

Research Article:

Optimizing Precipitation Prediction at Dukan Dam through a Comparative Climate and Temporal Modeling with Multi-Factor Statistical Analysis"

Hemn Abdullah Wasta Ahmed¹

¹Dukan Technical Institute, Sulaimani Polytechnic University, Sulaymaniyah, Kurdistan Region, Iraq
<https://doi.org/10.31530/cjnst.2026.2.2.6>

Received: 26 March 2026

Revised: 29 June 2026

Accepted: 07 July 2026

***Corresponding author info:** Hemn Abdullah Wasta Ahmed, Email: (hemn.ahmed@spu.edu.iq)
Copyright © 2026 Hemn Abdullah Wasta Ahmed, Charmo Journal of Natural Sciences (CJNST).

This is an open access article published under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits use, distribution, and reproduction in any medium, provided the original work is properly cited.



Abstract

Rainfall forecasting serves as a vital tool for mitigating droughts and floods and managing water resources in areas where rainfall predictability is prone to being inadequate. This study evaluates the effectiveness of the precipitation forecast models for the Dukan Dam catchment area in northern Iraq by combining historical rainfall data with key meteorological factors, including temperature, humidity, and wind speed. To achieve this objective three modeling approaches were applied: SARIMA, Prophet, and XGBoost, using monthly data from 2015 to 2025. With feature engineering and lag factors included, the XGBoost model performs superior to the others, obtaining the lowest RMSE (normalized 0.0416; 1.13 mm) and MAE (normalized 0.0296; 0.80 mm). The forecast accuracy is found to be higher when using multi-factor climatic inputs. In areas with limited resources, the suggested framework serves as a scalable tool for climate adaptation, planning with XGBoost significantly outperforming statistical models.

Keywords: Machine Learning, XGBoost, Hydrological Modeling, Climate Change, statistical modeling

1. Introduction

In Iraq's Kurdistan Region, Dukan Dam is an essential reservoir for drinking water, irrigation, and a source of hydroelectric electricity. Accurate rainfall forecasting supports long-term water security planning, flood mitigation, and efficient reservoir management. Traditionally, researchers have relied on physical models based on atmospheric processes for this purpose [1-3]. The climatic and atmospheric conditions in the Dukan Dam basin influence the precipitation regime. Variations in air temperature affect evaporation rates and the atmosphere's capacity to retain water vapor, where warmer air masses can hold greater amounts of moisture and potentially intensify precipitation events [4]. Relative humidity also plays a significant role in cloud formation and rainfall generation, as higher humidity levels enhance condensation processes and increase the likelihood of precipitation. Furthermore, atmospheric pressure and wind circulation patterns influence the transport of moist air masses from the Mediterranean region toward

northern Iraq [5]. Seasonal variations in these climatic elements contribute to temporal fluctuations in rainfall patterns across the Kurdistan Region. Recent climate change impacts, including rising temperatures and irregular humidity distributions, have increased the uncertainty of precipitation behavior, emphasizing the importance of accurate rainfall forecasting for sustainable water resource management [6].

Since the relationship between the input factors (humidity, wind, and temperature) and precipitation is not a straight-line relationship, several models were explored to identify the most accurate one to predict future precipitation. The reliability of traditional forecasting methods in essential operational contexts is compromised by the fact that they are frequently not able to capture the nonlinear nature of precipitation patterns, as they are physical process dependent [7]. Sophisticated forecast methods have been introduced as hydrological cycles are increasingly being complicated due to climate change to reduce the hazards of events like droughts and floods. [8, 9]. Researchers have employed Innovative Trend Analysis (ITA) and Innovative Polygon Trend Analysis (IPTA), contributing to a deeper

understanding of rainfall variability and its implications throughout a set of long datasets. Machine learning models are increasingly prominent [10, 11].

The primary contribution of this study is the establishment of a multi-factor, machine learning-based forecasting framework customized for local data that are operationally crucial, revealing that many meteorological factors could potentially be effectively combined for improved forecasting.

The remainder of this paper is organized as follows: Section 2 examines the related previous research to this study. Section 3 describes the study area, data acquisition, preprocessing steps, and modeling framework. Section 4 presents the results of the three forecasting models and a comparative discussion. Section 5 concludes the paper, summarizes key findings, and suggests directions for future research.

2. Related work

Statistical time-series models remain a fundamental part of hydrological prediction. The autoregressive integrated moving average (ARIMA) was widely applied to rainfall forecasting, as it is able to model temporal dependence. However, precipitation shows a strong seasonal cycle, and in order to address this limitation, the seasonal ARIMA was introduced by merging autoregressive and moving average elements. Studies have shown that SARIMA can capture periodic hydrologic behavior for short time intervals [12]. SARIMA (Seasonal Autoregressive Integrated Moving Average) extends ARIMA to handle seasonality components for both non-seasonal and seasonal patterns. It is based on a combination of autoregressive (AR), integrated (I), and moving average (MA) components, both non-seasonal and seasonal. In general terms, the model predicts the current value based on past values and past forecast errors. The model results in a satisfactory output for short-term prediction [13]. The seasonality cycle is a witnessed pattern for rain in the region, so the model is adopted from three comparative models to predict the precipitation.

The second model is the Prophet forecasting framework developed by Meta Platforms, which provides an additive regression model developed by Facebook that decomposes time series into trend, seasonality, and holiday components, performing efficiently with strong seasonality. Prophet has been increasingly applied because it handles missing data efficiently [14–16]. Studies have applied it for long-time climate history, as it decomposes the time series into long-term trend, yearly seasonality, and irregular variations. However, researchers have observed that Prophet

may struggle with nonlinear shifts, and thus it was applied with support vector regression (SVR) and random forest (RF) models; these models provide satisfactory predictions of nonlinear hydrological and meteorological processes [17, 18].

Despite these advantages, statistical models assume linear relationships, limiting their ability to represent nonlinear atmospheric interactions. When applied to short historical records, which are usual in hydrological contexts, multiple investigations show overfitting, and their ambiguous nature makes them tricky for water managers to grasp. [19]. Machine learning models are increasingly prominent in hydrological forecasting due to their capacity to model nonlinear hydrometeorology. Following the emergence of computational capability. However, machine learning models require large datasets and are often seen as "black boxes," limiting their adoption in operational contexts where interpretability is crucial [20]. For example, through the employment of ANN models, specifically Artificial Neural Networks (ANN) and Neuro-Fuzzy Inference Systems (ANFIS), precipitation was predicted based on the temperature, wind, and pressure; the results were accepted, and ANFIS outperformed the ANN [21]. Other researchers have used ANN modeling to predict monthly precipitation at one station using measurements from neighboring stations [22]. Higher predictive accuracy is achieved in hydrological forecasting through the extreme Gradient Boosting algorithm (XGBoost) due to gradient boosting optimization, regularization against overfitting, ability to learn nonlinear interactions. [23]. Hydrological applications demonstrate that XGBoost performs particularly well when multiple climate predictors are available, enabling extraction of complex relationships between atmospheric variables and rainfall patterns [24].

Despite these advances, several gaps remain, such as fewer studies having focused on integrated rainfall forecasting for the Dukan region. Most models incorporate only one or two climatic indices, neglecting the combined effects of multiple factors. A comparative analysis between statistical, decomposable, and machine learning paradigms provides an optimized precipitation framework for the region.

3. Materials and Methods

3.1. Study Area and Climate

The Dukan Dam is located at the coordinates (35°57'15"N, 44°57'10" E) in north Iraq as shown in Figure 1.

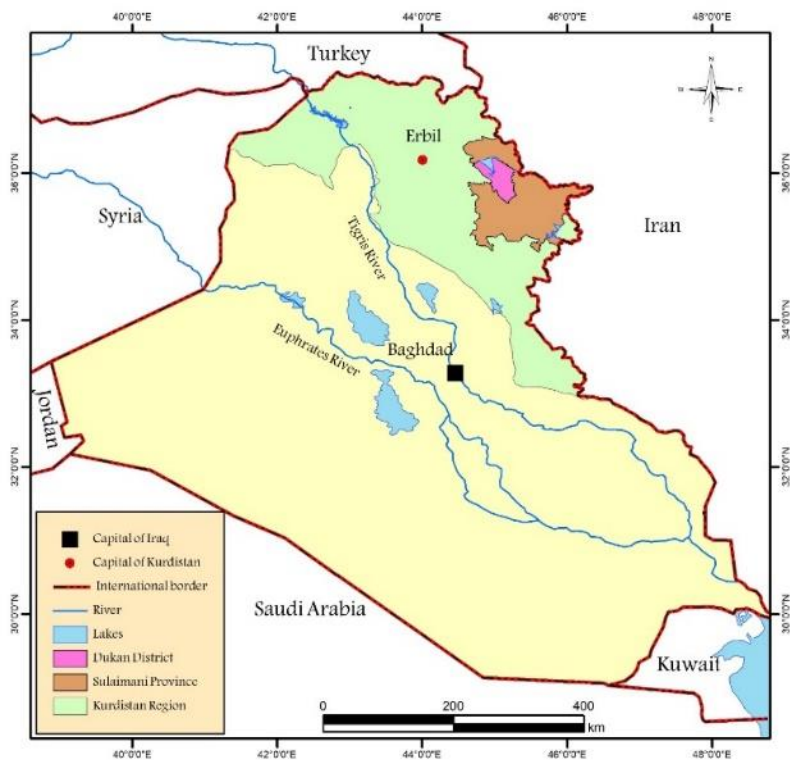


Figure 1: Dukan District boundary within Republic of Iraq

The district is located in Sulaymaniyah governorate. The region has a semi-arid climate with very hot, dry summers and cool, wet winters. The location of the district

within Kurdistan region in North of Iraq is shown in Figure 2.

