ACCURATE LDA-SPECTRA BY RESAMPLING AND ARMA-MODELING

S. de Waele and P.M.T. Broersen
Department of Applied Physics Delft University of Technology
P.O.Box 5046 2600 GA Delft The Netherlands
Phone + 31 15 2786419, Fax + 31 15 2784263, E-mail: broersen@tn.tudelft.nl

ABSTRACT

With Laser-Doppler Anemometry (LDA) the velocity of gases and liquids is measured without disturbing the flow. The velocity signal is sampled at irregular intervals; a regularly resampled signal is extracted using Nearest Neighbor Resampling or Linear Interpolation. From the resampled data the spectrum is estimated using AR-MA time series modeling. AR-MA modeling yields a more accurate description of the spectral structure than the best Windowed Periodogram. The accuracy of the spectral description is established with an objective measure: the model error at time scale T.

1. INTRODUCTION

Laser-Doppler Anemometry (LDA) is used to measure the velocity of gases and liquids with observations irregularly spaced in time. The interval between observations can depend on the current value of the velocity [1], which gives rise to bias problems. Those irregular intervals and bias problems are challenges for modern statistical signal processing.

Variance considerations give a preference for equidistant processing algorithms. A regularly sampled *signal* can be extracted using resampling, for which three methods will be compared: linear interpolation, zero order hold and nearest neighbor resampling. Resampling will be compared to slotting, which provides a regularly spaced estimate of the *autocovariance* function [2].

The power spectrum of the velocity signal is estimated using time series modeling. The application of time series models to irregularly sampled signals has been studied before [3]. In this paper, the time series model is estimated using the AR-MA algorithm [4]. With the AR-MA algorithm, a single time series model is selected using only the measured data. The power spectrum is calculated from the parameters of the selected model. This estimate is an alternative to the windowed periodogram.

The performance of resampling and AR-MA-modeling is evaluated in simulations with a new quality criterion: the model error at time scale T. This is a useful statistical measure for resampled signals.

2. DIRECT METHODS OR RESAMPLING

The techniques that have been proposed for estimating LDA-spectra can be divided into two classes [5]: direct methods and resampling.

Every couple of LDA observations provides an estimate of the autocovariance function at one specific lag time equal to the time between the two observations. The direct methods are based on these estimates of the autocovariance function. The direct transform [6] is a Fourier transform of all individual estimates of the autocovariance function which are found in the manner described above. When this method is applied to regularly sampled data it is similar to the periodogram estimate of the spectrum. The variance of the periodogram is equal to the square of the true spectrum [7]. The direct transform includes this variance, but on top of that it has a constant contribution to the variance [6]. This means that, in areas of the spectrum with low power, the standard deviation of the spectral estimate is much larger than the expectation of the spectrum. This explains statistically why negative values can be found as estimates for the spectral density. Therefore, the direct transform is incapable of accurately estimating the spectrum in areas with low power.

The slotting technique is based on the direct transform. In order to get a more accurate estimate of the autocovariance function, the estimates in the interval $[(k-1/2)\Delta, (k+1/2)\Delta]$ around lag time $k\Delta$ are averaged. This is an improvement over the direct transform. All direct methods have in common that the estimate for the spectrum can take on negative values. This estimate is no power spectrum because power spectra are positive by definition. In order to allow some kind of comparison we took the absolute value of the spectrum estimated with slotting to obtain the power spectra shown in Fig.1. Slotting, like the direct transform, is incapable of accurately estimating the spectrum in areas with low power. Fig.1 shows that the level of the slotting spectrum is about the same for both examples in the whole frequency range, whereas the true spectrum and the spectra of resampled signals give a significant difference at higher frequencies.

The second class of methods for estimating LDA-spectra are resampling methods. With these methods an equidistantly sampled representation of the signal is constructed at the resampling times. For a statistical interpretation, the maximum resampling rate should not be higher than the average irregular sampling rate. Otherwise the number of resampled observations exceeds the available degrees of freedom. The spectrum estimated after resampling has a much smaller variance than the spectrum obtained with direct methods. Also, the estimated spectrum has a positive value under all conditions. Therefore, signal reconstruction is

