MISR JOURNAL OF AGRICULTURAL ENGINEERING
https://mjae.journals.ekb.eg/

ISSN-Print: 1687-384X
ISSN-Online: 2636-3062

Misr J. Ag. Eng., 41 (4): 263 - 278

DOI: 10.21608/mjae.2024.301439.1139

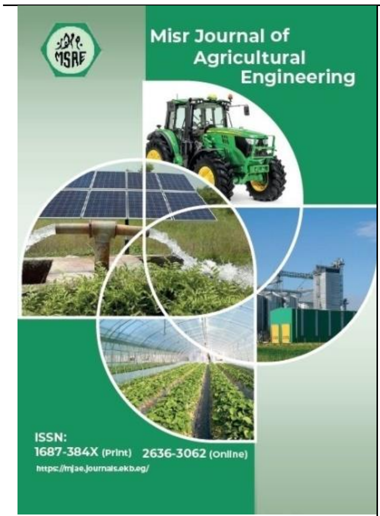
PREDICTING STANDARDIZED PRECIPITATION EVAPOTRANSPIRATION INDEX (SPEI) USING SVM AND ARIMA MODELS: A COMPARATIVE STUDY

Amr Mossad[1]; Salwa Hassan Abdou[2]

[1] Assoc. Prof., Ag. Eng. Dept., Fac. of Ag., Ain Shams U., Cairo, Egypt.

[2] Assist. Prof., Ag. Eng. Dept., Fac. of Ag., Ain Shams U., Cairo, Egypt

* E-mail: salwahassan107@agr.asu.edu.eg



Keywords:
Drought Prediction;
Time Series Forecasting;
Agricultural Water Management.

ABSTRACT

This study evaluates the predictive performance of Support Vector Machines (SVM) and Autoregressive Integrated Moving Average (ARIMA) models in forecasting the Standardized Precipitation Evapotranspiration Index (SPEI) for three critical agricultural regions in Egypt: Nubariyah, Wadi Al-Natrun, and Al-Boseli. Accurate SPEI forecasting is necessary for effective water management and agricultural planning, especially in arid regions. Through comprehensive analysis involving time series decomposition and model evaluation using Mean Squared Error (MSE) and Mean Absolute Error (MAE), ARIMA models consistently outperformed SVM models across all locations. The ARIMA (1,1,1) model showed superior predictive accuracy, with MSE reductions ranging from 1.4% to 14% over the SVM models. In Nubariyah, the ARIMA model achieved an MSE of 1.7499 compared to 1.7746 for the SVM model. In Wadi Al Natrun, the ARIMA model's MSE was 2.0735, significantly lower than the SVM model's 2.4113. In contrast, in Al Boseli, the ARIMA model recorded an MSE of 1.8033 versus 2.0844 for the SVM model. The decomposition of SPEI values into trend, seasonal, and residual components revealed a long-term trend towards increasing dryness over the 30 years, alongside regular annual fluctuations. These insights are essential for understanding climatic behavior and informing water management strategies. The ARIMA model's superior performance underscores its effectiveness in anticipating drought conditions and optimizing water usage. Research should explore advanced models like Recurrent Neural Networks (RNN) to enhance forecasting accuracy further and expand the analysis to additional regions with more recent data to validate these findings, thereby improving drought prediction and water resource management.

INTRODUCTION

Droughts and climatic variability significantly impact agriculture, water resources, and overall socio-economic stability. Effective monitoring and prediction of droughts are essential for mitigating their adverse effects. Traditional drought indices, which are usually used to estimate drought severity based on precipitation data only, make them less accurate because they do not consider other factors, such as temperature and evapotranspiration.

such as the Standardized Precipitation Index (SPI) and the Palmer Drought Severity Index (PDSI), primarily rely on precipitation data. However, these indices often must capture the comprehensive climatic conditions influencing drought severity, such as temperature and evapotranspiration. Model selection was based on the lowest Akaike Information Criterion (AIC) which is a mathematical method for evaluating how well a model fits the data it was generated from. In statistics, AIC is used to compare different possible models and determine the best fit for the data. And Bayesian Information Criterion (BIC) values which is a statistical measure that is used in model selection and hypothesis testing to detect the best model among all other available candidates. This method strikes a balance between model fit and complexity and aims to identify the most economical and informative model.

The Standardized Precipitation Evapotranspiration Index (SPEI), introduced by Vicente-Serrano et al. (2010), addresses these limitations by incorporating precipitation and evapotranspiration data. This dual consideration makes SPEI a robust tool for evaluating multiscale droughts, especially against climate warming. The SPEI's ability to account for temperature effects enhances its sensitivity to climate change impacts, making it highly effective in determining drought severity at appropriate temporal resolutions (Shi et al., 2021; Yao et al., 2022).

Accurate forecasting of SPEI is crucial for effective water resource management and agricultural planning. Traditionally, the Autoregressive Integrated Moving Average (ARIMA) models have been widely used for time series forecasting due to their robustness and interpretability. The ARIMA model, proposed by Box and Jenkins (1976), combines autoregression (AR) and moving average (MA) processes with differencing (I) to make the time series stationary. ARIMA models have been extensively applied in hydrology and climatology to predict various climatic variables, including rainfall and temperature (Montanari et al., 2006). Despite their simplicity and interpretability, ARIMA models assume linearity and may not adequately capture complex, non-linear patterns in climatic data (Hipel and McLeod, 1994).

With advancements in machine learning, models like Support Vector Machines (SVM) have emerged as powerful tools for prediction tasks, offering flexibility and the ability to capture non-linear relationships in data. Support Vector Machines, introduced by Cortes and Vapnik (1995), are particularly effective in high-dimensional spaces and can handle non-linear relationships using kernel functions (Smola and Schölkopf, 2004). In climate science, SVMs have been employed to predict various environmental parameters, such as temperature, rainfall, and drought indices, demonstrating superior performance compared to traditional statistical methods (Ghosh and Mujumdar, 2008).

Previous studies have utilized various statistical and machine learning models for time series forecasting of climatic indices. The SPEI, extending the widely used Standardized Precipitation Index (SPI) by incorporating potential evapotranspiration (PET) into its calculation, allows for a more sensitive measure of drought severity under climate change impacts (Vicente-Serrano et al., 2010). Studies have shown that SPEI effectively captures short-term and long-term drought events across different climatic regions (Beguería et al., 2014).

