

Analysis of Cough Time Stamps from COPD Patients Using Markov Chain Analysis

Tsega Kahsay Gebretekle^{1,*}, Stef van Eijndhoven², Bert den Brinker³

¹Department of Mathematics, Kotebe Metropolitan University, Addis Ababa, Ethiopia

²Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands

³Philips Electronics Nederland BV, Philips Research, Eindhoven, The Netherlands

Email address:

tsegaka@yahoo.com (T. K. Gebretekle), S.J.L.v.Eijndhoven@tue.nl (S. van Eijndhoven), Bert.den.Brinker@philips.com (B. den Brinker)

*Corresponding author

To cite this article:

Tsega Kahsay Gebretekle, Stef van Eijndhoven, Bert den Brinker. Analysis of Cough Time Stamps from COPD Patients Using Markov Chain Analysis. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 3, 2019, pp. 108-124. doi: 10.11648/j.ajtas.20190803.13

Received: May 22, 2019; **Accepted:** July 9, 2019; **Published:** August 5, 2019

Abstract: The objective of this study was to identify patterns of cough events for COPD patients. Simultaneously, the study was used to develop a Matlab based graphical user interface (GUI) that enables the user to analyze time-stamps of cough data. The time stamp data was received from Philips Research. They cover 17 data sets of 16 COPD patients and were determined using a semi-automated cough detection algorithm. Cough detection ran for multiple days in the living and bed rooms of the patients. The time stamp marks the event that a cough is assumed to occur. A descriptive statistics and a Markov Chain Model was used for analysis. A pattern of cough events was described by the probability that a COPD patient is in one of three possible states at a specific hour and in another state at the next hour. To define the states, the following three characteristics were used: 1) relative frequency, 2) average value-three times standard deviation band, 3) average value-three times inter-quartile range band. Relaxation time was determined to describe the dynamics of the cough event patterns. To be precise, pattern changes were characterized by considering the time it takes for the probabilities to reach stationarity. To reduce noise, the daily dynamics of the cough events over five day periods with a four day overlap were considered. From the results, we concluded that the distribution of cough events for all data sets was skewed to the right. The developed Matlab based graphical user interface allows the user to analyze the cough events of COPD patients together with their medical history. We conclude that the relaxation time and the stationary distribution of the Markov chain representation were typical characteristics of the patterns of cough events and the cough behavior of COPD patients was patient specific and varies over time.

Keywords: COPD, Cough, Markov Chain Analysis, Relaxation Time

1. Introduction

World Health Organization [1] defines Chronic Obstructive Pulmonary Disease (COPD) as a lung disease. It is characterized by chronic obstruction of lung airflow that causes a shortage of breath. From the total deaths in the Netherlands each year, about 4%, which is the equivalent of 6,000 people, is caused by COPD. It is a major cause of mortality worldwide. It is predicted to become the third top cause of death in 2030 by the World Health Organization [2]. Better means are clearly needed for the prevention and treatment of COPD, and more scientific research is needed to enable improvements in its clinical management. This

disease is a fully irreversible disease that gets worse over time but can be treated. The treatment is aimed at reducing the symptoms and preventing worsening. COPD has different symptoms such as cough, breathlessness, fatigue, production of sputum, and wheezing. Since the symptoms of COPD develop slowly, patients may not know that they suffer from COPD.

It is important to predict the worsening of the disease by only considering one of the symptoms called cough. Different researchers only analyze cough of COPD patients using descriptive statistics. However, a detailed modeling and analysis of COPD using coughs signal is needed.

There are four stages of COPD. The first stage, Mild COPD, occurs when the Pulmonary Function Test (PFT) with

a Forced Expiratory Volume in 1 second (FEV1) is greater than 80% of the predicted normal values. The second stage of COPD, Moderate COPD, is when the FEV1 of the patient is between 50% and 79%. The third stage of COPD is called Severe COPD. This happens when the FEV1 of the patient is between 30% and 49%. The fourth stage is called Very Severe COPD, which is when the FEV1 of the patient is less than 30% [3]. The progress of the disease means going from the first stage to the fourth stage. Those underlying stages are interrupted by exacerbations of symptoms, which vary in severity and frequency both between patients and during the course of the patient's illness.

As Chung [4], cough is a characteristic sound that can be distinguished by the human ear and that is unique from other sounds. He explained that measurement of cough may indicate tools that measure the number of coughs or may indicate the tools that measure the severity of coughs. He also explained that, since most cough observers use a microphone to record the cough sound and most software programs are currently unable to distinguish coughs from other closely related sounds, it is useful to find those features of coughs that could be used to differentiate coughs from other sounds.

When a patient coughs repeatedly within a few seconds, it is considered as a spontaneous cough. Most investigators count each discrete event of a cough because spontaneous coughs usually occur in a series of cough sounds within a few seconds. Therefore, it is reasonable to count all the coughs and all the cough epochs. The intensity and duration of a cough can be examined by recording the coughs placing the microphone as near as possible to the patient.

Stable COPD patients cough much more during day time periods than night time periods. There is also a significant but moderate correlation between times spent coughing during the day time and the night time: patients coughing more during the day time tended to cough more at night [5].

1.1. Statement of the Problem

There is little information on the frequency and severity of cough in COPD or its impact on quality of life at different stages of the disease. Cough is usually the first and most common symptom of COPD patients [6]. It has the greatest predictive value for subsequent hospital admission due to COPD as compared to other symptoms. Cough also had the highest predictive value for treatment of airflow obstruction [7]. According to Rennard et al. [8], who asked patients to report frequency of COPD symptoms, Cough was the most commonly reported symptom. That means, 70% of participants reported that cough was their symptom. Patients with symptoms such as cough are classified as a Stage 0 disease patients suggesting that these patients are at risk of developing COPD and may be a useful factor in case of finding patients at risk of progressive airflow obstruction in the community [9-11].

Possible indicators for an exacerbation in COPD patients are chronic cough, sputum production, a change in the color or thickness of mucus, and breathlessness. There is evidence that COPD patients with chronic cough production have, on

average, more exacerbations per year than COPD patients without this symptom [12]. Hurst *et al.* [13] also reported that chronic cough at study entry was associated with the occurrence of exacerbations. The research in this field is a very important topic for innovation-driven medical companies.

One of the activities that Philips Research considers is the development of a cough analyzer. To make such analyzer relevant for diagnosis, it can be investigated whether the number of coughs can be used as an indicator for a COPD exacerbation.

Studying these exacerbations is important not only because of their impact on the individual's life but also because of their long term effects on health status. In fact, exacerbation frequency is one of the most important determinants of health related quality of life. Exacerbations are a significant cause of hospital admission and re-admission; the burden placed on health resources is large.

It is important to determine symptoms that can predict or identify the dynamics of the disease. One of these symptoms is cough. A particular interest is to look into the pattern of cough events of COPD patients. The medical specialist can relate the characteristic of the patient with the diagnosis.

1.2. Objective of the Study

The purpose of this research is to build statistical models that have the potential to describe the behavior of cough events for COPD patients and the exacerbation of the disease. One particular objective is to find adequate descriptions of cough occurrence patterns. The deliverable of the project is a Matlab based graphical user interface (GUI) that renders the analysis results in easily interpretable ways.

2. Materials and Methods

2.1. Description of the Study Area

Philips Research is one of the three main divisions of Philips International and was founded in Eindhoven in 1914. Philips is a global organization, which has laboratories in the Netherlands, England, Germany, China, and the United States. It gives technology options for innovations in the area of health and well-being. It produces innovative solutions that meet the needs of both patients and health professionals combining expertise in medical technology with clinical know-how of its customers with the aim to apply its innovation to the needs, desires, and goals of people [14]. The company pays great attention to the development of new medical equipment.

2.2. Data

Cough time stamp data of 16 patients were received from Philips Research, Eindhoven, The Netherlands. The trial for these data sets was conducted in England at Hull University. One of the patients was involved in two sessions and that was why there were 17 data sets of 16 COPD patients. There were 34 files of data, which were recorded in the living room and the bedroom of the patient's home. The data from the living room contained 16, which were marked "A" recordings

whereas the data from the bedroom contained 18 files marked “B” recordings (See Table 1). The cough counts were started at the moment the systems were operational. The sound in these rooms was recorded continuously for 45 days. The patients were examined on first, fifth, 20th, and 45th of the trial. The start of the experiment for each patient was different. Patients were recorded in periods with duration of 33 days up to 46 days. The researcher who conducted the trial knew when patients went back to the hospital. For two data sets, there was no enough information from the living room. Therefore, only the time stamps from the bedroom were considered. For one of the data sets, It is observed that there was insufficient number of cough events to determine their pattern. One of the patients was not a COPD patient and the corresponding data set was not considered in this study. The trials of all patients considered in this study were between

January 2012 and May 2013. Patients at more or less the same stage of COPD were involved. Those were severe COPD stage patients.

Some of the patients (P7, P11, and P12) were absent during some days (from four up to 12 days). During the days without cough, for patients P2, P5, P6, and P7, It is not known if the patient was not at home or did not cough. Patient P9 has two periods of recording (data sets P9_1 and P9_2). Each of those periods was considered separately in the descriptive statistics analysis (See in Table 1).

Six patients (P1, P6, P8, P12, P16, and P17) had an exacerbation but only three of them were considered for the analysis of exacerbation in relation to number of cough events in this paper. The period of recording is patient specific and presented in number of days.

Table 1. Summary of COPD patients.

