THE LEAST SQUARES SPECTRUM, ITS INVERSE TRANSFORM AND AUTOCORRELATION FUNCTION: THEORY AND SOME APPLICATIONS IN GEODESY

by

Michael Ruthven Craymer

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy, Graduate Department of Civil Engineering, University of Toronto

ABSTRACT

To realize the full potential of increasingly more accurate measurements, scientists are now faced with the task of modelling ever smaller effects on their observations to improve their results. The problem, however, is that there is often little understanding of the cause and effect relation between these so-called systematic effects and the measurements. Spectra and autocorrelation functions can be used to help diagnose and improve the modelling of these systematic effects in measurements. However, standard techniques for computing spectra and autocorrelation functions require the data to be evenly spaced, which is often not satisfied in practice.

The approach taken here is to develop a general technique for determining autocorrelation functions for data which are unevenly spaced. This is an indirect method whereby the systematic effects, represented by the residuals from an incomplete a priori deterministic model, are transformed into a power spectrum and then into an autocorrelation function. To accommodate unevenly spaced data, a general least squares transform and its inverse are developed. The inverse transform is used to obtain the autocorrelation function from the least squares spectrum originally developed by Vaníček [1971]. This formulation can accommodate unequally spaced data, random observation errors, arbitrary frequency selection, arbitrarily weighted and correlated observations, as well as the presence of any a priori deterministic model. The conventional Fourier transform and spectrum are shown to be just special cases of this more general least squares formulation. It is also shown how the individual spectral components in the least squares spectrum and inverse transform can be estimated either independently of or simultaneously with each other.

The advantages and limitations of the least squares transforms and spectra are illustrated through tests with simulated data. The technique of using autocorrelation functions to model systematic effects is also illustrated with two real applications; one

based on the precise measurement of the extension of a baseline spanning the San Andreas fault in California, and another based on the measurement of ellipsoidal heights using a GPS receiver under the influence of the effects of Selective Availability. These tests show that the use of fully populated weight matrices generally results in an increase in the value of the standard deviations of the estimated model parameters, thereby providing more realistic estimates of the uncertainties. On the other hand, the effect of correlations among the observations on the least squares estimates of model parameters was found not to be very significant.

ACKNOWLEDGMENTS

To Mary, Sarah, Lisa and Samuel.

This work is dedicated to my family, Mary, Sarah, Lisa and Samuel. It simply would not have been possible without their sacrifice and unfailing support and understanding for so many years. I owe them a huge debt of gratitude.

I am also deeply indebted to my supervisor, Professor Petr Vaníček for all his guidance, advice, generosity and persevering support. His tireless and meticulous efforts in reviewing my manuscripts and patience in dealing with my stubbornness are gratefully appreciated. He is truly the quintessential supervisor. I could not have been more fortunate to have him as my mentor.

I also thank the members of my Examination Committee, especially my Internal Appraiser, Professor Ferko Csillag (Geography), and my External Examiner, Professor Douglas E. Smylie (Earth and Atmospheric Science, York University). Their constructive comments and recommendations, together with those from the other members of my Examination Committee, are greatly appreciated.

The GPS data used in my tests were kindly provided by William Prescott of the U.S. Geological Survey in Menlo Park. I especially thank John Langbein, also of the U.S. Geological Survey in Menlo Park, for supplying the EDM data and for generously taking the time to discuss some of the results of my analyses.

I express my sincere gratitude to my employer, the Geodetic Survey Division of Geomatics Canada, and in particular Norman Beck and Lloyd Nabe for giving me time and support to complete this work.

Portions of this research were also funded by various Natural Sciences and Engineering Research Council of Canada Operating Grants held by Prof. Petr Vaníček during the years 1986 to 1990.

Finally, I thank the Department of Civil Engineering for giving me the opportunity to finish my dissertation after so many years. I especially thank Professor Robert Gunn for his help in this regard.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
Introduction

Recent advances in technology have produced extremely precise and accurate measuring systems that are affected by even the smallest effects that were once much too small to be noticed. In the past these effects were considered to be random noise to be averaged out. To realize the full potential of their measurements, scientists are now faced with the task of modelling these small effects in order to improve their predictions. The problem, however, is that there is often little understanding of the cause and effect relation between these so-called systematic effects and the measured observables.

There are basically two approaches to describing or modelling the measured observations. Deterministic models are used to explicitly describe the behaviour of the observations in terms of a mathematical model of the physical process. These deterministic models consist of constants and parameters to be estimated. Often, however, there is little understanding of the physical processes underlying the behaviour of the measurements. In the other approach, stochastic models treat the measurements, or what remains after removing a deterministic part, as unpredictable random (i.e., stochastic) quantities. Stochastic models describe the dependencies between the data and the incomplete deterministic model in terms of mathematical correlations. These correlations can be represented by filters, polynomials, correlation functions and spectral density functions. Because deterministic modelling is usually the preferred approach, correlations are often used to help diagnose and improve the deterministic model. In cases where this is not possible, the correlations, if carefully constructed, can be used to help describe the residual systematic effects within the deterministic model.

The least squares estimation technique is primarily used for fitting deterministic models to the measurements. However, it is also able to accommodate stochastic models through the use of a fully populated covariance matrix for the observations. There are different methods of determining the variances and covariances that form the observation covariance matrix. The most direct method involves determining an autocovariance function that describes the behaviour of various systematic effects. The problem with this, and the main motivation for this work, is that traditional techniques for computing autocorrelation functions require the data to be evenly spaced. This may not be the case, especially when looking for correlations with some (physically meaningful) parameters given as numerical functions. In practice such functions are often known only for values of the argument that are unevenly spaced.

The usual way of handling unevenly spaced data is to interpolate or approximate the original series to get an evenly spaced one. Because this approach tends to model the lower frequency content in the data, the low frequency behaviour of the measurements must be known. Moreover, the high frequency components can be lost by the smoothing effect of the interpolation or approximation.

The approach taken here is to develop a more general technique for determining autocorrelation functions for data which are unevenly spaced with respect to quantities describing the systematic effects. As will be seen later, there are two basic approaches to estimating autocorrelation functions. The most direct is to compute the autocorrelation function directly from the data. In this case, however, there is no satisfactory method of handling unevenly spaced points. There are methods based on averaging over larger, evenly spaced intervals or bins, but using these results in a loss of resolution.

The alternative approach is to estimate the autocorrelation function indirectly by first representing the systematic effects in terms of a power spectrum and then transforming this into an autocorrelation function. This is the approach taken here. Again the problem is that most techniques for computing the power spectrum require evenly spaced data as do those

for transforming the power spectrum to the autocorrelation function. The aim here is to find a more general technique that does not require evenly spaced data. To this end, a general least squares transform is developed.

Other methods are also available for determining the variances and covariances of the observations. The most popular of these are the methods of analysis of variance and variance-covariance component estimation. The "analysis of variance" (ANOVA) method (also called factor analysis in statistics) can be found in most standard texts on statistics. Geodetic applications of the technique are described in detail by Kelly [1991] and in a series of articles by Wassef [1959; 1974; 1976]. Essentially the aim of the method is to divide the measurements into separate groups (factors which contribute to the overall variation in the data) and to estimate the variance components for each. The difficulty in applying the method is defining a scheme of dividing the observations into separate groups which characterize some behaviour of the systematic effect being modelled. Often, the factors describing the systematic effect cannot be so discretely defined, rather they are often of a continuous nature that precludes lumping them together into separate and distinct groups.

Variance-covariance component estimation, on the other hand, is based on modelling deterministically the residual variation in the measurements. The variances and covariances are expressed in terms of linear models relating these components to various factors describing the systematic effect. The coefficients (variance and covariance components) in the variance-covariance model are estimated together with the parameters in a least squares solution. The technique is described in detail in Rao and Kleffe [1988] and has been applied to many geodetic problems (see, e.g., Grafarend et al. [1980], Grafarend [1984], Chen et al. [1990]). It can be shown that the analysis of variance method is just a special case of this more general approach [Chrzanowski et al., 1994] The problem with applying the method is that the estimation of the variance-covariance model coefficients usually needs to be iterated which can result in biased estimates of the variances and

covariances [Rao and Kleffe, 1988]. This can lead to negative variances, which is unacceptable.

The approach taken here is to model any residual systematic effects remaining after removing a deterministic model, using autocorrelation functions derived from a power spectral density function of the residuals. This idea was first proposed by Vaníček and Craymer [1983a; 1983b] and further developed by Craymer [1984]. To accommodate unevenly spaced data, the least squares spectrum, developed by Vaníček [1969a], was used and converted to an autocorrelation function using the inverse Fourier transform. However, the inverse Fourier transform is not completely compatible with the more general least squares spectrum. Consequently, a more general least squares transform and its inverse are developed here which are completely compatible with the least squares spectrum and can provide correct autocorrelation functions for data that are unevenly spaced. Although applied only to geodetic problems here, this technique should have wide application in many areas of science where one needs to model or analyze measured data.

Before describing the technique, a review of the basic concepts of stochastic processes and the conventional Fourier transform and spectrum are given. This is followed by the development of a new "least squares" transform and its inverse, and the reformulation of the least squares spectrum, originally developed by Vaníček [1969a; 1971], in terms of this new transform. It is then shown how an autocorrelation function can be derived from the least squares spectrum using the inverse least squares transform, and how this can be used in a procedure for stochastically modelling residual systematic effects. These developments are followed by tests with simulated data to examine numerically some of the limitations of the technique. It is also applied to a couple of examples in geodesy; the modelling of residual systematic effects in electronic distance measurement (EDM) data and point positioning data from the Global Positioning System (GPS). Finally, conclusions and recommendations for further investigations are given.

Throughout the sequel the following notation is used:

variables/observables	italic
vectors	lower case, boldface letters
matrices/operators	upper case, boldface letters
functions	upper or lower case letters, no boldface

Chapter 2
Basic Concepts of Stochastic Processes

2.1 Types of Processes

A process can be considered to be any kind of physical phenomenon that varies in some way. We examine such processes by taking measurements of them; i.e., by describing their physical behaviour in terms of numerical quantities that can then be analysed mathematically. These processes are most commonly represented as series of measurements (observations), often taken with respect to time (time series) or space (spatial processes). When regarded more generally as series with respect to any other argument, these processes are referred to here as simply data series.

Processes $\phi(t)$ are usually thought of as one-dimensional; that is, varying with respect to a one dimensional argument (t) such as time. However, a process may also be multidimensional; i.e., a function $\boldsymbol{\phi(t)}$ of a vector of arguments ($\boldsymbol{t}$) — e.g., processes $\boldsymbol{\phi(t)}$ which are functions of three-dimensional position ($\boldsymbol{x}$) in space or four-dimensional position in space-time. One may also encounter multiple processes $\boldsymbol{\phi(t)}$ of multiple arguments ($\boldsymbol{t}$).

Processes can be classified as either continuous or discrete. Examples of continuous processes are the crustal motions of land masses due to tectonic deformations or the motions of satellites in orbit about the Earth. On the other hand, the accumulated errors from point to point in a geodetic network would be classified as a discrete process (in space). Generally, one is only able to obtain discrete samples of continuous processes, primarily due to the nature of data acquisition systems.

2.2 Deterministic and Random Processes

Processes can also be classified as deterministic and random. What is random? "Everything and nothing" according to Kac [1983, pp. 405-406]. There is no one test for determining whether a process is either random or deterministic. The definitions most often used are only subjective or heuristic and a matter of philosophical debate. One person may consider a process to be random noise to be filtered out while another may consider the same random process to be a deterministic signal to be modelled.

The most straightforward definition is that deterministic implies predictability while random implies unpredictability. Thus what is considered deterministic and what is considered random depends on what one wishes to model . The deterministic part is what is being predicted or estimated exactly while the random or stochastic part is that which one can only predict or estimate with some degree of uncertainty. In the last century, instruments had rather limited precision and much of the variability in a process was considered to be random or stochastic, so that one could only predict with a great deal of uncertainty. More recently, however, new measuring techniques have become more precise so that it is now possible to attempt to model ever smaller variations in the data in an effort to improve the prediction power of the deterministic model.

2.3 Stationarity and Egodicity

Different realizations of a random process will, in general, not be identical. A single realization of a process is called a sample record. The collection or ensemble of all sample records is called a random or stochastic process (see Figure 2.1). In a random process, all sample records are different while in a deterministic process all samples are identical.

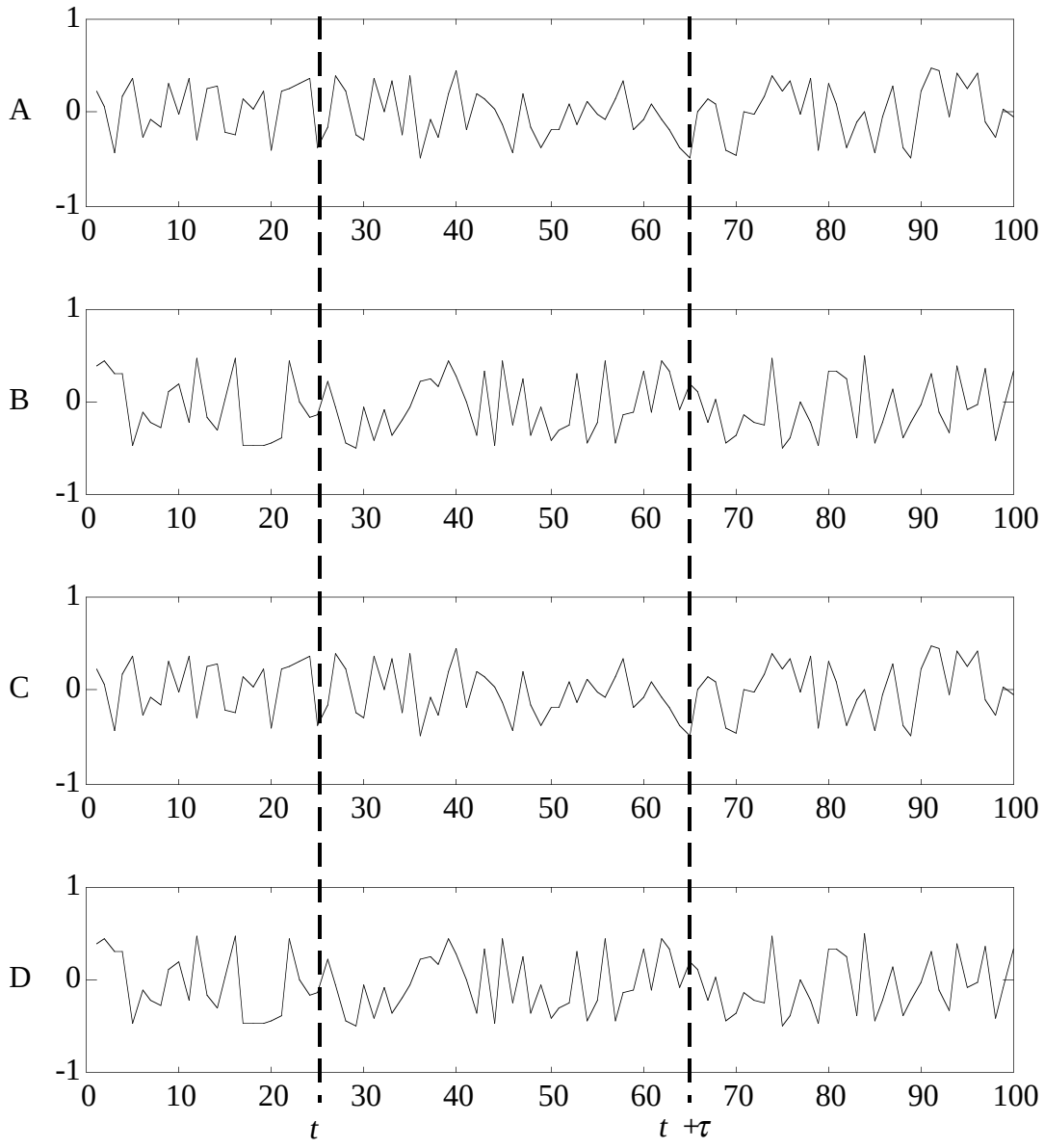

Figure 2.1: A single random process consisting of an ensemble of 4 sample records (A, B, C, D). There are 100 values of the argument ranging from 1 to 100. Ensemble or sample averages are taken over the four different sample records for each value (e.g., t or $t+\tau$) of the argument; i.e., there are 100 sample averages. Argument averages are taken over the arguments for each sample record; i.e., there are 4 argument averages.

Random or stochastic processes can be classified as being either stationary or non-stationary. A process is stationary if the statistical properties of the process, defined over the ensemble, are independent of the argument(s) (usually time or space). That is, the statistical moments over all realizations (e.g., ensemble or sample averages) are the same for all values of the argument. A non-stationary process is one for which this property is not satisfied. Such processes require special techniques to model their behaviour (see, e.g., Bendat and Piersol [1971] and Priestley [1981]).

In practice, different degrees of stationarity exist. If the complete statistical description of the process (i.e., all possible statistical moments) is independent of the argument, the process is said to be completely stationary. If only the first few moments are independent of the argument, the process is considered to be weakly stationary. Processes with a Gaussian probability distribution are completely described by only the first two moments. In this case, stationarity in only the first two moments infers complete stationarity.

Stationarity can be further classified on the basis of ergodicity. A process is ergodic if the statistical properties taken over the argument (e.g., time averages) are identical to the statistical properties taken over different realizations (e.g., ensemble or sample averages). The assumption of ergodicity allows for a considerable reduction in the number of observations and computations required to determine the statistical properties of a random process. For the sake of simplicity, convenience and, most importantly, costs, most random processes are assumed to be ergodic in practice, even though there may be evidence to the contrary.

When dealing with multidimensional (i.e., multi-argument) spatial processes $\phi(\boldsymbol{x})$ whose arguments ($\boldsymbol{x}$) define location and orientation in space, stationarity is often considered in terms of homogeneity and isotropy. A process is homogeneous if it is invariant with respect to its location in space and isotropic if it is invariant with respect to its orientation [Grafarend, 1976].

Throughout this work all processes are assumed to be stationary and ergodic. Any nonstationarity and nonergodicity is assumed to be explicitly modelled deterministically and is assumed to disappear when the model is selected properly.

2.4 Statistical Moments

The properties of a random processes can be described by the statistical moments of their probability density functions. For a single continuous random process (or variable) $\phi(t)$, at a particular argument t (hereafter called time for convenience), the kth-order moment is given by

$$\mathrm{E}\left[\phi(t)^k\right] = \int_{-\infty}^{\infty} \phi(t)^k \, P(\phi(t)) \, d\phi \,, \quad \forall\ t \in (-\infty,\infty), \tag{2.1}$$

where E[•] is the mathematical expectation operator and $P(\phi(t))$ is the probability density function of the random variable ϕ at time t. The integration is performed over all sample records at time t. This implies that $\phi(t)^k \in (-\infty,\infty)$ must be integrable.

Generally, only the first two moments are useful in practice. The first moment or mean value is the simplest and most common measure of a random process. It provides a measure of the central tendency in the data series. For random processes with discrete sample records $\phi_i(t)$, i=1,...,n, the mean $\mu(t)$ at argument t is defined by

$$\mu(t) = \mathrm{E}\left[\phi(t)\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \phi_i(t) \,, \quad \forall\ t \in (-\infty,\infty), \tag{2.2}$$

where n is the total number of sample records (infinite in the limit).

The second-order moment is a measure of the variation in the random process and is defined by

$$\mathrm{E}\left[\phi(t)^2\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \phi_i(t)^2 , \quad \forall\ t \in (-\infty,\infty). \tag{2.3}$$

The second-order central moment is a measure of the variation about the mean and is also called the variance $\sigma(t)^2$. The discrete form of the variance can be written as

$$\sigma(t)^2 = \mathrm{E}\left[\left(\phi(t)-\mu(t)\right)^2\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(\phi_i(t)-\mu(t)\right)^2 , \quad \forall\ t \in (-\infty,\infty). \tag{2.4}$$

2.5 Covariance and Correlation Functions

Covariance and correlation functions are generic terms for the more general second-order moments which provide a measure of the linear dependence between observations at different values of the argument *t*. Autocovariance and autocorrelation functions represent the linear dependence within a single random process. Cross-covariance and cross-correlation functions represent the linear dependence between a pair of different random processes.

The autocovariance function $C(t,t')$ is defined by

$$C(t,t') = \mathrm{E}\left[\left(\phi(t)-\mu(t)\right)\left(\phi(t')-\mu(t')\right)\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(\phi_i(t)-\mu(t)\right)\left(\phi_i(t')-\mu(t')\right) ,$$

$$\forall\ t,t' \in (-\infty,\infty). \tag{2.5}$$

When the times are the same (i.e., $t=t'$), eqn. (2.5) reduces to the variance $\sigma(t)^2$. The cross-covariance function $C_{\phi\gamma}$ between two random processes $\phi(t)$ and $\gamma(t)$ is defined similarly as

$$C_{\phi\gamma}(t,t') = \mathrm{E}\left[\left(\phi(t)-\mu_\phi(t)\right)\left(\gamma(t')-\mu_\gamma(t')\right)\right] = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \left(\phi_i(t)-\mu_\phi(t)\right)\left(\gamma_i(t')-\mu_\gamma(t')\right), \quad \forall\ t,t' \in (-\infty,\infty). \tag{2.6}$$

The autocorrelation function $R(t,t')$ is defined as the normalized autocovariance function; i.e.,

$$R(t,t') = \frac{C(t,t')}{\sqrt{C(t,t)\ C(t',t')}} = \frac{C(t,t')}{\sigma(t)\ \sigma(t')}, \quad \forall\ t,t' \in (-\infty,\infty). \tag{2.7}$$

Similarly, the cross-correlation function $R_{\phi\gamma}(t,t')$ is the normalized autocorrelation function:

$$R_{\phi\gamma}(t,t') = \frac{C_{\phi\gamma}(t,t')}{\sqrt{C_{\phi\gamma}(t,t)\ C_{\phi\gamma}(t',t')}} = \frac{C_{\phi\gamma}(t,t')}{\sigma_{\phi\gamma}(t)\ \sigma_{\phi\gamma}(t')}, \quad \forall\ t,t' \in (-\infty,\infty). \tag{2.8}$$

The autocorrelation function is limited to the range

$$-1 \le R(t,t') \le 1, \quad \forall\ t,t' \in (-\infty,\infty) \tag{2.9}$$

for all t and t'. When the times t and t' are the same, the autocorrelation function is equal to one. The same holds for the cross-correlation function.

If the random process is stationary, the moments are independent of the value of the argument. Thus, in the above definitions, the expressions are dependent only on the time difference or lag, $\tau=t'-t$. The moments then reduce to the follow forms:

$$\mu = \mathrm{E}\left[\phi(t)\right], \quad \forall\, t \in (-\infty,\infty), \tag{2.10}$$

$$C(\tau) = \mathrm{E}\left[\left(\phi(t)-\mu\right)\left(\phi(t+\tau)-\mu\right)\right], \quad \forall\, t \in (-\infty,\infty), \tag{2.11}$$

$$R(\tau) = \frac{C(\tau)}{C(0)}, \quad \forall\, t \in (-\infty,\infty). \tag{2.12}$$

Similar expressions to eqns. (2.11) and (2.12) can be written for the cross-covariance and cross-correlation functions.

The following two properties of these functions are consequences of the assumption of stationarity:

1. The auto/cross-covariance and auto/cross-correlation functions are even functions of τ; i.e.,

$$C(\tau) = C(-\tau), \tag{2.13}$$

$$R(\tau) = R(-\tau). \tag{2.14}$$

2. At lag $\tau=0$ the autocovariance function is positive and the autocorrelation function is equal to one; i.e.,

$$C(0) > 0, \tag{2.15}$$

$$R(0) = 1 . \tag{2.16}$$

Ergodicity is probably the most important and often used assumption in practical data analysis applications, even when the process is known to be non-ergodic or even non-stationary. This is done to simplify the data acquisition and handling procedures. Stochastic processes are ergodic if their sample moments (e.g., mean, autocovariance, etc.) can be determined from averaging over the argument (e.g., time) instead of averaging over the sample records (see Figure 2.1). For the mean and autocovariance function,

$$\mu = \mathrm{E}\left[\phi(t)\right] = \int_{-\infty}^{\infty} \phi(t)\, P(\phi)\, dt \ , \tag{2.17}$$

$$C(\tau) = \mathrm{E}\left[\left(\phi(t)-\mu\right)\left(\phi(t+\tau)-\mu\right)\right] = \int_{-\infty}^{\infty} \left(\phi(t)-\mu\right)\left(\phi(t+\tau)-\mu\right) P(\phi)\, dt \ . \tag{2.18}$$

The discrete forms of these expressions (for discrete random processes) are given by:

$$\overline{\phi} = \frac{1}{n}\sum_{i=1}^{n} \phi(t_i) \ , \tag{2.19}$$

$$C(\tau_k) = \frac{1}{n-k}\sum_{i=1}^{n-k} \left(\phi(t_i)-\mu\right)\left(\phi(t_i+\tau_k)-\mu\right) , \tag{2.20}$$

where lag $\tau_k = k\,\Delta t$ and Δt is the sampling interval. This expression gives an unbiased estimate of the autocovariance function. Although unbiasedness is desirable, this function is not positive definite. Constructing covariance matrices from this leads to singular matrices. It also exhibits so-called "wild" behaviour at large lags. For these reasons, the biased estimate is recommended by Bendat and Piersol [1971, pp. 312-314] and Priestley

[1981, pp. 323-324], where the denominator $n–k$ in eqn. (2.20) is replaced by the constant n. This results in a function that tapers off as the lag increases. An example of this is given in the numerical simulations in Chapter 7.

Similar expressions can also be written for the cross-covariance functions. Note that the integrations and summations are performed over the argument t rather than over the sample records; i.e., under the assumption of ergodicity the moments can be computed from a single sample.

2.6 Decomposition of the Observable

In the real world, processes cannot be modelled as purely deterministic or stochastic. Instead, one is faced with a mixture of both. Clearly, there are many factors which prevent us to model in a deterministic way. Most are due to either measurement errors or systems that are simply too complex to be modelled entirely deterministically. According to Priestley [1981, p. 14] "almost all quantitative phenomena occurring in science should be treated as random processes as opposed to deterministic functions."

The expected value of a random process may be computed from some deterministic model describing the expected behaviour of the series. However, this model will probably not describe the series exactly as mentioned above. A stochastic model may then be used to account for the resulting lack of fit. It is therefore convenient to decompose the observable $\phi(t)$ into a deterministic or trend component $\hat{\phi}(t)$ and a random or stochastic component $e(t)$; i.e.,

$$\phi(t) = \hat{\phi}(t) + e(t) , \quad \forall\, t \in (-\infty,\infty). \tag{2.21}$$

The random component $e(t)$ may also be decomposed into two components:

$$e(t) = s(t) + \varepsilon(t), \quad \forall\, t \in (-\infty,\infty). \tag{2.22}$$

where $s(t)$ is a statistically dependent (correlated) component and $\varepsilon(t)$ is a statistically independent (uncorrelated) component. The observable may then be represented in the form

$$\phi(t) = \hat{\phi}_i + s(t) + \varepsilon(t), \quad \forall\, t \in (-\infty,\infty). \tag{2.23}$$

The statistically dependent component is often due to effects neglected or incompletely accounted for in the deterministic model defined by $\hat{\phi}_i$. Both random components are assumed to have a zero mean. This is enforced when the trend component is estimated by least squares. However, due to the statistical dependence, there is a correlation among the $s(t)$ components. This statistically dependent component can be thought of as the residual deterministic part remaining after removing the postulated deterministic model. Thus, this component is often referred to as a systematic error or systematic effect.

Chapter 3
The Fourier Transform and Spectrum

3.1 Fourier Series and Integrals

It is well known in mathematics that any continuous, periodic function can be represented by an infinite series of trigonometric functions, called a Fourier series. If $\phi(t)$ is a function of period T, it can then be expressed in the form

$$\phi(t) = \frac{1}{2}a_0 + \sum_{i=0}^{\infty} (a_i \cos 2\pi f_i t + b_i \sin 2\pi f_i t \quad) , \tag{3.1}$$

where a_i and b_i are the Fourier coefficients corresponding to frequency f_i. The frequency f_i can also be expressed in terms of the natural or fundamental frequency f_o as $f_i = if_o$, where $f_0 = 1/T$. Note that if angular frequencies (ω) are to be used, ω_i (in radians per unit of t) should be substituted for $2\pi f_i$.

Using the fact that the cosine and sine functions form an orthogonal basis over the interval $(-T/2, T/2)$, the Fourier coefficients for all $i = 0,...,\infty$ are given by [Priestley, 1981, p. 194]

$$a_i = \frac{2}{T} \int_{-T/2}^{T/2} \phi(t) \cos 2\pi f_i t \, dt , \tag{3.2}$$

$$b_i = \frac{2}{T} \int_{-T/2}^{T/2} \phi(t) \sin 2\pi f_i t \, dt . \tag{3.3}$$

If $\phi(t)$ is an even function, i.e., $\phi(t) = \phi(-t)$, the b_i coefficients are zero. When $\phi(t)$ is an odd function, i.e., $\phi(t) = -\phi(-t)$, the a_i coefficients are zero. Note that writing the constant term in eqn. (3.1) as $\frac{1}{2}a_0$ rather than as a_0 makes the expressions (eqns. (3.2) and (3.3)) for the coefficients valid even for i=0.

For non-periodic functions, there is no such Fourier series representation. However, according to Priestley [1981, pp.198-200], a new periodic function may be defined which is the same as the non-periodic one over a finite interval, say, $(-T/2, T/2)$ but repeats itself and is thus periodic outside this interval. This new function will have a period T and can now be represented as a Fourier series. By letting $T \to \infty$, the discrete set of frequencies in the Fourier series becomes a continuous set of frequencies; i.e., an integral. The non-periodic function can then be represented by the so-called Fourier integral which has the form [Priestley, 1981, pp. 198-199]

$$\phi(t) = \int_{-\infty}^{\infty} (a(f) \cos 2\pi ft + b(f) \sin 2\pi ft)\, df , \quad \forall\ t , \tag{3.4}$$

where the Fourier coefficients over the continuous range of frequencies are defined by

$$a(f) = \frac{2}{T} \int_{-\infty}^{\infty} \phi(t) \cos 2\pi ft\, dt , \quad \forall\ f , \tag{3.5}$$

$$b(f) = \frac{2}{T} \int_{-\infty}^{\infty} \phi(t) \sin 2\pi ft\, dt , \quad \forall\ f . \tag{3.6}$$

This representation of a non-periodic function in terms of a continuous set of frequencies holds only when the function is absolutely integrable over the infinite interval $(-\infty, \infty)$ [Priestley, 1981, p. 200]; i.e.,

$$\int_{-\infty}^{\infty} |\phi(t)|\, dt \;<\; \infty\,, \quad \forall\, t\,. \tag{3.7}$$

This happens when $\phi(t)$ decays to zero as t goes to infinity.

So far only periodic and non-periodic deterministic functions have been considered. However, in practice one usually deals with random or stochastic functions (processes) where the application of the above representations is not so apparent. Clearly, stochastic functions may not necessarily be periodic and thus they cannot be represented by Fourier series. Furthermore, stochastic functions are not absolutely integrable since, by the definition of stationarity, they do not decay to zero at infinity. It would then appear that we cannot represent them as Fourier integrals either. Nevertheless, according to Priestley [1981, p. 207] it is possible to circumvent this problem by simply truncating the stochastic process at, say, $-T/2$ and $T/2$ as done for non-periodic functions. Outside this interval the function is defined to be zero, thereby satisfying the absolutely integrable condition. As long as the stochastic function is continuous, it can be represented by the Fourier integral as in eqn. (3.4) but with coefficients defined by finite Fourier integrals using integration limits $(T/2,-T/2)$ instead of $(-\infty,\infty)$; i.e.[Priestley, 1981, p. 207],

$$a(f) \;=\; \frac{2}{T} \int_{-T/2}^{T/2} \phi(t) \cos 2\pi f t\, dt\,, \qquad \forall\, f\,, \tag{3.8}$$

$$b(f) \;=\; \frac{2}{T} \int_{-T/2}^{T/2} \phi(t) \sin 2\pi f t\, dt\,, \qquad \forall\, f\,. \tag{3.9}$$

Unfortunately, we cannot take the limit $T\to\infty$ as before since, by the property of stationarity, the above integrals would not be finite.

Although all of the expressions for the Fourier series and integrals were given in terms of trigonometric functions, it is more common to use complex notation for a more compact representation of the series and integrals. Assigning the cosine term to the real component and the sine term to the imaginary component, each trigonometric term can be replaced by a complex exponential function using Euler's formula [Bronshtein and Semendyayev, 1985, p. 474]

$$\cos 2\pi ft + j \sin 2\pi ft = e^{j2\pi ft}, \tag{3.10}$$

where $j=\sqrt{-1}$ is the imaginary unit.

Using this notation, the Fourier series in eqn. (3.1) can be re-written as [Priestley, 1981, p. 199]

$$\phi(t) = \sum_{k=-\infty}^{\infty} A_k \, e^{j2\pi f_k t} \quad , \tag{3.11}$$

where

$$A_k = \begin{cases} \frac{1}{2}(a_{|k|} - j\, b_{|k|}) \, , & k > 1 \\ \frac{1}{2} a_0 \, , & k = 0 \\ \frac{1}{2}(a_k + j b_k) \, , & k \leq -1 \end{cases} \tag{3.12}$$

Substituting for a_k and b_k, using eqns. (3.2), (3.3) and (3.10),

$$A_k = \frac{1}{T} \int_{-T/2}^{T/2} \phi(t) \, e^{-j2\pi f_k t} \, dt \quad . \tag{3.13}$$

Putting this in the Fourier series in the continuous form of eqn. (3.11) and letting $T \to \infty$ gives the so-called Fourier integral over a continuous range of observations; i.e.,

$$\phi(t) = \int_{-\infty}^{\infty} F(f)\, e^{j2\pi ft} df \,, \qquad (3.1614$$

where

$$F(f) = \begin{cases} \int_{-\infty}^{\infty} \phi(t)\, e^{-j2\pi ft} dt & \text{for non-periodic functions} \\ \int_{-T/2}^{T/2} \phi(t)\, e^{-j2\pi ft} dt & \text{for stochastic functions} \end{cases} . \qquad (3.15)$$

3.2 Fourier Transform

Given the Fourier integral representation of a non-periodic or stochastic function, the transformation from $\phi(t)$ to $F(f)$ in eqn. (3.15) is called the (direct) Fourier transform, or the finite Fourier transform if dealing with stochastic functions. The transformation from $F(f)$ to $\phi(t)$ in eqn. (3.14) is called the inverse Fourier transform. $\phi(t)$ and $F(f)$ are referred to as a Fourier transform pair, denoted by $\phi(t) \Leftrightarrow F(f)$. Note that the complex conjugate form is used in the direct transform and not in the inverse transform. In some texts (e.g., Press et al. [1986]), the conjugate form is used in the inverse transform and not in the direct transform.

In practice one rarely deals with continuous stochastic processes of infinite length but rather with actual discrete processes or discretely sampled data from continuous processes of finite length. Although such discrete samples are often evenly spaced in time (or any other argument), this may not always be the case. Nevertheless, the application of

traditional Fourier transform techniques requires the processes to be discretely and evenly sampled. This is because the trigonometric functions are not orthogonal over an unevenly spaced domain.

For a discretely and evenly sampled stochastic process or data series $\{\phi(t_i), i=0,1,...,n-1\}$, the discrete Fourier transform is obtained by approximating the Fourier integral in eqn. (3.15) with a summation; i.e.,

$$F(f_k) = \int_{-\infty}^{\infty} \phi(t)\, e^{-j2\pi f_k t}\, dt \cong \Delta t \sum_{i=0}^{n-1} \phi(t_i)\, e^{-j2\pi f_k t_i}, \qquad (3.16)$$

where n is the number of "observations" (samples), Δt is the sampling interval and f_k is one of the frequencies belonging to the set of frequencies estimable from the discrete process (see below). Note also that the summation index extends from 0 to $n-1$ (instead of 1 to n) following the usual convention. If $T=n\Delta t$ is the length of the data series, the discrete set of frequencies are given by

$$f_k = \frac{k}{T} = \frac{k}{n\Delta t} = k f_o, \quad \forall\, k=-\frac{n}{2} \ldots \frac{n}{2}, \qquad (3.17)$$

where $f_o=1/T=1/(n\Delta t)$ is the fundamental frequency. To make matters simpler, n is assumed to be even (the data series is truncated to an even number of points). This set of integer multiples of the fundamental frequency will be simply called "Fourier" frequencies here because they are always used in the Fourier transform and Fourier spectrum.

By convention, it is only the final summation in eqn. (3.16) (without the Δt in front) that is commonly referred to as the discrete Fourier transform, denoted by F_k for frequency f_k. The discrete Fourier transform (DFT) is then defined by

$$F_k = \sum_{i=0}^{n-1} \phi(t_i)\, e^{-j2\pi f_k t_i} , \quad \forall\ k . \tag{3.18}$$

The inverse discrete Fourier transform is obtained similarly by approximating the integral in eqn. (3.14) with a summation and substituting for the discrete Fourier transform. This gives

$$\phi(t_i) = \frac{1}{n} \sum_{k=0}^{n-1} F_k\, e^{j2\pi f_k t_i} , \quad \forall\ i . \tag{3.19}$$

The discrete sampling of a stochastic process has an important consequence known as the aliasing effect, whereby some high frequency information will be lost or, more precisely, hidden (aliased) in the lower frequencies. This can be seen by examining the exponential term in eqns. (3.17) and (3.18) as a function of the discretely sampled process $\phi(t_i)$, $i = -\infty,...,\infty$, where $t_i = i\,\Delta t$ and Δt is the sampling interval. Re-writing the exponential function as

$$e^{j2\pi f_k t_i} = \cos 2\pi f_k t_i + j \sin 2\pi f_k t_i , \tag{3.20}$$

the effect of discrete sampling on each sine and cosine term can be seen. For example, substituting $i\Delta t$ for t_i in the cosine term gives

$$\cos 2\pi f_k t_i = \cos 2\pi i f \Delta t \tag{3.21}$$

The same can be written for a new frequency f+Δf;

$$\cos 2\pi (f+\Delta f) t_i = \cos(2\pi i f \Delta t + 2\pi i \Delta f \Delta t \tag{3.22}$$

These two cosine terms are equivalent only if $2\pi i \Delta f \Delta t$ is an integer multiple of π. This occurs when Δt is an integer multiple of $f_N = \frac{1}{2\Delta t}$, called the Nyquist or critical frequency. Thus the cosine terms will look the same for frequencies $f \pm \frac{k}{2\Delta t}$, $f \pm \frac{2k}{2\Delta t}$, ... All appear to have frequency f. The same holds for the sine terms. All frequencies outside of the Nyquist frequency range $(-f_N, f_N)$ will be aliased to (i.e., moved to and superimposed on) frequencies in this range. If possible, Δt should be chosen small enough to avoid aliasing. However, this requires a knowledge of the upper frequency limit of the information contained in the process being sampled or, at least, knowledge that only negligible information exists beyond the Nyquist frequency and our willingness to neglect this information.

There are some special properties of Fourier transforms that are of particular importance. These are summarized as follows (* indicates the complex conjugate operator):

$\phi(t)$ is real	$F(-f) = F(f)^*$,	(3.23)
$\phi(t)$ is imaginary	$F(-f) = -F(f)^*$,	(3.24)
$\phi(t)$ is even	$F(-f) = F(f)$ (i.e., $F(f)$ is even) ,	(3.25)
$\phi(t)$ is odd	$F(-f) = -F(f)$ (i.e., $F(f)$ is odd) .	(3.26)

Note that when dealing with real functions, the series of trigonometric terms of cosines and sines reduce to a series of only cosine terms; i.e., by eqn. (3.19) the sine terms are all zero. In this case the Fourier transform reduces to the so-called cosine transform.

The following are some other properties of the Fourier transform (from Press et al. [1992, p. 491]). Recall that $\phi(t) \Leftrightarrow F(f)$ indicates that $\phi(t)$ and $F(f)$ are a Fourier transform pair.

Time shifting	$\phi(t-t_o)$	$\Leftrightarrow$	$F(f)\, e^{j2\pi f t_o}$,	(3.27)
Time scaling	$\phi(at)$	$\Leftrightarrow$	$\frac{1}{\lvert a \rvert} F\left(\frac{f}{a}\right)$,	(3.28)
Frequency shifting	$\phi(t)\, e^{-j2\pi f_o t}$	$\Leftrightarrow$	$F(f-f_o)$,	(3.29)
Frequency scaling	$\frac{1}{\lvert b \rvert} \phi\left(\frac{t}{b}\right)$	$\Leftrightarrow$	$F(bf)$.	(3.30)

3.3 Fourier Spectrum

The representation of functions in terms of Fourier series has a special physical interpretation in terms of power (cf. Priestley [1981, p. 194-195] and Press et al. [1992, p. 492]). Consider an absolutely integrable non-periodic function f(t). The total "power" of $\phi(t)$ is customarily defined by

$$\text{Total power} = \int_{-\infty}^{\infty} \phi(t)^2 \, dt \quad . \tag{3.31}$$

Substituting the inverse Fourier transform in eqn. (3.14) for one of the $\phi(t)$ gives

$$\int_{-\infty}^{\infty} \phi(t)^2 \, dt = \int_{-\infty}^{\infty} \phi(t) \left[\int_{-\infty}^{\infty} F(f)\, e^{j2\pi f t} \, df \right] dt \quad . \tag{3.32}$$

Interchanging the order of the integrals and substituting for the direct Fourier transforms results in

$$\int_{-\infty}^{\infty} \phi(t)^2 \, dt = \int_{-\infty}^{\infty} F(f) \left[\int_{-\infty}^{\infty} \phi(t)\, e^{j2\pi f t} \, dt \right] df$$

$$= \int_{-\infty}^{\infty} F(f)\ F^*(f)\ df \qquad (3.33)$$

$$= \int_{-\infty}^{\infty} |F(f)|^2\ df\ ,$$

where $F^*(f)$ denotes the complex conjugate of $F(f)$. The total power can therefore be expressed either in terms of the integral of the original function or its Fourier transform; i.e.,

$$\text{Total power} = \int_{-\infty}^{\infty} \phi(t)^2\ dt = \int_{-\infty}^{\infty} |F(f)|^2\ df\ . \qquad (3.34)$$

This is known as Parseval's relation [Jenkins and Watts, 1969, p. 25; Priestley, 1981, p. 201] or Parseval's theorem [Press et al., 1992, p. 492]. Note that the total power is equal to n times the variance σ^2.

It can be seen from eqn. (3.34) that the total power is divided among a continuous set of frequencies in the representative Fourier integral. Each term $|F(f)|^2 df$ represents the contribution to the total power in $\phi(t)$ produced by the components with frequencies in the interval $(f, f+df)$. The so-called power spectral density $s(f)$ for frequency f is thus defined by

$$s(f) = |F(f)|^2\ . \qquad (3.35)$$

The plot of $s(f)$ versus frequency f is also called the power spectrum, or simply the spectrum. Theoretical power spectral density functions for some special functions are illustrated in Figure 3.1.