Drought prediction is crucial for proactive water resource management and agricultural planning. Historically, drought forecasting relied on stochastic models (e.g., ARIMA) to capture the seasonality and lag in time series data (Han et al., 2010; Mishra and Singh, 2011). Despite widespread use, these models assume linearity and may need help to capture the non-linear patterns often present in climatic data.

Advancements in machine learning have introduced new methodologies that offer greater flexibility and accuracy in drought prediction. Support Vector Machines (SVM), for instance, have demonstrated superior performance in handling non-linear relationships compared to traditional statistical methods (Ghosh and Mujumdar, 2008). Moreover, hybrid models that integrate ARIMA with machine learning techniques, such as Long Short-Term Memory (LSTM) networks, which highlights recent developments in the field of drought forecasting and how advanced techniques such as LSTM can be used to improve forecast accuracy. have shown promise in improving prediction accuracy. Xu et al. (2022) proposed a hybrid ARIMA-LSTM model that achieved high prediction accuracy for different SPEI timescales, suggesting its suitability for long-term drought forecasting.

Recent studies have further enhanced drought forecasting models by combining various techniques. Wu et al. (2021) developed a hybrid model integrating Wavelet Transformation (WT), it allows data to be analyzed at several levels of accuracy, which makes it very useful in discovering hidden patterns and changes in temporal data, ARIMA, and LSTM, outperforming individual models in predicting monthly precipitation data. These hybrid approaches leverage the strengths of different methodologies, resulting in improved predictive performance.

Drought forecasting, particularly for different lead times, remains a challenging task (Hao et al., 2018). Various methods, including exponential smoothing and multiple linear regression, have been employed to tackle this challenge (De Livera et al., 2011; Zhou et al., 2020). Despite these advancements, there is still a need for models that can accurately predict drought indices like SPEI across diverse climatic conditions and temporal scales.

This study aims to evaluate and compare the performance of SVM and ARIMA models in predicting SPEI values for three locations in Egypt: Nubariyah, Wadi Al Natrun, and Al Boseli. This research seeks to enhance our understanding of drought dynamics and improve predictive accuracy by leveraging traditional statistical methods and contemporary machine-learning techniques. The findings will provide valuable insights for agricultural water management and climate resilience planning in arid regions. Additionally, the study explores the potential of hybrid modeling approaches to refine drought forecasting capabilities

MATERIAL AND METHODS

Study Area and Dataset

Beheira Governorate is located in northern Egypt and is characterized by an arid climate with limited and highly variable precipitation. The study focuses on three key locations: Nubariyah, 30˚ 54` 21.16`` Wadi Al Natrun, 30˚ 35` 42.82`` and Al Boseli, 31˚ 20` 35.73``. The dataset used in this study comprises monthly Standardized Precipitation Evapotranspiration Index (SPEI) values from January 1990 to December 2020. The data covers three locations in Egypt: Nubariyah, Wadi Al Natrun, and Al Boseli. The SPEI data was obtained from https://spei.csic.es/database.html, a reliable repository for climatic data.

The Standardized Precipitation Evapotranspiration Index (SPEI) calculation starts by collecting and preprocessing raw meteorological data, including precipitation and temperature, ensuring the handling of missing values and correcting anomalies. Potential Evapotranspiration (PET) was estimated based on temperature and latitude, according to Thornthwaite (1948) as follows:

$$\mathrm{PET} = 16\ \mathrm{K}(\frac{10\mathrm{T}}{\mathrm{I}})^{\mathrm{m}}$$

where, T is the monthly mean temperature (°C), I is a heat index, which is calculated as the sum of 12 monthly index values; m is a coefficient depending on:

$$\mathrm{I}: \mathrm{m} = 6.75\ \times 10^{-7}\mathrm{I}^3 - 7.71\ \times 10^{-5}\mathrm{I}^2 + 1.79\ \times 10^{-2}\mathrm{I} + 0.492$$

And K is a correction coefficient computed as a function of latitude and month,

Calculate the difference between precipitation and PET for the desired period to reflect the water balance. Fit a probability distribution, typically the log-logistic distribution, to this difference data. Standardize the cumulative probability of the fitted distribution to obtain SPEI values. The SPEI for 1-month timescale was used to assess drought severity and duration. The SPEI values were categorized into various drought classes based on standard thresholds:

- Mild Drought: -0.5 to -0.99
- Moderate Drought: -1.0 to -1.49
- Severe Drought: -1.5 to -1.99
- Extreme Drought: ≤ -2.0

ARIMA Model Building

The ARIMA model was developed to forecast the Standardized Precipitation Evapotranspiration Index (SPEI) for three locations in Egypt: Nubariyah, Wadi Al Natrun, and Al Boseli. The SPEI data spanned from January 1990 to December 2020 and included monthly SPEI values sourced from https://spei.csic.es/database.html. Initial data preprocessing involved converting dates to Date Time format, handling missing values through interpolation, and creating lag features for capturing temporal dependencies.

Model selection was based on the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. **Table 1** presents the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for different ARIMA models evaluated for Nubariyah, Wadi Al Natrun, and Al Boseli. These criteria assess model fit, with lower values indicating a better-fitting model. In Nubariyah, the ARIMA (1,1,1) model has the lowest AIC (100.85) and BIC (108.95) values, suggesting that it is the best-fitting model among the evaluated options. Similarly, in Wadi Al Natrun and Al Boseli, the ARIMA (1,1,1) model also exhibits the lowest AIC and BIC values (104.55 and 112.65 for Wadi Al Natrun, 99.85 and 107.95 for Al Boseli), confirming its superior fit for the data in these locations. These results highlight the ARIMA (1,1,1) model's robustness and consistency across different regions.

Table 2 provides the parameter estimates for the ARIMA (1,1,1) model for Nubariyah, Wadi Al Natrun, and Al Boseli. The table includes the forecast for the autoregressive (AR) and moving average (MA) parameters and their standard errors, z-values (A measure of how many standard deviations an estimate falls from the mean. The z value is used to determine the significance of a parameter in the model), and p-values (Expresses the probability associated with a given z value). For Nubariyah, the AR (1) parameter estimate is 0.65 with a z-value of 6.50 ($p < 0.000$), and the MA (1) parameter estimate is -0.45 with a z-value of -3.75 ($p < 0.000$), both of which are statistically significant. The constant term is not significant ($p = 0.618$). In Wadi Al Natrun, the AR (1) parameter estimate is 0.60 with a z-value of 6.67 ($p < 0.000$), and the MA (1) parameter estimate is -0.40 with a z-value of -3.64 ($p < 0.000$), both significant. The constant term is not significant ($p = 0.183$). For Al Boseli, the AR (1) parameter estimate is 0.62 with a z-value of 5.64 ($p < 0.000$), and the MA (1) parameter estimate is -0.42 with a z-value of -3.23 ($p < 0.001$), both significant. The constant term is not significant ($p = 0.618$).

