4 Empirical cumulative distribution function for the noisy and filtered signals

11

4.1.5 Assessment of the Method Using Signals Produced by a Three-dimensional Numerical Model

Another method of assessment, using a three-dimensional numerical model which is an extension to the two-dimensional vertical (2DV) numerical model developed by Hejazi et al. (2013), has been deployed to generate a time series of the velocities at the points where measurements were undertaken. The simulated time series are free of noise and represent the velocity variations at each and every point of the measurement grids described in Section 2. The simulated signals were generated for 7080 samples, which equates to ten periods of the tidal period of the measurements, which was calculated to be equal to 708 seconds.

Noise and spikes were added to the simulated signals, and the autocorrelation algorithm was then applied to the contaminated signals and the performance of the new algorithm was examined. Figure (5) shows the performance of the correlation algorithm for two velocity signals, originally generated by the numerical model. The figures are plotted for 200 signals to produce higher resolution graphs, but the computations included all samples. It can be seen that the spikes are removed from the signals, and the filtered signals almost lie on the original simulated signals. To better visualise this observation, figure (6) exhibits a shorter period of time, which demonstrates the second series of signals (b), and shows that the variations of the filtered signal for this fluctuating series of signals are quite close to the variations of the originally simulated clean signals.

The empirical cumulative distribution functions (ECDF) for the contaminated and filtered signals are plotted in figure 7 using a vertical logarithmic scale. It is noticeable that the ECDFs for the filtered signals approach 1 much faster than the ECDFs for noisy signals. Table (1) tabulates the squared error of the contaminated and filtered signals for a range of probabilities. It is evident that a larger percentage of the filtered signals possess much less of a square error compared to the noisy signals. Also for the value of the squared error of the filtered signals the corresponding probability of the contaminated signals are tabulated. To clarify this, for the probability of 0.50 (for example), the squared error is equal to 0.0043 for the filtered signal, while this value corresponds to the probability of 0.251 for the contaminated signal. The average difference in the probabilities for the contaminated and filtered signals, for the range of probabilities tabulated in Table (1), is equal to 0.299 and 0.248 for the signals (a) and (b) respectively.

Table 1 Squared error of contaminated and filtered signals and equivalent probability of contaminated signals for the values of squared error of filtered signals
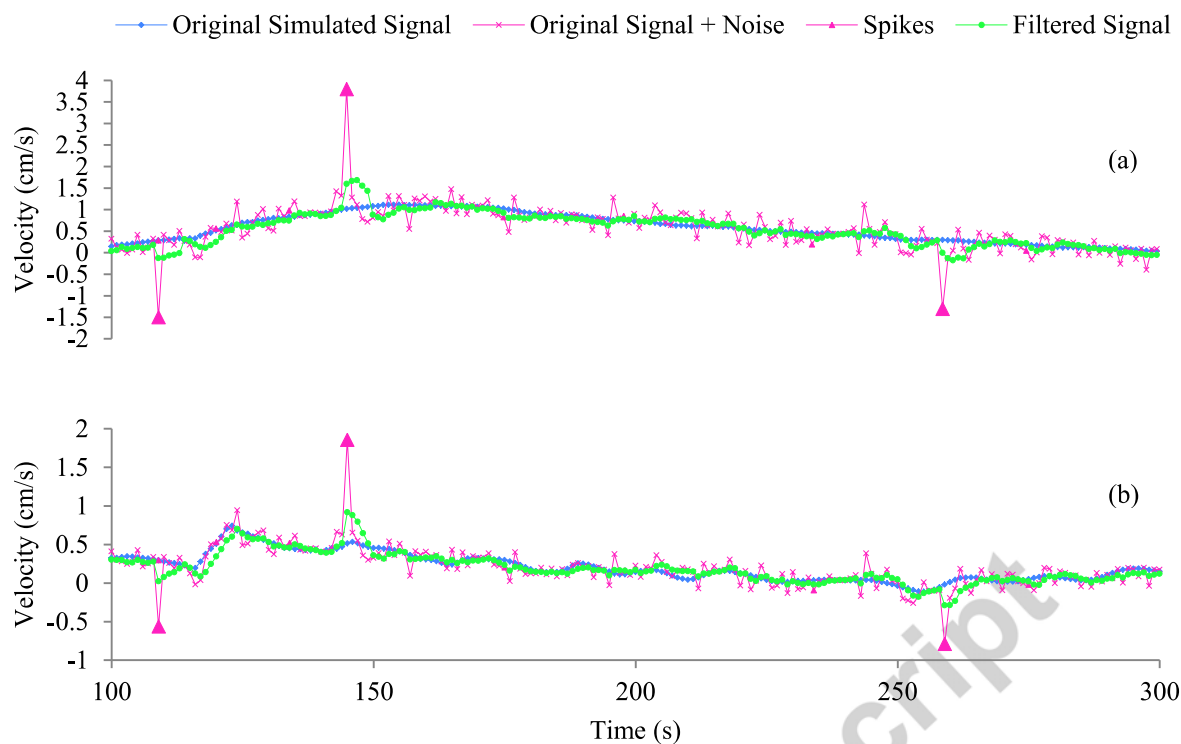
Probability	Squared error contaminated signal (a)	Squared error filtered signal (a)	Equivalent probability of contaminated signal (a)	Squared error contaminated signal (b)	Squared error filtered signal (b)	Equivalent probability of contaminated signal (b)
0.50	0.0197	0.0043	0.251	0.0046	0.0012	0.276
0.60	0.0312	0.0067	0.304	0.0072	0.0019	0.336
0.70	0.0489	0.0103	0.371	0.0113	0.0029	0.413
0.80	0.0738	0.0158	0.454	0.0171	0.0046	0.503
0.90	0.1244	0.0284	0.577	0.0288	0.0087	0.639
0.95	0.1840	0.0492	0.701	0.0427	0.0169	0.797

