Journal of Agrometeorology
(A publication of Association of Agrometeorologists)
ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 28 (3) : 345-355 (September - 2026)
https://doi.org/10.54386/jam.v28i3.3250
https://journal.agrimetassociation.org/index.php/jam

Research paper

A Methodological Approach for Selecting Climate Datasets for Drought Prediction Using ROC Analysis and Youden's Index

ESEQUIEL ROLANDO JIMÉNEZ ESPINOSA[1] and DONESSA LUCINDA DAVID[2*]

[1] *Department of Agriculture, Faculty of Agriculture and Forestry,* [2] *Department of Environmental Studies, Faculty of Earth and Environmental Sciences, University of Guyana, Turkeyen Campus, Greater Georgetown, Guyana, South America.*
[*] *Corresponding author email*: donessa.david@uog.edu.gy

ABSTRACT

This study presents a methodological approach for evaluating and selecting climate datasets for drought prediction in Guyana, based primarily on the combination of Receiver Operating Characteristic (ROC) analysis and Youden's Index (J). Traditional hydrological metrics—Nash–Sutcliffe Efficiency (NSE) and Kling–Gupta Efficiency Modified (KGEM) were applied only as comparators to highlight their limitations for event detection. Ten climate datasets available in CARiDRO were assessed at two Standardized Precipitation Index (SPI) time scales—3 months (SPI3) and 12 months (SPI12) using records from two contrasting stations: Georgetown (coastal) and Timehri (inland). ROC–Youden rankings were stable across stations and time scales and revealed clear leaders. At Georgetown, the highest J values were obtained by RCP 4.5 for both SPI3 (J = 0.44) and SPI12 (J = 0.48), followed by aexsk (J = 0.22 and 0.34, respectively). At Timehri, aexsk achieved the highest performance for both SPI3 (J = 0.21) and SPI12 (J = 0.35), while RCP 4.5 showed strong but second-best performance only at SPI12 (J = 0.32). In contrast, NSE and KGEM yielded inconsistent rankings between SPI3 and SPI12, underscoring the limitations of continuous-value error metrics for classifying drought versus non-drought events. By prioritizing classification skill (true/false event discrimination) over aggregate fit, the ROC–Youden framework provides an operationally relevant basis for dataset selection in agricultural drought monitoring. The SPI3-focused findings are directly applicable to short-term crop-water decisions, whereas SPI12 results inform longer-term water allocation.

Keywords: Climate datasets, Drought assessment tools, Drought forecasting, ROC analysis, Standardized Precipitation Index (SPI), Youden's Index

Droughts are considered as one of the most destructive natural hazards worldwide, with impacts on agricultural productivity, water resources, and socio-economic stability (Wilhite & Pulwarty, 2018). In the Caribbean Region and particularly in countries such as Guyana, agricultural drought poses a major challenge due to the strong dependence of farmers on rainfed agricultural systems as well as the vulnerability of irrigation schemes to seasonal and interannual rainfall variability (Farrell *et al.,* 2010; Trotman *et al.,* 2018). Additionally, due to the projected intensification of drought events under climate change scenarios, the timely detection and monitoring of drought conditions have become essential for agriculture disaster risk management (IPCC, 2021).

There are four types of droughts namely: meteorological, hydrological, socio-economic, and agricultural (Orimoloye *et al.,* 2022; Wilhite & Glantz, 1985). According to Paredes & Guevara, (2010) drought occurs when the magnitude of rainfall is significantly below the historical average. Agricultural droughts normally occur when soil moisture becomes insufficient to meet crop water demands, often as a consequence of meteorological anomalies in precipitation and temperature (Mishra & Singh, 2010). Drought impacts are closely linked to soil water availability, plant physiological responses, and yield losses, making its monitoring a priority for agricultural planning (Vicente-Serrano *et al.,* 2010).

There are several drought indices globally, however, the Standardized Precipitation Index (SPI) remains widely used in many countries due to its simplicity, temporal flexibility, and standardized comparability across various climates (McKee *et al.,* 1993). The SPI at short timescales (e.g., SPI-3) has the ability to capture soil

Article info - DOI: https://doi.org/10.54386/jam.v28i3.3250
Received:19 October 2025; Accepted: 18 June 2026; Published online : 7 September 2026

moisture deficits relevant to farming cycles of crops, while longer timescales (e.g., SPI-12) reflect hydrological storage and recharge patterns (Hayes *et al.,* 2011). Many studies have examined the role of climate datasets in agricultural drought modeling including Venturi *et al.,* (2024) in Mediterranean cropping and Nandgude *et al.,* (2023) in semi-arid India. Moreover, in the Latin America Region, Paredes-Trejo & Barbosa, (2017) analyzed satellite-derived precipitation data for SPI based agricultural drought monitoring. A study conducted in northern Mexico by Ortega-Gaucín *et al.,* (2021), also assessed the suitability of various climate products for drought early warning. Still, few studies have systematically ranked climate datasets according to their effectiveness in distinguishing drought from non-drought events using operational thresholds.

Notably, traditional metrics like the Nash–Sutcliffe Efficiency (NSE) and Kling–Gupta Efficiency Modified (KGEM) are widely used in hydrology, but they do not work as well for classifying events such as the start and end of droughts. The Receiver Operating Characteristic (ROC) analysis, on the other hand, measures how well events are detected and how often false alarms are avoided, making it better suited for monitoring droughts based on thresholds. The Area Under the Curve (AUC) from the ROC analysis provides an aggregate measure of classification performance, while Youden's Index (J) identifies the optimal trade-off between detection and false alarm rates. This combination has been applied in various environmental studies, including drought and flood prediction, but has received limited attention in the evaluation of climate datasets in agricultural drought detection studies (Hao & AghaKouchak, 2013; Venturi *et al.,* 2024).