Data set	Patient	Period of recording	Place of recording	Hospital visits	Dates without cough/absence	Number of days
“0234”	P1	07.06.12 – 31.07.12	Only bedroom	Exacerbation: 02.07.12 – 25.07.12	No	33 days
“0575”	P2	08.08.12 – 23.09.12	Both in living room and bedroom	No	Without cough: 20.08.12, 09.09.12, 20.09.12, 22.09.12	43 days
“1422”	P3	01.05.12 – 15.06.12	Both in living room and bedroom	No	No	46 days
“2057”	P4	12.04.12 – 27.05.12	Both in living room and bedroom	No	No	46 days
“2551”	P5	31.05.12 – 15.07.12	Both in living room and bedroom	No	Without cough: 3.06.12, 26.06.12, 05.07.12, 09.07.12	42 days
“4677”	P6	28.03.12 – 08.05.12 (A) 28.03.12 – 04.04.12 (B)	Both in living room and bedroom	06.04.12 – 12.04.12, 18.04.12 – 06.05.12, 09.05.12 – 11.05.12, Exacerbation: 16.05.12 – 17.05.12	Without cough: A: 29.03.12, 01.04.12, 02.04.12, 04.04.12 Absence: 24.05.12 – 26.05.12, 31.05.12 – 05.06.12, Without cough: 12.05, 15.05, 20.05, 23.05, 27.05 (A), 10.05, 15.05, 18.05 (B)	14 days (A) and 8 days (B)
“5352”	P7	02.05.12 – 15.06.12	Both in living room and bedroom	No	Without cough: 12.05, 15.05, 20.05, 23.05, 27.05 (A), 10.05, 15.05, 18.05 (B)	34 days (A) and 38 (B)
“5580”	P8	23.03.12 – 07.05.12	Joint living- and bedroom	Exacerbation: 25.03.12 – 28.03.12	No	46 days
“6867_1” “6867_2”	P9	24.01.12 – 08.02.12 12.03.12 – 25.04.12	Both in living room and bedroom	No	No	61 days
“8891”	P10	20.03.12 - 4.05.12	Both in living room and bedroom	No	No	46 days
“9693”	P11	29.05.12 – 19.06.12 (A) 29.05.12 – 13.07.12 (B)	Both in living room and bedroom	No	13.06.12, 14.06.12, 16.06.12, 18.06.12, 20.06.12, 21.06.12, 23.06.12 – 25.06.12, 27.06.12, 30.06.12, 02.07.12, 04.07.12, 06.07.12, 07.07.12, 11.07.12, 12.07.12 (B)	22 days (A) and 29 days (B).
“9816”	P12	27.04.12 – 13.05.12 (A) 27.04.12 – 08.06.12 (B)	Both in living room and bedroom	Exacerbation: 30.04.12 – 03.05.12	19.05.12, 21.05.12 – 24.05.12, 29.05.12, 04.07.12, 06.07.12 (B)	17 days (A) 35 days (B)
“0442”	P13	01.11.12 – 16.12.12	Both in living room and bedroom	No	No	46 days
“8606” (not COPD Patient)	P14	31.03.12 – 15.05.12	Both in living room and bedroom	No	No	46 days
“0333”	P15	04.03.13- 22.04.13	Both in living room and bedroom	No	No	50 days
“7762”	P16	12.03.13-30.04.13	Both in living room and bedroom	Exacerbation: 18.03.13	No	50 days
“9436”	P17	20.03.13-04.05.13	Both in living room and bedroom	Exacerbation: 21.04.13	No	46 days

2.3. Descriptive Statistics

Since the received data for each patient were in two files (time stamps of occurrences of coughs in the living room and bedroom), those were combined into one set. Then, descriptive statistics was carried out for each data set by using tables and graphs. A complication was that different data sets were recorded in different periods of time and that some had many hours with zero values. This case was taken into consideration in the methodology part.

2.4. One - Way Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA) was applied to identify if the day can be split into time intervals that have significantly different average relative number of coughs. ANOVA is used to compare the averages of two or more groups of data. Generally speaking, there are two hypotheses to be tested.

H_0 : The mean of the relative number of coughs in an hour at different grouped hours are equal, i.e. $\mu_i = \mu_j$ for $i \neq j$ ($\mu_1 = \mu_2 = \dots = \mu_h$) versus H_1 : The mean relative number of coughs at different hours is different; for at least one hour, the means are different, i.e. $\mu_i \neq \mu_j$ for at least one $i \neq j$.

The detail explanation about ANOVA is given by Penny and Henson [15].

The hypothesis is tested based on the p-value or based on the computed F-test statistic ($F = MSG/MSE =$ Mean square of group/ Mean square of error). That means, the F-test statistic is compared with the F-tabulated value or the p-value is compared with the selected level of significance to reject or not reject the null hypothesis.

To use this method, the discrete or count data points have to be transformed to continuous data points called the relative number of coughs in an hour, which is defined by:

$$r = \frac{\text{number of cough events at hour } h}{\text{total number of cough events during the day}} \quad (1)$$

The relative numbers of coughs in an hour of the whole experimental period were grouped into 24 hour groups, one for each hour in the day. The average difference of the relative number of coughs in an hour between day time and night time of a day were considered. The relative numbers of coughs in an hour were also grouped into three. The first group consists of the first 8 hours of the day (hours 0:00-8:00), the second group of the next eight hours (hours 8:00-16:00), and the third group of the last eight hours (hours 16:00-24:00).

An ANOVA test was made based on the p-value to see if there is a significant difference among the means of the hours of the relative number of coughs in an hour. If the p-value is smaller than 5% or 1%, the null hypothesis that states all samples of the groups are drawn from the population with the same mean is rejected.

2.5. Markov Chain Analysis

Markov Chain Analysis was applied to identify a cough pattern and its dynamics. For this, states were defined on the following characteristics of the cough signals: (1) Relative

frequency of number of coughs in an hour relative to the other hours of the day, (2) The average values of number of coughs in an hour of a given hour of a day over the whole experimental periods as lower bounds and three times of the standard deviation values as upper bounds, and (3) The average values of number of coughs in an hour of a given hour of a day over the whole experimental periods as lower bounds and three times of the inter-quartile range values as upper bound.

The whole experimental period was divided into sub periods with overlap. As the characteristic of the dynamics, the relaxation time was introduced, i.e., the average time that a patient is in a stable condition after disturbance.

Assume that there is a process that can be measured in discrete time. This process consists of occurrences of a simple event. In addition, there is a time period T during which the time when an event occurred was marked. The timestamps at which the event occurred were considered as data that are related to this process. Let us take one unit of time during which there are n possibilities of event occurrences. For example, if an hour is considered as one unit and event occurrences every second are counted, $n = 3600$ can be set. For each unit, in correspondence a number that determines the sum of event occurrences can be placed. For example,

$$m^{th} \text{ unit} \Rightarrow \sum_{i=1}^n x_i = k_m, \quad m = 1, 2, \dots, T \quad (2)$$

where

$$X_i = \begin{cases} 1, & \text{if event happens} \\ 0, & \text{if event does not happen} \end{cases}, i = 1, 2, \dots, n \quad (3)$$

In this, the event occurred $\sum_{m=1}^T k_m$ times during period T. d units can be merged together into one period and those larger periods can be considered. For instance, if one unit is one hour, then, every 24 hours ($d = 24 = 1$ day) are merged and event occurrences in a day are considered. There is a set of periods $\Omega = \{y_1, y_2, \dots, y_N\}$, where $N = T/d$. For each of those periods, in correspondence a number is placed that means how many times the event happened during this specific period:

$$y_m = \sum_{i=(m-1)d+1}^{md} k_i, \quad m = 1, 2, \dots, N \quad (4)$$

To do analysis of the process, a characteristic that aggregates events in a time period have to be chosen and that will help us to compare periods.

Let us denote X_j^l , $j = 1, 2, \dots, d$; $l = 1, 2, \dots, N$ as a characteristic of j^{th} unit in l^{th} period. For example, a relative frequency of the event can be determined as this characteristic:

$$X_j^l = \frac{k_j^l}{\sum_{i=(l-1)d+1}^{ld} k_i} \quad (5)$$

where k_j^1 is the number of event occurrences during j^{th} unit in 1^{st} period.

According to this characteristic, for instance, each j^{th} unit for every period from Ω can be compared. Of course, there is full freedom to interpret a time unit, values of T, n, d, and X. In any case, as a result, a process discretized in time with time periods that are comparable according to one rule is obtained. In general, the values X_{j1} have a huge variance, which makes the analysis complex. That is why it is better to make the process discrete not only in time but also in states. This means, the process is quantized. To do this, the characteristic is used again. This means that the process in the j^{th} time unit of the period l is in State i if $\alpha_i \leq \chi_j^l \leq \beta_i$, where the parameters α_i and β_i can be chosen with respect to the model properties or some other conditions. The process can have different number of states but this number has to be finite.

Now the considered process is discretized in time and in state. For further analysis, the method of discrete Markov Chains analysis was used, which is described below. The purpose of applying this method is to determine event patterns. With Markov Chains the process starts in one of the states and moves successively from one state to another according to some transition probabilities.

To make the data sets discrete in states, three possible ways were considered. In the first case, the relative frequency of number of coughs in an hour was calculated as a characteristic of the process as it is given in (1). In this case, the relative frequency defines the probability of cough in every hour for each day. In the second and third cases, threshold bands were considered. That means, the average of the relative number of coughs in a given hour was calculated and that average was considered as a lower bound. Three times the standard deviation and three times the inter-quartile range were calculated and they were considered as upper bounds. There are three events: on band, below band, and above band. State 0 means that the relative number of coughs is below band, State 1 means that it is on band, and State 2 that it is above band. Three ways were exploited to define the Markov states:

- 1) Relative frequency,
- 2) Average and θ time the standard deviation band,
- 3) Average and θ times the Inter-quartile range band

Three states according to the relative frequency r were indicated, where r is a relative frequency at any hour of the experiment.

$$X(t) = \begin{cases} 0, & r(t) \leq \alpha, \\ 1, & \alpha < r(t) \leq \beta, \\ 2, & r(t) > \beta \end{cases} \quad (6)$$

Three states according to a relative frequency r_h were also indicated in the band ($average_h < r_h \leq \theta.stdv$ or $average_h < r_h \leq \theta.stdv$),

$$X(t) = \begin{cases} 0, & r_h(t) \leq average_h, \\ 1, & average_h < r_h(t) \leq \theta.stdv, \\ 2, & r_h(t) > \theta.stdv \end{cases}$$

or

$$X(t) = \begin{cases} 0, & r_h(t) \leq average_h, \\ 1, & average_h < r_h(t) \leq \theta.IQR, \\ 2, & r_h(t) > \theta.IQR \end{cases} \quad (7)$$

where r_h is a relative frequency in a given hour of a day.

The parameters α and β might be chosen according to the properties of the real process or by the doctor's suggestions. The parameter θ is also chosen by the user or the number of data points above three times standard deviation or three times inter-quartile range.

As a result of the performed analysis, an estimate of the transition probability matrix is obtained that describes the probabilities of moving from one state to another. The Eigen row of eigenvalue 1 is the stationary distribution of probabilities of the Markov process, i.e., the probability that the process is in one of the three states. If the process is regarded as a dynamic system, one more characteristic of this process can be defined, typically, the relaxation time is considered (the time during which the system moves without any external factor), as a characteristic.

The dynamics of the event occurrences during time T can be also considered. Let us divide the whole time period T into smaller time periods D^j , $j = 1, 2, \dots, N$, with length of d.h time units and overlap by $d*(h-s)$ time units, where $N=[T:h]*h$, for any integers h and s, $s < h$. Although there is a free choice to define the value of h and s, they have to be optimal according to some presumed periodicity of the process. Every D^j is discretized in time and in state and hence, a Markov Chain method can be used. For each smaller period, a stationary probability distribution and a relaxation time were calculated. Then, the dynamics of those two characteristics were considered as a descriptor for the dynamics of the patterns of the process.

Given a discrete time stochastic process X_n , $n = 0, 1, 2, \dots$, to know the probability of this discrete stochastic process at different time periods ($P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)$), a Markov chain analysis can be assumed. A Markov chain is a stochastic process in which the distribution of future states depends only on the present state and not on how it arrived in the present state. Thus, the process is fully described by the transition probability and the initial state.

$$P(X_n = i_n | X_{n-1} = i_{n-1}) \quad (8)$$

In a Markov chain, a transition probability is a probability of a given state going to another state. A one-step transition probability, $p_{ij}^{n,n+1}$, is a probability of state i at time n going to State j at time n+1. If the one-step transition probability is time independent, then $P = p_{ij}$.

