

Application of neural networks for drought forecasting based on the standardised precipitation index

Rana Z. Fathiyana¹⁾ , Deny Haryadi²⁾ , Naufal Ananda*³⁾ , Dinda J. Hidayat³⁾ , Haryas Subyantara Wicaksana³⁾ , Rahma Wafii Azahra⁴⁾ , Hakim Giraldi Saputra⁴⁾ 

- ¹⁾ Telkom University, Faculty School of Computing, The University Center of Excellence Intelligent Sensing-IoT, Jl. Halimun Raya No. 2, RT.15/RW.6, Guntur, Kecamatan Setiabudi, Kota Jakarta Selatan, Jakarta 12980, Indonesia
- ²⁾ Telkom University, Faculty School of Computing, The University Center of Excellence Human Centric Engineering, Jl. Halimun Raya No. 2, RT.15/RW.6, Guntur, Kecamatan Setiabudi, Kota Jakarta Selatan, Jakarta 12980, Indonesia
- ³⁾ Meteorological, Climatological, and Geophysical Agency, Jl. Angkasa I No. 2 Kemayoran, Jakarta Pusat 10610, Indonesia
- ⁴⁾ Telkom University, Faculty School of Computing, Information Technology, Jl. Halimun Raya No. 2, RT.15/RW.6, Guntur, Kecamatan Setiabudi, Kota Jakarta Selatan, Jakarta 12980, Indonesia

* Corresponding author

RECEIVED 03.12.2024

ACCEPTED 07.03.2025

AVAILABLE ONLINE 06.06.2025

Abstract: Based on data from the National Disaster Management Agency, South Sumatra is one of the provinces with a reasonably large drought-affected area, totalling 8,853,691.009 ha. Drought is a hydrometeorological disaster, characterised by anomalous rainfall below normal levels. Reduced rainfall can lead to decreased soil moisture, reduced river flows, and a general scarcity of water, which limits availability of water both on the surface and in the soil. To anticipate and mitigate the impacts of drought, an accurate forecasting system is essential for effective disaster management and mitigation. This research focuses on forecasting drought using the standardised precipitation index (*SPI*) based on Long Short-Term Memory (LSTM) and Multilayer Perceptron (MLP) algorithms. It compares LSTM and MLP algorithms by integrating rainfall data from the FY-4A satellite and observational rain gauges, which are processed to generate *SPI* values. These data are employed to train and test MLP and LSTM models in predicting future drought conditions. The results indicate that drought can be effectively predicted using both MLP and LSTM. However, the MLP outperforms the LSTM, as reflected by a higher Nash–Sutcliffe efficiency (*NSE*) value, a lower error rate, and a predicted date trend that more closely aligns with actual observations.

Keywords: drought, forecasting, FY-4A satellite, long short-term memory (LSTM), multilayer perceptron (MLP), rainfall

INTRODUCTION

Climate change has resulted in several hydrometeorological disasters worldwide, such as floods, landslides, and droughts (Legionosuko *et al.*, 2019). It is expected to increase the frequency, intensity, and duration of drought events (Peterson *et al.*, 2014). Drought is a hydrometeorological disaster characterised by anomalously low rainfall below normal levels (Hartanto *et al.*, 2023; Yulizar *et al.*, 2024). Reduced rainfall can lead to declining soil moisture, reduced river flows, and a general scarcity

of water, which limits its availability both on the surface and in the soil (Saidah *et al.*, 2019).

The consequences of drought have significantly increased in both developed and developing countries. While agriculture is the first and most severely affected industry, other industries also face significant disadvantages. These include energy sector, public, tourism, transportation, urban water supply, and the environment (Wilhite, Sivakumar and Pulwarty, 2014).

Indonesia is heavily dependent on agricultural and plantation products. The extensive impacts of drought on rural areas

include irrigation water shortages, limited cultivation areas, reduced land productivity, decreased crop yields, and lower farmer incomes. Based on data from the National Disaster Management Agency (Ind.: Badan Nasional Penanggulangan Bencana – BNPB), South Sumatra is one of the provinces with a relatively large drought-affected area, totalling 8,853,691.009 ha. Drought conditions can also increase the risk of forest and land fires. To anticipate and reduce the impacts of drought, an accurate forecasting system is essential for effective disaster mitigation and management.

The World Meteorological Organization (WMO), a United Nations agency responsible for meteorology, states that drought can be monitored and classified using the standardised precipitation index (*SPI*) (WMO, 2021). The *SPI* is a simple, robust, easy-to-understand indicator that is independent of climatic factors. It was created by McKee *et al.* in 1993 and 1995 to more accurately represent a region's wetness and dryness (Guttman, 1999). The *SPI* is widely used for analysing, measuring, and monitoring drought levels based on rainfall data (Xu *et al.*, 2020). It enables early identification, monitoring, and modelling of drought by accounting for spatial and temporal distribution of rainfall across multiple time scales (Prabowo *et al.*, 2024).

Previous research by Ali *et al.* (2017) used the MLP-ANN algorithm to predict drought in the Pakistan region for the period 1975–2012. In this study, MLP-ANN was able to estimate the index and capture variations in the one-month time-scale drought index (Ali *et al.*, 2017). Drought forecasting using the *SPI* at a 3-month time scale was conducted by Mohammed Salisui and Shabri, (2020) for the region of Iran. The data used in their research consisted of observational records covering the period of 1980–2014. The methods used included Adaptive Neuro-Fuzzy Interface System (ANFIS), Multilayer Perceptron (MLP), Radial Basis Function Neural Network (RBFNN), and Support Vector Machine (SVM). Among these, the ANFIS model yielded the best accuracy, achieving the lowest root mean squared error (*RMSE*) of 1.16 and the lowest mean absolute error (*MAE*) of 1.10.

Drought forecasting research was also conducted by Bouaziz, Medhioub and Csaplovisc (2021). This study focused exclusively on annual drought forecasting using *SPI*-12, based on Extreme Learning Machine (ELM). The data used consisted of rainfall records from the CHIRPS satellite for the Sfax region, Tunisia, from 1981 to 2019. The study developed a machine learning model that uses remote rainfall data and standard rainfall indices to monitor and forecast drought events. In a separate study, Coşkun and Citakoglu (2023) compared the performance of LSTM and the EML algorithms. The LSTM model produced the best evaluation results, with *MAE* of 0.11 and R^2 of 0.97 (Coşkun and Citakoglu, 2023).