Most drought research in the Caribbean region has primarily addressed seasonal rainfall forecasting, SPI trend analysis, and the impacts of climate variability (Beharry *et al.,* 2019; Miller & Ramseyer, 2020). However, the application of event based classification metrics such as ROC and Youden's Index, to systematically compare and rank climate datasets for agricultural drought prediction remains limited. Further, several predictive models have been proposed to forecast drought as part of climate monitoring efforts. The Cuban Institute of Meteorology developed the Caribbean Assessment of Regional Drought (CARiDRO) tool (Centella *et al.,* 2014), aimed at assessing drought conditions at regional and grid levels. The tool facilitates the evaluation of datasets derived from Regional Climate Models and gridded observational data. It uses SPI and SPEI indices to project various types of droughts (Centella *et al.,* 2017). The novelty of this study lies in applying ROC–Youden analysis not just to drought indices but to climate dataset evaluation itself, an approach rarely explored in agricultural contexts.

Hence, the objective of this study is to develop and apply a methodological framework that combines ROC analysis and Youden's Index to evaluate the predictive performance of multiple climate datasets for agricultural drought monitoring. This approach will address the current gap in operational drought monitoring by focusing on event detection accuracy under defined SPI thresholds, providing a replicable and statistically robust basis for dataset selection in agricultural and other applications.

MATERIALS AND METHODS

The study was conducted in Guyana which is located in the northern part of South America. The country experiences two distinct wet seasons (May–July and November–January) and two dry seasons (February–April and August–October) (Centella *et al.,* 2014; Centella *et al.,* 2017). Annual rainfall averages between 1,500 and 3,000 mm, with marked spatial variability between the humid coastal areas and the drier inland regions (WMO, 2012). Agricultural activity is concentrated along the fertile coastal plain, supported by irrigation infrastructure for crops such as rice and sugarcane, whereas hinterland agriculture relies mainly on rainfed systems.

Although the CARiDRO platform supports several drought indices, this study employed the Standardized Precipitation Index (SPI) (McKee *et al.,* 1993) to generate forecasts and assess observed drought conditions in Guyana.

Method for Dataset Selection

Rainfall stations and SPI Computation

For this study, monthly SPI values for the period 2010–2019 were calculated using rainfall records from two meteorological stations managed by the Hydrometeorological Service of Guyana: Georgetown (coastal lowland) and Timehri (inland). These locations were selected for the study due to their contrasting agro-climatic conditions and completeness of their datasets. Both stations are part of the Caribbean Assessment of Regional Drought (CARiDRO) platform, which ensures data comparability and quality control in the study.

The SPI values were computed using the "SPI Generator" software (National Drought Mitigation Center, 2018). Two SPI timescales were selected for the study. The SPI3 was selected to capture short-term agricultural impacts such as soil moisture deficits, while SPI12 was used to represent long-term hydrological conditions affecting groundwater and reservoir storage (WMO, 2012; WMO, 2016).

Georeferencing and dataset configuration

In carrying out the study, the geographical coordinates of the Georgetown and Timehri stations were determined and used to locate the corresponding grid cells. These coordinates were entered into the CARiDRO platform to generate SPI predictions for 2010–2019 using ten historical datasets covering the years 1961 to 2009. These datasets were: echam5, RCP 8.5, RCP 2.6, RCP4.5, aexsk, aexsc, aexsa, ENSMEAN, aexsl and aexsm. According to the IPCC AR5 (2013), climate projections under Representative Concentration Pathways (RCPs) are typically expressed as anomalies relative to a reference period. In these scenarios, projected changes in variables such as temperature or precipitation represent deviations from that baseline. Furthermore, RCP scenarios begin to diverge from the year 2006 due to differing greenhouse gas emission trajectories.

Datasets 1–4 correspond to outputs from General Circulation Models (GCMs) and Representative Concentration Pathway (RCP) scenarios, which project future climate conditions under different greenhouse gas emission trajectories. The RCP 2.6, 4.5, and 8.5 scenarios represent low, intermediate, and high emissions pathways, respectively. Dataset 1 (echam5) is a GCM simulation used for historical climate variability analysis. The datasets 5 to 7 and 9 to 10 are statistically downscaled projections generated within the CARiDRO platform which provides higher spatial resolution for application in the Caribbean Region. Additionally, dataset 8 (ENSMEAN) represents the ensemble mean of all available projections, offering a more robust estimate by averaging across models represented in the the CARiDRO platform. IPCC AR5 (2013) report notes that climate projections under Representative Concentration Pathways (RCPs) are generally expressed as anomalies relative to a reference period.

In this study, the selection of these datasets was based primarily on their regional relevance, validation in Caribbean and South American contexts, availability through the CARiDRO system, and previous use in drought risk assessments (Paredes-Trejo & Barbosa, 2017; Ortega-Gaucin *et al.,* 2018). The utilization of a mix of single-model simulations, multiple RCP scenarios, and an ensemble mean allows for a comprehensive evaluation of both historical variability and future projections for the areas under study.

Receiver Operating Characteristic (ROC) Analysis

The evaluation of the accuracy of drought event predictions and the predictive performance of the ten CARiDRO datasets was assessed using the Receiver Operating Characteristic (ROC) statistical method. The Receiver Operating Characteristic (ROC), was developed during World War II by electrical and radar engineers to detect enemy targets (Junge & Dettori, 2024). Presently, ROC analysis is widely applied in forecasting natural hazards, meteorology, and model performance assessment (Peres & Cancelliere, 2014; Murphy, 1996; Peres *et al.,* 2015). This approach was selected for application in this study due to its suitability in representing the cyclical behavior of drought indices, enabling the establishment of clear criteria to distinguish between wet and dry conditions over a given period.