The one-step transition probability $P = p_{ij}$, has to satisfy the following properties.

$$0 \leq p_{ij} \leq 1, i, j = 1, \dots, n,$$

$$\sum_{j=1}^n p_{ij} = 1, \text{ for all } i = 1, \dots, n$$

P has an eigenvalue 1 with right eigenvector $e = (1 \dots 1)^T$: $Pe = e$

There is long run probability distribution, row vector q , in which qP is also a probability distribution. It is given as:

$$(qP)_i = \sum_{j=1}^n q_j p_{ij} \tag{9}$$

It satisfies the following property as:

$$0 \leq (qP)_i \leq 1$$

$$\sum_{i=1}^n (qP)_i = \sum_{i=1}^n q_i = 1$$

The k -step transition probability, $P^k = p_{ij}^k$, is the probability of going to State j from State i in exactly k steps. It is also the k^{th} power of the matrix P , which is given as:

$$p_{ij}^k = \sum_{l=1}^n p_{il}^{k-1} p_{lj} \tag{10}$$

By calculating the k -step transition probability, the Chapman-Kolmogorov equations are obtained.

There is a long run proportion of time that the process spends in State j , $q^{[∞]} \neq 0$, where $q^{[∞]}P = q^{[∞]}$. It is called stationary probability, which is given as follows:

$$q^{[0]}P^n = q^{[∞]} + c_2 \lambda_2^n q_2 + c_3 \lambda_3^n q_3 \tag{11}$$

where c_2 and c_3 are constants, λ_2 and λ_3 are eigenvalues of the transition probability matrix, and q_2 and q_3 are eigenvectors. If $n = \infty$, the two right side expressions, $c_2 \lambda_2^n q_2$ and $c_3 \lambda_3^n q_3$, in Equation (11) becomes zero, then the whole expression will be only the long run probability distribution.

The relaxation time is calculated by taking the maximum of the eigenvalues. It is given as:

$$\tau = -\frac{\Delta t}{\text{Log}(|\lambda_2|)}, \lambda_2 > \lambda_3 \tag{12}$$

where Δt is the sampling time of the process

3. Results

The time stamps of cough events were aggregated into hourly bases and it was observed that there were hours for which cough events were not available. All observation times were made to be equal. This means, all coughs in an hour counts for those days where all 24 hours were measured by excluding the first and last (partly-measured) days were taken. The percentage of hours for which cough data was available was calculated. It cannot be explained the hours with zero data in a data set, because there is no information about whether the patient was not at home or did not cough during those hours. For patient P13 and patient P9, the percentage was above 80%. For detailed analysis, P13 is considered, the data set with the highest percentage (88%) of hours for which cough data was available (see in Figure 1). In this study, the discussion is about the number of coughs not about the severity of the disease. Most literatures discussed only the descriptive statistics of measurement of coughs but in this study, the numbers of coughs were modeled and the patterns were identified.

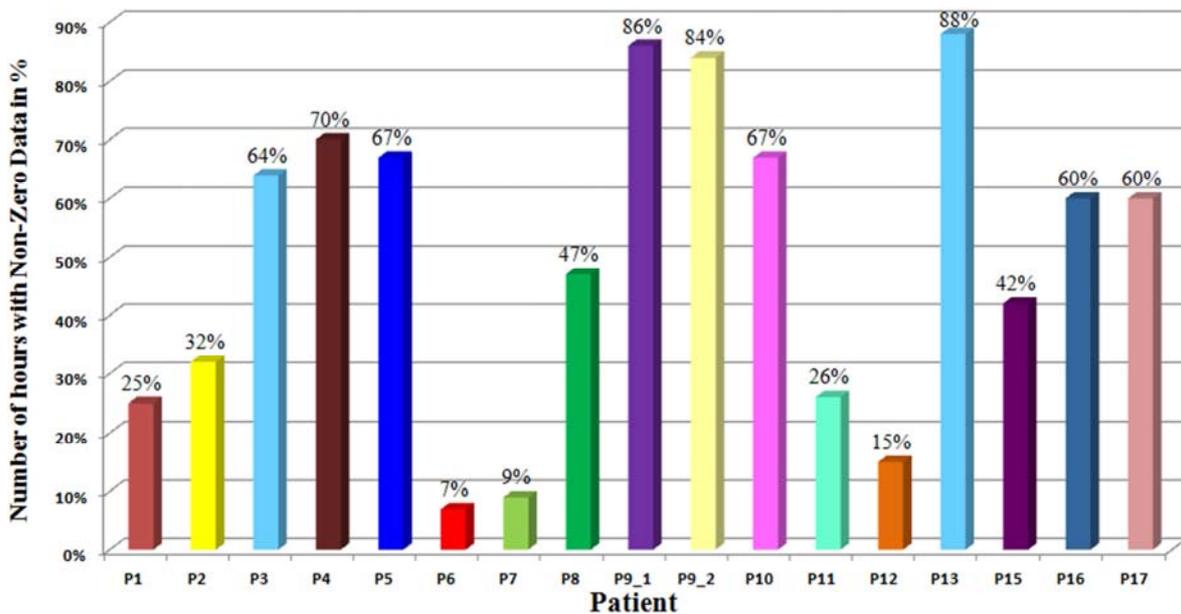


Figure 1. Bar graph of the percentage of hours for which Cough data is available.

3.1. Descriptive Statistics Results

Three highest and three lowest number of coughs on a day and also their respective experiment days were calculated (See Figure 2 and Figure 3). The three highest values are referred to be as first, second and third maximum and, similarly, the three lowest numbers are called the first, second, and third lowest number of coughs on a day. In Figure 3, there are large variations among patients. There are also large differences among the three highest numbers of coughs on a day for 5 patients (for patients P1, P4, P5, P8, and P13). For the other patients, these differences are smaller. Typically P13 distinguishes from the others with respect to the highest number of coughs on a day.

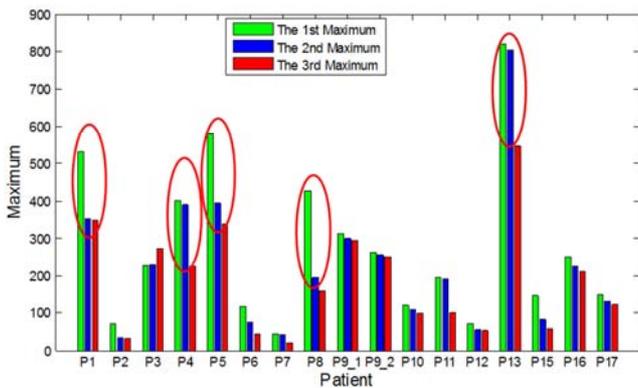


Figure 2. Bar graph of three highest numbers of coughs on a day for all patients.

We can see from Figure 3 that for 12 data sets, the highest number of coughs on a day occurred after the 10th day of the experiment (Data sets P1, P2, P3, P5, P7, P8, P9_1, P9_2, P10, P13, P15, and P17).

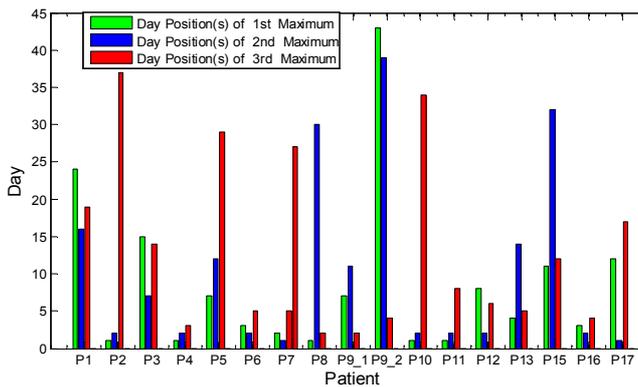


Figure 3. Bar graph of the day positions of the three maxima of number of coughs of each patient.

From Figure 4, the three lowest number of coughs on a day for patient P9 and patient P13 are higher than that of for the other patients.

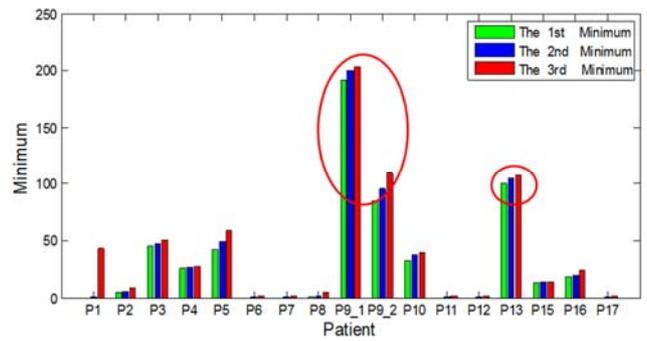


Figure 4. Bar graph of three lowest numbers of coughs on a day for all patients.

From Figure 5, for almost all data sets, the three lowest number of coughs occurred between the 27th and the 35th days of experiments. Most of the patients have a relatively small number of coughs after the 25th day of the experiment but note that some patients have less than 25 day of experiment, for example, data set P9_1.

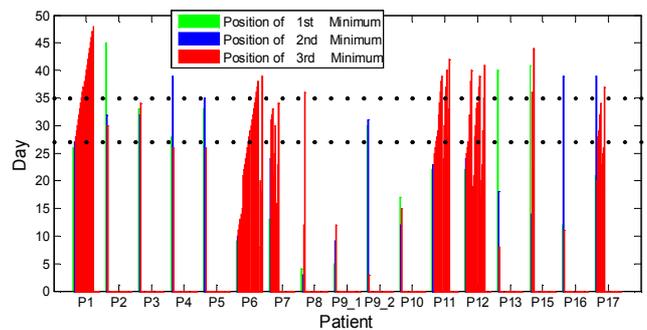


Figure 5. Bar graph of the day positions of the three minima of number of coughs of each patient.

The day time coughing is higher than the night time coughing for patient 13 but this is not true for all patients (See Figure 6). We can also see from Figure 7 that for almost all data sets, the day time coughing is higher than the night time coughing.

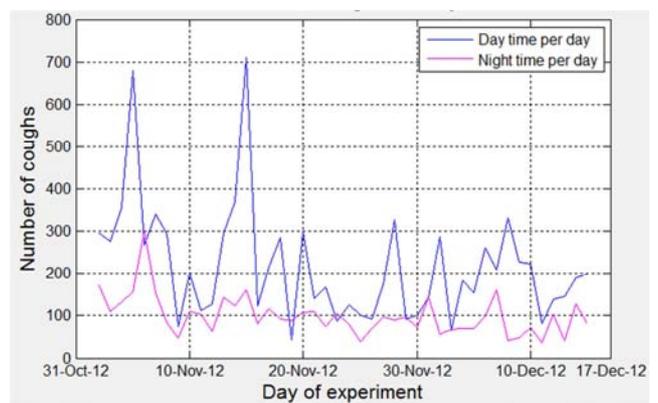


Figure 6. Number of coughs on a day time and Night time for P13.

