TECHNICAL RESEARCH REPORT

On Periodic Pulse Interval Analysis with Outliers and Missing Observations

by B.M. Sadler and S.D. Casey

T.R. 96-6

Sponsored by
the National Science Foundation
Engineering Research Center Program,
the University of Maryland,
Harvard University,
and Industry

On Periodic Pulse Interval Analysis with Outliers and Missing Observations

Brian M. Sadler[1] and Stephen D. Casey[2,3]
January 23, 1996

Abstract

Analysis of periodic pulse trains based on time of arrival is considered, with perhaps very many missing observations and contaminated data. A period estimator is developed based on a modified Euclidean algorithm. This algorithm is a computationally simple, robust method for estimating the greatest common divisor of a noisy contaminated data set. The resulting estimate, while not maximum likelihood, is used as initialization in a three-step algorithm that achieves the Cramer-Rao bound for moderate noise levels, as shown by comparing Monte Carlo results with the Cramer-Rao bounds. An extension using multiple independent data records is also developed that overcomes high levels of contamination.

EDICS: 3.6.1, 3.1.1

[1]Army Research Laboratory, AMSRL-IS-TA, Adelphi, MD 20783, *sadler@arl.mil,* 301-394-2520, (f) x4214.
[2]Institute for Systems Research, University of Maryland, College Park, MD 20742–3311, *scasey@isr.umd.edu.*
[3]Research partially supported by Air Force Office of Scientific Research Grant F49620–94–1–0196.

1 Introduction

Analysis of pulse trains is a long standing problem, with applications including radar [7,8,10,17,18], communications [8,9], neurology [1,5,22], and astronomy [21]. For example, in radar it may be desired to accurately estimate the pulse repetition interval (PRI) [19,25]. A similar problem is frequency estimation in additive white Gaussian noise at moderate to high signal-to-noise ratio using phase data [12,23]. At its most fundamental, pulse train analysis is based on time of arrival information only, as might be obtained from a matched filter or other detector. We assume time is highly resolved and ignore any time quantization error. In this work we are primarily concerned with a single periodic pulse train with (perhaps very many) missing observations that may be contaminated with outliers. Our data model for this case, in terms of the arrival times t_j, is given by

$$t_j = \phi + k_j T + \eta_j, \quad j = 1, \ldots, N, \tag{1}$$

where T is the unknown period, $\phi \sim \mathcal{U}[0, T)$ is a uniformly distributed random phase, the k_j's are positive non-repeating integers, and η_j is zero-mean additive white Gaussian noise with variance σ_η^2 such that $3\sigma_\eta \leq \frac{T}{2}$. We let S be a sample set of N arrival times, i.e.,

$$S = \{t_j\}_{j=1}^{N}. \tag{2}$$

Without loss of generality we assume the elements of S are indexed in descending order, so $t_j > t_{j+1}$ for $j = 1, \ldots, N-1$. This assumption will be useful later and does not represent any practical difficulty. This model allows for missing observations through the distribution of the k_j's, e.g., with N even, $k_j = N, N-2, N-4, \ldots, 2$ implies observations at $k_j = N-1, N-3, \ldots$ are missing. Outliers are not explicitly included in (1), although they are considered later. The assumption of Gaussian noise can be relaxed to include iid non-Gaussian noise, although the results are no longer optimal. Later the assumption on ϕ is relaxed to allow for any positive constant, and given a realization of S we regard it as an unknown constant. Equation (1) is also useful for modeling sinusoidal zero-crossing times at moderate to high signal-to-noise ratio

[20]. Alternative data models, commonly used in neurology, assume "integrate and fire" pulse generation, e.g., see Brillinger [5] and references therein. Cumulative error terms may also be included in some cases [10].

Given the sample set S the problem is to recover the period T and possibly the phase ϕ. The minimum variance unbiased estimates for this linear regression problem take a least-squares form (see equations (3), (4), and (7)). However, this requires knowledge of the k_j's. We therefore propose a multi-step procedure that proceeds by (i) estimating T directly, (ii) estimating the k_j's, and (iii) refining the estimate of T using the estimated k_j's in the least-squares solution. This estimate is shown to perform well, achieving the Cramer-Rao bound in many cases, despite many missing observations and contaminated data.

The direct estimate of T (step (i) above) is obtained using a modified Euclidean algorithm [6]. This approach is motivated by the fact that, in the noise-free case, the value of T is very likely to be the greatest common divisor (gcd) of a set of samples $\{t_j - \phi\}_{j=1}^{N}$, for $N \geq 10$. The modified Euclidean algorithm is a computationally simple, robust method for estimating the gcd from a noisy contaminated data set.

The paper is organized as follows. The next section gives the maximum likelihood solution and Cramer-Rao bounds for estimating T and ϕ. Our analysis has led us to work with the data set $\{t_j - t_{j+1}\}_{j=1}^{N-1}$ so as to avoid estimating ϕ, which can be unreliable. For completeness, related work based on a point process (zero-one time series) viewpoint is briefly reviewed in Section 3. Next, we describe our modified Euclidean algorithms for estimating T, including a robust version for use with contaminated data. In section 5 the three-step refined estimation procedure mentioned above is developed. Its performance is compared to Cramer-Rao bounds via Monte Carlo simulation, revealing the very good performance of the algorithm with many missing observations and contaminated data. Also, an approach using multiple independent data records is developed that overcomes very high levels of contamination. Before concluding, we mention in Section 6 the "common oscillator" problem and its solution using our approach.

2 Maximum-Likelihood Estimation

Given the sample data set S from (2) we may write

$$\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix} = \begin{bmatrix} 1 & k_1 \\ 1 & k_2 \\ \vdots & \vdots \\ 1 & k_N \end{bmatrix} \begin{bmatrix} \phi \\ T \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_N \end{bmatrix}, \tag{3}$$

where $k_j \in \mathcal{N}$ (the natural numbers) and $k_j > k_{j+1}$. In compact form this is

$$\mathbf{t} = X\boldsymbol{\beta} + \boldsymbol{\eta}, \tag{4}$$

where $\boldsymbol{\beta} = [\phi, T]^T$ and the definitions of $\mathbf{t}$, $\boldsymbol{\eta}$, and X follow from (3). We eliminate ϕ by forming the differences $y_j = t_j - t_{j+1} = (k_j - k_{j+1})T + (\eta_j - \eta_{j+1})$, yielding

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} k_1 - k_2 \\ k_2 - k_3 \\ \vdots \\ k_{N-1} - k_N \end{bmatrix} T + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{N-1} \end{bmatrix}, \tag{5}$$

where $\delta_j = \eta_j - \eta_{j+1}$. As in (4), we may write (5) compactly in an obvious notation as

$$\mathbf{y} = X_d T + \boldsymbol{\delta}. \tag{6}$$

Using (6) relaxes the assumption on ϕ, allowing it to be any positive constant.

Equations (4) and (6) are linear regression problems whose least-squares solutions yield the minimum-variance unbiased estimate when the noise is zero-mean Gaussian, e.g., see Kay [13]. Denoting the noise covariance matrix as $R_\eta = E[\boldsymbol{\eta}\boldsymbol{\eta}^T]$, the solution to (4) corresponds to maximum likelihood (ML) estimation and takes the form of a least-squares estimate

$$\widehat{\boldsymbol{\beta}} = (X^T R_\eta^{-1} X)^{-1} X^T R_\eta^{-1} \mathbf{t}. \tag{7}$$

For white noise (7) simplifies further with $R_\eta = \sigma_\eta^2 I$. Similarly, the solution to (6) is

$$\widehat{T} = (X_d^T R_\delta^{-1} X_d)^{-1} X_d^T R_\delta^{-1} \mathbf{y}, \tag{8}$$

where $R_\delta = E[\boldsymbol{\delta}\boldsymbol{\delta}^T]$. We have assumed white noise so that

$$E[\delta_n \delta_{n+k}] = \begin{cases} 2\sigma_\eta^2, & k = 0 \\ -\sigma_\eta^2, & k = \pm 1 \\ 0, & \text{else} \end{cases} \tag{9}$$

and $R_\delta = \sigma_\eta^2 \tilde{R}_\delta$ where $\tilde{R}_\delta$ has 2's on the main diagonal, -1's on the first upper and lower diagonals, and zeros elsewhere. In general R_δ is full rank and its inverse can be expressed element-wise as $[R_\delta^{-1}]_{ij} = \sigma_\eta^{-2}(\min(i,j) - ij/(N+1))$, and is therefore easily computed [12]. Note also that knowledge of σ_η^2 is not necessary for calculating (7) and (8) due to cancellation.

