



UNIVERSITAS INDONESIA

**CHANNEL DISTRIBUTION MAPPING USING
PRINCIPAL COMPONENT SPECTRAL ANALYSIS**

UNDERGRADUATE THESIS

**ANDREW JO
0606068026**

**FACULTY OF MATHEMATICS AND NATURAL SCIENCES
DEPARTMENT OF PHYSICS
DEPOK
MAY 2010**



UNIVERSITAS INDONESIA

**CHANNEL DISTRIBUTION MAPPING USING
PRINCIPAL COMPONENT SPECTRAL ANALYSIS**

UNDERGRADUATE THESIS

**Submitted in Partial Fulfillment of the Requirements for the Degree of
Sarjana Sains in Physics**

**ANDREW JO
0606068026**

**FACULTY OF MATHEMATICS AND NATURAL SCIENCES
DEPARTMENT OF PHYSICS
DEPOK
MAY 2010**

HALAMAN PERNYATAAN ORISINALITAS

Skripsi/Tesis/Disertasi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama : Andrew Jo
NPM : 0606068026
Tanda Tangan :
Tanggal : 31 Mei 2010

SHEET OF APPROVAL

This thesis is submitted by

Name : Andrew Jo
Nomor Pokok Mahasiswa : 0606068026
Major : Physics
Thesis Topic : Channel Distribution Mapping Using
Principal Component Spectral Analysis

Has been successfully defended before the Board of Examiners and declared acceptable as a partial fulfillment of the requirements for the Degree of Sarjana Sains in Physics at Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia

BOARD OF EXAMINERS

Thesis Advisor I : Dr. Leonard Lisapaly ()

Thesis Advisor II : Dr. rer. nat. Abdul Haris ()

Examiner I : Dr. Syamsu Rosid ()

Examiner II : Befriko Murdianto, M.Sc ()

PREFACE

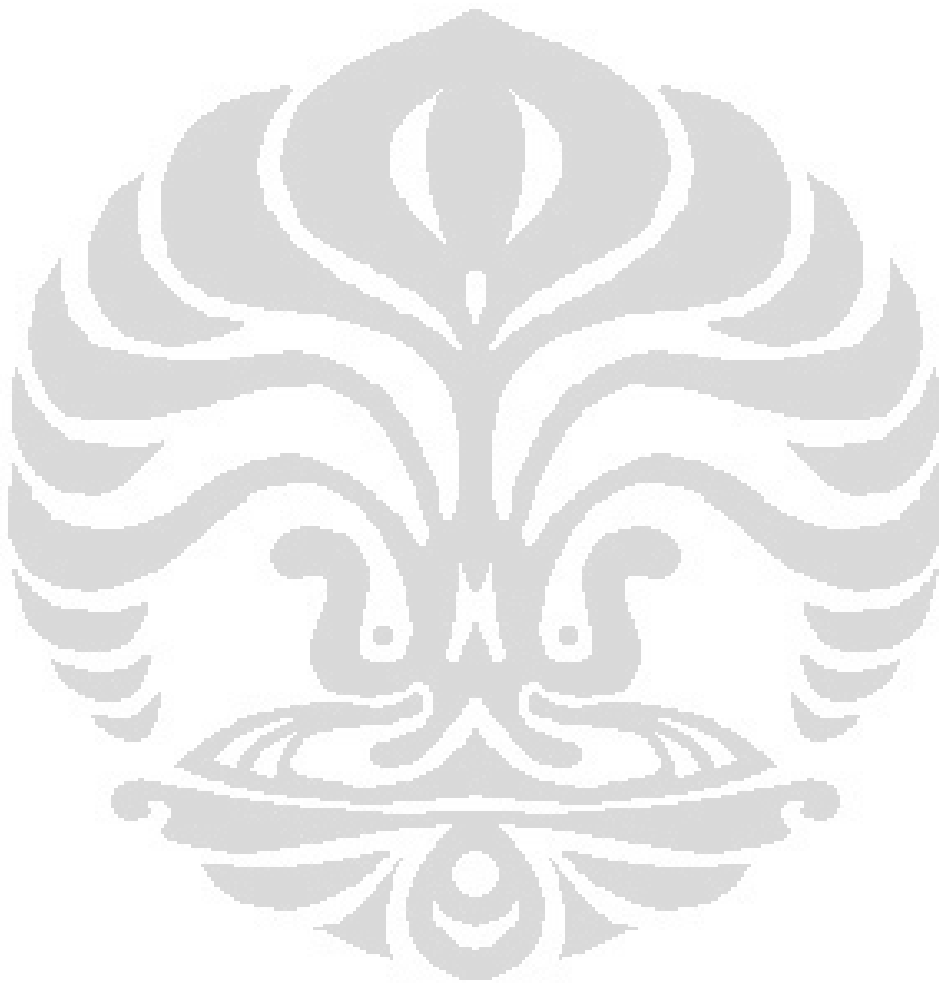
Nowadays, oil and gas exploration might come to a hard time, considering the urge of discovering new oil wells, on the other hand, the earth lacks for the resources or else the farther they are located. As we know, oil and gas are both unrenewable and take up a long period to produce one. Therefore, petroleum geologists and geophysicists are obliged to discover quick and yet efficient ways to overcome this matter. Creative and new ideas to explore the subsurface are surely the vital thing in oil industry. Hence, this thesis work is dedicated to work on a fresh idea to better understand the earth deep down there. I realize that without the assistance of many parties during my study in University of Indonesia, it is impossible for me to finish this work; therefore I would like to express my gratitude for:

- (1) Dr. Leonard Lisapaly as my thesis advisor, for the discussion, valuable guidance, and creative insights. He is a very resourceful geophysicist and I am very grateful could have an opportunity to work with him.
- (2) Dr. Abdul Haris, my thesis co-advisor for the tremendous support, encouragement, and enriching discussion.
- (3) Dr. Syamsu Rosid and Befriko Murdianto, M.Sc, as the examiners, who spared some time for brainstorming session and giving input to improve the thesis.
- (4) Friends in Department of Physics, University of Indonesia, and particularly in Geophysics Program and also Student Chapter of AAPG members, I thank you all for the support and help during my thesis work, and run the student chapter together with me,
- (5) Staffs in Department of Physics, Bu Ratna and Pak Mardi, thanks for the help regarding administrative stuff.

(6) Lastly, I would like to thanks my family for the motivation and support during my study.

Jakarta, 31th May 2010,

Author



**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Andrew Jo

NPM : 0606068026

Program Studi : Geofisika

Departemen : Fisika

Fakultas : MIPA

Jenis karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :

**CHANNEL DISTRIBUTION MAPPING USING
PRINCIPAL COMPONENT SPECTRAL ANALYSIS**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 31 Mei 2010

Yang menyatakan

Andrew Jo

ABSTRACT

Spectral decomposition is one of the method in advance seismic analyses to map geologic body. Usually, an interpreter will generate a lot of spectral components which are usually scanned every 1 Hz to look for anomalies. Those spectral components are highly correlated and redundant, same feature can occur in several components with different quality. To reduce the multiplicity of spectral data and enhance the most energetic trends inside the data, principal component analysis (PCA) is applied to the spectral components. PCA is a multivariate statistics method that works by rotating the axes to align with natural extension of the swarms of points and ordering the principal components by decreasing variance.

In the first experiment to wedge model, PCA is very excellent in mapping all wedge body continuously regardless of the thickness. All significant features are plotted in the first principal component which accounts for more than 85% data variance. We also verify the algorithm to Stratton seismic data, target at channel which provide a comparable image from spectral decomposition. But we notice that some parts of the channel are slightly poorly imaged due to the contamination of noise and low significance of the data. PCA is robust when the trend in the data coexists in other spectral components and it doesn't smeared by noise. By using blending technique, we sum up the first three principal component bands and shows that the channel is better imaged. Less significant features that show up in principal component bands two and three are all mapped to the composite image.

Keywords: *spectral decomposition, spectral component, principal component analysis, variance*

ABSTRAK

Dekomposisi spectral merupakan salah satu metoda analisis seismik lanjutan untuk memetakan benda geologi. Biasanya seorang *interpreter* akan menghasilkan sangat banyak komponen spektral yang di-*scan* setiap 1 Hz untuk mencari anomali tersebut. Komponen spektral tersebut saling berkorelasi dan berulang, suatu fitur yang sama dapat muncul di beberapa komponen tetapi dengan kualitas berbeda. Untuk mengurangi perulangan data spektral dan meningkatkan trend di dalam data, *principal component analysis* (PCA) diaplikasikan kepada komponen spektral. PCA adalah metode statistik multivariable yang bekerja dengan cara merotasi sumbu agar sejajar dengan arah penyebaran data dan mengurutkan *principal component* berdasarkan besar nilai varians.

Dalam eksperimen pertama kepada model baji, PCA dapat memetakan seluruh bagian baji dengan menerus tidak peduli dengan ketebalannya. Semua fitur yang signifikan diplot pada *principal component* pertama yang mempunyai lebih dari 85% varians data. Kami juga memverifikasi algoritma ini pada data seismik Stratton, dengan target channel dimana citra yang diperoleh sebanding dengan hasil dekomposisi spektral. Tetapi kami juga memperhatikan ada beberapa bagian dari channel yang dicitrakan sedikit kurang bagus akibat dari kontaminasi noise dan juga kurang signifikannya data tersebut. PCA akan berfungsi dengan baik jika trend dalam data koeksis dalam komponen spektral yang lain dan tidak didominasi oleh noise. Dengan menggunakan teknik *blending*, kami menjumlahkan tiga *principal component band* pertama dan menunjukkan bahwa channel dapat dipetakan dengan lebih baik. Fitur-fitur yang kurang signifikan yang muncul pada *principal component band* dua dan tiga semua dipetakan dalam citra komposit.

Kata kunci: *dekomposisi spektral, komponen spektral, principal component analysis, varians*

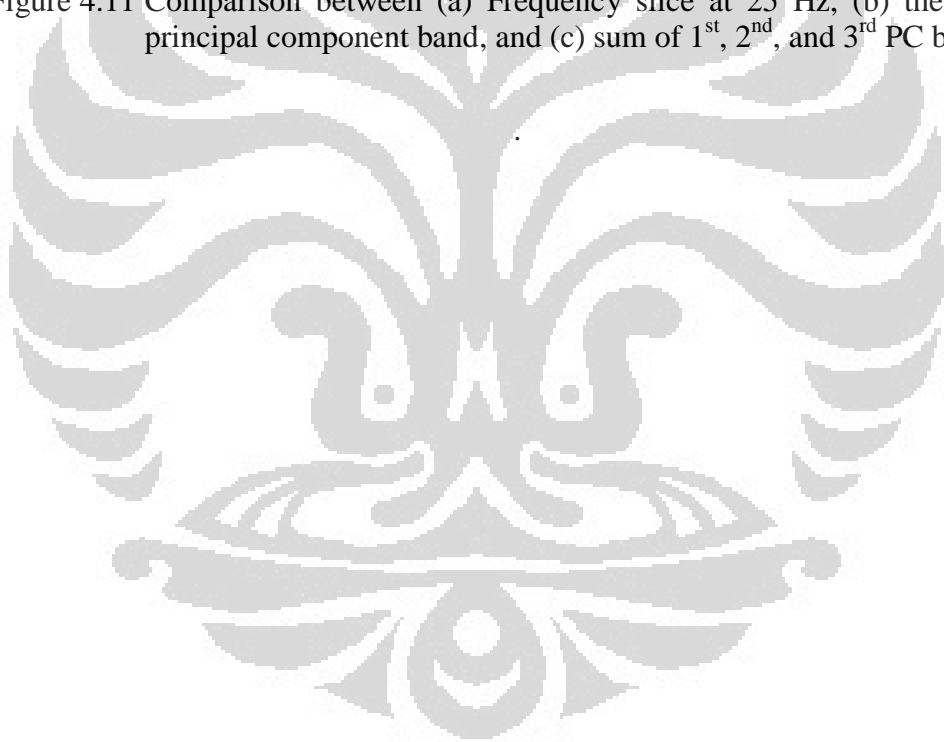
CONTENTS

	Pages
COVER PAGE.....	i
HALAMAN PERNYATAAN ORISINALITAS.....	ii
SHEET OF APPROVAL.....	iii
PREFACE	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	vi
ABSTRACT.....	vii
CONTENTS.....	ix
LIST OF FIGURES.....	x
LIST OF APPENDICES.....	xii
CHAPTER 1. INTRODUCTION	1
1.1. Background.....	1
1.2. Problem Statement.....	2
1.3. Objective.....	2
1.4. Methodology.....	2
1.5. Writing Systematic.....	2
CHAPTER 2. THEORETICAL BACKGROUND.....	3
2.1. Seismic Attribute.....	3
2.2. Seismic Vertical Resolution.....	4
2.3. Spectral Decomposition.....	6
2.4. Fourier Series.....	9
2.5. Fourier Transform.....	10
2.6. Fast Fourier Transform.....	11
2.7. Eigenvalue and Eigenvector.....	12
2.8. Covariance Matrix.....	12
2.9. Correlation Matrix.....	14
2.10. Principal Component Analysis.....	16
CHAPTER 3. METHODOLOGY.....	20
3.1. Processing of Synthetic Data.....	20
3.2. Processing of Real Seismic Data.....	22
CHAPTER 4. RESULT AND ANALYSIS.....	26
4.1. Synthetic Data.....	26
4.2. Three Dimension (3D) Seismic Data.....	30
CHAPTER 5. CONCLUSION.....	38
References	39

