

# Five Separate Bias Contributions in Time Series Models for Equidistantly Resampled Irregular Data

Piet M. T. Broersen

**Abstract**—The use of time series models for irregular data requires resampling of the data on an equidistant grid. Slotted resampling transforms an irregular randomly sampled process into an equidistant signal where data are missing. An approximate maximum-likelihood time series estimator has been developed to estimate the power spectral density and the autocorrelation function of multishift slotted nearest-neighbor (NN) resampled data sets. Resampling always causes bias in spectral estimates due to aliasing in the frequency domain and to shifting the observation times to an equidistant grid. Furthermore, orders of the time series models that are too low can cause a significant truncation bias and, probably, an additional missing-data bias, both of which disappear if the model orders are taken high enough. Finally, a special bias is present if the probability of making an observation at a certain time depends on the instantaneous amplitude of the observed signal. All five bias types are independent of the sample size and will not diminish if more data can be used for the estimation.

**Index Terms**—Autoregressive (AR) models, nearest-neighbor (NN) resampling, slotting, spectral analysis, time series, uneven sampling.

## I. INTRODUCTION

IRREGULAR sampling occurs, in practice, if continuous-time processes are observed at times determined by arbitrary triggering events. Irregular sampling may naturally arise in numerous fields, including geophysics, heart-rate analysis [1], astronomy [2], climate research, and turbulent flow, if that is observed with laser Doppler anemometry (LDA) [3]. A survey of popular methods for the spectral analysis of irregularly observed LDA data has been given [3], which includes direct Fourier transforms, slotted correlation estimates with variants [4], and sample and hold (SH) interpolation. A continuous-time maximum-likelihood approach has been developed for autoregressive (AR) models of irregular data [5]. However, the likelihood had many local minima, and convergence to the global minimum strongly depended on very accurate starting values for the nonlinear search [6].

Sampled data models are often used for an approximate description of continuous-time systems. The possibility of retrieving the system dynamics is improved by the use of relative error

measures [7]. Additional problems are encountered if the continuous system includes stochastic disturbances [8]. The choice of error definitions [7] and a sampling method [8] is very important if a physical interpretation is the final purpose of the system identification. In particular, if no filtering can be applied before sampling, the continuous-time models are very sensitive to the noise-generating mechanism in the sampling procedure [9].

Many continuous-time complications can be neglected if the goal of signal processing is to obtain an estimate for the power spectral density (PSD) of the data in a finite frequency range, without trying to separate the system and noise. However, bias problems are very important in all spectral estimates for irregular data, even in this simplified problem. SH replaces irregular data by a contiguous equidistantly resampled signal. It can give accurate results if the resampling frequency is lower than  $f_0/10$ , where  $f_0$  is the mean data rate [10]. SH is accurate if all grid points are close to actual irregular observations. However, the same irregular observation can sometimes be substituted for lower data rates at more than one equidistant grid point. Refined SH methods can be accurate until frequencies slightly higher than  $f_0/2$ , but the refined methods explicitly use the probability distribution function of the irregular observation times. They can only be applied if the observation times have a Poisson distribution. Nearest-neighbor (NN) resampling substitutes the closest observation on each equidistant grid node instead of the last preceding that is used in SH. The properties of NN and SH are similar.

In NN, multiple uses of single observations can be avoided with the slotting principle. An irregular observation is only substituted at an equidistant grid point if it is less than half the slot width away. The grid point is left empty if no observation falls within the time slot. At first, the slotting principle has been used in slotted autocorrelation estimation [4]. The product of two irregular observations contributes to a certain lag  $k\Delta$  of the slotted autocorrelation function if their time distance is between  $(k - 0.5)\Delta$  and  $(k + 0.5)\Delta$ , where  $\Delta$  is the slot width. However, the number of pairs that contribute to different slots will have variations. Therefore, no variant of the estimated slotted autocorrelation functions is positive semidefinite. Hence, its Fourier transform as spectral estimate is not positive everywhere. No satisfactory solution for this problem has been reported in the literature. In practice, it is common to present only the parts of the spectra with positive estimates while omitting the negative estimates. Most estimated spectra of this method look very irregular and unreliable at higher frequencies [4].

The slotting principle has been applied to the NN resampling of irregular data on a regular time grid [11]. Slotted NN can give

Manuscript received June 23, 2008; revised November 10, 2008. First published January 19, 2009; current version published April 7, 2009. The Associate Editor coordinating the review process for this paper was Dr. Jerome Blair.

The author is with the Department of Multi-Scale Physics, Delft University of Technology, 2628 Delft, The Netherlands (e-mail: p.m.t.broersen@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2009.2012928

positive semidefinite estimates for the autocorrelation function, with estimated spectra that are positive everywhere. It has bias components, some of which can be reduced by taking a higher resampling frequency. A second option for bias reduction is a slot width smaller than the resampling distance [11]. The slotting principle gives an equidistant signal, with data missing at those grid nodes that are more than half the slot width away from the closest actual irregular observation.

Spectral estimation for equidistant observations with missing data [12] is much simpler than the spectral analysis of continuous irregular data. Jones described an efficient method to calculate the exact likelihood for time series models in missing-data problems [13]. An automatic time series algorithm [12] used this missing-data method and outperformed all other methods.

If time series models of too-low model orders are estimated, this can cause a significant truncation bias and, probably, an additional missing-data bias [12]. Truncation bias can occur because the estimation of models with more than 15 or 20 parameters may become cumbersome or very time consuming in missing-data problems. Fortunately, often, lower-order models are sufficient. Missing-data bias is the influence of higher order nonzero parameters on the lower order estimates. Those bias contributions will disappear both if the model orders can be taken high enough.

LDA measures the velocity of small reflecting particles that move with the liquid or gas flow. The number of particles per second that arrive in the measurement volume depends on the flow velocity if the particles are uniformly distributed in space. Therefore, the probability of making an observation at a certain time depends on the instantaneous amplitude of the observed velocity signal [14]. This bias type is generally called velocity bias. Simple corrections have been given for 1-D flow [15]. The difficult 3-D problem has been described for a known Gaussian flow pattern [16]. An additional problem is that the expected number of observations per second with a certain velocity in a certain direction depends on all three velocity components, whereas only one component is measured.

This paper studies bias properties with the intention of bias reduction. It has been shown that spectral estimates converge to their biased expectation for increasing sample sizes [17]. Three groups of bias contributions can be distinguished for the five separate bias causes: the bias during data collection, the resampling or interpolation bias with the inevitable aliasing, and finally, the bias due to the model choice in processing the resampled data. The aliasing bias is determined by the resampling frequency and the bias of shifting times to an equidistant grid by the slot width. Truncation and missing-data biases will only occur for low-order time series models if the overall spectral shape is not described with acceptable accuracy. So far, no adequate solution can be given for the velocity bias in LDA data.

