THE ASTRONOMICAL JOURNAL VOLUME 93, NUMBER 4 APRIL 1987

TIME SERIES ANALYSIS WITH CLEAN. I. DERIVATION OF A SPECTRUM

DAVID H. ROBERTS
Physics Department, Brandeis University, Waltham, Massachusetts 02254

JOSEPH LEHÁR AND JOHN W. DREHER
Physics Department, 26-315, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Received 7 October 1986; revised 25 November 1986

ABSTRACT

We present a method of time-series spectral analysis which is especially useful for unequally spaced data. Based on a complex, one-dimensional version of the CLEAN deconvolution algorithm widely used in two-dimensional image reconstruction, this technique provides a simple way to understand and remove the artifacts introduced by missing data. We describe the method, give several examples, and point out various analogies with the conventional use of CLEAN.

I. INTRODUCTION

In this paper, we present a new approach to the problem of estimating the complex frequency spectrum of a continuous function of time which can be measured only at a finite number of discrete times. Our method is particularly useful for data samples unequally spaced in time. However, the technique we describe is also useful in the case of equally spaced data. We will deal only with deterministic signals.

Fourier inversion of a finite representation of a continuous function leads to a frequency spectrum distorted by (i) the limited frequency resolution due to the finite time span of the data sample and (ii) spurious apparent responses that are caused by the incompleteness of the sampling. By adapting the two-dimensional CLEAN algorithm developed for use in aperture synthesis (Högbom 1974), we show that there is a very intuitive way of handling the difficulties associated with the latter effect. The CLEAN algorithm performs a nonlinear deconvolution in the frequency domain (equivalent to a least-squares interpolation in the time domain). The algorithm is especially suitable for functions whose spectra are dominated by a small number of components at discrete frequencies. A good example of the application of CLEAN to such a spectrum can be seen in Fig. 6, where Fig. 6(c) shows the results of a direct inversion of a discrete sample of a function consisting of two components (at 31 and 57 Hz), and Fig. 6(d) shows the spectrum recovered by CLEAN.

In the next section we set out the problem. Section III describes the application of CLEAN to spectral analysis. Several examples are presented and discussed in Sec. IV, and a brief comparison with other techniques is made in Sec. V. Appendix A is a brief summary of aperture synthesis and the use of CLEAN, and appendices B and C contain some calculational details. The reliability of the technique (how well it recovers the amplitude and phase of a signal in the presence of noise and/or multiple signals) will be discussed later (Lehár, Roberts, and Dreher 1987; Paper II). CLEAN was used in the search for periodic variability in the BL Lacertae object OJ 287 by Dreher, Roberts, and Lehár (1986).

II. SPECTRAL ANALYSIS

We begin by considering a function $f(t)$ known for all time t. It is well known that f may be constructed by the superposition of contributions from all frequencies ν, both positive and negative,

$$f(t) = \int_{-\infty}^{+\infty} d\nu\, F(\nu) e^{+2\pi i\nu t},\ -\infty \leqslant t \leqslant +\infty;$$

here, as elsewhere, we follow the notation of Jenkins and Watts (1968). The *spectrum* of f, which defines the contribution of each ν to f, is given by the Fourier transform (FT) of f,

$$F(\nu) \equiv FT[f] \equiv \int_{-\infty}^{+\infty} dt\, f(t) e^{-2\pi i\nu t},\ -\infty \leqslant \nu \leqslant +\infty.$$

We assume that $f(t)$ is real, so $F(\nu)$ satisfies

$$F(-\nu) = F^*(\nu), \tag{1}$$

where the * represents complex conjugation. Rayleigh's theorem, which relates the variances of f and F,

$$\int_{-\infty}^{+\infty} dt\, |f(t)|^2 = \int_{-\infty}^{+\infty} d\nu |F(\nu)|^2, \tag{2}$$

leads one to define the *power spectrum* of f as

$$P(\nu) = |F(\nu)|^2,\ -\infty \leqslant \nu \leqslant +\infty.$$

$P(\nu)$ is the contribution of ν to the variance of f; since P is a square it is generally denoted a "power." Since $f(t)$ is real, half of the power at a given $|\nu|$ occurs at a *negative* frequency.

a) Real Data—Sampling a Continuous Function

In most experimental or observational settings, all that is available are the values of f at a finite number of sample points $\{t_r\}$, i.e., a finite number N of data points,

$$\{f_r, t_r\} \equiv \{f(t_r), t_r\},\ r = 1,2,...,N. \tag{3}$$

These N data points may be regarded as having been sampled from f by multiplication with a sampling function $s(t)$ consisting of a weighted sum of N Dirac delta functions $\delta(x)$. A convenient definition is

$$s(t) = C \frac{\sum_{r=1}^{N} w_r \delta(t - t_r)}{\sum_{r=1}^{N} w_r}, \tag{4}$$

where the rth data point is assigned weight w_r, and the normalization and dimensions of s are determined by the constant C. In this paper, we will use equal weights $w_r = 1$ and take $C = 1$; thus the *sampled signal* is

 0004-6256/87/040968-22$00.90

$$f_s(t) \equiv f(t)s(t) = \frac{1}{N}\sum_{r=1}^{N} f(t)\delta(t-t_r)$$

$$= \frac{1}{N}\sum_{r=1}^{N} f_r\delta(t-t_r). \quad (5)$$

It follows from the Fourier convolution theorem that the Fourier transform of the sampled signal $D(\nu) \equiv FT[f_s]$ is the convolution of the spectrum with the Fourier transform of the sampling function $W(\nu) \equiv FT[s]$:

$$D(\nu) = F(\nu) \otimes W(\nu) \equiv \int_{-\infty}^{+\infty} d\nu' F(\nu') W(\nu - \nu'). \quad (6)$$

We will call the observable functions $D(\nu)$ and $W(\nu)$ the *dirty spectrum* and *spectral window function*, respectively. Their calculation from the data is straightforward, as the integrals collapse to discrete Fourier transforms:

$$D(\nu) = \int_{-\infty}^{+\infty} dt f_s(t) e^{-2\pi i\nu t} = \frac{1}{N}\sum_{r=1}^{N} f_r e^{-2\pi i\nu t_r}, \quad (7)$$

and

$$W(\nu) = \int_{-\infty}^{+\infty} dt\, s(t) e^{-2\pi i\nu t} = \frac{1}{N}\sum_{r=1}^{N} e^{-2\pi i\nu t_r}. \quad (8)$$

The symmetries of $D(\nu)$ and $W(\nu)$ are the same as that of $F(\nu)$:

$$D(-\nu) = D^*(\nu), \quad (9)$$

$$W(-\nu) = W^*(\nu). \quad (10)$$

As a result of our choice of normalization, the zero-frequency (DC) component of D is the average value of the data,

$$D(0) = \frac{1}{N}\sum_{r=1}^{N} f_r = \langle f_r \rangle, \quad (11)$$

and the window function is unity at the origin,

$$W(0) = 1. \quad (12)$$

Note that solution of Eq. (6) directly using the convolution theorem,

$$F(\nu) = FT[f] = FT[f_s/s] = FT[f_s] \otimes FT[1/s], \quad (13)$$

is not possible. Precisely because we know $f(t)$ only at discrete times, $s(t)$ is zero almost everywhere, and Eq. (13) is nonsense. In fact, the solution for F is an example of an inverse problem known not to have a unique solution (Bracewell and Roberts 1954); no finite set of data $\{f_r\}$ is enough to specify fully the underlying function $f(t)$.

b) The Consequences of Sampling

The limitation to a finite number of sample points has two important consequences. The first is that the total length of the data span, $T = t_N - t_1$, is finite. A finite data span can be considered to be generated by a *blanking function* $h(t)$,

$$h = \begin{cases} 1 & \text{if } t_1 \leqslant t \leqslant t_N, \\ 0 & \text{otherwise.} \end{cases}$$

Continuous data blanked by this function may be represented by

$$f_h(t) = f(t)h(t),$$

and the corresponding dirty spectrum is the convolution of the spectrum and the *smearing function* H,

$$FT[f_h] = F(\nu) \otimes H(\nu),$$

where

$$H(\nu) \equiv FT[h] = \frac{\sin(\pi\nu T)}{\pi\nu} e^{-\pi i\nu(t_1 + t_N)}$$

is the window function for a continuous data set of finite length. The convolution of F with H results in the smearing (also called "spectral leakage") of a given feature in F over the "width" of H (the range of frequencies, around $\nu = 0$, where H is appreciable). This smearing occurs in D since the sampled data are by assumption limited to $t_1 \leqslant t \leqslant t_N$. The *frequency resolution* $\delta\nu$ which results from a discrete data set is the width of the main peak of W (at $\nu = 0$); as long as the data sampling is not too nonuniform, the width of W will be similar to the width of H, so

$$\delta\nu \simeq \frac{1}{T}.$$

The second consequence of a finite data set occurs because the data are not continuous over (t_1, t_N), but occur only at the $\{t_r\}$. The window function will not be H, but rather W, and there will be features in W which are due to the detailed distribution of sample points. Convolution of F with W to produce D will result in "false" features in D (features due to the structure of W), which will confuse the identification of the "true" features in D (those due to the structure of F).

Our goal is to *estimate F from our knowledge of D and W,* in other words, *to 'undo' the damage inflicted by our incomplete knowledge of f,* as far as possible, given the finite frequency resolution imposed by a limited time span.

1) Equally Spaced Data

If the sampling occurs at $N = 2n$ points separated by equal intervals Δ, so that one knows

$$\{f_r, t_r\} = \{f_r, r\Delta\}, \quad r = -n, \ldots, n-1, \quad (14)$$

the theory of Fourier series shows that one may reproduce $\{f_r\}$ exactly from a finite Fourier series of N terms,

$$\tilde{f}(t) = \sum_{s=-n}^{n-1} F_s e^{+2\pi i s t/N\Delta}, \quad -\infty \leqslant t \leqslant +\infty. \quad (15)$$

The Fourier coefficients required for $\tilde{f}(t_r) = f_r$ are

$$F_s = \frac{1}{N}\sum_{r=-n}^{n-1} f_r e^{-2\pi i s r/N}, \quad s = -n, \ldots, n-1, \quad (16)$$

and the N frequencies required are

$$\nu_s = \frac{s}{N\Delta}, \quad s = -n \ldots, n-1. \quad (17)$$

For discrete, equally spaced data and frequencies, Rayleigh's theorem (Eq. (2)) becomes Parseval's theorem,

$$\frac{1}{N}\sum_{r=-n}^{n-1} |f_r|^2 = \sum_{s=-n}^{n-1} |F_s|^2. \quad (18)$$

The contribution $|F_s|^2$ to the power at frequency ν_s is called the *intensity* at this frequency, and the plot of intensity versus frequency index s is the *Fourier line spectrum*,

$$P_s = |F_s|^2, \quad s = -n, \ldots, n-1. \quad (19)$$

In the case of equally spaced data, the relationship of $\{f_r\}$ to f is given by the sampling theorem (Shannon 1949; Bracewell 1965): *Any function whose Fourier transform is zero for $|\nu| \geqslant \nu_N$ is fully specified by values spaced at equal intervals not exceeding $1/(2\nu_N)$.* Thus the highest frequency which may be recovered from samples at intervals Δ is one-half the sampling rate; this is the Nyquist frequency

$$\nu_N = \frac{1}{2\Delta}.$$

Thus it is possible to reconstruct, *over the continuous domain of the time variable,* any function whose spectrum is zero for $|\nu| \geqslant \nu_N = 1/2\Delta$.* Since the Fourier transform of equally spaced data can be evaluated very economically with the FFT or other similar algorithms, and since equally spaced sampling is often natural and/or convenient, estimation of the spectrum from the Fourier line spectrum is a very common practice. However, the Fourier line spectrum is *not* the spectrum of f, not even at the frequency points ν_s. Rather, the Fourier line spectrum represents the spectrum of $\tilde{f}$, which is

$$FT[\tilde{f}] = \sum_{s=-n}^{n-1} F_s \delta(\nu_s - \nu), \quad -\infty \leqslant \nu \leqslant +\infty. \tag{20}$$

Unless f (like $\tilde{f}$) is a periodic function with period equal to the span of the data samples ($N\Delta$), and unless f is actually bandwidth limited, Eq. (20) is not its spectrum.

An important consequence of equally spaced data is that the coefficient F_s contains not only the contribution from ν_s, but also from $\nu_s \pm l2\nu_N$, where l is an integer. This "aliasing" of the spectrum occurs because the F_s are periodic functions of frequency with period $1/\Delta = 2\nu_N$ and are conjugate symmetric about $\pm l\nu_N$ (see the nice discussion by Brault and White 1971). Aliasing will occur unless the sampling interval can be made sufficiently small that there is no signal in F beyond the Nyquist frequency. In the absence of *a priori* information about the spectrum of f it can be difficult to know if this requirement has been met. (Note that it is the blanked function which must satisfy the sampling theorem; various procedures can be used to ensure that the endpoints of the data interval do not contribute power beyond the Nyquist frequency (e.g., Jenkins and Watts 1968). The problem of aliasing can be eliminated by use of randomly spaced sampling (Paper II).

Finally, even equally spaced sampling can lead to Fourier line spectra which are hard to interpret due to the confusion of features due to F with those arising from the features in W. The technique we describe below, which is designed to remove these latter features, is therefore useful for equally spaced as well as unequally spaced data.

2) Unequally Spaced Data

In astronomy, the rotation and revolution of the Earth impose diurnal and annual cycles on (Earth-based) observations, and often introduce large, roughly equally spaced gaps in the data sequence. Even though this will result in a very complex spectral window W, the temptation is just to plug ahead with the calculation of D from Eq. (7), and to hope that this gives a reasonable approximation to the spectrum F. As we have seen, *this is a very dangerous and misleading procedure because the convolution in Eq. (6) can introduce spurious features into D.* This difficulty is well known, and procedures to help recognize spurious features in D have been widely used (Deeming 1975 a,b). However, it would be more useful to be able to "clean up" the dirty spectrum by removing the spurious features. In the next section we describe an approximate, nonlinear method by which this may be accomplished.

*It is worth noting that, strictly speaking, the sampling theorem *never* applies to an experimental or observational setting, since the theorem (implicitly) requires that data samples be available over all times.

III. APPLICATION OF CLEAN TO TIME-SERIES ANALYSIS

a) Synthesis Maps and CLEAN

Synthesis maps in radio astronomy are produced by Fourier transformation of interferometer visibility data to produce a map of the sky. Because real interferometers gather only a finite number of visibility points, the problems of sampling described in Sec. II*a* apply to interferometer maps. The CLEAN algorithm (Högbom 1974) was devised to perform an approximate deconvolution of the true map of the sky from the dirty map produced from the data—in essence, removing the false features due to the finite sampling. The CLEAN algorithm works by subtracting from the dirty map the response expected for a point source located at the map maximum, producing a residual map from which both the point source and the spurious features due to sampling have been removed. This procedure is repeated on successive residual maps until nothing is left but noise. The set of point sources so obtained (the "clean components") is used as a model for the source. To preserve the angular resolution of the observation, this model is convolved with a "clean beam" which has the same resolution as the original "dirty beam" but no sidelobes. Finally, to preserve the noise level on the map, and (more importantly) to include any flux which was not recovered by CLEAN (either because the cleaning is not sufficiently "deep" or because there is extended emission not easily represented by point components), the final residual map is added to the convolved clean components to produce the "clean map." A fuller discussion of synthesis mapping and the application of CLEAN to interferometer maps is given in Appendix A.

