STATISTICAL APPROACHES TO DE-ALIASING WITH APPLICATION TO EARTH SYSTEM MONITORING

by

Yihao Yin

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
October 2015

I remember all the stories you told me.

Table of Contents

List of Tables

List of Figures

Abstract

Space-borne altimeters typically have a repeat time of order 10 days and so alias high frequency sea level variability. State-of-the-art de-aliasing methods are presently based on tidal and atmospheric corrections from dynamical models. However, analysis shows that significant high frequency variability remains after such corrections that could cause aliasing problems. In order to further de-alias the altimetry products, a statistical de-aliasing model is designed. Three methods are designed to fit the model (i) in the lag domain, (ii) in the frequency domain, and (iii) in the time domain using the lasso to limit the number of predictors. The three methods are first tested in two simulation-based studies and shown to be both effective and interpretable. The methods are then applied to the altimetry products. The lasso-based method performs best and reduces the standard deviation of the satellite altimetry products in the Gulf of St. Lawrence from about 8 cm to 6 cm.

List of Abbreviations and Symbols Used

$R(\tau)$	Auto-covariance function
$R_{ij}(s)$	Cross-covariance function
$X(t)$	Continuous parameter time series
X_t	Discrete parameter time series
Δ	Sampling interval
$\Gamma(\omega)$	Transfer function
Σ	Covariance matrix
$\alpha_{ij}(\omega)$	Gain-spectrum
η	Sea level variable
$\mathbf{X}_t$	Discrete parameter bivariate time series
$\lvert w_{ij}(\omega)\rvert$	Coherency-spectrum
$\phi_{ij}(\omega)$	Phase-spectrum
$\rho(\tau)$	Auto-correlation function
$\rho_{ij}(s)$	Cross-correlation function
$X \mid G_1$	Observations of time series X_t on the grid G_1
X^H	High frequency component of time series X_t
X^L	Low frequency component of time series X_t
$\{W(\theta)\}$	Spectral window
$\{\lambda(k)\}$	Lag window
$c_{ij}(\omega)$	Co-spectrum
$f(\omega)$	Normalized spectral density function
$h(\omega)$	Non-normalized integrated spectrum
$h(\omega)$	Non-normalized spectral density function
$h_{ij}(\omega)$	Non-normalized cross spectrum
$q_{ij}(\omega)$	Quadrature-spectrum
ACF	Auto-Correlation Function
ACVF	Auto-Covariance Function

AR(k)	Autoregressive model of order k
CCF	Cross-Correlation Function
CCVF	Cross-Covariance Function
DAC	Dynamic Atmospheric Correction
FFT	Finite Fourier Transform
PCA	Principal Component Analysis
SC1	Simulation Case 1
SC2	Simulation Case 2
SLA	Sea Level Anomaly

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof Keith R. Thompson, for his guidance and encouragement from the beginning. Besides my supervisor, I would like to thank my committee members, Prof. Michael Dowd, Prof. Joanna Mills Flemming and Prof. David Hamilton for their kind comments and advice on my thesis. I would also like to thank Prof. Emmanuel Devred for his help with this research.

Last but not least, I would like to thank my families for their support and love. My thanks also goes to my professors and colleagues in Dalhousie University. It has been a wonderful year for me because of you all.

Chapter 1

Introduction

The stability of the Earth's environment is a topic of great concern to society, in large part because it has a major influence on human activity. This has resulted in a significant effort by research scientists to reconstruct and understand past changes, and also project future states of the Earth system. The changes occur on both long and short time scales. The long-term changes (e.g., sea level rise, changing coastlines, advance and retreat of polar ice caps) have shaped human society throughout history. The short-term changes (e.g., hurricanes and storm surges) can have an immediate influence on human activity (e.g., marine transportation, retreat of coastal populations to higher land in the event of an impending large storm surge). The changes on both time scales are particularly important for sea level.

People have been observing the ocean for their own practical needs for thousands of years. For example, Gill (1982) notes that Aristotle spent the final a few months of his life observing tide conditions of the strait between the island of Evvoia (or Euboea) and the Greek mainland. These early observations were generally anecdotal, unorganized and limited to specific areas and periods of time (Pouliquen, 2006). The most scientifically-valuable observations of sea level over the last few centuries have been made by tide gauges. Tide gauges are distributed along the coast over the globe and they provide measurements of local sea level height at regular intervals of time. The oldest tide gauge records can be traced back 300 years in Amsterdam and the earliest records were made by human observers using tide staffs. Thanks to the installation of self-recording tide gauges over the last century, tide gauges have now become the major source of sea level information. Normally tide gauges provide hourly observations for an extended period of time (e.g., decades and longer); the newest tide gauges can provide observations with a time spacing of minutes or even seconds. Even though tide gauge observations are point measurements and thus limited in their spatial distribution, they provide excellent information through time.

All available tide gauge observations have been organized and made accessible to the general public (e.g., through University of Hawaii Sea Level Center).

The launch of TOPEX/Poseidon satellite in 1992 started a new era of observing the ocean. Along with its successors including Jason-1, 2 and 3, TOPEX/Poseidon provides global sea level observations of high accuracy. The satellites orbit around the earth and remotely sense the ocean using radar altimeters. By measuring the time delay of the radar pulse, the distance between the altimeter and the ocean surface can be obtained. Given the position of the satellite, the sea surface height can be then estimated. Satellite altimetry has revolutionized ocean research and started the era of 'big data' for global ocean measurements. The sea surface height observations from a satellite altimeter over one cycle (less than 10 days) exceeds one hundred years of tide gauge observations. These satellite measurements of sea level are not only vast (with global coverage) but also accurate (within a few centimeters). Satellite altimetry observations have been validated with various in-situ observations including tide gauge observations (e.g., Chelton et al., 2001).

Satellite altimetry also has limitations. One of the most important problems is the sampling constraint (Robinson, 2006). It normally takes approximately ten days for a satellite to complete a cycle and repeat a track. Hence the sampling interval for a given position on the Earth's surface is approximately ten days and slow enough to miss most high frequency sea level variations. This results in a severe "aliasing problem" where high frequency signals are misrepresented and can contaminate low frequency signals. Statistical description of aliasing effect is found in Chapter 2. Figure 1.1 illustrates the effect with an extreme example of aliasing: the hourly Halifax tide gauge record is sub-sampled at the satellite altimetry sampling rate, leading to the illusion of a 60-day period in sea level variability.

In order to reduce the aliasing problem, signals of frequency higher than the Nyquist frequency (the highest frequency that can be captured at a given sampling rate) should be removed from the observations. There is a large body of literature on the de-aliasing of satellite altimetry. The state-of-the-art methods are based on dynamical models that provide corrections for tides and storm surges. Tides are the dominant high frequency signals in sea level variations. Driven by astronomical forces, they are highly deterministic and periodic, and tide predictions are mostly of

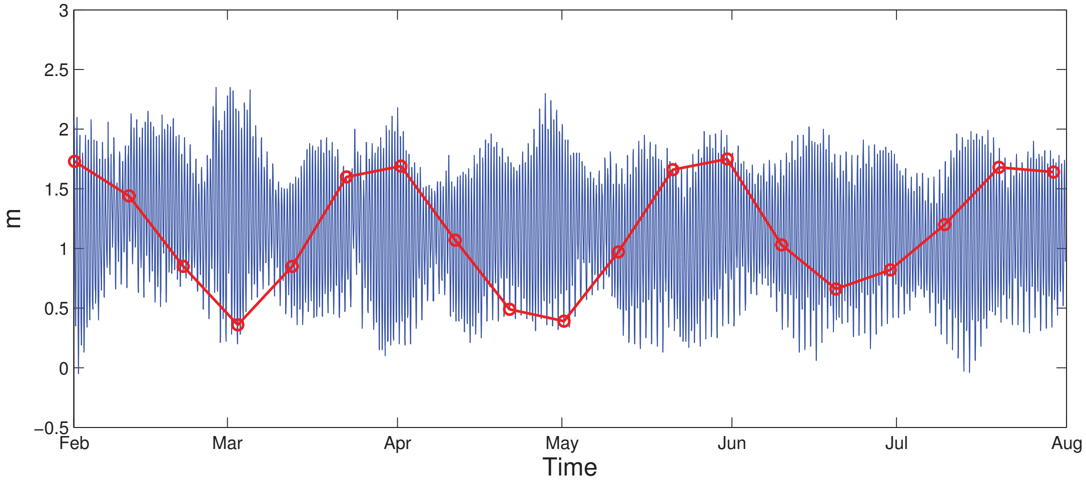

Figure 1.1: The effect of aliasing illustrated using a tidally dominated sea level record. The blue line shows hourly sea level observed by the Halifax tide gauge for six months in 2010. The red dots are the hourly values sub-sampled with a period of 10 days (similar to the repeat time of a satellite altimeter). The aliasing of the hourly record leads to an apparent oscillation (an alias) with a period of about 60 days.

high accuracy (e.g., FES2012 model of Carrère et al. (2012)). Surges also contribute to major high frequency components in the sea level variations, especially over continental shelves. Dynamically-based surge predictions from Carrère and Lyard (2003) are used widely to reduce the high frequency sea level variability and this correction is usually referred to as the dynamic atmospheric correction. The standard deviation of the sea level after the tide and surge corrections is typically reduced from meters to about ten centimeters. However, if these corrections do not account for all the variations, the aliasing problem is still a concern. This leads to the question: are there high frequency signals left after all the dynamical tide and surge corrections have been made and, if so, can we improve the corrections and further reduce the aliasing effect?

In the present study, the hourly tide gauge records are used to help answer this question. An analysis of several tide gauge records along the east coast of North America and Canada indicates that there are significant, spatially coherent high frequency signals remaining after the tidal and dynamic atmospheric corrections have been made. In order to de-alias these high frequency signals that could be misrepresented by satellite altimetry, a statistical model is designed. Unlike the deterministic corrections given by ocean models, the statistical model extracts information on the aliased signals with the help of the tide gauge records and then provides an alias correction for the satellite altimetry observations. The statistical method provides improved de-aliasing results, but as importantly, motivates ocean research into the causes of these aliased signals.

The statistical model assumes a linear relationship between the tide gauge measurements and the satellite altimeter measurements. In this study, three methods are developed: a lag domain method, a frequency domain method and one based on the lasso. The lag and the frequency domain methods (Priestley, 1981) estimate the parameters on different domains but are similar in principle. The lasso method (Hastie et al., 2009) treats the problem as a simple regression problem; it is easy to implement and also effective. With a fitted model, the high frequency signals aliased by the altimeter can be predicted from the tide gauge observations within a certain number of lags. The predictions for the aliased high frequency signals are used as corrections to further reduce the sea level variability.

The structure of the thesis is as follows: Chapter 2 provides the statistical background of the study. Chapter 3 describes the aliasing problem mathematically and introduces the three statistically based methods to solve the problem. Two simulation cases are presented to demonstrate the implementation of the methods and to test their performance. The tide gauge and altimetry data are introduced and their initial processing is described in Chapter 4. Chapter 5 provides an analysis of the sea level anomalies and motivates the de-aliasing study. The de-aliasing of the sea level anomalies is conducted and the results are presented in Chapter 6. Finally, Chapter 7 summarizes the thesis and discusses topics for future research.

Chapter 2

Statistical Background

In this chapter, the statistical background for the study is provided. First of all, I discuss the spectral representation of a stationary random process and some spectral estimation methods. Then, I discuss linear transformations of a stationary random process and related filtering techniques. The aliasing effect is discussed and two examples are provided.

2.1 Preliminaries

This section summarizes important theoretical aspect of the spectral analysis of stationary random processes. The theory was initially built on univariate continuous parameter processes and then generalized to the discrete case and bivariate processes.

2.1.1 Spectral Representation of Stationary Processes

The spectral representation theorem (Priestley, 1981, p. 246) is the basis for spectral analysis and is described in Theorem 1.

Theorem 1. *Let* $\{X(t)\}$, $-\infty < t < \infty$, *be a zero-mean stochastic continuous stationary process. Then there exists an orthogonal process,* $\{Z(\omega)\}$, *such that, for all* t, $X(t)$ *may be written in the form,*

$$X(t) = \int_{-\infty}^{\infty} e^{it\omega} \mathrm{d}Z(\omega) \tag{2.1}$$

the integral being defined in the mean-squared sense. The process $\{Z(\omega)\}$ *has the following properties: (i)* $\mathrm{E}\left[\mathrm{d}Z(\omega)\right] = 0$, *all* ω *(ii)* $\mathrm{E}\left[|\mathrm{d}Z(\omega)|^2\right] = \mathrm{d}H(\omega)$, *all* ω, *where* $H(\omega)$ *is the (non-normalized) integrated spectrum of* $X(t)$ *(iii) for any two distinct frequencies* $\omega \neq \omega'$,

$$\mathrm{cov}\left\{\mathrm{d}Z(\omega), \mathrm{d}Z(\omega')\right\} = \mathrm{E}\left[\mathrm{d}Z(\omega) \cdot \mathrm{d}Z(\omega')\right] = 0$$

Note that the zero-mean process $X(t)$ is stationary to second order if $\operatorname{cov}\{X(t), X(s)\}$ is a function of $(t-s)$ only. Theorem 1 provides a canonical form for the spectral decomposition (or transformation) of any stationary processes, and is the basis for spectral analysis. The theorem carries over to discrete parameter process as with a limited frequency range from $-\pi$ to π. The reason for the frequency range restriction is that any frequency outside this range has an 'alias' in the range (the aliasing effect is discussed in Section 2.4).

If $\{X(t)\}$ has a purely continuous spectrum, then

$$\mathrm{d}H(\omega) = h(\omega)\mathrm{d}\omega$$

$h(\omega)$ is known as the spectral density function of $\{X(t)\}$, or more simply, the spectrum of $\{X(t)\}$.

2.1.2 Spectrum and Auto-covariance Function

Suppose $\{X(t)\}$ is a zero-mean continuous parameter stationary process with spectral density function, $h(\omega)$, and auto-covariance function (ACVF) defined by

$$R(\tau) = \operatorname{cov}\{X(t), X(t+\tau)\} = \mathrm{E}\left[X(t)X(t+\tau)\right] \tag{2.2}$$

where τ is the lag. Then $h(\omega)$ is simply the Fourier transform of $R(\tau)$, and $R(\tau)$ is the inverse Fourier transform of $h(\omega)$ (Priestley, 1981, p.211):

$$h(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}R(\tau)\mathrm{d}\tau \tag{2.3}$$

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}h(\omega)\mathrm{d}\omega \tag{2.4}$$

Note that the spectral density function $h(\omega)$ exists for all ω if $R(\tau)$ is absolutely integrable, i.e.,

$$\int_{-\infty}^{\infty} |R(\tau)|\,\mathrm{d}\tau < \infty$$

In practice this limits how slowly the 'tails' of $R(\tau)$ tend to zero with increasing lag τ.

It follows that $h(\omega)$ and $R(\tau)$ form a Fourier transform pair. They provide equivalent information of the process $\{X(t)\}$: one in the "frequency domain" and the other in the “lag domain”. Equations (2.4) and (2.3) are crucial in moving from one to the other.

This leads to the normalized spectral density function of $\{X(t)\}$ and the auto-correlation function (ACF) of $\{X(t)\}$. Let σ_X^2 denote the variance of $\{X(t)\}$ and divide both sides of (2.3) and (2.4) by σ_X^2. We obtain

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}\rho(\tau)\mathrm{d}\tau \tag{2.5}$$

$$\rho(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} f(\omega)\mathrm{d}\omega \tag{2.6}$$

Clearly the normalized spectrum and auto-correlation function also form a Fourier transform pair.

Equations (2.3)-(2.6) give the spectral representations of the auto-covariance and auto-correlation functions. The spectrum is a decomposition of variance into different frequency components. Specifically, take lag $\tau = 0$ in (2.4) and we have,

$$\operatorname{var}\{X_t\} = \int_{-\infty}^{\infty} h(\omega)\mathrm{d}\omega$$

In addition, it is obvious that $h(\omega) \geq 0$ and $h(\omega) = h(-\omega)$. The spectrum of a process describes how the variance (often in physics referred as power or energy) is distributed over different frequency ranges and, in this sense, is analogous to a probability distribution function.

2.1.3 Bivariate Processes

Theorem 1 can be carried over to bivariate processes. Equations (2.3)-(2.6) have natural generalizations correspondingly where the auto-spectrum and auto-covariance functions are replaced by cross-spectrum and cross-covariance functions.

Cross-covariance Function: Suppose $\mathbf{X}_t = [X_{1,t}, X_{2,t}]^T$ is a discrete parameter zero-mean stationary bivariate process. Bivariate stationarity to second order means $\{X_{1,t}\}$ and $\{X_{2,t}\}$ are jointly stationary:

1. $\{X_{1,t}\}$ and $\{X_{2,t}\}$ are each univariate stationary, and

2. $\operatorname{cov}\{X_{1,t}, X_{2,s}\}$ is a function of $s - t$ only.

Then we may define the auto-covariance function (ACVF) in the usual way as in (2.2), namely,

$$\begin{aligned} R_{11}(s) &= \mathrm{E}\left[X_{1,t}X_{1,t+s}\right] \\ R_{22}(s) &= \mathrm{E}\left[X_{2,t}X_{2,t+s}\right] \end{aligned} \tag{2.7}$$

The ACVF describes the correlation *within* each process. It is natural to generalize it to a cross-covariance function (CCVF) to describe the correlation *between* the two processes. The CCVF between $\{X_{1,t}\}$ and $\{X_{2,t}\}$ is defined by

$$R_{21}(s) = \mathrm{cov}\left\{X_{1,t}, X_{2,t+s}\right\} = \mathrm{E}\left[X_{1,t}X_{2,t+s}\right] \tag{2.8}$$

The cross-correlation function (CCF) is then given by

$$\rho_{21}(s) = \frac{R_{21}(s)}{\sqrt{R_{11}(0)R_{22}(0)}} \tag{2.9}$$

Unlike auto-covariance and correlation functions, the cross-covariance and correlation functions are normally not symmetric.

Cross-spectrum: For the bivariate process $\mathbf{X}_t = \left[X_{1,t}, X_{2,t}\right]^T$, suppose $\{X_{1,t}\}$ and $\{X_{2,t}\}$ have purely continuous spectra with spectral density functions $h_{11}(\omega)$ and $h_{22}(\omega)$:

$$\begin{aligned} h_{11}(\omega) &= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} R_{11}(s)e^{-is\omega} \\ h_{22}(\omega) &= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} R_{22}(s)e^{-is\omega} \end{aligned} \tag{2.10}$$

Then the (non-normalized) cross spectral density function (cross-spectrum) exists for all ω provided that $\sum_s |R_{21}(s)| < \infty$. It is defined (Priestley, 1981, p.655) as

$$h_{21}(\omega) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty} R_{21}(s)e^{-is\omega} \tag{2.11}$$

Generalizing the spectral representation of Theorem 1 it can be shown

$$h_{21}(\omega)\mathrm{d}\omega = E[\mathrm{d}Z_1^*(\omega)\mathrm{d}Z_2(\omega)] \tag{2.12}$$

where * denotes complex conjugation. Note that $h_{21}(\omega)$ is usually complex-valued and Hermitian: $h_{21}(\omega) = h_{21}^*(-\omega)$. Cross-spectrum is a natural generalization of auto-spectrum and describes how the cross-covariance between two processes is distributed over different frequencies.

Co- and Quadrature spectra: Write $h_{21}(\omega)$ in terms of its real and imaginary parts:

$$h_{21}(\omega) = c_{21}(\omega) - iq_{21}(\omega) \tag{2.13}$$

$c_{21}(\omega)$ is called the co-spectrum and $q_{21}(\omega)$ is called the quadrature spectrum.(Priestley, 1981, p.659)

For real valued processes,

$$\begin{aligned} c_{21}(\omega) &= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\frac{1}{2}\{R_{21}(s) + R_{21}(-s)\}\cos s\omega \\ q_{21}(\omega) &= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\frac{1}{2}\{R_{21}(s) - R_{21}(-s)\}\sin s\omega \end{aligned} \tag{2.14}$$

In this case it is obvious that $c_{21}(\omega)$ is even function whereas $q_{21}(\omega)$ is odd function and $h_{21}(0) = c_{21}(0)$, $q_{21}(0) = 0$.

Gain and Phase: Write $h_{21}(\omega)$ in polar form:

$$h_{21}(\omega) = \alpha_{21}(\omega)e^{i\phi_{21}(\omega)} \tag{2.15}$$

$\alpha_{21}(\omega)$ is called the cross-amplitude spectrum or gain spectrum, and $\phi_{21}(\omega)$ is the phase spectrum.(Priestley, 1981, p.660)

The gain and phase spectra can lead to useful physical interpretations of the cross-spectrum and are frequently used in practice. At a given frequency ω, the gain spectrum $\alpha_{21}(\omega)$ measures how $\{X_{2,t}\}$ is attenuated or amplified relative to $\{X_{1,t}\}$ while the phase spectrum measures the phase shift of $\{X_{2,t}\}$ relative to $\{X_{1,t}\}$.

It is also easy to show that

$$\begin{aligned} \alpha_{21}(\omega) &= |h_{21}(\omega)| = \sqrt{c_{21}^2(\omega) + q_{21}^2(\omega)} \\ \phi_{21}(\omega) &= \tan^{-1}\{-q_{21}(\omega)/c_{21}(\omega)\} \end{aligned} \tag{2.16}$$

and it follows that the gain spectrum is also an even function and the phase spectrum is an odd function.

Coherence: Complex coherency at frequency ω is defined by

$$w_{21}(\omega) = \frac{h_{21}(\omega)}{\sqrt{h_{11}(\omega)\cdot h_{22}(\omega)}} \tag{2.17}$$

$|w_{21}(\omega)|$ is known as coherency spectrum. It is easy to show (Priestley, 1981, p.661) that $0 \leq |w_{21}(\omega)| \leq 1$.

From (2.12), an alternative representation of coherency is

$$w_{21}(\omega) = \frac{\mathrm{cov}\left\{\mathrm{d}Z_1(\omega), \mathrm{d}Z_2(\omega)\right\}}{\sqrt{\mathrm{var}\left\{\mathrm{d}Z_1(\omega)\right\} \cdot \mathrm{var}\left\{\mathrm{d}Z_2(\omega)\right\}}} \tag{2.18}$$

By analogy with the correlation between two variables, $w_{21}(\omega)$ is the 'correlation' coefficient between the random coefficients of the components in $X_{1,t}$ and $X_{2,t}$ at frequency ω.

2.2 Covariance and Spectrum Estimation

In this section we discuss briefly the estimation of covariance functions and spectral density functions from sample time series. Specifically, suppose we have a realization of a zero-mean stationary process $\{X_t\}$, where $t = 1, 2, \cdots, N$, and N is the length of the sample time series.

2.2.1 Sample Covariance Functions

Estimates of the auto-covariance function of a zero-mean random process $\{X_t\}$ are known as the sample auto-covariance function (sample ACVF) and are given by

$$\hat{R}(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} X_t X_{t+|k|} \tag{2.19}$$

It is easy to show that the bias of the estimator is $O(\frac{1}{N})$ and the variance is also $O(\frac{1}{N})$. So $\hat{R}(k)$ is an asymptotically unbiased estimator for $R(k)$ and it is also consistent.

Sample cross-covariance function between zero-mean random processes $\{X_{1,t}\}$ and $\{X_{2,t}\}$ is similarly defined:

$$\hat{R}_{21}(k) = \frac{1}{N} \sum_{t} X_{1,t} X_{2,t+k} \tag{2.20}$$

where the summation goes from $t = 1$ to $(N - k)$ for $k \geq 0$, and from $t = (1 - k)$ to N for $k < 0$.

2.2.2 Periodogram

Suppose $\{X_t\}$ has a purely continuous spectrum and the observed time series contains N data points. Given that the spectrum is the Fourier transform of the auto-covariance function (ACVF), a natural estimate for the spectrum is to plug in the sample ACVF in (2.3):

$$I_N^*(\omega) = \frac{1}{2\pi} \sum_{k=-(N-1))}^{N-1} \hat{R}(k) e^{-i\omega k} \tag{2.21}$$

$I_N^*(\omega)$ is known as the (raw) periodogram. In practice the periodogram can be calculated through a finite Fourier transform (FFT). The FFT of the observations is defined by

$$\zeta_X(\omega) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} X_t e^{-i\omega t}, \quad -\pi \le \omega \le \pi \tag{2.22}$$

Hence the periodogram can be written

$$I_N^*(\omega) = |\zeta_X(\omega)|^2 = \zeta_X(\omega)\zeta_X^*(\omega) \tag{2.23}$$

where * denotes the complex conjugation.

The periodogram $I_N^*(\omega)$ is the natural estimate of the spectrum $h(\omega)$ and it is unbiased. However, the periodogram is usually not a satisfactory estimate in that it is not consistent (Priestley, 1981, p.420-429).

If we substitute in (2.21) and (2.22) the spectral representation $X_t = \int_{-\pi}^{\pi} e^{it\theta} \mathrm{d}Z(\theta)$ and then take expectations of $I_N^*(\omega)$, then we have,

$$\mathrm{E}\left[I_N^*(\omega)\right] = \int_{-\pi}^{\pi} F_N(\theta - \omega) h(\theta) \mathrm{d}\theta \tag{2.24}$$

where $F(\theta)$ is known as a Fejer kernel:

$$F_N(\theta) = \frac{1}{2\pi} \cdot \frac{\sin^2(N\theta/2)}{N \sin^2(\theta/2)} \tag{2.25}$$

Equation (2.24), however, is the form that inspires applying a spectral window (or smoothing kernel) in the estimation process.

2.2.3 Consistent Estimators and Spectral Window

One reason for the inconsistency of the periodogram is that $I_N^*(\omega)$ uses $\hat{R}(k)$ with very large k. When k gets large, the estimates for $\hat{R}(k)$ deteriorates as it is using a decreasing portion of the sample time series. In this respect, we can consider "controlling" the tails of $\hat{R}(k)$ by applying a lag window when estimating the spectrum. A formal representation (Priestley, 1981, p.432) is as follows:

$$\hat{h}(\omega) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{(N-1)} \lambda(k)\hat{R}(k)e^{-ik\omega} \tag{2.26}$$

where $\{\lambda(k)\}$ is known as the lag window. $\lambda(k)$ is designed to control the amount of information used from high lags. Some common windows are introduced in the next section. In fact, (2.26) gives a weighted version of the periodogram as shown below.

Inverting (2.21) gives,

$$\hat{R}(k) = \int_{-\pi}^{\pi} I_N^*(\theta)e^{ik\theta}\mathrm{d}\theta \tag{2.27}$$

and substituting into (2.26) gives,

$$\hat{h}(\omega) = \int_{-\pi}^{\pi} I_N^*(\theta)W(\omega-\theta)\mathrm{d}\theta \tag{2.28}$$

where $W(\omega-\theta)$ is the Fourier transform of the lag window, known as the spectral window:

$$W(\omega-\theta) = \frac{1}{2\pi} \sum_{k=-(N-1)}^{(N-1)} \lambda(k)e^{-ik\omega} \tag{2.29}$$

Equation (2.28) gives a smoothed version of the periodogram with the smoothing kernel $W(\omega-\theta)$. This smoothing technique is often used in function fitting or estimation. Moreover, same techniques can be carried over to cross-spectral estimation.

With a fixed spectral window, the smoothed periodogram is an asymptotically unbiased estimator for the spectrum. Compared to the raw periodogram, the smoothed periodogram sacrifices bias to achieve consistency.

2.2.4 Choice of Window

THE 'TRUNCATED PERIODOGRAM' WINDOW The simplest window is the 'truncated periodogram' window. It uses $2M+1$ lags to estimate the spectrum with uniform weights. The lag window is defined by

$$\lambda(k) = \begin{cases} 1, & |k| \leq M, \\ 0, & |k| > M \end{cases} \tag{2.30}$$

where $M < N-1$ is the truncation point in (2.26). The corresponding spectral window is given by

$$W(\theta) = \frac{1}{2\pi} \sum_{k=-M}^{M} \cos k\theta = D_M(\theta) \tag{2.31}$$

The function $D_M(\theta)$ is known as 'Dirichilet kernel'.

THE DANIELL WINDOW: The Daniell window is also known as the 'rectangular' window. The 'truncated periodogram' window is a rectangular window on the lag domain, while the Daniel window is a rectangular window on the frequency domain. The spectral window is defined by

$$W(\theta) = \begin{cases} M/2\pi, & |\theta| \leq \pi/M, \\ 0, & \text{otherwise} \end{cases} \tag{2.32}$$

with the corresponding lag window

$$\lambda(k) = \frac{M}{2\pi} \int_{-\pi/M}^{\pi/M} e^{ik\theta} \mathrm{d}\theta = \frac{\sin \pi k/M}{\pi k/M} \tag{2.33}$$

Note that the lag window does not vanish after a certain lag as the 'truncated periodogram' window, so there is no truncation point.

In the frequency domain, the Daniell window gives an average at each frequency ω within the window $-\pi/M \leq \omega \leq \pi/M$. M is called 'window parameter', or 'width' of the window. Smaller M means greater smoothing. The M in the 'truncated periodogram' window plays a similar role.

THE PARZEN WINDOW: The Parzen lag window, suggested by Parzen (1961), takes the following form:

$$\lambda(k) = \begin{cases} 1 - 6(k/M)^2 + 6(|k|/M)^3, & |k| \leq M/2, \\ 2(1 - |k|/M)^3, & M/2 \leq |k| \leq M \\ 0, & |k| > M \end{cases} \tag{2.34}$$

with the corresponding spectral window

$$W(\theta) = \frac{3}{8\pi M^3} \left(\frac{\sin M\theta/4}{\frac{1}{2}\sin\theta/2} \right)^4 (1 - \frac{2}{3}\sin^2\theta/2) \tag{2.35}$$

The Parzen window always produces non-negative estimates of the spectral density function (Priestley, 1981, p.444). M is the truncation point. Figure 2.1 is the general form for the Parzen lag window.

