1-1-2013

Mathematical Modeling and Analysis of Asthma Stability and Severity

Arezoo Hanifi
University of Denver

Mathematical Modeling and Analysis of Asthma Stability and Severity

---

A Dissertation

Presented to

The Faculty of Engineering and Computer Science

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Arezoo Hanifi

March 2013

Advisor: Dr. Mohammad A. Matin

Author: Arezoo Hanifi
Title: Mathematical Modeling and Analysis of Asthma Stability and Severity
Advisor: Dr. Mohammad A. Matin
Degree Date: March 2013

Abstract

Asthma is one of the most common chronic conditions in the United States. Asthma affects about one in fifteen people. It affects children more than adults and blacks more than whites. People with asthma experience attacks of wheezing, breathlessness, chest tightness, and coughing. Asthma can be fatal and the costs for the disease (direct and indirect) are approximated to be tens of billions of dollars each year.

There is no cure for asthma. However; for most people if asthma is controlled well they can lead normal, active lives. Therefore asthma controllability is a main factor in clinical practice. In order to control asthma, the disease has to be completely understood. Asthma is very heterogeneous and this makes the exact diagnosis and control procedures difficult. To better evaluate and study asthma, mathematical tools can be very beneficial.

In this study we first develop a complete system for lung impedance analysis of laboratory models of asthma. Our designed system is capable of precisely diagnosing the diseased models and predicting the severity of their condition. We also evaluate the treatment progress in mouse models of asthma. We then study an asthma database of humans including measurements of four related laboratory parameters and cluster patients based on inherent properties of the study variables. This mathematical approach

clustered patients with specific characteristics and segregated the unstable asthmatic patients in a single group. Our method is very promising in predicting the instability of asthma, which is highly correlated with frequent asthma attacks and increased utilization of care.

Acknowledgements

I would like to thank my advisor Dr. Mohammad Matin for making my Ph.D. experience productive. I would like to thank my co-advisor Dr. Rafeul Alam for his trust and believe in me. Your love and motivation for true research despite any hardship has been the most amazing source of inspiration for me. Thanks for supporting me as a Research Assistant and for being there for me at all difficult times, whether academically or related to my personal life. I am also very thankful to his lab members, Shaila, Nick, JD, Dipa, Kavita, Qiaoling and Magda .

I would like to express my sincere gratitude to Dr. Roger E. Salterswho passed away a year ago. His deep mathematical understanding of biological systems was remarkable. May his soul rest in peace.

I would like to thank my wonderful husband, Dr. Amir Babazadeh for his unconditional love and support during the difficulties of the graduate studies. I am extremely grateful to him for technical discussions and his very helpful guidances, for reviewing my work drafts and making great comments and always redirecting me to a better path. I would like to devote my very special thanks to my first teachers in life, Vida Asgari and RezaAli Hanifi. Thanks for taking care of Arman and Armin for me and for all the sacrifices you and my dear brothers, Omid and Hamid made for me to make my educational dreams come true. Thank you for all your love, support and inspiration.

Table of Contents

Chapter 1 1
INTRODUCTION 1
1. 1 Background 1
1.2 Problem Statement 2
1.3 Objective 3
1.4 Scope of Research Work 4
1.5 Methodology 4
1.6 Organization of the Dissertation 5

Chapter 2 6
DATA MINING 6
2.1 Background 6
2.2 Data Mining Tasks 7
2.3. Medical Applications of Data Mining 9

Chapter 3 12
MOUSE RESPIRATORY IMPEDANCE 12
3.1 Introduction and Problem Statement 12
3.2 Forced Oscillation Technique and Constant Phase Model 18
3.3 Experimental Procedures and Data Acquisition 22
    3.3.1 Experiment 1 22
    3.3.2 Experiment 2 25
3.4 Methods 26
    3.4.1 Linear Parametric Identification of Lung Input Impedance 26
    3.4.2 Semi- Impulse Response of the Respiratory System 32

Chapter 4 35
HUMAN DATABASE 35
4.1 Background 35
4.2 Problem Statement 38
4.3 Parameter Selection 40
4.4 Methods 44
    4.4.1 Principal Component Analysis 45

4.4.2 Hierarchical Clustering 49
4.4.3 Self-Organizing Maps 50

Chapter 5 53
RESULTS AND DISCUSSIONS 53
5.1 Mouse Lung Impedance 53
5.1.1 Experiment 1: Model Development 54
5.1.2 Experiment 2: Model Validation 65
5.1.3 Discussion 69
5.2 Human Database 69
5.2.1 PCA 69
5.2.2 Hierarchical Clustering 73
5.2.3 SOM Results 73
5.2.4 Scoring System 75
5.2.5 Discussion 82

Chapter 6 87
CONCLUSIONS and FUTURE WORK 87

Bibliography 91

Appendix A 98

Publication list 98

Appendix B 99

List of Abbreviations

ACT: Asthma control test

CPM: Constant Phase Model

FENO: Fractional exhaled nitric oxide

FEV1: Forced expiration volume in 1 second

MAPK: Mitogen-activated protein kinase

PCA: Principal component analysis

PC20: Provocation concentration needed to induce a 20% fall in FEV1

SOMs: Self-organizing maps

List of Tables

Table V.I Linear parametric models (5.1) coefficients for Normal, Acute, Tolerant and Chronic Asthma models…….. 55

TableV.II. Coefficient matrix from Principal Component Analysis…… 71

Table V.III Asthma instability scoring system……..... 76

Table V.IV Asthma instability scores by clusters…… 77

Table V.V Comparison of baseline characteristics of the five asthma clusters…… 81

Table V.VI Demographics of the population study…….. 82

List of Figures

Figure 3.1 The simplest model of lung is an elastic balloon at the end of a rigid pipe. The balloon represents the tissues and the pipe represents the airways……15

Figure 3.2 (a) Real portion of mouse lung impedance and (b) Imaginary portion of mouse lung input impedance for the four models developed in the laboratory in experiment 1……17

Figure 3.3 The experimental set up for measuring the mouse lung impedance in the lab……...19

Figure 3.4 A simplified block diagram of forced oscillation maneuvers by flexivent…..20

Figure 3.5 A sample of the broad band flow input to the mouse lung……..21

Figure 3.6 General linear model structure…….27

Figure 3.7 Lung system identification……...28

Figure 3.8 Comparison of the FIT function for OE and ARX structures…….32

Figure 3.9 Realistic versus ideal impulse response……...33

Figure 3.10 The *semi-impulse* waveform……..34

Figure 4.1 A patient is blowing into a spirometer under supervision……41

Figure 4.2 Histograms of the four parameters for the (a) male patients, (b) female patients, (c) whole population……....44

Figure 5.1: Bode plots of the real data (R *) and estimated transfer function (G --) for chronic asthma model……56

Figure 5.2: Bode plots of the real data (R *) and estimated transfer function (G --) for Acute asthma model……..57

Figure 5.3: Bode plots of the real data (R *) and estimated transfer function(G --) for Tolerant asthma model……..58

Figure 5.4: Bode plots of the real data (R *) and estimated transfer function (G --) for Normal model……....59

Figure 5.5 the result of simulation of a) Chronic b) Acute c) Tolerant d) Normal with semi-impulse waveform at the baseline…….60

Figure 5.6 The result of simulation of a) Chronic b) Acute c) Tolerant d) Normal with semi-impulse waveform at dose 3.125 mg/ml of methacholine……60

Figure 5.7 The result of simulation of a) Chronic b) Acute c) Tolerant d) Normal with semi-impulse waveform at dose 12.5 mg/ml of methacholine……..61

Figure 5.8: the result of simulation of a) Chronic b) Acute c) Tolerant d) Normal models with semi-impulse waveform at dose 25 mg/ml of methacholine…….61

Figure 5.9 Semi-impulse responses for the Chronic asthma model. ……63

Figure 5.10 Semi-impulse responses for the Tolerant asthma model……63

Figure 5.11 Semi-impulse responses for the Normal model……..64

Figure 5.12 Semi-impulse responses for the Acute asthma model……64

Figure 5.13 The 'ap' for each mouse model dose-wise a) Chronic, b) Acute, c) Tolerant, d) Normal……...66

Figure 5.14 Semi-impulse response for the mouse models from experiment 2, a) Normal; b) Chronic c) MEK inhibitor- treated at baseline……..67

Figure 5.15 Semi-impulse response for the mice groups from experiment 2, a) Normal; b) chronic c) MEK inhibitor- treated at dose 50 mg/ml…….67

Figure 5.16 Scores in the space of the first three principal components……..71

Figure 5.17 Variance vector……..72

Figure 5.18 Dendrogram generated using Ward's minimum variance algorithm…….73

Figure 5.19 clusters of patients generated by SOMs……74

Figure 5.20 The distribution of FEV1 in each cluster……...77

Figure 5.21 The distribution of PC20 for methacholine in each cluster……77

Figure 5.22 Total serum IgE distribution for each cluster…….78

Figure 5.23 Blood eosinophil count in each cluster……...78

Chapter 1

INTRODUCTION

1. 1 Background

Asthma is a lifelong disease with several symptoms such as wheezing, breathlessness, chest tightness, and coughing. Asthma limits the patient's life quality and activities Asthma continues to be a serious and common public health problem. The Centers for Disease Control and Prevention (CDC) statistics show that 25.7 million people, including approximately 7.1 million children, have asthma. Asthma; accounts for over fifteen million physician office, hospital outpatient department visits and nearly two million emergency room visits every year. Asthma is one of the most common serious chronic diseases of childhood and the third-ranking reason for young children hospitalization (EPA, http://www.epa.gov/asthma/pdfs/asthma_fact_sheet_en.pdf).

Asthma affects about 5% of adults and about 10% of children. Half of the people with asthma develop it before age 10 and most develop it before age 30. The number of people diagnosed with asthma grew by 4.3 million from 2001 to 2009 among all ages, races, and gender groups; however, it is slightly more common in blacks and Hispanics than in whites. Asthma was linked to 3,447 deaths (about nine per day) in 2007(http://www.cdc.gov/VitalSigns/pdf/2011-05-vitalsigns.pdf).

Asthma is very heterogeneous and can be difficult to diagnose, as its symptoms are sometimes similar to other conditions such as allergic rhinitis, lung infection, and even cardiac problems.

1.2 Problem Statement

Asthma specialists are trying hard to prevent, control and treat asthma with medications and teach patients how to avoid triggers of asthma attacks. In this process and by the help of new technologies they are generating large databases of their patients' clinical information on a daily basis. There is something equally significant that mathematicians and biomedical engineers could do to shed more light on asthma unknowns. Studying these databases and establishing mathematical models that are simple and practical is of great importance and can contribute significantly to the areas of asthma diagnosis, asthma categorization, treatment evaluation and prediction of a patient's future condition. As one of the strong mathematical tools, modeling approaches have been widely utilized in studying the respiratory system in both human and animal models.

Asthma is a disease that affects the lung and immune system. In order to perform complete study of this disease, we have to take both aspects into considerations.

In animal models, respiratory system impedance (as a representation of lung function) has been extensively studied and there are well established models capable of

differentiating between airway and tissue mechanics in the lung. We propose a simple linear model capable of investigating the severity of lung conditions and also evaluating the treatment progression. We confirm the effectiveness of the analysis technique with experimental findings.

In humans, despite the existence of several clustering studies of asthma (parameters that are related to the lung function and the immune system response), an appropriate clustering methodology capable of reflecting heterogeneity of asthma and asthma instability is missing. We propose a new model that predicts asthma instability probability based on very well known asthma related clinical parameters.

1.3 Objective

This research focuses on the mining of data and modeling of asthma in general. Specifically, we tackle asthma modeling from two aspects of asthma. As mentioned, asthma affects the lung function and mechanics and also involves the immune system. In the first modeling, we model the mechanical effects of asthma in mouse models and in the second analysis, we model both aspects by studying a database of human asthma. The datasets were both collected at National Jewish Health (NJH), Denver, Colorado.

The mouse database was produced by Nicholas Goplen at Dr. Alam's lab at NJH. The human database includes measurements from NJH clinic patients.

For the mouse lung impedance study we develop a linear parametric model. The objective is to better define asthma severity through the use of lung mechanics.

In our second modeling, we apply several mathematical approaches including principal component analysis (PCA) and self-organizing maps (SOM) to cluster our patients into groups that represent them by their stability status of the disease.

1.4 Scope of Research Work

The focus of this research is on exploring and applying advanced mathematical analysis and data mining techniques to asthma databases. The goal is to have a thorough and multi-dimensional analysis of asthma modeling and clustering that would benefit the patients and physicians in understanding asthma and its stages and categories by the use of the currently available technologies and laboratory data.

1.5 Methodology

To achieve our objectives we apply several data mining approaches to our asthma databases including:

- identification approaches more specifically, linear parametric modeling and impulse response simulation;

- clustering methodologies such as PCA, SOM, and hierarchical clustering;

- statistical methodologies such as correlation analysis, chi square test, and ANOVA.

1.6 Organization of the Dissertation

Chapter 2 of this dissertation introduces data mining approaches. Major tasks of data mining are explained and the roots of data mining are explored. Following Chapter 2, the rest of the dissertation is divided into two parts. The first part presents mouse lung impedance analysis. The second part is on the analysis of the human asthma database. Chapter 3 provides the details of the mouse lung impedance study. It discusses the techniques we apply to model the impedance in detail. Chapter 4 examines the human asthma database and introduces the methodologies applied in this study. The results of the analyses introduced in Chapters 3 and 4 are provided in Chapter 5.

In Chapter 6 we provide the discussion and conclusion of our two models and provide our suggestions for future work.

Chapter 2

DATA MINING

2.1 Background

In recent years advances in medical technologies have resulted in the generation of extensive amounts of data which requires extraction and analysis for clinical use. Traditional manual data analysis is currently inadequate and new analysis methods are required. Due to this demand a new branch of science has developed called data mining.

Han et al(2001), Hand et al. (2001) and Roiger et al. (2003) define data mining as extracting or mining knowledge from large database or more briefly as knowledge discovery in databases (KDD).

Medical data mining is the process of applying mathematical methodologies to discover information from data that is then convertible to medical knowledge applicable for medical doctors.

In the mining of data, one iteratively searches for new information in a database. The major goals of data mining are prediction and description (Kantardzic, 2011).

In predictive data mining, one uses some variables of a data set to predict unknown or future values of variables of interest. This type of data mining generates a model of the system. In descriptive data mining, one tries to find patters in the dataset to make the data interpretable to doctors or others. This type of data mining generates new knowledge for the medical society (Kantardzic, 2011).

2.2 Data Mining Tasks

The primary data mining tasks include:

Classification – the discovery of a predictive learning function that maps (classifies) a data item into one of several predefined classes (Weiss and Kulikowski, 1991). Among classification techniques are the Neural networks (NNs), probability approaches, Naive Bayes and Adaptive Bayes Network supporting decision trees. There are two steps in classification. In the first step we build a classification model which contains a set of classes. In the second step, the model is used for prediction of new or future data. Han et al. (2001) defined classification as a model used for describing a pre-specified set of data classes. Roiger et al. (2003) defined classification as a technique with categorical output variable. The common point in all the definitions is the "building of a model for assigning data to categorical classes". (Kantardzic, 2011).

Regression/prediction – the discovery of a predictive learning function that maps a data item to a real-value prediction variable (Kantardzic, 2011).

Clustering - a descriptive task in which one uses unsupervised learning (unlike classification), such as associations and clustering algorithms, to find appropriate groupings of elements for a dataset without any a priori assumption.

A cluster is therefore a collection of objects which are "similar" but are "dissimilar" to the objects belonging to other clusters (Tayal and Raghuwanshi, 2010).

According to Berger (2004) and Berry et al. (2000) clustering is the segmenting of a diverse group into a number of more similar subgroups or clusters. Contrary to classification, clustering does not rely on predefined classes. (Mdzingwa, 2005).

Summarization - a descriptive task to find a compact description for a dataset (Kantardzic, 2011).

Dependency Modeling – finding a local model capable of describing the significant dependencies between variables in a dataset (Kantardzic, 2011).

Change and Deviation Detection – includes discovering the most significant changes in a dataset (Kantardzic, 2011).

Because of the roots of data mining in statistics, machine learning and control theory, there is emphasis on models and algorithms. The problem of determining a mathematical model for an unknown system by observing its input-output data is called system identification.

System identification generally involves two steps (Kantardzic, 2011):

1. *Structure identification* - In the first step, a class of models that is most suitable to our system is determined. This class of models is denoted by a parameterized function y = *f(u,t),* where y is the model's output, *u* is the input vector, and *t* is the parameter vector. The function *f* depends on the problem, and the function is related to the designer's expertise and the natural laws governing the system.

2. *Parameter identification* - In the second step, the structure of the model is known and we need to seek for the optimal parameters, such that the newly defined model can appropriately and precisely (based on an error criteria) describe the system.

In general, system identification, including both structure and parameter identification, needs to be done repeatedly until a satisfactory model is found.

2.3. Medical Applications of Data Mining

Raw medical data are voluminous and heterogeneous. Medical data is collected through several different means such as diagnostic tests, laboratory data, images, interviews with the patient, and the physician’s observations and interpretations. Every day gigabytes of data are generated.

Review of literatures shows that various fields of medical sciences have benefited from data mining and many medical breakthroughs have occurred by the application of this analytical method. Data mining has the potential to be applied to healthcare for objectives such as diagnosis, treatment progression evaluation, management of healthcare, and predictive medicine.

Below are some of the effective medical studies that have applied data mining approaches (Khajehei and Etemady, 2010).

- (Li et al., 2004) used data mining techniques for the detection and diagnosis of cancer. They applied genetic algorithm based methods for feature selection and were able to differentiate those with ovarian cancer from healthy subjects.

- (Tiffin et al., 2005) described a bioinformatics approach that selects candidate disease genes according to their expression profiles. They successfully selected the known disease gene for fifteen out of seventeen diseases and reduced the candidate gene set to 63.3% (±18.8%) of its original size. They declared that the data mining approach facilitated direct association between gene expression data and the disease phenotype, and successfully prioritized candidate genes according to their expression in disease-affected tissues.

- (Coulter et al., 2001) used data mining (Bayesian method) to study the relation between antipsychotic drugs and cardiac disease and showed that data mining was capable of finding relationships and identifying the potentially risky drugs.

- (Breault et al., 2002) used the classification tree approach and regression trees with a binary target to study a diabetes database. They declared that data mining could discover novel associations that would be very useful to clinicians and administrators.

- (Kusiak et al., 2005) used data preprocessing, data transformations, and a data mining approach to study patients' survival on hemodialysis. These rules were used by a decision-making algorithm, which predicted survival of new unseen patients. Important parameters identified by data mining were interpreted for their medical significance. They showed their analyses were useful for survival prediction of dialysis patients.

In Chapter 3 we will take advantage of identification for modeling the lung impedance in mice. In Chapter 4 we will utilize data clustering.

Chapter 3

MOUSE RESPIRATORY IMPEDANCE

3.1 Introduction and Problem Statement

Breathing is a spontaneous mechanical process. Thorax and abdomen muscles work under the control of the brain and produce the pressures required to expand the lung to suck the air in. These pressures must overcome the lung and the chest wall tendency to recoil. Pressure also drives air along the pulmonary airways beginning at the mouth and ending deep in the lungs at the point where air and blood exchange oxygen and carbon dioxide. The mechanical properties of the lung determine the relation between lung volume, muscular pressures and airway flow. In other words, it determines how much effort is needed to take in a breath and how comfortable it is to breathe. Therefore, these properties have an important bearing on how we experience our daily lives (Bates, 2009).

Assessment of respiratory mechanics is the process of uncovering relationships between pressures, flows, and volumes measured at appropriate sites. Appropriate assessment of the lung mechanics can play a crucial part in disease diagnosis and evaluation of the treatment progression (Bates, 2009).

Our knowledge of lung mechanical function is completely dependent on what we can measure. A great deal is known about lung mechanical function, due to the ongoing efforts of scientists beginning in the late 1800s. In the beginning, lung mechanics was largely of interest to physiologists and physicians. Over the past several decades however,the field has become highly quantitative by the help of electronic sensors and digital computers. These devices have provided accurate experimental data related to lung function which is very appealing to scientists equipped with sophisticated methods of data analysis. This means that the lung mechanic field is now attracting the attention of biomedical engineers, physicists, and mathematicians (Bates, 2009).

For the purpose of assessing the lung system we need to use the system identification (inverse modeling) approach because the lung system is not known a priori and the approach needs to be built based on input output data from measured experiments with the help of insights gained through previous research.

The structure of an inverse model has to correspond to the structure of the real system. This means that when the model mimics the behavior of the original system, this similarity in behavior has to be related to the system's internal mechanisms responsible for that behavior. Determination of model structure is much related to the experience and knowledge of the system modeler. Modeling is a dynamic process. Models of complex systems, such as the lung, are constantly being tested and improved in the light of new knowledge and new data(Bates, 2009).

Assuming the lung to be a linear dynamic system, the mechanical properties of the lung can be encapsulated in terms of its input impedance. Conventionally, the input impedance is achieved by measurements of pressure and flow made at certain sites. Input impedance is then interpreted using an appropriate mathematical model. Animal models are frequently used to measure input impedance. Among animal species, mice are more often used because of the advantages they provide (Tu et al., 1995).

Extensive research has been done over the past thirty years to find appropriate mathematical models for the interpretation of the input impedance of the mouse lung. Several experts from different fields have applied different methodologies. Some have proposed complicated nonlinear models of the lung (Tomalak et al., 1993) or just linear models (Diong et al. 2009). Some people have used fractional approach (Ionescu et al. 2011), or recursive least squares method (Lauzon and Bates 1991), or modeling with an electrical circuit (Baswa et al. 2005, Bates and Allen 2006, Goldman et al. 2010). These models are all appropriate models from a specific aspect.

Mathematical models do not need to be complicated in order to be useful. In fact, inverse models are rather simple as the number of adjustable parameters is limited. Due to this fact, the lung models achieved so far are not capable of encapsulating everything we know about the organ. Yet, they are still capable of mimicking many of the details of its global behavior. An example of a very simple yet acceptable model of lung mechanics is an elastic balloon sealed over a rigid pipe. The balloon represents the expandable lung

tissues and the pipe is the representative of the pulmonary airways (Fig. 3.1). A real lung is a lot more complicated than this, even though it still embodies much that is key to the ventilation process (Bates, 2009).

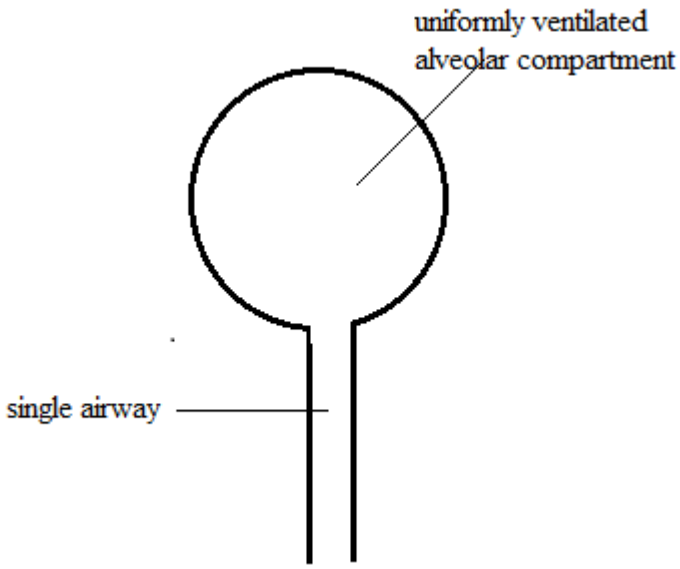


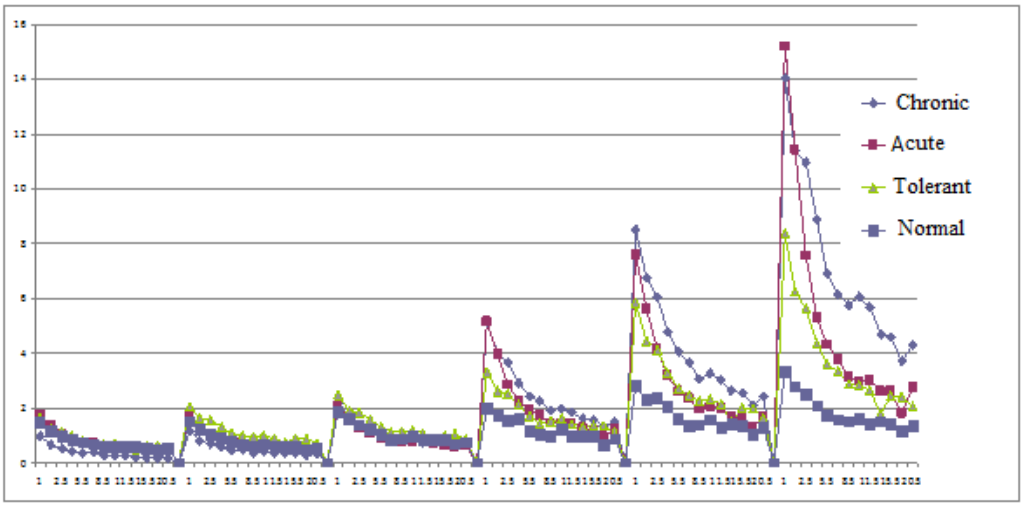
Figure 3.1 The simplest model of lung is an elastic balloon at the end of a rigid pipe. The balloon represents the tissues and the pipe represents the airways.