Schwarz (1978) has shown that CLEAN is a statistically correct method of (least-squares) fitting sinusoidal terms to the observations (visibility data). The CLEAN procedure in the map domain is equivalent to an interpolation and extrapolation in the visibility domain. The convolution of the point-source model with the clean beam serves to weight down the extrapolated visibility in regions outside those sampled.

b) Spectral Clean Components

To apply CLEAN to the deconvolution of Eq. (6) we begin by considering a single *harmonic component,* by which we mean a single cosinusoid of *harmonic amplitude A,* frequency $\hat{\nu}$, and phase constant ϕ,

$$f(t) = A \cos(2\pi\hat{\nu}t + \phi). \tag{21}$$

The spectrum of f is

$$F(\nu) = a\delta(\nu - \hat{\nu}) + a^*\delta(\nu + \hat{\nu}), \tag{22}$$

where

$$a = \frac{A}{2} e^{+i\phi}.$$

The spectrum is complex, and, as expected for a real signal, contains a pair of *spectral components* at $\pm\hat{\nu}$ with equal *spectral amplitudes* of $A/2$. The phases of the positive and negative spectral components are $\pm\phi$, respectively. The sampled data lead to a dirty spectrum consisting of two terms,

$$D(\nu) = aW(\nu - \hat{\nu}) + a^*W(\nu + \hat{\nu}). \tag{23}$$

As an example, Figs. 1(a)–1(c) show the time series, spectral window, and dirty spectrum of a cosinusoidal signal with $A = 1.0$, $\hat{\nu} = 31.25$ Hz (period 0.032 s), and phase

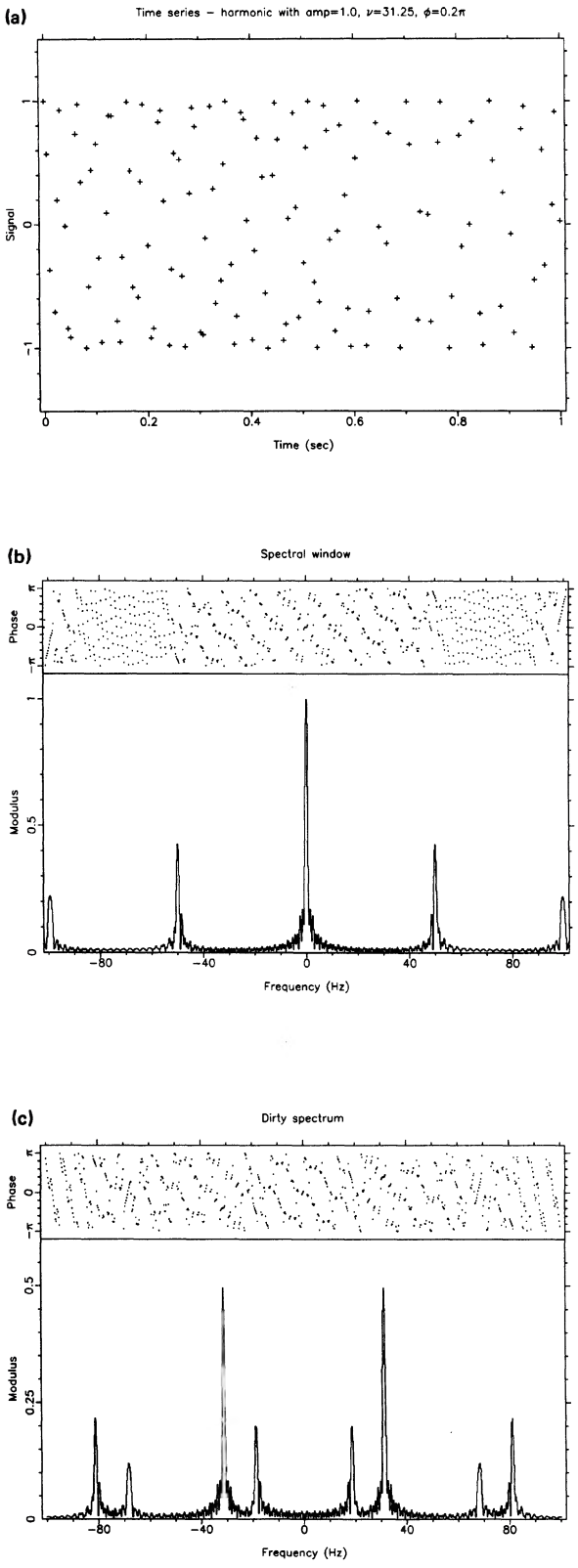

FIG. 1. Spectral analysis of a single cosinusoidal component of 31.25 Hz sampled at one hundred discrete times over 1 s. No noise has been added to the signal; (a) the time series, (b) the window function, and (c) the dirty spectrum.

(with respect to the mean sample time*) $\phi' = +\pi/2$, sampled at ~ 100 points over 1 s. (We used uniformly sampled data with a few points removed in order to produce a set of sidelobes suitable for the purposes of demonstration). *No noise was added to the data* (the role of noise will be discussed in Paper II). The spectral window consists of a central peak at $\nu = 0$, plus secondary peaks, called the *sidelobes*, which are displaced from the central peak. Note that while the central peak is always the largest feature in the spectral window, the largest sidelobes can be a substantial fraction of the central peak. The dirty spectrum contains both the true positive and negative frequency peaks (at ± 31.25 Hz) and false peaks due to the sidelobes of these peaks. The strongest false peaks (displaced from the true peaks by $\pm \sim 50$ Hz and $\pm \sim 100$ Hz) are of substantial amplitude, and confuse the interpretation of the dirty spectrum. Note especially that two prominent sidelobes of the negative frequency peak, as well as one sidelobe of the positive frequency peak, occur at positive frequencies.

By analogy to the application of CLEAN to aperture-synthesis maps, we would like to subtract functions of the form of the spectral window from the dirty spectrum. In the map case, a single point source produces a single maximum on the dirty map. In the dirty spectrum, however, a single harmonic component at frequency $\hat{\nu}$ produces two maxima, at $\pm\hat{\nu}$. Because each peak is contaminated by the sidelobes of the other, the amplitude and frequency of the component cannot be determined merely by using the amplitude and frequency of the maxima on the dirty spectrum. However, we may take the contamination into account by noting that in the case of a spectrum containing a single harmonic component, the dirty spectrum takes the value

$$D(\hat{\nu}) = aW(0) + a^*W(2\hat{\nu})$$

at the frequency of that component. Using Eqs. (9), (10), and (12), we can find the spectral amplitude of that component in terms of its frequency,

$$a = \frac{D(\hat{\nu}) - D^*(\hat{\nu})W(2\hat{\nu})}{1 - |W(2\hat{\nu})|^2}. \tag{24}$$

Because of the contamination of the positive and negative frequency peaks by the sidelobes of the other, it is not in general possible to determine precisely the frequency of even a single spectral component from a dirty spectrum. However, the peaks $\tilde{\nu}$ of D usually lie well within $\delta\nu$ of the corresponding spectral components. Given the frequency of a clean component estimated from D, its (complex) amplitude is found from a function defined analogous to Eq. (24),

$$\alpha(D;\nu) = \frac{D(\nu) - D^*(\nu)W(2\nu)}{1 - |W(2\nu)|^2}, \tag{25}$$

as $a \simeq \alpha(D;\tilde{\nu})$. If CLEAN is done judiciously (see below), the small errors in the frequency, amplitude, and phase of the clean components selected in this way will be corrected in subsequent iterations. As a practical matter, this procedure works in most cases we have tried. If multiple components are present, it is possible that D will take its maximum value at a point away from any component. In this case, CLEAN will try to subtract a component that does not exist, and it

*Since it is convenient to subtract the average of the sampled times from each time (see Appendix B), the natural variable is $t' = t - \langle t_r \rangle$, and here (and in all other examples) the phase constant ϕ' is the phase of the cosine at the mean sample time ($\langle t_r \rangle$), so $f(t) = A\cos(2\pi\hat{\nu}t' + \phi')$. Thus ϕ' is related to the phase constant ϕ in Eq. (21) by $\phi = \phi' - 2\pi\hat{\nu}\langle t_r \rangle$.

may fail. (An analogous problem can occur in map CLEANing if the sum of two sidelobes happens to be the map maximum.)

c) The CLEAN Algorithm for Spectral Analysis

To find all of the spectral components in a dirty spectrum D, remove the sidelobes due to the structure of W, and construct a "clean" spectrum, we iterate the procedure for finding the spectral component for a single cosinusoid. Beginning with D, find the amplitude and frequency of the first spectral component from Eq. (25). Then remove the contribution of this component, including its sidelobes, by subtracting response of the form of Eq. (23) from D, forming the first *residual spectrum* R^1. In order to prevent small errors from destabilizing the CLEAN procedure, it has been found helpful to subtract from D only a fraction g of the response due to this component at one time—removal of (nearly) all of the response will occur as the process is iterated. For convergence, the *clean gain* g must satisfy $0<g<2$; typical values for g lie between 0.1 and 1. Values at the bottom of this range require more iterations, but should provide more stability—experience in the synthesis-mapping case suggests that low values may be advisable in the analysis of spectra containing significant power over a band of frequencies considerably larger than the frequency resolution.

Next, repeat this procedure on $R^1, R^2, \ldots$. After completion of the ith step of this iterative process we have a set of i *spectral clean components* and a *residual spectrum* R^i, which is the result of removing these clean components and their sidelobes from the dirty spectrum. The iteration must be halted by some *stopping condition*, which might include one or more of the following: The maximum of R^i is less than some predetermined "noise" value, the number of iterations has reached some reasonable limit, or the sum of the moduli of the clean components is not increasing significantly. If computing time is not a constraint, we recommend continuing to iterate until (the gain) $\times$ (the number of clean components) is comparable to the number of frequencies for which D has been calculated.

After all the spectral clean components have been determined, they are convolved with the *clean beam* $B(\nu)$. This makes the frequency resolution of the clean spectrum $\sim 1/T$, rather than the (artificial) resolution of the clean components (the frequency spacing of the discrete representation used in calculation). B has a Gaussian shape in amplitude, $B(0)=1$, and a linear slope in phase, with the parameters chosen by numerical fitting to match the central peak of W (B can always be chosen to be real—see Appendix B). Finally, to preserve the noise level and to include any spectral features not well represented by the spectral clean components, the residual spectrum is added in to form the *clean spectrum* $S(\nu)$.

With the notation $R^0 \equiv D$, we can summarize:

(1) Starting with $i=1$, find the frequency ν^i of the ith clean component from the maximum of $|R^{i-1}(\nu)|$. Using clean gain g, the amplitude of the component is $c^i = g\alpha(R^{i-1};\nu^i)$.

(2) Form the ith residual spectrum,

$$R^i(\nu) = R^{i-1}(\nu) - (c^i W(\nu-\nu^i) + (c^i)^* W(\nu+\nu^i)).$$

(3) Examine R^i and the accumulated clean components. If the stopping condition is not met, continue iterating at step 1. Otherwise, exit at step 4.

(4) Fit the clean beam B to the window function W.

(5) Construct the clean spectrum $S(\nu)$ by convolving the K clean components with $B(\nu)$ and adding the final residual spectrum,

$$S(\nu) = \sum_{i=1}^{K} (c^i B(\nu-\nu^i) + (c^i)^* B(\nu+\nu^i)) + R^K(\nu).$$

A realization of these procedures appropriate for computation is given in appendices B and C.

Application of CLEAN to the dirty spectrum considered in the preceding section (Fig. 1) is illustrated in Figs. 2(a)–2(e), where we show the clean spectra which result from one, two, three, five, and one hundred iterations of CLEAN (at a gain of $g=0.5$), the clean components generated in these iterations, and the residual spectra after these iterations. After n iterations at a gain of $g=0.5$, the peaks due to the sidelobes are reduced to a fraction $\sim(0.5)^n$ of their original height; after one hundred iterations all of the false peaks visible in Fig. 1(c) are removed by this procedure, leaving only the true positive and negative frequency peaks. The spectral amplitude of each of the peaks is 0.5 and their phases are $+\pi/2$ and $-\pi/2$. Of course, in this case of a single cosinusoid without noise, a single iteration at gain $g=1$ also removes the sidelobes completely, leading to the spectra in Fig. 2(f). (The low level noise seen in (d)–(f) is numerical in origin.)

d) Some Requirements for CLEAN

Accurate determination of the clean components and the removal of their sidelobes requires that the window function W satisfy two simple criteria: (i) In order to make it possible to remove a component anywhere in the dirty spectrum ($-\nu_{\max} \leqslant \nu \leqslant \nu_{\max}$), W must be known over $-2\nu_{\max} \leqslant \nu \leqslant 2\nu_{\max}$, and (ii) in order to locate accurately each clean component, the central peak of the beam must be well defined. This requires that in the discrete realization of W and D there be several frequency points spanning $\delta\nu$ (see Appendix B).

