2010

A Time Series Analysis: Exploring the Link between Human Activity and Blood Glucose Fluctuation

Eric A. Sadowski
Wilfrid Laurier University

Your file Votre référence
ISBN: 978-0-494-75378-1
Our file Notre référence
ISBN: 978-0-494-75378-1

A Time Series Analysis: Exploring the Link between Human Activity and Blood Glucose Fluctuation

By

ERIC A. SADOWSKI

BA, WLU, 2007

THESIS

Submitted to the Department of Geography and Environmental Studies

in partial fulfillment of the requirements

for the Master of Arts

Wilfrid Laurier University

2010

Abstract

In this thesis, time series models are developed to explore the correlates of blood glucose (BG) fluctuation of diabetic patients. In particular, it is investigated whether certain human activities and lifestyle events (e.g. food and medication consumption, physical activity, travel and social interaction) influence BG, and if so, how. A unique dataset is utilized consisting of 40 diabetic patients who participated in a 3-day study involving continuous monitoring of blood glucose (BG) at five minute intervals, combined with measures for sugar; carbohydrate; calorie and insulin intake; physical activity; distance from home; time spent traveling via public transit and private automobile; and time spent with other people, dining and shopping. Using a dynamic regression model fitted with *autoregressive integrated moving average* (ARIMA) components, the influence of independent predictive variables on BG levels is quantified, while at the same time the impact of unknown factors is defined by an error term. Models were developed for individuals with overall findings demonstrating the potential for continuous monitoring of diabetic (DM) patients who are trying to control their BG. Model results produced significant BG predicting variables that include food consumption, exogenous insulin administration and physical activity.

Acknowledgments

This thesis is based on collaboration with Dr. S. Doherty, who has provided inspiration and attended to this work in various ways. I am especially grateful for him continually challenging me in this pursuit and serving as an exemplary mentor.

Additionally, I would like to thank the faculty and staff from the Department of Geography and Environmental Studies at Wilfrid Laurier University who have assisted and supported me while providing nurturing guidance throughout my career. Specifically, I would like to thank Dr. B Sharpe for his sagely advice in the early stages of my thesis development; Dr. C. Carmichael for her critiques of this work; Pam Schaus for her cartographic input; and Luke Cwik for sharing his computer science skills.

Finally, I would like to express my appreciation to my family for their loving support. I dedicate this work to my parents.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1 Introduction

Enhancing our understanding of human behaviour and related health impacts is of interest to researchers from the fields of geography, psychology, planning and health. Despite modest improvements to data collection techniques over the last couple of decades, the challenge remains to assemble data collection instruments that go beyond simulated experiments under laboratory experiments and capture high quality real world events.

Recent advances in wearable technology hold promise for ushering in an era of renaissance for improving our understanding of human behaviour under real-world conditions. The merger of geographic location and body-motion tracking devices small enough to fit in a cell phone is affording the opportunity to passively collect detailed spatial/temporal real-world data over extended periods of time. Being able to collect detailed real-world data allows for the examination of human behaviour geographies (e.g. activities, travel, exercise, social interaction). Combining tracking devices with wearable biophysical sensors allows for further analysis of specific health outcomes. In practical terms, this translates to allowing the identification of specific events and or activities in real space and time that may have an impact on a variety of lifestyle events, while maintaining the ability to aggregate such data in order to examine their cumulative affects on health.

Obesity has reached pandemic proportions in North America resulting in exorbitant health care costs (between 2-7% of total health care costs (WHO 2007)) and has become a major public health concern around the world. A number of health complications have been linked to obesity, including: psychosocial problems, certain cancers, cardiovascular problems, and type 2 diabetes (Saarloos et al. 2009). Furthermore, research suggests that behaviour

modification – healthy dieting and increasing physical activity (Sallis et al. 1998, Giles-Corti et al. 2002, Boarnet 2005, Lytle 2009) – can help individuals properly manage their weight and mitigate the risk of developing health complications such as diabetes of which there is currently no cure for and the management of is extremely challenging. Enhancing our understanding of the effect of place and other activities on people suffering with a chronic disease, such as diabetes, may assist in improved patient outcomes and disease management. A better understanding may assist in the development of 'smart-agent' health monitoring/management systems, the planning of healthy urban environments and compliment existing health policies and practices. Moreover, by combining spatio-temporal and health information data with time-series analysis techniques, provides researchers with the opportunity to better understand how to manage a chronic disease like diabetes.

This thesis explores the unique opportunity to dynamically analyze diabetic patients' BG levels in relationship to lifestyle events such as: food consumption, physical activity, and travel. Additionally, this thesis will evaluate the effectiveness of time-series analysis in which to predict for BG.

The rationale for this work is to explore the possibility of predicting for BG in diabetic patients. It is important to better understand how or what causes BG to change in diabetic patients in order to properly manage this chronic disease. Results from this thesis will contribute toward better understanding what variables (which includes a variety of both previously tested and new exploratory variables) affect BG in diabetic patients.

1.1 Objectives:

The overall purpose of this thesis is to gain a better understanding of how diabetic patient's BG may be influenced by a variety of factors (daily activities,

food and medication consumption) as they change over time and space. In particular, the objective of this study is threefold:

(1) To employ a time-series analysis technique to a unique real-world data set in order to explore BG variation on a patient-by-patient basis;

(2) To explore how a wide range of variables influence BG variation, including traditional variables related to food and exercise, and new variables related to lifestyle and geography;

(3) To evaluate the effectiveness of such data collection and analysis techniques for the purpose of predicting BG and other time lagged health outcomes.

1.2 Overview of thesis

The thesis is organized into five chapters. A *Literature Review* following this introduction provides a synopsis of theoretical frameworks and past data collection methodologies to highlight how this thesis will contribute to the existing literature. Chapter 3 presents the data collection *Methodology* and *Analysis Technique* employed in this thesis and identifies the strengths and weaknesses of these approaches. In Chapter 4 *Results* are shared while Chapter 5 provides a *Discussion* and *Conclusion* and an overview of the implications for future work, the challenges and limitations of the approach.

Chapter 2: Literature Review

2 Literature Review

2.1 Behavioural geography health foundations

Human behaviour is influenced by environmental factors through the presence of individuals and their activities behaving in response to both social and physical settings (Lindheim et al. 1983, Moudon et al. 2003, Aitken et al. 2006). In fact, people not only adapt to their social and physical environments but also are part of other people's environment (Grimm et al. 2005). A multitude of forces affect both individuals and the environment; for instance, the location of businesses influence travel including distances travelled. Understanding the multi-layered system of interaction is necessary when environmental design is meant to establish desired behaviour or lifestyle changes; for example, to encourage cycling and walking, or to create safer and healthier places (Timmermans et al. 1990, Saarloos et al. 2009).

In the last decade, studies have been able to demonstrate how the environment influences overweight and obesity in individuals. The highlight of this claim is an association between the built environment and physical activity (Hill et al. 1998, Frank et al. 2004, Boarnet 2005, Booth et al., Humpel et al. 2002, Papas et al. 2007) resulting in the recent designation of obesogenic environments. Obesity strongly depends on behavioural issues and can be traced to an increase in sedentary lifestyles (Sallis et al. 1998, Frank et al. 2001, Bauman et al. 2002, Sallis et al. 2006, Colabianchi et al. 2007) and nutritionally poor diets (Sobal 2002, Booth et al. 2005, Cummins et al. 2006). Adapted from Saarloos et al. (Saarloos et al. 2009), Figure 1 demonstrates how a variety of variables may influence an individual's health.

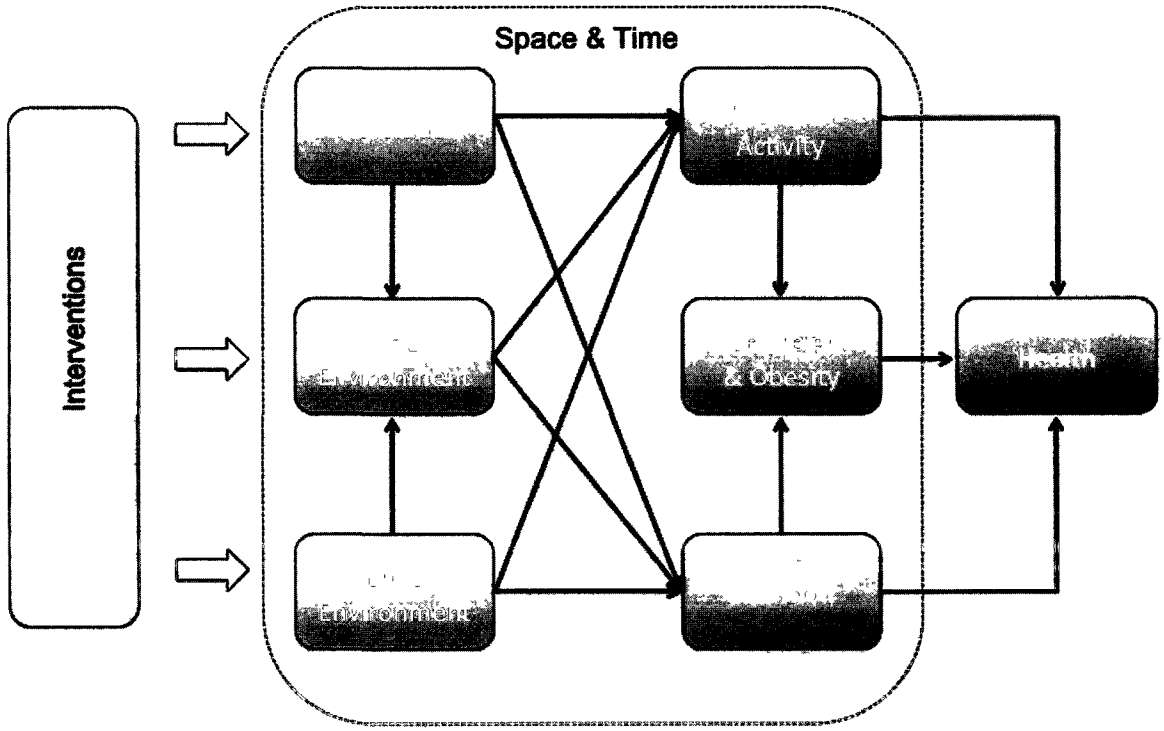


Figure 1 Space-time behaviour as the underlying concept for studying the health impacts of environmental interventions (here related to overweight and obesity)(Saarloos et al., 2009)

Saarloos et al (2009 p. 1735) describe public health interventions as:

"establishing health-enhancing changes at the population (or system) level ... actual changes to the built environment (e.g., providing more opportunities for recreation, improving access to healthy food, and creating places to accommodate social events for community members), interventions that address individuals (e.g., campaigns aimed at improving attitudes to physical activity, nutrition and social interactions, or increasing awareness of the health benefits or risks involved), and interventions to encourage social activities (e.g., community programs to promote physical activity and healthy food consumption). To support decision-making about health promotion interventions, studies will need to aim at understanding and predicting how such interventions will change individual health behaviours in space and time."

In geography, behaviour as a response to an individual's surrounding environment is a topic generally assigned to the realm of behavioural geography

with discourse overlapping with other disciplines such as environmental psychology and behaviour. Since the early 1970s, this field has contributed models and theories of preference and choice widely used to predict the impact of policy decisions on spatial behaviour as well as to assess the plausibility of envisaged projects (Timmermans et al. 1990). Although health has been discussed in land use and transportation planning (Brennan Ramirez et al. 2006), a lack of discussion of health exists in the behavioural geography realm highlighting the need for a multidisciplinary approach that integrates views and methods from health promotion, behavioural geography and planning disciplines (Saarloos et al. 2009). However, despite the agreement regarding the legitimacy of a holistic approach, epidemiological research is firmly oriented towards isolating the effects of specific attributes instead of using more integrative and systematic approaches that may map out causal relationships between interventions and health outcomes (Diez Roux 2007). This piecemeal approach results in a gap between policy and practice, and research. A current challenge faced is how to effectively align urban planning, environment design and behaviour modification with public health, whilst optimally incorporating the strengths of each field in order to create healthy and sustainable environments (Duhl et al. 1999). However, while environmental and transportation planning have had a focus on health in the past it is difficult to see the success of their efforts in improving health, especially with a rise in obesity suggesting perhaps more of an emphasis on behaviour modification.

2.2 Space-time behaviour and health

Capturing simultaneously where and when individuals are engaged in activities allows for the quantification of both the spatial extent and time intensity of different activities (Rainham et al. 2008, Duncan et al. 2009). In the past, information regarding time, activity and location were usually collected through observations (shadowing), personal interviews, or through time-diaries (Frank et al. 2004) opening a myriad of data quality and privacy issues. Today,

more sophisticated data collection techniques are being used to emulate, by way of automation, these past efforts (Doherty 2009, Papinski et al. 2009). Data loggers and electronic sensors have been employed with limited success within the context of time-activity and transportation research studies (Ermes et al. 2008). These techniques have been incorporated in health studies for the purpose of detecting physical activity information (Elgethun et al. 2003, Elgethun et al. 2007, Yang et al. 2009) and with some improvement these techniques have the potential for advancing individual-based modeling and allowing for a better understanding of human health behaviour.

In the past, regression-based analysis techniques have been applied to detect associations between environmental measures and health-related measures. Typically, studies have indexed the built environment by way of GIS procedures and used non-spatial data collected from residents regarding their nutrition habits (e.g. type, amount and frequency) or weight and physical activity habits (e.g. intensity, duration and frequency) to identify associations between these health-related measures and environmental measures using regression-based techniques (Figure 2). The contemporary approach circumvents how health behaviours really occur due to individuals interacting with their environment and each other (Auchincloss et al. 2008, Diez-Roux 2008). In order to model and forecast environmental impacts on health behaviour, such as nutrition and physical activity, it is necessary to understand how activities are linked in space and time at the individual level (King et al. 2008). In particular, it needs to be understood what health-related activities people are engaging in, where, when, the duration, with whom and so on.

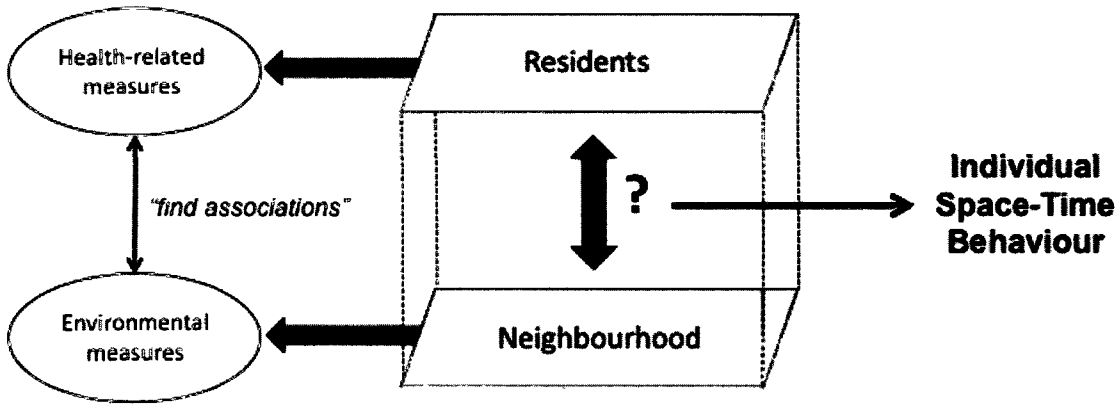


Figure 2 The contemporary approach to study environmental impacts on health behaviours (left), and the missing link of individual space-time behaviour (right)(Saarloos et al., 2009)

More recently, the advent of new technologies has sparked an evolution from aggregate to individual-level models with the adoption of new data collection techniques. Making use of geographical positioning system (GPS) instruments has granted access to data at a fine-grained scale required for study at a level of high detail. Adoption of these new methods provide the impetus for the greatly needed shift from the general thesis that individual neighbourhoods are responsible for affecting health to the specifics of why and how this may occur (Diez Roux 2002). Oriented at individual behaviour, an integrated approach would now be possible in which the focus is not solely on statistical relationships between characteristics of particular health issues and the built environment, but rather on how the built environment influences, among other variables, the way in which people organize their travel and activities in space and time and how these subsequently affect their health (Saarloos et al. 2009).

2.2.1 Emerging Data collection approaches for Tracking Human Activity

GPS has emerged as the most popular location detection device owing to its its small size, frequent updating (every second), wide-scale accuracy (generally accurate to within 5 meters), availability and low cost. Main outputs include longitude, latitude, altitude, speed, time, and signal strength. GPS receivers acquire information from a system of satellites allowing for location to be calculated using triangulation (Bajaj et al. 2002). GPS technology for the purpose of human tracking offers a tremendous "opportunity for improving the understanding of how characteristics of environmental context and place come to influence human behaviour, health and well-being" (Rainham et al. 2008). Furthermore, recent technological advances have made possible wireless sensor networks for the purpose of human health monitoring. Wearable sensors are able to monitor a wide range of biophysical metrics, including electro-cardio (heart-rate), respiration, BG, body temperature, and body motion/acceleration. Moreover, an increasing number of implantable devices are also evolving, for instance, defibrillators, visual/auditory sensory aides, drug administrators, pacemakers, and emerging implantable nanotechnologies. Such sensors can provide instantaneous feedback to medical personnel (Miyazaki et al. 2009) and are expected to revolutionize health monitoring in the near future (Lymberis 2004, Baert et al. 2006, Otto et al. 2006, Pandian et al. 2008, Meyfroidt et al. 2010, Pantelopoulos et al. 2010).

As reported in recent review by Doherty (2009), accelerometers have long been used in the health sciences for assessing physical activity/motion owing to their small size, non-invasive nature, low respondent burden, and ability to provide objective measures of both physical activity intensity, duration and type over long periods of time. Under lab conditions, estimates of intensity (energy expenditure) and type (e.g. walking, running, rowing, cycling, sitting, standing, and lying down) have been accurate to 95%+; however, accuracy

typically drops to below 67% under out-of-lab free-living conditions (Foerster et al. 1999, Welk 2002, King et al. 2004, Parkka et al. 2006, Ermes et al. 2008) owing in large part to the increased complexity/diversity of physical activity in daily life.

In light of these new developments, an opportunity exists for both geographers and health scientists to combine body-based biophysical sensors and automated sensing of human activities to examine the influence of human behaviour geographies (activities, travel, exercise, social interaction) as they relate to specific health outcomes (Doherty 2009).

2.3 Diabetes Mellitus

The medical condition known as Diabetes Mellitus (DM) has attracted much attention from a variety of disciplines. Research efforts tend to focus on improving the treatment and/or management (Mshelia et al. 2007, Towfigh et al. 2008, Parkin et al. 2009) of the disease with a new focus on prevention gaining momentum.

DM is a metabolic disorder distinguished by high concentrations of BG (hypoglycemia) that have the potential to result in total or partial failure of the pancreas to secrete and produce insulin – the hormone responsible for distributing glucose to cells (Souto et al. 2010). There are two types of DM: DM type I where the body fails to produce insulin and; DM type II where the body is inefficient at regulating insulin. Various forms of treatment exist and usually involve some degree of patient self-management and collaboration with health practitioners. A variety of medications are available that can be taken orally or intravenously to help regulate BG. Additionally, typical disease management recommendations and behaviour modification include lifestyle adjustments in the form of weight loss, increase in physical activity and adaption of healthy eating habits. Furthermore, the prevalence of DM is increasing and serves as an

additional financial strain on health care systems. Diseases associated with DM include: chronic renal failure, maligned tumors, stroke, lower limb amputation, myocardial infarction, retinopathy and coma. However, early diagnosis and management can prevent such complications (Allemann et al. 2009).

Past studies have linked lifestyle – including diet and physical activity levels – to diabetes, thereby suggesting behaviour modification as a form of prevention, treatment and or management (Schwedes et al. 2002, Murata et al. 2004), (Guerci et al. 2003, Evans et al. 2004, Glasgow et al. 2006a, Glasgow et al. 2006b). Existing evidence shows the benefits of eating healthy for those with diabetes, which include weight loss or maintenance, BG control and maintenance of blood pressure within a specified reference range (Sargrad et al. 2005). Such evidence has prompted the American Diabetes Association to depict medical nutritional therapy as "an integral component of diabetes self-management education (or training)", the primary goal of which is to "assist and facilitate individual lifestyle and eating behaviour changes," leading to "improved metabolic control, a reduced risk in complications and improved health." Additionally, researchers suggest that the maintenance and adoption of eating healthy has been recognized as a key topic that should be given priority in future research (Jones et al. 2003, Murata et al. 2004, Mshelia et al. 2007, with particular research attention devoted to issues concerning how to modify unhealthy behaviours and develop healthy eating habits {Sargrad, 2005 #289, WHO 2007, French et al. 2008, Bellazzi et al. 2009, Jovanovic 2009, Parkin et al. 2009, Poolsup et al. 2009).

