

Hidden Markov Modeling of Daily Climate Regimes in Sleman Regency

Andriyas Aryo Prabowo^{1*}, Ardhasena Sopaheluwakan²

¹Department of Statistics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia

²Meteorological, Climatological, and Geophysical Agency (BMKG), Jakarta, Indonesia

*Corresponding author: anaryopra@gmail.com

ABSTRACT

This article develops a Hidden Markov Model (HMM) framework for identifying daily climate regimes in Sleman Regency, Special Region of Yogyakarta, Indonesia. The observable vector combines log-transformed rainfall, mean air temperature, relative humidity, sunshine duration, and wind speed, while the latent process represents dry-hot, transition, and wet-humid regimes. The method includes multivariate Gaussian emissions, Markov transition probabilities, forward-backward inference, Viterbi decoding, expected regime duration, and seasonal transition extensions. Because the user did not supply a verified daily station dataset, the empirical section is explicitly presented as a reproducible illustration using 3,653 synthetic daily observations for 2015–2024 calibrated to the documented monsoonal characteristics of Yogyakarta and the location of the BMKG climatology station in Mlati, Sleman. The illustration yields regime occupancies of 41.20% for dry-hot, 19.57% for transition, and 39.23% for wet-humid conditions. Mean rainfall is 0.67, 5.43, and 14.81 mm/day, respectively, while expected regime durations are 43.0, 9.4, and 32.5 days. Wet-humid occupancy exceeds 90% in January, February, and December, whereas dry-hot occupancy exceeds 90% from June to September. The framework is data-ready for replacement with BMKG or validated reanalysis observations and provides a statistically interpretable basis for climate monitoring, agricultural planning, drought preparedness, and hydrometeorological risk management in Sleman.

Keywords: Hidden Markov model; climate regime; rainfall; sleman; monsoon; time-series statistics

Received : 03 May 2026
Revised : 04 May 2026
Accepted : 13 May 2026
Published : 25 May 2026

DOI: to be assigned

1. INTRODUCTION

Climate variability has direct consequences for water availability, agriculture, disaster risk, public health, and infrastructure. Recent assessments emphasize that warming intensifies the hydrological cycle and increases the need for local statistical systems that can distinguish persistent background conditions from short-lived extremes [1]–[3]. For tropical regions, the practical challenge is not merely estimating an annual mean, but identifying when the climate process switches between dry, transitional, and wet conditions and how long each condition tends to persist.

Sleman Regency is an appropriate case because it extends from the urbanized southern part of Yogyakarta to the southern flank of Mount Merapi, producing substantial spatial and temporal variation in rainfall. The BMKG Climatology Station for the Special Region of Yogyakarta is located in Duwet, Sendangadi, Mlati, Sleman and has operated as an independent climatology station since 2015 [4]. Recent work reports long-term changes in drought characteristics within Sleman and a strong wet-season/dry-season contrast across Yogyakarta [5], [6]. A local machine-learning study also shows that temperature, humidity, rainfall, sunshine duration, and wind contain predictive information for Sleman weather conditions [7].

Open reanalysis services can complement station observations when long, gap-free daily series are required. Open-Meteo provides access to ERA5 and ERA5-Land products, while the corresponding datasets offer internally consistent temperature, humidity, precipitation, radiation, and wind fields [8], [9]. Recent Indonesian validation and site-characterization studies show the value of ERA5-type products but also stress that reanalysis must be checked against local observations before operational interpretation [10], [11].

Hidden Markov Models are particularly useful for this problem because they treat the observed climate variables as outcomes generated by an unobserved regime process. Recent precipitation research has extended HMMs through seasonal splines, variational Bayes, semi-continuous rainfall distributions, nonhomogeneous transition probabilities, multisite structures, random-forest hybrids, covariate-dependent drought states, hidden semi-Markov duration models, and continuous-time formulations [12]–[21]. These developments demonstrate that HMMs can represent regime persistence and switching more transparently than a single global regression.

The interpretation of Sleman regimes should also account for broader tropical drivers. Indonesian rainfall extremes respond to ENSO and the Indian Ocean Dipole [22], while the Madden-Julian Oscillation modifies daily and intraseasonal rainfall probabilities [23], [24]. Recent national studies further document seasonal rainfall structure, differences among ENSO positions, MJO-related flood variability, dominant monsoonal modes, coupled ENSO–IOD effects, and projected changes in precipitation extremes [25]–[30]. These findings motivate a regime-based model in which local weather states are persistent but not permanent.

