Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 3, 2025
DOI: 10.30829/zero.v9i3.25443
Page: 703 - 714

Rainfall Risk Modelling for Rice Farming Using Continuous Hidden Markov Models

[1] **David Vijanarco Martal**
School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

[2] **Berlian Setiawaty**
School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

[3] **Retno Budiarti**
School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

Article Info

Article history:

Accepted 26 December 2025

Keywords:

Agricultural risk;
Hidden Markov Model;
Rainfall Simulation;
Time Series Analysis;

ABSTRACT

Climate change has increased rainfall variability and unpredictability, significantly impacted agricultural productivity, and raised the risk of crop failure, particularly in rain-fed rice farming systems. This study models rainfall data from Tabanan, Bali, using a continuous-time Hidden Markov Model (HMM) to identify latent weather states and assess the associated risk of rice crop failure. The model assumes four hidden states, each generating rainfall observations following a Gamma distribution. Simulation results produced Mean Absolute Percentage Error (MAPE) values below 5% for training and testing sets, indicating strong model performance in replicating rainfall patterns. Risk analysis compared simulated rainfall with rice crop water requirements across three planting periods. The second planting period (July–October) exhibited the highest risk at 3.75%. Compared to other predictive models, HMM offers superior capability in capturing temporal rainfall structure and identifying critical transition phases, making it highly suitable for agricultural risk assessment and climate-adaptive planning.

Corresponding Author:

Berlian Setiawaty,
School of Data Science, Mathematics, and Informatics,
IPB University, Bogor, Indonesia
Email: berlianse@apps.ipb.ac.id

1. INTRODUCTION

Indonesia is one of the biggest agricultural countries in the world, with rice as one of its major commodities. However, the increasing variability in rainfall patterns caused by global climate change has become a major challenge for the agricultural sector. Uncertainty in the amount and distribution of rainfall can significantly affect climate change [1]. This increasing variability of rainfall patterns also impacts the uncertainty of the planting schedules, crop growth, and overall agricultural productivity [2]. In agrarian countries like Indonesia, where many farmers rely heavily on rain-fed agriculture, the risks posed by irregular rainfall patterns can lead to reduced yields, economic losses, and food insecurity [3].

Rainfall plays a crucial role in the success of the agricultural sector. Its instability due to climate change poses a serious challenge, especially in water risk management issues [4]. Excessive and insufficient rainfall can reduce agricultural productivity [5]. This situation is increasingly concerning as changes in rainfall patterns in tropical regions like Indonesia are expected to become more extreme and unpredictable. Through modelling approaches, policymakers, researchers, and stakeholders can thoroughly comprehend how climate change may impact different dimensions of agriculture, such as plant development, water resources [6], soil quality, and agricultural productivity [7].

To address these challenges, predictive models that can capture the stochastic nature of rainfall are essential for supporting data-driven agricultural planning. One promising statistical approach is the hidden Markov model (HMM), which is particularly effective for modelling time series data with underlying hidden states [8]. In climate and weather analysis, HMM can represent unobserved weather, such as wet and dry conditions, and associate them with observable rainfall data distributed to a specific continuous distribution [9]. The application of HMM in time series analysis has become increasingly widespread, demonstrating success in diverse fields such as modelling trend prediction in financial markets [10], disease analysis [11], Ecology [12], and human behaviours [13]. In the analysis of weather data itself, HMM is widely applied to uncovering hidden patterns in atmospheric dynamics. For example, Hmm can be implemented to capture characteristics of rainfall and temperature, as well as seasonal or daily transitions between weather states [14], [15], [16].

This study aims to develop a hidden Markov model for rainfall data to uncover hidden weather patterns and measure the associated risks to agriculture. By applying HMM to rainfall data, we can estimate transition probabilities between different weather states and characterise rainfall distribution within each state. These allow for a quantitative assessment of agricultural risk, such as delayed rainy seasons, extreme rainfall events, or uneven rainfall distribution throughout the growing period. This modelling framework enables probabilistic forecasting of extreme events, making it valuable for agricultural planning, disaster risk reduction, and water resource management. Such insights are valuable for early warning systems and designing more resilient farming strategies in the face of climate uncertainty. The outcomes of this research will likely provide helpful information for policymakers and farmers in managing climate-related risks and improving the sustainability of agricultural practices.

2. RESEARCH METHODE

2.1 Continuous Hidden Markov Model (HMM)

A hidden Markov model is a probabilistic model that assumes the observed data are generated by an underlying sequence of unobservable (hidden) states, which evolve according to a Markov process [17]. Consider a pair stochastic process (S, O) where S is an unobserved process and O is the observed process. The hidden process S is a Markov chain and O is a process whose distribution depends on the value of S_t [18]. Based on the observation data type, HMM is divided into two types, discrete HMM and continuous HMM. In a continuous HMM, the observation O_t is a continuous random variable and the conditional distribution of O_t given $S_t = i$ for $i = 1,2,\dots,N$ come from a specific family of parametric distributions.

A continuous HMM can be parametrized by $\lambda = (A, B, \pi)$ and has the following characteristics:

- $A = \left[a_{ij}\right]_{N\times N}$ is a transition probability matrix, with $a_{ij} = P(S_{t+1} = j \,|S_t = i)$, for $i, j = 1,2,\dots,N$ and $\sum_{j=1}^{N} a_{ij} = 1$, for $i = 1,2,\dots,N$.
- $B = [f(o_t \,|\theta_i)]_{N\times 1}$is a probability density function matrix, with $f(o_t \,|\theta_i)$ is the conditional probability density function of $O_t \,|S_t = i$, for $i = 1,2,\dots,N$.
- $\pi = [\pi_i]_{N\times 1}$ is an initial probability matrix, with $\pi_i = P(S_1 = i)$ for $i = 1,2,\dots,N$ and $\sum_{i=1}^{N} \pi_i = 1$.