Recently, a large number of studies (Giskes, 2007, Allen et al. 1990, Glasgow et al. 1997, Kumanyika et al. 2000, Hu et al. 2001, Schwedes et al. 2002, Welschen et al. 2005, Tarnus et al. 2008) have presented dietary change and healthy eating interventions in a variety of settings. These studies employ many different mechanisms that promote modifying diet, ranging from a prescriptive approach to patient-centered empowerment approaches (Allemann et al. 2009).

Such studies use a range of outcome measures examining the effectiveness of interventions, varying from clinical outcomes such as hemoglobin A1c (HbA1c) and weight to behavioural outcomes such as alterations to nutrients (e.g. fat and fiber intake).

2.3.1 Blood glucose monitoring

By checking a patient's BG level at different times throughout the course of a day, researchers are able to examine the effects of various interventions, including, for instance: medication, insulin, caloric intake, sugar intake, carbohydrate intake or physical activity (Wen et al. 2004, Istepanian et al. 2009, Poolsup et al. 2009). Measurement of HbA1C is an alternative way to assess average BG levels over a longer period of time. Red blood cells in the blood stream are composed of hemoglobin, and BG adheres to hemoglobin to form HbA1C; thus, the higher the BG, the higher the HbA1C. Given that red blood cells have a life span of 8-12 weeks, measuring HbA1C serves as an indicator of the average BG level over the last 8-12 weeks. A non-diabetic HbA1C should be 3.5-5.5 and in diabetic patients 6.5 is considered good. Both short-term and long-term monitoring is essential for a diabetic, as an increase in blood glucose levels may result in further health complications (Souto et al. 2010).

2.3.2 Blood glucose influencing factors

Past studies have identified many factors as having an influence on BG, including diet, physical activity, travel burden, stress, depression, weight, and medication. The two most widely studied are food consumption and physical activity (Evans et al. 2004, Lustman et al. 2007, Barnett et al. 2008, Healy et al. 2008, Sardinha et al. 2008, Bellazzi et al. 2009, Helmerhorst et al. 2009, Istepanian et al. 2009, Parkin et al. 2009).

Regular physical activity is beneficial to a diabetic as it results in an increased uptake of glucose by skeletal muscle and is independent of insulin creation. The time period in which glucose is synthesized is debatable, however the effect on insulin sensitivity is suggested to last 12 to 48 hours after exercising (Parkin et al. 2009) . According to Healy et al. (2008) when engaged in physical exercise, whole-body oxygen consumption can increase by as much as 20-fold, with even greater increases occurring in working muscles. In order to meet energy needs under these conditions, skeletal muscle consume at a greatly increased rate, stored glycogen, triglycerides, and free fatty acids (FFAs) that result from the decomposition of glucose and adipose tissue triglycerides released from the liver. In an effort to preserve optimal central nervous system function, the body's blood glucose levels are maintained remarkably well during physical exercise. Hypoglycemia while engaged in physical activity hardly ever occurs in non-diabetic individuals. The metabolic adjustment responsible for preserving normoglycemia during physical exercise is substantially regulated hormonally. These hormonal adaptations are lost in insulin deficient patients suffering with type 1 DM, resulting in increased levels of ketone and BG and serve as a precursor to diabetic ketoacidosis. Conversely, the presence of elevated levels of insulin, resulting from exogenous insulin administration, can attenuate or even prevent the increased mobilization of other substrates and BG induced by physical exercise, and hypoglycemia may ensue. Similar concerns exist amongst patients with type 2 diabetes; however, typically, hypoglycemia during physical exercise tends to be less of an issue in the type 2 DM population group. Certainly, in patients with type 2 DM, physical exercise may improve insulin sensitivity in addition to assisting in lowering elevated BG levels to normalcy (Healy et al. 2008)

Consumption of carbohydrates and sugar most greatly influence blood glucose as both are converted to glucose. Research suggests that this conversion may take 15 minutes to 2 hours (Souto et al. 2010). Overall caloric intake is also associated with metabolic control having an impact on blood glucose (Murata et

al. 2004). Hu et al. (2001) examined the relationship between different types of dietary fat and carbohydrates – non-hydrogenated polyunsaturated and trans-fatty acids – emphasizing the quality of fats and carbohydrates as opposed to quantity as having an impact on reducing the risk of DM. In the same vein, Hung et al. (2003) examined insulin resistance in response to various types of foods, and verifies that diets higher in monounsaturated fatty acids, fiber and low glycemic index foods appear to have advantages in insulin resistance and BG control.

Closely observing BG response to food consumption and exercise over time has been shown to be an effective way of exploring BG change (Wallace et al. 2008). Tarnus et al. (Tarnus et al. 2008) examine food impacts on BG by calculating BG response after meals using blood-testing strips. BG was found to increase rapidly after meals and was also found to have a significant positive correlation with fat mass percentage. Boule et al. (2001) review the effect of exercise on HbA1C and body mass index in DM patients concluding that exercise reduces HbA1C by a significant enough amount that could decrease the risk of future health complications. Marcus et al. (2008) compare the outcomes between a diabetes exercise program that makes use of aerobic and high force eccentric resistance exercise and aerobic only exercise. Results from this study conclude significant improvements in long-term glycemic control were demonstrated in both groups.

Apart from food intake and exercise, other correlates of BG fluctuation have been less extensively studied. Exceptions include Strauss et al. (2006) who identify 'travel burden' as a contributing factor to the inability to control BG concluding that those who live closer to their centre of primary care tend to have better control over their BG than those living farther away. Additionally, Wiesli et al. (2005) identify stress as causing a delay in lowering BG, and Lustman et al. (2007) discovered DM patient depression as having a negative effect on the ability to control for BG. In order to better understand the effects of such

variables as they occur in real time, an appropriate time series analysis technique is required.

2.4 Time-series analysis

When using statistical methods there is an inherent assumption that observed data are realizations of independent random variables. However, if data are observed repeatedly over time, such as physiological or physical processes, it is likely that successive observations are dependent partly on previous observations and other possible explanatory variables in a time lagged fashion. Box and Jenkins (1979) were the first to develop a method for constructing time series models in practice that accounted for these dependencies (Helfenstein 1991, 1996, Imhoff et al. 1996). They introduced a method of identifying, estimating, and diagnosing what is now commonly referred to as 'the Box-Jenkins approach' and *Autoregressive Integrated Moving Average* (ARIMA) models.

After introducing their book in 1970 their method has been used to analyze a variety of data from fields such as industry, economics, water resource research and health. For example, this approach has been used to analyze daily concentrations of atmospheric pollutants. The challenge of such analysis is the dependence of consecutive measurements reflecting the tendency for high levels of pollutant concentration on one day to lead to high levels the next day (positive autocorrelation). The Box-Jenkins method of time series analysis became popular for its ability to account for such autocorrelation by assessing relations between a 'target', 'output' or 'endogenous' series and an 'explanatory', 'input', 'exogenous' or 'predictor' series; these situations are represented by a 'transfer function' (TF) model (Greaves et al. 2007).

Time series analysis of physiological phenomena presents more of a challenge. They require techniques that are able to control for seemingly chaotic signal artifacts, temporal dynamics (non-stationarity) and nonlinearity. Examples of physiological time series include: heart rate, skin conductance, diastolic and systolic blood pressure, oxygen consumption, ventilation, electromyograph, electroencephalograph, and electrocardiograph. Researchers believe that these complex time series data sets may possess "hidden information" that promises to be of clinical value, such as forecasting a hypoglycemic episode (Harvey et al. 1996, Goldberger et al. 2000, Duchene et al. 2007, Bellazzi et al. 2009, Takeuchi et al. 2009, Meyfroidt et al. 2010). The advent of powerful methods for storing and disseminating vast amounts of health information has lead many researchers to believe a major breakthrough is eminent at both the basic levels and clinical levels of investigation. Although sophisticated computational tools help with the selection process, a challenge for researchers lies in choosing an appropriate analysis technique.

The selection of an approach for heterogeneous multivariate physiological time series analysis is highly contentious. Some suggest that physiological time series metrics are chaotic (stochastic) and require nonlinear quantitative methods (Saalasti 2003). However, more recent findings (Glass 2009) that physiological metrics are indeed dependent and predictable based upon various explanatory variables. Past efforts have involved rudimentary attempts to employ techniques to better understand what events may influence BG where the validity of results is questionable due to the experimental nature of the data collected under laboratory conditions, small sample size and inappropriate analysis techniques. Takeuchi et al. (2009) examined one subject over 12 days using a Pearson Correlation function to compare BG levels 30 minute intervals before and after eating and exercise and concluded that exercise resulted in a decrease in BG within 30 minutes after exercising and food consumption resulted in an increase in BG an hour after consumption. Bellazzi and Abu-Hanna (2009) describes the application of 'data mining' for the purpose of

assessing BG and DM influencing variables and suggest the time series data mining approach is well suited for optimizing control of BG and DM.

2.5 Summary and Context of Current Study

Advances in technology have afforded the opportunity to enhance our understanding of how individuals make use of time and space and explore the role these factors may have on health and well-being. In this thesis, the impact of human activity on BG is explored at a very fine spatial-temporal scale made possible by data collected using the latest automated tracking technologies. In addition to exploring well known correlates of BG related to food and exercise, this approach allows for the analysis of the impacts of travel, activity engagement, socializing, and distance from home – which have scarcely been explored to date. Owing to the nature of blood glucose fluctuation over time, an appropriate time-series analysis approach that controls for autocorrelation between residuals is to be adopted, the details of which are discussed at length in the next chapter.

Chapter 3: Methods

3 Methodology

This chapter presents the data collection and analysis methods. This thesis utilizes an existing dataset collected in 2007 as part of a research study at Toronto Rehab funded by the Canadian Institutes of Health Research, and led by Dr. Sean Doherty and Dr. Paul Oh. The description of data collection process and instruments used is derived from previous descriptions of this project, including Doherty and Oh (2006), and Doherty and Oh (forthcoming). Details and images are reproduced here with permission.

3.1 Multi-sensor Diabetic Patient Monitoring System

The data collection methodology consists of a set of wearable sensors wirelessly connected to a BlackBerry smartphone with a continuously running software program that compresses and transmits the data to a central server for storage, processing, display and for purposes of further interaction with patients. The smartphone and sensors, shown in Figure 3, include: a GPS receiver; 3-axis accelerometer and ECG and heart rate monitor; and a continuous Blood Glucose (BG) monitor. Figure 3 demonstrates how all the instruments can be worn.

For this methodology, an off-the-shelf 3-axis accelerometer with Bluetooth wireless data transfer capabilities was sought, resulting in choice of the Alive Heart Rate Monitor. In general terms, an electronic component in 3-axis accelerometers measures acceleration of the body-part it is attached to in three directions - side-to-side (X axis), up/down (Y axis), and back-forth (Z axis), as shown in Figure 4. Measurements are taken at frequent sub-second intervals – in this case 25 times a second. The data are transmitted over a Bluetooth serial port as an interleaved stream containing the accelerometer and ECG measurements. The custom software developed on the BlackBerry reads the wireless stream of

data and takes the min, max, average and standard deviation of the data over every second.

In a similar fashion, the GPS receiver transmits second-by-second data via Bluetooth to the Blackberry. This includes the raw NMEA (National Marine Electronics Association) 0183 standard sentence that includes longitude, latitude, altitude, speed, heading, and signal quality variables (# satellites, displacement). Both the accelerometer and GPS data are then transmitted to a central server using a Java software program developed for the BlackBerry. Data were logged to the BlackBerry until it reached a set size (e.g. 50KB), then compressed and transmitted via the BlackBerry Enterprise Server to the project server (e.g. every 5 minutes), and stored in a MySQL database under the set username. To the best of the researchers' knowledge, this is the first application of such a varied set of sensors; additionally unique is the connection to a smartphone for purposes of wireless data transmission under real-world conditions.

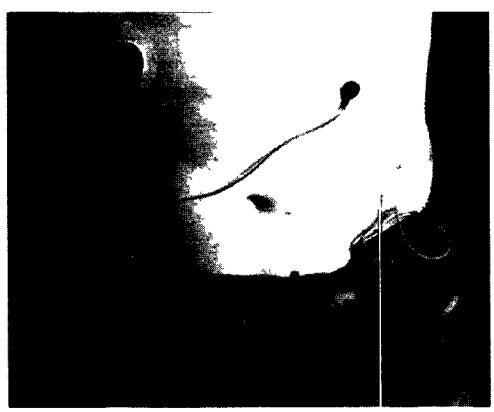


Figure 3 Example Placement of Sensors (c) Sean Doherty, 2006

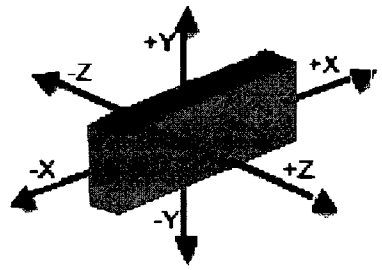


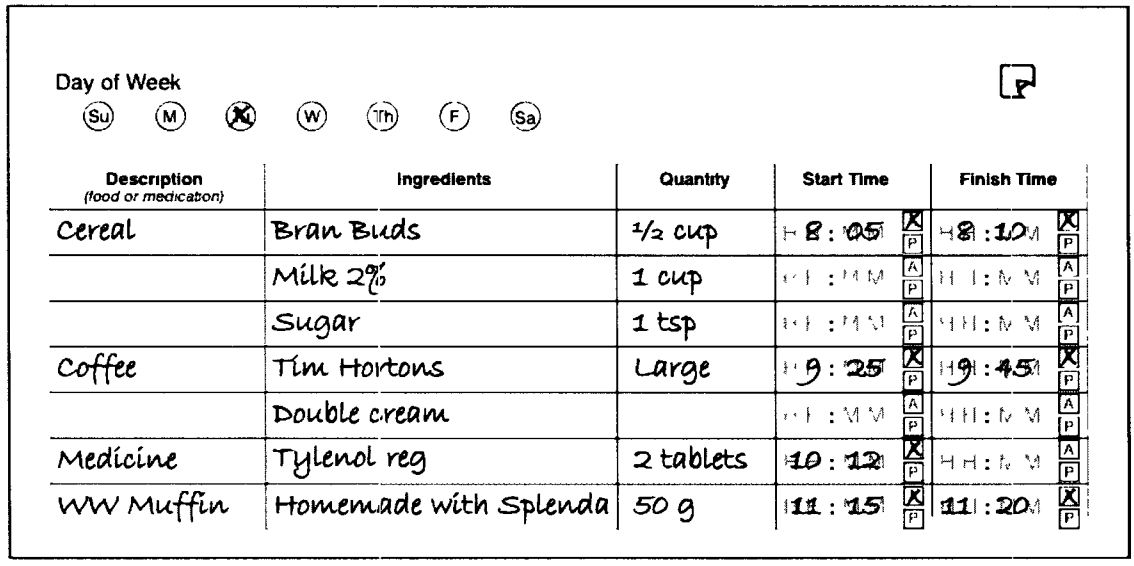
Figure 4 Accelerometer showing axis orientation and values (+ or -)(c) Sean Doherty, 2006

Two additional data collection instruments were developed requiring manual input from subjects: 1) A food/medicine diary; and 2) An internet-based interactive "prompted recall diary" for displaying patient's activity-travel patterns on a nightly basis for confirmation and supplemental information prompts.

The food/medicine diary was designed in consultation with a nutritional therapist and specifically targeted the types, amounts and *timing* of consumption throughout the day, as shown in Figure 5. Instructions to users are shown in the figure. A Bluetooth enabled digital pen was used to capture digitally exactly what was written on each page of the diary. The digital data was then transmitted wirelessly via Bluetooth to the BlackBerry and subsequently transmitted to a central server for storage and analysis. During post analysis, the data on types and amounts of food were entered into the Nutritionist Pro™ Diet Analysis software. Using a knowledgebase of over 35,000 + foods, recipes and ingredients, the software outputs a thorough nutrient analysis, including sugars, carbohydrates, calories, vitamins, etc. These outputs were then time stamped for each subject and uploaded to the central server.

By combining GPS, accelerometer, prompted recall, and food/medicine data an automated activity detection algorithm was employed to identify a wide variety of events subjects may have been engaged in. For instance, the algorithm

was able to detect whether a subject was travelling by private automobile or public transit. Additionally, the algorithm could detect if one was at school/work, home, or shopping. Additionally, the combination of accelerometer and GPS data can be used to detect the mode of travel (walk, run, bicycle etc.) regarding speed (Duncan et al. 2009), while intensity of activity can be measured with accelerometer data measured in vector magnitude (Vmag). Furthermore, prompt to recall data allows for the incorporation of social activities. Intermixing these various forms of data allows for an automated compilation of an activity diary.



Day of Week

Su M ~~T~~ W Th F Sa

Description (food or medication)	Ingredients	Quantity	Start Time	Finish Time
Cereal	Bran Buds	½ cup	8:05	8:10
	Milk 2%	1 cup		
	Sugar	1 tsp		
Coffee	Tim Hortons	Large	9:25	9:45
	Double cream			
Medicine	Tylenol reg	2 tablets	10:12	
WW Muffin	Homemade with Splenda	50 g	11:15	11:20

Figure 5 Food Diary (c) Sean Doherty, 2006

3.2 Diabetic Patient Pilot study

A pilot study utilizing the above multi-sensor data collection system was completed in 2007 involving 40 diabetic patients from the Toronto Rehabilitation Institute (TRI) in collaboration with caregivers and medical doctors. The goals were threefold: 1) to develop and test the value of a continuous system of multi-instrument/sensor data capture and wireless data transfer; 2) assess the practicality of the prompted recall diary; and 3) explore how the data could be used to assess correlates of Blood Glucose (BG)

fluctuation as a function of specific spatial-temporal events (exercise, working long hours, food, etc.). This thesis represents the first attempt to address the latter goal, following from early efforts at preparing the data for analysis.

Forty diabetic patients were recruited via advertisements and word-of-mouth within the Toronto Rehabilitation Institute. There was one pre-diabetic, four type I and thirty-five type II DM patients. Patients were given gift certificates valued at $30 CAD as an incentive and token of appreciation.

A 72-hour monitoring period was decided for the pilot study, reflecting the upper limit of the duration to which the blood glucose monitor could be worn continuously. Glucose readings are extracted every 10 seconds and recorded in the monitor as a five minute average. Unlike the other sensors however, no capabilities existed at the time to transmit these data wirelessly. Instead, the data were downloaded off the device at the end of the study period using software provided by the manufacturer. This study period covered a four-day period – on the first day, participants were set up at the rehab clinic at a time convenient for them, resulting in a partial day of monitoring. The next two days (day 2 and 3) were full days of monitoring. On the fourth day, they returned to the clinic for debriefing, again resulting in a partial day of monitoring.

A two-hour upfront interview on the first day was used to hook up the various sensors and train subjects on their use. A full hour of this time was spent with the blood glucose monitor, owing to its invasive nature (a sensor is implanted under the skin by a nurse). Patients were given the option of wearing remaining devices on their belt (as shown in Figure 3) in a small "fanny pack" style pouch provided, or in their own purse/bag.

Outside of recharging the BlackBerry and GPS receiver, the only time patients were instructed to manually interact with the devices was if the BlackBerry issued a long "buzz". In this case, an automated message would

appear on screen (generated by the on-board software) instructing patients to manually turn on one of the sensors which may have inadvertently become disabled, recharge a device, or call a research assistant (which was done by selecting "OK" and using the BlackBerry; no dialing was needed).

A full-time research assistant monitored all incoming data, addressed any equipment problems, and assisted patients with all aspects of data collection. Two to three patients were setup and monitored per week. In support of the study, detailed notes were taken on sensor performance, user issues, data quantity/quality, and other challenges that were faced (although not analyzed here).

3.3 Analysis Techniques

The following subsections introduce basic time-series concepts that serve as the foundation for data analysis in this thesis, including the multistep process of ARIMA modeling and its fundamental principles. Much of the technical material in the following section is derived from previous works (Pankratz 1983, Wei 1990, Pankratz 1991, Yaffee et al. 2000, Greaves et al. 2007, Box et al. 2008, Montgomery et al. 2008).

3.3.1 Time Series Data

Wei (1990) describes a time series as "an ordered sequence of observations. Although the ordering is usually through time, particularly in terms of some equally spaced intervals, the ordering may also be taken through other dimensions such as space." Time series may be measured either discretely or continuously. Continuous time series are generally recorded steadily and instantaneously; for example, an oscillograph records harmonic oscillations of an audio amplifier. Measurements made at regular intervals are considered to be discrete; for instance, rainfall accumulations measured at a regular interval. Alternatively, data may be pooled together, taken from individual observations in order to create a summary statistic that has been measured at regular intervals over time. As such, particular linear series that are not chronologically ordered might be adaptable to time series analysis. Ideally, however, the time series should consist of observations equidistant from one another in time. This thesis examines raw data and summary statistics measured at regular intervals over time, for which time series analysis is most appropriate.

