WORLDLINK

JOURNAL OF PHYSICAL AND ENGINEERING SCIENCES

https://www.worldlinkpub.com/journals/journal-of-physical-and-engineering-sciences/

ISSN: 3141-5997

OPTIMAL ORDER SELECTION IN MARKOV CHAIN MODELING OF RAINFALL OCCURRENCE IN A TROPICAL CLIMATE: EVIDENCE FROM MAKURDI, NIGERIA

***Adah, V., Ogbuagu, E.M. and James, S.**
Department of Statistics, Joseph Sarwuan Tarka University, Makurdi, Nigeria
*Correspondence: adahvictoria14@yahoo.com, https://orcid.org/0000-0002-8984-7333

Received: 12 May 2026
Accepted for publication: 15 June 2026
Published: 01 July 2026

ABSTRACT

This study investigates rainfall occurrence dynamics in Makurdi, Benue State, Nigeria, using discrete-time Markov chain models. Daily rainfall data were transformed into a binary sequence of wet and dry states using a 0.1 mm threshold. First, second, and third-order Markov chain models were developed to capture rainfall persistence and temporal dependence. Model parameters were estimated using maximum likelihood methods, and model performance was evaluated using log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and chi-square goodness-of-fit tests. Results show strong persistence in rainfall states, particularly wet-to-wet and dry-to-dry transitions. The second-order model produced a log-likelihood of −3017.9245 (AIC = 6043.8489, BIC = 6074.4747), while the third-order model improved significantly with a log-likelihood of −2940.3531 (AIC = 5896.7062, BIC = 5957.9577). Model selection criteria indicate that the third-order Markov chain provides the best fit, capturing more complex rainfall memory effects. The chi-square test confirms excellent agreement between observed and expected rainfall frequencies. The study demonstrates that higher-order Markov chains effectively model rainfall dynamics in tropical climates and can support stochastic weather generation and hydrological planning.

Keywords: Rainfall, Transition Probability, AIC, BIC, Model Selection

1.0 INTRODUCTION

Rainfall variability is one of the most important climatic factors influencing agriculture, hydrology, water resource management, and environmental sustainability, particularly in tropical regions such as Nigeria. Daily rainfall occurrence exhibits stochastic behavior characterized by alternating wet and dry periods, making probabilistic models highly suitable for its analysis. Among the various stochastic approaches, Markov chain models have been widely used for modeling daily rainfall occurrence because they effectively capture the temporal dependence between consecutive rainfall events (Wilks, 1989; Agada and Adah, 2026)

In Markov chain rainfall models, rainfall occurrence is typically represented as binary states, namely wet and dry days, where the probability of rainfall on a particular day depends on the rainfall condition of preceding days. A major challenge in such modeling is determining the optimum order of the Markov chain. The order specifies how many previous days significantly influence the current rainfall state. While first-order Markov chains assume dependence only on the immediately preceding day, higher-order models allow rainfall occurrence to depend on multiple previous days. Selecting an inappropriate order may lead to underfitting or overfitting, thereby reducing predictive accuracy and limiting the reliability of rainfall simulations. The determination of optimum Markov chain order has attracted considerable attention in climatology and hydrology. Information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are commonly employed to identify the most suitable order by balancing model fit and complexity. However, the optimum order often varies across climatic regions due to differences in rainfall persistence, monsoon dynamics, and atmospheric circulation patterns.

Despite extensive studies on rainfall occurrence modeling globally, limited research has specifically examined the optimum order of Markov chain models for daily rainfall in Makurdi, Nigeria. Existing Nigerian studies have largely focused on generalized national rainfall patterns or a few selected stations without detailed emphasis on the Middle Belt region. Makurdi experiences a tropical wet-and-dry climate strongly influenced by the West African monsoon system, resulting in distinct wet and dry spell persistence that may not be adequately represented by conventional first-order models.

Furthermore, climatic variability and recent changes in rainfall characteristics associated with climate change may alter rainfall dependence structures over time. Therefore, there is a need to investigate the optimum order of Markov chain models for daily rainfall occurrence in Makurdi using contemporary rainfall data and objective model selection criteria such as AIC and BIC. Such a study will contribute to improved rainfall prediction, agricultural planning, drought monitoring, flood forecasting, and water resource management within the region (Adah and Agada, 2019).

