LOW ORDER MODELING OF SEEMINGLY RANDOM SYSTEMS WITH APPLICATION TO STOCK MARKET SECURITIES

A Dissertation

by

ARUN SURENDRAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Aerospace Engineering

LOW ORDER MODELING OF SEEMINGLY RANDOM SYSTEMS WITH APPLICATION TO STOCK MARKET SECURITIES

A Dissertation

by

ARUN SURENDRAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Othon Rediniotis
Committee Members,	Rodney Bowersox
	Tamas Kalmar-Nagy
	Luis San Andres
Head of Department,	Dimitris Lagoudas

December 2009

Major Subject: Aerospace Engineering

ABSTRACT

Low Order Modeling of Seemingly Random Systems with

Application to Stock Market Securities. (December 2009)

Arun Surendran, B.Tech., Indian Institute of Technology, Mumbai;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Othon K. Rediniotis

Even simple observation of stock price graphs can reveal dominant patterns. In our work, we will refer to such re-occurring, dominant patterns as “coherent structures”, a term borrowed from the theory of turbulence in fluid dynamics. Stock price performance exhibits coherent structures, which by definition make it non-random, although a price-versus-time graph might seem totally chaotic to the naked eye.

A novel low order modeling technique for systems that are seemingly random has been developed. Though stock market data is used for the formulation and verification of the technique, its application in diverse fields is verified. The dissertation discusses some of the salient features of the novel technique along with a dynamic system analogy. The technique reduces many of the significant limitations associated with traditional methods like Fourier analysis and digital filters. Application of the technique to a nonlinear dynamical system and meteorological data are presented as well as the primary application on stock market securities.

ACKNOWLEDGMENTS

No academic work would be complete without acknowledging *the guru*, a duty I am glad to fulfill by expressing my heartfelt gratitude to Dr. Othon Rediniotis. I am also grateful to my committee members Dr. Luis San Andrés, Dr. Tamas Kalmar-Nagy and Dr. Rodney Bowersox for sparing time from their busy schedules to be part of my academic endeavor.

It would be borderline arrogance to consider it humanly possible to list the names of everyone who has helped me along the way. Innumerable good souls have been of immense help. So I would borrow the opening verse of the famous classical Carnatic devotional prayer song that goes, "*Entharo Mahanubhavalu, Anthariki Vandhanamulu*" meaning "My Salutations to All the Magnificent Manifestations of Life".

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

A new low order modeling technique for systems that are seemingly random has been developed. Though stock market data was used for the formulation and verification of the technique, its application in diverse fields has also been tested. This introductory chapter is divided into three sections. Section 1.1 provides a glimpse into the historical connection between the field of finance and time series analysis. Alongside we shall also discuss the limitations of existing tools and the significant limitations the new technique has managed to eliminate. In the subsequent sections of this introduction, the fields of technical analysis and the rapidly growing discipline of behavioral finance from which provides the theoretical understanding for the success of the technique are introduced.

The next chapter provides an outline of the Proper Orthogonal Decomposition and discusses a dynamical system that provides an analogy for the essential working of the new technique. Chapter III showcases the results of the application of the mathematical tool on a nonlinear dynamical system. Chapter IV details the application to stock market securities along with the graphical interface developed for presentation of results that can assist in decision making. In Chapter V, we present the application of the technique on meteorological data.

This dissertation follows the style of *Journal of Spacecraft and Rockets*.

1.1 A Brief History of Time-Series Analysis

Since the new technique is essentially one of time-series analysis and was developed primarily on stock market data, it is interesting to find from a brief historical survey how closely tied time series analysis has always been to markets and money. The second purpose of this discussion is to bring to light how activities in diverse fields have contributed to development of mathematical tools. Hidden in such diversity is the danger of applying what was fairly logical and purposeful in one context in some other area with disastrous effect.

Statistical historian Judy Klein acknowledges in "Statistical Visions in Time" [1] that history of time series indicates a primacy of "monetary arithmetic" developed to comprehend or participate in monetary exchanges in several cases and specific techniques were later appropriated for statistical method and theory.

The roots of modern time-series analysis seem to lie in the simple Rule of Three documented by London cloth merchant John Graunt in his "Mathematics of my Shop-Arithmetique" in 1662. The Rule of Three or Merchant's rule is essentially a comparison by division. Every school student knows it today as the simple ratio technique which forms the core of several types of data manipulation in modern statistics. Though the technique was used extensively in India and China as far back as 250 A.D. and Italian merchants of 15th century considered it crucial, Graunt receives credit for meticulous data collection over decades and using their manipulation to affect the realms of business and politics. Graunt was also familiar with comparison by subtraction, which forms the essence of all regression analysis. But the full potential of first differences had to wait for 19th century financial speculators [2].

By 1860, The Economist was publishing several tables which included columns showing increases and decreases of commercial data with positive and negative signs on top of absolute values. For speculators and political economists this short term view was sufficient and on such oscillations they could apply the new tools of frequency distributions, correlation and regression. The 19th century also saw the application of the method of averaging as a method of smoothing, as used by Bank of England in its publishing of bullion values, Jevons' study of periodic commercial fluctuations, and Poynting's study of drunkenness [3]. It appears that the "drunkard's random walk" used to describe modern day stock market movements has fairly mundane origins!

Time-series analysis continued to grow through the studies of Norton, Hooker, Yule and the like in early twentieth century all the way up to the current standard Box-Jenkins approach of Seasonal Integrated AutoRegressive/Moving Average (SARIMA) analysis [4, 5, 6].

Parallel to this growth was the development of probability theory or studies of chance. The pioneering works of Bernoulli, Pascal and others on the law of large numbers and games of chance gave way to studies of target practice (Stochastic means to aim at a mark) by prominent Russian scientists like Kalmogorov (famous for his work on turbulent flows). In fact almost all of the prominent names in the field of science like Laplace, Fermat, Lagrange, Gauss and Botlzmann contributed to the field of probability and estimation at some point in their career [1, 4, 7].

But the trouble with all the sophisticated tools that statistics has developed from primitive algorithms of real life practitioners is the reduction of time-series into what Klein rightly calls "a measurement without history"[1]. The algorithms are decomposed till a stationary process is obtained which can be subjected to analysis. We casually make assumptions about the existence of mean or seasonality. The circular nature of statistical approach is also evident in use of probability distribution where the nature of distribution is already assumed in the sampling, outliers removed and then distribution considered valid. Curve fitting inherently assumes the existence of a curve which may not exist in reality. The information discarded along the way of data decomposition severely limits the utility of anything produced using the tools. There exist severe problems in reconciling methods based on static laws of deviation with dynamically fluctuating data streams.

Specifically in the case of application of cutting edge mathematical and statistical tools to financial and economic data, this glaring defect has come into being time and again. In the late 1990s, the entire economy of United States went into crisis because of the failure of the mega hedge fund called Long Term Capital Management [8, 9]. The fund was created by the gentlemen behind the Nobel Prize winning Black-Scholes equation which "estimates" the pricing of financial instruments called options. This was not an isolated incident. There will always be an element of “miscalculation” in hindsight involved in every financial boom and burst. Such experiences provide ample humility and cautiousness to anyone setting out to develop new techniques for modeling and forecasting.

The approach of the novel technique goes back to the roots of financial algorithms. It limits itself to short-term utility and stays flexible by exposing itself mercilessly to real time data without resorting to any underhand “smoothening” beyond what is necessary for computational ease. It acknowledges the limitations of knowledge when it comes to profoundly complex systems like the stock market or weather. Therefore there are no equations or formulas to fit all cases. It sticks to a single input data stream from which all further computations are made without any other data used for comparison. The salient features of the technique are further discussed in the next chapter.

The intention of this section was to emphasis the need for caution at every step in the development and application of new mathematical techniques. The next two sections on technical analysis and behavioral finance intend to show why it is crucial to have mathematical models and forecasting techniques for navigating extremely complex dynamical systems like the stock market. *Technical Analysis*, the field that mathematically approaches the market is discussed in the next section. In section 1.3, we briefly survey the new field of “Behavioral finance” that brings home the need to trust mathematics more than our own minds when it comes to decision making!

1.2 Technical Analysis of Stock Market Securities

The utility of the stock market to the developing of mathematical modeling technique is succinctly captured in the following statement from Benoit Mandelbrot, the father of fractal geometry: “Data, a gold mine of data.”[10]

With electronic trading networks and globally interlinked stock markets, the amount of data generated with every passing moment in the market is enormous [11]. Stock market provides an inexhaustible source of data for anyone interested in exploring new tools for mathematical/statistical modeling, pattern recognition, probability theories etc. The field that develops and uses mathematical techniques for the modeling and subsequent prediction of stock prices is called *Technical Analysis* [12, 13, 14].

Any stock can be considered as a piece of ownership of the company behind that stock. The analysis of the value of stock based on the fundamentals of the company behind it like the annual revenue, quality of management, profit/earnings ratio, market capital, growth rate etc is called *Fundamental Analysis*. In contrast, Technical Analysis completely ignores the company and treats the stock price simply as a historical time series. Stock traders and investors usually use a combination of both fundamental and technical analysis to make their decisions to buy, sell or short sell their holdings.

The existence of 24/7 news channels and apparently infinite amount of information out there would naturally create doubt about the ability of mathematical tools to forecast stock market movements. A casual observation of the news headlines and market behavior will put any such doubts to rest. In November of 2008 after the announcement of the largest unemployment figures in the US, the market was up by a couple of percentages. In June, 2009 on the day, the iconic American company General Motors announced its bankruptcy; the Dow Jones Industrial average climbed over 200 points!

Stock market news reports routinely suggest that market behaved in a certain way "because" of some news or "despite" some news. Intelligence begs one to throw away any notion of causality after observing such statements. However, when something specific happens to a company, clearly there will be movement of money in and out of its stock. So any use of technical analysis should come with a caveat of not completely depending on it alone on days the company announces specific information like its earnings.

Technical Analysis could be as simple as the recognition of commonly recurring patterns or shapes in the price graph and using it to anticipate the next movement [14]. There are common patterns like "dead cat bounce" shown in figure 1 and "head and shoulders" pattern shown in figure 2 in the next page. Simplistic mathematical curves like short and long term moving averages are also commonly used. More technically minded traders and investors make use of other tools like Elliot waves and Bollinger bands. It may not be much of an exaggeration to say that almost all stochastic modeling tools if not developed for and from the stock market at least have been tried there!

Figure 1: Example of a "dead cat bounce" pattern in the market. Figure courtesy: Investopedia.com.

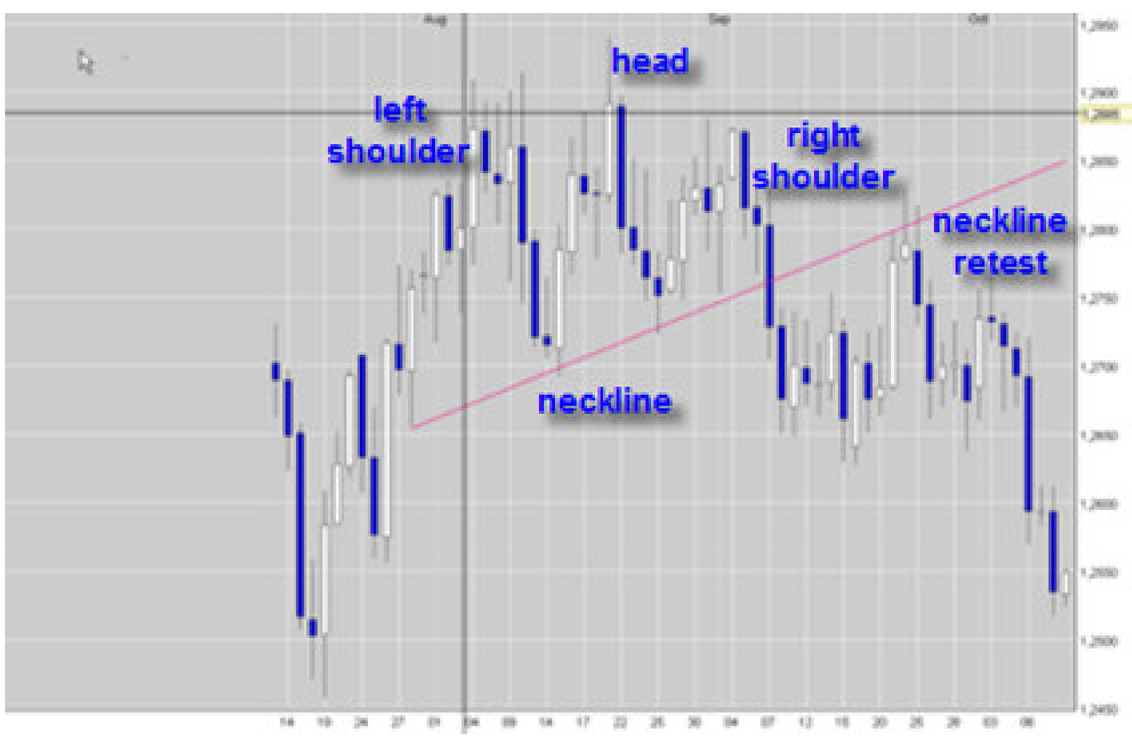

Figure 2: Head and shoulders pattern.

Gaussian curve is priceless in engineering applications and games based probability models but all it takes is one outlier point in economics to wipe out entire livelihoods and industries. Our cognitive flaws which lead to the disappearance of the "rational" when it comes to money will be discussed in the next section on behavioral finance. In contrast, rapidly fluctuating markets have provided incredible profit making opportunities for short term traders. Without getting into a moral or ethical argument about the merits of short term profit making, one can clearly see that mathematical tools based on recent historical data can be created to provide excellent assistance to a disciplined trader to predict and benefit from impending swings in the price.

The current (2008-9) economic downturn has brought home the realization to everyone with their retirement funds tied to the market that they are better off using the market as a wealth creator rather than a preserver. For those with engineering aptitude and simple mathematical background, technical analysis provides tools which can diminish the psychological drama associated with managing one's money through frequent decision making. Stock market is a gamble only if you treat it like one. But then so is any other activity or decision in life if we choose to perceive it so!

With rapid advancements in computational speeds and memory sizes, technical analysis has taken off to dizzying heights in the last couple of decades. It has come a long way from simple trend line fitting through Hurst's "combing filters" to Jim Simon's Renaissance Technology which manages close to $100 billion in funds and accounts for nearly 10% of volume traded in NASDAQ on some days [10, 15, 16].

The rapidly generated streams of data can at once provide a challenge and an opportunity to any modeling technique. With electronic trading, every tick up or down in the price change is permanently noted. Thus high resolution time series is readily available. However, keeping in mind the processing speeds for both modeling software and execution speed of trading decisions, price data collected every few minutes is a reasonable highest resolution. This can then be clubbed with end of day data, weekly, monthly and even annual data which have been traditionally used for developing techniques before the advent of computers.

With that brief introduction to Technical Analysis of the market, we can approach the emerging field of behavioral finance which provides theoretical insights into why mathematical models and computerized trading have a better chance of success and will be the way of the future.

1.3 Behavioral Finance

> "Perception is a choice of which we are not aware, and we perceive what has been chosen."
>
> Daniel Kahnemann, Nobel Prize lecture, 2002 [17]

Daniel Kahnemann, a psychologist at Princeton University, was awarded the Nobel Prize for Economics in 2002. Along with Dr. Amor Tversky, Dr. Kahnemann's path breaking research in the early 70s led to the new subject of behavioral finance. Their work in the topics of heuristics of judgment, risky choice and framing effect has led to the complete abandoning of Bernoulli's Expected Utility Theory, the model which held sway in the field of decision making for over 250 years.

Bernoulli theorized that human beings are perfectly rational when it comes to predicting the utilities and assigning values. As exposed by Kahnemann and Tversky [17], the fundamental flaw in that theory was that the model was reference independent. It takes very little observation to find out that everything is relative about the way we perceive the world around us. But according to Bernoulli both a millionaire and a mendicant should and will assign equal value to one hundred dollars and subsequently base their decisions on such valuation.

There was no reference point in utility function of wealth in Bernoulli's model, but the new prospect theory redraws the curve exposing the glaring loss aversion of human mind as shown in figure 3. Prospect theory attempts to understand decision making involving risk dealing with real life situations rather than optimal decisions or probabilistic games.

The curve reproduced below at once explains the reason for the two questions authors of "Sway: Irresistible pull of Irrational Behavior" [18] ask, "Why do we find it so difficult to let go of a failing relationship? Why is it so tough to sell a stock losing value?" The answer to both is the incredible loss aversion our mind has. We are terrified at the prospect of a loss which guides our actions much more than an objective calculation of the true probabilities of the outcome.

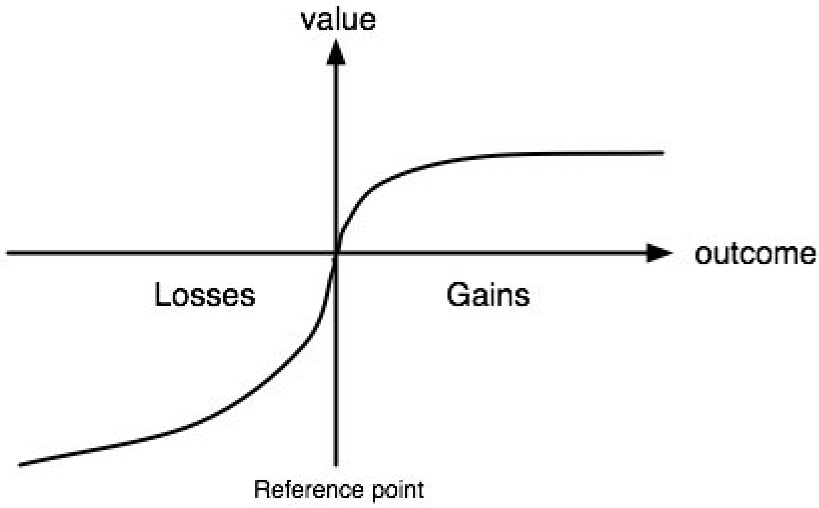

Figure 3: Value function according to prospect theory. Notice the rapid change in the value associated with losses.

As the opening quotation of this section suggests, we are all bundles of cognitive flaws. A brief survey of cognitive flaws will draw up a list of over fifty very common ones. We will mention a couple of them here not only because they are hugely entertaining but are relevant in the field of stock market and help us see the significance of having mathematical models.

Anchoring: In an experiment conducted in 1974 by Kahnemann and Tversky [19], subjects were asked to estimate the percentage representing African states in the United Nations. They were divided in several groups with each group assigned a random number between 0 and 100. One group received number 10 and another number 65. The subjects had to indicate whether the percentage of African nations falls above or below the number assigned to them. In the next stage of experiment, they were asked what the

actual figure they had in mind was. The average estimate of the first group was 25%, close to the random number 10 and the second group was 45%, close to the random 65.
The influence of random numbers or first sources (the numbers at the beginning of the sequence) on our subsequent decision making is substantial because we rarely adjust our estimates based on additional information when it comes available. This flaw severely affects decision making in the stock market because for every trade, there is an entry price and there are hundreds of other numbers floating around at the same time into any of which we can unknowingly get anchored to.

Representativity: In a 1972 experiment researchers asked subjects which order of birth is more likely in the case of a young couple with six children. Using G for girl and B for boy, which of the following order is more likely:
GGGGGG or GBBGGB?

Most people consider the first sequence less probable but from a mathematical standpoint they are clearly wrong. This cognitive flaw comes from our quickness to think in patterns. When an observation fits a pattern or a schema that we carry in our mind, we overestimate its probability [19, 20].

Though spiritual giants since Siddharatha Guatama, "The Buddha" (circa 500BC) have insisted on the importance of maintaining a silent mind, modern psychology finds almost all us with "diagnosis bias" where we label something in our mind which completely shades the clarity of all future observations we make about it. Similarly there is the fallacy of "value attribution" where the initial perceived value of something makes us imbue a person or thing with qualities they do not possess.

Just by considering the above mentioned cognitive problems, it is clear that machines will be able to do a better job at financial decision making like frequent buying and selling of stocks. Mathematical schemes fed into computers, they will be executed without any of the perceptional errors so common to man. This makes the development of mathematical modeling techniques and their application in the stock market almost a necessity for anyone hoping to generate wealth.

The stock market is a system where hundreds of thousands of minds are colliding with all their cognitive flaws and decision making about the stock's value every moment. So it is no surprise that a stock price is in continuous state of fluctuation. In the market, mass psychology is in action and the patterns in human behavior manifest themselves as patterns in the price movement which can be mathematically mined out. This does not mean that we are challenging the concept of free will or reducing human beings into mechanical entities. Our discussion is clearly limited to human behavior with regards to financial markets.

Equally important is the mechanics (or rather electronics these days) of the activity of trading. Though millions of shares get traded every day, the average size of a trade still remains close to a few thousand shares. Even the big funds that move millions of dollars in and out of the market with each trade have to operate in small chunks. This is important because if massive movement of money is noticed somewhere, then that itself will lead to change in behavior and price and thus the funds will no longer be able to exploit the price movement they were aiming for. Thus we can imagine several channels of different speeds and volume through which money enters and exits the market. To borrow an analogy from legendary investor Phil Town, the big funds are like ocean liners who need a big radius to turn around [21]. A smaller investor can consider herself to be on a jet ski and quickly turn around once she notices the ocean liner turning.

The time invariant smooth scales which the new technique, hereby developed, extracts can be looked at as such channels of money of different frequencies (ocean liners, ships, boats and jet-skis) instead of harboring notions about the mathematical predictability of stock price behavior!

At the outset, it might seem like a contradiction that we are arguing for the predictability of a system while maintaining that it has randomness. Since caveman began painting the nature around him some 30,000 years ago, we have used static imaging to study the world around us. This propensity of using this tool has been reflected in mathematical techniques as well where we use time-freeze on dynamic data streams to extract patterns and identify usable hidden information. However with the advent of video filming and other special effects like time-lapse and slow-motion, we are beginning to look at the world around us as a dynamical system that “behaves”. Today, we are seeking mathematical tools that can capture adaptive behavior.

From the second law of thermodynamics and Heisenberg's uncertainty principle, we are well aware of the inherent randomness in nature and our inability to observe with absolute accuracy. However, probabilistic behavior does not deny causality. It is true that when we estimate probability of future life events in our mind especially in money matters, we are by mistake measuring the strength of our confidence rather than objective probability. Behavioral finance experiments vividly bring home this point. But consider an experiment like the simple coin toss. The probability of the outcome being a head or tail is an inherent property of the coin. Thus there are systems for which inherent properties help us quantify possibility of future events. In our technique, we are searching for such inherent time-invariant properties of the system that manifest as the scales. Once we have identified such scales and verified their relative time invariance, we can venture forth to quantify how the system will respond to the external environment through these modes of behavior.

With advancements in the field of genetics and the introduction of the "Ontophylogenisis" theory [22], it appears that evolution through natural selection operates even at a cellular level. Man could be fundamentally a random machine. But this randomness is constantly subjected to an external and internal environment where the chance of erratic behavior is much reduced by the existence of memory, physical evolution, habits and patterns in behavior.

It has been proved for quite a while now that almost all life has a biological clock that leads to phenomena like circadian rhythm. We know that several species of plant seeds monitor the external temperature and other conditions to determine the right time to sprout. A two thousand year old Magnolia seed found in Japan had patiently waited for modern day scientists to provide it with the right environment to germinate [23]. A human being by the sheer size of his brain and complexity of organ system is definitely the product of much more of nature's investment than a seed or a crow. So clearly the range of behavior exhibited by man will be much more complex. Yet it is a very pertinent question: under a particular environment or in the face of a specific stimulus, how much freedom does even man have to vary his behavior? And how much of that freedom survives when humans behave as a group?

The purpose of this introduction has been to simultaneously stress the need for new mathematical techniques and the importance of the care to be taken while transferring techniques from one discipline to another. The single most important feature of the new technique is that a very tangible physical dynamical system analogy exists for its operation. This analogy shows that the level of abstraction and data manipulation has refrained from "idealizing" reality. The next chapter discusses this dynamical model analogy along with some other salient features of the new technique.