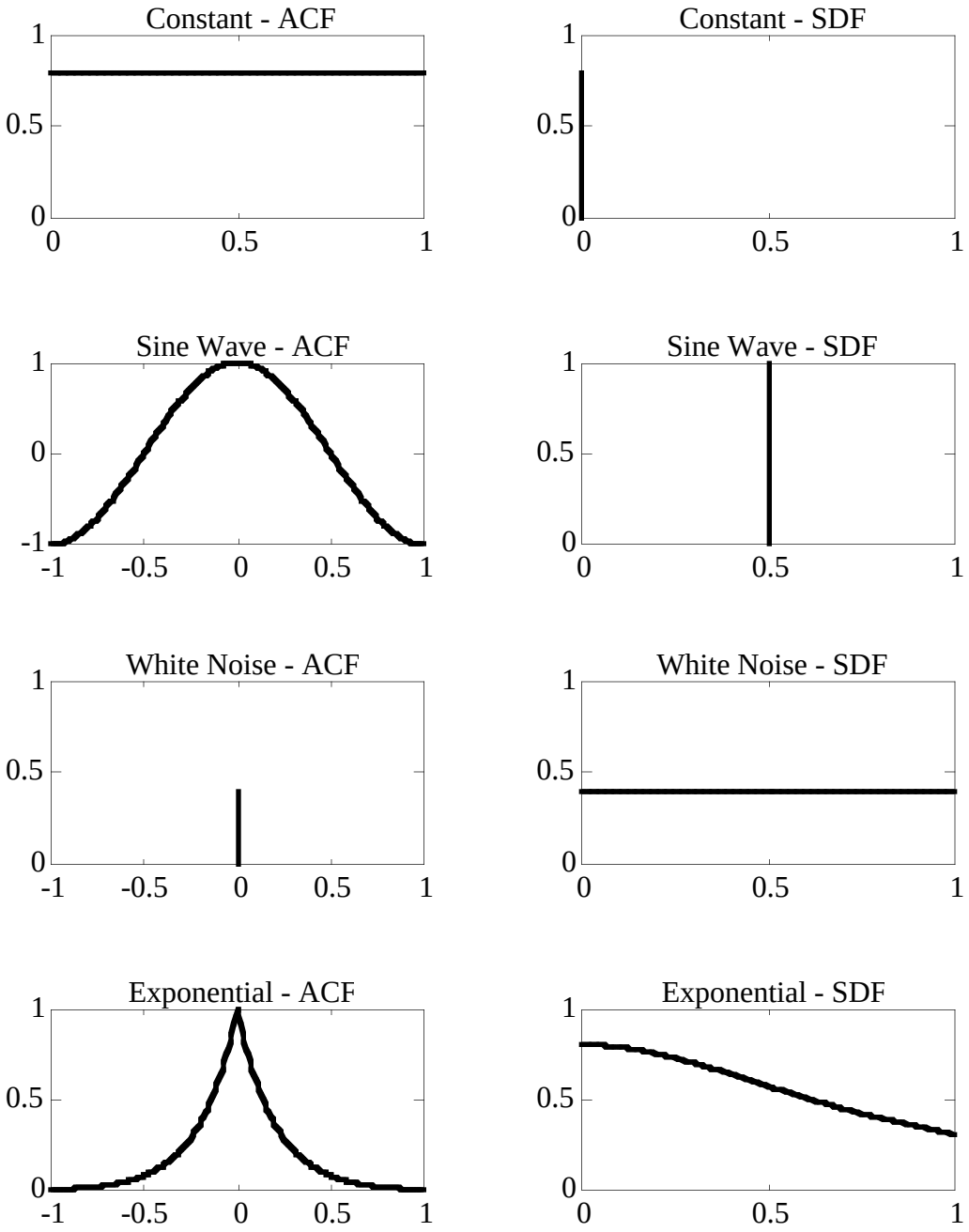

Figure 3.1: Autocorrelation functions (ACF) and power spectral density functions (SDF) for some special functions.

For periodic functions, the total power over the entire interval $(-\infty, \infty)$ is infinite [Priestley, 1981, pp. 195,205]. Although it is only needed to describe the power over the finite interval $(-T/2, T/2)$ in order to characterize it for the entire infinite interval, it is usually more convenient to use the total power per unit of time over the finite interval. This is obtained by dividing the total power by the period T; i.e.,

$$\begin{aligned} \text{Total power per unit of time}\left(-\frac{T}{2}, \frac{T}{2}\right) &= \frac{\text{Total power}\left(-\frac{T}{2}, \frac{T}{2}\right)}{T} \\ &= \frac{1}{T}\int_{-T/2}^{T/2} \phi(t)^2\, dt \\ &= \frac{1}{T}\sum_{k=-\infty}^{\infty} |F(f_k)|^2 \\ &= \sum_{k=-\infty}^{\infty} s(f_k) \quad . \end{aligned} \tag{3.36}$$

The total power over $(-T/2, T/2)$ is divided among the infinite set of discrete frequencies in the representative Fourier series. The contribution $s(f_k)$ to the total power per unit of time of each "Fourier" frequency $f_k = \frac{k}{T}$ is called the spectral value for frequency f_k and is defined by

$$s(f_k) = \frac{1}{T}|F(f_k)|^2 \, . \tag{3.37}$$

Similarly, for stationary stochastic functions (random processes), the total power is also infinite by the definition of stationarity (i.e., a steady state process from $t = -\infty$ to ∞ requires infinite energy or power). Using again the truncation approach, stochastic processes can also be represented by finite Fourier integrals in the finite interval $(-T/2, T/2)$. The total power in this finite interval will then be finite.

For both non-periodic and stochastic functions over a finite interval *(–T/2, T/2)*, it is generally more convenient to also use the power per unit of time. As for periodic functions, the power per unit time is obtained by dividing the total power over the finite interval by the length of the interval; i.e.,

$$\begin{aligned}\text{Total power per unit of time}\left(-\frac{T}{2}\,,\ \frac{T}{2}\right) &= \frac{\text{Total power}\left(-\frac{T}{2}\,,\ \frac{T}{2}\right)}{T} \\ &= \frac{1}{T}\int_{-T/2}^{T/2}\phi(t)^2\ dt \\ &= \frac{1}{T}\int_{-\infty}^{\infty}|F(f)|^2\ df \\ &= \int_{-\infty}^{\infty}s(f)\ df\quad .\end{aligned} \tag{3.38}$$

Here *s(f)* represents the power spectral density function. For a process of finite length *T* it is defined by

$$s(f) = \frac{1}{T}|F(f)|^2\ . \tag{3.39}$$

The spectrum defined above is a function of both positive and negative frequencies and is called a "two-sided" spectral density function. However, one does not usually distinguish between positive and negative frequencies. Moreover, when $\phi(t)$ is real, the Fourier transform is an even function; i.e., *F(f)=F(–f)*. It is therefore customary to express the spectrum as a function of only positive frequencies. Such a spectrum is called a "one-sided" spectral density function. Because the total power in the process must remain the same, the spectral values for the one-sided spectrum are defined as

$$s(f) = |F(f)|^2 + |F(-f)|^2\ ,\quad \forall\ \ 0 \le f < \infty\ . \tag{3.40}$$

For real $\phi(t)$, $F(f) = F(-f)$ and

$$s(f) = 2|F(f)|^2 , \quad \forall \ 0 \le f < \infty . \tag{3.41}$$

Hereafter, the one-side spectral density function will be used since only real $\phi(t)$ will be considered.

It is also convenient to normalize the spectral values so that they express the percentage of the total power or variation in the process contributed by each frequency. The normalized spectral values $\tilde{s}(f)$ are given by

$$\tilde{s}(f) = \frac{s(f)}{\sum_{k=0}^{n-1} s(f)} . \tag{3.42}$$

A couple of important properties for power spectra are obtained from the properties of Fourier transforms. One of the most important is the invariance of the spectrum to time shifting. Given a process $\phi(t)$ shifted by t_o, the new Fourier transform is

$$\begin{aligned} F'(f) &= \int_{-\infty}^{\infty} \phi(t+t_o)\, e^{-j2\pi f(t+t_o)}\, dt \\ &= e^{-j2\pi f t_o} \int_{-\infty}^{\infty} \phi(t+t_o)\, e^{-j2\pi f t}\, dt \\ &= e^{-j2\pi f t_o}\, F(f) . \end{aligned} \tag{3.43}$$

The spectrum $s'(f)$ for this process is then given by

$$s'(f) = F'(f)\, F'^*(f) = F(f)\, F^*(f) = s(f) , \tag{3.44}$$

which is identical to the spectrum of the original series. Note that the constant exponential term in eqn. (3.43) cancels with its complex conjugate.

The spectrum is not invariant with respect to time scaling, however. Intuitively, expanding time effectively results in shrinking frequencies, and vice versa. The relation between two spectra with different time scales can be obtained from eqn. (3.28). Given a function $\phi(at)$ which is scaled in time by a factor a, the new Fourier transform $F'(f')$ is, by eqn. (3.28),

$$F'(f') = \int_{-\infty}^{\infty} \phi(at)\, e^{-j2\pi fat}\, dt = \frac{1}{|a|} F\left(\frac{f}{a}\right) \tag{3.45}$$

where $f' = \frac{f}{a}$. The spectrum is then given by

$$\begin{aligned} s'(f') &= F'(f')\ F'^*(f') \\ &= \frac{1}{|a|^2} F\left(\frac{f}{a}\right) F^*\left(\frac{f}{a}\right) \\ &= \frac{1}{|a|^2} s\left(\frac{f}{a}\right). \end{aligned} \tag{3.46}$$

This results in both a scaling of the frequencies as well as the Fourier transform and spectrum.

For discretely sampled, infinite length processes, the Fourier transform and spectral values are defined only for the discrete set of "Fourier" frequencies $f_k = \frac{k}{n\Delta t}$, $k = -n/2,...,n/2$ (see discussion of discrete Fourier transform). The discrete form of Parseval's relation for the total power in a process is obtained in the same way as for the Fourier integral except that the discrete Fourier transform is used instead. Following the same substitution and reordering of summations in eqn. (3.33) gives [Press et al., 1986, p. 390]

$$\text{Total power} = \sum_{k=0}^{n-1} \phi(t)^2 = \frac{1}{n} \sum_{k=0}^{n-1} |F(f_k)|^2 = \sum_{k=0}^{n-1} s(f_k) \quad . \tag{3.47}$$

The individual spectral values $s(f_k)$ for the power spectral density function are then given by

$$s(f_k) = \frac{1}{n} |F(f_k)|^2 \, . \tag{3.48}$$

The normalized spectral values are obtained by dividing by the total power as in eqn. (3.42). For the discrete case, this gives

$$\tilde{s}(f_k) = \frac{s(f_k)}{\sum_{k=0}^{n-1} s(f_k)} \, . \tag{3.49}$$

Realizing that the variance σ^2 is the total power divided by n ($\sigma^2 = \frac{1}{n} \sum_{i=0}^{n-1} \phi(t_i)^2$), the normalized spectral values can also be written as

$$\tilde{s}(f_k) = \frac{s(f_k)}{\sum_{k=0}^{n-1} s(f_k)} = \frac{s(f_k)}{\sum_{k=0}^{n-1} \phi(t_k)^2} = \frac{|F(f)|^2}{n^2 \sigma^2} \, . \tag{3.50}$$

Sample estimates of the spectrum can be obtained by evaluating the discrete Fourier transform for frequencies $f_k = 0, \ldots, \frac{1}{2\Delta t}$ and computing the spectral values $s(f_k)$ using eqns. (3.48) or (3.49). It is important to note for later that this is equivalent to (i) evaluating the Fourier coefficients a_k and b_k for the discrete frequencies f_k using least squares estimation and (ii) computing the (amplitude) spectrum from $(a_k^2 + b_k^2)$. For real-valued functions, only positive frequencies need be considered because the negative

frequency part of the spectrum is the mirror image of the positive part. However, the negative frequencies will be aliased as positive ones and, combined with the (identical) positive ones, will result in spectral values twice those computed using eqn. (3.48), except for the zero frequency. This gives the one-side spectrum rather than a two-sided spectrum. The spectrum computed in this manner is generally referred to as the periodogram [Priestley, 1981, p. 394; Press et al, 1986, p. 421] and forms the basis of the least squares spectrum.

3.4 Convolution and Correlation

Another application of Fourier transforms is in the concept of convolution and correlation. Given two functions $\phi(t)$ and $\gamma(t)$ and their Fourier transforms $F(f)$ and $G(f)$, we can combine these two functions together in what is called a convolution. For the continuous case the convolution of $\phi(t)$ and $\gamma(t)$, denoted $\phi(t)*\gamma(t)$, is defined by [Bronshtein and Semendyayev, 1985, p. 582]

$$\phi(t) * \gamma(t) = \int_{-\infty}^{\infty} \phi(\tau)\, \gamma(t-\tau)\, d\tau \ , \quad \forall\ t \in (-\infty,\infty), \tag{3.51}$$

where τ is thought of as an argument (time) difference or lag. The convolution theorem then states that the Fourier transform of the convolution of two functions is equal to the product of the Fourier transforms of the individual functions [ibid, p. 582]; i.e.,

$$\phi(t) * \gamma(t) \ \Leftrightarrow\ F(f)\, G(f) \ . \tag{3.52}$$

where the symbol $\Leftrightarrow$ again signifies that the functions on either side are Fourier transform pairs. The Fourier transform is used to go from left to right while the inverse transform is used from right to left.

For discretely and evenly sampled processes $\phi(t_i)$ and $\gamma(t_i)$, $i = -n/2,...,n/2$, the discrete convolution is defined by

$$\phi(t_i) * \gamma(t_i) \;=\; \sum_{k=-n/2+1}^{n/2} \phi(t_i)\, \gamma(t_{i-k}) \;, \qquad \forall\; i \in\; (0,\infty), \tag{3.53}$$

where the lags $t_i - t_{i-k}$ are evenly spaced. The discrete version of the convolution theorem is then

$$\phi(t_i) * \gamma(t_i) \;\Leftrightarrow\; F_k\, G_k \,. \tag{3.54}$$

for frequencies f_k, $k = 0,...,n-1$.

Closely related to the convolution theorem in eqn. (3.51) is the correlation theorem. It can be shown that the product of a Fourier transform with the complex conjugate of another Fourier transform can be reduced to the form [Priestley, 1981, p. 211]

$$F(f)\, G^*(f) \;=\; \int_{-\infty}^{\infty} K(\tau)\, e^{-j2\pi f\tau}\, d\tau \,, \qquad \forall\; f \in\; (-\infty,\infty), \tag{3.55}$$

where K is called the kernel:

$$K(\tau) \;=\; \int_{-\infty}^{\infty} f(t)\, g(t-\tau)\, dt \,, \qquad \forall\; \tau \in\; (-\infty,\infty). \tag{3.56}$$

In the context of spectral analysis, the kernel $K(\tau)$ represents the cross-covariance function. Multiplying the Fourier transform by its own complex conjugate gives the autocovariance function $C(\tau)$ (cf. Section 2.5) as the kernel; i.e.,

$$F(f)\,F^*(f) = \int_{-\infty}^{\infty} C(\tau)\, e^{-j2\pi f\tau}\, d\tau\,, \qquad \forall\ f \in (-\infty,\infty), \tag{3.57}$$

Realizing that this multiplication gives the spectral value for frequency f, the covariance and the spectrum function can be expressed as a Fourier transform pair; i.e.,

$$C(t) \quad \Leftrightarrow \quad s(f)\,. \tag{3.58}$$

This is known as the Wiener-Khinchin theorem. Furthermore, the normalized spectrum $\tilde{s}(f)$ is the transform pair of the autocorrelation function $R(t)$ (cf. Section 2.5) so that

$$R(t) \quad \Leftrightarrow \quad \tilde{s}(f)\,. \tag{3.59}$$

When computing the convolution of two functions care must be exercised to avoid so-called "end effects" or "wrap around effects" caused by assuming the functions to be periodic. For example, when convolving a function with itself (i.e., autocorrelation), data from the end of the series are effectively wrapped around to the beginning of the series thereby forming a periodic function with period T. This can have adverse effects but can be prevented by simply "padding" the data series with enough zeros to avoid any overlap of original data. To estimate all possible frequencies up to the Nyquist frequency (defined in Section 3.2), a data series of n points must be padded with n zeros to completely avoid any wrap around affect. There is a trade off when doing this, however; the more zeros

appended to the series, the greater the errors in the sample estimates of the Fourier transforms. See Press et al. [1992, pp. 533] for more information on end effects.

These indirect expressions in terms of the spectrum are often used as the basis for the efficient computation of autocovariance and autocorrelation functions using the FFT. It will also be used as the basis for developing autocovariance functions for unevenly spaced data to provide objective a priori estimates of covariances and weights that account for residual systematic effects in least squares modelling. However, it must be realized that this indirect procedure gives the biased estimate of the autocovariance and autocorrelation functions [Bendat and Piersol, 1971, pp. 312-314; Priestley, 1981, pp. 323-324].

3.5 Fast Fourier Transform

Any discussion of the Fourier transform would not be complete without mentioning the so-called Fast Fourier Transform (FFT). Although the term is often used synonymously with the Fourier transform itself, it is really only a numerical algorithm used to compute the discrete Fourier transform (DFT) in an extremely efficient manner. The algorithm, popularized by Cooley and Tukey [1965], revolutionized the way in which the DFT had been used. Up to that time the DFT was restricted to only small data sets. With the advent of the FFT algorithm, however, it was quickly employed in a multitude of applications.

The basic idea behind the FFT is a bisection and recombination process. First the data is repeatedly bisected into pairs of points by recursively dividing the data into odd and even numbered points. The Fourier transforms are then computed for each of these pairs of points and subsequently recombined to form the Fourier transform of the entire data series. Because the Fourier transform of a pair of data points is a trivial and very fast computation (no multiplications are needed), the algorithm results in a dramatic increase in computational efficiency, especially for large data sets. The number of (complex)

multiplications involved in the direct evaluation of the discrete Fourier transform is of the order of n^2 whereas the number of such operations in the FFT algorithm (in the recombination of the individual transforms) is only of the order of $n \log_2 n$ [Press et al., 1992]. This general strategy was first used by Gauss to reduce the computational effort in determining planetary orbits and also derived by as many as a dozen others since (see Brigham [1974] and Bracewell [1989] for more information).

The main limitation of both the discrete Fourier transform and its FFT algorithm is that the data must be equally spaced. The expression for the Fourier coefficients, and thus the Fourier transform, are valid only for equally spaced data. Moreover, the FFT algorithm uses certain properties of the sine and cosine functions for evenly spaced data to reduce the number of terms that need to be evaluated. For the investigation of systematic effects which can be functions of many different kinds of arguments that are usually very irregularly spaced, this precludes the use of the FFT, at least in the computation of the discrete Fourier transform. A similar problem also arises when there are large gaps in an otherwise equally spaced data series.

To circumvent the problem of unevenly spaced or "gappy" data, interpolation schemes are sometimes used where the original data are interpolated to give an evenly spaced series. This then allows one to use traditional techniques such as the FFT. However, the accuracy of the interpolating function to represent the original data series depends on the form of the interpolating function, the smoothness of the original data series and the presence of large gaps in the data. This presents a dilemma since in order to properly interpolate the data we must have a good knowledge of their behaviour, but the lack of this knowledge is usually the reason for computing FFTs in the first place. Another problem with interpolation is that in the presence of large gaps, they often result in disastrous results.

A second limitation of the FFT is that the number of data points to be transformed must be a power of 2 for the FFT to be most efficient. Alternate and mixed radix

formulations of the FFT also exist but they are much less efficient. The conventional method of dealing with a number of points that are not a power of two is to again "pad" the data series with enough zeros to obtain the required number of points for the FFT. This clearly inflates the number of points to process thereby increasing not only processing time but also storage requirements. It is most inefficient when dealing with large data sets. In these cases, one usually only takes the first power of two number of points and omits the rest. More importantly, zero padding also increases the error in the FFT with respect to the continuous Fourier transform.

One more limitation of the FFT is that it is restricted to only the set of "Fourier" frequencies. If frequencies other than these standard ones are present, a phenomenon known as spectral leakage can degrade the results . To compensate for this, so-called window functions are employed to reduce this leakage by convolving a tapered Gaussian-like function with the data series in the Fourier transform. For more on window functions see, e.g., Priestley [1981, Chapter 7] and Press et al. [1992, Chapter 13.4].

3.6 Other Transforms

The preceding developments have been based on the use of Fourier (trigonometric) series to approximate functions and stochastic processes. The advantage of using Fourier series is that the periodic terms are usually easier to interpret physically. Nevertheless, other approximation or basis functions can be used.

One popular alternative approximation function is the so-called "cas" function which forms the basis of the Hartley transform [Hartley, 1942; Bracewell, 1986]. This function is defined as

$$\mathrm{cas}2\pi ft = \cos 2\pi ft + \sin 2\pi ft, \tag{3.60}$$

and is used in place of $e^{j2\pi ft} = \cos 2\pi ft - j \sin 2\pi ft$ in the usual Fourier expressions. Note that the difference between the two is that the Fourier expressions separate the cosine and sine terms while the Hartley expressions combine them.

In spite of the different functions used in the Fourier and Hartley transforms, they are similar in shape. In fact, the Fourier transform can be deduced from the Hartley transform, although this is considered unnecessary because either transform provides a pair of numbers at each frequency that represents the oscillation of the series in amplitude and phase [Bracewell, 1989]. Moreover, the amplitude and phase spectra obtained from either transform are identical, although they are derived in a slightly different manner [ibid., 1989].

As for the Fourier transform, Bracewell [1986] has also developed a fast Hartley transform in much the same way as the FFT. The advantage is that the fast Hartley transform has been shown to be twice as fast as the FFT and uses half as much computer memory [O'Neill, 1989].

Chapter 4
The Least Squares Transform

4.1 Introduction

A significant limitation of the traditional techniques for the estimation of autocorrelation functions, either directly or indirectly via the inverse of the Fourier spectrum, is that they always require the data to be equally spaced in the argument. Although the data might be evenly spaced with respect to some basic sampling parameter such as time, it will generally not be evenly spaced with respect to other parameters that may better characterize the behaviour of any systematic effects to be modelled by correlation functions. Some typical parameters that might be used to model such systematic effects in geodetic problems include spatial distance, satellite elevation angle, atmospheric temperature, temperature gradient, pressure, etc.; cf. Vaníček and Craymer [1983a,b], Craymer [1984; 1985], Vaníček et al. [1985], and Craymer and Vaníček [1986] Clearly it would be very difficult to obtain a data series evenly spaced in even some of these randomly fluctuating parameters.

Other reasons for seeking alternative techniques are concerned with the limitations of the discrete Fourier transform and FFT described in the preceding chapter. These include the use of only the set of standard "Fourier" frequencies (integer multiples of the fundamental frequency), and the requirement of 2^n data points for the FFT algorithm. In addition, a deterministic model is often estimated and removed from the data prior to any spectral analysis. Traditional spectral techniques do not consider any interaction or linear dependence (correlation) between the a priori deterministic model and the implied periodic

components modelled in the spectrum and in the correlation function. Moreover, the data cannot be weighted in the Fourier transform computation in accordance with their assumed probability density function. Thus, some observations with relatively large random errors will be treated the same as other observations that may be many times more precise.

The aim here is to formulate a more general transform that is capable of handling such unevenly spaced arguments. The transform is based on the least squares spectrum computation developed by Vaníček [1969a; 1971] and is referred to here as the least squares transform and its inverse. Note that this least squares approach is developed here for real-valued data and, consequently, positive frequencies. It cannot cope with complex data or negative frequencies, which are useful in distinguishing between prograde and retrograde motions.

4.2 Matrix Form of Fourier Transform

The basic form of the least squares transform can be derived by first expressing the discrete Fourier transform (DFT) in terms of matrices of complex exponential functions. Rewriting eqn. (3.18) in matrix form gives (the superscript "c" denotes a complex matrix)

$$F_k = \boldsymbol{A}^{\mathrm{c}}{}_{f_k}^{\mathrm{T}}\, \phi\,, \quad \forall\ k = 0,\ldots,n-1\,, \tag{4.1}$$

where

$$\boldsymbol{A}^{\mathrm{c}}{}_{f_k} = e^{2\pi j f_k \boldsymbol{t}} = \begin{bmatrix} e^{2\pi j f_k t_0} \\ e^{2\pi j f_k t_1} \\ \vdots \\ e^{2\pi j f_k t_{n-1}} \end{bmatrix}, \tag{4.2}$$

$$\phi = \begin{bmatrix} \phi(t_o) \\ \phi(t_1) \\ \vdots \\ \phi(t_{n-1}) \end{bmatrix}. \tag{4.3}$$

Note that the transpose in eqn. (4.1) is the complex conjugate transpose for complex matrices (see Golub and Van Loan [1983, p. 9]); i.e.,

$$\boldsymbol{A}^{c}{}_{f_k}^{T} = \begin{bmatrix} e^{-2\pi j f_k t_o} & e^{-2\pi j f_k t_1} & \dots & e^{-2\pi j f_k t_{n-1}} \end{bmatrix}, \tag{4.4}$$

This matrix form of the discrete Fourier transform can be written for each of the discrete "Fourier" frequencies in eqn. (3.17).

Combining all frequencies together gives the simultaneous transform for all the standard Fourier frequencies; i.e.,

$$\boldsymbol{F}^{c} = \boldsymbol{A}^{cT}\,\phi\,, \tag{4.5}$$

where

$$\boldsymbol{F}^{c} = \begin{bmatrix} F_o \\ F_1 \\ \vdots \\ F_{n-1} \end{bmatrix}, \tag{4.6}$$

$$\boldsymbol{A}^{c} = \begin{bmatrix} \boldsymbol{A}^{c}{}_{f_o}\, \boldsymbol{A}^{c}{}_{f_1} \dots \boldsymbol{A}^{c}{}_{f_{n-1}} \end{bmatrix} = \begin{bmatrix} e^{2\pi j f_0 t_o} & e^{2\pi j f_1 t_o} & \dots & e^{2\pi j f_{n-1} t_o} \\ e^{2\pi j f_0 t_1} & e^{2\pi j f_1 t_1} & \dots & e^{2\pi j f_{n-1} t_1} \\ \vdots & \vdots & \dots & \vdots \\ e^{2\pi j f_0 t_{n-1}} & e^{2\pi j f_1 t_{n-1}} & \dots & e^{2\pi j f_{n-1} t_{n-1}} \end{bmatrix}. \tag{4.7}$$

The transpose in eqn. (4.5) again indicates the complex conjugate transpose where

$$\boldsymbol{A}^{c\,T} = \begin{bmatrix} \boldsymbol{A}^{c}{}_{f_o}{}^{T} \\ \boldsymbol{A}^{c}{}_{f_1}{}^{T} \\ \cdots \\ \boldsymbol{A}^{c}{}_{f_{n-1}}{}^{T} \end{bmatrix} = \begin{bmatrix} e^{-2\pi j f_o t_o} & e^{-2\pi j f_o t_1} & \cdots & e^{-2\pi j f_o t_{n-1}} \\ e^{-2\pi j f_1 t_o} & e^{-2\pi j f_1 t_1} & \cdots & e^{-2\pi j f_1 t_{n-1}} \\ \vdots & \vdots & \cdots & \vdots \\ e^{-2\pi j f_{n-1} t_o} & e^{-2\pi j f_{n-1} t_1} & \cdots & e^{-2\pi j f_{n-1} t_{n-1}} \end{bmatrix}. \quad (4.8)$$

Note that $\boldsymbol{A}^{c}{}_{f_k}$ in eqn. (4.1) is the k-th column of $\boldsymbol{A}^{c}$ corresponding to the specific frequency f_k.

The inverse discrete Fourier transform expresses each observation $\phi(t_i)$ in terms of the Fourier transforms F_k for all of the discrete "Fourier" frequencies $f_k = k/(n\Delta t)$, k=0,...,n–1. This can also be written in matrix form as for the direct transform. Rewriting eqn. (3.19) in matrix notation gives

$$\phi(t_i) = \frac{1}{n} \boldsymbol{A}^{c}{}_{t_i} \boldsymbol{F}^{c}, \quad \forall\ i = 0,...,n-1, \quad (4.9)$$

where

$$\boldsymbol{A}^{c}{}_{t_i} = \begin{bmatrix} e^{2\pi j f_0 t_i} & e^{2\pi j f_1 t_i} & \cdots & e^{2\pi j f_{n-1} t_i} \end{bmatrix}. \quad (4.10)$$

Combining all observations together gives the simultaneous inverse transform; i.e.,

$$\phi = \frac{1}{n} \boldsymbol{A}^{c}\, \boldsymbol{F}^{c}, \quad (4.11)$$

where $\boldsymbol{A}^{c}$ is defined as in eqn. (4.7) and ϕ is defined by eqn. (4.3). Note that the design matrix is not transposed in the inverse transform and a factor of 1/n is included as in the complex form. Expanding this in terms of the Fourier transforms for the individual "Fourier" frequencies gives

$$\phi = \frac{1}{n} \sum_{k=1}^{n-1} \boldsymbol{A}^{c}{}_{f_k} \boldsymbol{F}^{c}{}_{k} \ . \tag{4.12}$$

Before developing a more general least squares form of the above transforms, it is necessary to replace these complex expressions with their real-valued trigonometric forms. It will be shown later that this is because, for unequally spaced data, the real and imaginary components can, in general, no longer be treated independently of each other. Using Euler's formula (eqn. (3.10)), the discrete Fourier transform in eqn. (3.18) becomes

$$F_k = \sum_{i=0}^{n-1} \phi(t_i) \left(\cos 2\pi f_k t_i - j \sin 2\pi f_k t_i\right), \quad \forall\ k = 0,\ldots,n-1 \tag{4.13}$$

and the inverse discrete Fourier transform is

$$\phi(t_i) = F_0 \left(\cos 2\pi f_0 t_i\right) + \sum_{k=1}^{n-1} F_k \left(\cos 2\pi f_k t_i + j \sin 2\pi f_k t_i\right), \quad \forall\ i = 0,\ldots,n-1. \tag{4.14}$$

Note that the sine term is zero for the zero frequency component (k=0) in the above expressions. Realizing that the real (cosine) and imaginary (sine) terms are two separate quantities that are independent of each other, the complex expression can be rewritten as two separate real expressions for each term. That is, for the real term,

$$\mathrm{Re}(F_k) = \sum_{i=0}^{n-1} \phi(t_i) \cos 2\pi f_k t_i \tag{4.15}$$

and for the imaginary term,

$$\mathrm{Im}(F_k) = \sum_{j=0}^{n-1} \phi(t_i) \sin 2\pi f_k t_i \ . \tag{4.16}$$

The discrete Fourier transform in eqn. (4.1) can now be expressed in real matrix notation using eqn. (4.1), with separate columns in the design matrix $\boldsymbol{A}$ for the real (cosine) and imaginary (sine) terms. The transform is then given by eqn. (4.1), where $\boldsymbol{F}^{\mathrm{C}}{}_k$ and $\boldsymbol{A}^{\mathrm{C}}{}_{f_k}$ are replaced with $\boldsymbol{F}_k$ and $\boldsymbol{A}_{f_k}$, respectively, which are defined as

$$\boldsymbol{F}_k = \begin{bmatrix} \mathrm{Re}(F_k) \\ \mathrm{Im}(F_k) \end{bmatrix}, \tag{4.17}$$

$$\boldsymbol{A}_{f_k} = \begin{bmatrix} \cos 2\pi f_k t_0 & \sin 2\pi f_k t_0 \\ \cos 2\pi f_k t_1 & \sin 2\pi f_k t_1 \\ \vdots & \vdots \\ \cos 2\pi f_k t_{n-1} & \sin 2\pi f_k t_{n-1} \end{bmatrix}. \tag{4.18}$$

Note that for zero frequency (k=0), Im(F_o)=0 and all the sine terms in $\boldsymbol{A}_{f_o}$ are also zero, so that

$$\boldsymbol{F}_o = \mathrm{Re}(F_k) \ , \tag{4.19}$$

$$\boldsymbol{A}_{f_o} = \begin{bmatrix} \cos 2\pi f_o t_0 \\ \cos 2\pi f_o t_1 \\ \vdots \\ \cos 2\pi f_o t_{n-1} \end{bmatrix}. \tag{4.20}$$

The simultaneous direct and inverse Fourier transforms for all the “Fourier” frequencies are then given by eqns. (4.5) and (4.11), respectively, with $\boldsymbol{F}^{\mathrm{C}}$ and $\boldsymbol{A}^{\mathrm{C}}$ replaced by, respectively,

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_0 \\ \boldsymbol{F}_1 \\ \vdots \\ \boldsymbol{F}_{n-1} \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(F_0) \\ \mathrm{Re}(F_1) \\ \mathrm{Im}(F_1) \\ \vdots \\ \mathrm{Re}(F_{n-1}) \\ \mathrm{Im}(F_{n-1}) \end{bmatrix}, \tag{4.21}$$

$$\boldsymbol{A} = \begin{bmatrix} \cos 2\pi f_0 t_0 & \cos 2\pi f_1 t_0 & \sin 2\pi f_1 t_0 & \cdots & \cos 2\pi f_{n-1} t_0 & \sin 2\pi f_{n-1} t_0 \\ \cos 2\pi f_0 t_1 & \cos 2\pi f_1 t_1 & \sin 2\pi f_1 t_1 & \cdots & \cos 2\pi f_{n-1} t_1 & \sin 2\pi f_{n-1} t_1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \cos 2\pi f_0 t_{n-1} & \cos 2\pi f_1 t_{n-1} & \sin 2\pi f_1 t_{n-1} & \cdots & \cos 2\pi f_{n-1} t_{n-1} & \sin 2\pi f_{n-1} t_{n-1} \end{bmatrix}. \tag{4.22}$$

Note that there are n observations and only n–1 coefficients to solve for.

4.3 Least Squares Transform

A more general least squares transform (LST) can be obtained from the above matrix form of the discrete Fourier transform (DFT) by realizing that the DFT and its inverse are equivalent to least squares interpolation or approximation using trigonometric functions (i.e., Fourier series) as the basis functions (see, e.g., Vaníček and Krakiwsky [1986, Chapter 12] for a detailed exposition of least squares theory). Specifically, a vector of observations ϕ can be approximated in terms of a Fourier series by eqn. (3.1), which can be written in matrix notation as

$$\phi = \boldsymbol{A}\,\boldsymbol{x}\,, \tag{4.23}$$

where

$$\boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ b_1 \\ \vdots \\ a_{n-1} \\ b_{n-1} \end{bmatrix} \tag{4.24}$$

is the vector of Fourier coefficients to be estimated and $\boldsymbol{A}$ represents the basis (trigonometric) functions as defined in eqn. (4.22). Note that for f_o=0, there is no imaginary term and thus no b_o coefficient. The Fourier coefficients $\boldsymbol{x}$ can be estimated by solving for them using the least squares minimization criterion (cf. Vaníček and Krakiwsky [1986, pp. 204-207]). The solution is given by

$$\hat{\boldsymbol{x}} = \boldsymbol{N}^{-1}\,\boldsymbol{A}^{\mathrm{T}}\,\phi\,. \tag{4.25}$$

where $\boldsymbol{N} = \boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A}$ is the normal equation coefficient matrix.

Note that in the above equation $\boldsymbol{A}^{\mathrm{T}}\phi$ is the matrix form of the (simultaneous) discrete Fourier transform in eqn. (4.5). Thus, the least squares transform for all frequencies simultaneously is given by eqn. (4.5) and the transform for each frequency f_k by

$$\boldsymbol{F}_k = \boldsymbol{A}_{f_k}{}^{\mathrm{T}}\,\phi\,, \tag{4.26}$$

where $\boldsymbol{A}_k$ that part of $\boldsymbol{A}$ corresponding to only frequency f_k.

The estimated Fourier coefficients in eqn. (4.25) can then be written as

$$\hat{\boldsymbol{x}} = \boldsymbol{N}^{-1}\,\boldsymbol{F}\,. \tag{4.27}$$

Substituting this in eqn. (4.23) gives the estimated observations

$$\hat{\phi} = \boldsymbol{A}\,\boldsymbol{N}^{-1}\,\boldsymbol{F}\,, \tag{4.28}$$

which represents the simultaneous inverse least squares transform for all frequencies. The individual observations $\phi(t_i)$ are then given by

$$\hat{\phi}(t_i) \;=\; \boldsymbol{A}_{t_i}\; \boldsymbol{N}^{-1}\, \boldsymbol{F}\;, \tag{4.29}$$

where $\boldsymbol{A}_{t_i}$ represents the i-th row of $\boldsymbol{A}$ corresponding to time t_i.

The conventional Fourier transforms are just a special case of these more general least squares definitions for equally weighted and equally spaced data. Although the direct least squares and Fourier transforms are equivalent by definition, the equivalence of the inverse transforms is not easy to see from the matrix expressions. This equivalence can be shown by examining the elements of $\boldsymbol{N}^{-1}$. Realizing that the Fourier expressions are valid only for equally spaced data and the discrete set of "Fourier" frequencies, it can be shown that the columns of $\boldsymbol{A}$ form an orthogonal basis under these assumptions. The elements of $\boldsymbol{N}$ (summations of trigonometric products) reduce to

$$\sum_{i=0}^{n-1}\left(\cos 2\pi f_k t_i\;\cos 2\pi f_l t_i\right) = \begin{cases} n \text{ for } k=l=0 \text{ or } n/2 \\ n/2 \text{ for } k=l\neq 0 \text{ or} n/2 \\ 0 \text{ for } k\neq l \end{cases}\;, \tag{4.30}$$

$$\sum_{i=0}^{n-1}\left(\sin 2\pi f_k t_i\;\sin 2\pi f_l t_i\right) = \begin{cases} 0 \text{ for } k=l=0 \text{ or } n/2 \\ n/2 \text{ for } k=l\neq 0 \text{ or} n/2 \\ 0 \text{ for } k\neq l \end{cases}\;, \tag{4.31}$$

$$\sum_{i=0}^{n-1}\left(\cos 2\pi f_k t_i \sin 2\pi f_k t_i\right) = 0\;, \quad \forall\; k\;. \tag{4.32}$$

Substituting these in $\boldsymbol{N}^{-1}$ in eqn. (4.28) and expanding in terms of the Fourier transforms for the individual frequencies gives

$$\hat{\phi} = \frac{1}{n} \boldsymbol{A}_{f_o} \boldsymbol{F}_o + \frac{2}{n} \sum_{k=1}^{n/2-1} \boldsymbol{A}_{f_k} \boldsymbol{F}_k \quad . \tag{4.33}$$

The difference between this and the inverse Fourier transform in eqn. (4.11) is the use of $n/2$ in place of n for non-zero frequencies (n is assumed to be even, otherwise n/2 is truncated down to the nearest integer). This is because for real data the transform for negative frequencies is identical to that for positive frequencies. The columns of $\boldsymbol{A}$ corresponding to these frequencies will be identical thus making **N** singular when simultaneously estimating all frequencies. Including only the positive frequencies will implicitly account for the identical response for both negative and positive frequencies, thereby effectively doubling the least squares transform with respect to the Fourier transform (i.e., it gives a transform which results in a one-sided spectrum as derived in the next chapter) . Note that the Nyquist frequency (at $k=n/2$) is also excluded from the summation since this is aliased with the zero frequency.

It is important to realize that for unequally spaced data the inverse least squares transform in eqn. (4.28) cannot in general be expressed as a summation of independent contributions from individual frequencies. This is because $\boldsymbol{N}$ in general contains off-diagonal elements between frequencies and even between the sine and cosine components for the same frequency; i.e., these Fourier components are mathematically correlated with each other (i.e., they are no longer orthogonal or linearly independent).

4.4 Weighted Least Squares Transform

The above developments have implicitly assumed the observations to be equally weighted. A more general form of the least squares transforms can be derived by weighting the observations using their associated covariance matrix $\boldsymbol{C}_{\phi}$. This also allows one to model any known correlations among the observations. The general expressions for

a weighted least squares interpolation or approximation are given by (cf. Vaníček and Krakiwsky [1986, pp. 204-207])

$$\hat{\boldsymbol{x}} = \boldsymbol{N}^{-1}\,\boldsymbol{u}\ . \tag{4.34}$$

$$\hat{\boldsymbol{\phi}} = \boldsymbol{A}\,\hat{\boldsymbol{x}}\ , \tag{4.35}$$

where $\boldsymbol{N} = \boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{A}$ is the normal equation coefficient matrix, $\boldsymbol{u} = \boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{\phi}$, is the normal equation constant vector and $\boldsymbol{P} = \boldsymbol{C}_{\boldsymbol{\phi}}^{-1}$ is the weight matrix of the observations.

Following the same development as for the unweighted (i.e., equally weighted) least squares transforms, the more general weighted least squares transform for all frequencies simultaneously is given by (cf. eqn. (4.26))

$$\boldsymbol{F} = \boldsymbol{u} = \boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{\phi}\ . \tag{4.36}$$

and the transform for each individual frequency f_k by (cf. eqn. (4.1))

$$\boldsymbol{F}_k = \boldsymbol{u}_k = \boldsymbol{A}_{f_k}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{\phi}\ . \tag{4.37}$$

Using this in the least squares estimation of the Fourier coefficient in eqn. (4.34) and then substituting into eqn. (4.35) gives the inverse least squares transform (cf. eqn. (4.28))

$$\hat{\boldsymbol{\phi}} = \boldsymbol{A}\,\hat{\boldsymbol{x}} = \boldsymbol{A}\ \boldsymbol{N}^{-1}\,\boldsymbol{F}\ . \tag{4.38}$$

The individual observations $\phi(t_i)$ are then

$$\hat{\phi}(t_i) = \boldsymbol{A}_{t_i}\ \boldsymbol{N}^{-1}\,\boldsymbol{F}\ . \tag{4.39}$$

Although the symbolic form of these expressions are identical to those for the unweighted inverse transform in eqns. (4.26) and (4.27), $\boldsymbol{N}$ and $\boldsymbol{F}$ are defined differently (they include the weight matrix $\boldsymbol{P}$). Note that the inverse transform is essentially just a least squares approximation of ϕ in terms of a Fourier series.

As stated at the end of Section 4.3, it is not possible in general to separately estimate the individual Fourier transform values for different frequencies because of the possible existence of mathematical correlations (non-orthogonality) among the Fourier components (trig functions) due to unequal data spacing or correlations among the observations. If, however, the observations are equally spaced, equally weighted and uncorrelated (i.e., $\boldsymbol{P} = \boldsymbol{I}$), and the set of "Fourier" frequencies are used, the normal equation matrix becomes a diagonal (i.e., $\boldsymbol{N}$ = diag(n, $n/2$, $n/2$,...) and the direct and inverse least squares transforms become identical to eqns. (4.26) and (4.33), respectively, and are thus equivalent to the standard Fourier ones. The Fourier transform is thus just a special case of the least squares transform.

An attractive feature of the least squares transform is that the covariance matrix for the Fourier coefficients and the inverse least squares transform are provided by the least squares theory as by-products of inverting the normal equation matrix $\boldsymbol{N}$ (cf. Vaníček and Krakiwsky [1986, pp. 209-210]). The covariance matrix for the estimated Fourier coefficients $\hat{\boldsymbol{x}}$ is given by

$$\boldsymbol{C}_{\hat{x}} = \boldsymbol{N}^{-1} \tag{4.40}$$

while that for the inverse transform (interpolated/approximated observations) is

$$\boldsymbol{C}_{\hat{\phi}} = \boldsymbol{A}\, \boldsymbol{C}_{\hat{x}}\, \boldsymbol{A}^{\mathrm{T}} . \tag{4.41}$$

It is recalled that only frequencies up to, but not including, the Nyquist frequency should be included in the Fourier series in order to avoid singularities in $\boldsymbol{N}$ due to the aliasing effect. In addition, if the data are equally spaced, only the set of standard "Fourier" frequencies should be used (see Section 3.2). Moreover, if the data are real, only the positive Fourier frequencies should be included (see property in eqn. (3.23)). This then allows for a total of n–1 terms (n/2–1 cosines and n/2 sines) to be estimated from n observations, which gives a nearly unique solution for the Fourier coefficients and enables the observations to be reproduced exactly using the inverse transform.

In addition to accepting unequally spaced data, another advantage of the least squares transforms are that they are not restricted to only the set of standard Fourier frequencies $f_k = k/(n\,\Delta t) = k/T$ for k=0,...,n–1. Any set of frequencies in the range (0, f_N) can be used in the expressions. However, only a maximum of n/2 frequencies (n Fourier coefficients) can be estimated simultaneously from only n observations. Moreover, some serious repercussions can also arise if the selected frequencies result in some of the Fourier components (trig functions) becoming nearly linearly dependent with each other, thereby producing an ill-conditioned or near singular $\boldsymbol{N}$. To avoid such ill-conditioning it becomes necessary to either select a different set of frequencies to be estimated (e.g., equally spaced frequencies) or simply neglect the correlations in $\boldsymbol{N}$ (i.e., the off-diagonal blocks) and estimate the inverse least squares transform separately for the individual frequencies using eqn. (4.39).

