

# Optimal Sampling Distribution and Weighting for Ensemble Averages

Blair H. Brumley and Kent L. Deines, Member, IEEE

RD Instruments

9855 Businesspark Ave.

San Diego, CA 92131 USA

**Abstract**—In many current-measuring applications, limited energy constrains the number of velocity measurements that can be taken, and limited data storage capacity further constrains the number of measurements that can be recorded. Consequently, samples are combined into ensemble averages before recording, and the number of samples available per ensemble average is limited. Unlike traditional inherently-integrating current meters, which ideally record a uniformly weighted average of the velocity over the interval between recorded measurements, a sampling instrument like an acoustic Doppler current profiler (ADCP) is free to arbitrarily weight the contribution of the velocity over time to each recorded measurement. This is possible through the choice of sample timing and the use of weighting factors (windowing) in the ensemble average, subject to the constraint on the ratio of the average sample rate to the average recording rate. Examples of choices of sampling distributions include uniform intervals, pseudorandom intervals, and short bursts of uniformly spaced samples with uniform or pseudorandom intervals between bursts. This paper attempts to put the choice of sampling distribution and weights on a rational basis. An "optimal" choice should minimize either expected or worst-case spectral aliasing into the frequency range of interest to the data user, given limited *a priori* knowledge of the typical shape of the velocity spectrum and the statistics of the parameters that describe it. An optimal choice should also be robust to data dropout at a given probability, assuming independence of dropout among samples. An example is given where waves must be averaged out to measure tidal currents as accurately as possible.

## I. INTRODUCTION

Fluctuations in velocity occur over a huge range of time scales. The historical period over which oceanographers have been making measurements is the longest useful scale, while the observation period of a particular experiment is often the longest scale being considered. The shortest time scale available to the data user is determined by the "recording interval" at which velocities are recorded (i.e., by the Nyquist criterion). For a traditional inherently-integrating current meter, the number of rotor turns in the recording interval represents a uniformly-weighted average of the velocity over the recording interval (or would if it performed ideally).

Acoustic current meters are sampling instruments, typically having a low duty cycle, and hence are not inherently integrating devices. They record an ensemble average velocity over a set of samples distributed over the recording interval, or possibly a somewhat greater period. Each sample can be considered to be essentially an instantaneous

velocity measurement, since the averaging period of a single sample is on the order of milliseconds. Often battery limita-

tions constrain the number of samples that can be devoted to each ensemble average. Velocity fluctuations on time scales shorter than the recording interval (due to waves and turbulence, for example,) may not average out as completely as they would for an inherently-integrating current meter, aliasing energy from higher frequencies into the frequency band of interest to the data user. This aliased energy may pose a greater problem than instrument noise. However, the greater flexibility of a sampling instrument makes it potentially possible to reduce aliasing below the level of an inherently-averaging instrument.

This paper examines the question of how a limited number of instantaneous velocity samples can be optimally distributed in time and weighted into an ensemble average to minimize aliasing. A plausible formulation of this question as a well-posed optimization problem is presented, solution methods are discussed, and results for a typical oceanographic application are given.

## II. POSING THE PROBLEM

We assume that a geographic coordinate system is used, so that the ensemble averaging process acts independently on each velocity component, and only consider one component. Suppose that ensemble averages are desired at a uniform recording interval  $T_r$ . We compute the  $n^{\text{th}}$  ensemble average velocity  $V_n$  from a set of  $M$  samples  $v_{n,k}$  distributed about the nominal ensemble time  $nT_r$ .

$$V_n = \sum_{k=1}^M w_k v_{n,k}(nT_r + t_k) \quad (1)$$

$$\sum_{k=1}^M w_k = 1, \quad \sum_{k=1}^M w_k t_k = 0 \quad (2)$$

where  $w_k$  are weighting factors, and  $t_k$  are sample times relative to the nominal ensemble time  $nT_r$ .

