UNIVERSITY OF CALIFORNIA, SAN DIEGO

Human Blood Glucose Dynamics

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Bioengineering

by

Farbod N. Rahaghi

Committee in charge:

Professor David Gough, Chair
Professor Alberto Hayek
Professor William Hodgkiss
Professor Marcus Intaglietta
Professor Kenneth Kreutz-Delgado

2007

The dissertation of Farbod N. Rahaghi is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2007

DEDICATION

This thesis is dedicated to my late father, Akbar Rahaghi. His commitment to asking tough questions, questioning tough people, and living his life as a scientist and a thinker have been a continual inspiration to me throughout my life. But perhaps more importantly, he taught me how to be a good human being, and that being in pursuit of knowledge is very much a human enterprise and that it comes with many grave responsibilities. May his soul rest in peace, with the knowledge that his ideas and his form persist.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

Generally, it may be looked down upon to thank too many people for their help in creating something because there is a tendency to associate that with an overstatement of the importance of the work. However, in thanking all the people in this section, I hope to instead avoid understating the degree to which I have benefited from their help in creating this document.

I would like to thank professor Gough for his patience and faith in me, giving me ample space for me to make my own mistakes and learn from them, and yet also the support for me to not lose confidence. I would also like to thank him for his vision and making my Ph.D. experience beyond what I could have dreamed would be possible. Most of all, I would like to thank him for treating me like a colleague and nurturing in me the desire to rise above the trivialities that inhibit progress.

I would like thank the members of my committee for their guidance and patience, and for forcing me to think from different perspectives when I thought about the different facets of the projects. I have learned from each and every one of them and was both challenged and encouraged by their advice. I would like to thank the staff at the UCSD bioengineering department who removed many concerns (such as finances) from my conscious mind allowing me to focus on my work.

I would like thank Professor Abarbanel for getting me interested in the whole field of time-series analysis and dynamic analysis. I would like to thank Harry Luithardt, Mathew Kennel, Michael Buhl and David Tarkowski for their help in the journey through nonlinear dynamics and music.

I would like to thank my uncles, Asghar Rahaghi and John Ghaznavi without whom I would not have had much of a childhood, nor the peace of mind to pursue intellectual work as a student for such a long time. Without them I would not have the luxury of spending time doing research and would instead have had to pursue financial security for my family.

I would like to thank my brother who handed me my first science book, gave me my first flask stand, showed me my first science experiment, gave me all his old books and LEGOs and who inspired me to think about inventing and engineering. I would like to thank my mother for teaching me that failure is not an option and for her encouragement to have faith in my pursuits and not to give up.

I would like to thank my good friend Richard Murphy with whom I begun this intellectual journey many years ago and who has always been a source of encouragement and insight. He has always been an inspiration and a source of light in my life. I would like to thank Christine Chandra Bexley for believing in me, standing by me and loving me. Most of all I would like to thank her for lending me her sword.

I would like to thank all the individuals who made their code and knowledge available for the progress of science and knowledge, and who went that extra mile to help further other individual's research. In particular I would like to thank Daniel Kaplan, the OpenTSTOOL and the Tisean folks for their time in making tools available for public usage. Thank you for putting out products and documents that may not have helped you directly but you thought may advance the field. They certainly made my work significantly easier.

I would like to thank Dr. Bruce Buckingham and the DirecNet group for making their data available .I would like to thank my colleagues for their help in teaching me and giving me wisdom. In particular I would like to thank Jared Goor without whom the Ph.D. experience would have been a lonely one, and Yashar Behzadi with whom I had many intellectual discussions and adventures which introduced me to some of the concepts in this thesis. I wish to thank Matthew Strain for his continued infusion of wisdom, long nights of discussions and scientific insight.

I would like to thank the MSTP program at UCSD for taking a chance on me, Dr. Paul Insel for his beyond the call of duty support, Christine Moran for putting up with my growing up process, Natalie Price for dealing with all my requests and Joyce Felder for being a calming force in an otherwise chaotic world.

Last but not least, I would like to thank the people of the state of California and the United States of America for accepting me as one of their own, paying for me to pursuit my interests and believing that their hard-earned dollars would be well invested in me. I will work very hard to repay my debt to them for their trust.

VITA

Studies:

- B.S., Physics, University of California, San Diego 1998.
- M.S., Biomedical Engineering, University of California, San Diego 2005.
- Ph.D., Biomedical Engineering, University of California, San Diego 2007.

Teaching:

Teaching Assistant, Physics Department 1997-1998.

Teaching Assistant, Bioengineering, 2001-2004.

Instructor, UCSD Extension, “Matlab for Scientists and Engineers”, Spring 2004, Spring 2005.

Publications:

New Glucose Sensors (Book Chapter) Encyclopedia of Medical Devices and Instrumentation, 2006.

ABSTRACT OF THE DISSERTATION

Human Blood Glucose Dynamics

by

Farbod N. Rahaghi

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2007

Professor David Gough, Chair

The control of blood glucose concentration has become central to the prevention of morbidity in diabetes. Currently sensors are becoming available to make available near continuous measurements of tissue glucose concentrations. Frequently measured values provide an opportunity to analyze the dynamics of these measurements in addition to statistical analysis. The dynamics can be used to verify sensor validity, to provide a physiologic control target, and serve as a tool to diagnose and monitor disease progression as well as therapeutic interventions. In this document, analysis methods from a diverse set of physical and engineering sciences are applied to blood glucose data that has been published in the literature, and measured in clinical

studies, in humans. The objective is to evaluate the utility of different techniques for time-series analysis, as well as to pave the way towards more data intensive studies to further the applications mentioned above. Data from nondiabetics, type I diabetics, type II diabetics as well as other disease is presented and analyzed using methods which appeared most promising during the author's research. Based on the limited available data, observations are made regarding the characteristics of dynamics in each population groups, and the potential utility of dynamic measurements in diagnosis and assessment of patient metabolic state are demonstrated. In particular, the utility and challenges associated with various methods of time-series analysis as applied to the human blood glucose signal are explored. It is hoped that this can provide a beginning to a very promising future of human glucose time-series analysis and that it will help in sensor and controller design.

Chapter I: Motivation, history and approach

I.A Motivation

Blood glucose values play a central role in the diagnosis and treatment of diabetes. Diabetes is comprised of a constellation of physiologic dysfunctions, which were first diagnosed historically by an increase in urine output and consequently thirst. It was later discovered that these symptoms were caused by uncontrolled increase in blood glucose values. The discovery of insulin led to the understanding that insulin is the primary molecule involved in stimulating cells to uptake glucose. As understanding of the underlying physiology has grown, it has been understood that diabetes can arise from multiple pathways including the loss of beta cells and loss of insulin function on the target cells.

In type I diabetes, where loss of beta cells which are the primary producers of insulin lead to In both cases, blood glucose can become inappropriately high or low depending on the extent of control Currently, patients measure their blood glucose 2-4 times a day and the dosage and timing of insulin administration is estimated based on personal experience as well as blood glucose samples and knowledge of food intake, and daily activities As a result, inappropriate insulin doses leads to life-threatening hypoglycemic episodes.

Recently, an alarming rise in the rate of diagnosis and early onset of Non-Insulin Dependent Diabetes Mellitus, the most common form of diabetes has been reported throughout the world. In some large demographic groups in the U.S, the prevalence of this condition has risen to nearly 25% of that population [1]. Trends towards increasing prevalence of NIDDM have been also noted in Europe [2] as well as in Japan [3, 4] previously thought to have been spared from the epidemic. This increase has been accompanied with a decrease in the age of onset of NIDDM, which was previously thought to be only present in the maturing adult population. Factors such as obesity, sedentary life style and family history have emerged as strong correlates of the phenomenon[1]. In addition to NIDDM, Insulin Dependent Diabetes Mellitus continues to be plague children, remaining one of the most prevalent chronic diseases of children. It is thus postulated that by 2010 the prevalence of diabetes will rise to 50 million individuals in North American and Europe, and 210 million worldwide [3].

Diabetes is all about control. While it has long been hypothesized that control plays an important part in the therapy of diabetes [5], The Diabetes Control and Complications Trial (DCCT), as well counterpart studies such as UKPDS, have demonstrated the central role of control in the prevention and inhibition of diabetic illness [6]. The results have led clinicians to the conclusion that regulation of blood glucose levels is of outmost importance in reducing incidence and progression of complications. Armed with this knowledge, clinicians and scientists have developed,

over the past two decades, the concept of home glucose monitoring systems which allow patients to sample their blood glucose levels, generally a few times a day, and regulate their responses accordingly. Furthermore, long-term measurements of markers of glucose levels (see for example HbA1c) as well as advances in information technology have allowed clinicians to better assess the success of their clinical management.

Until this point most treatment and assessment tools have relied on surrogates of blood glucose such as HbA1c which represents a long-term average of the individual's glycemic state[7]. Frequent blood glucose measurements are rarely made and the standard of care calls for a few daily measurements to properly dose insulin usage and meal portions. Thus the control sought after in diabetes patients is in essence managed by a personal heuristic algorithm based on 3-4 daily glucose measurements. This has been in part due to the lack of availability of blood glucose sensors. In fact, even the availability of home-based multiple daily sampling has only been made possible in the last two decades. The diagram that follows (figure 1-1) shows the historical development of blood glucose sensors.

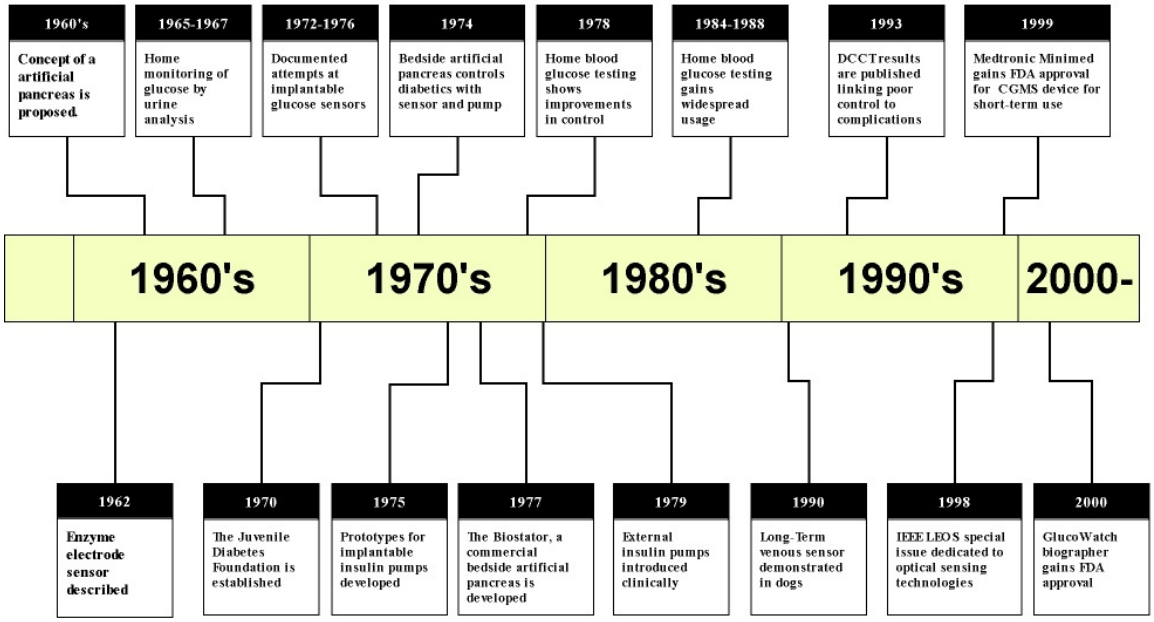


Figure 1-1: A Time-line of some important developments relating to glucose sensors.

As continuous glucose monitors begin to become clinically available, the question arises of what the output can be utilized for in addition to simply being a convenient measurement device. From an engineering perspective, knowledge of the signal content can be used for improving sensor behavior and design of controllers which can be used to inject insulin into the patient automatically. From a clinical standpoint, knowledge of the signal content can lead to screening tools, quantifiable methods of assessing the relative success of treatment and quantitative assessment of progression of disease over time. Additionally, the glucose signal may provide insights into the underlying physiologic processes governing human metabolism.

In anticipation of the increasing availability of frequently sampled glucose data, this dissertation takes a look at the range of dynamics observed in the blood glucose signal, explores possible ways of quantification of these behaviors and

potential uses of them. The methods are drawn from many disciplines of study which have faces a similar challenge, which is to extract as much information as possible from the signal available to them.

I.B Defining the term “dynamics”

The word dynamics is used in different fields with slightly altered meanings. Strictly speaking, the field of dynamics arises from physics where it is the study of the change in the state of the system over time. To properly study dynamics in the physics context, one must understand the state space (that is the possible states that can be assumed by the system) and the equations of motion (which describe the pathways and transitions between these states). The difficulty arises when certain variables necessary to construct this state space are not measurable or perhaps not even known. For example, the exact physiology of glucose metabolism depends on a myriad of known and unknown hormones, most which are unlikely to be continuously measured in the near future. This thesis focuses on the practical and clinical extreme in which the blood glucose time-series measurement, and possible a few other pieces of information (like meal times) are the only known data with which to study the system. At this extreme, the exact or perhaps even partial reconstruction of the state-space becomes improbable.

Limiting ourselves to the knowledge contained in one single variable being measured, in this case glucose, the term dynamics refers to the way in which change

occurs in the variable being measured, in this case blood glucose. More specifically, it refers to the specific patterns that are encountered in the evolving signal being measured. It differs vastly from a statistical analysis, in that unlike a statistical interpretation in which variability is looked at with each sample being independent, dynamics require that sufficient measurements be made such that the all intermediate states the signal can assume between samples are captured. In practice, it requires that the sampling be sufficiently fast such that states do not sufficiently evolve between samples such that significant information about their transition is missing. This fundamental requirement arises in many forms when considering how to sample the variable in question, and the criterion depend on some level on the definition of the states, the degree of accuracy needed and the speed with which the state of the signal evolves.

An additional requirement is that the variable is measured for long enough of a duration, and under variable conditions such that the system travels through all possible states. This requirement, which is more abstract, is much harder to implement in practice because checking for its fulfillment requires a priori knowledge of the system's state space. In the simplest case, an impulse input into the system produces a response that is sufficient to characterize the system. In practice, however, the presence of nonlinearities and multivariate nature of physiologic systems prevents such a simple tests from being implemented. This leads to the heuristic algorithm often used in physiologic analysis that the system should be observed under as many

extreme conditions as possible in order to minimally delimit the boundaries of its behavior.

This analysis attempts to gather together as many well sampled data sets under differing input conditions to reach a complete picture as possible, but with the understanding that this may fall short of the system's true possible state. This however, points the way for further studies that may be required to reach a more complete picture.

I.C The case for dynamics

Blood glucose regulatory system is a dynamic system with multiple major organ systems involved in its control. These include the pancreatic islet cells (which produce insulin, glucagons and somatostatin), the liver (which stores and outputs glucose), the digestive system (which perturbs the system by adding glucose to the blood stream) and a host of neuro-endocrine modulators. The rise and fall of blood glucose levels is deeply intertwined with other hormonal oscillations such as growth hormone and cortisol [8, 9]. Regardless of the evolution of the approaches to the treatment of diabetes, (e.g.. beta cell therapy, pharmaceuticals, mechanical controllers) it is desirable to restore complete normal control as treatment for diabetes. Markers such as HbA1c, though useful in assessing the mean and variance of glucose levels, do not include information about whether the system is behaving normally in a dynamic

sense. Thus by understanding the dynamics one can better assess degree of control and design new control approaches that better simulate the nondiabetic state.

The case for dynamics can thus be summarized as follows:

- To design and assess treatment methods (such as controllers or artificial beta cells) in such a way as to better resemble the nondiabetic state.
- To develop clinical applications for emerging sensors such as screening diagnosis, finer assessment of degree of control, and patient education.
- To better understand the underlying physiology of the system.

I.D Previous work

I.D.1 Underlying physiology

Human blood glucose is regulated by the action of insulin as primary reducer of blood glucose. On the opposing side, meals, basal glucose production and the action of glucagons and other neuro-endocrine factors lead to the rise in the glucose level. The human insulin response consists of a basal as well as reactive (in response to meals) component. In response to a rise in the levels of glucose in the blood supply being exposed, the β-cell secretes insulin in a multiphasic manner, that is in multiple bursts following the initial event[10]. The net effect is the large-scale release of insulin that is directed to the portal vein and thus the liver. The liver is the first organ to observe the insulin response, and as much as 60% of the release insulin is extracted by the liver before it enters peripheral circulation. The liver, which is the primary site of

glucose storage is responsible for a basal production of glucose and its input to the blood stream, and increases such production through a complex set of modulators. Insulin is enters the blood stream where it increases the uptake of glucose by target cells such as muscle cells and adipose tissue. Production of insulin itself is also regulated by factors such as glucagon, which is released by the pancreatic islet alpha cells. Complex interrelationships are described with oscillatory patterns of cortisol and growth hormone as well [9].

The cellular effect of insulin on most peripheral tissues includes the promotion of the uptake of glucose via pathways that include the increased translocation of glucose transporters to the cell surface [11]. Thus insulin essentially acts as a cellular "feeding" signal, promoting the uptake and storage of energy. Furthermore, insulin increases the rate of glycolysis in muscle tissue.

I.D.2 Sampling rate and prediction

Previous work by our group has used some limited data sets to show, using the autocorrelation function that the blood glucose signal contains determinism, and that this determinism can be used for the purposes of prediction. To this end, an Auto-Regressive Moving Average model was constructed and used to predict blood glucose values into the future [12]. Markov models and Hidden Markov Models were also applied to the system with similar success in prediction of future blood glucose values [13]. Additionally, using spectral estimates of the frequency content of the data sets, it

was concluded that the Nyquist sampling period for the data available was on the order of 10-15 minutes [14].

I.D.3 Mathematical modeling of system

Because this analysis takes a time-series based view of the analysis of the data, an extensive review of the mathematical modeling of the physiologic system is not presented, though they were extensively reviewed by the author. An increasing number of mathematical models have been devised on the years to account for blood glucose dynamics. The models vary in their mathematical structure, the number of physiologic variables considered and the specific approach to determination of parameters. As an example, attempts have been made to measure beta cell functionality by continuous infusion of glucose and assessment using simple models leading to an estimate of beta cell function [15] Simple indices of insulin resistance have been proposed and are assessed in a review paper on the subject [16]. Four early mathematical models were assessed in one study using 182 patients [17] which serves as a good early introduction to the subjects.

Of all the mathematical models, perhaps the one most often utilized is the Bergman model which deserves mention. The family of models developed by this group has served as a tool for better understanding the response of insulin to glucose input. Using various perturbations and “clamps” (that is an artificial system that keeps a variable at a steady state value), iterations of the “minimal models” have been developed to account for glucose and insulin dynamics. They are termed minimal

models in part because of attempt to create few observable values as the main variables. This is in contrast to multi-compartment models developed which contains a series of differential equations with many physiologic variables that are clinically difficult to assess. The reader is referred to this model as a good entrée point into the subjection of mathematical modeling of the glucose regulation system [18-21]. For a early review on the general approaches to modeling, the reader is referred to an early review on the subject[22].

I.D.4 Responses to perturbations

A long history of using meal as a perturbation to the glucose regulatory system exists within the field of endocrinology. The most developed of these is the glucose tolerance test (GTT), which has been, used heavily in the diagnosis of type II diabetes. In this instant, a fixed carbohydrate bolus is administered, generally after a fasting period [23] and the response of blood glucose is recorded. Though a tremendous body of literature exists on the subject, the quantity that has been traditionally used in the clinical setting has been the extent to which glucose values have returned to normal after a set period of time (frequently 2 hours). A tremendous body of literature exists on this subject and it has been thoroughly studied in various populations [24, 25].

Because of the limitations in sensor development, long-term continuous monitoring of blood glucose levels has been very infrequent in comparison of other types of analysis performed on the diabetic population. However, some groups have

taken such data with small population groups for the purposes of studying various types of perturbations to the state of diabetics and nondiabetics. Multiple attempts have been made at measuring the variability in the time-series derived from diabetics that have led to the two measures, mean amplitude of glycemic excursions (MAGE) and the “M” coefficient of Schlichtkrull which is a measure of insulin efficiency [26-30]

Glucose dynamics have also been studied in the context of continuous enteral nutrition (nutrition delivered directly through a feeding tube at a constant rat), as well as through an intravenous route. In both cases oscillatory patterns of the scale ~1-2 hours have been noted[8, 31]. Similarly, when given a glucose infusion with a sinusoidal rate of glucose delivery, the blood glucose levels display a phenomenon of entrainment, that is the system begins to oscillate at the same frequency as the input [32] . Both these phenomenon have been noted to be disturbed in patients with impaired glucose tolerance or type II diabetes.

I.D.5 Summary of current knowledge

(1) The glucose signal contains deterministic information that can be used for prediction of future values. Specifically, both ARMA and Hidden Markov Models can be utilized for this purpose.

(2) The Nyquist sampling requirement for blood glucose is closer to 10-15 minutes than the currently clinical practice of measuring blood glucose values a few times a day.

(3) Metrics of dynamics such as MAGE and M values can be used to differentiate between patient groups with stable and unstable control.

(4) Constant infusion of glucose leads to oscillations in blood glucose which are less robust in Type II and IGT patients.

(5) Dynamic infusion of glucose leads to entrainment on multiple time-scales, and on those scales the entrainment is less robust in type II and IGT patients.

I.E Thesis approach and objectives

The purpose of this thesis is to extract information from the available collected data in such a way as to direct future research on the subject. The paucity of data along with difficulty in collection of new data that is currently being overcome make it difficult to make general population based statements on dynamics of blood glucose based on time-series data. But, the diversity of available data does permit an assessment of possible approaches to characterization of the dynamics and points the way to large population based studies that can support the observations based on limited data.

The analysis approach is to look at the signal being produced at face value and without many assumptions about the underlying system. This approach is chosen in

part because it has the potential to produce useful engineering and clinical approaches that are based on the most likely signal available to clinicians and engineers in the foreseeable future and thus likely to find utility in the near future. Additionally, the condensation and extraction of salient features of blood glucose dynamics is likely to find utility in verification and design of mathematical models.

The thesis objectives are to apply methods of engineering to a diverse group of data and to assess the utility of such methods as applied to this type of data. While none of the data sets are from a large enough of a population to make population based claims, this thesis paves the way for methods involving larger subject groups. The methods are organized in order to answer the following three fundamental questions:

- What is normal blood glucose dynamics?
- Are there different classes of diabetic dynamics, and how do they differ from normal?
- What maximal sampling interval that is acceptable for capturing dynamic information useful to answering the above questions? How does sampling duration effect various measures of dynamics?

Chapter II: Data

As mentioned in the introductory chapter, the number of data sets and frequency of sampling has generally been less than ideal in the datasets available in the literature. Nonetheless, a thorough search of the literature was undertaken and a number of data sets were found. The most impressive aspect of the data is the diversity of patient types and perturbations under which they were collected, but each individual type of data is often involves only a few subjects, limiting statistical, population based conclusions. Unlike the common scientific perception that glucose levels are fairly constant, a very rich variety of behaviors were found under differing perturbations to the system. An example of four data sets are shown in figure 2-1

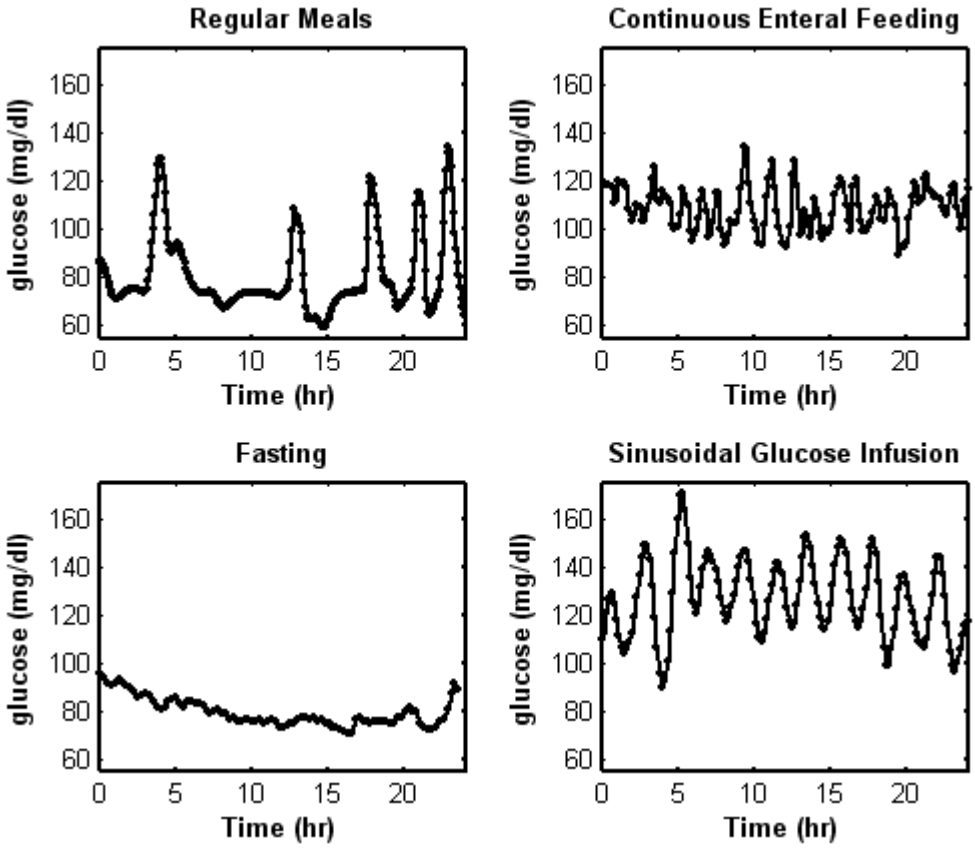


Figure 2-1: Blood Glucose Time-Series from four nondiabetic subjects. Literature data from representing responses to different glucose challenges. Sampling was at regular intervals ranging from 5 to 15 minutes over 24-hour periods. Data points are based on reverse digitization, and are connected by straight lines. Note that the response to fasting contains a rapid, small amplitude signal with a slowly varying baseline. The sinusoidal challenge had a period of 128 minutes, an amplitude of 1/3 of the mean, and an average infusion rate of 6 mg/kg per min.

II.A Sources of data

The literature was searched using pubmed (a service of the US National Library of Medicine) as an initial search path, and follow-up of the references included in the papers acquired from the search. The initial searches were based on keywords such as ‘blood glucose’ and authors known in the field to have studied this

subject. Based on references acquired from the resulting papers, new keywords and authors were then searched. After reviews of abstracts, approximately two hundred papers were evaluated for utility as a data sources. Most were rejected because they did not list data, did not sample frequently, did not use verified laboratory measurements, used experimental sensors, or displayed only average results between individuals.

Over 30 references spanning 1963-1999 met the criteria for utility for our studies. In a few cases, a study showed only a single time-series. In such cases the data was digitized for possible use, although when larger datasets were available for the question at hand, the inclusion of a single time-series from an experiment had to be weighed against the difficulty with interpretation that comes from such meta-analysis.

II.B Methods of data acquisition and conversion

When possible, original electronic copies of the articles were acquired in the .pdf format. In all other cases, however, the location of the reference material required that they be photocopied and then subsequently scanned to create an image from which the data would be reconstructed.

The scanning was performed by a brother MFC 9600 Copier and Scanner. The images were stored in a lossless .jpg format for further processing. Realizing that despite all attempts to make photocopies, there was a likelihood that the images would

be rotated and warped, a software algorithm was written and tested for the rotation of the images, and is described in more detail in Appendix A.

The reconstruction of the data from the graph had to be tailored to the specific graph type available. The graphs could be divided into three categories: graphs consisting of straight lines, graphs containing individual data point symbols and graphs consisting of disconnected symbols not representing original data-points but in lieu of a solid line. Most graphs were of the first two kinds discussed above.

Because of the importance of the data to the conclusions made in this document, multiple measures were taken to insure data quality. They are listed as follows and are further discussed in appendix A.

1) Verification with respect to the image: The acquired data set was plotted on the original image from which it was acquired (including the possible rotations) and visually verified to correspond to the lines in the images.
2) Verification with respect to the experimental results: In some cases statistical measures were also reported along with the data. In such cases, statistics for the acquired data was computed and checked against the reported values. A summary of these results can be found in the appendix A. No cases were found where there was a significant difference between the acquired data and the original reported statistics.

3) Repetition verification: In some selected cases, the paper was rescanned, the image was rotated, digitized and data extracted and compared. Agreement between the two data sets was noted.

II.C Summary of literature datasets

The data sets can be placed in one of the following six categories, in terms of the type of study from which they were collected:

- Data with multiple meals. This is the typical data set encountered and desirable for analysis because it most resembles the daily conditions of the patient. In some cases, exercise is included in the patient's daily regimen while in others, the patient is kept in a hospital bed. For diabetics, this data also includes their course of treatment which in almost all the data sets is a set of insulin injections.
- Fasting Data. This is data in which the subject is consumes a meal prior to the study. After that meal, a set period of time elapses and the investigators then begin measuring blood glucose levels.
- Continuous Enteral Feeding. In these studies nutrition is given through the gastrointestinal tract but using a continuous infusion which flows prior to being absorbed. This is the basis for a constant input into the system, which goes through the same pathway as food normally does.

- Continuous IV Feeding. In these studies, nutrition is given through an intravenous route, at a constant level. This is similarly a constant input into the system but does not go through the pathway as normal feeding.
- Dynamic IV Feeding. In these studies, nutrition is delivered intravenously but is modulated so that the input into the system is a sinusoid with a set period.
- Continuous insulin input. In these studies insulin is intravenously delivered at a constant rate while nutrition is delivered simultaneously.

Another way to think of the data is the state of the patient. Here the following groups are used to characterize subjects:

- Nondiabetics. These are individuals who have not been diagnosed with diabetes and in the case of those with risk factors such as age and obesity, have been tested using the oral glucose tolerance test (OGTT) and have not met the criterion for impaired glucose tolerance (IGT) nor Type II diabetes.
- Type I diabetics. These are individuals who have been diagnosed with type I diabetes characterized with childhood onset accompanied by complete or near complete loss of insulin production. In one specific data set, they are also further subcategorized as stable and unstable, with the stable individuals being easier to “manage” than the unstable ones. Difficulty of management is defined by the difficulty of achieving good diabetic regulation.
- Individuals with impaired glucose tolerance (IGT). These are individuals who in response to an oral glucose tolerance test (OGTT) display abnormalities in

the form of elevated blood glucose for a longer period after the test than what is normal.

- Individuals with Type II diabetes. These are individuals who either through an OGTT test and/or having developed diabetes at a later age and with the common constellation of symptoms (obesity) have been diagnosed with insulin resistance and type II diabetes. In most cases they are treated without insulin but in some cases insulin injections are included in the time-series period.
- Individuals with a GCK deficiency. These individuals have a defect in the enzyme GCK which leads to essentially similar symptoms as type II diabetes but through a different pathway.
- Individuals with insulin secreting tumors. These individuals experience abnormally low blood glucose because of abnormally high blood insulin levels. They are treated with frequent meals meant to counter the insulin levels, and ultimately with the removal of the insulin secreting tumor.

A summary of these data sets are presented in table 2-1. Further descriptions of selected experiments can be found in Appendix B.

Table 2-1: A summary of the data sets used in this document.

# Of Data Sets	Names	Duration/Sampling	Patients/Condition	Refs
Nondiabetics				
3	Sernorm1-3	48 Hrs/5 Minutes	Nondiabetics, feeding, exercising	[27]
4	Vannorm6-9	24 Hrs/20 Minutes	2 Nondiabetics with different Meal Spacings	[8]
7	Mal1-7	24 Hrs/60 Minutes	Nondiabetics W/ Meals	[33]
5	Mal8-12	13 Hrs/30 Minutes	Nondiabetics W/ Meals	[33]
5	Mej1-5	24 Hrs/30 Minutes	Nondiabetics W/ Meals	[34]
3	Shanorm1-3	24 Hrs/15 Minutes	Nondiabetics Fasting	[8]
9	Simnorm1-9	24 Hrs/10 Minutes (3-6 3pt moving av)	Nondiabetics W/ Continuous Enteral Feeding	[31, 35, 36]
6	Polnorm1-3,9,10,12	53 Hrs/ 20 Minutes, 9-12 detrended	Continuous IV Infusion	[37, 38]
2	Polnorm6-7	20 Hrs/15 Minutes	Continuous IV Infusion	[8]
5	Vannorm1-5	24 Hrs/15 Minutes	Continuous IV Infusion (Different Rates)	[8]
1	Poldynnorm1	20 Hrs/15 Minutes	Sinusoidal Infusion	[39]
9	Poldynnorm2-10	24 Hrs/10 Minutes	Continuous Or Sinusoidal IV Infusion	[40, 41]
4	Mealcha1-4	9 Hrs/4 Minutes	Single Meal After Fasting	[23]
4	Mealcha5-8	14 Hrs/4 Minutes	Single Meal After Fasting	[23]
4	Polinsul1-4	24 Hrs/ 15 Minutes	Continuous or sinusoidal insulin infusion	[32]
1	Byrn1	12 Hrs/ 15 Minutes	Sinusoidal Infusion	[42]
1	Pfenorm1	24 Hrs/ 1 Minutes	Fed/IV Combo	[43]
1	polentrain1	12 Hrs/ 10 Minutes	Sinusoidal Infusion	[8]
4	Polfat1-4	12 Hrs/ 15 Minutes	Meals	[44]
1	Simvf1	8 Hrs/ 2 Minutes	Single Meal	[31]
1	Polvf1	8 Hrs/ 2 Minutes	Complete Fasting	[45]

Table 2-1: A summary of the data sets used in this document (continued).

Type I Diabetic Subjects				
15	Serone1-15	48 Hrs/ 5 Minutes	Meals/Exercise/Insulin	[27]
4	Mir1-4	24 Hrs/ 1 Minute	Meals/Insulin	[26]
3	Mir 5-7	24 Hrs/ 5 Minute	Meals/Insulin	[26]
2	Mir 8-9	22-42 Hrs/ 5 Minutes	Meals/Insulin	[26]
8	Sanpregone1-8	12-24 Hrs/ 1 Minute	Pregnant, Some W/ Biostator	[46]
Impaired Glucose Tolerance and Type II Diabetics				
8	Polfat5-12	12 Hrs/ 15 Minutes	Meals	[44]
5	Polentrain2-6	12 Hrs/ 10 Minutes	Sinusoidal Infusion	[8]
1	Pfeniddm1	24 Hrs/ 1 Hour	Meals	[43]
2	Simniddm1-2	12 Hrs/ 10 Minutes	Constant Enteral Fed	[47]
3	Poldynigt1-3	24 Hrs/ 10 Minutes	Dynamic IV Infusion	[40]
3	Poldynniddm 1-3	24 Hrs/ 10 Minutes	Dynamic IV Infusion	[40]
9	Shaniddm1-9	24 Hrs/ 15 Minutes	Fasting	[8]
10	Sanpregtwo1-10	16-24 Hrs/ 1 Minute	Pregnant, Some W/ Biostator	[46]
Insulinoma				
6	Vil1-6	28 Hrs/ 15 Minutes	Meals (8 per)	[48]
GCK Deficiency				
1	Byrn2	12 Hrs/ 15 Minutes	Sinusoidal Infusion	[42]

II.D Clinical datasets

A unique, large data set that consisted of quasi-frequently sampled blood glucose was also obtained from the DirecNet study group. DirecNet (Diabetes Research in Children Network) was established in part to assess the accuracy of systems intended for continuous glucose monitoring in children. This was accomplished by using these devices (which at the time were not FDA approved for

insulin dosing but for assessment purposes only) in children while taking quasi-frequent glucose measurements with accelerated sampling during times in which change was expected (during meals, insulin injections, etc..). The gold standard data (blood samples) were acquired from the group after gaining their permission and after all patient related data that would identify the patients was removed. The studies themselves are described in detail in [24, 49] and were focused on the sensors but produced this useful data as a byproduct.

The difficulty with the DirecNet data set is that the sampling is not uniform and well above the limit set forth by assessing the Nyquist criterion previously described in the literature [14]. An example of the quality of data obtained from this dataset is shown in figure 2-2. Nonetheless, this data set represents a large number of patients and thus the data set was used to the extent possible by the constraints stated above. This was limited to the studies of rates of change. In particular, because of the recent nature of the data and the age of the other type I data being used, the rates of change obtained from this data set serve to confirm that observations on type I diabetics are not simply the result of outdated treatment approaches.

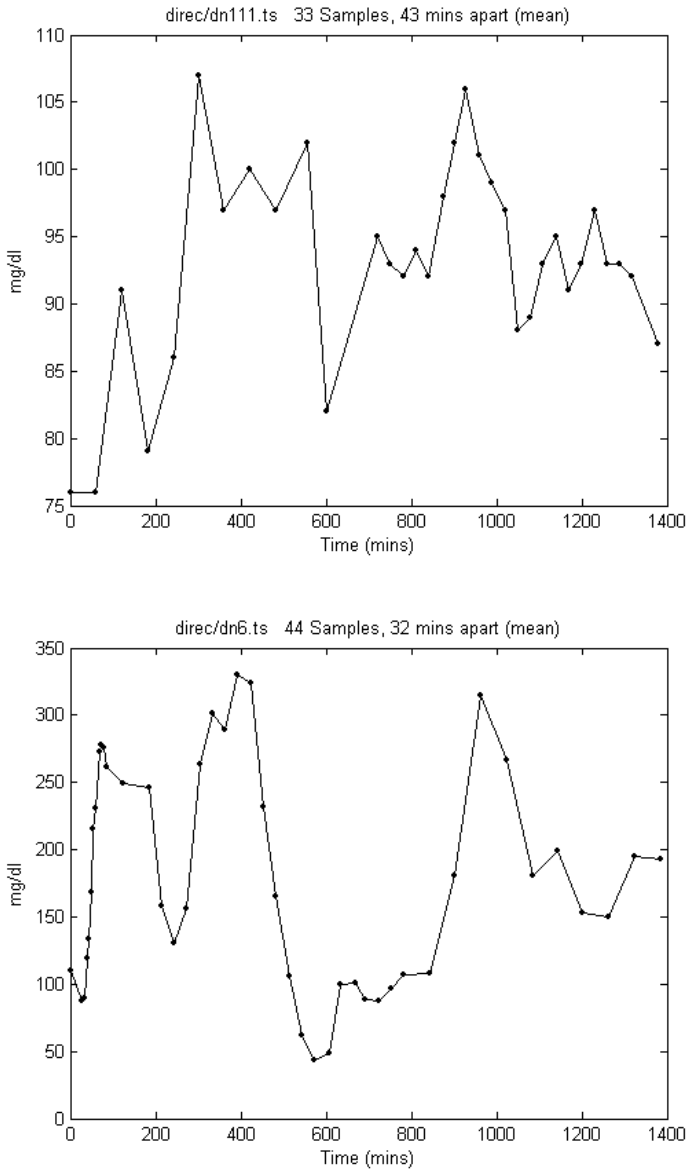


Figure 2-2: Two time-series from the DirecNet data set. Top: nondiabetic child. Bottom: Type I diabetic child. Note the difference in scales.

Chapter III: Methods of analysis

Dynamical system information in the form of time-series or spatial series (in the case of images) has been analyzed for over a century and a very diverse toolbox of such methods has been developed. A particular characteristic of the approaches to time series analysis is the high-level development of specific approaches by individual scientific and engineering communities. Thus when presented with a novel signal, it behooves the analyzer to draw from experiences with various fields of study. The methodologies presented have been developed in the fields of signal processing and have been applied to a diverse set of naturally occurring signals with the complexities present in the glucose time-series signal. In this section, they will be briefly reviewed as they form the basis of the examination of the data.

III.A Types of methods considered

A very wide net was cast in the process of searching for ways to study the blood glucose time-series. Over the past few decades there has been a significant proliferation of approaches to time-series analysis and modeling, in particular those taking advantage of the processing powers of computers. The approaches may be loosely divided into three categories, in part based on the scientific communities using such approaches:

1) Signal Decomposition

2) Model Independent Signal Modeling

3) Model Dependent Signal Modeling

In signal decomposition, and attempt is made break the time-series down into components, in hopes that the components can be used to more succinctly describe the time-series, or that specific components can be then utilized in other analysis. These methods include Fourier analysis, parametric spectral estimations, time-frequency analysis (short time Fourier transform Wavelets, Wigner-Ville Distribution), and component analysis techniques such as empirical mode decomposition, principal component analysis, independent component analysis and singular value decomposition. The resulting representation can be directly used to reach conclusions (such as by looking at the presence or absence of a component), or can be used itself as a components of a model using methods below.

In model independent signal modeling, a model is constructed using the signal without making assumptions that come from other sources than the signal itself. Another words, if assumptions about the signal structure are made, they are derived from the signal, and not based on some physical insight into the functioning of the underlying system. Examples of such approaches include various flavors of auto-regressive moving average models, Markov models, curve-fit based models, Statistical classification methods (neural networks, support vector machines), Measures of determinism (autocorrelation function, mutual information, entropy), nonlinear state-

space representations (correlation dimensions, phase space maps, Lyapunov exponents) as well as other single parameters that have been proposed over the years to reflect signal complexity (approximate entropy, sample entropy) The output of such models is generally a set of coefficients, an illustrative map or a single parameter which can then be used for the purposes of diagnosis, monitoring and design.

In model dependent signal modeling, other information in addition to the signal itself are utilized to construct the model: an underlying governing set of equations, known physical mechanisms of certain components of the signal, the timing and extent of perturbations to the system generating the signal, and known physical limitations and boundaries to the value the data can assume. The general class of such models encompass most of physical sciences and are too numerous to mention. Specific to the field of glucose monitoring, are single and multi-compartmental differential equation models of the human endocrine system, single compartmental differential equation models, differential equation models of blood glucose response to insulin and meals, parametric models of response to outside perturbations (requiring additional information in addition to the time-series). These models make many general assumptions about underlying physiology and almost universally require measurements of other substances such as insulin. Unlike glucose sensors, insulin sensors are not yet near clinical feasibility. For these reason, this type of modeling was not pursued in the studies in this dissertation.

III.B Constraints on applicable methods

To limit the methods applied constraints of the thesis objectives and the properties of the data are considered. The following limitations, in particular, played a large role in the choice methods utilized.

- Limited Length Of Tracings W/ Respect Time-Scales of Dynamics
- Limited number of individuals in each subgroup
- Multiple physiologic systems interacting through a complex mechanism

For example, in the case of higher order statistics, sufficient samples are required to populate the distribution and thus measures such as skewness and kurtosis require larger number of samples for sufficiently accurate calculation. Measures using higher order cumulants require much more population in the distribution to yield stable results.

Time-frequency analysis looks at the shift in the frequency characteristics of the signal as a function of time itself. The various methods of time-frequency analysis, discussed below provide varying degrees of tradeoffs between certainty in frequency versus time [50], as well as differing degrees of performance under specific data lengths and signal makeup. While discussions in the time-frequency community continue as to the optimal approach to time-frequency analysis that is generally applicable, each sub-community (defined generally by the type of signal of interest)

has adopted a subset of these for use, generally because of constraints imposed on the data.

III.C Signal rate of change

The most basic expression of a signal's dynamics that is not conveyed by its statistics is the rates of change. The study of the signal's rate of change, and how it evolves as a function of time is thus the one of the first approaches to time-series analysis. The simplest expression of this is the estimated signal derivative. In this analysis, a definition of derivative is used of the following form:

$$g'(k) = \frac{g(k+1) - g(k-1)}{2 * f_s}$$

Where f_s denotes the sampling frequency. This is implemented as is except at the end-points where a one-sided version of the above is computed.

More complex estimates of derivatives using higher order splines are possible but may introduce artifacts and cannot be applied with the same generality to signals with varying sampling intervals. The result of the above simply generates another time-series for the purpose of analysis which in itself may not seem like a great simplification. However, the statistics of this new signal can be studied in addition to the statistics of the original time series. Thus this represents the simplest way of analyzing the dynamics in the system using the well described tools of statistics.

Because hypoglycemic detection and controller design are amongst the top priorities guiding this analysis, maximum rates of change and in particular fall of glucose are of great interest.

The dynamics of a system can be studied in what is known as state space. A true state space reconstruction, however, requires an understanding of the underlying variables involved in the dynamics. Despite this limitation, there have been successful approaches at visualizing the evolution of the system having only the time series available. These generally involve the projection of the information from one dimension (the time series) to multiple dimensions. An example of such analysis is time-delay embedding (which will be discussed in the nonlinear time-series analysis section) and partial phase space reconstruction. The latter, when applied without the knowledge of the equations of motion, is generally an estimate of what the key variables *might* be. For example, in many mechanical systems position and momentum are used to describe the evolution of the state of the system. In simple cases, it can be shown that this represents a complete description of the phase space. In the case of a unknown system, a reasonable starting point may be the signal plotted against its rate of change. An example of this is shown displayed in figure 3-1.

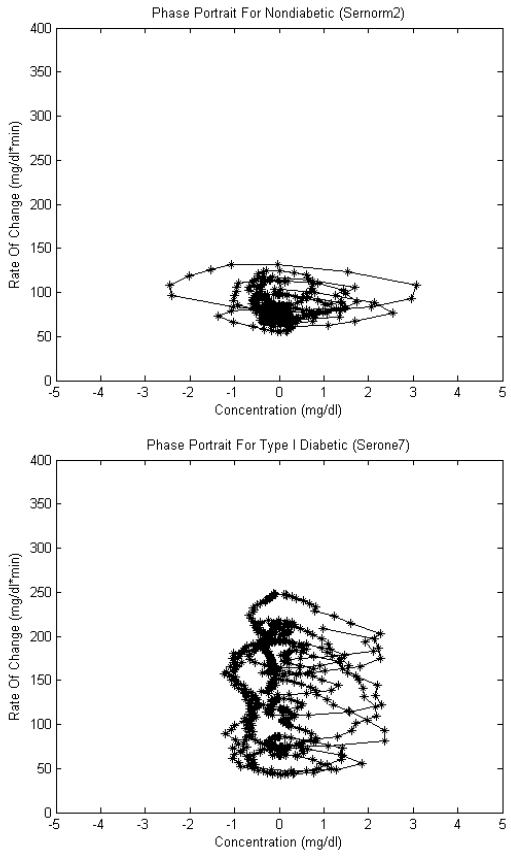


Figure 3-1: Two phase portraits constructed by plotting the value of blood glucose against the rate of change. The first (top) is for a nondiabetic patient whereas the second (bottom) is for a type I diabetic patient with less than desired control.

In many cases, this type of plot leads to insights about the evolution of the system. An example is shown below for a system implemented in our laboratory based on the chaotic Chua's circuit. Measurements from two of the electrical components can be plotted against each other to reveal an intricate geometry with two regions around which the system "circulates". These regions are called attractors because the state of the system seems to be "attracted" to them, and circulates around them. The transition between different behaviors in the system can be studied in terms of the number and exact geometry of these attractors [51]. This type of analysis is

particularly utilized in the case of nonlinear systems exhibiting chaos, which is a property where the exact evolution of the system seems to be very sensitive to initial conditions. An example of a plot from such a system designed from a circuit in our laboratory is shown in figure 3-2. This system, a Chua's circuit, exhibits highly nonlinear dynamics and chaos in certain parameter regimes. Data from this system was used in addition to simulated data for the purpose of testing methods.

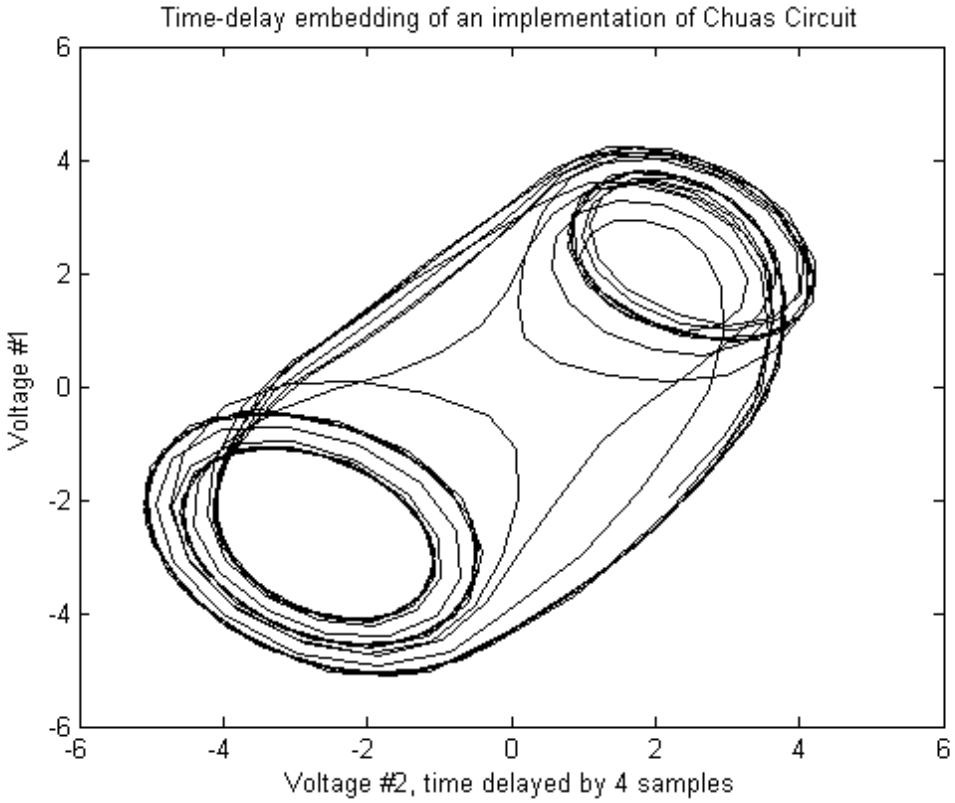


Figure 3-2: An attractor geometry for time-series obtained from a Chua's circuit, a circuit exhibiting highly nonlinear behavior and chaos.

In the absence of clear attractor geometry, statistical methods can be used to analyze the geometry of this reconstruction. In this analysis a number of methods analyzing the density, shape and symmetry of these attractors as a way of

characterizing glucose dynamics are proposed. Difficulties arise due to digitization which causes samples to lie in the exact same point in this phase space. To make the analysis more systematic, a vector quantization approach is applied where samples are simply placed in a two dimensional histogram based on the value of the time-series and the derivative estimate. This is shown in figure 3-3.

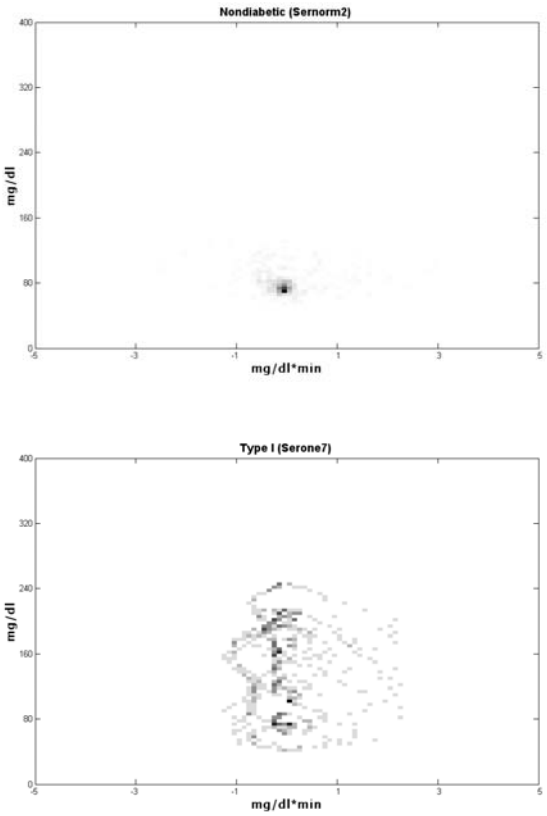


Figure 3-3: Phase portraits from two different types of patients. A nondiabetic is shown on top and a diabetic is shown on bottom. Darkness corresponds to more samples falling in that square.

This quantized phase-space estimate can then be analyzed. The first proposed step is to characterize the boundaries of the attractor. This is done by using a simple image processing algorithm whereby the image of the attractor is rotated and the first non-zero value is used as a "boundary". The boundaries are then used to create a solid object, which represents the space of occupied by the system (figure 3-4). Characteristics of this object can then be used to define the system's behavior. Of particular significance may be the size and asymmetry of the attractor.

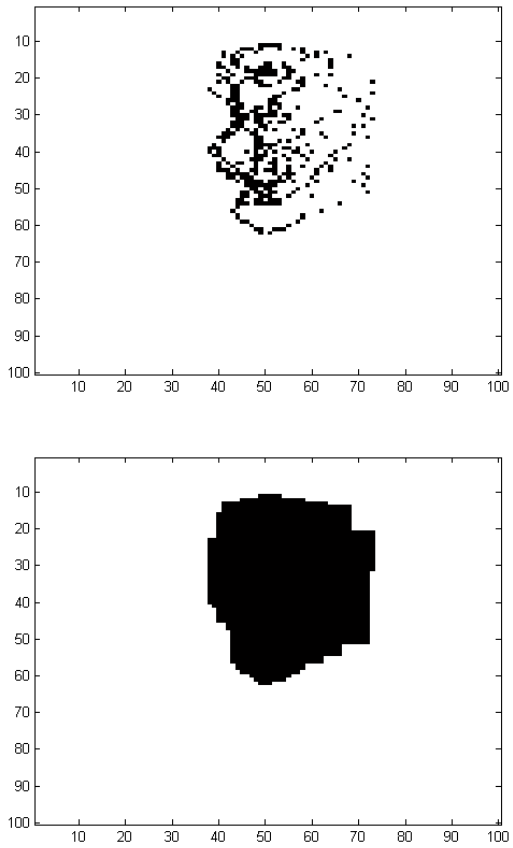


Figure 3-4: The phase portrait and the extracted geometrical shape of the phase portrait, the boundary of which is used to determine the size of the attractor.

On further analysis, however, it is clear that the same geometry can be distributed in different ways. For example, many of the samples may be concentrated at the center, or at the boundaries leading to the same geometric description. To make the analysis more thorough, additional considerations of the distribution must be considered. One simple consideration is density of the attractor object, or put another way its compactness. One simple approach is to form concentric rings of sample concentration in the attractor, but this requires an ellipsoid geometry with only one focal center where the values are concentrated which while present in nondiabetic

individuals, is not necessarily present in other time-series analyzed. To overcome this limitation a more general method is considered. The number of samples in the densest bin is added to the second densest bin and so on. In a very dense attractor, very few bins account for most of the samples, and thus the curve is shifted to the left. To account for attractor size, the curve is normalized for the number of bins in the attractor. The number of bins required to account for 90% of the samples (arbitrarily chosen boundary for the purpose of comparison) divided by the total area of the attractor will be referred to as the attractor density.

Another approach is to look at the center of the attractor object, and view all other samples from its perspective. This is achieved by using another image processing routine whereby the lines are drawn radiating away from the center of the attractor, and the distribution is sampled along this radial lines. This leads to a conversion from a circular geometry to one in which the attractor can be view from the perspective of excursions from the center. This can be then used to study the profile of such excursions. The degree of difference between what the radiating lines sample can be used to study the radial symmetry of the attractor. An example of such radiating sampling lines and the resulting portrait is shown in figure 3-5.

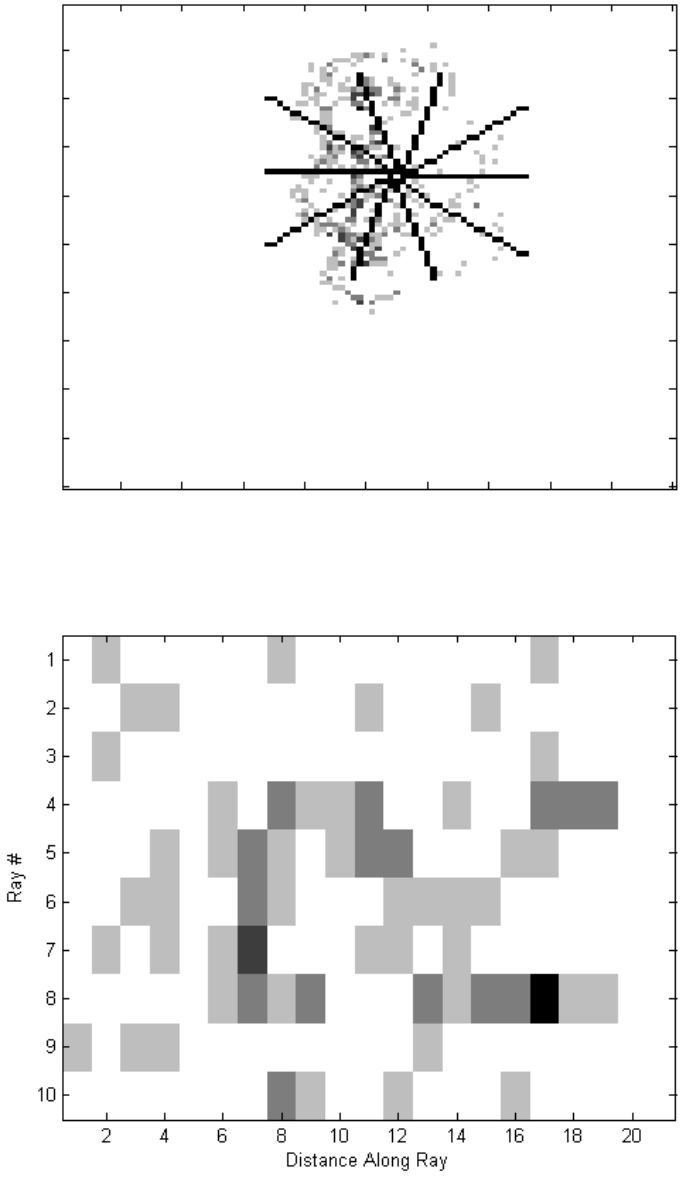


Figure 3-5: (Top) A set of radiating lines from the center of the attractor (center is chosen by locating the center of the geometric object containing the whole attractor) is used to sampled along various directions leading to the sample profile along those lines in different directions. The resulting sampling is shown in the bottom plot.

The profile generated from this radial sampling can then be studied by comparing the samples along each line. In a perfectly symmetric case, the number of samples would be the same along each line. This is measured by the finding the scaled range, defined by

$$\frac{\text{Maximum}(\#\text{Samples}) - \text{Minimum}(\#\text{Samples})}{\text{Maximum}(\#\text{Samples})}$$

This quantity can then be evaluated for each step away from the center of the attractor. Once this quantity is computed for each distance, they are averaged to yield a number which represents a rough estimate of the symmetry of the object. If the object is perfectly symmetric, then the value is zero and if it perfectly asymmetric, then the value will tend to one. The main weakness of this approach is the difficulty in defining a center and the computational intensity involved, which increases very rapidly as the radius of the attractor grows.

In conclusion, the attractor geometry will be assessed by three measures: area, density and symmetry in hopes that these will help differentiate patients and subgroups.