Another problem in dealing with unequally spaced data is that the Nyquist frequency is not well defined, if at all. It was thought that, because a single cycle of a periodic function can be defined with only 3 points, the smallest time interval of a triplet of adjacent points would represent the smallest period which can be estimated. Care would also be need to ensure that no pair of points in the triplet are so close together that the triplet is essentially only a pair of points for all practical purposes. In practice, however, this triplet interval does not appear to define a Nyquist frequency. As will be shown in the

numerical tests of Chapter 7, spectra computed to frequencies well beyond this implied Nyquist frequency do not exhibit the expected mirror image about any Nyquist frequency.

4.5 Effect of Deterministic Model

So far it has been assumed that the original data is stationary and can be modelled completely by a Fourier series. In general this is hardly ever the case. It is more common to first remove the non-stationarity by modelling some known a priori deterministic trends using, e.g., least squares fitting and to analyse the residual (stochastic) series using the above techniques. The problem, however, is that there may be linear dependence between the deterministic model and the periodic components in the Fourier series (the stochastic model) which may significantly affect the Fourier transform and spectrum.

To account for such effects, it is necessary to reformulate the preceding developments to accommodate both the deterministic model as well as the stochastic model (periodic Fourier series components) in the estimation of a least squares transform. Partitioning $\boldsymbol{A}$ and $\boldsymbol{x}$,

$$\boldsymbol{A} = [\boldsymbol{A}_D \quad \boldsymbol{A}_S] , \tag{4.42}$$

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_D \\ \boldsymbol{x}_S \end{bmatrix} , \tag{4.43}$$

the data series (observation) vector ϕ is modelled in terms of both deterministic ϕ_D and stochastic (Fourier series) ϕ_S components as

$$\phi = \boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{A}_D\,\boldsymbol{x}_D + \boldsymbol{A}_S\,\boldsymbol{x}_S = \phi_D + \phi_S . \tag{4.44}$$

For the deterministic model, $\boldsymbol{A}_D$ is the design matrix and $\boldsymbol{x}_D$ is the parameter vector to be estimated, and for the stochastic (Fourier series) model, $\boldsymbol{A}_S$ is the matrix of cosine and sine basis functions as defined in eqn. (4.22) and $\boldsymbol{x}_S$ is the vector of Fourier coefficients to be estimated as defined in eqn. (4.24). The aim is to account for the effect of estimating $\boldsymbol{x}_D$ in the estimation of $\boldsymbol{x}_S$.

The weighted least squares estimates of the combined parameter vector $\hat{\boldsymbol{x}}$ and the approximated observation vector $\hat{\phi}$ are given by eqns. (4.34) and (4.35), where the matrices are defined as above. Substituting the above partitioned forms of $\boldsymbol{A}$ and $\boldsymbol{x}$ into these expressions gives

$$\begin{bmatrix} \hat{\boldsymbol{x}}_D \\ \hat{\boldsymbol{x}}_S \end{bmatrix} = \begin{bmatrix} \boldsymbol{N}_{DD} & \boldsymbol{N}_{DS} \\ \boldsymbol{N}_{SD} & \boldsymbol{N}_{SS} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{u}_D \\ \boldsymbol{u}_S \end{bmatrix}, \tag{4.45}$$

$$\hat{\phi} = [\boldsymbol{A}_D \quad \boldsymbol{A}_S] \begin{bmatrix} \hat{\boldsymbol{x}}_D \\ \hat{\boldsymbol{x}}_S \end{bmatrix} = \boldsymbol{A}_D \hat{\boldsymbol{x}}_D + \boldsymbol{A}_S \hat{\boldsymbol{x}}_S = \hat{\phi}_D + \hat{\phi}_S , \tag{4.46}$$

where

$$\boldsymbol{N}_{DD} = \boldsymbol{A}_D^{\mathrm{T}} \boldsymbol{P} \boldsymbol{A}_D , \tag{4.47}$$

$$\boldsymbol{N}_{DS} = \boldsymbol{A}_D^{\mathrm{T}} \boldsymbol{P} \boldsymbol{A}_S , \tag{4.48}$$

$$\boldsymbol{N}_{SD} = \boldsymbol{A}_S^{\mathrm{T}} \boldsymbol{P} \boldsymbol{A}_D , \tag{4.49}$$

$$\boldsymbol{N}_{SS} = \boldsymbol{A}_S^{\mathrm{T}} \boldsymbol{P} \boldsymbol{A}_S . \tag{4.50}$$

$$\boldsymbol{u}_D = \boldsymbol{A}_D^{\mathrm{T}} \boldsymbol{P} \phi , \tag{4.51}$$

$$\boldsymbol{u}_S = \boldsymbol{A}_S^{\mathrm{T}} \boldsymbol{P} \phi . \tag{4.52}$$

Although, for stochastic modelling, we are really only interested in ϕ_S, it is necessary to account for any effect of the deterministic model on the estimation of $\hat{\phi}_S$ by $\hat{\boldsymbol{x}}_D$. This is obtained by making use of some well-known matrix identities in the evaluation

of $\hat{\boldsymbol{x}}_S$. Specifically, the inversion of the normal equation matrix $\boldsymbol{N}$ can be written as [Vaníček and Krakiwsky, 1986, p. 28]

$$\begin{bmatrix} \boldsymbol{N}_{DD} & \boldsymbol{N}_{DS} \\ \boldsymbol{N}_{SD} & \boldsymbol{N}_{SS} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{M}_{DD} & \boldsymbol{M}_{DS} \\ \boldsymbol{M}_{SD} & \boldsymbol{M}_{SS} \end{bmatrix}, \tag{4.53}$$

where

$$\begin{aligned} \boldsymbol{M}_{DD} &= (\boldsymbol{N}_{DD} - \boldsymbol{N}_{DS}\,\boldsymbol{N}_{SS}^{-1}\,\boldsymbol{N}_{SD})^{-1} \\ &= \boldsymbol{N}_{DD}^{-1} + \boldsymbol{N}_{DD}^{-1}\,\boldsymbol{N}_{DS}\,\boldsymbol{M}_{SS}\,\boldsymbol{N}_{SD}\,\boldsymbol{N}_{DD}^{-1}\,, \end{aligned} \tag{4.54}$$

$$\boldsymbol{M}_{DS} = -\boldsymbol{M}_{DD}\,\boldsymbol{N}_{DS}\,\boldsymbol{N}_{SS}^{-1} = \boldsymbol{M}_{SD}^{\mathrm{T}}\,, \tag{4.55}$$

$$\boldsymbol{M}_{SD} = -\boldsymbol{M}_{SS}\,\boldsymbol{N}_{SD}\,\boldsymbol{N}_{DD}^{-1} = \boldsymbol{M}_{DS}^{\mathrm{T}}\,, \tag{4.56}$$

$$\begin{aligned} \boldsymbol{M}_{SS} &= (\boldsymbol{N}_{SS} - \boldsymbol{N}_{SD}\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{N}_{DS})^{-1} \\ &= \boldsymbol{N}_{SS}^{-1} + \boldsymbol{N}_{SS}^{-1}\,\boldsymbol{N}_{SD}\,\boldsymbol{M}_{DD}\,\boldsymbol{N}_{DS}\,\boldsymbol{N}_{SS}^{-1}\,. \end{aligned} \tag{4.57}$$

Substituting into eqn. (4.45) and gives for $\hat{\boldsymbol{x}}_S$

$$\begin{aligned} \hat{\boldsymbol{x}}_S &= \boldsymbol{M}_{SD}\,\boldsymbol{u}_D + \boldsymbol{M}_{SS}\,\boldsymbol{u}_S \\ &= (\boldsymbol{N}_{SS} - \boldsymbol{N}_{SD}\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{N}_{DS})^{-1}\,(\boldsymbol{u}_S - \boldsymbol{N}_{SD}\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{u}_D) \end{aligned} \tag{4.58}$$

where the so-called “reduced” normal equation matrix and constant vector are

$$\boldsymbol{N}^{*} = \boldsymbol{N}_{SS} - \boldsymbol{N}_{SD}\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{N}_{DS}\,, \tag{4.59}$$

$$\boldsymbol{u}^{*} = \boldsymbol{u}_S - \boldsymbol{N}_{SD}\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{u}_D\,. \tag{4.60}$$

Defining the "reduced" weight matrix $\boldsymbol{P}^*$, which accounts for the effect of the deterministic model, by

$$\boldsymbol{P}^* = \boldsymbol{P} - \boldsymbol{P}\,\boldsymbol{A}_D\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{A}_D^{\mathrm{T}}\,\boldsymbol{P}\,, \tag{4.61}$$

the normal equations in eqn. (4.58) can be written in the same general form as that without the deterministic model; i.e.,

$$\hat{\boldsymbol{x}}_S = \boldsymbol{N}^{*-1}\,\boldsymbol{u}^*\,, \tag{4.62}$$

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}_S} = \boldsymbol{N}^{*1}\,, \tag{4.63}$$

where

$$\boldsymbol{N}^{*-1} = \left(\boldsymbol{A}_S^{\mathrm{T}}\,\boldsymbol{P}^*\,\boldsymbol{A}_S\right)^{-1}\,, \tag{4.64}$$

$$\boldsymbol{u}^* = \boldsymbol{A}_S^{\mathrm{T}}\,\boldsymbol{P}^*\,\boldsymbol{\phi}\,. \tag{4.65}$$

The simultaneous least squares transform $\boldsymbol{F}^*$ (for all frequencies simultaneously) which accounts for the deterministic model is then defined in the same manner as in eqn. (4.36):

$$\boldsymbol{F}^* = \boldsymbol{A}_S^{\mathrm{T}}\,\boldsymbol{P}^*\,\boldsymbol{\phi}\,. \tag{4.66}$$

The transform for each individual frequency f_k is then (cf. eqn. (4.37))

$$\boldsymbol{F}_k^* = \boldsymbol{A}_{f_k}{}^{\mathrm{T}} \boldsymbol{P}^* \phi \,. \tag{4.67}$$

Similarly, the inverse transform for all observations is defined by eqn. (4.38), using the reduced forms of $\boldsymbol{N}$ and $\boldsymbol{F}$, as

$$\hat{\phi} = \boldsymbol{A} \;\; \boldsymbol{N}^{*-1} \boldsymbol{F}^* \,, \tag{4.68}$$

and for individual observations $\phi(t_i)$ by

$$\hat{\phi}(t_i) = \boldsymbol{A}_{t_i} \;\; \boldsymbol{N}^{*-1} \boldsymbol{F}^* \,. \tag{4.69}$$

The expressions for independently estimated frequency components are simply obtained by ignoring the off-diagonal terms between different frequencies in $\boldsymbol{N}^*$ and $\boldsymbol{P}^*$

When there is no deterministic model, $\boldsymbol{A}_D = \boldsymbol{0}$, $\boldsymbol{P}^* = \boldsymbol{P}$ and the above expressions reduce to the same form as in the previous section. Note that the weighted inverse transform is essentially just a weighted least squares approximation of ϕ in terms of the a priori deterministic model and the individual periodic (Fourier series) components.

4.6 Vector Space Interpretation

The least squares transform can be more elegantly interpreted using the concept of Hilbert spaces and commutative diagrams using the language of functional analysis. The fundamental component of functional analysis is the space, in which we want to work. The elements in a space can be real numbers, complex numbers, vectors, matrices as well as functions of these. Here we consider the more restrictive case of vector spaces consisting of sets of vectors which can be visualized as positions in the space. A brief

review of functional analysis as it applies to the geometrical interpretation of the least squares transform is given. For more on functional analysis see, e.g., Kreyszig [1978].

There are various classifications of spaces. The most general type of space is the metric space in which the concept of a distance (or metric) $\rho(x,y)$ between two elements x and y in the space is defined. A normed space is a metric space in which a norm $\|\bullet\|$ may be induced as the distance from the null element. The norm $\|x\|$ of a single element x is just its length $\rho(x,0)$. A Hilbert space is a normed space in which a scalar (or inner) product $\langle x,y \rangle$ for a pair of elements x and y may be induced by the relations

$$\|x\| = \langle x, x \rangle^{1/2} , \tag{4.70}$$

$$\rho(x,y) = \|x-y\| = [\, \langle (x-y), (x-y) \rangle \,]^{1/2} . \tag{4.71}$$

There are many ways of defining a scalar product. For vector spaces of finite dimension the most common is the simple linear combination of vector elements; i.e., for vectors $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y} = \sum_i x_i \, y_i . \tag{4.72}$$

For compact vector spaces the analogous form of the scalar product is

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \int x(t) \, y(t) \, dt . \tag{4.73}$$

A more general definition of the discrete scalar product, and the one used here, is the norm defined by

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}\, \boldsymbol{y} \quad , \tag{4.74}$$

where $\boldsymbol{P}$, the weight matrix for the vector space, is generally the inverse of the covariance matrix of the vector elements. This corresponds to a generalization of Euclidean space with metric tensor $\boldsymbol{I}$, into a Riemanian space with metric tensor $\boldsymbol{P}$. Note that for compact matrices, the vectors and matrices will also be compact, and contain continuous functions.

An interpretation of basic least squares theory in terms of functional analysis is given by Vaníček [1986]. The theory is interpreted using commutative diagrams which describe the various transformations between probabilistic (Hilbert) spaces. The same diagram can be used to interpret the least squares transform. In this diagram ϕ is the observation vector belonging to the observation space $\boldsymbol{\Phi}$, $\boldsymbol{C}_\phi$ is the observation covariance matrix (not necessarily diagonal) defining the scalar product (and norm and distance) in this space, $\boldsymbol{x}$ is the parameter vector of Fourier coefficients to be estimated belonging to the parameter space $\boldsymbol{\mathbb{X}}$ and $\boldsymbol{A}$ is the design matrix transforming the observations to the parameters, which contains the sines and cosines functions (basis functions).

The commutative diagram is set up by first defining the transformation (i.e., the observation equations) $\phi = \boldsymbol{A}\boldsymbol{x}$ from the parameter space $\boldsymbol{\mathbb{X}}$ to the observation space $\boldsymbol{\Phi}$. The weight matrices $\boldsymbol{P}_{\boldsymbol{x}}$ and $\boldsymbol{P}_\phi$ define the transformations to the dual parameter space $\boldsymbol{\mathbb{X}}^*$ and dual observation space $\boldsymbol{\Phi}^*$, respectively. The transformation from the dual observation space $\boldsymbol{\Phi}^*$ to the dual parameter space $\boldsymbol{\mathbb{X}}^*$. is defined by $\boldsymbol{A}^{\mathrm{T}}$. Assuming the design matrix $\boldsymbol{A}$ and covariance matrix $\boldsymbol{C}_\phi$ are known, the remaining transformations can be obtained from the commutative diagram using the following steps.

1. $\boldsymbol{P}_\phi = \boldsymbol{C}_\phi^{-1}$ (4.75)

2. $\boldsymbol{P}_{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{f}}\, \boldsymbol{A} \quad \Rightarrow \quad \boldsymbol{C}_{\boldsymbol{x}} = \boldsymbol{P}_{\boldsymbol{x}}^{-1}$ (4.76)

3. $\boldsymbol{A}^{-} = \boldsymbol{C}_{\boldsymbol{x}} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{P}_\phi = (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{P}_\phi\, \boldsymbol{A})^{-1} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{P}_{\boldsymbol{f}}$ (4.77)

4. $\boldsymbol{F} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{P}_\phi$ (4.78)

$$5.\quad \boldsymbol{F}^{-1} \;=\; \boldsymbol{A}\,\boldsymbol{C}_{\boldsymbol{x}} \;=\; \boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{P}_{\phi}\,\boldsymbol{A})^{-1} \tag{4.79}$$

These steps are illustrated in Figure 4.1. Here, $\boldsymbol{F}$ is defined slightly differently than in the preceding developments. It represents the transform operator that acts on the observations, and not the entire transform itself as defined in Section 4.3. Similarly, $\boldsymbol{F}^{-1}$ is the inverse operator.

It can be seen from the commutative diagram that the least squares Fourier transform $\boldsymbol{F}$ is a transformation from the observation space $\boldsymbol{\Phi}$ to the dual parameter space $\mathbb{X}^*$ via the dual observation space $\boldsymbol{\Phi}^*$. The inverse least squares Fourier transform $\boldsymbol{F}^{-1}$ is then derived by proceeding from the dual parameter space $\mathbb{X}^*$ to the observation space $\boldsymbol{\Phi}$ via the parameter space $\mathbb{X}$.

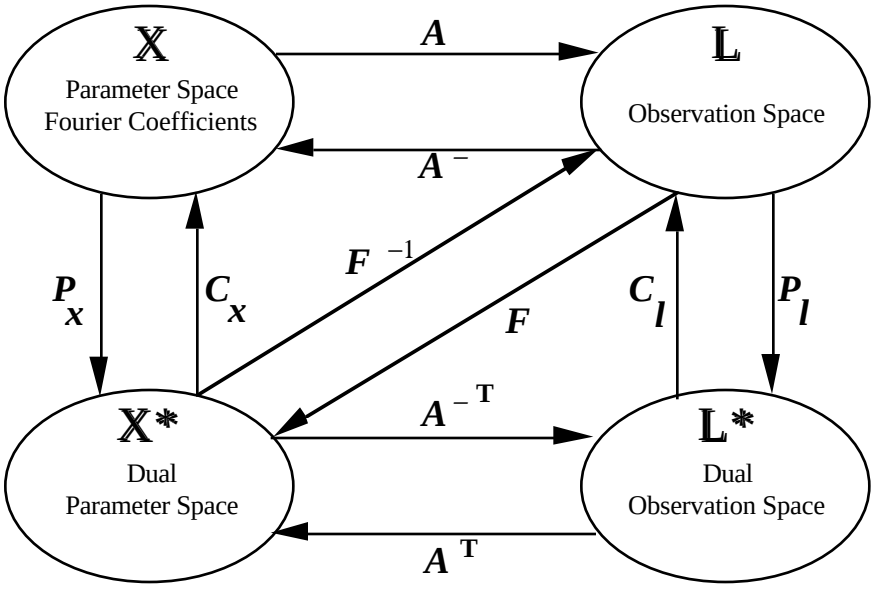

Figure 4.1: Commutative diagram for the direct and inverse least squares transform, where $\boldsymbol{F}$ denotes the direct transform and $\boldsymbol{F}^{-1}$ the inverse transform.

The design matrix $\boldsymbol{A}$ contains the trigonometric functions defining the Fourier series representation of the observations. The individual sine and cosine terms (columns of $\boldsymbol{A}$) form a basis for the observation space. For the standard Fourier transform, the data are equally spaced, equally weighted and uncorrelated so that the columns of $\boldsymbol{A}$ form an orthogonal basis. The normal equation matrix $\boldsymbol{N} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{A}$ then becomes a diagonal matrix as does the covariance matrix of the parameters. In the more general least squares transform, the base functions are not necessarily orthogonal, although, in practice, this is usually the case even with unequally spaced data.

4.7 Applications

The above least squares transform can be applied in the same manner as the traditional Fourier one, with the added advantage that they can be used not only for equally spaced data series, but also for unequally spaced series and for any arbitrary set of frequencies. One of the most important applications (to be discussed in the next chapter) is the determination of the power spectral density for unequally spaced data that also accounts for a deterministic model. In this case there is no need to determine a frequency response function for the deterministic model in order to remove its effect from the spectrum of the model residuals. The correct spectrum is obtained directly when the deterministic model is accounted for in the formulation of the spectrum.

Another important application of the least squares transform is the indirect estimation of autocovariance/autocorrelation functions using the correlation theorem (see Chapter 6). Instead of transforming the effect of all the spectral values, a smoother autocovariance function can be obtained by using only the significant spectral values. Because these significant spectral components are not likely to be evenly spaced, it is necessary to use the inverse least squares transform to convert them into an autocorrelation function.

The inverse least squares transform can also be used in data series approximation and interpolation problems. In these applications the direct Fourier transform is used to estimate Fourier series coefficients, which are then used in the inverse transform to approximate or interpolate the original series. The degree of smoothing of the original series can be increased by including only frequencies corresponding to highly significant Fourier coefficients (or spectral peaks).

Chapter 5
The Least Squares Spectrum

5.1 Introduction

As discussed in the previous chapters, traditional methods of determining power spectral density and autocorrelation functions are significantly limited in their application because they always require the data to be equally spaced in the argument. Other reasons for seeking alternative techniques are concerned with the limitations of the discrete Fourier transform and FFT commonly used to generate spectra as well as autocorrelation functions (transformed from the spectrum). These include the use of only the set of "Fourier" frequencies (integer multiples of the fundamental frequency), and the requirement of 2^n data points (for the FFT algorithm). In addition, the traditional techniques do not consider any interaction (correlation) between the deterministic model and the implied periodic components modelled in the spectrum. Moreover, the data cannot be weighted in the transform computation in accordance with their assumed probability density function. Thus, some observations with relatively large random errors will be weighted the same as other observations that may be many times more precise.

Traditional methods of computing power spectral density functions from unequally spaced data have often been based on interpolation or approximation. That is, the original unequally spaced data series was interpolated or approximated to an equally spaced series for which the standard Fourier techniques could then be applied. The problem, however, is that this approach really creates a new data series that depends on the smoothness of the original series, the presence of data gaps and the subjective choice of the interpolating or

approximating function. The interpolation also tends to smooth out any high frequency components of the original data series.

To overcome these limitations and difficulties, Vaníček [1969a] developed a method of spectrum computation based on least squares estimation. This method was further developed in Vaníček [1971], Steeves [1981] and Wells et al. [1985] and forms the basis of other similar techniques in slightly different forms promoted by various authors since (e.g., Rochester et al. [1974], Lomb [1976], Ferraz-Mello [1981], Scargle [1982], Horne and Baliunas [1986]). In this Chapter, the same basic least squares spectrum is reformulated in terms of the newly developed least squares transform. A new "simultaneous" spectral estimation procedure, somewhat similar to that used by Rochester et al. [1974], is also developed.

5.2 Matrix Form of Fourier Spectrum

Before giving the expressions for the least squares spectrum, the Fourier spectrum is first expressed in matrix form. This is done by simply using the matrix expressions for the Fourier transform (eqns. (4.9) and (4.21)) in the definition of total power (eqn. (3.47)) and the individual Fourier spectral estimates (eqn. (3.48)). Parseval's relation in eqn. (3.47) can then be written in matrix notation as

$$\text{Total power} \;=\; \phi^{\mathrm{T}}\phi \;=\; \frac{1}{n}\,\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\;. \tag{5.1}$$

where

$$\phi^{\mathrm{T}}\phi \;=\; \sum_{i=0}^{n-1} \phi(t_i)^2 \quad , \tag{5.2}$$

$$\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} \;=\; \sum_{k=0}^{n-1} \boldsymbol{F}_k{}^{\mathrm{T}}\boldsymbol{F}_k \;=\; \sum_{k=0}^{n-1} |F(f_k)|^2 \;. \tag{5.3}$$

The individual spectral components for the two-sided power spectral density function (eqn. (3.48)) are then given by

$$s(f_k) \;=\; \frac{1}{n}\,\boldsymbol{F}_k{}^{\mathrm{T}}\,\boldsymbol{F}_k \;=\; \frac{1}{n}\,|F(f_k)|^2 \;, \quad \forall\; k = 0,\ldots,n-1 \;. \tag{5.4}$$

The one-sided spectral density function is twice the two-side function and is defined by

$$s(f_k) \;=\; \begin{cases} \dfrac{1}{n}\,\boldsymbol{F}_k{}^{\mathrm{T}}\,\boldsymbol{F}_k \;=\; \dfrac{1}{n}\,|F(f_k)|^2 & \text{for } k = 0 \\[2ex] \dfrac{2}{n}\,\boldsymbol{F}_k{}^{\mathrm{T}}\,\boldsymbol{F}_k \;=\; \dfrac{2}{n}\,|F(f_k)|^2 & \text{for } k = 1\ldots n/2-1 \end{cases} \;. \tag{5.5}$$

5.3 Least Squares Spectrum

The least squares spectrum was originally developed by Vaníček [1969a; 1972] (see also Steeves [1981] and Wells et al. [1985]). The expressions for this form of the least squares spectrum (referred to here as the "conventional" form) can be developed in terms of the (unweighted) least squares transform. First, the total power is given by

$$\text{Total power} \;=\; \phi^{\mathrm{T}}\,\phi \;. \tag{5.6}$$

Substituting for the inverse least squares transform in eqn. (4.28) results in

$$\text{Total power} \;=\; \phi^{\mathrm{T}}\,\phi \;=\; \boldsymbol{F}^{\mathrm{T}}\,\boldsymbol{N}^{-1}\,\boldsymbol{F} \;. \tag{5.7}$$

Note that, generally, the total power can not be expressed as a sum of individual contributions from the different frequency components. As with the inverse least squares transform, the problem is that with unequally spaced data, $\boldsymbol{N}$ is not a diagonal matrix because the Fourier components (trig functions) are not orthogonal to (linearly independent of) each other. As explained above, this problem is avoided by simply examining one frequency at a time, independently (out of context) of the others. This is equivalent to ignoring the linear dependence between different frequency components in $\boldsymbol{N}$ and amounts to defining the spectrum as the independent contribution of each frequency component to the total power.

Following this approach, the spectral component $s(f_k)$ (for the one-sided least squares power spectral density function) is defined by

$$s(f_k) = \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{N}_k^{-1} \boldsymbol{F}_k , \tag{5.8}$$

where $\boldsymbol{N}_k$ is the k-th diagonal block of $\boldsymbol{N}$ corresponding to frequency f_k. The normalized spectral values $\tilde{s}(f_k)$ are then

$$\tilde{s}(f_k) = \frac{s(f_k)}{\phi^{\mathrm{T}}\phi} = \frac{\boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{N}_k^{-1} \boldsymbol{F}_k}{\phi^{T} \phi} . \tag{5.9}$$

The normalized spectrum represents the percentage of variation in the original data series independently explained by each spectral component. In its basic philosophy, this corresponds to the R^2 statistic in regression analysis [Draper and Smith, 1981, p. 33].

One of the most significant advantages of the least squares spectrum, other than handling unequally spaced data, is the ability to estimate spectral components for any real (arbitrary) frequency, not just the set of "Fourier" frequencies. The expressions in eqns. (5.8) and (5.9) essentially provide continuous estimates for any set of frequencies. The

usual procedure is to take a set of equally spaced frequencies between zero and the estimated Nyquist or maximum frequency (note that the Nyquist frequency is undefined for unevenly spaced data as discussed Section 4.4). The precise frequency location of significant peaks can then be determined by "zooming" in on that frequency area of the spectrum. This allows one to locate frequencies for significant peaks to any resolution, within the limits of the data sampling.

5.4 Weighted Least Squares Spectrum

The more general weighted least squares power spectrum is obtained in a similar way except that the general (weighted) least squares transforms are used in the above developments. In this more general situation of an observation weight matrix, the total power is defined by the weighted sum of squares as

$$\text{Total power} = \phi^{\mathrm{T}} \boldsymbol{P} \phi , \tag{5.10}$$

where $\boldsymbol{P}$ is the inverse of the observation covariance matrix. Substituting for ϕ using the weighted least squares inverse transform in eqn. (4.38) and noting that $\boldsymbol{A}^{\mathrm{T}} \boldsymbol{P} \boldsymbol{A} = \boldsymbol{N}$ gives

$$\text{Total power} = \phi^{\mathrm{T}} \boldsymbol{P} \phi = \boldsymbol{F}^{\mathrm{T}} \boldsymbol{N}^{-1} \boldsymbol{F} . \tag{5.11}$$

Vaníček [1969a] defines the spectrum as the independent frequency contributions to this total power (cf. eqns. (5.8) and (5.9)). That is, each frequency component is estimated independently, or out of context, of the others. Steeves [1981] extends this approach by incorporating the weight matrix ($\boldsymbol{P}$) of the observations. The independent estimate of each spectral component is then obtained using the weighted least squares

transform from eqn. (4.36) in the spectral estimates given by eqns. (5.8) and (5.9), where the weighted normal equation matrix $\boldsymbol{N}_k$ for the k-th spectral component is defined by

$$\boldsymbol{N}_k = \boldsymbol{A}_k^{\mathrm{T}} \boldsymbol{P} \boldsymbol{A}_k \, . \tag{5.12}$$

This type of spectral estimation is referred to here as “independent” or “out-of-context” spectral estimation.

An alternative approach to least squares spectral estimation can be developed in which all spectral components are estimated simultaneously; i.e., in the context of the others being present. This approach takes into account the non-orthogonality (mathematical correlations) between the spectral components. It is effectively equivalent to the geometrical projection of the total multidimensional quadratic form representing the total power, onto the subspace for each individual spectral component. This is analogous to the way in which quadratic forms and confidence regions are defined for station coordinates in geodetic networks. This estimation method is developed by first realizing that in eqn. (5.11) for the total power the inverse of the normal equation matrix $\boldsymbol{N}^{-1}$ is equivalent to the covariance matrix $\boldsymbol{C}_{\hat{x}}$ for the simultaneously estimated Fourier coefficients (cf. eqn. (4.40)). The total power can then be written as

$$\text{Total power} = \boldsymbol{F}^{\mathrm{T}} \boldsymbol{C}_{\hat{x}} \boldsymbol{F} \, . \tag{5.13}$$

Substituting for the weighted least squares transform in eqn. (4.36), the total power can be expressed in terms of the estimated Fourier coefficients:

$$\text{Total power} = \hat{\boldsymbol{x}}^{\mathrm{T}} \boldsymbol{C}_{\hat{x}}^{-1} \hat{\boldsymbol{x}} \, . \tag{5.14}$$

The weighted least squares spectrum is defined as the contribution $s(f_k)$ of the individual frequency components (Fourier coefficients) to the total power. That is, the quadratic form of the estimated Fourier coefficients for individual frequencies is

$$s(f_k) = \hat{\boldsymbol{x}}_k^{\mathrm{T}} \boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1} \hat{\boldsymbol{x}}_k \,, \tag{5.15}$$

where $\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}$ is the k-th diagonal block of covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$. Substituting back in the weighted least squares transform in eqn. (4.37) for individual frequencies, gives the weighted least squares spectral values

$$s(f_k) = \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{C}_{\hat{\boldsymbol{x}}_k} \boldsymbol{F}_k \tag{5.16}$$

which account for any non-orthogonality (mathematical correlations) among the different spectral components. Note that $\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}$ is not the same as $\boldsymbol{N}_k^{-1}$ in the expression for the independently estimated (out-of-context) least squares spectrum. Using the k-th diagonal block from $\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}$ is the same as extracting the k-th diagonal block from $\boldsymbol{N}^{-1}$ instead of from $\boldsymbol{N}$ as in the conventional expressions (cf. Steeves [1981]). Thus, eqn. (5.16) may also be written as

$$s(f_k) = \boldsymbol{F}_k^{\mathrm{T}} (\boldsymbol{N}^{-1})_k \boldsymbol{F}_k \tag{5.17}$$

The normalized spectral value $\tilde{s}(f_k)$ for frequency f_k is obtained by dividing by the total power; i.e.,

$$\tilde{s}(f_k) = \frac{s(f_k)}{\phi^{\mathrm{T}} \boldsymbol{P} \phi} = \frac{\boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{C}_{\hat{\boldsymbol{x}}_k} \boldsymbol{F}_k}{\phi^T \boldsymbol{P} \phi} = \frac{\boldsymbol{F}_k^{\mathrm{T}} (\boldsymbol{N}^{-1})_k \boldsymbol{F}_k}{\phi^T \boldsymbol{P} \phi} \,. \tag{5.18}$$

This type of spectral estimation is referred to here as “simultaneous” or “in-context” spectral estimation.

All linear dependence (mathematical correlation) among the frequency components are accounted for in this simultaneous estimate of the weighted least squares spectrum. When the correlations between the frequency components are ignored, $\boldsymbol{N}$ becomes a diagonal matrix of normal equation matrices $\boldsymbol{N}_k$ for each individual frequency f_k and $\boldsymbol{C}_{\hat{x}_k} = \boldsymbol{N}_k^{-1}$. The expressions given here are then equivalent to those in Steeves [1981], for the independent estimation of spectral components where no deterministic model is considered. When the data are also equally weighted, these expressions are identical to those in Vaníček [1969a; 1972]. When the data are equally spaced and the set of “Fourier” frequencies are used, $\boldsymbol{N}^{-1} = \mathrm{diag}(2/n)$, and the weighted least squares spectral values are then equivalent to the standard one-side Fourier ones given by eqn. (3.41).

Vaníček [1969a; 1972] also includes some simplifying trigonometric identities that make the evaluation of the elements in $\boldsymbol{N}^{-1}$ more efficient for equally spaced data (see also Wells et al. [1985]). These have been omitted from the developments here for the sake of simplicity, although any routine application of these should include these optimizations to reduce the required computational effort.

This approach is also similar to that used by Rochester et al. [1974] in that correlations between different frequencies are accounted for. However, the correlations among the coefficients for same frequency are implicitly ignored in their expressions because of the use of complex notation. The real (cosine) and imaginary (sine) terms for the same frequency are treated independently. Only when the data are equally spaced is their approach equivalent to the preceding ones.

The same comments on the Fourier transform regarding frequencies greater than the Nyquist frequency also apply here for the simultaneous estimate of the fully weighted least squares spectrum. Singularities in $\boldsymbol{N}^{-1}$ should be avoided by using only frequencies up to

the Nyquist frequency. Frequencies that are too closely spaced can also cause ill-conditioning problems in the simultaneous estimation of different spectral values.

Finally, it should be emphasized that these definitions of the least squares spectrum do not satisfy Parseval's relation. That is, the sum of the these spectral values does not equal the total power in eqn. (5.7). Because of the correlation among the frequencies, there is no equivalent to Parseval's relation for unequally spaced data.

5.5 Effect of Deterministic Model

In the developments thus far, the mathematical correlations (linear dependence) between the spectral components and any deterministic model have been ignored, as they are in the traditional Fourier method. One of the most significant contributions of Vaníček [1969a; 1972] was the incorporation of the effect of any a priori deterministic model in the determination of the spectral values. An important consequence (advantage) of this is that it alleviates the need to determine frequency response functions for the deterministic model. In the context of spectrum estimation, frequency response functions are used to account for the effect of the deterministic model on the spectrum. Here, the deterministic effects are modelled explicitly in the formation of the expressions for the estimation of the spectral components.

The effect of the deterministic model on the spectrum is obtained in the same way as for the inverse least squares transform in the previous chapter. The spectrum is defined as the contribution of each frequency component to the total power. This can be expressed in terms of the quadratic form of the estimated Fourier coefficients $\hat{\mathbf{x}}$ as in eqn. (5.15). However, to account for the effects of the deterministic model, the quadratic form must be based on estimates from the combined deterministic and "spectral" model as explained in Section 4.5. That is, the spectral component $s(f_k)$ for frequency f_k is given by

$$s(f_k) \;=\; \hat{\boldsymbol{x}}_k^{\mathrm{T}}\, \boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1}\, \hat{\boldsymbol{x}}_k \;, \tag{5.19}$$

where the matrix components for frequency f_k are, from eqn. (4.62),

$$\hat{\boldsymbol{x}}_k \;=\; \boldsymbol{N}^{*-1}\, \boldsymbol{u}^* \;=\; \boldsymbol{N}^{*-1}\, \boldsymbol{A}_S^{\mathrm{T}}\, \boldsymbol{P}^*\, \boldsymbol{\phi} \;, \tag{5.20}$$

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}_k} \;=\; (\boldsymbol{N}^{*-1})_k \;. \tag{5.21}$$

$\boldsymbol{N}^{*-1}$ and $\boldsymbol{P}^*$ are defined in eqns. (4.64) and (4.61), respectively. Note that these expressions are formally identical to those without a deterministic model, except that the "reduced" weight matrix $\boldsymbol{P}^*$ in eqn. (4.61) is used in place of $\boldsymbol{P}$. The effect of the deterministic model is therefore completely contained within $\boldsymbol{P}^*$.

Following the same substitution procedure as in the previous section, the least squares estimates of the spectral values can be written in terms of the weighted least squares transform $\boldsymbol{F}_k^*$ in eqn. (4.67) as (cf. Vaníček [1971, eqn. (2.4)])

$$s(f_k) \;=\; \boldsymbol{F}_k^{*\mathrm{T}}\, \boldsymbol{C}^*_{\hat{\boldsymbol{x}}_k}\, \boldsymbol{F}_k^* \;=\; \boldsymbol{F}_k^{*\mathrm{T}}\, (\boldsymbol{N}^{*-1})_k\, \boldsymbol{F}_k^* \;. \tag{5.22}$$

The normalized spectrum is defined as before to be the percentage of the variation in the data explained by the each spectral component. In the presence of an a priori deterministic model, this represents the variance explained by each spectral component which is not accounted for by the deterministic model. The part that is not explained by the deterministic model is just the residuals $\boldsymbol{r}_D$ from the deterministic model alone. That is, using the notation of Section 4.5,

$$\boldsymbol{r}_D \;=\; \boldsymbol{\phi} - \boldsymbol{A}_D\, \hat{\boldsymbol{x}}_D \;, \tag{5.23}$$

where

$$\hat{\boldsymbol{x}}_D \;=\; \boldsymbol{N}_{DD}^{-1}\,\boldsymbol{u}_D\,, \tag{5.24}$$

and $\boldsymbol{N}_{DD}$ and $\boldsymbol{u}_D$ are defined by eqns. (4.47) and (4.51), respectively. Expanding $\boldsymbol{r}_D$ and rearranging gives

$$\boldsymbol{r}_D \;=\; (\boldsymbol{I} - \boldsymbol{A}_D\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{A}_D^{\mathrm{T}}\,\boldsymbol{P})\;\phi\,. \tag{5.25}$$

Substituting this in the quadratic form of $\boldsymbol{r}_D$ and simplifying results in

$$\boldsymbol{r}_D^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{r}_D \;=\; \phi^{\mathrm{T}}\,(\boldsymbol{P} - \boldsymbol{P}\,\boldsymbol{A}_D\,\boldsymbol{N}_{DD}^{-1}\,\boldsymbol{A}_D^{\mathrm{T}}\,\boldsymbol{P})\;\phi \;=\; \phi^{\mathrm{T}}\,\boldsymbol{P}^{*}\,\phi\,, \tag{5.26}$$

where $\boldsymbol{P}^{*}$ is the "reduced" weight matrix accounting for the deterministic model. Dividing the spectral values eqns. (5.22) by (5.26), the normalized spectrum that accounts for the deterministic model is

$$\tilde{s}(f_k) \;=\; \frac{s(f_k)}{\phi^{\mathrm{T}}\boldsymbol{P}^{*}\phi} \;=\; \frac{\boldsymbol{F}^{*}{}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{x}_k}\,\boldsymbol{F}^{*}{}_k}{\phi^{T}\;\boldsymbol{P}^{*}\;\phi} \;=\; \frac{\boldsymbol{F}^{*}{}_k^{\mathrm{T}}\,(\boldsymbol{N}^{*-1})_k\,\boldsymbol{F}^{*}{}_k}{\phi^{T}\;\boldsymbol{P}^{*}\;\phi}\,. \tag{5.27}$$

The consideration of which frequencies to include in the weighted least squares spectrum must be done very carefully when accounting for the effects of a deterministic model (it is effectively undefined in the spectrum estimation). This is especially important if periodic trends are present in the deterministic model. In that case, the spectral value for the same frequency is undefined because it has effectively been accounted for in the deterministic model and is therefore undefined in the least squares spectrum; i.e., the periodic component in the deterministic model and the same component in the spectral

model will be perfectly linearly dependent. Evaluating spectral components for the same frequencies as the periodic trends will result in a singular normal equation matrix $\boldsymbol{N}^*$. Present algorithms for the least squares spectrum (e.g., Wells et al. [1985]) check for this situation by inspecting the determinant of $\boldsymbol{N}_k^*$; a zero or near zero value indicates a singularity and thus an undefined spectral value.

Ignoring correlations between spectral components is perfectly acceptable within the context of improving the deterministic model. In this case the objective is to iteratively search for only the largest spectral component in a residual data series from a deterministic model. Any significant spectral values can then be incorporated into the deterministic model, either explicitly as a periodic trend or implicitly as part of a more complex model of the underlying physical processes. In this way the method effectively accounts for the correlations among only the most significant spectral components that are iteratively included in the deterministic model.

5.6 Statistical Tests

Another great advantage of the least squares spectrum is that the significance of the least squares spectral values can be tested statistically in a rigorous manner. The following statistical tests are based on Steeves [1981].

It is well known in statistics that a quadratic form has chi-square distribution with degrees of freedom equal to the rank of the weight matrix. Expressing the estimated spectral values in terms of the quadratic form of the estimated Fourier coefficients $\hat{\boldsymbol{x}}_k$ in eqn. (5.15), this quantity then has a Chi-square distribution $\chi^2(u; 1-\alpha)$ with u=2 degrees of freedom (representing the rank of the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}$ for the two Fourier coefficients for frequency f_k) [Vaníček and Krakiwsky, 1986]. A statistical test of the null hypothesis H_o: $s(f_k) = 0$ can then be made using the decision function

$$s(f_k) \begin{cases} \leq \chi^2(2;\ 1-\alpha);\ \text{accept } H_o\text{: } s(f_k) = 0 \\ > \chi^2(2;\ 1-\alpha);\ \text{reject } H_o \end{cases} \tag{5.28}$$

where α is the significance level of the test (usually 5%).

If the scale (i.e., a priori variance factor σ_o^2) of $\boldsymbol{C}_\phi$ is unknown, the estimated value $\hat{\sigma}_o^2$ can be obtained from

$$\hat{\sigma}_o^2 = \frac{\hat{\boldsymbol{r}}^T \boldsymbol{P} \hat{\boldsymbol{r}}}{\upsilon}, \tag{5.29}$$

where $\upsilon = n-2$ is the degrees of freedom (two degrees of freedom lost to the estimation of the two Fourier coefficients). This estimated variance factor is used to scale the covariance matrix $\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}$, which then has a Fisher distribution $F(\upsilon,u;\ 1-\alpha)$ with $\upsilon=n-2$ and $u=2$ degrees of freedom. A statistical test of the null hypothesis H_o: $s(f_k) = 0$ can then be made using the decision function

$$s(f_k) \begin{cases} \leq F(\upsilon,2;\ 1-\alpha);\ \text{accept } H_o\text{: } s(f_k) = 0 \\ > F(\upsilon,2;\ 1-\alpha);\ \text{reject } H_o \end{cases} \tag{5.30}$$

The distribution of the normalized spectral values is obtained by first rewriting the quadratic form $\phi^T \boldsymbol{P} \phi$ in terms of the residuals $\hat{\boldsymbol{r}}$ and estimated observations $\hat{\phi}$ from the spectral model. Realizing that

$$\boldsymbol{r} = \phi - \boldsymbol{A} \hat{\boldsymbol{x}} = (\boldsymbol{I} - \boldsymbol{A}\, \boldsymbol{N}^{-1} \boldsymbol{A}^T \boldsymbol{P})\, \phi, \tag{5.31}$$

the quadratic form of the residuals can be expressed as

$$\boldsymbol{r}^T \boldsymbol{P}\, \boldsymbol{r} = \phi^T (\boldsymbol{P} - \boldsymbol{P}\, \boldsymbol{A}\, \boldsymbol{N}^{-1} \boldsymbol{A}^T \boldsymbol{P})\, \phi = \phi^T \boldsymbol{P}^*\, \phi. \tag{5.32}$$

Noting that $\boldsymbol{P} = \boldsymbol{C}_{\phi}$ and rearranging,

$$\begin{aligned} \boldsymbol{r}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{r} &= \phi^{\mathrm{T}}\,\boldsymbol{P}\,\phi - (\phi\,\boldsymbol{P}\,\boldsymbol{A})\,\boldsymbol{N}^{-1}\,(\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{P}\,\phi) \\ &= \phi^{\mathrm{T}}\,\boldsymbol{P}\,\phi - \hat{\boldsymbol{x}}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1}\,\hat{\boldsymbol{x}}_k\ . \end{aligned} \tag{5.33}$$

Thus, the quadratic form of the observations is

$$\boldsymbol{f}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{f} = \hat{\boldsymbol{x}}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1}\,\hat{\boldsymbol{x}}_k + \boldsymbol{r}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{r}\ , \tag{5.34}$$

which represent the total power. The quadratic forms on the right side of eqn. (5.34) are well known (see, e.g., Vaníček and Krakiwsky [1986]). The quadratic form $\hat{\boldsymbol{x}}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1}\,\hat{\boldsymbol{x}}_k$ of the estimated Fourier coefficient has a Chi-square distribution with 2 degrees of freedom (the number of Fourier coefficients for frequency f_k). The quadratic form $\boldsymbol{r}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{r}$ of the residuals has a Chi-square distribution with $\upsilon = n - u$ degrees of freedom, where u is the total number of Fourier coefficients being simultaneously estimated (if the spectral values are being estimated independently, then $u=2$).