Research on drought disaster forecasting using the *SPI* has been widely conducted. These studies used time series data from several previous years, with rainfall serving as the primary input parameter. However, artificial neural network-based models for drought prediction are still not widely used in Indonesia. Most drought models in Indonesia still use conventional statistical methods. The *SPI* values produced by the Climate Early Warning System (CEWS) in Indonesia are based on statistical data processing. Although previous research has used machine learning methods for drought modelling, few have utilised the LSTM and MLP algorithms, especially in the context of *SPI*-based drought forecasting in Indonesia. This study aims to

develop a neural network algorithm to predict drought events based on *SPI* values derived from ground observations and satellite data.

MATERIALS AND METHODS

MATERIALS STUDY

This research focuses on the South Sumatra Province of Indonesia, which has an area of 91,592.43 km² and is located between latitudes 1.00°S–4.75°S and longitudes 102.00°E–106.00°E. Based on data from the National Disaster Management Agency, South Sumatra is one of the provinces with a relatively large drought-affected area, totalling 8,853,691.009 ha. This study uses monthly total rainfall data for the 2019–2022 period. The data have been sourced from QPE FY-4A satellite data and corrected using data from five rain gauges. Rain gauges provide daily total rainfall data for the same period, which are then resampled into monthly data to align with QPE FY-4A satellite data. The study area, along with the distribution of rainfall measurement stations in the South Sumatra, is presented as a case study on the map (Fig. 1). The study area is divided into five technology grids, according to rain gauge locations.

In this research, rainfall estimates from FY-4A will be assessed for accuracy against the monthly rainfall data from rain gauges. The comparison will produce a bias correction to improve the accuracy of the FY-4A rainfall estimates (Hartanto *et al.*, 2024). The corrected FY-4A rainfall data with then be used as input for drought estimation in the South Sumatra region using LSTM and MLP algorithms. Locations of rain gauges can be seen in Table 1.

Fengyun-4A (FY-4A) is an advanced geosynchronous radiation imager (AGRI) satellite developed by China. One of FY-4A products is Quantitative Precipitation Estimation (QPE), which provides data with temporal resolutions of 1, 3, 6 and 24 h (Song *et al.*, 2024). The FY-4A QPE estimates precipitation by deriving transfer characteristic parameters from infrared brightness temperature, using probability density matching with real-time rolling updates (Song *et al.*, 2024). The spatial resolution of FY-4A QPE product is 4 km (Yin, Baik and Park, 2022). The FY-4A has 14 channels, comprising visible light, near-infrared, shortwave, mid-wave, water vapour, and longwave infrared spectra (Ren *et al.*, 2021).

Data obtained from FY-4A can be used for forecasting drought based on the *SPI*. The FY-4A satellite provides high-resolution QPE data, which can be integrated with other meteorological data to monitor and predict drought conditions (Luo *et al.*, 2023). The high temporal and spatial resolution of FY-4A data make it suitable for drought forecasting and monitoring (Ren *et al.*, 2021).

Based on research by McKee, Doesken and Kleist (1995) and Khan *et al.* (2020), the *SPI* is a drought index derived from long-term precipitation data. The *SPI* is widely used due to its robustness, flexibility, and ease of calculation. It enables early identification, monitoring, and modelling of drought by accounting for the spatial and temporal distribution of rainfall over multiple time scales.

The precipitation data conforms the Gamma distribution, which is then adjusted to a normal distribution. The Gamma

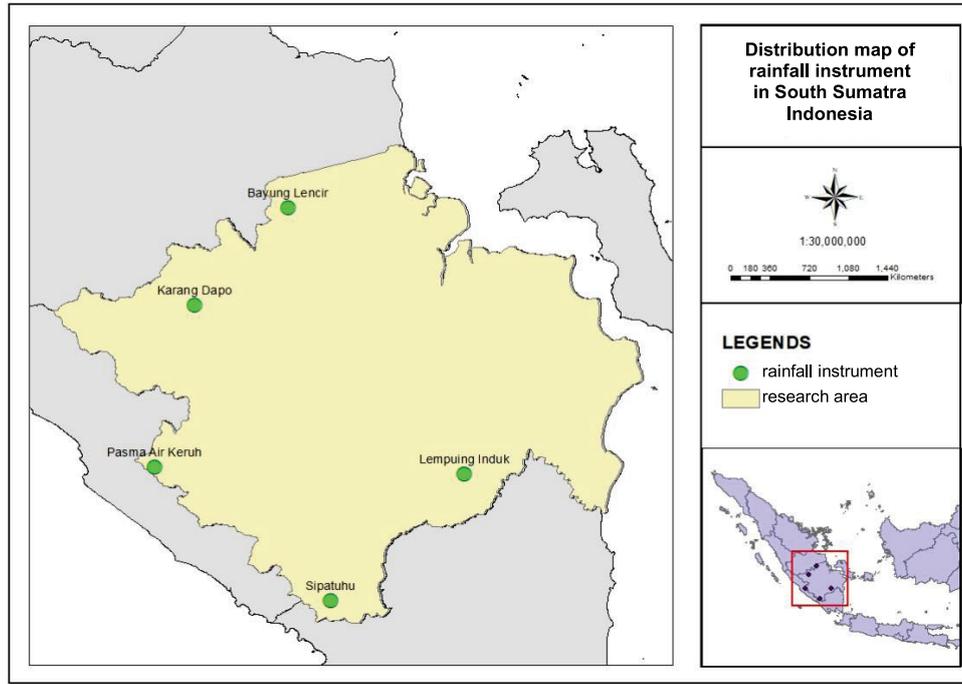


Fig. 1. Distribution of rain gauge in this study; source: own elaboration

Table 1. Rain gauge locations used in this study

Grid	Site instrument	Elevation (m)	Coordinates	
			latitude (S)	longitude (E)
1	Bayung Lencir	15	2.063	103.679
2	Karang Dapo	58	2.733	103.034
3	Pasma Air Keruh	22	3.851	102.761
4	Sipatuhu	720	4.767	103.972
5	Lempuing Induk	167	3.895	104.884

Source: own study.

distribution is well-suited for representing sequential rainfall data over time periods (Zhang *et al.*, 2020). The Gamma distribution probability function is calculated as follows:

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad (1)$$

where: α = total rainfall, β = shape variables, x = Gamma function.