Estimating the transition matrix A involves determining the probabilities of transitioning between hidden states based on the observed data. Accurate estimation of A is essential, as it governs the probability of remaining in or moving between different states. In principle, this transition makes it possible to move from one condition to a different condition characterized by the distribution in matrix B. Equally important are the initial state probabilities π, which define the probability of the system starting in each hidden weather state. These initial probabilities play a critical role in short-term simulations and risk assessment, particularly at the beginning of the simulation. However, over a longer time horizon, the influence of initial conditions diminishes as the system approaches a stationary distribution determined by the transition matrix A. Therefore, a solid understanding of both A and π is essential for accurate short-term forecasting and long-term analysis of the data behaviour. Together, they form the backbone of the HMM's dynamic structure, driving the evolution of hidden state regimes while anchoring the model's starting assumptions, ensuring reliable simulation, forecasting, and interpretation of rainfall patterns.

The parameters of the model $\lambda = (A, B, \pi)$ including the initial state probabilities (π), transition probabilities (A), and the parameters of the continuous distributions (B) are estimated using the Expectation-

Maximization (EM) algorithm, specifically the Baum-Welch algorithm [19]. The EM algorithm is used to find $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ that maximize the joint density of observations $o_1, \ldots, o_T$

$$p(o_1, \ldots, o_T \mid \hat{\lambda}) = \max_{\lambda \epsilon \Lambda} p(o_1, \ldots, o_T \mid \lambda).$$

This iterative method refines the model by maximizing the likelihood of the observed rainfall data given the hidden states. It updates the transition matrix by calculating the expected number of transitions between states and adjusts the emission distribution parameters to best fit the observed rainfall associated with each state. Accurate estimation of these parameters is crucial, as the transition matrix governs how likely the system is to move between different weather conditions over time. This directly influences the model's ability to simulate realistic rainfall sequences and identify periods of dry or wet conditions, which are essential for assessing agricultural risk and predicting rainfall-driven crop failures.

The EM algorithm consists of two stages as follow [20].

- E step

  For a given $\lambda = (A, B, \pi)$ define two functions G and H in Λ which is a set of all possible parameter for HMM. For all $\lambda' = (A', B', \pi') \in \Lambda$, define $G(\lambda, \lambda'), H(\lambda, \lambda')$, and $l(\lambda')$ as follow.

$$G(\lambda, \lambda') = \sum_{s_T, \ldots, 1}^{N} p(s_1, \ldots, s_T \mid o_1, \ldots, o_T, \lambda) \times \log[p(o_1, \ldots, o_T, s_1, \ldots, s_T \mid \lambda')]. \quad (1)$$

$$H(\lambda, \lambda') = \sum_{s_T, \ldots, 1}^{N} p(s_1, \ldots, s_T \mid o_1, \ldots, o_T, \lambda) \times \log[p(s_1, \ldots, s_T \mid o_1, \ldots, o_T, \lambda')]. \quad (2)$$

$$l(\lambda') = \log[p(o_1, \ldots, o_T \mid \lambda')] = G(\lambda, \lambda') - H(\lambda, \lambda'). \quad (3)$$

- M step

  On this stage, we will find $\hat{\lambda} \in \Lambda$ such that

$$G(\lambda, \hat{\lambda}) = \max_{\lambda' \in \Lambda} G(\lambda, \lambda').$$

Based on the equations in the E step, the following things can be shown [21].

$$\begin{aligned} l(\hat{\lambda}) - l(\lambda) &= G(\lambda, \hat{\lambda}) - H(\lambda, \hat{\lambda}) - (G(\lambda, \lambda) - H(\lambda, \lambda)) \\ &= (G(\lambda, \hat{\lambda}) - G(\lambda, \lambda)) + (H(\lambda, \lambda) - H(\lambda, \hat{\lambda})) \end{aligned}$$

Since

$$H(\lambda, \lambda) \geq H(\lambda, \hat{\lambda}), \qquad \hat{\lambda} \in \Lambda.$$

So that

$$l(\hat{\lambda}) - l(\lambda) \geq G(\lambda, \hat{\lambda}) - G(\lambda, \lambda) \geq 0.$$

This result prove that

$$l(\hat{\lambda}) \geq l(\lambda).$$

Because logarithmic is an increasing function, it can be concluded that

$$p(o_1, \ldots, o_T \mid \hat{\lambda}) \geq p(o_1, \ldots, o_T \mid \lambda).$$

Now, the optimization problem becomes finding the parameter $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ that maximize G with two constrains. The problem is

$$\max_{\lambda' \in \Lambda} G(\lambda, \lambda')$$

with $G(\lambda, \lambda')$ in equation (1) satisfies the following equation.

$$\begin{aligned} G(\lambda, \lambda') = & \sum_{s_t=1}^{N} p(s_t \mid o_1, \ldots, o_T, \lambda) \log \pi'_{s_t} \\ & + \sum_{s_t=1}^{N} \sum_{s_{t+1}=1}^{N} \sum_{t=1}^{T-1} p(s_t \mid o_1, \ldots, o_T, \lambda) a'_{s_t s_{t+1}} \\ & + \sum_{s_t=1}^{N} \sum_{t=1}^{T-1} p(s_t \mid o_1, \ldots, o_T, \lambda) \log f'(o_t \mid \theta_{s_t}) \end{aligned} \quad (4)$$

with constrains

$$\sum_{s_t=1}^{N} \pi'_{s_t} = 1$$

$$\sum_{s_{t+1}=1}^{N} a'_{s_t s_{t+1}} = 1, \quad s_t = 1,2,\ldots,N.$$

This problem can be solved using the multiplier Lagrange. From equation (4), define the Lagrange equation for $\lambda' \in \Lambda$ as

$$L(\lambda') = G(\lambda,\lambda') + \eta_1\left(\sum_{s_t=1}^{N}\pi'_{s_t} - 1\right) + \sum_{s_t=1}^{N}\eta_2\left(\sum_{s_{t+1}=1}^{N}a'_{s_t s_{t+1}} - 1\right) \quad (5)$$

with η_1 and η_2 are Lagrange multiplier variables. By solving three equations below

$$\frac{\partial L}{\partial \pi'_{s_t}} = 0, \qquad \frac{\partial L}{\partial a'_{s_t s_{t+1}}} = 0$$

And we obtain the estimate parameters $\hat{A}$ and $\hat{\pi}$ as follows.

