***Original Research Article***

# ESTIMATION OF TRANSITION PROBABILITY MATRIX AND OBSERVATION PROBABILITY DISTRIBUTION MATRIX OF MALARIA SYMPTOM DATA SET USING HIDDEN MARKOV MODEL

---

***Abstract:*** Clinical study of malaria presents a modeling challenge as patients disease status and progress is partially observed and assessed at discrete clinic visit times. Since patients initiate visits based on symptoms, intense research has focused on identification of reliable prediction for exposure, susceptibility to infection and development of severe malaria complications. Despite detailed literature on malaria infection and transmission, very little has been documented in the existing literature on malaria symptoms modeling, yet these symptoms are common. Furthermore, imperfect diagnostic tests may yield misclassification of observed symptoms. The main objective of this study is to develop a Bayesian Hidden Markov Model of Malaria symptoms in Masinde Muliro University of Science and Technology student population. An expression of Hidden Markov Model is developed and the parameters estimated through the forward-backward algorithm.

***Keywords:*** *Hidden Markov Model, Forward Variable, Backward Variable,*

## 1. INTRODUCTION

The term Malaria was first used by Dr. Fransisco Torti, but it was not until 1880 that scientist discovered that it was a parasitic disease caused by a unicellular protozoan of the genus *Plasmodium* which is transmitted by the anopheles mosquito. The most important vectors of malaria in Africa are members of the *An. Gambiae* complex and *An. funestus*. Five species of the *An. gambiae* complex are vectors of malaria, two species *An. gambiae sensu stricto* and *An. arabiensis* are vectors of malaria parasites widely distributed throughout sub-Saharan Africa. They often occur together but *An. arabiensis* dominates in drier areas whilst *An gambiae* dominates in more humid areas [66].

Malaria still remains a huge public health issue regardless of how many years of research has been con- ducted on how to combat this disease. According to WHO [79], the latest world malaria report released in November 2017 shows that the number of malaria cases reported in the year 2016 was 216 million up from 211 million cases reported in 2015. The report also shows that malaria death estimates in 2016 stood at 445,000 compared to 446,000 deaths in 2015. The high burden of malaria cases in 2016 was in Africa at 90% with 91% cases of deaths reported. In Kenya, malaria is one of the leading causes of morbidity with about 3.5 million children at risk of developing severe malaria out of which an estimated 34,000 children under five years dies every year. The disease is also responsible for 30% of out-patient visits at health centers, economically, it is estimated that 170 million working hours are lost each year because of malaria illness [80].

Symptoms are experienced deviations from an individual's perception of his or her normal healthy state of being, yet not necessarily an indicator of illness. A symptom can emerge from sensitivity to certain combinations of biological, social and environmental processes and vary in magnitude, severity, persistence and character. Symptoms can be subjectively reported or objectively observed. Depending on the disease, the scope and intensity, the duration of symptoms can vary over time. The malaria symptoms can be grouped into two; symptoms for uncomplicated malaria (suspected malaria) and symptoms for complicated malaria (severe malaria). Malaria is considered uncomplicated when symptoms are present but there are no clinical or laboratory signs to indicate severity or vital organ dysfunction. The symptoms of uncomplicated malaria are non-specific i.e. they are self-reported symptoms that do not indicate a specific disease process, they are initial symptoms and include fever (temperature), chills, headache, pains (joint, muscle, abdominal), muscle aches, loose stool, tiredness, nausea, high pressure, vomiting and diarrhea. Infection with *Plasmodium falciparum* if not promptly treated can quickly progress to complicated malaria (severe malaria). The main symptoms of severe malaria include coma, severe breathing difficulties, low blood sugar, hallucination, prostration, immobility, confusion and incoherent speech, seizure, loss of consciousness, hyperparasitaemia, black quarter urine and low blood hemoglobin [17]

## 2. LITERATURE REVIEW

In many studies of medical treatment, symptoms are measured repeatedly over time in observation called longitudinal observation. Though we cannot observe directly latent variables, we learn about it by measuring symptom. For the longitudinal models, two latent variables govern disease, one for the probability of experiencing a particular symptom and another for the severity of the experienced symptom. Thus the probability of a symptom and the severity of it depends on both latent variables and observed variables [83]. Latent variables are variables that are not directly observed but are inferred through a mathematical model from other variables that are directly observed or measured. A latent variable model is a statistical model that contains latent i.e. unobserved variables. These variables can either be discrete or continuous. Sometimes latent variables corresponds to aspects of physical reality which could in principle be measured but may not be for practical reason thus in this situation the term hidden variable is commonly used. One advantage of using latent variables is that they can serve to reduce the dimensionality of data. Latent variable link observable data in the real world to symbolic data in the model. Bayesian statistics is often used for inferring latent variables, the common method used inferring latent variables in Bayesian statistics is often used for inferring latent variables, the common method used inferring latent variables in Bayesian statistics are; Hidden Markov Model (HMM), factor analysis, principal component analysis and Expectation Maximization (EM) algorithm [83].

Zammit *et al* [84] developed an intra-individual consistency model using a logistic- type latent variable model. The latent variable in the model was used to represent the propensity of symptoms and intensity of episodes as these could not be observed directly and needed to be estimated through observation of symptoms episodes in hypoglycemia. The model results showed that there was individual difference in symptom reporting and that adults exhibit distinct intra-individual variability in symptom reporting. Hans *et al* extended on the model developed by Zammit *et al* by allowing for different forms of symptom experiencing thresholds between groups variability when symptoms are classified in groups and performing variable selection to determine a predictive model for the effect of patient characteristics and their interactions on symptom consistency. The study was conducted in several health centers in the United Kingdom and data collected from 381 participants aged between 17-75 years. Bayesian estimation was performed for all coefficients in the developed model without grouped symptoms and with grouped symptoms. The analysis shows that a multiplicative form of symptom propensity and episode intensity provides the most suitable symptom experiencing threshold and groups of symptoms show distinct propensity and that gender subjects had significant impact on the consistency of symptom reporting.