3.3.2 Box-Jenkins Models

Introduced by Box and Jenkins in 1970, the *autoregressive integrated moving average* (ARIMA) model may be performed on one variable, or multiple input variables without much complication in the form of preprocessing, as is typically the case with other methods. ARIMA models are considered to be superior to traditional time-series and multivariate regression models as these latter models' error residuals are often correlated with their own lagged values. This correlation violates standard assumptions of regression theory including that disturbances are therefore not correlated with other disturbances. Indeed, as part of early development for this thesis, multivariate regression models were developed and were indeed found to have goodness-of-fit and error measures that were unsatisfactory in comparison to the results obtained using ARIMA models that followed.

ARIMA models have three modeling parameters: 1) the *Autoregressive* or AR (p) process which takes into consideration historical data serving as a memory of past events; 2) an *Integrated* or I (d) process that stabilizes the time series by making the data stationary and assists in making the forecasting process easier; and 3) a *Moving Average* or MA (q) process which accounts for forecast errors and controls for autocorrelation. The more historical data, the more accurate the forecast will be, as the model learns over time. This model is sometimes referred to as a p,d,q or Box-Jenkins model.

3.3.3 The Basic Uni-variate ARIMA Model

Univariate models are sometimes referred to as noncausal models. Basic in nature, their objective is to describe the time-series by way of residual correlation and to make predictions on the formulated model as opposed to explain what influences them. In general, a univariate time series will reflect the

reality in which observations occurring close in time have a greater relationship than observations that are farther apart, looking only at the single (uni) variable. It is the purpose, therefore, of univariate time-series methods to statistically measure the degree of this relationship. The standard form of ARIMA (p,d,q) is:

$$\nabla^d y_t = \mu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + a_t - \Theta_1 a_{t-1} - \Theta_2 a_{t-2} - \ldots - \Theta_q a_{t-q}$$

where μ serves as a constant; Φ_p serves as an autoregressive component indicating the relationship of y_t to y_{t-p}; Θ_q is a moving average coefficient indicating how y_t relates to a_{t-q}; a_t is a random shock component; and ∇^d is the difference operator used when the time series is not stationary. In the case of a first order difference ($y_t - y_{t-1}$), d is equal to one; in the case of a second order difference ($(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$), d equals two. It is possible to use a backshift operator ($B^i y_t = y_{t-i}$) to make the basic form equation more compact:

$$\Phi(B)\nabla^d Y_t = \mu + \Theta(B)a_t$$

where:

$\nabla^d = (1 - B)^d$ (d-order differencing operator)

$\Phi(B) = (1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p)$ (p-order AR operator)

$\Theta(B) = (1 - \Theta_1 B - \Theta_2 B^2 - \ldots - \Theta_q B^q)$ (the q-order MA operator)

3.3.4 Mulitvariate ARIMA Model

A multivariate time-series is a combination of multiple univariate time-series; multivariate time-series analysis involves examining the impact of an event, or independent variable, on a dependent variable or time-series.

Alternatively, multivariate time-series methods investigate the relationship between an exogenous (forcing) series, X_t, and an endogenous (response) series, Y_t, where a dynamic system may exist – in which an input series appears to be related to an output series. Transfer function models are used to assess this relationship with input series and response series cross-correlated by way of a transfer function (TF). These models are commonly referred to as causal models, where a variation in the exogenous series or independent variable, is utilized to explain a variation in the endogenous variable or dependent variable series. The exogenous variables can include continuous variables or dummy indicators highlighting the presence of an intervention or a stochastic series, which drives the response series. These types of models are used to test explanatory relationships between time-dependent processes that are hypothesized to exist (Yaffee et al. 2000).

3.3.5 The Linear Transfer Function Model:

To overcome possible problems of omitted time-lagged input terms, autocorrelation in the disturbance series, and common correlation patterns among the input and output series that yield spurious correlations, the dynamic or linear transfer function (LTF) model can be very useful (Pankratz 1991) and is employed for analysis purposes in this thesis. The LTF combines aspects of the ARIMA and allows for input and output series to be separate time series, where the endogenous response output series (Y_t) is a function of the exogenous forcing input series (X_t); hence the name transfer function. The output series (Y_t) or transfer function model is considered to contain two components and is typically formulated as:

$$y_t = V(b) + \eta_t$$

where $V(b)$ may be explained using one or several exogenous input series (X_t) and consists of a response regressed on lagged autoregressive endogenous (Y_t)

variables and lagged exogenous (X_t) variables. η_t is the noise model component ARIMA time series error model and represents the unexplained part of (Y_t) ('noise process') adopted from Helfenstein (1996) as shown in Figure 6.

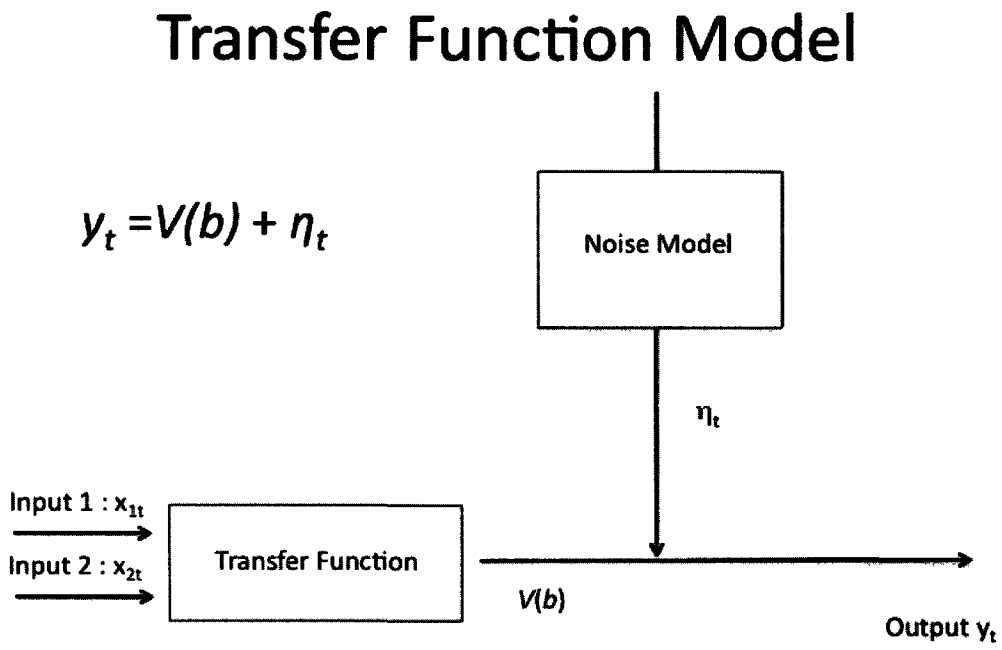


Figure 6 Schematic representation of a transfer function model with multiple inputs (Helfenstein, 1996)

3.3.6 Structure of the linear transfer function

If y_t depends on the lagged polynomial $x_{t,I}$ (or $V(b)$), in anyway, the relationship can be represented by:

$$y_t = v_{0,1}X_{t,1} + v_{1,1}X_{t-1,1} + \ldots + v_{0,2}X_{t,2} + v_{1,2}X_{t-1,2} + \ldots + v_{0,I}X_{t,I} + v_{1,I}X_{t-1,I} + \ldots = v(B)_I X_{t,I}$$

where, $\upsilon(B)$ is the transfer function or impulse response function [$\upsilon(B)_I = \upsilon_{0,1} + \upsilon_{1,I}B + \upsilon_{2,I}B^2 + ...$]. Individual weights, v_I in $v(B)_I$ are referred to as impulse response weights and represent changes in the endogenous output series as a result of a unit variation in the explanatory variable at the specified time. Each of the impulse response weights (v_I) can be understood as a response to an input at a given point in time. At time $i = t$, the degree of change to the output variable is relative to the magnitude of the input variable coefficient, v_t. According to Yaffee et al. (2000) for instance, "after one time period has elapsed, the response of the endogenous output (Y_t) variable is equivalent to the product of the coefficient times the value of the input variable at that time plus the same products at previous time periods" (Yaffee et al. 2000 p. 355). That is, the output response is thereby equal to the sum of the products from the impulse weights, multiplied by the value of the input variables from the inception of the process through the current time period. When $v_I < 0$, the direction of the impulse response from the current time is the opposite of the value of the input variable. On the other hand, if $v_I > 0$, than the direction of the response from the current time is the same as that of the value of the input variable. $v(B)_I$ may have a large number of v_I weights – possibly infinity – and so we represent the transfer function $v(B)_I$ as Wei (1990) and Box et al. (1994) have in the following rational parsimonious form:

$$v(B)_I = \frac{\omega(B)_I (B)^b}{\delta(B)_I}$$

where

$$\omega(B)_I = \omega_{0,I} - \omega_{1,I}B - \omega_{2,I}B^2 - ... - \omega_{h,I}B^s$$

$$\delta(B)_I = 1 - \delta_{1,I}B - \delta_{2,I}B^2 - ... - \delta_{r,I}B^r$$

ω is a regression weight, δ represents decay weights and B^b incorporates dead time input (or delay) from i. For instance, y_t might not react immediately to a

variation in $x_{t,I}$ but later responds at y_{t+I}. In this example, $v_{0,1}$ equals zero and dead time, noted by b, equals one. x_t may not be just a stochastic process but can also be a binary variable, which is usually assigned unity to depict the presence of events and zero to note the absence of an event.

Therefore, based on the previous equation, an Auto-regressive Integrated Moving Average (ARIMA) model can be expressed as:

$$\Phi(B)\nabla^d Q = \mu + \sum_{i=1}^{n} \frac{\omega(B)_i (B)_i^{b}}{\delta(B)_i} \nabla^d x_{i,t} + \Theta(B)a_t$$

or, more simply

$$y_t = v(B)x_t + \eta_t$$

3.3.7 ARIMA Modeling: Linear Transfer Function Approach

The type of data a researcher wishes to analyze will dictate ultimately which model parameters should be used for forecasting purposes. In certain applications, an output-input type of transfer function, or intervention model, may not be appropriate as a result of complex feedback and feedforward – lag or lead – relationships may exist among variables. In other applications it is quite possible that a number of different models may fit the data well. Wei (1990), Yaffee (2000), Box et al. (1979) and Montgomery et al. (2008) attest that in these circumstances it is up to the researcher to decide, based on their historical knowledge and familiarity with the data, which model is most appropriate.

When addressing such complex situations as in this study, such as a large sample size and multiple inputs, suitable statistical software is required. Here we employ Predictive Analytics Software (SPSSv18), which has the ability to

perform ARIMA – also called Box-Jenkins Models – TF models. This software is able to perform a number of automated procedures to assist with the modeling process; for example, automatically detect outliers and adjust accordingly. These and other automated procedures will be discussed in more detail in later sections of this chapter.

In general, the modeling process is iterative and involves an initial estimation of parameters *p,d,q*; followed by a performance evaluation of the model; a re-identification of the parameters; and a comparison of results via goodness-of-fit testing. These stages require considerable work and they themselves may not be exhaustive. This process is often repeated in order to find the best-suited parameters and goodness-of-fit measures for each specific model. Previously, much of this work was done manually and repeatedly; the software used in this thesis is able to select the best-fitting model, saving the researcher a considerable amount of time making this research possible and cutting edge; otherwise, the extent of analysis conducted in this thesis would not be plausible. Albeit, a number of intrinsic modeling steps and concepts need to be discussed in order to fully conceptualize the modeling process.

In this study a multivariate time-series ARIMA model is employed to take into account the effect of independent variable inputs ($x_{t,I}$) as well as their previous disturbances on predicting future values of a dependent variable (y_t) while controlling for autrocorrelation between residuals. The dependent variable examined in detail is the 5-minute average BG level on a subject-to-subject basis. As such, a transfer function (TF) will be utilized similar to a study conducted by Helfenstein (1996) where this technique was used to examine the effect of insulin treatment on BG of a single DM patient. A general weakness of the technique involves the use of historical information where a sharp change in the dependent variable has taken place. These changes therefore will only be taken into account through the certain interval of time skewing the models accuracy up until that point.

The following sections outline in detail the steps of data preprocessing, identification, estimation, and diagnosing and present how an appropriate time series model is developed using the steps of parameter estimation, tests for residual correlation to ensure stationarity and independence of variables using correlation functions and model forecasting accuracy using error functions and goodness-of-fit indicators.

3.3.8 Data Preprocessing

Preprocessing is an important early step in the modeling process. It provides an opportunity to identify the plausibility of analyzing particular time series data and allows for a background check of the integrity of the data. At this stage it is possible to identify and correct for missing values and outliers and deem whether or not corrective measures are necessary.

A number of missing BG values were identified in the original data set. There are a variety of methods for addressing this issue (Yaffee et al. 2000) a combination of which were used to ensure the forecasting integrity of the data was not compromised. Depending on the situation, missing values were dealt with in a variety of ways: 1) if values were missing at the beginning or end of a subject's time series these values were omitted and the time series was simply shortened by the number of missing values; 2) single missing values were assigned the same value as a previous record; and 3) a group > 1 but < 39 missing values were imputed based upon the average BG for the subject in question. Overall, there were 10 segments of missing BG values totaling 196 points from 7 subjects representing only 0.65% of the entire data set. Time series modeling requires that there not be any breaks or interruption in observations (Pankratz 1983, Box et al. 2008). By controlling for missing values and outliers, goodness of fit measures are only marginally affected and are not

considered to dramatically compromise models to the point of misinterpretation (Sanchez et al. 2003).

There are a number of ways in which outliers may be controlled for. As mentioned previously, SPSS is able to detect and manage outliers. In this methodology an automated process is employed by SPSS to detect and control for two types of typical outliers, additive (AO) and innovations (IO): 1) an additive outlier usually affects a single observation. For instance, a data coding error might be responsible for an additive outlier; and 2) an innovational outlier serves as an addition to the noise term at a specific point in the time series. For stationary series, an innovational outlier may affect several observations. For nonstationary series, it might affect every subsequent observation beginning at a specific point in the time series. Outliers may be represented at time *t* within *Q-series* as:

$$Q_t = f(t) + Z_t$$

where, $f(t)$ is an outliers function, Q_t serves as a "contaminated" influenced by outliers, and Z_t is an "uncontaminated" series free of outliers. Analogous to $v(B)$, $f(t)$ may take on a variety of forms. Box-Jenkins (1994) present a simplified form as:

$$f(t) = \frac{\omega(B)}{\delta(B)} I_t$$

where, $\frac{\omega(B)}{\delta(B)}$ is a rational distributed lag function and I_t serves as a pulse point (0,1).

For IO, $\omega(B) = \omega_0$, $\delta(B) = 1 - \delta_1$ and δ_1 is equal to ϕ_1. For AO, $\omega(B) = \omega_0$ and $\delta(B) = 1$ (Greaves et al. 2007).

3.3.9 Identification of ARIMA Parameters *(p,d,q)*

According to Chatfield (1991) a time series is considered to be *deterministic* if inherent future values are determined by a mathematical function of its previous values. *Stochastic*, or statistical, time series can be explained by some probability of distribution. A time series is considered to be *stationary* if its statistics, usually the variance and mean, remain constant over time. Conversely, nonstationary time series may contain one or several of five characteristics: outliers, random walk, drift, trend, or changing variance. The series is assessed in order to ascertain whether or not any of these phenomena exist and subsequently determine if transformation and parameter fitting of the time series is necessary. Stationarity or nonstationarity can be determined by a visual inspection of the time series ACF and PACF residual plots.

According to Box (1970b) and Bowerman et al. (1993), a time series is *nonstationary* when its autocorrelation function neither cuts off nor decreases quickly, but rather decreases slowly. Unfortunately, when discerning what is considered quick or slow, the literature fails to provide a quantifiable answer. Pedagogically, the rate of which a function decreases is dependent upon the frequency of the signal in comparison to the sampling rate. Parametric approaches presume a certain level of familiarity with the data, allowing the researcher to visually observe in a time series plot whether data might be considered stationary or nonstationary.

The characteristics presented here apply mainly to classical time series analysis. In general, physiological time series do not solely belong to any of these specified classes. The classification is dependent upon the observed time series, its length and nature of recording. For example, a time series of heart rate produced by a metronome-spaced oxygen intake test under steady conditions may appear to be stationary for some individuals who have a strong respiratory component related to heart rate; with short-term oscilliations relating to the stationary breathing frequency displaying a well-behaved signal. However, in an ambulatory recording setting or during a physically activity episode, the results are typically quite different: all components of the signal contain considerable temporal variation. Such ambulatory recordings are conducted outside of a laboratory environment and are therefore free of laboratory protocols, allowing for nonstationary changes in signal to be considered a rule as opposed to an exception. Furthermore, movements in posture for example, under such experimental conditions result in a change in muscular blood circulation and blood pressure. Subsequently, these actions cause fluctuations in heart rate. Therefore, if only heart rate time series is observed, the alterations would appear to be unpredictable and nonstationary in nature, which is why some researchers consider this type of series to be chaotic. Adding various inputs or interventions when modeling physiological time series may offer insight into their fluctuating yet non-chaotic characteristics. This is one of the objectives of the exploratory analysis undertaken in this thesis.

A series exhibiting a trend is considered to be a nonstationary time series. It is assumed that in order to model and forecast for a time series, any trend must be eliminated and the time series rendered stationary. As such, the methods for testing for stationarity and transforming time series in order to achieve stationarity are discussed next.

Stationary series exist in stable environments (Farnum et al. 1989) and are considered to be either local or globally stationary (Harvey 1991). Global stationarity is related to the entire time span of the time series whereas local stationarity only relates to a portion of the time series. Furthermore, there is strong (strict) and weak stationarity. When a time series is considered weakly stationary, stationarity exists in the autocovariance structure, the mean, and homogeneity. Alternatively put, both the variance and mean remain constant over time, and the autocovariance relies only on temporal reference points and the number of time lags between them. Weak stationarity is also referred to as stationarity in the second sense or covariance stationarity. In order for strict stationarity to be achieved another condition must be met, that is, distributions of the observations must be normally distributed; this issue is addressed next.

Two general approaches exist when testing for stationarity parametric and nonparametric. A review of the literature indicates parametric approaches tend to be used by researchers working in the time domain; for instance, economists who make certain assumptions regarding the nature of their data and require normal population distribution. Nonparametric approaches tend to be used by researchers working in the frequency domain. For instance, electrical engineers, who frequently treat the system as a "black box" and are unable to make basic assumptions regarding the nature of the system. Furthermore, nonparametric tests are not founded on the assumption or knowledge that the population is normally distributed (Bethea et al. 1991). In making no assumptions regarding the nature of the data, it is found that nonparametric tests are more broadly applicable than parametric tests as they often require normality – stationarity – in the data. Whilst more widely applicable, the resulting trade-off is such that nonparametric tests are far less powerful than parametric tests. As such, in order to arrive at the same statistical conclusion using the same confidence level, nonparametric tests demand 5% to 35% more data than parametric tests (Bethea et al. 1991). Since the data used in this theses is not stationary and is lacking normal distribution, the parametric approach is

utilized. The parametric approach corrects for abnormal distribution in order to achieve stationarity, which in turn allows for the data to be modeled. The adjustment of p,d,q parameters ensures stationarity and that residuals are not correlated with one another. As such, stationarity and independence of residuals are two assumptions of the parametric approach correlation. The next section presents how to test for stationarity.

3.3.9.1 Autocovariance and Autocorrelation functions:

Autocorrelation is used to approximate periodic behaviour of a time series. If a time series is deemed to be stationary, this means the joint probability distribution of any two subsequent random observations of a time series, say for instance, y_t and y_{t+k}, will be the same for any two time periods t and $t + k$. Additionally, if the time series is stationary, variance will not depend on time, and a random sample should indicate stationarity. Important information regarding this joint distribution, and consequently the nature of the time series, can be attained by plotting a scatter diagram of all data pairs, y_t and y_{t+k}, separated by the same interval k. The covariance between y_t and its value at another time period, say, y_{t+k} is referred to as the autocovariance at lag k, and can be defined according to Montgomery (2008) by:

$$\gamma_k = \text{Cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)].$$

As mentioned previously, before a series can be modeled it must be stationary in mean, variance, and autocovariance. A time series has a constant mean defined as:

$$\mu_y = E(y) = \int_{-\infty}^{\infty} yf(y)dy$$

A time series has a constant variance defined as:

$$\sigma_y^2 = Var(y) = \int_{-\infty}^{\infty} (y - \mu_t)^2 f(y) dy$$

The collection of values γ_k, k = 0, 1, 2, ... is referred to as the autocovariance function and the autocovariance at lag k = 0, which is just the variance of the series; that is, $\gamma_0 = \sigma_y^2$.

The corresponding autocorrelation coefficient at lag k is:

$$\rho_k = \frac{e[(y_t - \mu)(y_{t+k} - \mu)}{\sqrt{E[(y_t - \mu)^2]E(y_{t+k} - \mu)^2]}}$$

If the time series is stationary, variance will not depend on time and the correlation between (y_t - μ) and (y_{t+k} - μ) reduces to:

$$\rho_k = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} = \frac{\gamma_k}{\gamma_0}$$

where γ_0 is a calculation of the standard deviation and variance.