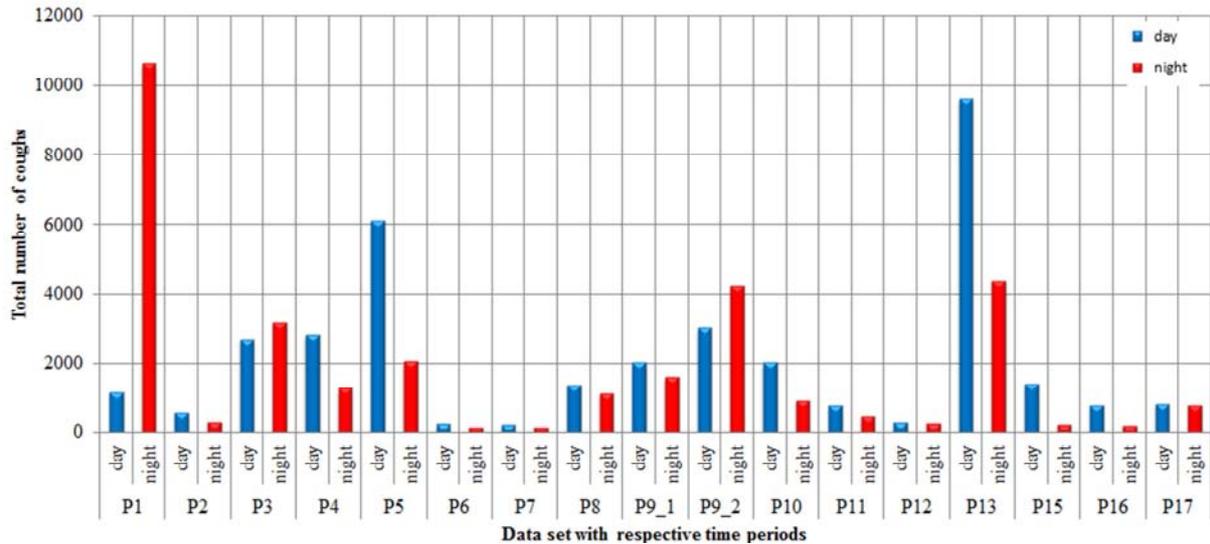


Figure 7. Bar graph of the total number of coughs during day time and night time for all patients.

Each day was divided into intervals with a length of three hours. According to the results illustrated in Table 2, approximately 53% (9 out of 17) of the data sets have a highest number of coughs between the hours 9:00-11:00 and

between the hours 21:00-23:00. This indicated that the highest number of coughs occurs for most of the patients before they go to sleep and after they woke up from their sleep.

Table 2. Time intervals of each day with maximum and minimum numbers of coughs in an hour, which occurs more often.

Data set	Time interval with maximum number of coughs	Data in the interval (%)	Time interval with minimum number of coughs	Data in the interval (%)
P1	21:00-23:00	25	12:00-14:00	16
P2	9:00-11:00	30	0:00-5:00	17
P3	6:00-8:00	33	12:00-17:00	16
P4	12:00-14:00	21	3:00-5:00	26
P5	9:00-11:00	25	3:00-5:00	28
P6	0:00-2:00	15	12:00-14:00	13
P7	15:00-17:00	22	0:00-5:00	14
P8	6:00-8:00	19	0:00-2:00	18
P9_1	21:00-23:00	40	0:00-2:00	33
P9_2	21:00-23:00	46	3:00-5:00	27
P10	9:00-11:00	24	3:00-5:00	33
P11	0:00-2:00	17	3:00-5:00	15
P12	15:00-17:00	21	0:00-5:00	14
P13	18:00-20:00	30	3:00-5:00	39
P15	9:00-11:00	33	6:00-8:00	21
P16	9:00-11:00	24	3:00-5:00	20
P17	21:00-23:00	33	3:00-5:00	22

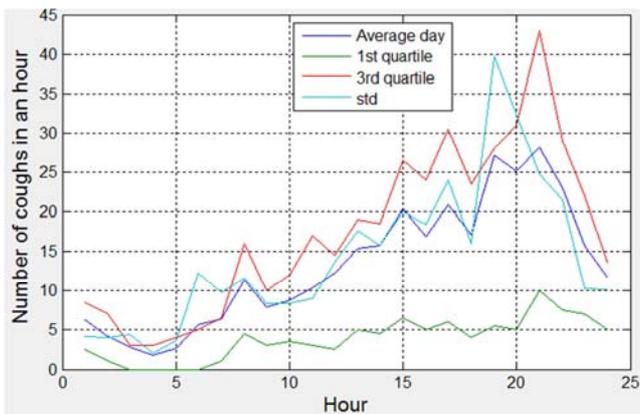


Figure 8. Average number of coughs in an hour during a day with its respective standard deviation and 1st and 3rd quartiles over 44 days of experiment for data set P13.

The average values of the number of coughs in an hour were also calculated for the whole experimental period. This day is called the average number of cough day. The blue curve represents the average values of the number of coughs in an hour over 44 days. The red curve represents the third quartile values of number of coughs in an hour and the green curve the first quartiles. The maximum number of coughs in an hour on the average number of cough day is about 28 for patient P13 at hour 20:00. The minimum average number is about two at hour 3:00 of the average day (See Figure 8).

3.2. Results of Data Analysis for Patients with Exacerbation

The data of three patients with exacerbation were analyzed. These were P1, P8, and P12. It is a first attempt to see if there is a cause-effect relation between the pattern of cough events

and the startup of an exacerbation and end of exacerbation. The data set does not reveal when exactly the exacerbation startup took place and also when it ends. Of course this complicates the analysis.

Data Set P1

The data set P1 is represented in Figure 9 - Figure 12. The data was recorded only in the bedroom. The maximum number of coughs in an hour for this patient was observed on July 01, 2012 between 7:00 – 8:00. This observation occurred one day before an onset of an exacerbation. It is known that this patient had a hospital visit from July 02, 2012 to July 25, 2012 and the visit was not at the same day at which the maximum number of cough counts per hour occurred. There were 532 coughs in this day (see Figure 10).

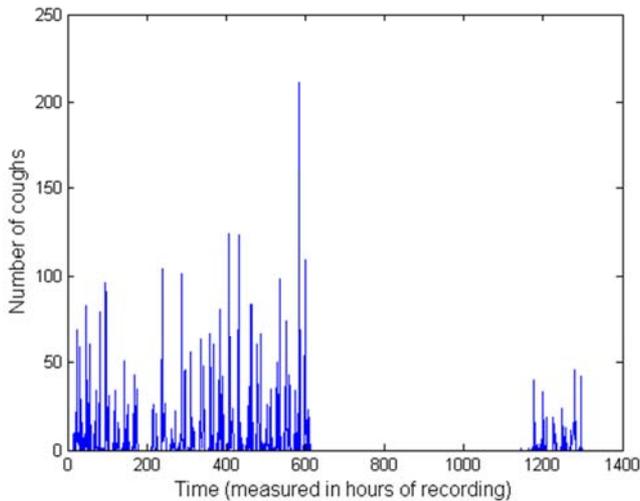


Figure 9. Number of coughs in an hour of the data set P1.

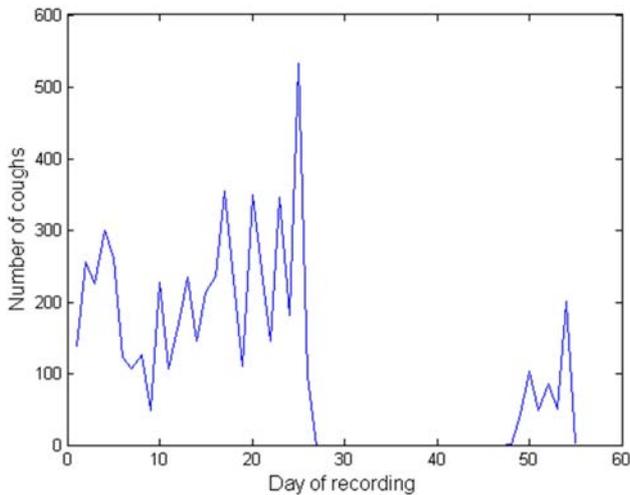
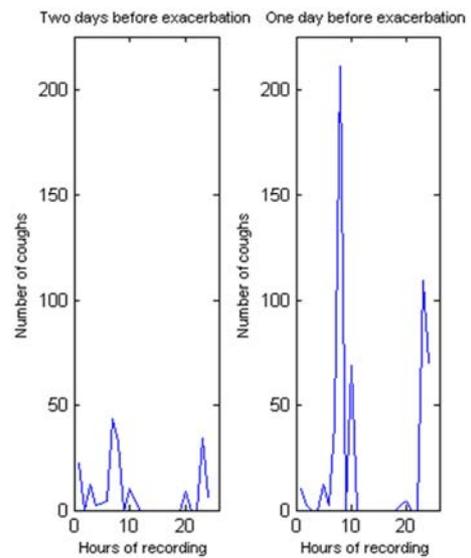


Figure 10. Number of coughs in a day of the data set P1.

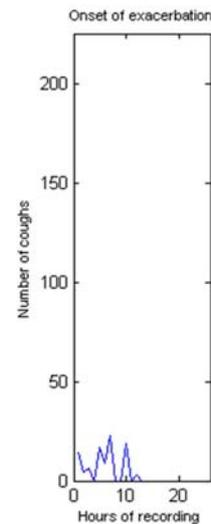
The number of coughs two days before the onset and two days after the end of the exacerbation were plotted. Note that there is no definite time of start of exacerbation and end of exacerbation but it can only be said when the patient went to hospital and when he/she returned.

From Figure 11(a), the patient had a maximum of 43 coughs in an hour two days before the onset of the exacerbation. One

day before the onset of the exacerbation, the patient had a maximum of 211 coughs in an hour. The total number of coughs during the day two days before an onset of the exacerbation was 181 coughs. The total number of coughs one day before the onset of the exacerbation was 532 coughs. On the day when the exacerbation started, shown in the right hand side figure of Figure 11(a), the total number of coughs was only 95 coughs; the patient went to the hospital and he/she was no longer at home. There was a maximum of 23 coughs in an hour on this day. The reason for the sudden drop in the number of coughs is not clear; it may be that the patient already received some medication before he/she went to hospital.



(a)



(b)

Figure 11. Number of coughs in an hour two days before the onset of an exacerbation and on the first day of an exacerbation of data set P1.

Figure 12 shows the number of coughs on the last day of the exacerbation and two days after the end of an exacerbation. On the day the exacerbation ended, the patient came back home from the hospital, the total number of coughs during the

hours that the patient was at home was 43. There were maximum 40 coughs in an hour on this day. However, on the first day after the end of the exacerbation the total number of coughs increased to 102 with a maximum of 33 in an hour. On the second day, after the end of the exacerbation, the total number of coughs was 49 with a maximum of 19 in an hour.

