

Multi-Scale Long-Term Forecasting of Meteorological Drought Using Wavelet-Decomposed LSTM Models across Six Climatic Regions of Tanzania

Paul T.S. Limbu

paul.limbu@gmail.com

Environmental and Atmospheric Sciences Group, Physics Department, University of Dar es Salaam
Andongwisye John Mwakisisile

Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam

Research Article

Keywords: Wavelet-LSTM, Meteorological drought forecasting, Standardized Precipitation Index, Deep learning, Tanzania

Posted Date: September 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10747820/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Multi-Scale Long-Term Forecasting of Meteorological Drought Using Wavelet-Decomposed LSTM Models across Six Climatic Regions of Tanzania

Paul T.S. Limbu¹ and Andongwisye John Mwakisisile ²

¹ Environmental and Atmospheric Sciences Group, Physics Department, University of Dar es Salaam,
P.O. Box 35063, Dar es Salaam, Tanzania.

² Department of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam,
P.O. Box 35091, Dar es Salaam, Tanzania.

*Corresponding author Email: Paul T.S. Limbu, Email: paul.limbu@ymail.com

Co-author Email: johnandongwisye@gmail.com

Abstract

Meteorological drought is a recurrent climate hazard that threatens agriculture, water resources, and socioeconomic development in Tanzania. However, accurate drought forecasting remains challenging due to the nonlinear and non-stationary nature of hydroclimatic processes, despite advances in deep learning. This study proposes a hybrid Wavelet-Long Short-Term Memory (Wavelet-LSTM) model for multi-scale forecasting of meteorological drought across six climatically distinct regions of Tanzania. Meteorological variables from the ERA5 reanalysis dataset, including precipitation, temperature, wind components, wind speed, dew-point temperature, and mean sea-level pressure, were used to predict the Standardized Precipitation Index (SPI) at 3- and 6-month accumulation periods. Wavelet decomposition was applied to reduce noise and extract dominant temporal features before model training. The dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets, and model performance was evaluated using the coefficient of determination (R^2), correlation coefficient (R), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), Nash-Sutcliffe efficiency (NSE), Willmott's index of agreement (d), and analysis of variance (ANOVA). The proposed model achieved consistently high predictive accuracy across all climatic regions, with testing R^2 ranging from 0.86 to 0.97, correlation coefficients between 0.94 and 0.99, NSE values of 0.86-0.97, and WI values of 0.96-0.99, while maintaining low prediction errors. ANOVA results ($p > 0.05$) indicated no statistically significant differences between observed and predicted SPI values, confirming the robustness and generalization capability of the model. These findings demonstrate that integrating wavelet decomposition with LSTM substantially improves long-term drought forecasting and provides a reliable framework for drought early warning, climate adaptation, and water resources management in Tanzania and other hydroclimatically diverse regions.

Keywords: Wavelet-LSTM; Meteorological drought forecasting; Standardized Precipitation Index; Deep learning; Tanzania.

41

42 **Introduction**

43 Drought is among the most complex and damaging natural hazards worldwide, with far-reaching impacts on water resources,
44 agriculture, ecosystems, energy production, and socio-economic development (Mishra & Singh, 2010; Soomro et al., 2025).
45 Unlike rapid-onset hazards such as floods and tropical cyclones, drought develops slowly, evolves over extended periods, and
46 may persist for months or even years, making its onset and termination difficult to identify accurately (Yang et al., 2025).
47 Meteorological drought, commonly defined as a sustained period of below-normal precipitation, represents the first stage of
48 drought development and often propagates into agricultural, hydrological, and socio-economic droughts (Mishra & Singh,
49 2010). Consequently, reliable forecasting of meteorological drought is essential for enhancing drought preparedness,
50 improving early warning systems, and supporting sustainable water resources management. East Africa region, including
51 Tanzania is highly vulnerable to drought owing to its strong dependence on rain-fed agriculture, limited adaptive capacity,
52 and increasing climate variability (IPCC, 2023). The region has experienced recurrent severe droughts over recent decades,
53 including the devastating events of 2008-2009 and 2010-2011, which resulted in widespread food insecurity and humanitarian
54 crises affecting millions of people (Funk, 2011; Mwangi et al., 2014). Studies have shown that the frequency and severity of
55 droughts in East Africa have increased over recent decades, posing significant challenges to livelihoods, agricultural
56 productivity, and regional economic growth (AghaKouchak, 2015; Nicholson, 2017). In Tanzania, droughts frequently affect
57 both unimodal and bimodal rainfall regimes, disrupting crop production, reducing water availability, and threatening food and
58 energy security (Limbu & Guirong, 2019; King'uza et al., 2025). The impacts are particularly severe in semi-arid regions
59 such as central Tanzania, where rainfall variability is inherently high and water resources are already constrained.

60 The characterization and monitoring of meteorological drought are commonly based on standardized drought indices. Among
61 these, the Standardized Precipitation Index (SPI) is the most widely adopted because of its simplicity, flexibility across
62 multiple timescales, and minimal data requirements (World Meteorological Organization, 2012). SPI has been extensively
63 used for drought assessment and monitoring at regional and global scales and is recommended by the World Meteorological
64 Organization as the standard indicator for meteorological drought monitoring (Hayes et al., 2011). Multi-timescale SPI
65 analysis is particularly important because short accumulation periods (e.g., SPI3) capture seasonal moisture deficits, whereas
66 longer timescales (e.g., SPI6, SPI12, and beyond) reflect prolonged drought conditions relevant for water resources
67 management (Vicente-Serrano et al., 2010). Although drought monitoring has substantially advanced, accurate long-term
68 drought forecasting remains challenging because drought processes are inherently nonlinear, nonstationary, and influenced
69 by temporal persistence across multiple timescales (Belayneh & Adamowski, 2013; Shafik, 2026). Traditional statistical
70 approaches, including autoregressive models and regression-based techniques, often fail to adequately capture these complex
71 temporal dynamics (Mishra & Singh, 2011). Consequently, machine learning and deep learning techniques have increasingly
72 been employed in drought prediction studies due to their ability to learn nonlinear relationships from historical data without
73 requiring explicit assumptions regarding the underlying physical processes (Mosavi et al., 2018).

74 Several studies have demonstrated the potential of machine learning techniques for drought forecasting. Artificial Neural
75 Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), Decision Trees (DT), and Long Short-Term
76 Memory (LSTM) networks have shown improved predictive performance compared with conventional statistical models

(Bouaziz et al., 2025; Piri et al., 2023). In South Africa, ANN-based approaches successfully reproduced drought variability and enhanced predictive accuracy relative to traditional techniques (Onyeuwaoma et al., 2024). Similarly, studies in Iran employed machine learning algorithms integrated with SPI and SPEI to classify drought severity and demonstrated the superiority of ensemble and deep learning approaches over single models (Dehghani & Molavi-Arabshahi, 2026). In Thailand, Humphries et al. (2024) reported that hybrid wavelet-based models consistently outperformed standalone Bi-LSTM and Random Forest models for forecasting SPI and other drought indices across multiple timescales. Among deep learning methods, LSTM networks have attracted considerable attention because they are specifically designed to capture long-term temporal dependencies in sequential data while overcoming the vanishing gradient problem associated with traditional recurrent neural networks (Mienye, 2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9), 517.). Recent studies have successfully applied LSTM and its variants in hydrometeorological forecasting, including precipitation prediction, drought monitoring, and seasonal forecasting (Waqas et al., 2024). Nevertheless, standalone LSTM models may exhibit reduced performance when applied to highly nonstationary drought time series, since meteorological observations often contain noise and multiple temporal scales that are difficult to represent using a single learning architecture (Charjan et al., 2026). Wavelet decomposition has emerged as an effective preprocessing technique for addressing nonstationarity in hydroclimatic time series. By decomposing a signal into approximation and detail components at different frequencies, wavelet analysis facilitates noise reduction and extraction of dominant temporal features while preserving the intrinsic characteristics of the original series (Addison, 2017).

Previous studies have demonstrated that integrating wavelet decomposition with machine learning or deep learning models substantially improves forecasting performance. For example, wavelet-enhanced ANN and LSTM models significantly improved rainfall and drought predictions in Malaysia, Thailand, and Australia (Khan et al., 2018; Waqas et al., 2024; Seifi et al., 2026). Furthermore, recent studies reported that Wavelet-LSTM hybrid frameworks outperform conventional LSTM models in forecasting SPI at multiple timescales because they effectively capture both short-term fluctuations and long-term persistence in drought processes (Humphries et al., 2024). Despite these advances, significant research gaps remain in Tanzania and the broader East African region. First, most drought studies in Tanzania have primarily focused on drought monitoring, characterization, and trend analysis rather than long-term forecasting (Limbu et al., 2026). Second, existing forecasting studies have largely employed conventional machine learning techniques, with limited application of hybrid deep learning frameworks capable of addressing the nonstationary nature of drought time series. Third, few studies have investigated multi-scale long-term drought forecasting simultaneously across contrasting climatic zones in Tanzania. Consequently, the potential of wavelet-decomposed deep learning models for long-term meteorological drought forecasting remains insufficiently explored in the country and the region. This study addresses these knowledge gaps by developing a hybrid Wavelet-LSTM framework for multi-scale long-term forecasting of meteorological drought across six climatically distinct regions of Tanzania. Specifically, the study aims to: (i) characterize meteorological drought using SPI at multiple accumulation periods; (ii) develop a Wavelet-LSTM model for forecasting drought conditions at extended lead times; and (iii) evaluate the capability of the proposed framework in reproducing drought variability across different climatic regimes. By integrating wavelet decomposition with LSTM networks, this study is expected to improve long-term drought forecasting skill in Tanzania. The findings of this study have important implications for drought early warning, agricultural planning, water resources management, and climate-risk reduction in Tanzania and East Africa. Scientifically, the study contributes to

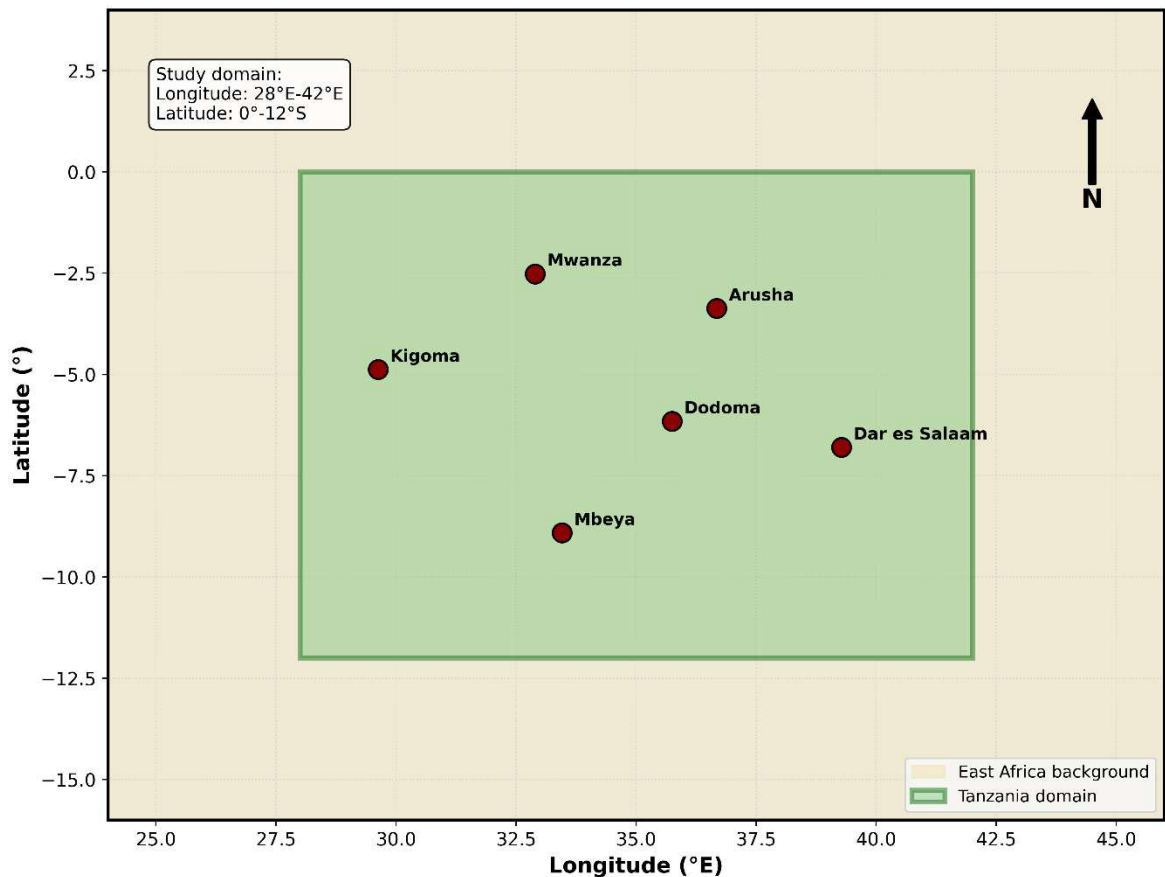
114 the growing body of knowledge on hybrid deep learning applications in hydrology by demonstrating the effectiveness of
115 wavelet-decomposed LSTM models for forecasting nonstationary drought processes in data-constrained environments.

116

117 **Material and methods**

118 **Study Area**

119 The study focuses on six regions of Tanzania with different climatic zones (refer to Fig. 1). Tanzania is situated
120 in East Africa at Latitudes 0-12 ° South of the Equator and 28-42 ° East. The Map below shows the location of
121 Tanzania and the study regions.
122



123

124 **Fig. 1 The Map of six different region across Tanzania under this study**

125

126 Table 1 summarizes the geographical characteristics of the six regions selected to represent Tanzania's principal climatic
127 zones. The study sites cover the northern highlands, coastal belt, semi-arid central plateau, southern highlands, and the Lake
128 Victoria and Lake Tanganyika basins, thereby capturing the country's diverse rainfall regimes and hydroclimatic conditions.
129 These climatic contrasts are expected to influence drought development and forecasting performance, making the selected
130 regions appropriate for evaluating the robustness of the proposed Wavelet-LSTM model under varying environmental
131 conditions. The selection is consistent with previous studies that have documented the pronounced spatial variability of rainfall
132 and drought across Tanzania and East Africa (Limbu & Guirong, 2019).
133

134 **Table 1:** Geographical information of the six locations selected for this study

135

Region	Longitude	Latitude	Geographical information
Arusha	-3.3869° S	36.6830° E	Cool sub-humid highland climate or Northern highland climate characterized by bimodal rainfall.

