PRINT ISSN 1119-8362
Electronic ISSN 2659-1502

Full-text Available Online at
https://www.ajol.info/index.php/jasem
http://www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 27 (2) 277-282 February 2023

A Note on Hidden Markov Models with Application to Criminal Intelligence

*[1]OSATOHANMWEN, P; [2]OMOTAYO-TOMO, MS; [3]OYEGUE, FO; [4]MAZONA, V; [3]EWERE, F; [5]BILESANMI, A; [3]OSAWE, NL

*[1]*School of Science and Technology, Pan-Atlantic University, Ibeju-Lekki, Lagos, Nigeria*
[2]*Department of Mathematical Sciences, Olusegun Agagu University of Science and Technology, Okitipupa, Ondo State, Nigeria*
[3]*Department of Statistics, University of Benin, Benin City, Edo State, Nigeria*
[4]*Department of Statistics, Federal University of Technology Akure, Ondo State, Nigeria*
[5]*Department of General Studies, Petroleum Training Institute, Effurun, Delta State, Nieria*

**Corresponding Author Email: profpato2014@gmail.com*
Co-Authors Email: tomomodupe309@gmail.com; francis.oyegue@uniben.edu; vmmazona@gmail.com; friday.ewere@uniben.edu; *bilesanmi_ao@pti.edu.ng; nosakhare.osawe@uniben.edu.ng*

ABSTRACT: Hidden Markov Models (HMMs) which fall under the class of latent variable models have received widespread attention in many fields of applications. HMMs were initially developed and applied within the context of speech recognition. The theoretical framework underpinning the formalism of HMMs has also evolved over time and has found an exalted place in the theory of stochastic processes. The three problems HMMs are used to resolve were discussed alongside their solutions in this paper. An application to criminal intelligence in unraveling the culprit in a situation involving theft was also carried out and results obtained indicated that the HMMs approach offered a similar result with that of the well-established Dynamic Programming approach.

DOI: **https://dx.doi.org/10.4314/jasem.v27i2.13**

Cite this paper as: OSATOHANMWEN, P; OMOTAYO-TOMO, M.S; OYEGUE, F.O; MAZONA, V; EWERE, F; BILESANMI, A; OSAWE, N.L**.** (2023). A Note on Hidden Markov Models with Application to Criminal Intelligence. *J. Appl. Sci. Environ. Manage*. 27 (2) 277-282

Dates: Received: 19 May 2022; Revised: 01 February 2023; Accepted: 14 February 2023
Published: 28 February 2023

Keywords: Markov model; hidden Markov model; transition probabilities; emission probabilities

Several models exist in the literature for characterizing the behavior of many real life processes and they can be dichotomized into deterministic and stochastic models. Deterministic models rely on some known specific properties of the observed process and hence it is straight forward to implement. On the other hand, stochastic models try to grasp the statistical properties of the process and as such they are more complex to deal with than deterministic models. Examples of such stochastic models are Markov models. Markov models have found great relevance in applications especially when the states of a process are fully observable but the existence of fully observable states is a very simplistic version of reality and in many complex situations, real life processes have one or more hidden states which can only be observed through another stochastic process which is fully observable. These kinds of processes with hidden states can best be modeled by a class of models called *latent-variable models* which include Hidden Markov Models (HMMs). The original works carried out on finite state space Markov models and works carried out on the theory of Gaussian linear state – space models, date back to the 1960s (Baum and Petrie, 1966; Baum and Egon, 1967; Baum and Sell, 1968). Since then, the successful application of these models in several distinct domains has spurred an ever-increasing interest in HMMs and in some other models derived from the principles of HMMs. HMMs have been applied to so many aspects of real life problems and