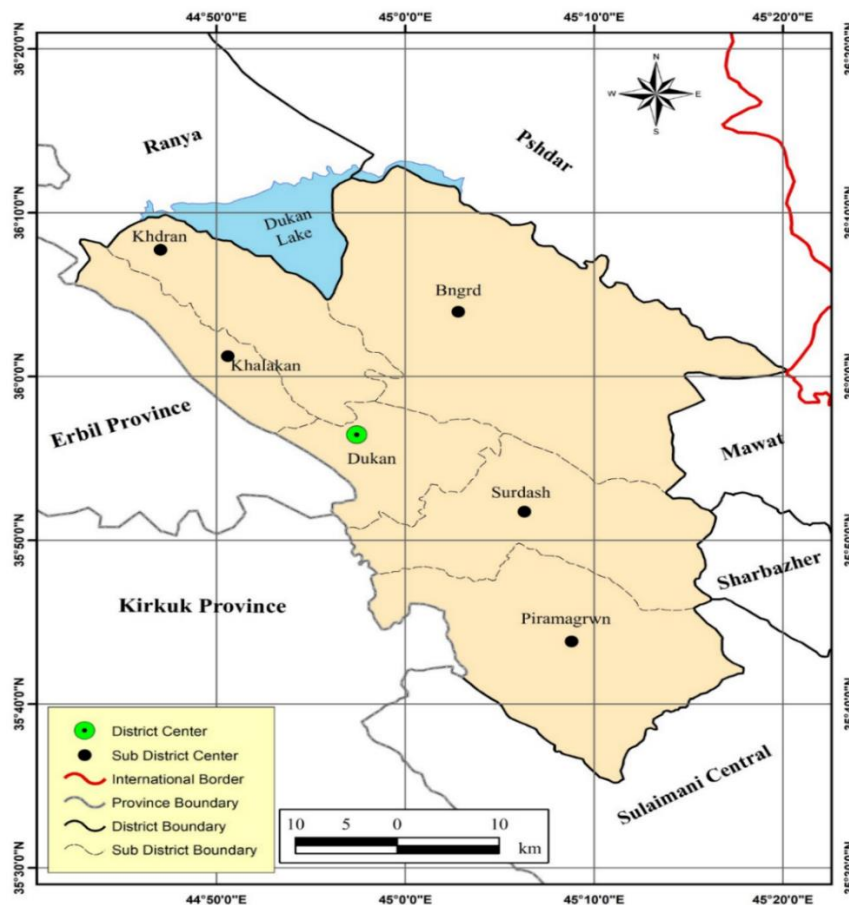


Figure 2 Dukan District boundary within the Kurdistan region [25]

3.2. Data Acquisition

The target is to obtain the most accurate predictive model among different machine learning

through exploring; the generalized steps are shown in Figure 3.

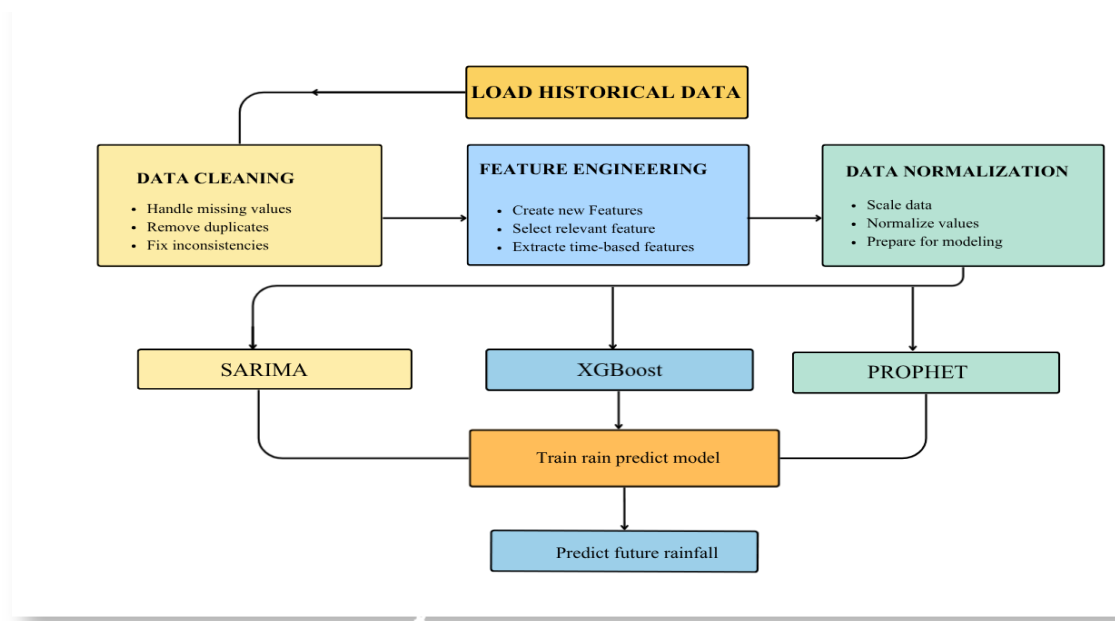


Figure 3 Generalized steps for rainfall in Dukan District

In order to achieve higher accuracy in machine-learning prediction, a wide range of data is required to capture the seasonality and non-linearity. The access to the data was through the weather station department in Dukan Dam directory. The data consists of monthly records from January 2015 to December 2025 (132 months). The data is classified into four metrics, which are the amount of rainfall (mm), average temperature ($^{\circ}\text{C}$), humidity (%), and wind speed (m/s) as shown in Figure 4 prior to normalization. It is generally acknowledged that temperature influences the

atmospheric behavior and its grand effect on evaporation; humidity is the amount of moisture in the air that is impacted by the wind movement direction. These physical elements provide a solid ground to establish a rain-predictive and forecasting model.

Prior to any transformation like standardization, this raw data delivers an essential overview for each indicator and initial interconnection as seen in Figure 4. The plot allows visual inspection of scale differences, missing values, and preliminary seasonal patterns prior to preprocessing.

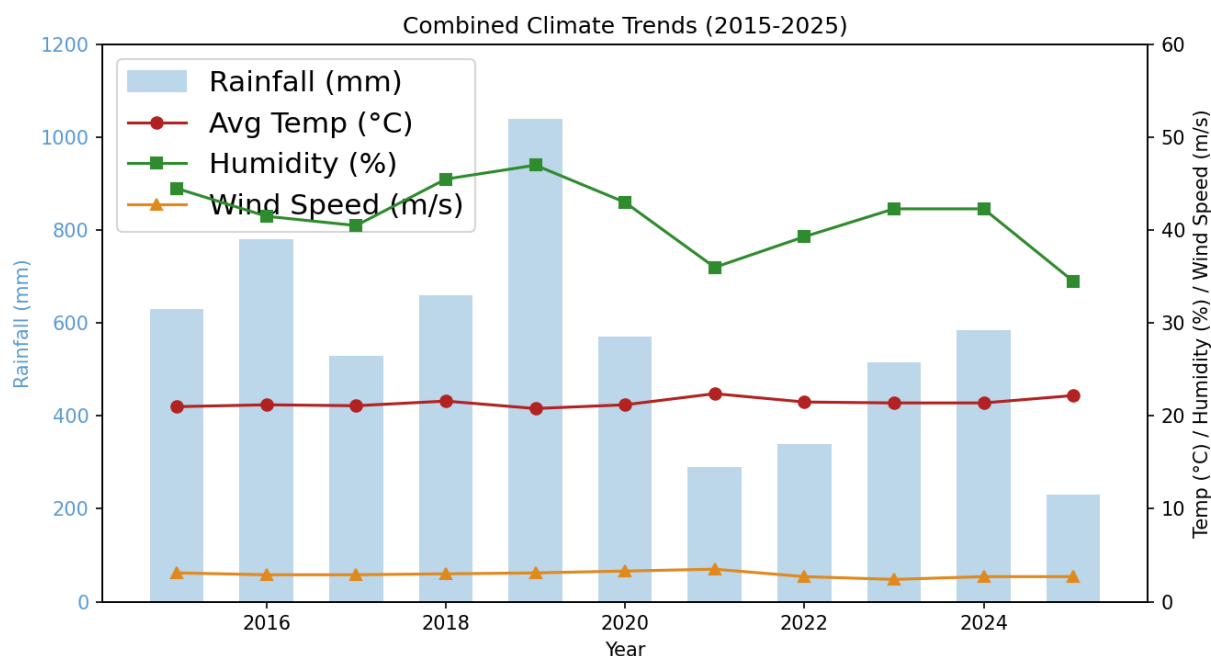


Figure 4 Raw monthly time series of rainfall, temperature, humidity, and wind speed before normalization.