e) The Frequency Range

The maximum and minimum frequencies for which spectral information may be extracted from a given data set are well determined in the case of equally spaced data, but for arbitrary $\{t_r\}$, the sampling theorem tells us nothing. If the data samples are otherwise equally spaced but with missing points, the theorem says that the data completely determine a function whose FT is zero for $|\nu| > 1/(2\Delta_{\max})$, where $\Delta_{\max}$ is the *largest completely sampled* data spacing. However, there are smaller spacings, and these certainly carry information about frequencies greater than $1/(2\Delta_{\max})$; some information is available about frequencies as high as $1/(2\Delta_{\min})$, where $\Delta_{\min}$ is the smallest spacing between data points. Furthermore, if the $\{t_r\}$ are more or less randomly distributed, so that a wide range of spacings are present and there is little redundancy in the spacing between various points, tests have shown (Paper II) that significant information is available on frequencies greater than $1/(2\Delta_{\min})$. Nonetheless, in the present paper we will restrict ourselves to frequencies obeying

$$|\nu| \leqslant \nu_{\max} = 1/(2\Delta_{\min}).$$

The minimum frequency is roughly that for which one cycle takes place during the data span,

$$\nu_{\min} \simeq 1/T.$$

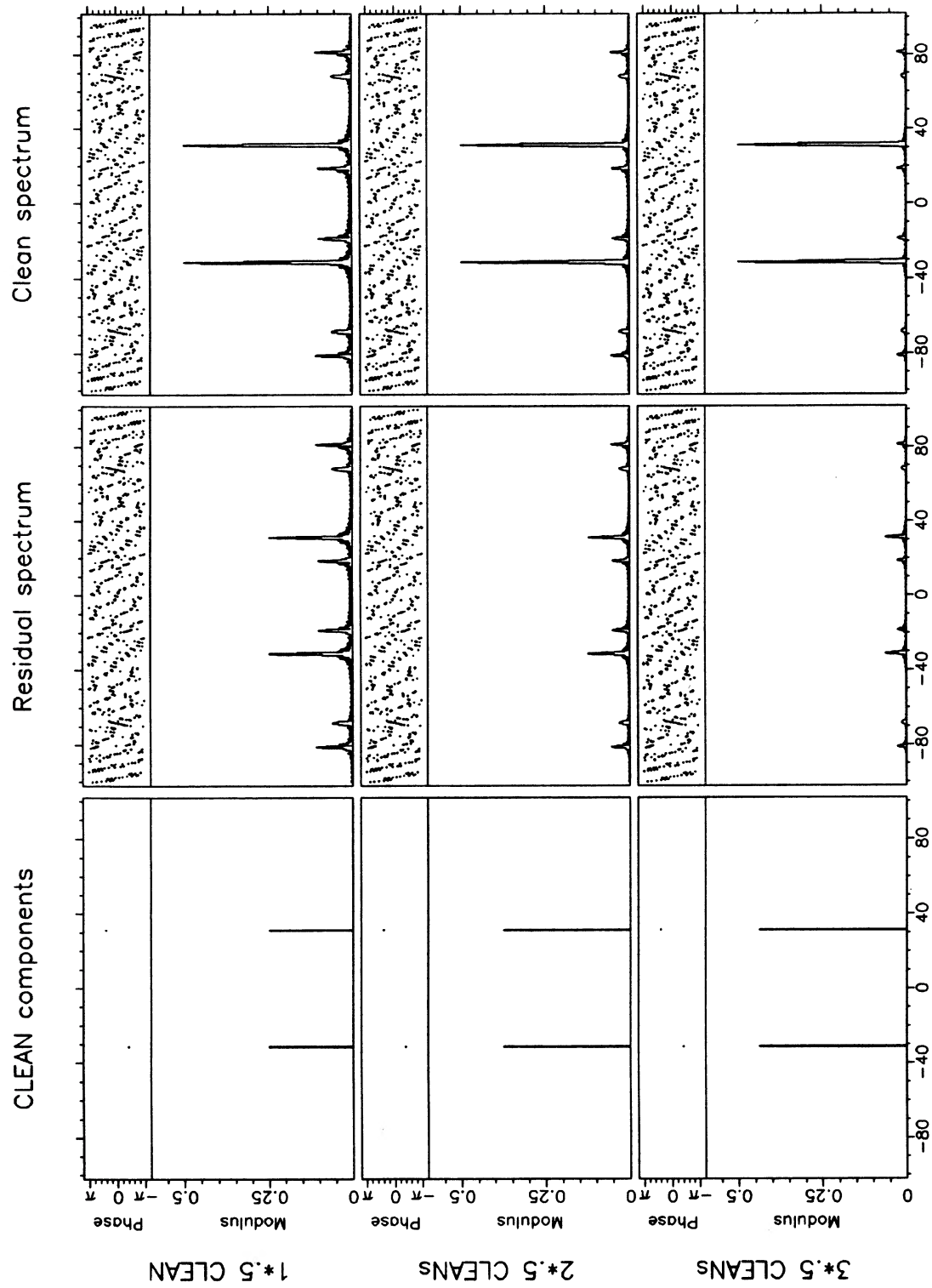

FIG. 2. Analysis of the time series in Fig. 1. (a)–(e). The clean components, residual spectra, and clean spectra after one, two, three, five and one hundred iterations with gain $g = 0.5$, and (f) after one iteration with $g = 1$. Note the change to a logarithmic scale for (d)–(f).

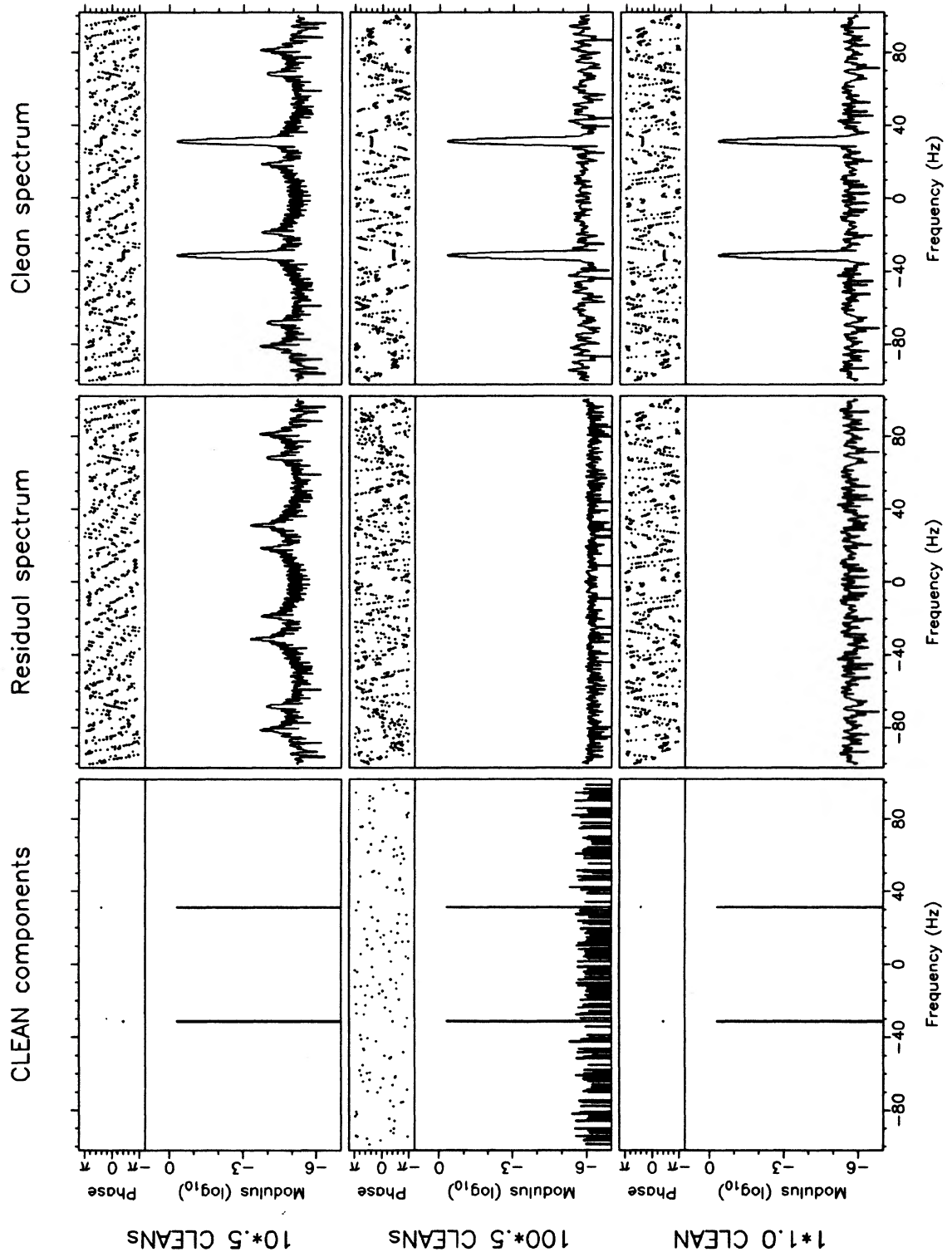

FIG. 2. (continued)

In the case of no gaps, this is related to the maximum frequency by

$$\nu_{\min} = 1/N\Delta = (2/N)\nu_{\max},$$

the same relationship as in Eq. (17).

f) Interpretation of D and S

It is worthwhile here to reconsider the interpretation of D and S and their relationship to F. Both D and S are obtained from convolutions over F, and can be thought of as averages of F over their respective response windows (W and B). It is readily seen that the dimensions of D and S are the same as those of the input function f, whereas the dimensions of F are those of the input divided by frequency. *Thus even in the limit of continuous sampling neither D nor S approaches F.* We have chosen a normalization for W and B (both of which are dimensionless) such that they are unity at $\nu = 0$. As a consequence the areas of these "beams" are not normalized, and in the limit of infinite sampling they do not approach Dirac delta functions. This normalization is, however, the most suitable for our main purpose, finding harmonic components: If the signal contains an isolated harmonic component of amplitude A, then features with modulus $A/2$ will be found at the appropriate frequencies in both D and S. The spectrum F may then be constructed according to Eq. (22). The same normalization is used in the map synthesis case, where a point source of flux density I_0 located at (x_0,y_0), whose brightness distribution is $I_0\delta(x-x_0)\delta(y-y_0)$, is represented on a clean map by a smooth peak at (x_0,y_0) with height I_0 and dimensions equal to those of the clean beam. Further discussion of the normalization of the sampling function can be found in Deeming (1975a,b).

The situation is more complex if the input signal has a harmonic content which extends over a band of frequencies. To find the contribution to the power between frequencies ν_1 and ν_2, the functions $|D|^2$ or $|S|^2$ must be integrated over the appropriate positive and negative frequency intervals and then divided by $\Delta\nu$, the effective area of the beam. For the CLEAN spectrum S, this area is the integral of B, which is $\sim\delta\nu$. For the dirty spectrum, D, however, this procedure is ambiguous, since the integral of W will typically be either zero or infinity. The same problem arises in the case of synthesis mapping, where the usual procedure is to use an effective (two-dimensional) beam area $\Delta\Omega$ defined by integrating the response only over the first completely positive "main lobe." This procedure will be roughly correct only if, as is usually the case, the amplitude of the main lobe of the dirty beam is much larger than that over the rest of the beam and if the area over which the map is integrated is not too large compared to the effective $\Delta\Omega$. For the spectral-analysis case, the same procedure could be followed, but would be rendered even less reliable because the window function W usually has secondary responses that are much larger with respect to the maximum response than encountered in the synthesis case. Note that in both cases, integration over the CLEAN function is the preferable technique. An alternative, if the residual spectrum is negligible, is to directly sum the clean components, in which case no normalization correction will be necessary.

IV. EXAMPLES AND TESTS OF THE ALGORITHM

a) Equally Spaced Data

To illustrate the use of CLEAN we start with a function $f(t)$ consisting of a constant A_0 plus two cosinusoids, $f(t) = A_0 + \Sigma_i A_i \cos(2\pi\hat{\nu}_i t + \phi_i')$, where

$$A_0 = 1.0,$$
$$A_1 = 1.0, \quad \hat{\nu}_1 = 31.0, \quad \phi_1' = +\pi/2,$$
$$A_2 = 0.4, \quad \hat{\nu}_2 = 57.0, \quad \phi_2' = -\pi/2.$$

No noise was added to the data. Figure 3 shows 201 equally spaced samples of f taken over 1 s ($\nu_N = 100$ Hz), along with the dirty spectrum and window function which result. Although there are sidelobes present, this dirty spectrum is easily interpreted. One hundred iterations of CLEAN at gain $g = 0.5$ yield the spectra and clean components shown; as expected, the sidelobes are completely gone.

In Fig. 4 we have (randomly) dropped half of the time samples from the previous example. The spectral window which results has larger sidelobes, and the dirty spectrum is more complicated. Although the two true peaks are present, the amplitude of the smaller one is somewhat low, and there are a few false peaks which might be taken to be real. One hundred iterations of CLEAN at $g = 0.5$ do a fine job of removing the sidelobes and recovering the correct amplitude and phase of both true peaks.

If we randomly drop 80% of the original time samples of Fig. 3, we get the time series in Fig. 5. The spectral window is now quite ragged, and the dirty spectrum has ceased to be interpretable—the weaker peak is completely lost in sidelobes of the stronger peak. CLEAN has little trouble with this case, recovering the correct amplitude and phase of the two peaks, but a small amount of "scruff" contaminates the clean spectrum. CLEAN is not infallible, and although there was no noise in the input time series, there are false features in the clean spectrum. The reason for this, of course, is the very sparse time sampling.

b) Randomly Spaced Data

If the $f(t)$ of Sec. IV*a* is sampled at one hundred random times (generated from uniformly distributed random numbers) over 1 s, we get the data in Fig. 6 (compare Fig. 4, where the number of data points is identical). Once again, no noise was added to the data. The spectral window is rather messy, and the smaller peak has disappeared completely from the dirty spectrum. However, CLEAN has no trouble with this case.

If we drop the number of samples down to 40 we get the data in Fig. 7 (compare Fig. 5). Once again, the smaller peak is completely lost in the dirty spectrum, but CLEAN is able to "find" it. Note that this clean spectrum is somewhat better than the one produced by the same number of equally spaced points.

Under some circumstances, samples made at random times over a given interval can be more advantageous than the same number of samples made at equally spaced times over the same interval (Paper II). In addition to superior rejection of aliasing, a wider range of frequencies can be sampled, and the window function resulting from unequal sampling will usually have smaller secondary maxima than will the window function for regular sampling. Thus it may be beneficial to design experiments to yield unequally spaced data. (The analogy in aperture synthesis is that a random distribution of interferometer baselines, generated by an array of randomly placed antennas, may be quite superior to the distribution of baselines generated by a regular array. The latter can produce strong "grating" response, whereas in the random array the secondary responses tend to cancel.)