CHAPTER II

DYNAMICAL SYSTEM ANALOGY FOR THE TECHNIQUE

This chapter introduces the dynamical system that provides an analogy for the operation of the mathematical technique. In section 2.1 we present a brief overview of the commonly used Proper Orthogonal Decomposition (POD) technique. POD and digital filtering techniques were studied during the very initial stages of the research [24]. The idea of extracting coherent structures or modes hidden in seemingly random data comes from POD and similar techniques [25, 26]. But as is explained in section 2.2, the new technique starkly differs from the existing ones in the kind of structures that it seeks to extract. That section lists the other important features of the technique as well. We follow that discussion with the introduction of the dynamical system analogy for the technique in Section 2.3. This analogy leads into the first application of the technique into a nonlinear dynamical oscillator presented in Chapter III.

2.1 Proper Orthogonal Decomposition – An Overview

POD first formulates a set of basis functions that are specific to the system of interest. They are not like the generic sine and cosine basis functions that a Fourier series uses to approximate any system/response. The POD functions are specific to the given system and should not be used to approximate another system.

In our applications [24], the POD technique is based on a collection of "snapshots" of the time series. Let us considered N (N=80) L-point (L=200) snapshots, regularly shifted with respect to each other. Each snapshot is a column vector $\vec{V}_i$, $i = 1,...,N$, with L elements. After the snapshots are prepared, we compute the correlation matrix defined as

$$C_{ij} = (1/N)(\vec{V}_i \cdot \vec{V}_j)$$

where $i, j = 1, 2, ..., N$ and

N: total number of snapshots

$\vec{V}_i$: snapshot vector i

The eigenvalues and eigenvectors of the above matrix are solved for:

$$CA = \lambda A, \qquad A = (a_1, a_2, ..., a_N)^T$$

$$A^K \cdot A^{K'} = \begin{cases} 1/(N\lambda_K), & K = K' \\ 0, & K \neq K' \end{cases}$$

By construction, C is a nonnegative symmetric real matrix and has a complete set of orthogonal eigenvectors $A^1, A^2, ..., A^N$ with the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$. The POD basis functions then are expressed as linear combinations of the snapshots:

$$\vec{\Phi}_K = \sum_{i=1}^{N} a_i^K \vec{V}_i, \quad 1 \le K \le N$$

$$(\vec{\Phi}_K, \vec{\Phi}_{K'}) = \begin{cases} 1, & K = K' \\ 0, & K \ne K' \end{cases}$$

where a_i^K is the i element of the eigenvector A^K corresponding to the eigenvalue λ_K. The basis functions, or modes $\vec{\Phi}_i$, can be interpreted as the "coherent structures" hidden in the system's behavior. The sum of all the eigenvalues is regarded as the total "energy" of the time series. Thus if most of the "energy" is contained in the first *M* POD modes (*M<<N*), i.e.

$$\sum_{i=1}^{M} \lambda_i \cong \sum_{i=1}^{N} \lambda_i$$

then we may achieve order reduction by approximating the time series with only the first *M* POD basis functions in the following approximation for the exact series $\vec{S}$

$$\vec{S} \approx \vec{S}_p = \sum_{k=1}^{M} y_k \vec{\Phi}_k$$

where $\vec{S}$ is an L-point section (L=200 here) of the exact time series, while $\vec{S}_p$ is its approximation or "projection" by using only the first M modes. The scalars y_k are the projection coefficients and are derived by taking the inner product of the exact price time series $\vec{S}$ with the basis function $\vec{\Phi}_k$ and normalizing the result with the square of the magnitude of $\vec{\Phi}_k$:

$$y_k = (\vec{S} \cdot \vec{\Phi}_k)/(\vec{\Phi}_k \cdot \vec{\Phi}_k)$$

2.2 Some Salient Features of the New Technique

In this section, we will introduce the salient features of the new technique. POD, digital filtering and Fourier analysis techniques will be frequently used for comparison to expose their limitations which have been avoided.

Complex phenomena, such as turbulence, stock market behavior, behavior of biological systems, etc., have a wide spectral range and therefore a wide range of time scales. Moreover, it is unwise to truncate the high frequencies; behavior at the higher frequencies is essential and affects the behavior at the lower frequencies. In Reynolds Averaged Navier-Stokes (RANS) for example, problems arise from our attempt to average out the fast scales (closure problem). In biological systems (let us consider specifically humans), protein folding usually happens at much faster time scales (down to a few microseconds) than the time scales human behavior is typically perceived at (from seconds to years). However, for many proteins the correct three dimensional structure resulting from the folding process is essential to function [27]. Failure to fold into the intended shape usually produces inactive proteins with different properties. Several neurodegenerative and other diseases are believed to result from the accumulation of mis-folded (incorrectly folded) proteins [28].

Stock market behavior is no different: modeling using data down to the minute time scale results in much better prediction than modeling using only data on the daily time scale (i.e. considering only daily closing prices). In order to model the entire spectrum, however, a single POD analysis of the available time series (in the stock market case, the most time-resolved time series we consider consist of minute data) results in “obscuring” of the fast scale dynamics. This makes sense since a single POD analysis is asked to model time scales differing by several orders of magnitude. The faster modes resulting from this single POD analysis are “buried” by the slower modes with significantly larger amplitude. This single POD analysis still captures the majority of the system’s energy in only a few dominant modes. But the problem stems from the fact that even the faster and low energy modes that would be thrown away in a low order modeling approach, are essential for the system dynamics even in the slow/large scales; not unlike receptivity problems in fluid mechanics.

In our modeling we have utilized a multi-resolution approach. In this approach, the entire spectrum is first split into sub-spectra (the sum of which results in the original spectrum). Each sub-spectrum is then POD modeled, but now the retained modes (typically from 2 to 6 modes) are all on the same order of magnitude (both in amplitude and frequency). One of the challenges in determining the optimal “splitting” of the spectrum is to determine the range and boundaries/cut-offs of each sub-spectrum, such that the non-retained modes/dynamics are not essential in affecting the behavior of the system in the slower frequencies/spectra. As demonstrated in the next chapters, we have developed a rather rigorous procedure for identifying the optimal number and range of the sub-spectra; a procedure that is independent of the specific system, since it, itself, identifies the important and non-important system dynamics and accordingly splits the spectrum. Another challenge we have faced relates to the proper “bridging” of the sub-spectra, in the synthesis phase (ultimately the dynamics of the individual sub-spectra have to be synthesized to yield a comprehensive system model over the entire spectrum). We now have a reliable procedure for merging the scales/spectra, irrespective of their number or range.

Besides physics challenges, we have also faced mathematical challenges. For example, the “splitting” of the spectrum requires band-pass filtering with very sharp drop-offs and minimal phase lag, for obvious reasons. None of the available filtering techniques worked sufficiently well for our purposes. We thus developed a spectral decomposition technique with significant advantages over existing techniques. For example, compared to Fourier decomposition, our technique does away with the periodicity assumption of Fourier and has infinite frequency resolution, unlike Fourier which has very limited frequency resolution in the low frequencies [29, 30]. The significant improvement on issues like end-effects and frequency leakage are demonstrated in the next chapter.

Coherent structures are simply the result of underlying laws (physical, psychological, financial, etc.). However, these laws are often very hard to model and express in a deterministic manner, due to their complexity/nonlinearity and sensitivity to initial and boundary conditions. Therefore, we often resort to the next best thing, which is a phenomenological model of a system's behavior.

Consider an individual’s behavior. There are laws governing neural networks and neurotransmitter behavior, laws about the transfer of energy-electrical signals in the nervous system, heat production and transfer from metabolic processes, mass transfer, e.g. blood flow- in the body, hormonal secretions, protein production, etc., along with the accompanying initial conditions (genetic heritage) and boundary conditions (the stimuli from the individual's environment). However, these laws are so complex and often so sensitive to initial and boundary conditions that the individual's behavior is (to-date) impossible to deterministically model. But futurists like Ray Kurzweil and others continue bold attempts [31, 32]. Nevertheless, an individual's behavior exhibits "coherent structures" which are repeated and often predictable, for example, anger, compassion, hunger, fear, hope, etc. Therefore, in this sense, psychology can be thought of as a phenomenological study of these coherent structures in order to model and predict an individual’s behavior in different situations.

Now consider the theory of turbulence in aerodynamics. Until a few decades ago the task of modeling turbulence and predicting its behavior seemed so overwhelmingly complicated that it seemed impossible. "Coherent Structures" was the most important tool in turbulence modeling and prediction. This was a phenomenological approach. Today, with the dramatic advances in computational power, we almost don't need to bother with coherent structures any more. Why? Because our understanding of turbulence has gone deeper; we have modeled the basic physical processes that give birth to these coherent structures. We can simulate a turbulent flow on powerful computers and a flow that to the naked eye seemed random and chaotic can now be computationally modeled and predicted with unprecedented accuracy. Does this mean that a few hundred years from now we might have accurate computational models for individual’s behavior with predictive capabilities?!

In our research, it became obvious, relatively early on, that only mathematical treatment of a stock's movements as a time series, in "black box" approaches and in the absence of a physical model, was not going to achieve our predictability performance goals. To address that, we spent the next few years modeling trader/investor psychology in the context of stock markets as a highly non-linear dynamical system. We discuss a simplified version of such a system and how the technique operates in the next section.

2.3 The Dynamical System Analogy

Real life physical systems like stock market and weather systems are without doubt extremely complex. The motivation for low order modeling techniques is that we find a physical model of reduced complexity that can capture the essential behavior of a complex system.

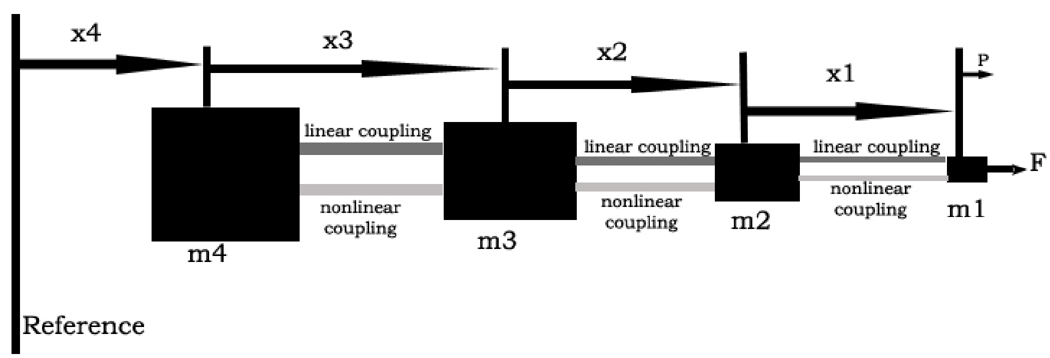

Figure 4: Dynamical system analogy.

Consider the dynamical system in the figure 4. The masses are progressively heavier from right to left. The masses are connected through different pairs of linear and nonlinear couplings. The entire system is free to move but we measure its motion about a reference line as shown. This reference line can represent something as simple as the mean of the data being considered for modeling. The force F indicated is an impulse force that is applied to the smallest mass. A reasonable question to ask here is why aren't we considering forces applied to the other masses? We have found that, in the stock market modeling at least, the energy flows primarily from the faster to the slower scales and that the effective forces applied to the heavier masses are significantly small than that applied to the small mass. The more profound the "fundamental" event, the larger the forces applied to the heavier masses become. However, in our work we are neglecting the forces on the larger masses. To account for the error introduced by this assumption, we:

1. "Stay Alert" by performing impact force applications in regular intervals. Most times, the magnitude of the force is small to practically zero, but we go through the process anyway, in order to catch "hints" of on-coming changes early and prevent error accumulation. Typically we perform this process every 7 or so 15-minute intervals. And vice versa: even if an impact application was last performed, two intervals ago for example, but the data shows significant deviation of the system's response from its unforced response, then a new impact application is performed. It is important to note that the system remains unforced between impact applications.
2. Re-optimize the dynamical system's coupling coefficients. These coefficients physically represent inherent physiological and psychological processes and

normally change only slowly. In case of significant to profound events, they can change; and are thus re-calculated. However the masses do not change and also each mass' resonance frequencies remain within a small frequency band.

The displacement P is the sum of the individual displacements:

$$P = \sum_{i=1}^{n} x_i$$

The system is already in motion while we encounter it. The data that is to be modeled is P. In the case of the stock market it is the price. As detailed below, we identify the dynamics of the system by first identifying and extracting "coherent structures" for the system output. These correspond to dominant physical processes, typically of hierarchical resonance frequencies, with nonlinear interactions between them. For the stock market system, since we extract close to a dozen "coherent structures" or "scales" for appropriate quality of overall modeling, our low order physical equivalent consist of close to a dozen masses.

Let us consider the details of the modeling process step by step.

1. First we do the spectral decomposition of the data to find the dominant frequencies inherent in the system. Note that our decomposition method differs from Fourier techniques and digital filters in that we have minimized end-effects and frequency leakages. Chapter III provides discussion illustrating the comparative advantage of the spectral decomposition part of the new technique over Fourier-based modeling. For a system that has embedded coherent structures, the sub-spectra obtained are consistent

regardless of the location of the time-series window considered in the data. The frequencies obtained from the decomposition present themselves in distinct bands. Chapter IV illustrates the existence of such bands in the case of stock market price data. Thus in the analogous nonlinear dynamical system of masses that can be considered to model the data, there will be as many masses as frequency bands found in the data.

2. Once the distinct bands of frequencies are identified in the system, we apply an adapted version of the POD technique elaborated in Section 2.1 to extract one coherent scale per band. This scale can be made up of different POD type modes. We can consider these scales to represent each mass and its motion in the dynamical system model of Figure 4.

3. In the next step, we use the extracted scales to determine the linear and nonlinear coupling coefficients of the system so that the overall motion matches the data being modeled. The motivation here is to obtain coefficients such that the *unforced* motion of the dynamical system follows the actual data. These masses and the coefficients capture the internal dynamics of the system, which in the case of stock market, is the essential dynamics of trader/investor psychology. As expected, drastic changes seldom happen to these quantities once they are identified. But obviously they are not set in stone since our model is deterministic and also impact of the external environment is crucial. Changes, when they happen to the internal dynamical quantities of the dynamical system, especially in the coupling coefficients, are slow and small.

4. The impact of the external environment on the dynamical system is captured using the model of an impulse force. When the real data diverges from the modeled behavior, we adjust the model either by optimizing the coefficients or introducing an impulse force or a combination of both. The need for an adjustment via an impulse force manifests when the fastest scales that correspond to the motion of the smallest mass display significant divergence. It is crucial not to neglect these high frequencies. Just like in turbulence, the small scales have paramount significance and cannot be truncated. The terms $u_i'u_j'$ in the RANS equations, even at the fine scales, do affect the mean.

The effects of an impact force are first felt by the fastest scales. The divergence of our fast scales from their unforced evolution is used to quantify the magnitude and sign of the impact force. In the case of scales of stock market data, we can tell from the behavior of the fast ones whether something "fundamental" has happened in the market. So at that point we represent the system as

$$M_i\ddot{x}_i(t)+C_i\dot{x}_i(t)+K_ix_i(t)+\sum_{j,k,l}a_{ijkl}(\ddot{x}_i)^j(\dot{x}_i)^k(x_i)^l=F$$

where

F is defined by

$$F=\begin{cases}0, i>1\\ \delta_i(t), i=1\end{cases}$$

Figure 5 shows typical application of such impulses. As mentioned above we typically perform this process every 7 or so 15-minute intervals, or as needed if a significant fundamental event has occurred as indicated by the divergence of the faster scales from their unforced motion.

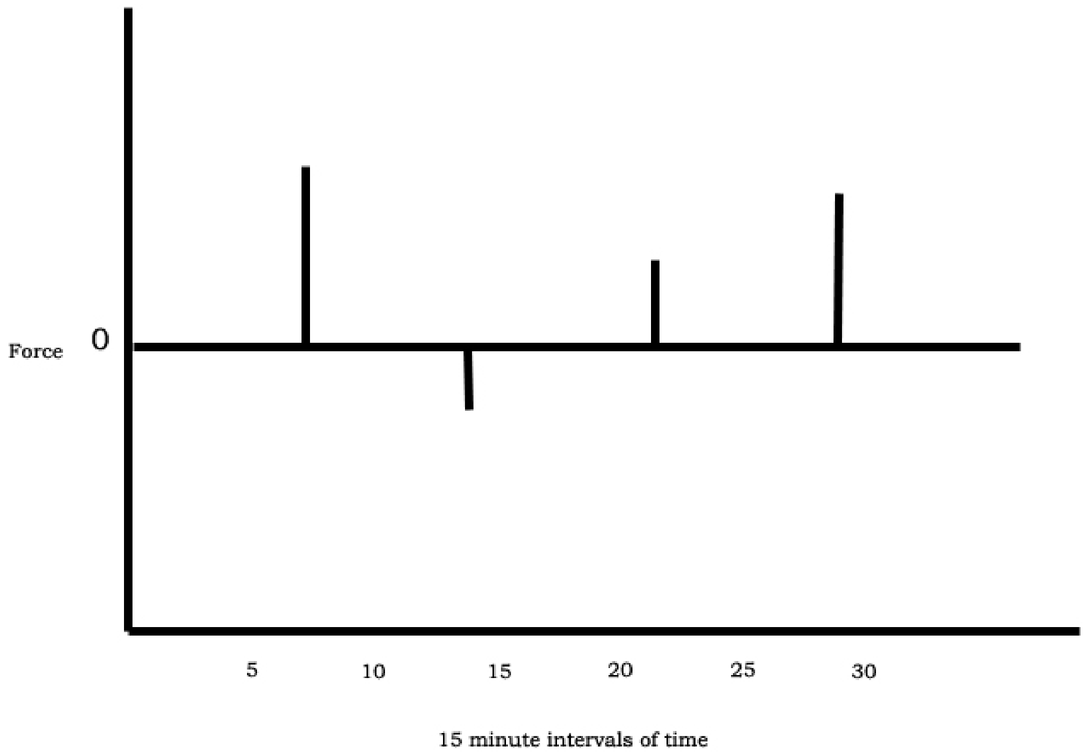

Figure 5: Application of impulses with time.

CHAPTER III

APPLICATION TO A NONLINEAR DYNAMICAL SYSTEM

This chapter presents the results from the application of the new technique to a nonlinear dynamical system. Before presenting the system, in Section 3.1 we demonstrate the application of the technique to a synthetic signal. Then the method is tested on a totally random signal to ensure that it does not “create” coherent modes where they do not inherently exist. Section 3.2 discusses the creation and preprocessing of the nonlinear dynamical system. Section 3.3 features the results of the application of the low order modeling technique.

3.1 Applications to Synthetic and Random Signals

A synthetic signal was generated by the summation of 9 cosine waves. The components had periods of 250, 180, 125, 90, 62, 45, 31, 22, 15 points and corresponding amplitudes were 10, 8, 6, 5, 4, 3, 2, 1 and 1. Figure 6 shows the time series plot of the signal and figure 7 shows its Fourier spectrum clearly revealing the components.

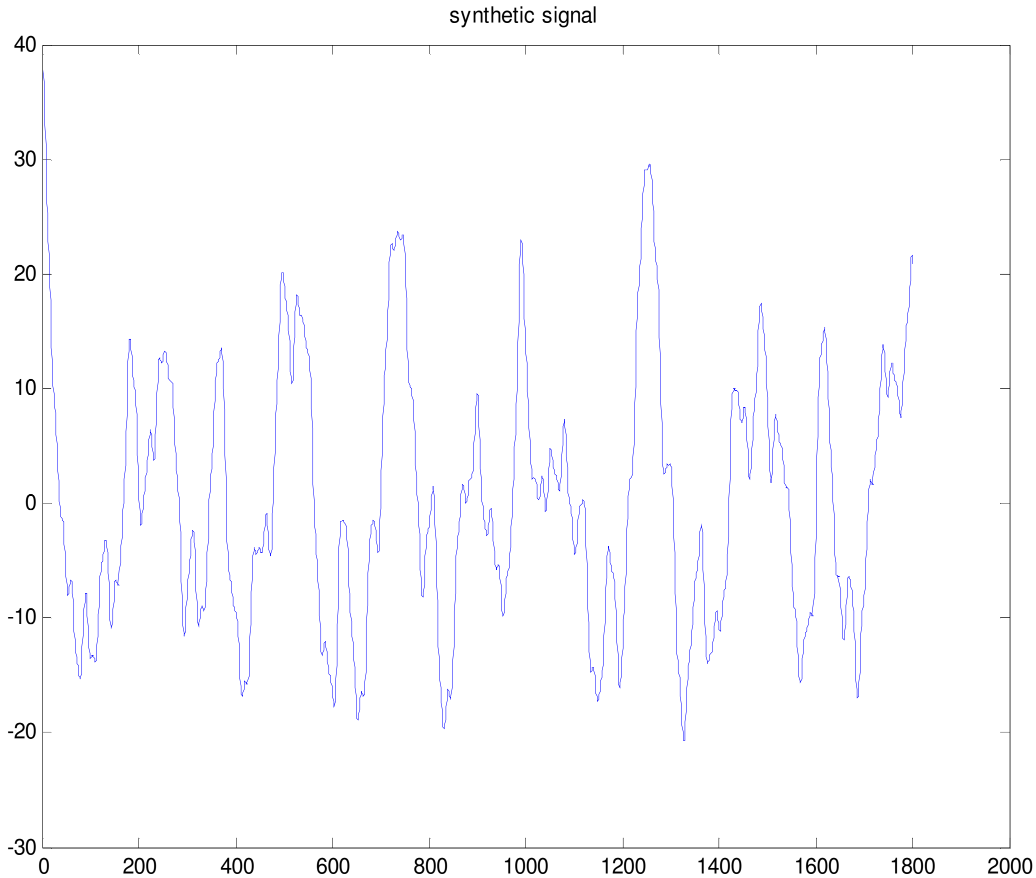

Figure 6: Synthetic signal generated by the summation of 9 cosine waves.

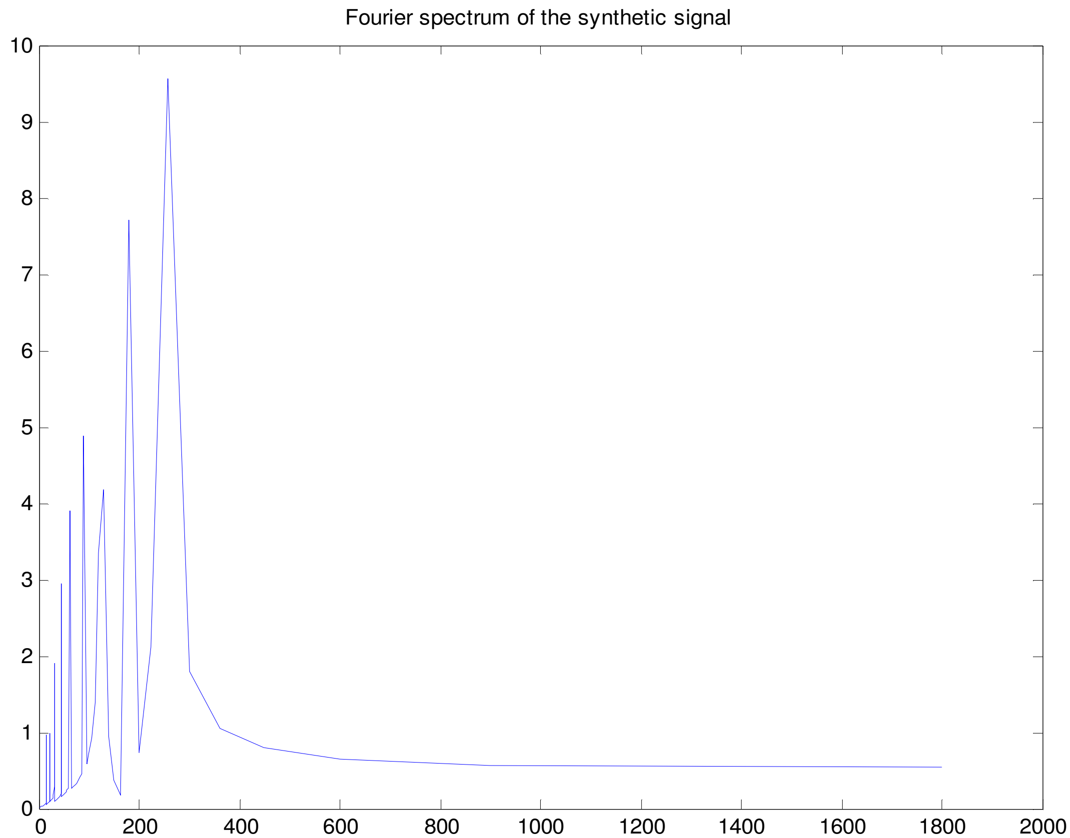

Figure 7: Fourier spectrum of the synthetic signal.