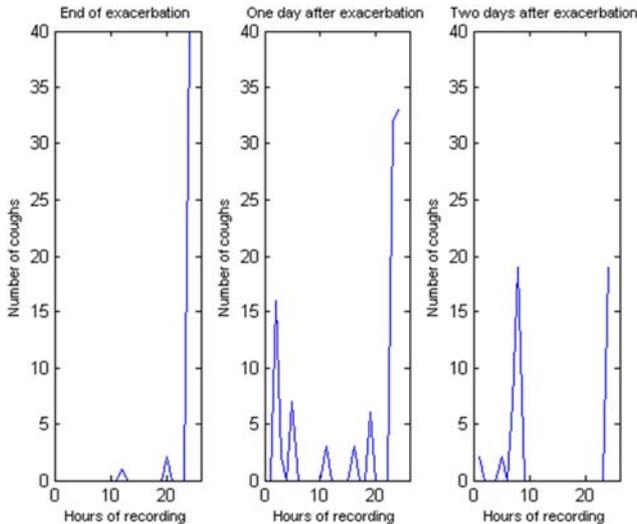


Figure 12. Number of coughs in an hour on the last day of exacerbation and two days after that of data set P1.

Data Set P8

The data set P8 covers the period from March 23, 2012 until May 7, 2012. The discharge date is March 21, 2012. The patient had a hospital visit from March 25, 2012 until March 28, 2012, which was caused by a moderate exacerbation. The onset of the exacerbation was at 20:13 on March 25, 2012 and it ended at 15:00 on March 28, 2012.

From Figure 13, the maximum number of coughs on a day was observed on March 24, 2012. It was about 429 coughs, which occurred the day before the onset of the exacerbation.

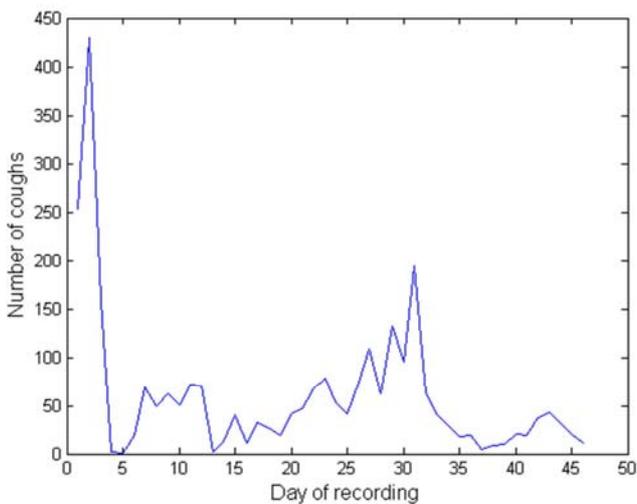


Figure 13. Number of coughs on a day of the data set P8.

The maximum number of coughs in an hour shown in Figure 14 was about 74, which was observed on March 24,

2012 between 1:00 and 2:00.

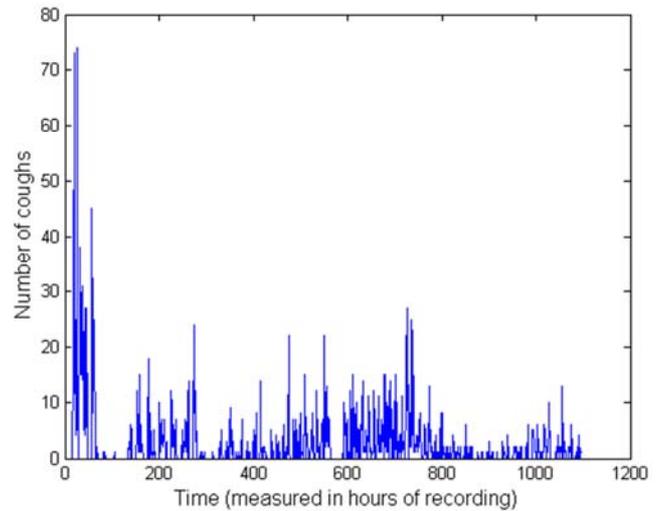


Figure 14. Number of coughs in an hour of the data set P8.

From Figure 15, two days before the onset of the exacerbation, the patient had a maximum of 73 coughs in an hour. One day before the onset of the exacerbation, the patient had a maximum of 74 coughs in an hour. The total number of coughs two days before the onset of the exacerbation was 253. The total number of coughs, one day before the onset of an exacerbation was 429 coughs. On the day the exacerbation started, the total number of coughs was only 159, because at that day the patient went to the hospital. There was a maximum of 45 coughs in an hour on this day.

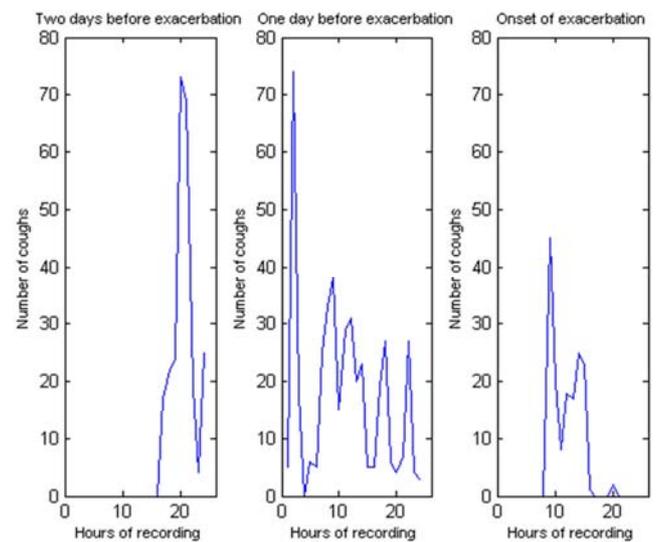


Figure 15. Number of coughs in an hour two days before and on the first day of an exacerbation of the data set P8.

Figure 16 shows the number of coughs on the last day of the exacerbation and two days after it ended. The patient came back from hospital on the day the exacerbation ended; the total number of coughs on this day was 19. There were maximum 6 coughs in an hour on this day. The total number of coughs on the first day after the end of an exacerbation increased to 69

with a maximum of 15 coughs in an hour. On the second day after the end of the exacerbation, the total number of coughs was 49 coughs with a maximum of 18 coughs in an hour.

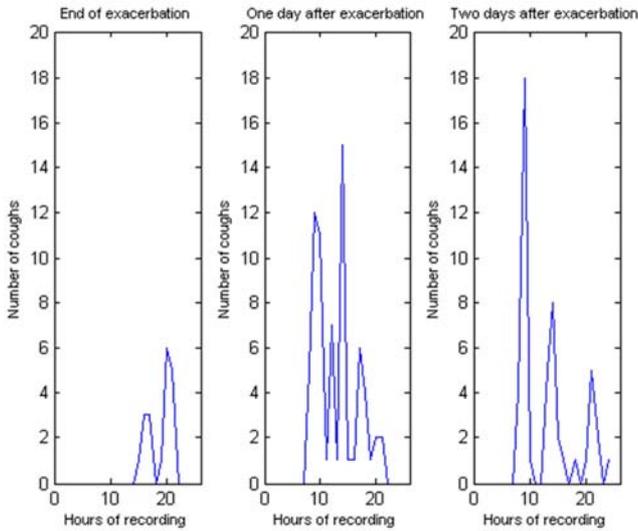


Figure 16. Number of coughs in an hour on the last day of exacerbation and two days after that of data set P8.

Data Set P12

The data set P12 was recorded from April 27, 2012 until June 8, 2012. The discharge date was on April 27, 2012. The patient had a hospital visit from April 30, 2012 until May 3, 2012 caused by a severe exacerbation. The onset of the exacerbation was at 6:47 on April 27, 2012, and it ended at 17:00 on May 3, 2012. From Figure 17, the maximum number of coughs on a day was observed on May 5, 2012 in which there were 72 coughs on this day.

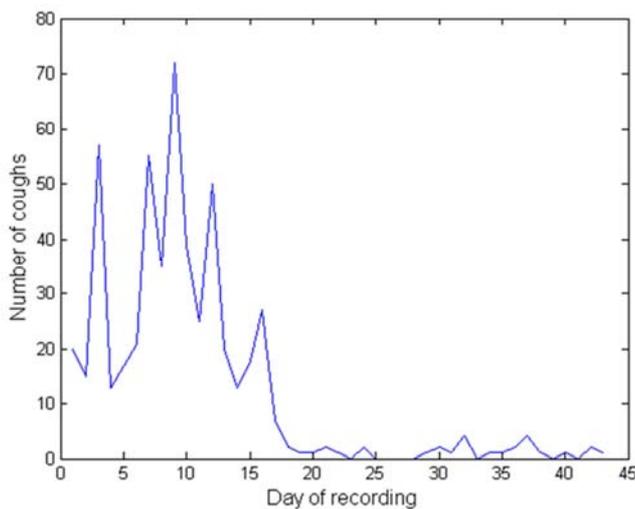


Figure 17. Number of coughs in a day of data set P12.

The number of coughs in an hour was shown in Figure 18. The maximum number of coughs equals 15 and occurred on April 29, 2012, between 11:00 and 11:59 and on May 3, 2012 between 20:00 and 21:00.

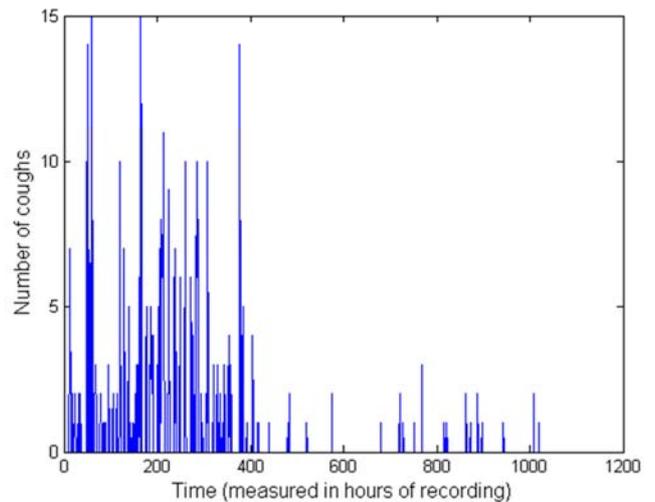


Figure 18. Number of coughs in an hour of data set P12.

From Figure 19, the patient had a maximum of 10 coughs in an hour two days before the onset of the exacerbation. One day before the onset of the exacerbation, the patient had a maximum of 15 coughs in an hour. The total day number of coughs two days before the onset of the exacerbation was 15. The total number of coughs one day before the onset of the exacerbation was 57. On the day that the exacerbation started, the total number of recorded coughs was only 13 because the patient went to the hospital and was not at home the whole day. There was a maximum of 3 coughs in an hour during the hours that the patient was at home.