Markov chain models have been extensively applied in daily rainfall occurrence modeling because rainfall sequences exhibit temporal dependence, where the probability of rainfall on a given day depends on previous days' weather conditions. A fundamental challenge in rainfall modeling is determining the optimum order of the Markov chain, that is, identifying the number of preceding days that significantly influence the current rainfall state. Several studies have investigated whether first-order, second-order, or higher-order Markov chains provide the most appropriate representation of rainfall dynamics. One of the earliest contributions was made by Richardson (1981), who applied first-order Markov chains to daily precipitation occurrence and demonstrated that rainfall persistence could be effectively captured using transition probabilities between wet and dry states. The study showed that first-order models are computationally simple and sufficiently accurate for many climatic regions. However, the author acknowledged that higher-order dependence may exist in areas with prolonged wet or dry spells.

Stern and Coe (1984) extended rainfall occurrence modeling by examining tropical rainfall data and observed that first-order models sometimes underestimate the duration of wet and dry sequences. Their work suggested that second-order Markov chains provide improved representation of rainfall persistence because the probability of rainfall may depend not only on the immediate previous day but also on rainfall conditions two days earlier. This finding was particularly important for tropical climates characterized by clustered rainfall events. Katz (1981) investigated statistical criteria for estimating the order of Markov chains and highlighted the importance of balancing model complexity with predictive performance. Using likelihood ratio tests and information criteria such as Akaike Information Criterion (AIC), the study showed that higher-order models improve fit but may introduce overfitting when sample sizes are limited. Katz concluded that optimal order selection should depend on both statistical adequacy and climatic characteristics of the rainfall process. Mimikou (1983) investigated seasonal-order Markov chains for daily precipitation occurrence in Greece and found that the optimum order varies seasonally. The study reported that higher-order chains tend to dominate during wet seasons, while lower-order models are adequate during dry periods. This demonstrated that rainfall persistence and dependence structures are influenced by climatic seasonality.

Wilks (1999) examined stochastic daily precipitation models across several climatic regions in the United States and concluded that first-order Markov chains are generally adequate for many temperate regions. However, the study also found that more complex alternatives perform better in regions exhibiting strong persistence of wet and dry spells. The author emphasized that the suitability of Markov chain order depends strongly on regional climatic characteristics. Research on rainfall occurrence modeling in Nigeria has similarly demonstrated variability in optimum order selection. A notable contribution was made in the study "The Optimum Order of a Markov Chain Model for Daily Rainfall in Nigeria. Jimoh, and Webster (1996), which analyzed rainfall data from five Nigerian stations using AIC and BIC criteria. The study found that optimum order varies across months and rainfall thresholds. While AIC frequently selected higher-order chains, BIC generally favored first-order or zero-order models. The authors concluded that rainfall dependence structures in Nigeria are spatially and seasonally variable.
Subsequent studies on stochastic rainfall generation have continued to support the use of Markov chains in modeling rainfall occurrence. Comparisons of traditional and modern stochastic rainfall models have shown that Markov chains remain among the most widely applied approaches due to their simplicity and ability to preserve rainfall persistence characteristics. Nevertheless, researchers have increasingly recognized limitations associated with fixed-order models, especially in regions with highly variable climatic conditions.

Recent hydrological studies have explored higher-order and hidden Markov approaches to better represent rainfall persistence. Thyer and Kuczera (2003) proposed hidden Markov models capable of capturing long-term persistence in rainfall sequences, while more recent studies have evaluated first-, second, and third-order chains using BIC to identify optimal rainfall occurrence structures. These studies generally report that higher-order models improve the representation of prolonged wet and dry spells but may suffer from parameter explosion and sparse transition matrices. Additionally, recent developments in rainfall occurrence modeling have incorporated Markov chain structures into climate bias-correction and forecasting systems. Parsons *et al.* (2026) demonstrated that incorporating first-order Markov chains into rainfall correction methods improves the representation of rainfall persistence, onset dates, and wet and dry spell characteristics in African rainfall datasets.