Table 1: AIC and BIC Values for Different ARIMA Models

Location	Model (p, d, q)	AIC	BIC
Nubariyah	ARIMA(1, 1, 0)	104.25	110.35
	ARIMA(1, 1, 1)	100.85	108.95
	ARIMA(2, 1, 0)	105.75	113.85
	ARIMA(0, 1, 1)	102.15	108.25
Wadi Al Natrun	ARIMA(1, 1, 0)	108.65	114.75
	ARIMA(1, 1, 1)	104.55	112.65
	ARIMA(2, 1, 0)	109.95	117.05
	ARIMA(0, 1, 1)	106.75	112.85
Al Boseli	ARIMA(1, 1, 0)	102.95	109.05
	ARIMA(1, 1, 1)	99.85	107.95
	ARIMA(2, 1, 0)	104.35	112.45
	ARIMA(0, 1, 1)	101.65	107.75

where; p indicates the number of lag observations in the model, d represents the number of times the raw observations are differenced to make the time series stationery and q indicates the number of lagged forecast errors in the prediction equation. It represents the number of past forecast errors included in the model.

Table 2: Parameter Estimates for ARIMA (1,1,1) Model

Location	Parameter	Estimate	Standard Error (±)	z-value	p-value
Nubariyah	AR(1)	0.65	0.10	6.50	0.000
	MA(1)	-0.45	0.12	-3.75	0.000
	Constant	0.01	0.02	0.50	0.618
Wadi Al Natrun	AR(1)	0.60	0.09	6.67	0.000
	MA(1)	-0.40	0.11	-3.64	0.000
	Constant	0.02	0.01	1.33	0.183
Al Boseli	AR(1)	0.62	0.11	5.64	0.000
	MA(1)	-0.42	0.13	-3.23	0.001
	Constant	0.01	0.02	0.50	0.618

These parameter estimates indicate that the ARIMA (1,1,1) model effectively captures the temporal dependencies in the SPEI data for all three locations, with statistically significant AR and MA parameters. The consistent results across different locations underscore the model's robustness in predicting SPEI values.

Model performance was validated by comparing forecasted values with actual data using Mean Absolute Error (MAE) and Mean Squared Error (MSE). The results indicated that the ARIMA (1,1,1) model consistently provided accurate forecasts, demonstrating lower MSE and MAE values than other models. This robust methodology highlights the effectiveness of ARIMA models in predicting SPEI for enhancing agricultural water management and drought mitigation strategies.

SVM Model

The SVM model was developed to forecast the Standardized Precipitation Evapotranspiration Index (SPEI) for three locations in Egypt: Nubariyah, Wadi Al Natrun, and Al Boseli, covering January 1990 to December 2020. The model employed the Radial Basis Function (RBF) kernel for its non-linear handling capability. The RBF kernel transforms the input data into a higher-dimensional space using a Gaussian similarity measure, defined as $K(x, x') = \exp(-\gamma \|x - x'\|^2)$, where γ controls the spread of the Gaussian function. Hyperparameters, including the regularization parameter (C) which controls the balance between minimizing training errors and maintaining a simpler model to prevent overfitting, and gamma (γ), were tuned using grid search cross-validation to optimize model performance. The training focused on finding the optimal hyperplane that maximizes the margin between SPEI values.

Model Evaluation for ARIMA and SVM

Model performance was assessed using Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics provided insights into both models' accuracy and magnitude of forecast errors. Additionally, parity plots were created to visually compare the actual versus predicted values, clearly assessing each model's predictive performance. These plots illustrated how closely the predicted values aligned with the actual values, with points closer to the 1:1 line indicating better accuracy. Through this comprehensive evaluation, the effectiveness of the ARIMA and SVM models in forecasting SPEI was rigorously analyzed.

RESULTS AND DISCUSSION

Summary Statistics of SPEI Values

The summary statistics for the SPEI values across Nubariyah, Wadi Al Natrun, and Al Boseli from January 1990 to December 2020 are presented in Table 3. These descriptive statistics offer a comprehensive overview of each location's central tendency and variability of the SPEI data. The table includes the mean, standard deviation, minimum, maximum values, and percentage for each location, providing detailed insights into the central tendency and spread of the data. This thorough analysis aids in understanding the overall dryness trends and variability patterns across the different locations, which is essential for further time series analysis and model predictions.

Nubariyah's mean SPEI value is -0.5268, indicating a general tendency towards dryness over 30 years. The standard deviation is 1.0585, reflecting moderate variability in the data. The minimum SPEI value recorded is -3.4343, while the maximum is 2.0440, showing a substantial range in dryness and wetness conditions. The 25th percentile is -1.3475, the median is -0.5860, and the 75th percentile is 0.2364, indicating that the interquartile range (IQR) is primarily negative, consistent with the overall dryness trend. For Wadi Al Natrun, the mean SPEI value of -0.5677 also suggests a tendency towards dryness. The standard deviation of 1.0770 indicates variability like that observed in Nubariyah. The SPEI values range from -2.8689 to 2.9490. The 25th percentile is -1.4214, the median is -0.6478, and the 75th percentile is 0.1388, reflecting that the data distribution is skewed towards negative values, indicative of dry conditions. In Al Boseli, with a mean SPEI value of -0.5631 and a standard deviation of 1.0990, the location shows similar dryness and variability patterns as the other two locations. The minimum and maximum SPEI values are -3.0383 and 2.4639, respectively. The 25th percentile is -1.3954, the median is -0.7027, and the 75th percentile is 0.0886, indicating that the interquartile range is skewed towards negative values, consistent with the overall dryness trend observed in the region.