## II. TIME SERIES MODELS

A brief introduction to the theory of time series models will give the notation and clarify how parameters determine the

PSD of a discrete-time signal. An AR moving average (ARMA)  $(p, q)$  discrete-time process  $x_n$  can be written as [18], [19]

$$x_n + a_1x_{n-1} + \dots + a_px_{n-p} = \varepsilon_n + b_1\varepsilon_{n-1} + \dots + b_q\varepsilon_{n-q} \quad (1)$$

where  $\varepsilon_n$  is a purely random discrete-time white-noise process of independent identically distributed stochastic variables with zero mean and variance  $\sigma_\varepsilon^2$ . For resampled continuous-time data with a resampling distance  $T_r$ , the signal  $x_n$  is the observation at time  $nT_r$ . Other values for  $T_r$  would give processes with different parameters and  $\sigma_\varepsilon^2$  in (1). The parameters of an estimated ARMA  $(p, q)$  model represent a parametric estimate of the autocorrelation function and of the PSD for the noisy data. Generally, it is assumed that the data  $x_n$  represent a stationary stochastic process [18]. In this paper, no additional assumptions considering the underlying continuous-time process are necessary. The only assumption is that the equidistantly resampled data at times  $nT_r$  can be described as a regular stationary stochastic process. This process describes the measured observations, and it can be a noiseless signal or the sum of a desired signal and additive noise. The estimated spectrum will then represent the PSD of signal plus noise in the discrete-time frequency range.

The PSD  $h(\omega)$  of the model and the frequency range depend on the resampling time  $T_r$ . The spectrum is fully determined by the parameters in (1) together with the variance  $\sigma_\varepsilon^2$  and  $T_r$

$$h(\omega) = \frac{T_r\sigma_\varepsilon^2}{2\pi} \frac{\left| 1 + \sum_{i=1}^q b_i e^{-j\omega i} \right|^2}{\left| 1 + \sum_{i=1}^p a_i e^{-j\omega i} \right|^2}, \quad -\frac{\pi}{T_r} < \omega \leq \frac{\pi}{T_r}. \quad (2)$$

The autocorrelation function at lags that are multiples of  $T_r$  can, in principle, be computed by an inverse continuous-time Fourier transform of (2). In practice, that would become an approximation because a discrete transform will replace the integral by a finite summation. However, exact formulas that relate the autocorrelation directly to the parameters of (1) are available and more accurate [19].

The accuracy of equidistant time series models can be evaluated with the prediction error  $PE$ . This can be seen as the accuracy of forecasts of the future values of the observations [20]. Care should be taken in using  $PE$  as the accuracy to predict independent new realizations of the same process [21]. The prediction error  $PE(p', q')$  of an ARMA  $(p', q')$  model with arbitrary orders  $p'$  and  $q'$ , respectively, is defined as the mean square error of the one-step-ahead prediction with that model in new fresh contiguous equidistant data. A normalized scaled version  $PE_s(p', q')$  can be computed without the availability of new data in simulation experiments where the true ARMA  $(p, q)$  process parameters are known [19]. This can be used in missing-data problems and is independent of the variance of the data [11], [19]

$$PE_s(p', q') = \frac{PE(p', q')}{\sigma_\varepsilon^2}. \quad (3)$$

It can be used for pure AR models where  $q' = 0$  and for MA models with  $p' = 0$ . The minimum for the expectation of the  $PE_s(p', q')$  for efficiently estimated *unbiased* ARMA  $(p', q')$  models from contiguous equidistant data is given by the number of estimated parameters and the sample size  $N$  [19]

$$E\{PE_s(p', q')\} \approx 1 + (p' + q')/N. \quad (4)$$

Unbiased models are characterized here by  $p' \geq p$  and  $q' \geq q$ . Every additionally estimated parameter gives at least a contribution of  $1/N$  to (4), due to its estimation variance. Models of orders lower than  $p$  and/or  $q$  are truncated and biased. They will often have a higher  $PE_s(p', q')$  than that given by (4), and that bias contribution does not depend on  $N$ . The prediction error measure is a very general measure [20]. It will be used here as a distance measure between biased ARMA  $(p', q')$  models and a true ARMA  $(p, q)$  process. It is related to the spectral distortion, which is the integral of the squared difference of the logarithms of the true and the estimated spectra [19].

### III. BIAS OF MSSNNR

Bias is present in all methods that use interpolation or resampling techniques to transform irregularly sampled continuous-time data into an equidistant discrete-time sequence. The interpolation techniques that have been investigated include zero-order hold, linear interpolation, polynomial interpolation, splines, Shannon interpolation, fractal interpolation, and many more [22]. The conclusion was that the visual appearance of the reconstructed signal may look promising in some examples, but the spectral bias could not significantly be improved in comparison to SH. SH substitutes true observations at shifted time instants. The bias of SH is only less than 10% for frequencies below  $f_0/20$  [10]. Hence, most methods can only be applied for extremely high data rates  $f_0$ . All attempts to approximate the signal at the resampling instants by interpolation, without slotting, failed for higher frequencies [22]. However, it has been shown that multishift *slotted* NN resampling (MSSNNR) can give reliable spectra for frequencies up to  $100f_0$  and still higher. Only the bias of this last method with slotting is treated in detail in this paper.

Resampling irregular data inevitably gives two sources for bias: aliasing and the shift of the time instants to a grid. After equidistant resampling, the highest frequency  $f = 1/(2T_r)$  or  $\omega = \pi/T_r$  in the discrete-time spectrum is determined by the resampling frequency  $f_s = 1/T_r$ . The discrete-time autocorrelation function is the sampled version of the continuous correlation  $R(\tau)$ . The aliased true spectrum is given by [19]

$$h(\omega) = \sum_{k=-\infty}^{\infty} R(kT_r)e^{-j\omega k}, \quad -\frac{\pi}{T_r} < \omega \leq \frac{\pi}{T_r}. \quad (5)$$

Aliasing has a strong influence on the spectrum if the spectral shape is rather flat at high frequencies. The smallest possible influence of aliasing at  $\omega = \pi/T_r$ , i.e., the end of the frequency range in (5), is an error of a factor of two. The aliasing error can be rather small and concentrated in a narrow range close to  $\omega = \pi/T_r$  for spectra with a steep decay at high frequen-

cies. It is not possible to reduce the influence of aliasing in irregular data by using antialiasing filters. The reason is that neither continuous-time nor the usual discrete-time filters can be applied to irregularly sampled signals.

NN resampling replaces an unevenly sampled continuous-time signal by a contiguous equidistant discrete-time signal at the resampling grid points  $nT_r$ . At every grid point, the closest irregular observation is substituted. The bias of NN will be large unless the resampling frequency  $f_s$  is much lower than the mean data rate  $f_0$ . Like many other resampling methods, the spectral bias is already about 10% at  $f_0/20$ . One of the reasons is that the same observation is substituted at more grid points if there is a gap between irregular observations that is larger than twice the resampling distance  $T_r$ . This causes a distortion of the autocorrelation function at lag  $T_r$  and a spectral bias. The bias can be reduced and made independent of  $f_0$  with the slotting principle. In the simplest case, the slot width is taken equal to the resampling distance  $T_r$ . Slotted NN accepts only an observation if it is within half the slot width from the resampling time. This excludes the possibility that a single irregular observation is used at different resampling points and reduces the bias, at the cost of creating a missing-data signal with gaps. Several methods exist for spectral estimation when some data are missing [12]. However, most methods can only be used for small missing fractions. Fortunately, one time series method can still be used if more than 99% of the data are randomly missing [12].