The different steps mentioned in the previous chapter were carried out on the first 1800 points of the synthetic signal and the modes were obtained. Using these modes, future points of the signal were predicted 10 points at a time. Figures 8-10 show the comparison of these 10 point predictions with the actual behavior of the synthetic signal. It is clear that the technique provides reliable predictions for the synthetic signal.

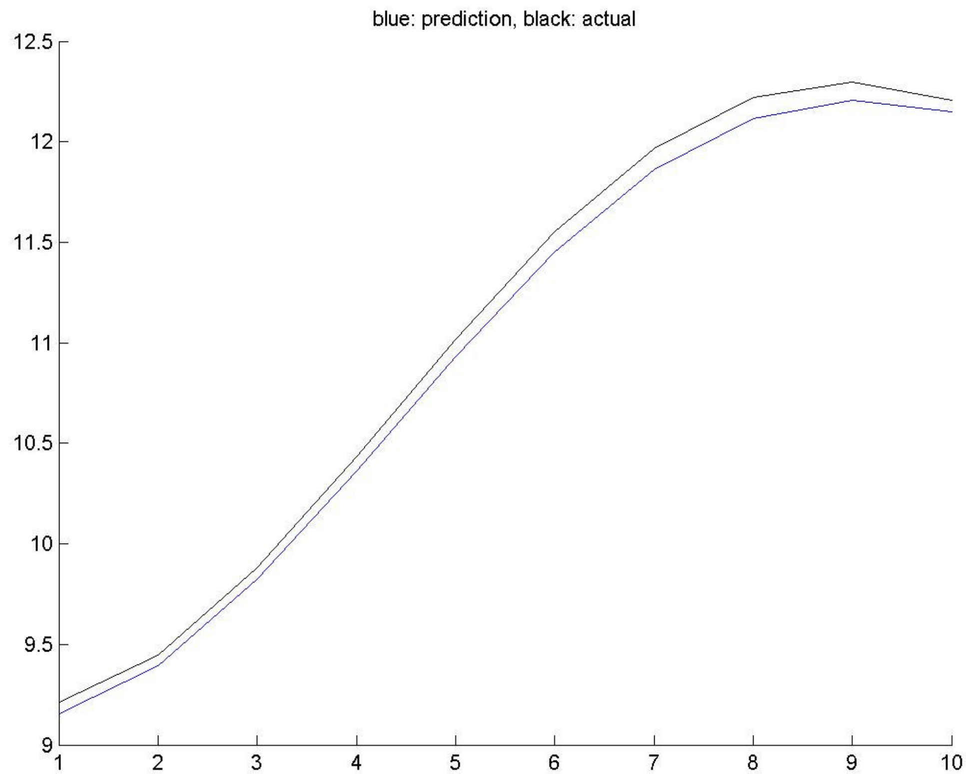

Figure 8: Example 1 of a 10 point prediction (blue) compared to the actual signal (black).

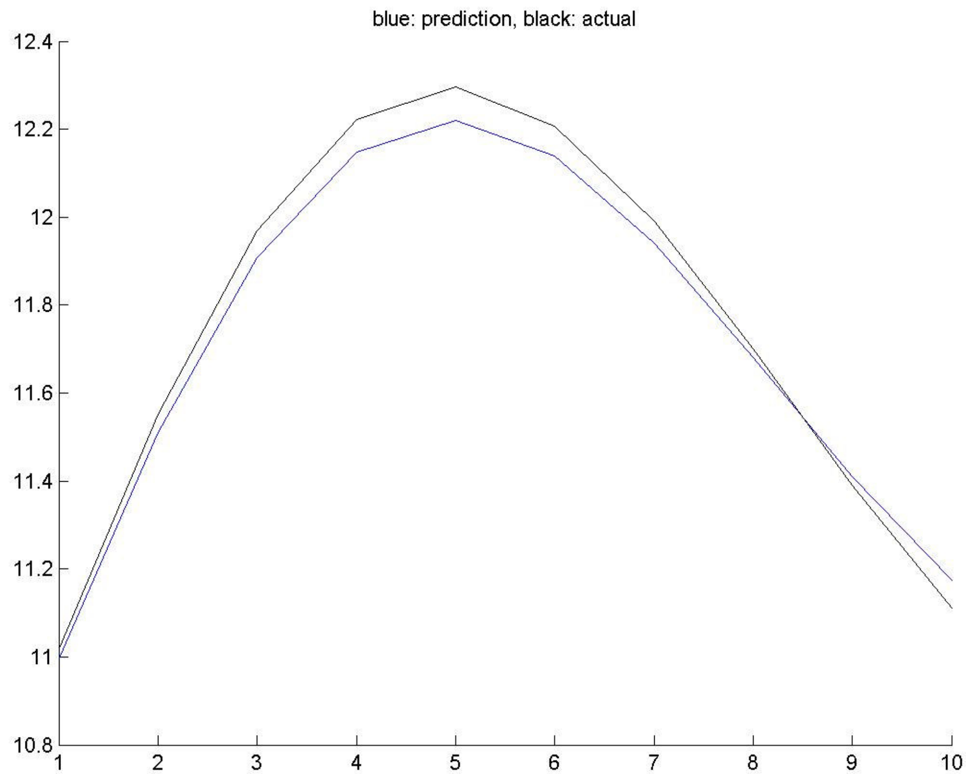

Figure 9: Example 2 of a 10 point prediction (blue) compared to the actual signal (black).

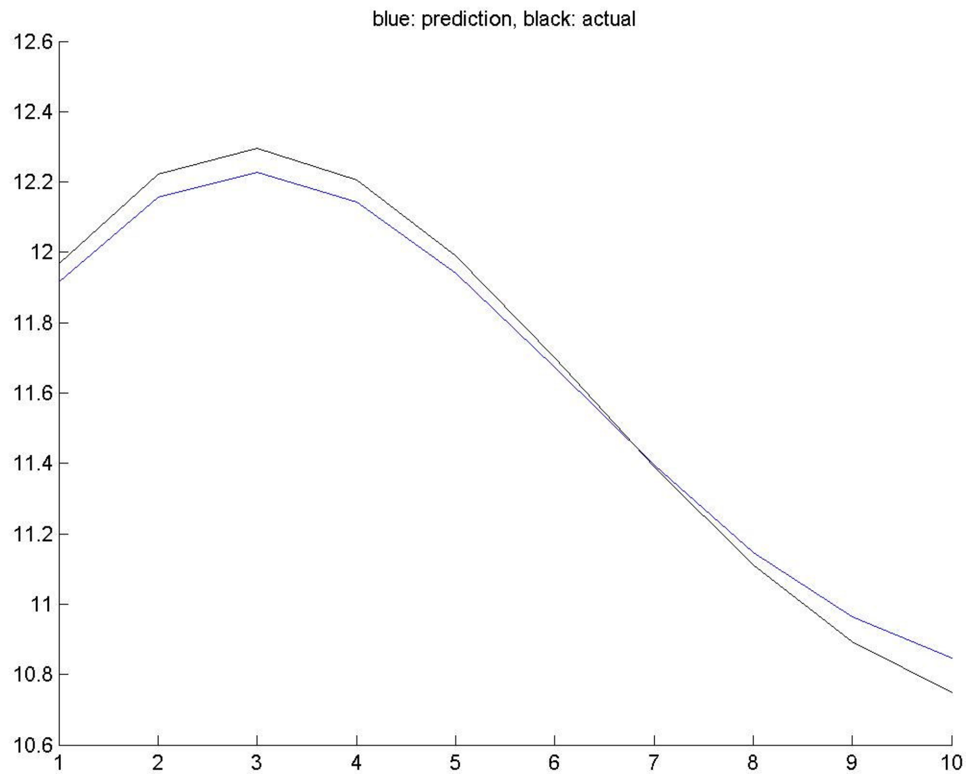

Figure 10: Example 3 of a 10 point prediction (blue) compared to the actual signal (black).

Once the technique was successfully tested on a synthetic signal with known cosine wave components, it was applied to a random signal. The random signal was generated using the MATLAB® commands *rand* and *randn.* Figure 11 shows the raw random signal thus generated. Since the technique seeks a minimum period of 12 points for its fastest mode, this raw signal was filtered using our indigenous technique to remove all periods below 12 points. Figure 12 shows this filtered signal on which further application of the technique was done. Figure 13 and 14 show the Fourier spectrum of the raw random signal and the filtered signal. Figure 14 illustrates the efficiency of the indigenous filter.

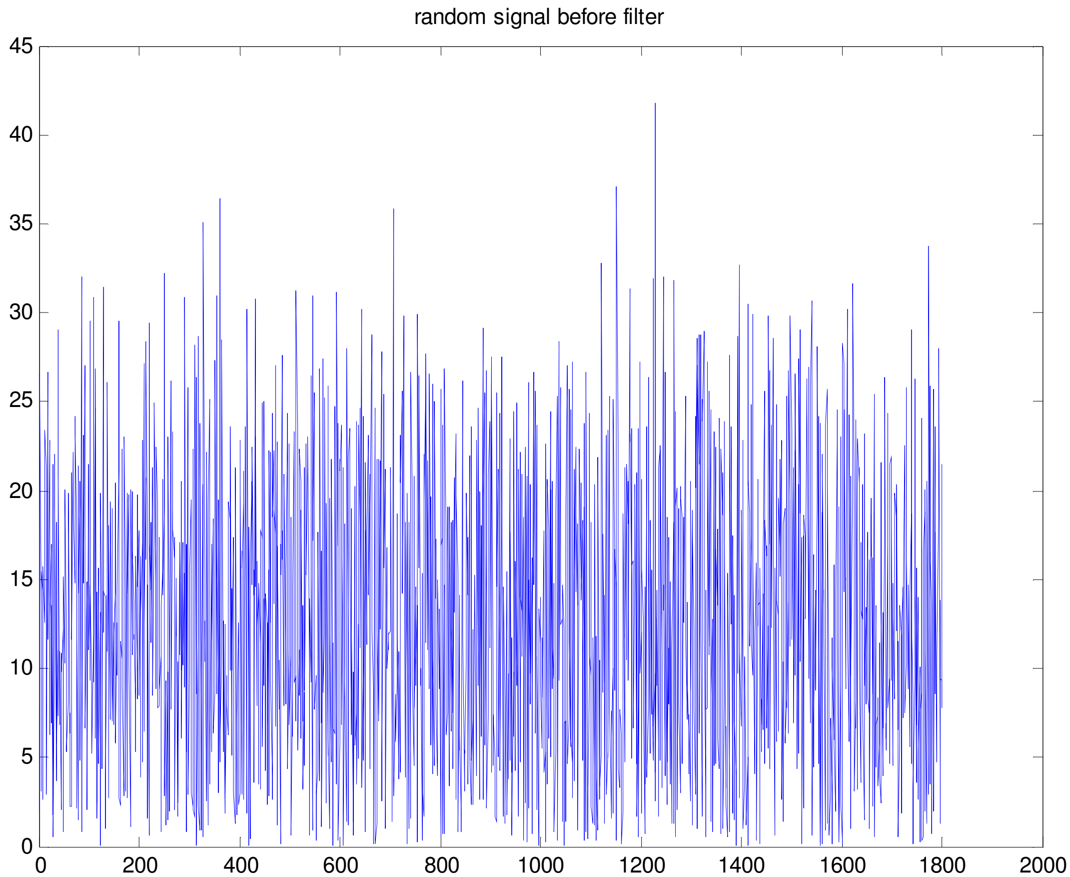

Figure 11: Random signal before the application of the filter.

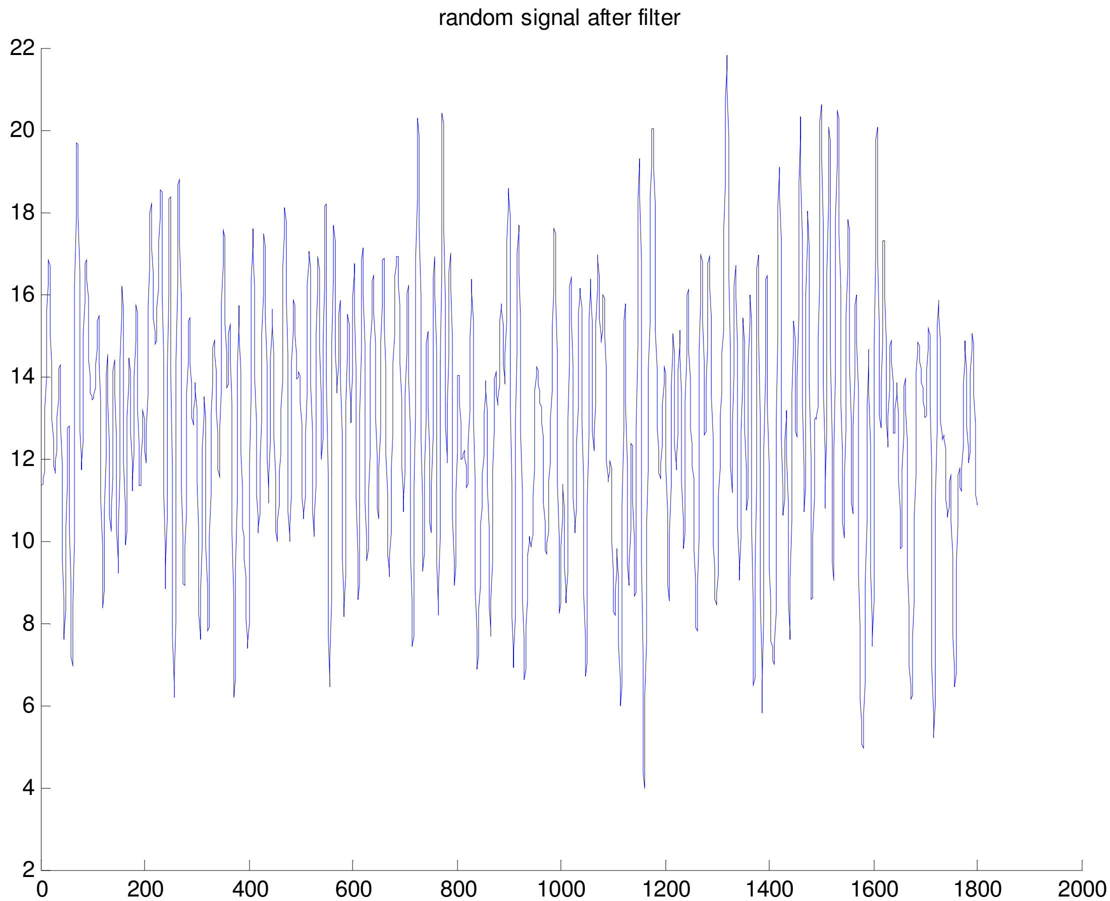

Figure 12: Random signal after filtering.

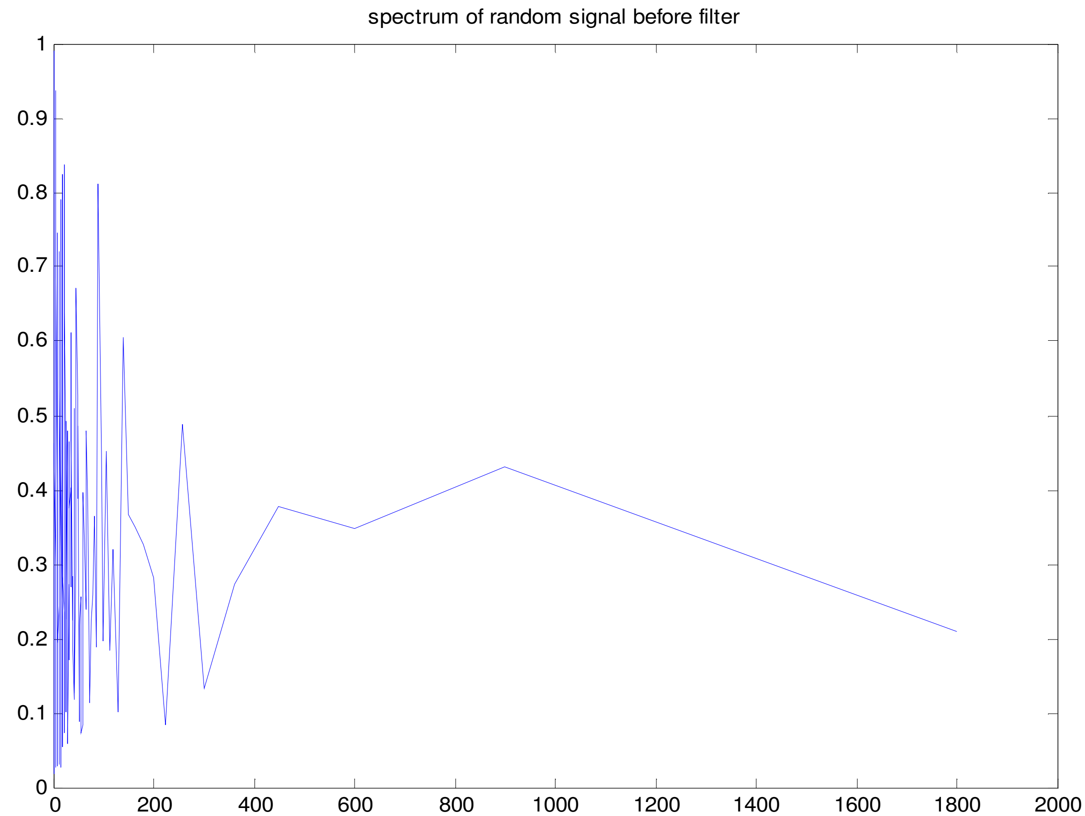

Figure 13: Fourier spectrum of the random signal before filtering. X-axis: time period.

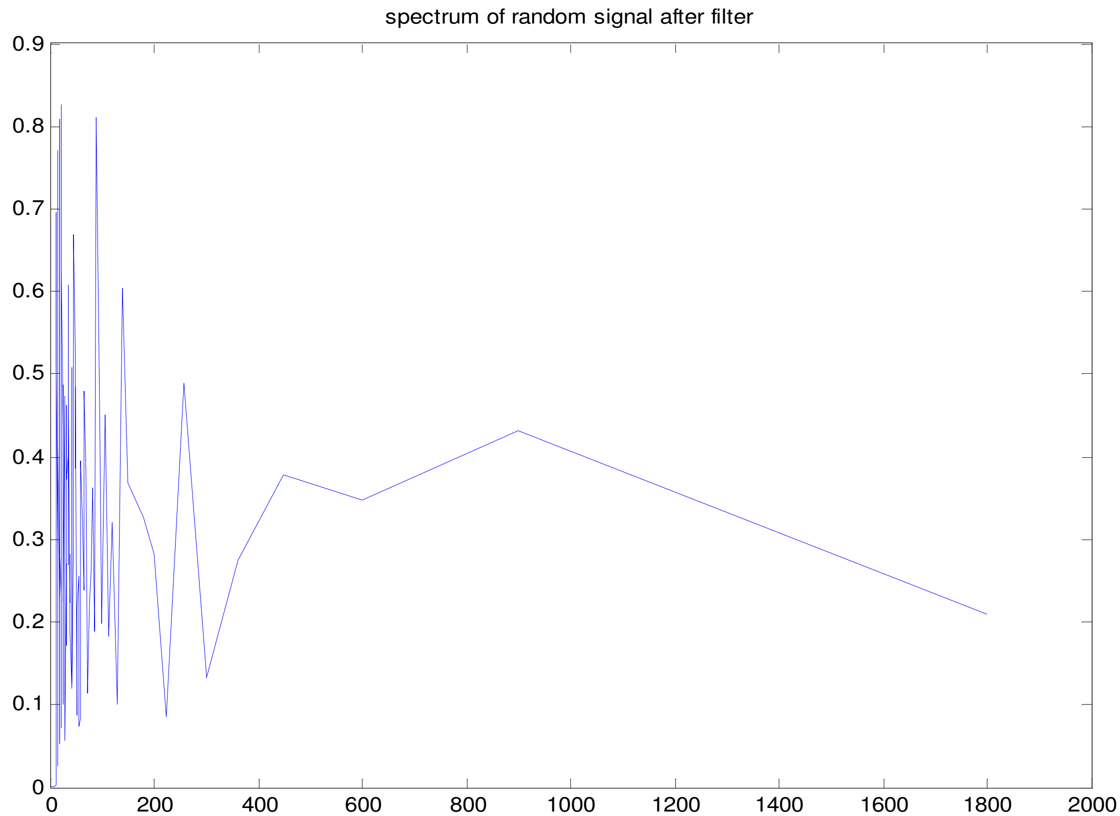

Figure 14: Fourier spectrum of the random signal after filtering. Notice the removal of component with periods below 12 in the x-axis.

On the filtered random signal the technique was applied to see if could yield predictions. Like in the case of synthetic signal presented earlier, 10 point predictions were attempted. Figures 15-17 show the comparison of these predictions to the actual signal. It is clear that when coherent structures are absent in a signal and it is random, the technique cannot make predictions with reliability. The technique does not generate pseudo coherent structures in signals where they are absent.

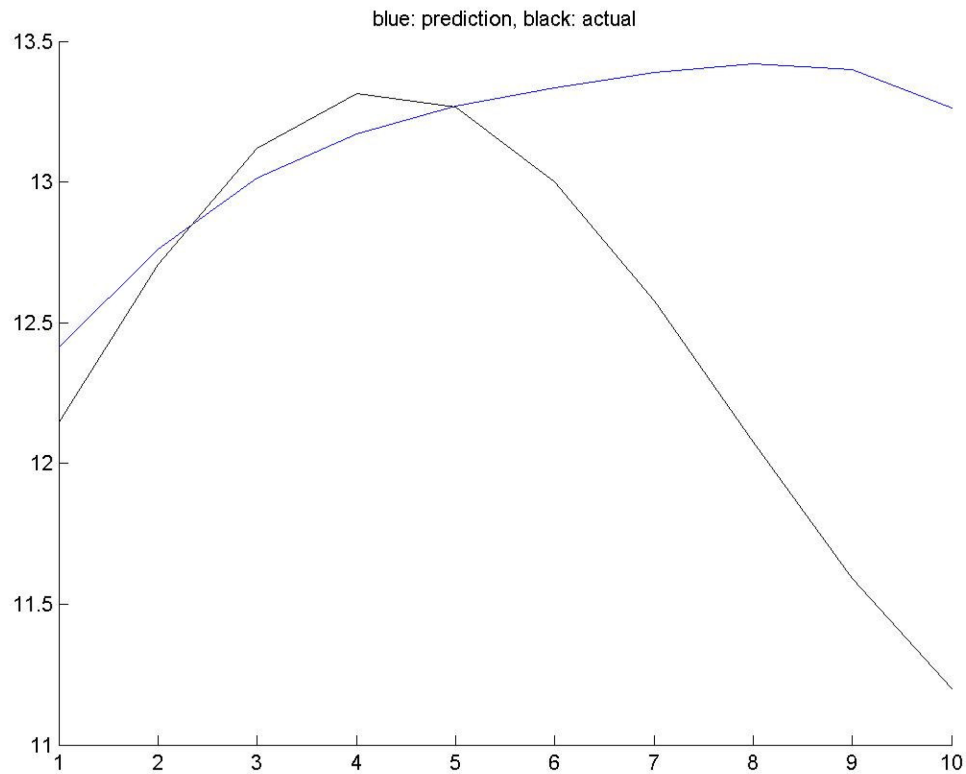

Figure 15: Example 1 of 10 point prediction (blue) compared to filtered signal (black).