The parameters α and β need to be estimated to model the Gamma distribution probability function (Jalalkamali, Moradi and Moradi, 2015). According to Khan, Muhammad and El-Shafie (2020), McKee, Doesken, and Kleist, these two parameters can be approximated using Thom's equation:

$$\alpha = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \quad (2a)$$

$$\beta = \frac{\bar{x}}{\alpha} \quad (2b)$$

$$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n} \quad (2c)$$

The parameter α is calculated using the variable A , derived from the natural logarithm of rainfall data. This parameter is then used to calculate β by dividing the mean rainfall by α . After approximating β and α , the function $g(x)$ is integrated over x to yield the cumulative probability function $G(x)$. This step is explained as follows:

$$G(x) = \int_0^x g(x) dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \quad (3)$$

If rain does not happen, then the $G(x)$ equation needs to be adjusted. If q is the probability of dry day, then the cumulative probability function equation is stated as:

$$H(x) = q + (1 - q)G(x) \quad (4)$$

The function $H(x)$ is then converted into a standard normal distribution. $H(x)$ is utilised to calculate the t value under two conditions, which yields the SPI value. The SPI equation is calculated as follows:

$$SPI = - \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right), t = \sqrt{\ln \frac{1}{[H(x)]^2}}; \quad 0 < H(x) < 0.5 \quad (5)$$

$$SPI = \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right), t = \sqrt{\ln \frac{1}{1 - [H(x)]^2}}; \quad 0.5 < H(x) < 1 \quad (6)$$

where: c = coefficient of nominator and d = coefficient of denominator; both have certain specific values: $c_0 = 2.515517$, $c_1 = 0.802853$, $c_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, and

$d_3 = 0.001308$. This index is quite compatible to seasonal zone variations and defines rainfall deficits compendiously. The WMO also uses *SPI* as a drought monitoring tool (Mohammed Salisu and Shabri, 2020).

The classification of *SPI* index as a drought type is shown in Table 2.

Table 2. Classification of standardised precipitation index (*SPI*)

<i>SPI</i> index	Drought type
≥ 2.0	extremely wet
$1.5 \leq SPI \leq 1.99$	very wet
$1.0 \leq SPI \leq 1.49$	moderately wet
$-0.99 \leq SPI \leq 0.99$	near normal
$-1.49 \leq SPI \leq -1.0$	moderately dry
$-1.99 \leq SPI \leq -1.5$	severely dry
≤ -2.0	extremely dry

Source: WMO (2012).

Using *SPI*, drought is forecasted in 1 month (*SPI1*), 3 months (*SPI3*), 6 months (*SPI6*) and 12 months (*SPI12*). In this research, we focused using *SPI3* (3 months) for forecasting drought because *SPI3* is capable of monitoring drought (Xu et al., 2020).

STUDY METHODS

This research aims to design an early warning model for drought in South Sumatra Province using *SPI3*, based on machine learning algorithms. The algorithms include Long Short-Term Memory (LSTM) and Multilayer Perceptron (MLP). The overall research framework is presented in Figure 2.

Based on Figure 2, the research steps include a literature study, rainfall dataset collection, dataset pre-processing, *SPI3*-based drought forecasting model design, and model evaluation. The literature study involved reviewing previous relevant research. It examined the theoretical basis of drought, data mining, *SPI*, rainfall, satellites, and machine learning algorithms. Rainfall data were collected by downloading satellite rainfall data and observational rain gauge data. FY-4A QPE satellite data were downloaded from the website <http://fyearth.nsmc.org.cn/>, while observational rain gauge data were obtained from the BMKGSoft page (<https://bmgksoft.bmkg.go.id/Met-View/#default>).

Dataset pre-processing was done by preparing FY-4A satellite data and observational rain gauge data to generate processed *SPI* data. The pre-processing includes resampling the observational rainfall data into monthly intervals, performing a range check to validate the rainfall data, calculating *SPI* values from the rainfall data, and transforming and segmenting *SPI* data into training and prediction model testing data (Akbar et al., 2024).