For clarity, we have shown redundant information in Eqn. 1. by both indicating the time at which each velocity sample is taken and also indicating the two-dimensional indexing scheme we have chosen, in which we group all of the velocity samples that contribute to the  $n^{\text{th}}$  ensemble average into the  $n^{\text{th}}$  row of the data array. We allow samples to be used in more than one ensemble average and  $|t_k|$  to be greater than  $T_r/2$  or even  $T_r$ , though they need not be.

We wish to choose an optimal set of weights  $w_k$  and sample times  $t_k$ . To do this we must define what we mean by optimal and constrain the problem by stating what information is

available *a priori* about the statistics of the velocity samples. The information we need to deal with is contained in the power spectrum, as well as its Fourier transform, the autocorrelation function. More specifically, we are interested in the effect of the sampling/ensemble averaging scheme on the power spectrum and on the autocorrelation function of the continuous velocity time series.

We can view recorded ensemble averages as the result of a two-step process, digital (i.e., discrete-time) filtering followed by down-sampling (decimation). To make the distinction between the steps clear, suppose that we were to have two separate data channels using the same ensemble averaging scheme, the only difference being an arbitrary time lag  $\mathbf{t}$  between them. The (continuous) autocorrelation function  $R_v(\mathbf{t})$  is the expected value of the product of the ensemble averages recorded by these two channels.

$$\begin{aligned} R_v(\mathbf{t}) &= E[V_n(nT) V_n(nT_r + \mathbf{t})] \\ &= \sum_{p=1}^M \sum_{q=1}^M w_p w_q E[v_{n,p}(nT_r + t_p) v_{n,q}(nT_r + \mathbf{t} + t_q)] \\ &= \sum_{p=1}^M \sum_{q=1}^M w_p w_q R_v(\mathbf{t} + t_q - t_p) \end{aligned} \quad (3)$$

where  $R_v(\mathbf{t})$  is the underlying autocorrelation function of the velocity component. In depicting  $R_v(\mathbf{t})$  as independent of time, we have made an assumption of statistical stationarity, which will be discussed further below. We can replace the double sum in Eqn. 3 with a single one by remapping the weight products  $w_p w_q$  to a single sequence  $W_k$  corresponding to the time differences  $s_k = (t_q - t_p)$ , and adding up all products corresponding to the same time difference. (Let  $s_0 = 0$ .)

$$R_v(\mathbf{t}) = \sum_{k=1}^M W_k R_v(\mathbf{t} + s_k) \quad (4)$$

where

$$W_k = \sum_{p=1}^M \sum_{q=1}^M w_p w_q \mathbf{d}(s_k - t_q + t_p) \quad (5)$$

and  $\mathbf{d}$  is the Kronecker delta function having the property  $\mathbf{d}(0)=1$ . There is an  $s_k$  and a corresponding  $W_k$  for every distinct time lag between all possible pairs of the sample times  $t_q$ .  $W_k$  is a kind of autocorrelation function of the weights  $w_k$ , in the sense that these sequences can be considered to be continuous-time trains of impulses (Dirac delta functions) [1], which will be denoted  $W(t)$  and  $w(t)$ .

In a similar sense, Eqn. 4 shows that the effect of the first step (“digital filtering”) is to convolve the underlying autocorrelation function with  $W(t)$ . In the frequency domain, therefore, the effect of the first conceptual step on the power spectrum is to multiply it by  $S_W(f)$ , the Fourier transform of  $W(t)$ . The function  $S_W(f)$  is the squared-magnitude of the filter

frequency response, which we will call the “power gain” for short. The relationships are as follows:

$$S_v(f) = 4 \int_0^\infty R_v(\mathbf{t}) \cos(2\mathbf{p}f\mathbf{t}) d\mathbf{t}, \quad R_v(\mathbf{t}) = \int_0^\infty S_v(f) \cos(2\mathbf{p}f\mathbf{t}) df \quad (6)$$