The initial signal-to-noise ratio for the contaminated signals after adding noise and spikes was calculated to be 5.28 dB. After applying the autocorrelation algorithm, the SNR was computed to be equal to 10.325 and 9.324 for the signals (a) and (b) respectively. An analysis was also undertaken to investigate the response of the filter to different values of SNR. The initial signal-to-noise ratios were set equal to 0, 3 and 10, and the corresponding filtered SNRs were computed. This analysis, which was conducted for 17 signals, showed that the filter exhibits maximum filtering gain when the initial SNR is equal to 0 dB. The study also showed that the average increase of the SNR for the case of initial SNR = 0, was calculated to be 6.88 dB.
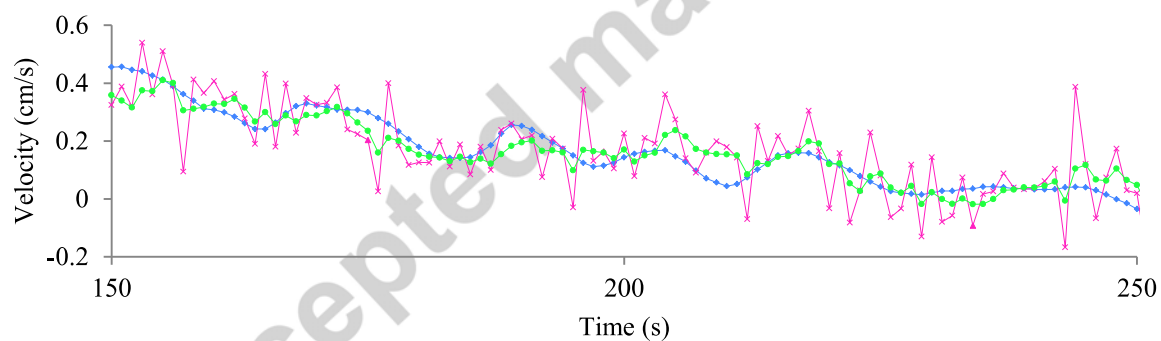