**Corresponding Author Email: profpato2014@gmail.com*

more profoundly in speech recognition (see Baker, 1975; Rabiner, 1989; Juang and Rabiner, 1991). A HMM is a Markov chain observed in noise. The model compose of a Markov chain denoted $\{X_k\}_{k\geq 0}$ where k is an integer index. The Markov chain $\{X_k\}_{k\geq 0}$ is often assumed to take values in a finite set, even though this restriction is not general and hence arbitrary number of states can be allowed. Furthermore, the Markov chain $\{X_k\}_{k\geq 0}$ is hidden and not observable. What is available to the observer is another stochastic process $\{Y_k\}_{k\geq 0}$ connected to the Markov chain $\{X_k\}_{k\geq 0}$ such that X_k controls the distribution of the corresponding Y_k. In fact, all statistical inference, even on the Markov chain $\{X_k\}_{k\geq 0}$ itself, has to be done in terms of $\{Y_k\}_{k\geq 0}$ only, since $\{X_k\}_{k\geq 0}$ is not observed. Again, it is also assumed that X_k is the only variable affecting the distribution of Y_k. Thus a HMM is a bivariate discrete-time stochastic process $\{X_k, Y_k\}_{k\geq 0}$, where $\{X_k\}$ is a Markov chain and $\{Y_k\}$ is a sequence of independent random variables conditioned on $\{X_k\}$ such that the conditional distribution of Y_k only depends on X_k. Fig 1 demonstrates the process diagrammatically

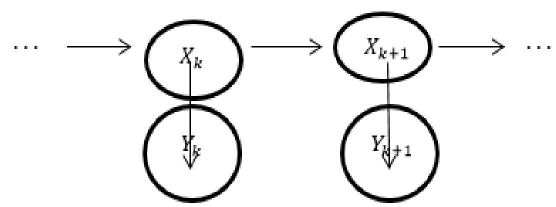

Fig 1: Diagrammatic representation of the dependence structure of a Markov switching model

Fig 1 demonstrates that the distribution of a variable X_{k+1} conditional on the history of the process $X_0, \dots, X_k$ is determined by the value taken by the preceding one, X_k; this is called the *Markov property.* Also, the distribution of Y_k conditionally on the past observations $Y_0, \dots, Y_{k-1}$ and the past values of the state $X_0, \dots, X_k$ is determined by X_k only. Indeed, even though the $Y_{k's}$ are conditionally independent given $\{X_k\}$, $\{Y_k\}$ is not an independent sequence because of the dependence in $\{X_k\}$. In short, $\{Y_k\}$ is not a Markov chain either even though the joint process $\{X_k, Y_k\}$ is a Markov chain because, $\{Y_k\}$ does not possess the *memoryless property* of Markov chains in the sense that the conditional distribution of Y_k given $Y_0, \dots, Y_{k-1}$ generally depends on all the conditional variables. Thus a HMM can be defined by a functional representation known as a latent state-space model

$$X_{k+1} = f(X_k, U_k),$$
$$Y_k = g(X_k, V_k),$$

Where $\{U\}_{k\geq 0}$ and $\{V_k\}_{k\geq 0}$ are mutually i.i.d sequences of random variables that are independent of X_0, and f and g are measurable functions. The first equation is the state equation, while the second equation is the observation equation. The two equations correspond to a recursive, generative form of the model but in this paper, our interest will be based on the specification of the joint probability distribution of the variables. Which approach is most natural and result-yielding, wholly depends on what the HMM is intended to model and for what purpose it is designed for. It is worthy to note that in this discussion, the Markov model is homogenous (the transition kernel does not depend on the time index k) and that the conditional law of Y_k given X_k does not depend on k too. A formal discussion on Markov and HMMs is presented in the next section. The problems and solutions of HMMs are contained in following section. An application of HMMs follows. The paper closes with a conclusion.

Markov and Hidden Markov Models: Consider a $Q-$ valued stochastic process $\{X_k\}_{k\geq 0}$, i.e., each X_k is a $Q-$ valued random variable on a common underlying probability space $(\Omega, \Sigma, \boldsymbol{P})$ where Q is some measure space. We take X_k to be the state of a process at time k and that Q is the *state space* of the process $\{X_k\}_{k\geq 0}$. The process $\{X_k\}_{k\geq 0}$, is said to be a Markov process if

$$\boldsymbol{P}(X_{k+1} \in A | X_0, \dots, X_k) = \boldsymbol{P}(X_{k+1} \in A | X_k) \quad \forall A, k. \quad (1)$$

Thus the process is a Markov process if the future evolution of the process depends only on its present state and not on its past history. Central to the evolution of the states in a Markov process are set of fixed probabilities called *transition probabilities* which are the probabilities of moving from one state of the process to another. To fix ideas mathematically, we will need the concept of a *transition kernel*.

