A Stochastic Model for Rice Yields Forecast

M. Yahaya* and A. Lawal

Department of Mathematics, Federal University of Technology Minna, Nigeria

ABSTRACT

Rice (*Oryza sativa*) is one of most consumed food crops in Nigeria and in around the word. Demand for rice has been increasing at fast rate in Nigeria due to population growth and the banning of its importation into the country has further led to its shortage in supply. This research is poised to provide necessary information to the government to boost rice production in the Nigeria. In view of this, a stochastic mathematical model that makes forecast of rice yields with respect to climatic elements (Rainfall and Temperature) has been developed and implemented in Niger State. The relevant data used in this research were collected from the Niger state Bureau of statistics over a period of 16years (2001-2016). Validity test for the model was conducted which indicates 80% accuracy, this shows that the model is reliable and dependable. Both the validity test and the future forecast of the model showed prevalence of high rice yields, this shows that Niger State is a better place for rice farming. Therefore, results from study is an important information to Federal Government to boost local rice production in country.

Keywords: Hidden Markov Model, Stochastic Model, Rice, Niger State, Baum-Welch Algorithm

INTRODUCTION

The banning of importation of rice into the country coupled with population growth has led to severe shortage in its supply. Hence, this research is aimed at developing a stochastic model for rice yields forecast with a view of providing necessary information to the government to boost local rice production in the country. Rice is the staple food for more than three billion people worldwide. Rice area accounted for approximately 11.5% of the world’s arable land area and it provides almost 19% of the global dietary energy in recent times and its annual average consumption per capita is about 65 kg (Ajith et al., 2017). Therefore, rice area mapping and forecasting is vital for food security, where demands often exceed production due to an ever-increasing population. The aim of yield forecasting, is to give a precise, scientifically sound and independent forecasts of crops’ yield as early as possible during the crops’ growing season by considering the effect of the weather and climate (Basso, 2013). Timely and accurate estimation of rice areas and forecasting can provide important information for governments, planners, and decision makers in formulating policies with regard to import or export and in the event of shortfall or surplus (Mostafa, 2015).

Rice (*Oryza sativa* L.) is a primary food for more than three billion people worldwide (Khush, 2005). It is cultivated on about 11.5% of the world’s arable land during 2012, however more than 88% production is observed in Asian countries. In the recent decades, two major issues like population growth in particular to the rice consuming countries where ~60% of the world’s population lives and climate change put enormous pressure on the rice demand (IPCC, 2014). As such, understanding (in other words, forecasting) the amount of rice production prior to the end of growing season is critical in order to ensure food security (Bastiaanssen & Ali, 2003; Nuarsaet et al., 2012; Huang et al., 2013; Son et al., 2013). This sort of forecasting may help the governments, planners, and decision makers to formulate appropriate policies to: quantify either how much to import in the event of shortfall or optionally to export in case of surplus (Noureldin et al., 2013). Rice is one of the most important crops in terms of human consumption (as opposed to animal feed) and is

Received 28 February, 2020
Accepted 30 June, 2020
Address Correspondence to: yahayamohammed1986@gmail.com,
produced in 95 countries across the world (Maclean et al., 2002). It is the staple food in many countries, accounting for more than 40% of global food production. People in the majority of countries in Asia depend on rice as their main source of nutrition, as well as for income and employment (Maclean et al., 2002; Makino, 2011). Rice has been a major contribution to the financial income of many Asian countries as shown in the statistics, 50% per annum is contributed to countries economy (Fahmiet al., 2013). Henceforth, numerous experiments were carried out to create a good quality paddy and to expand its yield due to increase demand in food. Due to consistent climate change some part of the paddy soil are prone to metal contamination (Sow et al., 2013).

Reported in (Dahikaret al., 2014) is a study of basic requirements for applications of ANNs in yield prediction. Simple network architectures, with one hidden layer and back propagation of errors were tested for different predictors and crops, like cotton, sugarcane, wheat, rice and others. Soil parameters detected to be relevant for crop yield prediction were PH and concentrations of nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper and iron. In terms of atmospheric predictors, temperature, rainfall and humidity were the relevant features detected.