LIST OF FIGURES

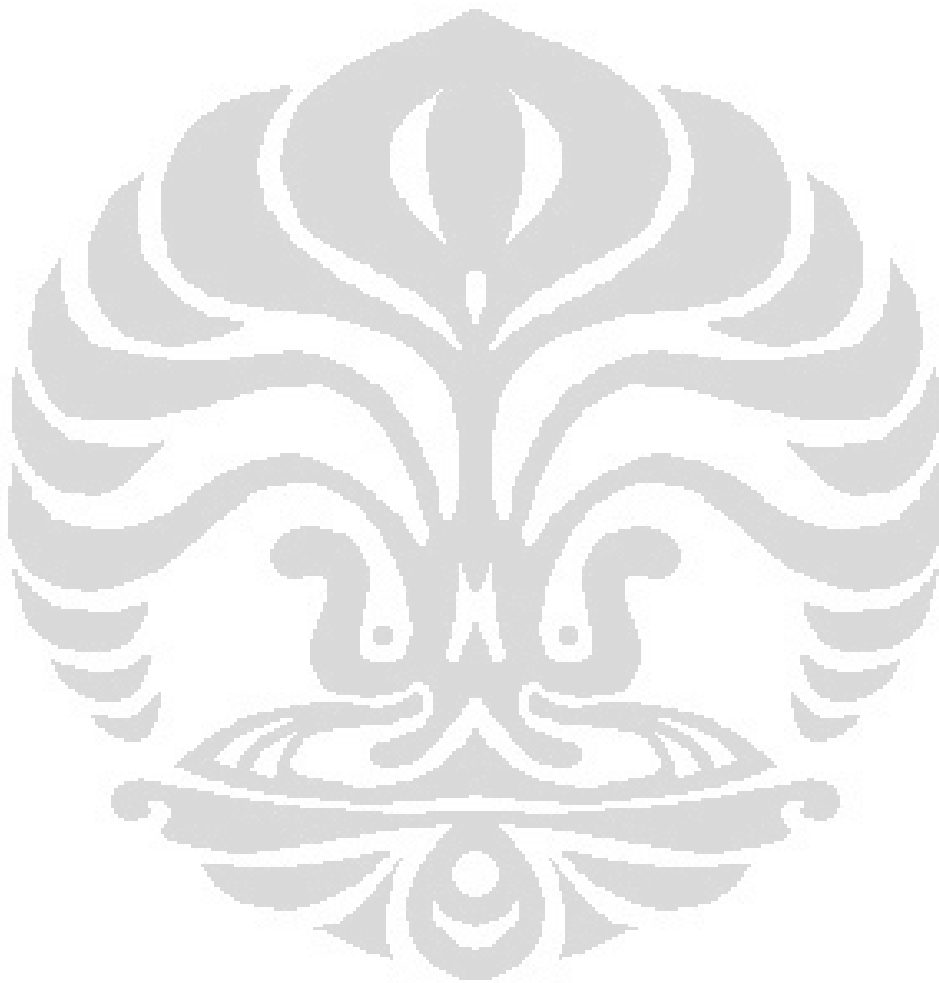
	Pages
Figure 2.1 Seismic attribute classification, derived from time, amplitude, frequency, and attenuation which are further group into pre-stack and post-stack categories (Brown, 2001)	4
Figure 2.2 Seismic response of beds with different thickness. Note that at thickness less than $\lambda/8$, the bed is not seperable in time (Widess, 1973).....	5
Figure 2.3 Spectral interference from thin bed reflection (Partyka <i>et al</i> , 1999)...	6
Figure 2.4 Illustration showing that higher frequency amplitude map will highlight thinner reservoir on amplitude map while lower frequency amplitude map will highlight the thickest part (Laughlin, <i>et al</i> , 2002).....	6
Figure 2.5 Spectral decomposition of seismic trace over (a) long window and (b) short window (Partyka <i>et al</i> , 1999).....	7
Figure 2.6 Wedge model. (a) Reflectivity; (b) filtered reflectivity; (c) spectral amplitudes (Partyka <i>et al</i> , 1999).....	8
Figure 2.7 Pleistocene-age delta map in GOM (a) 16 Hz (b) 26 Hz frequency slice, and (c) conventional full bandwidth energy extraction (Partyka, <i>et al</i> , 1999)	8
Figure 2.8 Fourier series approximation of a square signal (Wikipedia).....	10
Figure 2.9 Cross-correlation between two images (from MATLAB Help).....	15
Figure 2.10 PCA transformation of a variable y_1 and y_2 (Rencher, 2002)	17
Figure 2.11 Cartoon showing frequency slice of seismic data. Numbers written in the box denote trace number. PC band 1 is plot in (d).....	18
Figure 2.12 Principal component of data consisting of three components from the previous channel model (Figure 9), with red dots represent three spectral components from each trace.....	18
Figure 3.1 (a). Reflectivity model of wedge, (b). Synthethic seismogram from wedge model at crossline 60.....	21
Figure 3.2 Three dimension image of wedge body.....	21
Figure 3.3 Seismic slice at crossline 39, inline 39, and time slice (47) ms, notice the channel pattern (blue).	22
Figure 3.4 Stratton seismic data at crossline 75	23
Figure 3.5 Cartoon showing workflow of Principal Component Analysis.....	25
Figure 4.1 Time slice of 3D synthetic seismogram wedge model every 0.2 ms, starting at 2 ms.....	27
Figure 4.2 Amplitude spectrum of traces in crossline 50	27
Figure 4.3 Frequency slice of wedge model showing bright and dark pattern at different frequencies: (a) 1 Hz, (b) 2 Hz, (c) 3 Hz, (d) 4 Hz, (e) 7 Hz, (f) 9 Hz, (g) 11 Hz, (h) 15 Hz, (i) 17 Hz, (j) 20 Hz, (k) 23 Hz, (l) 32 Hz, (m) 40 Hz, (n) 48 Hz, and (o) 55 Hz	28
Figure 4.4 Spectrum projected onto (a) first, (b) second, (c) third, (d) fourth, (e) 17 th , and (f) 163 rd principal component.....	29

Figure 4.5	Scree graph of variance percentage explained by principal components.....	29
Figure 4.6	Time slice of Stratton seismic cube at (a) 0.06 ms, (b) 0.062 ms, (c) 0.064 ms, (d) 0.088 ms, (e) 0.09 ms, (f) 0.092 ms, (g) 0.094 ms (h) 0.1 ms, (i) 0.102 ms, (j) 0.104 ms, (k) 0.106 ms, (l) 0.108 ms, (m) 0.11 ms, (n) 0.112 ms, (o) 0.114 ms	31
Figure 4.7	Stratton data frequency slice at (a) 0 Hz, (b) 5 Hz, (c) 10 Hz, (d) 15 Hz, (e) 20 Hz, (f) 25 Hz, (g) 30 Hz, (h) 35 Hz, (i) 40 Hz, (j) 45 Hz, (k) 50 Hz, (l) 55 Hz, (m) 60 Hz, (n) 65 Hz, (o) 70 Hz, (p) 75 Hz, (q) 80 Hz, (r) 85 Hz, (s) 90 Hz, and (t) 95 Hz.	32
Figure 4.8	Covariance matrix of (a) 91 original spectral components, (b) 91 principal components displayed in form of colored image. Hot color represent high correlation.....	34
Figure 4.9	Graph showing variance explained by PC. The first principal component accounts for more than 97% variance.....	34
Figure 4.10	Spectral components projected to (a) 1 st , (b) 2 nd , (c) 3 rd , (d) 4 th , (e) 5 th , and (f) 16 th PC bands.	35
Figure 4.11	Comparison between (a) Frequency slice at 25 Hz, (b) the first principal component band, and (c) sum of 1 st , 2 nd , and 3 rd PC bands	37



LIST OF APPENDICES

	Pages
A. Wedge Model	40
B. Seismic 3D Data	42



CHAPTER 1

INTRODUCTION

1.1 Background

By definition, spectral decomposition is all method that produces continuous time-frequency analysis, so it is possible to have a frequency spectrum for every time sample from the seismic trace. By transforming the time domain into frequency domain via Fourier Transform, the amplitude spectra delineate temporal bed thickness while the phase spectra delineate geological discontinuities. This technique is an effective tool to delineate stratigraphic feature such as channel sands and structural settings involving complex fault systems in 3D surveys (Partyka, 1999).

Because spectral decomposition technique typically generates a large number of outputs compared to the input, it is a challenge for the interpreter to understand the meaning of these data. Usually, an interpreter may generate 80 or more spectral amplitude and phase components from the entire usable seismic bandwidth at 1 Hz intervals (Guo *et al*, 2009).

To reduce the multiplicity of spectral data and enhance the most energetic trends inside the data, principal component analysis (PCA) is applied to the spectral component. The most significant spectral components are sorted by the corresponding eigenvalue. Unlike the input spectral magnitudes and phase components, PCA does not indicate bed thickness. By projecting three largest principal components using three primary colors of red, green, and blue, the image can represent more 80% of spectral variance of the data. Principal component analysis generates less output volume and also rejects spectral components that are identified as noise.

1.2 Problem Statement

There are a great number of spectral components generated in spectral decomposition, but not all of them represent the signal from specific geologic bodies which are expected. So, what interpreters usually do is tune the frequency every 1 Hz, which means an interpreter should analyze more than 90 images and look for anomaly that may occur.

1.3 Objective

Because there are a lot of spectral components are generated in spectral decomposition, the purpose of this thesis work is to reduce spectral component data multiplicity using principal component analysis. Using this technique, we hope that trends in our data can be captured and plotted in the first few principal components, regardless of the thickness of the geologic body of interest.

1.4 Methodology

The pursuit of better imaging of channel system is investigated using principal component spectral analysis technique.

1.5 Writing systematic

The first Chapter contains introduction to this thesis work, motivation, and objectives. The theoretical frameworks are explained in Chapter 2. Methodologies and workflow used in this work are discussed in Chapter 3, while Chapter 4 contains analysis of the results. Conclusions and recommendations are given in Chapter 5.

CHAPTER 2

THEORETICAL FRAMEWORK

2.1 Seismic Attribute

A seismic attribute is a quantitative measure of a seismic characteristic of interest, and there are more than 50 different attributes that can be derived from seismic data to aid interpretation of geologic structure, stratigraphy, and rock/fluid properties. Good seismic attribute is either sensitive to geologic feature or reservoir property, or enable us to map the structural or depositional environment that is useful to infer some feature or properties of interest (Chopra *et al*, 2005).

There were many attempts to classify seismic attributes into families to better understand the application of them. Taner et al (1994) grouped attributes into geometrical and physical types. Geometrical attributes are used to enhance geometrical characteristics of seismic data, while physical attributes are for defining physical parameters of subsurface.

In a 2001 *Geophysics* journal, Brown classify attributes base on the information we have: time, amplitude, frequency, and attenuation. Each of them is further divided into pre-stack and post-stack categories. Post-stack attributes are then grouped into horizon and time window based (Figure 2.1). Time attributes are for structural definition and amplitude for reservoir and stratigraphic properties.

According to Liner *et al*, 2004, attributes are grouped into general and specific categories. Attributes that have strong physical and mathematical basis, which give the same behavior to basins around the world, are grouped into general categories, for example: horizon peak amplitude, time structure, complex amplitude and frequency, generalized Hilbert attributes, horizon dip and azimuth, illumination, edge-preserved smoothing, edge detection, AVO,

coherence, and spectral decomposition. In the other hand, special attributes are not universally valid, but only in special cases. It depends on the statistical well correlation in a field so the methods may not be useful for another field. Special attributes attract less attention since geoscientists need new fundamental information about the subsurface.

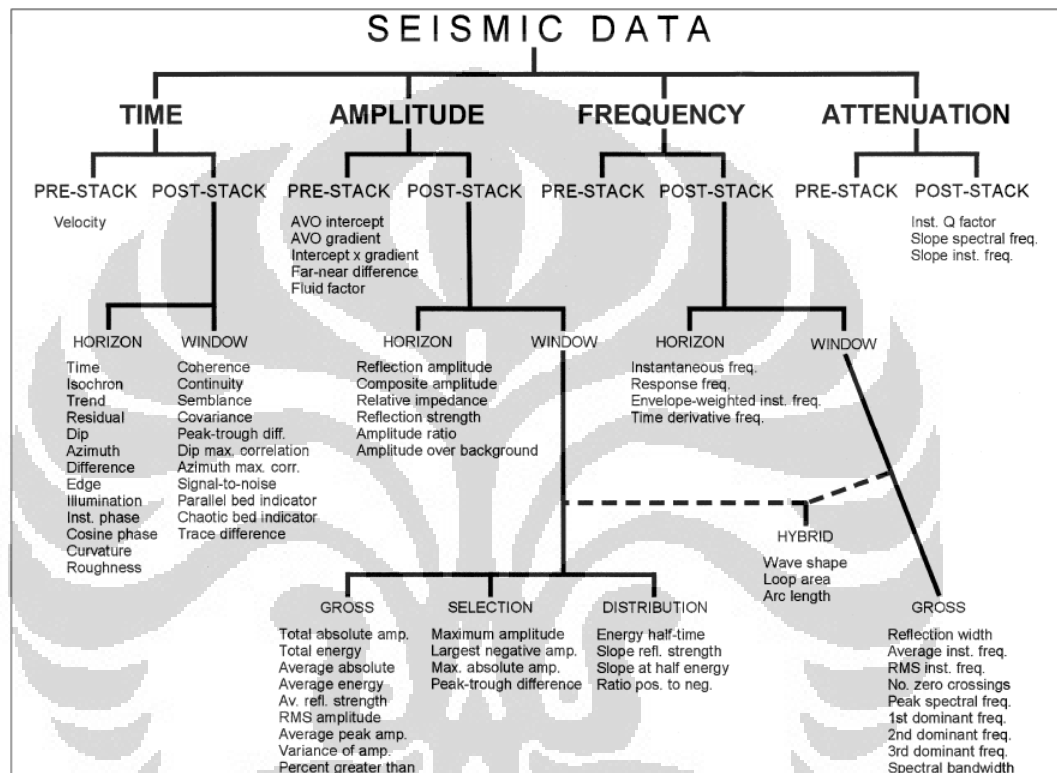


Figure 2.1. Seismic attribute classification, derived from time, amplitude, frequency, and attenuation which are further group into pre-stack and post-stack categories (Brown, 2001).

2.2 Seismic Vertical Resolution

Since seismic wave that travel through the earth is band-limited, experience attenuation and also absorption, it has limitation in imaging geological objects that are too subtle. When the wavelet interfere, tuning effect will occur, which is a phenomenon of constructive or destructive interference of waves from closely spaced events or reflections (Schlumberger glossary). It defines $\lambda/4$ as the tuning thickness, i.e. thickness limit when it is not distinguishable in time.

Rayleigh limits the resolution of two superpositioned wavelet when images are separated by the peak-to-trough time interval while Ricker's limit occurs when the composite waveform has a curvature of zero at its central maximum (flat spot). In a classic paper about thin bed by Widess (1973), it is defined that seismic resolving power is $\lambda/8$ or one eighth of seismic wavelength. This criterion is based on the calculation that the complex waveform from thin bed reflection will approximate the time derivative of incident wavelet. The complex waveform across a thin bed approach the time derivative of incident wavelet as the bed reaches zero thickness. He stated that to define thin bed, it involves the concept of resolving power. When a bed is thick enough, the reflection from the top and base can be distinguished in time. But when the thickness is diminished, the recorded reflection will be a composite of top and base reflections, i.e. less data of each interface separately, but in the form of combination of the two reflections. According to Widess, it continues until the bed thickness reach $\lambda/8$, and when the bed gets thinner, there is no information of each separate interface anymore (Figure 2.2). He also noted that resolving power does not depend only on dominant frequency, but also signal-to-noise ratio.

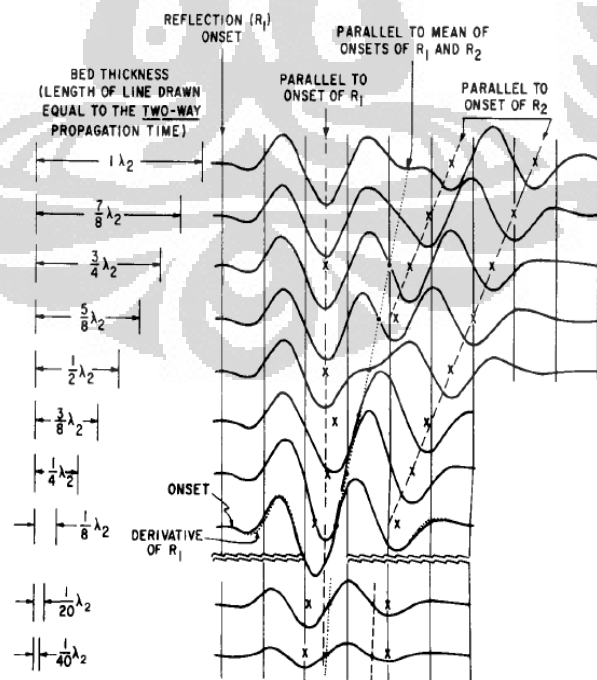


Figure 2.2. Seismic response of beds with different thickness. Note that at thickness less than $\lambda/8$, the bed is not separable in time (Widess, 1973)

2.3 Spectral Decomposition

Spectral decomposition has some applications in interpreting seismic data, such as mapping temporal bed thickness and geologic discontinuities. By transforming seismic data from time domain to frequency domain, the amplitude spectra can unveil temporal bed thickness variability while the phase spectra indicate geologic discontinuities (Partyka *et al*, 1999).