A threshold value of –0.99 was adopted for this study to define the onset of drought or wet conditions (see Table 1). Below is a confusion matrix that was developed for the threshold value that was adopted.

- **True Positives (TP):** Predicted values ≥ –0.99 and observed values ≥ –0.99.
- **True Negatives (TN):** Predicted values < –0.99 and observed values < –0.99.
- **False Positives (FP):** Predicted values ≥ –0.99 but observed values < –0.99.
- **False Negatives (FN):** Predicted values < –0.99 but observed values ≥ –0.99.

From these classifications, the following metrics were calculated:

$$TPR = \frac{TP}{TP + FN} \qquad [3]$$

$$FPR = \frac{FP}{FP + TN} \qquad [4]$$

$$SPC = \frac{TN}{TN + FP} \qquad [5]$$

According to Table 1, the threshold of –0.99 represents the lower limit of normal SPI conditions. This value was selected because, within the SPI classification system (McKee *et al.,* 1993; WMO, 2012), values between –0.99 and +0.99 are generally considered near-normal, whereas values ≤ –1.00 denote the onset of mild drought. It is also supported by operational studies in tropical regions such as Ortega-Gaucín *et al.,* (2021), where this value optimized early detection of agricultural drought without increasing false positives. Alternative thresholds (–1.0 and –1.5) were tested, with the former reducing sensitivity and the latter significantly increasing false alarms. Therefore, using SPI = –0.99 as the cut-off ensures that the threshold captures the transition point immediately before drought conditions are classified, while avoiding the inclusion of borderline anomalies within the drought category. This criterion was adopted to construct the confusion matrix, as it provides a consistent and operationally relevant baseline for evaluating the predictive performance of each dataset.

Finally, Youden's J statistic (Youden, 1950) was computed to quantify prediction skill. The index ranges from –1 to 1, where 0 indicates no discriminative ability (equal proportions of positive and negative results for drought and wet events) and 1 indicates perfect classification with no false positives or false negatives.

$$J = sensitivity + specificity - 1 \qquad [6]$$

The expanded formula is:

$$J = \frac{true\ positives}{true\ positive + false\ negatives} + \frac{true\ negatives}{true\ negatives + false\ positives} - 1 \qquad [7]$$

Table 1: SPI drought classification categories (McKee *et al.,* 1993)

SPI	Category
2.0 +	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2.0 and less	Extremely dry

Evaluating ROC analysis

Further, to demonstrate the effectiveness of the ROC analysis, a comparison of the results was carried out with other statistical analyses that have been used by several authors to measure the accuracy of a prediction model (Madrigal *et al.,* 2018; Kheyruri *et al.,* 2023). The methods selected for the study are: the Nash-Sutcliffe efficiency (NSE) and the Modified Kling-Gupta Efficiency (KGEM) which are indices used to quantify the extent to which datasets capture real-world observations.

Table 2: NSE and KGEM indicators for drought prediction using CARiDRO datasets - Georgetown station.

Datasets	3-months timescale		12-months timescale	
	NSE	KGEM	NSE	KGEM
echam5	-1.37	-2.71	-2.06	-0.53
RCP 8.5	-1.28	-6.72	-1.61	-6.81
RCP 2.6	-0.74	-1.79	-0.86	-2.19
RCP 4.5	-0.88	-7.16	-0.51	-5.69
aexsk	-1.49	-9.02	-2.75	-8.07
aexsc	-1.40	-2.20	-1.85	-1.61
aexsa	-0.37	-2.05	-0.24	-3.37
ENSMEAN	-2.73	-15.15	-4.43	-15.89
aexsl	-3.19	-15.85	-4.01	-17.38
aexsm	-4.07	-16.82	-5.79	-16.57

Table 3: NSE and KGEM indicators for drought prediction using CARiDRO datasets - Timehri station

Datasets	3-months timescale		12-months timescale	
	NSE	KGEM	NSE	KGEM
echam5	-0.80	-1.15	-0.58	-0.95
RCP 8.5	-0.96	-5.84	-0.92	-0.59
RCP 2.6	-0.41	-5.52	-0.61	-0.07
RCP 4.5	-0.95	-9.31	-0.15	-0.39
aexsk	-0.81	-13.16	-0.61	-0.86
aexsc	-1.04	-0.61	-1.20	-0.26
aexsa	-0.06	-5.72	0.17	-1.45
ENSMEAN	-1.68	-22.13	-1.12	-2.38
aexsl	-2.84	-25.76	-1.44	-2.77
aexsm	-2.91	-22.51	-2.19	-2.43

The Nash-Sutcliffe efficiency (Eq. 8) is a metric that determines the relative magnitude of variance of the modelled data relative to the observed variance (Nash & Sutcliffe, 1970).

$$NSE = 1 - \left[\frac{\sum_{n=1}^{N}(O_n - F_n)^2}{\sum_{n=1}^{N}(O_n - \overline{O_n})^2}\right] \qquad [8]$$
,

where N is the total number of time-steps, O_n is the forecasted value at time-step n, Fn is the observed value at time-step n, and (O_n) is the mean of the observed values. The Modified Kling-Gupta Efficiency (Eq. 9), as well as the NSE, has a range of -∞ to 1 and its optimal value is 1.

$$KGEM = 1 - \sqrt{(r-1)^2 + (\beta+1)^2 + (\gamma-1)^2} \qquad [9]$$,

where r is the correlation coefficient between forecasted (F) and observed (O) data, β is the ratio of means (μF/μO) and γ is the ratio of coefficients of variation (CVF/CVO). The ideal value of each of the three components is 1 (Gupta *et al.,* 2009; Kling *et al.,* 2012).