Dar es Salaam	-6.82349°S	39.26951°E	Hot, humid tropical coastal climate with bimodal rainfall.
Dodoma	-6.161184°S	35.745426°E	Semi-arid tropical climate or Semi-arid central plateau climate
Kigoma	-4.883°S	29.633°E	Humid tropical lake basin climate or Western tropical climate influenced by Lake Tanganyika
Mbeya	-8.9094°S	33.4608 °E	Southern highland temperate climate or Cool highland climate
Mwanza	-2.51667°S	32.9°E	Tropical lake basin climate or Humid tropical climate influenced by Lake Victoria

Data sources

The study employed meteorological variables obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis dataset, which provides continuous, physically consistent, and globally complete atmospheric information by assimilating observations from satellites, radiosondes, weather stations, and numerical weather prediction models (Hersbach et al., 2020). ERA5 has become one of the most widely used climate reanalysis datasets because of its high spatial and temporal resolution, improved representation of atmospheric processes, and demonstrated ability to reproduce precipitation variability and extreme climate events across different regions of the world. Previous investigations have shown that ERA5 performs well over East Africa, including Tanzania, making it a reliable data source for hydroclimatic studies conducted in areas where long-term observational records remain limited (Gleixner et al., 2020). The predictor variables selected for this investigation were chosen to represent the principal atmospheric mechanisms influencing rainfall variability and drought development. These variables included the 10-m zonal wind component (u_{10}), 10-m meridional wind component (v_{10}), 10-m wind speed (si_{10}), 2-m air temperature (t_{2m}), 2-m dew-point temperature (d_{2m}), mean sea-level pressure (MSLP), and total precipitation (tp). Together, these variables describe near-surface atmospheric circulation, thermodynamic conditions, moisture availability, and precipitation processes that regulate drought evolution across Tanzania. To improve model robustness, the available dataset was partitioned into three independent subsets. Seventy percent (70%) of the observations were allocated for model training, fifteen percent (15%) for validation during model development, and the remaining fifteen percent (15%) were reserved for independent testing. The training subset was used to estimate network parameters, the validation subset monitored model generalization and assisted in preventing overfitting, whereas the testing subset provided an unbiased assessment of predictive performance. The integration of multiple atmospheric variables within a unified modelling framework enables simultaneous representation of local meteorological processes and broader climate controls. Such integration has been shown to improve drought prediction relative to models relying exclusively on precipitation because drought development reflects interactions among atmospheric circulation, land-surface conditions, and moisture transport processes rather than precipitation anomalies alone (Ghosh et al., 2024). Consequently, the selected ERA5 variables provide a physically meaningful foundation for developing machine learning models capable of forecasting meteorological drought conditions across Tanzania.

Data analysis

Standardized Precipitation Index

The Standardized Precipitation Index (SPI) was used to quantify meteorological drought. For a given accumulation scale N , precipitation is accumulated as:

$$P_N(t) = \sum_{i=1}^N P(t-i) \quad (1)$$

Where $P(t)$ is the precipitation in the month t and $P_N(t)$ is the accumulated precipitation over N months. To obtain the values of SPI3 and SPI6, the accumulated scale is set $N = 3$ and $N = 6$.

SPI is computed by fitting the accumulated precipitation series to a probability distribution and transform the cumulative probability to a standard normal variable:

$$SPI = \phi^{-1}(G(P_N)) \quad (2)$$

Where ϕ^{-1} is the inverse standard normal distribution and $G(P_N)$ is accumulative probability of accumulated precipitation.

Wavelet Decomposition

Wavelet decomposition was used to the SPI series in order to reduce noise extract temporal pattern. We apply the discrete Wavelet decomposition for computational procedure since it is more straightforward and widely used in time series forecasting when combined with statistical and machine learning methods (Rezaei & Shabri, 2025). In this study, the discrete wavelet transform decompose the a signal through approximation and details components:

$$Y(t) = B_j(t) + \sum_{i=1}^j D_i(t) \quad (3)$$

Where $B_j(t)$ is the approximation component at level j and $D_i(t)$ are the detail components. Wavelet decomposition with Daubechies wavelets enables the segmentation of input signals into hierarchical subseries (Seifi et al., 2026).

Long Short-Term Memory

The LSTM network was used to forecast drought indices because it can capture temporal dependence in sequential data. The LSTM model uses memory cells controlled by input, forget, and output gates: According to Lamichhane et al (2026), the LSTM model is capable to capture temporal dependencies, making it particularly well-suited for identifying dynamic patterns in time-series climatic data. When well suited for time series forecasting, the LSTM efficiently learns and preserves long-term dependencies, allowing it to capture complex temporal patterns in sequential data (Krichen & Mihoub, 2025). The LSTM model uses memory cells controlled by input, forget, and output gates:

$$f_t = \sigma(\omega_f[h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(\omega_i[h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(\omega_c[h_{t-1}, x_t] + b_c) \quad (6)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (7)$$

$$o_t = \sigma(\omega_o[h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \tanh(C_t) \quad (10)$$

Where f_t , i_t and o_t are the forget, input, and output gates, respectively; C_t is the cell state, h_t is the hidden state, and x_t represents the input variables.

Hybrid Wavelet-LSTM Framework

We introduced a Wavelet-LSTM hybrid framework. Instead of feeding raw data straight into the machine learning model, the Wavelet Transforms (WT) act as a mathematical lens, decomposing complex and noisy climatic data into simpler, multi-resolution frequency components. On the other hand, the Long Short-Term Memory (LSTM) is effectively on capturing long-term temporal dependencies. The Hybrid Wavelet-LSTM Framework is a powerful time-series forecasting model. As Illustrated by Figure 2, The SPI series and Meteorological predictors are decomposed into simpler sub-series in Wavelet transforms. The main function of wavelet decomposition is to reduce noisy in SPI series. The decomposed data which are wallet-transformed SPI, meteorological predictors, seasonal indicators and trend variables are used as input in machine learning model lstm for forecasted. The LSTM models train SPI3 and SPI6 series separately. Finally, the LSTM generate 24-months forecasts for SPI3 and SPI6.

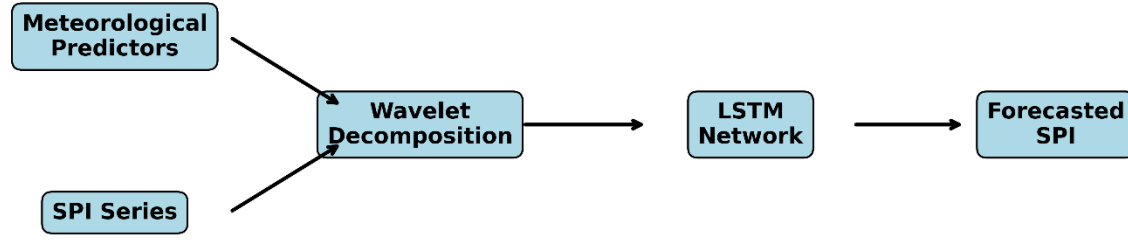


Fig. 2 The mechanism of the Hybrid Wavelet-LSTM framework used in this study

Models evaluation

The proposed framework integrates the wavelet decomposition and deep learning to improve the forecasts of the complex and non-stationary drought processes. Before model training, a Wavelet decomposition was first applied to the SPI series to extract dominant temporal features and reduce noise. The resulting wavelet-extracted indices are then combined with meteorological and seasonal predictors were used as inputs to Long Short-Term Memory networks. The performance of the model was evaluated using the coefficient of correlation (R), coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), Nash–Sutcliffe efficiency (NSE), and Willmott’s index of agreement. Analysis of variance (ANOVA) was further used to assess the statistical similarity between observed and modelled drought indices.

Mean Square Error (MSE):

Mean Squared Error (MSE) measures the average squared difference between observed and predicted SPI values. Smaller MSE values indicate improved predictive accuracy because prediction errors remain relatively small throughout the evaluation period.

$$MSE = \frac{1}{n} \sum_{i=1}^n (SPI_i - \widehat{SPI}_i)^2. \quad (4)$$

Where

n is the sum of observations

Y_i is the observed value of SPI

$\hat{Y}_i$ is the predicted value of SPI

Root Mean Square Error (RMSE):

The Root Mean Squared Error (RMSE) represents the square root of MSE and expresses prediction error in the same units as the observed SPI values. RMSE is widely used because it assigns greater weight to larger prediction errors.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (SPI_i - \widehat{SPI}_i)^2} \quad (5)$$

261 *Mean Absoute Error (MAE):*

262 Mean Absolute Error (MAE) calculates the average absolute deviation between observed and predicted drought indices.
263 Compared with RMSE, MAE is less sensitive to extreme prediction errors and therefore provides a complementary measure
264 of model accuracy.

$$265 \quad MAE = \frac{1}{n} \sum_{i=1}^n |SPI_i - \widehat{SPI}_i| \quad (6)$$

266

267

268 *Correlation Coefficient*

269 The Pearson correlation coefficient was computed to quantify the strength of the linear relationship between observed and
270 simulated SPI values. Correlation values approaching unity indicate strong agreement between observations and model
271 predictions.

$$273 \quad r = \frac{\sum_{i=1}^n (SPI_i - \overline{SPI})(\widehat{SPI}_i - \overline{\widehat{SPI}})}{\sqrt{\sum_{i=1}^n (SPI_i - \overline{SPI})^2} \sqrt{\sum_{i=1}^n (\widehat{SPI}_i - \overline{\widehat{SPI}})^2}} \quad (7)$$

272

274 Here, the $\overline{SPI}$ is the mean of observed SPI and $\overline{\widehat{SPI}}$ is the mean of predicted SPIs.

275 *Coefficient of Determination*

276 The coefficient of determination (R^2) was used to evaluate the proportion of observed SPI variability explained by the ANN
277 model. Larger R^2 values indicate greater explanatory capability and improved predictive performance.

$$278 \quad R = 1 - \frac{\sum_{i=1}^n (SPI_i - \widehat{SPI}_i)^2}{\sum_{i=1}^n (SPI_i - \overline{SPI})^2} \quad (8)$$

279 *Nash-Sutcliffe Efficiency*

280 Model efficiency was further assessed using the Nash–Sutcliffe Efficiency coefficient, which compares predictive
281 performance against the observed mean. Values approaching one indicate excellent predictive capability, whereas values
282 below zero suggest that the model performs worse than simply using the mean of the observations (Duc & Sawada, 2023).

$$283 \quad NSE = 1 - \frac{\sum_{i=1}^n (SPI_i - \widehat{SPI}_i)^2}{\sum_{i=1}^n (SPI_i - \overline{SPI})^2} \cdot \quad (9)$$

284

285 *Willmott Index of Agreement (d)*

286 The Willmott Index of Agreement provides an additional measure of similarity between observed and predicted SPI values
287 by simultaneously considering both systematic and random prediction errors.

$$288 \quad d = 1 - \frac{\sum_{i=1}^n (\widehat{SPI}_i - SPI_i)^2}{\sum_{i=1}^n (|\widehat{SPI}_i - \overline{SPI}| + |SPI_i - \overline{SPI}|)^2} \cdot \quad (10)$$

289

290 The optimal ANN configuration was selected primarily according to the highest testing R^2 , highest NSE, highest correlation
291 coefficient, and the lowest RMSE.

292 Analysis of Variance (ANOVA)

293 To determine whether statistically significant differences existed between observed and predicted SPI values, an Analysis of
294 Variance (ANOVA) was performed. The F-statistic was calculated using

$$295 F = \frac{\text{Between-group variance}}{\text{Within-group variance}} \quad (11)$$

296 A significance level of $\alpha = 0.05$ was adopted. ANN models satisfying the criterion $p > 0.05$ were considered statistically
297 consistent with the observed SPI values, indicating no significant difference between measured and simulated drought
298 conditions.

299 Results and discussion

300
301 Figure 3 shows the temporal variability of the SPI3 across Arusha, Dar es Salaam, Dodoma, Kigoma, Mbeya, and Mwanza
302 during the study period. The SPI3 series exhibits pronounced interannual and seasonal variability, with recurrent oscillations
303 between wet and dry conditions in all regions. Most drought events fall within the moderate drought category ($SPI \leq -1.0$),
304 although severe ($SPI \leq -1.5$) and extreme ($SPI \leq -2.0$) drought episodes are also evident, particularly in Dar es Salaam,
305 Kigoma, and Mwanza. The observed oscillatory behaviour demonstrates the highly variable nature of rainfall regimes in
306 Tanzania, reflecting the influence of both local climatic processes and large-scale atmospheric circulation systems. The SPI3
307 in Arusha shows frequent transitions between wet and dry conditions, with drought events generally remaining within
308 moderate severity. Similar fluctuations in SPI time series for northern Tanzania have been reported by Kalisa et al. (2021),
309 who found that seasonal rainfall variability in East Africa is primarily driven by ITCZ movements and ocean-atmosphere
310 interactions. The Dar es Salaam SPI3 series exhibits greater variability and larger amplitudes than most inland stations,
311 including several severe and extreme drought events. Such high variability reflects the strong sensitivity of coastal rainfall to
312 sea surface temperature anomalies in the western Indian Ocean and the ENSO. Coastal Tanzania is particularly influenced by
313 the IOD, which modulates moisture availability and rainfall intensity (Rohli, 2019). Inter-annual hydroclimatic variability in
314 coastal Tanzania. *International Journal of Climatology*, 39(12), 4736-4750.). The occurrence of intense dry and wet anomalies
315 in Dar es Salaam agrees with findings by Wainwright et al. (2019), who reported that coastal East Africa experiences
316 substantial rainfall fluctuations due to coupled ocean-atmosphere variability.

317
318 In Dodoma, the SPI3 series displays relatively regular seasonal oscillations with fewer extreme events compared with Dar es
319 Salaam. However, persistent moderate drought episodes are evident throughout the record. Dodoma lies within the semi-arid
320 central plateau, where rainfall is inherently low and highly variable. The recurrent drought conditions are therefore
321 characteristic of semi-arid climates, where small reductions in rainfall can readily trigger meteorological drought (Mishra &
322 Singh, 2010). Similar patterns of persistent moderate drought in central Tanzania have been documented by Makula et al.
323 (2024), who attributed them to declining rainfall reliability and increasing rainfall variability. The SPI3 for Kigoma reveals
324 repeated occurrences of severe and extreme drought episodes, particularly during the latter decades of the study period.
325 Western Tanzania is influenced by moisture transport from the Congo Basin and Lake Tanganyika, and fluctuations in these
326 moisture sources can lead to substantial rainfall anomalies (Nicholson, 2017). The increased frequency of negative SPI values
327 may indicate enhanced climatic variability or changing regional circulation patterns. Comparable findings were reported by
328 Spinoni et al. (2014), who observed increasing drought occurrence across several parts of East Africa during recent decades.
329 In Mbeya, the values fluctuate around normal conditions, although several severe drought episodes are apparent. The southern
330 highlands receive relatively high rainfall due to orographic uplift, which tends to moderate drought severity. Nevertheless,
331 interannual variability associated with ENSO and tropical circulation anomalies can still generate substantial rainfall deficits
332 (Rwambo et al., 2025). Similar SPI patterns over the southern highlands have been reported by Näschen et al. (2018), who
333 found that despite generally humid conditions, the region remains vulnerable to episodic drought events. The Mwanza region
334 exhibits some of the largest negative excursions among the study regions, with multiple severe and extreme drought events.
335 The Lake Victoria basin experiences complex interactions between lake-land breezes, regional convection, and large-scale
336 climate modes, resulting in considerable rainfall variability (Nyamweya et al., 2016). The pronounced drought episodes
337 observed in Mwanza may therefore be associated with disruptions in lake-induced convection and broader regional climate
338 anomalies. Similar observations were reported by Yang et al. (2014), who showed that rainfall variability around Lake Victoria
339 is highly sensitive to regional atmospheric circulation changes.

