Journal of Quality Measurement and Analysis *JQMA* 22(2) 2026, 279-302
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2202.2026.16

STATISTICAL ANALYSIS OF RAINFALL VARIABILITY AND WATER BALANCE COMPONENTS IN JOHOR DISTRICTS USING SEASONAL ORNSTEIN-UHLENBECK

(*Analisis Statistik Variabiliti Hujan dan Komponen Imbangan Air di Daerah-daerah Johor Menggunakan Model Ornstein-Uhlenbeck Bermusim*)

ANIZA AKARAM & ARIFAH BAHAR*

ABSTRACT

This study analyses rainfall dynamics across three key districts in Johor which are Kluang, Batu Pahat, and Senai using a water balance model integrated with Stochastic Differential Equations (SDEs). Daily rainfall data from 2005 to 2024 were examined to assess hydrological variability, seasonal behavior, and long-term rainfall trends. The water balance framework quantifies rainfall, evapotranspiration, runoff, and storage to capture the dynamic movement of water within the hydrological system. Evapotranspiration was estimated using the Penman-based method, which considers temperature, humidity, solar radiation, and wind speed to represent actual water loss. The stochastic component applies the Seasonal Ornstein–Uhlenbeck (SOU) process, a mean-reverting SDE that accounts for randomness and seasonal variation in rainfall patterns. This allows for realistic simulation of rainfall behavior under Malaysia's biannual monsoon influence. Results show that Kluang experiences higher rainfall amplitudes and more frequent extremes, while Senai remains consistently wet and Batu Pahat exhibits moderate but increasing rainfall intensity.The integrated SOU–water balance approach effectively represents both deterministic and stochastic processes, providing a robust framework for quantifying hydrological uncertainty. By coupling stochastic rainfall modeling with physical water balance processes, the study enhances understanding of rainfall–storage dynamics and supports adaptive water resource management, particularly under changing climate conditions.

Keywords: rainfall variability; water demand modelling; water balance model; stochastic modelling

ABSTRAK

Kajian ini menganalisis dinamik hujan di tiga daerah utama di Johor, iaitu Kluang, Batu Pahat dan Senai dengan menggunakan model imbangan air yang digabungkan dengan Persamaan Pembezaan Stokastik. Data hujan harian bagi tempoh 2005 hingga 2024 telah dianalisis untuk menilai kebolehubahan hidrologi, corak bermusim, dan trend hujan jangka panjang. Kerangka imbangan air digunakan untuk mengkuantifikasikan hujan, evapotranspirasi, larian permukaan dan simpanan air bagi menggambarkan pergerakan dinamik air dalam sistem hidrologi. Evapotranspirasi dianggarkan menggunakan kaedah berasaskan Penman yang mengambil kira suhu, kelembapan, sinaran suria dan kelajuan angin bagi mewakili kehilangan air sebenar. Komponen stokastik menggunakan proses Ornstein–Uhlenbeck Bermusim , iaitu suatu SDE yang bersifat berbalik kepada min dan mengambil kira unsur rawak serta variasi bermusim dalam corak hujan. Pendekatan ini membolehkan simulasi yang lebih realistik terhadap tingkah laku hujan di bawah pengaruh dua musim monsun tahunan di Malaysia. Hasil kajian menunjukkan bahawa Kluang mengalami amplitud hujan yang lebih tinggi serta kejadian hujan ekstrem yang lebih kerap, manakala Senai mengekalkan keadaan yang sentiasa lembap dan Batu Pahat menunjukkan intensiti hujan yang sederhana tetapi semakin meningkat. Pendekatan bersepadu SOU–imbangan air ini berupaya mewakili kedua-dua proses deterministik dan stokastik secara berkesan, sekali gus menyediakan kerangka yang mantap untuk mengkuantifikasikan ketidakpastian hidrologi. Dengan menggabungkan pemodelan hujan stokastik dengan proses fizikal imbangan air, kajian ini meningkatkan pemahaman terhadap dinamik hujan–simpanan air dan menyokong pengurusan sumber air yang adaptif, khususnya dalam menghadapi perubahan iklim.

Kata kunci: kebolehubahan hujan; pemodelan permintaan air; model imbangan air; pemodelan stokastik

1. Introduction

Statistical analysis plays a crucial role in understanding and interpreting rainfall behavior, particularly in regions experiencing complex and fluctuating climatic conditions. In Johor, Malaysia, where rainfall is the main contributor to the hydrological cycle, its variability significantly affects water availability. Given the dependence of water supply systems on consistent rainfall patterns, especially in the face of growing climate uncertainty, applying statistical tools to analyze rainfall data becomes essential. These tools help uncover patterns, detect anomalies, and provide insights into long-term trends that are critical for informed water resource planning. Rainfall variability directly affects water demand by influencing both supply availability and consumption behavior. During periods of high rainfall, reliance on supplementary water sources may decrease, while prolonged dry spells can lead to increased withdrawals for agricultural, domestic, and industrial use. This close relationship between rainfall and water consumption emphasizes the importance of integrated approaches that consider both natural variability and human response.

This study focuses on three key districts in Johor specifically Kluang, Batu Pahat, and Senai where each exhibits distinct hydrological characteristics and land use patterns. By analyzing twenty years of daily rainfall data (2005–2024) from these districts, this study aims to quantify and model rainfall variability and its hydrological implications under varying climatic and land-use conditions. The analysis integrates both statistical and stochastic approaches, allowing the identification of deterministic seasonal cycles and random fluctuations that influence water availability. Specifically, rainfall dynamics are modeled using a Stochastic Differential Equation (SDE) framework, which enables the explicit representation of randomness and mean-reverting behavior in rainfall processes. The SDE model is then coupled with a water balance framework to simulate how rainfall variability translates into changes in evapotranspiration, runoff, and soil moisture storage.

The integration of the Water Balance Model with SDEs, particularly the Ornstein–Uhlenbeck (OU) process, offers a reliable approach for simulating and forecasting water dynamics under uncertain conditions. The Ornstein–Uhlenbeck process is a stochastic model commonly applied to variables that fluctuate randomly but tend to return to a long-term average over time. This characteristic makes it well-suited for modeling rainfall, where values vary unpredictably but often stabilize around a mean due to seasonal or climatic influences. By combining a tendency toward equilibrium with random variability, the OU process effectively captures the irregular and mean-reverting nature of rainfall, supporting more realistic simulations of hydrological behavior across different time periods. Unlike traditional methods that mainly look at past patterns, the stochastic approach takes into account the natural unpredictability and sudden changes in rainfall. These ups and downs, also known as rainfall volatility, are caused by shifting weather conditions, local storms, and long-term climate changes. By including this randomness, stochastic models provide a more realistic way to understand and predict how rainfall behaves over time, thus yielding more realistic projections.

This is especially relevant in Johor, where the monsoonal climate and localized weather systems can lead to significant short-term variability that influences both surface water availability and consumption patterns.

Furthermore, the long-term analysis of daily data allows for the detection of gradual shifts in rainfall patterns that may not be evident in short-term studies. These shifts, if linked to climate change impacts, can have profound implications on future water demand profiles and supply resilience. By coupling statistical and stochastic approaches within a physically meaningful model like the water balance framework, this study contributes not only to academic knowledge but also to practical tools for integrated water resource management in Johor and similar tropical regions. However, despite the importance of long-term rainfall modeling, there remains a lack of studies that fully integrate statistical, stochastic, and hydrological perspectives within a unified analytical framework. Most existing research either focuses on statistical trend detection or employs models that fail to capture the probabilistic and dynamic nature of rainfall and its subsequent impact on water demand. This gap limits the ability to understand how long-term rainfall variability translates into real changes in water availability, consumption, and system performance. In regions such as Johor where effective water resource planning must consider both climate variability and rising water demand the absence of such integrated modeling reduces the reliability of management and forecasting strategies. Therefore, there is an urgent need for a holistic modeling approach that realistically simulates rainfall dynamics and translates these insights into practical implications for water demand forecasting and sustainable water supply planning.

In practice, the outputs of this long-term rainfall modelling can be used for strategic planning for the water supply infrastructure, reservoir operations, and drought mitigation strategies. For instance, water utilities can utilize the model's forecasts to adjust distribution schedules, optimize storage capacities, or implement early warning systems during periods of anticipated water scarcity. In agricultural zones such as those found in Kluang and Batu Pahat, rainfall-based forecasts can support irrigation planning and reduce dependency on treated water sources. Likewise, in more urbanized areas like Senai, understanding rainfall variability is crucial for stormwater management and preventing over-extraction from groundwater aquifers.

2. Literature Review

Rainfall serves as the cornerstone of the hydrological cycle, especially in tropical countries like Malaysia, where it directly sustains agriculture, ecosystems, and urban water supplies. Understanding how rainfall behaves over time, including its trends, extremes, and seasonal patterns, is crucial for managing water resources efficiently. In regions such as Johor, where population growth and land development are accelerating, the need to closely examine rainfall variability becomes even more urgent. In line with this, the broader influence of climate on Malaysia's water systems has been examined through hydro-climatic modeling. For instance, Amin *et al.* (2019) analyzed regional climate projections and identified significant changes in rainfall distribution and streamflow behavior across Peninsular Malaysia. Such changes are likely to intensify stress on existing water infrastructure and introduce new challenges in meeting future water demands. Their findings emphasize the importance of incorporating climate variability and projected changes into long-term water management and planning frameworks.