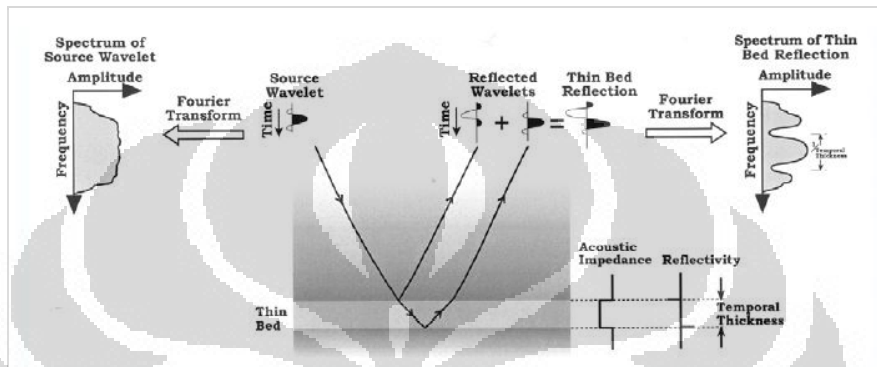


Figure 2.3. Spectral interference from thin bed reflection (Partyka *et al*, 1999)

The idea behind spectral decomposition is the unique characteristic in the frequency domain when the signal is reflected from a thin bed. Notch pattern in amplitude spectra directly indicate temporal bed thickness (Figure 2.3). Bed temporal thickness can be approximated by $1/d$, where d is the distance between notches. Using spectral decomposition, continuous bed with thickness variation will be mapped onto different frequency amplitude map (Figure 2.4).

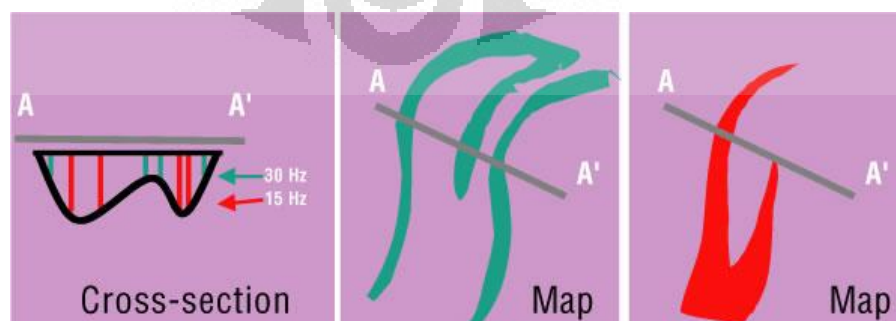


Figure 2.4. Illustration showing that higher frequency amplitude map will highlight thinner reservoir on amplitude map while lower frequency amplitude map will highlight the thickest part (Laughlin, *et al*, 2002).

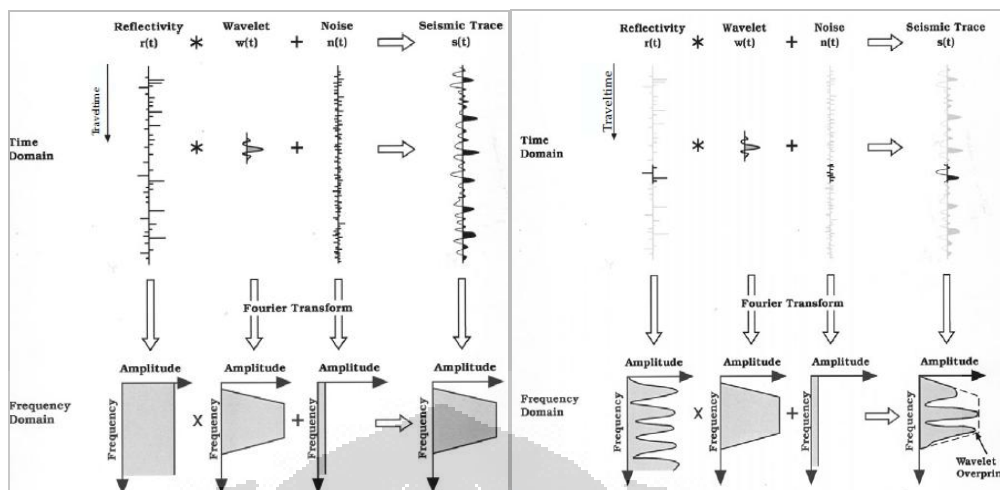


Figure 2.5. Spectral decomposition of seismic trace over (a) long window and (b) short window (Partyka *et al*, 1999).

Transformation from a long trace will give an amplitude spectrum that approximates the wavelet spectrum, while, over a short window, the spectrum contains information about wavelet overprint and also interferences from geologic layers over the selected window, exhibited in periodic notch pattern (Figure 2.5).

Using a wedge model, Partyka *et al*, show the relation between spectral decomposition and thin bed tuning. In the example, the temporal thickness of thin bed determines the period of notches in the amplitude spectrum with respect to frequency (Figure 2.6). From Figure 2.6.c, we can see that period of notch pattern is larger for thinner bed, and vice versa. Other way to look at the picture is that if we select/slice a certain frequency only, i.e. 50 Hz, we can image thin beds that have temporal thickness of 20 ms and 40 ms.

In Gulf of Mexico case study by Partyka, *et al* (1999), he showed that tuning cube frequency slice reveal the Pleistocene-age delta better than full bandwidth energy map. Note the different features that are highlighted by each frequency slice (Figure 2.7).

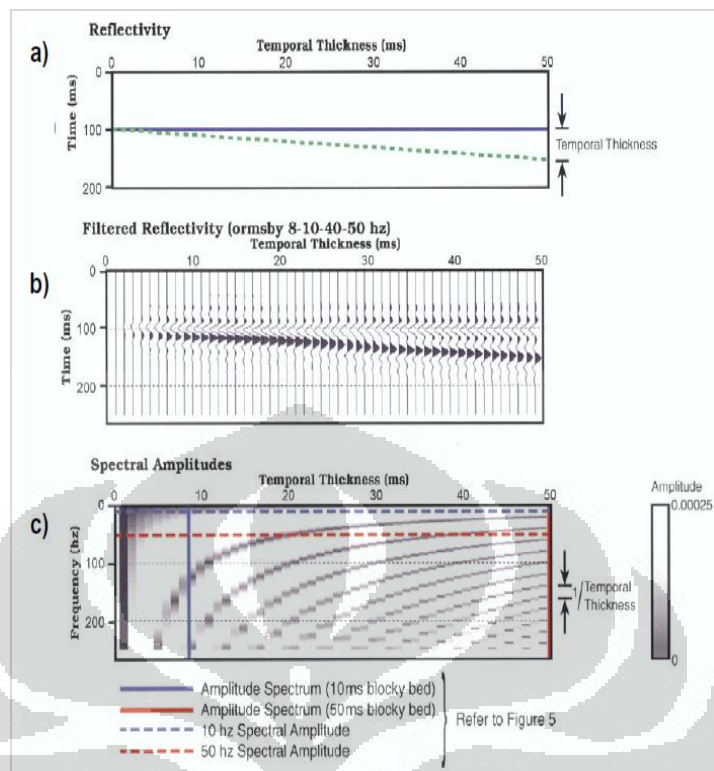


Figure 2.6. Wedge model. (a) Reflectivity; (b) filtered reflectivity; (c) spectral amplitudes (Partyka *et al.*, 1999).

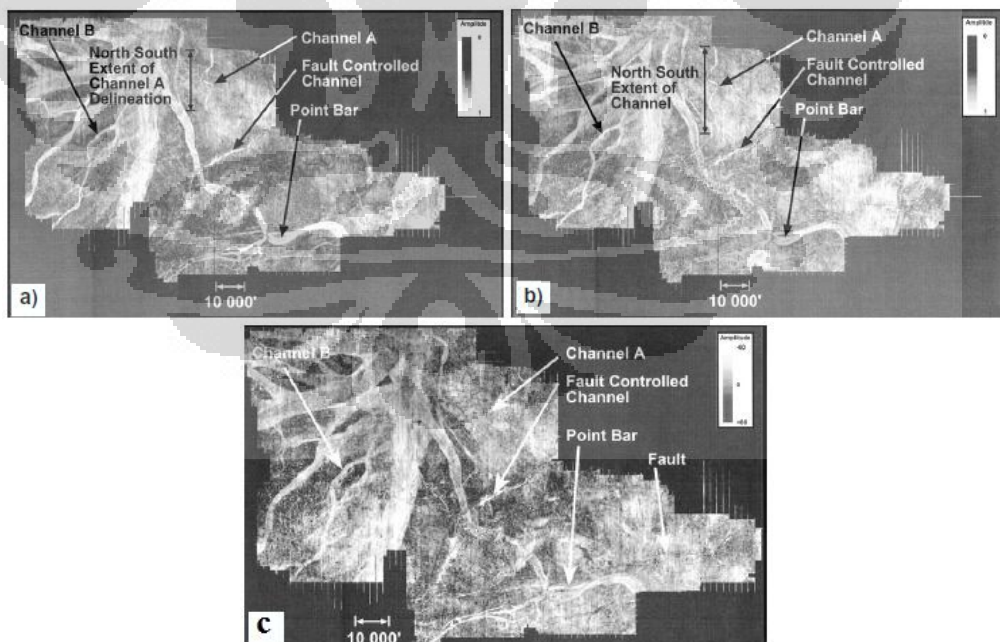


Figure 2.7. Pleistocene-age delta map in GOM (a) 16 Hz (b) 26 Hz frequency slice, and (c) conventional full bandwidth energy extraction (Partyka, *et al.*, 1999).

There are several algorithms that are used to decompose a time series into spectral components. Conventional DFT only produce amplitude spectral with no time information. Because seismic trace is not stationary and the spectral component varies significantly with time, non-standard method of spectral decomposition is required. Those methods include: short-time fast Fourier transform, continuous wavelet transform, and also the more sophisticated one, matching pursuit decomposition. These methods give better resolution in time and frequency domain but they are beyond the scope of this thesis work.

2.4 Fourier Series

According to Fourier theorem, a random periodic waveform can be represented by the sum of sines and cosines (Figure 2.8). Fourier series has application in electrical engineering, vibration analysis, acoustics, optics, signal processing, image processing, quantum mechanics, etc. Fourier series of a periodic 2π function $f(x)$ that is integrable on $[-\pi, \pi]$ is given by:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) \quad (2.1)$$

where: $a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx,$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx,$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx$$

It is possible to express Fourier series in exponential form using Euler's formula:

$$e^{inx} = \cos nx + i \sin nx \quad (2.2)$$

where i is the imaginary unit, to give a more compact formula:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx} \quad (2.3)$$

The Fourier coefficient is given by:

$$c_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx \quad (2.4)$$

while the Fourier coefficients a_n , b_n , c_n are related by:

$$a_n = c_n + c_{-n} \text{ for } n=0, 1, 2, \dots$$

$$b_n = i(c_n - c_{-n}) \text{ for } n=1, 2, \dots$$

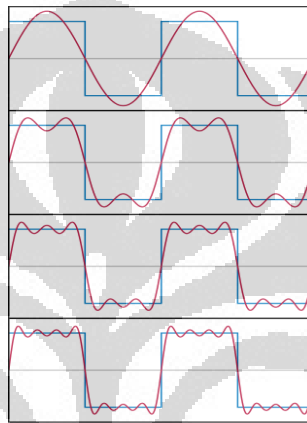


Figure 2.8. Fourier series approximation of a square signal (Wikipedia)

2.5 Fourier Transform

Fourier transform is one of the most important tools in signal processing until today. The concept of Fourier transformation is based on the Fourier series, which represents any functions with an infinite sum of cosine and sine waves. It transforms one complex valued function to another. The definition of Fourier transform is:

$$F(\xi) = \int_{-\infty}^{\infty} f(t) e^{-i2\pi t \xi} dt, \text{ for every real number of } \xi \quad (2.5)$$

In this case, t represents time (second) and ξ represents frequency (Hz) but it is also common to do transformation from time domain into angular frequency domain (ω). The inverse process is transformation from frequency domain into time domain, which formulated as:

$$f(t) = \int_{-\infty}^{\infty} F(\xi) e^{i2\pi t \xi} dt, \text{ for every real number of } t \quad (2.6)$$

An important property of Fourier transforms is their complementary nature, which can be stated as the following:

- a. If $F(\xi)$ is the Fourier transform of $f(t)$, then $f(t)$ is also the Fourier transform of $F(\xi)$
- b. If $f(t)$ is the Fourier transform of $F(\xi)$, then $F(\xi)$ is also the Fourier transform of $f(t)$.

Thus, Fourier transform can be thought as a measure of how much a certain frequency is present in a random waveform. Like Fourier series, it sums up or integrates the multiplication of the original waveform with a periodic wave with certain frequency and calculates how big the value is. Then we move to the next frequency, redo the same process and finally, we get a graph of frequencies versus constants representing how dominance they are in the waveform.

2.6 Fast Fourier Transform (FFT)

FFT is an algorithm to compute Discrete Fourier Transform (DFT) and its inverse. DFT decomposes a sequence of value into different frequencies. Let x_0, x_1, \dots, x_{N-1} are complex numbers, the definition of DFT is:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}} \quad (2.7)$$

where $k= 0, \dots, N-1$. FFT give the same result as by using the definition of DFT, but it performs much faster because instead of computing $O(N^2)$ arithmetical operation, it only needs to compute $O(N \log N)$ operations.

The most widely used algorithm of FFT is based on Cooley-Tukey FFT algorithm which breaks down the DFT into smaller DFTs and combines with other algorithm like Raders's and Bluestein's algorithm to handle large prime factors, and also prime factor algorithm to separate out relatively prime factors more efficiently.

The radix-2 decimation in time (DIT) is the simplest and common algorithm for Cooley-Tukey FFT. It first separates even indexed inputs x_m (x_0, x_2, \dots, x_{N-2}) and odd indexed inputs x_{m+1} (x_1, x_3, \dots, x_{N-1}), then perform DFT on each of them and later combines to form a complete sequence:

$$X_k = \sum_{m=0}^{N/2-1} x_{2m} e^{-i2\pi k \frac{2m}{N}} + \sum_{m=0}^{N/2-1} x_{2m+1} e^{-i2\pi k \frac{2m+1}{N}} \quad (2.8)$$

2.7 Eigenvalue and Eigenvector

The eigenvector of a matrix is defined as a vector that changes in magnitude but not in the direction if operated on a matrix. If the vector flipped to the opposite direction, it is still called the eigenvector of that matrix. A scalar which multiplied by the eigenvector and gives the same matrix as transformed eigenvector is called the eigenvalue of the corresponding eigenvector.

Given a square matrix \mathbf{C} , the eigenvalue λ and eigenvector \mathbf{v} can be found by solving:

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v} \quad (2.9)$$

Rearranging the formula above, we get $(\mathbf{C} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$, which imply that $\det(\mathbf{C} - \lambda\mathbf{I}) = 0$. The determinant can be expanded into a polynomial in term of λ . It is called characteristic polynomial of \mathbf{C} and to determine eigenvalue, one should find the roots of the polynomial.

This concept is very useful in principal component analysis because the largest eigenvalue corresponds with the most significant component (eigenvector) of the data.

2.8 Covariance Matrix

Covariance is statistical measurement to analyze relationship between several dimensions, i.e. how much the dimensions vary from the mean with respect to each other. If you do covariance calculation of a dimension with itself, you get the variance. The formula of covariance is similar to variance:

$$S_{xy} = \text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} = E[(X_i - \mu_x)(Y_i - \mu_y)] \quad (2.10)$$

where μ_x is $E(X)$ or expectation value of X and μ_y is $E(Y)$.

The result from the calculation above tells us something about the relation of dimensions X and Y . The exact value is not as important as its sign. If the covariance is positive, we can conclude that both dimensions increase together, and if the covariance is negative, while one dimension increases, the other decreases.

Covariance is always a measurement of more than one dimensions. For n dimensions data, we can calculate $\frac{n!}{(n-2)!*2}$ different covariance value. A

useful way to get all the possible covariance values is to calculate them and put them in a matrix. Covariance matrix of a set of data with n dimensions is:

$$C^{n \times n} = (c_{j,k}, c_{j,k} = \text{cov}(\text{dim}_j, \text{dim}_k)) \quad (2.11)$$

where $C_{j,k}$ is the j th, k th component of covariance matrix; dim_n is n th dimension.