$$\hat{a}_{s_t s_{t+1}} = \frac{\sum_{t=1}^{T-1}\xi_t(s_t,s_{t+1})}{\sum_{t=1}^{T-1}\gamma_t(s_t)}, \qquad \hat{\pi} = \gamma_1(s_t)$$

with

$$\xi_t(s_t,s_{t+1}) = p(s_t,s_{t+1}|o_1,\dots,o_T,\lambda) \quad (6)$$

and

$$\gamma_t(s_t) = p(s_t\,|o_1,\dots,o_T,\lambda). \quad (7)$$

Probability $\xi_t(s_t,s_{t+1})$ is the probability of state s_t at time t and state s_{t+1} at time $t+1$ given a series of observations sequence $o_1,o_2,\dots,o_T$ and model λ. Whereas $\gamma_t(s_t)$ is the probability of being in state s_t at time t given a series of observations sequence $o_1,o_2,\dots,o_T$ and model λ [22]. These two probabilities can be calculated using forward and backward algorithms [23].

Let $\alpha_t(s_t)$ is called forward variable and can be calculated as follow, for $t = 1,2,\dots,T$,

$$\alpha_t(x_t) = p(o_1,\dots,o_T,s_t\mid\lambda), \qquad s_t = 1,2,\dots,N. \quad (8)$$

Let $\beta_t(s_t)$ is called backward variable and can be calculated as follow, for $t = T-1,T-2,\dots,1$,

$$\beta_t(x_t) = p(o_{t+1},\dots,o_T,s_t\mid\lambda), \qquad s_t = 1,2,\dots,N. \quad (9)$$

Using equation (8) and (9), probability $\xi_t(s_t,s_{t+1})$ can be reform, for $t = 1,2,\dots,N$,

$$\xi_t(s_t,s_{t+1}) = \frac{\alpha_t(s_t)a_{s_t s_{t+1}}f\left(o_{t+1}\middle|\theta_{s_{t+1}}\right)\beta_{t+1}(s_{t+1})}{\sum_{s_t=1}^{N}\alpha_t(s_t)\beta_t(s_t)} \quad (10)$$

and $\gamma_t(s_t)$ can be reform, for $t = 1,2,\dots,N$,

$$\gamma_t(s_t) = \frac{\alpha_t(s_t)\beta_{t+1}(s_{t+1})}{\sum_{s_t=1}^{N}\alpha_t(s_t)\beta_t(s_t)}. \quad (11)$$

For parameter $\hat{B}$, the result depends on distribution of $f(o_t\,|\theta_{s_t})$. In this research, the distributions used is selected from family continuous distribution, such as Weibull, exponential, gamma, lognormal, and Pareto distribution. The parameters of each distribution are estimated using the maximum likelihood estimation (MLE).

Once trained, the model is used to infer the most likely sequence of hidden states via the Viterbi algorithm [21] and to simulate synthetic rainfall series that preserve the temporal dependence structure observed in the original data. Let $\delta_t(s_t)$ represents the highest probability in first t observations and ends in state s_t. This variable can be defined as follow.

$$\delta_t(s_t) = \max_{s_1,\dots,s_{t-1}} p(o_1,\dots,o_t,s_1,\dots,s_t\mid\lambda), \quad s_t = 1,2,\dots,T. \quad (12)$$

Using induction, will be obtain

$$\delta_{t+1}(s_{t+1}) = \max_{s_1,\dots,s_t} f(o_{t+1}\,|\theta_{s_{t+1}})\left[\delta_t(s_t)a_{s_t s_{t+1}}\right], \quad s_{t+1} = 1,2,\dots,T. \quad (13)$$

Let array $\psi_{t+1}(s_{t+1})$ represents the state at time t from which a transition to state s_{t+1} maximizes the probability $\delta_{t+1}(s_{t+1})$. This array can be define as follow.

$$\psi_{t+1}(s_{t+1}) = \arg\max_{s_t=1,\dots,N}\left[\delta_t(s_t)a_{s_t s_{t+1}}\right], \quad s_{t+1} = 1,2,\dots,T. \quad (14)$$

The steps in the Viterbi algorithm as follow.

- Initialization

$$\delta_1(x_1) = \pi_{s_1}f(o_1\,|\theta_{s_1}), \qquad s_1 = 1,2,\dots,N.$$
$$\psi_1(s_1) = 0, \qquad s_1 = 1,2,\dots,N.$$

- Recursion for $t = 1,2,\dots,T-1$

$$\delta_{t+1}(s_{t+1}) = \max_{s_t=1,\dots,N}\left[\delta_t(s_t)a_{s_t s_{t+1}}\right]f(o_{t+1}\,|\theta_{s_{t+1}}), \qquad s_{t+1} = 1,2,\dots,N.$$
$$\psi_{t+1}(s_{t+1}) = \arg\max_{s_t=1,2,\dots,N}\delta_t(s_t)a_{s_t s_{t+1}}, \qquad s_{t+1} = 1,2,\dots,N.$$

- Termination

$$p^* = \max_{s_T=1,\dots,N}[\delta_T(s_T)].$$

$$s_T^* = \arg\max_{s_T=1,\ldots,N}[\delta_T(s_T)].$$

- Backtracking

$$s_{t-1}^* = \psi_t(s_t^*), \qquad t = T, T-1, \ldots, 2.$$

From this step, will be obtain

$$\{s_T^*, s_{T-1}^*, \ldots, s_1^*\}.$$

2.2 Rainfall Data

The rainfall data used in this study were obtained from power.larc.nasa.gov in the form of 15-day rainfall time series covering the period from January 2010 to February 2023. The total number of data is 316, with 2 data each month. The data were collected from climatological or meteorological stations located in the region of Tabanan Regency, Bali. The data are divided into 292 training data points and 24 testing data points. The testing data is also used as simulation data to calculate risk probability. All computations are performed using a custom implementation in Mathematica 12.3, where both the estimation and simulation phases are handled through numerical optimization routines and stochastic sampling. The plot of the training and testing data used can be seen in **Figure 1** and **Figure 2** below.