Although numerous studies have examined rainfall behavior, many have focused on short-term timescales, often limited to specific seasons or events. These short-term analyses are typically employed to assess localized hydrological risks such as flash floods, stormwater runoff, or seasonal water shortages. Sun *et al.* (2021) predicted short-time rainfall using a deep learning approach that combines convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, which allowed for effective extraction of spatial and temporal features from meteorological data. Their model showed high accuracy in forecasting immediate rainfall events. Similarly, Prottasha *et al.* (2023) employed supervised machine learning techniques, including decision trees and support vector machines, to classify and forecast short-term rainfall, achieving promising predictive performance. Liao and Yin (2022) also proposed a deep learning model aimed at imminent rainfall prediction and evaluated its accuracy for short-term use. While these studies demonstrate the potential of artificial intelligence in short-range forecasting, their application is largely limited to event-based responses and does not fully address the complexity of long-term rainfall behavior. As a result, they may fall short in applications that require understanding broader climatic trends and their implications on future water demand and resource sustainability.

Long-term rainfall studies, by contrast, aim to capture gradual shifts in climatic patterns that can impact regional water availability and demand. Adib *et al.* (2022) projected future rainfall trends using CMIP6 climate scenarios for the Kurau River Basin, a major rice-growing area. Their research highlighted potential fluctuations in rainfall under different climate models, raising important considerations for future irrigation planning and agricultural water use. Suhaila *et al.* (2010) conducted a comprehensive analysis of three decades of rainfall data during the southwest and northeast monsoon seasons. They found notable variations across regions and seasons, indicating long-term changes in Malaysia's monsoonal rainfall behavior. Adding to this understanding, Muhammad *et al*. (2020) analyzed historical rainfall records and reported increasing variability along with a higher frequency of extreme rainfall events in Peninsular Malaysia. These studies collectively emphasize the significance of long-term analysis for detecting underlying climatic shifts and guiding resilient water management strategies. However, many of these investigations rely on deterministic or statistical models, which may not adequately represent the inherent uncertainty and randomness of rainfall processes.

To make sense of such variability and translate it into actionable planning, water balance models are often used. These models track the inflow and outflow of water within a system and help determine how much water is available for use. A case study by Mohd Faisal *et al.* (2022) applied the SWAT model to the Sungai Muda watershed in Malaysia. The study successfully estimated runoff, evapotranspiration, and baseflow which are the key components in understanding a region's water supply. Their work demonstrated the strength of water balance models in linking land use, rainfall, and hydrological responses in a practical, decision-support framework. Among these components, evapotranspiration plays a particularly important role in estimating actual water loss from the system, especially in agricultural regions. Tukimat *et al.* (2012) conducted a comparison of different methods for estimating evapotranspiration at the Muda Irrigation Scheme and found that the choice of method significantly affects water balance estimates. Similarly, Ahmad *et al.* (2017) analyzed evapotranspiration trends using the Penman-Monteith method in Alor Setar and emphasized its suitability for irrigation planning in regions with variable rainfall. However, Hao *et al.* (2018) highlighted that while the Penman-Monteith method is widely considered one of the most accurate for estimating crop evapotranspiration, it is subject to uncertainties, particularly related to the energy balance closure problem. They argued that although the method performs well under standard conditions, its accuracy may vary depending on data quality and local environmental factors. Therefore, while the Penman-Monteith method remains a preferred choice, it may require calibration or adjustments to improve reliability across different climatic and agricultural settings.

While water balance models provide a structured approach to simulating hydrological systems, they often rely on deterministic or statistical assumptions that may not adequately represent the randomness inherent in rainfall patterns. To overcome these limitations, SDEs offer a more dynamic framework by modeling rainfall and runoff as continuous-time random processes that incorporate both trend and variability components. Khan *et al.* (2023) applied the ARIMA model for forecasting short- and long-term rainfall. Although their study produced useful time-series predictions, ARIMA remains a deterministic model that does not explicitly account for the stochastic nature of rainfall, particularly the irregular fluctuations and extreme variability often driven by climatic factors. In contrast, Houénafa *et al.* (2025) introduced a stochastic model using a Langevin-type SDE driven by α-stable Lévy noise to simulate extreme river discharge events. Their study highlights the potential of non-Gaussian stochastic processes in representing the uncertainty and heavy-tailed behavior frequently observed in hydrological extremes. Despite such promising developments, the application of SDEs in rainfall–demand modeling is still relatively limited, presenting an opportunity to enhance water balance models by integrating stochastic representations of rainfall for more realistic and uncertainty-aware water resource planning.

Supporting this perspective, previous studies have demonstrated that SDE-based approaches are well suited for representing rainfall variability and uncertainty in hydrological systems. For instance, SDE formulations have been applied in rainfall–runoff analysis to explicitly quantify uncertainty and improve the interpretation of hydrological responses (Yoshimi *et al.* 2015). Evaluations of stochastic rainfall models further confirm their ability to reproduce key statistical characteristics required for hydrological applications (Nguyen *et al.* 2023), while hybrid stochastic models have been shown to capture rainfall dynamics across multiple temporal scales (Park *et al.* 2019). In addition, Ornstein–Uhlenbeck-type processes, including censored formulations, have been successfully used to model rainfall occurrence and intensity while accounting for intermittency (Tong & Liu 2021). These findings highlight the continued relevance of SDE-based methods as a flexible and physically interpretable approach for modelling rainfall processes, particularly when both seasonal structure and stochastic variability are important.

Bringing these approaches together, the use of stochastic models, particularly the Ornstein–Uhlenbeck (OU) process, adds greater realism to rainfall and water resource forecasting. When combined with a water balance framework, the OU model enhances flexibility and accuracy, offering valuable support for decision-making in regions like Senai, Kluang, and Batu Pahat, where complex interactions between land use and climate conditions influence water availability. Such integration allows for a more holistic understanding of how rainfall influences water availability and demand under uncertain future conditions. It bridges the gap between observed climate behavior and water management needs, providing tools that can be adjusted to changing environmental and socio-economic contexts.

Beyond the classical Ornstein–Uhlenbeck process, recent advancements in stochastic hydrology have introduced Langevin-type and Lévy-driven SDEs to better represent non-Gaussian rainfall dynamics and extreme hydrological behavior. These formulations capture sudden rainfall bursts or heavy-tailed fluctuations that conventional Gaussian-based models often underestimate (Houénafa *et al.* 2025; Sapozhnikov & Foufoula-Georgiou 2007). In tropical settings, such as Johor, where rainfall is both intermittent and intense, integrating these SDE variants within the water balance framework can improve the representation of rainfall–runoff responses and enhance uncertainty estimation. This refinement complements the OU-based approach used in the present study, extending its applicability to more extreme or nonlinear rainfall conditions relevant to Malaysia's monsoon-driven hydrological systems (Koutsoyiannis 2016; Northrop 2024).

In summary, the literature highlights the importance of a multi-faceted approach that integrates detailed rainfall analysis, stochastic modelling, and water balance assessment to capture both deterministic and random aspects of hydrological behavior. Such integration supports more resilient and adaptive water management strategies, enabling planners to anticipate variability in rainfall, runoff, and storage under evolving climatic and land-use conditions in Johor.

3. Methodology

3.1. *Data source*

Daily rainfall data from three meteorological stations located in Johor, Malaysia which are Senai, Kluang, and Batu Pahat were obtained from the Malaysian Meteorological Department (MetMalaysia) (Malaysian Meteorological Departmen n.d.). The dataset spans from January 2005 to December 2024, providing a long-term observation of rainfall dynamics across the state.

Each district represents a distinct geographical and climatic zone, enabling spatial assessment of rainfall variation across Johor:

1. Senai Station (southern zone) lies within an urban and coastal environment, where rainfall patterns are influenced by maritime air masses and localized convective storms.
2. Kluang Station (central zone) is situated in the interior highlands, typically experiencing higher rainfall amounts due to orographic uplift and stronger monsoon exposure.
3. Batu Pahat Station (western zone) represents lowland agricultural areas, often receiving comparatively lower rainfall totals and longer dry intervals during inter-monsoon periods.

These three stations collectively capture the spatial rainfall gradients across Johor from wetter central highlands to more variable coastal and lowland regions. Such variation reflects the combined effects of topography, land use, and monsoonal wind patterns on local rainfall distribution.

Rainfall records from different stations are occasionally affected by missing values, irregular time intervals, or non-standard data formats due to differences in observation practices. To ensure data consistency for stochastic modeling and water balance analysis, several preprocessing steps were applied:

1. Identification and treatment of missing values through interpolation or statistical imputation, depending on the duration of the gaps.
2. Temporal alignment to a fixed daily time resolution across all districts.
3. Standardization of units and date formats (rainfall in mm/day).

This standardized dataset ensures a continuous and reliable rainfall time series (2005–2024) suitable for modeling rainfall variability and water balance processes under uncertainty.

3.2. *Stochastic differential equation*

Stochastic differential equations (SDEs) are effective for modeling systems that exhibit both predictable patterns and random variability, making them well-suited for representing rainfall behavior. Since rainfall patterns are irregular and uncertain in both timing and intensity, purely deterministic models often fall short. SDEs address this by integrating a randomness component commonly represented by Brownian motion, which allows them to capture the inherent uncertainty and fluctuations observed in rainfall data.

A commonly used form of SDE in this context is the Ornstein–Uhlenbeck (OU) process, which is useful for capturing mean-reverting behavior in rainfall dynamics. However, rainfall processes typically exhibit both random fluctuations and pronounced seasonal patterns. To account for these characteristics, this study adopts a Seasonal Ornstein–Uhlenbeck (SOU) process, which extends the classical mean-reverting framework by allowing the long-term mean to vary over time. In this formulation, the mean level is treated as a periodic function to reflect recurring intra-annual variations in rainfall. The SOU process is expressed as:

$$dR(t) = \theta(\mu(t) - R(t))dt + \sigma dW(t) \tag{1}$$

where $R(t)$ represent rainfall intensity or cumulative rainfall at time t,θ represents the rate at which the process reverts to the time-varying seasonal mean function $\mu(t)$, while σ reflects the intensity of random fluctuations, and $W(t)$ is a Wiener process (standard Brownian motion).

