LEAST-SQUARES WAVELET ANALYSIS AND ITS APPLICATIONS IN GEODESY AND GEOPHYSICS

EBRAHIM GHADERPOUR

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN EARTH AND SPACE SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

JULY 2018

Abstract

The Least-Squares Spectral Analysis (LSSA) is a robust method of analyzing unequally spaced and non-stationary data/time series. Although this method takes into account the correlation among the sinusoidal basis functions of irregularly spaced series, its spectrum still shows spectral leakage: power/energy leaks from one spectral peak into another. An iterative method called AntiLeakage Least-Squares Spectral Analysis (ALLSSA) is developed to attenuate the spectral leakages in the spectrum and consequently is used to regularize data series. In this study, the ALLSSA is applied to regularize and attenuate random noise in seismic data down to a certain desired level. The ALLSSA is subsequently extended to multichannel, heterogeneous and coarsely sampled seismic and related gradient measurements intended for geophysical exploration applications that require regularized (equally spaced) data free from aliasing effects.

A new and robust method of analyzing unequally spaced and non-stationary time/data series is rigorously developed. This method, namely, the Least-Squares Wavelet Analysis (LSWA), is a natural extension of the LSSA that decomposes a time series into the time-frequency domain and obtains its spectrogram. It is shown through many synthetic and experimental time/data series that the LSWA supersedes all state-of-the-art spectral analyses methods currently available, without making any assumptions about or preprocessing (editing) the time series, or even applying any empirical methods that aim to adapt a time series to the analysis method. The LSWA can analyze any non-stationary and unequally spaced time series with components of low or high amplitude and frequency variability over time, including datum shifts (offsets), trends, and constituents of known forms, and by taking into account the covariance matrix associated with the time series. The stochastic confidence level surface for the spectrogram is rigorously derived that identifies statistically significant peaks in the spectrogram at a certain confidence level; this supersedes the empirical cone of influence used in the most popular continuous wavelet transform.

All current state-of-the-art cross-wavelet transforms and wavelet coherence analyses methods impose many stringent constraints on the properties of the time series under investigation, requiring, more often than not, preprocessing of the raw measurements that may distort their content. These methods cannot generally be used to analyze unequally spaced and non-stationary time series or even two equally spaced time series of different sampling rates, with trends and/or datum shifts, and with associated covariance matrices. To overcome the stringent requirements of these methods, a new method is developed, namely,

the Least-Squares Cross-Wavelet Analysis (LSCWA), along with its statistical distribution that requires no assumptions on the series under investigation. Numerous synthetic and geoscience examples establish the LSCWA as the method of methods for rigorous coherence analysis of any experimental series.

This dissertation is dedicated to my mother, Maryam Ayatipour, who has supported me throughout this whole journey and to the memory of my father, Masoud Ghaderpour. And to my wonderful and beloved wife, Tijana Vujadinovic, who has been with me all these years and has made them the best years of my life. And to all my friends, thank you for your friendship and encouragement.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Spiros D. Pagiatakis for his continuous support, patience, enthusiasm, and immense knowledge. He was generous and knowledgeable and always available for discussion when I needed assistance. I am very grateful to have him as my PhD supervisor.

I would like to thank my course instructors and committee members, Drs. Sunil Bisnath and Jianguo Wang, for their support, knowledge, guidance, and excellent comments that significantly improved the presentation of this dissertation.

Dr. Amir Asif was extremely generous in spending long late hours and making many essential comments that greatly improved the quality of this dissertation, and I appreciate Drs. Mathew Kyan and Costas Armenakis for their very useful and excellent comments on this dissertation.

Special thanks to my postdoctoral supervisors, Drs. Wenyuan Liao and Michael P. Lamoureux, who hired me as a Postdoctoral Scholar at the University of Calgary and supported me to complete part of this dissertation.

I would like to thank Dr. Petr Vaníček for his time in reviewing the essential parts of this work and providing very detailed and comprehensive comments. I would also like to thank Drs. Elmas Sinem Ince and Mahmoud Abd-El Gelil for useful discussions on the applications of the least-squares wavelet analysis.

My research was supported by NSERC (Natural Sciences and Engineering Research Council) grants to my supervisor and awards from the Faculty of Graduate Studies at York University as well as PIMS (Pacific Institute for the Mathematical Sciences) and the postdoctoral program at the Department of Mathematics and Statistics at the University of Calgary.

Table of Contents

List of Tables

List of Figures

Preface

This dissertation is mainly paper-based including extensive theories and applications.

- Chapters 3 and 4 are published in *Geophysics* (Ghaderpour et al., 2018b). Parts of these chapters were presented at the *Consortium for Research in Elastic Wave Exploration Seismology (CREWES)*, Calgary, Canada (March 2017). They were also presented in *GeoConvention*, Calgary, Canada (May 2017).

- Chapter 5 is under review in *Journal of Applied Geophysics* (Ghaderpour, 2018). This chapter was presented at the *International Conference on Diversification Trends in Engineering Technology and Applied sciences (DTETA)*, Tokyo, Japan (March 2018).

- Chapters 6 is published in *Mathematical Geosciences* (Ghaderpour and Pagiatakis, 2017). Some parts of this chapter were presented at the *Canadian Geophysical Union (CGU) Annual Meeting*, Montreal, Canada (May 2015) and published as a short paper in *CGU*, selected as the best student paper in Geodesy. They were also presented at the *International Union of Geodesy and Geophysics (IUGG)* in Prague, Czech Republic (June 2015).

- Chapters 7 is published in *Journal of Geodesy* (Ghaderpour et al., 2018a). Parts of this chapter were presented at the *International Symposium on the Geodynamics and Earth Tide*, Trieste, Italy (June 2016). They were also presented at the *European Geosciences Union (EGU)*, Vienna, Austria (April 2018).

- The author acknowledges with thanks Drs. Spiros D. Pagiatakis, Wenyuan Liao, Michael P. Lamoureux, and E. Sinem Ince for agreeing to have the papers mentioned above as parts of this dissertation.

Abbreviations

ALFT	AntiLeakage Fourier Transform
ALLSSA	AntiLeakage Least-Squares Spectral Analysis
ALLSS	AntiLeakage Least-Squares Spectrum
ASFT	Arbitrarily Sampled Fourier Transform
CHAMP	CHAllenging Mini-satellite Payload
CLSSA	Constrained Least-Squares Spectral Analysis
CWT	Continuous Wavelet Transform
XWT	Cross-Wavelet Transform
DFT	Discrete Fourier Transform
DTFT	Discrete-Time Fourier Transform
EGG	Electrostatic Gravity Gradiometer
ESD	Energy Spectral Density
EIC	Equivalent Ionospheric Currents
FFT	Fast Fourier Transform
GGT	Gravitational Gradient Tensor
GOCE	Gravity field and steady-state Ocean Circulation Explorer
GRACE	Gravity Recovery And Climate Experiment
HHT	Hilbert-Huang Transform
IMAP	Interpolation by MAtching Pursuit

LSCSA Least-Squares Cross-Spectral Analysis

LSCS Least-Squares Cross-Spectrum

LSCWA Least-Squares Cross-Wavelet Analysis

LSCWS Least-Squares Cross-Wavelet Spectrogram

LSSA Least-Squares Spectral Analysis

LSS Least-Squares Spectrum

LSWA Least-Squares Wavelet Analysis

LSWS Least-Squares Wavelet Spectrogram

MALLSSA Multichannel AntiLeakage Least-Squares Spectral Analysis

MALLSS Multichannel AntiLeakage Least-Squares Spectrum

MIMAP Multichannel Interpolation by MAtching Pursuit

MWNI Minimum Weighted Norm Interpolation

OMP Orthogonal Matching Pursuit

PDF Probability Distribution Function

PEF Prediction Error Filters

POCS Projection Onto Convex Sets

PSD Power Spectral Density

SEC Spherical Elementary Currents

STFT Short-Time Fourier Transform

UT Universal Time

VLBI Very Long Baseline Interferometry

WWA Weighted Wavelet Amplitude

WWZ Weighted Wavelet Z-transform

Symbols

c/a	cycles per annum
c/h	cycles per hour
c/km	cycles per kilometre
c/mo	cycles per month
Hz	cycles per second (Hertz)
cv	critical values
α	significance level
dB	decibel
f-k	frequency-wavenumber
J	Jacobian
mE	milli-Eotvos
Pa	Pascal
$\mathrm{W/m^2}$	Watt per metre square
σ	The standard deviation
i	The imaginary number (when used as part of the exponent in Fourier methods)
$\mathcal{N}(\mu, \sigma^2)$	The normal distribution with mean μ and variance σ^2
$\Gamma(x)$	The Gamma function
χ^2_r	The chi-squared distribution with r degrees of freedom
β_{p_1,p_2}	The beta distribution with parameters p_1 and p_2

ω_k	The cyclic frequency or cyclic wavenumber
$\mathbf{K}$	The preselected set of wavenumbers for the ALFT and ASFT
$\mathbf{\Omega}$	The preselected set of wavenumbers or frequencies for the least-squares methods
E(X)	The expected value of (complex) random variable X
$\mathbf{P}$	The weight matrix for entire data/time series
$\mathbf{C_f}$	The covariance matrix of data/time series $\mathbf{f}$
$\mathbf{P_y}$	The principal submatrix of the weight matrix corresponding to segment $\mathbf{y}$
$\underline{\mathbf{\Phi}}$	The design matrix containing the constituents of known forms such as trends and sinusoidal base function of known cyclic frequencies
$\mathbf{\Phi}$ or $\mathbf{\Phi_k}$	The design matrix containing the sinusoids of cyclic frequency/wavenumber ω_k
$\overline{\mathbf{\Phi}}$	The design matrix containing the constituents of known forms and the sinusoidal base functions of cyclic frequency ω_k
$\mathbf{N}$	The matrix of normal equations
M	The (average) number of data points per unit time (a window size parameter)
L_0	The fixed number of data points within the windows in the LSWA and LSCWA
L_1	The number of cycles per unit time within the windows in the LSWA and LSCWA
$R(j,k)$	The window size in the LSWA and LSCWA at time t_j and cyclic frequency ω_k
ℓ	The window length in the LSWA
q	The number of constituents of known forms
N_{eff}	The effective number of data points for the WWZ
$\hat{\mathbf{c}}$	The column vector containing the estimated amplitudes and phases
$s(t_j,\omega_k)$	The LSWS value at time t_j and cyclic frequency ω_k
$s(\omega_k)$	The LSS value at cyclic frequency ω_k
X_s	The LSCS or LSCWS
ζ	The stochastic surface of the LSWS

$\approx$	Approximation
$\sim$	Follows: a random variable follows a distribution
$\simeq$	Similar or equal to

1 Introduction

1.1 Motivation

Creatures intuitively process signals in various ways. For instance, human eyes can analyze various visible colours (may be described by electromagnetic waves), or human ears can identify various sounds. In speech communication, a listener may have a challenging time understanding the main message of a speaker due to environmental noise contaminating the speech signals. The brain intuitively concentrates and focuses (zooms) more to extract the important message from noisy environments. The more power (higher amplitude) the constituent of interest has, the better it can be heard by ears and vise versa. If the noise is dominant (very low signal-to-noise ratio), then the important message may be buried and not be heard.

A periodic oscillation can be described by a smooth sinusoidal function. It may be a function of time or distance. Consider the sinusoidal waves $f_1(t) = A\sin(2\pi\omega t + \theta)$ and $f_2(t) = A\cos(2\pi\omega t + \theta)$, where f_1 and f_2 are functions of time t. The parameters A, ω and θ are called the amplitude, cyclic frequency[1] (number of cycles per unit time) and phase of the sinusoid, respectively. These functions are periodic with period $T = 1/\omega$ (Mallat, 1999; Brown and Hwang, 2012). An aperiodic signal is a signal that does not repeat its pattern over a period. If time t is replaced by distance x, then ω is called wavenumber (number of cycles per unit distance). The sampling rate is the number of samples measured per unit time (unit distance).

In many areas of research, such as geodesy, geophysics, geodynamics, astronomy and speech communications, researchers deal with series of data points measured over time or distance to study the periodicity and/or power of certain constituents. For example, in a speech record, one may be interested in certain constituents (signals) in the record, such as voice of a person obscured by other simultaneously recorded constituents, noise from musical instruments and the environment. It is often difficult to study these constituents by direct analysis of the data series; however, decomposing the data into another domain (frequency) results in detecting the constituents of interest much easier.

Furthermore, the physical series of data are usually a combination of various sinusoidal waves, such as light emitted from a star, sound waves, ocean waves, and thus certain phenomena may be studied more

[1] In literature, symbol ω is often used for angular frequency, and f is used for cyclic frequency. However, in this dissertation, ω represents cyclic frequency or just frequency, and f is used for function or time series values.

accurately by decomposing the data into a new domain (e.g., frequency or time-frequency domains). For instance, the decomposition of a series of data points into sinusoidal waves of various frequencies may be imagined as dispersing sunlight to waves of different frequencies (different colours - like a rainbow) when it passes through a triangular prism. In other words, a triangular prism separates the sunlight to different wavelengths (the longer wavelength is red, and the shorter wavelength is blue). The word "spectrum" was first referred to the range of colours observed when white light was dispersed through a prism. Objects, such as eyes, ears and prisms may be viewed as decomposition operators that are somewhat similar to the decomposition operators introduced in this dissertation.

A "time series" is a sequence of data points measured at discrete time intervals that may not be equally spaced or it may contain data gaps, and so the time series is unequally spaced (irregularly or unevenly spaced). In certain experiments, measurements may also have uncertainties (variances) introduced by random noise and thus may also be unequally weighted. Similarly, a "data series" is a sequence of data points sampled along spatial directions, such as in seismic data applications.

In certain areas, such as seismology, researchers deal with data sets that are unequally (irregularly) sampled and want to reconstruct the data sets on any desired equally spaced series (or grids) because it is a requirement for other purposes. In Chapters 3, 4, and 5, I shall propose two new methods that can aid researchers in these areas to achieve this goal. Researchers are usually interested in studying certain periodic or aperiodic signals in time/data series that are usually obscured by noise. In Chapter 6, I shall introduce a new method that can help researchers to discover such signals in any time series. In many applications, researchers want to investigate the correlation or coherency between two time series, so that they can search for common constituents at different time segments. For instance, in geodesy, researchers may want to investigate the effect of atmospheric temperature on deformation of geodetic antennas; or in finance, a banker may like to investigate the coherency between two financial time series containing daily interest rates; or in medical sciences, physicians may want to study the effect of heart activities on brain function by analyzing their corresponding time series. In Chapter 7, I shall introduce a new method that can help researchers to rigorously investigate the coherency and phase differences between the components of two or more time/data series.

1.2 Fundamental Definitions and Concepts

Aliasing is an effect that causes different signals to be indistinguishable when sampled. In seismic data applications, aliasing in spatially sampled data series is called spatial aliasing. To understand the aliasing better, assume that a sinusoid is sampled at equally spaced intervals. If another sinusoid has the same amplitude and phase as the original and its frequency/wavenumber differs from the first one by a multiple of the sampling rate, then these two sinusoids will have exactly the same samples over the equally spaced

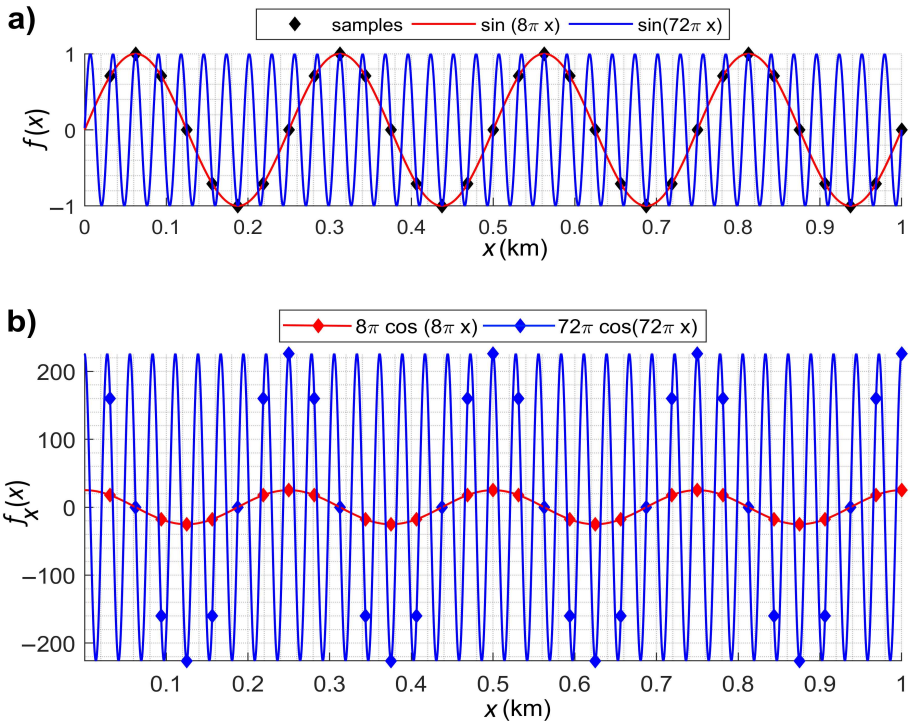

Figure 1.1: (a) Two aliased sinusoids with identical values at sample positions (32 samples/km), and (b) their gradients with different values at sample positions that can be used to identify the true signal.

intervals (Vassallo et al., 2010a). In other words, signals with different frequencies or wavenumbers may have the same samples over the equally spaced intervals. Aliasing occurs whenever a signal is sampled at a rate below its Nyquist rate, and so when decomposing a time/data series to the frequency/wavenumber domain, two or more different sinusoids identically fit the same set of samples (Craymer, 1998). Figure 1.1a shows an example of two aliased sinusoids that are indistinguishable from the samples (black diamonds). The two sinusoids are a sine wave of wavenumber 4 cycles per kilometre (c/km), shown in red, and a sine wave of wavenumber 36 c/km, shown in blue, that differs from the first sine wave by the sampling rate (32 samples/km).

For interpolation of time/data series, an incorrect selection of a signal may cause significant errors in the final interpolation result. Mitigating aliasing effect is crucial in many areas of research, such as interpolation of seismic data that are coarsely sampled (Vassallo et al., 2010a). Additional information or assumption must be provided to overcome the aliasing effect and identify the true signals. For instance, the gradients in Figure 1.1b (the derivative of the sinusoids) have different values at sample positions and can be used to identify the true sinusoid for interpolation purposes (Vassallo et al., 2010a). In down sampling (decimation), depending on the purpose of specific experiments, researchers may intentionally define a variable sampling rate to avoid aliasing (Shapiro and Silverman, 1960; Pagiatakis et al., 2007b).

If all statistical properties of a time series (i.e., the mean, variance, all higher order statistical moments and auto-correlation function) do not change in time, then the time series is called stationary. A time series is called non-stationary if it violates at least one of the assumptions of stationarity (Brown and Hwang, 2012). A non-stationary time series may contain systematic noise, such as trends (linear or exponential or others) and/or datum shifts (offsets) indicating that its mean value is not constant in time. The second moments (central and mixed moments) of the time series values form a symmetric matrix called the covariance matrix (Vaníček and Krakiwsky, 1986). In certain fields, such as geodesy, geophysics, and astronomy, a time series is usually associated with a covariance matrix which means that the time series is stationary or non-stationary in its second statistical moments. In geodynamics applications, seismic noise may contaminate the time series of interest or certain components of the time series may exhibit variable frequency, such as linear, quadratic, exponential or hyperbolic chirps. For more details, the reader is referred to Mallat (1999); Kay and Marple (1981); Brown and Hwang (2012).

To solve problems regarding heat transfer and vibrations, Fourier (1822) introduced the Fourier series and showed that any (periodic) function could be written as an infinite sum of harmonics (sinusoids). The Fourier transform and Fourier's law are also named after him. The Fourier transform basically decomposes a time series (usually aperiodic) into the frequency domain using sinusoidal basis functions (Mallat, 1999). The amplitudes of the sinusoids for a set of frequencies generate a spectrum similar to the light spectrum. The Fourier transform is defined for equally spaced time series without any data gaps (Appendix A).

Time series analyses have almost exclusively been performed using the classical Fourier method and its modifications, often incorrectly postulating that the time series under consideration is stationary, resulting in erratic results. It is not unusual to see that many researchers even attempt to modify the original raw measurements by interpolation, editing or offset removal, often using empirical and questionable approaches, simply to satisfy the stringent requirements of the Fourier transform (Kay and Marple, 1981; Stein and Shakarchi, 2003).

The Least-Squares Spectral Analysis (LSSA) was introduced by Vaníček (1969, 1971) for the purpose of analyzing non-stationary (in the first and second statistical moments) and unequally spaced time series. Conceptually, this method estimates a spectrum based on the least-squares fit of sinusoids to the entire time series similar to Fourier transform, but it is also defined for unequally spaced time series. The statistical properties of the least-squares spectrum were discussed later by Lomb (1976); Craymer (1998); Pagiatakis (1999). When analyzing unequally spaced data (time) series, the least-squares spectrum shows spectral leakages that is leaking power/energy from one spectral peak into another. These leakages are caused by non-orthogonality of the sinusoidal basis functions over unequally spaced series because the LSSA examines wavenumbers or frequencies one at a time (out-of-context), ignoring the correlations between the sinusoidal basis functions of different wavenumbers or frequencies (Craymer, 1998). For interpolation purposes, these

leakages cause error in the final interpolation results.

Furthermore, in the LSSA, although trends, datum shifts and the covariance matrix associated with a time series are parameterized and consequently estimated simultaneously (in-context) with the calculation of the spectrum (Wells et al., 1985), it cannot generally be used for time series whose components change amplitude (e.g., transient signals) and frequency (e.g., doppler-shifted signals) over time because the LSSA decomposes a time series into the frequency domain rather than into the time-frequency domain (similar to Fourier analysis). Craymer (1998) compared the LSSA to Fourier analysis (discrete) and showed that the Fourier analysis is a special case of the LSSA when the time series is strictly stationary, equally spaced, equally weighted, and with no gaps.

In order to measure the frequency variation of sounds, Gabor (1946) introduced the Short-Time Fourier Transform (STFT) to analyze piece-wise stationary and equally spaced time series. He endeavoured to obtain the spectrum of a time series, in short-time intervals (or within a window which translates through the entire time series) rather than considering the entire time series at once (Mallat, 1999). In the STFT, the window length is fixed during translation in time, which is not appropriate for analyzing time series with components of high frequency variability over time, presenting sudden changes in amplitude (jumps). Decomposing a non-stationary time series into time-frequency domain has many advantages over the decomposition into only frequency domain. These advantages include searching for the short duration hidden components in the time series or detecting constituents with variability of amplitude and/or frequency over time.

A wavelet is a wave-like function whose values start from zero and end at zero after a number of oscillations. A wavelet can be visualized as a short oscillation like one recorded by a seismograph or heart monitor. Wavelets are useful functions in signal and image processing. The Continuous Wavelet Transform (CWT) is a well-established method originated by Morlet (1983); Grossmann et al. (1985, 1986). The CWT computes a "scalogram" based on the dilations and translations of a (mother) wavelet. The scalogram is a three-dimensional analogue of the wavelet transform representation in which the *x-axis, y-axis*, and *z-axis* represent time, scale, and absolute value of Fourier coefficients at particular times, respectively. One may also convert the "scalogram" to a "spectrogram" using different approaches that each has its own advantages and disadvantages (Hlawatsch and Boudreaux-Bartels, 1992; Sinha et al., 2005). The spectrogram is slightly different from the scalogram in that its *y-axis* values represent frequencies rather than scales, and thus is more appropriate to study the periodicity of various components in the time series.

The CWT spectrogram has high-frequency resolution at low frequencies and high-time resolution at high frequencies and vice versa (Mallat, 1999; Sinha et al., 2005). There are many different types of wavelets introduced and used for particular applications (Mallat, 1999). In the CWT, if one uses an inappropriate wavelet, then the CWT spectrogram may give misleading and physically meaningless results (Qian, 2002). The CWT and its discrete version are not defined for unequally spaced time series, and so it cannot be used

to analyze unequally spaced and/or unequally weighted time series that may also have trends. Foster (1996) proposed a new method of time-frequency analysis, namely, the Weighted Wavelet Z-transform (WWZ), to analyze astronomical time series that are often unequally spaced and non-stationary. The WWZ basically obtains a signal-to-noise ratio spectrogram based on the least-squares principle and has several advantages over the CWT, yet it has its own shortcomings. The WWZ is described in detail in Chapter 2 and compared to the least-squares wavelet analysis developed in Chapter 6.

An underlying Probability Distribution Function (PDF) can be obtained for the spectrum or spectrogram of a time series by assuming that the time series has been derived from a population of random variables[2] following the multi-dimensional normal distribution. In other words, one can view each data point of a time series as a random variable following the normal distribution, and so the entire time series follows an n-dimensional normal distribution associated with a covariance matrix, showing the statistical dependency between the data points. It was shown that the least-squares spectrum (Pagiatakis, 1999) follows the beta distribution, and similarly the spectrogram follows the beta distribution (Ghaderpour and Pagiatakis, 2017). Using this distribution, one can identify the stochastic significance of the spectral peaks in the spectrum or spectrogram at certain confidence level. The critical value at certain confidence level is a straight horizontal line to frequency axis of the spectrum, and it is a 2D surface (stochastic surface) in the spectrogram in which its values increase when frequency increases.

In physics and signal processing, power is energy per unit time. In other words, energy is the time integral of power. If the time integral (or summation in discrete case) over the square of the absolute value of the signal values is finite, the signal is an energy signal, where for real signals (not complex), the absolute value can be omitted. The average power of the signal is defined as this integral (summation) divided by the time interval (from minus infinity to infinity). A power signal has infinite energy but finite average power. There are three types of signals: energy signals, power signals, or none of the two. A signal cannot be both energy and power signal. A signal that decays exponentially or a transient signal has finite energy, so it is an energy signal. The power of an energy signal is zero, because of dividing finite energy by infinite time (or length). On the other hand, a power signal is not limited in time. It always exists from beginning to end, and it never ends.

A stationary sine wave with infinite length and constant amplitude is a power signal. Since the energy of a power signal is infinite, it has no meaning. Therefore, one may use power for power signal, because the power of such signal is finite. The concept of energy and power sometimes is used loosely in the literature. A time series may contain both power and energy signals. Depending on the purpose of study, one may compute a power or energy spectral density (spectrum or spectrogram). In seismic signal processing, the

[2]A random variable is a variable that can take multiple different values, each with an associated probability. In practice, random variables are usually measurements.

seismic data contain energy signals and geophysicists usually compute an energy spectrum. In geodesy, geodesists are usually interested in studying the periodicities of certain power signals and usually compute a power spectrum.

In this dissertation, interpolation/extrapolation means filling missing data or gaps in a data series and is required in seismology for various purposes. Regularization is a process of introducing additional information in order to solve an ill-posed problem or to prevent overfitting. This process usually uses the spectrum of an equally/unequally spaced data series in an inverse mode to produce an equally spaced data series that will have the same spectrum as the original data series but with attenuated random noise. In other words, regularization places a data series onto any desired equally spaced series. The verb "regularize" means to reconstruct a data series on any desired equally spaced series.

To summarize this section, I should highlight the fact that the Fourier transform, STFT, and CWT have been extensively studied and used by researchers dealing with equally spaced time series (the state-of-the-art methods). However, in many applications, time/data series are not usually equally spaced and present trends and/or datum shifts. They are often non-stationary and have uncertainties. Many researchers propose Fourier based methods to analyze such time/data series that arise many questions and come up with several shortcomings and assumptions at the end. The Lomb method (Lomb, 1976), LSSA (Vaníček, 1969), and WWZ (Foster, 1996) are other robust methods considered the state-of-the-art methods for computing the spectrum/spectrogram of unequally spaced time/data series, yet they have some shortcomings (considerably less than the Fourier based methods) that I shall show analytically and by plethora of comparative examples.

1.3 Main Contributions

The main contributions of this dissertation are:

- Development of robust methods of regularization using the least-squares spectral analysis for data series that are irregularly spaced or coarsely sampled. Two methods, namely, the Antileakage Least-Squares Spectral Analysis (ALLSSA) and Multichannel Antileakage Least-Squares Spectral Analysis (MALLSSA), supersede the current methods of regularization in that they estimate the signal wavenumbers and amplitudes more accurately and consequently can regularize data series better and faster.
- Development of a robust method of analyzing any type of time series (stationary, nonstationary, equally spaced, unequally spaced, equally weighted, unequally weighted). This method, namely, the Least-Squares Wavelet Analysis (LSWA), supersedes the state-of-the-art methods of analyses of time series by setting no constraints on the type and characteristics of the series and can rigorously analyze any time series by decomposing it into time-frequency domain to obtain a (normalized) spectrogram, considering trends and/or datum shifts, and covariance matrix associated with the time series.

➘ Development of a robust method of analyzing any type of two or more time series together (possessing the same or different sampling rates, equally or unequally spaced, equally or unequally weighted, stationary or non-stationary). This method, namely, the Least-Squares Cross-Wavelet Analysis (LSCWA), obtains a (normalized) cross-spectrogram and can identify the statistically significant peaks in the cross-spectrogram by rigorous statistical testing.

With a plethora of comparative examples, I shall show that the ALLSSA, MALLSSA, LSWA, and LSCWA provide more accurate and more reliable results compared to the state-of-the-art methods. Note that in this dissertation, the ALLSSA and MALLSSA are proposed mainly for regularization purposes; whereas, the LSWA and LSCWA are proposed to study the periodicity of certain components and hidden signatures in time/data series. The ALLSSA and MALLSSA are based on the LSSA that use a refined spectrum to regularize data series, and the LSWA and LSCWA are extensions of the LSSA that decompose a time series to time-frequency domain rather than frequency domain. The LSWA spectrogram may also be used for regularization purposes that is subject to future work.

1.4 Outline

In this dissertation, several state-of-the-art methods of analyzing time series (data series) are reviewed in detail. Their advantages and shortcomings are studied, and new methods are proposed that can be applied to analyze any time series (data series) for various purposes without any need for editing. The outline of this dissertation is as follows.

In Chapter 2, many traditional and state-of-the-art methods of data series analysis are reviewed, and their advantages and shortcomings are discussed as well as the need for more sophisticated methods is emphasized, without the requirement of editing the data sets before analysis.

In Chapter 3, an antileakage method is proposed based on the LSSA which mitigates the spectral leakages in the spectrum significantly. Therefore, the estimated amplitudes and wavenumbers can be used to regularize data series, in certain applications. Two Fourier-based methods, namely, the AntiLeakage Fourier Transform (ALFT) (Xu et al., 2005) and Arbitrarily Sampled Fourier Transform (ASFT) (Guo et al., 2015) are revisited and compared to this method called the AntiLeakage Least-Squares Spectral Analysis (ALLSSA). Özbek et al. (2009) proposed an iterative algorithm called Interpolation by MAtching Pursuit (IMAP) to regularize irregularly sampled seismic data. The IMAP improves the original Lomb spectrum (least-squares spectrum) result (Lomb, 1976) and is compared to the ALLSSA. A synthetic data series example is given with its regularization results using the ALFT, ASFT, IMAP, and ALLSSA.

In Chapter 4, the ALLSSA is applied to a synthetic seismic data example to show its advantages in regularization and random noise attenuation. Since seismic data are often non-stationary and contain many nonlinear events (Crawley, 2000), a few windowing techniques with fixed window length are used to make

the ALLSSA adaptable to non-stationary data series. The ALLSSA is applied to regularize a marine 2D seismic data set, and its results are compared to the ALFT and nonlinear shaping regularization.

In Chapter 5, the Multichannel Interpolation by MAtching Pursuit (MIMAP), the generalization of the IMAP, is revisited. The MIMAP is conceived to interpolate a few irregularly spaced multicomponent measurements (Vassallo et al., 2010a). Then a new method, namely, Multichannel Antileakage Least-Squares Spectral Analysis (MALLSSA), is introduced that uses the spatial gradients of a data series (like the MIMAP) by extending the model of the ALLSSA to aid interpolation of the data series presenting severe aliasing. The MALLSSA is compared to MIMAP, and its outstanding performance is shown with synthetic and field seismic data examples.

In Chapter 6, a new method is proposed to analyze non-stationary time series in the first and second statistical moments that are also unequally spaced and comprise constituents of low/high amplitude and frequency variability over time. This method, namely, the Least-Squares Wavelet Analysis (LSWA), is a natural extension of the LSSA and can analyze rigorously any type of time series without any modifications or editing. The practical implementation steps of the LSWA for equally and unequally spaced time series are also demonstrated. The stochastic surfaces (confidence level surfaces) are obtained above which the peaks in the least-squares wavelet spectrograms are statistically significant at certain confidence level (usually 95% or 99%). Several advantages of the LSWA over the CWT are shown on two synthetic equally spaced time series. The WWZ is also compared to the LSWA on two synthetic examples, and as an application of the LSWA, a Very Long Baseline Interferometry (VLBI) baseline length time series is analyzed using both the LSWA and WWZ.

In Chapter 7, the Least-Squares Cross-Wavelet Analysis (LSCWA) is introduced, and the stochastic surfaces (confidence level surfaces) for the Least-Squares Cross-Wavelet Spectrogram (LSCWS) are derived, showing statistically significant peaks at a certain confidence level. A special case of the LSCWA, the Least-Squares Cross-Spectral Analysis (LSCSA), along with its Least-Squares Cross-Spectrum (LSCS) are studied as well as the phase differences in the LSCWS and LSCS. These methods are appropriate methods of analyzing two (or more) unequally spaced series together. The performances of the LSCWA and LSCSA are investigated on two synthetic time series. These methods are also applied to Gravity field and steady-state Ocean Circulation Explorer (GOCE) electrostatic gradiometer measurement disturbances to determine their coherence with the Poynting energy flux in the ionosphere along two successive ascending satellite tracks during a magnetic storm (Ince and Pagiatakis, 2016) to show their superior performances on the time series of different sampling rates. In another application, the coherency between the VLBI baseline length series from Westford to Wettzell with the temperature series (recorded since 1984) is investigated to show the effect of the temperature variation on the baseline length.

In Chapter 8, the proposed methods (ALLSSA, MALLSSA, LSWA, LSCWA) along with their advantages and limitations are briefly reviewed, and a table with their applications is provided. The future directions and next steps for processing time/data series and other applications are also discussed.

2 Preliminaries and Literature Review

In this chapter, several methods of analyzing time series (data series) are briefly reviewed in short sections, and their advantages and weaknesses are discussed. The review has the purpose of focusing on the most popular and state-of-the-art spectral analyses methods applied in almost every field of science, using spectral methods that require a number of assumptions and approximations. In the following chapters, a series of optimal statistical and mathematical approaches as well as a plethora of comparative synthetic and real time series analyses are used to show the robustness and effectiveness of the new proposed methods that do not require any assumptions or editing of the original data sets.

2.1 Fourier Transform

If f is an absolutely integrable function[3] on $\mathbb{R}$, then the Fourier transform of f for cyclic frequency ω, denoted $\hat{f}(\omega)$ is defined as

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i \omega t} \, \mathrm{d}t, \tag{2.1}$$

where "e" is the Euler's number, and "i" stands for the imaginary number[4]. The spectrum of f is defined as $s(\omega) = |\hat{f}(\omega)|^2$ which measures how much oscillations of cyclic frequency ω contribute to f, where for a complex number z, $|z|$ denotes its absolute value[5] (Abs or abs in short) (Mallat, 1999; Stein and Shakarchi, 2003; Brown and Hwang, 2012). If f is also a square-integrable function[6] on $\mathbb{R}$ (e.g., an energy signal), then the Parseval's theorem (Mallat, 1999, Chapter 2) states that

$$\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t = \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \, \mathrm{d}\omega. \tag{2.2}$$

[3] An absolutely integrable function f on $\mathbb{R}$ is a real- or complex-valued function such that $\int_{-\infty}^{\infty} |f(t)| \, \mathrm{d}t < \infty$.

[4] The imaginary number i is a solution of $x^2 = -1$.

[5] The absolute value of a complex number $z = a + ib$ is $|z| = \sqrt{a^2 + b^2}$.

[6] A square-integrable function f on $\mathbb{R}$ is a real- or complex-valued function such that $\int_{-\infty}^{\infty} |f(t)|^2 \, \mathrm{d}t < \infty$.

Equation (2.2) implies that the total energy may be determined either by integrating $|f(t)|^2$ over all time or by integrating $|\hat{f}(\omega)|^2$ over all cyclic frequencies. Therefore, $|\hat{f}(\omega)|^2$ is interpreted as energy spectral density (ESD) of f. For inverse of the Fourier transform to exist, function $\hat{f}$ should be absolutely integrable on $\mathbb{R}$ (Mallat, 1999, Chapter 2).

The average power P of f over all time is given by $P = \lim_{\tau\to\infty}(1/\tau)\int_0^\tau |f(t)|^2 \, \mathrm{d}t$. For many signals of interest, the Fourier transform does not formally exist. Therefore, one may use a truncated Fourier transform where the signal is integrated only over a finite interval:

$$\hat{f}(\omega) = \frac{1}{\sqrt{\tau}} \int_0^\tau f(t) e^{-2\pi i \omega t} \, \mathrm{d}t, \tag{2.3}$$

that is the amplitude spectral density. Then the Power Spectral Density (PSD) can be defined as $\lim_{\tau\to\infty} \mathrm{E}\big[|\hat{f}(\omega)|^2\big]$, where $\mathrm{E}[X]$ is the expected value of X (Miller and Childers, 2012, Chapter 10). The PSD describes how power of a signal or time series is distributed over cyclic frequency.

In the Discrete-Time Fourier Transform (DTFT), the integral in Equation (2.1) may be replaced with the summation multiplied by a factor, see Equation (2.4) (Xu et al., 2005). More precisely, suppose that $\mathbf{f} = \big[f(t_1), \ldots, f(t_n)\big]$ is a time series of size n. Define $\Delta t_\ell = \frac{1}{2}(t_{\ell+1} - t_{\ell-1})$ for $1 < \ell < n$. For the two special cases $\ell = 1$ and $\ell = n$, define $\Delta t_1 = t_2 - t_1$ and $\Delta t_n = t_n - t_{n-1}$, respectively. Let $\Delta t = \sum_{\ell=1}^n \Delta t_\ell$. The forward discrete summation of the Fourier integral (after dividing by the summation range) for cyclic frequency ω is defined as

$$\hat{f}(\omega) = \frac{1}{\Delta t} \sum_{\ell=1}^{n} \Delta t_\ell f(t_\ell) e^{-2\pi i \omega t_\ell}. \tag{2.4}$$

When t_ℓ's are equally spaced, Δt_ℓ is independent of ℓ, and $\Delta t_\ell / \Delta t = 1/n$. The Discrete Fourier Transform (DFT) can be seen as the sampled version (in frequency-domain) of the DTFT output because computers can only handle a finite number of values, i.e., in the DFT the frequencies are also discretized. The complexity of computing the DFT is in the order of n^2 floating operations for a time (data) series of n data points. The Fast Fourier Transform (FFT) is an optimized algorithm that computes the DFT (for equally spaced time series) but has an order of $n \log(n)$ floating operations instead of n^2 (Mallat, 1999). The DFT spectrum may be obtained by taking the absolute value of the Fourier coefficients in Equation (2.4).

The Fourier transform is not defined for unequally spaced time series. Even the DFT is not an appropriate method of estimating the signal frequencies and amplitudes of unequally spaced time series (Craymer, 1998; Xu et al., 2005). The Fourier transform and its modifications do not consider trends and/or datum shifts present in time series, and they are not appropriate for analyzing nonstationary time series. Moreover, there is no rigorous statistical testing for assessing the Fourier transform results.

2.2 Some Parametric Statistical Distributions

In this section, some of the most important underlying Probability Distribution Functions (PDFs) are briefly reviewed that will be used in this dissertation. They will play crucial roles in assessing the significance of spectral peaks in the least-squares spectrum and spectrogram. The statistical tests used in this dissertation are parametric, meaning that they assume underlying statistical distributions in the data. All the definitions and formulas in this section are taken from Craig et al. (2013). Before going into introducing the PDFs, it is worth mentioning some basic statistical definitions as follows.

The null hypothesis, denoted by H_0, is a hypothesis that researcher usually tries to reject; whereas, the alternative hypothesis is what the researcher really thinks is the cause of a phenomenon. In hypothesis testing, a critical value is a point on the test distribution that is compared to the test statistic[7] to determine whether to reject the null hypothesis. The significance level α is the probability of rejecting the null hypothesis given that it is true (a type I error). The common values of the significance level are 0.1, 0.05, and 0.01. For instance, $\alpha = 0.05$ implies that the null hypothesis is rejected 5% of the time when it is in fact true. The confidence level is the probability that the value of a parameter falls within a specified range of values.

In the following definitions, uppercase letters, such as X is used for a random variable of continuous type that can take infinitely many values (denoted by lowercase letters, such as x)[8], such that each value is associated with probability $f_X(x)$.

2.2.1 Normal distribution (one dimension and multi-dimensions)

A random variable X of continuous type has a normal distribution, denoted by $\mathcal{N}(\mu, \sigma^2)$, if its PDF has a Gaussian distribution as follows

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \quad -\infty < x < \infty, \tag{2.5}$$

where μ and σ^2 are the mean and variance of random variable X, respectively. Normal distributions are crucial in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known. Many physical quantities, such as random errors in measurements often have distributions that are nearly normal. The underlying PDF of many test statistics can be analytically derived, assuming that the random errors in measurements are normally distributed. Therefore, the theoretical distribution can be compared to the actual distribution of random errors (sample distribution) to make a decision about the validity of the H_0 using a certain level of confidence.

[7] A test statistic is a quantity derived from the sample used in statistical hypothesis testing. It is a function of one or more random variables.

[8] As a side note, $P(X \leq x)$ refers to the probability that random variable X is less than or equal to x (cumulative probability).

In multidimensional case, suppose that $\mathbf{X} = [X_1, \ldots, X_n]^{\mathrm{T}}$, where T denotes the transpose[9], is a random vector of continuous type with mean $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]^{\mathrm{T}}$ and covariance matrix $\mathbf{C_X}$ that is symmetric[10] and positive definite. Note that the covariance matrix is always required to be positive semi-definite (that is the quadratic form $\mathbf{A}^{\mathrm{T}}\mathbf{C_X}\mathbf{A}$ is greater than or equal to zero for all real column vector $\mathbf{A}$ of dimension n). For simplicity, it is assumed that the covariance matrix is positive definite, i.e., the quadratic form is greater than zero that is usually the case, and so its inverse exists. The random vector $\mathbf{X}$ has a multidimensional normal distribution, denoted by $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C_X})$, if its underlying PDF is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C_X}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{C_X^{-1}}(\mathbf{x}-\boldsymbol{\mu})}, \qquad \text{for } \mathbf{x} \text{ in } \mathbb{R}^n, \tag{2.6}$$

where $|.|$ is the determinant of a square matrix, and $\mathbb{R}^n$ is the real coordinate space of n dimensions.

2.2.2 Chi-squared distribution

A random variable X of continuous type has a chi-squared distribution with r degrees of freedom, denoted by χ^2_r, if its PDF is

$$f_X(x) = \begin{cases} \dfrac{1}{\Gamma(r/2)2^{r/2}} x^{r/2-1} e^{-x/2} & 0 < x < \infty, \\ 0 & \text{otherwise}, \end{cases} \tag{2.7}$$

where Γ is the gamma function[11]. Now suppose that $\mathbf{X}$ has a multidimensional normal distribution with a positive definite covariance matrix $\mathbf{C_X}$, and $\mathbf{A}$ is a symmetric matrix of rank r with the same dimension as $\mathbf{C_X}$. It is shown that $Q = \mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}$ is chi-squared distributed, χ^2_r, if and only if $\mathbf{A}\mathbf{C_X}\mathbf{A} = \mathbf{A}$ (Craig et al., 2013, Chapter 9). This is a crucial statement used to derive the underlying PDF of the least-squares spectrum and spectrogram (Appendix C).

2.2.3 F-distribution

Assume that X and Y are two independent random variables that each follows the chi-squared distribution with r_1 and r_2 degrees of freedom, respectively. Then the distribution of random variable $Z = (r_2/r_1)X/Y$ is called the F-distribution that is

$$f_Z(z) = \begin{cases} \dfrac{\Gamma\big((r_1+r_2)/2\big)(r_1/r_2)^{r_1/2}}{\Gamma(r_1/2)\Gamma(r_2/2)} \dfrac{z^{r_1/2-1}}{(1+r_1 z/r_2)^{(r_1+r_2)/2}} & 0 < z < \infty, \\ 0 & \text{otherwise}. \end{cases} \tag{2.8}$$

---

[9] The transpose of a matrix is a new matrix whose rows are the columns of the original matrix which also makes the columns of the new matrix the rows of the original matrix.

[10] A symmetric matrix $\mathbf{C}$ is a square matrix such that $\mathbf{C} = \mathbf{C}^{\mathrm{T}}$.

[11] The gamma function is defined as $\Gamma(r) = \int_0^\infty y^{r-1}e^{-y}\mathrm{d}y$. If r is a positive integer, then it can be verified that $\Gamma(r) = (r-1)!$ and $\Gamma(1+r/2)/\Gamma(r/2) = r/2$, where "!" is the factorial.

2.2.4 Beta distribution

A random variable X in $(0,1)$ has the beta distribution with shape parameters $p_1 > 0$ and $p_2 > 0$, denoted by β_{p_1,p_2}, if its PDF has the form

$$f_X(x) = \Gamma(p_1+p_2)/\big(\Gamma(p_1)\Gamma(p_2)\big)\ x^{p_1-1}(1-x)^{p_2-1}. \tag{2.9}$$

It is shown (Craig et al., 2013, Chapter 3) that if Z has an F-distribution with degrees of freedom r_1 and r_2, then the random variable $X = \big(1+(r_1/r_2)Z\big)^{-1}$ has a beta distribution. In Appendix C, it is shown that when a time series has been derived from a population of random variables following the multidimensional normal distribution, the least-squares spectrum and spectrogram follow the beta distribution.

2.3 Least-Squares Spectral Analysis (Simple Case)

The Least-Squares Spectral Analysis (LSSA) is an appropriate method of analyzing unequally spaced time series that comprises trends and/or datum shifts (Vaníček, 1969, 1971). The LSSA has been described in detail in many publications (Lomb, 1976; Wells et al., 1985; Craymer, 1998; Pagiatakis, 1999) and applied in many experiments (Scargle, 1982; Wu et al., 1995; Hui and Pagiatakis, 2004; El-Gelil et al., 2008; Psimoulis et al., 2008). Its basic definition for time series is described below.

Suppose that $\mathbf{f}$ is a real times series of n data points, $\mathbf{f} = \big[f(t_1), f(t_2), \ldots, f(t_n)\big]^{\mathrm{T}}$, where t_j's may not be equally spaced. Also, suppose that $\mathbf{C_f}$ is the covariance matrix of dimension n associated with the time series. Note that each data point $f(t_j)$ can be generally assumed to be a random variable with mean μ_j and variance σ_j^2, and so the covariance between any two data points may be calculated. The reader is referred to Chapter 3 and Craig et al. (2013) for more detail on the covariance matrix. Let $\mathbf{P} = \mathbf{C_f}^{-1}$, and $\mathbf{\Omega}$ be a set of cyclic frequencies (positive real numbers) that may be chosen based on the scope of analysis.

The LSSA seeks to find a vector $\mathbf{c} = [c_1, \ldots, c_m]^{\mathrm{T}}$ to represent $\mathbf{f}$ as $\mathbf{f} = \mathbf{\Phi c}$. Matrix $\mathbf{\Phi}$ of dimension $n \times m$, which models the physical relationship between the time series (the vector of measurements) and the unknown vector $\mathbf{c}$, is called the "design matrix" (Vaníček and Krakiwsky, 1986). If the dimension of $\mathbf{f}$ is larger than m, then the system $\mathbf{f} = \mathbf{\Phi c}$ is overdetermined, i.e., the system has more equations than unknowns. In order to solve the system, one can recall the least-squares principle to minimize quadratic form $\mathbf{\Psi} = (\mathbf{f} - \mathbf{\Phi c})^{\mathrm{T}}\mathbf{P}(\mathbf{f} - \mathbf{\Phi c})$. This results in the well-known expression

$$\hat{\mathbf{c}} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P}\mathbf{f}, \tag{2.10}$$

where $\hat{\mathbf{c}} = [\hat{c}_1, \ldots, \hat{c}_m]^{\mathrm{T}}$ (Wells et al., 1985; Pagiatakis, 1999). By the covariance law[12], the covariance of $\hat{\mathbf{c}}$

[12] If $\mathbf{Y} = \mathbf{BX}$, then by the covariance law, $\mathbf{C_Y} = \mathbf{BC_X B}^{\mathrm{T}}$ (Wells and Krakiwsky, 1971, Chapter 2).

can be obtained as

$$\mathbf{C}_{\hat{\mathbf{c}}} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi})^{-1}. \tag{2.11}$$

Design matrix $\mathbf{\Phi}$ comprises sine and cosine basis functions of various cyclic frequencies $\omega_k \in \mathbf{\Omega}$ as well as other basis functions that characterize systematic noise (e.g., trends and offsets). The simplest case (out-of-context) is when $\mathbf{c} = [c_1, c_2]^{\mathrm{T}}$, and $\mathbf{\Phi}$ is a $n \times 2$ matrix whose columns are the cosine and sine basis functions of a cyclic frequency ω_k. In this case, the least-squares spectrum (LSS) of $\mathbf{f}$ for ω_k is defined as

$$s(\omega_k) = \frac{\mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}\hat{\mathbf{c}}}{\mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{f}}, \tag{2.12}$$

which estimates how much of the wave of cyclic frequency ω_k contributes to $\mathbf{f}$, $s(\omega_k) \in (0, 1)$. Aliasing will not occur in time series that are inherently unequally spaced (Shapiro and Silverman, 1960) underlining yet another important advantage of the LSSA when applied to unequally spaced time series. The LSSA can also consider constituents of known forms, such as trends, datum shifts, and sinusoidal basis functions of particular frequencies. The number of these constituents is denoted by "q" in this dissertation. Simultaneous consideration of these constituents, when estimating the amplitudes of the sinusoidal basis functions, are described in detail in Chapter 3.

In order to find the underlying PDF of the LSS, suppose that $\mathbf{f}$ is a random vector of dimension n following the multidimensional normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{f}})$, where $\mathbf{C}_{\mathbf{f}}$ may be singular. Note that $\mathbf{f}$ can also have any arbitrary mean that may be modeled by constituents of known forms (Steeves, 1981). In Appendix C, it is shown that the LSS follows the beta distribution. The critical value at $(1 - \alpha)$ confidence level (α is the significance level) for the LSS given by Equation (2.12), without considering the constituents of known forms ($q = 0$), is $1 - \alpha^{2/(n-2)}$ (see Appendix C). Therefore, the spectral peaks in the LSS whose values are greater than the critical value are statistically significant at $(1 - \alpha)$ confidence level.

It is worth mentioning that the least-squares minimization has many applications in wireless networking and telecommunications. For instance, Seberry et al. (2005) showed that if $\mathbf{\Phi}$ is a weighing design whose entries are ± 1, and the associated covariance matrix to $\mathbf{f}$ is $\mathbf{C}_{\mathbf{f}} = \sigma^2\mathbf{I}$, then $\mathbf{\Phi}$ is optimal if $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi} = n\mathbf{I}$, and so from Equation (2.11), $\mathbf{C}_{\hat{\mathbf{c}}} = (\sigma^2/n)\mathbf{I}$. That means $\mathbf{\Phi}$ is optimal (the variances of the elements of $\hat{\mathbf{c}}$ are minimum) if it is a ± 1 matrix whose columns are orthogonal and are taken from a Hadamard matrix[13]. The existence of Hadamard matrices and their generalizations (orthogonal designs) were discussed comprehensively in the author's first PhD dissertation (Ghaderpour, 2013).

Craymer (1998) showed that the Fourier analysis is a special case of the LSSA when the time series is strictly stationary, equally spaced, equally weighted, and with no gaps (see also Appendix A). The Fourier transform of a function f for cyclic frequency ω is obtained by considering the entire domain of f similar to

[13] A Hadamard matrix is a square matrix $\mathbf{H}$ of order n with ± 1 entries such that $\mathbf{H}\mathbf{H}^{\mathrm{T}} = n\mathbf{I}$.

the LSSA that obtains the spectrum of a time series by fitting sinusoidal functions of cyclic frequency ω to the entire time series; this is not a satisfactory approach for analyzing non-stationary time series comprising constituents with variability of amplitude and frequency over time.

2.4 Spectral Leakages in Spectrum of Unequally Spaced Time/Data Series

Regularization is a crucial problem in many areas, such as seismology. Geophysicists often use the recorded seismic data created by sources (vibrators) and receivers to study the geologic structures beneath the Earth's surface or sea floor for several purposes, such as discovering oil and gas before drilling. Marine seismic data sets are usually irregularly sampled along spatial directions due to several reasons, such as editing bad data sets and economy. Regularly sampled seismic data are required for various purposes including wave equation migration, seismic inversion, amplitude versus azimuth or offset[14] analyses and surface-related multiple elimination (Scales, 1997; Crawley, 2000).

Since the sinusoidal basis functions in the Fourier transform are not orthogonal over irregularly spaced series, spectral leakages appear in the Fourier spectrum that is, power/energy leaks from one spectral peak into another (Xu et al., 2005, 2010). In the LSSA, since the selected frequencies are examined one by one (out-of-context), spectral leakages still appear which result in interpolation inaccuracy, although the non-orthogonality between the sine and cosine basis functions is taken into account for each frequency.

Example: To visualize the spectral leakages, a synthetic unequally spaced time series is generated as described below that may represent a geophysical time series obtained in one month (mo). Suppose that $\mathbf{f} = \big[f(t_1), \ldots, f(t_{70})\big]$, where $f(t_\ell) = 1.5\sin(2\pi \cdot 6t_\ell) + \sin(2\pi \cdot 15t_\ell)$ and t_ℓ's ($\ell = 1, ..., 70$) are random numbers in $[0, 1]$ generated by the MATLAB command 'rand' and sorted in ascending order (Figure 2.1a).

The time series contains two sine waves of cyclic frequencies 6 and 15 cycles per month (c/mo) with amplitudes 1.5 and 1, respectively. The ideal spectrum in terms of amplitudes is shown in Figure 2.1b that clearly shows the spectral peaks corresponding to the two sine waves. The LSS of this time series calculated by Equation (2.12) (suppose $\mathbf{P} = \mathbf{I}$, the identity matrix) at 99% confidence level is illustrated in Figure 2.1c. It can be seen that the spectral peaks corresponding to the two constituents are significant, and the other peaks are not statistically significant to mislead the identification of the dominant spectral frequencies.

In the LSSA, the estimated amplitudes of this time series is obtained by Equation (2.10), then the square root of $\hat{\mathbf{c}}^{\mathrm{T}}\hat{\mathbf{c}}$ is calculated, and its corresponding spectrum is illustrated in Figure 2.1d. Alternatively, the absolute values of the Fourier coefficients obtained from Equation (2.4) is also shown in Figure 2.1e, showing severe spectral leakages (see arrows). The absolute value of Fourier coefficients are approximately half of amplitudes shown in Figure 2.1d but with severe spectral leakages (see also Appendix A). Note that the

[14] In geophysics and reflection seismology, amplitude versus offset or amplitude variation with offset is the general term for referring to the dependency of the seismic attribute, amplitude, with the distance between the source and receiver (the offset).

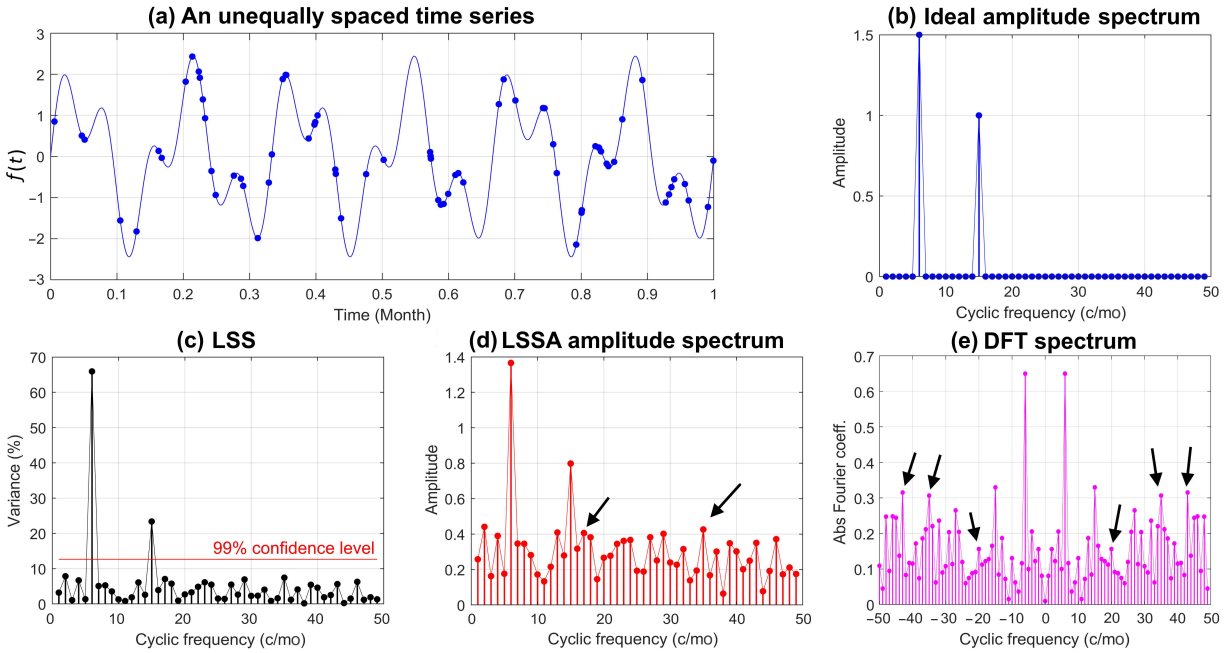

Figure 2.1: Spectral leakages in the DFT and LSSA for unequally spaced time series. Arrows show many of the leakages.

DFT of a real time series has identical response to both positive and negative frequencies.

The LSS peaks are cleaner and sharper in the percentage variance mode (Figure 2.1c) than amplitude mode (Figure 2.1d), and also the statistics of the significant peaks is on the percentage variance and not on the amplitude. This is a fundamental and crucial approach in the development of the antileakage LSSA algorithm for the purpose of accurate regularization (Ghaderpour et al., 2018b). Moreover, the identification of signal frequencies in the DFT is more difficult and challenging compare to the LSSA because the DFT does not consider the correlations among the sine and cosine basis functions (Pagiatakis, 1999).

Seismic data are well sampled in time (regularly sampled), but usually are poorly sampled along the spatial directions (irregularly spaced). Note that the LSSA and its applications are mainly discussed on unequally spaced time series in literature; however, its concept can concurrently be applied to the spatial direction that is based on wavenumber (cycles per unit distance) rather than frequency (cycles per unit time). Therefore, the term "data series" used in Chapters 3, 4 and 5 indicates a series of data points sampled along spatial direction.

2.5 Recent Methods of Mitigating Spectral Leakages in Spectrum

The AntiLeakage Fourier Transform (ALFT) (Xu et al., 2005, 2010, 2011) and Arbitrarily Sampled Fourier Transform (ASFT) (Guo et al., 2015) are two Fourier-based methods that try to mitigate the spectral

leakages in the Fourier spectrum of unequally spaced time series. The ALFT and ASFT estimate the Fourier coefficients of a data series, first by searching for the peak of maximum energy and subtracting its component from the data series and then repeating the procedure again on the residual data series until all the significant Fourier coefficients are estimated. This way the spectral leakages of Fourier coefficients emerging from the non-orthogonality of the global Fourier basis functions will be attenuated. However, these methods usually cannot find the correct location of a peak of maximum energy from a preselected set of wavenumbers because of the correlation between the sinusoids, and so these methods do not effectively reduce the leakage. This shortcoming becomes more severe when a data series has more spectral components. Moreover, the constituents of known forms, such as datum shifts and trends, are not explicitly considered in these algorithms. Therefore, if the trends or datum shifts are present in a data series, then the ALFT and ASFT implicitly approximate them by sinusoids of various wavenumbers (Xu et al., 2010). Hollander et al. (2012) proposed a method that uses an Orthogonal Matching Pursuit (OMP) to improve the ALFT results. In the OMP approach, the coefficients of all previously selected Fourier components are re-optimized, producing better interpolation results compare to the ALFT results.

The Interpolation by MAtching Pursuit (IMAP) is a least-squares based method proposed to mitigate the spectral leakages in the least-squares spectrum and consequently can be used to regularize seismic data (Özbek et al., 2009). In the IMAP, a wavenumber k_1 corresponding to the largest peak in the Lomb spectrum (Lomb, 1976) is selected, i.e., a wavenumber k_1 such that (2.12) (or just its numerator) is maximized. Then the contribution of the sinusoidal basis functions of wavenumber k_1 to the data series is subtracted from it to obtain the residual data series. In the next step, the residual data series is treated as a new input data series to choose a wavenumber k_2 corresponding to the largest peak in the new residual Lomb spectrum, and this procedure continues until the desired wavenumbers are estimated. This iterative method mitigates the spectral leakages in the Lomb spectrum and results in a faster convergence.

Unlike the ALFT, the IMAP has the ability to match the basis functions to the data rather than the data being mapped to the basis-function domain (Vassallo et al., 2010a). Since the correlation between the sine and cosine basis functions are considered in the IMAP, its spectrum generally shows much less leakages than the ALFT and ASFT, and so it performs better for regularization. Although the IMAP improves the spectral leakages in the Lomb spectrum, it does not explicitly account for the correlations among the sinusoids of different wavenumbers, and so its spectrum contains many wavenumbers due to leakages that increase the computational cost and reduce the accuracy of the regularization results. Vassallo et al. (2010b) have also studied estimating many wavenumbers per iteration and have documented several dealiasing schemes. The ALFT and ASFT will be mathematically described in Chapter 3 and will be compared to the new method, the AntiLeakage Least-Squares Spectral Analysis (ALLSSA).

2.6 Short-Time Fourier Transform

In order to measure the frequency variation of sounds, Gabor (1946) introduced a windowing technique called the Short-Time Fourier Transform (STFT), which is one way of analyzing piece-wise stationary and equally spaced time series. He basically endeavoured to obtain the spectrum of a time series in short-time intervals (or within a window which translates through the whole time series) rather than considering the entire time series at once. Mathematically, let f be a square-integrable function on $\mathbb{R}$ (e.g., an energy signal), and w be a normalized ($\|w\|^2 = \int_{-\infty}^{\infty} |w(t)|^2 \, \mathrm{d}t = 1$) real and symmetric ($w(t) = w(-t)$) window function (e.g., Hann or Gaussian). The continuous STFT is defined as

$$\hat{f}(\tau, \omega) = \int_{-\infty}^{\infty} f(t) w(t - \tau) e^{-2\pi i \omega t} \, \mathrm{d}t, \tag{2.13}$$

where w is centred at time τ, the translation parameter. The STFT energy density of f (spectrogram) is $|\hat{f}(\tau, \omega)|^2$, which measures the energy of f in the time-frequency neighbourhood of (τ, ω) (Mallat, 1999, Chapter 4). For a time series, a spectrogram can be obtained by calculating $|\hat{f}(\tau, \omega)|^2$ (in the summation form) for all pairs (τ, ω).

The rectangular boxes centred at (τ, ω) with side lengths σ_τ and σ_ω (standard deviation) are illustrated in Figure 2.2, and they are called the Heisenberg boxes (Mallat, 1999). The word "resolution" is referred to the side lengths of the Heisenberg boxes in the time-frequency domain (or time or frequency support). In other words, resolution means ability to resolve frequencies or times in a spectrogram. It is shown that σ_τ and σ_ω are independent from τ and ω, and so the STFT has the same resolution across the time-frequency domain, i.e., the area and the side lengths of the Heisenberg boxes do not change across the time-frequency domain, and the Heisenberg boxes do not rotate across the time-frequency domain (Mallat, 1999). If the selected window is wider, then σ_τ is larger and σ_ω is smaller, meaning that the spectrogram has poorer time resolution but better frequency resolution across the time-frequency domain (Figure 2.2a). Similarly, if the selected window is narrower, then σ_τ is smaller and σ_ω is larger, meaning that the spectrogram has better time resolution but poorer frequency resolution across the time-frequency domain (Figure 2.2b).

Since in the STFT the sinusoidal functions that do not complete an integer number of cycles within a window are not orthogonal, the segment of the time series within such window cannot be represented as linear combination of these sinusoids properly, resulting in spectral leakages. Note that the STFT does not consider the correlation between the sinusoidal functions when they are not orthogonal within a window (Mallat, 1999). Using an appropriate window function, the segment of the time series within the window can be tapered to zero at both ends to prevent leakages. It is shown (Weyl, 1931) that if w is Gaussian, then the area of the Heisenberg box is minimum, a property that is similar to the least-squares principle and highly desirable (optimal). In the STFT, the window length is fixed during translation in time, inappropriate for analyzing time series with components of high amplitude and frequency variability over time.