Using eqns. (5.15) and (5.34) in the expression for the normalized spectral value in eqn. (5.18) and rearranging gives

$$\tilde{s}(f_k) = \frac{\hat{\boldsymbol{x}}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1}\,\hat{\boldsymbol{x}}_k}{\hat{\boldsymbol{x}}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1}\,\hat{\boldsymbol{x}}_k + \boldsymbol{r}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{r}} = \frac{1}{1 + \dfrac{\boldsymbol{r}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{r}}{\hat{\boldsymbol{x}}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{\boldsymbol{x}}_k}^{-1}\,\hat{\boldsymbol{x}}_k}}\ , \tag{5.35}$$

where the ratio of two quadratic forms in the denominator has the following Fisher distribution

$$\frac{\boldsymbol{r}^{\mathrm{T}}\,\boldsymbol{P}\,\boldsymbol{r}}{\hat{\boldsymbol{x}}_k^{\mathrm{T}}\,\boldsymbol{C}_{\hat{x}_k}^{-1}\,\hat{\boldsymbol{x}}_k} \rightarrow \frac{\upsilon}{2} F_{\upsilon,2;\,\alpha} \,. \tag{5.36}$$

where "$\rightarrow$" means "is distributed as", and υ and 2 are the degrees of freedom of the numerator and denominator, respectively. Note the use of the α probability level instead of $1-\alpha$. This is because of the inverse relation between this F statistic and the spectral value (for which we want the $1-\alpha$ probability level). Given the distribution of the ratio of the quadratic forms in eqn. (5.35), the distribution of the normalized spectral value is then (cf. Steeves [1981, eqn. (3.19)])

$$\tilde{s}(f_k) \rightarrow \left(1 + \frac{\upsilon}{2} F_{\upsilon;2;\,\alpha}\right)^{-1} . \tag{5.37}$$

A statistical test of the null hypothesis $\mathrm{H_o}$: $\tilde{s}(f_k) = 0$ can then be made using the decision function

$$\tilde{s}(f_k) \begin{cases} \leq \left(1 + \frac{\upsilon}{2} F_{\upsilon;2;\,\alpha}\right)^{-1}; \text{ accept } \mathrm{H_o}\text{: } s(f_k) = 0 \\ > \left(1 + \frac{\upsilon}{2} F_{\upsilon;2;\,\alpha}\right)^{-1}; \text{ reject } \mathrm{H_o} \end{cases} . \tag{5.38}$$

The above Fisher distribution can be simplified further using the inverse relation for the Fisher distribution [Freund, 1971],

$$F_{\upsilon,2;\,\alpha} = F_{2,\upsilon;\,1-\alpha}^{-1} \,. \tag{5.39}$$

When the first degree of freedom is two, this can be approximated by [Steeves, 1981],

$$F_{2,\upsilon;\,1-\alpha} \approx \frac{\upsilon}{2}\left(\alpha^{-2/\upsilon} - 1\right) . \tag{5.40}$$

This results in a statistical test of the null hypothesis H_o: $\tilde{s}(f_k) = 0$ using the decision function

$$\tilde{s}(f_k) \begin{cases} \leq \left(1 + \left(\alpha^{-2/\upsilon} - 1\right)^{-1}\right)^{-1}; \text{ accept } H_o\text{: } s(f_k) = 0 \\ > \left(1 + \left(\alpha^{-2/\upsilon} - 1\right)^{-1}\right)^{-1}; \text{ reject } H_o \end{cases} . \qquad (5.41)$$

The statistical tests for spectral values that account for the presence of any deterministic model are exactly the same as above, except that the "reduced" observation weight matrix $\boldsymbol{P}^*$ is used in place of the actual weight matrix $\boldsymbol{P}$ in the computation of the quadratic forms.

The above tests are the so-called "out-of-context" tests, which test the individual spectral components out of context of the others being estimated (see Vaníček and Krakiwsky [1986, p. 229-231]). They are identical to those in Steeves [1981] and apply to the independent estimation of the spectral components, but not to the estimation of all the spectral values simultaneously. In that case the "in-context" test should be used which takes into consideration the estimation of the other spectral components. Two approaches can be used in this regard. The simplest one is to use the simultaneous confidence region for all m frequency components being estimated. This gives the same test as in eqn. (5.41) except that $2m$ degrees of freedom is used in place of 2. However, this approach usually results in too pessimistic (large) a limit to be any real value. A better approach is to use the relation between the simultaneous probability α for the joint test of all spectral components together and the "local" probability α_o for the test of each spectral component separately. Following Miller [1966], the relation is given to first-order approximation by $\alpha_o \approx \alpha/m$. The in-context test is then obtained by using α_o in place of α in the above tests. Note that Press and Rybicki [1989], Press et al. [1992, p. 570] also use the in-context test based on

simultaneous probability. However, they incorrectly apply it to the testing of the independently estimated spectral components, where the correlations among the different frequency components is ignored. The in-context test should only be used for the simultaneous estimates of the spectral values, where the correlations among all the frequencies used is accounted for.

5.7 Estimation Algorithms

As stated at the beginning of this chapter, there have been a variety of papers since Vaníček, [1969a, 1971] describing the same least squares spectrum (independently estimated spectral components) in slightly different forms; e.g., Lomb [1975], Ferraz-Mello [1981], Scargle [1982], Horne and Baliunas [1986]. It can be shown, however, that under the same assumptions all of these are identical to Vaníček's more general approach. The differences are only the use of slightly different normalization methods and different numerical methods for solving the normal equations.

In Vaníček [1969a], the direct inversion of the 2x2 normal equation matrix is optimized by using an analytical expression. In addition to being the fastest algorithm, it also accounts for the presence of a priori deterministic models and includes various trigonometric identities for greater efficiency, especially for equally spaced data. Compared to the FFT, however, the least squares transform and spectrum are computationally much slower. Unfortunately, a direct comparison of computational speed could not be made because of the software used. All tests were performed using the MATLAB software, which has a built-in (compiled) FFT function optimised for speed whereas the least squares spectrum algorithm was implemented as an external (interpreted) function. Because external functions execute much more slowly than built-in functions, no fair comparison between the FFT and least squares algorithms could be made in MATLAB. Nevertheless,

when confronted with unevenly spaced data, the least squares method is the only correct approach to use.

Lomb [1975] and Scargle, 1982] solve the normal equations using an orthogonalization (diagonalization) procedure based on time shifting (a different time shift is needed for each frequency). This approach is slower than the direct analytical solution of Vaníček. It also does not account for the presence of any a priori models, except for a mean. Ferraz-Mello [1981] uses Gram-Schmidt orthogonalization to diagonalize the normal equations. Again, this procedures is slower than direct analytical inversion and does not account for the presence of any a priori deterministic models.

Recently, Press and Rybicki [1989] have developed a novel approach to the fast computation of a least squares spectrum. It is based on the concept of "extirpolation" and the use of the FFT. Basically, extirpolation gives an equally spaced data series that, when interpolated to the original times, gives back exactly the original data series. This is also called reverse interpolation. The FFT is used to evaluate the evenly spaced (extirpolated) sine and cosine summations in the time-shifting algorithm of Lomb [1975]. The original extirpolation algorithm used two complex FFTs. The more efficient algorithm uses the same trigonometric identities used by Vaníček [1969a] to reduce the computations to only one FFT. The biggest disadvantage of this method is that it's limited to only the set of "Fourier" frequencies due to the use of the FFT. It is thus not possible to "zoom in" on significant peaks to better resolve the frequency. The FFT also requires 2^n data points, which necessitates zero-padding the data series. As for the other algorithms, the presence of a priori deterministic models cannot be accounted for. Finally, the extirpolation accuracy depends on the "oversampling factor" used in the extirpolation to generate many more data points than the original data series. Greater oversampling of the extirpolated series provides better accuracy but results in more computations. In spite of the above limitations, this algorithms works very well and very fast (on the order of $n \log n$, instead of n^2).

Chapter 6
Stochastic Modelling of Observation Errors

6.1 Introduction

The weighted least squares estimation model allows for the stochastic modelling of residual errors through the use of a fully populated covariance matrix. This can be used to account for those systematic effects that have not been modelled explicitly (deterministically) in the design matrix for the least squares model. The problem with using fully populated covariance matrices in this manner is the difficulty in determining the covariance or correlations among the observations in an objective way.

There are a few methods that can be used to determine the variance and covariance each with their own advantages and drawbacks. One of the most popular of these are the methods of analysis of variance and variance-covariance component estimation. The "analysis of variance" (ANOVA) method (also called factor analysis in statistics) can be found in most standard texts on statistics. Geodetic applications of the technique are described in detail by Kelly [1991] and in a series of articles by Wassef [1959; 1974; 1976]. Essentially the aim of the method is to divide the measurements into separate groups (factors which contribute to the overall variation in the data) and to estimate the variance components for each. The difficulty in applying the method is in defining a scheme of dividing the observations into separate groups which characterize some behaviour of the systematic effect being modelled. Often, the factors describing the systematic effect cannot be so discretely defined, rather they are often of a continuous nature that precludes lumping them together into separate and distinct groups.

Variance-covariance component estimation, on the other hand, is based on modelling deterministically the residual variation in the measurements. The variances and covariances are expressed in terms of linear models relating these components to various factors describing the systematic effect. The coefficients (variance and covariance components) in the variance-covariance model are estimated together with the parameters in a least squares solution. The technique is described in detail in Rao and Kleffe [1988] and has been applied to many geodetic problems (see, e.g., Grafarend et al. [1980], Grafarend [1984], Chen et al. [1990]). It can be shown that the analysis of variance method is just a special case of this more general approach [Chrzanowski et al., 1994]. The problem with applying the method is that the estimation of the variance-covariance model coefficients usually needs to be iterated which can result in biased estimates of the variances and covariances [Rao and Kleffe, 1988]. This can lead to negative variances, which is unacceptable.

The approach taken here is to model any residual systematic effects remaining after accounting for a deterministic model, using autocorrelation (ACF) or autocovariance (ACvF) functions derived from a power spectral density function of the residuals. This idea was first proposed for geodetic applications by Vaníček and Craymer [1983a; 1983b] and further developed by Craymer [1984]. To accommodate unevenly spaced data, a general least squares transform is developed to determine the normalized power spectrum. The inverse transform is then used to convert this to an ACF which is converted to an ACvF.

6.2 Direct Autocovariance Function Estimation

The autocovariance function of an equally spaced data series $l(t_i)$ can be estimated directly using the expressions given in Chapter 2. This gives the sample autocovariance function

$$C(\tau_m) = \frac{1}{n-m} \sum_{i=1}^{n-m} \big(l(t_i)-\mu\big)\big(l(t_i+\tau_m)-\mu\big) , \qquad (6.1)$$

where $m = \tau_m/\Delta t$ is the so-called lag number and Δt is the data series spacing. Note that, as in eqn. (2.20), the summation is divided by $n–m$ rather than by n, in order to provide an unbiased estimate of $C(\tau_m)$. The biased estimate is obtained by dividing by n.

For unequally spaced data which are relatively homogeneously distributed, an averaging procedure can be used. In this approach the unevenly spaced lags are divided into equally spaced lag intervals or bins, similar to the way in which histograms are constructed. All lags within the lag interval are summed together in (6.1) to give an average autocovariance for the lag interval. This method gives a smoothed estimate of the autocovariance function. The problem is that if the data have large gaps, the lag intervals may need to be relatively large, resulting in degraded resolution. See Vaníček and Craymer [1983a;b] and Craymer [1984] for more details of this technique.

6.3 Autocovariance Function Estimation via the Spectrum

The autocovariance function for an evenly spaced data series can be most conveniently derived from the power spectral density function using the Fourier transform. As discussed in Section 3.4, the autocovariance function can be expressed as the Fourier transform pair with the spectrum, and the autocorrelation function $R(t)$ as the transform pair with the normalized spectrum. These expressions in terms of the spectrum are often used as the basis for the efficient computation of autocovariance and autocorrelation functions of evenly spaced data using the FFT. It will also be used as the basis for developing autocovariance functions for unevenly spaced data to provide objective a priori estimates of

covariances and weights that account for residual systematic effects in least squares modelling.

As mentioned in Section 3.4, care must be exercised to avoid any “wrap around” or “end” effects when computing the autocovariance or autocorrelation function from the spectrum. This is most easily achieved by simply padding the data series with zeros out to double the length of the original series. Furthermore, this indirect estimation via the spectrum provides the biased estimate of the autocovariance/autocorrelation function. As recommended by Bendat and Piersol [1971, pp. 312-314] and Priestley [1981, pp. 323-324], this should be used in preference to the unbiased estimate because the biased one is a positive definite function which generates a positive definite covariance matrix. The unbiased ACF and ACvF are not positive definite and result in singular covariance matrices that are not suitable for generating weight matrices for least squares models.

6.4 Iteratively Reweighted Least Squares Estimation

The covariance matrix generated from the autocovariance function is used to stochastically model the residual errors in the deterministic least squares model. The basic idea is to begin with some a priori estimate of the covariance matrix, usually a diagonal matrix of known variances. A least squares solution is obtained for the deterministic model and the observation residuals provide an estimate of the random observation errors. The autocorrelation function is determined for these residuals in order to obtain a more realistic estimate of the correlations among the random observations errors. This autocorrelation function is then used together with the a priori variances to generate a new covariance function for the observations which is included in a new least squares solution for the deterministic model and new estimate of the residual observation errors. Another autocorrelation function is then computed and the whole estimation process is repeated (iterated) until the solution for the deterministic model and covariance matrix converge to a

stable form. This is referred to an iteratively reweighted least squares estimation and is identical to the iterated MINQUE technique except that a deterministic model is used there to model the variances and covariances (see Rao and Kleffe [1988]). The procedure is illustrated schematically in Figure 6.1.

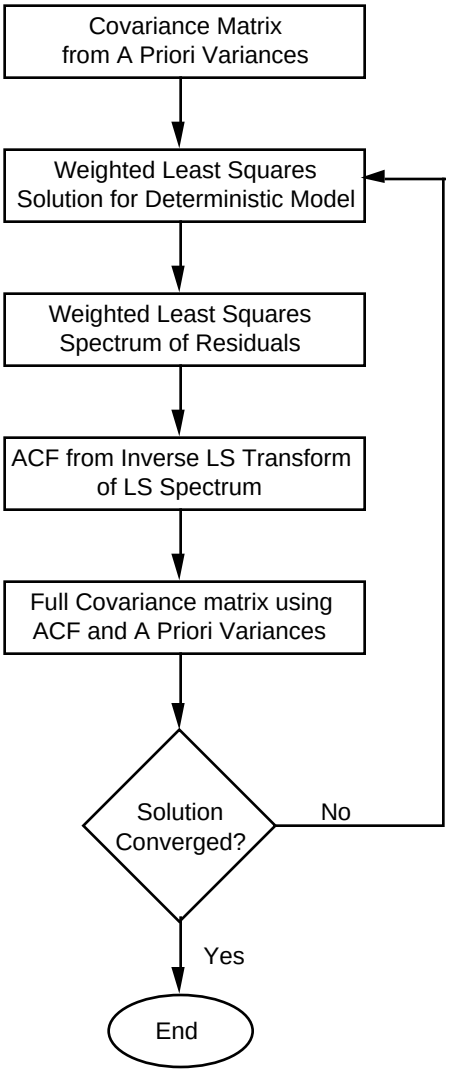

Figure 6.1: Iteratively reweighted least squares estimation process.

Chapter 7
Numerical Tests

7.1 Introduction

In this chapter, various numerical tests of the least squares transform and spectrum are given under a variety of different situations. Throughout, the following terminology and notation is used:

"Fourier" frequencies	Set of integer multiples of the fundamental frequency
LST	Least squares transform
ILST	Inverse least squares transform
LSS	Least squares spectrum
Independent LSS/ILST	Independent estimation of the LSS or ILST frequency components
Simultaneous LSS/ILST	Simultaneous estimation of the LSS or ILST frequency components
Unweighted LSS/ILST	Estimation of LSS or ILST using equally weighted observations (no weight matrix $\boldsymbol{P}$ used)
Weighted LSS/ILST	Estimation of LSS or ILST using weighted observations (weight matrix $\boldsymbol{P}$ used)
ACF	Autocorrelation function
Indirect ACF	Indirect estimation of the autocorrelation function via the ILST or the LSS

The tests presented here are based on simulated data using a pseudo-random number generator for normally distributed observation errors and uniformly distributed, unequally spaced times. Unless otherwise stated, these tests use a deterministic model consisting of a periodic trend with period 10 (frequency 0.1 Hz). All computations were performed using the MATLAB numerical and graphical software system.

Tests were performed to ascertain the effects of the following on the LSS and indirect estimation of the ACF:

- random observation errors
- correlations among observations
- random sampling (unequally spaced data)
- frequency selection
- deterministic model
- non-stationary random errors (random walk)

The effects on the LSS and ACF were determined by comparing the results to the known theoretical form for both functions.

7.2 Effect of Random Observation Errors

To study the effect of random observation errors, three data series of 100 equally spaced points were used. Each was composed of a periodic trend of amplitude 1 and period 10, i.e., frequency 0.1 Hz. The first series contained no observation errors. The second series contained normally distributed random errors with a standard deviation of 1/3. The third data series contained normally distributed random errors with a standard deviation of 2/3. The three data series are plotted in Figure 7.1.

The least squares spectra (for "Fourier" frequencies) of the three data series are given in Figure 7.2. Both the independently and simultaneously estimated spectral values

will be identical in these tests because the data are equally spaced, equally weighted and the set of "Fourier" frequencies is used. The effect of random observation errors on the LS spectrum is to reduce the magnitude of the largest spectral peak, which in all cases is correctly located at the frequency of the periodic trend. The larger the random error, the greater the reduction in the spectral value for the significant peak. The magnitude of the reduction in the peaks is equivalent to the inverse of the square of the signal to noise ratio (ratio of amplitude of periodic signal to standard deviation of noise).

The direct estimates of the autocorrelation functions for the three data series are given in Figure 7.3. These are unbiased estimates and were estimated using eqns. (2.20) and (2.12). The ACFs all exhibit the expected cosine form. However, the functions all display correlations larger than one at large lags, typical of the unbiased form. As explained in Section 3.4, this so-called "wild" behaviour is the main reason the unbiased estimate is not used.

The biased estimates of the autocorrelation functions are given in Figures 7.4 to 7.6 for the three data series, respectively. Both the direct estimate and the indirect estimate via the inverse LS transform of the LS spectrum are given as well as the difference between the two. The indirect estimates were derived following the procedure described in Section 6.3, where zero-padding is used to avoid any "wrap around" effects (see Section 3.4). As expected, all three ACFs exhibit the correct sinusoidal shape and tapering characteristic of the biased estimate. However, there is a reduction in the magnitude of the correlation as the random error increases. Although the differences between the direct and indirect estimates get larger in direct proportion to the magnitude of the random error, they are negligible for all three data series.

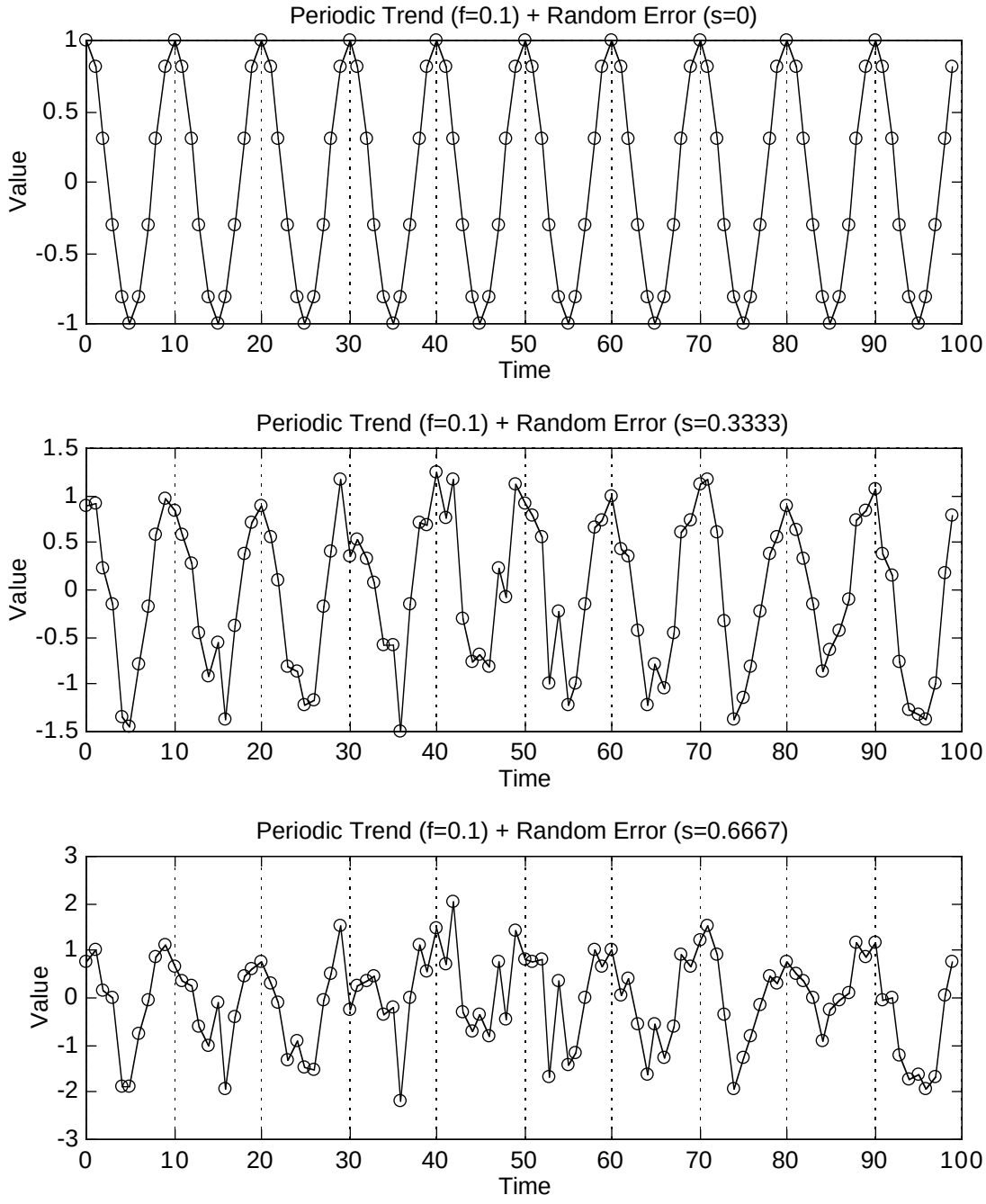

Figure 7.1: Periodic time series of 100 equally spaced points and period 10 (frequency 0.1 hz) with no observation errors and with normally distributed random errors (standard deviations 1/3 and 2/3).

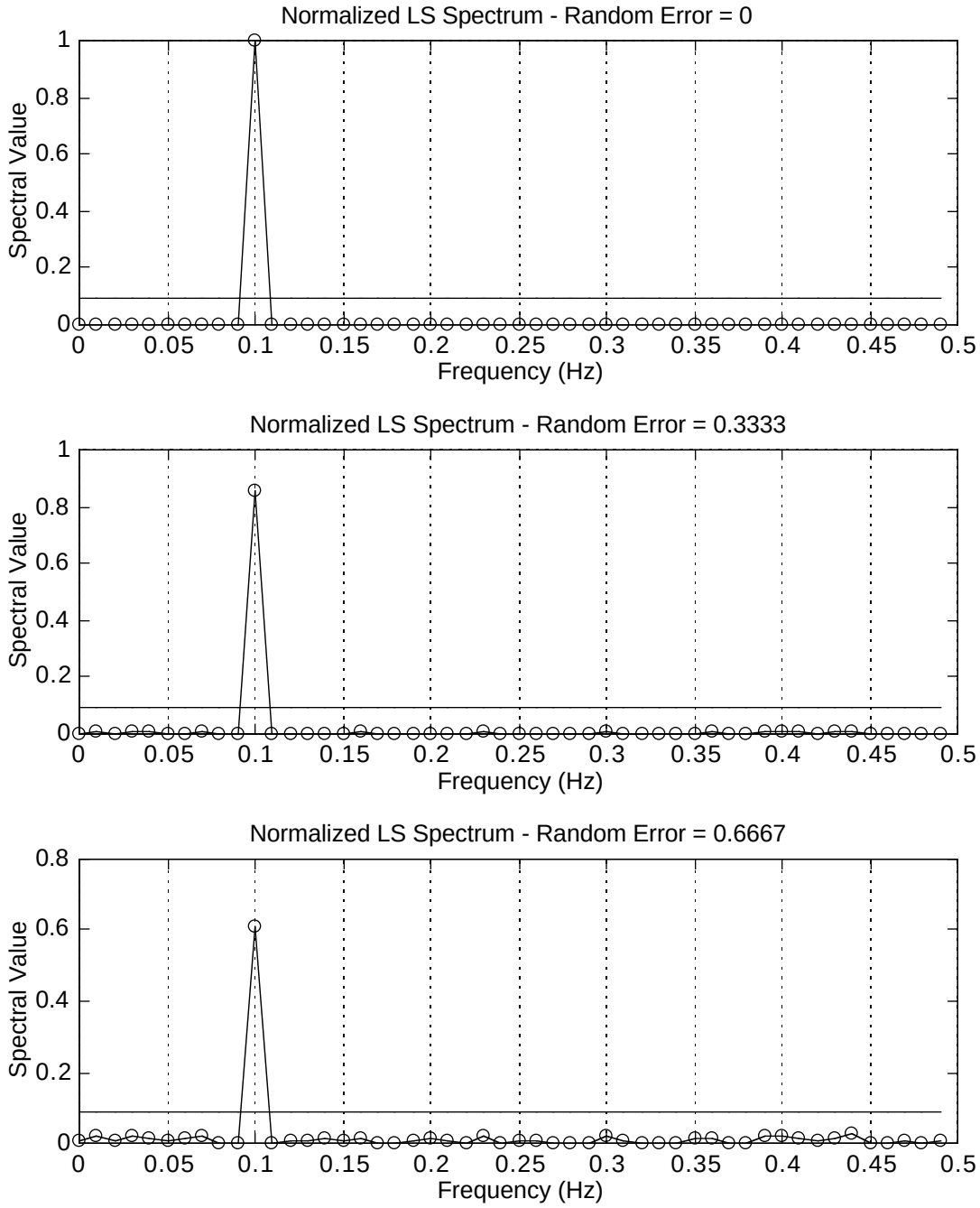

Figure 7.2: Least squares spectra of time series of 100 equally spaced points and period 10 (frequency 0.1) with no observation errors and with normally distributed random errors (standard deviations 1/3 and 2/3). The horizontal line indicates the 95% confidence limit for statistically significant spectral peaks.

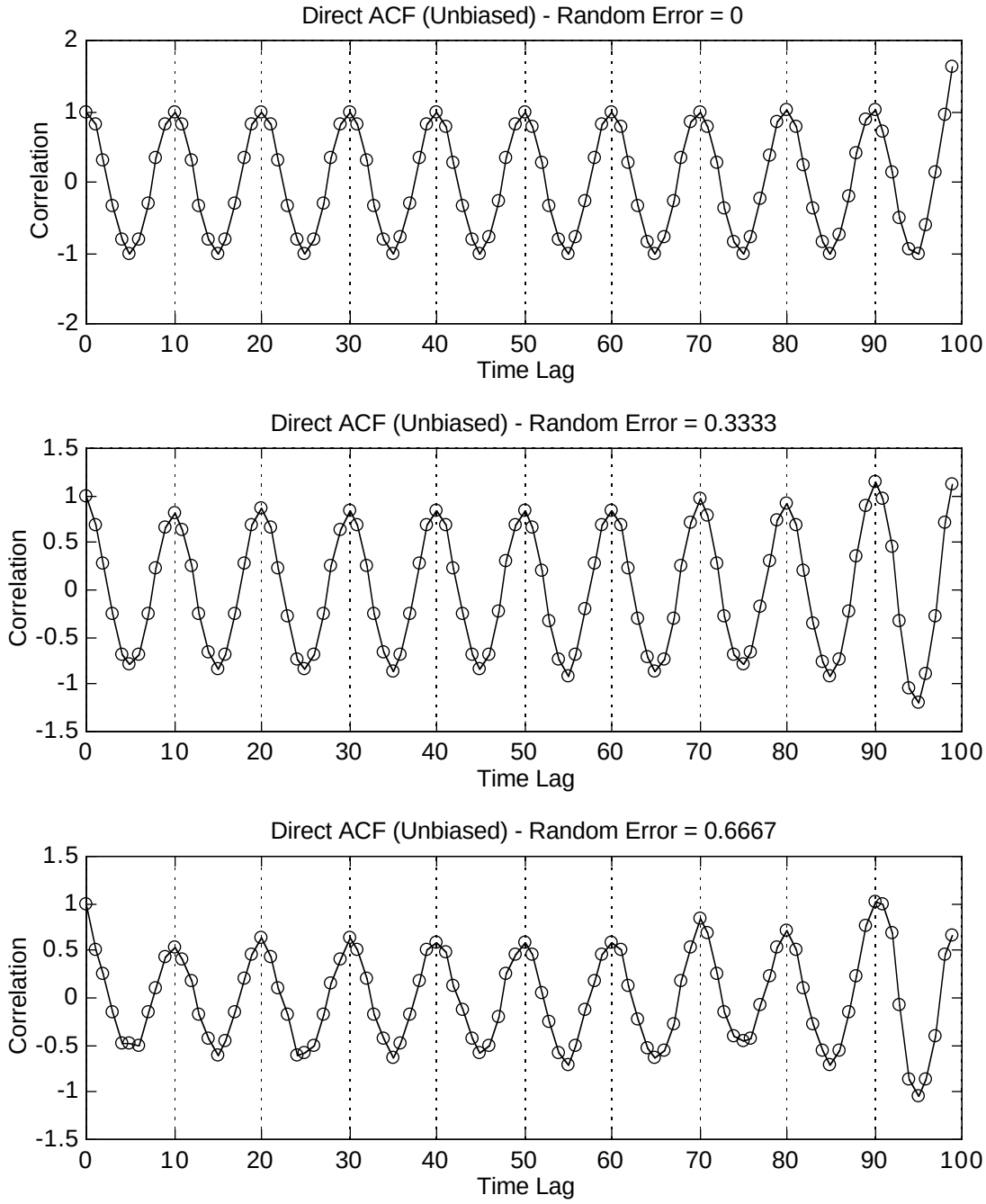

Figure 7.3: Direct estimation of unbiased autocorrelation functions of time series of 100 equally spaced points and period 10 (frequency 0.1) with no observation errors and with normally distributed random errors (standard deviations 1/3 and 2/3).

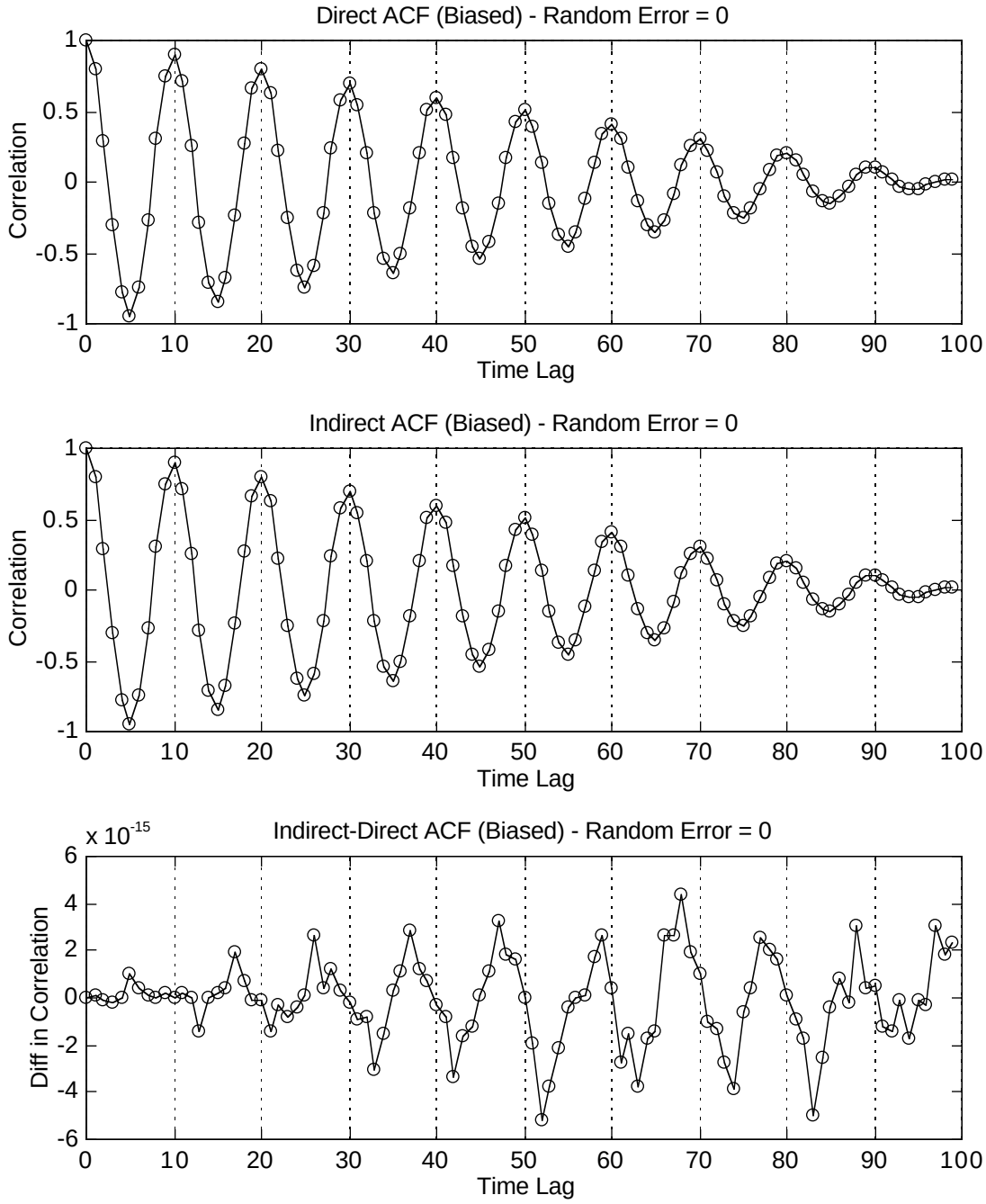

Figure 7.4: Comparison of direct and indirect (via LS spectrum) estimation of biased autocorrelation functions of time series of 100 equally spaced points and period 10 (frequency 0.1) with no observation errors.

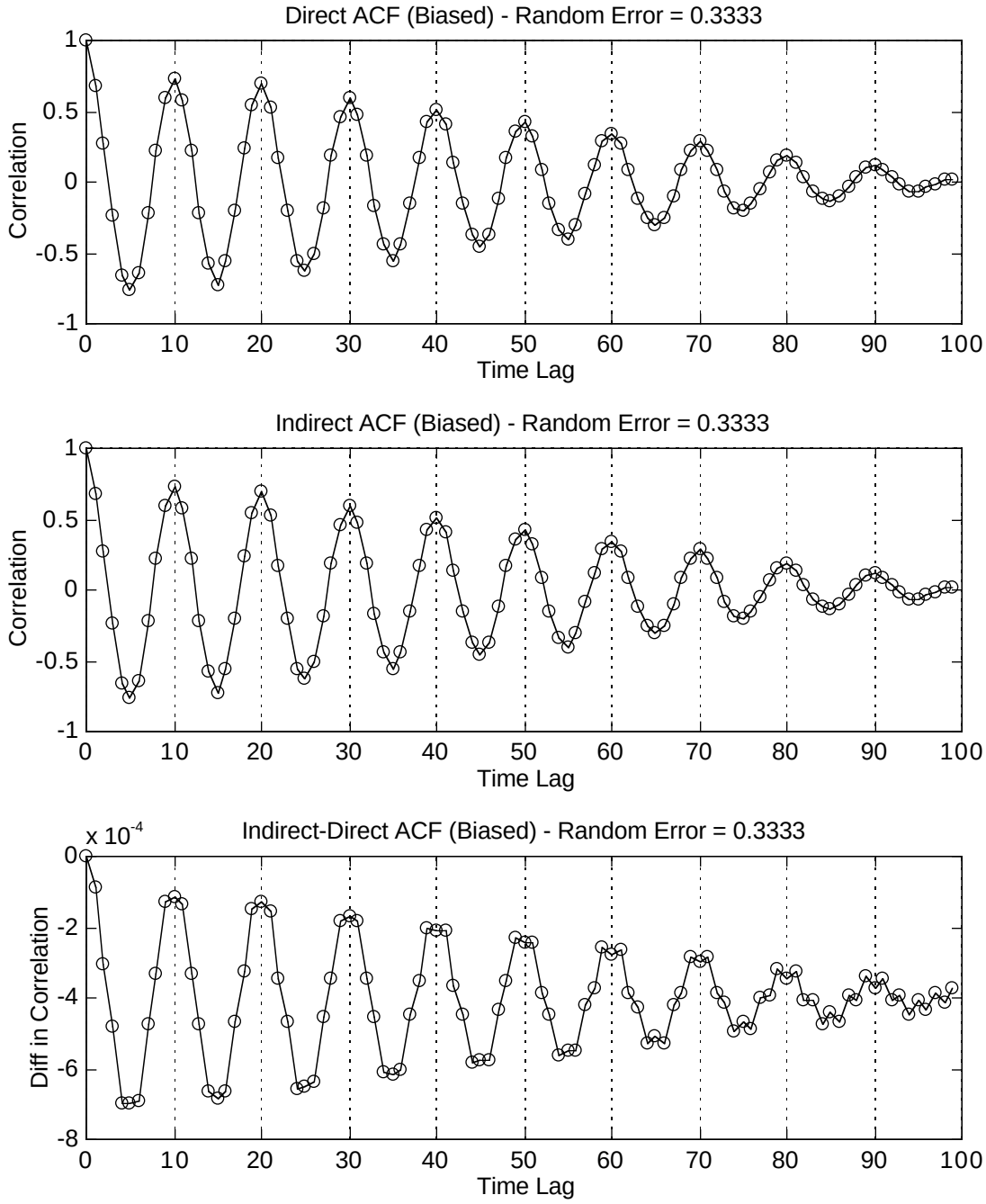

Figure 7.5: Comparison of direct and indirect (via LS spectrum) estimation of biased autocorrelation functions of time series of 100 equally spaced points and period 10 (frequency 0.1) with random observation errors (standard deviation 1/3).

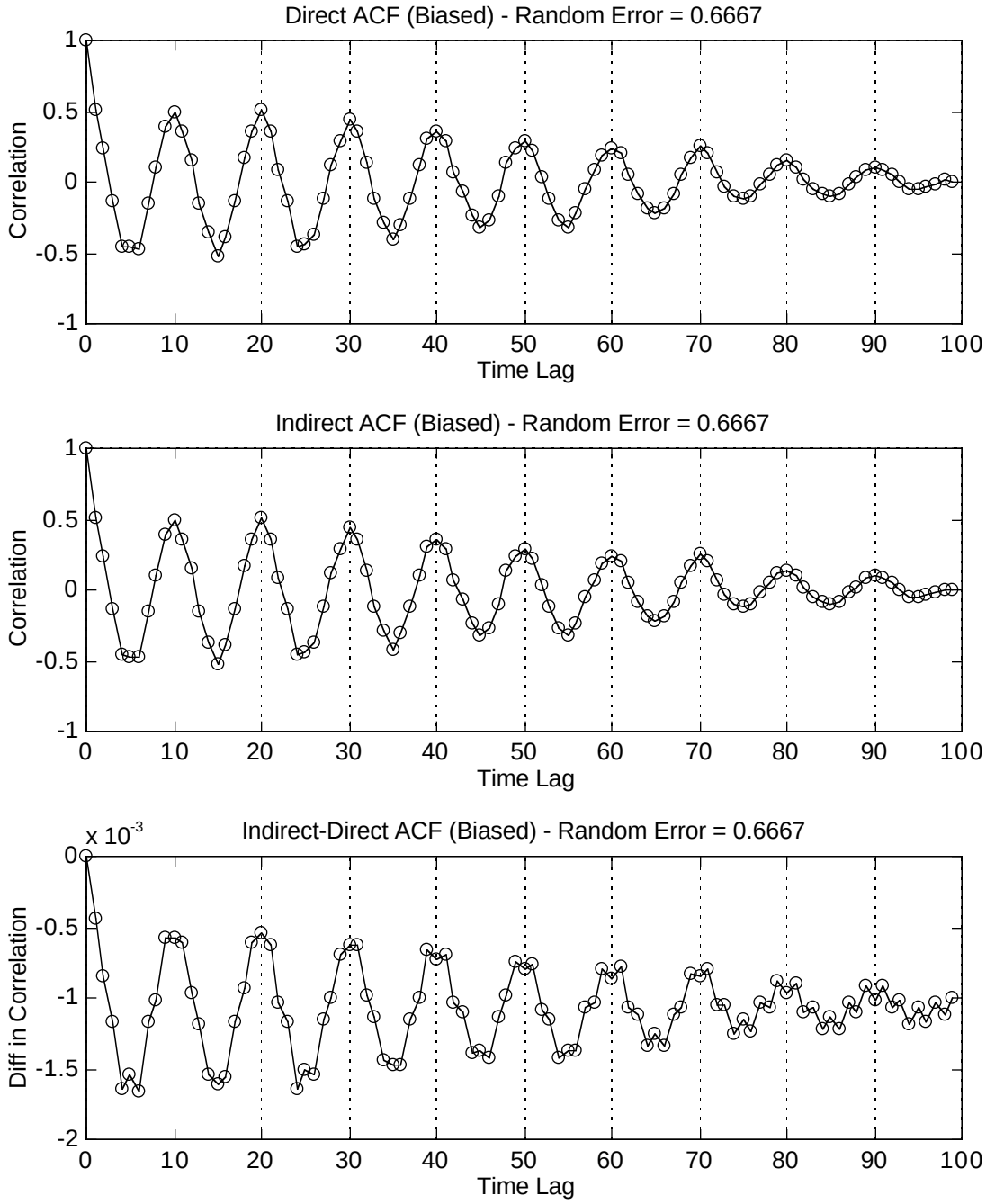

Figure 7.6: Comparison of direct and indirect (via LS spectrum) estimation of biased autocorrelation functions of time series of 100 equally spaced points and period 10 (frequency 0.1) with random observation errors (standard deviation 2/3).

