Error Measures for Resampled Irregular Data

Stijn de Waele and Piet M. T. Broersen

***Abstract*—With resampling, a regularly sampled signal is extracted from observations which are irregularly spaced in time. Resampling methods can be divided into simple and complex methods. Simple methods such as Sample&Hold (S&H) and Nearest Neighbor Resampling (NNR) use only one irregular sample for one resampled observation. A theoretical analysis of the simple methods is given. The various resampling methods are compared using the new error measure SD_T: the spectral distortion at interval T. SD_T is zero when the time domain properties of the signal are conserved. Using the time domain approach, an antialiasing filter is no longer necessary: the best possible estimates are obtained by using the data themselves. In the frequency domain approach, both allowing aliasing and applying antialiasing leads to distortions in the spectrum. The error measure SD_T has been compared to the reconstruction error. A small reconstruction error does not necessarily result in an accurate estimate of the statistical signal properties as expressed by SD_T.**

***Index Terms*—Interpolation, signal reconstruction, signal sampling, spectral analysis, time domain analysis.**

I. Introduction

IRREGULAR sampling occurs in several applications such as geophysics [1], Laser Doppler Anemometry (LDA) [2], and oscilloscopes [3]. The irregularly sampled signal can be analyzed by extracting a regularly sampled signal from the irregular data with resampling. Frequently used resampling methods are Sample&Hold (S&H), linear interpolation, and cubic spline interpolation [2]. It has been observed that cubic spline interpolation can lead to spurious peaks in the resampled signal for several types of data [1], [3]. With Sample&Hold (S&H), the resampled signal does not display spurious peaks. Therefore, it will be called a robust resampling method. Linear interpolation is also a robust resampling method. The reconstruction of the original signal will generally be more accurate than with S&H. However, the variance of the signal is estimated too low, which causes problems in spectral estimation.

The difference between the estimated and the true signal properties is evaluated using an error measure. For sampled continuous-time signals this is usually done by comparing the estimated power spectrum to the true power spectrum up to a certain frequency [4]. These error measures are called frequency domain error measures. If no antialiasing filter is used prior to sampling, aliasing will cause distortions in the spectrum; if an antialiasing filter is used, distortions in the spectrum will occur because practical antialiasing filters are non-ideal.

Manuscript received May 26, 1999; revised November 23, 1999.

The authors are with the Department of Applied Physics, Delft University of Technology, 2600 GA Delft, The Netherlands (e-mail: waele@tn.tudelft.nl).

Publisher Item Identifier S 0018-9456(00)02421-9.

In this paper, the alternative option is developed, namely a time domain approach. This approach results in the model error at interval T, ME_T [5] and the spectral distortion for the time domain approach, SD_T. In this approach, no antialiasing filter is used. The advantages of the time domain approach over the frequency domain approach are discussed. The new error measures are compared to the reconstruction error RE. Nearest Neighbor Resampling (NNR) is introduced as an improvement over S&H. Resampling methods are divided into simple and complex methods. A theoretical analysis of the influence of the simple methods on a stationary stochastic process is given. Finally, both the resampling methods and the various error measures are compared in a simulation study.

II. Simple and Complex Resampling Methods

With resampling, an equidistant signal x_R at times nT_r is derived from the irregularly spaced samples. A popular resampling method is S&H or zero order hold [2]. The resampled signal at nT_r, $x_R(nT_r)$, is set equal to the last irregular sample *prior* to nT_r. S&H is called a simple method because only one irregular observation is used for each resampled observation. NNR is a simple method which is an improvement over Sample & Hold. With NNR the resampled signal at nT_r is set equal to the irregular sample *nearest* to nT_r. As with S&H, only one irregular sample is used to determine a resampled observation. Complex resampling techniques use two or more irregular samples for each resampled observation. Examples of complex resampling methods are cubic spline interpolation and linear interpolation.

These resampling techniques are applied to irregular samples of the velocity of a turbulent flow as a function of time. The measurements were obtained with Laser-Doppler Anemometry (LDA) [6]. LDA is a non-intrusive way to measure the velocity of a fluid or gas. As a consequence of the measurement technique the samples of the velocity signal are irregularly spaced in time. The reconstructed signal obtained with cubic interpolation is shown in Fig. 1. Also after resampling the signal displays spurious peaks. As a result, the estimated variance is too high. Conversely, the variance estimated with S&H and NNR is correct because only one irregular sample is used for each resampled observation. The estimated variance for the various resampling techniques is given in Table I.