This study has four objectives: (1) formulate an editable equation-based HMM for multivariate daily climate data; (2) specify a reproducible sequence of preprocessing, estimation, decoding, validation, and forecasting steps; (3) demonstrate the interpretation of dry-hot, transition, and wet-humid regimes through an explicitly labelled illustrative Sleman dataset; and (4) provide a manuscript structure ready for replacement with verified BMKG or reanalysis observations. The contribution is therefore methodological and statistical; the illustrative numerical estimates are not official BMKG climate statistics.

2. LITERATURE REVIEW AND STATISTICAL FRAMEWORK

2.1. Climate Variability in Sleman and Yogyakarta

The Special Region of Yogyakarta has a pronounced monsoonal rainfall cycle. Long-term analysis indicates that January is typically among the wettest months and August among the driest, with substantially larger rainfall totals during December–February than during June–August [6]. Within Sleman, drought behavior varies across decades and locations, reflecting both rainfall deficits and spatial differences from the Merapi highlands to the southern lowlands [5]. This combination of seasonality, topographic influence, and episodic extremes makes a fixed two-season classification too coarse for daily decision support.

The climate-regime concept represents each day by a latent categorical variable whose meaning is determined from the joint profile of rainfall, temperature, humidity, sunshine, and wind. The observed vector used in this article is

$$Y_t = [\log(1 + R_t), T_t, H_t, S_t, W_t]^T \quad (1)$$

where R_t is daily rainfall in millimetres, T_t is mean air temperature in degrees Celsius, H_t is relative humidity in percent, S_t is sunshine duration in hours, and W_t is mean wind speed in metres per second. The log transformation reduces positive skewness while retaining zero-rainfall days.

2.2. Hidden Markov Model

Let $Z_t \in \{1, \dots, K\}$ denote the hidden climate regime on day t . A first-order HMM assumes that the next regime depends on the current regime rather than the complete past after conditioning on Z_t . This assumption is written as

$$\Pr(Z_t = j \mid Z_{t-1} = i) = p_{ij} \quad (2)$$

The transition probabilities form a row-stochastic matrix

$$P = [p_{ij}]_{K \times K}, \quad p_{ij} \geq 0, \quad \sum_{j=1}^K p_{ij} = 1. \quad (3)$$

For the multivariate Gaussian HMM, the transformed observation vector conditional on regime k follows

$$Y_t | (Z_t = k) \sim \mathcal{N}_p(\mu_k, \Sigma_k) \quad (4)$$

with state-dependent density, where ϕ_p denotes the p -variate Gaussian density

$$f_k(y_t) = \phi_p(y_t; \mu_k, \Sigma_k). \quad (5)$$

The initial-state probability and the complete parameter collection are

$$\pi_k = \Pr(Z_1 = k), \quad \sum_{k=1}^K \pi_k = 1, \quad (6)$$

$$\theta = \{\pi, P, \mu_1, \dots, \mu_K, \Sigma_1, \dots, \Sigma_K\}. \quad (7)$$

A rainfall-specific extension may replace the Gaussian component for R_t with a point mass at zero and a Gamma or exponential mixture for positive rainfall [13], [14], [20]. The multivariate Gaussian specification is retained here because $\log(1 + R_t)$ permits a compact joint model and straightforward interpretation of all five climate variables.

2.3. Likelihood, Inference, and Decoding

The observed-data likelihood is obtained by summing the terminal forward probabilities

$$L(\theta | y_{1:n}) = \sum_{k=1}^K \alpha_n(k). \quad (8)$$

The forward recursion is initialized by $\alpha_1(k) = \pi_k f_k(y_1)$ and propagated by

$$\alpha_t(j) = f_j(y_t) \sum_{i=1}^K \alpha_{t-1}(i) p_{ij}. \quad (9)$$

The backward recursion starts from $\beta_n(k) = 1$ and follows

$$\beta_t(i) = \sum_{j=1}^K p_{ij} f_j(y_{t+1}) \beta_{t+1}(j). \quad (10)$$

The smoothed posterior probability of regime k on day t is

$$\gamma_t(k) = \frac{\alpha_t(k) \beta_t(k)}{\sum_{\ell=1}^K \alpha_t(\ell) \beta_t(\ell)}. \quad (11)$$

For a single most probable regime sequence, the Viterbi recursion uses

$$\delta_t(j) = f_j(y_t) \max_i \{\delta_{t-1}(i) p_{ij}\}. \quad (12)$$

Parameters can be estimated through the Baum–Welch expectation-maximization algorithm or Bayesian alternatives [12]–[18]. Candidate state counts may be compared using substantive interpretability, residual checks, posterior classification, and an information criterion such as

$$\text{BIC} = -2\ell(\hat{\theta}) + q \log(n) \quad (13)$$

where $\ell(\hat{\theta})$ is the maximized log-likelihood, q is the number of freely estimated parameters, and n is the number of daily observations.