The literature shows that determining the optimum order of Markov chains remains an important issue in rainfall occurrence modeling. First-order models are computationally simple and widely applied, but second-order and higher-order chains often provide improved representation of rainfall persistence, particularly in tropical and monsoon climates. Despite these advances, there remains limited localized research focusing specifically on Makurdi, Nigeria. Most Nigerian studies have considered broader regional analyses without detailed investigation of rainfall dependence structures in the Middle Belt region. Consequently, there is a clear research gap regarding the identification of the optimum Markov chain order for daily rainfall occurrence in Makurdi using recent rainfall records and objective model selection techniques such as AIC and BIC.

Despite the comprehensive analysis by Jimoh and Webster (1996) on the optimum order of Markov chains for rainfall modeling in Nigeria, their study focused on a limited number of stations and did not specifically isolate Makurdi as a study location. Given that rainfall characteristics in Nigeria vary significantly across ecological zones, there remains a need to re-evaluate the optimum Markov chain order using more recent and location-specific datasets. In particular, Makurdi, located in the Guinea Savannah zone, experiences distinct rainfall dynamics influenced by monsoon variability and convective activity. This justifies a localized reassessment of whether first- or higher-order Markov chains best represent its daily rainfall structure using updated data and modern selection criteria such as AIC and BIC. Following Jimoh and Webster (1996), this study evaluates multiple Markov chain orders (1st, 2nd, and 3rd order) and selects the optimal order based on likelihood-based information criteria such as AIC and BIC, as well as rainfall persistence diagnostics. Since Makurdi is one of the five stations in the original study, this research is particularly relevant as it enables a reassessment of rainfall dynamics using more recent observations, given that the dataset employed by Jimoh and Webster (1996) is now over three decades old. Over this period, rainfall behaviour may have evolved due to climate variability, land-use change, and an observed increase in the frequency and intensity of extreme rainfall events. These potential shifts provide a strong justification for re-evaluating the appropriate Markov chain order specifically for Makurdi using updated datasets, thereby enhancing the empirical relevance and originality of the present study.

2.0 METHODOLOGY

2.1. Study Area

The study was conducted in Makurdi, Benue State, Nigeria, located within the Guinea Savannah ecological zone. The region is characterized by a tropical wet-and-dry climate with a unimodal rainfall pattern strongly influenced by the West African Monsoon system and the seasonal migration of the Inter-Tropical Discontinuity (ITD). Rainfall typically begins in April, peaks between July and September, and ends around October. The strong seasonality and variability in rainfall make Makurdi an appropriate location for stochastic rainfall modeling using Markov chain approaches (Wilks, 1999; Stern and Coe, 1984).

2.2. Data Description and Preprocessing

Daily rainfall data (in millimeters) were obtained from the National Aeronautics and Space Administration (NASA) from 1983 to 2025. Let the rainfall series be:

$$R = \{R_1, R_2, \dots, R_T\}$$

where (R_t) represents the observed rainfall amount on day (t), and (T) is the total number of observed days.

2.3 Rainfall State Discretization

Following Jimoh and Webster (1996), a rainfall amount of 0.1 mm was adopted as the threshold for classifying wet and dry days, ensuring consistency with standard stochastic rainfall modeling practices. To apply Markov chain modeling, continuous rainfall values were transformed into a binary stochastic process representing rainfall occurrence.

$$X_t = \begin{cases} 1 & if \quad R_t \geq 0.1mm \\ 0 & otherwise \end{cases} \tag{1}$$

This yields the sequence: $X = \{X_1, X_2, \dots, X_T\}$

This step follows standard stochastic weather generator methodology (Richardson, 1981; Katz, 1981).

2. 4 First-Order Markov Chain Model

A first-order Markov chain assumes that rainfall occurrence depends only on the immediately preceding day. The rainfall occurrence process is assumed to satisfy the Markov property:

$$P(X_t = j \mid X_{t-1} = i, X_{t-2}, \dots, X_{t-k}) = P(X_t = j \mid X_{t-1} = i) \tag{2}$$

The transition probabilities are defined as:

$$p_{ij} = P(X_{t+1} = j \mid X_t = i), \qquad i, j \in \{0,1\} \tag{3}$$

The transition matrix is:

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}, \qquad \sum_j p_{ij} = 1 \tag{4}$$

Estimation of Transition Probabilities

Let n_{ij} denote the observed number of transitions from state i to state j. Then:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i.}}, \qquad n_{i.} = \sum_j n_{ij} \tag{5}$$

This maximum likelihood estimator is widely used in rainfall Markov chain studies (Katz, 1981; Wilks, 1999).