In general, an inverse model should always be considered as a work in progress, and lung mechanic identification is not an exception. Any model, no matter how successful will have some shortcomings. The ideal outcome is that our model could serve adequately for a particular purpose. Very accurate fits to measured impedance spectra below 20 Hz in a variety of species, including the mouse, have been achieved with the so-called constant-phase model (CPM) of impedance (Hantos et al., 1992). CPM has been widely applied in the frequency domain analysis but it is very difficult: 1) to analyze the respiratory system in the time domain (Moriya et al., 2003); and 2) to

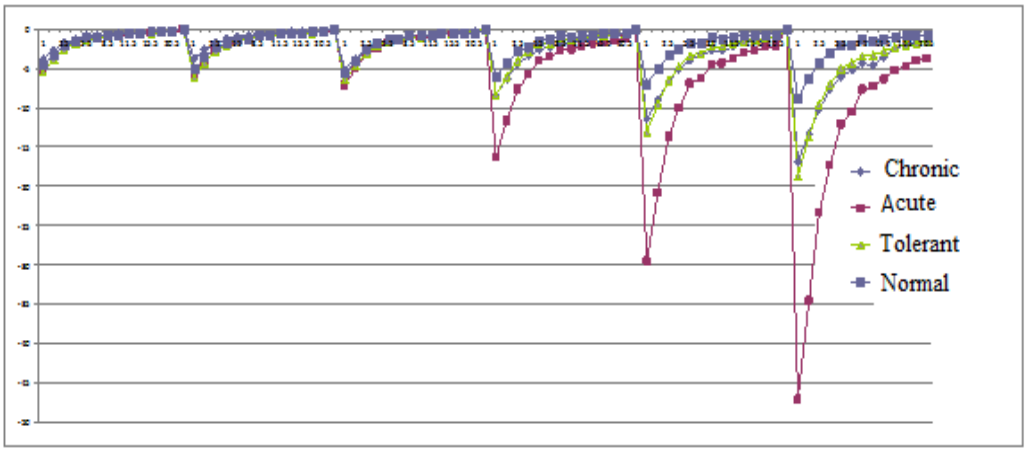
compare the qualitative behavior of the lung for disease diagnosis and evaluation of the treatment progression.

In Figure 3.2 the real and imaginary portions of the mouse respiratory impedance data for laboratory models of asthma (Goplen et al., 2009) in frequency domain are shown. In this research, the goal is to analyze the data from these mouse models of asthma and establish a simple but practical technique for evaluation of disease severity as well as treatment progression. We want our technique to be applicable in time domain analysis and to human data.

We also want to be able to simulate different inputs of interest through our model rather than do the experiment in reality, to find the response of the lung system to that input. For this purpose, a linear-parametric model in discrete frequency domain is established based on pole-zero approximation of the data. This model enables us to predict the time response of the lung system to various test signals, and to study the dynamic changes developing in the lung caused by a disease or a treatment.



(a)



(b)

Figure 3.2 (a) Real portion of mouse lung impedance (Z_{in}) and (b) Imaginary portion of mouse lung input impedance (Z_{in}) for the four models developed in the laboratory in experiment 1.

In this chapter we will introduce the methods for modeling the lung impedance. In Chapter 5 we will provide the result of the analysis of different laboratory models of asthma in mice, and show that the technique is capable of differentiating between

different models of asthma as well as stages of treatment. This established technique has the potential for application to human lung impedance.

3.2 Forced Oscillation Technique and Constant Phase Model

There are several invasive and non-invasive techniques for measuring the input impedance including the interrupter technique, impulse oscillometry, and forced oscillation technique (FOT) (Schuessler and Bates, 1995; Jablonski et al., 2011; Cavalcanti et al., 2006).

In impedance measurement there is a compromise between accuracy, noninvasiveness, and convenience. This means that the precision and the invasiveness of a method are correlated with each other. The less invasive the measurement technique applied, the less possibility of producing precise, reproducible data (Bates and Irvin, 2003)..

FOT is a general name for any approach that evaluates the breathing mechanics by superimposing small external pressure on the spontaneous breathing of the subject. FOT was developed over 50 years ago and is the subject of numerous studies (Peslin and Fredberg, 1986, Zwart and Woestijne, 1994). FOT requires minimum cooperation by the subject and no respiratory maneuvers. FOT is applied to the mice that have been anesthetized, paralyzed, and tracheotomized for measurement of their complex "lung input impedance Z(f)". The low-frequency input impedance Z(f) reflects the frequency

dependency of the airway and lung tissue compartments separately. Estimates of Z(f) achieved by FOT can be considered the most detailed measurements of pulmonary mechanics currently available (Glaab et al., 2007).

In Figure 3.3 the set up for mouse lung impedance measurement is shown. In a platform similar to Figure 3.3 in Dr. Alam's lab at National Jewish Health, all the input impedance data were collected by (Goplen et al., 2009).

Figure 3.3 The experimental set up for measuring the mouse lung impedance in the lab, Image source: http://phenome.jax.org/grpdoc/Berndt2/imgs/FlexiVent.jpg

The machine used for the impedance measurements is called Flexivent (Scireq, Montreal). A simple block diagram of the machine is shown in Figure 3.4. The flexivent is capable of producing:

- sine wave oscillations in the flow to the lungs to determine *R* and *E;*

- broad-band oscillations in the flow to the lungs to determine impedance *Z;*
- step changes in the volume to the lungs to determine the pressure-volume curve.

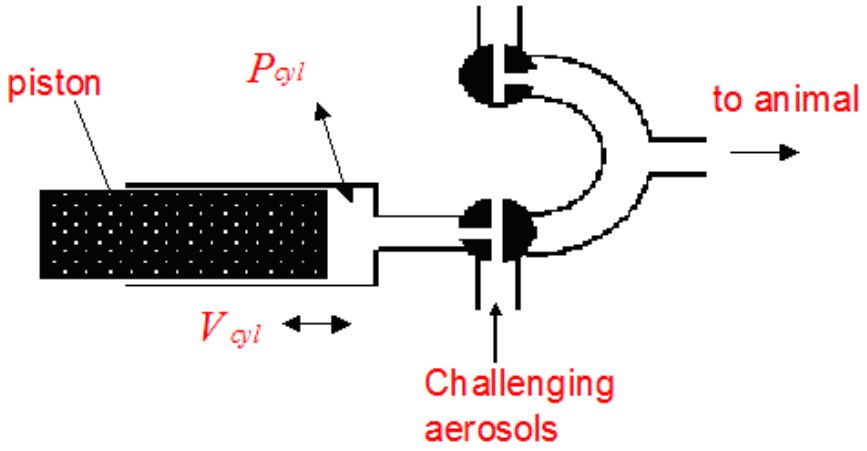


Figure 3.4 A simplified block diagram of forced oscillation maneuvers by flexivent, Image source: mbi.osu.edu/2006/tutmaterials/Mechanics%20Course_4.ppt

In FOT controlled broad-band perturbations (Figure 3.5) in flow ($\dot{V}$) are applied to the lungs via the trachea while the pressure (P) is measured. Consequently, the input impedance, (which is measured at the entrance to the lung) $Z(\omega)$, is achieved by the ratio

$$Z(\omega) = \frac{P(\omega)}{i\omega V(\omega)} \qquad (3.1)$$

where ω is angular frequency and $P(\omega)$ and $V(\omega)$ are the Fourier transforms of $P(t)$ and $V(t)$, respectively. $V(t)$ typically consists of a sum of discrete sinusoidal components spanning the frequency range of interest. The ratio in (3.1) is calculated between average cross-power and auto-power spectral densities determined by dividing the volume and flow data sets into a set of overlapping windows. The frequencies of the input sinusoids

are chosen to be mutually prime in order to reduce the harmonic distortion in $Z(\omega)$ that can arise as a result of system nonlinearities.

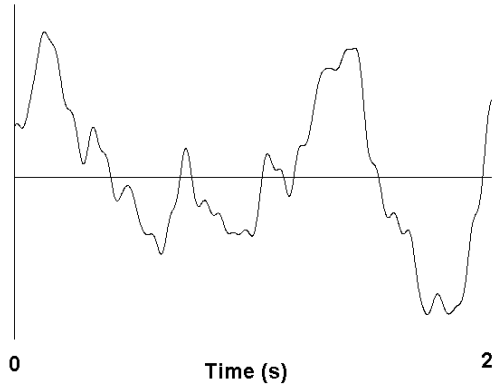


Figure 3.5 A sample of the broad band flow input to the mouse lung, Image source: mbi.osu.edu/2006/tutmaterials/Mechanics%20Course_4.ppt

Impedance data can then be analyzed by applying the Constant Phase Model. In this technique through assessment of oscillatory data from measurements of pulmonary mechanics airway and tissue mechanical components are identified. (Hantos et al., 1992). Constant-phase model of impedance is given by

$$Z(f) = R_N + i\omega I + \frac{G - iH}{\omega^{\alpha}} \tag{3.2}$$

and hence

$$\alpha = 2\pi \tan^{-1} \frac{H}{G} \tag{3.3}$$

R_N is a Newtonian resistance that has been shown to closely approximate resistance of the airway tree.

I is an inertance due to the mass of the gas in the central airways, and plays a

negligible role in Z for mice below 20 Hz. G and H are parameters that characterize, respectively, the elastic and dissipative properties of the lung tissue. The ratio of the real to the imaginary parts of the tissue component of Z (i.e. G/H) is independent of frequency, so tissue impedance has constant phase. CPM was briefly introduced as a fundamental model of the field but will not be utilized for our analysis. Next we will describe the process of acquiring the real data from mouse lung.

3.3 Experimental Procedures and Data Acquisition

Two sources of data were analyzed in this dissertation. Both are measurements of input impedance measured through FOT by use of Flexivent (Scireq, Montreal, Quebec). In order to demonstrate various applications of our method, data from Experiment 1 is used to compare the severity of the disease in a number of asthma models qualitatively, whereas the data from Experiment 2 is used for analyzing the treatment efficacy. The experimental procedures are described in detail in the following section.

3.3.1 Experiment 1

The development of chronic and tolerant models of asthma and their airway functional and histopathological findings in (Goplen et al., 2009) were reported previously. We used the data from those experimental models provided by Nicholas Goplen from Dr. Alam's lab for the mathematical analyses of this research. A brief outline of the models is given below.

- ***Saline control group****:* The development of asthma in mice requires subcutaneous sensitization with an allergen usually in conjunction with an adjuvant such as alum. The adjuvant boosts the immune response to the allergen. Following sensitization, the mice are exposed to the allergen intranasally and then examined for the presence of airway inflammation and hyperreactivity. As a negative control for these experiments, another group of mice is simultaneously immunized with saline in alum and then exposed to intranasal saline. The treatment of mice with saline in alum does not induce airway inflammation or alter airway mechanics. In our negative control experiments, we treated mice subcutaneously with saline in alum and challenged intranasally twice per week for eight weeks with saline. The mice were rested for 21 days and then used to measure airway hyperreactivity in response to increasing doses of methacholine (a bronchoconstrictor) inhalation using the Flexivent apparatus.

- ***Chronic asthma model***: This is a model of asthma that is severe in intensity and lasts weeks and months after the allergen exposure. We have reported that mice sensitized and then chronically exposed to a single allergen develop tolerance (see tolerant model). In contrast, mice sensitized and chronically exposed to multiple allergens resist tolerance and develop chronic asthma—i.e. sustained inflammation and airway hyperreactivity longer than three to four weeks after discontinuation of the allergen exposure. The specific approach for this protocol was as follows: mice were immunized with a combination of three allergens (dust mite, ragweed and Aspergillus extract) with

alum and challenged intranasally with the same allergens twice per week for eight weeks, rested for 21day and then subjected to the airway hyperreactivity measurement.

- ***Acute asthma model:*** This is a model of asthma that is severe in intensity but resolves in seven to ten days. We applied a commonly used protocol where mice develop acute asthma after sensitization and exposure to a single allergen. The model is called acute because airway inflammation and hyperreactivity resolve usually within seven to ten days after the last allergen exposure. The specific protocol is as follows: mice immunized with Alum plus Aspergillus allergen, challenged intranasally twice per week for two weeks, rested 72 hrs, and then subjected to the airway hyperreactivity measurement.

- ***Tolerant asthma model****:* This is a model where asthma develops during the initial phase of allergen exposure but then resolves despite further exposure to the allergen. As mentioned above, we induce tolerance by repeated exposure to a single sensitizing allergen. Mice are immunized with a single allergen, Aspergillus extract in alum, challenged intranasally with the allergen twice per week for eight weeks, rested 21days, and then used for airway hyperreactivity measurement. We studied six to ten mice per group for these experiments. The differences in airway inflammation and airway resistance to methacholine among the study groups were statistically significant (ANOVA and paired t test) as reported previously (Goplen et al., 2009).

3.3.2 Experiment 2

In this experiment, laboratory models of chronic asthma underwent a treatment.We are seeking the evaluation of treatment using the mathematical model of the lungs. The significance of this mathematical model is that sometimes in the laboratory models the treatment effects are not significant enough to be confirmed. However, a mathematical model is sensitive enough to differentiate between the treated groups and the untreated ones and to predict the efficiency of the treatment.

Various external factors and endogenous mediators activate cells through stimulation of the mitogen-activated protein kinase (MAPK) signaling pathway. A major member of the MAPK family is the extracellular signal-regulated kinase 1/2 (ERK1/2). The latter is activated by MEK1/2 (MAP ERK kinase 1/2). In order to determine if signaling from the MAPK pathway is important for induction of chronic asthma, we applied U0126, an inhibitor of MEK1/2. This inhibitor was given intranasally 10μg/mouse daily for five consecutive days in week 10, two weeks after the last allergen exposure. Dimethyl sulfoxide (15uL) was used as the vehicle for U0126 and was given intranasally to a control group of five mice. The airway hyperreactivity in response to methacholine was measured three days later. The difference in airway resistance between U0126- and vehicle-treated groups was statistically significant.

Complete dose-response studies with methacholine in Experiment 2 were done in the laboratory. The difference between the U0126- and vehicle-treated groups reached a

statistical difference only at the highest dose of mehacholine (50 mg/ml). For this reason we used FOT data for this dose in our analysis. The in-vivo experiments are complicated by the bioavailability of the medication. We expected to see a significant inhibition of airway resistance at the lower doses of methacholine following the treatment of mice with higher doses of U0126.

3.4 Methods

In this section we will go through the mathematical approach of developing our model of the impedance, and the design for simulating our model.

3.4.1 Linear Parametric Identification of Lung Input Impedance

System identification is a method to build a model to mathematically describe the relation between observed input and output data while estimating the transfer function. A model is useful for a variety of applications including prediction, simulation, and diagnostics. An important step in assessment of the lung mechanics is identification of a "suitable" model for representing the lung input impedance from the measured data. What suitable means depends strongly on the application, but constructing simple models with acceptable accuracy of fit between the model and the real system is always the primary concern.

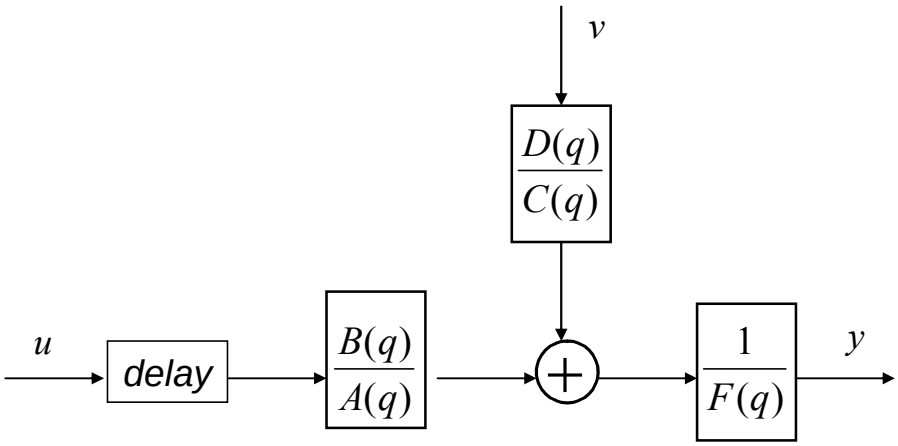


Figure 3.6 General linear model structure.

There are generally two approaches for identifying the lung input impedance; nonparametric and parametric. In nonparametric approaches, the frequency response is obtained directly from the experimental data. This model is very effective in the controller design. However, it is rarely used for the purpose of analysis in the respiratory system. The parametric identification approach is an iterative process which starts from the definition of the purpose of the model and uses existing knowledge to choose an appropriate framework to propose a first model structure. Given such structure and the set of experimental data, the objective of parametric identification is to estimate the non-measurable parameters, so as to reproduce the experimental results in the best possible way. In parametric identification the problem is formulated as a nonlinear optimization problem, where the objective is to find a set of parameters to minimize the function quantifying the goodness of the fit subject to the lung dynamics. The appropriateness of the mathematical model is best tested with a set of validation data which is different from the data used when the model was initially built.

This approach is widely used in the respiratory system where the general goal is to analyze and correlate these parameters with lung disorders.

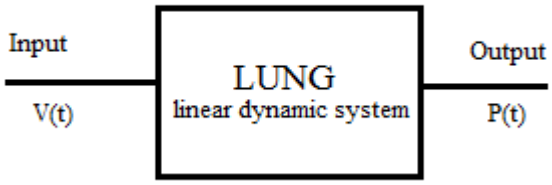


Figure 3.7 Lung system identification.

In this research, the objective of the parametric identification is to estimate a relation between input $u(t) = \dot{V}(t)$ and output of lung dynamic system $y(t) = P(t)$ by means of a *polynomial equation.* The general linear model structure, shown in Fig. 3.7, can be written in terms of the time-shift operator q. Note that this q description is completely equivalent to the Z-transform form.

Consider the following discrete-time difference equation (Ljung and Glad, 1994; Ljung, 1987; Söderström and Stoica , 1989; Li et al., 2009).

$$F(q)y(t) = \frac{B(q)}{A(q)}u(t-n_k) + \frac{D(q)}{C(q)}v(t), \tag{3.4}$$

where the polynomials A, B, C, D, and F are

$$A(q) = 1 + \alpha_1 q^{-1} + \alpha_2 q^{-2} + \ldots + \alpha_{n_\alpha} q^{-n_\alpha} \tag{3.5}$$

$$B(q) = \beta_0 + \beta_1 q^{-1} + \beta_2 q^{-2} + \ldots + \beta_{n_\beta} q^{-n_\beta} \tag{3.6}$$

$$C(q) = \gamma_0 + \gamma_1 q^{-1} + \gamma_2 q^{-2} + \ldots + \gamma_{n_\gamma} q^{-n_\gamma} \tag{3.7}$$

$$D(q) = \delta_0 + \delta_1 q^{-1} + \delta_2 q^{-2} + \ldots + \delta_{n_\delta} q^{-n_\delta} \tag{3.8}$$

$$F(q) = \eta_0 + \eta_1 q^{-1} + \eta_2 q^{-2} + \ldots + \eta_{n_\eta} q^{-n_\eta} \tag{3.9}$$

u is the input, n_k is the input delay that characterizes the delay response time, and v(t) is the white noise. It should be noted that t is the time instant at which the data is measured; so $t = N \cdot T_s + t_{0\,\text{and}}$where N is integer, T_s is the sampling time and t_0 is the initial time.

The structure given by (3.4) is stochastic and is able to incorporate random effects. Such effects are very common and often are the best available way to describe the difference between the ideal model and real observations known as identification error. This error depends on both the random noise and the choice of identification parameters (Franklin et al., 1998).

As shown in Figure 3.7, this model is able to represent a variety of parametric model structures, such as Auto Regressive Model with External Input (ARX), Average Model with External Input (ARMAX), Box-Jenkins (BJ), and Output Error (OE). These models, which differ by the number of included polynomials, provide insight into the system's physics and compact model structures. These structures can be achieved by the following substitutions correspondingly:

The ARX structure is obtained for A(q) = C(q) = D(q) = 1

$$F(q)y(t) = B(q)u(t - n_k) + \nu(t) \tag{3.10}$$

The ARMAX structure corresponds to A(q) = C(q) = 1

$$F(q)y(t) = B(q)u(t - n_k) + D(q)\nu(t) \tag{3.11}$$

The Output-Error model is obtained for F(q) = C(q) = D(q) = 1

$$y(t) = \frac{B(q)}{A(q)}u(t - n_k) + \nu(t) \tag{3.12}$$

The Box-Jenkins model is obtained for F(q) = 1

$$y(t) = \frac{B(q)}{A(q)}u(t - n_k) + \frac{D(q)}{C(q)}\nu(t) \tag{3.13}$$

Generally speaking, different structures provide varying levels of flexibility for modeling the dynamics and noise characteristics. It is often beneficial to test a number of structures to determine the best model. This process can be performed by evaluating several factors, such as simplicity, the relative order, and wellness of the fit between measured and predicted values.

In the respiratory system of mouse noise is usually not strongly coupled to the dynamics.Therefore, compared to ARMAX and BJ structures, which incorporate the noise dynamics, OE and ARX structures are considered better and simpler choices for modeling mouse respiratory system. The selection between ARX and OE can be quantitatively done by evaluating the following FIT function:

$$F = \frac{1 - \|y - \hat{y}\|}{\|y - \bar{y}\|} \tag{3.14}$$

Where $\hat{y}$ is the predicted value, $\bar{y}$ is the mean value of y, and $\| \ \|$ denotes the norm. Here, as the data we work with is real data measured from mice, we accept precision of 90% and above as being significant.

Based on evaluating 3.14 for the four asthma models in Experiment 1, the OE structure is selected as the base model for the analysis of the mouse respiratory system. A comparison of the performances from ARX and OE type based on 3.14 is provided in Figure 3.6. The OE structure outperformed the ARX model in three out of four mouse models of asthma. The laboratory models, other than normal, have been specifically manipulated based on a specific protocol to induce asthma. This could be a possible cause for the higher estimate accuracy in these models compared to the normal group.

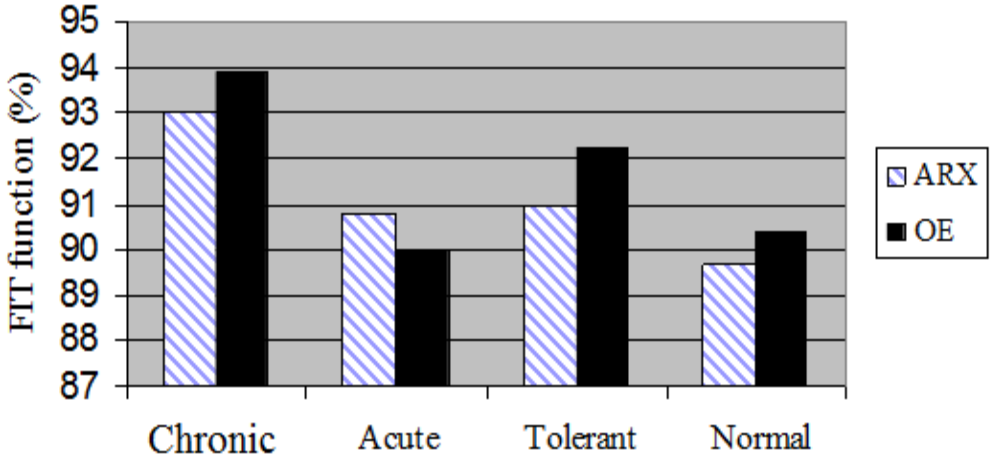


Figure 3.8 Comparison of the FIT function for OE and ARX structures for the four mouse models from experiment 1.

3.4.2 Semi- Impulse Response of the Respiratory System

The impulse response of a dynamic system is its reaction to an impulse signal (δ(t)) by which all modes of the system are stimulated. In real systems however, it is not possible to produce the exact impulse waveform. Therefore brief pulses, as approximations for δ(t), are used to stimulate the system. Provided that these pulses are short enough compared to the ideal impulse signal, the results will be acceptable approximations of the theoretical impulse response.

Inspiring from δ(t), we establish a *semi-impulse* input within the corresponding frequency range (from f_{min} to f_{max}). This will guarantee the tolerable speed of the signal for the respiratory system within which the laboratory models of asthma are valid. The *semi-impulse* will be used as the inputs for the OE model discussed previously. The OE

model's response to the *semi-impulse* is then found through simulation. This response can provide the basis for quantitative evaluation of lung diseases and treatment progression.

The ideal impulse response does not start from zero, since the input impulse gives initial conditions to the system's modes at t = 0+. The response is similar to the natural system's reaction to a specific initial condition which is created by ideal impulse input.

The semi-impulse response starts from zero since, in the respiratory system, the input cannot use an ideal impulse. The reason is that the model is valid only within a certain frequency range. We illustrate this effect by Figure 3.9.