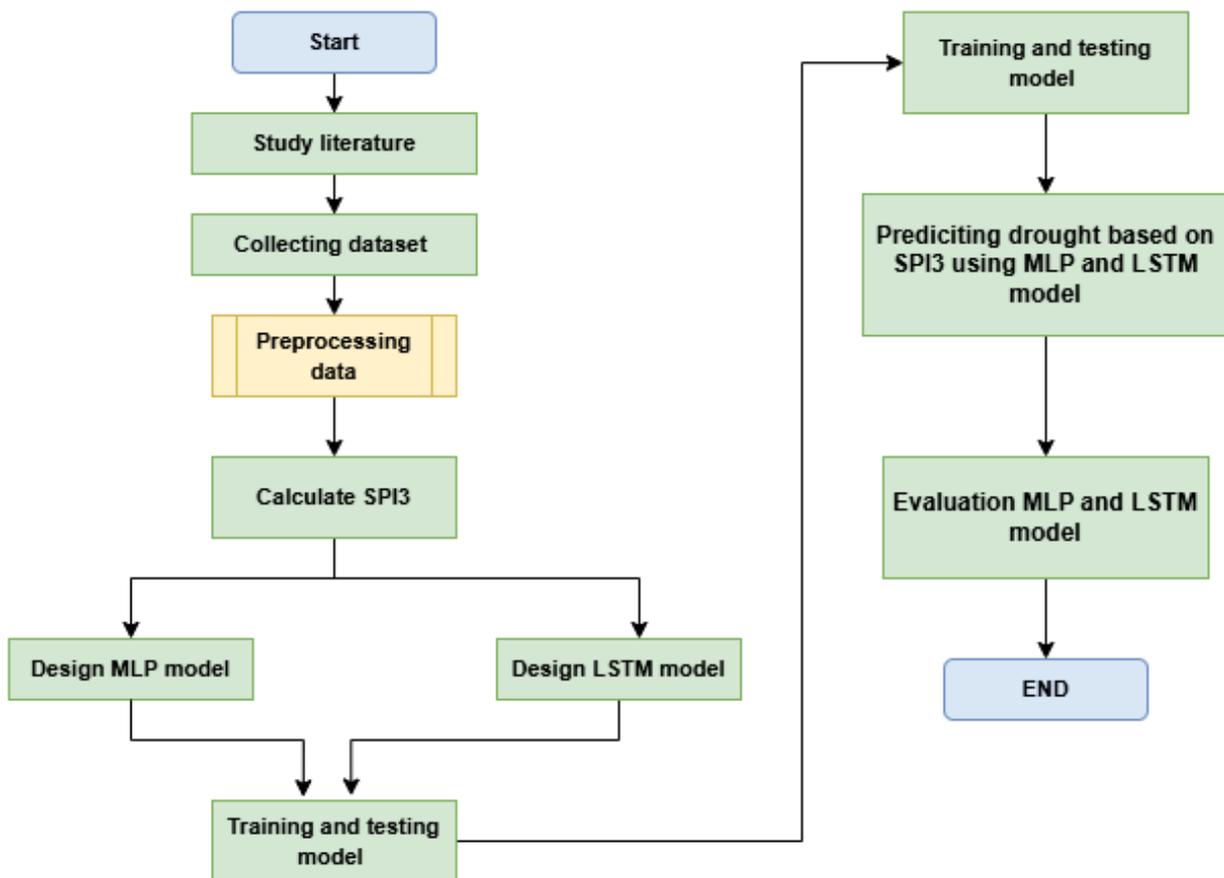


Fig. 2. Flowchart system in this study; *SPI3* = standardised precipitation index using 3 months, MLP = multilayer perceptron, LSTM = long short-term memory; source: own elaboration

The corrected FY-4A data were converted into *SPI3* values for the 2019–2022 period. For model development, 80% of the data were used as training data and 20% as testing data. The input data consisted of historical *SPI3* values. It comprises previous historical data at $SPI3(t - 1)$, $SPI3(t - 2)$ and $SPI3(t - 3)$. These inputs will predict current $SPI3(t)$ as the output data.

The algorithm design involves compiling machine learning models and hyperparameters. The algorithm parameters include weight and bias values obtained through the training process, while the hyperparameters consist of the number of neurons and number of hidden layers (Tab. 3). The *SPI* parameters drought forecasting model is based on LSTM and MLP algorithms. These two algorithms operate separately to evaluate accuracy of model.

Table 3. Multilayer perceptron (MLP) and long short-term memory (LSTM) hyperparameter model used in this study

Model	Type of hyperparameter	Value of hyperparameter
MLP	hidden layer	{1}
	neuron	{1–100}
	alpha	{0.0001, 0.001, 0.01, 0.1, 1}
	activation	{'relu', 'identity'}
	solver	'lbfgs', 'sgd', 'adam'
	max iteration	100
LSTM	hidden layer	{1, 2}
	neuron	{30, 35, 40, 45, 50}
	activation	{'sigmoid', 'tanh', 'relu'}
	epoch	{1000}

Source: own elaboration.

PARAMETER EVALUATION

The accuracy evaluation is conducted using testing data from the prediction model. The predicted *SPI* values are compared to the actual *SPI* data according to the Nash–Sutcliffe efficiency (*NSE*), root mean squared error (*RMSE*) and mean average percentage error (*MAPE*). The *NSE* is ratio of mean square error (*MSE*) and variance that indicates closeness of relationship between reference data and forecasting data (Ding *et al.*, 2022). This parameter is commonly used in hydrology and drought index forecasting (Alawsi *et al.*, 2022). The *NSE* ranges from $-\infty$ to 1, with an *NSE* value of 1 indicating a perfect match between the model output and the observed data. The *NSE* equation is expressed as follows:

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_1)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

where: n = total data, i = data index, \bar{y} = mean of actual data, y_i = actual data, \hat{y}_1 = predicted data.

Root mean squared error (*RMSE*) expresses error value of *SPI* forecasting model. The *RMSE* range has the same units as the *SPI* parameters. The *RMSE* equation is expressed as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (8)$$

The mean average percentage error (*MAPE*) expresses an error value of the *SPI* forecasting model. The percentage error scale is 0–100%. The *MAPE* equation is expressed as follows:

$$MAPE = \frac{100\%}{m} \sum_{i=1}^m \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (9)$$

RESULTS AND DISCUSSION

PRE-PROCESSING RESULTS CORRECTED FY-4A DATA

Pre-processing was applied to the FY-4A total rainfall data to obtain correction values by comparing them with direct measurements from observational rain gauges. Monthly rainfall data from FY-4A for the period 2019–2022 was corrected using corresponding monthly data from the rain gauges. This process results in a bias factor for each FY-4A grid, which was then utilised to correct FY-4A rainfall data. Furthermore, the corrected data were evaluated against initial data using correlation coefficient and *RMSE* parameters. The results of the pre-processing of the FY-4A monthly rainfall data are presented in Table 4.

Based on Table 4, the bias factor values for each grid vary considerably. Applying a bias correction is necessary to improve the spatial accuracy of rainfall measurements from FY-4A (Ouatiki, Boudhar and Chehbouni, 2023). Satellite-based rainfall measurements are biased due to factors such as sensor correction errors, evaporation, and wind speed circulation (Wang *et al.*, 2023). The results of the statistical analysis indicate that rainfall estimates from FY-4A show a moderate correlation with observational rainfall gauges readings and a significant decrease in *RMSE*.

Figure 3 shows the plotting of observational rainfall gauge values, satellite rainfall data, and corrected rainfall data. The application of the bias factor improves the accuracy of the FY-4A rainfall data, as evidenced by the decrease in the *RMSE* values on each grid when compared to the observational rainfall gauge data. The corrected FY-4A rainfall data closely align with the observational measurements. The increased accuracy supports the use of the corrected FY-4A rainfall data as input for *SPI3*-based drought prediction.