The Hidden Markov Model (HMM) is a statistical method based on Markov Chain. It is a powerful tool for random processing and modeling which is normally used to predict and classify data. The first step to develop the HMM was taken by Rabinner 1989 after the presentation of the educational article of HMM by revealing the details of the complex models [64]. Most of the investigation using HMM have been done in non-medical fields, for instance, Cholewa and Glomb [13] investigated on estimation of the number of critical points in time sequence while Farsi [19] investigated on implementation and optimization of speech recognition system based on HMM using genetic algorithm. Some of the investigation of the medical fields using HMM include the study by Vimala *et al* [75] which used HMM to identify and classify Electrocardiography (ECG) signals. The results of Vimala confirmed that the HMM could be used as a powerful tool for grouping ECG signals into three signals. Also Li *et al* [69] used HMM to predict the progression of lung cancer among 508 patients in one of the Chinese hospitals from 2010 to 2012. The results showed that HMM was able to predict 0.81 accuracy while Lee *et al* [68] used HMM to classify snoring sounds of 21 patients with sleeping disorders. Recent applications of these models are study by Barber *et al* for predicting short winds [5], study of Wu *et al* for using a Bayesian non-parametric vector ARHMM for testing robot performance [81] and Tuncel *et al* for using autoregressive forests for model multivariate time series [74]. Therefore as the literature review shows, the development of HMM is more in the field of continuous observation and less attention has been paid to relation of the hidden states in the Discrete Hidden Markov Models (DHMM). Therefore this study aims at using (DHMM) with underlying first order Markov Chain.

A Hidden Semi-Markov Model (HSMM) is an extension of HMM designed to allow general (i.e. non-geometric or non-exponential) distribution for the state duration. A major drawback with HMM is the inflexibility in describing the time spent in a given state which is geometrically distributed. A discrete Hidden Semi-Markov chain is composed of non-observable state process which is a semi-Markov chain and a discrete output process which is an embedded first-order Markov chain representing the transitions between distinct states and discrete state occupancy distribution representing sojourn times in a non-absorbing states [20] An HSMM is constructed by adding a temporal component (duration) into HMM. Unlike a state in a standard HMM, a state in an HSMM generates a sequence of observational as opposed to a single observation in HMM [64].

## 3. MATERIALS AND METHODS

The data used in this study was part of records of patients (students) with different type of malaria severity disease who visited the health service in the time period of 1st January 2015- 20th December, 2015 and displayed different types of malaria related symptoms. Thirteen different malaria symptoms observed by the health officer for each individual student was recorded in students file. The recorded symptoms are; temperature, blood pressure, pulse beat/min, prostration, vomiting, rigors, diarrhea, pallor, convulsions, jaundice, coma, sweating and dark urine. Depending on the number of each symptoms an individual has, the observed symptoms are the recorded on an ordinal scale from 0 to 3 with 0 being no symptoms and 3 being maximum symptoms. Modeling is then performed directly at the level of observed symptoms.

### 3.1 THE STATISTICAL ANALYSIS

The statistical method used in this study was the Hidden Markov Model which is explained as follows;

### 3.1.1 THE MODEL

Let Z = $\{Z_1, Z_2, \ldots, Zn\}$ be the number of possible disease states an individual can be at any given time point. In this study, the disease states are divided into four different states (n=4) i.e. Z = $\{Z_1, Z_2, Z_3, Z_4\}$ where ($Z_1$)-Normal/healthy state, ($Z_2$)-mild state, ($Z_3$)-moderate state and ($Z_4$)-severe state as shown in figure 1. Modeling of the disease (malaria) is performed directly at the level of observed symptoms and these symptoms can only be observed at infectious state of an individual which in this study is represented by mild state, moderate state or severe state as shown in Figure 1. In Figure 1, each state is shown as a circle and state transmission represented by directed graph edge between states. The arrow goes from the hidden states (Z= $\{Z_1, Z_2, Z_3, Z_4\}$) to observed symptoms because the state of disease at which an individual is causes a particular symptom(s) to be observed. The hidden states are interconnected in such a way that any state can be reached from any other state. Thus, the state transmission from one state to the next state is a Markov process of order one and the next state depends on the current state and fixed probabilities. Let O = $\{x_1 ... x_p\}$ be the sequence of observed symptoms (observation sequence) in each state and $x_p$ the observed symptoms. Thirteen (p=13) different symptoms observed and recorded by the health officer from different students with malaria related symptoms are used in this study, the symptoms are; $x_1$ - body temperature, $x_2$ - blood pressure, $x_3$ - pulse beat/min, $x_4$ - sweating, $x_5$ - rigors, $x_6$ - vomiting, $x_7$- pallor, $x_8$ - diarrhea, $x_9$ - prostration, $x_{10}$- dark urine, $x_{11}$- coma, $x_{12}$- convulsions and $x_{13}$- jaundice .

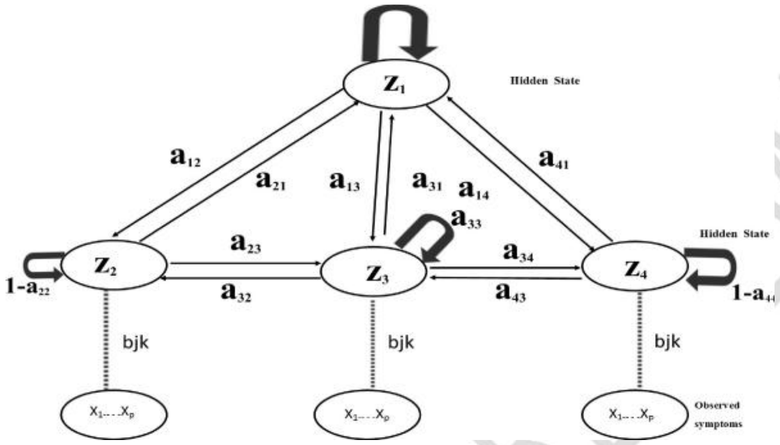

Figure 1: Hidden Markov Model with two disease states

Let $\pi = \{\pi_0(1), ..., \pi_0(n)\}$ be the initial state distribution vector; where $\pi$ is an initial probability for each state of the disease by which the Markov Chain begins to work. $\pi$ is shown as edge entering into disease state from state zero (start) which is not shown in figure 1, because we imagine that there is a silent state zero which all state originates from and therefore the system cannot transits to state zero but can only transit out of it.

$$\sum_{Z} \pi = 1 \tag{3.1}$$

Let $a_{ij}$ be the transition probability of the disease transiting from state i to state j i.e.

$$a_{ij} = p(Z_n = j | Z_{n-1} = i) \tag{3.2}$$

$a_{ij} \geq 0$, $\sum_{j=1}^{n} a_{ij} = 1$ and $1 \leq i, j \leq n$.