Monisha et al. (2005) developed a model for corn and soybean yield forecasting with climatic aspect by applying artificial neural network. They have considered the rainfall, Maryland corn and soybean yield data and predict the corn and soybean yield at state, regional and local levels by applying both the artificial neural network technique and the multiple linear regression model. Lastly they compared both the techniques and conclude that the ANN model gives more accurate yield prediction than the multiple linear regressions. (Saran et al., 2014) proposed a new approach to estimate rice cultivation and harvest dates by using 8-day composite normalized difference vegetation index (NDVI) derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. In their work, they divided the rice growth state into 4 states, namely, nothing, growing, mature and harvest state and applied Hidden Markov Model (HMM). Then, they assigned the state to the NDVI time-series data by using the Viterbi algorithm. By using those derived state, they were able to estimate the rice cultivation and harvest dates. The date estimation results were compared with the ground truth data to access the accuracy and they found that the average cultivation dates and harvest dates to have errors of 15.48 days and 6.525 days, respectively. They derived a simple linear regression model for wheat yield estimate and forecast based on NDVI images during the wheat grain filling period. They validated their results against official data and found good correlations between the two.

**METHODOLOGY**

**Study Area and Data source**

The data used in this research work, were collected from the Niger State bureau of statistics for the period of 16 years (2001 – 2016). Niger state with a population of 5,556,247 million people (National population commission, 2016) is located in the North central zone along the Middle Belt region of Nigeria. It is classified as one of the largest states in the country, spanning over 86,000 km² in land area with 80% of the land mass conducive for agriculture (Tologbonse, 2008).

**Hidden Markov Model:** A Hidden Markov Model (HMM) is a double stochastic process in which one of the stochastic processes is an underlying Markov chain which is called the hidden part of the model, the other stochastic processes is an observable one. Also a HMM can be considered as a stochastic process whose evolution is governed by an underlying discrete (Markov chain) with a finite number of state which are hidden, i.e. not directly observable (Enza, Daniele, 2007 and Lawal, 2017). Hidden Markov Model is characterized by the following N= number of state in the model M=number of distinct observation symbols per state Q = a state sequence of length T taking values from S.

\[ Q = q_1, q_2, q_3, \ldots, q_T \]  \hspace{1cm} (1)

O = an observation sequence consisting of T observations.

\[ O = o_1, o_2, o_3, \ldots, o_T \]  \hspace{1cm} (2)

A = \( \{ a_{ij} \} \), a transition probability matrix A, where each \( a_{ij} \) represents the
probability of moving from state $i$ to state $j$,

$$\sum_{j=1}^{N} a_{ij} = 1$$

with $B = \{ b_j(o_i) \}$, observation probability matrix

Where

$$b_j(o_i) = p(o_i \mid q_i = s_j)$$

is the probability that the symbol $O_i$ is emitted when the system is in state $S_j$

If the observation is continuous a probability density function is used as follows:

$$\int_{-\infty}^{\infty} b_j(x) dx = 1$$

$$\pi = \{ \pi_j \}$$

An initial probability distribution, where $\pi_j$

Indicates the probability of starting in state $S_j$.

Also,

$$\sum_{i=1}^{N} \pi_i = 1.$$ 

The parameters of hidden of Markov model (HMM) denoted by

$$\lambda = (A, B, \pi)$$

(4)

**Model Formulation:** Hidden Markov Model is used to examine the impact of rainfall and temperature in forecasting rice yield. The quantity of rice yield to a great extent relies upon these climatic components they contribute colossally to rice yield in a growing years. But we can't ordinarily quantify how every one of them contribute to the general yield (the quantity of rice yield). The quantity of rice yield relies upon them and their measure are not static or deterministic but they vary randomly from year to year, which makes the measure of rice yield in each growing year to change.