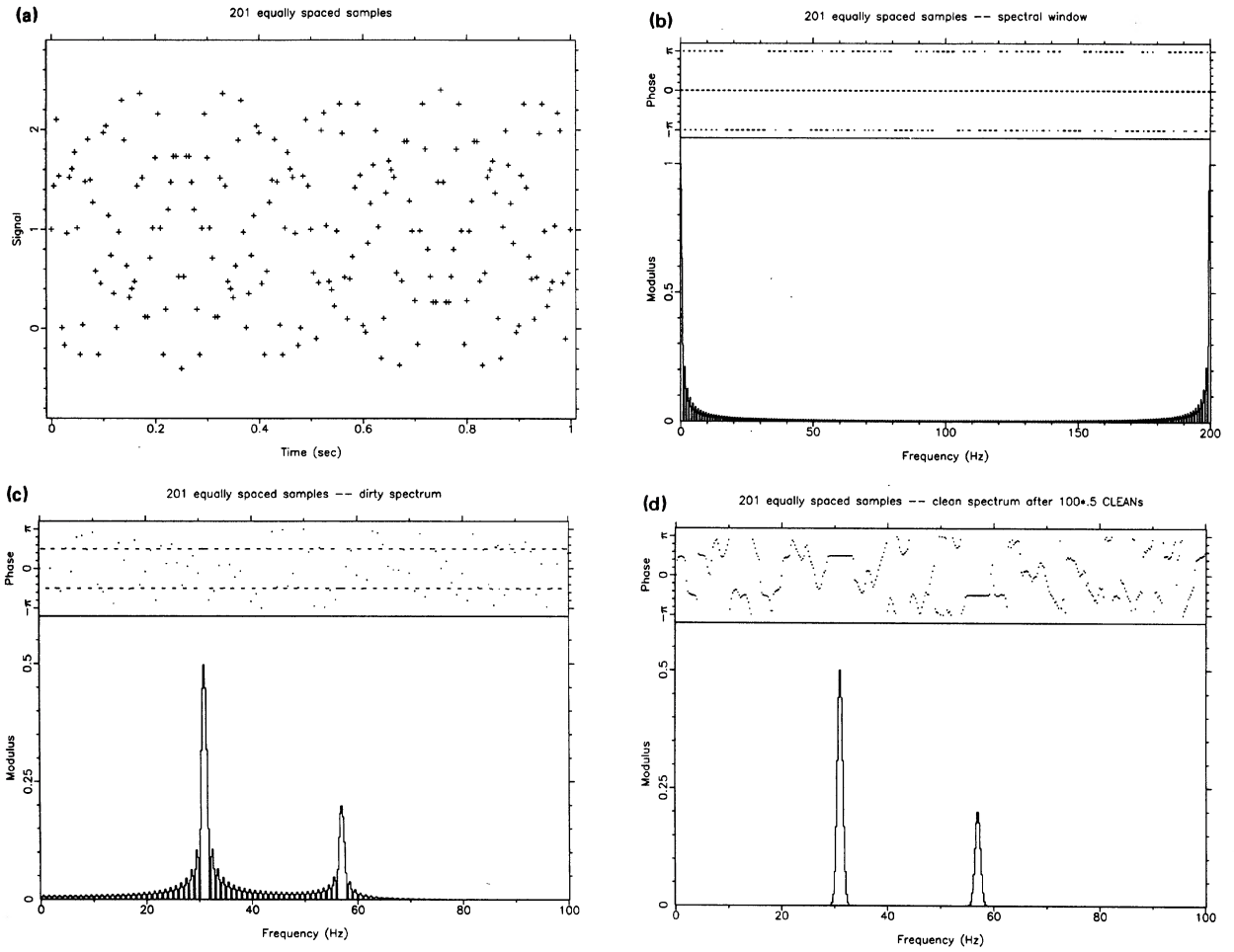

FIG. 3. Spectral analysis of 201 equally spaced noiseless data containing harmonic components at 31 and 57 Hz; (a) time series, (b) spectral window, (c) dirty spectrum, and (d) clean spectrum after one hundred iterations of CLEAN with $g = 0.5$.

c) Data with Large Gaps

The case of data with large gaps is quite typical in astronomy. In Fig. 8 we have imposed a 50% duty cycle with period 0.0715 s (14.0 Hz) on the data of Fig. 3. The resulting window function has large and periodic sidelobes due to the sharp, regular modulation of the sampling. The dirty spectrum is full of false features, making its interpretation very difficult. Once again, however, CLEAN works perfectly, and the two true peaks are easily recovered.

If we reduce the duty cycle of the regular sampling to 20% with the same period, we produce the data in Fig. 9. Here the window function is horrendous and the dirty spectrum hopeless, but CLEAN is still able to reconstruct the peaks correctly.

d) Inverting the Spectrum—How CLEAN Interpolates

The difference between dirty and clean spectra is that the dirty spectrum implies that where data are missing they are zero, while the clean spectrum interpolates across missing data points. This may be seen by inverting the spectra, i.e., Fourier transforming the spectra back into the time domain. To illustrate this we consider both equally spaced and unequally spaced data upon which a duty cycle of 50% with a period 0.23 s (4.35 Hz) has been imposed. In Fig. 10(b) we show a data set with regularly spaced data during the "on" periods. The function sampled consisted of four terms, a constant plus three cosinusoids of various amplitudes, frequencies, and phases, where

$$A_0 = 1.0,$$
$$A_1 = 0.2, \quad \hat{\nu}_1 = 1.0, \quad \phi_1' = -0.8\pi,$$
$$A_2 = 0.5, \quad \hat{\nu}_2 = 10.0, \quad \phi_2' = +0.6\pi,$$
$$A_3 = 0.1, \quad \hat{\nu}_3 = 34.0, \quad \phi_3' = 0.0.$$

The lowest frequency term has a period equal to the data span, and is thus at the low-frequency limit of the spectrum. The window function and dirty spectrum of this time series are shown in Figs. 10(a) and 10(c), while Figs. 10(e), 10(g), 10(i), and 10 (k) show the clean spectra produced after one, two, three, and one hundred iterations of CLEAN with $g = 1$. Note how the removal of successive clean components enables us to recognize weaker components which were previously confused in sidelobes.

Figure 11 shows a data set generated from the same four terms and data duty cycles as Fig. 10, but with the same number of time points distributed at random over the "on"

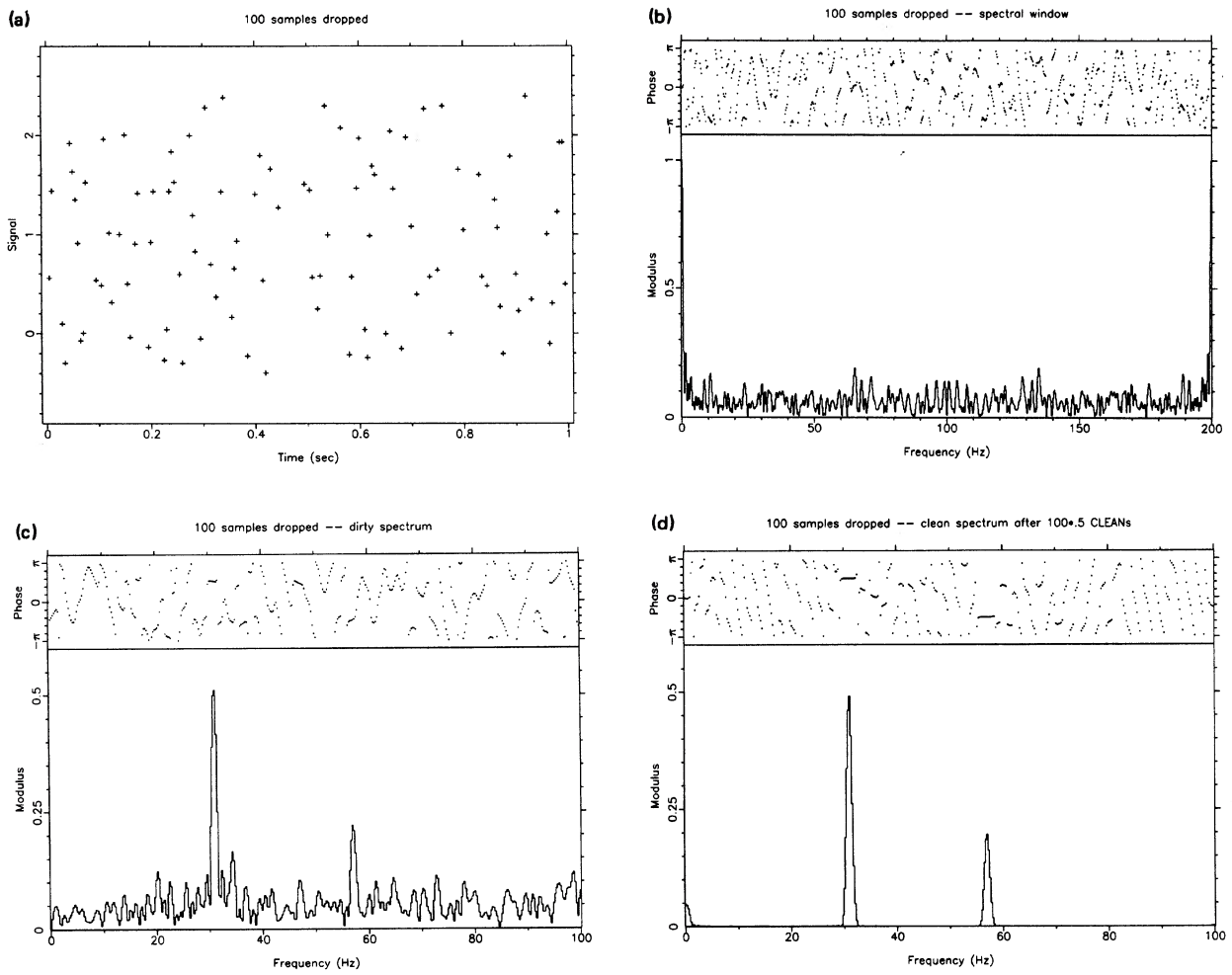

FIG. 4. Two-hundred-one equally spaced noiseless data from Fig. 3, less one hundred random drops. Same parts as Fig. 3.

periods (once again the time points were generated from uniformly distributed random numbers). CLEAN works just as well here as for the equal time spacing of Fig. 10.

Figures 10 and 11 also illustrate the inversion of these spectra. Inversion of the dirty spectra is shown in Figs. 10(d) and 11(d); this is accomplished with the generalization of Eq. (15) for unequal numbers of time and frequency points,

$$f(t) = \frac{N}{M} \sum_{s=-m}^{m-1} F_s e^{+2\pi i \nu_s t} = \frac{N}{M} \sum_{s=-m}^{m-1} D(\nu_s) e^{+2\pi i \nu_s t}, \tag{26}$$

where $N = 2n$ and $M = 2m$ are the numbers of data points and frequency samples, respectively. For equally spaced data, Eq. (26) reproduces the data exactly—the inverted dirty spectrum passes through each data point (Fig. 10(d)). In addition, *inversion of the dirty spectrum reproduces the sampling*—the inverted dirty spectrum also matches the sampling window. (That this will happen for equally spaced data with gaps may be seen by considering the Fourier series representation of the data, where the gaps consist of data points which are zero.) For the unequally spaced data of Fig. 11, the inverted dirty spectrum is a good (but not exact) fit to the data points and to the sampling window because although a large range of frequencies was included in evaluation of Eq. (26), only a Fourier transform (containing all frequencies) can reproduce unequally spaced data points exactly.

Figures 10 and 11 also demonstrate the inversion of the successive stages of the CLEANing of these spectra. The Fourier transform of a clean component is a single cosinusoidal term (Eq. (21)), so *inversion of the set of clean components is just summing up cosines of known frequency, amplitude, and phase.* Figure 10 shows the inversion of the DC and the first one, two, three, and one hundred clean components for the equally spaced data with gaps; Fig. 11 shows the result for the one hundred components for the unequally spaced data with gaps. From Figs. 10 and 11 it is clear that CLEAN *fits the data but not the sampling window*, for both the equally spaced and randomly spaced time points (Note that, as S is the convolution of the components with a Gaussian, inversion of S itself will be the transform of the clean components weighted by the corresponding Gaussian function of time.)

Figure 10 also illustrates the error correction built into the CLEAN algorithm. While there are only three components to

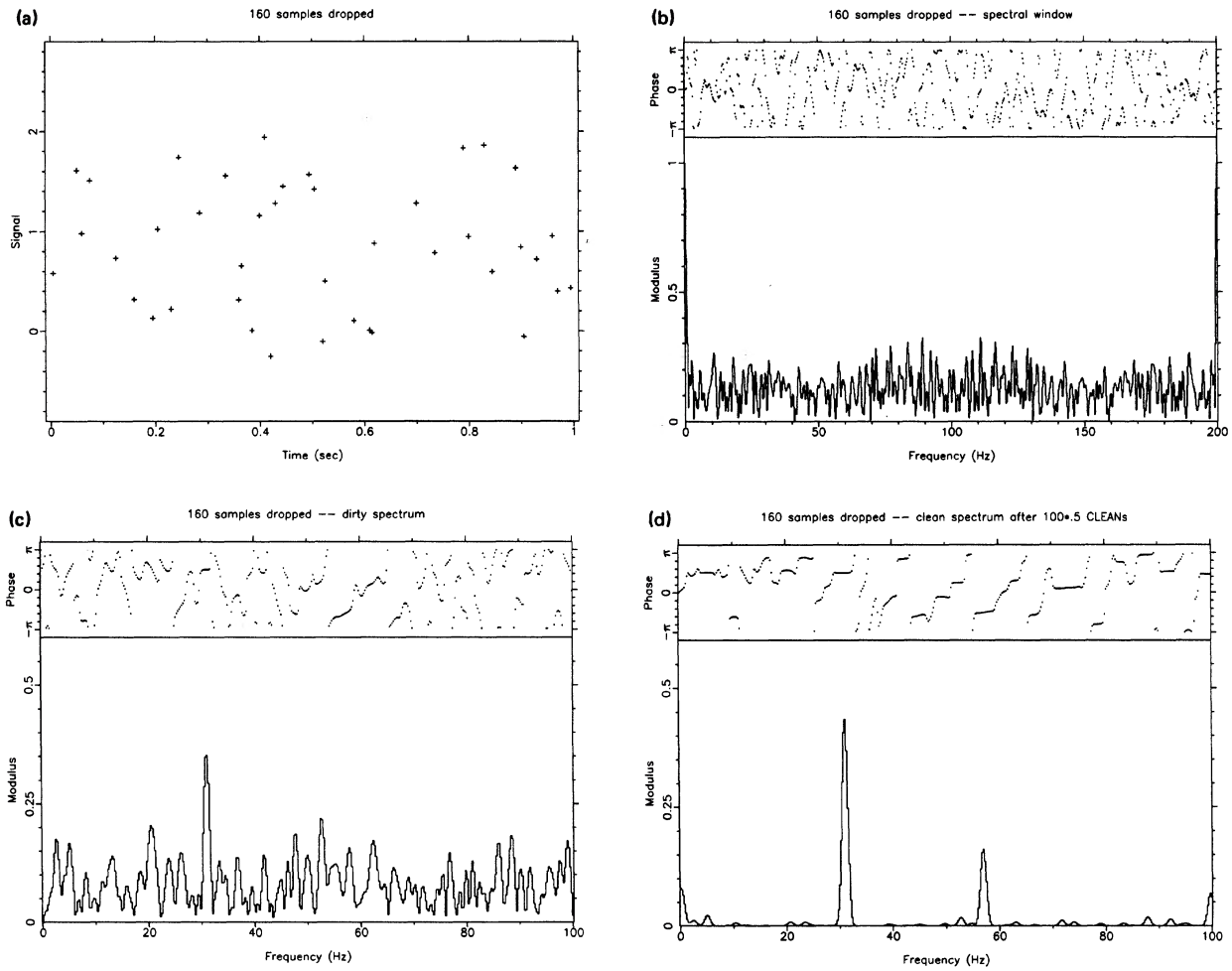

FIG. 5. Two-hundred-one equally spaced noiseless data from Fig. 3, less 160 random drops. Same parts as Fig. 3.

the time series in the figure, careful examination of Fig. 10 shows that inversion of the sum of the first three clean components (made with gain 1) does not quite reproduce the time series. This is because each of the clean iterations was somewhat in error, owing to the contribution at each spectral peak of the sidelobes of other cosinusoids not yet removed. However, successive iterations of CLEAN correct these small errors, and the inversion of the final result (after one hundred iterations) reproduces the input time series essentially exactly.

e) Resolution Tests

The ability to discriminate between two neighboring peaks in a spectrum is a function of the intrinsic resolution of the spectrum ($\delta\nu \simeq 1/T$), the signal-to-noise ratio of the spectrum, and the relative phase of the underlying maxima in the spectrum. Although CLEAN does not actually enhance the resolution of a spectrum, it can make it a great deal simpler to "separate" two close components since it removes the confusion of sidelobes. This is illustrated in Figs. 12–16, where two harmonic components whose amplitudes are in the ratio of 3:1 and whose phase constants differ by $\pi/2$ are moved from 4 Hz to 0.25 Hz apart in successive steps of a factor of 2. No noise was added to the spectrum. The data sampling is taken to be the same as in Fig. 1, where $T = 1$ s so $\delta\nu \simeq 1$ Hz. From these figures it is apparent that CLEAN makes it easy to "see" two components down to the nominal resolution of the spectrum. At a separation of 0.5 Hz $\simeq \delta\nu/2$, the existence of two components is still apparent from the asymmetry of the peak in the clean spectrum and from the strong phase gradient across the peak. Examination of the clean components shows that the frequencies, relative amplitudes, and phases of the two components are still recovered rather well. At a separation of 0.25 Hz $\simeq \delta\nu/4$ it is still apparent that the "single peak" is actually more complex, but nothing more quantitative may be said. It is possible to "super-resolve" the clean spectrum by choosing B to be narrower than W, or, in the extreme, by using the clean components as a direct model. We cannot recommend this technique, however, for it depends on the algorithm's extrapolation of the data to the region outside that measured, which cannot be expected to be very reliable.

