Science and Technology Indonesia
e-ISSN:2580-4391 *p*-ISSN:2580-4405
Vol. 5, No. 2, April 2020

Research Paper

A Hidden Markov Model for Forecasting Rainfall Data Availability at the Weather Station in West Sumatra

Rahmawati Ramadhan[1], Dodi Devianto[1]*

[1] *Department of Mathematics, Andalas University, Limau Manis Campus, Padang 25163, Indonesia*
***Corresponding author**: ddevianto@sci.unand.ac.id

Abstract

Indonesia is a maritime continent in Southeast Asian, laying between Indian Ocean and Pacific Ocean. This position intensely affects the level of rainfall in Indonesia, especially West Sumatra. The availability of rainfall data can form a Markov chain which its state is not able to be observed directly (hidden), is called the Hidden Markov Model (HMM). The purposes of this research are to predict the hidden state of the availability of rainfall data using decoding problems and to find the best state sequence (optimal) by using Viterbi Algorithm, and also to predict probability for the availability of rainfall data in the future by using the Baum Welch Algorithm in the Hidden Markov Model. This research uses secondary data with a period of one day from the availability of rainfall data at the Minangkabau Meteorological Station, Padang Pariaman Climatology Station, and Silaing Bawah Geophysics Station from January 2018 to July 2019. The results of the prediction show that the Hidden Markov Model can be used to predict the probability of rainfall data availability. The results for the availability of the highest rainfall data for one day ahead is at the Padang Pariaman Climatology Station, with a probability of 0.36, followed by Minangkabau Meteorological Station is 0.35, and Silaing Bawah Geophysics station is 0.29. The result has shown for the next one day period the probability of rainfall data available from the three stations will be available following the Viterbi algorithm.

Keywords

Hidden Markov Model, Rainfall, Decoding Problem

Received: 12 February 2020, Accepted: 03 April 2020
https://doi.org/10.26554/sti.2020.5.2.34-40

1. INTRODUCTION

A water particle that falls to the earth's surface through a series of hydrological processes is the occurrence of rain. Meanwhile, rainfall is the height of the rainwater that falls to the earth's surface to a flat within a certain period. Many say the least precipitation in an area influenced by several factors, including factors latitude, altitude, and distance from the source of the water, the wind, the mountain areas, the temperature difference, and the total land area.

In analyzing the availability of rainfall in the future maybe associated with a stochastic process, where the problem is related to the chance of future events that they cannot be predicted directly on the availability of rainfall data. States of availability of rainfall data are uncertain and subject to change, and it is assumed there are some unobserved circumstances; it can be modeled by Hidden Markov Model (HMM).

HMM is a broadening of the Markov chain in which the state cannot be observed directly (hidden), but can only be observed through a set of other observations. In HMM, there are three fundamental problems to be solved that problem evaluation, decoding problems, and learning problems. Based on research by Thyer and Kuczera (2003), they have discussed a calibration model with Bayesian approach on rainfall data by using HMM, the same thing was done by Sansom (1998) for the data breakpoint. Some research that uses inhomogeneous HMM and nonparametric model reduction of rainfall events has been carried out also by Mehrotra and Sharma (2005); Robertson et al. (2004); Greene et al. (2011); Pineda and Willems (2016).

Furthermore, there are some HMM models introduced by previous researchers, the abundance of species in the river with HMM with negative binomial model approach discussed by Spezia et al. (2014). According to research conducted in Xia and Tang (2019); Li et al. (2018); Bathaee and Sheikhzadeh (2016), HMM also associated with Bayesian analysis. Meanwhile, in the case of time series, there is a lot of research that studies the hidden Markov models, one of which discusses the most reliable classification of multivariate time series as described Antonucci et al. (2015), and categories with HMM multiple time series by Colombi and Giordano (2015). Also besides, HMM also is used in the financial case is to predict trends in the time series as discussed in Zhang et al. (2019) and to predict the probability of the changes of the exchange rate as discussed in Ramadhan et al. (2020). Based on the research of Devianto et al. (2015b), the techniques to build models using financial data can be used autoregressive fractionally integrated moving average (ARFIMA). Furthermore, the enumeration models have been developed with an exponential distribution characterization approach described in Devianto (2016); Devianto et al. (2015a,c).