Let A be the transition probability matrix i.e a set of transition probabilities among states. In this study, Malaria disease has four disease states, therefore the transition probability matrix is represented as

$$A = \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array} \begin{array}{c} \begin{array}{cccc} Z_1 & Z_2 & Z_3 & Z_4 \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \end{array}$$

Let $b_{jk}$ be the probability of observing symptom(s) in each of the disease states i.e., the probability of observed symptom in each hidden state is given by

$$b_{jk} = P(O = x_k | Z = Z_j) \tag{3.3}$$

where

$$b_{jk} \geq 0, \quad \sum_{k=1}^{p} b_{jk} = 1, \quad 1 \leq k \leq p, \quad 1 \leq j \leq n$$

Let B be the probability distribution observation matrix i.e. B= [ $b_{jk}$ ]. In this study, there are 13 observed symptoms and four different disease states. Therefore the probability distribution of observed symptom in each hidden state (observation matrix) given by;

$$B = \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array} \begin{array}{c} \begin{array}{ccccccccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} & x_{13} \end{array} \\ \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} & b_{1,5} & b_{1,6} & b_{1,7} & b_{1,8} & b_{1,9} & b_{1,10} & b_{1,11} & b_{1,12} & b_{1,13} \\ b_{2,1} & b_{2,2} & b_{2,3} & b_{2,4} & b_{2,5} & b_{2,6} & b_{2,7} & b_{2,8} & b_{2,9} & b_{2,10} & b_{2,11} & b_{2,12} & b_{2,13} \\ b_{3,1} & b_{3,2} & b_{3,3} & b_{3,4} & b_{3,5} & b_{3,6} & b_{3,7} & b_{3,8} & b_{3,9} & b_{3,10} & b_{3,11} & b_{3,12} & b_{3,13} \\ b_{4,1} & b_{4,2} & b_{4,3} & b_{4,4} & b_{4,5} & b_{4,6} & b_{4,7} & b_{4,8} & b_{4,9} & b_{4,10} & b_{4,11} & b_{4,12} & b_{4,13} \end{bmatrix} \end{array}$$

Therefore the Hidden Markov Model is specified using 3 parameters and defined as

$$\mu = (A, B, \pi) \tag{3.4}$$

Where A is the transition probability matrix, B is the probability distribution observation matrix and $\pi$ is the initial state distribution.

### 3.1.2 Computation of P $(Z|O)$

Let $Z_n = Z_1, Z_2, \ldots, Z_n$ where n=4 be the hidden states at all time points (days, hours, months, etc) . let $z = \{z_1, z_2, \ldots, z_T\}$ whose elements $z_t$ is a state at time point t where z follows the Markov property i.e given previous state $z_{t-1}$ of process z, the conditional probability of current state $z_t$ is only dependent on the previous state $z_{t-1}$ and is not relevant to any further state. ( $z_{t-3}, z_{t-2}, ..., z_1$ ) i.e.

$$P(z_{t-1}, z_{t-2}, ..., z_1) = P(z_t | z_{t-1})$$

(Equation 3.5)

Each vertical slice in Figure 3.2 represents a time step.

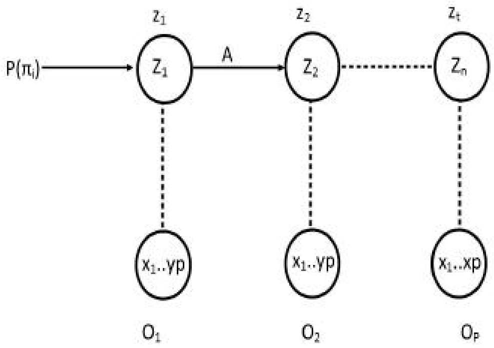

Figure 3.2: A representation of two-state HMM as a graphical model

The top node represents the variable $z_t$ and the bottom node represents the observable $0_t$ variable. Using Figure 2, Equation (3.5) can be written as

$$P(z_t = Z_j | z_{t-1} = Z_i) \;\; or \;\; P(z_{t+1} = Z_j | z_t = Z_i)$$

Equation (3.6)

Let O = $\{O_1, O_2, ..., O_T\}$ be the sequence of observation whose element $O_t$ is an observation at time point t and T is the length of the sequence.

Let $\pi_i = P(z_1 = Z_i)$ be the initial probability distribution.

From figure 2, we can read off various conditional independencies but the main conditional independency of interest is obtained by conditioning on a single state node. Conditioning on $z_t$ renders $z_{t-1}$ and $z_{t+1}$ independent. Conditioning on an output node on the other hand does not separate nodes in the graph and thus does not yield any conditional independencies. Therefore the joint probability is obtained by taking a product over the local conditional probabilities and for a particular sample point (z, O) we obtain the following joint probability

$$P(z, O) \;\; = \;\; P(z_1, z_2, ..., z_T, O_1, O_2, ..., O_T)$$

Equation (3.7)

which can be simplified as;

$$P(z,O) = P(z_1)[\prod_{t=1}^{T-1} p(z_t|z_{t+1})]\prod_{t=1}^{T} P(O_t|z_t)$$

Equation (3.8)

where $P(z_t \mid z_{t+1})$ is the transition probability A, $P(O_t \mid z_t)$ is the observation/emission probability and $P(z_1)$ is the initial state distribution.

Using the HMM $\mu$ defined in Equation (3.4), we have $A = [a_{ij}] = a_{z_t,z_{t+1}} = P(z_{t+1}, z_t)$, $B = [b_{jk}]$ and $\pi = z_1$

Equation (3.8) becomes;

$$P(z,O) = P(\pi_{z_1})[\prod_{t=1}^{T-1} a_{z_t,z_{t+1}}]\prod_{t=1}^{T} P(O_t|z_t,\mu)$$

Equation (3.9)

Where

$$\begin{aligned} a_{z_t,z_{t+1}} &\equiv [a_{ij}]z_t^i z_{t+1}^j \\ \pi_{z_1} &\equiv \prod_{i=1}^{Z}[\pi_i]z_1^i \qquad (z_1 = Z_1) \end{aligned}$$

Since we are required to compute the probability of a hidden vector z given an observation output O, we then solve the probability $P(z \mid O)$ using conditional probability as follows

$$P(z|O) = \frac{P(z,O)}{P(O)}$$

Equation 3.10

The numerator $P\,(z,\,O)$ is simplified by substituting the particular values of z and O into Equation (3.10) while the denominator $P\,(O)$ involves computing the sum across all the possible values of the hidden states as follow;

$$P(O) = \sum_{z} P(z, O)$$

Equation 3.11

upon simplification and substituting the value of *P* (*z, O*), we get

$$P(O) = \sum_{z_1} \sum_{z_2} \cdots \sum_{z_T} \pi_{z_1} \prod_{t=1}^{T-1} a_{z_t, z_{t+1}} \prod_{t=1}^{T} P(O_t | z_t, \mu)$$