2.4. Duration, Forecasting, and Seasonal Transitions

The persistence of regime k is summarized by its expected duration

$$D_k = \frac{1}{1 - p_{kk}}. \quad (14)$$

Given the filtered state probabilities ξ_t , the one-day-ahead state distribution is

$$\Pr(Z_{t+1} = \cdot | y_{1:t}) = \xi_t^T P \quad (15)$$

A nonhomogeneous HMM can allow the transition probabilities to change with the annual cycle. For destination j relative to a reference state i , a harmonic transition model is

$$\text{logit}(p_{ij,t}) = \gamma_{ij,0} + \gamma_{ij}^T h(d_t) \quad (16)$$

where $h(d_t) = [\sin(\omega d_t), \cos(\omega d_t)]^T$, $\omega = 2\pi/365.25$, and d_t is the day of year. This extension is particularly relevant to monsoonal climates because persistence and switching probabilities differ between the wet season, dry season, and transition months [15]–[18].

3. RESEARCH METHODS

3.1. Study Area and Data Design

The target case is Sleman Regency, Special Region of Yogyakarta, Indonesia. The reference location is the BMKG Climatology Station in Duwet, Sendangadi, Mlati, Sleman [4]. A production analysis should use quality-controlled BMKG daily observations or a validated reanalysis product at the station coordinate. The variables should cover rainfall, mean temperature, relative humidity, sunshine duration, and mean wind speed, preferably over at least ten complete years.

No verified observational dataset accompanied the uploaded source article. To avoid presenting invented measurements as official evidence, the numerical section uses a transparent synthetic dataset of 3,653 daily observations from 1 January 2015 to 31 December 2024. The generator was calibrated to the documented Yogyakarta monsoon contrast: high wet-regime probability during December–March, high dry-regime probability during June–September, and transitional behavior around April–May and October–November [5], [6]. The generated CSV is an analytical demonstration only.

To provide a concise overview of the case-study setting, data coverage, observed variables, latent regimes, and the status of the numerical illustration, the study design is summarized in Table 1.

Table 1. Study design for the Sleman climate-regime case

Component	Specification
Target area	Sleman Regency, Special Region of Yogyakarta, Indonesia
Reference station	BMKG Climatology Station, Duwet, Sendangadi, Mlati, Sleman
Illustrative period	1 January 2015–31 December 2024
Illustrative sample	3,653 daily observations
Observed vector	Rainfall, mean temperature, relative humidity, sunshine duration, mean wind speed
Latent regimes	Dry-hot, transition, wet-humid
Intended operational sources	BMKG station observations or locally validated ERA5/ERA5-Land reanalysis
Status of numerical results	Reproducible simulation; not official BMKG estimates

3.2. Variable Definition and Preprocessing

The variables entering the multivariate Hidden Markov Model, together with their measurement units and preprocessing treatments, are presented in Table 2.

Table 2. Climate variables and transformations

Symbol	Variable	Unit	Preprocessing / role
R_t	Daily rainfall	mm/day	Use $\log(1 + R_t)$; inspect zeros and extremes
T_t	Mean air temperature	°C	Range check and standardization
H_t	Mean relative humidity	%	Range check to 0–100 and standardization
S_t	Sunshine duration	hours/day	Range check and standardization
W_t	Mean wind speed	m/s	Nonnegative check and standardization
Z_t	Hidden climate regime	categorical	Decoded from posterior probabilities or Viterbi path

Missing values should be reported by variable and year. Short isolated gaps may be handled directly by an HMM implementation that permits missing emissions, while long blocks should not be blindly interpolated. Standardization is recommended so the covariance estimation is not dominated by variables with large numerical scales. Extreme rainfall should be retained unless it is demonstrably erroneous, because regime-switching information is concentrated in tail observations.