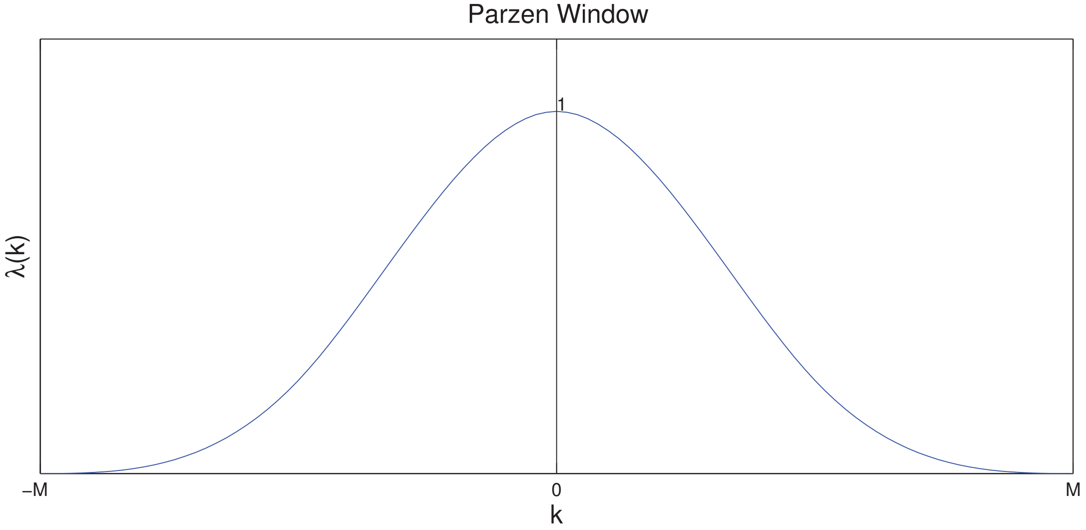

Figure 2.1: The general form for the Parzen lag window. M is the truncation point of the lag window in the sense that $\lambda(k) = 0$ outside the window. Within the window, the weight $\lambda(k)$ is 1 at $k = 0$ and decreases as $|k|$ increases.

2.3 Linear Transformation and Filters

In this section, the linear transformation of a process and the associated transfer function are introduced. The transfer function is essential in representing the linear relationship in the frequency domain. Following the discussion of linear transformations, some filtering techniques are described.

2.3.1 Transfer Function

Suppose a stationary process $\{X_t\}$ (consider discrete case for simplicity) admits the spectral representation

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} \mathrm{d}Z_X(\omega)$$

Now assume Y_t is a linear transformation of X_t with weights $\sum_k |a_k| < \infty$ thereby ensures the stationarity of Y_t. In signal processing, this gives a simple linear system where $\{X_t\}$ is often referred as an input and $\{Y_t\}$ as output:

$$Y_t = \sum_{k=-\infty}^{\infty} a_k X_{t-k} \tag{2.36}$$

It is easy to show that the spectral representation of Y_t is given by

$$Y_t = \int_{-\pi}^{\pi} \Gamma(\omega) e^{i\omega t} \mathrm{d}Z_X(\omega) \tag{2.37}$$

where $\Gamma(\omega)$ is called the transfer function. For the linear transformation defined by (2.36), the transfer function is the Fourier transform of the linear coefficients:

$$\Gamma(\omega) = \sum_{k=-\infty}^{\infty} a_k e^{-i\omega k} \tag{2.38}$$

Equation (2.37) also gives the connection between the two corresponding orthogonal processes by the transfer function:

$$\mathrm{d}Z_Y(\omega) = \Gamma(\omega)\mathrm{d}Z_X(\omega) \tag{2.39}$$

If we multiply both sides of (2.39) by $\mathrm{d}Z_X^*(\omega)$ and then take expectations, we have the connection between cross-spectrum of Y_t and X_t and auto-spectrum of X_t:

$$h_{XY}(\omega) = \Gamma(\omega)h_{XX}(\omega) \tag{2.40}$$

Suppose $h_{XX}(\omega) \neq 0$ for all ω and write (2.40) in form

$$\Gamma(\omega) = h_{XY}(\omega)/h_{XX}(\omega) \tag{2.41}$$

then the transfer function has the form of a frequency dependent regression coefficient.

Furthermore, the transfer function relates the spectrum of Y_t to X_t with

$$h_{YY}(\omega) = |\Gamma(\omega)|^2 \, h_{XX}(\omega) \tag{2.42}$$

This form describes the effect of a linear filter. Usually Y_t is referred to as a filtered version of X_t. By choosing different weights $\{a_k\}$, or equivalently the transfer functions $\Gamma(\omega)$, we can design different filters as desired. The most frequently used filters such as band-pass filter, low-pass filter and high-pass filter are introduced below.

2.3.2 Filters

HIGH-PASS FILTER: If we choose $\{a_k\}$ such that

$$|\Gamma(\omega)|^2 = \begin{cases} 1, & |\omega| > \omega_0, \\ 0, & |\omega| \leq \omega_0 \end{cases} \tag{2.43}$$

we can get a high-pass filter. High-pass filter passes frequency components above ω_0 and removes frequency components below ω_0. ω_0 is referred as the *cut-off* frequency.

Low-pass Filter: Similarly a low pass filter can be obtained by choosing $\{a_k\}$ such that

$$|\Gamma(\omega)|^2 = \begin{cases} 1, & |\omega| \leq \omega_0, \\ 0, & |\omega| > \omega_0 \end{cases} \tag{2.44}$$

By way of contrast to the high-pass filter, the low pass filter passes frequency components below ω_0 and removes frequency components above the cut-off frequency ω_0.

Band-pass Filter: If we choose $\{a_k\}$ such that

$$|\Gamma(\omega)|^2 = \begin{cases} 1, & \omega_1 \leq |\omega| \leq \omega_2, \\ 0, & \text{otherwise} \end{cases} \tag{2.45}$$

then we have a band-pass filter. In between high-pass filter and low-pass filter, a band-pass filter passes frequency components between ω_1 and ω_2 and suppresses frequency components outside this frequency band.

2.3.3 Orthogonality of Frequency Components

The spectral representation of a stationary process, $\{X_t\}$, transforms the process into orthogonal frequency components, $\{\mathrm{d}Z_X(\omega)\}$. Suppose $\{X_t\}$ is decomposed into low and high frequency components with a high-pass or low-pass filter with cut-off frequency ω_0,

$$X_t = X_t^L + X_t^H$$

The spectral representations of the two components are,

$$X_t^L = \int_{|\omega| \leq \omega_0} e^{it\omega} \mathrm{d}Z(\omega)$$

and

$$X_t^H = \int_{|\omega| > \omega_0} e^{it\omega} \mathrm{d}Z(\omega)$$

Since there is no overlapping frequency, the two parts are statistically uncorrelated.

Suppose Y_t is a linear transformation of X_t and a same high-pass or low-pass filter gives the low and high frequency components:

$$Y_t = Y_t^L + Y_t^H$$

Then the spectral representation of the two components are

$$Y_t^L = \int_{|\omega| \leq \omega_0} \Gamma(\omega) e^{it\omega} \mathrm{d}Z(\omega)$$

and

$$Y_t^H = \int_{|\omega| > \omega_0} \Gamma(\omega) e^{it\omega} \mathrm{d}Z(\omega)$$

$\Gamma(\omega)$ is the transfer function and is given by the Fourier transform of the linear coefficients as in (2.38). Since the same cut-off frequency is used in filtering of X_t and Y_t, there is no overlapping frequency between X_t^L and Y_t^H, nor between Y_t^L and X_t^H. As a result, both X_t^L and Y_t^L are uncorrelated with X_t^H and Y_t^H.

A consequence of the orthogonality of the frequency components is the separation of the cross-covariance function into two components. Given X^H is independent of Y^L, then the cross-covariance function between them, $\mathrm{E}\left[X_t^H Y_{t+k}^L\right] = 0$. Hence,

$$\mathrm{E}\left[X_t^H Y_{t+k}\right] = \mathrm{E}\left[X_t^H Y_{t+k}^L + X_t^H Y_{t+k}^H\right] = \mathrm{E}\left[X_t^H Y_{t+k}^H\right] \tag{2.46}$$

This result indicates that the cross-covariance between X^H and Y only depends on the high frequency component, namely the cross-covariance between X^H and Y^H. This result is crucial in our de-aliasing method discussed later.

2.4 The Aliasing Effect

In this section, the aliasing effect is described in detail. The aliasing theorem is the fundamental theorem for the aliasing effect.

2.4.1 The Aliasing Theorem

Sampling from a continuous process $\{X(t)\}$ at discrete time intervals Δ (or under-sampling from a discrete process $\{X_t\}$) will lead to some loss of information on its spectral properties. This is known as the aliasing effect. A mathematical statement of the aliasing effect (Priestley, 1981, p. 506) is as follows:

Theorem 2 (Aliasing theorem)**.** *Suppose* $\{X(t)\}$ *admits a spectral representation of the form*

$$X(t) = \int_{-\infty}^{\infty} e^{it\omega} \mathrm{d}Z_X(\omega) \tag{2.47}$$

Define the sampled process $Y_t \equiv X(t\cdot\Delta)$ *for* $t = 0, \pm 1, \ldots$. *The spectral representation of* Y_t *is then given by*

$$Y_t = \int_{-\pi/\Delta}^{\pi/\Delta} e^{it\omega\cdot\Delta} \mathrm{d}Z_Y(\omega) \tag{2.48}$$

where

$$\mathrm{d}Z_Y(\omega) = \sum_{k=-\infty}^{\infty} \mathrm{d}Z_X\left(\omega + \frac{2k\pi}{\Delta}\right) \tag{2.49}$$

Equation (2.48) indicates that the spectral representation of Y_t only extends over the frequency range, $(-\pi/\Delta, \pi/\Delta)$. π/Δ is called the Nyquist frequency of the sampled time series and the sample contains no information about variation at frequencies higher than Nyquist frequency. The components in $X(t)$ with frequencies $(\omega+2k\pi/\Delta)$ will all appear to have frequency ω. These frequencies are said to be *aliases* of ω. Note that the Nyquist period is 2Δ and signals with periods smaller than the Nyquist period are aliased.

If $X(t)$ has a (non-normalized) spectral density function $h_X(\omega)$, then the spectral density function of Y_t is defined by

$$h_Y(\omega) = \sum_{k=-\infty}^{\infty} h_X\left(\omega + \frac{2k\pi}{\Delta}\right), \quad |\omega| \leq \frac{\pi}{\Delta} \tag{2.50}$$

Equation (2.50) is a fundamental result of Theorem 2. From the equation we see that the value of $h_Y(\cdot)$ is a superposition of the values of $h_X(\cdot)$ at all alias frequencies $\{\omega, \omega \pm 2\pi/\Delta, \omega \pm 4\pi/\Delta, \omega \pm 6\pi/\Delta, \cdots\}$.

From a graphical view, the construction of $h_Y(\omega)$ can be seen as first folding back $h_X(\omega)$ over $(-\infty, \infty)$ at lines on $\{\omega, \omega \pm 2\pi/\Delta, \omega \pm 4\pi/\Delta, \omega \pm 6\pi/\Delta, \cdots\}$ and then superimposing all the folded portions additively. The Nyquist frequency, π/Δ is also called folding frequency for this reason.

2.4.2 Transformation of Frequency Range

We could usually treat the sampled process $\{Y_t\}$ as observations from a discrete parameter process with spectral density function $h_Y^*(\omega)$ defined on $(-\pi, \pi)$, and we can easily convert it back to $h_Y(\omega)$ with the original frequency range $(-\pi/\Delta, \pi/\Delta)$ (Priestley, 1981, p. 508).

$$h_Y(\omega) = \Delta \cdot h_Y^*(\omega \cdot \Delta), \quad -\pi/\Delta \leq \omega \leq \pi/\Delta \tag{2.51}$$

Most commonly used frequency ranges include $(-\pi, \pi)$ and $(-\frac{1}{2}, \frac{1}{2})$. In practice, the physical period is defined as $f = 2\pi/\omega$ on the range $(-\pi, \pi)$, while the frequency range $(-\frac{1}{2}, \frac{1}{2})$ makes it easy for conversion from frequency to period simply by inversion: $f = 1/\omega$.

2.4.3 Aliasing Theorem for Bivariate Processes

Suppose a bivariate process $\mathbf{X}_t = [X_{1,t}, X_{2,t}]^T$ is under-sampled as $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^T$ where $\{Y_{1,t}\}$ and $\{Y_{2,t}\}$ are sub-sampled at the same rate, say at a sampling interval of Δ. The aliasing theorem for bivariate processes is a straightforward extension of the univariate case. The spectra matrix of $\mathbf{Y}_t$ is a folded version of the spectra matrix of $\mathbf{X}_t$. This is obvious for the auto-spectra of $\{Y_{1,t}\}$ and $\{Y_{2,t}\}$ so we only provide a simple proof for the folding of the cross-spectrum below.

First of all, from Theorem 2 we have,

$$\begin{aligned} \mathrm{d}Z_{Y_1}(\omega) &= \sum_{k=-\infty}^{\infty} \mathrm{d}Z_{X_1}\left(\omega + \frac{2k\pi}{\Delta}\right) \\ \mathrm{d}Z_{Y_2}(\omega) &= \sum_{k=-\infty}^{\infty} \mathrm{d}Z_{X_2}\left(\omega + \frac{2k\pi}{\Delta}\right) \end{aligned} \tag{2.52}$$

Substituting (2.52) into (2.12) gives

$$\mathrm{d}H_{Y_2Y_1}(\omega) = \mathrm{E}\left[\sum_{k=-\infty}^{\infty} \mathrm{d}Z^*_{X_1}\left(\omega + \frac{2k\pi}{\Delta}\right) \cdot \sum_{k=-\infty}^{\infty} \mathrm{d}Z_{X_2}\left(\omega + \frac{2k\pi}{\Delta}\right)\right] \tag{2.53}$$

Theorem 1 gives that for any $\omega_1 \neq \omega_2, -\pi < \omega_1, \omega_2 < \pi$,

$$\mathrm{E}\left[\mathrm{d}Z^*_{X_1}(\omega_1) \cdot \mathrm{d}Z_{X_2}(\omega_2)\right] = 0 \tag{2.54}$$

It follows from (2.54) that for any $-\pi/\Delta < \omega < \pi/\Delta$ and any $k_1 \neq k_2$,

$$E\left[\mathrm{d}Z^*_{X_1}\left(\omega + \frac{2k_1\pi}{\Delta}\right) \cdot \mathrm{d}Z_{X_2}\left(\omega + \frac{2k_2\pi}{\Delta}\right)\right] = 0 \tag{2.55}$$

Hence all cross terms in (2.53) are zero and the right side reduces to:

$$\begin{aligned}
\mathrm{d}H_{Y_2Y_1}(\omega) &= E\left[\sum_{k_1=-\infty}^{\infty} \mathrm{d}Z^*_{X_1}\left(\omega + \frac{2k_1\pi}{\Delta}\right) \cdot \sum_{k_2=-\infty}^{\infty} \mathrm{d}Z_{X_2}\left(\omega + \frac{2k_2\pi}{\Delta}\right)\right] \\
&= E\left[\sum_{k_1=-\infty}^{\infty} \sum_{k_2=k_1} \mathrm{d}Z^*_{X_1}\left(\omega + \frac{2k_1\pi}{\Delta}\right) \cdot \mathrm{d}Z_{X_2}\left(\omega + \frac{2k_2\pi}{\Delta}\right)\right] \\
&= E\left[\sum_{k=-\infty}^{\infty} \mathrm{d}Z^*_{X_1}\left(\omega + \frac{2k\pi}{\Delta}\right) \cdot \mathrm{d}Z_{X_2}\left(\omega + \frac{2k\pi}{\Delta}\right)\right] \\
&= \sum_{k=-\infty}^{\infty} E\left[\mathrm{d}Z^*_{X_1}\left(\omega + \frac{2k\pi}{\Delta}\right) \cdot \mathrm{d}Z_{X_2}\left(\omega + \frac{2k\pi}{\Delta}\right)\right] \\
&= \sum_{k=-\infty}^{\infty} \mathrm{d}H_{X_2X_1}(\omega + \frac{2k\pi}{\Delta})
\end{aligned}$$

The cross-spectrum of the aliased process $\mathbf{Y}_t$ is then given by

$$h_{Y_2Y_1}(\omega) = \sum_{k=-\infty}^{\infty} h_{X_2X_1}\left(\omega + \frac{2k\pi}{\Delta}\right), \quad -\frac{\pi}{\Delta} \le \omega \le \frac{\pi}{\Delta} \tag{2.56}$$

Note (2.56) has a similar form to (2.50) for the univariate case.

Writing in terms of co- and quadrature spectra gives

$$h_{Y_2Y_1}(\omega) = \sum_{k=-\infty}^{\infty} c_{X_2X_1}\left(\omega + \frac{2k\pi}{\Delta}\right) - i\sum_{k=-\infty}^{\infty} q_{X_2X_1}\left(\omega + \frac{2k\pi}{\Delta}\right) \tag{2.57}$$

The co-spectrum and quadrature-spectrum are then given by

$$\begin{aligned}
c_{Y_2Y_1}(\omega) &= \sum_{k=-\infty}^{\infty} c_{X_2X_1}\left(\omega + \frac{2k\pi}{\Delta}\right) \\
q_{Y_2Y_1}(\omega) &= \sum_{k=-\infty}^{\infty} q_{X_2X_1}\left(\omega + \frac{2k\pi}{\Delta}\right)
\end{aligned} \tag{2.58}$$

The real and imaginary part of the cross-spectrum both conform to the folding theorem. This provides a graphic way to show the folding of cross-spectrum, which is often complex valued.

2.5 Examples of the Aliasing Effect

Two simple examples are given in this section to illustrate the aliasing effect. One is an univariate AR(1) process and another is a bivariate process known as linear regression with decay.

2.5.1 Discrete Parameter AR(1) Process

Consider a discrete parameter AR(1) process $\{X_t\}$:

$$X_t = aX_{t-1} + \epsilon_t \tag{2.59}$$

where $\{\epsilon_t\}$ is a purely random process (white noise) with $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$. It is easy to show that the normalized spectrum of $\{X_t\}$ is given by

$$f_X(\omega) = \frac{1 - a^2}{2\pi(1 - 2a\cos\omega + a^2)}, \quad -\pi \leq \omega \leq \pi \tag{2.60}$$

Suppose $\{Y_t\}$ is a sub-sample from $\{X_t\}$ with a sampling interval $\Delta = 4$. Hence $\pi/4$ is the Nyquist frequency of $\{Y_t\}$. From (2.50) the normalized spectrum of $\{Y_t\}$ is only defined on $-\pi/4 \leq \omega \leq \pi/4$ and is given by

$$\begin{aligned} f_Y(\omega) &= \sum_{k=-\infty}^{\infty} f_X\left(\omega + \frac{2k\pi}{\Delta}\right) \\ &= f_X(\omega) + f_X(\omega - \frac{\pi}{2}) + f_X(\omega + \frac{\pi}{2}) + f_X(\omega - \pi) \\ &= \frac{2(1 - a^8)}{\pi(1 - 2a^4\cos 4\omega + a^8)}, \quad -\pi/4 \leq \omega \leq \pi/4 \end{aligned} \tag{2.61}$$

Transforming the frequency scale using (2.51) gives the spectrum on the $-\pi \leq \omega \leq \pi$ scale. $f_Y^*(\omega)$ takes exactly the same form as $f_X(\omega)$ except that the parameter, a, is replaced by its fourth power and frequency, ω, is scaled by 1/4:

$$\begin{aligned} f_Y^*(\omega) &= \Delta \cdot f_Y(\omega \cdot \Delta) \\ &= \frac{(1 - a^8)}{2\pi(1 - 2a^4\cos\omega + a^8)}, \quad -\pi \leq \omega \leq \pi \end{aligned} \tag{2.62}$$

To show the folding of the spectrum, we take $a = 0.9$ and then plot and compare the normalized spectra of the two processes in Figure 2.5.1. The AR(1) process $\{X_t\}$ does not contain significant high frequency signals, so the aliasing effect is negligible.

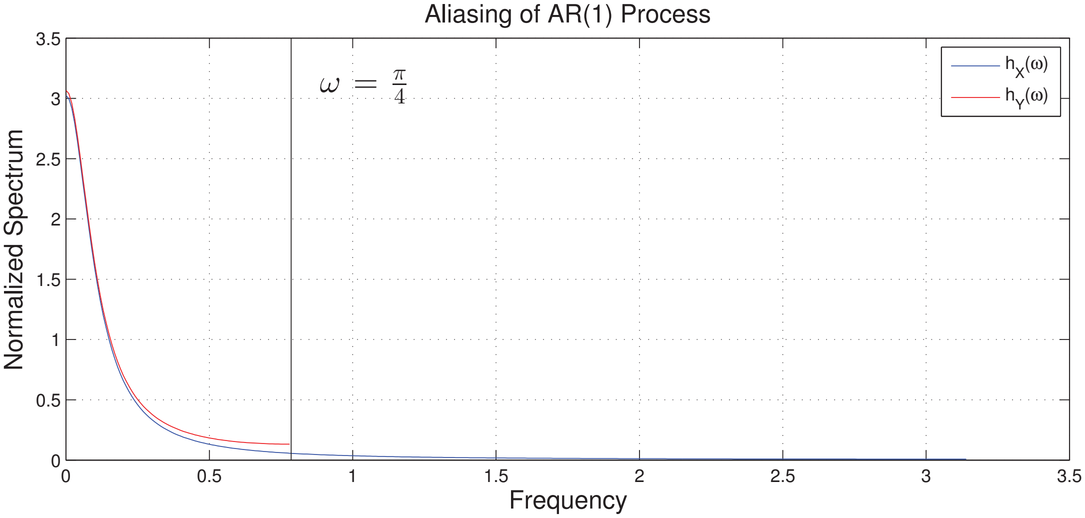

Figure 2.2: Power spectra of $\{X_t\}$ and $\{Y_t\}$ in the discrete parameter AR(1) process example. $\{X_t\}$ is an AR(1) process with $a = 0.9$ and $\{Y_t\}$ is sub-sampled from $\{X_t\}$ at sampling interval $\Delta = 4$. Nyquist frequency of $\{Y_t\}$ is $\pi/4$. The plot shows that the spectrum of $\{Y_t\}$ is a superposition of the folded spectrum of $\{X_t\}$. However, since the AR(1) process $\{X_t\}$ does not contain significant high frequency signals, the aliasing effect is negligible.

2.5.2 Linear Regression with Decay

Assume $\mathbf{X}_t = [X_{1,t}, X_{2,t}]^T$ is a bivariate stationary process. Let $\{X_{2,t}\}$ be a discrete parameter AR(1) process. Suppose $X_{1,t}$ and $X_{2,t}$ satisfy a linear regression relationship with a delay of d time units

$$\begin{aligned} X_{1,t} &= aX_{2,t-d} + \epsilon_{1,t} \\ X_{2,t} &= bX_{2,t-1} + \epsilon_{2,t} \end{aligned} \tag{2.63}$$

where $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are two independent white noise processes with $\epsilon_{1,t} \sim \mathcal{N}(0, \sigma_1^2)$ and $\epsilon_{2,t} \sim \mathcal{N}(0, \sigma_2^2)$.

It is straightforward to show that $R_{12}(k) = aR_{22}(k-d)$. Then with (2.61) from the last example, the cross-spectrum of $X_{1,t}$ and $X_{2,t}$ is given by

$$\begin{aligned} h_{X_1X_2}(\omega) &= \frac{a}{2\pi}\sum_{k=-\infty}^{\infty} R_{22}(k-d)e^{-i\omega k} \\ &= ae^{-i\omega d}h_{22}(\omega) \\ &= \frac{a\sigma_2^2}{2\pi} \cdot \frac{e^{-i\omega d}}{1 - 2b\cos\omega + b^2} \end{aligned} \tag{2.64}$$

Now suppose $\mathbf{Y}_t = [Y_{1,t}, Y_{2,t}]^T$ is a sub-sample of $\mathbf{X}_t$ at sampling interval Δ. From (2.56), the cross-spectrum between $Y_{1,t}$ and $Y_{2,t}$ is given by

$$\begin{aligned} h_{Y_1Y_2}(\omega) &= \sum_{k=-\infty}^{\infty} h_{X_1X_2}\left(\omega + \frac{2k\pi}{\Delta}\right) \\ &= \frac{a\sigma_2^2}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\exp\left(-i\left(\omega + \frac{2k\pi}{\Delta}\right)d\right)}{1 - 2b\cos\left(\omega + \frac{2k\pi}{\Delta}\right) + b^2} \\ &= \frac{a\sigma_2^2}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\cos\left(\left(\omega + \frac{2k\pi}{\Delta}\right)d\right) - i\sin\left(\left(\omega + \frac{2k\pi}{\Delta}\right)d\right)}{1 - 2b\cos\left(\omega + \frac{2k\pi}{\Delta}\right) + b^2} \end{aligned} \tag{2.65}$$

The co-spectrum and quadrature-spectrum are then given by

$$\begin{aligned} c_{Y_1Y_2}(\omega) &= \frac{a\sigma_2^2}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\cos\left(\left(\omega + \frac{2k\pi}{\Delta}\right)d\right)}{1 - 2b\cos\left(\omega + \frac{2k\pi}{\Delta}\right) + b^2} \\ q_{Y_1Y_2}(\omega) &= \frac{a\sigma_2^2}{2\pi} \sum_{k=-\infty}^{\infty} \frac{\sin\left(\left(\omega + \frac{2k\pi}{\Delta}\right)d\right)}{1 - 2b\cos\left(\omega + \frac{2k\pi}{\Delta}\right) + b^2} \end{aligned} \tag{2.66}$$

Assume $a = 0.9$, $b = 0.8$, $d = 2$ and $\Delta = 3$ to generate an $\mathbf{X}_t$ and its sub-sampled process $\mathbf{Y}_t$. To illustrate the folding of the cross-spectrum, the co- and quadrature-spectra of $\mathbf{X}_t$ and their aliased forms of $\mathbf{Y}_t$ are plotted in Figure 2.3.

2.6 Principal Component Analysis

The main objective of Principal Component Analysis (PCA) is to explain the variance-covariance structure with a set of linear combinations of the original variables, or principal components (Johnson and Wichern, 2007, p. 430). The principal components usually provide better interpretation and can be selected to reduce dimensions. This section gives a basic introduction of principal component analysis. The principal components are defined first and then the decomposition of the covariance matrix are explained. At last, the frequency dependent principal component analysis is introduced for this study.

2.6.1 Principal Components

Suppose the random vector $\mathbf{X} = [X_1, X_2, \cdots, X_p]^T$ is made up of a set of p random variables. Denote the covariance matrix by $\mathbf{\Sigma}$ and assume its eigenvalues

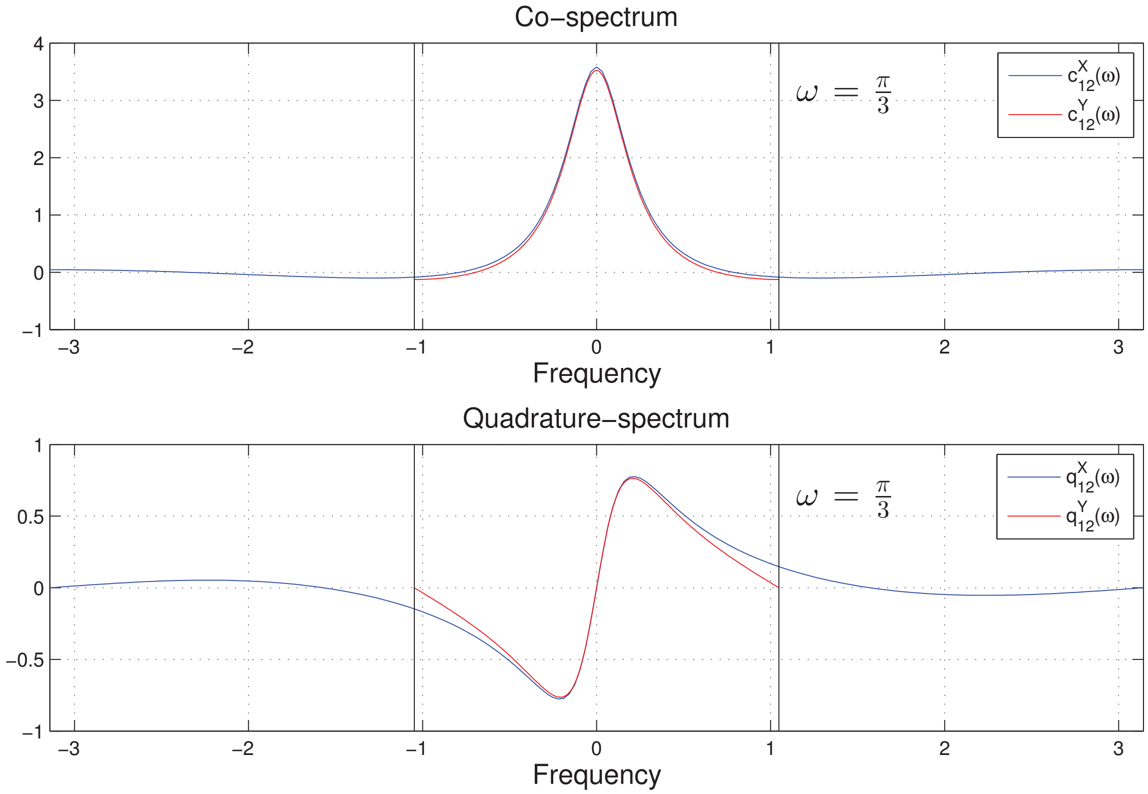

Figure 2.3: The co- and quadrature spectra of $\mathbf{X}_t$ and its aliased form $\mathbf{Y}_t$ in the linear regression with decay example. The co- (quadrature) spectra of $\mathbf{X}_t$ and $\mathbf{Y}_t$ are shown in the upper (lower) panel. Similar to Figure 2.5.1 for the univariate AR(1) example, the two plots show the superimposition of the spectra. Note that the co-spectra are even functions and the quadrature spectra are odd functions.

are $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$. Suppose $\mathbf{Y} = [Y_1, Y_2, \cdots, Y_p]^T$ is a new set of random variables, where each Y_k $(k = 1, 2, \cdots, p)$ is a linear combination of the original random variables:

$$Y_k = \mathbf{a}_k^T \mathbf{X} = a_{k1}X_1 + a_{k2}X_2 + \cdots + a_{kp}X_p \quad k = 1, 2, \cdots, p$$

This defines a linear transformation from $\mathbf{X}$ to $\mathbf{Y}$. It is easy to show that the variance and covariance structure of $\mathbf{Y}$ can be represented as follows,

$$\begin{aligned} \text{var}\{Y_k\} &= \mathbf{a}_k^T \mathbf{\Sigma} \mathbf{a}_k, \quad k = 1, 2, \cdots, p \\ \text{cov}\{Y_k, Y_j\} &= \mathbf{a}_k^T \mathbf{\Sigma} \mathbf{a}_j, \quad k, j = 1, 2, \cdots, p \end{aligned} \tag{2.67}$$

Based on the results above, the principal components are defined as the set of uncorrelated linear combinations with the following properties:

1. The first principal component $Y_1 = \mathbf{a}_1^T\mathbf{X}$ maximizes the variance of Y_1 subject to $\mathbf{a}_1^T\mathbf{a}_1 = 1$, or equivalently, maximizes $\dfrac{\mathbf{a}_1^T\mathbf{\Sigma}\mathbf{a}_1}{\mathbf{a}_1^T\mathbf{a}_1}$;

2. For $k = 2, 3, \cdots, p$, the kth principal component $Y_k = \mathbf{a}_k^T\mathbf{X}$ is uncorrelated with $Y_1, Y_2, \cdots, Y_{k-1}$ and maximizes the variance of Y_k subject to $\mathbf{a}_k^T\mathbf{a}_k = 1$. Equivalently, the kth principal component maximizes $\dfrac{\mathbf{a}_k^T\mathbf{\Sigma}\mathbf{a}_k}{\mathbf{a}_k^T\mathbf{a}_k}$ subject to $\operatorname{cov}\{Y_k, Y_j\} = 0$, for $j < k$.