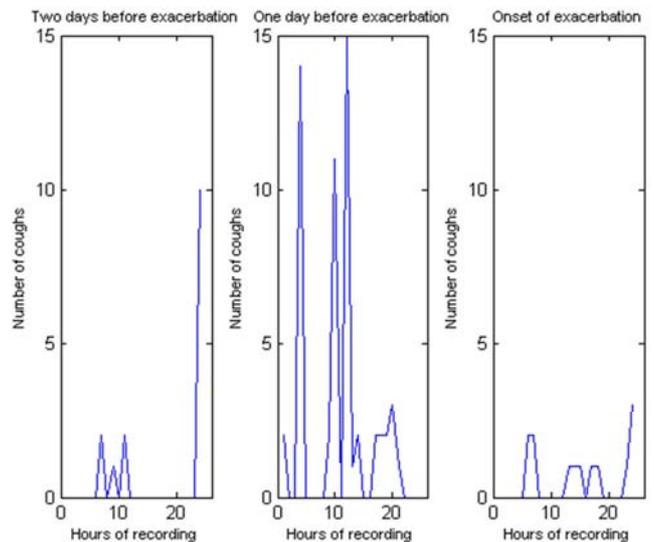


Figure 19. Number of coughs in an hour two days before and on the first day of an exacerbation of data set P12.

Figure 20 shows the number of coughs on the last day of the exacerbation and two days after its end. On the day the exacerbation ended and the patient came home, the total number of coughs was 55. There were maximum 15 coughs in an hour during the hours the patient spent at home. However, on the first day after the end of the exacerbation, the total number of coughs decreased to 35 with a maximum of 5

coughs in an hour. On the second day after the end of the exacerbation, the total number of coughs increased again to 72 with a maximum of 12 coughs in an hour.

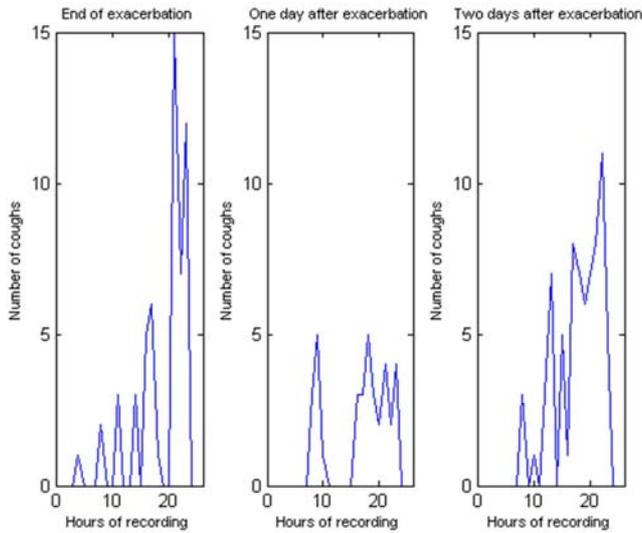


Figure 20. Number of coughs in an hour on the last day of exacerbation and two days after that of data set P12.

3.3. Results of Analysis of Variance (ANOVA) for P13

Analysis of variance is used to test whether the averages of the different groups of hours of a day are the same or not. From Table 3 and Table 4, the p-value is very small, which indicates that at 1% and 5% of significance, the null hypothesis, which states that all the means of the relative number of coughs of different hours of a day are identical, is rejected. Thus, from Table 4, at least one of the population means of the number of coughs in an hour of the three grouped hours is different from the rest means.

Table 3. One-way ANOVA between the numbers of coughs in an hour of the groups of hours 0:00-12:00 and 12:00-24:00 for data set P13.

Source	SS	df	MS	F	Prob>F
Columns	0.34207	1	0.34207	192.8	2.2155e-040
Errors	1.87004	1054	0.00177		
Total	2.21211	1055			

Table 4. One-way ANOVA among the number of coughs in an hour of hours 0:00-8:00, 8:00-16:00, and 16:00-24:00 for data set P13.

Source	SS	df	MS	F	Prob>F
Columns	0.3149	2	0.15745	87.39	7.70965e-036
Errors	1.89721	1053	0.0018		
Total	2.21211	1055			

3.4. Results of Markov Chain Analysis for P13

The cough events were considered as a discrete time process and the time stamps were counted hourly and daily. An hour (the smaller period) is important to compare events during different days (larger period). According to the description of Markov Chain in Section 2, the parameter values were fixed as follows: 1 time unit is 1 hour, n = 3600, d = 24. The value of ‘T’ is different for each data set because

it depends on the recording period for each patient. Thus, the number of cough occurrences were counted daily and each day were considered hourly.

When the relative frequency was used, a care has to be taken and the data upfront have to be screened. The numbers of coughs only during a specific day were compared. This means that the relative frequency of the day with low number of coughs during the whole day can be the same as the relative frequency of the day with a high number of coughs of another day.

In this case, $\alpha = 1/24$ and $\beta = 3/24$ were chosen. That means, in State 0, at a specific hour, the patient has less than 4% of the number of coughs during the day. In State 1, the patient has in specific hour between 4% and 13% of the number of coughs during the day. Finally, in State 2, the patient has in a specific hour more than 13% of the number of coughs during the day.

When a band is used to define the states, $\theta = 3$ was chosen. In particular, the considered band boundaries are:

- 1) The average value as lower bound and three times the standard deviation as upper bound.
- 2) The average value as lower bound and three times the inter-quartile range as upper bound

The figure of relative number of coughs in an hour suggests that roughly: mean = std = IQL, but this is not a general observation over all patients (See Figure 21).

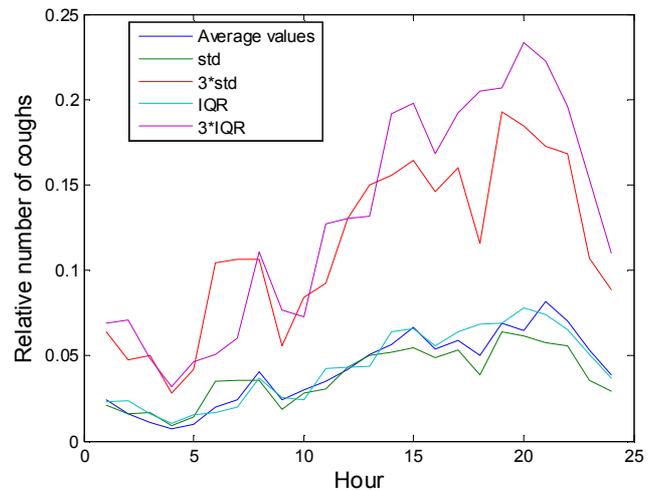
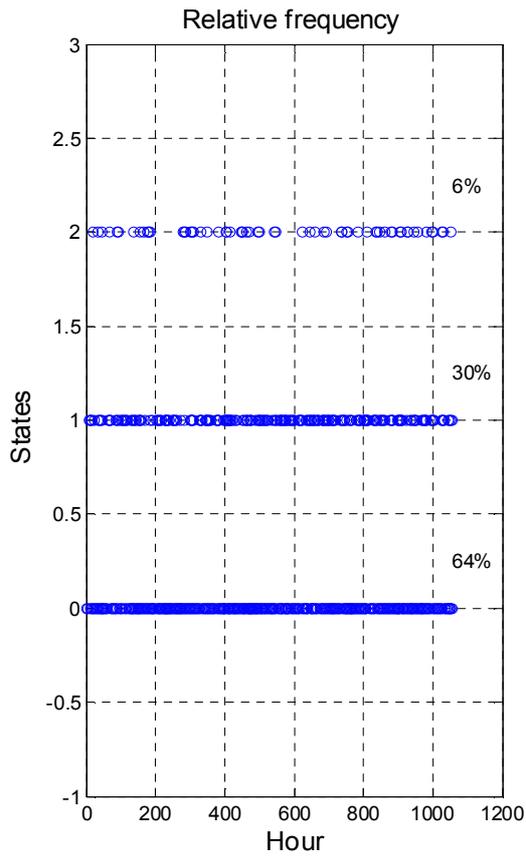
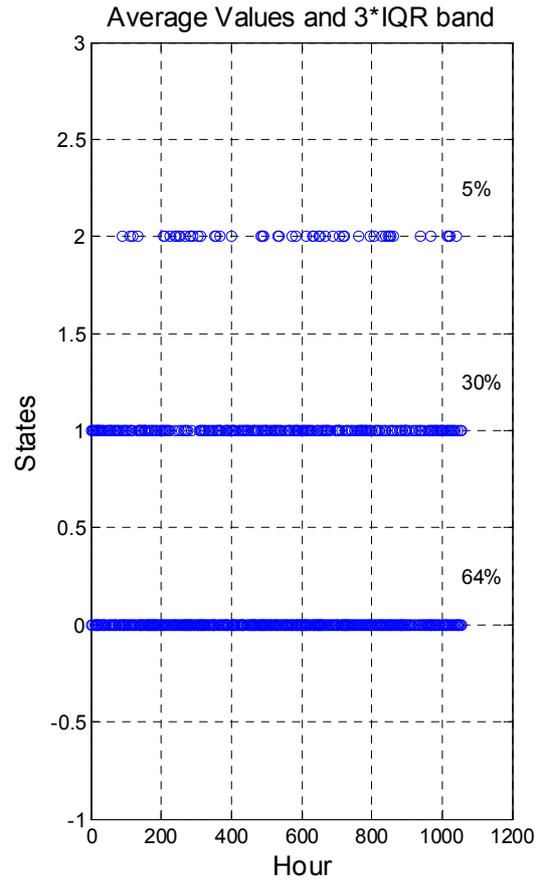


Figure 21. Average relative number of coughs in an hour during a day with its respective standard deviation and inter-quartile range over 44 days of experiment for patient P13.

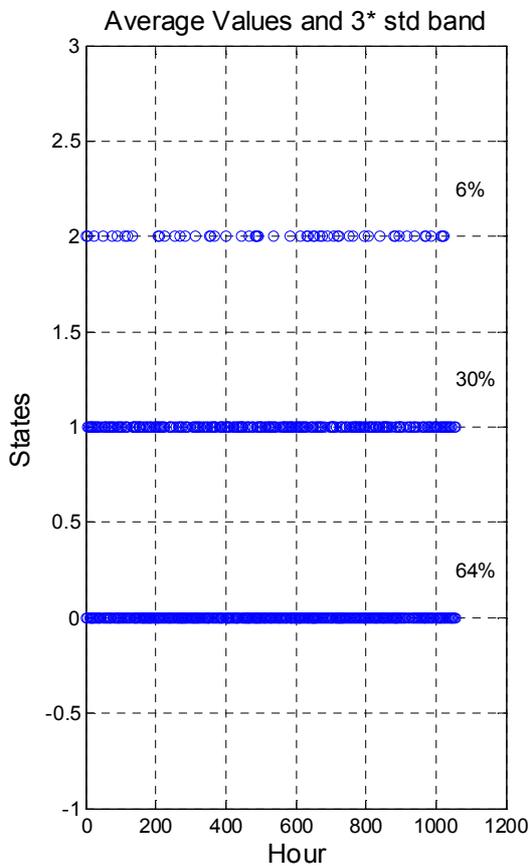
In both band cases, the lower bound is the average values of relative number of coughs in a given hour. In State 0, the patient has a relative number of coughs in a given hour less than the average relative number of coughs in that hour over the whole experimental period. The distribution of the data between different states is similar. i.e., 64% of them are in State zero, 30% of them are in State 1 and 6% of them are in State 2 (See (a), (b), and (c) of Figure 22).