Therefore Equation (3.10) can be written as,

$$P(z|O) = \frac{\pi_{z_1} \prod_{t=1}^{T-1} a_{q_t, z_{t+1}} \prod_{t=1}^{T} P(O_t | z_t, \mu)}{\sum_{z_1} \sum_{z_2} \cdots \sum_{z_T} \pi(z_1) \prod_{t=1}^{T-1} a_{z_t, z_{t+1}} \prod_{t=1}^{T} P(O_t | z_t, \mu)}$$

Equation 3.12

From Equation (3.12), each state node $Z_t$ can take on Z values since we have T state nodes, this implies that we must perform $Z^T$ sums to observe all the hidden states, calculation of $Z^T$ sums is infeasible. According to Rabinner [64] rather than calculating $P(z | O)$ for the entire sequence z, we focus on a particular state node $z_t$ and calculating its posterior probability i.e., $P(z_t | O)$

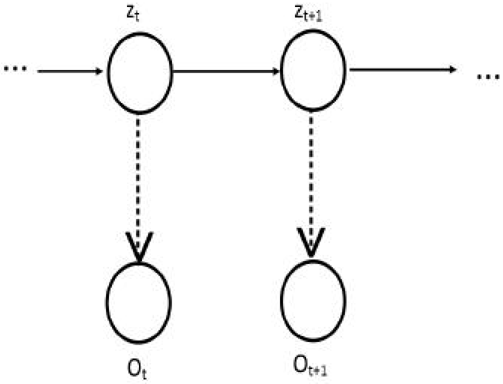

Figure 3.3: A fragment of the graphical model representation of an HMM.

To compute $P(z_t | O)$, we use Figure (3) to get the conditional probabilities by conditioning on a state node. This is achieved by reversing the terms in $z_t$ and O via an application of the Baye's rule as follow; using conditional probability, we have

$$P(z_{(t)}|O) = \frac{P(z_t, O)}{P(O)}$$

Equation (3.13)

applying the notion of independence, we get

$$P(z_t, O) = P(z_t|O)P(O) \quad and \quad P(O, z_t) = P(O|z_t)P(O)$$

Equation (3.14)

applying Bayes rule in Equation (3.14), we get

$$P(z_{(t)}|O) = \frac{P(O|z_{(t)})P(z_{(t)})}{P(O)}$$

Equation (3.15)

applying conditional independence using Figure 3, we obtain $P(z_t \mid O)$ as follow;

$$\begin{aligned} P(z_{(t)}|O) &= \frac{P(O_1, O_2 ..., O_t, O_{t+1}, O_{t+2}, ..., O_T|z_{(t)})}{P(O)} \\ &= \frac{P(O_1, ..., O_t|z_{(t)})P(O_{t+1}, ..., O_T|z_t)P(z_t)}{P(O)} \end{aligned}$$

Equation

(3.16)

Let $\alpha(z_t) \equiv P(O_1, ..., O_t, q_t)$ be the probability of emitting a partial sequence of output $O_1, ..., O_t$ and ending up in state $z_t$ and $\beta(z_t) \equiv P(O_{t+1}, ..., O_T|z_t)$ be the probability of emitting a partial sequence of output $O_{t+1}, ..., O_T$ given that the system starts in state $z_t$.

Then Equation 3.16 can be written as;

$$P(z_{(t)}|O) = \frac{\alpha(z_{(t)})\beta(z_{(t)})}{P(O)}$$

(Equation 3.17)

where $\alpha(z_t)$ and $\beta(z_t)$ are vector with component $\alpha(z_t^i)$ and $\beta(z_t^i)$.

Given that the sum $P(z_t|O)$ over the components of $z_t$ must equal to one, then we obtain

$$P(O) = \sum_i \alpha(z_t^i)\beta(q_t^i)$$

Equation (3.18)

Let $\gamma(z_t)$ be the posterior probability. Then $\gamma(z_t)$ is defined as

$$\gamma(z_{(t)}) \equiv \frac{\alpha(z_{(t)})\beta(z_{(t)})}{P(O)}$$

Equation (3.19)

where P(O) is computed once as normalization constant for a particular arbitrary choice of t.

Given that $\alpha(z_t)$ depends only on quantities up to time t and using the Markov properties in the model, we obtain recursion between $\alpha(z_t)$ and $\alpha(z_{t+1})$ in figure 3.3. Upon simplification, the forward recursion is obtained as follow;

$$\begin{aligned}
\alpha(z_{t+1}) &= P(O_1, ..., O_{t+1}, z_{t+1}) \\
&= P(O_1, ..., O_{t+1}|z_{t+1})P(z_{t+1}) \\
&= P(O_1, ..., O_t|z_{t+1})P(O_{t+1}|z_{t+1})P(z_{t+1}) \\
&= P(O_1, ..., O_t, z_{t+1})P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(O_1, ..., O_t, z_t, z_{t+1})P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(O_1, ..., O_t|z_t)P(z_t)P(z_t)P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(z_1, ..., O_t|z_t)P(z_{t+1}|z_t)P(_t)P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} P(O_1, ..., O_t, z_t)P(z_{t+1}|z_t)P(O_{t+1}|z_{t+1}) \\
&= \sum_{z_t} \alpha(z_t)a_{z_t,z_{t+1}}P(O_{t+1}|z_{t+1})
\end{aligned}$$

Equation (3.20)

For the beta variable we obtain "a backward" recursion by expressing $\beta(z_t)$ in terms of $\beta(z_{t+1})$. Upon simplification, the backward recursion is obtained as follows.

$$
\begin{aligned}
\beta(z_t) &= P(O_1, ..., O_{t+1}, z_t) \\
&= \sum_{q_{t+1}} P(O_{t+1}, ..., O_T, z_{t+1} | z_t) \\
&= \sum_{z_{t+1}} P(O_{t+1}, ..., O_T | z_{t+1}, z_t) P(z_{t+1} | z_t) \\
&= \sum_{z_{t+1}} P(O_{t+2}, ..., O_T | z_{t+1}) P(O_{t+1} | z_{t+1}) P(z_{t+1} | z_t)
\end{aligned}
$$

$$
= \sum_{z_{t+1}} \beta(z_{t+1}) a_{z_t, z_{t+1}} P(O_{t+1} | z_{t+1})
$$

Equation (3.21)

For the alpha recursion, the definition of alpha at the initial step yields

$$
\begin{aligned}
\alpha(z_1) &= P(O_1, z_1) \\
&= P(O_1 | z_1) P(z_1) \\
&= P(O_1 | z_1) \pi_{z_1}
\end{aligned}
$$