After the selection of the datasets to make the prediction, the data were processed with two timescales. The first was the 3 months (SPI3) to assess the impact of drought on agriculture, and the second was the 12 months (SPI12) to assess the influence of drought on water resources. A time period of 62 years (1961-2023) was considered to obtain improved precision in the prediction (2024-2034). Moreover, as it relates to the selected period analyzed, Guttman (1994, 1999) reported that, for time scales from 1 to 24 months, it is necessary to have 50 to 60 years of data to achieve greater statistical confidence. A threshold of SPI = -0.99 was established to delimit the areas of the prediction period where drought events occur.

Data analysis and workflow

The following workflow ensures that dataset selection is based not only on statistical agreement but also on event detection capability, aligning with the study's objective to improve operational drought prediction in Guyana.

1. Extract monthly precipitation data for Georgetown and Timehri stations from each dataset.
2. Calculate SPI3 and SPI12 for each dataset.
3. Define drought events based on SPI thresholds (SPI ≤ –0.99).
4. Compute NSE and KGEM scores comparing each dataset to the reference station data.
5. Perform ROC analysis for each dataset against observed drought events.
6. Calculate Youden's Index (J) for each ROC curve to identify the optimal classification threshold.
7. Rank datasets separately for J values.
8. Compare rankings to assess consistency and identify datasets most suitable for operational use.

RESULTS AND DISCUSSION

Tables 2 and 3 present the results of the NSE and KGEM statistical indicators for the Georgetown and Timehri stations, respectively. In most forecasts, both indicators yielded negative values, with the notable exception of the *aexsa* dataset at the 12-month timescale, which achieved the highest NSE value among all datasets (Table 3). While some datasets exhibited values approaching 1.0, suggesting good statistical agreement, these results were not consistent across timescales. For example, "*aexsa"* produced a near-optimal NSE for the 3-month timescale but lower performance under KGEM, highlighting discrepancies between traditional efficiency metrics.

This divergence between NSE and KGEM values aligns with the findings of Madrigal *et al.,* (2018), who reported that predictive quality can appear satisfactory under one metric yet fail to capture event-level accuracy. Such inconsistencies emphasize the need for classification-based metrics, especially for agricultural drought contexts where the timely detection of dry spells can directly inform irrigation scheduling and crop protection strategies (Paredes-Trejo & Barbosa, 2017; Ortega-Gaucin *et al.,* 2018). The results obtained are not unique to the present study. Similar findings have been reported in Mediterranean cropping systems (Venturi *et al.,* 2024) and in semi-arid regions of Latin America (Paredes-Trejo & Barbosa, 2017), where NSE and KGEM produced contradictory results across timescales. These cases confirm that while error-based

Table 4: Confusion matrix results from ROC analysis for SPI3 and SPI12 - Georgetown station.

Datasets	For SPI3				For SPI12			
	TP	FP	FN	TN	TP	FP	FN	TN
echam5	78	21	16	5	77	21	12	10
aexsa	88	25	6	1	88	31	1	0
aexsk	68	13	26	13	62	11	27	20
ENSMEAN	53	11	41	15	42	10	47	21
RCP 8.5	76	21	21	5	63	21	26	10
aexsc	74	23	20	3	60	30	29	1
aexsl	55	12	39	14	43	6	46	25
aexsm	48	15	46	11	42	19	47	12
RCP 2.6	82	25	12	1	76	27	13	4
RCP 4.5	70	8	24	18	80	13	9	18

Legend: TP – true positives; FP – false positives; FN – false negatives and TN – true negatives.

Table 5: Confusion matrix results from ROC analysis for SPI3 and SPI12 - Timehri station

Datasets	For SPI3				For SPI12			
	TP	FP	FN	TN	TP	FP	FN	TN
echam5	76	22	19	3	83	17	12	8
aexsa	89	21	6	4	94	22	1	3
aexsk	66	12	29	13	60	7	35	18
ENSMEAN	51	9	44	16	47	7	54	18
RCP 8.5	68	22	27	3	64	21	31	4
aexsc	75	24	20	1	69	25	26	0
aexsl	49	13	46	12	37	4	58	21
aexsm	54	11	41	14	47	13	48	12
RCP 2.6	80	24	15	1	76	20	19	5
RCP 4.5	70	17	25	8	72	11	23	14

Legend: TP – true positives; FP – false positives; FN – false negatives and TN – true negatives.

metrics may reflect overall variance reduction, they fail to capture threshold-dependent dynamics critical for agricultural drought decision-making. ROC-based approaches, by contrast, directly address these operational needs.

Similarly, Kheyruri *et al.,* (2023) developed a model to forecast agricultural drought, employing both the NSE and RMSE indicators to measure prediction performance, and obtained satisfactory outcomes.

Although the primary objective was to apply classification metrics such as ROC and Youden's Index, NSE and KGEM were included as references because they remain widely used in hydrological and climate validation. Comparing them with ROC highlights the limitations of continuous-value metrics in discriminating drought events, thus providing a necessary methodological contrast to justify the shift in approach.