The collection of ρ_k , k = 0, 1, 2, ... is called the *autocorrelation function* (ACF). By definition ρ_0 = 1. Furthermore, the ACF is independent of the scale of measurement of the series and is therefore a dimensionless quantity. Lastly, ρ_k = ρ_{-k} ; indicates that the autocorrelation is *symmetric* around zero, meaning it is only necessary to calculate the positive (or negative) half.

When checking for stationarity it is necessary to estimate for the autocovariance and autocorrelation functions of a finite time series, say, y_1 , y_2 , ... , y_T. The usual approximation of the autocovariance function is:

$$c_k = \hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y}),\ k = 0, 1, 2, \ldots, K$$

and the autocorrelation function is approximated by the *sample autocorrelation function* (or sample ACF) as suggested by Box et al. (2008):

$$r_k = \frac{\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

or,

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0},\ k = 0, 1, \ldots, K$$

Autocorrelation depends on the stationarity of the series; therefore, examining a time series' ACF plot is necessary to ensure stationarity.

3.3.9.2 Partial autocorrelation function

The PACF serves as a fundamental tool of Box-Jenkins time series analysis. Used in conjunction with the ACF, both can be used to distinguish between lower order and high order AR (p) processes. The PACF works similarly to a partial correlation; where, at k lags, controls for confounding autocorrelation in intermediate lags. The underlying effect is to partial out autocorrelations and leave only the autocorrelation that exists between the current observation and kth observation. Deriving the partial correlation is helpful in order to understand its source and meaning. To illustrate consider three random variables *X, Y,* and Z Montgomery (2008) illustrates a linear regression of *X on Z and Y* on *Z* as:

$$\hat{X} = a_1 + b_1 Z \text{ where } b_1 = \frac{Cov(Z,X)}{Var(Z)}$$

and

$$\hat{Y} = a_2 + b_2 Z \text{ where } b_2 = \frac{Cov(Z,Y)}{Var(Z)}$$

Errors can then be obtained from,

$$X^* = X - \hat{X} = X - (a_1 + b_1 Z)$$

and

$$Y^* = Y - \hat{Y} = Y - (a_2 + b_2 Z)$$

The *partial correlation* between X and Y after adjusting for Z can then be defined as the correlation between X^* and Y^*; $\text{corr}(X^*, Y^*) = \text{corr}(X - \hat{X}, Y - \hat{Y})$. That is, partial correlation can be observed as the correlation that exists between two variables after they have been adjusted for a common factor that might be affecting them. Generalization is possible by permitting adjustment for more than one factor.

Using the above definition, the *partial autocorrelation function* that exists between y_t and y_{t-k} is the autocorrelation between y_t and y_{t-k} after adjusting for $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$. Therefore, for an AR($p$) parameter, PACF between y_t and y_{t-k} for $k > p$ should be equal to zero.

The more formal definition, according to Montgomery (2008), is as follows. Considering a time series model (y_t) that is not an AR process. Further consider, for any fixed value of k, the ACF of an AR(p) process is given as:

$$\rho(j) = \sum_{i=1}^{k} \phi_{ik} \rho(j-1), j = 1, 2, \ldots, k$$

or

$$\rho(1) = \phi_{1k} + \phi_{2k}\rho(1) + \ldots + \phi_{kk}\rho(k-1)$$
$$\rho(2) = \phi_{1k}\rho(1) + \phi_{2k}\rho + \ldots + \phi_{kk}\rho(k-2)$$
$$\text{M}$$
$$\rho(k) = \phi_{1k}\rho(k-) + \phi_{2k}\rho(k-2) + \ldots + \phi_{kk}$$

These equations written in matrix notation are as follows:

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \text{K} & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \text{K} & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \text{K} & \rho(k-3) \\ \text{M} & \text{M} & \text{M} & \text{O} & \text{M} \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \text{K} & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \text{M} \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \text{M} \\ \rho(k) \end{bmatrix}$$

or

$$\text{P}_k \phi_k = \rho_k$$

where

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \text{K} & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \text{K} & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \text{K} & \rho(k-3) \\ \text{M} & \text{M} & \text{M} & \text{O} & \text{M} \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \text{K} & 1 \end{bmatrix}, \phi_k = \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \text{M} \\ \phi_{kk} \end{bmatrix}, \text{and } \rho_k = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \text{M} \\ \rho(k) \end{bmatrix}$$

To solve for ϕ_k, the equation is:

$$\phi_k = \mathrm{P}_k^{-1}\rho_k$$

For any given k, $k = 1, 2, \ldots$, the final coefficient ϕ_{kk} is referred to as the partial correlation at lag k. For an AR(p) process $\phi_{kk} = 0$ for $k > p$. Thus, when viewing a PACF residual, it is possible to identify when the PACF cuts off at a particular lag, say lag p, for an AR (p). This suggests that the PACF may be used to identify the order of an AR process in a similar manner in which the ACF may be used to indentify an MA process.

For sample computations, $\hat{\phi}_{kk}$, the sample approximation of ϕ_{kk}, is attained by using the sample ACF, $r(k)$. Moreover, in a sample of N observations in the case of an AR(p) process, $\hat{\phi}_{kk}$, for $k > p$ is approximately normally distributed by:

$$E\left(\hat{\phi}_{kk}\right) \approx 0 \text{ and } Var(\hat{\phi}_{kk}) \approx \frac{1}{N}$$

For this reason the 95% limits to judge as to whether $\hat{\phi}_{kk}$, is statistically significantly distinctive from zero are given by $\pm 2/\sqrt{N}$. For a more detailed review the reader can refer to Quenouille (1949), Jenkins (1954, 1956), and Daniels (1956).

3.3.9.3 Cross-Correlation Function:

The linear transfer function approach assumes that the transfer functions and noise model are independent of one another. As such, an additional transfer function metadiagnosis is required; that is, examining the cross-correlation

residuals of the transfer function and noise model residuals will validate this assumption. If the assumption holds, the residuals should appear as white noise.

The cross correlation coefficient at lag k is estimated by $r_{xy}(k) = \frac{C_{xy}(k)}{S_x S_y}$

where

$$C_{xy}(k) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(y_{t+k} - \bar{y}), k = 0,1,2,... \\ \frac{1}{n}\sum_{t=1}^{n+k}(y_t - \bar{y})(x_{t-k} - \bar{x}), k = -1,-2,... \end{cases}$$

$$s_x = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

$$s_y = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

The cross correlation function is not symmetric about k = 0. Approximate standard error of $r_{xy}(k)$ is

$$se(r_{xy}(k)) \cong \sqrt{\frac{1}{n-|k|}}, \; k = 0, \pm 1, \pm 2, ...$$

The standard error is also based on the assumption that the series are not cross correlated and one of the series is white noise. (The general formula for the standard error can be found in Box et al. (2008)).

3.3.9.4 Transformations

In the case that adjusting AR(p) and MA(q) parameters are unsuccessful in detrending and stabilizing the variance of the time series for the purpose of

stationarity, a data transformation is required. Much of this next section discusses how transformations are able to adjust for stationarity and is derived from Pankratz (1991).

When dealing with nonconstant variance, the *power family* transformation is employed, given by:

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda \dot{y}^{\lambda-1}}, & \lambda \neq 0 \\ \dot{y} \ln y, & \lambda = 0 \end{cases}$$

where $\dot{y} = \exp[(1/T)\sum_{t=1}^{T} \ln y_t]$ is the geometric mean of the observations. If $\lambda = 1$ there is no transformation. Typical values of λ used with times series data are $\lambda = 0.5$ (a square root transformation), $\lambda = 0$ (the log transformation), $\lambda = -0.5$ (reciprocal transformation), and $\lambda = -1$ (inverse transformation). The divisor $\dot{y}^{\lambda-1}$ is simply a scale factor that ensures that when different models are fit to investigate the utility of different transformations (values of λ), the residual sum of squares for these models can be meaningfully compared. The reason that $\lambda = 0$ implies a log transformation is that $(\lambda^{\lambda} - 1)/\lambda$ approaches the log of y as λ approaches zero. Often an appropriate value of λ is chosen empirically by fitting a model to $y^{(\lambda)}$ for various values of λ and then selecting the transformation that produces the minimal root mean square error.

3.3.9.5 Log Transformation

When standard deviation of the original time series increases in a linear fashion with the series mean, log transformation is the optimal variance-stabilizing transformation. This transformation does well to physically interpret

the variance percentage change and can be shown, according to Montgomery (2008), in a time series with y_1 , y_2 , ... , y_t where we are interested in the percentage change in y_T, say,

$$x_t = \frac{100(y_t - y_{t-1})}{y_{t-1}}$$

The approximate percentage change in y_t can be calculated from the differences of the log-transformed series $x_t \cong 100[\ln(y_t) - \ln(y_{t-1})]$ because:

$$100[\ln(y_t) - \ln(y_{t-1})] = 100\ln\left(\frac{y_t}{y_{t-1}}\right) = 100\ln\left(\frac{y_{t-1(y_t - y_{t-1})}}{y_{t-1}}\right)$$

$$= 100\ln\left(1 + \frac{x_t}{100}\right) \cong x_t$$

3.3.9.6 Differencing

Alternatively, researchers can detrend a time series through differencing, which involves setting the I (*d*) term as a preprocessing transformation step to control for autocorrelation and achieve stationarity. This is accomplished by subtracting each datum in a time series from its predecessor. That involves applying the difference operator to the original time series in order to obtain a new time series, say:

$$x_t = y_t - y_{t-1} = \nabla y_t$$

where, ∇ is the (backward) difference operator used when the time series is not stationary. An alternative way to write the differencing operator and make the equation more compact is to use the backshift operator where B is defined as ($By_t = y_{t-1}$). So,

$$x_t = (1-B)y_t = \nabla^d y_t = y_t - y_{t-1}$$

with $\nabla = (1\text{-}B)$. Difference can then be performed successively if required until a trend is removed allowing for stationarity. The second difference is therefore:

$$x_t = \nabla^2 y_t = \nabla(\nabla y_t) = (1-B)^2 y_t = (1-2B+B^2) = y_t - 2y_{t-1} + y_{t-2}$$

Generally, powers of the backward difference operator and the backshift operator are defined as:

$B^d y_t = y_{t-d}$

$\nabla^d = (1 - B)^d$

Differencing directs autocorrelation toward 0 or beyond in a negative direction. When single differencing results in an autocorrelation spike in an ACF residual plot >.5, *over-differencing* has occurred. Over-differencing may also be indicated in reviewing a residual plot where a change in sign (+ to - and vice versa) exists from one observation to the next. Optimal level of differencing will result in the lowest standard deviation - or RMSE value, as is the case for this thesis. In some cases, adding moving average (q) terms to the time series model will compensate for moderate over-differencing. Also, with each increase in term of differencing results in the time series being shortened by one.

Two advantages are associated with differencing in relation to possibly overfitting of AR and MA terms in efforts to detrend and achieve stationarity. First, it does not require an estimation parameter, thus providing a more *parsimonious* (e.g. simpler) approach. Secondly, fitting for AR and MA terms presume that a trend is constant over the duration of the time series history and will also be present in the future. Alternatively, the trend component, once estimated, is presumed to be *deterministic*; whereas, differencing allows for the

alteration of a trend over the time series. First differencing accounts for trends influencing a change in mean, and second order differencing controls for a change in the overall slope of a time series. Furthermore, if differencing is applied to the dependent variable, Pankratz (1991) shows that all corresponding explanatory (independent/predictor) variables should also be differenced.

Lastly, when differencing is applied the constant is excluded as it becomes irrelevant; when $d > 0$, differencing renders the constant unnecessary, ultimately removing it. When this is the case, the software automatically excludes the constant from modeling process. A number of models in this thesis are differenced and/or transformed, the extent of which are demonstrated further in the *Results* section.

3.3.10 Estimation

Parameter estimates are derived from two possible algorithms – the details of which are not discussed at length here as they are beyond the scope of this work, a more detailed review can be found in Yaffee (2000). Results are compared for optimal goodness-of-fit. Ultimately, parameter estimates should be of reasonable magnitude, and statistically significant with t-ratios > 1.96. Non-significant parameters are trimmed from the model. The general steps for the estimation of parameters are as follows and according to SPSS's *Time Series Algorithm Manual* (2010):

1) Computation of noise process N_t through the historical period.

2) Forecasting the noise process N_t up to the forecast horizon. This is a one step ahead forecasting during the historical period and multi-step ahead forecasting after that.

3) Final forecasts are obtained by first adding back to the noise forecasts the contributions of the constant term and the transfer function inputs and then integrating and back-transforming the result. The prediction variances of noise forecasts also may have to be processed to obtain the final prediction variances.

Let $\hat{N}(k)$ and $\sigma_t^2(k)$ be the k-step forecast and forecast variance respectively.

3.3.10.1 Conditional Least Squares

$$\hat{N}_t(k) = E(N_{t+k} \mid N_t, N_{t-1}, \ldots)$$

assuming $N_t = 0$ for $t < 0$.

$$\sigma_t^2(k) = \sigma^2 \sum_{j=0}^{k-1} \psi_j^2$$

where ψ_j are coefficients of the power series expansion of *MA/(Δ x AR).*

Minimize $S = \sum (N_t - \hat{N}_t(1))^2$.

3.3.10.2 Maximum Likelihood (ML) method (Brockwell et al. 1991)

$$\hat{N}_t = (k) = E(N_{t+k} \mid N_t, N_{t-1}, \ldots, N_1)$$

Maximize likelihood of $\{N_t - \hat{N}_t(1)\}_{t=1}^n$; that is,

$$L = -\ln(s/n) - (1/n) \sum_{j=1}^{n} \ln(n_j)$$

where $s = \sum (n_t - \hat{N}_t(1))^2 / n_t$, and $\sigma_t^2 = \sigma^2 n_t$ is the one-step ahead forecast variance.

3.3.10.3 Error Variance

$$\hat{\sigma}^2 = S/(n-k)$$

serves as the error variance in both methods. Here n is the number of non-zero residuals and k is the number of parameters (excluding error variance).

3.3.11 Diagnostic Statistics and Error Functions

Error functions are employed to measure the difference between actual and model-generated data and serve as goodness-of-fit indicators. When time series modeling, the error function is commonly referred to as the *objective function*, or the function we wish to minimize. For this reason, modeling is based on observed input data that is put into the model and target forecast data. The model-produced output can then be compared to target values allowing a measure of distance to be calculated. The error functions that follow are derived from Yaffee (2000) were used to determine the best fitting models based on time interval (5, 15 or 30 minute) and predictive independent variables.

3.3.11.1 Sum of Square Errors Mean Square Errors

While not used explicitly for measuring goodness-of-fit or forecast accuracy in this thesis, these measures are used to calculate other substantiating explanatory forecast accuracy measures: root mean square error and stationary r-squared. Sum of square errors (SSE) and mean square errors (MSE) may be used however, to attain a sense of dispersion error. As in the previous section, instead of using the absolute value of error to avoid cancelling of error resulting from adding positive and negative errors, it is possible to square the errors of each observation, add these squares for the entire forecast or time series in order to obtain the SSE:

$$\sum_{t=1}^{T} e_t^2$$

After this value is obtained the SSE can be divided by its degrees of freedom, resulting in the error variance or MSE:

$$\sum_{t=1}^{T} \frac{e_t^2}{T-k}$$

3.3.11.2 Root Mean Square Error

Calculating the square root of the MSE results in the standard deviation of error; also known as the root mean square error (RMSE):

$$\sqrt{\sum_{t=1}^{T} \frac{e^2}{T-k}}$$

Typically, a model with a lower RMSE indicates a good fit. The RMSE serves as an indicator of the difference between predicted and actual values. Alternatively, RMSE measures how much a dependent series differences from its model-predicted level.

3.3.11.3 R-Squared

Error variance must be accurately estimated or else other important aspects of the regression analysis would not be accurate. Selecting a model that maximizes the R^2 is the same as selecting the model that minimizes the sum of the root mean square error. Large values of the R^2 indicate a good fit with the historical data. However, since the root mean square error always decreases as more parameters are added to the model, relying solely on the R^2 value to select a forecasting model of best-fit supports applying more parameters than are necessary to obtain a good forecast (referred to as over-fitting).

$$R^2 = 1 - \frac{\sum_{t=1}^{T} e_t}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

3.3.11.4 Stationary R- Squared

Similar to the R-squared statistic, the stationary R-Squared uses a size adjustment component, where the number statistic adjusts according to the number of parameters used in the model and serves as a better indicator for goodness of fit when assessing multiple inputs.

$$SR^2 = \left(1 - \frac{k}{T}\right) r^2$$

where T = the number of observations and k = the number of parameters. If parameters are accurately accounted for in the model, residuals should consist of purely white noise or alternatively unsystematic random variation. Furthermore, the model that maximizes the stationary R^2 statistic also minimizes the RMSE. Stationary R-squared is an estimate of the proportion of total variation in the series explained by the model and is therefore preferable to ordinary R-squared when a trend exists. Stationary R-Squared is used in this thesis to select a forecasting model of best-fit.

3.3.11.5 Portmanteau test for autocorrealtion: Ljung-Box Statistic

The Ljung-Box statistic (Ljung et al. 1978), a time series Portmanteau test, and their p-values for lag k indicate whether or not autocorrelation exists up to the specified order of k. This is done in addition to examining ACF and PACF residuals for spikes that may indicate an erroneously inflated Q statistic.

$$Q(K) = n(n+2)\sum_{k=1}^{K} r_k^2 /(n-k)$$

where r_k is the kth lag ACF of residual.

Q(K) is approximately distributed as $x^2(K - m)$, where m is the number of parameters other than the constant term and predictor related-parameters. Ideal values for this statistic are < .05.

3.3.12 SPSS time Series Algorithm

Figure 7 (SPSS 2004) demonstrates the SPSS software algorithm employed in the Time Series Expert Model Builder function. It consolidates and simplifies a number of the aforementioned processes into a few steps. Outputs include: model summary, model statistics and ARIMA model parameters. The *(P,D,Q)* parameters, different from the *(p,d,q)* parameters, are an additional set of parameters used when a time series is considered to be seasonal – an explanation of which is beyond the scope of this thesis. Such is typically the case when evaluating and forecasting for economic time-series, the physiological time series data used in this thesis do not exhibit seasonal trends, thus the *(P,D,Q)* parameters are not applicable. Despite being included in the algorithm model in Figure 7, the software automatically omits these parameters from analysis as they only apply to seasonal time series.

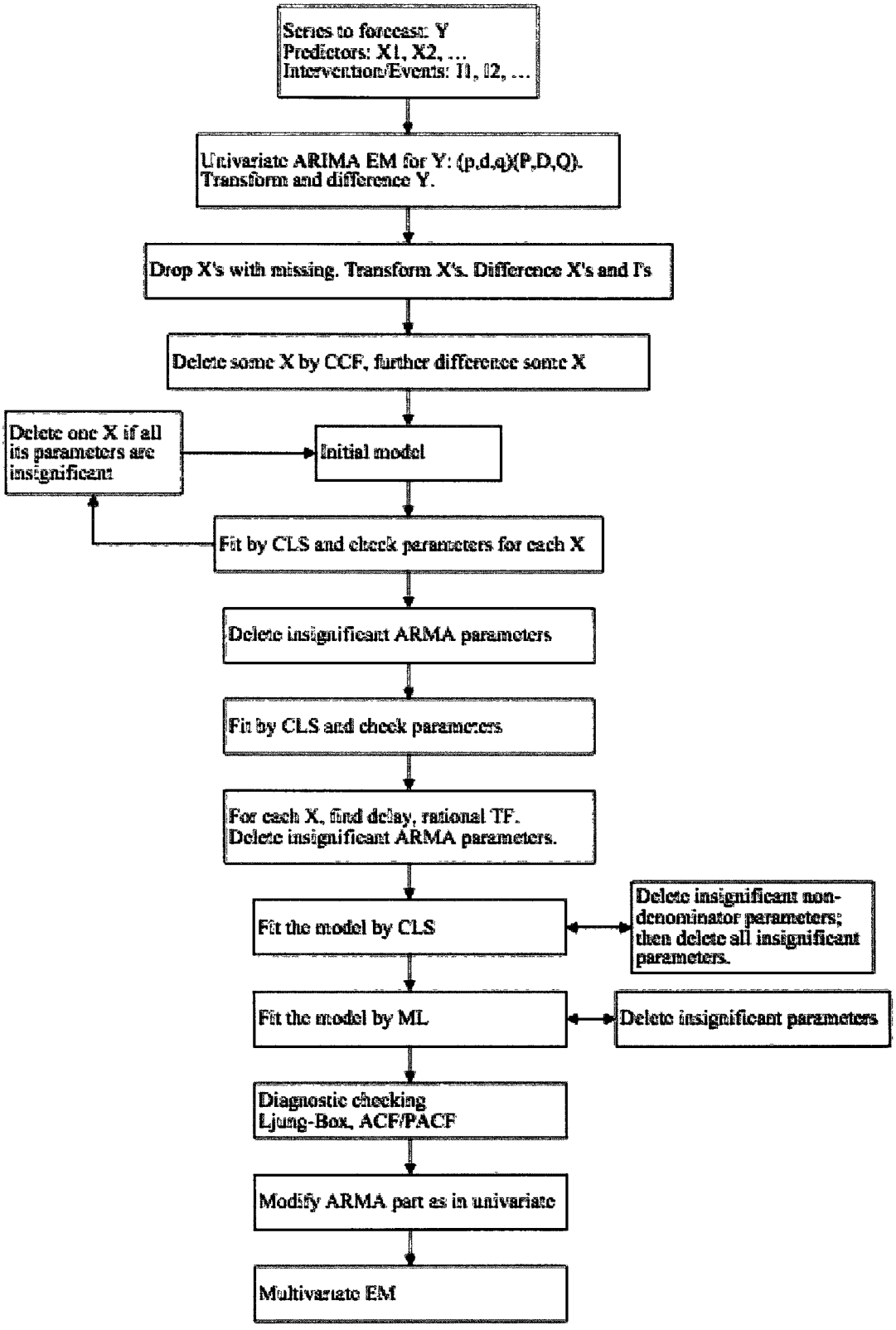


Figure 7 Software time series algorithm model concept (SPSS, 2004)

3.3.13 Summary and Remarks

The standard procedures of identification, estimation and diagnosis developed by Box and Jenkins can be summarized as follows:

1) ACF and PACF of the time series (including dependent and independent variables) are used to identify initial AR (p) and MA (q) parameters;

2) New parameters are estimated until stationarity in the variance and mean is achieved and;

3) Diagnostic statistics and error functions are checked for goodness-of-fit.