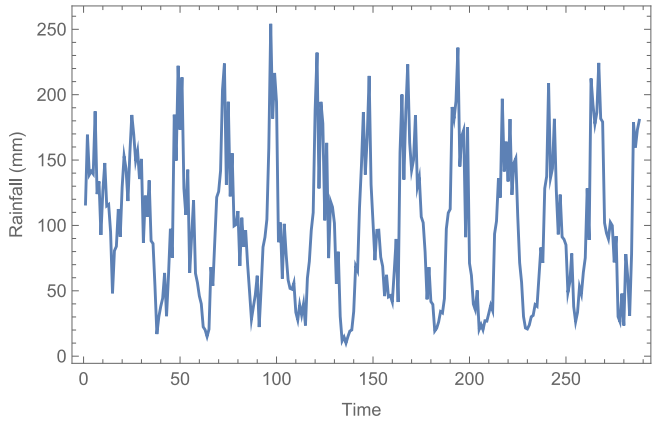

Figure 1. Plot of training data for the January 2010 - February 2022 period

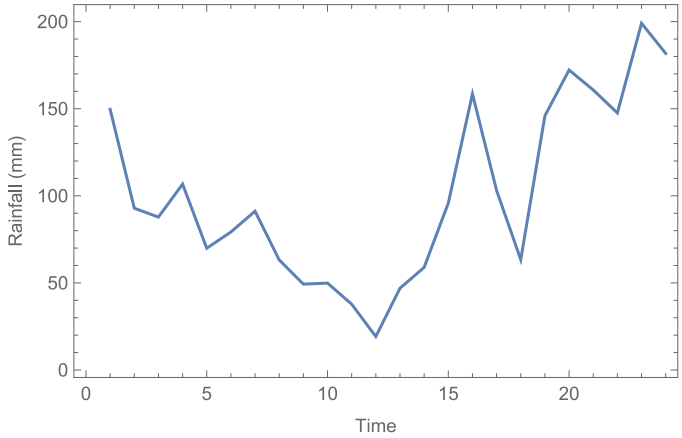

Figure 2. Plot of testing data for the March 2022 - February 2023 period

3. RESULT AND ANALYSIS

3.1 Hidden Markov Model Modelling

The initial stage of data modelling involves identifying the optimal number of hidden states with the appropriate probability distribution. One commonly used method for model selection is the Akaike information criterion (AIC). The AIC can be calculated as follow [24].

$$\text{AIC} = -2logL + 2(N^2 + kN - 1) \tag{15}$$

where L is the likelihood, N is number of hidden states, and k is the number of parameters in each distribution. The results of AIC calculations can be seen in the **Table 1** below.

Table 1. Akaike information criterion calculation result

Number of Hidden states	Weibull	Exponential	Gamma	Lognormal	Pareto
2	2928.41	3230.63	2919.94	2925.72	3270.13
3	2856.32	3242.68	2848.92	2856.82	3262.68
4	2848.45	3258.54	2831.92	2862.29	3073.91
5	2849.71	3277.75	2842.14	2847.71	2997.45

Based on the **Table 1** above, the yellow number indicate the smallest AIC value. From these results, the most optimal model is a continuous time HMM with 4 states and Gamma distribution. The estimated parameter $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ as follows.

$$\hat{A} = \begin{bmatrix} 0.555 & 0.213 & 0.029 & 0.202 \\ 0.249 & 0.751 & 4.429 \times 10^{-55} & 1.267 \times 10^{-5} \\ 4.766 \times 10^{-16} & 1.724 \times 10^{-132} & 0.858 & 0.142 \\ 0.192 & 2.436 \times 10^{-14} & 0.171 & 0.637 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} Gamma(14.477, 3.917) \\ Gamma(8.789, 3.183) \\ Gamma(23.822, 6.853) \\ Gamma(30.853, 3.152) \end{bmatrix} \quad \hat{\pi} = \begin{bmatrix} 7.17 \times 10^{-186} \\ 0. \\ 1. \\ 7.83 \times 10^{-19} \end{bmatrix}.$$

From the parameters above, each state tries to remain in that state without moving with a probability greater than 0.5. Each state communicates with each other even though there are pairs of states that are almost impossible to transition directly, such as states 2 and 3. Using the Viterbi algorithm, for the estimated parameter $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$, we can find the sequence of hidden states that generated the data. The data and the estimated hidden states are shown in **Figure 3** as below.

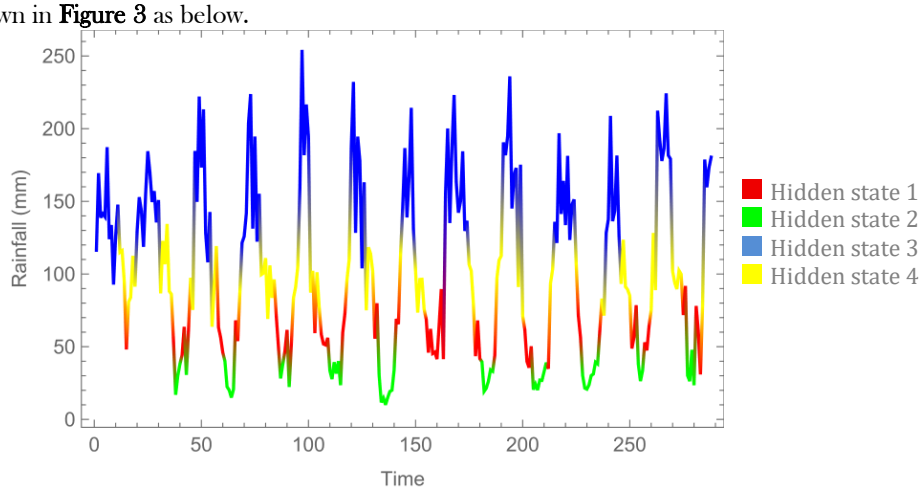

Figure 3. Plot of hidden state sequence

Each hidden state distribution on parameter B is then tested using the Kolmogorov-Smirnov test with the hypothesis:

H_0: the data is gamma distributed.

H_1: the data is not gamma distributed.

The results of the Kolmogorov-Smirnov test can be seen in **Table 2** below.