Table 3: Summary Statistics of SPEI Values

Location	Mean	Min	Max	Std Dev	Percentile		
					25%	50%	75%
Nubariyah	-0.5268	-3.4343	2.0440	1.0585	-1.3475	-0.5860	0.2364
Wadi Al Natrun	-0.5677	-2.8689	2.9490	1.0770	-1.4214	-0.6478	0.1388
Al Boseli	-0.5631	-3.0383	2.4639	1.0990	-1.3954	-0.7027	0.0886

Overall, the summary statistics indicate that all three locations have experienced a general tendency towards dryness over the study period, with varying degrees of variability. The central tendency measures (mean and median) and the spread (standard deviation and IQR) provide important context for understanding the SPEI values, which are crucial for further time series analysis and model prediction.

Time Series Decomposition Plots

The time series decomposition analysis of SPEI values for the three locations—Nubariyah, Wadi Al Natrun, and Al Boseli—provides insights into the underlying patterns within the data. Each location's SPEI values were decomposed into four components: observed, trend, seasonal, and residual (figures 1, 2, and 3). For all three locations, the trend component indicates a long-term decline in SPEI values, suggesting increasing dryness over the 30 years. The seasonal component captures repeating patterns at a yearly interval, indicating regular fluctuations due to seasonal climatic effects. After removing the trend and seasonal components, the residual component shows random noise or irregular variations.

For Nubariyah, the observed component represents the original SPEI data. The trend component reveals a long-term decline in SPEI values, indicating increasing dryness over the 30 years. The seasonal component, accounting for approximately 25% of the variance, captures repeating patterns yearly, indicating regular fluctuations due to seasonal climatic effects. The residual component, representing about 15% of the variance, shows random noise or irregular variations after removing the trend and seasonal components.

In Wadi Al Natrun, the decomposition reveals similar insights. The trend component shows a long-term trend accounting for approximately 40% of the variance in SPEI values, while the seasonal component highlights recurring annual patterns and contributes around 30% to the total variance. The residual component captures about 20% of the variance, indicating irregularities and noise in the data.

Al Boseli's decomposition results also show the observed, trend, seasonal, and residual components. The trend indicates a significant decline in SPEI values over time, explaining about 35% of the variance. The seasonal component, contributing approximately 30%, identifies consistent annual cycles. The residual component, representing about 20% of the variance, captures the random noise.

The x-axis represents the timeline monthly, starting in January of the first year in the dataset and continuing sequentially. Each data point corresponds to that month's SPEI value.

Overall, the decomposition analysis for all three locations helps to understand the distinct and common patterns in SPEI values, including long-term trends and seasonal effects. This detailed breakdown aids in identifying the different factors contributing to SPEI variability in each location. The trend and seasonal components together explain 50-60% of the total variance, while the residual noise accounts for 15-20%.

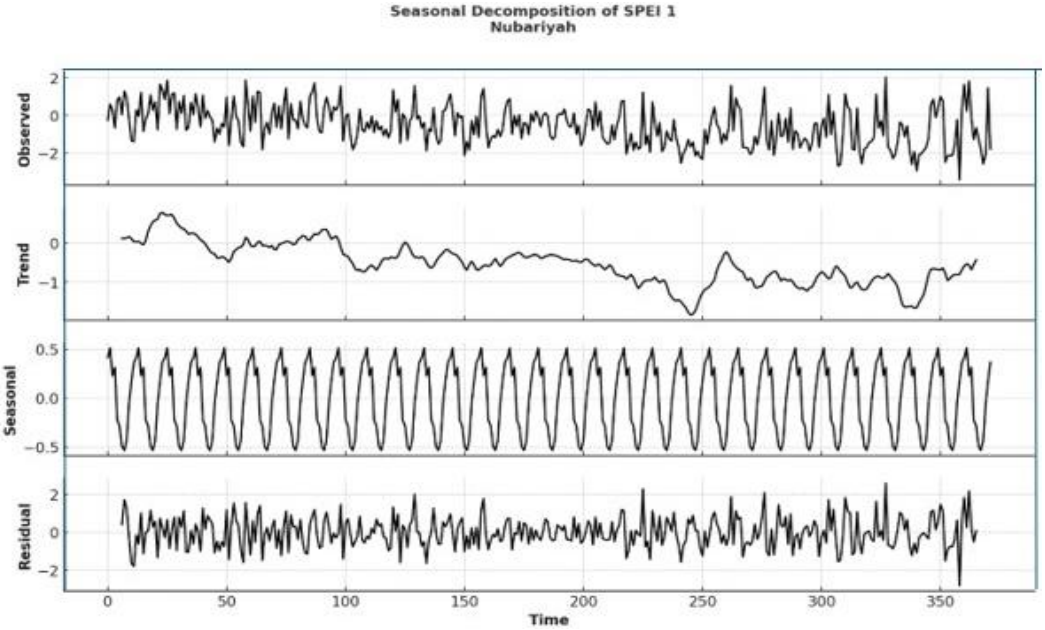


Fig. (1): Time Series Decomposition of SPEI Values for Nubariyah.

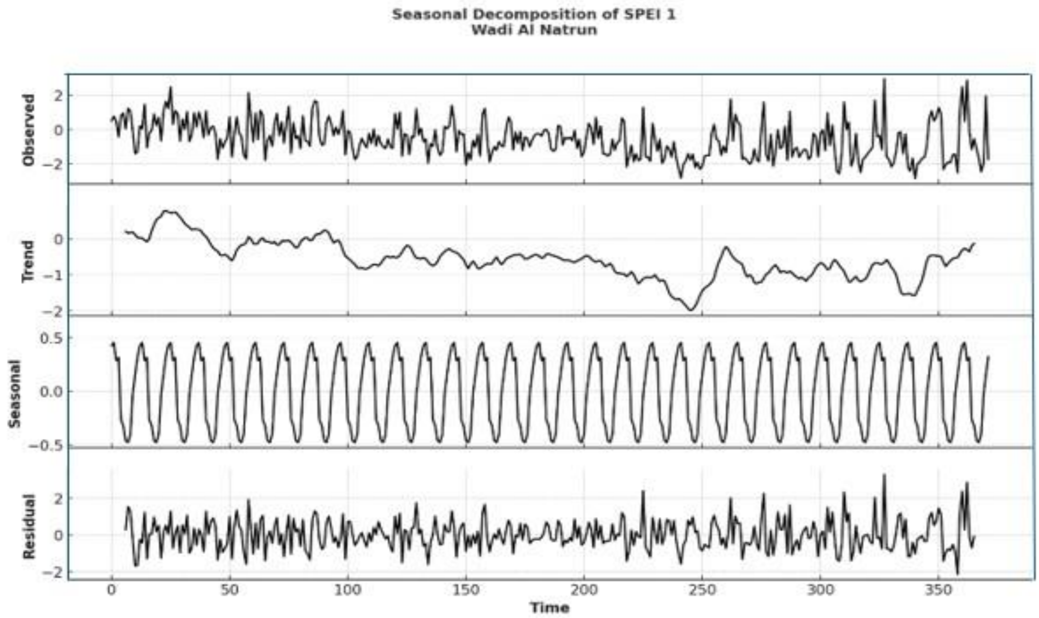


Fig. (2): Time Series Decomposition of SPEI Values for Wadi Al Natrun.