This circumstance follows a doubly stochastic process, with the measure of rice yield per hectare as the observation of the HMM depending on the state (measure of rainfall and temperature). In perspective on this, we have taken the measure of rice yield per hectare within a growing year as an emission of the HMM while the measure of rainfall and temperature within a similar period is taken as state of our model. We make the following assumptions.

1. The transition between the states is governed by first order Markov reliance as represented by equation (5)

$$P(X_{t+1} = j \mid X_0 = i, X_1 = i_{t-1}, X_t = i_t) = P_{ij}$$

(5)

2. The probability of conveyance of generating current observation symbol depends on current state, as represented by equation (6)

$$P(O \mid Q, \lambda) = \prod_{t=1}^{T} P(o_t \mid q_t, \lambda)$$

(6)

3. Quantity of rainfall is viewed as low in the event that it is underneath 193.65mm

4. Quantity of rainfall is viewed as high in the event that it is above 194.65mm

5. Measure of temperature is viewed as low in the event that it is underneath 27.95$^0$C

6. Measure of temperature is viewed as high in the event that it is above 28.95$^0$C

7. Quantity of rice yield is viewed as low on the off chance that it is underneath 2.986 metric tons

8. Quantity of rice yield is viewed as high in the event that it is above 3.992 metric tons

9. Quantity of rice yield is viewed as moderate in the event that it is in range of (2.986-3.992) metric tons

Following (Lawal, 2017), we have the following states and observations for our model

State 1: low rainfall and low temperature

State 2: low rainfall and high temperature

State 3: high rainfall and low temperature

State 4: high rainfall and high temperature

Observations.

The classification of states and the observations, and the assumption made in this work are based on the study area and the data obtained
Making Prediction with the Model: Making prediction with the model is done with the training of parameters of the model. The parameters of the model are initialized then trained using Baum-Welch algorithm to attend maximum likelihood. The forward probability of the training observation sequence is determined from time $t=1$ to time $T$ using Forward Algorithm (Rabiner, 1989 and Lawal, 2017). To predict the next state at time $T+1$ and its observation given the present state at time $T$, we calculate forward probability for every conceivable observation of the states, at that point, the sequence with most elevated value of the forward probability at time $T+1$ is taken as predicted state and its observation. The prediction is made for the following three years (year 2 at time $T+2$, year 3 at time $T+3$ and year 4 at time $T+4$).

RESULTS AND DISCUSSION

Application of the Hidden Markov Model for Rice Yields Forecast: The data used in this illustration was collected from the archive of Niger State Bureau of Statistics for the period of 16 years (2001-2016). The summary of the data is presented in Table 1.
Validity Test for the Model: To test for the validity of the model, we divide the data set into two sets, one of the sets was used to build the test model (HMM1)and the other set was used to test for the reliability of the model. We estimate the parameters of HMM1 utilizing Rainfall, Temperature and Rice yield data from 2001 to 2012, then test for the reliability of the model using data set for the following years 2013, 2014, 2015 and 2016.

The Transition Count Matrix, Pseudo count Observation matrix and Observation probability matrix are given in equations (10), (11) and (12), respectively.

\[
E = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 4 & 1 \\ 0 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix} \quad (10)
\]

\[
D = \begin{bmatrix} 2 & 2 & 2 \\ 1 & 5 & 2 \\ 1 & 1 & 3 \\ 1 & 2 & 2 \end{bmatrix} \quad (11)
\]

\[
B = \begin{bmatrix} 0.4000 & 0.2000 & 0.2222 \\ 0.2000 & 0.5000 & 0.2222 \\ 0.2000 & 0.1000 & 0.3333 \\ 0.2000 & 0.2000 & 0.2222 \end{bmatrix} \quad (12)
\]

The initial state probability is given below

\[
\pi = [0.25, 0.4167, 0.1666, 0.1666] \quad (13)
\]

The general HMM1 is represented by equation (14)

\[
\lambda = (A, B, \pi) \quad (14)
\]