Although optimal, use of these estimators requires knowledge of the coefficient matrices X or X_d. Obviously, this is not a problem if there are no missing observations, for then $k_j = N + 1 - j$ for $j = 1, 2, \ldots N$. However, when observations are arbitrarily missing and so the k_j's are not known in general, one is faced with more unknowns than equations in (4) and (6). One possible solution to this dilemma is to refine the estimate of T through a multi-step procedure. First use an alternative method for estimating T that does not require knowledge of the k_j's, use the estimate of T to estimate the k_j's, and then use the estimated k_j's in (4) or (6) in an attempt to refine the initial estimate of T. Later we take this approach using a modified Euclidean algorithm to obtain the initial estimate of T without knowledge of the k_j's.

Before ending this section we describe the Cramer-Rao bound (CRB) on the ML estimates produced by (7) and (8). The pdf of the observations $\mathbf{t}$ in (1) is multivariate Gaussian given by

$$p(\mathbf{t}|\boldsymbol{\beta}) = (2\pi)^{-N/2}\sigma_\eta^{-N} \exp\left\{\frac{-1}{2\sigma_\eta^2}(\mathbf{t} - X\boldsymbol{\beta})^T(\mathbf{t} - X\boldsymbol{\beta})\right\}, \tag{10}$$

yielding the log-likelihood function

$$\ln p(\mathbf{t}|\boldsymbol{\beta}) = \frac{-N}{2}(\ln 2\pi - \ln \sigma_\eta^2) - \frac{1}{2\sigma_\eta^2}(\mathbf{t} - X\boldsymbol{\beta})^T(\mathbf{t} - X\boldsymbol{\beta}). \tag{11}$$

The Fisher information matrix is found to be

$$J = -E[\nabla_\beta(\nabla_\beta \ln p(\mathbf{t}|\boldsymbol{\beta}))^T] = \frac{1}{\sigma_\eta^2} X^T X, \tag{12}$$

so that the CRB on β_i is

$$\text{var}\{\beta_i - \widehat{\beta}_i\} \geq [J^{-1}]_{ii}, \tag{13}$$

where β_i is the ith element of $\boldsymbol{\beta}$ and $[J^{-1}]_{ii}$ is the (i,i) element of J^{-1}. The inverse of X^TX is readily found to be

$$(X^TX)^{-1} = \frac{1}{N\sum_{j=1}^N k_j^2 - (\sum_{j=1}^N k_j)^2} \begin{bmatrix} \sum_{j=1}^N k_j^2 & -\sum_{j=1}^N k_j \\ -\sum_{j=1}^N k_j & N \end{bmatrix}, \tag{14}$$

so that

$$\text{var}\{T - \widehat{T}\} \geq \frac{N\sigma_\eta^2}{N\sum_{j=1}^N k_j^2 - (\sum_{j=1}^N k_j)^2}, \tag{15}$$

and

$$\text{var}\{\phi - \widehat{\phi}\} \geq \frac{\sigma_\eta^2 \sum_{j=1}^N k_j^2}{N\sum_{j=1}^N k_j^2 - (\sum_{j=1}^N k_j)^2}. \tag{16}$$

From (15) we observe that the CRB is reduced for smaller σ_η^2. Also, for fixed N, it is reduced when the spread of the k_j's increases.

From the model of (6) we have

$$p(\mathbf{y}|\boldsymbol{\beta}) = (2\pi)^{-N/2}\sigma_\delta^{-N} \exp\left\{\frac{-1}{2\sigma_\delta^2}(\mathbf{t} - X_d\boldsymbol{\beta})^T\tilde{R}_\delta^{-1}(\mathbf{t} - X_d\boldsymbol{\beta})\right\}, \tag{17}$$

with $\sigma_\delta = 2\sigma_\eta$. The Fisher information matrix for this case is

$$J = \frac{1}{\sigma_\delta^2} X_d^T \tilde{R}_\delta^{-1} X_d, \tag{18}$$

with the CRB given by

$$\text{var}\{T - \widehat{T}\} \geq \sigma_\delta^2 (X_d^T \tilde{R}_\delta^{-1} X_d)^{-1}, \tag{19}$$

which can be evaluated numerically given X_d.

Remark

1. We may form ML estimates of T from (7) or (8), with corresponding CRB's given by (15) and (19). However, we must be careful when comparing the CRB's because the two estimates are based on different data.

3 Analysis of Point Processes

Before describing the modified Euclidean algorithm (in section 4) we briefly discuss some alternatives to the least-squares ML estimates. The general approach is to view arrival times t_j as generating a zero-one time series or delta-train given by

$$x(t) = \sum_{j=1}^{N} \delta(t - t_j). \tag{20}$$

Using (1) we generate a corresponding $x(t)$ that is a periodic pulse train consisting of jittered data with missing observations. More complicated models include outliers and other pulse trains interleaved.

In searching $x(t)$ for the presence of a periodic delta-train $s(t)$ given by

$$s(t) = \sum_{j=1}^{N} \delta(t - jT), \tag{21}$$

it is natural and intuitively appealing to circularly convolve $x(t)$ and $s(t)$. This is equivalent to (circularly) applying a matched filter to $s(t)$. Note that the matched filter is not an optimal detector for $s(t)$ because the noise in $x(t)$ manifests itself in pulse-position rather than amplitude. Because we are dealing with zero-one time series the convolution is equivalent to counting pulses for a specific hypothesized phase ϕ and period T. Indeed, perhaps the conceptually simplest approach to detecting $s(t)$ is to use histograms [25]. While counting schemes are straightforward, they can be cumbersome to implement, and the resulting histograms may be difficult to interpret [25]. In cases of multiple interleaved pulse trains more sophisticated counting schemes rely on reducing the search space over T in clever ways [17,18].

Another approach is to perform spectral analysis of the point-process $x(t)$. This can be accomplished using a periodogram, given by

$$P_x(f) = \frac{1}{N} \left| \sum_{j=1}^{N} e^{2\pi i f t_j} \right|^2, \tag{22}$$

e.g., see Bartlett [2] and Brillinger [4]. A search for peaks of $P_x(f)$ produces candidate estimates of T. Peaks of $P_x(f)$ have been shown to yield approximate ML estimates of T with respect to

the point process $x(t)$ [8]. Note that this does not imply that use of $P_x(f)$ is ML for the data set S. Let the noise η_j have the pdf

$$f(\eta) = \frac{e^{\alpha \cos(2\pi\eta/T)}}{2\pi I_0(\alpha)}, \quad |\eta| \leq \frac{T}{2}, \tag{23}$$

with parameter α, where $I_0(\alpha)$ is the zero-th order Bessel function evaluated at α. This pdf was used by Van Trees for analyzing phase-lock synchronization schemes [24, ch. 3]. Note that $f(\eta)$ tends to a uniform pdf as $\alpha \to 0$, and tends to a Gaussian pdf as $\alpha \to \infty$. This is a fortuitous choice for the noise pdf, as it results in a maximum likelihood estimate of T given by (see the Appendix)

$$\widehat{T}_x = \text{argmax}_T \left| \sum_{j=1}^{N} e^{2\pi i \frac{t_j}{T}} \right|^2. \tag{24}$$

We use the notation $\widehat{T}_x$ so as not to confuse it (which is ML for (20)) with other period estimates, e.g., the ML estimate given by (7) for equation (3) corresponding to the data of (1). It may similarly be shown that the ML estimate of ϕ for this pdf is given by

$$\widehat{\phi}_x = \left(\frac{T}{2\pi}\right) \arg \left\{ \sum_{j=1}^{N} e^{2\pi i \frac{t_j}{T}} \right\}. \tag{25}$$

These estimators depend directly on the exponential sum (as does the periodogram), and do not require knowledge of the pdf parameter α. Use of (24) requires a fine search over $f = 1/T$. This may result in a significant computational load because the FFT is not a good computational option, versus direct computation of the exponential sum, given a large number of zeros in $x(t)$ [6]. Nevertheless, with no other information, the use of $P_x(f)$ is well founded, especially for multiple interleaved pulse trains without too many missing observations.