The effect of noise on the resolution of the clean spectrum will be discussed in Paper II.

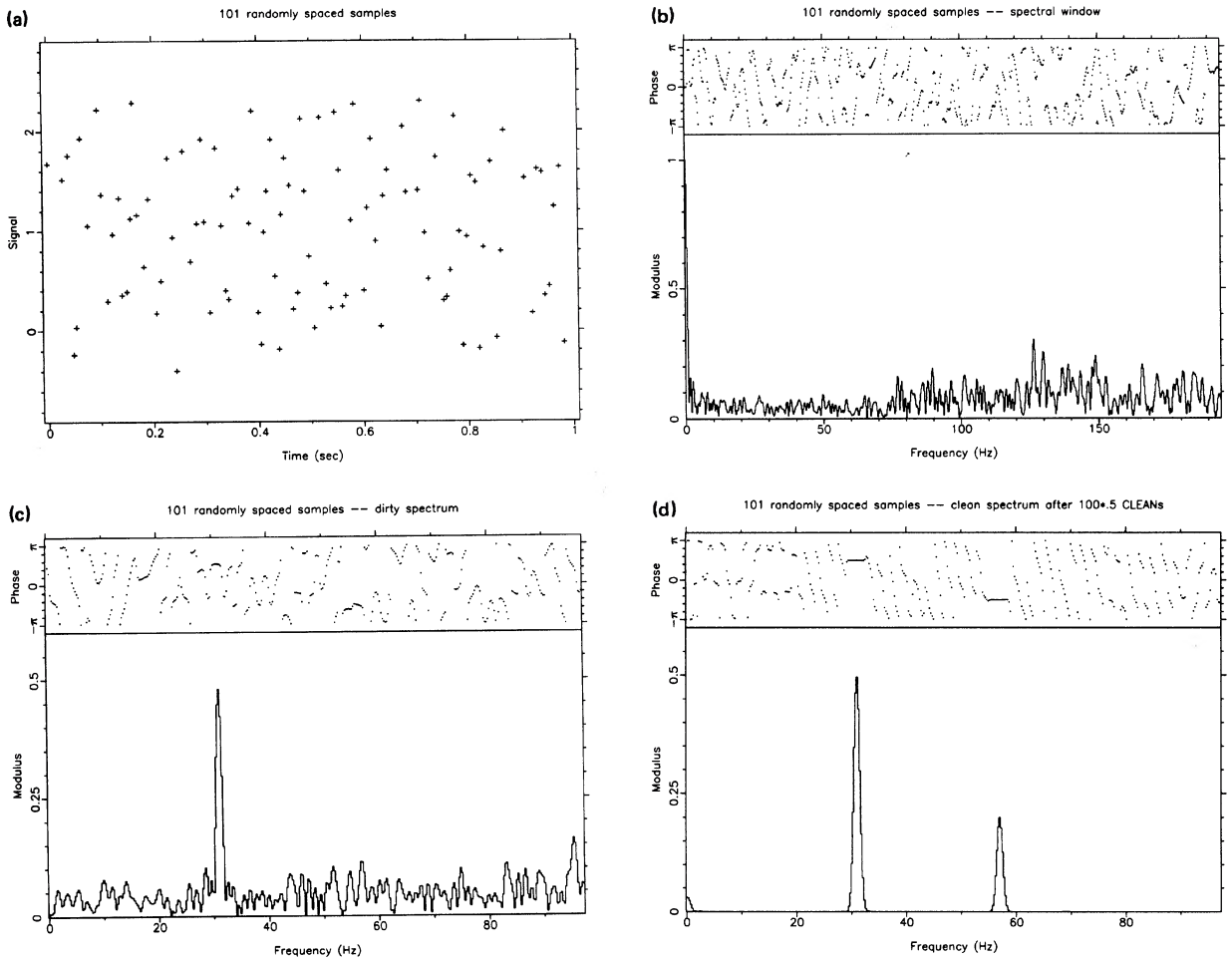

FIG. 6. One-hundred-one randomly spaced noiseless data containing harmonic components at 31 and 57 Hz. Same parts as Fig. 3.

V. DISCUSSION

We have shown that the CLEAN algorithm can be successfully applied to the problem of deconvolving the spectrum obtained from sampled data, within the resolution limits resulting from the finite time span of the observations. The method should be reliable for spectra consisting of a number of frequency components that is much less than the number of data samples. The reliability for more complex spectra, as well as the effect of noise on the deconvolution, will be considered in Paper II.

The CLEAN method compares favorably with the few alternatives for the analysis of unevenly sampled time series that can be found in the literature. Deeming (1975a,b) and Scargle (1982) discuss the formation of the discrete spectrum (our equations (7) and (8)) and the problems of interpreting $|D|^2$, the "periodogram," that results. These authors do not discuss any methods of deconvolution. A number of techniques for finding a single periodicity in the data have been suggested. The most obvious, and frequently applied, method is simply to use the peak of the periodogram. Jurkevich (1971) presents a quantitative method of searching for periodicity by binning. The "Jurkevich diagrams" that result have been shown to be equivalent, in most cases, to the periodogram (Swingler 1985). The phase-dispersion method is described and compared to the periodogram and other least-square methods by Pelt (1983). Dworetsky (1983) discusses a generalization of this method that is especially suitable for data whose sampling interval exceeds the period. All of these techniques are designed to find a single periodicity; their application to more complex signals may be impractical and is certainly not well understood. Kuhn (1982) and Swann (1982) have investigated more general techniques for deconvolving D. By considering only a finite set of frequencies, they reduce the deconvolution problem to one of matrix inversion. Unfortunately, the same problem discussed in Sec. II*a* shows up in their methods as an ill-conditioning problem with the matrices. At worst, this problem can cause these methods to fail entirely; in all cases, it amplifies the noise. Finally, the techniques generally known as "maximum entropy," or MEM, have often been applied to time series in order to extrapolate the data (see, for example, Haykin 1983). While, in principle, MEM could be used to interpolate data in an unevenly sampled time series, in practice, existing implementations all require interpolation to an evenly spaced sampling grid by other methods before MEM can be applied (McDonough 1983).

Thus CLEAN seems, at present, to be the only method suit-

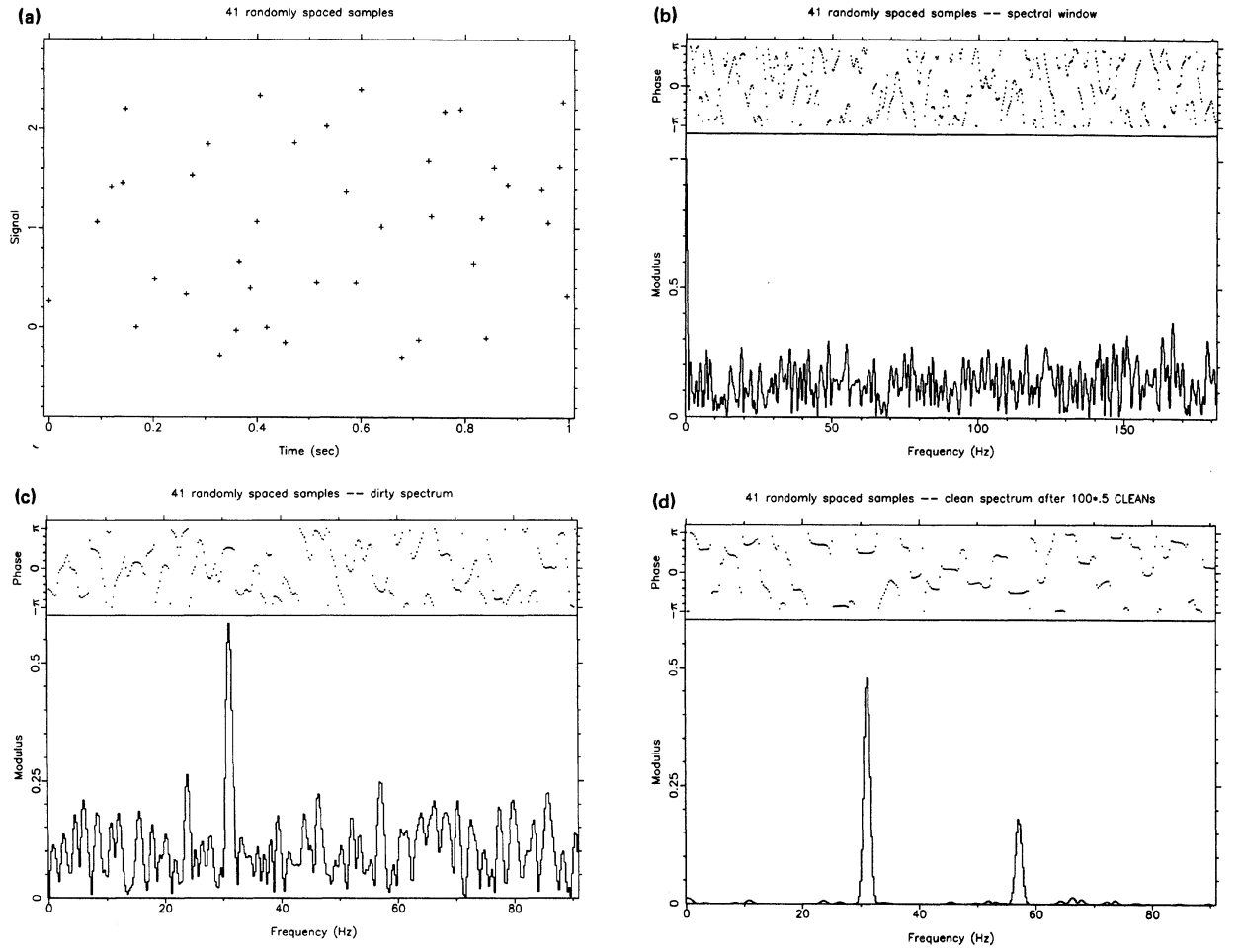

FIG. 7. Forty randomly spaced noiseless data containing harmonic components at 31 and 57 Hz. Same parts as Fig. 3.

able for the spectral analysis of irregularly sampled, deterministic time series for which the true spectrum may be multiply periodic or quasiperiodic. Even in those situations that the other methods can handle, the simplicity and intuitive nature of CLEAN render it an appealing alternative.

This research has been partly supported by the NSF under grants AST83-15945 and AST85-19529 (D.H.R.) and AST82-10966 (J.W.D.). We thank Jeff Scargle and John Wardle for some very helpful remarks. Part of this work has been submitted to Brandeis University as part of a B.A. thesis (Lehár 1985).

APPENDIX A: SYNTHESIS MAPPING AND CLEAN

a) Synthesis Mapping

Synthesis maps are produced by the Fourier transformation of interferometer visibility data, the complex cross-correlation of the voltage outputs of two elements of the interferometer (see Thompson, Moran, and Swenson 1986 for a complete discussion). The visibility of a celestial source whose brightness distribution on the sky is $I(x,y)$ is the Fourier transform of the brightness,*

$$V(u,v) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\, I(x,y) e^{+2\pi i(ux+vy)}. \quad \text{(A1)}$$

The variables u and v are the east (x) and north (y) components, expressed in units of the observing wavelength and viewed from the source, of the baseline **B** separating two elements of the interferometer. Since I is real, V satisfies $V(-u,-v) = V^*(u,v)$; thus measurement of the visibility at (u,v) also determines it at $(-u,-v)$. If we knew V for all (u,v) we could determine I by the inverse transform,

$$I(x,y) = \int_{-\infty}^{+\infty} du \int_{-\infty}^{+\infty} du V(u,v) e^{-2\pi i(ux+vy)}. \quad \text{(A2)}$$

However, real interferometers collect only a finite number $2N$ of data points, so the sampled data are

$$V_s(u,v) = \frac{1}{2N} \sum_{r=1}^{N} (V(u,v)\delta(u-u_r)\delta(v-v_r) + V^*(u,v)\delta(u+u_r)\delta(v+v_r)),$$

and their Fourier transformation yields a *dirty map*,

*This is strictly valid only in the paraxial approximation, where the so-called w term may be neglected.