$$S_{v1}(f) = 4 \int_0^\infty R_v(\mathbf{t}) \cos(2\mathbf{p}f\mathbf{t}) d\mathbf{t}, \quad R_v(\mathbf{t}) = \int_0^\infty S_{v1}(f) \cos(2\mathbf{p}f\mathbf{t}) df \quad (7)$$

$$S_W(f) = \sum_k W_k \cos(2\mathbf{p}fs_k) = \left| \sum_k w_k e^{i2\mathbf{p}fs_k} \right|^2 \quad (8)$$

$$W_k = \lim_{F \rightarrow \infty} \frac{1}{F} \int_{-F/2}^{F/2} S_W(f) \cos(2\mathbf{p}fs_k) df \quad (9)$$

$$S_{v1}(f) = S_W(f) S_v(f) \quad (10)$$

where  $S_{v1}(f)$  is the one-sided power spectrum resulting from the first conceptual step ( $S_W(f)$  is defined to be two-sided).

To implement the second conceptual step (“down-sampling”), we merely restrict  $\mathbf{t}$  to integer multiples of the recording interval  $T_r$ , so that we don’t actually require additional channels to supply the lagged measurements. The implication of this step in the frequency domain is that components of higher-frequency than the Nyquist frequency  $f_N=1/(2T_r)$  are aliased into the frequency band of interest between 0 and  $f_N$ . Since the data are real and the recording interval is constant, this aliasing occurs in the familiar “folding” manner, with reflections about odd multiples of  $f_N$ .

$$S_v(f) = 2T_r \sum_{n=-\infty}^{\infty} R_v(nT_r) \cos(2\mathbf{p}fnT_r) = \sum_{k=-\infty}^{\infty} S_{v1}(|f - 2kf_N|) \quad (11)$$

The advantage of this two-step conceptual approach is that no aliasing occurs until the final step. We do not have to worry about aliasing occurring during the sampling itself, which may be at irregular intervals and therefore may behave in ways beyond our experience and intuition.

We consider as “noise” all frequency components of the underlying spectrum  $S_v(f)$  at frequencies greater than  $f_N$ , since all of these unwanted components will be aliased down into the frequency band of interest after suppression by the power gain  $S_W(f)$ . Because the bandwidth of the interfering “noise” is usually large compared to the Nyquist frequency, the aliased noise will generally contaminate the entire useful frequency range almost uniformly. (However, “noise” just beyond the Nyquist frequency generally does not have a bandwidth this large.) Though other measures are possible, it is convenient to use as a cost function the aliased energy uniformly weighted at all frequencies, which is equivalent to  $R_v(0)$ , the variance of the ensemble averages considering only the unwanted “noise” signals above the Nyquist frequency  $f_N$ . Note that the down-sampling step that causes the aliasing does not affect this particular measure, except to the extent that it defines what part of the spectrum is considered “noise”.

We now have a well-posed problem that can be visualized in either the frequency or autocorrelation domain. We wish to choose sample times  $t_k$  and weights  $w_k$  as shown in Eqn. 1

so as to minimize the unwanted energy passed by the “digital filter” as shown in Eqn. 10 or, equivalently, so as to minimize the variance  $R_V(0)$  of the ensemble averages attributable to unwanted velocity signals (Eqns. 3 and 4).

### III. GENERAL PRINCIPLES

Before finding specific solutions, we can make some general statements that will apply to all or most situations. The filter power gain  $S_W(f)$  defined in Eqn. 8 is non-negative and symmetric, and has a peak value of 1 at the origin. It is periodic if and only if the sample time differences  $s_k$  are all rationally related to each other, in which case the period is the reciprocal of their greatest common divisor (or equivalently, the least common multiple of their reciprocals). In any case, by Eqn. 8 the average value of  $S_W(f)$  is  $W_0$ , equal to the sum of the squares of the  $w_k$ , which has a minimum value of  $1/M$  when the weights are uniform. Hence for the power gain to be less than  $1/M$  in some parts of the spectrum it must be greater than  $1/M$  in others so as to compensate.