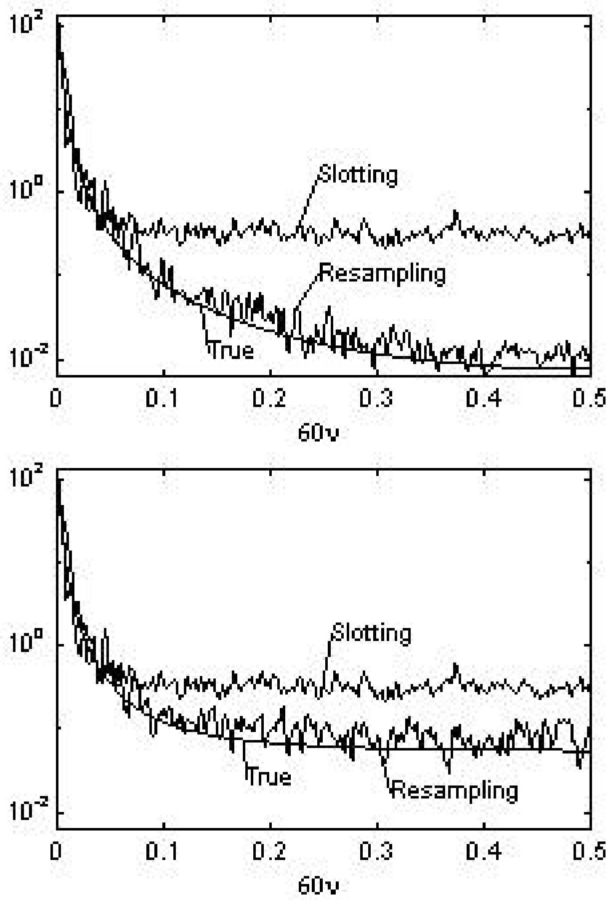

Fig.1 Normalized spectrum estimated with slotting and with NN resampling from a simulated AR(1) process (a) and an AR(1) process with an additional high frequency contribution (b). The level of the slotting spectrum is about the same for both examples in the whole frequency range, whereas the true spectrum and the spectra of resampled signals give a significant difference at higher frequencies.

preferred over direct methods. With signal reconstruction and resampling the estimate is not affected by velocity bias when the sampling rate is high enough.

Interpolation schemes for LDA Signal Reconstruction or Resampling have been studied extensively. It has been found that complicated interpolation algorithms may give good reconstructions in some examples but no scheme offered advantages over simple techniques in a comparison [5]. Resampling can take place by substituting the original observations, that are irregularly spaced in time, in three different ways at the resampling times nT_R:

- Linear Interpolation, LI
  a linear interpolation of the observation x_B at t_B before $t = nT_R$ and the observation x_A at t_A after nT_R.

$$x_{R,n} = x_B + (nT_R - t_B)\frac{x_A - x_B}{t_A - t_B}\,, \qquad (1)$$

- Zero Order Hold, ZOH
  the previous observation x_B
- Nearest Neighbor Resampling, NNR
  the nearest neighbor observation

Theory has a preference for NNR above ZOH because the expected value of the longest sequence of the same observation in the resampled signal is shorter. Precise theoretical results for irregular observations are difficult to obtain, also for LI. When using linear interpolation a new sample of the signal consists of a weighted average of two LDA observations. It is easily shown that, even for interpolation of a regularly sampled signal, the variance of the resampled signal is not equal to the variance of the original velocity signal. Also, the correlation structure of the resampled signal is different from the correlation structure of the velocity signal. With Zero Order Hold or Nearest Neighbor Resampling an unbiased estimate of the variance is obtained. Also, the correlation structure of the data may be conserved better, because a resampled data point is derived from only one irregular sample. The three methods will be compared in simulations. The quality of resampling will be evaluated exclusively on the discrete resampling moments and not on the interpolated continuous signal that might also be obtained with the three resampling schemes.

3. SPECTRAL ESTIMATION WITH AR-MA TIME SERIES MODELING

Time series models have three different types, autoregressive or AR, moving average or MA and combined ARMA. An ARMA(P,Q) process can be written as [7]

$$x_n + a_1 x_{n-1} + \ldots + a_P x_{n-P} = \varepsilon_n + b_1 \varepsilon_{n-1} + \ldots + b_Q \varepsilon_{n-Q}\,, \qquad (3)$$

where ε_n is a purely random process, so a sequence of independent identically distributed stochastic variables. This process is AR for Q=0 and MA for P=0. Any stationary stochastic process with a continuous spectral density can be written as an unique AR(∞) or MA(∞) process [7]. An estimated ARMA(p,q) model is given by

$$x_n + \hat{a}_1 x_{n-1} + \ldots + \hat{a}_p x_{n-p} = \varepsilon_n + \hat{b}_1 \varepsilon_{n-1} + \ldots + \hat{b}_q \varepsilon_{n-q}\,. \qquad (4)$$