Another useful expression of covariance matrix is:

$$C = \begin{pmatrix} c_{j,k} & \dots & c_{j,n} \\ \vdots & \ddots & \vdots \\ c_{n,k} & \dots & c_{n,n} \end{pmatrix}$$

Because covariance is dependent on the scale of X and Y , it is difficult to compare covariance of different pairs of variables. For example, if we change the scale from meter to kilometer, the covariance will change.

Suppose that we have a variable Z which is a transformation of Y and related by the following formula:

$$\mathbf{Z} = \mathbf{A}^T * \mathbf{Y} \quad (2.12)$$

where P is an orthonormal rotation matrix. Now we wish to find the covariance matrix of Z , S_z in term of covariance matrix of Y , S_y , we derive the equation by algebra substitution as below:

$$\begin{aligned} \mathbf{S}_z &= \mathbf{E}[\mathbf{Z}\mathbf{Z}^T] \\ &= \mathbf{E}[(\mathbf{A}^T\mathbf{Y})(\mathbf{A}^T\mathbf{Y})^T] \\ &= \mathbf{E}[(\mathbf{A}^T\mathbf{Y})(\mathbf{Y}^T\mathbf{A})] \\ &= \mathbf{A}^T\mathbf{E}[\mathbf{Y}\mathbf{Y}^T]\mathbf{A} \\ &= \mathbf{A}^T\mathbf{S}_y\mathbf{A} \end{aligned} \quad (2.13)$$

Now, we have $\mathbf{A}\mathbf{S}_z = \mathbf{S}_y\mathbf{A}$. Rewrite \mathbf{A} as $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$, \mathbf{S}_z as $\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$, we have $[\lambda_1 a_1, \dots, \lambda_n a_n] = [\mathbf{S}_y a_1, \dots, \mathbf{S}_y a_n]$, where λ_n is the eigenvalue and a_n is the eigenvector of covariance matrix \mathbf{S}_y (Further discussion at section 2.10).

2.9 Correlation Matrix

In signal analysis, correlation or cross-correlation is a standard method to measure the similarity of two waveforms. It finds many applications in the field of pattern recognition, single particle analysis, electron tomography averaging, cryptanalysis, and neurophysiology. Mathematically, cross-correlation is defined as:

$$(f * g)(t) = \int_{-\infty}^{\infty} f^*(\tau)g(t + \tau)d\tau \quad (2.14)$$

where $*$ here denotes cross-correlation, f and g are continuous functions, and f^* denotes complex conjugates of f . For discrete function, cross-correlation is defined as:

$$(f * g)(n) = \sum_{m=-\infty}^{\infty} f^*[m]g[n + m] \quad (2.15)$$

Cross-correlation of two functions is similar to convolution in the case of shifting the series and multiplies them, but without reversing the series. Consider two functions f and g , which differ by two shifts along the x-axis.

One could determine how much shifting is needed to make f identical to g , by sliding one of the function then multiply every components and sums them up. If the functions match, the correlation value will be maximized. In probability theory and statistics, cross-correlation usually refers to covariance $\text{cov}(X, Y)$ between two random vector X and Y .

In multivariate statistics, correlation and covariance are both measuring the relationship of variables, but as stated in previous section, covariance is sensitive to scale of variables. To overcome this problem, let's say that we need to compare a linear relationship of data in meter, and another in kilometer scale, we need to standardize the covariance by dividing by the standard deviation (σ) of the two variables. Correlation can be expressed in term of covariance as the following:

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (2.16)$$

where σ_x and σ_y are standard deviations of variable x and y .

2D cross-correlation is commonly used to measure similarity between two images that is quite useful in pattern recognition. It can easily computed by summing up the multiplication of pixels from first image with second image underneath (Figure 2.8).

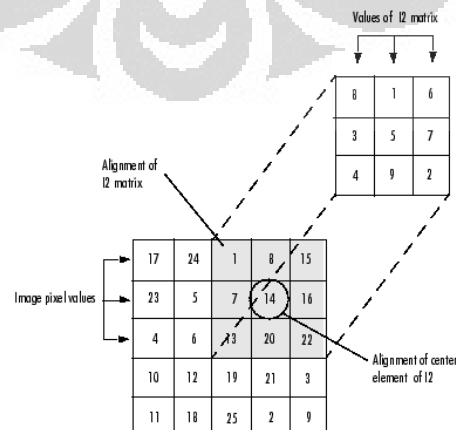


Figure 2.9. Cross-correlation between two images (from MATLAB Help)

2.10 Principal Component Analysis

Principal component analysis is a method of multivariate data analysis which aims to reduce dimensionality of interconnected variables and maximize the variance of linear combination of variables, i.e. searching for a dimension along which the data distribution is maximally spread out. It is a well established statistical method and has been used in analyzing seismic data. In a 1999 SPE publication, Scheevel *et al*, discussed about the application of PCA to seismic amplitude attribute for reservoir characterization. In SEG annual meeting presentation (2006), Guo *et al*, introduce principal component analysis of seismic spectral components. The purpose of PCA is to reduce number of spectral components from 80 or more to a more manageable number of principal components.

Assume that we have n variable vectors y_1, y_2, \dots, y_n that form a swarm of points in p -dimensional space. If all the variables are correlated, the ellipsoidal swarm of points is not oriented parallel to any of the axes represented by y_1, y_2, \dots, y_n . Then we would like to find the natural axes of the ellipsoidal swarm of points with origin at \bar{y} , that is the mean vector of y_1, y_2, \dots, y_n . This is done by translating the origin to \bar{y} and then rotating the axes. In this way, all the variables will be uncorrelated after the axes rotation. Axes rotation is done by multiplying y_i by an orthogonal matrix A , which transform y_i to z_i , with the same distance from the origin:

$$z_i = Ay_i \quad (2.17)$$

Finding the axes of ellipsoid data cloud is equivalent to find an orthogonal matrix A that rotate the axes to align with the natural extension of the swarm of points so that the principal components z_1, z_2, \dots, z_n are uncorrelated. So, we want the sample covariance matrix of z , $S_z = ASA'$ to be diagonal:

$$\mathbf{S}_z = \mathbf{A}\mathbf{S}\mathbf{A}' = \begin{pmatrix} s_{z_1}^2 & 0 & \cdots & 0 \\ 0 & s_{z_2}^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & s_{z_n}^2 \end{pmatrix}$$

where \mathbf{S} is the sample covariance matrix of $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$. Because \mathbf{A} is an orthogonal matrix whose columns are eigenvectors of \mathbf{S} , $\mathbf{S}_z = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, where λ_i are eigenvalues of \mathbf{S} . The principal components are the transformed variables $z_1 = \mathbf{a}_1' \mathbf{y}$, $z_2 = \mathbf{a}_2' \mathbf{y}$, ..., $z_n = \mathbf{a}_n' \mathbf{y}$. The principal component with largest eigenvalue, in this case, z_1 , has the largest variance and z_n has the smallest variance (Figure 2.10).

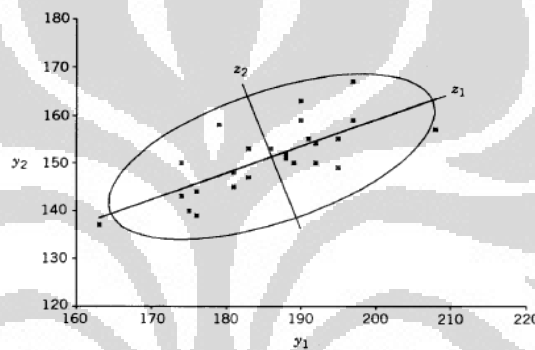


Figure 2.10. PCA transformation of a variable y_1 and y_2 (Rencher, 2002)

To illustrate how principal component analysis works on spectral components, consider the following frequency slice of an imaginary seismic data (Figure 2.11) and the traces plot in three spectral dimensions (Figure 2.12):

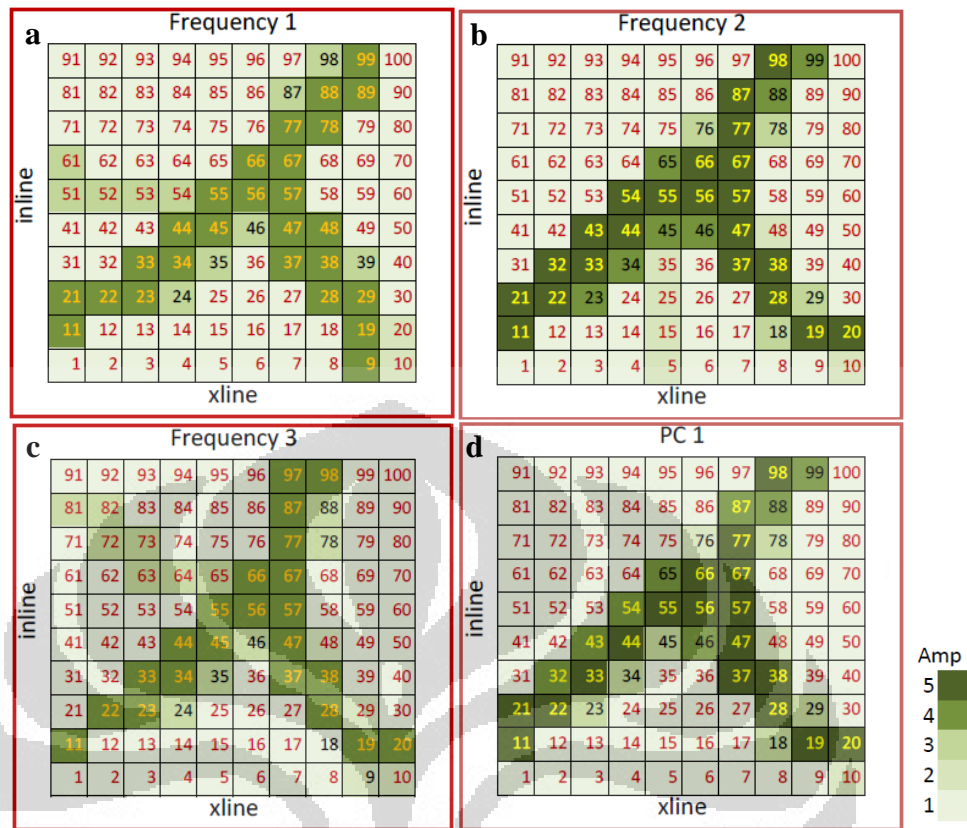


Figure 2.11. Cartoon showing frequency slice of seismic data. Numbers written in the box denote trace number. PC band 1 is plot in (d).

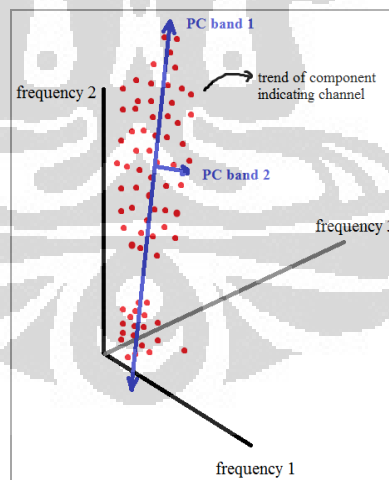
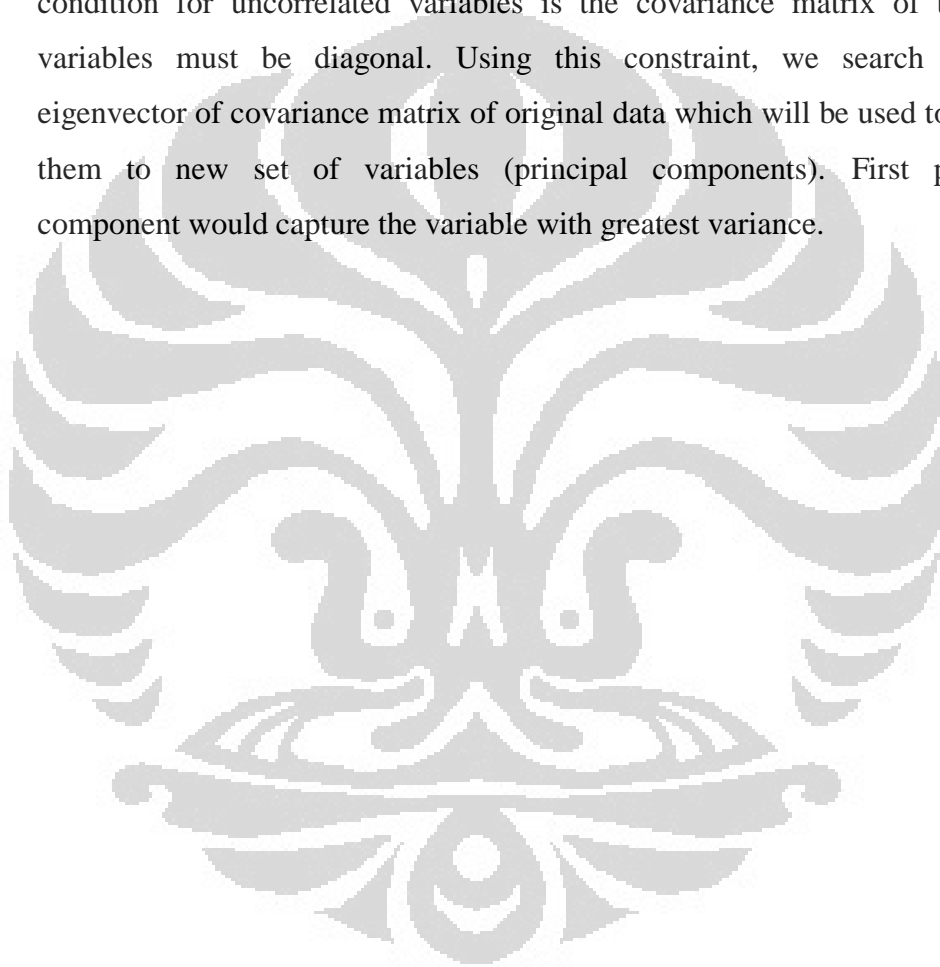


Figure 2.12. Principal component of data consisting of three components from the previous channel model (Figure 9), with red dots represent three spectral components from each trace.

From the frequency slice cartoon, we can see an obvious feature of the data represented by high amplitude (dark green colour) and it is most clearly delineated at frequency 2 (Figure 2.11.b). Each trace is plotted to three

dimension axes and we can see that traces that contain channel gathered near frequency 2 axes. We can also see that all the variables are correlated, in other word, we can see that feature in other spectral component, although at different clarity.

To remove the correlation, so that we can plot the feature in one image, we have to rotate the axes align with the natural extention of data cloud. One condition for uncorrelated variables is the covariance matrix of the new variables must be diagonal. Using this constraint, we search for the eigenvector of covariance matrix of original data which will be used to project them to new set of variables (principal components). First principal component would capture the variable with greatest variance.



CHAPTER 3

METHODOLOGY

After having a good grasp of how spectral decomposition and principal component analysis work, I am going to discuss about the methodology and workflow of PCA applied to synthetic and real spectral data. The algorithm is written and executed using MATLAB™ from Mathworks.

3.1 Processing of Synthetic Data

To evaluate PCA method, we use 3D wedge synthetic data. It is a good model to illustrate tuning effect at thin bed case and also hypothetical model of channel body which becomes thicker in one direction.

The purpose of constructing this synthetic model is to investigate its appearance in time slice, frequency slice using spectral decomposition, and the effectiveness of principal component analysis in delineating the whole wedge body.

The synthetic wedge model was constructed by defining reflectivity coefficients, -0.6 for the top and 0.6 for the base, and then convolved the reflectivity with 20 Hz Ricker wavelet which sampling interval is 2 ms (Figure 3.1). In this work, the model is in 3 dimension having 128 vertical samples, 70 crosslines, and 80 inlines (Figure 3.2). The top and base interface can be distinguished at relatively thick part but the reflections start to interfere destructively as the thickness diminishes.