A procedure related to the periodogram is the use of circular statistics, which is a method for synchronized averaging. This approach has been adopted for neuronal pulse train analysis [1], as well as for PRI detection [7]. For $\rho > 0$, a circular statistic is given by

$$\left\langle \frac{t_j}{\rho} \right\rangle = \frac{t_j}{\rho} - \left\lfloor \frac{t_j}{\rho} \right\rfloor, \tag{26}$$

where $\lfloor \cdot \rfloor$ is the floor function. Thus $\langle \cdot \rangle$ is the fractional part, and so $\left\langle \frac{t_j}{\rho} \right\rangle \in [0,1)$. If ρ is equal to T/k for positive integer k, then a histogram of $\left\langle \frac{t_j}{\rho} \right\rangle$ will exhibit a peak. The function $\left\langle \frac{t_j}{\rho} \right\rangle$ is a mapping of t_j onto a circle (or folding onto an interval). It is interesting to note that

$$\sum_{j=1}^{N} e^{2\pi i \frac{t_j}{\rho}} = \sum_{j=1}^{N} e^{2\pi i \left\langle \frac{t_j}{\rho} \right\rangle}, \tag{27}$$

so that the Fourier transform of $x(t)/\rho$ gives information similar to a histogram of t_j/ρ mapped onto the unit circle in the complex plane.

4 Modified Euclidean Algorithms for Estimating T

In this section we describe modifications to the Euclidean algorithm for estimating T in (1) that do not require knowledge of the k_j's [6]. Without knowledge of the k_j's the resulting estimates are not ML. However, as described in the next section, the estimates produced can be refined to yield results comparable to ML estimation for moderate noise levels. Motivation comes from the fact that, in the noise-free case with $\phi = 0$, T is very likely the greatest common divisor of the set S given by (2).

The Euclidean algorithm is the method for finding the gcd of a set of positive integers, e.g., see Hardy and Wright [11], Leveque [15], or Knuth [14]. The algorithm is a division process for the set of integers $\mathbf{Z}$. It is based on the property that, given two positive integers a and b, $a > b$, there exist two unique positive integers q and r such that

$$a = q \cdot b + r \,,\, 0 \leq r < b \,. \tag{28}$$

If $r = 0$, we say that b divides a. This property of the set of integers, combined with the fact that if $a, b \in \mathbf{Z} \setminus \{0\}$, then $a \cdot b \neq 0$ ($\mathbf{Z}$ has no zero divisors), make $\mathbf{Z}$ a unique factorization domain. Therefore, in $\mathbf{Z}$ every non-zero element may be written uniquely as the product of powers of irreducible integers, or primes. The Euclidean algorithm involves repeated application of (28) until the remainder is zero. This yields $\gcd(a, b)$, which is the product of the prime factors that divide both a and b.

The Euclidean algorithm for finding $\gcd(a, b)$ can be simply stated as follows, e.g., see Knuth [14, p.319].

1) Find remainder of a/b.

2) If remainder equals zero, done, and $\gcd(a, b) = b$.

3) Set $a \leftarrow b$, $b \leftarrow$ remainder, go to step 1.

We demonstrate the algorithm showing that $\gcd(252, 198) = 18 = 2 \cdot 3^2$.

$$\begin{aligned} 252 &= 1 \cdot 198 + 54 \\ 198 &= 3 \cdot 54 + 36 \\ 54 &= 1 \cdot 36 + 18 \\ 36 &= 2 \cdot 18. \end{aligned}$$

The algorithm can be extended to find the gcd of a set of more than two integers by finding the gcd pairwise [14, p. 324]. We can also compute the gcd of a set whose elements are multiples of a fixed real number, as is the case of interest here given the set S. It can be shown that

$$\gcd(k_1 T, \ldots, k_N T) = T \gcd(k_1, \ldots, k_N) \tag{29}$$

(see Leveque [15, pg. 16]). This more general notion of gcd is consistent with Euclid's original algorithm [14, pp. 317–320].

The standard Euclidean algorithm, involving repeated division, is intolerant to the presence of noise in (1). We have therefore developed a modified Euclidean algorithm (MEA) to gain robustness to noise. The MEA exploits the fact that [6]

$$\gcd(k_1, \ldots, k_N) = \gcd((k_1 - k_2), (k_2 - k_3), \ldots, (k_{N-1} - k_N), k_N)\,. \tag{30}$$

It uses repeated subtraction rather than division. The basic iterative MEA is stated as follows. Recall that S is assumed to be in descending order.

Modified Euclidean Algorithm

1) After the first iteration, append zero to the data set.

2) Form the new set with elements $t_j - t_{j+1}$.

3) Sort in descending order.

4) Eliminate elements in $[0, \eta_0]$ from end of the set (η_0 defined in text below).

5) Algorithm is done if left with a single element. Call it t_0, and declare $\widehat{T} = t_0$. Else, go to step 1.

The MEA works by repeatedly exploiting (29) and (30). The subtraction of adjacent pairs in the first iteration eliminates the phase, and results in the simpler data form

$$S' = \{t'_j\}_{j=1}^{N-1}, \text{ with } s'_j = K_j T + \delta_j, \tag{31}$$

where $K_j = k_j - k_{j+1}$ (see remark 1 at the end of this section). After the first iteration, we maintain the minimum in the data after differencing, just as in (30), by appending zero in step 1 before differencing. After step 2 differences that are separated by (approximately) T will yield small (ideally zero) values. These small values are eliminated from the data in step 5 with proper selection of threshold η_0. Choice of η_0 is dictated by the distribution of the noise.

The MEA will yield good results for small number of data samples N. We have shown that, given N ($N \geq 2$) randomly chosen positive integers $\{k_1, \ldots, k_N\}$,

$$P\{\gcd(k_1, \ldots, k_N) = 1\} = [\zeta(N)]^{-1}, \tag{32}$$

where $P\{\cdot\}$ denotes probability and $\zeta(n)$ is the Riemann Zeta function (Theorem 3.1 in [6]). We then get as a corollary

$$P\{\gcd(K_1, \ldots, K_{N-1}) = 1\} = [\zeta(N-1)]^{-1}, \tag{33}$$

(Corollary 3.1 in [6]). We also show that

$$\left[1 - 2^{1-N}\right]^{-1} \leq [\zeta(N)]^{-1} \leq 1, \tag{34}$$

(Proposition 3.1 in [6]). Thus, $[\zeta(N)]^{-1}$ converges to 1 from below very quickly as N increases. For example, $[\zeta(10)]^{-1} = 0.9990$ and $[\zeta(16)]^{-1} = 1.0000$. Combining these results with (29) implies that

$$\gcd(K_1 T, \ldots, K_{N-1} T) \longrightarrow T \tag{35}$$

with high probability for relatively small N, converging to T quickly as $N \longrightarrow \infty$. In the noise-free case this theory tells us that the algorithm will yield T with high probability from only $N = 10$ data samples. In practice the algorithm yields a good estimate of T for small ($N = 10$) to moderate ($N = 100$) sample sizes [6]. We note that the convergence rate of the iterative MEA depends on the spread of the k_j's, in that large spread requires more iterations before the differences approach T.

Other modifications of the MEA were developed and tested in [6]. One modification is a single iteration algorithm that is robust to the presence of outliers. The single iteration version of the MEA is possible when the data set S is such that $k_j - k_{j+1} = 1$ occurs for relatively many values of j. Assuming this is the case then, after step 3 in the first iteration, there will be a cluster of data that is symmetrically distributed about T (because $t_j - t_{j+1} \approx T$ when $k_j - k_{j+1} = 1$). Thus, a single iteration algorithm proceeds by identifying the cluster of data grouped around T (after step 3 above), and then averaging over this cluster to estimate T. The data cluster may be isolated using a gradient operator and data-adaptive thresholding. This approach has the additional advantage of being robust to the presence of arbitrary outliers in S. This robust single iteration algorithm is summarized as follows.

Robust Single Iteration Algorithm

1) Given the set S, form the new set with elements $t_j - t_{j+1}$.

2) Sort the new set in descending order.

3) Apply gradient operator and obtain $\tilde{T}$, with $y_0 \cong 0.6\tilde{T}$ (parameters defined in text below).

4) Obtain $\hat{T}$ by averaging the data over the lowest step, isolating the step based on the lowest two jumps of height y_0 with step-width greater than x_0.