for all indicators and their initial connections. Loading the historical data in the Google Colab environment, the general trends in the raw data before normalization are

visualized in Figure 5. Each subplot highlights trend direction, seasonality, and recent climatic changes (e.g., declining rainfall and humidity after 2020)

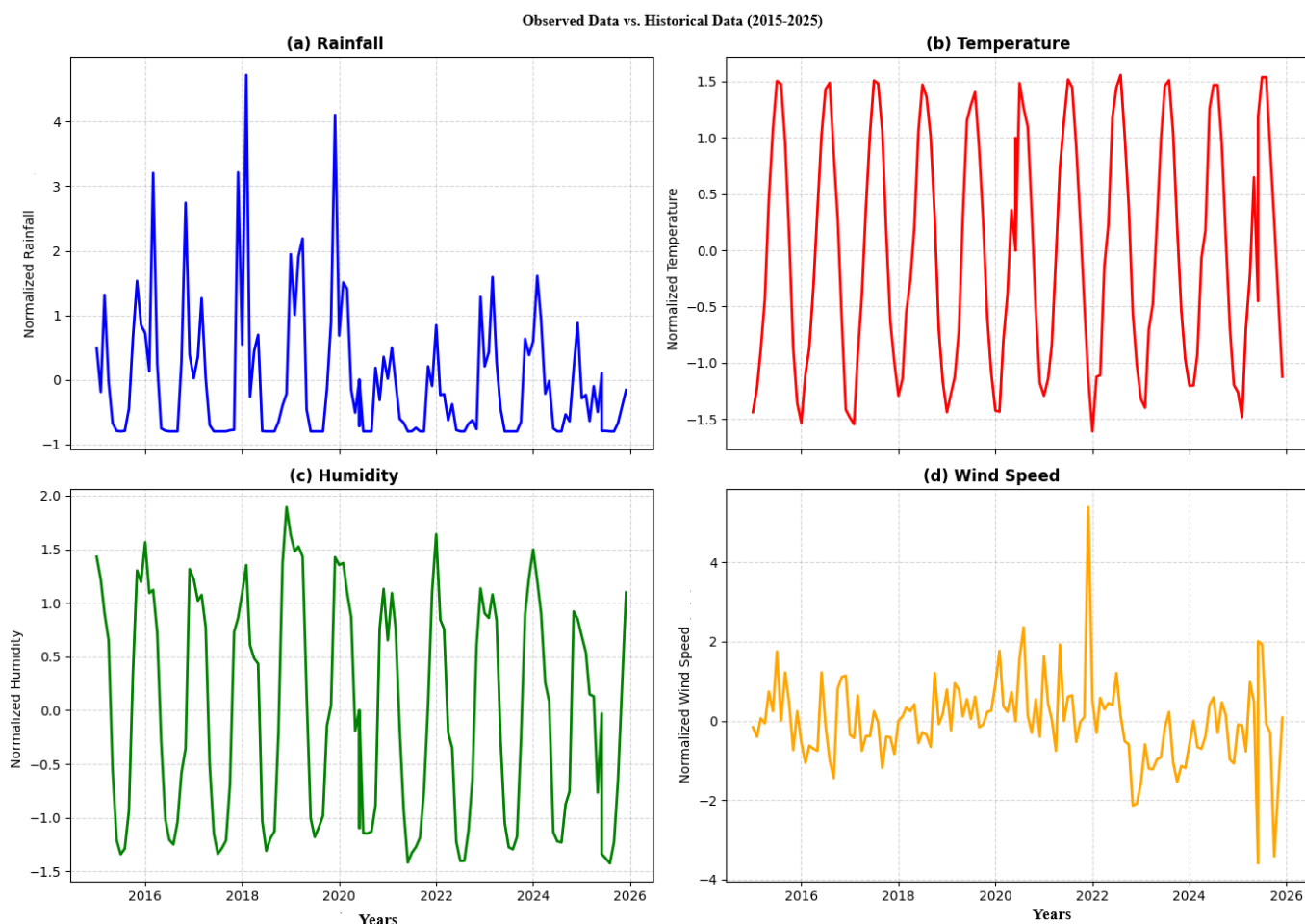


Figure 5 Historical time-series of normalized (a) Rainfall, (b) Temperature, (c) Humidity, (d) Wind Speed from 2015 to 2025.

The interdependence relationship between meteorological variables was evaluated:

- Temperature is the most visible one as it shows an annual cycle inversely correlated with rainfall.
- The precipitation metrics follow it. The figure clearly provides insight into patterns between earlier and recent dates, showing a declining trend after 2020.
- Humidity indicates the shift to a drier, warmer climate recently.
- The wind speed shows greater variability in recent years.

3.3. Data Preprocessing

Data preprocessing was performed through the employment of different libraries such as pandas to read the data

from the CSV file and convert it into a DataFrame structure (df and df-normalized), which is a basic way to recognize, label, and access the data. For efficient numerical calculations, the NumPy library is used in Google Colab. Missing values (less than 2% of the dataset) were handled using linear interpolation, and the data is normalized by z-score standardization to ensure the variables of different units contribute equally to the training process. The monthly data complies very well with time series for benchmarking a robust forecast model for weather evaluation in order to manage the future anticipated droughts, to evaluate the metrics in precipitation prediction data, and adopt the one with the most compatibility within the training session. In Figure 6, a sample of rainfall comparison before and after normalization is shown, so the preprocessing validity is verified.

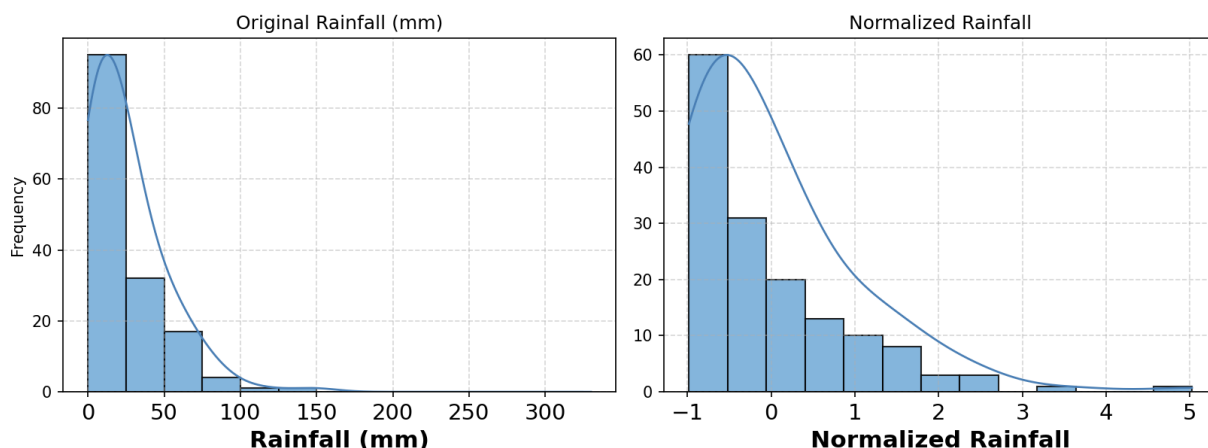


Figure 6 Comparison of rainfall data before and after normalization. Left side: Original rain value skewed distribution. Right: rain fall after z-score normalization

After preprocessing, the final dataset contained 132 months of complete records, which were used for training (2015–2023; 108 samples) and testing (2024–2025), 24 samples) after accounting for lag feature creation.