(a)



(c)



(b)

Figure 22. Distribution of states using relative frequency, average values and three times standard deviation bounds, and average value and three times of inter-quartile range bounds for data set P13.

For all three cases, the matrices of one-step transition probabilities of defining the Markov states were determined. These were denoted as follows: when using relative frequency, P_r^1 and P_r^n represent the one-step and n^{th} -step transition probability matrix respectively. P_{std}^1 and P_{IQR}^1 represent the one-step transition probability matrix when using standard deviation and inter-quartile range, respectively. P_{std}^n and P_{IQR}^n represent the n^{th} -step transition probability matrix when using standard deviation and inter-quartile range, respectively.

The entries in the first rows of the matrices (shown in Table 5) represent the probabilities for the various kinds of state following after State 0. Similarly, the entries in the second and the third rows represent the probabilities for the various kind of coughing state following State 1 and State 2, respectively.

For the relative frequency, the probability that the process stays in State 0 after it was in State 0 is about 75%. It is the highest probability. After nine hours ($n = 9$) the cough predictions can be assumed independent of the relative frequency of coughs in the initial hour. This means that the probability of being in State 0 nine hours after State 0 is the same as the probability of being in State 0 nine hours after State 1 or nine hours after State 2. The power 'n' refers to the

stable hour or the stable state of the process. Therefore, the process gets to a stable point with the stable distribution given as:

$$P_r^9 = (0.6360 \ 0.3014 \ 0.0626)$$

Here, it is assumed that the probabilities are equal if they have the same four numbers after the decimal places. Table 5,

Table 5. The 1st-step and nth-step transition probabilities using relative frequency, average and three times of standard deviation, and average and three times of inter-quartile range.

Method	1 st -step	n th -step
Using relative frequency	$P_r^1 = \begin{pmatrix} 0.7474 & 0.2214 & 0.0312 \\ 0.4511 & 0.4290 & 0.1199 \\ 0.3939 & 0.5000 & 0.1061 \end{pmatrix}$	$P_r^n = \begin{pmatrix} 0.6360 & 0.3014 & 0.0626 \\ 0.6360 & 0.3014 & 0.0626 \\ 0.6360 & 0.3014 & 0.0626 \end{pmatrix}$
Using average and 3.std	$P_{Std}^1 = \begin{pmatrix} 0.6593 & 0.2878 & 0.0529 \\ 0.6222 & 0.3238 & 0.0540 \\ 0.6000 & 0.3000 & 0.1000 \end{pmatrix}$	$P_{Std}^n = \begin{pmatrix} 0.6449 & 0.2993 & 0.0558 \\ 0.6449 & 0.2993 & 0.0558 \\ 0.6449 & 0.2993 & 0.0558 \end{pmatrix}$
Using average and 3.IQR	$P_{IQR}^1 = \begin{pmatrix} 0.6593 & 0.2819 & 0.0587 \\ 0.6062 & 0.3563 & 0.0375 \\ 0.6909 & 0.2545 & 0.0545 \end{pmatrix}$	$P_{IQR}^n = \begin{pmatrix} 0.6449 & 0.3030 & 0.0521 \\ 0.6449 & 0.3030 & 0.0521 \\ 0.6449 & 0.3030 & 0.0521 \end{pmatrix}$

To look into the dynamics of the number of coughs, the whole data collection period, 44 days, was divided into h day periods with (h-1) day overlap (See Figure 23). h=5 was chosen. Thus, five day periods with four day overlap was considered and hence, 40 periods were obtained. h=5 was chosen because after plotting the relative frequency for each day, It was observed that there were some cyclic behavior of five days.

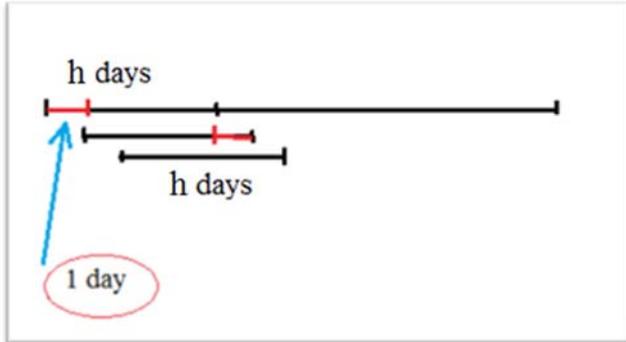
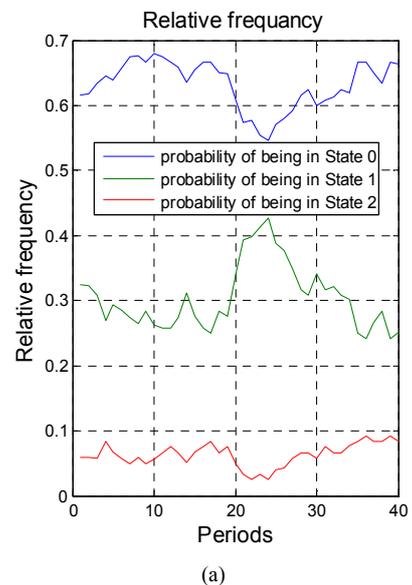


Figure 23. Scheme of looking into dynamics.

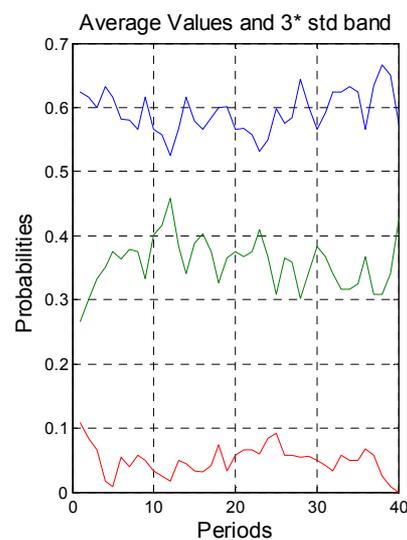
For each of the 5 day periods, the transition matrix, stable distribution, and the stable hour were determined. There is a higher probability of either being in State 0 or being in State 1 after State 2 in most of the periods. There is also a high probability to stay in State 0 after State 0 (Figure 24). As it can be seen from the dynamics of probabilities in Figure 24, the probabilities of being in State 0 and the probability of being in State 1 are symmetric. The probability of being in State 2 is almost stable for each definition of Markov states.

also presents the stable distribution when using average and three time standard deviation band,

$P_{std}^1 = (0.6449 \ 0.2993 \ 0.0558)$, which is stable at the 4th hour. If the average and three times inter-quartile range band is used, the system is stable at the 5th hour with a distribution $P_{IQR}^1 = (0.6449 \ 0.3030 \ 0.0521)$ (See Table 5).



(a)



(b)

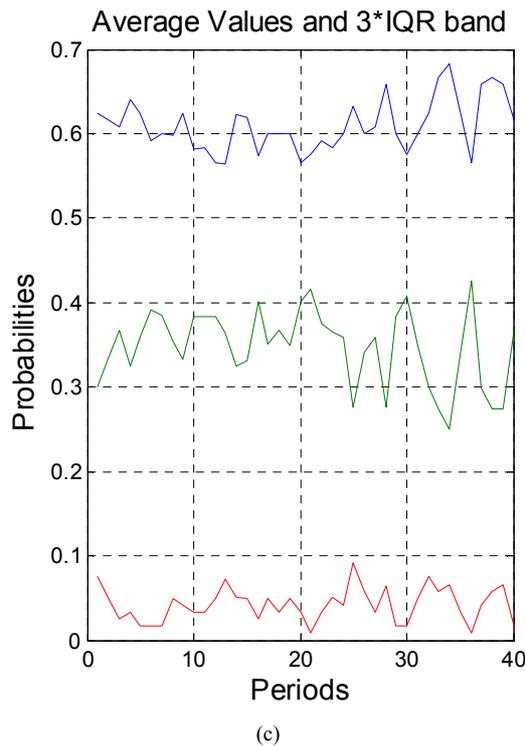
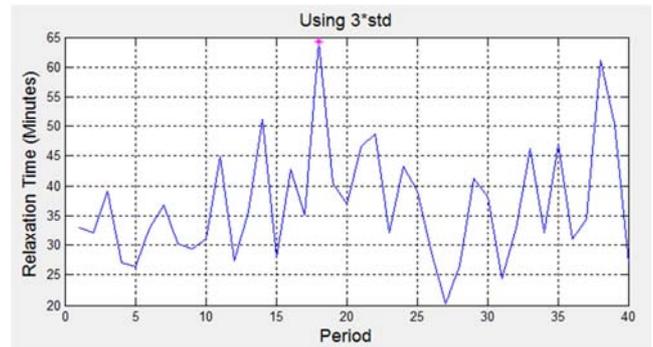


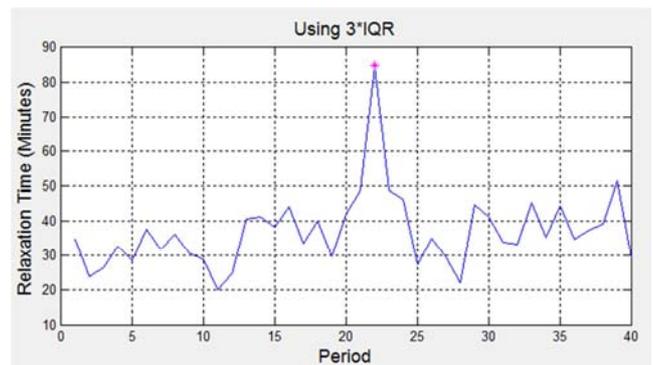
Figure 24. Dynamics of the probabilities in different states of the data set P13.

The eigenvalues of the one-step transition matrices were calculated and then the largest eigenvalue was taken to calculate the relaxation time.

If relative frequency is used to define states (Figure 25(a)), the longest relaxation time is 94 minutes in the 13th 5-day period and the shortest relaxation time is 30 minutes in the 20th period. If average and three times standard deviation is used to define the Markov states (Figure 25(b)), the longest relaxation time is 64 minutes in the 18th period and the shortest relaxation time is 20 minutes in the 27th period. The definition of states with inter-quartile range yields a longest relaxation time of 85 minutes and in the 22nd period and a shortest relaxation time of 20 Minutes in the 11th period (See Figure 25(c)). From all cases, the adding and removing of one day has large effect of the relaxation time. When the Markov states are defined using bands, similar dynamics pattern of the relaxation time can be observed (see (a), (b), and (c) of Figure 25).