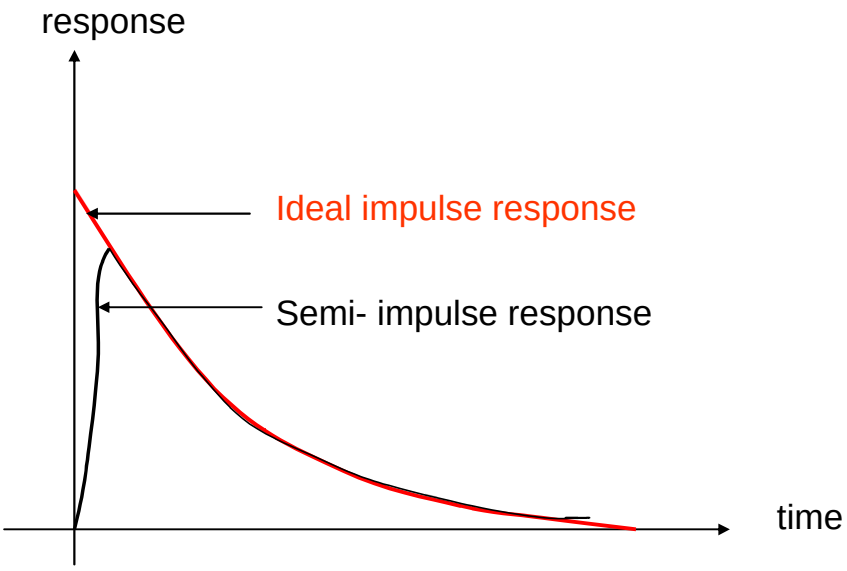


Figure 3.9 Realistic versus ideal impulse response.

The *semi-impulse* wave form is generated by

$$u = \frac{1}{N-1} \sum_{i=2}^{N} \cos(2\pi(\frac{i}{2}) f_{\min} t) \qquad (3.15)$$

where $N = \text{int}[2 f_{\max} / f_{\min}]$, f_{max}= *20.5 Hz* and f_{min}= *1 Hz* respectively. Figure 3.10 illustrates the proposed *semi-impulse* waveform.

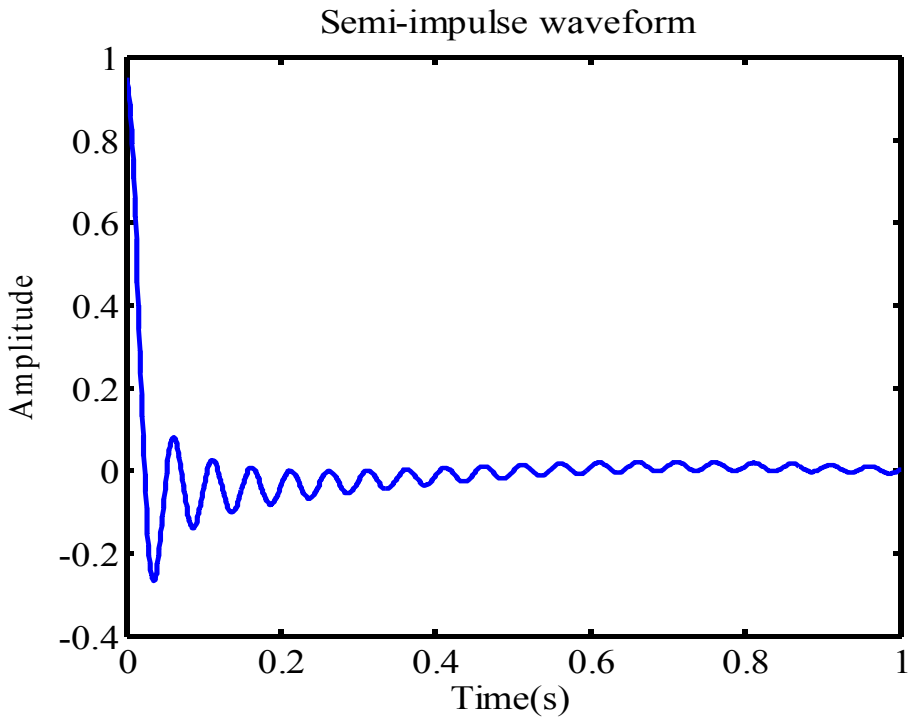


Figure 3.10 The *semi-impulse* waveform.

It should be noted that modeling the respiratory impedance with 3.12 makes it possible to study the time domain response of the system to not only the *semi-impulse* waveform but also any other input containing only the frequency components within the valid range for the laboratory models. The results of modeling experimental data from Experiments 1 and 2 with 3.12 and the simulation with the semi-impulse waveform are provided and discussed in Chapter 5.

Chapter 4

HUMAN DATABASE

4.1 Background

Asthma was long believed to be a relatively homogenous disease. It was characterized by mast cell– and eosinophil-dominant inflammation and its responsiveness to corticosteroids. Asthma severity was classified primarily by measures of airflow obstruction. This way severe asthma was a subtype of a disease with severe allergic inflammation and it required high dose corticosteroid treatment. Cellular or molecular defects in steroid responsiveness was thought to be responsible for poor or no response to high doses of steroids. This way the possibility that some asthma phenotypes could have airway pathology insensitive to steroids was neglected. Asthma now is considered a heterogeneous multidimensional disease in which, impairment and risk are the major determinants of the severity of the disease. Although eosinophilic asthma is still an important asthma phenotype characterized by systemic and airway markers of eosinophilia, subepithelial fibrosis, and corticosteroid responsiveness, importance of non-eosinophilic asthma is now well emphasized. Non-eosinophilic asthma is characterized by absence of eosinophilia and subepithelial fibrosis and poor responsiveness to corticosteroids. T-helper type 2 (Th2) cytokines have been proposed as the upstream molecular regulators of eosinophilic asthma. The molecular mechanism of noneosinophilic asthma is not well known however (Fahy, 2010).

Asthma can also be distinguished based on clinical features such as obesity, baseline lung function (low vs. high FEV1), degree of reversibility, presence or absence of confounding co-morbidities such as allergic rhinitis, chronic sinusitis with and without NSAID (non-steroidal anti-inflammatory drugs) sensitivity, and response to various treatment modalities such as leukotriene receptor antagonists, and inhaled and systemic steroids. Asthma within each sub-phenotype can be mild and severe. Furthermore, both mild and severe asthma can be clinically unstable despite being on proper controller medications. A major challenge has been to identify patients that are at risk of having an ***unstable*** clinical course of asthma.

Because of the complexity of etiology and pathogenesis, efforts have been made to apply unsupervised mathematical approaches to analyze multidimensional asthma parameters and group them into smaller clusters (phenotypes) based upon their interrelationship. Indeed, a number of studies have been published that applied unsupervised mathematical approaches to classifying asthma phenotypes (Haldar et al 2008, Moore et al 2010, Braiser et al 2008, Braiser et al 2010, Fitzpatrick et al 2011, Siroux et al 2011, Southerland et al 2012). These studies have identified not only new phenotypes but also illustrated the complexities of each phenotype. In one of the earliest clustering studies, Haldar and colleagues performed k-means cluster analyses of an asthma database separately in a primary and a secondary care setting (Haldar et al 2008). In the secondary care setting, the analysis identified four clusters of patients from a total

of 187 asthmatic patients. The clusters differed among themselves in age, sex, body mass index (BMI), reversibility, sputum eosinophils, fractional exhaled nitric oxide (FENO), dose of inhaled steroids, and number of steroid bursts. This study specifically identified and focused on two interesting clusters: one early onset symptom-dominant and another late onset inflammation-dominant. Both represented relatively refractory asthma. Inflammation-guided treatment adjustment allowed for the reduction in exacerbation and dose of inhaled steroid in these patient populations. This study did not include any airway hyperreactivity measurement (e.g. PC20 for methacholine) in the analysis.

In one of the most robust database analyses of severe asthmatic patients by Moore and colleagues a cluster analysis grouped patients in five clusters (Moore et al 2010). The major disease variables that determined the assignment of patients to the clusters came from the pulmonary function test (six variables—FVC, FEV1, FEV1/FVC at baseline and the best from the database), patient age, duration of disease, gender, and composite medication score. Two clusters concentrated patients with the highest health care utilization, i.e. patients with frequent steroid bursts, ER visits, and hospitalizations. These two clusters also had the lowest PC20 and one of the clusters had the highest sputum neutrophil count. Sputum and blood eosinophil counts and FENO did not show any difference among the clusters.

4.2 Problem Statement

The Global INitiative for Asthma (GINA); launched in 1993 produces guidelines called “Global Strategy for Asthma Management and Prevention” based on the evidence and recent publications. Reviewing the initial GINA guideline published in 1993, we observe that its emphasis was based on disease severity grading: intermittent; mild persistent; moderately persistent; and severely persistent asthma. Since then several studies have shown that “asthma control” defined by the GINA guideline is achievable in most patients. Achievement of good control in patients improves their health condition (Bateman et al. 2007, Pedersen et al. 2007, Pauwels et al. 2003). Therefore, in 2006 the global strategy for asthma management and prevention was revised to emphasize the management of the disease based on the level of asthma control rather than clustering the disease based on its severity. Thus, in our research, we concentrated on clustering the patients based on how "controllable" their asthma is rather than how "severe" it might be. The controllability of a patient's asthma is more informative about how he or she will be doing in the near future in comparison to how severe the disease currently is.

Asthma control takes into account the patient’s recent clinical state with respect to parameters such as day time and night time symptoms, frequency of rescue medication use, and lung function. Asthma control also considers a patient’s future potential risk for experiencing exacerbations, accelerated decline in lung function, -treatment-related side effects, or loss of control. Some pathologic and physiologic measures, independent of the level of current clinical control, also influence future risk (Reddel et al., 2009).

Based on Bateman et al. 2008, the components of control include daytime symptoms, limitations of activities, nocturnal symptoms/awakening, the need for rescue medication, lung function, and exacerbations.

The concept of asthma controllability (stability) - is a complicated one. Mathematical tools could significantly help physicians in understanding and clustering asthmatic patients based on this concept. Previously, researchers have tried to cluster asthmatic patients as discussed in the introduction of this chapter. However those studies did not focus on the controllability and stability of asthma and only identified the phenotypes of asthma. Patients with unstable asthma were scattered in different clusters.

For our study, we accessed the electronic research database at National Jewish Health. The protocol for this study was approved by the institutional Review Board. The initial query with the keyword "asthma" identified 3,000 patients. In our study, we selected only four very commonly available clinical parameters that had been known to highly correlate with asthma. These parameters are introduced in Section 4.3 of this chapter. A total of 174 patients with measurements for the selected four parameters were available in the National Jewish Health database. We determined the number of possible clusters within the study population by a dendrogram generated using Ward's minimum variance algorithm. We performed a cluster analysis by principal component analysis (PCA) and the application of the self-organizing maps (SOMs). We then assigned scores

to the major qualitative factors that determined asthma instability. Overall, the combination of our methods, PCA, SOM, and stability scoring generated an asthma stability analysis system predicting the possibility of a patient having only unstable asthma based on four very well known clinical parameters.

4.3 Parameter Selection

There are many laboratory parameters that tend to broadly correlate with the diagnosis of asthma but the correlation is rather poor and its positive predictive value is low. We focused on four laboratory parameters that were quantitative in nature — FEV1 (forced expiratory volume in 1 second), PC20 (provocative concentration 20) for methacholine, peripheral blood eosinophil count, and serum total IgE.

First measurement corresponds to *Forced Expiratory Volume (FEV1)*. FEV1 is the amount of air that one can forcibly blow out in one second. It is measured in liters. FEV1 is considered one of the primary measures of lung function. A spirometer is the machine that the patient blows into (Figure 4.1). The machine predicts an FEV1 value for the person according to his age, height, and weight. The real value is then measured. The result is two numbers; an absolute FEV1 value between 0 and 6 L; and a percentage showing the ratio of the real FEV1 over the predicted value. For a normal person the FEV1 percentage is above 80%.

Figure 4.1 A patient is blowing into a spirometer under supervision, Image source: http://www.nhlbi.nih.gov/health/health-topics/topics/copd/diagnosis.html

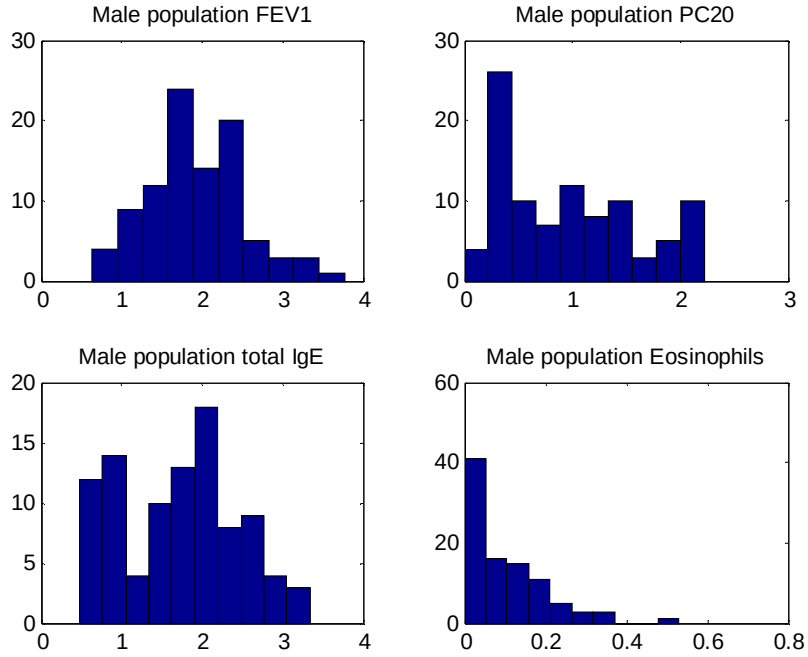
The second measurement corresponds to the *Methacholine Challenge test*. The Methacholine test is a breathing test that detects bronchial hyper reactivity that occurs in asthma. It is usually ordered after a routine pulmonary function test (FEV1) has been completed and the patient has shown a normal FEV1, yet has symptoms. In this test, the patient breathes in nebulized methacholine chloride through a mouthpiece. The patient is then asked to blow forcefully into a spirometer. A person with asthma reacts to lower doses of inhaled methacholine. A PC20 (provocative concentration 20) for methacholine test results below 8 mg is considered asthmatic and above 8 mg is normal. An alternative approach to diagnose asthma is to demonstrate a reduced baseline FEV1, which is reversed by a bronchodilator treatment. The latter approach is used more frequently.

The third measurement is the *eosinophil count*. Eosinophils are the white blood cells of the immune system that are associated with parasitic infections and allergic diseases such as asthma. A Eosinophils count requires a blood test.

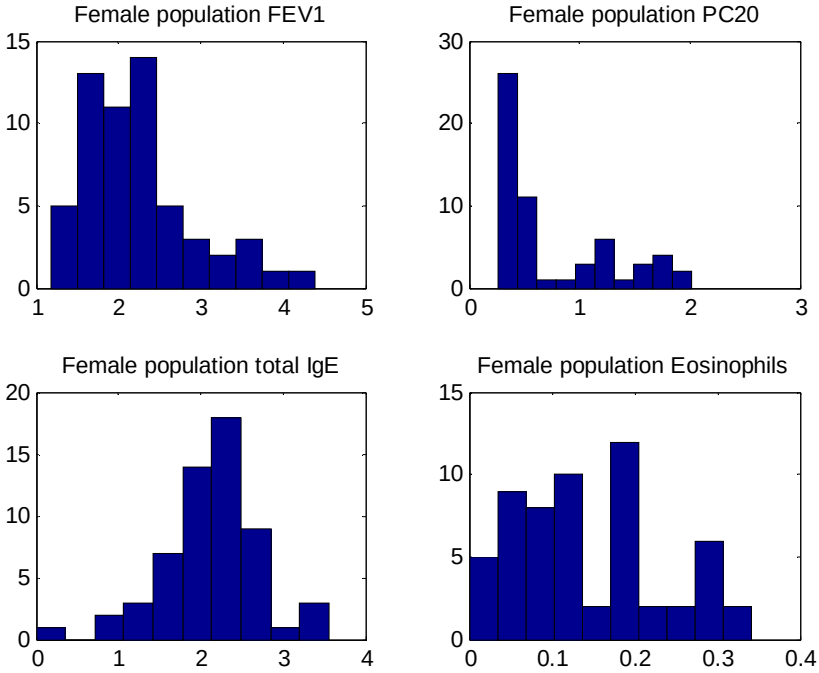
The last measurement is *immunoglobulin E (IgE)*. IgE is a class of antibodies that plays an important role in allergy and is especially associated with type 1 hypersensitivity (which occurs in asthma). IgE is capable of triggering the most powerful allergic reactions. IgE level is also deternined through a blood test.

Due to skewed distributions of Methacholine, IgE and Eosinophils we applied transformations to shift them closer to normal distributions. For PC20, a natural log was used, and for IgE and eosinophils, log10 was applied.

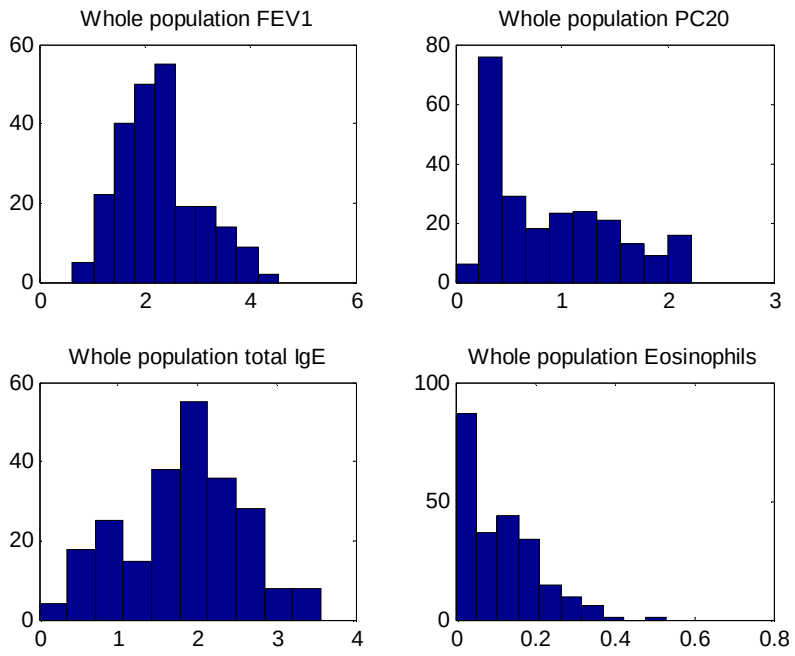
The histogram plots for the four parameters for males and females separately and the whole study population is shown in Figure 4.2.


Male population FEV1
Male population PC20
Male population total IgE
Male population Eosinophils

(a)


Female population FEV1
Female population PC20
Female population total IgE
Female population Eosinophils

(b)



(c)

Figure 4.2 Histograms of the four parameters for the (a) male patients, (b) female patients, (c) whole population.

As we want to focus on patients who meet the strictest objective criteria for diagnosis of asthma, we selected patients who had a PC20 for methacholine $\leq$ 8mg/ml. Asthma severity in study population ranges from intermittent mild to severe in the database.

4.4 Methods

In this section we introduce the mathematical approaches we applied in MATLAB to the database. We applied principal component analysis (PCA) as the first technique to the database. PCA extracts the structure of the database and therefore clusters of the

patients form. In the next step, we estimated the number of possible clusters by a hierarchical clustering tree (dendrogram). We then applied self-organizing maps from the neural network field to find the clusters resulted from PCA. We pre-specified the number of clusters achieved through generation of the dendrogram for the SOMs.

4.4.1 Principal Component Analysis

PCA (principal component analysis) is a modern tool for data analysis. PCA is non-parametric and extracts relevant information from large databases by linearly reducing the dimension of the data. The goal of PCA is to identify the most meaningful basis and transform the original dataset to this new basis. We expect this new basis to filter out noise and redundancy and reveal the hidden interesting data patterns. PCA can form clusters of data that can be used for making predictive models (Shlens, 2009). The main features of PCA are:

- principal components (PCs) are linear combinations of the original variablese;

- all of the principal components are orthogonal to each other. Therefore, there is no redundancy in principal component space;

- each principal component may replace several original variables as it is a linear combination of those original variables;

- the first component extracted in a PCA accounts for a maximal amount of total variance in the observed variables and the second PC will be correlated with some of the

observed variables that did not display strong correlations with PC1. The second PC will also be uncorrelated with the first component;

- the remaining PCs that are extracted in the analysis display the same two characteristics: each PC accounts for a maximal amount of variance in the observed variables that was not accounted for by the preceding components; and it is uncorrelated with all of the preceding PCs;

- total variance of the data set equals the number of original variables.

PCA involves the calculation of the eigenvalues of the data co-variance matrix or singular values of the data matrix after mean centering the data for each attribute. In our study, we implemented the PCA algorithm in Matlab based on finding the eigenvalues of the covariance technique. The fundamental mathematical concepts involved in PCA include:

Variance: is the original statistical measure of the spread of data in a dataset with mean $\bar{x}$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)} \tag{4.1}$$

Covariance: The covariance matrix is a symmetrical matrix presenting the variation of two vectors with respect to each other. The covariance of a vector with itself is its variance.

$$\mathrm{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)} \tag{4.2}$$

A positive value of covariance indicates that the two vectors increase or decrease together, whereas a negative value indicates that when one increases the other decreases. A zero covariance means that the two vectors are independent of each other.

Eigenvalues and Eigenvectors: consider the equation $Ax=\lambda x$

Certain exceptional vectors x are in the same direction as Ax. Those are the "eigenvectors" of matrix A. Multiply an eigenvector by A, and the vector Ax is a number λ times the original x; λ is called an eigenvalue of A. To find the solution one should solve:

$$\det (A-\lambda.I) = 0 \tag{4.3}$$

The steps in calculation of the principal components (PCs) are as following

1. data scaling: if the data are not normalized, a variable with a large variance will dominate. Therefore we have to scale the data. The most common scaling technique is the Unit variance scaling. Next we mean center the data;

2. the calculation of the covariance matrix.

3. the calculation of eigenvalues and eigenvectors of the covariance matrix. The covariance matrix is a symmetric matrix and the eigenvectors will be orthogonal;

4. the eigenvectors with the largest eigenvalues correspond to the dimensions that have the strongest correlation in the dataset. Choose the significant PCs and discard the rest;

5. map the data to the selected (significant) principal components' space.

Several criteria have been proposed for determining how many PCs should be investigated and how many should be ignored (step 4 in PCA). One common criterion is to ignore principal components at the point where the next PC shows little increase in the total variance explained. A second criterion is to include all those PCs up to a predetermined total percent variance explained, such as 90%. A third standard is to ignore components whose variance explained is less than one when a correlation matrix is used or less than the average variance explained when a co-variance matrix is used, with the idea that such a PC has less than one variable's worth of information. A fourth standard is to ignore the last PCs whose variance explained is all roughly equal. In our case we stopped at the point that the variance explained by a PC dropped below one.

4.4.2 Hierarchical Clustering

Optimal clustering can be defined as the one of all possible combinations of groupings, which provides the most meaningful associations in the clusters. In the context of our research, target groups (clusters) should reflect asthma heterogeneity identifiers.

Evaluation of clustering performance is a fundamental and difficult problem. As defined in Chapter 2, clustering is done without a priori understanding of the internal structure of the data.

If we use a fixed SOM size, the output of the SOM will only be a reduced representation of the input data. Depending on the network size and the data structure, some SOM nodes can remain unoccupied after the training. The original topological relations between target groups are preserved in the SOM output. This implies that the output of SOM depends on both the differences among groups and on the sizes of the groups. As a result we need to further process the SOM clustering to find a right number of clusters (Lehmann and Khawaja 2011).

Agglomerative hierarchical (bottom-up) clustering is usually used to determine the possible number of clusters in a population to be fed to the SOM algorithm. Agglomerative hierarchical clustering performs as follows (Lehmann and Khawaja 2011):

1. find the two closest clusters by using an appropriate distance measure;

2. merge these clusters and recalculate centroids and cluster measures;

3. go back to step one until only one cluster is left.

For this algorithm we used *Ward's linkage*. This method uses the incremental sum of squares, i.e. the increase in the total within-cluster sum of squares as a result of joining two clusters. The within-cluster sum of squares is defined as the sum of the squares of the distances between all objects in the cluster and the centroid of the cluster.

We apply the possible number of clusters generated by Ward's dendrogram to our SOM algorithm introduced in Section 4.4.3.

4.4.3 Self-Organizing Maps

Self-organizing maps (SOMs) is a very famous and powerful category of unsupervised (the correct answer is unknown) neural networks with competitive and cooperative learning abilities. They were developed in 1982 by Tuevo Kohonen. They are called "self organizing" because they learn on their own through unsupervised competitive learning. They are called "maps" because they try to map their weights to conform to the input data (Guthikonda, 2005).