According to research conducted in Stoner and Economou (2019), it is illustrated how the hidden Markov framework could be adapted to construct a compelling model for sub-daily rainfall, which is capable of capturing all of these essential characteristics well. Several homogenous Hidden Markov Models (HMMs) were developed to forecast droughts using the Standardized Precipitation Index, SPI, at short-medium term, as discussed in Khadr (2016). The hidden Markov sequence was assigned to represent the recurrence of mast years, as described in Tseng et al. (2020).

Model availability of rainfall data can be formed into a hidden state HMM with attention. In this study, there are three fundamental problems to be solved that are evaluations problem, decoding problems, and learning problems. The results can provide information about the availability of rainfall data in particular areas of West Sumatra in the future. Historical information about the availability and the accuracy of rainfall data will be very helpful in predicting climate change and irregularities. In addition, by obtaining an overview of future weather conditions, the specific policy considering water supply, plant performance and yield can be estimated for better preventive of resources for community.

2. EXPERIMENTAL SECTION

2.1 Materials

A sequence of events that fulfill the legal requirement of probability, where every value randomly changed against time. A Stochastic process predicted the properties of the future depend on the properties of the current conditions based on their characteristics in the past called a Markov chain (Ross, 1996). Stochastic process $X_n(t)$ is a series of random variables that change with the time of observation tϵT, and a stochastic process X_n is said to have Markov properties if,

$$P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) \quad (1)$$

$$= P(X_{n+1} = j|X_n = i)$$

for time and for every $n = 0, 1, 2, ...$ and for every $j, i, i_{n-1}, ..., i_1, i_0$

HMM is a stochastic model where the system is assumed to be a Markov Process with hidden states. If $X = (X_1, X_2, \dots)$ is a Markov process and $O = (O_1, O_2, \dots)$ is a function of X, then is a Hidden Markov Model which can be observed through O, or can be written to a function f. The parameter X represents the state process that is hidden, while parameter O states an observation room that can be observed. The elements of the Hidden Markov Model are:

1. The number of hidden state elements (hidden state) represented by N as the number of states which the probability of a denoted state space $S = (S_1, S_2, \dots, S_N)$ and the state at the time t denoted by Xt, $t = 1, 2, \dots, T$
2. The number of observations (observation) of each state represented by M, where probability every state represented by $v = v_1, v_2, \dots, v_M$ and space observation represented by $O = (O_1, O_2, \dots, O_T)$, which T is the length of the observation data.
3. The transition probability matrix $A = [aij]$ where aij is an element of A which is the conditional probability of the state at the time , given the state at the time , that is

$$a_{ij} = P(X_{n+1} = j|X_n = i) \quad (2)$$

for $1 \leq i, j \leq N$

4. Observation probability distribution at the time t, at state j, commonly known as emission matrix

$$B = [b_{ik}] \quad (3)$$

where

$$b_{ik} = P(O_t = v_k|X_t = i) \quad (4)$$

for $1 \leq j \leq N$, $1 \leq t \leq T$ and $1 \leq k \leq M$

5. The initial state distribution represented by π(i) where

$$\pi(i) = P(X_1 = i, 1 \leq i \leq N) \quad (5)$$

So HMM can be written in the notation λ=(A, B, π) where A is expressed by the matrix of transition probability, B a chance observation matrix known as emission matrix, and π is the distribution of the initial state. There are three special algorithms that can be solved by HMM method, namely:

a) Evaluation Problem

To calculate the probability of the observation sequence $P(O|\lambda)$ requires forward algorithms and backward algorithms (Bain and Engelhardt, 1992). Steps to resolve with the advanced algorithm are as follows:

i. Initial Step

In this initial step, we determined the initial observation probability α_1(i) which ends at state at the time if it is known a sequence of preliminary observations O_1 is as follows:

$$\alpha(i) = \pi(i)b_i(O_i) \quad (6)$$

for $1 \leq i \leq N$

ii. Induction step

In this induction step, we determined the total observation probability $\alpha_{t+1}(j)$ which end in state i at the time $i = 2, 3, 4, ...$, if known a sequence of observations $O_1, O_2, ..., O_r$ is as follows:

$$\alpha_{t+1}(j) = \{\sum_{i-1}^{N} \alpha_t(i)a_{ij}\}b_j(O_{t+1}) \quad (7)$$

for $j = 1, 2, ...N$, $t = 1, 2, ..., T$

iii. Termination step On termination of this step, we determined the total combined odds of observation and the hidden state when known a model so that it is known probability observation sequence $P(O|\lambda)$, as follows:

$$P(O|\lambda) = \sum_{i-1}^{N} \alpha_T(i) \tag{8}$$

Next, calculate the probability of observation by using a backward algorithm $\beta_t(i)$, namely with as following steps :

i. Initial step In this initial step, the initial observation probabilities otherwise equal to one, this is because it is assumed i is the final state, and zero for the other can be expressed as

$$\beta_t(i) = 1 \tag{9}$$

for $1 \leq i \leq N$

ii. Induction step

In this induction step, we determined the total observation probabilities for $t < 1$, as follows:

$$\beta_t(i) = \sum_{j-1}^{N} a_{ij} b_j(O_{t+1})\beta_{t+1}(j) \tag{10}$$

for $t = T - 1, T - 2, ...1$ and $i = 1, 2, ..., N$

iii. Termination step In this step, the total odds will be determined combination of observation and the hidden state when known a model so that it is known probabilities observation sequence $P(O|\lambda)$, as follows

$$P(O|\lambda) = \sum_{i-1}^{N} b_i(1)\pi(1)\beta_1(i) \tag{11}$$

b) Decoding Problem

Decoding problem, this decoding step is to find the best state sequence (optimal) associated with the observation of O and a model of λ, which has also been known. This problem can be solved by the Viterbi Algorithm. Steps in the Viterbi algorithm to determine the best state sequence are as folows

i. Initial step

In this initial step, it will be determined the greatest probabilities throughout t the first observation and ends in state i to $t = 1$, as follows:

$$\delta_1(j) = b_i(O_1)\pi(i)$$

$$\varphi_t(i) = 0, 1 \leq i \leq N \tag{12}$$

ii. Recursion Step

In this recursion step, it will be determined greatest probabilities throughout t the first observation and ends in state i for $t > 2$, as follows

$$\delta_t(j) = \max_{1\leq i\leq N}[\delta_{t-1}(i), a_{ij}]b_j(O_t)$$

$$\varphi_t(j) = arg \max_{1\leq i\leq N}[\delta_{t-1}(i), a_{ij}] \tag{13}$$

for $2 \leq t \leq T$ and $1 \leq j \leq N$

iii. Termination Step

On termination of this step, it will be determined the greatest probabilities throughout t the first observation and ends in state i, as follow

$$P^+ = \max_{1\leq i\leq N}[\delta_T(i)]$$

$$X_T^+ = arg \max_{1\leq i\leq N}[\delta_T(i)] \tag{14}$$

iv. Backtracking Step

In this last step, it will be determined the best state sequen ce, as follow

$$X_t^+ = \varphi_{t+1}(X_{t+1}^+), t = T - 1, T - 2, ..., 1 \tag{15}$$

c) Learning Problem

This problem estimates the best model to explain a sequence of observations, where the changing parameters of HMM, $\lambda = (A, B, \phi)$ so that $P(O|\lambda)$ becomes the maximum. In the Baum-Welch algorithm, also defined four variables, namely: variable forward (forward), variable backward (backward), variable $\xi_t(i, j)$, and the variable $\gamma_t(i)$. Forward variable and backward variable will be used in the calculation of the variable $\xi_t(i, j)$ and the variable $\gamma_t(i)$, so the estimation formula learning problem is as follows:

$$\hat{\pi}(i) = \gamma_t(i), 1 \leq i \leq N$$

$$\hat{a}_{ij} = \frac{\sum_{t=i}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, 1 \leq i \leq N, 1 \leq j \leq N$$

$$\hat{b}_{ij} = \frac{\sum_{t=1,O=j}^{t=1} \gamma_t(i)}{\sum_{t=1}^{T=1} \gamma_t(i)}, 1 \leq j \leq N, 1 \leq k \leq M \tag{16}$$

where $\bar{\pi}(i)$ is the estimated value of initial state, $\hat{a}_{ij}$ is the matrix of the estimated value of transition probabilities, and $\hat{b}_{ij}$ is the matrix of the estimated value of emission matrix.