This algorithm requires three parameters: a gradient threshold g_0, the step-height threshold y_0, and the step-width threshold x_0. The gradient is estimated in step 3, with large gradient values indicating a step or "edge" in the data. We have employed a simple gradient estimator by convolving with an impulse response given by $[-1, 0, 1]$. A data-adaptive gradient threshold g_0 is selected as 25% of the maximum gradient value, and data points above this threshold are assumed to correspond to the step edges. The choice of 25% is based on experience gained by extensive numerical results, and may typically be halved or doubled without greatly affecting the results. The rightmost (i.e., smallest in average amplitude) cluster in the data between edges corresponds to the cluster of data around T. We therefore take the rightmost step, whose height we denote $\tilde{T}$, as a coarse estimate of T. We then use $\tilde{T}$ to set the value of y_0 via $y_0 \approx 0.6\ \tilde{T}$. The desired data cluster is then obtained as the smallest (rightmost) cluster between two edges with height y_0 such that the number of data points in the cluster is greater than x_0.

The use of x_0 represents a tradeoff between robustness to noise and outliers versus the possibility that the algorithm will not yield a solution. For higher values of x_0, the single iteration algorithm may not always yield a solution, but will generally yield reliable estimates when it does. For lower values of x_0, in the presence of many outliers, the algorithm will likely always provide a solution but may yield an estimate that is not close to the true T. In our experience, these incorrect estimates are often less than the true T, caused by a falsely identified data cluster made up of outliers. Thus the number of data samples within the selected cluster is a good measure of the reliability of the resulting estimate $\hat{T}$.

Remarks

1. It is easily possible to generate low noise or even noise-free data sets such that the MEA does not yield the true period, even for an arbitrarily large number of data points. For example, consider $k_1 = N, k_2 = N-2, k_3 = N-4, \ldots$ with true underlying period $T = 1$. To eliminate the phase, we do not save the first minimum. For this data, the MEA will

give $T = 2$. However, for even 10 data points, Theorem 3.1 and Corollary 3.2 in [6] say that such data sets are extremely unlikely to occur. Another way of saying this is that if a large set of low noise or noise-free data leads the MEA to an incorrect solution, then the corresponding k_j's exhibit a consistent pattern.

2. The computational load of the robust single iteration algorithm is quite small, being roughly $2N$ additions plus $N \log N$ operations for a fast sort. This assumes convolution with the gradient operator $[-1, 0, 1]$, which can be accomplished with N additions only.

5 Achieving ML Performance with Missing Observations and Outliers

The MEA's described in the previous section provide low complexity methods for estimating T that are robust and perform reasonably well, but fall short of the CRB. To improve on the MEA estimate of T it is desired to estimate the coefficient matrices X or X_d in (4) or (6), respectively. The differenced-data equation (6) is preferred over equation (4) due to the difficulty of estimating the phase ϕ (see the remark below). If T were known then X_d could be estimated using $(1/T)$ $\mathbf{y}$. Ideally, this estimate is composed of positive integers, but imperfect knowledge of T and the presence of noise will generally yield an estimate of X_d that has non-integer components. We therefore propose to estimate X_d via

$$\widehat{X}_d = \text{round}\left[\frac{1}{\widehat{T}_{MEA}}\,\mathbf{y}\right], \tag{36}$$

where $\widehat{T}_{MEA}$ is the estimate of T obtained via the MEA, and $\text{round}[x] = \lfloor x + \frac{1}{2} \rfloor$ denotes rounding to the nearest integer. A refined estimate of T is then obtained by using $\widehat{X}_d$ in (8) yielding

$$\widehat{T} = (\widehat{X}_d^T R_\delta^{-1} \widehat{X}_d)^{-1} \widehat{X}_d^T R_\delta^{-1} \mathbf{y}. \tag{37}$$

This result approaches the optimal minimum variance performance when $\widehat{X}_d$ is close to X_d. The refinement algorithm is summarized as follows.

Refined Estimation Algorithm

1) Estimate T via MEA (see section 4), calling this estimate $\widehat{T}_{MEA}$.

2) Estimate $\widehat{X}_d$ via (36).

3) Refine the estimate of T using $\widehat{X}_d$ in (37), calling this estimate $\widehat{T}$.

Performance analysis of the estimate $\widehat{T}_{MEA}$ depends not only on the distribution of the noise η_j, but also on the distribution of the k_j's. Performance analysis of the iterative MEA is complicated because the analysis involves order statistics and, after the first iteration, the noise is no longer iid [6]. We focus here on the single iteration MEA, and utilize an iid Bernoulli process to model the k_j's.

Suppose $t_j = \phi + k_j T + \eta_j$ is an element of S for some k_j. Then, the probability that $k_{j+1} = k_j - 1$ is, or is not, an element of S is modeled by

$$\begin{aligned} \text{Prob}(k_{j+1} = k_j - 1) &= 1 - \lambda \\ \text{Prob}(k_{j+1} \neq k_j - 1) &= \lambda, \end{aligned} \tag{38}$$

where $0 \leq \lambda < 1$ is the Bernoulli parameter. If $k_j - 1$ is not accepted to the set, then $k_j - 2$ is accepted with probability $1 - \lambda$ or rejected with probability λ, and so on. Assuming that ideally $k_j = N + 1 - j$, $j = 1, 2, \ldots, N$ then (38) is regarded as a missing observations model. The Bernoulli model is commonly employed for missing observations in time series analysis, e.g., see [3]. For example, with $\lambda = 0.25$, we expect 25% of the possible observations to be missing. It follows that, after a single iteration of the MEA, we expect $(1 - \lambda)(N - 1)$ data samples to be symmetrically distributed around the true value of T. Recall that these samples arise when $t_j - t_{j+1} = (k_j - k_{j+1})T + (\eta_j - \eta_{j+1})$ is such that $k_j - k_{j+1} = 1$. The single iteration MEA estimates T by identifying and averaging the data cluster around T. Under the white Gaussian noise assumption we therefore have

$$E[\widehat{T}_{MEA}] = T, \tag{39}$$

with variance

$$\text{var}\{\hat{T}_{MEA}\} = \frac{\sigma_v^2}{(1-\lambda)(N-1)}. \tag{40}$$

Analysis of $\hat{T}$ produced by (37) is difficult due to the nonlinearity introduced when estimating X_d in (36). We therefore quantify the behavior of $\hat{T}$ by comparing Monte Carlo simulation with theoretical CR bounds. As can be expected, the nonlinearity of (36) leads to a threshold effect in the behavior of $\hat{T}$ that will be evident when comparing the sample variance of $\hat{T}$ with the CRB.

Data without outliers

First we consider noisy data without outliers. The data set S was generated with $N = 100$ data samples as in (2), where the k_j's were generated according to the Bernoulli model of (38) with parameter λ. The phase ϕ was selected randomly from $[0, T)$ for each realization. Without loss of generality we set $T = 1$ in all examples. The noise η_j was white Gaussian with variance σ_η^2. In order to express the effect of the noise as a percentage jitter in t_j we use the definition

$$\%\text{ jitter} = \left(\frac{3\sigma_\eta}{T}\right) \times 100 \tag{41}$$

and consider σ_η values such that $0 < \%$ jitter $\leq 50\%$. In all cases 200 Monte Carlo trials were conducted for each value of % jitter. Results are plotted for a normalized mean-square error, given by

$$\text{normalized mean-square error} = 10 \log\left(\frac{T^2}{\text{MSE}}\right), \tag{42}$$

where MSE was obtained experimentally comparing $\hat{T}$ with the known true value T. The MEA used in all cases was the robust single iteration algorithm described in Section 4.

Figure 1 compares three estimates of T with the CRB, with $\lambda = 0.25$ (25% missing observations) and MEA parameter $x_0 = 5$. The CRB was obtained from (19). Note that, using the normalization of (42), the CRB lies above the estimates in all our figures. Performance of $\hat{T}_{MEA}$ is always worse than the CRB, due to a lack of knowledge of the k_j's. Two refined estimates are also shown. Refinement based on (36) and (37) performs well, achieving the

CRB for up to 20% jitter and thereafter exhibiting a threshold effect, ultimately degrading back near the performance of $\widehat{T}_{MEA}$.

For comparison, performance of a refinement algorithm based on (7) is also shown in figure 1. We estimate X using

$$\widehat{X} = \text{ round } \left[\frac{1}{\widehat{T}_{MEA}} (\mathbf{t} - \widehat{\phi}_x) \right], \tag{43}$$

where $\widehat{\phi}_x$ is the phase estimate of (25). The estimate $\widehat{X}$ was then used in (7) to simultaneously refine $\widehat{T}_{MEA}$ and $\widehat{\phi}_x$. This approach yields relatively poor performance due to the poor quality of $\widehat{\phi}_x$. For higher values of λ we have observed little or no improvement over $\widehat{T}_{MEA}$ for even low values of % jitter, so further experiments are not reported for this approach.