SOM nodes try to become like the input data presented. The SOM algorithm is useful for extracting implicit, valuable information from large datasets. The principal advantages of SOM include the identification of clusters of similar sequences, projection and visualization of high dimensional data spaces to one or two dimension space, and the preservation of topological relationships between data vectors. The fact that SOM is

based on neural networks confers a series of advantages that makes it suitable to the clustering of large amounts of noisy data with outliers.

Retaining principal features of the data is an important property of SOMs. The topological relationships between input data are preserved when mapped to a SOM network. SOMs works similar to K-Means with the difference of not only choosing the size but also the shape for the network of clusters (to fit our data into). Similar to K-Means clustering, one should choose an initial value for expected number of clusters in the SOM algorithm.

SOM first populates its nodes by randomly sampling the data and then refining the nodes in a systematic fashion similar to K-means. However; SOM does not require the number of clusters and nodes to be the same. This means when the map is complete some nodes may remain without any associated items. SOM provides information on similarity between nodes which is another difference with K means (Schlosser and Wagner, 2004).

The SOM has an input layer fully connected to the output layer. When an input is received output nodes compete with each other to form the pattern. The winner node is the one that has the closest weight vector to the input pattern. The winner and surrounding nodes are updated for generating closer weight vector to the input pattern. This causes the close nodes to represent similar patterns and distant nodes to represent

distinct patterns as the training progresses. The clusters that are more dissimilar will be mapped in the SOM output layer further apart (Schlosser and Wagner, 2004).

Chapter 5

RESULTS AND DISCUSSIONS

5.1 Mouse Lung Impedance

In Chapter 3 we introduced the mouse lung impedance which is usually measured between 1 and 20.5 Hz. For this frequency range, the FIT function for Output Error (OE) and Auto Regressive Model with External Input (ARX) structures of the same orderwere evaluated. Based on the average FIT for each mouse group the OE model was proven to result in a better fit (Figure 3.8). The corresponding OE model can be written as

$$y(t) = \frac{\beta_0 + \beta_1 q^{-1} + \beta_2 q^{-2} + \beta_3 q^{-3}}{1 + \alpha_1 q^{-1} + \alpha_2 q^{-2}} u(t) + v(t) \tag{5.1}$$

The modeling process of experimental data with (5.1) is first performed for data from Experiment 1 (details in 3.3.1; i.e. Saline, Chronic asthma, Acute asthma, and Tolerant asthma models) and the results are provided in section 5.1.1. Section 5.1.2 corresponds to the results of analysis of experimental data from Experiment 2 (details in 3.3.2; Chronic asthma, Normal, and MEK treated models) with (5.1).

After estimating the real data with the third order model of (5.1), each model is simulated with the semi-impulse waveform (Figure 3.8) in the time domain. The results

of simulations for Experiments 1 and 2 is provided in 5.1.1 and 5.1.2 respectively. The discussion of these results is provided in Section 5.1.3. The results of mouse lung impedance modeling has been published in: Hanifi, A., Goplen, N., Matin, M., Salters, R. E., and Alam, R. (2012). "A Linear Parametric Approach for Analysis of Mouse Respiratory Impedance," IEEE Trans. Biomed. Circ. Sys., vol. 6 no. 3, pp. 287-294.

5.1.1 Experiment 1: Model Development

Table V.I presents the (5.1) coefficients of (α_i, β_i) for each mouse group and each dose of methacholine from Experiment 1, i.e. saline, chronic asthma, acute asthma, and tolerant asthma models. Studying Table V.I one can observe that the poles of the respiratory system in all of these four models are located in approximately the same location. But the zeros are changing locations with the methacholine dose change in each group (asthma model) and also are changing from one asthma model to another. This implies that the modes of the lung system are not changing as the system transitions from normal to any other asthma state.

Figures 5.1 to 5.4 show the bode plots from MATLAB for the model fitting process of data from Experiment 1. The red stars in the figure represent the real data from measurements of impedance in laboratory models. The dashed green lines correspond to the corresponding estimated third order linear model fit from (5.1). With the coefficients provided in Table V.I, the identification errors of the proposed linear parametric models lie in an acceptable range. Please note that we deal with a stochastic model, so the

identification problem is much more complicated in this case compared to that of the deterministic systems (Ljung, 1987; Franklin et al., 1998).

Group	Meth. dose	β_0	β_1	β_2	β_3	α_1	α_2
Normal	0	0.694	-1.45	0.869	-0.12	-1.98	0.977
	3	0.723	-1.28	0.464	0.095	-1.98	0.978
	6	1.285	-2.86	1.948	-0.37	-1.98	0.977
	12	1.133	-2.02	0.738	0.148	-1.98	0.977
	25	0.686	-0.57	-0.81	0.693	-1.98	0.98
Acute	0	0.897	-2.3	1.951	-0.55	-1.97	0.973
	3	1.56	-3.95	3.301	-0.91	-1.97	0.97
	6	3.613	-9.18	7.681	-2.12	-1.97	0.97
	12	6.038	-15.4	12.98	-3.6	-1.97	0.972
	25	12.47	-32.5	28.11	-8.03	-1.97	0.972
Tolerant	0	0.422	-0.65	0.088	0.136	-1.98	0.982
	3	0.466	-0.64	-0.04	0.215	-1.98	0.981
	6	-0.35	2.425	-3.62	1.552	-1.98	0.981
	12	-0.47	3.358	-5.02	2.135	-1.98	0.983
	25	-2.5	10.54	-13.1	5.095	-1.99	0.988
Chronic	0	1.534	-3.65	2.77	-0.66	-1.98	0.976
	3	1.344	-2.93	1.915	-0.33	-1.98	0.976
	6	1.432	-2.84	1.495	-0.09	-1.98	0.977
	12	2.381	-5.08	3.205	-0.5	-1.98	0.976
	25	1.661	-2.45	0.178	0.615	-1.98	0.977

Table V.I. Linear parametric models (5.1) coefficients for Normal, Acute, Tolerant and Chronic Asthma models, Hanifi et al. 2012

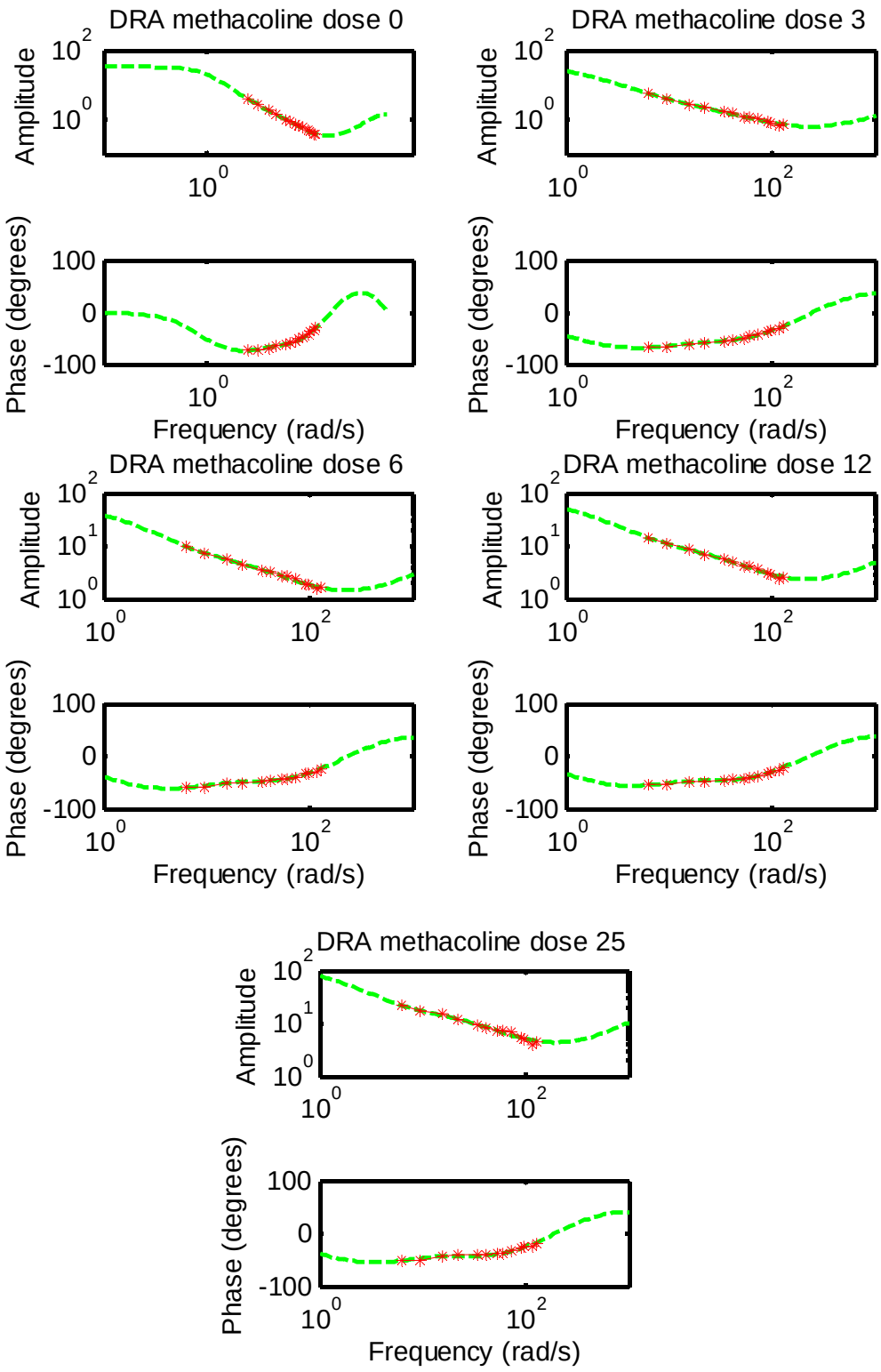


Figure 5.1: Bode plots of the real data (R *) and estimated transfer function (G --) for **chronic asthma** model Hanifi et al. 2012.

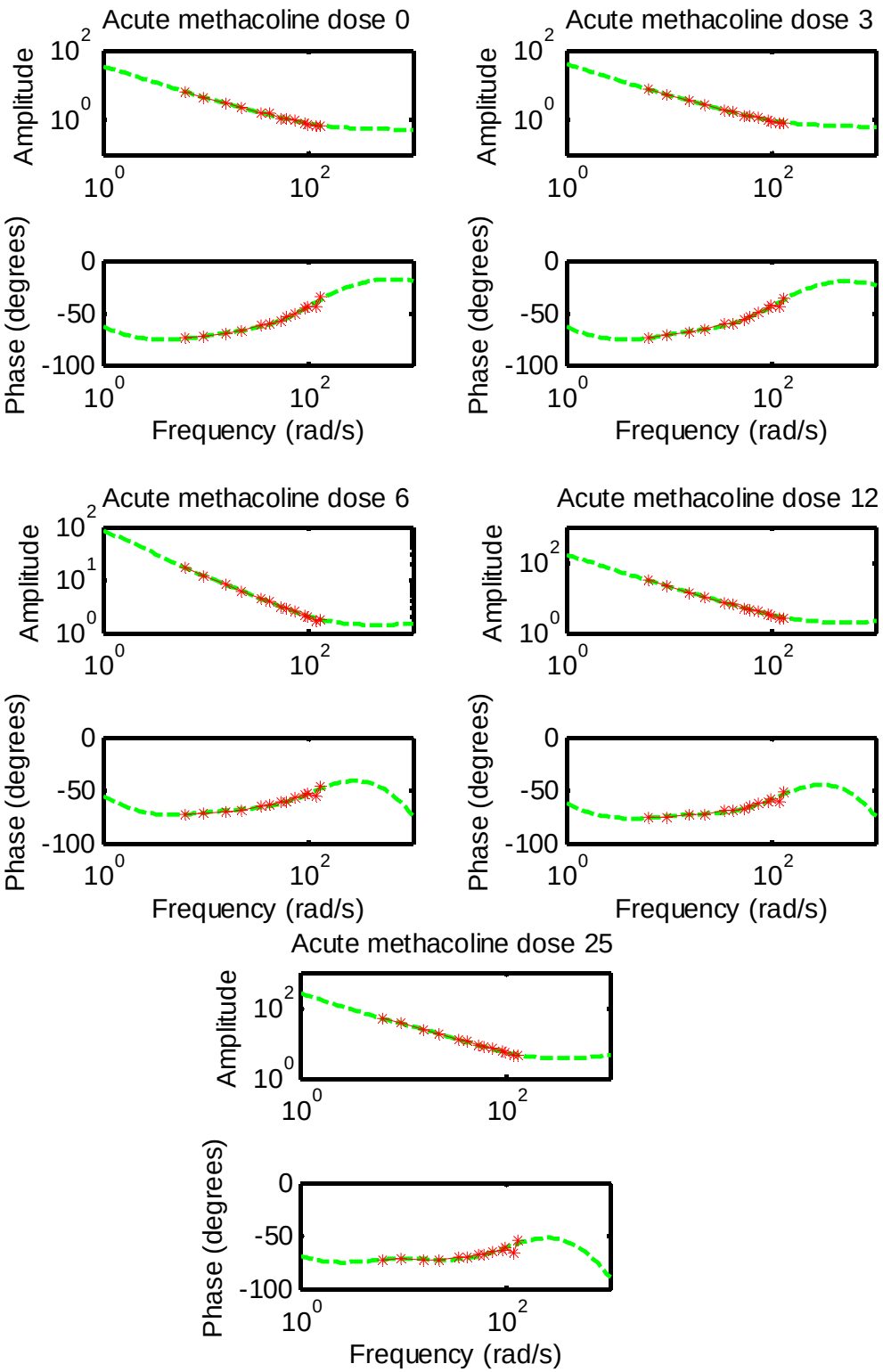


Figure 5.2: Bode plots of the real data (R *) and estimated transfer function(G --) for **Acute asthma** model, Hanifi et al. 2012.

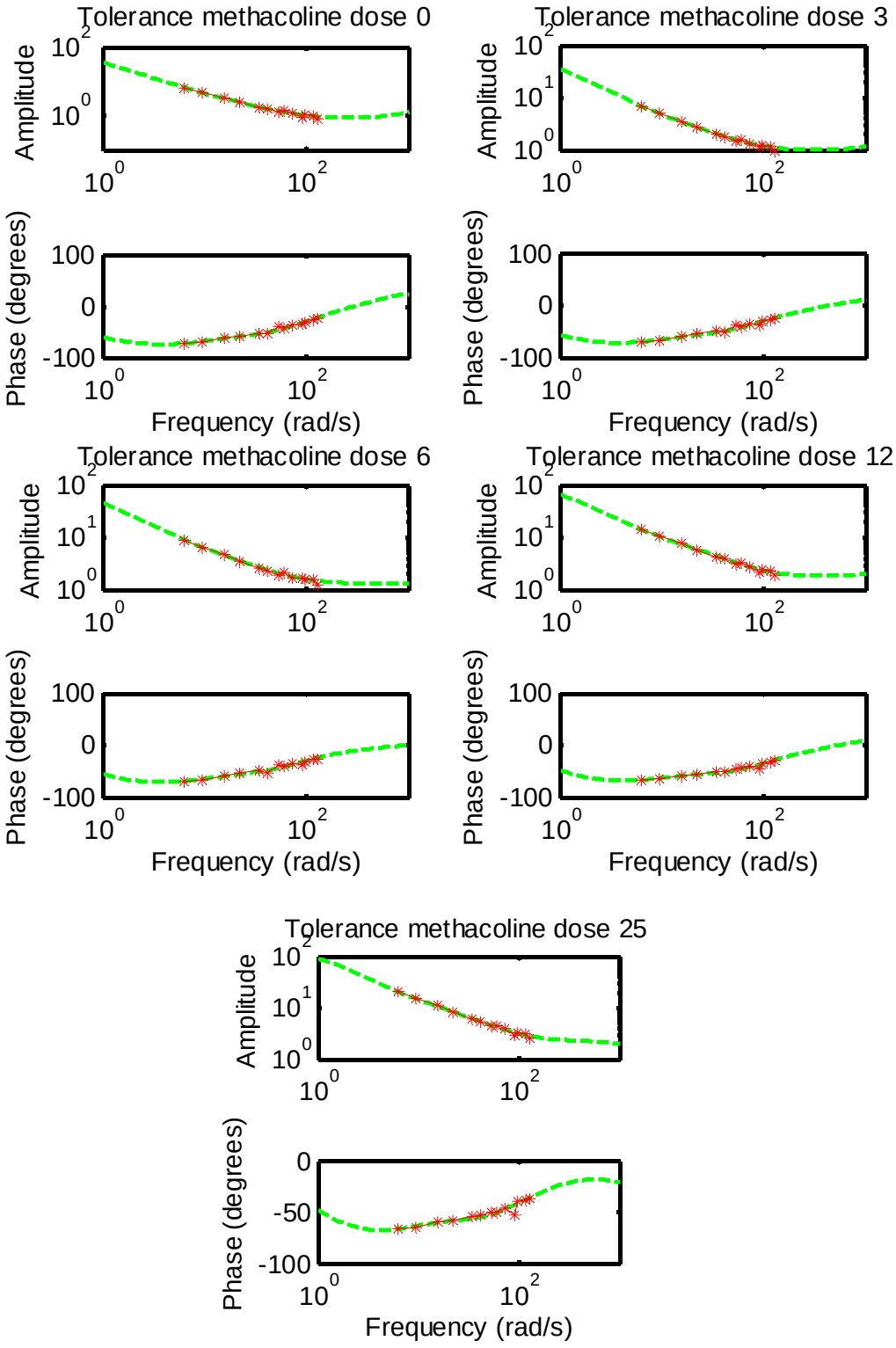


Figure 5.3: Bode plots of the real data (R *) and estimated transfer function (G --) for **Tolerant asthma** model, Hanifi et al. 2012.

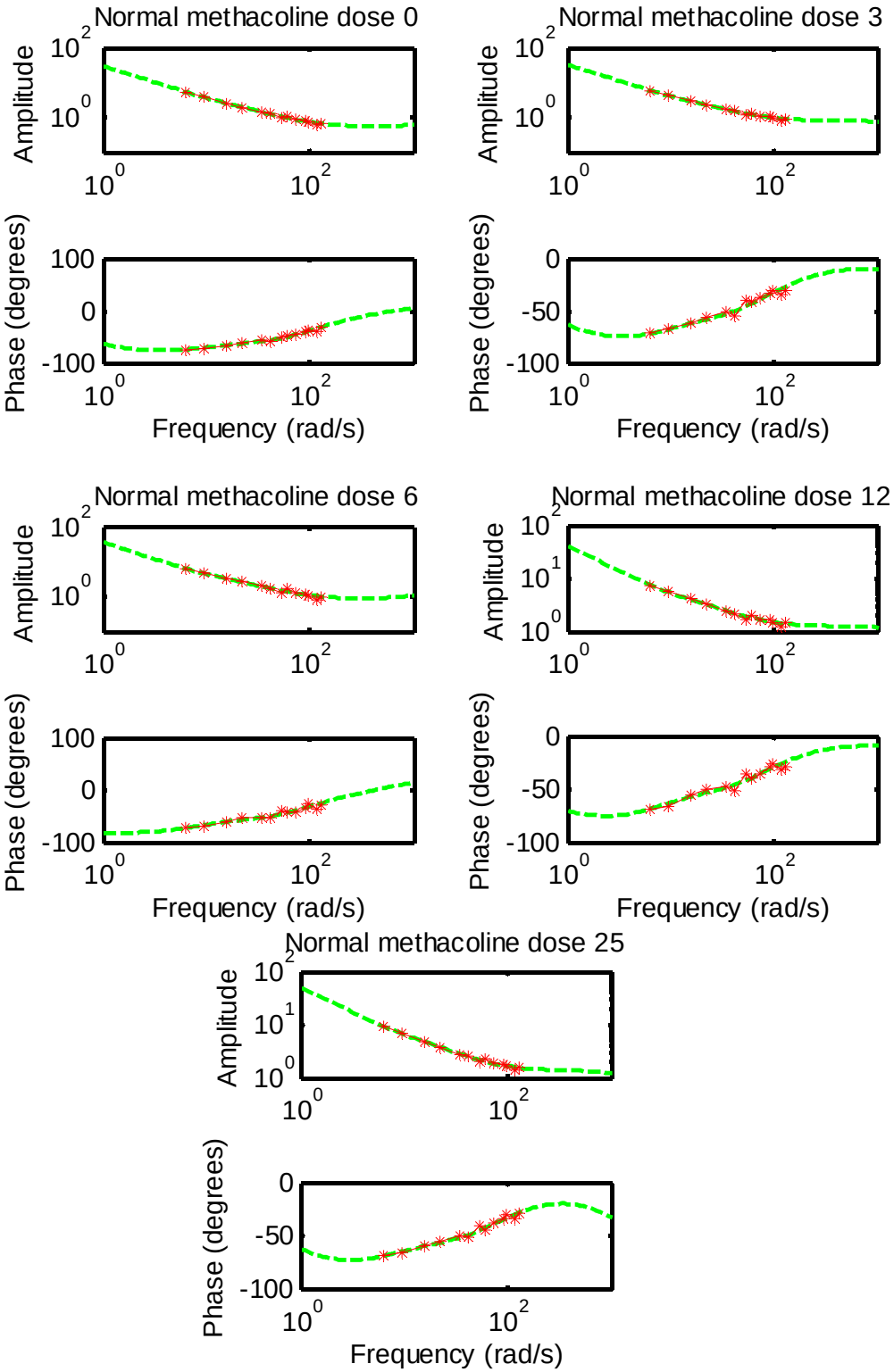


Figure 5.4: Bode plots of the real data (R *) and estimated transfer function (G --) for **Normal** model, Hanifi et al. 2012

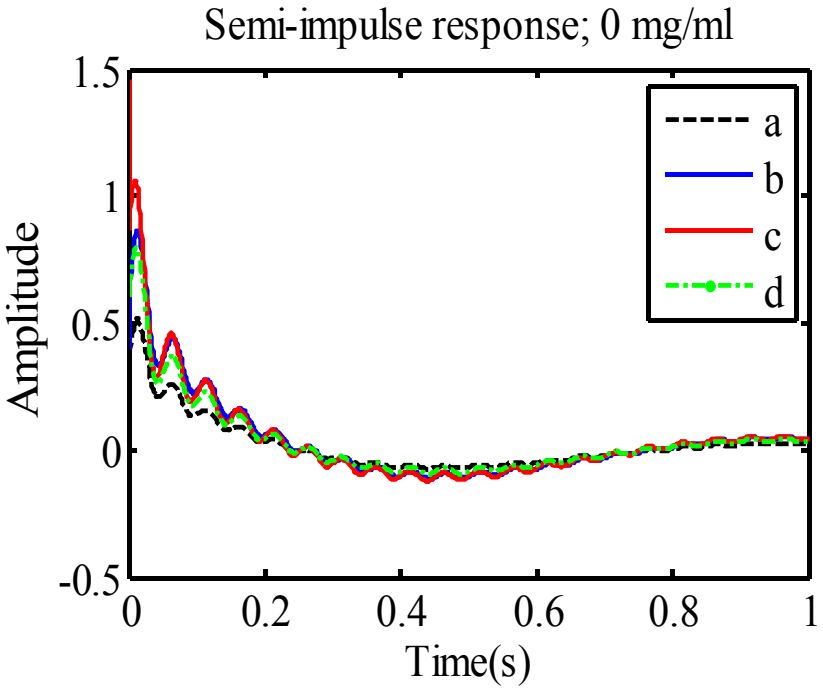


Figure 5.5 the result of simulation of a) Chronic b) Acute c) Tolerant d) Normal with semi-impulse waveform at the baseline Hanifi et al. 2012

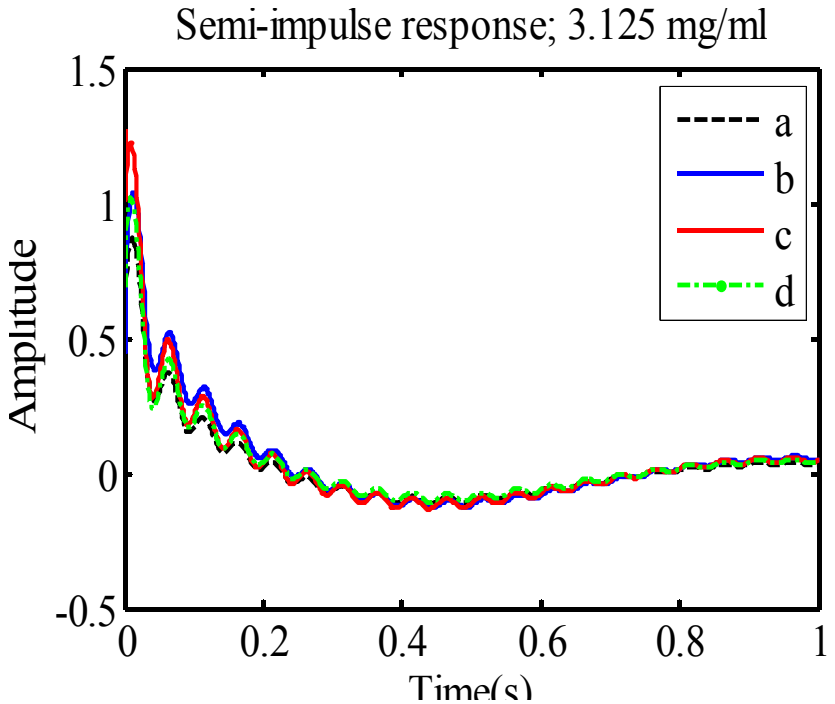


Figure 5.6 The result of simulation of a) Chronic b) Acute c) Tolerant d) Normal with semi-impulse waveform at dose 3.125 mg/ml of methacholine Hanifi et al. 2012.