As opposed to cubic interpolation, linear interpolation is a robust resampling method: the resampled signal does not display spurious peaks. However, with linear interpolation the variance is estimated too low. This can be understood by considering the fact that linear interpolation is a weighted average of two irregular observations. This will result in a resampled signal with a variance lower than the variance of the original signal. It is a general property of complex methods that the variance may

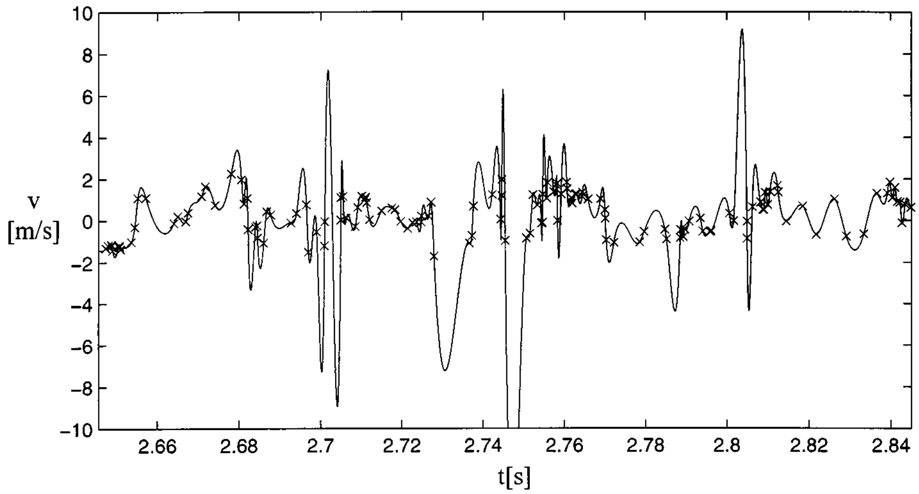


Fig. 1. Cubic interpolation applied to irregular samples of the velocity of a turbulent flow as a function of time. The resampled signal displays spurious peaks. $\times$ = Irregular observations. ——— = Cubic interpolation.

TABLE I
THE VARIANCE ESTIMATED WITH VARIOUS RESAMPLING TECHNIQUES FROM MEASURED LDA-DATA. THE FIRST COLUMN ARE THE ESTIMATES USING *all* IRREGULAR OBSERVATIONS. IN THE SECOND COLUMN THE AVERAGE IRREGULAR SAMPLING INTERVAL IS ARTIFICIALLY DOUBLED BY RANDOMLY SELECTING 50% OF THE IRREGULAR SAMPLES

	Estimated variance	
Method	all irregular observations	50 % of the irregular observations
Sample & Hold	0.76	0.78
Nearest Neighbor Resampling	0.76	0.78
Cubic interpolation	16.86	28.21
Linear interpolation	0.67	0.61

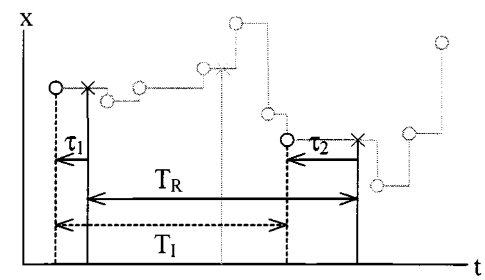


Fig. 2. Interval T_R between resampled data and the interval T_I between the irregular observations used for resampling. In this example the signal is resampled with S&H. The difference between the original signal properties and the signal properties of the resampled signal are caused by the fact that T_I deviates from T_R. $\circ$ = Irregular observations. $\times$ = Resampled signal.

be estimated erroneously. The data at hand can easily be used to demonstrate some of the properties of resampling methods. The average irregular sampling interval is artificially doubled by randomly selecting 50% of the irregular samples. Another estimate of the variance is determined from this set of irregularly spaced samples (see Table I). The estimated variance with S&H and NNR is only slightly different due to statistical variations. The bias in the estimate obtained with linear interpolation has increased. The estimate found with cubic interpolation is completely different from the initial estimate. Summarizing, it can be stated that the simple methods are robust and provide an unbiased estimate of the variance. This result justifies paying more attention to simple methods.