The term $\mu(t)$ governs the seasonal structure of the rainfall process by allowing the long-term level to evolve systematically over time. For daily rainfall data, $\mu(t)$ is specified using a sinusoidal form to reflect the recurring annual cycle observed in tropical climates. In Malaysia, rainfall patterns are largely controlled by the Northeast and Southwest monsoons, which produce alternating periods of relatively high and low precipitation within a year. A sinusoidal function provides a smooth and continuous approximation of this cyclic behaviour while remaining analytically tractable within the stochastic framework. Specifically,

$$\mu(t) = a_0 + a_1 sin\left(\frac{2\pi t}{T}\right) + a_2 cos\left(\frac{2\pi t}{T}\right) \tag{2}$$

where a_0 is the average annual rainfall level. The coefficients a_1 and a_2 jointly determine the strength and timing of seasonal variation where the amplitude reflects how pronounced the difference is between wetter monsoon periods and relatively drier intervals, while the phase shift determines when peak rainfall occurs during the year. The term $\frac{2\pi t}{T}$ enforces periodicity with cycle length T, corresponding to one year for daily observations. This formulation aligns with the physical behaviour of rainfall in Malaysia, where seasonal fluctuations are regular but influenced by monsoon timing, making it suitable for capturing both the intensity and timing of rainfall variation.

Following the specification of the seasonal mean $\mu(t)$, the model parameters are estimated using discretely observed rainfall data. Since the presence of a time-varying mean prevents a simple closed-form solution, numerical approximation is employed. The Euler–Maruyama method is used to simulate the process, allowing the continuous-time dynamics to be approximated at discrete time intervals.

Parameter estimation is carried out using both Maximum Likelihood Estimation (MLE) and the Method of Moments (MoM). In the MLE approach, the discretized process is treated as conditionally Gaussian, with its mean determined by the seasonal function $\mu(t)$. Specifically, the distribution of $R_{t+\Delta t}$ given R_t is characterized by a mean M_t and variance V_t, where M_t incorporates the contribution of the seasonal component over the interval $[t, t + \Delta t]$, while V_t retains the same structure as in the mean-reverting process. The parameters are obtained by maximizing the log-likelihood function:

$$\mathcal{L}(\theta, \mu, \sigma) = -\frac{1}{2}\sum_{t=1}^{n-1}\left[log(2\pi V_t) + \frac{(R_{t+1} - M_t)^2}{V_t}\right] \tag{3}$$

where the conditional mean is given by $M_t = R_t e^{-\theta\Delta t} + \int_t^{t+\Delta t} \theta\mu(s)\, e^{-\theta(t+\Delta t-s)} ds$, and the conditional variance is $V_t = \frac{\sigma^2}{2\theta}\left(1 - e^{-2\theta\Delta t}\right)$. This formulation accounts for the influence of the seasonal mean over time and enables parameter estimation through numerical maximization of the likelihood function.

Alternatively, the Method of Moments (MoM) can be applied by first removing the seasonal component. Let $X_t = R_t - \mu(t)$. The transformed series approximately follows a standard mean-reverting process with zero mean. Its theoretical properties can be used to relate the parameters to observable quantities. In particular, the lag-one autocorrelation satisfies $\rho(1) = e^{-\theta\Delta t}$, while the variance is given by $\frac{\sigma^2}{2\theta}$. Using the sample variance and autocorrelation of X_t, the parameters can be approximated as:

$$\hat{\theta} = -\frac{ln(\rho_1)}{\Delta t}, \hat{\sigma} = \sqrt{2\hat{\theta} s_X^2 (1 - \rho_1^2)} \tag{4}$$

These estimates may be used as initial values for likelihood optimization or as a consistency check for the MLE results.Once the parameters have been determined, the SOU process can be simulated numerically using the Euler–Maruyama discretization scheme. The continuous-time model is approximated in discrete form as:

$$R_{t+\Delta t} = R_t + \theta(\mu - R_t)\Delta t + \sigma\sqrt{\Delta t}\varepsilon_t \tag{5}$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$. This recursive formulation allows the generation of synthetic rainfall trajectories that preserve both the seasonal pattern governed by $\mu(t)$ and the stochastic variability inherent in the process.

3.3. *Water balance model*

The water balance model (WBM) is designed to quantify the interaction between water input, output, and storage within a specific area or watershed over a given time frame. It evaluates how rainfall influences water availability by incorporating key hydrological components such as evapotranspiration, surface runoff, infiltration, and changes in soil or groundwater storage. This approach is widely used in water resource management, particularly for applications in agriculture, urban water planning, and climate impact studies.

The general water balance model equation is:

$$\Delta S = P - ET - Q \tag{6}$$

where ΔS is the change in water storage, P is precipitation, ET is evapotranspiration, and Q is runoff (Chow *et al.* 1988). In Malaysia's equatorial climate, precipitation can be treated entirely as rainfall since the country does not experience snow or snowmelt. This simplification makes rainfall the sole input in the hydrological balance, emphasizing its critical role in determining water availability and supporting resource management. Through this formulation, the inflow and outflow components of a hydrological system which are rainfall, evaporation, runoff, and storage can be quantified, allowing for the assessment of water distribution under varying climatic and land-use conditions.

Evapotranspiration in Eq. (7), was estimated using the Penman–Monteith method, which integrates both energy balance and aerodynamic principles. The standardized FAO Penman–Monteith equation (Allen *et al.* 1998) is given as:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{7}$$

where ET_0 is the reference evapotranspiration (mm/day), Δ is the slope of vapor pressure curve $(kPa/°\mathrm{C})$, R_n is the net radiation $(MJ/m^2/day)$, G is the soil heat flux density $(MJ/m^2/day)$, γ is the psychrometric constant $(kPa/°\mathrm{C})$, T is the mean daily air temperature (°C), e_s is the saturation vapor pressure (kPa), e_a is the actual vapor pressure (kPa), and u_2 is the wind spped at $2\ meters$ (m/s). This method assumes a reference crop surface and provides a consistent standard for estimating evapotranspiration across various land uses.

Runoff was estimated using the Soil Conservation Service Curve Number (SCS-CN) method, a widely accepted empirical approach developed by the United States Department of Agriculture (USDA), which incorporates land use, soil type, and antecedent moisture conditions (USDA SCS 1972). The relationship is expressed as:

$$Q = \left(\frac{(P - 0.2F)^2}{P + 0.8F}\right), for\ P > 0.2F; else\ Q = 0 \tag{8}$$

where F is the potential maximum retention derived from the Curve Number (CN), which reflects the land surface's runoff potential. This approach is widely used for its simplicity and effectiveness in representing surface runoff behavior in tropical hydrological systems. Recent hydrological research in Malaysia has emphasized that the conventional CN method, originally developed for temperate catchments, may not accurately capture the hydrological response of tropical basins. Several studies have proposed that the CN formulation should be revised to reflect local rainfall intensity, soil texture, and infiltration behavior characteristic of Malaysian conditions (Khor *et al.* 2023). Moreover, GIS-based regional analyses have reported national-scale CN ranges of approximately 61, 78, and 89 under dry, average, and wet antecedent moisture conditions, respectively (Alcober & Macuha 2024). However, these values represent general national conditions and do not provide site-specific calibration for the districts of Senai, Kluang, or Batu Pahat.

Table 1: Curve Number assignment based on land use and hydrological characteristics

Location	Dominant Land Use	Assumed HSG	CN Value	Justification
Senai	Urban / industrial, high impervious surfaces	B–C	85	High runoff potential due to impervious surfaces; consistent with urban CN values (80–90)
Kluang	Mixed agriculture and forest	B	70	Moderate infiltration and runoff; aligns with mixed land CN range (65–75)
Batu Pahat	Predominantly agriculture (rural)	B	65	Lower runoff potential with higher infiltration; consistent with agricultural CN values (60–70)

As shown in Table 1, the CN values were assigned based on the dominant land use and the corresponding hydrological characteristics of each district, following the USDA Soil Conservation Service approach (1972). The classification also considers the Hydrologic Soil Group (HSG), which categorises soils according to their infiltration capacity and runoff potential. HSG is typically divided into four groups, namely A, B, C, and D, where Group A represents soils with high infiltration and low runoff potential, while Group D represents soils with very low infiltration and high runoff potential.

In this study, HSG B was assumed as representative of the study areas due to its moderate infiltration characteristics, which are commonly associated with mixed agricultural and semiurban regions in Johor. Based on this assumption, the assigned CN values reflect the varying degrees of surface runoff influenced by land use. Urbanised areas such as Senai correspond to higher CN values due to extensive impervious surfaces, whereas agricultural regions such as Batu Pahat are associated with lower CN values due to greater infiltration capacity. Kluang, with mixed land use, exhibits intermediate behaviour consistent with its assigned CN value. These parameter choices provide a reasonable basis for simulating runoff response within the stochastic rainfall–runoff framework.

The soil water retention parameter (F) was computed from the CN using the relationship $F = \frac{25400}{CN} - 254$, which represents the maximum potential storage before runoff begins (USDA SCS 1972). In this study, the CN values between 65 and 85 correspond to typical conditions in Malaysian agricultural and mixed-use catchments, including oil palm plantations and peri-urban areas. The rainfall input corresponds to daily observed precipitation, and when $P \leq 0.2F$, it is assumed that all rainfall is absorbed by the soil or lost to initial abstraction processes such as infiltration, interception, or surface storage, resulting in no surface runoff.