Table 4. Corrected monthly rainfall data of Fengyun-4A satellite data

Grid	Bias factor	Correlation	Root mean squared error (mm)		Decrease (%)
			unadjust bias	adjust bias	
1	1.13	0.67	187.14	109.40	41.54
2	2.05	0.68	164.80	63.52	61.46
3	0.91	0.60	220.36	139.61	36.46
4	2.12	0.69	185.67	125.85	32.22
5	1.06	0.65	185.83	104.89	43.56

Source: own study

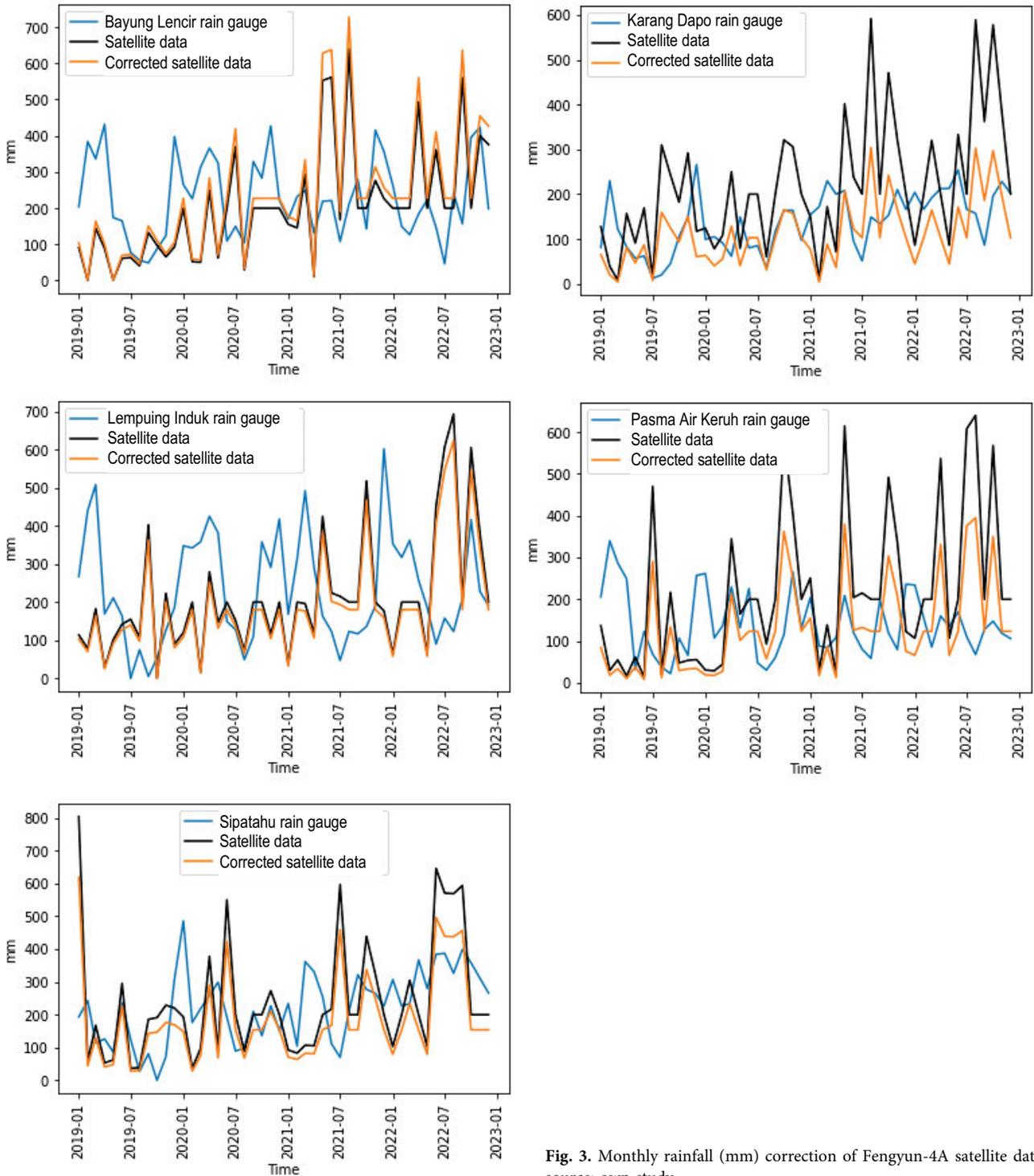


Fig. 3. Monthly rainfall (mm) correction of Fengyun-4A satellite data; source: own study

EVALUATION OF DROUGHT FORECASTING USING STANDARDISED 3-MONTHS PRECIPITATION INDEX

Drought prediction refers to the numerical forecasting of the drought index using Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM). This research focuses on predicting the SPI3 drought index for the three-month lead time. SPI3 information helps provide medium-term warnings regarding reduced water discharge in wells or other water sources (Li et al., 2023). The SPI3 prediction results from each algorithm are evaluated against SPI3 data using the NSE, RMSE, and MAPE

evaluation parameters. The accuracy of SPI3 predictions using Multiple Linear Regression (MLR), Support Vector Regression (SVR), MLP, and LSTM is presented in Table 5.

According to Table 5, the NSE values are positive across all models. This indicates a strong relationship between predicted and actual SPI3 data. A positive NSE value indicates that the variance of the prediction model is relevant to actual data. Applying MLP model as a drought predictor produces better NSE values than LSTM. The RMSE value <1 reduces the possibility of misclassification of drought information because the range of transition from wet to dry conditions is -1 to 1. The RMSE and

Table 5. Performance of accuracy estimator multiple linear regression (MLR), support vector regression (SVR), long short-term memory (LSTM) and multilayer perceptron (MLP)