A further improvement of the bias is found with MSSNNR, where the slot width  $w$  is made only a fraction of the resampling distance [11]. Just taking  $w = T_r/M$ , where  $M$  is an integer number, gives disjoint intervals. Several irregular observation times  $t_i$  will not be within the small slots around  $t = nT_r$ . They would be lost in the resampling. An irregular observation at  $t_i$  gives a nonempty place for the resampled signal at  $nT_r + mw$  if

$$nT_r + mw - 0.5w < t_i \leq nT_r + mw + 0.5w \\ m = 0, 1, \dots, M - 1. \quad (6)$$

All slots of width  $w$  are now connected in time, and every  $t_i$  is within a slot. If two irregular observations fall within the same slot, the one closest to the slot center is taken, and the other is discarded. The  $M$  shifted starting points in (6) give  $M$  different equidistant missing data sequences, each with the same resampling time step  $T_r$ .

An example is given here to demonstrate the meaning of  $M$  different sampled signals. If a continuous-time registration would be available for all  $t$ , a discrete-time representation can be made by sampling it at the instants  $nT_r$ . It is assumed that  $T_r$  is small enough to reduce the influence of aliasing.  $M$  different discrete-time signals could be extracted by using the resampling instants  $nT_r + mw$ ,  $m = 0, 1, \dots, M - 1$ , with  $w = T_r/M$ . Those  $M$  signals would be similar for a small value of  $T_r$ . One single signal for one value of  $m$  would contain almost all valuable information, and there is hardly anything gained by evaluating all  $M$  signals with the same  $T_r = Mw$ . Nothing is lost either, because all signals would give approximately the same time series model. This changes if the  $M$  signals are

obtained with the MSSNNR of (6) for irregular data. Now, each signal is a missing-data record, with many data missing if a small value for  $T_r$  or a large value for  $M$  is used. If the irregular observation times are random, it may be expected that close irregular observations at distances of approximately a multiple of  $T_r$  are present on different and independent time intervals for each of the  $M$  missing-data signals. This is certainly true for Poisson distributed irregular sampling instants and also for many other irregular sampling schemes. Nearby observations can always be predicted much better than observations that are far away. They have the largest influence on the negative log likelihood that is minimized in parameter estimation. Hence, the influential parts for the  $M$  MSSNNR signals may be at different places in time. Those  $M$  likelihoods are added to a single likelihood value in the ARMAse1-irreg algorithm [11], as if they were independent. It does no harm if they are dependent, because every signal would then give the same time series model. However, it can give better accuracy if they are independent. Hence, it will reduce the estimation variance of the parameters. AR parameters are estimated by maximizing the sum of the  $M$  likelihoods together as a function of the parameters [11]. The ARMAse1-irreg algorithm estimates AR, MA, and ARMA models from irregular data that are resampled with MSSNNR. The program automatically selects the best of all candidate models and is available on the Internet [23].

The analysis in this paper assumes that the irregular observation instants are randomly and freely distributed over the time axis. This applies to many practical problems. An exception is found in heart-rate analysis [1]. If the average distance between heartbeats is  $T$ , it will almost never occur that the actual distance is less than  $T/2$  or  $T/3$ .

Similar patterns of the interarrival times are found in measurements that are synchronized by the public Internet [24]. Fig. 1 shows an example where all irregular intervals are between 0.02 and 0.04 s. The highest frequency that can be evaluated in this example is 25 Hz, which is half the frequency belonging to the shortest interarrival time. Specific solutions have to be developed for those data. The integral pulse frequency modulation model for the heartbeat rate describes the instantaneous heart rate as a constant plus an irregular variable part [1]. In this paper, it is assumed that the histogram of the irregular interarrival times starts close to an interval with the length zero. No further restrictions about the shape of the histogram have been assumed; it may be Poisson distributed, uniform, or anything else.

For random sampling intervals, the bias of MSSNNR is reduced by increasing  $M$ , in comparison to the usual slotted NN that is found for the choice  $M = 1$ . The probability density function  $f(\tau)$  of the continuous-time lags  $\tau$  that contribute to the resampled autocorrelation  $R_{\text{res}}(kT_r)$  can be calculated if the density of the irregular sampling instants  $t_i$  is known [11]. The expectation of the resampled autocorrelation is then given by

$$R_{\text{res}}(kT_r) = \int_{-w}^w R(kT_r + \tau) f(\tau) d\tau, \quad n \neq 0$$

$$R_{\text{res}}(0) = R(0). \quad (7)$$

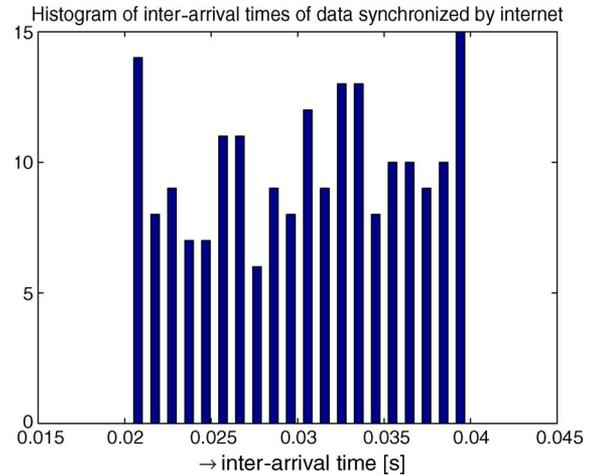


Fig. 1. Histogram of the interarrival times of measurements where a very precise clock sends synchronization pulses over the Internet.

This is a convolution in the continuous time domain of the true autocorrelation  $R(\tau)$  with the symmetric density function  $f(\tau)$ , followed by sampling at  $T_r$  and by adding white noise. The resampled autocorrelation at shift zero is given by the sum of squares of irregular observations and does not suffer from time shifts. It is favorable that the true autocorrelation at shift zero is left unchanged. The biased resampled discrete-time spectrum has both the bias caused by shifting irregular times to a grid and aliasing bias. It follows (5) as

$$h_{\text{res}}(\omega) = \sum_{k=-\infty}^{\infty} R_{\text{res}}(kT_r) e^{-j\omega k}, \quad -\frac{\pi}{T_r} < \omega \leq \frac{\pi}{T_r}. \quad (8)$$

For narrow functions  $f(\tau)$ , the outcome of (7) will converge to  $R(kT_r)$  and the influence on the spectral density is given by the aliasing only, like in (5). Shifting the time instants to a grid gives an effect that is additional to aliasing because the autocorrelations  $R_{\text{res}}$  are weighted averages of neighboring autocorrelation lags  $\tau$ .

For high data rates or dense irregular sampling, if the slot  $w = T_r/M$  is wider than the average distance  $T_0 = 1/f_0$  between the irregular observations, several observations will mostly fall within one slot, and only the one closest to the center of the slot  $nT_r + mw$  is accepted in MSSNNR [6], which means that the probability density  $f(\tau)$  in (7) is narrower than the slot width  $w$ , and  $f(\tau)$  depends on the probability density of the individual irregular observation times. This case is not further treated because slotting is not necessary for very dense irregular sampling. Ordinary NN or SH resampling without slotting will give accurate spectral estimates for very low frequencies. The error is less than 50% for frequencies below  $f_0/2\pi$  Hz [10]. The influence of the SH operation on the spectrum, without slotting, can be described by adding step noise to the true spectrum followed by filtering. Therefore, even the spectrum of broadband white noise becomes colored after SH resampling [10].