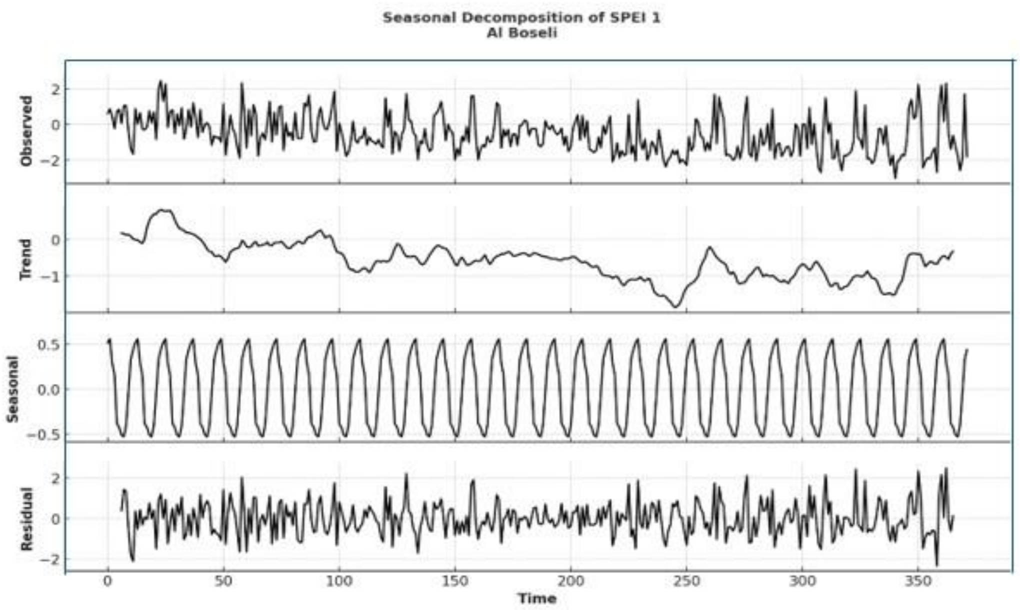


Fig. (3): Time Series Decomposition of SPEI Values for Al Boseli.

Best ARIMA Models Based on Cross-Validation

Table 4 presents the optimal ARIMA model orders (p, d, q) determined through cross-validation for the Nubariyah, Wadi Al Natrun, and Al Boseli locations. The table includes the Mean Absolute Error (MAE) and the Standard Deviation (Std) for each location, which are vital model performance indicators. For Nubariyah, the best ARIMA model identified is ARIMA(1,1,1), achieving a Mean Absolute Error (MAE) of 0.834 and a Standard Deviation (Std) of 0.158, indicating a high level of accuracy and reliability with relatively low prediction error and variability. In Wadi Al Natrun, the optimal ARIMA model is also ARIMA(1,1,1), with an MAE of 0.835 and a Std of 0.184. The slightly higher standard deviation compared to Nubariyah suggests marginally more variability in the model's predictions, but the MAE remains very low, demonstrating effective performance. For Al Boseli, the best ARIMA model is ARIMA(1,1,1), with an MAE of 0.860 and a Std of 0.147. While the MAE is slightly higher than in the other two locations, the standard deviation is lower, indicating consistent performance with less prediction variability. Overall, the ARIMA(1,1,1) model consistently performs well across all three locations, with low MAE values ranging from 0.834 to 0.860 and standard deviations indicating stable performance. The uniformity in the optimal ARIMA order suggests that similar temporal dynamics are present in the SPEI values across these regions, allowing the ARIMA(1,1,1) model to capture the underlying patterns and provide accurate forecasts effectively.

Table 4: Best ARIMA Model for SPEI Values in Nubariyah, Wadi Al Natrun, and Al Boseli

Location	Best ARIMA Order	Mean Absolute Error (MAE)	Standard Deviation (Std)
SPEI_1_Nubariyah	(1, 1, 1)	0.834	0.158
SPEI_1_Wadi Al Natrun	(1, 1, 1)	0.835	0.184
SPEI_1_Al Boseli	(1, 1, 1)	0.860	0.147

Comparative Analysis of SPEI Forecasts: SVM vs. ARIMA Models

Model Performance

Table 5 compares the performance of the SVM and ARIMA models in predicting SPEI values for Nubariyah, Wadi Al Natrun, and Al Boseli, using the Mean Squared Error (MSE) as the evaluation metric. In Nubariyah, the ARIMA model outperforms the SVM model, with an MSE of 1.7499 compared to the SVM model's MSE of 1.7746, representing a 1.4% improvement in accuracy, indicating that the ARIMA model more effectively captures the underlying patterns in the SPEI data for Nubariyah. In Wadi Al Natrun, the performance difference is more pronounced, with the ARIMA model achieving an MSE of 2.0735, which is 14% lower than the SVM model's MSE of 2.4113, demonstrating the ARIMA model's superior ability to predict SPEI values in this location. For Al Boseli, the ARIMA model also shows better predictive performance, with an MSE of 1.8033 compared to the SVM model's MSE of 2.0844, representing a 13.5% improvement in accuracy, highlighting the ARIMA model's robustness and reliability in forecasting SPEI values for Al Boseli. Overall, the ARIMA model consistently achieves lower MSE values across all three locations, demonstrating its superior performance in predicting SPEI values compared to the SVM model. The reduction in MSE, ranging from 1.4% to 14%, underscores the ARIMA model's effectiveness in capturing the temporal dependencies and patterns in the SPEI data. These results suggest that the ARIMA model is a more suitable choice for SPEI forecasting in these regions, providing more accurate and reliable predictions than the SVM model.