Experimental values corresponding to figure 1 are tabulated in table 1. Experimental mean and standard deviation (std) of $\widehat{T}_{MEA}$ and the refined estimate $\widehat{T}$ are shown. The CRB is given in terms of its square root for comparison with the experimental std's. The std of $\widehat{T}$ is quite close to the CRB for up to 20% jitter, while std$\{\widehat{T}_{MEA}\}$ is one to two orders of magnitude worse, showing the level of improvement achieved by refinement of $\widehat{T}_{MEA}$. Note however that, depending on the application, $\widehat{T}_{MEA}$ is a reasonable estimate in and of itself. For example, at 35% jitter std$\{\widehat{T}_{MEA}\}$ is $\approx 1\%$ of $T = 1$. Extensive numerical results for $\widehat{T}_{MEA}$ are given in [6].

Figures 2 and 3 show results for $\lambda = 0.5$ and $\lambda = 0.75$, respectively. All other parameters match those of figure 1. Beyond $\lambda = 0.5$ there is a lowering of the threshold for which $\widehat{T}$ achieves the CRB, as might be expected given the sparse nature of the data. Note the change in CRB with changing λ, such that better performance is possible for higher λ. This is explained by noting that for a fixed number of data samples N, a higher value of λ implies a greater spread of the k_j's (see the discussion following the CRB equations (15) and (16)).

Data with outliers

Further Monte Carlo experiments were conducted with outliers introduced into the data. The data set S was generated as before and then arbitrary outliers were added by selecting from a uniform distribution over the range of (outlier-free) S, i.e., over [min(S), max(S)]. For

example, with $N = 100$ and 5% outliers, S consists of 95 data samples obeying (1), and 5 arbitrary samples.

Naturally, we wish to somehow remove the outliers from the original data. However, with missing observations, noise, and arbitrarily placed outliers, an outlier removal strategy is not obvious. We propose using the MEA, because its robustness allows it to function in this case. So, instead of reusing all the data in step 3 of the refined estimation algorithm, we instead keep only the data that is utilized in forming $\widehat{T}_{MEA}$. Let the subset $S' \subset S$ denote this data, with elements t'_j, $j = 1, \ldots, N' < N$. The reduced set S' can be written in the matrix-vector form of (3) using $y'_j = t'_j - t'_{j+1}$, which gives

$$\mathbf{y}' = X'_d \widehat{T}'_{MEA} + \boldsymbol{\delta}'. \tag{44}$$

Following (36) we form

$$\widehat{X}'_d = \text{round}\left[\frac{1}{\widehat{T}_{MEA}}\,\mathbf{y}'\right], \tag{45}$$

and apply this to (44), whose solution is of the same form as (37). We denote this refined estimate based on S' as $\widehat{T}'$. Note that R_δ^{-1} is the same as before, except for a size change.

Figure 4 compares using S versus S' in the refinement algorithm. Here, $\lambda = 0.25$, $x_0 = 15$, and the data was contaminated with 5% outliers. The reuse of the full data set S does not yield significant improvement over $\widehat{T}_{MEA}$. Note that $\widehat{T}$ is actually worse than $\widehat{T}_{MEA}$ for jitter $< 8\%$. This is because the k_j values corresponding to outliers are not integers, but are forced to be so when forming $\widehat{X}_d$. As the noise is reduced this induced rounding error becomes dominant. Also shown in figure 4 is the normalized MSE of $\widehat{T}'$, whose performance mimics outlier-free results, in that a threshold effect is evident above 15% jitter. Below the threshold $\widehat{T}'$ very nearly achieves the CRB.

That $\widehat{T}'$ does not completely achieve the CRB is due to the inclusion of arbitrary outliers in S', which appears to be generally unavoidable. Let $s_m \in S'$ denote the mth outlier in S, and consider two cases. First, if two outliers are such that $s_m - s_n \approx T$, then they will be accepted by the MEA as good data. Second, if $s_m \approx kT + \phi$ for some integer k, then s_m will be

treated as valid by the MEA. The likelihood of such cases obviously increases with the number of outliers present. The first case will generally have more of a negative impact on $\hat{T}'$ due to contamination of $\widehat{X}'_d$. It is unlikely that outliers in either case can be identified and removed without further information. However, if other information is available, such as knowledge of their distribution, then exploiting this to remove them is likely to be both possible and worthwhile.

An examination of $\hat{T}'$ reveals interesting behavior near the threshold (around 15% jitter in figure 4). Plots of $\hat{T}'$ are shown in figures 5 and 6, with the % jitter fixed for each plot. Figure 5 shows $\hat{T}'$ over 200 Monte Carlo runs for 15% jitter, exhibiting low variance due to the noise level being below threshold. As the % jitter is increased through the threshold region the estimates first show occasionally poor results, as with 20% jitter in figure 5. These statistically unusual events correspond to excessive outlier contamination of S'. For 25% jitter, shown in figure 6, now above threshold, many more values of $\hat{T}'$ are not close to optimal. The situation increases more for 30% jitter. For yet higher % jitter, values of $\hat{T}'$ are statistically alike with variance comparable to $\hat{T}_{MEA}$, as shown in figure 4. The behavior depicted in figures 5 and 6 can be exploited to improve $\hat{T}'$ if multiple realizations are available, as we consider next.

Multiple Realizations

If multiple independent realizations of S are available then significant improvement in $\hat{T}'$ can be obtained in the transition region below the threshold. Let $\hat{T}'_{(i)}$ be the ith estimate based on the ith realization of S. Because of the behavior of $\hat{T}'$ due to outliers described above, we will eliminate values of $\hat{T}'_{(i)}$ that are excessively contaminated by outliers, and then average over the remaining $\hat{T}'_{(i)}$. This is accomplished by rejecting values of $\hat{T}'_{(i)}$ that are significantly distant from a robust measure of std$\{\hat{T}'_{(i)}\}$, where the set $\{\hat{T}'_{(i)}\}$ contains the estimates of T over all realizations.

Suppose there are N_r realizations of S. We form a robust estimate of the standard deviation as (see Ljung [16, p.400])

$$\hat{\sigma}_r = \frac{\overline{\sigma}}{0.7}, \tag{46}$$

where $\overline{\sigma}$ (the median absolute deviation) is the median of $\{|\widehat{T}'_{(i)} - \tilde{T}|\}$ and $\tilde{T}$ is the median of $\{\widehat{T}'_{(i)}\}_{i=1}^{N_r}$. Values of $\widehat{T}'_{(i)} > 3\hat{\sigma}_r$, say, are rejected as outlier contaminated, leaving $N'_r < N_r$ estimates of T. We then average over the remaining N'_r values of $\widehat{T}'_{(i)}$ to yield the final estimate

$$\widehat{T}_M = \frac{1}{N'_r} \sum_{i=1}^{N'_r} \widehat{T}'_{(i)}. \tag{47}$$

Figures 7 and 8 demonstrate the improvement possible with this scheme. Results in figure 7 are based on $\lambda = 0.25$, $x_0 = 15$, and 15% outliers. Shown are normalized MSE of $\widehat{T}_{MEA}$ and $\widehat{T}'$, and the CRB, all based on a single realization. Note that with this level of missing observations and contamination $\widehat{T}'$ does not approach the CRB until the % jitter is less than 5%. The variability in the MSE of $\widehat{T}_{MEA}$ is due to the relatively small number of Monte Carlo trials being inadequate to fully characterize all possible outlier realizations.

Also shown in figure 7 is $\widehat{T}_M$ for $N_r = 5$ realizations, averaged over 200 Monte Carlo runs. This approach yields good estimates of T with a threshold around 20% jitter, due to the judicious elimination of $\widehat{T}'_{(i)}$ when it is strongly outlier contaminated. Typically 2 or 3 of the set $\{\widehat{T}'_{(i)}\}_{i=1}^{N_r=5}$ were rejected. These results, with $\lambda = 0.25$ and 15% outliers, are similar to those of figure 1 ($\lambda = 0.25$, no outliers) and figure 4 ($\lambda = 0.25$, 5% outliers). Note that $\widehat{T}_M$, based on multiple records, can exceed the single record CRB shown in figure 7. Of course, the price paid in using $\widehat{T}_M$ to combat outliers is the necessity for multiple data records. This is to be contrasted with simply averaging over all N_r realizations without eliminating any of the $\widehat{T}'_{(i)}$, which exactly corresponds to the performance curve for $\widehat{T}'$ in figure 7, obtained by averaging over 200 realizations. Superior performance is obtained with $\widehat{T}_M$ using only five (versus 200) realizations.