Tables 4 and 5 summarize the confusion matrix results from the ROC analysis for both timescales. For SPI3 at Georgetown (Table 4), *aexsk* achieved 68 true positives (TP) and 13 true negatives (TN), while *RCP 4.5* produced 70 TP and 18 TN, both showing balanced detection of drought and non-drought events. At Timehri (Table 5), *aexsk* maintained a favorable balance—66 TP and 13 TN for SPI3, and 60 TP and 18 TN for SPI12—outperforming most other datasets. These patterns suggest that *aexsk* has a stable predictive capacity across varying agro-climatic conditions.

The agricultural implications of this balanced performance are significant. In operational drought monitoring for crops such as rice, sugarcane, and cassava in Guyana's coastal and inland zones, minimizing false positives (which could trigger unnecessary irrigation or early harvesting) is as important as maximizing true positives. Latin American studies, such as those by Paredes & Guevara*,* (2010) in Venezuela and Centella *et al.,* (2014) in the Caribbean, have similarly stressed the operational importance of dataset selection for irrigation system management.

Tables 6 and 7 summarize the ROC-derived performance metrics for SPI3 and SPI12 at the Georgetown and Timehri stations. The results showed that Georgetown, RCP 4.5 obtained the highest Youden's J values for both SPI3 (0.44) and SPI12 (0.48), followed by aexsk with 0.22 and 0.34, respectively. At Timehri, aexsk achieved the highest J for both SPI12 (0.35) and SPI3 (0.21), while RCP 4.5 maintained strong results only for SPI12 (J = 0.32).

These rankings were consistent across the two stations, indicating that the combination of ROC analysis and Youden's Index can provide a stable and reliable framework for dataset selection unlike the variability previously observed with NSE and KGEM metrics. Similar results have been documented in other studies in the Latin American Region. Research conducted in Mexico by Ortega-Gaucin *et al.,* (2018) reported that drought monitoring based on the SPI, when evaluated using classification metrics, significantly improved seasonal crop yield forecasts. Research by Paredes-Trejo

Table 6: ROC-derived performance metrics for SPI3 and SPI12 - Georgetown station

Datasets	For SPI3					For SPI12				
	TPR	FPR	SPC	J	Rank	TPR	FPR	SPC	J	Rank
echam5	0.83	0.81	0.19	0.02	6	0.87	0.68	0.32	0.19	4
aexsa	0.94	0.96	0.04	-0.03	7	0.99	1.00	0.00	-0.01	6
aexsk	0.72	0.50	0.50	0.22	2	0.70	0.35	0.65	0.34	2
ENSMEAN	0.56	0.42	0.58	0.14	3	0.47	0.32	0.68	0.15	5
RCP 8.5	0.78	0.81	0.19	-0.02	5	0.71	0.68	0.32	0.03	3
aexsc	0.79	0.88	0.12	-0.10	9	0.67	0.97	0.03	-0.29	10
aexsl	0.59	0.46	0.54	0.12	4	0.48	0.19	0.81	0.29	8
aexsm	0.51	0.58	0.42	-0.07	10	0.47	0.61	0.39	-0.14	9
RCP 2.6	0.87	0.96	0.04	-0.09	8	0.85	0.87	0.13	-0.02	7
RCP 4.5	0.74	0.31	0.69	0.44	1	0.90	0.42	0.58	0.48	1

Legend: TPR – true positives rate; FPR – false positives rate; SPC – specificity or true negatives rate and J – Youden Index

Table 7: ROC-derived performance metrics for SPI3 and SPI12 — Timehri station

Datasets	For SPI3					For SPI12				
	TPR	FPR	SPC	J	Rank	TPR	FPR	SPC	J	Rank
echam5	0.80	0.88	0.12	-0.08	7	0.87	0.68	0.32	0.19	4
aexsa	0.94	0.84	0.16	0.10	4	0.99	0.88	0.12	0.11	6
aexsk	0.69	0.48	0.52	0.21	1	0.63	0.28	0.72	0.35	1
ENSMEAN	0.54	0.36	0.64	0.18	2	0.43	0.28	0.72	0.15	5
RCP 8.5	0.72	0.88	0.12	-0.16	9	0.67	0.84	0.16	-0.17	9
aexsc	0.79	0.96	0.04	-0.17	10	0.73	1.00	0.00	-0.27	3
aexsl	0.52	0.52	0.48	0.00	6	0.39	0.16	0.84	0.23	10
aexsm	0.57	0.44	0.56	0.13	3	0.49	0.52	0.48	-0.03	8
RCP 2.6	0.84	0.96	0.04	-0.12	8	0.80	0.80	0.20	0.00	7
RCP 4.5	0.74	0.68	0.32	0.06	5	0.76	0.44	0.56	0.32	2

Legend: TPR – true positives rate; FPR – false positives rate; SPC – specificity or true negatives rate and J – Youden Index

& Barbosa, (2017) showed that ROC analysis could be used as a tool to guide the optimal choice of datasets for agricultural drought indices in northern South America. These findings strengthen the methodological robustness of the present study for the Caribbean–South American agro-ecological context.

The results obtained also show practical relevance for agricultural management in Guyana and other countries in the Caribbean, where crops such as rice, sugarcane, and cassava heavily depend on efficient water allocation. The selection of datasets with high discriminative capacity, such as aexsk and RCP 4.5, makes it possible to optimize irrigation schedules, reduce yield losses from water stress, and anticipate adjustments in drainage operations in coastal areas. Studies completed by Venturi *et al.,* (2024) and Ortega-Gaucín *et al.,* (2021) have demonstrated that integrating classification metrics into dataset selection improves the resilience of various production systems against climate variability.