340
341

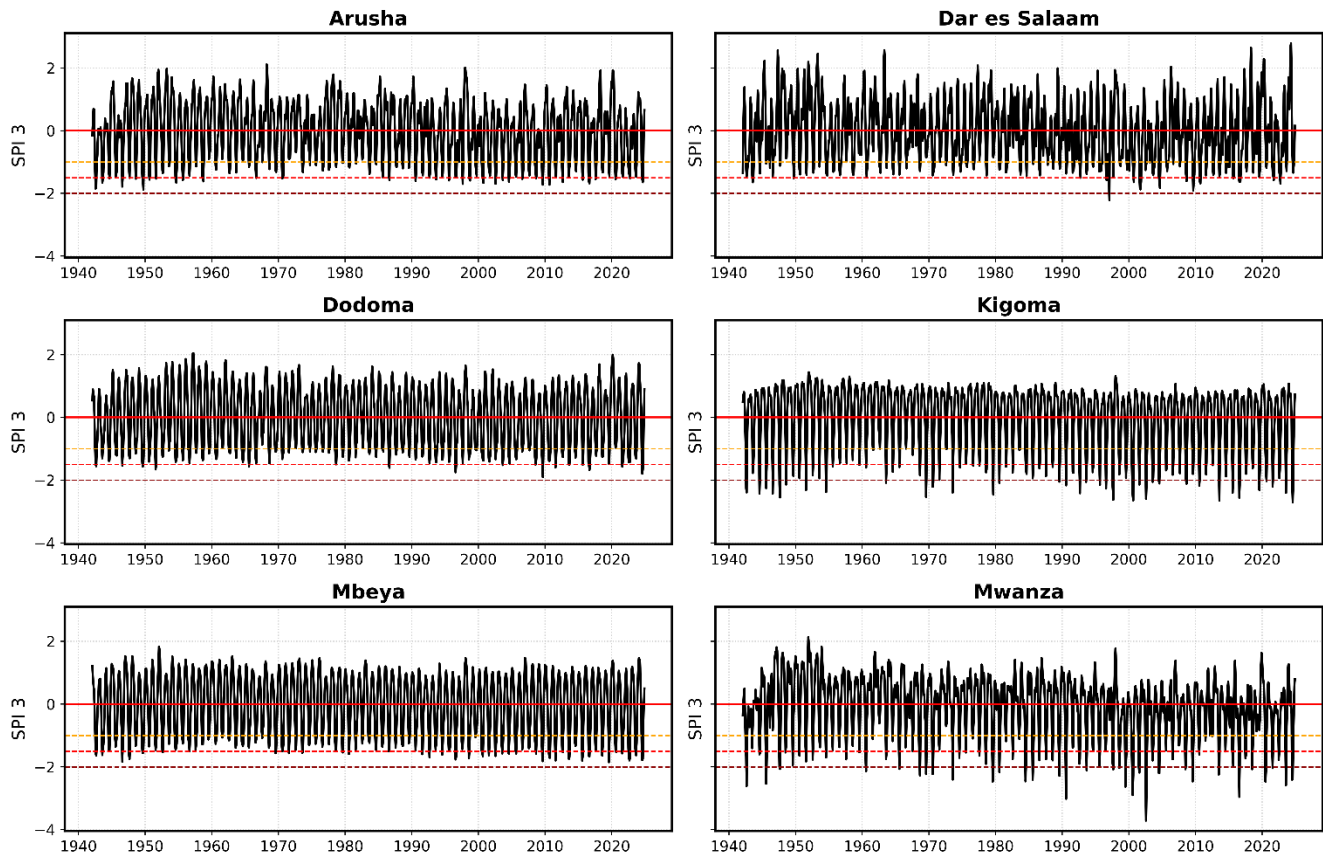


Fig.3 SPI3 time series for the study period

Compared with SPI3, the SPI6 series in Figure 4 exhibits smoother fluctuations and more persistent wet and dry phases, reflecting the longer accumulation period across all the six study regions. The reduced short-term variability and enhanced persistence indicate that SPI6 captures medium-term moisture conditions more effectively, making it particularly suitable for assessing agricultural and hydrological drought impacts (Svoboda et al., 2012). Across all regions, prolonged periods of negative SPI6 values are evident, indicating sustained drought conditions that could significantly affect soil moisture, streamflow, and agricultural productivity. In Arusha, the SPI6 series displays prolonged wet and dry episodes with fewer abrupt fluctuations than observed in SPI3. The smoothing effect observed in SPI6 reflects the integration of rainfall anomalies over a longer period, thereby reducing the influence of individual rainfall events. Similar findings were reported by Zhang et al., (2017), who demonstrated that longer SPI accumulation periods are more effective for detecting sustained drought conditions affecting agriculture. The Dar es Salaam SPI6 series continues to exhibit considerable variability, although extreme fluctuations are less pronounced than in SPI3. Several extended drought periods are evident, especially after the 1990s. The persistence of these drought episodes suggests that rainfall deficits in coastal Tanzania can extend over multiple seasons, potentially affecting water resources and urban water supply systems. This observation is consistent with Nicholson (2017), who noted that East African coastal rainfall is strongly influenced by large-scale oceanic forcing, resulting in prolonged hydroclimatic anomalies.

For Dodoma, SPI6 reveals recurrent medium-term drought episodes with extended durations, characteristic of semi-arid environments. The persistence of negative SPI6 values highlights the vulnerability of central Tanzania to prolonged rainfall deficits. Similar medium-term drought persistence in central Tanzania has been reported by Makula et al. (2024) and Gebrechorkos et al. (2020). The SPI6 series in Kigoma indicates increasingly persistent drought conditions during the study period. Several events exceed the severe drought threshold, suggesting prolonged moisture deficits. Such persistent drought conditions may be associated with changes in regional moisture transport and variability in the Congo air mass influence. Similar long-duration drought events in western Tanzania and neighbouring regions have been documented by Spinoni et al. (2014). The SPI6 series for Mbeya demonstrates relatively stable conditions compared with other regions, although prolonged drought episodes are still evident. The southern highlands benefit from orographic rainfall enhancement, which tends to reduce drought persistence. Nevertheless, sustained negative SPI6 values indicate that the region is not immune to prolonged rainfall deficits. Similar behaviour has been observed by Näschen et al. (2018), who reported periodic drought occurrences despite

the generally humid climate of southern Tanzania. SPI6 in Mwanza region reveals several extended drought episodes, particularly during the late twentieth and early twenty-first centuries. The persistence of these drought conditions suggests that cumulative rainfall deficits around Lake Victoria can extend over several seasons, potentially affecting lake levels, fisheries, and agricultural activities. Comparable findings have been reported by Yang et al. (2014), who highlighted the sensitivity of the Lake Victoria basin to regional climate variability and large-scale atmospheric circulation anomalies.

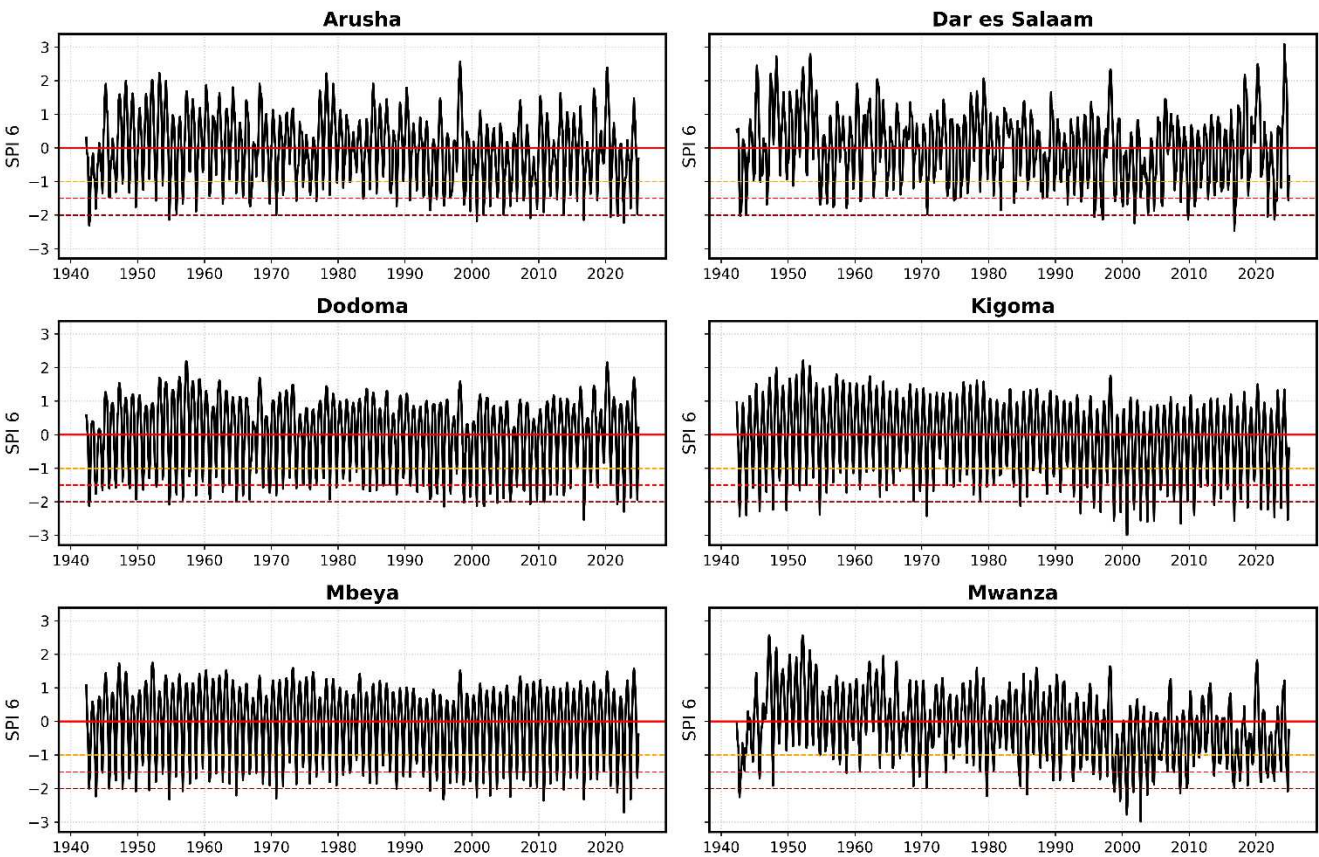


Fig.4 SPI6 time series for the study period

Table 2 presents the percentage occurrence of different drought severity categories based on SPI3 and SPI6 across the six study regions. Overall, mild drought was the most frequently observed category in all regions, accounting for approximately 17–37% of the study period. The highest occurrence was recorded in Mwanza at SPI6 (36.69%), followed by Arusha (31.75%) and Dar es Salaam (30.14%), indicating that short- to medium-term moisture deficits are common across Tanzania. This finding agrees with previous studies showing that mild drought events occur more frequently than severe and extreme droughts because they represent normal climatic fluctuations within semi-arid and sub-humid environments (Heim, 2002). Moderate drought occurred less frequently than mild drought but remained a common feature across all regions. Dodoma recorded the highest SPI3 occurrence (19.89%), reflecting the greater susceptibility of the semi-arid central plateau to rainfall deficits and prolonged dry spells. Similar findings have been reported for central Tanzania, where recurrent droughts are associated with high rainfall variability and limited moisture availability (Mwinuka et al., 2021). Severe and extreme droughts were comparatively rare, representing less than 9% and 5% of observations, respectively. Extreme drought was absent at SPI3 in Arusha, Dodoma, and Mbeya but occurred in Kigoma (4.62%) and Mwanza (4.32%), suggesting that regions influenced by large lakes are not immune to severe moisture deficits despite generally higher rainfall. Previous studies have similarly shown that East African droughts are spatially heterogeneous and are strongly influenced by large-scale climate drivers such as the ENSO and the IOD, which affect both the frequency and severity of drought events (Yang et al., 2015). A comparison between SPI3 and SPI6 indicates a shift in drought characteristics with increasing accumulation period. The frequency of mild drought generally increased at SPI6 in regions such as Arusha, Kigoma, and Mwanza, whereas moderate drought became less frequent in most regions. Conversely, severe and extreme drought occurrences increased slightly at the six-month timescale in several regions, particularly Dodoma and Mbeya, suggesting that prolonged rainfall deficits are more evident over longer accumulation periods. This behaviour is consistent with the characteristics of the Standardized Precipitation Index, whereby

longer time scales smooth short-term rainfall variability while capturing persistent moisture deficits that are more relevant for seasonal drought monitoring (Vicente-Serrano et al., 2010).

Table 2 Percentage of occurrence of the different drought categories for study period. Values are in percentage (%).

Drought category	Arusha		Dar es Salaam		Dodoma		Kigoma		Mbeya		Mwanza	
	SPI3	SPI6	SPI3	SPI6	SPI3	SPI6	SPI3	SPI6	SPI3	SPI6	SPI3	SPI6
Mild drought	25.33	31.75	31.26	30.14	24.42	21.77	17.19	27.02	23.12	22.78	25.83	36.69
Moderate drought	15.38	11.19	16.78	9.58	19.89	9.58	7.84	10.69	17.59	10.38	8.84	9.98
Severe drought	5.13	4.74	2.41	5.85	2.11	8.37	6.73	5.75	5.83	7.66	5.53	6.35
Extreme drought	0	0.91	0.10	1.11	0	1.41	4.62	3.13	0	2.22	4.32	1.41

Figure 5 presents the relationship between observed and predicted SPI3 values for Arusha generated by the Hybrid Wavelet-LSTM model during the training, validation, testing, and combined validation-testing stages. The model exhibited outstanding predictive capability, with correlation coefficients ranging from 0.979 to 0.992, coefficients of determination varying between 0.958 and 0.983, and RMSE values remaining low (0.122-0.197). These results are in agreement with Rezaei and Shabri (2025), who showed that integrating wavelet decomposition with LSTM networks enhances drought prediction by suppressing noise and extracting hidden temporal features from SPI time series. Likewise, the high correlations observed in this study are comparable to those reported by Nourani et al. (2015) using wavelet-based deep learning approaches for drought forecasting. The model also demonstrated excellent performance for SPI6 in Arusha (S1) in the supplementary materials, yielding correlation coefficients between 0.967 and 0.990, R^2 values ranging from 0.934 to 0.980, and RMSE values of 0.133-0.246. Although slightly larger errors were observed during the testing phase, most predicted values remained closely distributed around the ideal line, reflecting good generalisation ability. The modest decline in performance compared with SPI3 is likely associated with the longer accumulation period of SPI6, which incorporates additional low-frequency climatic fluctuations and stronger drought persistence. Similar observations have been documented by Hinge et al. (2022), who reported that hybrid wavelet-deep learning frameworks retain high forecasting accuracy even for longer SPI timescales characterised by greater temporal complexity. The Hybrid Wavelet-LSTM model produced robust forecasting results across all stations and timescales (SPI3 and SPI6), with correlation coefficients ranging from 0.939 to 0.995, R^2 values from 0.858 to 0.990, and generally low RMSE values. Detailed statistical measures and scatter plots for each station are provided in the Appendices. In Dar es Salaam (S2 & S3), the model successfully represented both short- and medium-term drought variability despite the high rainfall variability associated with coastal environments and the influence of large-scale climate drivers such as ENSO and the IOD. This suggests that wavelet decomposition effectively removed high-frequency noise while retaining important drought-related signals. In Dodoma, the model achieved particularly high accuracy for SPI3 (S4), which may be attributed to the pronounced temporal persistence of drought conditions in semi-arid regions. Although forecasting uncertainty increased slightly for SPI6 (S5), the model maintained satisfactory predictive skill. These findings are consistent with those of Hinge et al. (2022). Kigoma recorded some of the best forecasting performances, especially for SPI6 (S7), indicating strong persistence in medium-term drought behaviour over western Tanzania. The results further demonstrate the effectiveness of wavelet decomposition in isolating dominant climatic oscillations and enhancing LSTM learning. The model also performed exceptionally well in Mbeya (S8 & S9), where the relatively stable and highly seasonal rainfall regime likely contributed to improved drought predictability. Consequently, the Hybrid Wavelet-LSTM framework was able to accurately capture both short- and medium-term drought dynamics, supporting findings by Salimi et al. (2024). In Mwanza, prediction errors were comparatively higher (S10 & S11), possibly due to the complex hydroclimatic processes associated with lake-atmosphere interactions in the Lake Victoria basin (Nyamweya et al., 2016). Nevertheless, the model was still able to reproduce most drought episodes with high accuracy.