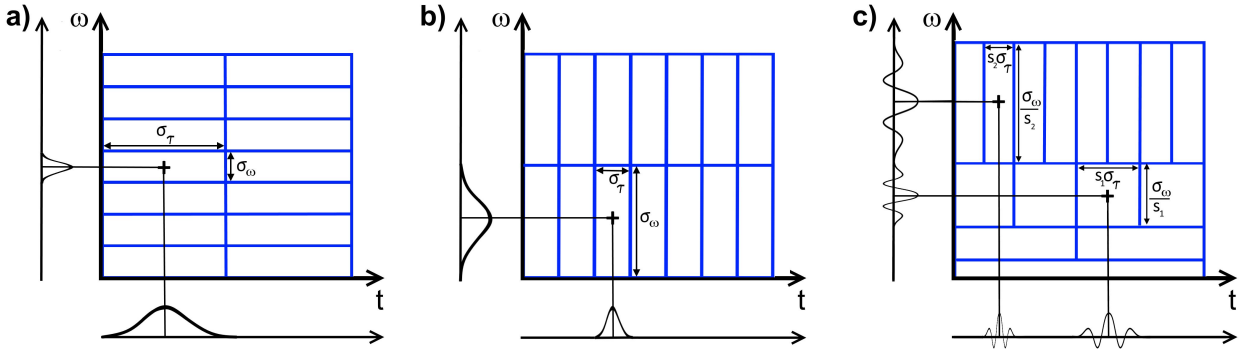

Figure 2.2: A few adjacent Heisenberg boxes of the STFT (a and b), and CWT (c).

2.7 Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) divides an equally spaced time series into wavelets to construct a time-frequency representation of the time series, providing a good time and frequency determination. In the CWT, the window length decreases as frequency increases allowing the detection of short-time (short duration) components of high frequencies, and vice versa: the window length increases as frequency decreases allowing the detection of the low frequency components in the time series.

In the CWT, the dilation (change in length) of a (mother) wavelet defines a window, and a scalogram (time-scale) is calculated instead of a spectrogram (time-frequency). However, one may convert the scalogram to a spectrogram using several approaches, such as conversion based on the centre frequencies of the scales (Hlawatsch and Boudreaux-Bartels, 1992). Mathematically, assume that the functions

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi\Big(\frac{t-\tau}{s}\Big), \tag{2.14}$$

are a family of (daughter) wavelets that are generated by dilating ψ (mother wavelet) with a scale parameter $s > 0$ and translating it by τ (time or position or location, etc.). For theoretical reasons, one may choose a special wavelet ψ (mother wavelet) from the space of all square-integrable functions over $\mathbb{R}$ such that it is normalized ($\|\psi\|^2 = \int_{-\infty}^{\infty} |\psi(\eta)|^2 \, \mathrm{d}\eta = 1$) and integrates to zero ($\int_{-\infty}^{\infty} \psi(\eta)\mathrm{d}\eta = 0$). The reader is referred to Daubechies (1990) and Farge (1992) for more details. The wavelet transform of a square-integrable function f on $\mathbb{R}$ (e.g., an energy signal) at time τ and scale s is then defined as

$$\hat{f}_\psi(\tau, s) = \big\langle f, \psi_{\tau,s} \big\rangle = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(t)\psi^*\Big(\frac{t-\tau}{s}\Big) \, \mathrm{d}t, \tag{2.15}$$

where ψ^* is the complex conjugate of ψ (Mallat, 1999, Chapter 4). Therefore, the values of time τ and scale s as well as the choice of wavelet affect the values of the coefficients $\hat{f}_\psi(\tau, s)$. An analytic wavelet transform defines a local time-frequency energy density $|\hat{f}_\psi(\tau, s)|^2$, which measures the energy of f in the Heisenberg box of each wavelet $\psi_{\tau,s}$. This energy density is called a scalogram (Mallat, 1999, Chapter 4). Note that the

Fourier transform in Equation (2.1) is a special case of the CWT in Equation (2.15) when $\psi_{\tau,s}(t) = e^{2\pi i\omega t}$ and f is absolutely integrable, ignoring scale s.

The CWT in MATLAB operates on discrete time series, where the integral in Equation (2.15) is implemented by a simple summation over the elements (Misiti et al., 2015). More precisely, suppose that $\mathbf{f} = \big[f(t_1), \ldots, f(t_n)\big]$ is an equally spaced time series of n data points, where $t_\ell = \ell/n$ ($\ell = 1, \ldots, n$). The CWT coefficients are calculated as

$$\hat{f}_\psi(t_j, s) = \frac{1}{\sqrt{ns}} \sum_{\ell=1}^{n} f(t_\ell)\psi^*\Big(\frac{t_\ell - t_j}{s}\Big), \tag{2.16}$$

where ψ^* is the complex conjugate of ψ, and the wavelet is stretched in time by varying its scale (s) and normalizing it to have unit energy (Torrence and Compo, 1998; Grinsted et al., 2004). For a time series, a scalogram can be obtained by calculating wavelet power $|\hat{f}_\psi(t_j, s)|^2$ for all pairs (t_j, s).

The time and frequency resolutions of the CWT depend on the scale s. It can be seen from Figure 2.2c that the CWT spectrogram has high-frequency resolution at low frequencies and high-time resolution at high frequencies (Mallat, 1999). The CWT is more useful for feature extraction and time series analysis due to its fine resolution; whereas, the discrete wavelet transform is a fast transform that represents a time series in a more redundant form and often used for data compression and noise attenuation of two dimensional signals, such as images (Mallat, 1999; Grinsted et al., 2004). Note that the discrete wavelet transform of a time series is calculated by passing it through a series of filters, calculated differently from the CWT for different purposes (Daubechies, 1990; Mallat, 1999) and is not in the scope of this dissertation.

Note that unequally spaced wavelets are not defined, so the CWT and discrete wavelet transform are not defined for an unequally spaced time series unless one interpolates to fill in the missing data or shrinks the wavelet (Hall and Turlach, 1997; Sardy et al., 1999). However, the interpolation is not appropriate as it damages the time series and results in erratic spectrograms in the time-frequency domain, especially for non-stationary time series. Moreover, the first two statistical moments of the time series are not considered in the STFT and CWT from their definitions.

There are a bewildering variety of wavelets to use in the CWT, and if one uses an inappropriate wavelet to correlate with a time series, then the CWT spectrogram may give misleading and nonphysical results (Qian, 2002). In order to study the periodicity of the constituents and to detect the periodic fluctuations in a time series, one may choose a (mother) wavelet which fluctuates based on the periodic sinusoidal basis functions (Grossmann et al., 1985; Foster, 1996). A simple and commonly used wavelet possessing this property is the Morlet wavelet that is defined as

$$\psi(\eta) = \pi^{-1/4} e^{-\eta^2/2} e^{2\pi i\omega_0 \eta}, \tag{2.17}$$

where the cyclic frequency ω_0 is chosen to be $6/(2\pi)$ cycles per unit time (s) to satisfy the wavelet admissibility

condition[15] of the mother wavelet and provide a good balance between time and frequency resolution (Farge, 1992; Torrence and Compo, 1998; Grinsted et al., 2004). For this choice of ω_0, the Fourier period of the Morlet wavelet ($T = 1/\omega_0$) is almost equal to 1.03 s.

To match the inverse of scale factor s exactly to frequency, Foster (1996) applied a simple transformation to define the Morlet wavelet as

$$\psi(z) = e^{-cz^2} e^{iz} = e^{-c(2\pi\omega(t-\tau))^2} e^{2\pi i\omega(t-\tau)}, \tag{2.18}$$

where ω is the cyclic frequency and $z = 2\pi\omega_0\eta$, so $c = 1/(8\pi^2\omega_0^2)$, disregarding coefficient $\pi^{-1/4}$ (cf., Equation (2.17)) because it does not affect the results of the least-squares methods. For MATLAB users, the command "scal2frq" may be used to convert the scales to frequencies in the CWT (Misiti et al., 2015). However, the Morlet wavelet given by Equation (2.18), used in the least-squares methods in Chapter 6, directly computes a spectrogram instead of a scalogram. Note that when $\omega_0 = 6/(2\pi)$, $c = 1/72$ that is approximately close to $c = 0.0125$ used in Foster (1996). Figure 2.3 shows the construction and analyses of a Morlet wavelet given by Equation (2.18) for cyclic frequency $\omega = 10$ cycles per second (Hertz or Hz). Figure 2.3a and c are generated using $e^{2\pi i\omega(t-\tau)}$ and $e^{-c(2\pi\omega(t-\tau))^2}$, respectively, that are the two components of the Morlet wavelet shown in Figure 2.3e. Comparing Figure 2.3b to 2.3f, one can see that the bandwidth[16] of the Morlet wavelet is larger (poor frequency resolution) than the pure sinusoids.

2.8 Cross-Wavelet Transform

In order to examine two equally spaced time series together (with the same sampling rates) and study the coherency and phase differences between their corresponding constituents, Torrence and Compo (1998) proposed a method called the Cross-Wavelet Transform (XWT). Briefly, suppose that $\mathbf{f_1}$ and $\mathbf{f_2}$ are two equally spaced time series of size n. For each pair (t_j, s), let $\hat{f}_{1\psi}(t_j, s)$ and $\hat{f}_{2\psi}(t_j, s)$ be their corresponding CWT coefficients calculated by Equation (2.16). The XWT is defined as

$$X_\psi(t_j, s) = \Big(\hat{f}_{1\psi}(t_j, s)\Big)\Big(\hat{f}^*_{2\psi}(t_j, s)\Big), \tag{2.19}$$

where $\hat{f}^*_{2\psi}(t_j, s)$ is the complex conjugate of $\hat{f}_{2\psi}(t_j, s)$, and the XWT power is defined as $|X_\psi(t_j, s)|$ (Torrence and Compo, 1998; Grinsted et al., 2004). The argument[17] $\arg[X_\psi(t_j, s)]$ defines the phase differences (local

[15]The condition $\int_0^\infty (|\Psi(\omega)|^2/\omega)\ \mathrm{d}\omega < \infty$ is called the wavelet admissibility condition, where $\Psi(\omega)$ is the Fourier transform of ψ. This condition is a necessary condition for the inverse of the CWT (Mallat, 1999).

[16]There are several definitions for bandwidth in the literature. The bandwidth is typically understood as the frequency interval in the spectrum where the main part (95% or 99%) of its power is located. Others use 50% to be compatible with the definition of the cutoff frequency of a filter when its power drops to -3 decibel (dB) (Mallat, 1999; Valkenburg, 2006).

[17]The argument of a complex number $z = a + ib$ is $\tan^{-1}(b/a)$.

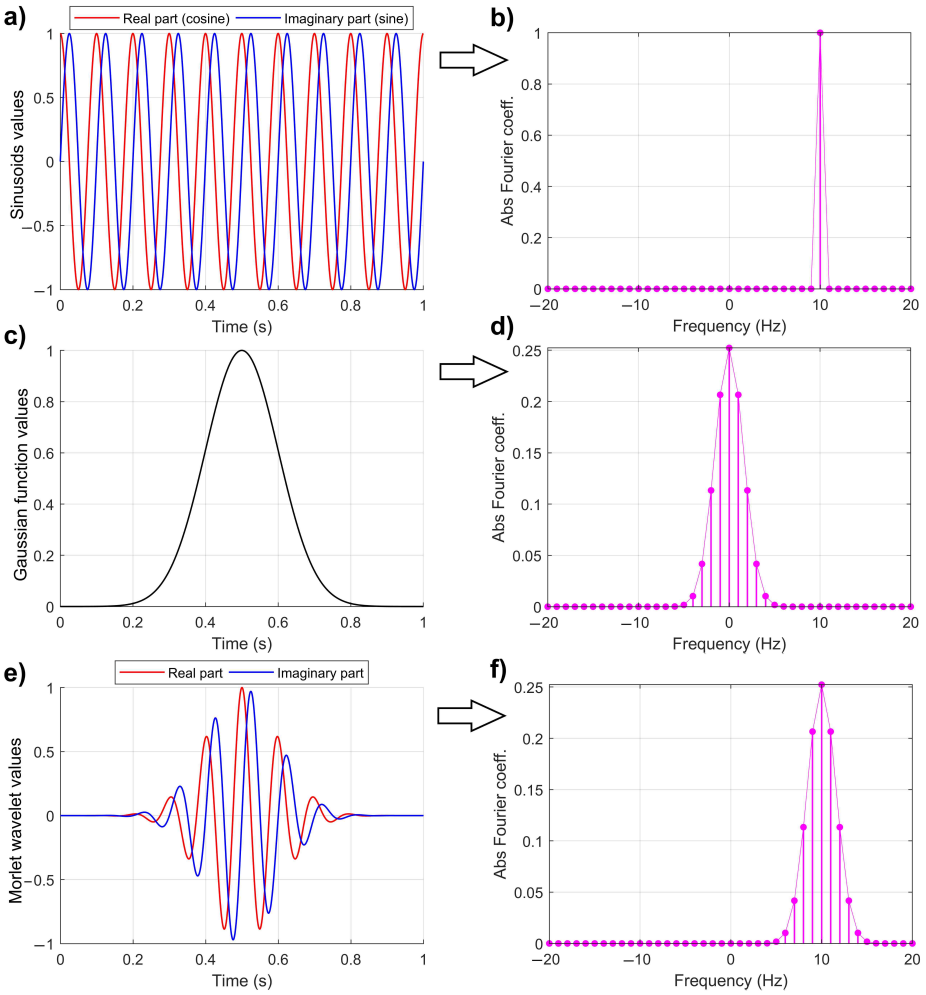

Figure 2.3: Morlet wavelet of frequency 10 Hz and its analyses (1000 samples per second). (a) Sinusoids of frequency 10 Hz, (b) DFT spectrum of panel a, (c) a Gaussian function, (d) DFT spectrum of panel c, (e) a Morlet wavelet, and (f) DFT spectrum of panel e.

relative phase or coherence phase) between the corresponding constituents of the two time series (Torrence and Compo, 1998; Grinsted et al., 2004).

Torrence and Compo (1998) showed that if a time series is derived from a population of random variables following the multidimensional normal distribution, then the real and imaginary parts of the Fourier coefficients and so the CWT coefficients are also normally distributed. Therefore, the continuous wavelet power, that is the sum of squares of two normally distributed random variables, follows χ^2_2 (chi-squared distribution with two degrees of freedom). Now if both time series are independently derived from populations of random variables following the multidimensional normal distribution, then the underlying PDF of the squared absolute value of XWT, i.e., $|X_\psi(t_j, s)|^2 = |\hat{f}_{1\psi}(t_j, s)|^2 |\hat{f}_{2\psi}(t_j, s)|^2$, is in fact the product of two chi-squared distributions (each with 2 degrees of freedom). Ge (2008) following Torrence and Compo (1998) showed that

the underlying PDF of the XWT power ($|X_\psi(t_j, s)|$) is $f(z) = zK_0(z)$, where $K_0(z)$ is the modified Bessel function of order zero[18].

Since in the locations toward the edges of the CWT spectrogram the wavelet is not completely localized, edge artifacts occur in those locations. Torrence and Compo (1998) introduce the cone of influence (COI), outside which the edge effects can be neglected. Grinsted et al. (2004) take the COI as the area in which the wavelet power caused by a discontinuity at the edge has decreased to e^{-2} of the value at the edge. The COI may be applied to the CWT or XWT spectrograms.

Although the CWT and XWT are well-established methods of time-frequency analysis, they cannot be used for analyzing unequally spaced series with associated covariance matrices. These methods cannot also handle the constituents of known forms, such as trends and/or datum shifts, and so are not appropriate for analyzing many time series in geophysical applications.

2.9 Weighted Wavelet Z-Transform

Foster (1996) proposed a robust method of analyzing astronomical time series that are non-stationary and unequally spaced, namely, the Weighted Wavelet Z-transform (WWZ). In this section, the same notation is used as in Foster (1996) to define the WWZ. Suppose that $\mathbf{f} = \big[f(t_j)\big]$ is a time series of n data points. For each τ (the window location; can be the t_j's or equally spaced times) and each $\omega_k \in \mathbf{\Omega}$, let $\mathbf{\Phi}_1 = [\mathbf{1}]$

$$\mathbf{\Phi}_2 = \big[\cos\big(2\pi\omega_k(t_1 - \tau)\big), \ldots, \cos\big(2\pi\omega_k(t_n - \tau)\big)\big], \tag{2.20}$$

$$\mathbf{\Phi}_3 = \big[\sin\big(2\pi\omega_k(t_1 - \tau)\big), \ldots, \sin\big(2\pi\omega_k(t_n - \tau)\big)\big], \tag{2.21}$$

where $[\mathbf{1}]$ is the all-ones vector of dimension n. The inner product of two vectors $\mathbf{u} = [u(t_1), \ldots, u(t_n)]$ and $\mathbf{v} = [v(t_1), \ldots, v(t_n)]$ is defined as

$$\langle \mathbf{u} | \mathbf{v} \rangle = \frac{\sum_{i=1}^{n} w_i u(t_i) v(t_i)}{\sum_{i=1}^{n} w_i}, \tag{2.22}$$

where w_i is the statistical weight chosen as $w_i = e^{-c(2\pi\omega_k(t_i - \tau))^2}$, and c is a window constant which may be selected to be $c = 0.0125$ as described in Section 2.7, see Equation (2.18). The constant c determines how rapidly the analyzing wavelet decays (Foster, 1996). Let $\mathbf{S}$ be the square matrix of order 3 whose ab entry is $\langle \mathbf{\Phi}_a | \mathbf{\Phi}_b \rangle$ $(1 \le a, b \le 3)$. The Weighted Wavelet Z-transform (WWZ) (power) is defined as

$$Z = \frac{(N_{\text{eff}} - 3) V_y}{2(V_x - V_y)}, \tag{2.23}$$

where $N_{\text{eff}} = (\sum_{i=1}^{n} w_i)^2 / (\sum_{i=1}^{n} w_i^2)$ is the effective number of data points, $V_x = \langle \mathbf{f} | \mathbf{f} \rangle - \langle \mathbf{1} | \mathbf{f} \rangle^2$ is the weighted variation of the data, and $V_y = \sum_{a,b} \mathbf{S}_{ab}^{-1} \langle \mathbf{\Phi}_a | \mathbf{f} \rangle \langle \mathbf{\Phi}_b | \mathbf{f} \rangle - \langle \mathbf{1} | \mathbf{f} \rangle^2$ is the weighted variation of the model function obtained by the least-squares method (Foster, 1996).

[18]The modified Bessel function of order zero is $K_0(z) = \int_0^\infty \cos\,(z \sinh t)\,\mathrm{d}t = \int_0^\infty \cos\,(xt)/\sqrt{t^2+1}\,\mathrm{d}t$ (Abramowitz and Stegun, 1972, Chapter 9).

The WWZ is based on the least-squares optimization that is a robust method of analyzing non-stationary and unequally spaced time series; however, the WWZ is a poor estimator of amplitude (Foster, 1996), and the spectral peaks in the WWZ lose their power toward higher frequencies similar to the CWT. Therefore, it may not detect possible hidden signatures in time series. Furthermore, accounting for constituents of known forms and covariance matrices associated with the time series are not discussed in Foster (1996) that may potentially alter the results. With several examples in Chapter 6, the CWT and WWZ are compared to my proposed method, namely, the Least-Squares Wavelet Analysis (LSWA) to show the advantages of the LSWA over the CWT and WWZ.

2.10 Hilbert-Huang Transform

The Hilbert-Huang Transform (HHT) is an alternative method of analyzing non-stationary time series with or without trends (Huang et al., 1998; Huang and Wu, 2008; Barnhart, 2011). The HHT is the result of the Empirical Mode Decomposition (EMD) and the Hilbert Spectral Analysis (HSA). The EMD decomposes the time series into Intrinsic Mode Functions (IMFs) with a trend via a procedure called sifting, and then HSA applies the IMFs to obtain instantaneous frequency data.

An IMF is a function whose number of extrema and zero-crossings are either equal or differ by at most one. Moreover, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at any point (Huang and Wu, 2008). The sifting process identifies all the local extrema in time series $\mathbf{f}$, then it connects all the local maxima by a cubic spline as the upper envelope and does the same for all the local minima to produce the lower envelope. Therefore, the upper and lower envelopes with mean $\mu(t)$ should cover all the data points between them (this generally may not occur). Then the difference between the time series values and $\mu(t)$ is ideally an IMF. Otherwise, the same process on the difference may be repeated; this is one the weaknesses of the HHT (Huang and Wu, 2008).

Denote this IMF by $\mathbf{c}_1$. The sifting process now will continue on $\mathbf{r}_1 = \mathbf{f} - \mathbf{c}_1$ (residual) to obtain $\mathbf{c}_2$ (IMF), and so $\mathbf{r}_2 = \mathbf{r}_1 - \mathbf{c}_2$. The sifting process stops when residual $\mathbf{r}_n$ becomes a monotonic function from which no more IMF can be extracted (Huang et al., 1998). Therefore, the time series is decomposed into IMFs, i.e., $\mathbf{f} = \sum_{j=1}^{n} \mathbf{c}_j + \mathbf{r}_n$. After obtaining the IMFs, $\mathbf{c}_j$'s, the instantaneous frequencies, $\omega_j(t)$'s, and instantaneous amplitudes, $a_j(t)$'s, of the IMFs can be obtained using the Hilbert transform, the convolution integral of $\mathbf{c}_j$ with $1/t$ (Huang et al., 1998; Huang and Wu, 2008; Barnhart, 2011). The distribution of amplitude a_j(t) in the time-frequency domain, $H(t, \omega)$, can be regarded as a skeleton form of that in the CWT (Huang and Wu, 2008).

Although the HHT is an empirical method based on the physical meaning of non-stationary time series, it has several weaknesses, such as limitation of the EMD in distinguishing different components in narrow-band time series, end effect, and mode mixing problem occurring in the EMD process (Guang et al., 2014;

Chen et al., 2015). Ädatig and Schlurmann (2004) comprehensively investigated the performance and limitations of the HHT with particular applications to irregular water waves. They showed that an appropriate interpolation technique is needed to determine better envelopes for this method.

2.11 Constrained Least-Squares Spectral Analysis

Puryear et al. (2012) introduced the Constrained Least-Squares Spectral Analysis (CLSSA) as a seismic spectral decomposition method and showed that it has spectral resolution advantages over the conventional methods, such as the CWT. Suppose that $\mathbf{\Phi}_w$ is a design matrix consisting of the sinusoidal functions truncated by the end points of a window in the time domain. More precisely, let

$$\mathbf{\Phi}_w = \begin{bmatrix} e^{2\pi i\omega_1(t_1-\tau)} & e^{2\pi i\omega_2(t_1-\tau)} & \dots & e^{2\pi i\omega_n(t_1-\tau)} \\ e^{2\pi i\omega_1(t_2-\tau)} & e^{2\pi i\omega_2(t_2-\tau)} & \dots & e^{2\pi i\omega_n(t_2-\tau)} \\ \vdots & \vdots & \dots & \vdots \\ e^{2\pi i\omega_1(t_m-\tau)} & e^{2\pi i\omega_2(t_m-\tau)} & \dots & e^{2\pi i\omega_n(t_m-\tau)} \end{bmatrix}, \tag{2.24}$$

be a $m \times n$ matrix, where m is the number of data points within the time window, n is the number of frequencies (equal to the size of the entire data series according to the DFT), and τ is the centre of the window (the translating parameter). Now consider system (model) $\mathbf{\Phi}_w\mathbf{u} = \mathbf{y}$, where $\mathbf{y}$ is the windowed seismic data series (complex segment) of size m and $\mathbf{u}$ is the column vector of unknown Fourier coefficients. Since m is less than n, this system is underdetermined[19], and specific constraints are needed to estimate the unknowns and achieve a unique solution.

The CLSSA uses two diagonal matrices $\mathbf{W_u}$ and $\mathbf{W_y}$ that are the model and data weights, respectively, to constrain the underdetermined system. Matrix $\mathbf{W_u}$ changes in iteration (starts with $\mathbf{I}$), and $\mathbf{W_y}$ is constant throughout the iteration. Applying these two matrices, the underdetermined system becomes $\mathbf{W_y\Phi}_w\mathbf{W_u W_u^{-1}u} = \mathbf{W_y y}$. Let $\mathbf{\Phi_w} = \mathbf{W_y\Phi}_w\mathbf{W_u}$ and $\mathbf{u_w} = \mathbf{W_u^{-1}u}$. The CLSSA minimizes the following cost function

$$\Upsilon = \|\mathbf{\Phi_w u_w} - \mathbf{W_y y}\|^2 + \alpha\|\mathbf{u_w}\|^2, \tag{2.25}$$

where $\|.\|$ is the L2 norm[20] and parameter α is selected empirically and varies at each iteration step. The solution of Equation (2.25) is

$$\hat{\mathbf{u}}_\mathbf{w} = \mathbf{\Phi}_\mathbf{w}^*(\mathbf{\Phi_w\Phi}_\mathbf{w}^* + \alpha\mathbf{I})^{-1}\mathbf{W_y y}, \tag{2.26}$$

where "$*$" is the conjugate transpose. So, the estimated Fourier coefficients are $\hat{\mathbf{u}} = \mathbf{W_u}\hat{\mathbf{u}}_\mathbf{w}$.

[19] An underdetermined system is a system that has more unknowns than equations.

[20] For n-dimensional complex vector $\mathbf{z} = [z_1, \cdots, z_n]$, the L2 norm of $\mathbf{z}$ is $\|\mathbf{z}\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2} = \sqrt{z_1\bar{z}_1 + \cdots + z_n\bar{z}_n}$.

Although the CLSSA results in a time-frequency analysis with improved frequency resolution compared to the STFT and CWT, no detailed discussion and analyses are provided for analyzing unequally spaced time series by Puryear (2012). Also, this method may not be able to handle the constituents of known forms, such as trends and/or datum shifts, and so it is not appropriate for detection of hidden signatures in the data (time) series. Furthermore, no discussion is made for incorporating the covariance matrix in the model as well as the rigorous statistical analyses for possible test on the significance of spectral peaks.

2.12 Other Methods

For seismic data regularization, several effective algorithms are proposed by researchers that have their own advantages and weaknesses, such as the *f-x* and *t-x* Prediction Error Filters (PEF) (Spitz, 1991; Crawley, 2000; Fomel, 2002; Wang, 2002), generalized frequency-wavenumber interpolation (Gülünay, 2003), Projection Onto Convex Sets (POCS) and its modifications (Abma and Kabir, 2006; Yang et al., 2012; Wang et al., 2016), Minimum Weighted Norm Interpolation (MWNI) (Liu and Sacchi, 2004; Trad, 2009), and nonlinear shaping regularization (Fomel, 2007; Chen et al., 2015).

In the PEF theory, the data must be in wide-sense stationary and regular. The *f-x* PEFs represent a spectral approach that uses the available samples to predict the linear events and is suitable for aliased seismic data. For seismic data, windowing is required to approximately fulfill the stationarity condition of the PEFs. Seismic waves would travel a straight path to the other side of the Earth if there were no changes with depth, i.e., seismic events would be linear. In reality, seismic waves travel a curving path through the Earth due to changes in composition, pressure, and temperature within the layers of the Earth, i.e., seismic events are curved. When the seismic waves encounter boundaries between different media, the waves will react according to the Snell's law, and the angle of refraction across a boundary will depend on the velocity of the second medium relative to the first medium. Nonstationary PEF has also been used in seismic data interpolation to solve the interpolation of curved events (Crawley, 2000; Liu and Chen, 2017).

The POCS method is based on the Gerchberg-Saxton iterative algorithm (Gerchberg and Saxton, 1972) widely used in seismic data processing. Its performance on noisy data and its computational efficiency have been recently improved (Ge et al., 2015; Wang et al., 2016). Abma and Kabir (2006) used a linear threshold in the iterations of the POCS method and pointed out that the threshold is the key parameter in this method. Thereafter, other works have been done on an appropriate selection of a threshold value to improve the result of POCS method (Gao et al., 2013).

The MWNI method solves an underdetermined system of equations subject to suitable prior information for the purpose of regularization. The linear shaping regularization is another approach for regularizing underdetermined geophysical inverse problem that uses a linear shaping operator. Fomel (2007) started the idea of shaping regularization by recognizing smoothing as a fundamental operation. In a more general sense,

smoothing implies mapping of the input model to the space of admissible functions. He called this mapping operator "shaping". The nonlinear shaping regularization allows the shaping operator to be nonlinear making the iterative framework more flexible.

There are a number of methods proposed and applied to non-stationary time series, such as Wigner-Ville distribution (Classen and Mecklenbrauker, 1980; Baydar and Ball, 2001), empirical wavelet transform (Gilles, 2013), tunable-Q wavelet transform (Selesnick, 2011), and ensemble empirical mode decomposition (Huang and Wu, 2009; Ren et al., 2014). Each of them has its own advantages and weaknesses (Huang et al., 1998; Chen and Feng, 2003; Guang et al., 2014), and they are not generally appropriate for analyzing non-stationary time series that are unequally spaced and unequally weighted. Foster (1996) proposed an effective method of analyzing astronomical time series that are non-stationary and unequally spaced. His method is described in detail in Chapter 6. Amato et al. (2006) used a wavelet-based Hilbert space reproducing kernel technique and showed that their technique is superior in terms of the mean squared error. Mathias et al. (2004) proposed a method that is based on the least-squares method to deal with the undesirable side effects of nonuniform sampling in the presence of constant offsets.

There are also several other methods for analyzing two time series together, such as the least-squares self-coherency analysis (Pagiatakis et al., 2007a; El-Gelil et al., 2008; El-Gelil and Pagiatakis, 2009) that splits the series into shorter and independent segments depending on the study to be conducted. Each of the segments is analyzed via the LSSA, and the resulting spectra are multiplied to derive the product spectrum that strengthens common spectral peaks, particularly if the phenomenon under study contains signals with signal-to-noise ratio in the vicinity of -40 dB. The least-squares self-coherency method may be considered the precursor of the least-squares wavelet analysis and the least-squares cross-wavelet analysis developed in this research.

2.13 Conclusions

We reviewed a number of time/data series analyses methods. Many of them are Fourier-based methods (widely used) and mainly designed for analyzing equally spaced time/data series but require certain assumptions and have several limitations. Table 2.1 shows many of the popular state-of-the-art methods of time/data series analysis with some of their advantages and weaknesses reviewed in this chapter. The new methods that I shall introduce in the next chapters are based on the least-squares optimization. They are fairly easy to implement and interpret and can consider trends and/or datum shifts, constituents of known forms, and covariance matrices associated with the time/data series. Comparing to the state-of-the-art methods, they can estimate the amplitudes and frequencies (wavenumber) of constituents in any the time/data series more accurately by decomposing them into either frequency (wavenumber) domain or time-frequency domain without any assumption or need for editing the original series.

Table 2.1: State-of-the-art-methods of time/data series analysis with some of their advantages and weaknesses.

Methods	**Advantages**	**Shortcomings**
FFT	Fast method of obtaining a spectrum of an equally spaced time series, and the order of its complexity is $\mathcal{O}(n \log n)$	Not defined for unequally spaced time series, and not appropriate for analyzing non-stationary time series
ALFT	Mitigates the spectral leakages in the Fourier spectrum of an unequally spaced time series, used for seismic data regularization	Not appropriate for analyzing non-stationary time series, and generally is a slow method, usually $\mathcal{O}(n^3)$
ASFT	Like the ALFT used for regularization, but converges faster than the ALFT and also performs better in regularization compared to the ALFT	Not appropriate for analyzing non-stationary time series
LSSA	Can rigorously analyze both equally and unequally spaced time series, and it also considers constituents of known forms and covariance matrices	Not appropriate for analyzing non-stationary time series, and is slower than the FFT, yet its complexity depends on the desired set of frequencies
IMAP	Mitigates the spectral leakages in the least-squares spectrum of unequally spaced time series, used for seismic data regularization. It performs better and faster than the ALFT and ASFT	Not appropriate for analyzing non-stationary time series
CWT	A popular time-frequency analysis and used for feature extraction. It obtains a scalogram of an equally spaced and non-stationary time series. Its fast method achieves $\mathcal{O}(n)$ complexity per scale	Does not consider constituents of known forms, and not defined for unequally spaced time series. Its spectral peaks loss power toward higher frequencies, and not appropriate to detect hidden signatures
WWZ	A robust time-frequency analysis and used for feature extraction. It obtains a least-squares spectrogram of an unequally spaced and non-stationary time series. Its spectral peaks have higher resolutions than the ones in the CWT	It is generally slower than the CWT. It is a poor measure of amplitude, and it is not appropriate to detect hidden signatures
HHT	A popular time-frequency analysis that is an empirical method based on the physical meaning of non-stationary and equally spaced time series. It has finer frequency resolution than the CWT	Not appropriate for analyzing unequally spaced time series, and it is limited in distinguishing different components in narrow band time series. It has end effect and mode mixing problem
CLSSA	A time-frequency analysis based on the least-squares optimization. Its spectrogram has finer frequency resolution than the CWT spectrogram	Not appropriate for analyzing unequally spaced time series with gaps, trends, and/or datum shifts, and does not consider covariance matrices

3 Antileakage Least-Squares Spectral Analysis

In the analyses of irregularly spaced data series, the Discrete Fourier Transform (DFT) and the Least-Squares Spectral Analysis (LSSA) examine wavenumbers independently (out-of-context), and thus their spectrum shows spectral leakages that is, leakage of power/energy from one spectral peak into another. These leakages are caused by the non-orthogonality of the sinusoidal basis functions over unequally spaced series (see Appendix A). In this chapter, the AntiLeakage Fourier Transform (ALFT) and Arbitrary Sampled Fourier Transform (ASFT) are mathematically described in detail. The least-squares spectral analysis is revisited and slightly revised for data series rather than time series, and the AntiLeakage Least-Squares Spectral Analysis (ALLSSA) is introduced and compared with the Interpolation by MAtching Pursuit (IMAP). A synthetic data series example is provided to show the advantages of the ALLSSA over the ALFT, ASFT, and IMAP. This chapter is an unabridged version of sections "METHODS" and "SYNTHETIC DATA SERIES EXAMPLE" in Ghaderpour et al. (2018b).

3.1 Antileakage Fourier Transform

Suppose that $\mathbf{f} = [f(x_\ell)]$, $\ell = 1, \ldots, n$ is a data series of size n, where the x_ℓ's may be irregularly sampled. Assume that x_ℓ is a real number in the closed interval $[0, 1]$, where '1' is the unit distance (e.g., one kilometre). Define $\Delta x_\ell = \frac{1}{2}(x_{\ell+1} - x_{\ell-1})$ for $1 < \ell < n$. For the two special cases $\ell = 1$ and $\ell = n$, define $\Delta x_1 = x_2 - x_1$ and $\Delta x_n = x_n - x_{n-1}$, respectively. Let $\Delta x = \sum_{\ell=1}^{n} \Delta x_\ell$. If $x_\ell = \left(\frac{\ell}{n}\right) d$ for $\ell = 1, \ldots, n$, then $\Delta x_\ell = \frac{d}{n}$ for $\ell = 1, \ldots, n$, where d is the distance between the first and last samples of the data series. Note that $\Delta x = d$, and for simplicity, one may consider $d = 1$. Let $\mathbf{K} = \{-\frac{n}{2}, \ldots, \frac{n}{2} - 1\}$ if n is even and $\mathbf{K} = \{-\frac{n-1}{2}, \ldots, \frac{n-1}{2}\}$ if n is odd. The forward discrete summation of Fourier integral (after dividing by the summation range) for wavenumber $\omega_k \in \mathbf{K}$ is defined as

$$\hat{f}(\omega_k) = \frac{1}{\Delta x} \sum_{\ell=1}^{n} \Delta x_\ell f(x_\ell) e^{-2\pi i \omega_k x_\ell}, \tag{3.1}$$

and the contribution of ω_k to the inverse is defined as

$$f^k(x_\ell) = \hat{f}(\omega_k) e^{2\pi i \omega_k x_\ell}. \tag{3.2}$$

The wavenumber ω_k means ω_k cycles of sinusoids per unit distance. The (spatial) sampling rate is defined as the number of samples per unit distance. Note that for regularly sampled data series the Nyquist wavenumber[21] is half of the (spatial) sampling rate; however, the Nyquist wavenumber does not explicitly exist for inherently unequally spaced data series. Therefore, the maximum wavenumber for the analysis may be implicitly chosen to be less than half of the sampling rate (the aliasing may not occur for the wavenumber beyond this selection). If a data series is assumed to be regularly spaced with certain samples missing, then the Nyquist wavenumber may be half of the sampling rate for the regularly spaced data series (Craymer, 1998).

If the x_ℓ's are regularly spaced, then the estimation of the Fourier coefficient $\hat{f}(\omega_k)$ has no effect on the estimations of other Fourier coefficients because the sinusoids are orthogonal, i.e., the dot product between the sinusoids is zero (see Appendix A). However, the orthogonality condition fails for irregularly sampled data series resulting in spectral leakages. To mitigate these leakages Xu et al. (2005) proposed the ALFT that has the following three steps:

1. Compute all Fourier coefficients of the input data series using Equation (3.1).
2. Select the Fourier coefficient with the maximum energy.
3. Subtract the contribution of this coefficient (Equation (3.2)) from the input data series to obtain the residual data series. Use this residual data series as the new input and go to Step 1. Repeat this until the L2 norm of the residual data series becomes approximately zero (below a threshold ϑ).

Although the ALFT is fairly efficient and effective in practical applications, it has several shortcomings. The main shortcoming of the ALFT is that it only uses wavenumbers in a preselected set of wavenumbers (e.g., set $\mathbf{K}$) and does not search for the actual wavenumbers (real numbers) of components in a data series that may not be in the preselected set of wavenumbers. In other words, the wavenumber context of the data may not fall on one of the wavenumber samples used in set $\mathbf{K}$. Therefore, for a better regularization result, a denser set of wavenumbers may be selected in the ALFT (Xu et al., 2011). For a data series of size n with N wavenumbers, the ALFT requires the computation of N Fourier coefficients for each iteration, which has an order of nN floating operations, and it usually takes N iterations to converge. Thus, the total computational cost in the ALFT is in an order of nN^2 (Xu et al., 2005, 2010). Therefore, a denser set of wavenumbers will greatly increase the computational cost, and the actual wavenumbers of the components cannot still be accurately estimated. In practice, however, one may need fewer iterations when dealing with data that have by nature only a few different spectral components. Figure 3.1 shows a flowchart of the ALFT algorithm for data series regularization, where $|.|$ and $\|.\|$ denote the absolute value and L2 norm, respectively.

[21] The maximum wavenumber beyond which aliasing will occur is called the Nyquist wavenumber.

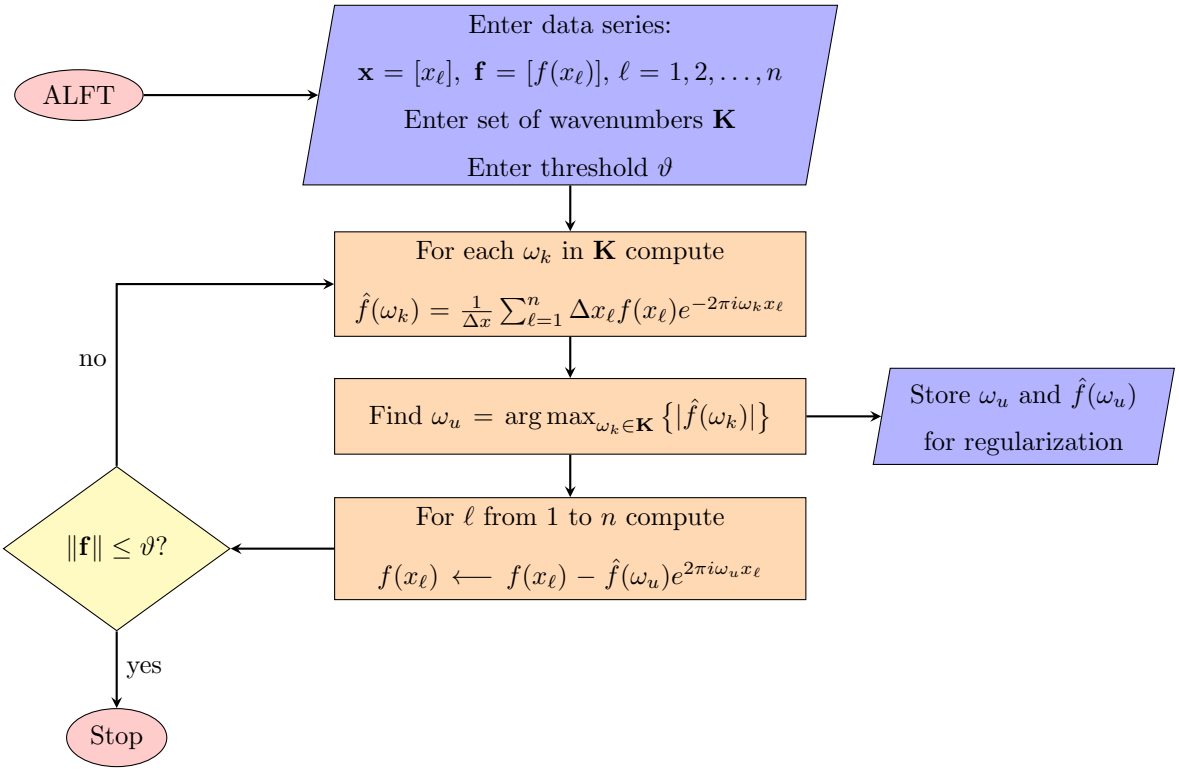

Figure 3.1: Flowchart of the ALFT algorithm.

3.2 Arbitrary Sampled Fourier Transform

Guo et al. (2015) proposed a method that is similar to the ALFT, but it can estimate the wavenumbers of the components in a data series more accurately. Their method, namely, the Arbitrary Sampled Fourier Transform (ASFT), is described in this section. Consider the following cost function,

$$\Lambda(\omega_k) = \sum_{\ell=1}^{n} \left| f(x_\ell) - \hat{f}(\omega_k) e^{2\pi i \omega_k x_\ell} \right|^2, \tag{3.3}$$

where $\hat{f}(\omega_k)$ is given by Equation (3.1), and $|.|$ denotes the absolute value. The ASFT has the following three steps (cf., Figure 3.2):

1. For each $\omega_k \in \mathbf{K}$, compute $\hat{f}(\omega_k)$ and find the wavenumber ω_u such that $\Lambda(\omega_u)$ is minimum.

2. Find wavenumber h in $I = [\omega_u - b, \omega_u + b]$ such that $\Lambda(h)$ is minimum. Since the wavenumbers in $\mathbf{K}$ are integer, one may choose $b = 0.5$.

3. Subtract the contribution $\hat{f}(h) e^{2\pi i h x_\ell}$ from the input data series to obtain the residual data series. Use this residual data series as the new input and go to Step 1. Continue this process until the L2 norm of the residual data series becomes approximately zero (below a threshold ϑ).

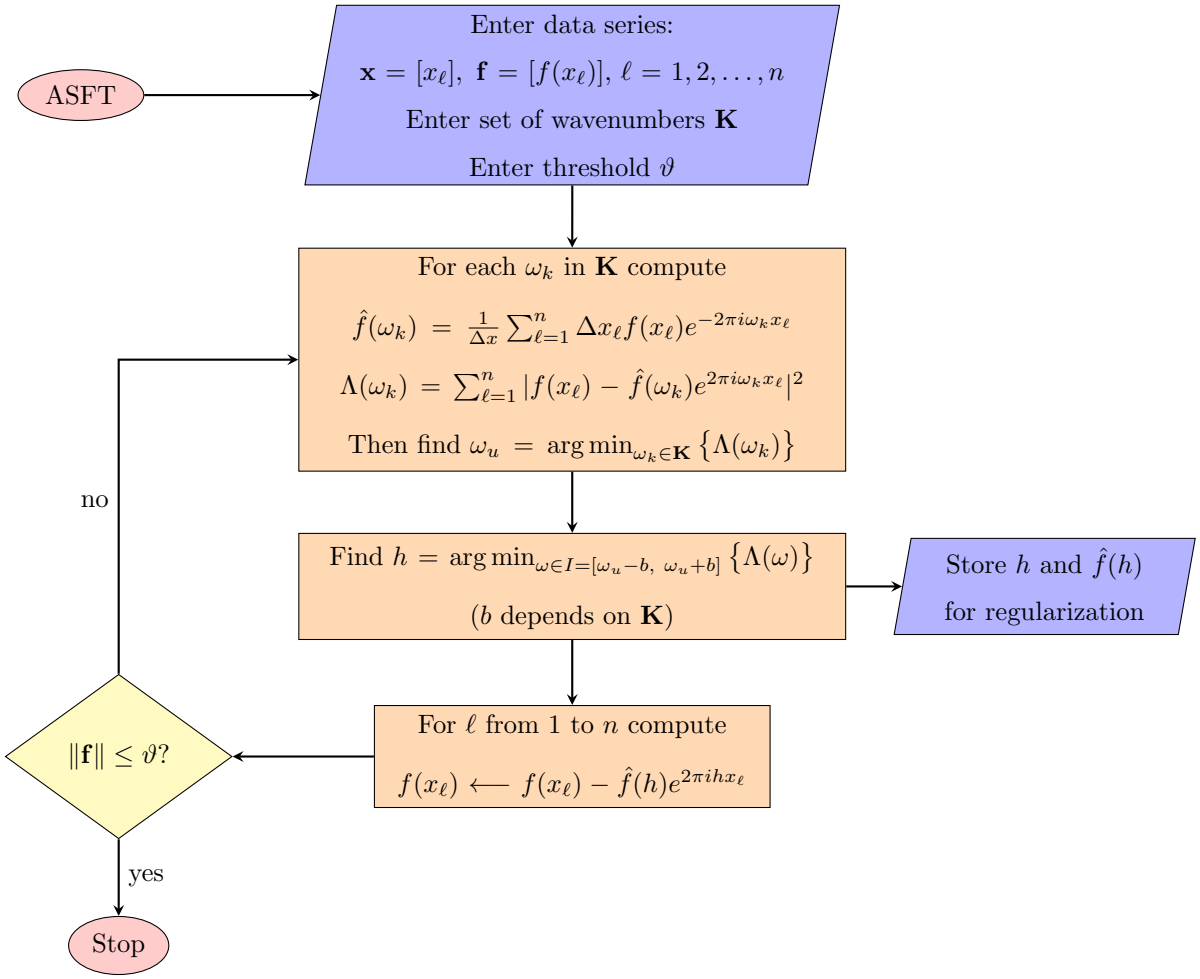

Figure 3.2: Flowchart of the ASFT algorithm.

Estimating h up to a chosen decimal place in Step 2 may simply be done by an appropriate partitioning of interval I as follows: For $b = 0.5$, choose wavenumber h_1 in set $\{\omega_u - 0.5 + 0.1j;\ j = 1, \ldots, 9\}$ such that $\Lambda(h_1)$ is minimum. Note that wavenumber h_1 is estimated up to one decimal place. Next choose wavenumber h_2 in set $\{h_1 - 0.1 + 0.01j;\ j = 1, \ldots 19\}$ such that $\Lambda(h_2)$ is minimum. Note that wavenumber h_2 is now estimated up to two decimal places. One may continue this process to estimate the wavenumber h up to any chosen decimal place. Estimating h up to four decimal places by using this technique requires $9 + 3(19) = 66$ times the use of Equation (3.3) that is computationally much faster than 9999 (equally spaced numbers in interval I) times the use of Equation (3.3), and it is also practically acceptable. For a fair comparison between the ASFT and ALLSSA, this partitioning technique is used in both methods. One may also apply other techniques, such as the conjugate gradient or the Broyden-Fletcher-Goldfarb-Shanno algorithm for Step 2 (Guo et al., 2015) or other robust minimizers, such as simulated annealing (Chen and Luk, 1999).

Note that wavenumbers estimated in the ASFT are real numbers obtained directly from the cost function minimization that is easier to interpret (Guo et al., 2015). One of the advantages of the ASFT over the ALFT is that the wavenumbers are estimated more accurately in each step. Therefore, if a threshold is defined for the L2 norm of the residual data series, then the ASFT is able to arrive at that threshold in a smaller number of iterations compared to the ALFT, resulting in more accuracy and less computational cost. In other words, for a data series of size n with n preselected wavenumbers, the number of ASFT floating operations is of the order of n^2m, where m is the number of iterations that is usually smaller than n by selecting an appropriate threshold. For an irregularly sampled data series with several components, although the ASFT can estimate the wavenumbers of the components more accurately than the ALFT, its spectrum still shows spectral leakages caused by the correlations between the sinusoids. Therefore, the wavenumbers of the data components cannot be estimated very accurately, and its spectrum contains many wavenumbers.

3.3 Least-Squares Spectral Analysis Revisited

Let $\mathbf{f} = [f(x_\ell)]$ be the column vector of n data points. Note that $x_\ell \in [0, 1]$, where '1' is the maximum distance. Let $\mathbf{\Phi}_k = \big[\cos(2\pi\omega_k x_\ell),\ \sin(2\pi\omega_k x_\ell)\big]$, $\ell = 1, \ldots, n$, be the $n \times 2$ matrix of a fixed wavenumber $\omega_k \in \mathbf{\Omega}$, where $\mathbf{\Omega}$ can be any set of real numbers. When a data series is regularly spaced, for the simultaneous estimation of the amplitudes and phases, $\mathbf{\Omega}$ may be selected as the set of all positive integers less than the Nyquist wavenumber (see Appendix A). Suppose that $\mathbf{f}$ has been derived from a population of (complex) random variables, and $\mathbf{C_f}$ is the (invertible and Hermitian) covariance matrix associated with $\mathbf{f}$ defined as

$$\mathbf{C_f} = \begin{bmatrix} \sigma^2_{f(x_1)} & \sigma_{f(x_1)f(x_2)} & \cdots & \sigma_{f(x_1)f(x_n)} \\ \sigma_{f(x_2)f(x_1)} & \sigma^2_{f(x_2)} & \cdots & \sigma_{f(x_2)f(x_n)} \\ \vdots & \vdots & & \vdots \\ \sigma_{f(x_n)f(x_1)} & \sigma_{f(x_n)f(x_2)} & \cdots & \sigma^2_{f(x_n)} \end{bmatrix}, \tag{3.4}$$

where $\sigma^2_{f(x_\ell)}$ is the variance of (complex) random variable $f(x_\ell)$, and $\sigma_{f(x_u)f(x_v)}$ is the covariance between two (complex) random variables $f(x_u)$ and $f(x_v)$. The covariance $\sigma_{f(x_u)f(x_v)}$ is a measure of the joint variability between random variables $f(x_u)$ and $f(x_v)$ defined as

$$\begin{aligned} \sigma_{f(x_u)f(x_v)} &= \mathrm{E}\Big[\Big(f(x_u) - \mathrm{E}\big[f(x_u)\big]\Big)\overline{\Big(f(x_v) - \mathrm{E}\big[f(x_v)\big]\Big)}\Big] \\ &= \mathrm{E}\Big[f(x_u)\overline{f(x_v)}\Big] - \mathrm{E}\Big[f(x_u)\Big]\mathrm{E}\Big[\overline{f(x_v)}\Big], \end{aligned} \tag{3.5}$$

where $\overline{f(x)}$ is the conjugate of $f(x)$, and $\mathrm{E}[X]$ is the expected value of (complex) random variable X (Eriksson et al., 2009).

Let $\mathbf{P} = {\mathbf{C_f}}^{-1}$. In many practical applications, $\mathbf{C_f}$ is approximately a diagonal matrix (the correlations between the data points are negligible), so one may treat $\mathbf{P}$ as a vector for computational effi-

ciency (Pagiatakis, 1999; Ghaderpour and Pagiatakis, 2017). In the LSSA, the cost function $\Psi(\mathbf{c}_k) = (\mathbf{f} - \mathbf{\Phi}_k\ \mathbf{c}_k)^{\mathrm{T}}\mathbf{P}(\mathbf{f} - \mathbf{\Phi}_k\ \mathbf{c}_k)$ is minimized to estimate $\mathbf{c}_k$, so one obtains $\hat{\mathbf{c}}_k = (\mathbf{\Phi}_k^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k)^{-1}\mathbf{\Phi}_k^{\mathrm{T}}\mathbf{P}\mathbf{f}$, where 'T' indicates transpose and conjugate transpose when $\mathbf{f}$ is complex. Note that $\hat{\mathbf{c}}_k$ is a 2×1 matrix (a column vector) whose elements are real or complex depending on $\mathbf{f}$. The (normalized) Least-Squares Spectrum (LSS) for $\omega_k \in \mathbf{\Omega}$ is defined as

$$s_0(\omega_k) = \frac{\mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k\hat{\mathbf{c}}_k}{\mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{f}}. \tag{3.6}$$

Note that $s_0(\omega_k) \in (0, 1)$, and it shows how much the sinusoids of wavenumber ω_k contribute to the entire data series. Since there is a very close relationship between the LSSA (simple case) and IMAP, a flowchart for the IMAP algorithm in matrix form is illustrated in Figure 3.3, that is an iterative algorithm to mitigate the spectral leakages in the LSS or Lomb spectrum (described in Chapter 2).

Now if there are constituents of known forms in the data series, then they may first be estimated and removed from the data series and then the residual data series may be studied. More precisely, let $\underline{\mathbf{\Phi}} = \left[\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_q\right]$ be the $n \times q$ matrix of constituents of known forms. The constituents can be represented by the column vector of all ones $[\mathbf{1}]$ and/or the column vector $[\mathbf{x}]$ containing the distances to explicitly account for any linear trend and/or the column vectors of the sinusoids of particular wavenumbers.

First minimize the cost function $\Psi_1(\underline{\mathbf{c}}) = (\mathbf{f} - \underline{\mathbf{\Phi}}\ \underline{\mathbf{c}})^{\mathrm{T}}\mathbf{P}(\mathbf{f} - \underline{\mathbf{\Phi}}\ \underline{\mathbf{c}})$ to estimate $\underline{\mathbf{c}}$ as

$$\hat{\underline{\mathbf{c}}} = \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f}, \tag{3.7}$$

where $\underline{\mathbf{N}} = \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\underline{\mathbf{\Phi}}$, and obtain the residual data series

$$\hat{\mathbf{g}} = \mathbf{f} - \underline{\mathbf{\Phi}}\ \hat{\underline{\mathbf{c}}}. \tag{3.8}$$

Next, minimize the cost function $\Psi_2(\underline{\mathbf{c}},\ \mathbf{c}_k) = (\mathbf{f} - \underline{\mathbf{\Phi}}\ \underline{\mathbf{c}} - \mathbf{\Phi}_k\ \mathbf{c}_k)^{\mathrm{T}}\mathbf{P}(\mathbf{f} - \underline{\mathbf{\Phi}}\ \underline{\mathbf{c}} - \mathbf{\Phi}_k\ \mathbf{c}_k)$ to estimate $\mathbf{c}_k$. Similar to Equation (6.10) (see also Appendix B), it can be seen that

$$\hat{\mathbf{c}}_k = \left(\mathbf{\Phi}_k^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k - \mathbf{\Phi}_k^{\mathrm{T}}\mathbf{P}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k\right)^{-1}\mathbf{\Phi}_k^{\mathrm{T}}\mathbf{P}\hat{\mathbf{g}}. \tag{3.9}$$

The term within the parentheses in Equation (3.9) is called the matrix of normal equations that is a real square matrix of order two (Craymer, 1998). If $\mathbf{f}$ is real, then $\hat{\mathbf{c}}_k$ is a real column vector of order two whose elements are the estimated amplitudes for cosine and sine basis functions. If $\mathbf{f}$ is complex, then $\hat{\mathbf{c}}_k$ is a complex column vector whose real and imaginary parts correspond to the estimated amplitudes for cosine and sine basis functions of the real and imaginary parts of $\mathbf{f}$, respectively. The normalized LSS of the residual data series for $\omega_k \in \mathbf{\Omega}$ is then defined by

$$s(\omega_k) = \frac{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k\hat{\mathbf{c}}_k}{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{g}}}. \tag{3.10}$$

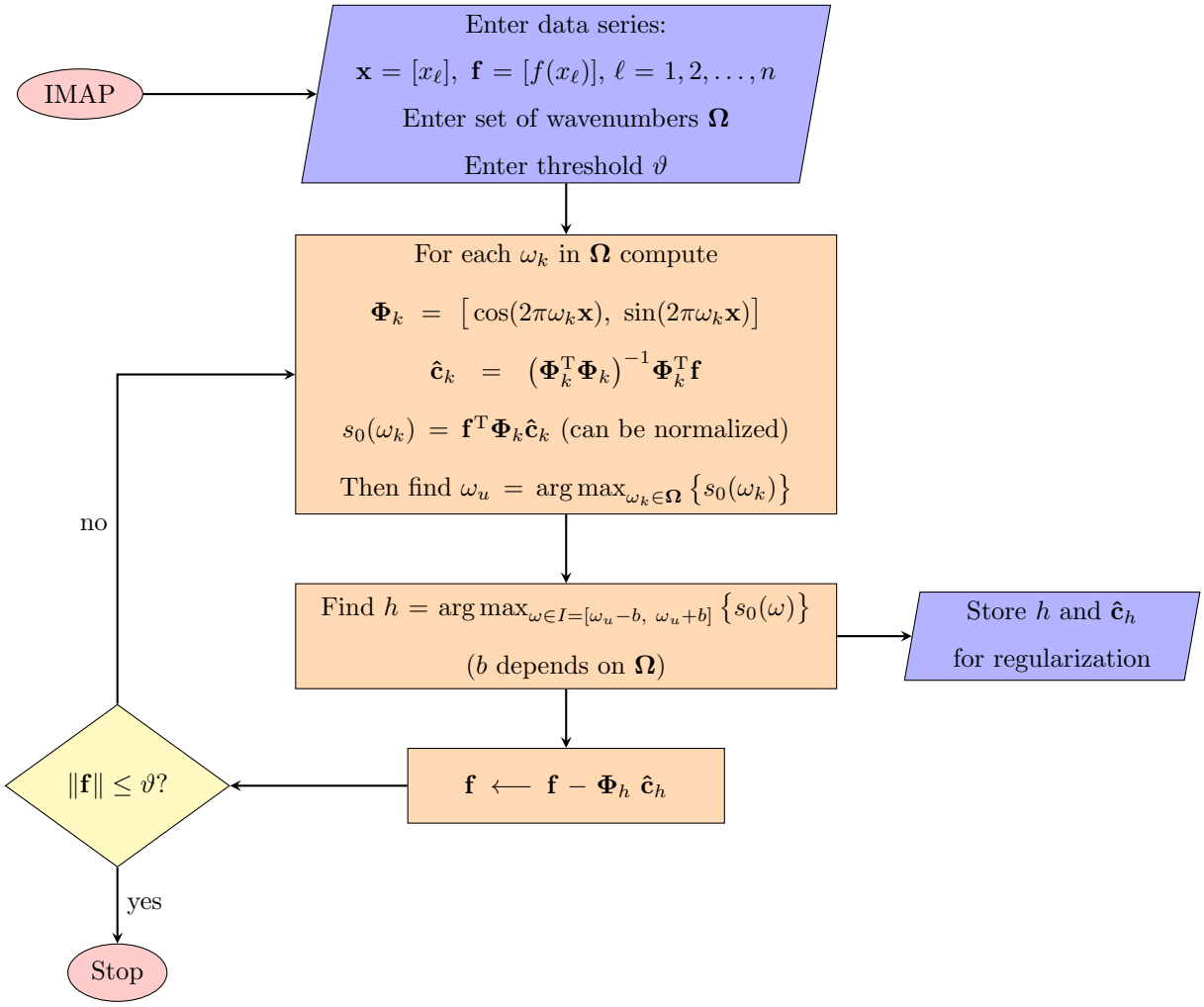

Figure 3.3: Flowchart of the IMAP algorithm in matrix form.

Note that $s(\omega_k) \in (0, 1)$ ("percentage variance" when multiplied by 100), and it shows how much the sinusoids of wavenumber ω_k contributes to the residual data series. Ideally, if $\underline{\mathbf{\Phi}}$ is an invertible square matrix of order n, then $\hat{\mathbf{g}} = \mathbf{0}$ (Craymer, 1998).

In order to demonstrate the estimated amplitude and phases simultaneously, the LSS is expressed in terms of wavenumber-amplitude by $\hat{\mathbf{c}}_k$ in Equation (3.9). In other words, for each $\omega_k \in \mathbf{\Omega}$, the square root of the sum of squares of the first and second elements of $\hat{\mathbf{c}}_k$ is calculated (separately for the real and imaginary parts of $\hat{\mathbf{c}}_k$ if $\mathbf{f}$ is complex). This way, the LSS can be compared with the Fourier spectrum expressed in terms of absolute values of the Fourier coefficients, described in a synthetic example in the next Chapter.

It is shown (Pagiatakis, 1999) that if $\mathbf{f}$ has been derived from a population of random variables following the multidimensional normal distribution, then the spectrum given by Equation (3.10) follows the beta

distribution with parameter values 1 and $(n-q-2)/2$, where q is the number of constituents of known forms, and n is the number of data points (see Appendix C). The critical value at $(1-\alpha)$ confidence level for the spectrum is then equal to $1-\alpha^{2/(n-q-2)}$, where α is the significance level (usually 0.05 or 0.01). The critical value is independent of the weight matrix and wavenumbers (Pagiatakis, 1999; Ghaderpour and Pagiatakis, 2017). Note that the 'critical value' in Ghaderpour et al. (2018b) is mistyped and should be 'significance level'.

3.4 Antileakage Least-Squares Spectral Analysis

One may apply the idea of maximum energy in the algorithm of the ALFT to develop a new algorithm for the LSSA that reduces the spectral leakages and consequently regularizes the irregularly sampled data series very accurately. Assume that $\mathbf{\Omega} = \{1, 2, \ldots, \eta\}$, where η may be chosen as the largest positive integer less than half of the sampling rate. Beginning with an initial $\underline{\mathbf{\Phi}}$ that may be selected as $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{x}]$, the algorithm for the ALLSSA has the following four steps:

1. Given $\underline{\mathbf{\Phi}}$, compute $\underline{\hat{\mathbf{c}}}$ and $\hat{\mathbf{g}}$ (Equations (3.7) and (3.8)).

2. For all $\omega_k \in \mathbf{\Omega}$, calculate $\hat{\mathbf{c}}_k$ (Equation (3.9)) and find ω_u in $\mathbf{\Omega}$ such that $s(\omega_u)$ (Equation (3.10)) is maximum.

3. Eliminate the cosine and sine basis functions of a wavenumber h in $\underline{\mathbf{\Phi}}$ such that $|h-\omega_u| < b$ (if it exists). Assuming that the difference between any two consecutive actual wavenumbers of the components in the data series is greater than one, one may choose $b = 0.5$ to resolve the peaks and avoid singularity of the matrix of normal equations.

4. Find h in $I = [\omega_u - b, \omega_u + b]$ such that $s(h)$ (Equation (3.10)) is maximum. Concatenate (add) the cosine and sine column vectors of wavenumber h to $\underline{\mathbf{\Phi}}$ and then go to Step 1. Finding h up to a chosen decimal place can be done by an appropriate partitioning of I. Terminate the process if $s(h)$ is no longer significant at certain confidence level (usually 95% or 99%).

Note that in the LSSA, set $\mathbf{\Omega}$ does not need to contain only integers, and so b in Step 3 can be variable. Therefore, at the computational step (practical step), one can define set $\mathbf{\Omega}$ to include as many wavenumbers as the length of the series allows so that there is no singularity. In other words, the user can have a very high resolution spectrum with b as small as possible, and therefore the peak(s) with the highest power/energy can be easily identified in the spectrum. Since there is no attempt to estimate two waves with very close wavenumbers, there is no danger of reaching any singularity. However, in this chapter and the next chapter, set $\mathbf{\Omega}$ contains only positive wavenumbers to compare the result of the ALLSSA directly to the results of the ALFT and ASFT.

When considering the weight matrix $\mathbf{P}$, the method is called the weighted ALLSSA. Removing the constituent of a particular wavenumber with maximum percentage variance (power/energy) from the data series reduces the spectral leakages in the residual spectrum. On the other hand, the elimination of the basis functions of a wavenumber h in Step 3 is crucial in the ALLSSA because considering the correlation between the sinusoids of different wavenumbers results in a more accurate h in the next step. This correlation will be taken into account as the column dimension of $\underline{\mathbf{\Phi}}$ increases in the process. The L2 norm of residual data series $\hat{\mathbf{g}}$ approaches zero rapidly because the wavenumbers are accurately estimated recursively that also prevents increasing the column dimension of $\underline{\mathbf{\Phi}}$ in many practical applications especially when applying the confidence level. The AntiLeakage Least-Squares Spectrum (ALLSS) in terms of wavenumber-amplitude is defined by $\hat{\underline{\mathbf{c}}}$ in Step 1 after the process is terminated.

Similar to the discussion in the previous section, $\hat{\underline{\mathbf{c}}}$ is a real or complex column vector (depending on $\mathbf{f}$) whose elements are the estimated intercept and slope of the linear trend (if the initial $\underline{\mathbf{\Phi}}$ is $[\mathbf{1}, \mathbf{x}]$) and the estimated amplitudes for the sine and cosine basis functions of the estimated wavenumbers. This $\hat{\underline{\mathbf{c}}}$ is used for the regularization of a stationary data series (reconstruction of the data over any given regularly spaced series) that will also result in attenuating the random noise present in the data series down to certain confidence level (all the statistically significant peaks at certain confidence level are used for regularization). For non-stationary data series, one may also treat the final residual data series as a new input data series then perform the ALLSSA again (by applying the confidence level) and repeat this procedure until the L2 norm of the final residual data series approaches zero[22] (below a threshold). Note that the actual wavenumbers of the data series will be better approximated if the initial $\underline{\mathbf{\Phi}}$ at least contains the column vector of all ones (Foster, 1996; Ghaderpour and Pagiatakis, 2017). Figure 3.4 shows the flowchart of ALLSSA algorithm.