In this way, the principal component analysis decomposes the variance of the original variables. This is explained in the following section.

2.6.2 Variance Decomposition

Suppose the eigen-decomposition of the covariance matrix $\mathbf{\Sigma}$ gives the eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e_1}), (\lambda_2, \mathbf{e_2}), \cdots, (\lambda_p, \mathbf{e_p})$, where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$. It can be shown that the kth principal component is given by $Y_k = \mathbf{e}_k^T\mathbf{X}$ and

$$\begin{aligned} &\operatorname{var}\{Y_k\} = \lambda_k, \quad k = 1, 2, \cdots, p \\ &\operatorname{cov}\{Y_k, Y_j\} = 0, \quad k \neq j \end{aligned} \tag{2.68}$$

If some λ_k are equal, then the choice of Y_k is not unique (Johnson and Wichern, 2007).

It is also easy to show that the total variance of the principal components is equal to the total variance of the original variables:

$$\sum_{k=1}^{p} \operatorname{var}\{X_k\} = \sum_{k=1}^{p} \operatorname{var}\{Y_k\} = \operatorname{tr}(\mathbf{\Sigma}) = \lambda_1 + \lambda_2 + \cdots + \lambda_p \tag{2.69}$$

The principal components provide a summary of the variance structure of the original variables. This is very useful in practice because principal components not only provide a direct interpretation of the sample variability of an observed data set but are also linked with the original variables. Principal component analysis is widely used in many areas to identify patterns of the data. In addition, if the original data are standardized before analysis, the principal components provides a summary of the correlation structure instead.

2.6.3 Frequency Dependent Principal Component Analysis

In order to identify patterns within several time series, principal component analysis (PCA) can be extended to the frequency domain. Various extensions exist including complex principal component analysis (CPCA, see Horel (1984)) and frequency dependent principal component analysis. In this study, frequency dependent principal component analysis is used.

Suppose we have N time series, each containing T observations: $x_k(t)$, $k = 1, 2, \cdots, N$ and $t = 1, \cdots, T$. Assume that the auto-spectrum of $x_k(t)$ is given by $h_{kk}(\omega)$ and the cross-spectrum between $x_k(t)$ and $x_j(t)$ is given by $h_{kj}(\omega)$. (The spectra can be normalized in order to emphasize the coherence-phase relationships among the time series.) At each frequency ω, this leads to an $N \times N$ matrix, denoted by $\tilde{H}(\omega)$. Note that $\tilde{H}(\omega)$ is Hermitian since $h_{kj}(\omega)$ is the conjugate of $h_{jk}(\omega)$.

Now if we apply PCA to $\tilde{H}(\omega)$, then we'll have a set of eigenvalue-eigenvector pairs $(\lambda_k(\omega), \mathbf{e}_k(\omega))$ $(k = 1, 2, \cdots, N)$ which are also defined at each frequency ω. Even though the elements of the spectral matrix are normally complex-valued, the PCA will give a set of real-valued eigenvalues and complex-valued eigenvectors (Petersen and Pedersen, 2012). At each frequency a PCA process is applied, so the variance decomposition equations (2.68)-(2.69) hold for every frequency with extension to an Hermitian matrix. As a result, the total variance accounted by the kth principal component is $\sum_{j=1}^{k} \int_{-\pi}^{\pi} \lambda_j(\omega) \mathrm{d}\omega / \sum_{j=1}^{N} \int_{-\pi}^{\pi} \lambda_j(\omega) \mathrm{d}\omega$. Note that in a discretized form, it is equivalent to $\sum_{j=1}^{k} \overline{\lambda}_j / \sum_{j=1}^{N} \overline{\lambda}_j$ where $\overline{\lambda}_j$ is the average of the kth principal component over the frequency range.

Compared to PCA in the time domain, except from the similar variance decomposition result above, the frequency dependent PCA can actually provide more frequency-specific information. For example, the first eigenvalue $\lambda_1(\omega)$ as a function of ω, is similar to a spectrum where the peaks indicate specific frequency ranges where the variance can be represented largely by one or a few principal component(s). This is useful in identifying certain frequencies where a simplified pattern of multiple time series lies. Moreover, the eigenvector usually preserves dynamic information of this pattern. This is illustrated in the analysis study in Chapter 5, where the eigenvalue and eigenvector of the first principal component is used to interpret important frequency signals.

Chapter 3

De-aliasing Methods

This chapter starts with an overview of the de-aliasing problem from a mathematical perspective and outlines of three possible de-aliasing methods. This is followed by a more detailed description of the de-aliasing methods and their evaluation using simulated data from idealized test cases.

3.1 Overview

Suppose we have two related zero-mean stationary processes. The sample time series associated with one of the processes, X say, contains N_1 data points on the grid $G_1 : \{t = 1, 2, \cdots, N_1\}$. The sample series from the other process, Y say, is under-sampled at an interval Δ and thus only contains N_2 valid data points defined on a coarser grid $G_2 : \{t = \Delta, 2\Delta, \cdots, N_2\Delta\}$. The two grids are illustrated in Figure 3.1. Let $X|G_1$ denote the sample time series on the fine grid: $\{X_t, t \in G_1\}$ and let $Y|G_2$ denote the sample time series of the related process on the coarse grid, $\{Y_t, t \in G_2\}$. Let $Y|G_1$ denote the second process defined on the fine grid: $\{Y_t, t \in G_1\}$. In general observations of $Y|G_1$ will not be available. The Nyquist frequency of $Y|G_2$ is π/Δ and we may anticipate that $Y|G_2$ will suffer from aliasing (see Section 2.4).

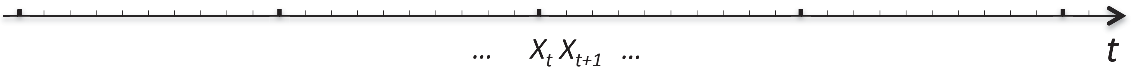

Figure 3.1: Spacing of the sample time series of the two related processes on the time axis. X is defined on G_1, the fine grid marked with thin tick marks. Y, corresponding to the sub-sampled process, is defined on G_2 with a grid spacing of Δ as indicated by the thick tick marks. In this case $\Delta = 10$.

For each process, assume a frequency decomposition that separates it into low (superscript L) and high frequency (superscript H) components:

$$\begin{aligned} X &= X^L + X^H \\ Y &= Y^L + Y^H \end{aligned} \tag{3.1}$$

where both X^L and Y^L are uncorrelated with X^H and Y^H (see Section 2.3.3). In practice, the frequency decomposition is accomplished by a high-pass filter with cut-off frequency that is set by the Nyquist frequency of G_2, π/Δ. In order to reduce the aliasing affecting $Y|G_2$, we consider removing the high frequency component $Y^H|G_2$ that is potentially aliased. Since $Y^H|G_2$ is a sub-sample of $Y^H|G_1$, the de-aliasing problem turns into the prediction of $Y^H|G_1$ given observations of $X|G_1$ and $Y|G_2$.

It is normally impossible to extract the aliased high frequency component from $Y|G_2$ alone. However, in some situations it may be possible to use information from $X|G_1$ which is sampled at a higher rate and can therefore provide high frequency information directly. Since the processes are assumed related, the high frequency information of X can be used to estimate the aliased high frequency information in Y. For simplicity, we assume a linear relationship between the two processes and write

$$Y_t = \sum_{u=-p}^{p} a_u X_{t-u} + \tilde{\epsilon}_t \tag{3.2}$$

where $\{\tilde{\epsilon}_t\}$ is a white noise process uncorrelated with the X for all time lags. We can rewrite the model by decomposing X into high and low frequency components as follows:

$$Y_t = \sum_{u=-p}^{p} a_u X^H_{t-u} + \epsilon_t \tag{3.3}$$

where $\epsilon_t = \sum\limits_{u=-p}^{p} a_u X^L_{t-u} + \tilde{\epsilon}_t$. Thus $\{\epsilon_t\}$ now includes both the low frequency information of X and the white noise term. It follows that $\{\epsilon_t\}$ is uncorrelated with X^H for all time lags. Model (3.3) is a general linear model with autocorrelated errors. The parameters in the model are denoted by $\theta = [\mathbf{a}, p]$, where $\mathbf{a} = [a_{-p}, \cdots, a_p]$ and p is the order of the model (system).

Model (3.3) can be considered as a regression model, where $X^H_{t-u}, \quad u = -p, -p+1, \cdots, p$ are the predictors and Y_t is the response. This means that the lasso method (Hastie et al., 2009, p. 68) can be used to estimate the linear coefficients. The main attraction of the lasso method is that it contains a penalty term (the L^1 norm of the regression coefficients) that is used to simplify the model. It also has a well developed cross-validation procedure that can be used to estimate p.

Two other de-aliasing methods can be developed based on model fitting in the lag and frequency domains. The lag domain method leads to a system of Yule-Walker equations which can be solved to estimate the linear coefficients. The lag domain method can also be transformed to the frequency domain, where the linear coefficients are estimated via a frequency-dependent transfer function. Both methods are conceptually similar but there are also differences as will be explained below.

Both the lag and frequency domain methods involve calculation of the sample auto-covariance function of X^H, and the sample cross-covariance function between X^H and Y^H. The high-pass filter already gives $X^H|G_1$ from $X|G_1$, so the sample auto-covariance function comes directly from $X^H|G_1$. The sample cross-covariance between X^H and Y^H is equivalent to the sample cross-covariance between X^H and Y because X^H and Y^L are uncorrelated. The latter, fortunately, can be readily estimated from the data available for arbitrary lags: even though data for the two time series, $X^H|G_1$ and $Y|G_2$, are not on the same grid, sample cross-covariance at arbitrary lag k can still be computed by shifting $X^H|G_1$ at each lag k, and then using pairs of common data points from the two time series.

All three methods (lasso, lag domain and spectral domain) lead to estimates of the parameter vector θ. Prediction of $Y^H|G_1$ follows from the fitted model,

$$\hat{Y}_t^H = \sum_{u=-p}^{p} \hat{a}_u X_{t-u}^H, \quad t \in G_1 \tag{3.4}$$

where $\hat{a}_u$ denotes the estimate of a_u. The aliased high frequency component in $Y|G_2$ can be estimated by sub-sampling the predictions $\hat{Y}^H|G_1$ to give $\hat{Y}^H|G_2$. Let Y^c denote the de-aliased time series of $Y|G_2$, then we have

$$Y_t^c = Y_t - \hat{Y}_t^H, \quad t \in G_2 \tag{3.5}$$

3.2 De-aliasing Methods

The three de-aliasing methods are now described in more details. We start with the lag domain and frequency domain methods. Although conceptually related, the implementation of these two methods leads to differences. The lasso method and its implementation are then described.

3.2.1 Lag Domain Method

The estimation of the sample auto- and cross-covariance functions is described first. (This is the basis for both the lag and frequency domain methods.) The Yule-Walker equations are then constructed followed by more information on parameter estimation.

Sample Auto- and Cross-covariance Functions: The auto-covariance function (ACVF) and cross-covariance function (CCVF) are fundamental in both the lag domain and frequency domain methods. Natural estimates of the ACVF and CCVF are the sample ACVF and the sample CCVF. For stationary zero-mean processes, the sample ACVF and CCVF at lag k are defined by (2.19) and (2.20) respectively.

In the present situation, the sample time series $X^H|G_1$ contains N_1 points and the sample ACVF of X^H is given by

$$\hat{R}_{X^H X^H}(k) = \frac{1}{N_1} \sum_{t=1}^{N_1-|k|} X_t^H \cdot X_{t+|k|}^H, \quad k = 0, \pm 1, \cdots, \pm(N_1 - 1) \tag{3.6}$$

The sample CCVF between X^H and Y^H cannot be estimated is a similar way for two reasons. First, observations of Y^H are not readily available; we only have sample time series $X^H|G_1$ and $Y|G_2$. Second, the two sample time series are not on the same grid.

To solve the first problem, Section 2.3.3 shows that the CCVF between X^H and Y is equivalent to the CCVF between X^H and Y^H because the low and high frequency components are uncorrelated. As a result, we can estimate the CCVF between X^H and Y^H by the sample CCVF between X^H and Y.

To solve the second problem, we can first shift $X^H|G_1$ by arbitrary lag k and then find the pairwise correlation between the shifted sample time series (as illustrated in Figure 3.2). This means that the CCVF can estimated for lags, $k = 0, \pm 1, \cdots$. Mathematically the sample CCVF is given by

$$\hat{R}_{YX^H}(k) = \frac{1}{N_2} \sum_{t} X_{t\cdot\Delta-k}^H \cdot Y_{t\cdot\Delta}, \quad k = 0, \pm 1, \cdots, \pm(N_2 - 1) \tag{3.7}$$

where the summation goes from $t = 1$ to $(N_2 - k)$, $k \geq 0$, and from $t = (1 - k)$ to N_2, $k < 0$.

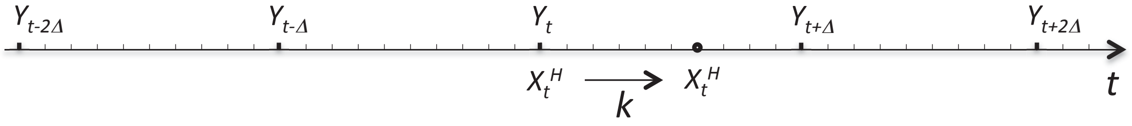

Figure 3.2: Calculation of the sample CCVF at lag k. First we shift the time series $X|G_2$ by k where X is originally defined on G_1 (thin tick marks). Then we find on the grid G_1 (thick tick marks) the times when both time series give valid data. We then use these pairwise data points to estimate the sample CCVF at lag k.

Yule-Walker Equations: Multiplying both sides of (3.3) by X_{t-k}^H and taking expectations gives

$$\mathrm{E}\left[Y_t^H X_{t-k}^H\right] = \mathrm{E}\left[Y_t X_{t-k}^H\right] = \sum_{u=-p}^{p} a_u \mathrm{E}\left[X_{t-u}^H X_{t-k}^H\right]$$

This leads to the following set of equations defined in terms of ACVF and CCVF:

$$R_{Y^H X^H}(k) = \sum_{u=-p}^{p} a_u R_{X^H X^H}(k-u), \quad k = -q, -q+1, \cdots, q-1, q \tag{3.8}$$

where q is flexible and normally $q \geq p$. Note the CCVF is a convolution of the predictor ACVF with the sequence of a_u coefficients. This leads to a system given by

$$\mathbf{r} = \mathbf{R}\mathbf{a} \tag{3.9}$$

where $\mathbf{r}$ is a $(2q+1) \times 1$ vector, $\mathbf{R}$ is a $(2q+1) \times (2p+1)$ matrix, and $\mathbf{a}$ is a $(2p+1) \times 1$ vector:

$$\mathbf{R} = \begin{bmatrix} R_{X^H X^H}(p-q) & R_{X^H X^H}(p-q-1) & \cdots & R_{X^H X^H}(-p-q) \\ R_{X^H X^H}(p-q+1) & R_{X^H X^H}(p-q) & \cdots & R_{X^H X^H}(-p-q+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_{X^H X^H}(p+q) & R_{X^H X^H}(p+q-1) & \cdots & R_{X^H X^H}(-p+q) \end{bmatrix}$$

$$\mathbf{a} = \left[a_{-p}, a_{-p+1}, \cdots, a_p\right]^T$$

$$\mathbf{r} = \left[R_{Y^H X^H}(-q), R_{Y^H X^H}(-q+1), \cdots, R_{Y^H X^H}(q)\right]^T$$

The form of (3.9) is similar to the traditional Yule-Walker equations (e.g., Shumway and Stoffer, 2010, p.121) but it involves two processes. The left side of the equation is replaced by the cross-covariances.

It is easy to see that $R_{Y^H X^H}(k)$ for k from $-q$ to q, and $R_{X^H X^H}(k)$ for k from $-p-q$ to $p+q$, are used in the estimation process. This means that up to $\pm(p+q)$ lags are involved in the estimation process.

Parameter Estimation: Substituting the sample ACVF and CCVF into (3.9) leads to estimates for $\mathbf{R}$ and $\mathbf{r}$, denoted by $\hat{\mathbf{R}}$ and $\hat{\mathbf{r}}$. Then the parameter estimation for model (3.3) is given by the solution of the above Yule-Walker like equations.

Note $\theta = [\mathbf{a}; p]$ includes two parts, the linear coefficients $\mathbf{a}$, and the order, p. First consider the estimation of $\mathbf{a}$ for a known p. The Yule-Waler equations usually take $q = p$ and the coefficient estimates are then given by the unique solution (assuming $\hat{\mathbf{R}}$ nonsingular)

$$\mathbf{a} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{r}}$$

Improved estimates can be obtained by including more terms of $\hat{R}_{Y X^H}(k)$ and $\hat{R}_{X^H X^H}(k)$. This leads to the 'modified Yule-Walker equations' with $q > p$. This forms an overdetermined system whose solution is usually given by regression. q is selected to be reasonably large to include enough information. Friedlander and Porat (1984) provides an overview of parameter estimation based on the modified Yule-Walker equations.

Since high lags often contain less information, the weighted least square solution (Friedlander and Porat, 1984) was adopted in this study. It is generally given by

$$\mathbf{a} = (\hat{\mathbf{R}}^T \mathbf{W} \hat{\mathbf{R}})^{-1} \hat{\mathbf{R}}^T \mathbf{W} \hat{\mathbf{r}} \tag{3.10}$$

where $\mathbf{W}$ is known as the weight matrix. For simplicity, the scalar weighting is often used, where $\mathbf{W}$ is a diagonal matrix. In this case, the covariance between estimators at different lags (Priestley, 1981, p. 326) is ignored. This gives a weight matrix with a set of weights $\{w_i, i = 1, 2, \cdots, 2q+1\}$ on the diagonal and zeros off-diagonal. The weights are usually positive. The weight sequence is often chosen to be non-increasing with $|k|$, based on the belief that higher lags contain less information and are more affected by noise. This is essentially similar to applying a lag window.

In general there is not a perfect method for determining the order of the modified Yule-Walker equations. The choice of p leads to a 'bias-variance' trade-off problem: large p increases the degrees of freedom and can cause an inflation of variance; small p can result in poor model fit and bias.

One method to estimate p is to monitor the variance of the residual series (de-aliased series). A decent model would give accurate predictions of the high frequency component and thus, after removing these aliased signals, the residual series would only include the low frequency component and the error term (white noise). The variance of the residual series is the sum of the variance of the low frequency component and the variance of the error term. Any model that doesn't remove the aliased signals adds additional variance to the residual variance, or equivalently, the optimal model is supposed to give minimum residual variance. As a result, it is natural to use cross-validation to select the optimal model that minimizes the residual variance.

3.2.2 Frequency Domain Method

Theorem 1 (Page 5) gives the spectral representation of a zero-mean stationary process $\{X_t\}$. This theorem shows that $\{\mathrm{d}_X(\omega)\}$ provides equivalent information on the process. The Fourier transform makes it easy to transform a time series from the time domain to the frequency domain, and the inverse Fourier transform to move from the frequency to time domain. The basis of the frequency domain approach to de-aliasing is to first transform (3.3) to the frequency domain where the transfer function is estimated. Inverse Fourier transforming the transfer function leads to estimates of the linear coefficients in the time domain.

Estimating the Transfer Function: The auto-spectrum of X^H can be estimated from the sample ACVF of X^H, and the cross-spectrum between X^H and Y^H can be estimated from the sample CCVF. With the sample ACVF and CCVF as defined above, the auto-spectrum and cross-spectrum are given by

$$\begin{aligned}\hat{h}_{X^H X^H}(\omega) &= \frac{1}{2\pi}\sum_{k=-M}^{M} \lambda(k)\hat{R}_{X^H X^H}(k)e^{-i\omega k}\\ \hat{h}_{Y^H X^H}(\omega) &= \frac{1}{2\pi}\sum_{k=-M}^{M} \lambda(k)\hat{R}_{Y X^H}(k)e^{-i\omega k}\end{aligned} \tag{3.11}$$

where $\{\lambda(k)\}$ is the lag window and M is the width of the lag window (see Section 2.2 for definitions).

The lag window has a similar effect as the weight matrix W in the lag domain method. W and q in the lag domain method together control the number of lags and their corresponding contributions in estimating the ACVF and CCVF. Similarly, the window $\lambda(k)$ and window width M work together for the same purpose. More specifically, the frequency domain method is conceptually the same as the lag domain method when $p = q$ where M is equivalent to $p + q$ in that M controls the number of lags involved in the estimation process, and $\lambda(k)$ is equivalent to the weights in the matrix W in that $\lambda(k)$ controls the weight or contribution at each lag. The difference is that the two methods are implemented in different domains. In practice, we will show later they give similar results using simulation cases.

Having estimated the auto- and cross-spectra, the estimated transfer function $\hat{\Gamma}(\omega)$ is then given by

$$\hat{\Gamma}(\omega) = \frac{\hat{h}_{Y^H X^H}(\omega)}{\hat{h}_{X^H X^H}(\omega)} \tag{3.12}$$

Parameter Estimation: Equation (2.37) defines the transfer function from X^H to Y^H. For the linear transformation assumed here, the transfer function can also be derived directly from the linear coefficients using (2.38). This means that we can inverse Fourier transform the estimated transfer function $\hat{\Gamma}(\omega)$ to estimate the linear coefficients for model (3.3):

$$\hat{a}_u = 2\pi \int_{-\pi}^{\pi} \hat{\Gamma}(\omega) e^{i\omega u} \mathrm{d}\omega \tag{3.13}$$

The problem left is the estimation of the model order, p. The same method as suggested for the lag domain method can be used, considering that the two methods are only different in estimating linear coefficients **a** and both methods arrive at the same point. That is, with a fixed model order, both methods give residual series and then cross-validation can be used to select the optimum model order with minimum residual variance.

However, in practice, the implementation of the cross-validation process is on M instead of p since p is inherent in M. More specifically, with a fixed M, the transfer function can be estimated and then the linear coefficients of any order can be derived.

That is to say, if we consider the true linear coefficients a_u with $|u| > |p|$ as zero, M determines the estimations of all the linear coefficients including those outside $|p|$. Nevertheless, a satisfying model would give trivial estimations for a_u with $|u| > |p|$. In this way, the choice of p is inherent in the choice of M.

In the frequency domain method, M controls the number of lags taken into account to fit the model. As the window width (Page 12 in Section 2.2.4), M also controls the 'smoothness' of the estimated spectra and the corresponding transfer function. It is obvious that if M is inappropriately small, the delay information would be lost due to the weights of the window. Whereas if M is overwhelmingly large, variance from high lags will accumulate in the model and influence the model adaptability. Similar to the choice of p in the lag domain method, the cross-validation process could be applied to select an optimum M.

3.2.3 Lasso Method

As noted above, model (3.3) can be considered as a regression problem. Y is the response and data for Y contains N_2 points: $\{Y_t, t = \Delta, 2\Delta, \cdots, N_2\Delta\}$. There are $(2p+1)$ predictors, $\{X^H_{t-u}, u = -p, -p+1, \cdots, p\}$ and data for each predictor also contains N_2 points: $\{X^H_{t-u}, t = \Delta, 2\Delta, \cdots, N_2\Delta\}$. Let $\mathbf{Y}$ denote the response data vector and $\mathbf{X}$ denote the design matrix, then we have:

$$\mathbf{Y} = \mathbf{X}\mathbf{a} + \epsilon \tag{3.14}$$

Parameter estimation also include the linear coefficients and the model order p.

The lasso method (Hastie et al., 2009, p. 68) is used in this study to estimate the parameters, by solving the following minimization problem:

$$\hat{\mathbf{a}} = \operatorname*{argmin}_{\mathbf{a}} \left\{ \frac{1}{2} \sum_{t \in G_2} (Y_t - \sum_{u=-p}^{p} a_u X^H_{t-u})^2 + \lambda \sum_{u=-p}^{p} |a_u| \right\} \tag{3.15}$$

Compared to ordinary least squares regression, the lasso method includes a penalty term on the L^1 norm of the the linear coefficients, $\lambda \sum_{u=-p}^{p} |a_u|$. The effect of this penalty term is equivalent to applying a constraint on the linear coefficients. But what makes the lasso method useful in this study is that this penalty term can actually suppress the less significant contributors to zeros and therefore select variables according to their contributions. More specifically, with a cross-validation process on λ to choose the model that gives the minimum residual variance, the linear coefficients can be selected according to their significance. In this way, only a small number of coefficients are preserved but they are the most significant ones. The lasso method gives a parsimonious model which is often more interpretable than usually least square regression models.

Note that the intercept term in (3.14) is included in ϵ so that the $\mathbf{a}$ in the lasso model (3.14) is consistent with $\{a_u, u = -p, -p+1, \cdots, p\}$ in the original model (3.3).

The lag domain and the frequency domain methods both apply windows to select the lags or predictors within the window and suppress these outside the window. The lasso method, however, select the predictors according to their contributions. All these methods add prior informations.

3.3 Simulation Case 1

In this section, a simple simulation example (referred as SC1, Simulation Case 1) is designed to describe and test the de-aliasing methods introduced above. These methods are applied to the simulation data and results are presented and discussed. Note that the frequency ranges used in the simulation cases are all from -0.5 to 0.5 (See Section 2.4.2 for transformation of frequency range). Therefore the frequencies are in terms of cycle per unit time.

3.3.1 Simulation Data

Let X_t be defined by the following AR(1) process,

$$X_t = 0.8X_{t-1} + \delta_t \tag{3.16}$$

where $\{\delta_t\}$ is a white noise process with $\delta_t \sim \mathcal{N}(0,1)$. Suppose $X|G_1$ is a realization of X. In order to satisfy the conditions at the method, I spun up the model to approximate stationarity, and then removed the mean. The sample time series of X is on the grid $G_1 : \{t = 1, 2, \cdots, 10000\}$.

Let Y be a linear transformation of X defined by the following constant delay model with noise:

$$Y_t = 2X_{t-4} + \epsilon_t \tag{3.17}$$

where $\{\epsilon_t\}$ is a white noise process independent of X with $\epsilon_t \sim \mathcal{N}(0,1)$. Corresponding to the sample $X|G_1$ we have $Y|G_1$. Sub-sampling $Y|G_1$ at $\Delta = 20$ gives the aliased series $Y|G_2$, which is on the grid $G_2 : \{t = 1, 21, 41, \cdots, 9981\}$ and contains 500 data points.

Figure 3.3 presents the simulated data. It is obvious that $Y|G_2$ aliases any signals in $Y|G_1$ with frequency above 0.05π $(= \frac{\pi}{20})$.

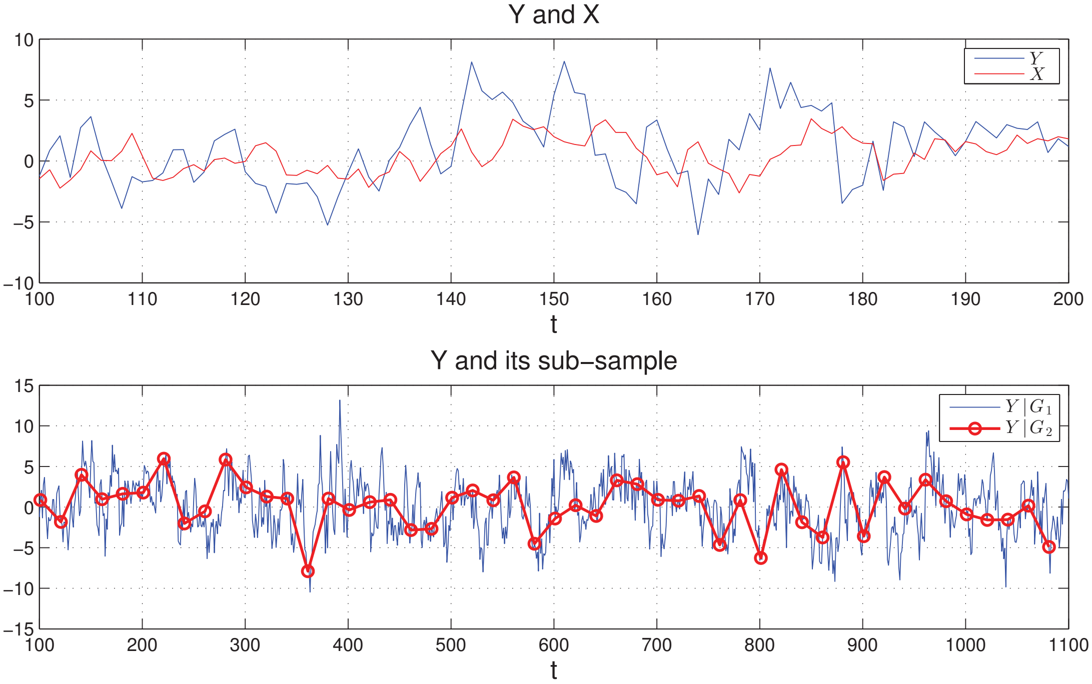

Figure 3.3: Time series plot of the simulated data in SC1. The upper panel shows $X|G_1$ and $Y|G_1$ for 100 time steps. Y is a linear transformation of X. The lower panel shows $Y|G_1$ and its sub-sampled series $Y|G_2$ for 1000 time steps. $Y|G_2$ is sub-sampled at $\Delta = 20$ and therefore aliases the high frequency signals.