III.D Time-scales of dynamics

Frequency spectrum attempts to capture the properties of signal by decomposing the signal in terms of a sum of sinusoids of various frequencies. This gives a direct interpretation in frequency domain in terms of time-scales of dynamics,

allows for the detection and removal of noise, and can be correlated to the time-scale of other physical phenomenon which can then be tied to the analysis. In the most theoretical framework, these are composed of sinusoids of infinite duration and infinite number of such sinusoids can be used to construct a signal. In the digital world, signals are composed of finite number of samples and finite sampling intervals, limiting the construct in terms of its universality. However, it is still possible to approximate the ideal behavior of this decomposition using a variety of methods. Two general types of methods exist to this end: the first group utilize direct methods of computing the frequency decomposition using an implementation of the discrete Fourier transform which is defined by the following equation.

$$F_j = \sum_{k=0}^{N-1} f_k e^{2\pi i j k / N}$$

Where F represents the frequency domain representation of the signal and f represents the time domain, and N denotes the number of samples. The second group creates a model for the signal and then studies the frequency response of the model. The former are called nonparametric methods and the latter are termed parametric methods because of the intermediate need to approximate the parameters of a model prior to the computation of the coefficients associated with each sinusoidal time-scale[52]. Because of the fundamental elegance of interpretation of the ideal approximation in terms of sinusoids which are essentially one of the best understood mathematic functions in physics and engineering, methods which attempt to decompose the signal in terms of such have been ubiquitous in physics and engineering, and thus analysis of

the frequency content of the signal has become synonymous with this type of decomposition. As a result dozens of computational methods have been developed for approximating this analysis from sampled, short and even noisy time-series. Each method is, in essence trying to capture the same underlying distribution but because of the realities associated with real signals (limited duration, limited samples per period at high frequencies and noise) end up approximating that distribution with different errors. The various approaches are then, in essence, a way to reduce specific types of errors in specific types of situations and thus no one method has been able to be shown superior to the others without considering the context. These impose fundamental limitations on the methods of analysis.

Limited duration of the signal leads to reduced ability to approximate the coefficients associated with each frequency. At higher frequencies (higher with respect to the signal length, such that many oscillations occur inside the length of the time-series) the reduced resolution may come from reduced statistical strength, where is at the lower frequencies, because even a single oscillation is not complete, low frequency oscillations may not be defined and may simply serve as distortions. For this reason, the longest possible sequence of data is used when possible and mean and linear trends are removed prior to analysis, a practice followed for all frequency analysis in this document. The limitation of length also affects different algorithms differently. For example certain methods rely on decomposing the signal into eigen-modes which may require multiple periods to be detected whereas some parametric methods may be able to characterize the frequency response of the system by fitting a small amount of data.

Presence of noise, which can be defined in the most general sense as “signal which is not part of the information of interest” is detrimental to all analysis, but depending on the nature of the noise, affects different routines to different degrees. For example, if a long time-series is available, averaging methods can often eliminate normally distributed frequency nonspecific (“white”) noise, whereas pre-filtering may be more effective against frequency specific (“colored”) noise [53]. Parametric methods can often yield very good results if the a signal model is well understood and thus making the assumptions in the model estimating the parameters are highly valid, whereas the presence of nonlinearity and nonstationary behavior can effect parametric methods detrimentally.

Both parametric and non-parametric methods of spectral analysis were applied to the human glucose time-series in order to study the frequency content of the signal. The result of such analysis is frequently expressed in a plot called a power spectrum, where the signal energy per unit of the frequency domain per sample is graphed against the frequency. A set of such methods was explored for the purpose of spectral analysis, and a brief description of them is given in table 3-1.

Table 3-1: Various methods of estimating the frequency spectrum of a time-series.

Routine	Implementation	Comment
FFT w/ windowing	Author (using Matlab FFT)	Simplest most general purpose approach
FFT w/ windowing, incoherent averaging (50% segment overlap)	Author (using Matlab FFT)	Increased optimality at high frequencies, w/ loss of information at low.
Multi-taper method	MATLAB	Uses frequency domain windowing
Eigenanalysis	MATLAB	
Covariance	MATLAB	
Modified covariance	MATLAB	Slightly different (uses bidirectional prediction error)
Burg model	MATLAB	Method of AR estimation
Yule-AR	MATLAB	Method of AR estimation
Direct AR	Author (using Matlab ARMAX)	Box-Jenkins model
Prony model	Author (using Matlab Prony)	Complex exponential model

FFT spectral estimation is the most widely used method and is based on the Fast Fourier Transform, which is a computationally fast method of estimating the discrete Fourier transform of a uniformly sampled signal. The DFT and FFT are the most direct computational methods of moving from time to frequency domain without a change in the overall signal energy. As mentioned above, the shortness of the time-series for the purpose of analysis introduces distortions at the end point where the signal abruptly ends. To overcome this, the signal is attenuated at the beginning and end using a windowing function which multiplies by the signal prior to processing. Care then has to be taken to take this multiplication into account in calculating signal energies. In essence windowing reduces the importance of the beginning and the end of the time-series to the calculation while emphasizing the middle. Each window, like

each method of spectral estimation has advantages and disadvantages and the same window, a rather general choice, the Kaiser window with a coefficient of 8.0 was used for all analysis for the purpose of consistency.

Signal energy in time-domain

$$E = \sum_{n=1}^{N} a_n^2$$

Signal energy in frequency domain

$$E = \sum_{w=1}^{W} a_w^2$$

FFT spectral estimation can be further refined by estimating the Fourier transform for smaller snippets of the data and averaging the resulting distribution. This can lead to an improved estimate of the higher frequency components of the signal, but because the lowest frequency estimated by the method is dependent on the total length of the signal, frequency estimates at lower frequencies are sacrificed for this purpose. Thus the data is divided into multiple snippets, with variable amount of overlap between the snippets. The overlap is useful because as discussed above, the windowing lessens the importance of parts of the snippets during the FFT process, and thus without overlapping, not all the information in the time-series is used optimally for the estimate. The use of multiple snippets does not warrant discontinuation of windowing as the algorithm still only sees and computes the FFT one snippet at a time and has no knowledge of the fact that the snippet end is not the end point in the full time-series. A typical value of 50% overlap was utilized for the purpose of the analysis. In averaging the resulting frequency domain representation, care has to be taken to average the energies in each band (termed incoherent averaging) and not the actual direct frequency coefficients (which are complex numbers) as phase cancellations in the

complex domain may lead to cancellation of frequency content (termed coherent averaging).

The multi-taper method (MTM) performs a similar averaging and windowing (tapers in frequency domain) as the method described above in the frequency domain. Like the above two methods, it uses FFT at its core but modifies the FFT process using windowing and averaging of snippets but in the frequency domain instead of the time domain [54].

Another nonparametric approach to estimating the frequency content of the signal is by the assessment of the correlation matrix, computed from the signal [55]. The autocorrelation function, which can be derived from the signal, can be directly transformed using the FFT algorithm into the power spectrum estimate (this, however, in itself offers little advantage). However, analysis of the eigenvectors of the correlation matrix, and their translation into frequency domain estimates of signal energy generates what is known as the psudeospectrum. This approach is particularly advantageous in that in theory, eigenvectors associated with noise may be separated and not analyzed.

Parametric approaches attempt to estimate a model for the signal and then study the frequency response of that model in order to estimate the frequency spectrum of the data. A simple model of the covariance matrix can be used to compute a model whose frequency response can be then be estimated. The covariance matrix

can be used to estimate the coefficients of a nth order autoregressive (all-pole) model. The pcov() routine in Matlab minimizes the forward prediction error (Least Square) while the modified covariance algorithm, pmcov() uses both the forward and backward prediction error to estimate the coefficients of the AR model. The power spectrum is then computed by studying the frequency response of the model. The Yule-Walker method can also be used to compute the AR coefficients pyulear(), as can the burg method which uses both forward and backward prediction errors to optimize the estimate of the AR model using the same solution technique (Levinson-Durbin recursion), and is implemented in pburg(). Another approach to calculating the frequency response is by using models which generate both zero and pole coefficients. One such model is the Prony model. The Prony model is based on using complex exponentials as the basis for finding the frequency behavior of the system. Their use in a biological signal processing context is discussed in [56]

To evaluate the utility of various methods was tested using simulated time-series with known spectral content. Because the time-series in the data set are short relative to the slowest possible time-scales (such as circadian rhythms and meals) one of the key criterion of methods employed must be that they can make computations using short time-series. Additionally, noise and distortion must not excessively degrade the estimate. The spectral estimates using multiple methods were applied to a variety of different time-series under different sampling conditions to better understand the limitations of each routine. Three composite figures generated from such multiple approaches are shown in figures 3-6 to 3-8.

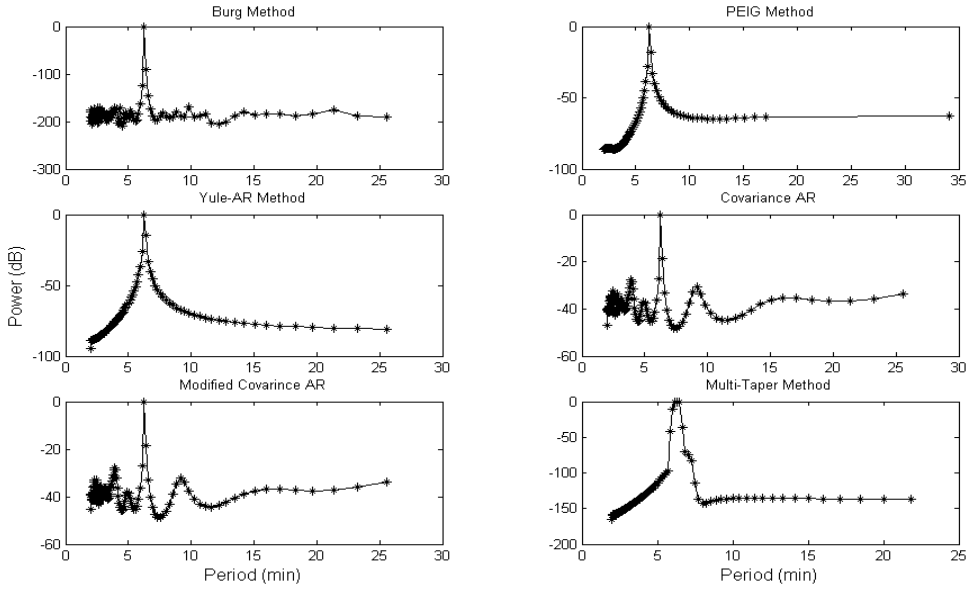


Figure 3-6: A demonstration of six of the methods discussed applied to a single sinusoid w/ a period of 6 minutes, with 240 minutes of data and sampling of 1 per minute. This represents an ideal data set. Model order of 100 minutes was used for AR models and 25 minutes for the covariance methods.

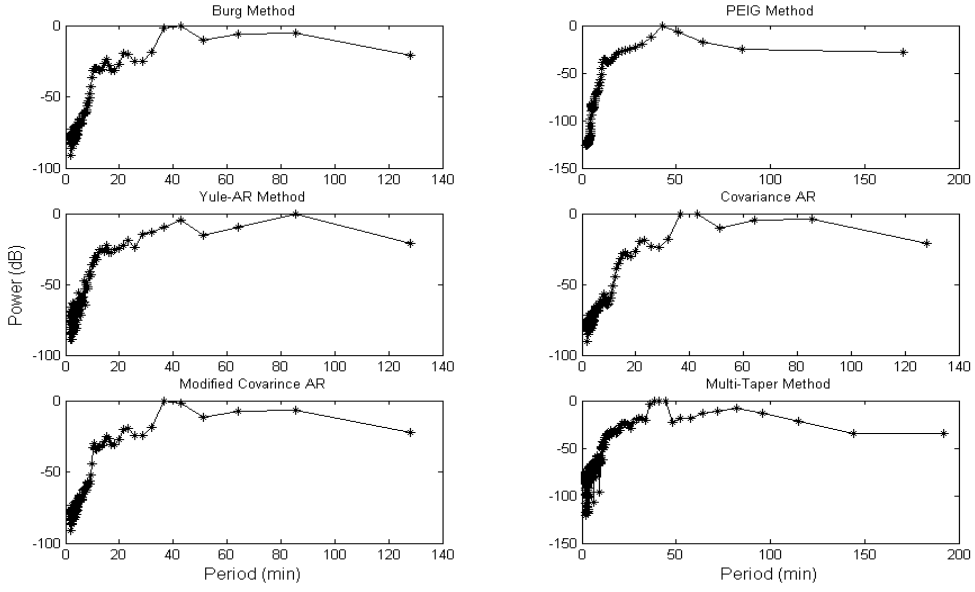


Figure 3-7: A demonstration of six of the methods discussed applied to a single sinusoid w/ a period of 40 minutes, mixed with a sinusoid with nonlinearly varying frequency, 576 minutes of data and sampling of 1 per minute. This is meant to simulate data from constant infusion of glucose. Model order of 50 minutes was used for AR models and the covariance methods. The Peig routine calculated a maximum of 10 eigenvectors.

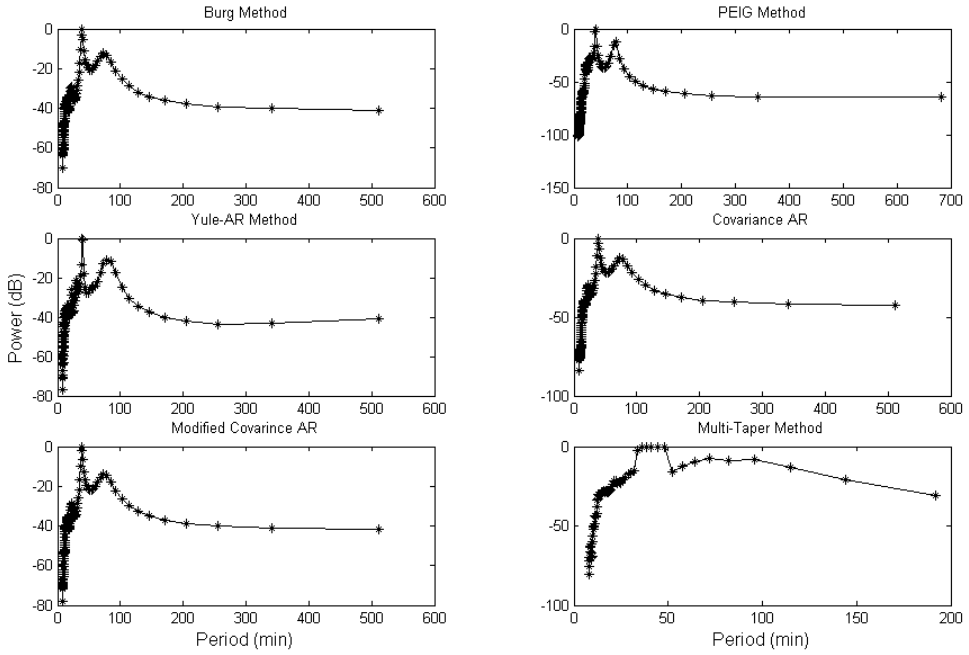


Figure 3-8: A demonstration of six of the methods discussed applied to a single sinusoid w/ a period of 40 minutes, mixed with a sinusoid with nonlinearly varying frequency, 576 minutes of data and sampling of 4 per minute. This is meant to simulate data from constant infusion of glucose. Model order of 200 minutes was used for AR models and the covariance methods. The Peig routine calculated a maximum of 10 eigenvectors.

All methods performed well in this context in finding the peak related to the constant sinusoid, however, the multi-taper method tended to create a broader peak, the covariance and modified covariance tended to produce very similar results and the peig algorithm is very sensitive to the estimated number of sinusoids in the signal (which translates into the maximum eigenvectors decomposed). Because the analysis here is to be applied to a wide variety of signal, sensitivities to parameters are difficult to manage because they require a-priori knowledge of the signal and changing it depending on the tracing. The Yule-AR and the Burg AR also seem to perform similarly and in other tests performed that are not shown. Figures 3-9 and 3-10 show

some of the tests already discussed and new ones that were not in the previous figures. These include the Prony method based estimate as well as the averaged FFT method.

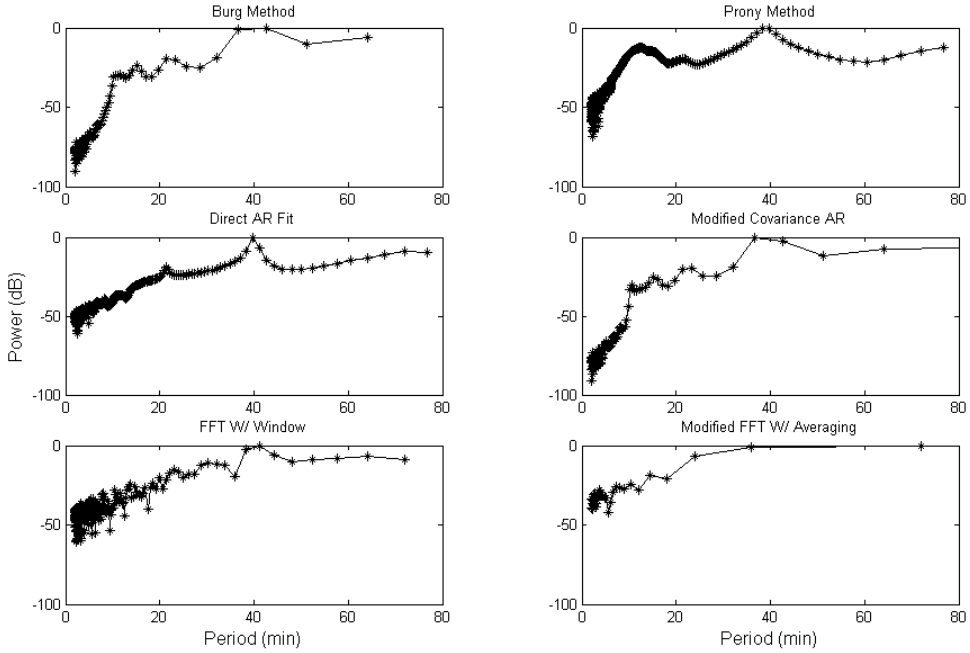


Figure 3-9: A demonstration of six of the methods discussed applied to a single sinusoid w/ a period of 40 minutes, mixed with a sinusoid with nonlinearly varying frequency, 576 minutes of data and sampling of 1 per minute. This is meant to simulate data from constant infusion of glucose. Model order of 50 minutes was used for AR models and the covariance methods.

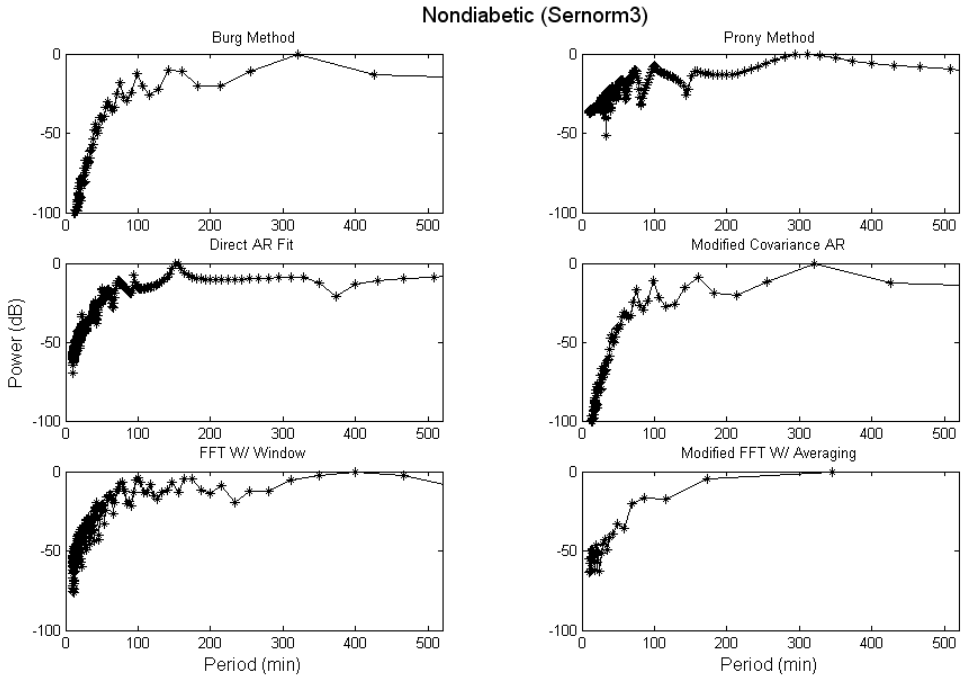


Figure 3-10: A demonstration of six of the methods discussed applied to a tracing of nondiabetic blood glucose levels sampled every 5 minutes for 48 hours.. Model order of 350 minutes (70 samples) was used for AR and a model order of 1400 minutes (280 samples) was used for the burg, Prony and covariance methods.

After an exhaustive search, three methods which seemed sufficiently "different" from each other were chosen and have been applied throughout this text. The burg model was chosen because of the ease of implementation and because other AR based methods did not seem to produce significantly different estimates in the range of samples and time-scales of interest. The Prony method was taken as a different model of frequency response and the simple windowed FFT was used as a nonparametric measure. These three methods (along with two that were not used) are shown once again applied to a different type of signal from nondiabetics in figure 3-11.

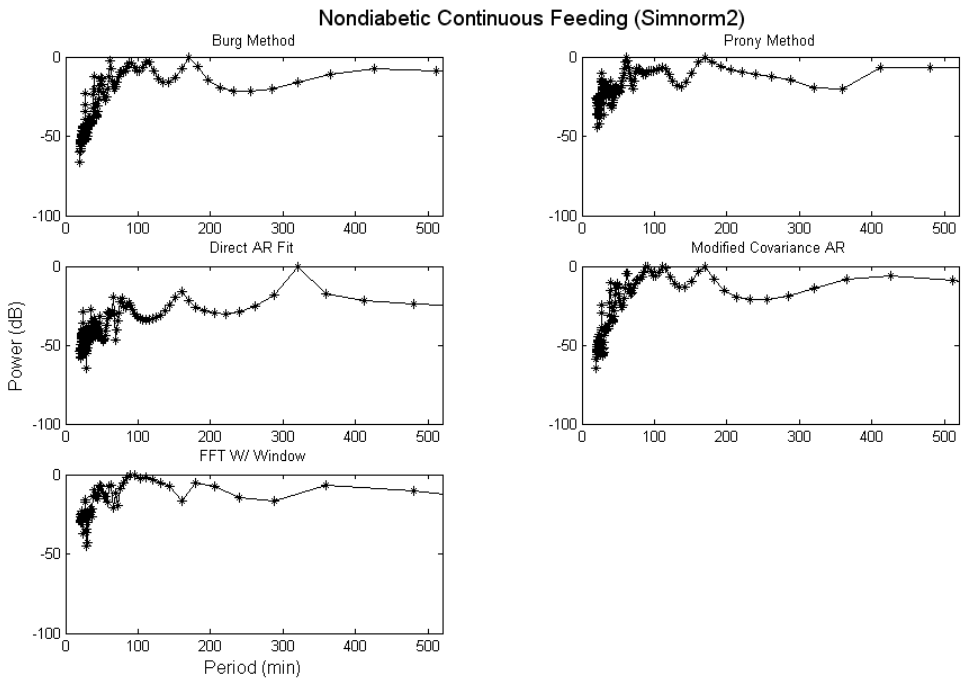


Figure 3-11: A demonstration of five of the methods discussed applied to a tracing of nondiabetic blood glucose levels during continuous feeding sampled every 10 minutes for 28 hours.. Model order of 350 minutes (70 samples) was used for AR and a model order of 1400 minutes (140 samples) was used for the burg, Prony and covariance methods.

In order to look at multiple individuals, a method of looking at multiple spectra for the specific purposes of analyzing relative and mean distributions is to show each spectra as a single row of pixels with different intensities. Each power spectrum is re-sampled in order to facilitate visual comparison between datasets. The sampling is determined by the lowest sampling frequency between all datasets. Furthermore, because the interest in this case is the relative distribution of energies, all spectral estimates are normalized to make comparison across data possible. They are displayed on a color scale ranging from white to black, denoting values from zero to one with one representing the highest peak in that spectrum, for that specific time series. These

are then averaged to yield a representative spectrum on the bottom of the colored map. An example of such a figure, demonstrating the effect of the model order on the burg estimator is shown in figure 3-12.

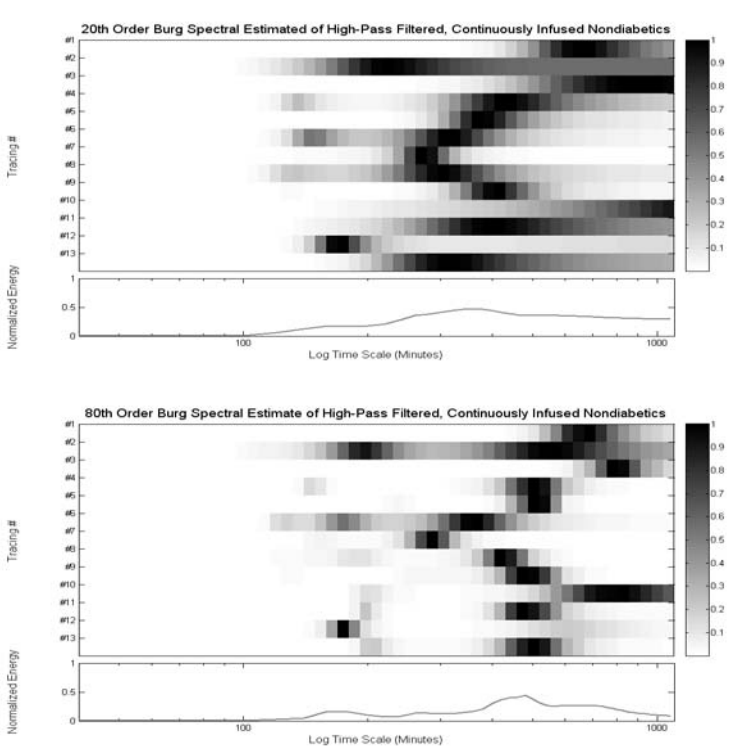


Figure 3-12: The comparison of two nondiabetic subgroups with the spectrum estimated using a 20^{th} order (top) and an 80^{th} order (bottom) Burg estimator. Note the increase in localization of the peaks.

III.E Time-frequency approaches

The methods in the previous section assume that the frequency content of the signal is constant and thus assess the signal as a single entity. It is critical to realize that the relationship between time and frequency is governed by an uncertainty principle: the more exact one tries to localize the signal in the frequency domain, the less the resolution can be defined in the time domain [50]. This uncertainty principle is a consequence of the properties of Fourier transform pairs, a class to which time and frequency belong to. In the case of spectral analysis, the time resolution is fully "traded" in return for maximizing the resolution in the frequency domain. Thus to study the frequency behavior of the signal over time, loss of frequency resolution is a necessary cost.

The results from time-frequency approaches are shown in a time-frequency plot where signal energy is denoted by color, the vertical axis represents some measure of frequency (or period, time-scale, length-scale) and the horizontal axis represents the progression of the window in time. The difference in the approaches, in essence amounts to approach taken to define frequency and the specific trades that are made in time-frequency analysis. The short time Fourier transform, for example, directly trades resolution in frequency for time at the same scale for all frequencies whereas wavelet decomposition makes a different trade depending on the frequency scale of the analysis. The methods are summarized in the table 3-2.

Table 3-2: A short list of the general class of time-frequency approaches considered for this analysis. They are discussed individually in the following sections.

	Basis	Advantages	Disadvantages
Short Time Fourier Transform	Fourier Series	Ease of Interpretation	Poor Resolution
Wavelets	Wavelet Library		Hard to Interpret, dependent on wavelet chosen
Wigner-Ville and Pseudo-Wigner-Ville Distributions	None	Simple, gives a possible "instantaneous frequency"	Distortions resulting from cross terms
Emperical Mode Decomposition	Derived from the data	Decomposes signal into modes, gives a possible instantaneous frequency	Hard to interpret

III.E.1 Short-time Fourier transform

This is the classical approach to time-frequency analysis, which involves a moving window which is short in comparison to the time-series, and in which the Fourier transform is computed. Essentially, each column represents a spectral estimate with the window starting at that time-point. Clearly, the resolution in time is thus further limited (in addition to constraints of the fundamental uncertainty discussed above) by the resolution of the spectral estimates

III.E.2 Wavelet decomposition

Wavelet decomposition is similar to Fourier decomposition in that the signal is decomposed into a set of coefficients by taking the inner product of the signal with a basis function. But unlike the Fourier transform where the basis set consists of sinusoids which extend towards infinity towards both directions, the basis set of wavelet transform consist of functions which are localized in time and tend to zero in either directions of infinity. Each new basis is derived by moving in time (translation) and rescaling of the original wavelet function. An example of such a decomposition is shown in figure 3-13.

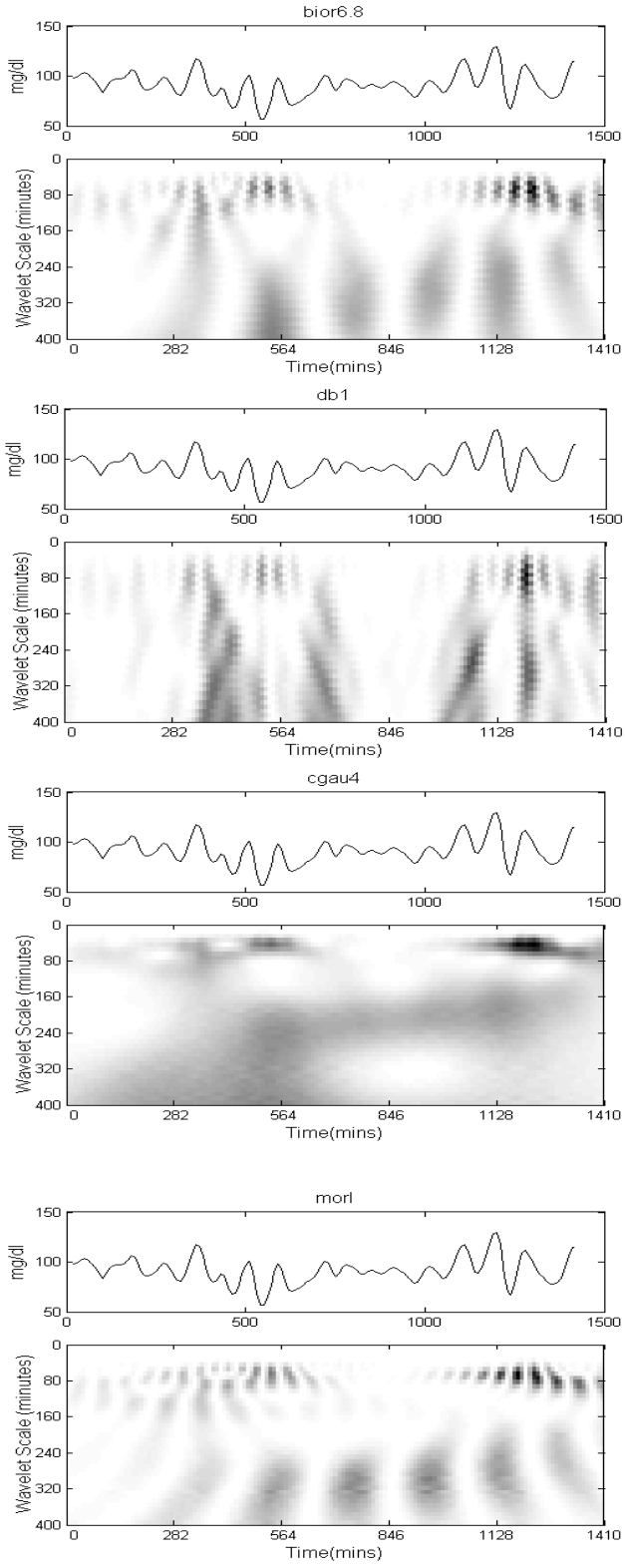


Figure 3-13: The representation of a signal with variable time-frequency components represented in time-frequency domain using various wavelet basis sets.

In theory, the more compact the wavelet representation is, the more the wavelet is able to capture the signal and hence it is considered optimal for that signal. The strength of wavelets in their diversity also then leads to their main drawback which is that the optimal wavelet for a given signal is highly dependent on the signal in question, and in particular is sensitive to sampling periods and data length[57].

An algorithm was devised to systematically search through a library of continuous wavelet transforms to search for a compact wavelet representation. This algorithm did not find a wavelet basis that was universally produced a compact representation. Thus two wavelets were chosen for the purpose of analysis. They were chosen because they represented two very different types of wavelets. The first one, a first order Daubechies wavelet was chosen because of the very compact support (that is high time-localization) whereas the Gaussian wavelets provide non-compact support but are symmetric. Both these were implemented in the Matlab wavelet toolbox (Matlab 6.5.1).

The resulting plot from a time-frequency representation can be informative about the different shifts in time-scales in the signal. However, when trying to reach population based conclusions, it becomes necessary to be able to represent the time-frequency in terms of a few parameters. To this end a method was devised which would locate the “center” of signal energy concentration in each time point. This was achieved by taking a weighted (by signal energy) average of the location of the top signal energy contributors for that time. In the case of wavelets time points during

which no dense signal energy concentration was detected (which is a side-effect of the nature of wavelets) were not included in the analysis. The location of these central points were tracked and averaged to yield a mean frequency peak. The variations about this mean were taken as the primary surrogate for time-frequency variability. An example such a plot for the STFT and wavelet decomposition is shown in figure 3-14 and 3-15.

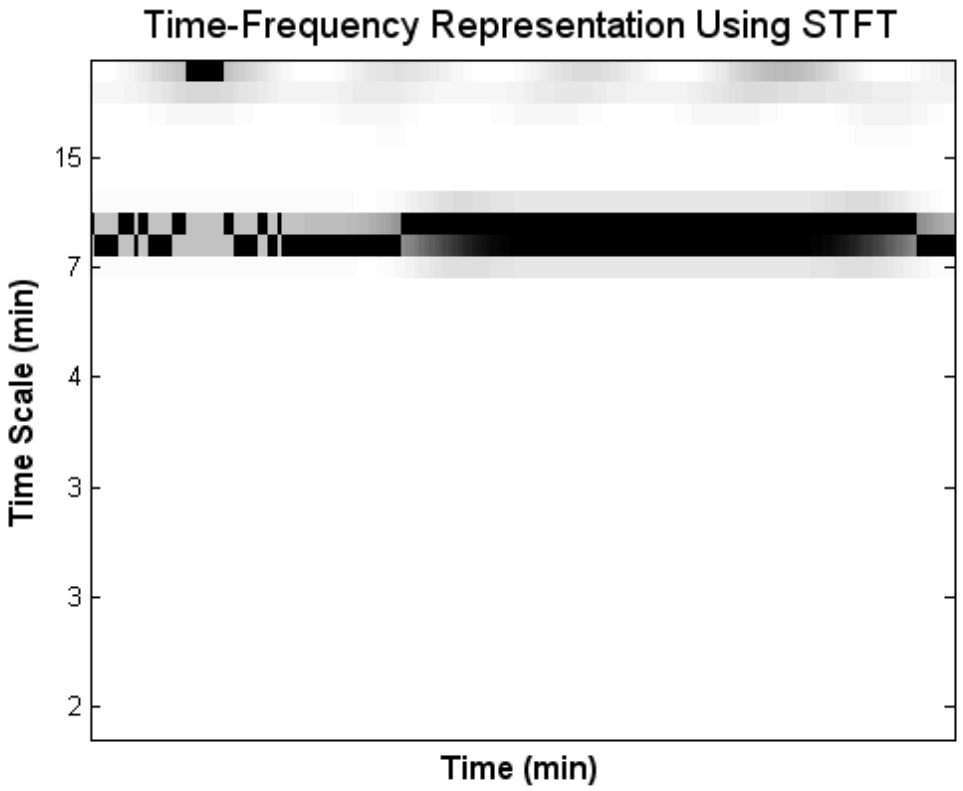


Figure 3-14: Dark squares superimposed on the gray scale image of the time-frequency response computed using the short-time Fourier transform indicate the "center" of energy localization in the time-frequency plot.

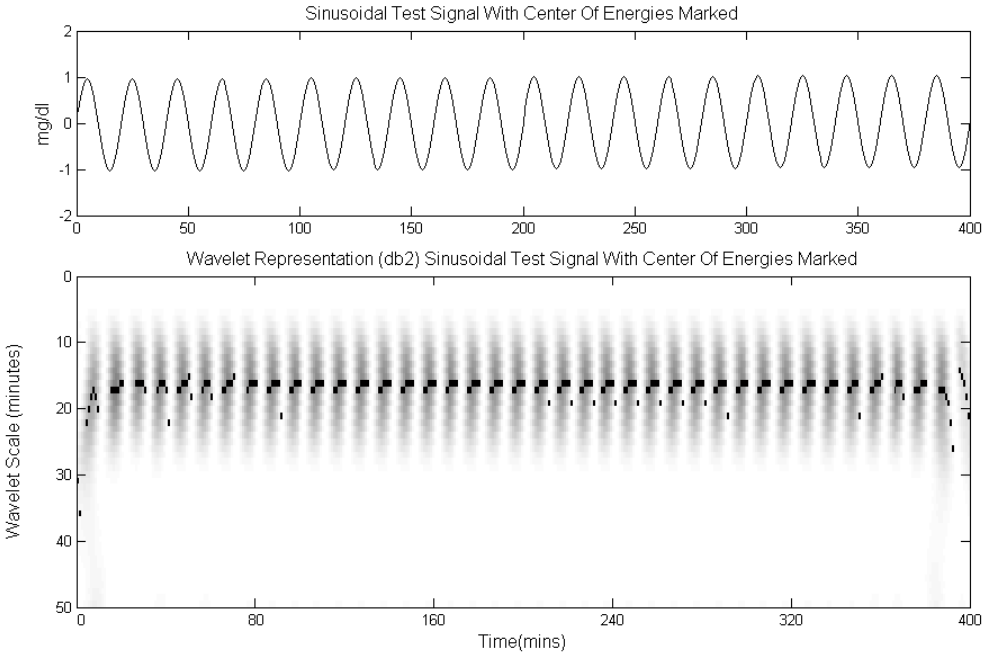


Figure 3-15: The center of energy localization shown by the very dark squares superimposed on the time-frequency plot, for a sinusoid with a wavelength of ~20.

III.E.3 Empirical mode decomposition and Hilbert-Huang transform.

Empirical mode decomposition represents an attempt to decompose the signal into different time-scales without the use of any predefined basis function. In other words, the time-scales upon which the data is projected in order to decompose the signal are not predetermined but are rather derived from the signal itself. In this way, this method (and other basis independent methods such as Singular Value Decomposition and Eigen-analysis), fundamentally differ from the ones discussed so far because they make no assumption about the nature of the time-scales that the signals being decomposed into. This provides these methods with a tremendous advantage in terms of analyzing time-series generated from signals containing

nonlinearities and possible nonstationarities. The tradeoff comes in the form of uniqueness and interpretation. Even a slight change in the make of the time-series (for example phase related changes) may lead to an alternative though albeit correct decomposition. This makes the interpretation of the results difficult in particular for the purpose of comparison between individual tracings. Additionally, the interpretation of the modes themselves can be difficult, because they are not related to each other in a strict sense (although there is a progression in the time-scale of each mode with the latter modes containing slower oscillations).

The process of generating the EMD is discussed thoroughly in multiple references [58] but can be summarized as follows: lower and upper envelopes are computed for the signal by interpolation of lines connecting the maxima and minima of the signal. The mean envelope is then computed between these two envelopes and is subtracted from the signal. This is repeated until a mean envelope of nearly zero results from the signal. This new signal is called a mode. The mode is then subtracted from the signal and the process is repeated to the resulting signal with the mode removed. The decomposition is terminated after a few repetitions or when the only remaining mode has few if any oscillations. An example of such decomposition, applied to the nondiabetic time series using code implemented by the author is shown in figure 3-16.

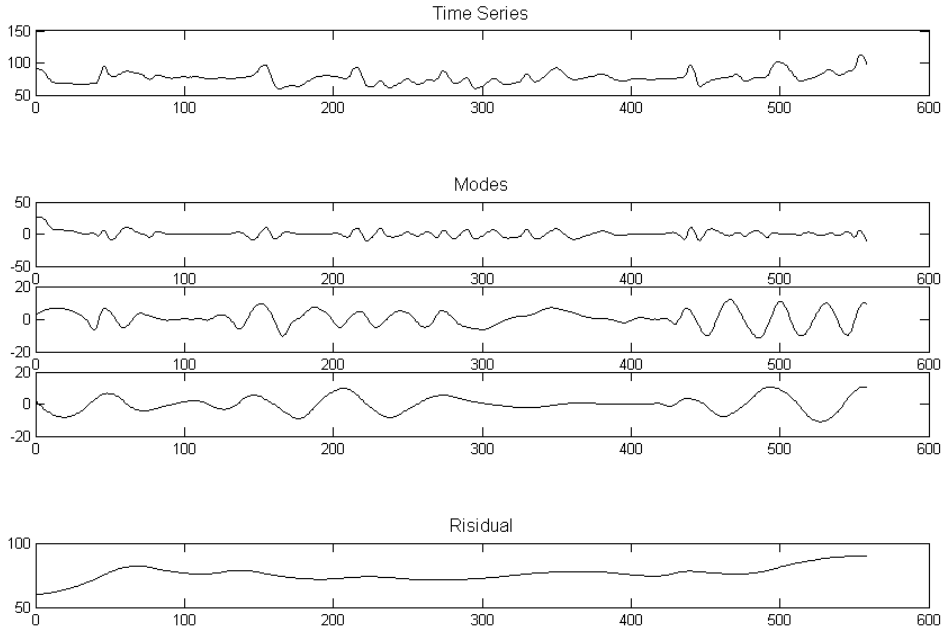


Figure 3-16: A time-series from a nondiabetic is decomposed into three modes and a residual using an implementation of the EMD. Notice that the time-scales of each mode become progressively longer.

Aside from not relying on any external definition of a basis set, the EMD can be used, after conversion to an analytic signal, to calculate an instantaneous frequency. An In signal terms, this means that the signals have zero signal energy in the negative frequencies, which result when using the Fourier transform to translate most real signal into frequency domain. An analytic signal is created by taking the real component and adding, in quadrature the Hilbert transform of the signal [59] . Analytic signals have the critical property that the negative frequency components vanish, permitting the computation of the concept of instantaneous frequency.

The Hilbert transform is defined by:

$$H(t) = s(t) + i\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{s(\tau)}{t-\tau}d\tau$$

The quantity H(t) represents the analytical signal obtained by taking the Hilbert transform, which combined with the EMD processing is termed the Hilbert-Huang transform. This is performed on each separate EMD signal component individual. The phase component of this analytic signal fluctuates in a manner such that its derivate represents a quantity with the dimensions of frequency and significant literature debate exists as to whether this represents a real quantity. However, it has been proposed that this frequency like component represents a good estimate of the instantaneous frequency as there is no negative frequency component and thus a single frequency like number is computer per each unit time. Applying this to each component of the EMD yields a group of instantaneous frequency plots which represent the frequency fluctuations of each of the empirical modes.

As mentioned, difficulty with interpretation persists as the main problem with this method. Because a significant amount of the signal energy is concentrated in the meal events, insufficient meal events in a 48 hour period exist to really take advantage of this and other “independent mode” decompositions which rely on many wave iterations to separate signal components. Additionally, the instantaneous frequency component did not yield insights not available with the simpler short-time Fourier transform approaches discussed above. Thus while promising, this method was not used in the final analysis, but may promise to be a useful tool in the prediction or

categorization of signals with larger datasets and may offer superior results in the context of large-scale data analysis.

III.E.5 Pattern finding in time-frequency analysis

Another approach to looking at the temporal evolution of signals, and in particular ones which contain pulsatile behavior is to identify the pulses in the signal using the peaks of the signal and subsequently analyze the properties of these signals. These were implemented on certain time-series which exhibited pulsatile behavior which was specifically of interest. Peaks were detected by looking at maxima in the time-series which were significant and not in the near neighborhood of other maxima to prevent “false” detection. In reality, because no truth-table exists, it is very hard to judge the performance of such pulse detectors in terms of detecting all the pulses in the signal. However, the pulses that are detected can be characterized very well in terms of their temporal characteristics. This method, in essence is a time-frequency based method in which certain singularities (peaks) are used to gain very good temporal resolution, but require knowledge about the specific features that are of interest.

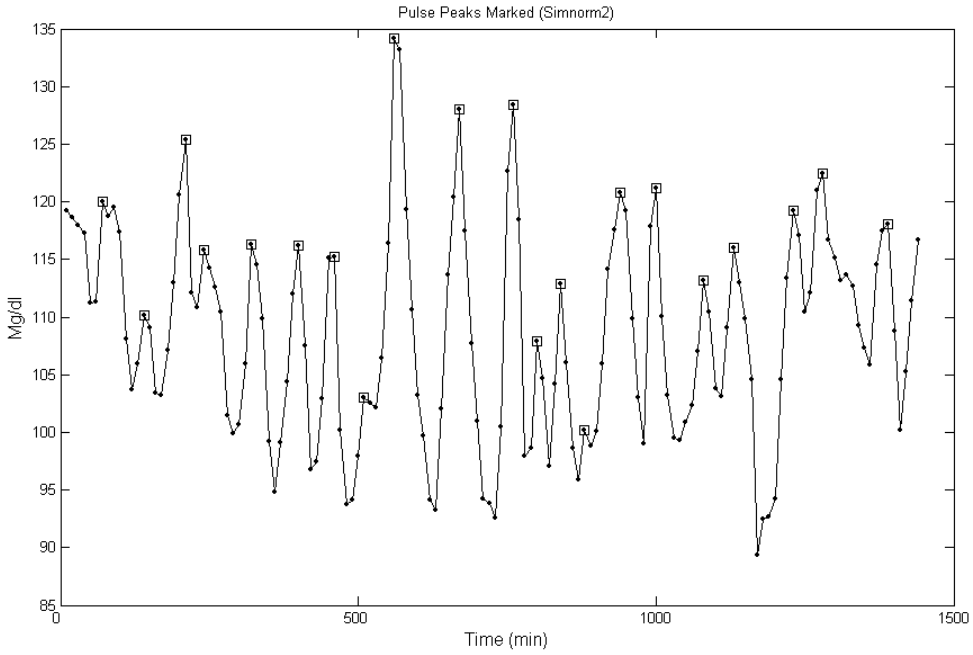


Figure 3-17: the top of pulses are marked by looking for points where the signal changes direction. To avoid marking small fluctuations, certain limits (such as distance from the previous pulse are used to reject some of the candidate points. The distance between these points forms the basis for pulse width analysis.

As shown in figure 3-17, by varying parameters such as the required rate of change and the inhibition of pulses in the vicinity of each other (as discussed above), good agreement, at least with visual perception is reached. The distribution, in terms of the number of pulses per unit time is then used to characterize the time-series.

III.F Information half-life

Though intimately related to the time-scales of dynamics, another approach to better understanding the signal is to look at the information content of the signal. A metric of interest is the rate at which information disappears with time. That is, how much does the information at current time predict the future values of the signal. This rate of disappearance is intimately related to the way information is extracted from the signal. To this end, three different approaches are used to assess information dissipation with time. These are based on the average mutual information, autocorrelation function and the difference in sample energies. The autocorrelation function, which will be revisited in multiple contexts throughout this document is defined by the following equation:

$$R_{xy}(m) = E\{x_{n+m} x_n\}$$

where E denotes the expectation value. The average mutual information is similar to the autocorrelation function but instead of simply looking at the product of the samples, looks at the correlation between the vector quantized states of the system. Strictly speaking, the mutual information of two discrete independent variables X and Y is given by the following equation:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)},$$

where p(X;Y) is the joint probability density function of the two variables while p(X) and p(Y) are the marginal probability density functions of each variable. In applying this to a time series, he process of vector quantization is used to create discrete states

using which the probability densities are calculated between the variable at time t and t-n. The average mutual information at lag n is the amount of information known about the variable at time t based on the distribution at time t-n. Much like the autocorrelation function, this is a measure of the information content of the signal framed in terms of time. Looking at the way in which this value changes as the delay between the samples increases gives us another way of looking at the information half-life which is more based on the information contents of the signal rather than a linear correlation.

Finally, the simple difference in energies can be used as the simplest energy based measurement of the change of the time-series value. The energy at time t and t-n are calculated and the time delay n is compared to the change in the signal energy during that time. Combined, these three measurements and the corresponding calculation of information half-life (That is the time required for the value to change of these metrics to be reduced by 50%) can be used to assess the time-scales of change of the information content of the signal.

III.G Entropy, Approximate Entropy and Sample Entropy

Vector quantization is a process by which continuous values are assigned discrete “bins”. The simplest form of vector quantization is the generation of histogram, in which case evenly sized bins are used to summarize the data. Most forms of vector quantization contain loss in information as the data is reduced in

dimensionality. Nonlinear complexity measures such as entropy use vector quantization to describe the complexity of a signal [60, 61]

Approximate entropy uses the concept of patterns to measure the “regularity” of a time-series. Patterns are defined by their length and the tolerance of difference between patterns that can fall into a category. This method has been successfully developed and applied in assessing physiological dynamics in the human cardiovascular system [62] and in particular in terms of disease processes [10].

In a similar way as approximate entropy, sample entropy creates bins based on a similarity parameter r, which is a percentage difference between two values [63, 64]. If the percentage difference between the two values is less than r, then they are considered to be in the same bin. Sample entropy procedures rely on finding the patterns of m consecutive bin transitions that are similar and comparing the frequency of similar transitions to each other. For example, with m = 2 and the time-series starting with the value of 100 and then 80, the routine will count the number of times a value of 90-110 is followed by a value of 72-88 (two bins of +10% and -10% around the value of 90 and 100 respectively). This will constitute a pattern and the number of these cases will be counted. The next two values in the time-series are then evaluated as a pattern template and the number of times similar transitions are detected are noted. Thus, in time-series where similar transitions exist between bins, the counts will be high. The procedure is then repeated with three sample long patterns. The relative number of patterns detected at different pattern lengths form the basis for the entropy

calculation. Approximate entropy provides a good dynamic measure of complexity as it is not sensitive to the simple mean and standard deviation of the time-series as compared to the standard information theoretic entropy measurement (table 3-3).

Table 3-3: Some values of entropy and approximate entropy calculated for some test signals. The signals are detrended and normalized based on their range, and the entropy is recalculated. Approximate entropy does not change because of the rescaling of the time-series.

	Entropy	**Detrended & Normalized**	**Apen**	**Detrended & normalized**
Gaussian Noise	2.98	1.42	1.37	1.37
Sines W/ Nonlinearly varying frequency	3.5	1.33	.44	.45
Chaotic Series (Chuas)	3.0	1.3	.61	.61
Chaotic Series (Chuas 2)	3.08	1.2	.49	.49
Sine + Noise	4.5	1.41	1.42	1.45
Two Sines	4.2	1.31	.17	.18
Single Sine	3.33	.87	.20	.21
Sine Shifted	3.43	.94	.20	.21

III.H Study of the underlying system

Biological systems tend to contain many nonlinear elements, a characteristic discussed extensively in the literature ([60, 61, 65]). The presence of nonlinear signal structure has strong implications for development of models and also the validity of linear approaches to treating the signal. The sources of nonlinearity may result from the underlying physiology, transport of glucose as well as measurement techniques

(which presumably play a negligible part in the data sets in question). Testing for nonlinear behavior in signals is not a very well developed field for many reasons which include lack of approaches to signal processing once the nonlinearity has been detected, lack of computational power, and increased complexity in defining nonlinearity and separating it from noise, measurement distortions and artifacts. The approaches to testing signals for nonlinear content have been largely developed by the physiology and economics communities which both deal with very complex time-series. Much of the algorithms and simulations have been proposed and implemented in the past two decades, and largely rely on simulated data (as opposed to real, noisy data). For this reason interpretation of these tests must be taken with the awareness that they may suggest nonlinear signal contents rather than proving it. This is not simply caused by the weakness of the test statistics but rather by the fundamental computational implementation and its susceptibility to error, in particular with smaller data sets and with operator input of parameters which must be estimated using human intervention. These difficulties, as will be discussed, also reappear in the nonlinear time-series analysis after nonlinearities are suspected. The only prescribed remedy, it appears, is to approach the problem using multiple tools and interpret the evidence in a holistic sense.

III.H.1 Tests for presence of nonlinearity

Multiple tests were examined for the purposes of testing the degree of nonlinearity in the signal content. An introduction to these methods can be found in a

single reference [66] where they were used in a blind comparison on simulated data in order to determine whether the data was generated by a nonlinear system of equation. Two tests which were not used are described below to illustrate the diversity of approaches to this problem.

The Hinich bispectrum test uses higher order statistics to test for nonlinearity and gaussian profile. The term higher order statistics, derives from the equations that give rise to the common statistical descriptors of distributions such as the mean and standard deviation. The power spectrum can be computed by taking the Fourier transform of the autocorrelation function and similarly, the bispectrum can be computed by taking the Fourier transform of the higher order cumulants. The output is a two dimensional relationship between the frequency in the data. Properties of this result can be used to test for Gaussian distribution and by extension for the presence of a Gaussian linear process. This test suffers from lack of specificity in terms of the null hypothesis as well as requiring a significant number of samples to compute the higher order cumulants.

The BDS method does not provide a direct test for nonlinearity or chaos, as the sampling distribution of the test statistic is unknown. The test statistics are characterized, however, for a null hypothesis of independence of samples. Thus, if for example all linearity is removed from the signal by fitting of a general purpose linear model (like an ARMA model) then the remaining signal dependence, which can be tested using BDS may be attributed to nonlinearities. This off course leads to the

challenge of completely removing all linear dependence, or linear pre-whitening. In one implementation, an ARMA model estimated using the Box-Jenkins method was used to pre-filter the signal. This method was not implemented because it relies heavily on having optimized a linear model

The methods that seemed most promising with smaller data sets were based on the method of surrogates. Here the goal is to create alternative signals, which are statistical equivalents of the original signal and examine various forms of determinism in these signals. In the case of detecting nonlinearities, the objective is to reject the null hypothesis that the signal is a deterministic linear signal (assuming determinism is based on a non-rapidly falling autocorrelation function). Since the determinism in a linear signal is “carried” in the autocorrelation function, surrogate signals with the same autocorrelation functions should contain the same amount of determinism. Thus the test amounts to construction of linear-equivalent surrogates and searching to see if determinism is significantly reduced. If it is not, then the null hypothesis cannot be rejected, because a linear surrogate to the time-series can be constructed that contains the same degree of determinism. The Kaplan delta-epsilon test was used for the purpose of assessing determinism. This test is well described extensively in multiple references [67]. Briefly, a test statistic for determinism is computed in multiple dimensions using time-delay embedding. This test statistic, generated by the Kaplan delta-epsilon method is then compared to the test statistic generated for the linear surrogates. If there is a linear surrogate that has a similar degree of determinism to the

original signal as displayed by the K statistic, then the linearity of the signal cannot be rejected.

This type of analysis has a clear interpretation and framework but relies heavily on the correct construction of surrogates and a good test for determinism that does not involve simple linear predictions (these predictors, would, after all only see the linear determinism to begin with). Two different codes for linear surrogate generation were used as well two different tests of determinism. Creation of linear equivalent surrogates is achieved by shuffling the phase in the Fourier domain without changing the signal energy at each frequency, which changes the essence of the signal without changing the autocorrelation function. A glucose signal and its linear equivalent surrogate are shown below in figure 3-18.

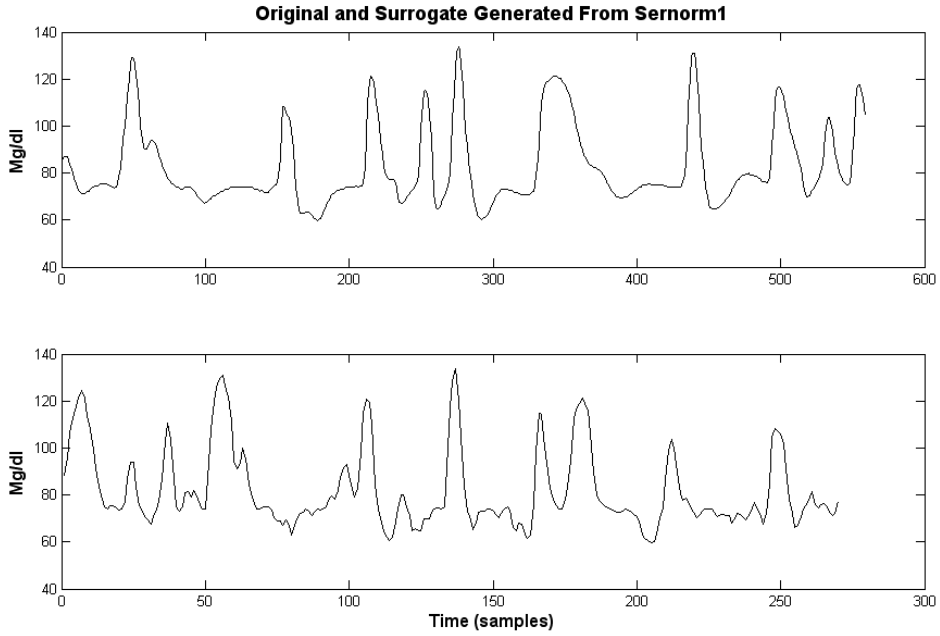


Figure 3-18: A glucose time-series and a linear surrogate with equivalent autocorrelation function.

Five models were used as described in [66], for the purposes of validating the tests for nonlinearity. Additionally, other signals with known but diverse set of properties were utilized. These include time series from simulations of the Lorenz attractor, simulate sine waves and normally distributed noise, as well as experimentally designed Chua's circuits (described in appendix C). These models are summarized in the table 3-4, and the Kaplan delta-epsilon test as applied to these models is shown in table 3-5.

Table 3-4: Core models utilized to test nonlinearity using various tests.

Model Name	Model Type	Equation	Initial Condition	# Samples
Model 1	Feigenbaum Chaotic	$Y(t) = 3.57y(t-1)(1-y(t1))$	$y(0) = .7;$	2000
Model 2	GARCH process	$Y(t) = h(t)^{1/2} u(t)$ $H(t) = 1 + .1y(t-1)^2 + .8h(t-1)$	None	2000
Model 3	Nonlinear ARMA	$Y(t) = u(t) + .8u(t-1)u(t-2)$	None	2000
Model 4	ARCH Process	$Y(t) = (1+.5y^2(t-1))^{1/2} u(t)$	$y(0)= 0;$	2000
Model 5	Linear ARMA	$Y(t) = .8y(t-1)+.15y(t-2) + u(t) + .3u(t-1)$	$y(0) = 1$ and $y(1) = .7$	2000
Sinusoid	Sinusoid (Nonlinear)	$Y(t) = \sin(2* \pi *t/100)$	None	1440
Noise	Gaussian Noise	Matlab Randn()	None	576
Lorenz	Chaotic Nonlinear	See appendix C		1250
Chuas1	Chaotic Nonlinear	Experimental Circuit (See Appendix C)	None	1250
Chuas2	Chaotic Nonlinear	Experimental Circuit (See Appendix C)	None	1250

Table 3-5: Delays were estimated for the various test signals including the models described above. The mean and minimum values of the test statistic K is shown for various dimensions, for 200 surrogates and the time-series. If the time series has a k smaller than the minimum surrogate K, then the null hypothesis of linearity can be rejected.