Grid	Algorithm	NSE (-)		RMSE (-)		MAPE (%)	
		TR	VAL	TR	VAL	TR	VAL
1	MLR	0.75	0.72	0.66	0.98	5.6	10.7
	SVR	0.77	0.69	1.94	1.77	16.5	14.8
	MLP	0.46	0.69	0.74	0.87	5.8	9.8
	LSTM	0.18	0.16	1.42	1.43	11.7	12.3
2	MLR	0.72	0.73	0.88	0.91	7.6	10.1
	SVR	0.77	0.71	2.16	2.22	19.0	21.2
	MLP	0.43	0.73	0.81	0.86	6.5	9.2
	LSTM	0.21	0.22	1.45	1.47	17.9	16.7
3	MLR	0.82	0.78	0.86	0.90	7.4	9.8
	SVR	0.84	0.76	2.32	2.61	20.9	26.9
	MLP	0.52	0.75	0.78	0.85	6.3	6.8
	LSTM	0.19	0.18	1.51	1.53	16.1	15.8
4	MLR	0.79	0.74	0.95	1.02	8.6	17.8
	SVR	0.79	0.71	2.45	2.27	22.2	21.7
	MLP	0.44	0.74	0.93	0.96	7.8	17.2
	LSTM	0.21	0.19	1.43	1.42	25.1	24.5
5	MLR	0.75	0.75	0.74	0.96	5.4	11.3
	SVR	0.76	0.71	1.82	2.49	15.3	21.9
	MLP	0.32	0.76	0.68	0.94	5.6	10.2
	LSTM	0.22	0.23	1.53	1.67	10.9	12.0

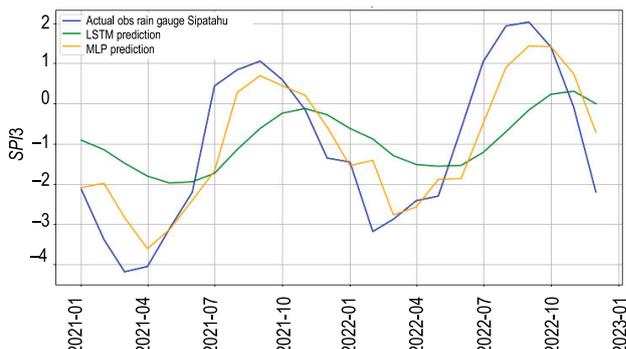
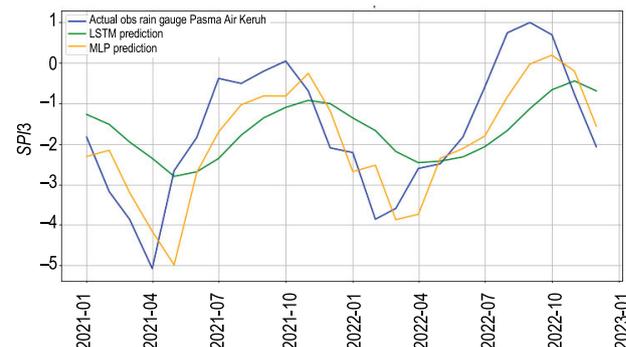
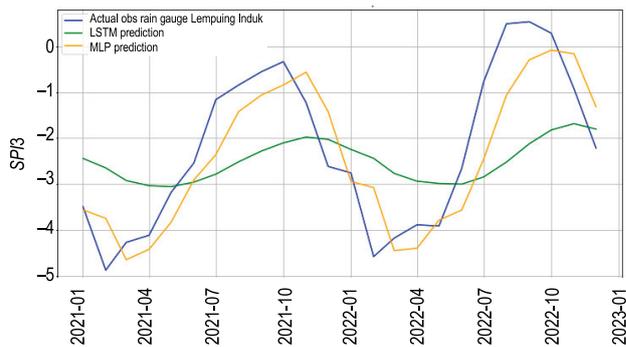
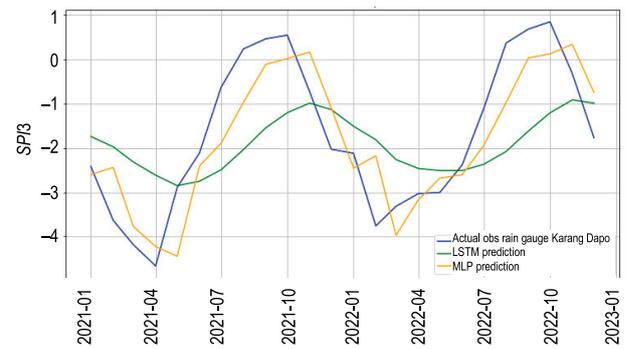
Explanations: TR = training, VAL = validation/testing.
Source: own study.

MAPE values further show that the error produced by the LSTM model is higher than that of the MLP. This shows that MLP predictor performance is better than that of the LSTM in predicting drought in the South Sumatra region. The regression-based algorithm achieved a high NSE value, demonstrating performance that was nearly comparable to that of the MLP. However, the lowest RMSE and MAPE values were consistently achieved using the MLP during the testing process. This proves that the MLP algorithm provides the best performance in predicting SPI3 compared to statistical regression methods. Grid 4 produced the lowest NSE values and the highest RMSE and MAPE, likely due to its significantly higher elevation compared to geographical contours in the other grids. The MLP hyperparameters may need to be further adjusted when applied to high-elevation topographies, such as hilly areas, as the temporal characteristics of SPI3 in highlands tend differ from those in lowlands.

The SPI3 prediction plots for each grid in South Sumatra Province are presented in Figure 4. The MLP-based prediction data appear closer to actual data. The LSTM model tends to be smoother and less responsive to sudden changes than actual data. This is particularly evident in areas with sharp peaks and valleys in the SPI3 data, where LSTM predictions often produce gentler peaks that do not fully match the observed values. The MLP

model is more responsive and closer to predicting sudden changes in SPI3. However, in some areas, especially in sharp changes, this model can predict overshooting fluctuations or vice versa. The performance of the MLP is considered superior to that of LSTM in predicting drought, as evidenced by the closer alignment of predicted and actual data trends, as well as higher NSE values achieved by the MLP model.

The upper quartile, lower quartile, and median values of MLP and SVR based predictions show minimal deviation from



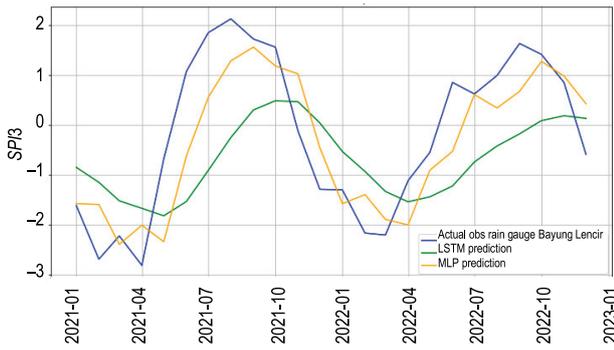


Fig. 4. Prediction of standardised precipitation index for 3-months (*SPI3*) using multilayer perceptron (MLP) and long short-term memory (LSTM); source: own study

the actual data (Fig. 5). In contrast, the corresponding values of the *SPI3* predictions produced by LSTM and MLR models deviate significantly from the actual data. This indicates that MLP model maintains superior performance compared to LSTM and MLR, as it better preserves the statistical characteristics of the original prediction dataset.