The computational complexity of the ALLSSA depends on the estimation of $\hat{\underline{\mathbf{c}}}$ and the matrix of normal equations in each iteration. For a data series of size n with η wavenumbers (e.g., $\eta = \frac{n}{2} - 1$ if n is even and $\eta = \frac{n-1}{2}$ if n is odd), the process has an order of $n\eta m$ floating operations, and m mainly depends on the number of significant spectral peaks. When applying the confidence level, the total number of the significant peaks is usually small, making m small, so the computational efficiency can be much better than the ALFT and ASFT for stationary data series.

In the ALLSSA, the sinusoidal basis functions of different wavenumbers, corresponding to the largest statistically significant peaks in the spectrum, are added to the columns of design matrix $\underline{\mathbf{\Phi}}$ in an iterative manner. Therefore, the wavenumbers will be more accurately estimated in the next step by considering the correlation among the columns. The accuracy of estimated wavenumbers significantly reduces the number of iterations and results in more accurate regularization compared to the IMAP (see example below).

[22]It is recommended that the researcher continues the analysis until the end, when the residual series has a white spectrum and the normality test shows that the residual series is random.

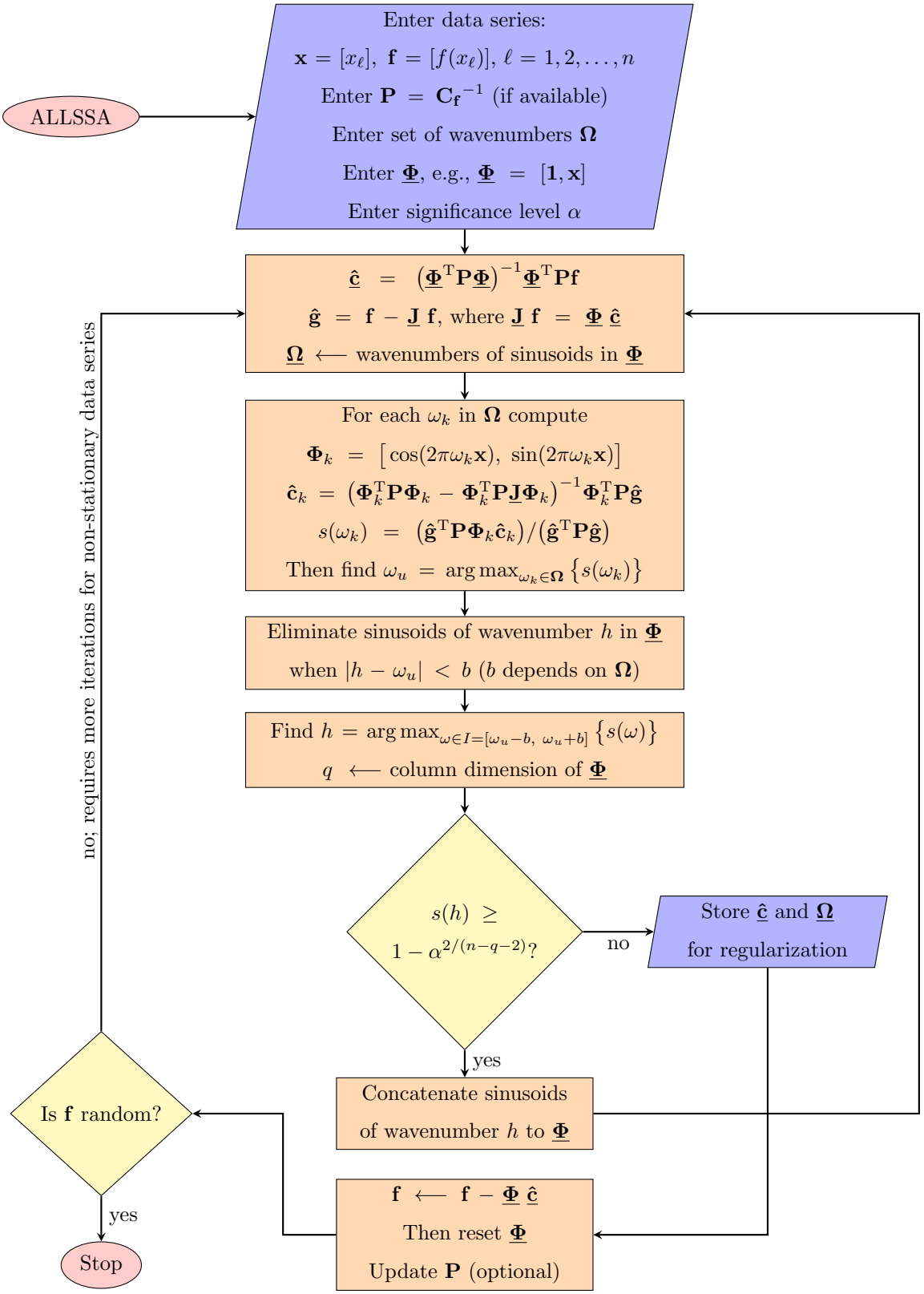

Figure 3.4: Flowchart of the ALLSSA algorithm.

3.5 Synthetic Data Series Example

In this example, the ALFT, ASFT, and ALLSSA are used to regularize the data series of size 128 given by Equation (13) in Xu et al. (2005) with an additional sine wave and a linear trend

$$f(x_\ell) = 5\sin(25.6x_\ell) + 2.5\sin(128x_\ell + 1) + \sqrt{3}\sin(140x_\ell) + \sqrt{2} + \pi x_\ell, \tag{3.11}$$
$$\simeq 5\sin(2\pi \cdot 4.0744\ x_\ell) + 2.5\sin(2\pi \cdot 20.3718\ x_\ell + 1) + \sqrt{3}\sin(2\pi \cdot 22.2817\ x_\ell) + \sqrt{2} + \pi x_\ell,$$

where x_ℓ $(l = 1, 2, 3, \ldots, 128)$ is a random number in $[0, 1]$ generated by the MATLAB command 'rand'. All the x_ℓ's are sorted in ascending order (see Table 3.1).

The data series obtained from the values in Table 3.1 is shown in Figure 3.5a. The goal is to reconstruct the data series on a series with regular spacing 1/128 (see the ideal reconstruction in Figure 3.5b). This data

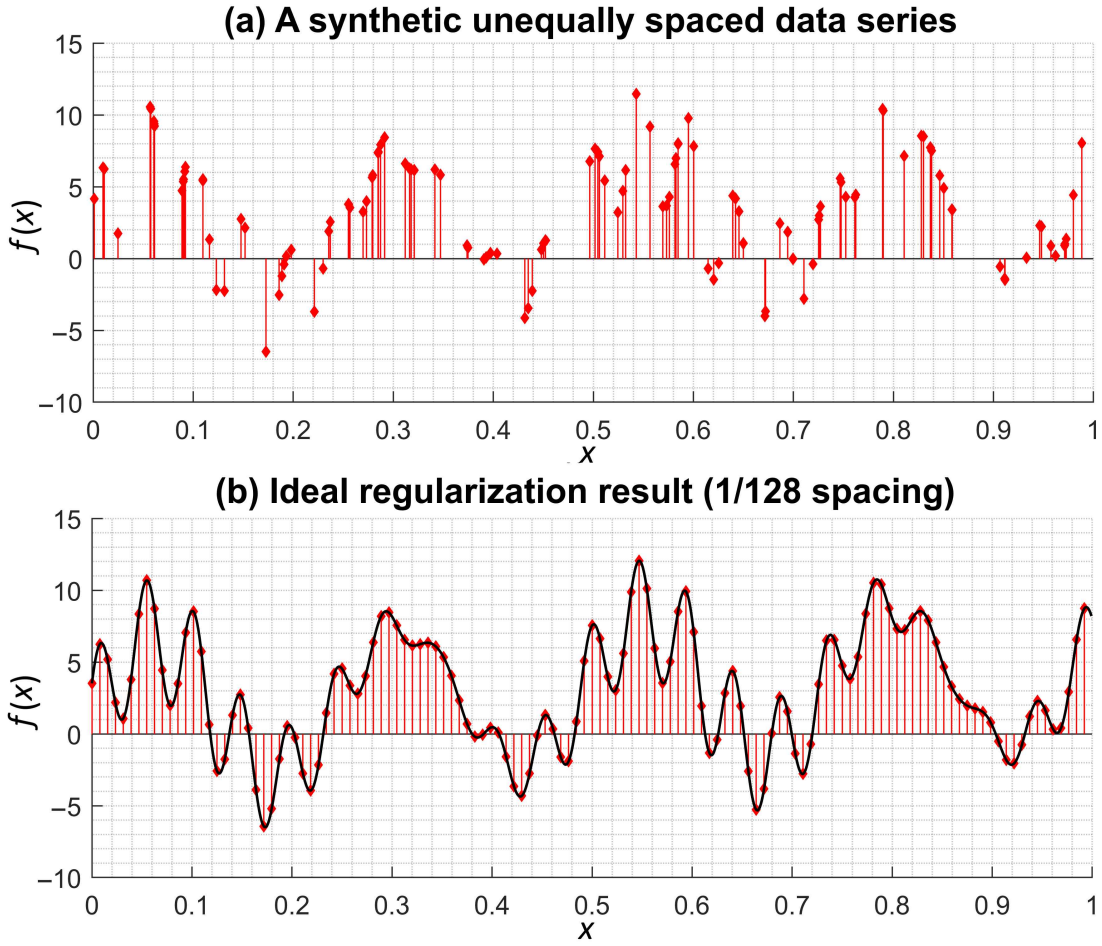

Figure 3.5: (a) The irregularly sampled data series (red diamonds) given by Equation (3.11), and (b) the ideal reconstructed data series on a series with regular spacing 1/128. Note that this is a deterministic series with no noise.

Table 3.1: The values of the data series given by Equation (3.11).

ℓ	x_ℓ	$f(x_\ell)$	ℓ	x_ℓ	$f(x_\ell)$	ℓ	x_ℓ	$f(x_\ell)$	ℓ	x_ℓ	$f(x_\ell)$
1	0.0012	4.1459	33	0.2554	3.7746	65	0.5114	5.4213	97	0.7257	2.9864
2	0.0098	6.3235	34	0.2568	3.5381	66	0.5246	3.206	98	0.7269	3.6399
3	0.011	6.2455	35	0.2699	3.2506	67	0.5294	4.7075	99	0.7466	5.59
4	0.0249	1.733	36	0.2732	3.9738	68	0.5323	6.158	100	0.7477	5.3266
5	0.0569	10.5501	37	0.2789	5.6463	69	0.5431	11.4665	101	0.7522	4.2909
6	0.0577	10.4264	38	0.2794	5.8081	70	0.5566	9.1851	102	0.7615	4.2617
7	0.0606	9.5492	39	0.2848	7.3645	71	0.5694	3.6382	103	0.7624	4.4445
8	0.061	9.38	40	0.2851	7.4331	72	0.5733	3.663	104	0.789	10.4141
9	0.0614	9.2229	41	0.2874	7.9229	73	0.5761	4.2949	105	0.7897	10.3048
10	0.0887	4.7172	42	0.2915	8.4397	74	0.5761	4.2957	106	0.8106	7.1337
11	0.09	5.3539	43	0.3119	6.6108	75	0.5816	6.5607	107	0.8277	8.5478
12	0.0903	5.5206	44	0.3164	6.2615	76	0.5825	6.9787	108	0.83	8.5057
13	0.0915	6.0724	45	0.3178	6.2017	77	0.5846	7.9723	109	0.8371	7.7299
14	0.0921	6.3641	46	0.3209	6.1423	78	0.5847	8.0061	110	0.8372	7.7104
15	0.1097	5.5084	47	0.3416	6.203	79	0.595	9.7637	111	0.8383	7.5168
16	0.1099	5.4367	48	0.3473	5.8137	80	0.6003	7.8217	112	0.8464	5.7735
17	0.1161	1.33	49	0.3738	0.9117	81	0.6146	-0.6803	113	0.8504	4.899
18	0.1232	-2.1801	50	0.3747	0.75	82	0.6203	-1.4524	114	0.8586	3.415
19	0.1311	-2.2501	51	0.3907	-0.06	83	0.6253	-0.323	115	0.8588	3.3873
20	0.1478	2.7437	52	0.3927	0.0796	84	0.6393	4.3755	116	0.9064	-0.5541
21	0.1519	2.1289	53	0.3971	0.3735	85	0.6423	4.1792	117	0.9111	-1.3885
22	0.1726	-6.4868	54	0.4038	0.3453	86	0.6455	3.2762	118	0.9116	-1.4815
23	0.1858	-2.5312	55	0.4315	-4.1425	87	0.65	1.0609	119	0.9329	0.0564
24	0.1888	-1.2075	56	0.435	-3.4527	88	0.6714	-4.0102	120	0.9462	2.2945
25	0.191	-0.4057	57	0.439	-2.2456	89	0.6723	-3.6663	121	0.9481	2.2368
26	0.193	0.1571	58	0.448	0.6196	90	0.6866	2.4392	122	0.9573	0.8809
27	0.1982	0.5938	59	0.4503	1.0558	91	0.6944	1.8507	123	0.9622	0.1906
28	0.2213	-3.6923	60	0.4522	1.2574	92	0.6996	-0.0216	124	0.9711	0.906
29	0.2299	-0.6893	61	0.4963	6.7645	93	0.6996	-0.0283	125	0.9715	1.0231
30	0.2354	1.8974	62	0.5017	7.6344	94	0.7104	-2.7925	126	0.9727	1.3772
31	0.2369	2.5595	63	0.5044	7.421	95	0.7194	-0.3853	127	0.9798	4.4223
32	0.2553	3.7928	64	0.506	7.1136	96	0.7252	2.71	128	0.9884	8.0317

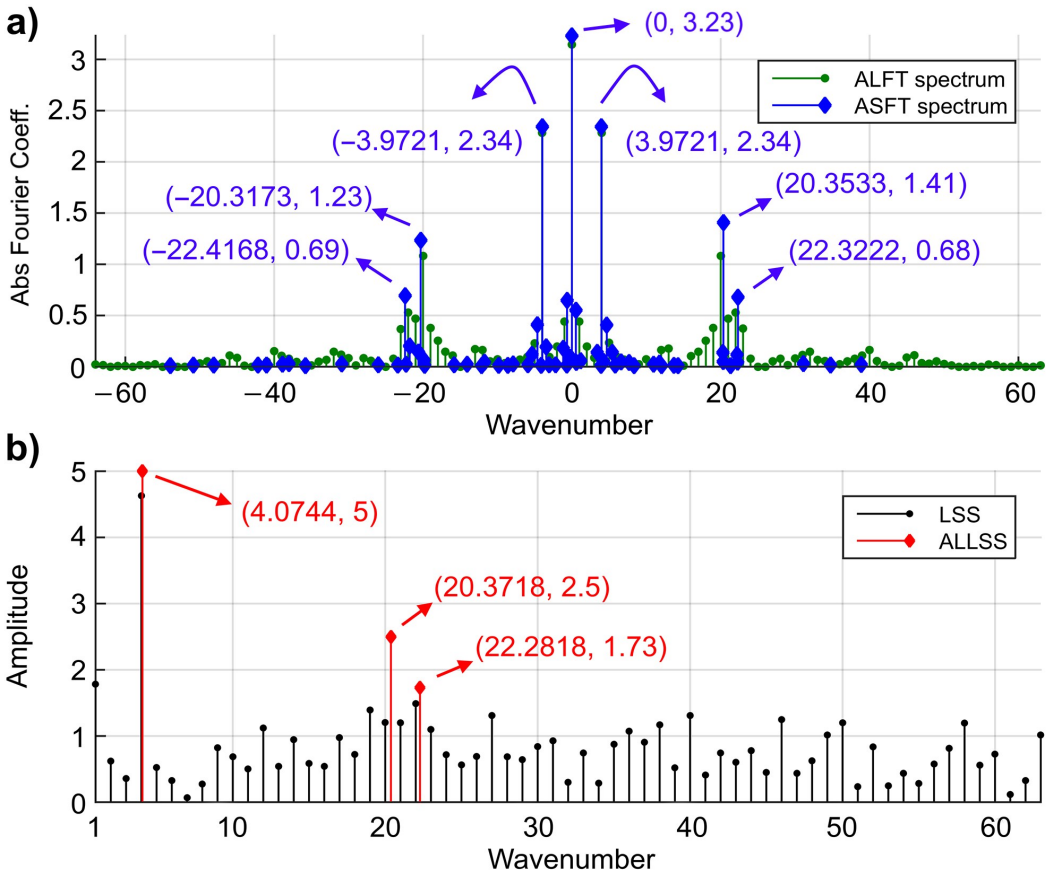

Figure 3.6: The spectra of the irregularly sampled data series given by Equation (3.11). (a) The ALFT (green) and ASFT (blue) spectra, and (b) the LSS (black) and ALLSS (red). Note that the wavenumbers in the ALFT spectrum (green) and LSS (black) are integers in this figure, and they are used as preselected wavenumbers in the ASFT and ALLSSA. The spectral leakages can be observed in the LSS, ALFT and ASFT spectra, resulting in more number of iterations and less accuracy.

series is inherently unequally spaced, and so the Nyquist wavenumber does not explicitly exist. Therefore, there is no restriction for the selection of maximum wavenumber in this example. However, since the goal is to regularize this data series to a series with regular spacing 1/128 (the Nyquist wavenumber of this regular series is 64), the initial set of wavenumbers is chosen as $\mathbf{\Omega} = \{1, 2, 3, \ldots, 63\}$ for the LSSA and ALLSSA, and the wavenumbers are estimated up to 4 decimal places. Note that the data series has three wavenumber components with wavenumbers $25.6/2\pi \simeq 4.0743665$, $128/2\pi \simeq 20.3718327$, and $140/2\pi \simeq 22.2816920$, rounded to 7 decimal places.

For the ALFT and ASFT, the initial set of wavenumbers is chosen as $\mathbf{K} = \{-64, -63, \ldots, 63\}$. Their spectra are illustrated with green and blue dots in Figure 3.6a, respectively. In the ALFT and ASFT, the fitting of sinusoids of various wavenumbers (including zero) to the data series can implicitly account for the linear trend. Figure 3.6a clearly shows the presence of the spectral leakages in both the ALFT and ASFT

Table 3.2: The result of wavenumber estimation of the ALLSSA algorithm on the irregularly sampled data series given by Equation (3.11) after each iteration.

Iteration number	1st wavenumber	2nd wavenumber	3rd wavenumber	Norm of residual
First	4.0384			23.5040
Second	4.0384	20.3057		13.3170
Third	4.0384	20.3057	22.2947	3.5079
Fourth	**4.0748**	20.3057	22.2947	2.2501
Fifth	4.0748	**20.3708**	22.2947	0.3068
Sixth	4.0748	20.3708	**22.2818**	0.0468
Seventh	4.0748	**20.3718**	22.2818	0.0319
Eighth (final)	**4.0744**	20.3718	22.2818	0.0036

(with threshold $\vartheta = 0.5$ for the L2 norm of residuals). However, the actual wavenumbers of the data series in the ASFT are estimated more accurately than the ALFT (see the numbers shown in blue).

For the LSSA and ALLSSA, the initial $\underline{\mathbf{\Phi}}$ is selected as $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{x}]$ to explicitly account for the linear trend present in the data series, where '$\mathbf{1}$' and '$\mathbf{x}$' are the column vectors of all ones and the position vector, respectively. Since there is no covariance matrix associated with the data series, the square weight matrix $\mathbf{P}$ is not considered in the calculation. The LSS (black) and the ALLSS (red) in terms of wavenumber-amplitude are illustrated in Figure 3.6b. The LSS shows one strong peak at wavenumber 4 that is the closest value to the actual value 4.0743665; however, the other two signal peaks are buried in the leakage spectrum. On the other hand, it can be seen from the ALLSS that only the actual signal peaks are estimated very accurately. Note that the signal peaks are estimated up to only four decimal places, and this accuracy is chosen for the synthetic and field data examples in the next chapter.

The preselected wavenumbers are all integers, and the LSS accurately detects the peak at the lowest wavenumber. Note that the resolution of the LSS spectrum is not enough to resolve the other two peaks at higher wavenumbers. A denser set of wavenumbers can estimate the signal wavenumbers more accurately. Typically, in the LSSA, for higher wavenumbers, one may select denser wavenumbers, so the peaks at higher wavenumbers that are naturally of much smaller bandwidth can be selected nicely (Pagiatakis, 1999).

Table 3.2 shows the result of the ALLSSA. In the first iteration, the wavenumber 4.0384 is estimated that is approximately 0.036 different from its actual value 4.0743665. This is a shortcoming of the LSSA (out-of-context) caused by the presence of other constituents in the data series. In the second and third iterations, the other two wavenumbers are estimated that by removing their corresponding components from the data series simultaneously, the first wavenumber is better approximated and so forth (see the highlighted numbers in Table 3.2).

After all the signal wavenumbers with their amplitudes are estimated as well as the estimated slope and intercept for the linear trend in the ALLSSA, they can be used to reconstruct the data series on any desired regular series (regularize). In this example, the desired series is regular with 1/128 spacing, and

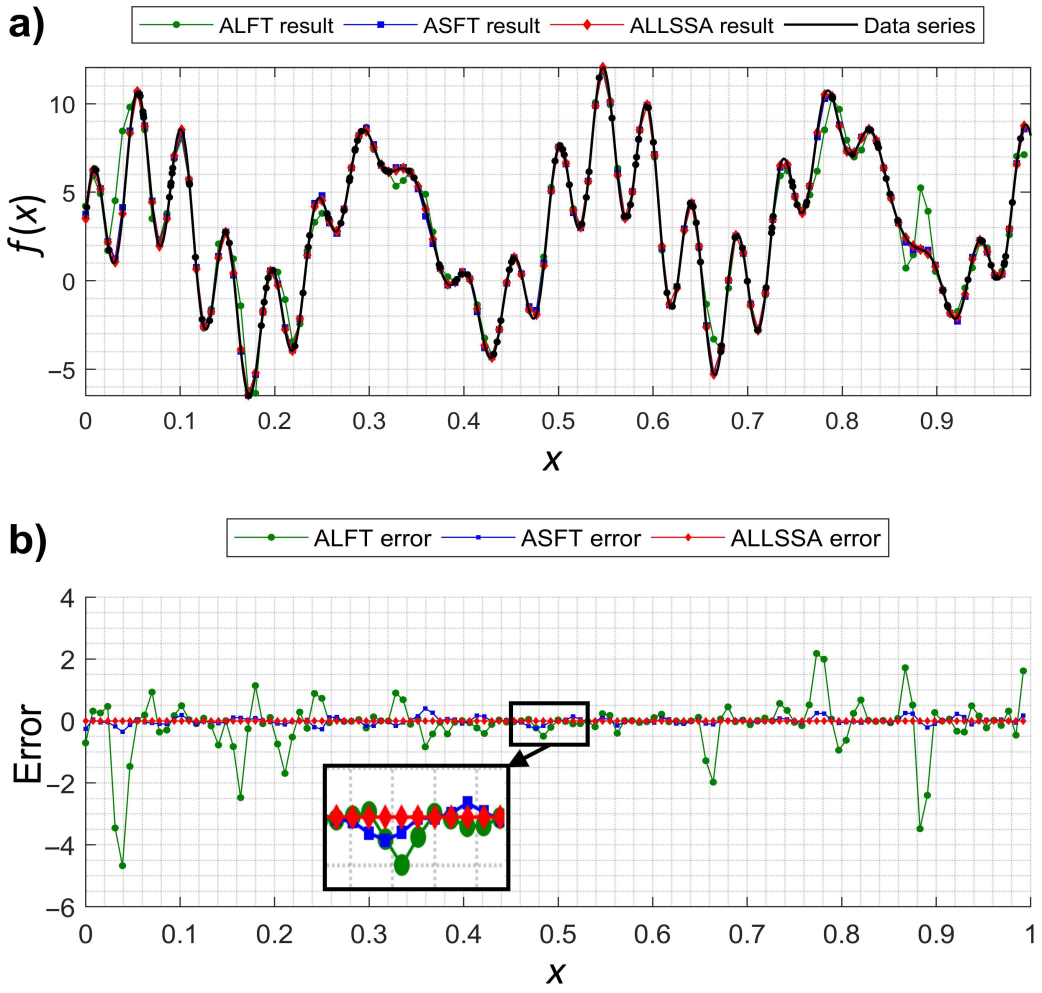

Figure 3.7: (a) The irregularly sampled data series (black) given by Equation (3.11) and its regularizations using the ALFT (green), ASFT (blue), ALLSSA (red), and (b) the difference between the ideal data series values and the regularized results (residual).

the ideal data series and the regularization results using the three methods are illustrated in Figure 3.7a. The differences between each ideal data point and the estimated one for the three methods are also shown in Figure 3.7b. The L2 norm of the ideal regularized data series is 57.382. The L2 norm of the difference between the ideal data series and the regularized data series for the ALFT, ASFT, and ALLSSA are 9.864, 1.26, and 0.004, respectively. For the ALLSSA, the estimated slope and intercept of the linear trend and the estimated amplitudes and phases of the three wavenumbers, shown in the last row of Table 3.2, are used for the regularization. The result of the ALLSSA regularization for the data series without the linear trend is also more accurate than the one for the ALFT and ASFT regularization (not shown here). Note that when a smaller threshold is selected for the ASFT, the computational time significantly increases, and its spectrum

will have much more wavenumbers with very low amplitudes (absolute value of Fourier coefficients), and it still cannot reach the accuracy of the ALLSSA.

To compare the results of the ALLSSA with the IMAP, the same data series is used for regularization as illustrated in Figure 3.5a, and the results are shown in Figure 3.8. To consider the linear trend using IMAP, the position vector and column vector of all ones are fitted to the data series and after estimating its slope and intercept, it is removed from the data series. It is interesting to note that the estimated slope and intercept are -0.7737 and 4.011, respectively, after this process (see the blue line in Figure 3.8b). However, when the slope and intercept are being estimated simultaneously with other constituents using the ALLSSA, the result is 3.1415 for the slope and 1.4143 for the intercept, approximately the same as the actual slope and intercept in Equation (3.11) (see the red line in Figure 3.8b). In the LSS, the slope and intercept are estimated simultaneously with sinusoids of a particular wavenumber (one at a time). The IMAP after several iterations tries to compensate for the error in estimating the linear trend by using wavenumbers sufficiently close to zero (see the IMAP peaks close to zero wavenumber in Figure 3.8a).

Since the correlations between the constituents are not simultaneously considered in the IMAP, its spectrum still shows leakages that increase the number of iterations and cause regularization inaccuracy (Figure 3.8c). However, the leakages in the IMAP spectrum are less than the ones in the ALFT and ASFT spectra shown in Figure 3.6a because the correlations between the sine and cosine basis functions are considered in the IMAP for each wavenumber. The L2 norm of the residual series in the IMAP is 0.83, indicating better regularization result compared to the ALFT (9.864) and ASFT (1.26).

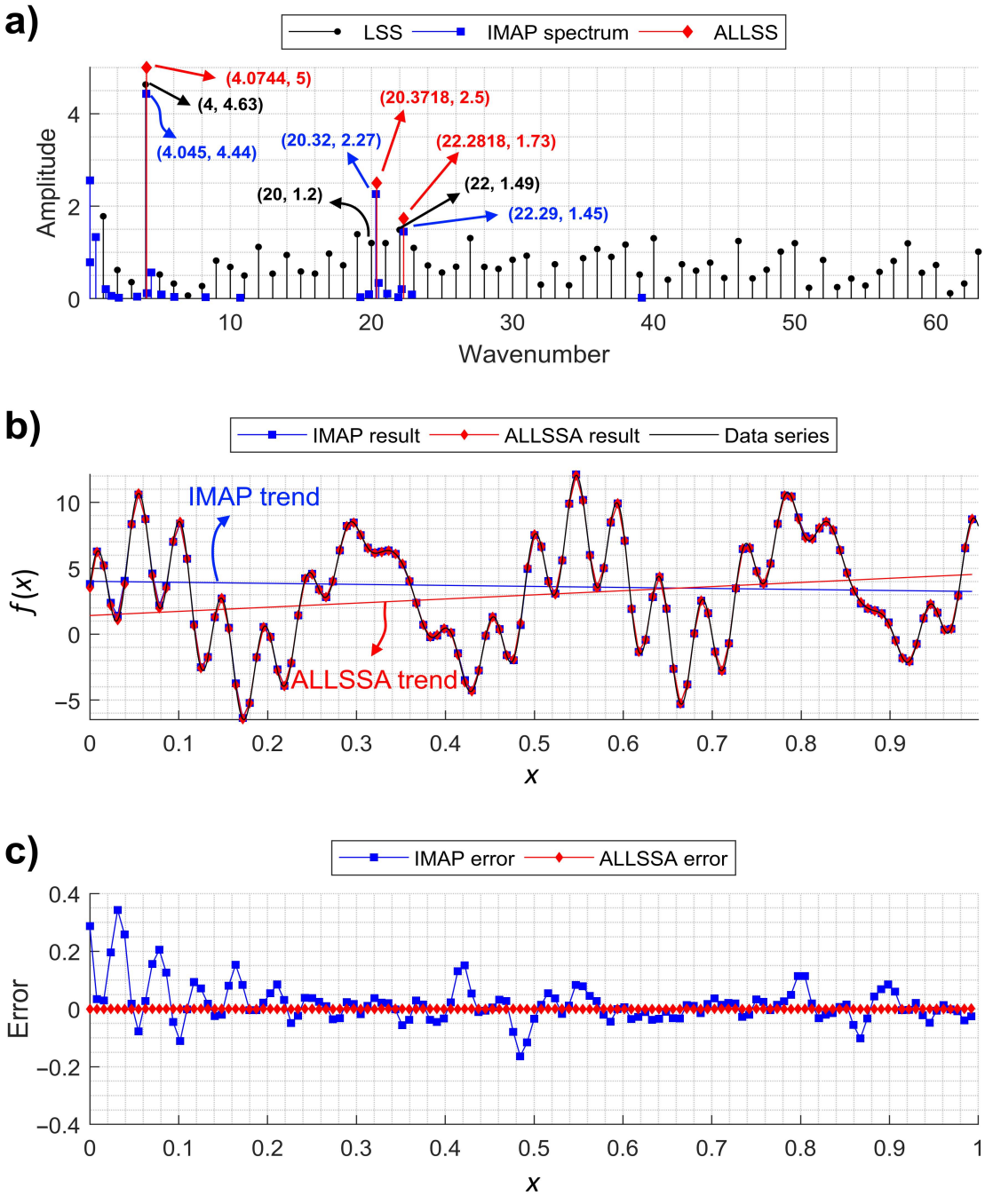

Figure 3.8: (a) The IMAP spectrum (blue squares) and the LSS and ALLSS (same as the ones in Figure 3.6b), (b) the regularization results using the IMAP and ALLSSA, and (c) the difference between the ideal data series and the regularized data series using the IMAP and ALLSSA.

3.6 Summary

I have proposed an antileakage method, the ALLSSA, to mitigate the spectral leakages in the least-square spectrum. I have shown the outstanding performance of the ALLSSA and compared it with the ALFT, ASFT, and the IMAP. The ALLSSA is a robust method that performs very well for analyzing data series and can be applied to non-stationary data series with more complex structure by an appropriate segmentation of the data series. The antileakage spectrum can estimate the exact location of the significant spectral peaks and consequently can be used to regularize unequally spaced data series. The ALLSSA is applied to analyze data series; however, it can be applied to analyze and regularize unequally spaced time series as well, where the spectrum is in term of frequency rather than wavenumber.

Simultaneous estimation of wavenumbers and amplitudes with constituents of known forms significantly reduces the number of iterations and improves the regularization results compared to the ALFT, ASFT, and the IMAP. Note that when the constituents of known forms and weight matrices are not considered in the iterative process, the ALLSSA is in fact the IMAP.

4 Applications of the Antileakage Least-Squares Spectral Analysis in Seismology

In this chapter, the ALLSSA is applied to regularize irregularly spaced synthetic 2D seismic data and 2D field seismic data given in time-distance (t-x) domain that could be a crossline (inline) section of a 3D seismic data or a 2D seismic data acquired from a 2D seismic survey. The results are compared with the ALFT, ASFT, and nonlinear shaping regularization. Sections 4.2 and 4.3 are unabridged versions of sections "SYNTHETIC DATA EXAMPLE" and "FIELD DATA APPLICATIONS" in Ghaderpour et al. (2018b), respectively.

4.1 Introduction to Marine Seismic Data Processing

In seismology, the seismic refraction[23] method uses the refraction of seismic waves on geologic layer and rock/soil to characterize the geologic structure. This method is frequently used in engineering geology, exploration geophysics, etc. For instance, geophysicists apply this method to find oil associated with salt domes. Seismic waves generated by a vibrator or air gun (source) travel through the Earth, and their energies partially reflect and partially refract through the interface. Geophysicists reconstruct the pathways of the waves by measuring the traveling time from the source to several receivers and the velocity of seismic waves depending on the medium in which they are traveling. In marine seismic data acquisition, the receivers are hydrophones[24] that convert sound waves[25] (pressure waves) into electrical signals and record the pressure amplitude. In other words, hydrophone sensors measure the pressure wave field[26]. The response of each receiver to a single shot from the source is known as a "trace", and it is recorded onto a data storage (generating a time series). As the vessel travels the shot location moves, and the process will be repeated.

[23]Seismic refraction is a geological principle based on the law of refraction (Snell's law).

[24]A hydrophone is a type of microphone designed to measure the underwater sound.

[25]Sound waves tend to travel faster in denser media. The speed of sound in the air is about 344 m/s (20 degrees Celsius) and in the sea water is about 1531 m/s (20 degrees Celsius).

[26]The voltage generated by the hydrophone is proportional to the amplitude of the pressure wave as it passes by the hydrophone (Hurrell, 2004).

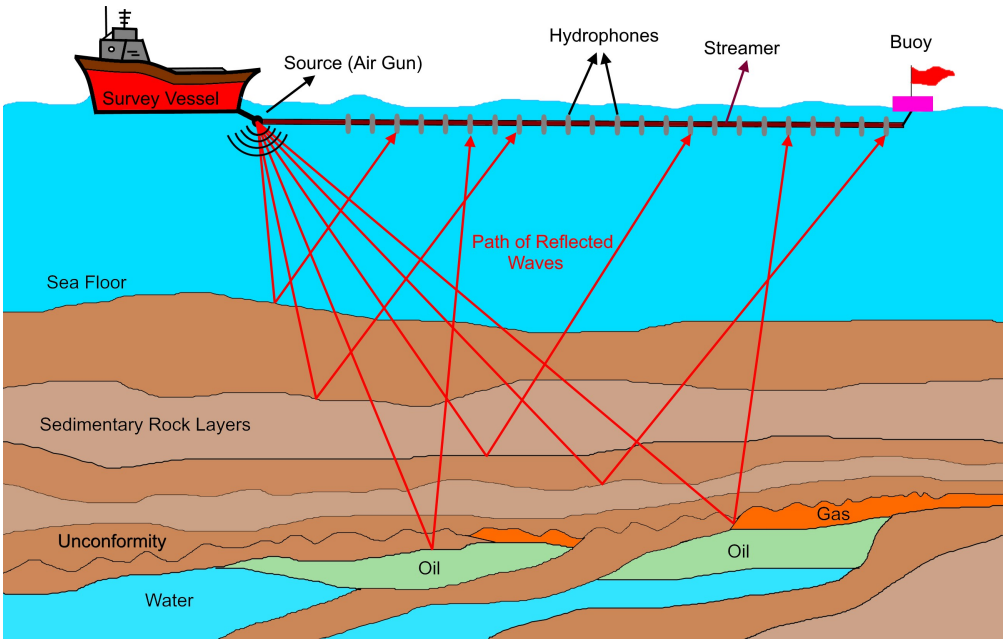

Figure 4.1: 2D visualization of seismic survey.

Depending on the area of interest, the seismic images may contain many linear and/or curved events[27].

In traditional marine seismic survey, a vessel carries one or more cables (streamers) containing a series of hydrophones separated by a constant distance. The cables are deployed beneath the surface of the water and are positioned at a set distance away from the vessel and can be as long as six to eight kilometres (Figure 4.1). The 2D surveys use only one cable and one source; whereas, the 3D surveys usually use six or eight or more cables and sources. The inexpensive 2D surveys usually take place for general understanding of the regional geologic structure before drilling or performing 3D surveys. In 3D surveys, the term "inline" refers to the direction in which the receiver cables are deployed (the shooting direction), and the term "crossline" refers to the direction that is perpendicular to the orientation of the receiver cables (parallel to the water surface). In other words, the inline is the direction of the vessel movement (parallel to receivers), and the crossline is perpendicular to vessel movement. The reader is referred to Scales (1997); Crawley (2000); Biondi (2006) for more details on seismic data acquisition and processing.

[27]In seismology the term "dip" refers to a slop angle of the foot-wall block measured clockwise from horizontal (0 to 90 degrees). A linear event is an event whose dip is constant. A curved event is an event whose dip gradually changes when observed at different times and offsets. A strongly curved event is an event whose dip changes over short ranges of offset (see also Chapter 2 for more detail).

Marine seismic data sets are usually irregularly sampled along spatial directions because of cable feathering[28], editing bad traces, economy (acquiring more data is expensive and difficult because of current marine-acquisition technology), etc. They are more coarsely sampled along the crossline direction than the inline direction. Regularly sampled seismic data are required for various purposes including wave equation migration, seismic inversion, amplitude versus azimuth or offset analyses and surface-related multiple elimination (Weglein et al., 1997; Dragoset et al., 2010). An ideal way of obtaining regularly sampled data is to acquire more data. In 3D seismic surveys, acquiring more data may be done by increasing the number of streamers with a wider azimuth range, but this is expensive and difficult to achieve, and so data regularization (reconstruction of data over any desired regular grids) becomes an important task.

The area spanned by a 3D seismic image is subdivided into small grids called stacking bins. Each trace in a 3D seismic data is positioned so that it passes vertically through the midpoint of each bin. The 3D seismic data can be viewed in terms of time-crossline-inline (t-x-y). The simplest 3D data regularization[29] is to bin the data and disregard the true data acquisition locations in the bin. This binning technique may cause artifacts that contaminate the final migrated images (Xu et al., 2005). Similar to the ALFT and ASFT, the true locations of the acquired data (using their latitudes and longitudes) may be used in the extension of the ALLSSA for regularization of 3D seismic data or higher orders (5D, inline-crossline-azimuth-offset-frequency) seismic data (Trad, 2009; Xu et al., 2010) that is subject to future work.

It is customary to transform each trace from the time domain to the frequency domain using the Fast Fourier Transform (FFT) (Spitz, 1991; Abma and Claerbout, 1995; Xu et al., 2010; Guo et al., 2015). Then for each frequency, generate a data series[30] whose data points, located at the trace locations, are the Fourier coefficients of that frequency (a temporal frequency slice). For irregularly spaced seismic data, each temporal frequency slice is an irregularly sampled data series, and the ALFT or ASFT or IMAP or ALLSSA method may be used to regularize the data series, and then each trace can be transformed back from the frequency domain to the time domain by taking the inverse FFT. Note that in this chapter and also the next chapter, all the analyses are performed on the temporal frequency slices.

In order to visualize the aliasing effects, the seismic data may be transformed from the t-x domain to frequency-wavenumber[31] (f-k) domain using the two-dimensional discrete Fourier transform (DFT). More precisely, suppose that $\{f(t_j, x_\ell) : \quad 1 \le j \le m,\ 1 \le \ell \le n\}$ contains the values of the seismic data in the

[28]The lateral displacement of the cables is often called cable feathering. In the 3D surveys, due to sideways drift, the cables are not parallel to each other, and so the hydrophones are not regularly spaced.

[29]A popular method of 3D data regularization is using the least-squares collocation that is computationally slow for typical seismic data applications (Fomel, 2001, Chapter 2).

[30]In this chapter, the data series is obtained from 2D seismic images. In 3D seismic images, an irregular set of data points over irregular grids may be obtained.

[31]Frequency is the number of cycles per unit time, and wavenumber is the number of cycles per unit distance.

t-x domain, where $t_j = j/m$ and $x_\ell = \ell/n$. The two-dimensional DFT for each pair $(k_1, k_2) \in \mathbf{K_1} \times \mathbf{K_2}$ is defined as

$$\hat{f}(k_1, k_2) = \sum_{j=1}^{m} \sum_{\ell=1}^{n} f(t_j, x_\ell) e^{-2\pi i k_2 x_\ell} e^{-2\pi i k_1 t_j}, \tag{4.1}$$

where $\mathbf{K_1} = \{-\frac{m}{2}, \ldots, \frac{m}{2} - 1\}$ and $\mathbf{K_2} = \{-\frac{n}{2}, \ldots, \frac{n}{2} - 1\}$ if m and n are even and $\mathbf{K_1} = \{-\frac{m-1}{2}, \ldots, \frac{m-1}{2}\}$ and $\mathbf{K_2} = \{-\frac{n-1}{2}, \ldots, \frac{n-1}{2}\}$ if m and n are odd. Since the two dimensional DFT is defined for data sampled in regular grids, in the following examples, when a trace at spatial location x_ℓ is removed from the seismic image, the value of $f(t_j, x_\ell)$ in Equation (4.1) is set to zero for all j's. The f-k spectrum is defined by calculating the absolute values of $\hat{f}(k_1, k_2)$ for all $k_1 \in \mathbf{K_1}$ and $k_2 \in \mathbf{K_2}$.

4.2 Synthetic Seismic Data Example

Four linear seismic events are simulated by using an Ormsby wavelet that is a common type of synthetic wavelet in reflection seismology. The wavelet is defined by the sinc function as

$$A(t) = \frac{f_4^2 \operatorname{sinc}^2(f_4 t) - f_3^2 \operatorname{sinc}^2(f_3 t)}{f_4 - f_3} - \frac{f_2^2 \operatorname{sinc}^2(f_2 t) - f_1^2 \operatorname{sinc}^2(f_1 t)}{f_2 - f_1}. \tag{4.2}$$

Ormsby wavelets have several sidelobes (see Figure 4.2a), unlike Ricker wavelets which only have two, one on either side. Let $f_1 = 5\pi$, $f_2 = 10\pi$, $f_3 = 20\pi$, and $f_4 = 30\pi$ Hz. The wavelet defines a trapezoidal shape in the frequency spectrum with low-cut, low-pass, high-pass, and high-cut frequencies f_1, f_2, f_3, and f_4, respectively (see Figure 4.2b). The wavelet shown in Figure 4.2a may be considered as a pressure wave with the unit of kiloPascal (kPa).

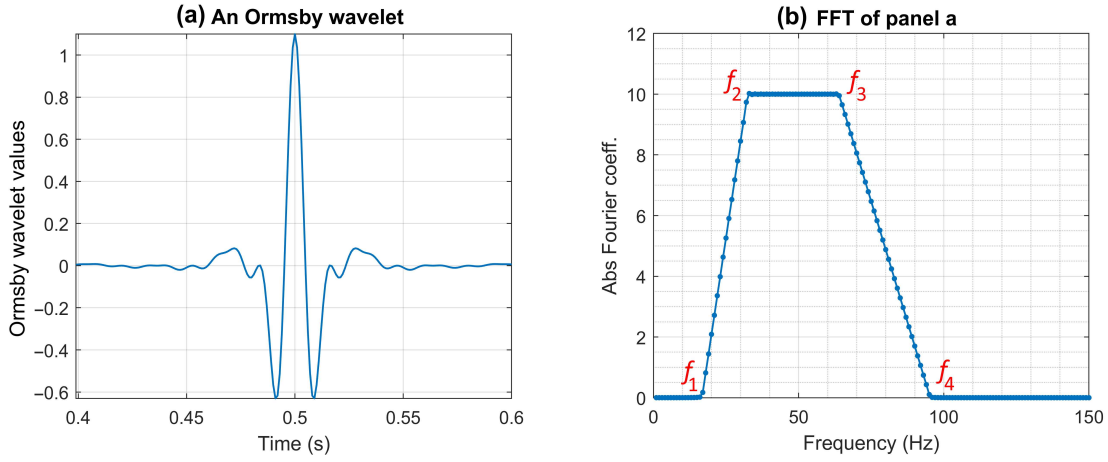

Figure 4.2: (a) An Ormsby wavelet given by Equation (4.2) multiplied by factor 0.01 with low-cut, low-pass, high-pass, and high-cut frequencies $f_1 = 5\pi$, $f_2 = 10\pi$, $f_3 = 20\pi$, and $f_4 = 30\pi$ Hz, respectively, and (b) the spectrum of panel a. The time sampling rate is 1000 samples per second.

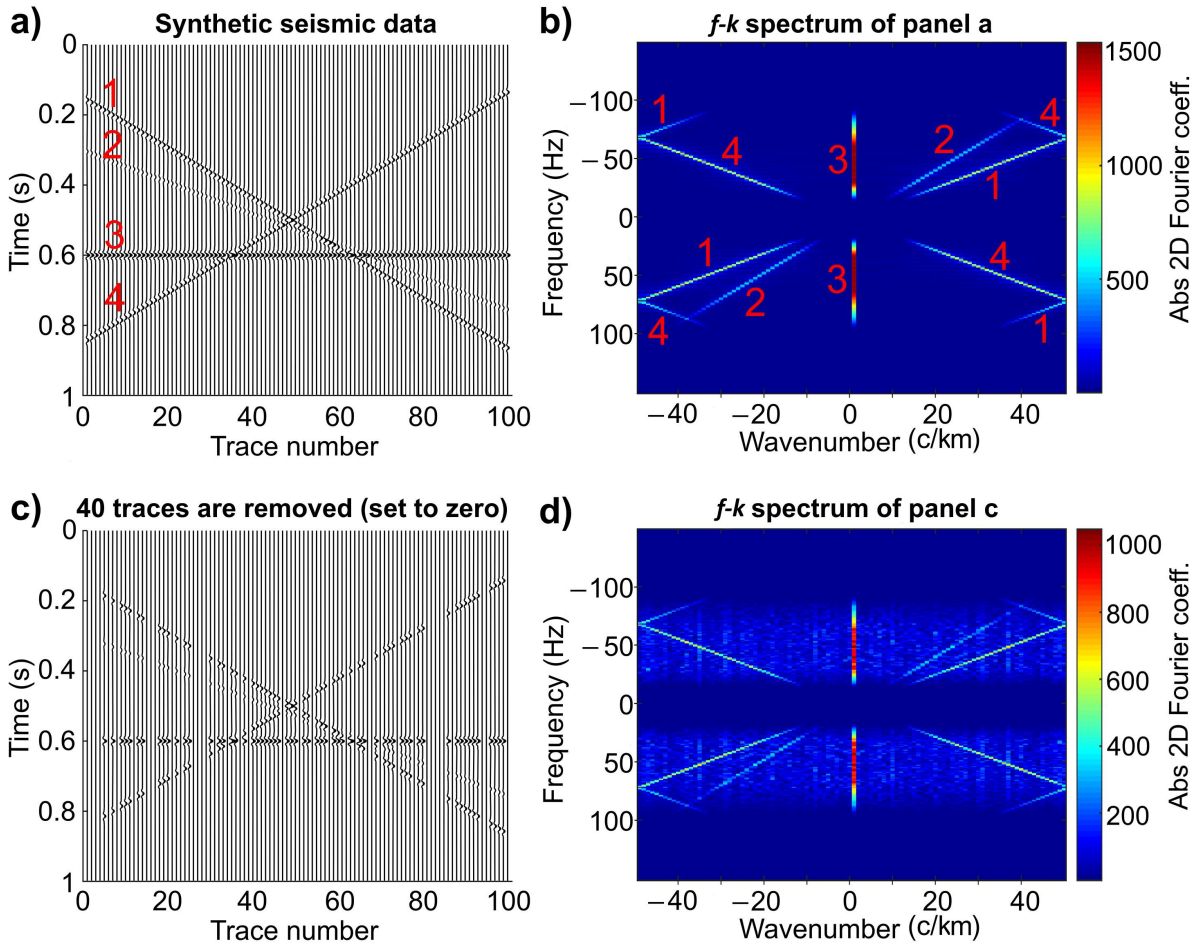

Figure 4.3: A synthetic seismic data containing 4 linear events. (a) Original data with 100 traces, (b) the *f-k* spectrum of panel a, (c) 40 traces are randomly removed, and (d) the *f-k* spectrum of panel c (the removed traces are set to zero to generate the *f-k* spectrum). In the *f-k* spectra (right panels), the vertical axes (in Hz) correspond to the vertical axes (in second) in the left panels (one sample per millisecond). Similarly, the horizontal axes (in c/km) in the right panels correspond to the horizontal axes in the left panels (10 m trace spacing), so for example, the vertical reddish lines in the *f-k* spectra at wavenumber zero correspond to the horizontal event. The spectral peaks corresponding to the linear events are shown with numbers 1 to 4 in panels a and b.

The linear events are designed to have different amplitudes and dips. There are 100 traces that are regularly spaced, out of which 40 traces are randomly removed (see Figure 4.3). The time sampling rate is 1000 samples per second, and assuming that the unit distance is in kilometre, the distance sampling rate is 100 samples per kilometre (10m trace spacing). Therefore, the Nyquist frequency is 500 Hz, and the Nyquist wavenumber is 50 cycles per kilometre (c/km) (40 traces are removed). The frequency-wavenumber (*f-k*) spectra of Figure 4.3a and c are shown in Figure 4.3b and d, respectively.

The ALFT, ASFT, and ALLSSA are applied to the edited data shown in Figure 4.3c (with the 40 traces removed) and to regularize the temporal frequency slices. For a fair comparison, the same threshold ($\vartheta = 0.05$

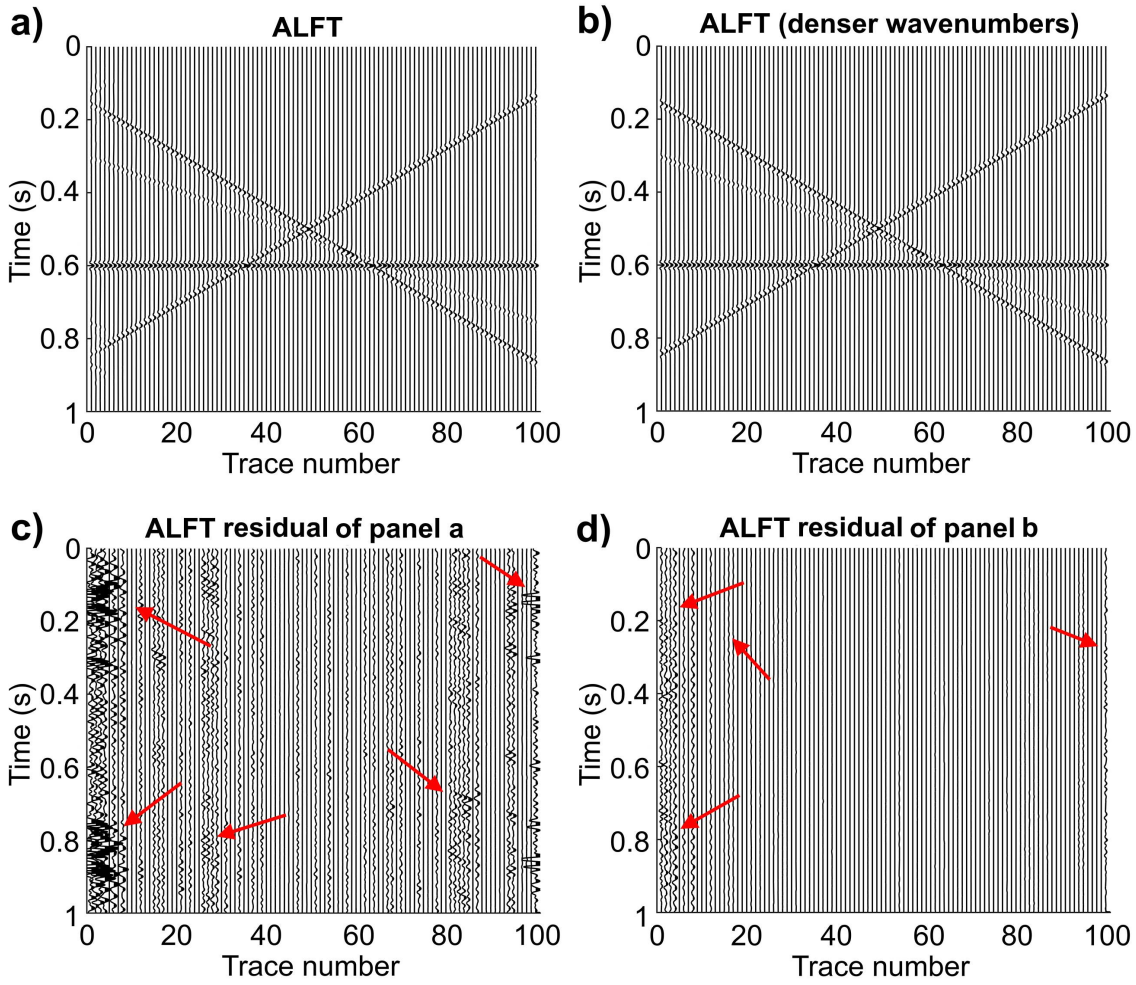

Figure 4.4: The regularization results using the ALFT with (a) 100 wavenumbers, (b) a denser set of wavenumbers (200 wavenumbers), and (c and d) the residuals of panels a and b, respectively (amplified 30 times). Arrows show places with significant difference, and panel d shows less difference compared to panel c.

kPa) for the ALFT, ASFT, and ALLSSA is used. The difference between the original synthetic data and the regularized data (the residual) is amplified by a factor of 30, and so the *f-k* spectra of the residuals will correspond to the amplified residuals for all the methods. The same partitioning for wavenumbers is used for both the ASFT and ALLSSA to estimate the wavenumbers to a resolution of 4 decimal places. All the methods are programmed in MATLAB language and run on a computer with two cores (processors). The sum of L2 norms of differences between the original and interpolated traces are denoted by 'ε'. The sum of the L2 norms of all traces in Figure 4.3a is 650.86. To have a better measure for the total residual, a percentage residual is also calculated, that is $\varepsilon/6.5086$ %.

First, $\mathbf{K} = \{-50, -49, \ldots, 48, 49\}$ is chosen for the ALFT, and its regularization result and residual (ripples) are shown in Figure 4.4a and c, respectively. The computational time is 7s and $\varepsilon = 37.29$ (5.73%).

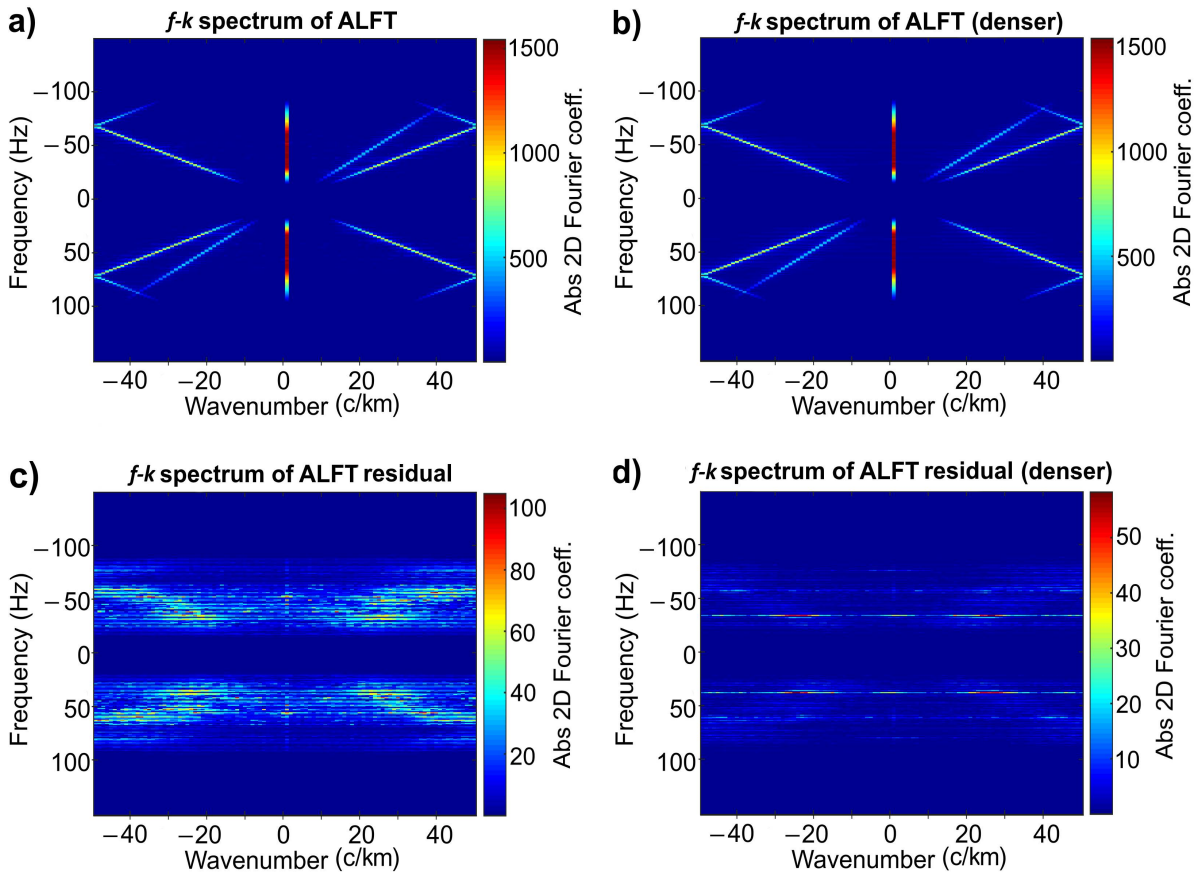

Figure 4.5: The *f-k* spectra corresponding to (a) Figure 4.4a, (b) Figure 4.4b, (c) Figure 4.4c, and (d) Figure 4.4d.

Figure 4.4a and c show that the missing traces are not very well constructed especially for the extrapolation (the traces toward the edges). This is mainly due to the lack of adequate resolution of the wavenumbers. Next the number of wavenumbers is doubled, i.e., $\mathbf{K} = \{-50, -49.5, -49, \ldots, 48.5, 49, 49.5\}$, and the ALFT regularization result and residual are shown in Figure 4.4b and d, respectively. The computational time increased to 14s, but $\varepsilon = 6.76$ (1.04%). One can observe that the traces are constructed more accurately than the previous choice for $\mathbf{K}$. The *f-k* spectra corresponding to Figure 4.4 are shown in Figure 4.5. Comparing Figure 4.5c to 4.5d, one can observe that the fuzziness of the spectrum is reduced when a denser set of wavenumbers is selected for the ALFT. Next, the preselected set of wavenumbers in the ASFT is chosen as $\mathbf{K} = \{-50, -49, \ldots, 48, 49\}$ and its regularization result and residual are illustrated in Figure 4.6a and c, respectively. The computational time is 13s, and $\varepsilon = 2.6$ (0.4%), that is more accurate and computationally faster than the ALFT with a denser set of wavenumbers.

For the ALLSSA, the preselected set of wavenumbers is chosen as $\mathbf{\Omega} = \{1, 2, \ldots, 49\}$. To ensure that all the real wavenumbers are considered up to the Nyquist wavenumber 50, the partitioning for wavenumber 49 is adjusted to cover the range $[49, 50)$, and we choose $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{x}]$ to explicitly account for any linear trend.

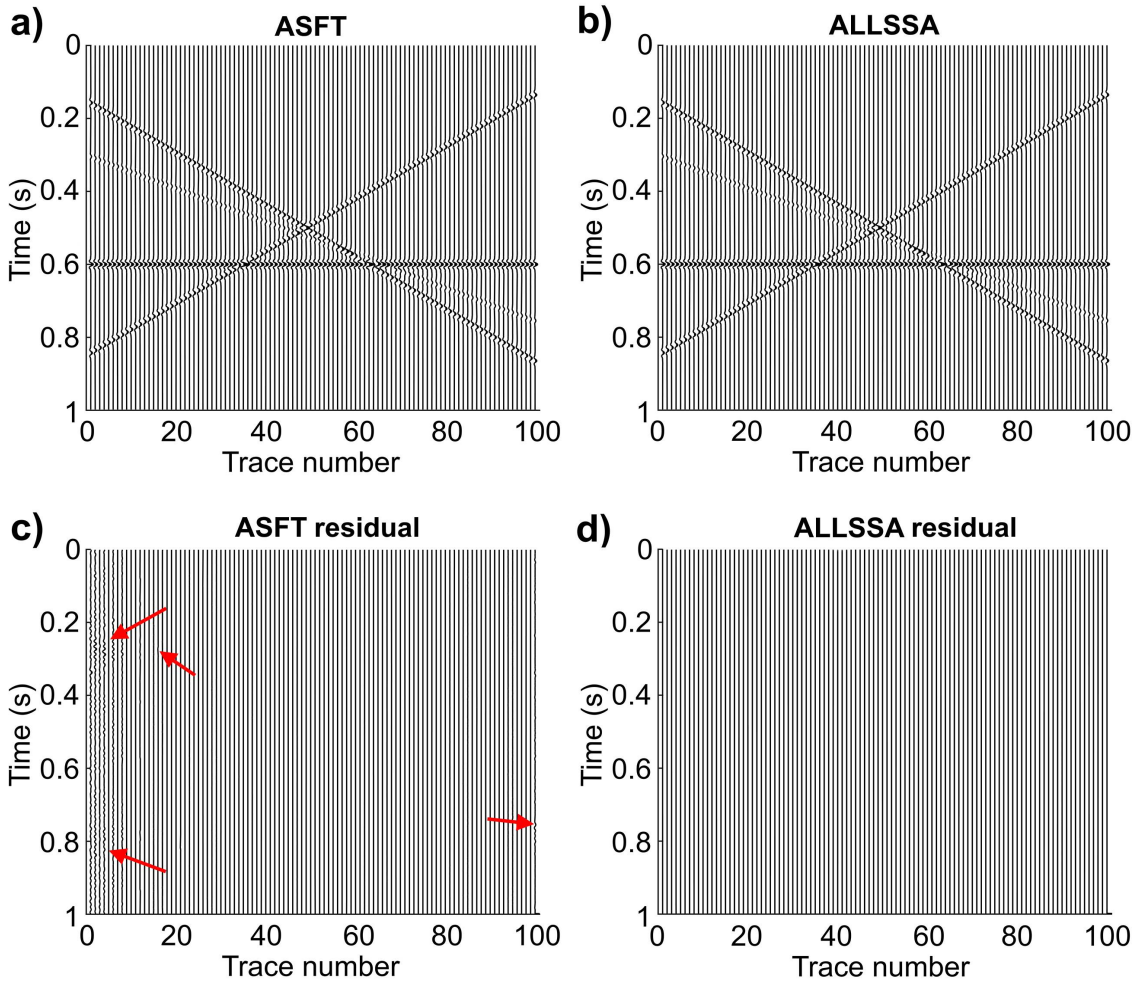

Figure 4.6: The regularization results using the (a) ASFT, (b) ALLSSA, and (c and d) the ASFT and ALLSSA residuals, respectively (amplified 30 times). Arrows in panel c show places with residual that are less than the ALFT residual. Panel d shows almost zero residual.

If one chooses $\underline{\mathbf{\Phi}} = [\mathbf{1}]$, then the partitioning for wavenumber 1 should be adjusted to cover the range $(0, 1]$, so the sinusoids of lower wavenumbers can approximate the trends. The ALLSSA regularization result and residual are shown in Figure 4.6b and d, respectively. The computational time is 3s, and $\varepsilon = 0.3$ (0.05%). Figure 4.6b and d show that the ALLSSA is able to fill in the missing traces much more accurately (0.05% vs 1.04% and 0.4%) and computationally faster than the ALFT and ASFT. The *f-k* spectra corresponding to Figure 4.6 are shown in Figure 4.7.