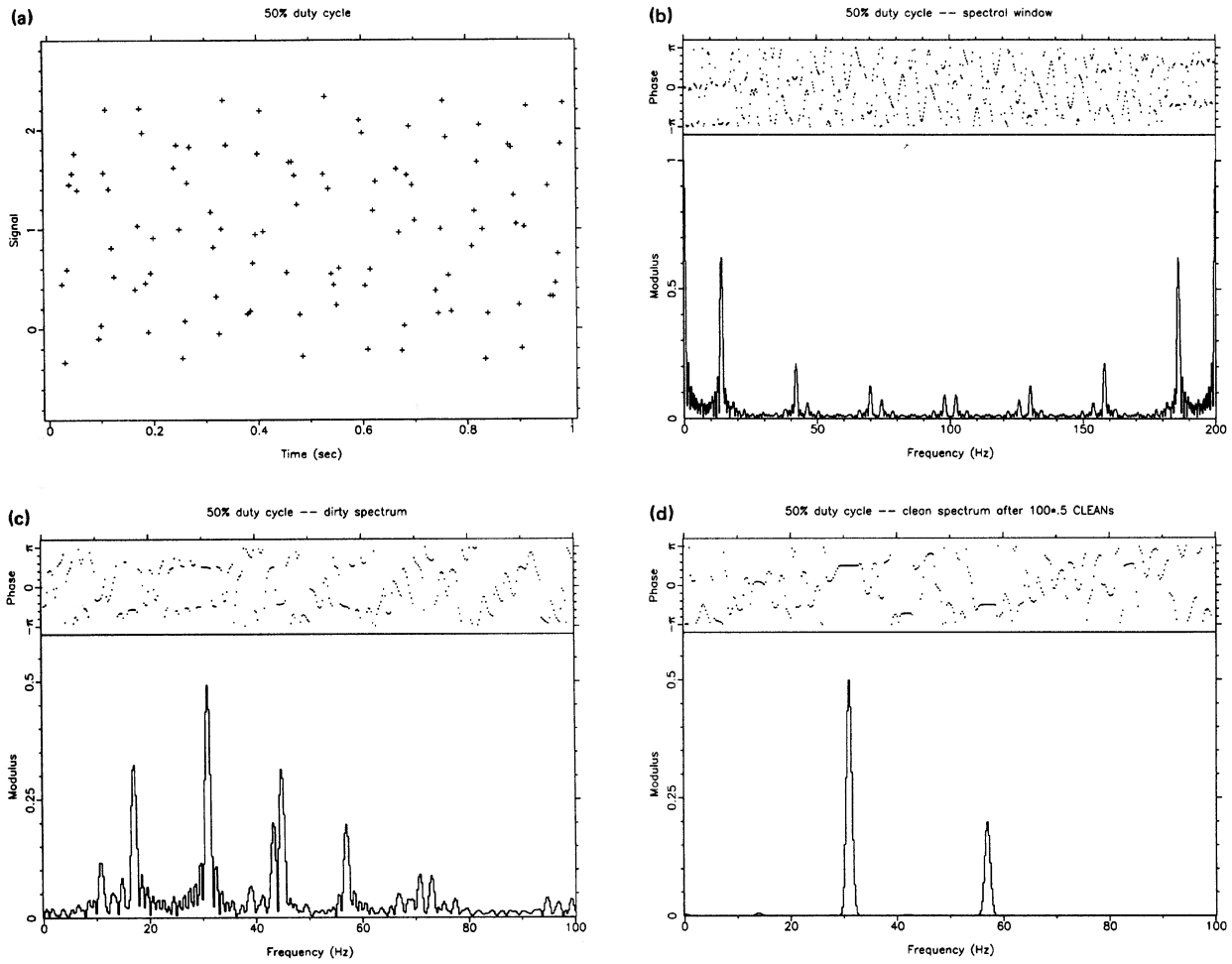

FIG. 8. Two-hundred-one equally spaced noiseless data containing harmonic components at 31 and 57 Hz, windowed with a 50% duty cycle. Same parts as Fig. 3.

$$FT[V_s] = M(x,y)$$
$$= \frac{1}{2N}\sum_{r=1}^{N}(V_r e^{-2\pi i(u_r x + v_r y)} + V_r^* e^{+2\pi i(u_r x + v_r y)}). \quad \text{(A3)}$$

Associated with this is a two-dimensional "window function" called the *dirty beam*,

$$W(x,y) = \frac{1}{2N}\sum_{r=1}^{N}(e^{-2\pi i(u_r x + v_r y)} + e^{+2\pi i(u_r x + v_r y)})$$
$$= \frac{1}{N}\sum_{r=1}^{N}\cos(2\pi(u_r x + v_r y)). \quad \text{(A4)}$$

The brightness distribution is related to the dirty map and dirty beam by the convolution

$$M(x,y) = I(x,y) \otimes W(x,y). \quad \text{(A5)}$$

Equations (A3), (A4), and (A5) are the two-dimensional analogs of Eqs. (7), (8), and (6). Here I, M, and W are all real functions, and W satisfies $W(x,y) = W(-x,-y)$.

b) The CLEAN Algorithm for Synthesis Maps

The CLEAN algorithm was devised to estimate the brightness on the sky I given a dirty map M and dirty beam W. The assumption of CLEAN is that I can be expressed as the sum of a small number p of point components,

$$I(x,y) \simeq I_{\text{mod}}(x,y) = \sum_{i=1}^{p} I_i \delta(x - x_i)\delta(y - y_i). \quad \text{(A6)}$$

Under this assumption the dirty map is the sum of p window functions "centered" at (x_i,y_i), weighted by the I_i:

$$M_{\text{mod}}(x,y) = \sum_{i=1}^{p} I_i W(x - x_i, y - y_i).$$

The CLEAN algorithm consists of identifying and removing the contributions of each model point from the peaks of M. This is accomplished by successively subtracting functions of the form $I_i W(x - x_i, y - y_i)$ from M until nothing is left but noise. Each subtraction removes not only the contribution of that point component at (x_i,y_i), but also its sidelobes

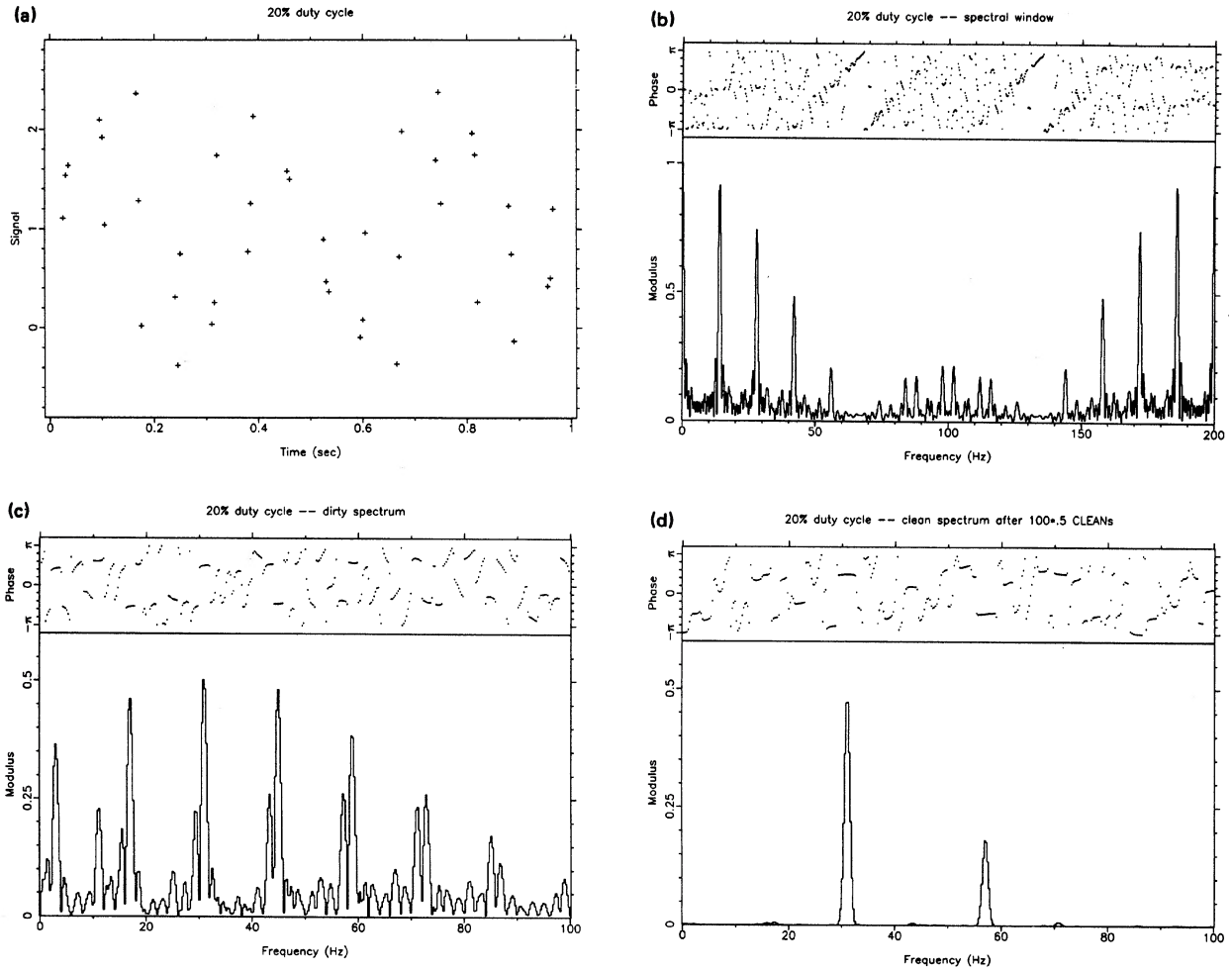

FIG. 9. Two-hundred-one equally spaced noiseless data containing harmonic components at 31 and 57 Hz, windowed with a 20% duty cycle. Same parts as Fig. 3.

at all (x,y). The resulting *clean components* $\{(x_i,y_i),I_i\}$ contain the positions and amplitudes of the real features of I, but none of the spurious features introduced into M by the sidelobes of W. To the extent that I satisfies Eq. (A6), we have succeeded in finding it. A map of the sky with the same resolution as the dirty map of the sky with the same resolution as the dirty map is made by convolving the clean components with a *clean beam*. The clean beam is a window function constructed to have the same frequency resolution $\delta\nu$ as W, but no sidelobes; it is usually taken to be an ellipsoidal Gaussian with the same major- and minor-axis full-widths at half-maximum and major-axis position angle as the central spike of the dirty beam.

In practice, each subtraction is not permitted to remove the full value of the highest peak of the map, but only a fraction g of the peak. This fraction g is called the *gain* of the CLEAN; for convergence, the gain must obey $0 \leqslant g \leqslant 2$ (Schwarz 1978). Typically, the gain is chosen in the range $0.1 \lesssim g \lesssim 1.0$. This restriction of the size of each step in the CLEAN is necessary because if there is more than one peak in the map, and anytime there is noise, each subtraction will be slightly in error. These errors are due to the presence at (x_i,y_i) of contributions from the sidelobes of components which have not yet been CLEANed away, and to the contributions from noise. The use of a gain less than one permits CLEAN to "correct" these small errors, often by putting in a "negative" clean component to compensate for a previous oversubtraction. Experience has shown that a gain of a few tenths is usually optimal. Gains too close to one can produce errors which cannot be recovered by subsequent iterations, while gains which are too small simply take too long.

A clear discussion of CLEAN is given by Thompson, Moran, and Swenson (1986); the mathematical basis of CLEAN has been studied by Schwarz (1978).

In the aperture-synthesis case, the map of the sky I (a real function) is the FT of V (a complex function), where $V(-u,-v) = V^*(u,v)$, and, due to sampling, the dirty map M satisfies $M = I \otimes W$. In the spectral-analysis case, the spectrum F (a complex function), where $F(-\nu) = F^*(\nu)$, is the FT of f (a real function), and, due to sampling, the dirty spectrum satisfies $D = F \otimes W$. Despite these differences, estimations of I and F by CLEAN are analogous as long as the presence in F and D of components at both positive and negative frequencies is taken into account. Thus much

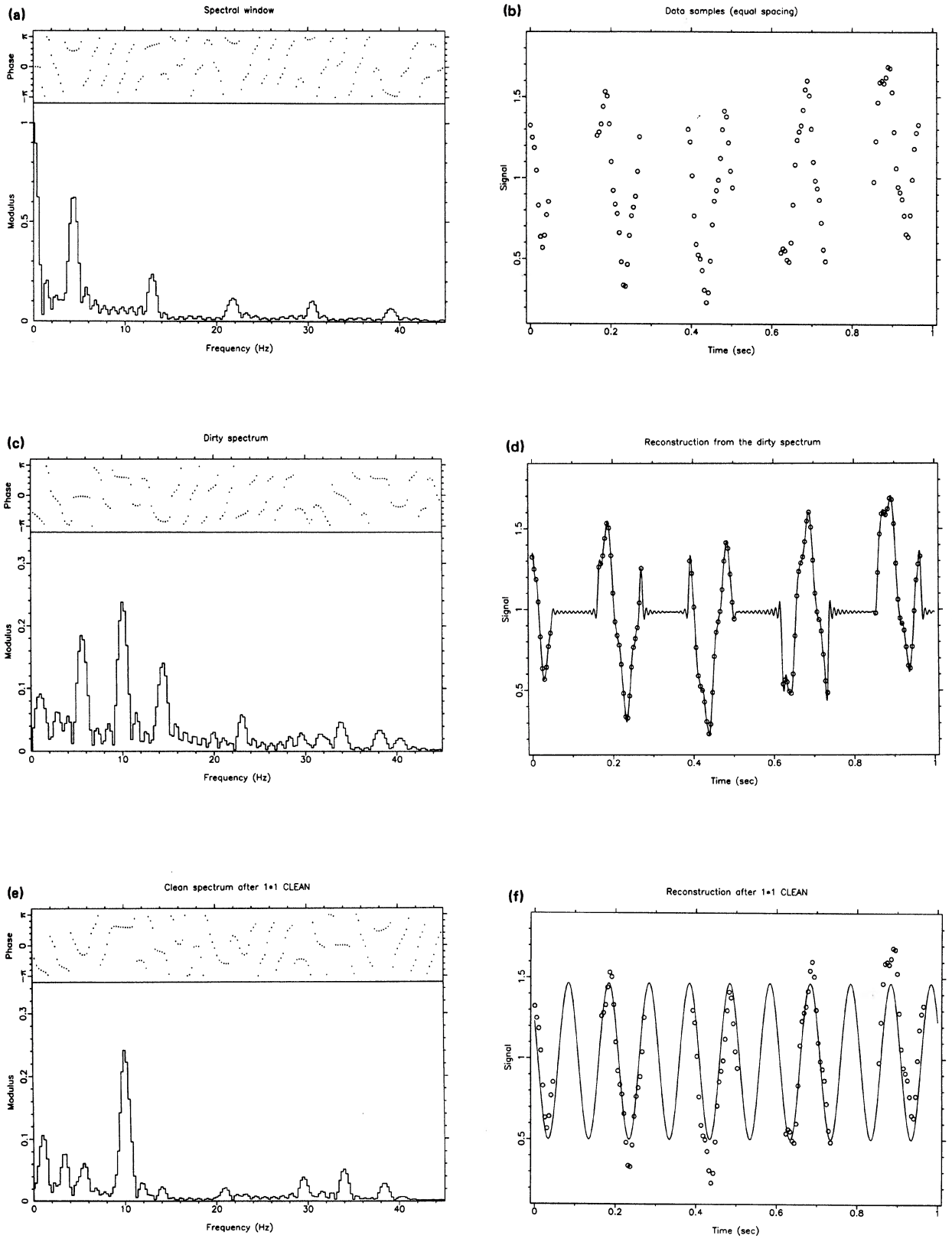

FIG. 10. Inversion of spectral analysis of data containing harmonic components at 1, 10, and 34 Hz, sampled with a duty cycle of 50% using otherwise equally spaced time points. (a) The spectral window, (b) the time series, (c) the dirty spectrum, (d) data reconstruction from dirty spectrum, and (e)–(h) clean spectra and data reconstruction after one, two, three, and one hundred iterations with gain $g = 1$.