Equation (3.22)

### 3.1.3 Computation of $P(O \mid \mu)$ using Forward-Backward Algorithm

Let $\alpha(z_t^i)$ be the joint probability of partial observation sequence $\{O_1, O_2, ..., O_t\}$ at state $z_t = Z_i$ where $1 \leq t \leq T$ is specified as

$$
\alpha(z_t^i) = P(O_1, O_2, ..., O_t, z_t = Z_i | \mu)
$$

Equation (3.23)

Multiplying Equation (3.23) by $a_{ij}$ where $a_{ij}$ is the transition probability from state i to state j and counts for probability of joint event that partial observation sequence exists and state $Z_i$ at time point t is changed to $Z_j$ at time point $t+1$. Upon simplification via multiplication rule and Markov property, we obtain

$$\begin{aligned}
\alpha(z_t^i) &= P(O_1,...,O_t,z_t=Z_i|\mu)P(z_{t+1}=Z_j|z_t=Z_i)\\
&= P(O_1,...,O_t|z_t=Z_i)P(z_t=Z_i)P(z_{t+1}=Z_j|z_t=Z_i)\\
&= P(O_1,...,O_t|z_t=Z_i)P(z_{t+1}=Z_j)P(z_t=Z_i)\\
&= P(O_1,O_2,...,O_t,z_{t+1}=Z_j|z_t=Z_i)P(z_t=Z_i)\\
&= P(O_1,O_2,...,O_t,z_t=z_i,z_{t+1}=Z_j)(Markov \quad property)
\end{aligned}$$

Equation (3.24)

Summing product over all n possible states of $z_t$ produces probability of joint event that the partial observation sequence exists and the next state is $z_{t+1}=Z_j$ regardless of the state $z_t$. By summing product we obtain

$$\begin{aligned}
\sum_{i=1}^{n}\alpha(z_t^i a_{ij}) &= \sum_{i=1}^{n}P(O_1,O_2,...,O_t,z_t=Z_i,z_{t+1}=Z_j)\\
&= P(O_1,O_2,...,O_t,z_{t+1}=Z_j)
\end{aligned}$$

Equation (3.25)

The forward variable at time t+1 and state $Z_j$ is calculated as follows using the multiplication rule

$$\begin{aligned}
(z_{t+1}^i) &= P(O_1,O_2,...,O_t,O_{t+1},z_{t+1}=Z_j|\mu) \qquad (\\
&= P(O_{t+1}|O_1,O_2,...,O_t,q_{t+1}=Z_j)P(O_1,O_2,...,O_t,z_{t+1}=Z_j)\\
&= P(O_{t+1}|z_{t+1}=Z_j)P(O_1,O_2,...,O_t,z_{t+1}=Z_j)
\end{aligned}$$

$$= b_j(O_{t+1})\sum_{i=1}^{n}\alpha(z_t^i)a_{ij}$$

Equation (3.26)

where $b_j(O_{t+1})$ is the probability of an observation $O_{t+1}$ when the markov state is in state $Z_j$.

Using the Forward recurrence Equation in (3.26), we obtain the observation sequence $O = (O_1, O_2, ..., O_T)$ of the Forward variable as

$$\alpha_T(z_t^i) = P(O_1, O_2, ..., O_T, z_T = Z_i|\mu)$$

Equation (3.27)

The probability $P(O|\mu)$ is sum of $\alpha_T(z_t^i)$ over all n possible states of $z_T$ specified by

$$\begin{aligned} P(O|\mu) &= P(O_1, O_2, ...O_T) \\ &= \sum_{n}^{i=1} P(O_1, O_2, ..., O_T, z_T = Z_i|\mu) \\ &= \sum_{i=1}^{n} \alpha_T(i) \end{aligned}$$

Equation (3.28)

Let $\beta(z_t^i)$ be the Backward variable which is a conditional probability of partial observation sequence $\{O_t, O_{t+1}, ..., O_T\}$ given state $z_t = Z_i$ where $1 \leq t \leq T$ specified as $\beta(z_t^i) = P(O_{t+1}, O_{t+2}, ..., O_T|z_t = Z_i, \mu)$
multiplying the transition probability $a_{ij}$ and $b_j(O_{t+1})$ the probability of the observation sequence $O_{t+1}$ when the Markov is in state $Z_j$ together with the Backward variable $\beta(z_{t+1}^j)$ at time point $t + 1$ we obtain

$$
\begin{aligned}
a_{ij}b_j(O_{t+1})\beta(z_{t+1}^{j}) &= P(z_{t+1}=Z_j|z_t=Z_i)\times P(O_{t+1}|z_{t+1}=Z_j)\\
&\times\ P(O_{t+2},O_{t+3},....,O_T|z_{t+1}=Z_j,\mu)
\end{aligned}
$$

Equation

(3.29)

because observation $(O_{t+2},O_{t+3},....,O_T)$ are mutually independent, we have

$$
\begin{aligned}
a_{ij}b_j(O_{t+1})\beta(z_{t+1}^{j}) &= P(z_{t+1}=Z_j|z_t=Z_i)\\
&\times\ P(O_{t+1},O_{t+2},...,O_T|z_{t+1}=Z_j\mu)
\end{aligned}
$$

Equat

ion (3.30)

because partial observation sequence $(O_{t+2},O_{t+3},....,O_T$ is independent from state $z_t$ at time point t, we have

$$
\begin{aligned}
a_{ij}b_j(O_{t+1})\beta(z_{t+1}^{j}) &= P(O_{t+1},O_{t+2},...,O_T|z_{t+1}=z_j,\mu)\\
&= P(z_{t+1}=Z_j|z_t=z_i)\\
&\times\ P(O_{t+1},O_{t+2},...,O_T|z_t=Z_i,z_{t+1}=Z_j,\mu)
\end{aligned}
$$

Equation (3.31)

due to multiplication rule, we have

$$
a_{ij}b_j(O_{t+1})\beta(z_{t+1}^{j}) = P(O_{t+1},O_{t+2},...,O_T,z_{t+1}=Z_j|z_t=Z_i,\mu)
$$