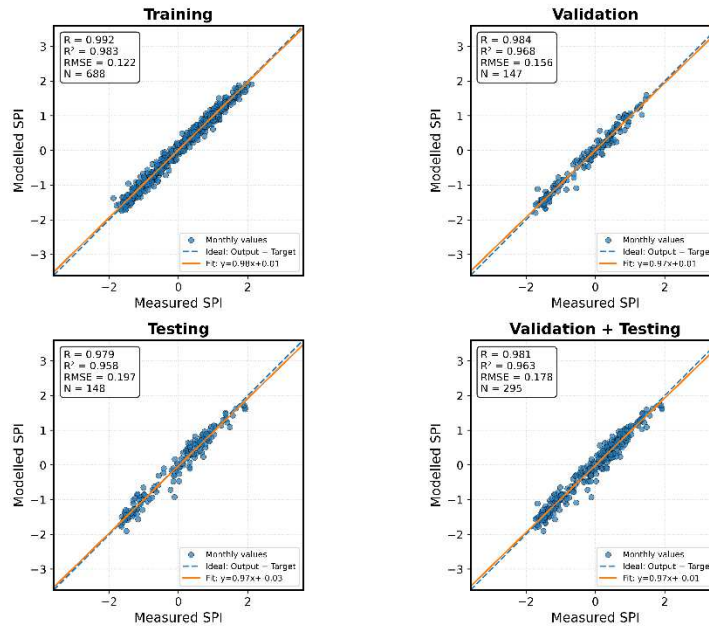


Fig.5 Sample statistic of the input combinations for forecasting of SPI3, showing the relationship between the output and target in Arusha region

Table 3 summarizes the performance of the proposed Hybrid Wavelet-LSTM model for drought forecasting across the six study regions using SPI3 and SPI6. Overall, the model demonstrated excellent predictive performance during the training, validation, and testing phases, with high correlation coefficients ($R = 0.94$ - 1.00), coefficients of determination ($R^2 = 0.86$ - 0.99), Nash–Sutcliffe efficiency ($NSE = 0.86$ - 0.99), and Willmott's index of agreement ($d = 0.96$ - 1.00), accompanied by consistently low MSE, RMSE, and MAE values. The slight reduction in performance from the training to testing datasets indicates good model generalization without substantial overfitting. Among the study regions, Mbeya and Kigoma exhibited the highest prediction accuracy, whereas Dodoma and Mwanza showed comparatively lower performance, particularly for SPI6, likely reflecting the greater rainfall variability and complexity of drought dynamics in semi-arid environments. These findings demonstrate that integrating wavelet decomposition with LSTM effectively captures both nonlinear and non-stationary characteristics of meteorological drought, consistent with previous studies reporting improved forecasting accuracy using hybrid wavelet–deep learning approaches (Charjan et al., 2026).

Table 3 Statistical details for Hybrid Wavelet_LSTM. The measurement includes (R), coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), Nash–Sutcliffe efficiency (NSE), and Willmott's index of agreement (d).

Region	SPI	Dataset partition	MSE	RMSE	MAE	NSE	R	R ²	d
Arusha	SPI3	Training	0.01	0.12	0.09	0.98	0.99	0.98	1
		Validation	0.02	0.16	0.12	0.97	0.98	0.97	0.99
		Testing	0.04	0.2	0.15	0.96	0.98	0.96	0.99
	SPI6	Training	0.02	0.13	0.11	0.98	0.99	0.98	0.99
		Validation	0.03	0.16	0.13	0.96	0.98	0.96	0.99
		Testing	0.06	0.25	0.18	0.93	0.97	0.93	0.98
Dar es Salaam	SPI3	Training	0.03	0.17	0.13	0.97	0.98	0.97	0.99
		Validation	0.05	0.22	0.16	0.94	0.97	0.94	0.99
		Testing	0.11	0.33	0.25	0.91	0.96	0.91	0.98
	SPI6	Training	0.03	0.19	0.14	0.96	0.98	0.96	0.99
		Validation	0.04	0.21	0.16	0.94	0.97	0.94	0.98
		Testing	0.12	0.35	0.26	0.91	0.96	0.91	0.98
Dodoma	SPI3	Training	0.02	0.13	0.1	0.98	0.99	0.98	1
		Validation	0.03	0.17	0.13	0.97	0.98	0.97	0.99
		Testing	0.07	0.27	0.2	0.93	0.97	0.93	0.98

Kigoma	SPI6	Training	0.04	0.19	0.15	0.96	0.98	0.96	0.99
		Validation	0.06	0.25	0.17	0.93	0.97	0.93	0.98
		Testing	0.15	0.38	0.28	0.86	0.94	0.86	0.96
	SPI3	Training	0.02	0.15	0.11	0.98	0.99	0.98	0.99
		Validation	0.03	0.18	0.13	0.97	0.99	0.97	0.99
		Testing	0.06	0.25	0.19	0.94	0.97	0.94	0.98
	SPI6	Training	0.01	0.1	0.08	0.99	1	0.99	1
		Validation	0.02	0.12	0.09	0.98	0.99	0.98	1
		Testing	0.04	0.19	0.13	0.96	0.98	0.96	0.99
Mbeya	SPI3	Training	0.01	0.1	0.07	0.99	1	0.99	1
		Validation	0.02	0.13	0.09	0.98	0.99	0.98	1
		Testing	0.04	0.19	0.14	0.97	0.99	0.97	0.99
	SPI6	Training	0.02	0.14	0.1	0.98	0.99	0.98	0.99
		Validation	0.04	0.19	0.12	0.96	0.98	0.96	0.99
		Testing	0.08	0.28	0.19	0.92	0.97	0.92	0.98
	SPI3	Training	0.03	0.16	0.12	0.97	0.99	0.97	0.99
		Validation	0.04	0.21	0.16	0.95	0.98	0.95	0.99
		Testing	0.1	0.32	0.23	0.88	0.94	0.88	0.97
Mwanza	SPI6	Training	0.02	0.15	0.12	0.97	0.99	0.97	0.99
		Validation	0.06	0.24	0.18	0.92	0.96	0.92	0.98
		Testing	0.05	0.23	0.18	0.93	0.97	0.93	0.98

461

462 The time series in Figure 6 reveal that the simulated SPI values closely follow the observed drought fluctuations throughout
463 the entire study period. For both SPI3 and SPI6, the model successfully captures the timing, magnitude, and persistence of
464 wet and dry episodes, as evidenced by the substantial overlap between the measured and modelled series. Although minor
465 discrepancies are visible during some extreme wet and drought events, the overall agreement remains remarkably high. The
466 close agreement between observed and modelled SPI values demonstrates that the integration of wavelet decomposition with
467 LSTM effectively captures both high-frequency and low-frequency rainfall variability. The slightly improved alignment
468 observed for SPI6 relative to SPI3 suggests that longer accumulation periods smooth short-term precipitation fluctuations,
469 thereby enhancing predictability. Longer SPI timescales generally exhibit stronger temporal persistence and lower stochastic
470 variability, making them easier for deep learning models to forecast (Vicente-Serrano et al., 2010). Similar findings have been
471 reported by Rezaei and Shabri (2025), who demonstrated that hybrid wavelet-based models substantially improved drought
472 prediction accuracy by decomposing precipitation signals into multiple frequency components. Likewise, Seifi et al. (2026)
473 observed that integrating wavelet decomposition with machine learning algorithms significantly enhanced forecasting
474 performance for long-term drought conditions. Comparable improvements using Wavelet-LSTM frameworks have also been
475 documented by Hao et al. (2024) and Alawsi et al. (2022), who reported that hybrid deep learning models outperform
476 standalone neural networks in modelling hydroclimatic extremes.

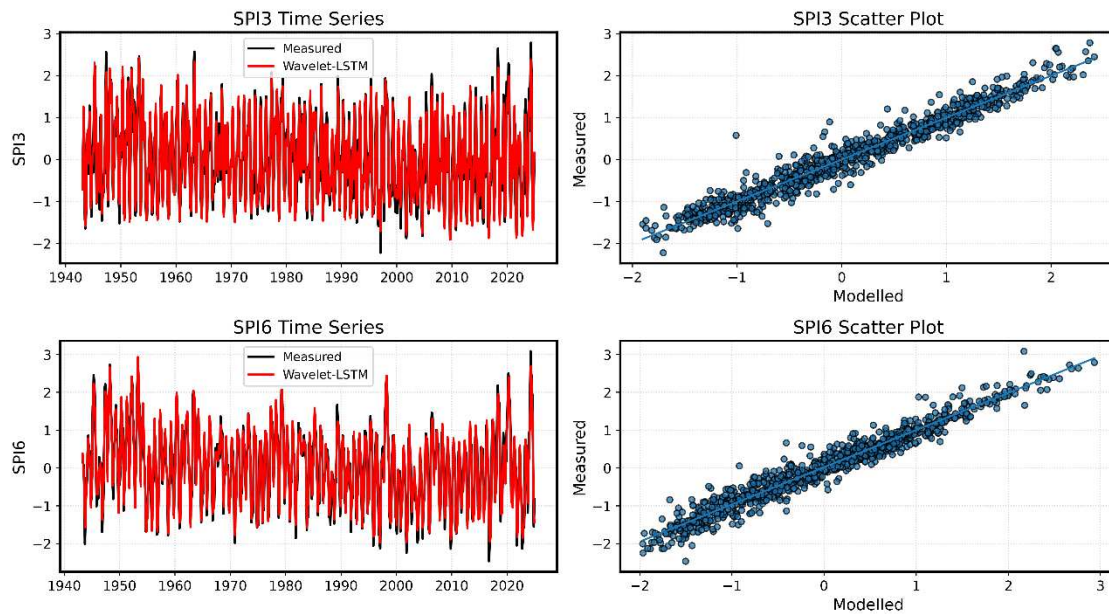
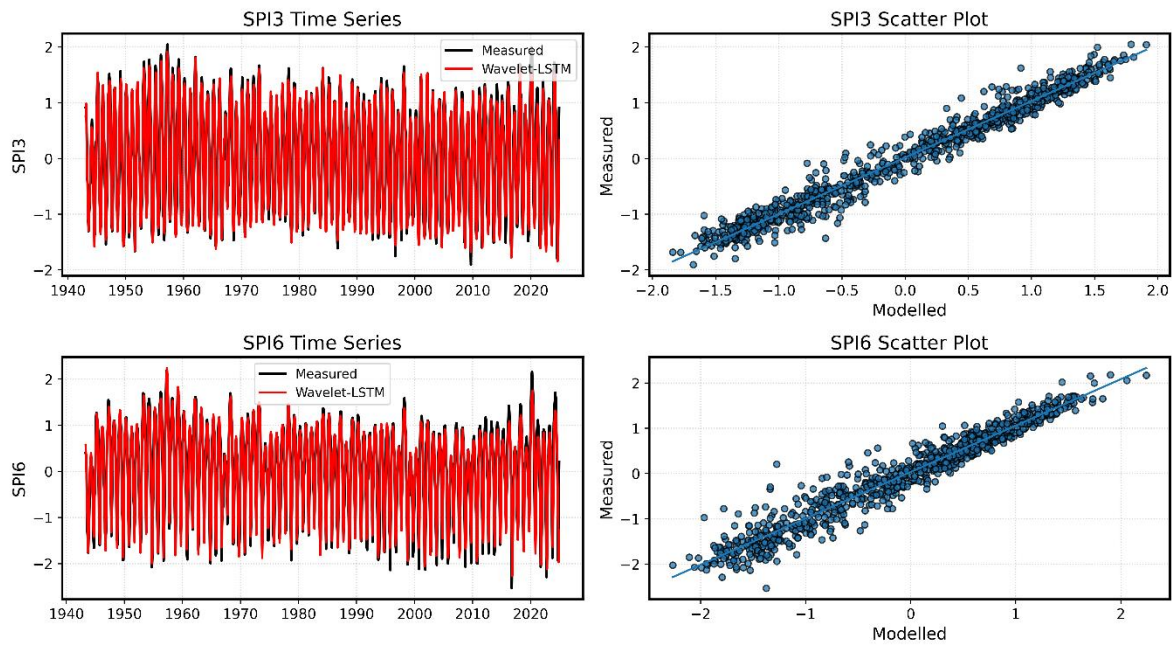


Fig. 6 Sample of model results for Dar es Salaam showing SPI3 and SPI6 time series alongside scatter plots for comparison

The modelled and observed time series in Figure 7 exhibit a strong correspondence across the entire study period, with the Wavelet-LSTM model successfully reproducing the recurrent oscillatory behaviour of drought conditions characteristic of the semi-arid central plateau. The model accurately captures both wet and dry anomalies, including the seasonal transitions and prolonged drought episodes. The scatter plots reveal a pronounced linear relationship between measured and modelled SPI values, with most observations concentrated near the regression line, indicating high model reliability and limited prediction error. The strong model performance observed in Dodoma can be attributed to the relatively persistent drought behaviour associated with semi-arid climates. In such regions, precipitation variability tends to exhibit pronounced temporal autocorrelation, allowing recurrent neural networks such as LSTM to effectively learn sequential dependencies from historical observations. Compared to SPI3, the slightly larger dispersion observed in the SPI6 scatter plot reflect the cumulative influence of rainfall variability over extended periods, which introduces additional uncertainty during exceptionally dry or wet conditions. Nevertheless, the high concentration of points around the regression line suggests that the model remains robust across drought timescales. Similar behaviour has been documented by Belayneh et al. (2014), who found that recurrent neural networks effectively model drought persistence in semi-arid regions.



493

494 **Fig. 7** Sample of model results for Dodoma showing SPI3 and SPI6 time series alongside scatter plots for comparison

495 Figure 8 illustrates the observed and simulated SPI3 and SPI6 series for Arusha together with corresponding scatter plots. The
 496 results indicate that the model successfully reproduces the temporal evolution of meteorological drought across both
 497 timescales. The modelled series closely overlap with the observed records, demonstrating the capability of the Wavelet-LSTM
 498 framework to capture seasonal variability, drought onset, recovery phases, and intermittent wet periods. The scatter plots
 499 exhibit a dense clustering of points along the 1:1 line for both SPI3 and SPI6, indicating strong agreement between measured
 500 and predicted SPI values and confirming the high predictive skill of the model. Arusha experiences highly variable rainfall
 501 influenced by complex topographic effects associated with Mount Meru, the East African Rift system, and interactions
 502 between large-scale atmospheric circulations such as the Intertropical Convergence Zone (ITCZ) and Indian Ocean variability.
 503 Despite this complexity, the Wavelet-LSTM model effectively captures drought variability, suggesting that the hybrid
 504 architecture successfully extracts dominant climatic signals while preserving long-term memory characteristics. The stronger
 505 concentration of points around the regression line in the SPI6 scatter plot indicates that drought conditions aggregated over
 506 longer timescales are more predictable than short-term anomalies. This observation is consistent with drought theory, which
 507 suggests that longer accumulation periods exhibit greater persistence and reduced random variability (Vicente-Serrano et al.,
 508 2010). Similar findings were reported by Tuğrul et al. (2025), who showed that hybrid Wavelet-LSTM models accurately
 509 forecast drought under heterogeneous climatic conditions. Likewise, Babrekar et al. (2026) demonstrated that integrating
 510 wavelet decomposition with deep learning substantially improves forecasting accuracy in regions characterized by strong
 511 topographic and climatic variability. Comparable improvements in drought prediction using hybrid machine learning models
 512 have also been reported by Alawsi et al. (2022).