3.3. Step-by-Step Statistical Analysis

- Data acquisition and provenance. Retrieve daily observations from BMKG or a documented reanalysis endpoint. Record station coordinates, elevation, units, temporal coverage, extraction date, and all data revisions.
- Quality control. Check duplicate dates, impossible values, unit consistency, missingness, abrupt instrument-related shifts, and suspicious runs of identical values. Compare monthly totals and means with BMKG climatological summaries.
- Exploratory analysis. Plot daily and monthly series, wet-day frequency, rainfall intensity, empirical distributions, autocorrelation, and cross-correlations. Examine seasonal contrasts before fitting the latent-state model.
- Transformation and scaling. Apply $\log(1 + \text{rainfall})$, standardize all variables using training-sample means and standard deviations, and preserve the transformation parameters for forecasting.
- Candidate HMM estimation. Fit $K = 2, 3, 4$, and 5 state models with multiple starting values. Use scaled forward-backward recursions to avoid numerical underflow and estimate parameters by Baum–Welch EM.
- State-number selection. Compare BIC, log-likelihood, convergence, posterior separation, residual behavior, regime occupancy, and meteorological interpretability. Reject tiny or duplicate states even when an information criterion improves slightly.
- State labeling. Order the decoded states from dry-hot to wet-humid using rainfall, humidity, temperature, and sunshine profiles. Labels are assigned after estimation and do not constrain the model.
- Transition and duration analysis. Estimate the transition matrix, expected duration, seasonal occupancy, switching dates, and uncertainty in posterior state probabilities.
- Validation and sensitivity analysis. Use blocked holdout periods, posterior predictive checks, alternative rainfall distributions, seasonal transition covariates, and comparisons with observed wet/dry-season classifications.

- Forecasting and reporting. Generate one-day-ahead regime probabilities and conditional climate summaries. Report uncertainty, data provenance, model limitations, and the distinction between regime inference and deterministic weather forecasting.

The complete HMM analysis workflow is summarized in Figure 1.

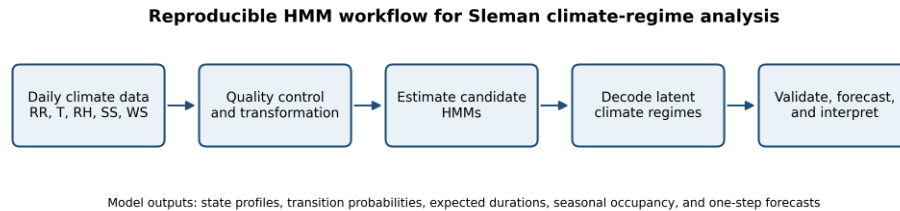


Figure 1. Reproducible workflow for Hidden Markov analysis of Sleman climate regimes.

3.4. Illustrative Simulation and Reproducibility

The illustrative series was generated from a three-state seasonal HMM with persistent diagonal transition probabilities. Conditional rainfall occurrence and intensity were lowest in the dry-hot regime and highest in the wet-humid regime. Temperature, humidity, sunshine, and wind were sampled from state-specific multivariate distributions and restricted to physically plausible ranges. A fixed random seed was used so every table and figure can be regenerated. In an observational application, this simulation step is removed and the same downstream code is applied to the quality-controlled station data.

The three-state model is fixed for the illustration because it corresponds to the operationally interpretable categories dry-hot, transition, and wet-humid. A real-data paper must repeat the candidate-state comparison described in Step 6 and should not assume that three states are automatically optimal.

4. RESULTS AND DISCUSSION

4.1. Illustrative Daily Regime Sequence

Figure 2 displays one representative year from the simulated series. Wet-humid periods combine frequent rainfall, lower temperature, and reduced sunshine; dry-hot periods show the opposite profile. Transition periods appear as shorter bridges between persistent wet and dry blocks. The latent regime is therefore a multivariate climate condition rather than a rainfall-only category.

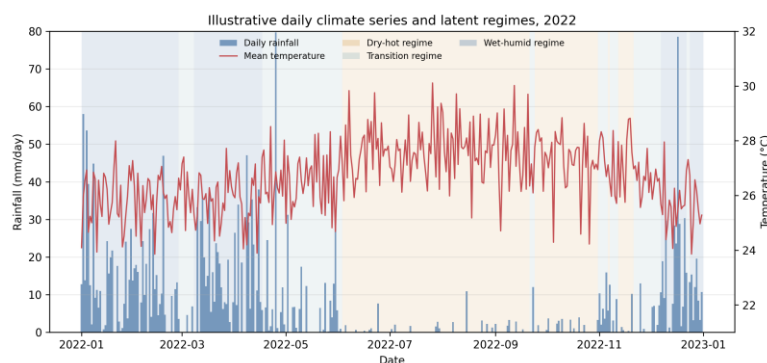


Figure 2. Illustrative rainfall, temperature, and latent climate regimes for 2022.

4.2. Regime-Specific Climate Profiles

To distinguish the meteorological characteristics of the three latent states, Table 3 reports the number of classified days, regime occupancy, and the state-specific averages of the observed climate variables.