Summing the product $a_{ij}b_j(O_{t+1})\beta(z_{t+1}^{j})$ over all n possible states of $z_{t+1}=Z_j$ and applying total probability rule, we have

$$
\begin{aligned}
\sum_{j=1}^{n}a_{ij}b_j(O_{t+1})\beta(z_{t+1}^{i}) &= \sum_{j=1}^{n}P(O_{t+1},O_{t+2},...,O_T,z_{t+1}=z_j|z_t=Z_i,\mu)\\
&= P(O_{t+1},O_{t+2},...,O_T|z_t=Z_i,\mu)\\
&= \beta(z_t^{i})
\end{aligned}
$$

therefore the Backward recurrence equation is specified as

$$\beta(z_t^i) = \sum_{j=1}^{n} a_{ij} b_j(O_{t+1}) \beta(z_{t+1}^j)$$

Equation (3.32)

where $b_j(O_{t+1})$ is the probability of observation $O_{t+1}$ when the Markov state is in state $Z_j$

## RESULTS AND DISCUSSION

The data used in the study was part of records of patients (students) with different type of malaria severity disease who visited the health service in the time period of 1st January 2015- 20th December, 2015 and displayed different types of malaria related symptoms. Thirteen different malaria symptoms observed by the health officer for each individual student was recorded in students file. The recorded symptoms are; temperature, blood pressure, pulse beat/min, prostration, vomiting, rigors, diarrhea, pallor, convulsions, jaundice, coma, sweating and dark urine. Depending on the number of each symptoms an individual has, the observed symptoms are the recorded on an ordinal scale from 0 to 3 with 0 being no symptoms and 3 being maximum symptoms.

### Analyzing the Hidden Markov Model

To estimate observation matrix, we used *hmmestimate* which corresponds to sequence of states that the model went through to generate sequence (seq). The command takes the emission (observation sequences), seq and states, and returns estimates of the transition and emission matrix. With the help of initial probability distribution we were able to estimate the transition matrix and observation matrix using the *hmmtrain* command in MATLAB. Estimation of transition probability matrix and observation probability matrix was computed using MATLAB. The optimal value obtained for the transition probability matrix A between states of malaria disease was as shown below;

$$A = \begin{array}{c} \\ Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{array} \begin{array}{c} \begin{array}{cccc} \quad Z_1 & \quad\; Z_2 & \quad\; Z_3 & \quad\; Z_4 \end{array} \\ \begin{bmatrix} 0.0034 & 0.9966 & 0.0000 & 0.0000 \\ 0.1297 & 0.1511 & 0.7192 & 0,0000 \\ 0.0002 & 0.0645 & 0.4235 & 0.5118 \\ 0.0000 & 0.0542 & 0.1192 & 0.8266 \end{bmatrix} \end{array}$$

The results of matrix A shows that the probability of an individual remaining in infectious state after displaying malaria related symptoms is 83% ($a_{44} = 0.8266$) in case of severe malaria, 42% ($a_{33} = 0.4235$) the case of moderate malaria, 15% ($a_{22} = 0.1511$) the case of mild malaria. The results also shows that the individual remaining in healthy state after displaying malaria related symptoms is 0.3% ($a_{11} = 0.0034$). The probability of an individual transiting from healthy state to mild state of the disease is 99% ($a_{12} = 0.9966$), from mild state to moderate state is 72% ($a_{23} = 0.7192$) and from moderate state to severe state is 51% ($a_{34} = 0.5118$). The results also shows that there is 0% ($a_{41} = 0.0000$) transition from severe state of the disease to healthy state of the malaria disease.

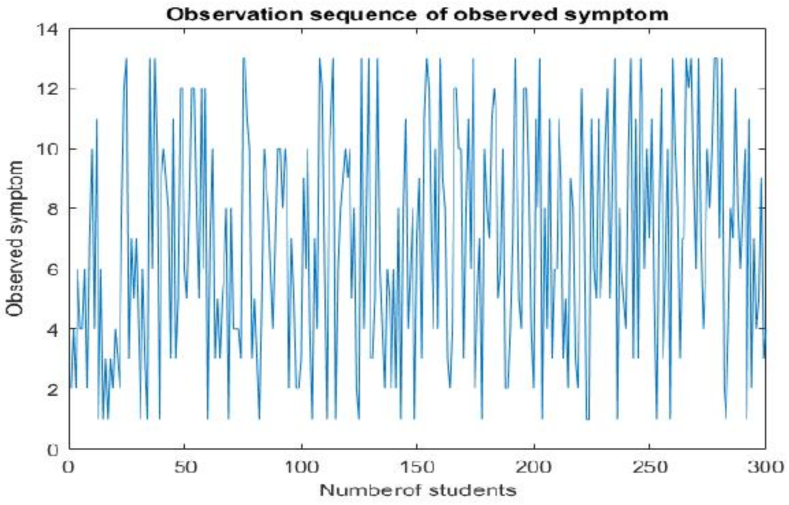

Figure 1: Observation sequence

The results in Figure 5 shows the observation sequence of symptoms as displayed by different students who visited the health facility. From the results it is clear that each student had more than two symptoms with majority having between 4 to 8 malaria symptoms.

$$B = \begin{array}{cccccccccccccc} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} & x_{13} \\ Z_1 & 0.1251 & 0.0723 & 0.0702 & 0.7033 & 0.0291 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ Z_2 & 0.0260 & 0.0649 & 0.0909 & 0.1299 & 0.0260 & 0.1034 & 0.0390 & 0.1039 & 0.1169 & 0.0779 & 0.0779 & 0.0390 & 0.1039 \\ Z_3 & 0.0704 & 0.1549 & 0.1268 & 0.0845 & 0.0563 & 0.0563 & 0.0704 & 0.1268 & 0.0282 & 0.0563 & 0.0704 & 0.0704 & 0.0282 \\ Z_4 & 0.0286 & 0.0714 & 0.0857 & 0.1143 & 0.0429 & 0.1286 & 0.0714 & 0.0429 & 0.1143 & 0.0429 & 0.0571 & 0.1429 & 0.0571 \end{array}$$