Suppose we wish to design our filter to reject a signal that lies in the frequency band  $f_{min}$  to  $f_{max}$ , where  $f_{min}$  is well above the Nyquist frequency. The width of the envelope of the autocorrelation function  $R_v(\mathbf{t})$  is roughly the reciprocal of the signal bandwidth, and this envelope modulates an oscillation with a frequency somewhere in the frequency band. The autocorrelation function is symmetric and has zero area under it, the negative portions canceling out the positive ones.

The worst possible choice is to have samples so close together that they are all perfectly correlated. No matter how many samples are averaged nor what weights are chosen, the cost function (ensemble average variance) will always be equal to the signal variance  $\mathbf{s}^2$ . The (low-pass) pass band of  $S_W(f)$  is so wide that it passes the whole signal. In order for the pass band to be narrower than  $f_{min}$ , the samples must be spread over a “dwell” interval  $T_{dwell}$  greater than  $1/(2f_{min})$ .

Another strategy is to choose uniformly-weighted samples taken so far apart that they are uncorrelated. Since  $W_0=1/M$ , the cost function will be  $\mathbf{s}^2/M$ . If the samples are uniformly spaced,  $S_W(f)$  is a comb with teeth spaced closer together than the signal bandwidth, so as to cut out most of the signal but never all of it.

Suppose we happened to choose uniform sample intervals at a sampling frequency equal to the centroid of the “noise” signal spectrum. Since the autocorrelation function will be sampled only at the positive peaks, the digital filter will effectively be integrating the *envelope* of the autocorrelation function, or to put it another way, the signal peak will be aliased to zero frequency. In this “worst case” the cost function will be roughly  $\mathbf{s}^2$  divided by the product of the signal bandwidth and the dwell time  $T_{dwell}$  spanned by the samples (assuming this product is greater than 1). For narrowband signals the cost function will have the worst-case value of  $\mathbf{s}^2$ .

We can reduce the cost function far below the nominal value of  $\mathbf{s}^2/M$  for uncorrelated samples by sampling in bursts of samples that are closely spaced enough to include some of the negative values of the autocorrelation function in the convolution. It is remarkable that for narrowband signals (less than 100% bandwidth), the dwell time of a burst can be far shorter than the correlation time (reciprocal bandwidth) and still achieve this cancellation to some degree. Hence a sensible strategy would be to space out a set of bursts so that they are uncorrelated. It doesn’t really matter whether the bursts are spaced uniformly or in some other pattern (e.g., time-jittered [2] or pseudo-Poisson process [3,4]); the fine structure of  $S_W(f)$  will be affected by this pattern, but not the average over the stop band. In all cases, the cost function will be that of a single burst divided by the number of independent bursts.

In fact, we are free to choose a pattern of bursts of bursts; that is, the convolution of two burst patterns of much different dwell times. The power gain  $S_W(f)$  will then be the product of those corresponding to the two burst patterns. If our “noise” consists of two relatively narrowband signals (100% or less each) at widely separated frequencies, we can design the burst characteristics at each scale to optimally block the corresponding signal. Since  $M$  is the product of the two burst sizes, we would like to keep both burst sizes as small as possible, particularly that of the inner burst since it does nothing to help average out the lower-frequency signal.

Let us now consider a single burst of samples. As mentioned above, the dwell time for the burst must be greater than  $1/(2f_{min})$ . With equally-spaced samples,  $S_W(f)$  will be periodic at intervals equal to the sample rate, which should therefore be greater than  $f_{max}+1/(2T_{dwell})$ . For other sample spacings, a similar constraint applies to the average sample rate. Combining these constraints, to have a stop band wide enough to cover the entire signal requires a minimum number of samples per burst  $M_b$ :

$$M_b \gg \frac{1}{2} \left( \frac{f_{max}}{f_{min}} + 1 \right) \quad (12)$$

This criterion suggests that with only a handful of samples per burst, we can block a signal of order 100% bandwidth or less.