3.4. Feature Engineering and Modeling Framework

The following preprocessing steps were implemented:

Creating lagged features at (1,3,6,12) months. They were chosen based on partial correlation to capture short-term (1,3) months and long-term (6–12) months to capture characteristic dependencies of the region's precipitation regime. All features are normalized with the target variables. Rolling window statistics, mean and standard deviation over 3 and 6 months were computed to smooth out short-term fluctuations and capture recent trends.

Time-based features, including month, year, and seasonal indicators, were extracted. These features enable explicit learning of seasonal patterns driven by Mediterranean weather systems, which produce distinct wet winters and dry summers in the Dukan Dam catchment. The models are trained in three separate files: Sarima, Prophet, and X-Boost. The trained model is employed to predict the future rain and generate a plot comparing the historical rain data with the predicted future rain. The data set contains 132 monthly samples. Because lag feature 12 observation only, the first 12 observations were removed, leaving 120 samples. The training/testing split was chronological: 2015–2023 for training, 2024–2025 for testing. SARIMA and Prophet were implemented using their standard workflows with the same chronological split for fair comparison. The generalized steps for both models are shown side by side to illustrate different mechanisms of both models in Figure 7

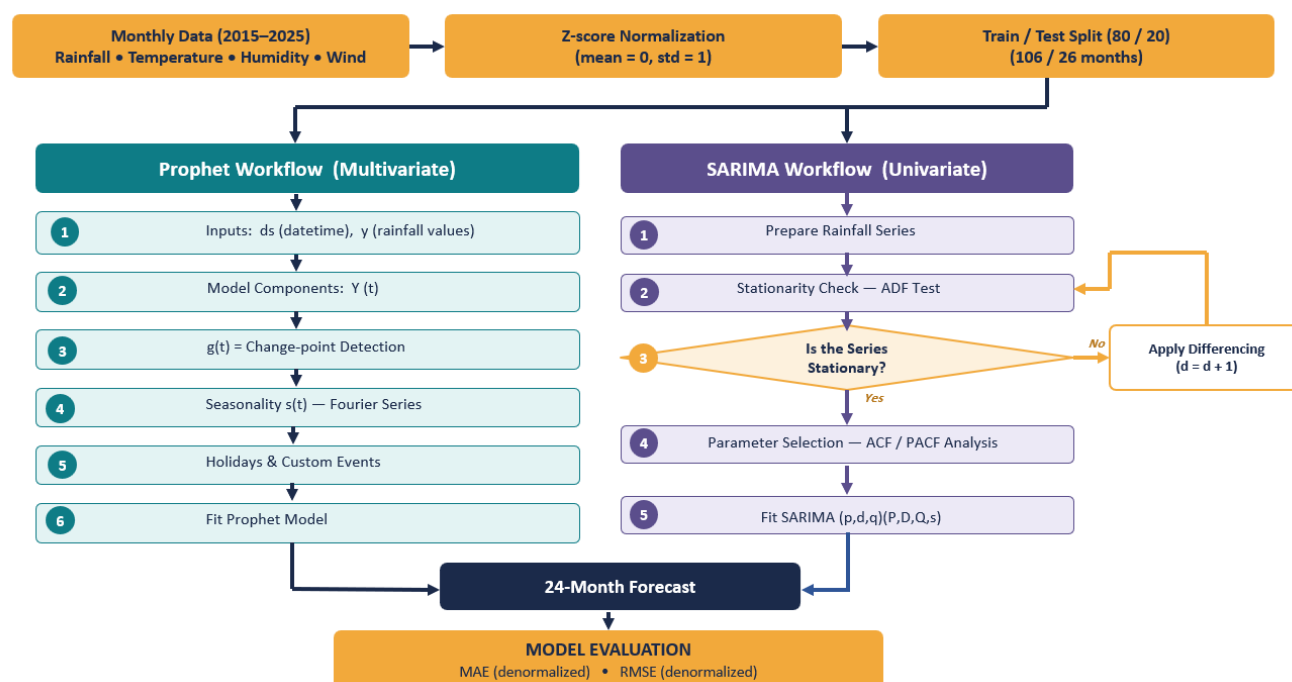


Figure 7: Prophet and SARIMA modeling framework.

The XGBoost model was trained with hyperparameter tuning using time series cross-validation with a three-fold split.

3.4.1. Prophet model

It optimizes parameter tuning for time-series forecasting and requires a data frame with date-indexed 'ds' and numerical objective 'y' columns. [16], It incorporates seasonality and exogenous predictors and performs efficiently. [15, 26]. The workflow is initiated by targeting the rainfall series and converting it into a data frame with two columns: ds (for dates), and y (for the rainfall values), which is the format required by Prophet. Prophet uses an additive regression model with three main components: trend, seasonality, and holidays. As shown in Equation 1 [27] is shown.

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t) \quad (1)$$

Where:

$y(t)$: the observed value of the time series at time t .

$g(t)$: the trend component, the non-periodic changes in the time series

$s(t)$: the seasonal component, representing monthly changes

$h(t)$: the holiday component, which accounts for the impact of holidays or custom events

$\epsilon(t)$: is the error(residua) term

The trend parameters are governed by the flexibility of the trend and the numbers of possible changepoints. Higher values allow more frequent trend changes, while lower values produce smoother trends. The value of 0.05 was selected to balance trend adaptation with overfitting prevention, and the changepoints of 25. The number of potential changepoints in the trend is automatically determined based on the time series length. A higher value of seasonality allows more flexible patterns, and in this case, it is equal to 10. Uncertainty parameters are set to 80% prediction intervals, reflecting the model's confidence in future predictions.

3.4.2. SARIMA Model

The SARIMA model extends the ARIMA model to handle seasonality. It is denoted as SARIMA(p,d,q)(P,D,Q)_s, where:

- (p, d, q) are the non-seasonal AR, I, and MA orders.
- (P, D, Q) are the seasonal AR, I, and MA orders.
- S is the length of the seasonal cycle.

The general form of a SARIMA model can be quite complex, but conceptually, it involves differencing the series (both non-seasonal and seasonal) to make it stationary and then fitting autoregressive and moving average terms to the differenced series.

A simplified representation of a SARIMA model (without explicitly writing out the polynomial lags) can be seen in Equation 2 [13]

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - \phi_1 L^s - \dots - \phi_p L^{ps})(1 - L)^d(1 - L^s)^D y_t = (1 + \theta_1 L + \dots + \theta_q L^q)(1 + \theta_1 L^s + \dots + \theta_q L^{qs}) \epsilon_t \quad (2)$$

Where:

y_t : is the time series value at time t .

L : is the lag operator ($Ly_t = y_{t-1}$).

(ϕ, θ) : Seasonal coefficient

$(1 - \phi_1 L - \dots - \phi_p L^p)$ represents the non-seasonal autoregressive (AR) component at p (nonseasonal AR order)

$(1 - \phi_1 L^s - \dots - \phi_p L^{ps})$ represents the seasonal autoregressive (SAR) component of order p

$(1 - L)^d$ performs non-seasonal differencing of order d (nonseasonal differencing order)

$(1 - L^s)^D$ performs seasonal differencing of order D (Seasonal differencing order) with periodicity S (The seasonal period).

$(1 + \theta_1 L + \dots + \theta_q L^q)$ represents the non-seasonal moving average (MA) component of order q (highest lag of past error).

$(1 + \theta_1 L^s + \dots + \theta_q L^{qs})$ represents the seasonal moving average (SMA) component of order Q (highest seasonal lag).

ϵ_t is white noise error term at time t .

The general steps to predict the future rainfall in the Dukan district are, firstly, to load rainfall, humidity, wind direction, and temperature data into a Pandas DataFrame. Then separate the features (humidity, wind direction, temperature) from the target variable (rainfall). Then, split the data into 108 training samples (starting from 2015 to 2023) and 18 testing validation sets for the years 2024 and 2025 to prepare it for the XGBoost model.

The basic mathematical equation for this model is shown in Equation (3) [13, 17]

$$Fk(x) = Fk - 1(x) + \alpha hk(x) \quad (3)$$

(x) is the final prediction for input x after k iterations.

$Fk-1(x)$ is the prediction from the ensemble of the previous $k-1$ trees.

$hk(x)$ is a new weak learner (a decision tree) that attempts to predict the residual error from $Fk-1(x)$.

α is the learning rate, controlling the contribution of each new tree.

The goal at each step is to minimize a loss function, $L(y_i, Fk-1(x_i) + hk(x_i))$, where y_i is the actual value. This optimization is performed using gradient descent principles.