In the stochastic water balance framework, this CN–F relationship serves as a key component connecting rainfall variability to surface runoff generation. By integrating the CN-based runoff estimation within the Seasonal Ornstein–Uhlenbeck (SOU) rainfall simulation, the model dynamically captures how stochastic fluctuations in rainfall influence infiltration, evapotranspiration, and storage over time. This coupling allows the water balance model to respond realistically to both seasonal wet-dry transitions and random rainfall extremes, providing a more robust representation of hydrological processes under Malaysia's tropical monsoon climate.

4. Result and Discussion

4.1. *Descriptive statistics*

The rainfall statistics for Senai, Kluang, and Batu Pahat show notable differences in daily rainfall patterns as in Table 1. Senai records the highest average rainfall at 7.43 millimetres, with a standard deviation of 15.9 millimetres, indicating considerable variation. It also has high skewness and kurtosis values, reflecting frequent dry days and occasional intense rainfall. Kluang has a slightly lower average rainfall of 6.46 millimetres but displays even higher skewness and kurtosis, suggesting a stronger tendency for rare but extreme rain events. Batu Pahat shows the lowest average rainfall at 6.26 millimetres and the least variability, though the distribution still indicates occasional heavy rainfall. In all three districts, the median rainfall is close to zero, meaning most days are relatively dry. For consistency in data analysis, rainfall values below 0.1 millimetres, including zero, were standardized to 0.09 millimetres across all districts.

Table 2: Descriptive statistics of daily rainfall in Senai, Kluang, and Batu Pahat

District	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Median	Maximum
Senai	7.43	15.9	4.50	35.6	0.09	0.40	264.00
Kluang	6.46	15.6	5.67	58.9	0.09	0.20	315.80
Batu Pahat	6.26	14.0	3.97	22.4	0.09	0.20	183.20

An exploratory analysis of daily rainfall from 2005 – 2024 showed that near-dry conditions ($\leq$ 0.1 mm) occurred on 45.78% of days in Senai, 49.75% in Kluang, and 48.53% in Batu Pahat. These high proportions indicate strong rainfall intermittency influenced by alternating monsoon and dry periods in Johor. Identifying the frequency of near-dry days is essential for understanding rainfall persistence and its effect on runoff, evapotranspiration, and soil moisture. This variability supports the application of stochastic models, such as the SOU process, which can realistically capture both the dry and wet fluctuations that deterministic models often overlook.

The daily rainfall pattern in Senai from 2005 to 2024 exhibits significant variability, with numerous spikes indicating days of heavy rainfall. The highest rainfall events occurred between 2006 and 2008, where the daily rainfall exceeded 250 mm. After this period, while rainfall remained consistent in frequency, the intensity of extreme events slightly decreased but continued to show regular occurrences of high rainfall days, often between 50 mm and 150 mm. Notably, there are periodic peaks every few years, suggesting recurring intense rainfall possibly due to seasonal or climatic cycles.

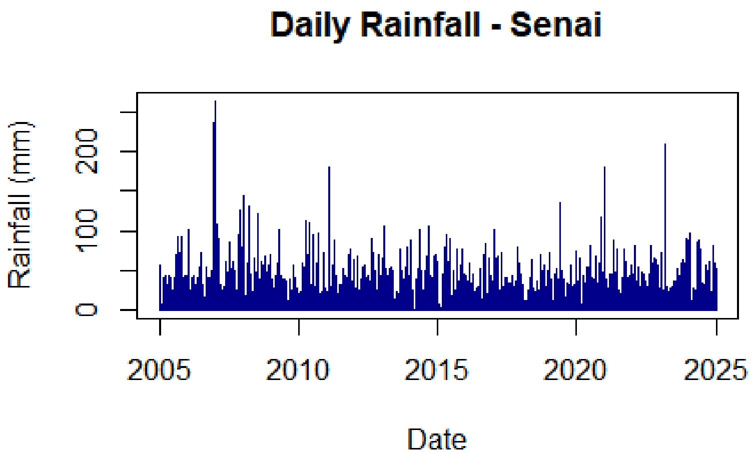

Figure 1: Daily rainfall (mm) in Senai

In Kluang, the daily rainfall data from 2005 to 2024 shows a pattern of frequent and occasionally intense rain events. The graph highlights major rainfall spikes, especially around 2007, 2011, and again after 2020, with rainfall surpassing 300 mm on certain days. These extreme weather events point to episodes of torrential rainfall. Compared to Senai, Kluang experienced slightly higher peak rainfall, indicating a greater propensity for extreme weather events. The frequency of high rainfall days also appears steady over the two decades, reflecting a persistently wet climate with a few particularly intense years.

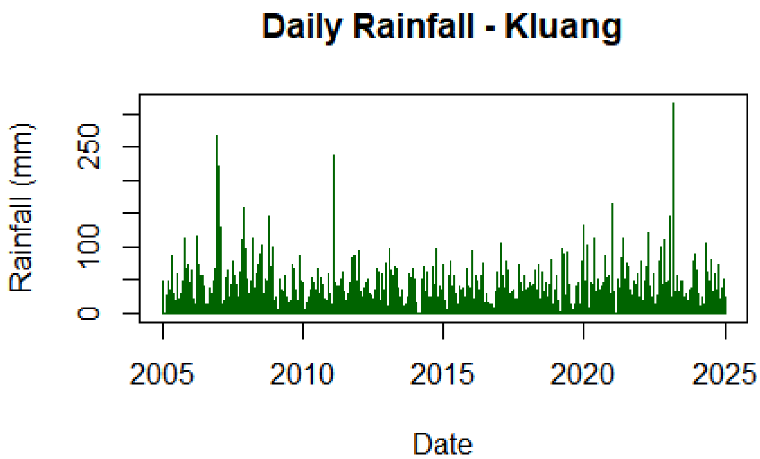

Figure 2: Daily rainfall (mm) in Kluang

The rainfall pattern in Batu Pahat over the same period is characterized by more moderate daily rainfall compared to Senai and Kluang. Most rainfall events range below 100 mm, with fewer instances of extreme rainfall. There is a noticeable increase in the frequency and intensity of rainfall after 2018, culminating in a significant spike around 2022–2023 where rainfall neared 160 mm in a day. Despite these events, Batu Pahat generally shows a more stable and less volatile daily rainfall pattern, suggesting it may experience less frequent but still impactful heavy rain events.

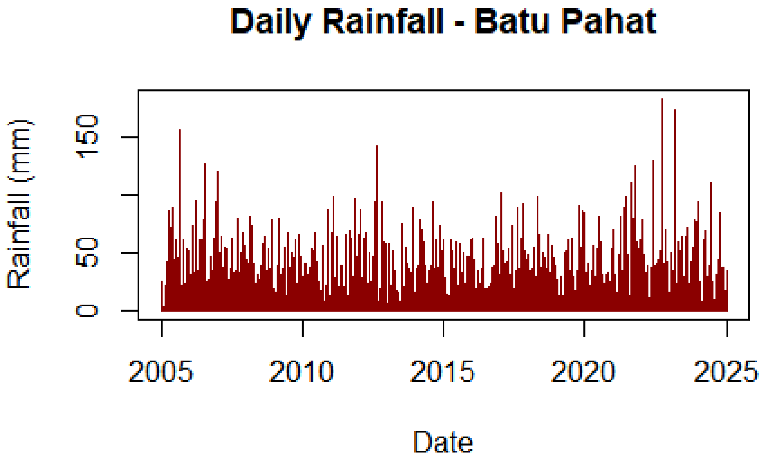

Figure 3: Daily rainfall (mm) in Batu Pahat

While all three locations show active rainfall patterns, Kluang is most prone to extreme rainfall events. Senai maintains a consistently wet climate, with less extreme peaks in recent years. Batu Pahat, although historically more moderate, is now showing signs of increasing rainfall intensity over the last few years. These spatial differences in rainfall behavior have direct implications for SDE modeling. In areas such as Kluang, where extreme rainfall events occur more frequently, the stochastic component (i.e., the diffusion term, $\sigma\, dW(t)$) must capture higher variability and heavier tails to represent these fluctuations accurately. Conversely, in Senai, where rainfall is more consistent with fewer extremes, the model's mean-reverting parameter (θ) plays a stronger role in stabilizing fluctuations around the long-term average. Batu Pahat, showing a recent upward shift in rainfall intensity, may require a time-varying or seasonally adjusted mean $\mu(t)$ to reflect evolving climatic patterns. Together, these regional contrasts emphasize the importance of calibrating the SDE parameters $(\theta, \mu(t), \sigma)$ to local hydrological characteristics so the model can realistically represent both stable and extreme rainfall regimes across districts.

4.2. *Amplitude*

In this study, the amplitude represents the strength of the seasonal cycle in rainfall, reflecting how much daily rainfall typically deviates from the annual mean due to seasonal effects. A larger amplitude indicates a more pronounced wet–dry contrast, while a smaller amplitude suggests more consistent rainfall throughout the year. This measure is valuable because it quantifies the degree of seasonal variability, allowing us to assess how strongly each district's rainfall is influenced by monsoon cycles.

The amplitude of the seasonal component is determined from the sinusoidal representation in Eq. (2). Since the seasonal mean is expressed as a combination of sine and cosine terms, the overall magnitude of seasonal variation is given by the resultant amplitude $A = \sqrt{\alpha_1^2 + \alpha_2^2}$. This quantity represents the extent to which rainfall deviates from the annual average level α_0 over the course of a year. In physical terms, a larger amplitude indicates a more pronounced contrast between wetter and drier periods, which is consistent with stronger monsoon influences. Conversely, a smaller amplitude reflects more uniform rainfall distribution throughout the year.