3.3.2 High-pass Filter

Our goal is to de-alias $Y|G_2$ using information from $X|G_1$ based on the assumption of a linear relationship between X and Y. More specifically, if we suppose X^H is the high frequency component of X, and Y^H is the high frequency component of Y, then our goal is predict $Y^H|G_1$ using observations of $X^H|G_1$ and $Y|G_2$.

The first step is to extract $X^H|G_1$ from $X|G_1$. This is accomplished using a high-pass filter and a cut-off frequency of 0.05π. With $X^H|G_1$ and $Y|G_2$, model (3.3) can be fit using the methods described in Section 6.4 and then predictions for the aliased high frequency component can be given with (3.4). Let $\hat{Y}^H|G_2$ denote the predictions.

In addition, the same high-pass filter was applied to $Y|G_1$ to get the true aliased high frequency component. Let $Y^H|G_2$ denote the true aliased high frequency component. In this way, the predictions $\hat{Y}^H|G_2$ can be validated with $Y^H|G_2$ in order to evaluate the model. The filter results of $X|G_1$ and $Y|G_1$ are presented in Figure 3.4. Comparison of the auto-spectra of X and Y shows that the amplification is independent of the lag information.

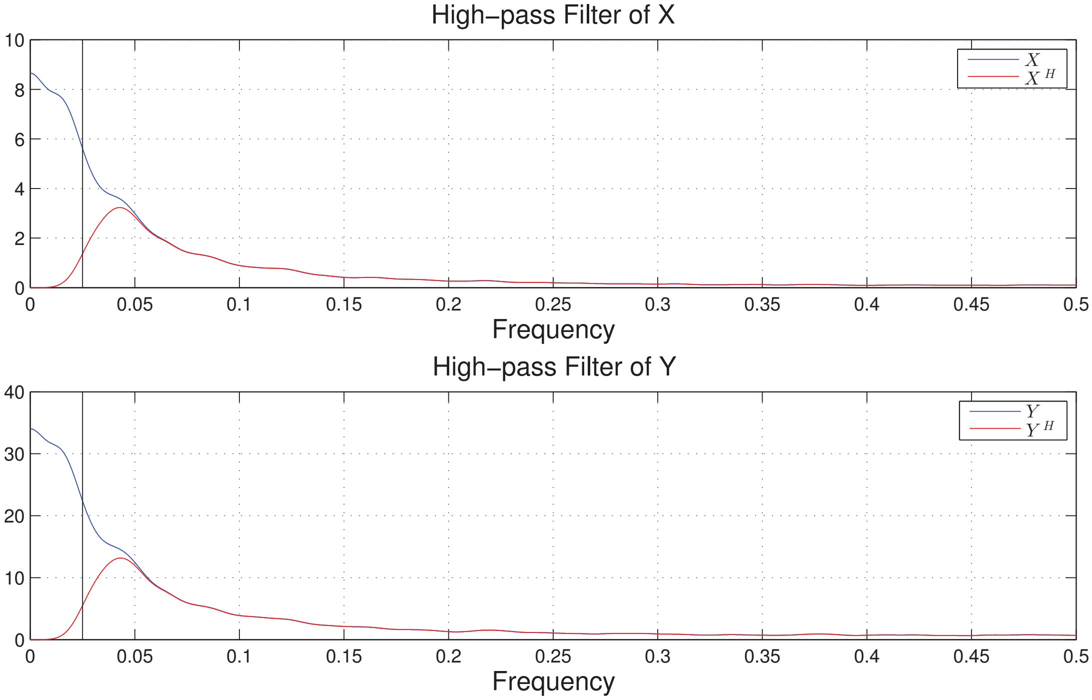

Figure 3.4: High-pass filtering of the time series in SC1. The upper panel shows the sample spectrum of X (blue line) and its high frequency component X^H (red line). The lower panel shows the sample spectrum of Y (blue line) and its high frequency component Y^H (red line). The black line marks the cut-off frequency, which is the Nyquist frequency of Y. The high-pass filter passes the high frequency component so the two spectral lines coincide above cut-off frequency. Note that the spectral slopes of X (X^H) and Y (Y^H) are identical.

Before presenting results of the two de-aliasing methods, note that in this simulation example, the true parameters in model (3.3) are given by

$$\theta = [a_{-4} = 2, a_{-3} = a_{-2} = \cdots = a_4 = 0; p = 4] \tag{3.18}$$

3.3.3 Lag Domain Method

Sample Cross-covariance Function: The sample ACVF and CCVF are estimated and plotted in Figure 3.5. The sample ACVF decays with increasing lags. The sample CCVF, though estimated from $X^H|G_1$ and $Y|G_2$, preserves most of the properties, including the decaying property of an AR(1) process and the time delay information: the peak is at lag $k = -4$, which indicates that Y^H lags X^H by four.

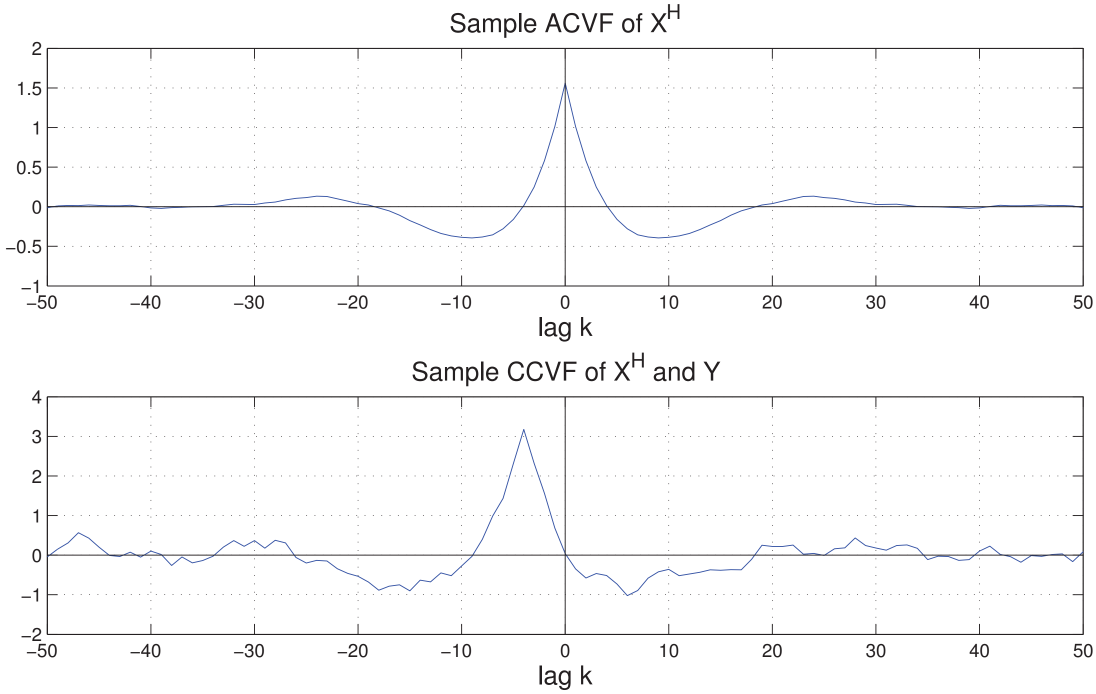

Figure 3.5: The estimations for the ACVF of X^H and the CCVF between X^H and Y in SC1. The CCVF is estimated using data of $X^H|G_1$ and $Y|G_2$ but the estimation preserves the 4-hour time delay information.

The sample ACVF and CCVF are both small after ±20 lags. Most of the information should be contained within these lags. I chose $q = 20$ for the construction of the Yule-Walker equations in (3.8).

Parameter Estimation: With a given model order p, the linear coefficients were estimated using the weighted least square method based on the modified Yule-Walker equations discussed earlier (Page 32 in Section 3.2.1). For comparison with the frequency domain method later, the weight matrix W was specified with a Parzen window weight sequence on the diagonal and zeros elsewhere. The Parzen window weight sequence can be derived from (2.34).

The choice for p is based on a 5-fold cross-validation process. Specifically for each p, the residual variance is calculated according to the following steps:

1. $Y|G_2$ is randomly split into five equal portions;
2. Linear coefficients are estimated using the training set (four of the five portions) and then used to fit model (3.3);
3. Y^H is predicted for the test set (the one portion left) using the fitted model;
4. Subtract the predicted Y^H from the Y and get the de-aliased series, which are residuals for the test set;
5. Repeat steps 2-4 for each of the five portions and we have the residual series from which the residual variance is computed.

Figure 3.6 shows the result of the cross-validation. The minimum residual variance is given at $p = 4$, which is exactly the true model order.

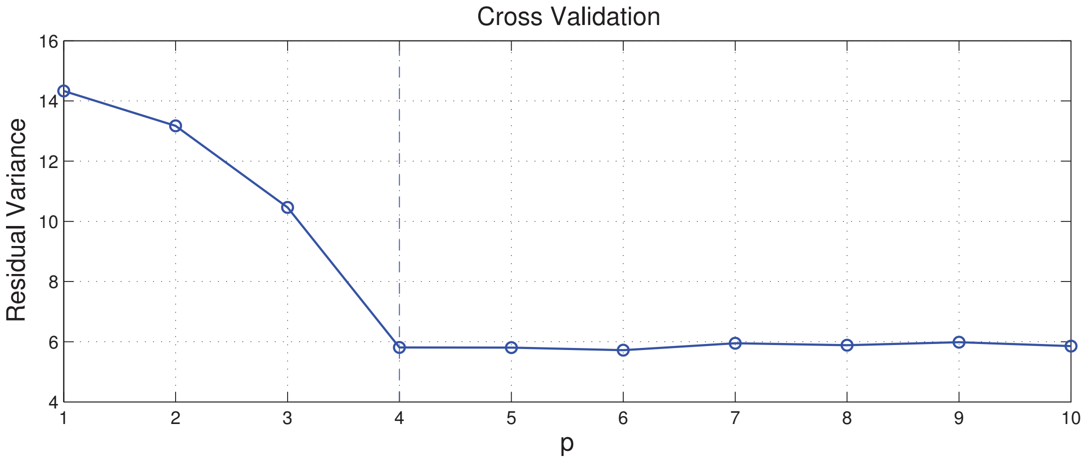

Figure 3.6: The cross-validation result of the lag domain method in SC1. For each p, the residual variance through a 5-fold cross-validation is computed. The residual variance reaches minimum at $p = 4$, which is exactly the true model order.

We use $p = 4$ and then estimate the linear coefficients. The parameter estimation are given by

$$\hat{\theta} = [2.1, -0.1, 0.3, 0, -0.1, 0, -0.1, 0.3, -0.1; p = 4] \tag{3.19}$$

Compared to the true parameter θ in (3.18), the determining parameters, p and a_{-4} are both well captured by the estimation. However, for other trivial coefficients, $\hat{\theta}$ seems to give an oscillating end effect.

De-aliasing Result: With the fitted model, the aliased high frequency component $Y^H|G_2$ can be predicted. Denote the predictions by $\hat{Y}^H|G_2$. Then the de-aliased series, denoted by Y^c, can be derived by (3.5). The de-aliasing results are presented in Figure 3.7. In addition, since the true $Y^H|G_2$ can be derived from the high-pass filter and sub-sampling of $Y|G_1$, the predictions $\hat{Y}^H|G_2$ and the true $Y^H|G_2$ are presented together in Figure 3.7 in order to evaluate the predictions. In this simple simulation example, the lag domain method gives satisfying results.

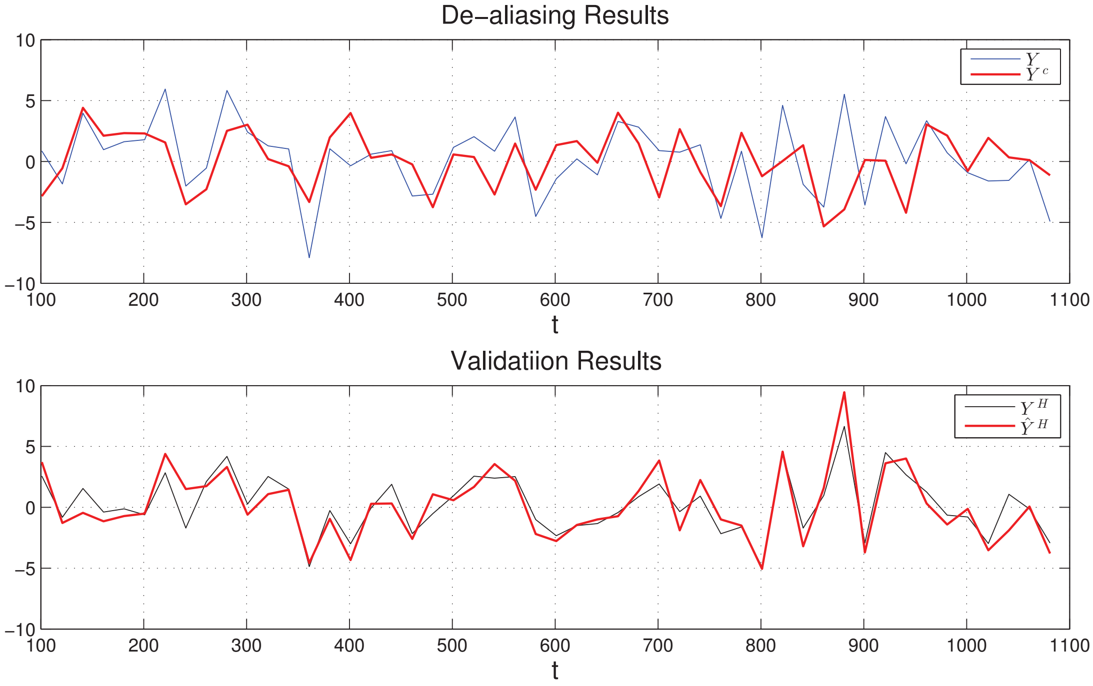

Figure 3.7: The de-aliasing results for the lag domain method in SC1. The upper panel shows the result of de-aliasing using the lag domain method. The de-aliased series (red line) contains less spikes and has a smaller variance, because the high frequency signals have been removed. The lower panel shows the validation result. The predicted high frequency component (red line) is compared with the true high frequency component (black line). Result shows that the predictions are very precise.

3.3.4 Frequency Domain Method

Parameter Estimation: The model order was first selected by a cross-validation procedure where M is chosen to give the minimum residual variance. This is similar to the lag domain method. Cross-validation results indicate that $M = 25$ is the optimal choice. The cross-validation result is in Figure 3.8.

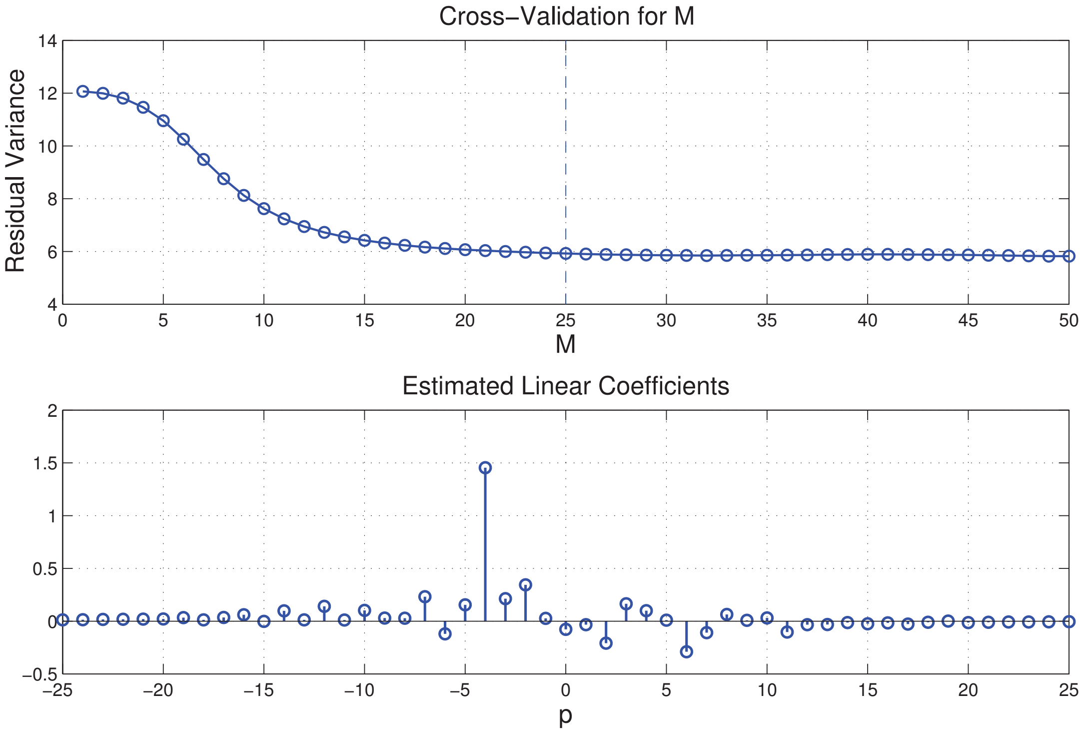

Figure 3.8: The cross-validation result and the estimated coefficients using the frequency domain method. The upper panel shows the cross-validation results. In the simulation example, the residual variance becomes stable after $M = 25$. The lower panel shows the estimated linear coefficients for this choice of window width M. The coefficients within the window width are presented since those outside are zeros. Note that a_{-4} is recovered well and other coefficients are much smaller as expected.

With $M = 25$, the auto- and cross-spectra and the transfer function can be estimated. The linear coefficients are derived from the inverse Fourier transform of the estimated transfer function $\hat{\Gamma}(\omega)$ using (3.13). The coefficients within $\pm M$ are plotted in Figure 3.8. The deterministic coefficient a_{-4} is recovered and all other coefficients are small, consistent with (3.18).

Transfer Function: In order to better illustrate the frequency domain method, we provide a detailed description of the estimation of the transfer function in this section.

The estimates for the sample ACVF and CCVF are given in the lag domain method. The auto-spectrum and cross-spectrum can be estimated and denoted by $\hat{h}_{X^H X^H}(\omega)$ and $\hat{h}_{Y^H X^H}(\omega)$. Note that in the spectral estimation, a Parzen window was applied and the window width $M = 25$. With $\hat{h}_{X^H X^H}(\omega)$ and $\hat{h}_{Y^H X^H}(\omega)$, the transfer function can be estimated by (3.12). In addition, the theoretical transfer function is derived from (3.17). The estimated transfer function $\hat{\Gamma}(\omega)$ and theoretical transfer function are presented in Figure 3.9.

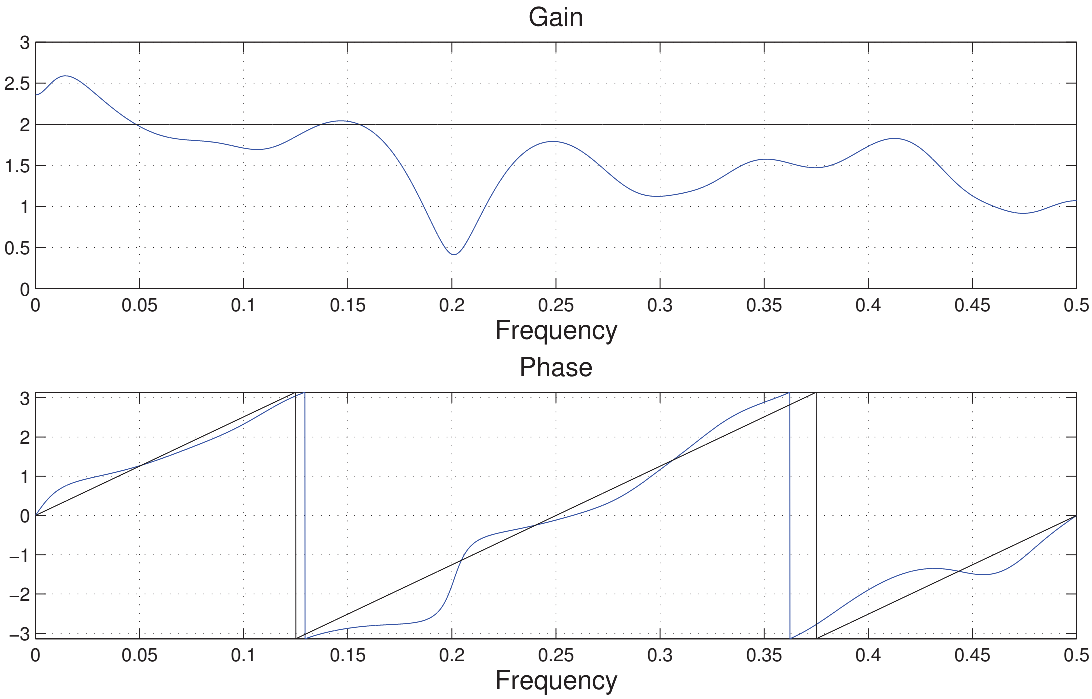

Figure 3.9: The gain and phase of the estimated transfer function (blue line) and theoretical transfer function (black line). The $[-1/2, 0]$ range is omitted because the gain is an even function and the phase is an odd function. The transfer function provides information on the transformation from X^H to Y^H. In this simple simulation example, the transformation consists of two parts, the lag and the amplification. It is obvious that the amplification is contained in the phase spectrum and the lag is contained in the phase spectrum.

The gain spectrum provides the information of the amplification from X^H to Y^H, which is constantly close to two. The phase spectrum provides the time delay information. The phase spectrum in Figure 3.9 is close to a straight line. Empirically this indicates a constant time delay at all frequencies. An approximation for this time delay is given by the slope, which is close to 4 ($\approx \dfrac{\pi}{2\pi \cdot 0.14}$) as expected.

De-aliasing Result: The de-aliasing results for the frequency domain method are presented in Figure 3.10. They are comparable to the results in the lag domain method (see Figure 3.7). Thus in SC1, the frequency domain method also works well.

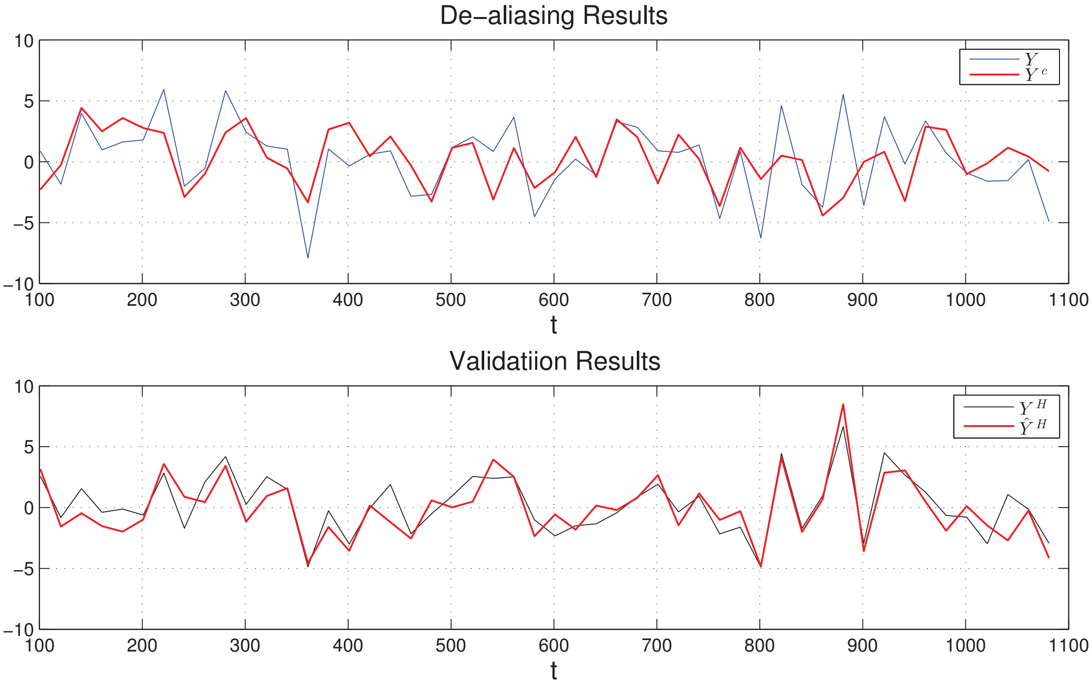

Figure 3.10: The de-aliasing results for the frequency domain method in SC1. The results are presented in the same format as in Figure 3.7. The validation of predicted high frequency component and true high frequency component of $Y|G_1$ in the lower panel indicates that the frequency domain method also performs well.

3.3.5 Lasso Method

Parameter Estimation: In the lasso model (3.14), we use lags within ±20 to construct the design matrix $\mathbf{X}$ and the response data vector $\mathbf{Y}$. In our case, $\mathbf{Y} = Y_t, t \in G_2$, where $G_2 = [\Delta, 2\Delta, \cdots, N_2\Delta]^T$. $\mathbf{X}$ contains 42 columns: the first column are all ones and the j-th column ($2 \leq j \leq 42$) corresponds to the data of X^H_{t-u}, with $u = j - 22$ ($-20 \leq u \leq 20$) and $t \in G_2$.

The lasso method was applied to the design matrix and the response data vector, with a 5-fold cross-validation process to select the optimal λ as in (3.15). After several tests, we chose the optimal $\lambda = 0.03$ which gives the minimum residual variance. Corresponding linear coefficient estimates of the lasso method are presented in Figure 3.11. Similar to the frequency domain method, the main coefficient a_{-4} is captured well but other coefficients are not perfectly suppressed.

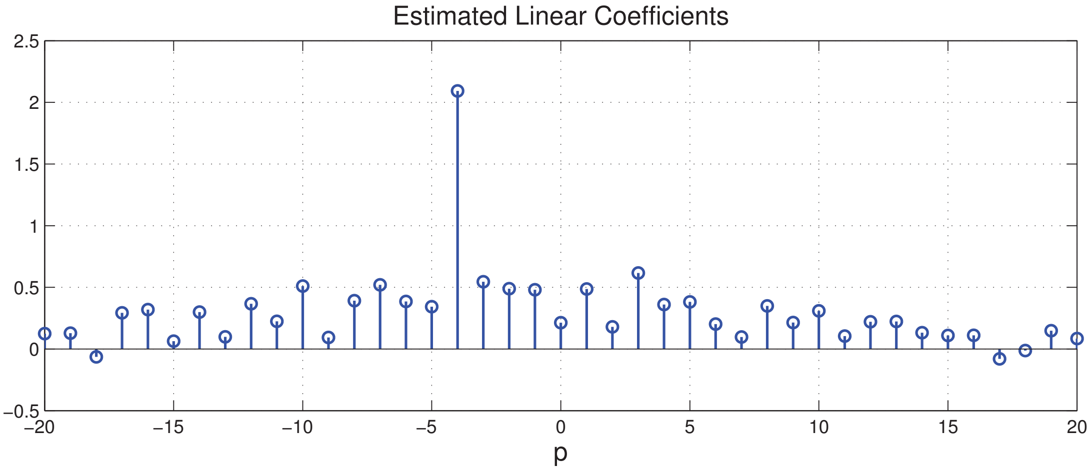

Figure 3.11: Plot of the linear coefficients estimated by the lasso method. Lags within ±20 are used as predictors so a_u with u within ±20 are estimated. The determining coefficient a_{-4} is well recovered.

De-aliasing Results: We now fit the original model with the parameter estimations and predict the aliased high frequency component. As in the lag domain and frequency domain methods, the de-aliasing results and the validation results are presented in Figure 3.12. The lasso method also gives satisfying predictions.

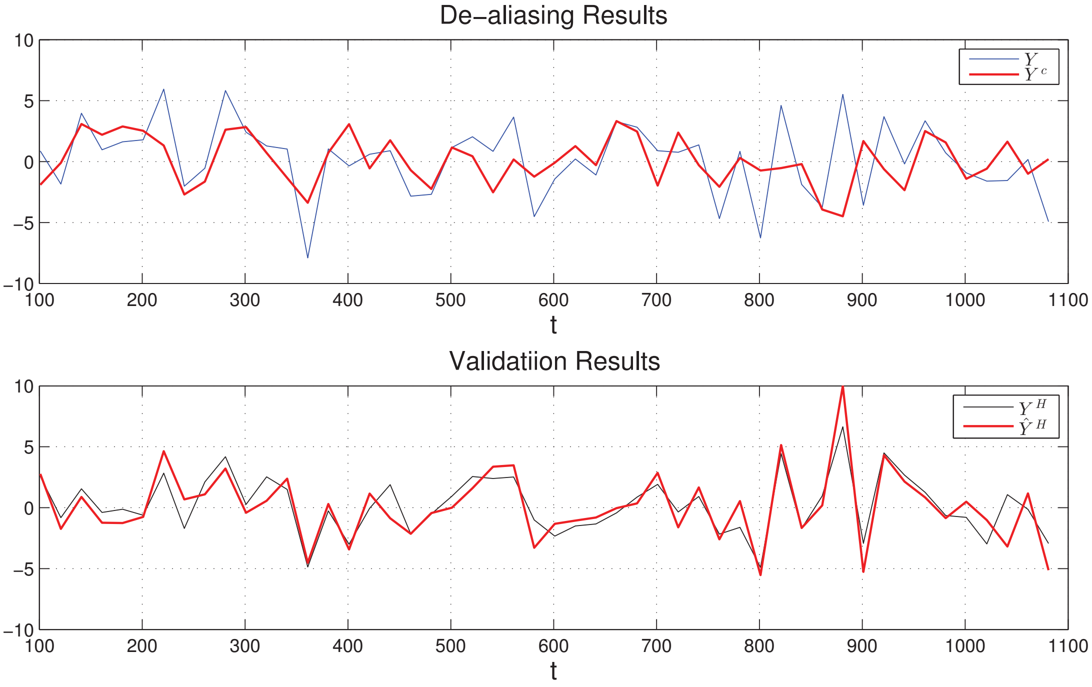

Figure 3.12: The de-aliasing results for the lasso method in SC1. The results are presented in the same format as in Figure 3.7. The predictions are also very precise.