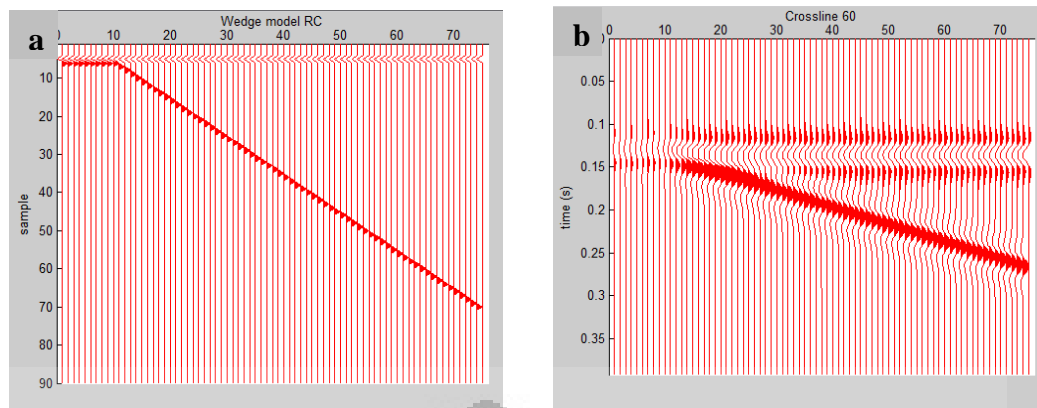


Figure 3.1. (a). Reflectivity model of wedge, (b). Synthetic seismogram from wedge model at crossline 60

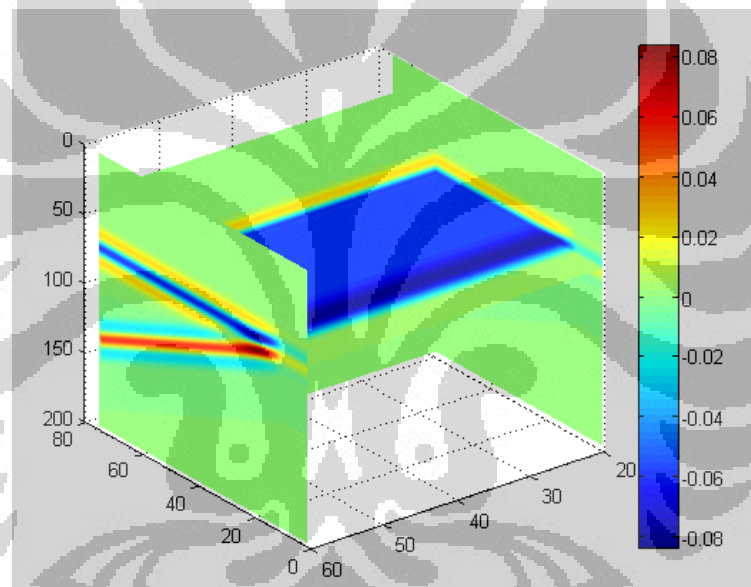


Figure 3.2. Three dimension image of wedge body.

The second step is transforming all traces into frequency domain with Fast Fourier Transform using MATLAB built-in function (fft) with 200 points output. This domain transform generated 200 spectral components with interval every 2.5 Hz. These components are then cross-correlated to get 200 by 200 covariance matrix. To perform PCA, eigenvectors are decomposed from the covariance matrix. Original data is projected to the eigenvector to obtain new components.

3.2 Processing of Real Seismic Data

We continue with a real 3D seismic data to evaluate the effectiveness of principal component analysis to enhance channel distribution map, which sand thickness is unknown. Seismic data used in this project is from Stratton Field, Texas. The data was already in form of matrix format (*.mat), readily loaded using MATLAB. It has 101 time samples with sampling rate 2 ms, 200 crosslines, and 100 inlines (Figure 3.3). Crossline 50 is shown in Figure 3.4.

Because we are not provided with information from well log, we first scanned the time slice to determine zone of interest where geologic feature we want to map does appear. Localizing zone of interest will give better approximation of frequency domain representation because it takes into account subtle information within the window. Otherwise, the frequency spectrum will approximate seismic wavelet and could not resolve interesting spectra from subtle feature.

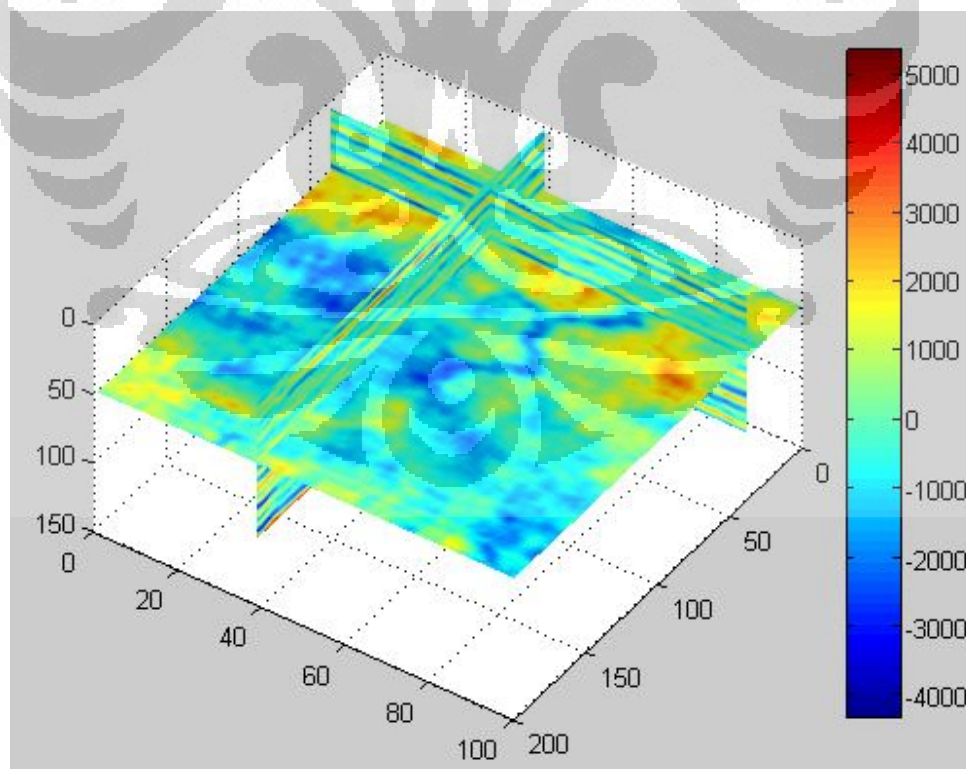


Figure 3.3. Seismic slice at crossline 39, inline 39, and time slice (47) ms, notice the channel pattern (blue).

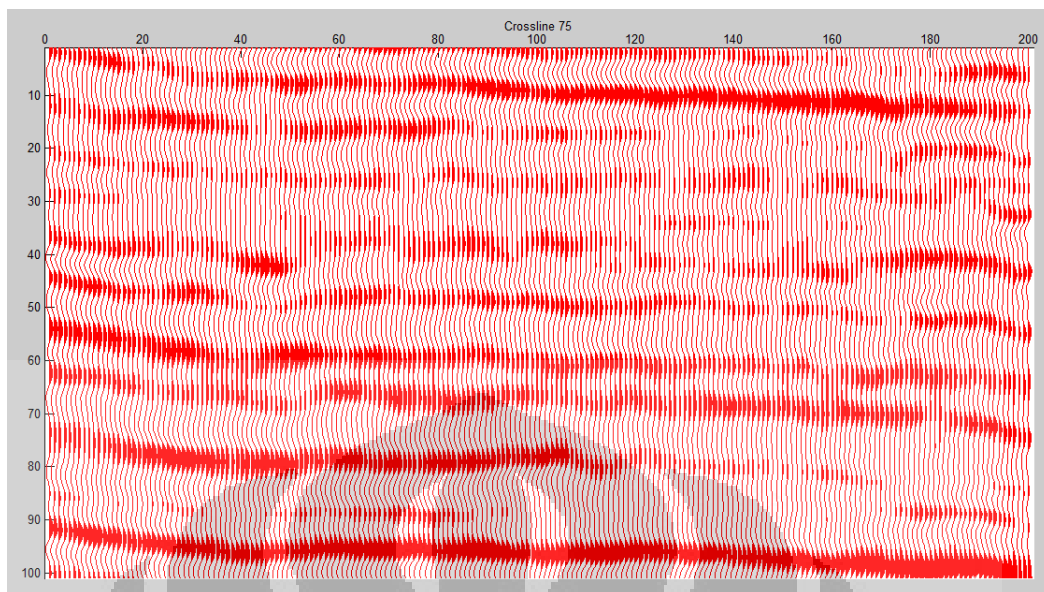


Figure 3.4. Stratton seismic data at crossline 75

Time to frequency domain transformation of traces from every inline n and crossline m is done using Fast Fourier Transform algorithm in MATLAB (*fft* function) with 501 points output. This generated 501 spectral components with 1 Hz increment. The results of transformation become the input to covariance matrix, which is the cross-correlation of every frequency slices with formula as describe in Guo, *et al*:

$$C_{j,k} = \sum_{n=1}^N \sum_{m=1}^M d_{mn}^{(j)} d_{mn}^{(k)} \quad (3.1)$$

where C_{jk} is the jk th element of covariance matrix C ; N is the number of seismic lines in the survey; M is the number of seismic crosslines in the survey; and $d_{mn}^{(j)}$ and $d_{mn}^{(k)}$ are spectral magnitudes of j th and k th frequencies at line n and crossline m . The cross-correlation values are not normalized; it means that they are not divided by the standard deviation of each variable, because our spectral data have already had the same scale. We also do not subtract the spectral magnitude from its mean value (as the covariance and correlation formula) because every frequency component the same range of spectral magnitude value. So we can get a quick and meaningful result with minimum computation.

The next step is decomposing covariance matrix into eigenvector v_p and eigenvalue λ_p by solving equation:

$$Cv_p = \lambda_p v_p \quad (3.2)$$

In MATLAB, the eigenvector is obtained by applying *pcacov* function to a p-by-p covariance matrix **C** which returns three outputs: *coeff*, *latent*, and *explained*. *coeff* is a p-by-p matrix which each column containing coefficients of a principal component, that are sorted in order of decreasing variance. *coeff* is different from eigenvector produced by *eig* function in MATLAB, since *coeff* is the normalized eigenvector of covariance matrix, so all the vectors are unit length. *latent* is a vector containing principal component variances, that is the eigenvalues of covariance matrix **C** and the percentage of every principal component variances are stored in *explained*. The variable *explained* is the fraction of: variance of a principal component and total variances. This value shows how much a PC band represents the entire data because large variance means the principal component captures the major trends of data cloud. Eigenvectors with largest eigenvalues represent dimensions with largest variance in the data set.

The last step is transforming original spectra d into new variables z using equation:

$$z = v_p d \quad (3.3)$$

which linear combination has maximum variance. The principal components are sorted in order of decreasing eigenvalue; with largest eigenvalue represent linear combination with maximum variance. The comprehensive workflow is shown in Figure 3.5.

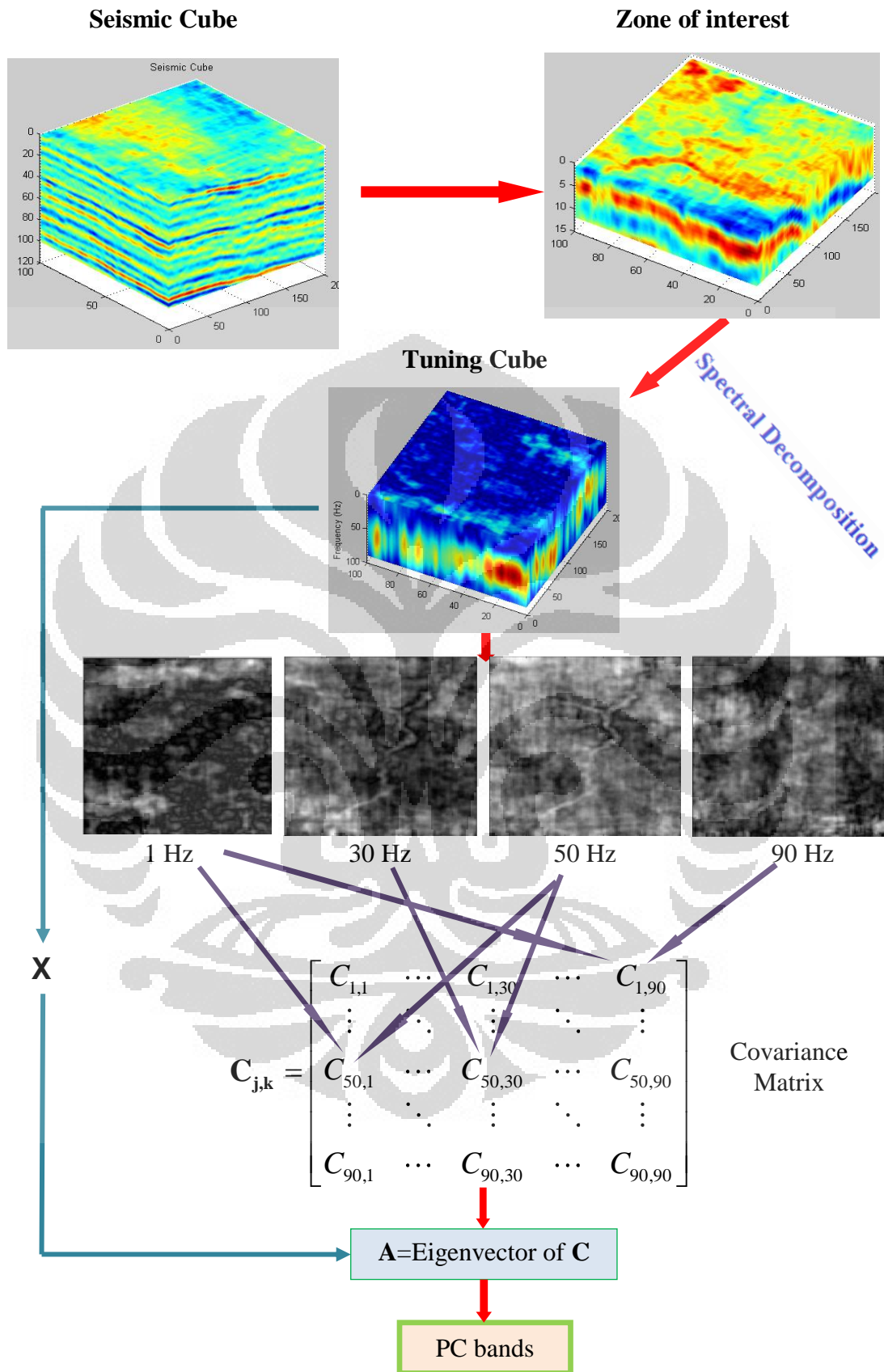


Figure 3.5. Cartoon showing workflow of Principal Component Analysis

CHAPTER 4

RESULTS AND ANALYSIS

In this section, I will discuss about the result and analysis of the performance of principal component spectral analysis applied to a 3D synthetic wedge model and also a real 3D seismic data cube.

4.1 Synthetic Data

From this 3D synthetic seismogram cube, we perform time slice to map the wedge body. As predicted, it appears that time slice only map the partial part of the body because it is very sensitive to interface position in time domain. From time slice, it is obvious that the wedge is thickening downward (Figure 4.1).

To map the wedge in frequency slice, which can give an indication of bed temporal thickness, we first transform all traces into frequency domain using Fast Fourier Transform.

From the amplitude spectrum of crossline 60, we can see that different wedge thickness has different amplitude spectrum response. So, when we see the frequency slice, we expect to see bright and dark bands pattern that indicates particular temporal thickness. For example, from Figure 4.2, we can see that frequency slice at 25 Hz will map wedge with temporal thickness 0.025, 0.05, 0.1 s.