### IV. OPTIMAL SHORT BURSTS

To find conditions for an optimal solution, let us augment the cost function  $\Phi=R_V(0)$  from Eqn. 3 (at  $\mathbf{t}=0$ ) with the constraints of Eqn. 2 using Lagrangian multipliers:

$$\Phi = \sum_{p=1}^{M_b} \sum_{q=1}^{M_b} w_p R_{pq} w_q + I_1 \left( 1 - \sum_{k=1}^{M_b} w_k \right) - I_2 \sum_{k=1}^{M_b} w_k t_k \quad (13)$$

where  $R_{pq}$  represents the symmetric matrix  $R_v(t_q-t_p)$ . We can find local extrema of the cost function  $\Phi$  by setting the derivatives of Eqn. 13 with respect to the parameters to zero:

$$\frac{\mathcal{I}\Phi}{\mathcal{I}w_k} = 0 \Rightarrow w_k = \frac{\sum_{n=1}^{M_b} R_{kn}^{-1}}{\sum_{p=1}^{M_b} \sum_{q=1}^{M_b} R_{pq}^{-1}} \quad (14)$$

$$\frac{\mathcal{I}\Phi}{\mathcal{I}t_k} = 0 \Rightarrow \sum_{n=1}^{M_b} R'_{kn} w_n = \frac{1}{2} \mathbf{I}_2 = 0, k = 1, 2, \dots, M_b \quad (15)$$

$$\Phi_{\min} = \left[ \sum_{p=1}^{M_b} \sum_{q=1}^{M_b} R_{pq}^{-1} \right]^{-1} \quad (16)$$

where  $R_{kn}^{-1}$  represents the inverse matrix of  $R_{nk}$  and  $R'_{kn}$  represents the skew symmetric matrix  $\partial R_v(\tau)/\partial \tau$  at  $\tau = t_n - t_k$ , which by Eqn. 15 must be singular. Eqn. 15 requires that the weights satisfying Eqn. 14 also form an eigenvector of  $R'_{kn}$  (corresponding to a zero eigenvalue) when the sample times are locally optimal. All real skew symmetric matrices of odd rank are singular, but when the number of samples in the burst  $M_b$  is even, the requirement that the determinant of  $R'_{kn}$  be zero may give a useful constraint. The weights and sample times need not be symmetric about the origin, but in the special case of equally-spaced samples they must be.

In the case of sample pairs ( $M_b = 2$ ), Eqn. 14 requires that  $w_1 = w_2 = 0.5$  in all cases, and Eqn. 15 (or zeroing the determinant) requires that the pair spacing ( $t_2 - t_1$ ) be chosen at an extremum of the autocorrelation function of the “noise” signal. The highest pair variance will be at the peaks, the lowest at the valleys. At all extrema the pair variance will be  $(1 + \mathbf{r})/2$  times the signal variance, where  $\mathbf{r}$  is the pair correlator  $R_v(t_2 - t_1)/R_v(0)$ . For narrowband signals sample pairs at half-period spacing, the factor  $(1 + \mathbf{r})/2$  is approximately 0.445 times the square of the -3dB fractional width.

For bursts with more than two samples, it is not possible in general to sample the correlation function only at the extrema though if it were possible such solutions would of course be locally extremal since they would satisfy Eqn. 15.

Eqns. 14-16 were solved numerically for 2-, 3-, and 4-sample bursts for a Pierson-Moskowitz wave orbital velocity spectrum normalized to unit variance [5]:

$$S_v(f) = \sqrt{\frac{20}{p}} (fT)^{-3} \exp\left[-\frac{5}{4}(fT)^{-4}\right]$$

where the constant  $T$  is the wave period corresponding to the peak of the wave height spectrum.

This shape has a -3 dB bandwidth of  $0.9/T$  and a moment bandwidth of  $1.2/T$ , characteristic of a developed wave field. Though wave spectra in coastal areas are typically narrower than this (the Donelan-Hamilton model [6] is more realistic), the wide bandwidth may be considered to be representative of spreading due to a degree of variability in the wave period that we wish to account for. Table I gives optimal parameter values for this spectrum, and shows the degree to which averaging a single burst suppresses the “noise” from wave energy.