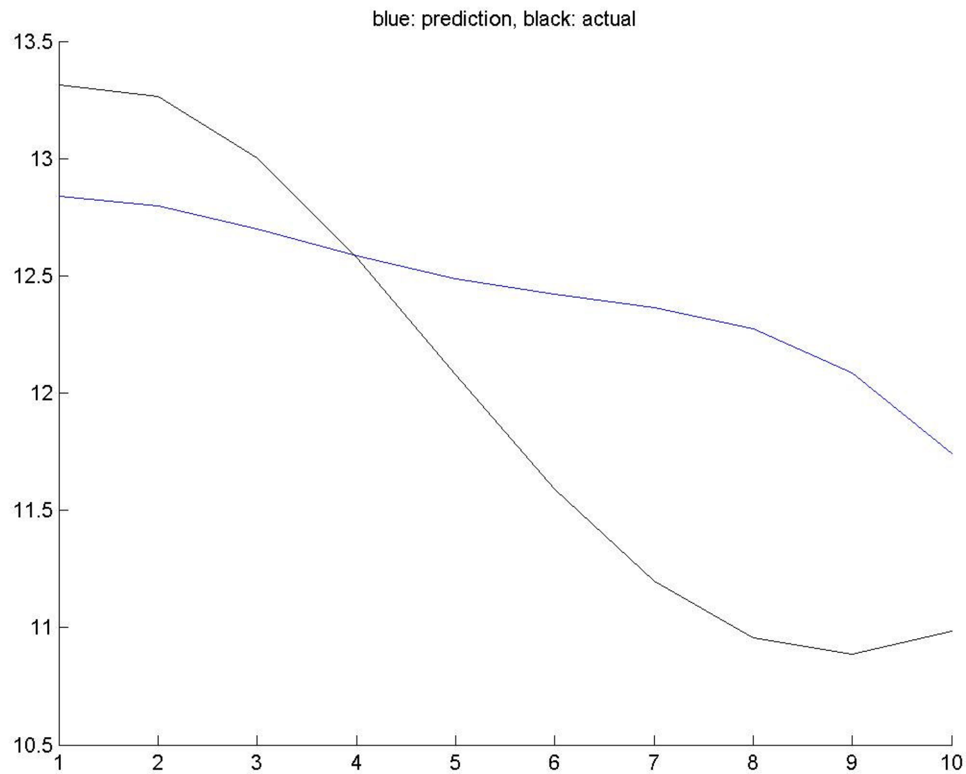

Figure 16: Example 2 of 10 point prediction (blue) compared to filtered signal (black).

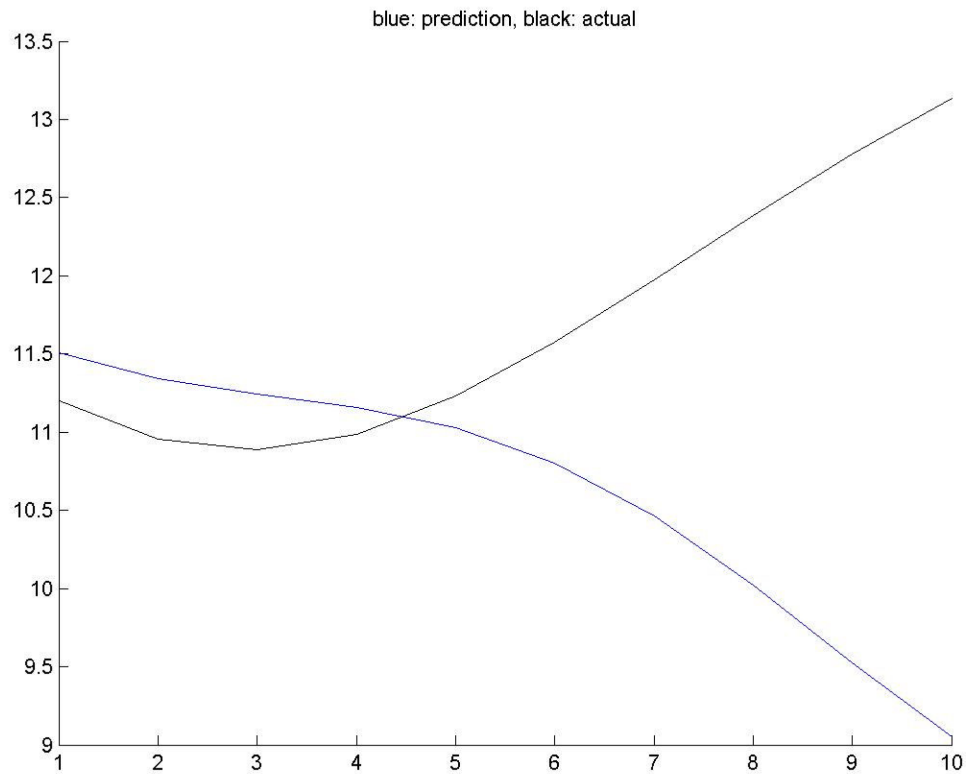

Figure 17: Example 3 of 10 point prediction (blue) compared to filtered signal (black).

3.2 The Nonlinear Dynamical System

A combination of two classical nonlinear oscillators was used to create a seemingly random system with single output as the displacement of a mass. Dynamics of the moving mass was produced from a linear combination of

(a) The classical Duffing's oscillator model (hard spring) which is a periodically forced oscillator represented by

$$\ddot{x} + \delta\dot{x} + \beta x + \alpha x^3 = \gamma\cos\omega t$$

Where the damping coefficient is always greater than or equal to zero.

and

(b) the Van -der Pol's oscillator model [33] with non-linear damping and the following governing equation

$$\ddot{x} - \epsilon(1 - x^2)\dot{x} + x = 0.$$

The governing equation for the combined system being

$$\ddot{x} = -kx - c\dot{x} - \alpha x^3 - \mu(x^2 - 1)\dot{x}$$

The objective was to create an output that resembled the seemingly random systems like the stock price or weather data. So the constants in the governing equation were arrived at using trial and error and Gaussian Nonlinear Least Squares method. The values of the constants used were

$$k = 1; c = 0.05; \alpha = 0.5; \mu = 0.6$$

The equation was solved using ode45 subroutine in MATLAB®. The output y was created as:

$$y = \sqrt{x^2 + \dot{x}^2}$$

The measurements are taken at discrete time instances and were further corrupted by a zero mean Gaussian random variable using *randn* function of MATLAB®.

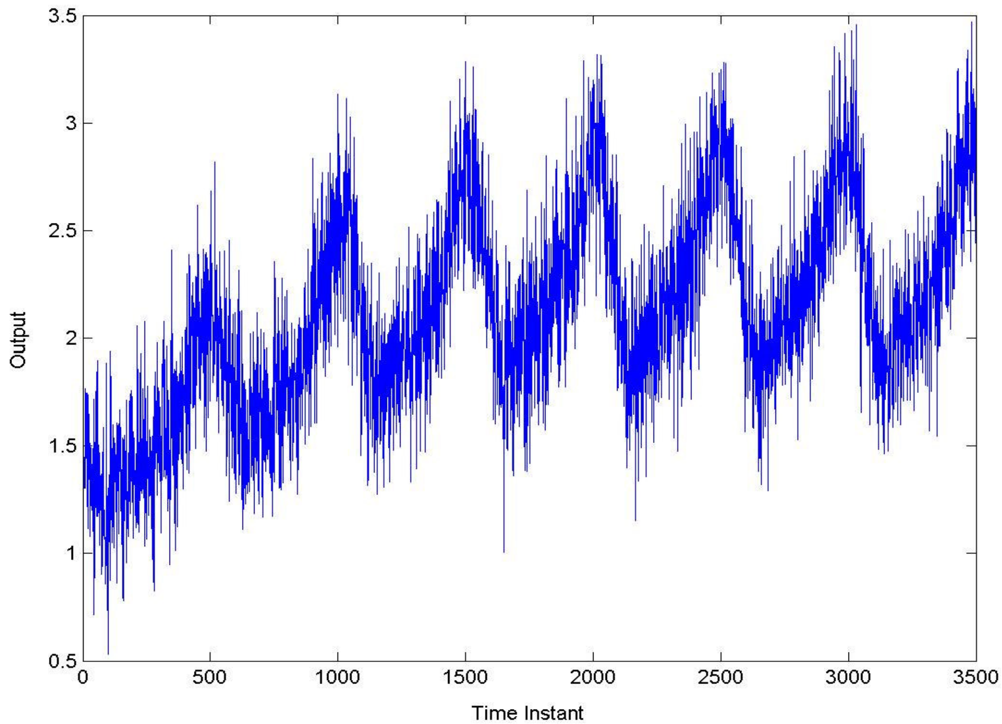

Figure 18: Raw data representing the nonlinear dynamical system.

Figure 18 above shows the raw data produced. Since the minimum period of a scale is 12 points in the way the technique is set up, we apply our indigenous new filtering technique on the above data to obtain the smoother version shown in figure 19. 2800 initial points shown in that figure were used for "training" or extracting the scales. 10 point predictions were tested on the remaining points.

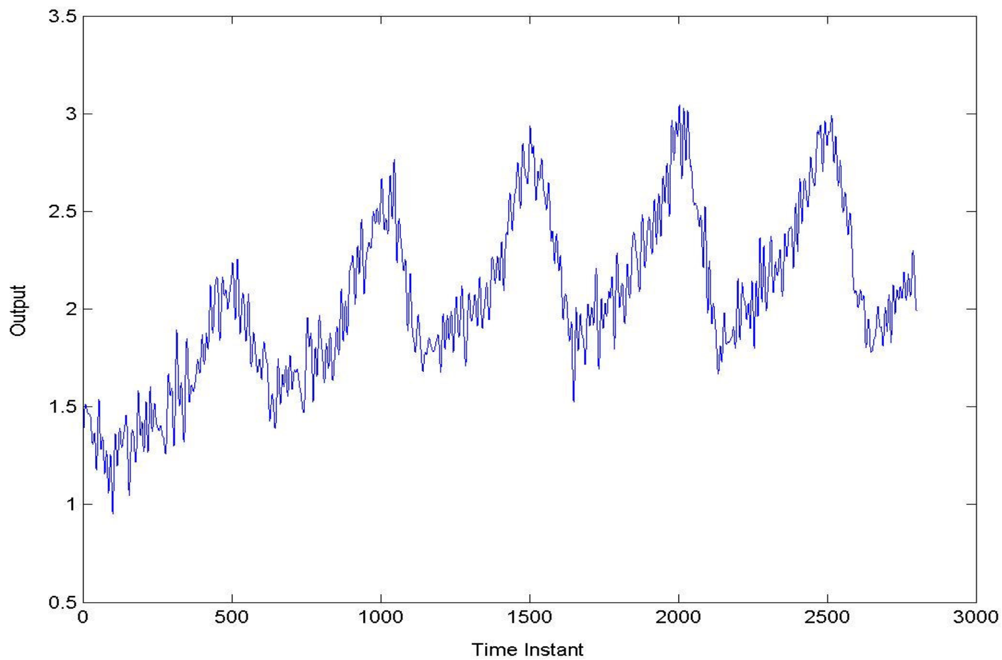

Figure 19: 2800 points of smoothened data used to extract scales.

3.3 Application of the New Technique

As mentioned in chapter II, Fourier analysis makes assumptions about the periodicity of the data and is also limited by finite frequency resolution through Nyquist limit. We developed a low order spectral decomposition, as a result to the mathematical problem: For a given time series, and a specified band-pass, find the minimum number of cosine functions (not necessarily orthogonal), of any frequency (infinite resolution), the sum of which best represents the signal in the specific band-pass. The technique does not assume periodicity of the signal and has infinite frequency resolution. Also, its evolution in time is much more continuous than the Fourier representation.

Since we cannot assume periodicity of the time series, Fourier techniques are not justified. Figure 20 shows the Fourier analysis based filtration that has a minimum retained period of 200 points on consecutive windows of the data staggered by 100 points starting at 1800th point purely to show contrast with our method. The end effects of Fourier techniques are obvious. The models correct themselves drastically as future data becomes available. The danger of using such models, though they may "fit" past data, for any kind of forecasting is obvious.

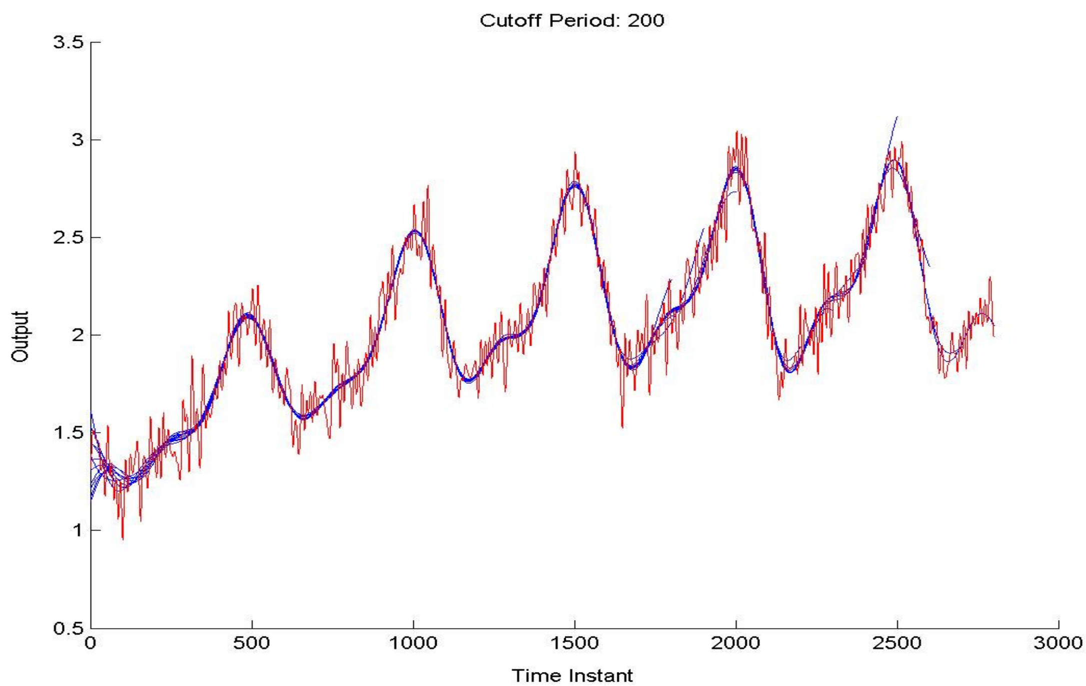

Figure 20: Fourier analysis based filtration with a cutoff period of 200 points.

In the figure above data used is in red. Note the mismatch of end-points of the filter outputs. Contrast this with the outputs of our lower order modeling technique shown in Figure 21. Obviously, the end effects are significantly reduced.

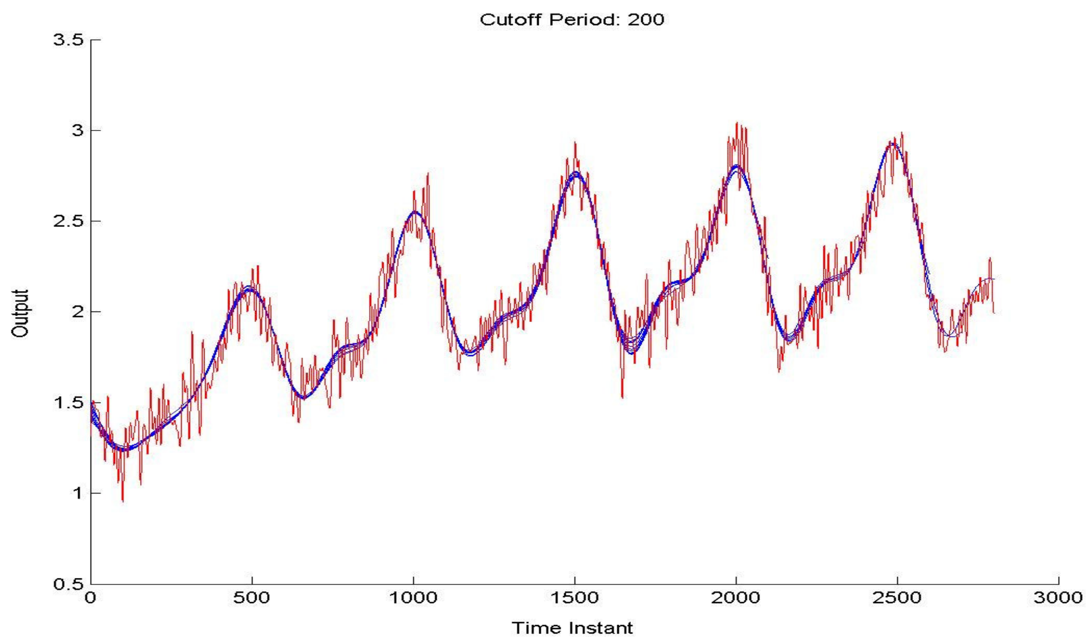

Figure 21: Low order model filter output with significantly reduced "end-effects".

Instead of Fourier based filters, if we consider an Infinite Impulse Response digital filter like Butterworth filter, we run into the trade off problem between lag and frequency leakage. Increasing the order of the digital filter would result in reduced leakage effects but that also leads to increased lag which makes the model not useful for future prediction. Reducing lag demands using lower order filter which in turn increases the frequencies leaked into the band that is being filtered from both sides.

Observation of the data used for modeling as shown in Figure 19 shows that there is a significant component with a period of 520 points in the data. Ideally thus a filter that is asked to retain 550 points as minimum time period should not capture this time of 520 points. Figure 22 shows this significant period leaking into a filter with cutoff 550 created through Fourier analysis. In contrast as figure 23 shows, there is minimal leakage issues with the low order modeling technique filter.

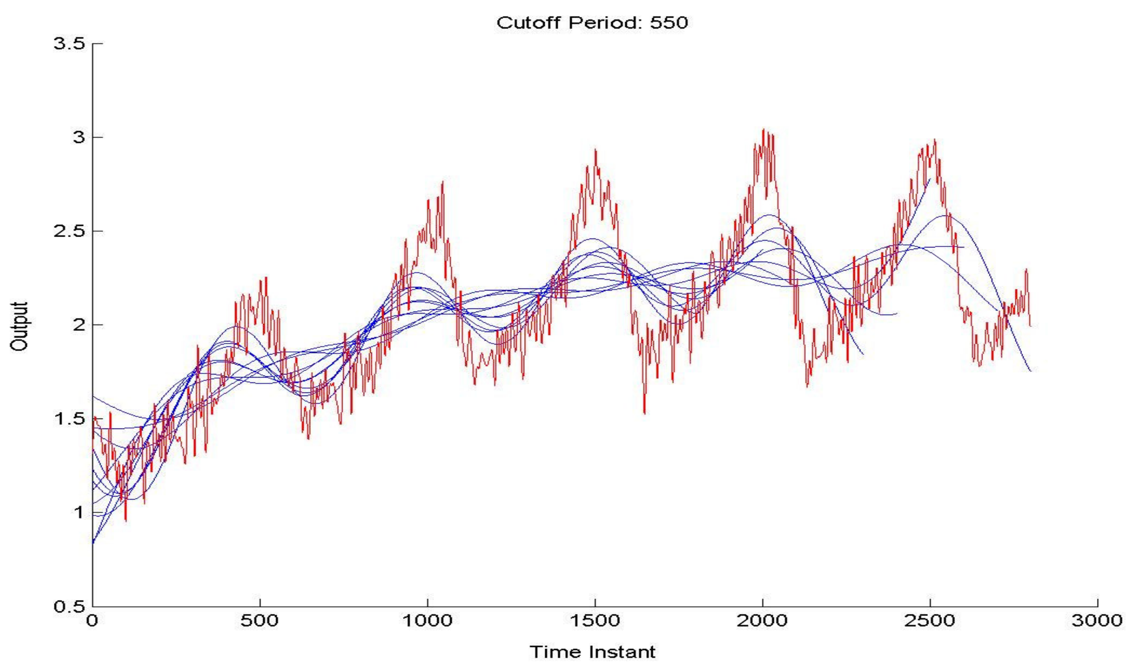

Figure 22: Fourier based filtration with cutoff period 550 showing leakage.

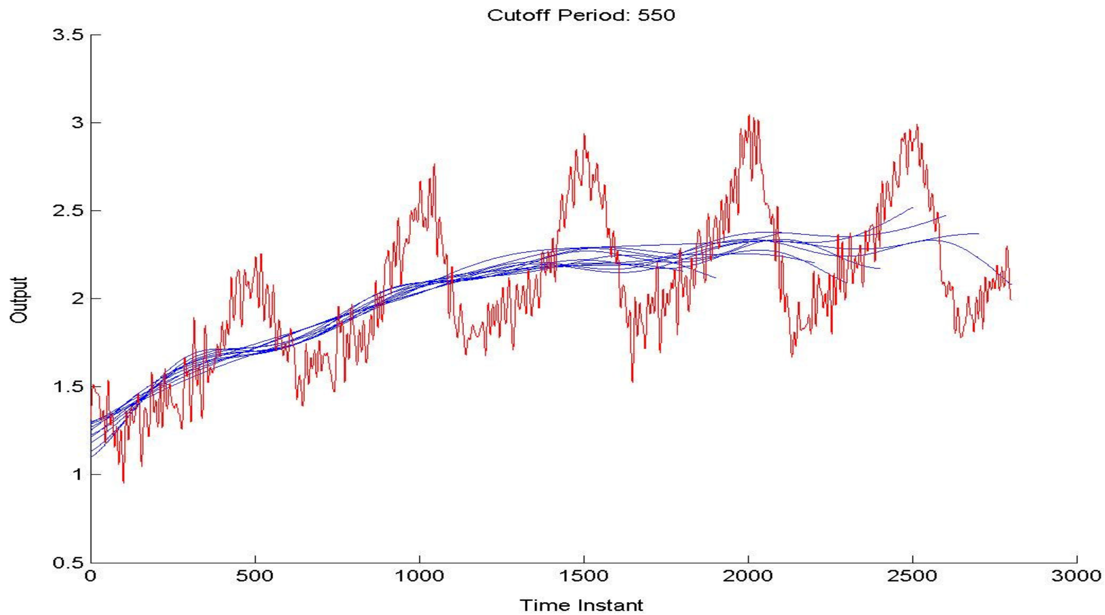

Figure 23: Low order model based filtering shows very little leakage effects in contrast.

The single greatest feature of the scales provided by the method is their time invariance. This can be best visualized by watching successive figures of the data, the model and the scales for consecutive points in time as a movie. The mpg file Movie 1 attached provides this visualization. The data is represented in black, the model closely follows the data in red and the various scales are shown in other distinct colors. After examining the time invariance of the scales from the previous figures, we present the comparison of 10 point predictions made with the technique compared to actual system behavior. It is clear that the reduction in end effects and frequency leakage lead to much better modeling and subsequently forecasting. Figures 24-27 show some examples of predictions.

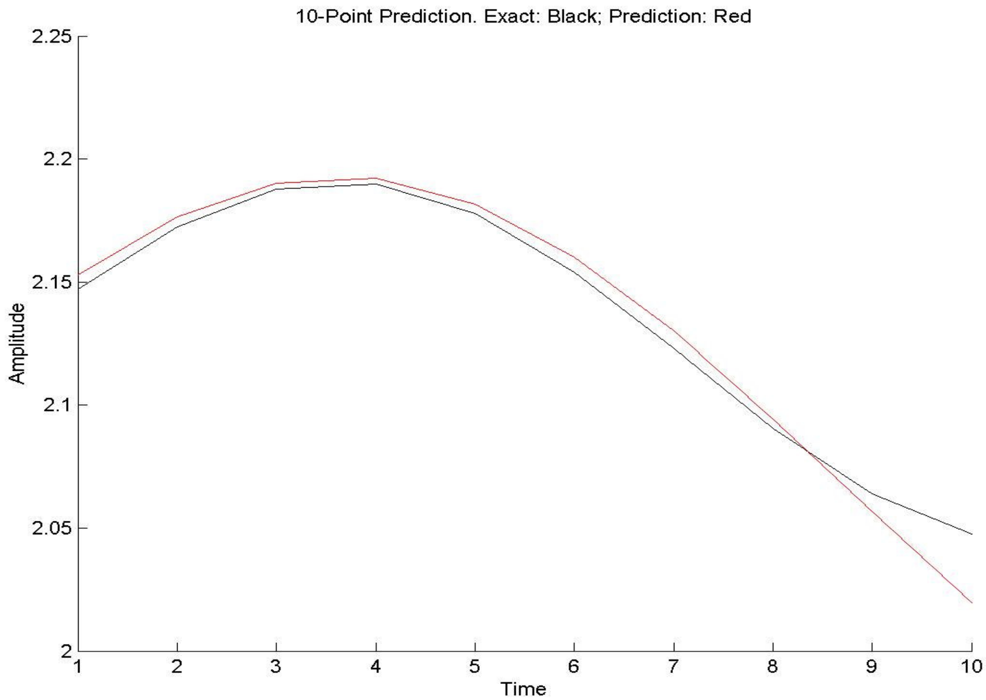

Figure 24: Example 1 of a 10 point prediction of the nonlinear dynamical system.