III. THEORETICAL ANALYSIS OF SIMPLE METHODS

In this section the influence of resampling with simple methods on the signal properties of stationary stochastic processes is analyzed in more detail. The interval T_R between resampled data is different from the various intervals T_I between the irregularly spaced samples used for resampling (see Fig. 2). With NNR, the mean square deviation between T_R and T_I will be smaller than for S&H. The difference between T_R and T_I causes deviations between the properties of the resampled signal and the properties of the original signal.

Irregular samples obtained with LDA can exhibit a velocity bias. This is a dependency between the interval between samples and the value of the original signal: when the signal is high the average irregular sampling rate is high [2]. This effect is disregarded in this analysis. It is assumed that the sampling times are equally distributed with mean irregular sampling time T_i. This means that the time between samples has an exponential distribution [7]. The resampling time T_r is assumed to be considerably greater than the average irregular sampling time T_i. This means that the number of times the same irregular observation is used for two resampled observations is negligible in this theoretical analysis. The possibility $P_{S\&H}$ that the same observation is used for two resampled observations using S&H is equal to the possibility that the time between two irregular samples is larger than the resampling time

$$P_{S\&H} = \exp(-T_r/T_i). \quad (1)$$

For NNR this possibility is smaller than for S&H.

The autocovariance function R_{x_R} of the resampled signal is

$$R_{x_R}(T_R) = E[x_R(t)x_R(t+T_R)]. \quad (2)$$

For simple methods the resampled signal at time $t, x_R(t)$, is equal to the original signal x at time $t-\tau_1$: $x(t-\tau_1)$ (see Fig. 2). The resampled signal at $t+T_R$ is equal to the original signal x

at time $t+T_R-\tau_2$: $x(t+T_R-\tau_2)$. Substitution in expression (2) and some rearranging yields

$$R_{x_R}(T_R) = E[x(t)x(t+T_I)] \tag{3}$$

with $T_I = T_R - \Delta\tau, \Delta\tau = \tau_2 - \tau_1$. At this point a distinction has to be made between $T_R = 0$ and $T_R = kT_r, k \neq 0$. For $T_R = 0$ expression (3) reduces to

$$R_{x_R}(0) = E[x(t)^2] = R(0) \tag{4}$$

where R is autocovariance function of x. This means that the variance $R_{x_R}(0)$ of the resampled signal is equal to the variance $R(0)$ of the original signal. For $T_R = kT_r, k \neq 0$, the covariance function of the resampled signal is

$$R_{x_R}(T_R) = \int_{\Delta\tau=-\infty}^{+\infty} f_{\Delta\tau}(\Delta\tau) R(T_R + \Delta\tau)\, d(\Delta\tau) \quad \text{for } T_R = KT_r, k \neq 0 \tag{5}$$

where $f_{\Delta\tau}$ is the distribution function of $\Delta\tau$. This can be re-written as a convolution of $f_{\Delta\tau}$ and R

$$R_{x_R}(T_R) = (f_{\Delta\tau} * R)(T_R). \tag{6}$$

Using the assumption that different irregular observations are used for $x_R(t)$ and $x_R(t+T_R)$, it can be stated that τ_1 and τ_2 are independent stochastic variables with the same distribution function f_τ. The distribution function of $\tau_1 - \tau_2$ can be written as the convolution of f_τ and the mirrored version f_τ^+ of f_τ [8]

$$f_{\Delta\tau} = f_\tau^+ * f_\tau, \tag{7}$$

where f_τ^+ is the distribution function of $-\tau$. Substitution of this expression in (6) yields

$$R_{x_R}(T_R) = (f_\tau^+ * R * f_\tau)(T_R) \quad \text{for } T_R = kT_r, k \neq 0. \tag{8}$$

To include $k = 0$, (4) and (8) are combined to yield the following expression:

$$R_{x_R} = f_\tau^+ * R * f_\tau + n\delta. \tag{9}$$

Since the delta-function δ is zero for $T_R = kT_r, k \neq 0$ (8) is satisfied. Equation (4) is satisfied by setting n equal to

$$n = R(0) - (f_\tau^+ * R * f_\tau)(0). \tag{10}$$

$(f_\tau^+ * R * f_\tau)(0)$ can be interpreted as the expectation of $R(\Delta\tau)$. This expectation will never be greater than the maximum value of $R(\Delta\tau)$ which is $R(0)$. This means that n will always be positive. Note that the first part $f_\tau^+ * R * f_\tau$ of (9) can be interpreted as the covariance function of the signal x filtered by a LTI with pulsresponse f_τ [5]. The second part $n\delta$ of (9) is additive white noise. This means the resampled signal x_R has the same autocovariance function as an artificial signal $x_{R'}$, where $x_{R'}$ is a filtered version of x plus white noise (see Fig. 3). Examples of the influence of these operations on turbulence spectra are given in Section V.