Definition 1: A kernel from a measurable space (Q, ξ_1) to a measurable space (F, ξ_2) is a map $\boldsymbol{P}: Q \times \xi_2 \to \mathbb{R}_+$ such that
i. for every $x \in Q$, the map $A \longmapsto \boldsymbol{P}(x, A)$ is a measure on F; and
ii. for every $A \in \xi_2$, the map $x \longmapsto \boldsymbol{P}(x, A)$ is measurable.

If $P(x, \xi_2) = 1\, \forall x \in Q$, the kernel $\boldsymbol{P}$ is called a transition kernel.

Thus the stochastic process $\{X_k\}_{k\geq 0}$, on the state space (Q, ξ_1) is an homogenous Markov process if there exist a transition kernel $\boldsymbol{P}$ from Q to itself such that

$$\boldsymbol{P}(X_{k+1} \in A | X_0, \dots, X_k) = \boldsymbol{P}(X_k, A) \quad \forall A, k, \quad (2)$$

Where $\boldsymbol{P}(x, A)$ is the probability that the process will be in the set $A \subset Q$ in the next time step, given that it is currently in the state $x \in Q$. The probability measure $\boldsymbol{\pi}$ on Q defined as $\boldsymbol{\pi}(A) = \boldsymbol{P}(X_0 \in A)$ is called the initial measure of $\{X_k\}_{k\geq 0}$.

In distinction to a Markov process is a HMM which is a Markov process decomposed into two components: an observable component and an unobservable or latent component. Thus a HMM is a Markov process $\{X_k, Y_k\}_{k\geq 0}$ on the state space $Q \times \mathcal{O}$, where Y_k is observable and X_k is not. In this case, Q is the state space of the unobserved X_k and $\mathcal{O}$ is the state space of the observed Y_k. It is worthy to note that the process $\{X_k\}_{k\geq 0}$ is a Markov process but $\{Y\}_{k\geq 0}$is not given that it is only a noisy functional of $\{X_k\}_{k\geq 0}$.

Definition 2: A stochastic process $\{X_k, Y_k\}_{k\geq 0}$ on the product state space $(Q \times \mathcal{O}, \xi_1 \otimes \xi_2)$ is an HMM if there exist transition kernels $\boldsymbol{A}: Q \times \xi_1 \to [0,1]$ and $\boldsymbol{B}: Q \times \xi_2 \to [0,1]$ such that

$$\boldsymbol{E}(g(X_{k+1}, Y_{k+1}) | X_0, Y_0, \dots, X_k, Y_k) = \int g(x,y)\boldsymbol{A}(X_k, dx)\boldsymbol{B}(x, dy),$$

and a probability measure $\boldsymbol{\pi}$ on Q such that

$$\boldsymbol{E}\big(g(X_0, Y_0)\big) = \int g(x,y)\boldsymbol{\pi}(dx)\boldsymbol{B}(x, dy),$$

for every bounded measurable function $g: Q \times \mathcal{O} \to \mathbb{R}$. In this situation $\boldsymbol{\pi}$ is called the initial measure, $\boldsymbol{A}$ the transition kernel and $\boldsymbol{B}$ the observation kernel of the hidden Markov model $\{X_k, Y_k\}_{k\geq 0}$.

In non-formal mathematical term, we can describe an HMM as follows: consider a Markov process $\{X_k\}_{k\geq 0}$ with N distinct number of states $= \{q_1, \dots, q_N\}$. The process can transit from one state to another as well as to itself with probabilities a_{ij} such that

$$a_{ij} = P\big(\text{state } q_j \text{at } k+1 | \text{state } q_i \text{at } k\big),$$

and hence the process has a transition matrix $\boldsymbol{A} = \{a_{ij}\}_{N\times N}$. The process $\{X_k\}_{k\geq 0}$ is not observable and can only be observed through another non-Markov process $\{Y_k\}_{k\geq 0}$ with observations sequence, $\mathcal{O} = \{\mathcal{O}_1, \dots, \mathcal{O}_K\}$ and a set of possible observation $V = \{1, \dots, M\}$, where K and M are the length of the observation sequence and number of observation symbol respectively. Suppose the probabilistic relationship between the observable Y_k and the unobservable X_k is given by $b_j(t)$ where