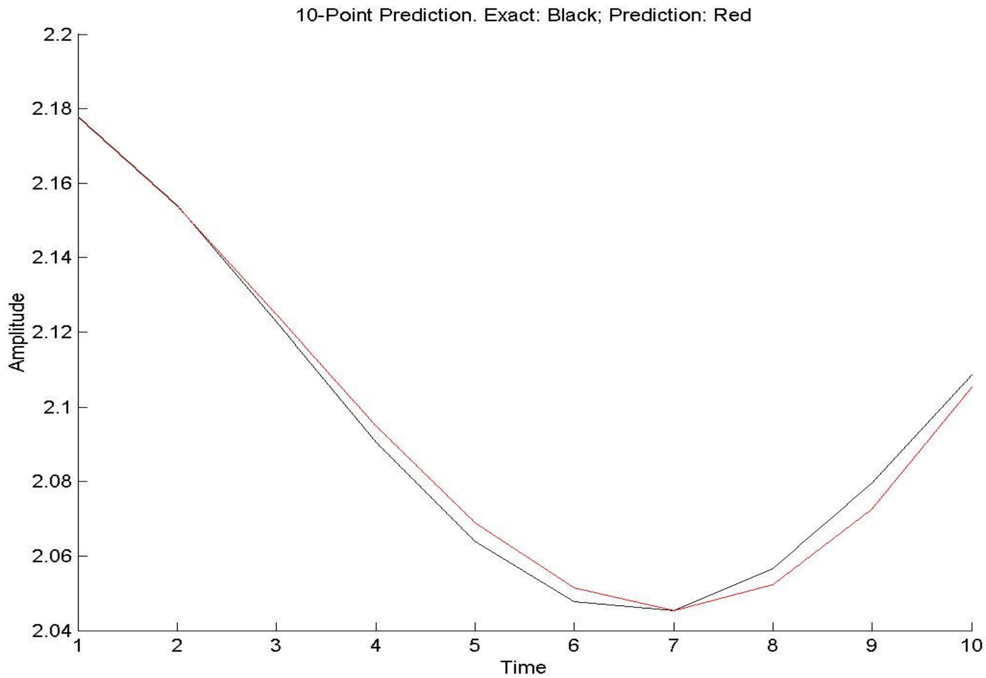

Figure 25: Example 2 of a 10 point prediction of the nonlinear dynamical system.

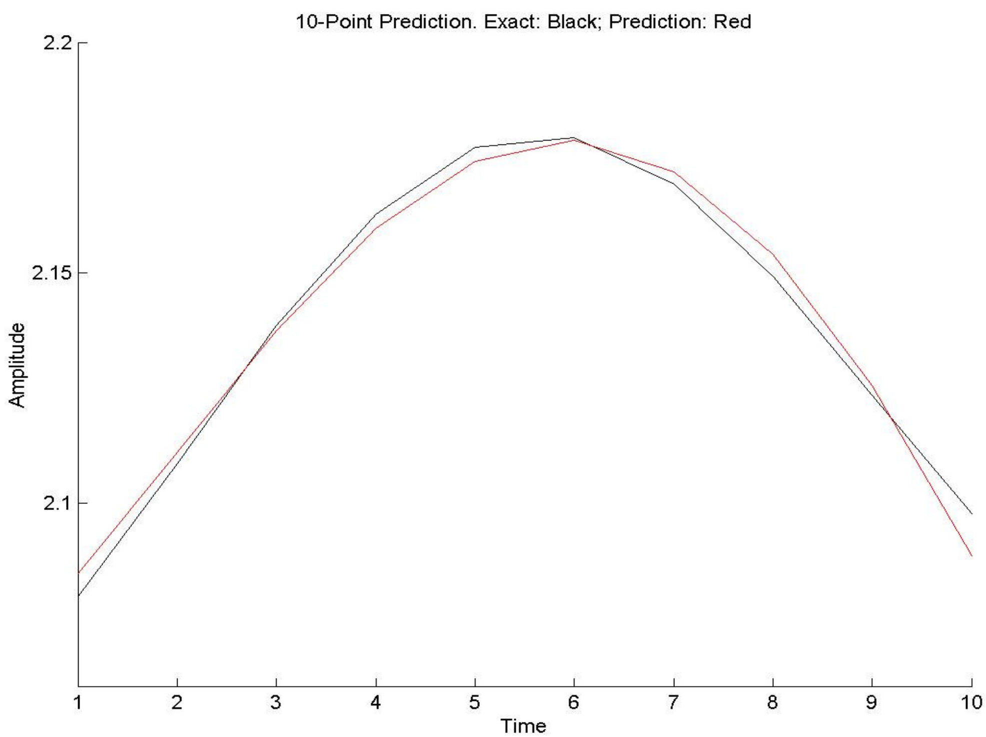

Figure 26: Example 3 of a 10 point prediction of the nonlinear dynamical system.

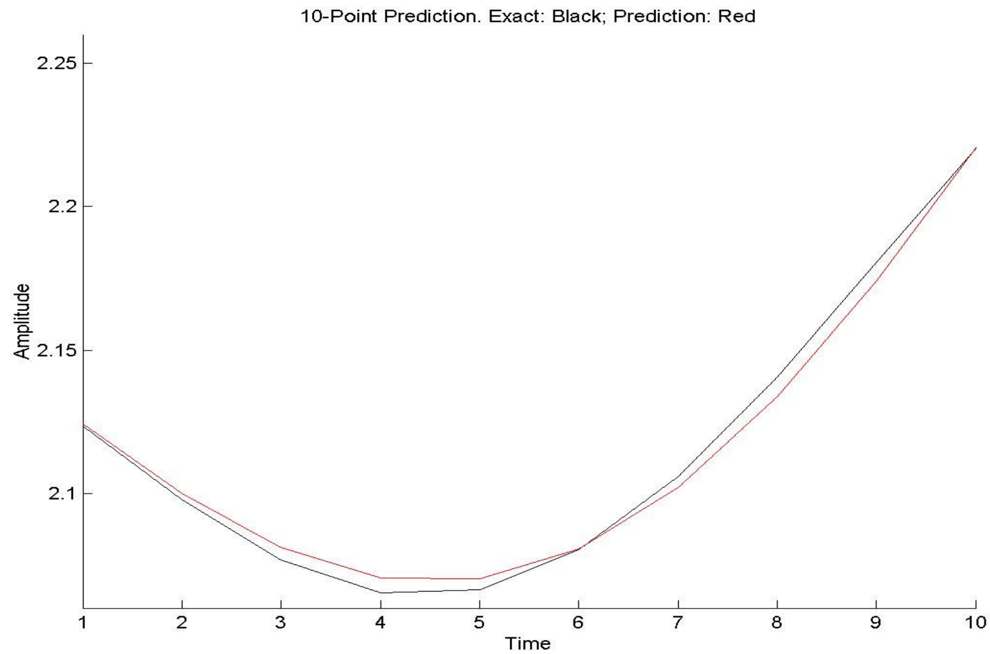

Figure 27: Example 4 of a 10 point prediction of the nonlinear dynamical system.

CHAPTER IV

APPLICATION TO STOCK MARKET SECURITIES

In this chapter, we present the application of the technique on stock market securities. The first section deals with the selection of stocks and preprocessing technique. Section 4.2 presents the results of the application of the technique to a stock price data. In section 4.3, we discuss the flash-based Graphical User Interface that was developed to assist investment and trading decision making based on the technique. The final section of this chapter gives an example of a successful trading decision made using the GUI.

4.1 Selective Modeling

Over 4000 stocks are listed and traded in the two large American stock exchanges, the NYSE and NASDAQ alone. Though our technique is a fairly good "hammer", as the introduction was meant to bring home, we should not consider every single of these stocks to be a nail!

For modeling of stock market securities using the technique, a two stage preprocessing is applied.

1. Market based selection: Technical analysis is supposed to treat every stock as a time series with no regard to the company behind it. But since profit making is the fundamental motivation of anyone interested in the market, clearly all stocks cannot be

considered equal. Specifically from the point of view of our technique which focuses on oscillations, we are interested in stocks whose prices show volatility (market jargon refers to this as beta). Also we would like to have some safety associated with putting our money in a stock. So besides considering stocks with good volatility, our pre-selection process looks for stocks that have fairly good volume of shares being traded and have a price range that is greater than $5. The condition on price is a numerical reflection of company fundamentals. It can be easily verified that fluctuations in very low priced stocks tend to be large and being on the wrong side can be costly. Besides, very low share price could also mean bad health for the company behind it.

2. Mathematical preprocessing: As mentioned before, our technique does not attempt to create coherent structures or scales in every price history. We seek the existence of scales which can then be exploited for modeling. For the mathematical preprocessing of the stocks, firstly we require at least five years of historical data. If the stock has that much available history, then we put it through the preprocessing program to check for existence of coherent structures so that further analysis and modeling can be initiated. No assumption is ever made of the existence of the scales till they reveal themselves.

Using this two stage process, we have so far encountered a few hundred stocks that are amenable to modeling using our technique. This is a fairly large universe of stocks from which profit opportunities can be perceived if not on an hourly basis, easily on a daily basis.

Five years of history is sought because two distinct streams of data is collected for every stock.

1. End-of-day price
2. Intraday price every 15 minutes during the trading day

End-of-day price refers to price at which a stock closes its trading for the day. This price is significant because the final minutes of the market usually display frantic activity. With computerized execution available, it is now well known that there are several funds that automatically trigger their trade executions in the final hour after watching the market during the day. There are also several investors who do not pay any attention to the intraday fluctuations of the market, but make up their decisions based on closing prices. Keeping this in mind, end-of-day price history has its own significance. For our technique, it provides low resolution data which has critical influence on decision making during intraday.

Intraday data used for modeling is collected every 15 minutes during the six and a half hours that the stock markets are open for regular business. 15 minute interval was chosen as a good resolution considering the available processing speeds of computers and trading platforms for execution. As faster and faster processing power becomes available, data can be collected at shorter interval from which price swings of a higher frequency can be exploited using faster trades. But with the current state of available resource, 15 minutes is the optimal time interval.

Modeling using 15 minute day and end-of-day both leave behind residue after modeling which are clearly the frequencies that cannot be captured without higher resolution of data. However, the residue left behind by the modeling of end-of-day prices is invaluable when used in conjunction with intraday model.

Before we present an example of extracted scales from a price history, a couple of their properties is worth mentioning. The scales associated with a stock price exhibit what can be termed as “life-like” quality. Over time, we can observe them gaining and losing strength even dying off and regaining life in some cases. This further strengthens the belief that the physical realities behind what the scales represent are the different channels of money entering and exiting the market. Funds can dry up, come alive, take active interest or lose interest in a stock from time to time.

Figure 28 shows the amplitude-time period spectrum generated using the new spectral decomposition technique for the intraday stock price of ticker APH. Bands of frequency are clearly visible as separated by the lines. A 5000 point signal was split in 10 sections, 500 points long each. Then, using our spectral decomposition technique, the spectral content of each section was calculated, in the form of a set of cosine functions of different frequencies. The figure plots the amplitude versus period of all these cosine functions, from all 10 sections. Figure 29 and figure 30 show the zoomed in views.

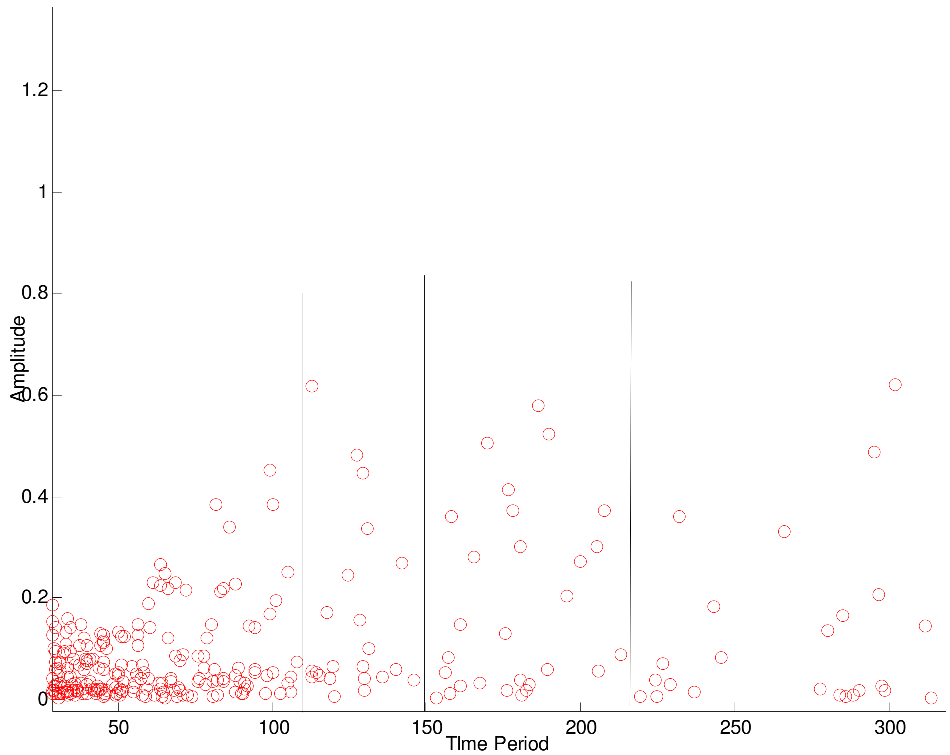

Figure 28: Time Period – Amplitude spectrum for stock ticker APH.

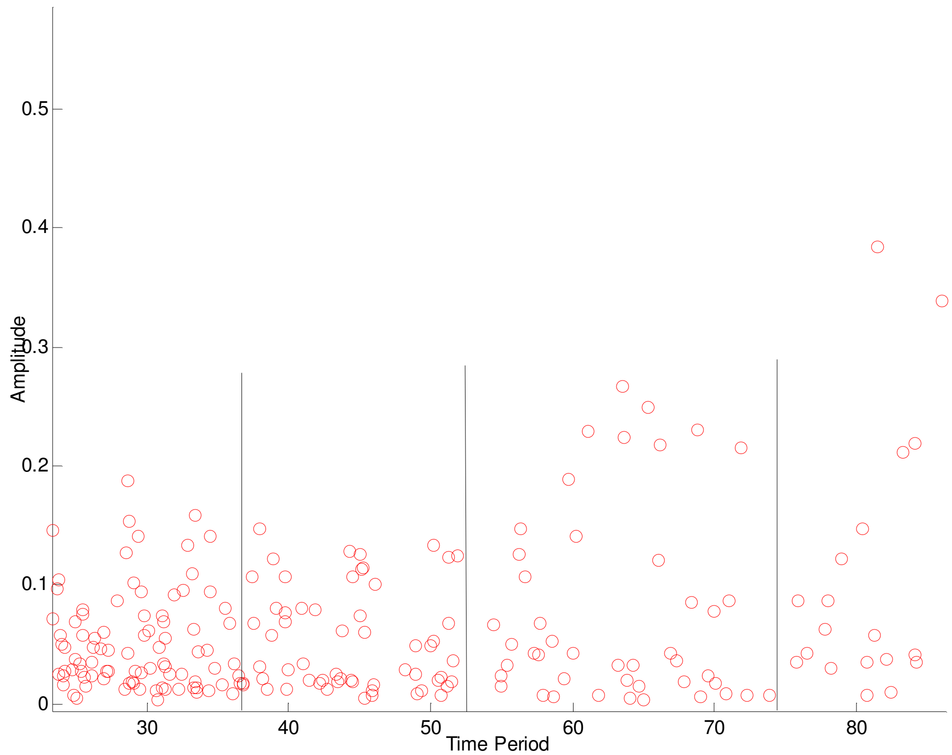

Figure 29: Zoomed in version of the spectrum showing more bands.

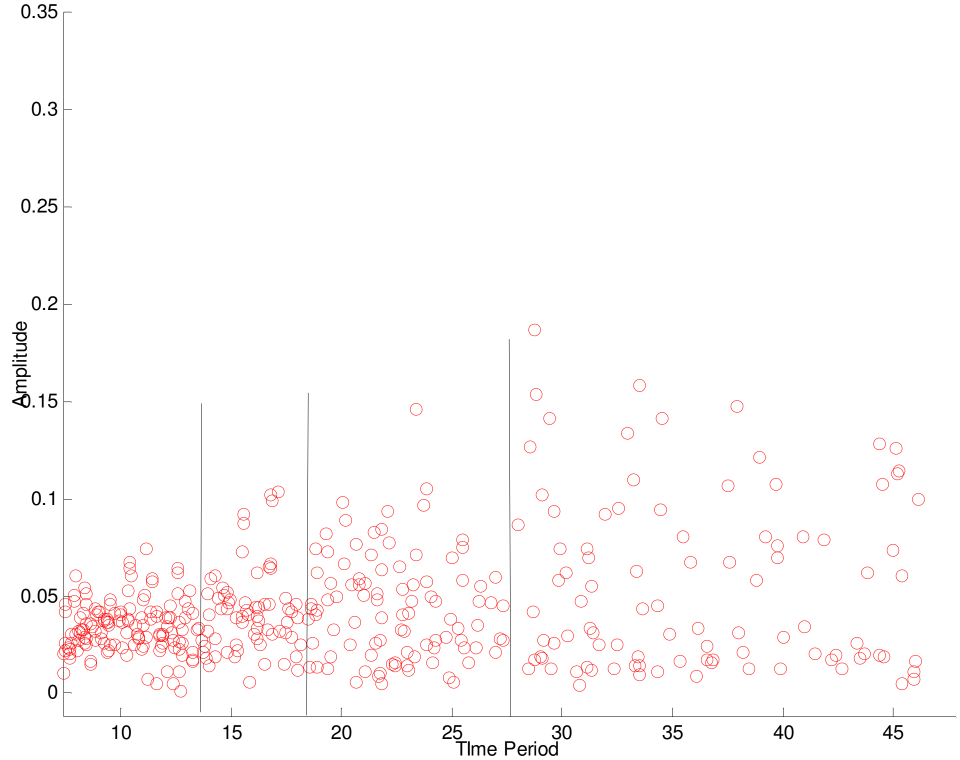

Figure 30: Further zoomed in version of spectrum showing more bands.

Figure 29 and figure 30 show further zoomed in versions of the spectrum. In this specific case of the stock ticker APH, the "break" periods which can be used to determine the band-pass filter are 13.7, 18, 27, 37, 54, 76, 108, 150 and 218. In general, we have observed that using the technique the break periods follow the thumb rule of:

Period(i+1)=2*Period(i-1).

4.2 Scales from the Price of Stock Ticker CRNT

As in the previous chapter, the smooth evolution and time invariance of the scales are presented as successive figures in a movie form in Movie 2 that is attached. The evolution of the scales along with the data over the period of one trading day, May 29,2009 is presented. Thus the movie is composed of 27 figures each separated by 15 minutes in real time. The data is presented in black with the model overlaid in red. The part of the data left as residue after the extraction of the scales is shown in cyan. The figure below shows a sample snapshot of the movie.

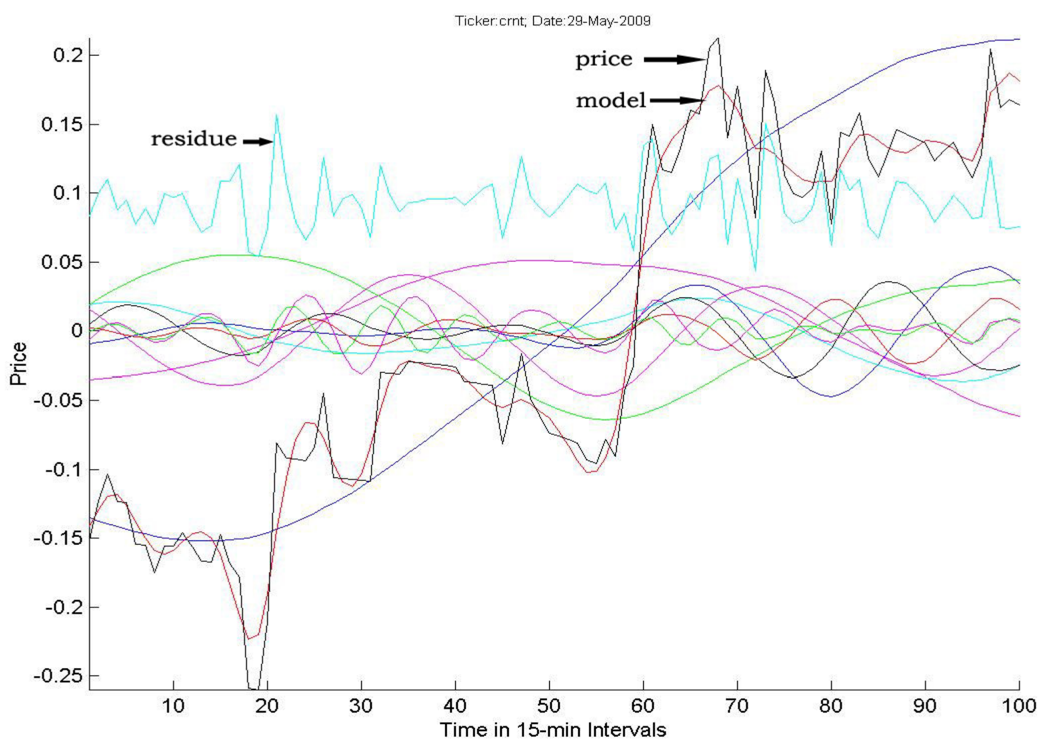

Figure 31: CRNT price, model and scales at 9:30am EST on 05/29/2009.

To bring home the consistency and time-invariance in the scales, we present the time history of the individual scales during the whole day. Figures 32-37 are created by arranging the history of each scale staggered by one point at a time. We can clearly see the clubbing together of these histories showing relatively no change especially in the low frequency scales.

In the higher frequency curves however we find four distinct bunching of the histories. This corresponds to the “reoptimization” process explained in chapter performed four times during the day. But it can be see that the divergence in the histories is stark at the peaks and valleys of the history of the scale. From a trading stand point, these “turning around” points in history correspond to very little actual contribution to the price change. Keep in mind that the price at any time is sum of these scales and the residue. So any scale contributes maximum to a change in price when it has the maximum slope. We can see that even after “reoptimization” the scale histories overlap during those crucial times of rapid change.

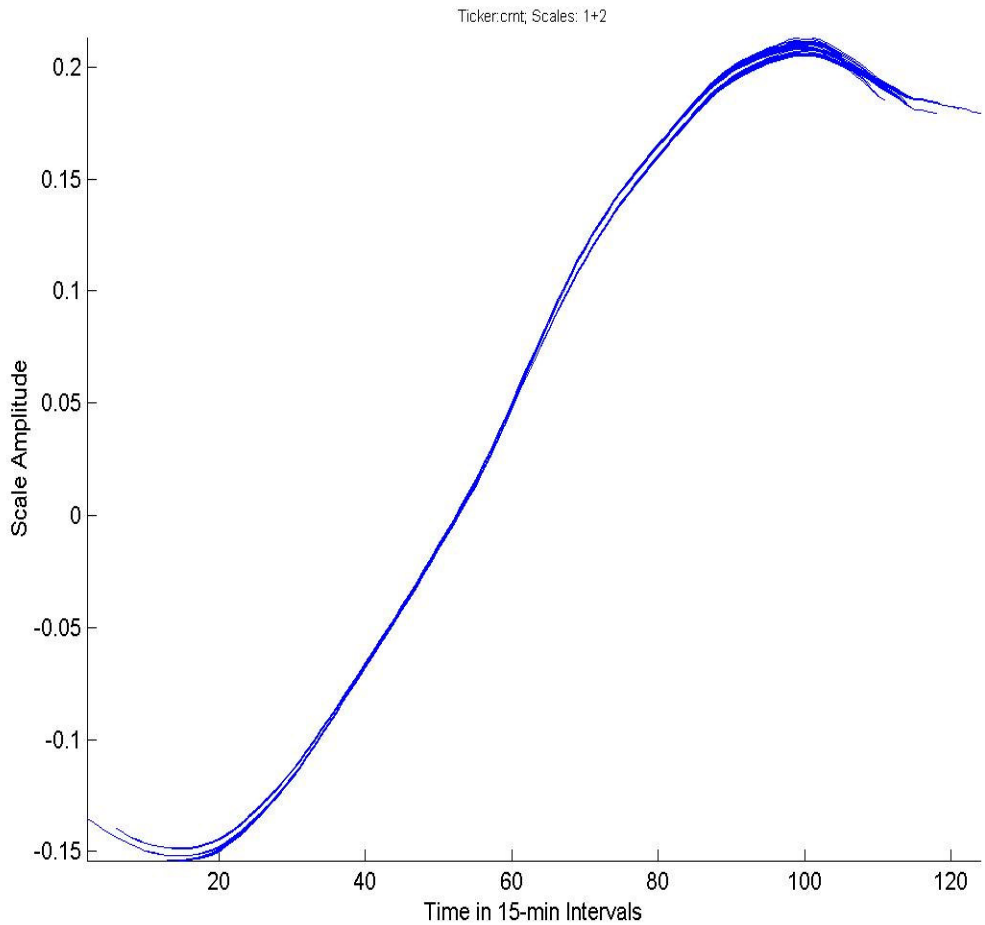

Figure 32: Overlapped histories of the sum of scales 1& 2 during a trading day.