	Delay	Mean Surrogate K	Min Surrogate K	Std Dev Surrogate K	Time Series K	Dims	Result
Model 1	3	.1	.05	.02	1e-4	1	Strongly Reject Linearity
		.05	.03	.008	1.3e-4	2	
		.04	.03	.008	7.8e-5	3	
		.04	.02	.007	3e-5	4	
Model 2	2	3.5	3.5	.02	3.5	1	Weakly Reject Linearity
		3.5	3.3	.05	3.2	2	
		3.6	3.3	.08	2.9	3	
		3.5	3.2	.09	2.7	4	
Model 3	3	1.4	1.4	.008	1.4	1	Weakly Reject Linearity
		1.4	1.4	.02	1.3	2	
		1.4	1.3	.03	1.2	3	
		1.4	1.3	.04	1.1	4	
Model 4	3	1.5	1.5	.008	1.4	1	Weakly reject Linearity
		1.5	1.5	.02	1.3	2	
		1.5	1.4	.03	1.2	3	
		1.5	1.4	.04	1.1	4	
Model 5	36	1.2	1.2	.03	1.1	1	Cannot Reject Linearity
		1.2	1.0	.05	1.0	2	
		1.0	.8	.06	.9	3	
		.8	.7	.07	.8	4	

Table 3-5: Delays were estimated for the various test signals including the models described above. The mean and minimum values of the test statistic K is shown for various dimensions, for 200 surrogates and the time-series. If the time series has a k smaller than the minimum surrogate K, then the null hypothesis of linearity can be rejected (Continued).

Sinusoid	100	.09	.07	.007	.04	1	Cannot reject linearity
		.05	.03	.007	.04	2	
		.03	.01	.08	.04	3	
		.03	.001	.01	.04	4	
Noise	1	5.4	5.2	.1	5.4	1	Cannot Reject Linearity
		5.3	4.8	.2	5.4	2	
		5.4	4.8	.2	5.6	3	
		5.4	4.6	.3	5.7	4	
Lorenz	32	1.4	1.2	.08	1.0	1	Reject Linearity
		.8	.5	.1	.5	2	
		.4	.1	.1	.01	3	
Chuas 1	55	.16	.15	.003	.14	1	Reject Linearity
		.13	.1	.005	.1	2	
		.12	.1	.07	.07	3	
		.1	.09	.008	.06	4	
		.1	.06	.01	.05	5	
		.08	.05	.009	.03	6	
		.06	.03	.01	.04	7	
Chuas 2	49	.9	.8	.03	.6	1	Reject Linearity
		.9	.7	.06	.5	2	
		.9	.7	.06	.5	3	
		.9	.7	.08	.4	4	
		.8	.6	.09	.4	5	
		.7	.5	.09	.4	6	
		.6	.4	.1	.3	7	

As shown in table 3-5, the Kaplan delta-epsilon test performs very well in extreme cases (such as noise, model 1 (highly nonlinear and chaotic), Chua's (highly nonlinear and chaotic) but has more trouble with intermediate systems which are linear but complex.

III.H.2 Nonlinear time-series analysis

An essential first step in nonlinear time series analysis is the concept of the dimensionality of the data. The underlying idea is that the time-series is a projection to a single dimensional data set from a multidimensional evolutionary pathway. This concept was discussed briefly in the earlier section on rates of change and attractor reconstruction, where the single dimensional data (time-series) was re-embedded into two dimensions to study the evolution of the system. Visualization is generally limited in 2 or 3 dimensions, but in reality, the system may require more dimensions to fully "unfold" the attractor structure. To "unfold" the attractor is essentially to reverse the process of the projection to one dimension. This process is known as embedding the data set in multiple dimensions [68]. Many methods exist for embedding the data (that is recreating the multidimensional data set that was projected into one dimension to create the time-series), but the first step is to estimate how many dimensions are required to achieve this. Two general approaches exist for this: one is to embed the

time-series in multiple dimensions and study the quantity known as the correlation integral. The second relies on the concept of false nearest neighbors.

All methods however, rely on a good estimate of a delay parameter. This is a rough estimate, based on the time-series of the duration, in samples, of one excursion of the system around the multidimensional attractor discussed above. Two methods have been proposed [68] and utilized in the literature to estimate this parameter. The first relies on the well-understood concept of the autocorrelation function, which can be used, and will be used later in this text, to estimate the dissipation of information with time. In theory, the ability to predict the future values of the time-series, at least using linear methods, falls off as the autocorrelation function decreases in magnitude. The delay can be roughly estimated using the point at which the predictability reaches the level of randomness. In this case, data were detrended to remove slow varying estimation biases, and the threshold was set at the autocorrelation function reaching 5% of its maximum, or within 5% of the minimum (in cases where residual estimation ability exist because of long-timescale oscillations). Another approach proposed for estimating this delay is based on the average mutual information. Here, a nonlinear correlation between information (after vector quantization of the data) is used to generate a similar curve to the autocorrelation function with similar implications in terms of dissipation of information and predictability, but in this case, the nonlinearity of the information contents is taken into account. In this case similar criterion for the crossing and delay estimation (5% of maximum, or within 50% of the value of the minimum itself) were used. Estimates of these delays and subsequent embedding

dimensions are shown in table 3-6 for a few of the time-series, using various dimension estimators [69]. Note that noise is theoretically infinite dimensional.

Table 3-6: Lags and various estimators of a sufficient embedding dimension.

	Lag (Autocorrelation)	Lag (AMI)	False Nearest Neighbor's Embedding Dimensions	Embedding dimensions using Cao's Method	Taken's Dimension Estimator
Noise	2	2	None	>5	3.1
Lorenz	67	33	3	3	3.64
Chuas1	20	55	4	6	4.5
Chuas2	32	49	5	6	38

The Lyapunov (also spelled Liapunov in some texts) exponent quantifies the degree of divergence of the system along a trajectory in its state space given a shift in the initial vector specifying the state of the system. This can be stated by the following equation:

$$|\delta Z(t)| = e^{\lambda t} |Z_0|$$

λ is a group of exponents, one for each dimension of the phase space, each which account for the evolution from the initial vector Z_0 of the trajectory in phase space. Thus if λ is less than zero, initial perturbations of along that vector in phase space eventually disappear. Similarly, positive Lyapunov exponents denote instability that is the growth of the initial perturbation along that trajectory. Stable systems, contain a balance of Lyapunov exponents such that their total is negative or zero. Otherwise,

perturbations can exist which lead the system into a continuous growth in phase space and thus instability. In practice, the largest positive Lyapunov exponent is often used as a metric of stability, as having large positive Lyapunov exponents hint at a direction of instability in phase space (though in reality this should be tempered with the knowledge of the full spectrum, which is difficult to calculate). Two different methods were used to calculate the Lyapunov exponent [70], with the results compared in table 3-7. While the numbers are different, their relative magnitudes are fairly similar for these data sets.

Table 3-7: Two different methods used to calculate the largest positive Lyapunov exponent.

	Largest Lyapunov Exponent (Max Estimate)	Largest Lyapunov Exponent (Mean Estimate)	Largest Lyapunov Exponent (Max Estimate) Method 2	Largest Lyapunov Exponent (Mean Estimate) Method 2
Noise	1.1	.16	1.08	.46
Lorenz	.03	.015	.06	.034
Chuas1	.11	.054	.072	.048
Chuas2	.087	.034	.069	.034

III.I Meal models and process models

Meals, as shall be discussed, represent the most omnipresent event in the time-series of nondiabetic time series. The meal event will be analyzed in part by computing the signal energy, which is the signal amplitude squared (with the mean removed) during meal events and non-meal events. In the time-series where meal

times are known, the meal event is defined by the two hours period immediately following a meal perturbation (figure 3-19). The signal energy during these periods is summed and compared to the signal energy during the other periods of the time-series.

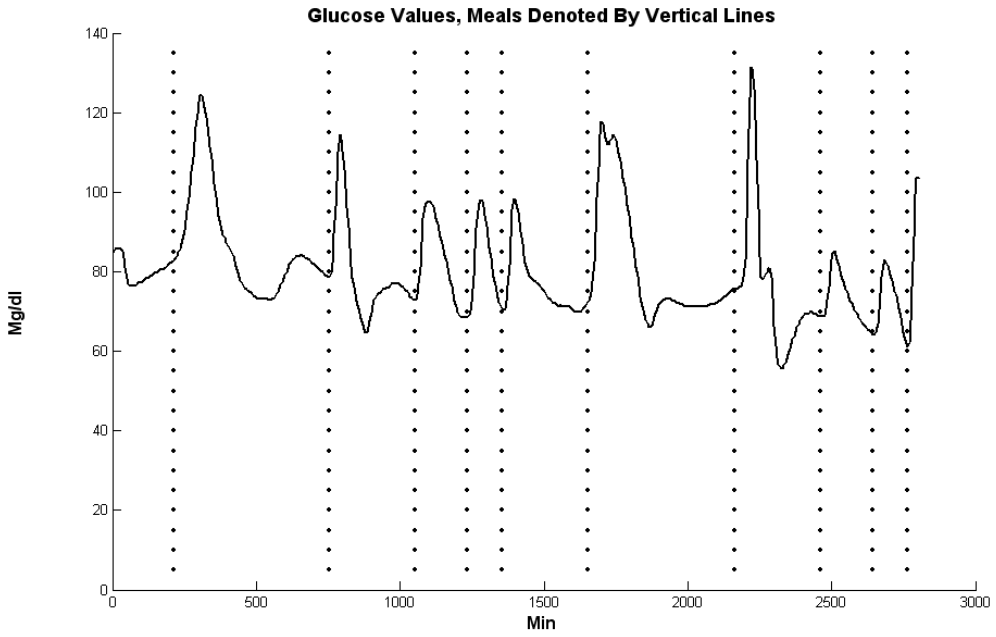


Figure 3-19: An example of a glucose time-series with meals denoted by vertical lines marked by dots. In this example, this represents dynamics from a nondiabetic.

Additionally, meals can be averaged to yield a meal profile. The only significant difficulty comes from averaging meals from time-series, which are sampled differently, which in this analysis are avoided. Another small problem arises because the meals are of different lengths (some meal profiles are interrupted by another meal after which the meal profile is cut short). Thus different parts of the meal profile (after the first two hours) use different number of meals in the average. This can lead to spike-like discontinuities which though small can be misinterpreted as events. To lessen this effect, discontinuities in the averages are shifted to create a smooth meal

profile. The meal profile can then be computed for an individual, across individuals at the same meal and between populations. The rate of change of the glucose values during the meal event can be used to view the dissipation of the meal disturbance over time.

Various metrics of the meal response can be then quantified. These include the time to maxima, maximum rates of rise and fall and other metrics. Because there are many of them, they will be discussed in the section with the results, along with the method of meal detection. Automated detection of the meal event, while not strictly a method of analysis, highlights both the potential in utilizing dynamics and the singular nature of the meal event which lends it to detection without significant effort.

Going beyond the whole meal event, one can look at the processes underlying the events themselves: the rise of glucose values and the fall. Looking at these two different processes is one of the few places where physiologic insight enters the analysis: the knowledge that the rise and fall of glucose values are governed by interjections from two different underlying physiologic mechanisms. This is true both in the case of meals (gut absorption versus beta cell response), fasting (liver generated glucose versus beta cell response) and continuous input into the system either through the gut or direct IV injection. The breaking of meals into process models is the simple process of dividing the meal event into the rise and subsequent fall. In the case of the other time-series were events are not clearly denoted (fasting), events are divided into

rises and falls based on snippets of the time-series where there is a sustained rise or fall in the glucose values (figure 3-20).

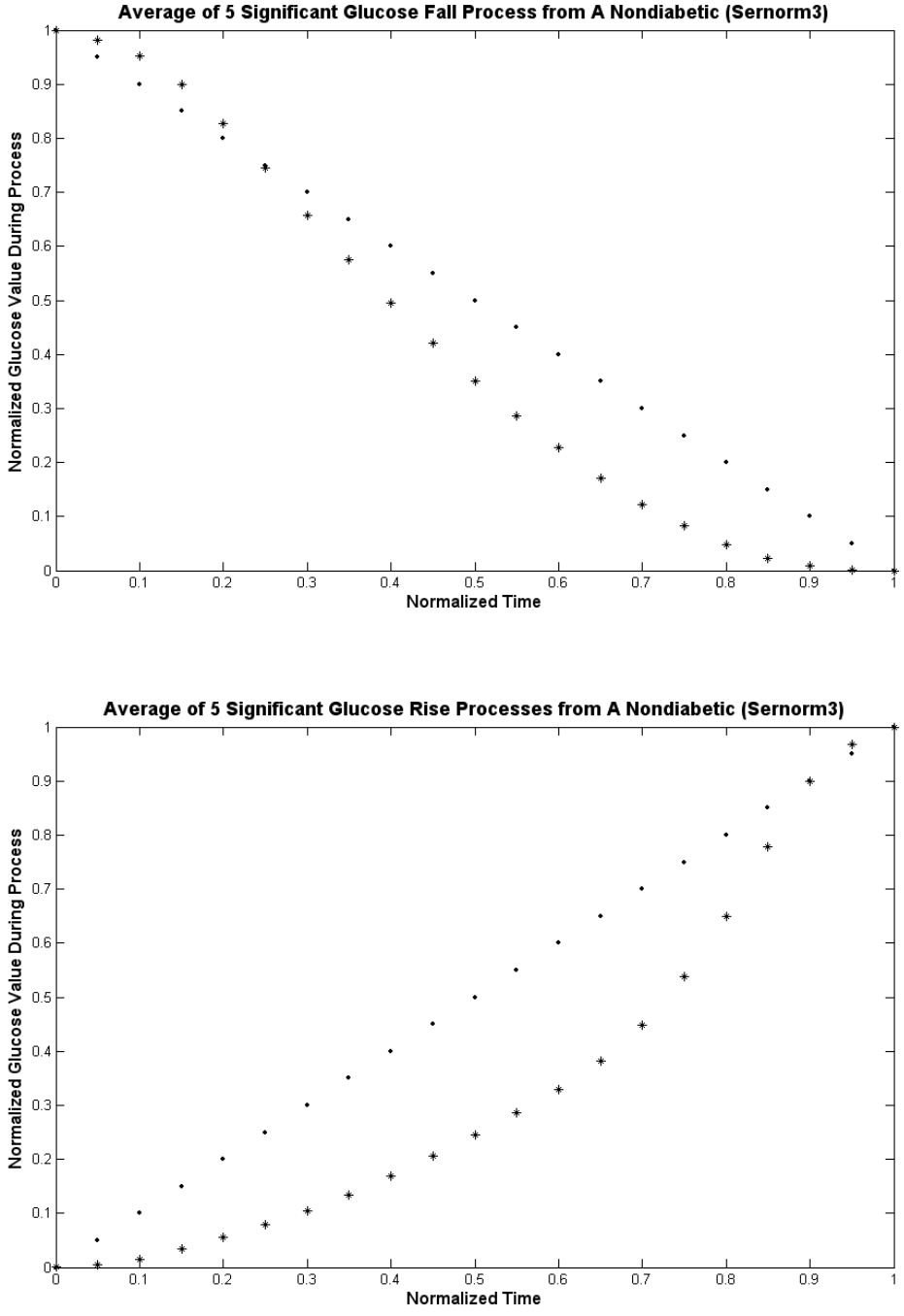


Figure 3-20: Two meal process models from averaged from 5 meals in a nondiabetic. The process is shown by stars. The line consisting of circles is a simple line provided for comparison.

III.J Sampling requirements

The methods for degrading sampling are twofold: one involves resampling the signal at a lower sampling rate, while the other simply drops samples. The former has the advantage of generating more continuous study of sampling rate while the latter is confined to successive halving of the sampling rate. However, it is impossible to resample the signal at lower frequencies without making fundamental assumptions about the underlying signal structure. This is caused by the need to fit successive data points in order to interpolate what a supposed sample would be between them. In order to minimize the complexity of the assumption in signal structure, simple linear interpolation and/or zero-order hold (where the interpolated sample is simply the previous sample) are employed, but although while simple, these also contain assumptions about signal structure. Thus when possible, both methods (interpolation and halving of the sampling rate) are utilized in under-sampling the signal.

Once the signal is undersampled, simulating the condition where fewer samples are collected from the same system, various metrics can be deployed to study the effect on degradation of dynamic measurements. Because the rate of change is the simplest expression of the underlying dynamics that is not simply a function of signal statistics, the effect of undersampling on the estimate of the signal derivative is used as a first approach. This is presumably the single most important expression of the signal dynamic information contents and thus subsequently effects most other measurements indirectly.

Specific examples of performance degradation are explored but not simply in the context of error but rather in the degradation of the utility of the measure. This type of analysis complements, rather than replaces, the previous work measuring the Nyquist sampling requirement, which is the most commonly used and global measurement of sampling requirement.

Chapter IV: Nondiabetics

In the introduction, the importance of understanding normal physiology to the long-term development of therapies for metabolic dysfunction was described. Briefly, the study of nondiabetic subjects not only enhances our understanding of the underlying physiology, but also can be a key in ingredients in enhanced diagnosis and monitoring of patients. Perhaps most importantly, it also allows for a more complete set of parameters that describe normoglycemia as well as setting more physiologically precise control targets. These insights can be applied regardless of the methods of therapy employed in the future, which may include modalities such as beta cells, electronics or pharmaceuticals. Significantly larger number of data sets representing individuals from a diversity of genetic and environmental backgrounds is necessary to come to a consensus definition of normoglycemia, if in fact such a consensus can be reached. But the methods described and their applications to the limited data may suggest possible avenues to pursue in the path towards defining normoglycemia.

IV.A Rates of change

As mentioned in the methods section, the rate of change is the simplest byproduct arising from the dynamic analysis. It has implications for sensors (in terms of the effects of sensor lag and sampling), controllers (in terms of their speed of response) as well as being a potential clinical tool that captures the simplest aspect of dynamics. Rates of change can be most simply analyzed by looking at the distribution of these rates of change (that is the statistics of rates of change). Therefore, in this section the statistical distribution of the rates of change and in particular the maximum rates of change in the glucose signal are studied. A 2D histogram comparing blood glucose values and the rates of change are then used, a tool developed specifically for this application as a way to quantify estimated phase space projection geometry. These methods were discussed in chapter three of this document.

IV.A.1 Nondiabetics eating meals and exercising

Three nondiabetic subjects were sampled every 5 minutes for up to 48 hours during which they ate 10 meals and were forced to exercise. These three time-series represent the most clinically relevant data set because they may most resemble a typical nondiabetic blood glucose series. Additionally, 4 other data sets were multiple meals were administered are available, but suffer from more idealized patient situations (non-mobile) and also infrequent (~30 minutes) and/or non-uniform (in the case of the DirecNet dataset) sampling. Knowing that the latter data sets are less than

ideal, they are treated separately but are included because they add a large number of subjects to the analysis and can serve as confirmatory observations.

For the sernorm1-3 data set, the maximal rates of change were slightly asymmetric depending on the direction, but were between 1.5-3 mg/dl per minute. The figure that follows shows a single distribution and then the combined distribution for all three individuals in the group (figure 4-1).

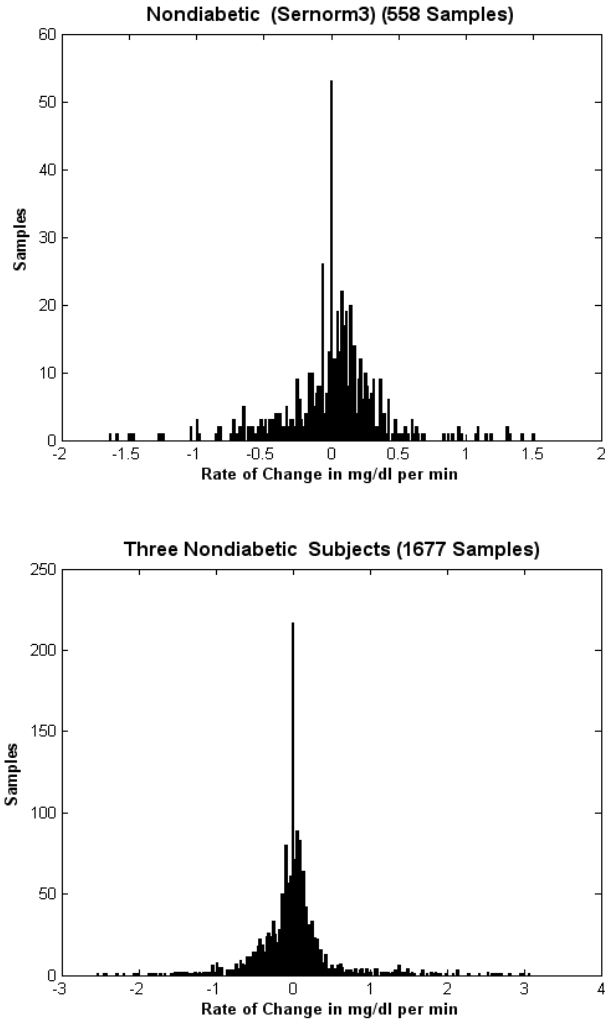


Figure 4-1: Top: Sample Distribution from a single nondiabetic. Bottom, distribution from three nondiabetics combined.

A characteristic asymmetry is noted in the distribution, specifically in that the negative derivatives seem to have a wider distribution close to the center of the histogram, but both sides do contain a number of outliers which seem evenly distributed. Table 4-1 summarizes this analysis.

Table 4-1: Statistical Analysis of rates of change in three nondiabetic subjects. All non dimensionless results are expressed in mg/dl per minute.

Patient ID	Max (-) Rate	Max (+) Rate	Mean Rate	Skewness	Kurtosis
Sernorm1	-2.6	2.9	.39	.9	7.6
Sernorm2	-2.5	3.1	.28	1.3	12.7
Sernorm3	-1.6	1.5	.25	-.4	6.9

In the case of the vannorm6-9 datasets, which contain far fewer samples, similar distributions and numerical results are also noted. The distributions and observed statistics are shown in figure 4-2 and table 4-2.

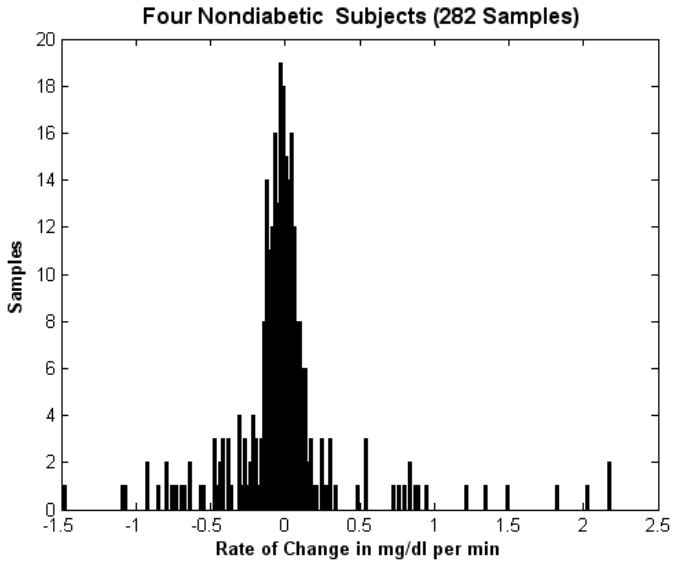


Figure 4-2: Distribution of rates of change for two patients with two different days of testing where the number of meals was varied (4 data sets total). A similar pattern is seen in this set to the previous one.

Table 4-2: Statistical analysis of rates of change for two patients on two different days with varying meal timing. All non dimensionless results are expressed in mg/dl per minute.

Tracing ID	Max (-) Rate	Max (+) Rate	Mean Rate	Skewness	Kurtosis
Vannorm6	-.7	.9	.15	.9	7.2
Vannorm7	-.7	1.2	.17	1.6	9.4
Vannorm8	-1.1	2.0	.2	2.4	13.8
Vannorm9	-1.5	2.2	.33	1.4	7.6

Unfortunately, the relative lack of data points in the remaining data sets significantly reduces the accuracy of derivative estimation and makes statistical analysis less meaningful, particularly in the case of higher order statistics. To make use of the data, the derivative estimates from the three data sets are pooled into one

large data set containing 31 different individuals (mal1-12, mej1-5 and DirecNet set). The summary of these data are shown in figure 4-3 and table 4-3.

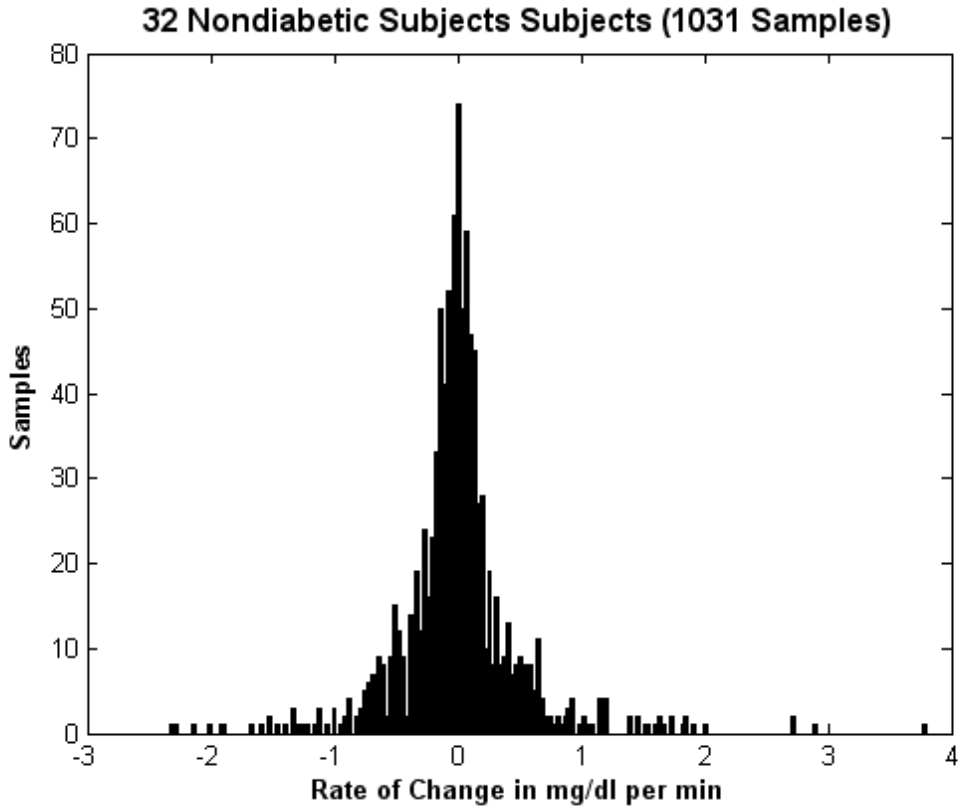


Figure 4-3: The distribution of the rates of change from 31 Subjects sampled ~30 minutes with multiple meals.

Table 4-3: Statistical measures of rates of change from 31 subjects from three different studies.

Measure	Max (-) Rate	Max (+) Rate	Mean Rate	Skewness	Kurtosis
Mal1-12 (356 Samples)	-2.3	3.8	.02	.9	9.3
Mej1-5 (239 Samples)	-1.6	1.7	0	.9	7.5
Direcnet (436 Samples)	-2.13	2.0	-.01	0	9.3
Combined (Above)	-2.3	3.8	0	.86	11.6
All 39 data set combined (1871 Samples)	-2.5	3.8	0	.9	11

It is important, in interpreting these results to consider the fact that these rates of change are likely to be linked significantly to the diet status of the individual, which may account for differences between subgroups. Additionally, in the last group where sampling was less than ideal, estimates of the derivative are likely to underestimate maximal rates of change. From this analysis the following observations are made for nondiabetics:

1) Non diabetic maximal rates of change range between -.7 to -2.6 on drop and .9 to 3.1 on the rise
2) The distribution of the rates of change from the seven "ideal" data sets containing meals and activities contain a positive skewness (6 out of 7)
3) The maximal rate of rise in 39 data sets was 3.8 mg/dl*min and the maximum rate of fall was -2.5 mg/dl*min. Skewness remained positive or zero for various groups of data combined.
4) The rates of change in a nondiabetic are below 1 mg/dl per min during most samples obtained, and rarely exceed 2 mg/ml per min in either down or up directions.
5) In most cases, the maximum positive rate of change is larger than the maximum rate of change in the negative direction.
6) The distributions are slightly asymmetric with a very long "tail"and a significant concentration of samples near zero.

The asymmetry in the distribution is likely to be a result of the fact that the process of ingestion and absorption are physiologically different than the process of

insulin production and glucose uptake by the cells. That is to say, the distribution under study is fundamentally a result of two very different processes. This perhaps makes the relative lack asymmetry surprising. There also appears to be limits on the rates of change in both processes. The long tail in the distribution is largely due to the meal process, during which the dynamics of the system become more pronounced.

After considering the rates of change as a single variable, one can combine it with the value at which the rate of change was computed. Plotted against each other, as discussed in the methods section, this phase space representation attempts to give a two dimensional representation of the relationship between two important variables in the system, the glucose levels and the rates of change. This object is often called an "attractor" because of its tendency in dissipative systems to be confined in a limited region, although strictly speaking this term is much more specific to certain dynamic behaviors. Here one can use a 2D histogram to quantify the distribution of points in the attractor geometry. The plot for the five data sets is shown in figure 4-4, 4-5, 4-6, 4-7 and 4-8.

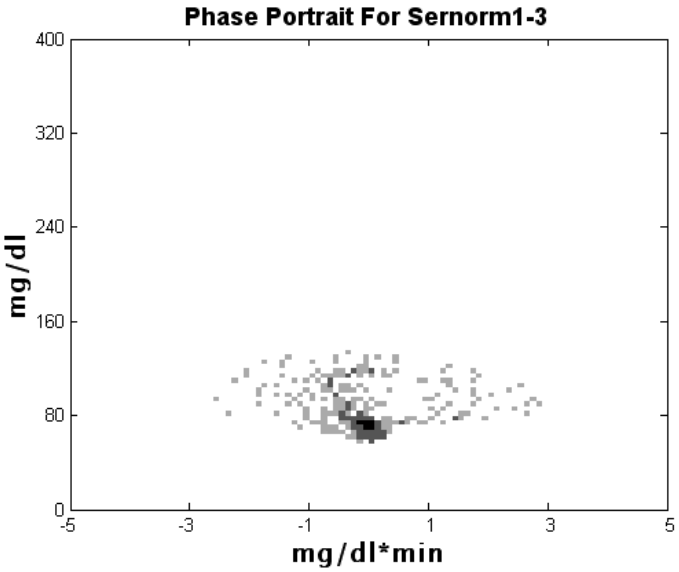


Figure 4-4: Two Dimensional Histogram representation of the rates of change and the values at which they are computed. Darker squares indicate that a higher number of samples fall into that location.

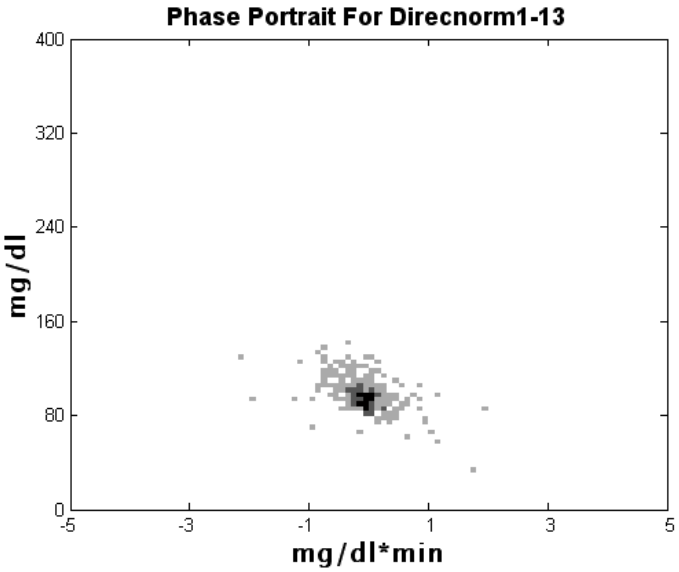


Figure 4-5: Two Dimensional Histogram representation of the rates of change and the values at which they are computed. Darker squares indicate that a higher number of samples fall into that location.

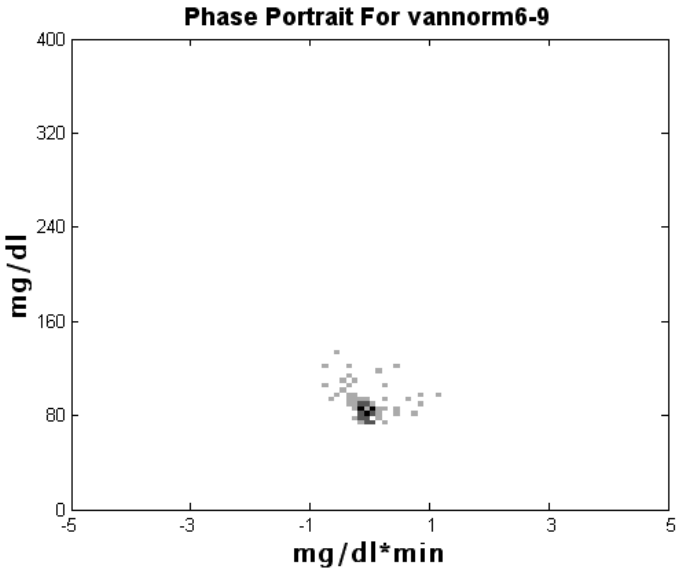


Figure 4-6: Two Dimensional Histogram representation of the rates of change and the values at which they are computed. Darker squares indicate that a higher number of samples fall into that location.

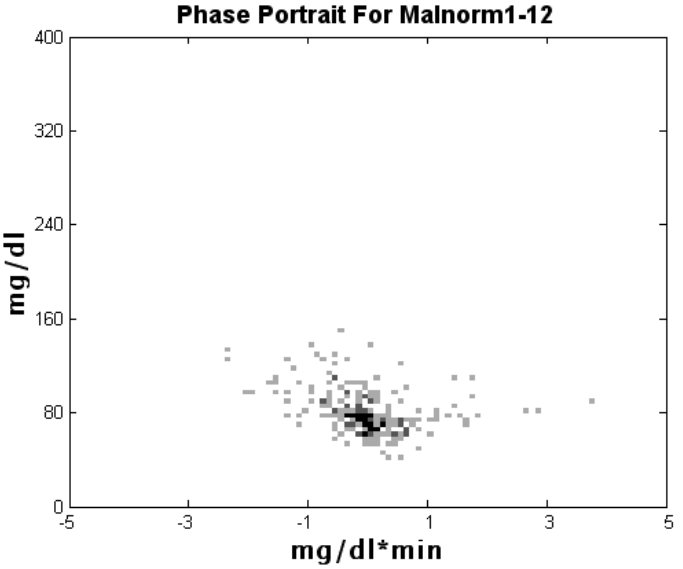


Figure 4-7: Two Dimensional Histogram representation of the rates of change and the values at which they are computed. Darker squares indicate that a higher number of samples fall into that location.

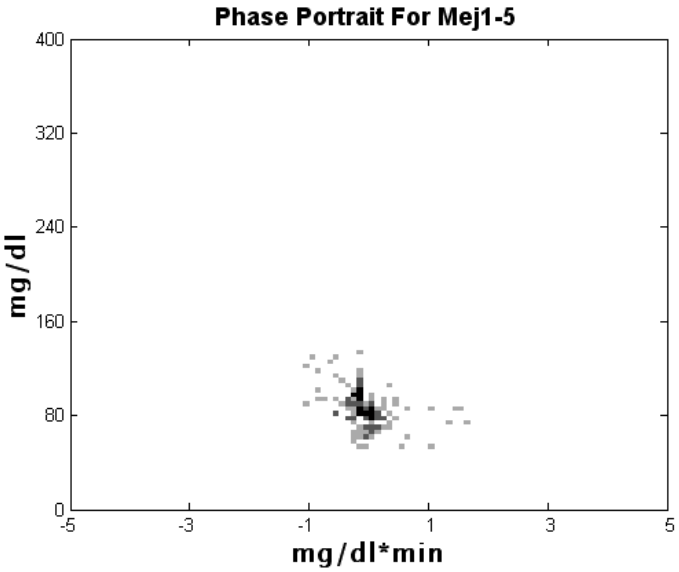


Figure 4-8: Two Dimensional Histogram representation of the rates of change and the values at which they are computed. Darker squares indicate that a higher number of samples fall into that location.

Looking at these plots, much of the insight gained by looking at the histogram in the previous section is confirmed: the distributions are highly concentrated but contain a significant number of outliers. They are also not symmetric. The asymmetry (about the zero rate of change axis) is additionally dependent on the value of glucose. As would be expected, high rates of positive change occur more often at lower glucose values and higher rates of drop occur more often at higher glucose values. This makes sense from control standpoint were it is expected that physiologic mechanisms would counter high rates of change at both ends of the control target region when these would act to move the system further away from control.

Quantifying the geometry of the attractor, metrics discussed in the methods section will be utilized. The size of the attractor (in number of bins), the compactness (the number of bins accounting for 90% of the samples, and the attractor radial symmetry calculated by sampling the distribution along radial lines from the center of the attractor. These tests were once again applied to each of the ideal data sets and the collective subsets of less than ideal data, shown in table 4-4 and 4-5.

Table 4-4: Attractor geometrical descriptions for the ideal data sets with multiple meals.

Data Set	Attractor Area	Compactness	Symmetry
Sernorm1	665	1.8%	.6
Sernorm2	683	2.8%	.7
Sernorm3	309	7.8%	.9
Vannorm6	126	3%	.7
Vannorm7	136	4%	.6
Vannorm8	263	2%	.4
Vannorm9	400	1%	.4
Average	369	3.2%	.6

Table 4-5: Attractor geometrical descriptions, applied to pooled data from less than ideal data sets containing multiple meals.

Data Set	Attractor Area	Compactness	Symmetry
Sernorm1-3	743	4.4%	1.0
Vannorm6-9	513	3.3%	.6
Mej1-5	449	4.4%	.9
Mal1-12	671	3.4%	.7
Direcnet	564	4.1%	.6

A few preliminary considerations must be applied before making observations on these results. Both the compactness and area are functions of the bin size and thus these numbers are only truly useful in terms of comparisons using exact same

algorithmic applications. Thus the true test for the utility of these measurements is in comparison to the other groups discussed in the coming chapters. Having set these factors as constant, it is observed that 90% of the samples fall within 2%-5% of the attractor area. That is to say the system spends majority of its time inside a small area within its attractor. This is not surprising, however, it allows us to quantify what is known intuitively about this system: the majority of the time is spent in a confined glycemic boundary. The symmetry measurement used appears to vary widely between datasets and thus is likely to be a poor differentiating tool and requires further study.

IV.A.2 Continuous feeding, infusion and fasting

Continuous enteral feeding removes the meal perturbation as a large factor in the dynamics of blood glucose control. IV infusion removes most of the influence of the GI system on the glucose values altogether. Looking at the rates of change (figure 4-9, 4-10, table 4-6) a much more compact distribution is noted where a range of approximately -2 to 2 is observed (in comparison to the normal feeding range of -2.3 to 3.8 (table 4-3 and figure 4-3). Notably, despite the fact that meals are removed from set of perturbations, the rates of change, and in particular the maximal rates of fall are not as different as one might have expected.

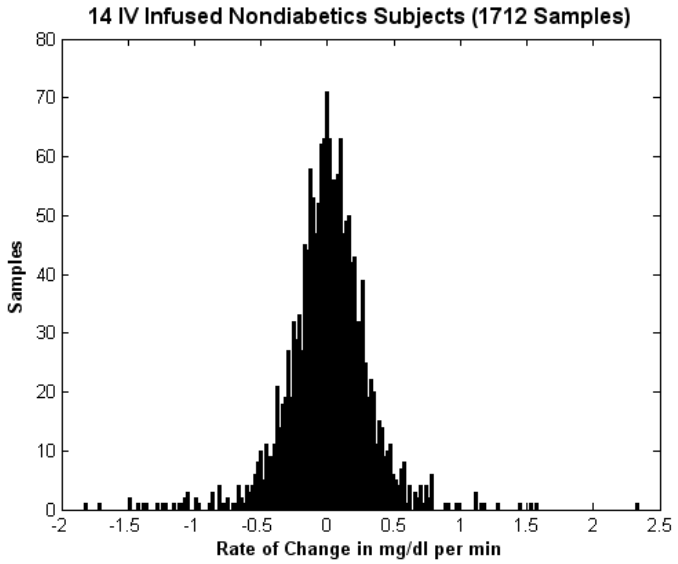


Figure 4-9: Distribution of rates of change in continuously infused nondiabetics. Note the relative symmetry in the distribution. Also note that despite the fact that the large perturbations from meals are absent, occasional fast rates of fall and rise are still nonetheless present.

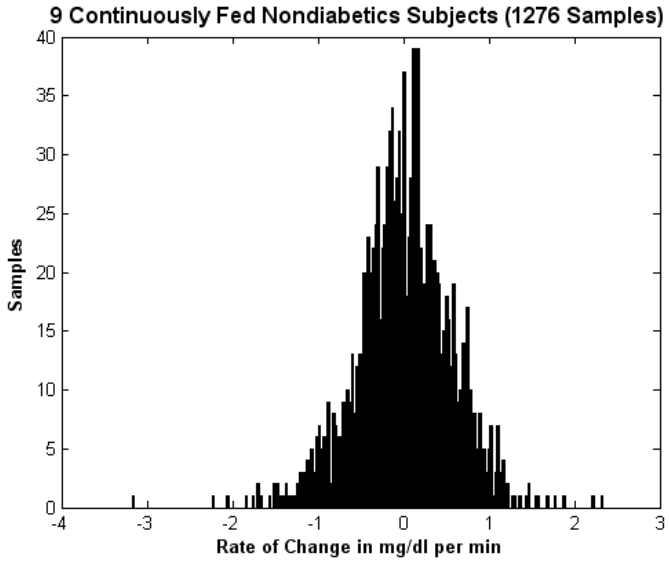


Figure 4-10: Distribution of rates of change in 9 nondiabetics with continuous enteral nutrition.

Table 4-6: Summary statistics for continuous nutrition administered through two different routes at differing rates, as well as fasting nondiabetics.

Measure	Max (-) Rate	Max (+) Rate	Mean Rate	Skewness	Kurtosis
IV Infused 13 Sets 1569 Samples	-1.3	1.1	0	-.3	4.9
Enteral Feeding 9 Sets 1276 Samples	-3.2	2.3	0	-.2	4.6
Fasting 3 sets 282 Samples	-.4	.5	0	.6	5.1

Furthermore, the observation can be made that the skewness for the infusion/enteral feeding sets are now negative whereas in the fasting state the distribution remains positively skewed for rates of change. This is markedly different than the data sets containing meal and exercise responses. Looking at the phase space geometry (figure 4-11, 4-12 and table 4-7) it is noted that the compactness of the attractor has remained fairly high with most samples falling within 7% of the attractor size, despite containing a significant amount of data.

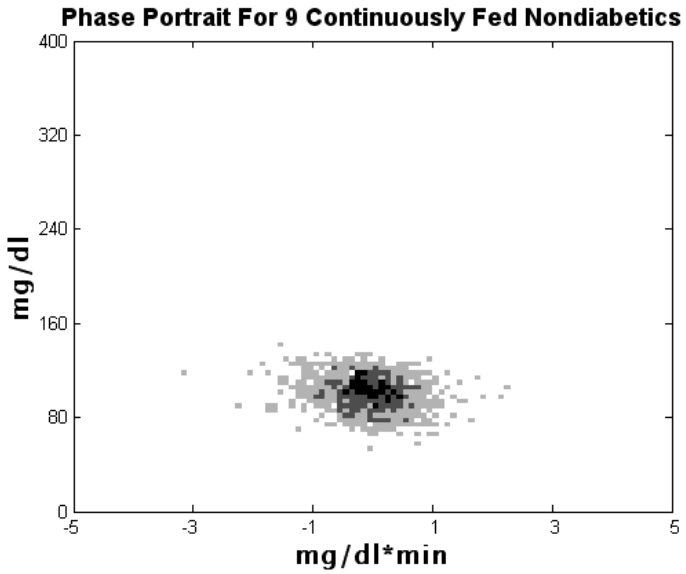


Figure 4-11: attractor portrait for 9 nondiabetics with continuous enteral feeding.

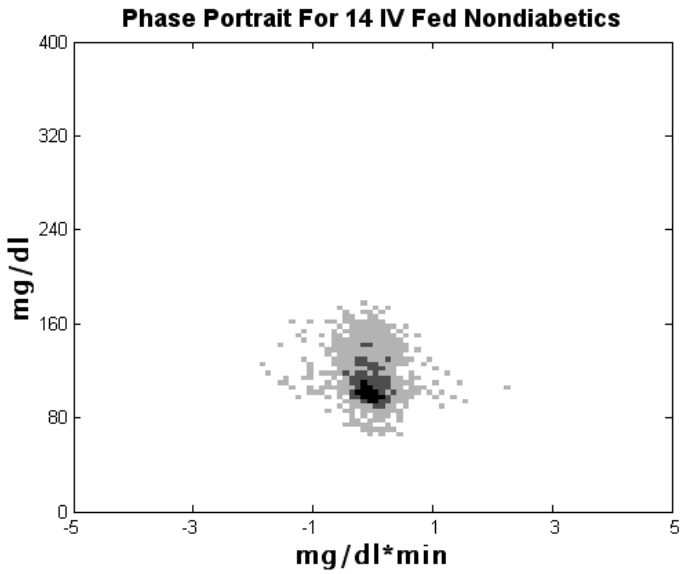


Figure 4-12: attractor portrait for 14 IV fed nondiabetics.

Table 4-7: Relative comparison of attractor geometry between two different modes of continuous nutrition delivery.

Data Set	Attractor Area	Compactness	Symmetry
Continuous Enteral Feeding	706	6%	.8
IV Feeding	585	7%	.5

IV.B Time-scales of dynamics

Time scales define the expected time over which significant changes or repetitions are observed. In the simplest case of a sinusoid, the signal can be defined by two scales, amplitude and the frequency, the latter representing the time-scale of the evolution of the signal. As signals become more complex, it is harder to divorce these two scales and analyzing one leads to insights about the other. Nonetheless, in the previous section the rates of change of the signal and their relationship to the amplitude of the signal were analyzed. In this section, the time-scale, an in specific the time-scales of repetition, as analyzed in the frequency domain will be studied. Because of the awareness in the literature of multiple time-scales, and the realization that nonlinear dynamics cannot be ruled out at this point, multiple tools are utilized to analyze the signal in an attempt to gain insight into the time-scales that are evident, under different perturbations to the system.

IV.B.1 Why is time-scales of dynamics important?

The time-scales over which the signal evolves, is the most fundamental constituent of the understanding of the signal because it dictates sampling frequency and duration of observation necessary to characterize the system. Additionally, it provides insight into possible mechanisms underlying the dynamics of the system. For example, hormonal cycles with periods of 24 hours can often be connected to rhythmic variations tied to other physiologic variables (circadian rhythms). Significantly different time-scales of dynamics within a signal may also suggest multiple quasi-independent mechanisms generating and contributing to the signal. As mentioned circadian rhythms in the glucose signal may have a fundamentally different origin than the dynamics associated with pulsatile insulin release. Attempting to model both processes using the same model type, let alone the same model, may be suboptimal as the underlying generating process may be different.

Most importantly, understanding fundamental time-scales of dynamics, affect the approach and validity of any time-series analysis technique. For example, various tests require that the time-series being analyzed include a complete or multiple complete evolutions of the system through its prescribed state space. This makes intuitive sense: in order to characterize the behavior of the system, one needs to observe it for long enough in order to be able to categorize the various behaviors the system is capable of. This question of “long enough” is critical to the approach of dynamical system analysis both in terms of how frequently the system has to be sampled but also how long it needs to be observed to characterize certain behaviors.

The distribution of signal energy in the various bands of time-scales has implication for analysis, filter design as well as estimation/control algorithms. In this section the time-scales of change based on available signals from nondiabetics will be studied and comment on the implications of these observations based on the limited data will be made.

IV.B.2 Very fast time-scales

Dynamics on the order 5-15 minutes have been reported in the way in which beta-cells secrete insulin ([71]). These oscillations in insulin release have been the subject of intense study, and may be reflected in the glucose signal. Most data sets in the literature and in this study are not sampled frequently enough to capture such perturbations. In order to ascertain possible contribution of very high frequency signal dynamics, data sets which are sampled at 2 or 4 minute intervals were studied for their high-frequency content. Two differing methods were utilized to study the relative contribution of the high-frequency band to the overall signal. One arose from the computation of the spectral estimate using FFT algorithm. The other arose from high-pass filtering of the signal so as to isolate these components. In both cases the time-scale considered to be high-frequency was that with period shorter than 30 minutes. In the case of spectral method, all energies with a period of greater than 30 minutes were summed. The 10^{th} order high pass Butterworth filter was designed with corner period of 30 minutes. The results are shown in table 4-8 and 4-9.

Table 4-8. Contribution of High Frequency Oscillation Component to Total Blood Glucose Signal energy (mean and linear trend removed).

Reference	Sampling Period, min	Conditions	Signal energy in < 30 min Component	Total Signal energy	% of Total Signal energy
Polonsky et al [45]	2	Fasting	120	700	16.8%
Simon et al [31]	2	Continuous Infusion	500	20000	2.5%
Simon et al [23]	4	Meal Challenge	290	7240	4.0%
#2	4	Meal Challenge	120	5880	2.0%
#3	4	Meal Challenge	850	22300	3.8%
#4	4	Meal Challenge	820	24100	3.4%
#5	4	Meal Challenge	270	27500	1.0%
#6	4	Meal Challenge	920	32700	2.8
#7	4	Meal Challenge	160	43400	0.4%
#8	4	Meal Challenge	430	94500	0.5%

Table 4-9: Percentage of signal energy in the very high frequency component of the signal (short time-scale). Notice that in the first two signals, the nature of the perturbation to the system (fasting and continuous infusion) exclude the high signal energy meal response and thus the high frequency components account for a larger percentage of the signal energy.

Reference	Sampling Period, min	Conditions	Total Signal energy in High Pass Signal	Total Signal energy	% of Total Signal energy
Polonsky et al [45]	2	Fasting	180	700	26%
Simon et al [31]	2	Continuous Infusion	2600	20100	13%
Simon et al [23]	4	Meal Challenge	1200	7200	16%
#2	4	Meal Challenge	1100	5900	19%
#3	4	Meal Challenge	2400	22000	11%
#4	4	Meal Challenge	1200	24000	5%
#5	4	Meal Challenge	2400	27000	9%
#6	4	Meal Challenge	2100	33000	6%
#7	4	Meal Challenge	1000	43000	2%
#8	4	Meal Challenge	2100	94000	2%

Note that total signal energy is dependent on the number of samples in the data set, and thus a percentage was used to study the relative signal energy content. Based on these results, it can be concluded that the very high frequency oscillations constitute a very small portion of the signal energy content, particularly in the context of the presence of larger perturbations such as meals and infusions. The average signal energy of high-pass signal, in the units of mg^2/dl^2 is given in the table 4-10.

Table 4-10: The average signal energy in the very high frequency component of the glucose signal. Units of signal energy, in this context are in mg^2/dl^2.

Reference	Sampling Period, min	Average Signal energy in High Pass Signal
Polonsky et al [45]	2	.75
Simon et al [31]	2	11
Simon et al [23]	4	9.4
#2	4	8.7
#3	4	19
#4	4	9.5
#5	4	13
#6	4	10
#7	4	5.3
#8	4	10

The implications of the very-high frequency component will be revisited in subsequent chapters in terms of loss of signal energy due to lower sampling rates.

IV.B.3 Normal feeding

Seven ideal data sets were considered for the analysis of the frequency content 3 of these data sets were sampled every five minutes for durations of nearly 50 hours and 4 were sampled every 20 minutes and were followed for 24 hours. Spectral Analysis was undertaken for these data sets using the three techniques discussed in the methods sections.

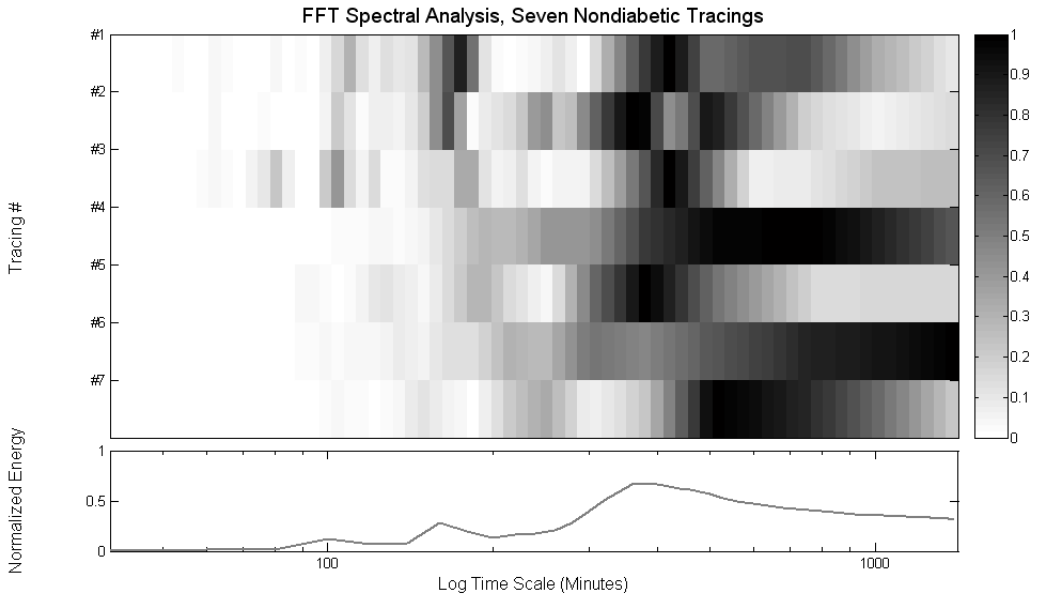


Figure 4-13: FFT Spectral Estimates from seven nondiabetic subjects. A Kaiser window with a parameter of 8.0 was used.

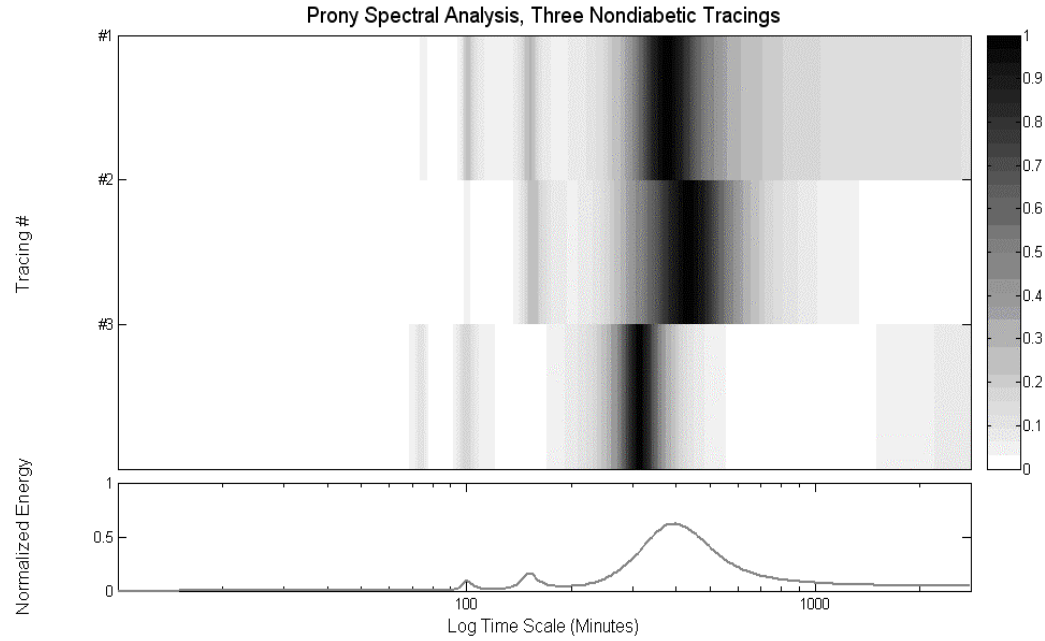


Figure 4-14: Prony spectral estimate using a Prony estimator of order 60 (300 minutes) for three nondiabetic subjects sampled every 5 minutes for 48 hours.

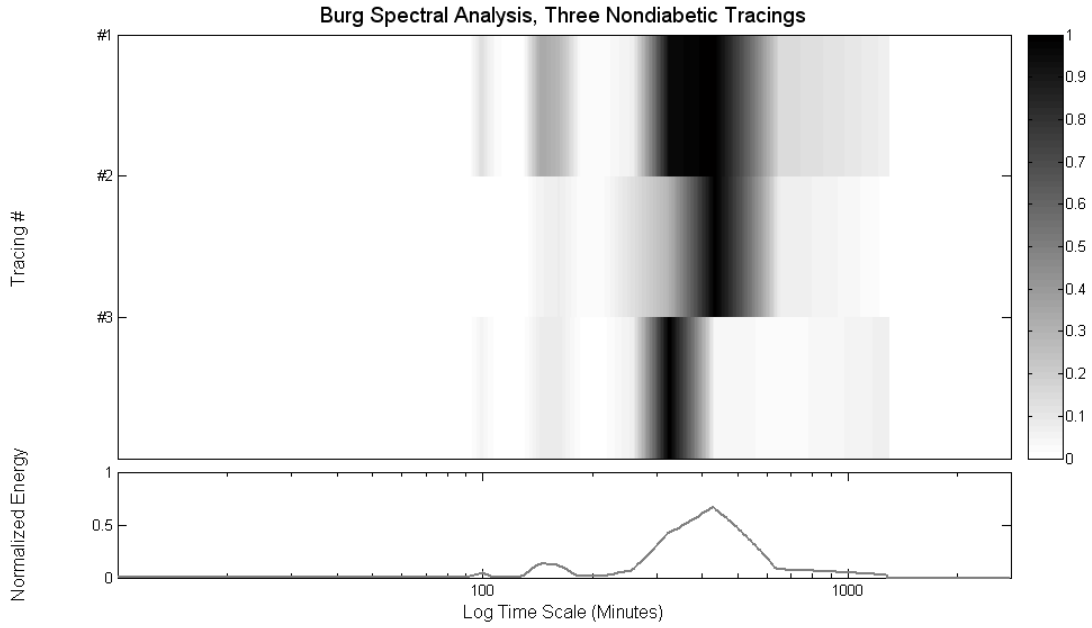


Figure 4-15: Spectral estimate using Burg's parametric method with an order of 80 (400 minutes) of three nondiabetic subjects sampled every 5 minutes for 48 hours.

Analysis of the frequency content of the best sampled sets shows three distinct peaks (Figures 4-13, 4-14, 4-15 and table 4-11). The smallest peak, near 100 minutes as well as the peak near 120-150 minutes, may represent intrinsic oscillations having similar in nature to the ones observed in the case of continuous nutrition. These are likely to be the result of glucose production and insulin production cycles not having to do with meals, which are seen also in fasting. Another possible source for these peaks is the meal pulse length, which is on the order of ~120 minutes. This is likely to contribute to the second peak, which is located at that range. The largest peak, however, occurs at a period of 5-7 hours. This peak is very likely to reflect the spacing

of the meals, most of which are spaced 4-8 hours apart. In essence, the frequency study of the time-series is capturing the timing and spacing of the meals, in part because the size of the meal pulse is relatively short in the nondiabetics, making the meal event (which is the signal energy containing event as shall be discussed next) appear like narrow pulse and thus the spacing between the meals a large component of the frequency estimate.