Table 2. Kolmogorov-Smirnov test on the distribution of each hidden state

Hidden state	parameter	Statistic	*p*-value
1	(14.477, 3.917)	0.079	0.836
2	(8.789, 3.183)	0.089	0.747
3	(23.822, 6.853)	0.073	0.604
4	(30.853, 3.152)	0.072	0.826

Based on the **Table 2**, the p-value obtained for each hidden state data is greater than 0.05. This means there is insufficient evidence to reject H_0, so it can be accepted that all hidden state distributions have been estimated well.

The obtained HMM parameter $\hat{\lambda} = (\hat{A}, \hat{B}, \hat{\pi})$ is then used for simulations on training and testing data. The simulations were conducted using Mathematica 12.3 software by generate random data. In this research, the accuracy metric mean absolute percentage error (MAPE) is used as a benchmark for model error relative to actual data. MAPE can be calculated as follow.

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{o_t - \hat{o}_t}{o_t}\right| \tag{16}$$

with o_t is actual data, $\hat{o}_t$ is prediction data, and T is number of data [25].

The simulation results show MAPE results for the training data are 4.973% and 1.177 %. The plot comparing the simulated data using HMM on the training data with the actual data can be seen in **Figure 4** below.

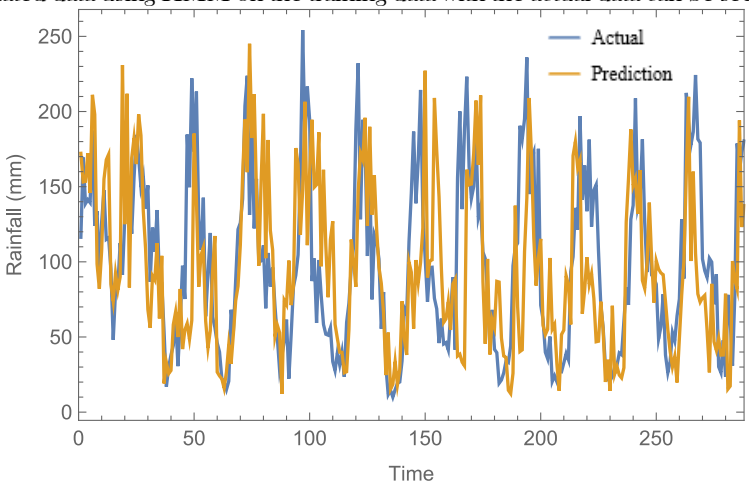

Figure 4. Plot of prediction and actual data for the January 2010 - December 2021 period

Figure 4 demonstrates that HMM successfully captures the underlying pattern of the data. The yellow line, representing the actual data, closely follows the blue line, which representing the simulated output. Nevertheless, there are still certain time intervals where the model fails to accurately replicate the original data. The simulation results show MAPE results for the testing data is 0.994%. A comparison between the HMM-simulated data and the actual data during the testing period is illustrated in **Figure 5** below.

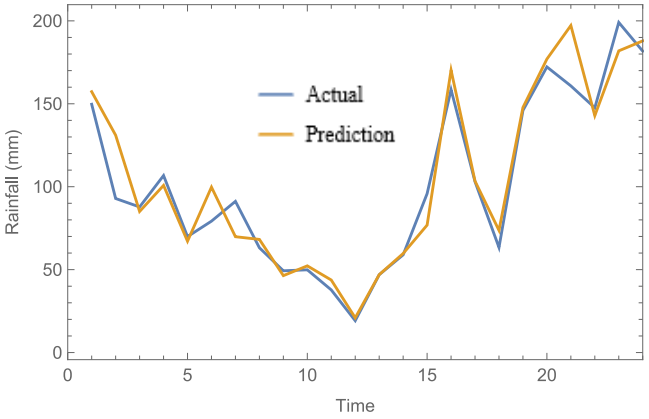

Figure 5. Plot of prediction and actual data for the January 2022 - December 2022 period

Based on **Figure 5**, simulation results demonstrate that the HMM successfully predicts testing data. The plots of actual and prediction almost coincide in some places, but there are still some errors at certain times. Simulations on training and testing data yielded a MAPE of less than 5%. This indicates that HMM can model rainfall data accurately. Based on these results, risk calculations on rainfall data can be performed using a hidden Markov model.

3.2 Model Interpreted

The HMM constructed in this study consists of 4 hidden states representing the primary weather conditions, namely the dry, very dry, wet, and transition state. Each state is assumed to generate rainfall data following a

gamma distribution with different parameter, which allows the model to capture the continuous and asymmetric characteristics of rainfall distribution.

The parameter estimation results indicate that:

- State 1 ("dry") is associated with a low average rainfall, with a narrow and consistent distribution close to zero. This corresponds to dry season conditions or days without significant precipitation.
- State 2 ("very dry") is associated with extremely low or zero rainfall, with a very narrow distribution close to zero. It represents days with no precipitation at all, often occurring during the peak of the dry season.
- State 3 ("wet") is characterized by moderate average rainfall and increased variability, this state represents regular rain events, commonly seen during the wet season.
- State 4 ("very wet") shows high mean rainfall with large variability, indicating periods of heavy and intense precipitation, such as during peak rainy season or extreme weather events.

The estimated transition matrix reveals a strong tendency toward state persistence. The probability of remaining in the same state from one time to the next is relatively high. This suggests that the model successfully captures the seasonal structure of rainfall patterns, which tend to occur in extended periods rather than shifting randomly. The distribution of hidden states obtained through simulation also shows a clear seasonal pattern. The wet state appears more frequently during specific months (e.g., December to March), while the dry state dominates the other periods. This aligns with the typical tropical seasonal climate patterns observed in Indonesia and surrounding regions.

3.3 Risk Analysis

In rice farming activities, water availability is the most important aspect. Rice cultivation in this Tabanan region is generally divided into three planting periods: March–June, July–October, and November–February, with different water requirements for each growth phase. Water availability is greatly influenced by rainfall, so fluctuations in rainfall can cause a risk of water shortages or excesses, impacting crop productivity.