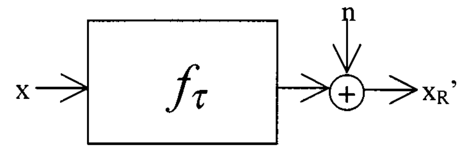


Fig. 3. Relation between the original signal x and the signal x'_R. The resampled signal x_R has the same autocovariance function as x'_R. The variance of x'_R is equal to the variance of x.

The distribution functions for S&H and NNR are

$$\begin{aligned} f_{\tau,\text{S\&H}} &= 0, \qquad \tau < 0, \\ &= \frac{1}{T_i}\exp(-\tau/T_i), \qquad \tau > 0 \end{aligned} \tag{11a}$$

$$f_{\tau,\text{NNR}} = \frac{1}{T_i}\exp(-|\tau|/(T_i/2)). \tag{11b}$$

The standard deviations of T_I for S&H and NNR are

$$sd(T_{I,\text{S\&H}}) = T_i\sqrt{2} \tag{12a}$$

$$sd(T_{I,\text{NNR}}) = T_i. \tag{12b}$$

The deviation between T_R and T_I is smaller for NNR than for S&H. Therefore, it is expected NNR will conserve the character of the original signal better than S&H.

The resampling noise level $n_{\text{S\&H}}$ in S&H can be related to a measurable quantity, namely the mean square difference between subsequent irregular samples $x_{i,n}$ and $x_{i,n-1}$

$$n_{\text{S\&H}} = \tfrac{1}{2}E\{(x_{i,n} - x_{i,n-1})^2\}. \tag{13}$$

Thus, an indication of the resampling noise level for S&H, $n_{\text{S\&H}}$, can be obtained from the measurements. This expression remains useful when measurement noise is present besides the influence of resampling. For NNR the noise level will generally be lower, because the distortions caused by the filter $f_{\tau,\text{NNR}}$ will be smaller.

IV. Error Measures

The difference between the estimated and the true signal properties is evaluated using an error measure. The signal properties of a stationary process can be expressed by the autocorrelation function, the power spectrum or a stochastic differential or difference equation. The model error ME for discrete-time [9] is the starting point for the development of an error measure for sampled continuous signals. The model error can be expressed in the time domain or in the frequency domain. Both expressions are reformulated so they can be applied to sampled continuous-time signals.

The model error ME is defined as a normalized version of the one step ahead prediction error PE

$$\text{ME} = N\frac{\text{PE} - \sigma_\varepsilon^2}{\sigma_\varepsilon^2} \tag{14}$$

where σ_ε^2 is the minimal value for PE. The prediction error PE is a measure for how well x_n can be predicted using all previous observations $x_n, x_{n-1}, \cdots$ with the estimated signal characteristics. Expression (14) for the model error is called the time domain formulation of the model error.

The frequency domain interpretation of the model error is that it is asymptotically equal to the spectral distortion SD

$$\mathrm{SD} = \frac{N}{4\pi} \int_{-\pi}^{+\pi} (\ln\{h(\omega)\} - \ln\{\hat{h}(\omega)\})^2 \, d\omega \tag{15}$$

for unbiased models. h and $\hat{h}$ are the true and the estimated power density function, respectively. The advantage of the spectral distortion is that the variance is taken into account. This expression can easily be reformulated for sampled continuous signals by comparing the estimated and true spectra up to half the sampling frequency. This results in the spectral distortion SD_f

$$\mathrm{SD}_f = \frac{NT}{4\pi} \int_{-\pi/T}^{+\pi/T} \ln\{h(\omega)\} - \ln\{\hat{h}(\omega)\})^2 \, d\omega. \tag{16}$$