The amplitude analysis reveals clear differences in seasonal rainfall variability across the three districts. Kluang records the highest amplitude (1.11 mm/day), reflecting a stronger contrast between wet and dry periods, while Senai follows closely (1.06 mm/day), also showing considerable seasonal fluctuation. In contrast, Batu Pahat has the lowest amplitude (0.79 mm/day), indicating relatively stable rainfall throughout the year.

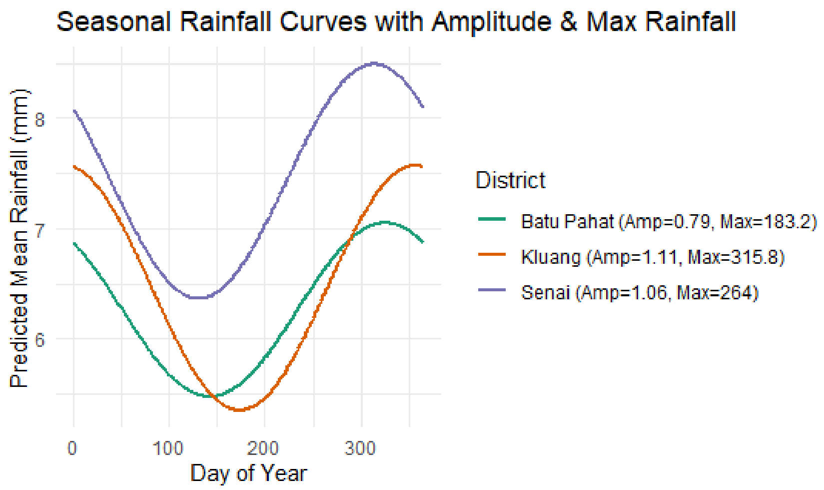

Figure 4: Seasonal rainfall curves for Senai, Kluang, and Batu Pahat

The maximum daily rainfall intensity is computed from the observed data as the maximum value of daily rainfall in each districts. This further highlight these differences, with Kluang reaching 315.8 mm/day compared to Batu Pahat's 183.2 mm/day, suggesting that Kluang experiences more extreme rainfall events. These patterns correspond with Malaysia's biannual monsoon influence, where the Northeast Monsoon (November–March) brings heavier rainfall to eastern and central Johor, enhancing Kluang's seasonal variation. The Southwest Monsoon (May–September) is generally drier but interspersed with localized storms, while inter-monsoon months (April and October) contribute to rainfall peaks in Senai and Batu Pahat. The higher amplitudes in Kluang and Senai point to stronger sensitivity to monsoon transitions, whereas Batu Pahat's lower amplitude may reflect reduced exposure to monsoon extremes due to geographic and topographic factors.

Understanding the amplitude of rainfall variation is crucial, as it reflects the magnitude of seasonal oscillations and the system's sensitivity to climatic forcing. A higher amplitude, such as observed in Kluang and Senai, indicates stronger seasonal contrasts between wet and dry periods signifying that these districts experience pronounced fluctuations in water input throughout the year. This has important implications for hydrological modeling and water resource management: larger amplitudes suggest greater variability in runoff, storage, and potential flooding risks during monsoon peaks. In contrast, Batu Pahat's lower amplitude suggests a more moderate hydrological response, with fewer extremes and a more stable water balance over time. From a stochastic modeling perspective, amplitude helps determine the seasonal mean function $\mu(t)$ in the SOU process, allowing the model to accurately reproduce observed rainfall cycles and capture district-specific climatic dynamics.

4.3. *Normality test*

Assessing the normality of rainfall data is a crucial preliminary step in statistical analysis, particularly when the chosen modeling framework involves SDEs with components that assume underlying Gaussian or log-normal behavior. The Anderson-Darling test is employed in this study to evaluate whether the distribution of daily rainfall data follows a normal distribution. Compared to other normality tests, Anderson-Darling is more sensitive to deviations in the tails of the distribution, which is especially relevant for rainfall data that often includes extreme or heavy rainfall events. The test provides a formal statistical basis to confirm or reject the assumption of normality, thereby guiding subsequent choices in data transformation or model selection. If the rainfall data are found to be non-normally distributed, this justifies the use of nonparametric methods or log-transformed data when integrating the rainfall process into the SDE framework.

Table 3: The Anderson-Darling test

District	A Statistic	p-value
Senai	1176.6	< 2.2e-16
Kluang	1325.3	< 2.2e-16
Batu Pahat	1289	< 2.2e-16

The Null Hypothesis (H_0) asserts that the data are drawn from the normal distribution. Conversely, the Alternative Hypothesis (H_A) posits that the data do not originate from that distribution. If the p-value is less than 0.05, the null hypothesis is rejected since there is evidence that the data does not follow the normal distribution. When the p-value is more than 0.05, we cannot reject the null hypothesis because the data does not significantly deviate from normal distribution. The result of the Anderson-Darling test proved that the data for all districts does not follow normal distribution. The deviation from normality has important implications for model selection, as deterministic or conventional statistical approaches that assume normally distributed data may fail to represent the occurrence of rainfall extremes accurately. Hence, employing a SDE framework specifically the SOU process is more suitable. This approach effectively models non-normal, intermittent, and mean-reverting rainfall dynamics, allowing for a more realistic representation of both random fluctuations and seasonal variability observed in tropical climates.

4.4. *Stationarity test*

Stationarity, which implies that statistical properties such as the mean and variance remain constant over time, is a key assumption in many time series models and stochastic simulations. Applying the Augmented Dickey-Fuller (ADF) test allows us to assess whether the rainfall data is stationary or if it exhibits unit roots, indicating non-stationarity. This is particularly relevant when determining whether differencing or transformation is required before model fitting. In the context of SDE modeling, non-stationary data may necessitate additional considerations for the drift and volatility terms to account for time-varying behavior. Establishing stationarity ensures that the model outputs are interpretable and that future simulations are not distorted by structural shifts in the data.

The Null Hypothesis (H_0) for the ADF test is that the time series contains a unit root, indicating that it is non-stationary and may follow a stochastic trend over time. While the Alternative Hypothesis (H_A) of the test is that the time series does not contain a unit root, which implies it is stationary and its statistical properties do not change over time.

The stationarity result for Senai, Kluang, and Batu Pahat is the same, with *p*-values equal to 0.01 for all districts. As the p-value is below 0.05 for each case, the null hypothesis of a unit root is rejected, suggesting that the rainfall time series exhibits stationarity.

4.5. *Trend analysis*

Identifying the presence of a trend in rainfall data is critical for understanding long-term climatic shifts and their potential impact on regional water resources. The Mann-Kendall test is a non-parametric statistical method used to detect trends in a time series dataset without requiring the data to follow a specific distribution. It evaluates whether there is a consistent upward or downward trend over time by analyzing the ranks of the data rather than their actual values. The Mann-Kendell test suggest that when p-value less than 0.05, the trend in the data is statistically significant either upward or downward. Meanwhile, when p-value is greater or equal to 0.05, there is no significant trend in the data.

Table 4: Mann-Kendell test result

District	*p*-value	Z
Senai	0.5111	0.6572
Kluang	0.1687	1.3763
Batu Pahat	0.5194	0.6443

Based on the results of the Mann-Kendall trend test, there is no statistically significant trend in daily rainfall for any of the three districts. For Senai, and Batu Pahat, with p-values 0.5215, and 0.5195, respectively, which are well above the common significance level of 0.05. This indicates that there is no clear increasing or decreasing trend in rainfall over time. Similarly, Kluang has a p-value of 0.1687, which, while lower than Senai and Batu Pahat, still does not suggest a significant trend. In general, the Mann-Kendall test suggests that rainfall in all three districts has remained relatively consistent over the observed period, with no significant upward or downward trend.

4.6. *Seasonality analysis*

Rainfall in tropical regions like Johor often exhibits strong seasonal patterns driven by monsoonal influences. It is therefore necessary to isolate and quantify seasonal components to distinguish between short-term variability and periodic climatic cycles. STL decomposition offers a flexible and robust approach to separating the seasonal, trend, and irregular components of time series data, even in the presence of noise and missing values. This decomposition aids in understanding the regularity and magnitude of seasonal fluctuations, which is important when calibrating the seasonal demand side of the water balance model or when simulating rainfall behavior in SDE frameworks.

The STL (Seasonal-Trend decomposition using Loess) decomposition analysis reveals that rainfall patterns in all districts display consistent and repeating seasonal trends. By separating the seasonal component from the overall time series, the analysis shows that rainfall tends to vary predictably throughout the year. This indicates that seasonal factors significantly influence rainfall distribution across the districts. Recognizing this regularity is crucial for improving the accuracy of rainfall forecasting and supporting effective water resource management strategies. The seasonal patterns identified through STL decomposition align closely with Malaysia's biannual monsoon cycles, specify that rainfall variability across the districts is primarily driven by the Northeast and Southwest Monsoons. This relationship underscores the importance of accounting for monsoon-driven seasonality in rainfall forecasting and stochastic modeling.