7.3 Effect of Correlated Random Errors

To test the effect of correlations among the random observation errors, it is necessary to generate a correlated set of errors ε. This can be accomplished by finding a transformation $\boldsymbol{L}$ of a set of uncorrelated random errors η with diagonal covariance matrix $\boldsymbol{C}_\eta$, which, by the law of propagation of errors, gives a set of correlated random errors ε with the desired covariance matrix $\boldsymbol{C}_\varepsilon$, i.e., for identically normally distributed random errors ($\boldsymbol{C}_\eta=\boldsymbol{I}$),

$$\boldsymbol{C}_\varepsilon = \boldsymbol{L}\,\boldsymbol{C}_\eta\,\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{L}\,\boldsymbol{L}^{\mathrm{T}} . \tag{7.1}$$

The above decomposition (factorization) of a matrix into another matrix times the transpose of itself is known as Cholesky decomposition, where $\boldsymbol{L}$ is a lower triangular matrix called the Cholesky triangle or square root [Dahlquist and Björck, 1974, p. 158; Golub and Van Loan, 1983, pp. 88; Press et al., 1992, pp. 89]. Using the Cholesky triangle, the transformed set of correlated random errors can then be obtained from

$$\varepsilon = \boldsymbol{L}\,\eta . \tag{7.2}$$

In the following tests, the periodic data from the previous section is used with a standard deviation of 2/3. A fully populated covariance matrix for the observations was constructed from the autocorrelation function

$$\rho(t_i,t_j) = e^{-\Delta t^2/25} . \tag{7.3}$$

where $\Delta t = t_j - t_i = 1$. A plot of the time series and correlation function are given in Figure 7.7 using a standard deviation of 2/3.

Three different types of least squares spectrum were computed for this data series: (1) the unweighted independent estimate, (2) the weighted independent estimate, and (3) the weighted simultaneous estimate. The different spectra all provide good results, each clearly identifying the periodic component correctly at frequency 0.1 (see Figure 7.8). Although the unweighted independent LS spectrum displays slightly larger noise at the lower frequencies than the other spectra, the noise is well within the 95% confidence interval. The weighted LS spectra provide almost identical results, although the peak at frequency 0.1 is slightly larger. These results verify the claim by Steeves [1981] that correlations among the observations have little effect on the resulting spectra.

The direct and indirect (via the unweighted inverse LS transform of the unweighted LS spectrum) estimates of the autocorrelation function are given in Figure 7.9. The two ACFs are identical and agree well with the expected form for the periodic data set (see Figure 7.6), although those here display slightly larger correlations at lower frequencies due to the a priori correlation function. The weighted indirect ACFs are shown in Figure 7.10. Both exhibit the correct shape for the periodic signal, but that based on the independently estimated spectrum gives larger correlations than for the unweighted estimates. On the other hand, the ACF based on the simultaneously estimated spectrum displays much smaller correlations and thus gives the poorest estimate of the ACF.

Another check on the estimation of the autocorrelation functions was performed by computing the ACFs only for the correlated errors (the periodic signal was not included). The ACFs should agree closely with the a priori one used in constructing the correlated errors (see bottom plot of Figure 7.7). Figure 7.11 shows both the direct and indirect (via the unweighted inverse LS transform of the unweighted LS spectrum) estimates of the biased autocorrelation function. Both are identical and agree well with the theoretical correlation function in Figure 7.7. The departures from the true ACF are due to the limitations of the random number generator. The indirect weighted estimates via the inverse weighted LS transform of both the independently and simultaneously estimated LS

spectra are given in Figure 7.12. All these ACFs display the same shape, except for the weighted simultaneous estimate which has slightly larger correlations.

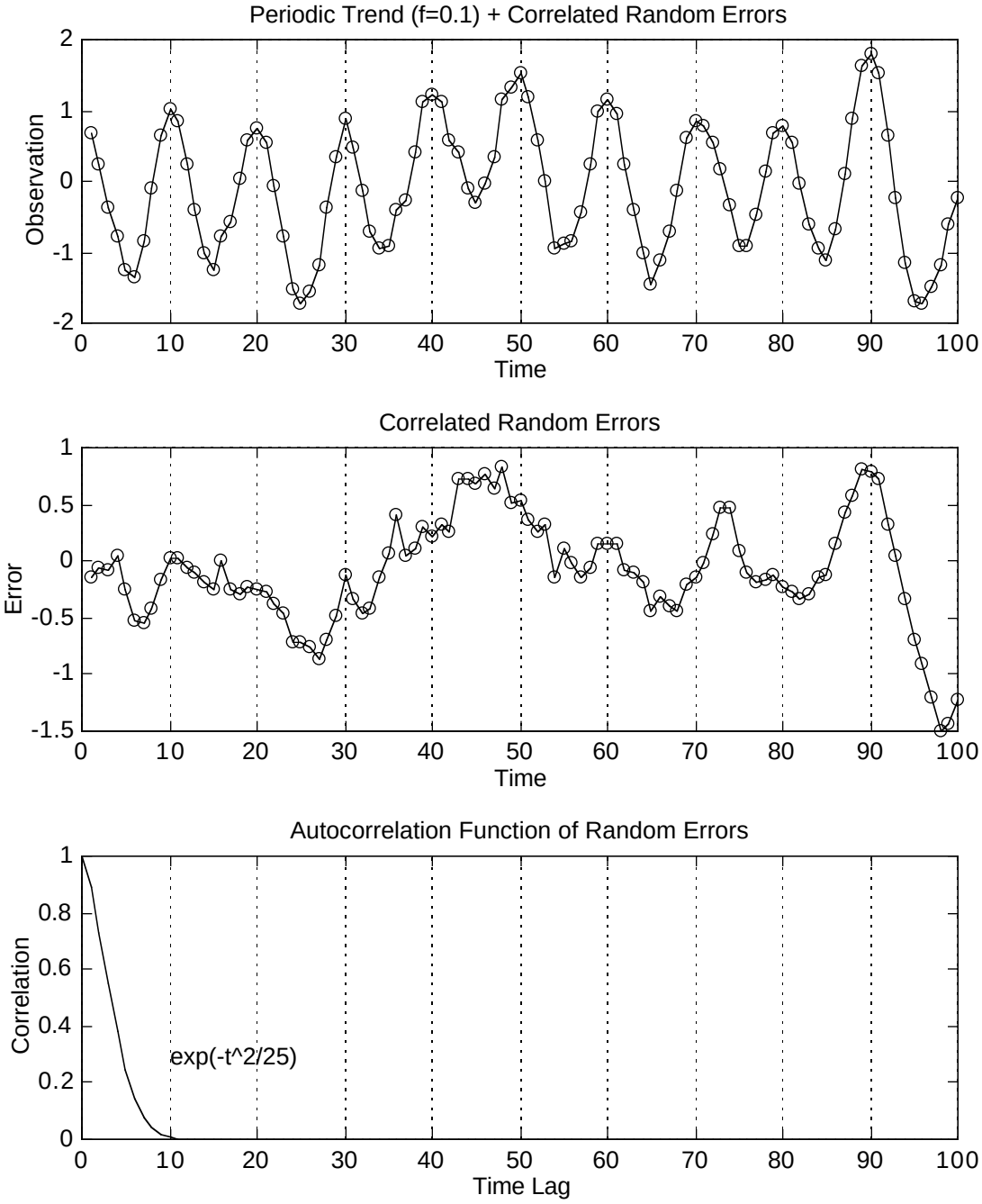

Figure 7.7: Periodic time series of 100 equally spaced points with period 10 (frequency 0.1) and correlated random observation errors (standard deviation 2/3).

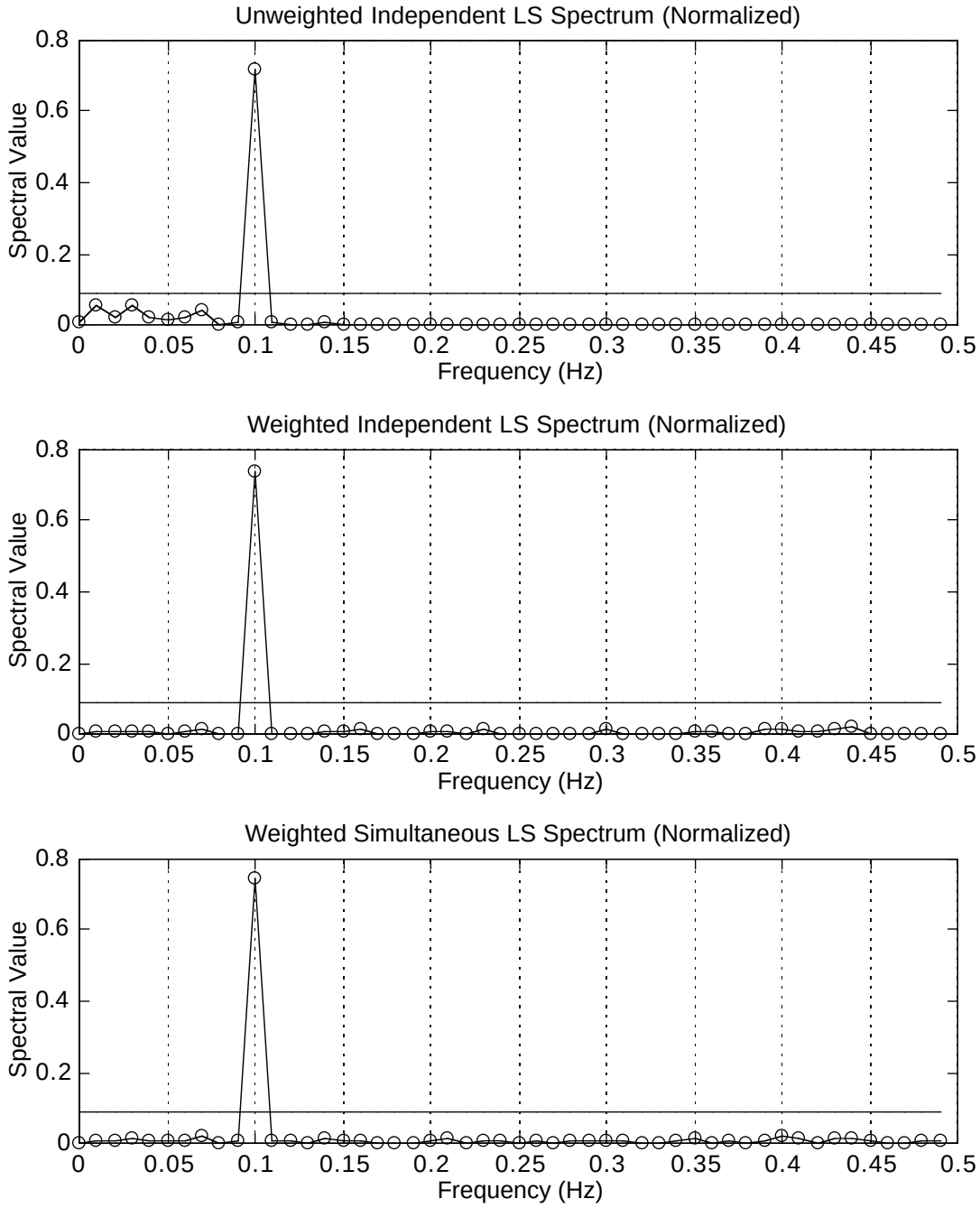

Figure 7.8: Unweighted and weighted LS spectra (both independent and simultaneous estimation) for periodic time series of 100 equally spaced points with period 10 (frequency 0.1) and correlated random observation errors (standard deviation 2/3).

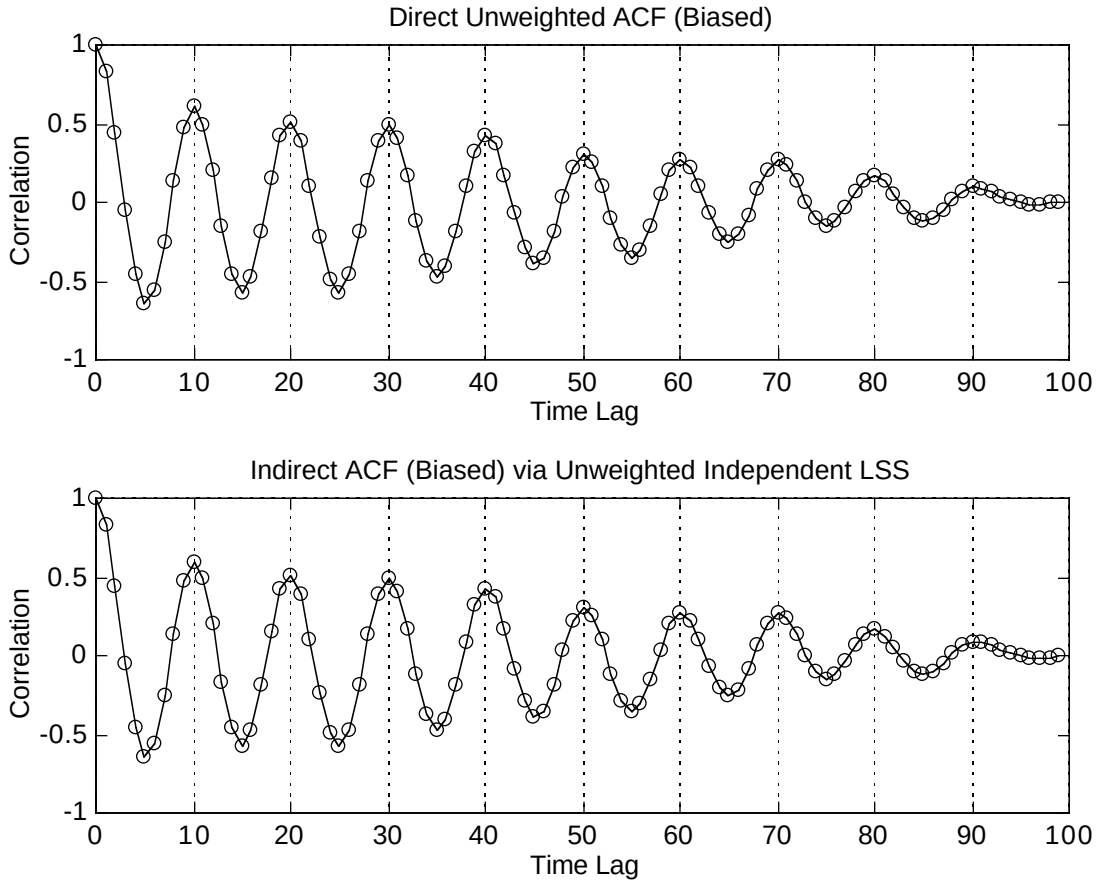

Figure 7.9: Direct and unweighted indirect (via unweighted inverse transform of unweighted LS spectrum) estimates of biased autocorrelation function for periodic time series of 100 equally spaced points with period 10 (frequency 0.1) and correlated random observation errors (standard deviation 2/3).

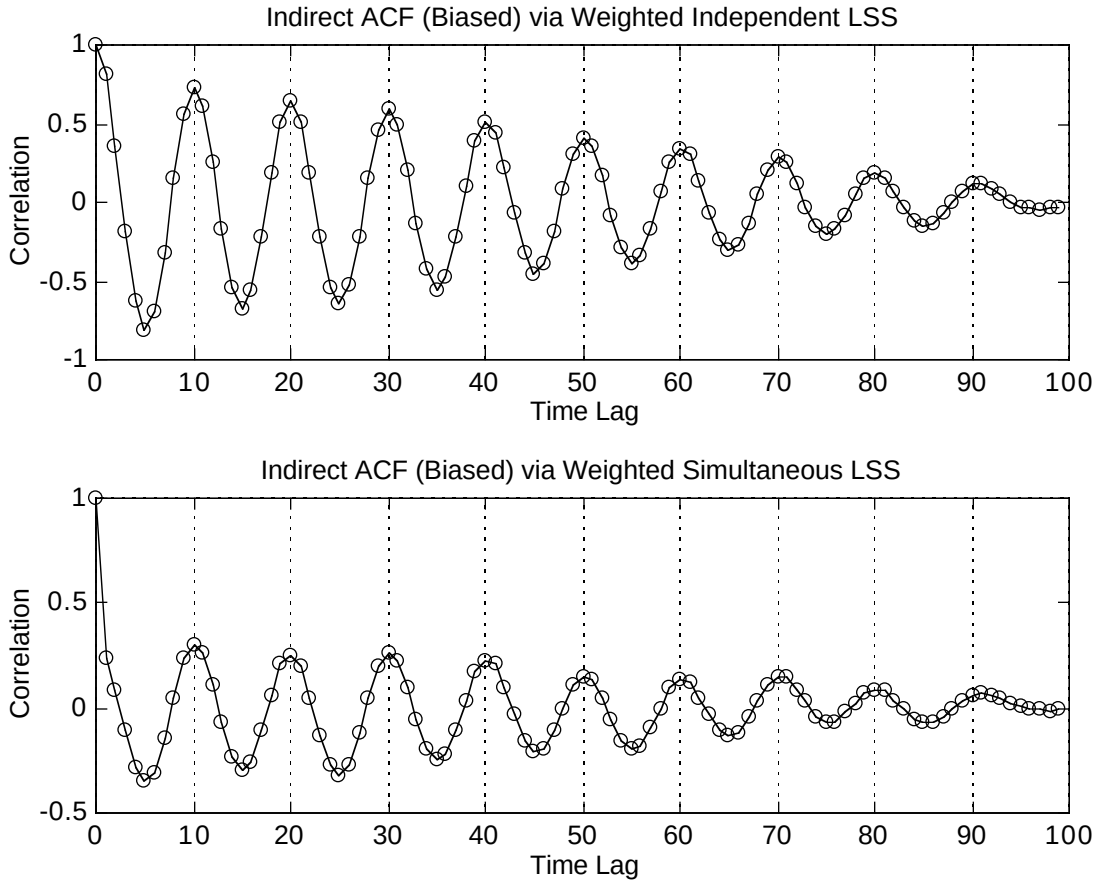

Figure 7.10: Weighted indirect estimates of biased autocorrelation function via weighted inverse LS transform of both independent and simultaneously estimated LS spectra for periodic time series of 100 equally spaced points with period 10 (frequency 0.1) and correlated random observation errors (standard deviation 2/3).

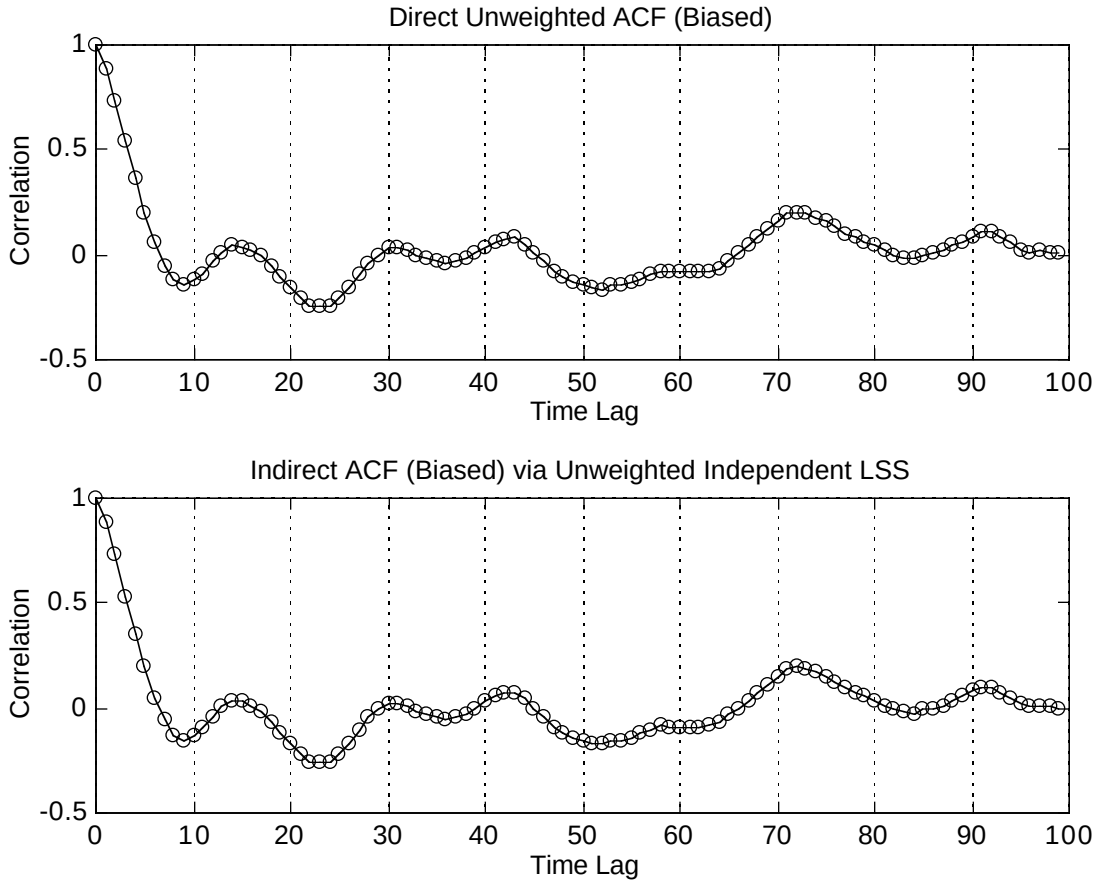

Figure 7.11: Direct and unweighted indirect (via unweighted inverse transform of unweighted LS spectrum) estimates of biased autocorrelation function for time series of 100 equally spaced points with correlated random observation errors only (standard deviation 2/3).

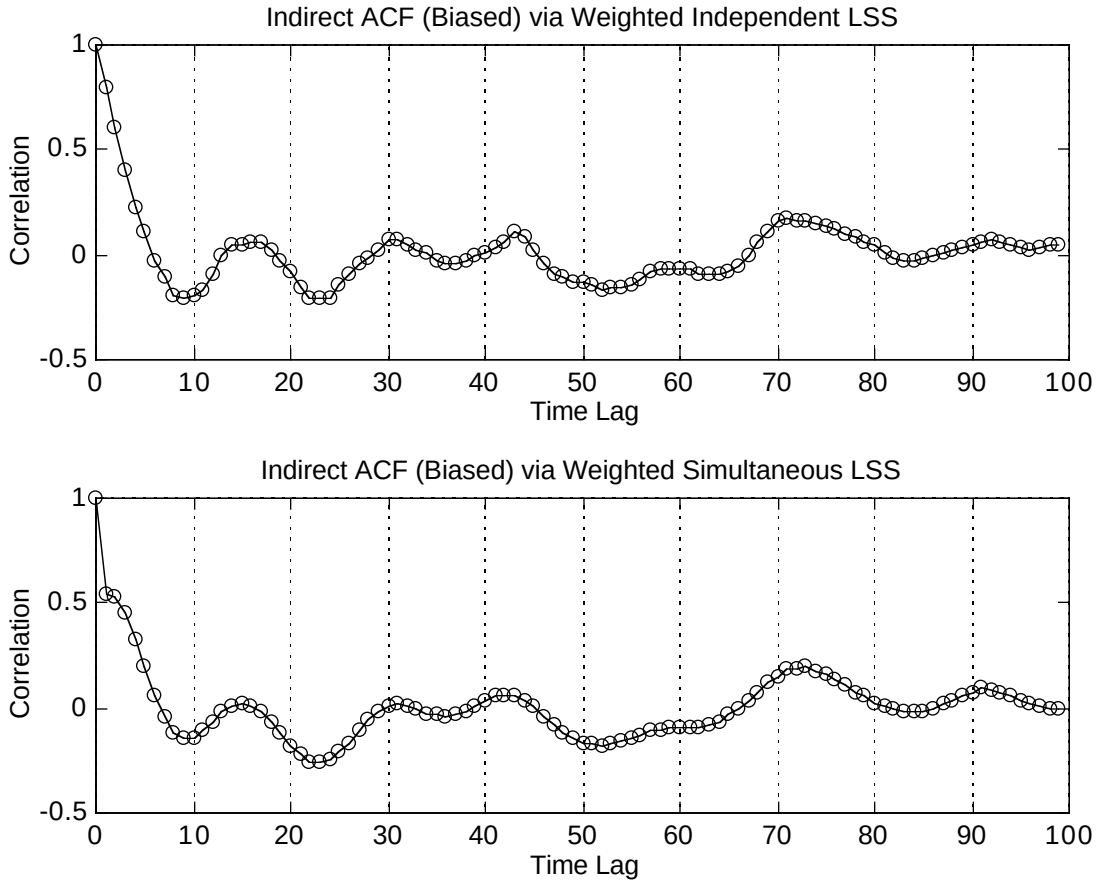

Figure 7.12: Weighted indirect estimates of biased autocorrelation function via weighted inverse LS transform of both independent and simultaneously estimated LS spectra for time series of 100 equally spaced points with correlated random observation errors only (standard deviation 2/3).

7.4 Effect of Random Sampling

Random observation sampling results in an unequally spaced data series in which case the conventional Fourier expressions are no longer valid. This is the primary reason for using the least squares transform and spectra. To test the effect of random sampling on the LS transform and spectra, unequally spaced periodic data series were constructed.

Different lengths of data series were used to examine the effect of the finiteness and sparseness of the data. The unequally spaced time arguments were created using a pseudo-random number generator with a uniform distribution (see Press et al. [1991] for an explanation of the uniform distribution). Three unequally spaced (errorless) data sets with a periodic trend of period 10 (frequency 0.1 Hz) were generated with 100, 60 and 20 points (see Figure 7.13).

The spectra were computed independently for integer multiples of the fundamental frequency (0.01 hz), up to frequency 0.5 hz. Because the Nyquist frequency is undefined for randomly data spacing, the spectra were computed only up to an arbitrarily selected frequency of 0.5 hz. The absence of a Nyquist frequency is illustrated in Figure 7.14a, which gives the spectra of the data series up to maximum frequencies of 0.5, 6 and 25 hz. There is no evidence of a mirror image in these spectra that would indicate the presence of a possible Nyquist frequency. Also, because of the large correlations between the frequency components, it is not possible to estimate the simultaneous inverse LS transform due to ill-conditioning. This will be investigated further in the next section.

The spectra for the three data series are given in Figure 7.14b. The effect of unequal sampling on the independent LS spectrum is negligible. The spectral component at frequency 0.1 is correctly located with a normalized spectral value of 1. The correct location of the spectral peak is also unaffected by the finiteness or sparseness of the data series. Even with only 20 points the LS spectrum is practically unchanged, except for greater noise in the spectrum and a larger 95% confidence level.

The indirect (biased) estimates of the autocorrelation function via the independent LS spectrum are given in Figure 7.15 for the three data series. Zero-padding was used prior to computing the spectrum to which the inverse LS transform was applied. All ACFs display the correct shape and tapering for the periodic signal in the data series. The effect of the random sampling is to reduce the magnitude of maximum correlation for non-zero lags (compare top plot in Figure 7.15 with Figure 7.4). The maximum correlation is about half of the theoretical ±1 value for all plots; i.e., the magnitude does not change as a function of the finiteness or sparseness of the data. The correct shape of the theoretical ACF is also preserved even with only 20 points.

For comparison, Figure 7.16 gives direct estimates of the autocorrelation functions computed for the same unequally spaced data series using the interval averaging method described by Vaníček and Craymer [1983a; 1983b] and Craymer [1984]. All ACFs display the same periodic component as the indirect estimates (overlay Figure 7.16 with Figure 7.15). However, the direct ACF for the 100 point series clearly does not follow the expected tapered shape (compare with Figure 7.4). Instead, the correlations at both small and large time lags are significantly attenuated, while correlations at the middle lags are equal to one. It appears more like a modulated unbiased ACF. The other ACFs agree well with both the indirect estimates; they are closer in magnitude to the theoretical ACF (compare with Figure 7.4).

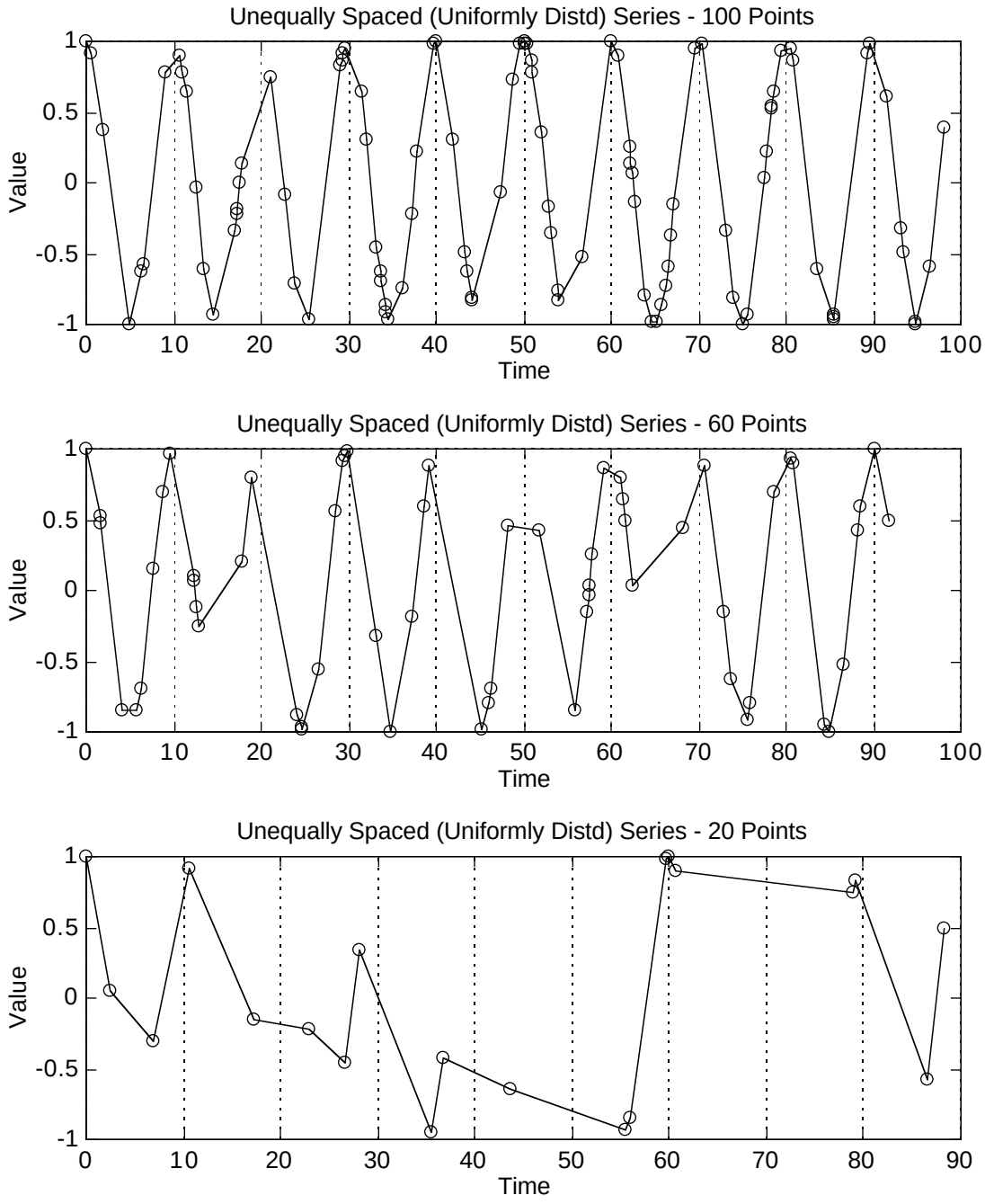

Figure 7.13: Periodic time series of different lengths of randomly spaced points (uniformly distributed) with period 10 (frequency 0.1) and no random observation errors.

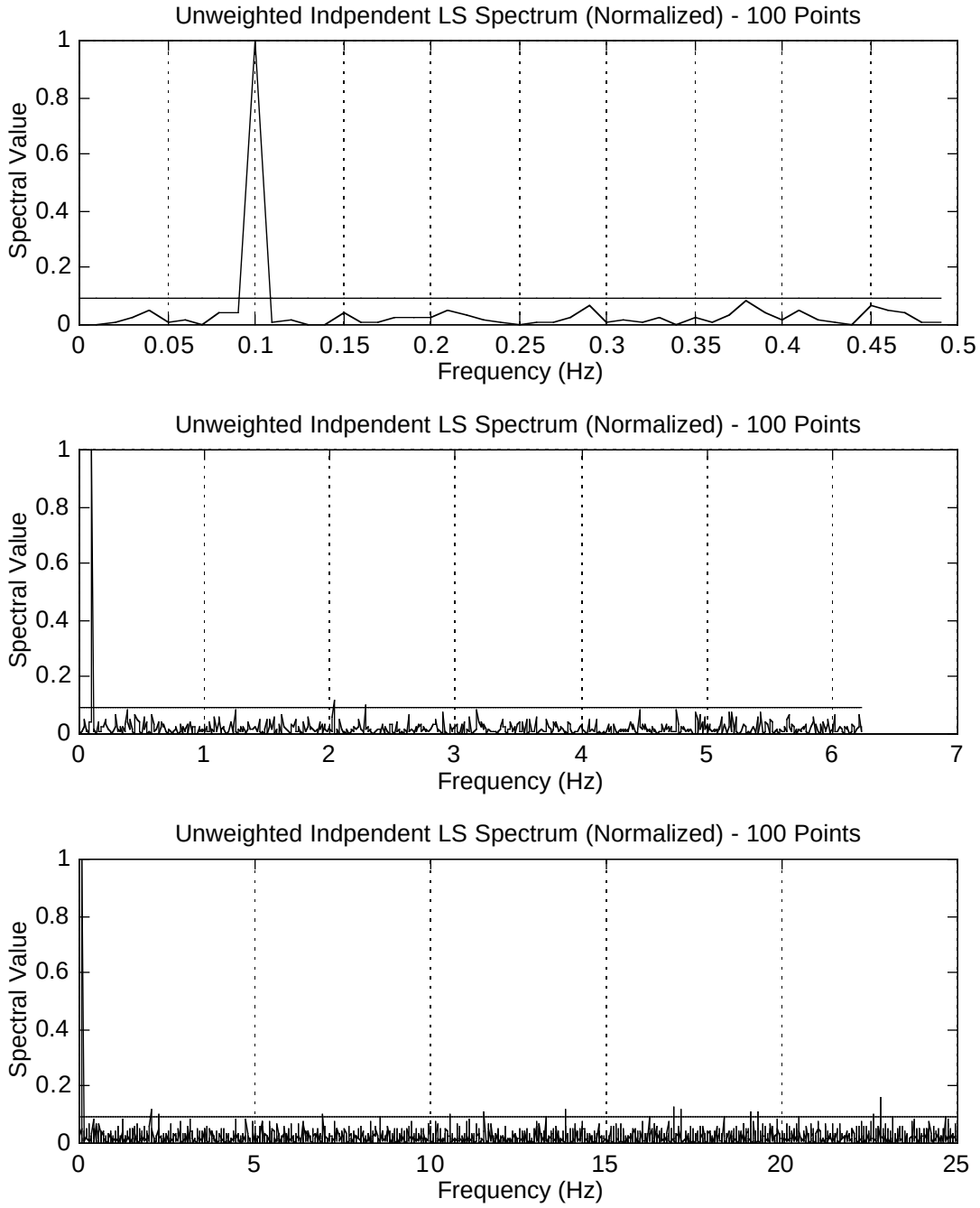

Figure 7.14a: LS spectra (independently estimated frequency components) up to different maximum frequencies for periodic data series of unequally spaced points with period 10 (frequency 0.1) and no random observation errors.

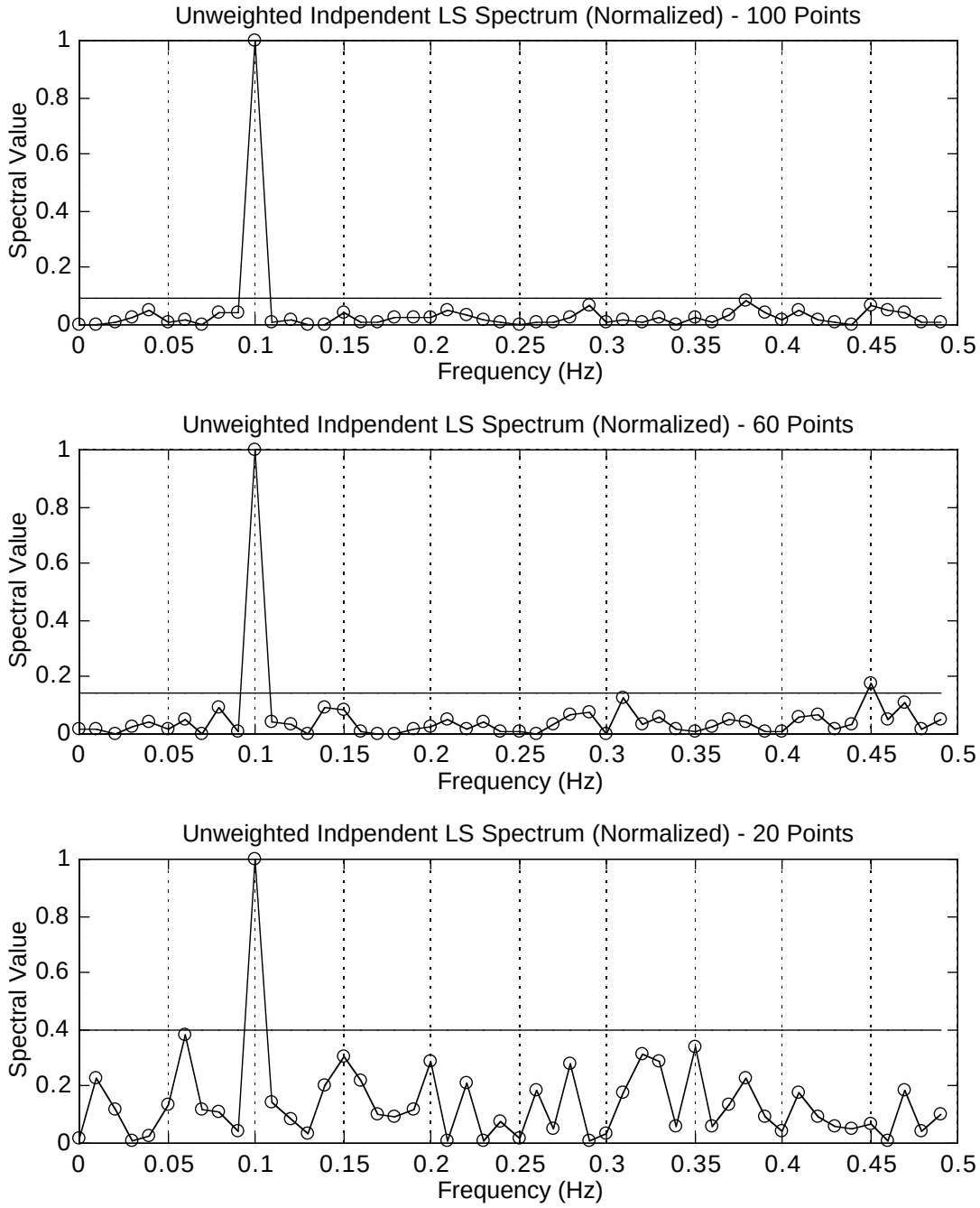

Figure 7.14b: LS spectra (independently estimated frequency components) for different lengths of periodic data series of unequally spaced points with period 10 (frequency 0.1) and no random observation errors.

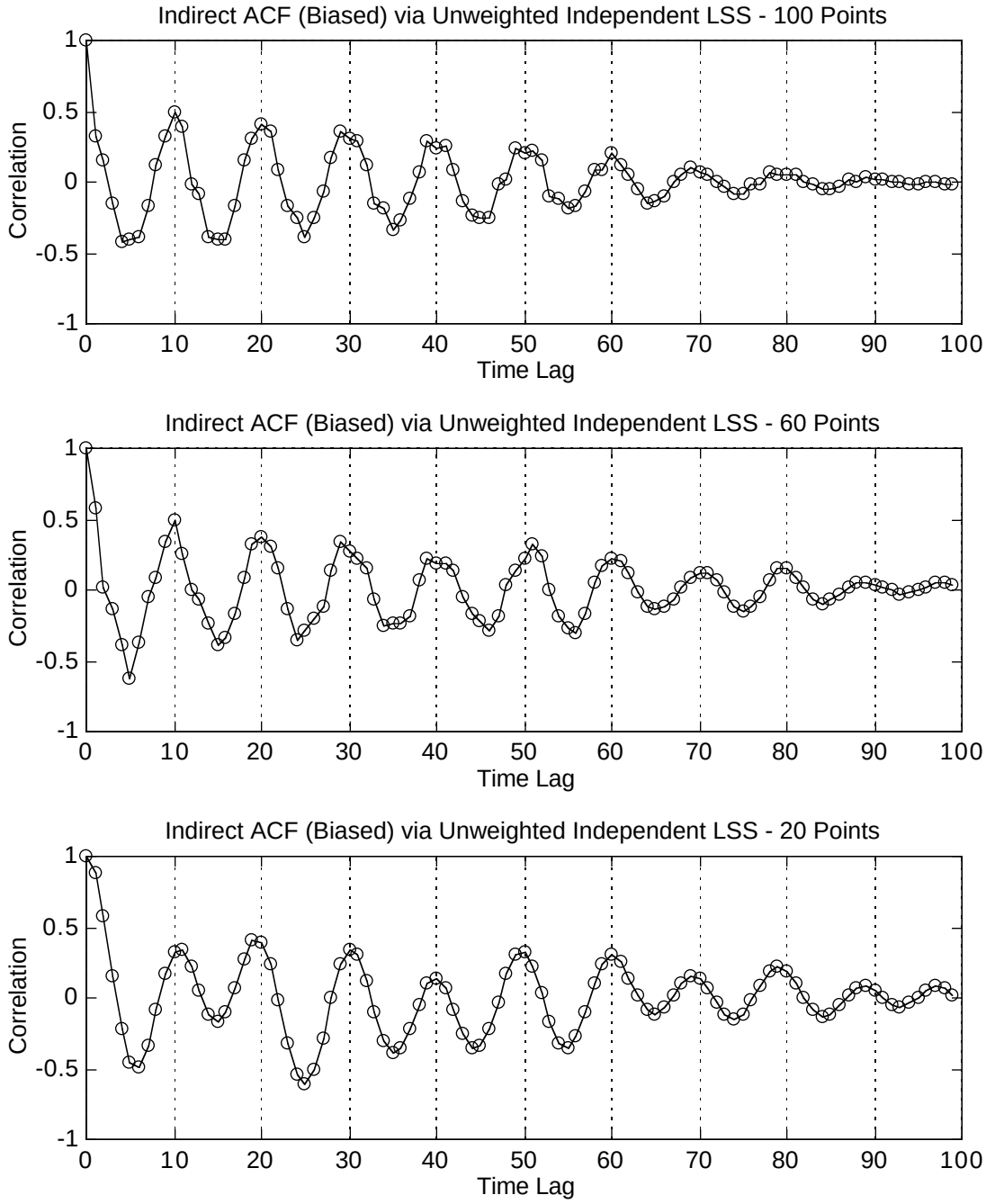

Figure 7.15: Indirect estimates (via unweighted inverse LS transform of unweighted LS spectrum) of biased autocorrelation functions for different lengths of periodic data series of unequally spaced points with period 10 (frequency 0.1) and no random observation errors.