If  $w$  is much smaller than  $T_0$ , however, most slots will remain empty. The total width of  $f(\tau)$  cannot be greater than  $2w$  if all individual irregular observations are less than  $w/2$  away from

the center of the slot. The shape of  $f(\tau)$  turns out to become almost triangular for Poisson distributions of the individual observation times. In the limit for small values of  $w$ , it will become independent of the original density of the observation times. The probability density of the contributing continuous time lags can then be approximated with [11]

$$f(\tau, M) = \frac{T_r/M - |\tau|}{(T_r/M)^2}, \quad 0 < |\tau| \leq T_r/M$$

$$f(\tau) = 0, \quad |\tau| > T_r/M. \quad (9)$$

The bias of the MSSNNR signal in time domain depends on the convolution  $R(\tau) * f(\tau)$ . The white-noise correction is given by

$$\Delta R_{\text{res}}(0) = R(0) - \int_{-w}^w R(\tau) f(\tau) d\tau. \quad (10)$$

The equivalent operation in the frequency domain is the multiplication of the PSD with the transfer function that belongs to (9), followed by adding white noise with power  $\Delta R_{\text{res}}(0)$  to the filtered spectrum.

The filter factor for the slotted spectral density is the Fourier transform of the triangle of (9). That is the real symmetrical Fejer kernel [18] with the main lobe between  $+M/T_r$  and  $-M/T_r$  Hz. Hence, the main lobe is always wider than the total discrete-time frequency range of the resampled signal, which extends from  $-1/(2T_r)$  to  $1/(2T_r)$  Hz.

The effect of a slot width that is much smaller than the resampling distance  $T_r$  is shown in Fig. 2. Only the frequency range below  $1/(2T_r)$  is shown in the figure. Taking a smaller slot width diminishes the spectral filtering. Furthermore, a smaller slot gives a smaller value of  $\Delta R_{\text{res}}(0)$  in (9). The numerical value of the transfer function at  $f = 1/(2T_r)$  is 0.409, 0.812, 0.950, and 0.9873 for  $M = 1, 2, 4$ , and 8, respectively. The magnitude of the transfer function can be compared with the bias due to aliasing. That is at least a factor of two at the end of the frequency domain. The maximum filter effect gives 0.409 for  $w = T_r$ , and the smallest aliasing would give a factor of two. Aliasing adds power, and the filter removes power; therefore, in theory, both effects could counteract. It will turn out that the white-noise correction (10) can be more important than the filter effect.

In comparison with the values in Fig. 2, the spectral multiplication factor  $\alpha$  for SH resampling is given by [10]

$$\alpha = \frac{1}{1 + (2\pi f/f_0)^2}. \quad (11)$$

This factor is about 0.9 for  $f = f_0/20$ . The factor is about the same for many other methods that do not use slotting [22]. For that reason, those methods can only accurately be used for frequencies that are much lower than  $f_0$ . The mean data rate  $f_0$  is not important in Fig. 2. Its only role is that the resampling distance  $T_r$  must be much smaller than the mean distance  $T_0 = 1/f_0$ . This is the same as requiring a resampling frequency that is much higher than the mean data rate.

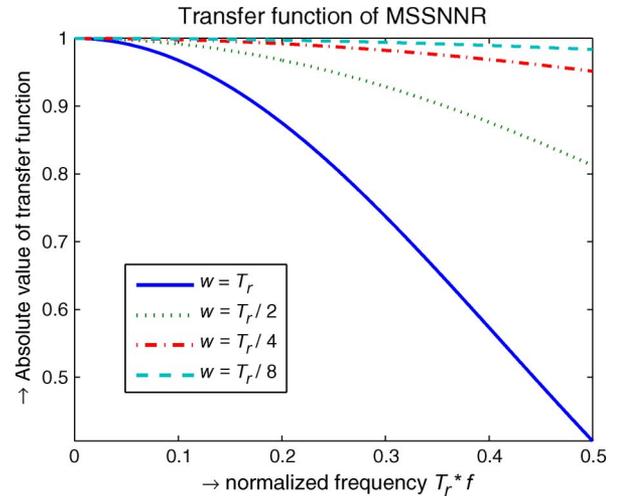


Fig. 2. Real symmetric transfer function that is the Fourier transform of the triangular impulse response  $f(\tau, M)$  of MSSNNR in (9). The true continuous-time spectrum is multiplied with this transfer function factor.

If aliasing is not very small, filtered contributions from higher continuous-time frequencies should be added to those in Fig. 2. These contributions can be important. A typical result follows for a broadband white-noise example with irregular sampling, which has a frequency range that is much greater than  $f_0$ . Aliasing is very important then. The usual SH or NN resampling without slotting produces a colored spectrum [3]. However, MSSNNR has no effect on the autocorrelation function for broadband white noise if that is zero for all lags greater than  $T_r - w$  in (7). Hence, there is no filtering operation. In simulations with various sample sizes, the ARMAse1-irreg algorithm selected the white-noise model. The presentation in Fig. 2 for a limited frequency range gives a misleading impression in this example due to the very large aliasing that takes place.

The filter effect of shifting the time instants can be made much smaller than the effect of aliasing. Folding the infinitely wide continuous-time spectrum gives an error of spectral aliasing of at least a factor of two at the end of the discrete-time frequency range, given by  $1/(2T_r)$ . However, the aliasing error may be concentrated in a small part of the frequency range, for steep spectra. The shift bias becomes already very small for  $w = T_r/8$  in Fig. 2 because the real filter transfer function is almost equal to one. Therefore, in many applications, using  $w = T_r/2$  or  $w = T_r/4$  may be advisable to reduce the filter effect of shifting observation times to a grid.

The bias due to aliasing can only be diminished by using a higher resampling frequency. Afterward, the bias of shifting slotted time instants to a grid can still be made as small as desired by using denser and denser grids, with a smaller slot width  $w$ . The disadvantage is that the total number of grid points at distance  $w$  increases for the same number of irregular observations, and more grid points are left empty. This gives a greater missing fraction, and the estimation of time series parameters becomes less accurate and takes more computing time [12]. The theoretical limit for  $f(\tau)$  is a delta function, and the spectrum finally approaches the aliased spectrum of (5). Therefore, it is sufficient if the bias due to shifting is made small

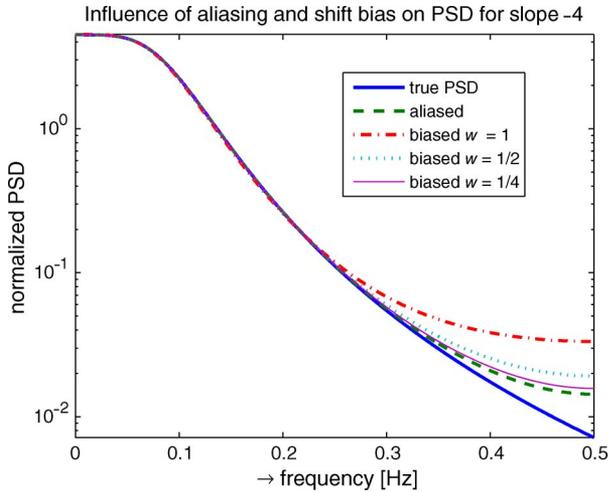


Fig. 3. True PSD with constant log-log slope until very high frequencies, the aliased spectrum, and the effect of slotted shifting for a moderate slope.

in comparison with the inevitable aliasing bias. However, the white-noise contribution of (10) to the spectral bias depends on the shape of the true autocorrelation function, and no general expression for its magnitude can be given.