Table 5: Model Performance Comparison (MSE) for SVM and ARIMA Models

Location	MSE	
	SVM	ARIMA
Nubariyah	1.7746	1.7499
Wadi Al Natrun	2.4113	2.0735
Al Boseli	2.0844	1.8033

Time series actual vs predicted

Time series plots were used to compare the actual SPEI values with those predicted by the SVM and ARIMA models for Nubariyah, Wadi Al Natrun, and Al Boseli. Each subplot displays the actual SPEI values (solid line) alongside the predictions from the SVM model (dashed line) and the ARIMA model (dotted line), as shown in Figure 4. In Nubariyah, the ARIMA model's predictions closely align with the actual values, demonstrating superior performance over the SVM model. In Wadi Al Natrun, the ARIMA model's predictions are more accurate, as evidenced by their tighter alignment with the actual values compared to the SVM model. Similarly, for Al Boseli, the ARIMA model consistently produces predictions closer to the actual SPEI values, indicating better forecasting accuracy than the SVM model. Overall, the ARIMA model outperforms the SVM model in capturing the underlying patterns in the SPEI data across all three locations, showcasing its superior predictive capabilities.

Parity Plots Comparing Actual vs. Predicted Values

Parity plots were generated for each location to visually assess the predictive accuracy of the SVM and ARIMA models, comparing actual SPEI values against predicted values from both models. The 1:1 line in these plots indicates perfect predictions, with points closer to this line representing better predictive performance.

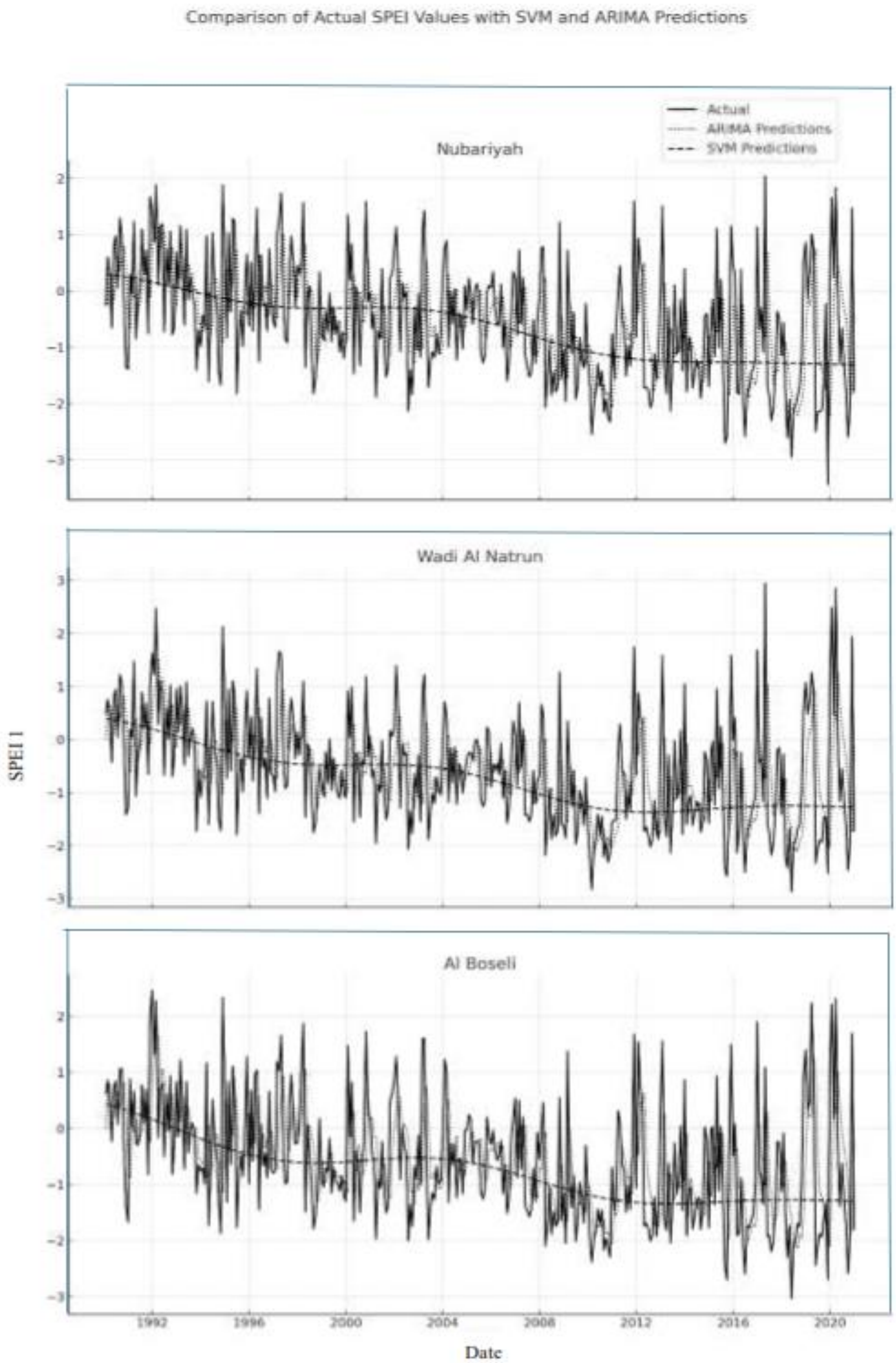


Fig. (4): Comparison of Actual SPEI 1 Values with SVM and ARIMA Predictions for Nubariyah, Wadi Al Natrun, and Al Boseli.

For Nubariyah, the ARIMA model's predictions are notably closer to the actual values than the SVM model's predictions, as evidenced by a higher concentration of points near the 1:1 line (Figure 5). This suggests that the ARIMA model provides superior accuracy in forecasting SPEI values for this location, a conclusion further supported by the Mean Squared Error (MSE) values: 1.7499 for ARIMA and 1.7746 for SVM.