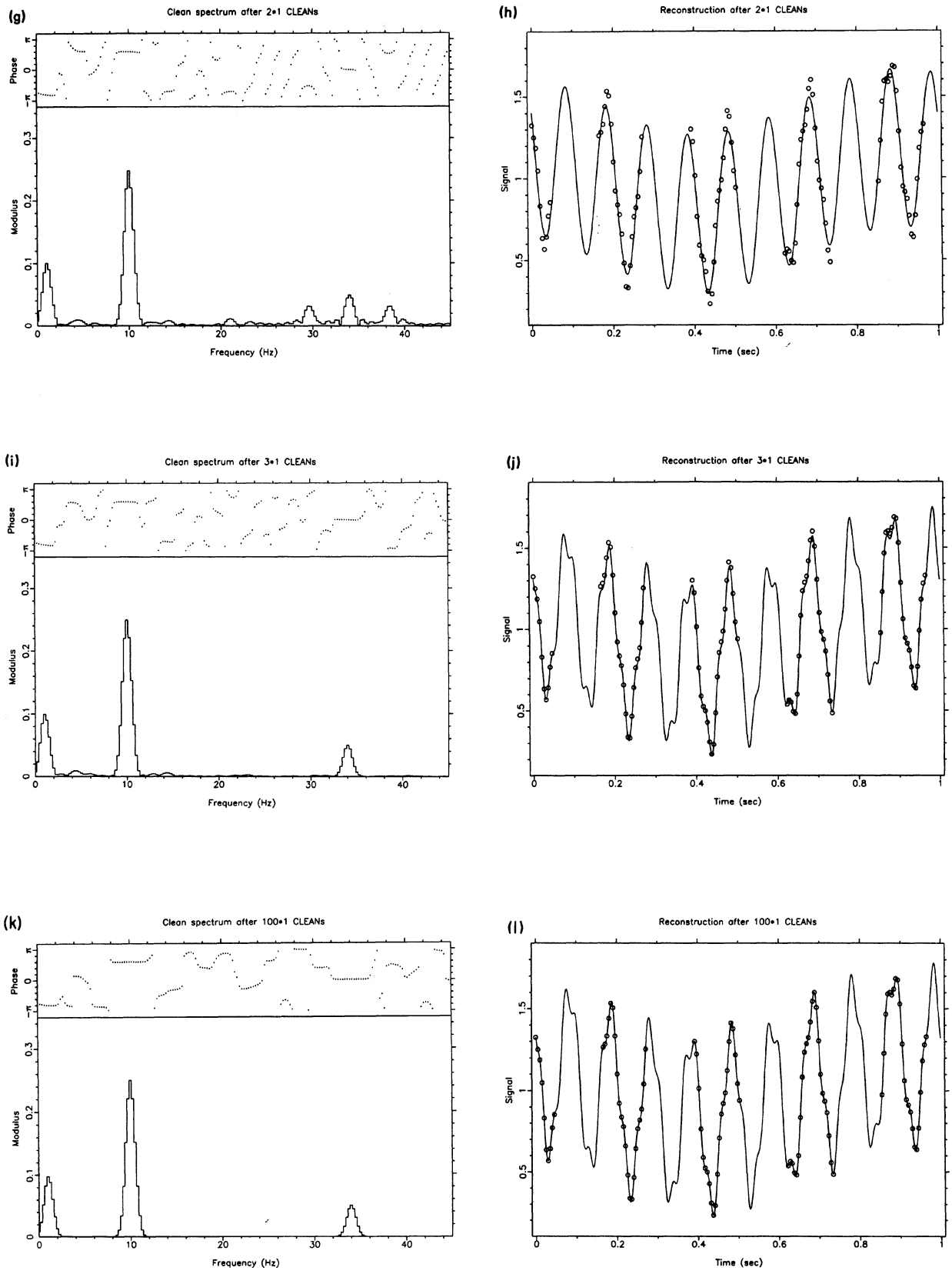

FIG. 10. (continued)

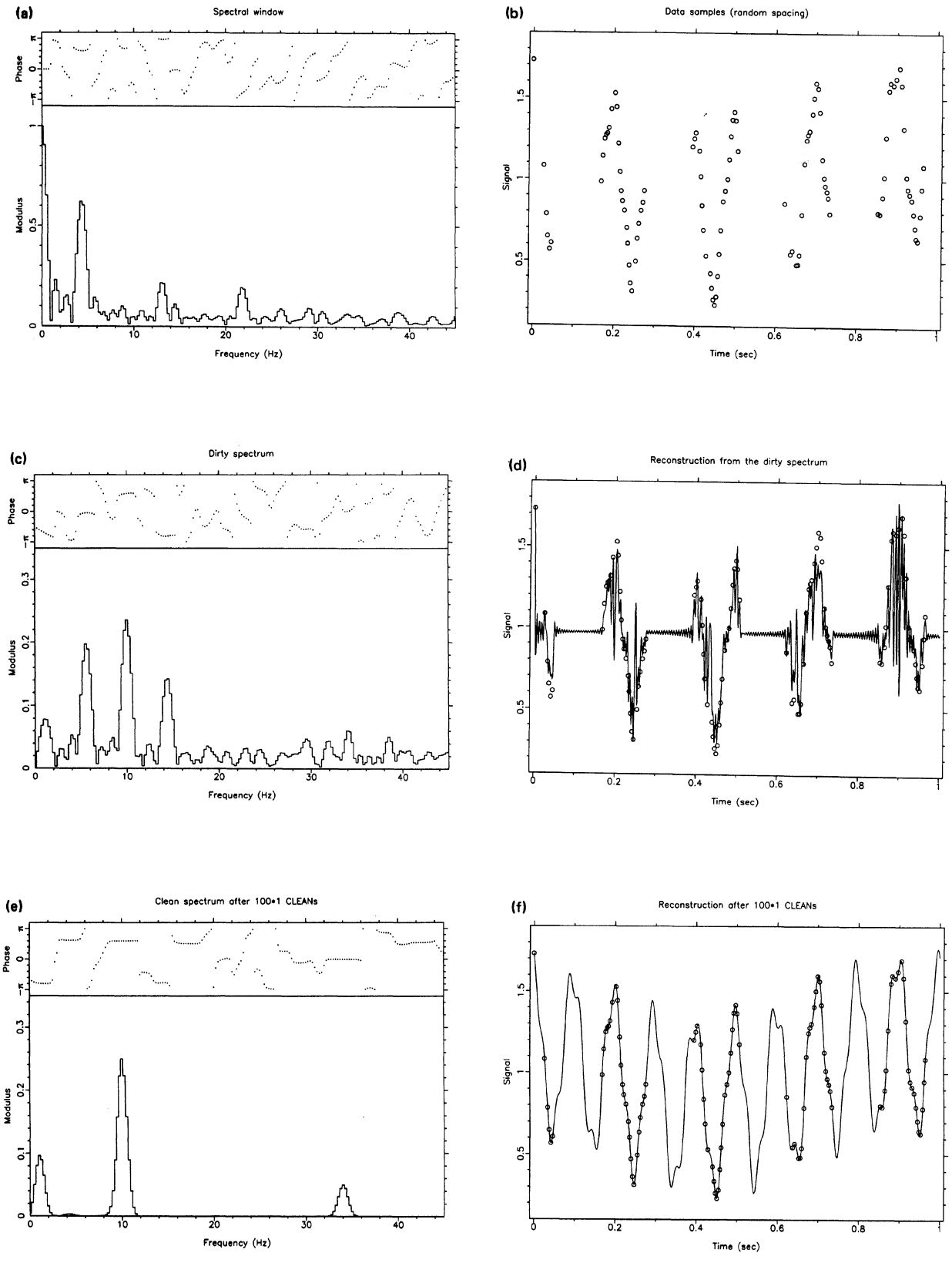

FIG. 11. Inversion of spectral analysis of data containing harmonic components at 1, 10, and 34 Hz, sampled with a duty cycle of 50% using unequally spaced time points. (a) The spectral window, (b) the time series, (c) the dirty spectrum, (d) data reconstruction from dirty spectrum, and (e) and (f) clean spectrum and data reconstruction after one hundred iterations with gain $g = 1$.

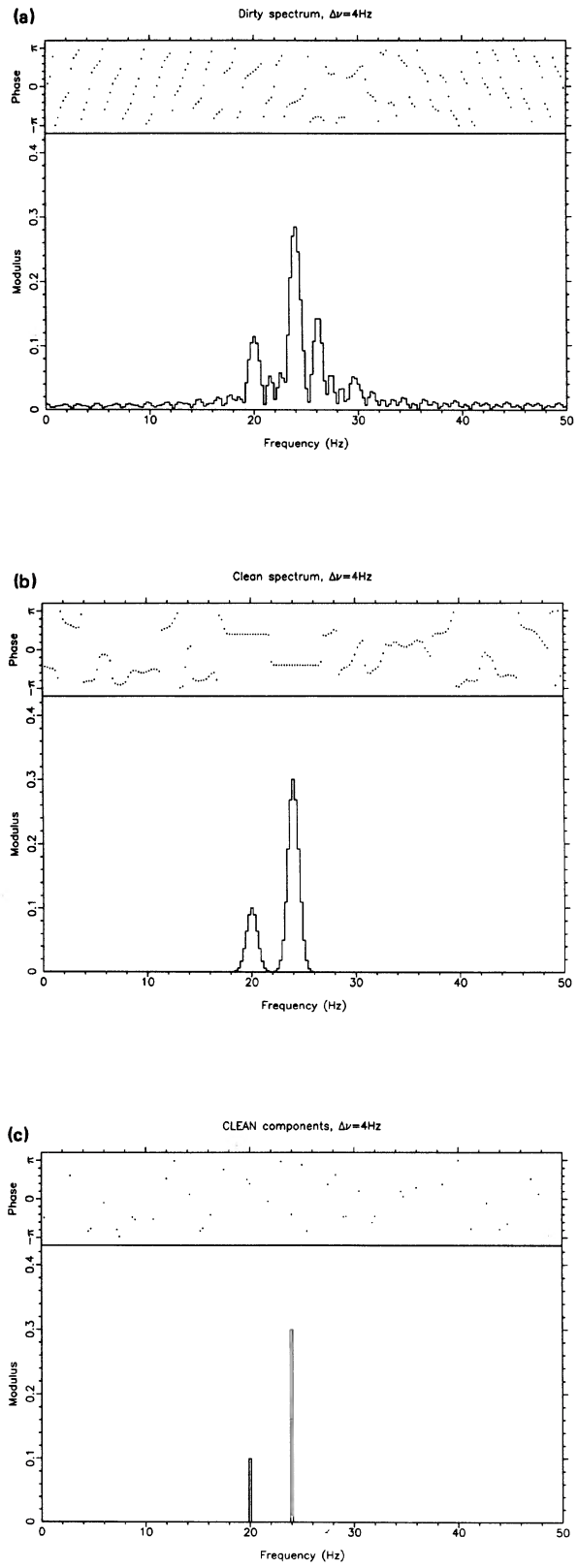

FIG. 12. Test of the resolution of a clean spectrum. The two components have frequencies, amplitudes, and phases of (20.0,24.0) Hz, (0.2,0.6), and $(+\pi/2, -\pi/2)$, respectively. Parts (a), (b), and (c) are the dirty spectrum, clean spectrum, and clean components, respectively.

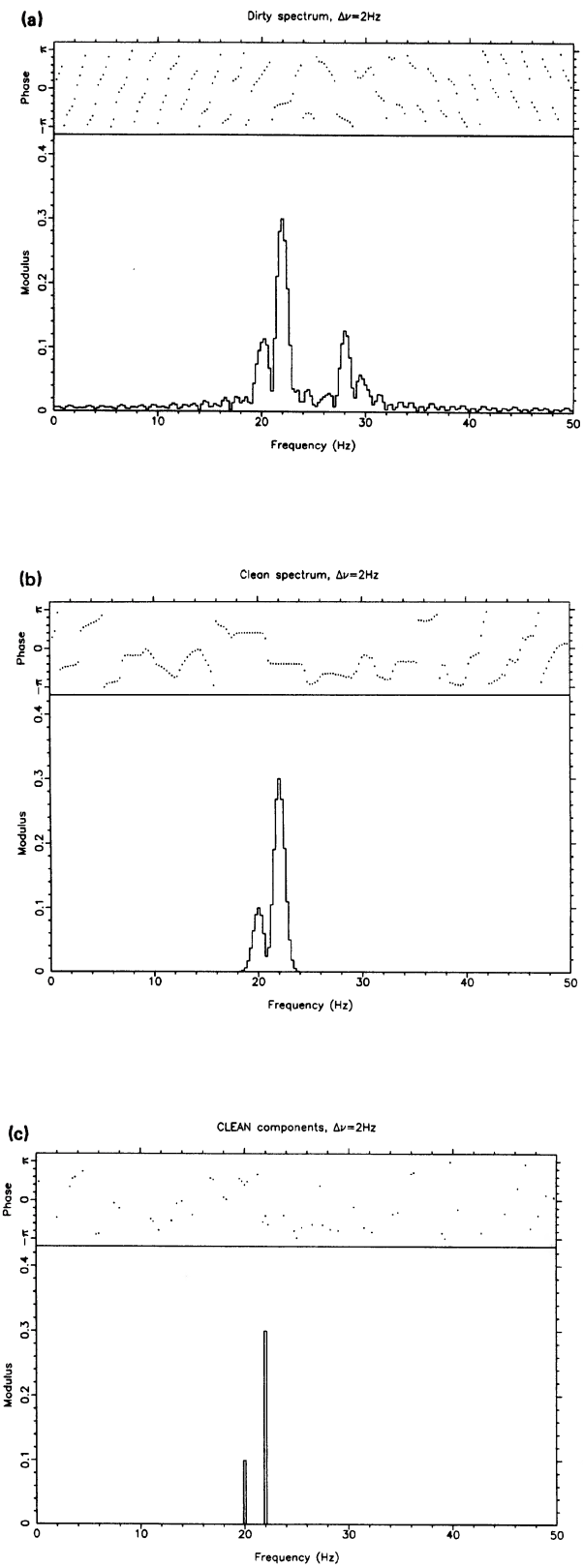

FIG. 13. Same as Fig. 12, but for frequencies of 20.0 and 22.0 Hz.

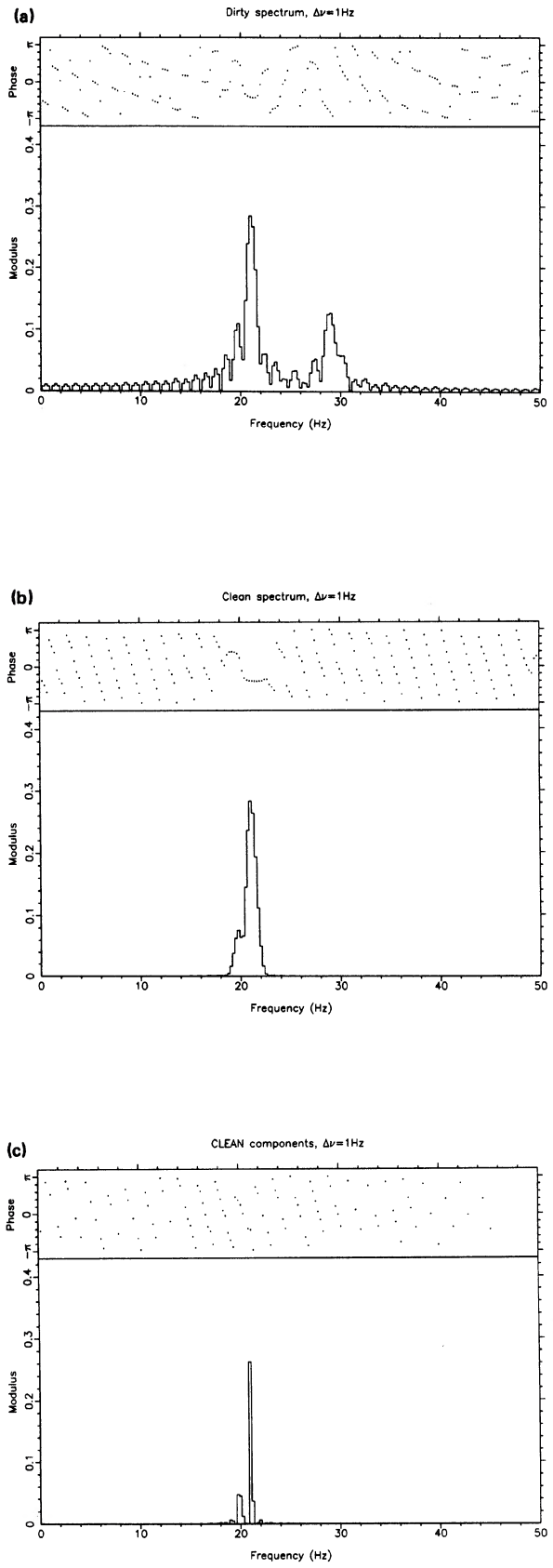

FIG. 14. Same as Fig. 12, but for frequencies of 20.0 and 21.0 Hz.