2.2 Methods

In this study, the data to be used only calculates rainfall data for 17 months on January 1, 2018, to July 31, 2019, with a period of one day, from the contribution of rainfall data at the Minangkabau Meteorological Station (MM station), Padang Pariaman Climatology Station (PPC station), and Silaing Bawah Geophysics Station (SBG station), it is obtained at the website address dataonline.bmkg.go.id with the amount of data used in one station is 570. Presentation of rainfall data in diagram form can be seen from the Figure 1 as follows

Based on Figure 1, it can be seen that the highest availability of rainfall data is at the MM station at 287 data, followed by the PPC station at 231 data and the SBG station at 219 data. The highest number of unavailable data of rainfall during the period

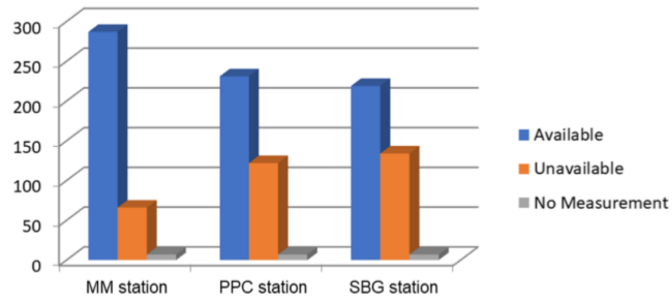

Figure 1. Rainfall data availability

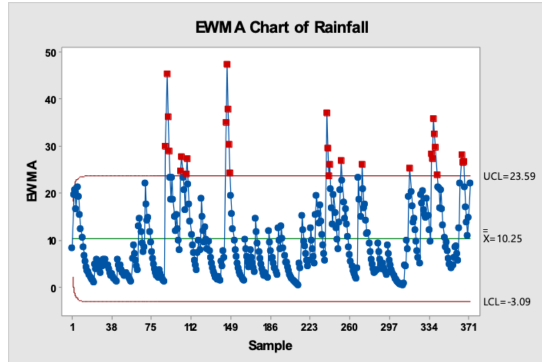

Figure 2. Rainfall EWMA Control Charts

of research is at SBG station with 134 data, followed by PPC station with 122 data, and MM station with 66 data. From the three stations, each has 7 data that were not measured.

This research uses HMM for forecasting the availability of rainfall data. That will be analyzed is rainfall data with predictions of probability rainfall on the next period, as follows:

1. Take rainfall data by a period of one day. The numbers of data have been observed in the range of time of 570 days.
2. Determine the transition probability matrix needed exchange rate by using the following formula as a probability.

$$P(A) = \frac{n(A)}{n(S)} \tag{17}$$

   where $n(A)$ is the number of elements in A and $n(S)$ is the total number of elements in S [22].
3. Determine elements of HMM.
4. Analyze the elements of HMM that have been able to use where the HMM is calculated the probability of availability of an observation by using Forward-Backward algorithm, then it is followed by determining the sequences of hidden state by using the Viterbi algorithm, and it is predicting the HMM parameters by using the Baum-Welch algorithm.
5. Make the interpretations or conclusions from the results that have been obtained.

3. RESULT AND DISCUSSION

3.1 The Importance of Rainfall Data Availability by Controller Graph of Exponentially Weighted Moving Average (EWMA)

In this study, the EWMA control chart is used to see whether there is extreme rainfall or non-controlled data, such that the availability of rainfall data will be very crucial to observe. Rainfall data had been taken from the weather stations in West Sumatera in 2018. The following chart is EWMA control of rainfall data.

Based on Figure 2. This shows that there is extreme rainfall on the day of observation to-88, 89, 90, 91, 102, 103, 107, 108, 144, 145, 146, 147, 148, 238.239, 240, 241, 252, 271, 316, 336, 337, 338, 339, 340, 341, 364, 365, and 366. This means that there are 28 days with rainfall that is not controlled or extreme. This condition have indication there are changes in rainfall throughout the day, where there is a condition of very high rainfall or otherwise. The availability of rainfall data in these circumstances is very important in order to be taken to minimize losses related to changes in precipitation. Therefore, it is necessary to build a model of the availability of rainfall data in every weather station is West Sumatera by using HMM.