The time domain expression (14) for the model error can also be reformulated for sampled continuous signals. This is done by introducing a new prediction error, the prediction error at interval T, PE_T. This prediction error is a measure for how well $x(t)$ can be predicted using all previous observations at lag time kT: $x(t-T), x(t-2T), \cdots$ with the estimated signal characteristics. This prediction error is denoted PE_T. The model error at interval T is again defined as a normalized version of the prediction error at interval T [5]

$$\mathrm{ME}_T = N \frac{\mathrm{PE}_T - \sigma_{\varepsilon'}^2}{\sigma_{\varepsilon'}^2} \tag{17}$$

where $\sigma_{\varepsilon'}^2$ is the minimal value for PE_T. As the sampled observations will become uncorrelated for large T, $\sigma_{\varepsilon'}^2$ will tend to σ_X^2 if T tends to infinity. It can be shown that ME_T is equal to the ME of the sampled model with respect to the sampled process if N is set to the same value in both measures [5]. The sampled process x_T is given by

$$x_{T,n} = x(nT). \tag{18}$$

No antialiasing filter is applied prior to sampling, because optimal predictions are obtained by using the data themselves. This is illustrated by the following example. Suppose a sine with frequency 5 Hz is sampled with a sampling frequency of 4 Hz. When antialiasing is applied before sampling a zero-signal is all that is left. Consequently, no prediction of the original signal can be made. When the signal is sampled without antialiasing, the sampled signal looks like a sine of 1 Hz. Using these samples, the next value of the original signal can be predicted perfectly. The model error at interval T is asymptotically equal to the spectral distortion of the *aliased* spectrum h_a, SD_T

$$\mathrm{SD}_T = \frac{NT}{4\pi} \int_{-\pi/T}^{+\pi/T} (\ln\{h_a(\omega)\} - \ln\{\hat{h}_a(\omega)\})^2 \, d\omega. \tag{19}$$

The spectrum ha is the aliased version of the original spectrum h or, equivalently, the Discrete Fourier Transform of the sampled autocovariance function $R(kT)$

$$h_a = \sum_{m=-\infty}^{+\infty} h\left(\omega + m\frac{2\pi}{T}\right) = \frac{T}{2\pi} \sum_{k=-\infty}^{+\infty} R(kT) e^{-j\omega kT}. \tag{20}$$

ME_T and SD_T are error measures for the time domain approach, because both error measures are zero when the time domain properties given by the sampled autocovariance function $R(kT)$ are conserved. Nevertheless, the autocovariance functions are not plotted because ME_T and SD_T are relative error measures. Autocovariance functions are not a good indication of relative error measures, since differences with low power are not noticed [11]. Instead, the *aliased* spectra h_a are plotted with a logarithmic power axis. The advantage of SD_T is that the variance is included in the error measure. The advantage of ME_T is that it has a clear interpretation as a normalized version of the prediction error at interval T [(17)]. For the comparison of resampling methods SD_T will be used, as the variance may be estimated erroneously with complex methods.

The advantage of the time domain approach is that no processing is required prior to sampling. The frequency domain approach requires the use of an ideal antialiasing filter. Even when most of the power is concentrated at low frequencies compared to half the sampling frequency, an antialiasing filter is necessary for relative error measures. As a result of aliasing, the difference between the original spectrum and the spectrum of the sampled signal at half the sampling frequency is at least a factor of 2. The contribution of the high-frequency part to a relative error measure is not diminished if the power at these frequencies is low.

The antialiasing filter must be applied to the continuous signal. Often, the continuous signal is not available. Then, the time domain approach is the operational alternative. When the continuous signal is available an antialiasing filter can be used. However, all antialiasing filters are non-ideal. This introduces distortions in the estimated spectrum, which hampers the analysis of the data. This is a fundamental problem of the frequency domain approach which is not experienced with the time domain approach. Hence, for many applications the time domain approach has advantages over the frequency domain approach.