In order to show how the ALLSSA statistically attenuates the random noise, a distinct Gaussian random noise is added to each trace in Figure 4.3b. Therefore, approximately 20% of the synthetic data is random noise. For display purposes, only trace 1 for the extrapolation, trace 15 for the interpolation, and trace 92 for noise attenuation are illustrated in Figure 4.8. The set $\mathbf{K} = \{-50, -49, \ldots, 48, 49\}$ is selected for

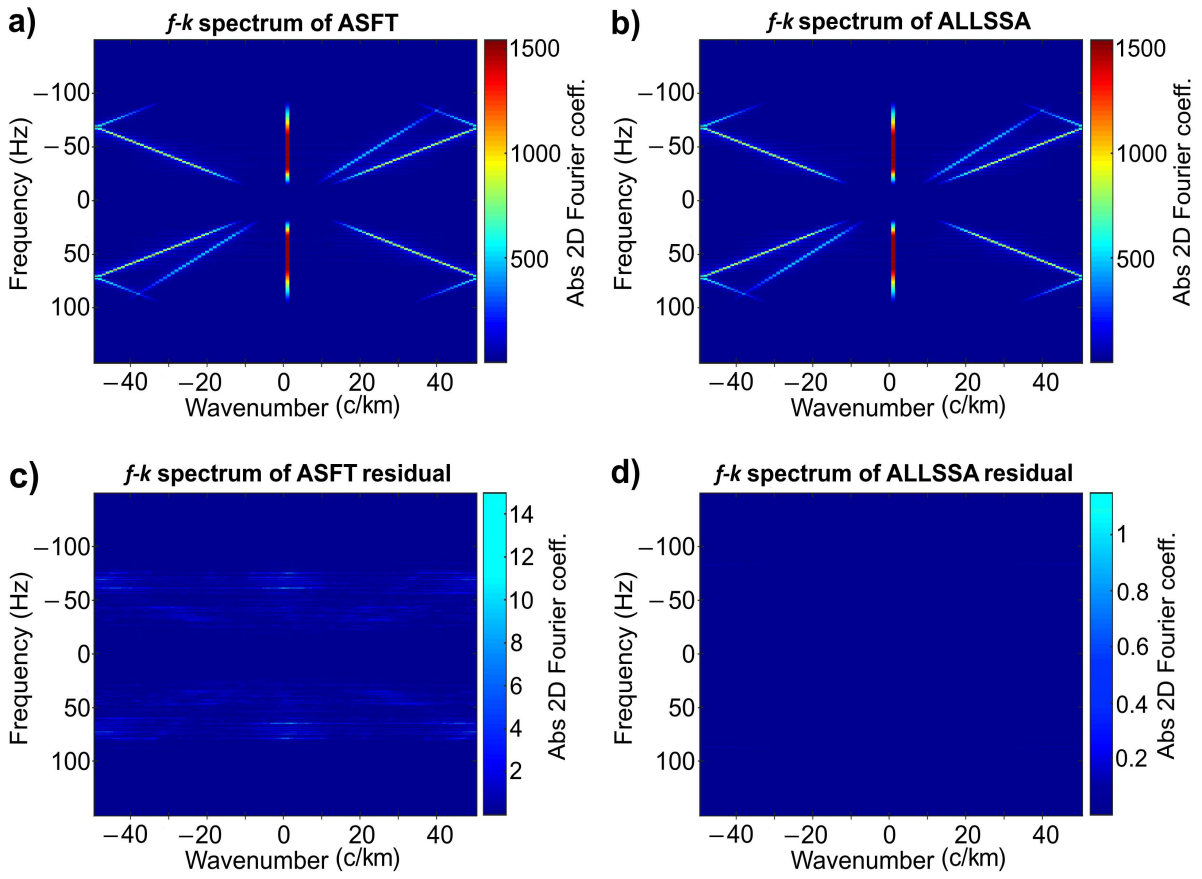

Figure 4.7: The *f-k* spectra corresponding to (a) Figure 4.6a, (b) Figure 4.6b, (c) Figure 4.6c, and (d) Figure 4.6d.

both the ALFT and ASFT. Comparing the ALLSSA with the ALFT and ASFT, one can observe that the ALLSSA performs much better in reconstructing the signals (the events) and attenuates the random noise down to a certain confidence level. The result of the ALFT and ASFT for trace 1 and 15 looks noisier because aside from the accuracy of the actual wavenumbers, more noise is interpolated in the traces. In the ALLSSA, when the 99% confidence level is used, the noise attenuated slightly more than when the 95% level is used; however, it is recommended using the 95% confidence level for real seismic data because of their non-stationarity behaviour so that more wavenumbers will be estimated simultaneously at each round of iterations that may correspond to short duration signals. Note that there is no rigorous statistical property of choosing a confidence level for both the ALFT and ASFT, so a threshold was selected based on the magnitude of random noise for these methods.

In order to see the sensitivity of the ALLSSA to spatially aliased events, the traces of odd numbers are removed except for a few of them (e.g., trace numbers $5, 21, 83, 91$) and show the result in Figure 4.9a. A dense set of wavenumbers $\mathbf{K} = \{-50, -49.9, \ldots, 49.8, 49.9\}$ (1000 wavenumbers) is selected for the ALFT. In this case, the ALFT and ASFT have approximately the same computational time (57s) with $\varepsilon = 109.3$

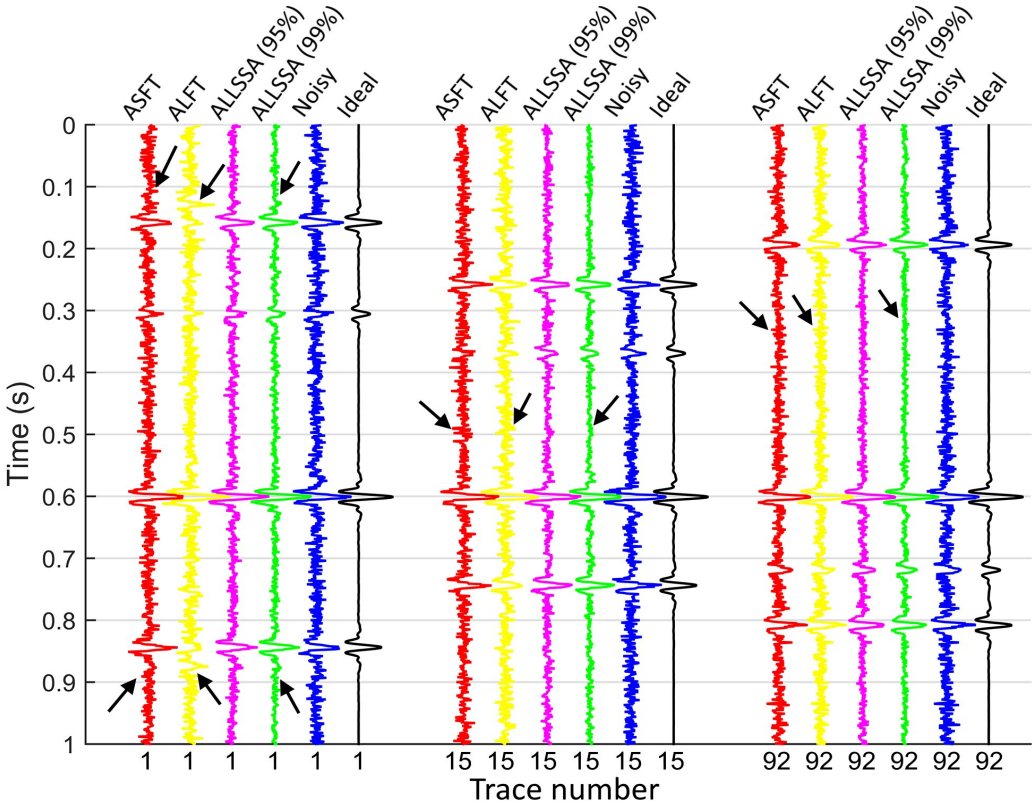

Figure 4.8: The comparison of three regularized traces using the ALFT, ASFT, and ALLSSA. Original traces without random noise in black, with random noise in blue, the ALLSSA results at 99% and 95% confidence levels in green and pink, respectively, the ALFT result in yellow and the ASFT result in red. The left six traces are for trace 1, the middle ones are for trace 15, and the right ones are for trace 92. Arrows show several positions in time where the ALLSSA performed better in random noise attenuation.

(16.79%) and $\varepsilon = 93.4$ (14.35%), respectively. Only the ALFT result is illustrated because the ASFT result is also very similar (cf., Figure 4.9). Note that the residual is amplified by a factor of 10 in Figure 4.9e. To show aliasing, the *f-k* spectrum of this synthetic data is illustrated in Figure 4.9b. Since the dominant trace spacing is 20m, all of the events are spatially aliased. The *f-k* spectra of the ALFT regularization result and its residual are shown in Figure 4.9d and f, respectively. One can clearly observe the effect of the alias events in the spectra (e.g., see the arrows in Figure 4.9f). The computational time for the ALLSSA is 2s, and $\varepsilon = 0.9$ (0.14%) which is significantly smaller than the ALFT and ASFT. The final regularization result is almost the same as Figure 4.6b (not shown here). The reason for this accuracy is that the ALLSSA very accurately estimates the positive wavenumbers, and these few extra traces (four traces) prevents the matrix of normal equations to be singular. In the event that this matrix is singular, one may compute its pseudo-inverse (Rao and Mitra, 1972). Note that if one removes the traces of odd numbers, the Nyquist wavenumber will be 25 c/km, and the aliasing will occur for the wavenumbers beyond 25 c/km. In this situation, an anti-alias technique similar to the technique in Schonewille et al. (2009) may be used.

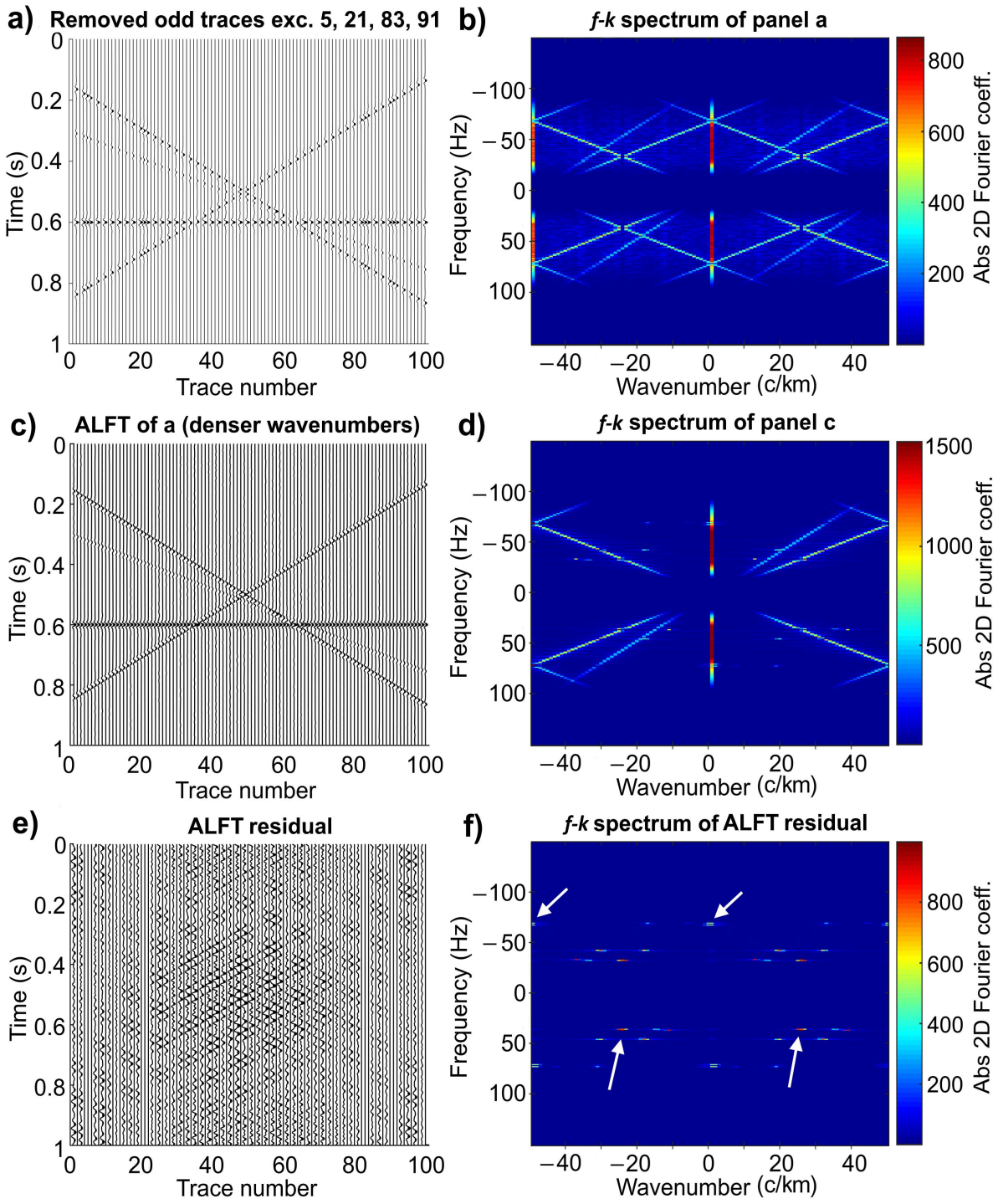

Figure 4.9: (a) The synthetic data in Figure 4.3a after setting the traces of odd numbers to zero except trace numbers $5, 21, 83, 91$, (b) the f-k spectrum of panel a, (c) the regularization result using the ALFT with a denser set of wavenumbers (1000 wavenumbers), (d) the f-k spectrum of panel c, (e) the ALFT residual (amplified 10 times), and (f) the f-k spectrum of the ALFT residual. Arrows show the effect of aliased events in the ALFT residual.

4.3 Field Seismic Data Applications

The field data example is a marine 2D shot gather from a deep water of Gulf of Mexico (pressure data set) that is the same as the one used by Crawley (2000); Fomel (2002); Chen et al. (2015). The shot gather has long-period multiples and complicated diffraction events caused by a salt body. The time sampling rate is 250 samples per second. To compare the ALLSSA with the nonlinear shaping regularization, the same traces (30%) as in Chen et al. (2015) are removed (see Figure 4.10). Note that the normalized wavenumber axis in the f-k spectra is obtained by considering $\Delta x = 1$. The nonlinear shaping regularization result (as described in Chapter 2) after 15 iterations and its residual are illustrated in Figure 4.11a and c, respectively.

Since stationarity of data series is the main assumption of the ALFT, ASFT, and ALLSSA (i.e., the seismic events must be approximately linear), one may use the windowing strategy. The seismic data are divided to six spatial non-overlapping windows (see blue windows in Figures 4.11b and 4.13a), and the same threshold is applied for both the ALFT and ALLSSA. Within each window, $\mathbf{K} = \{-15, -14.9, \ldots, 14.8, 14.9\}$ (300

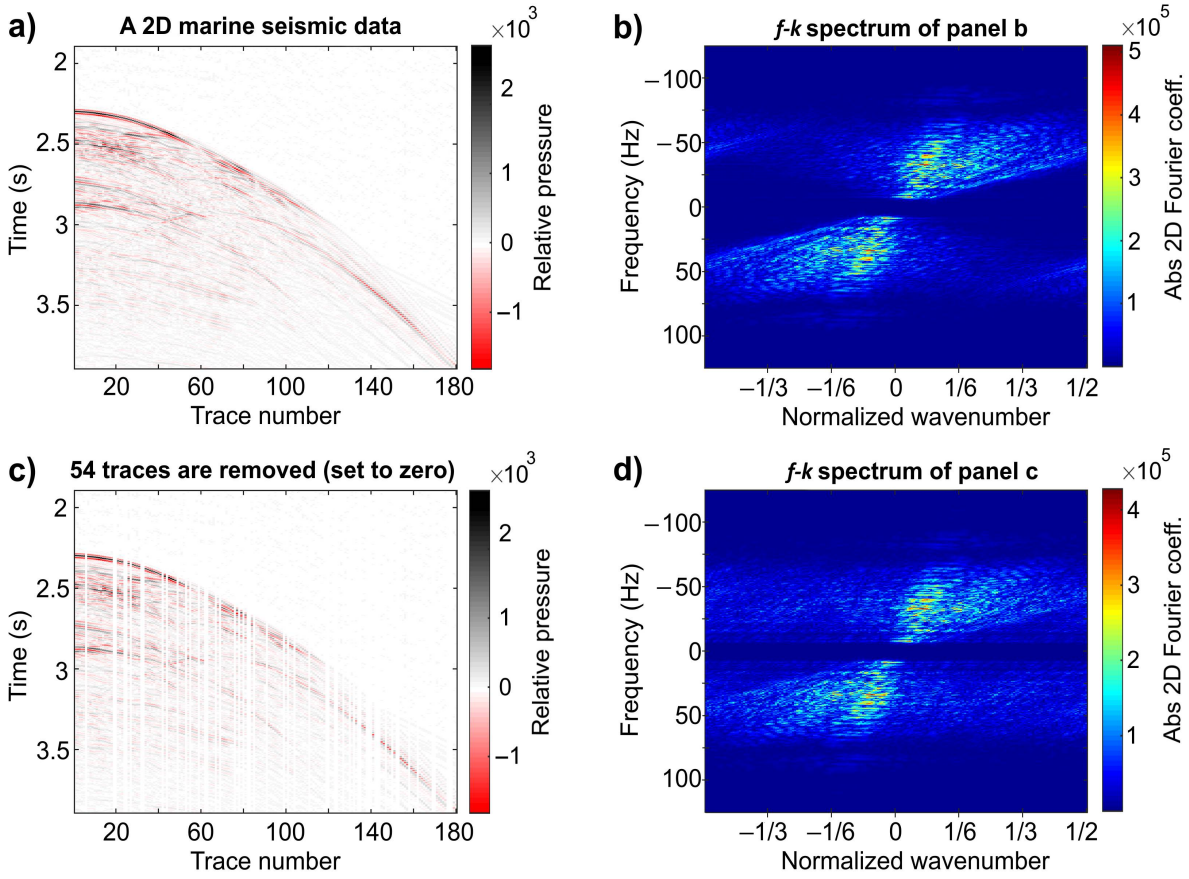

Figure 4.10: (a) A marine 2D shot gather from a deep water of Gulf of Mexico, (b) the f-k spectrum of panel a, (c) 30% of the traces are randomly removed, and (d) the f-k spectrum of panel c.

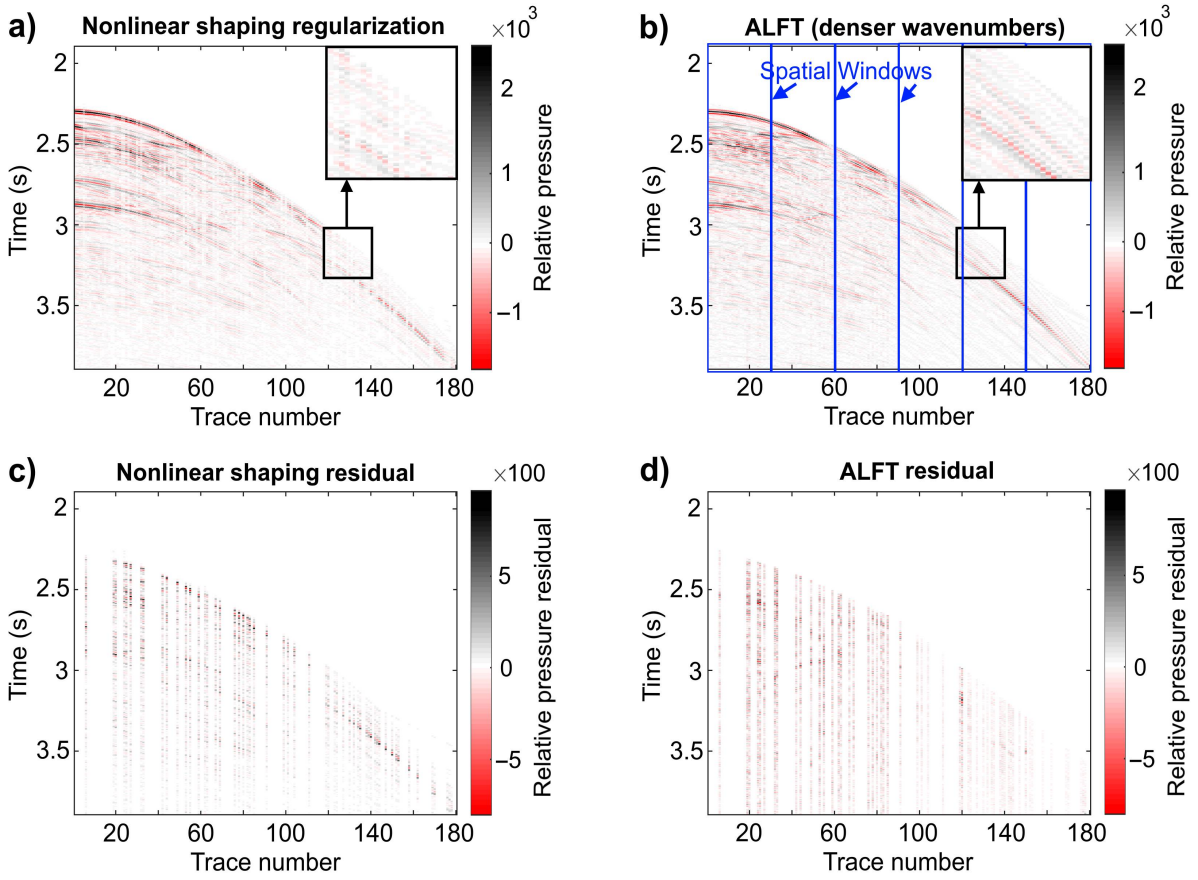

Figure 4.11: (a) The result of the nonlinear shaping regularization after 15 iterations, (b) the ALFT result, (c) the residual of panel a, and (d) the residual of panel b.

wavenumbers) for the ALFT, and $\mathbf{\Omega} = \{1, \ldots, 14\}$ and initial $\underline{\mathbf{\Phi}} = [\mathbf{1}]$ for the ALLSSA. The regularization result of the ALFT and its residual are illustrated in Figure 4.11b and d. The *f-k* spectra corresponding to the nonlinear shaping regularization and ALFT are shown in Figure 4.12. The regularization result using the ALLSSA and its residual are shown in Figure 4.13a and c, respectively. Figure 4.14a and c show the *f-k* spectra corresponding to Figure 4.13a and c, respectively.

One can observe from these figures that the ALFT and ALLSSA performed well in interpolating the missing traces as compared to the nonlinear shaping regularization; however, they still have small amount of artifacts (e.g., compare the magnified sections in Figures 4.11b and 4.13a). The data are very poorly sampled along the spatial direction within a few of the windows (e.g., half of the traces are missing within the magnified section), and so the spatial aliasing mainly caused these artifacts. An anti-aliasing scheme within the windows may further improve these results (Schonewille et al., 2009; Xu et al., 2010). An alternative approach is now proposed for a better interpolation result.

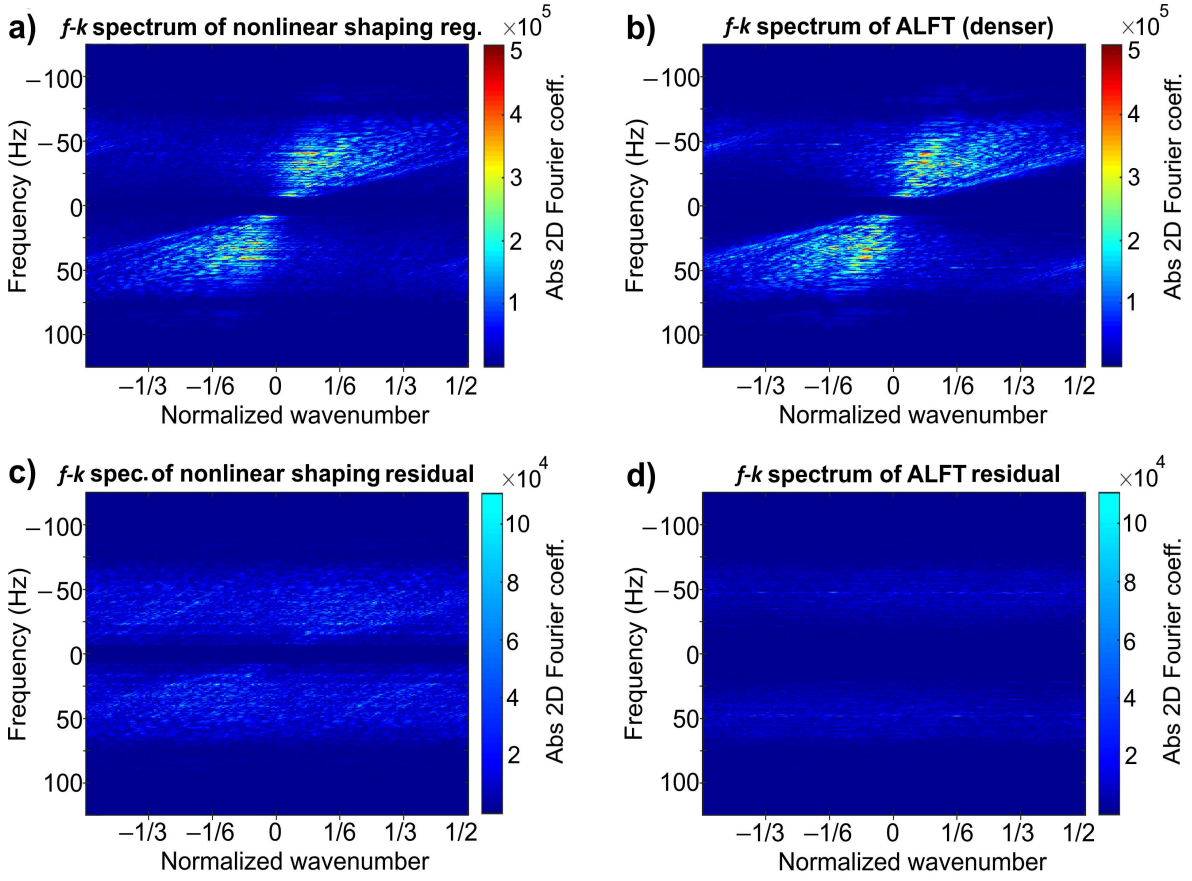

Figure 4.12: The *f-k* spectra corresponding to (a) Figure 4.11a, (b) Figure 4.11b, (c) Figure 4.11c, and (d) Figure 4.11d.

In the weighted ALLSSA, the larger the values of the diagonal entries $\mathbf{P}$ are, the more the fitting process is focused around the locations of those values and vice versa. For a faster computation, one may treat the weight matrix in Equation (3.10) as a vector $\mathbf{P} = [w(\ell)]$ $(1 \leq \ell \leq n)$ whose entries are the values of the following Gaussian function

$$w(\ell) = e^{-\wp(x_\ell - x_i)^2}, \tag{4.3}$$

where $\wp$ determines the effective width of the Gaussian function (length of the window), and x_i is the location of a trace being approximated. By this selection of $\mathbf{P}$, the sinusoidal basis functions will be adapted to the Morlet wavelet in the least-squares sense, smoothing the seismic image (Foster, 1996; Ghaderpour and Pagiatakis, 2017). This selection also allows us to incorporate more wavenumbers into the algorithm resulting in a better interpolation result (Crawley, 2000). A similar discussion for selection of the weight function in the ALFT is also given in Xu et al. (2010).

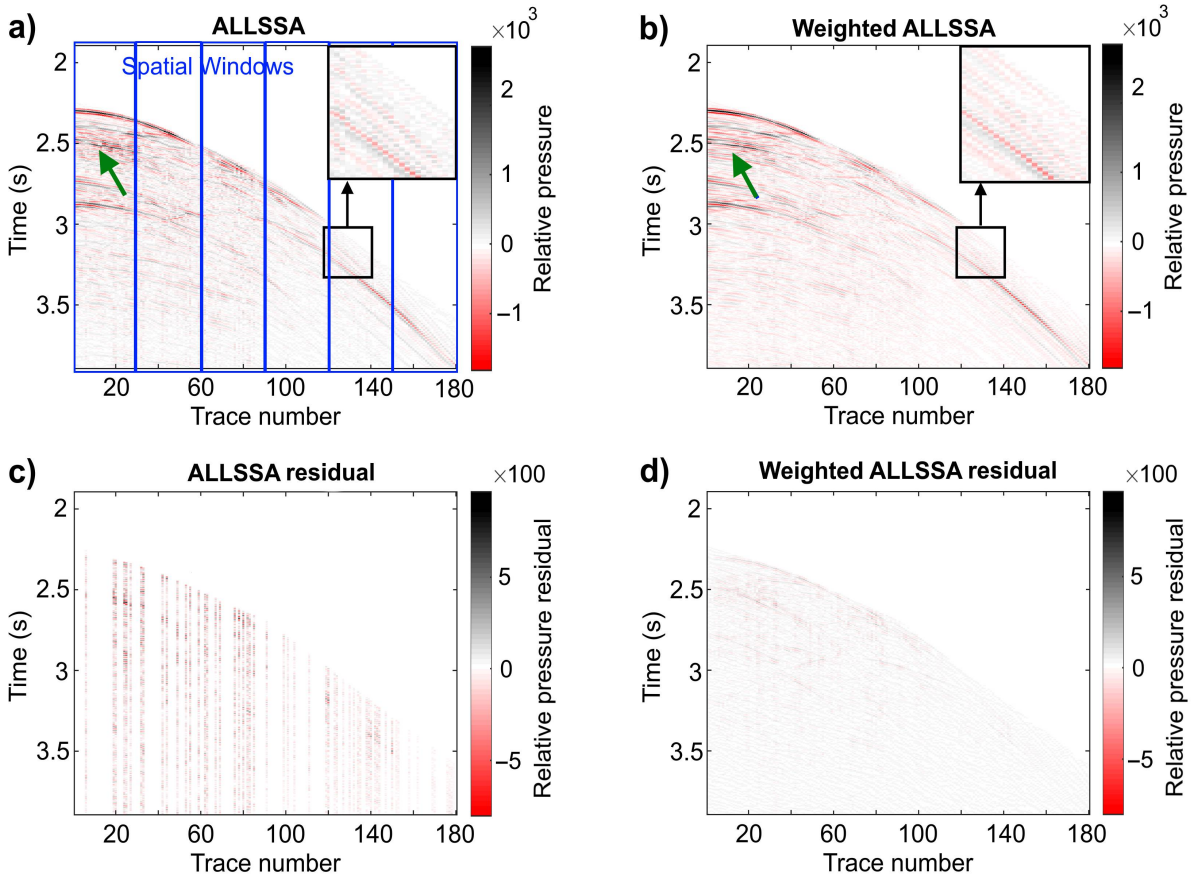

Figure 4.13: (a) The ALLSSA result, (b) the weighted ALLSSA result, (c) the residual of panel a, and (d) the residual of panel b. The events in panel b are constructed more smoothly than in panel a (e.g., the green arrows and magnified sections).

One may apply the weighted ALLSSA to estimate each trace individually using the data within the spatial window containing that trace which may not necessarily fall in the centre of the window. For existing traces, one may still take advantage of the computational speed of the ALLSSA and apply it to attenuate the random noise of the traces down to a certain confidence level and so smooth the entire seismic data (Gaussian overlapping windows).

For better comparison between the ALLSSA and the weighted ALLSSA, $\wp$ in Equation (4.3) is selected in such a way that the Gaussian bell curve is almost truncated to zero in the margins of the translating windows whose lengths are the same as the window length used above. The result of the weighted ALLSSA and its corresponding residual are shown in Figure 4.13b and d, respectively[32]. The *f-k* spectra corresponding to Figure 4.13b and d are illustrated in Figure 4.14b and d, respectively. One can observe that the events are smoothly constructed, and the random noise is also attenuated down to the 95% confidence level (e.g., green

[32]The basis functions in the weighted ALLSSA are still the sinusoidal basis functions, but the Gaussian weights given by Equation (4.3) adapt the sinusoids to the morlet wavelet in the least-squares sense.

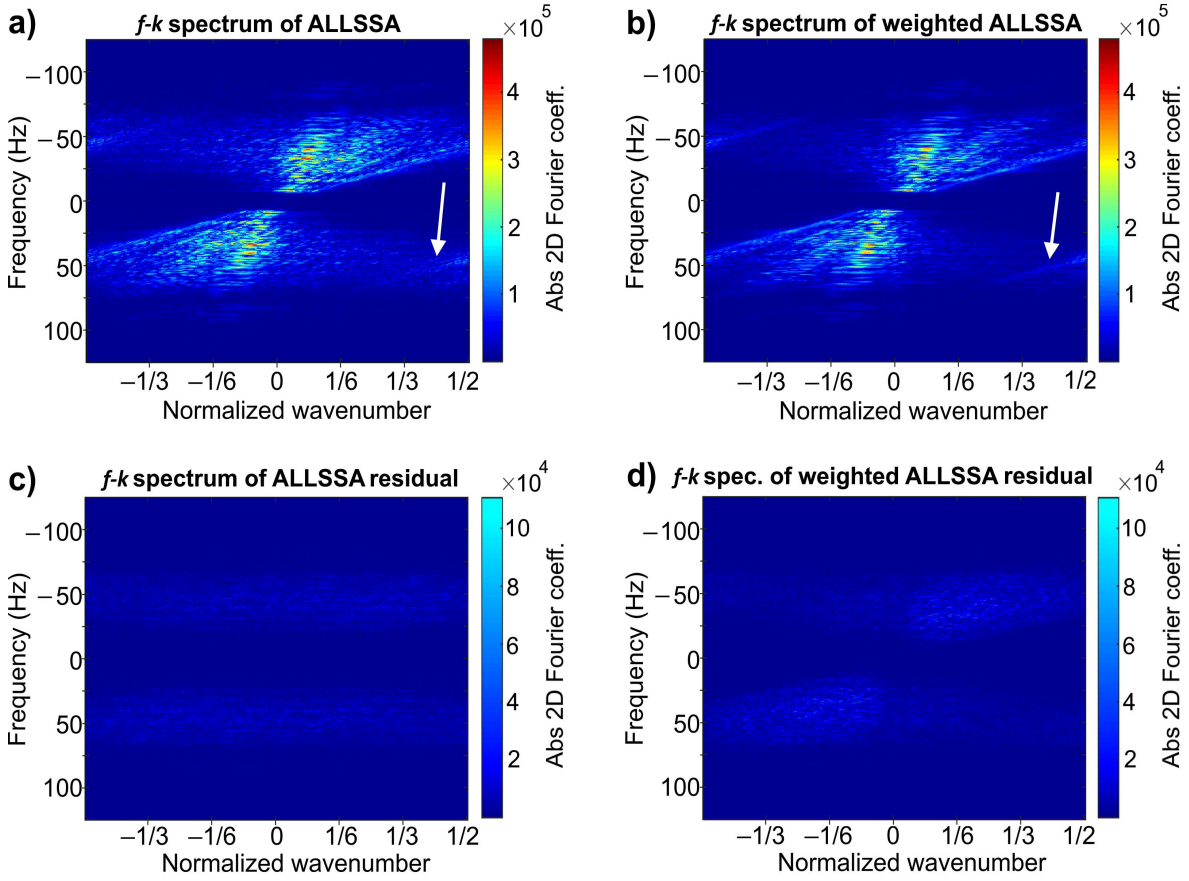

Figure 4.14: The *f-k* spectra corresponding to (a) Figure 4.13a, (b) Figure 4.13b, (c) Figure 4.13c, and (d) Figure 4.13d. The white arrows in panels a and b refer to the linear event shown in the magnified sections in Figure 4.13a and b, respectively.

arrows in Figure 4.13). The almost linear event in the south-east part of the seismic data is clearly constructed (compare the events within the magnified sections shown in Figure 4.13a and b and the corresponding *f-k* spectra shown in Figure 4.14a and b, white arrows). All the methods show small artifacts above the water-bottom reflections that are not significant. A mask is used to set the values of the north-east part of the images shown in Figures 4.11 and 4.13 to zero after the interpolation.

Note that the traces within each of the six spatial windows are simultaneously estimated in the ALFT and ALLSSA, and so the existing traces are almost fully constructed, resulting in error propagation to other traces (Figures 4.11b and 4.13a). However, in the weighted ALLSSA, each trace is individually approximated by applying 95% confidence level (the LSSA critical value at 95% confidence level is used to identify all the statistically significant peaks in the spectrum), so the events are smoothly constructed, and random noise is also attenuated down to 95% confidence level for the entire seismic data.

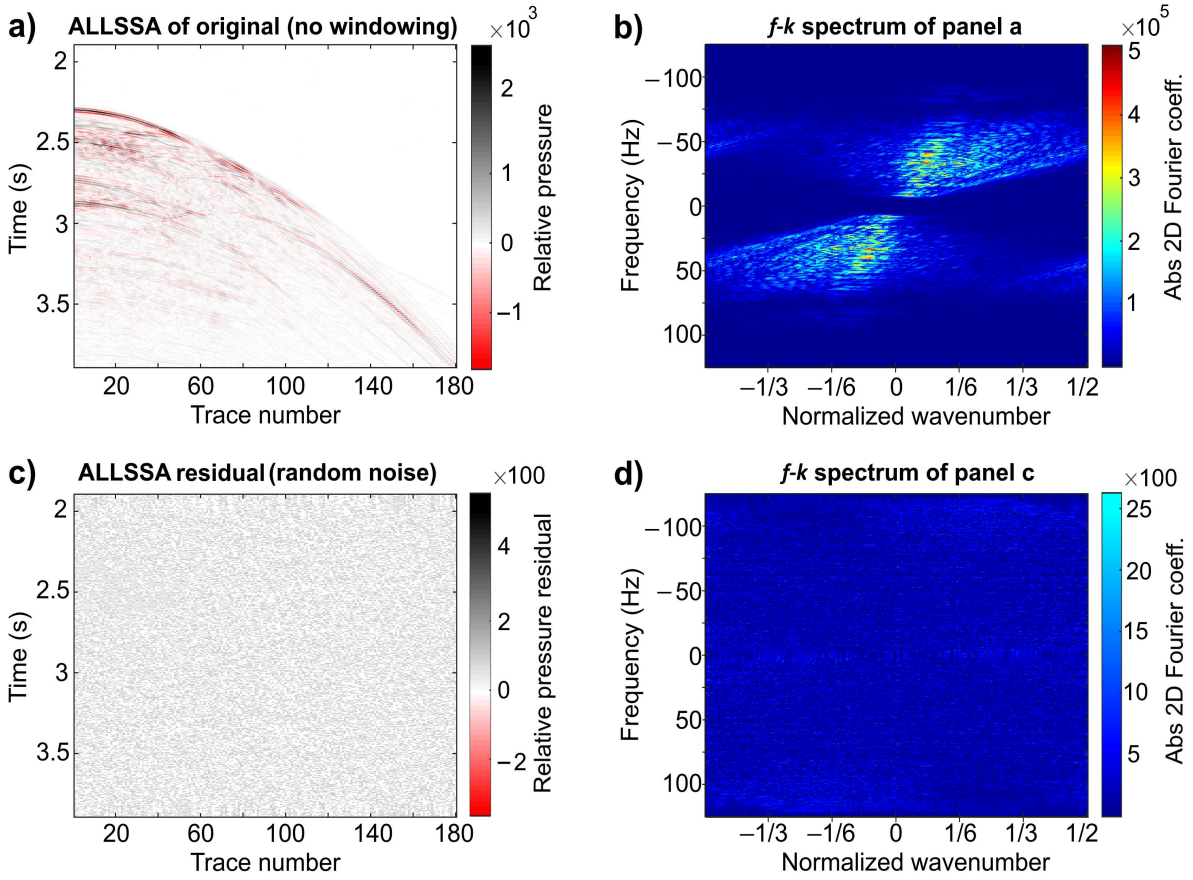

Figure 4.15: (a) The noise attenuated result of Figure 4.10a using the ALLSSA, (b) the *f-k* spectrum of panel a, (c) the ALLSSA residual, and (d) the *f-k* spectrum of panel c.

Finally, for each temporal frequency slice of the original seismic data (regularly spaced), one may repeat the ALLSSA until the L2 norm of the residual series becomes less than approximately 10% of the total L2 norm of the data series and use the inverse FFT to reconstruct the traces. The result of this process, its difference from the original seismic data, and their corresponding *f-k* spectra are shown in Figure 4.15. One can observe that the geologic structures are clearly reconstructed, and the random noise is attenuated (A 95% confidence level is used in each iteration).

4.4 Summary

Two windowing techniques, namely, the rectangular non-overlapping and Gaussian overlapping windowing techniques, are introduced and applied to regularize 2D non-stationary marine seismic data. Within each translating window over distance, the (weighted) ALLSSA is applied to regularize segments of data series within that window. In the field data example, I showed that applying the Gaussian window can regularize the seismic data smoother and simultaneously can attenuate the random noise down to a certain confidence level. In 3D seismic data regularization, the Gaussian window may be considered as the two-dimensional Gaussian window for regularization and random noise attenuation. As demonstrated in Figures 4.8 and 4.14, the ALLSSA is able to interpolate/extrapolate the seismic traces more accurately than the state-of-the-arts methods while it also attenuates the random noise down to a certain confidence level that is a breakthrough in signal processing.

There are several other sophisticated windows that may be used in the weighted ALLSSA, such as Dolph-Chebyshev and Kaiser windows. The Dolph-Chebyshev window minimizes the Chebyshev norm of the side-lobes for a given main-lobe (Dolph, 1946; Harris, 1978). In other words, it minimizes the L_∞ norm of the local maxima in the transform magnitude outside of the main-lobe (the central local maximum). The Kaiser window also approximates a window that maximizes the energy concentration in the main-lobe (Kaiser and Schafer, 1980). These windows are more difficult to compute and implement compared to the Gaussian window. One of the advantages of the least-squares optimization in the LSSA, IMAP, and ALLSSA is that it considers the correlation between the sinusoids, and so it performs much better with much less spectral leakages compared to the Fourier based methods, such as the DFT, ALFT, and ASFT, when segmenting data series. Therefore, no window, such as Gaussian or Dolph-Chebyshev or Kaiser window, is required in the least-squares based methods developed in this dissertation. However, it may be interesting to investigate the effect of using Dolph-Chebyshev and Kaiser windows on the regularization results, that is subject to future work.

5 Multichannel Antileakage Least-Squares Spectral Analysis

As was shown in Chapter 4, when seismic data present mild aliasing in the f-k spectrum, the ALLSSA can regularize the aliased data effectively. However, certain assumption, such as linearity of seismic events (Schonewille et al., 2009) or additional information, such as spatial gradients of the seismic data (Vassallo et al., 2010a) is further needed to aid regularization beyond aliasing. For example, in the crossline direction, only a few irregularly spaced samples are available, and the data series may be strongly aliased. Or, the samples of a data series are regularly spaced, but the spectrum of the data series presents severe aliasing when the maximum wavenumber is chosen as the Nyquist wavenumber of the desired regularly spaced series having finer resolution. In both situations, the data series is referred to as "coarsely sampled".

Recent marine vessels usually tow true multicomponent streamers that can measure the full particle velocity vector in addition to the pressure. The particle velocity measurements can increase the effective Nyquist wavenumber by a factor of two or three, depending on how they are used (Robertsson et al., 2008). Because particle motion can easily be converted to a pressure gradient by using the equation of motion[33], such streamers would enable acquisition of both pressure and the gradient of pressure simultaneously. The Multichannel Interpolation by MAtching Pursuit (MIMAP) is based on the IMAP that uses multicomponent seismic measurements to reconstruct the seismic wavefield at any desired crossline position between towed streamers (Vassallo et al., 2010a).

In this chapter, it is assumed that data series $\mathbf{f}(\mathbf{x})$ and its spatial gradient $\mathbf{f_x}(\mathbf{x})$ are available and similar techniques proposed by Vassallo et al. (2010a) are used to regularize coarsely sampled data series. The new method developed in this chapter, namely, Multichannel AntiLeakage Least-Squares Spectral Analysis (MALLSSA), incorporates the spatial gradients of seismic data into the ALLSSA to regularize the data series beyond aliasing. The MALLSSA is able to estimate the wavenumbers of data series accurately, resulting in smaller number of iterations and better regularization results. The outstanding performance of this method is shown on synthetic data series, synthetic and marine seismic data sets.

[33]The equation of motion states that the particle acceleration vector $\mathbf{a}$ (measured by an accelerometer) is proportional to the gradient of pressure P by equation $\nabla P = -\rho\mathbf{a}$, where ρ is the density of the medium. In particular, in the crossline direction (denoted by x here), $P_x = -\rho\dot{V}_x$, where the dot above V_x denotes the temporal derivative (Vassallo et al., 2010a).

5.1 Multichannel Interpolation by Matching Pursuit

In this section, the Multichannel Interpolation by MAtching Pursuit (MIMAP) is briefly described in matrix form, and the reader is referred to Vassallo et al. (2010a) for more details. Suppose that $\mathbf{\Omega}$ is a set of wavenumbers and let $\mathbf{f} = [f(x_\ell)]$ be a column vector of n samples ($\ell = 1, \ldots, n$), and $\mathbf{f_x} = [f_x(x_\ell)]$ be its spatial gradient, where the x_ℓ's may be irregularly spaced. Note that $\mathbf{f}$ and $\mathbf{f_x}$ do not need to have the same size. For each wavenumber $\omega_k \in \mathbf{\Omega}$, let $\mathbf{\Phi}_k$ and $\mathbf{\Phi}_{k_\mathbf{x}}$ be design matrices of orders $n \times 2$ defined as

$$\mathbf{\Phi}_k = \big[\cos(2\pi\omega_k x_\ell),\ \sin(2\pi\omega_k x_\ell)\big], \tag{5.1}$$

$$\mathbf{\Phi}_{k_\mathbf{x}} = \big[-2\pi\omega_k \sin(2\pi\omega_k x_\ell),\ 2\pi\omega_k \cos(2\pi\omega_k x_\ell)\big]. \tag{5.2}$$

Also, let $\mathbf{c}_k = [a_k\ b_k]^\mathrm{T}$, where T is the transpose, and a_k and b_k are the coefficients of the cosine and sine basis functions being estimated, respectively. For each wavenumber $\omega_k \in \mathbf{\Omega}$, the MIMAP minimizes the following cost function with respect to $\mathbf{c}_k$

$$\Psi_k(\mathbf{c}_k) = (\mathbf{f} - \mathbf{\Phi}_k\ \mathbf{c}_k)^\mathrm{T}(\mathbf{f} - \mathbf{\Phi}_k\ \mathbf{c}_k) + \lambda(\mathbf{f_x} - \mathbf{\Phi}_{k_\mathbf{x}}\ \mathbf{c}_k)^\mathrm{T}(\mathbf{f_x} - \mathbf{\Phi}_{k_\mathbf{x}}\ \mathbf{c}_k), \tag{5.3}$$

where T is the transpose, and so it estimates $\mathbf{c}_k$ as

$$\hat{\mathbf{c}}_k = \Big(\mathbf{\Phi}_k^\mathrm{T}\mathbf{\Phi}_k + \lambda\ \mathbf{\Phi}_{k_\mathbf{x}}^\mathrm{T}\mathbf{\Phi}_{k_\mathbf{x}}\Big)^{-1}\Big(\mathbf{\Phi}_k^\mathrm{T}\mathbf{f} + \lambda\ \mathbf{\Phi}_{k_\mathbf{x}}^\mathrm{T}\mathbf{f_x}\Big). \tag{5.4}$$

The derivation of Equation (5.4) is shown in Appendix D in a more general case (i.e., for the MALLSSA). The weighting parameter λ in Equation (5.3) adjusts the relative contributions of data and gradient residuals to the cost function. The proper selection of λ should consider the expected energy difference between the two signals and the signal-to-noise ratio of the respective multicomponent measurements (Vassallo et al., 2010a). When $\lambda = 0$, the gradient measurements are discarded, and the MIMAP degenerates to the IMAP as is briefly described in Chapter 2.

After substituting $\hat{\mathbf{c}}_k$ given in Equation (5.3), the optimization problem becomes estimating ω_k that maximizes the following cost function that corresponds to the extension of the Lomb spectrum to the case of multichannel reconstruction (Vassallo et al., 2010a, Appendix A):

$$s_\lambda(\omega_k) = \mathbf{f}^\mathrm{T}\mathbf{\Phi}_k\hat{\mathbf{c}}_k + \lambda\ \mathbf{f_x}^\mathrm{T}\mathbf{\Phi}_{k_\mathbf{x}}\hat{\mathbf{c}}_k. \tag{5.5}$$

After estimating ω_k and $\hat{\mathbf{c}}_k$, the MIMAP obtains the residual data series ($\mathbf{f} - \mathbf{\Phi}_k\ \hat{\mathbf{c}}_k$) and its residual gradient ($\mathbf{f_x} - \mathbf{\Phi}_{k_\mathbf{x}}\ \hat{\mathbf{c}}_k$) and treat them as new input data series and gradient and performs the entire process on these residual series. When the energy of the overall residual is sufficiently reduced, the iterations are terminated, and the estimated wavenumbers and amplitudes (ω_k's and $\hat{\mathbf{c}}_k$'s) will be used to reconstruct the data series on any desired equally spaced series.

In the IMAP and MIMAP, the wavenumbers and corresponding amplitudes are estimated one at a time (out-of-context), ignoring the correlations between the sinusoids of different wavenumbers (see Appendix A). Therefore, the estimated wavenumbers and amplitudes are not very accurate, resulting in large number of iterations and less accuracy compared to when these correlations are considered. In the next section, I show how these correlations can be considered as well as the constituents of known forms and covariance matrices associated with the data series.

5.2 Multichannel Antileakage Least-Squares Spectral Analysis

Suppose that $\mathbf{\Omega} = \{1, 2, \ldots, \eta\}$ is a preselected set of wavenumbers, where η is the Nyquist wavenumber of the desired regular series. Let $\mathbf{f} = [f(x_\ell)]$ be a column vector of n samples ($\ell = 1, \ldots, n$), and $\mathbf{f_x} = [f_x(x_\ell)]$ be its spatial gradient, where the x_ℓ's may be irregularly spaced. Suppose that $\mathbf{C_f}$ and $\mathbf{C_{f_x}}$ are the covariance matrices associated with $\mathbf{f}$ and $\mathbf{f_x}$, respectively (see Equation (3.4)). Let $\mathbf{P} = {\mathbf{C_f}}^{-1}$ and $\mathbf{P_x} = {\mathbf{C_{f_x}}}^{-1}$ that can identify parameter λ in the MIMAP in a more general and rigorous way. Note that $\mathbf{f}$ and $\mathbf{f_x}$ do not need to have the same size; however, their sizes are assumed to be the same in this chapter. For each wavenumber $\omega_k \in \mathbf{\Omega}$, the MALLSSA minimizes the following cost function

$$\Psi_k(\overline{\mathbf{c}}) = \big(\mathbf{f} - \overline{\mathbf{\Phi}}\ \overline{\mathbf{c}}\big)^{\mathrm{T}}\mathbf{P}\big(\mathbf{f} - \overline{\mathbf{\Phi}}\ \overline{\mathbf{c}}\big) + \big(\mathbf{f_x} - \overline{\mathbf{\Phi}}_{\mathbf{x}}\ \overline{\mathbf{c}}\big)^{\mathrm{T}}\mathbf{P_x}\big(\mathbf{f_x} - \overline{\mathbf{\Phi}}_{\mathbf{x}}\ \overline{\mathbf{c}}\big), \tag{5.6}$$

where T is the (conjugate) transpose, and $\overline{\mathbf{\Phi}}$ and $\overline{\mathbf{\Phi}}_{\mathbf{x}}$ are the design matrices of orders $n \times (q+2)$ defined as

$$\overline{\mathbf{\Phi}} = \Big[\underline{\mathbf{\Phi}},\ \mathbf{\Phi}_k\Big], \quad \overline{\mathbf{\Phi}}_{\mathbf{x}} = \Big[\underline{\mathbf{\Phi}}_{\mathbf{x}},\ \mathbf{\Phi}_{k_{\mathbf{x}}}\Big] \tag{5.7}$$

such that $\underline{\mathbf{\Phi}} = \big[\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_q\big]$ contains the constituents of known forms, $\underline{\mathbf{\Phi}}_{\mathbf{x}} = \big[\mathbf{\Phi}_{1_{\mathbf{x}}}, \ldots, \mathbf{\Phi}_{q_{\mathbf{x}}}\big]$ contains their gradients, and $\mathbf{\Phi}_k$ and $\mathbf{\Phi}_{k_{\mathbf{x}}}$ are given by Equations (5.1) and (5.2), respectively. Minimizing the cost function in Equation (5.6) with respect to $\overline{\mathbf{c}}$, it is shown in Appendix D that

$$\hat{\overline{\mathbf{c}}} = \begin{bmatrix} \mathbf{v} \\ \hat{\mathbf{c}}_k \end{bmatrix} = \Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + \overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big)^{-1}\Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\mathbf{f_x}\Big), \tag{5.8}$$

where $\mathbf{v}$ is the estimated amplitudes of the constituent of known forms, and $\hat{\mathbf{c}}_k$ is the estimated amplitudes of the sine and cosine basis functions of wavenumber ω_k. Now in the process of estimating the wavenumbers, at each step, a wavenumber $\omega_k \in \mathbf{\Omega}$ will be selected such that the following normalized ratio (multichannel spectrum) is maximum

$$s_{\mathrm{multi}}(\omega_k) = \frac{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\mathbf{\Phi}_k\hat{\mathbf{c}}_k + \hat{\mathbf{g}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\mathbf{\Phi}_{k_{\mathbf{x}}}\hat{\mathbf{c}}_k}{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}\hat{\mathbf{g}} + \hat{\mathbf{g}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\hat{\mathbf{g}}_{\mathbf{x}}}, \tag{5.9}$$

where $\hat{\mathbf{g}} = \mathbf{f} - \underline{\mathbf{\Phi}}\ \underline{\hat{\mathbf{c}}}$ and $\hat{\mathbf{g}}_{\mathbf{x}} = \mathbf{f_x} - \underline{\mathbf{\Phi}}_{\mathbf{x}}\ \underline{\hat{\mathbf{c}}}$, and

$$\underline{\hat{\mathbf{c}}} = \Big(\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\underline{\mathbf{\Phi}} + \underline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\underline{\mathbf{\Phi}}_{\mathbf{x}}\Big)^{-1}\Big(\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \underline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\mathbf{f_x}\Big). \tag{5.10}$$

Next, a wavenumber will be chosen in a small neighbourhood that maximizes Equation (5.9), and then the sine and cosine basis functions of this wavenumber and their derivatives will be added to the design matrices $\underline{\mathbf{\Phi}}$ and $\underline{\mathbf{\Phi}}_{\mathbf{x}}$ in Equation (5.7), respectively (as additional constituents of known forms).

The sinusoidal basis functions and their derivatives may be eliminated from the design matrices to be estimated more accurately in the next step like the ALLSSA. This process continues until $s_{\text{multi}}(\omega_k)$ given by Equation (5.9) is no longer statistically significant at certain confidence level, and so $\hat{\underline{\mathbf{c}}}$, given by Equation (5.10), will comprise the estimated amplitudes of the final constituents in the first round of iterations, and it will produce a spectrum called the Multichannel Antileakage Least-Squares Spectrum (MALLSS). When the data series is wide-sense stationary, one round of iterations will provide sufficient wavenumbers to regularize the data series. However, one may repeat the entire process on the new residuals in the next round of iterations and continue until the L2 norm of residuals goes below a threshold (the residuals are random).

The weight matrices $\mathbf{P}$ (associated with $\mathbf{f}$) and $\mathbf{P}_{\mathbf{x}}$ (associated with $\mathbf{f}_{\mathbf{x}}$) balance the relative importance between the data series and its gradient (the noiser the gradient is, the less it will be considered). After each round of iterations, these matrices may be updated using the covariance law. Note that when $\mathbf{P}_{\mathbf{x}} = \mathbf{0}$, Equation (5.9) will be the same as Equation (3.10), following the beta distribution. In Equation (5.7), the x_ℓ's in $\overline{\mathbf{\Phi}}$ do not need to be the same as the ones in $\overline{\mathbf{\Phi}}_{\mathbf{x}}$; however, in the following examples, they are considered to be the same. A few of the constituents of known forms can be preselected, such as the column vector of all ones or the distance vector. Other basis functions can be considered, such as the sinusoidal basis functions, whose wavenumbers will be estimated in an iterative manner using the same partitioning of the wavenumbers described in Chapter 3.

When a data series is coarsely sampled, several sinusoids of various wavenumbers may identically fit the constituents of the data series; however, the amplitudes of their derivatives are different, resulting in selecting the correct wavenumbers when minimizing Equation (5.6). In other words, the derivative measurements can alter the sampling criteria (Vassallo et al., 2010a). Note that the MALLSSA is slightly different from the MIMAP in that the wavenumbers are simultaneously being estimated in an iterative manner, resulting in higher accuracy and a smaller number of iterations. If one does not consider $\underline{\mathbf{\Phi}}$ and $\underline{\mathbf{\Phi}}_{\mathbf{x}}$ in Equation (5.7) that contain the constituents of known forms and ignores the weight matrices $\mathbf{P}$ and $\mathbf{P}_{\mathbf{x}}$ (simply replace them by parameter λ), then the MALLSSA is in fact the MIMAP.

In the following examples, it is shown that when $\mathbf{\Phi}_1$ in $\underline{\mathbf{\Phi}}$ is selected as the column vector of ones, the horizontal linear event will be well constructed. This selection is crucial in the regularization of seismic data beyond aliasing because it also considers the trace averaging in the least-squares sense. Note that the gradient of this vector is zero, and it cannot be useful for horizontal events whose spectra are vertical in the *f-k* spectrum. This selection helps further in more accurate estimation of the wavenumbers.

5.3 Synthetic data series regularization beyond aliasing

Consider the following data series and its gradient series

$$f(x_\ell) = 5\sin(25.6x_\ell) + 2.5\sin(128x_\ell + 1) + \sqrt{3}\sin(140x_\ell) + 4\cos(350x_\ell), \tag{5.11}$$

$$f_x(x_\ell) = 5(25.6)\cos(25.6x_\ell) + 2.5(128)\cos(128x_\ell + 1) + \sqrt{3}(140)\cos(140x_\ell) - 4(350)\sin(350x_\ell), \tag{5.12}$$

where $x_\ell = \ell/32$, $\ell = 1, \ldots, 32$. The data series is shown by black stars in Figure 5.1b. The goal is to regularize the data series on a series with regular spacing 1/128 whose Nyquist wavenumber is 64. The actual wavenumbers of the data series are real numbers: $25.6/(2\pi) = 4.0743665$, $128/(2\pi) = 20.3718327$, $140/(2\pi) = 22.2816920$, $350/(2\pi) = 55.7042301$. Since the distance between every two consecutive samples is 1/32 and the goal is to reconstruct the data series on a series with regular spacing 1/128, the data series presents three Nyquist wavenumbers at 16, 32, and 48 (cf., Figure 5.1a). The actual locations of two wavenumbers are between 16 and 32, and one is after 48.

The LSSA amplitude spectrum (the square root of the sum of squares of estimated sine and cosine coefficients of a wavenumber) is shown in Figure 5.1a. The estimated wavenumbers using the LSSA show severe aliasing. For example, in the LSS (see black peaks in Figure 5.1a), the sinusoids of wavenumbers 4, 28, 36, and 60 fit the data series identically, and so the true signal that has wavenumber close to 4 is indistinguishable from wavenumbers 28, 36, and 60. However, since the amplitude of the gradient series for these wavenumbers are different, the MALLSSA is able to accurately estimate the true wavenumber that is 4.0743 (see the red peak at wavenumber 4.07 in Figure 5.1a). The estimated wavenumbers using the MALLSSA are shown by red in Figure 5.1a that are very accurately estimated up to four decimal places.

The MIMAP spectrum is also shown by blue in Figure 5.1a, showing spectral leakages and has many wavenumbers in the spectrum that increase the computational cost (approximately 1 s for the MIMAP and 0.1 s for the MALLSSA in this example) and reduce the accuracy of the regularization. The L2 norm of the ideal data series with 1/128 sample spacing is 56.2436, and the L2 norm of the MIMAP and MALLSSA residuals are 1.863 and 0.006, respectively. The differences between the ideal and the regularized data series using the MIMAP and MALLSSA are illustrated in Figure 5.1c. Note that since the amplitudes of the components of the gradient series depend on the wavenumber, minimization of Equation (5.6) resulted in picking the correct wavenumbers in the spectrum.

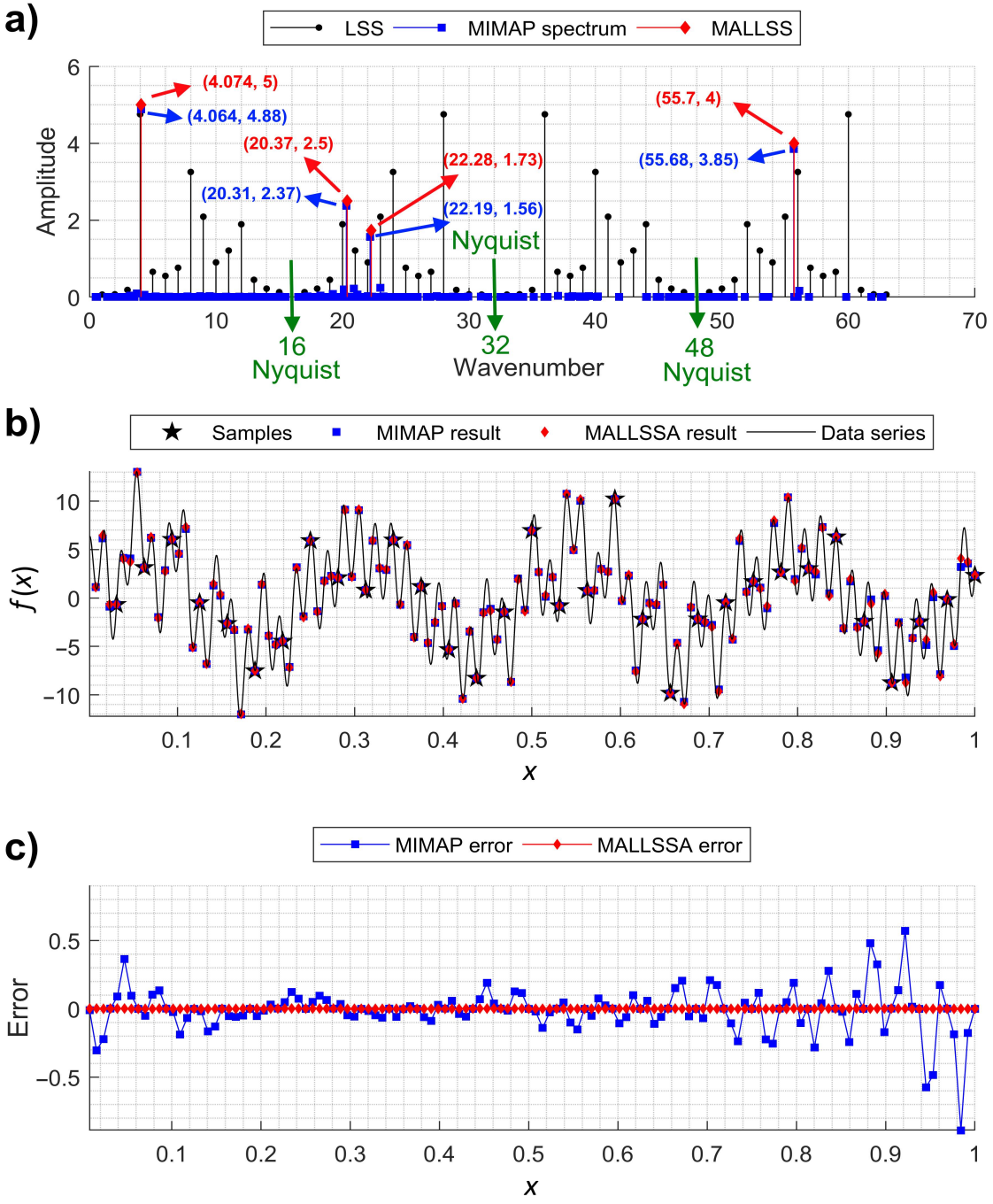

Figure 5.1: A coarsely sampled data series (black stars in panel b) presenting severe aliasing (black circles in panel a) and its regularization on series with spacing 1/128 using the MIMAP (blue squares in panel b obtained from wavenumbers shown by blue squares in panel a) and MALLSSA (red diamonds in panel b obtained from the four wavenumbers shown by red diamonds in panel a), and (c) the difference between the ideal data series and the regularized data series using the MIMAP (blue squares) and MALLSSA (red diamonds). Note that the MIMAP spectrum in panel a shows spectral leakages and has many wavenumbers in the spectrum that increase the computational cost and reduce the accuracy of the regularization (see panel c).