4.2 Exploiting Correlations in Space

The same approach can be used to take advantage of correlation of the velocity samples located at different positions in the grid and/or at different depths. In this case the parameter to be estimated is, for example, $v(p_1, i)$, based on a collected noisy random signal $\{x(p_1, i), x(p_2, i), \ldots, x(p_n, i)\}$, where $(p_., i)$ denotes the position and time indices respectively. The symmetric matrix $\overline{R}_{xx}$ is then as follows:
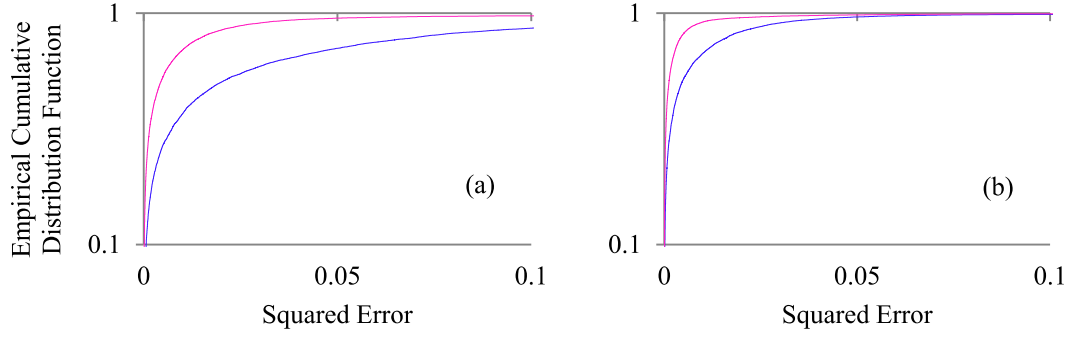
$$\overline{R}_{xx}(p_m, p_n) = \mathrm{E}\left[\left(x(p_m, .) - \overline{\mathrm{E}}_x\right) \cdot \left(x(p_n, .) - \overline{\mathrm{E}}_x\right)\right] = \frac{1}{N-1} \sum_{i=0}^{N-1} \left(x(p_m, i) - \overline{\mathrm{E}}_x\right) \cdot \left(x(p_n, i) - \overline{\mathrm{E}}_x\right) \qquad (18)$$

Fig. 5 Performance of the correlation algorithm for velocity signals originally generated by a numerical model

Fig. 6 Performance of the correlation algorithm for the velocity signal (b) generated by a 3D numerical model.

Fig. 7 ECDFs for the contaminated and filtered signals originally generated by a numerical model

Again the correlation matrix $\overline{R}_{nn}$ is diagonal. However, diagonal elements may not necessarily be equal, as the signal to noise ratio might vary along different axes/depths/locations. Accordingly, the best linear estimate for $\hat{v} = \left[v(p_1,i), v(p_2,i), \ldots, v(p_n,i) \right]$, based on noisy data $x = \left[x(p_1,i), x(p_2,i), \ldots, x(p_n,i) \right]$, at each and every time instance, can be computed as follows:

$$\hat{v} = \left(\overline{R}_{xx} - \overline{R}_{nn} \right) R_{xx}^{-1} \, x \tag{19}$$

The LMMSE estimator relies on the correlation between random variables, where the uncorrelated parameters cannot be estimated linearly. Therefore, to achieve a more effective noise reduction, the elements of vector *x* should be chosen such that they establish a strong correlation. The methodology can be expanded to two, three, and four dimensions, to include the integrated one-, two- and three-dimensional spatial and temporal correlations.

4.3 Interpolation

In many experimental programmes, the number of points to be measured is chosen with economic considerations in mind, or the sampling frequency may be limited due to instrumental specifications. The smoothing filter may also be used to interpolate missing data points by computing the correlation coefficients between the existing and missing data, either to obtain a higher frequency for a quantity which has been recorded with a low frequency (e.g. velocity) or to estimate a variable in a different place in space (e.g. velocity at a different depth) where no measurements have been taken.

In the following, performance of a smoothing interpolation filter is examined. For a set of data points and for a window length $(2w + 1)$, the correlation between all odd time indexed and even time indexed velocity points are calculated to obtain interpolation coefficients. Then assuming, for example, even time indexed data points are missing, they are estimated as follows:

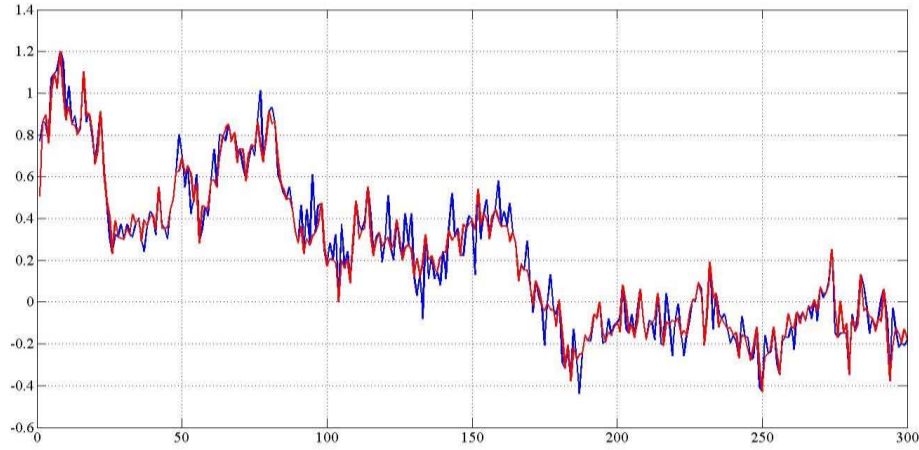$$\hat{v}_{2k} = \sum_{j=2i+1, i=-w,\ldots,+w} h_{2k, 2k+j} \, v_{2k+j} \tag{20}$$