The normalized spectral density of this model is given by:

$$\hat{f}(\omega) = \gamma \left| 1 + \sum_{k=1}^{q} \hat{b}_k \exp(-j\omega k) \right|^2 \Big/ \left| 1 + \sum_{k=1}^{p} \hat{a}_k \exp(-j\omega k) \right|^2 . \qquad (5)$$

where γ is a factor normalizing the total power to 1. The goal of AR-MA modeling is to find the time series which gives the best description of the data at hand without using prior knowledge about the true process. This means that not only the model parameters have to be estimated but also the model type(AR, MA or ARMA(p,p-1)) and model order have to be selected.

When the search for the most accurate model is restricted to AR models the best AR order can be found using order selection [8]. In a similar fashion the best MA and ARMA(p,p-1) model is selected. From these three models the single model is selected using a procedure similar to order selection. It is known that using only one model type cannot under all circumstances provide an accurate model of the data. Only by comparing the three model types AR, MA and ARMA an accurate description of the data can be found under all circumstances. For a more detailed discussion of AR-MA modeling the reader is referred to the accompanying

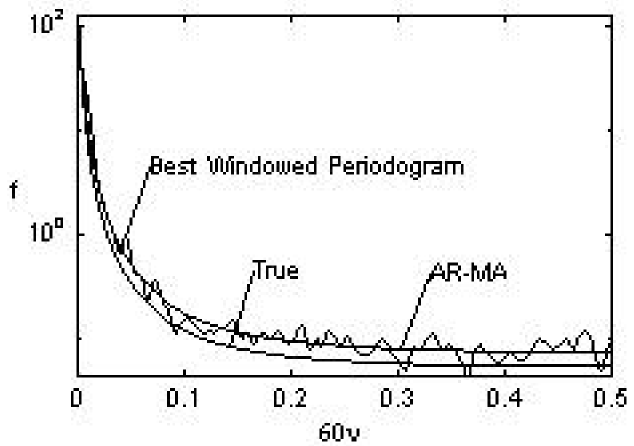

Fig.2 The spectrum of the automatically selected AR-MA model and the best Windowed Periodogram compared to the true spectrum for r=0.05; NN resampling has been used.

paper at this conference by Broersen [4]. The order selection criterion which are used in the AR-MA algorithm take into account the behavior of estimation methods in short samples [9]. This makes it particularly useful for situations where only a short data record is available.

The method of spectral estimation used to be windowed periodograms of tapered data [7]. However, on regularly sampled data AR-MA modeling performs better than windowed periodograms, even when the optimal window size is chosen [4]. Therefore AR-MA modeling is used to estimate the spectrum from the resampled data.

4. SIMULATIONS

To determine the performance of a spectral estimator the estimated spectrum must be compared to the true spectrum. In simulations the true spectrum is known, so they can be used to test the various spectral estimators. The true velocity signal x in the simulations is a slowly varying AR(1) process v with an additional high frequency component w: $x = v + w$.

$$v_n = a v_{n-1} + \varepsilon_n \tag{6}$$

with a = $e^{-1/4000}$ and w is represented by a purely random process with variance $\sigma_w^2 = r\sigma_v^2$. By a suitable choice of σ_ε^2 it is realized that σ_x^2 equals 1 in all simulations. This signal is similar to the practical data which are analyzed in chapter 6. Of this true velocity signal M = 100 000 equidistant observations are generated. Afterwards, this signal is sampled at irregularly spaced intervals with mean sampling distance 10. In the sampling scheme a dependency between the interval between samples and the current value of the velocity can be incorporated. P[S|x] describes the probability a sample S is taken, for a given value of the velocity x: P[S|x] = $\alpha|x|$.

The quality of an estimate is expressed using the model error ME [10]. The model error is a scaled version of the prediction error PE that can be interpreted as the fit of a model to new data of a known correlation or spectral structure:

$$ME = N\frac{PE - \sigma_\varepsilon^2}{\sigma_\varepsilon^2} \tag{7}$$

where N is the number of observations that has been used for estimation. The one step ahead prediction error PE is a measure for how well x_n can be predicted using all previous observations x_{n-1}, x_{n-2}… and the estimated correlation structure. The prediction error includes the additional error caused by estimating the mean. This contribution can become quite significant when the estimate of the mean is biased. The minimal value of the prediction error σ_ε^2 is subtracted which means that the model error is zero when the estimated mean equals the true mean and the estimated correlation function equals the true correlation function.