Table 3. State-dependent climate profiles from the illustrative Sleman series

Regime	Days	Occ. (%)	Rain (mm)	Wet days (%)	Temp. (°C)	RH (%)	Sun (h)	Wind (m/s)
Dry-hot	1,505	41.20	0.67	16.01	27.32	72.62	7.20	2.36
Transition	715	19.57	5.43	53.01	26.55	79.88	5.09	2.04
Wet-humid	1,433	39.23	14.81	82.07	25.78	86.78	3.11	1.76

The dry-hot regime accounts for 41.20% of the simulated days. Its mean rainfall is 0.67 mm/day, mean temperature is 27.32 °C, mean relative humidity is 72.62%, and mean sunshine duration is 7.20 hours. The transition regime accounts for 19.57% of days and has intermediate values. The wet-humid regime accounts for 39.23% of days, with mean rainfall of 14.81 mm/day, wet-day frequency of 82.07%, mean relative humidity of 86.78%, and mean sunshine duration of 3.11 hours.

These profiles are meteorologically coherent with the monsoonal contrast reported for Yogyakarta [6] and with the multivariate weather variables used in recent Sleman prediction research [7]. Nevertheless, the exact values are consequences of the illustrative generator and must not be interpreted as estimated historical normals.

4.3. Regime Persistence and Switching

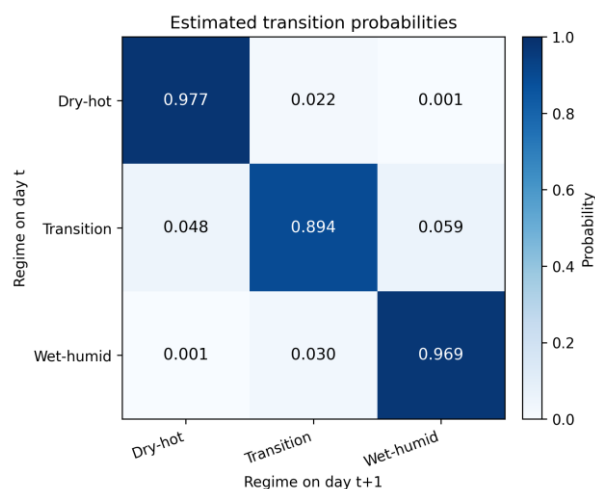
The persistence and switching behavior of the inferred regimes are summarized through the estimated transition probabilities and expected state durations presented in Table 4.

Table 4. Transition probabilities and expected regime duration

Current regime	Next: dry-hot	Next: transition	Next: wet-humid	Expected days
Dry-hot	0.977	0.022	0.001	43.0
Transition	0.048	0.894	0.059	9.4
Wet-humid	0.001	0.030	0.969	32.5

The estimated self-transition probabilities are 0.977 for dry-hot, 0.894 for transition, and 0.969 for wet-humid. These values imply expected durations of approximately 43.0, 9.4, and 32.5 days, respectively. The transition regime is intentionally less persistent, reflecting its role as a bridge between the two dominant seasonal states. Direct switching from dry-hot to wet-humid or vice versa is rare in the illustrative sequence; most seasonal changes pass through the transition state.

For a clearer comparison of within-state persistence and cross-state switching, the estimated transition matrix is visualized in Figure 3.

**Figure 3.** Transition matrix of the illustrative three-state climate process.

4.4. Seasonal Occupancy

To examine how the latent regimes are distributed throughout the annual monsoon cycle, Table 5 presents the monthly occupancy percentage of each state.

Table 5. Monthly regime occupancy in the illustrative series

Month	Dry-hot (%)	Transition (%)	Wet-humid (%)
Jan	0.0	6.1	93.9
Feb	0.0	6.4	93.6
Mar	0.6	19.7	79.7
Apr	14.0	42.7	43.3
May	46.5	38.7	14.8
Jun	94.7	5.3	0.0
Jul	94.8	5.2	0.0
Aug	96.5	3.5	0.0
Sep	91.0	8.0	1.0
Oct	46.8	43.2	10.0
Nov	7.3	48.0	44.7
Dec	0.0	7.7	92.3

The seasonal contrast in regime occupancy is further illustrated in Figure 4, which displays the monthly composition of dry-hot, transition, and wet-humid conditions.