2.5 Higher-Order Markov Chain Models

To examine memory effects beyond one day, higher-order Markov chains of order (k) were considered:

$$P(X_t \mid X_{t-1}, \dots, X_{t-k}) \tag{6}$$

For a second-order chain:

$$p_{hij} = P(X_t = j \mid X_{t-1} = i, X_{t-2} = h), \tag{7}$$

This increases the number of parameters from 4 (first-order) to 8 (second-order), reflecting the trade-off between model flexibility and parsimony (Hughes *et al*., 1999).

2.6 Likelihood Function

For a Markov chain of order (k), the likelihood of observing the sequence is:

$$L(\theta) = \prod_{t=k+1}^{T} P(X_t = j \mid X_{t-1}, X_{t-2}, \dots, X_{t-k}) \quad (8)$$

and the log-likelihood:

$$logL(\theta) = \sum_{t=k+1}^{T} P(X_t = j \mid X_{t-1}, X_{t-2}, \dots, X_{t-k}) \quad (9)$$

For an order-k Markov chain, the maximized likelihood is obtained by substituting the MLE estimates

into the log-likelihood function:

$$\widehat{p}_{hj} = \frac{n_{hj}}{n_{h.}}$$

$$(10) logL(\hat{\theta}) = \sum_{h\in S^k} \sum_{j\in s} n_{hj} \log\left(\frac{n_{hj}}{n_{h.}}\right) \quad (11)$$

where: n_{hj} is the number of transitions from history h to state j, $n_{h.}$ is the total number of occurrences of history h.
The maximized log-likelihood is then incorporated into the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

2.7 Model Order Selection Criteria

Akaike Information Criterion (AIC)

$$AIC = -2log(\hat{\theta}) + 2m \quad (12)$$

Bayesian Information Criterion (BIC)

$$BIC = -2log(\hat{\theta}) + mlog(T) \quad (13)$$

where:

- L: likelihood, m: number of parameters, and T: sample size.

since each conditional probability row sums to one.
Thus:

- first-order model: m=2
- second-order model: m=4
- third-order model: m=8, etc.

As the order increases, the likelihood generally improves because the model captures longer rainfall memory. However, AIC and BIC penalize this increased complexity to avoid overfitting.

The optimal order is:

$$k^* = ar\,g\min_{k}\{AIC_k, BIC_k\} \quad (14)$$

AIC prioritizes predictive accuracy, while BIC penalizes complexity more strongly (Akaike, 1974; Schwarz, 1978).

For a binary rainfall process with state space $S = \{0,1\}$.The number of free parameters in an order-k Markov chain is:

$$m = 2^k(2-1) = 2^k \quad (15)$$

2.8 Stationary Distribution (Model Stability Check)

To assess long-term rainfall behavior, the stationary distribution (π) was computed. The stationary distribution π\piπ satisfies:

$$\pi P = \pi, \quad \sum_i \pi_i = 1 \quad (16)$$

This provides the long-run proportion of wet and dry days and ensures ergodicity of the process (Norris, 1997).

2.9 Model Validation

Model performance was evaluated by comparing observed and simulated rainfall characteristics:
Goodness-of-fit was assessed using chi-square statistics:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \quad (17)$$

2.10 Software Implementation

All computations were implemented using Python. Transition matrices, likelihood functions, and information criteria were computed programmatically to ensure reproducibility. Matrix operations were used to derive stationary distributions and validate ergodicity.

3.0 RESULTS AND DISCUSSION

The results of this study are presented in the following tables and analyses. These results summarize the estimated Markov chain parameters, model selection criteria, and validation outcomes used to evaluate rainfall occurrence dynamics in the study area. The interpretation focuses on rainfall persistence, model performance, and the optimal order of the Markov chain for representing the observed rainfall process.