Table 4-11: Summary of peak spectral estimates using three methods for estimating the frequency spectrum, in three nondiabetic subjects sampled every 5 minutes for 48 hours.

Data Set	FFT Estimate	Prony Estimate	Burg Estimate
Sernorm1	400	370	320
Sernorm2	350	430	430
Sernorm3	400	310	320

Looking at the signal energy distribution in the signal (recalling that signal energy is the square of the amplitude with the mean removed in this case), the meal period is defined to be the 2 hours subsequent to the commencing of the ingestion of the meal. Looking at the signal energy during the meal period versus the non-meal period, one can clearly see the meal period accounts for a significant portion of the signal energy. One can quantify this with the ratio of the energy accounted for by the meal periods to the time accounted for by meal periods (table 4-12). This ratio appears to be significantly greater than one, ranging from 1.8 to 4.9, which one summarize by stating that the meal periods contain about 2-5 times as much energy as would be expected given their duration.

Table 4-12: Percentage of energy during mealtimes computed for nondiabetics. In the second column, the percentage of time which is considered "meal time" is calculated by considering mealtime as the two hour interval after each meal event. The percent of signal energy (defined by the mean subtracted amplitude squared) during this period is shown in the third column, and the same quantity is shown in the fourth column for the nonmeal times. The last column shows a ration of the former versus the latter. Note that the % signal energy in the meals is systematically larger than the % time of the meals.

Data Set	% Time Meals (2 hr horizon)	% Signal energy Meals	% Signal energy Not Meals	Ratio (Meal Vs. Nonmeal)
Sernorm1	42%	77%	23%	3.3
Sernorm2	40%	69%	31%	2.2
Sernorm3	40%	71%	29%	2.5
Vannorm1	18%	79%	21%	3.7
Vannorm2	29%	78%	22%	3.5
Vannorm3	18%	82%	18%	4.5
Vannorm4	27%	83%	17%	4.9
Mejnorm1	51%	69%	31%	2.3
Mejnorm2	51%	65%	35%	1.8
Mejnorm3	63%	75%	25%	3.0
Mejnorm4	40%	77%	23%	3.0
Mejnorm5	47%	80%	20%	3.9

Looking at the evolution of the various components of the signal, one can turn to the time-frequency methodologies described in the methods section. By this point it is expected that on some time-scale, the signal will have varying frequency components because the largest component of the signal is the meal event which itself can vary in duration timing and content. A short Fourier transform map of a sample meal is shown in figure 4-16. Note that time encompasses a two-day period. The solid black squares represent centers of the frequency distribution as discussed in the methods section.

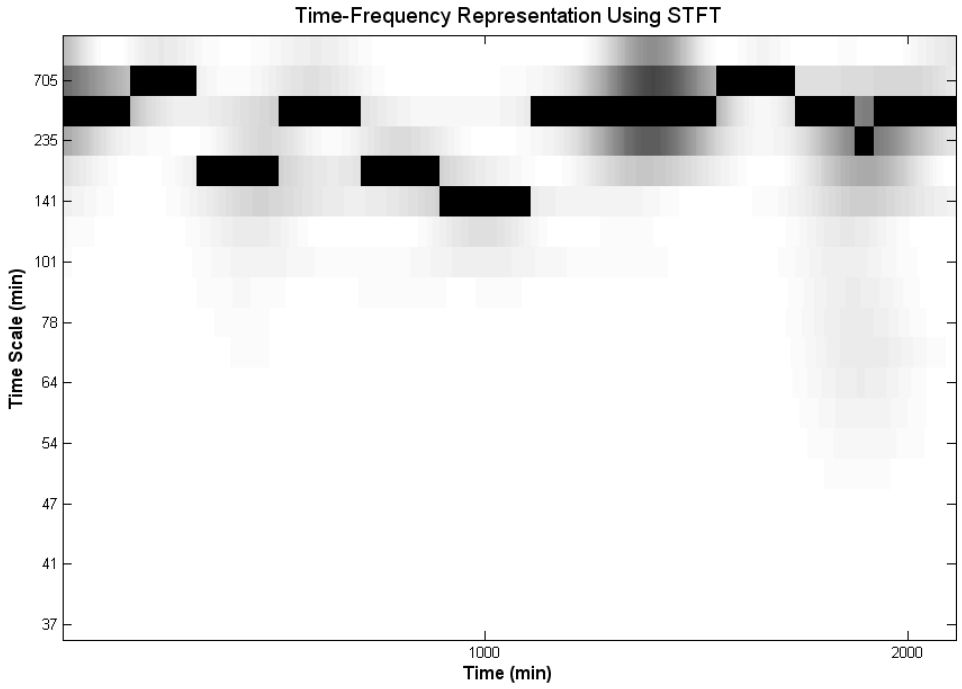


Figure 4-16: An example of a time-frequency representation of 48 hours of data from a nondiabetic. Note the changes in frequency content having to do with the meal events.

In order to facilitate analysis, and because most of the signal energy and thus changes in signal energy distribution are meal related, time-frequency analysis was performed on the three longest best sampled data sets. As mentioned in the methods section, two wavelet basis were chosen because of the different properties they exhibit. The wavelet analysis shows a wide variety of frequency components and great variations. This in part leads to the conclusion that significant frequency content variation does exist in the time-series (table 4-13). However, as mentioned before, this suffers from the fact that the time-series contains relatively large time-scale components which are not sampled often (as the time-series is short in duration). This

does, however, point to the requirement for further analysis of longer time-series to characterize exact nature of such variations and whether they can all be accounted for by differences in the timing and size of meals and activities.

Table 4-13: Summary of time-frequency observations on nondiabetics consuming 10 meals over a two day period. Note that significant deviation exists in the detected time-scales of the centers of signal energy concentration. This is because of the differences between different basis ability to focus on a certain frequency components of the signal.

Patient	STFT Mean	STFT STD	DB 2 Wavelet Mean	DB 2 Wavelet STD	GAU 3 Wavelet Mean	GAU 3 Wavelet STD
Sernorm1	398	216	500	362	224	147
Sernorm2	353	175	438	316	211	151
Sernorm3	478	237	823	482	932	356

IV.B.4 Fasting

During fasting, which constitutes a constant zero external perturbation state, two intrinsic dynamic time-scales are high-lighted. These consist of the oscillatory intrinsic dynamics on a time-scale similar to those seen during continuous infusion, and a long-term circadian component. Two methods were used to extract these circadian rhythms from the original data. The first was a 40 point (in this case equivalent to 480 minutes or 8 hours) moving average, with adjustments to the endpoints (less points used in moving average at the edges of the signal). A circadian rhythm of amplitude of approximately 20 mg/dl is observed in all three subjects (figure 4-17 and 4-18).

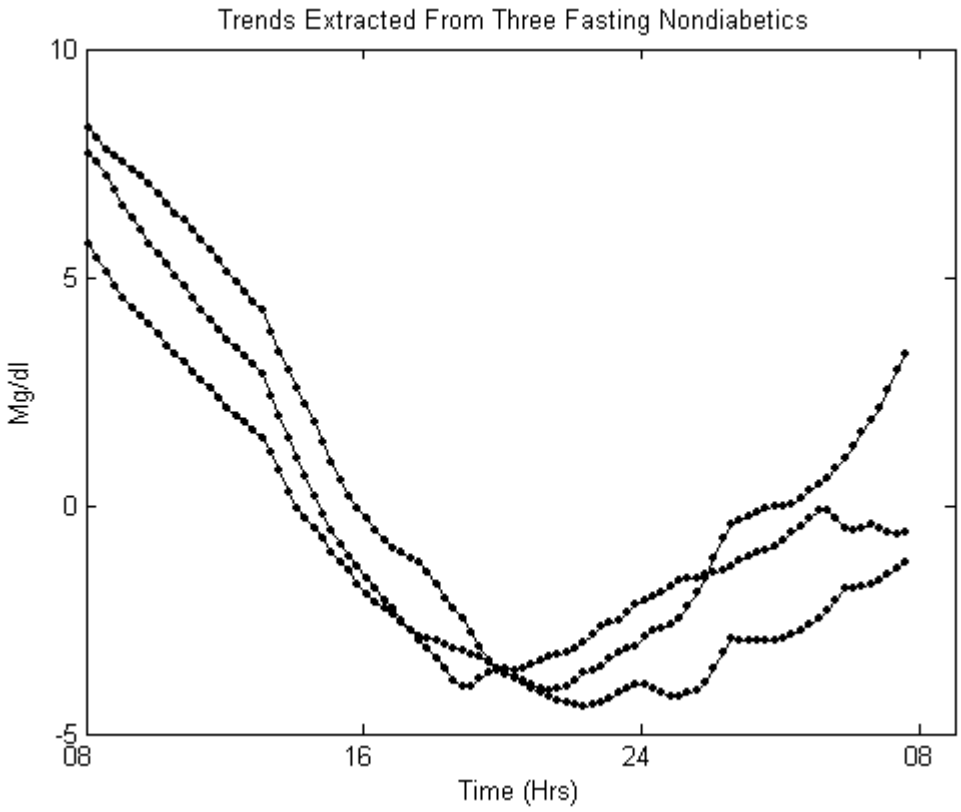


Figure 4-17: Extracted circadian rhythms from three fasting nondiabetics using a moving average process.

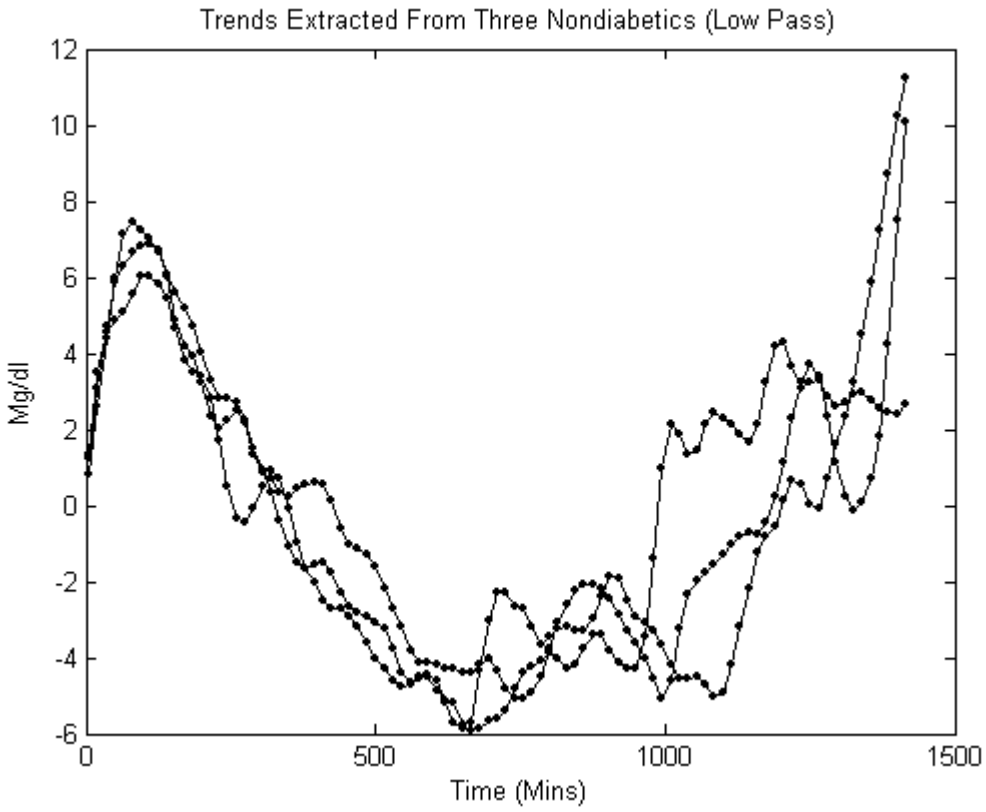


Figure 4-18: Extracted circadian rhythm using a butterworth digital low pass filter.

Spectral and time-frequency analysis was performed only on the high-frequency components removed by applying the low-pass filter described above to the three available data sets (4-19). The resulting pulses were very small in amplitude as expected, since fasting represents the zero input state and it has already been observed that decreased continuous infusion reduces the pulse amplitude in these cases.

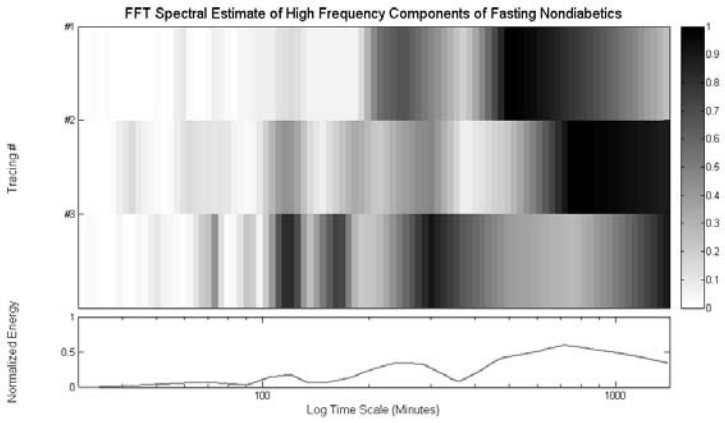


Figure 4-19: FFT based spectral analysis of the fasting nondiabetic signal. The circadian component is the most prevalent source of signal energy but the ultradian oscillations are also noted.

Results of pulse width analysis yielded a set of pulses with widths that are similar to the ones observed in continuous feeding (see the next section) but with a relatively small amplitude and are displayed in table 4-14.

Table 4-14: Pulse-width analysis of fasting time-series from nondiabetic individuals.

	Number of Full Pulses	Mean Pulse Width	Mean Pulse Amplitude
Shanorm1	9	147	5.7
Shanorm2	9	128	5.0
Shanorm3	11	117	4.2

The pulse widths noted are well within the range of those noted with continuous IV and enteral nutrition. This is not surprising as fasting can be thought of as a constant zero input, and considering the continued release of glucose into the blood stream by the liver during fasting, similarities may be shared in the dynamics.

IV.B.5 Continuous enteral feeding & IV infusion

As mentioned in the data section, in the presence of constant input into the system, either by continuous infusion through the gut or through the blood stream, oscillatory behavior is noted. In this section data collected through various studies are pooled together and the nature of this oscillatory behavior is quantified. At first one glance the methods applied in the previous section to these data sets. In order to isolate the high-frequency component related to the oscillations in question, the data is first filtered using a moving average filtering scheme with a period of half a day in order to remove circadian influences. The results for continuous enteral feeding are shown in figure 4-20 while the results for intravenous feeding is shown in figure 4-21.

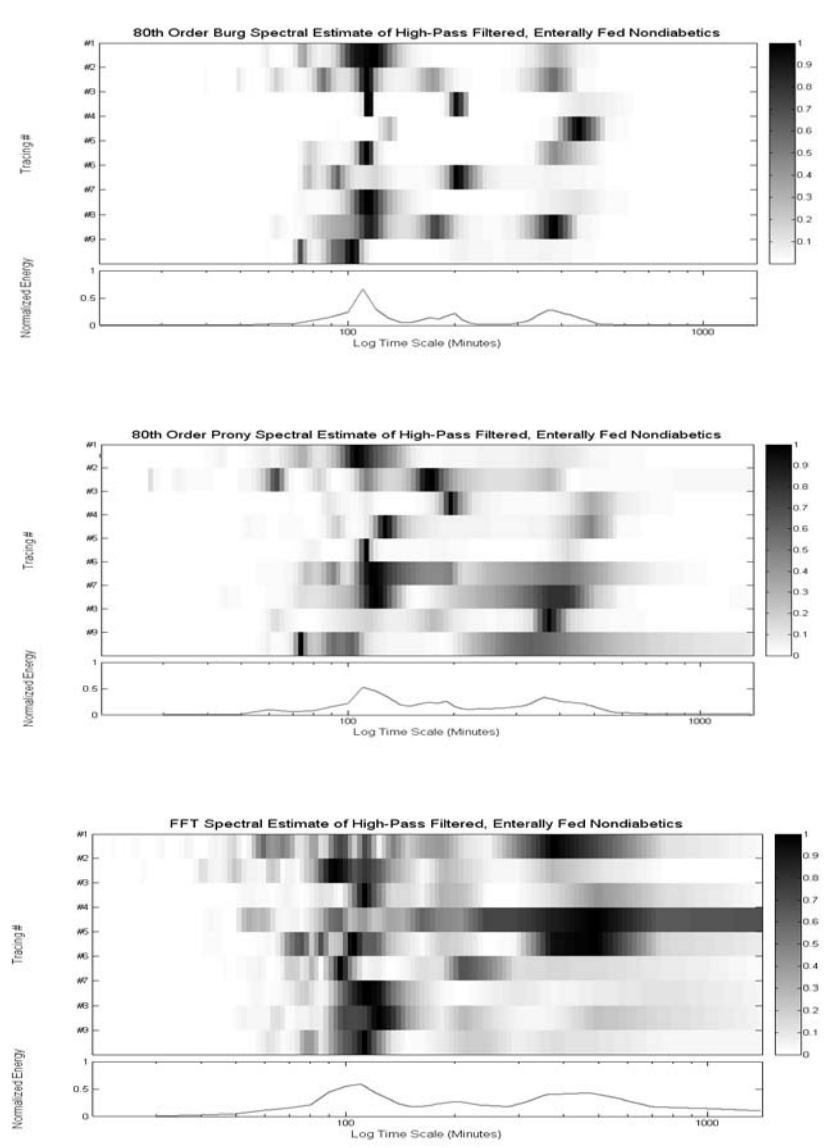


Figure 4-20: Three different methods are used to perform a spectral estimate of nondiabetics with continuous enteral feeding. A large peak at around 100 minutes is noted corresponding to the very evident oscillations in the signal.

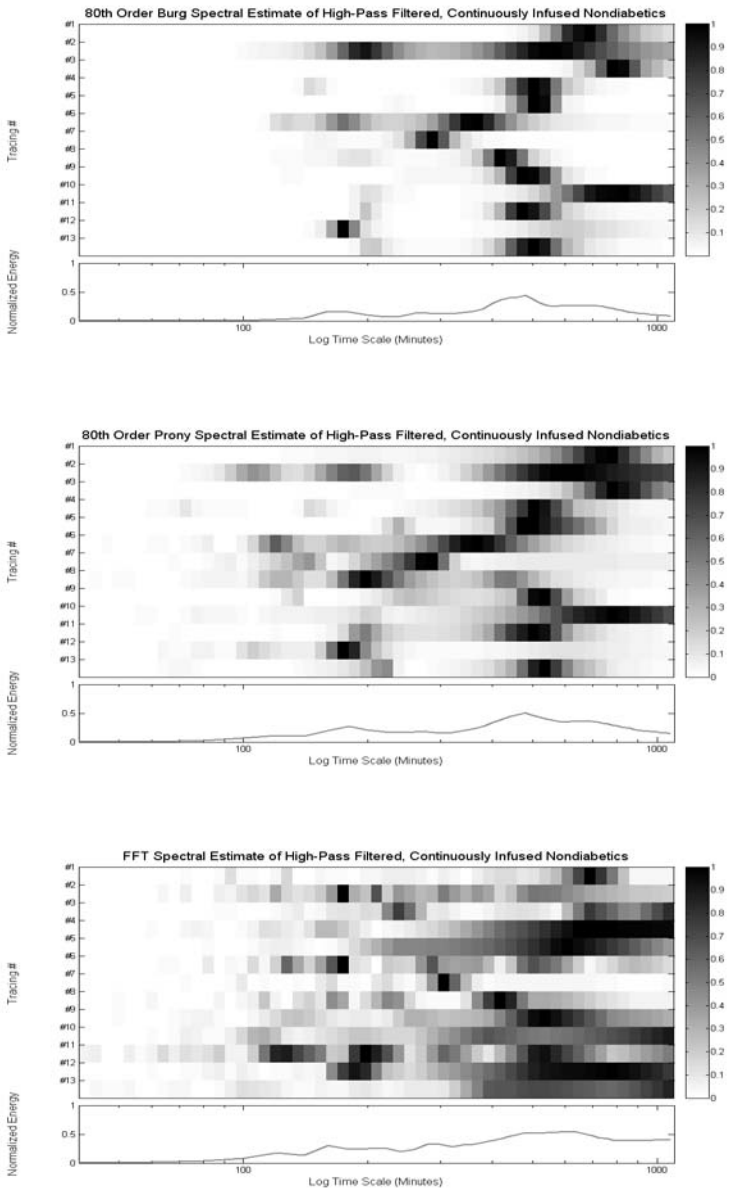


Figure 4-21: Three different methods are used to perform a spectral estimate of nondiabetics with continuous Intravenous feeding. The peaks are far more distributed and less robust than enteral feeding. Because of the nature of the experiments, it is difficult to ascertain whether the differences arise from the mode of feeding alone or if other experimental conditions play a part in the differences, but clearly medium scale dynamics are present in the intravenous case in addition to the pulses which are observed.

Looking at the spectral content of both subgroups a few observations can be made. One is that in the case of the continuous enteral feeding experiments (all of

which were performed by the same group, but over multiple experiments) appear to have a stronger relative peak in the region of the expected oscillatory behavior. The IV data, on the other hand, appears to have strong mid-scale dynamics (~6-12 hours) in addition to the oscillatory behavior. In fact, while many of the absolute energy peaks in the case of the continuous enteral feeding correspond to the oscillatory infusion, the IV case is generally dominated by the longer time-scales as shown in table 4-15 which summarizes the spectral observations using three different methods.

Table 4-15: The time-scales estimated using three methods for different types of continuous feeding including continuous enteral nutrition as well as IV feeding. Simnorm1-9 are continuous enteral feeding while the rest of the data sets represent continuous IV feeding.

Patient	Peak Period (FFT)	Peak Period (Prony)	Peak Period (Burg)
Simnorm1	480	93	512
Simnorm2	90	85	116
Simnorm3	109	473	512
Simnorm4	1410	470	512
Simnorm5	473	406	102
Simnorm6	95	95	95
Simnorm7	111	115	111
Simnorm8	120	95	116
Simnorm9	110	57	98
Polnorm1	1590	1060	1024
Polnorm2	166	790	732
Polnorm3	226	903	732
Polnorm6	1185	593	549
Polnorm7	555	555	226
Polnorm9	158	162	165
Polnorm10	290	262	256
Polnorm12	395	372	366
Vannorm1	480	262	256
Vannorm2	1440	960	960
Vannorm3	1440	960	480
Vannorm4	720	720	768
Vannorm5	1455	224	549

Pulse-width analysis performed by pulse detection can be used to characterize the oscillatory behavior of the time-series. This is chosen to analyze the data set because of the contamination with larger-scale dynamics mentioned above, and the fact that given the limited size of the data, frequency analysis can often be misleading as the similarity between these time-scales in frequency terms is difficult to distinguish using short time-series. Mean pulse-width varies from 76 to 175 minutes and interestingly varies between experiments, while amplitude varies from 7-16 mg/dl and is much more spread out between the experiments (table 4-16). The correlation between the size of the pulse and its width is not strong.

Table 4-16: Pulse analysis of IV infused Nondiabetic. Once pulses were detected using the peaks and valleys, their width (in minutes) and amplitude was determined and recorded for each time-series. Note the variability within each experiment and between experiment. The second group of experiments (vannorm1-5) show much faster pulse width.

	Number of Full Pulses	Mean Pulse Width (min)	Mean Pulse Amplitude (mg/dl)
Polnorm1	23 (53 Hour Period)	127	9.2
Polnorm2	23 (53 Hour Period)	127	15.9
Polnorm3	17 (53 Hour Period)	173	16.3
Polnorm6	9	117	14.9
Polnorm7	9	83	14.7
Polnorm9	20 (53 Hour Period)	135	7.2
Polnorm10	22 (53 Hour Period)	175	7.8
Polnorm12	14 (53 Hour Period)	135	6.8
Vannorm1	14	81	7.2
Vannorm2	12	95	9.2
Vannorm3	15	83	7.8
Vannorm4	12	99	10.7
Vannorm5	17	76	8.6

Time-frequency analysis of the continuous feeding scenarios reveals a pronounced change in the signal frequency content in the larger time-scale. An example is shown in figure 4-22.

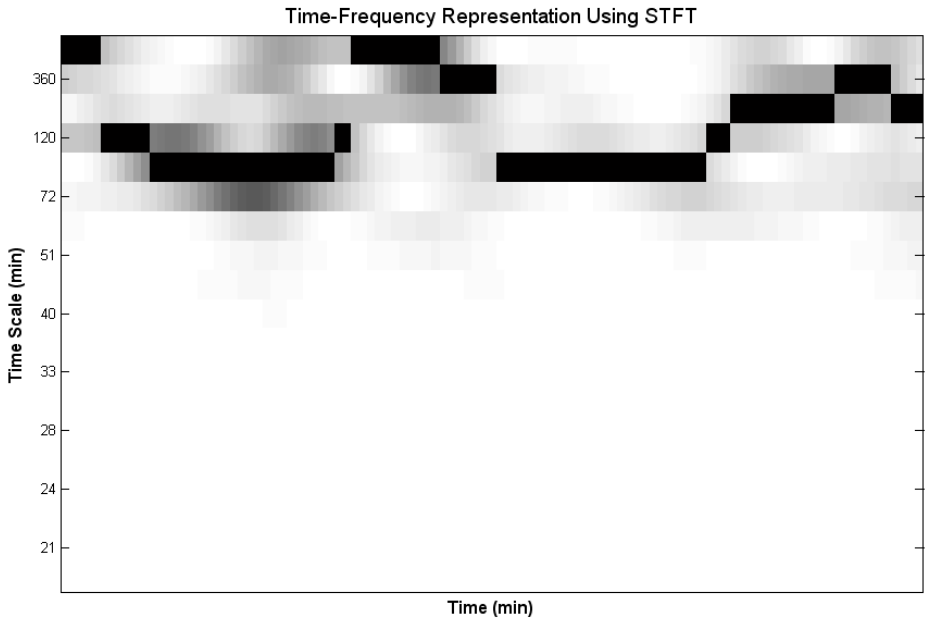


Figure 4-22: The time-frequency representation of two days worth of dynamic data from an IV continuous infusion. The circadian rhythm can be clearly noted in this analysis.

In fact, this is noted by the investigators in the case of continuous enteral feeding, but is noted in the form of amplitude modulation and not a change in frequency. Of particular interest is the appearance and disappearance of longer-time-scale energies. As mentioned in the methods section, to summarize the magnitude of this effect, the energy centers were located and their position measured in time-scale units for each of the methods used (Short-Time Fourier Transform, Wavelets). The resulting statistics of the location of these peaks are reported in the table 4-16.

Significant inter-individual differences as well as the large standard deviations in the centers point to the observation that although these time-series clearly do contain oscillatory behavior, these pulses are not fixed in wavelength, and are accompanied by other time-scales of dynamics that remain unstudied in terms of mechanism (table 4-17).

Table 4-17: Results of various methods of time-frequency analysis applied to the continuous input data sets.

Patient	STFT Mean	STFT STD	DB 2 Wavelet Mean	DB 2 Wavelet STD	CGAU Wavelet Mean	CGAU Wavelet STD
Simnorm1	167	119	741	404	434	220
Simnorm2	129	78	444	355	322	250
Simnorm3	259	111	596	280	263	144
Simnorm4	297	98	577	245	289	134
Simnorm5	211	131	726	335	379	179
Simnorm6	183	115	504	366	278	179
Simnorm7	237	126	643	309	314	163
Simnorm8	181	105	575	405	420	263
Simnorm9	186	132	1084	159	564	122
Polnorm1	780	88	1021	211	774	364
Polnorm2	520	320	1126	355	753	261
Polnorm3	709	217	997	278	709	304
Polnorm6	251	86	702	178	246	89
Polnorm7	272	41	689	112	176	59
Polnorm9	251	188	302	198	259	159
Polnorm10	249	37	237	163	265	261
Polnorm12	310	105	632	588	369	300
Vannorm1	275	121	511	292	263	124
Vannorm2	330	72	870	109	447	99
Vannorm3	330	69	734	206	344	123
Vannorm4	287	104	690	235	354	115
Vannorm5	283	105	763	359	380	151

IV.B.6 Constant insulin infusion

As can be seen in figure 4-23, and similar to glucose infusion case, constant insulin infusion appears to stimulate similar periodic behavior in blood glucose. One data set with four subjects was acquired and analyzed using both spectral and pulse analysis methods.

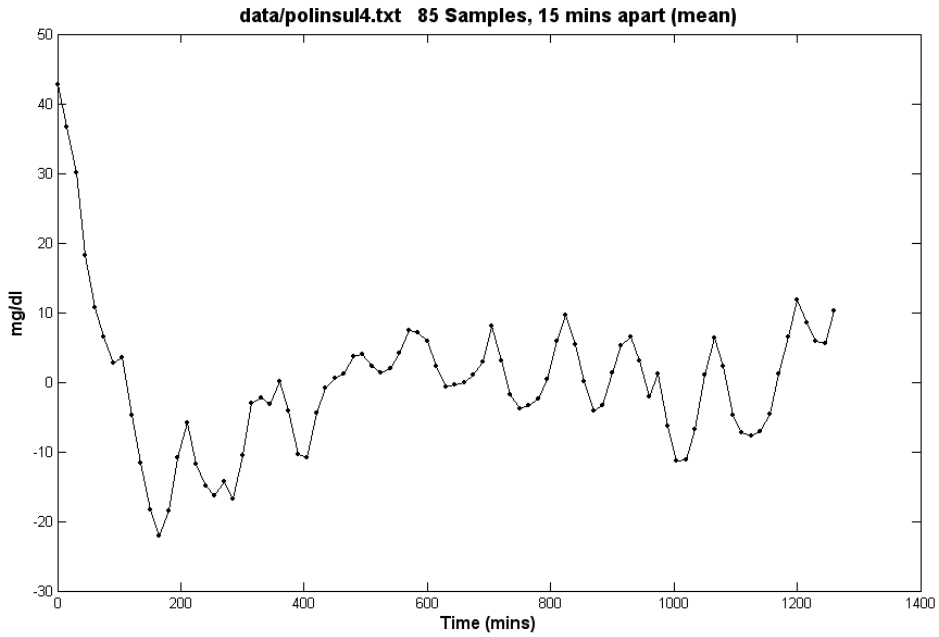


Figure 4-23: Glucose time-series with constant infusion of insulin in a nondiabetic subject.

Table 4-18 shows the results of the pulse analysis, delineating pulses which are very similar in period to the ones observed in enteral and IV feeding and similar in amplitude. An example of a spectral estimates derived from the data is shown in figure 4-24. Here, as before a high-pass filter with a corner frequency of 720 minutes was applied to remove circadian rhythms. Note that once again a significant portion of the signal energy is concentrated in slower time-scales of dynamics.

Table 4-18: Pulse analysis of the insulin infusion time-series

	Number of Full Pulses	Mean Pulse Width (min)	Mean Pulse Amplitude (mg/dl)
Polinsul1	15	77	8.3
Polinsul2	16	83	9.6
Polinsul3	15	85	7.4
Polinsul4	10	99	10.1

Spectral analysis of the high-pass filtered data yields a similar pattern to the IV infused individuals with a significant localization of energy in the longer time-scales near 600-700 minutes, while a peak is present in the vicinity of the pulsatile behavior (figure 4-24).

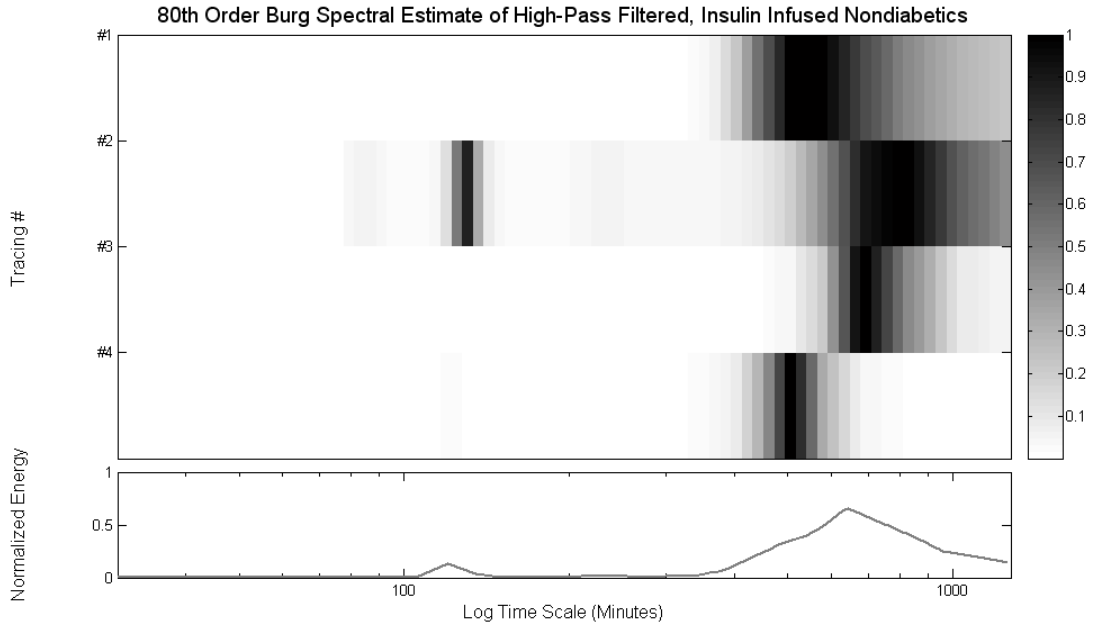


Figure 4-24: 80th order burg spectral estimate of the high-passed filtered (to remove circadian components) signal. Despite the oscillations, much of the signal energy is concentrated in the long-term fluctuations.

IV.B.7 Entrainment to glucose and insulin infusion

Entrainment describes a phenomenon in which a system begins oscillating in phase and at the same frequency as an input signal. Studies have been performed to look at this phenomenon in particular as a tool for diagnosis of insulin resistance, which is hypothesized to lead to the loss of entrainment. Six 24 hour time-series with a constant frequency infusions were extracted from the literature ([40, 41]. To assess degree of entrainment, simulated data was created with the same sampling frequency and infusion characteristics. Degree of entrainment can be quantified by consideration of the correlation between the simulated signal or in frequency domain using the coherence function. Difficulty arises, however, because both of these functions consider the phase in the input and resulting signal and even slight phase differences can lead to underestimating the degree to which the system is responding to the input. An FFT based spectral estimate of the simulated data (which represents 100% entrainment) was studied and the significant signal energy band delimited. The times-series with infusion from the patient was then also analyzed using the same technique and the percentage of signal energy contained in the same significant band was measured. This percentage was termed the degree of entrainment. An example of such a study is shown in figure 4-25.

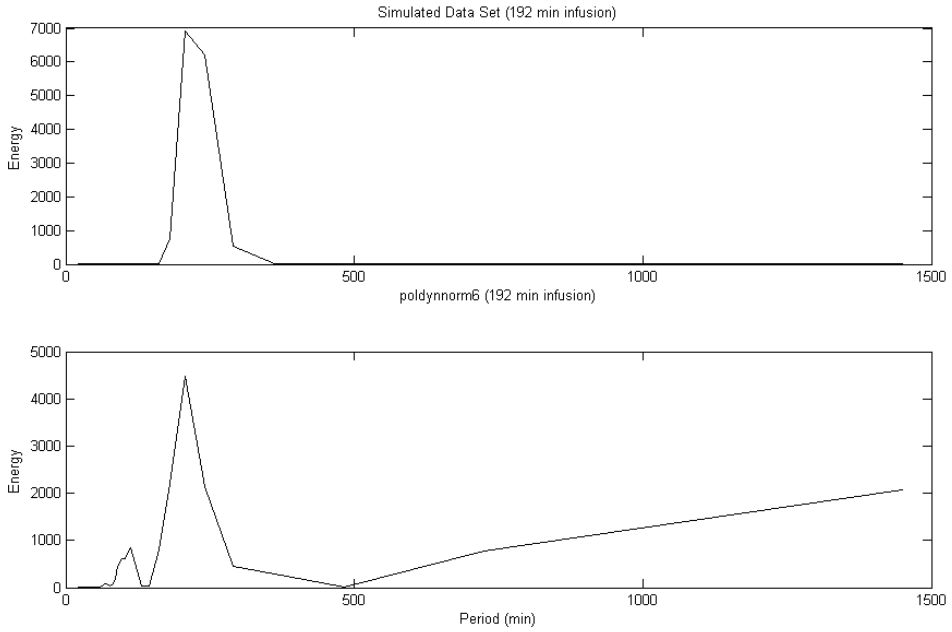


Figure 4-25: FFT Spectral estimate of simulated data with the same frequency as the dynamic input (top) and the FFT spectral estimate of the blood glucose values in the patient with dynamic infusion. The percentage of signal energy contained in the band that is significant in the simulated data is utilized as a marker for entrainment.

This analysis was performed for the available datasets and is summarized in table 4-19. Significant entrainment was noted in the nondiabetic patients and is reduced in the progression toward impaired glucose tolerance and diabetes.

Table 4-19: Table showing the percentage of entrainment in different patients. Patients 1-3 are nondiabetic, patient 4 has impaired glucose tolerance (IGT) and patient 5 has type II diabetes.

Patient ID #	Series	"Wavelength" of Infusion Sinusoid	% Entrainment
1	Poldynnorm3	144 min	79%
1	Poldynnorm4	96 min	55%
2	Poldynnorm6	192 min	52%
2	Poldynnorm7	128 min	81%
3	Poldynnorm9	144 min	58%
3	Poldynnorm10	96 min	37%
4	Poldynigt2	144 min	50%
4	Poldynigt3	96 min	23%
5	Poldynnidm2	144 min	41%
5	Poldynnidm3	96 min	16%

IV.C Measures of signal complexity

Many approaches have been proposed for quantification of signal complexity. This concept has evaded an agreement on the exact general definition but has found utility in time-series analysis in specific fields. Such measures have arisen from the field of information theory (Signal Entropy), nonlinear dynamics and physiologic time-series analysis of cardiodynamics [10] In general it is agreed that complexity is a measure of the multiple time scales in which information exists in the signal, such that a complexity random signal (white noise) is extremely complex whereas a simple signal like a sinusoid is "regular" and thus not complex.

The advantage of such methods is that they produce a signal number that communicates essential characteristics of the signal. The disadvantage is, however, rooted in the inherent difficulty of summarizing an entire time-series in a single

number, which in essence forces many undesirable qualities on such measures. These include lack of uniqueness (meaning that two different time-series can produce the same number) and difficulty with interpretation. Also the sensitivity to changes in uninteresting signal parameters (such as sampling rate) are not very well described in the literature.

For these reasons such measurements have found their utility in comparing signals with a single variable in the system changed, that is comparing the "relative" entropy between two similar signals. In this section some of these tests are applied to the blood glucose time-series. The objective is to capture, using a single parameter, the complexity and regularity of the time-series. The results are shown in table 4-20 and 4-21. These shall be revisited in the context of comparison to other subjects.

Table 4-20: Entropy and measures of complexity described for three nondiabetic, ideally sampled subjects.

Data Set	Entropy	Apen	Median Sampent
Sernorm1	3.9	.35	.21
Sernorm2	3.6	.36	.24
Sernorm3	3.5	.51	.36

Table 4-21: Entropy and measures of complexity described for nondiabetics during continuous enteral feeding.

Data Set	Entropy	Apen	Median Sampent
Simnorm1	3.7	.78	5.2
Simnorm2	3.6	.81	5.3
Simnorm3	3.7	.76	5.3
Simnorm4	3.7	.77	5.3
Simnorm5	4.0	.60	4.9
Simnorm6	3.6	.71	5.3
Simnorm7	3.5	.69	4.9
Simnorm8	3.3	.61	.9
Simnorm9	3.4	.64	5.5

IV.D Evidence for nonlinear determinism

As mentioned in the previous chapter, testing of time-series for the presence nonlinearity is a complex and highly debated topic. As discussed, one can choose to approach the problem using the method of surrogate construction. Briefly, this is a method in which linear surrogates (that is, time-series which have the same linear autocorrelation structure as the data) are created. These surrogates are tested for degree of determinism. The logic is that if the system is a linear one, linear equivalent systems should be able to capture the determinism stored in the autocorrelation function. If after generating a “sufficient” number of surrogates the determinism in the signal remains higher than the surrogates then the null hypothesis of linearity is rejected.

The most complex aspect of such tests is that they must measure determinism in a way that is beyond linear measures of determinism, as by design, the surrogates

should have the same autocorrelation function. This generally requires embedding the time-series in multiple dimensions and performing nonlinear analysis, itself which is complex and involves multiple estimation steps. Further difficulty arises with the proper construction of linear-stochastic surrogates: because the signal may be non-stationary on some time-scales, generation of surrogates with proper auto-correlation functions become difficult as non-stationarity can give rise to a changing autocorrelation function. However, focusing on small regions may in effect serve as a "linear" approximation thus limiting the utility of assessment of smaller, stationary time-scales. Because of the lower sampling period and the limited number of meals included in the vannorm6-9 datasets they were not used for this analysis.

Two methods of measuring determinism were utilized. One is the Kaplan delta-epsilon test and the other was a nonlinear predictor. Both are described in the methods section. The results for the Kaplan test are shown in table 4-22. The K statistic is the result of the test, which tends to zero as the degree of determinism increases. In the latter two tables, the maximally different K values are shown between the 100 surrogates and the data.

Table 4-22: The results of the Kaplan delta-epsilon test to nondiabetic time-series and 100 linear surrogates. The second column contains the delay computed from the autocorrelation in order to perform time-delay embedding. The dimensions (third column) represent the dimensions of embedding. In the fourth column the minimum K statistic (the most deterministic linear surrogate) is shown and compared to the data in column five. For certain dimension parameters, linearity can be rejected, particularly for the third data set. The results, however, are highly dependent on the choice of delay and embedding dimension considered.

Data Set	AMI based estimated Delay	Dimensions	Min Surrogate K	Data K
Sernorm1	18	1	2.6	2.1
		2	1.2	1.5
		3	.6	1.0
Sernorm2	23	1	2.1	1.5
		2	1.3	1.2
		3	.7	1.5
Sernorm3	13	1	1.9	1.6
		2	1.2	.9
		3	.6	.9
		4	.2	.4
	Autocorrelation based estimated Delay			
Sernorm1	12	1	2.5	2.1
		2	1.4	1.4
		3	.6	1.6
Sernorm2	15	1	2	1.6
		2	1.2	.9
		3	.9	1.1
		4	.5	1.5
Sernorm3	20	1	1.9	1.6
		2	1.1	1.0
		3	.7	.7
		4	.1	.6
	Physiologic based estimated Delay			
Sernorm1	50	1	2.5	2.1
		2	1.3	1.1
		3	.4	.2

Table 4-22: The results of the Kaplan delta-epsilon test to nondiabetic time-series and 100 linear surrogates. The second column contains the delay computed from the autocorrelation in order to perform time-delay embedding. The dimensions (third column) represent the dimensions of embedding. In the fourth column the minimum K statistic (the most deterministic linear surrogate) is shown and compared to the data in column five. For certain dimension parameters, linearity can be rejected, particularly for the third data set. The results, however, are highly dependent on the choice of delay and embedding dimension considered (Continued)

Sernorm2	50	1	2.1	1.6
		2	1.0	1.1
Sernorm3	50	1	1.8	1.6
		2	1.1	.6
		3	.4	.13

Note that dimensions were omitted as insufficient samples led to incoherent results (such as negative values). In the case of physiologic delay, the null hypothesis can be rejected for Sernorm1 and Sernorm3, as no higher dimensional linear surrogate had a smaller K than the nonlinear model. This could however be the result of not being able to go to higher dimensions based on sample insufficiency. Based on these preliminary observations, the null hypothesis of linearity cannot be rejected as the choice of delay remains critical to the outcome, and as the system has shifting frequency contents, this method of detecting nonlinearities may not be easily applied in this case. While the null hypothesis cannot be strictly rejected, it is also clear that this does not imply that the system is linear as the null hypothesis can in fact be rejected depending on certain delay choices (such as the arbitrarily chosen 50 minutes based on physiologic insight derived from time-scale analysis). In conclusion evidence

for nonlinearity exists in two of the time-series studied but the conclusion cannot be made based on firm computational grounds.

IV.E Nonlinear time-series analysis

The presence, quantification and understanding of the nature of nonlinearity present in the system can suggest further signal processing approaches and possible approaches to controlling the system. Nonlinear time-series methods are both difficult to implement in a consistent manner and also difficult to interpret. Unfortunately, in the biological realm there is generally good reason to look for nonlinearities as most biological systems contain bounds and thresholds, which introduce these nonlinearities into the system.

The procedure most often followed for nonlinear time-series analysis, and described in the methods section was applied to a selected number of data sets from nondiabetics. Briefly, this involves the determination of the lag using either autocorrelation function or average mutual information (both are included, the AMI was used in subsequent computations). This lag is then used to embed the time-series in multiple dimensions and various tests are used to determine whether the number of dimensions is “sufficient”. These include the method of false nearest neighbors as well as Cao’s method. Taken’s dimension estimator may be used to confirm the data’s dimensionality. The results are shown in table 4-23. Unfortunately, limited agreement exists between the different dimensional estimators. Some computations yielded erroneous results perhaps due to lack of data in the time-series. The problem, as

anticipated once the time-scales of dynamics are known is that the system does not have enough orbits around its attractor because of limited duration of sampling.

Table 4-23: two different time lag estimates using autocorrelation function and average mutual information is shown in columns two and three for different nondiabetics under different perturbation regimes. The lag computed from AMI is used to embed the time-series and determination of proper dimensionality of the dynamics is made using multiple methods.

	Autocorrelation Lag (Samples)	AMI Lag (Samples)	Embedding Dimensions (False Nearest Neighbors)	Embedding Dimensions (Cao's Method)	Taken's Dimension Estimator
Sernorm1	12	13	5	4	5.0
Sernorm2	15	13	4	3	3.7
Sernorm3	13	14	4	4	6.3
Simnorm1	8	4	N/A	4	5.6
Simnorm2	6	4	N/A	3	2.5
Simnorm3	9	4	N/A	3	7.7
Simnorm4	8	5	N/A	4	N/A
Simnorm5	15	4	N/A	3	6.5
Simnorm6	8	4	N/A	3	N/A
Simnorm7	15	4	N/A	3	2.6
Simnorm8	9	4	N/A	3	3.8
Simnorm9	14	5	N/A	4	3.8
Polnorm1	14	10	N/A	4	N/A
Polnorm2	14	6	N/A	4	N/A
Polnorm3	13	5	N/A	3	2.1
Polnorm9	11	3	N/A	3	3.0
Polnorm10	18	3	N/A	3	1.9
Polnorm12	25	3	N/A	3	3.5

Most of the data sets did not have sufficient number of points to attempt the calculation of the dominant Lyapunov exponent. The Lyapunov exponents represent the degree to which the system evolves in each dimension of its state in an exponential sense. Thus negative exponents represent a shrinking in energy in that dimension.

Systems with a more negative Lyapunov profile tend to be dissipative, whereas if too many large positive Lyapunov exponents exist, the system tends to grow in volume in state space and "blow up". The most positive Lyapunov exponent then can be interpreted as the representation of the degree of instability in the least stable dimension. These were calculated for the three nondiabetics and the results are shown in table 4-24. These will be compared to the diabetic counterparts in chapter V.

Table 4-24: Lyapunov exponent estimation for three nondiabetics.

	False Nearest Neighbor's Dimensions	Cao's Method Embedding Dimension	Largest Lyapunov Exponent (Max Estimate)	Largest Lyapunov Exponent (Mean Estimate)
Sernorm1	5	4	.53	.32
Sernorm2	4	3	.47	.24
Sernorm3	4	4	.34	.20

IV.F Meal event detection and analysis

V.F.1 Meal detection

In the normal subject, the meals are accompanied by a very pronounced change in glucose value. In particular, as discussed, the rate of change of glucose values undergoes significant positive increase. This can be used to detect meals automatically from an existing time-series from nondiabetics. Two approaches were used to test this process.

The first approach relies on the concept of dividing the signal into “runs” that is, continuous segments where the signal derivative does not change sign. These runs can then be analyzed in terms of their length (both in time and amplitude) as well as other characteristics such as maximum velocity during the run. A program can then analyze such runs and based on characteristics, determine if they are likely to have been caused by a meal event. In this case, the amplitude of the positive runs and the maximum velocity were selected and kept the same throughout the process although it is likely that an optimal parameter can be estimated for each individual to yield maximum detection. For non-diabetics a maximum velocity of .5 mg/dl*min and a minimum run-depth of 10 mg/dl was used. For diabetics, a maximum velocity of .25 mg/dl*min and a minimum run-depth of 20 mg/dl was used.

The second method is based on a similar concept except that run detection is performed using a rolling ball algorithm. Briefly, a physical simulation of a rolling ball being effected by gravity and friction is used to detect peaks and valleys in the time-series. The advantage of this method is that as the ball gains momentum, it is able to overcome small “bumps”, that is small reversals in the derivative of the signal. Thus, larger runs are not simply terminated by small reversals that are temporary. This method was originally developed by our group to detect vessels in images and was modified for use in this context. Examples are shown in figure 4-26 and results are summarized in table 4-25.

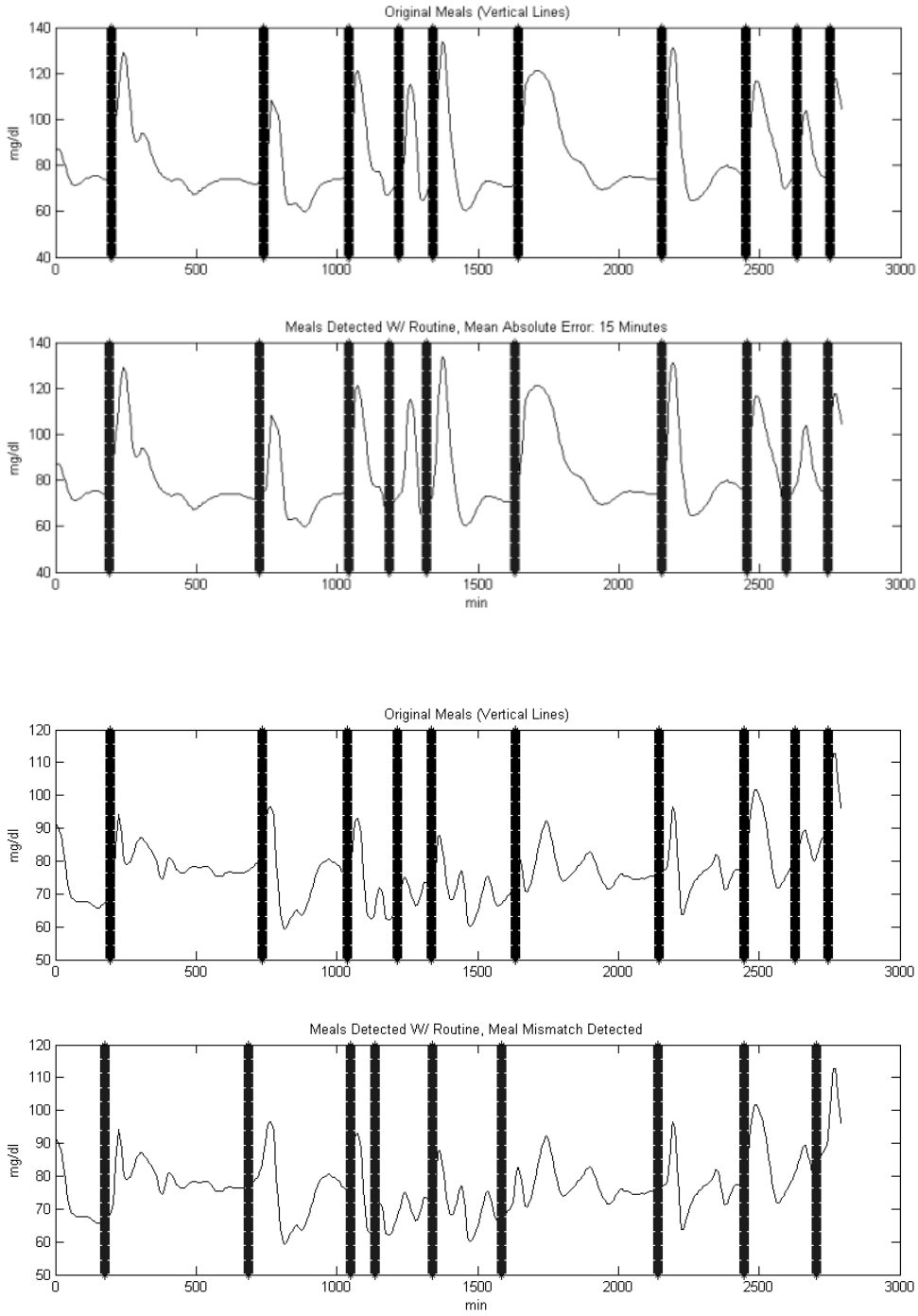


Figure 4-26: Two meal detection sessions. The very top panel shows a time-series with meals marked with vertical lines. The second panel shows the detected meals and the average error. The third and fourth panel demonstrates an attempt on another data set where false meals were detected.

Table 4-25: Summary of meal detection success in a group of time-series. The first column represents the fraction of meals correctly detected. The second column shows the number of spurious meals detected. In the case where meals were correctly detected and no spurious meals were detected, the mean absolute error is given.

	Meals Correctly Detected (Approach #1)	False Positives	Mean Absolute Error)	Meals Correctly Detected (Approach #2)	False Positives	Mean Absolute Error
Sernorm1	10/10	0	15 min	10/10	0	10 min
Sernorm2	9/10	0	-	10/10	0	25 min
Sernorm3	8/10	0	-	8/10	0	-
Serone1	10/10	0	27 min	10/10	0	33 min
Serone2	7/9	0	-	8/9	0	-
Serone3	10/10	0	26 min	10/10	0	20 min
Serone4	9/10	1	-	8/10	2	-
Serone5	7/10	3	-	10/10	1	-
Serone6	10/10	0	37 min	10/10	0	26 min
Serone7	10/10	1	-	10/10	1	-
Serone8	9/10	0	-	7/10	0	-
Serone9	10/10	2	-	10/10	2	-
Serone10	10/10	2	-	10/10	2	-
Serone11	8/10	1	-	7/10	2	
Serone12	10/10	1	-	9/10	2	-
Serone13	8/10	2	-	10/10	2	-
Serone14	7/10	1	-	9/10	0	-
Serone15	10/10	1	-	10/10	1	-

For each specific case, meal detection could be improved by specifically modifying the constants in the program. This suggests that individual dynamic constants may be extractable to train systems to correctly identify meals. Doing so, however, is likely to require monitoring without standardized meals and significant data for training such an algorithm. These routines demonstrate the potential that

exists in event detection based on study of dynamics which can be ultimately incorporated into analysis and control algorithms.

IV.F.2 Multiple meals

Most meals in normal individuals are not conducted after long periods of fasting, are of different sizes, are consumed at different speeds, and are often followed by movement and exercise. Thus while the fasting state can give us insight into the meal response process, it is not necessarily a good representation of the individual's typical dynamics. For this reason, an analysis of the meals in the context of multiple meals events and in some cases exercise was performed. Seven parameters were chosen primarily because they were easy to understand, compute and seemed to best capture the meal event. The first two capture the amplitude of the meal perturbation. The rise is computed as the difference between the glucose value prior to the meal and the maximum value during the next four hours or prior to the next meal if that meals occurs before four hours. During this rise, the maximum rate of change is the maximum rise velocity. The time it takes to reach the maximum value is the time to m value. Similarly, the fall is the difference between the maximum glucose value and the minimum glucose value prior to the next meal, or in the next four hours. The maximum fall velocity is the rate of fall during that period of time between the maximum and minimum glucose values. The other two variables, time to maximum velocity and time to minimum velocity, measure the time between the start of the rise and fall process, and when the maximum speed is reached in those processes.

Table 4-26: Various meal parameters (see text) are listed for the various meals for three nondiabetics with non-identical meals.

Subject	Rise	Fall	Time to Max	Max Rise Vel	Max Fall Vel	Time to Max RV	Time to Max FV
Sernorm1							
Supper	47	62	40	1.4	1.5	25	25
Breakfast	33	49	25	2.2	1.7	20	35
Lunch	47	54	30	2.7	1.2	15	35
Snack	42	51	40	2.3	2.4	25	25
Dinner	59	73	35	2.5	1.9	15	15
Supper	47	52	70	1.7	.6	20	70
Breakfast	56	66	40	2.8	2.1	25	20
Lunch	41	47	40	1.9	.7	20	85
Snack	25	29	30	1.3	.9	15	20
Sernorm2							
Supper	42	52	95	.8	.7	70	45
Breakfast	36	50	35	1.4	1.0	15	30
Lunch	24	29	45	1.2	.4	20	85
Snack	29	27	45	1.4	.6	20	20
Dinner	28	28	40	1.4	.6	30	30
Supper	45	52	45	1.6	.6	25	100
Breakfast	56	76	55	3	2.4	45	25
Lunch	16	21	45	.7	.3	25	15
Snack	19	21	40	.8	.4	25	55
Sernorm3							
Supper	25	20	25	1.4	.9	15	10
Breakfast	10	37	25	.7	1.4	5	15
Lunch	18	31	35	.9	1.4	15	20
Snack	6	9	20	.4	.3	10	15
Dinner	14	20	25	1.1	.7	15	20
Supper	22	20	105	.4	.4	65	25
Breakfast	20	33	45	1.3	1.6	35	20
Lunch	23	30	35	.9	.6	10	55
Snack	8	9	30	.3	.4	15	15

The following observations are made from table 4-26. One striking observation is the variations in the same person, given the same exact meal during two consecutive days. This outlines the difficulty associated with generalizing the results from a single oral glucose tolerance test to the general reaction of the system in a nondiabetic individual.

Analysis of the differences between specific meals is not possible as the meals were non-identical in this data set, but quite a variation exists between the extent of the meal process, the height of the excursion and the maximum rates of change within the same person as well. The fall process tends to exceed in amplitude the rise process, which is not surprising because after a meal a gradual rise in blood glucose is often seen prior to the next meal. In two of the subjects the rise process systematically has a higher maximal rate highlighting the spike like effect of meal ingestion. On average the time to maximal fall velocity is greater than the time to the maximum rise velocity, highlighting that the rise of glucose is a “faster” process. This is once again clearly observed in the first two subjects but not the third. The peak value of the meal is reached within 40-60 minutes of the initiation of the meal.

The vannorm6-9 datasets result from two patients with identical meals spaced at different intervals (2 meals the first day after fasting and three meals the second day after fasting). These data sets are much less frequently sampled (20 min) and thus do

not yield the same precision particularly in regards to the parameters having to do with time. These results are shown in table 4-27. In some cases the rate of fall was maximal immediately after the peak value of glucose and thus the time to maximal fall velocity is zero. Once again, falls are clearly larger and the rises reach a higher maximum than the falls do. There exist significant difference between the characteristics of the meals within the same individual, which is even more pronounced considering that the meals are all identical.

Table 4-27: Meal parameters extracted from vannorm6-9 which represents two individuals with two identical meals and then three identical meals. These data sets where sampled every 20 minutes which limits the resolution of the analysis.