Vector *h* is calculated as follows:

$$h = \overline{R}_{oo}^{-1} \, R_{eo} \tag{21}$$

in which $\overline{R}_{oo}$ is the correlation matrix of odd time indexed velocity points and R_{eo} is the vector of correlation between odd and even indexed velocity points. The signal to noise ratio, based on the definition given in equation (21) assuming a sequence with zero mean value, for the interpolated signal was calculated to be about 10.96 dB. This meant that the interpolation error variance was about 0.1 signal power. Figure (8) shows the interpolated signal.

14

$$\text{SNR} = \frac{\text{var}\left(x^e - \hat{x}^e\right)}{\text{var}\left(x^e\right)} = \frac{\sum_i (x_i^{\,e} - \hat{x}_i^{\,e})^2}{\sum_i (x_i^{\,e})^2} = \frac{\sum_i (x_{2i} - \hat{x}_{2i})^2}{\sum_i (x_{2i})^2} \qquad \begin{array}{ll} x_i^{\,e} = x_{2i}, & \forall i \\[2mm] x_i^{\,o} = x_{2i-1}, & \forall i \end{array} \tag{22}$$
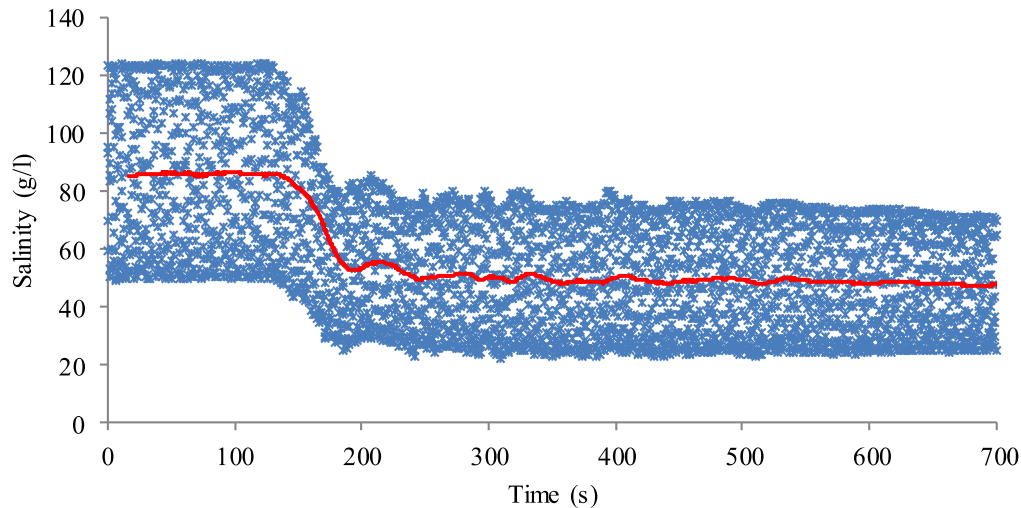
The SNR is a measure to assess the accuracy of the filter estimation (interpolation) for the missing data points. To compute the SNR, using a sequence of data points, denoted by x, with all even and odd indexed data points available, denoted by x^e and x^o respectively, it was assumed that the even data points were missing and then an estimate of the even indexed data points, $\hat{x}^e$, were calculated based on equation (20) and using only the odd indexed points, namely x^o. The SNR is an indication to show the degree of difference of the estimated $\hat{x}^e$ from the actual x^e measured values.