A final Monte Carlo experiment is shown in figure 8, repeating the experiment used to generate figure 7 but now using $\lambda = 0.50$, $x_0 = 10$, and 15% outliers. Here, with a very high level

of missing observations and outliers, $\widehat{T}'$ has failed to achieve the CRB for even small percentages of jitter. However, $\widehat{T}_M$ (again with $N_r = 5$ realizations) shows good performance for up to 20% jitter. In this difficult case, when a single realization is insufficient to gain improvement over $\widehat{T}_{MEA}$ with the refinement algorithm, multiple realizations can be successfully exploited to obtain good results. It can also be noted that the MEA continues to yield a decent estimate, based on a single realization. Numerical experiments for the MEA with values up to $\lambda = 0.80$ and 10% outliers are described in [6].

Remark

1. Estimation of ϕ and T is a linear regression problem. For example, if $k_j = N + 1 - j$, $j = 1, 2, \ldots, N$ then the problem is analogous to line fitting with ϕ playing the role of intercept and T the role of slope. The observed noisy data are much more sensitive to changes in the slope than in the intercept, making estimation of the intercept considerably more difficult. The CRB of the ML estimate of ϕ is $\mathcal{O}(N^{-1})$, while the CRB for ML estimation of T is $\mathcal{O}(N^{-3})$, e.g., see Kay [13, p.43]. The estimate $\widehat{\phi}_x$ is not ML and has high variance for even moderate noise levels. Thus, if it is desired to estimate T only, it may be advisable to work with $y_j = t_j - t_{j+1}$, eliminating the need to estimate ϕ.

6 The Common Oscillator Problem

Many radar and communications systems rely on a very stable oscillator to provide an accurate time reference. Pulse rates will then generally be multiples of the common oscillator rate, obtained via countdown circuitry, e.g., see Wiley [25, sect. 8.6.2]. Examples include radar PRI switching and pseudo-randomly jittered hop times in frequency hopping communications. In such cases it may be desired to estimate the underlying period of the common oscillator. This can be accomplished using the MEA approach provided the observations are of the form (1).

One such case occurs when successive pulse intervals are multiples of the common oscillator period. Now, (1) is applicable with $k_j - k_{j+1} = \xi T$, where ξ is a positive integer. For example, ξ might be generated pseudorandomly. Thus the MEA will yield an estimate of T and the refinement algorithm may be used to enhance the MEA estimate.

7 Conclusions

The MEA provides an efficient tool for finding the period T from the data set S arising from (1). It is especially useful with many missing observations and contaminated data. With missing observations it is not straightforward to apply the least-squares minimum variance solution, and the addition of outliers makes the problem that much worse. Without prior knowledge of their distribution an outlier removal strategy is not obvious. Because the MEA is able to work robustly in this situation it may also be exploited to provide a subset $S' \subset S$ that may be used to enhance the MEA period estimate. Although not in general outlier-free, S' is outlier reduced.

Nonlinear estimates of the k_j's are used together with the subset S' in the least-squares solution, yielding an estimate of T that is close to or attains the CRB for moderate noise levels and is in general an improvement over the MEA estimate. The enhanced estimate of T exhibits a threshold effect due to the nonlinearity in estimating the k_j's, with estimates reaching the CRB below the threshold. In the case of contaminated data it is possible to extend the threshold to higher noise levels if multiple independent data records are available. This cannot be accomplished by simply averaging over records, but instead relies on deleting those estimates of T that are highly contaminated. This is achieved by robustly estimating the standard deviation of the estimates over records, and eliminating those that are of significantly high statistical variation.

Further improvement is likely to be achieved if sort parameters other than simply the time of arrival can be incorporated, which is a topic for further research. When dealing with more

complex problems such as radar PRI estimation in the presence of random period changes, it is likely that parameters such as direction of arrival, pulse duration, pulse amplitude, etc., will be necessary, e.g., [17,18]. Also interesting might be the combination of the MEA with deinterleaving problems. For example, a period may be postulated by time differencing arrival times, and then selected data points within a window analyzed. In this sense the MEA provides another tool to use in combination with existing methods.

8 Appendix

Assume a noise pdf of the form

$$f(\eta) = \frac{e^{\alpha \cos(2\pi\eta/T)}}{2\pi I_0(\alpha)}, |\eta| \leq \frac{T}{2}. \tag{48}$$

Here, $f(\eta) = f(-\eta)$ for all α implies zero-mean. Assuming that $\boldsymbol{\beta} = [\phi, T]^T$ is known, the pdf of t_j given $\boldsymbol{\beta}$ is

$$f(t_j|\boldsymbol{\beta}) = f(t_j - k_j T - \phi), \tag{49}$$

and for iid noise,

$$f(\mathbf{t}|\boldsymbol{\beta}) = \prod_{j=1}^{N} f(t_j|\boldsymbol{\beta}). \tag{50}$$

Now the log-likelihood function is

$$L(\boldsymbol{\beta}) = \ln f(\mathbf{t}|\boldsymbol{\beta}) = \sum_{j=1}^{N} [\alpha \cos \frac{2\pi}{T}(t_j - \phi) - \ln 2\pi I_0(\alpha)]. \tag{51}$$

So, the ML estimate is found by maximizing $L(\boldsymbol{\beta})$ with respect to T, yielding

$$\widehat{T}_{ML} = \text{argmax}_T \sum_{j=1}^{N} \cos\left(2\pi \frac{(t_j - \phi)}{T}\right). \tag{52}$$

Using the fact that

$$\cos \frac{2\pi}{T}(t_j - \phi) = \text{Re}\left\{e^{2\pi i \frac{(t_j - \phi)}{T}}\right\}, \tag{53}$$

then, for ϕ constant, maximizing $L(\boldsymbol{\beta})$ is equivalent to

$$\widehat{T}_x = \text{argmax}_T \left| \sum_{j=1}^{N} e^{2\pi i \frac{t_j}{T}} \right|^2, \tag{54}$$

which is in the form of spectrum analysis of a point process with occurrence times t_j.

References

[1] Arnett, D. W., and Ellert, B. M., "A real-time cross correlator for neurophysiological research," *IEEE Transactions on Biomedical Engineering,* Vol. BME-23, No. 1, pp. 65–70, 1976.

[2] Bartlett, M. S., "The spectral analysis of point processes," *Journal of the Royal Statistical Society B*, Vol. 25, No. 2, pp. 264–280, 1970.

[3] Bloomfield, P., "Spectral analysis with randomly missing observations," *Journal of the Royal Statistical Society B,* Vol. 32, No. 3, pp. 369–380, 1970.

[4] Brillinger, D. R., "The spectral analysis of stationary interval functions," *Proc. of the 6th Berkeley Symposium on Mathematical Statistics and Probability,* pp. 483–513, 1972.

[5] Brillinger, D. R., "Maximum likelihood analysis of spike trains of interacting nerve cells," *Biological Cybernetics,* Vol. 59, No. 3, pp. 189–200, 1988.

[6] Casey, S. D., and Sadler, B. M., "Modifications of the Euclidean algorithm for isolating periodicities from a sparse set of noisy measurements," *IEEE Signal Processing*, to appear 1996; also "A modified Euclidean algorithm for isolating periodicities from a sparse set of noisy measurements," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95)*, Vol. 3, pp. 1764–1767, 1995.

[7] Elton, S. D., and Gray, D. A., "The application of circular statistics to specific radar pulse train detection," *Proceedings EUSIPCO-94.*

[8] Fogel, E., and Gavish, M., "Parameter estimation of quasi-periodic sequences," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '88)*, Vol. 4, pp. 2348–2351, 1988.

[9] Fogel, E., and Gavish, M., "Performance evaluation of zero-crossing-based bit synchronizers," *IEEE Transactions on Communications*, Vol. 37, No. 6, pp. 663–665, 1989.

[10] Gray, D., Slocumb, B., and Elton, S., "Parameter estimation for periodic discrete event processes," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, Vol. 4, pp. 93–96, 1994.

[11] Hardy, G. H., and Wright, E. M., *An Introduction to the Theory of Numbers*, Oxford University Press, New York, 1982.

[12] Kay, S. M., "Statistically/computationally efficient frequency estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '88)*, pp. 2292–2295, 1988.

[13] Kay, S. M., *Fundamentals of Statistical Signal Processing,* Prentice-Hall, New Jersey, 1993.