The basic idea of time series models is that they are dynamic regression models wherein explanatory variables are functions of time with coefficients that change over time. They are created by the past available observations of the series itself.

3.4 Disclaimer

It is important to note that at the onset of the results section, the exploratory nature of these results, and to remind the reader that they are based on a small pilot study from a single rehabilitation facility. It is also important to highlight that the results were prepared by a non-medical professional; thus, any discussion of medical implications and related conclusions should be limited and interpreted with caution. To further offset the non-medical knowledge of the researcher, results were shared with three medical professionals at Toronto Rehab in an attempt to gain insight and suggestions for some of the unusual findings presented.

That said, to the knowledge of the author, these data represents the only such data of its kind to date. And whilst the sample of patients is small, the number of observations was extensive: greater than 800 measurements of BG and predictor variables per subject (or >22,000 total observations). This would seem to warrant at least some exploratory analysis to examine if any patient-by-patient variability exists (even if not representative of all diabetic patients), to test the possible significance of typical and new explanatory variables, to test the usefulness of the analysis technique deployed, and to at least provide a means to assess the potential usefulness of larger samples in the future; all in support of the objectives of this thesis.

Chapter 4: Results

4 Results

In this chapter, the main results of subject-by-subject time series modeling analysis are presented. The response of each subject's BG (dependent variable) to possible predictive inputs (independent variables) is assessed using the Linear Transfer Function Modeling Method described in the previous chapter. As a background to this, subject profiles and related descriptive statistics are first presented.

4.1 Subject Profiles and Descriptive Statistics

Overall, 37 of the 40 subjects in the original study were monitored for the full 72 hour study period. The manner in which subjects are labeled is by no means an indication of the total number of subjects who participated in the pilot study (e.g. Subject 196 does not mean there were 196 subjects in the study). The reason why this labeling method was used is unknown to the author. Two subjects ceased monitoring after day 2 (providing 1.5 days monitoring), and one after day 3 (providing 2.5 days of monitoring), all reporting irritation with the tape associated with the blood glucose monitor as the reason. Additionally, four of the first eight subjects GPS data were deemed unusable – owing to a technical problem. Closer examination of the data prior to this thesis revealed a further 6 subjects whose data contained missing values that prevented time series modeling. Four of these had missing food diary data, one had missing heart-rate and accelerometer data, and one had missing activity diary data. This left a total of 27 subjects available for analysis in this thesis.

Table 1 presents basic information of these 27 subjects, including age, gender (1 = male, 2 = female), weight, waist size, type of DM, year of DM

diagnosis, HBA1C, total number of BG readings, average BG value and the standard deviation of BG. Participants were split evenly by gender, ages ranged between 32 and 74 with a mean age of 56. Participants weighed between 45 and 147 kilograms, with a mean weight of 86.5 kilograms. Average BG for each subject over the course of the study ranged from 5.1 – 13.6 mmol/L. The overall average across all subjects being 7.3 mmol/L, indicating a wide variation, with some subjects demonstrating extremely high BG and other near normal (non-diabetic) BG. The standard deviation of the mean BG ranged from 0.71 – 5.28 with a mean of 2.01, indicating that some subjects have little BG variation while others BG seem to be far less stable.

Table 2 displays the drug type, action and specific name of insulin medication used by each subject. Three types of insulin medication were used amongst subjects: insulin, an insulin sensitizer, and an insulin releaser – insulin sensitizers allow the body to respond more normally to insulin secretion and insulin releasers stimulate endogenous release of insulin. Specific types of insulin medication used by subjects included: Actos, Diabeta Diamicron, Humalin, Humalog, Lantus, Levemir, Novarapid, and Metaformin; each having a different response time varying between slow, medium and rapid. Subjects may have used one, or a combination of different medications.

User_ID	Hours of continuous BG Data	Age	Gender	Weight (kg)	Waist (cm)	Diabetes Type	Diagnosis year	HBAIC	Blood Glucose (mmol/L), ORIGINAL		
									Count	Mean	SD
61	69.6	64	1	123.1	123	2	.	8.1	835	9.41	2.12
66	71.8	45	2	106.5	124	2	1993	7.1	861	11.43	3.5
70	101 3	49	2	55	71	2	1992	6.4	1216	7 28	1.25
73	71.8	67	2	59 1	88	2	1987	6.4	861	8.66	2 4
74	71 8	64	1	91.8	100	1	1987	7 8	861	13.59	5.28
77	72.2	69	1	94.5	113	2	1996	8.1	866	6.74	2.27
80	70 8	44	2	59	74	1	1986	7 2	849	7.31	4 7
81	68 6	56	1	104 5	117	2	2006	6.5	823	8.75	1 53
84	71 7	49	2	67 3	94	2	2002	6.2	860	6 12	1.53
86	73 3	42	1	94.8	112	2	2006	7.9	879	5.13	1.41
87	73.2	52	2	66.8	88	2	2006	5.8	878	7.38	1.81
88	73 0	59	2	53 6	75	2	2006	10	876	6.47	1 6
90	63 8	60	1	99.1	107	2	2006		765	6 63	1.54
91	70.3	69	2	67.7	92 5	2	2006	5 1	843	6 74	1.74
92	67 4	64	2	82 3	93	2	1998	6.8	809	6.45	1.59
93	72 1	35	2	84.5	97 5	2	1985	5.6	865	5.78	0.75
153	63.3	50	2	129 5	125	Pre	2006	6 6	760	6.65	1.1
163	70 1	55	2	85.9	106	2	2006		841	7.52	1.54
167	71.3	50	2	127 3	117	2	2006	5.3	856	5.36	0 71
170	49.8	46	1	77 7	93	2	2006	.	597	6.41	1.41
171	67.7	66	2	45	83	2	1993	7.7	812	6.48	2.86
172	68.3	68	1	86.4	115	2	1993	7.2	819	6 91	2.87
173	72 0	48	1	93 6	105	2	.	7.6	864	7.15	1.41
175	75.2	32	1	71	91	1		.	902	6.96	3.57
176	71.5	68	1	90	113	2	.	.	858	5 78	0 86
177	72 5	51	2	147		2		.	870	7.61	2.4
196	70 0	65	1	90 5	109.5	2	1995		840	6.12	2.07

Table 1 Subject Profile

User_ID	Drug Type (Action) and Specific Name
61	INSULIN RELEASER (Slow) - Diamicron
	INSULIN SENSITIZER (Medium) - Metformin
66	INSULIN (Rapid) - NovaRapid
	INSULIN (Slow) - Levemir
	INSULIN SENSITIZER (Medium) - Metformin
70	INSULIN SENSITIZER (Medium) - Metformin
73	INSULIN SENSITIZER (Medium) - Metformin
	INSULIN SENSITIZER (Slow) - Actos
74	INSULIN (Rapid) - Humalin
	INSULIN (Rapid) - Humalog
77	INSULIN (Rapid) - NovaRapid
	INSULIN (Slow) - Lantus
80	INSULIN (Medium) - Humalin
	INSULIN (Rapid) - NovaRapid
81	INSULIN SENSITIZER (Medium) - Metformin
84	INSULIN SENSITIZER (Medium) - Metformin
90	INSULIN SENSITIZER (Medium) - Metformin
92	INSULIN SENSITIZER (Medium) - Metformin
93	INSULIN SENSITIZER (Medium) - Metformin
153	INSULIN SENSITIZER (Medium) - Metformin
167	INSULIN SENSITIZER (Medium) - Metformin
171	INSULIN RELEASER (Rapid) - Diabeta
	INSULIN SENSITIZER (Medium) - Metformin
172	INSULIN RELEASER (Rapid) - Diabeta
	INSULIN SENSITIZER (Medium) - Metformin
	INSULIN SENSITIZER (Slow) - Actos
173	INSULIN SENSITIZER (Medium) - Metformin
175	INSULIN (Rapid) - NovaRapid
	INSULIN (Slow) - Lantus
176	INSULIN SENSITIZER (Medium) - Metformin
178	INSULIN SENSITIZER (Medium) - Metformin
	INSULIN SENSITIZER (Slow) - Actos
196	INSULIN SENSITIZER (Medium) - Metformin

Table 2 Subject insulin medication

Table 3 provides a summary of total subject food and activity data over the course of the 72-hour monitoring period. Carbohydrate consumption ranged from 330-1854 grams with a study average of 642 grams. Sugar consumption ranged from 46-335 grams with a study average of 187 grams. Total calorie consumption ranged from 2476-14055 with a study average of 5424. Activity

diary data also varied widely amongst subjects with the most time spent at home followed by work/school. Total exercise varied between 0-6.8 hours with an average of 2.3 hours. Notice also the number of subjects who did not exercise or exercised very little. Lastly, sugar consumption for subject 84 appears to be unusually high and could be the result of a coding error.

User_ID	Carbohydrates consumed (Total, grams)	Sugar consumed (Total grams)	Calories consumed (Total)	Insulin taken in last 5 minutes (yes/no)		Total hours in diary	Activity time (hours), all types)	Trip time (hours, (all types)	Exercise time (hours) (all types)	Time spent with people (hours)	Activity-At home time (hours)	Activity-Work/school time (hours)	Activity-Shopping time (hours)	Trip-Automobile time (hours)
				No	Yes									
61	628 27	240 94	4669 94	778	85	75 8	73 6	2 2	4 4	2 2	62 0	0 0	1 2	2 2
66	731 46	202 46	5474 17	839	22	76 4	69 9	6 5	2 9	34 1	55 5	0 0	0 4	2 2
70	505 95	115 18	4967 42	1184	32	95 1	93 1	2 0	4 4	20 0	46 2	17 4	3 5	1 9
73	450 77	113 3	2476 24	853	8	101 4	99 6	1 9	0 0	9 3	61 2	0 0	1 9	1 9
74	856 7	335 53	8403 97	838	23	57 1	47 4	9 7	4 3	5 2	30 1	10 9	0 4	2 8
77	454 68	46 74	5073 23	859	7	75 6	72 8	2 8	2 0	0 3	55 9	7 8	0 3	2 6
80	576 48	200 63	5126 45	829	20	72 5	70 4	2 1	0 8	8 5	60 2	0 0	2 2	2 1
81	439 75	133 46	3363 96	819	4	72 2	64 3	7 9	2 1	5 3	38 4	19 5	0 1	5 8
84	841 9	0	6868 4	852	8	63 1	59 8	3 3	4 8	6 1	33 0	11 5	0 1	2 1
86	602 13	190 64	5175 43	879	0	84 2	77 6	6 6	6 8	7 3	43 0	5 0	1 3	3 2
87	453 75	177 46	3995 54	878	0	101 1	96 1	5 0	3 9	5 3	38 2	27 2	0 0	0 9
88	431 69	137 22	5691 14	876	0	102 1	97 3	4 8	5 5	5 6	47 8	12 0	0 2	0 7
90	450 34	120 84	3247 32	748	17	65 2	61 9	3 3	0 0	1 7	35 9	22 7	0 2	3 0
91	364 62	128 66	2534 98	843	0	73 6	72 1	1 5	0 0	2 5	59 0	0 0	0 4	1 4
92	687 34	235 08	4949 45	772	37	63 7	61 6	2 1	0 3	1 8	58 4	0 0	2 4	2 1
93	875 19	315 09	5775 34	859	6	70 4	64 3	6 1	2 0	7 9	34 3	24 5	0 5	0 3
153	734 52	193 53	5542 43	752	8	64 1	58 9	5 2	2 2	6 3	30 7	17 9	3 4	0 0
163	747 03	220 03	7942 67	841	0	73 2	67 6	5 6	1 4	13 0	49 1	4 6	2 0	4 2
167	577 21	137 02	7298 01	846	10	61 0	59 3	1 6	2 0	1 0	46 8	0 0	1 5	1 1
170	330 94	63 87	3350 63	597	0	90 0	89 4	0 6	0 1	1 0	45 9	0 0	0 0	0 0
171	464 03	148 5	4331 78	779	33	74 8	72 0	2 8	0 7	10 0	54 5	0 0	0 0	2 8
172	656 71	197 75	5943 99	798	21	71 6	65 5	6 1	1 4	9 7	44 7	4 6	0 1	5 8
173	741 6	139 09	5374 27	860	4	72 9	65 3	7 6	2 8	25 4	28 6	13 1	3 0	7 2
175	1854 08	616 6	14055 1	878	24	96 1	88 6	7 5	3 3	2 1	45 5	3 4	5 1	5 3
176	476 05	144 78	3626 29	853	5	84 1	80 6	3 5	2 2	11 7	55 7	0 0	2 2	1 0
177	597 54	195 19	5359 09	870	0	71 8	66 8	5 0	0 0	7 9	62 6	0 0	2 0	5 0
178	502 25	166 98	4066 16	859	11	67 8	66 6	1 3	0 0	0 0	30 7	10 5	0 5	1 3
196	963 7	264 52	7197 49	817	23	68 9	63 0	6 0	4 1	3 2	57 8	0 0	1 6	3 8

Table 3 Subject summary data

4.2 Time interval and independent variable selection

An extensive number of ARIMA time series models were estimated on a patient-by-patient basis using possible predicting independent variables, and three different time averaging intervals: 5, 15, and 30-minute. Several 15 and 30-minute models were considered adequate, however, after reviewing model statistics and goodness-of-fit measures, 5-minute interval models performed best overall while testing for the predictability of all the independent variables presented in Table 5. Overall, the 5-minute interval models suggested the 'optimal' averaging period for all subjects based on having produced: 1) the most significant BG predicting variables; and 2) the lowest standard deviation (RMSE) of dependent series differences from its model-predicted level. A summary of the different time interval models' performance is shown in Table 4 (n = 28).

	5 min	15 min	30 min
PREDICTORS	2.17	1.25	.792
RMSE	.181	.509	.609

Table 4 Model Comparison

Table 4 demonstrates the average decrease by subject in the number of BG predicting variables when modeling with larger time intervals. Following this trend is a decrease in the ability to accurately predict future values of BG indicated by larger RMSE values at larger time intervals.

In this study, a *multivariate* time-series model is employed to examine the effect of a wide range of independent variable inputs, as well as their previous disturbances, on BG, the dependent variable. To do so, a TF using the independent/explanatory variables presented in Table 5 is employed. All variables are measured with respect to their values over the course of 5 minutes and the preceding measurement of BG. For example, "Calories_5m" indicates the

number of calories consumed in the 5 minutes period prior to the given BG measurement; "Event23_time_5m" indicates how many of the previous 5 minutes were spent shopping. Vmag is a measurement of the intensity of physical activity and is calculated by taking the square root of the sum of accelaration of the x, y and z-axes.

Given past research and expectations, variables reflecting food intake, insulin medication and physical activity are included in the analysis. Additionally, this thesis incorporates several unique variables related to activity and travel engagement, social engagement, and distance from home. A number of variables were considered for model input, and include:

- Distance home (kms)
- Average speed last 5 minutes
- Average heart rate last 5 minutes
- Carbohydrates consumed in last 5 minutes (grams)
- Sugar consumed in last 5 minutes (grams)
- Calories consumed in last 5 minutes
- Medicine consumed in last 5 minutes (yes/no)
- Insulin Medication taken in last 5 minutes (yes/no)
- Time spent with people in last 5 minutes (seconds)
- Activity time in last 5 minutes (all types)
- Trip time in last 5 minutes (all types)
- Exercise time in last 5 mins (all types)
- Activity-At_home time in last 5 mins
- Activity-Indoor (Unspecified) time in last 5 mins
- Activity-Exercise-Aerobic/Intense workout time in last 5 mins
- Activity-Exercise-Low intensity workout time in last 5 mins
- Activity-Exercise-Weight training time in last 5 mins
- Activity-Exercise-Walking around time in last 5 mins
- Activity-Exercise-Jogging around time in last 5 mins
- Activity-Exercise-Other time in last 5 mins
- Activity-Work/school time in last 5 mins
- Activity-Shopping time in last 5 mins
- Activity-Service/care time in last 5 mins
- Activity-Social time in last 5 mins
- Activity-Leisure/recreation time in last 5 mins
- Activity-Restaurants time in last 5 mins
- Activity-Short event stop time in last 5 mins
- Trip-Automobile time in last 5 mins

- Trip-Public Transit (bus, streetcar, subway, train) time in last 5 mins
- Trip-Exercise-Walk time in last 5 mins
- Trip-Public Transit-Train time in last 5 mins
- Trip-Public Transit-Intercity Transit time in last 5 mins
- Trip-Public Transit-Subway time in last 5 mins
- Trip-Public Transit-Bus time in last 5 mins

These variables may be broadly categorized as: nutrition intake, medication intake, physical activity, social activity, and travel activity. The notion behind testing such a wide range of variables serves as an effort to identify episodes, events or activities that may be stressful. Research suggests that stress may affect an individual's metabolism and subsequently their BG level (Hennessy et al. 1999). The activity specific variables are a direct attempt at enhancing the understanding of how the spatial/temporal structuring of daily activities potentially influence health and find their genesis in behavioural geography theories. Preliminary analysis however indicated a limited number of these variables found to be predictors, those of which were included in final models and can be found in Table 5.

Input Variables	Description	Type	Value
sugar_5m	sugar consumed	continuous	grams
carbs_5m	carbohydrates consumed	continuous	grams
Calories_5m	calories consumed	continuous	grams
med_ins_5m	insulin medication taken	dichotomous	yes/no
avgavgVmag_5m	square root of the sum of acceleration	continuous	vector magnitude
event21x_time_5m	time spent involved in physical activity	continuous	minutes
event12_time_5m	time traveling in an automobile	continuous	minutes
event30x_time_5m	time spent commuting via public transit	continuous	minutes
Distance_home	distance to home	continuous	km
event27_time_5m	time spent at a restaurant	continuous	minutes
event23_time_5m	time spent shopping	continuous	minutes
ppl_time_5m	time spent with people	continuous	minutes

Table 5 Independent Variables in previous 5 minutes

4.3 Example Full Model output for Subject 84

This section provides an example of the full model output for one of the 27 subjects as produced by the SPSS software. A key summary of all subjects models is presented in the next section.

Table 6 displays the model description for subject 84. The reader may recall from the *Methods* section that ARIMA models have three modeling parameters: 1) the *Autoregressive* or AR (p) process which takes into consideration historical data serving as a memory of past events; 2) an *Integrated* or I (d) process that stabilizes the time series by making the data stationary and assists in making the forecasting process easier; and 3) a *Moving Average* or MA (q) process which accounts for forecast errors and controls for autocorrelation. More specifically from Table 6, the model parameters (p,d,q) for the dependent variable BG are presented. AR(2) indicates the BG value of the current time period is regressed upon the previous two values of itself plus some random error. The differencing and moving average, D(2) and MA(9) indicate that a differencing order of 2 and moving average involving 9 lags were required to achieve stationarity and eliminate autocorrelation between residuals.

Model Description

			Model Type
Model ID	Blood Glucose (mmol/L; with Imputed values)	Model_1	ARIMA(2,2,9)

Table 6 Model description for subject 84

Model Statistics

Model	Number of Predictors	Model Fit statistics		Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	Statistics	DF	Sig.	