The symptoms used in the study were; $x_1$ - body temperature, $x_2$ - blood pressure, $x_3$ - pulse beat/min, $x_4$ - sweating, $x_5$ - rigors, $x_6$ - vomiting, $x_7$- pallor, $x_8$ - diarrhea, $x_9$ - prostration, $x_{10}$- dark urine, $x_{11}$- coma, $x_{12}$- convulsions and $x_{13}$- jaundice. The observation matrix B shows the emission (observation) probabilities and their relationship with the hidden state (status of the individual) of the model as provided by symptom dataset. Each emission probability represent the chance of a particular observation, for instance, the results shows that there is 70% $x_4$ chance of observing sweat in healthy state than in coma 0%.To compute the most likely state to be observed after displaying the malaria related symptoms, we used the function *hmmviterbi* in MATLAB which uses the Viterbi algorithm to compute the most likely sequence of state that the model would go through to generate the given sequence of observation using the function *likelystates = hmmviterbi (seq,A,B).* The results shows that the most likely state sequence is 2. To test the accuracy of hmmviterbi, we compute the percentage of the time that the actual sequence states agrees with the sequence of observation by writing the function. *Sum (states==likelystates/300.* The results obtained from running this function in MATLAB is 0.8467 which shows that the most likely sequence of states agrees with the actual sequence by 85%. The posterior state probabilities of an emission (observation) sequence are the conditional probabilities that the model is in a particular state when it generates a particular sequence. To compute the posterior state probabilities, we use the function,
*PSTATES =hmmdecode (seq, A_{EST}, B_{EST})*
The output *PSTATES* is an n by T matrix, where n is the number of states and T is the length of sequence (seq). PSTATES (i, j) is the conditional probability that the model is in state *i* when it generates the $jth$ symbol of sequence. The actual probability of a sequence tends to 0 as the length of the sequence increases, therefore we use the function *hmmdecode* which gives the logarithm of the probability.
*[PSTATES, logpseq] = hmmdecode ([13], A_{EST}, B_{EST});*
exp(logpseq). The results of *PSTATES* is -2.5649 and its probability is 0.0769 which is the logarithm of the probability. The forward

algorithm evaluates how well the model predicts the given observation sequence. Using the MATLAB function, the results shows that the model is able to predict 87% (0.8719) of the observation sequence

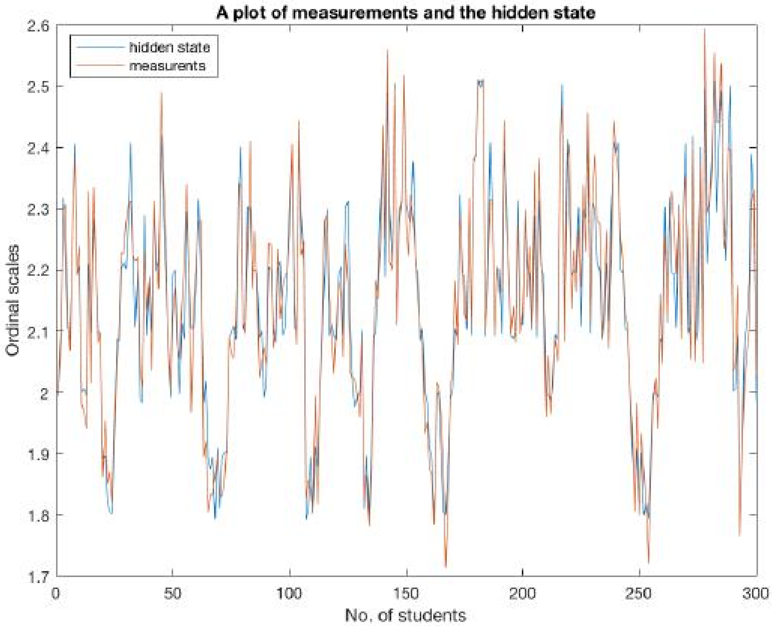

Figure 2: A plot of ordinal scale measurement and Hidden state

The results shows that the measurement are not far much removed from the hidden states i.e. measurements have small error or noise indicating high precision.

The results of Figure 3 shows that the filter estimates are close to the true values and the error is small, this is also true for sequential Monte Carlo (Bootsrap filter particle)

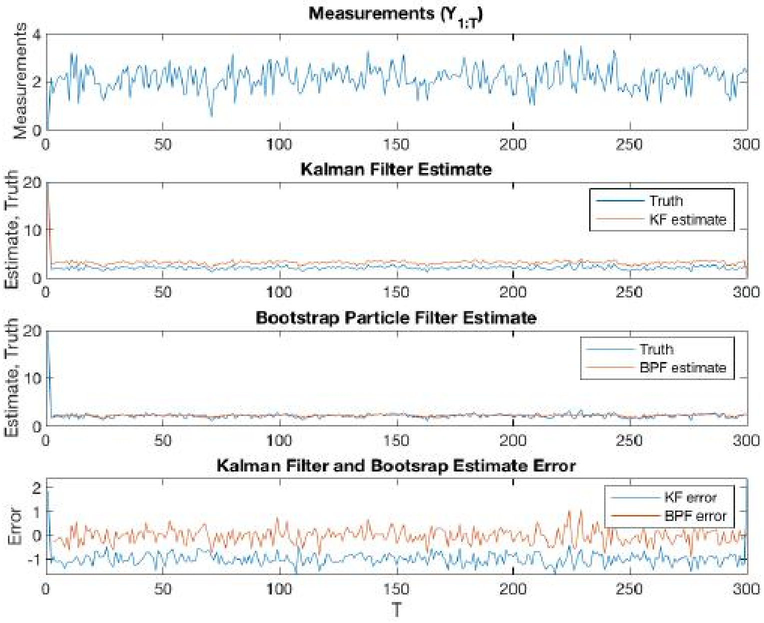

Figure 3: A plot of Kalman Filter estimates, Bootsrap Particle Filter estimates

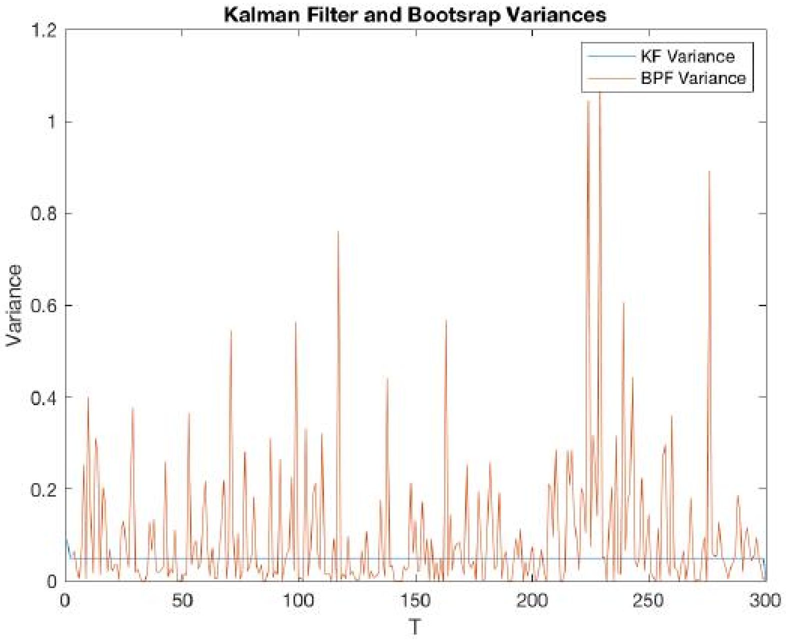

Figure 4: A plot of Kalman Filter and Monte Carlo variances

The results shows that the Kalman Filter produces an optimal estimate which is shown by the lowest variance of the two filters. The more the number of samples in the Monte Carlo filter, the lower the variance.