Figure 4.3 shows spectral decomposition of wedge model. Notice that at frequencies 3 – 7 Hz, although spectral decomposition is able to delineate most part of the wedge without bright and dark bands, it is not very descriptive because spectral decomposition is sensitive to thickness. So, only parts of the wedge that have temporal thickness correspond to a certain frequency are highlighted.

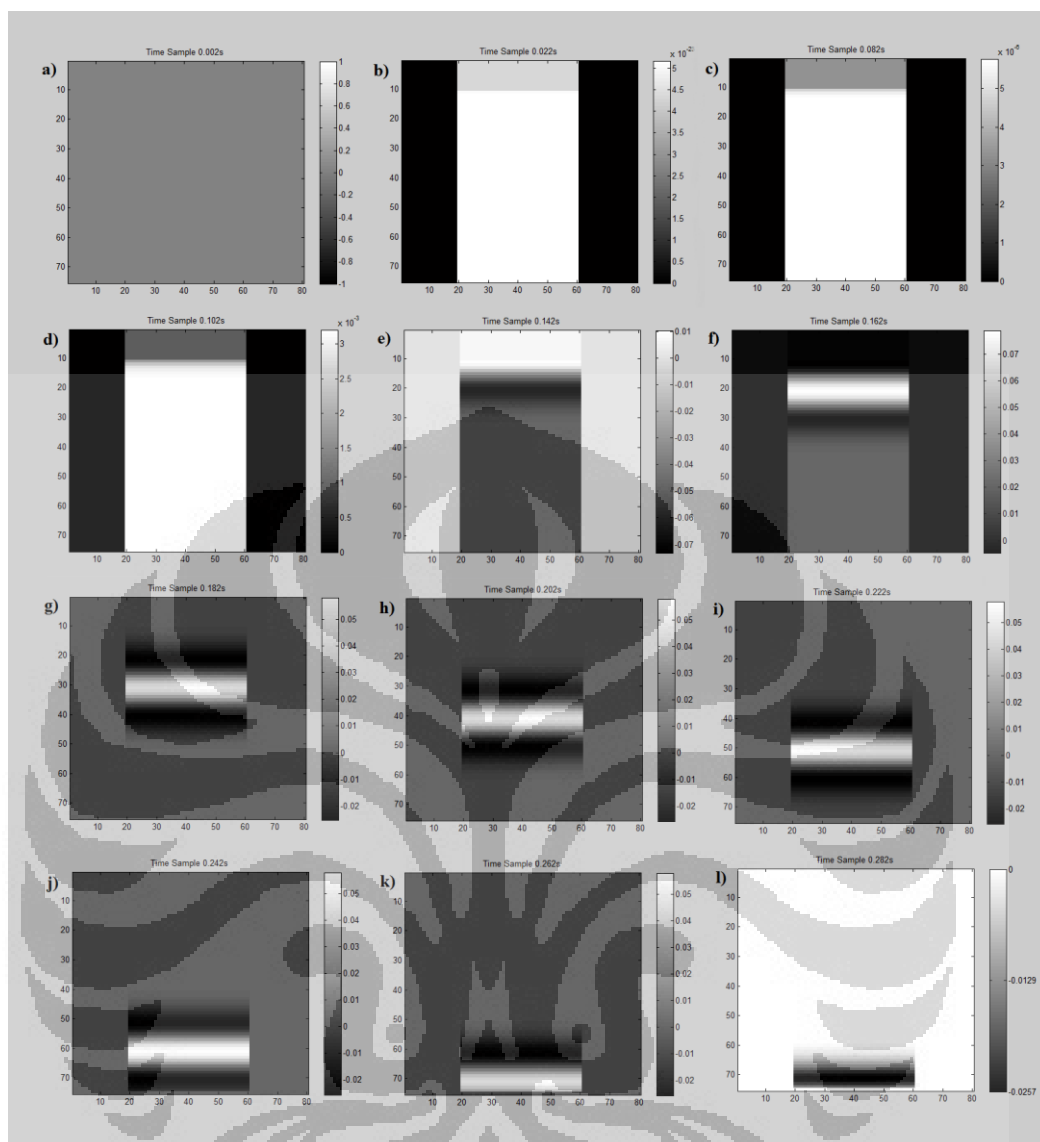


Figure 4.1. Time slice of 3D synthetic seismogram wedge model every 0.2 ms, starting at 2 ms

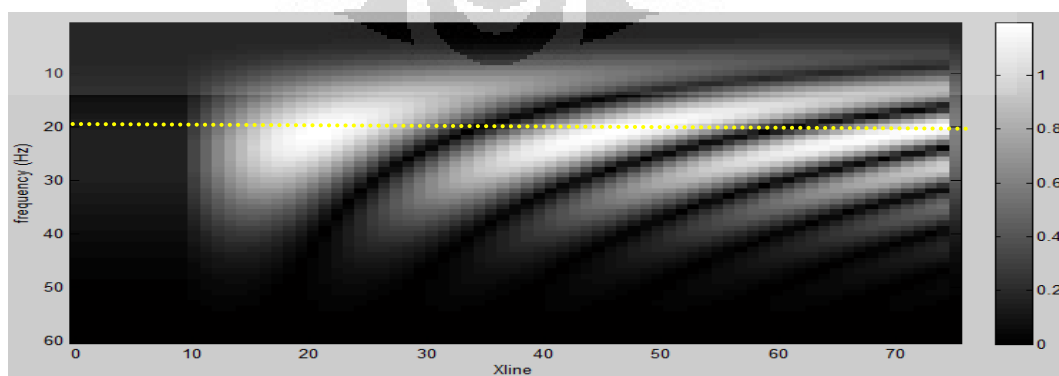


Figure 4.2. Amplitude spectrum of traces in crossline 50

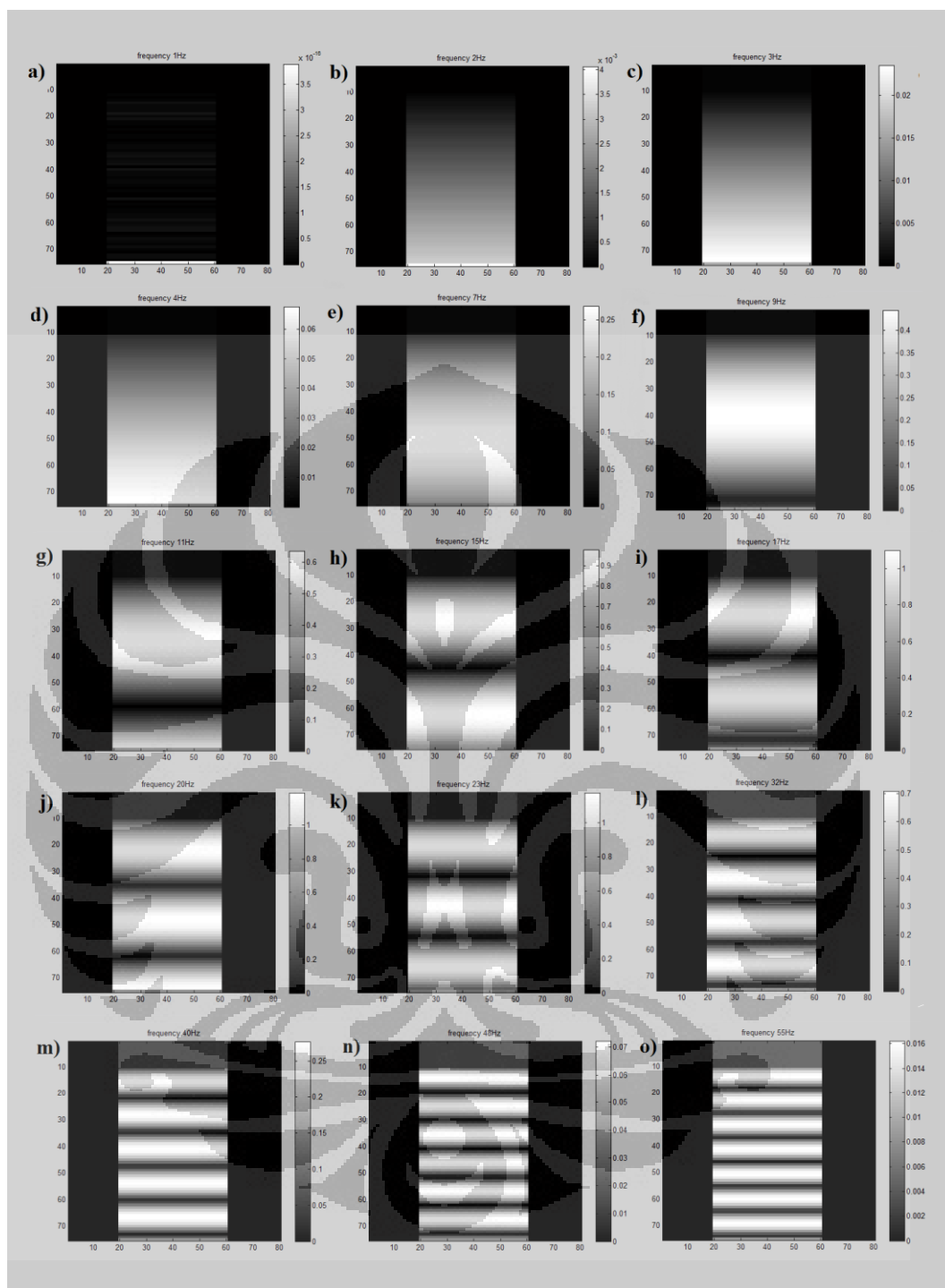


Figure 4.3. Frequency slice of wedge model showing bright and dark pattern at different frequencies: (a) 1 Hz, (b) 2 Hz, (c) 3 Hz, (d) 4 Hz, (e) 7 Hz, (f) 9 Hz, (g) 11 Hz, (h) 15 Hz, (i) 17 Hz, (j) 20 Hz, (k) 23 Hz, (l) 32 Hz, (m) 40 Hz, (n) 48 Hz, and (o) 55 Hz

To overcome the sensitivity problem, we try to run principal component analysis which looks for the trend in our data and then map it on the first few principal component bands. In this way, we expect to see all the wedge body in one image. Figure 4.4 shows first four PC bands, also band 17th and 163rd as comparison. We can see that the wedge is mapped to the first PC band regardless of the thickness. First principal component accounts for 85% variance and the entire first five principal components account for 95% variance; this is plotted on Figure 4.5.

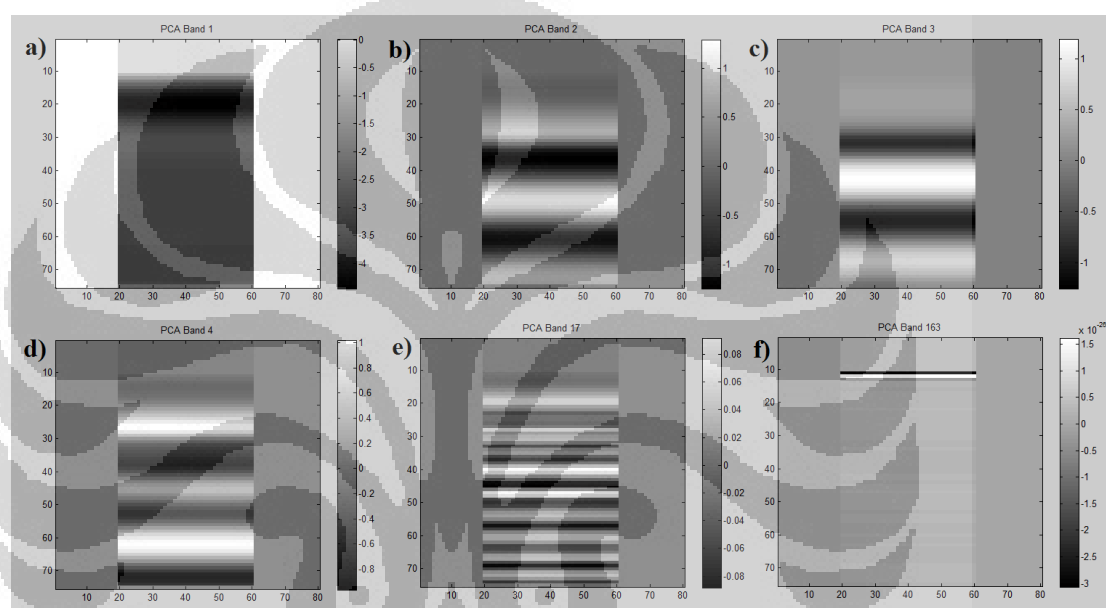


Figure 4.4. Spectrum projected onto (a) first, (b) second, (c) third, (d) fourth, (e) 17th, and (f) 163rd principal component.

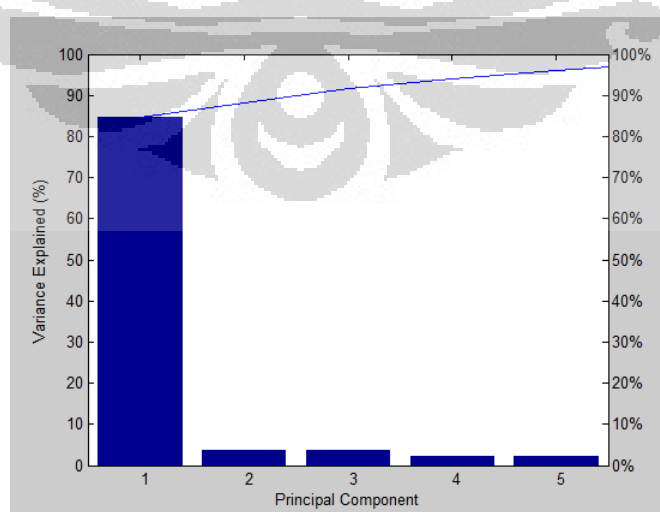


Figure 4.5. Scree graph of variance percentage explained by principal components

4.2 Three Dimension (3D) Seismic Data

As discussed in Chapter 3, we have to locate zone of interest (channel deposit) by scanning time slices because we are not provided with log data. From Figure 4.6, the channel is visible starting at 0.06 ms and diminishes at 0.104 ms. Because of structural variation, some part of the channel appears to have different polarity. At time sample 0.06 until 0.064 ms (Figure 4.6.a-c), northern part of the channel has positive polarity and becomes negative at 0.066 until 0.094 ms (Figure 4.6.d-g), while the southern part is visible in positive polarity from 0.066 until 0.092 ms (Figure 4.6.d-f).

Next, we perform spectral decomposition using FFT algorithm to the zone of interest cube to see at what frequency the channel is clearly imaged. Choosing FFT transformation window is very critical because it affects the shape of amplitude spectrum greatly and in turn determining whether the geological subtleties are captured in the spectrum.

This frequency decomposition generated 501 spectral components from 0 until 500 Hz. From Figure 4.7, part of channel pointed by yellow arrow tune in first at 5 Hz and then gradually tune out at 30 Hz, indicating the thickest part. Also notice that the boundary of channel (red arrow) appears to be dark at low frequency, and then tune in at higher frequency

The whole channel is quite well imaged at frequency slice 20 – 30 Hz. Other nearby spectral components also give almost the same image with different part that tune in and out since the thickness of the channel is not homogenous.