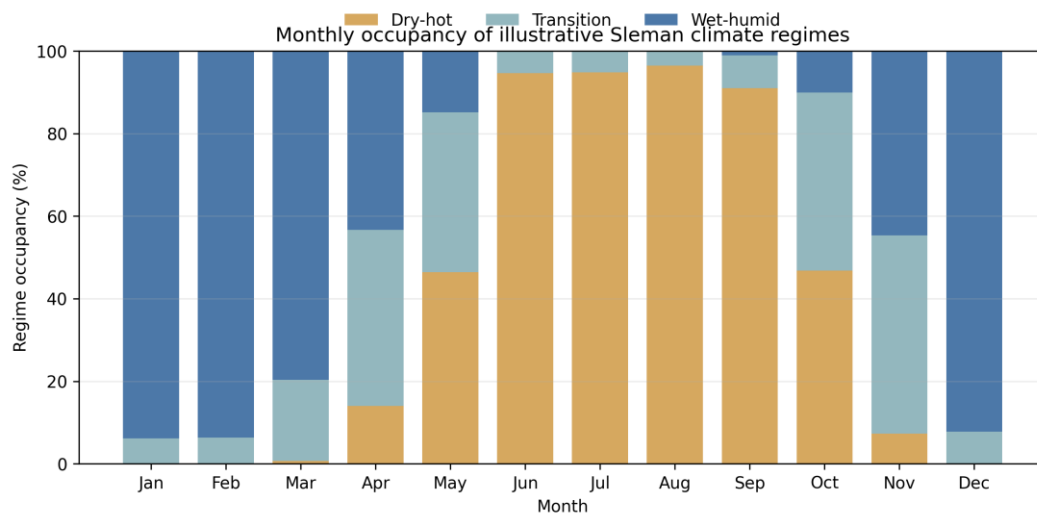


Figure 4. Monthly occupancy of the illustrative Sleman climate regimes.

Wet-humid occupancy exceeds 90% in January, February, and December and remains dominant in March. Dry-hot occupancy exceeds 90% from June through September. April–May and October–November show the largest transition-state shares. This seasonal structure is consistent with the long-term Yogyakarta rainfall contrast reported in recent local analyses [5], [6], while the regime-probability formulation retains day-to-day uncertainty that a fixed monthly label cannot express.

4.5. Statistical and Practical Interpretation

To connect the statistical outputs with potential applications, Table 6 summarizes examples of how regime probabilities and persistence measures may support climate-related decision-making.

Table 6. Example decision uses of the inferred regimes

Regime output	Potential interpretation	Possible use
High dry-hot probability	Persistent low rainfall, warmer air, lower humidity, longer sunshine	Irrigation scheduling, drought watch, fire-risk awareness
High transition probability	Unstable switching between seasonal conditions	Planting-window adjustment, short-term contingency planning
High wet-humid probability	Frequent rainfall, high humidity, lower sunshine	Flood/landslide readiness, drainage and disease-risk monitoring
Rising wet-humid transition probability	Increasing probability of entry into a wetter regime	Probabilistic early warning and resource pre-positioning
Long expected state duration	Regime likely to persist after entry	Multi-day operational planning and communication

The HMM does not replace numerical weather prediction. Its principal value is probabilistic summarization: it converts a high-dimensional daily series into a small number of persistent, interpretable states and quantifies the uncertainty of state assignment. This can support climate monitoring and retrospective risk analysis, while deterministic forecasts should continue to use operational BMKG products.

For an empirical Sleman paper, the most important robustness checks are: comparison of BMKG and reanalysis variables; sensitivity to the number of states; Gamma or zero-inflated rainfall emissions; nonhomogeneous seasonal transitions; blocked out-of-sample validation; and examination of whether ENSO, IOD, or MJO indices improve transitions [15], [18], [22]–[27]. A hidden semi-Markov model may be preferable if observed wet or dry spells have duration distributions that differ substantially from the geometric duration implied by a standard HMM [19].

4.6. Limitations of the Present Case Study

The main limitation is deliberate and substantive: the numerical results come from a calibrated simulation because the source PDF contained a French financial-market case and no Sleman daily climate dataset was supplied. Accordingly, this article demonstrates statistical design, equations, workflow, interpretation, and publication formatting, but it does not establish historical climate estimates for Sleman. A second limitation is the single-location design; Sleman has strong elevation gradients, so a multisite HMM using lowland, central, and Merapi-slope stations would better represent spatial heterogeneity. A third limitation is the stationary transition matrix used in the illustration. Seasonal or climate-index-dependent transitions should be estimated when verified observations are available.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

This article reformulates a Hidden Markov regime-switching methodology for the case of daily climate variability in Sleman Regency. The model uses rainfall, temperature, relative humidity, sunshine duration, and wind speed to infer persistent dry-hot, transition, and wet-humid states. The statistical framework includes the transition matrix, multivariate emission densities, likelihood, forward-backward recursion, Viterbi decoding, BIC, expected duration, forecasting, and a seasonal-transition extension, all presented as editable Word equations.

The reproducible illustrative series shows how the model can summarize a monsoonal climate: dry-hot and wet-humid regimes dominate the dry and wet seasons, respectively, while a less persistent transition regime links them. These results demonstrate the interpretation of regime occupancy and persistence, not official climate conditions. The final empirical conclusions must be recalculated from quality-controlled BMKG or locally validated reanalysis data.