$$b_j(t) = P\big(\text{observation } t \text{ at } k | \text{state } q_j \text{at } k\big),$$

and hence the process has an emission probability matrix $\boldsymbol{B} = \{b_j(t)\}_{N\times M}$. Understand that the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are both row stochastic and the probabilities $b_j(t)$ is independent of k. The Markov process $\{X_k\}_{k\geq 0}$ also possess an initial state distribution π_i such that $\boldsymbol{\pi} = \{\pi_i\}_{1\times N}$ is a row stochastic set of initial probabilities of the states of the Markov process$\{X_k\}_{k\geq 0}$. It follows that an HMM is defined by $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{\pi}$ (and, implicitly by the dimensions N and M). The HMM is denoted by

$$\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}).$$

To fix ideas, consider a state sequence of length four

$$X = (x_1, x_2, x_3, x_4)$$

With corresponding observations

$$\mathcal{O} = (\mathcal{O}_1, \mathcal{O}_2, \mathcal{O}_3, \mathcal{O}_4).$$

The probability of the state sequence $P(X, \mathcal{O})$ is given by

$$P(X, \mathcal{O}) = \pi_{x_1} b_{x_1}(\mathcal{O}_1) a_{x_1,x_2} b_{x_2}(\mathcal{O}_2) a_{x_2,x_3} b_{x_3} \times (\mathcal{O}_3) a_{x_3,x_4} b_{x_4}(\mathcal{O}_4), \quad (3)$$

Where $\pi_{x_1}$is the probability of starting in state x_1, $b_{x_1}(\mathcal{O}_1)$ is the probability of initially observing $\mathcal{O}_1$ and $a_{x_1,x_2}$is the probability of transiting from state x_1 to x_2. The process continues in that manner accordingly. Thus we can compute the probability of each possible state sequence of length four, given the observation sequence.

Problems and Solutions of Hidden Markov Models: Three fundamental problems exist that HMMs are used to address. Here we present these problems and their attendant solutions.

Problem one and solution: The first problem centers on given a model $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ and a sequence of observation $\mathcal{O}$, how do we find the probability of the observed sequence given the model i.e. to determine $P(\mathcal{O}|\lambda)$. This problem is called the *evaluation problem.*

To find a solution to the problem above, suppose $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ is a model and let $\mathcal{O} = (\mathcal{O}_1, \dots, \mathcal{O}_K)$ be a series of observations. Our goal is to find $P(\mathcal{O}|\lambda)$. Let $X = (x_1, \dots, x_K)$ be a state sequence. Then by the definition of $\boldsymbol{B}$ one realizes

$$P(\mathcal{O}|X,\lambda) = b_{x_1}(\mathcal{O}_1)b_{x_2}(\mathcal{O}_2)\dots b_{x_K}(\mathcal{O}_K)$$

and by the definition of $\boldsymbol{A}$ and $\boldsymbol{\pi}$ it holds that

$$P(X|\lambda) = \pi_{x_1}a_{x_1,x_2}a_{x_2,x_3}\dots a_{x_{K-1},x_K}.$$

Also,

$$P(\mathcal{O},X|\lambda) = \frac{P(\mathcal{O}\cap X\cap\lambda)}{P(\lambda)}$$

And

$$P(\mathcal{O}|X,\lambda)P(X|\lambda) = \frac{P(\mathcal{O}\cap X\cap\lambda)}{P(X\cap\lambda)}.\frac{P(X\cap\lambda)}{P(\lambda)} = \frac{P(\mathcal{O}\cap X\cap\lambda)}{P(\lambda)}.$$

It follows that

$$P(\mathcal{O},X|\lambda) = P(\mathcal{O}|X,\lambda)P(X|\lambda).$$

Summing over all possible state sequences gives

$$P(\mathcal{O}|\lambda) = \sum_X P(\mathcal{O},X|\lambda) = \sum_X P(\mathcal{O}|X,\lambda)P(X|\lambda)$$
$$= \sum_X \pi_{x_1}b_{x_1}(\mathcal{O}_1)a_{x_1,x_2}b_{x_2}(\mathcal{O}_2)\dots a_{x_{K-1},x_K}b_{x_K}(\mathcal{O}_K). \quad (4)$$

It is computationally very demanding to evaluate (4) especially when N is very large. In fact, the computation in (4) requires $2KN^K$ multiplications. This computational tediousness can be greatly reduced by using the so-called *forward - backward algorithm* (see Rabiner, 1989 for a description of this algorithm). More so, the forward algorithm only requires about N^2K multiplications in contrast to more than $2KN^K$ multiplications in the naïve approach. However, when N is small, the naïve approach works very efficiently.