The application of the Hidden Markov Model allows for a more structured understanding of rainfall variability and its implications for agricultural risk. Through the identification of distinct hidden weather states particularly dry and wet periods, this model provides insight into both the frequency and intensity of rainfall events over time. Such information is crucial in assessing the level of exposure that agricultural systems have to weather-related risks. This risk analysis aims to evaluate the potential mismatch between agricultural crop water requirements and predicted rainfall. By understanding these risks, mitigation strategies can be formulated to minimize the negative impacts of water shortages or excesses on agriculture. In this risk assessment, upper and lower thresholds are set at 50% of the effective requirement to reflect the plant's tolerance to water availability.

The data used for comparison was from March 2022 to February 2023, covering 3 planting periods. Data randomly generated, with 40000 samples with MAPE less than 15%. This generated data then reviewed to determine whether there were specific periods exceeding the upper or lower threshold for water requirements. The first planting period, from March to June 2022, was compared with simulated data. An overview of water requirements for the first planting period can be seen in **Figure 6** below.

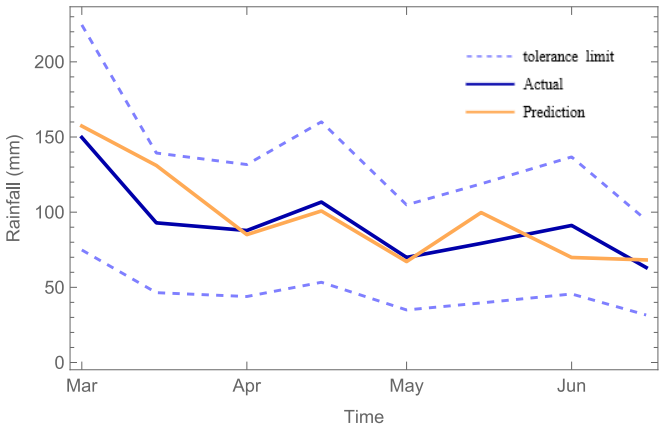

Figure 6. Comparison plot of water requirements and estimated the first planting period

Figure 6 is an example of a comparison of one of the datasets during the first planting period. The dashed lines represent the upper and lower threshold values, while the blue line represents the water requirement data for the first planting period. Risk calculations for the first planting period can be seen in **Table 3** as follow.

Table 3. Risk calculates for the first planting period

Period	Shortage of Water	Excess Water
Mar I	8	8
Mar II	1	18
Apr I	3	14
Apr II	15	12
Mei I	6	17
Mei II	4	8
Jun I	8	0
Jun II	1	13
Total	**46**	**90**

Based on **Table 3**, the exact number of points exceeding the threshold for each period can be identified. This number indicates the risk of crop failure during the first planting period. Crop failure during the first planting period due to water shortages was recorded at 46 out of a total of 4000 estimates, or 1.15%. Meanwhile, crop failure due to excess water was recorded at 90 out of a total of 4000 estimates, or 2.25%. The overall risk of crop failure during the first planting period was 145 out of 4000 estimates, or 3.4%. Next, we will show a comparison of the estimated data with the water requirements for the second planting period. The second planting period, from July to October 2022, was compared with simulated data. An overview of water requirements for the second planting period can be seen in **Figure 7** below.

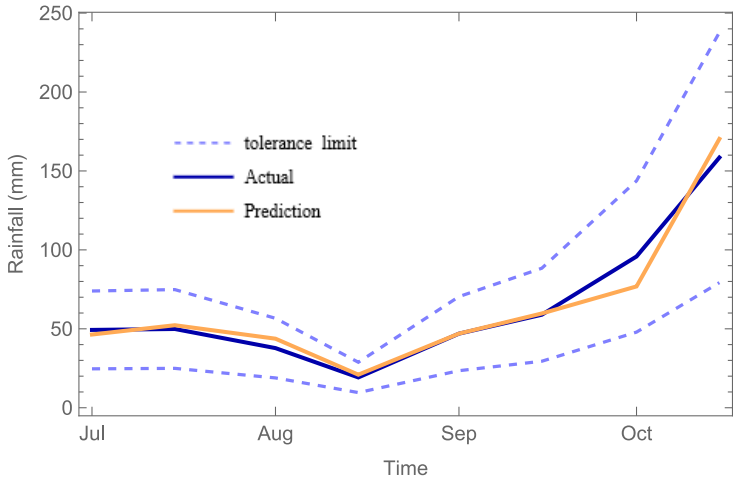

Figure 7. Comparison plot of water requirements and estimated the second planting period

Figure 7 is an example of a comparison of one of the datasets during the second planting period. As before, the dashed lines represent the upper and lower threshold values, while the blue line represents the water requirement data for the second planting period. In this example, there is no risk of water shortage or excess in rice farming. Risk calculations for the second planting period can be seen in **Table 4** as follow.

Table 4. Risk calculates for the second planting period

Period	Shortage of Water	Excess Water
Jul I	8	17
Jul II	17	6
Aug I	11	10
Aug II	1	48
Sep I	2	8
Sep II	2	7
Oct I	3	0
Oct II	10	0
Total	**54**	**96**

Based on **Table 4**, the exact number of points exceeding the threshold for each period can be identified. This number indicates the risk of crop failure during the second planting period. Crop failure during the second planting period due to water shortages was recorded at 54 out of a total of 4000 estimates, or 1.35%. Meanwhile, crop failure due to excess water was recorded at 96 out of a total of 4000 estimates, or 2.4%. The overall risk of crop failure during the second planting period was 150 out of 4000 estimates, or 3.75%. Next, we will show a

comparison of the estimated data with the water requirements for the third planting period. The third planting period, from November 2022 to February 2023, was compared with simulated data. An overview of water requirements for planting period 3 can be seen in **Figure 8** below.