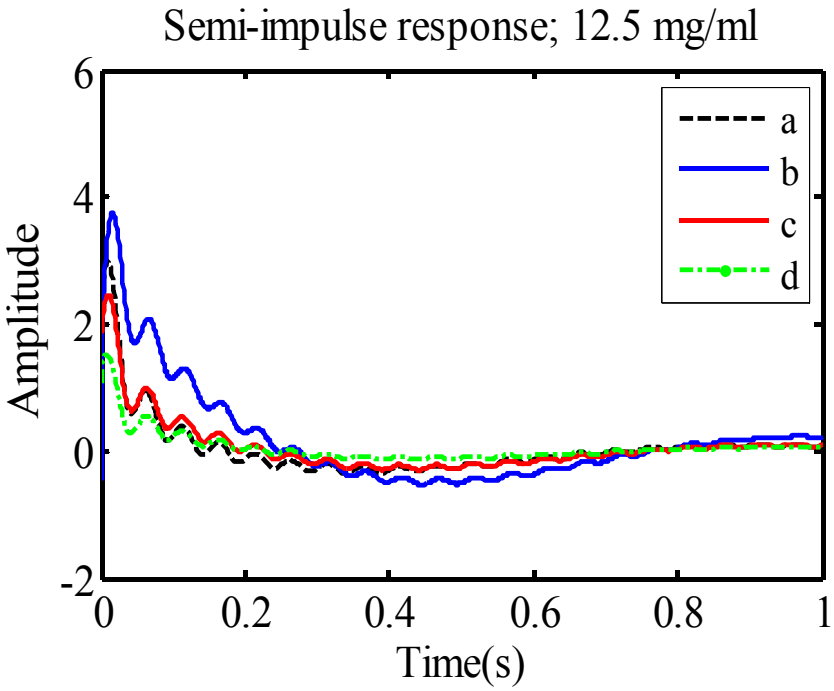


Figure 5.7 The result of simulation of a) Chronic b) Acute c) Tolerant d) Normal with semi-impulse waveform at dose 12.5 mg/ml of methacholine Hanifi et al. 2012.

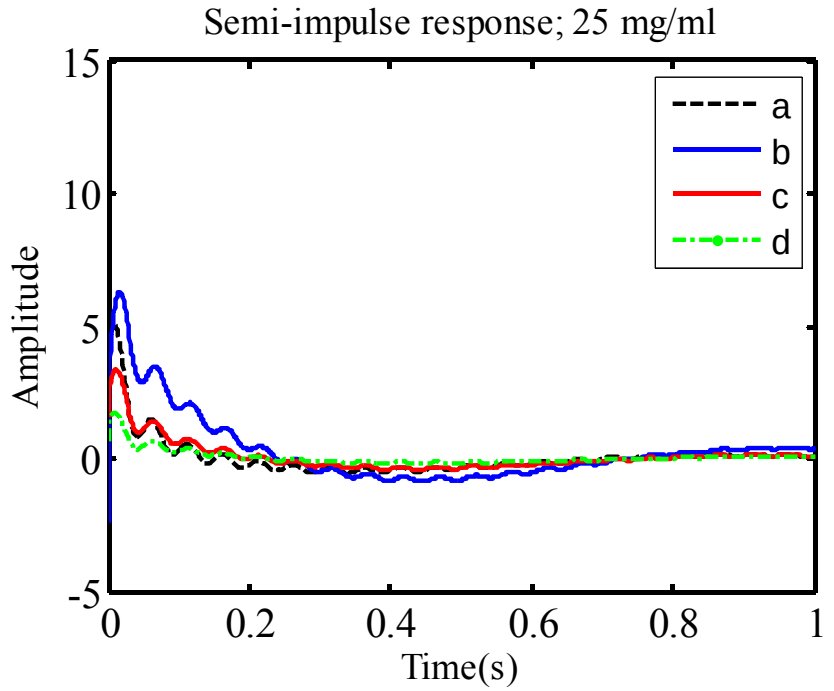


Figure 5.8: the result of simulation of a) Chronic b) Acute c) Tolerant d) Normal models with semi-impulse waveform at dose 25 mg/ml of methacholine Hanifi et al. 2012.

Figures 5.5 to 5.8 present the simulation results from MATLAB of models from Table V.I with the *semi- impulse* waveform shown in Figure 3.10. Figure 5.5 shows the simulation results of the four mouse models at baseline methacholine level (0mg/ml). At the baseline level the four models respond very closely and similarly to the *semi-impulse*. In Figure 5.6 the methacholine dose is increased to 3.125 mg/ml and the responses remain fairly close. Once the methacholine reaches 12.5 and above however; the models differentiate behaviors.

One can see that the Normal model is the one with the steadiest response with respect to the increase of methacholine dose among the four laboratory models from Experiment 1. Next is the Tolerant model. This model, when compared to Acute and Chronic models, is closest in behavior to a normal lung. The reason some of the responses (especially those with higher methacholine doses) show oscillations, or contain damped sinusoidal waveforms is that the *semi-impulse* input, given by 3.15 and illustrated by Figure 3.10 contains low frequency harmonics. Please note that the results verify that the respiratory system with higher methacholine doses reacts faster to the input pulse.

Another way to investigate the results shown in Figures 5.5 to 5.8 is to categorize the *semi-impulse* responses based on the asthma models (Figures 5.9 to 5.12) rather than the dose of methacholine. There is a great deal of information hidden in these figures.

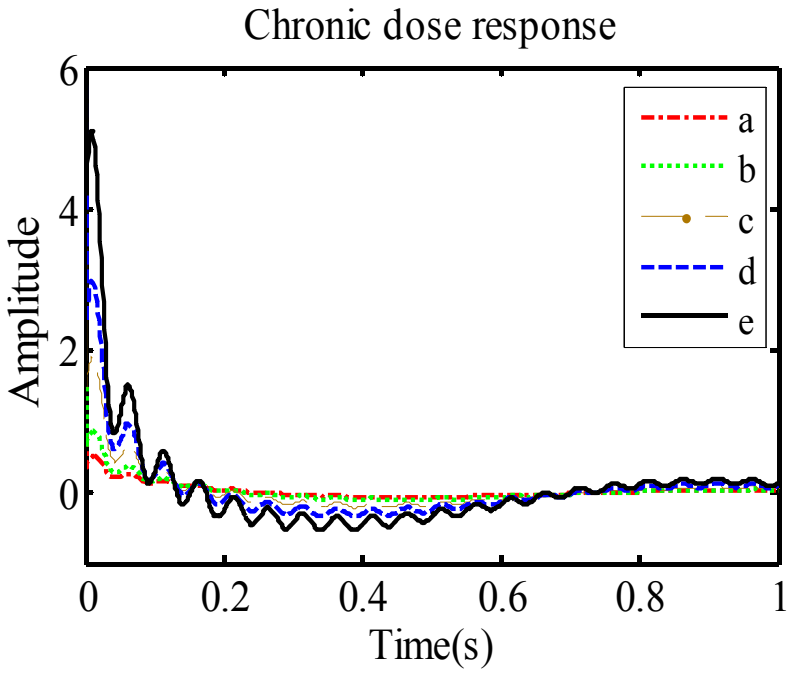


Figure 5.9 Semi-impulse responses for the Chronic asthma model, a) 0 mg/ml b) 3.125 mg/ml c) 6.25 mg/ml d) 12.5 mg/ml e) 25 mg/ml.

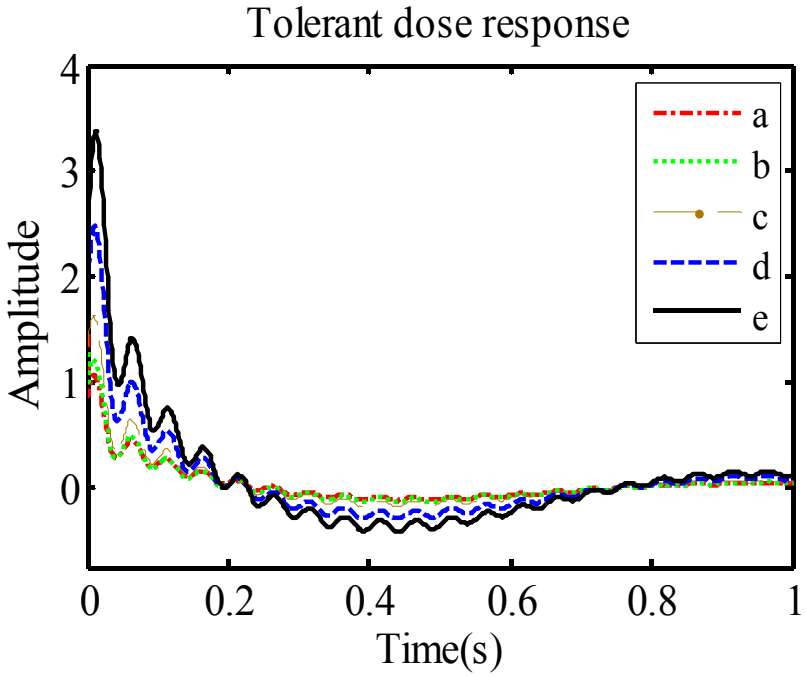


Figure 5.10 Semi-impulse responses for the Tolerant asthma model a) 0 mg/ml b) 3.125 mg/ml c) 6.25 mg/ml d) 12.5 mg/ml e) 25 mg/ml

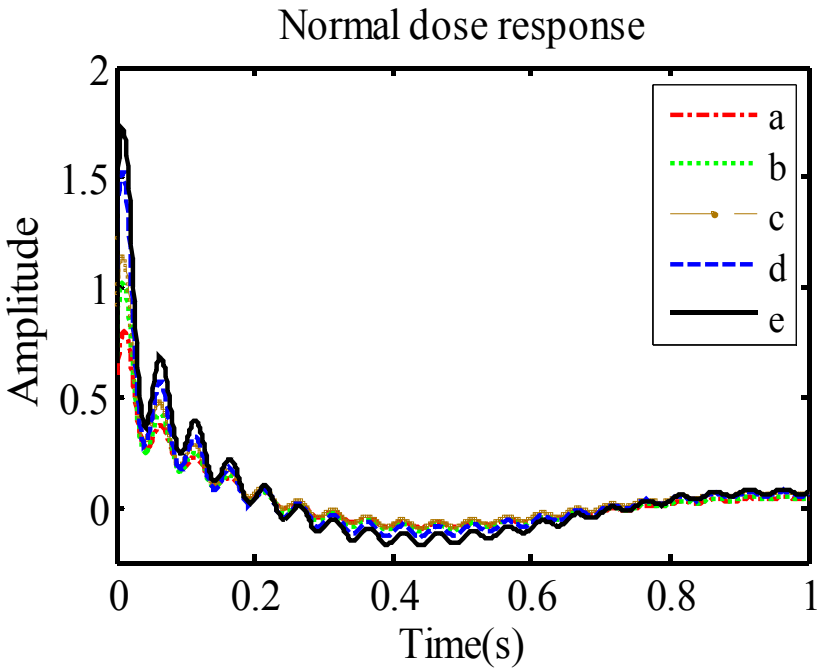


Figure 5.11 Semi-impulse responses for the Normal model, a) 0 mg/ml b) 3.125 mg/ml c) 6.25 mg/ml d) 12.5 mg/ml e) 25 mg/ml

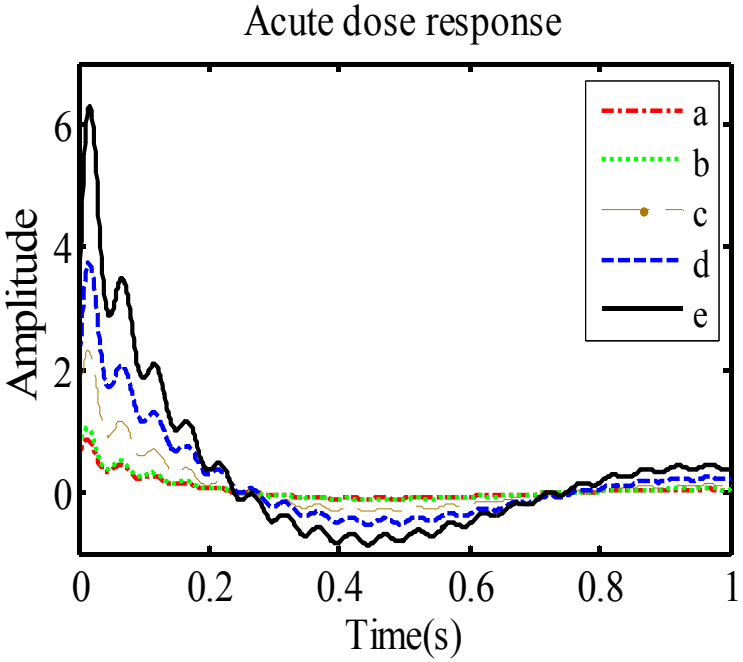


Figure 5.12 Semi-impulse responses for the Acute asthma model, a) 0 mg/ml b) 3.125 mg/ml c) 6.25 mg/ml d) 12.5 mg/ml e) 25 mg/ml.

To interpret and compare the *semi-impulse* responses from different mouse models and to visualize the information from Figures 5.9 to 5.12, we define an arbitrary quantity as

$$a_p = \frac{amp_{peak} - amp_{base}}{amp_{base}} \tag{5.2}$$

where amp_{peak} corresponds to the amplitude of the first peak of the *semi-impulse* responses for any dose other than the 0 mg/ml, and amp_{base} corresponds to the amplitude of the first peak of the *semi-impulse* response of the 0 mg/ml dose.

In Figure 5.13 'a_{p}' is shown for each asthma model based on increasing the methacholine dose. Interestingly enough, the parameter 'a_{p}' is capable of predicting the severity of asthma for the mouse models developed in the laboratory. In other words, it shows that the degree of similarity to the normal mouse decreases from 'Tolerant' to 'Acute' to 'Chronic'. Thus, this mathematical index provides quantitative information about the severity of the disease and it preserves all the information from the semi-impulse responses.

5.1.2 Experiment 2: Model Validation

The third order linear model from (5.1) was fitted to Normal, Chronic and the MEK inhibitor-treated mouse models from Experiment 2 at two available doses of methacholine (0 and 50 mg/ml).

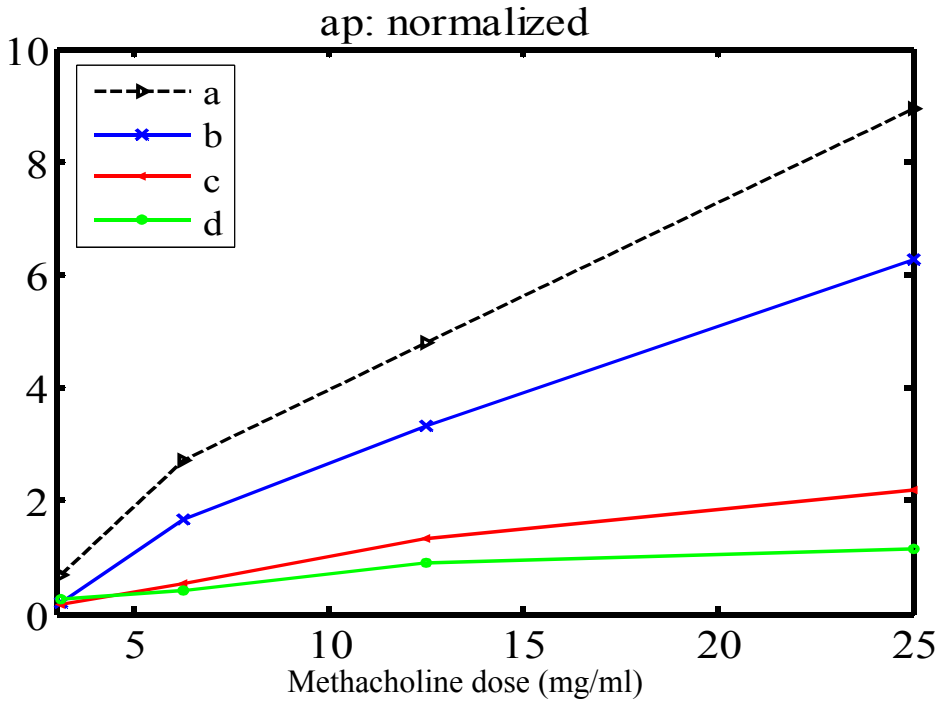


Figure 5.13 The 'a_p' for each mouse model dose-wise a) Chronic, b) Acute, c) Tolerant, d) Normal Hanifi et al. 2012.

In Figure 5.14 the *semi-impulse* response at the baseline level shows that the Normal model and the MEK inhibitor-treated have a similar pattern and they are slightly different from the Chronic asthma model. After the methacholine challenge, one can observe that the three models show more discrimination. The methacholine challenge magnified the difference in behavior. The result is interesting because it shows that the treatment has shifted the Chronic asthma model to a state that is closer to the Normal model.

The significance of analysis provided in Experiment 2 is 1) it validates our proposed model and, 2) the model is applicable to a single dose of methacholine. This

can be very promising for human studies where the complete methacholine dose-response frequently cannot be measured due to the poor lung condition of the patients.

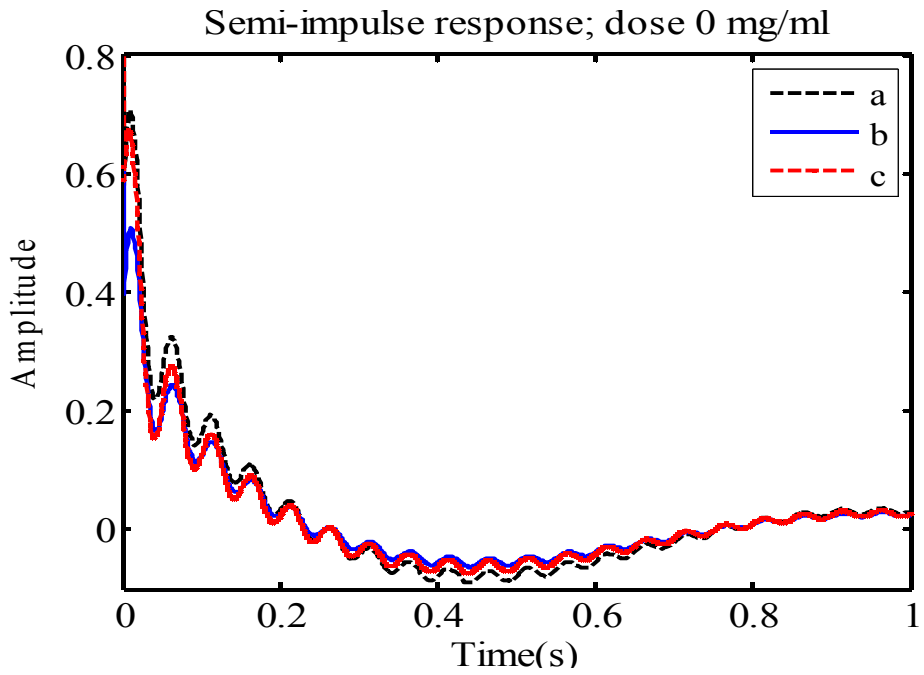


Figure 5.14 Semi-impulse response for the mouse models from experiment 2, a) Normal; b) Chronic c) MEK inhibitor- treated at baseline Hanifi et al. 2012.

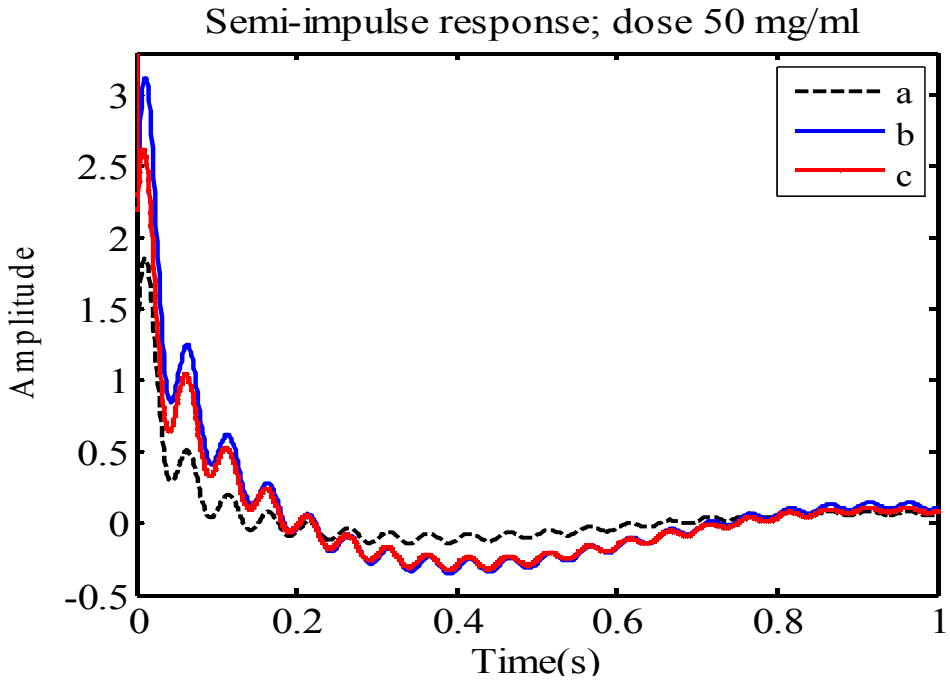


Figure 5.15 Semi-impulse response for the mice groups from experiment 2, a) Normal; b) chronic, c) MEK inhibitor- treated at dose 50 mg/ml Hanifi et al. 2012.

5.1.3 Discussion

In Section 5.1, we designed a complete new analysis system for studying the mouse lung impedance. We:

- identified the respiratory impedance of mouse with a linear parametric model of order 3 (5.1) that fits the data from Experiment 1 and 2 precisely in MATLAB;
- designed a *semi-impulse* waveform that would represent the impulse in the valid frequency range for a mouse lung system;
- simulated the mathematical models with the *semi-impulse* in MATLAB*;*
- designed a parameter 'a_p' to interpret the results of time domain simulations of models with *semi-impulse* based only on the peak values.

We showed that 'a_p' and the time domain responses are capable of providing valuable insight into the systems' behavior. It was concluded that the low order model and the simple time domain analyses are mathematically predictive of the severity of asthma and its response to treatment.

5.2 Human Database

In this part, the results of analysis of human database based on the methods described in Chapter 4 are provided.

5.2.1 PCA

PCA as introduced in Chapter 4, follows two major objectives:

- reducing the number of variables of a dataset while retaining as much variability as possible;
- revealing the hidden patterns in the data, and classifying them according to how much information they account for.

In this research, the main objective for applying the PCA has been to discover the possible hidden pattern of the original databases before clustering them with a secondary method. This results in better clustering of the data compared to the case of clustering the dataset with only a method like K-means or SOMs. Also, for a practical clustering system it is essential to suitably visualize the clustering results.

As stated in Chapter 4, the output of the PCA algorithm includes three matrices in the new space defined by the PCs, coefficient matrix, score matrix and variance vector:

Coefficient matrix – contains coefficients summarizes the linear combinations of the original variables which generate the principal components. It represents the new space and its relation to the original space. Here, it is a 4 by 4 matrix. Table V.II shows the coefficient matrix:

	PC1	PC2	PC3	PC4
1. FEV1	0.14	0.97	-0.13	0.067
2. PC20	0.54	-0.14	-0.68	-0.45

3.Eosinophils	-0.62	-0.14	-0.008	-0.77
4. IgE	-0.54	-0.04	-0.71	0.43

TableV.II. Coefficient matrix from PCA in MATLAB

Scores matrix - This matrix contains the coordinates of the original data in the new coordinate system defined by the principal components. This matrix has the same size as the input data matrix. Figure 5.16 is a plot of the first three columns of scores. It showsthe patients' data projected onto the first three principal components.