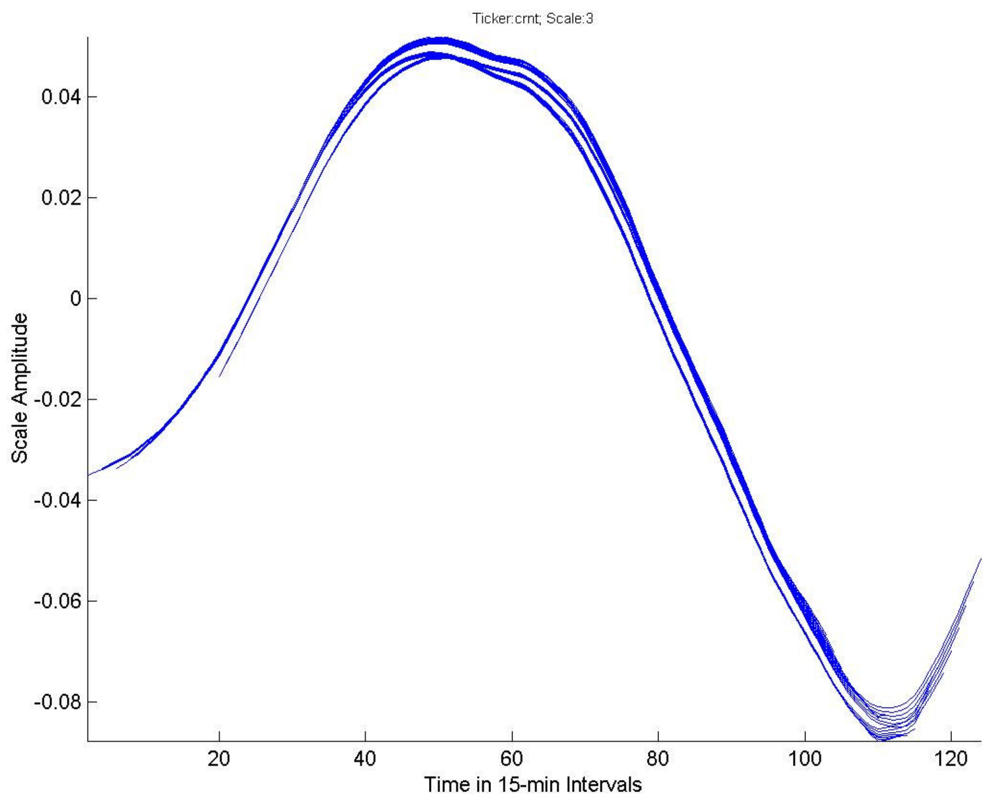

Figure 33: Overlapped histories of scale 3 during the progress of a trading day.

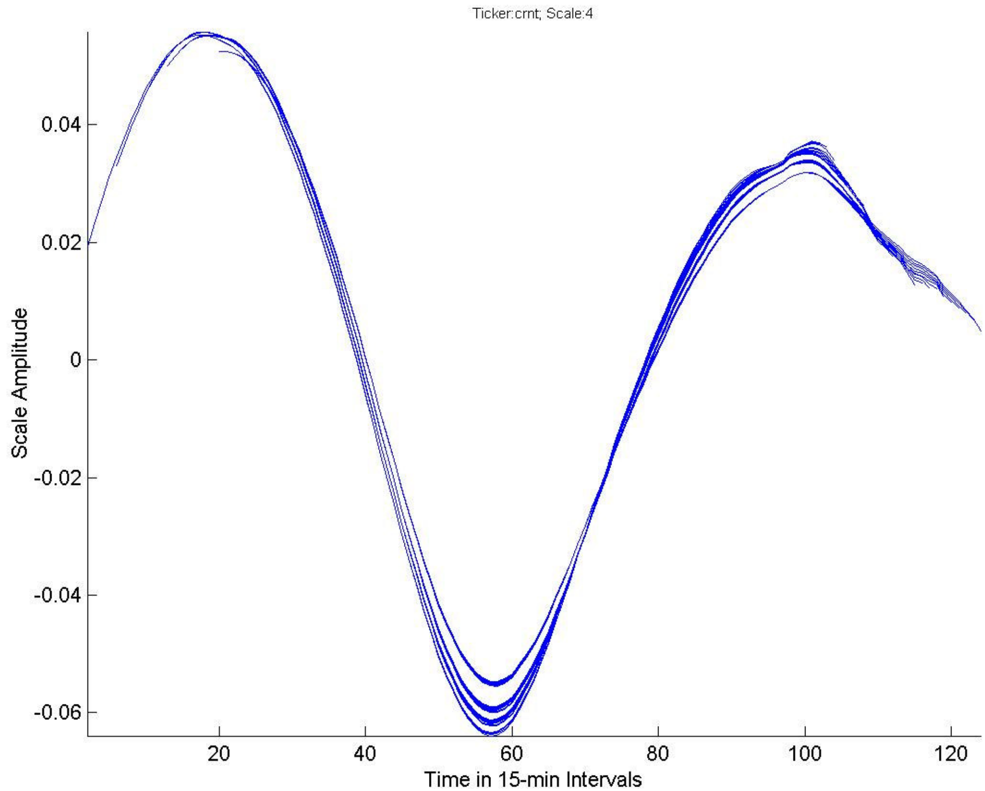

Figure 34: Overlapped histories of scale 4 during the progress of a trading day.

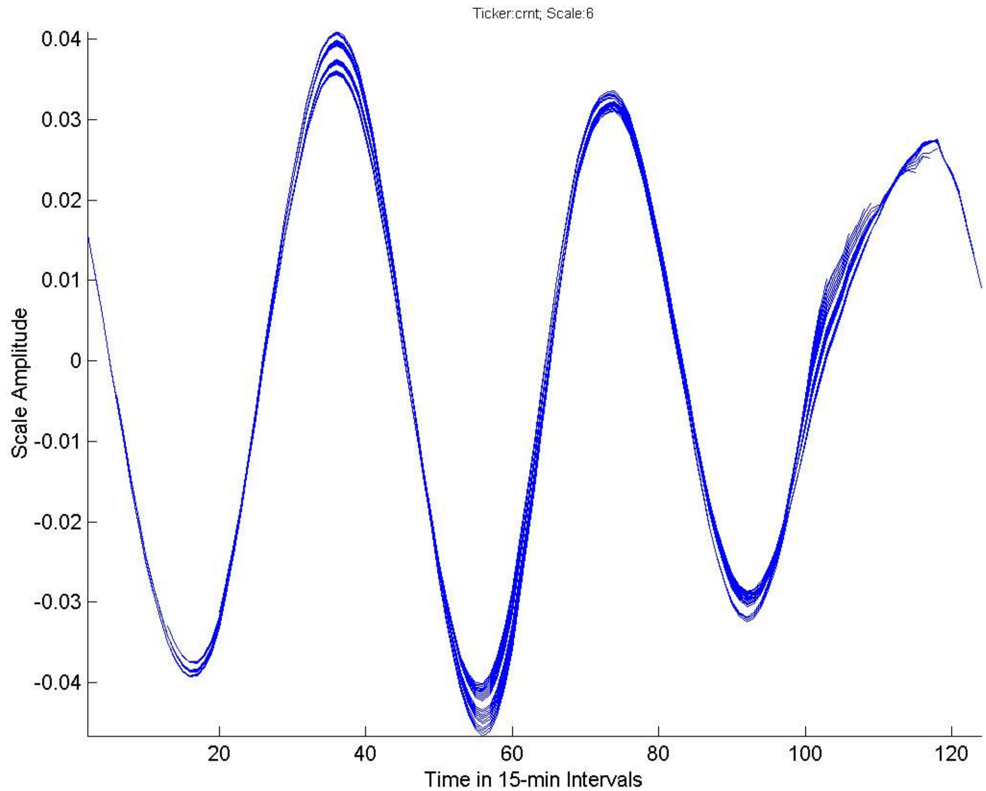

Figure 35: Overlapped histories of scale 6 during the progress of a trading day.

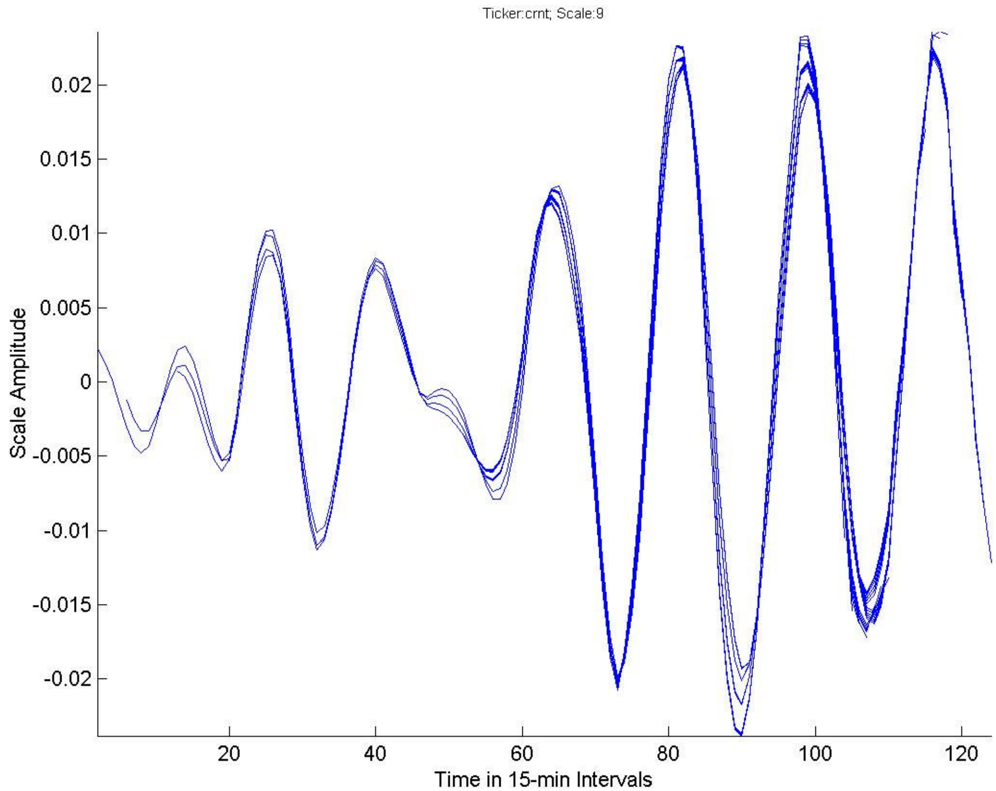

Figure 36: Overlapped histories of scale 9 during the progress of a trading day.

In the above figure, we can notice the scale coming "alive" in terms of magnitude. It will be good to recollect the market action of different channels of money becoming active and dynamical model analogy of energy transfer resonating through different frequencies of the system. Figure 37 shows the high frequency scale 11.

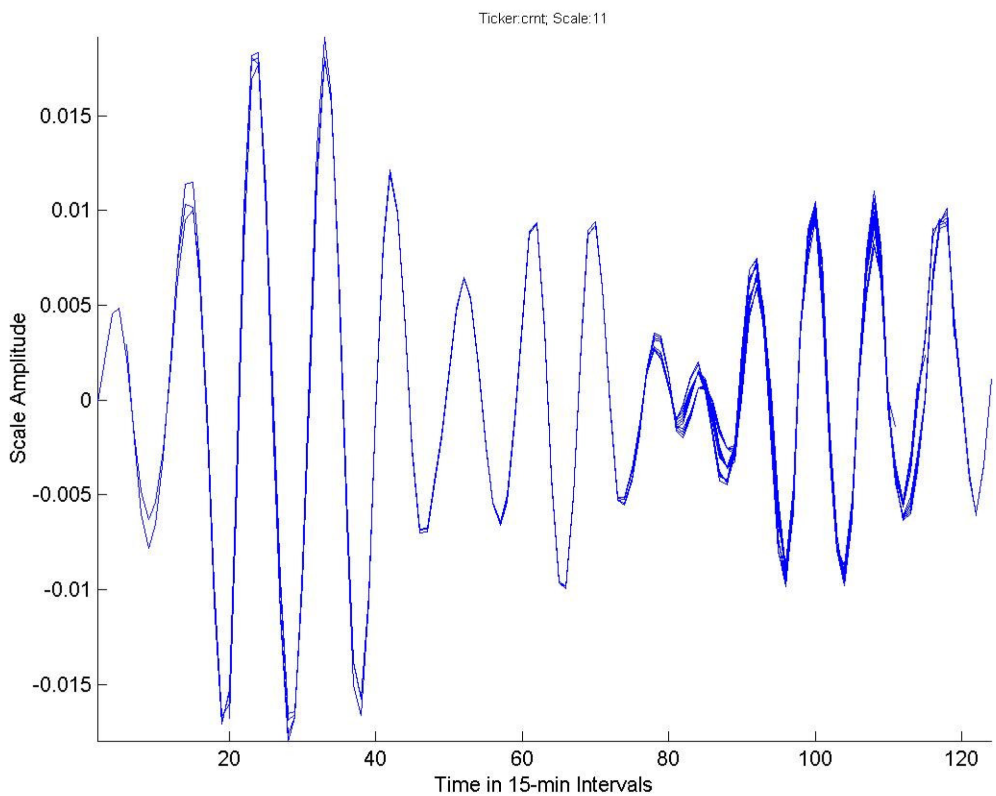

Figure 37: Overlapped histories of high frequency scale 11 during a trading day.

The purpose of technical analysis is to provide mathematical tools for trading and investment decision making. This demands that the products of any such technique be presented in a manner that can lead to easy and quick decision making. Towards this end, an Adobe Flash™ based Graphical User Interface was developed. This GUI is discussed in the next section.

As we conclude this section, it is worthwhile to recap two distinct features of the scales created by the new technique that sets it apart from existing technical analysis products.

1. The scales (oscillators) are not just indicators. They are components of the price movement. In other words, the sum of our oscillators is the price time history.

2. The scales (oscillators) are constructed such that they have a smooth and continuous evolution in time, with a large degree of predictability.

It is well known that natural systems have natural frequencies and exhibit resonance. The scales of a system can be considered as the channels of energy transfer available to it. In turbulence the changes in the finest level can lead to divergence and bifurcations at the coarsest level. But if we are aware of the levels in between through which such energy transfer is going to take place, we can better model the system's internal dynamics be able to anticipate its behavior. The changes in the smallest levels can move up the system through a kind of "hierarchical resonance". In the case of the scales identified for a stock market security, by considering the changes that happen to the fastest scales when a "fundamental" event, such as news released, occurs, and how this input energy is transferred to the slower scales in succession we can quantify the "magnitude" of the external excitation as well as monitor possible changes in the internal dynamics, which however happens at a much slower rate than the occurrence of fundamental events.

4.3 The GUI to Assist Decision Making

Going back to the idea of different channels of money that lie hidden in the seemingly random fluctuations of a stock price, we can say that what the scales capture is “the state of the market” for the particular stock price. The direction and slope of a scale can be considered to indicate whether money is coming in or going out of the stock there by leading to an upward or downward movement in the price. Clearly the surest movement in the price comes when there is a synchronized movement of the scales in a particular direction. Such synchronized movement also implies lowest risk for the associated trade. Thus the technique provides the invaluable “low risk: highest reward” combination.

To allow, traders and investors to operate at speeds most convenient to them, the GUI presents a 4 panel layout clubbing scales according to their relative speed of oscillation into Slow, Mid and Fast.

Figure 38 and Figure 39 show this layout for end-of-day data and intraday data respectively.

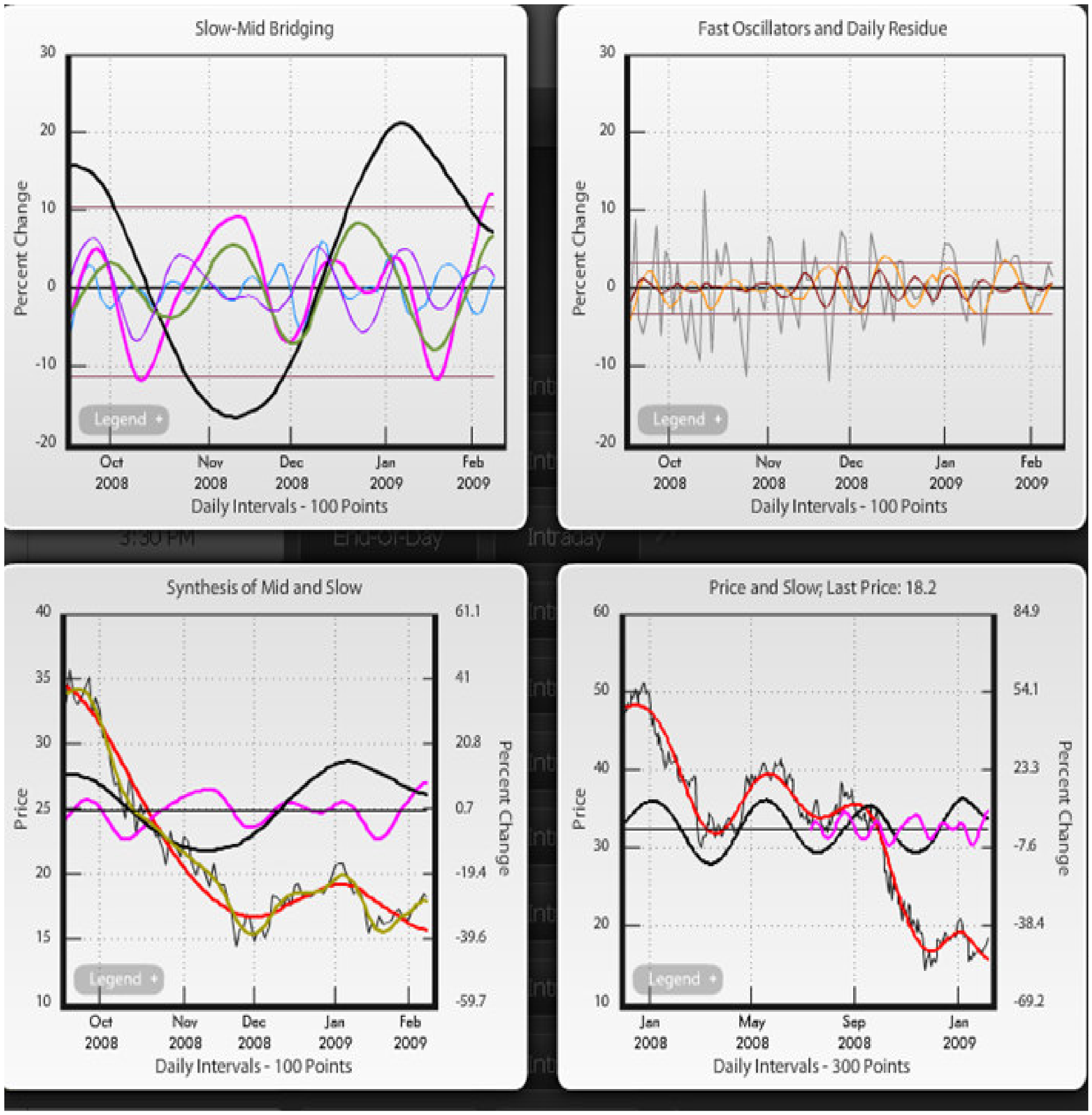

Figure 38: Four panel layout for end-of-day data.

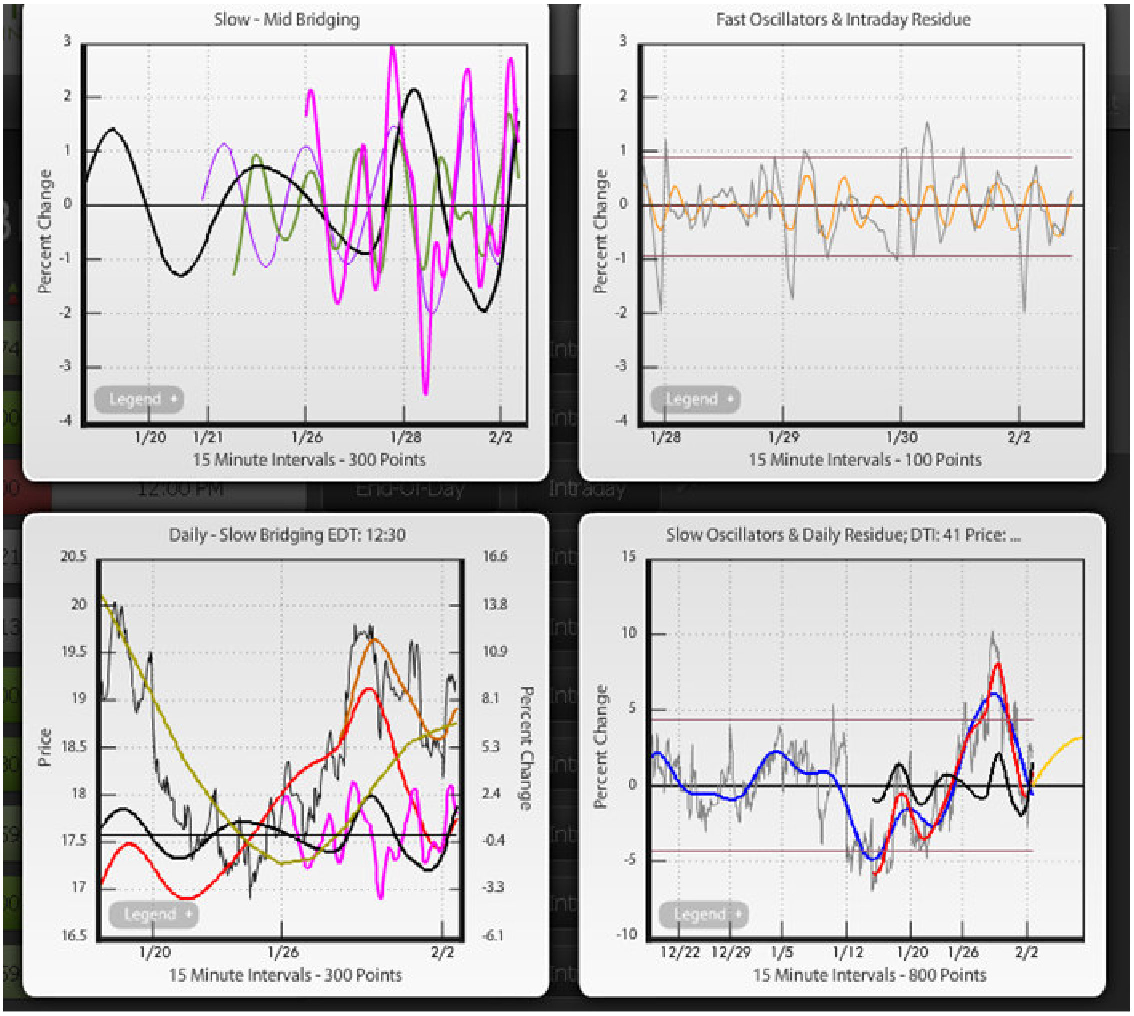

Figure 39: Four panel layout for intraday data.

The Flash™ based GUI allows the user to enlarge any of the panels for further study. Also the data and scales in any of them can be zoomed into because the graphs are generated from the corresponding numerical data upon such requests. Each chart also has a legend that can be used to identify the various lines and turn them on and off from the display. Each scale is identified by its particular color so that it is easy to spot it when it appears in more than one chart in the layout.