3.4 Simulation Case 2

Simulation Case 1 is very simple. It helps describe the implementation of the methods, but in order to test the performances of the methods we need a more complex example. In this section, another simulation case (SC2, Simulation Case 2) is designed to test the de-aliasing methods and results are presented and discussed.

3.4.1 Simulation Data

First of all, we design a set of linear coefficients to give an amplification of signals within a certain frequency band. Figure 3.13 presents the linear coefficients and the corresponding (theoretical) transfer function. The gain spectrum shows that this linear system corresponds to a band-pass filter where there is an amplification of signals over a certain frequency band. This transfer function is more realistic and similar to the real data examined in this study.

SC1 is now updated with this new set of linear coefficients. More specifically, the process X remains an AR(1) process and the sample time series $X|G_1$ is the same as

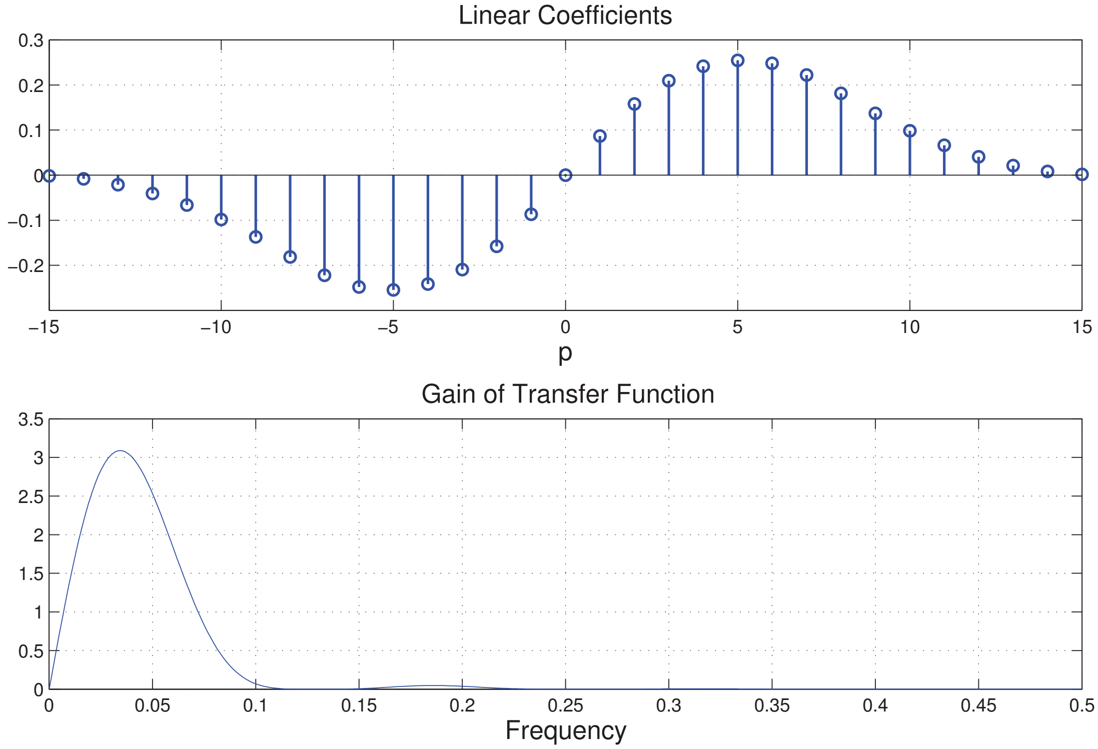

Figure 3.13: Linear coefficients a_u and the gain spectrum $|\Gamma(\omega)|$ corresponding to the linear coefficients for SC2. The set of 31 linear coefficients together creates an amplification at a frequency band, which is similar to a band-pass filter.

in SC1; the process Y is a linear transformation of X using the new transformation system; the sub-sample procedure and the implementations of the de-aliasing methods are also the same as in the simulation case 1.

The specifics of the implementations of the de-aliasing methods have already been described in SC1, and so I focus below on the estimation and de-aliasing results.

3.4.2 Parameter Estimation

For the lag domain method, I chose $q = 30$ and the cross-validation gives the optimal $p = 11$. For the frequency domain method, the cross-validation gives the optimal $M = 25$. For the frequency domain method, the ± 20 lags are used as predictors. the cross-validation gives the optimal $\lambda = 0.013$. For all three methods, the linear coefficients are estimated and plotted in Figure 3.14, along with the true coefficients.

The true linear coefficients are smooth but all three methods give erratic shapes,

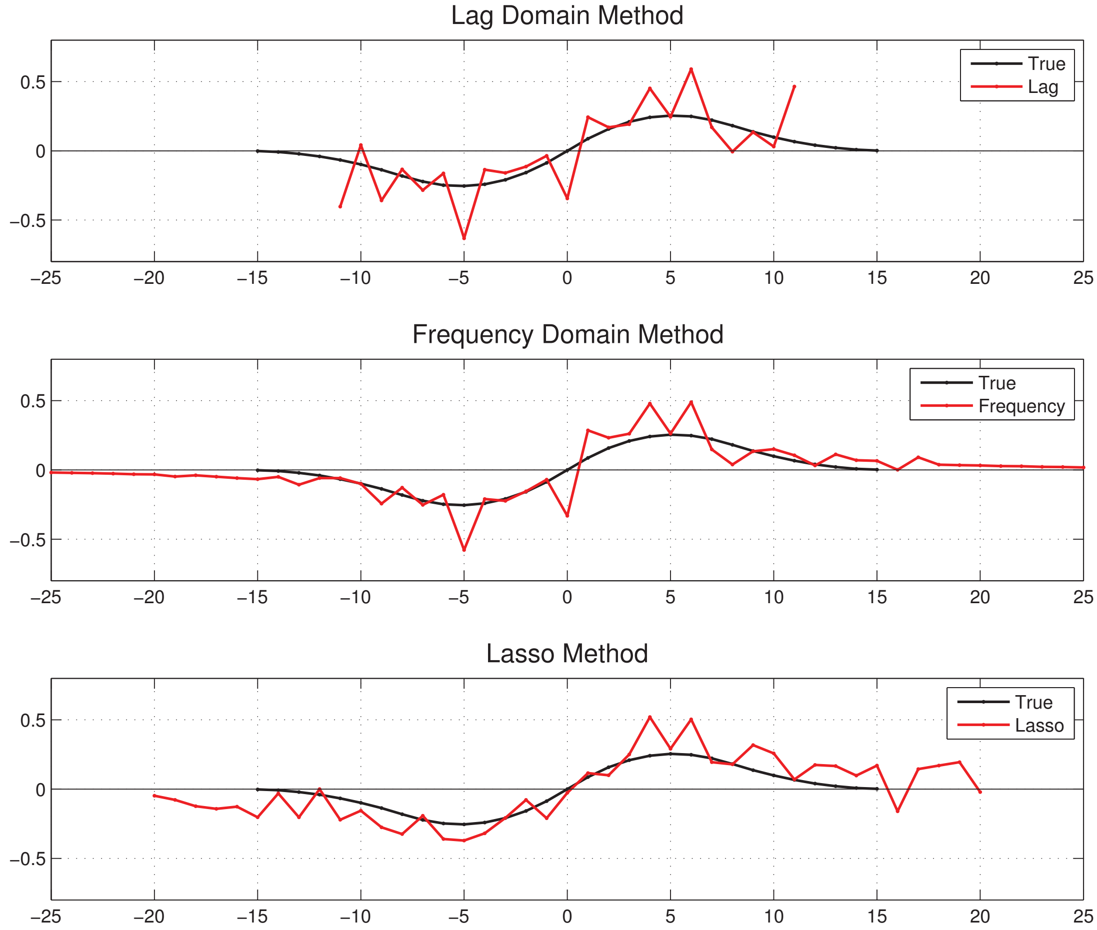

Figure 3.14: The true linear coefficients (black line) and the estimated linear coefficients (blue line) for the lag domain, frequency domain and lasso methods. All three methods give similar results to some extent. However, each of the three methods has its own advantages.

especially at the most significant coefficients. The lag domain method gives the least number of parameters and thus the simplest model, but the coefficients at the end are not well suppressed. The frequency domain method gives the most number of coefficients and the estimates best approximate the original shape, with both tails close to zeros. Note that the lag and the frequency domain methods are very similar within ±10. This is because both the two methods are from a same origin. On the other hand, the lasso method seems to perform best at keeping the relative contributions of the coefficients and thus gives the smoothest shape.

3.4.3 De-aliasing results

In order to evaluate the three methods, the predictions for the aliased high frequency component, $\hat{Y}^H|G_2$ is compared with the true high frequency component, $Y^H|G_2$. Results are presented in Figure 3.15. All the methods seem to give satisfying, and very similar predictions.

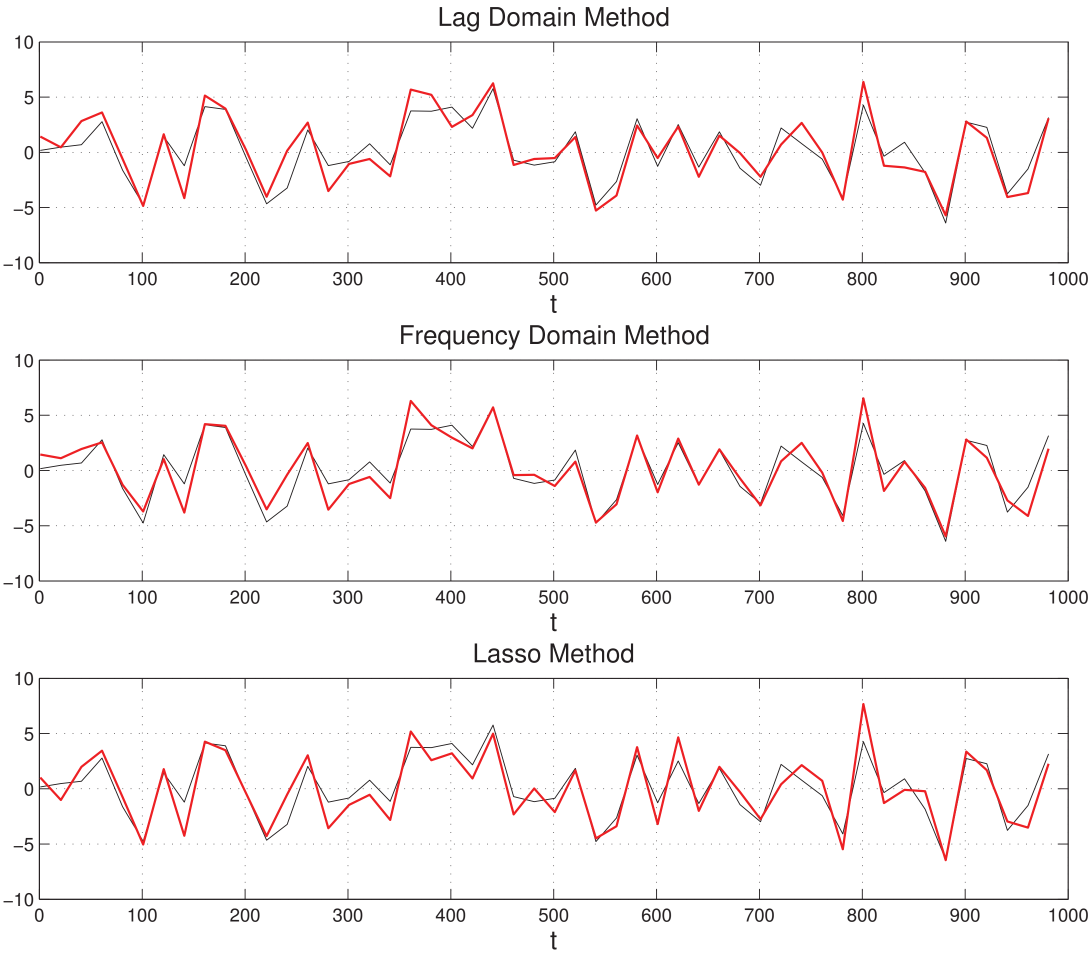

Figure 3.15: Predictions (red lines) and truth (black lines) for the aliased high frequency component of $Y|G_2$ for SC2. The first 50 data points in the time series are plotted. All three methods give satisfying predictions.

Chapter 4

Description of Observations and Initial Processing

The de-aliasing methods described in the previous chapter are now applied to sea level observations made in the Gulf of Maine, Scotian Shelf and Gulf of St. Lawrence. This chapter describes the sea level observations made by coastal tide gauges and space-borne altimeters. The altimeter observations are made with a minimum time spacing of about 10 days and so severely alias the strong tides of the region. (The Bay of Fundy in the Gulf of Maine has the largest tides in the world with a range of over 15 m.) The tide gauge observations are made hourly and they form the high frequency supplementary information (Chapter 3) used to de-alias the altimeter observations.

Before applying the de-aliasing method, the tide gauge and altimeter data sets were processed and organized. This chapter describes this initial processing including the removal of (i) tides and (ii) the effect of atmospheric forcing using a dynamically-based ocean model. Table 4.1 defines the processed variables used in the rest of this thesis.

Table 4.1: Notation for the processed sea level variables. The cut-off frequency of the high pass filter is defined by the Nyquist frequency of the satellite observations ($1/2\Delta$ cycles per day where $\Delta = 9.9156$ days). All sea levels are in meters.

Notation	Description
η^{Obs}	Observed sea level
η^{Tide}	Tidal component of η^{Obs}, used as a tidal correction
η^{Surge}	Surge component of η^{Obs} from the dynamical ocean model
η^{SLA}	Sea level anomaly, $\eta^{Obs} - \eta^{Tide} - \eta^{Surge}$
η^{AH}	High frequency (above $1/2\Delta$) component of η^{SLA}
η^{AL}	Low frequency (below $1/2\Delta$) component of η^{SLA} defined by $\eta^{SLA} - \eta^{AH}$

4.1 Tide Gauge Observations

High frequency signals are normally aliased in satellite altimetry records due to their long sampling interval. To correct for the aliasing effect, use is made of the high frequency information contained in hourly tide gauge observations of sea level. In general, tide gauge observations are also more precise and reliable than altimetry measurements although there spatial coverage is not as good. In the present study hourly sea level observations were used from the 22 tide gauges shown in Figure 4.1 and listed in Table 4.2.

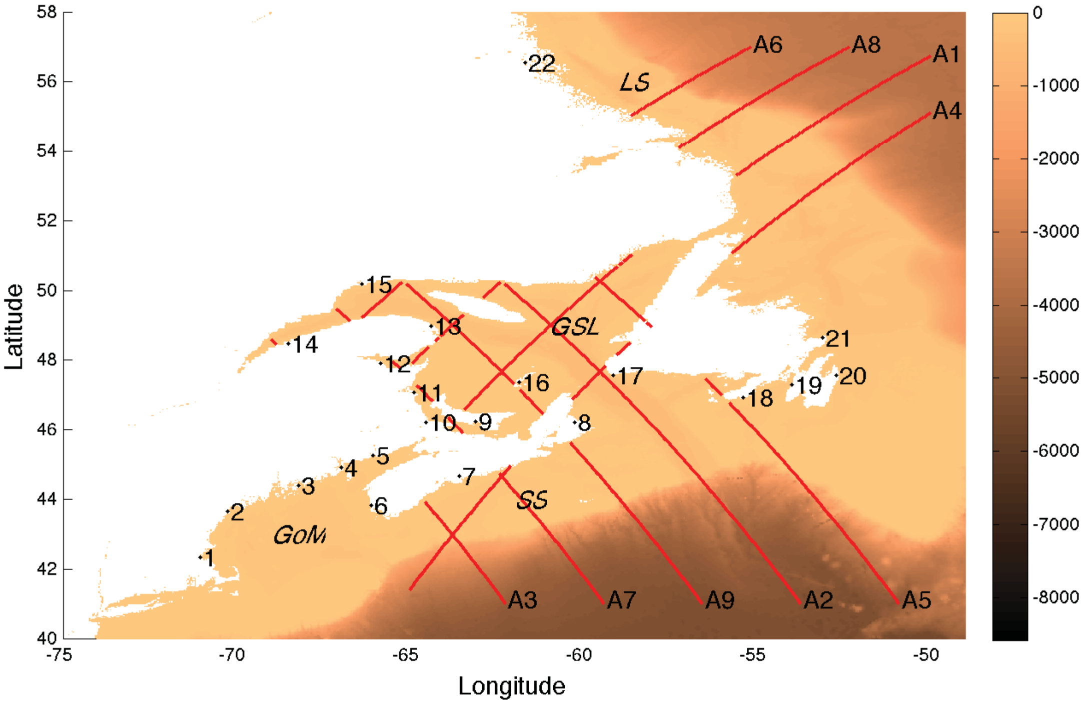

Figure 4.1: Distribution of tide gauge stations and satellite altimeter tracks. The location of the 22 stations are shown by the black dots and marked with numbers 1 to 22. The satellite altimeter tracks are shown as red lines and marked as A1 to A9. These observations cover the Gulf of Maine (GoM), Scotian Shelf (SS), Gulf of St. Lawrence (GSL) and Labrador Shelf (LS).

The hourly tide gauge records came from the Global Sea Level Observing System (GLOSS)[1]. The core network of GLOSS consists of 290 sea level stations around the world and is designed to provide measurements of global coastal sea level variations. The 22 tide gauges used in the present study are distributed along the Gulf of

---

[1] http://www.gloss-sealevel.org/

Maine, Gulf of St. Lawrence and coast of Newfoundland, Canada. Records for the 18 Canadian stations were downloaded from Fisheries and Oceans Canada[2]. For the US stations, Boston, Portland and Eastport records were obtained from the University of Hawaii Sea Level Center[3] and the Bar Harbor record came from National Oceanic and Atmospheric Administration of United States[4].

The original tide gauge records contained numerous gaps and had many other problems making it necessary to provide a thorough and comprehensive initial processing. The following corrections were made:

1. Repeat records were removed.
2. All records were converted to universal time (UTC) and interpolated to hourly values.
3. Records were trimmed to the time period of the satellite observations (1 January, 1993 to 31 December, 2014).
4. Missing values were denoted by NaN ("not a number").
5. Systematic one-point gaps were filled by interpolation (e.g., observations at 4:00 UTC are missing every day from the Nain record after 2012/10/30)
6. Each of the 22 stations were carefully scrutinized for possible datum shifts, outliers and other problems.

A brief summary of the tide gauge observations, and the results of the initial processing, are given in Table 4.2.

4.2 Tidal Correction

Tides dominate the variablity of all of the 22 tide gauge records (see, for example, Figure 4.2). The tides result from the gravitational pull of the Sun and Moon and also the rotation of the Earth. These tidal signals can be treated as deterministic and periodic with known frequencies. For example, the largest tidal constituent at most

---

[2] http://www.dfo-mpo.gc.ca/

[3] http://uhslc.soest.hawaii.edu/

[4] http://www.noaa.gov/

Table 4.2: Summary statistics of tide gauge observations and results of the initial processing. The first three columns give the name and location of tide gauges. The fourth column is the percentage of valid hourly data. The remaining columns give the standard deviations (in meters) of η^{Obs}, η^{Tide}, η^{Surge} and η^{SLA}. Data are for the period 1 January, 1993 to 31 December, 2014.

Station name	Lon	Lat	Data	σ_{Obs}	σ_{Tide}	σ_{Surge}	σ_{SLA}
1 Boston	-71.05	42.35	99.37	1.035	1.027	0.102	0.074
2 Portland	-70.25	43.66	99.14	1.032	1.026	0.101	0.069
3 Bar Harbor	-68.21	44.39	71.97	1.171	1.166	0.101	0.065
4 Eastport	-66.99	44.91	97.69	1.967	1.964	0.096	0.071
5 Saint John	-66.06	45.25	86.34	2.231	2.228	0.097	0.094
6 Yarmouth	-66.12	43.83	93.06	1.238	1.233	0.107	0.076
7 Halifax	-63.58	44.67	89.42	0.490	0.477	0.099	0.074
8 North Sydney	-60.25	46.22	98.10	0.324	0.301	0.103	0.080
9 Charlottetown	-63.12	46.23	93.62	0.619	0.597	0.130	0.105
10 Shediac Bay	-64.55	46.23	38.82	0.355	0.299	0.145	0.119
11 Lower Escuminac	-64.88	47.08	93.58	0.331	0.291	0.140	0.093
12 Belledune	-65.85	47.90	63.34	0.539	0.516	0.137	0.096
13 Rivière-au-Renard	-64.38	49.00	91.00	0.448	0.428	0.122	0.083
14 Rimouski	-68.51	48.48	96.09	1.013	1.001	0.144	0.092
15 Sept-Iles	-66.38	50.19	97.16	0.745	0.730	0.135	0.088
16 Cap-aux-Meules	-61.86	47.38	34.12	0.246	0.213	0.114	0.072
17 Port-aux-Basques	-59.13	47.57	88.80	0.357	0.337	0.104	0.066
18 St. Lawrence	-55.39	46.92	36.87	0.501	0.484	0.117	0.090
19 Argentia	-53.98	47.30	93.74	0.538	0.523	0.116	0.083
20 St John's	-52.72	47.57	97.01	0.325	0.295	0.127	0.078
21 Bonavista	-53.12	48.65	39.80	0.283	0.255	0.118	0.072
22 Nain	-61.68	56.55	46.54	0.603	0.591	0.118	0.069

of the tide gauges is the principle lunar semi-diurnal constituent (usually denoted by M_2). It has a period of about 12 hours and 25.2 minutes. These high frequency tidal signals are aliased by the satellites and need to be corrected.

Tides can be estimated in several ways. For example, bandpass filters can be used but, in general, they are not efficient because of difficulties in filter design (the tides have a discrete spectrum to first order) and also handling gaps in the sea level record. The most popular way of estimating the tides is based on classical harmonic analysis. Given the specific tidal frequencies, classical harmonic analysis assumes a regression model which is essentially a sum of sinusoids at the given frequencies. Least squares is then used to estimate the tidal amplitude and phase and thus the time-varying tide. It is generally accepted that harmonic analysis is more effective than bandpass filters. However, classical harmonic analysis also has some drawbacks and this has led to more sophisticated and effective forms of regression model as implemented for example in the T_Tide package (Pawlowicz et al., 2002). This package allows for "inference" of tidal constituents that are so close in frequency to dominant constituents like M_2 that they cannot be resolved with just the available data record. The package also allows for nodal variations in tidal amplitude and phase which occur with a period of 18.6 years (due to slow variations in the Moon's orbit).

For each of the 22 tide gauge records, above package was used to predict the tide. For a better performance, the following steps were taken:

1. Each record was divided into yearly time series for separate tidal analysis by T_Tide.

2. Long gaps at the head and tail of each yearly time series were removed prior to tidal analysis as suggested in Pawlowicz et al. (2002).

3. Periods with insufficient data were replaced by missing values. Insufficient was defined by at least 764 hourly values. This ensures that the major tidal constituents will be reliably estimated by T_Tide.

Sea level variations were reduced significantly after removal of the tides as shown in Figure 4.2 and Table 4.2 (compare the standard deviation of the observations and tide).

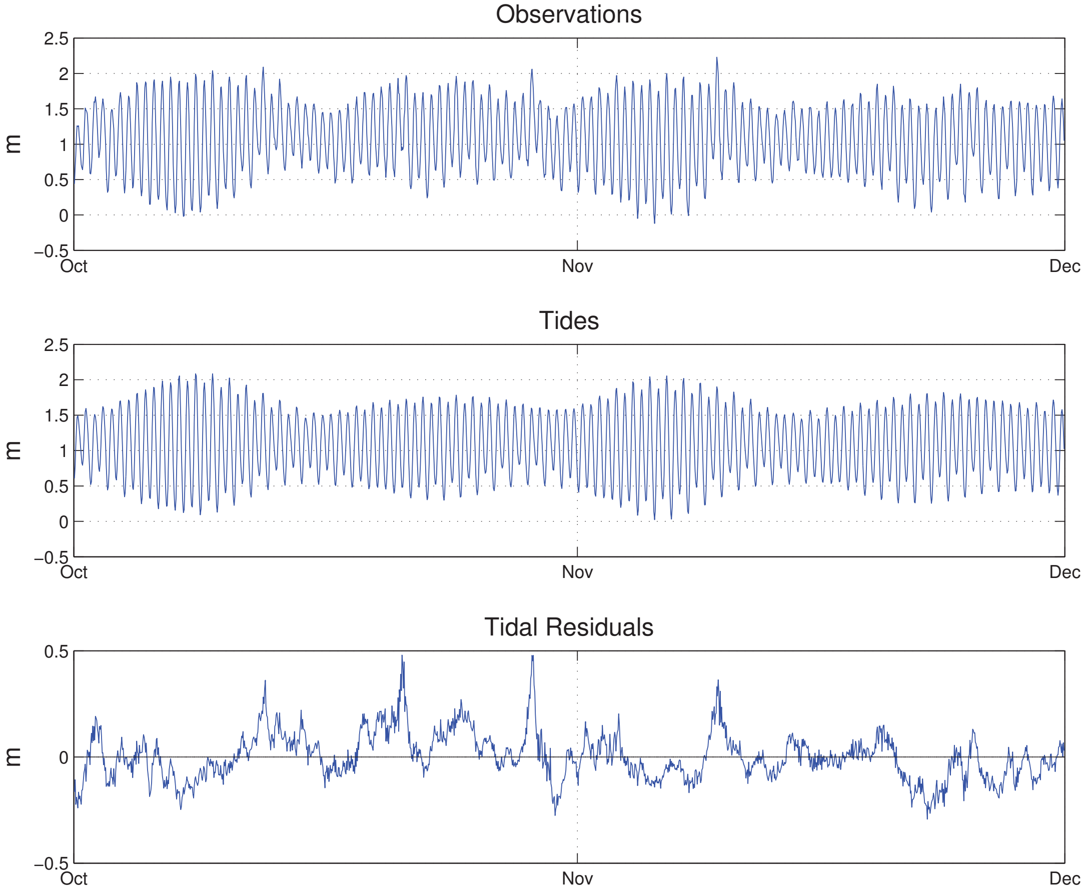

Figure 4.2: Decomposition of observed hourly sea level series (top panel) into tide (middle panel) and residuals (bottom panel). Tides were estimated using T_Tide, a MATLAB package designed by Pawlowicz et al. (2002). The tidal residuals are given by subtraction of the tides from the observations. Note the expanded vertical scale for the bottom panel, illustrating the dominance of the tides over the residual variability. The observations are for Halifax, Nova Scotia for two months in 2006.

4.3 Dynamic Atmospheric Correction

Changes in wind and air pressure can cause storm surges over the shallow water close to shore (i.e., the continental shelf, normally defined as water shallower than 200 m). Large storm surges can cause sea level changes of meters over a short period of time, as seen in Figure 4.2. These high frequency signals, like tides, are normally aliased by altimeters and need to be removed. One way to estimate and remove wind and atmospheric effects is by using a dynamically-based model of the ocean forced by wind and air pressure. The result is the so-called “Dynamic Atmospheric Correction (DAC)”.

In the 1990s, the DAC simply came from the inverted barometer correction (IB). The IB approximation assumes a static ocean response to atmospheric pressure for low frequencies (i.e., periods longer than about 20 days). The IB associates a 1 millibar increase in atmospheric pressure with a 1 cm drop in sea level. Wind effects were ignored. However, more recent studies have shown that the ocean response to atmospheric forcing is more complicated at high frequency and the effect of the wind cannot be ignored.

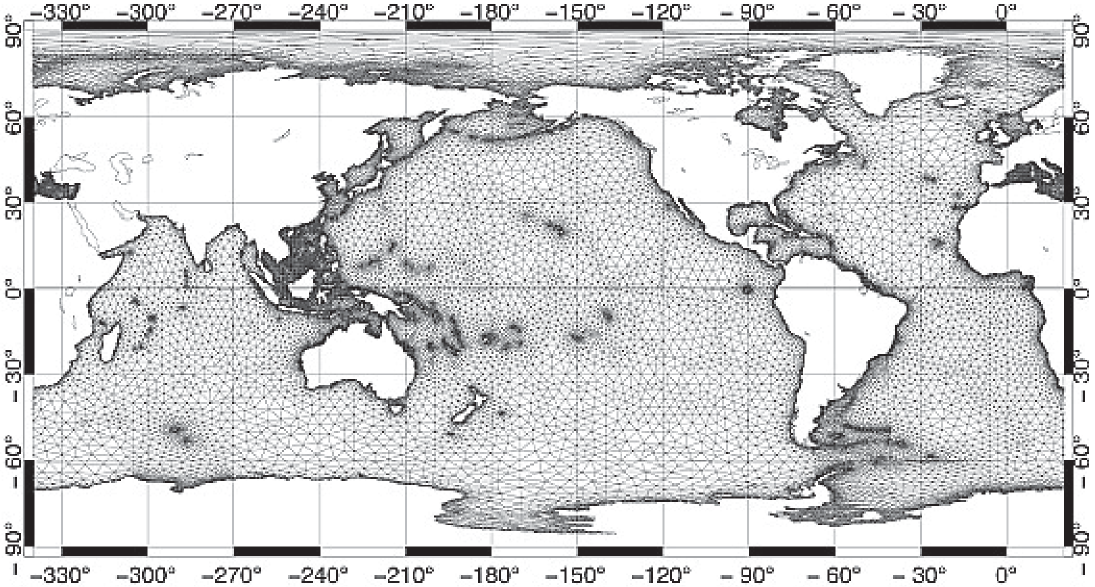

Figure 4.3: Finite element mesh of the MOG2D-G model used to predict storm surges. Smaller grid elements (higher resolution) are used in coastal areas where the storm surges are generally largest. From Carrère and Lyard (2003).