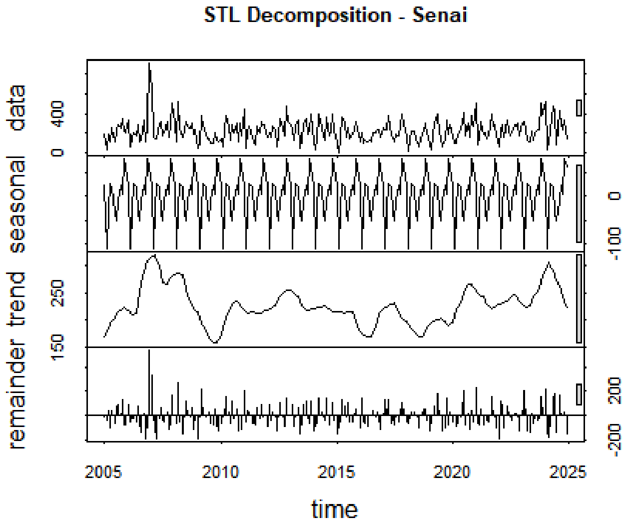

Figure 5: STL decomposition for Senai

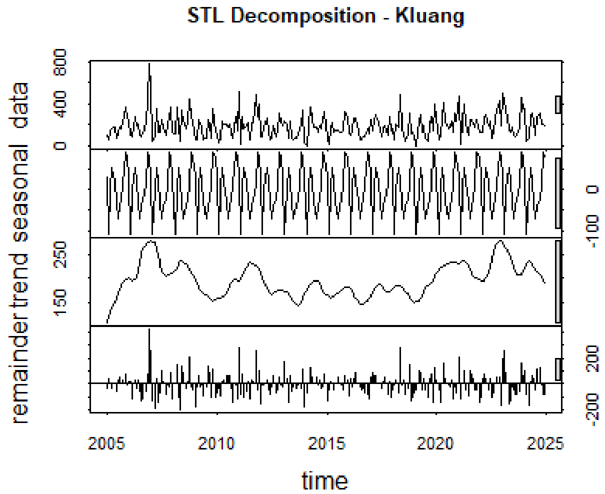

Figure 6: STL decomposition for Kluang

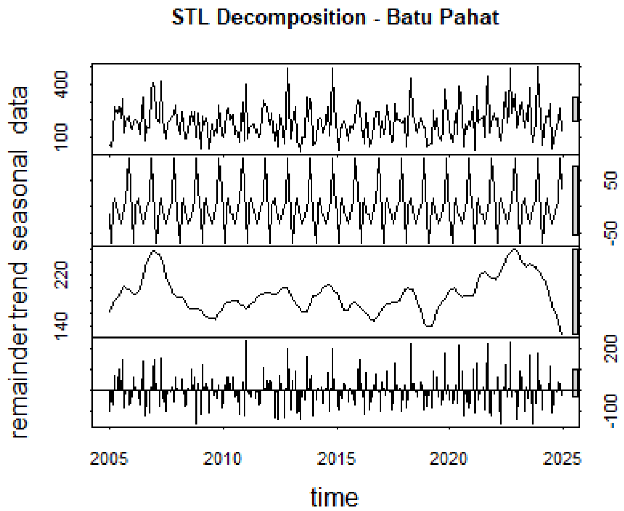

Figure 7: STL decomposition for Batu Pahat

4.7. *Stochastic differential equations*

The statistical analysis of the rainfall data reveals that the series is stationary and shows a clear seasonal pattern. However, it does not follow a normal distribution and lacks a significant long-term trend. Given these characteristics, the standard Ornstein-Uhlenbeck (OU) process which assumes a constant mean and normally distributed residuals is not suitable for modeling this dataset. The presence of strong seasonality requires a model that incorporates periodic behavior into the mean structure.

To better capture these seasonal dynamics while maintaining the stationary and mean-reverting nature of the process, the Seasonal Ornstein-Uhlenbeck (SOU) model is more appropriate.

The estimated parameters of the SOU model for the three districts indicate distinct rainfall dynamics influenced by local climatic and geographical conditions. Senai (θ = 1.7829), and Kluang (θ = 1.777) exhibit relatively similar mean-reversion rates, suggesting that deviations from the seasonal mean tend to stabilise at a comparable speed in both locations. In contrast, Batu Pahat shows a slightly higher value of $\theta = 2.0027$, indicating a faster adjustment of rainfall towards its mean level following fluctuations. In terms of variability, the diffusion parameter σ is highest in Batu Pahat ($\sigma = 2.4574$), followed closely by Senai ($\sigma = 2.4287$), while Kluang records a slightly lower value ($\sigma = 2.3327$). This suggests that Batu Pahat experiences greater random variability in daily rainfall, whereas Kluang exhibits comparatively more stable fluctuations. Overall, while the persistence behaviour is broadly similar across the districts, differences in σ highlight variations in the intensity of stochastic rainfall variability. This reflect how local topography and exposure to monsoon systems influence rainfall dynamics across Johor.

The actual rainfall data for Senai demonstrates a pattern of frequent precipitation events, with several significant peaks reaching above 150 mm. These peaks indicate intense rainfall days scattered throughout the year. The simulated rainfall generally follows the trend of the actual data but tends to underestimate the magnitude of higher rainfall events. Although the model captures the timing and frequency of rainfall to a reasonable extent, it lacks accuracy in replicating extreme values. This suggests that while the simulation is adequate for estimating average daily rainfall, it requires further refinement to better handle extreme weather scenarios common in Senai.

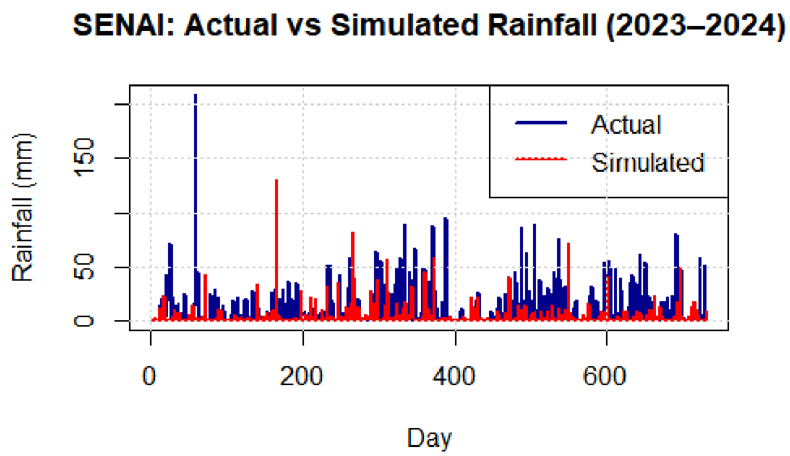

Figure 8: Senai actual and predicted rainfall

In Kluang, the actual rainfall exhibits a highly variable distribution with multiple extreme spikes, particularly early in the period, where rainfall exceeds 300 mm. These extremes highlight the region's vulnerability to intense storm events. The simulated data, however, remains relatively modest and smoother, rarely surpassing 100 mm. This gap indicates that the simulation struggles to replicate the intensity and variability of real rainfall conditions in Kluang. The model seems to underrepresent severe events, which could lead to underestimation in water resource planning or flood risk assessments if used without adjustments.

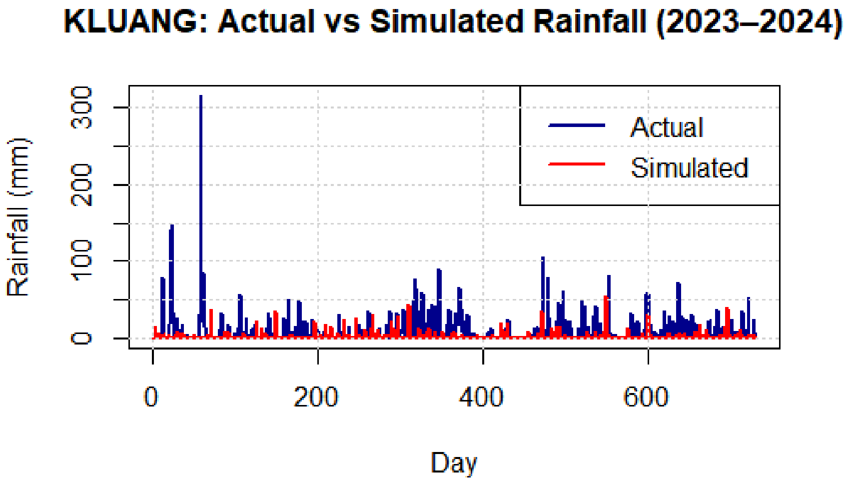

Figure 9: Kluang actual and predicted rainfall

For Batu Pahat, the actual rainfall pattern shows moderate variability, with occasional peaks surpassing 150 mm. The simulated values tend to mirror the general rainfall distribution and frequency, but still underestimate the intensity of heavier rain days. Compared to Kluang and Senai, the match between actual and simulated data in Batu Pahat appears relatively closer, though discrepancies remain. The simulation captures the broader pattern of rainfall occurrence but requires enhancement to reflect extreme events more accurately, which are essential for precise hydrological modeling.

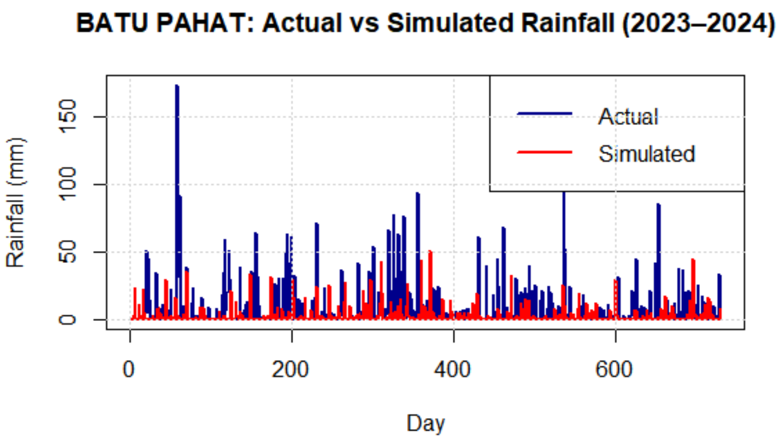

Figure 10: Batu Pahat Actual and Predicted Rainfall

To evaluate the accuracy of the rainfall predictions used in the water balance model, four statistical error metrics were applied: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These indicators help assess the model's predictive performance and its ability to reproduce observed rainfall patterns.