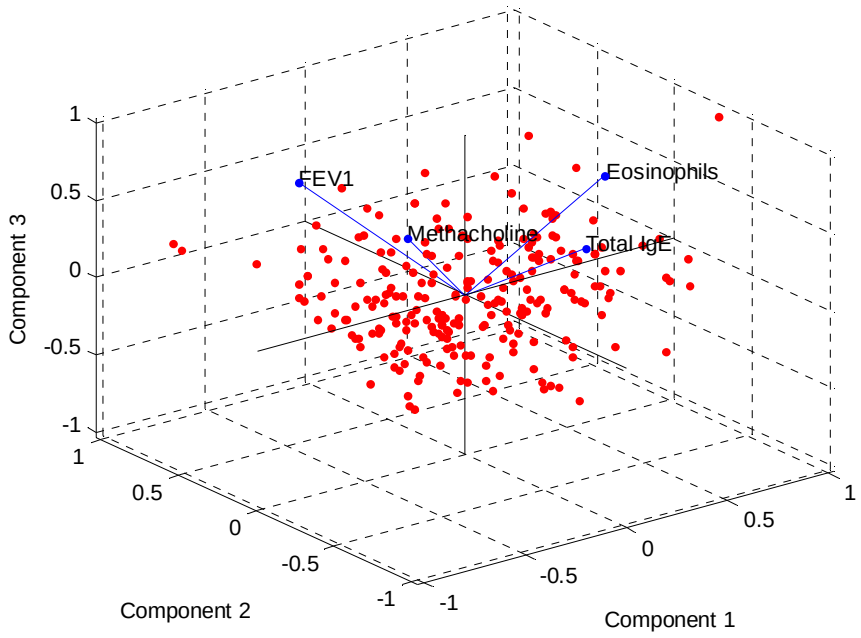


Figure 5.16: Scores in the space of the first three principal components in MATLAB.

Variance vector: This vector contains the variance which is explained by each of the principal components. Each column of *scores* has a sample variance equal to the corresponding element of *variances*. Figure 5.17 shows the percentage of variance (variance (i) over the sum of variances) measured by each principal component.

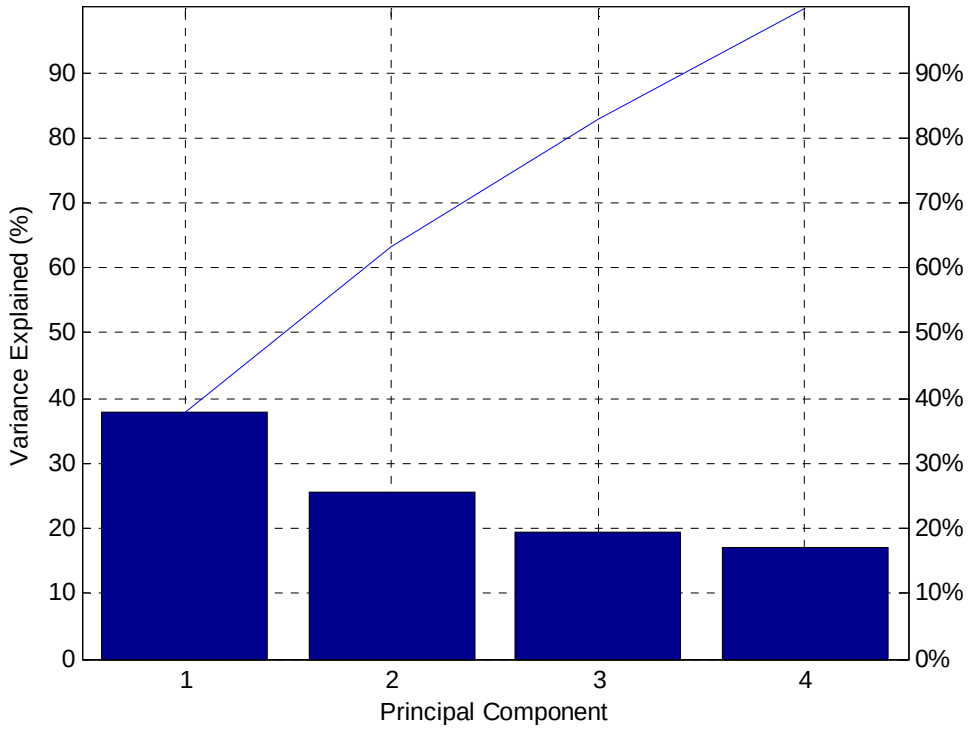


Figure 5.17 Variance vector in MATLAB.

After applying the PCA to the database, the first two PCs were chosen as the dominant PCs for SOM analysis.

5.2.2 Hierarchical Clustering

As described in the SOM introduction, this algorithm needs the number of clusters as an input. To find an appropriate number for the possible clusters we generated a dendrogram shown in Figure 5.18. The dendrogram was generated by the Ward's minimum variance algorithm. As seen on the dendrogram a good choice for number of clusters is five.

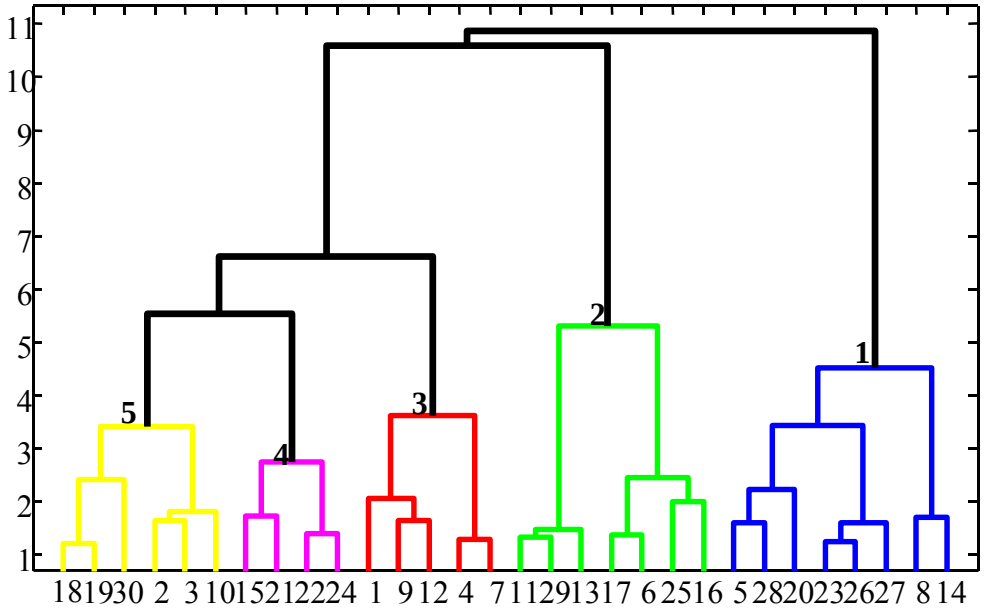


Figure 5.18 Dendrogram generated using Ward's minimum variance algorithm in MATLAB.

5.2.3 SOM Results

Running the SOM algorithm with five clusters, Figure 5.19 was generated. In this step the performance of K-Means was compared to that of the SOMs. SOM was superior. The way SOMs populated its nodes always made the patients that are more "similar" in

their asthma condition cluster together. In Figure 5.19 the five clusters that are automatically generated by the SOM algorithm in MATLAB are manually separated from each other by circles. The circles aim to include the majority of the population and are drawn solely for the purpose of visualizing the relative location of clusters with respect to each other. This solely indicates that the patients out of the circle have not been close to the majority of population. In the case of increasing the number of patients under investigation, we expect the size of the circles to increase as well.

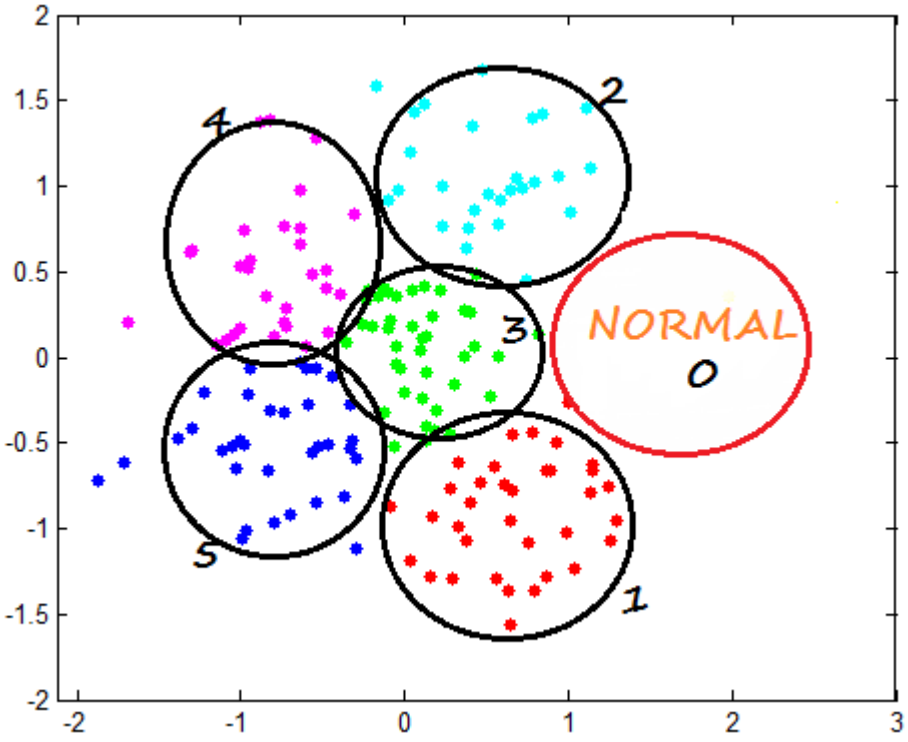


Figure 5.19 clusters of patients generated by SOMs in the space of SOM weight vectors in MATLAB.

Cluster 0: Normal healthy subjects
Cluster 1: Mild asthma
Cluster 2: Non-atopic female-dominant asthma
Cluster 3: Poorly reversible asthma
Cluster 4: Poor lung function- and eosinophilia-associated asthma
Cluster 5: High IgE and hyperreactive airway-associated asthma

5.2.4 Scoring System

In order to find a clinical relevance for the mathematically generated clusters we designed a scoring system. The scoring system is based upon parameters related to the patients' ability to control asthma. The details of scoring system are provided in Table V.III.

Each of the clusters from Figure 5.19 was then analyzed based on the corresponding patients' instability score. Each patient was assigned an instability score based on Table V.III. A patient with an instability score above three is considered an unstable asthmatic patient. The average instability score for each cluster, as well as unstable asthmatic percentage for each cluster, is computed. The average instability score and the instability percentage in clusters are highly correlated. Cluster 5 which includes the subjects with highly eosinophilic asthma has the highest percentage of instability and highest average score. The results of the analyses are summarized in Table V.IV.

Score	0	1	2	3
Use of rescue bronchodilators per week	0 to 2	2 to 5	>5	>5 plus nebulizer
Frequency of systemic steroid treatment per year	None	1 to 2	3 to 4	≥5 or chronic
Frequency of ER/urgent care visit and hospitalization per year	None	1 to 2	3 to 5	≥6
Frequency of >20% drop in FEV1 (variability) per year	None	1	2 to 4	≥5

Table V.III: Asthma instability scoring system

Cluster number	1	2	3	4	5
Cluster name	Mild asthma	Low IgE & female dominant asthma	Poorly reversible asthma	Poor lung function- & eosinophilia associated asthma	High IgE-, eosinophilia & hyperreactive airway-associated asthma
% patients with unstable asthma	0%	9.5%	17.9%	31%	80%
Mean instability score	0.8	1.52	1.52	3.47	4.93
ACT score	20	18	19	14	8
% patients on omalizumab treatment	3	3	4	4	46

Table V.IV Asthma instability score and ACT score by clusters and treatment with omalizumab

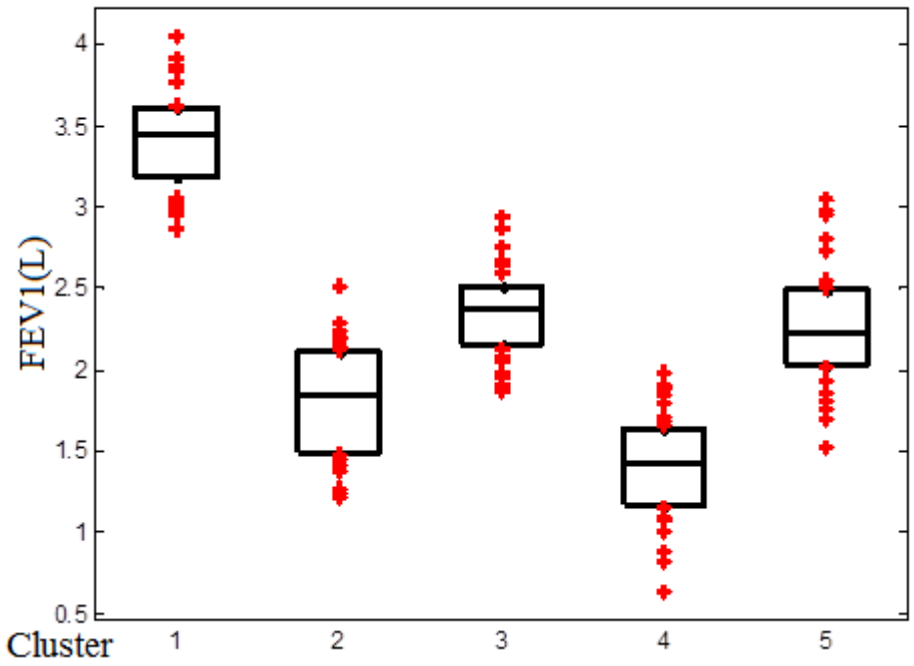


Figure 5.20 The distribution of FEV1 in each cluster

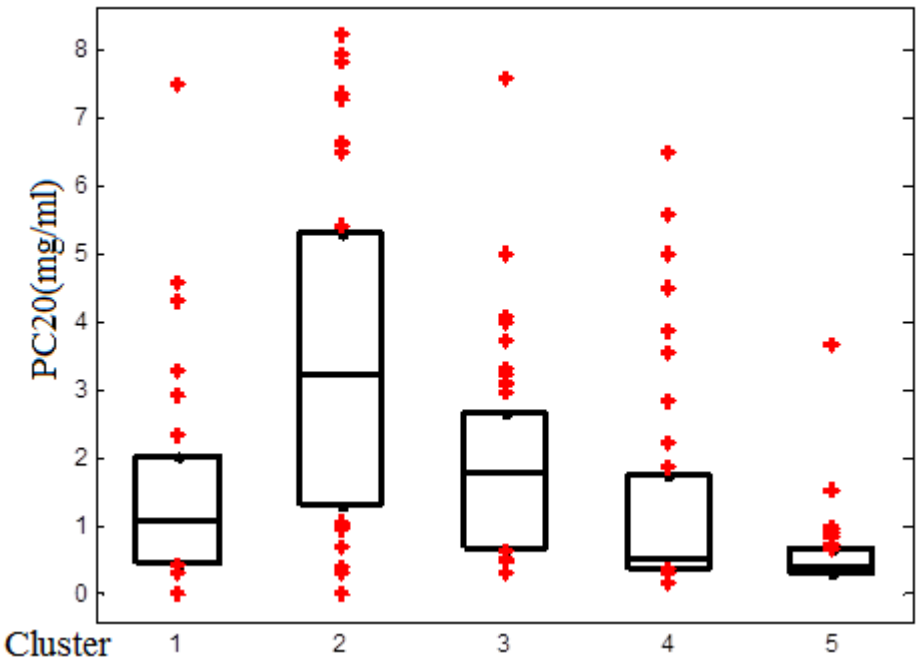


Figure 5.21 The distribution of PC20 for methacholine in each cluster

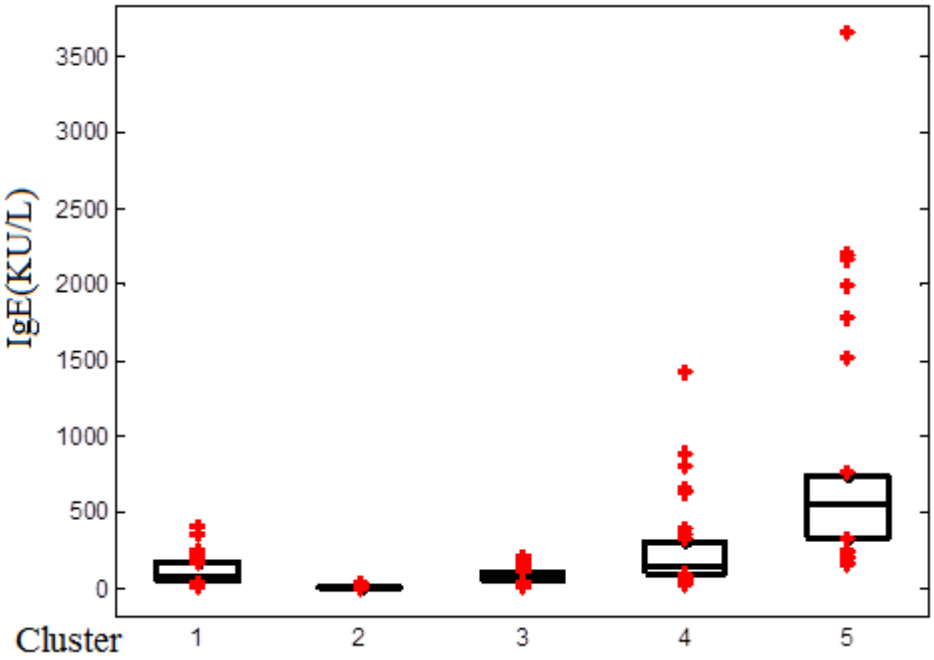


Figure 5.22 Total serum IgE distribution for each cluster

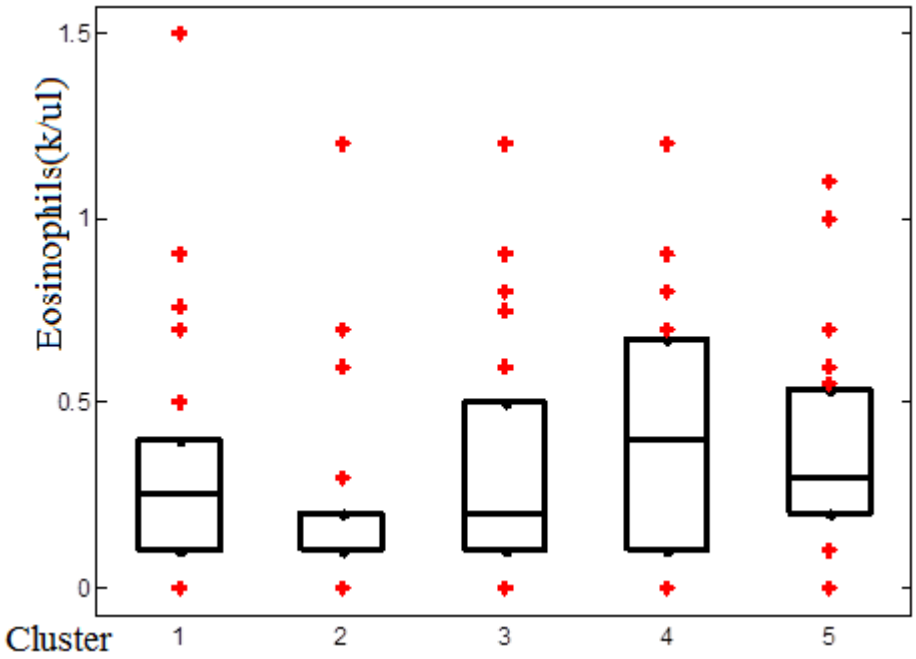


Figure 5.23 Blood eosinophil count in each cluster

Cluster		1	2	3	4	5
N		30	35	42	33	28
Age	Mean	33.96	58	50.1	52	48.35
	Median	33	61	50	55	49.5
Female		55.5%	91%	71.2%	68.5%	50%
BMI	Mean	31.6	30.5	27.68	28.65	30.7
	Median	32	28.75	28	28	30
FEV1 (L)	Mean	3.43	1.8	2.35	1.39	2.26
	Median	3.44	1.85	2.37	1.42	2.2
Reversibility (%)	Mean	9.3	11.7	11	23.93	19.11
	Median	6	12	9	19.5	17
PC20 for Mch (mg/ml)	Mean	1.58	3.6	1.89	1.4	0.67
	Median	1.08	3.21	1.76	0.51	0.39
IgE KIU/L	Mean	114.5	9.25	83	263.95	811.03
	Median	78	7	74.5	152	552
Eosinophils K/μL	Mean	0.33	0.20	0.31	0.44	0.44
	Median	0.25	0.2	0.2	0.4	0.3
Skin test/RAST		30%	30%	40%	53%	86%
FENO (ppb)	Mean	16.6	35	32.5	38	50
	Median	17	35	30	37	47
Descriptive Name of the Cluster		Mild asthma	Low IgE & female dominant asthma	Poorly reversible asthma	Poor lung function- & eosinophilia associated asthma	High IgE-, eosinophilia- & hyperreactive airway-associated asthma

Table V.V Comparison of baseline characteristics of the five asthma clusters.

In Figures 5.20 to 5.23, the distribution of patients within each of the five clusters is shown for each of the four clinical parameters by help of a boxplot. The upper and lower margins of the box represent the 75^{th} and 25^{th} percentile of the data. The patients residing outside this percentile range are shown with plus signs. The line inside the box represents the median value. This demographical presentation helps with understanding how the natural distribution for each of the four clinical parameters is different from each other.

	Mean	Median
Age	48.59	50
Gender	68% female	
BMI	29.28	28.75
FEV1	2.22	2.15
PC20	1.97	1.1
IgE	226.4	90
Eosniophils	0.33	0.2
Reversibility		

Table V.VI Demographics of the population study

5.2.5 Discussion

Through the application of PCA and SOMs we identified five different clusters of asthmatic patients based upon four quantitative variables of asthma—FEV1, PC20 for methacholine, eosinophil count and total IgE. The cluster analysis identified five clinically distinguishable subtypes of asthma: 1) Mild asthma; 2) Low IgE and female-dominant asthma; 3) Poorly reversible asthma; 4) Poor lung function- and eosinophilia-associated asthma; and 5) High IgE-, eosinophilia- and hyperreactive airway-associated asthma. The first two subtypes had stable asthma. Despite poor reversibility, cluster 3

patients also had largely stable asthma. Asthma instability was increased in cluster 4 with poor lung function- and eosinophilia-associated asthma. However, the highest percentage (80%) of unstable asthmatic patients was segregated in cluster 5, which was associated with high IgE, eosinophilia and hyperreactive airways. Thus, our mathematical approach was able to identify patients, who had poorly controlled asthma. The baseline FEV1 of cluster 5 patients was in the middle range for our patient cohort. Cluster 5 also showed the second best reversibility. The foregoing findings suggest that these patients, despite being highly unstable, did not have severely remodeled airways.

Clinical Relevance for our mathematical cluster analysis: we have used the arbitrary score of four variables, weekly usage of rescue bronchodilators, annual systemic steroid usage, ER visits and hospitalization, and frequency of >20% drop in FEV1, to generate a composite annual score of instability. The annual instability score was generated from a retrospective observation period of three or more years. The first three variables are frequently used as a measure of asthma severity. Peak flow variability is also used as a measure of unstable asthma. We used a threshold instability score of 3 from a scale of 0-12 to define unstable asthma. Increasing or decreasing the threshold by one point did not significantly alter the dominant feature of cluster 5 in representing unstable asthma. The ACT score is a validated measure of asthma stability. About 35% of our study patients had ACT score available in the database for analysis. We examined the ACT score in the clusters and compared it with our composite instability score. The ACT score was highest in cluster 1 with mild asthma and lowest in cluster 5. The ACT

score negatively and strongly correlated (r = -0.98, p = 0.001, Pearson correlation test) with the composite instability score(Table V.IV). The composite instability score of asthma correlated (r = 0.93, p = 0.01, Pearson correlation test) with the frequency of omalizumab treatment in the clusters. The frequency of current omalizumab treatment was 3%-4% in clusters 1 through 4. In contrast, 46% of cluster 5 patients were on active omalizumab treatment and 18% of patients tried the treatment but discontinued it due to lack of benefits. The omalizumab treatment reflects physician's judgment about the refractoriness of their patients. Its strong correlation with the instability score provides further evidence that the PCA and SOMs based analysis is a powerful approach to identifying patients with unstable patients.

A low PC20 for methacholine and a high eosinophil count have traditionally been considered high risk factors for severe and unstable asthma (Moore et al., Woodruff et al.). Our mathematical model confirmed this tenet. Table V.V presents the most unstable group with the highest eosinophils count and lowest PC20 value. Asthma is generally considered to be associated with mildly elevated IgE. High total IgE is generally associated with allergic bronchopulmonary aspergillosis (ABPA), atopic dermatitis, food allergy, Churg-Strauss syndrome and parasitic illnesses. Outside the foregoing conditions, a high total IgE level is not widely considered a risk factor for unstable asthma. However, epidemiological studies have shown that an increased total of IgE highly correlates with airway hyperreactivity (low values of PC20 for methacholine) (Burrows et al.). Our results are in agreement with a recent study of steroid therapy-

resistant asthma in children (Bossley et al.). The steroid therapy-resistant unstable asthmatic patients in this cohort had elevated IgE (386 KIU/L) and mild eosinophilia (4.1%), which is similar to what we observed in cluster 5. The major function of an IgE antibody is to activate mast cells. The airway smooth muscle-associated mast cell number has been reported to be increased in asthma (Brightling et al.). Thus, it is possible that IgE contributes to asthma instability primarily by increasing airway hyperreactivity. The seminal role of IgE in controlling asthma stability is supported by the clinical efficacy of omalizumab. Our result suggests that an increase in total IgE above 300 KIU/L (25^{th} percentile of cluster 5) PC20 for methacholine below 0.7 mg/ml (75^{th} percentile of cluster 5) and an eosinophil count above 200/μL (25^{th} percentile of cluster 5) put the patient at risk of having unstable asthma.