Another type of error measure which is often used to compare resampling methods is the reconstruction error RE [3]. This is the mean square difference between the resampled signal and the original signal

$$\mathrm{RE} = \frac{\sum_k (x,(kT_r) - x(kT_r))^2}{\sigma_x^2}. \tag{21}$$

A resampling method which provides accurate reconstruction of the original signal does not necessarily lead to an accurate estimate of the statistical properties of the signal. An example is given by linear interpolation and nearest neighbor resampling. Linear interpolation will generally provide a more accurate reconstruction. However, the variance of the signal is estimated too low with linear interpolation, while nearest neighbor resampling provides an unbiased estimate of the variance. This can result in a more accurate covariance structure with a lower SD_T. An example of this is given in the next section. Another difference between the reconstruction error and the model error is that the error of reconstruction is an absolute error measure in the frequency domain while ME_T and SD_T are relative error measures. A comparison of relative and absolute error measures

TABLE II
THE INFLUENCE OF AN ANTIALIASING FILTER FOR SPECTRA A AND B. THE ERROR AS A RESULT OF USING A NONIDEAL ANTIALIASING FILTER IS COMPARED TO THE STATISTICAL ERROR IN AR, MA, AND ARMA$(p, p-1)$ PARAMETER ESTIMATION AND ORDER SELECTION IN THE TIME DOMAIN APPROACH

frequency domain approach (SD_f)	A	B
applied to true spectrum:		
sampling after anti-aliasing	300	28
time domain approach (SD_T)		
applied to true spectrum:		
sampling without anti-aliasing	0	0
estimated AR-model	10	3
estimated MA-model	71	40
estimated ARMA(p,p-1)-model	5	4

suggests that relative error measures are generally more appropriate for the analysis of stationary stochastic processes [11].

V. SIMULATIONS

Simulations are used to show the influence of an antialiasing filter and to determine the performance of the various resampling techniques. Two different turbulence spectra are used in these simulations [12]:

1) Spectrum A with a –5/3 slope in the log-log spectrum followed by a –7 slope
2) Spectrum B with only the –5/3 slope.

High order autoregressive (AR) processes have been determined with power spectra that are very accurate approximations of the turbulence power spectra. These AR processes are used to generate the data.

A. The Influence of an Anti-Aliasing Filter

The influence of an antialiasing filter on the turbulence spectra A and B is examined. Most of the power in these spectra is concentrated at the lower frequencies. This signal is sampled at regularly spaced intervals with sampling time $T = 100$. Using the frequency domain approach, the signal is filtered prior to sampling by a fifth order Butterworth antialiasing filter with the cut-off frequency at half the sampling frequency. The influence of this filter on the *true* power spectrum is expressed using the spectral distortion SD_f as defined for the frequency domain approach (see Table II). Using the time domain approach, no processing is required prior to sampling. This means that the application of the time domain approach to the true spectrum results in a perfect result: SD_T is zero.

The spectral distortion of the frequency domain approach is compared to the spectral distortion of estimated time series models in the time domain approach. The models are estimated from 10^7 simulated observations regularly sampled at interval 100 without antialiasing. For spectrum A, the error of the antialiasing filter applied to the true spectrum is up to 60 times larger than the error as a result of parameters estimation and order selection in the time domain approach. For spectrum B, the differences are somewhat smaller.

TABLE III
THE SPECTRAL DISTORTION SD_T AT INTERVAL 500 AND THE RECONSTRUCTION ERROR RE FOR THE VARIOUS RESAMPLING METHODS FOR IRREGULAR SAMPLING TIME T_i EQUAL TO 100. THE RECONSTRUCTION ERROR CANNOT BE CALCULATED FOR LINEAR INTERPOLATION WITH CORRECTED VARIANCE σ^2, BECAUSE THIS METHOD CANNOT BE EXPRESSED AS A METHOD WHICH RECONSTRUCTS THE ORIGINAL SIGNAL

	Spectrum A		Spectrum B	
Method	SD_T	RE	SD_T	RE
NNR	802	0,0085	14	0,17
S&H	1631	0,0390	6	0,27
linear with corrected σ^2	1596	-	14	-
linear interpolation	167	0,0012	445	0,12
cubic interpolation	8	$6 \cdot 10^{-5}$	7193	0,75

B. Comparison of Resampling Methods

Of the high-order AR-process, $M = 10^7$ equidistant observations are generated. Afterwards, this signal is sampled at irregularly spaced intervals with mean irregular sampling interval T_i equal to 100. To show the influence of the irregular sampling interval T_i, both signals are also sampled with T_i equal to 50. The quality of the estimated signal properties is determined at interval $T = 500$ using SD_T. From the irregularly spaced data, the signal properties are estimated using S&H, NNR, cubic and linear interpolation. The resampling time T_r equals 500. From the resampled signal the signal properties are estimated using ARMAsel time series analysis. ARMAsel time series analysis is used instead of the windowed periodogram because ARMAsel time series analysis provides a more accurate estimate of the signal properties [10]. The resulting values of SD_T and the reconstruction error RE for T_i equal to 100 are given in Table III. The estimated spectra with NNR, S&H, and cubic interpolation for T_i equal to 100 are shown in Fig. 4. The values of SD_T for T_i equal to 50 are given in Table IV.