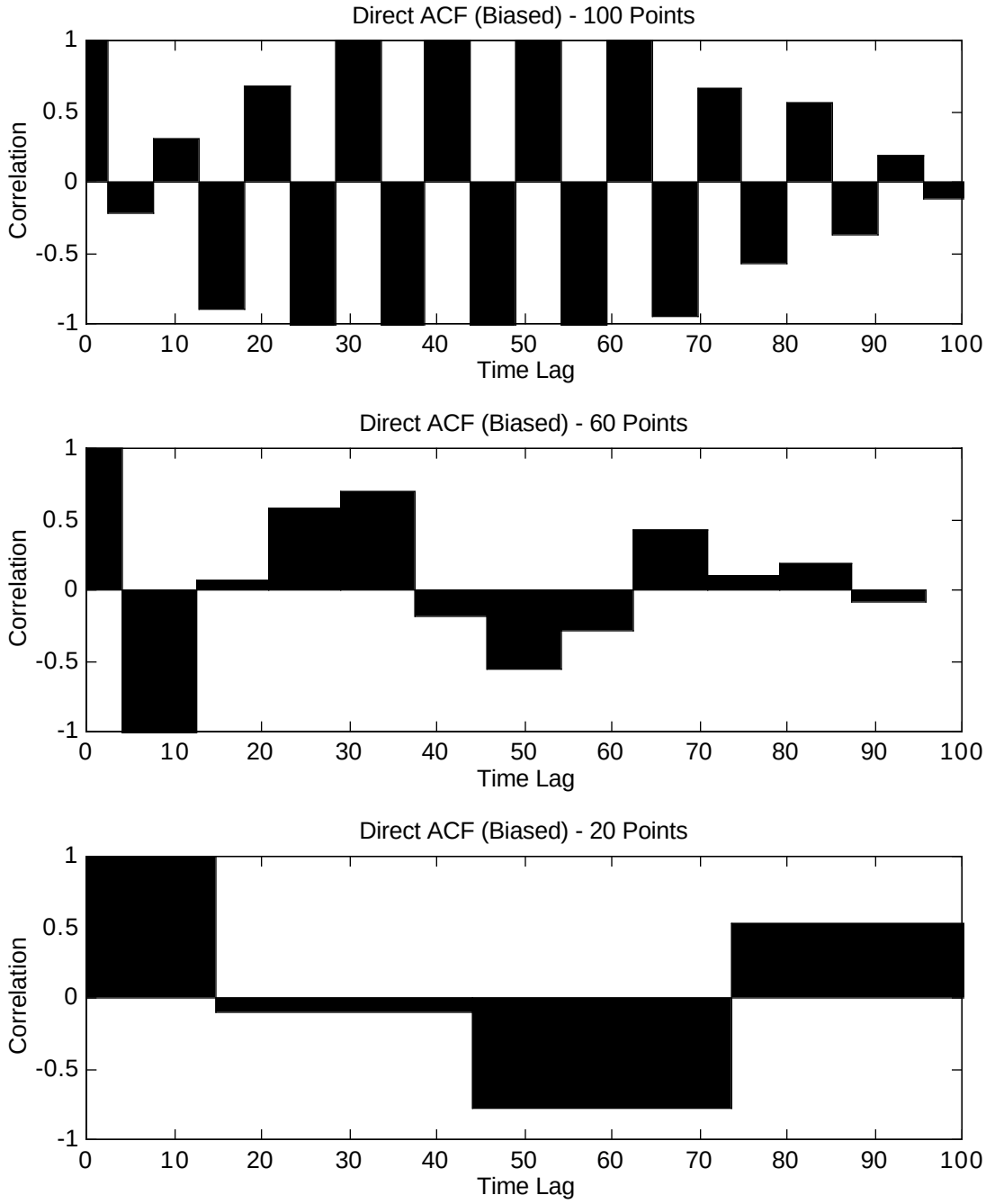

Figure 7.16: Direct estimates (via interval averaging) of biased autocorrelation functions for different lengths of periodic data series of unequally spaced points with period 10 (frequency 0.1) and no random observation errors.

7.5 Effect of Frequency Selection

The effect of different selections of frequencies for the simultaneous LS spectrum was also examined. Note that frequency selection only affects the simultaneous estimation of the spectral components. It has no effect on the independently estimated LS spectrum where each spectral component is treated out-of-context of the others (no correlations arise) and any set of frequencies may be used to correctly locate the significant spectral peaks in a data series, within the limitations of the sampling theorem (see Section 5.3). This effectively provides a continuous spectrum, although spectral leakage may affect the result. The significant spectral components can then be used in the indirect estimation of the ACF via the simultaneously estimated LS transform or in an improved deterministic model.

On the other hand, the selection of frequencies is of critical importance for the simultaneously estimated LS spectrum. In this case the correlations among the spectral components must be carefully considered, otherwise ill-conditioning in the normal equations for the simultaneous solution of all spectral components can produce completely wrong results. For example, consider the same data series used in the previous section (top plot in Figure 7.13), containing 100 unequally spaced (uniformly distributed) points with a periodic trend of period 10 (frequency 0.1 Hz) and no random errors. Using the entire set of 50 “Fourier” frequencies in the simultaneous LS spectrum, results in an ill-conditioned solution. The resulting spectrum fails to detect the periodic trend at frequency 0.1 hz even with no random errors present (see top plot in Figure 7.17).

The correlations among the frequencies can be reduced and the ill-conditioning in the spectral transform removed by decreasing the frequency sampling to only every other frequency; i.e., 25 of the original set of 50 frequencies. Although the periodic component is now visible in the simultaneous LS spectrum, it is still relatively small and only just statistically significant (see middle plot in Figure 7.17). This is improved further by taking

every 5th frequency so that only 10 of the original 50 frequencies are used. The spectral peak at 0.1 is now highly significant.

The same behaviour is also displayed by the indirect estimate of the autocorrelation function. Note, however, that the original data series needs to be zero-padded to avoid "wrap around" effects in the ACF. This doubling of the series length results in a fundamental frequency that is half of that for the original series and twice as many frequencies. This results in even more severe ill-conditioning and a completely erroneous ACF where correlations are much greater than 1 (see top plot in Figure 7.18). Decreasing the frequency sampling to only 50 frequencies improves the ACF but there are still some correlations greater than 1 (see middle plot of Figure 7.18). The situation is improved when only 10 frequencies are used. The ACF has the correct cosine form and the maximum correlations are only slightly larger than 1 (they could be truncated to 1 in practice).

The problem with decreasing the frequency sampling is that some peaks may be missed. Clearly, great care must be exercised when selecting the frequencies to use with the simultaneous estimation of the LS spectrum and the inverse LS transform. Note that by reducing the number of simultaneously estimated frequencies, one is approaching the method of independent estimation of the spectral components (the extreme or limiting case of reducing the number of frequencies).

A better approach may be to instead search for and use only statistically significant spectral components from the independent estimation of the LS spectrum. These frequencies can then be used in a simultaneous estimation of the LS spectrum and in the simultaneous inverse LS transform for the indirect ACF. The results following this procedure are illustrated in Figures 7.19 and 7.20 for a randomly sampled data series with two periodic components (frequencies 0.1 and 0.25 hz) and no random errors. The independent estimation of the LS spectrum correctly identifies the two periodic components as shown in Figure 7.19. Using only these significant periodic components in the

simultaneous estimation of the spectrum and the subsequent simultaneous inverse transform gives an indirect ACF that agrees with the theoretical form of the unbiased, rather than the biased, ACF, rather than the biased as shown in Figure 7.20. On the other hand, the ACF derived from the inverse transform of the entire independently estimated LS spectrum provides the expected biased form ACF. It appears that reducing the number of frequencies in the inverse transform gives an ACF that more closely agrees with the unbiased estimate. The biased ACF can be obtained by simply using n in place of the divisor $(n–k)$ in the expression for the unbiased ACF in eqn. (2.20).

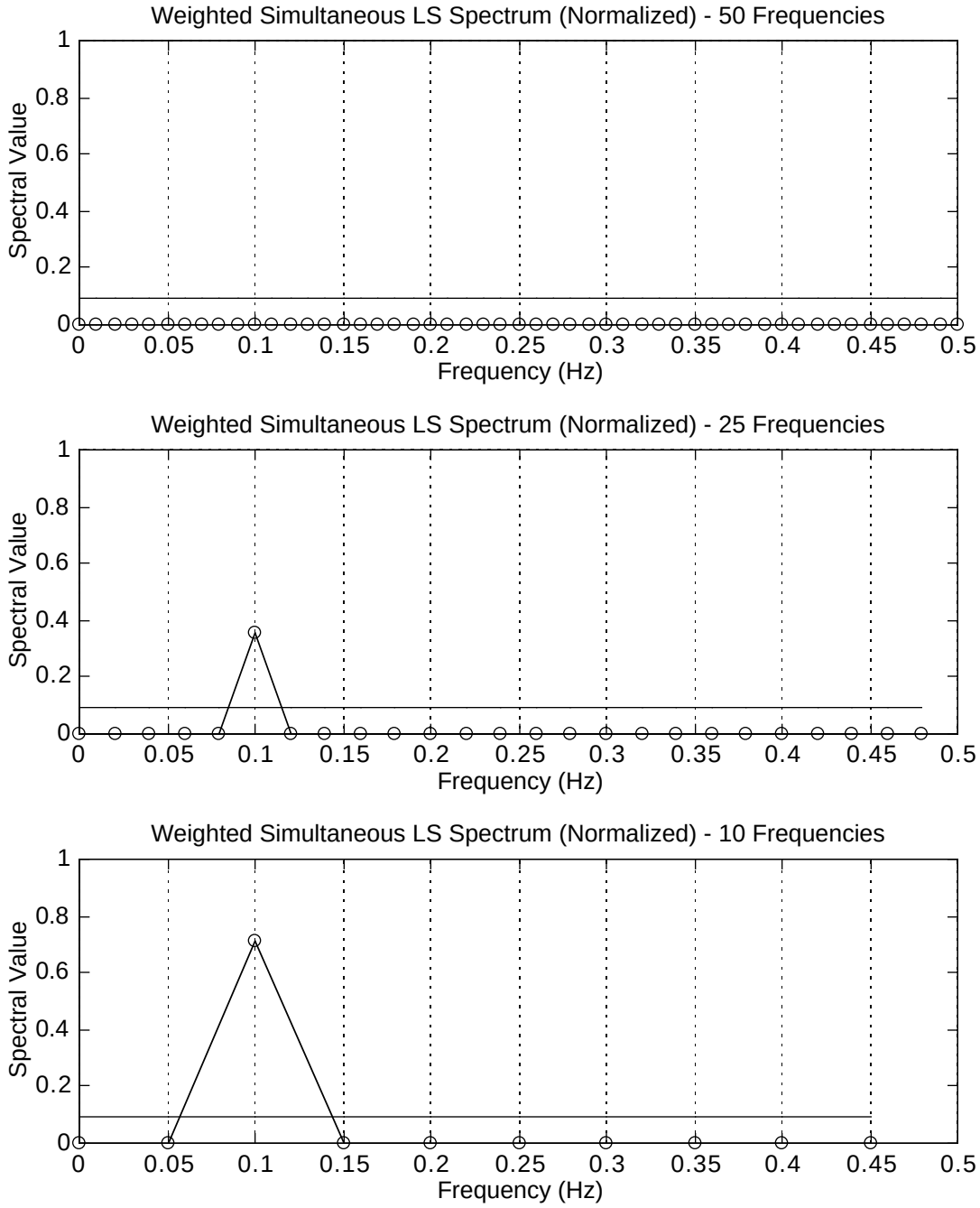

Figure 7.17: LS spectra for different sets of simultaneously estimated frequencies for periodic data series of 100 unequally spaced points with period 10 (frequency 0.1) and no random observation errors.

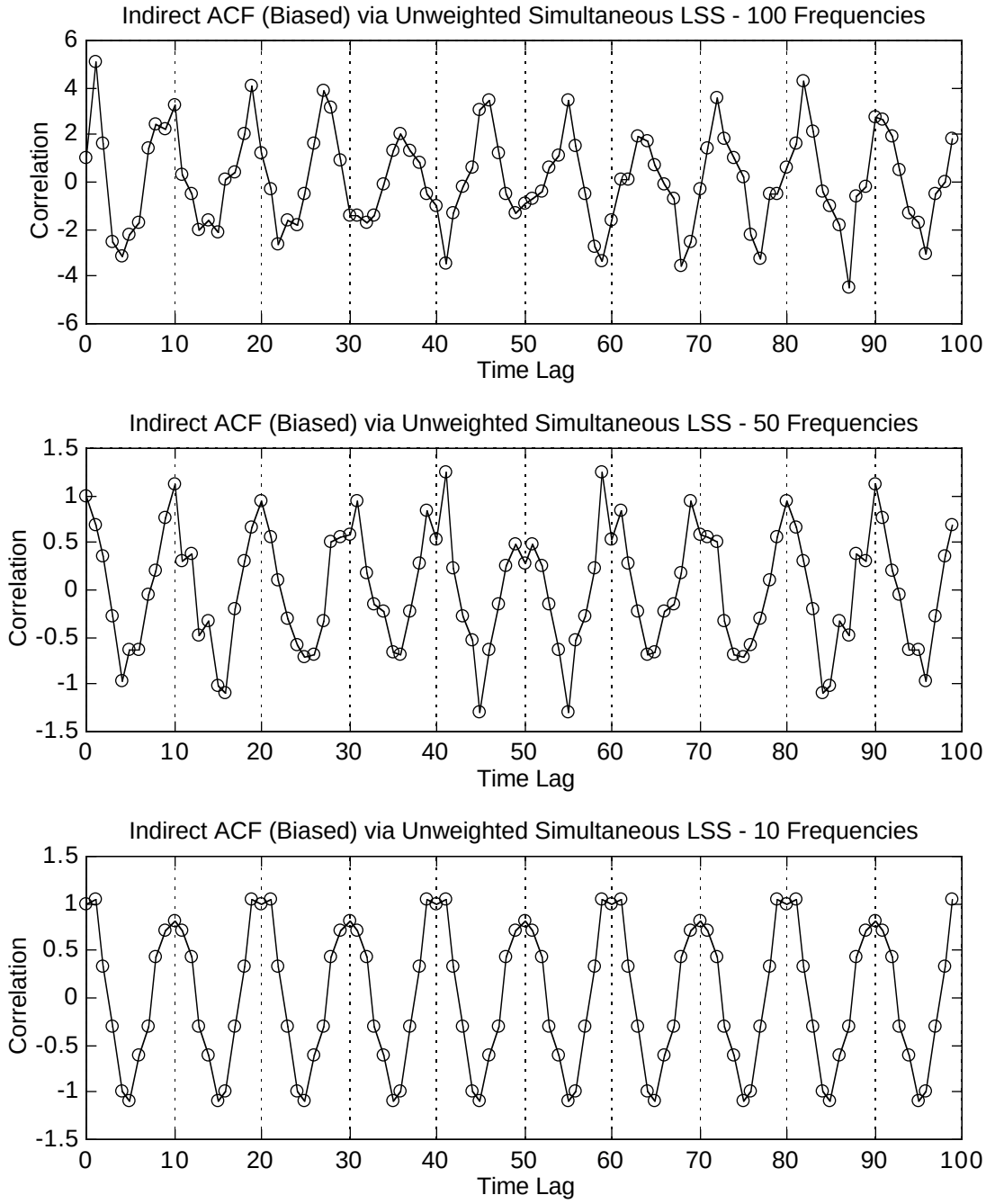

Figure 7.18: Indirectly estimated LS autocorrelation functions via the LS spectrum using different sets of simultaneously estimated frequencies for periodic data series of 100 unequally spaced points with period 10 (frequency 0.1) and no random observation errors.

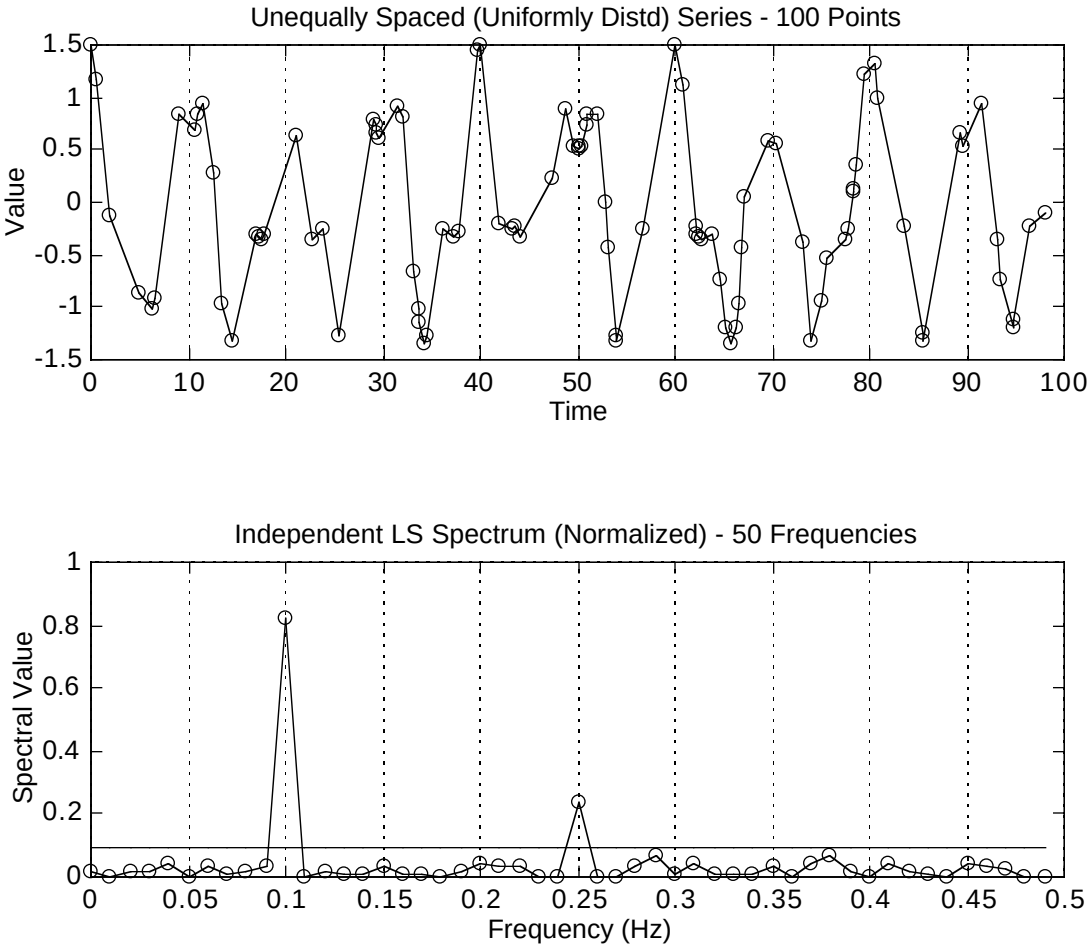

Figure 7.19: Periodic time series of randomly spaced points with frequencies 0.1 and 0.25 hz and no random observation errors (top), and independent estimation of the LS spectrum (bottom).

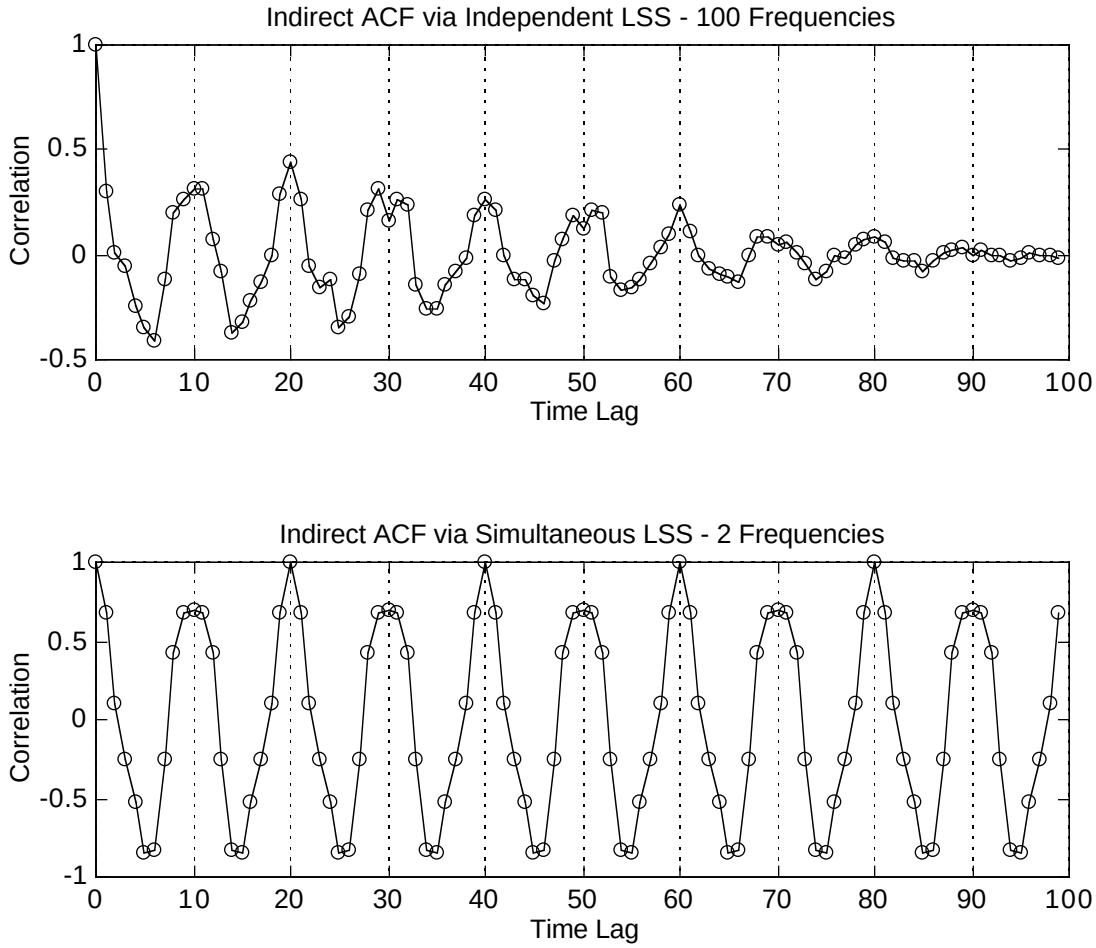

Figure 7.20: Indirectly estimated ACF via the inverse LS transform of the independent LS spectrum using all frequencies (top) and of the simultaneous LS spectrum using only the two significant spectral peaks at 0.1 and 0.25 hz (bottom).

7.6 Effect of Deterministic Model

The effect of the deterministic model on the LS spectrum and indirectly estimated autocorrelation function is to absorb any spectral components that are highly correlated with the deterministic model. These spectral components are usually at the lower frequencies, unless some high frequency periodic trends are included in the deterministic model. The deterministic model is accommodated by accounting for its effect within the estimation of the LS spectrum and inverse LS transform following the approach described in Chapters 4 and 5.

To test the effect of a deterministic linear trend model, a 100 point equally spaced data series consisting of a quadratic trend ($1 + 0.02\ t + 0.00005\ t^2$) and a periodic residual trend of frequency 0.01 hz was generated with no random errors (see top plot in Figure 7.21). The quadratic trend will tend to alias as a long period trend which may result in erroneous estimates of the spectrum of the residuals if the correlations with the quadratic model are not accounted for. This is evident in the middle plot of Figure 7.21, where the LS spectrum displays a peak at 0.02 hz while the actual periodic signal should be at 0.01 hz. There is also some spectral leakage into the neighbouring frequencies at 0.01 and 0.03 hz. Accounting for the correlations with the deterministic model results in a spectrum that correctly identifies the 0.01 hz peak and eliminates the spectral leakage (bottom plot in Figure 7.21).

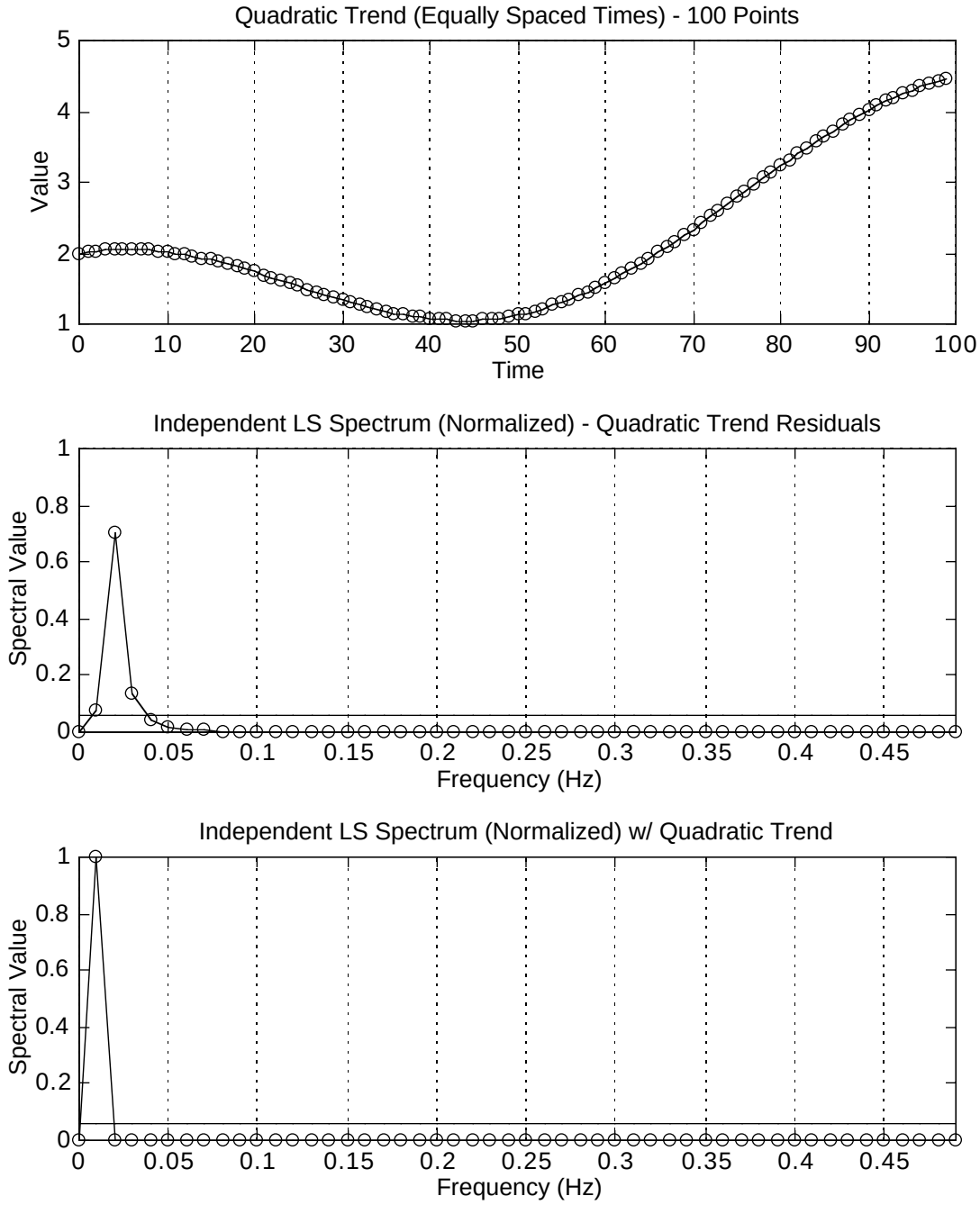

Figure 7.21: Quadratic trend time series with periodic component (frequency 0.01 hz) and no random errors (top); LS spectrum of residuals from quadratic trend model (middle); LS spectrum accounting for effects of quadratic model (bottom).

7.7 Effect of Non-Stationary Random Errors (Random Walk)

Another kind of correlated error are non-stationary random errors. One example of this is the simple random walk model where the error ε_i at time t_i is the accumulation of a white noise process [Papoulis, 1965]; i.e.,

$$\varepsilon_i = \sum_{j=1}^{i} \eta_j \,, \tag{7.4}$$

where the η_i are normally distributed random variables with zero mean. One such equally spaced random walk data series with a unit standard deviation is displayed in Figure 7.22 (top plot). This 100 point data series is actually a evenly sampled subset (every fifth point) of a much larger 500 point random walk data series using a white noise process with unit standard deviation. The theoretical spectrum for such a process is inversely proportional to the square of the frequency [Zhang et al., 1997]. The computed LS spectrum is given in the middle and bottom plots of Figure 7.22. The bottom plot uses a log scale for both axes and exhibits a linear trend with a slope of about –2 corresponding to the expected f^{-2} relation for a random walk model. The direct and indirect autocorrelation functions are given in Figure 7.23. The indirect estimate via the LS spectrum (zero-padding is used) agrees well with the direct estimate. The differences between them shown in the bottom plot of Figure 7.23 increase in direct proportion to the lag. The indirect ACF departs from the direct ACF to about 0.5 at the highest lag.

To test the effect of the data sampling, an unevenly spaced random walk data series was generated by randomly sampling the same 500 point random walk series used above (see Figure 7.24). (A uniform random number generator was again used to generate the random selection of 100 points; see Section 7.4.) The LS spectrum is given in the bottom two plots. The effect of the random sampling is to flatten out the spectrum at the higher

frequencies. The inverse square frequency relation only holds at the lower frequencies. This behaviour was also found by Zhang et al. [1997]. The indirect estimate of the autocorrelation function via the independent LS spectrum (with zero-padding) is also significantly affected by the random sampling (see Figure 7.25). It now drops off much more rapidly in comparison to the direct estimate in Figure 7.23).

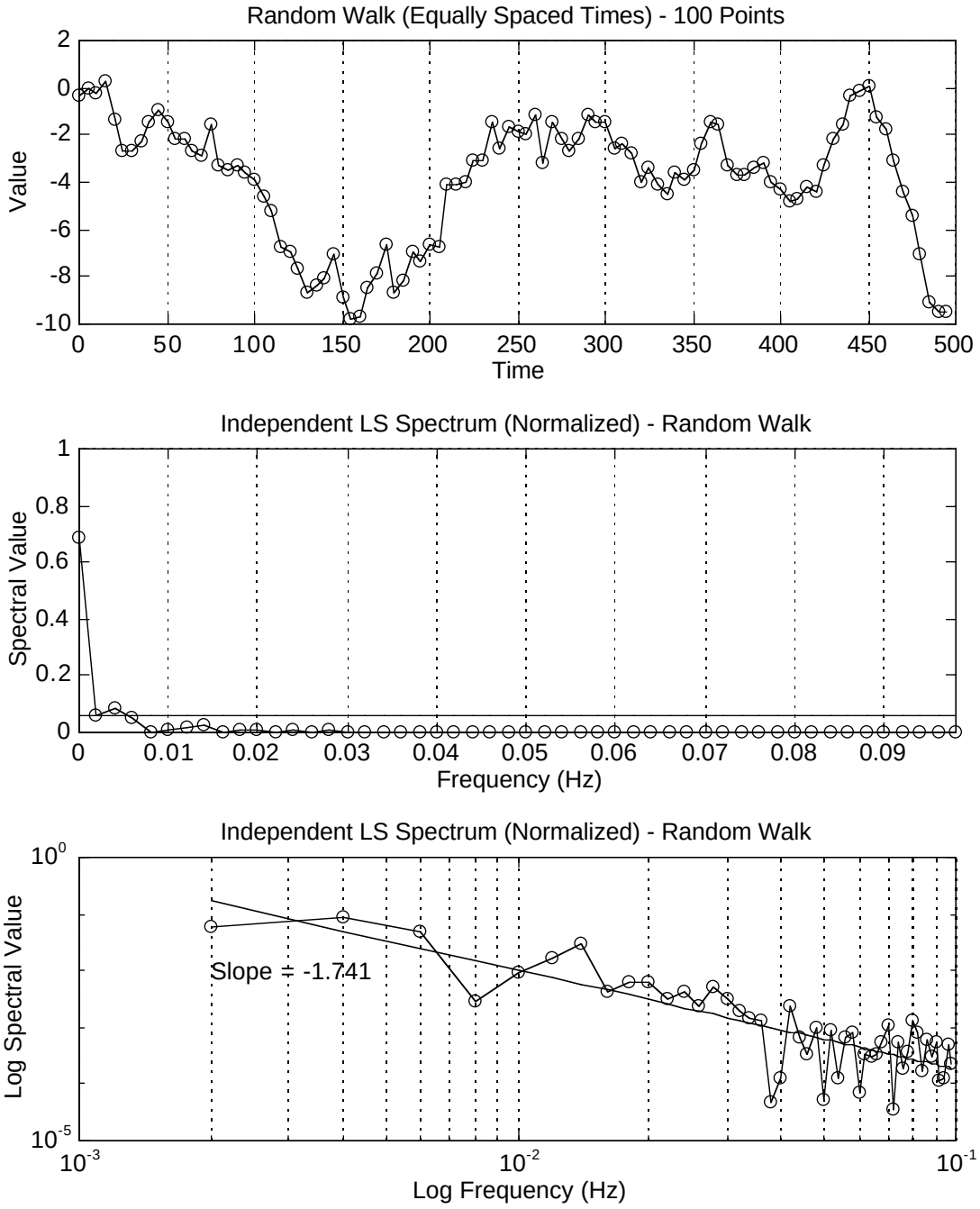

Figure 7.22: Evenly sampled 100 point random walk time series (standard deviation 1) (top) and its corresponding LS spectrum.

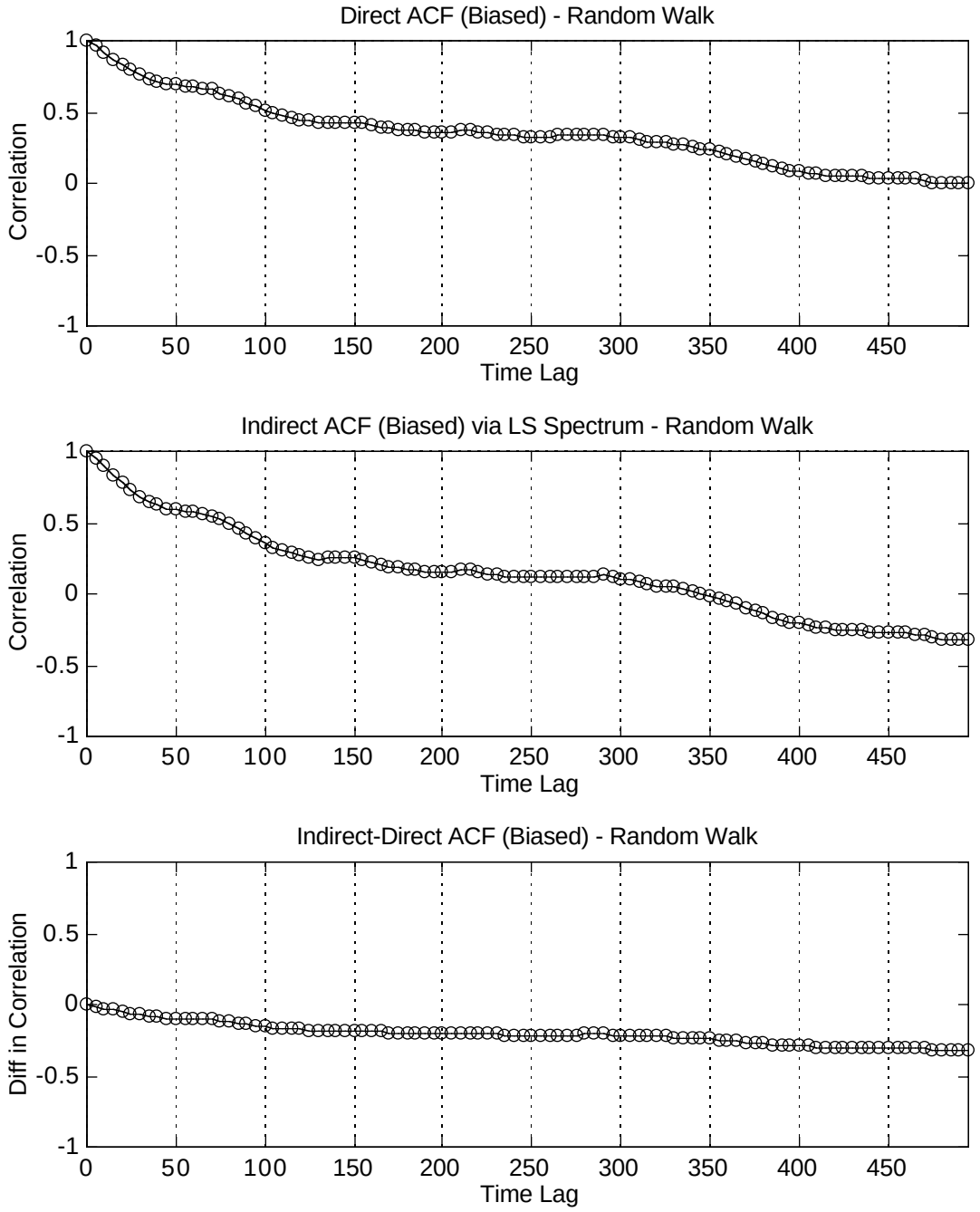

Figure 7.23: Direct (top) and indirect (bottom) autocorrelation functions for 100 point random walk data series.

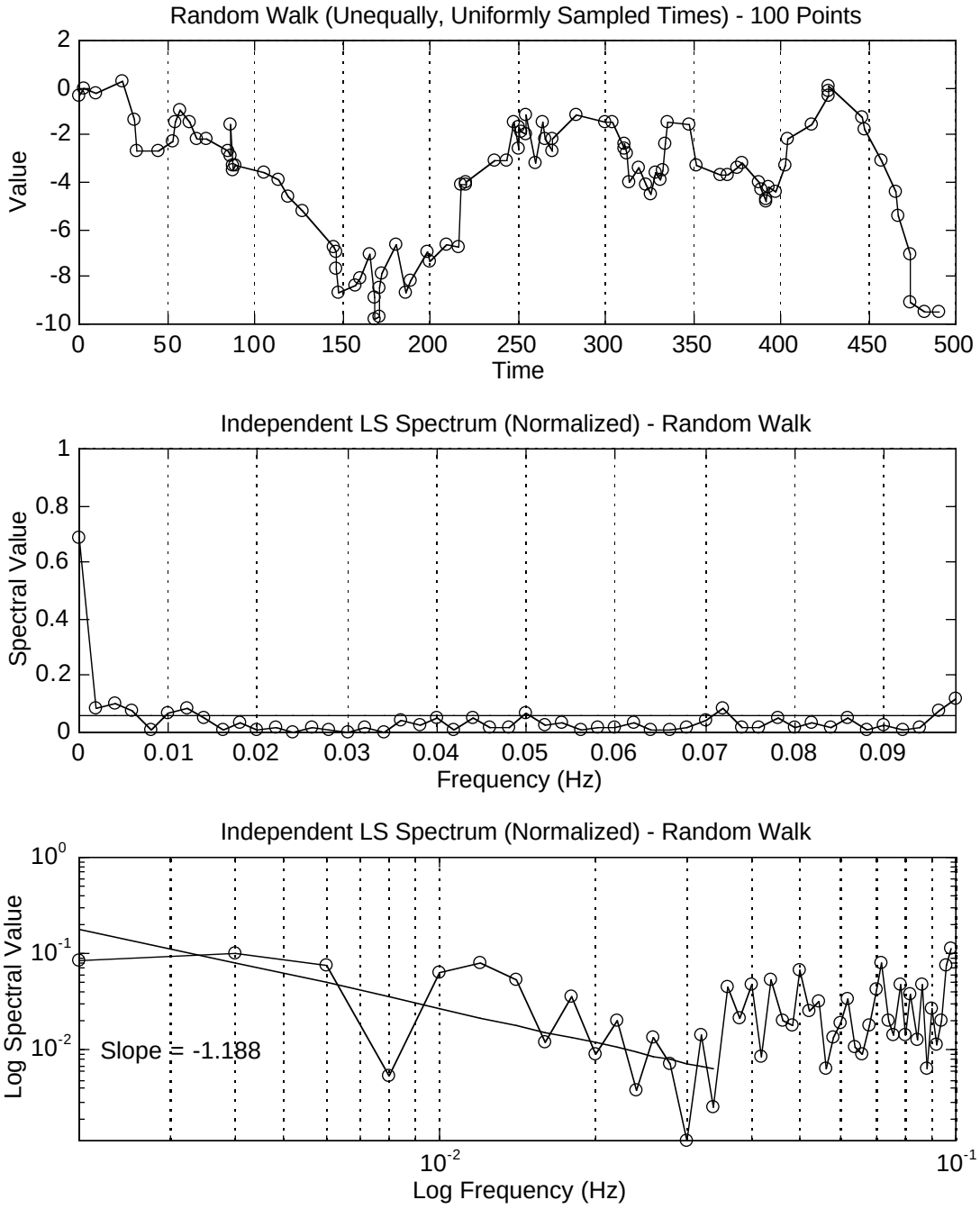

Figure 7.24: Unevenly sampled 100 point random walk time series (top) and its corresponding LS spectrum.

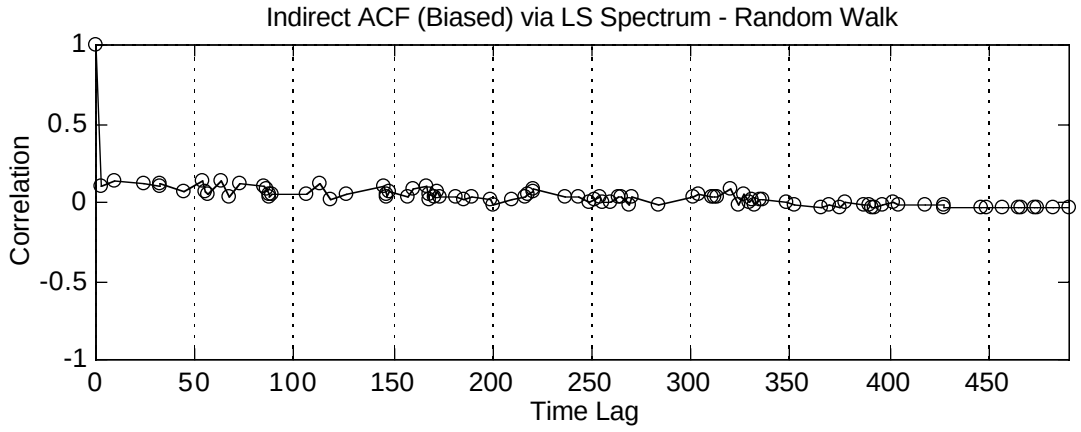

Figure 7.25: Indirect estimate of autocorrelation via the independently estimated LS spectrum for the unevenly sampled 100 point random walk time series.

Chapter 8
Some Applications in Geodesy

8.1 Introduction

There have been many applications of time series analysis in geodesy to the study of tide gauge data, gravity data and geodynamics. In particular, the method of least squares spectral analysis has been applied to studies of the Earth-pole wobble by Vaníček [1969b] and Rochester et al. [1974]. However, there have been few applications of time series analysis techniques to other kinds of geodetic data. The few studies employing these techniques have been mostly applied to levelling data (see, e.g., Vaníček and Craymer [1983a, 1983b], Craymer [1984], Vaníček et al. [1985], Craymer [1985], Craymer and Vaníček [1985, 1986, 1990]). More recently time series analysis techniques have also been applied to electronic distance measurement (EDM) data by Langbein et al. [1990] and Langbein and Johnson [1997], and to Global Positioning System (GPS) data by El-Rabbany [1994], King et al. [1995] and Zhang et al. [1997]. In El-Rabbany [1994], only standard Fourier (and FFT) methods in the equally spaced time dimension are considered. The study by King et al. [1995] also assumed equally spaced time arguments. Only the recent work of Langbein and Johnson [1997] and Zhang et al. [1997] have considered unequally spaced data. In particular, Zhang et al. [1997] have used the periodogram as defined by Scargle [1982], which can be shown to be a special case of Vaníček's original method (see Section 5.7). Estimation of covariance and correlation functions for stochastic modelling of errors, however, was still based on traditional methods assuming equally spaced data.

The studies by Craymer et al. have applied time series techniques more generally to arguments that are not necessarily equally spaced in order to search for systematic errors that depend on these quantities. All these studies have used the unweighted form of the independently estimated least squares spectrum to search for systematic errors in precise levelling. Here, the weighted form of the least squares approach to spectrum and autocovariance function estimation are applied to the stochastic modelling of errors using two real examples: estimation of the deformation of an EDM baseline across the San Andreas fault using the same data as in Langbein and Johnson [1997], and GPS single point positioning using pseudo-range observations (the typical positioning data used by most handheld GPS receivers).