5.4 Simple synthetic stationary seismic data regularization

A synthetic seismic data (pressure) containing four linear events with different amplitudes and dips is shown in Figure 5.2a. The events are generated using an Ormsby wavelet that is a common type of synthetic wavelet in reflection seismology. The wavelet is the same as the one used in Chapter 4, and the time sampling rate is 1000 samples per second. Suppose that the trace spacing is 10 metres, toward the crossline direction. Assume that the unit distance is one kilometre in this and the next example, so the wavenumbers are the number of cycles per kilometre (c/km). The *f-k* spectrum of this data is shown in Figure 5.2b. The vertical peaks at zero wavenumber in the *f-k* spectrum correspond to the horizontal event in Figure 5.2a. The spatial gradient data of the pressure data is shown in Figure 5.2c. The *f-k* spectrum of the spatial gradient data is shown in Figure 5.2d. From Figure 5.2d, one can observe that the crossline gradient amplitude is very low at low wavenumbers, and it linearly grows toward higher wavenumbers. Note that the horizontal linear

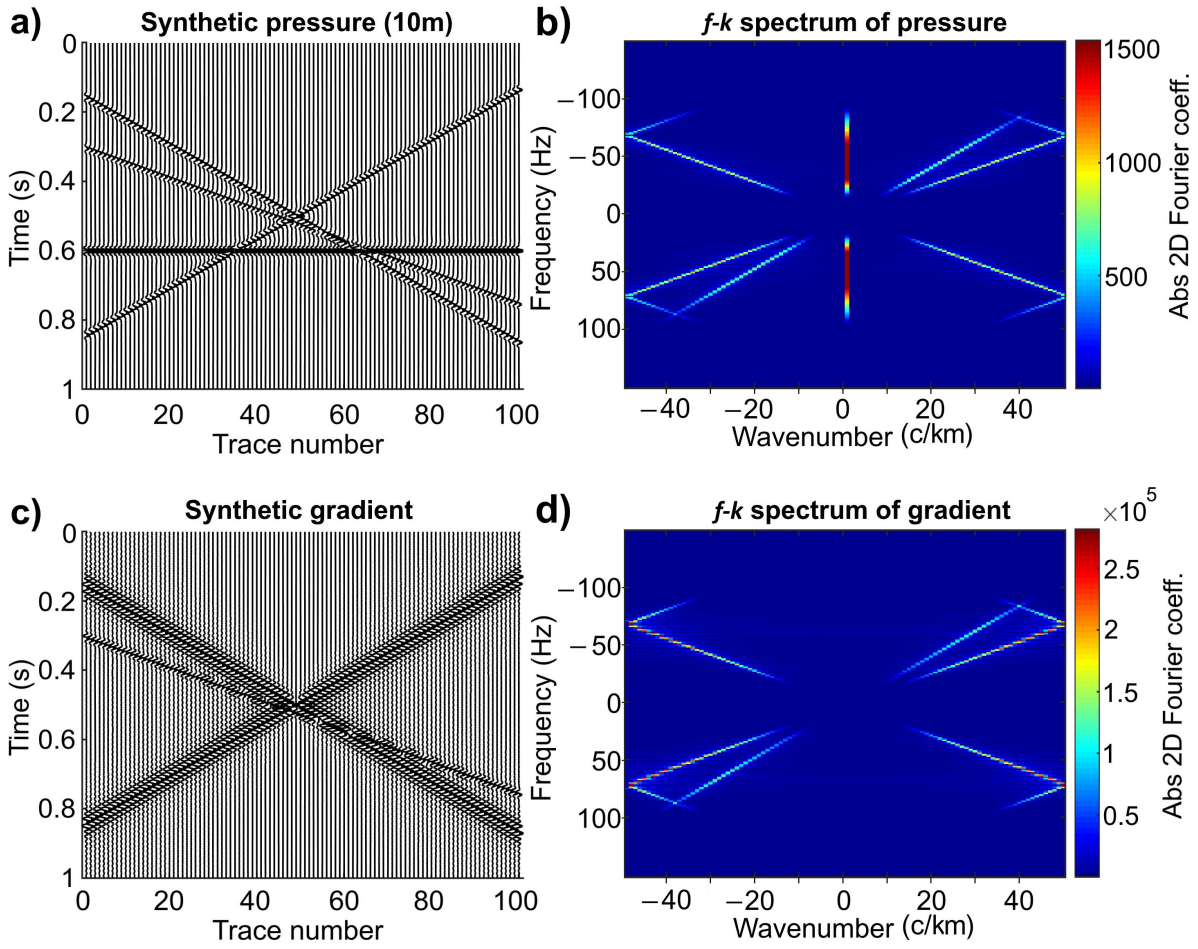

Figure 5.2: Simple synthetic seismic data set. (a) Pressure (10 m trace spacing), (b) the *f-k* spectrum of panel a, (c) gradient, and (d) the *f-k* spectrum of panel c.

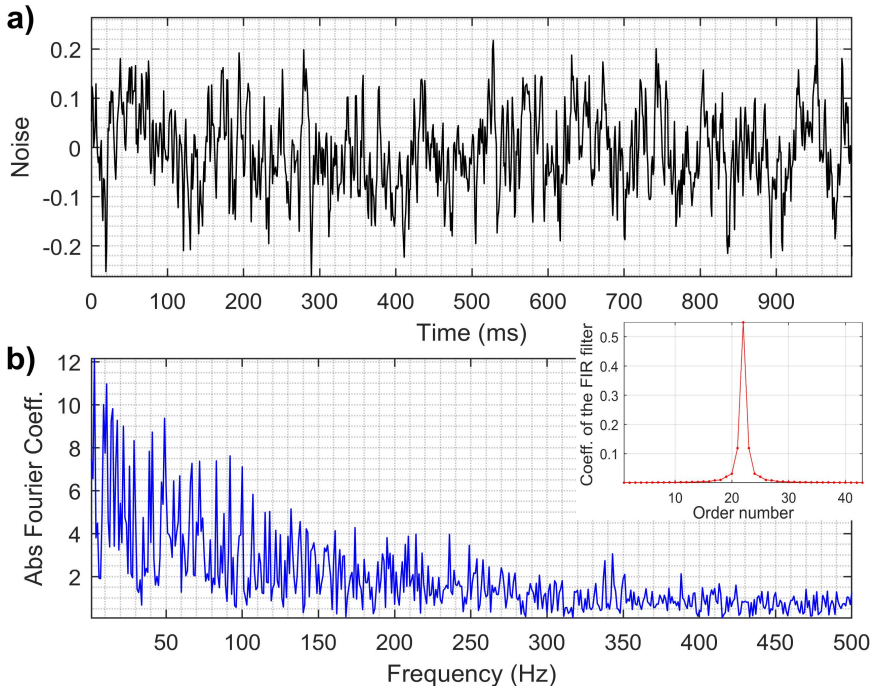

Figure 5.3: (a) Random noise ($1/f$ noise, where f is frequency), and (b) its Fourier spectrum. It is noisier at lower frequencies than higher frequencies. The small panel on the right in red is generated by implementing the MATLAB command 'firls' to create random noise.

event in the seismic data disappeared in its crossline gradients as seen in Figure 5.2c and d.

It is well known that in a real marine acquisition data sets, noise is expected to be higher at low frequencies than at high frequencies (Vassallo et al., 2010a). The MATLAB command filtfilt(b, a, ns) is used to generate $1/f$ noise, where a and b are transfer function coefficients, ns is a vector containing normally distributed random numbers (can be generated by MATLAB command 'randn'), and f is frequency. Figure 5.3a shows such noise when a = 1 and b is generated using the MATLAB command 'firls' that creates a series containing the coefficients of the finite-duration impulse response filter[34] closely matching the $1/f$ passband (cf., the red graph in Figure 5.3). To see the performance of the MALLSSA in higher frequencies, random noise (cf., Figure 5.3b) is generated in such a way that it is more noisier in higher frequencies than the random noise used in Vassallo et al. (2010a). The mean of random noise is approximately zero, and its relative amplitude is about 30% of the signals used in the synthetic seismic data examples in this chapter.

In this example, 75 traces are removed such that the remaining traces are equally spaced with 40 metres spacing, and random noise as shown in Figure 5.3 is introduced to each trace and illustrated with its f-k spectrum in Figure 5.4a and b, respectively. Note that the crossline gradients are kept in the samples

[34]A finite-duration impulse response (FIR) filter is a filter whose impulse response (or response to any finite length input) is of finite duration since it goes to zero in finite time.

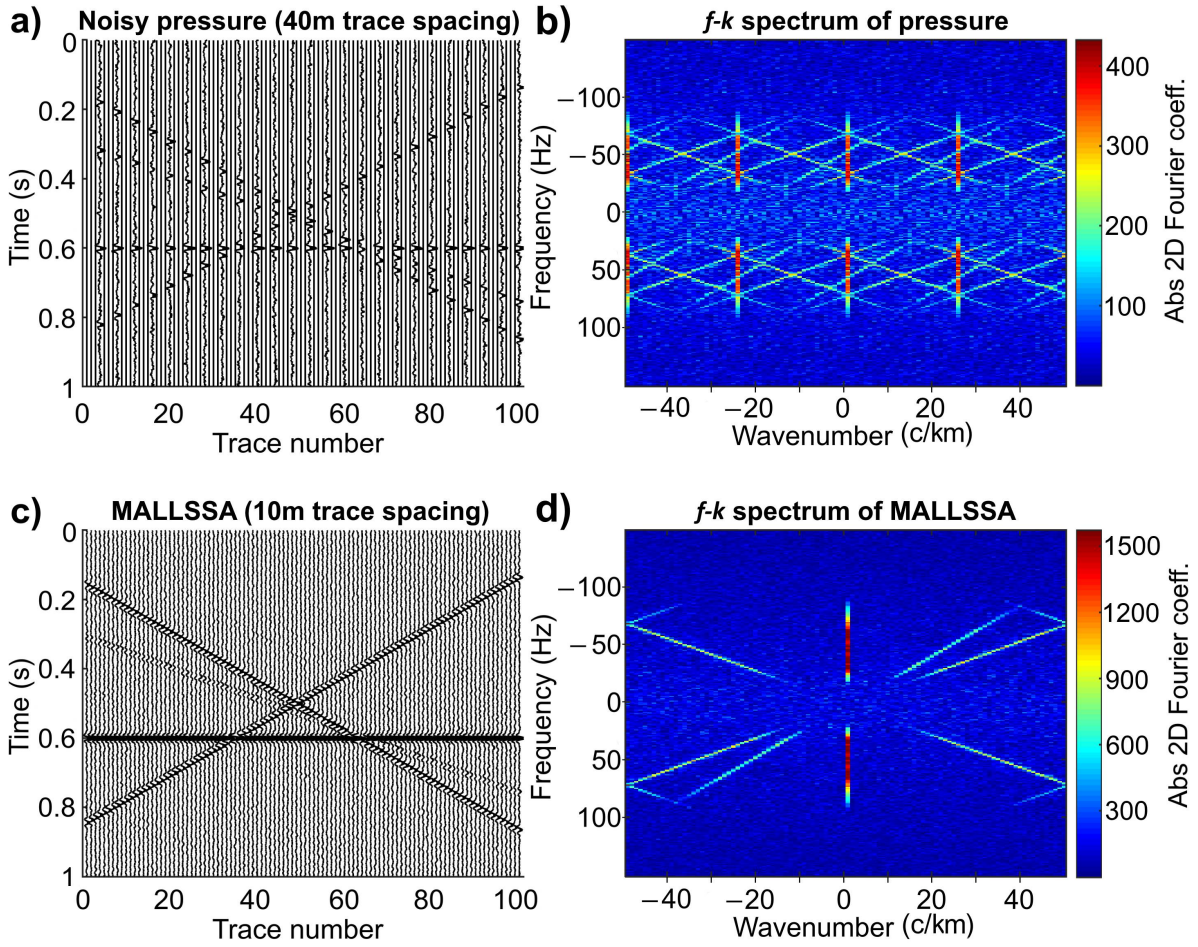

Figure 5.4: (a) The noisy seismic data with 40 metres trace spacing, (b) the *f-k* spectrum of panel a, (c) the MALLSSA result with 10 metres trace spacing, and (d) the *f-k* spectrum of panel c.

positions in the analysis and are not shown here. Also, note that the removed traces are set to zero only to generate the *f-k* spectra. It can be seen from Figure 5.4b that the events are spatially aliased.

The frequency slices (data series) for both pressure and gradient data are obtained, and then the MALLSSA is applied to regularize the frequency slices. Finally, the regularized pressure series from the *f-x* domain are transformed back to the *t-x* domain using the inverse FFT. The real and imaginary parts of the frequency slices can be regularized individually or simultaneously in the MALLSSA. The MALLSSA regularization result is obtained without windowing after one round of iterations, and its *f-k* spectrum are shown in Figure 5.4c and d, respectively. The MALLSSA regularization result clearly constructed all the events in 10 metres trace spacing, and its *f-k* spectrum is approximately the same as the ideal (noise-free) spectrum shown in Figure 5.2b. The MALLSSA regularization result without the random noise is almost the same as Figure 5.2a and b, and so there is no need to illustrate it again.

5.5 Simple synthetic non-stationary seismic data regularization

A synthetic seismic data (pressure) containing three curved and two linear events with different amplitudes and dips is shown in Figure 5.5a. The events are generated using the same Ormsby wavelet as the previous example, and the time sampling rate is 1000 samples per second. Also, random noise is introduced to each trace, and it is assumed that the trace spacing is 10 metres. The *f-k* spectrum of this data is shown in Figure 5.5b. In the *f-k* spectrum, the vertical peaks at zero wavenumber correspond to the horizontal event, and the slanted peaks correspond to the other linear event in Figure 5.5a. The spatial gradient data of the pressure data is shown in Figure 5.5c. The *f-k* spectrum of the spatial gradient data is also shown in Figure 5.5d.

All the traces of even numbers are removed so that the trace spacing is 20 metres (cf., Figure 5.6a). The *f-k* spectrum of the result is shown in Figure 5.6b. Figure 5.6b shows all the events are spatially aliased (e.g.,

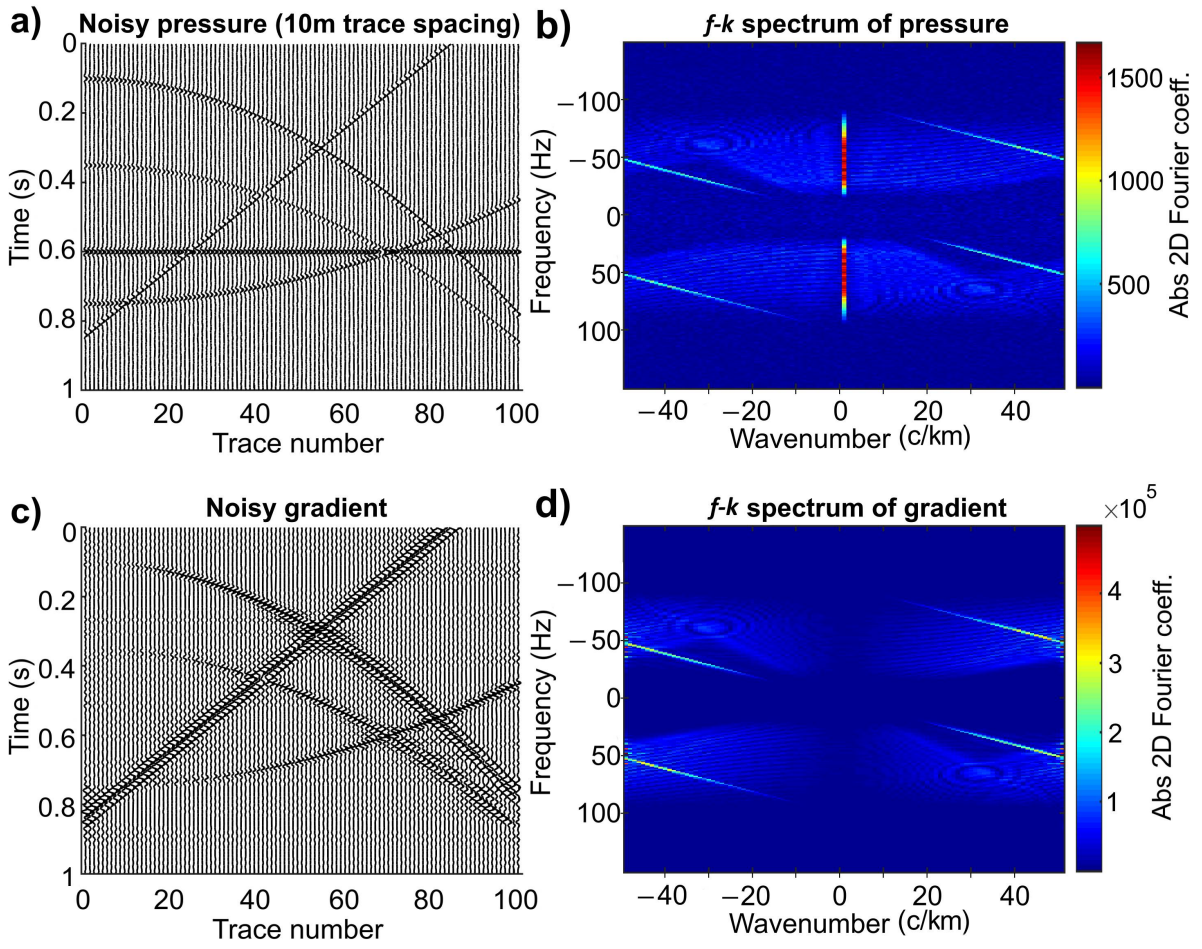

Figure 5.5: Simple synthetic seismic data set. (a) Pressure, (b) the *f-k* spectrum of panel a, (c) gradient, and (d) the *f-k* spectrum of panel c. Note that the vertical peaks at zero wavenumber (reddish peaks) in panel b, corresponding to the horizontal event, disappeared in panel d.

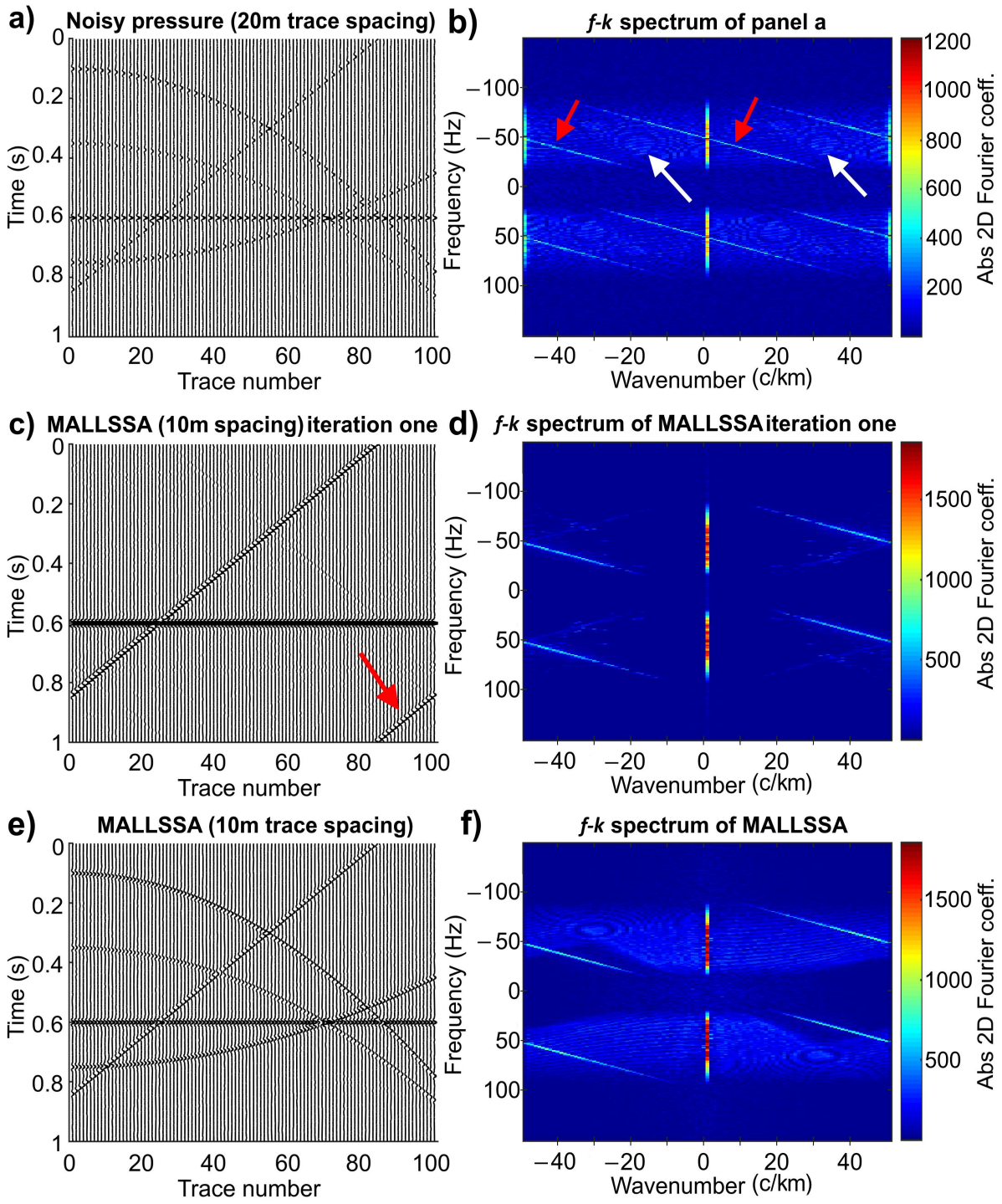

Figure 5.6: (a) The noisy seismic data with 20 metres trace spacing, (b) the *f-k* spectrum of panel a, (c) the MALLSSA result after first iteration with 10 metres trace spacing (the linear events are constructed), (d) the *f-k* spectrum of panel c, (e) the MALLSSA result after a few more iterations with 10 metres trace spacing, and (f) the *f-k* spectrum of panel e. Panel b shows that all the events are spatially aliased (the red and white arrows in panel b show some of the aliasing effects of the linear and curved events, respectively). Arrow in panel c shows some artifacts from the linear event that will be mitigated after a few more iterations (cf., panel e).

the red and white arrows show some of the aliasing effects of the linear and curved events, respectively). Note that the crossline gradients are kept in the samples positions in the analysis and are not shown here. The linear events can be well constructed by the first round of iterations in the MALLSSA (cf., Figure 5.6c and d), yet the curved events cannot be properly constructed and produce artifacts in the traces along with random noise. Arrow in Figure 5.6c shows some artifacts from the linear event that is due to the non-stationary behaviour of the event. However, after a few rounds of iterations the curved events can also be well constructed and interpolated (cf., Figure 5.6e). The *f-k* spectrum of the regularization result is shown in Figure 5.6f that is approximately the same as the one shown in Figure 5.5b. Note that no windowing technique is used in this example.

5.6 A marine seismic data regularization

The field data example is the marine 2D shot gather from a deep water of Gulf of Mexico used in Chapter 4 (Figure 5.7a). The *f-k* spectrum of this data is shown in Figure 5.7b. The spatial gradient data of the marine data is estimated using the ALLSSA as would be approximately acquired in multicomponent seismic acquisition, and it is shown in Figure 5.7c. The *f-k* spectrum of the spatial gradient data is also shown in Figure 5.7d. Note that the normalized wavenumber axis in the *f-k* spectra is obtained by considering the offset of the far offset trace equal to one.

All the traces of even numbers are removed from the marine seismic data and its gradient data. Only the result of marine seismic data and its *f-k* spectrum are illustrated in Figure 5.8a and b, respectively. The effect of aliased events can be clearly seen from the *f-k* spectrum. For better performance, the marine seismic data is divided into two spatial non-overlapping windows of the same size. The MALLSSA regularization result and its *f-k* spectrum are shown in Figure 5.8c and d, respectively. For better comparison between the original marine seismic data and its regularization result, their difference is calculated and illustrated it along with its *f-k* spectrum in Figure 5.8e and f, respectively. A mask is used to set the values (minor artifacts) of the north-east part of the images shown in Figure 5.8c and e to zero after the interpolation. As it can be seen, the final regularization result is pleasant and acceptable for practical applications.

Note that in this example, the traces were equally spaced and removing the traces of even numbers also makes the remaining traces equally spaced with a Nyquist wavenumber that is half of the original Nyquist wavenumber. When the distance between traces varies, the ALLSSA and MALLSSA perform better because these methods consider the correlations among the sinusoidal basis functions of various wavenumbers.

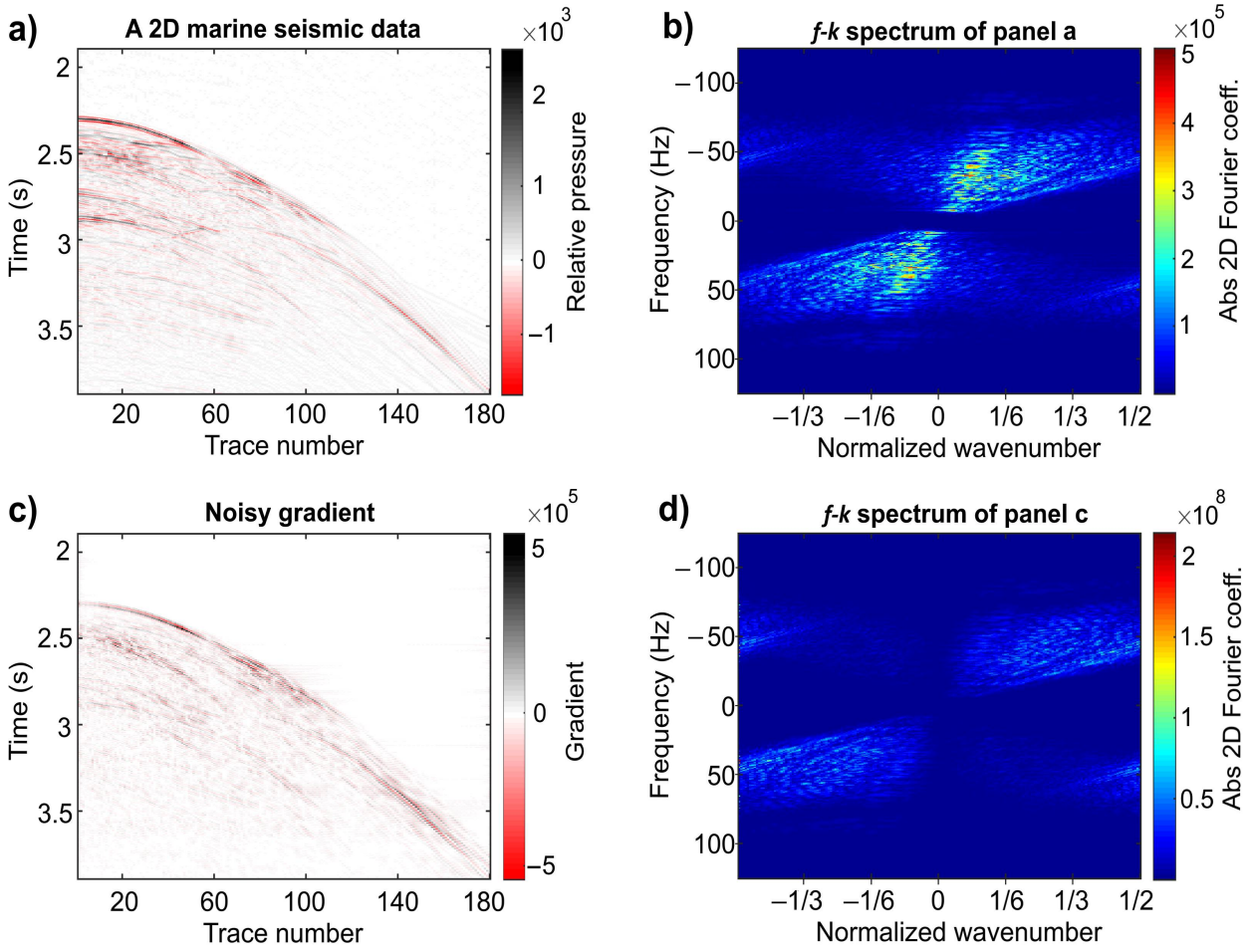

Figure 5.7: (a) A 2D marine seismic data obtained from a deep water of Gulf of Mexico, (b) the *f-k* spectrum of panel a, (c) the estimated gradient data, and (d) the *f-k* spectrum of panel c.

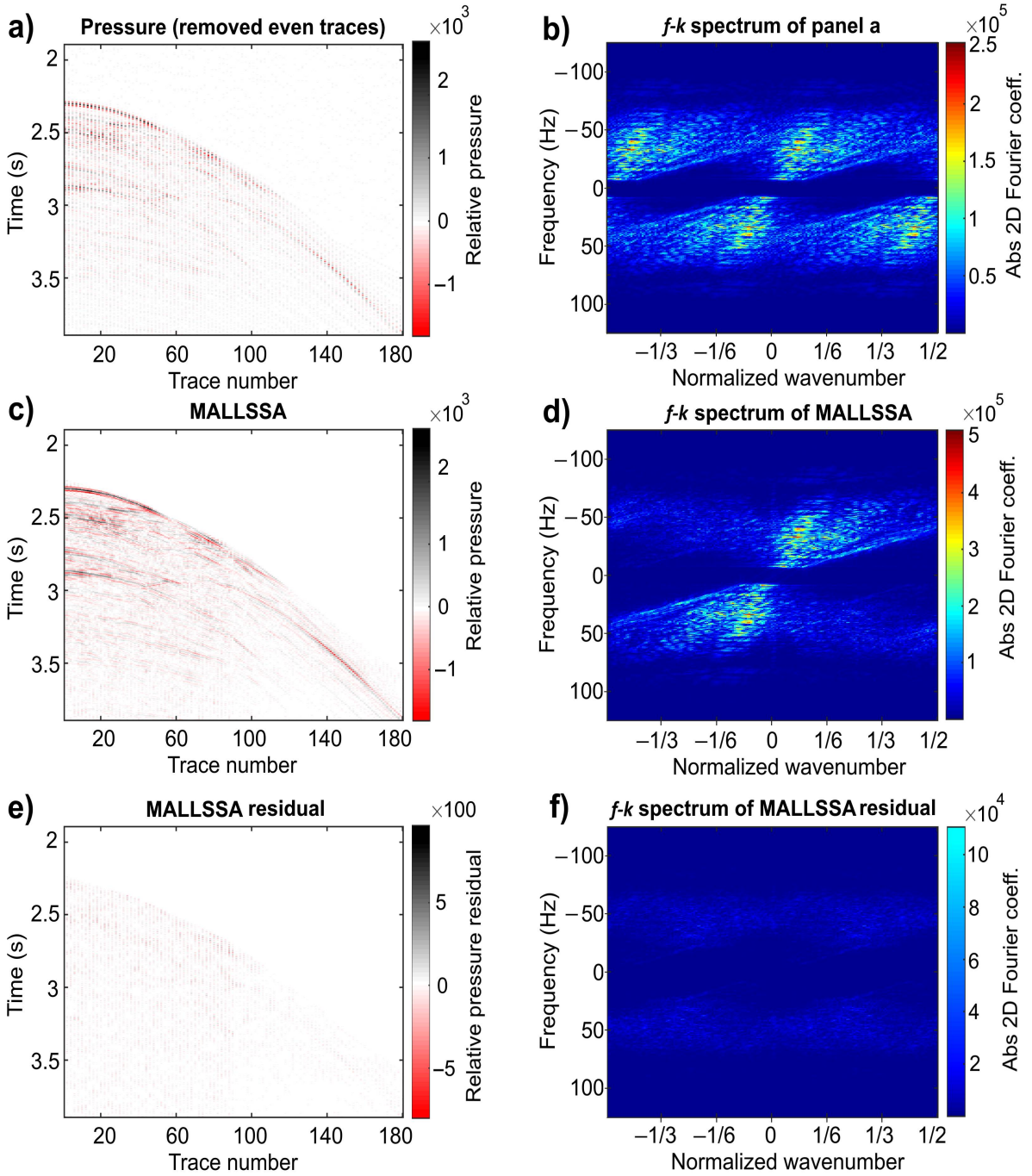

Figure 5.8: (a) The traces of even numbers in Figure 5.7a are removed, (b) the *f-k* spectrum of panel a, (c) the MALLSSA regularization result, (d) the *f-k* spectrum of panel c, (e) the difference (residual) between panel c and Figure 5.7a, and (f) the *f-k* spectrum of panel e.

5.7 Summary

The ALLSSA is a robust method of regularizing seismic data presenting mild aliasing in the f-k spectrum. However, when the seismic data is coarsely sampled, additional information is needed to aid regularization. This is crucial in regularization because incorrect selection of the spectral peaks results in inaccurate regularization. I discovered that the spectral peaks of the wide-sense stationary seismic data can be accurately estimated by incorporating the spatial gradients of the seismic data into the ALLSSA model and showed its robustness on synthetic data series and seismic data. The MALLSSA is designed for wide-sense stationary data series; however, an appropriate windowing technique can be used to adapt it to seismic data with more complex geologic structure.

The ALLSSA and MALLSSA consider the correlations between the sinusoids of different wavenumbers simultaneously along with the constituents of known forms and covariance matrices associated with data series. Therefore, the ALLSSA and MALLSSA estimate the signal wavenumbers and amplitudes more accurately than the IMAP and MIMAP. The ALLSSA and MALLSSA may be considered as extensions of the IMAP and MIMAP, respectively.

6 Least-Squares Wavelet Analysis and Its Applications

Analyzing unequally spaced and non-stationary time series is a challenging problem in many areas, such as geodesy and geophysics. Many time series present datum shifts and/or have trends and have uncertainties (variances, covariances) associated with them. The current methods of time series analyses, such as the FFT, DFT, STFT, CWT, CLSSA, HHT, and WWZ, have several limitations and cannot explicitly consider the constituents of known forms and covariance matrices associated with time series. In this chapter, a novel method is developed that can analyze any time/data series (to study the periodicity of the components and detect hidden signatures) without any assumptions and preprocessing of the original time/data series.

Conceptually, this method, namely, the Least-Squares Wavelet Analysis (LSWA), decomposes a time series into the time-frequency domain by an appropriate segmentation of the series and calculates spectral peaks based on the least-squares fit of sinusoids (other functions may be chosen as well) to each segment. The segmentation is similar to the one in the CWT, and computing the spectral peaks is similar to the LSSA, so the approach is a combination of the CWT and the LSSA. The LSWA explicitly considers the constituents of known forms, such as trends and/or datum shifts, as well as the covariance matrices associated with the time series and obtains reliable spectrograms that can be statistically assessed by stochastic surfaces (confidence level surfaces) defined for spectrograms. This chapter is an unabridged version of a paper published in Mathematical Geosciences (Ghaderpour and Pagiatakis, 2017).

6.1 Theoretical Part of the Least-Squares Wavelet Analysis

Suppose that $\mathbf{f} = \big[f(t_j)\big]$, $1 \leq j \leq n$, is a discrete time series of n data points; here the t_j's are not necessarily equally spaced. Let $\mathbf{\Omega} = \{\omega_k;\ k = 1, \ldots, \kappa\}$ be a set of spectral frequencies, and let $\mathbf{y}$ be an arbitrary segment of the time series that contains $L(\omega_k)$ samples such that ω_k can be resolved within it

$$\mathbf{\Phi} = \Big[\cos\big(2\pi\omega_k t_{i+j-\frac{1}{2}(L(\omega_k)+1)}\big),\ \sin\big(2\pi\omega_k t_{i+j-\frac{1}{2}(L(\omega_k)+1)}\big)\Big] \tag{6.1}$$

$$\mathbf{y} = \Big[f\big(t_{i+j-\frac{1}{2}(L(\omega_k)+1)}\big)\Big]; \tag{6.2}$$

where $1 \leq i \leq L(\omega_k)$ and

$$L(\omega_k) := \begin{cases} \left\lfloor \dfrac{L_1 M}{\omega_k} \right\rfloor + L_0 & \text{if } \left\lfloor \dfrac{L_1 M}{\omega_k} \right\rfloor + L_0 \text{ is odd,} \\ \left\lfloor \dfrac{L_1 M}{\omega_k} \right\rfloor + L_0 + 1 & \text{otherwise,} \end{cases} \tag{6.3}$$

where symbol $\lfloor * \rfloor$ indicates the floor function that gives the largest integer no greater than $*$, L_0 is a fixed number of data points, L_1 is a selected number of cycles of the sinusoidal base functions, M is the average number of data points per unit time, and ω_k is the number of cycles per unit time ($\omega_k \in \mathbf{\Omega}$). For instance, for an equally spaced time series recorded at a rate of one sample per millisecond, if frequency is in Hertz, then $M = 1000$ (data points per second), and if $L_1 = 2$, then two cycles of sinusoidal base functions of cyclic frequency ω_k will be fitted to a segment of the time series with $L(\omega_k)$ data points. Also, L_0 is the selected number of additional samples considered in the least-squares fit to achieve the desired time and frequency resolution in the spectrogram (see Section 6.2 for more details).

For simplicity of notations, L is used instead of $L(\omega_k)$ from now on. For each k, $1 \leq k \leq \kappa$, and each j, $1 \leq j \leq n$, if $1 \leq i + j - (L+1)/2 \leq n$, then the size of $\mathbf{y}$ in Equation (6.2) which is the same as the number of rows in $\mathbf{\Phi}$ given by Equation (6.1) is

$$R(j,k) := \begin{cases} (L-1)/2 + j & \text{if } 1 \leq j < (L+1)/2, \\ L & \text{if } (L+1)/2 \leq j \leq n - (L-1)/2, \\ (L+1)/2 + n - j & \text{if } n - (L-1)/2 < j \leq n. \end{cases} \tag{6.4}$$

Note that when $L_1 = 0$, $R(j,k)$ is independent from the choice of ω_k. Assume that $\mathbf{P} = {\mathbf{C_f}}^{-1}$ is the associated weight matrix of $\mathbf{f}$, and $\mathbf{P_y}$ is the principal submatrix of $\mathbf{P}$ of dimension $R(j,k)$ (see below for more details).

If $\mathbf{f}$ contains constituents of known forms $(\phi_1, \ldots, \phi_q)$ but of unknown amplitudes (e.g., trends or sinusoids of constant frequencies), then for each (t_j, ω_k), let

$$\underline{\mathbf{\Phi}} = \left[\phi_1(t_{i+j-\frac{1}{2}(L+1)}), \ldots, \phi_q(t_{i+j-\frac{1}{2}(L+1)}) \right], \tag{6.5}$$

$$\overline{\mathbf{\Phi}} = \left[\underline{\mathbf{\Phi}} \;\; \mathbf{\Phi} \right], \tag{6.6}$$

where $\mathbf{\Phi}$ is given by Equation (6.1). Model $\mathbf{y} = \overline{\mathbf{\Phi}}\, \overline{\mathbf{c}}$ is used to determine $\hat{\overline{\mathbf{c}}}$ using the least-squares method, so

$$\hat{\overline{\mathbf{c}}} = {\overline{\mathbf{N}}}^{-1} \overline{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{P_y} \mathbf{y}, \tag{6.7}$$

where T signifies transpose and $\overline{\mathbf{N}} = \overline{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{P_y} \overline{\mathbf{\Phi}}$ (matrix of normal equations). Let $\overline{\mathbf{J}} = \overline{\mathbf{\Phi}}\, {\overline{\mathbf{N}}}^{-1} \overline{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{P_y}$. Define the weighted Least-Squares Wavelet Spectrogram (LSWS) as

$$\overline{s}(t_j, \omega_k) = \frac{\mathbf{y}^{\mathrm{T}} \mathbf{P_y} \overline{\mathbf{J}} \mathbf{y}}{\mathbf{y}^{\mathrm{T}} \mathbf{P_y} \mathbf{y}}, \tag{6.8}$$

where $1 \leq j \leq n$, $1 \leq k \leq \kappa$, and $\overline{s}(t_j, \omega_k) \in (0, 1)$. The justification of the name comes from the fact that when $L_1 \neq 0$, the size of $\mathbf{y}$ decreases as frequency increases in the calculation of the spectrogram using the least-squares method similar to the CWT. Notice that if $\mathbf{f}$ is equally spaced, equally weighted and $\overline{\mathbf{\Phi}} = \mathbf{\Phi}$, then $\hat{\overline{\mathbf{c}}}^{\mathrm{T}} \hat{\overline{\mathbf{c}}}$ has similar properties to the STFT with a rectangular window when $L_1 = 0$, and it has similar properties to the CWT when $L_1 \neq 0$. However, Equation (6.8) shows how much of $\overline{\mathbf{\Phi}}\ \hat{\overline{\mathbf{c}}}$ is contained in $\mathbf{y}$.

A more practical approach used in this contribution is to first remove (suppress) the known constituents from each segment $\mathbf{y}$, and then analyze the residual segment $\hat{\mathbf{g}}$. More precisely, first use the model $\mathbf{y} = \underline{\mathbf{\Phi}}\ \underline{\mathbf{c}}$ to estimate $\underline{\mathbf{c}}$ as

$$\hat{\underline{\mathbf{c}}} = \underline{\mathbf{N}}^{-1} \underline{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \mathbf{y}, \tag{6.9}$$

where $\underline{\mathbf{N}} = \underline{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \underline{\mathbf{\Phi}}$, and so $\hat{\mathbf{g}} = \mathbf{y} - \underline{\mathbf{\Phi}}\ \hat{\underline{\mathbf{c}}}$. Then use the model $\mathbf{y} = \overline{\mathbf{\Phi}}\ \overline{\mathbf{c}} = \underline{\mathbf{\Phi}}\ \underline{\mathbf{c}} + \mathbf{\Phi}\ \mathbf{c}$ to estimate $\mathbf{c}$ as

$$\hat{\mathbf{c}} = \mathbf{N}^{-1} \mathbf{\Phi}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \hat{\mathbf{g}}. \tag{6.10}$$

where $\mathbf{N} = \mathbf{\Phi}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \mathbf{\Phi} - \mathbf{\Phi}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1} \underline{\mathbf{\Phi}}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \mathbf{\Phi}$ (see Appendix B). Let $\mathbf{J} = \mathbf{\Phi} \mathbf{N}^{-1} \mathbf{\Phi}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}}$. Define the weighted LSWS as

$$s(t_j, \omega_k) = \frac{\hat{\mathbf{g}}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \mathbf{J} \hat{\mathbf{g}}}{\hat{\mathbf{g}}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}} \hat{\mathbf{g}}}, \tag{6.11}$$

where $1 \leq j \leq n$ and $1 \leq k \leq \kappa$. Note that $s(t_j, \omega_k) \in (0, 1)$. This spectrogram shows how much the sinusoids of cyclic frequency ω_k contributes to each residual segment of $\mathbf{f}$ and called percentage variance when multiplied by 100. This definition is very useful in searching for hidden signals (short duration signals with very low amplitude with respect to the total amplitude of the data points) in a time series. The ordinary LSWS does not consider the weight matrix $\mathbf{P}_{\mathbf{y}}$ in the calculation of spectrogram.

To better understand how to determine the weight matrix $\mathbf{P}_{\mathbf{y}}$ for each segment $\mathbf{y}$, a simple example is provided. Let $\mathbf{P}$ be of order $n = 10$, and $M = 10$, $\omega_1 = 2$, $L_1 = 1$ and $L_0 = 0$. It is easy to see that $L(\omega_1) = 5$, and so for $j = 1$ and $j = 10$, the top left and the bottom right principal submatrices of $\mathbf{P}$ (margins) have dimension 3 (green matrices in Figure 6.1), and the submatrices shown by blue have dimension 5. Note that the correlations between data points in the entire time series are considered in each $\mathbf{P}_{\mathbf{y}}$ due to the inversion of the entire $\mathbf{C}_{\mathbf{f}}$. In practice, $\mathbf{P}$ is usually diagonal that means the correlations of the time series values are not given. Note that $\mathbf{P}_{\mathbf{y}}$ is a positive definite matrix because $\mathbf{P}$ is positive definite (Harville, 2008), and so Equations (6.7), (6.9) and (6.10) are optimal.

The known constituents contaminate the spectrogram given by Equation (6.8), but they do not contaminate the spectrogram given by Equation (6.11) because they are removed from the segments of the time series. However, Equation (6.8) shows how much the constituents of known forms and the sinusoids of a particular frequency simultaneously contribute to the segments. Note that the Least-Squares Spectrum

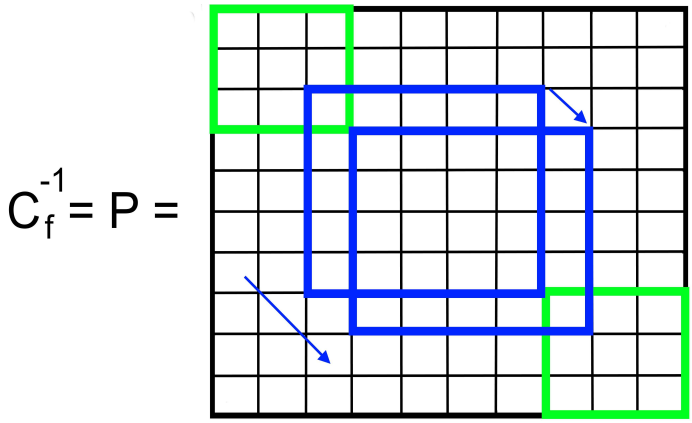

Figure 6.1: The weight matrix **P** with a few of its principal submatrices.

(LSS) is a special case of the LSWS that is independent of time and has only one segment (the entire time series). The reader is referred to Wells et al. (1985); Pagiatakis (1999) for more details.

6.2 Practical Implementation of the Least-squares Wavelet Algorithm

To visualize the concept of the segmentation of a time series better, the term window is used. In this contribution, two different characteristics for the window are used, namely, its size and its length. For a time series of n data points, a window of size $R(j,k)$ given by Equation (6.4) located at t_j, $1 \leq j \leq n$, is a window that contains $R(j,k)$ data points comprising $(L-1)/2$ data points on either side of t_j. If $1 \leq j < (L+1)/2$, then a window of size $R(j,k)$ located at t_j is a window that contains $(L-1)/2$ data points on the right hand side and $j-1$ data points on the left hand side of t_j; this window is referred to as the left marginal window. Similarly, one can identify the number of data points for the right marginal windows. From Equation (6.1), when j runs from j_0 to j_1, a window of size $R(j_0,k)$ located at t_{j_0} translates to a window of size $R(j_1,k)$ located at t_{j_1}. In the LSWA, at each step, the sinusoids and the constituents of known forms simultaneously are fitted to each segment of the time series within a window of size $R(j,k)$ located at t_j to estimate their amplitudes and phases in the segment.

The length of a window located at t_j, denoted by $\ell(t_j)$, can be calculated as follows. If $(L+1)/2 \leq j \leq n-(L-1)/2$, then

$$\ell(t_j) = t_{j+(L-1)/2} - t_{j-(L-1)/2}. \tag{6.12}$$

If $1 \leq j < (L+1)/2$ (the left marginal windows), then

$$\ell(t_j) = t_{j+(L-1)/2} - t_1, \tag{6.13}$$

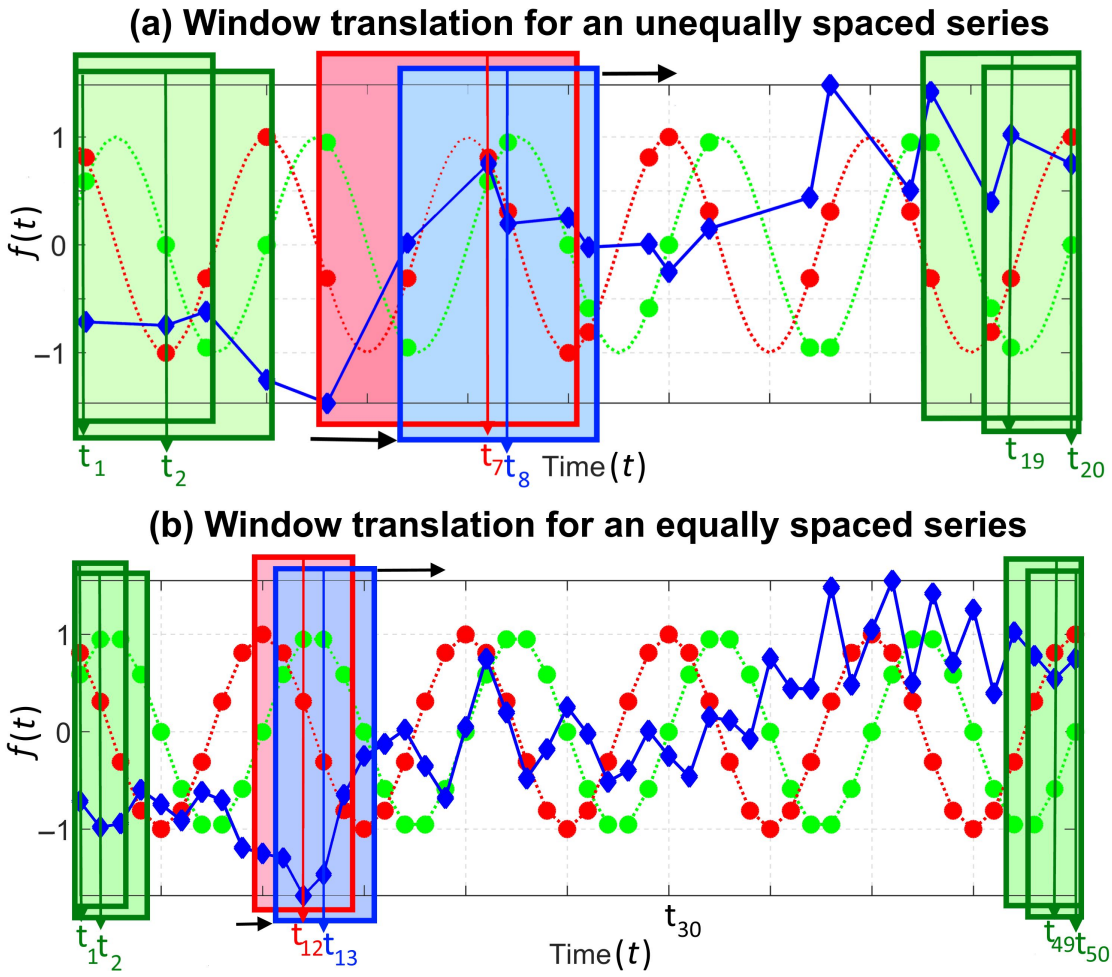

Figure 6.2: (a) An unequally spaced and (b) an equally spaced time series (blue diamonds) along with sinusoids (cosine with red circles and sine with green circles) and several windows.

and if $n - (L-1)/2 < j \leq n$ (the right marginal windows), then

$$\ell(t_j) = t_n - t_{j-(L-1)/2}. \tag{6.14}$$

From Equations (6.12), (6.13) and (6.14), it is clear that the length of the window located at t_j depends on the distribution of the t_i's.

In order to illustrate these concepts, an unequally spaced and an equally spaced times series along with the sine and cosine base functions of cyclic frequency ω_k are illustrated in Figure 6.2a and b, respectively. In Figure 6.2a, the red window is of size 5 (includes 5 data points) located at t_7, the median time mark within the window. Similarly, the blue window is of size 5 located at t_8 which is again the median of the window. The length of the red window is $\ell(t_7) = t_9 - t_5$, and the length of the blue window is $\ell(t_8) = t_{10} - t_6$. In this case $\ell(t_7) > \ell(t_8)$. For the equally spaced time series (Figure 6.2b), it is clear that the red and blue

windows have the same size and length. The left and right marginal windows with the green windows and the translations are shown in these figures.

For an unequally spaced time series, the window length varies but the window size remains constant during translation except for the marginal windows. For unequally spaced time series, if the length of the window is fixed during translation rather than its size, then there may not exist enough data points within the window located at t_j to calculate $s(t_j, \omega_k)$. Thus, it is more appropriate to fix the size of the translating window. Since in Equation (6.1), j runs from 1 to n, one unit at a time, the window located at t_j overlaps with the window located at t_{j+1} by $L-1$ data points when $(L+1)/2 \leq j \leq n-(L-1)/2$ (red and blue windows in Figure 6.2a and b). The overlaps for the marginal windows are also shown in Figure 6.2a and b (green windows).

Figure 6.3a and b show the changes in the window size for an equally spaced time series when $L_1 = 1$ cycle with a fixed L_0. The window size decreases when the frequency increases allowing the detection of short duration waves (signal, transients, etc.), and L_0 (a constant number of samples) increases the size of the window so that the resolution may be adjusted based on the specific scope of analysis.

For a time series (equally or unequally spaced), the sinusoidal functions (and the constituents of known forms) are generally not orthogonal within a window of size $R(j,k)$ or length $\ell(t_j)$ located at t_j. The LSWA considers their correlations by the off-diagonal elements of the matrix of normal equations (similar to the LSSA). For each t_j, Equation (6.11) shows how much the sinusoids of cyclic frequency ω_k contribute to the residual segment of the time series within a window of size $R(j,k)$ or length $\ell(t_j)$ located at t_j.

Puryear et al. (2012) simultaneously include the sinusoidal functions with different frequencies in the columns of design matrix $\mathbf{\Phi}$, which makes the system $\mathbf{y} = \mathbf{\Phi c}$ underdetermined, and then they use constraints to calculate the spectrogram (see the CLSSA described in Chapter 2). However, the LSWS is calculated by trying each cyclic frequency ω_k one at a time, so the system is highly overdetermined, and the matrix of normal equations remains invertible by selecting appropriate L_0 and L_1, window size parameters.

When analyzing an unequally spaced time series, spectral leakages (peaks that do not correspond to signals) appear in the LSS and LSWS. However, the leakages are far less than the leakages in the Fourier transform because the correlation among the sinusoids is considered for each frequency in the calculation of the LSS and LSWS. Since the frequencies are examined one at a time (out-of-context), the leakages appear in the LSS and LSWS. By considering the correlation among the sinusoids of different frequencies (the frequencies of the constituents of known forms), the spectral leakages will be mitigated (Craymer, 1998), and this can be achieved by simultaneously removing the constituents of known forms from the segments.

The selection of parameters L_0 and L_1 for the analysis of a time series mainly depends on whether the time series is weakly or strongly non-stationary. In addition, it depends on the time and frequency resolution as well as the number of constituents being estimated within the window whose size should be adequate to

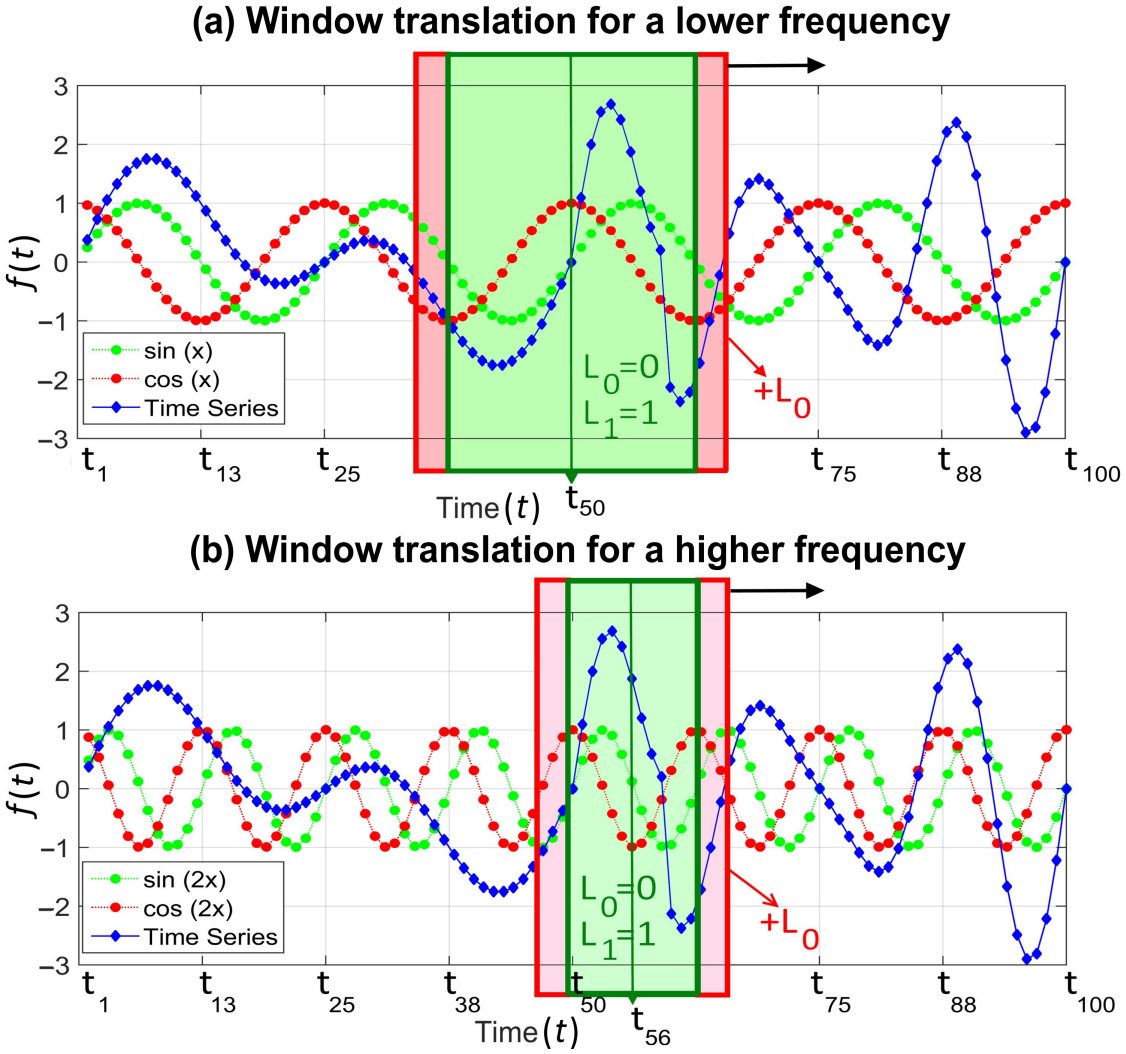

Figure 6.3: Window translation of an equally spaced time series (blue) for (a) $\sin(x)$ and $\cos(x)$, and (b) $\sin(2x)$ and $\cos(2x)$.

avoid singularity of the matrix of normal equations in a desired frequency band. In many applications, for time series with series more than 50 data points per unit time, it is recommended that L_0 be 20 to 30 samples, and L_1 be 3 or 4 cycles to avoid the singularity of $\underline{\mathbf{N}}$ and $\mathbf{N}$ when removing the constituents of known forms (usually less than 10). In practical applications, especially for unequally spaced time series, choosing $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{t}]$ makes the mean values of the segments approximately zero because even if there is no systematic trend present in an unequally spaced time series, the sine and cosine functions no longer have zero mean causing an error in determining the zero point of the signals and reducing the percentage variance of the signal peaks in the spectrogram.

6.3 Comparison Between the Least-Squares Wavelet Spectrogram and the Weighted Wavelet Z-Transform

The least-squares method used in the LSWS and WWZ is a great tool that estimates the exact amplitude of a physical fluctuation much more accurately than the traditional Fourier transform. The amplitude of signals in the LSWA can be estimated using Equation (6.10) after finding the frequency or period of the signals using the LSWS that is similar to the Weighted Wavelet Amplitude (WWA) (Foster, 1996). Both the LSWS and WWZ are excellent in frequency determination of the signals; however, the WWZ is a poor estimator of amplitude (Foster, 1996), and the spectral peaks in the WWZ lose their power toward higher frequencies similar to the CWT. The latter shortcoming is caused mainly by the effective number of data points that decreases when the frequency increases, making the numerator of Equation (2.23) smaller. Note that the term $V_y/(V_x - V_y)$ in Equation (2.23) is the estimated signal-to-noise ratio (Foster, 1996), and the numerator of Equation (6.11) is the estimated signal and its denominator is the sum of the estimated signal and noise (Pagiatakis, 1999). Subtracting V_y from V_x in the denominator of Equation (2.23) alters significantly the true power of the signals that may not allow one to see the behaviour of the time series, especially when searching for possible hidden signals with low power in a time series.

In both the LSWS and WWZ, the locations of the windows can be equally spaced; however, the locations are chosen to be at the times corresponding to the time series values simply because there might be signatures in the time series gaps that cannot be detected using the available data points. An unequally spaced pure sine wave of cyclic frequency 10 cycles per annum (c/a) and amplitude 3 is generated by randomly selecting 100 data points per year ($M = 100$) and demonstrated in Figure 6.4a. The ordinary LSWS is shown in Figure 6.4b with $L_1 = 4$, $L_0 = 10$ and $\mathbf{\Phi}_1 = [\mathbf{1}]$. The percentage variance of the peak at 10 (c/a) is 100% at all times. Note that the spectral leakages in the spectrogram are not significant at certain confidence level defined for the distribution of the LSWS (Section 6.4). Figure 6.4c shows the WWZ with $c = 0.0125$. The peaks at the cyclic frequency 10 (c/a) have variable power over time, but the spectral leakages are mitigated (they exist but with lower power) because of the subtraction in the denominator of Equation (2.23).

The LSWA can detect and suppress significant peaks (that takes into account the correlations among the sinusoids of different frequencies), and so it can be used to search for the possible hidden signals in a time series. Suppressing the significant peaks simultaneously mitigates the spectral leakages in the LSS and LSWS of the unequally spaced residual series and results in strengthening the peaks of possible hidden signals. Another advantage of using the ordinary LSWS is the computational efficiency that is much better than the WWZ because the weights are not involved in the computation. Moreover, the window size in the LSWS decreases when frequency increases; whereas, the window size in the WWZ is the same at all times and frequencies (Equations (2.20) and (2.21)), fact that significantly increases the computational cost.

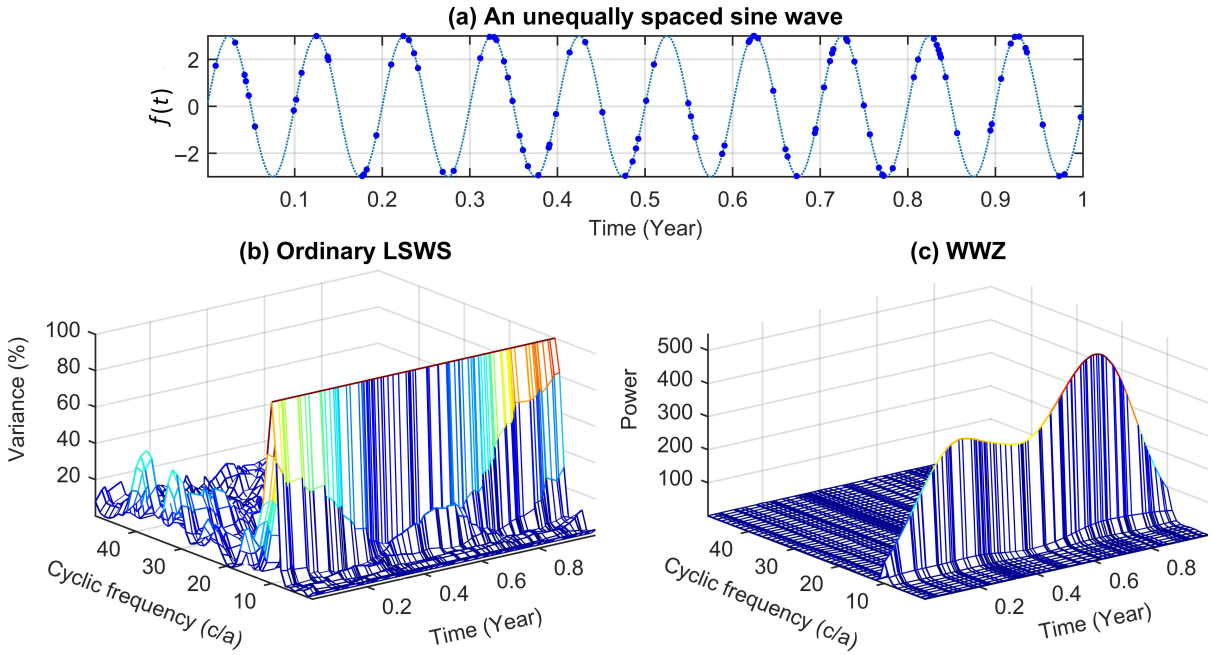

Figure 6.4: An unequally spaced pure sine wave of cyclic frequency 10 (c/a) and its analyses.

6.4 Stochastic Surfaces in the Least-Squares Wavelet Spectrograms

In this section, a similar methodology is used as in Pagiatakis (1999) to determine stochastic surfaces for the Least-Squares Wavelet Spectrogram (LSWS) that define confidence levels (usually 95% or 99%) above which spectral peaks in the LSWS are statistically significant. It is also shown that the stochastic surfaces of the LSWS are independent of the a-priori variance factor σ_0^2, that is, the stochastic surfaces are independent of any scale defect of the covariance matrix associated with the time series. The reader is referred to Vaníček and Krakiwsky (1986) for more details on the variance factor.

A remarkable property of the least-squares wavelet spectrogram defined by Equations (6.8) and (6.11) is that it provides the percent of a specific component(s) (as defined by $\overline{\mathbf{\Phi}}$) contained in $\mathbf{f}$. This ratio includes the weight matrix in both numerator and denominator and as such, any scale defect in the covariance matrix cancels out. This leads to the conclusion that the Probability Distribution Function (PDF) of the spectrogram does not depend on either the a-priori or a-posteriori variance factors. Therefore, the derivation of the PDF of the spectrogram will be based on previous work by Pagiatakis (1999) considering only the case of a-priori variance factor being known since the treatment of the case of unknown a-priori variance factor (a-posteriori variance factor used) was in error.

Suppose that $\mathbf{f}$ has been derived from a population of random variables following the multidimensional normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{C_f})$, where $\mathbf{C_f}$ may be singular. It is shown in Appendix C that the LSWS given

by Equation (6.11) follows the beta distribution with parameters 1 and $\Re/2$, where $\Re = R(j,k) - q - 2$, and so the stochastic confidence level surface at significance level α is $\zeta(t_j, \omega_k) = 1 - \alpha^{2/\Re}$.

Following similar techniques in Pagiatakis (1999), it is not difficult to see that the spectrogram given by Equation (6.8) follows $\beta_{(q+2)/2,\Re/2}$ with the mean $(q+2)/(\Re+q+2)$ and variance $2(q+2)\Re/\big((\Re+q+2)^2(\Re+q+4)\big)$; however, in practice, the spectrogram given by Equation (6.11), following $\beta_{1,\Re/2}$ with the mean $2/(\Re+2)$ and variance $4\Re/\big((\Re+2)^2(\Re+4)\big)$, provides better results. The reader is referred to Johnson et al. (1995) for more details on the mean and variance of random variables following the beta distribution.

In the following sections, whenever the CWT is applied for analysis, the MATLAB command cwt($\mathbf{f}$, scales, wname) is used, where $\mathbf{f}$ is time series values, scales is vector of positive scales for the CWT scalogram, and wname is name of a wavelet (e.g., morl for Morlet wavelet). One may convert the scales to frequencies using centre frequencies of scales by calling MATLAB command scal2frq(scales, wname, DELTA) to obtain a spectrogram, where DELTA is sampling period. This is not a correct approach of analyzing unequally spaced time series. One of the advantages of the LSWA over the CWT is that the LSWA does not convert any scale to frequency to obtain a spectrogram, and it directly decomposes a time series into the time-frequency domain.

6.5 Synthetic Equally Spaced Time Series

In this example, the sensitivity of the LSWS for different window size parameters is examined. Consider the following hyperbolic chirp signal (Mallat, 1999)

$$f(t_j) = \cos\Big(\frac{230}{2.3 - t_j}\Big), \tag{6.15}$$

where $t_j = (0.001j)$s, and j runs from 1 to 2000 (Figure 6.5a). Note that the instantaneous frequency of the hyperbolic chirp is $230/\big(2\pi(2.3-t_j)^2\big)$ Hz. Since there are 1000 data points per second, $M = 1000$.