Figure 40 presents the enlarged view of the bottom left chart for the intraday data. This chart contains the actual price history and some of the most significant scales used for decision making. Each trading decision can be broken down to an intuitive three step process. First we can pick the dominant oscillator for a particular stock using its correlation with the actual price and use it to determine whether we will buy or short sell.

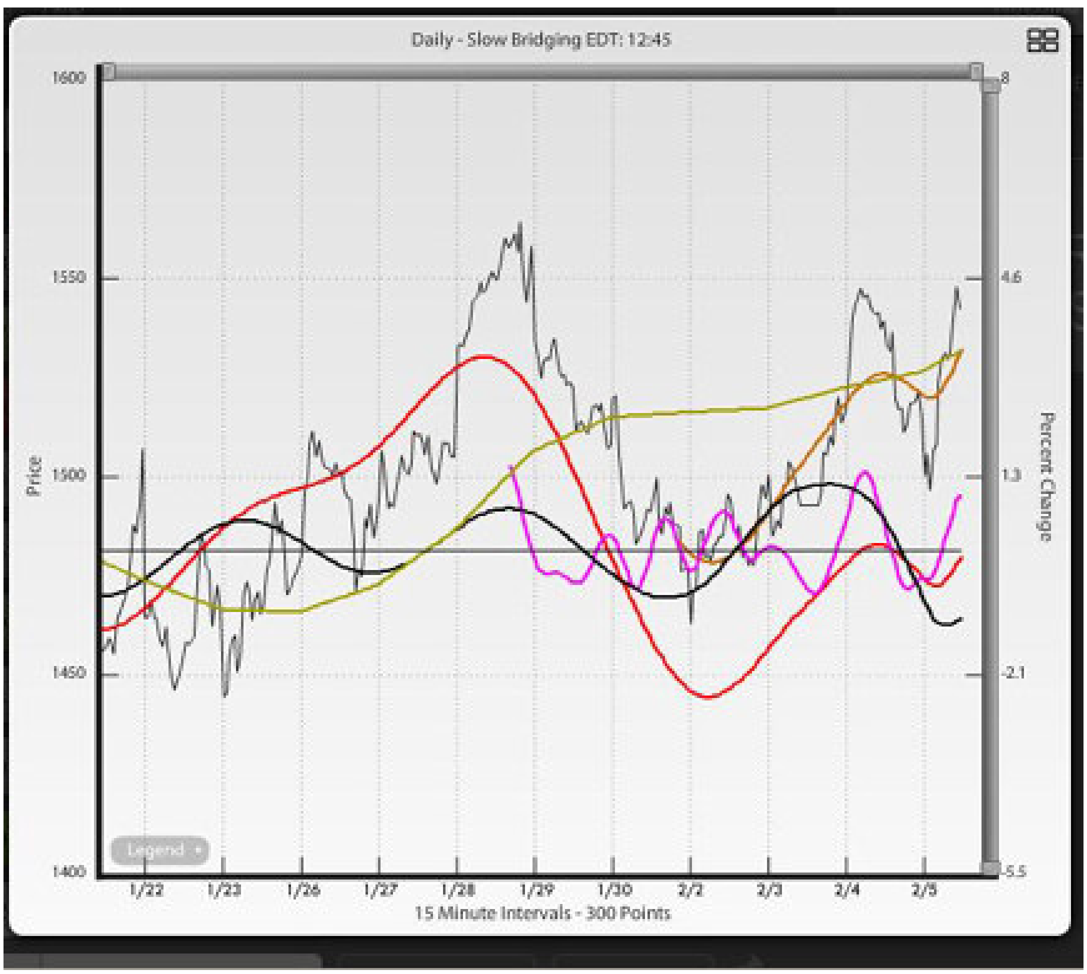

Figure 40: Enlarged view of a chart with the price history and some significant scales.

If the duration of the trade is our primary focus then hunt for tickers where the dominant oscillator swings in our preferred time frame be it few hours, couple of days or even weeks. Then to reduce the risk of trading in the direction of the dominant oscillator we check the oscillators slower than it for disagreement. Finally, entries and exits can be timed by using the faster oscillators.

4.4 Example of a Successful Trading Decision

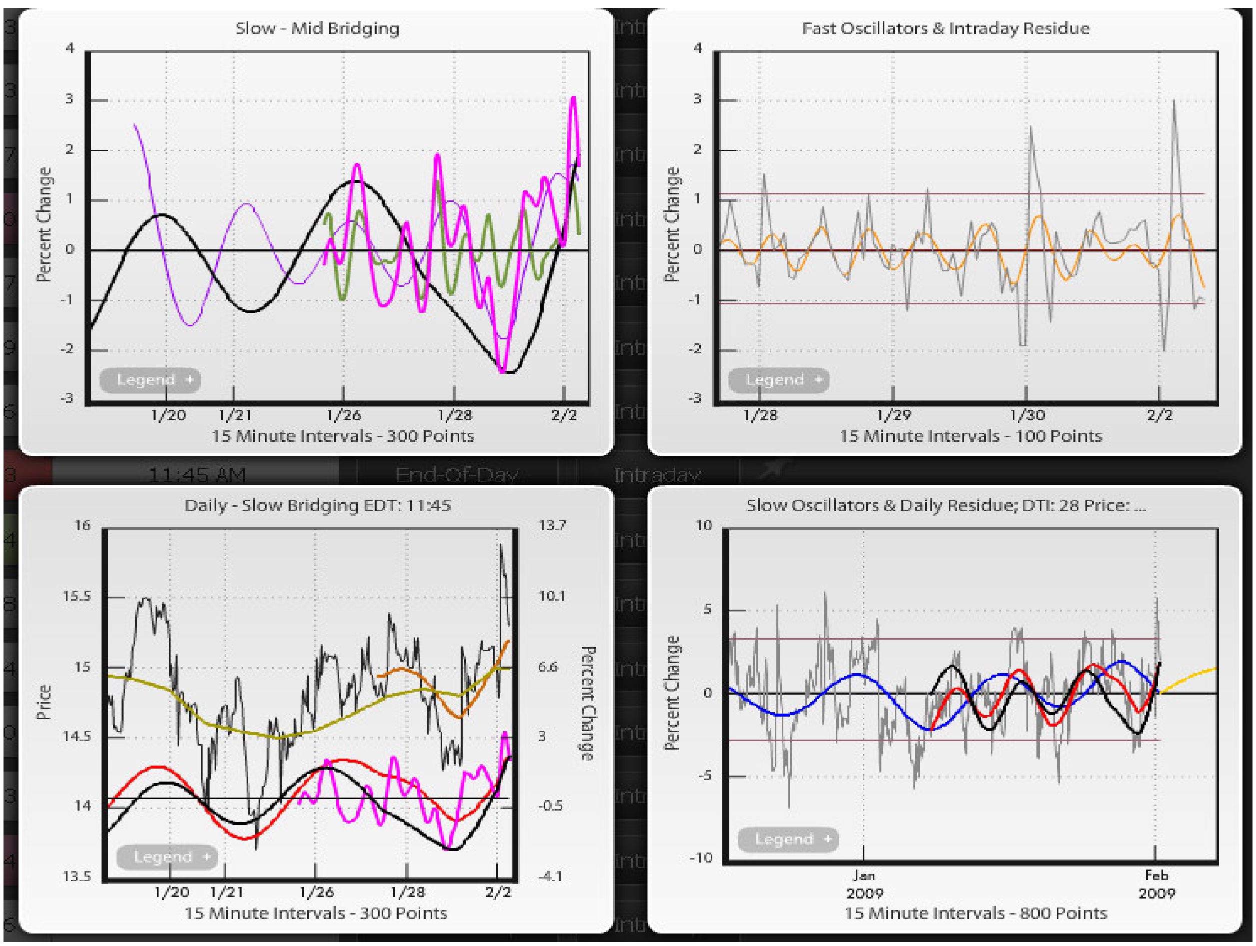

Figure 41: Intraday 4-panel layout for stock ticker CMN at 11:45am EST on 02/02/09.

Consider the intraday chart of ticker CRNT shown in figure 41. We see that the Mid Oscillator which is plotted in pink color on the both the charts on the left can be chosen as dominant oscillator based on its correlating well with the actual price movements shown in the bottom left chart. It is bigger in magnitude and better correlated than the Slow oscillator (red) with the price. The fast oscillator seen on the top right chart has small magnitude. Since the mid oscillator has turned around to move downwards we will consider this situation to be suited for short selling the stock. The short sell here is safe because Slow Oscillator is turning (Slow B in blue is already moving down and Slow A in black has exhausted its upward motion).

Since we are focused on the Mid oscillator, we shouldn't be worried about upward movement that may occur for a short period in the price. If we execute a trade based on this graph, our entry price will be $15.30.

Figure 42 on the next page shows the situation on the next day, 02/03/09 by 14:00 EST, we can see that the dominant Mid oscillator (pink) has turned around to the upward direction. The Slow oscillator is also turning around. This would indicate a right time to buy-to-cover and exit. Our exit price would be $14.50 which corresponds to a profit of 5 % in a span of 24 hours.

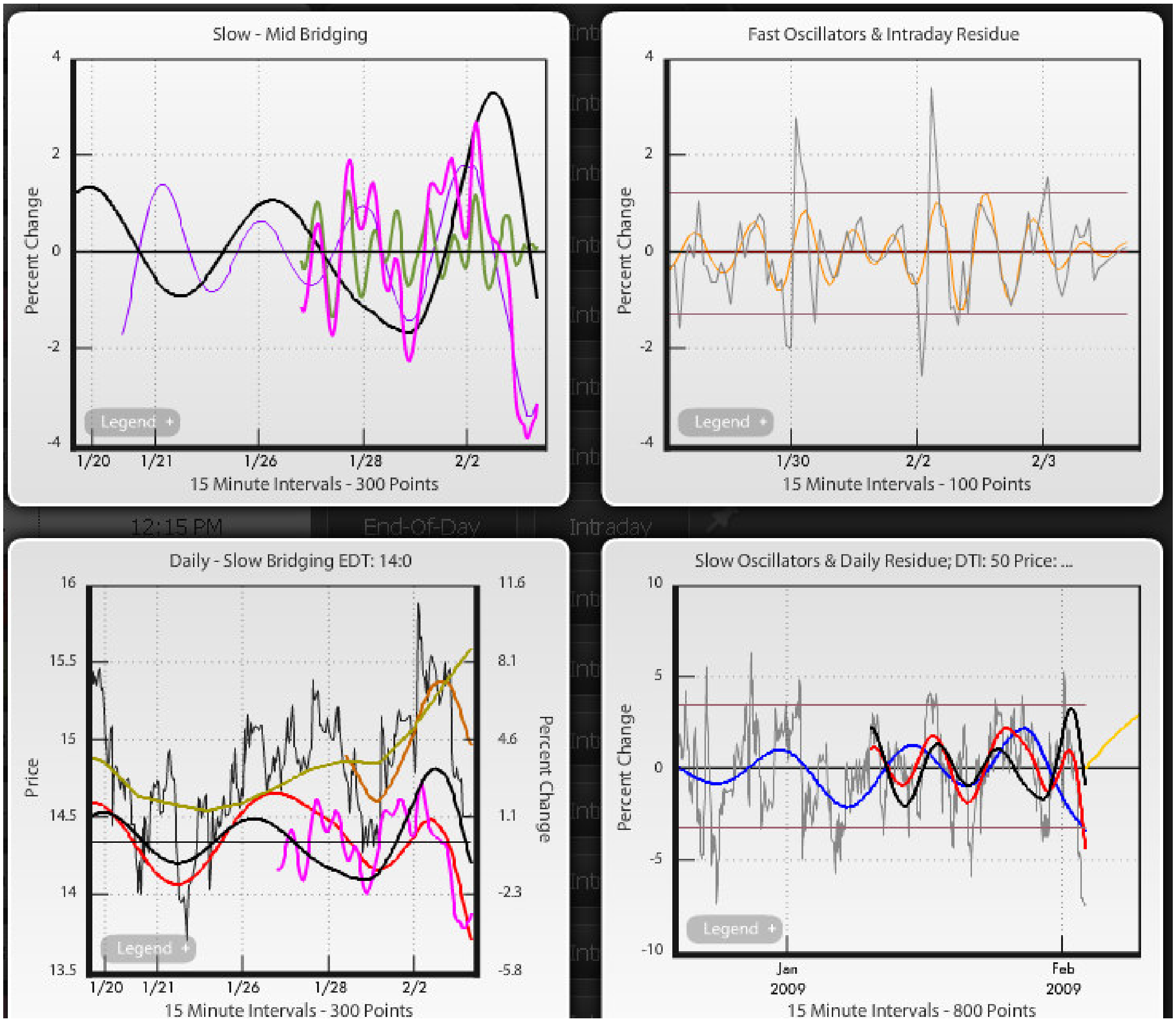

Figure 42: Time to exit based on turning around of the dominant oscillator.

In the structured environment of the stock market where the apparently diverse stimuli can be grouped into a handful of adjectives (like the media does using words “good”, “bad”, “indifferent” etc.) and the channels of behavior are severely limited (“buying”, “selling” or “refraining”), it should not be a big wonder that mass psychology has developed time invariant rhythmic patterns which can be mathematically extracted.

Our confidence in the fact that the cup on the table will not bounce off by itself though its individual atoms are in random chaotic behavior comes from the law of large numbers and knowledge of chemical bonding. Each individual might be an unpredictable system moment by moment during his life time, but when he assumes the role of a trader or an investor and becomes part of a massive system like the stock market, it appears that we can create and apply tools that can estimate future behavior in context.

CHAPTER V

APPLICATION TO METEOROLOGICAL DATA

In this chapter, we present the results of the application of the new technique to historical temperature data. Section 5.1 presents briefly the philosophy behind the approach. In section 5.2 we present the data that was used together with the extracted scales and the predictions.

5.1 Approaching Weather Systems

Weather systems remain out of human control except for instances of cloud seeding, which is still in its early stages. However, prediction of meteorological variables is paramount. Though it is strictly not an aerospace application, predicting atmospheric temperature, pressure, wind speed and direction, etc., are crucial in aerospace engineering experiments and applications.

The inspiration for our research was the similarity between turbulent flow behavior and stock market securities price movements. Former Federal Reserve Chairman Alan Greenspan even calls ours an economic “Age of Turbulence”. Thus, though our technique was developed and tested on stock market securities, it can be applied to seemingly random atmospheric variables. Atmospheric activity has a very large amount of parameters influencing it which makes reliable modeling hard to impossible. With millions of traders making investment decisions, the situation is similar in the stock market. But with sufficient past data available, it can be mined to find the existence of coherent structures which can then be used to make reliable short term predictions of good accuracy. The failure of most physical models developed in meteorology stems from missing out on crucial parameters that wax and wane in importance over time. Our models differ from these in two crucial aspects.

1. We don't try to force fit the data into predetermined set of parameters. Our parameters are numerical, determined from the data and are allowed to evolve over time.
2. Long term predictions are not attempted so that the higher order random nature of the system is never challenged. The hugely popular documentary, 'An Inconvenient Truth', on the harmful effects of the phenomenon labeled 'global warming' had the unforgettable scene where Mr. Al Gore uses a crane to raise him to the highest point of a graph projected on the screen. The method of extrapolation used to create that graph by taking Paleoclimatology data and extending it to the future is similar to the use of extrapolation many hedge fund managers use in their promotional brochures to entice potential customers with a graph of spiking future profits!

In our technique changing the discrete time interval in which data is collected can lead to predictions on a different scale which can then be used to fine-tune the short term predictions. For example, generating predictions based on hourly data can be crossed checked for reliability and accuracy with daily data which in turn can be made to fall back on weekly, monthly data.

5.2 Modeling Temperature Data

Hourly temperature data for the city of Des Moines, Iowa for five years from the year 2000 to the year 2004 was obtained from the Iowa Environmental Mesonet website maintained by the Department of Agronomy of the Iowa State University. The website provides the historical data for several different weather variables at different time interval. Hourly temperature was chosen based on the preview graph that can be viewed on the website. Temperature data resembled both stock data and nonlinear dynamical system data that had been successfully modeled by the technique. It must be kept in mind that temperature is just a convenient variable that we use to describe a weather system or climate both because we have convenient ways of measuring it through other easily measurable quantities like length of a mercury column or expansion or change in conductivity of other material that occur due to change in temperature. Temperature is neither an independent quantity in the context of the weather system nor does it provide an overarching blanket that can be used to make qualitative assessment about the weather or climate of a large geographical area. Nevertheless at a particular spot where the measurement is made knowing the pattern in temperature is very useful for the planning of several activities like farming, aquaculture etc. We have to remember that just like in the context of the stock market where our technique is used to make an assessment of the state of the market for a particular stock at a given time and there by assisting short term decision-making, we have be very clear of the context of modeling a quantity like temperature and limits of its applicability.

For the hourly temperature data provided by the website, a flat line was used to interpolate the data points wherever they were missing in the original available data. From the hourly temperature data collected, daily data was extracted for the purposes on modeling. Figure 43 shows this data.

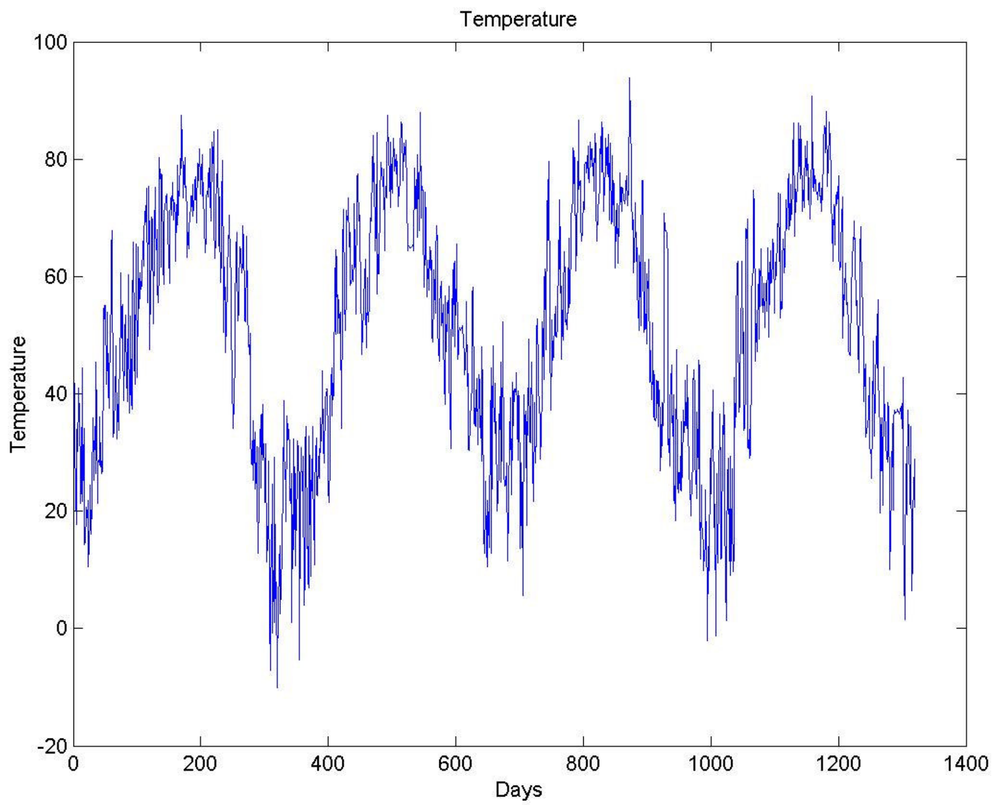

Figure 43: Temperature data used for modeling.

After the daily model was created, it was applied to the hourly data so that the residue created from their difference could be used to extract the finer scales. A combination of these scales from the two sets of data from different time scales were used for the final prediction. One hundred day windows of mean subtracted data were used to extract the scales. As in previous chapters, we now present the data in black, model in red and extracted scales for consecutive days as a movie in file Movie 3 to show the invariability and continuous nature of the scales over a period of a month. In figures 44-49, we present the 10 day predictions produced using the technique. The predictions are provided along with the actual temperature changes in the predicted period.

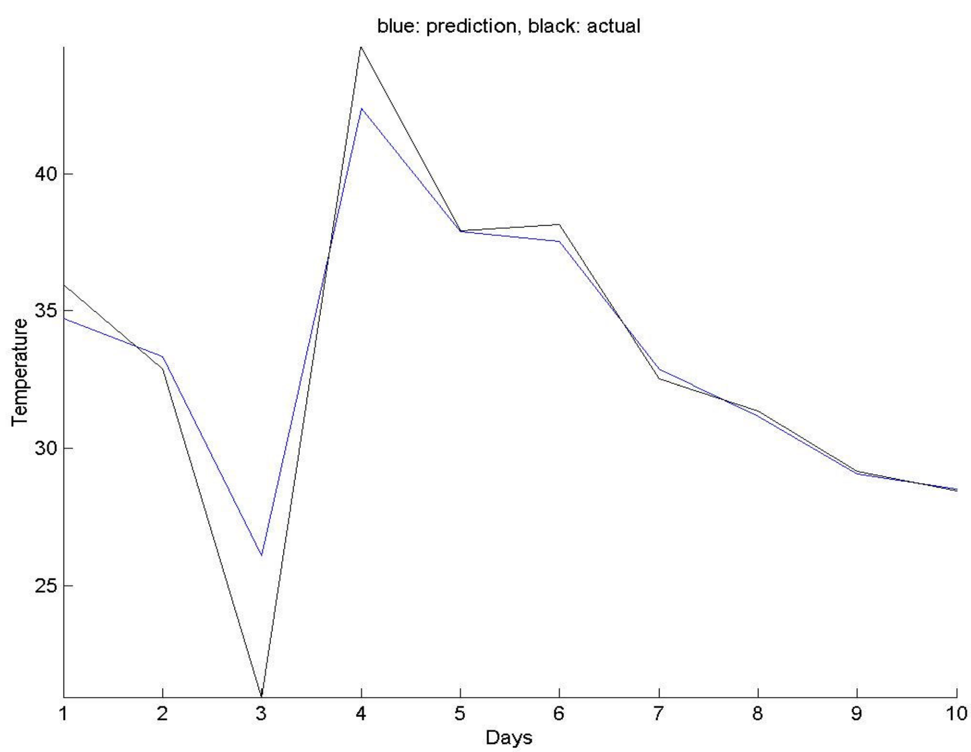

Figure 44: Example 1 of 10 day prediction compared to actual temperature change.

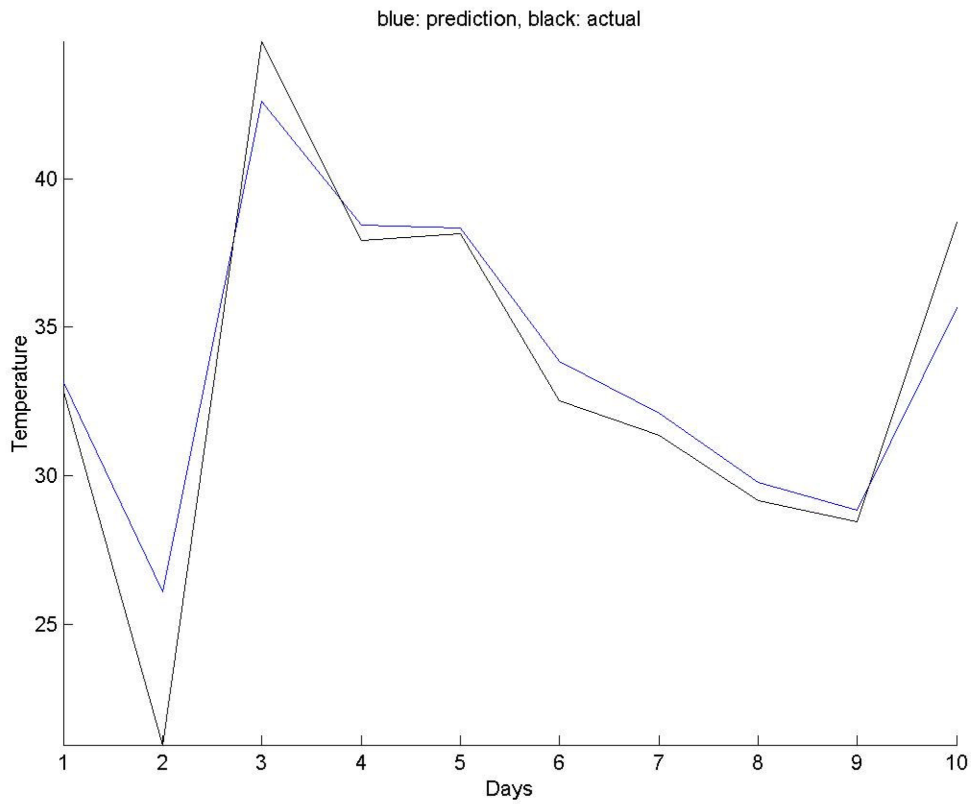

Figure 45: Example 2 of 10 day prediction compared to actual temperature change.