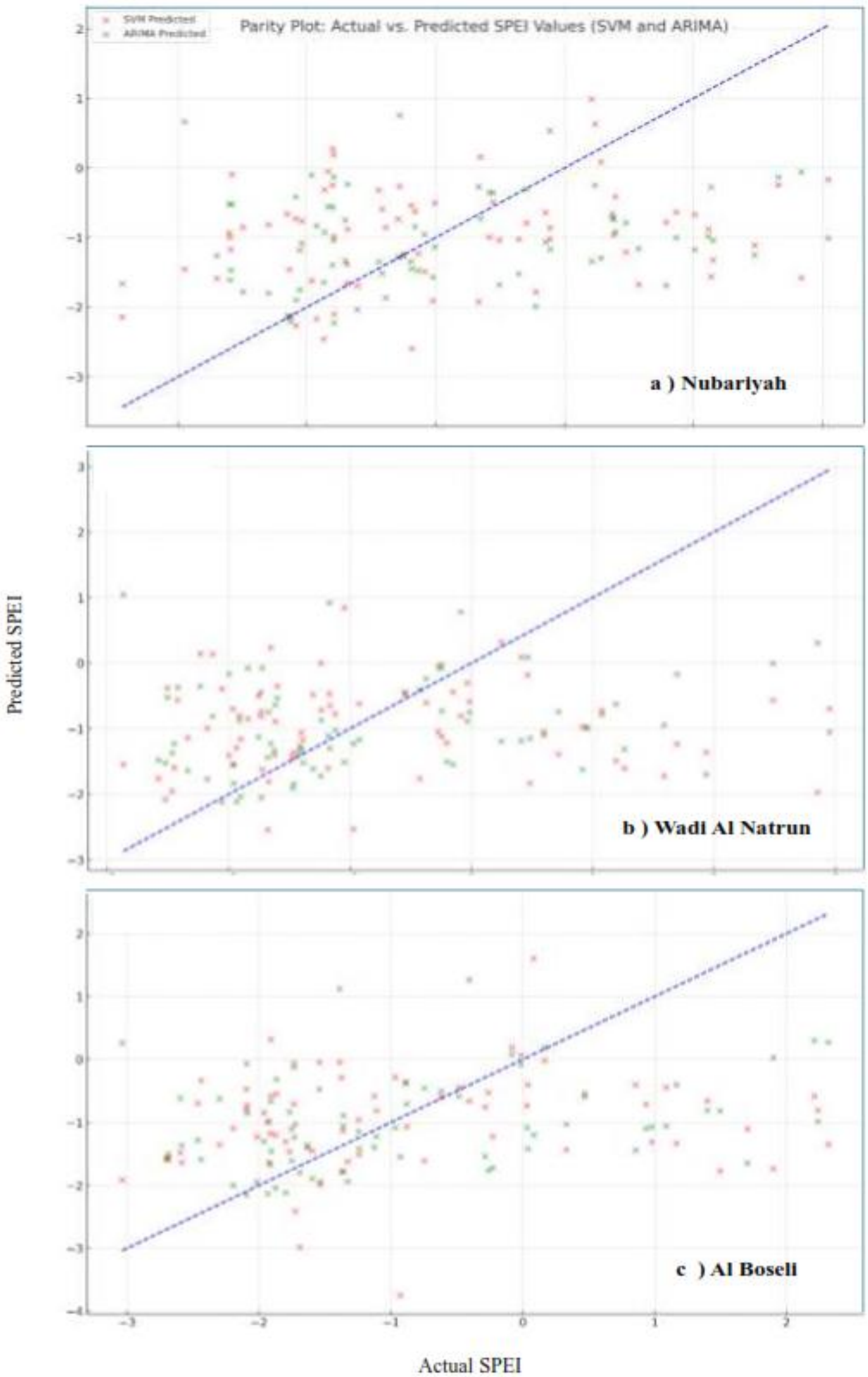


Fig. (5): Parity Plot of Actual vs. Predicted SPEI Values for Nubariyah, Wadi Al Natrun, and Al Boseli Using SVM and ARIMA Models.

Similar trends are observed in the parity plots for Wadi Al Natrun and Al Boseli. In Wadi Al Natrun, the ARIMA model again outperforms the SVM model, with its predictions more tightly clustered around the 1:1 line and an MSE of 2.0735, compared to 2.4113 for the SVM model. For Al Boseli, the ARIMA model consistently demonstrates higher predictive accuracy, with points closer to the 1:1 line and an MSE of 1.8033 versus 2.0844 for the SVM model. Overall, the parity plots for all three locations—Nubariyah, Wadi Al Natrun, and Al Boseli—clearly indicate that the ARIMA model provides more accurate predictions of SPEI values than the SVM model. The lower MSE values for ARIMA across all locations underscore its effectiveness in capturing the underlying patterns in the SPEI data, making it a more reliable tool for climatic forecasting in these regions.

CONCLUSION

- This study aimed to compare the predictive performance of Support Vector Machines (SVM) and AutoRegressive Integrated Moving Average (ARIMA) models in forecasting the Standardized Precipitation Evapotranspiration Index (SPEI) for three locations in Egypt: Nubariyah, Wadi Al Natrun, and Al Boseli. The analysis, which included time series decomposition and model evaluation using Mean Squared Error (MSE) and Mean Absolute Error (MAE), revealed that the ARIMA models consistently outperformed the SVM models across all three locations. The ARIMA(1,1,1) model demonstrated lower MSE and MAE values, with MSE reductions ranging from 1.4% to 14% compared to the SVM models, highlighting ARIMA's superior predictive accuracy and robustness.

- Additionally, the study found that the ARIMA(1,1,1) model was the best fit for all three locations, indicating similar underlying temporal dynamics of SPEI values across these regions. The decomposition of SPEI values into trend, seasonal, and residual components provided further insights: the trend components tended to increase dryness over the 30-year period, while the seasonal components highlighted regular annual fluctuations.

- Accurate forecasting of SPEI values is critical for effective agricultural water management. The ARIMA model's superior performance offers a reliable tool for anticipating drought conditions and planning water usage. Implementing ARIMA-based forecasts in water management strategies can lead to more informed decision-making, ensuring sustainable water use and enhancing agricultural resilience to climatic variability. Thus, the ARIMA(1,1,1) model is recommended for forecasting SPEI values in Nubariyah, Wadi Al Natrun, and Al Boseli. Future research should explore advanced models, such as Recurrent Neural Networks (RNN), to further enhance forecasting accuracy and consider expanding the analysis to additional regions and more recent data to validate and extend the findings.

REFERENCES

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B. (2014). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. International Journal of Climatology, 34(10), 3001-3023. https://doi.org/10.1002/joc.3887.

Box, G. E. P., and Jenkins, G. M. (1976). Time series analysis: Forecasting and control. Holden-Day.

Ghosh, S., and Mujumdar, P. P. (2008). Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Advances in Water Resources, 31(1), 132-146. https://doi.org/10.1016/j.advwatres.2007.07.005.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273-297. https://doi.org/10.1007/BF00994018.