3.2 Elements of the Hidden Markov Model

The elements that must be determined to solve the case of forecasting the availability of rainfall data with HMM, as follows:

1. Suppose N, is the number of hidden states, with state space $A = (S_1, S_2, \ldots, S_N)$ and the state at time t is expressed by X_t. In this case of rainfall data availability in MM station, PPC station, and SBG station. The hidden state is available, unavailable, and not measurement. So in this case study $N = 3$ or can be written as $s_1 = P$(available), $s_2 = P$(unavailable), $s_3 = P$ (no measurement). For example , $X_t = 1$, it states that the state is in a state of rainfall data available.
2. Let M, is the number of observations of each state, the observation space $O = (O_1, O_2, \ldots, O_T)$ and the probability every observations expressed by $v = (v_1, v_2, \ldots v_M)$, in this study $M = 3$, the MM station as v_1, PPC station as v_2, and SBG station as v_3.
3. Let

$$A = a_{ij} = P(X_{t+1} = j | X_t = i)$$

   where A is the probability of availability of rainfall data over a range of values to j on day $t+1$ if it is known on day t is in the range value to i thus formed matrix probability are:

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

   a) Transition probability matrix MM station Data

$$A = [a_{ij}] = \begin{bmatrix} 0.79 & 0.19 & 0.02 \\ 0.58 & 0.41 & 0.01 \\ 0.46 & 0.18 & 0.36 \end{bmatrix}$$

b) Transition probability matrix PPC station Data

$$A = [a_{ij}] = \begin{bmatrix} 0.74 & 0.25 & 0.01 \\ 0.46 & 0.52 & 0.02 \\ 0.42 & 0.25 & 0.33 \end{bmatrix}$$

c) Transition probability matrix SBG station Data

$$A = [a_{ij}] = \begin{bmatrix} 0.64 & 0.35 & 0.01 \\ 0.47 & 0.50 & 0.03 \\ 0.33 & 0.42 & 0.25 \end{bmatrix}$$

4. Emission matrix $B = [b_{ik}]$, the conditional probability matrix observation v_k if the process is in state j, matrix emissions observations MM station, PPC station, and SBG station are as follows:

$$A = [a_{ij}] = \begin{bmatrix} 0.74 & 0.64 & 0.56 \\ 0.24 & 0.34 & 0.42 \\ 0.02 & 0.02 & 0.02 \end{bmatrix}$$

5. Suppose $\pi(i)$ is the initial state distribution, in case of the availability of rainfall data are assumed: $\pi(1) = P(\text{available})$, $\pi(2) = P(\text{unavailable})$, $\pi(1) = P(\text{no measurement})$. Initial matrix for MM station, PPC station, and SBG station are as follows:

a) Initial matrix for MM station: $\pi = \begin{bmatrix} 0.70 \\ 0.27 \\ 0.03 \end{bmatrix}$

b) Initial matrix for PPC station: $\pi = \begin{bmatrix} 0.62 \\ 0.36 \\ 0.02 \end{bmatrix}$

c) Initial matrix for SBG station: $\pi = \begin{bmatrix} 0.64 \\ 0.34 \\ 0.02 \end{bmatrix}$

3.2.1 Evaluation Problem with Forward and Backward Algorithm

For the first problem on HMM, will be calculated the probability models $\lambda = (A, B, \pi)$ that represented by $P(O|\lambda)$ or probabilities of observation sequence O = (O1, O2, O3). This probability can be determined using the Forward and Backward algorithm.

$$P(O = MKG|\lambda) = \sum_{i=N}^{N} a_t(i) = a_4(1) + a_4(2) + a_3(3) = 0.091$$

$$P(O = MKG|\lambda) = \sum_{i=N}^{N} \beta_1(i)\pi(i)b_i(O_1)$$

$$= \beta_1(1)\pi(i)b_1(O_1) + \beta_1(2)\pi(2)b_2(O_1) + \beta_1(3)\pi(3)b_3(O_1) = 0.091$$

Based on the results of the backward algorithm obtained are consistent with the obtained solution of the forward algorithms, namely the observation probabilities at 0.091.

3.2.2 Decoding Problem with Viterbi Algorithm

For this problem decoding problem is how to determine the optimal sequence hidden state, in this case is available, not available, or no measurement with the sequence of observations that have been assumed. The Viterbi algorithm consists of three steps, including the following:

$$X_1^* = 1, X_2^* = 1, X_3^* = 1$$

It means the most suitable sequence of available, unavailable, or no measurement rainfall data sequence for August 2019 is more widely available.