TABLE I  
OPTIMAL VALUES FOR PIERSON-MOSKOWITZ SPECTRUM

$M_b$	$t_k/T$	$w_k$	Ensemble	Variance
			dB re $\mathbf{s}^2$	dB re $\mathbf{s}^2/M_b$
2	-0.16	0.5	-5.7	-2.7
	+0.16	0.5		
3	-0.27	0.311	-9.3	-4.5
	0.00	0.378		
	+0.27	0.311		
4	-0.345	0.221	-12.0	-6.0
	-0.105	0.279		
	+0.105	0.279		
	+0.345	0.221		
4 even spacing	-0.345	0.213	-11.9	-5.9
	-0.115	0.287		
	+0.115	0.287		
	+0.345	0.213		

For 4-sample bursts, the optimal scheme is symmetric but with slightly non-uniform sample spacing. The more practical uniform  $0.23T$  spacing shown gives performance only 0.06 dB worse. Using uniform weighting as well as uniform spacing (not shown) degrades the performance by another 0.52 dB. Also note that the (average) interval between samples decreases with  $M_b$ , while the dwell time increases.

As explained above, an ensemble average may consist of several bursts so as to average out lower-frequency energy not represented by the Pierson-Moskowitz spectrum. The vari-

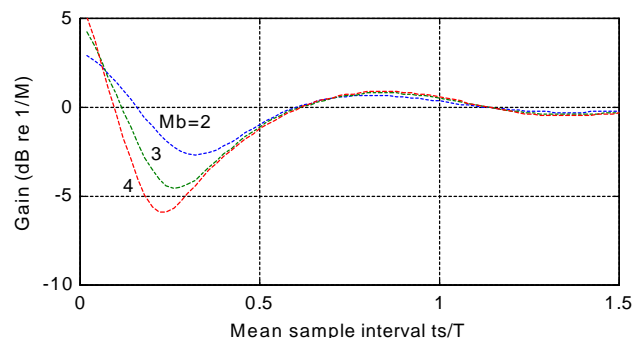


Fig. 1. Sensitivity of ensemble variance  $\times M$  for optimal 2-, 3-, and 4-sample bursts from Table I.

same thing as equalizing the contributions from the two noise sources after taking into account the extra wave suppression shown in the “dB re  $\mathbf{s}^2$ ” column of Table I.

If the sample rate is chosen poorly relative to the wave period, the potential benefit of burst sampling may not be achieved. This sensitivity to sample interval (or equivalently, to wave period) is shown in Fig. 1.

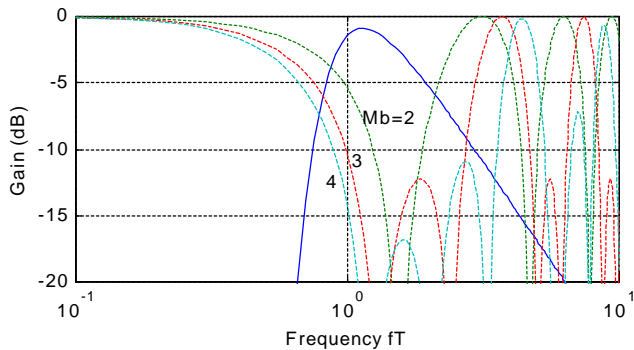


Fig. 2. Pierson-Moskowitz spectrum and filter power gain (Eqn. 8) for optimal 2-,3-,and 4-sample bursts from Table I.

The performance of a burst average can be considerably degraded if even one sample is missing, particularly if  $M_b$  is small. In a conventional ensemble averaging scheme that divides the cumulative weighted sum of the valid samples by the cumulative sum of the corresponding weights, this degradation is partly offset by the reduced representation of the burst when data is missing (i.e., the sum of the valid weights is less than one). Since the burst variance for each case of missing data can be estimated and stored, in principle an optimal approach would be to apply a reciprocal-variance weighting to each burst average, and at the end divide by the cumulative sum of these burst weights.