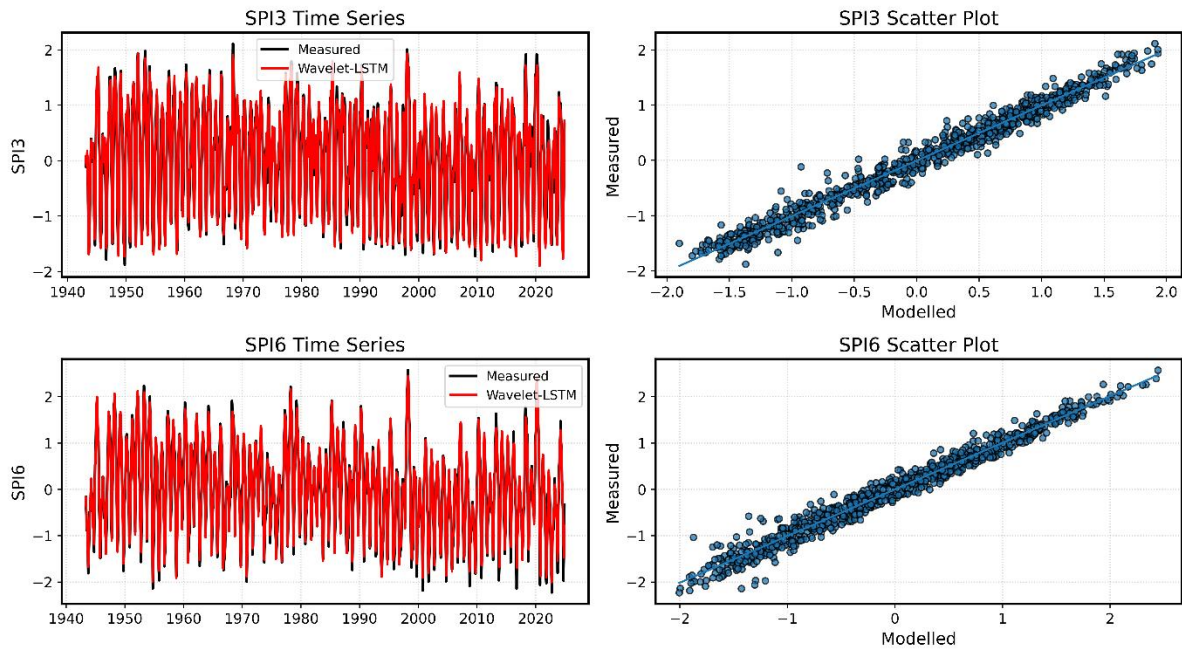
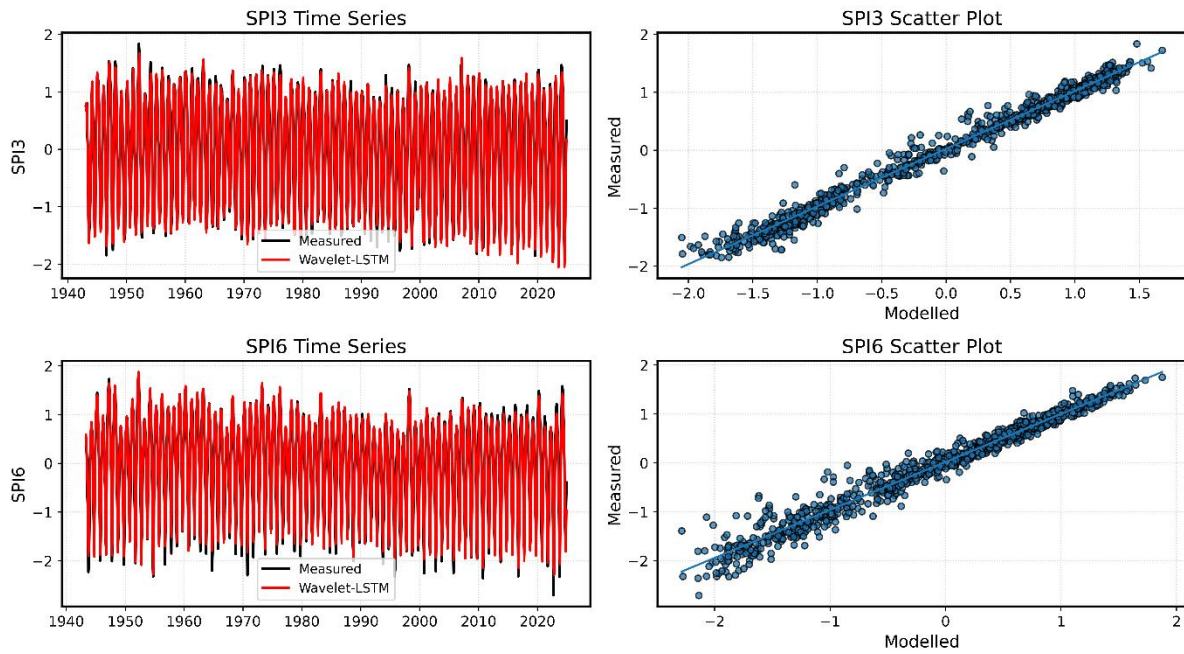


Fig. 8 Sample of model results for Arusha showing SPI3 and SPI6 time series alongside scatter plots for comparison

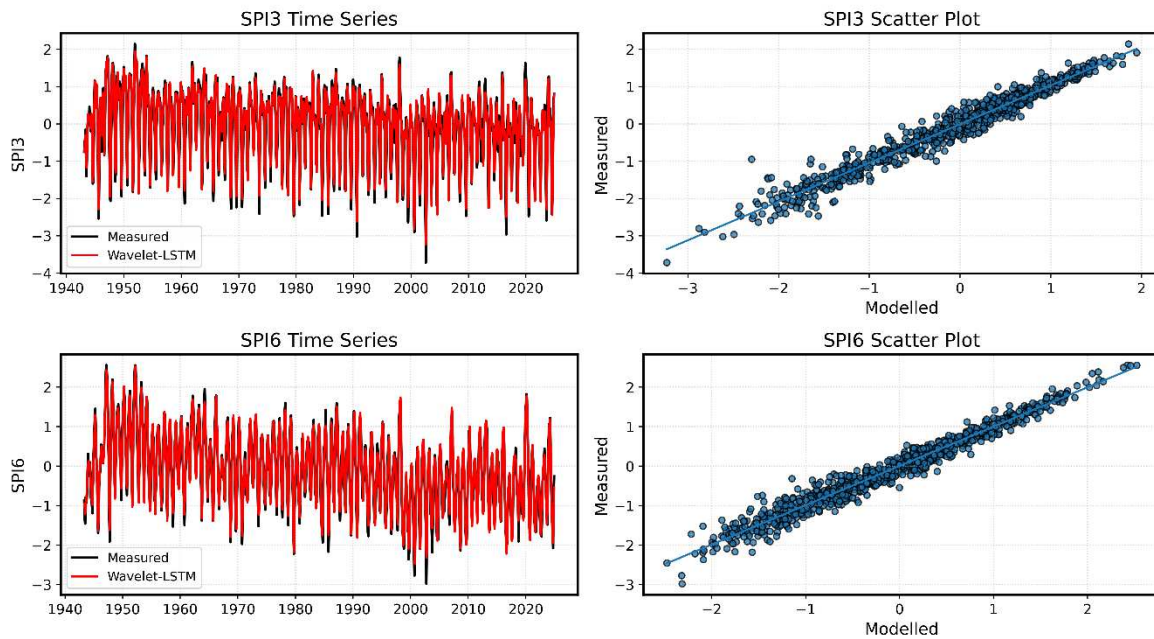
The results in Figure 9 show that the Hybrid Wavelet-LSTM model reproduces the temporal variability of meteorological drought with a high degree of accuracy. In both SPI3 and SPI6 time series, the modelled values closely track the observed drought fluctuations, capturing the timing and magnitude of wet and dry conditions throughout the study period. The overlap between the observed and simulated series indicates that the model effectively represents drought onset, duration, and recovery phases. The scatter plots further demonstrate this strong agreement, as the majority of points are densely clustered along the regression line with minimal dispersion, suggesting high predictive skill and low forecasting errors. The excellent performance observed in Mbeya can be attributed to the ability of the hybrid framework to capture the nonlinear and multiscale characteristics of precipitation variability. Mbeya is characterized by complex rainfall regimes influenced by topographic effects associated with the Southern Highlands and regional atmospheric circulation systems. Such hydroclimatic processes often exhibit non-stationary behaviour, making drought forecasting challenging using conventional approaches. Consequently, the combined Wavelet-LSTM framework enhances the representation of both short-term fluctuations and persistent drought signals. The SPI6 results exhibit slightly tighter clustering around the regression line than SPI3, indicating improved forecasting performance at longer accumulation timescales. This behaviour is expected because longer timescales smooth high-frequency rainfall variability and increase drought persistence, thereby enhancing predictability (Vicente-Serrano et al., 2010). Similar findings have been reported by Tuğrul et al. (2025), who demonstrated that Wavelet-LSTM models effectively forecast drought across climatically heterogeneous regions. Likewise, Rezaei and Shabri (2025) found that hybrid wavelet-based models significantly improve drought prediction accuracy by preserving both temporal persistence and multiscale variability. Comparable results were also documented by Hao et al. (2024), who reported superior performance of hybrid deep learning approaches for drought forecasting compared with standalone neural network models.



535

536 **Fig. 9** Sample of model results for Mbeya showing SPI3 and SPI6 time series alongside scatter plots for comparison

537 Figure 10 illustrates the observed and simulated SPI3 and SPI6 series for Mwanza using the Hybrid Wavelet-LSTM model.
 538 The results indicate that the model successfully reproduces the temporal evolution of drought conditions over the study period.
 539 Compared to the other regions, a slightly larger discrepancy between observed and simulated values is evident during some
 540 extreme drought events, especially in SPI3. This behaviour is reflected in the scatter plots, where a greater spread of points is
 541 observed around the regression line, although the overall linear relationship remains strong. The relatively larger variability
 542 observed in Mwanza may be associated with the strong climatic influence exerted by Lake Victoria. The lake significantly
 543 modifies local rainfall patterns through lake-land breeze circulations, enhanced convection, and mesoscale atmospheric
 544 interactions, resulting in highly variable and localized precipitation events. Such processes introduce additional uncertainty
 545 and nonlinearity into drought evolution, thereby increasing forecasting complexity. Despite these challenges, the close
 546 correspondence between observed and modelled SPI values demonstrates that the Wavelet-LSTM model effectively captures
 547 the dominant drought signals. The capability of the model to reproduce complex drought dynamics can be attributed to wavelet
 548 decomposition, which separates rainfall variability across different temporal scales, and LSTM networks, which retain
 549 important information from past observations through memory cells (Mienye et al., 2024). This finding is consistent with
 550 previous studies showing that longer SPI timescales exhibit greater temporal persistence and reduced random variability
 551 (Mishra & Singh, 2011). Similar observations have been reported by Alawsi et al. (2022), who found that hybrid deep learning
 552 models performed better for longer drought timescales. Likewise, Babrekar et al. (2026) reported that Wavelet-LSTM
 553 frameworks substantially improve forecasting skill in regions characterized by highly variable precipitation regimes.



554

555 **Fig. 10** Sample of model results for Mwanza showing SPI3 and SPI6 time series alongside scatter plots for comparison

556 The results in Figure 11 demonstrate excellent agreement between observed and modelled drought conditions across both
557 timescales in Kigoma region. The simulated series accurately reproduces the oscillatory drought patterns, including wet spells,
558 drought episodes, and transition periods. The overlap between measured and predicted values is particularly strong for SPI6,
559 where the observed and modelled series are almost indistinguishable for much of the study period. The scatter plots reveal a
560 very dense concentration of points along the regression line with minimal dispersion, indicating high model efficiency and
561 limited forecasting uncertainty. The outstanding performance observed in Kigoma may be linked to the relatively persistent
562 rainfall regime associated with western Tanzania. Rainfall in this region is strongly influenced by the migration of the
563 Intertropical Convergence Zone (ITCZ) and moisture transport from the Congo Basin, which often results in coherent and
564 persistent precipitation patterns. Such persistence enhances the predictability of drought conditions because future states
565 depend strongly on antecedent hydroclimatic conditions. LSTM networks are particularly effective in exploiting these
566 temporal dependencies owing to their long-memory structure, while wavelet decomposition improves the extraction of
567 dominant oscillatory modes embedded within the rainfall series. The stronger correspondence observed in SPI6 relative to
568 SPI3 further confirms that longer drought timescales are easier to forecast because they contain less high-frequency noise and
569 exhibit greater temporal continuity (Vicente-Serrano et al., 2010). Likewise, Hao et al. (2024) demonstrated that combining
570 wavelet decomposition with deep learning substantially enhances the prediction of both short-term and long-term drought
571 conditions. Comparable findings were also documented by Rezaei and Shabri (2025), who showed that multiscale hybrid
572 models outperform conventional machine learning approaches in capturing drought dynamics.

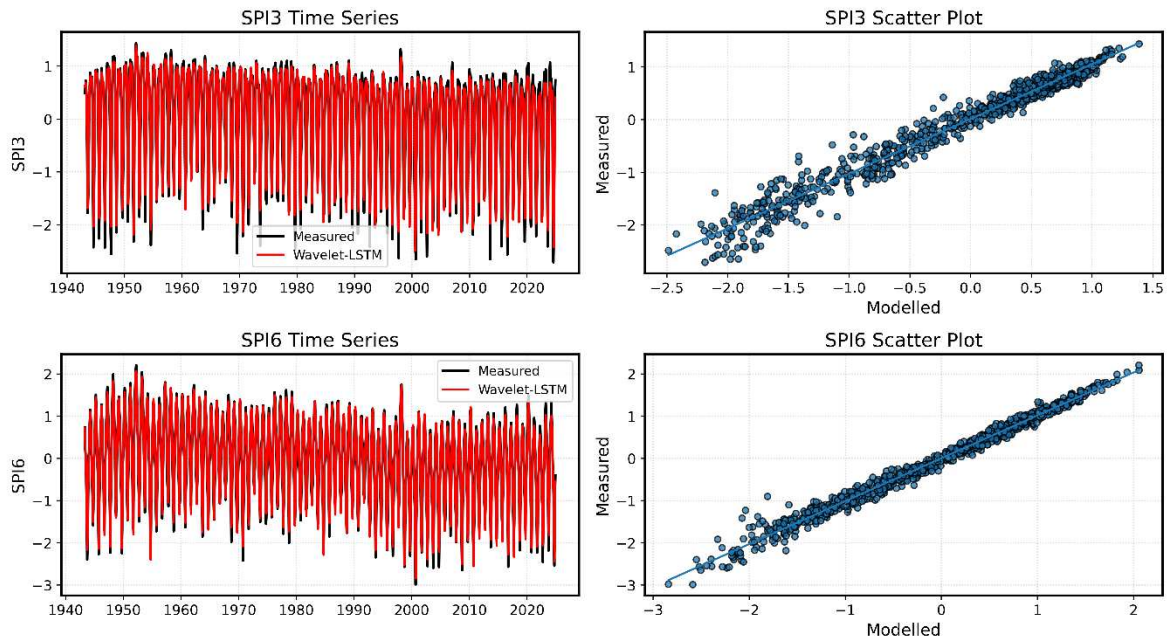
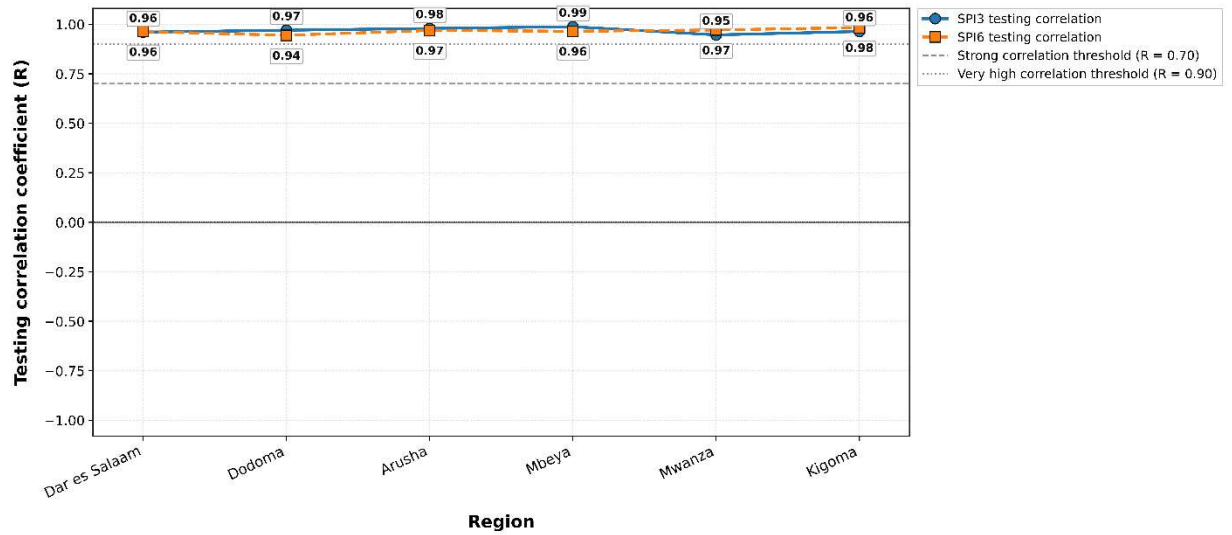