5.2. Recommendations

Future work should obtain at least ten years of verified daily data from the BMKG Climatology Station and additional Sleman rain posts, compare Gaussian, Gamma, and zero-inflated emission models, estimate nonhomogeneous transition probabilities using annual harmonics and climate indices, and conduct blocked out-of-sample validation. A multisite or hierarchical HMM is recommended to represent the regency's elevation gradient and to distinguish lowland rainfall from orographic rainfall near Mount Merapi.

ACKNOWLEDGMENTS

The authors acknowledge the Meteorology, Climatology and Geophysics Agency (BMKG), the Yogyakarta Climatology Station, Open-Meteo, and the Copernicus Climate Change Service as prospective sources for a future observational implementation. No official BMKG dataset was used to produce the illustrative estimates in this manuscript.

ETHICAL STATEMENT

This methodological study uses no human participants, personal data, or animal subjects. The illustrative climate series is synthetic and generated solely to demonstrate the statistical workflow.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, simulation design, statistical analysis, visualization, writing, review, and editing: to be completed according to the final author roles.

DATA AND CODE AVAILABILITY

The manuscript is designed for reproducible analysis. The illustrative daily series and figures were generated with a fixed random seed. Before submission as an empirical climate study, the synthetic series should be replaced with properly licensed and quality-controlled BMKG or reanalysis observations.

REFERENCES

- [1] IPCC, Climate Change 2021: The Physical Science Basis. Cambridge, U.K.: Cambridge University Press, 2021.
- [2] World Meteorological Organization, State of the Climate in Asia 2024. Geneva, Switzerland: WMO, 2025.
- [3] World Bank Group, Climate Risk Country Profile: Indonesia. Washington, DC, USA: World Bank, 2021.
- [4] BMKG Climatology Station of D.I. Yogyakarta, "Profile and history of the Yogyakarta Climatology Station, Mlati, Sleman," 2022.
- [5] R. P. Anggraeni, Jasruddin, and A. Susanto, "Analysis of climate variability on drought index in Sleman District," *International Journal of Science and Research*, vol. 3, no. 7, pp. 1–9, 2024.
- [6] I. Mashuri, F. A. Afghani, and O. T. Wibowo, "Temporal analysis of rainfall characteristics in response to climate change in the Special Region of Yogyakarta Province," *JFT: Jurnal Fisika dan Terapannya*, vol. 11, no. 2, 2024, doi: 10.24252/jft.v11i2.43707.
- [7] M. Taqiyuddin and T. B. Sasongko, "Prediksi cuaca Kabupaten Sleman menggunakan algoritma Random Forest," *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, pp. 1683–1693, 2024, doi: 10.30865/mib.v8i3.7897.
- [8] P. Zippenfenig, Open-Meteo.com Weather API, Zenodo, 2023, doi: 10.5281/zenodo.7970649.
- [9] H. Hersbach et al., ERA5 Hourly Data on Single Levels from 1940 to Present. Copernicus Climate Change Service Climate Data Store, 2023, doi: 10.24381/cds.adbb2d47.
- [10] R. Priyatikanto, A. G. Admiranto, T. Djamaluddin, A. Rachman, and D. D. Wijaya, "Weather conditions at Timau National Observatory from ERA5," *arXiv:2406.11027*, 2024.
- [11] A. Aprilia et al., "Comparative validation of NASA POWER and ERA5 satellite data against meteorological measurements," *Info Sains: Jurnal Manajemen Informatika dan Komputer*, 2025.
- [12] C. J. Paciorek, "Analyzing trends in precipitation patterns using Hidden Markov model stochastic weather generators," *arXiv:2207.08649*, 2022.