## V. PRACTICAL DESIGN CONSIDERATIONS

In the previous section we calculated some optimal sampling schemes applicable when the peak wave period is relatively well known. More typically, at the time of deployment planning the wave period is only known to likely fall within a relatively broad confidence band. Moreover, the confidence bands for wind waves and swell typically overlap, and therefore must be treated as one large band. In practical situations, therefore, the needed spectral stopband will have a ratio  $f_{max}/f_{min}$  that is 10 or more (e.g., from 20 s to 2 s period). By Eqn. 12, allowing a factor of two for filter rolloff, we will need at least a dozen or so samples per burst to filter out all kinds of waves that are likely to be encountered.

There are two different approaches to specifying the optimal filter. One approach is to specify a noise spectrum representing the probability-weighted spectrum of “noise” signals that are predicted to occur during the deployment using available *a priori* information. The Fourier transform of this spectrum can be used in Eqns. 14-16 just as was done with the Pierson-Moskowitz spectrum above.

However, not everyone is happy with a scheme whose average performance is optimal. Those who are risk-averse

would prefer to minimize the worst-case performance over some range of reasonably likely conditions. A filter with a stop band with large ripples may have a low average level over the climatological range of wave conditions, but if a narrowband swell happens to fall on a tall ripple, the performance at that time might be poor. We can idealize this approach to consider an infinitely narrowband wave field at any one time that might fall anywhere in the stop band, leading immediately to an equal-ripple filter design specification. That is, we wish to minimize the largest peaks in the stop band, rather than the weighted-average value. The gain in the stop band need not be uniform, but rather can be specified arbitrarily. If the sampling scheme is symmetrical in time with rationally-related intervals, the equiripple design can be carried out numerically using an iterative approach similar to the Parks-McClellan algorithm, except that the solution must be constrained to allow only a small number ( $M$ ) of non-zero coefficients (the weights  $w_k$ ) out of a large number of possible sampling times.

Now let us consider the related but conceptually separate issue of near-Nyquist aliasing, the aliasing that occurs if there is “noise” just beyond the Nyquist frequency and if the filter passband is too wide to filter out this noise. Near-Nyquist aliasing may occur for any scheme with non-overlapping sampling windows for consecutive ensembles, which includes inherently-integrating current meters. With a sampling instrument, it is quite feasible to include each sample in two or three consecutive ensembles, shaping the filter rolloff and narrowing the passband by applying a weighting envelope that is somewhat wider than the recording interval. There is inherently a tradeoff between near-Nyquist aliasing and attenuation (reduced signal-to-instrument-noise spectral level ratio) of the desired velocity signal at high frequencies just below the Nyquist (in other words, reduced time resolution). In addition there is another tradeoff between high sidelobes in the frequency domain and ringing in the time domain. Tradition apparently dictates a choice that simulates an inherently-integrating current meter.

In deriving Eqn. 3, we made an assumption of statistical stationarity, which is really only an approximation that the variance of the “noise” signal does not change significantly during the recording interval. This assumption is closely related to the assumption that there is no cross-correlation between separate spectral components. An example of failure of this assumption is the case of tidal harmonics in an estuary that are phase-locked to the fundamental frequency. If the Nyquist frequency is poorly chosen such that one of these phase-locked harmonics folds over on top of the fundamental, the autocorrelation function of the ensemble averages may diverge from that of the continuous velocity signal due to the interaction of the different spectral components. Aside from such unlikely cases, however, the approximation of statistical stationarity is quite adequate.