Fig. 11 Sample of model results for Kigoma showing SPI3 and SPI6 time series alongside scatter plots for comparison

Figure 12 illustrates the testing correlation coefficients (R) between observed and simulated SPI values produced by the Hybrid Wavelet-LSTM model across the six climatic regions of Tanzania. The results reveal exceptionally high correlations in all regions, with correlation coefficients ranging from approximately 0.945 to 0.976. Kigoma exhibited the highest predictive skill ($R = 0.976$), followed by Arusha ($R = 0.970$), Mbeya ($R = 0.965$), Mwanza ($R = 0.963$), Dar es Salaam ($R = 0.957$), and Dodoma ($R = 0.945$). Notably, all correlation values are well above the strong-skill threshold ($R = 0.70$) and exceed the very high skill threshold ($R = 0.90$), indicating excellent agreement between measured and simulated drought conditions across all climatic zones. The consistently high correlation coefficients suggest that the Hybrid Wavelet-LSTM framework successfully captures both short-term and long-term drought variability irrespective of regional climatic differences. This strong performance demonstrates the model's capability to reproduce the temporal evolution of drought by effectively learning nonlinear relationships and long-memory characteristics embedded within precipitation series. The superior performance observed in Kigoma and Arusha may be associated with relatively persistent rainfall regimes, which enhance drought predictability due to stronger temporal autocorrelation in hydroclimatic processes (Mishra & Singh, 2011). In contrast, the slightly lower correlations in Dodoma and Dar es Salaam may reflect the influence of higher rainfall variability and localized convective processes that introduce additional uncertainty into drought evolution. The high correlation values obtained in this study are consistent with previous drought forecasting studies employing hybrid deep learning approaches. Belayneh et al. (2014) reported correlation coefficients exceeding 0.90 when using Wavelet-Artificial Neural Network models for SPI forecasting. Similarly, Khan et al. (2018) and Hao et al. (2024) demonstrated that Wavelet-LSTM frameworks achieve superior predictive performance compared with standalone machine learning models. Comparable findings were also reported by Seifi et al. (2026), who observed consistently high correlations across different drought timescales using hybrid Wavelet-Kalman-LSTM models.



597

598 **Fig. 12** Correlation values of the Hybrid Wavelet-LSTM simulation results at the different regions

599 Figure 13 presents the ANOVA analysis comparing measured and modelled SPI3 values across the six study regions. The
600 boxplots in Figure 13(a) describe the distribution and median values of observed and simulated SPI3 datasets, whereas Figure
601 13(b) compares their respective means. The results indicate a remarkable similarity between observed and modelled SPI3
602 distributions in all regions. The medians of simulated SPI3 closely align with those of the observed data, as evidenced by the
603 near coincidence of the red median lines within each pair of boxplots. Moreover, the interquartile ranges and whisker lengths
604 of the simulated datasets closely resemble those of the measured datasets, suggesting that the model successfully reproduces
605 both central tendency and variability of drought conditions. Likewise, the comparable spread of the boxplots demonstrates
606 that the model adequately reproduces the variability and frequency of both wet and dry events. Although a few outliers are
607 present, particularly in Mwanza and Mbeya, these are associated with extreme drought episodes, which are inherently more
608 difficult to predict due to their rarity and nonlinear dynamics (Mishra & Singh, 2011). Figure 13(b) further confirms the
609 excellent model performance by showing that the means of simulated SPI3 values closely correspond to those of the observed
610 series across all regions. The mean values of both datasets lie close to the reference line representing the overall mean of
611 measured data, indicating minimal bias in model simulations. The slight differences observed in Mwanza likely reflect the
612 strong influence of Lake Victoria on local rainfall variability, which introduces additional uncertainty through mesoscale
613 convective processes and localized precipitation events. The findings agree with previous studies that reported significant
614 improvements in preserving drought statistics using hybrid wavelet-based models. Belayneh et al. (2014) found that Wavelet-
615 ANN models accurately reproduced SPI distributions and reduced forecasting bias. Similarly, Alawsi et al. (2022) observed
616 that hybrid wavelet-machine learning frameworks effectively maintain both the variability and statistical structure of drought
617 indices.

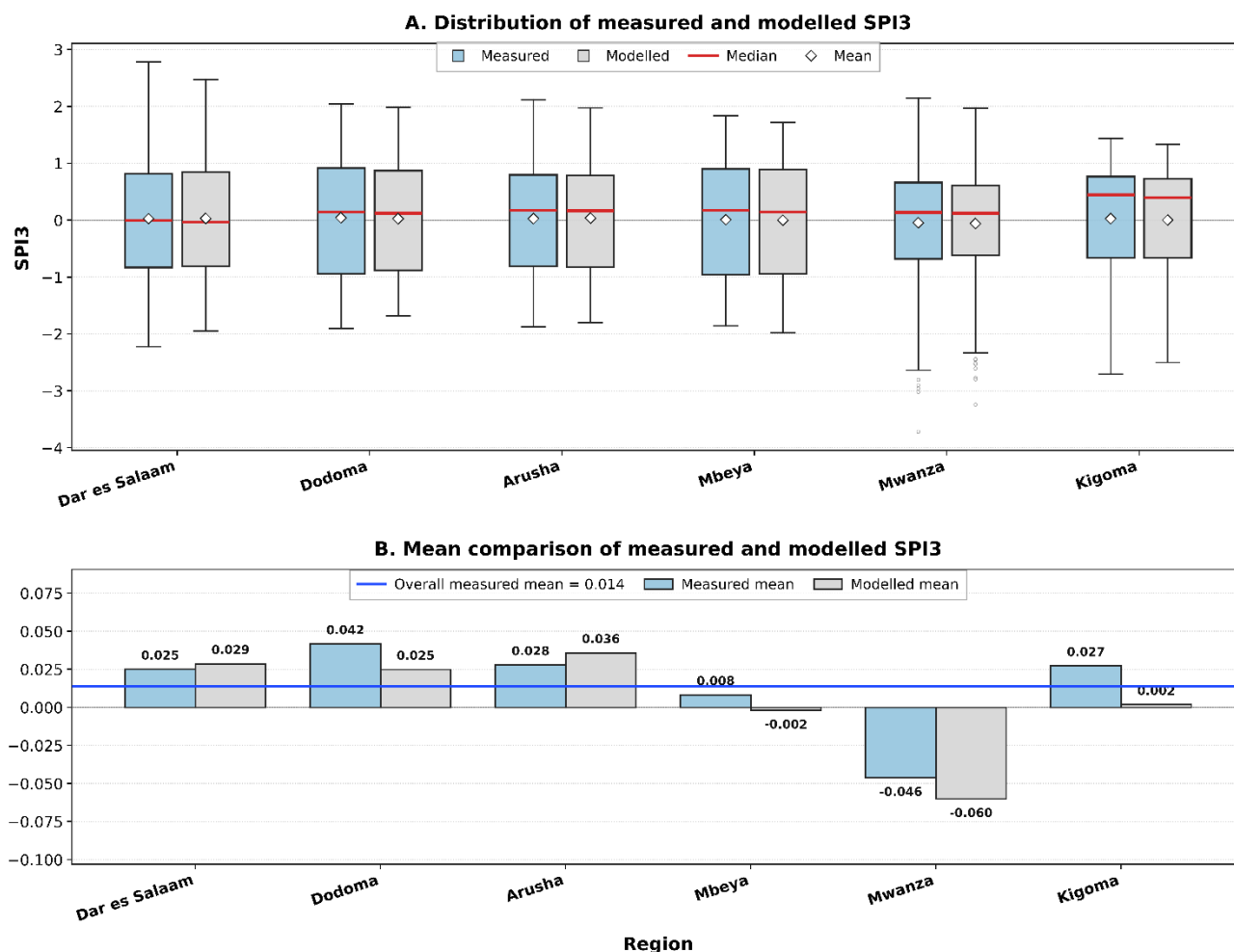


Fig. 13 ANOVA plots representing the measured and modelled SPI3

Similar to SPI3, the boxplots in Figure 14(a) reveal a close correspondence between measured and simulated SPI6 distributions. The medians of the simulated datasets almost coincide with those of the measured data across all regions, while the interquartile ranges and whisker extents exhibit very similar patterns. This close agreement indicates that the model successfully captures the statistical behaviour of medium-term drought conditions represented by SPI6. Compared with SPI3, the SPI6 distributions appear slightly more compact, with fewer extreme outliers. This observation is physically expected because longer accumulation timescales smooth short-term rainfall fluctuations, thereby reducing variability and emphasizing persistent drought conditions (Li et al., 2024). Consequently, SPI6 exhibits stronger temporal persistence and lower stochastic variability than SPI3, enhancing model predictability. The close match between observed and simulated SPI6 distributions therefore indicates that the model effectively reproduces the persistence characteristics inherent in medium-term drought processes. Figure 14(b) demonstrates that the means of simulated SPI6 values closely follow those of the observed datasets. In most regions, the differences between measured and modelled means are negligible, confirming the absence of significant systematic bias. Although Mwanza exhibits relatively lower mean values compared with other regions, the model successfully reproduces this regional characteristic, highlighting its ability to adapt to diverse climatic regimes. The preservation of regional drought characteristics is essential for operational drought forecasting because it ensures realistic representation of local hydroclimatic conditions. The results are consistent with previous studies demonstrating that hybrid deep learning models perform particularly well at longer drought timescales. Tuğrul et al. (2025) reported improved drought forecasting accuracy for SPI6 and SPI12 compared with shorter timescales. Likewise, Hao et al. (2019) found that Wavelet-LSTM models effectively preserve both the temporal and statistical characteristics of medium- and long-term drought indices. Similar conclusions were also reached by Babrekaret al. (2026), who reported that LSTM-based models exhibit superior performance in forecasting persistent drought conditions.

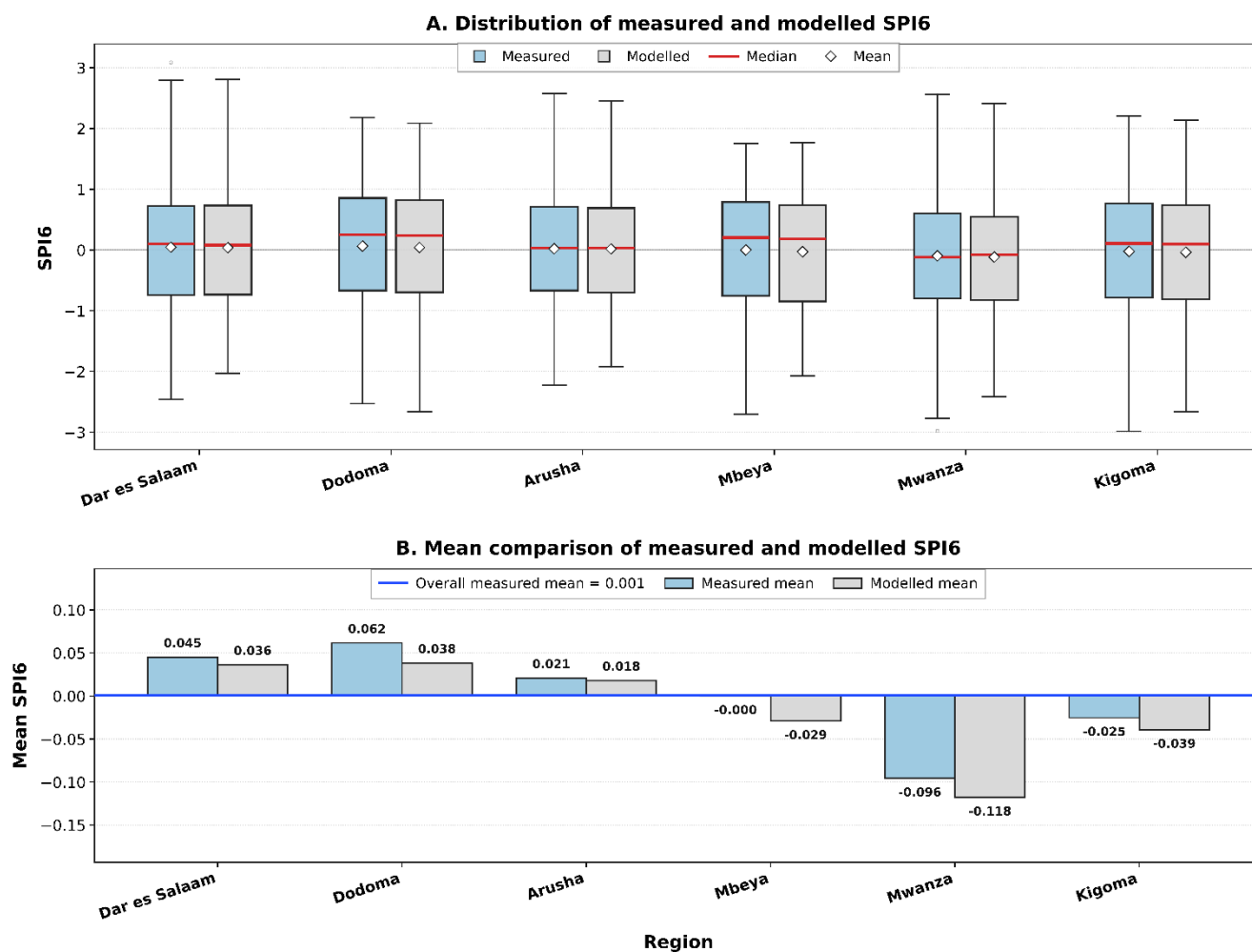


Fig. 14 ANOVA plots representing the measured and modelled SPI6

Table 4 presents the ANOVA results used to evaluate the statistical consistency of the Hybrid Wavelet-LSTM model across the six study regions for both SPI3 and SPI6. All p-values exceeded the 0.05 significance level, and the corresponding "Pass" decisions indicate that the null hypothesis (H_0) was not rejected. This implies that there were no statistically significant differences between the observed and predicted SPI values at the 95% confidence level, demonstrating that the model accurately reproduced the observed drought conditions. Furthermore, the relatively low F-statistics across all regions indicate minimal variation between observed and predicted values, confirming the model's robustness and absence of systematic prediction bias. The consistent "Pass" results across diverse climatic zones, including coastal, semi-arid, highland, and lake basin regions, highlight the strong generalization capability of the Hybrid Wavelet-LSTM model for meteorological drought forecasting. These findings are consistent with previous studies reporting that hybrid wavelet-deep learning models improve forecasting accuracy while maintaining statistical agreement between observed and predicted drought indices (Mokhtarzad et al., 2017; Hinge et al., 2022).