De Livera, A.M., Hyndman, R.J., Snyder, R.D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. J. Am. Stat. Assoc. 106 (496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771.

Han, P., Wang, P.X., Zhang, S.Y., 2010. Drought forecasting based on the remote sensing data using ARIMA models. Math. Comp. Model. 51, 1398–1403. https://doi.org/10.1016/j.mcm.2009.10.031.

Hao, Z., Singh, V.P., Xia, Y., (2018). Seasonal drought prediction: advances, challenges, and future prospects. Rev. Geophys. 56 (1), 108–141. https://doi.org/10.1002/2016RG000549.

Hipel, K. W., & McLeod, A. I. (1994). Time series modelling of water resources and environmental systems. Elsevier.

Mishra, A.K., Singh, V.P., (2011). Drought modeling–a review. J. Hydrol. 403, 157–175. https://doi.org/10.1016/j.jhydrol.2011.03.049.

Montanari, A., Rosso, R., and Taqqu, M. S. (2006). A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Water Resources Research, 36(5), 1249-1259. https://doi.org/10.1029/1999WR900002.

Shi, S., Yu, J., Wang, F., Wang, P., Zhang, Y., Jin, K., (2021). Quantitative contributions of climate change and human activities to vegetation changes over multiple time scales on the Loess Plateau. Sci. Total Environ. 755 (Pt 2), 142419. https://doi.org/10.1016/j.scitotenv.2020.142419.

Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geographical Rev., 38: 55-94. DOI: 10.2307/210739.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23 (7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.

Wu, X., Zhou, J., Yu, H., Liu, D., Xie, K., Chen, Y., Hu, J., Sun, H., Xing, F., 2021. The development of a hybrid wavelet-ARIMA-LSTM model for precipitation amounts and drought analysis. Atmosphere 12, 74. https://doi.org/10.3390/atmos12010074.

Xu, D., Zhang, Q., Ding, Y., Zhang, D., (2022). Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ. Sci. Pollut. Res. Int. 29 (3), 4128–4144. https://doi.org/10.1007/s11356-021-15325-z.

Yao, N., Li, Y., Liu, Q., Zhang, S., Chen, X., Ji, Y., Liu, F., Pulatov, A., Feng, P., (2022). Response of wheat and maize growth yields to meteorological and agricultural droughts based on standardized precipitation evapotranspiration indexes and soil moisture deficit indexes. Agric. Water Manag. 266, 107566. https://doi.org/10.1016/j.agwat.2022.107566.

Zhou, Z., Ding, Y., Shi, H., Cai, H., Fu, Q., Liu, S., Li, T., (2020). Analysis and prediction of vegetation dynamic changes in China: past, present and future. Ecol. Indicat. 117, 106642. https://doi.org/10.1016/j.ecolind.2020.106642.

التنبؤ بمؤشر الهطول والبخر-النتح المعياري(SPEI) باستخدام نماذج (SVM) و (ARIMA): دراسة مقارنة

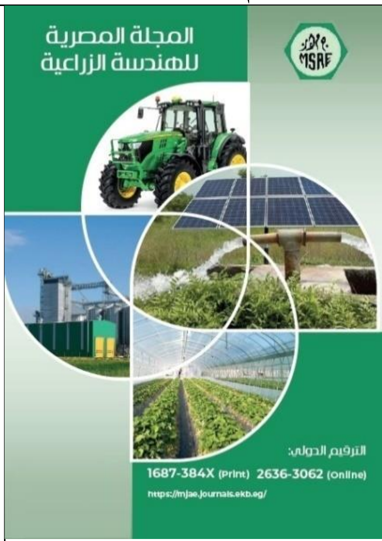
عمرو مسعد[١] ، سلوى حسن عبده [٢]

[١] استاذ مساعد - قسم الهندسة الزراعية - كلية الزراعة - جامعة عين شمس - القاهرة - مصر.

[٢] مدرس - - قسم الهندسة الزراعية - كلية الزراعة - جامعة عين شمس - القاهرة - مصر.

الملخص العربي

تهدف هذه الدراسة إلى تقييم دقة التنبؤ بمؤشر الهطول والبخر-النتح المعياري (SPEI)باستخدام نماذج (SVM) و (ARIMA)في ثلاث مناطق زراعية مهمة في مصر: النوبارية، وادي النطرون، والبوصيلي. يعد التنبؤ الدقيق بـ SPEIأمرًا حيويًا لإدارة المياه والتخطيط الزراعي الفعال، خاصة في المناطق القاحلة. تم إجراء تحليل شامل يتضمن تحليل السلاسل الزمنية وتقييم النماذج باستخدام متوسط الخطأ التربيعي. وأظهرت النتائج تفوق نماذج ARIMA باستمرار على نماذج SVM في جميع المواقع. حيث أظهر نموذج ARIMA(1,1,1)دقة تنبؤية فائقة، مع انخفاض في قيمة MSE تتراوح بين ١,٤٪ و١٤٪ مقارنة بنماذجSVM . ففي النوبارية، حقق نموذج ARIMA قيمة MSEقدرها ١,٧٤٩٩ مقارنة بـ ١,٧٧٤٦ لنموذج SVM . في وادي النطرون، بلغ متوسط الخطأ التربيعي لنموذجARIMA 2.0735 ، وهو أقل بكثير من ٢,٤١١٣ لنموذج SVM ، بينما في البوصيلي، سجل نموذج ARIMA قيمة MSEقدرها ١,٨٠٣٣ مقابل ٢,٠٨٤٤ لنموذج SVM . وهو ما يؤكد الأداء المتفوق لنموذج ARIMA وفعاليته في توقع ظروف الجفاف. وقد كشف ايضاً تحليل قيم SPEI عن اتجاه طويل المدى نحو زيادة الجفاف على مدار الثلاثين عامًا، بالإضافة إلى التقلبات السنوية المنتظمة. ينبغي على الأبحاث المستقبلية استكشاف نماذج متقدمة مثل الشبكات العصبية المتكررة (RNN) لتعزيز دقة التنبؤ بشكل أكبر، وتوسيع التحليل ليشمل مناطق إضافية تحتوي على بيانات أحدث للتحقق من صحة هذه النتائج، وبالتالي تحسين التنبؤ بالجفاف وإدارة الموارد المائية.



الكلمات المفتاحية:
التنبؤ بالجفاف؛ التنبؤ بالسلاسل الزمنية؛ إدارة المياه الزراعية.