To improve on the static assumption, Carrère and Lyard (2003) developed a dynamically-based model called MOG2D-G. It is a nonlinear, time-stepping, dynamical ocean model based on the approach of Lynch and Gray (1979). It uses a finite element space discretization of the global ocean and adjacent shelves. The use of finite elements allows increased resolution in coastal and shallow water areas. Figure 4.3 shows the global grid. The grid size is approximately 400 km in the deep ocean, reducing to about 20 km approaching coastal areas. Using such a grid it is possible to achieve high resolution in important coastal areas at relatively low computational cost. Carrère and Lyard (2003) performed a global simulation for the Topex/Poseidon (T/P) period (1992-2002) and compared the sea level predictions to de-tided sea level records from tide gauges and satellite altimeters. It was shown that MOG2D-G simulates the high frequency, atmospherically-forced sea level variability of the global ocean with unprecedented accuracy. For example, the model correction reduces the variance of the tidal residuals by typically 50%.

The DAC predictions produced by MOG2D-G are widely adapted for oceanographic research. The data are distributed by AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic data) with support from CNES (Centre National d'Etudes Spatiales) [5]. Data are archived in year-by-year directories starting from 1992, with day-by-day groups of files with four files per day (0, 6, 12 and 18 hour, UTC). Each file stores global DAC data for a given time on a $0.25^\circ \times 0.25^\circ$ grid.

For each of the 22 tide gauges, DAC data are needed to further correct the tidal residuals. However, the stations are not distributed precisely on the surge model grid. To get DAC data for each station it is sometimes necessary to interpolate and extrapolate the model output. Unfortunately this is not straightforward because the $1/4^\circ$ grid distributed by Aviso is sometimes more than 20 km from the coast. To obtain the optimal DAC predictions for a given tide gauge, and avoid complicated interpolation and extrapolation procedures, the following statistical approach was used. For each tide gauge, all of surge model grid points less than 0.5° are identified. The correlation between each selected grid point time series and the observed residual at the tide is calculated. The grid point time series with the highest correlation is then used to provide the DAC for the given tide gauge. Finally this 6-hour DAC time series is linearly interpolated to the hourly time spacing of the sea level observations. Dynamic atmospheric corrections and sea level anomaly (SLA) after corrections are shown in Figure 4.4.

For further study, each sea level anomaly series was high-pass filtered to separate low and high frequency components. The cut-off frequency is the Nyquist frequency of the satellite altimetry record, π/Δ. The high frequency component is denoted by η^{AH} and the low frequency component is denoted by by η^{AL}.

---

[5] http://www.aviso.altimetry.fr/

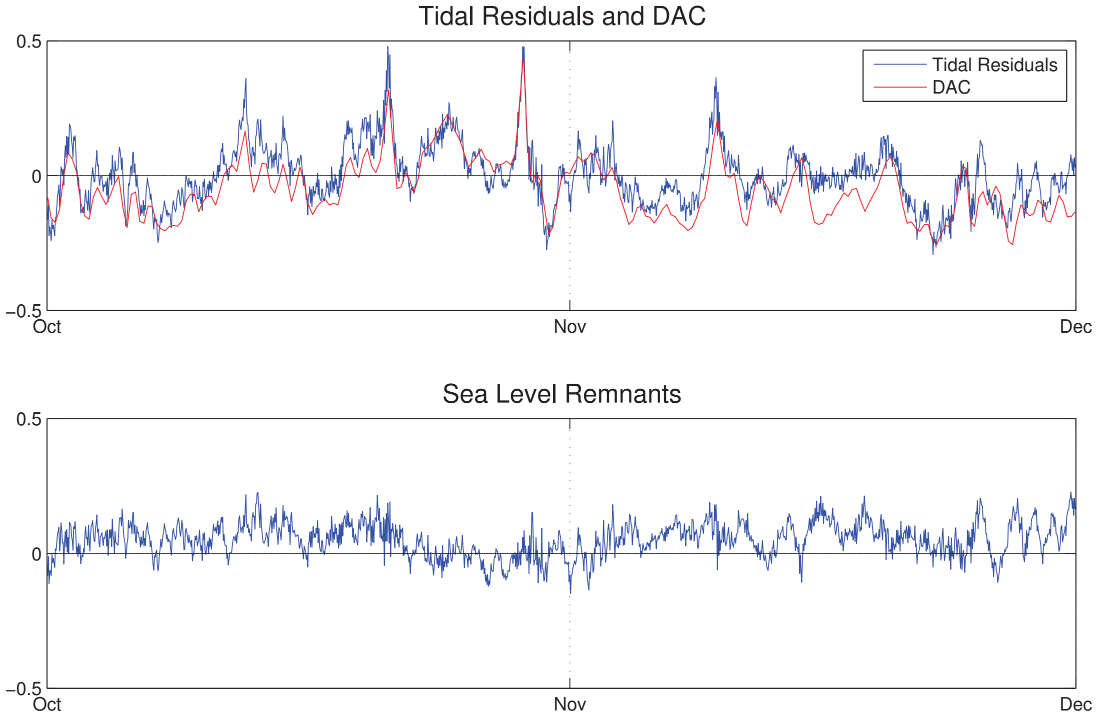

Figure 4.4: Dynamic atmospheric correction and tidal residuals (top panel) and their difference (bottom panel). The agreement between the red and blue lines in the upper panel illustrates the accuracy of the surge model predictions. The smoothly varying character of the sea level anomaly plotted in the lower panel shows that much of the high frequency variability in the tidal residual ($\eta^{Obs} - \eta^{Tide}$) has been removed by subtraction of η^{Surge}, i.e., the DAC. Data are for Halifax, Nova Scotia for two months in 2006.

4.4 Altimeter Observations

Satellite altimetry has revolutionized oceanography research since the early 1990s. The observations are accurate (within cm) and provide unprecedented global monitoring of the deep ocean and the adjacent continental shelves. The amount of data is vast: the number of sea level observations from satellite altimeters over one one cycle (repeated every 10 days) exceeds a hundred year's tide gauge observations from the present global array.

There are however some issues with satellite altimetry. A critical issue is the subject of the present research: aliasing. The satellite altimeter orbits the earth and measures the sea surface along fixed tracks with a revisit time of roughly 10 days.

For each point on the track, the sampling interval is thus roughly $\Delta = 10$ days. Consequently any information below the Nyquist period of 20 days is aliased and will contaminate the altimeter record. Figure 1.1 illustrates the aliasing effect using the observed Halifax tide gauge record. Another issue is the accuracy of the observations close to shore. Although the altimeters perform well over the open ocean their accuracy is significantly degraded in coastal areas: the presence of land increases instrumental error and additional geophysical corrections are required. Several preprocessing techniques used to extract more accurate altimetric information in coastal areas are described by Vignudelli et al. (2011).

Regional along-track sea level anomaly (SLA) products from CTOH/LEGOS are used in this study. These SLA products are based on a combination of data from multiple satellite missions (Topex/Posiden, Jason-1 and Jason-2) and they have been corrected by state of the art processing techniques. For example, in coastal areas, outliers were removed using criteria described in AVISO (1996). Several geographic corrections were then applied by CTOH/LEGOS to improve product quality. Deterministic high frequency components such as tides and surges were removed using X-track tidal correction (Carrère et al., 2012) and dynamic atmospheric correction (Carrère and Lyard, 2003). Finally, potential remaining outliers in SLA were removed using a 3 standard deviation filter. See Vignudelli et al. (2005), Roblou et al. (2011), and Durand et al. (2008) for details of processing.

Regional along-track sea level anomaly (SLA) products distributed by Aviso are archived by track. Each track file contains satellite information about the track, latitudes and longitudes for points along the track, and along-track data, including observation time, sea level anomaly, tidal correction, dynamic atmospheric correction and so on. In this study, nine tracks that cover the Gulf of St. Lawrence were selected for study (Figure 4.1). Figure 4.5 shows a representative altimetric time series for a position in the Gulf of St. Lawrence and compares it to the sea level anomalies for a nearby tide gauge.

4.5 Organization and Storage of Observations

For both datasets, all intermediate results and final results of the initial processing were organized into a MATLAB data structure for further analysis. The tide

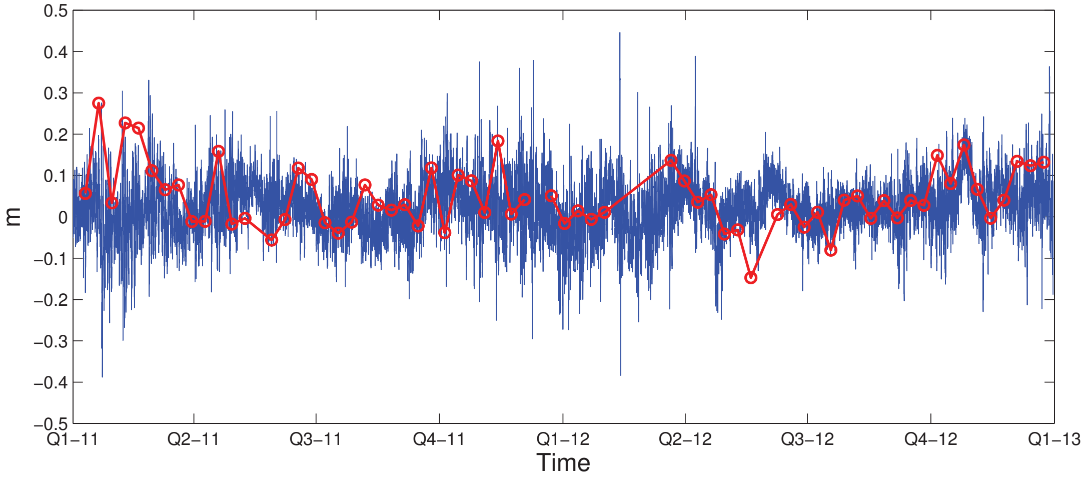

Figure 4.5: Comparison of tide gauge and altimeter observations. Both sets of data have been corrected for tides and the DAC. The hourly sea level anomalies (blue line) are from the Cap-aux-Meules tide gauge (-61.86°E, 47.38°N). The altimeter sea level anomalies, defined every 10 days, are from position (-62.40°E, 47.71°N), close to the Cap-aux-Meules tide gauge. Data are for 2011 and 2012.

gauge data structure is organized by station. For each station, the original sea level observations and their observation times, the tidal corrections, dynamic atmospheric corrections, sea level anomalies, low and high frequency part of sea level anomalies were saved, along with basic information of the station including station number, station name, latitude and longitude. The satellite altimetry data structure is organized by track. For each track, sea level observations, tidal corrections, dynamic atmospheric corrections and sea level anomalies are saved according to each point on track. Observation times, latitudes and longitudes for each point are also stored in the structure.

Chapter 5

Spatial Structure of Sea Level Anomalies of Tide Gauges

Removing the effect of tides and storm surges reduces significantly the variance of the sea level observations. However, the remaining variance is not negligible. It corresponds to a standard deviation of 6 to 12 cm at the 22 tide gauges (Table 4.2, column labeled σ_{SLA}). If the sea level anomalies ($\eta^{SLA} = \eta^{Obs} - \eta^{Tide} - \eta^{Surge}$) contain high frequency signals, they will be aliased by the altimeter and this will contaminate the signal the altimeter is attempting to observe. This leads to the following question: Does η^{SLA} contain high frequency signals that will cause significant aliasing problems? To answer this question, principal component analyses have been performed on the hourly time series of η^{SLA} at the 22 tide gauges for the period 1993 to 2014 inclusive. The basic idea is that if it can be shown that most of the η^{SLA} variance can be accounted for by a small number of principal components then the anomalies are not just noise and there is the potential to improve upon the tidal and surge corrections.

The structure of this chapter is as follows. A principal component analysis (PCA) is first performed on the high frequency component of the sea level anomalies (η^{AH}) at the tide gauge sites. (Overviews of PCA, and its frequency dependent generalization, are given in Chapter 2.) This is followed by a frequency dependent PCA of η^{SLA}. Both analyses lead to the conclusion that the tide and surge corrections are inadequate and this motivates the statistical approach to de-aliasing described in Chapter 6.

5.1 Principal Component Analysis of η^{AH}

The correlation and covariance matrices of the high frequency component of the hourly sea level anomalies (η^{AH}) at the 22 tide gauges is shown in Figure 5.1. Both matrices suggest that most tide gauges can be placed into one of two groups based on the similarity of their time series of η^{AH}. One group (stations 1 to 6) corresponds to tide gauges in the Gulf on Maine (see Figure 4.1 and Table 4.2 for locations). The other group (stations 8 to 17) corresponds to tide gauges in the Gulf on St. Lawrence.

The within-group correlations are typically 0.7 and the Gulf of St. Lawrence group is the most energetic (consistent with the column labelled σ_{SLA} in Table 4.2). Based on this visual examination of the correlation and covariance matrices, it is concluded that large scale signals remain in the time series of η^{AH}.

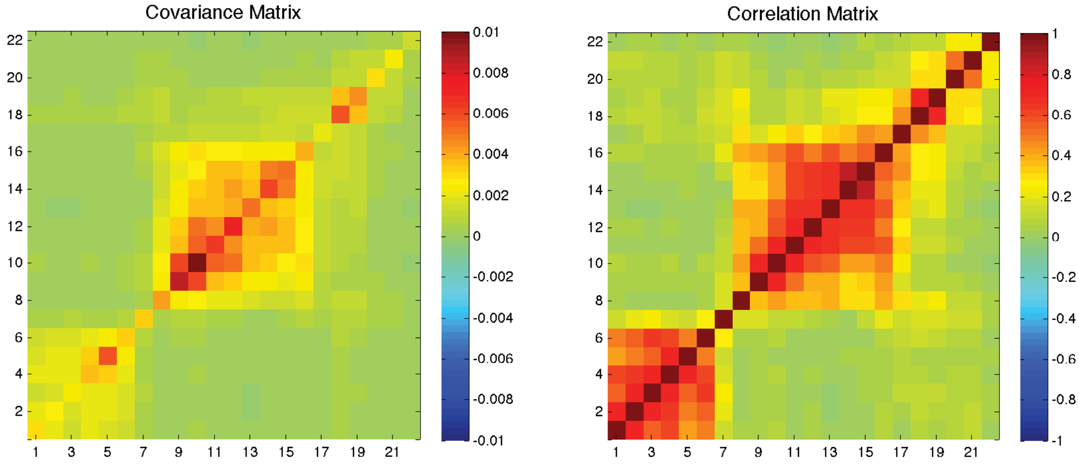

Figure 5.1: Covariance and correlation matrices of the hourly time series of η^{AH} at the 22 tide gauges. Based on data for the period 1/1/1993 to 12/31/2014. The numbers $1, 2, \cdots 22$ refer to the tide gauge station codes listed in Table 4.2

To define the dominant spatial modes of variability of η^{AH}, and the associated time-varying amplitudes, a PCA was performed on the correlation matrix shown in Figure 5.1. The first two principal components account for 44% of the total (standardized) variance and the first four account for 60% (upper panel of Figure 5.2). The lower panels show the first principal component mainly picks up information from stations 8 to 17 and the second mode mainly picks up information from stations 1 to 6. This is consistent with the visual interpretation of the correlation matrix given above.

The time-varying amplitudes of the first two modes (i.e., the first two principal components) are plotted in the upper panels of Figure 5.3. Both time series are dominated by quasi-periodic variations. Power spectra of the first two modes are plotted in the lower panels of Figure 5.3. Each time series was normalized by its standard deviation prior to spectral analysis, thereby emphasizing the relative energy distribution with regard to frequency. The major peak of the power spectrum of the first principal component has a period of roughly 32.8 hours and the spectral peaks

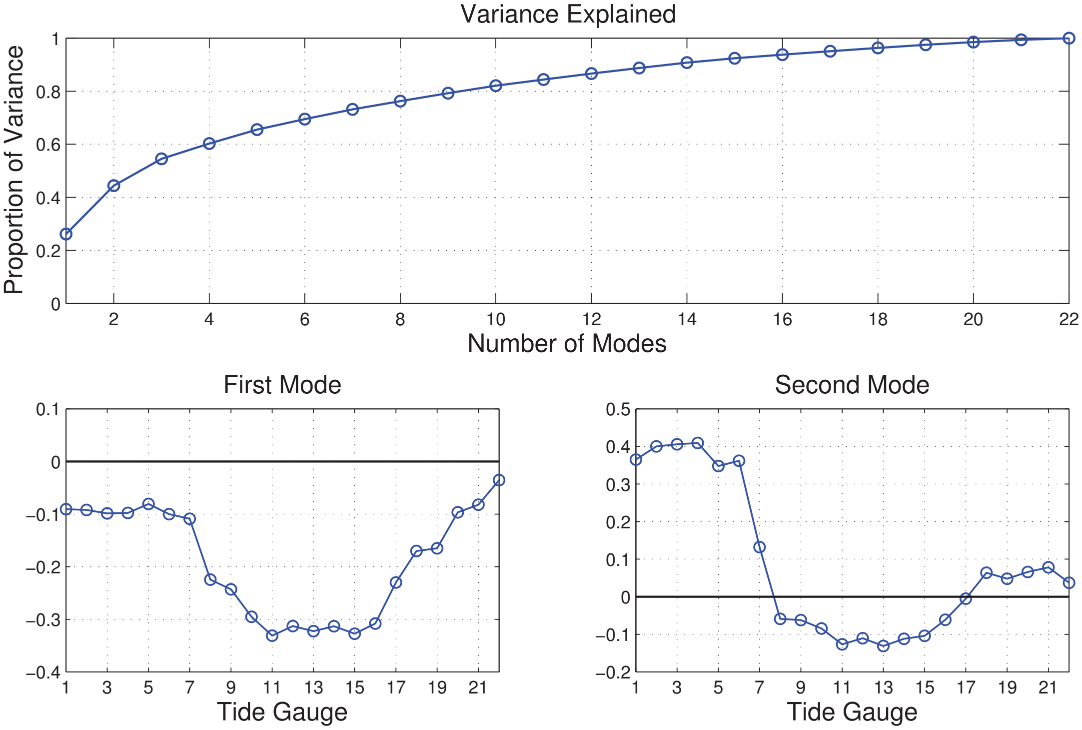

Figure 5.2: Results of a principal component analysis of the hourly time series of η^{AH} at the 22 tide gauges. The correlation matrix was used. The upper panel shows the cumulative proportion of explained variance as a function of mode number. The bottom panels show the first and second eigenvectors of the correlation matrix plotted against tide gauge index.

of the second principal component has a period of about 12.7 hours. This means that both modes of variability contain significant signals above the Nyquist frequency of satellite altimetry and that will lead to aliasing.

5.2 Frequency Dependent Principal Component Analysis of η^{SLA}

To better define the structure of the sea level anomalies as a function of frequency, a principal component analysis was performed on the 22 time series of hourly η^{SLA} in the frequency domain. (For details of frequency dependent PCA see Chapter 2.)

The 22 × 22 cross spectral matrix of the η^{SLA} time series at a given frequency consists of power spectra on the diagonal and cross spectra off the diagonal (similar to a covariance matrix). The cross spectral matrix is Hermitian. To estimate the power and cross spectra, missing values in each of the η^{SLA} time series were replaced

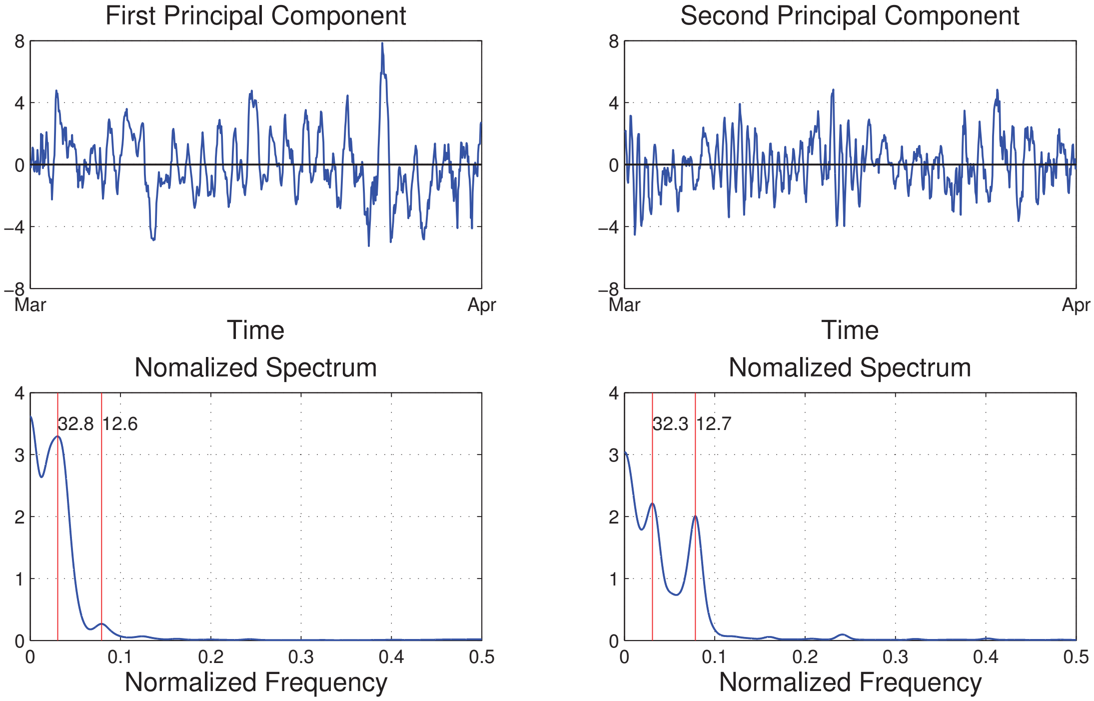

Figure 5.3: The first two principal components of the correlation matrix of η^{AH} and their spectra. The principal components were reconstructed from the high-passed hourly sea level anomalies at the 22 tide gauges for the period 1993 to 2014. Frequencies corresponding to the two major peak periods are shown by the red lines. Data for time series plot are one month from 2010.

by their means, and a Parzen lag window of length 100 (corresponding to a spectral window with a bandwidth of 1/52.4 cycles per hour) was used. The result of the frequency dependent PCA is a total of 96421 cross spectral matrices, one for each frequency resolved by the spectral analysis. (The record length is $N = 192840$ and this leads to $N/2 + 1 = 96421$ spectral estimates at equispaced frequencies between 0 and 0.5 cycles per hour.)

Let $\lambda_k(\omega)$ denote the kth eigenvalue of the cross spectral matrix at frequency ω. As discussed in Chapter 2, $\lambda_k(\omega)/\sum_{k=1}^{22}\lambda_k(\omega)$ is the proportion of total power explained by the kth principal component at frequency ω. A frequency-averaged version of this proportion is given by $\overline{\lambda}_k/\sum_{k=1}^{22}\overline{\lambda}_k$ where $\overline{\lambda}_k = \int \lambda_k(\omega)\, d\omega$. The cumulative version of this quantity is plotted in Figure 5.4. It can be seen that, when averaged with respect to frequency, the first principal component explains 46.5% of total variance and the first two explain almost 58.7%. As expected there values are higher than proportions

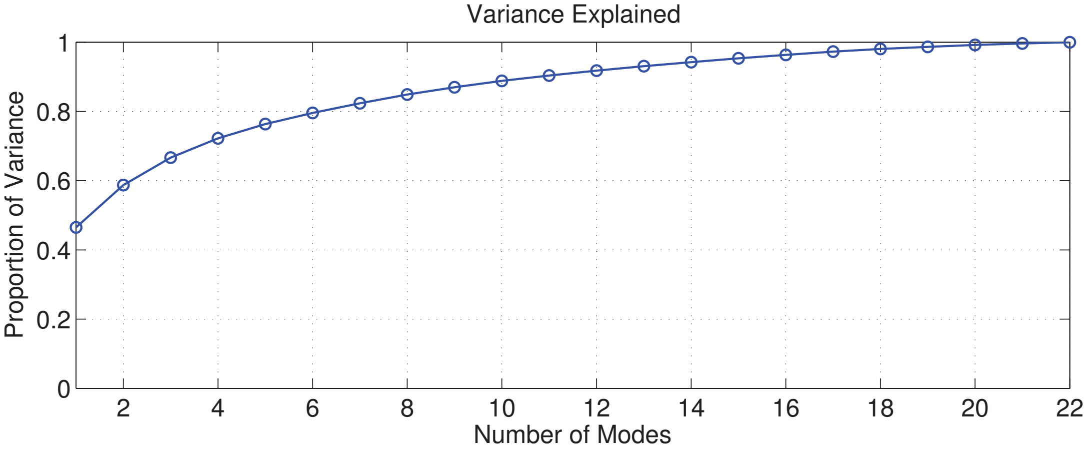

Figure 5.4: Cumulative proportion of total power at frequency ω explained by the first k principal components of η^{SLA}. The quantity $\sum_{j=1}^{k} \overline{\lambda}_j / \sum_{j=1}^{22} \overline{\lambda}_j$ is plotted as a function of k.

of explained variance obtained with the standard PCA because the modes can change with frequency.

The first principal component contains most of the information. The first eigenvalue of the cross spectral matrix, $\lambda_1(\omega)$, is plotted as a function of frequency in the upper panel of Figure 5.5. Two large peaks at nonzero frequency are evident at periods of approximately 30.3 hours and 12.7 hours. This is broadly consistent with the standard PCA of η^{AH} and, in particular, the quasi-periodic variations of the first two principal components shown in Figure 5.3 with periods of about 30 and 12.7 hours. According to the first eigenvalue, the variance explained by the first principal component with regard to each frequency, $\lambda_1(\omega)$, is plotted in the lower panel of Figure 5.5. Accordingly at the two large peaks, most variance is explained by the first principal component. There are other peaks in $\lambda_1(\omega)$, but peak frequencies only correspond to minor energy.

In addition, the 30-hour signal has a wide band. The Nyquist frequency of the altimetry record is comparatively very small which results in a narrow frequency band. Consequently if the spectrum in Figure 5.5 is the truth, then the aliases of the 30-hour signal will appear at each frequency of the narrow altimetry frequency band. That is, in the frequency domain, the aliasing effect could end up with a resemblance to an overwhelming white noise.

Each cross spectral matrix has 22, complex-valued eigenvectors. The amplitude

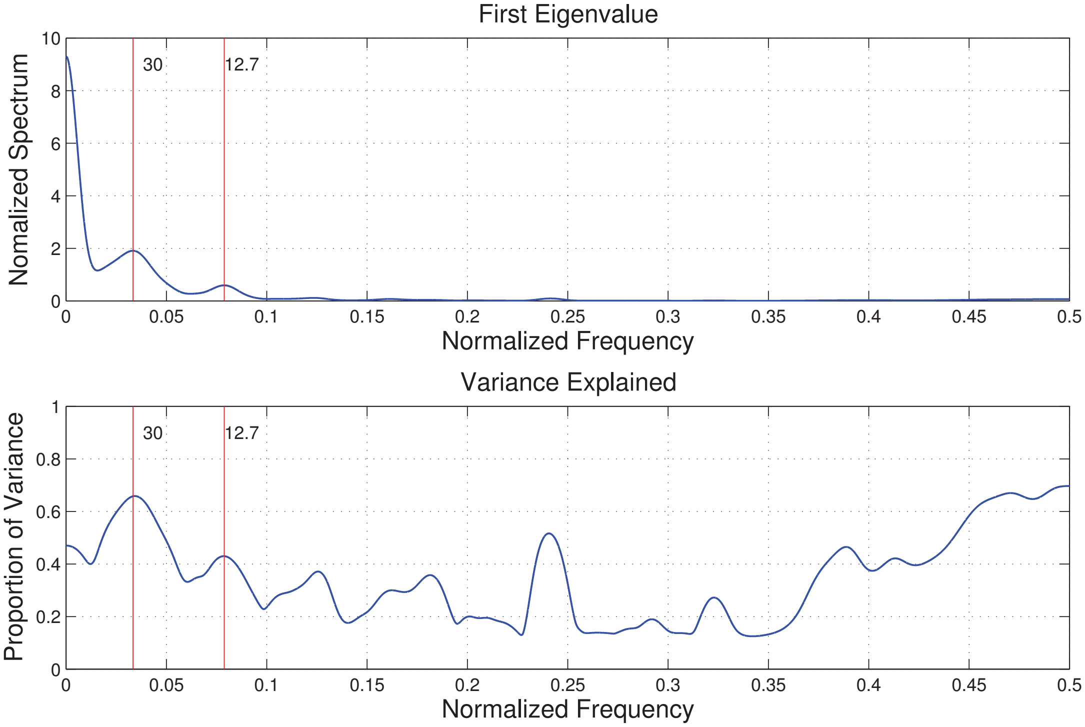

Figure 5.5: The first eigenvalue of the cross spectral matrix of η^{SLA} as a function of frequency (top panel) and the proportion of variance explained by the first principal component at each frequency (bottom panel). The frequency is in cycles per hour. Frequencies corresponding to periods of 30 and 12.7 hours are shown by the red lines.

and phase of the first eigenvector are plotted in Figure 5.6. As discussed in Chapter 2, the amplitude is analogous to the coefficients of principal components estimated in the time domain and provide information on the contribution of the original variables to each principal component. The phase, on the other hand, provides information on lags between series and allows propagating features to be described by a single mode (Chapter 2). To concentrate on the two dominant structures described above, the upper limit of the frequencies plotted in Figure 5.6 is taken to be 0.02 to 0.1 cycles per hour.