[14] Knuth, D. E., *The Art of Computer Programming, Volume 2: Seminumerical Algorithms (Second Edition)*, Addison-Wesley, Reading, Massachusetts, 1981.

[15] Leveque, W. J., *Topics in Number Theory, Volumes 1 and 2*, Addison-Wesley, Reading, Massachusetts, 1956.

[16] Ljung, L., *System Identification: Theory for the User* (Prentice-Hall, 1987).

[17] Mardia, H. K., "New techniques for the deinterleaving of repetitive sequences," *IEE Proceedings-F,* Vol. 136, No. 4, pp. 149–154, 1989.

[18] Milojevic, D. J., and Popovic, B. M., "Improved algorithm for the deinterleaving of radar pulses," *IEE Proceedings-F,* Vol. 139, No. 1, pp. 98–104, 1992.

[19] Sadler, B. M., and Casey, S. D., "PRI analysis from sparse data via a modified Euclidean algorithm," *Proceedings Twenty-Ninth Annual Asilomar Conference on Signals, Systems, and Computers (invited),* Monterey, CA, October 1995.

[20] Sadler, B. M., and Casey, S. D., "Frequency estimation via sparse zero crossings," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96)* May 1996.

[21] Staelin, D. H., "Fast folding algorithm for detection of periodic pulse trains," *IEEE Proceedings,* Vol. 57, pp. 724–725, 1969.

[22] Tam, D. C., Ebner, T. J., and Knox, C. K., "Cross-interval histogram and cross-interspike interval histogram correlation analysis of simultaneously recorded multiple spike train data," *Journal of Neural Science Methods,* Vol. 23, pp. 23–33, 1988.

[23] Tretter, S. A., "Estimating the frequency of a noisy sinusoid by linear regression," *IEEE Transactions on Information Theory,* Vol. IT-31, pp. 832–835, 1985.

[24] Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part II,* Wiley, New York, 1971.

[25] Wiley, R. G., *Electronic Intelligence: The Analysis of Radar Signals (Second Edition),* Artech House, Norwood, Massachusetts, 1993.

Table 1: Comparison of data from figure 1.

% jitter	$\hat{T}_{MEA}$ mean(std)	$\hat{T}$ mean(std)	$\sqrt{\text{CRB}}$
2	1.0000 (0.0007)	1.0000 (1.826×10^{-5})	1.829×10^{-5}
4	1.0000 (0.0011)	1.0000 (3.587×10^{-5})	3.669×10^{-5}
6	1.0003 (0.0016)	1.0000 (5.056×10^{-5})	5.532×10^{-5}
8	1.0000 (0.0022)	1.0000 (7.427×10^{-5})	7.366×10^{-5}
10	1.0002 (0.0028)	1.0000 (9.275×10^{-5})	9.228×10^{-5}
15	0.9998 (0.0043)	1.0000 (1.438×10^{-4})	1.387×10^{-4}
20	1.0000 (0.0054)	1.0000 (1.907×10^{-4})	1.836×10^{-4}
25	0.9999 (0.0072)	1.0000 (2.210×10^{-4})	2.308×10^{-4}
30	1.0007 (0.0084)	1.0000 (1.583×10^{-3})	2.758×10^{-4}
35	0.9999 (0.0110)	0.9994 (3.401×10^{-3})	3.222×10^{-4}
40	1.0008 (0.0139)	0.9996 (6.216×10^{-3})	3.672×10^{-4}
45	1.0052 (0.0426)	0.9999 (3.285×10^{-2})	4.178×10^{-4}
50	1.0133 (0.0741)	1.0041 (5.830×10^{-2})	4.792×10^{-4}

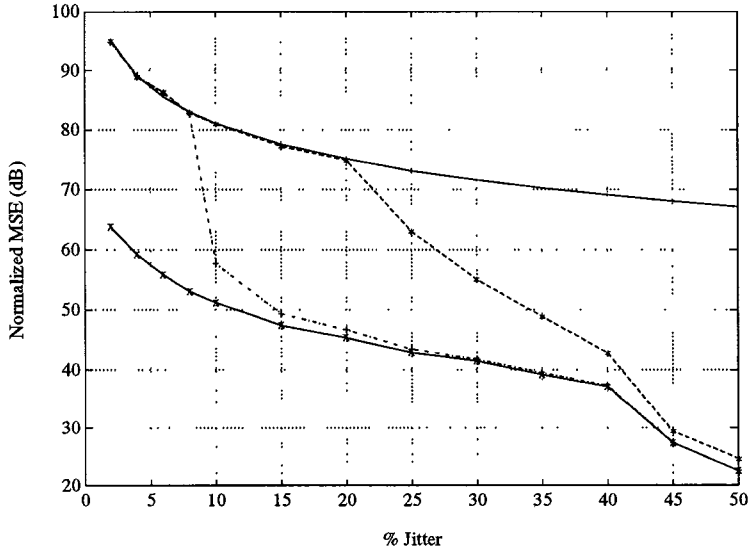

Figure 1: Monte Carlo estimation results (without outliers, $\lambda = 0.25, N = 100, x_0 = 5$). Solid = CRB, x = $\hat{T}_{MEA}$, dash (*) = $\hat{T}$, and dash-dot (+) = $\hat{T}$ incorporating $\hat{\phi}_x$.

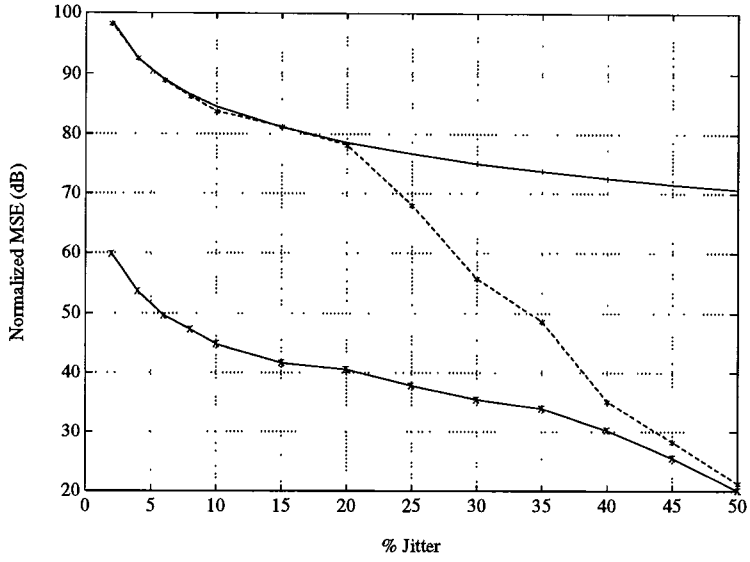

Figure 2: Monte Carlo estimation results (without outliers, $\lambda = 0.50, N = 100, x_0 = 5$). Solid = CRB, x = $\widehat{T}_{MEA}$, and dash (*) = $\widehat{T}$.

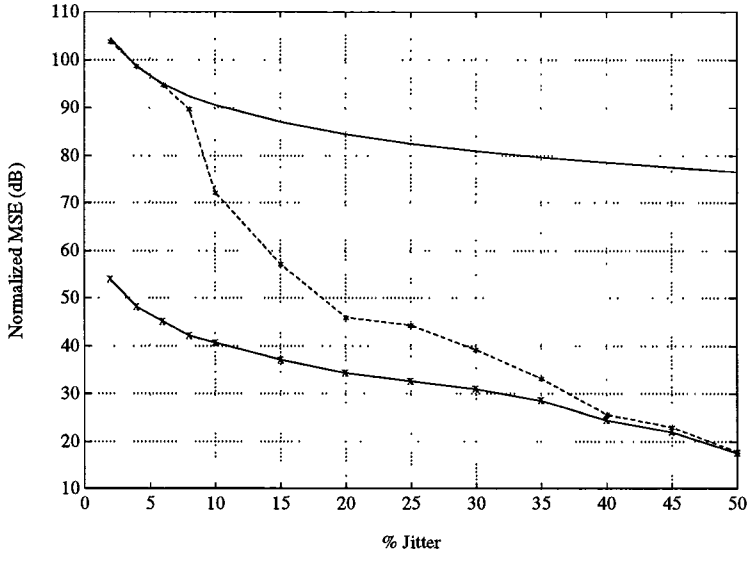

Figure 3: Monte Carlo estimation results (without outliers, $\lambda = 0.75, N = 100, x_0 = 5$). Solid = CRB, x = $\widehat{T}_{MEA}$, and dash (*) = $\widehat{T}$.