Model Statistics

Model	Number of Predictors	Model Fit statistics		Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	Statistics	DF	Sig.	
Blood Glucose (mmol/L; with Imputed values)	1	.389	.092	25.868	13	.018	5

Table 7 Model statistics for subject 84

ARIMA Model Parameters

				Estimate	SE	t	Sig.
Blood Glucose (mmol/L)	No Transformation	AR	Lag 1	-.505	.034	-14.7	.000
			Lag 2	-.245	.061	-4.0	.000
		Difference		2			
		MA	Lag 2	-.223	.068	-3.3	.001
			Lag 7	.131	.035	3.8	.000
			Lag 9	.087	.035	2.5	.014
Average Vmag for time period	No Transformation	Delay		3			
		Numerator	Lag 0	-.124	.036	-3.4	.001
		Difference		2			

Table 8 ARIMA Model Parameters for subject 84

Table 7 indicates one (1) BG predicting variable (Vmag, which the software produces in a separate ARIMA Model Parameters table) the software found to be significant. The stationary R^2 value serves as a criterion when comparing other competing models and selecting a forecasting model of best-fit; the value presented here translates to the model being able to explain about 39% of the observed variation in the time series. The RMSE serves as an indicator of the difference between predicted and actual values. Alternatively, RMSE measures how much a dependent series differences from its model-predicted level. The lower the RMSE the greater the predicting ability of the model; .092 serves as indication that the model does a good job at predicting

future values of BG. Ljung-Box statistics provides an indication of whether the model is correctly specified and whether or not autocorrelation exists amongst residuals and the assumption of stationarity is violated. The Ljung-Box statistic should have significance levels ≤ .05 (within the 95% confidence interval) for the time series under analysis. If autocorrelation is within these bounds, it is not considered to be statistically different from zero and the time series is deemed to be stationary (Ljung et al. 1978). In this case, the significance value of .018 does not violate this assumption; additionally, a check of ACF and PACF plots verifies that the assumption of stationarity is upheld.

Table 8 provides t-test results for the dependent variable BG and any predicting variable significant at the 95% confidence level. Significant AR (e.g. AR1 for the first-order autoregressive component $p = 1$), and MA (e.g. MA1 for a first-order moving average component where $q = 1$) estimates of these components (that is AR or MA) and any predicting independent variable (denoted as numerator and denominator where denominator and numerator estimates for the independent variable are respectively like the AR and MA part of the dependent variable ARIMA model) reveal which variables meaningfully contribute to predicting future values of the dependent variable with non-significant variables being excluded – this is similar to significance testing of b coefficients for ordinary regression models. Predicted values are calculated using the linear TF ARIMA equation, repeated here for convenience.

$$\Phi(B)\nabla^d Q = \mu + \sum_{i=1}^{n} \frac{\omega_i(B)_i(B)_i^{b}}{\delta(B)_i} \nabla^d x_{i,t} + \Theta(B)a_t$$

The model fit statistic RMSE, presented previously in Table 7, is a result of comparing the predicted values calculated using the TF ARIMA equation to actual observed values. The software calculates the RMSE based on solving for the abovementioned equation to determine predicted values by inputting the

significant variable parameters found in Table 8. The ARIMA equation in this instance would then look like:

$$BG_t = -.124\nabla^{4}Vmag_1 + \frac{(1 + .223 - .131 - .087)}{(1 + .245 + .505)\nabla^{d}} a_t$$

Adding a TF adds continuous variables to the right-hand side of the time series equation. The objective of adding transfer functions is to see how the independent variable influences the dependent variable rather than simply observing how previous values of the dependent variable are related to itself. The parameters (*p,d,q*) of the dependent variable were discussed previously with Table 6. The reader may recall from the *Methods* chapter that the transfer function equation contained a polynomial numerator in the form $\omega(B)_I = \omega_{0,I} - \omega_{1,I}B - \omega_{2,I}B^2 -...- \omega_{h,I}B^s$. The numerator parameters are used to establish the magnitude of the effect of a predictor variable, x_t, on the output, y_t. Similarly, the denominator polynomial from the transfer function, $\delta(B)_I = 1 - \delta_{1,I}B - \delta_{2,I}B^2 -...- \delta_{r,I}B^r$, determines the shape of the response.

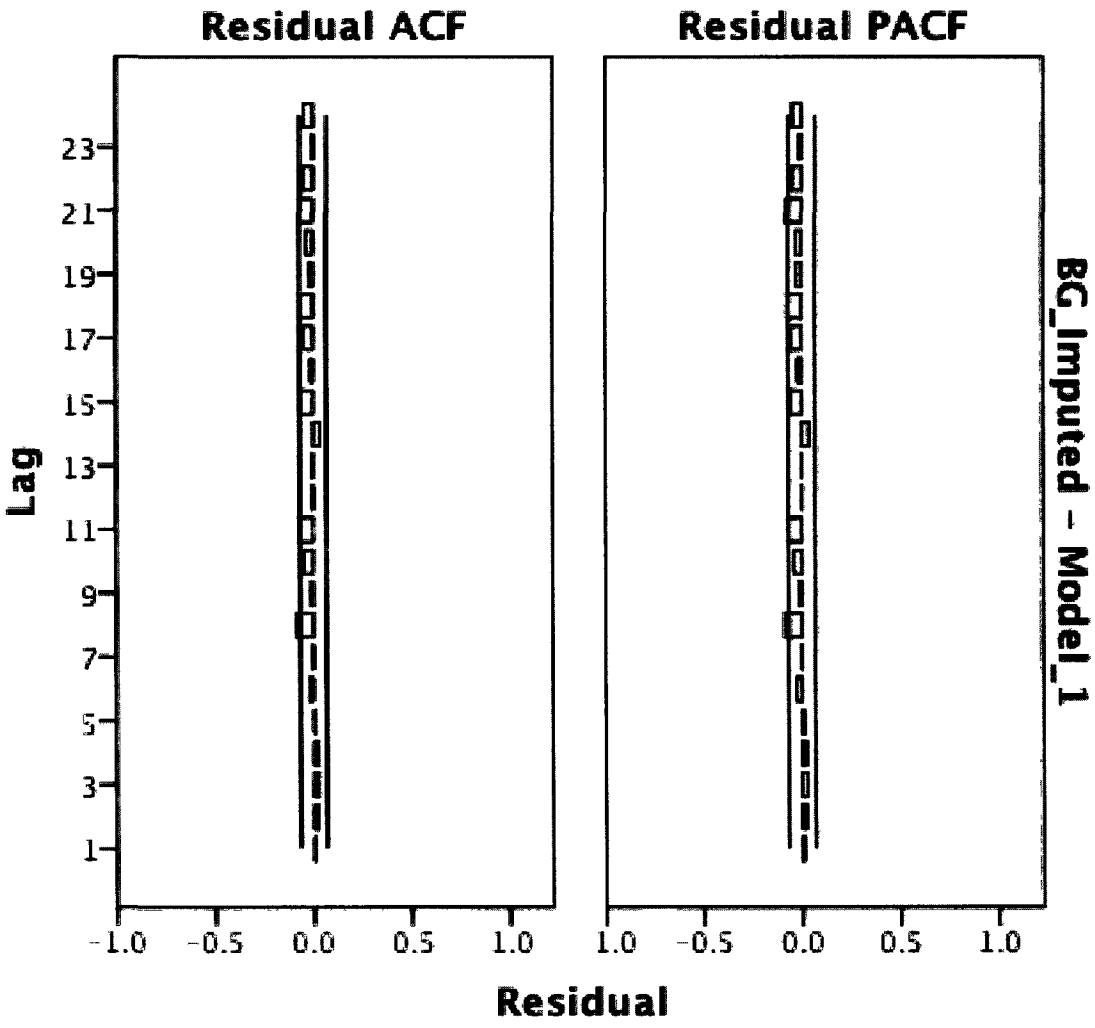


Figure 8 ACF and PACF residual plots for subject 84

Figure 9 displays the ACF and PACF plots for subject 84. Visual examination of ACF and PACF plots serves as a supplementary test to the Ljung-Box statistic to ensure that there is no autocorrelation between residuals and stationarity within the time series exists. Multiple significant spikes beyond the confidence interval limit would indicate the existence of autocorrelation requiring the re-parameterization of the model – that is a re-estimate of the (*p,d,q*) parameters to ensure the assumption of stationartiy is upheld. In this case, a spike at lag 8 in both the ACF and PACF plots and lag 21 in the PACF plot appears to have occurred randomly by chance as residual normality

(randomness amongst residuals) exists for all other lags, thereby upholding the assumption of stationarity rendering the model acceptable.

The combination of the numerator and denominator and review of the respective CCF therefore determine/confirm the direction of the TF effect on BG, that is, whether or not the predicting variable results in an increase or decrease in BG (+ or –). Results from Table 8 indicate that the independent predicting variable Vmag has a decreasing affect on BG after 3 lags (15 min). This is verified by examining the CCF of BG and Vmag. The reader may recall that Vmag may serve as an indicator for the intensity of physical activity.

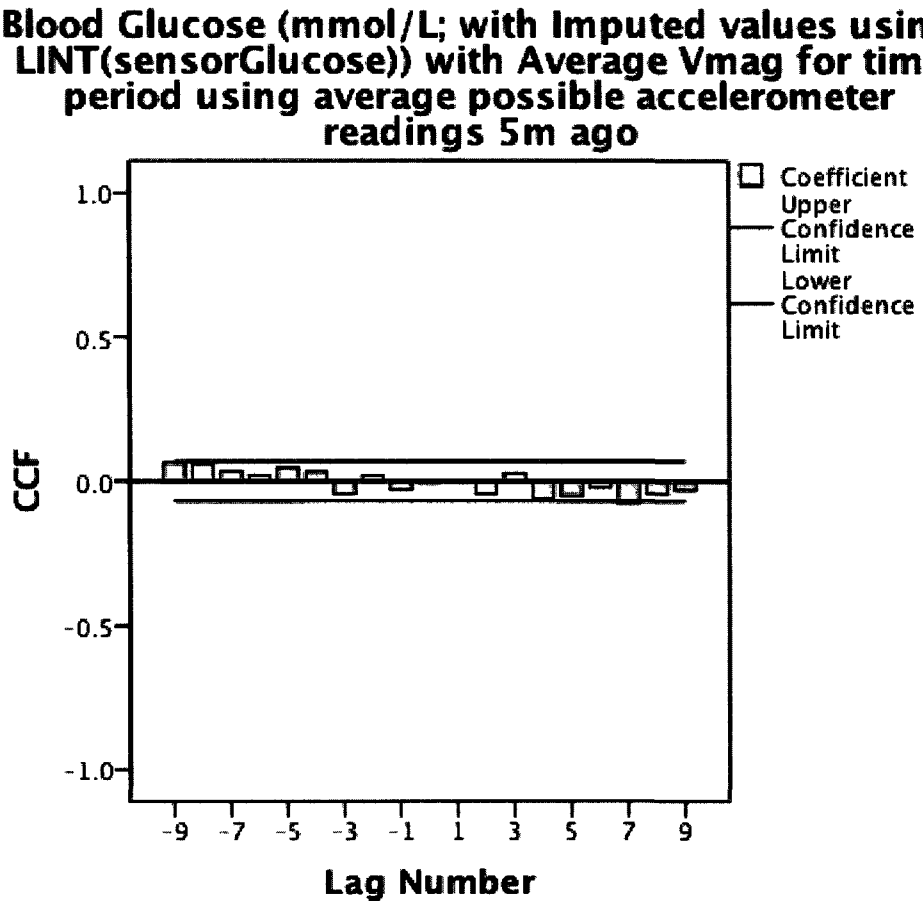


Figure 9 Cross-Correlation Function of blood glucose and Vmag for subject 84

Upon examination of the CCF for BG and Vmag for subject 84 in Figure 15, it is clear to see that Vmag after a delay of 3 lags has a decreasing effect on BG which continues through to lag 9. More simply, the overall modeling results for subject 84, as presented in Tables 6-8, Figures 14 and 15, demonstrate Vmag as having a diminishing effect on BG after 15 minutes plus some degree of unaccounted for error.

4.4 Summary of model outputs by subject

In an effort to simplify and provide only the pertinent model statistics, the outputs for all individual subject models (in other words, important statistics presented in the full example output from Tables 6 – 8 and visual verification of results from Figures 9 and 10) have been consolidated and presented in Table 9. Each row in the table contains significant modeling components pertaining to one subject. The first row (*p, d, q*) indicates the AR (*p*), *differencing/integrated* (*d*) and MA (*q*) parameters in addition to any transformation performed (N indicating a natural log transformation SR indicating a square root transformation, the reader may recall from the methods section that transformations may be required in order to achieve stationarity), followed by the Stationary R^2, RMSE and Ljung-Box (labeled L-B) statistics and parameter estimates of predicting input variables tested for significance amongst subjects. Numbers in bold represent predicting variables that had a significant effect on BG with the time delay of their impact in parenthesis. A positive number indicates the predicting variable causing an increase in BG, whereas a negative (–) sign indicates the predicting variable causing a decrease in BG.

User_ID	(p,d,q)	Predicting Inputs	S R^2	RMSE	L-B	Sugar	Carb	Cal	Insulin	Meds *	Vmag	PA	Social	Shop	Dine	Auto	Public Transit	Mean home km
61	(1,1,1)	3	620	085	272	-	**-.002(0)**	-	**-.05(0)**	**DM**	-	-	-	-	-	-	-	**.018(7)**
66	(2,1,6)SR	2	386	356	008	-	-	-	-	LMN	-	-	-	-	-	**.001(8)**	**.001(2)**	-
70	(0,1,1)	7	247	215	029	-	**-.023(1)**	-	-	M	**-.345(2)**	**.001(3)**	**.001(3)**	**.001(5)**	-	**.001(4)**	-	**-.057(1)**
73	(1,1,6)	0	527	258	352	-	-	-	-	AM	-	-	-	-	-	-	-	-
74	(1,1,2)	8	608	282	365	**-.018(0)**	**-.022(0)**	**-.774(2)**	-	HnHg	-	-	**-.01(4)**	**1 00(1)**	-	**.939(1)**	**-.971(1)**	**-.025(0)**
77	(0,2,13)	0	628	152	149	-	-	-	-	LN	-	-	-	-	-	-	-	-
80	(1,1,0)SR	4	241	293	177	-	-	**.001(0)**	-	HnN	-	-	**.607(2)**	-	**-.001(2)**	-	-	**.015(2)**
81	(0,1,2)	3	502	138	000	-	**-.020(0)-**	-	-	M	-	-	-	-	-	**.001(0)**	-	**.030(0)**
84	(2,2,9)	1	389	092	018	-	-	-	-	M	**-.345(3)**	-	-	-	-	-	-	-
86	(2,1,2)	1	589	089	415	-	**.947(1)**	-	-		-	-	-	-	-	-	-	-
87	(0,2,15)	4	723	214	007	-	-	-	-	-	**.764(0)**	**-.01(7)**	-	-	-	**.001(0)**	-	**.036(2)**
88	(1,1,3)	4	616	174	000	**.016(3)**	**.893(4)**	-	-	-	**-.387(2)**	-	-	**-.001(3)**	-	-	-	-
90	(2,1,2)	0	675	144	295	-	-	-	-	M	-	-	-	-	-	-	-	-
91	(2,1,12)	3	629	172	043	-	-	**.865(2)**	-	-	-	-	**-.03(1)**	-	-	-	-	**.-166(0)**
92	(1,1,14)	0	627	140	021	-	-	-	-	M	-	-	-	-	-	-	-	-
93	(1,1,13)	3	262	153	026	-	-	-	-	M	**-.181(5)**	**-.01(8)**	-	**.002(5)**	-	-	-	-
153	(1,1,8)	2	433	133	035	-	**-.004(0)**	-	-	M	-	-	-	-	-	-	-	**-.085(4)**
163	(2,1,0)	1	182	244	222	-	-	-	-	-	-	-	-	-	-	**-.001(1)**	-	-
167	(0,1,4)	1	622	090	055	-	-	-	-	M	**-.738(5)**	-	-	-	-	-	-	-
170	(2,1,2)	2	712	152	008	-	-	-	-	-	-	**-.01(0)**	**-.88(0)**	-	-	-	-	-
171	(1,1,13)N	1	476	217	357	**.008(5)**	-	-	-	DM	-	-	-	-	-	-	-	-
172	(0,2,2)	3	735	114	241	-	-	-	-	DMA	**-.087(0)**	**-.01(1)**	-	-	**-.740(0)**	-	-	-
173	(0,1,4)N	5	619	148	002	**-.007(3)**	-	-	-	M	-	**-.01(2)**	**.001(1)**	**.001(8)**	-	**-.001(0)**	-	-
175	(1,1,3)SR	4	681	237	229	-	**-.706(0)**	**-.610(0)**	**-.28(0)**	**Ln**	**-.129(0)**	-	-	-	-	-	-	-
176	(2,1,0)	2	579	099	078	-	**-.007(3)**	-	-	M	-	-	-	-	-	-	**.001(7)**	-
177	(0,2,9)	3	793	163	031	-	**-.060 (4)**	**-.010(4)**	-	-	**-.616(1)**	-	-	-	-	-	-	-
178	(2,1,3)N	1	361	220	006	-	-	-	-	AM	-	-	-	-	-	**-.001(3)**	-	-
196	(1,1,10)	0	636	129	000	-	-	-	-	M	-	-	-	-	-	-	-	-

Meds

Actos (A)	Diabeta (D)	Humalin (Hn)	Humalog (Hg)	Lantus (Ln)	Levemir (L)	Metamorfin (M)	NovaRapid (N)

Table 9 Consolidated model output including significant predicting variable

Stationary R^2 values range from .182 – .793 with an average of .539, where the higher the value the greater the model explanation of variance. RMSE values varied from .085 – .356 with an average of .175, where the lower the value, the greater the models ability to predict for future values of BG. The Ljung-Box values ranged from .000 – .415, with an average of .123. Values less than .05 (within the 95% confidence limit) indicate residual randomness and stationarity amongst the time series. In cases where the Ljung-Box value is greater than .05 a visual inspection of ACF and PACF plots verified that indeed these values occurred by chance and that residual normality (or residual randomness indicating stationarity) exists for remaining lags, upholding the assumption of stationarity, thus, deeming the models to be acceptable.

Generally, no two models performed identically. That is, all models presented unique statistical values and (p, d, q) parameters suggesting each subject having unique BG fluctuation and correlating factors. Note particularly, the difference in AR(p) values by subject, referring to the regressed nature (or relatedness) of previous BG values to current BG values indicating perhaps a unique difference in subjects metabolic response to BG change. Additionally, two basic groups can first be differentiated: those whose BG appears to be little influenced by outside factors (including even food and exercise) and thus appear to have their BG largely under control, versus those whose BG is sensitive to a variety of factors, and thus appear to have less control of their BG. For instance, subjects 86, 88 and 91 who are not medicating appear to demonstrate an increase in BG shortly after eating. Other subjects (61, 70, 74, 81, 153, 176, 177) appear to be able to control for BG, likely with the help of exogenous insulin that appears to even decrease BG shortly after eating. Also interesting is subject 80 – the only subject to use insulin medication that appears to have an increase in BG shortly after eating and may have something to do with the combination of insulin medication they are using, as they are the only one to use Humalin and Humalog.

Frequently occurring combinations of predicting variables amongst subjects are presented in Table 9, including Physical Activity (PA) and Vmag. Not surprisingly a variety of food variables combine as predicting for BG. Additionally, PA for subjects range from 0 – 2.3 hours a day with an average of 0.7. Note for the subjects who have PA serving as a predicting variable (that is, values in bold under the column with the heading title 'PA' and the delayed response prior to having an influence on BG in parenthesis, that is an increasing or a decreasing influence on BG, indicated by a + or – symbol) that only two of the six (170 and 172) engage in less than the average PA time. Lastly, significant parameter estimates are related with both exercise variables (Vmag and PA) and nutrition variables.

Independent Variables	Description	Number of Significant Predictors Female	Number of Significant Predictors Male	Total Number of Significant Predictors
sugar_5m	sugar consumed	2	2	4
carbs_5m	carbohydrates consumed	4	6	10
calories_5m	calories consumed	3	2	5
med_ins_5m	insulin medication taken	0	2	2
avgavgVmag_5m	square root of the sum of acceleration	7	2	9
event21x_time_5m	time spent involved in physical activity	3	3	6
event12_time_5m	time traveling in an automobile	5	3	8
event30x_time_5m	time spent commuting via public transit	1	2	3
Distance_home	distance to home	5	3	8
event27_time_5m	time spent at a restaurant	1	1	2
event23_time_5m	time spent shopping	3	2	5
ppl_time_5m	time spent with people	3	3	6

Table 10 Number of significant independent BG predicting variables

In terms of significant explanatory/predictor variables, both the *types* of variables and the *number* of them varied substantially from patient to patient, as summarized in Table 10. Generally, the results indicate a wide variety of combinations of BG predicting variables amongst subjects. This further suggests that the process of BG fluctuation is not a straightforward one and instead appears quite subject-specific. Well-known BG correlates such as carbohydrates/sugar consumed and exercise, are not even present in all patients models, as somewhat expected according to existing literature. The most frequent predictor variables were carbohydrate and Vmag, the least frequent were distance from home, and surprisingly, insulin. Interestingly, models that had the greatest predicting ability for BG (that is, the lowest RMSE value) indicated physical activity in the form of time spent or Vmag as the predicting variables. With respect to non-traditional explanatory variables, it was quite interesting to find that geographic-related variables such as time spent travelling by automobile, public transit, and distance from home significantly affected many subjects, on par even with the number affected by expected variables

related to food intake (such as sugar consumed). Equally interesting were the significant affects of time use variables, such as time spent at restaurants, shopping or with other people. Lastly, when observing the impacts of gender, most variables seem to be relatively split between male and female except for Vmag, where a high proportion of female subjects (7) experienced this a predicting variable. Further interpretation of these affects and the predicting variables presented in Table 10 is explored in the next chapter.