8.2 EDM Deformation Measurements

Electronic distance measurements (EDM) is the most precise distance measuring technique at close to moderate ranges (about 1 km). The most accurate EDM instruments, such as the Kern ME5000, can routinely obtain submillimeter repeatability. The most accurate EDM instrument is based on dual frequency ("two-colour") lasers (see Slater and Huggett [1976]). The two measuring frequencies allow one to more directly determine and correct for the refraction effect (which is a function of the frequency of the laser). For this reason, two-colour EDM instruments are often used in southern California by Earth Scientists to monitor the crustal deformation around the San Andreas fault (see, e.g., Savage and Lisowski [1995]).

Here the least squares spectral analysis technique is applied to the same data used by Langbein and Johnson [1997] to search for possible systematic signals in their two-colour EDM data. Traditional spectral techniques were used by Langbein and Johnson for this purpose. Because the observations are at irregular time intervals, some necessary approximations, specifically interpolation, had to be made to estimate their spectra. No

such approximations are needed for the least squares technique which is an ideal application of this method.

The data used in this analysis are part of the Pearblossom network, near Palmdale in southern California and were provided by J. Langbein (personal communication, 21 February 1997) of the U.S. Geological Survey, Menlo Park, CA. The network is radial in design, where all distances (baselines) are measured from Holcomb to twelve surrounding monuments at distances from 3 to 8 km (see Figure 8.1). Only the Holcomb-Lepage baseline with a nominal distance of 6130 m was used in this analysis. Initially the baseline measurements at Pearblossom were made several times per week for 4 years (1980-1984). Since about 1987 they have been reduced to about once every 3 or 4 months, although each baseline is measured twice during each network re-observation. In addition, different instruments and monuments have been used over the years and there have been a number of earthquakes. Consequently, the data have been reduced to changes in baseline length from the nominal value and grouped into sets sharing common EDM instrumentation and monuments between earthquakes. The time series of the Lepage baseline measurements is given in Figure 8.2. Note the different offsets between each data group and the consistent linear trend (expansion of the baseline) for all groups. The different datum offsets represent biases in the measured differences due to the different instrument/monument combinations or the occurrence of earthquakes. It was also noted that several observations were repeated within a couple of hours of each other (two within 15 minutes!). To avoid excessively large temporal correlations under these circumstances, only the second (repeat) observations were used.

The different biases between measurement groups necessitate accounting for a separate datum offset for each. Likewise, the consistent trend for all groups necessitates modelling a common linear trend for all groups. Least squares estimates of these model parameters are given in Table 8.1, where the datum offsets are all referenced to the first measurement epoch. The 1.72 ± 0.07 mm/year linear trend (extension of the baseline)

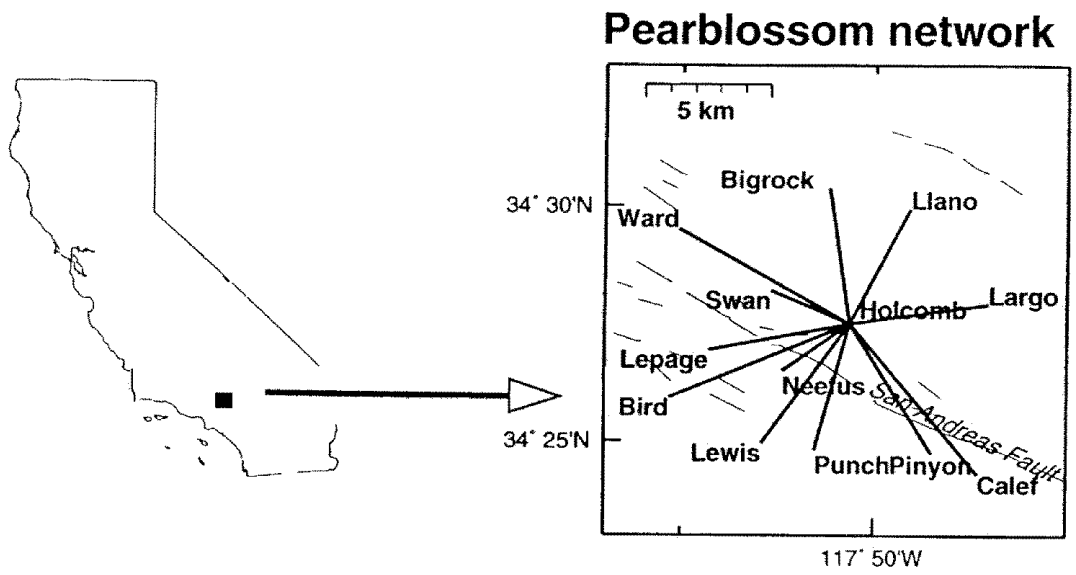

Figure 8.1: Location of the Pearblossom network in California used to measure crustal deformation with a two-colour EDM instrument and location of the Holcomb-Lepage baseline spanning the San Andreas fault running through this network [after Langbein and Johnson, 1997, Figure 1].

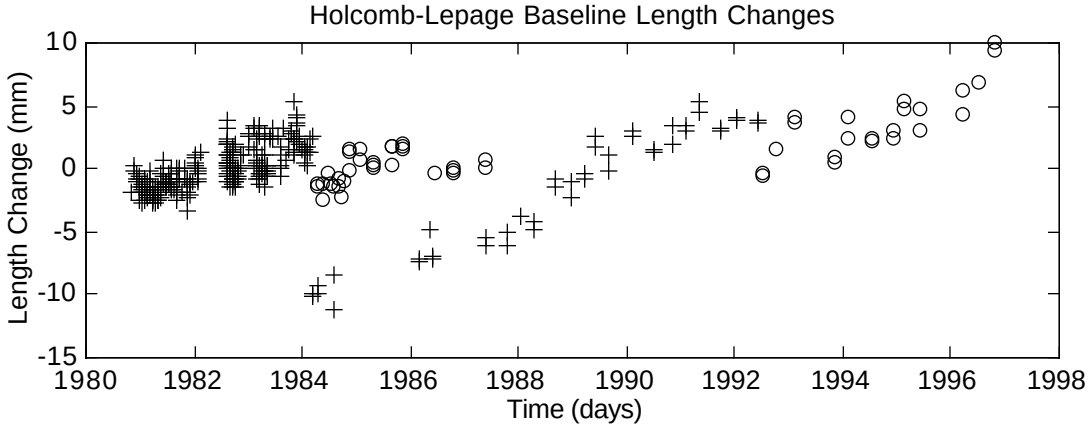

Figure 8.2: Changes in length of Holcomb-Lepage baseline. Different observation groups are denoted by different symbol colour/type combinations.

Table 8.1: Least squares estimates of linear trend and datum offsets.

	Estimate	Std	t Statistic
Offset #1 (mm)	-2.3	0.1	22.4
Offset #2 (mm)	-3.2	0.2	17.0
Offset #3 (mm)	-4.2	0.2	17.5
Offset #4 (mm)	-4.8	0.5	9.5
Offset #5 (mm)	-15.4	0.4	35.0
Offset #6 (mm)	-20.1	0.7	27.6
Offset #7 (mm)	-7.1	0.4	20.3
Offset #8 (mm)	-10.5	0.5	19.5
Linear Trend (mm/yr)	1.72	0.05	34.4

agrees well with the 1.67 value determined by Langbein and Johnson [1997]. In the least squares solution, the data were weighted using standard deviations provided by J. Langbein (personal communication, 21 February 1997). All estimated model parameters were statistically significant at any reasonable significance level and were removed from the data leaving the residual series in Figure 8.3. It is this data series that is used in the following spectral analysis.

Before performing a spectral analysis, appropriate frequencies (i.e., frequency spacing and range) must be chosen. The total length of the data series defines the smallest frequency spacing that can be resolved without spectral "leakage" from adjacent peaks. The frequency interval (Δf) is defined by

$$\Delta f = f_o = \frac{1}{T_o}, \tag{8.1}$$

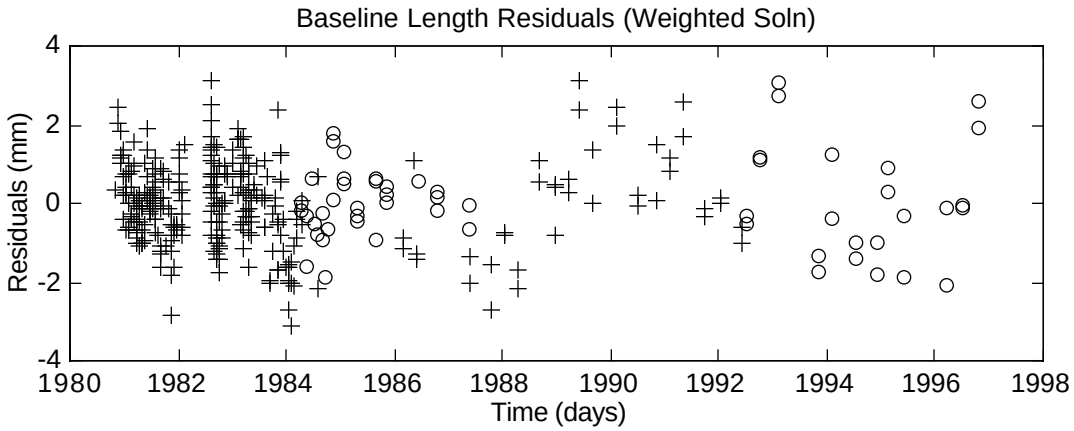

Figure 8.3: Comparison of residual baseline length changes after removal of estimated distance offsets for each observation group and a common linear trend. Different observation groups are denoted by different symbol colour/type combinations.

where $T_o = (t_{max} - t_{min})$ is the fundamental period and f_o is the fundamental frequency (see Section 3.2, eqn. (3.17)). The largest frequency that can be determined by the data series is defined by the Nyquist frequency f_N. It corresponds to the time interval over a triplet of adjacent points, the minimum number of points for the unambiguous determination of a periodic component.

The Nyquist frequency is not clearly defined for unevenly spaced data. For evenly spaced data, it is simply twice the time interval between any pair of adjacent points (i.e., twice the sampling interval Δt). The Nyquist frequency is then defined as $f_N = 1 / (2\Delta t)$ (cf. Section 3.2). This represents the largest frequency (smallest period) the data series is capable of reliably estimating without aliasing effects. For unevenly spaced data series, the distribution of possible triplets of points can vary significantly and thus there is no well defined Nyquist frequency present. In theory, the highest frequency (that can be estimated from a data series) will correspond to the smallest point triplet interval. This interval corresponds to the smallest period (maximum frequency) that can possibly be determined

from the data series. However, in practice, the spectra generally exhibit no mirror image about this or any other frequency when the data are unevenly and randomly spaced. The exception is when dealing with data that are regularly spaced as multiples of some common interval or evenly spaced except for gaps.

For the baseline length residuals in Figure 8.3, the variation in possible Nyquist frequencies is illustrated in Figure 8.4 in terms of histograms of the lengths (time intervals) of all possible point triplets ("Nyquist periods"). The smallest triplet interval is about 1 day corresponding to a Nyquist frequency of 1 cy/day. This is because the measurements were collected on a regular daily basis in the beginning. In the following analyses, spectra are therefore estimated at integer multiples of the fundamental frequency up to a Nyquist frequency of 1 cy/day.

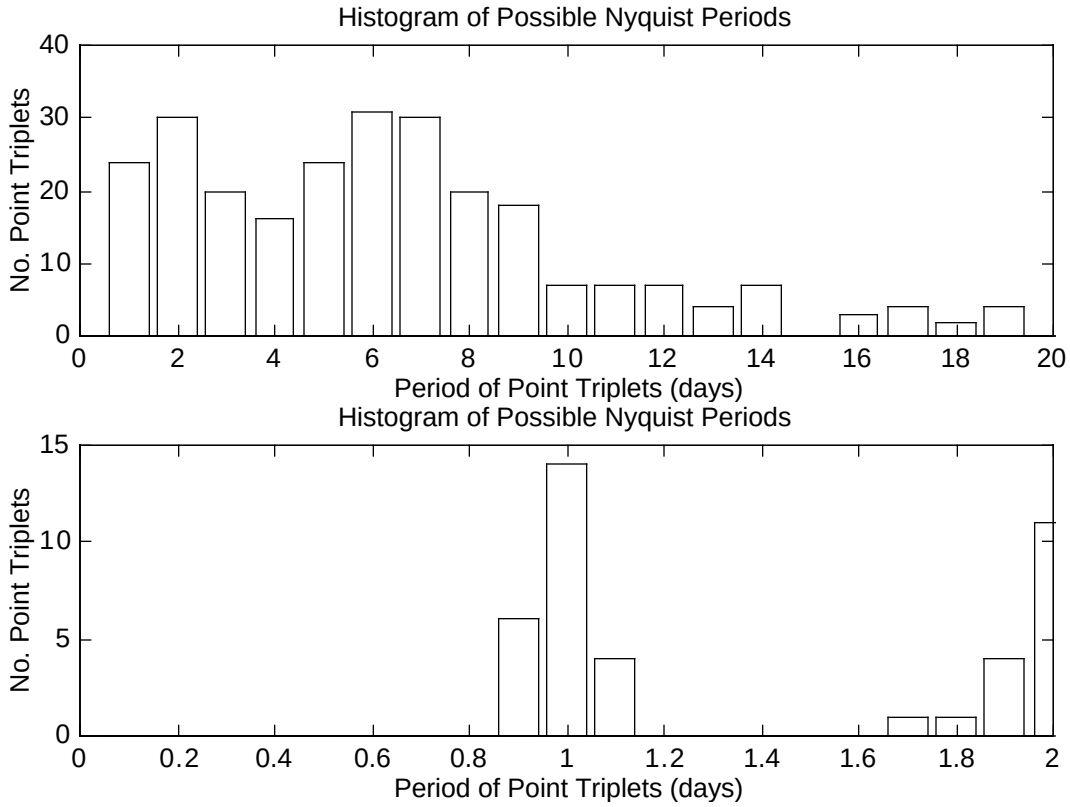

Figure 8.4: Histograms of lengths of point triplets ("Nyquist periods") corresponding to possible Nyquist frequencies. Bottom plot gives a more detailed histogram at 1 day.

In the estimation of the weighted least squares spectrum of the baseline length residuals, any linear dependence (mathematical correlation) with the estimated deterministic model (distances offsets and linear trend) are taken into account as described in Sections 4.5 and 5.5. The spectrum is plotted in Figure 8.5 with respect to period instead of frequency for easier interpretation. There are clear significant spectral components at periods of 2 and 8 years, in addition to several peaks at periods shorter than a year. The lower plot in Figure 8.5 enlarges the short period range and shows significant spectral components at periods of about 1, 100, 150 and 200 days.

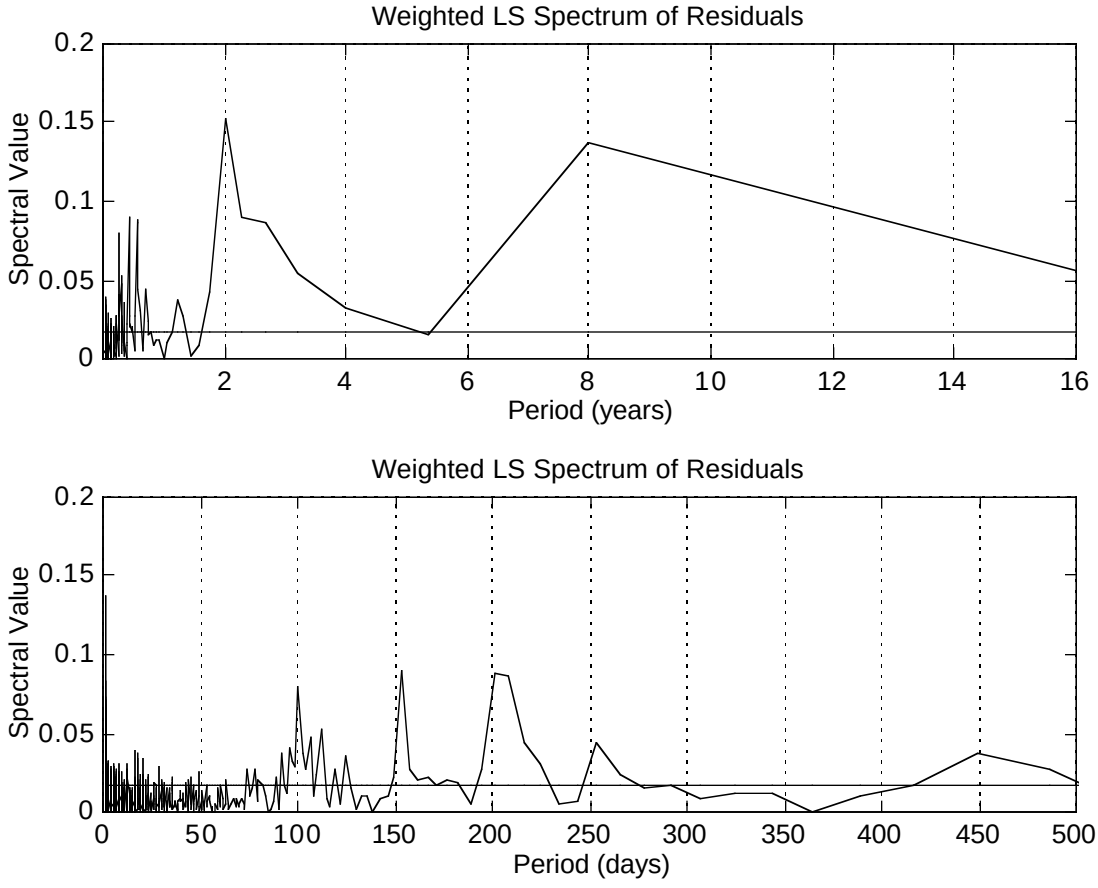

Figure 8.5 Weighted least squares spectra (independently estimated) of baseline length residuals from the deterministic model in Table 8.1. The horizontal line is the 95% confidence interval for detecting significant spectral values.

The 8 year period is interesting because it is also visible in the residuals between about 1984 and 1996 (see Figure 8.3). It was thought that this might be due to a possible additional datum offset at about 1988.7 in the data group between 1984.2 and 1992.5 (see Figure 8.2). Apparently, the instrumentation had been taken down and set up again at this time but it was thought that this was done accurately so as not to produce any additional bias in the distance measurements (J. Langbein, personal communication, 21 March 1997). To check for the significance of such a bias, an additional datum offset was estimated at 1988.7. This resulted in replacing the 1984.2-1992.5 group (with datum offset #5) with two new groups; 1984.2-1988.7 with datum offset #5 and 1988.7-1992.5 with new datum offset #5a. Figure 8.6 shows these two new groups together with the time series of length changes. The least squares estimates of the model with the additional offset (#5a) are given in Table 8.2 and the residual series after removing the model is given in Figure 8.7. It was found that the datum offsets #5 and #5a for the two new groups were statistically different from each other at any reasonable significance level (t statistic = 7.0) and both biases were therefore modelled in the following analyses.

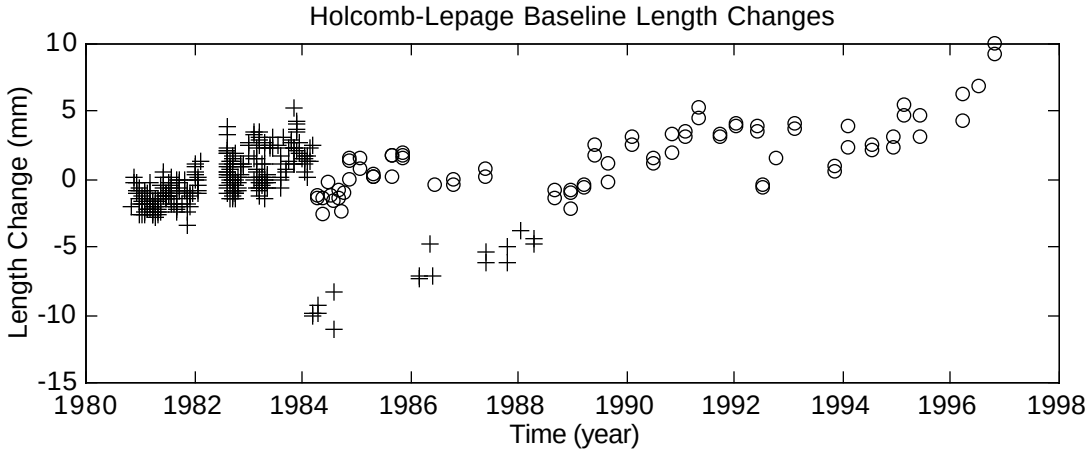

Figure 8.6: Changes in length of Holcomb to Lepage baseline with additional datum offset in observation group from 1984 to mid-1992. Different observation groups are indicated by different symbol colour/type combinations.

Table 8.2: Least squares estimates of linear trend and datum offsets, including additional datum offset (#5a).

	Estimate	Std	t Statistic
Offset #1 (mm)	-2.0	0.1	19.3
Offset #2 (mm)	-2.8	0.2	15.1
Offset #3 (mm)	-3.6	0.2	15.5
Offset #4 (mm)	-4.2	0.5	8.7
Offset #5 (mm)	-15.3	0.4	36.9
Offset #5a (mm)	-12.6	0.6	21.9
Offset #6 (mm)	-17.1	0.8	21.3
Offset #7 (mm)	-6.2	0.4	17.4
Offset #8 (mm)	-9.2	0.5	17.0
Linear Trend (mm/yr)	1.51	0.06	26.9

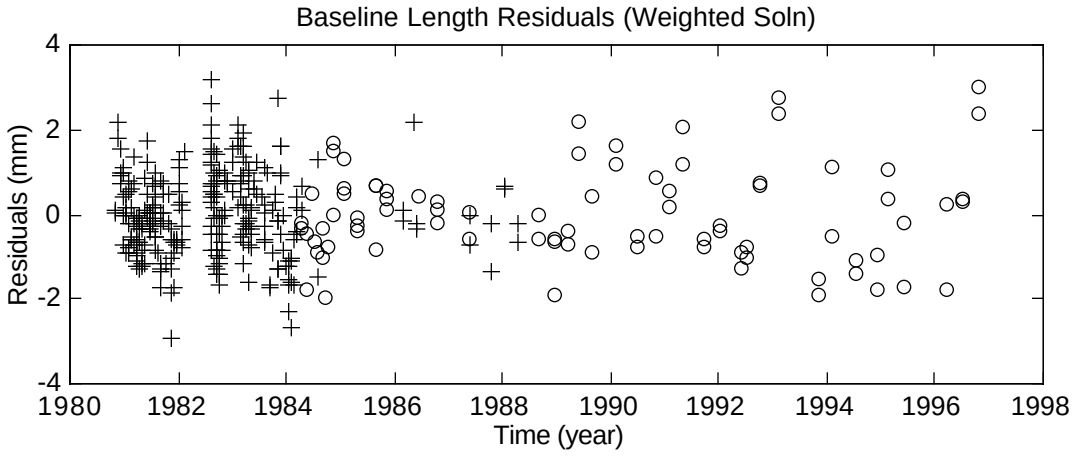

Figure 8.7: Comparison of residual baseline length changes after removal of estimated datum offsets, including additional offset, for each observation group and a common linear trend for all groups. Different observation groups are denoted by different symbol colour/type combinations.

The weighted least squares spectrum for the residuals after removing the estimated deterministic model with the additional datum offset is given in Figure 8.8. The most obvious difference from the previous spectrum is that the peak at 8 years has now been significantly reduced by the introduction of the additional datum offset in the model. However, there still remains a large peak at about 1000 days (2.5 years) that accounts for 15% of the noise in the residual data series. One possible explanation for such an interannual behaviour may be an El Niño warming effect, which has frequencies of between 2 and 4 years during this time period. The warming effect is generally

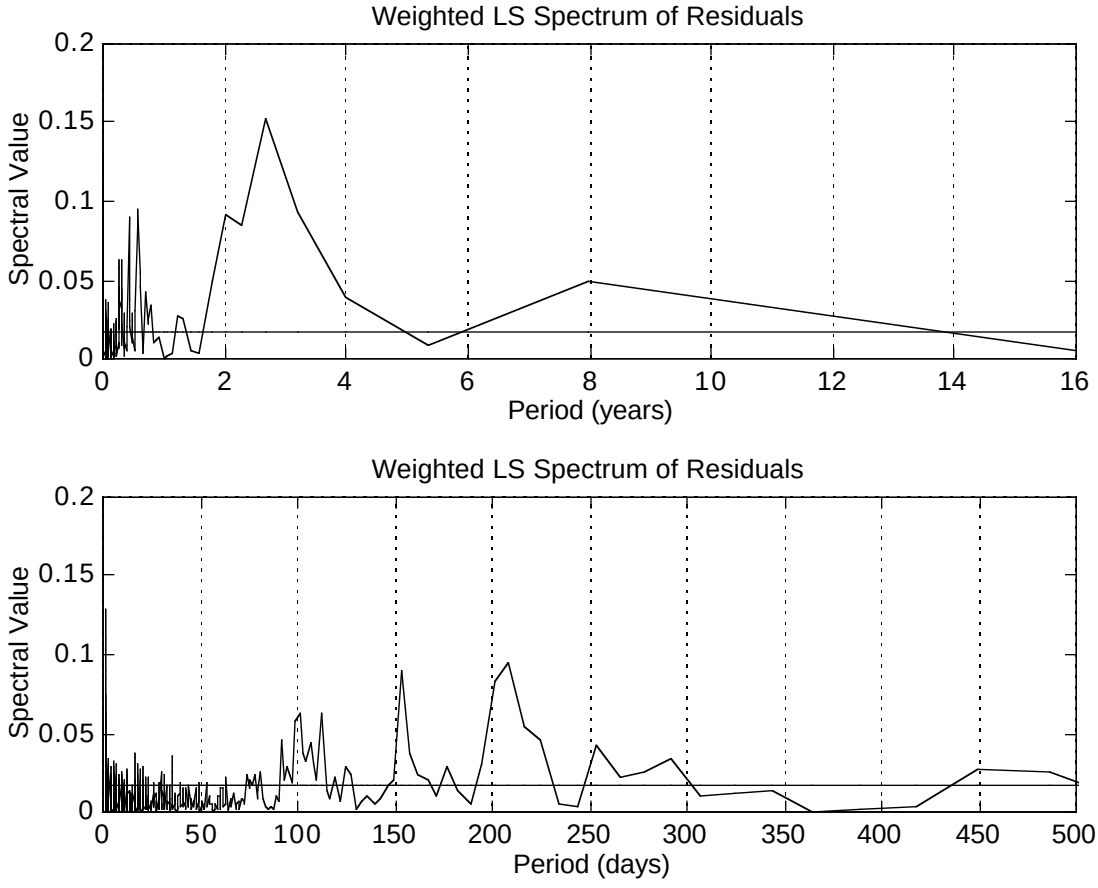

Figure 8.8 Weighted least squares spectra of baseline length residuals from the deterministic model with additional distance offset. The horizontal line is the 95% confidence interval for detecting significant spectral values.

accompanied by more frequent and severe wet weather which could cause monument motion due to higher levels of ground water. In addition, the "piling up" of warmer waters in the eastern Pacific could also possibly led to additional crustal loading on the western seaboard of North America. The other significant peaks are at short periods and are more clearly identified in the lower plot of Figure 8.8. The largest peaks in this frequency range are at about 150 and 210 days. Curiously, these peaks are symmetrical (±30 days) about small central peak with a semi-annual period (180 days). According to Vaníček [1969b], this corresponds to a possible modulation of a semi-annual period by a 30 day period. The semi-annual period may be related to weather. For example, it is well known that southern California generally has wet spring and fall and a dry summer and winter which could conceivably cause a semi-annual period in the presence of ground water, thus possibly contributing to a semi-annual behaviour of the motions of the geodetic monuments. The 30 day period may be related to lunar tidal effects. Other peaks evident in the spectrum are at period of about 110 days and 1 day. The diurnal period is believed to be a consequence of the usual diurnal behaviour of many systematic effects related to atmospheric conditions, such as atmospheric refraction and heating (expansion) of the ground and monuments. The other notable feature of the spectrum is the absence of an annual period. In fact, the spectral value for this period is almost exactly zero, indicating that such a period had already been removed from the data. This was denied by Langbein (personal communication, 21 March 1997), however.

Langbein and Johnson [1997] also argue for the presence of a random walk signal in the residual data series. Their spectrum for the Holcomb-Lepage baseline was computed by first interpolating the unevenly spaced measurement series to an evenly spaced one by averaging the data spanning 15-35 days either side of the missing point. White noise was also added to their interpolated value. The power spectrum was then computed using the FFT technique and plotted against the log of the frequency (see Langbein and Johnson [1997, Figure 3]). Their plots display a clear trend proportional to $1/f^{-2}$ as expected for a

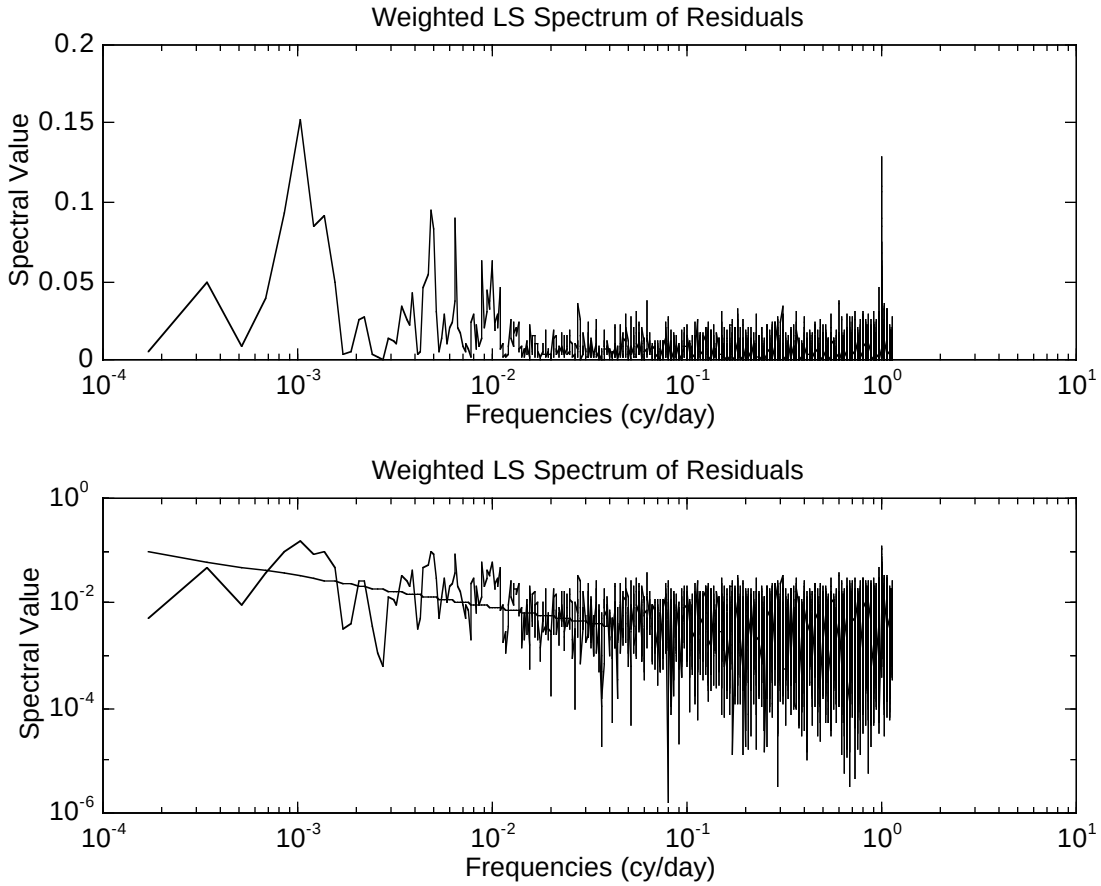

Figure 8.9 Semi-log (top) and log (bottom) plots of weighted least squares spectra of baseline length residuals from the deterministic model with additional datum offset. The straight line represents a –0.60 linear trend at low frequencies ($f<4x10^{-2}$).

random walk process (see Section 7.7). For comparison, the weighted least squares spectra is displayed in Figure 8.9 (top plot) using the same semi-log frequency plot. No clear $1/f^{-2}$ trend is apparent in this spectrum. The spectrum is also displayed in Figure 8.9 (bottom plot) using a full log plot, where the presence of random walk noise should produce a negative linear trend at low frequencies, as discussed in Section 7.7. A small negative trend (–0.60 ± 0.08) is visible in the least squares spectrum at frequencies below 4

x 10^{-2} cy/day, which grows even smaller for higher frequencies. However, this linear trend is proportional to $1/f^{-0.6}$, rather than $1/f^{-2}$ as characteristic of a random walk process.

The autocorrelation function for the observations was indirectly estimated from the inverse least squares transform of the independently estimated, weighted least squares spectrum following the iterative procedure outlined in Section 6.4. The a priori standard deviations of the data were used to generate a priori observation weights.. The data series was also zero-padded prior to computing the spectrum to avoid any wrap around effects in the autocorrelation function as described in Section 3.4. The main difficulty encountered was with the large number of possible time lags for which the autocorrelation needed to be computed. For unevenly and randomly spaced data, there are in general as many different lags as there are combinations of observation pairs. For the Holcomb-Lepage distance measurements, there are 361 observations for which there are 65,341 unique possible time lags (number of off-diagonal elements in the observation covariance matrix). It was therefore impractical to compute the autocorrelation function for all lags at once. Instead, the ACF was computed separately for the lags corresponding to each row of the observation covariance matrix. Only the autocorrelations for the upper triangular part of each row needed to be computed. The entire correlation matrix $\boldsymbol{R}$ for the observations was assembled in this way and the full covariance matrix $\boldsymbol{C}$ was obtained using the a priori standard deviations of the observations (which were also used in the computation of the weighted spectrum); i.e.,

$$\boldsymbol{C} = \boldsymbol{S}\,\boldsymbol{R}\,\boldsymbol{S}\,, \tag{8.2}$$

where $\boldsymbol{S}$ is a diagonal matrix of the a priori standard deviations. The autocorrelation function for is plotted in Figure 8.10 together with an enlargement at short lags. Although there is a periodic behaviour in the enlarged plot, the magnitude of the correlations are small even for short lags. No explanation was found for small correlation “spikes”.

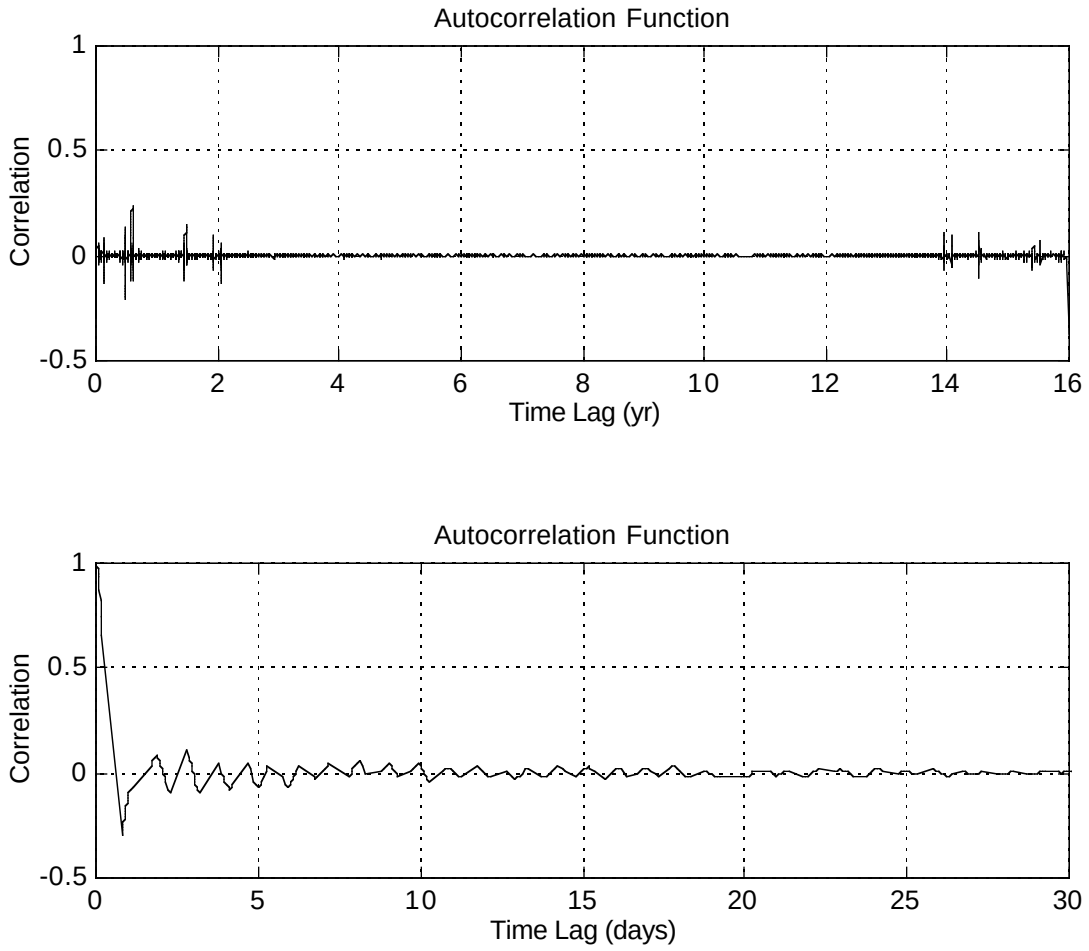

Figure 8.10: Indirect ACF, and enlargement at short lags, estimated from zero-padded time series of Holcomb-Lepage length changes with additional datum offset.

The deterministic model of the datum offsets and linear trend were re-solved using the new full covariance matrix. The solution is given in Table 8.3 with the additional datum offset (#5a) at 1988.7 included. Because of the small correlations, there is little difference in the estimated offsets and trend between this solution and that based on only a diagonal covariance matrix (Table 8.2); all are statistically compatible. However, in most cases the estimated standard deviations of the offsets and trend are larger when the full covariance matrix is used, indicating greater uncertainty in the estimated parameters.

Table 8.3: Least squares estimates of linear trend and datum offsets, including additional offset (#5a) and using estimated full observation covariance matrix based on computed ACF.

	Estimate	Std	t Statistic
Offset #1 (mm)	-1.7	0.2	7.8
Offset #2 (mm)	-3.6	0.2	16.2
Offset #3 (mm)	-4.6	0.2	20.8
Offset #4 (mm)	-4.1	1.0	3.9
Offset #5 (mm)	-14.0	0.3	51.3
Offset #5a (mm)	-11.9	1.0	12.3
Offset #6 (mm)	-16.2	1.3	12.4
Offset #7 (mm)	-5.8	0.6	9.3
Offset #8 (mm)	-8.9	1.2	7.4
Linear Trend (mm/yr)	1.44	0.09	16.3

Specifically, the standard deviation for the linear trend is increased from 0.06 to 0.09 mm/yr. This is thought to be caused by a slight reduction in the overall redundancy due to the linear dependence (mathematical correlations) among the observations. There were also some significant differences in the correlations between the estimated parameters. For example, the correlation between offsets #5 and #5a was reduced from 0.75 to 0.44. This caused the difference between the two offsets to become less statistically significant (t statistic reduced from 7.0 to 2.4). Nevertheless, the difference is still statistically significant at the 95% confidence level, leading us to still consider the possibility that the addition datum offset is real.

Table 8.4: Summary of estimated linear trends with and without extra offset and correlations.

	Linear Trend ± Standard Deviation (mm/yr)	
	Without Corr.	With Corr.
Without extra offset	1.72 ± 0.05	1.61 ± 0.07
	1.67 [Langbein & Johnson]	
With extra offset	1.51 ± 0.06	1.44 ± 0.09

Finally, the estimated linear trends (baseline expansion) are summarized in Table 8.4. The main difference with the estimate from Langbein and Johnson [1997] is due to the use of the additional datum offset #5a. When the offset is not used, the estimated trend with or without the observation correlations is not significantly different from Langbein and Johnson's. The differences are well within the 95% confidence intervals. When the extra offset is used, the linear trends are reduced by about 0.2 mm/y with or without the use of correlations. These are significantly different at the 95% confidence level. The standard deviation of the linear trend is only slightly increased by the additional offset.

The use of observation correlations derived from the estimated autocorrelation function also reduces the magnitude of the linear trends both with and without the extra offset. However, the reduction is only about 0.1 mm/yr in both cases and is not statistically significant at the 95% confidence level. The correlations also increase the estimated formal standard deviations of the linear trends by about 50%, even though the magnitude of the autocorrelation is relatively small. This increase is thought to be due to a implied reduction in the total redundancy (the existence of correlations means there are effectively fewer truly independent observations).

Finally, it is noted that the estimated linear trend (1.44 ± 0.09 mm/yr) with the extra offset and correlations agrees better with the linear trend (1.46 mm/yr) estimated by

Langbein and Johnson for the baseline from station Holcomb to station Bird, which is in the same general vicinity and direction as station Lepage (see Figure 8.1). The baselines to these two station should therefore behave similarly in terms of its motion relative to Holcomb. The apparent agreement therefore supports the existence of an extra datum offset in the measurements to Lepage.

8.3 GPS Point Positioning

The use of the Global Position System (GPS) has grown greatly in recent years, largely owing to the wide availability of small, low cost receivers. For an in depth explanation of the concepts involved in GPS, see Wells et al. [1986] or Dana [1997]. In its most basic mode of operation, referred to as the Standard Position Service, users can obtain their position to an accuracy of only about 100 metres horizontally and about 150 metres vertically. In this mode, GPS receivers make use of the so-called C/A code pseudo-range (measured satellite-to-receiver distance), which is obtained by timing the satellite-to-receiver travel time of the basic C/A (coarse acquisition) code that is superimposed on the L1 carrier frequency. The satellite-to-receiver ranges are used to solve for the receiver's position in what is essentially known as a 3-dimensional resection problem in surveying. This mode of positioning is called “point positioning” to distinguish it from other, more accurate, methods based on relative or differential positioning between receivers.

Although the pseudo-range observable is capable of providing point positioning accuracies of about 10 to 30 metres, the US Department of Defense intentionally degrades the observable to the 100 m level for security reasons. This degradation is called Selective Availability (S/A) and involves the introduction of systematic errors in the form of a mathematical algorithm (called “dithering”) into the broadcast satellite ephemeris and clock. To date only clock dithering has apparently been applied. This error propagates directly into the signal travel time from which the pseudo-range observable is derived. However,

because the S/A error is fairly systematic, there exist very large autocorrelations in the pseudo-range data and thus also in the estimated positions derived from them. Here, only the vertical position is examined with the aim of investigating the degree of autocorrelation in the data and the effect of using the autocorrelation function to weight the point positions when using time averaging to reduce the effects of S/A and improve the accuracy. The analysis of the two horizontal components can be done in an analogous fashion.

The data used in this study were provided by W. Prescott of the US Geological Survey (personal communication, 19 May 1994). They were obtained from a geodetic quality Ashtech L-XII GPS receiver and included, for each measurement epoch, the receive time, computed WGS-84 Cartesian coordinates of the receiver's antenna and computed receiver clock bias. The time series of instantaneous point positions refers to station Chabot in the south part of Oakland, California. The point positions were recorded every 30 seconds for a total of 24 hours on April 6, 1994. Plots of the variation in the horizontal and vertical position estimates over this 24 hour period are given in Figure 8.11, and for only the first hour in Figure 8.12. The high degree of autocorrelation at short time intervals is readily apparent from the very systematic way in which the positions slowly vary.

As already stated, the most common method of reducing the effects of S/A is to average the point positions over time. Generally, users average their positions over intervals as short as 5 minutes and at most about an hour. Here, one hour averaging is used to examine the effectiveness of this in reducing the effects of S/A. This provides for 24 independent hourly means.