As expected, NNR performs better than S&H. The theory describing the influence of NNR and S&H has been compared to the simulation results using SD_T. The SD_T is equal to the value expected as a result of statistical errors in the estimated parameters. This means the theoretical description is correct for these examples. The indication of the noise level $n_{S\&H}$, which can be obtained from the data by using (13), provides an accurate estimate of the true noise level. The noise level for spectrum A with T_i equal to 100 is plotted in Fig. 4. For this signal the noise level is greater than the true spectrum for the high frequencies: as a result the spectrum found with S&H and NNR deviates from the true spectrum. The noise level can be reduced by increasing the average irregular sampling frequency. This results in a lower SD_T, as can be seen by comparing Tables III and IV.

The variance of the resampled signal with linear interpolation is too low. Since the correct value of the variance is known from the simple methods, this error can be corrected. The variance found with linear interpolation is replaced by the value found with NNR. Results of linear interpolation with correct variance are similar to those found with S&H. Using the erroneous value for the variance in linear interpolation happens to result in a lower SD_T for spectrum A; the result for spectrum B is worse. Cubic interpolation has a smaller SD_T than NNR for spectrum A and a higher SD_T for spectrum B. The behavior of resampling

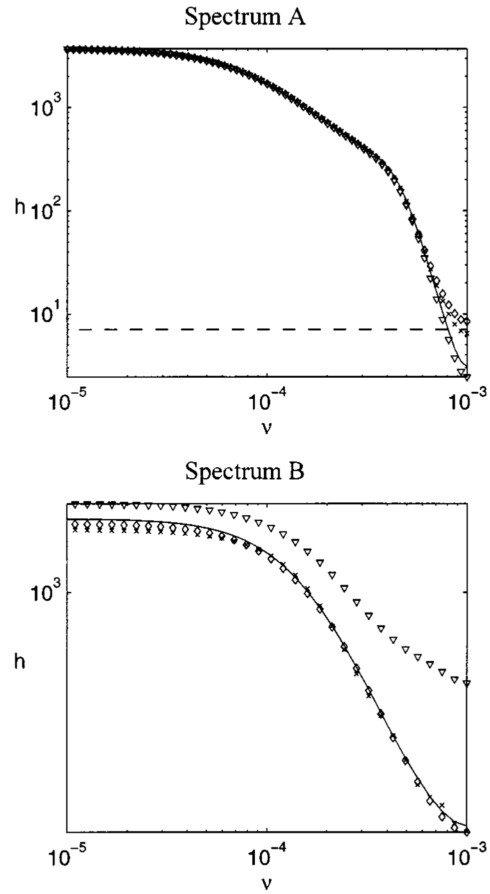


Fig. 4. Power spectra estimated with Nearest Neighbor Resampling, Sample&Hold and Cubic interpolation. The mean irregular sampling interval T_i is equal to 100. For spectrum A the estimated noise level $n_{\mathrm{S\&H}}$ is given. ——— = true spectrum. × = Nearest Neighbor Resampling. ◇ = Sample&Hold ∇ = Cubic interpolation. - - - - - - - - = estimated noise level.

TABLE IV
THE SPECTRAL DISTORTION SD_T AT INTERVAL 500 FOR THE VARIOUS RESAMPLING METHODS FOR IRREGULAR SAMPLING TIME T_i EQUAL TO 50

Method	Spectrum A	Spectrum B
NNR	81	7
S&H	286	2
linear with corrected σ^2	167	2
linear interpolation	20	74
cubic interpolation	14	1906

methods in measured data is similar to the behavior found in these simulations.

The reconstruction error RE is not a good indication for the quality of an estimate as expressed by SD_T. Comparing the results for NNR for the two spectra for T_i equal to 100 (see Table III), the reconstruction error is considerably smaller for spectrum A, while the SD_T is smaller for spectrum B. This is a result of the fact that the errors in spectrum A occur in a region of the spectrum with low power (see Fig. 4). Another example is given by the results for linear interpolation and NNR for spectrum B (see Table III). The reconstruction error RE is smallest for linear interpolation, while the spectral distortion SD_T is smallest for NNR. These simulations show that an accurate reconstruction of the original signal does not necessarily result in an accurate estimate of the statistical signal properties.