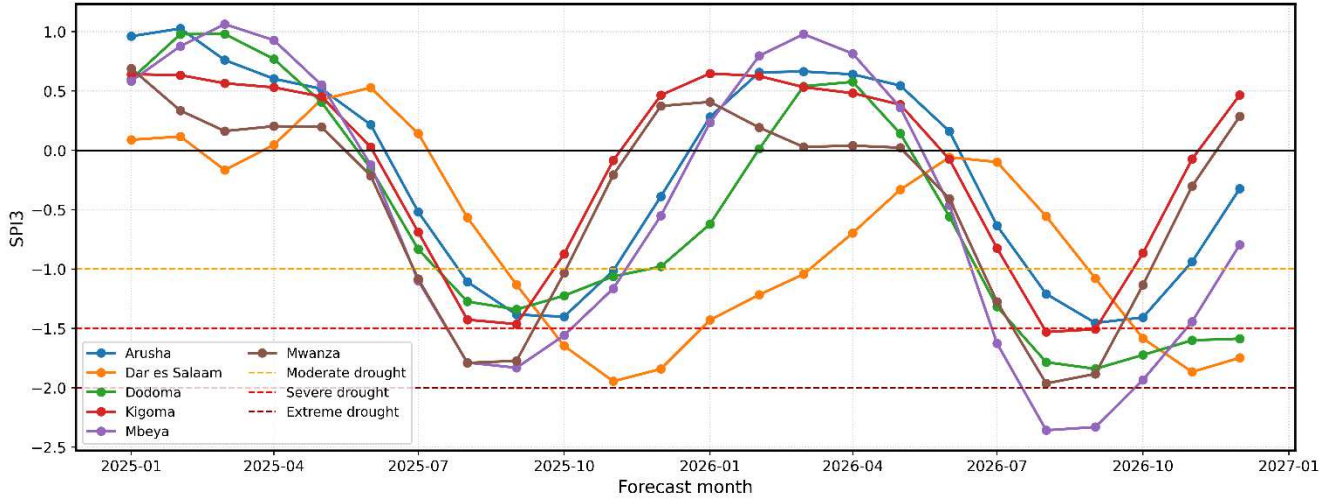
Table 4 ANOVA statistic. F, F-statistic, Prob > F, p-value at 0.05 percentage confidence level

Location	SPI	F-statistic	Prob > F	Decision
Arusha	SPI3	0.048	0.826	Pass
	SPI6	0.037	0.847	Pass
Dar es Salaam	SPI3	0.163	0.687	Pass
	SPI6	0.721	0.396	Pass
Dodoma	SPI3	0.494	0.483	Pass
	SPI6	1.893	0.170	Pass

Kigoma	SPI3	0.678	0.411	Pass
	SPI6	0.384	0.536	Pass
Mbeya	SPI3	0.475	0.491	Pass
	SPI6	1.579	0.210	Pass
Mwanza	SPI3	0.438	0.509	Pass
	SPI6	0.001	0.973	Pass

656

657 Figure 15 presents the 24-month forecast of SPI3 generated by the Hybrid Wavelet-LSTM model for six representative
658 climatic regions of Tanzania. The forecast indicates substantial temporal variability in short-term meteorological drought
659 conditions across all regions, characterized by alternating wet and dry phases. Most regions begin the forecast period under
660 near-normal to moderately wet conditions, with SPI3 values generally exceeding zero during the first half of 2025. Arusha,
661 Dodoma, Mbeya, Kigoma, and Mwanza exhibit positive SPI values ranging between approximately 0.5 and 1.1, suggesting
662 favourable rainfall conditions. However, a pronounced decline is observed during the second half of 2025, with SPI3 values
663 dropping below the moderate drought threshold ($SPI \leq -1.0$) in nearly all regions. The most severe drought conditions are
664 projected in Mbeya, Mwanza, and Dar es Salaam during late 2025 and again during the second half of 2026. The forecast
665 further reveals a cyclical drought pattern, with drought intensification occurring around September to November 2025 and
666 July to October 2026. Mbeya is projected to experience the most intense drought event, with SPI3 values falling below -2.0
667 during August-September 2026, indicating extreme drought conditions. Mwanza also exhibits severe drought episodes
668 approaching the extreme drought threshold, whereas Dar es Salaam displays relatively smoother transitions between wet and
669 dry conditions. Following these drought episodes, most regions show gradual recovery toward near-normal conditions by the
670 end of the forecast period, although Dodoma and Dar es Salaam remain under moderate drought conditions. The differences
671 observed among regions can also be attributed to local climatic controls. For example, the relatively persistent drought
672 projected in Dodoma is consistent with the semi-arid nature of central Tanzania, where rainfall is highly variable and
673 evapotranspiration rates are high (Shemsanga et al., 2016). Conversely, Arusha and Kigoma exhibit faster transitions between
674 drought and wet conditions due to the influence of orographic rainfall and proximity to major water bodies, which enhance
675 atmospheric moisture availability. The severe drought projected over Mwanza may partly result from fluctuations in Lake
676 Victoria-induced convection, which can amplify rainfall variability within the Lake Victoria basin. The cyclical drought
677 patterns projected in this study are consistent with previous studies that identified recurrent drought episodes in East Africa
678 associated with coupled ocean-atmosphere interactions (Lyon & DeWitt, 2012). Similar oscillatory SPI forecasts using hybrid
679 deep learning models have been reported by Hao et al. (2019), who demonstrated that Wavelet-LSTM frameworks effectively
680 capture seasonal drought cycles and long-term hydroclimatic variability.



681

682 **Fig. 15** Twenty-four-month SPI3 forecast for the selected regions using the Hybrid Wavelet-LSTM

683 Figure 16 illustrates the 24-month SPI6 forecasts generated by the Hybrid Wavelet-LSTM model for the selected Tanzanian
684 regions. Compared with SPI3, the SPI6 forecasts exhibit smoother temporal variations and greater persistence, reflecting the
685 longer accumulation period represented by SPI6. During the early months of the forecast period, most regions experience

moderately wet conditions, with SPI6 values exceeding 0.5 between March and June 2025. Arusha records the highest wetness intensity, reaching SPI6 values above 1.0. However, a gradual transition toward drought conditions occurs during the second half of 2025, culminating in widespread moderate-to-extreme drought conditions between October and December 2025. Furthermore, the forecast indicates that Kigoma, Mbeya, and Mwanza experience the most severe drought conditions during late 2025, with SPI6 values falling below -2.0, corresponding to extreme drought conditions. A second drought cycle emerges during the latter half of 2026, particularly affecting Arusha, Kigoma, and Mbeya. Notably, Kigoma and Mbeya again register SPI6 values below -2.0 during October-November 2026, indicating prolonged and persistent drought. In contrast, Dar es Salaam maintains comparatively milder drought conditions throughout the forecast period, rarely exceeding the severe drought threshold. Regional differences in drought severity observed in Figure 16 are also consistent with the climatic characteristics of Tanzania. Semi-arid regions such as Dodoma exhibit prolonged moisture deficits due to inherently low rainfall and high evaporative demand, whereas lake and highland regions, including Kigoma and Mbeya, display stronger drought persistence associated with complex topographic and mesoscale climatic influences. The relatively moderate drought intensity projected over Dar es Salaam likely reflects the moderating influence of the Indian Ocean, which supplies additional atmospheric moisture to the coastal zone (King’uza et al., 2025). The forecast behaviour shown in Figure 16 agrees with previous drought forecasting studies that reported improved predictability and stronger persistence at longer SPI timescales. Khan et al. (2018) found that hybrid Wavelet-LSTM models produced more stable and accurate forecasts for SPI6 and SPI12 than for SPI3. Similar findings were reported by Babrekar et al. (2026), who demonstrated that long-timescale drought indices exhibit stronger temporal memory and are therefore more amenable to deep learning-based forecasting.

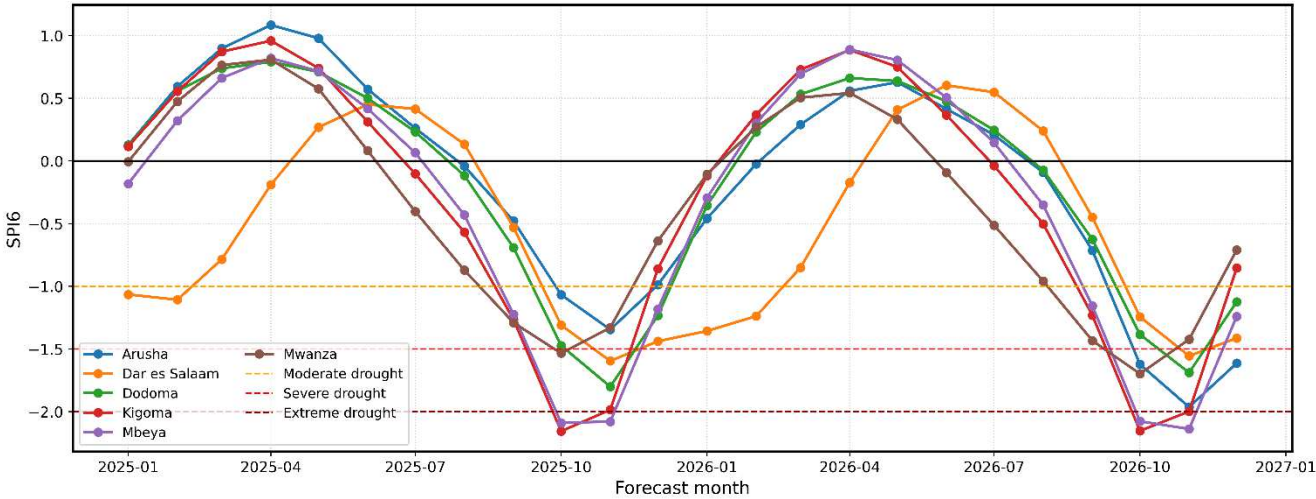


Fig.16 Twenty-four-month SPI6 forecast for the selected regions using the Hybrid Wavelet-LSTM

Conclusion

This study investigated the applicability of a Hybrid Wavelet-LSTM model for long-term meteorological drought forecasting in Tanzania using the SPI at 3-month (SPI3) and 6-month (SPI6) time scales. The model was evaluated across six climatologically distinct regions, namely Arusha, Dar es Salaam, Dodoma, Kigoma, Mbeya, and Mwanza, representing diverse rainfall regimes and hydroclimatic conditions. The conclusion of this study is highlighted as follows:

- i. The Hybrid Wavelet-LSTM model effectively captured both short-term and medium-term drought variability across all study regions.
- ii. The strong agreement between observed and simulated SPI series, supported by the close clustering in the scatter plots, demonstrates the model's ability to accurately reproduce historical drought patterns. Correlation coefficients exceeded 0.94 across all regions, ranging from 0.945 to 0.976, highlighting the excellent predictive performance of the proposed framework.
- iii. The highest model performance was achieved in Kigoma, while slightly lower, yet still robust, performance was observed in Dodoma.
- iv. The ANOVA analyses further revealed no substantial differences between the statistical distributions of observed and modelled SPI values, suggesting that the model preserved both central tendencies and variability of drought conditions.
- v. The twenty-four-month forecasts revealed considerable spatial and temporal heterogeneity in drought evolution across Tanzania. Several regions, particularly Mbeya, Dodoma, Mwanza, and Kigoma, exhibited episodes of

727 moderate to severe drought during the forecast period, while Dar es Salaam showed relatively less severe drought
728 conditions.
729 vi. The results confirm that integrating wavelet decomposition with LSTM networks substantially enhances drought
730 forecasting performance by effectively decomposing non-stationary hydroclimatic signals and improving the
731 learning of temporal dependencies.
732

733 Despite the encouraging results, this study has some limitations. The forecasting framework relied exclusively on
734 precipitation-derived SPI and historical rainfall information, without explicitly incorporating large-scale climatic and oceanic
735 drivers known to influence rainfall variability over Tanzania and East Africa. Consequently, some aspects of drought
736 variability associated with climate teleconnections may not have been fully represented. Future studies should therefore extend
737 the proposed framework by incorporating relevant climatic drivers such as the ENSO, IOD, Madden–Julian Oscillation
738 (MJO), sea surface temperature anomalies, and other atmospheric circulation indices. Integrating these large-scale climate
739 predictors into hybrid deep learning architectures may further improve forecast accuracy, particularly for longer lead times
740 and extreme drought events.
741

742 **Acknowledgements**

743 The authors acknowledge ERA5 global reanalysis dataset through European Centre for Medium-Range Weather Forecasts
744 (ECMWF) for providing a high-resolution gridded precipitation dataset necessary resources for this work.

745 **Funding**

746 The authors did not receive support from any organization for the submitted work.

747 **Contributions**

748 The study conception and design, material preparation and data collection were done by P.T.S.L. Data analyses were
749 performed by A.J.M. The first draft of the manuscript was written by P.T.S.L. The manuscript was reviewed and edited by
750 P.T.S.L and A.J.M. All authors read and approved the final manuscript.

751 **Declarations**

752 All authors have read, understood, and have complied as applicable with the statement on “Ethical responsibilities of Authors”
753 as found in the Instructions for Authors”

754 **Ethical approval**

755 Not applicable.

756 **Data Availability**

757 The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable
758 request.

759 **Conflict of interest**

760 The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest
761 in the subject matter or materials discussed in this manuscript.

762 **References**

763

764 Addison, P.S. (2017). *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science,*
765 *Engineering, Medicine and Finance.* CRC Press, Boca Raton.

766 AghaKouchak, A. (2015). A multivariate approach for persistence-based drought prediction: Application to the 2010–2011
767 East Africa drought. *Journal of Hydrology*, 526, 127–135. <https://doi.org/10.1016/j.jhydrol.2014.09.063>

768 Alawsi, M. A., Zubaidi, S. L., Al-Bdairi, N. S. S., Al-Ansari, N., & Hashim, K. (2022). Drought Forecasting: A Review and
769 Assessment of the Hybrid Techniques and Data Pre-Processing. *Hydrology*, 9(7), 115.
770 <https://doi.org/10.3390/hydrology9070115>

771 Babrekar, V. R., Gaikwad, S. V., Shah, S., Vibhute, A., & Saini, J. R. (2026). Meteorological drought prediction using long
772 short-term memory (LSTM) model. *Procedia Computer Science*, 280, 495–502. <https://doi.org/10.1016/j.procs.2026.04.063>

773 Belayneh, A., & Adamowski, J. (2013). Drought forecasting using new machine learning methods. *Journal of Water and*
774 *Land Development*, 18, 3–12. <https://doi.org/10.2478/jwld-2013-0001>

775 Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River
776 Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, 508,
777 418–429. <https://doi.org/10.1016/j.jhydrol.2013.10.052>

778 Bouaziz, M., Abid, M. A., Medhioub, E., & John, A. (2025). A Century of Data: Machine Learning Approaches to Drought
779 Prediction and Trend Analysis in Arid Regions. *Water*, 17(24), 3567. <https://doi.org/10.3390/w17243567>

780 Charjan, O., Gajbhiye, K., Warhade, J., & Wankhade, S. (2026). A novel hybrid approach to drought forecasting: Leveraging
781 feature engineering and ensemble methods. *Scientific Reports*, 16(1), 7972. <https://doi.org/10.1038/s41598-026-37206-6>

782 Salimi, A., Ghobrial, T., Samakosh, J. M., & Bonakdari, H. (2024). Drought forecasting based on machine learning
783 techniques. In *Intelligence Systems for Earth, Environmental and Planetary Sciences* (pp. 113–145). Elsevier.
784 <https://doi.org/10.1016/B978-0-443-13293-3.00009-9>

785 Dehghani, F., & Molavi-Arabshahi, M. (2026). Classifying drought severity in Northern Iran using machine learning and
786 integrated climate indices. *International Journal of Climatology*, 46(2), e70188. <https://doi.org/10.1002/joc.70188>.