Fig. 8 Interpolated data (red) based on the recorded samples (blue) for the central point of the harbour at a depth of 2 cm above bed, for a tidal range of 100 mm and saline water

5 Moving Average Strategy for the Salinity Concentration Data

The salinity recorded data also suffered from very noisy signals. However, the nature of the noise for the salinity data was such that the signal magnitudes fluctuate around a mean value, with almost a constant band of fluctuations (Fig. 9), therefore the filtering process was less complicated. A backward moving average strategy was adapted for the signals of conductivity recorded with a sampling frequency of 5 Hz. Before the tidal water flushed into the harbour, it was expected that the salinity remained almost constant, as no forcing influenced the water inside the harbour, and only slight local movements due to stirring the water were expected to produce small motions, which would not change the salinity patterns. Therefore the span for the moving average algorithm was chosen such that for each set of data the first part showed a uniform behaviour. The result for a point at a distance of 24 cm from the centre of the harbour, at layer 180 mm above bed and which exhibited the most salinity exchange is plotted for the tidal range of 100 mm in figure (9).

15

Fig. 9 Moving average curve of the noisy salinity samples collected for a point at a distance of 24 cm from the centre of the harbour, and at 180 mm above the bed, for a tidal range of 100 mm

6 Conclusions

A laboratory tidal basin was designed, set up and deployed for measuring the velocity components in three directions, water surface elevations and salinity concentrations, for an idealised vertically distorted model harbour, for fresh water and salinity stratified conditions, and with various barrier heights, tidal ranges and mean water depths. Constant amplitude and constant period sinusoidal model tides were reproduced in the tidal tank by means of a vertically oscillating weir. ADVs and control software were used for measuring the velocity components. For measuring the water surface elevations and conductivity, GENIE data acquisition and control software were adapted and used accordingly.

In turbulent flow field measurements or where the shear layers due to stratification in combination with turbulence cannot be fully recognised, noisy data may lead to incorrect conclusions. For despiking and denoising the velocity data a linear correlation algorithm was established, as reported herein, and a filter was designed and adapted for the temporal correlation of the signals, which successfully lowered the noise level and removed the spikes where the stationary assumption of the statistical characteristics of the velocities were valid. The filter can be extended to the spatial domain by establishing one-, two- or three-dimensional correlations.

For the assessment of the method, an autoregressive model was used to generate a clean velocity signal. The spikes were generated with a uniformly random time index and a Gaussian distributed value. White Gaussian noise was added to this simulated signal. After filtering, the spikes were smoothed and the SNR for the filtered signal indicated about 47% improvement in the SNR. Assessment was also undertaken on the signals which were generated by a three-dimensional numerical model. After adding noise and spikes to the simulated signals, by applying the autocorrelation algorithm to the contaminated signals the spikes were removed, and the filtered signals almost lay on the original simulated signals, and the SNR increased by about 5 dB.

In many experimental programmes, the number of points to be measured is chosen with economic considerations in mind, or the sampling frequency may be bound by instrument limitations. An interpolation method for producing the missing data has been provided, which may be deployed for increasing the sampling frequency or producing data for the spatial domain. The smoothing filter also may be incorporated for smoothing the numerical simulation outputs as a post-processing tool.

For the conductivity a moving average strategy has been amended as the data did not suffer from spikes, but exhibited almost a constant band of noisy fluctuations. The water elevation data showed reasonably clean signals and no filtering procedure was carried out for these data.

7 Acknowledgements

16

References

ADVANTECH (1995) GENIE reference manual, Version 2.0, American Advantech Corporation, USA.

Chanson H, Trevethan M, Aoki S-i (2008) Acoustic Doppler velocimetry (ADV) in small estuary: field experience and signal post-processing. Flow Measurement and Instrumentation 19:307-313

de Nijs MAJ, Winterwerp JC, Pietrzak JD (2009) On harbour siltation in the fresh-salt water mixing region. Continental Shelf Research 29:175-193

Dombroski DE, Crimaldi JP (2007) The accuracy of acoustic Doppler velocimetry measurements in turbulent boundary layer flows over a smooth bed. Limnol Oceanogr: Methods 5:23-33

Doroudian B, Bagherimiyab F, Lemmin U (2010) Improving the accuracy of four-receiver acoustic Doppler velocimeter (ADV) measurements in turbulent boundary layer flows. Limnol Oceanogr: Methods 8:575-591

Falconer RA, Li G (1994) Numerical modelling of tidal eddies in coastal basins with narrow entrances using the k-ε turbulence model. In: Beven KJ, Chatwin PC, Millbank JH (eds) Mixing and transport in the environmental. John Wiley & Sons, London, pp 325-350