Figure 6.5b shows the ordinary LSWS with $L_0 = 201$ and $L_1 = 0$. It can be seen that the true peaks at higher frequencies are not well resolved (the peaks lose their percentage variance in higher frequencies, arrows). The reason is that the window size is fixed to 201 (except for the marginal windows), and the time series is changing its frequency rapidly as time advances, so the sinusoidal base functions do not fit the segments of the time series at higher frequencies as well as they do at lower frequencies.

The ordinary LSWS with $L_0 = 0$ and $L_1 = 1$ is shown in Figure 6.5c. Although the percentage variance of the true peaks in the spectrogram does not change across the time-frequency domain, the spectrogram has low-frequency resolution at high frequencies. Furthermore, the spikes appear in the spectrogram at the times when the series has zero magnitude (arrows) because one cycle of sinusoids also fits those segments that cross the horizontal axis. These spikes may also have applications as they show the behaviour of the time series.

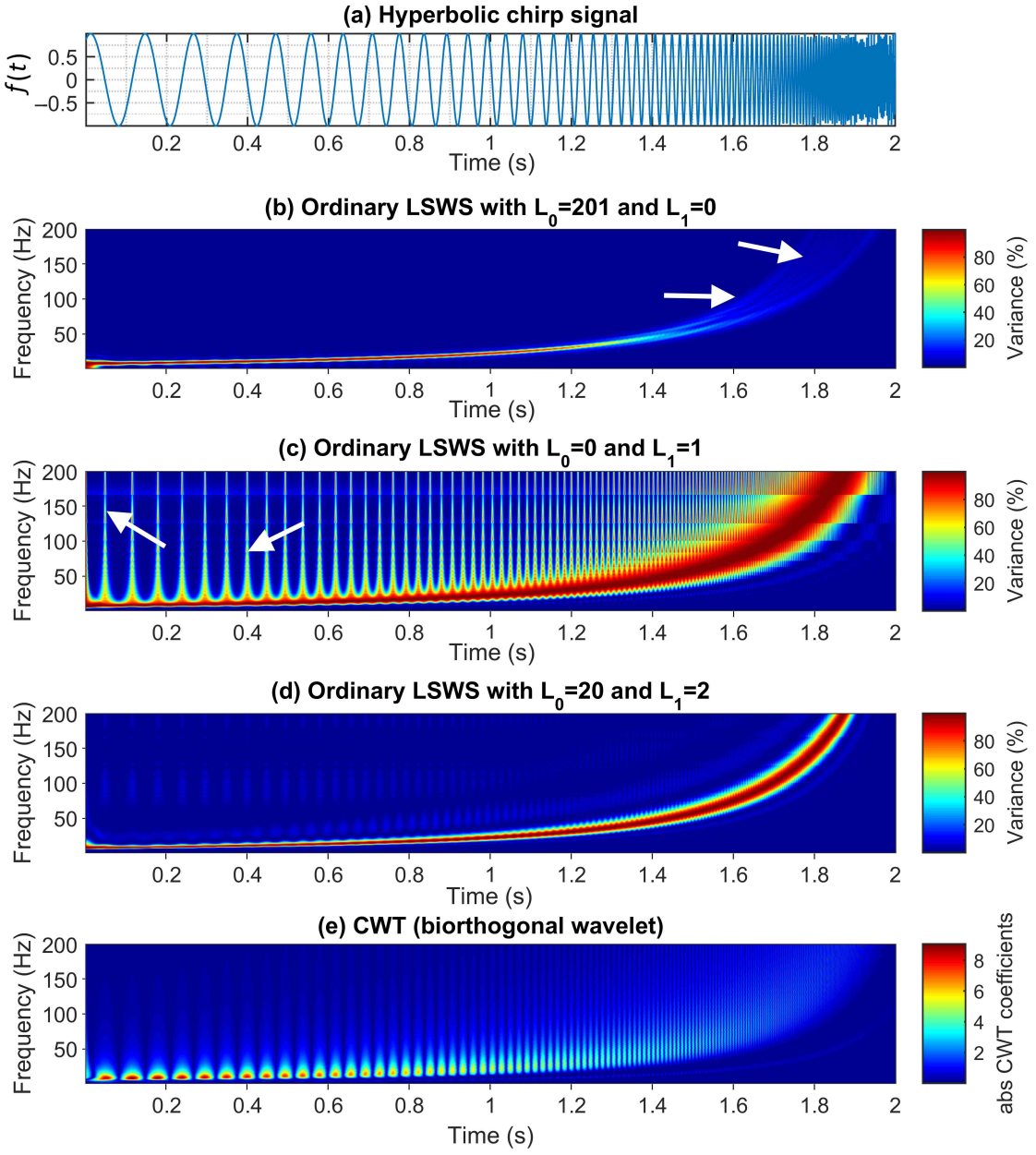

Figure 6.5: The equally spaced chirp signal given by Equation (6.15) and its analyses.

If one selects $L_0 = 20$ and $L_1 = 2$, then the percentage variance of the spikes reduces because two cycles and many additional data points are incorporated in the fitting process resulting in a better resolution (Figure 6.5d). In fact, the selection of the window size parameters is based on the purpose of analysis that is similar to the selection of window size in the STFT. The magnitude of $\hat{\mathbf{c}}$ in Equation (6.10) is approximately 1 at the signal peaks.

Figure 6.5e shows the CWT after applying the MATLAB command cwt($\mathbf{f}$, scales, wname) with wname=bior1.3, a biorthogonal wavelet that has similar structure to the sinusoids (Cohen et al., 1992). Note that the scalogram is converted to spectrogram using the centre frequencies of scales. The colour bar values represent the absolute values of the CWT coefficients (abs CWT coefficients). The higher frequencies in the CWT spectrogram are not well resolved and lose power (indicating that the amplitude of the chirp signal is decreasing over time) because the spectrogram is computed in terms of frequency bands (i.e., scales) that overlap each other more when frequency increases (Sinha et al., 2005).

The spikes in the CWT spectrogram appear at the same positions as in the LSWS. However, since the LSWS is normalized and the correlation among sinusoids is taken into account, the spikes in the LSWS are sharper. Another advantage of the LSWS is that the true peaks are not scattered as the CWT spectrogram because the least-squares minimization is used in translation. Note that the LSWS shows how much the sinusoids of a particular frequency within a window contributes to the segment of the chirp signal within the window.

Now consider the following equally spaced time series which is the sum of two hyperbolic chirp signals

$$f(t_j) = \cos\Big(\frac{230}{2.3 - t_j}\Big) + \cos\Big(\frac{500}{2.3 - t_j}\Big), \tag{6.16}$$

where $t_j = (0.001j)$s, and j runs from 1 to 2000 (Figure 6.6a). Note that the instantaneous frequencies of the hyperbolic chirps are $230/\big(2\pi(2.3 - t_j)^2\big)$ Hz and $500/\big(2\pi(2.3 - t_j)^2\big)$ Hz.

The CWT spectrogram with a Morlet wavelet (most commonly used; similar to a sine wave in attenuation) is shown in Figure 6.6b. One can see that the CWT peaks lose power toward higher frequencies (white arrows). This might be misinterpreted as the amplitudes of the hyperbolic chirps are decreasing over time (attenuating) which is not true because the hyperbolic chirps constructed in this example have constant amplitude over time. Comparing Figures 6.5e and 6.6b, one can observe that using the Morlet wavelet mitigated the effect of the spikes in the spectrogram.

Figure 6.6c shows the ordinary LSWS of the time series given by Equation (6.16) for $L_0 = 20$ and $L_1 = 2$. Using these window size parameters, the true signal peaks are very well resolved, and their percentage variances do not change significantly across the time-frequency domain. The effect of the spikes (appears toward the locations in time where the time series has zero magnitude) is also mitigated in the spectrogram (white arrows).

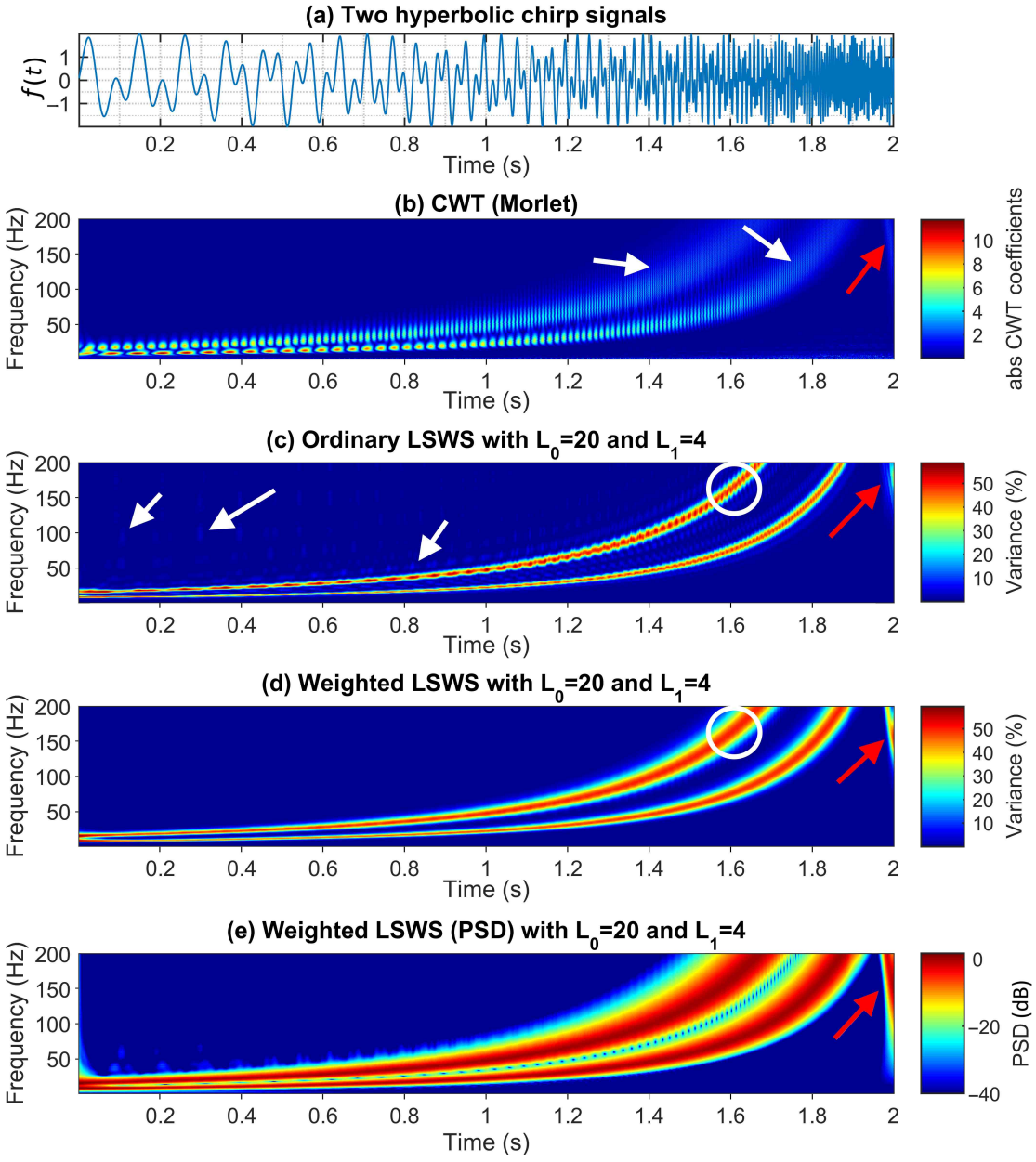

Figure 6.6: The time series given by Equation (6.16) and its analyses.

The spikes can be mitigated further by defining an appropriate weight function, such as the Gaussian function. For instance, let $\mathbf{P_y}$ be the diagonal matrix of order $R(j,k)$ given by Equation (6.4) for the window located at τ (choose $\tau = t_j$) whose diagonal entries are given by the following Gaussian function

$$G(i) = e^{-c(2\pi\omega_k(t_i-\tau))^2}, \tag{6.17}$$

where c is a window constant, and the t_i's are the times of the data points within the window of size $R(j,k)$. In astronomical applications, a popular choice for c would be 0.0125 (Foster, 1996). By this selection of $\mathbf{P_y}$, the sinusoidal basis functions will be adapted to the Morlet wavelet in the least-squares sense for analysis. The weighted LSWS is illustrated in Figure 6.6d by considering the weights defined by Equation (6.17). It can be seen that the effect of the spikes are mitigated compared to Figure 6.6c, but the bandwidth of the true spectral peaks increased (poor frequency resolution, inside the circles in Figure 6.6c and d). Defining the weights using Equation (6.17) may also change the true peak locations (Foster, 1996). Thus, it is recommended that the ordinary LSWS be used when there is no covariance matrix associated with a time series.

Figure 6.6e shows the LSWS given by Equation (6.11) in terms of Power Spectral Density (PSD) in deciBel (dB) defined by

$$\mathrm{PSD_{LSWA}} = 10\log_{10}\left(\frac{s(t_j,\omega_k)}{1-s(t_j,\omega_k)}\right). \tag{6.18}$$

The values less than -40 dB are set to -40 dB. Note that the peaks shown by red arrows in Figure 6.6b to e are the alias effect of the signal $\cos\big(500/(2.3-t_j)\big)$.

6.6 Synthetic Unequally Spaced Time Series

Consider the following inherently unequally spaced time series simulated for one month

$$\begin{aligned} f(t_j) &= 5 + 20t_j + (8-5t_j)\sin(2\pi\cdot 60t_j) + 2t_j\sin(2\pi\cdot 5t_j) \\ &\quad + \cos\left(\frac{150}{1.3-t_j}\right) + h(t_j) + wgn(t_j), \end{aligned} \tag{6.19}$$

where $1 \le j \le 720$ ($0 \le t_j \le 1$), wgn is white Gaussian noise generated by the MATLAB function wgn, and

$$h(t_j) = \begin{cases} (9-0.03j)\sin(2\pi\cdot 120t_j) & \text{if } \ 200 \le \mathrm{j} \le 300, \\ 0 & \text{otherwise.} \end{cases} \tag{6.20}$$

The linear trend, the two sine waves of variable amplitudes and constant cyclic frequencies 60 and 5 cycles per month (c/mo), the hyperbolic chirp, the short duration sine wave $\mathbf{h}$ given by Equation (6.20) of variable amplitude and constant cyclic frequency 120 c/mo, the white Gaussian noise and the time series $\mathbf{f}$ are illustrated in Figure 6.7a to g, respectively.

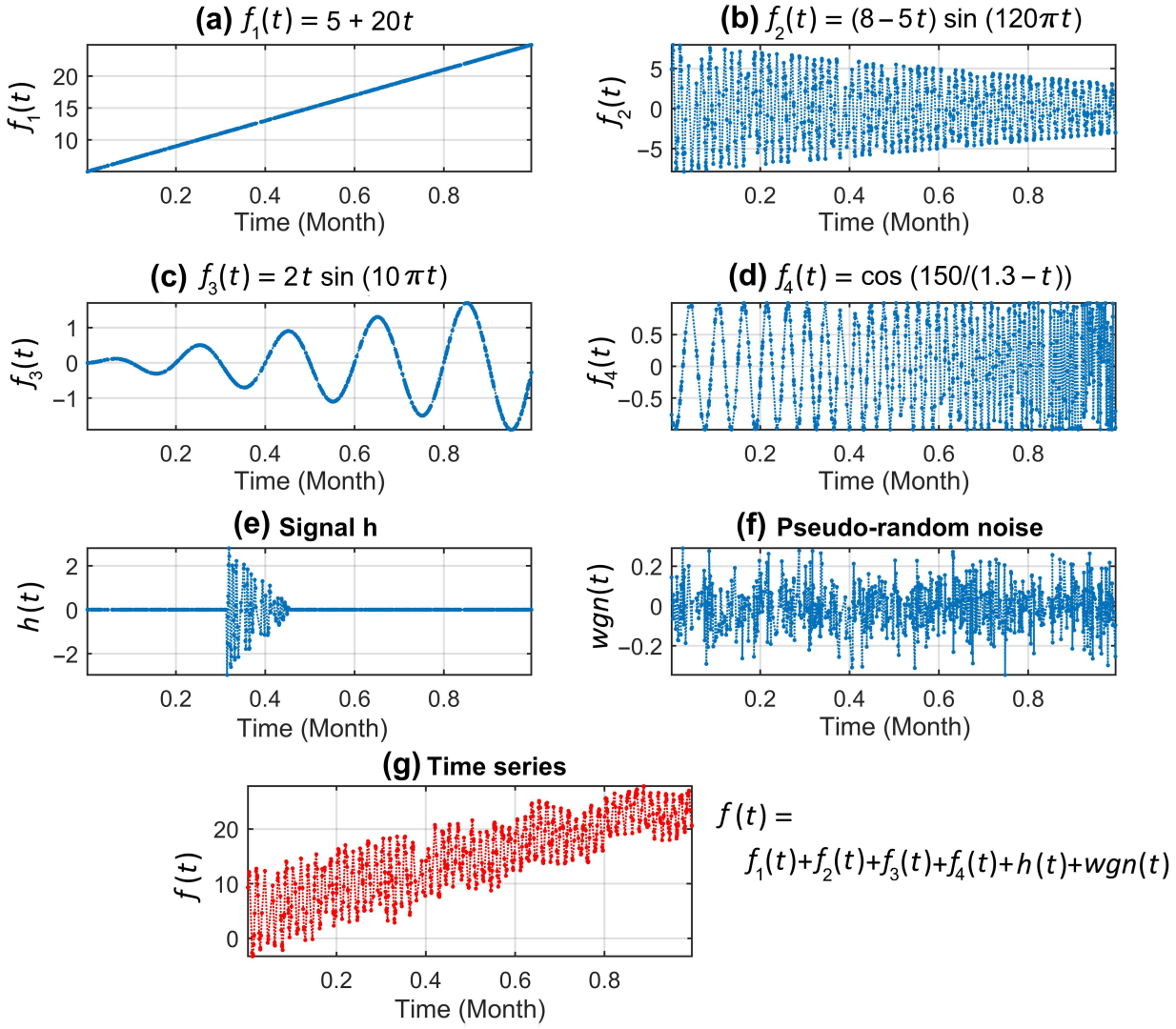

Figure 6.7: The unequally spaced time series given by Equation (6.19) and its constituents. (a)-(f) The constituents, and (g) the time series (the sum of the constituents).

In this example, it is intended to search for signals **h**, the hyperbolic chirp, and the sine wave of constant cyclic frequency 5 c/mo. Note that the white Gaussian noise, the linear trend and the sine wave of constant cyclic frequency 60 c/mo are considered noise in this example that highly contaminate the time series. A notable advantage of the LSSA, LSWA and WWZ is that one may select any set of spectral frequencies based on the purpose of analysis. The set of cyclic frequencies chosen for the analysis in this example is $\mathbf{\Omega} = \{1, 2, 3, \ldots, 200\}$ c/mo.

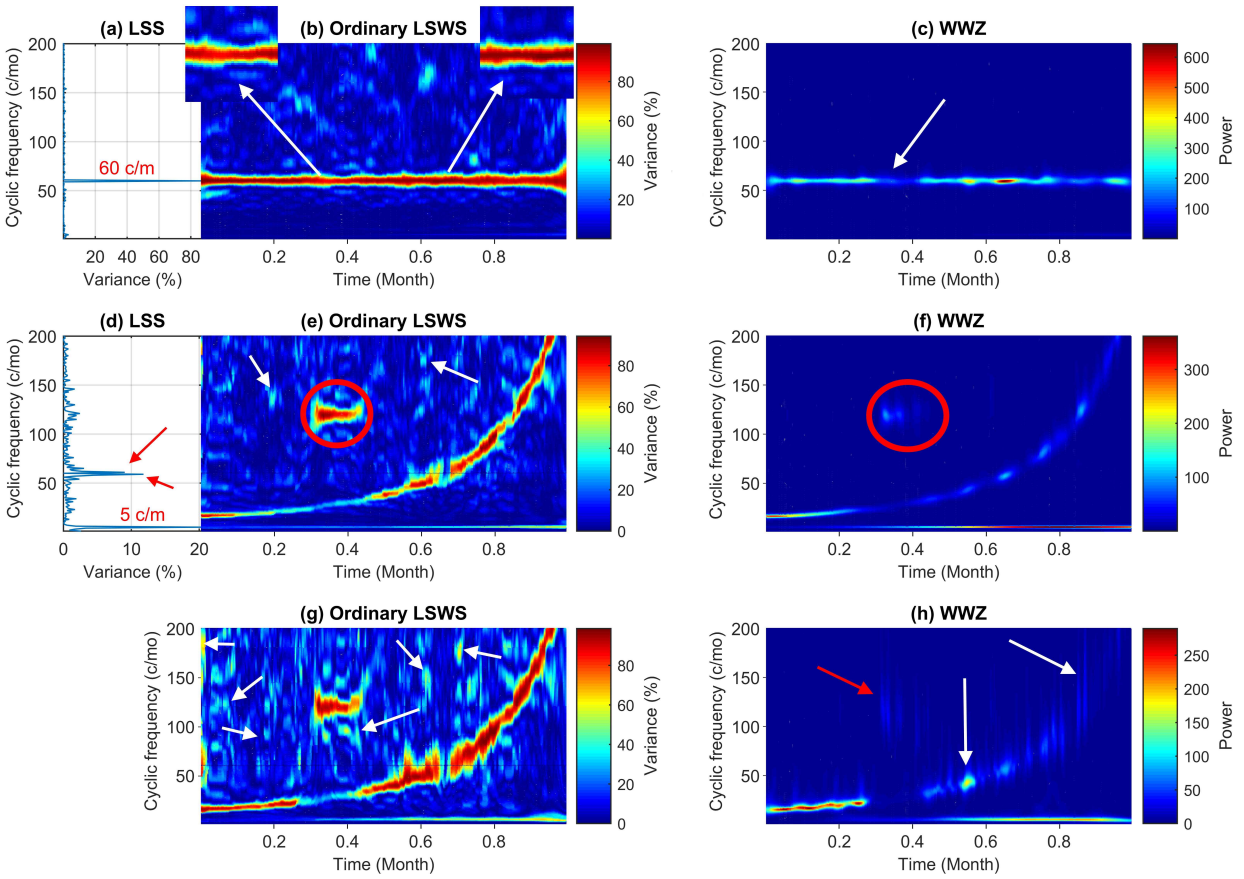

Figure 6.8: The analyses of the unequally spaced time series shown in Figure 6.7g.

Figure 6.7g shows that the time series has a trend. The LSS is illustrated in Figure 6.8a (white panel) after removing the trend. The LSS detects one strong peak at 60 c/mo, but it does not explain the nature of the constituent in $\mathbf{f}$ that creates this peak, in other words, the LSS does not show whether this peak is for a wave of variable or constant amplitude over time, or whether the duration of the wave is short or long.

Since there are 720 samples per month, $M = 720$. Figure 6.8b shows the ordinary LSWS with $L_0 = 10$ and $L_1 = 4$ after removing the trend (i.e., by setting $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{t}]$). The peaks corresponding to sine wave of constant cyclic frequency 60 c/mo can be observed in Figure 6.8b (horizontal reddish band in the spectrogram). The percentage variances of the spectral peaks in the spectrogram show how much each residual segment of $\mathbf{f}$ of size $R(j,k)$ contains a constituent of cyclic frequency ω_k. For instance, the peaks shown by the left magnified section have lower percentage variance than the peaks shown by the right magnified section indicating that there might be other constituents in the residual segments approximately from time 0.3 to 0.4 month.

Figure 6.8c shows the WWZ of the time series after removing the linear trend estimated by the least-squares method (approximately $5 + 20t_j$). Similar to the LSWS, the WWZ shows the peaks corresponding

to the sine wave of constant cyclic frequency 60 c/mo, and the effect of the spectral leakages are mitigated. Although the peaks from 0.3 to 0.4 month lose their power (arrow), it cannot ascertain that there might be certain hidden components of different frequencies in that time period (see Figure 6.4c).

In the next step, one may suppress the peak at 60 c/mo (remove the sinusoidal wave of 60 c/mo) to search for other possible signals. The LSS of the residual series after suppressing the peak is shown in Figure 6.8d. The LSS now shows several significant peaks. Note that the sine wave of 60 c/mo has variable amplitude. Since the sine wave of constant amplitude is used to fit the entire $\mathbf{f}$, the wave will not fit well to $\mathbf{f}$, and so several peaks appear around 60 c/mo in the LSS for the residual series (red arrows). If researchers are not aware of a phenomenon that causes variable amplitude waves in a time series, this can be misinterpreted. Furthermore, the hyperbolic chirp and series $\mathbf{h}$ cannot be studied from the LSS, and the LSS does not show the amplitude variation for the sine wave of 5 c/mo.

Selecting $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{t}, \cos(2\pi \cdot 60\mathbf{t}), \sin(2\pi \cdot 60\mathbf{t})]$ in the LSWA removes the trend and the constituent of cyclic frequency 60 c/mo simultaneously from the segments of $\mathbf{f}$ (Figure 6.8e). From Figure 6.8e, one can study the spectral peaks corresponding to signals $\mathbf{h}$ (inside the circle), the hyperbolic chirp with variable frequency over time and the sine wave of variable amplitude and constant cyclic frequency 5 c/mo. Note that the actual nature of these signals cannot be studied from Figure 6.8d. The spectral peaks corresponding to the sine wave of variable amplitude and constant cyclic frequency 60 c/mo are completely removed from the spectrogram shown in Figure 6.8e because sinusoids within the windows fit better to the segments of the series, but this is not achieved in the LSS (arrows in Figure 6.8d).

Figure 6.8f shows the WWZ of the sum of series shown in Figure 6.7c to f. One can clearly see that signal $\mathbf{h}$ is not completely detected (inside the circle). The reason is that the effective numbers of data points are small in certain time periods of $\mathbf{h}$ that reduced the power significantly, and also the WWZ is defined in terms of signal-to-noise ratio that reduced the power of false peaks (mitigated the spectral leakages). A stochastic surface (to be defined in the next section) can, in fact, flag the peaks shown by arrows in Figure 6.8e and g as noise at certain confidence level (usually 99%).

To see the sensitivity of the LSWS and WWZ to the window width, the window size parameters are chosen as $L_1 = 2$ and $L_0 = 20$ with the same $\underline{\mathbf{\Phi}}$ as in Figure 6.8e, and the ordinary LSWS is shown in Figure 6.8g. In addition, the WWZ is shown in Figure 6.8h by selecting $c = 0.1$. The frequency resolution in both spectrograms are now poor, but the WWZ shows the significant variability of power compare to Figure 6.8f, and signal peaks are not well localized (arrows).

In Figure 6.8e to h, since the window size decreases when frequency increases, the spectral peaks corresponding to signal $\mathbf{h}$ with cyclic frequency 120 c/mo have higher time resolution but poorer frequency resolution, and the spectral peaks corresponding to signal of cyclic frequency 5 c/mo have higher frequency resolution but poorer time resolution.

6.7 Analysis of a Very Long Baseline Interferometry Length Time Series

Very Long Baseline Interferometry (VLBI) is a geometric technique that collects signals from an astronomical object, such as a quasar[35] using many antennas (radio telescopes) mounted on the Earth. To understand this technique better, two VLBI antennas (one in Westford in the USA and one in Wettzell in Germany) with geometry are illustrated in Figure 6.9. Since a quasar is so far from the Earth, the spherical wavefronts of the radio signals (shown by yellow curves) emitted from the quasar can be considered as flat planes as they reach the Earth. Since the radio wavefronts from the quasar are planar, rays drawn perpendicular to the arriving wavefronts are essentially parallel. Thus, the antennas, oriented to receive signals from the quasar, are pointed in the same direction, indicated by the angle θ in Figure 6.9. The distance between the two antennas is called baseline. The radio signal from the quasar is shown as wiggly lines. When the signal arrives at the Westford antenna, the same signal is a distance L away from the Wettzell antenna. The incoming signals are tagged with an atomic clock. The distance L can then be calculated as the time

[35] A quasar is a compact region at the centre of a galaxy with high amount of emitting energy per unit time.

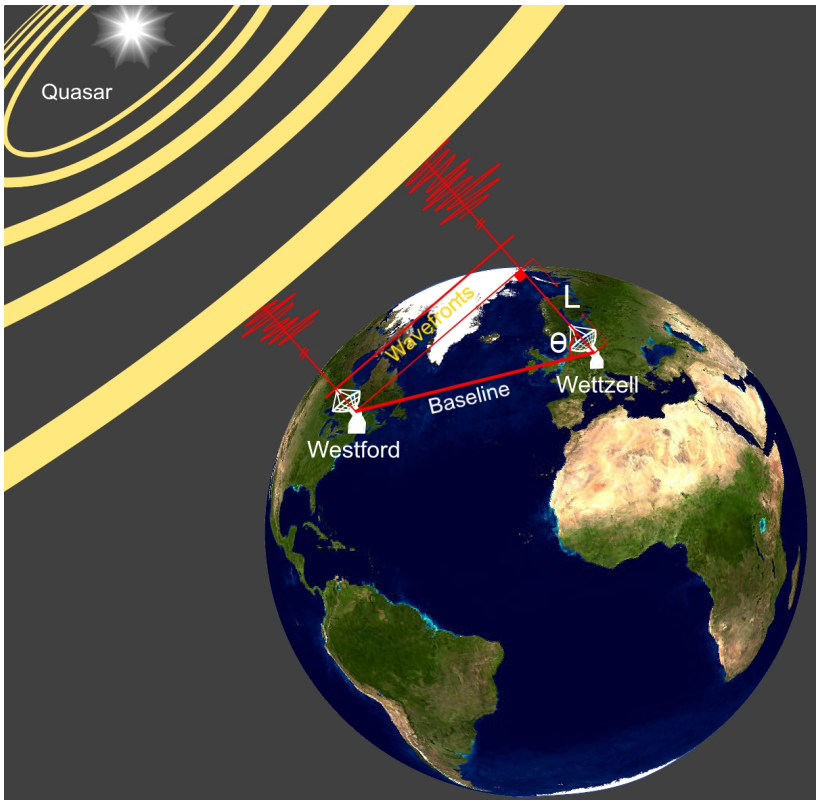

Figure 6.9: The VLBI technique for two antennas in Westford and Wettzell.

Figure 6.10: The main VLBI stations in the world. This image is downloaded from www.sinet.ad.jp.

difference between identical radio signals arriving at the two antennas multiplied by the speed of light[36]. Therefore, the baseline length may be calculated as $L/\cos(\theta)$. Repeating this process over time generates a time series that is often unequally spaced due to several reasons, but primarily due to experiment scheduling among participating countries.

This technique is useful to define the celestial reference frame, to monitor the motion of the Earth's tectonic plates, to improve the atmospheric models, to verify the length-day, and to measure the fundamental speed of gravity (Fomalont and Kopeikin, 2003). Measuring time differences between arrivals from many quasars observed with a global network of antennas can identify the inertial reference frame defined by the quasars and can estimate the relative positions of the antennas (Vaníček and Krakiwsky, 1986; Campbell, 2004). Since the antennas are fixed to the Earth, their locations track the instantaneous orientation of the Earth in the inertial reference frame. In practice, it is often required to express positions on a flat surface (map) in form of grid coordinates, considering the distortion in length, shape, and size of the area of interest (Ghaderpour, 2016). Relative displacements of the locations of the antennas indicate tectonic plate motion, antenna deformation, and local uplift or subsidence. Figure 6.10 shows the main VLBI stations in the world used for many purposes, such as the ones mentioned above.

[36]The wavefront is moving at the speed of light that is 3×10^8 metres per second.

In this section, a VLBI length time series analysis is demonstrated as an application of the LSWA. The VLBI length series is between stations Westford and Wettzell, and the length changes over longer time due primarily to tectonic plate movement and other reasons, such as atmospheric temperature (due to deformation of the antennas), tides, etc. (Titov, 2002; Campbell, 2004). The time series was obtained from www.ccivs.bkg.bund.de. The series comprises 1733 unequally spaced and unequally weighted baseline length estimates via the least-squares method, from January 9th 1984 at 19:12:00 Universal Time (UT) to September 3rd 2014 at 16:48:00 UT (Figure 6.11a).

Since there are 1733 samples over 30.6 years (unit time is a year), there are approximately 57 samples per annum ($M = 57$). The window size parameters chosen for the analyses are $L_1 = 4$ cycles and $L_0 = 30$ samples. The set of cyclic frequencies selected for this analysis is $\mathbf{\Omega} = \{0.1, 0.2, 0.3, \ldots, 12\}$ c/a. Note that the linear trend expresses the lengthening of the baseline due to the relative tectonic plate movement on which the two VLBI antennas are mounted.

Figure 6.11b shows the ordinary LSWS of the VLBI series and its stochastic surface at 99% confidence level (gray surface) after removing the trend (by setting $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{t}]$). The two-dimensional representation of Figure 6.11b is shown in Figure 6.11d. The constant annual cyclic frequency (1 c/a or $\approx$ 365 days, white arrows) and the semiannual cyclic frequency (2 c/a or $\approx$ 182.5 days) are significant in certain years (red arrows), where $\approx$ indicates approximation. The ordinary LSS also shows a significant peak of variance 4.5% at period of $\approx$ 365 days (Figure 6.11c). The linear trend estimated by the least-squares method as $5998325.356 + 0.017t$ is removed from the time series, and the result of the WWZ for the residual series is shown in Figure 6.11e. Similar to the ordinary LSWS, the WWZ shows the annual and semiannual peaks (arrows in Figure 6.11d and e).

The weighted LSS and LSWS after removing the trend are shown in Figure 6.11f and g. Many peaks that are not significant in Figure 6.11d, are now significant at 99% confidence level (e.g., arrows). In the next step, the annual peaks (having the highest percentage variance) are suppressed, and the spectrum and the spectrogram of the residual series are shown in Figure 6.11h and i, respectively. It can be seen that peaks shown by arrows from year 1993 to year 1997 are not significant anymore (the spectral leakages). The peaks at cyclic frequency 1.6 c/a ($\approx$ 228 days, red arrow) are suppressed, and the spectrum and spectrogram of the new residual series are shown in Figure 6.11j and k, respectively. One can observe that the spectral peaks shown by white arrows in Figure 6.11i are no longer significant in Figure 6.11k. However, the semiannual peaks shown by red arrows and the peaks inside the circles remained significant after simultaneously removing the linear trend and sinusoids of cyclic frequencies 1 and 1.6 c/a.

In order to investigate what phenomena could possibly be causing the significant peaks in the spectrogram, the atmospheric temperature records at both stations are also analyzed. They are obtained from http://ggosatm.hg.tuwien.ac.at/DELAY/SITE/VLBI/ almost at the same time period of the VLBI series

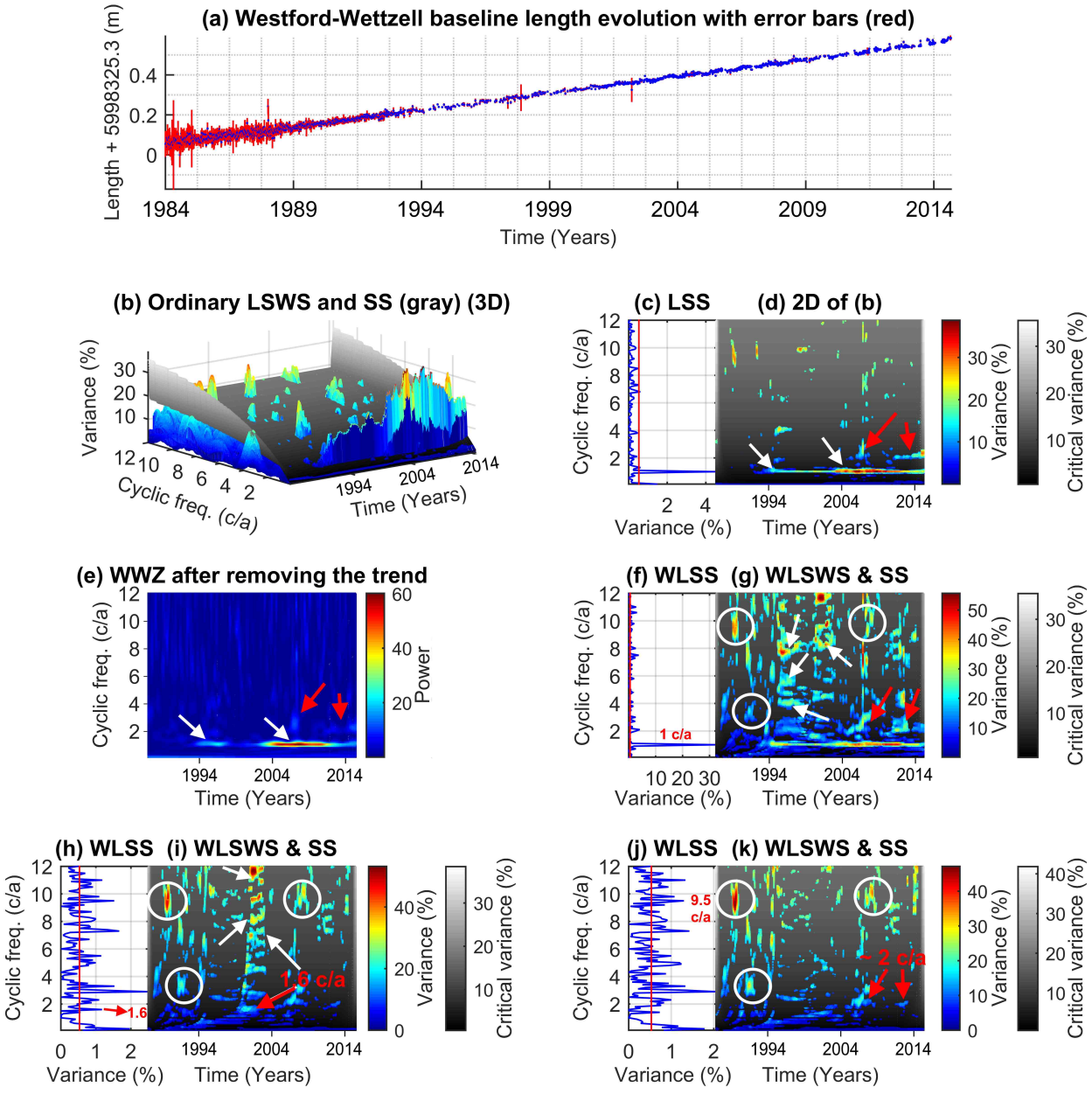

Figure 6.11: The VLBI time series and its analyses using the LSS, LSWS and WWZ. Note that WLSS, WLSWS, SS, 3D and 2D are the short forms for the weighted LSS, weighted LSWS, stochastic surface, three-dimensional and two-dimensional representations, respectively.

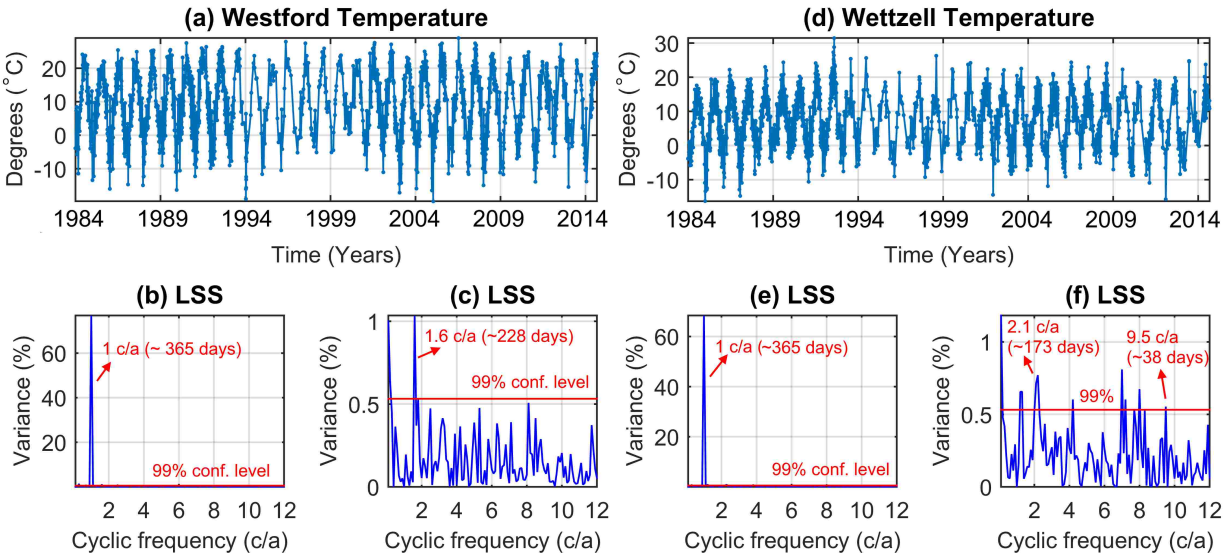

Figure 6.12: The unequally spaced temperature records at the VLBI sites (a) Westford (d) Wettzell and their LSSA (b) and (c) for Westford and (e) and (f) for Wettzell.

(Figure 6.12a and d). The least-squares spectra of the Westford and Wettzell temperature series after removing the trend are shown in Figure 6.12b and e, respectively. The annual peaks in the spectra show that the annual peaks in Figure 6.11d, e and g are mainly due to the temperature variation causing deformation of the antennas.

The annual peak is suppressed for each time series, and the least-squares spectra of the residual temperature series of Westford and Wettzell are shown in Figure 6.12c and f, respectively. The peak at 1.6 c/a in Figure 6.12c shows that the Westford antenna is being deformed with a period of $\approx$ 228 days (unusual atmospheric temperature variations), and the evidence of its effect was observed in both spectrograms in Figure 6.11g and i from year 1997 to year 2001. The LSS of the Westford temperature series from year 1997 to year 2001 (not shown here) has a significant peak at 1.6 c/a ($\approx$ 228 days). The semiannual peaks in the spectrograms that are significant since 1997 could also be physically linked to the semiannual temperature variation at the Wettzell station as observed in Figure 6.12f. Similarly, the peaks shown inside the circles in Figure 6.11k at cyclic frequency of 9.5 c/a ($\approx$ 38 days) could possibly be caused by unusual temperature variation at the Wettzell station.

6.8 Summary

The LSWA is a natural extension of the LSSA that decomposes a time series from the time domain into the time-frequency domain using an appropriate segmentation of the time series. This segmentation is linked to a translating window (rectangular or Gaussian) whose size decreases as frequency increases within a frequency band of interest. I mathematically showed how this segmentation results in detection of constituents with amplitude and/or frequency variability over time as well as detection of short-duration signatures buried in the spectrogram by simultaneously removing the constituents of known forms from the original time series. The LSWA can statistically analyze any non-stationary and unequally spaced time series without any need for editing of the original time series, that is indeed a breakthrough in signal processing and time series analysis. The flowchart of the LSWA algorithm is illustrated in Figure 6.13. Note that the locations of the translating windows may also be chosen to be equally spaced, and the translating windows may also be set up in such a way that they do not overlap each other for the sake of faster computation.

I applied the LSWA to analyze synthetic and real time series. The spectrogram in the LSWA shows the percentage variance of the spectral peaks corresponding to the segments of the time series, and unlike the WWZ, it is not a poor measure of amplitude. In the synthetic examples, I showed that the percentage variance of the constituents with frequency variability (but constant amplitude) over time remain constant relatively in the spectrogram, using an appropriate selection of the window size parameters. Selection of the Gaussian window (the weight matrix whose diagonal elements are the Gaussian values) generates smoother spectral peaks with wider bandwidth in the spectrogram. Simultaneous estimating and removing the constituents of known forms in the spectrogram allows one to search for any hidden signatures in the residual time series. The spectral peaks in the spectrogram can be statistically estimated at certain confidence level. I have also investigated and discovered several short-duration signatures in the VLBI baseline length series, many of which are generated by deformation of the antennas due to temperature variations.

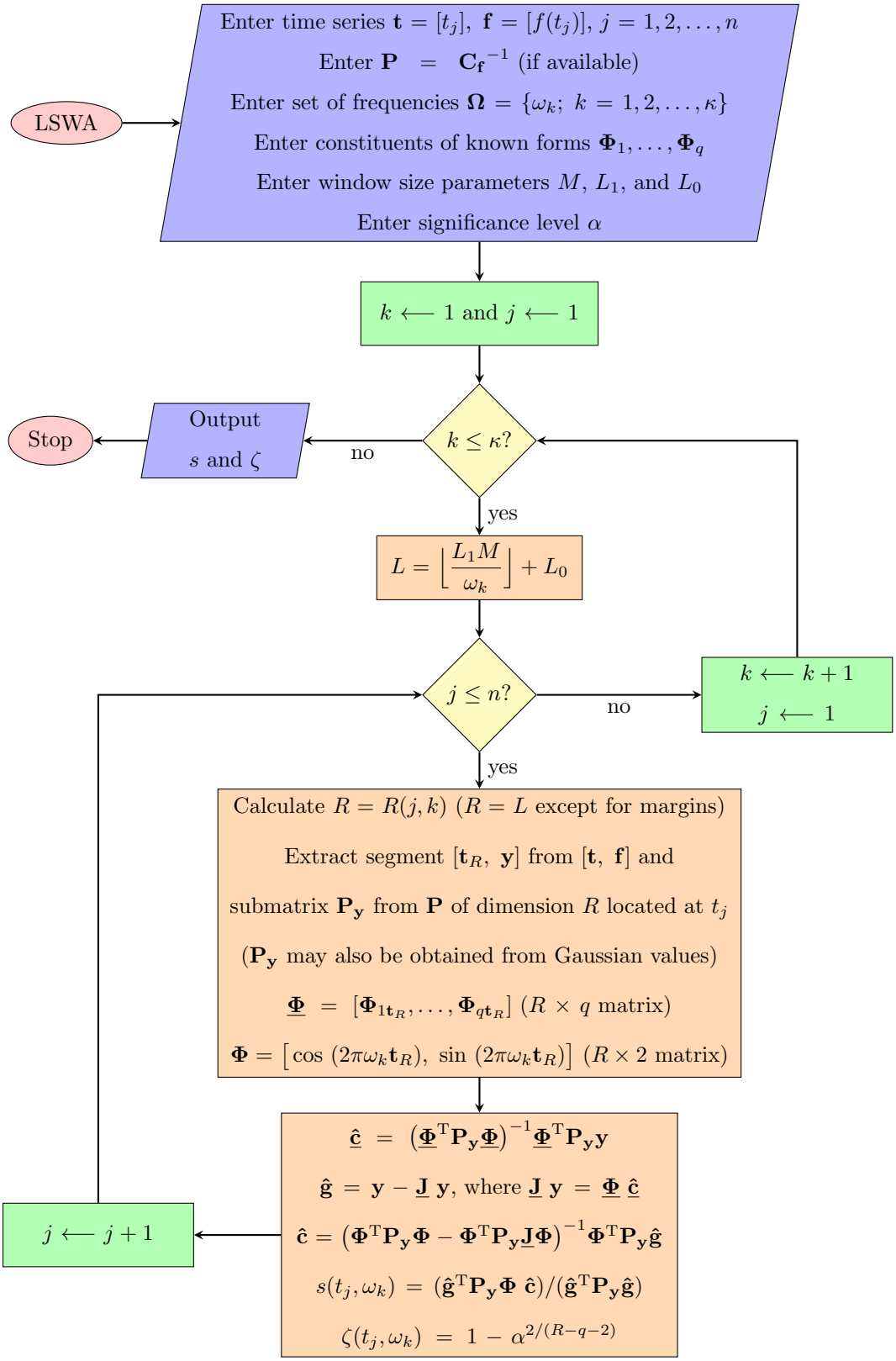

Figure 6.13: Flowchart of the LSWA algorithm.

7 Least-Squares Cross-Wavelet Analysis and Its Applications

In this chapter, cross-spectrogram of two time series (may be extended to more) with its stochastic surface at a certain confidence level (usually 95% or 99%) is introduced as well as the phase differences of common constituents of the two time series. To obtain the cross-spectrogram, the spectrogram of each time series first needs to be calculated using the LSWA algorithm illustrated in Figure 6.13. I will show the outstanding performance of my proposed method on synthetic and real time series. This chapter is an unabridged version of a paper published in Journal of Geodesy (Ghaderpour et al., 2018a).

7.1 Cross-Spectrogram and Its Stochastic Surface

It is shown in Appendix C that the LSWS of time series $\mathbf{f}$ given by Equation (6.11) follows ($\sim$) the beta distribution, i.e., $s \sim \beta_{1,\Re/2}$, where β stands for the beta distribution, and $\Re = R(j,k) - q - 2$. Suppose that time series $\mathbf{f_1}$ and $\mathbf{f_2}$ have been derived from two statistically independent populations of random variables following respectively the multidimensional normal distributions $\mathcal{N}(\mathbf{0}, \mathbf{C_{f_1}})$ and $\mathcal{N}(\mathbf{0}, \mathbf{C_{f_2}})$, where the covariance matrices $\mathbf{C_{f_1}}$ and $\mathbf{C_{f_2}}$ may be singular. Let $\mathbf{\Omega} = \{\omega_k;\ k = 1, \ldots, \kappa\}$ be a set of common spectral frequencies for the two time series under consideration.

Let s_1 and s_2 be the spectrograms corresponding to $\mathbf{f_1}$ and $\mathbf{f_2}$, respectively. By the assumption, s_1 and s_2 are statistically independent. For each pair (t_j, ω_k), the Least-Squares Cross-Wavelet Spectrogram (LSCWS) of the two time series are defined as the product of their spectrograms:

$$X_s = s_1 s_2. \tag{7.1}$$

Note that $0 < X_s < 1$ because $0 < s_1, s_2 < 1$. Since the sampling rates and/or the times of the two time series may not be the same, one may set the time vector in X_s as the union of the sets of times of the first and second time series. If t_j is a time in the first time series but not in the second (or vice-versa), then $s_1(t_j, \omega_k)$ is calculated within a window of size $R_1(j,k)$ located at t_j, and $s_2(t_j, \omega_k)$ is calculated within a window of size $R_2(j,k)$ located at t_j, emphasizing that the centre of the window does not have to be located at a time in which there exists a sample. For a pair (t_j, ω_k), if the value of X_s is closer to one, then the two segments of the two time series within the windows of sizes $R_1(j,k)$ and $R_2(j,k)$ are highly coherent, and they are highly incoherent if the value of X_s is closer to zero.

Suppose that q_1 and q_2 are the number of constituents of known forms being considered in the calculation of the spectrograms for $\mathbf{f_1}$ and $\mathbf{f_2}$, respectively. Ghaderpour and Pagiatakis (2017) showed that $s_1 \sim \beta_{1,\Re_1/2}$ and $s_2 \sim \beta_{1,\Re_2/2}$, where $\Re_1 = R_1(j,k) - q_1 - 2$ and $\Re_2 = R_2(j,k) - q_2 - 2$. More precisely, from the definition of the beta distribution, the Probability Distribution Functions (PDF) of the independent random variables s_1 and s_2 are respectively as follows

$$f_{s_1}(u) = \frac{\Re_1}{2}(1-u)^{\Re_1/2-1}, \qquad 0 < u < 1, \tag{7.2}$$

$$f_{s_2}(v) = \frac{\Re_2}{2}(1-v)^{\Re_2/2-1}, \qquad 0 < v < 1. \tag{7.3}$$

Following similar methodologies in Glen et al. (2004), the PDF of random variable X_s can be obtained. The following transformation is a one-to-one mapping from $A = \big\{(u,v);\ \ 0 < u < 1,\ 0 < v < 1\big\}$ to $B = \big\{(x,y);\ \ 0 < x < 1,\ 0 < y < 1\big\}$

$$X_s = s_1 s_2, \qquad Y = s_2. \tag{7.4}$$

Let T and K denote the transformation and the inverse transformation, respectively. Therefore,

$$\begin{aligned} x = T_1(u,v) = uv, \qquad & y = T_2(u,v) = v, \\ u = K_1(x,y) = x/y, \qquad & v = K_2(x,y) = y. \end{aligned} \tag{7.5}$$

The Jacobian of the transformation is

$$\begin{aligned} J = \frac{\partial(u,v)}{\partial(x,y)} &= \Big(\frac{\partial u}{\partial x}\Big)\Big(\frac{\partial v}{\partial y}\Big) - \Big(\frac{\partial u}{\partial y}\Big)\Big(\frac{\partial v}{\partial x}\Big) \\ &= \Big(\frac{1}{y}\Big)(1) - \Big(\frac{-x}{y^2}\Big)(0) = \frac{1}{y}. \end{aligned} \tag{7.6}$$

Using Equation (7.5) and the independency of s_1 and s_2,

$$f_{X_s,Y}(x,y) = |J| f_{s_1,s_2}(u,v) = \frac{1}{y} f_{s_1}(x/y) f_{s_2}(y). \tag{7.7}$$

Integration of Equation (7.7) with respect to y over the appropriate interval and using Equations (7.2) and (7.3) yield the PDF of X_s

$$\begin{aligned} f_{X_s}(x) &= \int_x^1 f_{X_s,Y}(x,y)\ \mathrm{d}y = \int_x^1 \frac{1}{y} f_{s_1}\Big(\frac{x}{y}\Big) f_{s_2}(y)\ \mathrm{d}y \\ &= \frac{\Re_1 \Re_2}{4} \int_x^1 \frac{1}{y}\Big(1 - \frac{x}{y}\Big)^{\Re_1/2-1} (1-y)^{\Re_2/2-1}\ \mathrm{d}y, \end{aligned} \tag{7.8}$$

where $0 < x < 1$. One may numerically calculate this integral using the well-known methods, such as Simpson quadrature and Gauss-Kronrod quadrature (Hazewinkel, 2001; Moler, 2008; Shampine, 2008). Figure 7.1 shows the PDF of X_s for selected values of $\Re_1$ and $\Re_2$.

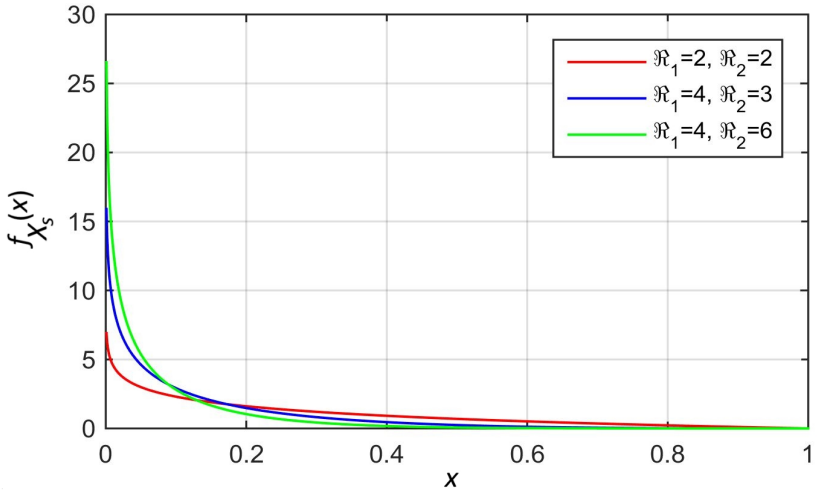

Figure 7.1: Illustration of the PDF of X_s given by Equation (7.8) for several values of $\Re_1$ and $\Re_2$.

Table 7.1 shows the critical values (cv) corresponding to two significance levels $\alpha = 0.01$ and $\alpha = 0.05$ for selected values of $\Re_1$ and $\Re_2$. It can be seen that the critical values approach zero rapidly when $\Re_1$ and $\Re_2$ increase. Note that the area under the density curve from its lower right tail to the left is computed to find the critical value for a certain significance level because the PDF becomes singular when approaching zero. Now let $\mathbf{f_1}$ and $\mathbf{f_2}$ be two time series of dimensions n_1 and n_2, respectively. For each $\omega_k \in \mathbf{\Omega}$, assume that s_1 and s_2 are the least-squares spectra corresponding to $\mathbf{f_1}$ and $\mathbf{f_2}$, respectively. Analogous to the cross-spectrogram, the Least-Squares Cross-Spectrum (LSCS) is defined as $X_s = s_1 s_2$, where $\Re_1 = n_1 - q_1 - 2$ and $\Re_2 = n_2 - q_2 - 2$. The PDF of X_s is given by Equation (7.8). Note that the LSCS is independent of the sampling rates because it decomposes the time series into the frequency domain and not into the time-frequency domain.

Table 7.1: The critical values (cv) corresponding to the significance levels (α) and $\Re_i$.

$\Re_1$	$\Re_2$	$\alpha = 0.01$	$\alpha = 0.05$
2	2	$cv = 0.862$	$cv = 0.701$
4	3	$cv = 0.624$	$cv = 0.436$
4	6	$cv = 0.452$	$cv = 0.290$
10	10	$cv = 0.186$	$cv = 0.105$
20	30	$cv = 0.043$	$cv = 0.022$
100	101	$cv = 0.003$	$cv = 0.002$

7.2 Phase Differences in the Cross-Spectrogram

When analyzing two time series together, researchers are also interested in the phase differences between any two constituents of interest of the same time-frequency spectral value. To find the phase differences, note that

$$\mathbf{y} = a \sin\left(2\pi\omega\mathbf{t} + \theta\right) = \left[\cos\left(2\pi\omega\mathbf{t}\right),\ \sin\left(2\pi\omega\mathbf{t}\right)\right]\mathbf{c} = \mathbf{\Phi}\mathbf{c}, \tag{7.9}$$

where $\mathbf{c} = a\ \left[\sin\theta,\ \cos\theta\right]^{\mathrm{T}}$ and $\mathbf{\Phi}$ is the design matrix given by Equation (6.1). Following the same procedure in the LSWA, one can see that Equation (6.10) is the estimate for $\mathbf{c}$. For each pair (t_j, ω_k), let $\hat{c}_1$ and $\hat{c}_2$ be the two elements of $\hat{\mathbf{c}}$ given by Equation (6.10). Therefore, $\hat{a}\sin\hat{\theta} = \hat{c}_1$ and $\hat{a}\cos\hat{\theta} = \hat{c}_2$, and so $\hat{a} = (\hat{c}_1^2 + \hat{c}_2^2)^{0.5}$. Thus

$$\hat{\theta} = 2\arctan\left(\frac{1 - \hat{c}_2/\hat{a}}{\hat{c}_1/\hat{a}}\right). \tag{7.10}$$

Now for pair (t_j, ω_k), if $\hat{\theta}_1$ and $\hat{\theta}_2$ are the phase differences of the constituents of $\mathbf{f_1}$ and $\mathbf{f_2}$ calculated by Equation (7.10), respectively, then define the phase difference corresponding to pair (t_j, ω_k) as $\psi = \hat{\theta}_2 - \hat{\theta}_1$, where $-2\pi < \psi < 2\pi$. The phase difference for a common constituent of a particular frequency indicates how much the constituent of the second time series lags (if $-2\pi < \psi < -\pi$ or $0 \le \psi < \pi$) or leads (if $-\pi \le \psi < 0$ or $\pi \le \psi < 2\pi$) the constituent of the first time series. Let σ_1^2 and σ_2^2 be the variances associated with $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively, estimated from the least-squares estimation process. Applying the covariance law to $\psi = \hat{\theta}_2 - \hat{\theta}_1$, one may calculate the standard deviation of ψ as $\sigma_\psi = \left(\sigma_1^2 + \sigma_2^2\right)^{0.5}$.

7.3 Analysis of Two Synthetic Unequally Spaced Time Series

The outstanding performance of the LSCSA and LSCWA on two synthetic and inherently unequally spaced time series are demonstrated below. The time series are given by

$$f_1(t_i) = \cos\left(\frac{30}{1.2 - t_i}\right) + g(t_i) + w_1, \quad 0 \le t_i \le 1,\ 1 \le i \le 600, \tag{7.11}$$

where $g(t_i) = \cos\left(25 \cdot 2\pi t_i\right)$ for $200 \le i \le 300$ and $g(t_i) = 0$, otherwise, and w_1 is Gaussian white random noise, and

$$f_2(t'_j) = \cos\left(5 \cdot 2\pi t'_j\right) + \cos\left(25 \cdot 2\pi t'_j + \frac{\pi}{2}\right) + w_2, \quad 0 \le t'_j \le 1,\ 1 \le j \le 400, \tag{7.12}$$

where w_2 is Gaussian white random noise. Note that in the LSSA, LSWA, LSCSA, and LSCWA, the time series do not have to contain white noise. In this example, a white noise is considered to see the performance of these analyses because measurements in practice typically contain white noise. Time series $\mathbf{f_1} = [f_1(t_i)]$ and $\mathbf{f_2} = [f_2(t'_j)]$ are illustrated in Figure 7.2a with blue and red colours, respectively. Assume that the unit

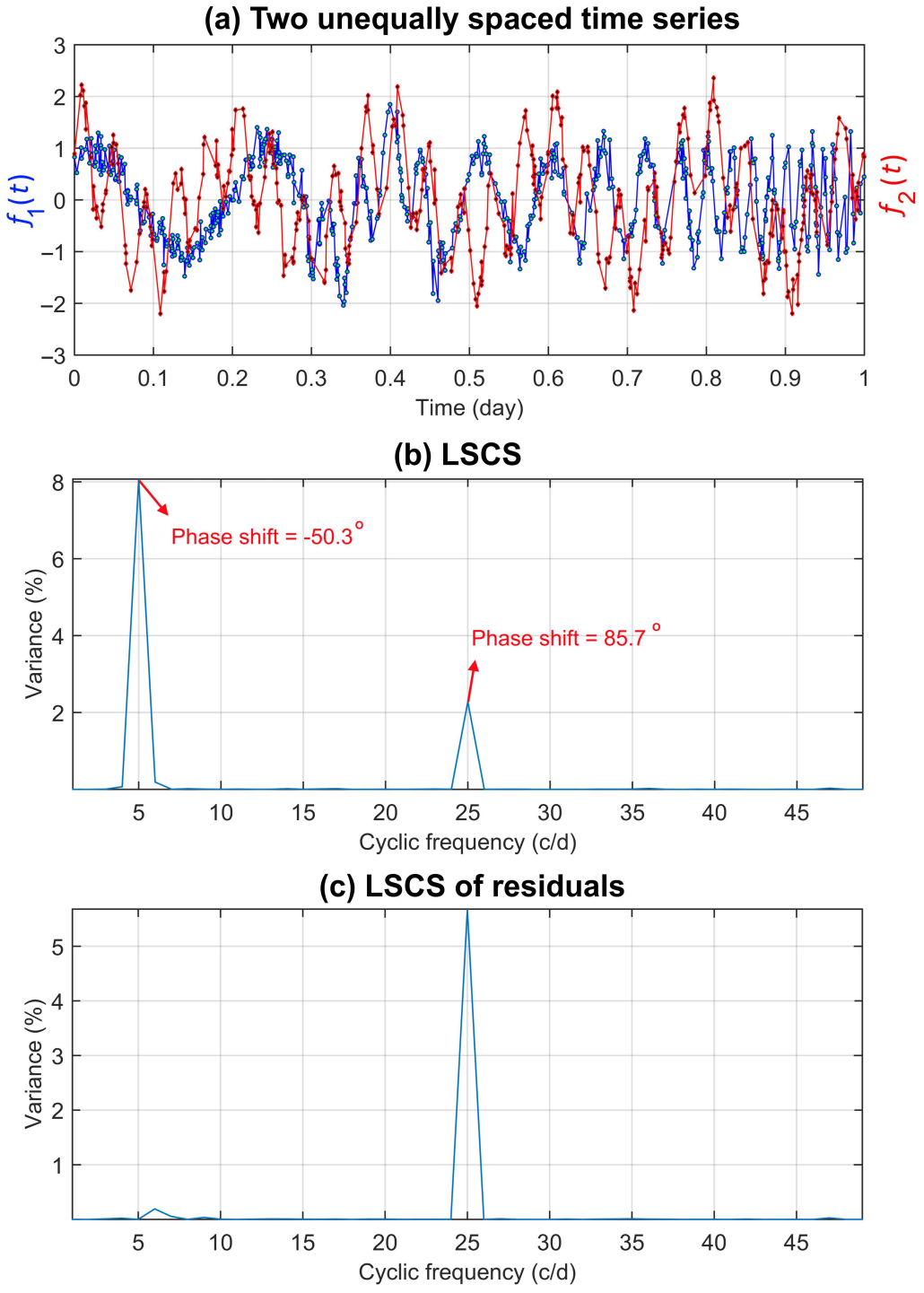

Figure 7.2: (a) Two unequally spaced series given by Equations (7.11) and (7.12) shown in blue and red, respectively, (b) the LSCS of the two time series, and (c) the LSCS of the two residual time series after suppressing the peak at 5 c/d.

time for $\mathbf{f_1}$ and $\mathbf{f_2}$ is in days. The set of common frequencies chosen in the analyses of the two time series is $\mathbf{\Omega} = \{1, 2, \ldots, 49\}$ cycles per day (c/d). The LSCS of the two time series detects two peaks at cyclic frequencies 5 and 25 c/d (Figure 7.2b). Figure 7.2b also shows the phase differences of the constituents of cyclic frequencies 5 and 25 c/d in $\mathbf{f_1}$ with respect to the ones in $\mathbf{f_2}$ which are approximately close to $-50°$ and $+90°$, respectively, as expected from the synthetic time series. In the next step, the peak at 5 c/d is suppressed and the LSCS of the residual time series is shown in Figure 7.2c. It can be seen that the percentage variance of the peak at 25 c/d increases, indicating more coherency in the residual time series.