4. Result and Discussion

The monthly average data were normalized to prevent feature dominance for every parameter, then analyzed for seasonal patterns. The month was extracted from the 'Date' field and, grouped by month, after which the mean for each normalized weather metrics. Rainfall, average temperature, humidity, and wind speed was plotted as a separate time series. And a line plot was generated against date. In addition, a single 2D line chart plotting all four normalized metrics against date for direct comparison. As shown in Figure 8, when temperature (orange) peaks, humidity (green) and rainfall (blue) typically bottom out.

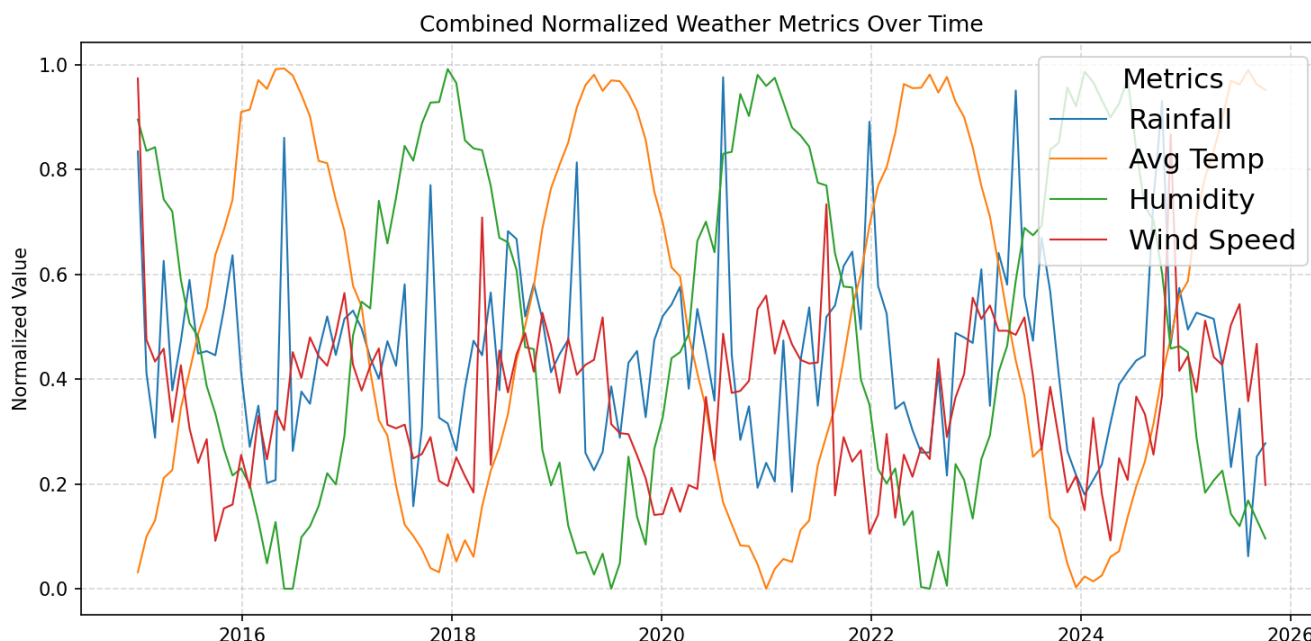


Figure 8 Normalized time-series plots for weather metric

4.1. Linear Regression Model

The simple linear Regression model with only two delayed features fails to adequately capture the natural variability or more complex patterns in this type of time-series data, as it fails to capture the seasonality and the noteworthy highs and lows, in contrast to the actual data. as, seen in

Figure 9, more sophisticated time series should be employed in order to capture seasonality. Linear regression is employed to emphasize the superior outcome of the other time series.

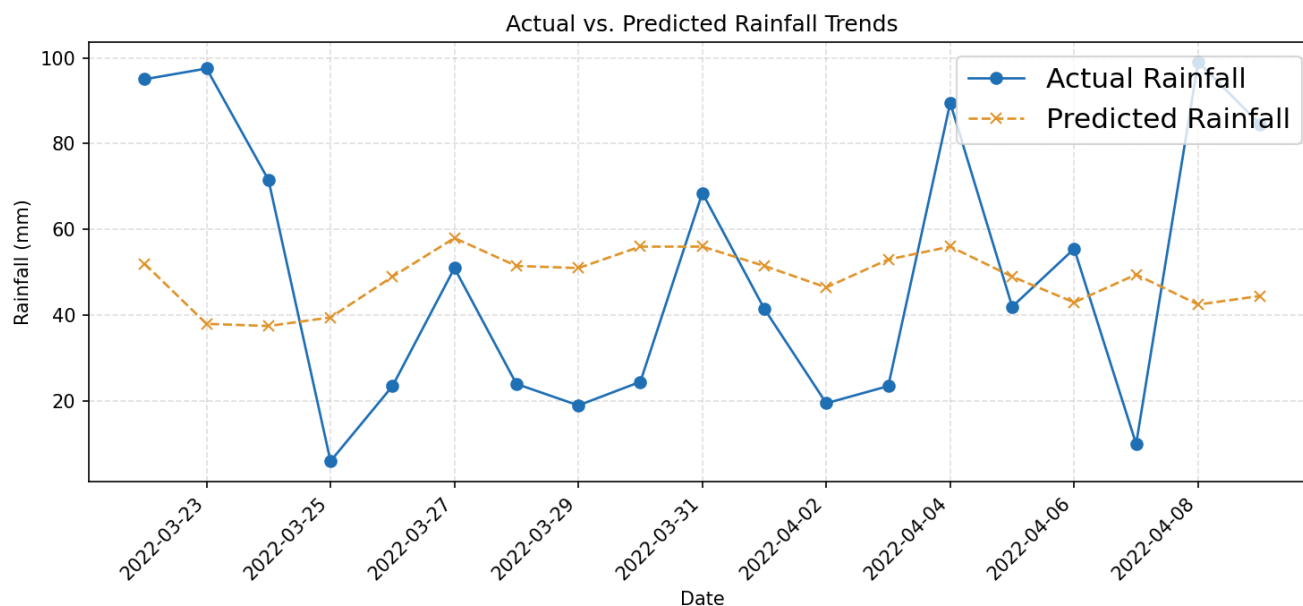


Figure 9 Linear regression model performance

4.2. Time Series Models

Exploring different modeling mechanisms provides a broader perspective to interpret and grasp the future perception and its influence on the economy, agriculture, and climate in general. The forecast models that are examined to

predict the future rainfall around the Dukan district are Prophet, to capture the strong seasonality presented in hydrological properties during winter and summer. SARIMA provides a base statistical line, and XGBoost incorporates

the lag feature as the moisture rises and the precipitation occurs a while after. The starting point is common for all the models: the preprocessed time series data, specifically the target rainfall series, which is a univariate series of normalized monthly rainfall values indexed by date (datetime format). Throughout the use of metrics, the trained models and evaluate their performance with relevant regression metrics such as Mean Absolute Error (MAE) or R-squared.

4.2.1. Prophet Model Flow

The model is initiated and fitted to the specific required data frame. Prediction for future values is established within 24 months including the confidence intervals (yhat_lower and yhat_upper). The model is plotted against actual historical data as shown in the Figure 10.