Tracing Tag	Rise	Fall	Time to Max	Max Rise Vel	Max Fall Vel	Time to Max RV	Time to Max FV
Vannorm6							
Meal One	24	33	40	.7	.7	40	0
Meal Two	39	42	60	.9	.4	20	60
Vannorm7 (Same Subject as Vannorm6)							
Meal One	11	15	40	.3	.3	40	0
Meal Two	28	31	60	.6	.3	40	40
Meal Three	46	52	60	.9	.6	20	20
Vannorm8							
Meal One	52	59	40	1.8	1.0	40	0
Meal Two	57	59	40	2	.5	20	0
Vannorm9 (Same Subject as Vannorm8)							
Meal One	30	53	20	1.5	1.5	20	0
Meal Two	62	85	40	2.2	.8	20	40
Meal Three	76	80	60	1.8	.9	40	60

One can also consider not the amplitude or timescale of the meal but rather the general shape of the progression. A simple approach, in this case, where the meal can be divided into a rise portion and a fall portion, is to select each of these processes (rise and fall) and then subsequently rescale them so their range in time and energy goes from zero to one. These processes can be averaged to yield an average shape of rise and fall of glucose, which can then be compared between subgroups. The average rise process is shown in a single individual and three individuals in figure 4-27. Notice the characteristic shape of the meal rise process yielding and S shaped curve below the linear line. The fall process is shown in figure 4-28 where averages from a single nondiabetic, three nondiabetics and nondiabetics from a different study are shown for comparison. Note that the fall process also maintains a sigmoid profile, which is observed in all three panels. This is to be expected as the initial insulin pulse is integrated leading to a S-shaped curve.

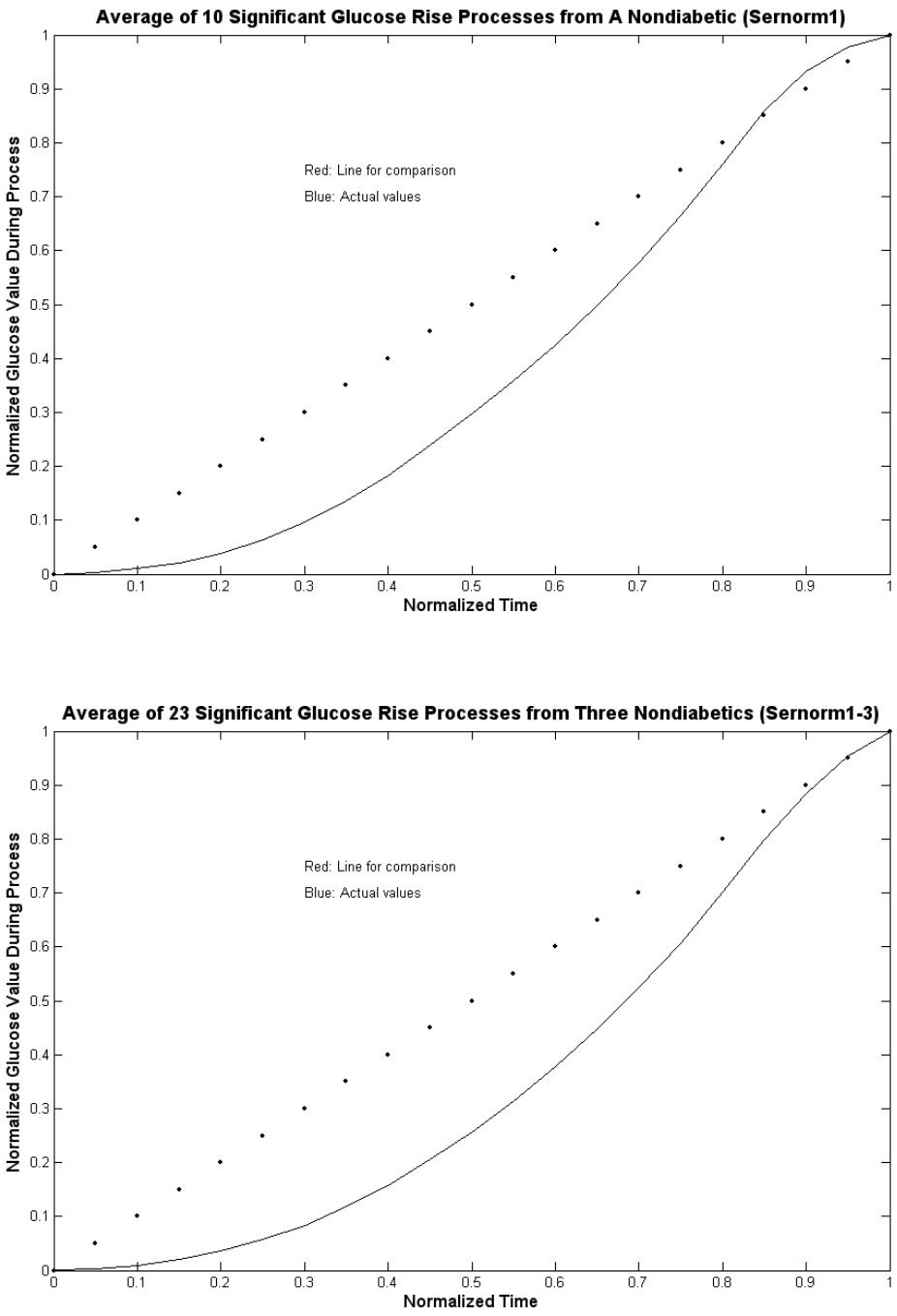


Figure 4-27: the average normalized rise process from a single nondiabetic individual and averaged between three nondiabetics. The process has a characteristic deviation with respect to the linear process shown for comparison.

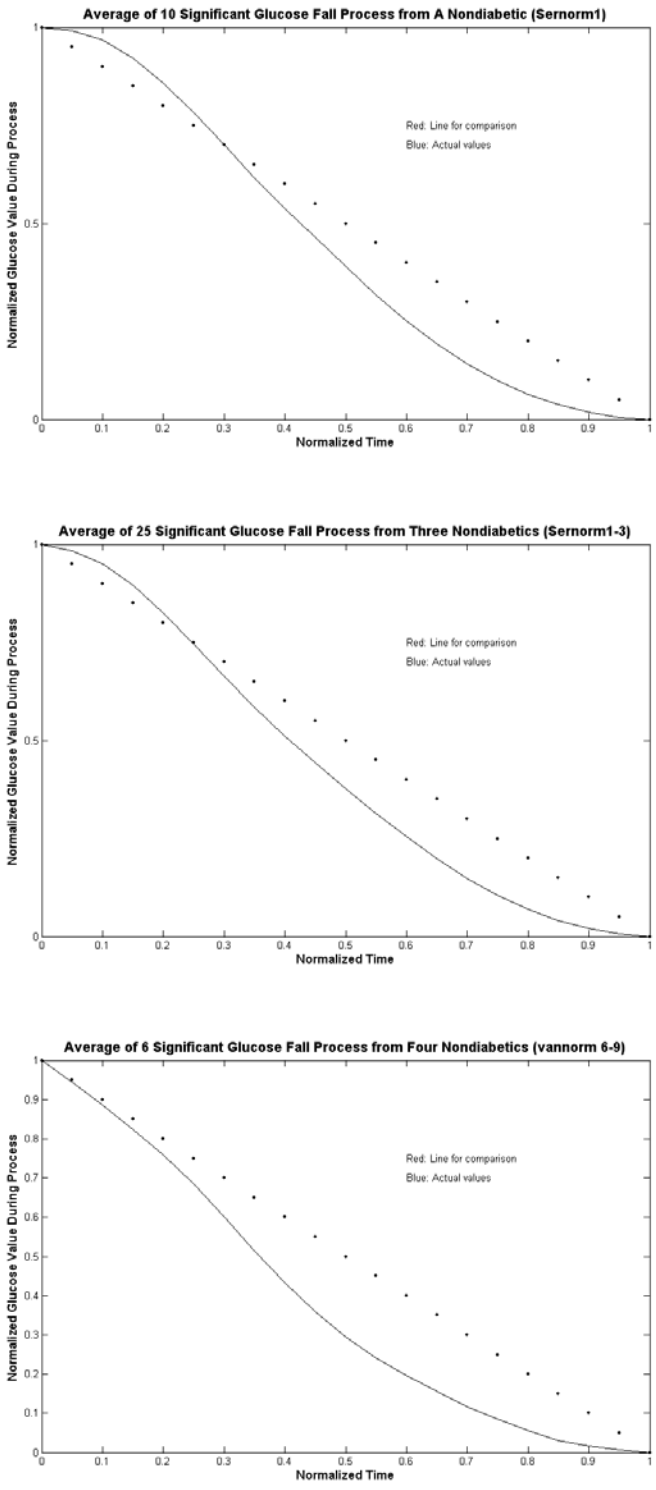


Figure 4-28: the fall process from a single nondiabetic (top), three nondiabetics (middle) and 2 nondiabetics from a different dataset (bottom). The process follows the same type of deviation from the linear reference line in each case.

The shape of the meal process will be revisited in the chapter on type I diabetes where it will be contrasted between the two groups.

IV.G Individual dynamics

One key question that is pertinent particularly to design of sensors and controllers is whether individuals within a subgroup have their own dynamic signatures, which can be used to develop individualize therapies for them. Given the limited duration of datasets, the best analysis can ask this simple question: is the day to day variation of dynamic measurements the same in the same person given identical (or in some cases different) conditions? If not, is the variability less than the variability in the group?

In the case of nondiabetics, sernorm1-3 presented two days of data under the same conditions. Vannorm6 and Vannorm7 were repeated using a different meal frequency in the same two individuals. Additionally, six of the continuously IV fed data sets were of a greater than 48 hour duration. Thus data sets were split into two in order to compare signal characteristics in the same individual under identical, and in the case of two of the data sets, different conditions.

Looking at spectral analysis, the peak of the power spectrum was calculated for both days in nine nondiabetic subjects under differing condition (normal feeding,

continuous feeding). The power spectrum peaks for each day was computed for the nine nondiabetic subjects. These spectra are shown in figure 4-29.

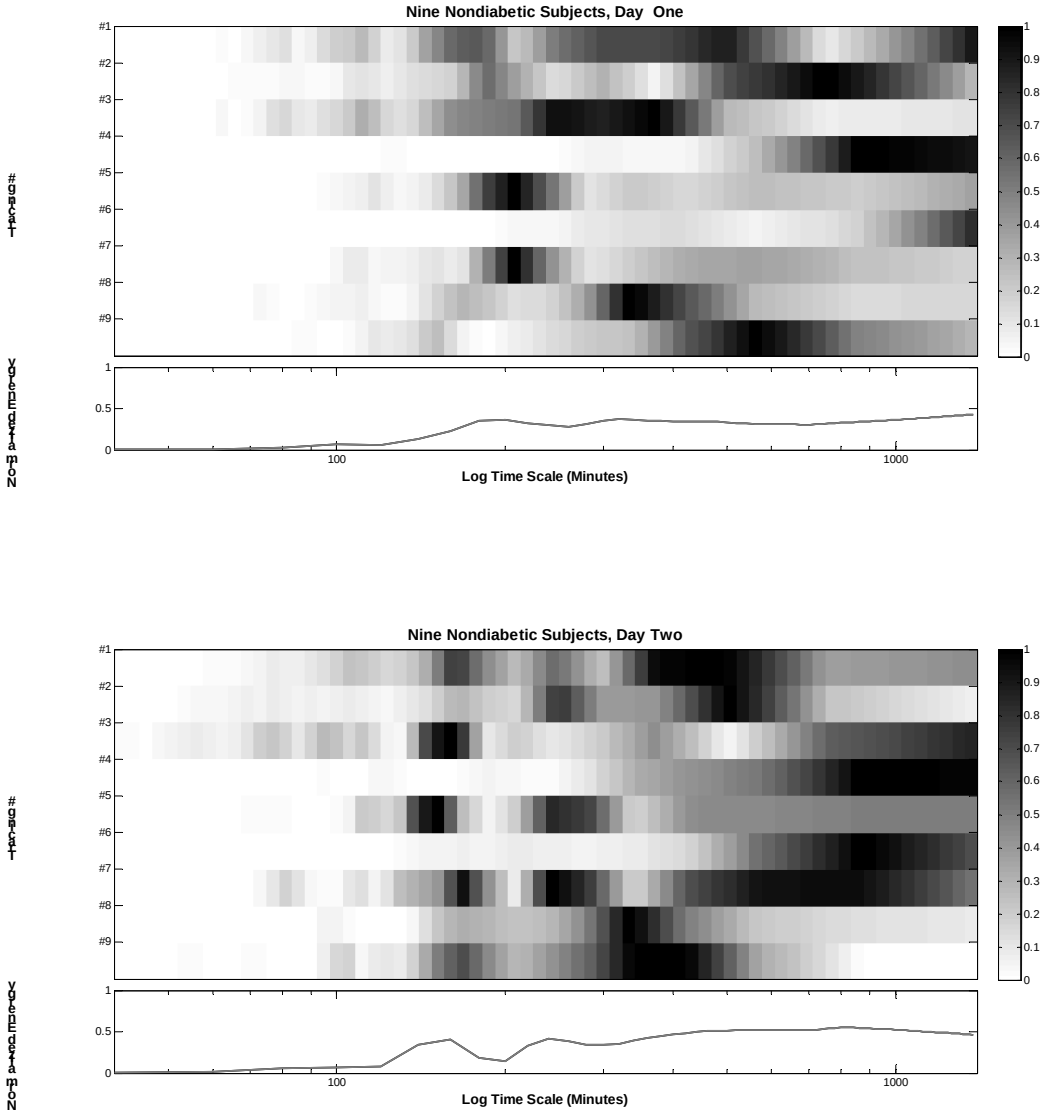


Figure 4-29. Spetra from nine nondiabetic subjects from two consecutive days. Note similar patterns of power distribution in each subject (denoted by the tracing number #) and the resulting average spectra.

Looking at the correlation between the full spectrum shows a varying relative correlation in frequency space, otherwise known as the coherence function. The coherence appears on average > 50% for the non-diabetics studied (Table 4-28). This is very limited and further studies, using multiple days are required to further examine the significance of this correlation.

Table 4-28: Correlation coefficient between the two spectra from each day for each subject. This is the "coherence" of the two time-series with respect to each other.

Subject	Coherence
Sernorm1	.68
Sernorm2	.49
Sernorm3	.21
Polnorm1	.93
Polnorm2	.31
Polnorm3	.32
Polnorm9	.37
Polnorm10	.96
Polnorm12	.45

Other measurements of dynamics such as the rate of change and its statistics can be calculated for different days of the data. In table 4-29, various statistics of rates of change are shown for the different days of the data sets. A look at the similarities of the values reveals good agreement between values in most of the individuals. Notably, those whose values did not agree, had significant disagreements throughout the various statistical measures. While a positive correlation does exist, the number of measurements is not sufficient to lead to a statistical conclusion. However, it can be said that at least in certain individuals, the variation between the two days are

minimal. Table 4-30 shows a similar trend with certain individuals (the same ones as in table 4-29) showing very similar characteristic attractor values while other individuals show differing dynamic characteristics between the two days.

Table 4-29: Statistical analysis of the rate of change as compared between two days in nine nondiabetic subjects.

Set	Max ROC day I	Max ROC day II	Min ROC day I	Min ROC day II	Skew ROC day I	Skew ROC day II	Kurt ROC day I	Kurt ROC day II
Sernorm1	2.8	2.9	-2.5	-2.2	.8	1.0	7.4	7.8
Sernorm2	1.5	3.1	-1.0	-2.5	1.1	1.3	5.9	13
Sernorm3	1.4	1.5	-1.6	-1.6	-.5	-.4	7.1	6.8
Polnorm1	.5	.8	-.5	-.80	-.1	.30	2.7	5.2
Polnorm2	.8	.4	-1.1	-.7	-.4	-.5	.2	2.5
Polnorm3	.5	.5	-.8	-.7	-.4	-.5	2.9	3.1
Polnorm9	.4	.4	-.3	-.3	.1	0.0	2.8	2.9
Polnorm10	.2	.3	-.5	-.3	-1.1	-.1	4.6	2.9
Polnorm12	.3	.4	-.4	-.4	-.3	-.1	3.8	3.1

Table 4-30: Statistical analysis of the attractor geometry as compared between two days in nine nondiabetic subjects.

Set	Attractor Area Day I	Attractor Area Day II	Attractor Concentration Day I	Attractor Concentration Day II	Attractor Symmetry Day I	Attractor Symmetry Day II
Sernorm1	650	543	.04	.04	.9	.9
Sernorm2	281	683	.11	.04	1.0	.8
Sernorm3	198	284	.17	.13	.8	.9
Polnorm1	142	110	.20	.25	1.0	.8
Polnorm2	162	77	.16	.27	1.0	1.0
Polnorm3	190	142	.08	.2	.75	.81
Polnorm9	30	34	.53	.56	.89	.89
Polnorm10	31	25	.61	.64	.83	1
Polnorm12	44	36	.30	.5	.75	1

Similarly, other measures of complexity and information can be looked at from one day to the next. Table 4-31 shows an example of these characteristics over two day period. Intra-day variations remain small compared to the intra-data-set variations in approximate entropy, underlying the importance of sampling and experimental conditions to such complex measures which are sensitive to these changes. However, the first minimum of the average mutual information and autocorrelation function do not exhibit these properties and appear widely different between the samples.

Table 4-31: Three chosen variables tracked across multiple days in three nondiabetics with normal meals and six nondiabetics with continuous intravenous infusion. A correlation does exist between the two days within individuals but the correlation coefficient is not very strong.

	Apen (Day I)	(Day II)	Autocorrelation 1st minimum (Day I)	(Day II)	AMI 1st minimum (Day I)	(Day II)
Sernorm1	.30	.32	40 min	60 min	45 min	65 min
Sernorm2	.39	.28	65 min	70 min	55 min	70 min
Sernorm3	.43	.44	70 min	55 min	45 min	65 min
Polnorm1	.58	.62	95 min	45 min	140 min	160 min
Polnorm2	.58	.55	55 min	15 min	80 min	120 min
Polnorm3	.54	.63	70 min	50 min	120 min	80 min
Polnorm9	.48	.64	25 min	10 min	80 min	100 min
Polnorm10	.45	.50	10 min	10 min	60 min	60 min
Polnorm12	.65	.59	10 min	35 min	140 min	60 min

IV.H Summary of observation

- The attractor geometry is asymmetric, and contains possible structures that can be studied with larger numbers of similar data sets. For example, there appear to be higher negative rates of change at higher glucose levels and higher positive rates of change at lower glucose levels. This confirms the fundamental

nonlinear nature of the dynamics which try to restore the value to the equilibrium, and do so with greater “effort” as values reach the extremes.

- We observe that 90% of the samples fall within 2%-5% of the attractor area. That is to say the system spends majority of its time inside a small area within its attractor. This again is not surprising, however, it allows us to quantify what is known intuitively about this system: the majority of the time is spent in a confined glycemic boundary. The symmetry measurement used appears to vary widely between datasets and thus is likely to be a poor differentiating tool and requires further study.
- Looking at the rates of change in continuous feeding without meal perturbations(figure 4-9, table 4-5), a much more compact distribution is noted. A range of approximately -2 to 2 is observed (in comparison to the normal feeding range of -2.3 to 3.8 (table 4-3 and figure 4-3). Notably, despite the fact that meals are removed from set of perturbations, the rates of change, and in particular the maximal rates of fall, are not as different as one might have expected.
- Furthermore, the observation can be made that the skewness for the infusion sets are now negative whereas in the fasting state the distribution remains positively skewed for rates of change. This is markedly different than the data sets containing meal and exercise responses. Looking at the phase space geometry (figure 4-11 and table 4-6) we see that the compactness of the attractor has remained fairly high with most samples falling within 7% of the attractor size, despite containing a significant amount of data.

- Time-scale analysis of data sets with very high sampling rates (2 mins) leads to the conclusion that very high frequency oscillations constitute a very small portion of the signal energy content, particularly in the context of the presence of larger perturbations such as meals and infusions.
- Time-scale analysis yields multiple peaks, including a peak in the time scale of meals (~6 hours), a peak corresponding to the duration of the meal perturbation (~120-150 minutes) and a peak at ~100 minutes which may come from intrinsic oscillatory behavior as seen in constant infusion or fasting.
- We can quantify this with the ratio of the energy accounted for by the meal periods to the time accounted for by meal periods. This ratio appears to be significantly greater than one, ranging from 1.8 to 4.9, which we summarize by stating that the meal periods contain about 2-5 times as much energy as would be expected given their duration.
- The frequency content is not constant in the time-series including meals (easily explained by the nonconstant frequency of meals and activities) but also in the constant feeding which is a bit more surprising.
- A circadian component with an amplitude of approximately 20 mg/dl is observed in fasting in nondiabetics.
- An ultradian component with a frequency of approximately ~100 minutes can be observed with a small amplitude in fasting (~5 mg/dl).
- Continuous infusion of IV glucose contains larger scale dynamics in addition to the ~100 min time-scale observed in the continous enteral feeding case.

- These larger time-scales can be removed yielding pulses of significantly varying duration 76-175 minutes and amplitude of 7-16 mg/dl.
- Insulin infusion shows similar findings to the IV infusion with two time-scales of dynamics, on a pulsatile behavior of 77-99 minutes and another nearly 600-700 minutes.
- Entrainment with pulsatile infusion is noted an ranges from 40-80% of the energy band.
- Evidence for nonlinearity exists in two of the time-series studied but the conclusion cannot be made based on firm computational grounds.
- Nonlinear dynamics analysis will require significantly longer samplings to allow for attractor characterization.
- Meals can be detected using simple algorithms with fairly good accuracy.
- Significant variation exists between meal parameters even in the same individual with the same meal between two days.
- One average, the time to maximum rise velocity in the meals is shorter than the time to maximum fall velocity.
- The average meal profile presents a sigmoid rise and fall process.
- Correlation between the time-scales of dynamics existed between day one and day two of time-series from nondiabetics, but the degree of the correlation was inconsistent. This correlation was computed in the frequency space using the coherence function.

- Statistics of rates of change showed differences between individuals with some individuals showing good intra-day correlation while others did not. The same inconsistency was seen in the statistics of the shapes of the attractors.
- Measures of complexity were inconsistent between individuals from day one to day two.

Chapter V: Type I diabetes

Type I diabetes is characterized by a generally rapid loss of beta cell function leading to little or no insulin being secreted in response to elevated blood glucose. Without treatment, these individuals will generally perish from unchecked high glucose concentration. With insulin injections and proper diet and monitoring, they can often live fairly long lives but the impact in terms of difficulty of life is tremendous[72]. The DCCT established that the occurrence and rate of progression of the complications associated with diabetes are highly correlated with the control of blood sugars as estimated using HbA1c. From the curve associating complications and HbA1c levels, it appears that there is no point at which the benefit of lowering blood glucose is not realized, thus essentially setting the target as glucose levels as “low as possible”[73, 74] . Attempts to realize aggressive treatment are almost always universally blocked by increased incidence of hypoglycemia[75-79]. Hypoglycemia is accompanied by loss of awareness and coma, and is very much a deterrent to aggressive treatment. This matter is made more complex by the tendency for decrease sensitivity and the reduction of the person’s awareness of oncoming hypoglycemia, making the likelihood of acute events more pronounced after a few years of treatment. All these problems lay as a barrier to the effective control of type I diabetic patients (and to some extent type II diabetic patients, particularly ones who have insulin injections).

V.A Rates of change

A look at the phase portraits of the type I diabetic subjects reveal larger excursions of blood glucose as would be expected, in comparison to nondiabetic subjects (figure 5-1) . This is reflected in the number of samples in the higher glucose values in the phase portrait. Additionally, points are also distributed in a wider space in the rate of change axis. Overall, the type I diabetic occupies a larger “space” in the phase portrait in both dimensions. Interestingly, while the well controlled diabetic subgroup does show larger excursions in the value of blood glucose, the distribution of the rates of change are comparable to the nondiabetics (Table 4-1). Also of note, as shall be examined next, the maximum rates of change in nondiabetics and non-well controlled diabetics are similar. Similar observations are not noted in the DirecNet data set (Figure 5-2), which represents the clinical data set composed of many children (see chapter 2 for a detailed explanation of the data set). Based on figure 5-1 it appears that diabetic subjects have a larger range of values but not necessarily consistently larger rates of change, and that the well controlled diabetic subjects have lower rates of change than the either nondiabetics or poorly controlled diabetics. The direcnet data, however, does not support the conclusion of similarities in rates of change between nondiabetics and type I diabetic children. It remains unclear what differences (including age, sampling, experimental set-up, population) may have led to this difference.

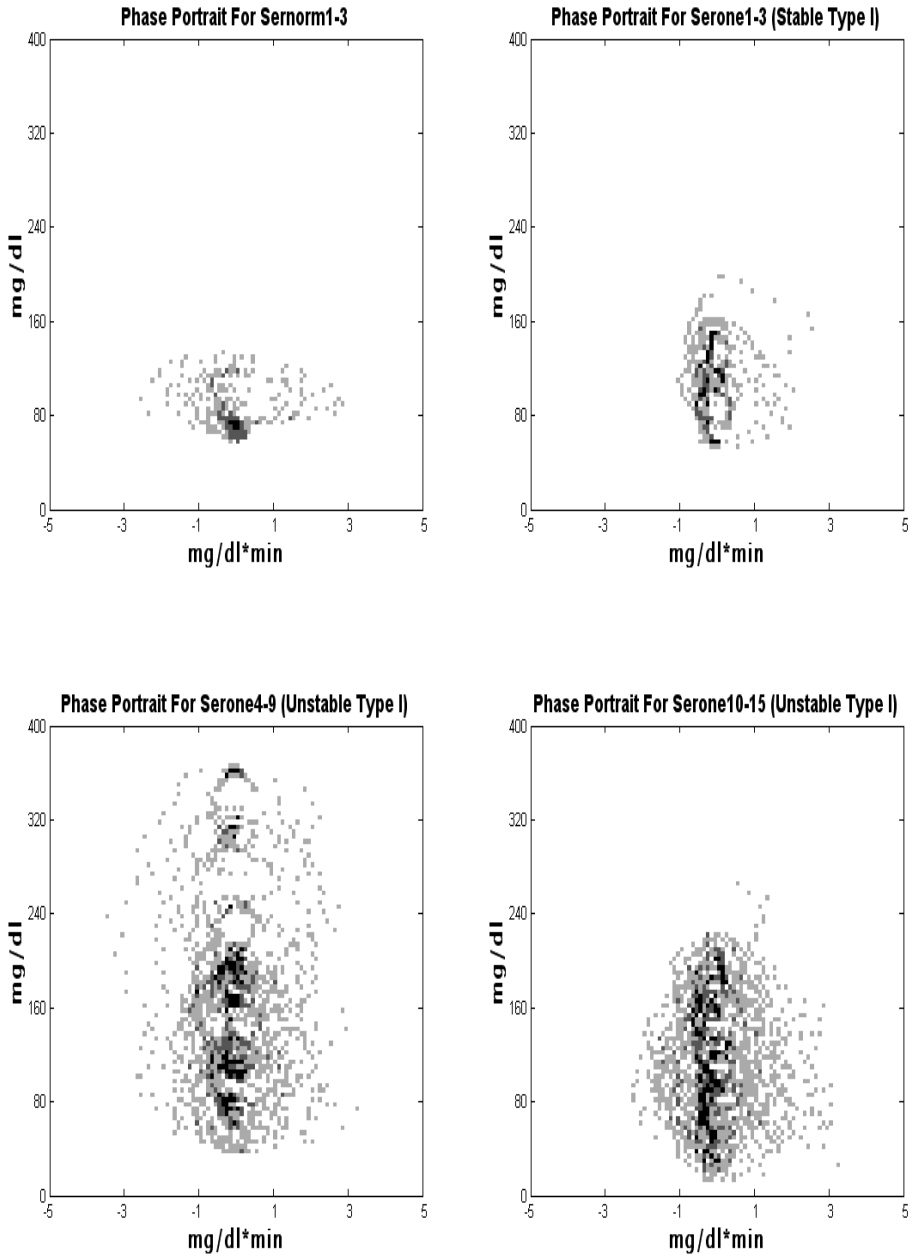


Figure 5-1: A phase portrait from three nondiabetic individuals (upper left corner) and three phase portraits for stable type I diabetics, unstable type I diabetics and the same six unstable type I diabetics with a four insulin doses instead of two (more spread out insulin injections). The maximal rates of change remain fairly bounded, while the distribution changes characteristics.

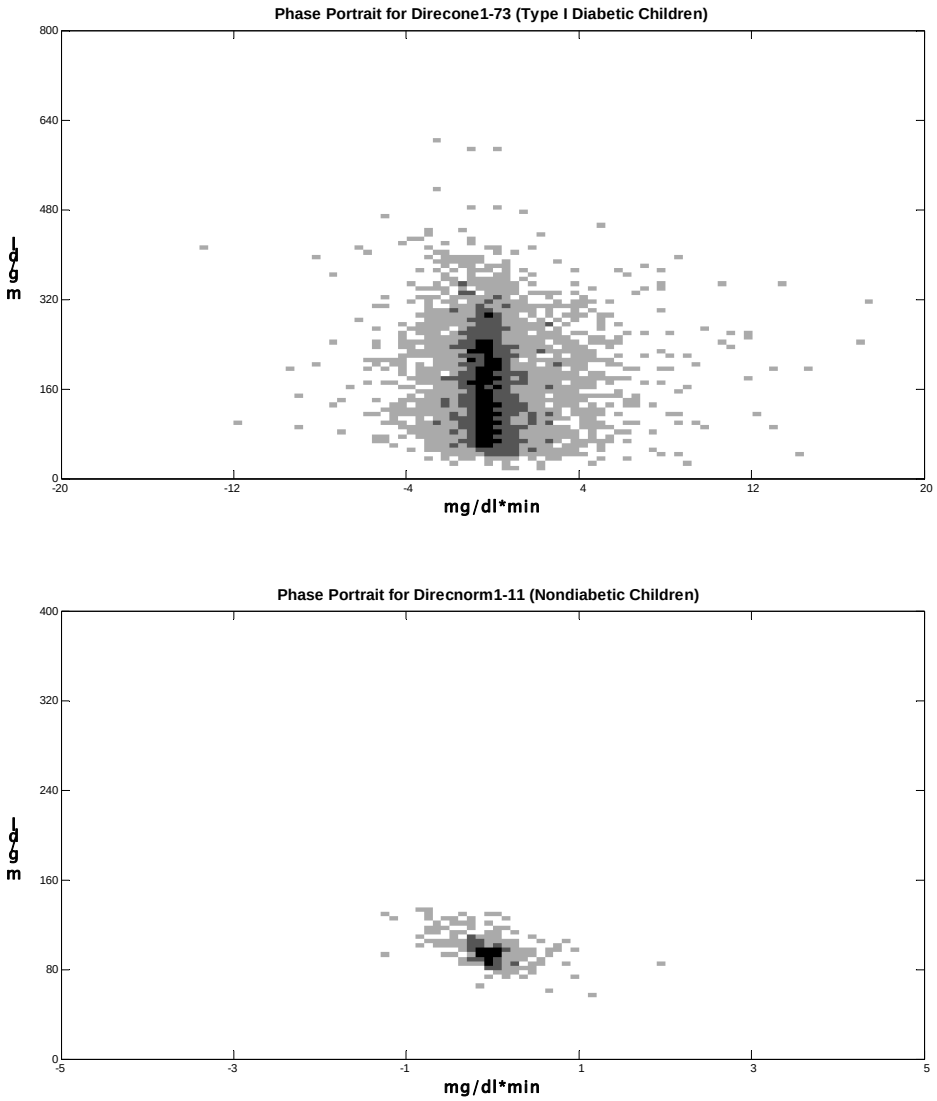


Figure 5-2: A phase portrait from the type I diabetic children from the DirecNet Study (top) and the nondiabetic children (bellow). Note the similarity in the differences between these two groups versus the literature data sets.

The rates of change are expected to be elevated for type I diabetics as it has been noted that diabetic dynamics include larger excursions of blood glucose. Surprisingly, however, in the data sets analyzed, there was not a systematic, significant increase in the maximal rates of change in the type I diabetic subgroup as compared to the nondiabetic subjects (table 5-1). In the DirecNet data, however, there

was a very large difference between the two population groups, with maximal rates of change well into the range of 10 mg/dl*min. Another visible finding when comparing each subgroup is the decrease in the Kurtosis of the rate of change (compare to Table 4-1) between the nondiabetic subjects, the well controlled diabetic subjects and the poorly controlled subjects. This Kurtosis reflects the “flatness” of the distribution. Once again, in the DirecNet data set, this relationship is reversed (compare to Table 4-2 and 4-3). Skewness of the rate of change remains positive for type I diabetic data-sets as it was in the nondiabetic data sets.

Table 5-1: Rates of change characteristics from multiple type I diabetics, including pregnant type I diabetics (sanpregone2,6,8) as well as the clinical DirecNET data which spans 73 children with type I diabetes).

Data Set	Max (-) Rate	Max (+) Rate	Mean Rate	Skewness	Kurtosis
Serone1	-1.0	2.7	.4	1.7	7.0
Serone2	-1.5	2.6	.4	1.1	6.6
Serone3	-1.6	3.1	.4	1.0	5.7
Serone4	-3.4	3.3	.7	.05	4.3
Serone5	-2.6	2.3	.6	-.2	4.2
Serone6	-1.8	3.0	.6	.7	4.0
Serone7	-1.2	2.4	.5	1.2	4.9
Serone8	-1.9	2.6	.5	.3	3.8
Serone9	-4.9	3.7	.9	.15	4.2
Serone10	-2.2	3.3	.7	.9	3.8
Serone11	-1.8	3.1	.5	.7	4.8
Serone12	-2.2	3.2	.6	.8	4.9
Serone13	-1.9	2.1	.5	.2	3.5
Serone14	-2.2	2.1	.6	.3	3.3
Serone15	-4.4	4.4	1.1	.6	3.4
Mir1	-3.9	4.1	.9	.7	3.9
Mir2	-3	4.3	.8	.7	4.7
Mir3	-3	2.6	.7	-.6	3.6
Mir5	-3	3.3	1.0	.4	2.5
Sanpregone2	-2.3	3.2	.6	.8	3.9
Sanpregone6	-3.4	7.8	.6	2.2	14.5
Sanpregone8	-3.7	2.0	.3	-.7	11.2
DirecNet (3701 Samples)	-13.2	25.8	.3	1.8	15.3
All Sets (95 sets, 22193 samples)	-13.2	25.8	.05	2.3	32
Mal1-12 (356 Samples)	-2.3	3.8	.02	.9	9.3
Mej1-5 (239 Samples)	-1.6	1.7	0	.9	7.5
Direcnet (436 Samples)	-2.1	2.0	-.01	0	9.3
Combined (Above)	-2.3	3.8	0	.86	11.6
All 39 data set combined (1871 Samples)	-2.5	3.8	0	.9	11

Attractor characteristics are significantly different between type I diabetics and nondiabetics. In particular, the area of the attractor is significantly larger for type I diabetics versus nondiabetics in all data-sets, as would be expected. Interestingly, not all the contribution comes from the increased range of glucose values (the y-axis) in the case of the DirecNet data as noted. Once again the progression between well controlled diabetics (serone1-3) and poorly controlled diabetics are noted. Interestingly, the compactness is highest in the well-controlled type I data set, reflecting the restricted range for rates of change. Symmetry failed to show consistent changes between subgroups.

Table 5-2. Summary of the attractor geometric characteristics for the various data sets from type I diabetics including pregnant type I diabetics.

Data Set	Attractor Area	Compactness	Symmetry
Serone1	909	4.40%	0.88
Serone2	876	6.10%	0.82
Serone3	943	5.80%	0.94
Serone4	3658	1.20%	0.82
Serone5	2835	1.20%	0.77
Serone6	2751	1.50%	0.83
Serone7	2029	1.70%	0.8
Serone8	1976	2.60%	0.88
Serone10	1663	3.10%	0.78
Serone11	1468	4.00%	0.95
Serone12	1586	3.50%	1
Serone13	2456	1.40%	0.86
Serone14	3768	1.20%	0.8
Serone15	4302	1.50%	0.73
Mir1	3307	1.40%	.66
Mir2	3325	.8%	.52
Mir3	2779	1.40%	.7
Mir5	654	8.5%	1
Sanpregone2	2039	2.20%	.92
Sanpregone6	610	12%	1
AVERAGE:	2200	3.3%	.833

V.B Time-scales

The spectral estimates were computed for groups of type I diabetics. The type I diabetics from the Service experiment (serone1-15) were analyzed using the FFT algorithm. The results are shown in figure 5-3. In the case of serone1-serone3 data sets, which are well controlled diabetics, spectral content contains the peaks near 2 hours and 5 hours just as the nondiabetic dynamics. The small peak near 100 is not as apparent, as would be expected because the lack of beta cell function would lead to the cessation of the ultradian dynamics discussed in chapter IV. Serone10-serone15 represent six diabetic patients characterized as unstable by their physician, with four insulin injections throughout the days and showing less wings in blood glucose values. Serone4-serone9, displays the same patients but with only two injections. A progression is seen in the shifting of the time-scales of dynamics towards longer periods from the nondiabetic individual, to the type I diabetic individual with poor control and only two insulin shots. This is well illustrated in figure 5-3, which is summarized in figure 5-4.

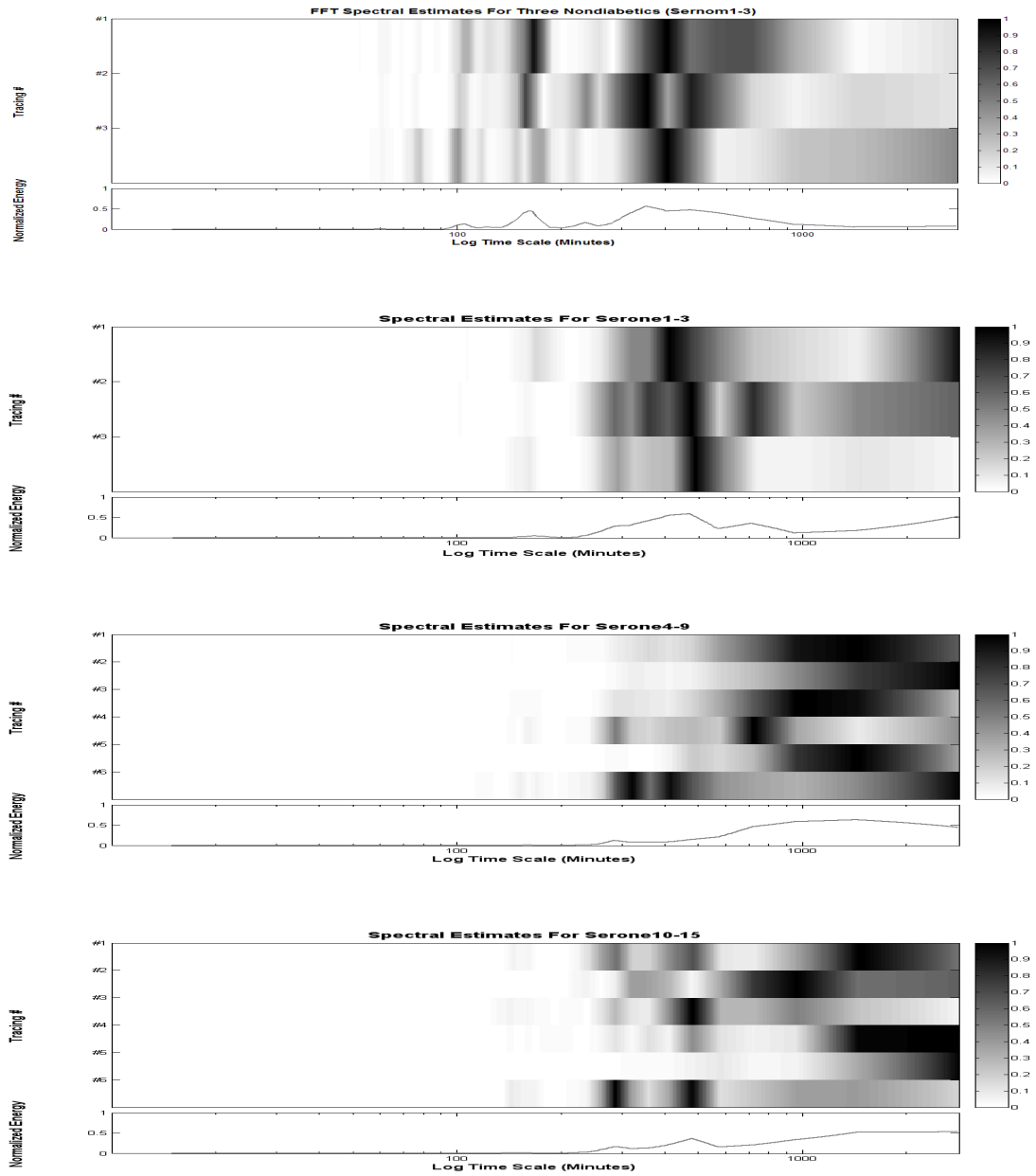


Figure 5-3: FFT based spectral estimates of nondiabetic (top), well controlled diabetic (second panel), poorly controlled diabetic (third panel) and poorly controlled diabetics with increased number of insulin injects (bottom). A progression is noted towards slower time-scales with partial restoration in the last panel where more insulin shots are used to improve control.

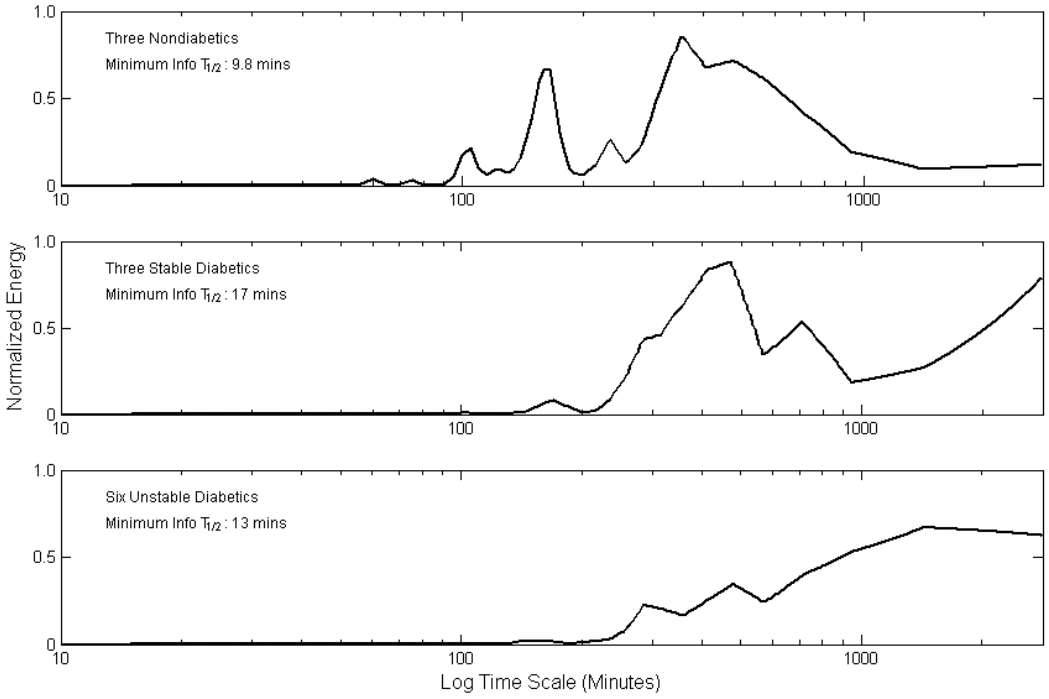


Figure 5-4: Summarized results from figure 5-3. The top panel shows average of the FFT Spectrums from the three nondiabetics in the Service data set (sernorm1-3), the middle panel shows the same for three well-controlled diabetics and the bottom panel reflects six poorly controlled ("unstable) diabetics. Note the progression towards longer time-scales.

V.C Complexity and nonlinear analysis

The measures of complexity were taken for a group of type I diabetics and are shown in the table 5-3. Measures of entropy ranged from 3.5 to 3.9 for nondiabetic subjects but was much higher for diabetic subjects (4.6-5.6). This difference may be cause largely by the statistical difference in the signals as discussed in the methods section. Looking at approximate entropy and sample entropy, which are much less statistically driven (See methods), the nondiabetic approximate entropy ranged from .35-.51 while the diabetics ranged from .22 to .40. Unfortunately, the limited size of the nondiabetic population prevents significant generalization of the results (Because

of the fact that entropy measurements are easily effected by sampling periods, it is not generally recommended to compare measures between different experiments). However, it appears that the signal complexity is increased in the nondiabetic individual as a population group. The same observation is noted in the case of the median sample entropy where the nondiabetic range is .21-.36 whereas the diabetic range is from .16 to .29. This data is noted in table 5-3.

Table 5-3: Three different measures of entropy as a way of assessing signal complexity, applied to data collected from a single experimental protocol involving normals and type I diabetics. The first three data sets are normals and the rest are type I diabetics with various degrees of control. The first column shows the entropy of the signal, while the other two columns show the approximate entropy which is another form of entropy measurement and the median sample entropy which is yet another measurements.

Data Set	Entropy	Apen	Median Sampent
Sernorm1	3.9	.35	.21
Sernorm2	3.6	.36	.24
Sernorm3	3.5	.51	.36
Serone1	4.7	.33	.24
Serone2	4.7	.31	.23
Serone3	4.6	.38	.27
Serone4	5.6	.23	.17
Serone5	5.3	.21	.2
Serone6	5.3	.25	.18
Serone7	5.1	.3	.23
Serone8	5.3	.22	.16
Serone9	5.6	.34	.19
Serone10	5.2	.37	.26
Serone11	5.1	.3	.22
Serone12	5.0	.4	.25
Serone13	5.2	.29	.22
Serone14	5.5	.16	.18
Serone15	5.6	.37	.29

Estimates of dimension show a very subtle decrease in the average dimensional estimates of the less controlled type one diabetics versus nondiabetics, particularly in the case of Cao's embedding dimensional estimates (table 5-4). Estimates of the largest Lyapunov exponents are markedly different for type I diabetics and nondiabetics (table 5-5), with both the maximum estimate and the mean estimate being significantly different. Notably, unlike in the rates of change, the well controlled diabetics are not as similar to the nondiabetics as they appear in rates of change and time-scale analysis. This is likely to reflect the multi-scale presence of dynamics in the healthy, nondiabetic time-series.

Table 5-4: Dimensional estimates for type I diabetics. Average mutual information (AMI) was used to measure a lag. This was done by measuring the first delay in the firsm minimum of the cross mutual information between the series and itself. The embedding dimensions (columns 3 and 4) are computed using two different methods. The last two columns list the estimates of the lyapunov exponents (both the max and the mean) based on the previous columns values.

	AMI	False Nearest Neighbor's Dimensions	Cao's Method Embedding Dimension
Sernorm1	18	4	3
Sernorm2	23	3	4
Sernorm3	13	3	5
Serone1	20	3	3
Serone2	14	2	3
Serone3	18	3	4
Serone4	30	3	3
Serone5	24	3	4
Serone6	21	3	3
Serone7	22	3	3
Serone8	25	3	3
Serone9	36	3	2
Serone10	15	3	4
Serone11	20	3	3
Serone12	12	3	3
Serone13	15	3	3
Serone14	18	3	3
Serone15	15	4	3

Table 5-5: Two estimates of the largest Lyapunov exponent one based on the maximum and the other based on the mean slope of the error estimate along the fastest direction of error growth, once the time-series has been embedding using the number of dimensions required (in this case using cao's estimate).

	Largest Lyapunov Exponent (Max Estimate)	Largest Lyapunov Exponent (Mean Estimate)
Sernorm1	.53	.32
Sernorm2	.47	.24
Sernorm3	.34	.20
Serone1	.17	.06
Serone2	.16	.07
Serone3	.38	.18
Serone4	.9	.05
Serone5	.41	.2
Serone6	.42	.19
Serone7	.24	.1
Serone8	.05	.03
Serone9	.26	.13
Serone10	.16	.07
Serone11	.12	.06
Serone12	.15	.08
Serone13	.09	.04
Serone14	.08	.04
Serone15	.19	.1

V.D Meals

An analysis of the importance of the meal event to the signal energy in type I diabetics can be performed in a similar way as was performed in the case of nondiabetics. The results can be seen in table 5-6. Although the amount of time that is spent in the meal episode is similar to the ones observed in nondiabetics, the signal energy content during those meal periods no longer accounts for as significant of the total signal energy content. This is likely to occur in part because of the increase in length of the system's response to the meal and in part because of the introduction of the time-scales of insulin dynamics into the system.

Table 5-6. Percentage of energy during meal-times computed for type I diabetics. In the second column, the percentage of time which is considered "meal time" is calculated by considering meal time as the two hour interval after each meal event. The percent of signal energy (defined by the mean subtracted amplitude squared) during this period is shown in the third column, and the same quantity is shown in the fourth column for the non meal times. The last column shows a ration of the former versus the latter.

Data Set	% Time Meals	% Signal energy Meals	% Signal energy Not Meals	Ratio (Meal Vs. Nonmeal)
Serone1	39%	36%	64%	.56
Serone2	40%	46%	54%	.85
Serone3	41%	49%	51%	.94
Serone4	44%	30%	70%	.44
Serone5	43%	58%	42%	1.38
Serone6	41%	59%	41%	1.44
Serone7	39%	51%	49%	1.06
Serone8	44%	37%	63%	.59
Serone9	39%	40%	60%	.66
Serone10	40%	39%	61%	.64
Serone11	42%	58%	42%	1.37
Serone12	40%	49%	51%	.98
Serone13	39%	36%	64%	.57
Serone14	40%	36%	64%	.56
Serone15	41%	39%	61%	.65
Mir5	36%	22%	78%	.28
Mir6	36%	49%	51%	.95
Mir8	28%	39%	61%	.63
Average	40%	43%	57%	.81

The progression for the meal event behaving as a pulse of signal energy to the meal event being less and less clear as the major contributor of dynamics can be visualized in the progression shown in figure 5-5.

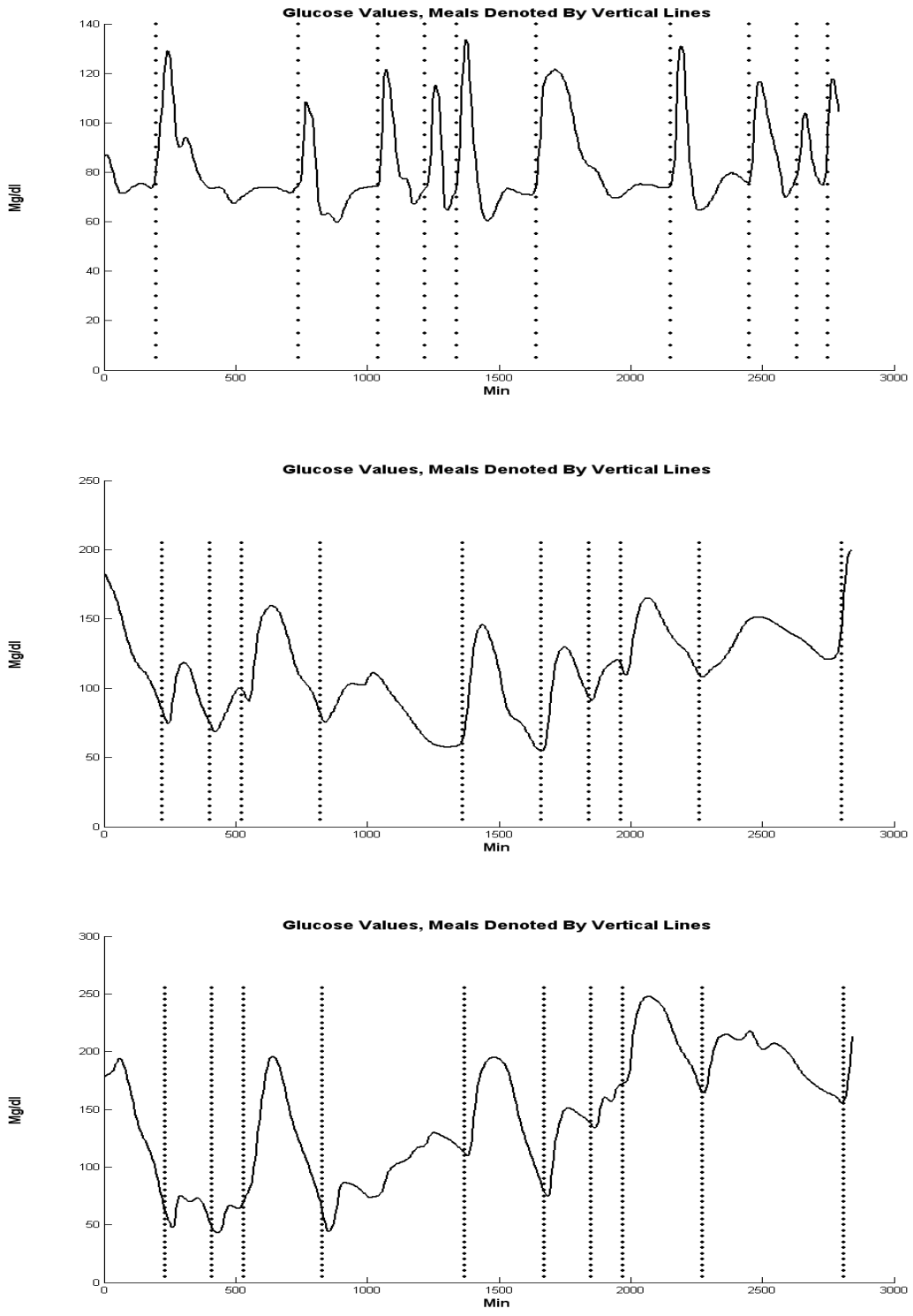


Figure 5-5: Meals are less localized and accompanied by other time-scales of dynamics in type one diabetes (middle and bottom tracing) as compared to a nondiabetic. The bottom figure represents an unstable diabetic while the middle figure represents a stable one.

Both the amplitude and the time-scale of the meal are thus increased. An average of the signal following a four hour period (or less if another meal occurred within 4 hours) is shown in figure 5-6, for nondiabetics, stable diabetics and unstable diabetics. Both the amplitude as well as the time-scale of meal response are increased as the degree of control is lessened in these cases.

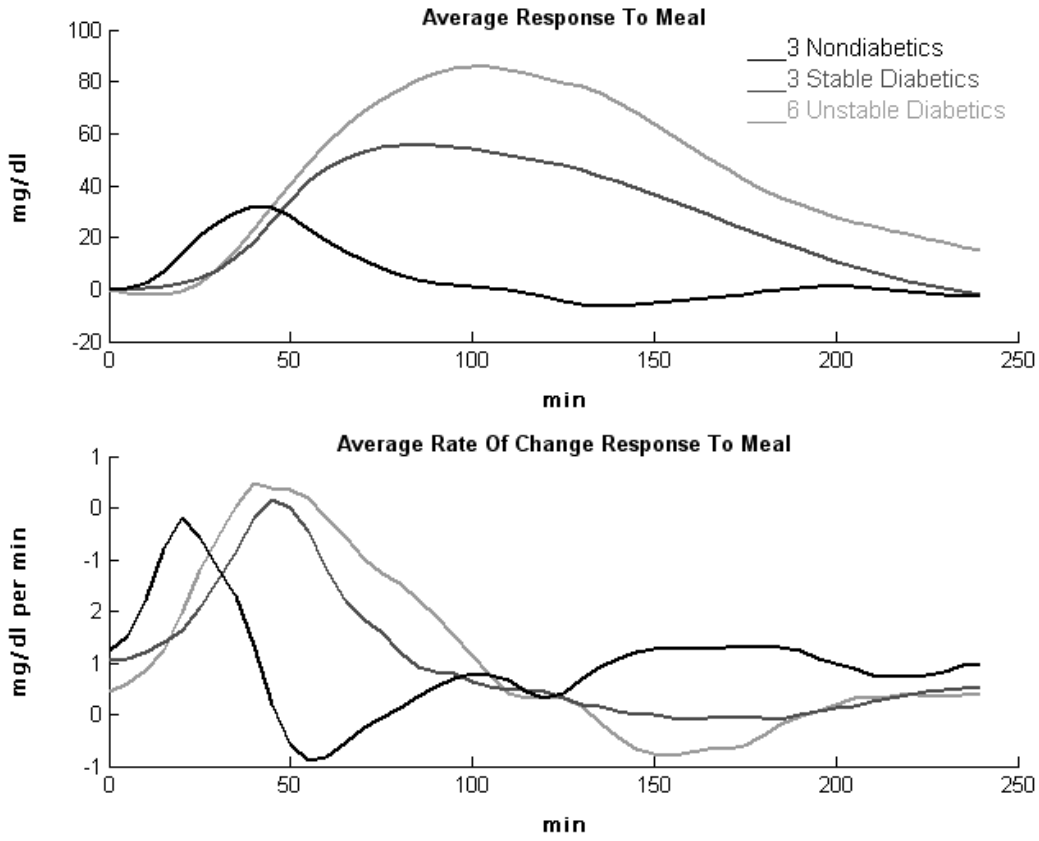


Figure 5-6: the response to a meal in terms of blood glucose values as well as the rate of change assessed in three patient groups.

As was done with the nondiabetic case, one may also examine the normalize shape of the rise and fall of the glucose values during the time series. As was done in the previous sections, the time-series were scanned for large rise and falls in glucose

values. These are then averaged and normalized to yield a characteristic nondimensional representation of the rise and fall of glucose values. The results are shown in figure 5-7. The rise and fall process has lost a significant amount of similarity to an s-shaped process and now appears to be much more linear.