Mathematical evaluation of the clustering efficacy with k fold cross validation:

Mathematical evaluation of a predictive model can be done through *cross validation.* It includes methods such as: holdout method, k-fold cross validation and leave one out cross validation. In K-fold cross validation the data set it divided into k subsets and every time one of these subsets is kept for test and the other k-1 sets are considered as training sets. At last the average error for all test sets is computed. k fold cross validation has good performance and not much complexity. For mathematical validation of our clustering accuracy, we performed k fold cross validation. It is very common for to select k = 10 in k fold cross validation. The numerical error returned by this algorithm for

k = 10 in our study is 0.057. This means that our clustering analysis has high accuracy in practice as the error values are significantly low.

Chapter 6

CONCLUSIONS and FUTURE WORK

In the presented research we were able to provide an analysis approach for the mouse lung impedance. We developed a linear model for the mouse lung system and tested the model with several sets of real data from laboratory mouse models of asthma. This technique does the diagnosis and treatment evaluation by simulation of mouse impedance data and through defining an index called 'a_p'. It is simple and provides the possibility of time domain analysis that was not previously available. In comparing the proposed method with the previously reported methods, we observe that the previously reported methods, such as CPM, often study a group of parameters for evaluating the disease severity. Whereas the method we propose is a novel approach in which, by defining a single parameter called '$a_{p,}$' it can predict the disease severity and treatment response.

Our method also provides the possibility of investigating the time domain response for input tests, such as semi-impulse, while in the previously reported methods, the time domain analyses is a very complicated procedure, if at all possible.

This analysis can be extended to human data because similar measurements have been achieved through FOT in humans. In humans the lung input impedance is also

measured in the frequency domain. Using the general model given by (3.4) one can identify the impedance with a different ordered model compared to the mouse model. The proposed technique of predicting semi-impulse response can also be applied for analysis of efficacy of a treatment intervention. It would be interesting to generate a GUI based on our analysis for diagnosis and evaluation of treatment.

In human studies, asthma has traditionally been classified based upon clinical and laboratory features. However, these classifications do not always predict the therapeutic response or long-term clinical stability. Attempts have been made to analyze multidimensional clinical and laboratory data by mathematical approaches to identify asthma subtypes. Clustering data through unsupervised learning approaches is a popular approach. Clustering methods partition data into subsets (clusters) that are thought to exhibit internal cohesion and/or external isolation.

The use of SOMs in our approach identified a single cluster of patients that is distinct not only in baseline disease severity but also in instability. This is different from other previous studies. Severe refractory patients were distributed in more than one cluster in the foregoing studies (Haldar et al., Moore et al.). The identification of a single cluster with unstable asthma is important because it will have therapeutic and management implications. It is of note that all four variables that we used in our analyses were also used in the k-means cluster analyses in the Haldar study. Therefore, the input

data alone cannot explain the difference. There are two major differences between our study and the previous studies with regards to the mathematical analysis as well as the evaluation of the clusters. First, we used a PCA- and SOMs-based cluster analysis method. Second, we applied an annual composite score for unstable asthma.

Unlike other studies, we focused only on a few highly studied clinical parameters. The future stability or controllability of a patient's asthma is a major question for doctors. In this study, by designing a scoring system for our mathematically driven clusters we assigned an instability score to each patient. We used a threshold of 3 from a scale of 0-12 to define unstable asthma. Increasing or decreasing the threshold by one point did not significantly alter the dominant feature of cluster 5 in representing unstable asthma.

A major weakness of this study is its relatively low database size. Increasing the study population size could further strengthen the correlation and the level of statistical significance. Adding normal people to our study is also of great importance for cluster 0. A future goal would be to follow the patients from cluster 5 and prospectively assess their clinical course. Of note, patients from cluster 5 had the second highest median eosinophil count in addition to the highest level of total IgE. Nearly half of the patients from cluster 5 are treated with omalizumab and 18% failed to respond to this treatment. Those who partially benefitted remain unstable despite being on omalizumab. An anti-IL5 antibody has been shown to improve asthma control in refractory eosinophil-rich asthma. It would

be of interest to determine if an anti-eosinophil intervention would reduce the risk of instability in this patient population.

Bibliography

Baswa, S., Diong, B., Nazeram, H., Nava, P., and Goldman, M. (2005). "Evaluation of respiratory system models based on parameter estimates from impulse oscillometry data," in Proc. IEEE Engineering in Medicine and Biology Society 27th Annu. Int. Conf., pp. 2958–2961.

Bateman ED, Bousquet J, Keech ML, Busse WW, Clark TJ, Pedersen SE.(2007)," The correlation between asthma control and health status: the GOAL study," Eur Respir J 29(1):56-62.

Bates, J. H. T. (2009). "Lung Mechanics: An Inverse Modeling approach," (Cambridge University Press, New York).

Bates, J. H. T., Irvin, C. G. (2003). "Measuring lung function in mice: the phenotyping uncertainty principle," J. Appl. Physiol., vol. 94, pp. 1297–1306.

Bates J. H., and Allen, G. B. (2006). "The estimation of lung mechanics parameters in the presence of pathology: A theoretical analysis," Ann. Biomed. Eng., vol. 34, pp. 384–392.

Berry, M. J. A., and Linoff, G. S. (2000). "Mastering Data Mining: The Art and Science of customer relationship management," (Wiley Computer Publishing, USA)

Berger, C. (2005)." Oracle Data Mining, Know More, Do More, Spend Less," An Oracle White Paper, http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/bwp-db-odm-10gr2-0905-132293.pdf?ssSourceSiteId=otncn

Bossley CJ, Fleming L, Gupta A, Regamey N, Frith J, Oates T, Tsartsali L, Lloyd CM, Bush A, Saglani S. (2012). "Pediatric severe asthma is characterized by eosinophilia and remodeling without T(H)2 cytokines," J Allergy Clin Immunol. 2012;129(4):974-82.e13.

Brasier AR, Victor S, Boetticher G, Ju H, Lee C, Bleecker ER, Castro M, Busse WW, Calhoun WJ. (2008). "Molecular phenotyping of severe asthma using pattern recognition of bronchoalveolar lavage-derived cytokines," J Allergy Clin Immunol. Vol. 121, pp. 30-37

Brasier A.R., Victor S., Ju, H., Busse, W. W., Curran-Everett, D., Bleecker, E., Castro, M., Chung, K. F., Gaston, B., Israel, E., Wenzel, S. E., Erzurum, S.C., Jarjour N. N., Calhoun, W. J. (2010). "Predicting intermediate phenotypes in asthma using bronchoalveolar lavage-derived cytokines," Clin Transl Sci.,3(4):147-57.

Breault, J. L., Goodall, C.R., Fos, P.J. (2003) “Data mining a diabetic data warehouse,” J. Artif. Intell. Med., vol. 26, pp. 37-54.

Brightling C. E., Bradding P., Symon F. A., Holgate S.T., Wardlaw A .J., Pavord ID.(2002). “Mast-cell infiltration of airway smooth muscle in asthma,” The New England j. of Med., 346(22):1699-705.

Burrows B, Martinez FD, Halonen M, Barbee RA, Cline MG. “Association of asthma with serum IgE levels and skin-test reactivity to allergens,” The New England J. of Med. 1989;320(5):271-7.

Cavalcanti, J. V., Lopes, A. J., Jansen, J. M., and Melo, P. L. (2006). “Detection of changes in respiratory mechanics due to increasing degrees of airway obstruction in asthma by the forced oscillation technique,” J. Respir. Med., vol. 100, no. 12, pp. 2207–2219.

Coulter, D. M., Bate, A., Meyboom, R.H.B., Lindquist, M., Edwards, I.R.(2001). ”Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study,” British Medical Journal vol. 322, pp. 1207-1209.

Diong, B., Grainger, J., Goldman, M., and Nazeran, H. (2009). “A comparison of linear respiratory system models based on parameter estimates from PRN forced oscillation data,” in Proc. IEEE Engineering in Medicine and Biology Society Annu. Int. Conf., pp. 2879–2882.

Fahy, J. V.(2010). “Identifying clinical phenotypes of asthma, steps in the right direction,” American J. Resp. and critic. Care med. Vol. 181.

Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, Wenzel SE, Aujla S, Castro M, Bacharier LB, Gaston BM, Bleecker ER, Moore WC; National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. J Allergy Clin Immunol. 2011 Feb;127(2):382-389.e1-13.

Franklin, G., Powell, J., and Workman, M. (1998). “Digital Control of Dynamical Systems,” 3rd ed, (Ellis-Kagle, Half Moon Bay, CA).

Frey U, Brodbeck T, Majumdar A, Taylor DR, Town GI, Silverman M, et al. (2005). “Risk of severe asthma episodes predicted from fluctuation analysis of airway function,” Nature. 438(7068):667-70.

Glaab, T., Taube, C., Braun, A., and Mitzner, W. (2007). "Invasive and noninvasive methods for studying pulmonary function in mice," Respir. Res. 8, 63.

Goldman, M. D., Nazeran, H., Ramos, C., Toon, E., Oates, K., Bilton, D., Meraz, E., Hafezi, N., and Diong, B. (2010). "Electrical circuit models of the human respiratory system reflect small airway impairment measured by impulse oscillation (IOS)," in Proc. IEEE Engineering in Medicine and Biology Society 32nd Annu. Int. Conf., Buenos Aires, Argentina.

Goplen, N., Karim, M. Z., Liang, Q., Gorska, M. M., Rozario, S., Guo, L., and Alam, R. (2009). "Combined sensitization of mice to extracts of dust mite, ragweed, and Aspergillus species breaks through tolerance and establishes chronic features of asthma," J. Allergy Clin. Immunol., vol. 123, no. 4, pp. 925–32.e11.

Guthikonda, S. M. (2005) "Kohonen Self-Organizing Maps," http://www.shy.am/wp-content/uploads/2009/01/kohonen-self-organizing-maps-shyam-guthikonda.pdf

Haldar P, Pavord, ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green, R. H. (2008). "Cluster analysis and clinical asthma phenotypes," Am J Respir Crit Care Med., vol. 178, pp. 218-24.

Han, J., and Kamber, M. (2001). "Data Mining: Concepts and Techniques",Morgan Kaufmann Publishers.

Hand, D., Mannila, H., and Smyth, P. (2001)." Principles of data mining," (MIT Press, Cambridge, MA).

Hantos, Z., Daroczy, B. Suki, B., Nagy, S., and Fredberg, J. J. (1992). "Input impedance and peripheral inhomogeneity of dog lungs," J. Appl. Phys., vol. 72, pp. 168–178.

Haselkorn T, Zeiger RS, Chipps BE, Mink DR, Szefler SJ, Simons FE, et al. (2009). "Recent asthma exacerbations predict future exacerbations in children with severe or difficult-to-treat asthma," J. of allergy and clinical immunology. 124(5):921-7.

Ionescu, C., Desager, K., and Keyser, R. D. (2011). "Fractional order model parameters for the respiratory input impedance in healthy and in asthmatic children," J. Comput. Methods Programs Biomed., pp. 315–323.

Ionescu, C. M., Machado, J. A. T., and Keyser, R. D. (2011). "Modeling of the lung impedance using a fractional-order ladder network with constant phase elements," IEEE Trans. Biomed. Circuits Syst., vol. 5, no. 1, pp. 83–89.

Jablonski, I., Polak, A. G., and Mroczka, J. (2011). "Preliminary study on the accuracy of respiratory input impedance measurement using the interrupter technique," J. Comput. Methods Programs Biomed., pp. 115–125.

Kantardzic, M., Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition, IEEE Press & John Wiley, August 2011.

Khajehei, M., and Etemady, F. (2010). "Data Mining and Medical Research Studies," CIMSim, IEEE, Bali, Indonesia.

Kohonen, T. (1982). "Self-organized formation of topologically correct feature maps," Biological cybernetics, Vol. 43, pp. 159-69, DOI: 10.1007/BF00337288

Kusiak, A., Dixon, B., Shah, S. (2005). "Predicting survival time for kidney dialysis patients: a data mining approach," J. Comput. Biol. Med., vol. 35, pp. 311-327.

Lauzon, A. M., and Bates, J. H. (1991). "Estimation of time-varying respiratory mechanical parameters by recursive least squares," J. Appl. Phys., vol. 71, pp. 1159–1165.

Lehmann, C., Khawja, A., (2011),"A Novel Multi-lead Method for Clustering Ventricular Ectopic Heartbeats," IEEE Computing in Cardiology, pp. 749 - 752

Ljung, L., and Glad, T. (1994). "Modeling of Dynamic Systems," (Prentice-Hall, Englewood Cliffs,NJ).

Ljung, L. (1987). "System Identification Theory for the User," (Prentice-Hall, Englewood Cliffs, NJ).

Li, L., Tang, H., Wu, Z., Gong, J., Gruidl, M., Zou, J., Tockman, M., Clark, R. A. (2004). "Data mining techniques for cancer detection using serum proteomic profiling," Artificial Intelligence in Medicine, vol. 32, pp. 71-83.

Li, M., Chen, C., and Liu, W. (2009). "Identification based on MATLAB," presented at the Int. Workshop Information Security and Application , Qingdao, China, http://www.academypublisher.com/proc/iwisa09/papers/iwisa09p523.pdf

Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R Jr, Castro M, Curran-Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair SA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleecker ER; National Heart, Lung, and Blood Institute's Severe Asthma Research Program (2010). "Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program," Am J Respir Crit Care Med. 181(4):315-23.

Moriya, H. T., Moraes, J. T. B., and Bates, J. H. T. (2003). "Nonlinear and frequency-dependent mechanical behavior of the mouse respiratory system," Ann. Biomed. Eng., vol. 31, pp. 318–326.

Mdzingwa N., (2005), Literature Review Data Mining with Oracle 10g using Clustering and Classification Algorithms, http://www.cs.ru.ac.za/research/g05m5125/documents/Literature%20Review.pdf

Pauwels, R. A., Pedersen, S., Busse, W. W. (2003), "Early intervention with budesonide in mild persistent asthma: a randomised, double-blind trial." Lancet, 361(9363):1071-6.

Pedersen S., (2010). "From asthma severity to control: a shift in clinical practice", Prim Care Respir J. , 19(1):3-9.

Pedersen SE, Bateman ED, Bousquet J, Busse WW, Yoxall S, Clark TJ. (2007),"Determinants of response to fluticasone propionate and salmeterol/fluticasone propionate combination in the Gaining Optimal Asthma controL study," J Allergy Clin Immunol;120(5):1036-42.

Peslin, R., and Fredberg, J. J. (1986). "Oscillation mechanics of the respiratory system," In Handbook of Physiology. Section 3: The Respiratory System. P. Macklem and J. Mead, Editors, Bethesda, MD: American Physiological Society, p. 145–178.

Reddel H. K., Taylor D. R., Bateman E. D. (2009), "An Official American Thoracic Society/European Respiratory Society Statement: Asthma Control and Exacerbations: Standardizing Endpoints for Clinical Asthma Trials and Clinical Practice," Am J Respir Crit Care Med, 180:59-99.

Roiger, J., and Geatz, M. W. (2003). "Data Mining: A Tutorial-Based Primer", (International Edition), Pearson Education, USA.

Schuessler, T. F., and Bates, J. H. T. (1995). “A computer-controlled research ventilator for small animals: Design and evaluation,” IEEE Trans. Biomed. Eng., vol. 42, pp. 860–866.

Schlosser, G., Wagner, G. P., (2004).”Modularity in development and evolution,” (University of Chicago Press).

Shlens, J., (2009). “A Tutorial on Principal Component Analysis,” http://www.snl.salk.edu/~shlens/pca.pdf

Siroux, V., Basagaña, X., Boudier, A., Pin, I., Garcia-Aymerich, J., Vesin, A., Slama, R., Jarvis, D., Anto, J. M. , Kauffmann, F., Sunyer, J. (2011). “Identifying adult asthma phenotypes using a clustering approach,” Eur Respir J., 38(2):310-7.

Sutherland E. R., Goleva, E., King T. S., Lehman E., Stevens A. D., Jackson L. P., Stream, A. R., Fahy, J. V., Leung, D. Y. (2012). “ Asthma Clinical Research Network. Cluster analysis of obesity and asthma phenotypes,” PLoS One, 7(5):e36631.

Söderström, T., and Stoica, P. (1989). “System Identification,” (Prentice-Hall International, London, U.K).

Taval, M. A., Raghuwanshi, M. M., (2010). “Review on Various Clustering Methods for the Image Data,” J. of Emerging trends in computing and information, vol. 2, pp. 34-38.

Tiffin, N., Kelso, J. F., Powell, A. R., Pan, H., Bajic, V. B., and Hide, W. A. (2005). “Integration of text- and data-mining using ontologies successfully selects disease gene candidates,” Nucl. Acids Res., vol. 33, pp. 1544-1552.

Tomalak, W., Peslin, R., Duvivier, C., and Gallina, C. (1993). “Optimal frequency range to analyze respiratory transfer impedance with sixelement model,” J. Appl. Phys., vol. 75, pp. 2656–2664.

Tu, Y. P., Larsen, G., and Irvin, C. G. (1995). “Utility of murine systems to study asthma pathogenesis,” J. Eur. Respir. Rev., vol. 29, pp. 224–230.

Weiss, S. M., and Kulikowski, C. A. (1991). “Computer Systems That Learn”, Morgan Kaufmann Publishers, San Mateo.

Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, et al. (2009). “T-helper type 2-driven inflammation defines major subphenotypes of asthma,” American j. resp. and crit. care med. 180(5):388-95.

Zwart, A., and Vaqn deWoestijne, (1994). “Mechanical respiratory impedance by forced oscillation,” Eur.Resp. Rev., vol. 4, p. 114–237.

Appendix A
Publication list

Hanifi, A., Goplen, N., Matin, M., Salters, R. E., and Alam, R. (2012). "A Linear Parametric Approach for Analysis of Mouse Respiratory Impedance," IEEE Trans. Biomed. Circ. Sys., vol. 6 no. 3, pp. 287-294.

Hanifi A., Salters R. Matin M. A., and Alam R. (2010). "On Extracting Information from Mine-able Asthma Database", Proceedings, International Conference on Bioinformatics, Computational Biology, Genomics and Chemo informatics, ISBN: 978-1-60651-017-9, pp. 20-24.

Hanifi, A., Sheth, D., Hum, R., Salters, R. E., Katial, R., Matin, M.A., and Alam, R., (2012). "A Multifaceted Mathematical Analysis of an Asthma Database Identifies Factors Associated with Asthma Instability," to be submitted.