Chapter 5: Discussion and Conclusion

5 Discussion

This thesis has utilized a combination of unique data collection and analysis techniques at a very fine spatial and temporal scale. The intent was to explore the various possible real-world predictive variables on blood glucose, including well-known correlates related to food and exercise, and new correlates related to daily life and geography. As such, this thesis demonstrates key findings that relate to the fields of geography and health. The discussion below first highlights several themes related to the modeling results, followed by methodological contributions, challenges and limitations, and future work.

5.1 Time Series Models: Emergent Themes

All models are wrong, some are useful.

George E. P. Box, *British statistician (Date unknown)*

5.1.1 Impacts of food and medication

Consistent with existing literature, the most telling results from this thesis come from food and medication input variables. In all cases where more than one food variable reports as a predicting variable for BG, each variable has the same effect on BG; that is, collectively they either indicate an increase or a decrease in BG all with the same time lag, except for one case where the delay of more complex sugar in the form of carbohydrates occurs one lag later.

A surprising mix of responses to food consumption was observed. In the 4 cases where food consumption resulted in an increase in BG (subjects 80, 86, 88, 171), three were not medicating to control for their BG. In the other case,

subject 80 was using a unique combination of medications. All other subjects medicating to control for BG saw a decrease in their BG in 0-4 lags after food consumption; presumably owing to the effects of medication that is most often taken in conjunction with meals. These findings are unique in highlighting the mixed impacts of medication in optimizing BG control, revealed in part as a result of the detailed nature of the data.

5.1.2 Inconsistent impacts of transportation, distance and activity-time variables

In contrast to above, variables associated with time spent travelling and distance from home had much less consistent impacts on subjects with found to be significant in the model. Lag effects varied very widely from between 0-8, and the direction in which they altered BG (+ or -) was not constant amongst subjects. This conforms to the past findings (Riddle et al. 2003, Sagburg 2006, Strauss et al. 2006) that have demonstrated distance traveled and time spent commuting as having irregular and unique effects on BG control citing that regular routines are disrupted inhibiting DM patients to control for BG. Interestingly, subject 74 demonstrated an increase in BG after 1 lag while driving and a decrease after 1 lag while taking public transit. This may be attributed to the stress associated with driving (Stokols 1978, Novaco 1979, Gulian 1990, Novaco 1990, Howard 1994, Apparies 1998, Hennessy 1999, Healey 2005, Hill 2006) and its ability to result in an increase in BG (Cox 2002, Wiesli et al. 2005). This evidence suggests that driving may be more stressful for this particular subject as he/she i better able to control for his/her BG when using public transit.

Although there is a lack of specific literature to suggest that certain activities such as shopping or socializing have a definitive impact on BG, results from this thesis support the notion that they may have unique idiosyncratic effects on patients as subjects 70, 74, 80, 88, 91, 93, 170 and 173 show these

variables as predicting for BG. In all but one case (subject 88) shopping results appear to have an increasing effect on BG varying between 1-8 lags. Perhaps shopping for subjects 70, 74, 93, and 173 is a stressful experience resulting in an inability to control for BG and a subsequent increase in BG. An alternative explanation may be that regular patterns of controlling for BG are disrupted while shopping. Conversely, time spent with people in all but one case demonstrated as having a decreasing effect on BG in 0-4 lags. Maybe for these subjects socializing does not hinder their ability to control for BG. Also interesting are the two dining cases where subjects 80 and 172 demonstrate dining as having a decreasing effect on BG in two or less lags. In each case subjects are medicating to control for BG which is consistent with food consumption results.

5.1.3 The complexity of physical activity

Modeling results of the effects of physical activity on BG are not exactly straightforward. According to a number of past studies (Boule et al. 2001, Healy et al. 2008, Marcus et al. 2008, Ekelund et al. 2009) generally an increase in physical activity will result in a decrease in BG levels over time, however none of these studies are able to predict with much certainty when this might occur. Alternatively, physical activity deemed stressful may even result in an increase in BG (Wiesli et al. 2005). There are also a variety of other factors to consider for that may have an impact on BG, namely food and medication and the time of their consumption.

In this thesis, Vmag was significant for 8 subjects and 'exercise time' 6 cases, all exhibiting a decrease in BG ranging from 0-8 lags, except for subject 172 where Vmag and exercise time report as having an increasing effect on BG. This may be due to the fact that subject 172 reacts differently to exercise, or they may consider exercise to be stressful. Nonetheless, these results conform to the

literature suggesting that the effects of exercise on patients with DM are not straightforward. Also interesting are the results for subject 87, as both exercise time and Vmag report as predicting variables, however, Vmag indicates as having an increase in BG while exercise time a decrease. The results for subject 87 could conform with the notion that although exercise measured in Vmag has an effect with no delay and initially may be deemed to be stressful resulting in an immediate increase in BG, after 7 lags (or 35 minutes) the effects of exercise result in decreasing BG. This idea would be consistent with the existing literature.

5.1.4 Overall Discussion

Past studies have suggested there to be an idiosyncratic nature of controlling for BG (Kovatchev et al. 2002, Wearden et al. 2006). In other words, BG variation and subsequent attempts to control BG are specific and unique for each patient. The results from this thesis would appear to support and extend this notion from three perspectives: firstly, it was clear that some subjects' BG varied much more than others, suggesting less overall control; secondly, the types of factors influencing BG varied widely, suggesting that not all patients are sensitive to (or control effectively for) the same factors; and thirdly, even well known factors such as food and medicine did not have consistent affects on all patients.

This thesis also elaborates on several new correlates of BG fluctuation that have impacts on par with those of traditional variables such as food and exercise (at least with respect to how many subjects were influenced by these variables). These new correlates, in the form of activity and travel engagement, social engagement, and distance from home are associated with behavioural geography and, as Timmermans and Golledge (1990) and Saarloos et al. (2009) have suggested, it is necessary to examine such variables to see how they may

influence health. Over 40 years ago Tobler (1970) invoked the first law of geography, "everything is related to everything else, but near things are more related than distant things," (p. 234) when he developed a demographic model for the purpose of forecasting population growth and its geographical distribution. This thesis contributes to the behavioural geography research call and utilizes Tobler's law by enhancing our understanding of how the structuring of spatial/temporal daily activities influence health.

Coincidently around the same time as Tobler, Box and Jenkins (Box et al. 1970a) released their highly influential work used for forecasting. It is Tobler's precept and Box-Jenkins method that form the basis of analysis in this thesis. The time series approached used here examines the relationship and influence of events near in time to a dependent variable, BG. Additionally, Tobler emphasizes avoiding the use of over complex models, stating this should be one of the researchers objectives. However, and as stated previously in Chapter 3, ordinary regression models did not compete for the purpose of this thesis, as the data employed required more robust time sensitive dynamic regression models for proper interpretation.

This thesis employs the combination of data assumed using new data collection technologies and an old analysis technique, under the guise of a new tool for the purpose of addressing a current health problem. That is, to attend to a progressing disease, diabetes, by observing possible lifestyle changes or behaviour modification in order to promote an increase in quality of life for those who suffer with this ailment. As noted in the disclaimer in this chapter, while the modeling results from this thesis are by no means substantiated by the medical community and effort is made to avoid discussion of the biochemical and psychological aspects relating to how metabolism and the autonomic nervous system influences BG – as this would probably require a treatise of some sort – they do however demonstrate a number of key geographical findings to be discussed further in the next section.

Although the technique employed in this thesis is not simple and indeed, at the onset of examining the complex nature of BG fluctuation in relation to a wide range of possible influencing variables, the results may seem chaotic, random or complex. However, findings from this thesis suggest that in fact, BG change may be explained by a wide variety of variables that may be unique to certain groups of DM patients. This notion suggests grounds for a larger study in order to possibly identify with greater certainty specific groups of DM patients where effective disease management and control of BG is possible through specific behavioural modification.

The results of this thesis also reflect and expand upon more recent key aspects of conceptual frameworks introduced by Helfenstein (Helfenstein 1991, 1996), Imhoff (Imhoff et al. 1998) Bellazzi et al. (Bellazzi et al. 2000, Bellazzi et al. 2009) and Takeuchi et al. (Takeuchi et al. 2009) regarding time series analysis of health information; (Lymberis 2004, Baert et al. 2006, Otto et al. 2006, Pandian et al. 2008, Meyfroidt et al. 2010, Pantelopoulos et al. 2010) involving the continuous monitoring of health information by way of wearable sensors; the culmination of which as identified by Saarloos et al. (2009) and Doherty et al. (2006) as providing *key opportunities* for geographers to contribute to the health sciences by linking space and time at the individual level. Furthermore, this thesis has made use of powerful methods for storing and disseminating vast amounts of health information contributing to the idea that many researchers believe a major breakthrough is eminent at both the basic levels and clinical levels of investigation that will result in revolution of health monitoring in the not so distant future.

5.2 Key methodological contributions

Despite several challenges and limitations (see also, next section), results from this thesis suggest that it is indeed possible to obtain spatial and physiological data at regular even sub-second intervals over long periods of time for the purpose of extracting or 'data-mining' for significant health influencing factors. The analysis undertook in this thesis are considered especially novel as physiological habits specific to individual subjects make an unsupervised learning approach under real world conditions ideal – in contrast to laboratory conditions (Duchene et al. 2007). Although reluctant to draw any specific medical conclusions, the uniqueness of the data and analysis technique used in this thesis suggests a potential clinical value for these types of methodologies.

More specifically, the findings from this thesis are able to build upon previous attempts to examine the relationship of BG with various external factors. Past studies have had to rely on data observed at extremely large intervals measured in days (Helfenstein 1996), make use of a limited number BG predicting variables (Ekelund et al. 2009) or rely on data at larger intervals (30 minutes) and then apply analysis techniques that do not control for autocorrelation of residuals (Takeuchi et al. 2009). This thesis was able to contribute to the existing approaches by:

1. Demonstrating the ability to dynamically analyze BG fluctuation effectively at short intervals and control for autocorrelation of residuals;

2. Testing the BG predicting ability of a wide range of daily life event variables including the first attempt to examine the relationship between time-measured activity variables on BG and;

3. Demonstrate the effectiveness of such data collection and analysis techniques for the purpose of predicting BG and other possible time lagged health outcomes.

Furthermore, this thesis incorporated traditional diary style time measured activity data resulting in hybridization with human monitoring physiological data. This opportunity for collaboration between data collection techniques has allowed for the chance to explore spatial/temporal associations and how social relations may support or weaken health and well-being (Rainham et al. 2008). Further to this, Saarloos et. al (2009) suggest that past approaches to examine health at the individual level previously circumvent how health behaviours really occur due to individuals interacting with their environment and each other. In order to model and forecast environmental influences on health behaviour, such as nutrition and physical activity, better understanding how activities are linked in space and time at the individual level is required. This thesis serves as an early attempt at doing so.

5.3 Challenges and limitation

The accuracy of some of the input variables may have distorted results. Whereas some variables were measured passively (e.g. BG, location) in order to limit the amount of human error, some were self-reported by subjects, such as food intake. It is likely that some degree of error regarding food diary input existed, whether it be a subject not eating what they had recorded/intended, or simply not recording what they had eaten. Such errors may contribute to error in the models and possible misinterpretation. Also, over medicating or not reporting medicating may add ambiguity to modeling results.

Additionally, a few equipment related challenges existed during data collection. Some devices suffered from some data outage issues for technical and

user-related reasons, but that overall capture rates were reasonably high. In the end, 13 subjects were excluded from analysis for these reasons, in an attempt to reduce the impact of missing/poor quality data on the results. Other devices suffer from additional measurement issues, the most relevant being the accelerometer - a high Vmag recording in this thesis was presumed to be associated with engagement in physical activity. This may in fact be artificial and the result of mechanically assisted travel velocity, or 'non-active' transport in the form of vehicular travel. Such affects were not accounted for in this thesis.

5.4 Future Work

Recently, the increasing amounts of health and human activity data with high dimensionality has provided researchers the opportunity to discover new patterns and build models from large datasets; also referred to as data-mining or knowledge discovery. Such applications dealing with temporal sequences, encourage the further development of the "time-series mining" research field. New initiatives may include the development of discovery algorithms that serve to extract important patterns and identify similarities or trends for the purpose of description and/or prediction. Using data collection methods and time sensitive analysis techniques presented in this thesis makes possible for the grouping of traditional health determinants to examine for spatial/temporal patterns among similar population groups to examine for places in time that may be responsible for adverse effects on health and well-being. This may be most useful for monitoring purposes, like in a home-care setting, when dealing with detecting abnormal behaviours, threats to the loss of independence and the need to develop personalized health management strategies.

Additional future work could involve the examination of each time measured activity episode, treating each like the input predicting variable insulin used in this thesis. That is, isolate in the form of a 'time window' with

start and end times for each specific activity episode, like 'time spent with people'. This would allow for each activity episode to then be treated as an event that results in an increase or a decrease in BG. Additionally, future studies may want to include a more sophisticated acceleration monitoring system. Although current studies are aimed at improving the accuracy of a singular accelerometer unit in detecting and monitoring physical activity, until then, one device provides limited information leaving much to speculation.

5.5 Conclusions

This thesis explored a key new opportunity for geographers to contribute to the health sciences by using data collected from emerging human location/activity tracking technology combined with body-based biophysical sensors and dynamically analyzing diabetic health information using a time-series sensitive analysis technique. This has allowed for an unprecedented opportunity to examine the affects of human geographies (activities, travel, exercise, social interaction, exposure, etc.) on health outcomes at a very fine spatial/temporal scale over long periods of time under real-world conditions in the built environment.

Furthermore, this thesis was able to demonstrate the potential for continuous monitoring of DM patients who are trying to optimize their BG control and echoes the sentiment of Dr. P. Oh, data collection research collaborator from TRI, by combining a time-series sensitive analysis technique to a unique data set:

> "We all know that exercise, heart rate and diet are important factors in successful management of a diabetic patient's condition [and] we anticipate that this GPS-supported monitoring and alerting system will place greater control of health management in the hands of the patient. At a minimum, it will allow individuals to gain a

greater appreciation of the effects their, work, travel, recreation and eating habits have on their health" (WLU 2006).

Findings on a subject-to-subject basis certainly suggest that the uniqueness and complexity of treating for DM. Albeit these findings are unlikely to guide immediate policy and health practice change, they do highlight the opportunity and need for a variety of disciplines to continue to collaborate in order to improve our understanding of human health from a behavioural context.

References:

Aitken, S. & G. Valentine. 2006. *Ways of knowing and ways of doing geographic research.* London Sage Publications

Allemann, S., C. Houriet, P. Diem & C. Stettler (2009) Self-monitoring of blood glucose in non-insulin treated patients with type 2 diabetes: a systematic review and meta-analysis. *Current Medical Research Opinion,* 25, 2903-13.

Allen, B. T., E. R. DeLong & J. R. Feussner (1990) Impact of glucose self-monitoring on non-insulin-treated patients with type II diabetes mellitus. Randomized controlled trial comparing blood and urine testing. *Diabetes Care,* 13, 1044-50.

Apparies, R. J., T.C. Riniolo and S.W. Porges (1998) A psychophysiological investigation of the effects of driving longer-combination vehicles. *Ergonomics,* 41, 581-592.

Auchincloss, A. H. & A. V. Diez Roux (2008) A New Tool for Epidemiology: The Usefulness of Dynamic-Agent Models in Understanding Place Effects on Health. *American Journal of Epidemiology,* 168, 1-8.

Baert, K., B. Gyselinckx, T. Torfs, V. Leonov, F. Yazicioglu & S. Brebels (2006) Technologies for highly miniaturized autonomous sensor networks. *Interaction with Computers,* 18, 171-186.

Bajaj, R., S. L. Ranaweera & D. P. Agrawal (2002) GPS: location-tracking technology. *Computer,* 35, 92-94.

Barnett, A. H., A. J. Krentz, K. Strojek, J. Sieradzki, F. Azizi, M. Embong, S. Imamoglu, J. Perusicova, V. Uliciansky & G. Winkler (2008) The efficacy of self-monitoring of blood glucose in the management of patients with type 2 diabetes treated with a gliclazide modified release-based regimen. A multicentre, randomized, parallel-group, 6-month evaluation (DINAMIC 1 study). *Diabetes Obes Metab,* 10, 1239-47.

Bauman, A. E., J. F. Sallis, D. A. Dzewaltowski & N. Owen (2002) Toward a better understanding of the influences on physical activity: The role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American Journal of Preventive Medicine,* 23, 5-14.

Bellazzi, R. & A. Abu-Hanna (2009) Data mining technologies for blood glucose and diabetes management. *Journal of Diabetes Science and Technology,* 3, 603-12.

Bellazzi, R., C. Larizza, P. Magni, S. Montani & M. Stefanelli (2000) Intelligent analysis of clinical time series: an application in the diabetes mellitus domain. *Artif Intell Med,* 20, 37-57.

Bethea, R. M. & R. R. Rhinehart. 1991. *Applied Engineering Statistics.* New York, NY: Marcel Dekker.

Boarnet, M. G. 2005. The Built Environment and Physical Activity: Empirical Methods and Data Resources. In *Does the Built Environment Influence Physical Activity? Examining the Evidence.* Washington, D. C.: Transportation Resarch Board and the Institute of Medicine Committee on Physical Activity, Health, Transportation, and Land Use.

Booth, K. M., M. M. Pinkston & S. C. P. Walker (2005) Obesity and the built environment. *Journal of American Dietetic Association,* 105, 110-117.

Boule, N. G., E. Haddad, G. P. Kenny, G. A. Wells & R. J. Sigal (2001) Effects of exercise on glycemic control and body mass in Type 2 Diabetes Mellitus: a meta-analysis of controlled clinical trials. *Journal of the American Medical Association,* 286.

Bowerman, B. & R. T. O'Connell. 1993. *Forecasting and Time Series.* Belmont, CA: Wadsworth.

Box, G. E., G. M. Jenkins & G. C. Reinsel. 2008. *Time Series Analysis: Forecasting and Control.* New York: John Wiley and Sons.

Box, G. E. P. & G. M. Jenkins. 1970a. *Time series analysis: Forecasting and control.* San Francisco: Holden-Day.

Box, G. E. P. & D. A. Pierce (1970b) Distribution of Residual Autocorrelation in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association,* 65.

Brennan Ramirez, L. K., C. M. Hoehner, R. C. Brownson, R. Cook, C. T. Orleans, M. Hollander, D. C. Barker, P. Bors, R. Ewing, R. Killingswort, K. Petersmarck, T. Schmid & W. Wilkinson (2006) Indicators of activity friendly communities: An evidence-based consensus process. *American Journal of Preventive Medicine,* 31, 515-524.

Brockwell, P. J. & R. J. Davis. 1991. Time Series: Theory and Methods. 256-273. New York: Springer-Verlga.

Chatfield, C. 1991. *The Analysis of Time Series: An Introduction.* Great Britain: Chapman and Hall.

Colabianchi, N., M. Dowda, K. A. Pfeiffer, D. E. Porter, M. J. Almeida & R. R. Pate (2007) Towards an understanding of salient neighborhood boundaries: adolescent reports of an easy walking distance and convenient driving distance. *Int J Behav Nutr Phys Act,* 4, 66.

Cox, D. J., L.A. Gonder-Frederick, B.P. Kovatchev, and W.L. Clarke (2002) The metabolic demands of driving for drivers with type 1 diabetes mellitus. *Diabetes/Metabolism Research and Reviews,* 18, 381-385.

Cummins, S. & S. Macintyre (2006) Food environments and obesity -- neighbourhoods or nation. *International Journal of Epidemology,* 35, 100-104.

Daniels, H. E. (1956) The approximate distribution of serial correlation coefficients. *Biometrika,* 43, 169-185.

Diez Roux, A. V. (2002) Invited commentary: places, people, and health. *American Journal of Epidemiology,* 155, 516-9.

--- (2007) Integrating social biologic factors in health research: A system view. *Annals of Epidemology* 17, 569-574.

Diez-Roux, A. V. (2008) Next steps in understanding the multilevel determinants of health. *Journal of Epidemiology and Community Health,* 23, 15-25.

Doherty, S. T. 2009. Emerging methods and technologies for tracking physical activity in the built environment. In *Transport Survey Methods: Keeping up with a Changing World*, eds. P. Bonnel, M. Lee-Gosseling, J. Zmud & J. -L., 153-190. Bingley, U.K.: Emerald.