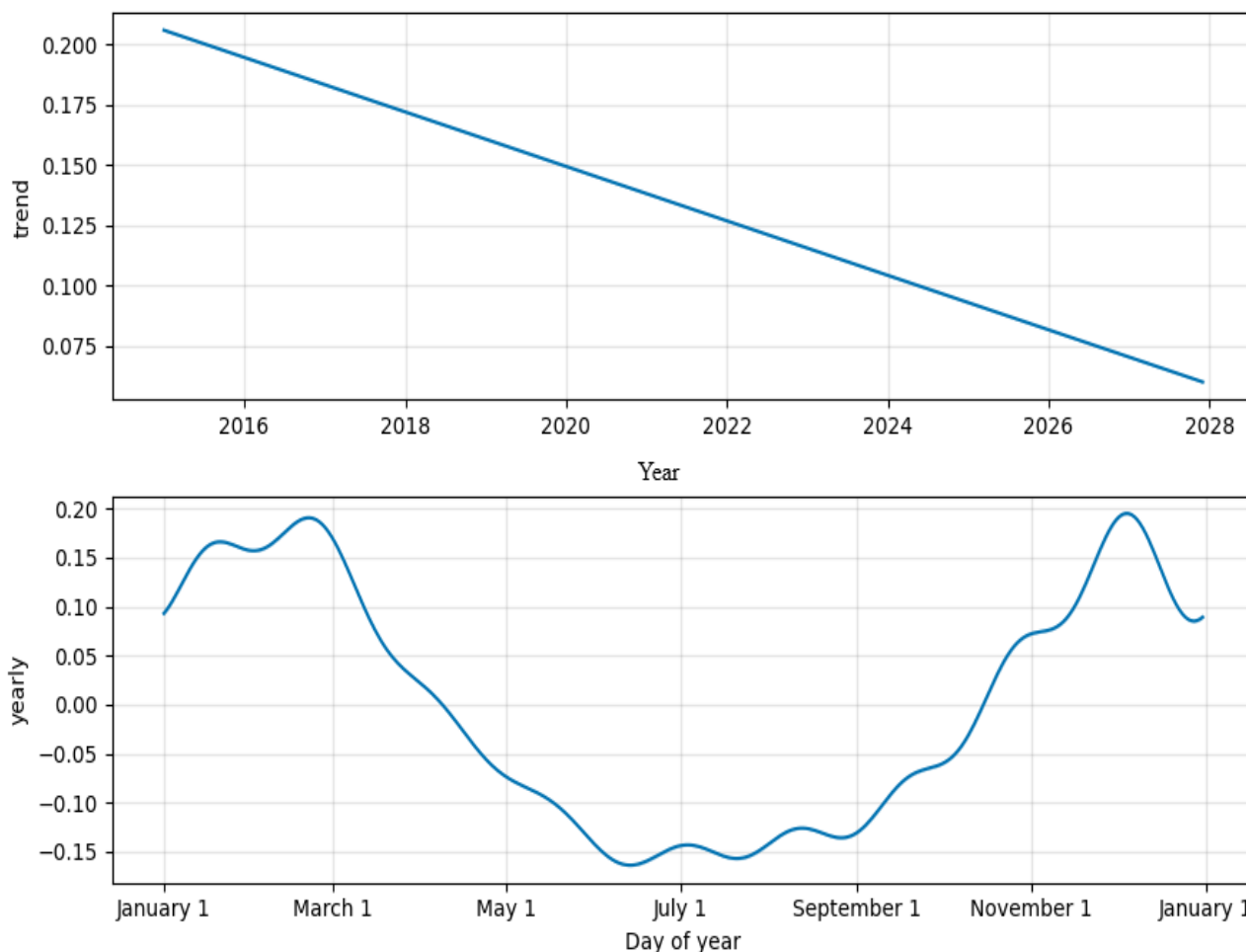


Figure 10 : Rainfall forecast generated by the Prophet model.

The model performance is evaluated by the metrics MAE (0.0653) and RMSE (0.0878). It achieved an intermediate result where the model successfully captured the annual precipitation cycle and recent declining trends post-2020. The reasons are attributed to the fact that it is a decomposable model and provides flexibility in handling the data. However, Prophet underestimated the drying trend in 2022 due to the fact that the changing trend may not fully capture the abrupt weather change, and the Fourier series oversimplify the irregular precipitation patterns common in semi-arid regions.

4.3. SARIMA Model Flow:

The autocorrelation function (ACF) and Partial correlation function (PACF) plots of the original data are analyzed

to determine the optimal (p, d, q) and (P, D, Q, S). In this notebook, the order= (1, 1, 1) and seasonal order= (1, 1, 1, 12), this pattern shows a single spike at lag 12, while the other lags were in the confidence interval. A seasonal difference (D) was applied to remove the yearly pattern and a first difference (d) to make the series stationary. The non-seasonal order were chosen by comparing candidate model. The SARIMA model is established with the identified parameters and then fitted to the target rainfall series. The fitted SARIMA prediction results object is utilized to generate forecasts for 24 months. The forecasted values are plotted alongside the actual target rainfall values as seen in Figure 9.

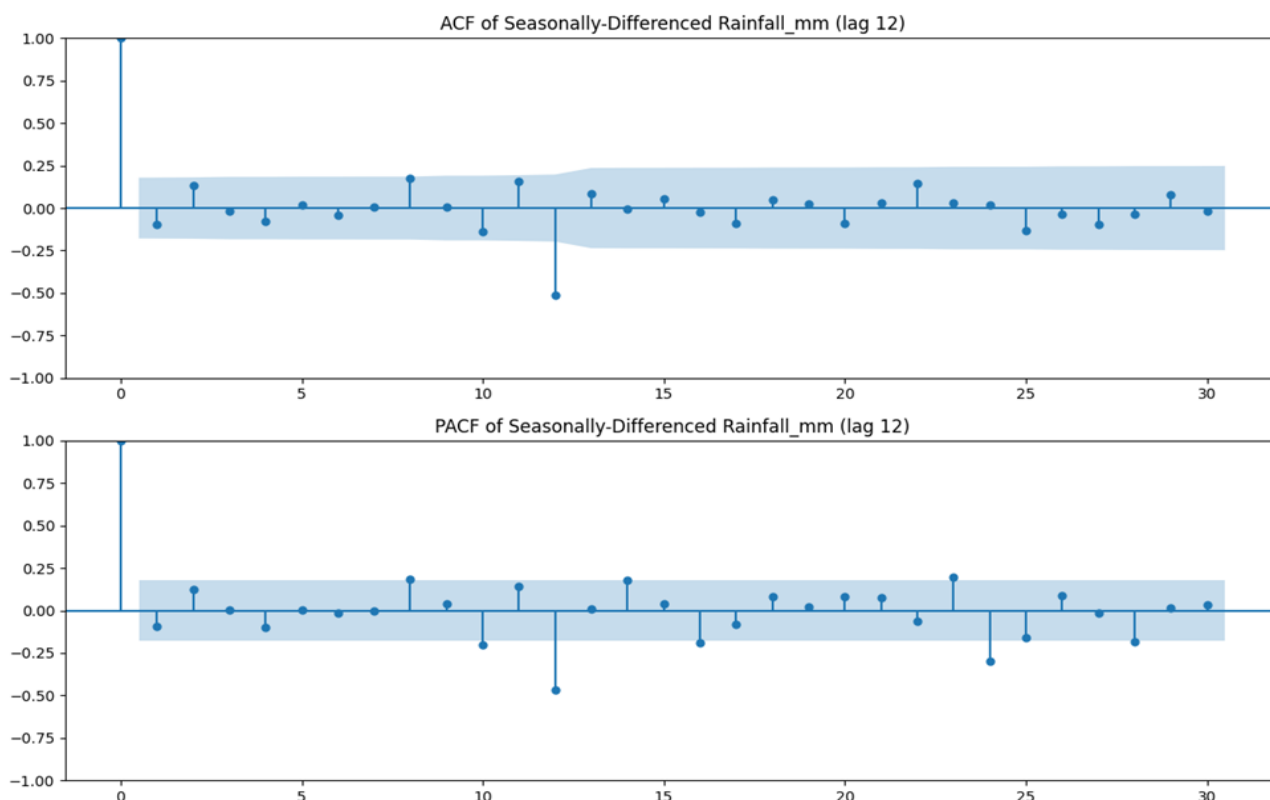


Figure 11. The autocorrelation function (ACF) and Partial correlation function (PACF) For 12 months period

The underlying temporal dependence structure of the seasonally differenced monthly rainfall series (lag 12) ACF) and (PACF) plots were analyzed as illustrated in Figure 11. The 95% confidence interval is illustrated by the shaded confidence band in both parts; spikes that go beyond these bounds signify statistically significant autocorrelation at the relevant lag. Most of the autocorrelation values in the ACF

plot are inside the confidence intervals, proving that seasonal differencing has effectively decreased the seasonal reliability. There is still some seasonal link to the annual cycle, though, as evidenced by the notable rise that is seen around lag 12. Significant spikes at early lags in the PACF plot indicate the existence of short-term autoregressive activity.