Since there are 600 and 400 samples per day for $\mathbf{f_1}$ and $\mathbf{f_2}$, respectively, $M_1 = 600$ and $M_2 = 400$. Note that since the time series are inherently unequally spaced, no aliasing is expected to occur at high frequencies, however, if at least one of the time series was equally spaced, then it is recommended to choose half of the sampling rate as the Nyquist frequency for the cross-spectrogram. The cross-spectrograms of the two time series and their residuals along with many selected phase differences (arrows) are shown in Figure 7.3 ($L_1 = 2$ and $L_0 = 10$ are selected for both time series). The angles of white arrows on the cross-spectrogram with respect to the horizontal axis show the phase differences. Note that an arrow is in the first or second quadrant of the trigonometric circle when $0° \leq \psi < 180°$, and it is in the third or fourth quadrant when $-180° \leq \psi < 0°$.

The coherency of the two original time series can be observed in more details in Figure 7.3a. The cross-spectrogram shows that the hyperbolic chirp signal in Equation (7.11) is coherent with the cosine wave of 5 c/d in Equation (7.12) from 0 to 0.5 day interval. The percentage variance of the corresponding peaks decreases as time increases because the frequency of the hyperbolic chirp in $\mathbf{f_1}$ increases rapidly over time, showing less coherency with the cosine wave of 5 c/d in $\mathbf{f_2}$ as time increases. It can be seen from Figure 7.3a that signal $g(t_i)$ in $\mathbf{f_1}$ and the cosine wave of 25 c/d in $\mathbf{f_2}$ are coherent in the time period of approximately $0.3 - 0.5$ day. Part of the hyperbolic chirp at 25 c/d in $\mathbf{f_1}$ and the cosine wave of 25 c/d in $\mathbf{f_2}$ are also coherent in the time period of approximately $0.7 - 0.8$ day. Figure 7.3b is obtained by suppressing (removing) the peaks at 5 c/d in the LSCWS that can be done by adding the sine and cosine base functions of 5 c/d as two columns to $\underline{\mathbf{\Phi}}$ in Equation (6.5) for $\mathbf{f_2}$. The coherency between signal $g(t_i)$ in $\mathbf{f_1}$ and the cosine wave of 25 c/d in the residual $\mathbf{f_2}$ is significantly increased (compare the percentage variance of the peaks in Figure 7.3a and b from time 0.3 to 0.5 day at 25 c/d).

The phase difference of almost 90° can be observed in the spectrogram from time 0.35 to 0.5 day at 25 c/d as expected. Note that the phase differences in the LSCS may not be accurate for non-stationary time series as the sinusoids are being fitted to the entire time series. For instance, the phase difference in the LSCWS corresponding to time 0.77 d and cyclic frequency 25 c/d is 43.2°, but it cannot be detected as such in the LSCS. Moreover, the coherency between signal $g(t_i)$ in $\mathbf{f_1}$ and cosine wave of cyclic frequency 25 c/d in $\mathbf{f_2}$ is more localized and significant in Figure 7.3b.

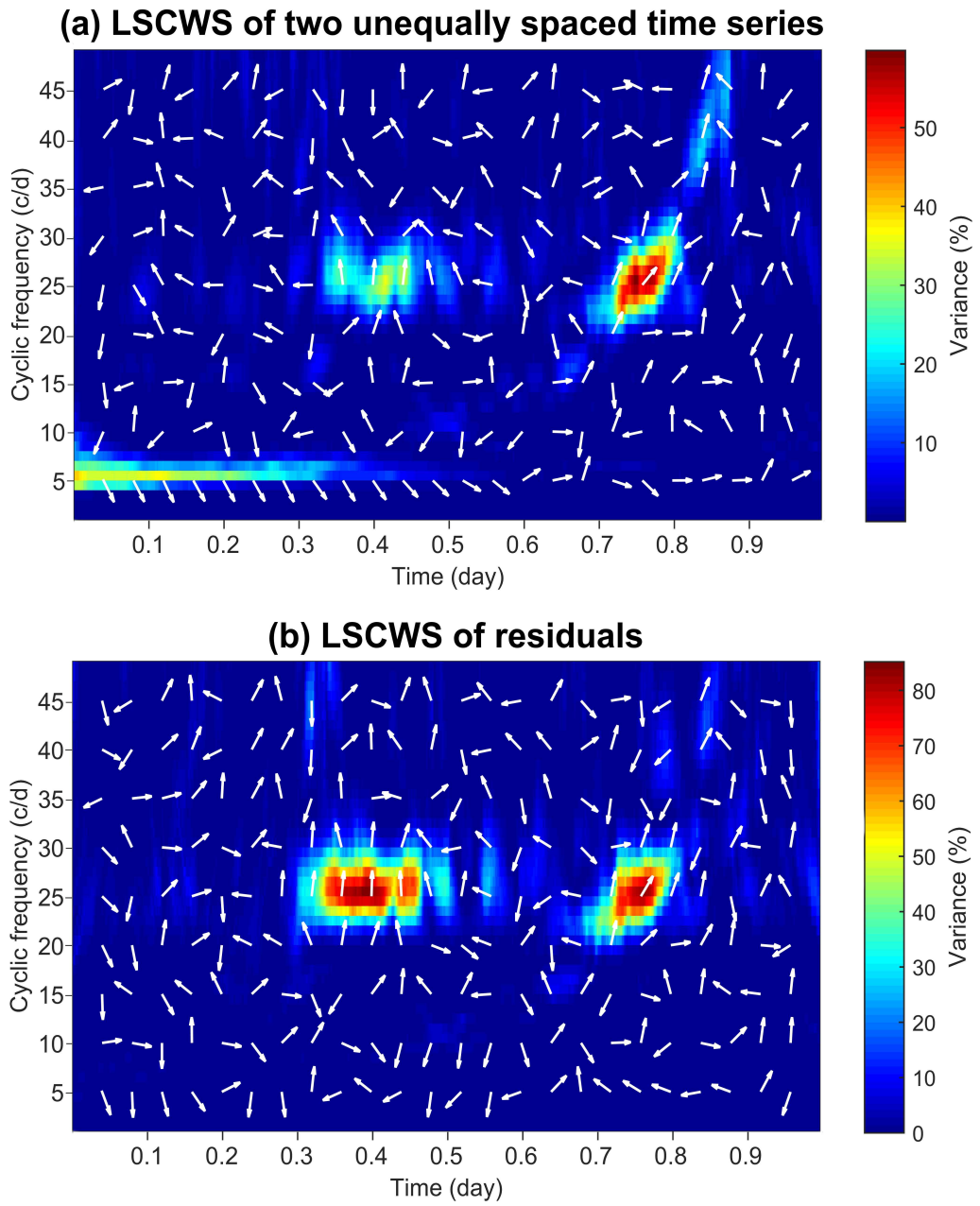

Figure 7.3: (a) The LSCWS of the time series shown in Figure 7.2a, and (b) the LSCWS of the residual time series. The arrows in the spectrogram show the phase differences.

As an application of the LSCWA, the Gravity field and steady-state Ocean Circulation Explorer (GOCE) electrostatic gradiometer measurement disturbances are analyzed to determine their coherence with the Poynting energy flux along two successive ascending satellite tracks during a magnetic storm (Ince and Pagiatakis, 2016), and part of the results are compared to the XWT. The possible effect of the temperature on the VLBI baseline length series was investigated in Chapter 6. As another application of the LSCWA, the coherency between the VLBI baseline length series from Westford to Wettzell with the temperature series (recorded since 1984) is investigated to show the effect of the temperature variation on the baseline length more rigorously.

7.4 GOCE Electrostatic Gravity Gradiometer Measurements and Poynting Energy Flux

Two different geophysical time series are analyzed to demonstrate the performance of LSCWA when is applied to real, unequally spaced and unequally weighted time series. The first time series is composed of gravitational gradient disturbances observed by GOCE satellite, whereas the second time series comprises Poynting electromagnetic energy flux (plasma flow) in the ionosphere derived from equivalent horizontal ionospheric currents and vertical currents. The hypothesis of this analysis is that the electromagnetic energy flux in the Earth's thermosphere introduces undesirable disturbances in GOCE Electrostatic Gravity Gradiometer (EGG) measurements as shown in Ince and Pagiatakis (2016); Ince (2017); Ince and Pagiatakis (2017). The data sets are described and their analyses are presented in this section.

7.4.1 GOCE Electrostatic Gravity Gradiometer

The latest gravity field mission GOCE was one of the three satellite missions, after CHAMP (CHAllenging Mini-satellite Payload) and GRACE (Gravity Recovery And Climate Experiment), which has helped remarkably to improve the accuracy and resolution of the Earth's global geopotential models from space. In order to map the Earth's gravitational field from satellite measurements, the influence of all temporal gravitational and non-gravitational accelerations acting on the satellite should be measured and removed from the observations.

Preliminary analyses and literature search have shown that the (GOCE) EGG measurements (Rummel et al., 2011) are affected by unknown non-gravitational sources around the magnetic poles (Peterseim et al., 2011; Stummer et al., 2012; Siemes et al., 2012; Yi et al., 2013; Ince, 2017; Ince and Pagiatakis, 2016). Ince (2017) demonstrated that the source of these disturbances are related to solar activity dependent ionospheric dynamics. Moreover, these disturbances can reach up to a magnitude of about 3 to 5 times larger than the expected noise level (approximately 11 milli-Eotvos, mE) of the gravity field components at specific epochs.

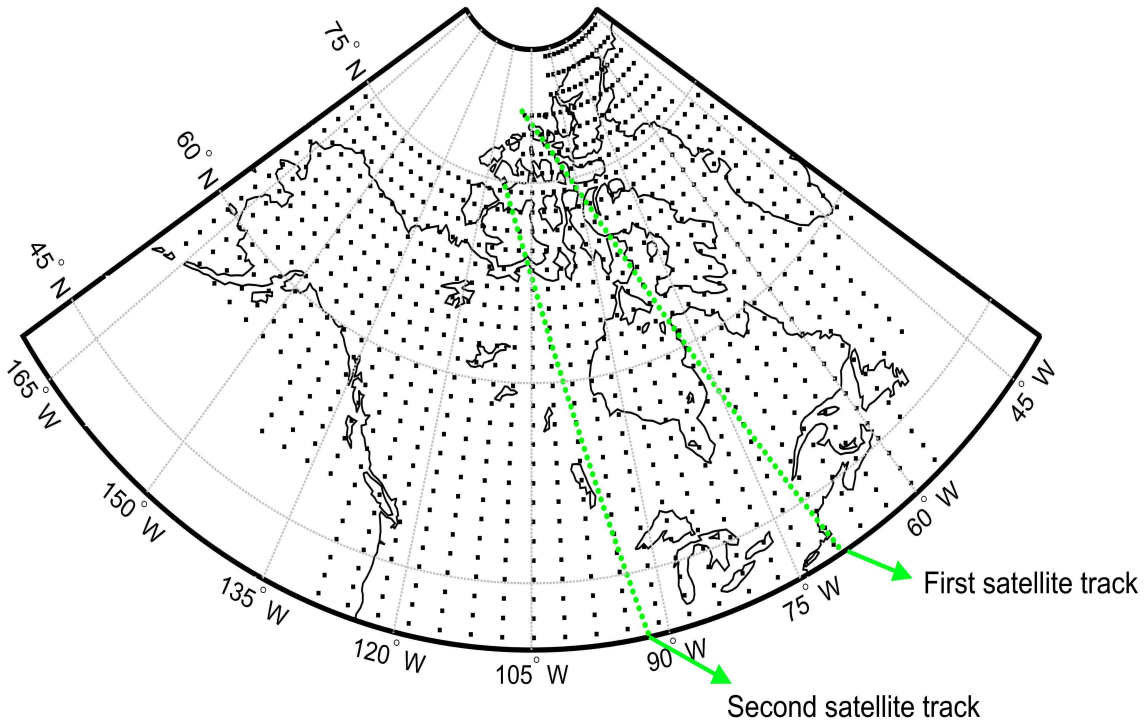

Figure 7.4: Two satellite ascending tracks (green curves) over Canada. The grid points represent the SEC grid.

Ince and Pagiatakis (2016) found that these specific epochs correspond to geomagnetic storms due to the interaction between the interplanetary magnetic field and geomagnetic field.

In order to understand the relation between the undesirable disturbances observed in gravitational gradients and plasma flow variations due to solar activity, Spherical Elementary Currents (SEC) are used for the coherency analysis (Amm, 1997; Amm and Viljanen, 1999; Weygand et al., 2011). Based on the Laplace condition, the summation of the diagonal components of the Gravitational Gradient Tensor (GGT) (trace) should be zero. It is worth mentioning that these variations affect the EGG measurements around the auroral oval when the ionospheric dynamics (plasma flow) are more intense. In this example, the trace on which these spurious signals are more visible is investigated, and the selected GOCE tracks are examined individually in which the trace is subject to rapid changes (Figure 7.4).

The sampling period of GOCE measurements is one sample per second, and along track correlated errors are not considered in this study (error in the GOCE gradients outside of the measurement bandwidth is colored). The data sets analyzed here do not have any gaps, and they are equally spaced. However, it is possible that GOCE data in level zero or level one may have gaps.

7.4.2 Equivalent Ionospheric Currents and Poynting Vector

The SEC system methodology given in Amm (1997); Amm and Viljanen (1999) are used to derive Equivalent Ionospheric Currents (EIC) which are based on the magnetic field disturbances measured at the Earth's surface (Weygand et al., 2011). The SEC system methodology furthermore upward continues the anomalous magnetic field measurements from the ground to an altitude of about 110 km. Equivalent currents (horizontal currents) and spherical elementary current amplitudes (vertical currents) are provided in 10 s temporal resolution in a gridded format over North America and West Greenland (see the grid points in Figure 7.4). The kriging method (using a variogram) is applied to interpolate the grid values into the satellite ground track position (see green lines in Figure 7.4) for the analyses (Ince, 2017).

By using SEC values provided on a grid (see dots in Figure 7.4) interpolated into the satellite position, the Poynting vector (Kelley, 2009; Ince, 2017) is computed in the north-south and east-west directions along the satellite track. The units of Poynting vector are Watt per metre square (W/m^2). In general, the direction of the Poynting vector indicates the direction of the flow of the electromagnetic energy that drives the ionospheric currents.

The Gravitational Gradient Tensor (GGT) trace and the Poynting vector of the first satellite track are shown in Figure 7.5a. The unit time for our analyses is in hours and the sampling rate for original GGT trace series is $M_1 = 3600$ samples per hour. The sampling rate for the cross-track Poynting vector component series is $M_2 = 360$ samples per hour. Note that the Nyquist frequency of the cross-track Poynting vector component series is 180 cycles per hour (c/h) which is different from the one for the original GGT trace series of 1800 c/h. Therefore, the maximum frequency for the analysis must not be greater than 180 c/h. In this analysis, $\mathbf{\Omega} = \{1, 2, \ldots, 100\}$. The LSCS after removing the datum shift of the original GGT trace series is shown in Figure 7.5b. The original GGT trace and the cross-track Poynting vector component series of the second satellite track and their LSCS (after removing the datum shift) are also shown in Figure 7.6.

Note that the XWT (Grinsted et al., 2004) would require interfiling and editing of both time series to be identical in their times and equally spaced, and it does not consider the covariance matrices associated with the time series. However, the LSCWA does not require the series to have the same sampling rates and to be equally spaced at identical times. The comparison between the XWT and LSCWA for equally spaced time series of the same sampling rates are shown below. This comparison is made only for the first satellite track.

In order to apply the XWT, the GGT trace series of the first satellite track should be decimated from 3600 to 360 samples per hour ($M_1 = 360$) because the cross-track Poynting vector component series has the sampling rate of 360 samples per hour. To keep the useful information of the GGT trace series as much as possible, a Gaussian filter is used in this process (see Appendix E). The result of the decimated GGT trace series is illustrated in black in Figure 7.5a, and it can be observed that the decimated data points match the

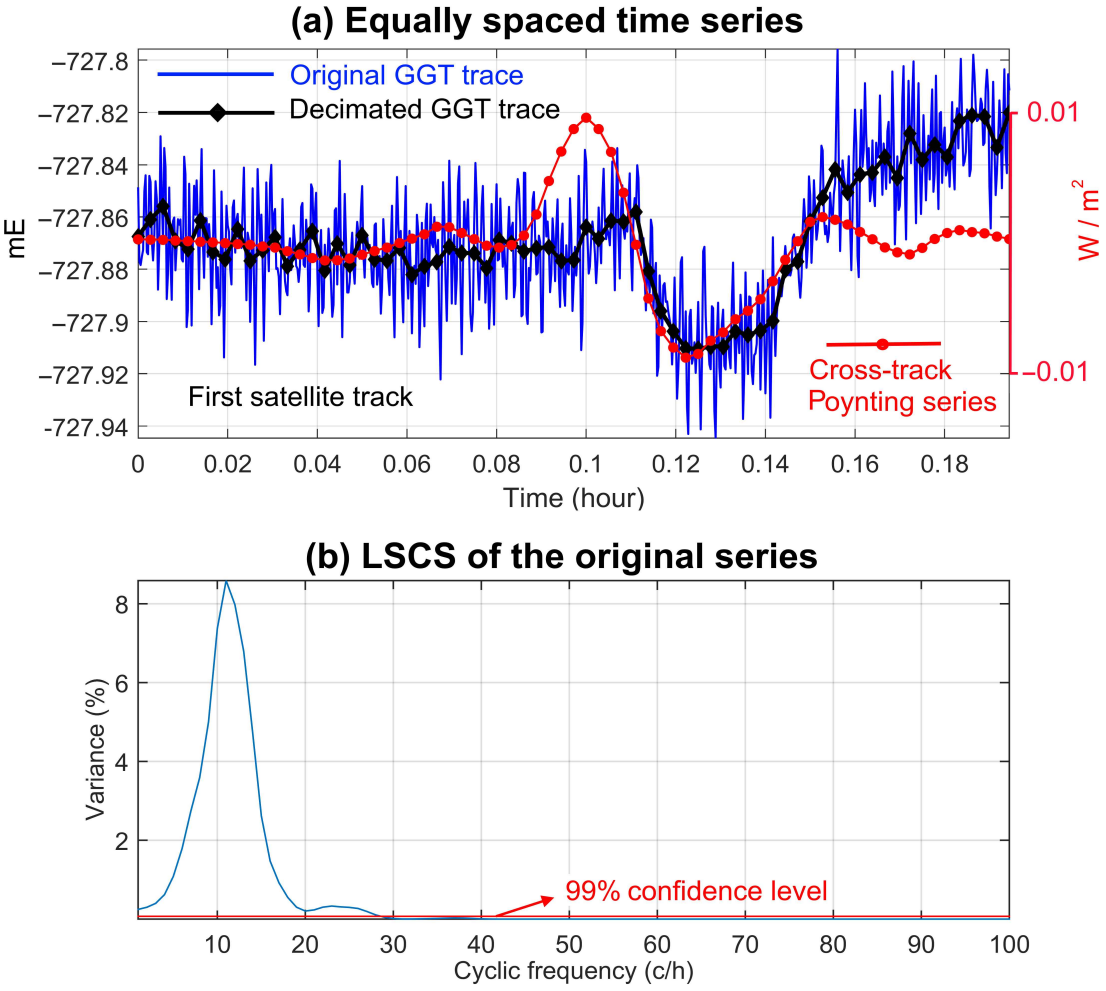

Figure 7.5: (a) The cross-track Poynting vector component series in W/m^2, original and decimated GGT trace series in mE for the first satellite track, and (b) the LSCS of original GGT trace and cross-track Poynting vector component series with 99% confidence level.

data points of cross-track Poynting vector component series in time.

Figure 7.7 shows the XWT of the decimated GGT trace series and the cross-track Poynting vector component series of the first satellite track (the Morlet wavelet is used). The MATLAB code to generate this figure was obtained from www.mathworks.com/matlabcentral/fileexchange/ 47985-cross-wavelet-and-wavelet-coherence. Note that the XWT is in time-scale, and the scales are converted to frequencies in this figure. The confidence level $Z_2(p)$ associated with a probability p in the XWT may be calculated by inverting the following integral

$$p = \int_0^{Z_2(p)} z\, K_0(z)\, \mathrm{d}z, \tag{7.13}$$

where $K_0(z)$ is the modified Bessel function of order zero (Torrence and Compo, 1998; Ge, 2008). From

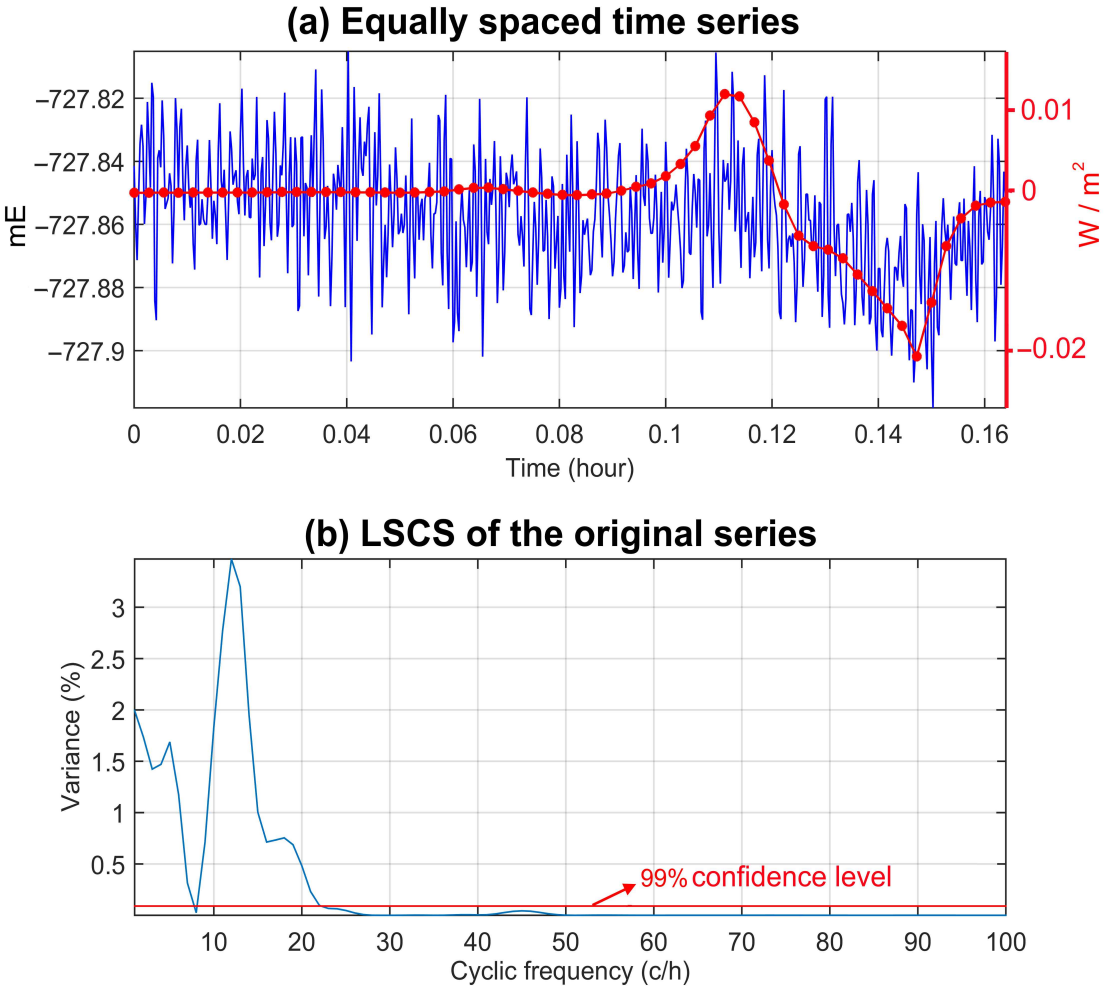

Figure 7.6: (a) The original GGT trace series shown by blue and the cross-track Poynting vector component series shown in red for the second satellite track, and (b) the LSCS of the original GGT trace and cross-track Poynting vector component series with 99% confidence level.

Equation (7.13), one can obtain $Z_2(0.95) = 3.999$ and $Z_2(0.99) = 5.767$. $Z_2(0.99)$ is used in Figure 7.7.

The LSCWA for the first satellite track is applied to three different cases, namely, the decimated GGT without (as in the XWT above) and with considering the covariance matrix and the original GGT trace series. Based on the constituents of low frequency and short duration empirically observed in the time series, the window size parameters are chosen as $L_1 = 1$ and $L_0 = 5$ for both time series and $\underline{\mathbf{\Phi}} = [\mathbf{1}]$ for the decimated GGT trace series to remove its datum shifts.

Figure 7.8a shows the LSCWS of the decimated GGT trace series by considering the covariance matrix generated by the Gaussian filtering process described in Appendix E. The significant peaks in the LSCWS are those that are above the stochastic surface shown by gray at 99% confidence level. The two-dimensional representation of the LSCWS and its stochastic surface (gray) and phase differences are shown in Figure

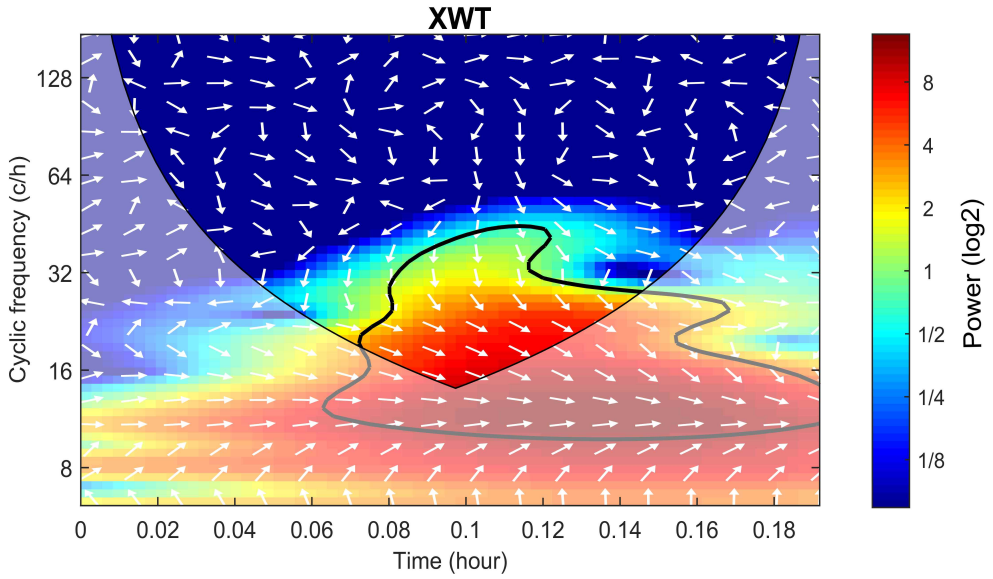

Figure 7.7: The XWT of the decimated GGT trace and cross-track Poynting vector component series of the first satellite track along with phase shifts (arrows). The area within the thick contour is the significant area at 99% confidence level. The cone of influence where the edge effects might distort the results is shown as a light shade.

7.8b. It can be seen that the two time series are highly coherent with approximately zero phase differences (see the reddish area around 0.12 hour in Figure 7.8b). This fact can also be observed in Figure 7.7. The very low coherence in the interval $0.02 - 0.1$ hours is also expected due to the lower power random character of the signals (Figure 7.5a). However, the significant peaks are much more localized in the LSCWS than in the XWT because the LSCWS is obtained from the spectrograms, decomposing the time series directly into the time-frequency domain using the least-squares method. Although the Morlet wavelet used in the XWT smooths the signal peaks in the spectrogram, it increases the bandwidth of the signal frequencies in the spectrogram (poor time-frequency resolution).

To verify the importance of the covariance matrix in the analysis, an equally weighted analysis without considering the covariance matrix is also performed (Figure 7.8c), and many of the cross-spectrogram values (denoted by $X_s(t_j, \omega_k)$) are shown in Table 7.2 and compared to the values when the covariance matrix is considered (denoted by $X_s^e(t_j, \omega_k)$). In Table 7.2, $cv(t_j, \omega_k)$ denotes the critical values (%) corresponding to pair (t_j, ω_k) at 99% confidence level. Comparing the values in the third and fourth columns, one can observe the changes in percentage variances for the LSCWS. The first five and last nine rows in this table correspond to many of the values inside the left and right circles, respectively, shown in Figure 7.8b and c. Many of the peaks that are significant in Figure 7.8b are insignificant in Figure 7.8c (inside the right circle). On the other hand, many of the peaks that are insignificant in Figure 7.8b are significant in Figure 7.8c (inside the left circle). For the original GGT trace series, the same window size parameters are selected but $M_1 = 3600$

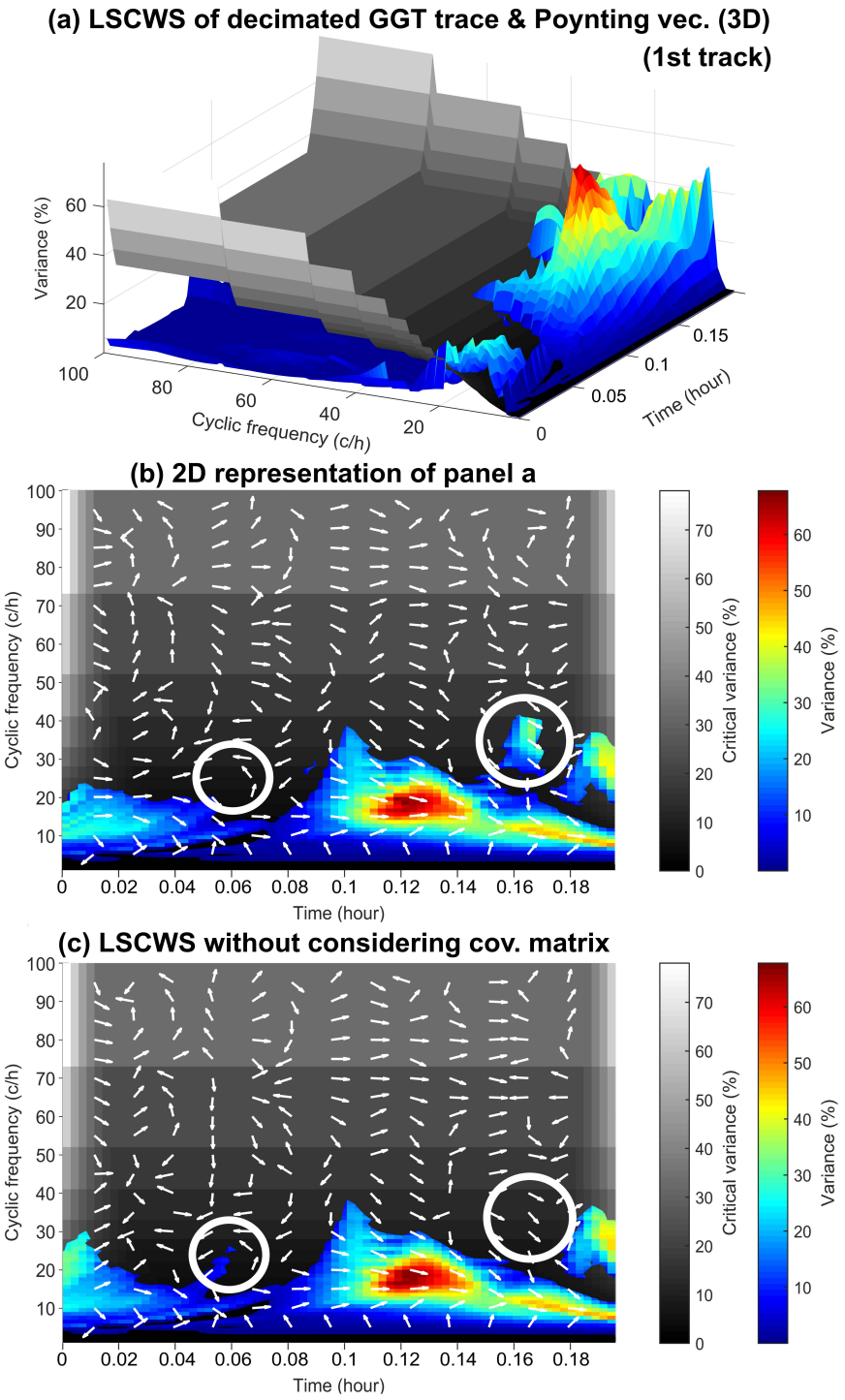

Figure 7.8: (a) Three-dimensional representation of the LSCWS of the cross-track Poynting vector component and decimated GGT trace series shown in Figure 7.5a considering the covariance matrix with the stochastic surface at 99% confidence level (gray), (b) its two-dimensional representation with phase differences (arrows), and (c) the LSCWS without considering the covariance matrix with phase differences (arrows).

Table 7.2: Comparison between many of the values in the LSCWS with (third column) and without (fourth column) considering the covariance matrix associated with the decimated GGT trace series and the critical value (fifth column) at 99% confidence level.

t_j	ω_k	$X_s^e(t_j,\omega_k)$	$X_s(t_j,\omega_k)$	$cv(t_j,\omega_k)$
0.058	19	3.82	9.16	5.9
0.058	20	3.84	9.01	5.9
0.058	22	4.37	9.45	7.02
0.061	25	8.03	10.13	8.5
0.061	26	7.45	8.98	8.5
0.153	23	11.73	6.45	7.02
0.158	21	5.94	4.95	5.9
0.158	28	16.36	9.02	10.5
0.164	36	29.4	7.71	13.28
0.164	37	29.32	8.47	13.28
0.167	36	33.99	7.52	13.28
0.167	37	34.22	8.65	13.28
0.169	36	20.72	6.61	13.28
0.169	37	22.49	7.97	13.28

(Figure 7.9). Since the window size for the original GGT trace series is larger than for the decimated one, the critical values of the LSCWS of the original GGT trace series are smaller than the ones for the decimated GGT trace series (Figures 7.8b, 7.9a). This means that the spectral peaks in the LSCWS for the decimated series are expected to be stronger than the ones for the original series in order to be statistically significant. Comparing Figure 7.8b and c with Figure 7.9a, one can observe that Figure 7.9a is closer to Figure 7.8b in terms of significant peaks (e.g., see inside the circles). This indicates the crucial importance of considering the covariance matrix associated with the series when computing the spectrogram (the weighted LSCWA).

The analysis of the original GGT trace series (Figure 7.9a) shows that the effects of the marginal windows are improved and more reliable results for the coherency and the phase differences are obtained. Note that various filtering and decimating techniques may alter the result of analysis, and so analyzing the original series (raw data) will give more reliable results. Therefore, for the second satellite track, only the LSCWS of the original GGT trace and cross-track Poynting vector component series (shown in Figure 7.6a) is illustrated in Figure 7.9b.

There is a small negative phase difference around cyclic frequency 20 c/h and time $0.12-0.14$ h in Figure 7.9a and b indicating that the ionospheric variations are leading the disturbances in GGT trace series. The analyses show that the variations observed in the two series have a period of about 5 minutes. This corresponds to spatial variations of around 2000 km. The ionospheric variations of this spatial resolution were investigated and presented in Ince and Pagiatakis (2016).

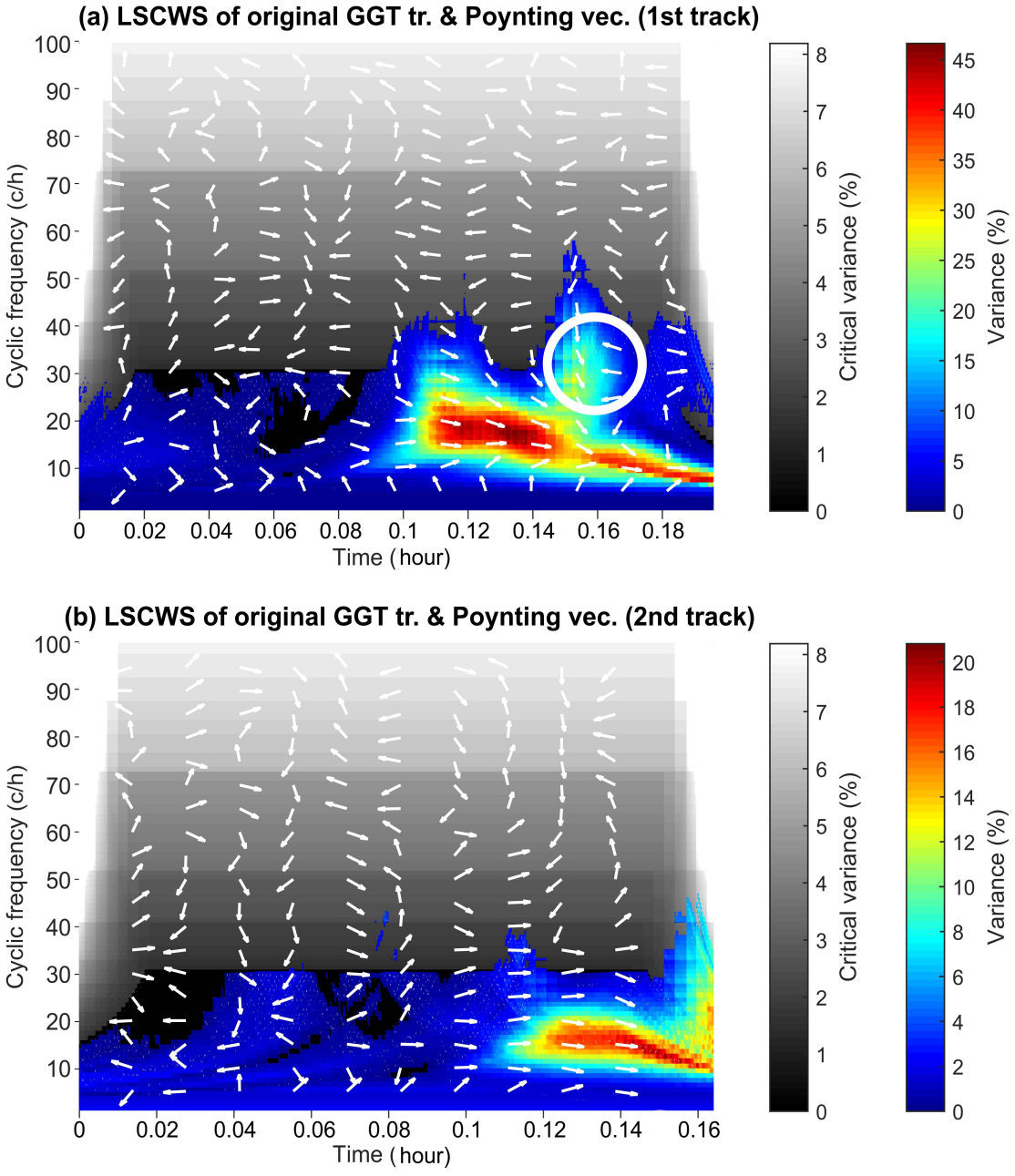

Figure 7.9: The LSCWS of the original GGT trace series and the cross-track Poynting vector component series with stochastic surface at 99% confidence level (gray) along with the phase differences for (a) the first satellite track, and (b) the second satellite track.

7.5 Westford-Wettzell VLBI Baseline Length and Atmospheric Temperatures

In this section, the effect of atmospheric temperature on the VLBI baseline length estimates between Westford in the USA and Wettzell in Germany is investigated, and the coherency between the temperature variations at the two stations and the VLBI baseline are shown. The VLBI length series and the associated variances are obtained from www.ccivs.bkg.bund.de, and the temperature time series for Westford and Wettzell are obtained from http://ggosatm.hg.tuwien.ac.at/DELAY/SITE/VLBI/.

The VLBI series (cf., Figure 7.10a) comprises 1733 unequally spaced and unequally weighted baseline length estimates from January 9th, 1984 at 19:12:00 Universal Time (UT) to September 3rd, 2014 at 16:48:00

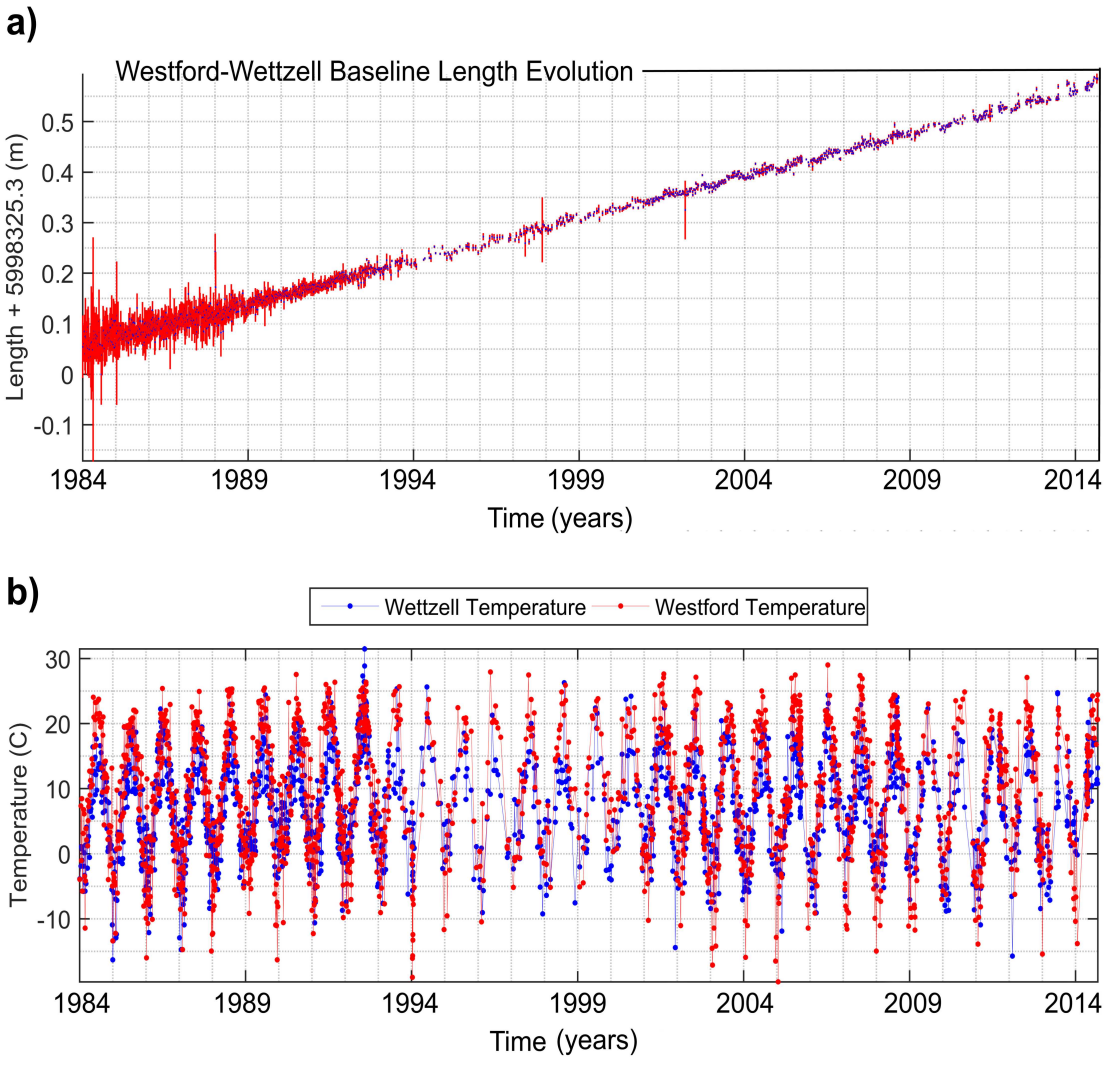

Figure 7.10: (a) The unequally spaced VLBI baseline length evolution since January 1984 with error bars (red), and (b) unequally spaced and equally weighted temperature series at stations Westford (red) and Wettzell (blue) since January 1984.

UT. In order to see the effect of the atmospheric temperature on the Westford-Wettzell baseline length, the temperatures are chosen close to, and in many cases identical to the times when the baseline length estimates are available (Figure 7.10b). Since there are 57 samples per year (on average), $M = 57$. In order to compare the results with the results in Chapter 6, the window size parameters are chosen as $L_1 = 4$ and $L_0 = 30$, and the set of cyclic frequencies is $\mathbf{\Omega} = \{0.1, 0.2, 0.3, \ldots, 12\}$ with unit of cycles per annum (c/a). Selecting smaller window size parameters will produce cross-spectrograms that are more localized in time and more accurate in phase shifts; however, frequency resolution will be poorer.

The linear trend observed in Figure 7.10a expresses the lengthening of the baseline due to the relative tectonic plate movement on which the two VLBI antennas are mounted (Campbell, 2004). Therefore, the linear trend is the known constituent, so in Equation (6.5), let $\underline{\mathbf{\Phi}} = [\mathbf{1}, \mathbf{t}]$. Figure 7.11 shows the (weighted) cross-spectrograms of the VLBI series and each of the temperature series (after removing the trends) along with their stochastic surfaces at 99% confidence level. The coherency between the two series at the cyclic frequency 1 c/a (365 days) can be seen in the figure. From the cross-spectrograms, one can observe that both temperature series show relatively small coherency at 1 c/a with the VLBI series for the time period from 1984 to 1991. This small coherency is due to large measurement errors for the VLBI series in that period (Figure 7.10a). As it is shown in Chapter 6, considering the error bars of the VLBI series improves the percentage variances of the spectral peaks in the LSWS, and it results in improved coherency with the temperature series.

The VLBI series has approximately the same phase shifts over time with both temperature series at 1 c/a (the constituents of the temperature series at 1 c/a have approximately zero phase shifts) indicating simultaneous contraction and expansion at both antennas which results in significant periodic change in the baseline length (see arrows). The small negative phase differences mean that the constituents of each temperature series lead the constituents of VLBI series at period of one year, indicating that the periodic change in the VLBI baseline length slowly occurs after the temperature variation. In other words, depending on the design of the antennas, the annual temperature variation can affect the geometry of the antennas by several millimetres (Wresnik et al., 2007). The constituents of the VLBI series and each temperature series at cyclic frequency 1 c/a have slightly smaller phase differences at certain time periods from 1989 to 2014 when selecting smaller window size parameters (e.g., $L_1 = 1$ and $L_0 = 10$).

For further investigation, the annual peaks are also suppressed from the VLBI baseline length series as well as from the Westford temperature series, and the LSCWS of the residual series is shown in Figure 7.12a. The weak peak at 1.6 c/a in Figure 7.12a may indicate that the Westford antenna is being deformed with a period of approximately 228 days (unusual atmospheric temperature variations) from year 1997 to year 2001 with almost zero phase shift (see inside the circle) as was discussed in Chapter 6.

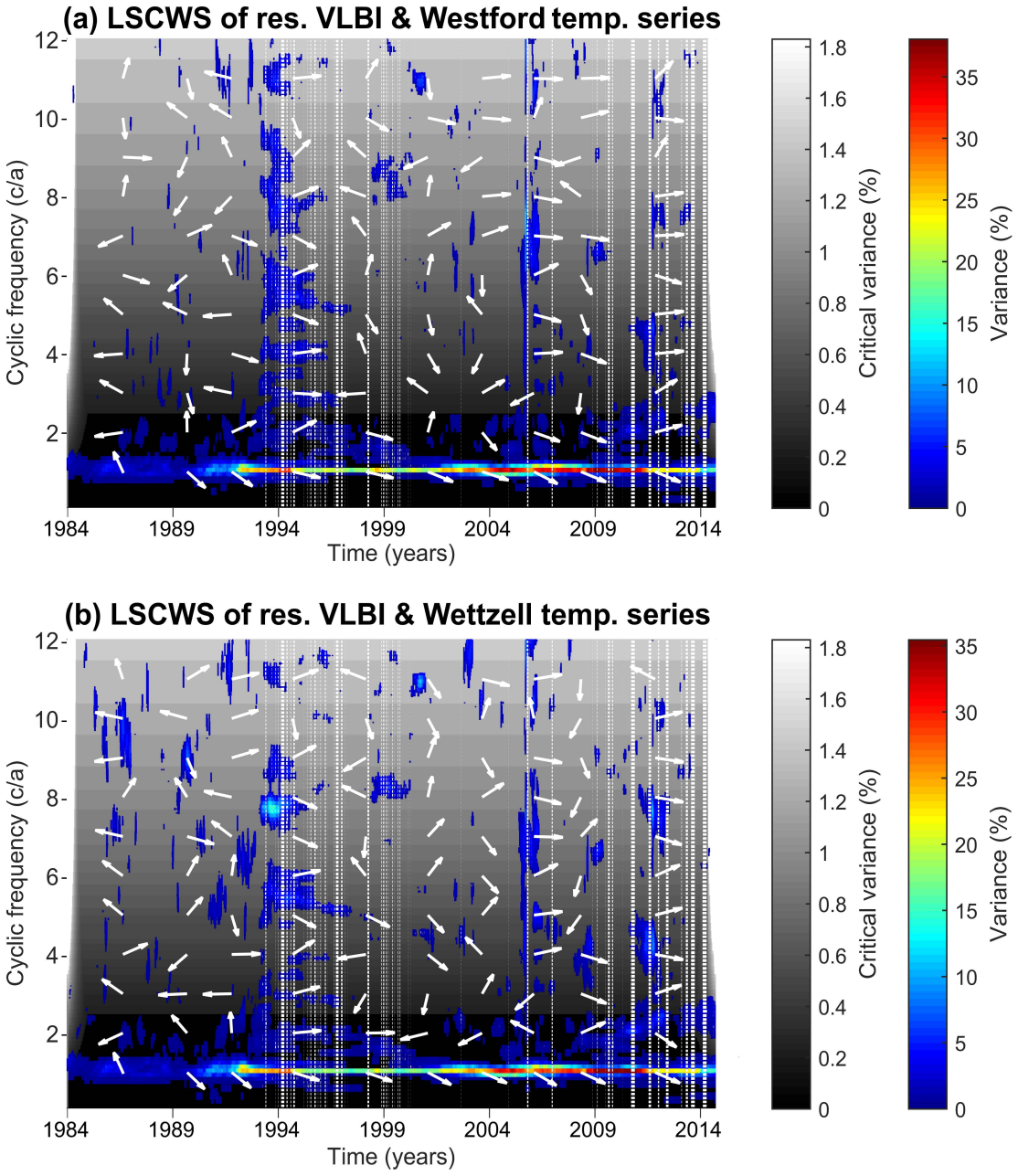

Figure 7.11: The LSCWS of the residual VLBI series and the (a) Westford and (b) Wettzell temperature series with their stochastic surfaces at 99% confidence level (gray) and phase differences (white arrows).

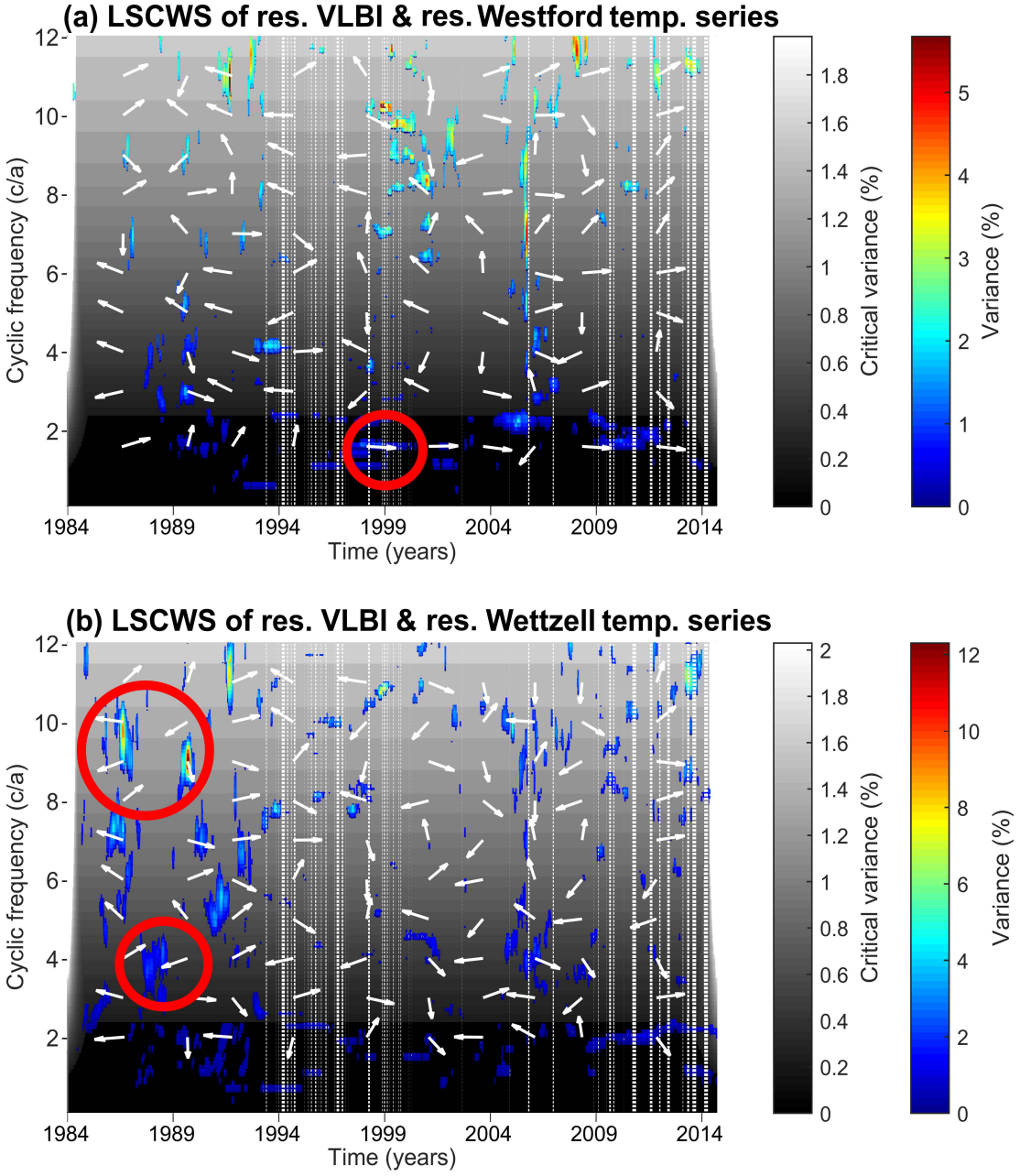

Figure 7.12: The LSCWS of the residual VLBI series and the residual (a) Westford and (b) Wettzell temperature series with their stochastic surfaces at 99% confidence level (gray) and phase differences (white arrows). Many of the significant peaks are shown inside the circles.

To compare the results with the result of Wettzell temperature series shown in Chapter 6, the trends, the annual peaks, and peaks at cyclic frequency 1.6 c/a are simultaneously removed from the VLBI series. The trends and annual peaks are also removed from the Wettzell temperature series and the LSCWS of the two residual series is shown in Figure 7.12b. The statistically significant coherency between the constituents of the two series in the LSCWS indicates the deformation of Wettzell antenna, possibly affecting the baseline length at certain periods (see inside the circles). The semiannual length variation of the baseline length can be possibly linked to the temperature variation as observed in the cross-spectrograms shown in Figure 7.12 (since 2004). The dashed vertical lines in the cross-spectrograms are due to significant data gaps.

7.6 Summary

I proposed a new and robust method (the LSCWA) of analyzing two (or more) time series simultaneously to search and estimate the amplitudes and phase differences of their common constituents. The LSCWA is able to estimate the coherency between the components of any two equally/unequally spaced time series statistically regardless of their sampling rates, opening a breakthrough in signal processing. I showed its outstanding performance on two synthetic time series.

The time series analyses presented in this chapter are concrete examples of real time series sampled at different rates and represent time-variant systems where the Fourier and wavelet transform or other traditional methods cannot be applied directly.

We applied the LSCWA to investigate the disturbances in the gravitational gradients observed by GOCE satellite that arise from plasma flow in the ionosphere represented by Poynting electromagnetic energy flux. We discovered that the ionospheric variations are leading the disturbances in GGT trace series, and they are coherent significantly at certain periods. Our analyses showed that the variations observed in the two series have a period of about 5 minutes, corresponding to spatial variations of around 2000 km.

The coherence analyses of the VLBI baseline length and the atmospheric temperature variations at the Westford and Wettzell stations showed that at certain time and periods the baseline length was affected by temperature variations. In particular, we observed that the annual temperature variations affected the baseline length significantly over time. I also showed that using the covariance matrix associated with the VLBI baseline length can improve the percentage variances of the spectral peaks in the cross-spectrogram and can produce more reliable results.

8 Discussion, Conclusions and Future Directions

The Least-Squares Spectral Analysis (LSSA) is an appropriate method for analyzing non-stationary and unequally spaced time series. Contrary to Fourier analysis, the LSSA can account for constituents of known forms, such as trends and/or datum shifts simultaneously with periodic constituents. The covariance matrix associated with the time series (if known) can also be accounted for in the analysis. No editing of the time series of any sort is necessary when using the LSSA. In the LSSA, frequencies are examined one by one (out-of-context), and so, when analyzing unequally spaced time series, the least-squares spectrum shows spectral leakages, although the signal frequencies may be estimated accurately by increasing the resolution of frequencies (zooming in). For data regularization purposes (e.g., seismic data regularization), these leakages slow down the process and may cause significant errors in the final regularization results. However, the spectral leakages can be effectively mitigated by simultaneously suppressing the statistically significant peaks in the spectrum (simultaneously estimating the signal frequencies) as done in the LSSA, when considering the constituents of known forms. On the other hand, if a non-stationary time series comprises constituents of low/high frequency variability over time, neither the LSSA nor the Fourier analysis are appropriate for the analysis because both methods decompose the time series into the frequency domain, disabling the nature of constituents in the time series.

Starting from the fundamental concept and very desirable properties of the LSSA, I have accomplished significant innovations in the way the analysis of a time/data series or a set of time/data series must be approached to achieve reliable results, the significance of which can be assessed by rigorous statistical analysis. The philosophy behind my contributions was that we needed a method or methods for the analyses of any type of data series that does not require any alteration or preprocessing of the actual measurements. This philosophy essentially dictated the development of new methods to meet the needs of the analyses and the type of the data, unlike the most popular approaches that necessitate the editing of the data to meet the stringent requirements of the method of analysis. To this end, beyond resolving the common problem of spectral leakages in the data series analyses, I developed two robust methods, namely, the Least-Squares Wavelet Analysis (LSWA) and Least-Squares Cross-Wavelet Analysis (LSCWA) that supersede all state-of-the-art Fourier-based methods by achieving superior results demonstrated with a plethora of synthetic and experimental time series.

Table 8.1: Appropriate analysis methods for various types of time series without the need for their editing.

Properties of one time series	**Appropriate analyses**
Equally spaced, equally weighted, stationary	LSSA, Fourier transform, wavelet transform, WWZ, LSWA
Equally spaced, equally weighted, non-stationary	Wavelet transform, WWZ, LSWA
Unequally spaced and/or unequally weighted, non-stationary	WWZ, LSWA
Properties of two or more time series	**Appropriate analyses**
Same sampling rates, equally spaced, equally weighted, stationary	LSCSA, XWT, LSCWA
Same sampling rates, equally spaced, equally weighted, non-stationary	XWT, LSCWA
Different sampling rates, equally/unequally spaced, equally/unequally weighted, stationary	LSCSA, LSCWA
Different sampling rates, equally/unequally spaced, equally/unequally weighted, non-stationary	LSCWA

After an exhaustive and detailed review and comparisons of the most popular methods of data analysis performed throughout this dissertation, I was compelled to develop Table 8.1 that summarizes all the appropriate methods of spectral analysis that could be used to analyze a series without prior editing (Ghaderpour et al., 2018a). The main innovations of this dissertations are summarized in the following four sections that are extensions of the LSSA. The first two methods (ALLSSA and MALLSSA) are established in the frequency or wavenumber domain, and the other two methods (LSWA and LSCWA) are established in the time-frequency or distance-wavenumber domain:

8.1 AntiLeakage Least-Squares Spectral Analysis (ALLSSA)

The ALLSSA is an iterative method based on the LSSA that usually uses an equally spaced preselected set of wavenumbers to accurately estimate the statistically significant spectral peaks. After simultaneously suppressing several significant spectral peaks with highest percentage variance (power/energy), this method revisits the previously estimated wavenumbers to estimate them more accurately in the next step, and so it significantly reduces the computational cost while it estimates the signal wavenumbers and amplitudes more accurately than the LSSA. Once the wavenumbers and amplitudes are estimated, they can be used to reconstruct the data series to any desired equally spaced series (regularization of data series). The ALLSSA is applied to regularize irregularly sampled seismic data more accurately, efficiently, and computationally faster than the state-of-the-art methods, such as the ALFT, ASFT, and IMAP (Ghaderpour et al., 2018b).

The ALLSSA is a fast and very accurate new method of regularizing data/time series that can also attenuate the random noise down to a certain confidence level. This method takes into account the corre-

lations among the sinusoidal basis functions as well as the constituents of known forms by considering the covariance matrix associated with the data/time series. I demonstrated how an appropriate selection of the weight matrix (the inverse of the covariance matrix) adapts the method for regularizing non-stationary and more complex data series.

Simultaneously considering the correlation among the sinusoidal basis functions is the key point in the ALLSSA. This correlation will be considered when the matrix of normal equations is being inverted. In practice, this matrix is often invertible because the seismic traces are usually inherently unequally spaced. However, one may calculate the pseudo-inverse of this matrix when it is singular. For seismic data with very complex geologic structures that have severe aliasing, the ALLSSA may not be reliable for regularization as this method is established mainly for the wide-sense stationary seismic data like the ALFT, ASFT, and IMAP.

8.2 Multichannel AntiLeakage Least-Squares Spectral Analysis (MALLSSA)

Aliasing is a crucial problem in many fields of science, such as seismology. When interpolating seismic data (demanded for various seismic exploration purposes) that are coarsely sampled, aliasing may introduce significant artifacts (may generate false events) in the final regularized seismic data. Additional information or assumption is a must when regularizing coarsely sampled data series. The MALLSSA is a robust method that uses the spatial gradients of data series to accurately estimate the true signal wavenumbers required for reconstruction of the seismic data on any desired regular grids. The MALLSSA is successfully applied to accurately regularize coarsely sampled seismic data presenting severe aliasing (Ghaderpour, 2018).

A limitation of the MALLSSA is the potential overlap of two or more spectral replicas at the same wavenumber, similar to the MIMAP. However, the effect of the replicas in the MALLSSA may be improved in practice because the wavenumbers are being estimated simultaneously and more accurately at each round of iterations. An appropriate windowing technique can be applied to adapt the MALLSSA to seismic data with complex geologic structures as described in Chapter 4.

8.3 Least-Squares Wavelet Analysis (LSWA)

The selection of sinusoidal basis functions in the LSWA allows one to directly decompose a time series into the time-frequency domain, considering the correlations among the sine and cosine functions and other constituents of known forms as well as any correlated noise. The LSWA allows one to compute spectrograms rigorously for equally or unequally spaced time series without any preprocessing, modification or editing. When the fully populated covariance matrix associated with the time series is provided, the LSWA considers the correlations between the entire time series values in each segment. For equally spaced and equally

weighted time series, when the dilation window size parameter is set to zero (i.e., $L_1 = 0$), the LSWS has similar time and frequency resolution to the STFT, and it has similar time and frequency resolution to the CWT when this parameter is greater than zero (i.e., $L_1 > 0$). In the LSWA, when the dilation window size parameter is greater than zero, the size of the translating window decreases when the frequency increases and vice versa, allowing the detection of all short- and long-duration constituents simultaneously, particularly when they are in attenuation (e.g., free oscillations of the Earth, resonant phenomena, etc.). The stochastic surfaces defined for the spectrogram show the significant spectral peaks at a certain confidence level (usually 95% or 99%). The LSWA is applied to analyze many time series with various characteristics to show it supersedes the state-of-the-art time series analysis methods, such as the very popular continuous wavelet transform and its empirical cone of influence (Ghaderpour and Pagiatakis, 2017).

In the LSWA, care must be taken when selecting the window size parameters: the dilation parameter (L_1) and additional fixed number of data points (L_0) in and around the window. For instance, if the window location approaches to a large gap in an unequally spaced time series, then the LSWA will borrow data points corresponding to the neighbouring time period when selecting a large number of additional data points (L_0). Taking these neighbouring data points into account may affect the percentage variance of the true signal peaks in the spectrogram and may not be reliable when the time series is strongly non-stationary. Therefore, the researchers need to know some basic information of the time series that are studying for the selection of the window size parameters. The LSWA offers such alternative because it is a robust method of analyzing time series containing energy and/or power signals. After detecting power signals (e.g., stationary sinusoidal signals of constant amplitudes) and trends, they can be simultaneously removed from the time series to search for hidden energy signals (e.g., transient signals).

The computational complexity of the LSWA will be depended on several factors, such as the window size parameters, covariance matrices, number of constituents of known forms, set of frequencies, and the translating windows (overlapping or non-overlapping). The LSWA can be set up to achieve $\mathcal{O}(n)$ complexity per frequency like the fast algorithm for CWT, when windows do not overlap and covariance matrices are ignored, or it can be as slow as $\mathcal{O}(n^3)$ for a time series of n data points with n frequencies to be examined when the windows overlap. The appropriate selection of the window size parameters can considerably reduce the computational cost compared to the WWZ. In many fields of science, the quality and reliability of the spectral peaks corresponding to periodic and/or aperiodic signals are far more important than the speed.