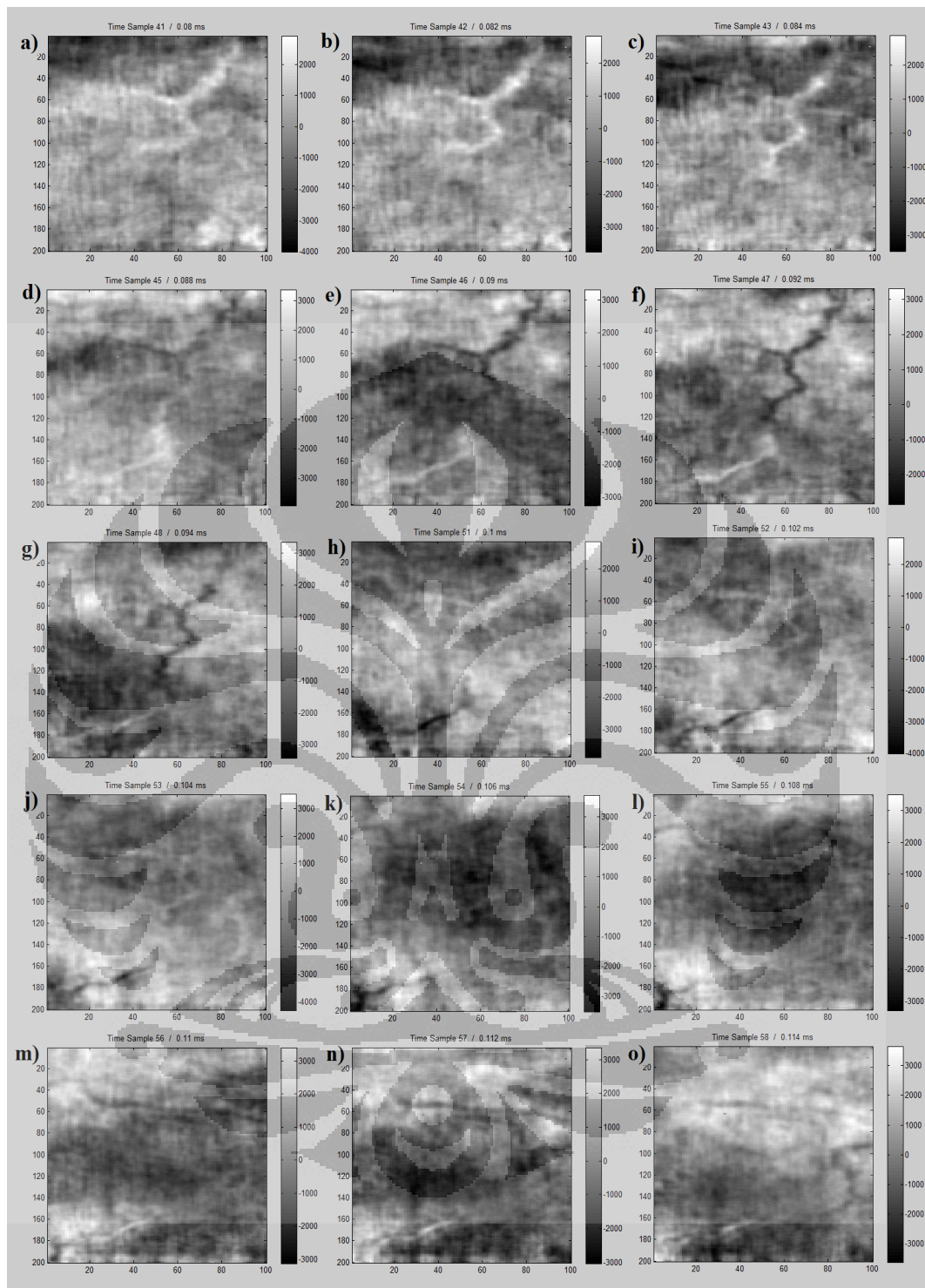


Figure 4.6. Time slice of Stratton seismic cube at (a) 0.06 ms, (b) 0.062 ms, (c) 0.064 ms, (d) 0.088 ms, (e) 0.09 ms, (f) 0.092 ms, (g) 0.094 ms (h) 0.1 ms, (i) 0.102 ms, (j) 0.104 ms, (k) 0.106 ms, (l) 0.108 ms, (m) 0.11 ms, (n) 0.112 ms, (o) 0.114 ms

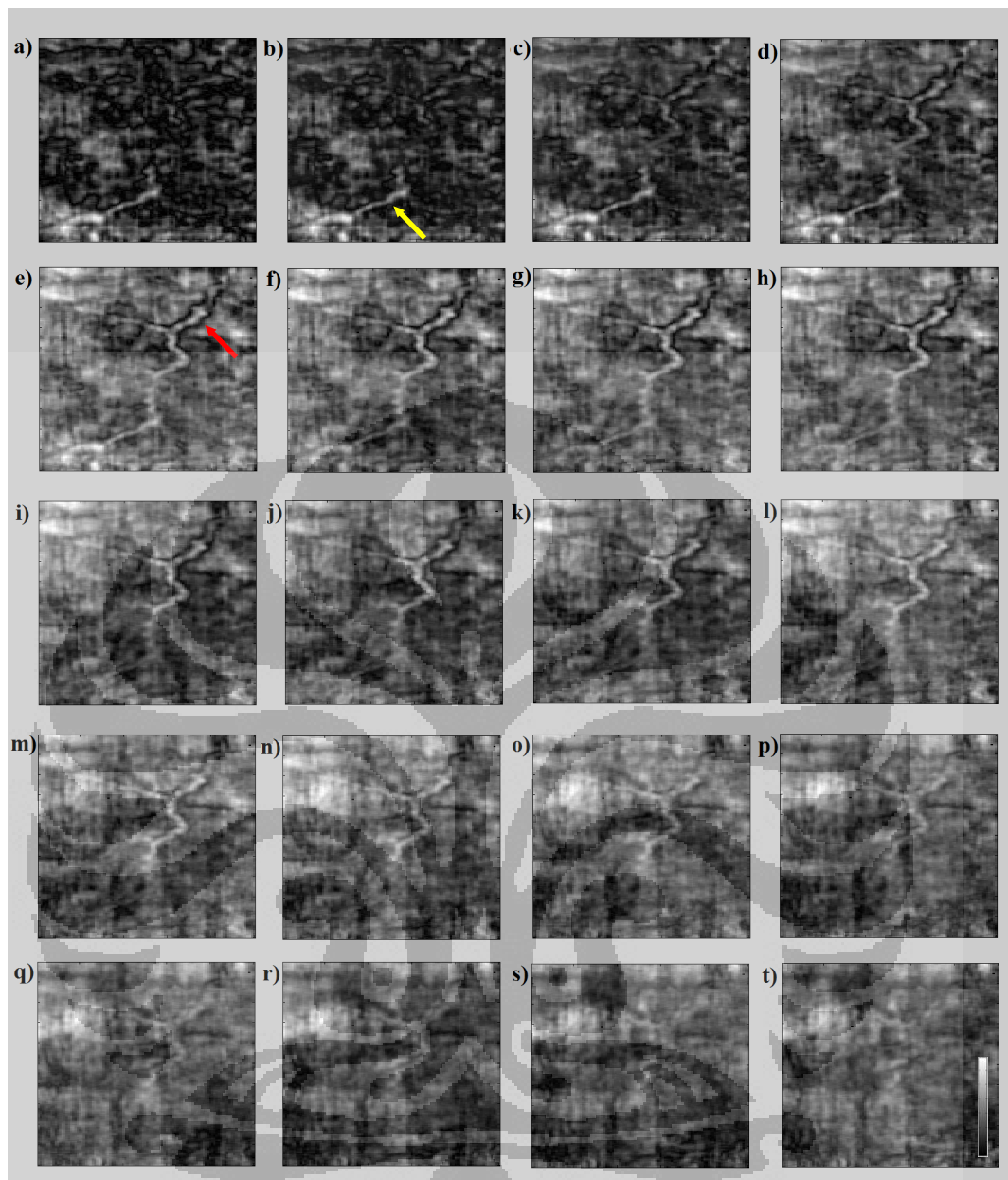


Figure 4.7. Stratton data frequency slice at (a) 0 Hz, (b) 5 Hz, (c) 10 Hz, (d) 15 Hz, (e) 20 Hz, (f) 25 Hz, (g) 30 Hz, (h) 35 Hz, (i) 40 Hz, (j) 45 Hz, (k) 50 Hz, (l) 55 Hz, (m) 60 Hz, (n) 65 Hz, (o) 70 Hz, (p) 75 Hz, (q) 80 Hz, (r) 85 Hz, (s) 90 Hz, and (t) 95 Hz.

We realize that most of the spectral components are highly correlated and carry almost the same information about the channel, like in Figure 4.7, we can see that major part of the channel is coexist from frequency slice 10 until 70 Hz. To capture the trend, we transform the original data to remove the correlation among the bands using principal component analysis (PCA). Figure 4.8.a shows the covariance matrix of original spectral components.

To perform PCA to our data, we utilize 86 spectral components from 5 to 90 Hz and then form a 86-by-86 covariance matrix using equation (3.1), compute the eigenvector and eigenvalue pairs by solving equation (3.2), and project the original spectra to eigenvectors to yield new sets of data. Figure 4.9 shows the amount of information from the data that is defined by each principal component. The first principal components alone accounts for more than 97% spectral variance. Note that the cross-correlation among principal components is zero, except correlation with themselves (Figure 4.8.b).

In Figure 4.10, we map the five largest principal components and also the 16th component. Channel feature is best represented by the first PC band (Figure 4.10.a), with decreasing significant feature map on the rest PC bands. Note that in large PC band number (Figure 4.10.f), the image mostly contains noise, thus, PCA can be an effective filter mechanism improve signal-to-noise ratio by throwing out PC bands that contain insignificant signal. PCA successfully capture the major trend (channel) in our data without needing to scan the entire frequency slice. This can be very useful for application in data compaction, since we are able to reject lots of insignificant components. Because PCA capture trend of data, it map the most energetic trend regardless of the thickness. So, while spectral decomposition is sensitive to thickness, PCA is the opposite.

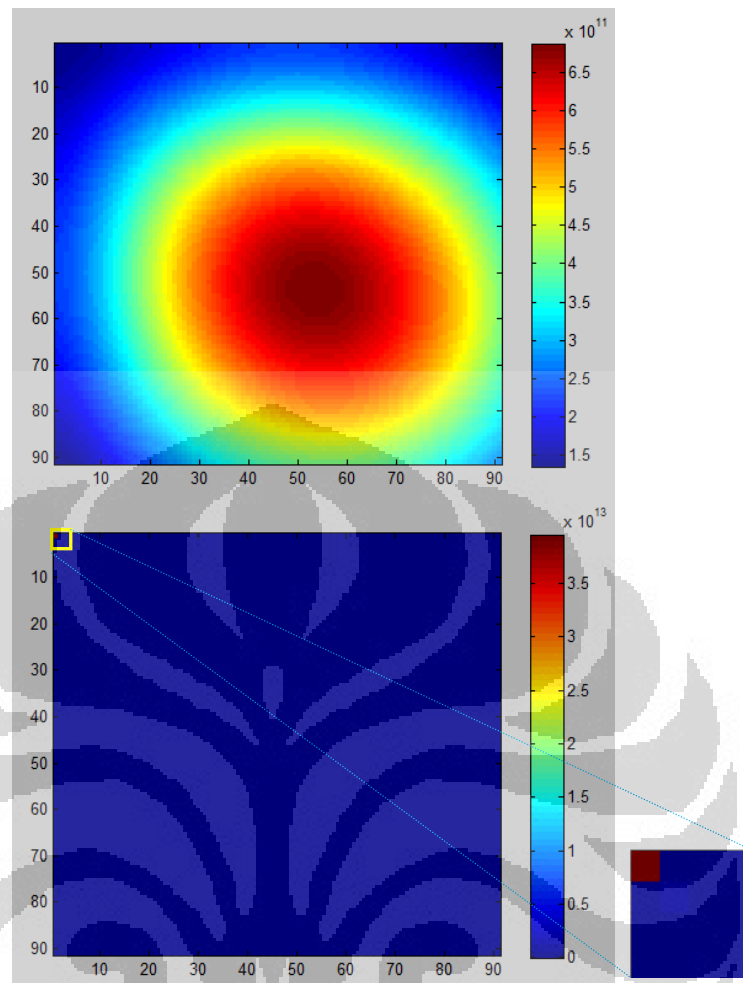


Figure 4.8. Covariance matrix of (a) 91 original spectral components, (b) 91 principal components displayed in form of colored image. Hot color represent high correlation.

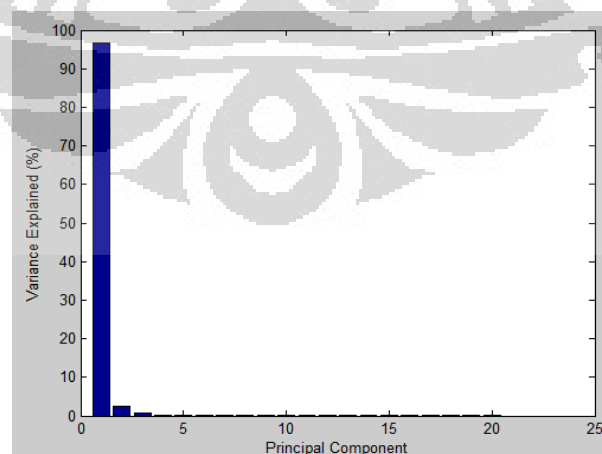


Figure 4.9. Graph showing variance explained by PC. The first principal component accounts for more than 97% variance.

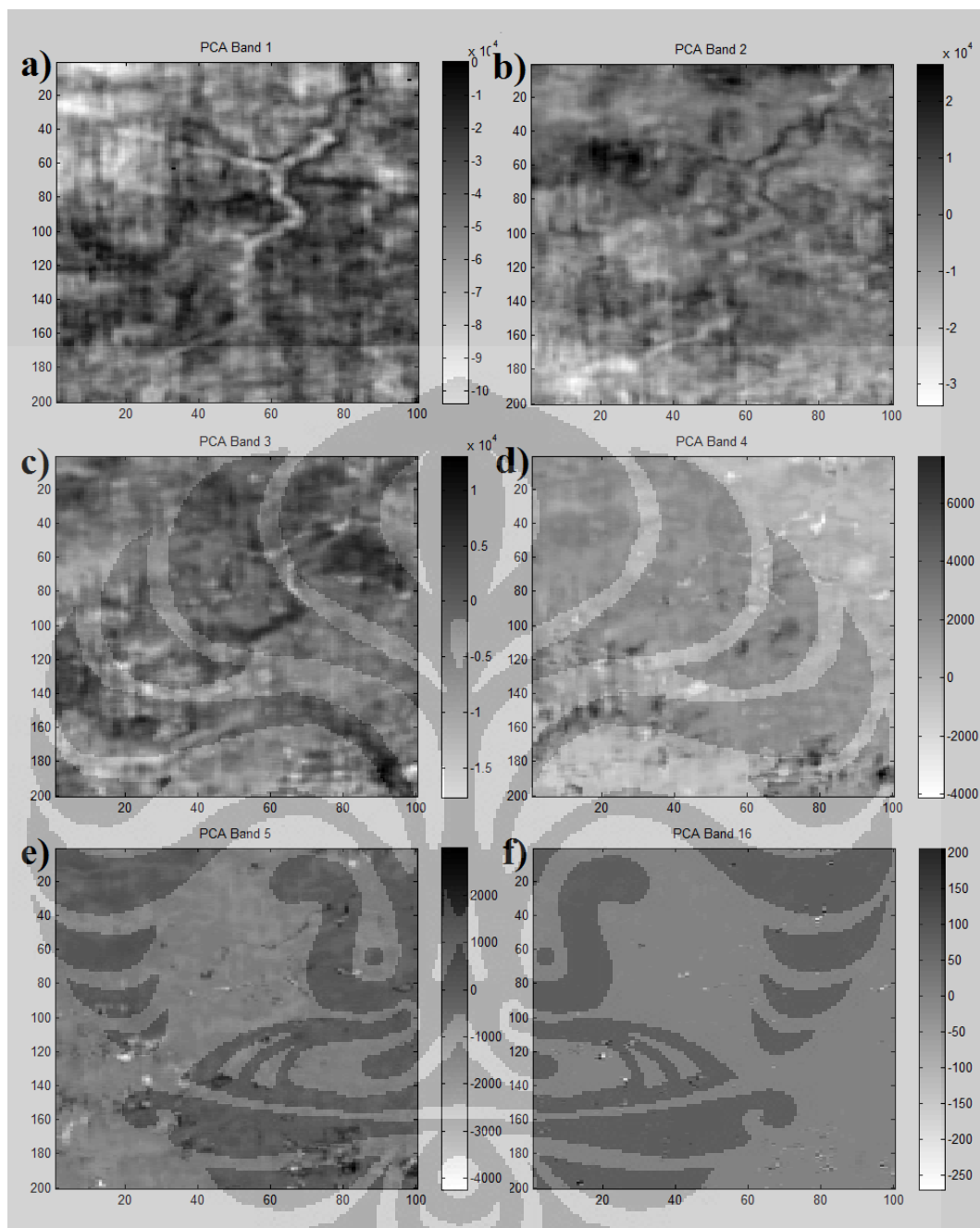


Figure 4.10. Spectral components projected to (a) 1st, (b) 2nd, (c) 3rd, (d) 4th, (e) 5th, and (f) 16th PC bands.