The amplitude plot clearly shows two main groupings of tide gauges that are defined in different frequency ranges. The first group includes stations 1 to 6 (Gulf of Maine, largest amplitudes in the Bay of Fundy) and the second group includes stations 9 to 15 (Gulf of St. Lawrence). The frequency ranges of the first and second groups are approximately 0.06 to 0.1 (periods of 17 to 10 hours) and 0.02 to 0.06 (50 to 17 hours) respectively. These frequency ranges are broadly consistent with the spectra of the

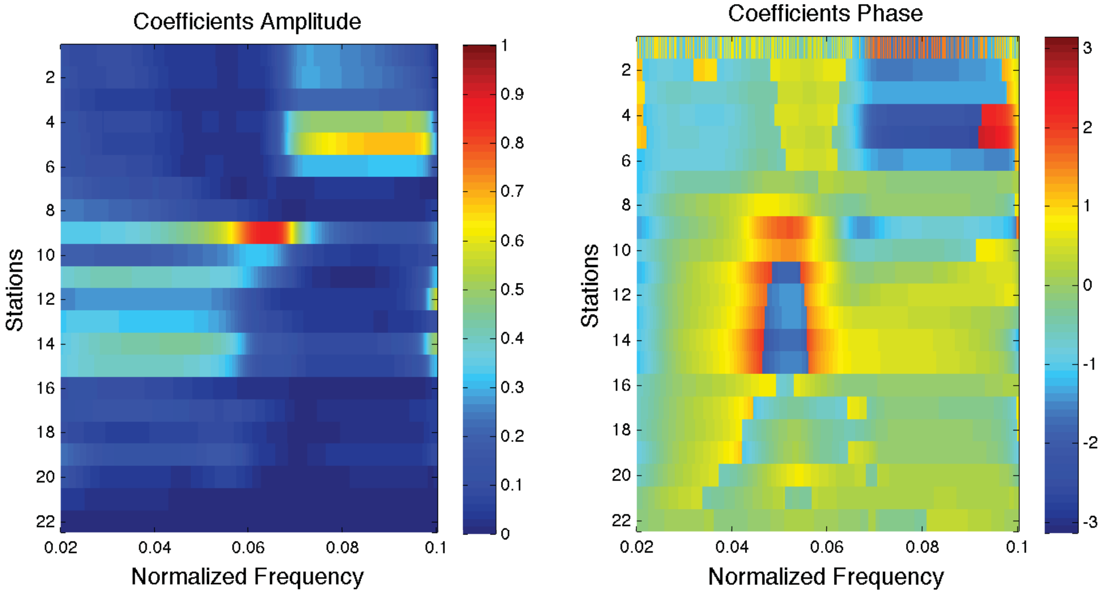

Figure 5.6: Amplitude and phase of the first eigenvector of the cross spectral matrix of η^{SLA} as a function of frequency. The plotted frequencies are between 0.02 and 0.1 cycles per hour. This is not true for the latest version of the plot.

first two principal components plotted in Figure 5.3.

The phases plotted in the right panel of Figure 5.6 suggests that the first group of stations (Gulf of Maine, periods of 17 to 10 hours) is approximately in phase, corresponding to a “standing mode”. The phase plot for the second group (Gulf of St. Lawrence, periods 50 to 17 hours) suggests a counterclockwise propagation around the Gulf of St. Lawrence. Station 16 is not in alignment with this propagation, presumably because this station is on an island near the center of the Gulf where the signal is weak while the other stations are distributed along the coastline where the signal is large.

In summary, PCA of the sea level anomalies in the time and frequency domain shows that the removal of high frequency variability using η^{Tide} and η^{Surge} is not sufficient: large scale signals remain at frequencies that are high enough to be aliased by the altimeter.

Chapter 6

De-aliasing of the Sea Level Anomaly

Spectral analysis of the sea level anomalies (SLA) in Chapter 5 showed that the tidal and dynamic atmospheric corrections do not remove all the high frequency signals in the hourly sea level variation observed at the tide gauges. There are high frequency signals left that could cause aliasing problems in the satellite altimetry. To de-alias the sea level anomaly products in the satellite altimetry, we need to remove these high frequency signals.

Table 6.1: Notations used in this chapter and their corresponding real data or variables (defined in Table 4.1).

Notation	Data/Variables Description
Δ	Sampling interval of satellite altimeter, $\Delta = 237.9754h$
G_1	tide gauge time grid, hourly interval from 01/01/1993 to 12/31/2014
G_2	satellite altimetry time grid, Δ interval from 01/01/1993 to 12/31/2014
X	tide gauge η^{SLA}
X^H	tide gauge η^{AH}
X^L	tide gauge η^{AL}
Y	altimetry η^{SLA}
Y^H	altimetry η^{AH}
Y^L	altimetry η^{AL}
$X\|G_1$	observed hourly tide gauge η^{SLA}
$Y\|G_1$	hourly altimetry η^{SLA}, not observed
$Y\|G_2$	observed altimetry η^{SLA} at sampling interval Δ

In this chapter, the de-aliasing methods introduce in Chapter 3 are applied to the sea level anomaly products. The methods are applied to observations from a single point on a given track. The de-aliasing results show the potential of the methods. The methods are then applied to the sea level anomaly products for a complete track. De-aliased sea level anomalies are presented as the final result of this study.

For clarity and convenience, the notations used in this chapter are inherited from

Chapter 3 although real data are used in this chapter. Table 6.1 defines these notations (also see Table 4.1).

6.1 Test of De-aliasing Methods

In this section, a single tide gauge and a single point on a given track are selected to test the three de-aliasing methods introduced in Chapter 3.

6.1.1 Data Description

The tide gauge station position and the track point position are marked in the map in Figure 6.1. For both tide gauge and altimetry records, time period from 01/01/2008 to 12/21/2012 is selected for maximum overlap. In this case, $X|G_1$ corresponds to the hourly observations of tide gauges and $Y|G_2$ corresponds to the altimetry observations with a sampling interval of Δ (see Table 6.1).

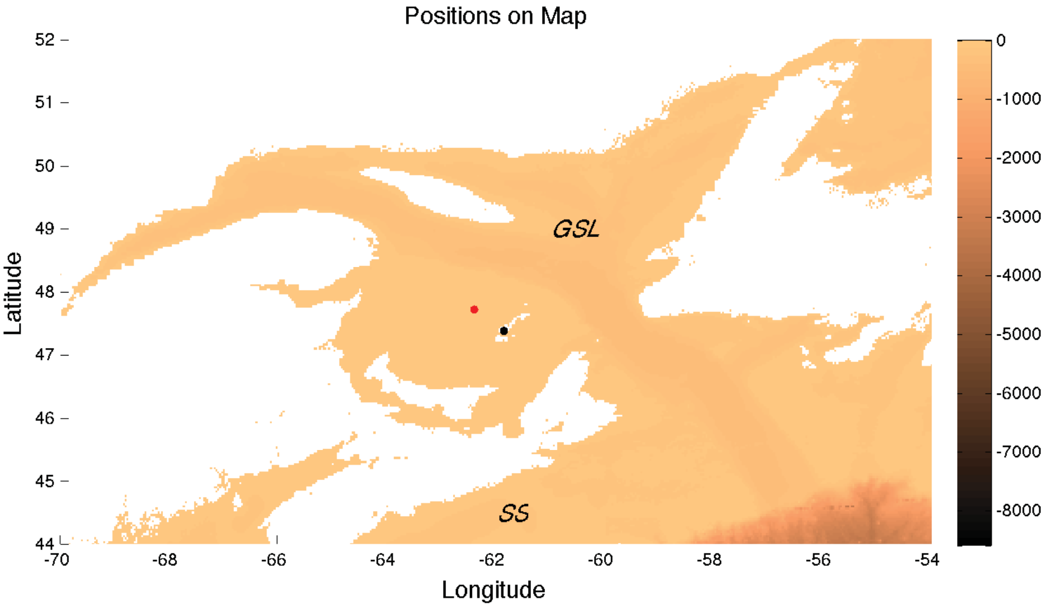

Figure 6.1: The tide gauge station position (black dot) and the satellite altimetry track point position (red dot). Tide gauge data are from the 16th station, Cap-aux-Meules. Satellite altimetry data are from one point on track 9, which is close to the station but still some distance off the coast. Both positions are inside the Gulf of St. Lawrence so that they both contain the 30-hour spatial-coherent signals.

Tide gauge data, $X^H|G_1$, used in this test are from the 16th station, Cap-aux-Meules, which is in the middle of the Gulf of St. Lawrence. Plot of the time series is given in Figure 6.2. This tide gauge record clearly contains many gaps. However, the 30-hour signal (see Chapter 5) is significant in the record as shown in the spectrum in Figure 6.2. As a result, the significant high frequency signals could be used as information for predicting the aliased high frequency signals in the altimetry. The number of valid hourly data points in the tide gauge record is 35465.

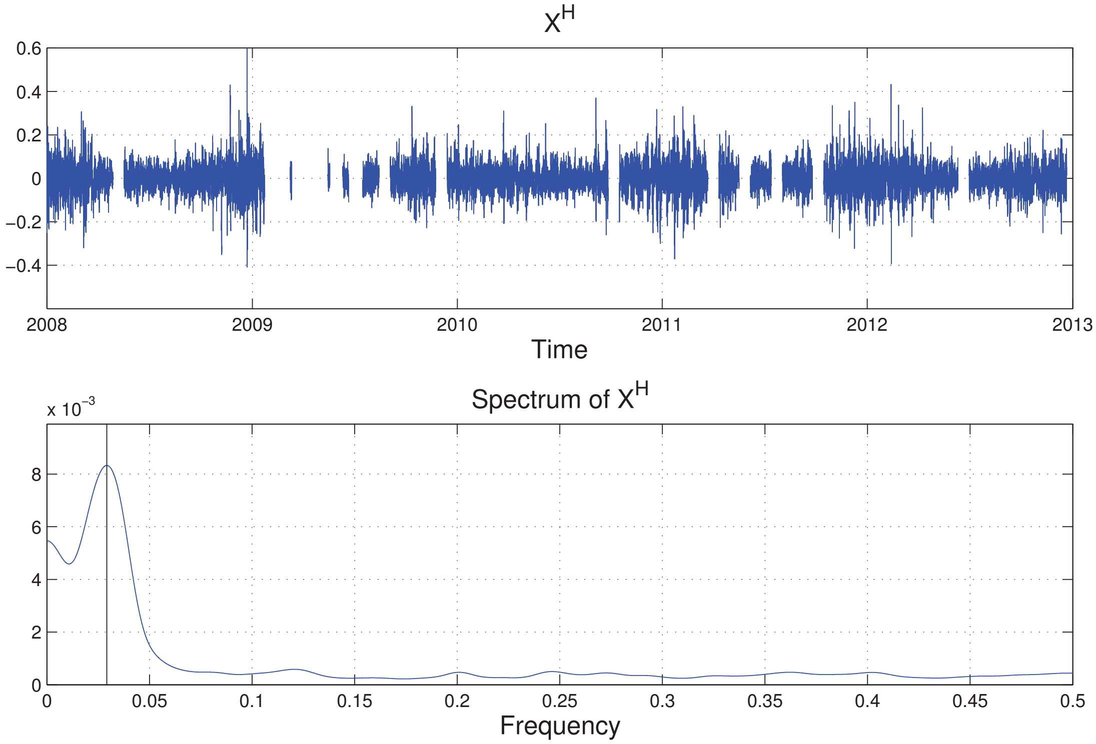

Figure 6.2: Time series plot (upper panel) and spectrum (lower panel) of $X^H|G_1$ for tide gauge record form station 16. Time period of data is from 01/01/2008 to 12/21/2012 for best overlap with altimetry data. There are numerous gaps in the record. The spectral estimation is based on a 100-point Parzen window. The 30-hour (marked with black line) signal is significant in this record.

Satellite altimetry record, $Y|G_2$ used in this test is from one point on track 9, which is close to the station, but some distance off the coast. The track point is not chosen to be the nearest to the station because satellite altimetry normally deteriorates very close to the coast. Plots of $Y|G_2$ and $X|G_1$ are shown in Figure 6.3. It is noticeable that $Y|G_2$ is aliasing $X|G_1$. The number of valid data points in the satellite altimetry record is only 157.

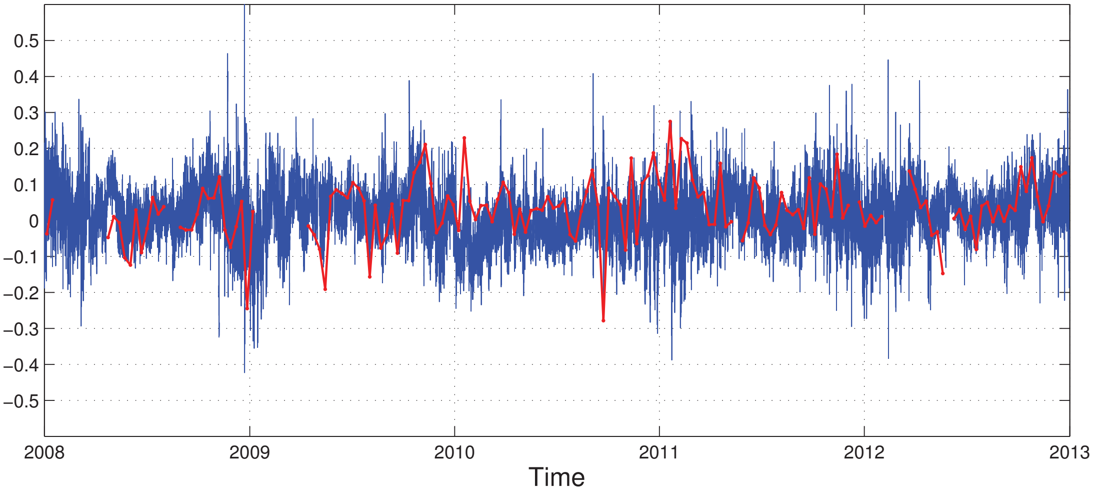

Figure 6.3: Time series plot of the satellite altimetry sea level anomaly (red line), and the hourly tide gauge sea level anomaly (blue line). Most high frequency variations are aliased by the satellite. Time period is from 01/01/2008 to 12/21/2012.

6.1.2 Lag Domain Method

Sample Auto-Covariance and Cross-Covariance: The lag domain method uses the sample auto-covariance function (ACVF) of $X^H|G_1$ to estimate the ACVF of X^H, and uses the sample cross-covariance function (CCVF) between $X^H|G_1$ and $Y|G_2$ to estimate the CCVF between X^H and Y^H. The sample ACVF and sample CCVF within 100 lags are plotted in Figure 6.4. The sample CCVF appears more erratic than the sample ACVF. This is due to different sample size used in estimation. The sample ACVF uses time series $X^H|G_1$ with data points up to 10^5, while the sample CCVF uses pairwise data points between $X^H|G_1$ and $Y|G_2$ where valid data points are only up to order 10^2.

Both the ACVF and CCVF are significant within a few lags and indicate periodicity of the processes. Moreover, the CCVF is highest at $k = 1$. This indicates that Y^H (the process at the track point) leads X^H (the process at the station point) by approximately one hour. Both ACVF and CCVF decay with increasing lag and become almost zeros after ±30 lags. This indicates that most of the information are contained within these lags. As a result, I chose $q = 20$ for parameter estimation and chose $p = 1, 2, \cdots, 15$ for the cross-validation process. That means we use the lags up to 35 to fit the model.

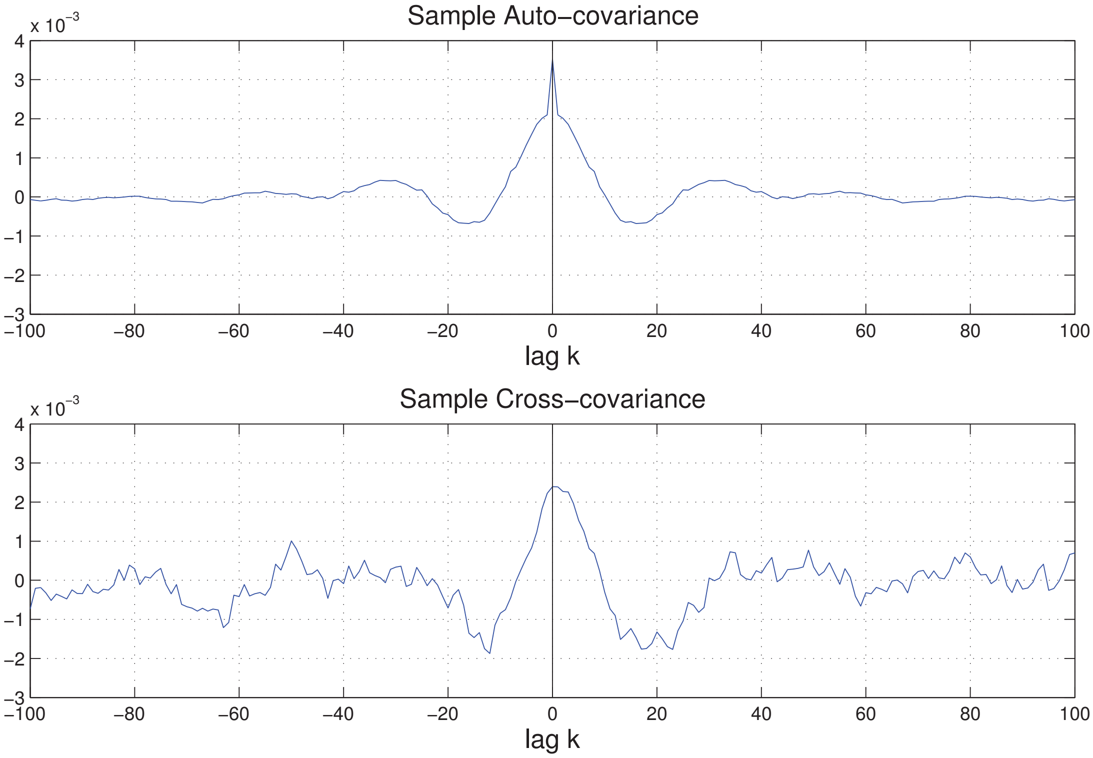

Figure 6.4: Plot of $\hat{R}_{X^H X^H}(k)$ in the top panel and $\hat{R}_{Y X^H}(k)$ in the bottom panel. $\hat{R}_{X^H X^H}(k)$ is estimated using hourly tide gauge records from station 16 and $\hat{R}_{Y X^H}(k)$ is estimated using this tide gauge record and the altimetry record from one point on track 9. Both are presented for ±100 lags. It is obvious that both $\hat{R}_{X^H X^H}(k)$ and $\hat{R}_{Y X^H}(k)$ decay for large lags and become effectively zero after ±30 lags. Note that the spike at $\hat{R}_{X^H X^H}(0)$ indicates noise in the record.

PARAMETER ESTIMATION: The model order p is first selected by a 5-fold cross-validation. Details of the implementation can be found in SC1 (Page 36 in Section 3.3). Plot of the residual variance resulting from the cross-validation procedures is given in Figure 6.5. This gives the optimal model order $p = 3$. Note that there are different ways of choosing the optimal order p in the cross-validation process. The common method is to simply choose the p giving minimum residual variance. Another typical method (Hastie et al., 2009, p. 244) also computes the standard deviation of a set of residual variance corresponding to each fold. Means and standard deviations for each possible p are both considered. In this method, the optimal p is chosen to give the most parsimonious model whose mean is no more than one standard deviation above the mean of the best model. In this study, the simple method is used.

With $p = 3$, the linear coefficients are then estimated by the weighted least square

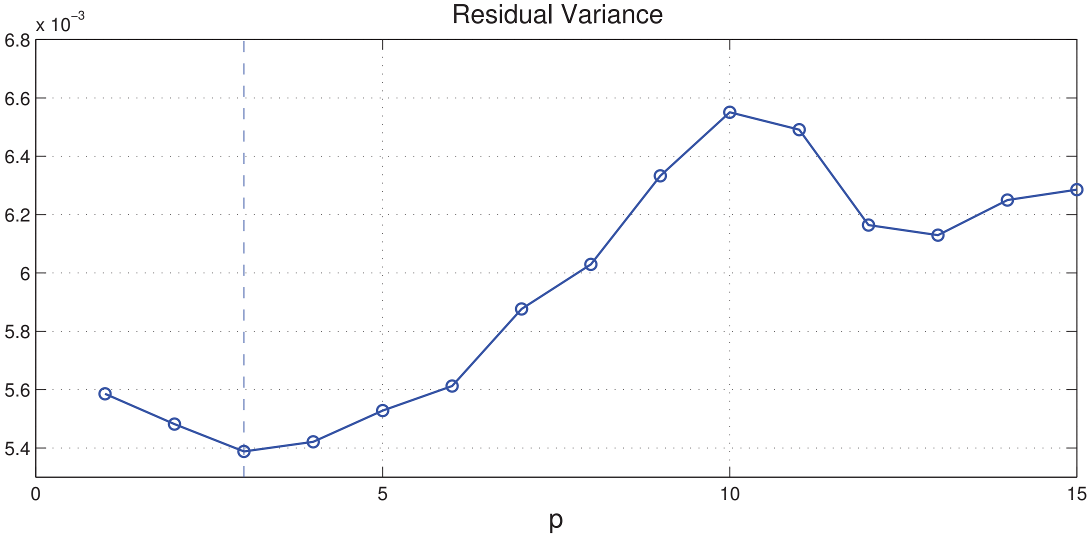

Figure 6.5: The cross-validated residual variance for the lag domain method. The residual variance has a minimum at $p = 3$.

method on the modified Yule-Walker Equations. The weight matrix W used here is a Parzen window weight sequence on the diagonal and zeros elsewhere. The Parzen window weight sequence can be derived from (2.34). The estimated linear coefficients of the lag domain method are plotted in the top panel of Figure 6.6.

The lag domain method gives a simple model. The predictors only involve ± 3 lags from the tide gauge records and the estimated linear coefficients has a waveform. The predictability decreases for large lags. This indicates a short 'memory'.

RESULTS: With the parameters estimated, model (3.3) can be used to predict Y^H from X^H, where X^H is the high frequency component of the tide gauge sea level anomalies. The predictions for Y^H is also a time series defined on the hourly grid G_1. Sub-sampling the predictions of $Y^H|G_1$ gives predictions on the satellite altimetry grid G_2, denoted by $\hat{Y}^H|G_2$. $\hat{Y}^H|G_2$ is subtracted from $Y|G_2$ to de-alias as in (3.5), where the de-aliased series is denoted by $Y^c|G_2$.

The de-aliased series given by the model is plotted in the top panel of Figure 6.7 on top of the original time series $Y|G_2$. It is obvious that $Y^c|G_2$ is less variable than $Y|G_2$ and contains less extreme values. The low frequency variability in $Y^c|G_2$ is thus more visible.

The standard deviations of $Y|G_2$, $\hat{Y}^H|G_2$ and $Y^c|G_2$ are listed in Table 6.2. The

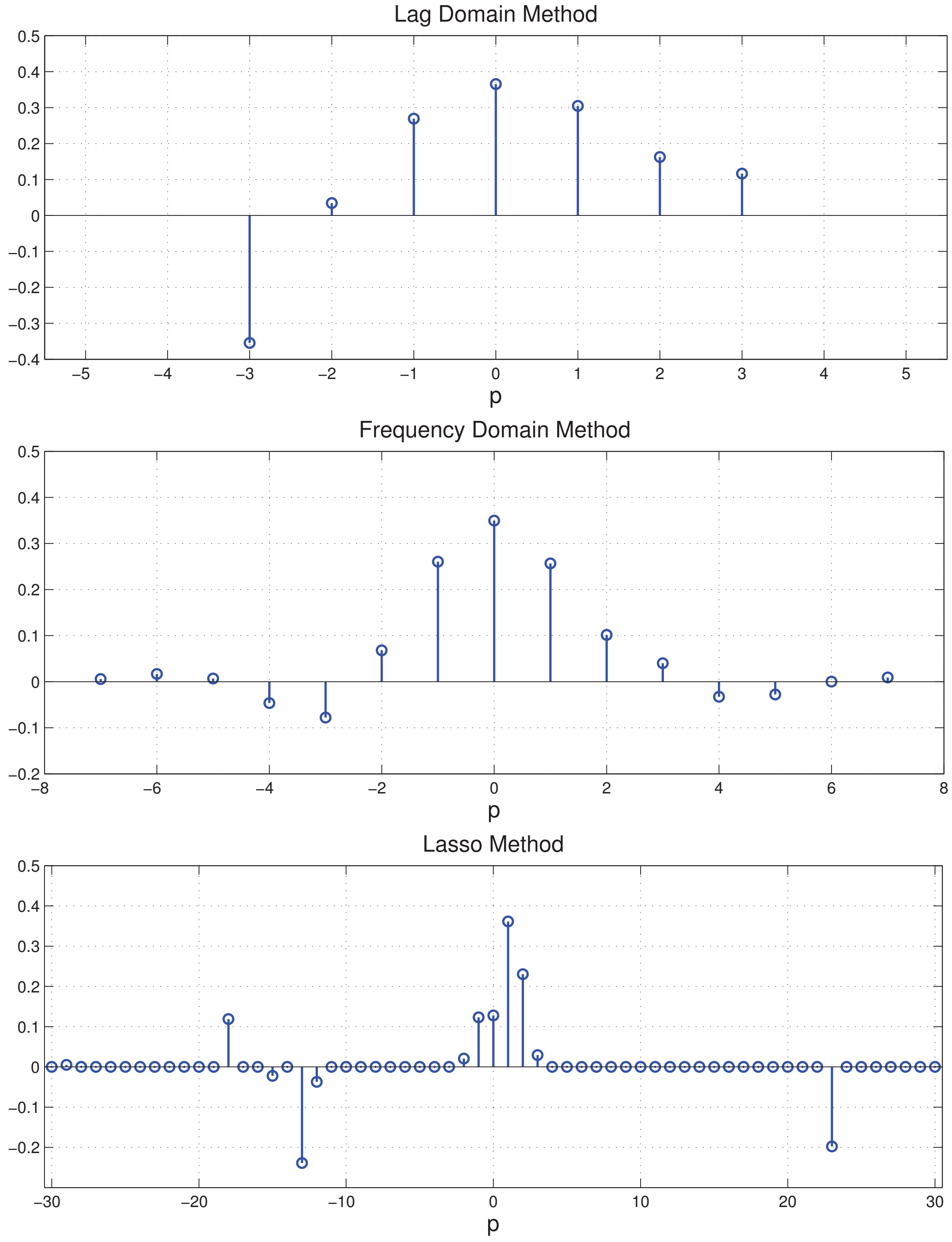

Figure 6.6: Plot of the estimated linear coefficients a_u $(u = -p, -p+1, \cdots, p)$ for the three de-aliasing methods. The lag domain and frequency domain methods both select predictors within a few small lags and the results are similar. But the frequency domain method gives a smoother shape of the coefficients. The lasso method also selects a small number of predictors but it can pick up some significant predictors corresponding to large lags.

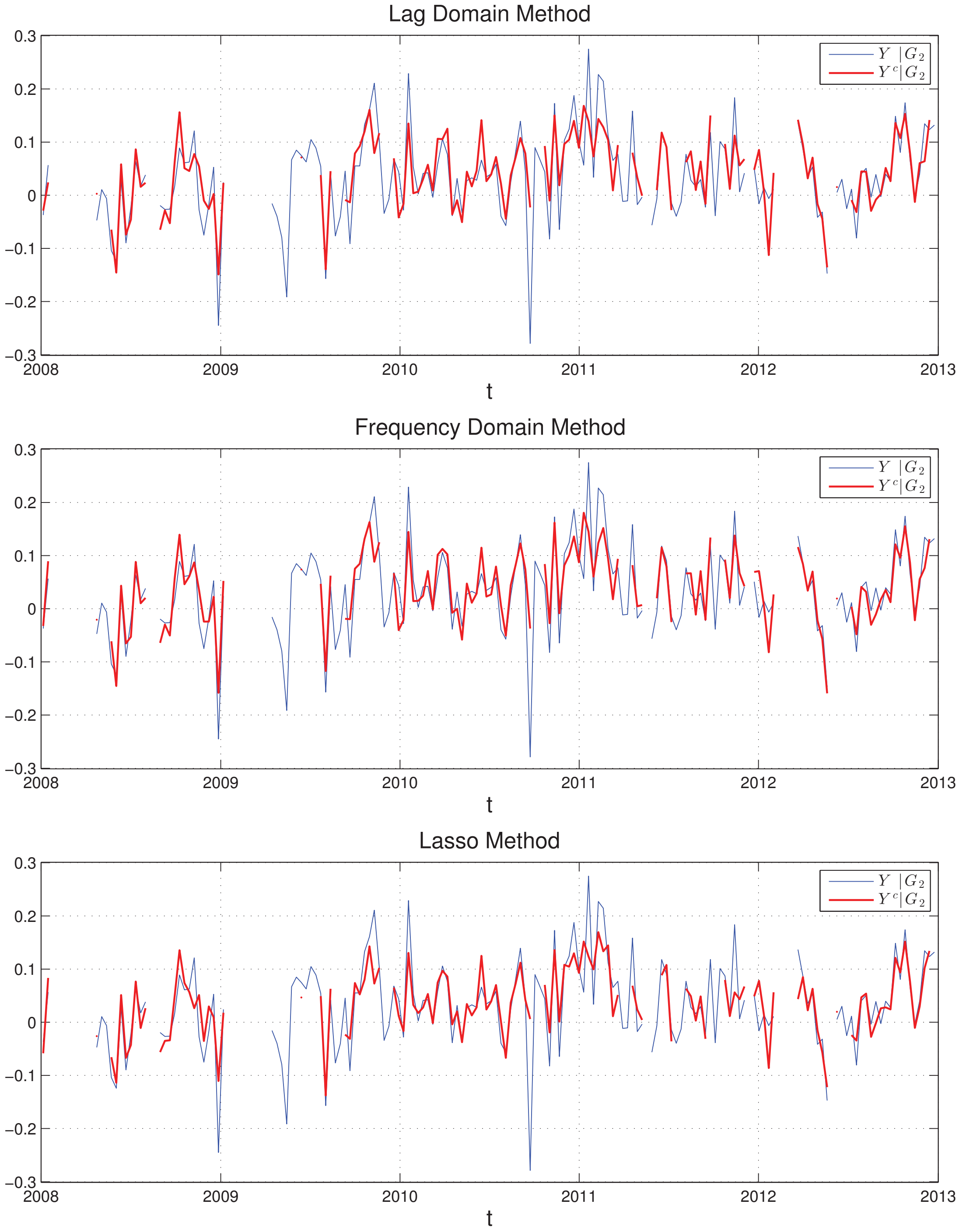

Figure 6.7: Plot of the de-aliased altimetry sea level anomalies (SLA) using the three de-aliasing methods (red line) and the original altimetry SLA (blue line). The de-aliased SLAs given by the three methods all reduce the variance of original SLA by reducing the spikes and extreme values. Results of the lag domain and the frequency domain methods are similar as expected. The lasso method performs better than the other two methods.

lag domain de-aliasing method reduces the standard deviation of the sea level anomaly by 1.55 cm.

Table 6.2: De-aliasing results of the three methods. Time series $Y|G_2$ corresponds to the original satellite altimetry sea level anomalies. $\hat{Y}^H|G_2$ is the predicted high frequency component in $Y|G_2$ and $Y^c|G_2$ is the de-aliased version of $Y|G_2$. The sample standard deviations (unit in meters) of these time series for the three de-aliasing methods are listed.