Regarding the unsatisfactory results obtained by Madrigal *et al.,* (2018) with ROC analysis, it could be that good predictions were not obtained, even when the KGEM and NSE statistical analyses were satisfactory. This could be due to the cyclical nature of drought.On the other hand, regarding the results of Kheyruri *et al.,* (2023), there could have been a risk of the same thing happening, since the statistical indicators used do not discriminate between wet and dry conditions of drought.

Other authors used correlation as a statistical tool to demonstrate the veracity of a new meteorological drought index based on fuzzy logic (FL) and neuro-fuzzy models to describe and predict droughts (Oyounalsoud *et al.,* 2023), and concluded that, overall, those developed soft computing drought indices outperformed conventional methods. Bhardwaj *et al.,* (2022) carried out an evaluation study of three different types of datasets according to their source: by satellite, from satellite observations, modelling, and in situ observations. Said authors used the indicators Mean systematic difference (MBE), Mean absolute difference (MAE), Spread or standard deviation of errors (RMSE), and a statistical metric that measures the strength and direction of correlation (R). However, they obtained mixed results, raising several questions about the inclusion of satellite SM (soil moisture) for proactive drought monitoring in Australia and the Pacific.

Vicente-Serrano *et al.,* (2022), conducted correlation analyses to measure the accuracy of a global drought dataset and a monitoring system based on the Standardized Precipitation Evapotranspiration Index (SPEI). However, these authors raise the need to carry out research to specify changes in drought to atmospheric evaporative demand (AED) in the ERA5 SPEI. Ceglar *et al.,* (2012) also conducted a correlation study to analyze the ability of the datasets to identify drought impacts on different drought-prone systems. But it found that some datasets failed to realistically detect the spatial patterns of specific drought episodes.

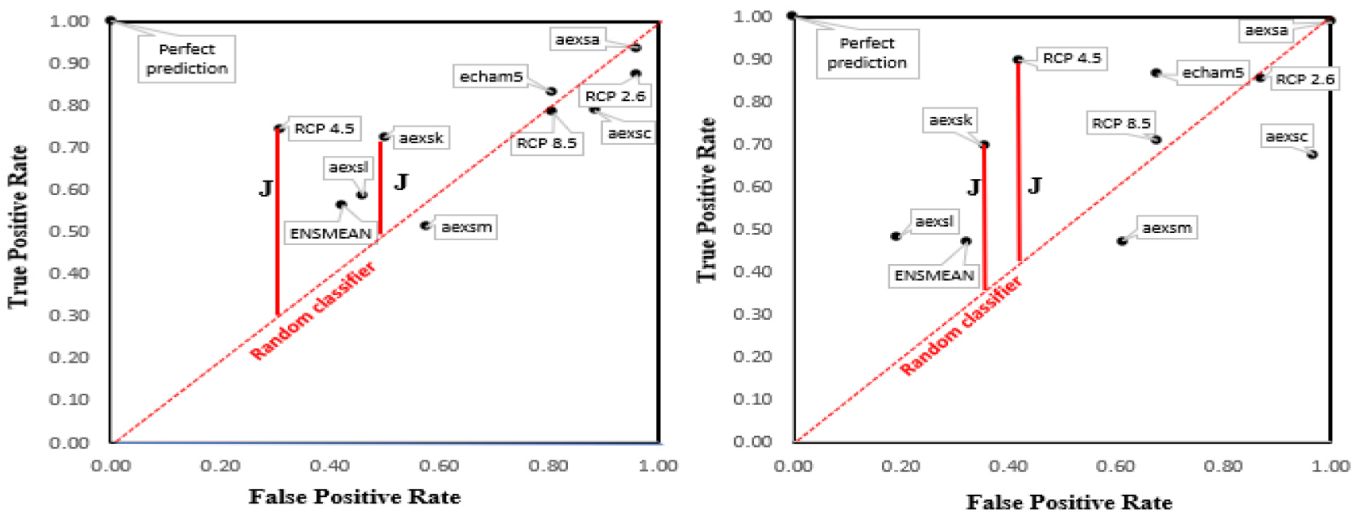

Fig. 1: ROC space showing predictive performance of datasets for SPI3 and SPI12 — Georgetown station.