Falconer RA, Yu G (1991) Effects of depth, bed slope and scaling on tidal currents and exchange in a laboratory model harbour. Proceedings of Institution of Civil Engineers, Part2, Research and Theory 91:561-576

Garbini JL, Forster FK, Jorgensen JE (1982) Measurement of fluid turbulence based on pulsed ultrasound techniques, Part 1. Analysis. J Fluid Mech 118:445-470

Goring DG, Nikora VI (2002) Despiking acoustic Doppler velocimeter data. J Hydraul Eng, ASCE 128(1):117-126

Hejazi K, Soltanpour M, Sami S (2013) Numerical Modelling of Wave-Mud Interaction Using Projection Method. Ocean Dynamics 63(9-10):1093-1111

HR Wallingford (1994) Wave probe monitor, operating instruction, Issue 5. HR Wallingford Ltd, Wallingford.

Hurther D, Lemmin U (2001) A correction method for turbulence measurements with a 3D acoustic Doppler velocity profiler. J Atmos Oceanic Technol 18(3):446-458

Jenway Ltd (1996) Model 4320 conductivity meter operating manual. Jenway Ltd, England.

Kay SM (1993) Fundamentals of statistical signal processing, estimation theory, Vol 1. Prentice Hall PTR, New Jersey.

Khorsandi B, Mydlarski L, Gaskin S (2012) Noise in turbulence measurements using acoustic Doppler velocimetry. J Hydraul Eng, ASCE 138(10):829-838

Kraus NC, Lohrmann A, Cabrera R (1994) New acoustic meter for measuring 3D laboratory flows, J Hydraul Eng 120:406-412

Lemmin U, Lhermitte R (1999) Discussion of "ADV measurements of turbulence: can we improve their interpretation?" By Nikora VI, Goring DG. J Hydraul Eng, ASCE 125(9):987-988

Lohrmann A, Cabrera R, Kraus NC (1994) Acoustic Doppler velocimeter (ADV) for laboratory use. In: Pugh CA(ed) Proc Symp on fundamentals and advancements in hydraulic measurements and experimentation, ASCE:351-365.

McLelland SJ, Nicholas AP (2000) A new method for evaluating errors in high-frequency ADV measurements. Hydrological Processes 14:351-366.

Mardapitta-Hadjipandeli L, Falconer RA (1990) Some observations on nested modelling of flow and solute transport in rectangular harbours. Proc Institution of Civil Engineers, Part 2, Research and Theory 89(1):15-38.

Nece RE, Falconer RA (1989) Hydraulic modelling of tidal circulation and flushing in coastal basins. Proc of the Institution of Civil Engineers, Part 1, Design and Construction 86(5):913-935.

Nikora VI, Goring DG (1998) ADV measurements of turbulence: can we improve their interpretation? J Hydraul Eng, ASCE 124(6):630-634.

NORTEK AS (1997a) ADV operation manual. NORTEK AS, Norway.

NORTEK AS (1997b) ADV software manual, ver. 2.6. NORTEK AS, Norway.

Parsheh M, Sotiropoulos F, Porté-Agel F (2010) Estimation of power spectra of acoustic-Doppler velocimetry data contaminated with intermittent spikes. J Hydraul Eng, ASCE 136(6):368-378.

Voulgaris G, Trowbridge JH (1998) Evaluation of the acoustic Doppler velocimeter (ADV) for turbulence measurements. Journal of Atmospheric and Oceanic Technology 15:272-289.

Wahl TL (2002) Despiking acoustic Doppler velocimeter data, Discussion. J Hydraul Eng, ASCE, 128: 117-126.

Wahl TL (2003) Despiking acoustic Doppler velocimeter data, Discussion. J Hydraul Eng, ASCE, 129(6):484-487.

Yin J, Lloyd PM, Falconer RA (2001) Tidal flow local mean velocities in noisy signals. Flow Measurement and Instrumentation 12:25-28.

Highlights

- An optimum filter has been designed, based on minimising the mean square error.
- Filter coefficients rely on empirical correlation for the data, averaged over time samples.
- Noisy parts of the frequency spectrum are attenuated and the desired parts are amplified.
- The method interpolates between data points in time, to increase the sampling frequency.
- The concept may be applied to estimate the missing data points in space.

18