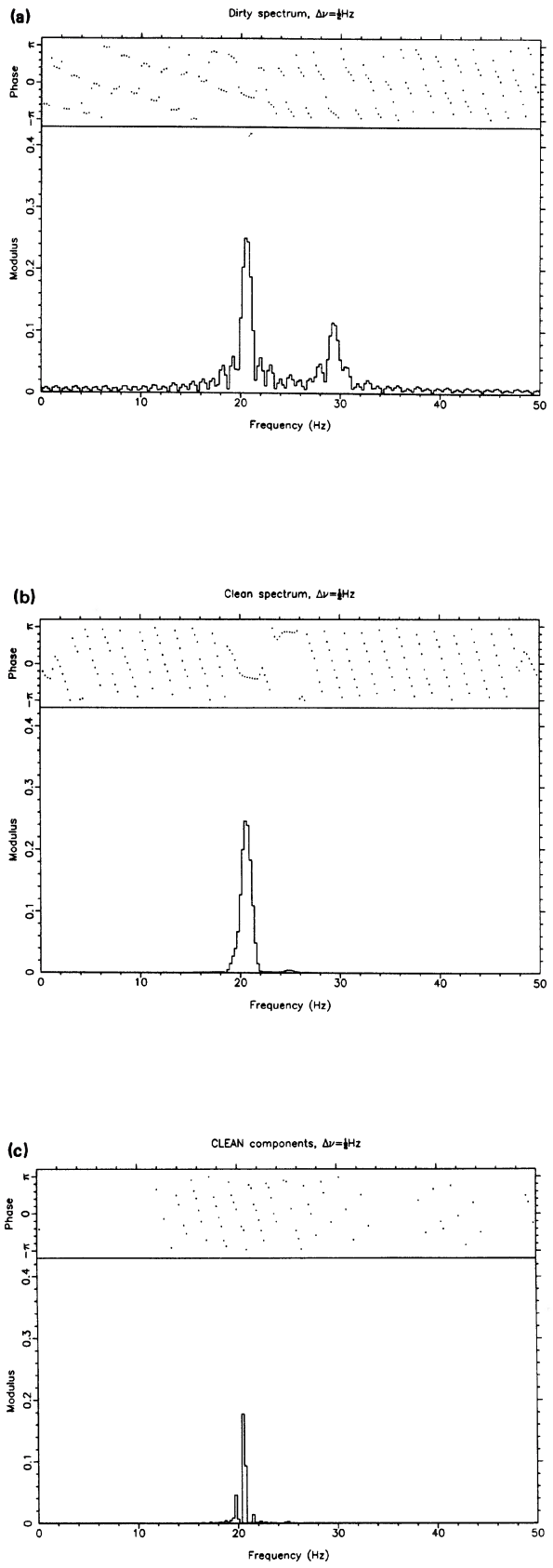

FIG. 15. Same as Fig. 12, but for frequencies of 20.0 and 20.5 Hz.

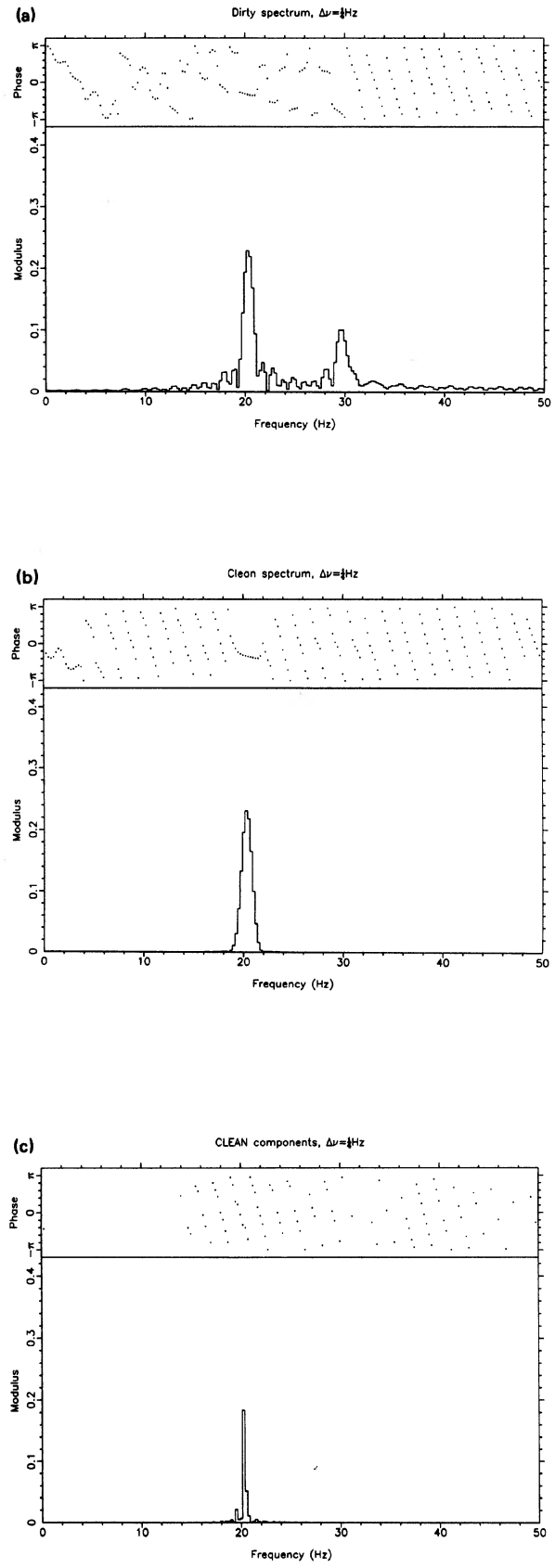

FIG. 16. Same as Fig. 12, but for frequencies of 20.0 and 20.25 Hz.

that has been learned about aperture synthesis applies to the spectral-analysis problem.

APPENDIX B: SOME CALCULATIONAL DETAILS

We evaluate the dirty spectrum and window function on finite arrays of discrete points. We assume that the dirty spectrum D has been determined at m positive frequency points, i.e., at $2m+1$ equally spaced points on $(-\nu_{\max},\nu_{\max})$,

$$\nu_j = \left(\frac{j}{m}\right)\nu_{\max},\quad j=-m,\ldots,m,$$

where $\nu_{\max}$ is the maximum frequency in the spectrum. Here, zero frequency (DC) is at $j=0$, and negative frequencies are labeled with $j<0$. In order to CLEAN D, the spectral window W must be determined on $4m+1$ points on $(-2\nu_{\max},2\nu_{\max})$. The clean spectrum S will be determined at the same frequency points as D. (For the sake of clarity we discuss evaluation of the various functions at an odd number of points distributed symmetrically about $\nu=0$. In actual fact, the symmetry of D, W, and S enables us to calculate only the positive-frequency parts.)

For simplicity we take the maximum frequency to be

$$\nu_{\max}=\frac{1}{2\Delta_{\min}},$$

where $\Delta_{\min}$ is the smallest interval between adjacent time samples. The number of positive-frequency points m is related to the desired number of points per beam n_B and the frequency resolution $\delta\nu\simeq 1/T$ by $n_B=\delta\nu/(\nu_{\max}/m)$, so

$$m=n_B\left(\frac{\nu_{\max}}{\delta\nu}\right)\simeq n_B\left(\frac{T}{2\Delta_{\min}}\right),$$

where T is the total length of the data span. In the case of $N=2n$ equally spaced data this reduces to $m\simeq n_B(N/2)\simeq n_B n$. In most cases we use $n_B=4$ and have $m\simeq 4n$.

The DC component of the dirty spectrum is usually removed by subtracting the data mean from the time series before Fourier transformation. While this is unnecessary (CLEAN will remove the DC and its sidelobes properly, as long as one remembers that only half of the zero-frequency component occurs in the positive-frequency part of the spectrum), we follow this convention in order to make fair comparison with standard techniques. It is convenient to arrange the phase of the window function to have zero slope at the origin since then the clean beam may be taken to be a real function. This is accomplished by shifting the times so that their average, too, is zero. Thus the discrete realizations of the continuous functions D and W are

$$D_j=\frac{1}{N}\sum_{r=1}^{N}(f_r-\langle f_r\rangle)e^{-2\pi i\nu_j(t_r-\langle t_r\rangle)},\quad j=-m,\ldots,m,$$

and

$$W_j=\frac{1}{N}\sum_{r=1}^{N}e^{-2\pi i\nu_j(t_r-\langle t_r\rangle)},\quad j=-2m,\ldots,2m.$$

All examples have been generated using the convention that the phase constant ϕ' is the phase of the cosine at the average time $\langle t_r\rangle$.

APPENDIX C: DISCRETE REALIZATION OF CLEAN

Using subscripts to denote array elements and superscripts to identify the order of the iteration, implementation of the CLEAN algorithm is:

(1) On the ith iteration, find the amplitude $\alpha(R^{i-1};\nu_{\rm peak})$ of the largest component of the previous residual spectrum ($R^0 \equiv D$); denote the positive-frequency bin which corresponds to $\nu_{\rm peak}$ by $I(i)$, so that $\nu_{\rm peak} = \nu_{I(i)}$. Call $\alpha(R^{i-1};\nu_{\rm peak}) = a^i$.

(2) Calculate the contribution of this component to the dirty spectrum from Eq. (23), and subtract a fraction g of the result from R^{i-1}, forming the residual spectrum R^i:

$$R^i_j = R^{i-1}_j - g(a^i W_{j-I(i)} + (a^i)^* W_{j+I(i)}),$$
$$j = -m,...,m.$$

Add this clean component (ga^i) to the $2m+1$ element clean components array C_j, at both positive and negative frequencies. This affects elements $j = \pm I(i)$, so with $C^0_j = 0$,

$$C^i_{+I(i)} = C^{i-1}_{+I(i)} + ga^i$$

and

$$C^i_{-I(i)} = C^{i-1}_{-I(i)} + g(a^i)^*.$$

Note that a zero-frequency (DC) clean component ($I(i) = 0$) is handled correctly by this prescription.

(3) Examine R^i and the accumulated clean components. If the peak of R^i is still above the "noise," or if the sum of the clean components is still changing significantly as more are added, the procedure has not yet converged: continue iterating at (1). When convergence is reached (after K iterations), exit at (4).

(4) Construct a clean beam B_j by fitting W_j with a Gaussian function with $B_0 = 1$. If the average time has been subtracted from each datum, the clean beam may be taken to be real; otherwise, fit a linear phase gradient to the peak of W_j and apply it to B_j.

(5) Form the clean spectrum S_i by convolving the clean-components array with the clean beam and adding the final residual spectrum,

$$S_j = \left(\sum_{k=-m}^{m} C^K_k B_{j-k}\right) + R^K_j, \quad j = -m,...,m.$$

A FORTRAN version of the CLEAN algorithm is available from the authors either as a listing or in machine-readable form.

Note added in proof: Since the acceptance of this paper, we have been informed that techniques apparently very similar to ours have been applied to the deconvolution of the frequency spectra of solar oscillations by Delache and Scherrer (1983) and by Duvall and Harvey (1984). Delache and Scherrer used a scheme consisting of estimating the properties of a harmonic component from the highest peak in the dirty spectrum, determining an accurate amplitude, frequency, and phase from a limited FFT or a least-squares fit to the data for frequencies near the estimate, subtracting the contribution of that harmonic from the data, transforming the resulting time series into a new spectrum, and iterating to convergence. Although equivalent to our procedure, theirs is apparently more time consuming since it requires finding a spectrum at each iteration. Few details are provided by Duvall and Harvey, but it appears that their technique was the same as that which we describe. We thank Drs. Harvey and Scherrer for bringing their work to our attention.

REFERENCES

Bracewell, R. N. (1965). *The Fourier Transform and its Applications* (McGraw-Hill, New York).

Bracewell, R. N., and Roberts, J. A. (1954). Aust. J. Phys. **7**, 615.

Brault, J. W., and White, O. R. (1971). Astron. Astrophys. **13**, 169.

Deeming, T. J. (1975a). Astrophys. Space Sci. **36**, 137.

Deeming, T. J. (1975b). Astrophys. Space Sci. **42**, 257 (E).

Delache, P., and Scherrer, P. H. (1983). Nature **306**, 651.

Dreher, J. W., Roberts, D. H., and Lehár, J. (1986). Nature **320**, 239.

Duvall, T. L., Jr., and Harvey, J. W. (1984). Nature **310**, 19.

Dworetsky, M. M. (1983). Mon. Not. R. Astron. Soc. **203**, 917.

Haykin, S. (1983). *Nonlinear Methods in Spectral Analysis*, edited by S. Haykin (Springer, Berlin).

Högbom, J. A. (1974). Astron. Astrophys. Suppl. **15**, 417.

Jenkins, G. M., and Watts, D. G. (1968). *Spectral Analysis and its Applications* (Holden-Day, San Francisco).

Jurkevich, I. (1971). Astrophys. Space Sci. **13**, 154.

Kuhn, J. R. (1982). Astron. J. **87**, 196.

Lehár, J. (1985). B. A. thesis, Brandeis University.

Lehár, J., Roberts, D. H., and Dreher, J. W. (1987). In preparation (Paper II).

McDonough, R. N. (1983). In *Nonlinear Methods in Spectral Analysis*, edited by S. Haykin (Springer, Berlin), p. 240.

Pelt, J. (1983). In *Statistical Methods in Astronomy*, edited by E. J. Rolfe (ESA Scientific and Technical Publishing, Noordwijk), p. 37.

Scargle, J. D. (1982). Astrophys. J. **263**, 835.

Schwarz, U. (1978). Astron. Astrophys. **65**, 345.

Shannon, C. (1949). Proc. IRE **37**, 10.

Swann, P. R. (1982). Astron. J. **87**, 1608.

Swingler, D. N. (1985). Astron. J. **90**, 675.

Thompson, A. R., Moran, J. M., and Swenson, G. W., Jr. (1986). *Interferometry and Synthesis in Radio Astronomy* (Wiley, New York).