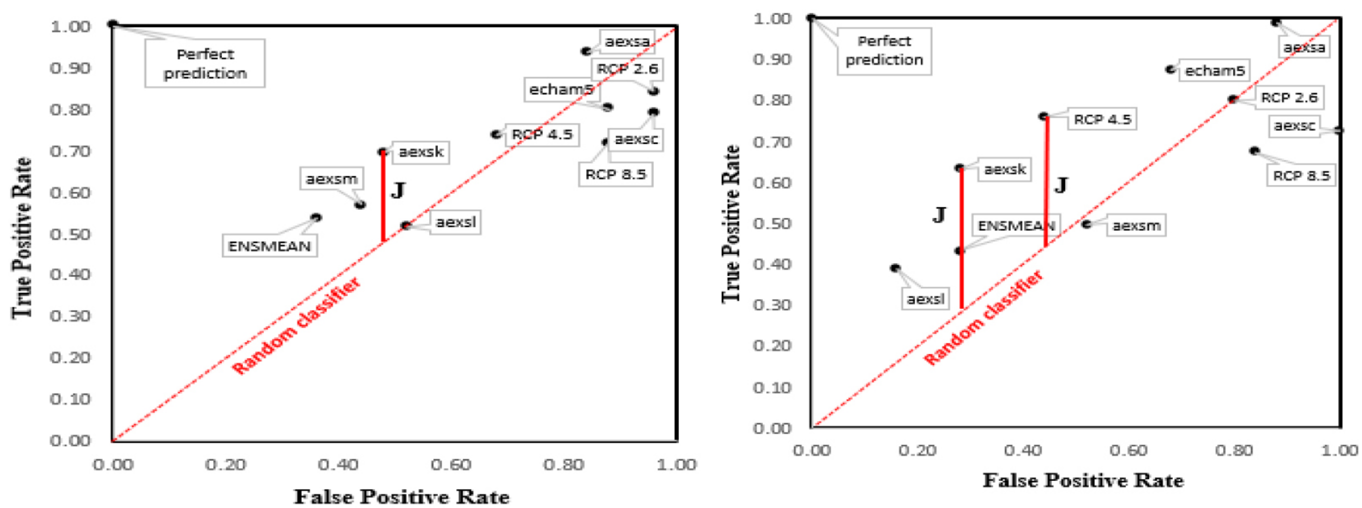

Fig. 2: ROC space showing predictive performance of datasets for SPI3 and SPI12 — Timehri station.

Fig. 1 and 2 illustrate the ROC space for Georgetown and Timehri, respectively. The *aexsk* and *RCP 4.5* datasets are plotted closest to the perfect prediction point (TPR = 1, FPR = 0), visually confirming their superior classification ability. The diagonal red line represents the random classifier, serving as a baseline for comparison. Datasets located farther above this line indicate stronger discriminative ability. Axis labels have been enlarged for readability, and legends were adjusted to highlight the top-performing datasets (aexsk and RCP 4.5). From an operational standpoint, such graphical diagnostics are crucial. They enable agricultural planners to visualize which datasets consistently capture drought onset without excessive false alarms. This is particularly valuable in Guyana's irrigation districts, where misclassification could either waste scarce water resources or leave crops vulnerable to stress. Comparable visualization-based selection approaches have been recommended by Venturi *et al.,* (2024) for European drought monitoring and by Paredes-Trejo & Barbosa, (2017) for Venezuelan agriculture.

The ROC analysis and the Youden index allow quantifying the efficiency of the prediction model for both positive and negative values of the SPI index, and both types of values on one side or the other of the threshold (-0.99) have the same importance. Therefore, this statistical method is adequate to measure the performance of the prediction of drought or wet events, but not the severity of these events. Up to this point it can be deduced that the dataset 'aexsk' consistently performed well. Therefore, all this led to the generation of the future prediction for the period 2024-2034, using the aforementioned data set.

The future forecast (Table 8) based on *aexsk* suggests that for 2024–2034, normal SPI conditions will dominate; however, extreme drought will occur in 8.5% of months for SPI3 and 18.3% of months for SPI12. Four discrete drought events are projected for SPI3, concentrated in mid-2026 and 2033–2034, whereas SPI12 shows no events exceeding 12 consecutive months.