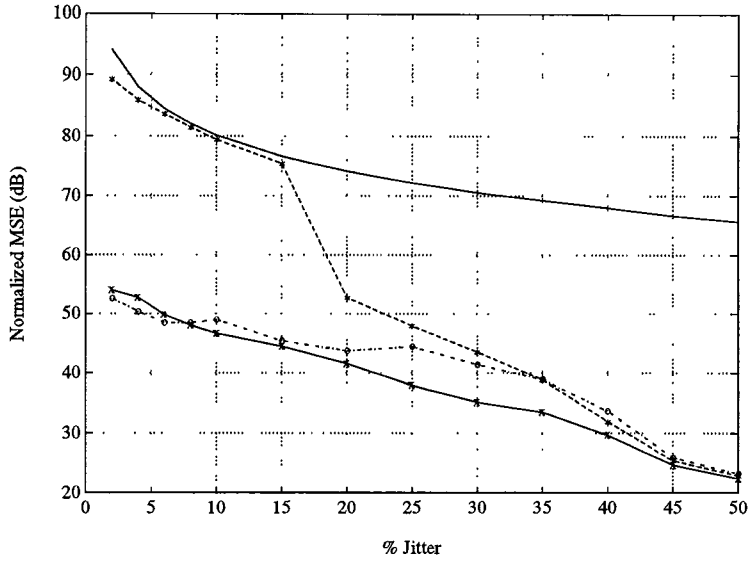

Figure 4: Monte Carlo estimation results (with 5% outliers, $\lambda = 0.25, N = 100, x_0 = 15$). Comparison of $\widehat{T}$ versus $\widehat{T}'$ (full versus selected data reuse in the refinement algorithm). Solid = CRB, x = $\widehat{T}_{MEA}$, dash-dot (o) = $\widehat{T}$, dash (*) = $\widehat{T}'$.

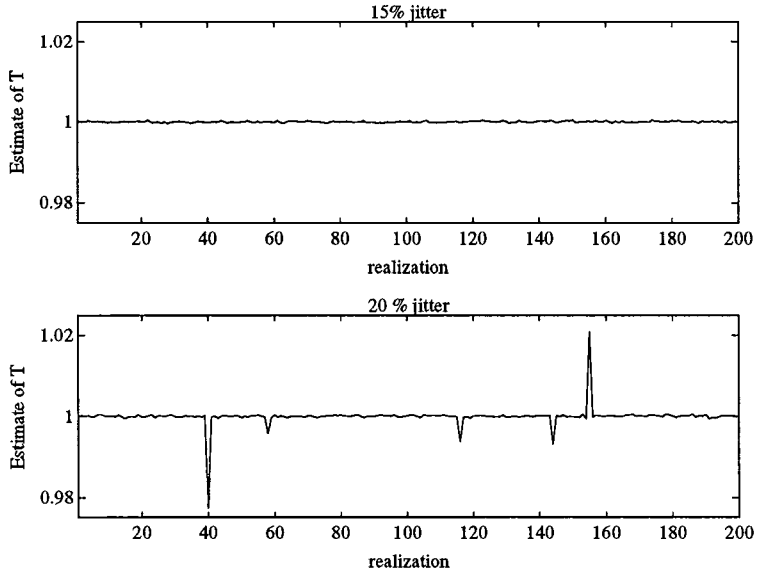

Figure 5: Realizations of $\widehat{T}'$ for different percentages of jitter near threshold, with experimental parameters matching those of figure 4.

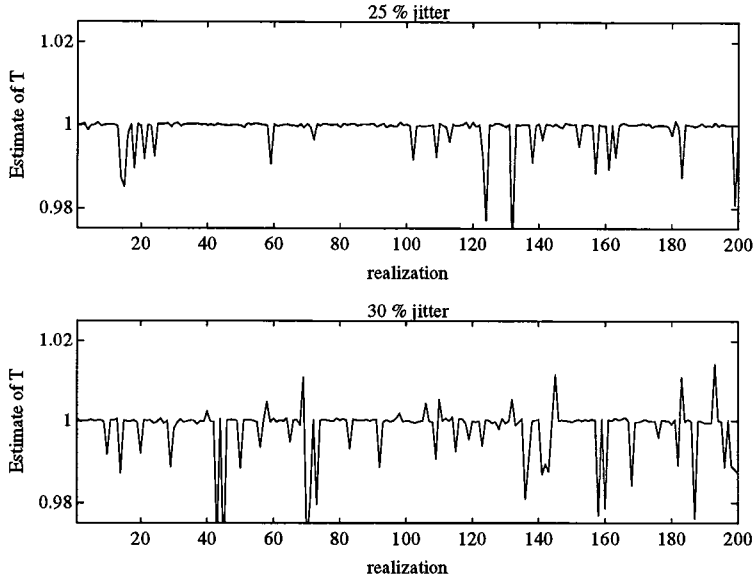

Figure 6: Realizations of $\widehat{T}'$ for different percentages of jitter above threshold, with experimental parameters matching those of figure 4.

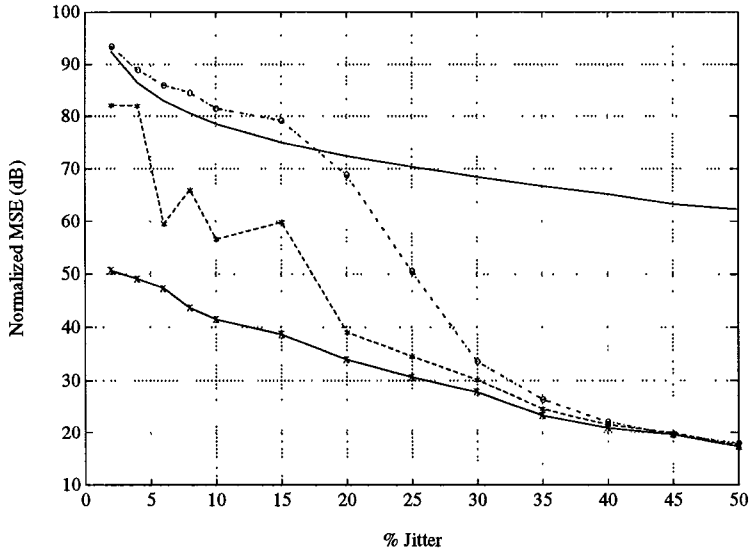

Figure 7: Multiple realization results. Solid = CRB, x = $\widehat{T}_{MEA}$, and dash (*) = $\widehat{T}'$, for single data records ($\lambda = 0.25$, $N = 100$, $x_0 = 10$, and 15% outliers). Also, dash-dot (o) = $\widehat{T}_M$ ($N_r = 5$ data records).

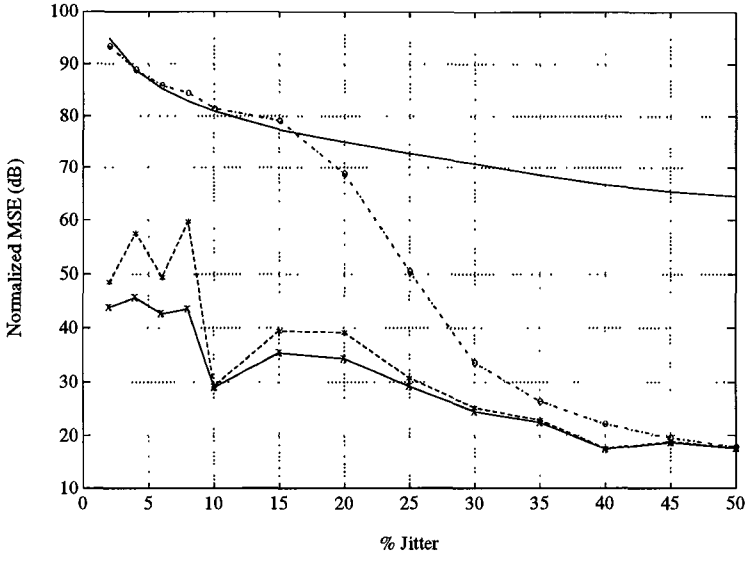

Figure 8: Multiple realization results. Solid = CRB, x = $\widehat{T}_{MEA}$, and dash (*) = $\widehat{T}'$, for single data records ($\lambda = 0.5$, $N = 100$, $x_0 = 10$, and 15% outliers). Also, dash-dot (o) = $\widehat{T}_M$ ($N_r = 5$ data records).