787 Duc, L., & Sawada, Y. (2023). A signal-processing-based interpretation of the Nash–Sutcliffe efficiency. *Hydrology and*
788 *Earth System Sciences*, 27(9), 1827–1839. <https://doi.org/10.5194/hess-27-1827-2023>

789 Funk, C. (2011). We thought trouble was coming. *Nature*, 476(7358), 7. <https://doi.org/10.1038/476007a>

790 Gebrechorkos, S. H., Hülsmann, S., & Bernhofer, C. (2020). Analysis of climate variability and droughts in East Africa using
791 high-resolution climate data products. *Global and Planetary Change*, 186, 103130.
792 <https://doi.org/10.1016/j.gloplacha.2020.103130>

793 Ghosh, A., Baidya, S., & Gupta, A. K. (2024). Climate change impact on land use and livelihood in Sundarbans: A case study
794 of Sagar Island. In A. K. Gupta, A. Gupta, & P. Acharya (Eds.), *Disaster resilience and green growth* (pp. 403–434). Springer
795 *Nature*. https://doi.org/10.1007/978-981-99-4105-6_20

796 Gleixner, S., Demissie, T., & Diro, G. T. (2020). Did ERA5 Improve Temperature and Precipitation Reanalysis over East
797 Africa? *Atmosphere*, 11(9), 996. <https://doi.org/10.3390/atmos11090996>

798 Hao, W., Cominola, A., & Castelletti, A. (2024). Combining wavelet-enhanced feature selection and deep learning techniques
799 for multi-step forecasting of urban water demand. *Environmental Research: Infrastructure and Sustainability*, 4(3), 035005.
800 <https://doi.org/10.1088/2634-4505/ad5e1d>

801 Hayes, M., Svoboda, M., Wall, N., & Widhalm, M. (2011). The Lincoln Declaration on Drought Indices: Universal
802 Meteorological Drought Index Recommended. *Bulletin of the American Meteorological Society*, 92, 485-488.
803 <https://doi.org/10.1175/2010bams3103.1>

804 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R.,
805 Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita,
806 M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A.,
807 Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G.,
808 de Rosnay, P., Rozum, I.,

809 Hinge, G., Piplodiya, J., Sharma, A., Hamouda, M. A., & Mohamed, M. M. (2022). Evaluation of Hybrid Wavelet Models
810 for Regional Drought Forecasting. *Remote Sensing*, 14(24), 6381. <https://doi.org/10.3390/rs14246381>

811 Vamborg, F., Villaume, S., & Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal*
812 *Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>

813 Humphries, U. W., Waqas, M., Hlaing, P. T., Dechpichai, P., & Wangwongchai, A. (2024). A deep learning perspective on
814 meteorological droughts prediction in the Mun River Basin, Thailand. *AIP Advances*, 14, 085026.
815 <https://doi.org/10.1063/5.0209709>

816 IPCC (2023) Climate Change 2023: Synthesis Report. Intergovernmental Panel on Climate Change, Geneva.
817 <https://doi.org/10.59327/IPCC/AR6-9789291691647>

818 Li, Y., Huang, Y., Li, Y., Zhang, H., Fan, J., Deng, Q., & Wang, X. (2024). Spatiotemporal heterogeneity in meteorological
819 and hydrological drought patterns and propagations influenced by climatic variability, LULC change, and human
820 regulations. *Scientific Reports*, 14(1), 5965. <https://doi.org/10.1038/s41598-024-56526-z>

821 Kalisa, W., Igbawua, T., Ujoh, F., Aondoakaa, S. I., Namugize, J. N., & Zhang, J.-H. (2021). Spatio-temporal variability of
822 dry and wet conditions over East Africa from 1982 to 2015 using quantile regression model. *Natural Hazards*, 106(3), 1–30.
823 <https://doi.org/10.1007/s11069-021-04530-1>

824 Khan, M. M. H., Muhammad, N. S., & El-Shafie, A. (2018). Wavelet-ANN versus ANN-Based Model for
825 Hydrometeorological Drought Forecasting. *Water*, 10(8), 998. <https://doi.org/10.3390/w10080998>

826 King'uza, P. H., Zhou, B., & Limbu, P.T.S. (2025). Understanding Possible Physical Mechanisms Associated With Variability
827 of March to May Seasonal Extreme Wettest Days Rainfall in Tanzania. *International Journal of Climatology*, 45, 10: e8887.
828 <https://doi.org/10.1002/joc.8887>.

829 Krichen, M., & Mihoub, A. (2025). Long Short-Term Memory Networks: A Comprehensive Survey. *AI*, 6(9), 215.
830 <https://doi.org/10.3390/ai6090215>

831 Lamichhane, M., Jha, A. N., Ankon, S. B., Suwal, N., Khatakho, R., & Ansari, A. (2026). Assessment and forecasting of
832 drought in Nepal using long short-term memory model. *Theoretical and Applied Climatology*, 157, 374.
833 <https://doi.org/10.1007/s00704-026-06298-5>.

834 Limbu, P. T. S., & Guirong, T. (2019). Influence of the Tropical Atlantic Ocean and Its Walker Circulation Cell on October-
835 December Rainfall Variability over Tanzania. *International Journal of Climatology*, 40, 5767-5782.
836 <https://doi.org/10.1002/joc.6550>.

837 Limbu, P.T.S., King'uza, P.H., & Okrah, A. (2026) Characteristics of the observed diurnal cycle of precipitation during long-
838 rain season over Tanzania. *Theoretical and Applied Climatology*, 157, 301. <https://doi.org/10.1007/s00704-026-06229-4>

839 Lyon, B., & DeWitt, D. G. (2012). A recent and abrupt decline in the East African long rains. *Geophysical Research Letters*,
840 39(2), L02702. <https://doi.org/10.1029/2011GL050337>.

841 Makula, E. K., Mangara, R. J., Kazimili, B., Mbigi, D., Mtewe, Z. F., Kebacho, L. L., Kessy, W., & Limbu, P. T. S. (2024).
842 Assessment of drought characteristics using SPEI and VHI in Tanzania and their associated climate factors. *Natural Hazards*,
843 121(2), 2071–2093. <https://doi.org/10.1007/s11069-024-06894-6>

844 Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures,
845 variants, and applications. *Information*, 15(9), 517. <https://doi.org/10.3390/info15090517>

846 Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, 391(1–2), 202–216.
847 <https://doi.org/10.1016/j.jhydrol.2010.07.012>

848 Mishra, A. K., & Singh, V. P. (2011). Drought modeling—A review. *Journal of Hydrology*, 403(1–2), 157–175.
849 <https://doi.org/10.1016/j.jhydrol.2011.03.049>

850 Mokhtarzad, M., Eskandari, F., Jamshidi Vanjani, N., & Arabasadi, A. (2017). Drought forecasting by ANN, ANFIS, and
851 SVM and comparison of the models. *Environmental Earth Sciences*, 76(21), Article 729. [https://doi.org/10.1007/s12665-017-](https://doi.org/10.1007/s12665-017-7064-0)
852 7064-0

853 Mosavi, A., Ozturk, P., & Chau, K.-W. (2018). Flood prediction using machine learning models: Literature review. *Water*,
854 10(11), 1536. <https://doi.org/10.3390/w10111536>

855 Mwangi, E., Wetterhall, F., Dutra, E., Di Giuseppe, F., & Pappenberger, F. (2014). Forecasting droughts in East Africa.
856 *Hydrology and Earth System Sciences*, 18, 611–620. <https://doi.org/10.5194/hess-18-611-2014>

857 Mwinuka, P., Uiso, C. B. S., Chang’a, L. B., & Kumwenda, M. J. (2021). Atmospheric instability conditions during rainy
858 seasons over Tanzania. *Tanzania Journal of Science*, 47(5), 1647–1659. <https://doi.org/10.4314/tjs.v47i5.14>

859 Näschen, K., Diekkrüger, B., Leemhuis, C., Steinbach, S., Seregina, L. S., Thonfeld, F., & Van der Linden, R. (2018).
860 Hydrological modeling in data-scarce catchments: The Kilombero Floodplain in Tanzania. *Water*, 10(5), 599.
861 <https://doi.org/10.3390/w10050599>.

862 Nicholson, S. E. (2017). Climate and climatic variability of rainfall over eastern Africa. *Reviews of Geophysics*, 55(3), 590–
863 635. <https://doi.org/10.1002/2016RG000544>

864 Nourani, V., Alami, M. T., & Daneshvar Vousoughi, F. (2015). Wavelet-entropy data pre-processing approach for ANN-
865 based groundwater level modeling. *Journal of Hydrology*, 524, 255–269. <https://doi.org/10.1016/j.jhydrol.2015.02.048>.

866 Nyamweya, C., Desjardins, C., Sigurdsson, S., Tomasson, T., Taabu-Munyaho, A., Sitoki, L., & Stefansson, G. (2016).
867 Simulation of Lake Victoria circulation patterns using the Regional Ocean Modeling System (ROMS). *PLoS ONE*, 11(3),
868 e0151272. <https://doi.org/10.1371/journal.pone.0151272>

869 Onyeuwaoma, N., Sivakumar, V., & Bade, M. (2024). Modelling drought in South Africa: Meteorological insights and
870 predictive parameters. *Environmental Monitoring and Assessment*, 196, 965. <https://doi.org/10.1007/s10661-024-13009-y>

871 Piri, J., Abdollahipour, M., & Keshtegar, B. (2023). Advanced machine learning model for prediction of drought indices using
872 hybrid SVR-RSM. *Water Resources Management*, 37(2), 683–712. <https://doi.org/10.1007/s11269-022-03395-8>

873 Rezaei, R., & Shabri, A. (2025). Integrating wavelet transform and support vector machine for improved drought forecasting
874 based on standardized precipitation index. *Journal of Hydroinformatics*, 27(2), 320–337.
875 <https://doi.org/10.2166/hydro.2025.292>

876 Rohli, R. V., Ates, S. A., Rivera-Monroy, V. H., Polito, M. J., Midway, S. R., Castañeda-Moya, E., ... & Suwa, M. (2019).
877 Inter-annual hydroclimatic variability in coastal Tanzania. *International Journal of Climatology*, 39(12), 4736–4750.
878 <https://doi.org/10.1002/joc.6103>

879 Rwambo, I., Fan, Y., Yu, P., Chu, C., Nyasulu, M., & King'uza, P. (2025). Interannual variability of short rains in Tanzania
880 and the influences from ENSO and the Indian Ocean Dipole. *Atmospheric and Oceanic Science Letters*, 100614.
881 <https://doi.org/10.1016/j.aosl.2025.100614>

882 Seifi, A., Ali, M., Madvar, H. R., Ali, A., & Khan, M. (2026). Wavelet-Kalman-LSTM fusion: Enhancing meteorological
883 drought prediction with robust feature selection and uncertainty quantification. *Journal of Hydrology*, 677(Part B), 135865.
884 <https://doi.org/10.1016/j.jhydrol.2026.135865>

885 Shafik, W. (2026). A Review of the Machine Learning and Deep Learning Approach for Drought Monitoring. In: Vyas, S.,
886 Jawak, S.D., Deb Burman, P.K., Patel, H., Kandekar, A., Sawant, S. (eds) *Application of Machine Learning in Earth Sciences.*
887 *Earth and Environmental Sciences Library*. Springer, Cham. https://doi.org/10.1007/978-3-032-11426-6_22

888 Shemsanga, C., Muzuka, A. N. N., Martz, L., Komakech, H., & Omambia, A. N. (2016). Statistics in climate variability, dry
889 spells, and implications for local livelihoods in semiarid regions of Tanzania: The way forward. *Handbook of climate change*
890 *mitigation and adaptation*. Springer Science and Business Media, New York. https://doi.org/10.1007/978-3-319-14409-2_66

891 Soomro, S. E. H., Boota, M. W., Guo, J., Li, Y., Qin, H., Ahmad Meahrayen, M., ... & Chen, J. (2025). Drought management
892 for long-term water sustainability and resilience. *Marine and Freshwater Research*, 76(5), MF24210.
893 <https://doi.org/10.1071/MF24210>

894 Spinoni, J., Naumann, G., Carrao, H., Barbosa, P., & Vogt, J. (2014). World drought frequency, duration, and severity for
895 1951–2010. *International Journal of Climatology*, 34(8), 2792–2804. <https://doi.org/10.1002/joc.3875>.

896 Svoboda, M., Hayes, M., & Wood, D. (2012). World Meteorological Organization, 2012: Standardized Precipitation Index
897 User Guide (M. Svoboda, M. Hayes and D. Wood). (WMO-No. 1090), Geneva.

898 Tuğrul, T., Oruç, S., Hall, J. L., Şenocak, A. U. G., & Hınıs, M. A. (2025). Hybrid wavelet-ML models for regional drought
899 forecasting in Norway. *Scientific Reports*, 15(1), 38573. <https://doi.org/10.1038/s41598-025-22416-1>

900 Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscale drought index sensitive to global warming:
901 The standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7), 1696–1718.
902 <https://doi.org/10.1175/2009JCLI2909.1>

903 Wainwright, C. M., Marsham, J. H., Keane, R. J., Rowell, D. P., Finney, D. L., Black, E., et al. (2019). ‘Eastern African
904 paradox’ rainfall decline due to shorter not less intense long rains. *npj Climate and Atmospheric Science*, 2, Article 34.
905 <https://doi.org/10.1038/s41612-019-0091-7>.

906 Waqas, M., Humphries, U. W., Hlaing, P. T., & Ahmad, S. (2024). Seasonal WaveNet-LSTM: A deep learning framework
907 for precipitation forecasting with integrated large-scale climate drivers. *Water*, 16(22), 3194.
908 <https://doi.org/10.3390/w16223194>

909 World Meteorological Organization (WMO) (2012) Standardized Precipitation Index: User Guide (WMO-No. 1090). WMO.

910 Yang, L., He, W., Zou, X., Wang, J., Xia, H., Jahangir, M. S., ... & Feng, Q. (2025). A probabilistic approach to enhance
911 drought modelling in alpine regions: Transition from meteorological to hydrological droughts. *Journal of Environmental*
912 *Management*, 390, 126297. <https://doi.org/10.1016/j.jenvman.2025.126297>

913 Yang, W., Seager, R., Cane, M. A., & Lyon, B. (2014). The East African long rains in observations and models. *Journal of*
914 *Climate*, 27(19), 7185–7202. <https://doi.org/10.1175/JCLI-D-13-00447.1>

915 Zhang, X., Chen, N., Li, J., Chen, Z., & Niyogi, D. (2017). Multi-sensor integrated framework and index for agricultural
916 drought monitoring. *Remote Sensing of Environment*, 188, 141–163. <https://doi.org/10.1016/j.rse.2016.10.045>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.docx](#)