Appendix B
Mouse Lung impedance modeling, MATLAB codes.

```
    clc;clear;
    close all;
    load realZ.mat;
    load imagZ.mat;
    f = [1 1.5 2.5 3.5 5.5 6.5 8.5 9.5 11.5 14.5 15.5 18.5
20.5];f = 2*pi*f;
    meth = [0 3.125 6.25 12.5 25];
    Ts = 0.001;
    mag0  =sqrt( meth0R.^2  + meth0I.^2);
    mag3  =sqrt( meth3R.^2  + meth3I.^2);
    mag6  =sqrt( meth6R.^2  + meth6I.^2);
    mag12 =sqrt( meth12R.^2 + meth12I.^2);
    mag25 =sqrt( meth25R.^2 + meth25I.^2);
    magBL =sqrt( BLpostR.^2 + BLpostI.^2);
    ph0  = atan(meth0I ./meth0R) .* 180/pi;
    ph3  = atan(meth3I ./meth3R) .* 180/pi;
    ph6  = atan(meth6I ./meth6R).* 180/pi;
    ph12 = atan(meth12I./meth12R).* 180/pi;
    ph25 = atan(meth25I./meth25R).* 180/pi;
    phBL = atan(BLpostI./BLpostR).* 180/pi;
    magSaline =
[mag0(:,4),mag3(:,4),mag6(:,4),mag12(:,4),mag25(:,4)];
    magDRA =
[mag0(:,1),mag3(:,1),mag6(:,1),mag12(:,1),mag25(:,1)];
    magAcute =
[mag0(:,2),mag3(:,2),mag6(:,2),mag12(:,2),mag25(:,2)];
    magTol =
[mag0(:,3),mag3(:,3),mag6(:,3),mag12(:,3),mag25(:,3)];
    subplot 121
    mesh((magSaline),'EdgeColor','black');hold on;
    mesh((magDRA),'EdgeColor','green');
    mesh((magAcute),'EdgeColor','red');
    mesh(magTol,'EdgeColor','yellow');title('magnitudes of
the four
groups');legend('Normal','DRA','Acute','Tolerance');
    phSaline =
[ph0(:,4),ph3(:,4),ph6(:,4),ph12(:,4),ph25(:,4)];
    phDRA =
[ph0(:,1),ph3(:,1),ph6(:,1),ph12(:,1),ph25(:,1)];
    phTol =
[ph0(:,3),ph3(:,3),ph6(:,3),ph12(:,3),ph25(:,3)];
```

```
    phAcute =
[ph0(:,2),ph3(:,2),ph6(:,2),ph12(:,2),ph25(:,2)];
    subplot 122
    mesh((phSaline),'EdgeColor','black');hold on;
    mesh((phDRA),'EdgeColor','green');
    mesh((phAcute),'EdgeColor','red');
    mesh(phTol,'EdgeColor','yellow');title('phases of the
four groups');legend('Normal','DRA','Acute','Tolerance');
    %%
    %set(gca,'DefaultLineLineWidth',10)
    set(0,'DefaultAxesLineStyleOrder','-|--|:|-.');
    %set(0,'DefaultAxesColorOrder',[0 0 0],...
    %       'DefaultAxesLineStyleOrder','-|--|:|-.')
    set(0,'defaultaxesfontsize',12);
    close all;

    %% DRA
    zfr= meth0R(:,1)+ i* meth0I(:,1);
    dra0= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(dra0,'g--',{1 ,1000});
    h = findobj(gcf,'type','line'); set(h,'linewidth',2);
    subplot 211
    title('DRA methacoline dose 0');hold
on;semilogx(f,(mag0(:,1)),'r*-');
    subplot 212
    hold on;semilogx(f,ph0(:,1) ,'r*-');current_axis =
gca;h = findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0);
    axes(current_axis);
    %%
    figure
    zfr= meth3R(:,1)+ i* meth3I(:,1);
    dra3 = oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(dra3,'g--',{1 ,1000});
    h = findobj(gcf,'type','line');set(h,'linewidth',2);
    subplot 211
    title('DRA methacoline dose 3');hold
on;semilogx(f,(mag3(:,1)),'r*-');
    subplot 212
    hold on;semilogx(f,ph3(:,1) ,'r*-'); current_axis =
gca;  h = findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth6R(:,1)+ i* meth6I(:,1);
```

```
    dra6= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(dra6,'g--',{1 ,1000});
    h = findobj(gcf,'type','line');set(h,'linewidth',2);
    subplot 211
    title('DRA methacoline dose 6');
    hold on
    semilogx(f,(mag6(:,1)),'r*-');
    subplot 212
    hold on;semilogx(f,ph6(:,1) ,'r*-'); current_axis =
gca;  h = findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth12R(:,1)+ i* meth12I(:,1);
    dra12= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(dra12,'g--',{1 ,1000});
    h = findobj(gcf,'type','line');
    set(h,'linewidth',2);
    subplot 211
    title('DRA methacoline dose 12');hold
on;semilogx(f,(mag12(:,1)),'r*-');
    subplot 212
    hold on;semilogx(f,ph12(:,1) ,'r*-'); current_axis =
gca;  h = findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth25R(:,1)+ i* meth25I(:,1);
    dra25= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(dra25,'g--',{1 ,1000});
    h = findobj(gcf,'type','line');
    set(h,'linewidth',2);
    subplot 211
    title('DRA methacoline dose 25');
    hold on
    semilogx(f,(mag25(:,1)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph25(:,1) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%
```

```
    %%
    % Acute
    figure
    zfr= meth0R(:,2)+ i* meth0I(:,2);
    acute0= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(acute0,'g--',{1 ,1000});
    h = findobj(gcf,'type','line');
    set(h,'linewidth',2);
    subplot 211
    title('Acute methacoline dose 0');
    hold on
    semilogx(f,(mag0(:,2)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph0(:,2) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth3R(:,2)+ i* meth3I(:,2);
    acute3= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(acute3,'g--',{1 ,1000});
    h = findobj(gcf,'type','line');
    set(h,'linewidth',2);
    subplot 211
    title('Acute methacoline dose 3');
    hold on
    semilogx(f,(mag3(:,2)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph3(:,2) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth6R(:,2)+ i* meth6I(:,2);
    acute6= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(acute6,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Acute methacoline dose 6');
    hold on
    semilogx(f,(mag6(:,2)),'r*-');
    subplot 212
    hold on
```

```
    semilogx(f,ph6(:,2) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth12R(:,2)+ i* meth12I(:,2);
    acute12= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(acute12,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Acute methacoline dose 12');
    hold on
    semilogx(f,(mag12(:,2)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph12(:,2) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth25R(:,2)+ i* meth25I(:,2);
    acute25= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(acute25,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Acute methacoline dose 25');
    hold on
    semilogx(f,(mag25(:,2)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph25(:,2) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    % Tolerance
    figure
    zfr= meth0R(:,3)+ i* meth0I(:,3);
    tol0= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(tol0,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Tolerance methacoline dose 0');
    hold on
    semilogx(f,(mag0(:,3)),'r*-');
    subplot 212
```

```
    hold on
    semilogx(f,ph0(:,3) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth3R(:,3)+ i* meth3I(:,3);
    tol3= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(tol3,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Tolerance methacoline dose 3');
    hold on
    semilogx(f,(mag3(:,3)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph3(:,3) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth6R(:,3)+ i* meth6I(:,3);
    tol6= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(tol6,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Tolerance methacoline dose 6');
    hold on
    semilogx(f,(mag6(:,3)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph6(:,3) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth12R(:,3)+ i* meth12I(:,3);
    tol12= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(tol12,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Tolerance methacoline dose 12');
    hold on
    semilogx(f,(mag12(:,3)),'r*-');
    subplot 212
```

```
    hold on
    semilogx(f,ph12(:,3) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth25R(:,3)+ i* meth25I(:,3);
    tol25= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(tol25,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');set(h,'linewidth',2);
    subplot 211
    title('Tolerance methacoline dose 25');
    hold on
    semilogx(f,(mag25(:,3)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph25(:,3) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    % Normal
    figure
    zfr= meth0R(:,4)+ i* meth0I(:,4);
    normal0= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(normal0,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');set(h,'linewidth',2);
    subplot 211
    title('Normal methacoline dose 0');
    hold on
    semilogx(f,(mag0(:,4)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph0(:,4) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth3R(:,4)+ i* meth3I(:,4);
    normal3= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(normal3,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Normal methacoline dose 3');
    hold on
    semilogx(f,(mag3(:,4)),'r*-');
```

```
    subplot 212
    hold on
    semilogx(f,ph3(:,4) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth6R(:,4)+ i* meth6I(:,4);
    normal6= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(normal6,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Normal methacoline dose 6');
    hold on
    semilogx(f,(mag6(:,4)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph6(:,4) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth12R(:,4)+ i* meth12I(:,4);
    normal12= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(normal12,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Normal methacoline dose 12');
    hold on
    semilogx(f,(mag12(:,4)),'r*-');
    subplot 212
    hold on
    semilogx(f,ph12(:,4) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%
    figure
    zfr= meth25R(:,4)+ i* meth25I(:,4);
    normal25= oe(idfrd(zfr,f,Ts),[4 2 0],'Focus','Stab');
    bode(normal25,'g--',{1 ,1000}); h =
findobj(gcf,'type','line');  set(h,'linewidth',2);
    subplot 211
    title('Normal methacoline dose 25');
    hold on
    semilogx(f,(mag25(:,4)),'r*-');
```

```
    subplot 212
    hold on
    semilogx(f,ph25(:,4) ,'r*-'); current_axis = gca;  h =
findall(gcf,'type','axes','visible','on');
    set(h,'linewidth',2.0); axes(current_axis);
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
    %%
    close all
    t = -8:.001:4;
    % k = size(t);
    ut=t.*0;

     for k=1:.05:20

    ut = ut+cos(2*pi*k*t);
     end
    % xx=0;
    % %  gy=16/.001+1;
    % %  for jj=1:gy
    % %
    % %  ut(jj) = xx+1*ut(jj);
    % %  xx=ut(jj);
    % %  end
    % % ut=-ut;
    %
    figure
    plot(t,ut);
    %  ut = 1/1*sin(2*pi*1*t)+ 1/3* sin(2*pi*3*t)+
1/5*sin(2*pi*5*t) ...
    %     +1/7*sin(2*pi*7*t)+ 1/9* sin(2*pi*9*t)+ 1/11*
sin(2*pi*11*t)...
    %         +1/13* sin(2*pi*13*t) + 1/15*
sin(2*pi*15*t)...
    %      +1/17*sin(2*pi*17*t)+ 1/19* sin(2*pi*19*t) ...
    %       +1/21*sin(2*pi*21*t)+ 1/23* sin(2*pi*23*t);
    %
    ut = cos(2*pi*1*t)+ cos(2*pi*2*t) + cos(2*pi*3*t)+
cos(2*pi*4*t)+ cos(2*pi*5*t)+cos(2*pi*6*t) ...
        +cos(2*pi*7*t)+ cos(2*pi*8*t) + cos(2*pi*9*t)+
cos(2*pi*10*t)+ cos(2*pi*11*t)...
            +cos(2*pi*12*t)+ cos(2*pi*13*t) +
cos(2*pi*14*t)+ cos(2*pi*15*t)+ cos(2*pi*16*t)...;
```

```
         +cos(2*pi*17*t)+ cos(2*pi*18*t) + cos(2*pi*19*t)+
cos(2*pi*20*t)...

         + cos(2*pi*1.5*t)+ cos(2*pi*2.5*t) +
cos(2*pi*3.5*t)+ cos(2*pi*4.5*t)+
cos(2*pi*5.5*t)+cos(2*pi*6.5*t) ...
        +cos(2*pi*7.5*t)+ cos(2*pi*8.5*t) +
cos(2*pi*9.5*t)+ cos(2*pi*10.5*t)+ cos(2*pi*11.5*t)...
            +cos(2*pi*12.5*t)+ cos(2*pi*13.5*t) +
cos(2*pi*14.5*t)+ cos(2*pi*15.5*t)+ cos(2*pi*16.5*t)...
         +cos(2*pi*17.5*t)+ cos(2*pi*18.5*t) +
cos(2*pi*19.5*t)+cos(2*pi*20.5*t);
    % % xx=0;
    %  gy=9/.001+1;
    %  for jj=1:gy
    %
    %  ut(jj) = xx+1*ut(jj);
    %  xx=ut(jj);
    %  end
    % ut=-ut;
    %

     ut = ut*.01;
     gg=size(t);
    gg =8000-8;
    %%
    figure
    subplot 321
    plot(t(gg:end),ut(gg:end),'m-
','LineWidth',2);title('test input');axis tight;grid on
    subplot 322
    ydra0 =     sim(dra0,ut');
plot(t(gg:end),ydra0(gg:end),'r.-','LineWidth',2);hold on;
%axis([0 10 -200 200])
    yacute0 =   sim(acute0,ut');
plot(t(gg:end),yacute0(gg:end),'g--','LineWidth',2);hold
on; %axis([0 10 -400 400])
    ytol0 =     sim(tol0,ut');
plot(t(gg:end),ytol0(gg:end),'b-','LineWidth',2);hold on;
%axis([0 10 -200 200])
    ynormal0 =  sim(normal0,ut');
plot(t(gg:end),ynormal0(gg:end),'c--','LineWidth',2);hold
on; %axis([0 10 -100 100])
    title('Baseline');axis tight;grid on% legend(' dra0','
acute0','tolerance0','normal0');
```

```
    legend(' dra',' acute','tolerance','normal');
    % figure
    subplot 323
    ydra3 =     sim(dra3,ut');
plot(t(gg:end),ydra3(gg:end),'r.-','LineWidth',2);hold on;
%axis([0 10 -200 200])
    yacute3 =   sim(acute3,ut');
plot(t(gg:end),yacute3(gg:end),'g--','LineWidth',2);hold
on; %axis([0 10 -400 400])
    ytol3 =     sim(tol3,ut');
plot(t(gg:end),ytol3(gg:end),'b--','LineWidth',2);hold on;
%axis([0 10 -200 200])
    ynormal3 =  sim(normal3,ut');
plot(t(gg:end),ynormal3(gg:end),'c--','LineWidth',2);hold
on; %axis([0 10 -100 100])
    % legend(' dra3',' acute3','tolerance3','normal3');
    title('Dose 3mg');axis tight;grid on
    % figure
    subplot 324
    ydra6 =     sim(dra6,ut');
plot(t(gg:end),ydra6(gg:end),'r.-','LineWidth',2);hold on;
%axis([0 10 -200 200])
    yacute6 =   sim(acute6,ut');
plot(t(gg:end),yacute6(gg:end),'g--','LineWidth',2);hold
on; %axis([0 10 -400 400])
    ytol6 =     sim(tol6,ut');
plot(t(gg:end),ytol6(gg:end),'b--','LineWidth',2);hold on;
%axis([0 10 -200 200])
    ynormal6 =  sim(normal6,ut');
plot(t(gg:end),ynormal6(gg:end),'c--','LineWidth',2);hold
on; %axis([0 10 -100 100])
    % legend(' dra6',' acute6','tolerance6','normal6');
    title('Dose 6mg');axis tight;grid on
    % figure
    subplot 325
    ydra12 =    sim(dra12,ut');
plot(t(gg:end),ydra12(gg:end),'r.-','LineWidth',2);hold on;
%axis([0 10 -200 200])
    yacute12 =  sim(acute12,ut');
plot(t(gg:end),yacute12(gg:end),'g--','LineWidth',2);hold
on; %axis([0 10 -400 400])
    ytol12 =    sim(tol12,ut');
plot(t(gg:end),ytol12(gg:end),'b--','LineWidth',2);hold on;
%axis([0 10 -200 200])
```

```
    ynormal12 = sim(normal12,ut');
plot(t(gg:end),ynormal12(gg:end),'c--','LineWidth',2);hold
on; %axis([0 10 -100 100])
    % legend(' dra12',' acute12','tolerance12','normal12');
    title('Dose 12mg');axis tight;grid on
    % figure
    subplot 326
    ydra25 =    sim(dra25,ut');
plot(t(gg:end),ydra25(gg:end),'r-','LineWidth',2);hold on;
%axis([0 10 -200 200])
    yacute25 =  sim(acute25,ut');
plot(t(gg:end),yacute25(gg:end),'g--','LineWidth',2);hold
on; %axis([0 10 -400 400])
    ytol25 =    sim(tol25,ut');
plot(t(gg:end),ytol25(gg:end),'b--','LineWidth',2);hold on;
%axis([0 10 -200 200])
    ynormal25 = sim(normal25,ut');
plot(t(gg:end),ynormal25(gg:end),'c--','LineWidth',2);hold
on; %axis([0 10 -100 100])
    % legend(' dra25',' acute25','tolerance25','normal25');
    title('Dose 25mg');axis tight;grid on
    %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%

    %%
    %figure
    % subplot 221
    % ydra0 = sim(dra0,ut'); plot(t(k(2)-
4000:k(2)),ydra0(k(2)-4000:k(2)),'r--','LineWidth',2);hold
on;
    % ydra3 = sim(dra3,ut'); plot(t(k(2)-
4000:k(2)),ydra3(k(2)-4000:k(2)),'g.-','LineWidth',2);
    % ydra6 = sim(dra6,ut');plot(t(k(2)-
4000:k(2)),ydra6(k(2)-4000:k(2)),'c--','LineWidth',2);
    % ydra12 = sim(dra12,ut');plot(t(k(2)-
4000:k(2)),ydra12(k(2)-4000:k(2)),'y-','LineWidth',2);
    % ydra25 = sim(dra25,ut');plot(t(k(2)-
4000:k(2)),ydra25(k(2)-4000:k(2)),'b-','LineWidth',2);
    figure
    % subplot 321

    ydra0 = sim(dra0,ut'); plot(t(gg:end),ydra0(gg:end),'r-
-','LineWidth',2);hold on;
```

```
    ydra3 = sim(dra3,ut');
plot(t(gg:end),ydra3(gg:end),'g.-','LineWidth',2);%axis([0
10 -50 50])
    % subplot 323
    ydra6 = sim(dra6,ut');plot(t(gg:end),ydra6(gg:end),'c--
','LineWidth',2);%axis([0 10 -50 50])
    % subplot 324
    ydra12 =
sim(dra12,ut');plot(t(gg:end),ydra12(gg:end),'y-
','LineWidth',2);%axis([0 10 -50 50])
    % subplot 325
    ydra25 =
sim(dra25,ut');plot(t(gg:end),ydra25(gg:end),'b-
','LineWidth',2);%axis([0 10 -50 50])

    title('DRA dose response to u(t)');legend(' dra0','
dra3','dra6','dra12','dra25');
    grid on;

    % figure
    % subplot 222
    % yacute0 = sim(acute0,ut'); plot(t(k(2)-
4000:k(2)),yacute0(k(2)-4000:k(2)),'r--
','LineWidth',2);hold on;
    % yacute3 = sim(acute3,ut'); plot(t(k(2)-
4000:k(2)),yacute3(k(2)-4000:k(2)),'g.-','LineWidth',2);
    % yacute6 = sim(acute6,ut');plot(t(k(2)-
4000:k(2)),yacute6(k(2)-4000:k(2)),'c--','LineWidth',2);
    % yacute12 = sim(acute12,ut');plot(t(k(2)-
4000:k(2)),yacute12(k(2)-4000:k(2)),'y-','LineWidth',2);
    % yacute25 = sim(acute25,ut');plot(t(k(2)-
4000:k(2)),yacute25(k(2)-4000:k(2)),'b-','LineWidth',2);
    % yacute0 = sim(acute0,ut'); plot(t,yacute0,'r--
','LineWidth',2);hold on;
    % yacute3 = sim(acute3,ut'); plot(t,yacute3,'g.-
','LineWidth',2);
    % yacute6 = sim(acute6,ut');plot(t,yacute6,'c--
','LineWidth',2);
    % yacute12 = sim(acute12,ut');plot(t,yacute12,'y-
','LineWidth',2);
    % yacute25 = sim(acute25,ut');plot(t,yacute25,'b-
','LineWidth',2);
    figure
    % subplot 321
```

```
    yacute0 = sim(acute0,ut');
plot(t(gg:end),yacute0(gg:end),'r--','LineWidth',2);hold
on; %axis([0 10 -400 400])
    % subplot 322
    yacute3 = sim(acute3,ut');
plot(t(gg:end),yacute3(gg:end),'g.-
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 323
    yacute6 =
sim(acute6,ut');plot(t(gg:end),yacute6(gg:end),'c--
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 324
    yacute12 =
sim(acute12,ut');plot(t(gg:end),yacute12(gg:end),'y-
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 325
    yacute25 =
sim(acute25,ut');plot(t(gg:end),yacute25(gg:end),'b-
','LineWidth',2);%axis([0 10 -400 400])
    title('Acute dose response to u(t)');legend(' acute0','
acute3','acute6','acute12','acute25');
    grid on;

    % %%
    % % figure
    % % subplot 223
    % % ytol0 = sim(tol0,ut'); plot(t(k(2)-
4000:k(2)),ytol0(k(2)-4000:k(2)),'r--','LineWidth',2);hold
on;
    % % ytol3 = sim(tol3,ut'); plot(t(k(2)-
4000:k(2)),ytol3(k(2)-4000:k(2)),'g.-','LineWidth',2);
    % % ytol6 = sim(tol6,ut');plot(t(k(2)-
4000:k(2)),ytol6(k(2)-4000:k(2)),'c--','LineWidth',2);
    % % ytol12 = sim(tol12,ut');plot(t(k(2)-
4000:k(2)),ytol12(k(2)-4000:k(2)),'y-','LineWidth',2);
    % % ytol25 = sim(tol25,ut');plot(t(k(2)-
4000:k(2)),ytol25(k(2)-4000:k(2)),'b-','LineWidth',2);
    % % ytol0 = sim(tol0,ut'); plot(t,ytol0,'r--
','LineWidth',2);hold on;
    % % ytol3 = sim(tol3,ut'); plot(t,ytol3,'g.-
','LineWidth',2);
    % % ytol6 = sim(tol6,ut');plot(t,ytol6,'c--
','LineWidth',2);
    % % ytol12 = sim(tol12,ut');plot(t,ytol12,'y-
','LineWidth',2);
```

```
    % % ytol25 = sim(tol25,ut');plot(t,ytol25,'b-
','LineWidth',2);

    figure
    % subplot 321
    ytol0 = sim(tol0,ut'); plot(t(gg:end),ytol0(gg:end),'r-
-','LineWidth',2);hold on; %axis([0 10 -400 400])
    % subplot 322
    ytol3 = sim(tol3,ut');
plot(t(gg:end),ytol3(gg:end),'g.-','LineWidth',2);%axis([0
10 -400 400])
    % subplot 323
    ytol6 = sim(tol6,ut');plot(t(gg:end),ytol6(gg:end),'c--
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 324
    ytol12 =
sim(tol12,ut');plot(t(gg:end),ytol12(gg:end),'y-
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 325
    ytol25 =
sim(tol25,ut');plot(t(gg:end),ytol25(gg:end),'b-
','LineWidth',2);%axis([0 10 -400 400])
    title('Tolerance dose response to u(t)');legend('
Tolerance0','
Tolerance3','Tolerance6','Tolerance12','Tolerance25');
    grid on;
    % % figure
    % % subplot 224
    % % ynormal0 = sim(normal0,ut'); plot(t(k(2)-
4000:k(2)),ynormal0(k(2)-4000:k(2)),'r--
','LineWidth',2);hold on;
    % % ynormal3 = sim(normal3,ut'); plot(t(k(2)-
4000:k(2)),ynormal3(k(2)-4000:k(2)),'g.-','LineWidth',2);
    % % ynormal6 = sim(normal6,ut');plot(t(k(2)-
4000:k(2)),ynormal6(k(2)-4000:k(2)),'c--','LineWidth',2);
    % % ynormal12 = sim(normal12,ut');plot(t(k(2)-
4000:k(2)),ynormal12(k(2)-4000:k(2)),'y-','LineWidth',2);
    % % ynormal25 = sim(normal25,ut');plot(t(k(2)-
4000:k(2)),ynormal25(k(2)-4000:k(2)),'b-','LineWidth',2);
    % % ynormal0 = sim(normal0,ut'); plot(t,ynormal0,'r--
','LineWidth',2);hold on;
    % % ynormal3 = sim(normal3,ut'); plot(t,ynormal3,'g.-
','LineWidth',2);
    % % ynormal6 = sim(normal6,ut');plot(t,ynormal6,'c--
','LineWidth',2);
```

```
    % % ynormal12 = sim(normal12,ut');plot(t,ynormal12,'y-
','LineWidth',2);
    % % ynormal25 = sim(normal25,ut');plot(t,ynormal25,'b-
','LineWidth',2);
    figure
    % subplot 321
    ynormal0 = sim(normal0,ut');
plot(t(gg:end),ynormal0(gg:end),'r--','LineWidth',2);hold
on; %axis([0 10 -400 400])
    % subplot 322
    ynormal3 = sim(normal3,ut');
plot(t(gg:end),ynormal3(gg:end),'g.-
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 323
    ynormal6 =
sim(normal6,ut');plot(t(gg:end),ynormal6(gg:end),'c--
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 324
    ynormal12 =
sim(normal12,ut');plot(t(gg:end),ynormal12(gg:end),'y-
','LineWidth',2);%axis([0 10 -400 400])
    % subplot 325
    ynormal25 =
sim(normal25,ut');plot(t(gg:end),ynormal25(gg:end),'b-
','LineWidth',2);%axis([0 10 -400 400])
    title('Normal dose response to u(t)');legend('
Normal0',' Normal3','Normal6','Normal12','Normal25');
    grid on;
```

Human data modeling, MATLAB codes

```
    data=[data(:,1),log(data(:,2)+1),log10(data(:,3)+1),(lo
g10(1+data(:,4)))];

    %%
    categories = {'FEV1','Methacholine','Total
IgE','Eosinophils'};
    ave     = mean(data);
    data = data- repmat(ave,length(data),1);
    stdr    = std(data);
```

```
    sr      = data./repmat(stdr,length(data),1);
    [coefs,scores,variances,tt] = princomp(sr);
    figure(4)
    scatter(scores(:,1),scores(:,2),'r');
    xlabel('First Principal Component');
    ylabel('Second Principal Component');
    title('Principal Component Scatter
Plot');%gname(comdiagnosis);
    % gname();
    figure(5)
    vlabs = categories;
    biplot(coefs(:,1:3),'scores',scores(:,1:3),'varlabels',
vlabs);
    figure(6)
    percent_explained = 100*variances/sum(variances);
    pareto(percent_explained);
    xlabel('Principal Component')
    ylabel('Variance Explained (%)');grid;
    %
    [pc, zscores, pcvars] = princomp(data);
    pcvars./sum(pcvars) .* 100;
    cumsum(pcvars./sum(pcvars) * 100)
    figure(7)
    scatter(zscores(:,1),zscores(:,2));
    xlabel('First Principal Component');
    ylabel('Second Principal Component');
    title('Principal Component Scatter Plot');%gname();

    corrDist = pdist(data, 'corr');
    clusterTree = linkage(corrDist, 'average');
    clusters = cluster(clusterTree, 'maxclust', 4);

    figure
    for c = 1:4
        subplot(2,2,c);
        plot(data((clusters == c),:)');
        axis tight
    end
    %%
    rand('state',0);
    [cidx, ctrs] = kmeans(data, 6, 'dist','corr',
'rep',2,'disp','final');
    figure
    for c = 1:6
        subplot(2,3,c);
```

```
        plot(data((cidx == c),:)');
        axis tight
    end
    suptitle('K-Means Clustering of Profiles');
    suptitle('Hierarchical Clustering of Profiles');
    figure
    for c = 1:6
        subplot(2,3,c);
        plot(ctrs(c,:)');
        axis tight
        axis off    % turn off the axis
    end
    suptitle('K-Means Clustering of Profiles');
    %%
    % clustergram(data)
    % figure;
    % Y = pdist(data(:,1:end),'seuclidean');
    % Z = linkage(Y,'average');
    % [H,T] = dendrogram(Z,'colorthreshold','default');
    % set(H,'LineWidth',2)


    %%
    P = zscores(:,1:2)';
    net = newsom(P,[2 2]);
    net = train(net,P);
    figure
    plot(P(1,:),P(2,:),'.g','markersize',20);
    hold on
    plotsom(net.iw{1,1},net.layers{1}.distances);
    hold off
    distances = dist(P',net.IW{1}');
    [d,cndx] = min(distances,[],2);
    % cndx gives the cluster index
    figure
    gscatter(P(1,:),P(2,:),cndx); legend off;
    hold on
    plotsom(net.iw{1,1},net.layers{1}.distances);gname(ori(
:,3))
    hold off

    %%
    orindx = [ori,cndx];
    sum1 = []; sum2 = []; sum3 = []; sum4 = []; sum5 = [];
sum6 = [];
```

```
    num1 =0;num2 =0;num3 =0;num4 =0;num5 =0;num6 =0;
    ind1=[];ind2=[];ind3=[];ind4=[];ind5=[];ind6=[];
    for i = 1 : length(ori)
        if orindx(i,5) ==1
            sum1 = [sum1; ori(i,:)];
            num1 = num1+1;
            ind1 = [ind1;i];
        elseif orindx(i,5) ==2
            sum2 = [sum2 ; ori(i,:)];
            num2 = num2+1;
            ind2 = [ind2;i];
        elseif orindx(i,5) ==3
            sum3 = [sum3 ; ori(i,:)];
            num3 = num3+1;
            ind3 = [ind3;i];
        elseif orindx(i,5) ==4
            sum4 = [sum4 ; ori(i,:)];
            num4 = num4+1;
            ind4 = [ind4;i];
        elseif orindx(i,5) ==5
            sum5 = [sum5 ; ori(i,:)];
            num5 = num5+1;
            ind5 = [ind5;i];
        elseif orindx(i,5) ==6
            sum6 = [sum6 ; ori(i,:)];
            num6 = num6+1;
            ind6 = [ind6;i];
        end
    end
    clus1 = (sum(sum1, 1))/num1;
    clus2 = (sum(sum2, 1))/num2;
    clus3 = (sum(sum3, 1))/num3;
    clus4 = (sum(sum4, 1))/num4;
    clus5 = (sum(sum5, 1))/num5;
    clus6 = (sum(sum6, 1))/num6;
    [clus1;clus2;clus3;clus4;clus5;clus6]
    %%

    fevgroups = [fev11;fev12;fev13;fev14;fev15];
    pcgroups  =  [pc201;pc202;pc203;pc204;pc205];
    igegroups = [ige1;ige2;ige3;ige4;ige5];
    eosgroups = [eos1;eos2;eos3;eos4;eos5];
    fevnames = {names1{:,1},names2{:,1},
names3{:,1},names4{:,1},names5{:,1}};
```

```
    pcnames  = {names1{:,2},names2{:,2}, names3{:,2},
names4{:,2}, names5{:,2}};
    igenames = {names1{:,3},names2{:,3}, names3{:,3},
names4{:,3}, names5{:,3}};
    eosnames = {names1{:,4},names2{:,4}, names3{:,4},
names4{:,4}, names5{:,4}};

    [p1,anovatab1,stats1] = anova1(fevgroups',fevnames);

    [p2,anovatab2,stats2] = anova1(pcgroups',pcnames);
    [p3,anovatab3,stats3] = anova1(igegroups',igenames);
    [p4,anovatab4,stats4] = anova1(eosgroups',eosnames);
    %[p5,anovatab5,stats5] = anova1(eosgroups',eosnames);
```