For a useful theoretical approximation of the bias with (7), it was not necessary to compute the exact *integral* of the true continuous autocorrelation function. The bias could be computed with a discrete-time approximation of the autocorrelation function. It becomes accurate if at least about ten lags of the true autocorrelation are available in the range from  $-w$  to  $w$ . This, in turn, means that the true continuous-time spectral density must be known up to frequencies of  $5M/T_r$  Hz to use a discrete-time version of (7) for the bias evaluation of resampling.

#### IV. BIAS FOR SPECTRAL SLOPES

The complete shifting and aliasing biases can only be given for specified examples. A simple example is a spectrum consisting of two straight lines in the double logarithmic representation, with a transient part. The spectrum is given by

$$S(f) = \frac{C}{(1 + f/10)^\alpha} \quad (12)$$

where a constant  $C$  is the low-frequency slope, and  $\alpha$  is the asymptotic slope in the double logarithmic plot.  $C$  is an arbitrary constant for the variance of the process. The influence of the slope will be investigated, with the resampling time  $T_r = 1$  s. The mean data rate  $f_0$  has no influence on the aliasing. It has no influence on the shifting bias either if it is much smaller than 1 Hz, for example, less than 0.1. Otherwise, the triangular shape in (9) is no longer applicable.

Fig. 3 shows the results for  $\alpha = 4$ . The effect of aliasing is visible above 0.3 Hz. The effect of the shift bias converges to the bias of aliasing for  $w < 0.25T_r$ . A further reduction of the slot width no longer has much influence. The filtering effect in Fig. 2 would diminish the spectral density at the end of the discrete-time frequency interval; the white-noise contribution of (10) adds most to the logarithmic spectrum at the frequencies where the spectrum is weak. It is clear that the latter effect is dominating in the shift bias of MSSNR.

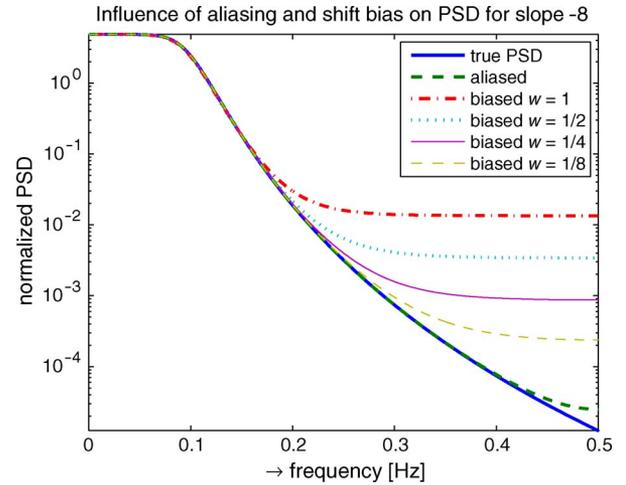


Fig. 4. True PSD with constant log-log slope until very high frequencies, the aliased spectrum, and the effect of slotted shifting.  $T_r = 1$  s.

TABLE I  
ALIASED FRACTION AND  $PE_s$  OF ALIASING, WHITE-NOISE FRACTION OF SHIFT BIAS, AND  $PE_s$  OF THEORETICAL ALIASED SHIFT SPECTRUM AS A FUNCTION OF THE SLOPE  $\alpha$ , WITH  $w = T_r/2$  IN ALL CASES;  $T_r = 1$  s

$\alpha$	Aliased fraction	$PE_s$ alias	$\Delta R_{\text{res}}(0)/R(0)$	$PE_s$ shift
2	$0.13 \cdot 10^{-1}$	1.032	0.0968	1.030
4	$0.24 \cdot 10^{-2}$	1.013	0.0075	1.013
6	$0.60 \cdot 10^{-6}$	1.009	0.0041	1.288
8	$0.18 \cdot 10^{-7}$	1.008	0.0034	2.735

Fig. 4 shows what happens if the slope is steeper and the spectral level at the end of the frequency range is weaker. The difference between the aliasing and the shift biases is much greater here. The shift bias will finally converge to the aliasing bias but for a much smaller slot width than  $T_r/8$ .

Table I shows that the aliased fraction quickly drops for steeper spectral slopes. The fraction is defined as the ratio of the integrated continuous-time PSD above the Nyquist frequency  $1/(2T_r)$  to the power below that frequency. If the spectrum decreases with  $f^{-\alpha}$ , the integral of the tail will decrease with  $f^{-\alpha+1}$ . The white-noise power  $\Delta R_{\text{res}}(0)/R(0)$  also becomes smaller, but the decrease is much slower. Therefore, the influence of the shift bias becomes stronger than that of aliasing for steeper spectra. The value of the prediction error caused by the shift bias is much greater than that caused by aliasing for slopes steeper than six. This agrees with the results in Figs. 3 and 4, with a limited and a large shift bias, respectively.

Table II shows that a smaller slot width can theoretically give a better bias accuracy of the spectral estimates, with a lower value of  $PE_s$ . The estimation variance is not considered here, but it would make the value of  $PE_s$  a little bit higher for the estimated spectra. The white-noise power (10) depends on the slot width  $w$ . The theoretical value of the  $PE_s$  of aliasing without time shifts is about 1.009. It follows that the accuracy of aliasing can finally be obtained for smaller values of the slot width. Taking it smaller than  $T_r/32$  will hardly improve the accuracy any further in Table II. A disadvantage of a small slot width is the increasing number of empty points in the resampled

TABLE II  
WHITE-NOISE FRACTION OF SHIFT BIAS AND PREDICTION  
ERROR OF SHIFTING AS A FUNCTION OF THE SLOT WIDTH  
FOR THE SPECTRUM WITH SLOPE SIX;  $T_r = 1$  s

$w/T_r$	$\Delta R_{\text{res}}(0)/R(0)$	$PE_s$
1	$0.16 \cdot 10^{-1}$	1.799
1/2	$0.41 \cdot 10^{-2}$	1.288
1/4	$0.10 \cdot 10^{-2}$	1.086
1/8	$0.26 \cdot 10^{-3}$	1.026
1/16	$0.64 \cdot 10^{-4}$	1.013
1/32	$0.16 \cdot 10^{-4}$	1.010

signal, which gives an increased estimation variance. A smaller slot gives more equidistant grid points for the same number of irregular data. This leads to more missing data. In practice, the slot width is often taken as  $T_r/2$  or  $T_r/4$ .

The correction of the SH bias with the refined SH method [3] can improve SH spectral estimates if they are obtained at observation times that are Poisson distributed. Those corrections undo the filtering effect with the inverse of (11). Furthermore, an estimate is made for the white noise that is involved in SH resampling. However, the accuracy will certainly not be acceptable for frequencies higher than  $f_0$ . So far, MSSNNR is the only resampling method that can give a low bias in spectra for frequencies above  $f_0$ .