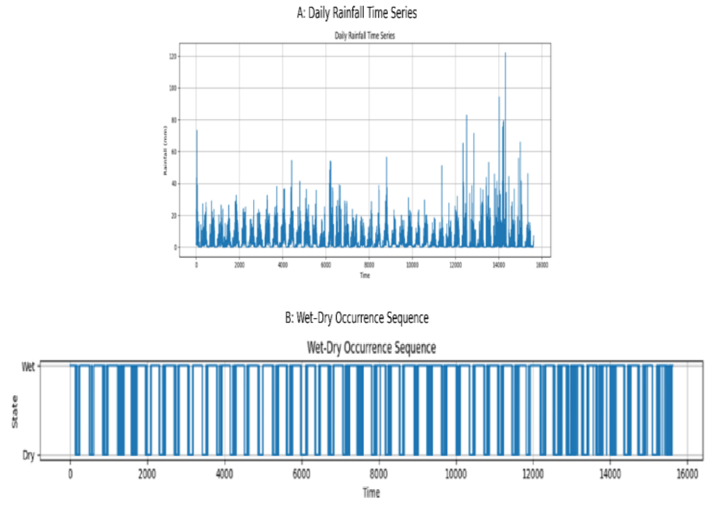

Figure 1: Rainfall Time series\Wet-Dry sequence

Figure 1, demonstrates that rainfall occurrence exhibits strong temporal dependence and persistence structures. The continuous rainfall series (A) shows clustered high-intensity events, while the corresponding binary sequence (B) confirms prolonged wet and dry spells. This justifies the application of Markov chain models for capturing rainfall dynamics and transition behavior.

Table 1: First-Order Markov Chain Transition Frequencies and Probabilities

Previous State (History)	Count to Dry (0)	Count to Wet (1)	(P(0))	(P(1))
0	4221	425	0.9085	0.0915
1	425	10550	0.0387	0.9613

The first-order Markov chain results indicate strong persistence in rainfall occurrence states. When the previous day was dry, the probability of remaining dry on the following day was very high, 0.9085, while the probability of transitioning from a dry day to a wet day was relatively low, 0.0915. Similarly, when the previous day was wet, the probability of remaining wet was extremely high, 0.9613, compared with a very small probability of transitioning to a dry day (0.0387.

These findings suggest substantial short-term persistence in rainfall occurrence within the study area. Wet days tend to cluster together during the rainy season, while dry spells also exhibit continuity. The high value of p_{11} particularly indicates strong wet-day persistence, which is characteristic of tropical rainfall systems influenced by monsoonal circulation.

Table 2: Second-Order Markov Chain Transition Frequencies and Probabilities

Previous States (History)	Count to Dry (0)	Count to Wet (1)	P(0)	P(1)
(0, 0)	3925	296	0.9299	0.0701
(0, 1)	99	326	0.2329	0.7671
(1, 0)	296	129	0.6965	0.3035
(1, 1)	326	10223	0.0309	0.9691

The occurrence of rainfall is dependent on the preceding two days, as shown by the second-order Markov chain. While successive wet states (1,1) demonstrated substantial wet persistence (0.9691), consecutive dry states (0,0) had a high probability of staying dry (0.9299). Moderate transition behavior was observed in mixed-state histories, suggesting that rainstorm events have memory effects lasting longer than 1 day. In general, the persistence of wet and dry spells is better represented by the second-order model than by the first-order model.

Table 3: Third-Order Markov Chain Transition Frequencies and Probabilities

Previous States (History)	Count to Dry (0)	Count to Wet (1)	P(0)	P(1)
(0, 0, 0)	3673	252	0.9358	0.0642
(0, 0, 1)	70	226	0.2365	0.7635
(0, 1, 0)	69	30	0.697	0.303
(0, 1, 1)	60	266	0.184	0.816
(1, 0, 0)	252	44	0.8514	0.1486
(1, 0, 1)	29	100	0.2248	0.7752
(1, 1, 0)	227	99	0.6963	0.3037
(1, 1, 1)	266	9956	0.026	0.974

The third-order Markov chain provides additional understanding of long-memory behavior in rainfall occurrence. Prolonged wet sequences (1,1,1) demonstrated an even greater persistence likelihood (0.9740), whereas extended sequences of dry days (0,0,0) demonstrated a very high probability of keeping dry (0.9358). These findings show that rainfall states are strongly clustered and dependent on time. Depending on recent rainfall sequences, different rainfall dynamics are revealed by the transition probabilities linked to mixed histories. For example, the likelihood of another wet day remained high (0.8160) following the sequence (0,1,1), whereas the process was more likely to remain dry (0.8514) following the sequence (1,0,0). These differences attest to the fact that past rainfall circumstances have a significant impact on the likelihood of future rainfall. AIC and BIC are required to choose the best model order because the third-order Markov chain captures more intricate rainfall persistence patterns than lower-order models, but its higher parameter count increases complexity and overfitting risk.