3.2.3 Learning Problem with Baum Welch Algorithm

To calculate the parameters of HMM prediction using Baum Welch algorithm, it can be defined a new variable $\xi_t(i, j)$, that probability process in state i at the time $t + 1$. It follows

$$\hat{\pi} = \begin{bmatrix} \gamma_1(1) \\ \gamma_1(2) \\ \gamma_1(3) \end{bmatrix} = \begin{bmatrix} 0.9765 \\ 0.1078 \\ 7.201x10^{-4} \end{bmatrix}$$

The value at $\gamma_t(i)$ for $t = 1$ is an estimated of early chance. That means the value of

$$P(O|\hat{\lambda}) \geqslant P(O|\lambda)$$

has completed, then the probability process in a state of rainfall data availability will be available by 0.9765, it will not be available is equal to 0.1078, and the estimation of the initial probability that no measurements is equal to 7.201 x 10-4.

Meanwhile, for a prediction of the transition matrix a_{ij} that was written with $\hat{a}_{ij}$ is the ratio between the amount of displacement of state probability to state the amount of displacement in the state probability in the state i , as follows:

$$\hat{a}_{ij} = \begin{bmatrix} \frac{\sum_{t-1}^{T} \xi_t(1,1)}{\sum_{t-1}^{T} \gamma_t(1)} & \frac{\sum_{t-1}^{T} \xi_t(1,2)}{\sum_{t-1}^{T} \gamma_t(1)} & \frac{\sum_{t-1}^{T} \xi_t(1,3)}{\sum_{t-1}^{T} \gamma_t(1)} \\ \frac{\sum_{t-1}^{T} \xi_t(2,1)}{\sum_{t-1}^{T} \gamma_t(2)} & \frac{\sum_{t-1}^{T} \xi_t(2,2)}{\sum_{t-1}^{T} \gamma_t(2)} & \frac{\sum_{t-1}^{T} \xi_t(2,3)}{\sum_{t-1}^{T} \gamma_t(2)} \\ \frac{\sum_{t-1}^{T} \xi_t(3,1)}{\sum_{t-1}^{T} \gamma_t(3)} & \frac{\sum_{t-1}^{T} \xi_t(3,2)}{\sum_{t-1}^{T} \gamma_t(3)} & \frac{\sum_{t-1}^{T} \xi_t(3,3)}{\sum_{t-1}^{T} \gamma_t(3)} \end{bmatrix}$$

$$\hat{a}_{ij} = \begin{bmatrix} \frac{2.3702}{2.7670} & \frac{0.3957}{2.7670} & \frac{1.013x10^{-3}}{2.7670} \\ \frac{0.3064}{0.4399} & \frac{0.1331}{0.4399} & \frac{3.621x10^{-4}}{0.4399} \\ \frac{1.157x10^{-3}}{2.273x10^{-3}} & \frac{10^{-3}}{2.273x10^{-3}} & \frac{10^{-5}}{2.273x10^{-3}} \end{bmatrix}$$

$$\hat{a}_{ij} = \begin{bmatrix} 0.85 & 0.14 & 0.01 \\ 0.69 & 0.3 & 0.01 \\ 0.51 & 0.14 & 0.35 \end{bmatrix}$$

The matrix $\hat{a}_(ij)$ is an estimator for the transition matrix a_{ij} . The matrix $\hat{a}_(ij)$ describes to reach a value $P(O \mid \hat{\lambda}) \geqslant$

$P(O \mid \lambda)$, then the transition probabilities of rainfall data at the state of availability will be the probability of the change conditions of "available" to "available" is 0.85, from "available" to "unavailable" is 0.14, from "available" to "no measurement" is 0.01. The probability of "unavailable" to "available" is 0.69, from "unavailable" to "unavailable" is 0.3, from "unavailable" to "no measurement" is 0.01. The probability for conditions as well as of "no measurement" to "available" is 0.51, "no measurement" to "unavailable" is 0.14, "no measurement" to "no measurement" is 0.35.

