DIRECT QUADRATIC SPECTRUM ESTIMATION WITH IRREGULARLY SPACED DATA

Donald W. Marquardt and Sherry K. Acuff
E. I. du Pont de Nemours and Company, Inc.
Engineering Department
Wilmington, DE 19898
USA

1. INTRODUCTION

The Direct Quadratic Spectrum Estimation (DQSE) method was defined in Marquardt and Acuff, 1982. Some of the theoretical properties of DQSE were explored. The method was illustrated with several numerical examples. The DQSE method is versatile in handling data that have irregular spacing or missing values; the method is computationally stable, is robust to isolated outlier observations in irregularly spaced data, is capable of fine frequency resolution, makes maximum use of all available data, and is easy to implement on a computer. Moreover, DQSE, coupled with irregularly spaced data, can provide a powerful diagnostic tool because irregularly spaced data are inherently resistant to aliasing problems that often are a limitation with equally spaced data.

In this paper the definition of the DQSE method is restated. A refinement is developed that reduces the small (and often negligible) negative offset of the spectrum ordinates, as reported in Marquardt and Acuff (1982), Section 13. This paper also comments on the relation between the "nonparametric" DQSE approach and the "parametric" approaches currently being pursued by others for irregularly spaced data.

2. IRREGULARLY SPACED TIME SERIES DATA

There are various causes by which data may be irregularly spaced: some data points may be missing in otherwise equally spaced data, the data point locations may vary unavoidably about equally spaced target locations, the series may have one or more extended gaps, the data points may inherently be obtained at random locations, eg, Poisson sampling, or the data points may deliberately be Poisson-sampled to provide "alias"-free data. The DQSE method can be employed whenever the individual observation times are known.
It will be assumed that the pattern of irregular spacing is "unrelated" to the stochastic properties of the process \( y(t) \) being sampled. The independent variable \( t \) may represent time units or distance units, depending on the application. It will be assumed that the process \( y(t) \) is at least weakly stationary. Thus, the covariance between two points at times \( t_i, t_j \) depends only on the time difference \( (t_j - t_i) \). It is also assumed that the mean of the process is zero, whence the covariance for time difference \( (t_j - t_i) \) is \( E[y(t_j)y(t_i)] \). Data obtained manually or data obtained over a long period of time are likely to suffer from irregular spacing. The Bibliography lists a number of references to applications as well as theory.

3. TYPES OF TIME SERIES ANALYSIS

Most work to date has been with equally-spaced data. During the 1950s and 1960s, stimulated by the Blackman and Tukey (1959) work, the emphasis turned to nonparametric spectral analysis methods. The subsequent widespread use of the Fast Fourier Transform (FFT) enhanced the practical usefulness of the nonparametric spectral analysis procedures. During the 1970s parametric models received heavy emphasis, stimulated by major advances in formulation (e.g., Box and Jenkins, 1970) and computing procedures. The use of parametric models for spectrum estimation is embodied in procedures such as the maximum entropy method (Findley, 1978, Haykin, 1979).

The recent emphasis on parametric models for equally spaced data has continued with the developing interest in irregularly spaced data. The papers in the present symposium undoubtedly represent a major cross section of current work in irregularly spaced time series; most take the parametric model approach. This paper and the paper of Masry (1983) are exceptions. We believe that the diagnostic insight available from the nonparametric spectral approach should be a part of almost all applications, including applications where a parametric model is needed and appropriate. The DQSE method makes this dual approach routinely practical.

4. NOTATION AND HISTORY

The spectral estimate is based upon a time series sample from the process \( y(t) \). The observations are made at known times \( t_i \) within a single continuous total period of observation \([0,T]\). The data are denoted

\[
(t_i, Y_i) \quad i = 1, 2, \ldots, n \quad (1)
\]

with \( 0 \leq t_i \leq T \).

For equally spaced data with interval \( \Delta t \), the estimated spectrum will cover the frequency range from zero cycles per unit time, to the Nyquist cutoff frequency \( 1/(2(\Delta t)) \). Any cyclic structure in \( y(t) \) that has higher frequency than the Nyquist frequency will be aliased with the computed power somewhere in the estimable frequency range 0 to \( 1/(2(\Delta t)) \). When the data are unequally spaced and there is a wide range of spacings throughout the total period of observation, the Nyquist effect is spread out, and does not create a sharp cutoff between the nominally estimable frequency range and the nominally aliased range. The more variable the interpoint spacing in the data, the more gradual the cutoff. Hence, the sensitivity for exhibiting high frequency cycles on the spectrum is diminished at frequencies