Problem two and solution: The second problem of HMMs border on determining the optimal state sequence for the underlying Markov process given a model $\lambda = (\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi})$ and an observation sequence $\mathcal{O}$. In this problem, we seek to uncover the hidden part of the HMM. This problem is also called the *decoding problem.*

The solution to this problem is to find the most likely state sequence. Now, there are different possible interpretations of "most likely". For example, using *dynamic programming*, the optimal state sequence is the sequence with the highest probability. In the HMM sense, the optimal state sequence is obtained by choosing the most probable symbol at each position. The solution due to *dynamic programming* can be different from that due to HMM. The well – known *Viterbi algorithm* is used to solve this problem (see Rabiner, 1989 for a description of the algorithm).

Problem three and solution: Given an observation sequence $\mathcal{O}$ and the dimensions N and M, the third problem of an HMM is to find the model $\lambda = (\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi})$ that maximizes the probability of $\mathcal{O}$. We may consider this as training a model to best fit the observed data. The problem is also called the *learning problem.*

The solution to problem three is determined by adjusting the model parameters to best fit the observations. The *Baum – Welch algorithm* is a very effective algorithm for handling problem three (also see Rabiner, 1989 for details).

Application: Here we shall apply the theory of HMM to a situation involving a theft in a bank, and use it to unfold who the culprit was.

Problem: A certain bank hired two security men ($S1$and $S2$) to be in charge of keeping watch over the belongings of customers which are prohibited from being taken into the floor of the bank by the customers. These belongings are usually one of a handbag, suitcase and a backpack. The two security men ($S1$and $S2$) work on shift in the sense that $S1$begins a day's operation 2 out of the 5 working days of the week, implying that $S2$ begin a day's operation 3 days out of the 5 working days. When on duty, $S1$ and $S2$ work for equal number of hours per day. Of every 100 items that customers bring to the bank $S1$ on average handles 10 handbags, 60 suitcases and 30 backpacks. Likewise, $S2$ handles on average 40 handbags, 25 suitcases and 35backpacks out of every 100 items customers bring to the bank. On an eventful day, a cellphone got missing from the suitcase of one of the customers and the bank was determined to know who stole the cellphone. $S1$ and $S2$ were declared as the only suspects to the theft since they are responsible for handling the customers' belongings and to make matter worse, there was no CCTV camera in the area where customers' items are kept. Furthermore, the CCTV camera footage at the entrance of the bank revealed that prior to the theft, the last four customers (including the one whose cellphone is missing) who came into the bank came in a manner where the first person came with a backpack, the second person came with a suitcase, the third person came with a backpack and the fourth person came with a handbag. Using the above information, how can we determine who the likely culprit is?

Solution: Clearly, an HMM can be used to solve the problem. The system has a finite set of hidden states $Q = \{S1, S2\}$ where the changes from state S1 to S2 are invisible to observers. Again the process is a Markov process because the current state is always dependent on the previous state. Denote a handbag by HB, a suitcase by SC and the Backpack by BP. Footage from the CCTV camera at the entrance of the bank can be taken as the observation sequence. Thus, the observation sequence is given by $\mathcal{O} = \{BP, SC, BP, HB\}$. Letting 1 represent HB, and 2 to represent SC, and allowing 3 to represent BP, we can redefine the observation sequence as $\mathcal{O} = \{3,2,3,1\}$. The goal is to determine the most likely state sequence given the observation sequence. From the problem one can infer that $\boldsymbol{\pi} = [0.4, 0.6]$, which is the initial distribution corresponding to the proportion of time S1 and S2 changes work shift in a week. The transition and emission matrices as inferred from the problem are given by

$$\boldsymbol{A} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},$$
$$\boldsymbol{B} = \begin{bmatrix} 0.1 & 0.6 & 0.3 \\ 0.4 & 0.25 & 0.35 \end{bmatrix}.$$

Given these information, the first task will be to compute the probability of all possible state sequence of length four given the observation sequence $\mathcal{O}$ using the relation in (3). A list of all these possible state sequence and their probabilities are given in Table 1 alongside the corresponding normalized probabilities.