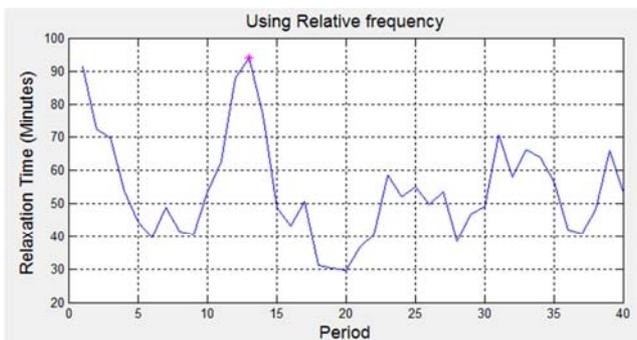
(b)



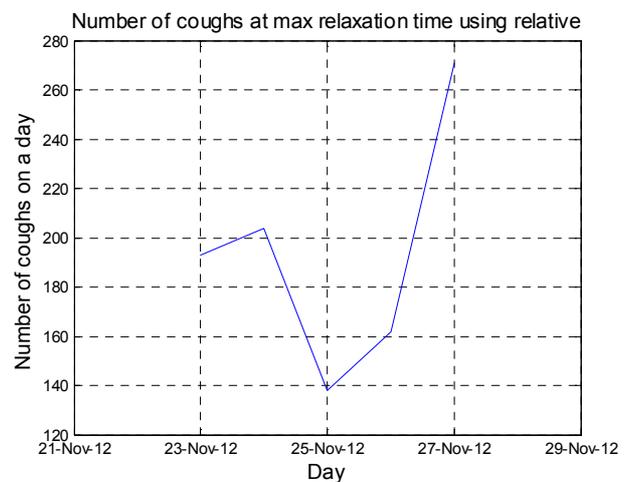
(c)

Figure 25. Dynamics of relaxation time of the data set data set P13.

From (a), (b), and (c) of Figure 26, the longest relaxation time is due to the highest relative number of coughs in an hour at the second day and the last day of the 5-day period when using relative frequency to define the Markov states. When using three times standard deviation, it is due to the highest relative number of coughs in an hour at the second day and the fourth day of the 5-day period. It is also seen that, when using three times inter-quartile range, the longest relaxation time is due to the highest relative number of coughs at the first day, 3rd day, and 5th day of the 5-day period.



(a)



(a)

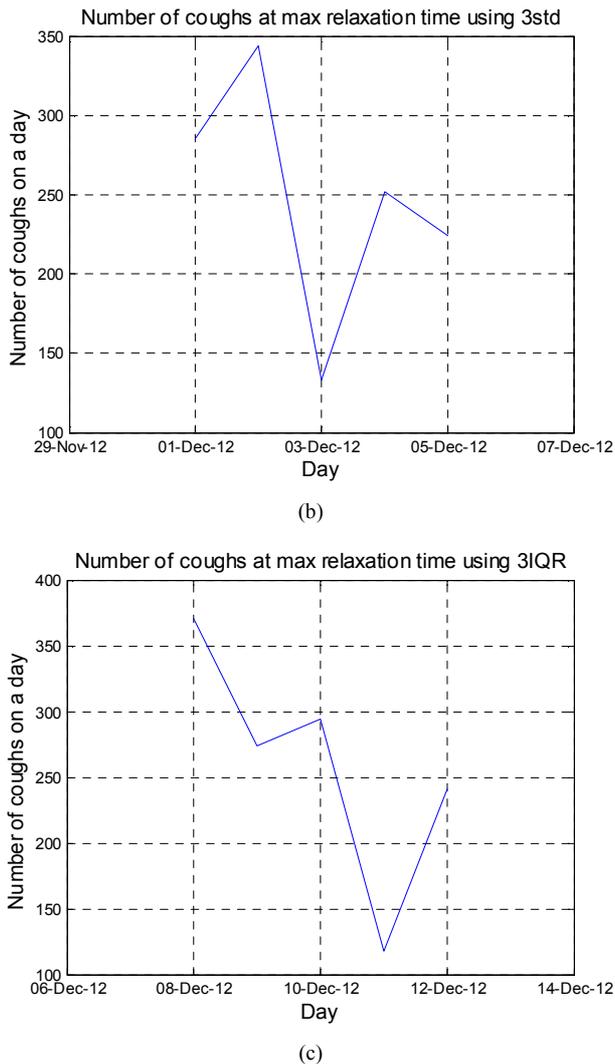


Figure 26. Number of coughs on a day for which the relaxation is the highest for P13.

4. Conclusions and Recommendations

Methods of statistics and stochastic processes were applied and tested to the time series that consist of the number of cough events per hour of 16 COPD patients. The data were recorded in the living room and in the bedroom of the patient's house and then combined.

From the general review of the data, we conclude that for each patient the periods of recording is patient specific and related to a specific number of days of recording with varies pattern over time. Some patients were hospitalized or data was not present at some days. Some data includes information that cannot be explained in logic way, for instance, there were hours of recording with zero data. The data does not review whether the patients were not at home or whether they did not cough. Some data sets consist of 80% of hours without coughs.

The summary statistics showed that the distribution of the number of coughs in an hour was skewed to the right. This indicated that the probability of having less cough is high. From the plotting of the relative frequency for each day, we

concluded that some cyclic behavior of five days period was observed.

Further, two cough distributions were considered: the distribution of the number of coughs per day and for any hour out of 24, the distribution of the number of coughs for that hour during the recording period. We conclude that these distributions show high variation.

From the analysis of patients with exacerbation, we conclude that the data sets P1 and P8 reveal a similar pattern. The maximum number of coughs in an hour and the maximum number of coughs on a day for both patients happened one day before the onset of the exacerbation. Two days before the onset of exacerbation, the patients had a higher number of coughs as compared to other days. For both patients, the number of coughs increased on the first day after the end of the exacerbation whereas on the second day after the end of the exacerbation, it became smaller. Therefore, the pattern of cough events can determine the exacerbation of the disease.

Analysis of variance (ANOVA) was used to identify if there is a difference between the average values of relative numbers of coughs, relative with respect to the total number of coughs per day, for period of hours in a day. The period referring to day time and night time are 8am - 8pm and 8:00pm - 8:00am respectively. From ANOVA results, we concluded that the night time coughing is not the same as the day time coughing. This means, the day time coughing is higher than the night time coughing for P13.

The Markov transition matrix represents the changing pattern of coughs. The transition matrix enabled us to distinguish a characteristic time of the pattern of coughs by overlapping five day periods with four day. The characteristic time shows the dynamics of the pattern. Generally, the patterns of cough events can be characterized by the relaxation time and stable distribution of a Markov Chain. These patterns are possible indicators of the underlying physical process and need to be compared to medical history of the patients in order to assess application scenarios based on this information.

A user friendly Matlab based graphical user interface (GUI) was developed that allows the user to identify the patterns of cough events of COPD patients together with their medical history. This tool is designed to display the detailed analysis of any time stamp data. All analysis methods used in this study to explore the cough event data were incorporated in the software application. The tool was ready for use but should be validated for its diagnostic properties.

We recommended that it is important to use the methods developed in this study to interpret the data and relate it to patient diagnostics and communicate the tool to potential users, doctors and researchers specialized in COPD. Thus, with the feedback generated in this way, steps can be set to professionalize the tool towards an instrument for diagnostic purposes.

Acknowledgements

We would like to thank the department of Mathematics for Industry at Eindhoven University of Technology (TU/e) and

the department of Smart Sensing and Analysis (SSA) at Philips Research for providing the timestamp data. We would like to thank Semira Mohammed Beyan and Ayele Taye Goshu for their valuable comments of this manuscript.

References

- [1] World Health Organization (WHO) (2019). Chronic obstructive pulmonary disease (COPD): What is COPD? Accessed from <http://www.who.int/respiratory/copd/en/> on May 22, 2019.
- [2] World Health Organization (WHO) (2008). A report on “World Health Statistics 2008”: ‘The top 20 causes of death in 2030’.
- [3] Dr Colin Tidy, (2014). Chronic Obstructive Pulmonary Disease. Accessed from: patient.info/health/chronic-obstructive-pulmonary-disease-leaflet.
- [4] Chung, K. F. (2006). Measurement of Cough. *Journal of Respiratory Physiology & Neurobiology*, 152, pp. 329-339.
- [5] Smith, J., Owen, E., Earis, J., Woodcock, A. (2006). Cough in COPD: Correlation of Objective Monitoring With Cough Challenge and Subjective Assessments; *Journal of Chest*, 130, pp. 379-385.
- [6] Kenny T. (2011). Chronic obstructive pulmonary disease. Available online.
- [7] Vestbo, J. and Rasmussen, F. (1989). Respiratory symptoms and FEV1 as predictors of hospitalization and medication in the following 12 years due to respiratory disease; *European Respiratory Journal*, 2, pp. 710-715.
- [8] Rennard, S., Decramen, M., Carverley, P., Pride, N., Soriano, J., Vermeire, P., Vestbo, J. (2002). Impact of COPD in North America and Europe in 2000: subjects’ perspective of confronting COPD international survey. *European Respiratory Journal*, 20, pp. 799-805.
- [9] Pauwels, R., Buist, A., Calverley, P. (2001). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *American Journal of Respiratory and Critical Care Medicine*; 163, pp. 1256-1276.
- [10] Vestbo J, Prescott E, Lange P. (1996). Association of chronic mucus hypersecretion with FEV1 decline and chronic obstructive pulmonary disease morbidity. Copenhagen City Heart Study Group. *American Journal of Respiratory and Critical Care Medicine*; 153, pp. 1530-1535.
- [11] Smith, J. and Woodcock, A. (2006). Cough and its importance in COPD. *International Journal of COPD*; 1 (3), pp. 305-314.
- [12] Burgel, P-R., Paillasseur, J-L., Caillaud, D., Escamillae, R., Court-Fortune, I., Perez, T. (2010). Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *European Respiratory Journal*; 36, pp. 531-539.
- [13] Hurst, J. R., Vestbo, J., Anzueto, A., Locantore, N., Müllerova, H., Tal-Singer, R., Miller, B., Lomas, D. A., Agusti, A., MacNee, W., Calverley, P., Rennard, S., Wouters, M., Wedzicha, J. A. (2010). Susceptibility to exacerbation in chronic obstructive pulmonary; *New England Journal of Medicine*, 363 (12), pp. 1128-1138.
- [14] Koninklijke Philips N. V. (2004-2017). About the company: Philips.
- [15] Penny, W. and Henson, R. (2006). Chapter 13: Analysis of Variance.