After 1000 iteration of Baum Welch Algorithm, equation (14) settled to (15)

\[
\lambda^* = \left( ^\wedge A, ^\wedge B, ^\wedge \pi \right) \quad (15)
\]

When

\[
^\wedge A = \begin{bmatrix} 0.0000 & 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.4283 & 0.0000 & 0.5717 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.5717 & 0.4283 \end{bmatrix} \quad (16)
\]

\[
^\wedge B = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \\ 0.0000 & 0.2102 & 0.7898 \end{bmatrix} \quad (17)
\]

Comparison of the Predicted States and Observations, and the Actual States and Observations from the Dataset.

Table 1: The States and Observations of the Hidden Markov Model for Period of sixteen years

<table>
<thead>
<tr>
<th>Years</th>
<th>States</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>2002</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>2003</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>2004</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>2005</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>2006</td>
<td>4</td>
<td>M</td>
</tr>
<tr>
<td>2007</td>
<td>3</td>
<td>H</td>
</tr>
<tr>
<td>2008</td>
<td>3</td>
<td>H</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
<td>H</td>
</tr>
<tr>
<td>2010</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>2011</td>
<td>2</td>
<td>H</td>
</tr>
<tr>
<td>2012</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>2013</td>
<td>2</td>
<td>H</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
<td>H</td>
</tr>
<tr>
<td>2015</td>
<td>3</td>
<td>H</td>
</tr>
<tr>
<td>2016</td>
<td>4</td>
<td>H</td>
</tr>
</tbody>
</table>

\[
C = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \quad (7)
\]

\[
S = \begin{bmatrix} 2 & 2 & 1 & 2 \\ 1 & 4 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 1 & 2 & 2 & 1 \end{bmatrix} \quad (8)
\]

\[
A = \begin{bmatrix} 0.2857 & 0.2857 & 0.1428 & 0.2857 \\ 0.1250 & 0.5000 & 0.1250 & 0.2500 \\ 0.3333 & 0.1666 & 0.3333 & 0.1666 \\ 0.1666 & 0.3333 & 0.3333 & 0.1666 \end{bmatrix} \quad (9)
\]
Predicted States and Observations:
Observations: M H H H H

Actual States and Observations:
Observations: M H H H H

The parameter of the HMM1 were determined utilizing rainfall, temperature and rice yield data from 2001 to 2012. After 1000 iterations of the Baum Welch Algorithm, \( \hat{\lambda}_1 \) settled to another model, \( \hat{\lambda}_2 \), this new model was used to test for rice yield for 2013, 2014, 2015, and 2016. From the test, the HMM1 was in state 2 at time T (2012) emitting Moderate rice yield, at that point, it then make move to state 2 at time T+1 (2013) emitting a Moderate rice yield, then move to state 4 at time T+2(2014) emitting a High rice yield, next, it make move to state 3 at time T+3(2015) emitting a High rice yield and move to state 3 at time T+4 (2016) emitting a High rice yield. The transition between the states are governed by the first order Markov dependence as referenced in the past section. The validity test shows 80% precision in the rice forecast when compared with the actual rice yield from the dataset.

**Hidden Markov Model (HMM2) for future forecast:** HMM2 was developed to forecast rice yield for future years, the parameters of the HMM2 were determined utilizing rainfall, temperature and rice yield data from 2001 to 2016, at the point, we made forecast for 2017, 2018, 2019, and 2020.