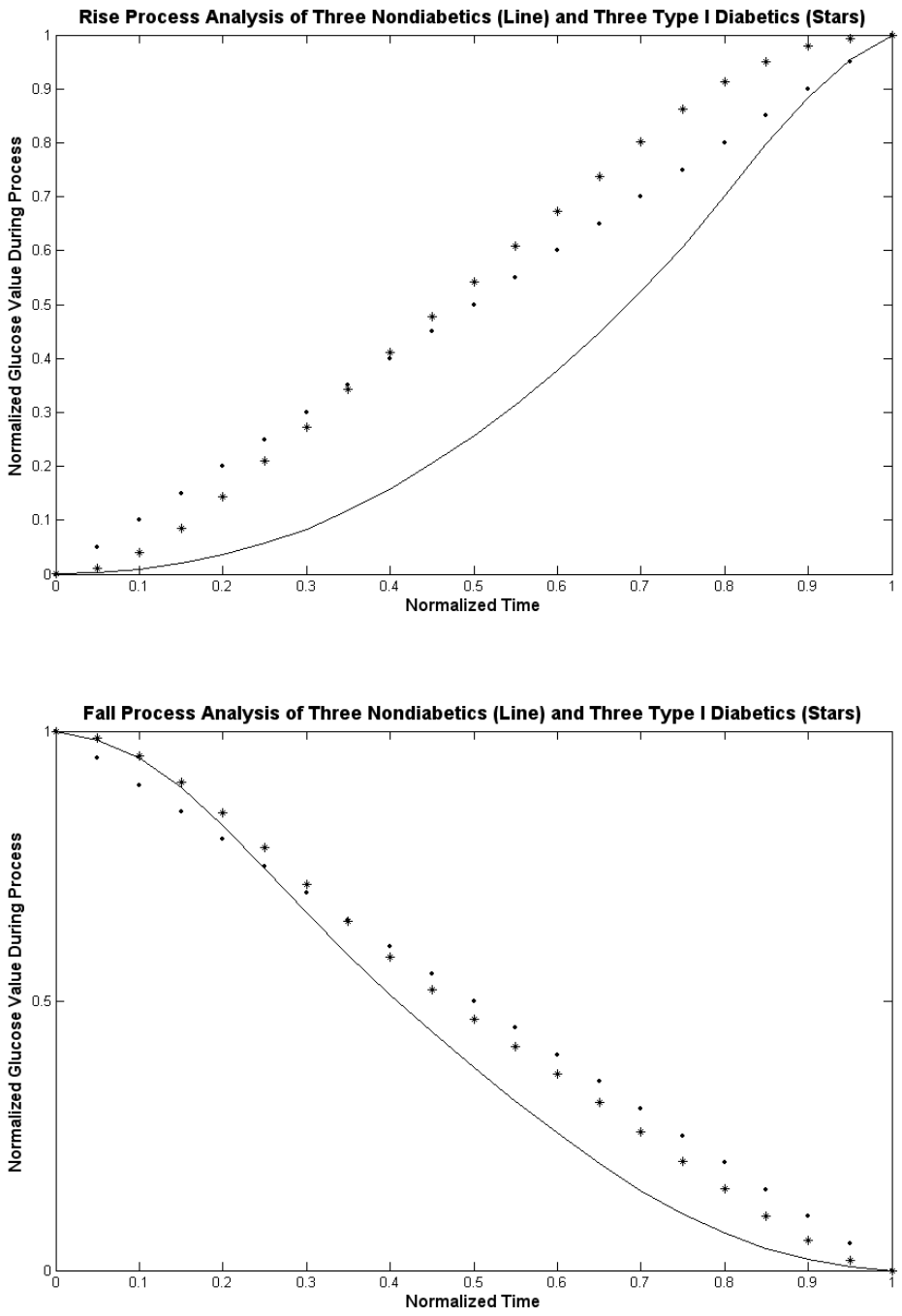


Figure 5-7: Meal process analysis for nondiabetic individuals (solid line) and three type I diabetics (stars). A straight line (circles) is provided as a linear reference. The diabetic process is much less similar to the S curve of the nondiabetic than a simple line.

V.E Individualized dynamics

It is once again of interest to question whether measures of dynamics remain similar in an individual from day to day. This can be assessed by looking at the relationship between the measurements between two days. Figure 5-8 shows two time-scale analysis plots for each day in 15 type I diabetic subjects from the Service data set. The summed spectrum appears similar as do many of the specific strips, but noticeable differences are apparent with a cursory glance.

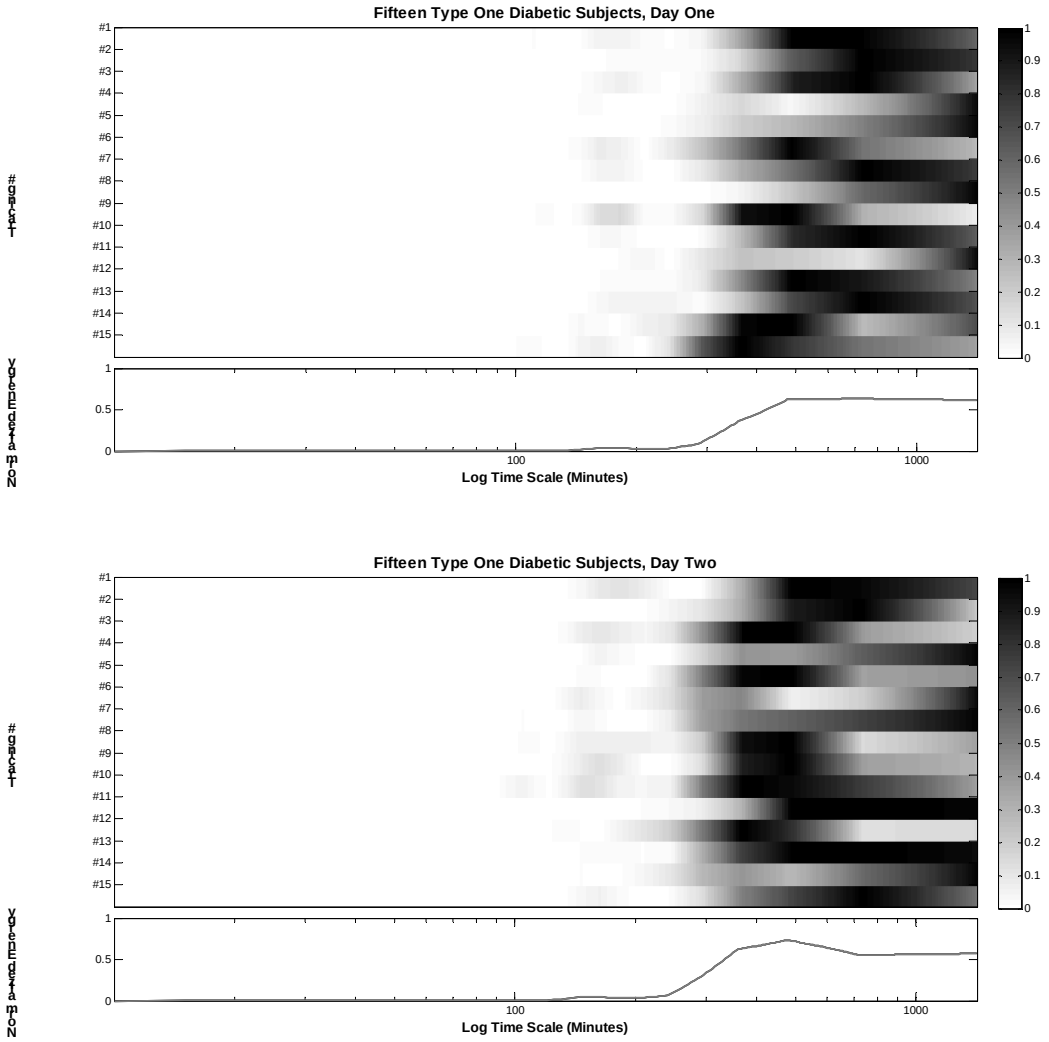


Figure 5-8: FFT spectral estimates for 15 diabetics on day one (top) and day two (bottom). Notice the similarities between the overall profile though it is clear that certain individuals display significant variations in their daily frequency content.

One method to quantify the similarities is to calculate the correlation coefficient in the frequency space, a function otherwise known as the coherence. The result of this analysis is shown in table 5-7. As was noted in the case of nondiabetic individuals, significant correlation exists between the two frequency contents but there are significant variations between individuals in the extent of this correlation which does not appear to be different for different subgroups.

Table 5-7: Frequency dimension correlation coefficient between each day of the time-series from type I diabetics, computed by FFT. This is otherwise known as the coherence function. Significant agreement exists for most data-sets, though some show notable differences (Serone8 and Serone5).

Serone1	.99
Serone2	.85
Serone3	.73
Serone4	.91
Serone5	.57
Serone6	.83
Serone7	.86
Serone8	.50
Serone9	.98
Serone10	.79
Serone11	.74
Serone12	.66
Serone13	.93
Serone14	.76
Serone15	.80

In looking at the measures of complexity (table 5-8), as was the case in nondiabetics, little consistency was noted in the measurements of information dissipation such as the autocorrelation function or the AMI between each day. Values for approximate entropy, however, showed fairly good agreement from day to day in most of the individuals. The significance of this agreement in the light of the general variability of the information loss functions is unclear.

Table 5-8: Comparison of three metrics in two different (consecutive) days in type I diabetics.

	Apen (Day I)	(Day II)	Autocorrelation 1st minimum (Day I)	(Day II)	AMI 1st minimum (Day I)	(Day II)
Serone1	.57	.61	95 min	75 min	90 min	70 min
Serone2	.55	.52	105 min	120 min	65 min	75 min
Serone3	.54	.62	85 min	80 min	55 min	75 min
Serone4	.49	.60	230 min	140 min	85 min	80 min
Serone5	.43	.49	225 min	85 min	70 min	75 min
Serone6	.62	.57	135 min	155 min	80 min	80 min
Serone7	.30	.35	120 min	130 min	55 min	75 min
Serone8	.28	.31	170 min	110 min	90 min	110 min
Serone9	.38	.34	90 min	80 min	85 min	60 min
Serone10	.25	.38	100 min	85 min	65 min	85 min
Serone11	.40	.31	175 min	105 min	60 min	70 min
Serone12	.22	.28	75 min	95 min	70 min	70 min
Serone13	.29	.32	90 min	140 min	75 min	90 min
Serone14	.2	.26	110 min	135 min	95 min	125 min
Serone15	.36	.36	90 min	100 min	45 min	65 min

Finally, table 5-9 and 5-10 show the rates of change and attractor characteristics for the same group from the Service data set. Once again, although some correlations exist between some of the metrics (in particular maximal rates of change, and attractor area) the correlation between the values is fairly inconsistent, as was the case in the nondiabetics.

Table 5-9: Statistics of rates of change from fifteen type I diabetic subjects between two different days.

Set	Max ROC day I	Max ROC day II	Min ROC day I	Min ROC day II	Skew ROC day I	Skew ROC day II	Kurt ROC day I	Kurt ROC day II
Serone1	2.2	2.7	-.8	-1.0	1.6	1.9	6.0	8.0
Serone2	2.6	1.6	-.8	-1.5	1.9	0.0	8.4	3.6
Serone3	1.8	3.1	-1.0	-1.6	1.0	1.0	3.9	6.1
Serone4	2.7	3.3	-3.4	-2.6	-.3	.2	5.8	3.3
Serone5	2.3	2.2	-1.5	-2.6	.4	-.4	3.8	3.9
Serone6	2.8	3.0	-1.6	-1.8	.7	.7	4.0	4.0
Serone7	2.3	2.4	-1.2	-1.1	1.0	1.5	4.2	5.7
Serone8	2.2	2.6	-1.9	-1.5	.2	.5	3.4	4.3
Serone9	3.0	3.7	-2.8	-5.0	.6	-.2	3.0	4.7
Serone10	3.2	3.0	-2.2	-1.6	.5	1.2	3.7	3.8
Serone11	3.1	2.4	-1.8	-1.6	.7	.7	4.4	4.9
Serone12	2.8	3.2	-2.2	-1.8	.6	1.0	5.3	4.5
Serone13	2.0	2.1	-1.9	-1.9	.3	.2	3.2	3.8
Serone14	1.9	2.1	-1.8	-2.2	.1	.6	3.6	3.3
Serone15	3.8	4.4	-2.6	-4.4	.6	.6	2.7	3.9

Table 5-10: Characteristics of the phase portrait from fifteen type I diabetic subjects between two different days.

Set	Attractor Area Day I	Attractor Area Day II	Attractor Concentration Day I	Attractor Concentration Day II	Symm Day I	Symm Day II
Serone1	158	228	.20	.12	1	1
Serone2	188	159	.14	.21	1	1
Serone3	159	268	.26	.14	1	1
Serone4	583	511	.05	.07	1	.9
Serone5	650	373	.04	.11	.9	.9
Serone6	669	527	.03	.07	.9	.9
Serone7	287	309	.11	.11	1	.9
Serone8	480	380	.07	.07	1	.9
Serone9	792	1214	.04	.03	.8	.9
Serone10	348	464	.50	.07	.4	.9
Serone11	413	273	.87	.15	.9	1
Serone12	343	346	.10	.13	.9	1
Serone13	328	336	.11	.11	1	.9
Serone14	436	466	.05	.60	1	1
Serone15	781	881	.04	.77	.9	.8

V.F Summary of observations

- In one experiment set, rates of change of blood glucose are more confined in well-control type I diabetics than in nondiabetic subjects, while they are similar between nondiabetics and non-well controlled diabetics.
- The DirecNet data sets, for unclear reasons shows much increased rates of change as high as 13 mg/dl * min. This may be due to populational characteristics, infrequent irregular sampling and different experimental set-ups.
- There is an overall decrease in the Kurtosis of the rate of change from nondiabetics to well controlled diabetics and finally to poorly controlled diabetics.
- Attractor size is increased from nondiabetic to well controlled diabetic to poorly controlled diabetic.
- Time-scales of dynamics are shifted towards longer time-scales of insulin injection, with well controlled diabetics showing faster time-scales of evolution.
- Measurements of complexity show decreased complexity in the nature of the signal from diabetics. This is exhibited in the measurements of approximate entropy which is the entropy measurement least effected by sampling and signal statistics as described in chapter 3.
- Nonlinear time-series analysis shows a decrease in the estimates of the largest positive Lyapunov exponent in the type I diabetics.

- Meals are larger, last longer and account for less of the total energy in the signal in type I diabetics, presumably because of the increased contribution of insulin dynamics to the signal intensity.
- Meal process models in type I diabetics show an increased linear profile in both the rise and fall of blood glucose levels in comparison to the S-shaped curves generated by nondiabetic profile.
- As was noted in the case of nondiabetic individuals, significant correlation exists between the signal time-scale content from one day to the next, but there is significant variations between individuals in the extent of this correlation which does not appear to be different for different subgroups.
- Little consistent correlation existed between measures of information half-life from day to day, but approximate entropy did show good correlation day to day. Statistics of rates of change and attractor geometry remain inconsistent between each day in type I diabetic subjects.

Chapter VI: Type II diabetes and other altered states

Type I diabetes amounts to the destruction of the insulin producing machinery in the human body. In type II diabetes, however, a gradual progression is noted which begins with decreased action of insulin at the target sites (cells uptaking glucose) [80]. This is thought to progress to a state high-insulin resistance coupled with hyperinsulinemia. Eventually the beta cells, which are the primary producers of insulin are thought to “burn out” and insulin production eventually falls off. Towards the latter stage of the disorder, insulin production may cease and type II diabetes may become more similar to type I diabetes except that it is generally accompanied by insulin resistance and associated dyslipidemias [81]. For a variety of reasons including the recent food supply increase in the western world, type II diabetes was not, until recently, a major problem but has now reached epidemic levels. For this reason the history of data collection and study of type II diabetes is much less well developed than type I diabetes. In the sections that follow the tools developed in the previous chapters will be used to study this disorder as well as a few other rare disorders. As the data is extremely scant, the analysis is particularly limited and is best thought of as illustrative of the methods described more so than any other part of this document.

VI.A Obesity, impaired glucose tolerance and aging

The study of the effect of aging and obesity on glucose dynamics is complicated by the correlation between aging, obesity and insulin resistance. Many

possible explanations exist as to this inter-relationship but at least in the case of aging and developing insulin resistance they are thought to be correlated in the general population outside of the diabetic subgroup. Additionally, some postulate that many consequences of uncontrolled diabetes essentially can be considered as accelerated aging. Obesity, while not always present, is also well correlated through multiple pathways with aging and impaired glucose tolerance. Increasing insulin resistance that occurs with aging and obesity is thought to be a central mediator of this correlation. Taking this point of view, individuals with high indices of obesity were also studied in terms of their glucose profile prior to and after weight loss [44]. Impaired glucose tolerance is diagnosed and defined by increased amount of time for the blood glucose to return to baseline levels following the administration of a standardized meal. In this way IGT is very similar to type II diabetes and differs only by the magnitude of the severity of the post-meal elevated glucose levels. It is also generally agreed [80, 82] that IGT is a pre-diabetic state and is part of the progression towards full pathogenesis of type II diabetes.

A data set was studied in which six individuals were monitored prior to and after weight loss under continuous IV glucose infusion for 12 hours. The sampling period of 15 minutes combined with short duration of the signal collection limited the utility of this data group. Two of the individuals were nondiabetic, while two were diagnosed with impaired glucose tolerance (IGT) and two had met criteria for type II diabetes. Results of two analysis methods, the analysis of time-scales and of rates of change are shown in figures 6-1 and 6-2. The phase space attractor portrait shows a

consistent reduction in blood glucose levels in the IGT and Type II with no such a change in the nondiabetic cases. There does appear to be an increase in the compactness of the attractor in the center which would be interesting to quantify. Maximal rates of change are not affected. There are no consistent findings in the time-scale analysis although at least in one nondiabetic subject (#2) a clear increase in the fast time-scale is noted, resulting in an increase in the ultradian peak near 100 minutes. Other strips show a variation in their response.

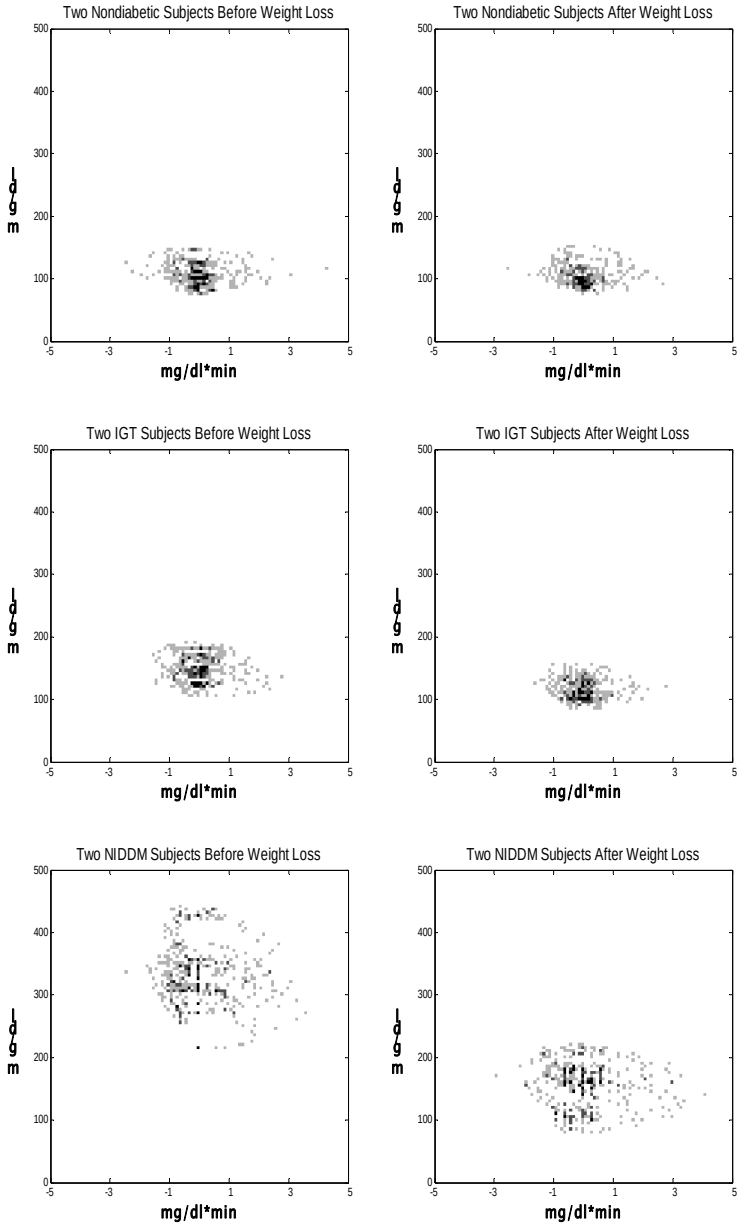


Figure 6-1: Phase space attractor profiles for three subject groups before and after weight loss. Note a general shift towards lower blood glucose values in both IGT and NIDDM groups but no change in the normal subjects.

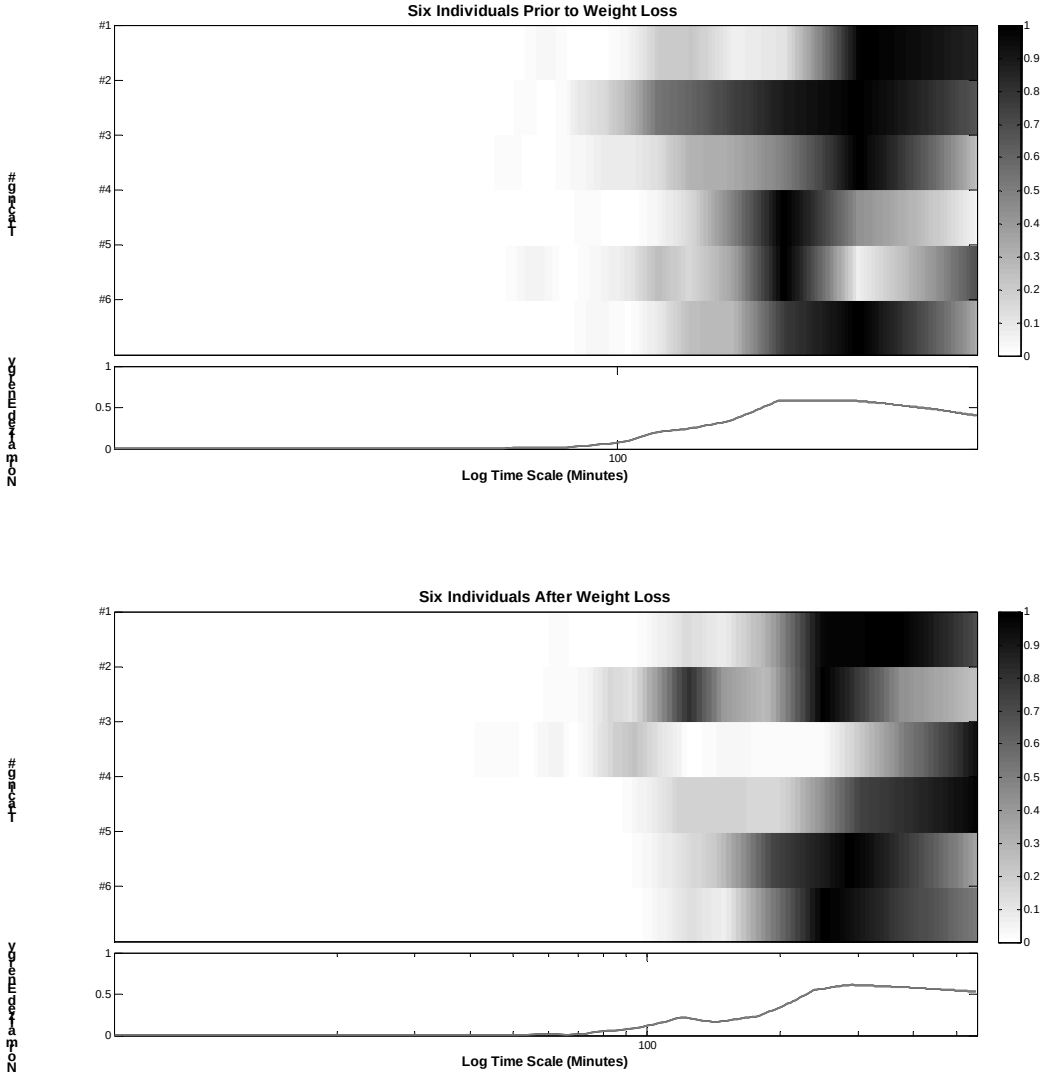


Figure 6-2: Time scale analysis of six individuals before (top panel) after weight loss (bottom panel). The first two strips in each panel are nondiabetics with the subsequent two strips obtained from IGT and the last two strips obtained from type II diabetics individuals.

Looking at the effect of age on the profiles of six nondiabetics with continuous enteral feeding, phase portraits of the two groups is shown in figure 6-3, and figure 6-4 shows the FFT based time-scale analysis. At least in these six subjects, there appears to be a decrease in the size of the attractor, as well a decrease in the long time-scales in the older subjects. As mentioned significant controlled data sets are lacking to make conclusions beyond sparking curiosity for further data.

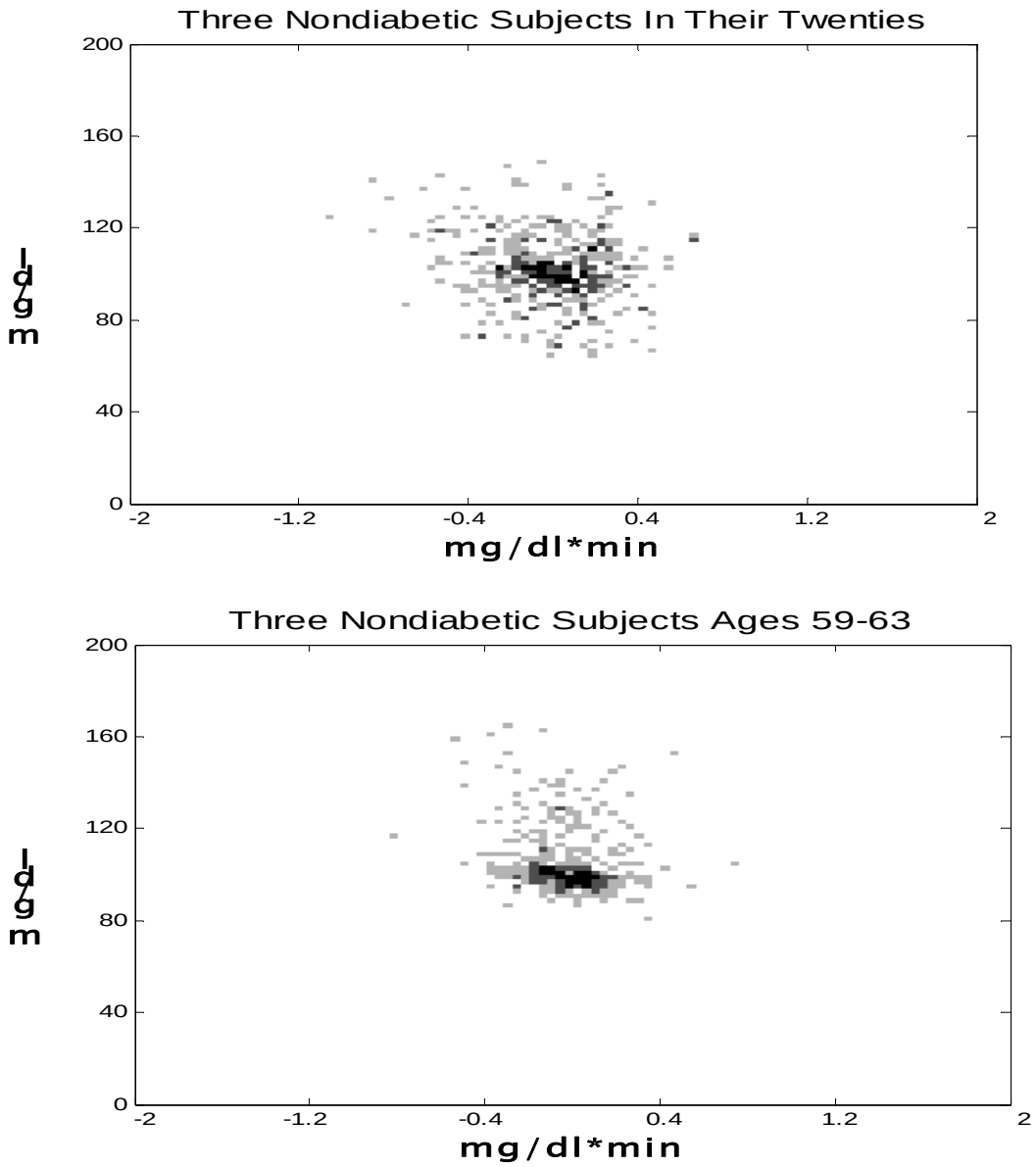


Figure 6-3: Phase space attractor profile for three nondiabetic subjects in their twenties receiving continuous IV glucose (top) and similar subjects who are older (bottom).

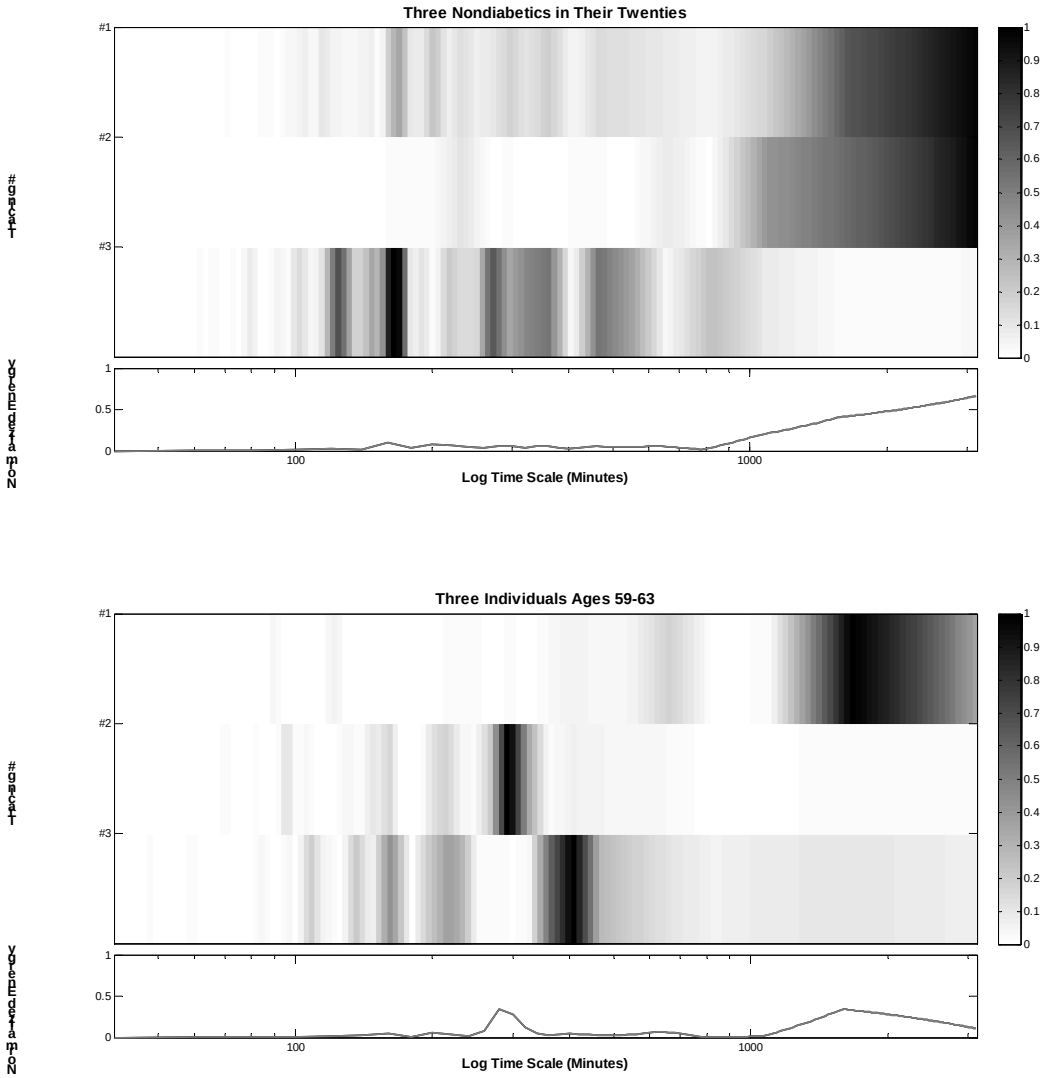


Figure 6-4: FFT based spectral components for three individuals in their twenties (top) and three older individuals (bottom).

VI.B Type II diabetes

Type II diabetes is characterized by loss of sensitivity to insulin and subsequent elevation of blood glucose. This leads to hyperinsulinemia and a feedback system of increased insulin production coupled by decreased sensitivity to insulin

([83]). In the end-progression of the disease state, insulin levels begin to drop as the beta cells are thought to “burn-out” in an attempt to produce more insulin to counter rising glucose levels. At this juncture in the disease progression, type II diabetics require insulin.

VI.B.1 Fasting

Nine subjects were observed during a twenty four hours of fasting, four of which begun their fast at a different time of the day. An example of the time-series generated is shown in figure 6-5. Trends were extracted as described in the section on the time-scales of nondiabetic dynamics, using a moving average process. These are shown in figure 6-6. The circadian rhythm is clearly present and is disturbed when the fasting occurs in a different time of the day. In two of the shifted subjects, the circadian rhythm is simply shifted accordingly and in two others it appears to be disturbed but present. A clear difference between these data sets and the nondiabetic case is the amplitude of the excursions in which such circadian rhythms had an amplitude of 20 mg/dl.

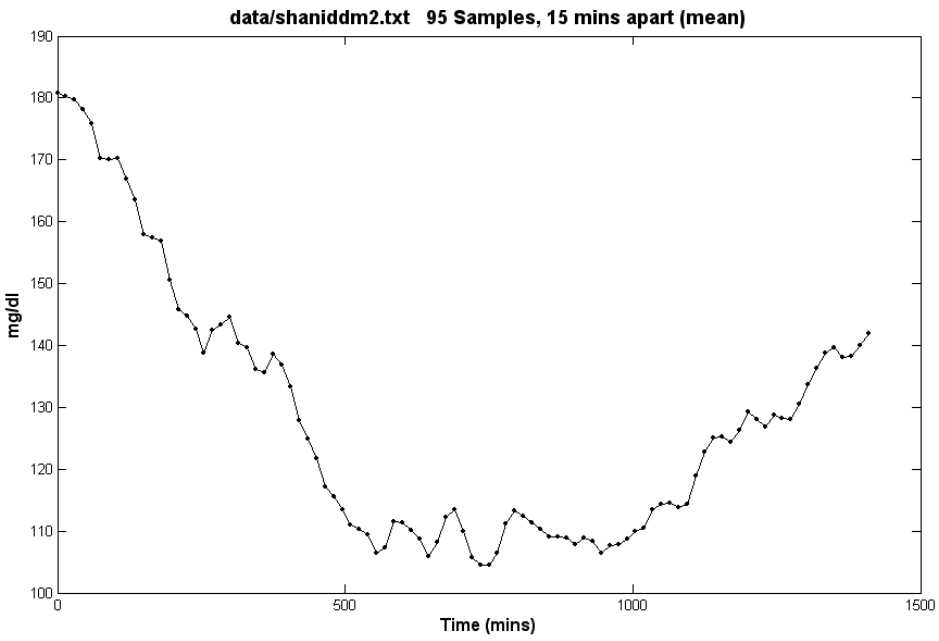


Figure 6-5: an example of a time-series from a type II diabetic fasting. This time-series is run through a moving average filter to create the circadian rhythms.

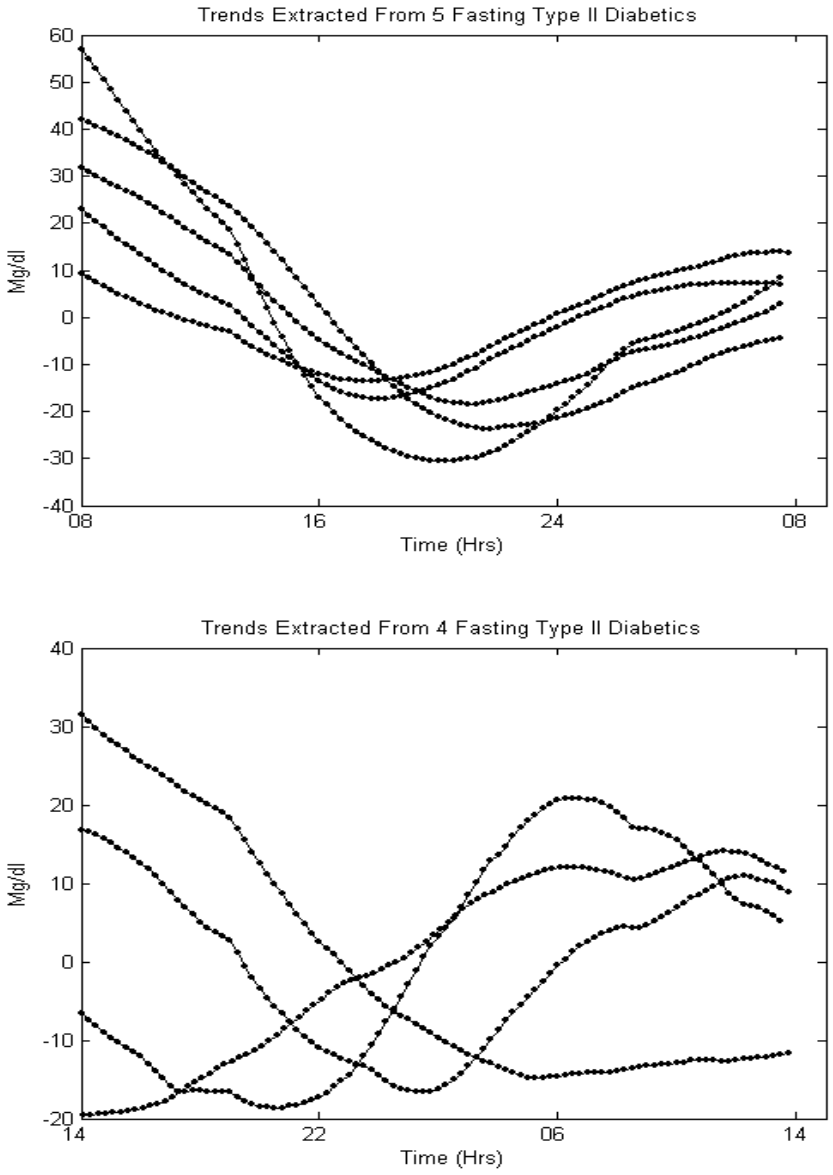


Figure 6-6: Trends extracted from 9 type II diabetics under fasting conditions. The first group (top) follows standard timing (8am to 8am the next day) whereas the second group (bottom) was tested with a six hour shift. Note that in the second group the circadian rhythm is disturbed and off phase, showing its relationship to the time of the day. Also note the magnitude of this circadian component is ~40-80 mg/dl whereas in the nondiabetic case the change was more on the order of 20 mg/dl.

VI.B.2 Constant input

In constant input into the system, specifically intravenous delivery of glucose, oscillations are noted as discussed in the previous chapters. The degree of contribution of these oscillations to the overall signal energy is decreased in the two data sets available from individuals with decreased insulin response: a patient with impaired glucose tolerance and one with type II diabetes. A look at the FFT based time-scale analysis yields decreased fast time-scale dynamics and increased long time-scale dynamics as would be expected. This is described as the loss of ultradian cycles in the case of impaired glucose tolerance . The finding itself is shown in figure 6-7.

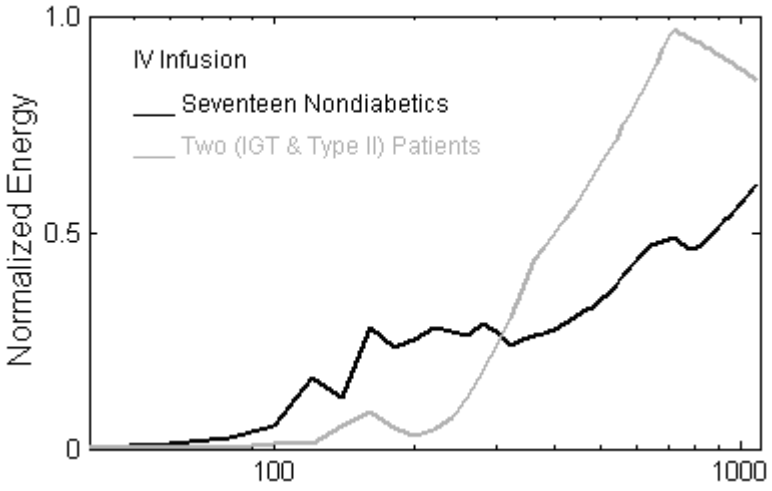


Figure 6-7: comparison of the average power spectrum evaluated via the FFT method, for seventeen nondiabetics at various intravenous infusion rates and two individuals with different degrees of insulin resistance (IGT and type II diabetes).

VI.B.3 Dynamic input

As in the case of nondiabetics, dynamic input can be used to observe the phenomenon of entrainment. In the case of insulin resistance and Type II diabetes, it has been observed that entrainment is less robust than in the case of nondiabetics. It is important to note that the original intent of these studies, and where the entrainment was mostly monitored was the insulin signal (which was also measured in these studies), and that the degree of entrainment is much more pronounced in the insulin signal. Table 6-1 and 6-2 described the % of entrainment. The methods for computing this were discussed in chapter 3. Table 6-1 described multiple nondiabetics and one patient with IGT and another with type II diabetes. As the disease state progresses, there is a clear decrease in the % of entrainment in the signal. In table 6-2 another experiment is shown in which a normal patient was compared with a type II diabetic patient who was then given medication. Another patient was treated with Placebo. In this case medication increased the percentage of entrainment but placebo did not. This particular experiment is included to highlight the possible utility of such measurements for assessing therapeutic utility. Such studies, if implemented on larger groups may be useful in assessing the therapeutic benefits of drugs meant to treat type II diabetes, as is shown in the case of patient #3. Without significantly more data, however, these results themselves cannot be generalized.

Table 6-1: Comparison of three normal patients (ID #1-#3) being infused at two different wavelengths of infusion and the percent entrainment noted. The % entrainment is the amount of signal energy that falls in the frequency bands associated with the input sinusoid. This percentage is lowered for the IGT (#4) and Type II patient (#5).

Patient ID #	Series	"Wavelength" of Infusion Sinusoid	% Entrainment
1	Poldynnorm3	144 min	79%
1	Poldynnorm4	96 min	55%
2	Poldynnorm6	192 min	52%
2	Poldynnorm7	128 min	81%
3	Poldynnorm9	144 min	58%
3	Poldynnorm10	96 min	37%
4	Poldynigt2	144 min	50%
4	Poldynigt3	96 min	23%
5	Poldynnidm2	144 min	41%
5	Poldynnidm3	96 min	16%

Table 6-2: Percentage of entrainment resulting from sinusoidal infusion in three patients, one normal, and two pre and post treatment with a drug and placebo respectively.

Patient ID #	Series	"Wavelength" of Infusion Sinusoid	% Entrainment
1 Normal	Polentrain1	144 min	89%
3 (Pre Treatment)	Polentrain3	144 min	77%
3 (Post Treatment)	Polentrain4	144 min	96%
4 (Pre Treatment)	Polentrain5	144 min	95%
4 (Post Placebo)	Polentrain6	144 min	90%

VI.B.4 Normal meals/ pregnancy

Unfortunately, no frequently sampled time-series of a type II diabetic individual that includes meals and exercise were found. This is in part because of the strong focus, until the recent epidemic, of research on type I diabetics, and the tendency for type I diabetics to require closer monitoring of glucose values, yielding a greater interest in the subject. This is truly unfortunate because type II diabetes is quickly becoming one of top health challenges facing western civilization [3].

The data sets that were found are of pregnant women with type II diabetes. Several such sets, in which the diabetes was controlled by diet and medications (but not insulin) were analyzed. The results for meal process models are shown in figure 6-8. Here, the loss of curvature of the fall model as was the case in the general comparison of type I diabetics and nondiabetics is noted. Differences between type I pregnant diabetics and type II pregnant diabetics is not as noticeable.

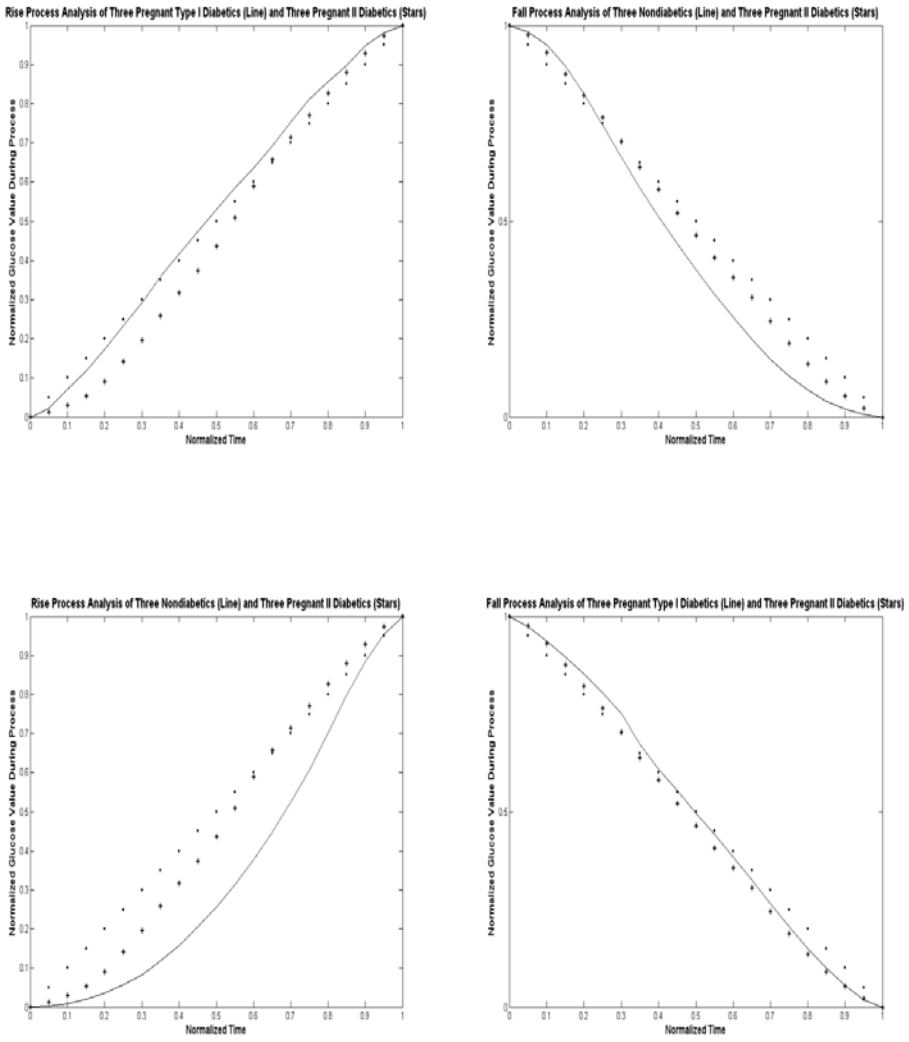


Figure 6-8: Meal process models for nondiabetics, type I diabetics and pregnant type II diabetics. In the upper panels, pregnant type II diabetics (stars) are compared with nondiabetics (solid line). In the lower panels the comparison is made between pregnant type II diabetics (stars) and pregnant type I diabetics (solid line). In both cases a line is provided for comparison (circles).

The time-scale analysis of the two different types of diabetes does not reveal a consistent significant difference between each other, but does show a shift towards the slower time-scale than the nondiabetic comparisons (figure 6-9). This is consistent with the observations made in the previous chapter and with the physiologic insight that insulin action is slowed in type II diabetics, and thus slower dynamics are expected particularly when insulin injections are not utilized. As before, multiple peaks are noted around the time-scale of meals and the natural oscillatory time-scale.

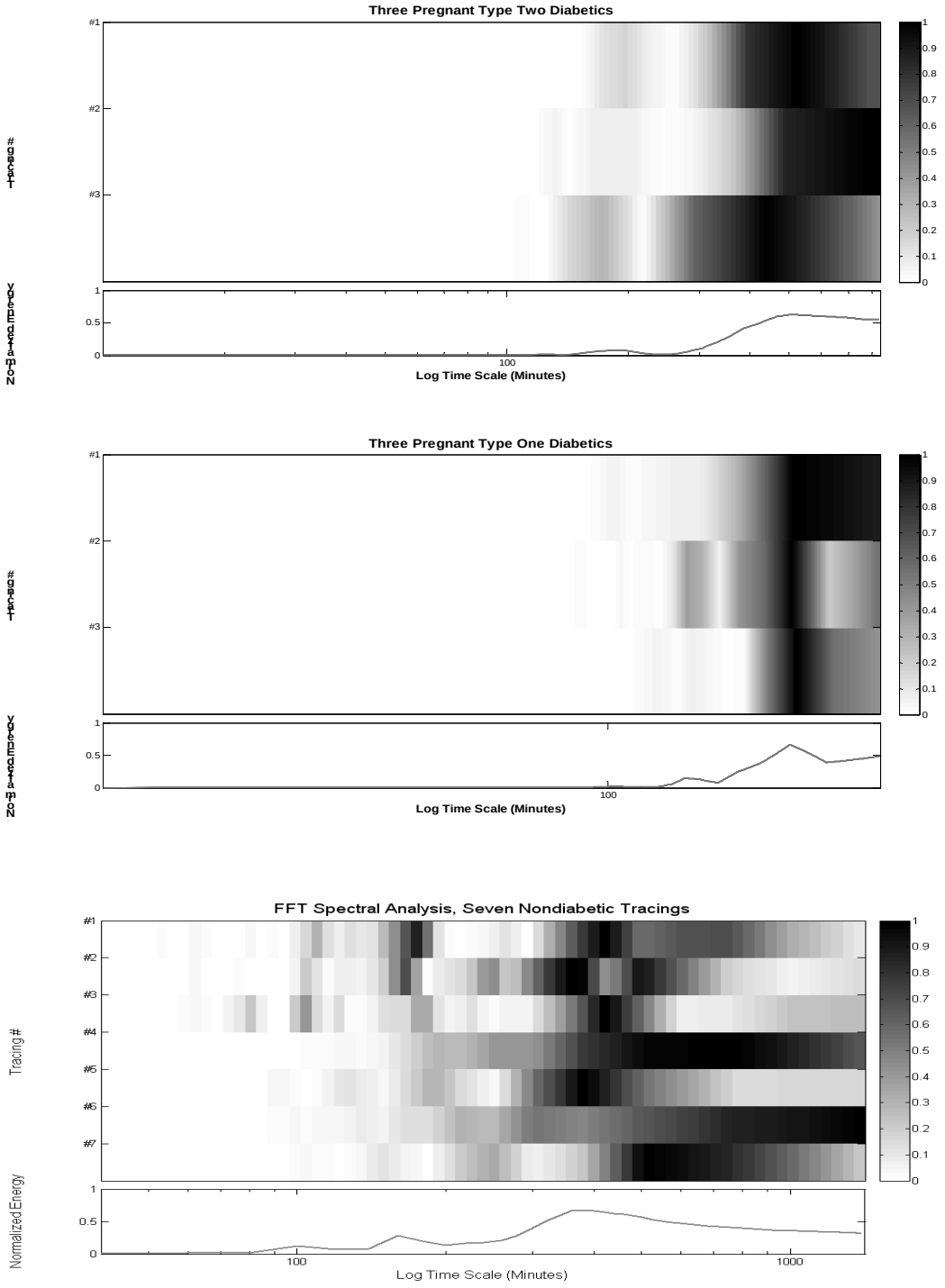


Figure 6-9: Time-scale spectral portraits based on FFT spectrum for pregnant type II diabetics, pregnant type I diabetics and nondiabetics (bottom).

A look at the statistics of the rate of change shows a consistent difference in the maximal rates of change for type I diabetics and type II diabetics. This is evident in both the rise and the fall maximal rates. This is as would be expected as insulin dynamics are greatly dampened by insulin resistance in the case of type II diabetics, leading to decreased maximal fall rate. The decreased rate of rise may be due to delayed insulin action in type I diabetes which may not "kick in" until well into the meal, whereas the type II diabetic has an immediate albeit dampened response as the meal is consumed. Skewness and kurtosis were not consistently changed, but attractor area for type I diabetics is significantly larger, again because of the increased speeds of rise and fall. These observations are summarized in table 6-3.

Table 6-3: Statistics of rates of change and its attractor for diabetic pregnant subjects. Maximal rates of change (in both directions) are larger for type I diabetics than type II diabetic subjects, as is the combined area of their attractors.

Subject	Max Rise	Max Fall	Skewness	Kurtosis	Attractor Area (combined)
Sanpregone	3.2	-2.3	.8	3.9	
Sanpregone	7.8	-3.4	2.2	14.5	
Sanpregone	3.2	-2.2	-.7	11.2	1852
Sanpregtwo	2.3	-1.7	.7	5.4	
Sanpregtwo	1.7	-2.0	-.1	6.1	
Sanpregtwo	2.6	-1.3	1.3	4.7	536

VI.C GCK deficiency

The enzyme GCK (Glucokinase) plays a role in the metabolism of glucose in the liver, as well as in the pancreatic beta cells, and is thought to be a key step in the glucose sensing abilities of the pancreatic beta cells [84]. Patients are characterized by mild fasting hyperglycemia. GCK deficiency provides a unique opportunity to study the effect of the exact mechanism of glucose imbalance on dynamics and further studies are necessary to extract information from such patients. In particular studies such as those discussed in the nondiabetic case would help differentiate the effect of dysfunction in this system on glucose dynamics as opposed to other mechanisms which do not involve beta cell loss (such as insulin resistance)

Six subjects with elevated fasting glucose levels and GCK mutations were studied, along with six normoglycemic controls. The insulin response of the GCK deficient subjects was tested and the first phase response was found to be the same as the controls, but abnormalities were noted in the insulin secretion rates (ISR). The data set obtained was from an IV infusion using a sinusoidal infusion at a frequency of 144 min. The last 12 hours of the infusion were recorded at fast intervals. Unfortunately this was the only data set available. The results for a normal and GCK deficient subject are shown in figure 6-10.

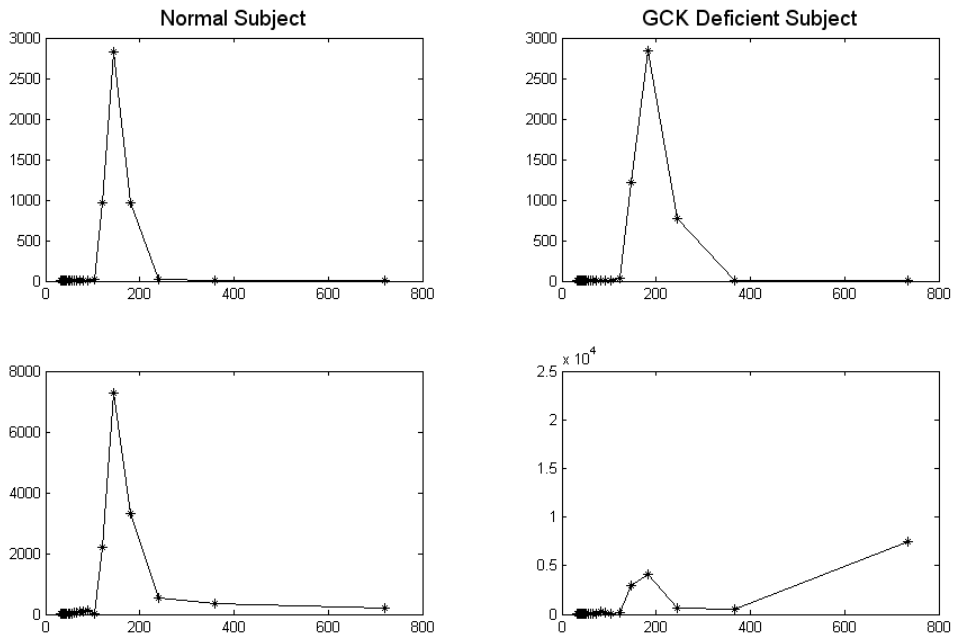


Figure 6-10: Comparison of a normal patient and a GCK deficient patient. The top two panels represent the time-scale analysis of the input signal in each case and the resulting signal in the patient is shown in the bottom two panels. The degree of entrainment measured was 92% in the normal patient and 20% in the GCK deficient patient.

VI.D Insulinoma

Tumors secreting insulin lead to metabolic imbalance, frequent episodes of hypoglycemia and side-effects resulting from hyperinsulinization of the body [85]. The course of therapy includes the surgical removal of the insulin secreting tumor. A study was performed which included frequent sampling of four individuals with a tumor secreting insulin. This is a very interesting subgroup in that it provides yet another disease state perturbation to the system.

Figure 6-11 shows a comparison of the phase space portraits for two patients before and after surgery. In both group an increase in compactness of the attractor is noted. Interestingly, the range of blood glucose values is not affected in the same way in both patients, with one patient showing a significant increase in blood glucose values while the other showing an unchanged maximum but an increased minimum. The range of rates of change is not significantly affected. Figure 6-12 shows the time-scale analysis of the same two pairs, showing a shift to the faster time-scales of dynamics. These two observations illustrate the potential use of these techniques to evaluation of the affect of procedures to the dynamics and may find utility in comparison of outcomes in the future.

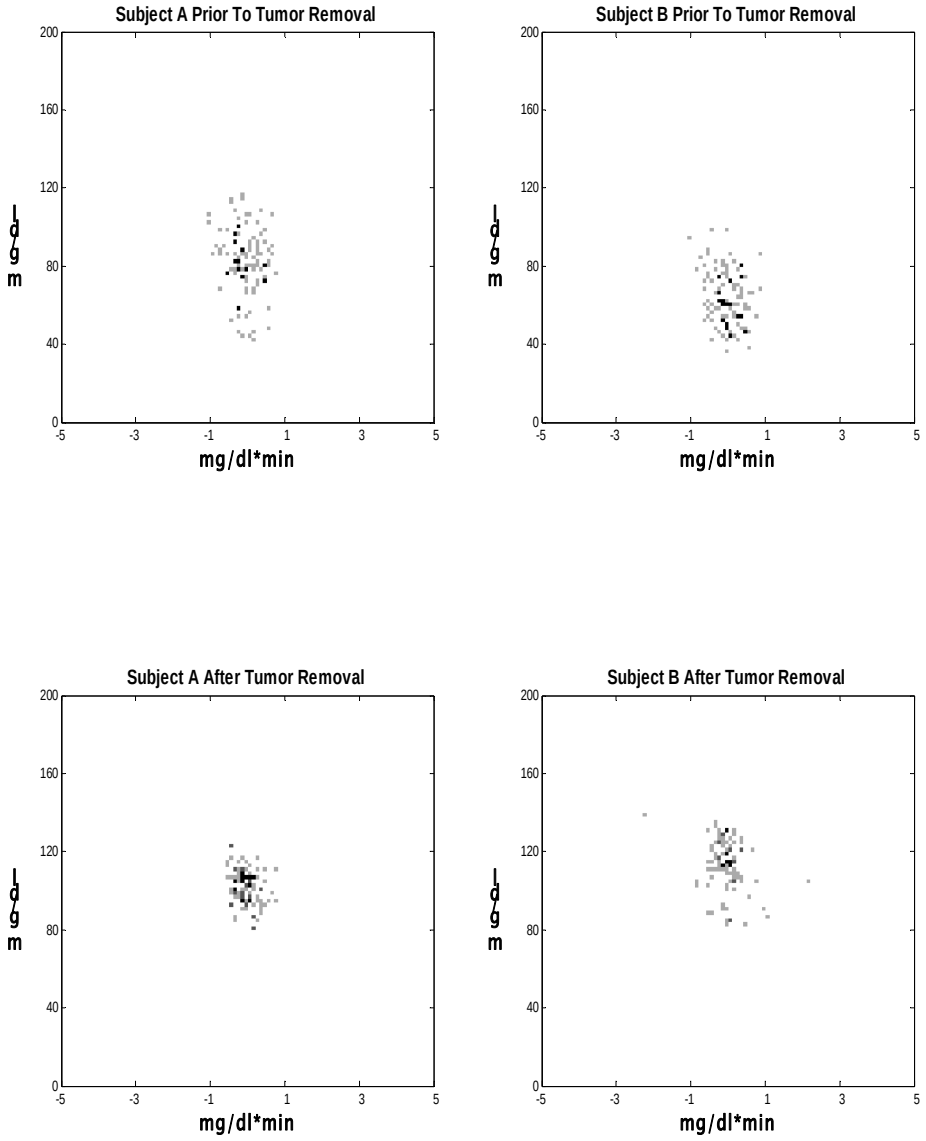


Figure 6-11: Before and after phase portraits for two patients with insulinomas prior to and after surgery to remove the tumor. Note a shrinking of the attractor size in both glucose values and rates of change.