Doherty, S. T. and Oh, P. (forthcoming) A continuous monitoring system of diabetic patients' daily activities and biophysical conditions.

Duchene, F., C. Garbay & V. Rialle (2007) Learning recurrent behaviors from heterogeneous multivariate time-series. *Journal of Artificial Intelligence in Medicine,* 39, 25-47.

Duhl, L. J. & A. K. Sanchez. 1999. Healthy cities and the city planning process: a background document on links between health and urban planning. ed. W. H. Organization. Copenhagen, Denmark.

Duncan, M. J., H. M. Badland & W. K. Mummery (2009) Applying GPS to enhance understanding of transport-related physical activity. *Journal of Science and Medicine in Sport,* 12, 549-56.

Ekelund, U., S. Brage, S. J. Griffin & N. J. Wareham (2009) Objectively measured moderate- and vigorous-intensity physical activity but not sedentary time predicts insulin resistance in high-risk individuals. *Diabetes Care,* 32, 1081-6.

Elgethun, K., R. A. Fenske, M. G. Yost & G. J. Palcisko (2003) Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. *Environmental Health Perspectives,* 111, 115-122.

Elgethun, K., M. G. Yost, C. T. Fitzpatrick, T. L. Nyerges & R. A. Fenske (2007) Comparison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time-location patterns. *Journal of Exposure Science and Environmental Epidemiology,* 17, 196-206.

Ermes, M., J. Parkka, J. Mantyjarvi & I. Korhonen (2008) Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions. *IEEE Transactions on Information Technology in Biomedicine* 26, 20-26.

Evans, E. & R. Patry (2004) Management of gestational diabetes mellitus and pharmacists' role in patient education. *American Journal of Health-System Pharmacy,* 61, 1460-5.

Farnum, N. R. & L. W. Stanton. 1989. *Quantitative Forecasting Methods.* Boston: PWS-Kent Publishing.

Foerster, F., M. Smeja & J. Fahrenberg (1999) Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. *Computers in Human Behavior,* 15 571-583.

Frank, L. D., M. A. Andresen & T. L. Schmid (2004) Obesity relationships with community design, physical activity, and time spent in cars. *American Journal of Preventive Medicine,* 27, 87-96.

Frank, L. D. & P. O. Engelke (2001) The built environment and human activity patterns: Exploring the impacts of urban form on public health. *Journal of Planning Literature,* 16, 202-218.

French, D. P., A. N. Wade, P. Yudkin, H. A. Neil, A. L. Kinmonth & A. J. Farmet (2008) Self-Monitoring of blood glucose changed non-insulin-treated type 2 diabetes patients' belief about diabetes and self-monitoring in a randomized trail. *Diabetic Medicine,* 25.

Giles-Corti, B. & R. J. Donovan (2002) The relative influence of individual, social and physical environment determinants of physical activity. *Social Sciences Medicine.*

Glasgow, R. E., P. A. La Chance, D. J. Toobert, J. Brown, S. E. Hampson & M. C. Riddle (1997) Long-term effects and costs of brief behavioural dietary intervention for patients with diabetes delivered from the medical office. *Patient Educ Couns,* 32, 175-84.

Glasgow, R. E., P. A. Nutting, D. J. Toobert, D. K. King, L. A. Strycker, M. Jex, C. O'Neill, H. Whitesides & J. Merenich (2006a) Effects of a brief computer-assisted diabetes self-management intervention on dietary, biological and quality-of-life outcomes. *Chronic Illness,* 2, 27-38.

Glasgow, R. E., L. A. Strycker, D. K. King, D. J. Toobert, A. K. Rahm, M. Jex & P. A. Nutting (2006b) Robustness of a computer-assisted diabetes self-

management intervention across patient characteristics, healthcare settings, and intervention staff. *American Journal of Managed Care,* 12, 137-45.

Glass, L. (2009) Introduction to Controversial Topics in Nonlinear Science: Is the Normal Heart Rate Chaotic. *Chaos,* 19.

Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng & H. E. Stanley (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation,* 101, E215-20.

Greaves, S. & T. Issarayangyun (2007) Analysis of minute-by-minute exposure to fine particulates inside a car – A time-series modelling approach *Transportation Research Part D,* 12, 346-357.

Grimm, V. & S. F. Railsback. 2005. *Individual-based modeling and ecology.* Princeton, NJ, USA: Princeton University Press.

Guerci, B., P. Drouin, V. Grange, P. Bougneres, P. Fontaine, V. Kerlan, P. Passa, C. Thivolet, B. Vialettes & B. Charbonnel (2003) Self-monitoring of blood glucose significantly improves metabolic control in patients with type 2 diabetes mellitus: the Auto-Surveillance Intervention Active (ASIA) study. *Diabetes and Metabolism,* 29, 587-94.

Gulian, E., Glendon A.I., Mathews G. Davies D.R., and Debney M. (1990) The stress of driving: a diary study. *Work and Stress,* 4, 7-16.

Harvey, A. & S. J. Koopman (1996) Structural time series models in medicine. *Statistical Methods in Medical Research,* 5, 23-49.

Harvey, A. C. 1991. *The Econometric Analysis of Time Series.* Cambridge: MIT Press.

Healey, J. A., and R. Picard (2005) Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions of Intelligent Transportation Systems,* 6, 156-167.

Healy, G. N., K. Wijndaele, D. W. Dunstan, J. E. Shaw, J. Salmon, P. Z. Zimmet & N. Owen (2008) Objectively measured sedentary time, physical activity, and

metabolic risk: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). *Diabetes Care,* 31, 369-71.

Helfenstein, U. (1991) The use of transfer function models, intervention analysis and related time series methods in epidemiology. *International Journal of Epidemiology,* 20, 808-15.

--- (1996) Box-Jenkins modelling in medical research. *Statistical Methods in Medical Research,* 5, 3-22.

Helmerhorst, H. J., K. Wijndaele, S. Brage, N. J. Wareham & U. Ekelund (2009) Objectively measured sedentary time may predict insulin resistance independent of moderate- and vigorous-intensity physical activity. *Diabetes,* 58, 1776-9.

Hennessy, D. A., and D.L. Wiesenthal (1999) Traffic Congestion, Driver Stress, and Driver Aggression. *Aggressive Behavior,* 25, 409-423.

Hill, J. & J. Peters (1998) Environmental contributions to the obesity epidemic. *Science,* 280, 1371-1374.

Hill, J. D., and L.N. Boyle (2006) Driver stress as influenced by driving maneuvers and roadway conditions. *Transportation Research Part F,* 10, 177-186.

Howard, A., M. Joint (1994) Fatigue and Stress in Driving. *Employee Counselling Today,* 6, 3-7.

Hu, F. B., R. M. van Dam & S. Liu (2001) Diet and risk of type II diabetes: the role of types of fat and carbohydrate. *Diabetologia,* 44.

Humpel, N., N. Owen & E. Leslie (2002) Environmental factors associated with adults' participation in physical activity. *American Journal of Preventive Medicine,* 22, 188-199.

Hung, T., J. L. Sievenpiper, A. Marchie, C. Kendal & D. Jenkins (2003) Fat versus carbohydrate in insulin resistance, obesity, diabetes and cardiovascular disease. *Current Opinion in Clinical Nutrition and Metabolic Care,* 6.

Imhoff, M. & M. Bauer (1996) Time series analysis in critical care monitoring. *New Horizons,* 4, 519-31.

Imhoff, M., M. Bauer, U. Gather & D. Lohlein (1998) Statistical pattern detection in univariate time series of intensive care on-line monitoring data. *Intensive Care Med,* 24, 1305-14.

Istepanian, R. S., A. Sungoor & K. A. Earle (2009) Technical and compliance considerations for mobile health self-monitoring of glucose and blood pressure for patients with diabetes. *IEEE Engineering Medicine Biology Society,* 2009, 5130-3.

Jenkins, G. M. (1954) Tests of hypotheses in the linear autoregressive model, I. *Biometrika,* 43, 186-199.

--- (1956) Tests of hypotheses in the linear autoregressive model, II. *Biometrika,* 43, 186-199.

Jones, H., L. Edwards, T. M. Vallis, L. Ruggiero, S. R. Rossi, J. S. Rossi, G. Greene, J. O. Prochaska & B. Zinman (2003) Changes in diabetes self-care behaviors make a difference in glycemic control: the Diabetes Stages of Change (DiSC) study. *Diabetes Care,* 26, 732-7.

Jovanovic, L. (2009) Using meal-based self-monitoring blood glucose (SMBG) data to guide dietary recommendations in patients with diabetes. *Diabetes Education,* 35, 1023-30.

King, A. C., W. A. Satariano & J. Marti (2008) Multilevel modeling of walking behavior: Advances in understanding the interactions of people, place, and time. *Medicine and Science in Sports and Exercise,* 40, S584-593.

King, G., N. Torres, C. Potter, B. T. & K. Coleman (2004) Comparison of Activity Monitors to Estimate Energy Cost of Treadmill Exercise. *Medicine and Science Sports and Exercise,* 36, 1244-1251.

Kovatchev, B. P., D. J. Cox, L. Gonder-Frederick & W. L. Clarke (2002) Methods for quantifying self-monitoring blood glucose profiles exemplified by an examination of blood glucose patterns in patients with type 1 and type 2 diabetes. *Diabetes Technology and Therapeutics,* 4, 295-303.

Kumanyika, S. K., L. Van Horn, D. Bowen, M. G. Perri, B. J. Rolls, S. M. Czajkowski & E. Schron (2000) Maintenance of dietary behavior change. *Health Psychol,* 19, 42-56.

Lindheim, R. & S. L. Syme (1983) Environments, People, and Health. *Annual Review of Public Health,* 4, 335-359.

Ljung, G. M. & G. E. P. Box (1978) On a Measure of Lack of Fit in Time Series Models. *Biometrika,* 65.

Lustman, P. J., M. M. Williams, G. S. Sayuk, B. D. Nix & R. E. Clouse (2007) Factors influencing glycemic control in type 2 diabetes during acute- and maintenance-phase treatment of major depressive disorder with bupropion. *Diabetes Care,* 30, 459-66.

Lymberis, A. (2004) Research and development of smart wearable health applications: the challenge ahead. *Studies in Health Technology and Informatics,* 108, 155-61.

Lytle, L. A. (2009) Measuring the food environment: State of the science. *American Journal of Preventive Medicine,* 36, S134-S144.

Marcus, R. L., S. Smith, G. Morrell, O. Addison, L. Dibble, D. Wahoff-Stice & P. LaStayo (2008) Comparison of Combined Aerobic and High-Force Eccentric Resistance Exercise With Aerobic Exercise Only for People With Type 2 Diabetes Mellitus. *Jounral of the American Physical Therapy Association,* 88, 1345-1354.

Meyfroidt, G., D. M. Keenan, X. Wang, P. J. Wouters, J. D. Veldhuis & G. Van den Berghe (2010) Dynamic characteristics of blood glucose time series during the course of critical illness: effects of intensive insulin therapy and relative association with mortality. *Critical Care Medicine,* 38, 1021-9.

Miyazaki, M., T. Ohyanagi, L. Liu, B. Dobbs, M. Rowe, S. Sutphen, J. Miller & A. Cook (2009) Mobile ICT support for the continuum of care. *Studies in Health Technology and Informatics,* 143, 241-7.

Montgomery, D. C., C. L. Jennings & M. Kulahci. 2008. *Introduction to time series analysis and forecasting.* Hoboken, N.J.: Wiley-Interscience.

Moudon, A. V. & C. Lee (2003) Walking and biking: An Evaluation of environmental audit instruments. *American Journal of Health Promotion* 18, 21-37.

Mshelia, D. S., O. M. Akinosun & F. M. Abbiyesuku (2007) Effect of increased patient-physician contact time and health education in achieving diabetes mellitus management objectives in a resource-poor environment. *Singapore Medical Journal,* 48, 74-9.

Murata, G. H., J. H. Shah, W. C. Duckworth, C. S. Wendel, M. J. Mohler & R. M. Hoffman (2004) Food frequency questionnaire results correlate with metabolic control in insulin-treated veterans with type 2 diabetes: the Diabetes Outcomes in Veterans Study. *Journal of the American Dietetic Association,* 104, 1816-26.

Novaco, R. W., D. Stokols, and Milanesi L. (1990) Objective and subjective dimensions of travel impedance as determinants of commuting stress. *American Journal of Community Psychology,* 18, 231-257.

Novaco, R. W., D. Stokols, J. Campbell, and J. Stokols (1979) Transportation, Stress, and Community Psychology. *American Journal of Community Psychology,* 7.

Otto, C., A. Milenkovic, C. Sanders & E. Jovanov (2006) System architecture of a wireless body area sensor network for ubiquitous health monitoring. *Journal of Mobile Multimedia,* 1, 307-326.

Pandian, P. S., K. Mohanavelu, K. P. Safeer, T. M. Kotresh, D. T. Shakunthala, P. Gopal & V. C. Padaki (2008) Smart Vest: wearable multi-parameter remote physiological monitoring system. *Medical Engineering and Physics,* 30, 466-77.

Pankratz. 1983. *Forecasting with univariate Box-Jenkins Models: Concepts and Cases.* New York: John Wiley and Sons.

Pankratz, A. 1991. *Forecasting with dynamic regression models.* New York: John Wiley and Sons.

Pantelopoulos, A. & N. Bourbakis (2010) Prognosis—A Wearable Health-Monitoring System for People at Risk: Methodology and Modeling *IEEE Transactions on Information Technology in Biomedicine,* 14.

Papas, M. A., A. J. Alberg, R. Ewing, K. J. Helzlsour, T. L. Gary & A. C. Klassen (2007) The Built Environment and Obesity. *Epidemology Review* 29, 129-143.

Papinski, D., D. Scott & S. T. Doherty (2009) Exploring the route choice decision-makingcomparison of planned and observed routes obtained using person-based GPS. *Transportation Research Part F: Traffic Psychology and Behaviour,* 12, 347-358.

Parkin, C. G. & J. A. Davidson (2009) Value of self-monitoring blood glucose pattern analysis in improving diabetes outcomes. *Journal of Diabetes Science and Technology,* 3, 500-8.

Parkka, J., M. Ermes, P. Korpipaa, J. A. Mantyjarvi, J. A. Peltola & I. A. Korhonen (2006) Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine,* 10, 119-128.

Poolsup, N., N. Suksomboon & S. Rattanasookchit (2009) Meta-analysis of the benefits of self-monitoring of blood glucose on glycemic control in type 2 diabetes patients: an update. *Diabetes Technology and Therapeutics,* 11, 775-84.

Quenouille, M. H. (1949) Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society Series B,* 11.

Rainham, D., D. Krewski, I. McDowell, M. Sawada & B. Liekens (2008) Development of a wearable global positioning system for place and health research. *International Journal of Health Geographics,* 7, 59.

Riddle, M. S., D. L. Blazes, M. R. Wallace & S. Tasker (2003) The effect of driving distance from medical treatment facility and short-term management outcomes in Department of Defense beneficiaries newly diagnosed with human immunodeficiency virus. *Military Medicine,* 168, 96-100.

Saalasti, S. 2003. Neural Networks for Heart Rate Time Series Analysis. In *Faculty of Information Technology,* 194. Jyvaskyla: University of Jyvaskyla.

Saarloos, D., J. E. Kim & H. Timmermans (2009) The built environment and health: introducing individual space-time behavior. *International Journal of Environmental Research and Public Health,* 6, 1724-43.

Sagburg, F. (2006) Driver health and crash involvement: A case-control study. *Accident Analysis and Prevention,* 38.

Sallis, J. F., A. Bauman & M. Pratt (1998) Environmental and policy interventions to promote physical activity. *American Journal of Preventive Medicine,* 15, 379-397.

Sallis, J. F., R. B. Cervero, W. Ascher, K. A. Henderson, M. K. Kraft & J. Kerr (2006) An ecological approach to creating active living communities. *Annual Review of Public Health,* 27, 297-322.

Sanchez, M. J. & D. Pena (2003) The Identification of Multiple Outliers in ARIMA Models. *Communications in Statistics - Theory and Methods,* 32.

Sardinha, L. B., L. B. Andersen, S. A. Anderssen, A. L. Quiterio, R. Ornelas, K. Froberg, C. J. Riddoch & U. Ekelund (2008) Objectively measured time spent sedentary is associated with insulin resistance independent of overall and central body fat in 9- to 10-year-old Portuguese children. *Diabetes Care,* 31, 569-75.

Sargrad, K. R., C. Homko, M. Mozzoli & G. Boden (2005) Effect of high protein vs high carbohydrate intake on insulin sensitivity, body weight, hemoglobin A1c, and blood pressure in patients with type 2 diabetes mellitus. *Journal of the American Dietetic Association,* 105, 573-80.

Schwedes, U., M. Siebolds & G. Mertes (2002) Meal-related structured self-monitoring of blood glucose: effect on diabetes control in non-insulin-treated type 2 diabetic patients. *Diabetes Care,* 25, 1928-32.

Sobal, J. 2002. Social and Cultural Influences on Obesity In *International Textbook of Obesity,* ed. P. Bjorntorp, 305-322. Hoboken, NJ, USA: Wiley.

Souto, D. L. & E. L. Rosado (2010) Use of carb counting in the dietary treatment of diabetes mellitus. *Nutricion Hospitalaria* 25, 18-25.

SPSS, I. 2004. *SPSS Trends 13.0.* Chicago, IL: SPSS Inc.

Stokols, D., R.W. Novaco, J. Stokols, and J Campbell (1978) Traffic Congestion, Type A Behavior, and Stress. *Journal of Applied Psychology,* 63, 467-480.

Strauss, K., C. D. MacLean, A. R. Troy & B. Littenberg (2006) Driving distance as a barrier to glycemic control in diabetes. *Journal of Genetic Internal Medicine,* 21.

Takeuchi, H., N. Kodama & K. Tsurumi (2009) Time-series data analysis of blood-sugar level of a diabetic in relationship to lifestyle events. *IEEE Engineering in Medicine and Biology Society,* 2009, 5195-8.

Tarnus, E. & E. Bourdon (2008) Exploring the glycemic response to food intake with undergraduate students at the University of La Réunion. *Advances in Physiology Education,* 32, 161-164.

Timmermans, H. & R. G. Golledge (1990) Applications of behavioural research on spatial problems II: preference and choice. *Progress in Human Geography,* 14, 311-354.

Tobler, W. R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography,* 46: Supplemental Proceedings. International Union. Commision on Quantitative Methods

Towfigh, A., M. Romanova, J. E. Weinreb, B. Munjas, M. J. Suttorp, A. Zhou & P. G. Shekelle (2008) Self-monitoring of blood glucose levels in patients with type 2 diabetes mellitus not taking insulin: a meta-analysis. *American Journal of Managed Care,* 14, 468-75.

Wallace, A. J., S. L. Eady, R. S. Scott, J. A. Willis & C. A. Frampton (2008) Considerable temporal variability in glucose reference curves in humans for a year period. *Nutrition Research,* 28

Wearden, A. J., K. Hynd, H. Smith, R. Davies & N. Tarrier (2006) Idiosyncratic personal explanations for blood glucose events are associated with poorer self-management and glycemic control in adult patients with long-standing type 1 diabetes. *Diabetes Care,* 29, 1208-13.

Wei, W. W. S. 1990. *Time Series Analysis: Univariate and Multivariate Methods.* Redwood City, California, USA: Addison-Wesley Publishing Company.

Welk, G. J. 2002. Use of accelerometry-based activity monitors to assess physical activity. In *Physical Activity Assessments for Health-related Research,* eds.

G. J. Welk & I. L. Champaign, 125-141. Champaign, IL: Human Kinetics Publishers.

Welschen, L. M., E. Bloemendal, G. Nijpels, J. M. Dekker, R. J. Heine, W. A. Stalman & L. M. Bouter (2005) Self-monitoring of blood glucose in patients with type 2 diabetes who are not using insulin. *Cochrane Database Syst Rev*, CD005060.

Wen, L., M. L. Parchman, W. D. Linn & S. Lee (2004) Association between self-monitoring of blood glucose and glycemic control in patients with type 2 diabetes mellitus. *American Journal of Health-System Pharmacy,* 61, 2401-5.

WHO. 2007. Tackling obesity by creating health residential environments. World Health Organization, Regional Office for Europe: Copenhagen, Denmark.

Wiesli, P., C. Schmid, O. Kerwer, C. Nigg-Koch, R. Klaghofer, B. Seifert, G. A. Spinas & K. Schwegler (2005) Acute psychological stress affects glucose concentrations in patients with type 1 diabetes following food intake but not in the fasting state. *Diabetes Care,* 28, 1910-5.

Yaffee, R. & M. McGee. 2000. *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS.* Orlando, Florida: Academic Press.

Yang, C. C. & Y. L. Hsu (2009) Development of a wearable motion detector for telemonitoring and real-time identification of physical activity. *Telemedicine Journal and E-health* 15, 62-72.