Using the *dynamic programming* approach, the optimal sequence is the most probable sequence and this corresponds to the sequence $S2\,S1\,S2\,S2$ as shown in Table 1, which clearly tells us that $S1$ is the most likely security personnel responsible for the theft since $S1$ is directly in the position of the Suitcase with symbol 2, bearing in mind that the observation sequence is $\mathcal{O} = \{3,2,3,1\}$. In the Hidden Markov sense, the process of determining the optimal state sequence is different from the one based on the *dynamic programming* approach.

To find the optimal state sequence using the Hidden Markov approach, we sum the probabilities of all possible sequence starting with $S1$ in first position; sum all probabilities with $S1$ in the second position; sum all probabilities with $S1$ in the third position and sum all probabilities with $S1$ in the fourth position. This is also done for $S2$. This is carried out using the *forward – backward algorithm.* The result is summarized in Table 2.

To obtain the optimal state sequence, we use the *Viterbi algorithm* to trace the path of the state with the highest probability for a given position number. From table 2, the sequence is $S2\ S1\ S2\ S2$. Interestingly, the Hidden Markov and the dynamic programming approaches are giving the same result in this application. This is not always the case. Indeed the Culprit is likely the security personnel $S1$!

Table 1: State sequence and probabilities

Sequence	Probability	Normalized Probability
$S1\,S1\,S1\,S1$	0.00027	0.02366
$S1\,S1\,S1\,S2$	0.00108	0.09465
$S1\,S1\,S2\,S1$	0.00032	0.02805
$S1\,S1\,S2\,S2$	0.00126	0.11043
$S1\,S2\,S1\,S1$	0.00011	0.00964
$S1\,S2\,S1\,S2$	0.00045	0.03944
$S1\,S2\,S2\,S1$	0.00013	0.01139
$S1\,S2\,S2\,S2$	0.00053	0.04645
$S2\,S1\,S1\,S1$	0.00047	0.04119
$S2\,S1\,S1\,S2$	0.00189	0.16564
$S2\,S1\,S2\,S1$	0.00055	0.04820
$S2\,S1\,S2\,S2$	0.00221	0.19369
$S2\,S2\,S1\,S1$	0.00020	0.01753
$S2\,S2\,S1\,S2$	0.00079	0.06924
$S2\,S2\,S2\,S1$	0.00023	0.02016
$S2\,S2\,S2\,S2$	0.00092	0.08063

Table 2: Hidden Markov Probabilities

	1	2	3	4
$P(S1)$	0.36371	0.70551	0.46099	0.19982
$P(S2)$	0.63629	0.29449	0.53901	0.80018

Conclusion: A study on Hidden Markov Models has been carried out in this paper. A formal mathematical as well as a non-formal mathematical presentation of the theory has been carried out. An application of the theory to unraveling a problem of criminal investigation has been performed. The major results from the application show that both the *dynamic programming* approach and the Hidden Markov approach detected the same optimal sequence of the states and hence suggested the same security personnel as being the most likely to carry out the theft in the facility under investigation.

REFERENCES

Baker, JK. (1975). The dragon system – An overview, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **ASSP-**23: 24 – 29.

Baum, LE; Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov Chains, *Ann Math Stat*, 37: 1554 – 1563.

Baum, L.E; Egon, J.A. (1967). An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology, *Bull* Am Meteorol Soc 73: 360 – 363.

Baum, L.E; Sell, G.R. (1968). Growth functions for transformations on manifold, *Pac J of Mathem*, **27**: 211 – 227.

Juang, B.H; Rabiner, L.R. (1991). Hidden Markov model for speech recognition, *Technometrics*, 33:251 – 271.

Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE*, **77**: 257 – 286.