Time series	$Y\|G_2$	$\hat{Y}^H\|G_2$	$Y^c\|G_2$
Lag	0.0835	0.0483	0.0680
Frequency	0.0835	0.0442	0.0683
Lasso	0.0835	0.0498	0.0621

6.1.3 Frequency Domain Method

Parameter Estimation: The frequency domain method first selects a window width M (instead of p) using a 5-fold cross-validation process. Figure 6.8 shows the result of the cross-validation, where the minimum residual variance is at $M = 7$.

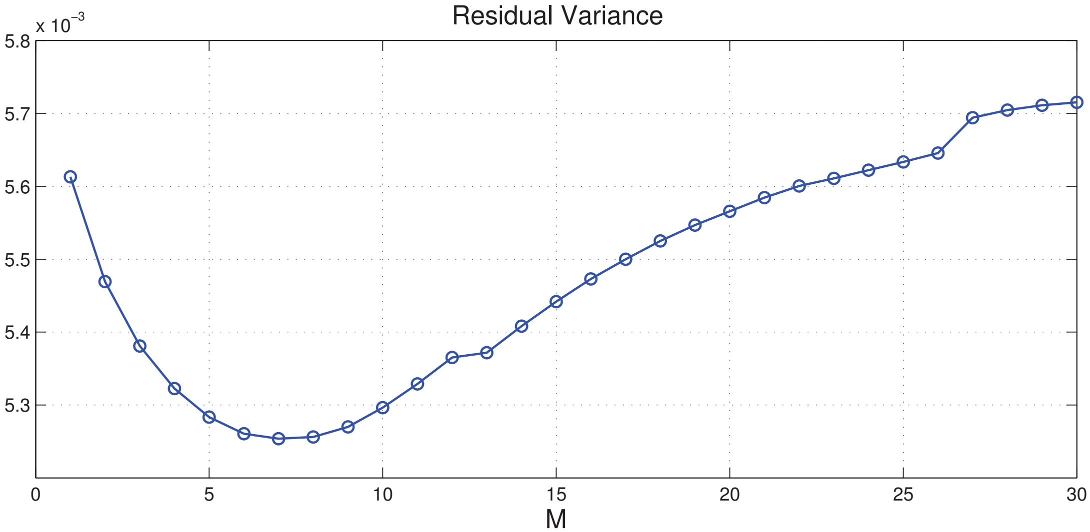

Figure 6.8: The cross-validation result for the frequency domain method. $M = 7$ gives the minimum residual variance and is used as the window width in the spectral estimations.

With $M = 7$, the linear coefficients are estimated from the inverse Fourier transform of the transfer function. The coefficients within $\pm M$ are estimated and plotted in the middle panel of Figure 6.6. The frequency domain method gives similar results as the lag domain method. The predictors are within a few small lags and the coefficients has a waveform. The difference is that the frequency domain methods gives a smoother shape at the 'tails'.

Transfer Function: The estimation for the transfer function is provided to give details of the frequency domain method and physical interpretations. The auto-spectrum $h_{XX}(\omega)$ is estimated from the sample ACVF, $\hat{R}_{X^H X^H}(k)$. The cross-spectrum $h_{YX}(\omega)$ is estimated from the sample CCVF, $\hat{R}_{YX^H}(k)$. For each spectrum estimation, a Parzen window with window width of $M = 7$ is used where M is selected by the cross-validation as explained above. Details of the estimation process is the same as in SC1 (Page 36 in Section 3.3). The transfer function is estimated and plotted in Figure 6.9.

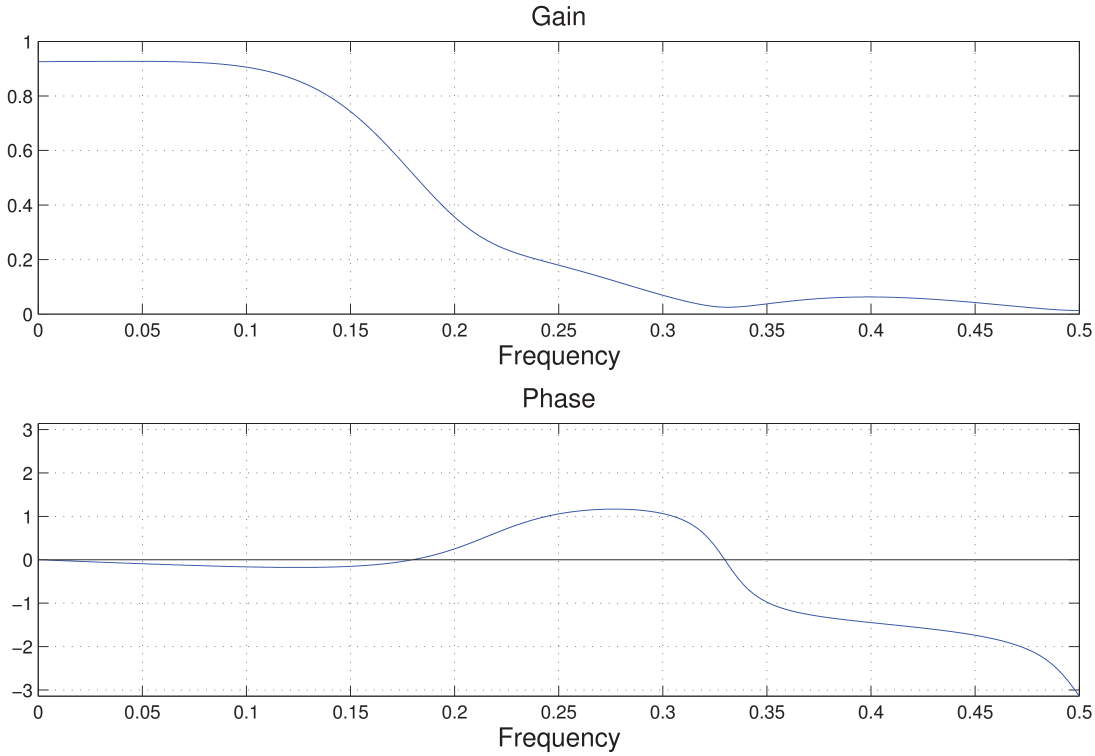

Figure 6.9: The gain and phase of the estimated transfer function on frequency range $[0, 1/2]$. The transfer function from the given tide gauge observations to the altimetry observations is estimated. The window width used in the estimation process is $M = 7$.

The gain of the transfer function indicates the amplification of signals at each frequency and the phase of the transfer function indicates the time delay relationship at each frequency. The amplification on the frequency band 0-0.15 (on a 0-1/2 scale) are relatively more significant than higher frequencies. Correspondingly on this frequency band, there is a small phase change. At frequencies higher than 0.15, the signals are attenuated. But noting that these signals contribute little to sea level variance (see spectrum of the tide gauge record in Figure 6.2), the overall effect is not significant.

Results: Similar to the lag domain method, predictions for $Y^H|G_1$ can be derived from the fitted model (3.3) and sub-sampling gives predictions for the aliased high frequency component $Y^H|G_2$. Subtracting $Y^H|G_2$ from $Y|G_2$ gives the de-aliased time series $Y^c|G_2$. De-aliasing results are in the middle panel of Figure 6.7 and the standard deviations of $Y|G_2$, $\hat{Y}^H|G_2$ and $Y^c|G_2$ are listed in Table 6.2. The frequency domain method reduces the standard deviation of the sea level anomaly by 1.52 cm. As expected, the results are similar to the lag domain method.

6.1.4 Lasso Method

Parameter Estimation: We now consider the problem as a regression problem and see how the lasso method works on real data. The response data vector is the satellite altimetry sea level anomaly $Y|G_2$, which includes 143 data points. The design matrix is constructed from the hourly tide gauge sea level anomaly $X^H|G_1$. The predictors include 61 lags, $X^H_{t-u}, u = -30, -29, \cdots, 30$, because higher lags are close to zeros (see Figure 6.4). As a result, the design matrix include 61 predictors and 143 data points for each predictors. However, as shown in Figure 6.3, there are some gaps and this leads to occasional NaNs for the predictors and response. Removing the missing data columns means that the number of valid data points is 149.

The lasso method chooses the significant predictors by constraints on the coefficients. A 5-fold cross-validation process was used to estimated λ. The optimal value was 0.006. In this way, the optimal model is selected to give minimum residual variance. The linear coefficients estimated by the lasso method are given in the bottom panel of Figure 6.6.

The lasso method selects the predictors according their effect on the residual variance. Only a few predictors are preserved. Similar to the lag domain and the frequency domain methods, the $a_u, u == -1, 0, \cdots, 4$ indicate a short 'memory' and also a small time delay of the signals. Different from the other two methods, the predictors corresponding to large lags are selected by the lasso method. Results show that there are some significant predictors corresponding to large lags. For example, the $a_u, u = -13, 23$ are not picked up by the other two methods. However, they indicate amplifications of signals at certain frequencies. Actually, analysis of the prediction for the aliased component show that the 30-hour signal as indicated in Chapter 5 is picked up by the lasso method. Overall, the lasso method performs better than the other two methods at detecting this specific signal.

RESULTS: Similar to the other two methods, $Y|G_2$ can be de-aliased with the fitted model. Results are in the bottom panel of Figure 6.7. The standard deviations of $Y|G_2$, $\hat{Y}^H|G_2$ and $Y^c|G_2$ are listed in Table 6.2. The lasso method reduces the variance of the sea level anomaly by 2.14 cm. The improvement compared to the other two methods is also explained by the 30-hour signal.

6.2 Application of De-aliasing Method

Results of the last section showed promise for these methods. In this section, the frequency domain method is applied to a full satellite track and the de-aliasing results are assessed.

6.2.1 Data Description

Sea level anomaly time series from tide gauge station 13, Rivière-au-Renard, was used to de-alias the sea level anomaly products from satellite altimetry track 9. This track contains 176 points. The tide gauge station position and the track points are plotted on the map in Figure 6.10. Tide gauge station 13 was selected because this record is relatively more complete (see Table 4.2 for proportion of valid data of each tide gauge record).

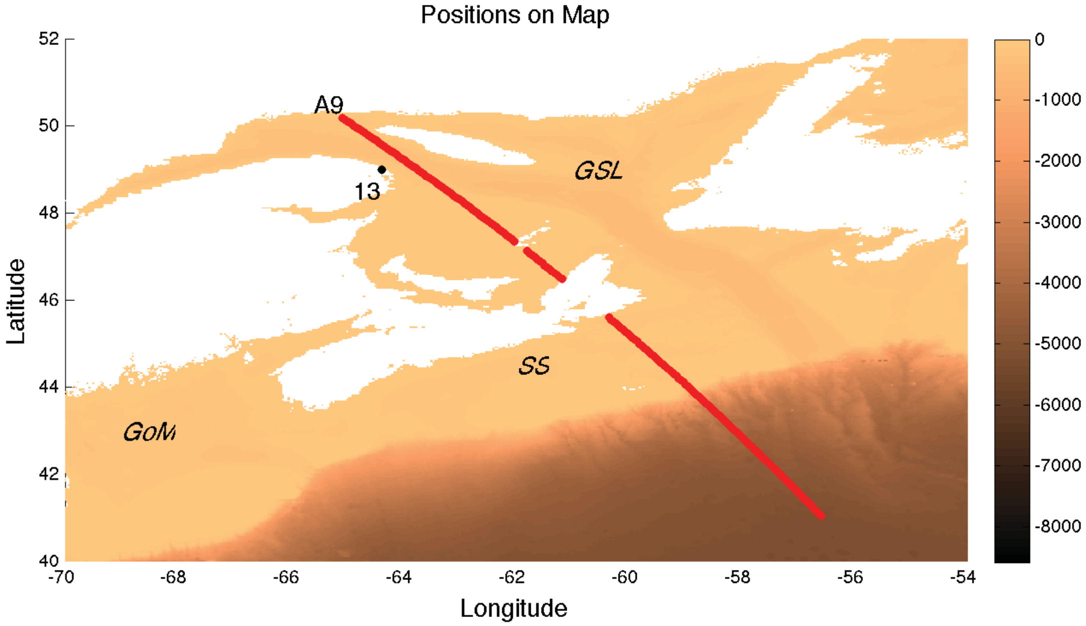

Figure 6.10: Plot of the tide gauge station position (black dot) and satellite altimetry track (red dot). Tide gauge data are from the 13th station, Rivière-au-Renard. Satellite altimetry data are from track 9 and this track contains 176 points. The tide gauge station is inside the Gulf of St. Lawrence. The satellite altimetry track, however, contains points both inside and outside the Gulf of St. Lawrence.

6.2.2 De-aliasing Results

The lasso method was applied to each point on track. Specific details of implementation are the same as in SC1 in Section 6.1.3. Figure 6.11 shows the de-aliasing results for track 9. For each point on track, the standard deviation of sea level anomaly (SLA) product η^{SLA} $(Y|G_2)$ and standard deviation of the de-aliased SLA $(Y^c|G_2)$ are plotted in the upper panel of Figure 6.11. The bathymetry corresponding to the track point positions are plotted in the lower panel.

The de-aliasing method normally reduces the standard deviations by 1-2 centimeters for track points inside the Gulf of St. Lawrence (longitude range between $-65°$ and $-62°$). SLA is not improved as significantly for track points outside the Gulf of St. Lawrence, especially in deep ocean area. This is expected. The main reason is that a large portion of the aliased high frequency variations came from the 30-hour signals in the Gulf of St. Lawrence (as indicated in Chapter 5). In addition, the points on the track have different distances from the tide gauge position; it is expected that

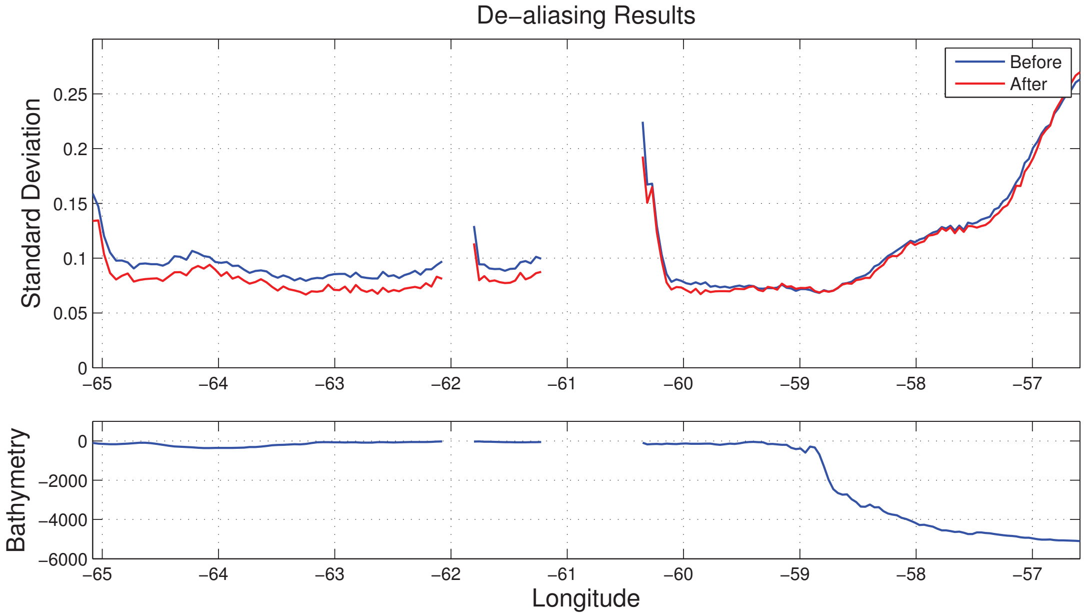

Figure 6.11: Standard deviation of sea level anomaly before and after de-alising at each track point (upper panel) and the bathymetry corresponding to each track point position (lower panel). The de-aliasing method reduces the variance of sea level anomaly, especially at positions inside the Gulf of St. Lawrence. The standard deviations went down by approximately 0.02 m. For positions outside the Gulf of St. Lawrence the de-aliasing results are not significant.

the information transferred from a distant position should attenuate and thus be less predictable. To illustrate this, we calculated the correlation between SLA of the tide gauge and SLA along the satellite track and results show that the correlation goes down as distance increases. Note that the gaps in the figures corresponds to lands. The variance of the sea level is high in coastal areas and deep ocean areas.

Figure 6.12 shows the SLA time series before and after de-aliasing for track points between −64.5° and −62.5° where the de-aliasing method gives good results. The de-aliasing technique smoothed out some of the extreme observations including e.g., a red part (extremely high values) in year 2011 between −63.5° and −62.5°, a blue part (extremely low values) in year 2002.

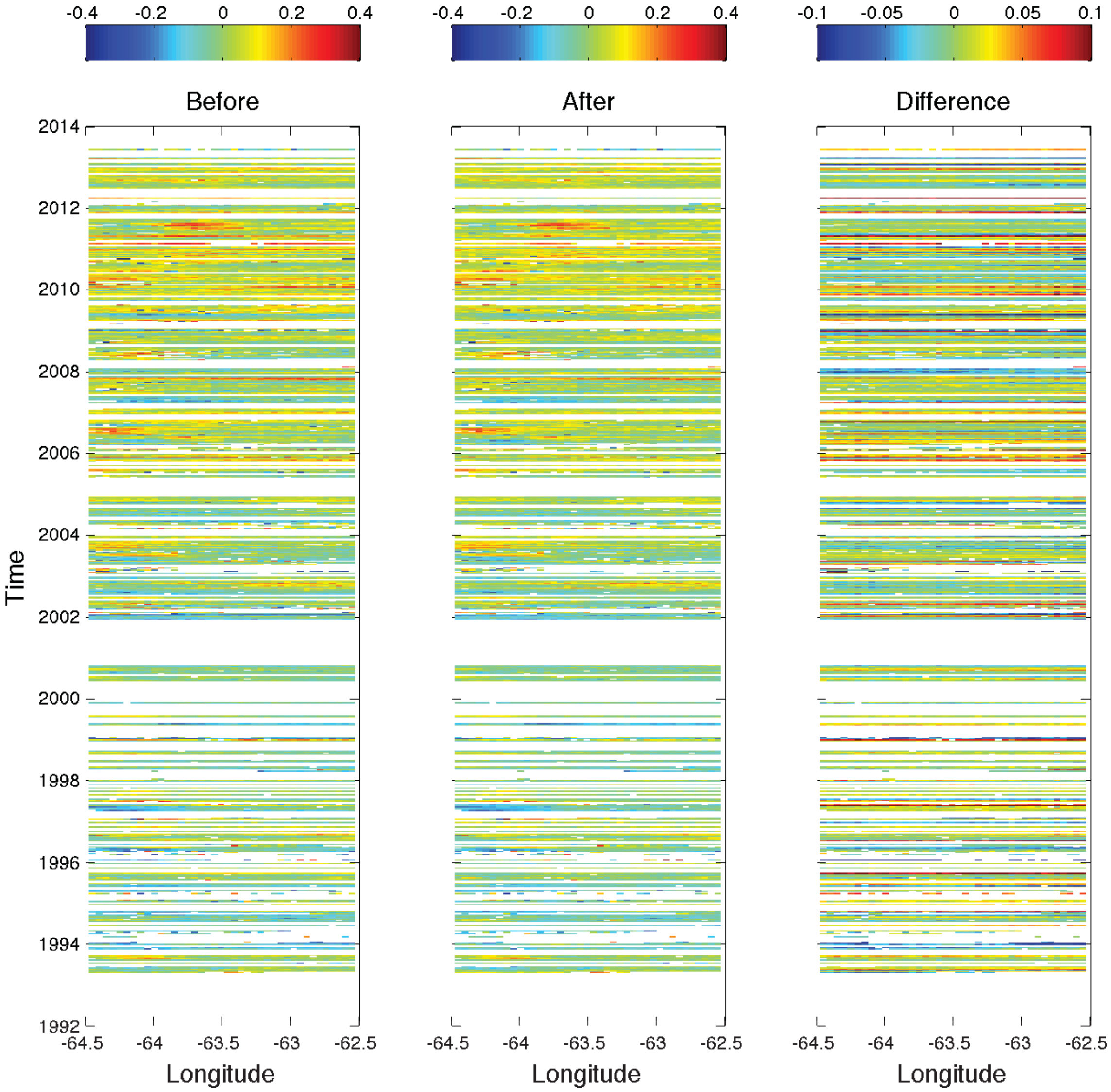

Figure 6.12: Altimetry sea level anomalies before (left panel) and after (middle panel) de-aliasing and the difference between them (right panel). The difference is also the predictions of the aliased high frequency component in the altimetry sea level anomalies. Data are for longitude range from −64.5° to −62.5°.

Chapter 7

Summary and Future Work

7.1 Summary

Satellite altimeters sample sea level about every ten days and thus alias the high frequency variability of sea level. Conventional de-aliasing techniques in oceanography remove the high frequency signals (above the Nyquist frequency of satellite altimetry) using dynamical model predictions. These predictions are ocean model outputs and account for the deterministic high frequency signals due, for example, to tides and surges. Tidal corrections (predictions for tides) and dynamic atmospheric corrections (predictions for surges) nowadays are very effective, usually reducing the standard deviation of sea level variability from meters to centimeters (Chapter 4).

However, these corrections do not account for all of the high frequency signals in sea level. The analysis presented in Chapter 5 indicates other significant high frequency signals, in addition to tides and surges, that are potentially aliased by satellite altimeters. Specifically, hourly tide gauge observations after corrections for tides and surges (termed as sea level anomaly, SLA) indicate there are large scale, high frequency signals left in the sea level which could cause an aliasing problem. Principal component analysis was applied to multiple tide gauge SLA time series to determine the frequency and spacial characteristics of the remaining signals. Several structures were found, including an amplified 30-hour signal within the Gulf of St. Lawrence.

In order to effectively de-alias the satellite altimetry SLA product, it was concluded additional techniques must be designed. It is natural to base the new techniques on the tide gauge records used in the principal component analysis above. Compared to the altimetry data, the tide gauge records are sparse in spatial distribution. However, as hourly records, they provide more information on the high frequency sea level variability. This inspired the development of three statistical de-aliasing methods, where tide gauge records are used to predict the aliased high frequency component in altimetry SLA.

The statistically-based prediction model assumes stationarity of sea level variations and a linear relationship between the altimetry and tide gauges measurements. This is equivalent to a linear relationship between the high frequency component of the tide gauge SLA, denoted by X^H, and the satellite altimetry SLA, denoted by Y. The predictors in the model are X^H within $\pm p$ lags and the response is Y. Parameters in the model include $2p+1$ linear coefficients associated with the predictors stepping from $-p$ to p (see Chapter 3).

Three methods were designed to fit the model (Chapter 3). The lag domain method adapts the modified Yule-Walker equations (Shumway and Stoffer, 2010). The frequency domain method transforms the problem to the frequency domain where the transfer function is estimated and then transformed back to get the coefficient estimates for each predictor. The lag domain and the frequency domain methods are similar but implemented differently. A critical step in the lag domain and frequency domain methods is the estimation of the cross-covariance function between X^H and Y. The estimation is possible using the special techniques discussed in Section 3.2.1. The lasso method treats the estimation as a regression problem. All three methods use cross-validation to determine the optimal model order.

The three methods all end up with a fitted model. Although the model response is the satellite altimetry SLA (or Y), the predictions are for the high frequency component of Y (denoted by Y^H) because the low and high frequency components are uncorrelated (see Section 2.3.3). With the fitted model and the hourly tide gauge SLA, the predictions are also hourly. To de-alias, the predicted time series is subsampled at the altimetry sampling interval and then subtracted from the satellite altimetry SLA products.

The mathematical statement of the statistical de-aliasing model including the three methods of parameter estimation is presented in Chapter 3. Two simulation cases are used to evaluate and compare the performance of the three methods. All methods give promising results as measured by the comparison of the prediction of the high frequency component with the truth (which is available in the simulation cases).

Finally, the statistical de-aliasing methods were applied to the satellite altimetry SLA and the tide gauge SLA. In order to show how these methods work, the methods were tested on a single tide gauge station and a single point on track. Both the tide gauge and altimetry positions are located within the Gulf of St. Lawrence so that they both contain the 30-hour signal. Details are given in Section 6.1. All methods gave similar de-aliasing results in terms of reducing the sea level variance. Specifically, the lag domain method reduced the standard deviation of sea level variation by 1.55 cm, the frequency domain method by 1.52 cm and the lasso method by 2.14 cm. Overall, the lasso method gives most effective and interpretable results in this study. The lasso method was then applied to a complete satellite altimetry track using the record of one tide gauge station. De-aliasing results are stable. The standard deviations at track points inside the Gulf of St. Lawrence are normally reduced by about 2 cm. The reductions in standard deviations outside the Gulf of St. Lawrence are small, as expected.

7.2 Future Work

The most obvious next step is to apply the de-aliasing methods to all the satellite altimetry tracks to improve the overall quality of the altimetry product. However, as mentioned in Section 6.2, it is not easy to determine a universal model for all the track points based on one tide gauge. This statistical de-aliasing method is a local method, where local tide gauge information is used to predict the satellite altimetry information. As a result, the quality of the predictions is highly dependent on the choice of tide gauge. Sea level at each track point may contain high frequency signals not present at the tide gauge station. As a result, more flexible and effectives technique need to be developed to extract specific information from certain tide gauge records, with which altimetry products for a specific track point can be de-aliased.

The statistically-based model in this study can be improved. One promising way would be to include more tide gauge stations as predictors. As shown in Figure 5.6, the dominant 30-hour signal among the aliased signals is propagating around the Gulf of St. Lawrence. A single tide gauge record may not fully capture the spatial characteristics of the propagating signal. Predictions based on multiple tide gauge stations will likely be much more effective.

Additional improvements may also include extension and generalization of the dealiasing model. The current model has a very simple form as is based on the assumptions of stationarity and linearity. Stationarity of the underlying processes admits spectral representation theorem discussed in Theorem 1. Linearity is the cornerstone for the relationship between the coefficients and transfer function (Equation 2.38). It is also the basis for the zero correlation between the low and high frequency components (Section 2.3.3). Both assumptions are used in the development of the present model. In practice, stationarity of the time series used in the application process was approximated by simply removing a linear trend. However, it would be more precise to take into account seasonality of the sea level variability. It would also be interesting to consider generalization to a nonlinear model.

Finally, my discovery of the spatially-coherent signal within the Gulf of Maine and Gulf of St. Lawrence is already stimulating further research in ocean modelling. Eventually it will be necessary to model these local high frequency signals deterministically, based on physical principles.

Bibliography

User Handbook AVISO. Merged topex/poseidon products. *Romonville St-Agne, France*, page 201, 1996.

Loren Carrère and Florent Lyard. Modeling the barotropic response of the global ocean to atmospheric wind and pressure forcing comparisons with observations. *Geophysical Research Letters*, 30(1275):4, 2003.

Loren Carrère, Florent Lyard, M Cancet, A Guillot, and Laurent Roblou. Fes2012: A new global tidal model taking advantage of nearly 20 years of altimetry. In *Proceedings of meeting*, volume 20, 2012.

Dudley B Chelton, John C Ries, Bruce J Haines, Lee-Lueng Fu, and Philip S Callahan. Satellite altimetry. *International Geophysics*, 69:1–ii, 2001.

Fabien Durand, Doraiswamy Shankar, Florence Birol, and S Satheesh Chandra Shenoi. Estimating boundary currents from satellite altimetry: A case study for the east coast of india. *Journal of oceanography*, 64(6):831–845, 2008.

Benjamin Friedlander and Boaz Porat. The modified yule-walker method of arma spectral estimation. *Aerospace and Electronic Systems, IEEE Transactions on*, (2):158–173, 1984.

Adrian E Gill. *Atmosphere-ocean dynamics*, volume 30. Academic press, 1982.

Trevor J.. Hastie, Robert John Tibshirani, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009.

JD Horel. Complex principal component analysis: Theory and examples. *Journal of climate and Applied Meteorology*, 23(12):1660–1673, 1984.

R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007. ISBN 9780131877153. URL `https://books.google.ca/books?id=gFWcQgAACAAJ`.

Daniel R Lynch and William G Gray. A wave equation model for finite element tidal computations. *Computers & fluids*, 7(3):207–228, 1979.

Emanuel Parzen. Mathematical considerations in the estimation of spectra. *Technometrics*, 3(2):167–190, 1961.

Rich Pawlowicz, Bob Beardsley, and Steve Lentz. Classical tidal harmonic analysis including error estimates in matlab using t_tide. *Computers & Geosciences*, 28(8): 929–937, 2002.

K. B. Petersen and M. S. Pedersen. The matrix cookbook, nov 2012. URL `http://www2.imm.dtu.dk/pubdb/p.php?3274`. Version 20121115.

Sylvie Pouliquen. In-situ observations: Operational systems and data management. In *Ocean Weather Forecasting*, pages 207–227. Springer, 2006.

M.B.M.B. Priestley. *Spectral Analysis and Time Series.* Probability and mathematical statistics. Academic Press, 1981. ISBN 9780125649223. URL `http://books.google.ca/books?id=0Ge0ngEACAAJ`.

Ian Robinson. Satellite measurements for operational ocean models. In *Ocean Weather Forecasting*, pages 147–189. Springer, 2006.

Laurent Roblou, Julien Lamouroux, Jérôme Bouffard, Florent Lyard, Matthieu Le Hénaff, A Lombard, Patrick Marsaleix, Pierre De Mey, and Florence Birol. Post-processing altimeter data towards coastal applications and integration into coastal models. In *Coastal altimetry*, pages 217–246. Springer, 2011.

R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications: With R Examples.* Springer Texts in Statistics. Springer, 2010. ISBN 9781441978653. URL `https://books.google.ca/books?id=NIhXa6UeF2cC`.

S Vignudelli, P Cipollini, Laurent Roblou, Florent Lyard, GP Gasparini, G Manzella, and M Astraldi. Improved satellite altimetry in coastal systems: Case study of the corsica channel (mediterranean sea). *Geophysical Research Letters*, 32(7), 2005.

Stefano Vignudelli, Andrey G Kostianoy, Paolo Cipollini, and Jérôme Benveniste. *Coastal altimetry.* Springer Science & Business Media, 2011.