Table 5: The measurement error

District	MAE	RMSE	MAPE
Senai	9.82	19.94	2.15%
Kluang	8.60	21.10	1.47%
Batu Pahat	7.99	16.98	1.75%

As shown in Table 4, the model demonstrates consistent performance across all districts, with MAE and RMSE values remaining within a reasonable range for daily rainfall prediction. The results indicate that the deviations between observed and predicted rainfall are moderate, reflecting the inherent variability of rainfall processes rather than systematic model bias.

In addition, the MAPE values for all districts fall between 1.47% and 2.15%, which are well below the commonly accepted threshold of 10% for good predictive accuracy. This indicates that the model is able to estimate rainfall with a satisfactory level of precision. Overall, the error metrics suggest that the model performs reliably across different locations, supporting its suitability for use in subsequent water balance analysis.

4.8. *Water Balance Model*

The water balance model is employed to estimate the dynamic interactions among key hydrological components, specifically evapotranspiration, runoff, and changes in water storage. The model integrates both climatic and hydrological parameters to provide a comprehensive understanding of the water cycle within the study area. The precipitation (P) in this study is represented solely by rainfall, which serves as the only input from the atmosphere. The climatic variables required for estimating evapotranspiration using the Penman-Monteith method, such as solar radiation, air temperature, wind speed, and relative humidity were obtained from the NASA POWER Data Access Viewer (NASA POWER Project n.d.).

For runoff estimation, the Soil Conservation Service Curve Number (SCS-CN) method was employed, using land use characteristics and soil data specific to each district. In Senai, the runoff calculation employed a CN of 85, reflecting its highly urbanized and impervious landscape, while for Kluang, which comprises mixed agricultural and forested land, a CN value of 70 was adopted, and Batu Pahat, characterized by predominantly rural and agricultural areas, was assigned a CN of 65. These parameter settings represent the distinct land surface characteristics and infiltration potentials of each district, ensuring a more realistic simulation of the stochastic water balance components. Generally, a higher CN value corresponds to greater surface runoff and lower infiltration, which is typical in urban or impervious regions, while lower CN values indicate higher infiltration capacity and are associated with forested or agricultural landscapes.

The initial storage, S_0 was estimated from the mean rainfall–evapotranspiration (ET) balance during the first 60 days of the simulation period:

$$S_0 = max\left(\overline{P_{(1:60)}} - \overline{ET_{(1:60)}}, 10\right) \quad (9)$$

A 60-day window was selected to represent the early monsoonal wetness that typically characterizes the beginning of the hydrological year in Johor. This extended averaging period smooths short-term fluctuations and captures the antecedent soil moisture conditions more realistically, leading to a stable initialization for the stochastic storage dynamics. The stochastic water balance model as in Figure 10, 11 and 12 integrates simulated rainfall (via the SOU process), evapotranspiration, and runoff using a dynamic storage equation without an upper bound constraint based on Eq. (6). This approach allows water storage to evolve naturally according to hydrological forcing and depletion through evapotranspiration and drainage.

The simulation for Senai (Figure 10) shows a steady accumulation of storage during the early simulation years, followed by stabilization around 1000 mm. This reflects Senai's consistently wet climatic regime and urbanized surface characteristics, where impervious areas promote rapid runoff but continuous rainfall sustains recharge. The stochastic rainfall inputs reproduce seasonal oscillations and interannual variability, while storage fluctuations remain moderate, indicating stable hydrological equilibrium under persistent rainfall.

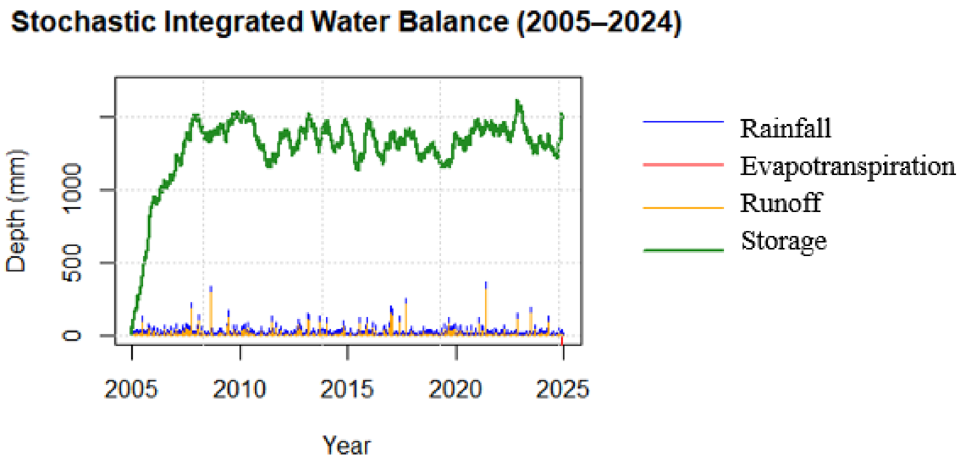

Figure 11: Senai water balance components

In Kluang (Figure 11), the model shows stronger seasonal fluctuations and higher storage amplitude, reaching approximately 1500 mm. This outcome aligns with Kluang's higher rainfall intensity and agricultural–forested land mix, which enhances infiltration and delayed drainage. The Seasonal OU-driven rainfall variability captures both wet-season recharge and dry-season depletion, producing a more dynamic storage pattern. The higher amplitude reflects strong sensitivity to monsoonal cycles, particularly the Northeast Monsoon, consistent with observed rainfall extremes in this district.

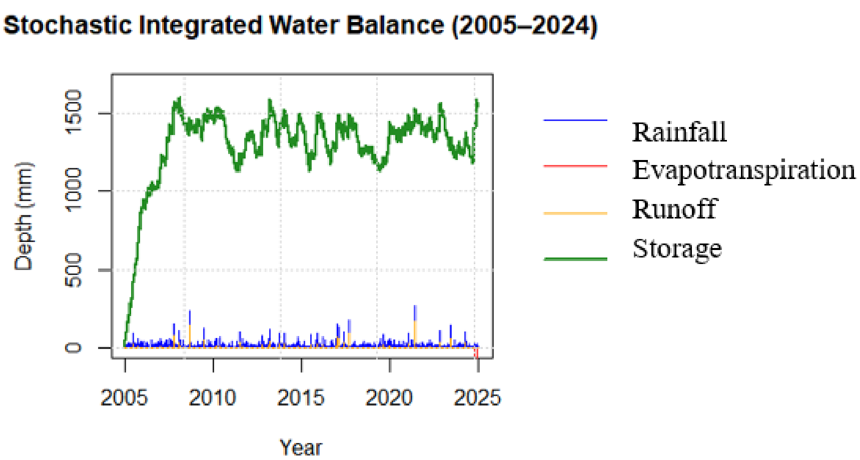

Figure 12: Kluang water balance components

The Batu Pahat simulation (Figure 12) exhibits a similar seasonal pattern to Kluang but with slightly lower overall storage levels, stabilizing between 1000–1400 mm. This pattern is influenced by the district's predominantly agricultural landscape and lower rainfall amplitude, leading to less pronounced recharge events compared to Kluang. The stochastic rainfall–ET balance reflects moderate hydrological variability, suggesting a balanced water regime where rainfall input and evapotranspiration losses are relatively well-matched over time.

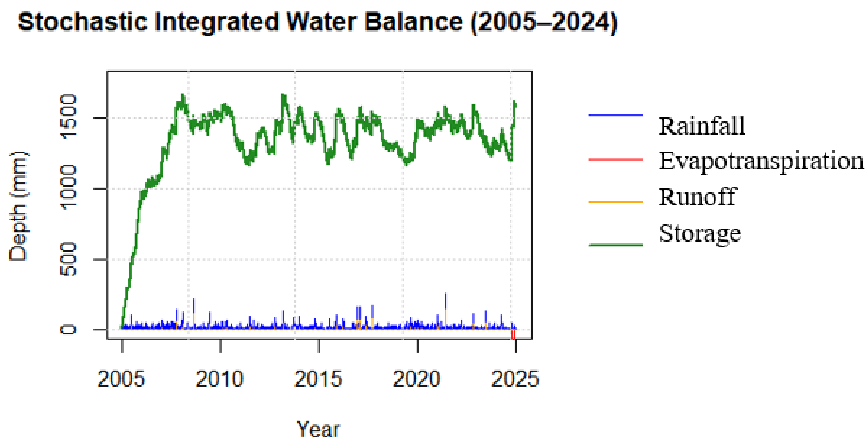

Figure 13: Batu Pahat water balance component

Across all three districts, the stochastic water balance model demonstrates the effectiveness of the Seasonal OU process in capturing realistic hydrological behavior, including seasonal storage oscillations, interannual variability, and stochastic rainfall effects. The results highlight spatial contrasts Kluang's high variability, Senai's consistent wetness, and Batu Pahat's moderate response which are consistent with observed climatic and land-use differences. This underscores the model's capacity to simulate district-specific hydrological dynamics under variable climatic conditions.

Practically, the integration of stochastic rainfall modeling with the deterministic water balance framework bridges the gap between statistical rainfall analysis and hydrological simulation. This coupling allows for dynamic estimation of the water balance under uncertain climatic conditions, enhancing the understanding of how stochastic rainfall variability propagates through evapotranspiration, runoff, and storage components. For instance, the amplitude-driven rainfall fluctuations observed in Kluang directly lead to higher variability in storage and runoff, emphasizing the district's sensitivity to monsoon transitions. Meanwhile, Senai's persistently wet regime supports more stable water storage but elevates runoff risk in urbanized areas, whereas Batu Pahat's moderate rainfall pattern reflects a more balanced but less resilient hydrological response during dry spells. This integrated approach provides valuable insights for managing water availability, reservoir operations, and irrigation scheduling under Malaysia's increasingly variable monsoon climate.