## VI. CONCLUSION

Aliasing of unwanted high-frequency "noise" signals such as wave orbital velocities into the frequency band of interest can be characterized by the variance of an individual recorded ensemble average. Significant reduction in ensemble variance can often be achieved by clustering the samples that contribute to the ensemble average into one or more bursts rather than spacing them out uniformly. For narrowband "noise", a remarkable reduction is possible with simple sample pairs with half-period spacing, but only for unusually predictable narrowband signals such as tides and harbor resonances is it possible to plan in advance what spacing to use. In practical situations, it is common to deal with "noise" signals such as waves that are either inherently broadband or must be treated as effectively so due to variability in the peak frequency. At least a dozen or so samples are needed to be assured of suppressing "noise" that might lie anywhere in a stop band of 10:1 frequency ratio. An equiripple approach to filter design minimizes the ensemble variance for worst-case conditions. It is desirable to include several bursts spread out either uniformly or pseudorandomly over the recording interval to provide at least some suppression of "noise" that may occur at frequencies between the Nyquist and the reciprocal of the burst duration.

In recent years, data storage technology has improved by several different performance measures (e.g., storage density, energy used to store one bit, cost per bit) at a dramatically

higher rate than that of battery technology. If that trend continues, ensemble averaging will soon give way to storage of raw sampled data for later post-processing in self-contained instruments with physical data recovery. Non-uniform sampling as discussed in this paper will still be needed to reduce aliasing in many cases. A second trend toward non-physical data recovery (e.g., via acoustic modem) will continue the need for the data compression provided by ensemble averaging for a long time to come.

## ACKNOWLEDGMENT

Eugene Terray of the Woods Hole Oceanographic Institution provided much useful advice, especially concerning wave spectrum models.

## REFERENCES

- [1] Alan V. Oppenheim and Ronald W. Schaffer, *Discrete-Time Signal Processing*. Prentice-Hall, 1989, p. 81.
- [2] A.V. Balakrishnan, "On the problem of time jitter in sampling," *IRE Trans. Inf. Theory*, vol. 8[3], pp. 226-236, April 1962.
- [3] Frederick J. Beutler, "Alias-free randomly timed sampling of stochastic processes," *IEEE Trans. Inf. Theory*, vol. 16[2], pp. 147-152, March 1970.
- [4] Elias Masry, "Random Sampling and Reconstruction of Spectra," *Inf. & Control*, vol. 19, pp. 275-288, 1971.
- [5] Pierson and Moskowitz, *JGR*, vol. 69, pp. 5181-5203, 1964.
- [6] Mark A. Donelan, *Phil. Trans. Royal. Soc. London*, vol. A315, 509-562.

??

It is convenient to restrict the ratios of the  $t_k$  to rational fractions (as opposed to irrational numbers) and imagine the acquired samples to be a subset of potential samples that could be taken at equal and arbitrarily short intervals.

Promised to discuss statistical stationarity.

(18)

This is a reference to Eqn. **Errore. Il segnalibro non è definito.**

[waves and wideband background]  
[figure showing spectrum, optimal filter function, and product]

[Boneyard

[It would be nice if the  $W_k$  sequence could be designed to “sample” the autocorrelation function at the negative peaks and avoid the positive ones, but alas that is only possible in the trivial case of sample pairs. (do this case)]

Suppose we construct a sampling pattern composed of bursts of bursts; that is, the convolution of two bursts. The filter  $S_w(f)$  will then be the product of the filters corresponding to the two bursts. If our “noise” consists of two relatively narrowband signals (100% or less each) at widely separated frequencies, we can design the burst characteristics at each scale to optimally block the corresponding signal. Since  $M$  is the product of the two burst sizes, we would like to keep both burst sizes as small as possible.]

Near-Nyquist aliasing will occur for any scheme with non-overlapping sampling windows for consecutive ensembles, which includes inherently-integrating current meters.

there is rarely any hope of confining the aliased “noise” to a small part of the useful spectrum conveniently far

»

It may be helpful to visualize  $W_k(t)$  as the product of a broad smooth “filter” function a sampling function composed of a series of impulses, in which case its Fourier transform can be thought of as a power gain function convolved with the Fourier transform of the sampling function.