- [13] R. Majumder, N. K. Neerchal, and A. Mehta, "Stochastic precipitation generation for the Chesapeake Bay Watershed using Hidden Markov Models with variational Bayes parameter estimation," arXiv:2210.04305, 2022.
- [14] S. Zhang, M. Tuerde, and X. Hu, "Stochastic precipitation generation for the Xilingol League using Hidden Markov Models," *Water*, vol. 15, no. 20, Art. no. 3600, 2023, doi: 10.3390/w15203600.
- [15] Q. Cao, H. Zhang, U. Lall, T. Holsclaw, and Q. Shao, "The predictability of daily rainfall during rainy season over East Asia by a Bayesian nonhomogeneous Hidden Markov Model," *Journal of Flood Risk Management*, vol. 17, no. 1, Art. no. e12942, 2024, doi: 10.1111/jfr3.12942.
- [16] E. Gobet, D. Métivier, and S. Parey, "Interpretable seasonal multisite Hidden Markov Model for stochastic rain generation in France," *Advances in Statistical Climatology, Meteorology and Oceanography*, vol. 11, pp. 159–201, 2025, doi: 10.5194/ascmo-11-159-2025.
- [17] N. H. M. Sani, S. M. Shaharudin, M. S. Ibrahim, U. Lall, and M. L. Tan, "Integrating a hybrid statistical downscaling-based HMM-RF model for enhanced rainfall prediction in Selangor," *Stochastic Environmental Research and Risk Assessment*, vol. 39, pp. 6001–6018, 2025, doi: 10.1007/s00477-025-03103-z.
- [18] B. Tezcan and M. Garcia, "Training a Hidden Markov Model with PMDI and temperature to create climate-informed scenarios," *Frontiers in Water*, vol. 7, Art. no. 1472695, 2025, doi: 10.3389/frwa.2025.1472695.
- [19] Y. Shi, A. Punzo, H. Otneim, and A. Maruotti, "Hidden semi-Markov models for rainfall-related insurance claims," *Insurance: Mathematics and Economics*, vol. 120, pp. 91–106, 2025, doi: 10.1016/j.insmathco.2024.11.008.
- [20] D. V. Martal and B. Setiawaty, "Rainfall risk modelling for rice farming using continuous Hidden Markov Models," *Zero: Jurnal Sains, Matematika, dan Terapan*, vol. 9, no. 3, pp. 703–714, 2025, doi: 10.30829/zero.v9i3.25443.
- [21] M. L. Gámiz, D. Montoro, and M. C. Segovia-García, "Modelling stochastic inflow patterns to a reservoir with a hidden phase-type Markov model," arXiv:2508.02522, 2025.
- [22] A. Kurniadi, E. Weller, S.-K. Min, and M.-G. Seong, "Independent ENSO and IOD impacts on rainfall extremes over Indonesia," *International Journal of Climatology*, vol. 41, pp. 3640–3656, 2021.
- [23] F. R. Muhammad, S. W. Lubis, and S. Setiawan, "The influence of boreal summer Madden–Julian Oscillation on precipitation extremes in Indonesia," arXiv:2112.00001, 2021.
- [24] S. Lestari et al., "Variability of Jakarta rain-rate characteristics associated with the Madden–Julian Oscillation and topography," *Monthly Weather Review*, vol. 150, no. 8, 2022.
- [25] A. Z. A. Zaini, M. Vonnisa, and M. Marzuki, "Seasonal variation of rainfall in Indonesia under normal conditions," *Journal of Penelitian Pendidikan IPA*, 2023.
- [26] A. Z. A. Zaini, M. Vonnisa, and M. Marzuki, "Impact of different ENSO positions and Indian Ocean Dipole events on Indonesian rainfall," *Vietnam Journal of Earth Sciences*, vol. 46, no. 1, pp. 100–119, 2024, doi: 10.15625/2615-9783/19926.
- [27] R. Ramadhan, M. Marzuki, W. Suryanto, S. Sholihun, H. Yusnaini, and R. Muharsyah, "Rainfall variability in Indonesia's new capital associated with the Madden–Julian Oscillation and its contribution to flood events," *Quaternary Science Advances*, Art. no. 100163, 2024, doi: 10.1016/j.qsa.2024.100163.
- [28] A. A. Prabowo, Z. N. Ruslana, and M. Y. Matdoan, "Ordinal logistic regression analysis on the nature of monthly rainfall against ENSO, IOD, monsoon and SST," *MITIGATION: Journal of Multidisciplinary Innovation and Technology*, vol. 1, no. 1, pp. 23–32, 2026.
- [29] M. Ariska, Suhadi, Supari, M. Irfan, and I. Iskandar, "Detection of dominant rainfall patterns in Indonesian regions using empirical orthogonal function and its relation with ENSO and IOD events," *Science and Technology Indonesia*, vol. 9, no. 4, pp. 1009–1023, 2024, doi: 10.26554/sti.2024.9.4.1009-1023.
- [30] A. F. Deshafa, "ENSO and IOD variability: Impacts on precipitation and sea surface temperature in Bali and NTB," *Journal of the Physical Society of Indonesia*, vol. 1, no. 1, pp. 13–25, 2025, doi: 10.35895/jpsi.1.1.13-25.2025.
- [31] N. F. Palisoa, L. J. Sinay, M. Y. Matdoan, Y. Yudistira, and L. Bakarbesy, "Penerapan Support Vector Machine (SVM) untuk klasifikasi kabupaten tertinggal di Provinsi Maluku," *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, vol. 2, no. 2, pp. 79–86, 2023.
- [32] M. Marzuki et al., "Future projections of extreme precipitation over Indonesia's new capital under climate change scenarios using CORDEX-SEA regional climate models," *Atmospheric Research*, 2025.