Likewise with predictions emission matrix b_{ik} , which is denoted by $\hat{b}_{ik}$ obtained from a comparison of the number of states that produce k observations when the process is in state i with the number of the entire process is in state i so that the emission matrix estimators obtained as follows

$$\hat{a}_{ij} = \begin{bmatrix} \frac{\sum_{t-1,O_t-1}^{T} \gamma_t(1)}{\sum_{t-1}^{T} \gamma_t(1)} & \frac{\sum_{t-1,O_t-2}^{T} \gamma_t(1)}{\sum_{t-1}^{T} \gamma_t(1)} & \frac{\sum_{t-1,O_t-3}^{T} \gamma_t(1)}{\sum_{t-1}^{T} \gamma_t(1)} \\ \frac{\sum_{t-1,O_t-1}^{T} \gamma_t(2)}{\sum_{t-1}^{T} \gamma_t(2)} & \frac{\sum_{t-1,O_t-2}^{T} \gamma_t(2)}{\sum_{t-1}^{T} \gamma_t(2)} & \frac{\sum_{t-1,O_t-3}^{T} \gamma_t(2)}{\sum_{t-1}^{T} \gamma_t(2)} \\ \frac{\sum_{t-1,O_t-1}^{T} \gamma_t(3)}{\sum_{t-1}^{T} \gamma_t(3)} & \frac{\sum_{t-1,O_t-2}^{T} \gamma_t(3)}{\sum_{t-1}^{T} \gamma_t(3)} & \frac{\sum_{t-1,O_t-3}^{T} \gamma_t(3)}{\sum_{t-1}^{T} \gamma_t(3)} \end{bmatrix}$$

$$\hat{a}_{ij} = \begin{bmatrix} \frac{0.9765}{2.7670} & \frac{0.9961}{2.7670} & \frac{0.7944}{2.7670} \\ \frac{0.1078}{0.4399} & \frac{0.1298}{0.4399} & \frac{0.2023}{0.4399} \\ \frac{7.201x10^{-3}}{2.273x10^{-3}} & \frac{5.074x10^{-4}}{2.273x10^{-3}} & \frac{1.045x10^{-5}}{2.273x10^{-3}} \end{bmatrix}$$

$$\hat{a}_{ij} = \begin{bmatrix} 0.35 & 0.36 & 0.29 \\ 0.25 & 0.29 & 0.46 \\ 0.32 & 0.22 & 0.46 \end{bmatrix}$$

The matrix $\hat{a}_{ik}$ is an estimator for the conditional probability of matrix $\hat{b}_{ik}$. The matrix describes that to achieve value $P(O \mid \hat{\lambda}) \geqslant P(O \mid \lambda)$, then the chances of availability of rainfall data for the period one day ahead for MM station is 0.35, for PPC station is 0.36, for SBG station is 0.29. Probabilities unavailability of rainfall data for the period of one day ahead for MM station is 0.25, for PPC station is 0.29, for SBG station is 0.46. Probabilities no measurement on rainfall data for the period of one day ahead for MM station is 0.32, for PPC station is 0.22, for SBG station 0.46. Based on the value of the probability of the rainfall data, SBG station has the greatest probability of unavailability and no measurements of rainfall data. It also impacts to sectors that have direct relation to the weather with rainfall conditions; one of them is in the agriculture sector that requires weather information to estimate an increase of output product. Hence, it is necessary to improve human resources and also a good device to do measurements of rainfall data.

4. CONCLUSIONS

A stochastic process that meets the properties of probability in which the properties of future events depend on the properties of events in the present and the past and assumed there are properties of events that cannot be observed is called the Hidden Markov Model (HMM). Hidden Markov model by the learning problems with Baum-Welch algorithm might be the probability for available data the highest rainfall in the period of the day to come in PPC station is 0.36. It is followed by probability available data is 0.35 in MM station and 0.29 in SBG station. As for decoding problem on HMM using the Viterbi algorithm can be concluded that for the period of one day ahead might be probabilities rainfall data availability in MM station, PPC station, and SBG station will be more widely available. The results of probability rainfall data availability can available. For the future, the results of probability rainfall availability in agriculture are to help anticipate extreme climate change, and can provide information and early warning to farm communities about drought or flooding. In addition, this information is also needed in disaster mitigation as a basis for determining the policy to be taken.