For each hour, the least squares spectrum is computed. Any linear dependence between the estimated mean and the spectral components is accounted for as described in Section 5.5. The hour long subsets are also zero-padded to avoid any wrap around effects in the derived autocorrelation functions (see Section 3.4). The systematic nature of S/A is revealed as statistically significant peaks in the spectra, mainly at lower frequencies. The independently estimated least squares spectrum for first hour of the height series is given in

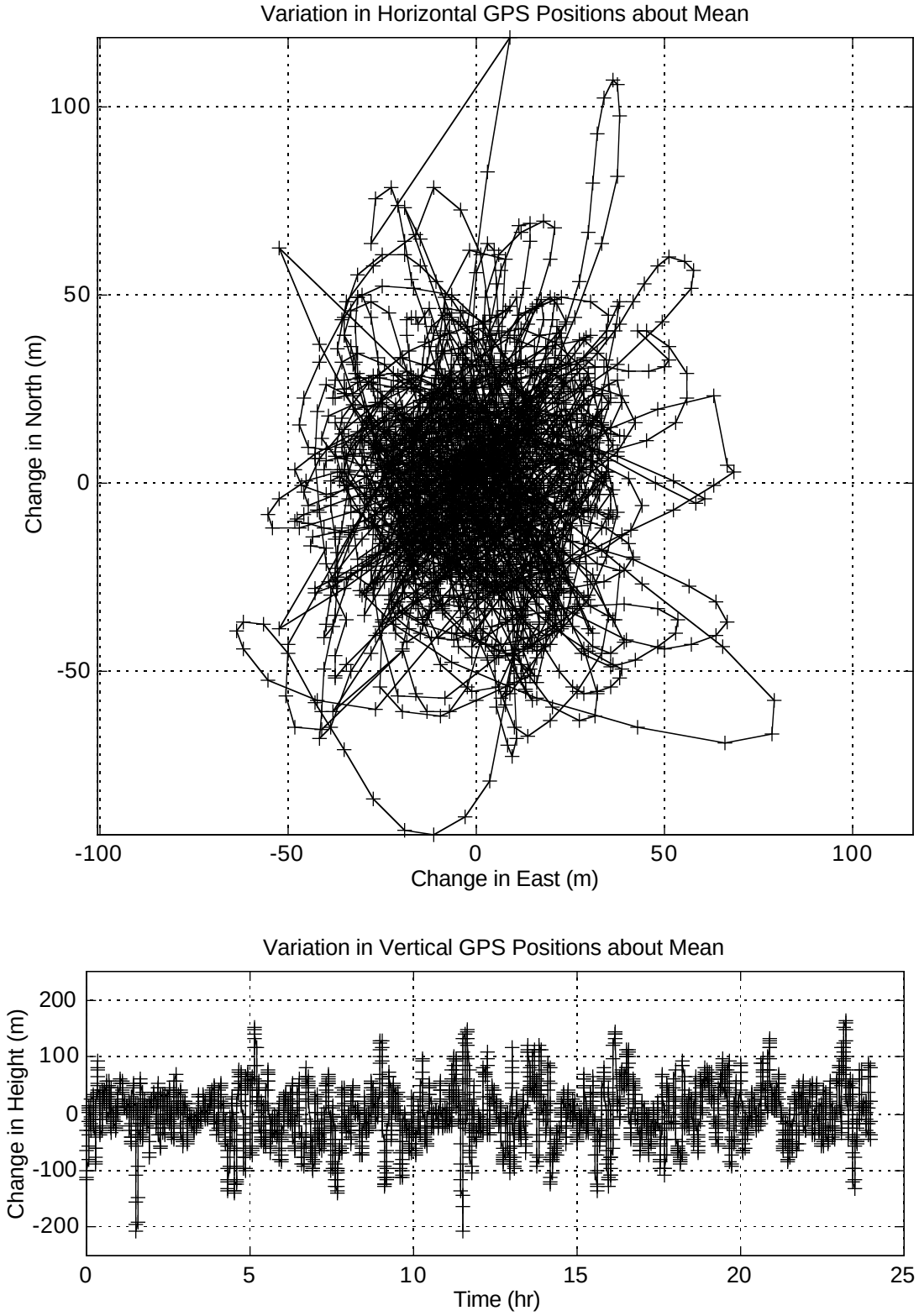

Figure 8.11: Variations in derived horizontal (top) and vertical (bottom) GPS positions over 24 hours at station Chabot. Variation is with respect to mean position.

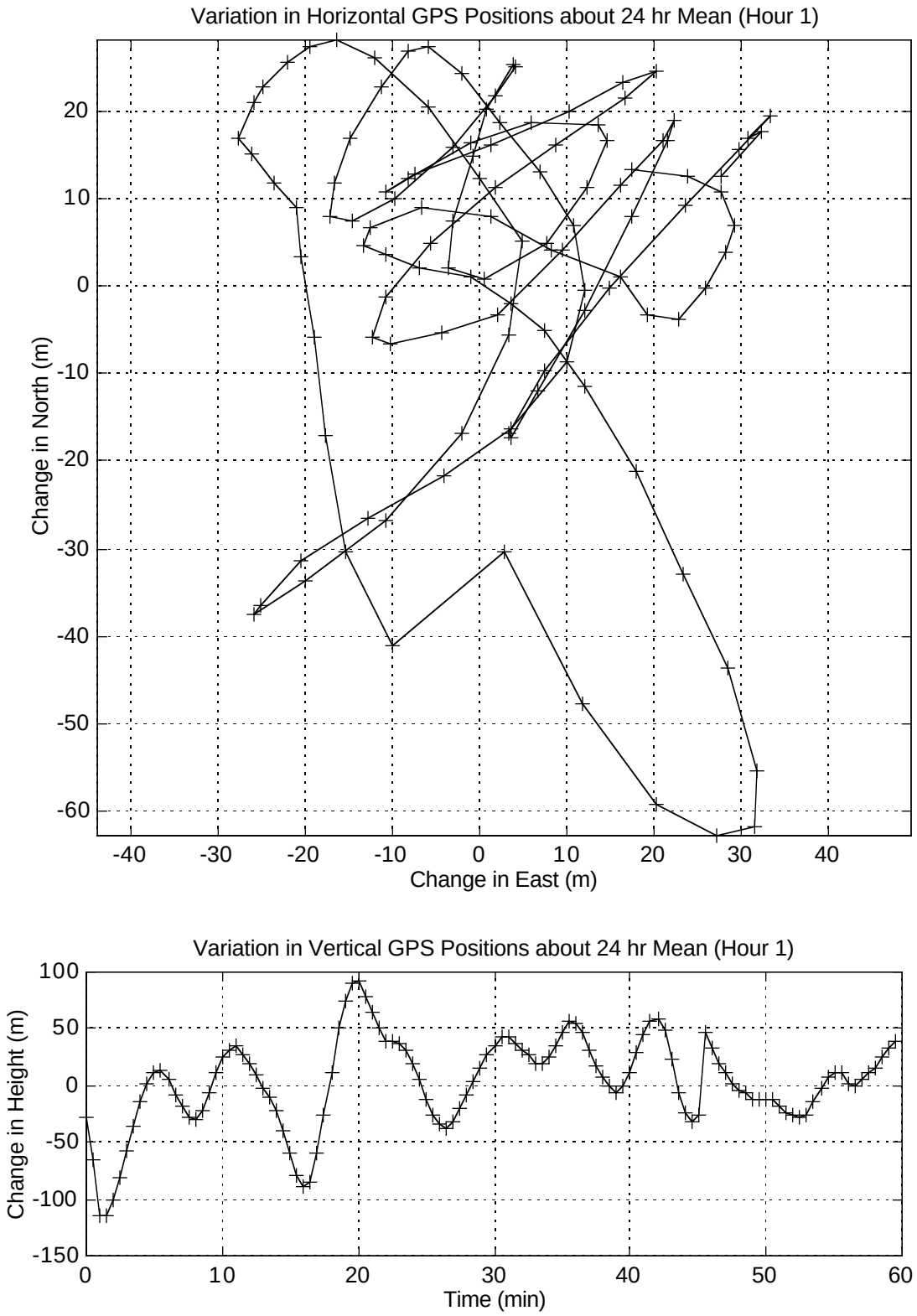

Figure 8.12: Variations in recorded horizontal (top) and vertical (bottom) GPS positions for the first hour at station Chabot. Variations are with respect to 24 hour mean position.

Figure 8.13, where the S/A effects are clearly evident. The peak around 0.1 cy/min (period = 10 min) is typical also of the other hourly data sets. The spectra are sufficiently different from hour to hour, however, so that it is not possible to predict the S/A algorithm with only one hour of data.

The inverse least squares transform of the zero-padded spectrum was used to obtain the (biased) autocorrelation function (ACF) for each hour. The biased estimate of the ACF for the first hour is given Figure 8.14. There are strong, periodic correlations present which are typical also of the other hourly subsets. For each hourly subset the estimated autocorrelation function is used to populate a full correlation matrix for the observations. These are then used as the a priori weight matrix in the re-estimation of the hourly means; referred to here as the "weighted" means. Note, however, that the ACF used in the weighted mean was computed by first removing the unweighted mean (using the identity matrix for the a priori weight matrix). Consequently, the weighted mean was re-estimated by iterating the computation of the ACF using the weighted mean from the previous iteration as described in Section 6.4 and illustrated in Figure 6.1. The standard deviations of the means were also scaled by the square root of the estimated variance factor of the height residuals from each hourly mean solution.

The unweighted means (equally weighted observations with no correlations) and weighted means (with correlations based on the estimated ACF) were obtained in the above manner for each hourly subset of data. The results are summarized in Table 8.5 and plotted in Figure 8.15. It is evident from these results that there is a systematic variation in the hourly means. The average of the formal standard deviations is only about 4 m whereas the RMS of the hourly means is about 13 m. This clearly indicates that the formal standard deviations of the estimated means are too optimistic. The use of observation correlations derived from the ACFs estimated from each hour of data does not improve the results. The average formal standard deviation is still much smaller than the RMS.

Table 8.5: Unweighted and weighted hourly means and their standard deviations (Std) of GPS height measurements over a 24 hour period. "Average", "Range" and "RMS" denote the arithmetic average (unweighted), range (max – min) and root mean square (about the "Average"), respectively, of these hourly values and are provide for illustration only.

	Unweighted		Weighted	
	Mean	Std	Mean	Std
Hour 1	229.394	3.639	231.817	5.004
Hour 2	222.816	4.610	221.524	6.561
Hour 3	231.493	2.684	231.041	0.746
Hour 4	226.876	1.930	226.870	3.375
Hour 5	213.093	5.597	208.093	5.871
Hour 6	223.762	4.923	226.615	9.201
Hour 7	214.982	4.171	215.731	5.191
Hour 8	201.366	4.476	199.889	5.600
Hour 9	230.347	3.321	231.344	4.800
Hour 10	208.072	5.017	210.205	2.675
Hour 11	219.493	3.016	221.364	3.082
Hour 12	235.390	5.804	237.343	5.825
Hour 13	225.044	4.125	228.568	8.124
Hour 14	242.865	4.951	247.190	6.306
Hour 15	211.883	3.892	210.867	4.932
Hour 16	215.865	4.335	216.867	1.704
Hour 17	248.885	5.524	252.355	8.125
Hour 18	213.339	3.651	212.270	3.928
Hour 19	228.925	4.484	233.407	6.505
Hour 20	245.503	4.188	246.744	3.477
Hour 21	244.350	3.941	247.259	8.296
Hour 22	221.701	3.576	218.515	3.165
Hour 23	227.920	3.058	226.854	2.021
Hour 24	244.776	5.562	241.927	9.138
Average	225.173	4.186	226.861	5.152
Range	47.519	3.874	52.466	8.455
RMS	12.576		13.691	

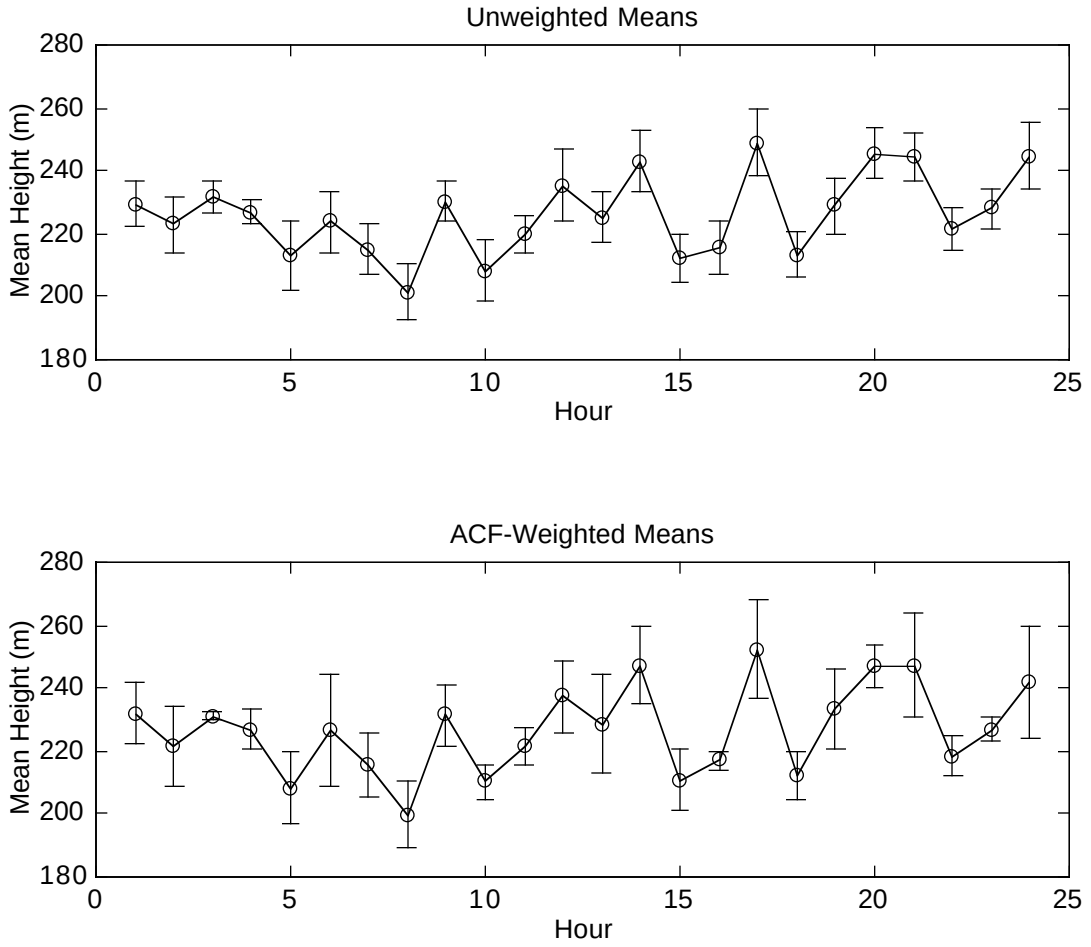

Figure 8.15: Unweighted (top) and weighted (bottom) hourly means of GPS height measurements over a 24 hour period. The ACF-weighted means are based on iterative estimation of the ACF. Error bars represent 95% confidence intervals.

From these results it is clear that the one hour averaging does not eliminate all the effects of S/A. To gain a better understanding of the nature of S/A for this data set, the least squares spectrum was computed for the entire 24 hour data set. The spectrum is plotted in Figure 8.16 and the largest peaks are tabulated in Table 8.6. It is clear from these results that there are many significant periodic components at periods longer than one hour. In fact, the largest peak has a period of about 1.5 hrs. There is also a peak at 24 hrs, which is the fundamental period of the data series. This could indicate a 24 hour period (possibly diurnal atmospheric effects) or spectral leakage from periodic components beyond 24 hours. Without a longer data series, it is not possible to distinguish between these two possibilities. Regardless, it is evident that the hourly means are not capable of "averaging out" all of the systematic effects. Longer averaging is clearly needed.

The hourly means were recomputed using the autocorrelation function from the entire 24 hour data set. The autocorrelation function is plotted in Figure 8.17 and the new weighted hourly means tabulated in Table 8.7 and plotted in Figure 8.18. The lower plot in Figure 8.18 shows the differences between these means and the unweighted ones. All differences are within the 95% confidence interval. Although the estimated means still exhibit the same systematic trend as for the unweighted ones, the formal standard deviations are much larger. The average of these formal standard deviations (12 m) is very nearly the same as the RMS of the means (13 m). This indicates that the formal standard deviations are now much more realistic, accounting for the uncertainty due to the systematic effects still present in the results.

As stated in the previous section, this increase in the formal standard deviations is thought to be caused by a reduction in the overall redundancy due to correlations among the individual position estimates. The observations can no longer be considered independent of each other. The degrees of freedom is thus somewhat less than the number of observations minus the number of unknowns. The larger standard deviations represent a

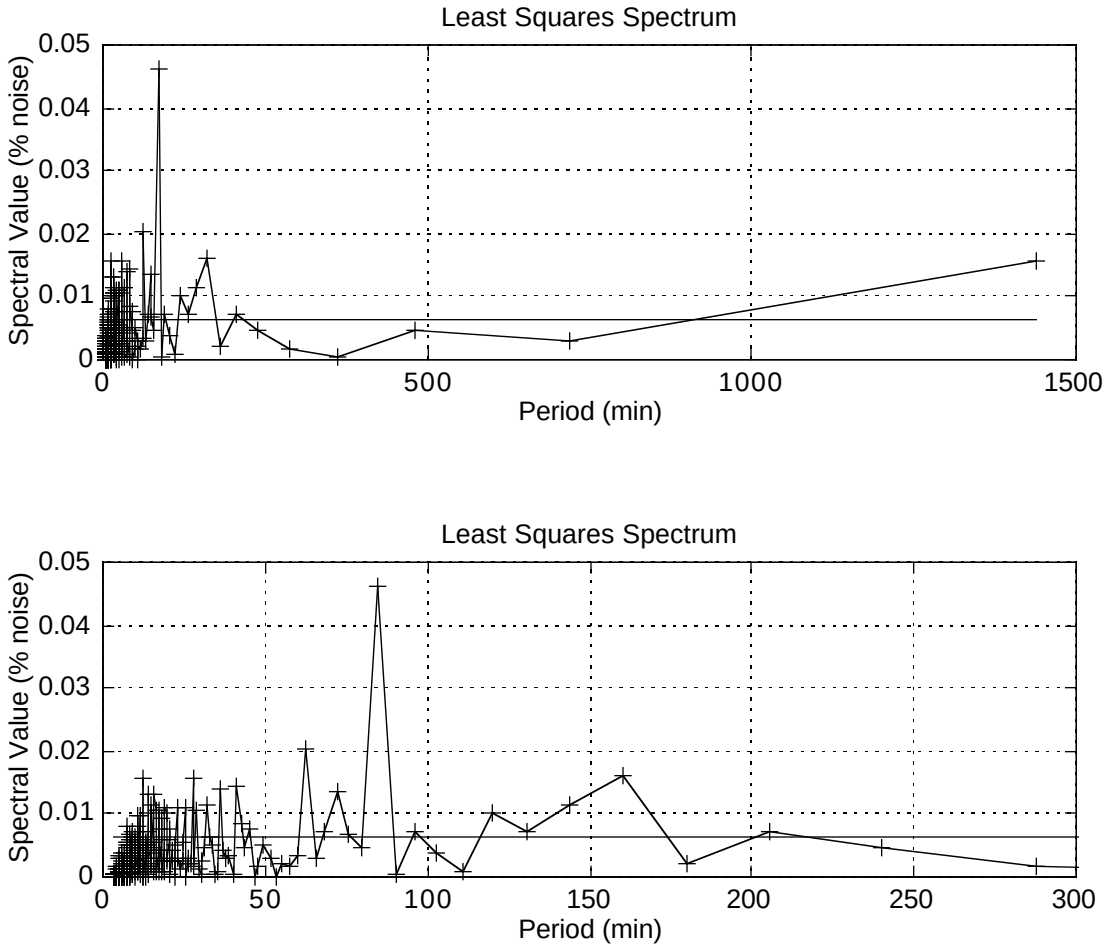

Figure 8.16: Least squares spectrum (independently estimated) for entire 24 hour data set. The lower plot is an enlargement of the short period range. The horizontal line is the 95% confidence interval for detecting significant spectral values.

Table 8.6: Twenty of the largest peaks in least squares spectrum in Figure 8.16.

Period (min)	Spectral Value
84.7	0.046
62.6	0.020
160.0	0.016
27.7	0.016
12.3	0.016
1440.0	0.016
41.1	0.015
36.0	0.014
72.0	0.013
16.0	0.013
13.7	0.013
32.0	0.011
15.2	0.011
144.0	0.011
19.2	0.011
22.9	0.011
25.3	0.011
17.3	0.011
16.2	0.011
28.8	0.010

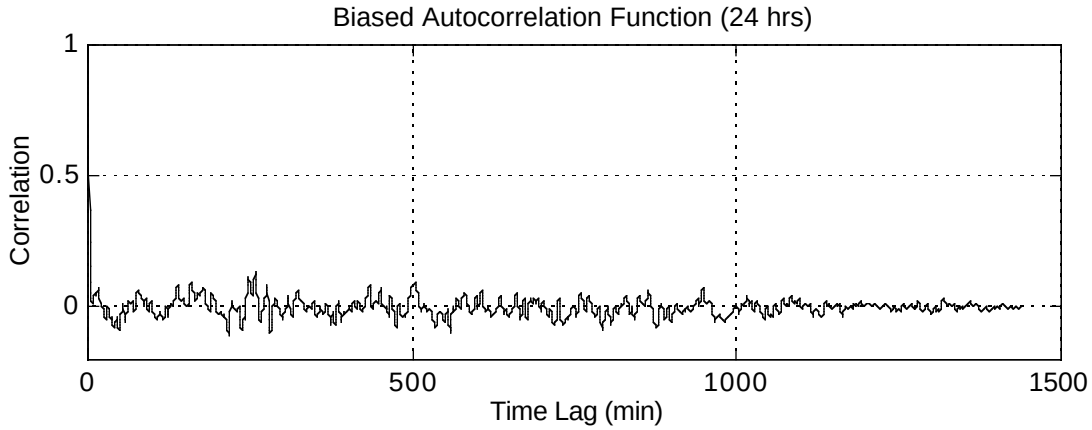

Figure 8.17: Autocorrelation function for entire 24 hour data set.

Table 8.7: Weighted hourly means of GPS height measurements and their standard deviations (Std) over a 24 hour period using correlations from ACF based on 24 hours of data. "Average", "Range" and "RMS" denote the arithmetic average (unweighted), range (max – min) and root mean square (about the "Average"), respectively, of these hourly values and are provide for illustration only.

	Mean	Std
Hour 1	231.277	12.598
Hour 2	227.152	14.071
Hour 3	232.981	6.455
Hour 4	226.417	7.193
Hour 5	215.888	11.572
Hour 6	225.890	11.139
Hour 7	214.706	7.610
Hour 8	203.544	14.945
Hour 9	234.686	11.630
Hour 10	210.003	9.070
Hour 11	218.894	10.390
Hour 12	234.644	22.224
Hour 13	229.946	18.557
Hour 14	246.397	11.733
Hour 15	213.002	7.875
Hour 16	214.925	10.002
Hour 17	246.554	16.185
Hour 18	209.304	11.909
Hour 19	231.577	6.678
Hour 20	249.461	9.964
Hour 21	247.885	11.907
Hour 22	227.453	7.648
Hour 23	232.254	12.413
Hour 24	249.421	21.283
Average	228.094	11.877
Range	45.917	15.769
RMS	13.289	

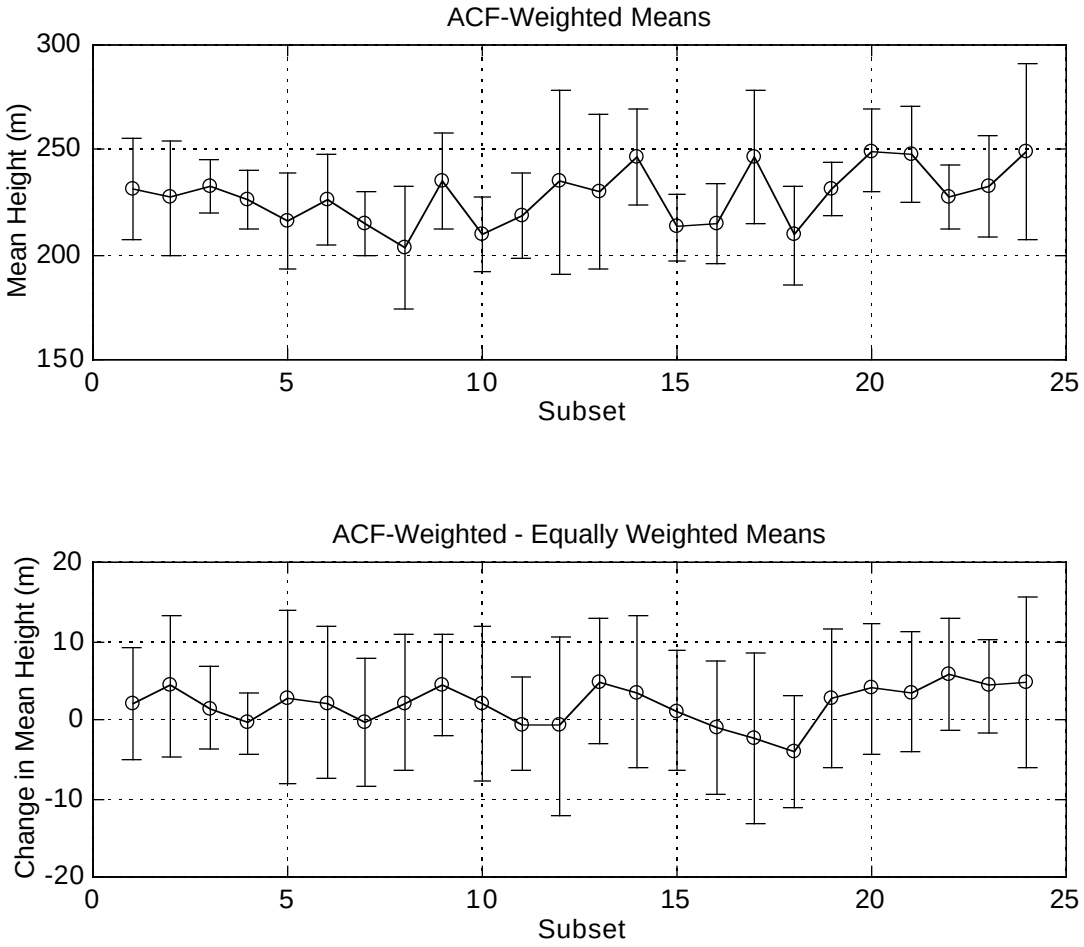

Figure 8.18: Weighted (top) hourly means of GPS height measurements over a 24 hour period using correlations obtained from ACF based on 24 hours of data, and difference with equally weighted means without correlations (bottom). Error bars represent 95% confidence intervals. Note different vertical scale in top plot from that used in Figure 8.15.

more realistic estimate of precision than that derived from the unweighted means without any correlations among the individual height observations.

It should be realized that the S/A algorithm is very complex and a closely guarded secret. It cannot be assumed the effect would repeat from day to day to enable its prediction. Furthermore, the systematic signals evident in the spectrum and autocorrelation function may also be due to other effects, such as tropospheric and ionospheric delays and signal multipath. Consequently, the systematic positioning errors can be expected to vary spatially so that, in general, an autocorrelation function computed for one point would not be applicable for any other point. Further investigations of spectra for longer time periods, at different times of the year and at different locations would be needed to make more general conclusions about these effects.

Chapter 9
Conclusions and Recommendations

A general least squares (LS) transform and its inverse have been developed which can accommodate unequally spaced data, random observation errors, arbitrary frequencies, arbitrarily weighted and correlated observations, as well as the effect of any a priori deterministic model on the spectral components. The LS spectrum has also been reformulated in terms of this LS transform and its inverse. The conventional Fourier transform and derived spectrum have been shown to be just a special case of the more general LS transform and spectrum.

The LS transform avoids many of the limitations associated with the traditional FFT algorithms. These include the requirements that the data be equally spaced and have 2^n data points (some FFT algorithms exist for handling other than 2^n data points, but these begin to lose the computational efficiency of the FFT). It is also not limited to providing information only about the "Fourier" frequencies (i.e., the integer multiples of the fundamental frequency). On the other hand, the LS transform is much more computationally demanding than the FFT.

The spectral components in the LS spectrum and inverse LS transform can be estimated either independently (out of context) of each other or simultaneously (in-context) with each other. However, the simultaneous estimation is very sensitive, in terms of ill-conditioning, to the choice of frequencies while the independent estimation is not. Independent estimation is thus a more robust estimator in general and was also found to provide a more reliable estimate of the autocorrelation function. Nevertheless, it would be possible to use the independent estimation for the identification of significant spectral

components, and to model only these components in the simultaneous estimation of the inverse LS transform (for LS approximation) or in the LS spectrum and the indirectly derived autocorrelation function. This would result in a more optimal estimation of the inverse transform and the spectrum, where the correlations among the estimated spectral components are taken into account. Care must be taken, however, when selecting these components to avoid any ill-conditioning in the simultaneous estimation due to multicollinearity among the components (see Draper and Smith [1981, p. 258] for a discussion of multicollinearity). In such cases, only the independently estimated spectral components can be used. It is recommended to further investigate the advantages and disadvantages of simultaneously estimating the spectral components.

The most significant application of the LS transform is in the rigorous determination of autocorrelation functions for unevenly spaced data series. Autocorrelation functions can be indirectly derived from the inverse LS transform of the LS spectrum. These can then be used to construct fully populated a priori weight matrices for more complete modelling of systematic effects in least squares estimation problems. The tests with real data show that the use of fully populated weight matrices generally results in an increase in the magnitude of the standard deviations of the estimated model parameters. This provides more realistic estimates of the uncertainties and is thought to be caused by an implicit reduction in redundancy due to the correlations among the observations.

The effect of correlations among the observations on the estimation of the LS spectrum was found not to be very significant, as claimed by Steeves [1981], even when the correlations are relatively large. This is an important consideration because accounting for the correlations requires the use of a fully populated observation covariance matrix in the weighted form of the LS expressions derived here, which imposes a great computational burden. Ignoring the correlations was found to have little effect on the results (estimated spectra and model parameters) but it significantly increases the computational efficiency. On the other hand, as stated above, the use of a priori

correlations among the observations generally gives more realistic estimates of the uncertainties associated with the estimated model parameters. Ignoring correlations generally results in overly optimistic estimates of the standard deviations.

Although the algorithm for the LS transform and spectrum is relatively slow compared to the FFT, the computational efficiency can be greatly improved by using the trigonometric identities employed by Vaníček [1971]. It is strongly recommended to use these in any practical use of the LS transform and spectrum. It is also recommended to investigate the use of "windowing" techniques widely used in traditional Fourier spectra to reduce the effects of aliasing (see, e.g., Press et al. [1992, pp.545-551]).

Finally, the direct and inverse LS transforms have been developed here only for one-dimensional problems. Thus, it is strongly recommended to investigate the possibility of extending the LS transform to multiple dimensions in order to allow for the estimation of multidimensional spectra and autocorrelation functions. This is especially important in geodesy, where most problems are of a two- or three-dimensional nature and where observations are affected by a wide variety of (multidimensional) systematic errors. In these cases, a multidimensional form of the LS transforms should ideally be used.

References

Bendat, J.S. and A.G. Piersol (1971). *Random Data: Analysis and Measurement Procedures*. John Wiley and Sons, Inc., New York.

Bracewell, R.N. (1989). The Fourier transform. *Scientific American*, Vol. 262, No. 6, June, pp. 86-95.

Bracewell, R.N. (1986). *The Hartley Transform*. Oxford University Press, London.

Castle, R.O., T.D. Gilmore, R.K. Mark, B.W. Brown, Jr. and R.C. Wilson (1983). An examination of the southern California field test for the systematic accumulation of the optical refraction error in geodetic leveling. *Geophysical Research Letters*, Vol. 10, No. 11, pp. 1081-1084.

Chen, Y.C. (1983). Analysis of deformation surveys – A generalized approach. Technical Report No. 94, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B.

Chen, Y.C., A. Chrzanowski and M. Kavouras (1990). Assessment of observations using unbiased minimum norm quadratic unbiased estimation (MINQUE). *CISM Journal*, Vol. 44, No. 4, pp. 39-46.

Chrzanowski, A., M. Caissy, J. Grodecki, J. Secord (1994). Software development and training for geometrical deformation analysis. Contract Report 95-001, Geodetic Survey Division, Geomatics Canada, Ottawa.

Cooley, J.W. and J.W. Tukey (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematical Computing*, Vol. 19, pp. 297-301.

Craymer, M.R. (1984). Data series analysis and systematic effects in levelling. M.A.Sc. Thesis, Dept. of Civil Engineering (Survey Science), University of Toronto,

Toronto; Technical Report No. 6, Survey Science, University of Toronto (Erindale Campus), Mississauga, Ontario.

Craymer, M.R. and P. Vaníček (1985). An investigation of systematic errors in Canadian levelling lines. Proc. Third International Symposium on the North American Vertical Datum, Rockville, MD, April 21-26, pp. 441-447.

Craymer, M.R. and P. Vaníček (1986). Further analysis of the 1981 Southern California field test for levelling refraction. *Journal of Geophysical Research*, Vol. 91, No. B9, pp. 9045-9055.

Craymer, M.R. and P. Vaníček (1990). A comparison of various algorithms for the spectral analysis of unevenly spaced data series. Presented at the Joint Annual Meeting of the Canadian Institute of Surveying and Mapping and Canadian Geophysical Union, Ottawa, May 22-25.

Craymer, M.R., P. Vaníček and R.O. Castle (1995). Estimation of rod scale errors in geodetic leveling. *Journal of Geophysical Research*, Vol. 100, No. B8, pp. 15129-15145.

Dahlquist, G. and A. Bjorck (1974). *Numerical Methods*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Dana, P.H. (1997). Global Positioning System Overview. The Geographer's Craft Project, Department of Geography, University of Texas at Austin. World Wide Web address <http://www.utexas.edu/depts/grg/gcraft/notes/gps/gps.html>

Deeming, T.J. (1975). Fourier analysis with unequally-spaced data. *Astrophysics and Space Science*, Vol. 36, pp. 137-158.

Draper, N.R. and H. Smith (1981). *Applied Regression Analysis*. Second Edition. John Wiley & Sons, New York.

El-Rabbany, A. and A. Kleusberg (1992). Physical correlations in GPS differential positioning. Proc. Sixth International Symposium on Satellite Positioning, Columbus, OH, March 17-20.

El-Rabbany, A.E-S. (1994). The Effect of Physical Correlations on the Ambiguity Resolution and Accuracy Estimation in GPS Differential Positioning. Technical Report No. 170, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B.

Ferraz-Mello, S. (1981). Estimation of periods from unequally spaced observations. *The Astronomical Journal*, Vol. 86, No. 4, pp. 619-624.

Freund, J.E. (1971). *Mathematical Statistics*. Second Edition. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Golub, G.H. and C.F. Van Loan (1983). *Matrix Computations*. The John Hopkins University Press, Baltimore, MD.

Grafarend, E.W. (1976). Geodetic applications of stochastic processes. *Physics of the Earth and Planetary Interiors*, Vol. 12, pp. 151-179.

Grafarend, E.W. (1984). Variance-covariance-component estimation of Helmert type in the Gauss-Helmert model. *ZfV*, Vol. 109, pp. 34-44.

Grafarend, E.W., A. Kleusberg and B. Schaffrin (1980). An introduction to the variance-covariance component estimation of Helmert type. *ZfV*, Vol. 105, pp. 161-180.

Hartley, R.V.L. (1942). A more symmetrical Fourier analysis applied to transmission problems. Proceedings of the IRE, Vol. 30, p. 144.

Holdahl, S.R. (1981). A model of temperature stratification for the correction of leveling refraction. *Bulletin Geodesique*, Vol. 55, No. 3, pp. 231-249.

Horne, J.H. and S.L. Baliunas (1986). A prescription of period analysis of unevenly sampled time series. *The Astrophysical Journal*, Vol. 302, pp. 757-763.

Kac, M. (1983). What is random? *American Scientist*, Vol. 71, pp. 405-406.

Kelly, K.M. (1991). Weight estimation in geodetic levelling using variance components derived from analysis of variance. M.A.Sc. Thesis, Dept. of Civil Engineering (Survey Science), University of Toronto, Toronto.

Kreyszig (1978). *Introductory Functional Analysis with Applications*. John Wiley and Sons, New York.

Langbein, J. and H. Johnson (1997). Correlated errors in geodetic time series: Implications for time-dependent deformation. *Journal of Geophysical Research*, Vol. 102, No. B1, pp. 591-603.

Langbein, J.O., R.O. Burford and L.E. Slater (1990). Variations in Fault Slip and Strain Accumulation at Parkfield, California: Initial Results Using Two-Color Geodimeter Measurements, 1984-1988. *Journal of Geophysical Research*, Vol. 95, No. B3, pp. 2533-2552.

Lomb, N.R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, Vol. 39, pp. 447-462.

Lucht, H. (1972). Korrelation in prazisionsnivellement. Wissenschafliche Arbeiten der Lehrstule fur Geodasie, Photogrammetrie und Kartographie, Nr. 48, Technische Universitat Hannover.

Lucht, H. (1983). Neighbourhood-correlation among observations in levelling networks. In H. Pelzer and W. Niemeier (Editors), *Precise Levelling: Contributions to the Workshop on Precise Levelling*, Dümmler Verlag, Bonn, pp. 315-326.

Maul, G.A. and A. Yanaway (1978). Deep sea tides determination from GEOS-3. NASA Contract Report 141435, National Aeronautics and Space Administration, Wallops Flight Center, Wallops Island, VA.

Miller, R.G. (1966). *Simultaneous Statistical Inference*. McGraw-Hill, New York.

Moritz, H. Covariance functions in least-squares collocation. Report No. 240, Department of Geodetic Science, The Ohio State University, Columbus, OH.

O'Neill, M.A. (1988). Faster than fast Fourier. *Byte*, Vol. 13, No. 4, pp. 293-300.

Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes*. 2nd Edition. McGraw-Hill Book Company, New York.

Press, W.H. and G.B. Rybicki (1989). Fast algorithm for spectral analysis of unevenly sampled data. *The Astrophysical Journal*, Vol. 338, pp. 277-280.

Press, W.H. and S.A. Teukolsky (1988). Search algorithm for weak periodic signals in unevenly spaced data. *Computers in Physics*, Vol. 2, No. 6, pp. 77-82.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing, Second Edition*. Cambridge University Press, Cambridge.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*. Academic Press, London.

Rao, C.R. and J. Kleffe (1988). *Estimation of Variance Components and Applications*. North-Holland, New York.

Rochester, M.R., O.G. Jensen and D.E. Smylie (1974). A Search for the Earth's 'Nearly Diurnal Free Wobble'. *Geophysical Journal of the Royal Astronomical Society*, Vol. 38, pp. 349-363.

Savage, J.C. and M. Lisowski (1995). Geodetic monitoring of the southern San Andreas fault, California, 1980-1991. *Journal of Geophysical Research*, Vol. 100, pp. 8185-8192.

Scargle, J.D. (1982). Studies in astronomical time series analysis II: Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, Vol. 263, pp. 835-853.

Schwarz, K.P., M. Sideris, E.G. Anderson, P. Stoliker and S.M. Nakiboglu (1986). A weight estimation scheme for the NAVD 88 readjustment. Contract Report 86-002, Geodetic Survey Division, Geomatics Canada, Ottawa.

Seeber, G. (1973). Data evaluation by covariance analysis, exercised on photographic satellite observations. Proc. Symposium on Earth's Gravitational Field and Secular Variations in Position, pp. 454-462.

Sjöberg, L.E. (1983). Unbiased estimation of variance-covariance components in condition adjustments with unknowns – A MINQUE approach. *ZfV*, Vol. 108, pp. 382-387.

Slater, L.E. and G.R. Huggett (1976). A multiwavelength distance-measuring instrument for geophysical experiments. *Journal of Geophysical Research*, Vol. 81, pp.6299-6304.

Steeves, R.R. (1981). A statistical test for significant peaks in the least squares spectrum. In Collected Papers, Geodetic Survey, Surveys and Mapping Branch, Energy, Mines and Resources Canada, Ottawa, Ontario.

Stein, R.S. and W. Thatcher (1982). Field test for refraction in leveling. *Eos Transactions, American Geophysical Union*, Vol. 63, No. 45, p. 1106.

Stein, R.S., C.T. Whalen, S.R. Holdahl, W.E. Strange and W. Thatcher. Saugus-Palmdale, California, field test for refraction in historical levelling surveys. *Journal of Geophysical Research*, Vol. 91, No. B9, pp. 9056-90??.

Taylor, J. and S. Hamilton (1972). Some tests of the Vaníček method of spectral analysis. *Astrophysics and Space Science*, Vol. 17, pp. 357-367.

Vaníček, P. (1969a). Approximate spectral analysis by least-squares fit. *Astrophysics and Space Science*, Vol. 4, pp. 387-391.

Vaníček, P. (1969b). New Analysis of the Earth-Pole Wobble. *Studia Geophysica et Geodaetica*, Vol. 13, pp. 225-230.

Vaníček, P. (1971). Further development and properties of the spectral analysis by least-squares. *Astrophysics and Space Science*, Vol. 12, pp. 10-33.

Vaníček, P. (1979). Tensor structure and the least squares. *Bulletin Geodesique*, Vol. 53, pp. 221-225.

Vaníček, P. (1986). Diagrammatic approach to adjustment calculus. *Geodezja*, Vol. 86, No. 999, pp. 29-39.

Vaníček, P. and M.R. Craymer (1983a). Autocorrelation functions as a diagnostic tool in levelling. In H. Pelzer and W. Niemeier (Editors), *Precise Levelling: Contributions to the Workshop on Precise Levelling*, Dümmler Verlag, Bonn, pp. 327-341.

Vaníček, P. and M.R. Craymer (1983b). Autocorrelation functions in the search for systematic errors in levelling. *Manuscripta Geodaetica*, Vol. 8, pp. 321-341.

Vaníček, P., G.H. Carrera and M.R. Craymer (1985). Corrections for systematic errors in the Canadian levelling network. Technical Report No. 10, Survey Science, University of Toronto (Erindale Campus), Mississauga, Ontario.

Vaníček, P. and E.W. Grafarend (1980). On the weight estimation in levelling. NOAA Technical Report NOS 86 NGS 17, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Survey, Rockville, MD.

Wassef, A.M. (1959). Note on the application of mathematical statistics to the analysis of levelling errors. *Bulletin Geodesique*, Vol. 52, pp. 19-25.

Wassef, A.M. (1974). On the search for reliable criteria of the accuracy of precise levelling based on statistical considerations of the discrepancies. *Bulletin Geodesique*, Vol. 112, pp. 149-163.

Wassef, A.M. (1976). Propagation of errors in precise levelling and its bearing on the assessment of recent crustal movements. *Bulletin Geodesique*, Vol. 53, pp. 53-60.

Wells, D.E., N. Beck, D. Delikaraoglou, A. Kleusberg, E.J. Krakiwsky, G. Lachapelle, R.B. Langley, M. Nakiboglu, K.-P. Schwarz, J.M. Tranquilla and P. Vaníček (1987). Guide to GPS Positioning. Canadian GPS Associates, Fredericton, N.B.

Wells, D.E., P. Vaníček and S. Pagiatakis (1985). Least squares spectral analysis revisited. Technical Report No. 84, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B.

Whalen, C.T. and W.E. Strange (1981). The 1981 Saugus to Palmdale, California leveling refraction test. NOAA Technical Report NOS 98 NGS 27, National Geodetic Information Center, Rockville, Md.

Zhang, J., Y. Bock, H. Johnson, P. Fang, J. Genrich, S. Williams, S. Wdowinski and J. Behr (1997). Southern California Permanent GPS Geodetic Array: Error Analysis of Daily Position Estimates and Site Velocities. *Journal of Geophysical Research,* in press.