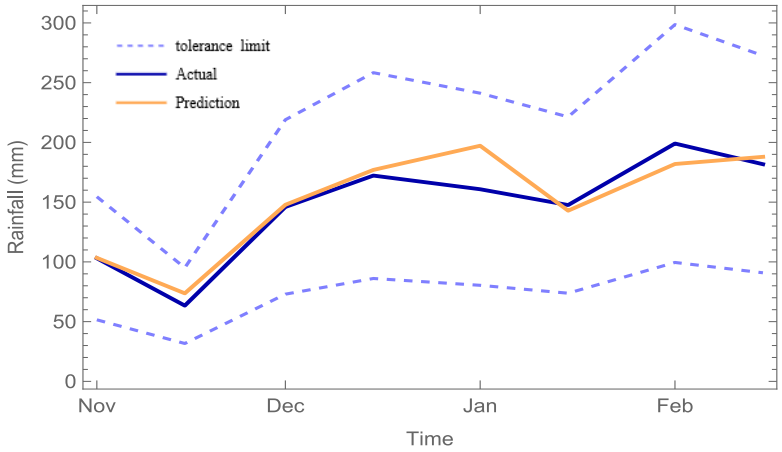

Figure 8. Comparison plot of water requirements and estimated the third planting period

Figure 8 illustrates a sample comparison from one of the datasets during the third planting period. The dashed lines indicate the upper and lower threshold values, while the blue line depicts the water requirement data for this period. The corresponding risk assessment for the third planting period is presented in **Table 5** below.

Table 5. Risk calculates for the third planting period

Period	Shortage of Water	Excess Water
Nov I	4	3
Nov II	2	20
Dec I	7	3
Dec II	9	0
Jan I	20	0
Jan II	8	0
Feb I	1	5
Feb II	2	13
Total	**53**	**44**

Based on **Table 5**, the exact number of points exceeding the threshold for each period can be identified. This number indicates the risk of crop failure during the third planting period. Crop failure during the third planting period due to water shortages was recorded at 53 out of a total of 4000 estimates, or 1.325%. Meanwhile, crop failure due to excess water was recorded at 44 out of a total of 4000 estimates, or 1.1%. The overall risk of crop failure during the third planting period was 97 out of 4000 estimates, or 2.425%.

To optimize agricultural yields that rely solely on rainfall, it is recommended to adopt planting patterns aligned with rainfall predictions to optimize agricultural yields that rely solely on rainfall. The first planting period (February–June) and the third planting period (November–February) are optimal, as rainfall is generally sufficient and the associated risk of crop failure remains below 3.5%. This aligns with risk analysis results, which show relatively low probabilities of failure compared to the second planting period (July–October), which is above 3.5% failure risk. These findings have direct implications for both farmers and policymakers. This information provides farmers with a data-driven basis for adjusting planting schedules or diversifying crop types to minimize losses. Alternative solutions for the second planting period can be implemented, namely planting secondary crops such as corn, soybeans, or peanuts, which require less water than crops like rice. For policymakers, it enables the development of targeted agricultural calendars and resource allocation strategies, such as prioritizing irrigation infrastructure during high-risk periods. Compared to findings in other regions, where crop failure risks can exceed 5–7% under similar rain-fed conditions, the risk levels observed in Tabanan, Bali, are relatively manageable, demonstrating the value of localized predictive modelling in supporting climate-resilient agriculture. This weather-prediction-based cropping pattern is expected to reduce the risk of crop failure while increasing the optimal planting frequency to three times a year. This will ensure more stable and sustainable farmer incomes throughout the year, while optimizing agricultural land use.

4. CONCLUSION

This study demonstrates the effectiveness of the Hidden Markov Model in capturing the stochastic nature of rainfall and its impact on agricultural risk. By modelling rainfall data using four hidden states representing dry, very dry, wet, and very wet conditions and assuming a Gamma distribution for rainfall intensity, the HMM could accurately reflect the temporal dynamics and variability of rainfall patterns. The results highlight the presence of seasonal structures in rainfall behaviour, where the wet and dry states tend to persist over time, aligning with the known climatic cycles in tropical regions. The model also provides valuable insight into the timing and duration of critical weather transitions, which are essential factors in agricultural planning. Based on these findings, several actionable strategies can be proposed. Policymakers should consider integrating HMM-based weather models into national agricultural planning tools, particularly for regions dependent on rain-fed farming. Adaptive cropping calendars that shift planting dates based on probabilistic rainfall forecasts could significantly reduce the risk of crop failure. Additionally, implementing early warning systems powered by real-time weather data could alert farmers to high-risk periods for drought or flooding.

HMM-based analysis presents a valuable approach to quantifying the uncertainties and risks of rainfall variability in agriculture. Its ability to uncover underlying weather regimes and their temporal dynamics makes it a practical tool for improving resilience in agricultural systems under changing climatic conditions. From an agricultural risk perspective, these results are highly significant. A delayed transition from the dry season to the wet season or a shortened duration of the wet season can disrupt planting schedules, lead to soil moisture deficits, and even result in crop failure. Additionally, extreme rainfall events within a short period (high rainfall bursts) observed under the wet state can increase the risk of flooding and soil erosion. In conclusion, the HMM model's ability to accurately predict rainfall patterns and associated agricultural risks provides a powerful foundation for climate-resilient farming. By enabling data-driven decision-making, it empowers farmers to optimize planting times, reduce vulnerability to extreme weather, and ultimately improve long-term productivity and sustainability in the face of climate uncertainty.

Future research is recommended in several directions to enhance the model's predictive power further. One is the development of hybrid models that integrate HMM with other probabilistic or deterministic frameworks, such as ARIMA-HMM or HMM coupled with physical climate models. Applying machine learning techniques such as recurrent neural networks (RNNs), LSTM, or ensemble methods could improve the model's ability to capture non-linear patterns and long-term dependencies in rainfall data. These approaches may offer better generalization, especially under rapidly changing climate scenarios. Exploring spatial extensions of HMM using geospatial data may also improve regional risk mapping and localized agricultural decision-making.

5. REFERENCES

[1] A. Ansari, Y. P. Lin, and H. S. Lur, "Evaluating and adapting climate change impacts on rice production in indonesia: A case study of the keduang subwatershed, Central Java," *Environments*, vol. 8, p. 117, Nov. 2021, doi: 10.3390/environments8110117.

[2] A. Berhane, "Climate Change and Variability Impacts on Agricultural Productivity and Food Security," *J. Climatol. Weather Forecast.*, vol. 6, no. 3, p. 240, 2018, doi: 10.4172/2332-2594.1000240.

[3] O. E. Olayide and T. Alabi, "Between rainfall and food poverty: Assessing vulnerability to climate change in an agricultural economy," *J. Clean. Prod.*, vol. 198, pp. 1–10, 2018, doi: 10.1016/j.jclepro.2018.06.221.