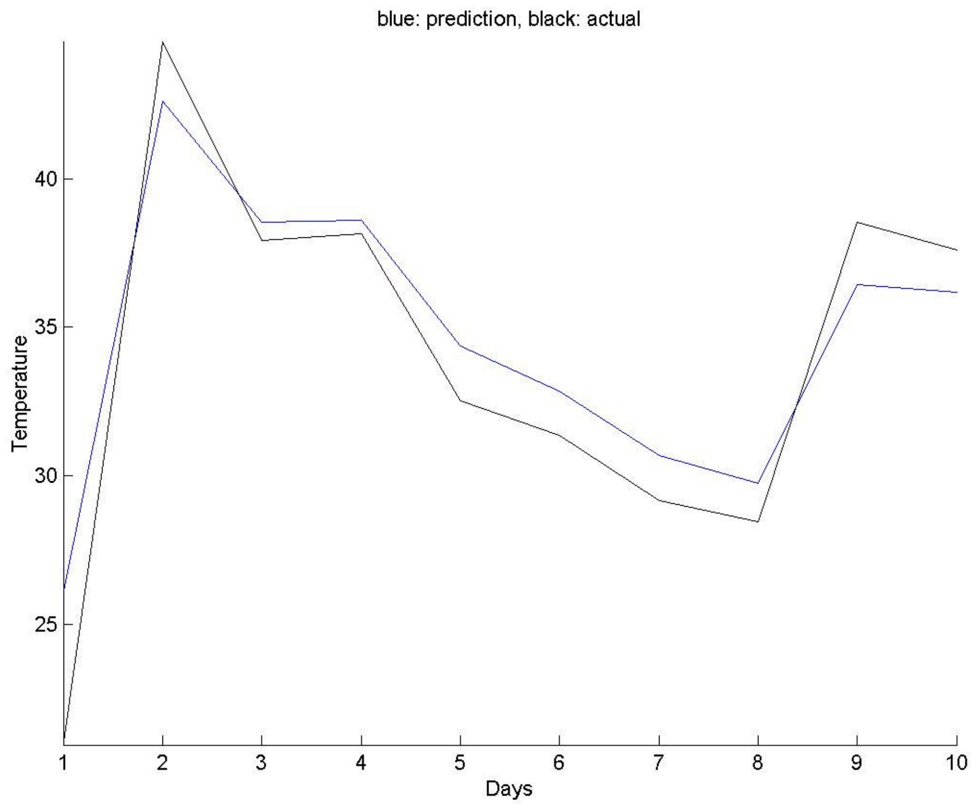

Figure 46: Example 3 of 10 day prediction compared to actual temperature change.

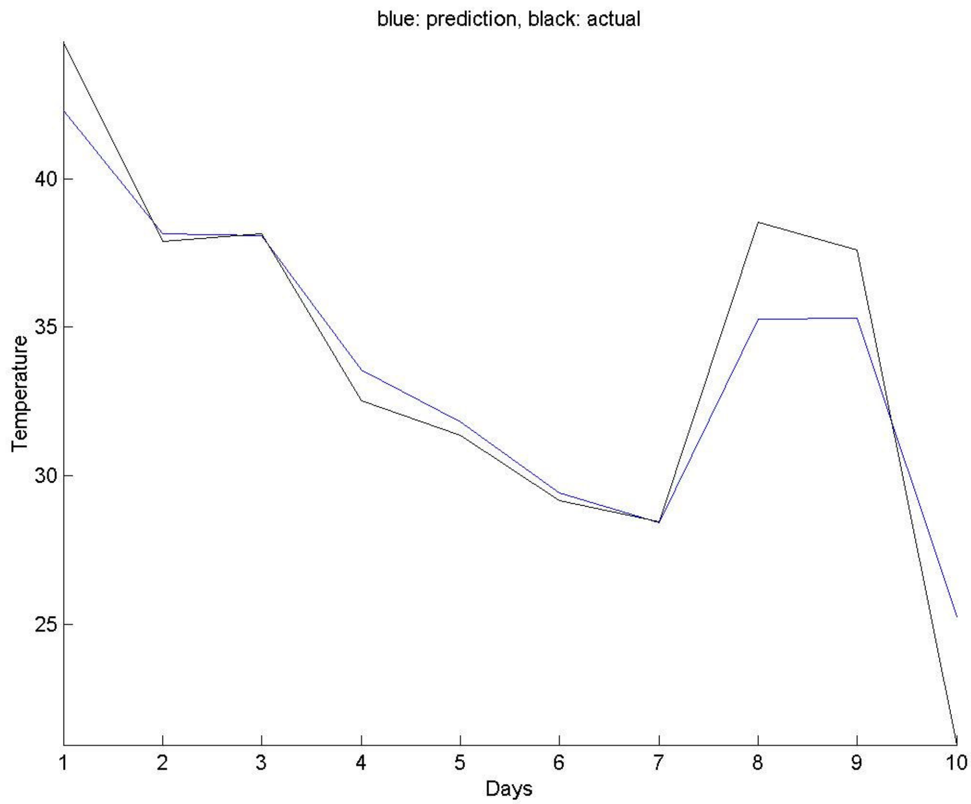

Figure 47: Example 4 of 10 day prediction compared to actual temperature change.

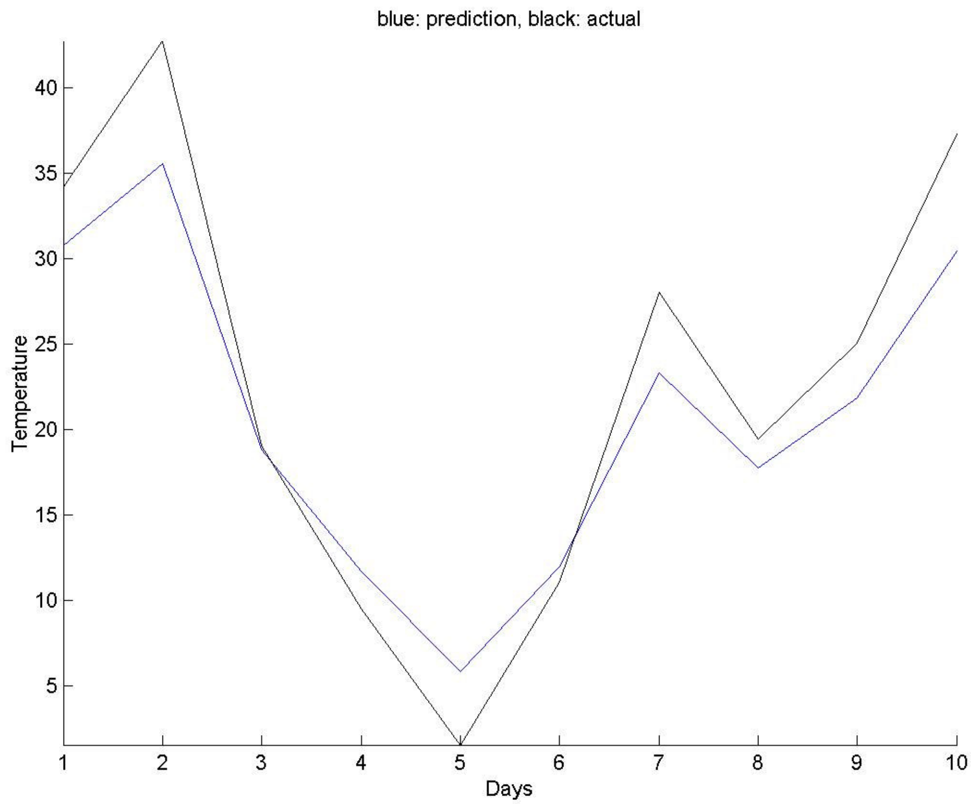

Figure 48: Example 5 of 10 day prediction compared to actual temperature change.

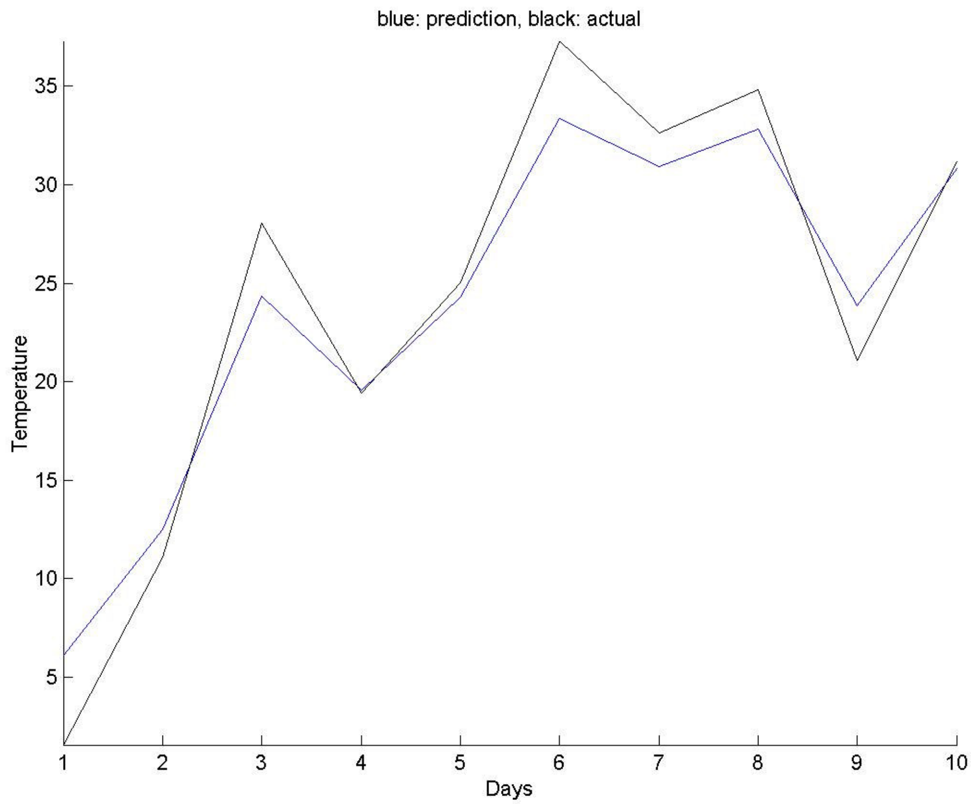

Figure 49: Example 6 of 10 day prediction compared to actual temperature change.

It is clear from the previous figures that the technique provides excellent predictions for a variety of temperature fluctuation scenarios. However, the prediction capabilities over 10 day periods are not universal as shown by figures 50-51. But in these figures the scale of the y-axis shows much higher resolution than previous figures, so the prediction are not as flawed as they appear to be.

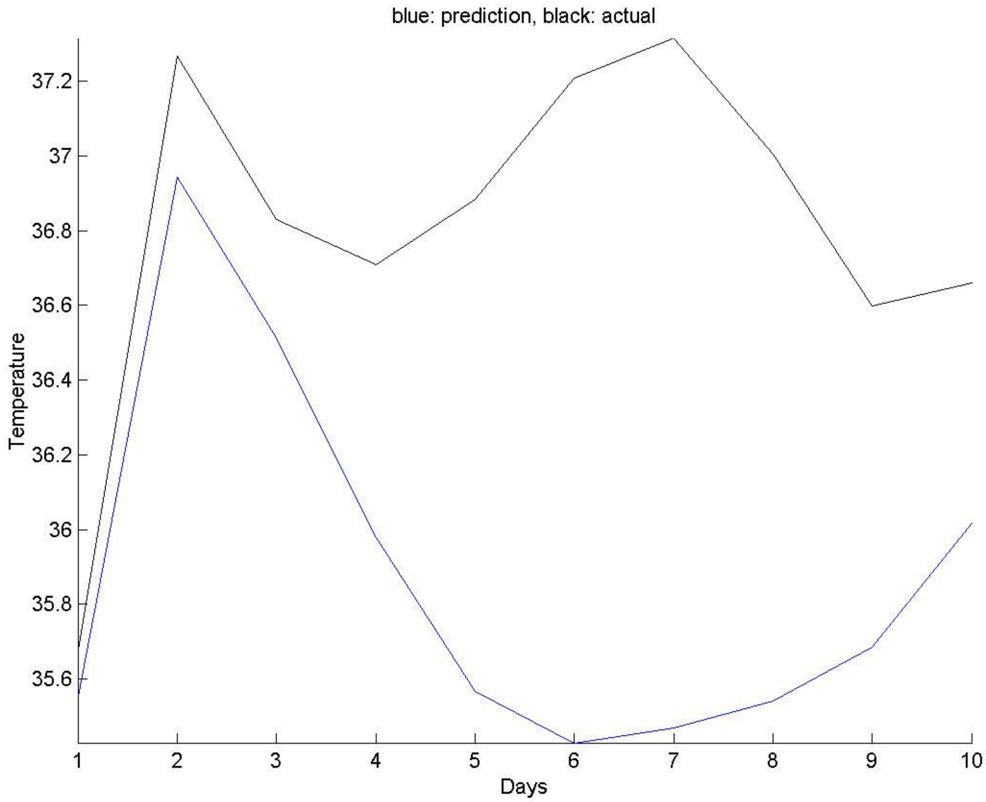

Figure 50: Example 7 of 10 day prediction compared to actual temperature change.

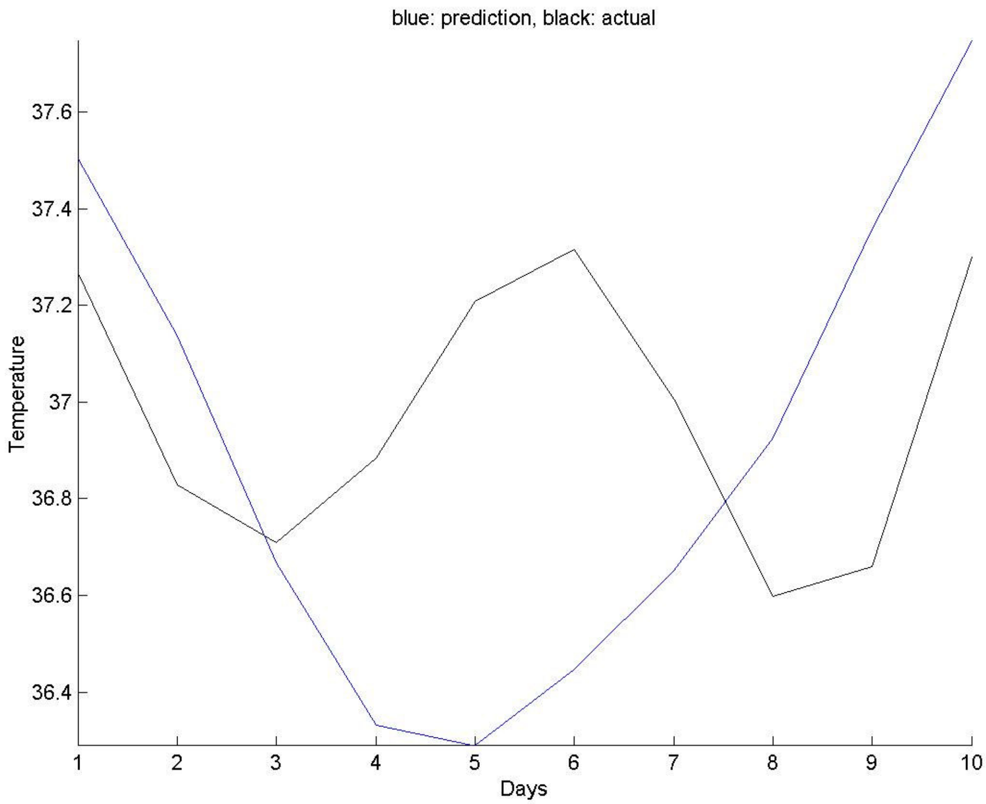

Figure 51: Example 8 of 10 day prediction compared to actual temperature change.

CHAPTER VI

CONCLUSIONS

Although highly complex systems may be perceived as seemingly random and hard or impossible to predict reliably, almost invariably "coherent structures" exist in its behavior. For example, the crowd activity in the financial markets is beginning to find causal explanations in the new field of behavioral finance. It has become clear that human beings suffer from several cognitive flaws which make it difficult for them to estimate probabilities and use them for decision making objectively while operating under different environments. As a result the fundamentals of human psychology become dominant factors in market movements. In the research presented in this dissertation we developed a method for modeling complex physical systems, even systems that would not ordinarily be perceived as mechanical or engineering systems, by identifying and extracting underlying "coherent behaviors" or "scales" with distinct time invariant properties that capture the main physical processes taking place in the system and their interactions. These "scales" are subsequently used to create a dynamical model for the system which is in turn used for forecasting.

Such a traditionally thought of as a non-engineering problem is the modeling of mass human psychology associated with the economy and the stock market in particular. The method was applied to the price behavior of stock market securities. It was found that not only do smooth coherent structures exist within the stock price but that they can be used to effectively capture the current state of the market from which predictions about future swings can be made that can be successfully exploited. With the single stream of input data, not only were the coherent structures of the system extracted, but we also estimate impulse forces which can model the external stimuli that act on the system. In the case of the stock market, when the fast scales diverge from their evolutionary path, we can sense that something "fundamental" has happened in the market and we use an impulse to model this into the system.

In the process of developing the technique, several mathematical tools were also developed. One of them was a spectral decomposition method that eliminates the drawbacks of commonly used frequency analysis techniques based on Fourier methods and digital filters. Frequency leakage in spectral analysis was significantly reduced and end effects nearly eliminated. The technique is geared towards short term prediction thereby respecting the inherent randomness of systems whilst formulating a practically useful model for forecasting purposes. Since the mathematical method has its foundations on a physical dynamical system analogy, it was also successful used to model and predict with good accuracy a real nonlinear dynamical system. It pays attention to smallest fluctuations so that unexpected bifurcations don't result from data that could have been dismissed as an "outlier". It borrows physical insight from the theory of turbulence and how processes in the finest of scales can influence the large scales through hierarchical resonance.

For practical application of the technique in the stock market, a graphical user interface, that can assist a trader or an investor to make decisions in real-time, was developed. By continuously adapting to the market conditions and price fluctuations without losing their time invariance properties, the scales provide an invaluable visual tool to the trader. Also, the initial "automation" of the technique (computer-made trading decisions) shows excellent performance.

The utility of the new technique in the analysis and modeling of natural complex systems was illustrated through its application to the forecasting of temperature data. The hourly data collected was split into two time scales of daily data and hourly. First from the daily data, a model was created and this was used to extract the finer intraday hourly change which was used to extract the finer scales. With the combination of scales coming from two different time scales of data, predictions were successfully made.

The successful application of the technique in both natural and man-made complex systems, like the weather and the stock market shows its versatility. The technique can without doubt be found useful in various other systems like airline reservation, restaurant supply chains etc., where human behavior needs to be modeled and anticipated. It can also be tried on seemingly complex natural systems provided we can isolate variables that can be measured and used for modeling.

REFERENCES

[1] Klein, J., *Statistical Visions in Time*, Cambridge University Press, Cambridge, 1997.

[2] Freedman, D., *Markets and Morals*, Palgrave Macmillan, New York, 2008

[3] Aldrich, J., "Probability and Depreciation: A History of the Stochastic Approach to Index Numbers," *History of Political Economy*, Vol. 24, 1992, pp. 657-687.

[4] Bell, E. T., *Men of Mathematics*, Touchstone, New York, 2008.

[5] Ubel, P. A., *Free Market Madness*, Harvard Business Press, Boston, 2009.

[6] Box, E. P., and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, 1970.

[7] Bernstein, P. L., *Against the Gods: The Remarkable Story of Risk*, Wiley, New York, 1998.

[8] Lowenstein, R., *When Genius Failed: The Rise and Fall of Long Term Capital Management*, Random House, New York, 2001.

[9] Dunbar, N., *Inventing Money: The Story of Long Term Capital Management and the Legends behind It*, John Wiley & Sons, Inc., New York, 1999.

[10] Taleb, N. N., *The Black Swan: The Impact of the Highly Improbable*, Random House New York, 2007.

[11] Freedman, R. S., *Introduction to Financial Technology*, Academic Press, Burlington, 2006.

[12] Brown, C.M., *Technical Analysis for the Trading Professional*, McGraw-Hill, New York, 1999.

[13] Murphy, J. J., *Technical Analysis of the Financial Markets*, New York Institute of Finance, Paramus, NJ, 1999.

[14] Murphy, J. J., *Intermarket Technical Analysis*, John Wiley & Sons, Inc., New York, 1991.

[15] Bulkowski, T.N., *Encyclopedia of Chart Patterns*, John Wiley & Sons, Inc., New York, 2000.

[16] Hurst, J. M., *The Profit Magic of Stock Transaction Timing*, Traders Press Inc., Greenville, SC, 1970.

[17] Kahneman, D., “Maps of bounded rationality: A perspective on intuitive judgment and choice,” *Les Prix Nobel*, edited by T. Frangsmyr, The Nobel Foundation, Stockholm, 2002.

[18] Brafman, O., and Brafman, R., *Sway: The Irresistible Pull of Irrational Behavior*, Broadway Business, New York, 2008.

[19] Goldberg, J., and Rudiger, N., *Behavioral Finance*, Wiley & Sons, Inc., New York, 1999.

[20] Gilbert, D., *Stumbling on Happiness*, Random House, New York, 2006.

[21] Town, P., *Rule #1*, Three Rivers Press, New York, 2007.

[22] Kupiec, J., *The Origin of Individuals*, World Scientific Publishing Company, Hackensack, NJ, 2009.

[23] Attenborough, D., *The Private Life of Plants*, Princeton University Press, Princeton, NJ, 1995.

[24] Rediniotis, O., Surendran, A., “Low Order Modeling and Forecasting of Stock Market Securities,” *Computational Stochastic Mechanics: Proceedings of the*

Fourth International Conference on Computational Stochastic Mechanics, IOS Press, Amsterdam, 2003.

[25] Aubury, N., Holmes, P., Lumley, J.L., and Stone, E., “The Dynamics of Coherent Structures in the Wall Region of a Turbulent Boundary Layer,” *Journal of Fluid Mechanics*, Vol. 192, 1998, pp. 115-173.

[26] Berkooz, G., Holmes, P., and Lumley, J.L., “The Proper Orthogonal decomposition in the Analysis of Turbulent Flows,” *Annual Review of Fluid Mechanics*, Vol. 25, 1993, pp. 539-575.

[27] Anfinsen, C., "The Formation and Stabilization of Protein Structure," *Biochemistry Journal*, Vol. 128 (4), 1972, pp. 737–749.

[28] Berg, J. M., Tymoczko, J. L., Stryer, L., *Protein Structure and Function*, W.H. Freeman, San Francisco, 2002.

[29] Proakis, J.G., Manolakis, D.G., *Digital Signal Processing-Principles, Algorithms and Applications*, Prentice-Hall Inc., Upper Saddle River, NJ, 1996.

[30] Porat, B., A *Course in Digital Signal Processing*, John Wiley & Sons, Inc., New York, 1997.

[31] Kurzweil, R., *The Singularity is Near: When Humans Transcend Biology*, Viking Adult., New York, 2005.

[32] Doyle, J., *Extending Mechanics to Minds: The Mechanical Foundations of Psychology and Economics*, Cambridge University Press, New York, 2006.

[33] Guckenheimer, J. and Holmes, P., *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York, 1983.

VITA

Arun Surendran received his Bachelor of Technology degree in aerospace engineering from the Indian Institute of Technology at Mumbai in 2000. He received his Master of Science degree in aerospace engineering from Texas A&M University in August 2002. Mr. Surendran was awarded The Eppright Outstanding International Student Award, the highest honor bestowed on an enrolled international student at Texas A&M University, in 2004.

Mr. Surendran can be reached at Texas A&M University, Department of Aerospace Engineering, College Station, TX 77843-3141. His email is mallusarun@gmail.com.