VI. Conclusions

Simple resampling methods have some desirable properties which are not found in complex resampling methods. They are robust and provide an unbiased estimate of the variance. NNR is an improved version of S&H. A theoretical analysis of NNR and S&H is given. The theoretical formulae have been verified in simulations.

Complex methods such as linear interpolation and cubic interpolation have as a common disadvantage that the variance can be estimated erroneously. When applied to practical Laser-Doppler Anemometry data, cubic interpolation displays spurious peaks. The variance found with linear interpolation can be replaced by the correct value. After this correction linear interpolation is similar to Sample&Hold. Hence, NNR is the preferred resampling method for irregularly spaced data.

The error measures for the time domain approach ME_T and SD_T have been introduced for sampled continuous signals. The advantage of the time domain approach over a frequency domain approach is that no antialiasing is required prior to sampling. An accurate reconstruction of the original signal does not necessarily result in an accurate estimate of the statistical signal properties as expressed by ME_T and SD_T.

References

[1] P. Yuou, E. Baert, and M. F. Loutre, "Spectral analysis of climate data," *Surv. Geophys.*, vol. 17, pp. 619–663, 1996.

[2] C. Tropea, "Laser doppler anemometry: Recent developments and future challenges," *Meas. Sci. Technol.*, vol. 6, pp. 605–619, 1995.

[3] Y. Rolain, J. Schoukens, and G. Vandersteen, "Signal reconstruction for non-equidistant finite length sample sets: A "KIS" approach," *IEEE Trans. Instrum. Meas.*, vol. 47, pp. 1046–1052, 1998.

[4] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic Press, 1981.

[5] S. de Waele and P. M. T. Broersen, "Reliable LDA-spectra by resampling and ARMA-modeling," *IEEE Trans. Instrum. Meas.*, vol. 48, pp. 1117–1121, Dec. 1999.

[6] E. R. Meinders, J. B. Campen, K. Hanjalic, and T. H. van der Meer, "LDA-analysis of the flow pattern around an array of cubical protrusions in low Reynolds number turbulent channel flow," in *Proc. 4th Int. Conf. Exp. Heat Transfer, Fluid Mechanics and Thermodynamics*, 1997, pp. 67–74.

[7] A. M. Mood, F. A. Grasybill, and D. C. Boes, *Introduction to Statistics*. New York: McGraw-Hill, 1974.

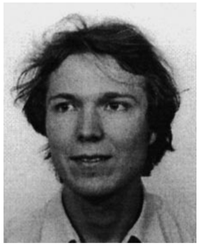
[8] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1965.

[9] P. M. T. Broersen, "The quality of models for ARMA processes," *IEEE Trans. Signal Processing*, vol. 46, pp. 1749–1752, 1998.

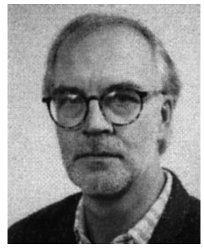
[10] ——, "Facts and fiction in spectral analysis of stationary stochastic processes," *IEEE Trans. Instrum. Meas.*, to be published.

[11] ——, "The performance of spectral quality measures," in *Proc. IMTC '99 Conf.*, Venice, Italy, 1999, pp. 751–756.

[12] E. Müller, H. Nobach, and C. Tropea, "Model parameter estimation from non-equidistant sampled data sets at low data rates," *Meas. Sci. Technol.*, vol. 9, pp. 435–441, 1998.



Stijn de Waele was born in Eindhoven, The Netherlands, in 1973. He received the M.Sc. degree in applied physics in 1998 from Delft University of Technology, Delft, The Netherlands, where he is currently pursuing the Ph.D. degree in the D Department of Applied Physics. His research interests are the development of new time series analysis algorithms and its application to radar processing.



Piet M. T. Broersen was born in Zijdewind, The Netherlands, in 1944. He received the M.Sc. degree in applied physics in 1968 and the Ph.D degree in 1976, both from Delft University of Technology, Delft, The Netherlands.

He is currently with the Department of Applied Physics, Delft University. His research interests are the selection of order and type of time series models and the application to spectral analysis, model building, and feature extraction.