[4] G. C. Matchaya, G. Tadesse, and A. N. Kuteya, "Rainfall shocks and crop productivity in Zambia: Implication for agricultural water risk management," *Agric. Water Manag.*, vol. 269, p. 107648, 2022, doi: 10.1016/j.agwat.2022.107648.

[5] B. Praveen and P. Sharma, "A review of literature on climate change and its impacts on agriculture productivity," *J. Public Aff.*, vol. 19, no. 4, pp. 1–15, 2019, doi: 10.1002/pa.1960.

[6] R. Tirtalistyani, M. Murtiningrum, and R. S. Kanwar, "Indonesia Rice Irrigation System: Time for Innovation," *Sustainability*, vol. 14, no. 19, p. 12477, 2022, doi: 10.3390/su141912477.

[7] A. Ansari *et al.*, "Evaluating the effect of climate change on rice production in Indonesia using multimodelling approach," *Heliyon*, vol. 9, p. e19639, Sep. 2023, doi: 10.1016/j.heliyon.2023.e19639.

[8] S. Ruiz-Suarez, V. Leos-Barajas, and J. M. Morales, "Hidden Markov and Semi-Markov Models When and Why are These Models Useful for Classifying States in Time Series Data?," *J. Agric. Biol. Environ. Stat.*, vol. 27, no. 2, pp. 339–363, 2022, doi: 10.1007/s13253-021-00483-x.

[9] O. Stoner and T. Economou, "An advanced hidden Markov model for hourly rainfall time series," *Comput. Stat. Data Anal.*, vol. 152, p. 107045, 2020, doi: 10.1016/j.csda.2020.107045.

[10] M. Zhang, X. Jiang, Z. Fang, Y. Zeng, and K. Xu, "High-order Hidden Markov Model for trend prediction in financial time series," *Phys. A Stat. Mech. its Appl.*, vol. 517, pp. 1–12, 2019, doi: 10.1016/j.physa.2018.10.053.

[11] O. Nikdelfaz and S. Jalili, "Disease genes prediction by HMM based PU-learning using gene expression profiles," *J. Biomed. Inform.*, vol. 81, pp. 102–111, 2018, doi: 10.1016/j.jbi.2018.03.006.

[12] R. Glennie, T. Adam, V. Leos-Barajas, T. Michelot, T. Photopoulou, and B. T. McClintock, "Hidden Markov models: Pitfalls and opportunities in ecology," *Methods Ecol. Evol.*, vol. 14, no. 1, pp. 43–56, 2023, doi: 10.1111/2041-210X.13801.

[13] Q. Deng and D. Söffker, "A Review of HMM-Based Approaches of Driving Behaviors Recognition and Prediction," *IEEE Trans. Intell. Veh.*, vol. 7, no. 1, pp. 21–31, 2022, doi: 10.1109/TIV.2021.3065933.

[14] A. M. Nyongesa, G. Zeng, and V. Ongoma, "Non-homogeneous hidden Markov model for downscaling of short rains occurrence in Kenya," *Theor. Appl. Climatol.*, vol. 139, no. 3, pp. 1333–1347, 2020, doi: 10.1007/s00704-019-03016-2.

[15] L. Guo, Z. Jiang, and W. Chen, "Using a Hidden Markov Model to Analyze the Flood-Season Rainfall Pattern and Its Temporal Variation over East China," *J. Meteorol. Res.*, vol. 32, no. 3, pp. 410–420, 2018, doi: 10.1007/s13351-018-7107-9.

[16] H. Wang, T. Asefa, and A. Sarkar, "A novel non-homogeneous hidden Markov model for simulating and predicting monthly rainfall," *Theor. Appl. Climatol.*, vol. 143, no. 1, pp. 627–638, 2021, doi: 10.1007/s00704-020-03447-2.

[17] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989, doi: 10.1109/5.18626.

[18] M. L. Gámiz, N. Limnios, and M. C. Segovia-García, "The continuous-time hidden Markov model based on discretization. Properties of estimators and applications," *Stat. Inference Stoch. Process.*, vol. 26, no. 3, pp. 525–550, 2023, doi: 10.1007/s11203-023-09292-0.

[19] Y. Du, T. Wu, and V. Makis, "Parameter estimation and remaining useful life prediction of lubricating oil with HMM," *Wear*, vol. 376–377, pp. 1227–1233, 2017, doi: 10.1016/j.wear.2016.11.047.

[20] O. Zadehbagheri, M. R. Salehizadeh, S. V. Naghavi, M. Moattari, and B. Moshiri, "Novel Adaptive Hidden Markov Model Utilizing Expectation–Maximization Algorithm for Advanced Pipeline Leak Detection," *Modelling*, vol. 5, no. 4, pp. 1339–1364, 2024, doi: 10.3390/modelling5040069.

[21] B. Setiawaty, A. Y. Lestari, D. C. Lesmana, and N. K. K. Ardana, "Predicting the conditions of stock market using hidden Markov models," *AIP Conf. Proc.*, vol. 3201, no. 1, 2024, doi: 10.1063/5.0230595.

[22] C. Puerto-Santana, P. Larranaga, and C. Bielza, "Autoregressive Asymmetric Linear Gaussian Hidden Markov Models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4642–4658, 2022, doi: 10.1109/TPAMI.2021.3068799.

[23] M. El Annas, M. Ouzineb, and B. Benyacoub, "Hidden Markov Models Training Using Hybrid Baum Welch - Variable Neighborhood Search Algorithm," *Stat. Optim. Inf. Comput.*, vol. 10, no. 1, pp. 160–170, 2022, doi: 10.19139/soic-2310-5070-1213.

[24] R. Kuiper, "AIC-type Theory-Based Model Selection for Structural Equation Models," *Struct. Equ. Model.*, vol. 29, no. 1, pp. 151–158, 2022, doi: 10.1080/10705511.2020.1836967.

[25] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *Int. J. Forecast.*, vol. 32, no. 3, pp. 669–679, 2016, doi: 10.1016/j.ijforecast.2015.12.003.