Transition Count Matrix

\[
C = \begin{bmatrix}
1 & 1 & 0 & 0 \\
0 & 4 & 0 & 2 \\
1 & 0 & 1 & 1 \\
0 & 1 & 2 & 0
\end{bmatrix}
\]  \hspace{1cm} (18)

Pseudo count Transition Matrix

\[
S = \begin{bmatrix}
2 & 2 & 1 & 1 \\
1 & 5 & 1 & 3 \\
2 & 1 & 2 & 2 \\
1 & 2 & 3 & 1
\end{bmatrix}
\]  \hspace{1cm} (19)

Transition Probability Matrix

\[
A = \begin{bmatrix}
0.2857 & 0.2857 & 0.1428 & 0.2857 \\
0.1000 & 0.5000 & 0.1000 & 0.3000 \\
0.2857 & 0.1428 & 0.2857 & 0.2857 \\
0.1428 & 0.2857 & 0.4285 & 0.1428
\end{bmatrix}
\]  \hspace{1cm} (20)

Observation Count Matrix

\[
C = \begin{bmatrix}
1 & 1 & 1 \\
0 & 4 & 2 \\
0 & 0 & 3 \\
0 & 1 & 3
\end{bmatrix}
\]  \hspace{1cm} (21)

Pseudo count Observation Matrix

\[
S = \begin{bmatrix}
2 & 2 & 2 \\
1 & 5 & 3 \\
1 & 1 & 4 \\
1 & 1 & 4
\end{bmatrix}
\]  \hspace{1cm} (22)

Observation Probability Matrix

\[
B = \begin{bmatrix}
0.4000 & 0.2222 & 0.1535 \\
0.2000 & 0.5555 & 0.2307 \\
0.2000 & 0.1111 & 0.3076 \\
0.2000 & 0.1111 & 0.3076
\end{bmatrix}
\]  \hspace{1cm} (23)

Initial State Probability

\[
\pi = [0.1875, 0.1875, 0.250, 0.375]
\]  \hspace{1cm} (24)

\( \hat{\lambda}_2 = (A, B, \pi) \)  \hspace{1cm} (25)

After 1000 iteration of Baum Welch Algorithm, equation (26) stabilized to (27)

\[
\hat{\lambda}^* = (\hat{A}, \hat{B}, \hat{\pi})
\]  \hspace{1cm} (28)

Where

\[
\hat{A} = \begin{bmatrix}
0.0000 & 1.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.7500 & 0.0000 & 0.2500 \\
0.0000 & 0.0000 & 0.8750 & 0.1250 \\
0.0000 & 0.0000 & 1.0000 & 0.0000
\end{bmatrix}
\]  \hspace{1cm} (27)

\[
\hat{B} = \begin{bmatrix}
1.0000 & 0.0000 & 0.0000 \\
0.0000 & 1.0000 & 0.0000 \\
0.0000 & 0.0000 & 1.0000 \\
0.0000 & 1.0000 & 0.0000
\end{bmatrix}
\]  \hspace{1cm} (28)
Future Forecast States and Observations:


Observations: H H H H H H

The parameter of the HMM2 were determined using rainfall, temperature and rice yield data from 2001 to 2016. After 1000 iteration of the Baum Welch algorithm, \( \lambda_2 \), settled to another model, \( \lambda_2' \), this new model was then used to make a forecast for future years. From the forecast, the HMM2 was in state 4 at time T (2016) emitting High rice yield, at that point, it then make move to state 3 at time T+1 (2017) emitting High rice yield. Similar interpretation is given to move to state 3 at time T+2 (2018), move to state 3 at time T+3 (2019), move to state 3 at time T+4 (2020) all emitting High rice yield.

Conclusion

A hidden Markov Model to forecast Rice yield with respect to major climatic elements (rainfall and temperature) has been developed and implemented in Niger State. The validity test for the model show 80% precision in Rice yield forecast. Both the validity test and the future forecast shows pervasiveness of high Rice yield, this shows that Niger State is a better place for rice farming. The model could likewise be used to forecast Yield of other crops with little or no modification.

REFERENCES


Maclean, J. L., Dawe, D. C., Hardy, B. & Hettel, G. P. (2002). Rice almanac source book for the most important economic activity on earth. Published by CABI (Oxon, UK) for IRRI (Manila) WARDA (Cote d’Ivoire), CIAT (Colombia) and FAO (Rome, Italy).


Oliver K (2009). Department of mathematics, Harvard University Cambridge, MA 02138, USA