8.4 Least-Squares Cross-Wavelet Analysis (LSCWA)

The LSCWA is a novel method of analyzing two (or more) time series together, based on the LSWA. The LSCWA can be applied to any type of time series and generates fairly simple and meaningful results including the confidence level and phase differences between the two time series, and it shows how much

the constituents of the time series in a particular frequency are coherent and whether the coherency is significant at certain confidence level using rigorous statistical inference. In the LSCWA, the time series do not have to be equally spaced and equally weighted, and they do not need to have the same sampling rate. The LSCWA can analyze non-stationary time series consisting of constituents of variable amplitude and frequency variability over time. The time series analyses presented in Chapter 7 are concrete examples of real time series which are sampled at different rates and represent time-variant systems, where the Fourier transform and wavelet transform or other traditional methods cannot be applied directly. Through a number of synthetic and real time series, I showed that the LSCWA does not require any preprocessing or editing of the original time series superseding the state-of-the-art methods (Ghaderpour et al., 2018a).

In many applied sciences, researchers would like to study the impact of one phenomenon on another by analyzing two or more time series (or data series that are measurements recorded over other quantities, such as distance instead of time, e.g., seismic data series) obtained for each phenomenon. The LSCWA is a powerful tool that can efficiently and directly show the coherency and phase differences between the time series to make a reliable decision. Note that the LSWA and LSCWA can be applied to data series instead of time series, and so the decomposition is in time-wavenumber domain instead of time-frequency domain. The properties of the time series analyses are listed in Table 8.1.

When analyzing unequally spaced time series, spectral leakages appear in the spectra and spectrograms, often statistically insignificant. However, the spectral leakages can be significantly mitigated by simultaneously suppressing the significant spectral peaks as done in the LSSA and LSWA. In the LSCWA, the effect of the leakages can be further mitigated because they appear randomly in the spectrograms of each time series. Therefore, their cross-spectrogram mainly shows the coherency of the true signals. Furthermore, if by any chance the random errors (e.g., colored noise) that might be present in two time series appear to have similar behaviour in certain time periods, then they may also contaminate the cross-spectrograms.

8.5 Future Directions

The potential applications of my proposed methods, the ALLSSA, MALLSSA, LSWA, and LSCWA are not restricted to seismology and geodesy, and their performances in other fields of science, such as astronomy, atmospheric sciences, medical sciences and finance are expected to enhance the results of the analyses. Here I provide several potential applications of my proposed methods in other fields of science that can be implemented in the future.

- In geophysical time series analysis, there are cases in which time series show discontinuities, for example in GPS (Global Positioning System) coordinate time series that may have discontinuities (i.e., sudden offset or "jump", instrument failure, antenna and reference frame changes) in its values or trend due

to tectonic phenomena. Such cases are also observed in time series of gravimeter measurements. Since the LSWA and LSCWA decompose time series segment-wise (depending on the time and frequency), selecting the column vector of all ones and the position vector for the constituents of known forms will automatically consider the discontinuities simultaneously with the periodic constituents thus, producing more reliable results for detecting and quantifying signals.

- Astronomers are usually interested in analyzing light curves of variable stars whose brightness as seen from the Earth fluctuates (Foster, 1996). Such astronomical time series are often unequally spaced and weighted and can be obtained from the American Association of Variable Star Observers (AAVSO), www.aavso.org/main-data. The LSWA is expected to show its robust performance in detecting and quantifying periodic and aperiodic signals in these time series because the LSWA is somewhat similar to the WWZ in that they are both least-squares based methods.

- Atmospheric scientists or agronomists may want to interpolate/extrapolate atmospheric temperature time series, recorded in certain stations located around a farm that can be irregularly spaced, to use it for various purposes, such as crop disease forecasting (Wolf et al., 2003). The ALLSSA can also be useful for these purposes.

- Physicians may be interested in studying the coherency between the Heart Rate Variability (HRV) and brain waves recorded by the Electroencephalography method (EEG) (Niedermeyer and da Silva, 2005). The LSCWA may be applied to investigate the coherency between the components of these time series rigorously, or the LSWA may be applied to the time series obtained by the EEG method to search for brain tumors and other space-occupying lesions.

- Financial analysts may be interested in analyzing certain time series to estimate trends in the overall unemployment rate as well as any periodicity in such time series, or occasional trends in price movement in intraday transaction prices of International Business Machines (IBM) stock and its coherency with transactions that resulted in price change, etc. (Tsay, 2010). The ALLSSA and LSWA may be able to accurately estimate such trends simultaneously with other periodic or aperiodic components in such time series for making a reliable decision.

The iterative antileakage technique in the ALLSSA may also be applied alongside the LSWA to further mitigate spectral leakages in the spectrogram for data regularization purposes. Using other sophisticated windows, such as Dolph-Chebyshev and Kaiser, instead of the Gaussian window, in the weighted ALLSSA and LSWA may improve the quality of the spectral peaks by further reducing spectral leakages, and it is subject to future work. I am planning to extend the ALLSSA and MALLSSA to higher dimension cases and show their capabilities in regularization. It is expected that the generalizations of the ALLSSA and MALLSSA

to higher dimensions will perform better in regularization of seismic data. A MATLAB software package including a GUI (Graphical User Interface) is being developed for the ALLSSA, LSWA, and LSCWA (follow the updates at www.ghader.org or www.researchgate.net/profile/Ebrahim_Ghaderpour), and a manuscript is currently being prepared to submit to Computers and Geosciences.

Bibliography

R. Abma and J. Claerbout. Lateral prediction for noise attenuation by *t-x* and *f-x* techniques. *Geophysics*, 60(6):1887–1896, 1995.

R. Abma and N. Kabir. 3D interpolation of irregular data with a POCS algorithm. *Geophysics*, 71(6): E91–E97, 2006.

M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. National Bureau of Standards Applied Mathematics Series, USA Department of Commerce, 1972.

U. Amato, A. Antoniadis, and M. Pensky. Wavelet kernel penalized estimation for non-equispaced design regression. *Statistics and Computing*, 16(1):37–55, 2006.

O. Amm. Ionospheric elementary current systems in spherical coordinates and their application. *Journal of Geomagnetism and Geoelectricity*, 49(7):947–955, 1997.

O. Amm and A. Viljanen. Ionospheric disturbance magnetic field continuation from the ground to the ionosphere using spherical elementary current systems. *Earth, Planets and Space*, 51(6):431–440, 1999.

B. L. Barnhart. *The Hilbert-Huang transform: Theory, application, development*. PhD thesis, University of Iowa, USA, 2011.

N. Baydar and A. Ball. A comparative study of acoustic and vibration signals in detection of gear failures using Wigner-Ville distribution. *Mechanical Systems and Signal Processing*, 15(6):1091–1107, 2001.

B. Biondi. *3d Seismic Imaging: Three Dimensional Seismic Imaging*. Society of Exploration Geophysicists, 2006.

R. G. Brown and P. Y. C. Hwang. *Introduction to random signals and applied Kalman filtering, 383*. John Wiley and Sons, Inc, New Jersey, USA, 2012.

J. Campbell. VLBI for geodesy and geodynamics, the role of VLBI in astrophysics, astrometry and geodesy, Chapter 22, NATO science series II. *Mathematics, Physics and Chemistry*, 135:359–381, 2004.

S. Chen and B. L. Luk. Adaptive simulated annealing for optimization in signal processing applications. *Signal Processing*, 79(1):117–128, 1999.

Y. Chen and M. Q. Feng. A technique to improve the empirical mode decomposition in the Hilbert-Huang transform. *Earthquake Engineering and Engineering Vibration*, 2(1):75–85, 2003.

Y. Chen, L. Zhang, and L. Mo. Seismic data interpolation using nonlinear shaping regularization. *Journal of Seismic Exploration*, 24(5):327–342, 2015.

T. A. C. M. Classen and W. F. G. Mecklenbrauker. The Wigner distribution: A tool for time-frequency analysis (Parts I, II, III). *Philips Journal of Research*, 35, 1980.

A. Cohen, I. Fechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 45(5):485–560, 1992.

A. T. Craig, R. V. Hogg, and J. McKean. *Introduction to mathematical statistics (7th Edition), 694*. Pearson, 2013.

S. Crawley. *Seismic trace interpolation with nonstationary prediction-error filters*. PhD thesis, Stanford University, USA, 2000.

M. R. Craymer. *The least-squares spectrum, its inverse transform and autocorrelation function: Theory and some application in geodesy*. PhD thesis, University of Toronto, Canada, 1998.

M. Ďatig and T. Schlurmann. Performance and limitations of the Hilbert-Huang transformation (HHT) with an application to irregular water waves. *Ocean Engineering*, 31(14–15):1783–1834, 2004.

I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990.

C. L. Dolph. A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level. *Proceedings of the IRE*, 34(6):335–348, 1946.

B. Dragoset, E. Verschuur, I. Moore, and R. Bisley. A perspective on 3d surface-related multiple elimination. *Geophysics*, 75(5):75A245–75A261, 2010.

M. Abd El-Gelil and S. D. Pagiatakis. Least-squares self-coherence for sub-ngal signal detection in the superconducting gravimeter records. *Journal of Geodynamics*, 48(3):310–315, 2009.

M. Abd El-Gelil, S. D. Pagiatakis, and A. El-Rabbany. Frequency-dependent atmospheric pressure admittance of superconducting gravimeter records using least-squares response method. *Physics of the Earth and Planetary Interior*, 170(1–2):24–33, 2008.

J. Eriksson, E. Ollila, and V. Koivunen. Statistics for complex random variables revisited. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009. doi: 10.1109/ICASSP.2009.4960396.

M. Farge. Wavelet transforms and their applications to turbulence. *Annual Review of Fluid Mechanics*, 24: 395–457, 1992.

E. B. Fomalont and S. M. Kopeikin. The measurement of the light deflection from Jupiter: experimental results. *The Astrophysical Journal*, 598(1):704–711, 2003.

S. Fomel. *Three dimensional seismic data regularization*. PhD thesis, Stanford University, USA, 2001.

S. Fomel. Applications of planewave destruction filters. *Geophysics*, 67(6):1946–1960, 2002.

S. Fomel. Shaping regularization in geophysical-estimation problems. *Geophysics*, 72(2):R29–R36, 2007.

G. Foster. Wavelet for period analysis of unevenly sampled time series. *The Astronomical Journal*, 112(4): 1709–1729, 1996.

J. B. J. Fourier. *The analytical theory of heat*. Translated by Alexander Freeman, Cambridge, The University Press, 1878, 1822.

D. Gabor. Theory of communication. *J. IEE*, 93(26):429–457, 1946.

M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, F. Rossi, and R. Ulerich. *GNU Scientific Library. Reference Manual, Edition 2.3, for GSL Version 2.3, pp 612*. 2016.

J. J. Gao, A. Stanton, M. Naghizadeh, M. D. Sacchi, and X. Chen. Convergence improvement and noise attenuation considerations for beyond alias projection onto convex sets reconstruction. *Applied Geophysics*, 61(s1):138–151, 2013.

Z. Ge. Significance tests for the wavelet cross-spectrum and wavelet linear coherence. *Annales Geophysicae*, 26(12):3819–3829, 2008.

ZJ. Ge, JY. Li, SL. Pan, and XH. Chen. A fast-convergence POCS seismic denoising and reconstruction method. *Applied Geophysics*, 12(2):169–178, 2015.

R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35(2):227–246, 1972.

E. Ghaderpour. *Asymptotic existence of orthogonal designs*. PhD thesis, University of Lethbridge, Canada, 2013.

E. Ghaderpour. Some equal-area, conformal and conventional map projections: A tutorial review. *Journal of Applied Geodesy*, 10(3):197–209, 2016.

E. Ghaderpour. Multichannel antileakage least-squares spectral analysis for seismic data regularization beyond aliasing. *Journal of Applied Geophysics*, 2018. doi: Under Review.

E. Ghaderpour and S. D. Pagiatakis. Least-squares wavelet analysis of unequally spaced and non-stationary time series and its applications. *Mathematical Geosciences*, 49(7):819–844, 2017.

E. Ghaderpour, E. S. Ince, and S. D. Pagiatakis. Least-squares cross-wavelet analysis and its applications in geophysical time series. *Journal of Geodesy*, 92(10):1223–1236, 2018a.

E. Ghaderpour, W. Liao, and M. P. Lamoureux. Antileakage least-squares spectral analysis for seismic data regularization and random noise attenuation. *Geophysics*, 83(3):V157–V170, 2018b.

J. Gilles. Empirical wavelet transform. *IEEE Transaction on Signal Processing*, 61(16):3999–4010, 2013.

A. G. Glen, L. M. Leemis, and J. H. Drew. Computing the distribution of the product of two continuous random variables. *Computational Statistics and Data Analysis*, 44(3):451–464, 2004.

A. Grinsted, J. C. Moore, and S. Jeverjeva. Application of the cross-wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11:561–566, 2004.

A. Grossmann, J. Morlet, and T. Paul. Transforms associated to square integrable group representations. I General results. *Journal of Mathematical Physics*, 26(10):2473–2479, 1985.

A. Grossmann, J. Morlet, and T. Paul. Transforms associated to square integrable group representations. II examples. *Ann. Inst. Henri Poincare, Physique Theorique*, 45(3):293–309, 1986.

Y. Guang, X. Sun, M. Zhang, X. Li, and X. Liu. Study on ways to restrain end effect of Hilbert-Huang transform. *Journal of Computers*, 25(3):22–31, 2014.

N. Gülünay. Seismic trace interpolation in the Fourier transform domain. *Geophysics*, 68(1):355–369, 2003.

Z. Guo, Y. Zheng, and W. Liao. High fidelity seismic trace interpolation. *SEG, Expanded Abstracts*, pages 3915–3919, 2015. doi: 10.1190/segam2015-5923716.1.

P. Hall and B. A. Turlach. Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *The Annals of Statistics*, 25(5):1912–1925, 1997.

F. J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978.

D. A. Harville. *Matrix algebra from a statistician's perspective, 630*. Springer Science and Business Media, LLC, 2008.

M. Hazewinkel. *Gauss-Kronrod quadrature formula, Encyclopedia of Mathematics*. Springer. ISBN 978-1-55608-010-4, 2001.

F. Hlawatsch and F. Boudreaux-Bartels. Linear and quadratic time-frequency signal representations. *IEEE Signal Processing Magazine*, 9(2):21–67, 1992.

Y. Hollander, D. Kosloff, Z. Koren, and A. Bartana. Seismic data interpolation by orthogonal matching pursuit. *74th EAGE Conference and Exhibition incorporating EUROPEC*, 2012. doi: 10.3997/2214-4609.20148144.

R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, 2nd edition, 2012.

N. E. Huang and Z. Wu. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2):1–23, 2008.

N. E. Huang and Z. Wu. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1):1–41, 2009.

N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A*, 454:903–995, 1998.

Y. Hui and S. D. Pagiatakis. Least-squares spectral analysis and its application to superconducting gravimeter data analysis. *Geo-spatial Information Science*, 7(4):279–283, 2004.

A. Hurrell. Voltage to pressure conversion: are you getting 'phased' by the problem? *Journal of Physics: Conference Series*, 1:57–62, 2004.

E. S. Ince. *GOCE gradiometer disturbances and their modeling by means of ionospheric dynamics*. PhD thesis, York University, Canada, 2017.

E. S. Ince and S. D. Pagiatakis. Effects of space weather on GOCE electrostatic gravity gradiometer measurements. *Journal of Geodesy*, 90(12):1389–1403, 2016.

E. S. Ince and S. D. Pagiatakis. GOCE gradiometer measurements response to ionospheric dynamics. *Journal of Geophysical Research: Space Physics*, 122(10):712–726, 2017.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions 2 (2nd ed.)*. Wiley. ISBN 978-0-471-58494-0, 1995.

J. Kaiser and R. Schafer. On the use of the i0-sinh window for spectrum analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(1):105–107, 1980.

S. M. Kay and S. L. Marple. Spectrum analysis- a modern perspective. *Proceedings of the IEEE*, 69(11): 1380–1419, 1981.

M. C. Kelley. *The Earth's Ionosphere*, volume 96. Plasma Physics and Electrodynamics. Academic press, 2009.

B. Liu and M. D. Sacchi. Minimum weighted norm interpolation of seismic records. *Geophysics*, 69(6): 1560–1568, 2004.

G. Liu and X. Chen. Seismic data interpolation using frequency domain complex nonstationary autoregression. *Geophysical Prospecting*, 66(3):478–497, 2017.

N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39 (2):447–462, 1976.

S. Mallat. *A Wavelet tour of signal processing, 637*. Academic Press, Cambridge UK, 1999.

A. Mathias, F. Grond, R. Guardans, D. Seese, M. Canela, and H. H. Diebner. Algorithms for spectral analysis of irregularly sampled time series. *Journal of Statistical Software*, 11(2):1–27, 2004.

S. Miller and D. Childers. *Probability and Random Processes*. Academic Press, 2012.

M. Misiti, Y. Misiti, G. Oppenheim, and J. M. Poggi. *Wavelet toolbox: User's Guide (MATLAB)*. 2015.

C. B. Moler. *Numerical computing with MATLAB*. SIAM. ISBN: 9780898716603, 2008.

J. Morlet. Sampling theory and wave propagation. *In: Chen C.H. (eds), Issues in Acoustic Signal–Image Processing and Recognition. NATO ASI Series (Series F: Computer and System Sciences)*, 1:233–261, 1983.

E. Niedermeyer and F. L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams and Wilkins, 2005.

A. Özbek, A. K. Özdemir, and M. Vassallo. Interpolation by matching pursuit. *SEG Technical Program Expanded Abstracts*, pages 3254–3258, 2009. doi: 10.1190/1.3255534.

S. Pagiatakis. Stochastic significance of peaks in the least-squares spectrum. *Journal of Geodesy*, 73(2): 67–78, 1999.

S. D. Pagiatakis, H. Yin, and M. Abd El-Gelil. Least-squares self-coherency analysis of superconducting gravimeter records in search for the slichter triplet. *Physics of the Earth and Planetary Interiors*, 160(2): 108–123, 2007a.

S. D. Pagiatakis, H. Yin, and M. Abd El-Gelil. Least-squares self coherency analysis of superconducting gravimeter records in search for the Slichter triplet. *Physics of the Earth and Planetary Interior*, 160(2): 108–123, 2007b.

N. Peterseim, A. Schlicht, C. Stummer, and Y. Wiyong. Impact of cross winds in polar regions on goce accelerometer and gradiometer data. *4th International GOCE User Workshop, Munich, Germany*, 2011.

P. Psimoulis, S. Pytharouli, D. Karambalis, and S. Stiros. Potential of Global Positioning System (GPS) to measure frequencies of oscillations of engineering structures. *Journal of Sound and Vibration*, 318(3): 606–623, 2008.

C. I. Puryear. *Constrained least-squares spectral analysis: Application to seismic data*. PhD thesis, University of Houston, USA, 2012.

C. I. Puryear, O. N. Portniaguine, C. M. Cobos, and J. P. Castagna. Constrained least-squares spectral analysis: Application to seismic data. *Geophysics*, 77(5):143–167, 2012.

S. Qian. *Introduction to time-frequency and wavelet transforms*. Prentice-Hall Inc. Upper Saddle River, NJ, 2002.

C. R. Rao and S. K. Mitra. Generalized inverse of a matrix and its applications. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1:601–620, 1972.

H. Ren, Y. L. Wang, M. Y. Huang, Y. L. Chang, and H. M. Kao. Ensemble empirical mode decomposition parameters optimization for spectral distance measurement in hyperspectral remote sensing data. *Remote Sensing*, 6(3):2069–2083, 2014.

J. O. Robertsson, I. Moore, M. Vassallo, A. K. Özdemir, D. J. van Manen, and A. Özbek. On the use of multicomponent streamer recordings for reconstruction of pressure wavefields in the crossline direction. *Geophysics*, 73(5):A45–A49, 2008.

R. Rummel, W. Yi, and C. Stummer. GOCE gravitational gradiometry. *Journal of Geodesy*, 85(11):777–790, 2011.

S. Sardy, D. B. Percival, A. G. Bruce, H. Gao, and W. Stuetzle. Wavelet shrinkage for unequally spaced data. *Statistics and Computing*, 9(1):65–75, 1999.

J. Scales. *Theory of seismic imaging*. Golden, Colorado: Samizdat Press, 1997.

J. D. Scargle. Studies in astronomical time series analysis: II - Statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal, Part 1*, 263:835–853, 1982.

M. Schonewille, A. Klaedtke, and A. Vigner. Antialias antileakage Fourier transform. *SEG Technical Program Expanded Abstracts*, pages 3249–3253, 2009. doi: 10.1190/1.3255533.

J. Seberry, B. J. Wysocki, and T. A. Wysocki. On some applications of Hadamard matrices. *Metrika*, 62 (2–3):221–239, 2005.

I. V. Selesnick. Wavelet transform with tunable Q-factor. *IEEE Transactions on Signal Processing*, 59(8): 3560–3575, 2011.

L. F. Shampine. Vectorized adaptive quadrature in MATLAB. *Journal of Computational and Applied Mathematics*, 211(2):131–140, 2008.

H. S. Shapiro and R. A. Silverman. Alias-free sampling of random noise. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):225–248, 1960.

C. Siemes, R. Haagmans, M. Kern, G. Plank, and R. Floberghagen. Monitoring GOCE gradiometer calibration parameters using accelerometer and star sensor data: methodology and first results. *Journal of Geodesy*, 86(8):629–645, 2012.

S. Sinha, P. S. Routh, P. D. Anno, and J. P. Castagna. Spectral decomposition of seismic data with continuous-wavelet transform. *Geophysics*, 70(6):19–25, 2005.

S. Spitz. Seismic trace interpolation in the *f-x* domain. *Geophysics*, 56(6):785–794, 1991.

R. R. Steeves. A statistical test for significance of peaks in the least-squares spectrum. *Collected Papers of Geodetic Survey, Dept. of Energy, Mines and Resources, Surveys and Mapping, Ottawa*, pages 149–166, 1981.

E. M. Stein and R. Shakarchi. *Fourier analysis*. Princeton University Press, USA, 2003.

A. Stuart and K. Ord. *Kendall's Advanced Theory of Statistics, Distribution Theory I, 6th ed.* Wiley. ISBN 0470665300, 2010.

C. Stummer, C. Siemes, R. Pail, B. Frommknecht, and R. Floberghagen. Upgrade of level 1b gradiometer processor. *Advances in Space Research*, 49(4):739–752, 2012.

O. Titov. Spectral analysis of the baseline length time series from VLBI data. *IVS General Meeting Proceedings*, pages 315–319, 2002.

C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61–78, 1998.

D. Trad. Five-dimensional interpolation: Recovering from acquisition constraints. *Geophysics*, 74(6):V123–V132, 2009.

R. S. Tsay. *Analysis of financial time series, 3rd edition*. Wiley, 2010.

M. E. Van Valkenburg. *Network analysis (3rd Edition)*. Pearson, 2006.

P. Vaníček. Approximate spectral analysis by least-squares fit. *Astrophysics and Space Science*, 4:387–391, 1969.

P. Vaníček. Further development and properties of the spectral analysis by least-squares. *Astrophysics and Space Science*, 12(1):10–33, 1971.

P. Vaníček and E. J. Krakiwsky. *Geodesy the concepts*. Amsterdam The Netherlands, University of New Brunswick, Canada, 1986.

M. Vassallo, A. Özbek, A. K. Özdemir, and K. Eggenberger. Crossline wavefield reconstruction from multi-component streamer data: Part 1-multichannel interpolation by matching pursuit (MIMAP) using pressure and its crossline gradient. *Geophysics*, 75(6):WB53–WB67, 2010a.

M. Vassallo, A. Özbek, A. K. Özdemir, WesternGeco, and K. Eggenberger. Crossline wavefield reconstruction from multicomponent streamer data: multichannel interpolation by matching pursuit. *SEG Technical Program Expanded Abstracts*, pages 3594–3598, 2010b. doi: 10.1190/1.3513597.

B. Wang, X. Chen, J. Li, and J. Cao. An improved weighted projection onto convex sets method for seismic data interpolation and denoising. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(1):228–235, 2016.

Y. Wang. Seismic trace interpolation in the *f-x-y* domain. *Geophysics*, 67(4):1232–1239, 2002.

A. B. Weglein, F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt. An inverse-scattering series method for attenuating multiples in seismic reflection data. *Geophysics*, 62(6):1975–1989, 1997.

D. E. Wells and E. J. Krakiwsky. *The method of least-squares*. Department of Surveying Engineering, University of New Brunswick, Canada, 1971.

D. E. Wells, P. Vaníček, and S. D. Pagiatakis. *Least-squares spectral analysis revisited*. Department of Surveying Engineering, University of New Brunswick, Canada, 1985.

D. H. D. West. Updating mean and variance estimates: an improved method. *Communications of the Association for Computing Machinery*, 22(9):532–535, 1979.

J. M. Weygand, O. Amm, A. Viljanen, V. Angelopoulos, D. Murr, M. J. Engebretson, H. Gleisner, and I. Mann. Application and validation of the spherical elementary currents systems technique for deriving ionospheric equivalent currents with the North American and Greenland ground magnetometer arrays. *Journal of Geophysical Research: Space Physics*, 116(A3):1978–2012, 2011.

H. Weyl. *The theory of groups and quantum mechanics*. Dutton, New York, USA, 1931.

E. D. De Wolf, L. V. Madden, and P. E. Lipps. Risk assessment models for wheat fusarium head blight epidemics based on within-season weather data. *Phytopathology*, 93(4):428–435, 2003.

J. Wresnik, H. Haas, J. Boehm, and H. Schuh. Modeling thermal deformation of VLBI antennas with a new temperature model. *Journal of Geodesy*, 81(6):423–431, 2007.

D. L. Wu, P. B. Hays, and W. R. Skinner. A least-squares method for spectral analysis of space-time series. *Journal of the Atmospheric Sciences*, 52:3501–3511, 1995.

S. Xu, Y. Zhang, D. Pham, and G. Lambaré. Antileakage Fourier transform for seismic data regularization. *Geophysics*, 70(4):V87–V95, 2005.

S. Xu, Y. Zhang, and G. Lambaré. Antileakage Fourier transform for seismic data regularization in higher dimensions. *Geophysics*, 75(6):WB113–WB120, 2010.

S. Xu, Y. Zhang, and B. Tang. 3D angle gathers from reverse time migration. *Geophysics*, 76(2):S77–S92, 2011.

P. Yang, J. Gao, and W. Chen. Curvelet-based POCS interpolation of nonuniformly sampled seismic records. *Journal of Applied Geophysics*, 79:90–99, 2012.

W. Yi, R. Rummel, and T. Gruber. Gravity field contribution analyses of GOCE gravitational gradient components. *Studia Geophysical et Geodetica*, 57(2):174–202, 2013.

A Comparison Between Simultaneous Least-Squares Spectral Analysis and Discrete Fourier Transform

The Discrete Fourier Transform (DFT) of a sequence of n regularly spaced complex numbers $f(x_1), \ldots, f(x_n)$ into another sequence of complex numbers (Fourier coefficients) is defined as

$$\hat{f}(\omega_k) = \frac{1}{n} \sum_{\ell=1}^{n} f(x_\ell) e^{-2\pi i \omega_k x_\ell}, \tag{A.1}$$

where $x_\ell = \ell/n$, $\omega_k \in \mathbf{K} = \{-\frac{n}{2}, \ldots, \frac{n}{2} - 1\}$ if n is even, and $\omega_k \in \mathbf{K} = \{-\frac{n-1}{2}, \ldots, \frac{n-1}{2}\}$ if n is odd. Note that ω_k is referred to wavenumber that is the number of cycles per unit distance. Considering all the wavenumbers simultaneously, the DFT can be represented in matrix form as $\mathbf{f} = \frac{1}{n}\mathbf{D}^{\mathrm{T}}\mathbf{f}$, where

$$\mathbf{F} = \begin{bmatrix} \hat{f}(\omega_1) \\ \vdots \\ \hat{f}(\omega_n) \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} e^{2\pi i \omega_1 x_1} & e^{2\pi i \omega_2 x_1} & \ldots & e^{2\pi i \omega_n x_1} \\ \vdots & \vdots & \ldots & \vdots \\ e^{2\pi i \omega_1 x_n} & e^{2\pi i \omega_2 x_n} & \ldots & e^{2\pi i \omega_n x_n} \end{bmatrix}, \tag{A.2}$$

and 'T' denotes the conjugate transpose (Craymer, 1998). It can be seen that $\mathbf{f} = \mathbf{DF}$ (when traces are regularly sampled in space). The Euler's formula implies that $e^{2\pi i \omega_k x_\ell} = \cos(2\pi\omega_k x_\ell) + i \sin(2\pi\omega_k x_\ell)$, and so the LSSA expresses the columns of $\mathbf{D}$ as cosine and sine basis functions. More precisely, let $\mathbf{\Omega} = \{1, \ldots, \eta\}$, where $\eta = \frac{n}{2} - 1$ if n is even ($\eta = \frac{n}{2}$ is the Nyquist wavenumber), and $\eta = \frac{n-1}{2}$ if n is odd. When the real-valued version of the normal equations considered, the least-squares minimization uses the model $\mathbf{f} = \mathbf{\Phi c}$ to estimate $\mathbf{c}$ as $\hat{\mathbf{c}} = \mathbf{N}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{f}$, where $\mathbf{N} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$,

$$\mathbf{\Phi} = \begin{bmatrix} 1 & \cos(2\pi\omega_1 x_1) & \sin(2\pi\omega_1 x_1) & \cdots & \cos(2\pi\omega_\eta x_1) & \sin(2\pi\omega_\eta x_1) \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 1 & \cos(2\pi\omega_1 x_n) & \sin(2\pi\omega_1 x_n) & \cdots & \cos(2\pi\omega_\eta x_n) & \sin(2\pi\omega_\eta x_n) \end{bmatrix}, \tag{A.3}$$

and $\omega_k \in \mathbf{\Omega}$. The dimensions of $\mathbf{\Phi}$, $\mathbf{N}$, and $\hat{\mathbf{c}}$ are $n \times (2\eta+1)$, $(2\eta+1) \times (2\eta+1)$, and $(2\eta+1) \times 1$, respectively.

Note that the first column of design matrix $\mathbf{\Phi}$ above corresponds to the zero wavenumber. In fact, we could have two column vectors associated to the zero wavenumber: the first column for cosine whose elements are $\cos(0) = 1$ and the second column for sine whose elements are $\sin(0) = 0$. However, we avoid writing the zero column vector because it will make matrix $\mathbf{N}$ singular. Note that matrix $\mathbf{N}$ is a real square matrix, and it is invertible as long as the column vectors of $\mathbf{\Phi}$ are linearly independent. One may also include the column vector of $\cos(2\pi\frac{n}{2}x_\ell) = \cos(2\pi\frac{n}{2}\frac{\ell}{n}) = \cos(\ell\pi) = \pm 1$, $\ell = 1, \ldots, n$, in matrix $\mathbf{\Phi}$ above when n is even. This way $\mathbf{f}$ will be reconstructed accurately if $\mathbf{f}$ contains constituents of the Nyquist wavenumber $\frac{n}{2}$. Note that the column vector of sine is not included because $\sin(2\pi\frac{n}{2}x_\ell) = \sin(2\pi\frac{n}{2}\frac{\ell}{n}) = \sin(\ell\pi) = 0$ for $\ell = 1, \ldots, n$. To avoid the singularity of square matrix $\mathbf{N}$, the negative wavenumbers are excluded from matrix $\mathbf{\Phi}$. Craymer (1998) discussed that when $\mathbf{f}$ is real, one may only consider the positive wavenumbers. However, not much work is done on the LSSA for the case when $\mathbf{f}$ is complex. The analyses show that if $\mathbf{f}$ is complex, one may still consider only the positive wavenumbers for regularization (see below for further details).

Assume that n is even (a similar discussion for n odd). When a data series is regularly spaced,

$$\sum_{\ell=1}^{n} \cos(2\pi\omega_k x_\ell)\cos(2\pi\omega_u x_\ell) = \begin{cases} n & \text{for } \omega_k = \omega_u = -\frac{n}{2} \text{ or } 0 \\ \frac{n}{2} & \text{for } \omega_k = \omega_u \neq -\frac{n}{2} \text{ or } 0 \\ 0 & \text{for } \omega_k \neq \omega_u \end{cases} \tag{A.4}$$

$$\sum_{\ell=1}^{n} \sin(2\pi\omega_k x_\ell)\sin(2\pi\omega_u x_\ell) = \begin{cases} 0 & \text{for } \omega_k = \omega_u = -\frac{n}{2} \text{ or } 0 \\ \frac{n}{2} & \text{for } \omega_k = \omega_u \neq -\frac{n}{2} \text{ or } 0 \\ 0 & \text{for } \omega_k \neq \omega_u \end{cases} \tag{A.5}$$

$$\sum_{\ell=1}^{n} \cos(2\pi\omega_k x_\ell)\sin(2\pi\omega_u x_\ell) = 0 \quad \text{for all} \quad \omega \in \mathbf{K}. \tag{A.6}$$

Therefore, including only the positive wavenumbers in the LSSA also implicitly accounts for the identical response to both positive and negative wavenumbers and produces a one-sided spectrum that is physically meaningful. This will also effectively double the amplitude of the least-squares transform with respect to the DFT (Vaníček, 1969; Craymer, 1998). More precisely, it can be seen that

$$\hat{\mathbf{c}} = \mathbf{N}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{f} = \begin{bmatrix} \frac{1}{n}\sum_{\ell=1}^{n} f(x_\ell) \\ \frac{2}{n}\sum_{\ell=1}^{n} \cos(2\pi\omega_1 x_\ell) f(x_\ell) \\ \frac{2}{n}\sum_{\ell=1}^{n} \sin(2\pi\omega_1 x_\ell) f(x_\ell) \\ \vdots \\ \frac{2}{n}\sum_{\ell=1}^{n} \cos(2\pi\omega_\eta x_\ell) f(x_\ell) \\ \frac{2}{n}\sum_{\ell=1}^{n} \sin(2\pi\omega_\eta x_\ell) f(x_\ell) \end{bmatrix}, \qquad \mathbf{\Phi}\hat{\mathbf{c}} \approx \mathbf{f}. \tag{A.7}$$

The elements of column vector $\hat{\mathbf{c}}$ in Equation (A.7) from top to bottom are the estimated coefficients for the columns of $\mathbf{\Phi}$ in Equation (A.3) from left to right, respectively. Note that Vaníček (1969); Craymer (1998); Pagiatakis (1999) discussed the LSSA mainly for real time series (i.e., $\mathbf{f}$ a real vector) with its applications in geodesy and geodynamics. However, since design matrix $\mathbf{\Phi}$ is a real matrix, $\mathbf{f}$ can also be a complex vector (e.g., a temporal frequency slice). In this case, $\hat{\mathbf{c}}$ in Equation (A.7) will be a complex vector whose real and imaginary parts contain the estimated amplitudes corresponding to the real and imaginary parts of $\mathbf{f}$, respectively. One may also use Equation (A.7) for the real part and imaginary part of $\mathbf{f}$ individually (treating each part as a real data series) and then form vector $\hat{\mathbf{c}}$ by combining the estimated amplitudes of each part, producing similar regularization results to the previous case.

For irregularly spaced data series, Equations (A.4), (A.5) and (A.6) are no longer valid, i.e., the sinusoidal basis functions are no longer orthogonal. However, the correlation between them is taken into account when matrix $\mathbf{N}$ is being inverted. The selection of wavenumbers is a critical task for the simultaneous estimation of the least-squares spectrum. The correlation among the sinusoids must be taken into account carefully; otherwise, ill-conditioning in $\mathbf{N}$ can produce wrong results (Craymer, 1998). In the ALLSSA, the wavenumbers are accurately estimated in an iterative manner, producing an accurate one-sided spectrum for stationary data series. Note that in the ALLSSA, only the coefficients of significant components of a data series of size n will be estimated, and so the number of columns in Equation (A.3) may be much smaller than n. Also, note that we are not using the fast methods of the FFT for the ALFT, ASFT, and ALLSSA.

B An Optimization Method for the Least-Squares Spectral Analysis

A similar methodology used as in Wells et al. (1985); Craymer (1998) to obtain Equation (6.10). From the model $\mathbf{y} = \overline{\mathbf{\Phi}}\ \overline{\mathbf{c}} = \underline{\mathbf{\Phi}}\ \underline{\mathbf{c}} + \mathbf{\Phi}\ \mathbf{c}$, estimate $\overline{\mathbf{c}}$ using the least-squares method as $\hat{\overline{\mathbf{c}}} = \overline{\mathbf{N}}^{-1}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}$, where $\overline{\mathbf{N}} = \overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\overline{\mathbf{\Phi}}$. Now write $\overline{\mathbf{N}}^{-1}$ as

$$\overline{\mathbf{N}}^{-1} = \begin{bmatrix} \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}} & \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi} \\ \\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}} & \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{M_1} & \mathbf{M_2} \\ \\ \mathbf{M_3} & \mathbf{M_4} \end{bmatrix}, \tag{B.1}$$

where

$$\begin{aligned} \mathbf{M_1} &= \Big(\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}} - \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}\ (\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\Big)^{-1} \\ &= \underline{\mathbf{N}}^{-1} + \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}\ \mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}, \qquad \text{(B.2)} \\ \mathbf{M_2} &= -\mathbf{M_1}\ \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}\ (\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi})^{-1} = -\underline{\mathbf{N}}^{-1}\ \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}\ \mathbf{M_4}, \qquad \text{(B.3)} \\ \mathbf{M_3} &= -\mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}, \qquad \text{(B.4)} \\ \mathbf{M_4} &= \Big(\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}\Big)^{-1}, \qquad \text{(B.5)} \end{aligned}$$

and $\underline{\mathbf{N}} = \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}$. Note that it is assumed that matrices $\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}$ and $\underline{\mathbf{N}}$ are invertible. Thus, from

$$\begin{bmatrix} \mathbf{x} \\ \hat{\mathbf{c}} \end{bmatrix} = \hat{\overline{\mathbf{c}}} = \overline{\mathbf{N}}^{-1}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} = \begin{bmatrix} \mathbf{M_1} & \mathbf{M_2} \\ \\ \mathbf{M_3} & \mathbf{M_4} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \\ \\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \end{bmatrix}, \tag{B.6}$$

one obtains

$$\begin{aligned} \hat{\mathbf{c}} &= \mathbf{M_3}\ \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} + \mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \\ &= -\mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} + \mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{y} \\ &= \mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\Big(\mathbf{y} - \underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}\Big) \\ &= \mathbf{N}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\hat{\mathbf{g}}. \end{aligned} \tag{B.7}$$

where $\mathbf{N} = {\mathbf{M_4}}^{-1} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}$, and $\hat{\mathbf{g}} = \mathbf{y} - \underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{y}$ is the estimated residual segment. Note that for each frequency, $\mathbf{N}$ is a square matrix of order two, and so its inverse is computationally efficient.

C Calculation of the Stochastic Surface of the LSWS

A similar methodology, used in Steeves (1981); Pagiatakis (1999), is applied to obtain the stochastic surface of the least-squares wavelet spectrogram given by Equation (6.11). Note that the critical value of the least-squares spectrum given by Equation (3.10) is a special case of this methodology that is independent of frequency or wavenumber. Assume that time series $\mathbf{f}$ has been derived from a population of random variables following the multidimensional normal distribution. The covariance matrix is always Hermitian (it is equal to its conjugate transpose) and positive semi-definite. For simplicity, assume that the covariance matrix $\mathbf{C_f}$ associated to the time series $\mathbf{f}$ is invertible and so positive definite. Thus, $\mathbf{P} = {\mathbf{C_f}}^{-1}$ is positive definite, and so all its principal submatrices $\mathbf{P_y}$ are positive definite, thus invertible (Horn and Johnson, 2012). Note that in case of singular $\mathbf{C_f}$, one may obtain the pseudo-inverse of $\mathbf{C_f}$ and obtain similar results (Rao and Mitra, 1972; Pagiatakis, 1999). A square matrix $\mathbf{M}$ is called idempotent if $\mathbf{MM} = \mathbf{M}$. Idempotent matrices have many nice properties. For instance, $\text{rank}(\mathbf{M}) = \text{trace}(\mathbf{M})$ (Horn and Johnson, 2012).

Let $\overline{\mathbf{J}} = \overline{\mathbf{\Phi}}\ \overline{\mathbf{N}}^{-1}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}$, $\underline{\mathbf{J}} = \underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}$ and $\mathbf{J} = \mathbf{\Phi}\mathbf{N}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}$, where $\overline{\mathbf{N}} = \overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\overline{\mathbf{\Phi}}$, $\underline{\mathbf{N}} = \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}$ and $\mathbf{N} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi} - \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{J}}\mathbf{\Phi}$ (same as defined in Appendix B). Note that $\overline{\mathbf{\Phi}} = \begin{bmatrix} \underline{\mathbf{\Phi}} & \mathbf{\Phi} \end{bmatrix}$ is the $R(j,k) \times (q+2)$ design matrix. It can be seen that $\overline{\mathbf{J}}$ and $\underline{\mathbf{J}}$ are idempotent, and

$$\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{J}} = \begin{bmatrix} \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \\ \\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \\ \\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \end{bmatrix}, \tag{C.1}$$

$$\begin{aligned} \overline{\mathbf{N}}^{-1}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{J}} &= \begin{bmatrix} \mathbf{M_1} & \mathbf{M_2} \\ \\ \mathbf{M_3} & \mathbf{M_4} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \\ \\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{M_1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} + \mathbf{M_2}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \\ \\ \mathbf{M_3}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} + \mathbf{M_4}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \end{bmatrix} \\ &= \begin{bmatrix} \left(\underline{\mathbf{N}}^{-1} + \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}\ \mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\right)\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} - \underline{\mathbf{N}}^{-1}\ \underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\mathbf{\Phi}\ \mathbf{M_4}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \\ \\ -\mathbf{M_4}\ \mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} + \mathbf{M_4}\mathbf{\Phi}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \end{bmatrix} \\ &= \begin{bmatrix} \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \\ \\ \mathbf{0} \end{bmatrix}. \end{aligned} \tag{C.2}$$

Therefore,

$$\overline{\mathbf{J}}\underline{\mathbf{J}} = \overline{\mathbf{\Phi}}\ \overline{\mathbf{N}}^{-1}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y}\underline{\mathbf{J}} = \begin{bmatrix} \underline{\mathbf{\Phi}} & \mathbf{\Phi} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} \\ \\ \mathbf{0} \end{bmatrix} = \underline{\mathbf{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P_y} = \underline{\mathbf{J}}. \tag{C.3}$$

Taking the transpose of both sides of Equation (C.3), one also can see that $\mathbf{P_y}\underline{\mathbf{J}}\overline{\mathbf{J}} = \mathbf{P_y}\underline{\mathbf{J}}$. Since $\overline{\mathbf{J}}\underline{\mathbf{J}} = \underline{\mathbf{J}}$

(Equation (C.3)) and

$$\begin{aligned}\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}} &= (\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y})^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}\\ &= (\mathbf{y}-\underline{\boldsymbol{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\mathbf{y})^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}\\ &= \mathbf{y}^{\mathrm{T}}(\mathbf{I}-\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}\ \underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}})\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}\\ &= \mathbf{y}^{\mathrm{T}}(\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}-\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}) = \mathbf{0}, \end{aligned} \tag{C.4}$$

it can be seen that

$$\begin{aligned}\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\hat{\mathbf{g}} &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\big(\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y}\big)\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\mathbf{y}-\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\underline{\mathbf{J}}\ \mathbf{y}\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\mathbf{y}-\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\underline{\mathbf{J}}\ \mathbf{y}\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\mathbf{y}-\big(\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}\big)\ \underline{\mathbf{N}}^{-1}\underline{\boldsymbol{\Phi}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\ \mathbf{y}\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\mathbf{y}\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\boldsymbol{\Phi}}\ \hat{\mathbf{c}}\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\underline{\boldsymbol{\Phi}}\ \mathbf{x}+\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\boldsymbol{\Phi}\ \hat{\mathbf{c}}\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\boldsymbol{\Phi}\ \hat{\mathbf{c}}\\ &= \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\mathbf{J}\hat{\mathbf{g}}, \end{aligned} \tag{C.5}$$

where the parameters are the same as the ones in Equation (B.6). Therefore, Equation (6.11) is identical to

$$s(t_j,\omega_k) = \frac{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\hat{\mathbf{g}}}{\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\hat{\mathbf{g}}}. \tag{C.6}$$

Since $\overline{\mathbf{J}}$ and $\underline{\mathbf{J}}$ are idempotent and $\overline{\mathbf{J}}\underline{\mathbf{J}} = \underline{\mathbf{J}}$ and $\mathbf{P}_{\mathbf{y}}\underline{\mathbf{J}}\overline{\mathbf{J}} = \mathbf{P}_{\mathbf{y}}\underline{\mathbf{J}}$,

$$\begin{aligned}\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\hat{\mathbf{g}} &= \big(\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y}\big)^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\big(\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y}\big)\\ &= \mathbf{y}^{\mathrm{T}}\big(\mathbf{P}_{\mathbf{y}}-\mathbf{P}_{\mathbf{y}}\underline{\mathbf{J}}\big)\big(\overline{\mathbf{J}}\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y}\big)\\ &= \mathbf{y}^{\mathrm{T}}\big(\mathbf{P}_{\mathbf{y}}\overline{\mathbf{J}}\mathbf{y}-\mathbf{P}_{\mathbf{y}}\underline{\mathbf{J}}\ \mathbf{y}\big)\\ &= \mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\overline{\mathbf{J}}-\underline{\mathbf{J}}\big)\ \mathbf{y}, \end{aligned} \tag{C.7}$$

and

$$\begin{aligned}\hat{\mathbf{g}}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\hat{\mathbf{g}} &= \big(\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y}\big)^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y}\big)\\ &= \mathbf{y}^{\mathrm{T}}\big(\mathbf{P}_{\mathbf{y}}-\mathbf{P}_{\mathbf{y}}\underline{\mathbf{J}}\big)\big(\mathbf{y}-\underline{\mathbf{J}}\ \mathbf{y}\big)\\ &= \mathbf{y}^{\mathrm{T}}\big(\mathbf{P}_{\mathbf{y}}\mathbf{y}-\mathbf{P}_{\mathbf{y}}\underline{\mathbf{J}}\ \mathbf{y}\big)\\ &= \mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\mathbf{I}-\underline{\mathbf{J}}\big)\ \mathbf{y}. \end{aligned} \tag{C.8}$$

Therefore, Equation (C.6) can be written as

$$s(t_j,\omega_k) = \frac{\mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\overline{\mathbf{J}}-\underline{\mathbf{J}}\big)\ \mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\mathbf{I}-\underline{\mathbf{J}}\big)\ \mathbf{y}}. \tag{C.9}$$

The denominator of Equation (C.9) can be written as $\mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\mathbf{I}-\underline{\mathbf{J}}\big)\ \mathbf{y} = \mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\overline{\mathbf{J}}-\underline{\mathbf{J}}\big)\ \mathbf{y} + \mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\mathbf{I}-\overline{\mathbf{J}}\big)\ \mathbf{y}$. Let $Q_s = \mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\overline{\mathbf{J}}-\underline{\mathbf{J}}\big)\ \mathbf{y}$ (signal) and $Q_n = \mathbf{y}^{\mathrm{T}}\mathbf{P}_{\mathbf{y}}\big(\mathbf{I}-\overline{\mathbf{J}}\big)\ \mathbf{y}$ (noise). Equation (C.9) can be written as

$$s(t_j,\omega_k) = \frac{Q_s}{Q_s+Q_n} = \left(1+\frac{Q_n}{Q_s}\right)^{-1}. \tag{C.10}$$

Since $\overline{\mathbf{J}} - \underline{\mathbf{J}}$ is idempotent and

$$\begin{aligned}\operatorname{rank}\big(\mathbf{P}_{\mathbf{y}}\big(\overline{\mathbf{J}} - \underline{\mathbf{J}}\big)\big) &= \operatorname{rank}\big(\overline{\mathbf{J}} - \underline{\mathbf{J}}\big) \\ &= \operatorname{trace}\big(\overline{\mathbf{J}} - \underline{\mathbf{J}}\big) \\ &= \operatorname{trace}\big(\overline{\mathbf{J}}\big) - \operatorname{trace}\big(\underline{\mathbf{J}}\big) \\ &= q + 2 - q = 2, \end{aligned} \tag{C.11}$$

random variable Q_s is distributed as chi-squared distribution with 2 degrees of freedom, i.e., $Q_s \sim \chi^2_2$, where symbol $\sim$ is used to show that a random variable 'follows' a distribution (Craig et al., 2013, Chapter 9.8). Similarly, since $\overline{\mathbf{I}} - \overline{\mathbf{J}}$ is idempotent and

$$\begin{aligned}\operatorname{rank}\big(\mathbf{P}_{\mathbf{y}}\big(\mathbf{I} - \overline{\mathbf{J}}\big)\big) &= \operatorname{rank}\big(\mathbf{I} - \overline{\mathbf{J}}\big) \\ &= \operatorname{trace}\big(\mathbf{I} - \overline{\mathbf{J}}\big) \\ &= \operatorname{trace}\big(\mathbf{I}\big) - \operatorname{trace}\big(\overline{\mathbf{J}}\big) \\ &= R(j,k) - (q+2) = \Re, \end{aligned} \tag{C.12}$$

random variable Q_n is distributed as chi-squared distribution with $\Re$ degrees of freedom, i.e., $Q_n \sim \chi^2_{\Re}$ (Craig et al., 2013, Chapter 9.8). Since $\mathbf{P}_{\mathbf{y}}\big(\mathbf{I} - \overline{\mathbf{J}}\big)\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{P}_{\mathbf{y}}\big(\overline{\mathbf{J}} - \underline{\mathbf{J}}\big) = \mathbf{P}_{\mathbf{y}}\big(\overline{\mathbf{J}} - \underline{\mathbf{J}} - \overline{\mathbf{J}} + \underline{\mathbf{J}}\big) = \mathbf{0}$, random variables Q_s and Q_n are statistically independent, and so Q_n/Q_s follows $(\Re/2)F_{\Re,2}$, where F is the F-distribution (Craig et al., 2013, Chapter 9.8). Therefore, the least-squares wavelet spectrogram given by Equation (C.10) that is identical to Equation (6.11) follows the beta distribution with parameters 1 and $\Re/2$, i.e., $s(t_j, \omega_k) \sim \beta_{1,\Re/2}$.

Now define the null hypothesis H_0 and alternative hypothesis H_1 as follows:

$$H_0: \ \frac{Q_n}{Q_s} \geq \frac{\Re}{2} F_{\Re,2,\alpha}, \tag{C.13}$$

$$H_1: \ \frac{Q_n}{Q_s} < \frac{\Re}{2} F_{\Re,2,\alpha}, \tag{C.14}$$

where α is the significance level (usually $\alpha = 0.01$ or $\alpha = 0.05$). Based on the null hypothesis, one may only consider the lower tail end of F because large values of ratio Q_n/Q_s mean that the noise is significantly larger than the signal, and so the signal is undetectable (Pagiatakis, 1999). The stochastic surface at $(1-\alpha)$ confidence level for the spectrogram given by Equation (C.10) is

$$\zeta_{(t_j,\omega_k)} = \left(1 + \frac{\Re}{2} F_{\Re,2,\alpha}\right)^{-1}, \tag{C.15}$$

It can be seen that if $s(t_j, \omega_k) > \zeta_{(t_j,\omega_k)}$, then H_0 is rejected, and spectral peak $s(t_j, \omega_k)$ is statistically significant at $(1-\alpha)$ confidence level. Note that the stochastic surface given by Equation (C.15) depends on the window size, and it does not depend on the frequency when $L_1 = 0$. The stochastic surface is also independent of the time (the location of the window) except for the marginal windows.

Using the relation $F_{\Re,2,\alpha} = F^{-1}_{2,\Re,1-\alpha} = (2/\Re)(\alpha^{-2/\Re} - 1)^{-1}$ (Rao and Mitra, 1972; Pagiatakis, 1999), Equation (C.15) can be written as

$$\zeta_{(t_j,\omega_k)} = \left(1 + \frac{1}{\alpha^{-2/\Re} - 1}\right)^{-1} = 1 - \alpha^{2/\Re}. \tag{C.16}$$

Following similar methodology as discussed above, one can see that the least-squares spectrum given by Equation (3.10) follows the beta distribution with parameter values 1 and $(n-q-2)/2$, and so the critical value at $(1-\alpha)$ confidence level for the least-squares spectrum is $1-\alpha^{2/(n-q-2)}$. In other words, in equations above assume that $R(j,k) = n$, where n is the size of time series $\mathbf{f}$, and $\mathbf{P}_{\mathbf{y}} = \mathbf{P}$.

D Derivation of Estimated Multichannel Coefficients

A similar methodology used as in Wells and Krakiwsky (1971) and Vaníček and Krakiwsky (1986) to obtain Equation (5.8) and to show that it minimizes Equation (5.6). Simplify Equation (5.6) using elementary matrix operations (e.g., Horn and Johnson (2012)) as follows

$$\begin{aligned}\Psi_k(\overline{\mathbf{c}}) &= \Big(\mathbf{f} - \overline{\mathbf{\Phi}}\ \overline{\mathbf{c}}\Big)^{\mathrm{T}}\mathbf{P}\Big(\mathbf{f} - \overline{\mathbf{\Phi}}\ \overline{\mathbf{c}}\Big) + \Big(\mathbf{f_x} - \overline{\mathbf{\Phi}}_{\mathbf{x}}\ \overline{\mathbf{c}}\Big)^{\mathrm{T}}\mathbf{P_x}\Big(\mathbf{f_x} - \overline{\mathbf{\Phi}}_{\mathbf{x}}\ \overline{\mathbf{c}}\Big)\\ &= \Big(\mathbf{f}^{\mathrm{T}} - \overline{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}^{\mathrm{T}}\Big)\mathbf{P}\Big(\mathbf{f} - \overline{\mathbf{\Phi}}\ \overline{\mathbf{c}}\Big) + \Big(\mathbf{f_x}^{\mathrm{T}} - \overline{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\Big)\mathbf{P_x}\Big(\mathbf{f_x} - \overline{\mathbf{\Phi}}_{\mathbf{x}}\ \overline{\mathbf{c}}\Big)\\ &= \mathbf{f}^{\mathrm{T}}\mathbf{P}\mathbf{f} - \mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\ \overline{\mathbf{c}} - \overline{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \overline{\mathbf{c}}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\Big)\ \overline{\mathbf{c}}\\ &+ \mathbf{f_x}^{\mathrm{T}}\mathbf{P_x}\mathbf{f_x} - \mathbf{f_x}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\ \overline{\mathbf{c}} - \overline{\mathbf{c}}^{\mathrm{T}}\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\mathbf{f_x} + \overline{\mathbf{c}}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big)\ \overline{\mathbf{c}}. \end{aligned} \tag{D.1}$$

To minimize Equation (D.1) with respect to $\overline{\mathbf{c}}$, one may calculate the derivative of $\Psi_k(\overline{\mathbf{c}})$ with respect to $\overline{\mathbf{c}}$ and set it equal to zero as follows

$$\begin{aligned}\frac{\partial\Psi_k(\overline{\mathbf{c}})}{\partial\overline{\mathbf{c}}} &= -\mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} - \mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + 2\ \overline{\mathbf{c}}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\Big)\\ &\quad - \mathbf{f_x}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}} - \mathbf{f_x}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}} + 2\ \overline{\mathbf{c}}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big) = 0. \end{aligned} \tag{D.2}$$

Rearranging and simplifying the terms in Equation (D.2), one obtains

$$\frac{\partial\Psi_k(\overline{\mathbf{c}})}{\partial\overline{\mathbf{c}}} = -2\ \Big(\mathbf{f}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + \mathbf{f_x}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big) + 2\ \overline{\mathbf{c}}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + \overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big) = 0. \tag{D.3}$$

Taking the transpose of both sides of Equation (D.3) and solve for $\overline{\mathbf{c}}$ yields Equation (5.8):

$$\hat{\overline{\mathbf{c}}} = \begin{bmatrix}\mathbf{v}\\ \hat{\mathbf{c}}_k\end{bmatrix} = \Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + \overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big)^{-1}\Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\mathbf{f} + \overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\mathbf{f_x}\Big), \tag{D.4}$$

where $\hat{\overline{\mathbf{c}}}$ is used to indicate that $\overline{\mathbf{c}}$, minimizing Equation (5.6), is an estimation. Using the second derivative test in Calculus, one may verify that $\hat{\overline{\mathbf{c}}}$ in fact minimizes Equation (5.6). To see this, one may calculate the second derivative of Equation (D.3) as

$$\frac{\partial}{\partial\overline{\mathbf{c}}}\Big\{\frac{\partial\Psi_k(\overline{\mathbf{c}})}{\partial\overline{\mathbf{c}}}\Big\} = 2\ \Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + \overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big), \tag{D.5}$$

that is positive definite because

$$\mathbf{z}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}} + \overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big)\mathbf{z} = \mathbf{z}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}\Big)\mathbf{z} + \mathbf{z}^{\mathrm{T}}\Big(\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}\Big)\mathbf{z} > 0, \tag{D.6}$$

for any $\mathbf{z} \neq \mathbf{0}$. Note that $\overline{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{P}\overline{\mathbf{\Phi}}$ and $\overline{\mathbf{\Phi}}_{\mathbf{x}}^{\mathrm{T}}\mathbf{P_x}\overline{\mathbf{\Phi}}_{\mathbf{x}}$ are positive definite.

E A Decimation Algorithm Using the Gaussian Filter

Since the time series in the XWT must have the same sampling rates, the GGT trace series described in Section 7.4 must be decimated from 3600 to 360 samples per hour. In this appendix, I will show how one may decimate the GGT trace series for the first satellite track shown in Figure 7.5a.

To decimate the M value of the original GGT trace series and show the importance of considering the covariance matrix of decimated data points, one may use the Gaussian function $g(\ell) = e^{-0.5\ell^2/\sigma^2}$, where $\sigma = 2.5$ and $-9 \leq \ell \leq 9$. In other words, the Gaussian window contains 19 samples (Figure E.1a). Thus, the cutoff frequency is (sampling rate)$/(2\pi \cdot \sigma)$ that is $3600/(2\pi \cdot 2.5) \approx 229$ c/h. The Gaussian function is normalized to define the weights of the filter as $w(\ell) = g(\ell)/s$ $(-9 \leq \ell \leq 9)$, where $s = \sum_{\ell=-9}^{9} g(\ell)$. It can be seen from Figure E.1a that the Gaussian windows overlap each other when translate 10 points over time (red arrows), and so one may obtain the weighted covariances between the decimated data points using the weights shown with solid black diamonds.

The weighted average and standard deviation of the GGT trace values within each window give the value of the decimated data point and its error bar, respectively (Pagiatakis et al., 2007b). More precisely, suppose that n is the number of data points in the original GGT trace series $f(t_i)$ $(1 \leq i \leq n)$, and m is the number of data points in the decimated GGT trace series $h(\tau_j)$ $(1 \leq j \leq m)$. In this example, $n = 701$ and $m = 71$. The values of the decimated series may be calculated as

$$h(\tau_j) = \sum_{\ell=-9}^{9} w(\ell) f(t_{10j+\ell-9}) \tag{E.1}$$

for $2 \leq j \leq 70$ (except for $j = 1$ and $j = 71$ that correspond to the margins), and their corresponding variances as

$$\mathrm{Var}(j) = \eta \sum_{\ell=-9}^{9} w(\ell)\big(f(t_{10j+\ell-9}) - h(\tau_j)\big)^2, \tag{E.2}$$

where $\eta = 19/18$ (West, 1979). From Equations (E.1) and (E.2), one may calculate $h(\tau_1)$ (for $j = 1$) and $h(\tau_{71})$ (for $j = 71$) and their variances by ranging ℓ from 0 to 9 and -9 to 0, respectively. One may also use $\eta = 1/\big(1 - \sum_{\ell=-9}^{9} w^2(\ell)\big)$ that is slightly smaller than 19/18, producing almost the same final results (Galassi et al., 2016). The minimum, maximum, and mean values of the standard deviations (error bars) are 0.006, 0.0257, and 0.0154, respectively. Now it can be seen from Figure E.1a that only the covariance between every two consecutive decimated data points is nonzero when $\sigma = 2.5$. One may also calculate the covariance between each two consecutive decimated data points using

$$\begin{aligned}\mathrm{Cov}(j, j+1) = \eta_c \sum_{\ell=1}^{9} \Big(& w_c(\ell)\big(f(t_{10j+\ell-9}) - h(\tau_j)\big) \\ & \big(f(t_{10j+\ell-9}) - h(\tau_{j+1})\big)\Big),\end{aligned} \tag{E.3}$$

for $1 \leq j \leq 70$, where $w_c(\ell) = w(\ell - 10)$ for $1 \leq \ell \leq 5$, $w_c(\ell) = w(\ell)$ for $6 \leq \ell \leq 9$, and η_c may be calculated as $\eta_c = 1/\big(1 - \sum_{\ell=1}^{9} w_c^2(\ell)\big)$ similar to η (Stuart and Ord, 2010; Galassi et al., 2016). The result of this

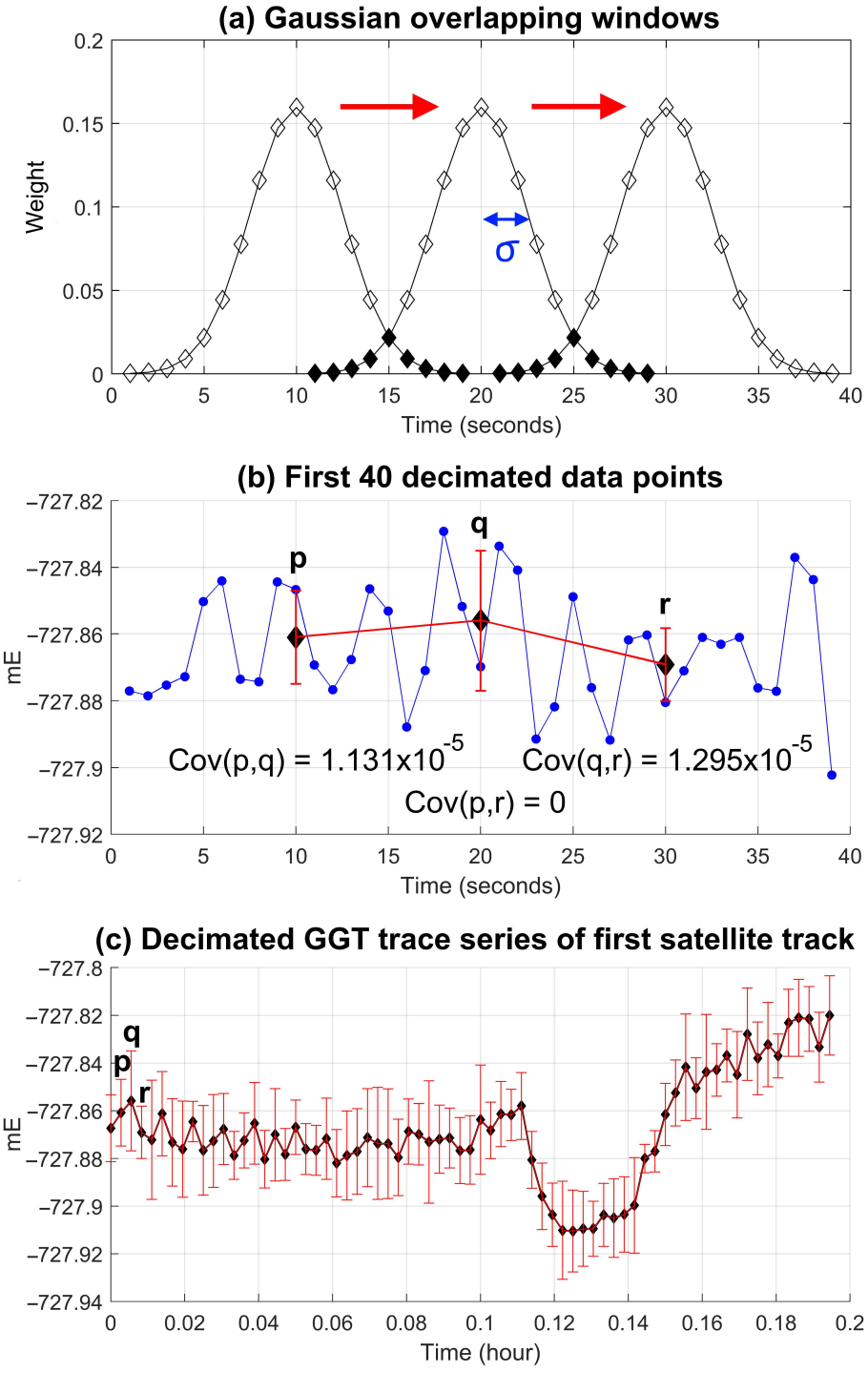

Figure E.1: The decimating process of the GGT trace series for the first satellite track shown in Figure 7.5a using a Gaussian window. (a) The translating Gaussian window, (b) the result of decimation of the first 40 data points of the GGT trace series using the Gaussian weights shown with diamonds in panel a (note that p, q, r are the second, third, and fourth decimated points, respectively), and (c) the decimated GGT trace series and its error bars in red.

decimation method for the first few data points and the entire decimated GGT trace series along with its error bars is illustrated in Figure E.1b and c, respectively (see decimated points p, q, r). The variances and covariances calculated from Equations (E.2) and (E.3) form a covariance matrix (symmetric and tridiagonal) associated with the decimated GGT trace series.