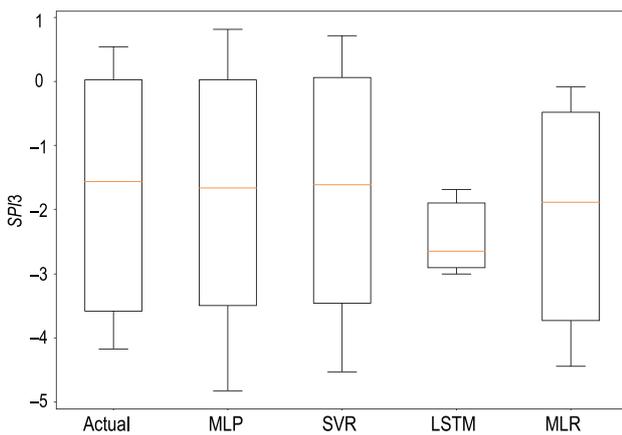


Fig. 5. Boxplot of standardised precipitation index for 3-months (*SPI3*) prediction on Grid 3 – Lempuing Induk site; MLP = multilayer perceptron, SVR = support vector regression, LSTM = long short-term memory, MLR = linear regression; source: own study

DISCUSSION

The MLP algorithm has optimal performance for long-term drought prediction, such as with the *SPI3* index. The MLP model is responsive and closely follows the actual *SPI3* patterns over long-term periods. Providing early warning information for drought is critical, and the implementation of MLP is highly appropriate. The statistical characteristics of the MLP prediction results closely align with those of the actual data. Neural network-based algorithms are capable to detect nonlinearity patterns in datasets compared to regression-based methods such as MLR and SVR. This is due to adaptive learning process during training, where network weights and biases are continuously updated, enabling the model to effectively learn and predict drought conditions in accordance with the characteristics of the *SPI3* index.

CONCLUSIONS

Drought prediction in this study was designed through pre-processing stages, including pre-processing of the Fengyun-4A satellite monthly rainfall dataset for the 2019–2022 period, transforming the rainfall dataset into standardised precipitation index (*SPI*) values, segmenting dataset, determining hyper-parameters of Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) algorithms, and testing the performance of standardised precipitation index using three-month (*SPI3*) prediction accuracy. Overall, MLP and LSTM models accurately predict drought conditions; however, MLP is considered superior compared to LSTM. This is evident from the higher *NSE* values and lower error metrics observed in the MLP model, as well as closer alignment of MLP predictions with the actual data trends. Nevertheless, the study has a limitation concerning the hyper-parameter tuning process. Certain hyperparameter configurations may only be suitable for certain drought cases in South Sumatera Province. It is therefore recommended to tune MLP hyper-parameters based on various optimisation methods such as genetic algorithm, simulated annealing, ant colony optimisation, stingless bee algorithm, or other approaches.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Telkom University, Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG), Region II of Indonesia for Meteorology, Climatology, and Geophysics, and the Banten Climatology Station for providing the data used in this study.

FUNDING

This research was supported by a grant from Telkom University under the Non-Collaborative Research Scheme, Period 2, Fiscal Year 2024, Grant Number 1052/LIT06/TUKJ-PMK/2024.

CONFLICT OF INTERESTS

The authors declare that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest. There are no competing interest with the funders.

REFERENCES

Akbar, G. *et al.* (2024) “Multivariate imputation chained equation on solar radiation in automatic weather station,” *Jurnal Penelitian Pendidikan IPA*, 10(7), pp. 3633–3639. Available at: <https://doi.org/10.29303/jppipa.v10i7.7679>.
 Alawsi, M.A. *et al.* (2022) “Drought forecasting: A review and assessment of the hybrid techniques and data pre-processing,” *Hydrology*, 9(7), 115. Available at: <https://doi.org/10.3390/hydrology9070115>.