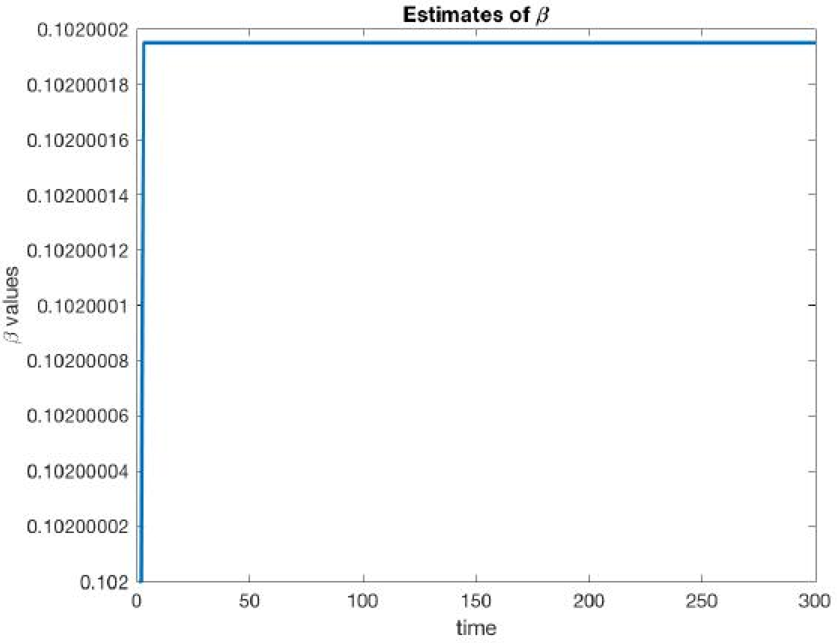

Figure 5: A plot of estimate of beta

The results shows that the filter estimates for parameter beta is 0.1020. The actual value of beta is 0.1 but the estimate value is 0.1020 which shows that the filter estimates is not far much removed from the actual value. Therefore using measurements only, we can obtain estimate of beta which can then be used in the model equation to provide a fit for the data and any other data.

## Conclusion

In the proposed HMM, health states of an individual are modeled by state transition probability matrix and observation probability matrix. To facilitate computation of HMM, forward and backward variables are defined and correspondingly re-estimation methods applied in estimating the model parameters. From the results in chapter 5, the results shows that there is a high probability of 83% of an individual remaining infectious after displaying malaria related symptoms and a low probability of 0.03% of an individual remaining healthy after displaying malaria related symptoms. The results also shows that an individual can remain infectious for a period of one day in case of mild state of the disease before progressing to the next state.

## References

[1] Abeku TA (2006) Malaria Epidemics in Africa: Prediction, Detection and Response. PhD thesis, Erasmus University, Rotterdam

[2] Ader H.J.(2008). Modeling. Advising on Research Methods: A consultants companion. Johannes van Kessel publishing.

[3] Anderson RM, May RM (1991). Infectious diseases of humans: dynamics and control. London: Oxford University Press

[4] Andrew S. Azman, James H. Stark, Benjamin M. Althouse, Charles J. Vukotich Jr. Samuel Stebbins, Donald S. Burke, Derek A.T. Cummings, (2013). Household transmission of in uenza A and B in a school-based study of non-pharmaceutical interventions. Epidemics (5) 181-186.

[5] Barber C., Bockhorst J, Roebber P. (2010). Auto-regressive HMM inference with incomplete data for short-horizon wind forecasting.Adv.Neural Inf Pro-cess Syst.

[6] Bartolucci F. , A. Farcomeni, and F. Pennoni (2012). Latent Markov Models for Longitudinal Data. Chapman and Hall/CRC Press, Boca Raton, FL.

[7] Borooah, Vani K. Logit and Probit (2002). Ordered and Multinomial Models. Thousand Oaks,CA: Sage Publications, Inc.

[8] Broet, P. and Richardson, S. (2006). Detection of gene copy number changes in CGH using a spatially correlated mixture model Bioinfor-matics, 22, 8:911-918.

[9] Chiyaka C., Tchuenche J.M., Garira W. and Dube S.(2008), A Mathematical Analysis of the effects of control strategies on the transmission dynamics of malaria Applied Mathematics and Computation, 641-662.

[10] Cholewa M and Gomb P.(2013), Estimation of the number of states for gesture recognition with Hidden Markov Models based on the number of critical points in time sequence. Recognition Letters. 34(5):574-9.

[11] Dietz K, L. Molineaux, and A. Thomas, (1974). A malaria model tested in the African savannah.Bull World Health Org, 50: 347357.

[12] Dondorp AM and Day NP (2007). The treatment of severe malaria. Trans R Soc Trop Med Hyg, 2007, vol. 101 (pg. 633-634)

[18] Drebel, T., Kueil, B.G. and Meyrowitsch, D.W. (2013) Prevalence of Malaria and Use of Malaria Risk Reduction Measures among Resettled Pregnant Women in South Sudan. International Health, 5, 211-216.

[19] Farsi H, Saleh R. (2014), Implementation and optimization of a speech recognition system based on hidden Markov model using genetic algorithm. Intelligent Systems (ICIS), 2014 Iranian Conference on; IEEE.

[20] Ferguson J.D (1980). Variable Duration Models for speech in Proceeding of the symposium on the application of HMM to Text and Speech ed, J.D Ferguson, Princeton NJ, pp 143-179

[21] Fraser C, Riley S, Anderson R M. (2004). Factors that make and infectious disease controllable. Proc Nat Acad Sci, 101,61466151.

[22] Gemperli A (2003). Development of spatial methods for modeling point-referenced spatial data in malaria epidemiology. PhD thesis, University of Basel, Switzerland

[23] Giacomo Giampieri, Mark Davis, and Martin Crowder (2005). Analysis of default data using hidden markov models. Quantitative Finance, 5(1):27-34.

[24] Gillaizeau Florence , Etienne Dantan, Magali Giral, and Yohann Foucher, (2015). A multistate Additive relative survival semi-markov model. Statistical methods in medical research