Table 4: Model Order Selection Using AIC and BIC

Order	Parameters	Log-Likelihood	AIC	BIC
1	2	-3219.86	6443.72	6459.033
2	4	-3017.92	6043.849	6074.475
3	8	-2940.35	5896.706	5957.958

The results show that model performance improves as the order increases, with the log-likelihood increasing (less negative values) from order 1 to order 3. Correspondingly, both AIC and BIC decrease across the models, indicating a better fit despite increasing complexity. The third-order Markov chain has the lowest AIC and BIC values, suggesting it provides the best balance between goodness-of-fit and model parsimony for the rainfall process. This finding is in agreement with Kuczera (2003), who established that a higher-order Markov chain model is optimum for rainfall occurrence. Though the BIC in Jimoh and Webster (1996) favored first-order or zero-order models, the AIC frequently selected higher-order chains.

Table 5: Chi-Square Goodness-of-Fit Validation

Category	Observed Frequency	Expected Frequency
State 0 (Dry)	4646	4646.3
State 1 (Wet)	10976	10975.7

Chi-square(χ^2) = 0.000

With a statistic of 0.0000, the chi-square test reveals nearly perfect agreement between observed and anticipated frequencies. This suggests that the marginal rainfall occurrence probabilities are closely reproduced by the fitted Markov chain model. As a result, there is no indication that the observed and simulated rainfall states differ significantly, indicating a very high goodness-of-fit.

4.0 CONCLUSION

This study applied Markov chain models of varying orders to analyze rainfall occurrence in Makurdi, Nigeria. The results revealed strong temporal dependence in rainfall, with clear persistence of both wet and dry spells. Higher-order models improved the representation of rainfall dynamics, with the third-order Markov chain providing the best overall fit based on AIC, BIC, and likelihood values. Validation results confirmed that the model reproduces observed rainfall frequencies with high accuracy. The findings show that rainfall occurrence in the study area is not random but strongly memory-dependent, making Markov chain approaches suitable for rainfall modeling and simulation. The third-order model is recommended for applications requiring more detailed representation of rainfall persistence, such as hydrological forecasting and climate impact studies.

REFERENCES

Adah, V., and Agada, P. O. (2019). A probabilistic modelling approach to quantify and assess the impact of climate change on the growth of yam in Makurdi, Nigeria. *FUDMA Journal of Sciences*, 3(4), 352–359.

Agada, I.O. and Adah, V. (2026) Characterization of Rainfall Regimes in the Middle Belt States of Nigeria Using a 3-State Hidden Markov Model. *Asian Journal of Physical and Chemical Sciences*, 14(1), 153-169

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Hughes, J. P., Guttorp, P., and Charles, S. P. (1999). A Markov chain approach to stochastic rainfall modeling. *Journal of Hydrology*, 223(1–2), 1–17.

Jimoh, O. D., and Webster, P. (1996). The optimum order of a Markov chain model for daily rainfall in Nigeria. *Journal of Hydrology*, 185(1–4), 45–69. https://doi.org/10.1016/S0022-1694(96)03015-6

Katz, R. W. (1981). On some criteria for estimating the order of a Markov chain. *Technometrics*, 23(3), 243–249. https://doi.org/10.2307/1267787

Norris, J. R. (1997). *Markov Chains*. Cambridge University Press.

Parsons, D., Stern, D., Sylla, M. B., Musyoka, J., Bagiliko, J., Clements, L., Mupuro, J., and Ndanguza, D. (2026). Improving bias correction methods for daily rainfall using a Markov chain approach.

Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resources Research*, 17(1), 182–190.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.

Stern, R. D., and Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society Series A*, 147(1), 1–34. https://doi.org/10.2307/2981736

Thyer, M., and Kuczera, G. (2003). A hidden Markov model for modelling long-term persistence in multi-site rainfall time series. 10.1016/S0022-1694(02)00411-0

Wilks, D. S. (1989). Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research*, 25(6), 1429–1439.

Wilks, D. S. (1999). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, 93(3), 153–169.