REFERENCES

Antonucci, A., R. De Rosa, A. Giusti, and F. Cuzzolin (2015). Robust classification of multivariate time series by imprecise hidden Markov models. *International Journal of Approximate Reasoning*, **56**; 249–263

Bain, L. and M. Engelhardt (1992). *Introduction to Probability and Mathematical Statistics*. Duxbury Press: California

Bathaee, N. and H. Sheikhzadeh (2016). Non-parametric Bayesian inference for continuous density hidden Markov mixture model. *Statistical Methodology*, **33**; 256–275

Colombi, R. and S. Giordano (2015). Multiple hidden Markov models for categorical time series. *Journal of Multivariate Analysis*, **140**; 19–30

Devianto, D. (2016). The uniform continuity of characteristic function from convoluted exponential distribution with stabilizer constant. In *AIP Conference Proceedings*. AIP Publishing LLC

Devianto, D., Maiyastri, L. Oktasari, and M. Anas (2015a). Convolution of generated random variable from exponential distribution with stabilizer constant. *Applied Mathematical Sciences*, **9**; 4781–4789

Devianto, D., M. Maiyastri, and S. Damayanti (2015b). Forecasting Long memory Time Series for Stock Price with Autoregressive Fractionally Integrated Moving Average. *International Journal of Applied Mathematics and Statistics*, **53**(5); 86–95

Devianto, D., L. Oktasari, and Maiyastri (2015c). Some properties of hypoexponential distribution with stabilizer constant. *Applied Mathematical Sciences*, **9**; 7063–7070

Greene, A. M., A. W. Robertson, P. Smyth, and S. Triglia (2011). Downscaling projections of Indian monsoon rainfall using a non-homogeneous hidden Markov model. *Quarterly Journal of the Royal Meteorological Society*, **137**(655); 347–359

Khadr, M. (2016). Forecasting of meteorological drought using Hidden Markov Model (case study: The upper Blue Nile river basin, Ethiopia). *Ain Shams Engineering Journal*, **7**(1); 47–56

Li, X., V. Makis, H. Zuo, and J. Cai (2018). Optimal Bayesian control policy for gear shaft fault detection using hidden semi-

Markov model. *Computers & Industrial Engineering*, **119**; 21–35

Mehrotra, R. and A. Sharma (2005). A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. *Journal of Geophysical Research: Atmospheres*, **110**(D16)

Pineda, L. E. and P. Willems (2016). Multisite downscaling of seasonal predictions to daily rainfall characteristics over Pacific–Andean River Basins in Ecuador and Peru Using a nonhomogeneous Hidden Markov model. *Journal of Hydrometeorology*, **17**(2); 481–498

Ramadhan, R., D. Devianto, and M. Maiyastri (2020). Hidden Markov Model for Exchange Rate with EWMA Control Chart. In *Proceedings of the Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia*. EAI

Robertson, A. W., S. Kirshner, and P. Smyth (2004). Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *Journal of climate*, **17**(22); 4407–4424

Ross, S. M. (1996). *Stochastic Processes P.* John Wiley & Sons

Sansom, J. (1998). A hidden Markov model for rainfall using breakpoint data. *Journal of Climate*, **11**(1); 42–53

Spezia, L., S. Cooksley, M. Brewer, D. Donnelly, and A. Tree (2014). Modelling species abundance in a river by Negative Binomial hidden Markov models. *Computational Statistics & Data Analysis*, **71**; 599–614

Stoner, O. and T. Economou (2019). *An Advanced Hidden Markov Model for Hourly Rainfall Time Series.* Cornell University Report: UK, arXiv: 1906.03846

Thyer, M. and G. Kuczera (2003). A hidden Markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach. *Journal of Hydrology*, **275**(1-2); 12–26

Tseng, Y.-T., S. Kawashima, S. Kobayashi, S. Takeuchi, and K. Nakamura (2020). Forecasting the seasonal pollen index by using a hidden Markov model combining meteorological and biological factors. *Science of The Total Environment*, **698**; 134246

Xia, Y.-M. and N.-S. Tang (2019). Bayesian analysis for mixture of latent variable hidden Markov models with multivariate longitudinal data. *Computational Statistics & Data Analysis*, **132**; 190–211

Zhang, M., X. Jiang, Z. Fang, Y. Zeng, and K. Xu (2019). High-order Hidden Markov Model for trend prediction in financial time series. *Physica A: Statistical Mechanics and its Applications*, **517**; 1–12