Possible corrections for the white-noise level in (10) have been considered for MSSNNR. However, the correction of the estimated spectrum would require too much *a priori* information about the data, e.g., assuming that the data are noiseless. Most practical data will be noisy, particularly at high frequencies. Many true spectra with different spectral slopes or additive noise above  $1/(2T_r)$  would give almost the same result for the filtered PSD plus noise. Hence, it is hardly possible to use noise correction to extract additional useful information from estimated spectra with shift bias. The idea of refined reconstruction is not useful for MSSNNR. Instead, using a smaller slot width is possible for every spectral shape and does not require any assumption about the data.

## V. MISSING-DATA AND TRUNCATION BIASES

A second group of bias terms is applicable to spectral estimation with time series models. Those models are used for the MSSNNR signal with ARMA<sub>sel</sub>-irreg. Two sources of bias have been defined for this group. The first is the truncation bias, which is caused by selecting underfitting models with a lower order than the true process. The reason for underfit is that no candidates with high-enough orders are considered. That might be for computational reasons. Not all details can be represented by such models. The magnitude of the truncation bias is independent of the sample size  $N$ , like the aliasing bias and the bias due to shifting irregular times to a grid. It depends on the magnitude of the omitted true parameters of higher orders. Fortunately, many different spectra can quite accurately be approximated with five or six AR parameters and often even less.

As an example, truncated low-order AR models of the process with the spectrum of (12) have been studied. The true process has an infinite AR order with many small parameters. Truncated AR models have been derived from the parameters of

TABLE III  
 $PE_s$  OF TRUNCATED TRUE AR MODELS AS A FUNCTION OF THE MODEL  
ORDER FOR THE SPECTRUM OF (12) WITH THE SLOPE  $\alpha = 6$

order	$PE_s$
0	28.499
1	5.067
2	1.611
3	1.110
4	1.032
5	1.012
6	1.005

the true process, and the  $PE_s$  has been computed as a function of the model order.

Table III shows that the truncation bias is quite important for orders one, two, and three. It is rather small for all AR orders greater than three. The truncation bias is then comparable to the aliasing and the shift biases and the inaccuracy that is caused by estimating the parameters.  $PE_s$  for order zero is the accuracy of the AR(0) or white-noise model; it is equal to  $\sigma_x^2/\sigma_\varepsilon^2$ .

This example shows that low-order AR models can describe this type of smooth discrete-time spectra with a high accuracy. In general, low-order AR models describe the overall shape of the PSD, with broad peaks. However, narrow peaks are smoothed in low-order models. Only very high order models give an accurate shape of those peaks. Spectra with narrow details may require much higher AR orders [25].

A second bias in this group is only relevant for missing-data problems [12]. Generally, AR models are estimated recursively, for increasing model orders. The parameters of the lower order model, together with an additional parameter zero, are used as the starting values for the AR model of one order higher [11]. AR(1) models are always computed for data with a higher order, as an intermediate step. Whereas only observations at distance  $T_r$  contribute to the AR(1) parameter of contiguous data, observations at multiples of  $T_r$  will also contribute to the estimate of that parameter if data are missing [12]. This causes an additional bias in estimated models if the order of the estimated model is lower than the true process order. This bias can only be large if the best possible approximation of the spectrum of the truncated model is still far from the true spectral shape. Therefore, it is not very important in practice, because those models will not be appropriate for the given data anyway. It explains why simulations with AR(1) and AR(2) models do not always converge to the truncated expectations of the  $PE_s$  values, like those in Table III for the orders one and two. This bias can only be present in models where the truncation bias is very important and disappears for higher orders.

## VI. VELOCITY BIAS

The bias that is treated so far is caused by operations on the irregular data, and the first four bias contributions are independent of the measurement equipment. The third group contains bias terms that are caused by the measurement equipment that delivers the irregular data. This may be a poor calibration, a poor measurement principle, or poor information exchange over long distances. Reduction of this bias, if possible, does not depend on the signal processing but on the specific equipment

that is used. No general methods to reduce that bias can be given.

The fifth bias that is discussed is important for data of LDA measurement devices. The application of time series models to the spectral analysis of irregular turbulence revealed an important bias in LDA observations [25]. Velocity bias is present if the probability of making an observation at a certain time depends on the instantaneous amplitude of the observed velocity signal [14]. Early corrections were limited to the 1-D probability density function of the velocity [15]. Recently, experimental evidence showed important differences between the histograms of hot-wire measurements and LDA observations of the same turbulent process [25].

Corrections for the velocity bias have been proposed in slotted autocorrelation estimators [14]. Weighting functions for the individual irregular observations depend on the transit time of a particle, the interarrival time, or the inverse of the instantaneous volume flux [14]. These can give a correction for the higher probability of finding irregular observations with a high velocity. Unfortunately, the velocity bias has the greatest impact in the high frequency range of the spectrum, which mostly is rather inaccurate for slotted correlation methods.

The influence of velocity bias can be small for a flow with a low turbulence level. With low turbulence, the mean value of the velocity is much higher than the standard deviation of the velocity variations around the mean. However, the velocity bias can be very important if the variations are large in comparison with the mean. This may happen if the mean velocity is zero. This is the situation where LDA data are most sensitive to velocity bias.

Fig. 5 shows a histogram of the velocity of observations of the same turbulent flow process with a normal or Gaussian amplitude distribution. The data have been produced with a benchmark generator [26]. It is obvious that the variance of the biased data differs from the unbiased variance. That difference will be even stronger for the PSD, which is the distribution of the total variance over the frequency interval. Furthermore, it is not possible to derive the multidimensional Gaussian distribution function of all data together from their individual histograms. It is useful to determine a histogram like that in Fig. 5 for irregular observations that are made with LDA equipment. If unexpected valleys or dips are found, it is likely that velocity bias is present. Spectral estimates obtained with any method should then be considered with care.

## VII. BENCHMARK DATA

The separate bias contributions can only be determined exactly for specific examples for which the observation procedure is known. In a turbulent flow, a Heisenberg spectrum is a well-known spectral prototype. It is given by [11], [26]

$$S(f) = \frac{C}{(1 + f/100)^{5/3}(1 + f/1000)^{16/3}} \quad (13)$$

with the two constant slopes of  $-5/3$  and  $-7$  in the double logarithmic plot.  $C$  is an arbitrary constant for the variance of the process.

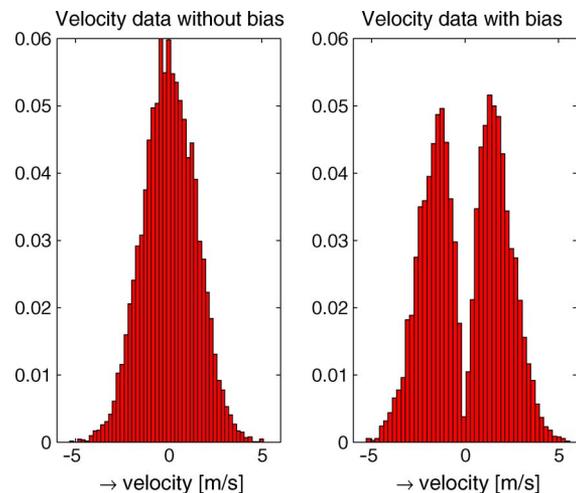


Fig. 5. Histograms of 10000 Gaussian benchmark data with mean velocity zero, without and with velocity bias.