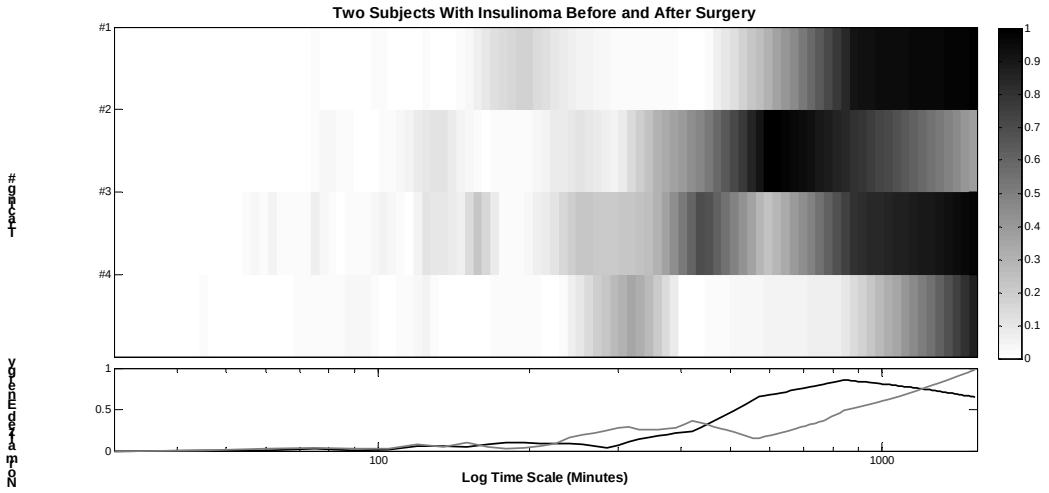


Figure 6-12: The combined spectral content of the time-series from two patients with insulinoma, prior to surgery (black line) and after surgery (gray line). A shift to faster time-scales in energy is noted.

VI.E Summary of observations

- In the few individuals studied, weight loss affected the compactness of the attractors in nondiabetic, patients with IGT and type II diabetic subjects but did not significantly alter the maximal rates of change.
- In the few individuals studied, time-scales were not affected in a consistent manner by the weight loss.
- Aging similarly showed an increase in the compactness of the attractor while increasing the energy in the shorter time-scales.

- The small number of data sets prevents generalization of these observations on aging and weight loss but illustrate the potential utility of these methods for assess impact on blood glucose dynamics.
- Fasting circadian rhythms are increased in amplitude in patients with type II diabetes, up to ~40-80 mg/dl as opposed to ~20 mg/dl in nondiabetic individuals. This effect is still present even if the sleep-wake cycle is altered, and is synchronized with the sleep-wake cycle rather than the time of the day.
- Constant input yielded decreased energy in the ultradian (~100min) time scale and an overall shift to the slower time-scale in comparison to nondiabetic subjects.
- Dynamic sinusoidal input showed decreased % of entrainment in IGT and Type II diabetics. In one case available there was an increase in an individual's % entrainment after medical treatment.
- Meal process models for type I and type II diabetic pregnant women showed similar linear process models as seen in type I diabetics in comparison to the S-shaped response of nondiabetics.
- Slower time-scales of dynamics were noted in both type I and type II diabetic pregnant women, as would be expected.
- Maximum rise and fall rates of change were lower in type II diabetic pregnant women and type I pregnant diabetic women. In type I diabetics this can be a result of slower insulin dynamics. In type II diabetics, it can be attributed to the damping effect of insulin resistance.

- Patients with GCK deficiency show an overall deceased % entrainment to ultradian time-scale sinusoidal input.
- Patients with removal of insulin secreting tumors display an increase in the compactness of the phase space portrait as well as a shift to faster time-scales

Chapter VII: Sampling rate and dynamic distortions

The most common approach to understanding sampling requirements is set forth by the need to prevent aliasing, a phenomenon in which under-sampled high frequency components are reflected back into the lower frequencies and cause distortions to the signal. The frequency necessary to avoid aliasing effects is set forth by the Nyquist sampling period. It roughly states that a signal should be sampled at twice the highest frequency component. This has been studied in the case of blood glucose dynamics [14]. The motivation behind extending this analysis further is twofold:

(1) To illustrate to clinical researchers not familiar with the Nyquist criterion the importance of frequent sampling using different dynamic results

(2) The Nyquist criterion may not be a sufficiently strong criterion in certain applications and may affect various dynamic parameters differently.

In doing so, a few different approaches will be taken to the analysis of the sampling requirement for dynamic analysis. One is to look at the reconstruction error. Simply put, if an attempt is made to replace the missing samples by using simple geometric fits, one can measure the error encountered in this reconstruction. While it is unlikely that this will be done directly in clinical applications, data points are often connect with lines on graphs and if not so, will be looked at in this way in the

clinician's mind. Thus here an attempt is made to capture the error that is introduced by "looking" at data points that represent an under-sampling of the actual signal.

A different approach is taken by extending the Nyquist concept using nonlinear information analysis. Here, the concept of mutual information is introduced as another measure of information. Information loss is then looked at as the samples become further and further along. The information "half-life" is defined as the sampling period after which half of the information has dissipated from the last sample. This is another approach to understanding the time-frame of information loss in the system and its relationship to the need for sampling.

Finally, the effect of sampling is discussed on two dynamic measures that serve as the underpinning for much of our previous discussions: the rate of change and the spectral estimate.

VII.A Sampling rate and reconstruction error

One of the simplest methods of assessing the degradation of signal information content is the relative success with which the original signal can be reconstructed from the undersampled signal. A simple linear interpolator is used to reconstruct the original (more frequently sampled) signal from the undersampled signal. In effect, the simple linear interpolator over estimates the success of reconstruction compared to the current algorithm often used clinically, which is to assume that the last value is valid

until the next measurement is made (a zero order hold). Other, more complex interpolation schemes in essence begin to introduce more model-based assumptions about the system and require further investigation and optimization. To accomplish this, the signal is undersampled and the reconstructed signal is compared with the original (frequently sampled) signal. For each resampled time-series, absolute maximum error is calculated. This measure gives a reasonable sense of the kinds of errors that would be expected in reconstructing data sets from undersampled data, and in particular highlights the possible worst-case scenario errors. This maximum error is averaged for each patient groups and the results are shown for the combined data sets sernorm1-3 and serone1-15 in figure 7-1. These data sets were chosen because they were from similarly collected experiments and measurements were taken every five minutes.

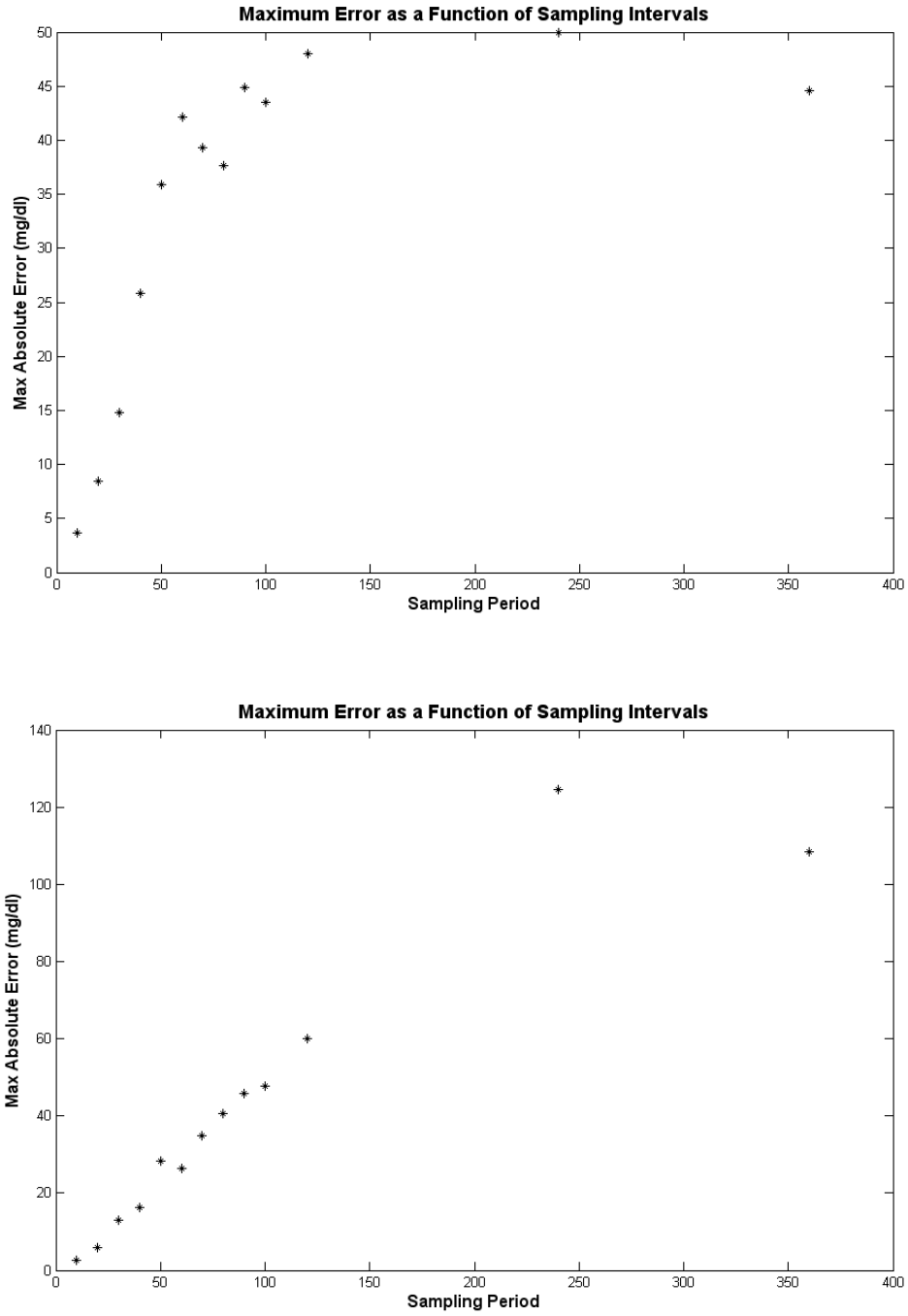


Figure 7-1: the maximum reconstruction error, expressed in mg/dl as a function of sampling period in 3 nondiabetics (sernorm1-3, top) and 15 diabetic time-series (bottom). Note that the diabetics time-series are of a larger amplitude which may explain the difference between the two subgroups.

Figure 7-1 shows a significant increase in error, even in the range of sampling below an hour. An average maximum error of approximately 30 mg/dl is seen in both data sets when attempting to reconstruct the data set from hourly sampled blood glucose values, a significant error in a clinical sense. Of interest is the observation that because of the slower time scales of dynamics in type I diabetics, coupled with the larger amount of signal energy in this group, the rate of increase of the error is comparable between the two groups as sampling increases to an hour, but levels off for the nondiabetics and continues to rise for the diabetics as the distance between samples increase. This highlights the fact that despite the large swings in diabetic time-series, significant error can be introduced into both signals if sampling intervals are increased, likely due to the faster dynamics in nondiabetics.

VII.B Sampling rate and information loss

As discussed in previous chapters, the dynamic information content of the signal becomes zero when samples do not share any information and thus are independent from each other. Two methods, the autocorrelation function and the average mutual information as ways of quantifying the sample to sample information cross-over were also discussed. Autocorrelation utilized the linear correlation (product) of samples to assess the common information between samples. This is plotted against the distant in time between the samples and the first minimum represents the sufficient time between samples to make them relatively independent. Similarly, mutual information uses a nonlinear vector quantized representation of the

signal to compare the common information between the samples and produces a similar output that allows us to estimate the amount of time that causes samples to become relatively independent of each other. Figure 7-2 shows two such curves for a single data set. It can be seen that both measures fall rather quickly to near zero as expected, because the information from a sample dissipates as time goes forward. Each measure approaches zero in a different manner, but the approximate time-scale is similar between the two measures.

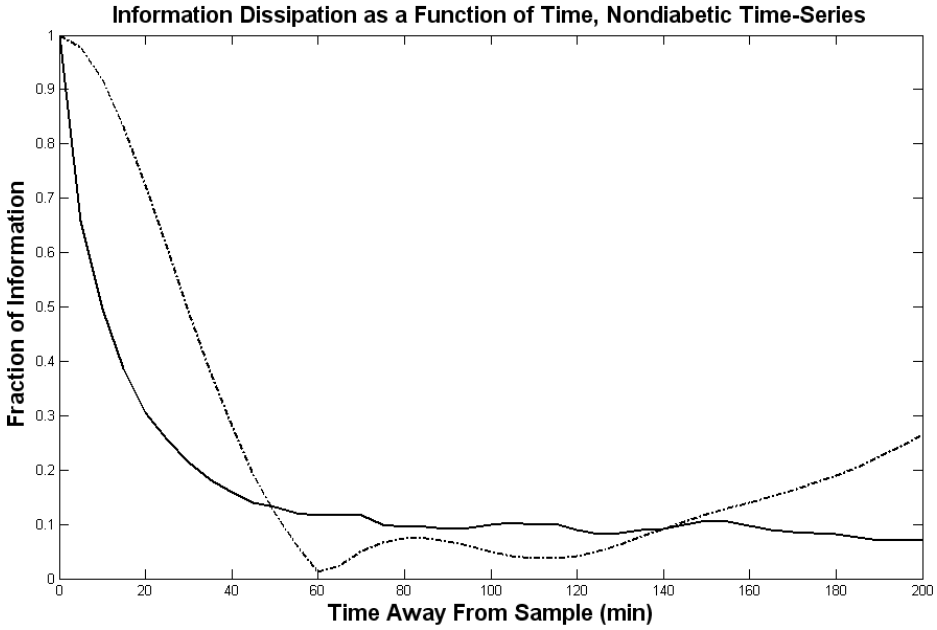


Figure 7-2: Information dissipation as a function of time is estimated using the autocorrelation function and the average mutual information.

Here, in addition to the above two metrics, another measure of similarity is used, which is the difference between the signal energy in each sample with respect to the other which serves as another comparison. Additionally, one can introduce the concept of information half-life. This is the time required for half of the information in each sample to disappear, which allows for a simple characterization of these curves

using a single number. The results applied to one complete data set that includes nondiabetics and various types of type I diabetics is shown in table 7-1.

Table 7-1: Various measures of information half-life computed for nondiabetics (sernorm1-sernorm3), stable type I diabetics (serone1-serone3) and unstable diabetics (serone4-serone15). Note that while the time-scales of information dissipation are longer on average for diabetics, some diabetics have relatively short information half-lives. The information half-life based on mutual information has the fastest decay rate in almost all the cases, owing in part to the way information is defined in this case.

Information Half-Lives (minutes) Calculated By:	Mutual Information	Auto-Correlation	% Change In Signal energy Value
Sernorm1	10	30	15
Sernorm2	15	35	14
Sernorm3	10	30	20
Serone1	20	65	35
Serone2	20	80	30
Serone3	15	55	29
Serone4	30	150	97
Serone5	25	130	32
Serone6	25	90	23
Serone7	25	80	41
Serone8	30	130	103
Serone9	20	70	58
Serone10	20	65	37
Serone11	25	85	29
Serone12	15	60	35
Serone13	25	85	51
Serone14	35	190	84
Serone15	15	65	46

Table 7-1 summarizes information dissipation for the most complete data set available in this thesis. Despite the different measures of information half-life, two common observations can be made. One is that the information dissipation is significantly faster for the nondiabetics and the well-controlled diabetics, and thus for

monitoring purposes, sampling must be fast enough to be well within these limits. Second, it can be noted that depending on the measurement of information, different criteria arise for the information dissipation. This highlights the importance of a priori knowledge of the dynamic metrics in calculating an optimal sampling frequency.

VII.C Sampling rate and rates of change

As mentioned before, the rate of change is the simplest expression of the underlying dynamics of the system. Thus the degradation of the estimate of the rate of change can serve as a good estimate of how undersampling is likely to effect all measurements derives from rates of change. An example such a study applied to nondiabetic and diabetic time-series is shown below in figure 7-3.

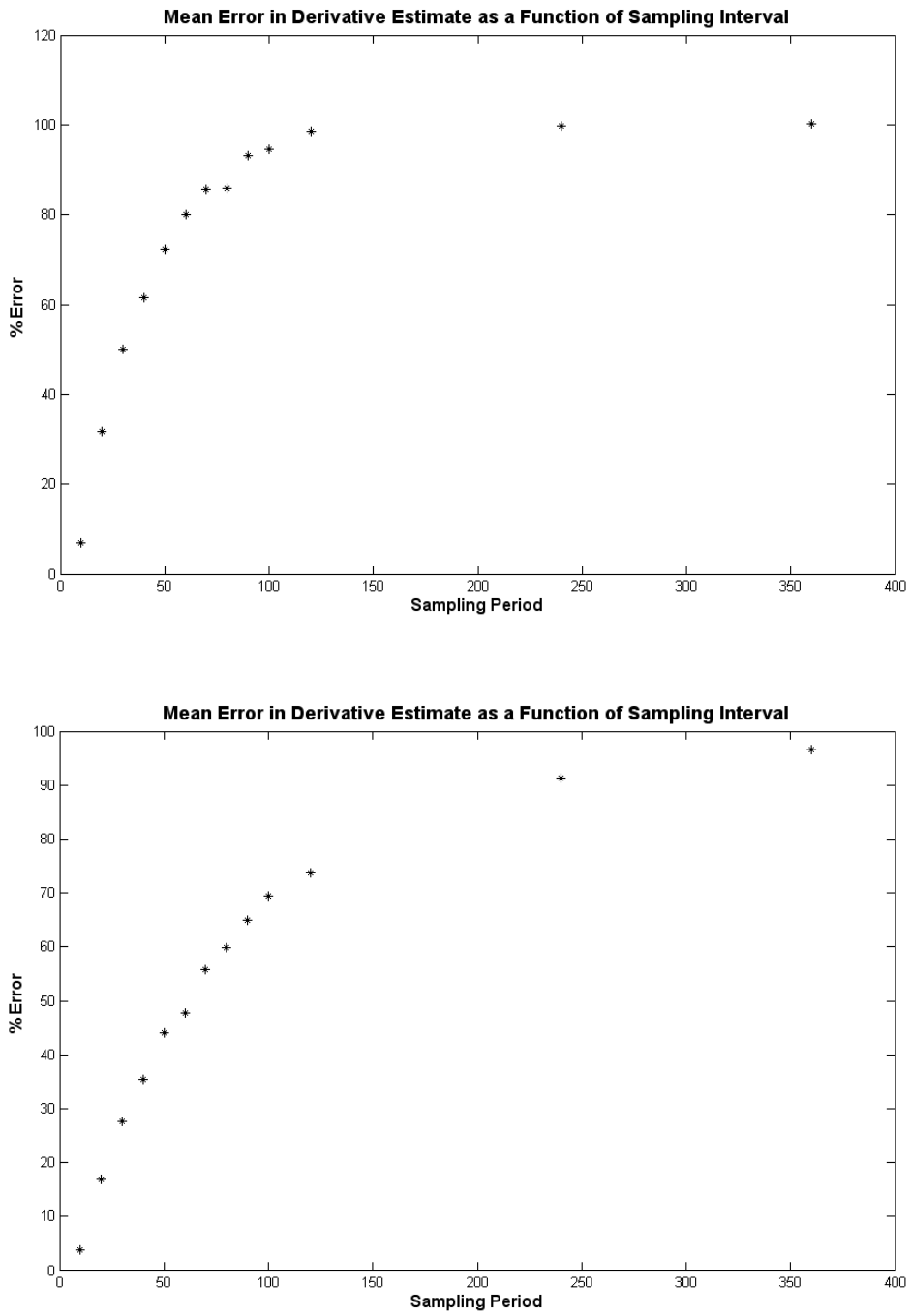


Figure 7-3: The mean error in the estimate of the derivative as a percentage of mean absolute value of the derivative estimates, from sernorm1-3 (top) and serone1-15 (type I diabetics). In the former, filter of period 40 minutes, there is a 50% error in the derivative estimate highlighting the importance of frequent sampling. In the case of the type I diabetics, the rate of degradation is significantly reduced, but nonetheless reaches unacceptable levels rather quickly.

As can be seen in figure 7-3, there is significant degradation of the estimate of the derivative as sampling periods are lengthened. In fact, for hour long sampling errors approach or exceed 50% of the value of the derivative estimate, on average. This highlights the loss of information in undersampling and its effect on the simplest aspect of dynamics, the rate of change.

VII.D Sampling rate and spectral estimate

The phenomenon of aliasing which is the central side-effect of sampling below the Nyquist frequency is best observed in the frequency analysis of the time-series and comparing it with the undersampled case. As was discussed, previous work by our group studied the Nyquist period and determined it to be near 10-15 minutes for most rapidly changing datasets. It is of key importance to recognize that sampling below the Nyquist frequency introduces distortions in the spectral estimate and does not simply make information in that frequency unavailable. This is important as there is an urge to undersample time-series in clinical settings with the rationalization that the fast time-scales are not of interest, which as discussed is not a sufficient cause for undersampling because of the distortions introduced in the slower time-scales. This is an extensively discussed topic in the literature [86] and in this section a simple example is given for illustration purposes in figure 7-4. Here, the top strip (#1) shows the spectrum calculated from the original data set. The resampled spectrum calculated from the less frequently sampled data is shown in strips 2-5. Note that not only does the peak in the spectrum becomes wider, but other peaks also disappear and a

significant shift of the energy to the slow time-scales occurs which indicates aliasing. This illustrates the detrimental effect of undersampling on time-scale energy estimations.

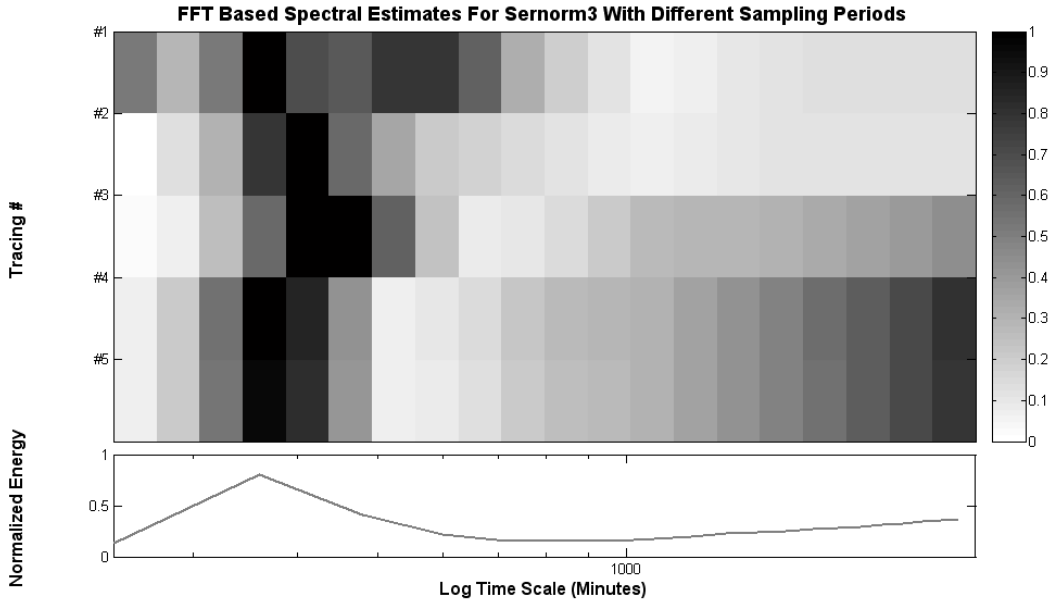


Figure 7-4: The gradual shift and degradation of the spectral estimate at sampling rates of 10, 20, 40, 60 and 120 minutes for tracing number 2-5 respectively. The first tracing represents the 5 minute sampled time-series.

VII.E Summary of observations

- Attempts at signal reconstruction using hourly samples leads to errors as large as 30mg/dl.
- The information half-life may serve as a more intuitive concept for understanding the need to sample and its effect on actual deterioration of signal dynamic concept. It was found to be smaller for nondiabetics and well controlled diabetics, an observation in agreement with the Nyquist frequency measurements in previous work.
- Estimates of rates of change are particularly sensitive to undersampling. In nondiabetics, 40minute sampling leads to a 50% error in rates of change estimates. In diabetics this error is closer to 30%
- As expected there is a gradual degradation of the time-scale analysis with decreasing signal frequency.

Chapter VIII: Conclusion & future directions

The objective of this thesis was to explore methods of analysis of dynamics developed in other disciplines of physics and engineering, and apply them to the limited amount of data currently available to provide tentative hypothesis and directions of further work. The objective was to improve the utility of this signal to the diagnosis and monitoring of disease processes, helping define various measures of "normal" dynamics as a target for therapy, and define dynamics that can be utilized to guide device characteristics. To the extent that numerical understanding of the dynamics leads into insight about the underlying system generating the signal was considered a possible added benefit. Thus, care was taken to minimize the number of assumptions about the underlying model generating the dynamics, which are sampled using the single blood glucose measurement. As mentioned, whether a single variable such as blood glucose can accurately capture the dynamics of the system is itself subject to debate. The goal then becomes to characterize, with minimal assumptions about the underlying model, as much of the dynamic behavior of the underlying system as possible in order to extend the utility of the signal beyond a statistical interpretation. To this end, methods of time-series analysis were explored thoroughly.

The simplest methods, relying on the statistical analysis of rates of change, which are the simplest expression of dynamics were used. Multiple statistics were used to accomplish this: mean rate of change, maximum rates of change in either direction, and the Skewness/Kurtosis measures that characterize, to some extent, the

shape of the distributions. Additionally, by plotting the value of blood glucose against its rate of change, a phase portrait is generated. Because values of glucose are governed by physiology that attempts to restore equilibrium, there is a tendency for clustering of values around a region, which creates what is termed an attractor. The path of the system around this geometric object is its “orbit”. We proposed a few measures to characterize this geometry in hopes that considering both values at the same time would lead to new insights not available by looking at the distribution of each measurement alone.

We then looked at the time-scales of dynamics. We chose the word time-scales instead of frequency because it is more intuitive for clinical purposes to consider time instead of frequency, particularly when the common units for frequency are on the order of 1 second whereas blood glucose dynamics evolves on minutes to hours. We explored various methods of time-scale energy estimation and used them to find time-scales of dynamics in different groups that were considered. We further observed that these time-scales are themselves subject to change and used time-frequency analysis to better understand their evolution. Methods such as pulse extraction or circadian rhythm extraction were utilized when these were the primary time-series “structures” available.

Methods borrowed from information theory, such as measures of signal complexity or entropy was applied as a way to study the signal from an information theoretical perspective. Tests of nonlinearity were examined in order to ask whether

nonlinearity in the system generating the signal could be detected by time-series analysis. Methods of nonlinear time-series analysis were themselves applied to the signals. Methods of testing and analyzing nonlinearity met both difficulty with application (signals were very short and did not go through sufficient orbits to characterize dynamics) and interpretation, which without an underlying mathematical model can be difficult.

We then turned to the meal event, a characteristic perturbation in the time-series whose timing was known in some of the time series. By introducing this additional component to the time-series, we examined whether the events themselves could be detected, what percentage of the signal energy they constitute, and whether their shape and characteristics could be better understood. We proposed extents both in values and time to characterize the meal, and also averaged the “shapes” of the meals to for a “meal process model”.

One characteristic of a dynamically useful measure as a way to categorize individuals is to require that dynamic characteristics remain stable for the same person. In data sets that contained two subsequent days of dynamics, we decided to measure some of the methods discussed above to see if dynamics were variable from day to day. Finally, we observed the degradation of various methods mentioned so far as the sampling frequency was lowered.

To test the above methods, we gathered data that was published in the literature. We also incorporated data from a clinical trial of sensors that included direct blood glucose measurements from diabetic and nondiabetic children. We used only direct measurements of blood glucose. It was assumed that these measurements reflect a similar dynamic system. Measurements using sensors that indirectly measure glucose in the subcutaneous tissue were not used because the tissue introduces its own complex dynamics, which remains to be characterized completely. These data were "extracted" using digitization of published figures with acceptable accuracy which were tested in multiple ways. Fortunately, the data sets included a multitude of patients with various conditions at various ages, under various perturbations. Unfortunately, many of these data sets did not include enough individuals even for a preliminary conclusion that can be generalized to a larger general subgroup. They did provide a way, however, to assess the utility of tools in application towards future, larger data sets.

In searching for a dynamic definition of "normoglycemia", rates of change appear very promising. The attractor geometry, looking at the compactness of the distribution of both blood glucose levels and their rate of change reveals characteristics that may be useful for defining normoglycemia. **We observe that 90% of the samples fall within 2%-5% of the attractor area. That is to say the system spends majority of its time inside a small area within its attractor. The size of the attractor and its compactness were useful consistent characteristics that were similar between subjects in the same experiments.** The attractor geometry is

asymmetric, and contains possible structures that can be studied with larger numbers of similar data sets. For example, there appear to be higher negative rates of change at higher glucose levels and higher positive rates of change at lower glucose levels. This confirms the fundamental nonlinear nature of the dynamics, which try to restore the value to the equilibrium, and do so with greater "effort" as values reach the extremes. The concept of symmetry as measured in this analysis failed to produce consistent results.

Time-scale analysis of data sets with very high sampling rates (2 mins) leads to the conclusion that very high frequency oscillations constitute a very small portion of the signal energy content, particularly in the context of the presence of larger perturbations such as meals and infusions.

The meal is the dominant source of signal energy in the blood glucose time-series. In nondiabetics we find that the period following a meal contains 2-5 times as much energy as the rest of the time-series. In fact, these periods account for a significant portion of the signal. This combined with the observation of energy in the time-scale of meals leads to the conclusion that in the an attribute of normoglycemia is the lack of other dominate time-scales and signal energy contributions from other events. Without knowledge of the meal event timing, large meals could be successfully detected from the time-series itself, increasing the utility of this type of analysis in real-time devices or clinical settings.

In nondiabetics consuming meals, time-scale analysis using a variety of methods including signal modeling and FFT based estimations show four time-scales of dynamics: The circadian rhythm (~24 hours), the meal spacing (~4-6 hours), the meal length (~2-3 hours) and faster time-scales possibly from background ultradian pulses (~100 mins). These observations were fairly consistent between the limited data available. The assigned causative source of these energy time-scales, is off course a matter of conjecture to a large extent. Time-series generated with complex perturbation regimes can lead to a more causative model of these contributions. We also demonstrated, by bringing together the data collected from these different types of experiments that **various time-scale can be highlighted using a different perturbation regime to the system.**

In the absence of meals, ultradian oscillations and a circadian component dominate the signal energy. **The ultradian pulses have a range of length scales based on the individual or the experiment but repeat about every 100 minutes**. The other time-scales of dynamics can be observed in the absence of meal perturbations. In fasting, two dominant time-scales expose themselves: the circadian rhythm, and a pulsatile “ultradian” rhythm, which seems to centered around 100 minutes. The same ultradian time-scale is noted when individuals are placed under continuous enteral feeding, continuous IV glucose infusion, or even continuous insulin infusion. The exact period of these ultradian pulses varies between experiments and individuals, and based on our time-frequency analysis, may even vary from pulse to pulse. Infusion of glucose in a pulsatile manner near this frequency leads to

entrainment, a phenomenon in which phased lock synchronization occurs, and blood glucose levels oscillate at the infusion frequency. Interestingly, despite the removal of meals from the perturbations, the maximal rates of change are not as different as one might have expected between time-series without meals and those with them. There was, however, a change in the shape of the distributions of rates of change, noted in the skewness of the distribution.

In fasting, a circadian component (~20 mg/dl) and an ultradian component similar to the continuous infusion experiments (~5mg/dl) becomes apparent.

Time-frequency analysis shows that there are changes in the frequency components of the nondiabetic regardless of the type of perturbation. In the case of multiple meals, these can be explained by the differences between the meal response itself, and in the case of constant input, these can be attributed in part to intermediate frequency processes which are not at the same frequency as the oscillatory behavior observed. These seem to be particularly pronounced in the case of intravenous infusion. Infusion of the nondiabetic subject with an oscillatory input within the same vicinity as the oscillatory frequency leads to entrainment, which means that the blood glucose values begin to oscillate at the same frequency as the input. In the nondiabetic, the meal process appears to be asymmetric, with a faster rising process, which can reach a maximal velocity of 2-3 mg/dl per minute. The fall process tends to reach a smaller maximum velocity but overall tends to be larger than the rise process. In addition no “shoulder” is observed in the rise process whereas a

shoulder is often noted in the fall process owing in part to the biphasic response of insulin.

Evidence for nonlinearity exists in two of the time-series studied but the conclusion cannot be made based on firm computational grounds. Nonlinear dynamics analysis will require significantly longer samplings to allow for attractor characterization. In hindsight, this observation is completely consistent with the finding of larger time-scales of dynamics (circadian rhythms), the nonstationary and highly nonlinear process of meal spacing and size, as well as changes between day-to-day responses to the same perturbations. That is to say, if there is something to be gained by nonlinear time-series analysis, it will require significantly larger time-series if it is to include perturbations such as meals.

The analysis of the meal event itself yields some consistent observations, including a sigmoid fall and rise process as well as the observation that on average, the time to maximum rise velocity in the meals is shorter than the time to maximum fall velocity. Further study of this technique is warranted, and in particular, parametric fitting of these curves may find future use.

Aside from a more complete characterization of normoglycemia, time-series analysis would not be clinically valuable unless is able to detected differences between dynamics in altered states. Due in part to historical significance of type I diabetes as the best studied example of such altered state, a significant amount of effort was

dedicated in this analysis to comparing data from type I diabetics to nondiabetics. Overall, the methods that found utility in the analysis of nondiabetics also found similar successes in this population. However, the differences between each experiment and each individual grew more significant. This is at least in part due to the introduction of yet another “extrinsic” dynamic factor, mainly the injection of insulin which varies tremendously in its dynamic characteristic as many different treatment types and regimens exist. Key observations are summarized below.

In the experiments with the best sampling and duration, rates of change of blood glucose are more confined in well-control type I diabetics than in nondiabetic subjects, while they are similar between nondiabetics and non-well controlled diabetics. The DirecNet data sets, for unclear reasons shows much increased rates of change as high as 13 mg/dl * min. This may be due to populational characteristics, infrequent irregular sampling and different experimental set-ups. There is an overall decrease in the Kurtosis of the rate of change from nondiabetics to well controlled diabetics and finally to poorly controlled diabetics. Attractor size is increased from nondiabetic to well controlled diabetic to poorly controlled diabetic.

In the frequency domain, the increased rates of change do not translate into a higher frequency behavior. Rather, they represent larger, wider pulses that result from the interaction of meals and insulin injections. **Time-scales of dynamics are shifted towards longer time-scales of insulin injection as well as the meals, with well controlled diabetics showing faster time-scales of evolution.**

Methods of signal characterization such as measures of complexity and nonlinear time-series analysis show differences between the type I diabetics and nondiabetics. Signal complexity is decreased in the type I diabetics, as shown in a decrease in the approximate entropy of the signal. Nonlinear time-series analysis shows a decrease in the estimates of the largest positive lyapunov exponent in the type I diabetics.

Meals are larger, last longer and account for less of the total energy in the signal in type I diabetics, and are shaped less like a sigmoid and more like a line if they are averaged together. This is presumably caused by the increased role of insulin injection as opposed to the biphasic, fast physiologic response available in nondiabetics.

It is very important to note that while these observations regarding type I diabetics are interesting, they cannot necessarily be attributed fully to the disease process, in the sense that the dynamics of type I diabetics, while clearly less dependent on meal timing is now more dependent on insulin timing and dosage. Thus the analysis performed is relevant only in the context that provides examples as it relates to this specific data set and the parameters of the treatment.

Insulin resistance, type II diabetes, obesity, pregnancy, GCK deficiency as well as tumors that secrete insulin provide challenges and opportunities to study the

glucose regulatory system and its dynamics. Data in these domains were difficult to find. This is unfortunate, as insulin resistance and type II diabetes have reached epidemic proportions in western societies and now eclipse type I diabetes as cause of morbidity and mortality. Very limited data exists in these domains, however significant work has been done in limited perturbation regimes in some of the cases described.

For example, oscillatory behavior at the ultradian time-scale has been extensively studied in IGT, Type II diabetes as well as obesity. **Fasting blood glucose dynamics is found to be markedly altered between nondiabetics and type II diabetics in the form of an increased amplitude of a sleep dependent circadian rhythm**. Fasting circadian rhythms are increased in amplitude in patients with type II diabetes, up to ~40-80 mg/dl as opposed to ~20 mg/dl in nondiabetic individuals. Type I diabetics during pregnancy reproduce findings noted in type I diabetics who are not pregnant, such as increased time-scales of dynamics as well as an expansion of the phase space attractor. Meals were altered in the same ways as type I diabetics studied. Pregnant type II diabetics showed similar increase in time-scales of dynamics but have less marked changes to the maximal rates of change than their type I counterparts. Maximum rise and fall rates of change were lower in type II diabetic pregnant women and type I pregnant diabetic women. In type I diabetics this can be a result of slower insulin dynamics. In type II diabetics, it can be attributed to the damping effect of insulin resistance.

Throughout the analysis, it is generally found that ultradian oscillations are mixed with slower time-scales as disease states progress, presumably as a result of a less robust regulatory system. Entrainment, the process of time-scale and phase synchronization with an input signal is found to deteriorate in disease related conditions. For example Dynamic sinusoidal input showed decreased % of entrainment in IGT and Type II diabetics. In one case available there was an increase in an individual's % entrainment after medical treatment. Both of the above mentioned findings provide consistent observations, which may serve as a diagnostic and assessment tool. Methods of treatment including removal of tumors, weight loss and drug treatments show changes in dynamic parameters in the few data sets available of this kind. Methods of time-series analysis applied to these observations allow quantification of these observations, though the specific optimal method must be further refined by considering larger population based studies. Few data sets that characterized age and weight status were also studied. In the few individuals studied, weight loss affected the compactness of the attractors in nondiabetic, patients with IGT and type II diabetic subjects but did not significantly alter the maximal rates of change. Aging similarly showed an increase in the compactness of the attractor while increasing the energy in the shorter time-scales.

Some of the more pronounced observations are shown in figures 8-1 through 8-3. The first two figures, using the Service data set which included frequently sampled data from nondiabetics, clinically easy to control type I diabetics ("stable") and difficult to control diabetics ("unstable"), shows the characteristics each subgroup

in terms of rates of change and time-scales of dynamics. Figure 8-3 shows a variety of situations in which time-scales of dynamics can be used to distinguish between two groups in an experiment. These observations, once applied to larger population groups appear promising in terms of diagnosis and monitoring of disease states.

Differences as well as correlations were found between each day of two day experiments, even though the perturbations to the system remained the same between each day. Correlation between the time-scales of dynamics existed between day one and day two of time-series from nondiabetics, but the degree of the correlation was inconsistent. This correlation was computed in the frequency space using the coherence function. Statistics of rates of change showed differences between individuals with some individuals showing good intra-day correlation while others did not. The same inconsistency was seen in the statistics of the shapes of the attractors. Significant variation exists between meal parameters, as measured using the methods described, even in the same individual with the same meal between two days. Measures of signal complexity were inconsistent between individuals from day one to day two.

Despite the differences between each day in the analysis, in general, variations between individuals in two consecutive days are less than the variations between individuals suggesting the system has deterministic reaction structure that can be further explored. This implies that certain dynamic measures may be more stable as a way to derive a personal “signature” of the individual’s dynamic response. These

differences must be further characterized in order to reliably characterize the expected day-to-day variance and can in a way be a method of defining the expected “error” in using each dynamic analysis method for a short period time-series (1 day).

In type I diabetics, similar observations were made about the day to day variations as was the case in nondiabetics. Time-scales show day to day correlations that vary between individuals. Little consistent correlation existed between measures of information half-life from day to day, but approximate entropy did show good correlation day to day. Statistics of rates of change and attractor geometry remain inconsistent between each day in type I diabetic subjects.

As expected, dynamic information degrades significantly with less frequent sampling. Estimates of rates of change are particularly sensitive to undersampling. In nondiabetics, 40 minute sampling leads to a 50% error in rates of change estimates. In diabetics this error is closer to 30%. As expected there is a gradual degradation of the time-scale analysis with decreasing signal frequency. Attempts at signal reconstruction using hourly samples lead to errors as large as 30mg/dl. The concept of the information half-life was introduced and explored as a framework for better communicating the importance of frequent sampling within each population. Its value was found to be smaller for nondiabetics and well controlled diabetics, which is in agreement with the Nyquist frequency measurements in previous works.

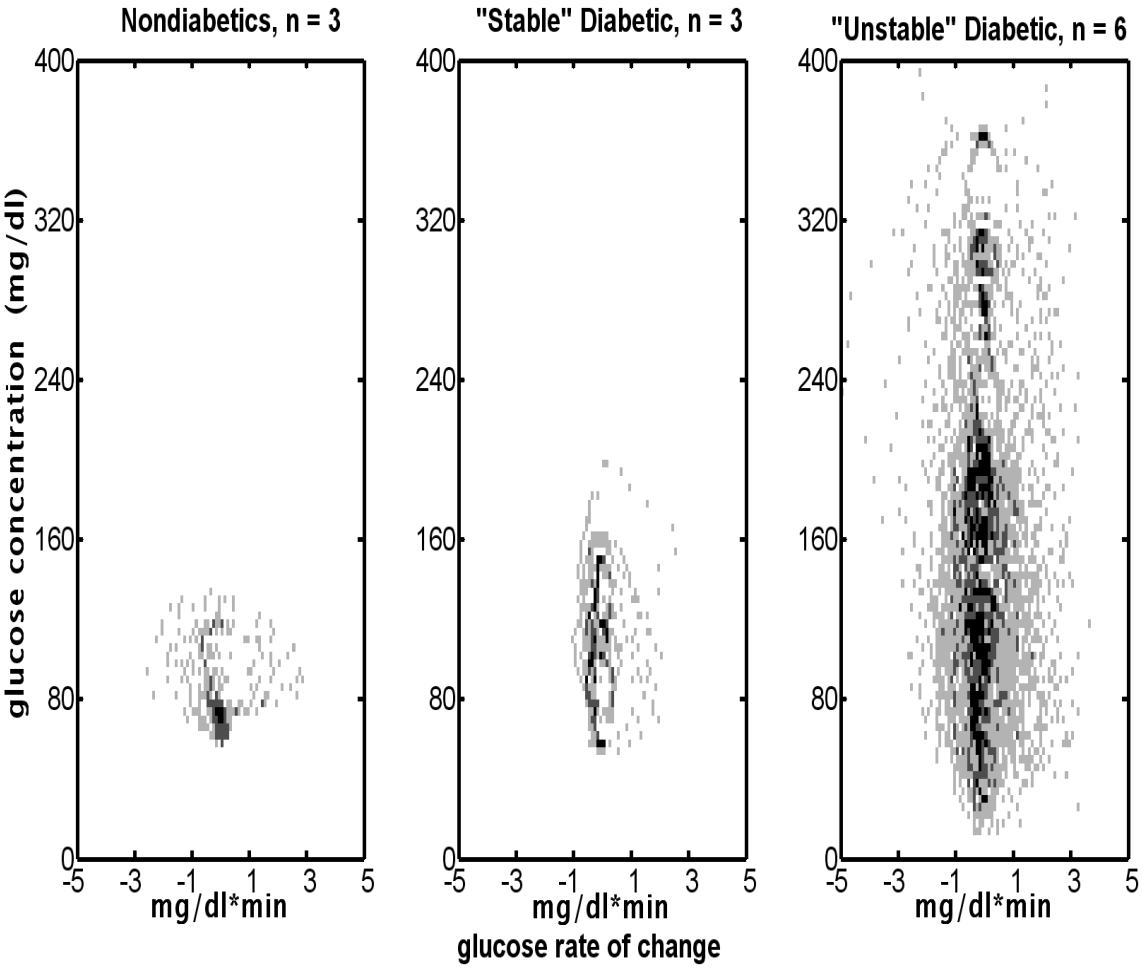


Figure 8-1: Three different subgroups of the same experiment shown in phase space attractor form. The darker the squares, the more samples fall in that two dimensional square. This view allows a concurrent visualization of values and their rates of change. Notice the difference in the shapes and sizes of the attractors. Maximal rates of change are not as affected as one may have expected between each group.

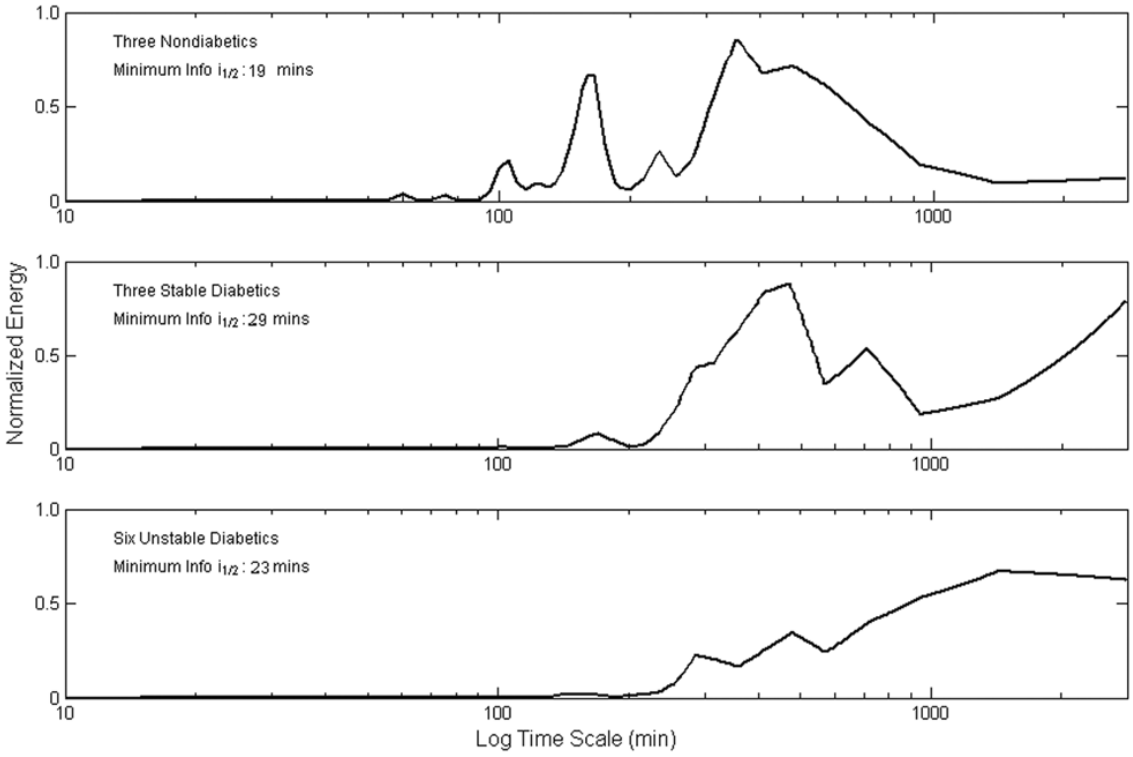


Figure 8-2: Time-scales of dynamics in the same three subgroups as in figure 8-1. Here the progression from faster time-scales and extent of the disease process are clearly visible.

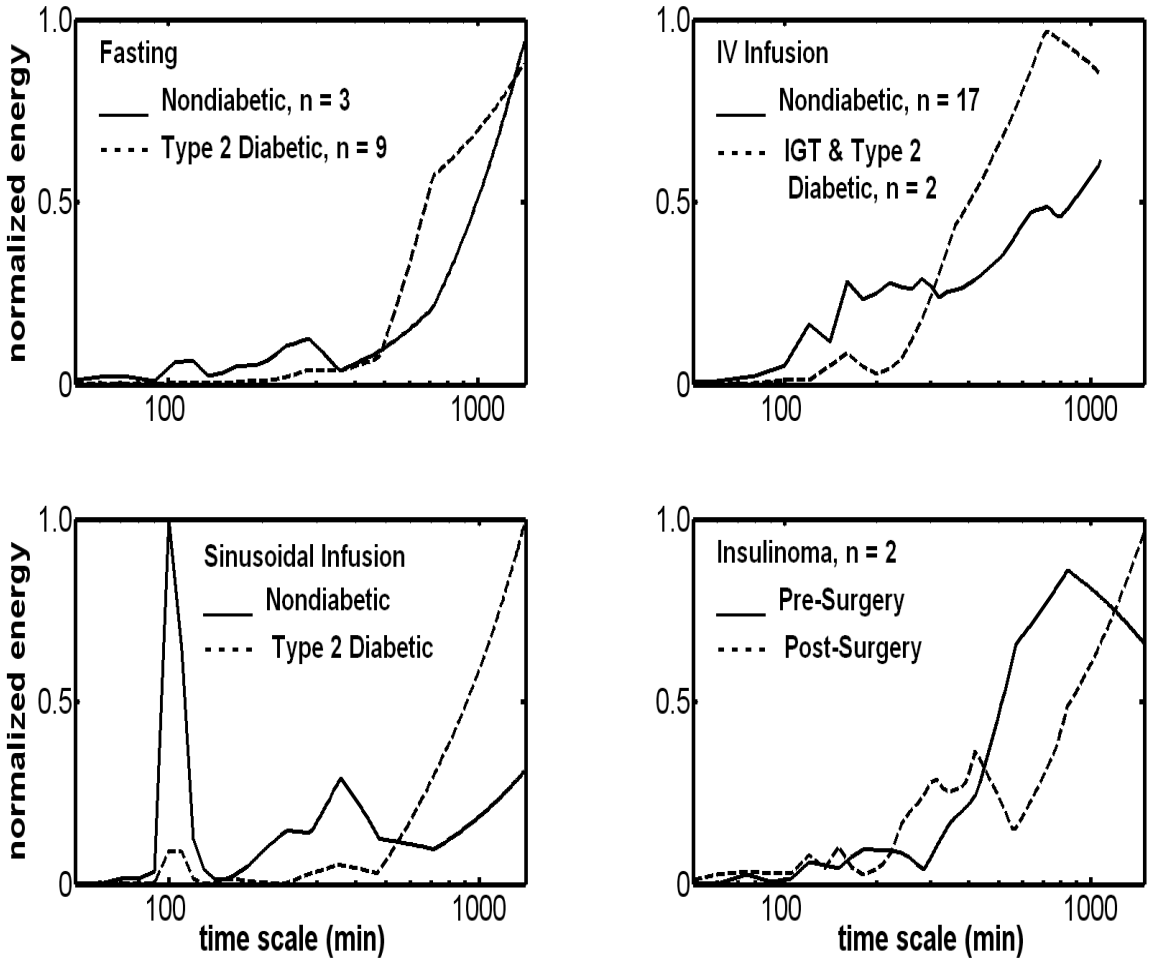


Figure 8-3: Time-scale analysis applied to four different situations where two subgroups exist. The first panel (top left) shows the time-scales of dynamics compared between nondiabetics and type 2 diabetics. In the top right, the time-scales between nondiabetics and those with IGT/Type 2 Diabetes are shown. Lack of entrainment is noted in type 2 diabetic individuals in comparison to nondiabetic individuals in the lower left panel. Finally, the same subjects are shown prior to and after surgery, removing an insulin secreting tumor.

.

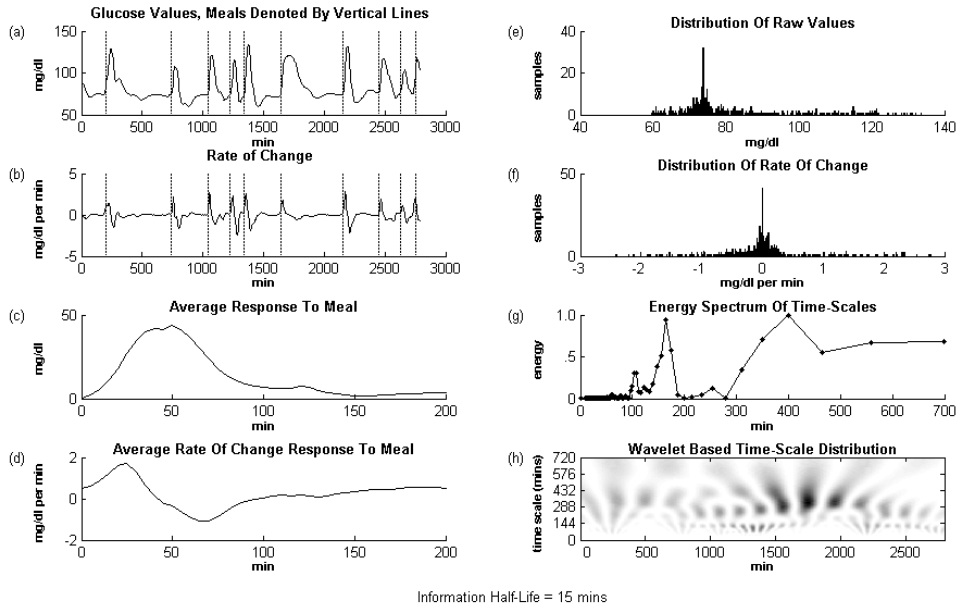


Figure 8-4: An individualized dynamic profile that includes multiple views of the time-series acquired from a two day frequently sampled blood glucose data set. Each panel represents a different piece of information derived from the top left panel.

Future directions in this study clearly require more data, particularly of the kind contained in the Service data set which include multiple days, frequent sampling, meals and exercise which are similar to the daily routine of individuals. It is the author's intention to continue to optimize and make available the software used to analyze these time-series as a way to stimulate further investigation into the subjects discussed. It is hoped that by a synergistic design of future experiments to collect appropriate data, while analytic tools are further developed to better analyze that data, the study of dynamics and assume a useful role in the diagnosis and treatment of metabolic disorders. This development may ultimately lead to a comprehensive glucose profile, which is a snapshot of the individual's glucose dynamics. An example

of such a snapshot is shown in the final figure of this document. The author looks forward to exciting future developments of tools associated with this new window into human metabolic dynamics.

Appendix A: Acquisition and verification of data-sets

Images are recovered either from computerized files in the .pdf format readable from Adobe Acrobat computer software, or from scanned in photocopies of the original journals. The .pdf format allows the selection of the image of the graph directly. The image is selected and saved as a .jpg file with. All images acquired in this manner were correctly aligned and did not require any rotation.

In the papers not available electronically, photocopies of the original journals were scanned in at 1200 X 1200 dpi. In most cases, the geometry of the book, the error of photocopying and subsequent scanning can introduce rotation in the graph. The angle of this rotation was calculated by using one of the axes of the graphs to measure the angle of rotation, based on the assumption that in the original the axes are either horizontal or vertical. The axes chosen was based upon the length of the axes, using the assumption that the longer the axis, the better the estimate of the rotation. The rotation angle was then used to re-rotate the image using the imrotate() function in MATLAB, and in specific using bicubic interpolation.

In all image types various points in the axis tick-marks were registered and recorded for use in the quantification of the final results. Once a time-series was generated, it was normalized to reflect the actual units of time and mg/dl based on measurements of the original axis of the graph. These allowed for a conversion

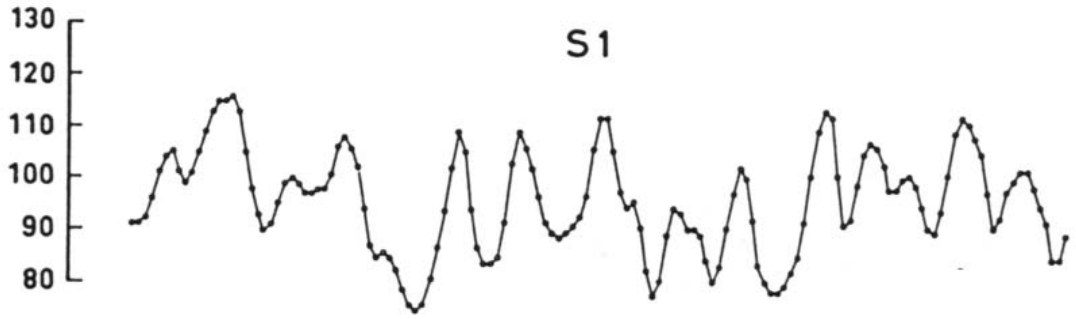
between pixels and relevant units. Units other mg/dl were converted to mg/dl using the appropriate conversion factors.

There were three different types of image processing necessary for digitization, each which was dictated by the way the data was graphed. All methods were verified visually as well as numerically to the extent possible and an estimate of the time and value uncertainty was computed. These values are available with the data set download by contacting the author. A description of the methods follows.The points sampled were then replotted against the original image in order to verify the integrity of the acquisition. This was performed in all cases and was visually verified. In the few cases were slight visual disagreement was clear, the pages were rescanned and the process repeated until discrepancy was minimized.

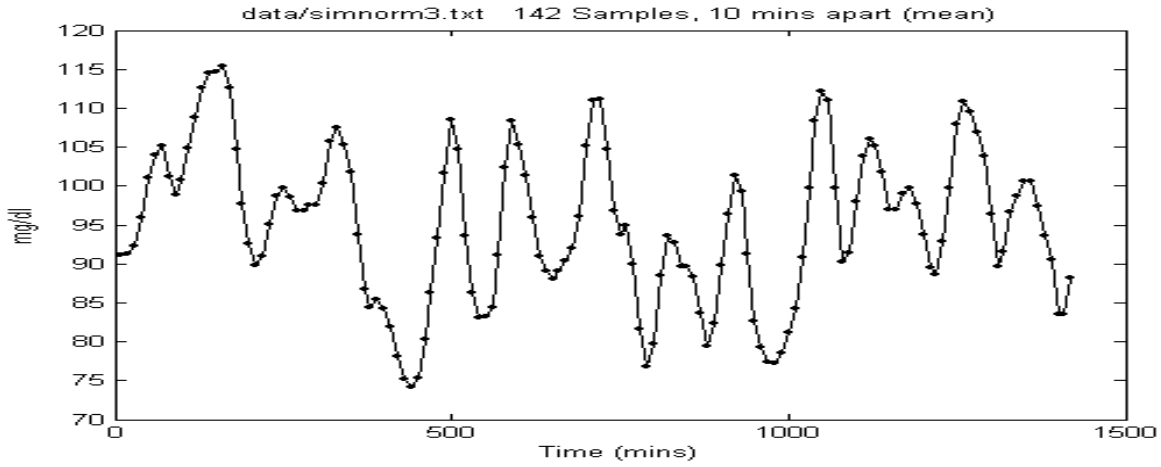
Type 1: Exact data points marked on graph, connected by lines

In these data sets, the exact data point is marked by small circles connected by a line. This gives the opportunity to closely approximate the actual original data used to generate the graph. Automated determination of the data point center was examined, but human determination performed well and was more time-efficient than creating an exact routine. Thus the center point of these circles was determined visually and marked by a single dot with different pixel color than the black circle background. A routine was then developed to extract the data. All registration and rotation algorithms were carried out in the same way as the other two techniques. An example of this conversion is shown below in figure A-1.