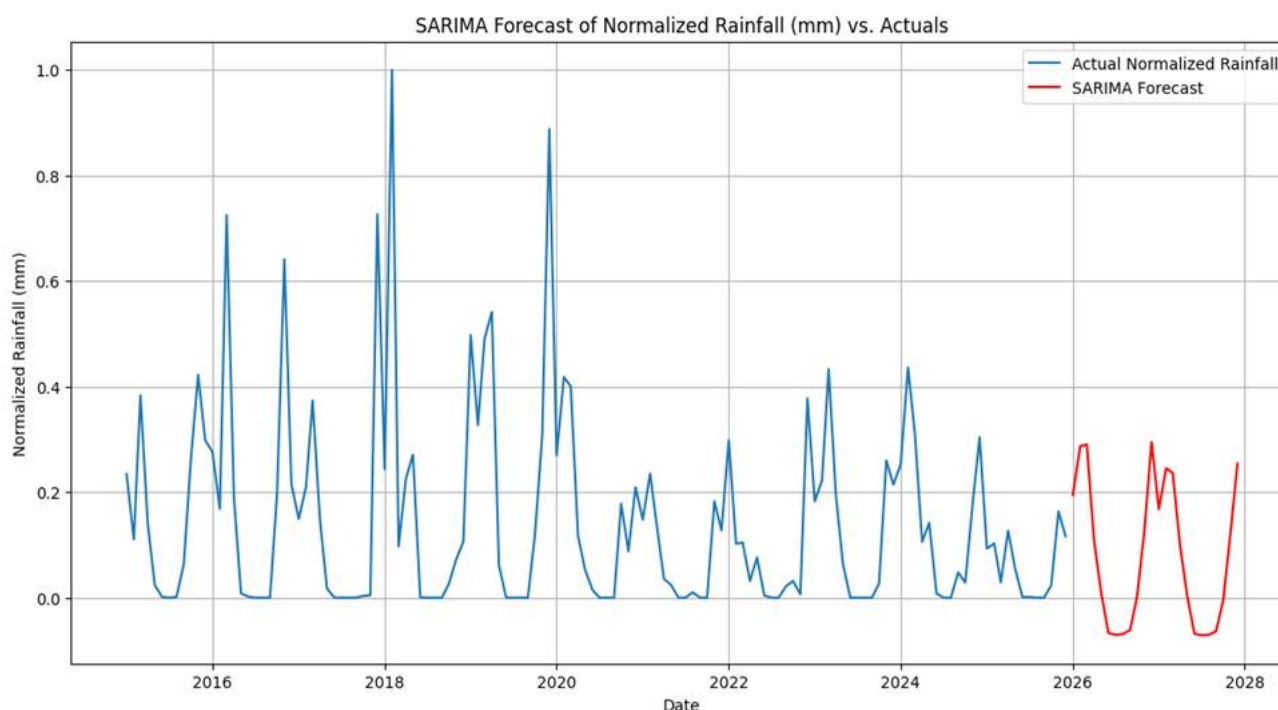


Figure 12 Sarima prediction results after fitting the model into the historical data

The model is evaluated through the use of metrics, where MAE is 0.0741, and RMSE is 0.0966. The SARIMA model is able to capture the seasonal cycle $S=12$. However, it is still not able to model the combined effect of the meteorological variables in the precipitation forecast because it relies on historical rainfall series, and the need for differencing ($d=1$, $D=1$) may remove some information content. Thus, reduce the model's ability to capture long-term climate shifts.

4.3.1. XGBoost Model Flow:

For time series forecasting with XGBoost, the 'input x' consists of the engineered features (e.g., lag values, rolling statistics, time-based features like month and year) that we

explicitly created. The model learns the complex, non-linear relationships between these features and the target variable (Rainfall_mm) to predict.

For future predictions, a `future_df_xgboost` is created with future dates, and the same feature engineering steps (lags, rolling stats, time-based) are applied. Due to the nature of lags and rolling windows, this often involves iteratively predicting and feeding predictions back, or careful handling of unknown future feature values (e.g., filling with means or last known values). The `predict()` method is then used on `X_future` to get forecasts. Actual `y_test` values are plotted against the `y_pred_xgboost` predictions (for the test set) and the `xgboost_future_forecast` (for the future period) to visualize performance and future trends, as shown in Figure 13.

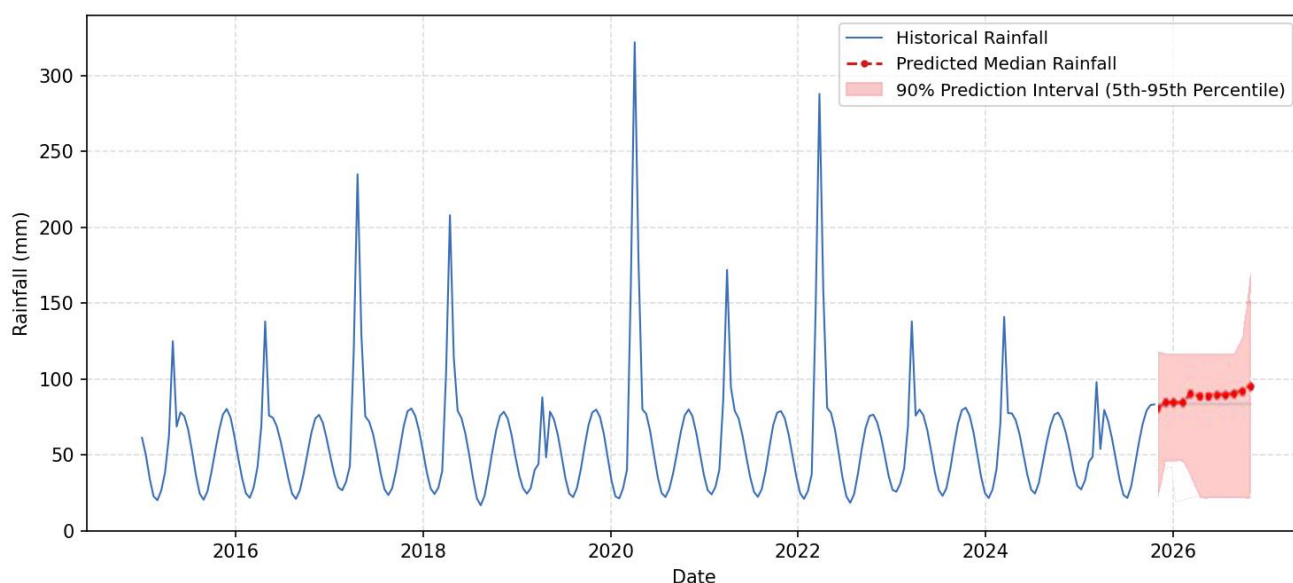


Figure 13: XG Boost prediction results

The historical rainfall records show strong seasonality and variation over time; it can be observed by the blue curve. The gray area depicts the forecast interval derived via the model indicating model uncertainty and possible rainfall variability scenarios, while the red dashed line indicates the predicted median rainfall for the future period to remain within a moderate range compared with historical peaks.

However, the prediction interval highlights the uncertainty in climatic processes and suggests that both wetter-than-average and drier-than-average conditions remain possible. The model performance is evaluated through the use of metrics where MAE is 0.0296 and RMSE is 0.0416, and the model's comparison summary is illustrated in Figure 15.

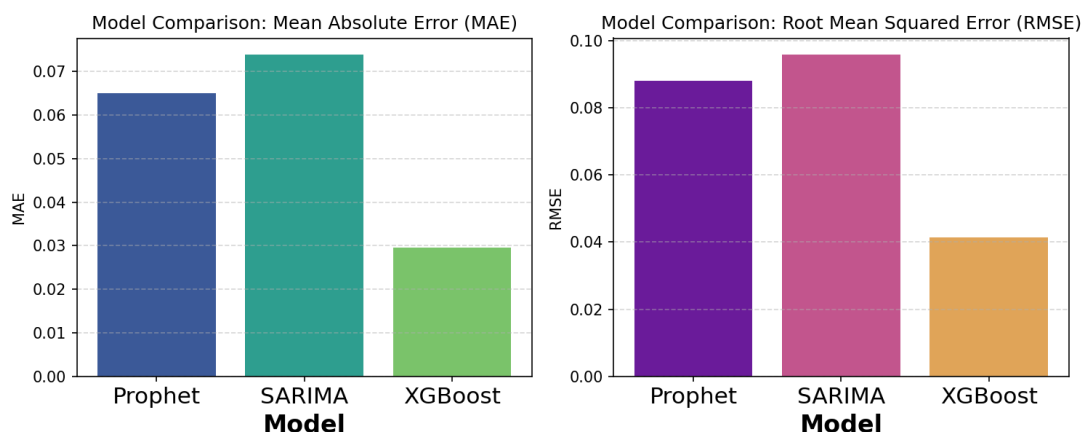


Figure 14 The comparison of the models' evaluation metrics (Normalized units)

The metrics used for evaluation were denormalized by multiplying the normalized MAE and RMSE values by the standard deviation of 27.17 mm of the original rainfall dataset because the models were trained using Z-score normalization. This allowed the errors to be reported in millimeters, where the values for MAE and RMSE are (0.80, 1.13), (2.01, 2.62), (1.76, 2.38) mm for XGBoost, SARIMA, Prophet, respectively. XGBoost outperformed the other models, which can be attributed to its advantageous algorithm for this task.

1. It shows the capability of modeling nonlinear behavior like the synergistic effect of temperature and humidity on precipitation possibility. For example, the model captures that high humidity combined with low temperature leads to rainfall, while high temperature with low humidity results in dry conditions.
2. The other models treat the variables as add-ons, while XGBoost leverages their combined effect; finally, the lag features capture the rainfall pattern from two to six months. These characteristics make XGBoost particularly well-suited for noisy, non-linear, and multivariate data.
3. XGBoost's sequential learning process, where each new tree is trained on the residual of the previous one and corrects the previous error. It maintained accuracy without showing signs of memorizing training patterns.