5. Conclusion

This study presents an integrated framework that combines statistical analysis, stochastic differential equations (SDEs), and water balance modeling to analyze rainfall variability and hydrological dynamics across Senai, Kluang, and Batu Pahat from 2005 to 2024. Statistical results show that rainfall in all three districts is stationary, non-normally distributed, and highly seasonal, with clear monsoonal influence but no significant long-term trend. These characteristics justify the use of the Seasonal Ornstein–Uhlenbeck (SOU) process, which effectively represents rainfall as a mean-reverting system with periodic variation and random fluctuations.

By coupling the SOU-based rainfall simulations with the water balance model, this study establishes a direct link between rainfall variability and hydrological response specifically evapotranspiration, runoff, and storage. The integration demonstrates how stochastic rainfall fluctuations propagate through the hydrological system, influencing water availability and balance across districts with differing land uses and climatic exposures.

Senai exhibited the highest overall rainfall and evapotranspiration, reflecting its humid, urbanized environment where impervious surfaces promote rapid runoff but consistent rainfall sustains storage around 1000 mm. Kluang experienced the strongest rainfall amplitude and most pronounced seasonal oscillations, with higher recharge and storage (up to 1500 mm) due to its mixed agricultural–forested landscape. In contrast, Batu Pahat displayed moderate rainfall intensity and the most balanced hydrological behavior, with storage fluctuating between 1000 and 1400 mm. These district-specific outcomes highlight that both climatic forcing and land surface characteristics significantly shape water balance responses.

Practically, the stochastic–deterministic integration bridges the gap between statistical rainfall analysis and physical hydrological simulation. It enables dynamic estimation of the water balance under uncertainty, which is vital for reservoir operation, irrigation scheduling, and drought management. For example, the amplitude-driven rainfall variability in Kluang directly translates to stronger fluctuations in storage and runoff, indicating the need for adaptive strategies to buffer seasonal water scarcity. Conversely, Senai's persistently high rainfall supports stable recharge but necessitates enhanced flood management, while Batu Pahat's moderate regime suggests relatively steady but less resilient storage during prolonged dry periods.

From a theoretical perspective, the incorporation of SDEs particularly the Seasonal Ornstein–Uhlenbeck process enhances the realism of hydrological modeling by capturing both deterministic seasonal behavior and stochastic variability inherent in tropical rainfall. This approach may be further extended by employing alternative stochastic formulations, such as Langevin-type or Lévy-driven SDEs, to better represent extreme rainfall or nonlinear recharge dynamics.

Overall, the findings underscore the effectiveness of integrating stochastic and deterministic models for comprehensive hydrological assessment. This synthesis improves understanding of rainfall–storage interactions and provides a robust foundation for sustainable water resource planning in monsoon-driven environments like Johor, where rainfall variability strongly governs water availability, demand, and long-term resilience.

Acknowledgments

The authors would like to thank Universiti Teknologi Malaysia, UTM Encouragement Research grant (Q.J130000.3809.31J69) and Ministry of Higher Education, Fundamental Research Grant Scheme (FRGS/1/2024/STG06/UTM/02/14) for their partial financial support.

References

Adib M.N.M., Harun S. & Rowshon M.K. 2022. Long-term rainfall projection based on CMIP6 scenarios for Kurau River Basin of rice-growing irrigation scheme, Malaysia. *SN Applied Sciences* **4**(3): 70.

Ahmad A.A., Yusof F., Mispan M.R. & Kamaruddin H. 2017. Rainfall, evapotranspiration and rainfall deficit trend in Alor Setar, Malaysia. *Desalin Water Treat* **13**: 400–404.

Alcober C.A. & Macuha R. 2024. Development and analysis of a GIS-based curve number map of the Philippines. *ASEAN Engineering Journal* **14**(2): 173–181.

Allen R.G., Pereira L.S., Raes D. & Smith M. 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Food and Agriculture Organization of the United Nations: Rome.

Amin I.M.Z. bin M., Ercan A., Ishida K., Kavvas M.L., Chen Z.Q. & Jang S.H. 2019. Impacts of climate change on the hydro-climate of peninsular Malaysia. *Water* **11**(9): 1798.

Chow V.T., Maidment D.R. & Mays L.W. 1988. *Applied hydrology*. McGraw-Hill.

Hao X., Zhang S., Li W., Duan W., Fang G., Zhang Y. & Guo B. 2018. The uncertainty of Penman-Monteith method and the energy balance closure problem. *Journal of Geophysical Research: Atmospheres* **123(**14): 7433–7443.

Houénafa S.E., Ronoh E.K., Johnson O. & Moore S.E. 2025. Lévy-induced stochastic differential equation models in rainfall–runoff systems for assessing extreme hydrological event risks. *Stochastic Environmental Research and Risk Assessment* **39**(4): 1537–1554.

Khan M.M.H., Mustafa M.R.U., Hossain M.S., Shams S. & Julius A.D. 2023. Short-term and long-term rainfall forecasting using ARIMA model. *International Journal of Environmental Science and Development* **14**(5): 292–298.

Khor J.F., Lim S. & Ling L. 2023. Evaluating the effect of deforestation on decadal runoffs in Malaysia using the revised curve number rainfall runoff approach. *Water* **15**(7): 1392.

Koutsoyiannis D. 2016. Generic and parsimonious stochastic modelling for hydrology and beyond. *Hydrological Sciences Journal* **61**(2): 225–244.

Liao Y. & Yin G. 2022. Study on Short-term and Imminent Precipitation Prediction Model and Accuracy Evaluation Based on Depth Learning. *Journal of Physics: Conference Series* **2410**(1): 012018.

Malaysian Meteorological Departmen. (n.d.). Official portal of the Malaysian Meteorological Department (MetMalaysia). https://www.met.gov.my/

Mohd Faisal M.S., Jamal M.H., Mohamad K.A. & Juneng L. 2023. Investigating SWAT Model Efficiency to Determine Water Balance Components (Case Study: Sungai Muda Watershed). *Proceedings of the 5th International Conference on Water Resources (ICWR) – Volume 1. Lecture Notes in Civil Engineering*, pp. 123–135. Springer: Singapore.

Muhammad N.S., Abdullah J. & Julien P.Y. 2020. Characteristics of rainfall in peninsular Malaysia. *Journal of Physics: Conference Series*, pp. 52014.

NASA POWER Project. (n.d.). NASA POWER data access viewer. https://power.larc.nasa.gov/data-access-viewer/

Nguyen T.H.T., Bennett, B. & Leonard, M. 2023. Evaluating stochastic rainfall models for hydrological modelling. *Journal of Hydrology* **627**: 130381.

Northrop P.J. 2024. Stochastic models of rainfall. *Annual Review of Statistics and Its Application* **11**: 51–74.

Park J., Onof C. & Kim D. 2019. A hybrid stochastic rainfall model that reproduces some important rainfall characteristics at hourly to yearly timescales. *Hydrology and Earth System Sciences* **23**(2): 989–1014.

Prottasha N.J., Tahabilder A., Kowsher M., Mia M.S. & Kobra K.T. 2023. Short-term rainfall prediction using supervised machine learning. *Advances in Technology Innovation* **8**(2): 111–120.

Sapozhnikov V.B. & Foufoula-Georgiou E. 2007. An exponential Langevin-type model for rainfall exhibiting spatial and temporal scaling. *Nonlinear dynamics in geosciences*, pp. 87–100. Springer: New York.

Suhaila J., Deni S.M., Zin W.Z.W. & Jemain A.A. 2010. Trends in peninsular Malaysia rainfall data during the southwest monsoon and northeast monsoon seasons: 1975–2004. *Sains Malaysiana* **39**(4): 533–542.

Sun D., Wu J., Huang H., Wang R., Liang F. & Xinhua H. 2021. Prediction of Short-Time Rainfall Based on Deep Learning. *Mathematical Problems in Engineering* **2021**(1): 6664413.

Tong Z. & Liu A. 2021. A censored Ornstein–Uhlenbeck process for rainfall modeling and derivatives pricing. *Physica A: Statistical Mechanics and its Applications* **566**: 125619.

Tukimat N.N.A., Harun S. & Shahid S. 2012. Comparison of different methods in estimating potential evapotranspiration at Muda Irrigation Scheme of Malaysia. *Journal of Agriculture and Rural Development in the Tropics and Subtropics* **113**(1): 77–85.

USDA Soil Conservation Service. 1972. *National engineering handbook, section 4: Hydrology.* U.S. Department of Agriculture: Washington.

Yoshimi K., Yamada T. & Yamada T.J. 2015. Assessment of uncertainty in rainfall-runoff analysis incorporating stochastic differential equation. *Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering)* **71**(4): I_259-I_264.

Department of Mathematical Sciences
Faculty of Science
Universiti Teknologi Malaysia Johor Bahru
81310 UTM Johor Bahru
Johor, MALAYSIA
*E-mail: anizaakaram@graduate.utm.my, arifah@utm.my**

UTM-Centre or Industrial and Applied Mathematics,
Ibnu Sina Institute of Scientific and Industrial Research
Universiti Teknologi Malaysia Johor Bahru
81310 UTM Johor Bahru
Johor, MALAYSIA
E-mail: arifah@utm.my

Received: 11 July 2025
Accepted: 18 May 2026

*Corresponding author