The MSSNNR has been used in an automatic computer program denoted as ARMAseI-irreg, which estimates time series models from irregular data [23]. In a first application of the ARMAseI-irreg algorithm to simulated irregular data, good results have been obtained for estimated AR models with a known and low order [6]. It has been demonstrated that using very high resampling rates is not a problem for the ARMAseI-irreg algorithm [11]. The spectra of practical bubbly flow data [27] have been analyzed until frequencies that are 250 times higher than the mean data rate  $f_0$ . The algorithm performed well until very high frequencies, even if the observation time instants are not Poisson distributed. In many circumstances, the probability density of the observation times has no direct influence on the spectral estimates. An exception has been shown in Fig. 1. The application of the irregular time series algorithm requires that the data are stationary and stochastic. In particular, variable data rates or dropouts with large gaps between the data cause no problems. The smallest interarrival distances between the irregular observations set a limit to the smallest useful resampling time  $T_r$ , which should at least be as large as that distance. There is no reason to suspect that the distribution function for the arrival times will strongly influence the spectral estimates of ARMAseI-irreg.

Attention is focused to the influence of the slot width in MSSNNR on the quality of the selected spectrum. It is assumed that the resampling frequency is high enough to reduce aliasing and that AR models of low orders can approximate the global spectral shape. This eliminates truncation and missing-data biases. The bias of MSSNNR can still be influenced by choosing the slot width. Benchmark data for (13) are generated with a program that is available on the Internet [26]. The true spectrum is shown in Fig. 6 for the chosen discrete-time frequency range. The constant slope in the log-log spectrum continues for higher frequencies. The slotting bias in the autocorrelation function has been computed with (7)–(9). The true biased spectra, with the theoretical biases of aliasing and shifting included, are shown in Fig. 6 for two values of the slot width  $w$ . The true biased spectrum for  $w = T_r/4$  is already close to the aliased spectrum computed with (5), with  $PE_s = 1.040$ ;  $PE_s = 1.408$

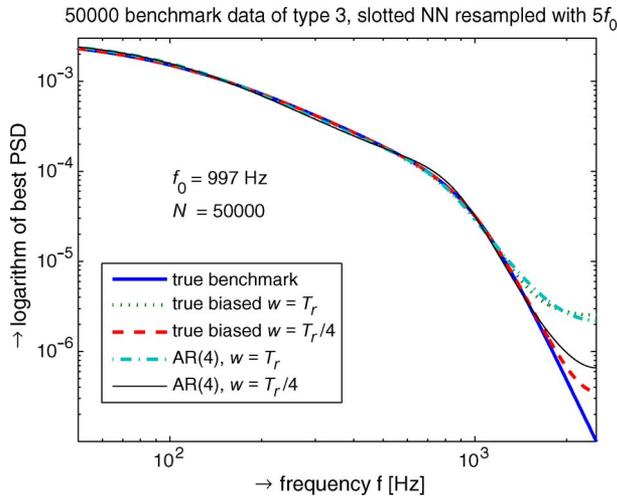


Fig. 6. True, biased true, and estimated spectra for  $T_r = 0.2$  ms and two slot widths  $w = T_r$  and  $w = T_r/4$ , both with selected AR order four for  $N = 50\,000$ .

for  $w = T_r$ . The bias can hardly be reduced further by taking a smaller slot width in this example. The true and biased spectra almost coincide for the largest part of the frequency range. A small difference with the true spectrum is seen for frequencies above 1400 Hz for  $w = T_r/4$ . The bias difference is somewhat greater for  $w = T_r$  and starts at about 1200 Hz.

The  $PE_s$  values for the estimated AR(4) models in Fig. 6 were 1.391 and 1.112 for  $w = T_r$  and  $T_r/4$ , both rather close to their theoretical minimal expectations. Fig. 6 shows that the estimated AR(4) spectra are close to their biased expectations. Therefore, choosing a slot width smaller than  $T_r$  can improve the accuracy. A similar investigation with  $T_r = 0.25$  s showed that sometimes, ten irregular observations could be sufficient to obtain a global spectral estimate [17]. If only few observations are available, it was better to choose  $w = T_r$  because then, the missing fraction was smaller.

Table IV gives the accuracy of the estimated AR models for increasing sample sizes of the benchmark data (13) and  $T_r = 0.25$  s. Less than ten irregular observations are generally not enough to obtain reliable estimates. It is remarkable that the accuracy of the estimated AR(1) and AR(2) models does not improve much if more data are available because the realized value for the  $PE_s$  of estimated models is already close to the biased expectation for less than 100 observations. It requires higher AR orders to get accurate spectra. The truncation bias is dominant for AR model orders one and two. The final row gives the expectation of  $PE_s$  for truncated true models, with the best possible parameters for the low-order aliased and shifted model. The estimated AR(2) models do not converge to this theoretical value. That is the effect of the special missing-data bias that can occur in models of orders that are too low. The effect is only rather strong for AR(2) models in this example.

For higher orders, the main bias contribution is the bias (7) of shifting the observation times to a grid. The expectations in the rows denoted by  $E\{PE_s\}$  do not become significantly smaller for orders higher than six for truncated true AR models with alias and shifting biases included; the results for those orders are not reported in this paper, and in practice, they would not

TABLE IV  
 $PE_s$  FOR THE ESTIMATED AR MODELS IN A SINGLE SIMULATION RUN AND ITS EXPECTATION FOR THE TRUNCATED BIASED AR PROCESSES AS A FUNCTION OF THE MODEL ORDER AND OF THE SAMPLE SIZE. SLOTTED NN USED  $T_r = 0.25$  ms AND  $w = T_r$  OR  $w = T_r/4$ . THE FINAL ROWS IN EACH BLOCK GIVE THE EXPECTATIONS OF  $PE_s$  FOR TRUNCATED BIASED TRUE MODELS IF  $w = T_r$  AND  $T_r/4$

$N$	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)
$w = T_r$						
10	1.73	1.50	1.46	2.87	8.57	32.92
100	1.77	1.33	1.27	1.59	1.68	2.30
1000	1.76	1.44	1.20	1.20	1.14	1.16
10000	1.75	1.43	1.16	1.22	1.20	1.21
$E\{PE_s\}$	1.73	1.31	1.18	1.19	1.19	1.19
$w = T_r/4$						
10	1.87	1.63	2.73	7.24	7.34	20.33
100	1.75	1.40	1.81	4.90	5.90	10.74
1000	1.77	1.40	1.09	1.73	1.64	1.57
10000	1.76	1.49	1.09	1.11	1.03	1.57
$E\{PE_s\}$	1.73	1.29	1.08	1.04	1.02	1.02

be selected by a good order-selection criterion [6]. The values 1.19 and 1.02 for the AR order six in the final rows are still the same for the AR(100) model. Therefore, the accuracy of the estimated model cannot increase much further by taking more than 10 000 observations unless the slot width is made smaller. However, the aliasing effect will then be the limiting factor for the accuracy, with a minimum value around 1.01 in Table II. It is very unlikely that AR orders higher than five will ever be selected for this benchmark example with the chosen resampling time and slot width.