5. Conclusions

The study analyzes three time series models, SARIMA, Prophet, and XGBoost, to predict the future monthly rainfall in the Dukan district. All models' prediction performance of predictions shows improvement in comparison to univariate baselines through the inclusion of exogenous variables such as temperature, humidity, and wind speed. XGBoost exhibited its capacity to capture nonlinear interactions and seasonal dependencies by consistently outperforming the others and achieving the lowest error metrics.

The results reflect the promise of machine learning techniques in operational water resource management and emphasize the importance of multi-factor statistical modeling in hydrological forecasting. For future investigation, the incorporation of satellite images from different seasons for the region could further improve the accuracy of the forecast, along with the inclusion of external large-scale climate indices (e.g., ENSO, NAO), which may further enhance model accuracy.

Contributor ship Statement

Conflict of interests

The authors declare that there is no conflict of interest regarding the publication of this paper or receipt of specific funding.

6. References

- [1] Z. Liu, "Comparison of versions 6 and 7 3-hourly TRMM multi-satellite precipitation analysis (TMPA) research products," *Atmospheric Research*, vol. 163, pp. 91-101, 2015/09/15/ 2015, doi: <https://doi.org/10.1016/j.atmosres.2014.12.015>.
- [2] F. Vitart and A. W. Robertson, "Chapter 1 - Introduction: Why Sub-seasonal to Seasonal Prediction (S2S)?," in *Sub-Seasonal to Seasonal Prediction*, A. W. Robertson and F. Vitart Eds.: Elsevier, 2019, pp. 3-15.
- [3] H. Al-Bazaz, "Assessment of Climate Change Impacts in Iraq Using Innovative and Polygonal Trend Analyses of Monthly Rainfall Data," *Jordan Journal of Civil Engineering*, vol. 20, 2026, doi: <https://doi.org/10.14525/JJCE.v20i2.07>.
- [4] K. Trenberth, "Changes in precipitation with climate change," *Climate Research*, vol. 47, no. 1, pp. 123-138, 2011, doi: 10.3354/cr00953.
- [5] J. D. Neelin et al., "Precipitation Extremes and Water Vapor," *Current Climate Change Reports*, vol. 8, no. 1, pp. 17-33, 2022, doi: 10.1007/s40641-021-00177-z.
- [6] J. Fan, R. Zhang, G. Li, and W. K. Tao, "Effects of aerosols and relative humidity on cumulus clouds," *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D14, 2007, doi: 10.1029/2006jd008136.
- [7] B. A. Sara Bouziane, Aniss Moumen, Ali Sahlaoui and Anas El Ouali, "Enhancing Precipitation Prediction in the Ziz Basin: A Comprehensive Review of Traditional and Machine Learning Approaches," presented at the E3S Web Conf. 4th International GIRE3D Congress "Participatory and Integrated Management of Water Resources in Arid Zones" (GIRE3D 2023) 09 February 2024, 04010.
- [8] F. Yazdandoost, M. Zakipour, and A. Izadi, "Statistical refinement of the North American Multi-Model Ensemble precipitation forecasts over Karoon basin, Iran," *Journal of Water and Climate Change*, vol. 14, no. 8, pp. 2517-2530, 2023, doi: 10.2166/wcc.2023.277.
- [9] T. A. Al-Qadi, K. D. Muslih, and A. G. Shiltagh, "Analysis of Correlation and Coupling between El Niño-Southern Oscillation and Dust Storms in Iraq from 1971 to 2016," *The Iraqi Geological Journal*, vol. 54, no. 1E, pp. 103-113, 2021, doi: 10.46717/igj.54.1e.9ms-2021-05-30.
- [10] B. K. Raoof, A. Rabia, U. Alameedy, P. Shakor, and M. Karakouzian, "Synergizing Machine Learning and Physical Models for Enhanced Gas Production Forecasting: A Comparative Study of Short- and Long-Term Feasibility," *Energies*, vol. 18, no. 5, p. 1187, 2025, doi: <https://doi.org/10.3390/en18051187>.
- [11] A. K. Shukla, M. Janmajaya, A. Abraham, and P. K. Muhuri, "Engineering applications of artificial intelligence: A bibliometrics analysis of 30 years (1988–2018)," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 517-532, 2019/10/01/ 2019, doi: <https://doi.org/10.1016/j.engappai.2019.06.010>.

- [12] H. Zhang, V. P. Singh, B. Wang, and Y. Yu, "CEREF: A hybrid data-driven model for forecasting annual streamflow from a socio-hydrological system," *Journal of Hydrology*, vol. 540, pp. 246-256, 2016/09/01/ 2016, doi: <https://doi.org/10.1016/j.jhydrol.2016.06.029>.
- [13] S. Guo, J. Guo, X. Li, and N. Zhang, "Rainfall Prediction in Henan Province based on the Prophet-SARIMA Model," 2024: ACM, pp. 286-290, doi: 10.1145/3735014.3735914. [Online]. Available: <https://dx.doi.org/10.1145/3735014.3735914>
- [14] S. J. Taylor and B. Letham, "Forecasting at scale," *PeerJ*, 2017.
- [15] [15] A. A. Dash and E. McBean, "Prophet-Based Artificial Intelligence Versus Seasonal Auto-Regressive Models for Flood Forecasting with Exogenous Variables," *Water*, vol. 17, no. 24, p. 3551, 2025, doi: 10.3390/w17243551.
- [16] D. Li et al., "Prediction of Rainfall Time Series Using the Hybrid DWT-SVR-Prophet Model," *Water*, vol. 15, no. 10, p. 1935, 2023, doi: <https://dx.doi.org/10.3390/w15101935>.
- [17] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting," *Machine Learning with Applications*, vol. 7, p. 100204, 2022/03/15/ 2022, doi: <https://doi.org/10.1016/j.mlwa.2021.100204>.
- [18] H. Aksoy and A. Dahamsheh, "Markov chain-incorporated and synthetic data-supported conditional artificial neural network models for forecasting monthly precipitation in arid regions," *Journal of Hydrology*, vol. 562, pp. 758-779, 2018/07/01/ 2018, doi: <https://doi.org/10.1016/j.jhydrol.2018.05.030>.
- [19] M. Shabbir, S. Chand, and F. Iqbal, "A novel hybrid framework to model the relationship of daily river discharge with meteorological variables," *Meteorology Hydrology and Water Management. Research and Operational Applications*, vol. 11, 2023.
- [20] P. K. Das, R. L. Sahu, and P. C. Swain, "Comparative analysis of machine learning models for rainfall prediction," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 264, p. 106340, 2024/11/01/ 2024, doi: <https://doi.org/10.1016/j.jastp.2024.106340>.
- [21] Y. A. Hamaamin, "Evaluation of Data Mining and Artificial Intelligence Methods to Predict Daily Precipitation," *Journal of Engineering*, vol. 31, no. 5, pp. 99-112, 2025.
- [22] R. H. Al-Suhaili and R. A. Karim, "Spatial Prediction of Monthly Precipitation in Sulaimani Governorate using Artificial Neural Network Models," *Journal of Engineering*, vol. 20, no. 3, pp. 15-27, 2023, doi: 10.31026/j.eng.2014.03.02.
- [23] Q. Ma, H. Wang, P. Luo, Y. Peng, and Q. Li, "Ultra-short-term Railway traction load prediction based on DWT-TCN-PSO_SVR combined model," *International Journal of Electrical Power & Energy Systems*, vol. 135, p. 107595, 2022/02/01/ 2022, doi: <https://doi.org/10.1016/j.ijepes.2021.107595>.
- [24] Y. Xia, S. Jiang, L. Meng, and X. Ju, "XGBoost-B-GHM: An Ensemble Model with Feature Selection and GHM Loss Function Optimization for Credit Scoring," *Systems*, vol. 12, no. 7, p. 254, 2024, doi: 10.3390/systems12070254.
- [25] S. S. D. Ministry of Planning, Departement of Mapping and GIS, 2025. (accessed).
- [26] S. J. Taylor and B. Letham, "Forecasting at Scale," *The American Statistician*, vol. 72, no. 1, pp. 37-45, 2018/01/02 2018, doi: 10.1080/00031305.2017.1380080.
- [27] M. Cintani, Y. A. Nizar, Y. Angraini, K. A. Noto-diputro, and L. N. A. Mualifah, "Evaluation of the SARIMA and Prophet Models in Forecasting Ship Passenger Numbers at Balikpapan Port," *Jambura Journal of Mathematics*, vol. 7, no. 2, pp. 120-128, 2025, doi: 10.37905/jjom.v7i2.31606.