The model error defined by equation (7) is a measure for discrete time models at the usual time scale 1, a time index without dimension. To apply a simular measure to LDA, two changes are required: a time t with dimension has to be introduced and it must be possible for the experimentator to choose the time scale T at which the model quality is determined. A usual choice for T is given by the resampling time ΔT. A proposal for such error measure is the model error at time scale T, ME_T. ME_T is based on the prediction error at time scale T, denoted PE_T. PE_T is a measure for how well an observation x(t) can be predicted using the previous observations *at lag times kT*: x(t-T), x(t-2T), … and the estimated correlation structure. The model error at time scale T is related to PE_T in the same way the original model error is related to PE:

$$ME_T = N\frac{PE_T - \sigma_{\varepsilon'}^2}{\sigma_{\varepsilon'}^2} \tag{8}$$

where $\sigma_{\varepsilon'}^2$ is the minimal value of PE_T. It can be proved the model error at time scale T is zero when the estimated mean equals the true mean and the estimated correlation function equals the true correlation function at all times kT. As the correlation function at these times is left unchanged, the corresponding spectrum is the *aliased* spectrum. This correlation structure is used in the definition because it is supposed to be the correlation with optimal predictive properties. When the LDA-signal is resampled, T must be an integer multiple of the resampling time ΔT.

Using ME_T the AR-MA spectral estimate is compared to several windowed periodograms for the AR(1) signal with an additional high frequency content r = 0.05 and P[S|x] is a constant independent of x. The spectrum was estimated after NNR with resampling time 30. A Parzen window has been used to estimate the Windowed Periodogram; the length of the window is βM. The model error at time scale 60, ME_{60}, averaged over 100 simulation runs is given in Table 1. The AR-MA spectrum is more accurate than the best windowed periodogram. So far, no example has been found where this result does not hold. In Fig.2 the spectral estimate of the automatically selected AR-MA model is compared to the best windowed periodogram (window size 0.1M) for a typical simulation run.

Linear Interpolation (LI), Nearest Neighbor Resampling (NNR) and Zero Order Hold (ZOH) are compared in simulations of the AR(1) signal with increasing high frequency content r. From the resampled data the spectrum is esti-

	P[S|x]=Constant			P[S|x]= α|x|		
r	LI	ZOH	NNR	LI	ZOH	NNR
0	19.2	4.9	5.3	33.2	4.8	5.1
0.05	40.5	7.6	7.0	85.5	20.4	15.0
0.20	34.4	8.2	7.6	74.8	21.2	15.4
1.00	26.4	9.1	9.0	55.4	21.8	21.8
5.00	15.9	10.1	10.1	59.3	45.5	44.9

mated

AR-MA	Windowed Periodogram				
	0.03M	0.05M	0.1M	0.15M	0.25M
7	180	80	70	90	150

Table 1 The model error ME_{60} averaged over 100 simulation runs of a signal with r=0.05 and P[S|x] a constant independent of x for the selected AR-MA model and windowed periodograms with several window sizes βM. The AR-MA model is more accurate than the best windowed periodogram.

Table 2: The model error ME_{60} averaged over 100 simulation runs for Linear Interpolation(LI), Zero Order Hold(ZOH) and Nearest Neighbor Resampling(NNR) for increasing power at high frequencies. The signal was sampled with P[S|x]=Constant and with P[S|x]= α|x|.

using AR-MA modeling. The values of ME_{60} for increasing high frequency content r are shown in Table 2. The standard deviation of the numbers in the table is about 0.6, so smaller differences than 1 in a row cannot be considered significant. It turns out that ZOH and Nearest Neighbor Resampling are more accurate than LI in this example.

The irregular sampling rate is still rather high in all simulations. That might be a reason that there is no significant difference in the performance of Nearest Neighbor Resampling and Zero Order Hold in most situations. In resampling a purely random process, without a low frequency AR part, it was found that ME_{10} was about 2 times higher for ZOH than for NNR. Therefore, a preference is given to NNR above ZOH. Also in Table 2, NNR was never significantly worse than ZOH.