The bias will not decrease if more observations are available; only the estimation variance can become smaller. Due to this estimation variance, the accuracy of the estimated AR(6) model is still significantly worse than its biased expectation for  $N$  equal to 10 000 for  $w = T_r/4$ . The accuracy of the estimated AR models of orders that are higher than five will become better for larger sample sizes but never better than 1.02 for this slot width. In particular, for the order six, it is clear that a smaller slot width will require many more observations to obtain accurate models. The missing fraction and, hence, the estimation variance increase for a smaller slot width.

## VIII. CONCLUSION

Only continuous-time spectra can be unbiased for irregularly sampled processes. In equidistantly resampled data, three groups of bias contributions can be recognized: the bias during data collection, the resampling bias with the inevitable aliasing, and, finally, the bias due to the model choice in signal processing.

A data-collection bias in LDA experiments is called velocity bias. This may occur if the turbulent velocity variations are not small in comparison with the mean flow velocity. Precautions about this bias type should be taken during data collection.

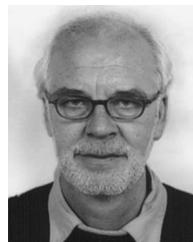
Resampling irregular data into an equidistant missing-data problem causes aliasing bias. This is made smaller by using a higher resampling frequency. Equidistant resampling without

slotting can be used for accurate spectra until frequencies that are less than 10% of the mean irregular data rate. Higher frequency ranges are only accurate if slotting is applied. A good resampling method is the MSSNNR procedure. The bias that is caused by shifting the irregular observation times to an equidistant grid is diminished in MSSNNR by using a slot width that is smaller than the resampling distance.

In processing the resampled signal with time series models, truncation and missing-data biases can appear. They become small if the model order is high enough to describe the significant spectral details. The influence is obvious if low-order models are studied, and it explains what happens then. However, order selection will not select the models of orders that are too low in practice, and those low orders need no further attention.

#### REFERENCES

- [1] J. Mateo and P. Laguna, "Improved heart rate variability signal analysis from the beat occurrence times according to the IPFM model," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 8, pp. 985–996, Aug. 2000.
- [2] C. Thiebaud and S. Roques, "Time-scale and time-frequency analyses of irregularly sampled astronomical time series," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 15, pp. 2486–2499, 2005.
- [3] L. H. Benedict, H. Nobach, and C. Tropea, "Estimation of turbulent velocity spectra from laser Doppler data," *Meas. Sci. Technol.*, vol. 11, no. 8, pp. 1089–1104, Aug. 2000.
- [4] M. J. Tummers and D. M. Passchier, "Spectral estimation using a variable window and the slotting technique with local normalization," *Meas. Sci. Technol.*, vol. 7, no. 11, pp. 1541–1546, Nov. 1996.
- [5] R. H. Jones, "Fitting a continuous time autoregression to discrete data," in *Applied Time Series Analysis II*, D. F. Findley, Ed. New York: Academic, 1981, pp. 651–682.
- [6] P. M. T. Broersen and R. Bos, "Estimating time-series models from irregularly spaced data," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 4, pp. 1124–1131, Aug. 2006.
- [7] G. C. Goodwin, J. I. Yuz, and J. C. Agüero, "Relative error issues in sampled data models," in *Proc. 17th World Congr. IFAC*, Seoul, Korea, 2008, pp. 5047–5052.
- [8] L. Ljung and A. Wills, "Issues in sampling and estimating continuous-time models with stochastic disturbances," in *Proc. 17th World Congr. IFAC*, Seoul, Korea, 2008, pp. 14 360–14 365.
- [9] R. Pintelon, J. Schouten, and P. Guillaume, "Continuous-time noise modeling from sampled data," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 6, pp. 2253–2258, Dec. 2006.
- [10] R. J. Adrian and C. S. Yao, "Power spectra of fluid velocities measured by laser Doppler velocimetry," *Exp. Fluids*, vol. 5, no. 1, pp. 17–28, Jan. 1987.
- [11] P. M. T. Broersen, "Time series models for spectral analysis of irregular data far beyond the mean data rate," *Meas. Sci. Technol.*, vol. 19, no. 1, pp. 015 103-1–015 103-14, Jan. 2008.
- [12] P. M. T. Broersen, S. de Waele, and R. Bos, "Autoregressive spectral analysis when observations are missing," *Automatica*, vol. 40, no. 9, pp. 1495–1504, Sep. 2004.
- [13] R. H. Jones, "Maximum likelihood fitting of ARMA models to time series with missing observations," *Technometr.*, vol. 22, no. 3, pp. 389–395, Aug. 1980.
- [14] M. J. Tummers and D. M. Passchier, "Spectral analysis of biased LDA data," *Meas. Sci. Technol.*, vol. 12, no. 10, pp. 1641–1650, Oct. 2001.
- [15] D. K. McLaughlin and W. G. Tiedeman, "Biasing correction for individual realization of laser anemometer measurements in turbulent flows," *Phys. Fluids*, vol. 16, no. 12, pp. 2082–2088, 1973.
- [16] T. J. McDougall, "Bias correction for individual realisation LDA measurements," *J. Phys. E, Sci. Instrum.*, vol. 13, no. 1, pp. 53–60, Jan. 1980.
- [17] P. M. T. Broersen, "Bias contributions in time series models for resampled irregular data," in *Proc. I2MTC*, Victoria, BC, Canada, 2008, pp. 882–889.
- [18] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic, 1981.
- [19] P. M. T. Broersen, *Automatic Autocorrelation and Spectral Analysis*. London, U.K.: Springer-Verlag, 2006.
- [20] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1976.
- [21] E. J. Hannan and M. F. Deistler, *The Statistical Theory of Linear Systems*. New York: Wiley, 1988.
- [22] D. Britz and R. A. Antonia, "A comparison of computing power spectra of LDA signals," *Meas. Sci. Technol.*, vol. 7, no. 7, pp. 1042–1053, Jul. 1996.
- [23] P. M. T. Broersen, *ARMASel for Irregular or Missing Data, Matlab toolbox*, 2001. [Online]. Available: [www.mathworks.com/matlabcentral/fileexchange/?term=armasa](http://www.mathworks.com/matlabcentral/fileexchange/?term=armasa)
- [24] L. Barford, "Filtering of randomly sampled time-stamped measurements," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 2, pp. 222–227, Feb. 2008.
- [25] P. M. T. Broersen, "Application of time series models to the spectral analysis of irregular turbulence data," in *Proc. I2MTC*, Victoria, BC, Canada, 2008, pp. 33–38.
- [26] H. Nobach, *LDA benchmark generator III*, 2001. [Online]. Available: <http://www.nambis.de/benchmark>
- [27] W. K. Harteveld, R. F. Mudde, and H. E. A. van den Akker, "Estimation of turbulence power spectra for bubbly flows from laser Doppler anemometry signals," *Chem. Eng. Sci.*, vol. 60, no. 22, pp. 6160–6168, 2005.



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

He is currently with the Department of Multi-Scale Physics, Delft University of Technology. He developed statistical measures to let the measured data speak for themselves in estimated time series models. This provides a practical and accurate solution for the spectral and autocorrelation analysis of

stochastic data. Software for the automatic estimation of spectra and autocorrelation functions of equidistant random data, for missing data problems, and for irregularly sampled observations, is available online. His main research interest is in the automatic and unambiguous identification of the character of stationary random measurement data.