- Ali, Z. *et al.* (2017) "Forecasting drought using multilayer perceptron artificial neural network model," *Advances in Meteorology*, 2017 (1), 5681308. Available at: <https://doi.org/10.1155/2017/5681308>.
- Bouaziz, M., Medhioub, E. and Csaplovics, E. (2021) "A machine learning model for drought tracking and forecasting using remote precipitation data and a standardized precipitation index from arid regions," *Journal of Arid Environments*, 189, 104478. Available at: <https://doi.org/10.1016/j.jaridenv.2021.104478>.
- Coşkun, Ö. and Citakoglu, H. (2023) "Prediction of the standardized precipitation index based on the long short-term memory and empirical mode decomposition-extreme learning machine models: The case of Sakarya, Türkiye," *Physics and Chemistry of the Earth, Parts A/B/C*, 131, 103418. Available at: <https://doi.org/10.1016/j.pce.2023.103418>.
- Ding, Y. *et al.* (2022) "Application of a Hybrid CEEMD-LSTM Model Based on the Standardized Precipitation Index for drought forecasting: The case of the Xinjiang Uygur Autonomous Region, China," *Atmosphere*, 13(9), 1504. Available at: <https://doi.org/10.3390/atmos13091504>.
- Guttman, N.B. (1999) "Accepting the Standardized Precipitation Index: A calculation algorithm," *JAWRA Journal of the American Water Resources Association*, 35(2), pp. 311–322. Available at: <https://doi.org/10.1111/j.1752-1688.1999.tb03592.x>.
- Hartanto *et al.* (2023) "Evaluation of Meteorological Radar Precipitation Forecast in Banten," *2023 International Conference on Information Technology and Computing (ICITCOM)*, 297–300. Available at: <https://doi.org/10.1109/ICITCOM60176.2023.10442051>.
- Hartanto *et al.* (2024) "Spatial evaluation rainfall estimation on weather radar using Marshall–Palmer reflectivity–rainfall rate in Banten," *Journal of Water and Land Development*, 62, pp. 193–200. Available at: <https://doi.org/10.24425/jwld.2024.151567>.
- Jalalkamali, A., Moradi, M. and Moradi, N. (2015) "Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index," *International Journal of Environmental Science and Technology*, 12(4), pp. 1201–1210. Available at: <https://doi.org/10.1007/s13762-014-0717-6>.
- Khan, Md. M.H., Muhammad, N.S. and El-Shafie, A. (2020) "Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting," *Journal of Hydrology*, 590, 125380. Available at: <https://doi.org/10.1016/j.jhydrol.2020.125380>.
- Legionosuko, T. *et al.* (2019) "Posisi dan strategi Indonesia dalam Menghadapi Perubahan Iklim guna Mendukung Ketahanan Nasional [Indonesian position and strategy in responding to climate change to strengthen national resilience]," *Jurnal Ketahanan Nasional*, 25(3), 295. Available at: <https://doi.org/10.22146/jkn.50907>.
- Li, H. *et al.* (2023) "Machine learning-based bias correction of precipitation measurements at high altitude," *Remote Sensing*, 15(8), 2180. Available at: <https://doi.org/10.3390/rs15082180>.
- Luo, H. *et al.* (2023) "Validation analysis of drought monitoring based on FY-4 satellite," *Applied Sciences*, 13(16), 9122. Available at: <https://doi.org/10.3390/app13169122>.
- McKee, T.B., Doesken, N.J. and Kleist, J. (1995) "Drought monitoring with multiple time scales," in *Proceedings of the Ninth Conference on Applied Climatology*, Dallas TX, Jan 15–20, 1995. American Meteorological Society, pp. 233–236.
- Mohammed Salisu, A. and Shabri, A. (2020) "A Hybrid Wavelet-Arima Model for Standardized Precipitation Index Drought forecasting," *MATEMATIKA*, 36(2), pp. 141–156. Available at: <https://doi.org/10.11113/matematika.v36.n2.1152>.
- Ouatiki, H., Boudhar, A. and Chehbouni, A. (2023) "Can bias correction techniques improve remote sensing-based rainfall estimates in a semi-arid context: Case of the Oum Er-Rbia River Basin in Morocco," *EGU General Assembly 2023*, Vienna, Austria, 24–28 Apr 2023, EGU23-8537. Available at: <https://doi.org/10.5194/egusphere-egu23-8537>.
- Peterson, T.C. *et al.* (2014) "Changes in weather and climate extremes: State of knowledge relevant to air and water quality in the United States," *Journal of the Air & Waste Management Association*, 64(2), pp. 184–197. Available at: <https://doi.org/10.1080/10962247.2013.851044>.
- Prabowo, M.A. *et al.* (2024) "Drought prediction based Standardized Precipitation Index Using Multilayer Perceptron Model," *2024 International Conference on Smart Computing, IoT and Machine Learning (SIML)*, pp. 262–267. Available at: <https://doi.org/10.1109/SIML61815.2024.10578245>.
- Ren, J. *et al.* (2021) "Evaluation and improvement of FY-4A AGRI quantitative precipitation estimation for summer precipitation over complex topography of Western China," *Remote Sensing*, 13(21), 4366. Available at: <https://doi.org/10.3390/rs13214366>.
- Saidah, H. *et al.* (2019) "Model Runtut waktu untuk peramalan indeks kekeringan daerah Lombok Utara [Time series model for North Lombok drought indices forecasting]," *Jurnal Sains Teknologi & Lingkungan*, 5(2), pp. 123–132. Available at: <https://doi.org/10.29303/jstl.v5i2.130>.
- Song, Y. *et al.* (2024) "Evaluation of Fengyun geosynchronous orbit and GPM satellites precipitation products over the southeastern Tibetan plateau," *International Journal of Remote Sensing*, 45(16), pp. 5616–5639. Available at: <https://doi.org/10.1080/01431161.2024.2377834>.
- Wang, L. *et al.* (2023) "Towards Bias correction of satellite precipitation retrievals in complex regions with deep learning: A case study over Taiwan," *IGARSS 2023 – 2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 3803–3806. <https://doi.org/10.1109/IGARSS52108.2023.10283452>.
- Wilhite, D.A., Sivakumar, M.V.K. and Pulwarty, R. (2014) "Managing drought risk in a changing climate: The role of national drought policy," *Weather and Climate Extremes*, 3, pp. 4–13. Available at: <https://doi.org/10.1016/j.wace.2014.01.002>.
- WMO (2012) *SPI index for different locations*. Available at: https://library.wmo.int/records/item/39629-standardized-precipitation-index-user-guide?language_id=13 (Accessed: November 7, 2024).
- WMO (2021) *Guide to instruments and methods of observation (WMO-No. 8)*. Geneva: World Meteorological Organization. Available at: https://community.wmo.int/en/activity-areas/imop/wmo-no_8 (Accessed: November 7, 2024).
- Xu, D. *et al.* (2020) "Application of a Hybrid ARIMA–SVR Model based on the SPI for the forecast of drought – A case study in Henan Province, China," *Journal of Applied Meteorology and Climatology*, 59(7), pp. 1239–1259. Available at: <https://doi.org/10.1175/JAMC-D-19-0270.1>.
- Yin, G., Baik, J. and Park, J. (2022) "Comprehensive analysis of GEO-KOMPSAT-2A and FengYun satellite-based precipitation estimates across Northeast Asia," *GIScience & Remote Sensing*, 59(1), pp. 782–800. Available at: <https://doi.org/10.1080/15481603.2022.2067970>.
- Yulizar, D. *et al.* (2024) "Visualization of rainfall classification using rain gauge based on website," in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Bandung, Indonesia, February 21–23, 2024, pp. 1–5. Available at: <https://doi.org/10.1109/AIMS61812.2024.10512600>.
- Zhang, Y. *et al.* (2020) "Comparison of the ability of ARIMA, WNN and SVM models for drought forecasting in the Sanjiang Plain, China," *Natural Resources Research*, 29(2), pp. 1447–1464. Available at: <https://doi.org/10.1007/s11053-019-09512-6>.