(A)



(B)



(C)

Figure A-1: (a) Original image scanned and rotated from publication. (b) a subsection showing the marking of the center point. (c) The final data displayed in time-series analysis program (TSV)

Error in these cases was computed by the size of the marker used to denote the data point. This was expected to lead to both a horizontal (time) as well as a vertical (value) error. Data points were re-sampled to correspond to uniform sampling using

interpolative methods, but no significant deviation in timing of the samples was noted that would lead to concern about the introduction of distortions.

Type 2: Continuous line drawing

These data sets were acquired from graphs containing one continuous line-drawing of the graph, presumably from connecting samples that were taken. Here, an automated program was developed to mark the “center” of the line segment after cropping and rotation were performed as mentioned above. In this method, applied to graphs containing continuous lines, the value at each horizontal pixel was recorded by moving a MXN window and finding the “center of mass” of the non-white pixels. N was generally the length of the smallest rectangle able to enclose all non-white pixels for the M columns. M was either 3 in the case of very finely measured graphs or 1 in the case of graphs with smaller resolution. In graphs with poor resolution (in comparison to the ones above) not enough samples exist per column to justify the averaging methods. Thus the darkest pixel for each column was selected to be the value pixel and it was quantified. .

This then created a single point estimate of the value in time from the graph. Generally the spacing between the pixels represented a shorter time than the spacing between the original samples used to generate the graph, so the time-series was re-sampled at the original slower frequency. There were no cases where the resulting time-series had to be up-sampled. It is critical to recognize that this particular approach amounted to linear interpolation between the original samples (used to

generate graph by the plotter/printer), and then phase incongruent re-sampling of the time-series at the original sample. This is expected to introduce a distortion in the very high frequencies in which very little energy presumably exists to begin with because only frequently sampled data were used. The other distortion is one associated with detecting the middle of the line, which becomes particularly difficult when there is a significant rate of change and thus significant vertical range in per a given time. An estimate of the maximum possible error can be made based on the range of the vertical line segment at that time. An example of these values is shown in figure A-4. The estimation of the error was based on the square root of these values. That these errors did not actually occur is discussed in the sections to follow as these types of data sets also included some pre-calculated statistics that allowed verification of the extracted data. Figure A-2 and A-3 show the progression from a scanned image, to a rotated and cleaned image, to the selection of the central points (zoomed in a section) and the ultimate time-series generated (prior to normalizing based on the original graph grid).

The average estimate of the error (keeping in mind that the error was calculated for the worst sample for each time-series) was ~ 3mg/dl averaged between all time-series. The maximum was ~6 mg/dl for a single sample in one of the time-series, keeping in mind that this is a worst-case scenario error.

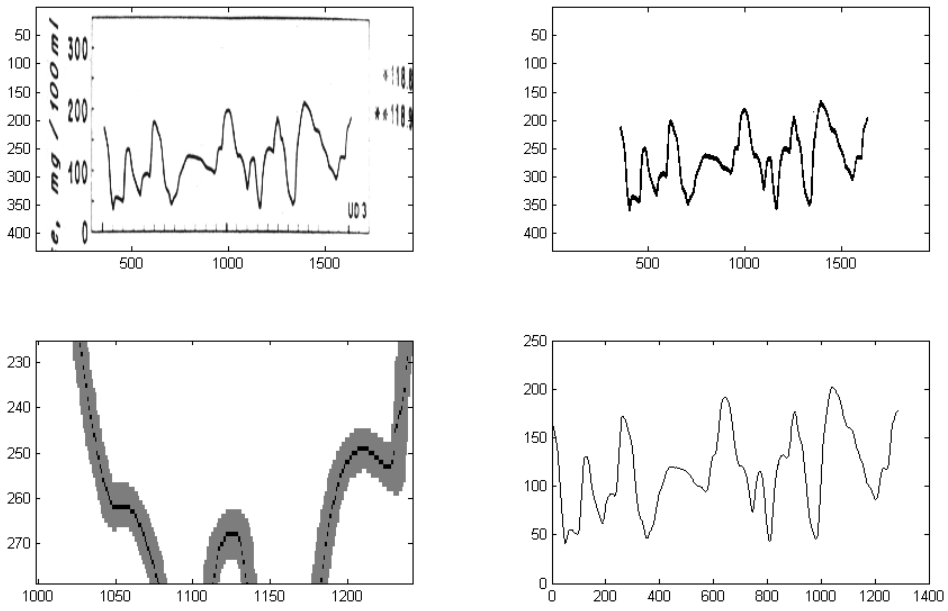


Figure A-2 : An example of the conversion of line graphs to data. (upper left) the original rotaed image. The axis denote pixels and not time or glucose values. (upper right) the cleaned line image. (lower left) a subsection showing the line center estimation technique. (lower right) final data displayed prior to conversion actual time and concentrations.

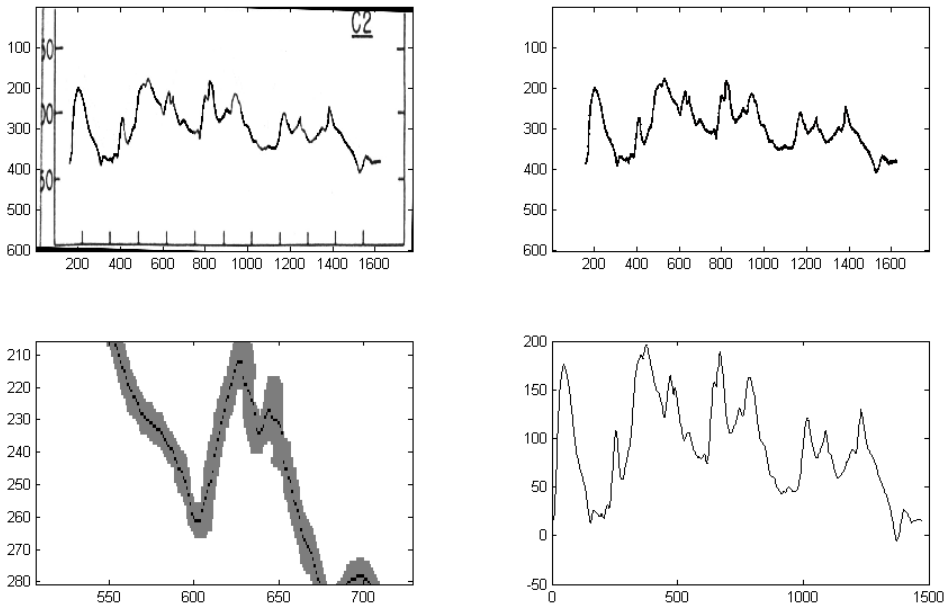


Figure A-3 : A second example of the conversion of line graphs to data. (upper left) the original rotaed image. The axis denote pixels and not time or glucose values. (upper right) the cleaned line image. (lower left) a subsection showing the line center estimation technique. (lower right) final data displayed prior to conversion actual time and concentrations.

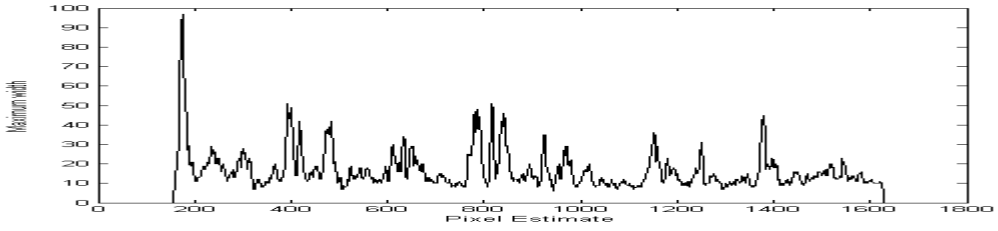


Figure A-4: Estimates of the maximum width of a line during the center estimation technique. Peaks correspond to areas of maximal rates of change where determining the exact sample point has a larger maximum possible error associated with it. These errors severely overestimate possible error as seen in data sets where values were verified using published statistics of the data set.

Type 3: Dashed and dotted lines

Few of the data sets fell into the categories where dashes or squares were used to denote data points which do not lend themselves to as logical of a data extraction approach as would be desired. These were treated by direct connection of the markers by computer based drawing which was then treated in a similar way as the line-center approach described above. The progression from the original scanned item to the data plotted is shown in figure A-5.

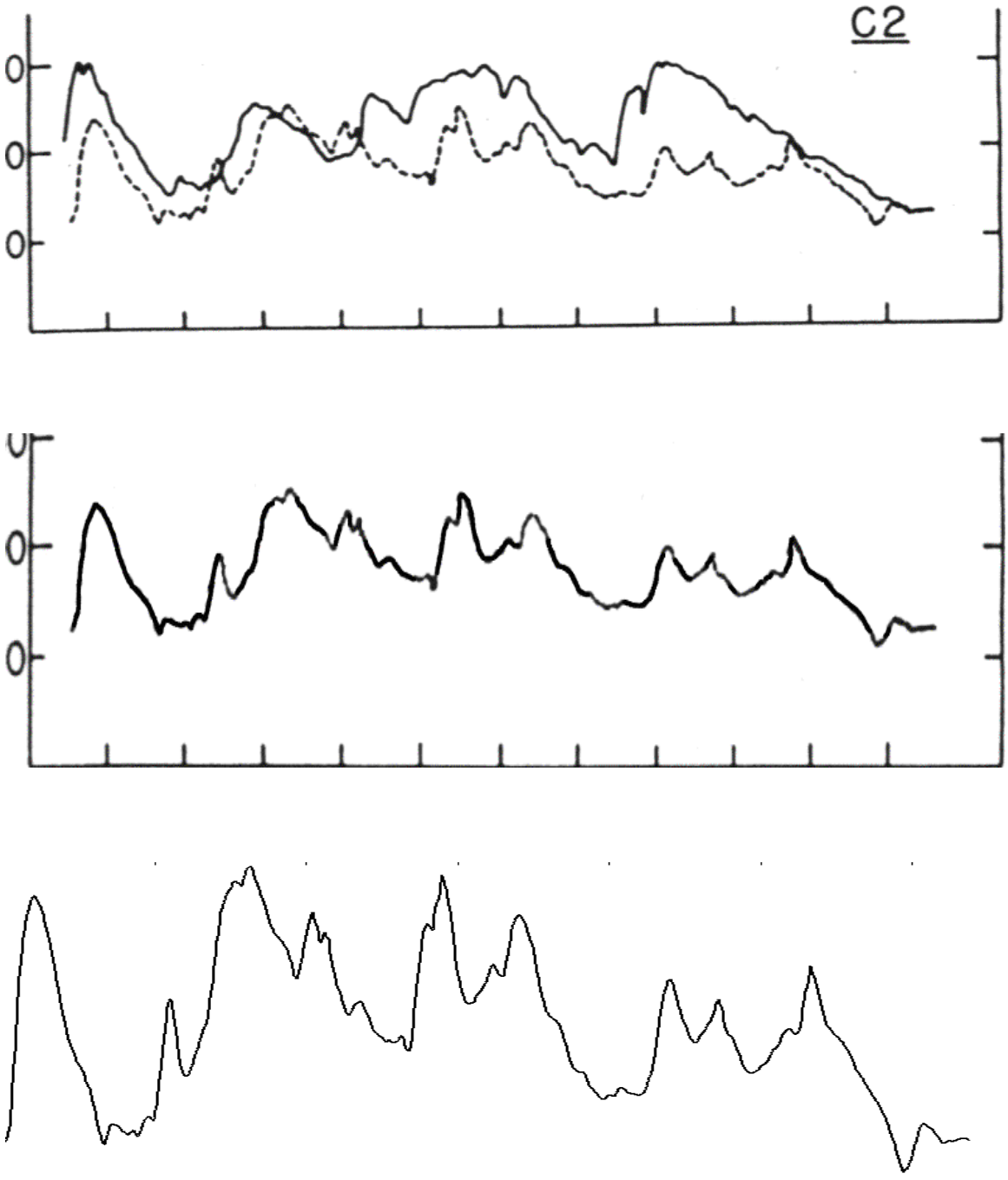


Figure A-5. The progression from a scanned figure connected by dashes (gray) in the first panel, to a hand-connected image which is then sampled and regenerated in the bottom panel. Very few such data sets existed in this analysis, but they were included by using this approach.

Verification with respect to reported experimental results

In some cases, means and standard deviations were reported in the paper. The agreement between the data and reported values are shown in table A-1. In general there was fairly good agreement between the two.

Table A-1: Reported and acquired data sets are compared and show good agreement based on their mean and standard deviation.

Set	Reported Mean (mg/dl)	Data Set Mean (mg/dl)	Reported SD	Data SD
Sernorm1	79.7	82.4	N/A	N/A
Sernorm2	82.7	80.4	N/A	N/A
Sernorm3	78.1	77.1	N/A	N/A
Serone1	110.7	112.9	N/A	N/A
Serone2	110.0	108.6	N/A	N/A
Serone3	114.6	114.6	N/A	N/A
Serone4	195.8	197.3	N/A	N/A
Serone5	170.5	168.9	N/A	N/A
Serone6	149.2	146.7	N/A	N/A
Serone7	146.2	142.7	N/A	N/A
Serone8	207.8	210.3	N/A	N/A
Serone9	244.1	240.3	N/A	N/A
Serone10	113.3	110.2	N/A	N/A
Serone11	101.1	97.7	N/A	N/A

Table A-1: Reported and acquired data sets are compared and show good agreement based on their mean and standard deviation (Continued).

Serone12	118.8	119.2	N/A	N/A
Serone13	140.9	138.6	N/A	N/A
Serone14	144.8	143.5	N/A	N/A
Serone15	152.0	147.9	N/A	N/A
Vannorm1	113	113	7	6.3
Vannorm2	140	140	17	17.0
Vannorm3	121	121	10	9.4
Vannorm4	144	143	13	12.9
Vannorm5	114	114	13	12.5

Appendix B: Description and Notes on Selected Data Sets.

Description of Chua's Circuit and Lorenz data set

The Lorenz Attractor was first described by Edward N. Lorenz in the description of atomospheric dynamics [87, 88]. It has since been often used as an investigational model for theoretical study of nonlinear dynamics and chaos. It is described by the coupled set of three differential equations:

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\tau - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$

Chua's circuit was introduced by Leon O. Chua in the 1980's as an experimental instance of a nonlinear system capable of transition to chaos[89]. It is described by the following equations:

$$\frac{dx}{dt} = \alpha[y - x - f(x)]$$
$$\frac{dy}{dt} = x - y + z$$
$$\frac{dz}{dt} = -\beta y$$

Nondiabetics with normal meals and exercise

Mal1-mal12

A total of 12 male subjects were involved in this investigation. Some were sampled every 30 mins and others every hour using direct blood draw. Patients were allowed to rest the night before. Plasma cortisol was measured 6 times as well. Meals were controlled. Some received identical meals others varying meals. Feeding times were all the same.Insulin was additionally measured.

Average Age:

Group A: 29.3 +/- 2.5

Group B: 23 +/- 1.2

Average Weight: 68 KG for both groups, all were male.

Sernorm1-3

Multiple representations of the data allows for testing the digitization algorithm.

Testing seems to have begun around 19:30 for all three cases.

Note the presence of exercise..

Breakfast at 8:00

Lunch at 13:00

Snack at 16:00

Dinner at 18:00

Supper at 21:00

Fasting data

Shanorm1-3

Shapiro et al (ref) present data representing three nondiabetics and nine type II diabetics from a group with matched Body Mass Indices (BMI). Those w/ type II diabetes were being treated with diet and or hypoglycemic agents. Subjects discontinued hypoglycemic agents two weeks prior to the study. The subjects were fasting ten hours prior to the beginning of the data sets. They were kept in a recumbent position throughout the studies. Four of the type II diabetes subjects were studied during different times of the day. The 24hr period of study started at 1400 instead of 800, with the fasting taking place at the same 10 hour period prior to examination (figure B-1).

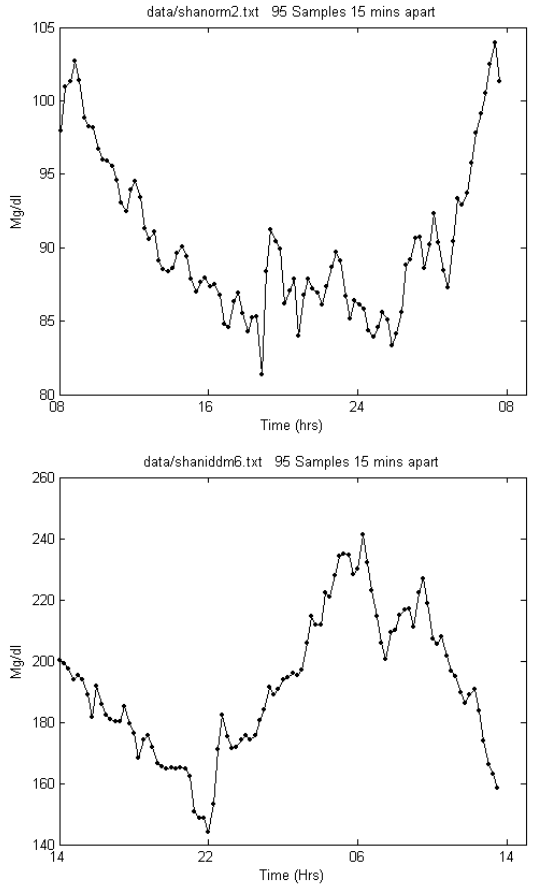


Figure B-1: Two time-series from fasting data. The top panel shows a nondiabetic fasting patient, while the bottom panel shows a type II diabetic patient fasting but whose sleeping cycle has been modified.

Vannorm1-5

This study was meant to look at decreased glucose tolerance during the night. It did so by looking at the nocturnal amplitude of the glucose pulses during continuous glucose infusion. Two rates of infusion 5 g/kg and 8 g/kg (per 24 hour period) were administered for 30 hours and glucose was measured during the last 24 hour period every 15 minutes. The subjects were nonobese healthy volunteers aged 18-25, with

BMI ranging from 18.1-24.4 kg/m^2. The subjects fasted for 12 hours prior to the infusion. The samples were analyzed using a Technicon autoanalyzer. The group applied ULTRA (a pulse analysis program) to the data in order to study the amplitude of the pulses and their correlation with the time of the day. The results were significant and pointed to increased amplitude during nocturnal periods, as well as increased amplitude of oscillations with increased infusion rates.

Polnorm1-3

Here the objective was to study the effect of aging on the changes in the pulses generated by continuous glucose infusion during sleep and wakefulness. In particular growth hormone was studied during sleep and correlated to the behavior of the time-series during sleep. There was a 57 hour infusion of 5 g/kg min (per 24 hour period) that started six hours after breakfast. 53 hours of monitoring sampled at 20 min intervals was initiated. Two periods of sleep were carried on, one from 11PM to 7AM and the latter, after skipping a night of sleep, from 11 AM to 7 PM. This was meant to tease out the effect of sleep as opposed to darkness on the glucose amplitude of the oscillations. The study found increase in amplitude of the oscillations correlates with sleep and not with time of the day, as have other studies, but additionally showed that aging effected the insulin response to this increase.

Polnorm6,7

This was one of the first studies geared towards detected and characterizing the oscillatory behavior of blood glucose levels during continuous infusions of glucose.

Eight subjects 21-33 yrs old were studied after a 10h overnight fast, for 24 hours and were sampled every 15 minutes. The infusion was at a rate of 4.5 mg/kg min, for a period of 24 hours. The study found that there was significant number of pulses in the glucose and insulin profiles.

DirecNet

Characteristics

Original Data Set Size: 112

Usable Subjects: 97

Age Ranges For Diabetics: 3-17, mean 9.4

11 Normals, 86 IDDM

Totals:

4837 Sampling Periods

1689 less than 15 minutes (35%)

3680 less than 36 minutes (76%)

4712 less than 62 minutes (97.5%)

15 durations of 2 hours or more

In scanning the valid data sets, 15 sample lengths of 2 hrs or more were found and they are as follows:

ID 19, 1:30-4 AM No data available

ID 19, 10AM-1PM No data available

ID 23, 7PM-9PM, Data Taken but marked NA

ID 35, 7PM-9PM, Data Taken but marked NA

ID 37, 12-2 PM, Data taken but marked NA at 1PM. QC value is available

ID 44, 5:30-8AM, No data available

ID 44, 8 AM-11 AM (IV insulin test and "other data point")

ID 51, 6PM-8PM, Data Taken but marked NA

ID 74, 1AM-5AM, No data available

ID 75, 6PM-8PM, Data Taken but marked NA, QC value is available, also first data pt

ID 85, 9AM-12:17, No data available after meal

ID 85, 7AM-9AM, Calibration/Other only thing available. END OF DATA SET

ID 85, 9AM-12AM, Calibration/Other only thing available, END OF DATA SET

ID 101, 6PM-8PM, Calibration Only available, Beginning of Dataset

ID 111, 7PM-9PM, No data available

ID 23, 1140-1260 mins, data was removed., END OF DATA SET

Summary Of literature data:

Lengths:

A Total of 185 Tracings

174 tracings 12 hours or longer

134 tracings 24 hours or longer

References:

Drawn from 33 references spanning 1963-1999.

Data Available:

- In many cases age, weight, height, sex and insulin schedule of patient is known.
- 51 data sets contain simultaneous insulin tracing, additional 35 contain ISR
- Additionally, a number of cortisol, and GH tracings are also available.

Sampling:

147 data sets sampled at 15 mins or less

167 data sets sampled at 30 mins or less

Tracing Types:

- 71 Tracings from Normals.
- 52 Tracings from IDDMs
- 18 Tracings from NIDDM
- 18 tracing from pregnant diabetics (IDDM and NIDDM)
- 9 patients with Imparied Glucose Tolerance (IGT)
- 4 patients with insulinomas
- 1 GCK deficiency

“Ideal” Data Sets

Data sets that are sampled at least every 20 minutes, for at least 20 hours, and are done under either normal diet/exercise, continuous enteral nutrition, IV nutrition or fasting:

- 43 Normals.
- 35 IDDMs
- 10 NIDDMs
- 5 IGTs
- 11 Pregnant Subjects with NIDDM and IDDM

Overall Data Set:

- Total of 282 tracings.
- 82 of them from normal individuals under various testing conditions

Appendix C: Selected verification of analysis methods

What follows are selected verifications that were not directly discussed in the methods section in order to keep focus on higher-level discussion of methods.

Spectral estimation techniques

Spectral estimation techniques were verified using sinusoids with known frequency content. The verification of the application of the methods to the more complex time-series was done by confirmatory estimation using other methods, as mentioned in the text. Spectral estimates were made using three methods: AR based model estimation, Prony spectral estimation and FFT based spectral estimation. The results of one such test is shown in figure C-1. Sinusoids of multiple frequency, as well as noise, chaotic time-series and signals with changing frequencies were also applied and agreed with expected results. These are shown in figure C-2.

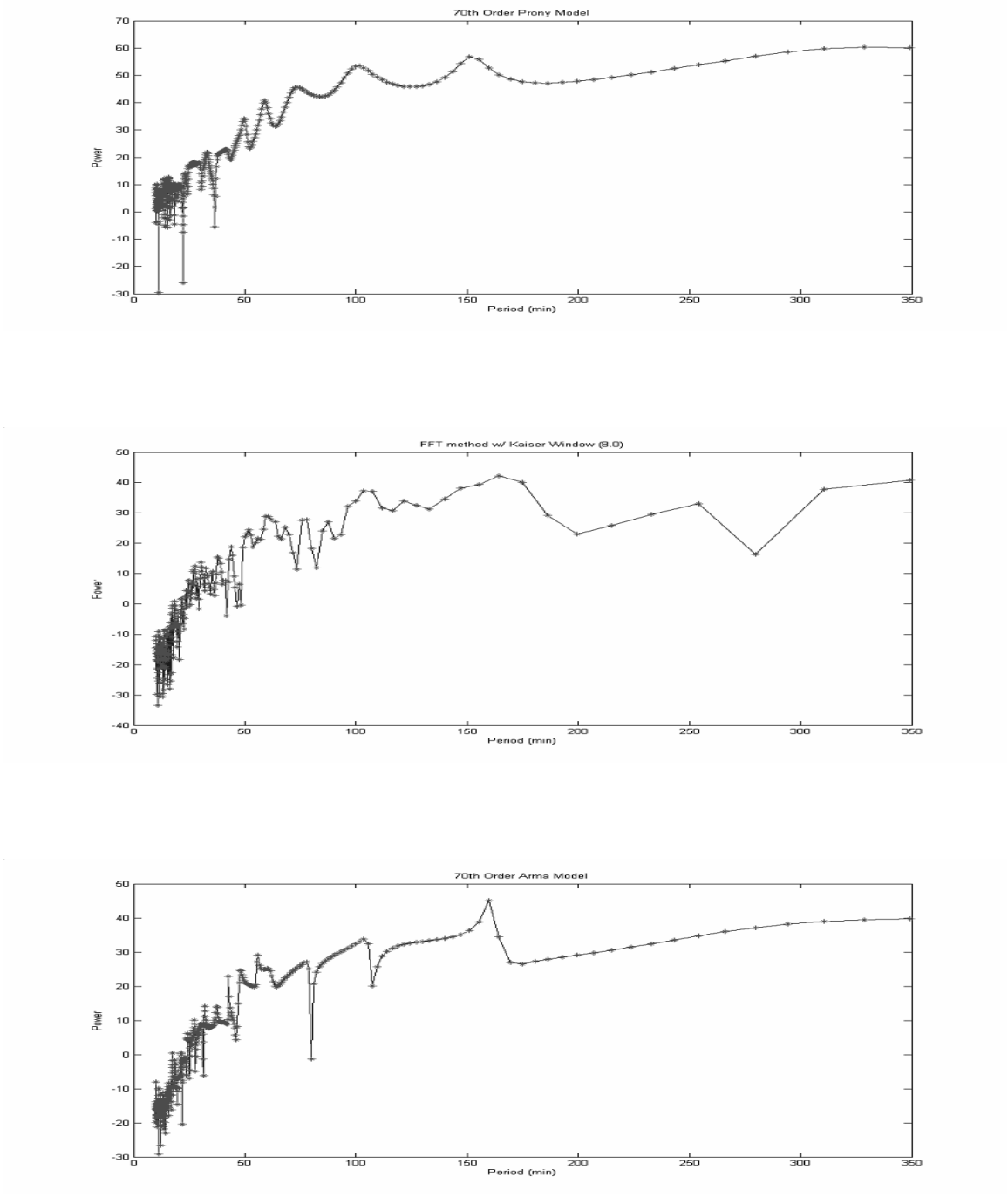


Figure C-1: Three different methods applied to the same time-series with known frequency. A prominent peak is noted in the expected frequency.

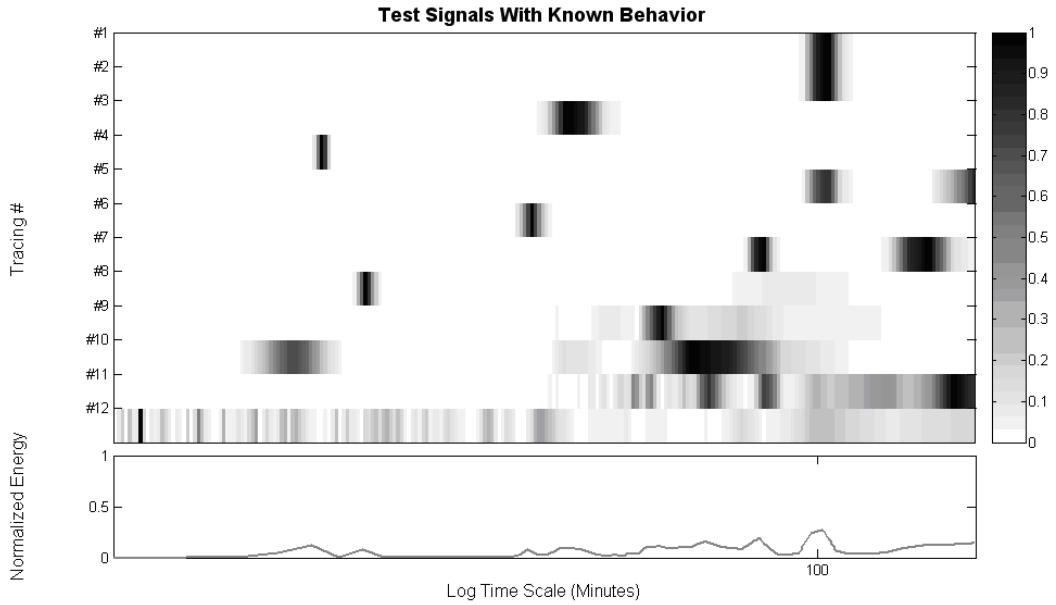


Figure C-2: Test signals with known behaviors tested in the FFT-based spectral analyzer. The set begins with sines with single frequency (#1-4), multiple frequencies (#5-7), changing frequencies and highly nonlinear behavior (#8-11) and ends with noise (#12).

Time-frequency methods

The Short-Time Fourier Transform (STFT) was implemented using the Matlab routine specgram(). A FFT length of ¼ of the length of the signal was used with almost no total overlap (in order to move the window only one sample at a time). A kaiser window of 8.0 was used. The routine was tested with sinusoids of known period and also signals with changing frequency content, including all the signals mentioned in the spectral analysis section preceding this one. An example of these techniques

applied to a signal with fixed frequency content are shown in figure C-3, which shows a constant diffuse band around the correct frequency. It is important to keep in mind that there is a loss of frequency resolution which occurs in trading frequency certainty for time certainty as discussed in the methods chapter.

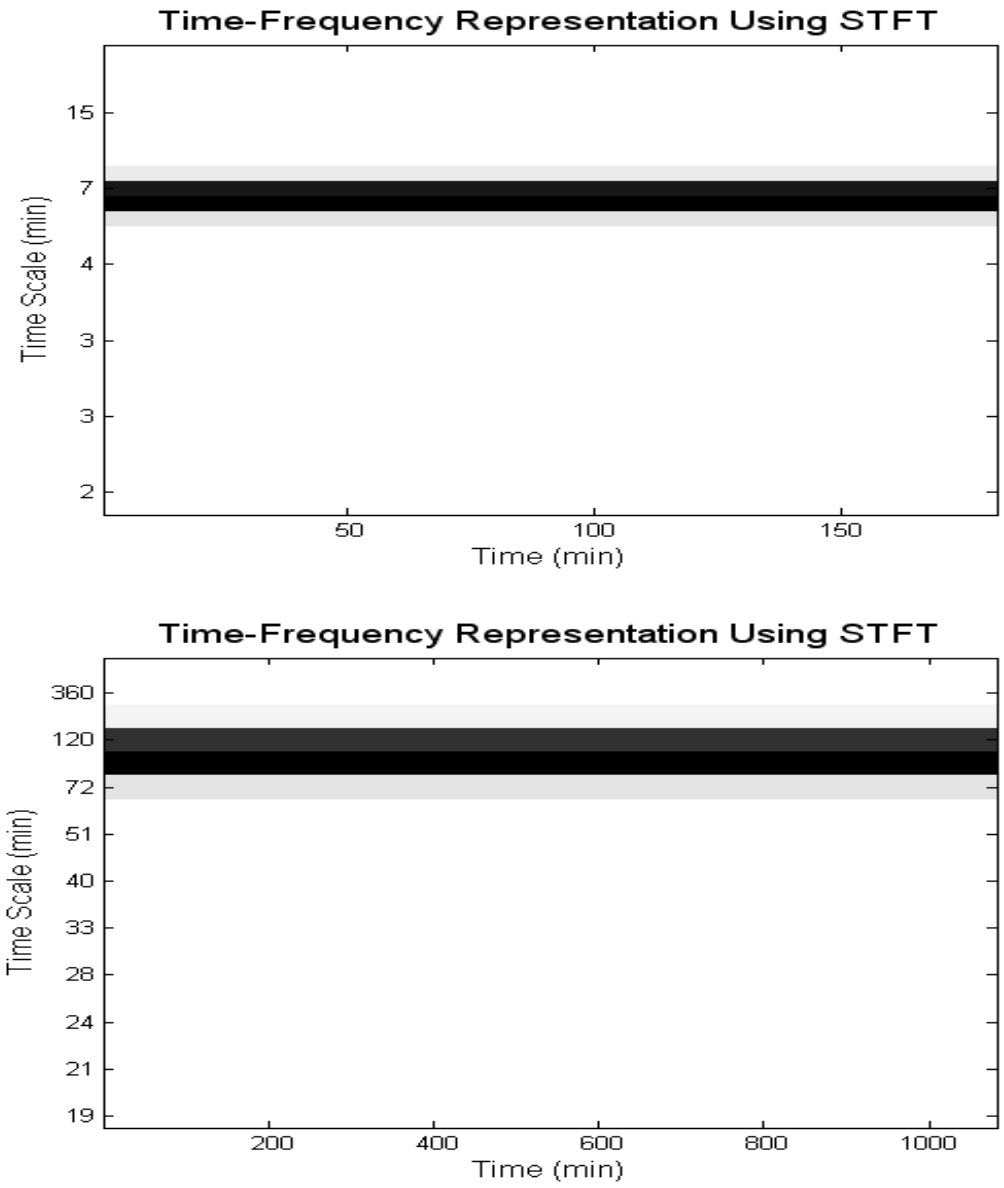


Figure C-3: Example of a sinusoid with a period of ~7 minutes (top) and 100 minutes (bottom). Analyzed using the STFT function. Note the constant but spread out nature of the peak, which is characteristic of all time-frequency analysis where certainty in the time scale domain is exchanged for certainty in the time domain.

Two changing frequencies were then tested and the results are shown in figure C-4. In this case, there is a slow changing high frequency component combined with a fast changing low frequency component.

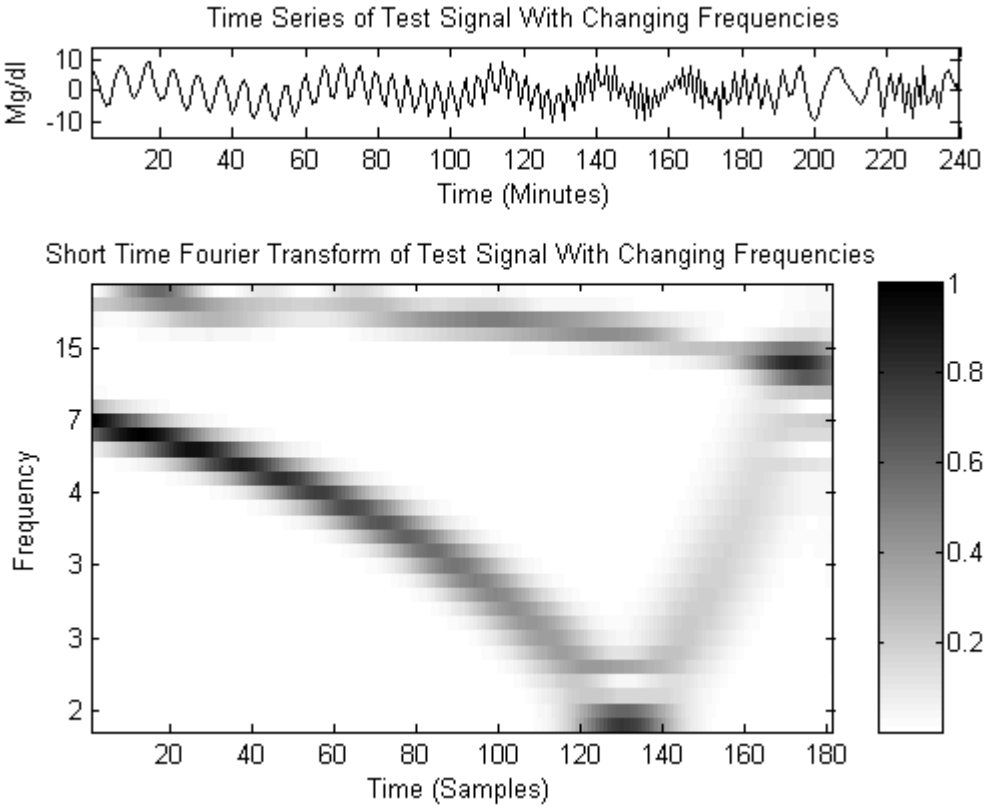


Figure C-4: STFT analysis of two combined signals both with changing frequencies.(tfchanging2.txt)

The same signals were used to test the functionality of wavelet based time-frequency analysis methods. As mentioned in the methods section, the wavelet based methods are largely dependent on the wavelet basis chosen and allow more flexible

“trading” of time and frequency localization. Figures C-5 shows a two constant frequency signals. The second faster frequency signal in particular illustrates the “ringing” of harmonics at lower frequencies as well as the distortions on the beginning and end of the time-series. The changing frequency used in figure C-4 is also shown as the third panel. Clearly, in this case the wavelet basis used is not as optimal as the STFT in showing the frequency progression in time but the irregularity in the signal results in much more of a localized effect (near time = 200) than in the STFT case illustrating the trade-off between frequency and time localization.

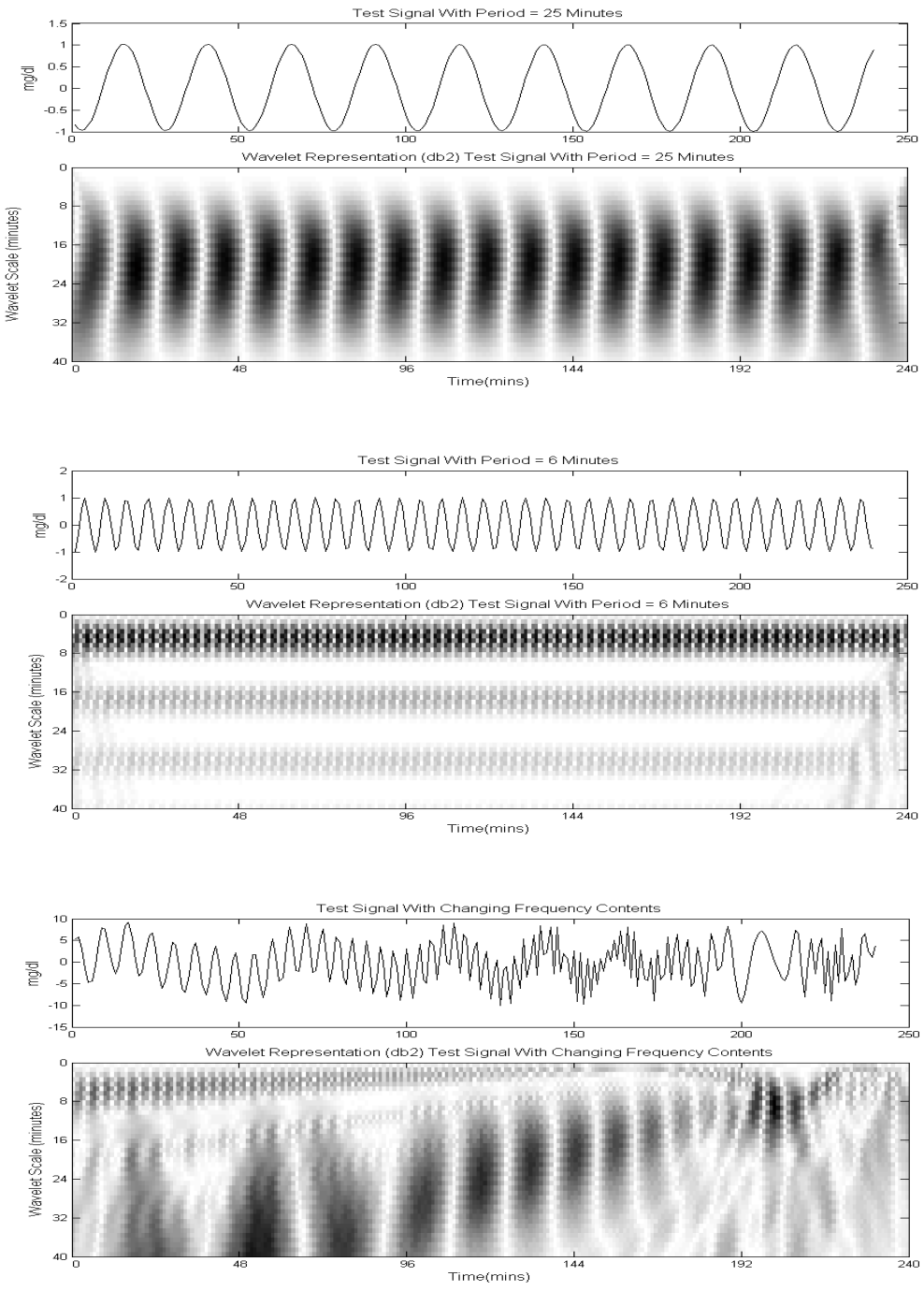


Figure C-5: Examples of wavelet based time-frequency analysis using the Debauchie wavelet basis of order 1. The first to panels represent constant frequency signals whereas the last panel is the same signal used in figure C-4. Loss of frequency localization as compared to the STFT is noted but the localized event of the two frequencies intersecting leads to a more pronounced distortion.

Meal analysis methods

Meals as discussed were analyzed by a priori knowledge of the meal time provided by the investigators, although as demonstrated meals could be automatically detected with significant success. When meals were averaged, they were “aligned” on the beginning. A meal event was taken to be a set period after the initiation of the meal. In cases where a second meal prevented the full length of a meal event to be utilized, the meal time-series snippet was cut short. This led to averaging between time-series of differing length, such that generally the first two hours were result of all meals averaged but the remaining meal profile was an average of progressively fewer time-series. In general, this did not lead to significant discontinuities in the profile, but a linear estimator was used to “patch” the a few noted discontinuities in the resulting time-series, by adjusting the running average to correspond between segments with differing number of meals utilized to compute the average.

Individual meal analysis was performed using simple measures described in the methods section and using the same definition of derivative in order to arrive at rates of change as used throughout this document. All meal analysis was tested and verified using artificially constructed glucose time-series with meal perturbations containing pre-defined maximal velocities, distributions and asymmetries. An example of some of the tools applied to two artificial data sets are shown in table C-1.

Table C-1: Two artificially constructed examples are shown below, with expected results from the construction of the data-set. (one with 20 minute and the latter w/ 30 minute sampling)

	Rise (mg/dl)	Fall (mg/dl)	Time to Max (mins)	Time to Max Rise Rate (mins)	Time to Max Fall Rate (mins)	Max Fall Rate (mg/dl)	Max Rise Rate (mg/dl)
Normfake1							
Meal1	89	66	60	20	0	1.3	1.8
	89	66	60	20	0	1.3	1.8
Meal2	83	122	40	20	0	1.65	3.8
	83	122	40	20	0	1.65	3.8
Meal3	66	96	20	20	80	3.3	.7
	66	96	20	20	80	3.3	.7
Normfake4							
Meal1	61	62	30	30	60	.4	2
	61	62	30	30	60	.4	2
Meal2	66	73	30	30	30	.4	2.2
	66	73	30	30	30	.4	2.2
Meal3	84	32	30	30	60	.5	2.8
	84	32	30	30	60	?	2.8
Meal4	55	101	90	60	60	.9	.75
	55	101	90	60	60	.9	.75

References

1. Skyler, J.S. and C. Oddo, *Diabetes trends in the USA.* Diabetes Metab Res Rev, 2002. **18 Suppl 3**: p. S21-6.

2. Passa, P., *Diabetes trends in Europe.* Diabetes Metab Res Rev, 2002. **18 Suppl 3**: p. S3-8.

3. Zimmet, P., K.G. Alberti, and J. Shaw, *Global and societal implications of the diabetes epidemic.* Nature, 2001. **414**(6865): p. 782-7.

4. Kawamori, R., *Diabetes trends in Japan.* Diabetes Metab Res Rev, 2002. **18 Suppl 3**: p. S9-13.

5. Cahill, G.F., Jr., L.D. Etzwiler, and N. Freinkel, *Editorial: "Control" and diabetes.* N Engl J Med, 1976. **294**(18): p. 1004-5.

6. Stratton, I.M., et al., *Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study.* Bmj, 2000. **321**(7258): p. 405-12.

7. Gabbay, K.H., et al., *Glycosylated hemoglobins and long-term blood glucose control in diabetes mellitus.* Journal of Clinical Endocrinology and Metabolism, 1977. **44**(5): p. 859-64.

8. Shapiro, E.T., et al., *Nocturnal elevation of glucose levels during fasting in noninsulin-dependent diabetes.* Journal of Clinical Endocrinology & Metabolism, 1991. **72**(2): p. 444-454.

9. Fatourechi, V., et al., *Growth hormone and glucose interrelationships in diabetes: studies with insulin infusion during continuous blood glucose analysis.* Journal of Clinical Endocrinology and Metabolism, 1969. **29**(3): p. 319-27.

10. Pincus, S.M. and A.L. Goldberger, *Physiological time-series analysis: what does regularity quantify?* Am J Physiol, 1994. **266**(4 Pt 2): p. H1643-56.

11. Kahn, B.B., *Lilly lecture 1995. Glucose transport: pivotal step in insulin action.* Diabetes, 1996. **45**(11): p. 1644-54.

12. Bremer, T. and D.A. Gough, *Is blood glucose predictable from previous values? A solicitation for data.* Diabetes, 1999. **48**(3): p. 445-51.

13. Bremer, T.M., *Investigation of the predictability of blood glucose dynamics, based solely on preceding blood glucose measurements and implications for the treatment of diabetes mellitus*. 1999.

14. Gough, D.A., K. Kreutz-Delgado, and T.M. Bremer, *Frequency characterization of blood glucose dynamics.* Ann Biomed Eng, 2003. **31**(1): p. 91-7.

15. Hosker, J.P., et al., *Continuous infusion of glucose with model assessment: measurement of insulin resistance and beta-cell function in man.* Diabetologia, 1985. **28**(7): p. 401-11.

16. Vaccaro, O., et al., *Comparative evaluation of simple indices of insulin resistance.* Metabolism, 2004. **53**(12): p. 1522-6.

17. Walton, C., et al., *Evaluation of four mathematical models of glucose and insulin dynamics with analysis of effects of age and obesity.* Am J Physiol, 1992. **262**(5 Pt 1): p. E755-62.

18. Bergman, R.N., et al., *Quantitative estimation of insulin sensitivity.* Am J Physiol, 1979. **236**(6): p. E667-77.

19. Bergman, R.N., L.S. Phillips, and C. Cobelli, *Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose.* J Clin Invest, 1981. **68**(6): p. 1456-67.

20. Bergman, R.N., *Minimal model: perspective from 2005.* Horm Res, 2005. **64 Suppl 3**: p. 8-15.

21. Bergman, R.N., *Lilly lecture 1989. Toward physiological understanding of glucose tolerance. Minimal-model approach.* Diabetes, 1989. **38**(12): p. 1512-27.

22. Bonadonna, R.C., et al., *Roles of glucose transport and glucose phosphorylation in muscle insulin resistance of NIDDM.* Diabetes, 1996. **45**(7): p. 915-25.

23. Simon, C., M. Follenius, and G. Brandenberger, *Postprandial oscillations of plasma glucose, insulin and C-peptide in man.* Diabetologia, 1987. **30**(10): p. 769-73.

24. Picardi, A. and P. Pozzilli, *Dynamic tests in the clinical management of diabetes.* J Endocrinol Invest, 2003. **26**(7 Suppl): p. 99-106.

25. Temelkova-Kurktschiev, T.S. and M. Hanefeld, *Oral glucose tolerance test: to be or not to be performed?* Clin Lab, 2002. **48**(3-4): p. 143-52.

26. Mirouze, J., et al., *[Insulin Efficiency Coefficient. M Coefficient Of Schlichtkrull Corrected And Simplified By The Continuous Blood Glucose Recording Technic.].* Diabete, 1963. **11**: p. 267-73.

27. Service, F.J., et al., *Mean amplitude of glycemic excursions, a measure of diabetic instability.* Diabetes, 1970. **19**(9): p. 644-55.

28. Schlichtkrull, J., O. Munck, and M. Jersild, *The M-Valve, An Index Of Blood-Sugar Control In Diabetics.* Acta Med Scand, 1965. **177**: p. 95-102.

29. Wojcicki, J.M., *Mathematical descriptions of the glucose control in diabetes therapy. Analysis of the Schlichtkrull "M"-value.* Horm Metab Res, 1995. **27**(1): p. 1-5.

30. Zanno, C., et al., *[Evaluation of diabetes control by new numerical indices of daily blood glucose values].* Minerva Med, 1978. **69**(11): p. 683-96.

31. Simon, C., G. Brandenberger, and M. Follenius, *Ultradian oscillations of plasma glucose, insulin, and C-peptide in man during continuous enteral nutrition.* J Clin Endocrinol Metab, 1987. **64**(4): p. 669-74.

32. Sturis, J., et al., *24-hour glucose profiles during continuous or oscillatory insulin infusion. Demonstration of the functional significance of ultradian insulin oscillations.* J Clin Invest, 1995. **95**(4): p. 1464-71.

33. Malherbe, C., et al., *Circadian variations of blood sugar and plasma insulin levels in man.* Diabetologia, 1969. **5**(6): p. 397-404.

34. Mejean, L., et al., *Circadian and ultradian rhythms in blood glucose and plasma insulin of healthy adults.* Chronobiol Int, 1988. **5**(3): p. 227-36.

35. Simon, C., et al., *Slow oscillations of plasma glucose and insulin secretion rate are amplified during sleep in humans under continuous enteral nutrition.* Sleep, 1994. **17**(4): p. 333-8.

36. Simon, C., L. Weibel, and G. Brandenberger, *Twenty-four-hour rhythms of plasma glucose and insulin secretion rate in regular night workers.* Am J Physiol Endocrinol Metab, 2000. **278**(3): p. E413-20.

37. Daneman, D., et al., *Severe hypoglycemia in children with insulin-dependent diabetes mellitus: frequency and predisposing factors [see comments]*. Journal of Pediatrics, 1989. **115**(5 Pt 1): p. 681-5.

38. Scheen, A.J., et al., *Alterations in the ultradian oscillations of insulin secretion and plasma glucose in aging.* Diabetologia, 1996. **39**(5): p. 564-72.

39. Sturis, J., et al., *Differential effects of glucose stimulation upon rapid pulses and ultradian oscillations of insulin secretion.* J Clin Endocrinol Metab, 1993. **76**(4): p. 895-901.

40. O'Meara, N.M., et al., *Lack of control by glucose of ultradian insulin secretory oscillations in impaired glucose tolerance and in non-insulin-dependent diabetes mellitus.* J Clin Invest, 1993. **92**(1): p. 262-71.

41. Sturis, J., et al., *Entrainment of pulsatile insulin secretion by oscillatory glucose infusion.* J Clin Invest, 1991. **87**(2): p. 439-45.

42. Porter, P.A., et al., *Incidence and predictive criteria of nocturnal hypoglycemia in young children with insulin-dependent diabetes mellitus [see comments]*. Journal of Pediatrics, 1997. **130**(3): p. 366-72.

43. Pfeiffer, E.F., et al., *On line continuous monitoring of subcutaneous tissue glucose is feasible by combining portable glucosensor with microdialysis.* Hormone and Metabolic Research, 1993. **25**(2): p. 121-124.

44. Gumbiner, B., et al., *Abnormalities of insulin pulsatility and glucose oscillations during meals in obese noninsulin-dependent diabetic patients: effects of weight reduction.* J Clin Endocrinol Metab, 1996. **81**(6): p. 2061-8.

45. Sturis, J., et al., *Abnormalities in the ultradian oscillations of insulin secretion and glucose levels in type 2 (non-insulin-dependent) diabetic patients.* Diabetologia, 1992. **35**(7): p. 681-9.

46. Santiago, J.V., W.L. Clarke, and F. Arias, *Studies with a pancreatic beta cell simulator in the third trimester of pregnancies complicated by diabetes.* American Journal of Obstetrics and Gynecology, 1978. **132**(4): p. 455-63.

47. Simon, C., et al., *Alteration in the temporal organisation of insulin secretion in type 2 (non-insulin-dependent) diabetic patients under continuous enteral nutrition.* Diabetologia, 1991. **34**(6): p. 435-40.

48. Villaume, C., et al., *28-hour profiles of blood glucose (BG), plasma immunoreactive insulin (IRI) and IRI/BG ratio in four insulinomas.* Ann Endocrinol (Paris), 1984. **45**(3): p. 155-60.

49. Mauras, N., et al., *Lack of accuracy of continuous glucose sensors in healthy, nondiabetic children: results of the Diabetes Research in Children Network (DirecNet) accuracy study.* J Pediatr, 2004. **144**(6): p. 770-5.

50. Cohen, L., *Time-Frequency Analysis*. 1995, Englewood Cliffs, NJ: Prentice Hall.

51. Strogatz, S.H., *Nonlinear dynamics and Chaos : with applications to physics, biology, chemistry, and engineering*. 1994, Reading, Mass.: Addison-Wesley Pub.

52. Oppenheim, A.V. and R.W. Schafer, *Digital signal processing*. 1975, Englewood Cliffs, N.J.,: Prentice-Hall.

53. DeFatta, D.J., J.G. Lucas, and W.S. Hodgkiss, *Digital signal processing : a system design approach*. 1988, New York: Wiley.

54. Percival, D.B. and A.T. Walden, *Spectral analysis for physical applications : multitaper and conventional univariate techniques*. 1993, Cambridge ; New York, NY: Cambridge University Press.

55. Kay, S.M., *Fundamentals of statistical signal processing*. Prentice Hall signal processing series. 1993, Englewood Cliffs, N.J.: PTR Prentice-Hall.

56. Felber, J.P., et al., *Role of lipid oxidation in pathogenesis of insulin resistance of obesity and type II diabetes.* Diabetes, 1987. **36**(11): p. 1341-50.

57. Mallat, S.G., *A Wavelet Tour of Signal Processing.* 1998.

58. Huang, W., et al., *Engineering analysis of biological variables: an example of blood pressure over 1 day.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 4816-21.

59. Marple, S.L., *Digital spectral analysis : with applications*. Prentice-Hall signal processing series. 1987, Englewood Cliffs, N.J.: Prentice-Hall.

60. Costa, M., A.L. Goldberger, and C.K. Peng, *Multiscale entropy analysis of complex physiologic time series.* Phys Rev Lett, 2002. **89**(6): p. 068102.

61. Costa, M., A.L. Goldberger, and C.K. Peng, *Multiscale entropy analysis of biological signals.* Phys Rev E Stat Nonlin Soft Matter Phys, 2005. **71**(2 Pt 1): p. 021906.

62. Pincus, S.M., *Irregularity and asynchrony in biologic network signals.* Methods Enzymol, 2000. **321**: p. 149-82.

63. Lake, D.E., et al., *Sample entropy analysis of neonatal heart rate variability.* Am J Physiol Regul Integr Comp Physiol, 2002. **283**(3): p. R789-97.

64. Richman, J.S. and J.R. Moorman, *Physiological time-series analysis using approximate entropy and sample entropy.* Am J Physiol Heart Circ Physiol, 2000. **278**(6): p. H2039-49.

65. Glass, L. and D. Kaplan, *Time series analysis of complex dynamics in physiology and medicine.* Med Prog Technol, 1993. **19**(3): p. 115-28.

66. Barnett, W.A., et al., *A Single-Blind Controlled Competition among Tests for Nonlinearity and Chaos.* Journal of Econometrics, 1998. **82**(1): p. 157-192.

67. Kaplan, D.T., *Exceptional Events as Evidence for Determinism.* Physica D, 1994. **73**: p. 38-48.

68. Abarbanel, H.D.I., *Analysis of Observed Chaotic Data*. 1996, New York: Springer.

69. Satchell, S., *A Bias Correction for Taken's Correlation Dimension Estimator.* Econometric Theory, 1994. **10**(2): p. 439-440.

70. Kantz, H. and T. Schreiber, *Nonlinear Time Series Analysis*. 1997, Cambridge: Cambridge University Press.

71. Porksen, N., *The in vivo regulation of pulsatile insulin secretion.* Diabetologia, 2002. **45**(1): p. 3-20.

72. Beaser, R.S., L.P. Krall, and Joslin Diabetes Center, *Joslin diabetes manual*. 12th / ed. 1989, Philadelphia: Lea & Febiger. xxxi, 406.

73. Diabetes, C. and G. Complications Trial Research, *The absence of a glycemic threshold for the development of long-term complications: The perspective of the Diabetes Control and Complications Trial.* Diabetes, 1996. **45**: p. 1289-1298.

74. Galloway, J.A. and R.E. Chance, *Improving insulin therapy: achievements and challenges.* Hormone and Metabolic Research, 1994. **26**(12): p. 591-8.

75. The DCCT Research Group, *Epidemiology of severe hypoglycemia in the diabetes control and complications trial.* American Journal of Medicine, 1991. **90**(4): p. 450-9.

76. Veneman, T., et al., *Induction of hypoglycemia unawareness by asymptomatic nocturnal hypoglycemia.* Diabetes, 1993. **42**(9): p. 1233-1237.

77. Bhatia, V. and J.I. Wolfsdorf, *Severe hypoglycemia in youth with insulin-dependent diabetes mellitus: frequency and causative factors.* Pediatrics, 1991. **88**(6): p. 1187-93.

78. Casparie, A.F. and L.D. Elving, *Severe hypoglycemia in diabetic patients: frequency, causes, prevention.* Diabetes Care, 1985. **8**(2): p. 141-5.

79. Cryer, P.E., *Hypoglycemia: The limiting factor in the management of IDDM.* Diabetes, 1994. **43**(11): p. 1378-1389.

80. Groop, L.C., E. Widén, and E. Ferrannini, *Insulin resistance and insulin deficiency in the pathogenesis of type 2 (non-insulin-dependent) diabetes mellitus: errors of metabolism or of methods?* Diabetologia, 1993. **36**(12): p. 1326-31.

81. Groop, L.C., et al., *Glucose and free fatty acid metabolism in non-insulin-dependent diabetes mellitus. Evidence for multiple sites of insulin resistance.* Journal of Clinical Investigation, 1989. **84**(1): p. 205-13.

82. Baron, A.D., *Impaired glucose tolerance as a disease.* Am J Cardiol, 2001. **88**(6A): p. 16H-9H.

83. DeFronzo, R.A., E. Ferrannini, and D.C. Simonson, *Fasting hyperglycemia in non-insulin-dependent diabetes mellitus: contributions of excessive hepatic glucose production and impaired tissue glucose uptake.* Metabolism: Clinical and Experimental, 1989. **38**(4): p. 387-95.

84. Matschinsky, F.M., *Regulation of pancreatic beta-cell glucokinase: from basics to therapeutics.* Diabetes, 2002. **51 Suppl 3**: p. S394-404.

85. Le Roith, D., *Tumor-induced hypoglycemia.* N Engl J Med, 1999. **341**(10): p. 757-8.

86. Lathi, B.P., *Signal Processing & Linear Systems*. 1998, Carmichael: Berkeley Cambridge Press.

87. Lorenz, N.E., *Deterministic non-periodic flows.* Journal of Atmospheric Science, 1963.

88. Matsumoto, T., Chua, L. O., and Komuro, *"The Double Scroll."* M. IEEE Trans. CAS, 1985. **32**: p. 798-818.

89. Chua, L.O., Desoer, Charles A. and Kuh, Ernest S. , *Linear and Nonlinear Circuits*. 1987: McGraw-Hill.