Different simulations with an AR(2) example showed that in that example the performance of LI was better. It seems that it depends on the characteristics of the true process which resampling scheme is most accurate.

When the signal has little power at high frequencies

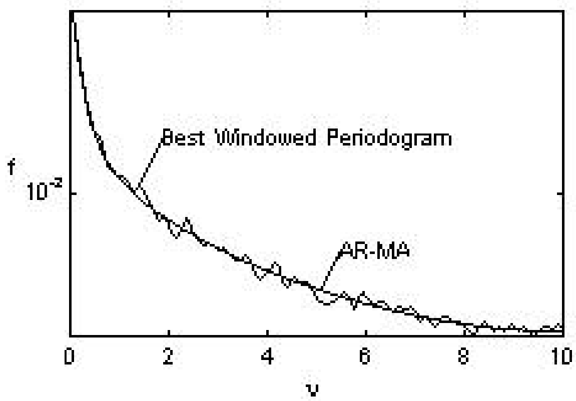

Fig.3 Spectral estimates from practical data estimated with the automatically selected AR-MA model and the Windowed Periodogram closest to the AR-MA estimate.

(small values of r), the performance of resampling is not affected by the dependence of the interval between samples with the current value of the velocity. For increasing r bias effects are increasingly prominent. As the estimate of the mean is biased, it's constribution to the ME is considerable. For r=5 and NNR the contribution of the mean to ME_{60} is 35.

5. PRACTICAL DATA

Practical data have been used that were measured in mixing layers of water. The data have been investigated before [11]. The LDA-data was resampled using Nearest Neighbor Resampling. From the resampled data the spectrum was estimated with AR-MA modeling as well as Windowed Periodogram estimation. The selected AR-MA model is an ARMA(5,4)-model. In Figure 3 the spectrum of this model is compared to the windowed periodogram with window size β=0.02. This is the window size for which the Windowed Periodogram is closest to the AR-MA estimate. The differences are similar to those found in simulations.

6. CONCLUDING REMARKS

Signal reconstruction is preferred over direct methods for spectral estimation from LDA-data. Both Nearest Neighbor Resampling and Linear Interpolation can be used. After resampling the spectrum is estimated using AR-MA times series modeling. The automatically selected AR-MA model provides a more accurate estimate of the spectrum than the best windowed periodogram. The performance of AR-MA on unknown practical data is similar to the calibrated performance in simulations.

The accuracy of a method is evaluated with the model error at time scale T, a measure for how well an observation can be predicted by using all previous observations at intervals kT.

REFERENCES

[1] R.V. Edwards, 'Report of the special panel on statistical particle bias problems in laser anemometry', *Trans. of the ASME, Journ. of Fluids Eng.*, vol. 109, p 89-93, 1987.

[2] L.H.J. Absil, *'Analysis of the Laser Doppler measurement technique for application in turbulent flows'*, dissertation, Delft University of Technology, 1995.

[3] E. Parzen Ed., *'Time series analysis of irregularly observed data'*, New York, Springer-Verlag, 1984

[4] P.M.T. Broersen, 'Facts and Fiction in Spectral Analysis of stationary stochastic processes', *Signal Processing IX, Proc. Eusipco Conf., Rhodes, Greece*, 1998.

[5] C. Tropea,' Laser doppler anemometry: recent developments and future challenges', *Meas. Sci. Technol.*, vol. 6, p 605-619, 1995.

[6] M. Gaster and J.B. Roberts, 'Spectral analysis of randomly sampled signals', *J.Inst. of Mathematics and its Applications*, vol. 15, p 195-216, 1975.

[7] M.B. Priestley, '*Spectral Analysis and Time Series'*, London, Ac. Press, 1981.
[8] P.M.T. Broersen, 'The ABC of autoregressive order selection criteria', *Preprints Sysid '97 Conf., Kitakyushu, Japan*, 8-11 July 1997, p 231-236.
[9] P.M.T. Broersen and H.E. Wensink, 'On Finite Sample Theory for Autoregressive Model Order Selection', *IEEE Trans. on Signal Processing*, vol SP-41,p 194-204, 1993.
[10] P.M.T. Broersen, 'The quality of models for ARMA processes', *IEEE Trans. on Signal Process.*, vol.46, in press, 1998.
[11] J. Tukker, '*Turbulence structures in shallow free-surface mixing layers'*, dissertation, Delft University of Technology, 1995.