To compare between the result of conventional spectral decomposition and principal component spectral analysis, we use PC band 1 and frequency slice 25 Hz (Figure 4.11). We notice that image of the first PC band manage to plot the major part of channel, with slightly better image at the northern part (blue arrow), the channel appears to be more continuous although other spectral components like 35 – 45 Hz also display a continuous feature there.

But the western (yellow arrow), and the southern part (red arrow) are slightly poorly imaged. Also point out that those parts are slightly smeared in frequency slice too. The reason slightly reduce quality in PC band is due to properties of PCA that seeks the most energetic trend in data, but if the channel has low reflectivity, it will not appear as a significant component in data and is not restored in the few first principal components. In other words, it overlooks subtle feature that may be interesting for a geoscientist. Conversely, if noise (such as acquisition footprint) has higher amplitude and seems to be coherent, it will show up in first few principal components.

By summing up or blending the first three principal components, i.e. 1st, 2nd, and the 3rd PC bands, we can enhance the channel distribution mapping. In Figure 4.11.c, southern part of the channel (green arrow) which appears strongly at frequency slice 0 – 15 Hz, is better imaged compared to using principal component band 1 alone. This blending technique will capture most parts of channel that only appear in several frequency slices.

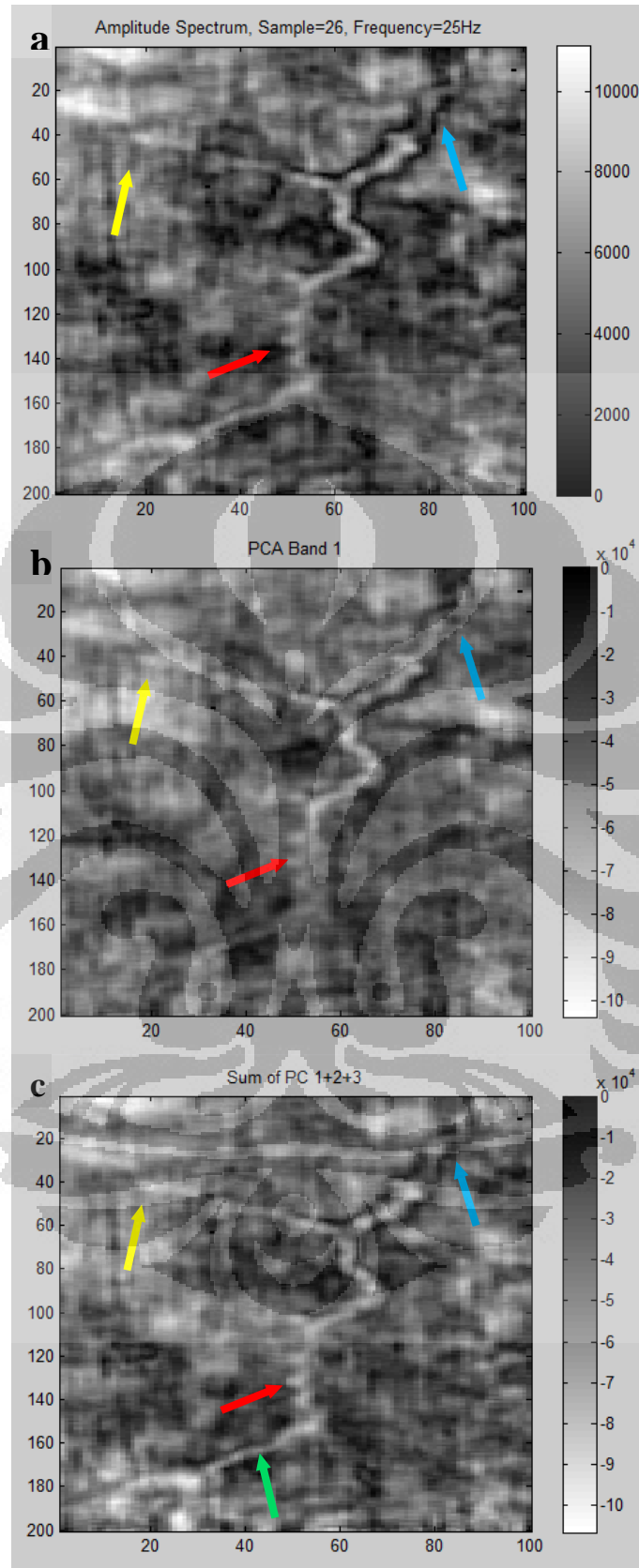


Figure 4.11. Comparison between (a) Frequency slice at 25 Hz, (b) the first principal component band, (c) sum of 1st, 2nd, and 3rd PC bands.

CHAPTER 5

CONCLUSION

From the result of this thesis work, we conclude that principal component analysis is proved to be an effective tool for geophysicists to track the trend of data they have. Principal component analysis is also very useful in data compaction since it reduce more than 90 spectral components that have to be analyzed into less than 5 principal components. The principle is to let the data tell something by themselves. In our case, i.e. channel mapping, principal component reject most of the insignificant signal, hence only project the main feature to the first few PC bands. While spectral decomposition is very sensitive to thickness, PCA is the opposite. In case we have a continuous feature but with different thickness, PCA will plot all the body in the first principal component band if it is also a trend in the data.

Since PCA basically look for statistical trend between spectral components, it has a major pitfall, i.e. subtle feature with little reflectivity or those which do not coexist in other spectral components will be overlooked or poorly imaged. In our work, some part of the channel is blurred, not because PCA fail to capture that feature, but because of the blurry part is not coexist in most of spectral components. By summing up the first three principal component bands, we show that the channel distribution can be better imaged.

References

- Chopra, S., K. J. Marfurt, 2005, Seismic attributes – A historical perspective: *Geophysics*, **70**, no. 5, 3SO – 28SO.
- Guo, H., K. J. Marfurt, and J. Liu, 2009, Principal component spectral analysis: *Geophysics*, **74**, no. 4, 35 – 43.
- Haris, A., 2008, *Komputasi Geofisika: Universitas Indonesia.*, Depok.
- Kallweit, R., S., and L. C. Wood, 1982, The limits of resolution of zero-phase wavelets: *Geophysics*, **47**, no. 7, 1035 – 1046.
- Laughlin, K., P. Garossino, and G. Partyka, 2002, Spectral Decomposition for Seismic Stratigraphic Patterns: *AAPG Explorer*.
- Partyka, G. A., J. Gridley, and J. Lopez, 1999, Interpretational applications of spectral decomposition in reservoir characterization: *The Leading Edge*, **18**, 353 – 360.
- Rodarmel, C., and J. Shan, 2002, Principal component analysis for hyperspectral image classification: *Surveying and Land Information Science*, **62**, 115-122.
- Rencher, A. C., 2002, *Methods of Multivariate Analysis 2nd ed*: John Willy and Sons, Inc., Canada.
- Scheevel, J. R. and K. Payrazyan, 1999, Principal component analysis applied to 3D seismic data for reservoir property estimation: *Annual SPE Technical Conference*.
- Smith, L. I., 2002, A tutorial on principal component analysis.
- Widess, M. B., 1973, How thin is a thin bed?: *Geophysics*, **38**, no. 6, 1176 – 1180.

Appendices

A. Wedge Model

```

clear all
clc
%ANDREW JO/0606068026
%RC wedge 3D
wed=zeros(70,75,80);
wed(5, :, 20:60)=-0.6;
wed(6, 1:10, 20:60)=0.6;
for i=6:70;
    wed(i, (i+4), 20:60)=0.6;
end

%Ricker 20 Hz
dt=0.002;
fdom=20;
tlength=127*dt;
[wave, tw]=ricker(dt, fdom, tlength);

%3D-->2D
wed2=reshape(wed, 70, 6000);
%wed2=wed2';

%convolv
seisw=conv2(wed2, wave);
[ntime, ntrace]=size(seisw);

%2D-->3D
seisw3=reshape(seisw, ntime, 75, 80);
%-----
%Plot crossline 60
r=seisw3(:, :, 60);
rt=0:dt:(ntime-1)*dt;
figure(1); plotseis(r, rt); title('Crossline 60'); ylabel('time (s)')

seisw=seisw';
seis2d=seisw;
seisw=reshape(seisw, 75, 80, 197);

%Slice
[ny, nx, nz]=size(seisw);
[x, y, z]=meshgrid(20:60, 1:ny, 1:nz);
figure(2);
slice(x, y, z, seisw(:, 20:60, :), [20, 60], 60, 70); colormap(jet); colorbar
;set(gca, 'xdir', 'reverse'); set(gca, 'zdir', 'reverse'); shading
interp

%FFT
seisf=abs(fft(seisw3, 500));
fx=0:1/(dt*499):1/dt;
figure(5); imagesc(fx, reshape(seisf(1:70, :, 50), 70, 75));

```

```

%time slice
figure(3);
for j=1:10:190;
    imagesc(reshape(seisw3(j, :, :), 75, 80)); colorbar; colormap(gray);
    title(['Time Sample ' num2str(j*0.002) 's']);
    pause;
end

%freq slice
for k=1:10:200;
    imagesc(reshape(seisf(k, :, :), 75, 80)); colorbar; colormap(gray);
    title(['frequency ' num2str(500/500*k) 'Hz']);
    pause;
end

%-----
%covariance matrix
[nf, il, xl]=size(seisf);
c(1:90, 1:90)=0;
%cross-correlation
for j=1:90;
    for k=1:90;
        for m=1:xl;
            for n=1:il;
                c(j, k)=c(j, k)+seisf(k, n, m)*seisf(j, n, m);
            end
        end
    end
end

%[V,D]=eig(c);
[V,D,variance]=pcacov(c);

seispca=reshape(seisf(1:90, :, :), 90, 6000);

seisfinal=zeros(90, 6000);
for g=1:6000;
    seisfinal(:, g)=V*seispca(:, g);
end

figure(6);
for o=1:20;

imagesc(reshape(seisfinal(o, :, :), 75, 80)); colormap(gray); colorbar;
    title(['PCA Band ' num2str(o)]);
    pause;
end

figure(8);
bar(variance(1:20));
xlabel('Principal Component')
ylabel('Variance Explained (%)')

```

B. Seismic 3D Data

```

clear all
clc
% %ANDREW JO/0606068026
load dataku

pts=501; %no of fft points
%
%=====
%Slice 3D
seissl=reshape(seis,101,20000);
seissl=seissl';
seissl=reshape(seissl,200,100,101);
[ny,nx,nz]=size(seissl);
[x,y,z]=meshgrid(1:nx,1:ny,1:nz);
figure(12);
slice(x,y,z,seissl,39,39,47);colormap(jet);
colorbar;set(gca,'xdir','reverse');
set(gca,'zdir','reverse');shading interp
%
%=====
%Plot Zone of Interest
seissl=reshape(seis,101,20000);
seissl=seissl';
seissl=reshape(seissl,200,100,101);seissl=seissl(:,:,41:52);
[ny,nx,nz]=size(seissl);
[x,y,z]=meshgrid(1:nx,1:ny,1:nz);
figure(112);
slice(x,y,z,seissl,[1 100],[1 200],1);colormap(jet);
colorbar;set(gca,'xdir','reverse');
set(gca,'zdir','reverse');shading interp
%
%=====
%Time Slice
figure(1);
for i=30:1:50;
    seis2=seis(i,:,:);
    seis2=reshape(seis2,200,100);
    imagesc(seis2);colorbar;colormap(gray);
    title(['Time Sample ' num2str(i) ' / ' num2str(i*0.002-
0.002) ' ms'])
    pause;
end
%
%=====
%FFT
freq=fft(seis(41:48,:,:),pts); %channel still visible in this
range, 100 pts
freqabs=abs(freq);
%
%=====
%Tuning Cube
seissl=reshape(freqabs,pts,20000);
seissl=seissl';
seissl=reshape(seissl,200,100,pts);
[ny,nx,nz]=size(seissl);
[x,y,z]=meshgrid(1:nx,1:ny,1:90);

```

```

figure(112);
slice(x,y,z,seissl(:,:,1:90),[1 100],[1 200],1);colormap(jet);
colorbar;set(gca,'xdir','reverse');
set(gca,'zdir','reverse');shading interp
%
%=====
%Tuning Cube Cross Section
ii=0;
for i=57:65;ii=ii+1;
figure(33);subplot(3,3,ii);imagesc(reshape(freqabs(1:100,i,:),100,
100));
title(['x1 ' num2str(i)]);colormap(jet);
end
%
%=====
%Frequency Spectrum of Traces
xl=62; %max 200
ii=0;figure(231);
for i=59:83;ii=ii+1;
subplot(5,5,ii);plot(1:200,freqabs(1:200,xl,i));xlabel('Hz');
title(['X1 ' num2str(xl) ', Trace #' num2str(i)])
end
%
%=====
%Frequency Slice
figure(16);
for n=10:10:100;
x=freqabs(n,:,:);
imagesc(reshape(x,200,100));colorbar;colormap(gray);
title(['Amplitude Spectrum, Sample=' num2str(n) ', Frequency='
num2str(501/pts*n-1) 'Hz']);
pause;
end
%
%=====
%Phase
phs=unwrap(angle(freq));
figure(98);
for i=1:5:100;
imagesc(reshape(phs(i,:,:),200,100));colorbar;
pause;
end
%
%=====
% %Covariance Matrix
[nf,il,xl]=size(freqabs);
comp=250; %number of spectral components to cross-correlate

% %Cross-Correlation
c=0;
c(1:comp,1:comp)=0;
for j=1:comp;
for k=1:comp;
for m=1:20000;
c(j,k)=c(j,k)+freqabs(k,m)*freqabs(j,m);
end
end
end
end

```



```

%
%=====
% %EigenV
fstart=6;
fend=200;
[V,D,variance]=pcacov(c(fstart:fend,fstart:fend));
seispca=reshape(freqabs(fstart:fend, :, :),fend-fstart+1,20000);
%
%=====
%Projection
seisfinal=zeros(fend-fstart+1,20000);
for g=1:20000;
    seisfinal(:,g)=V*seispca(:,g);
end
%
%=====
% %Covariance of new variables
d=0;
d(1:fend-fstart+1,1:fend-fstart+1)=0;
for j=1:fend-fstart+1;
    for k=1:fend-fstart+1;
        for m=1:20000;
            d(j,k)=d(j,k)+seisfinal(k,m)*seisfinal(j,m);
        end
    end
end
end
%
%=====
% %Plot PCA
figure(9);
for o=1:9;

subplot(3,3,o);imagesc(reshape(seisfinal(o, :, :),200,100));colormap
(flipud(gray));colorbar;
    title(['PCA Band ' num2str(o)]);
end
%
%=====
%Variance Explained
figure(8);
bar(variance(1:20));
xlabel('Principal Component')
ylabel('Variance Explained (%)')

```