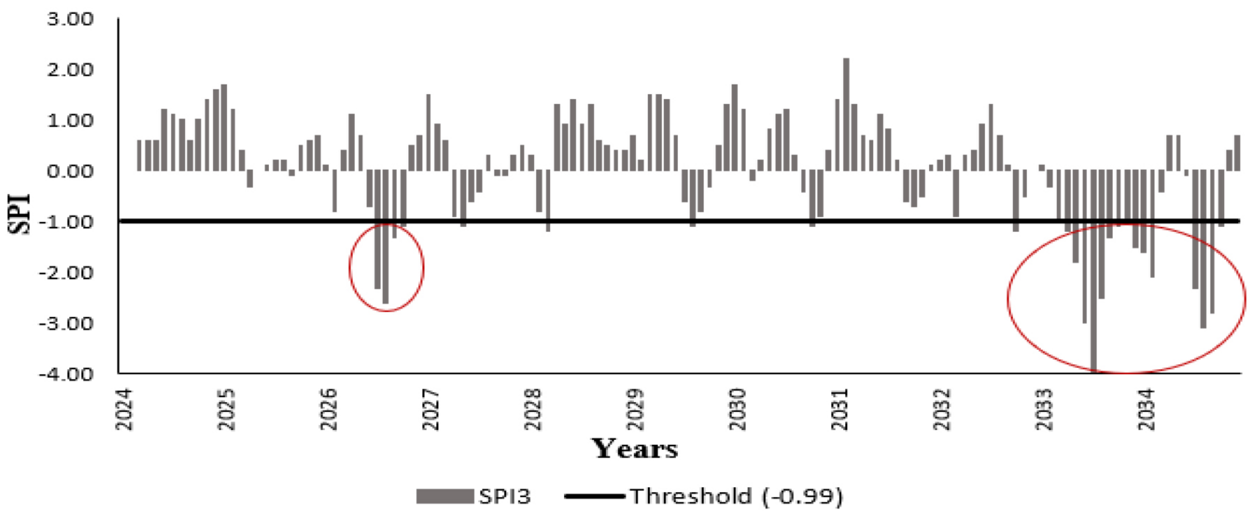

Fig. 3: Predicted SPI3 time series (2024–2034) using the 'aexsk' dataset

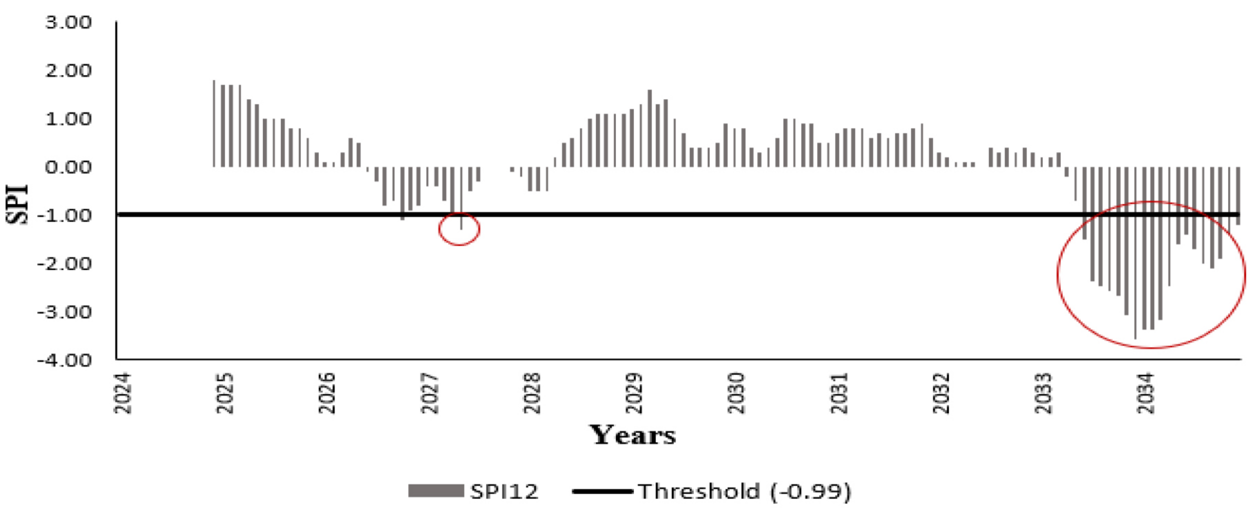

Fig. 4: Predicted SPI12 time series (2024–2034) using the 'aexsk' dataset

For farmers and policymakers, these temporal patterns carry clear implications: SPI3 events may signal the need for short-term irrigation adjustments and supplemental feeding of livestock and SPI12 events may reflect long-term water storage and allocation concerns, potentially requiring adjustments to national drought contingency plans.

In Latin America, similar dual-timescale approaches have been successfully applied, such as in Paredes & Guevara*,* (2010) for Venezuela's Llanos region and Centella *et al.,* (2017) for the Caribbean, to ensure that both immediate and strategic responses are informed by the most reliable climate datasets.

Fig. 3 and 4 describe the behavior of the SPI index through the ten years of prediction. As it relates to Fig. 3 (for a timescale of 3 months), the four drought events predicted by the CARiDRO tool are distributed in the middle of 2026 and then from 2033 until almost the end of 2034. In the case of Fig. 4, isolated drought events are observed in early 2027 and between 2033 and 2034. But none of those events are projected to occur for a period of 12 consecutive months or more.

It should be noted that drought indices, in general sense, indicate dry and wet conditions. On the other hand, the Manual of Drought Indicators and Indices (WMO, 2016) compiled a group of the most used drought indices and indicators. Some of these indices make drought predictions and, in different ways, establish classifications based on their severity. Taking this into account, the statistical methods used in these studies do not take into account this aspect, that is, the discrimination of drought categories. It is the case of tables 3 and 4, where there is no uniformity in the expected result, that is, a value closer to (1) for the same dataset in the two timescales". So, both the NSE and KGEM indicators do not explain the behavior of the SPI index for both dry and wet conditions. However, ROC analysis provides a valuable framework for interpreting predictions based on classifications generated

Table 8: Forecasted drought statistics for 2024–2034 using the selected 'aexsk' dataset.

Model Dataset	SPI Categories Statistics				General Drought Statistics			
	Normal (%)	Moderate drought (%)	Severe drought (%)	Extreme drought (%)	Months below threshold	Percentage (%)	Maximum consecutive	No. Drought events
For Threshold: -0.99, Time-scale: 3 months and Months drought: 3								
aexsk	62	10.9	1.6	8.5	25	19.4	12	4
For Threshold: -0.99, Time-scale: 12 months and Months drought: 12								
aexsk	69.2	5	4.2	18.3	22	18.3	2	0

by a given drought index. It enables the evaluation of predictive performance under both wet and dry conditions—even when these events are infrequent within the analyzed time period. In this sense, ROC analysis captures the nature and dynamics of drought with greater analytical precision. In this study, drought predictions were generated using multiple climate datasets, including "echam5," "RCP 2.6," "aexsa," "aexsc," and "RCP 4.5," and their performance was initially assessed using traditional statistical indicators such as NSE and KGEM. Nonetheless, when evaluated through the ROC analysis through the Area Under the Curve (AUC) analysis only the dataset "aexsk" demonstrated superior discriminatory power in distinguishing drought from non-drought conditions.

Numerous studies have primarily applied ROC analysis to compare drought indices or prediction algorithms (e.g., Hao & AghaKouchak, 2013; Naumann *et al.,* 2014), however, this research introduces a novel application of ROC metrics for evaluating the predictive performance of models built using different input datasets. With the application of this approach a robust, data-driven basis for selecting the most suitable dataset for operational drought forecasting was established. The methodological contribution enhances both model evaluation and dataset selection in primarily in agricultural drought studies.

CONCLUSIONS

This study introduces a novel methodological application of ROC analysis, combined with Youden's Index, for evaluating and selecting climate datasets for drought prediction. Further, unlike conventional approaches that rely solely on statistical efficiency indicators, this method emphasizes the discriminatory capacity of predictions between drought and non-drought conditions, directly addressing the needs of agricultural decision-making.

The results show that ROC based evaluation particularly through Youden's Index and ROC space provides a more nuanced and robust assessment of model performance. Added to this, while traditional metrics such as NSE and KGEM are useful for quantifying prediction error, they fail to consistently capture the balance between true positive and true negative classifications in the context of drought forecasting. Among the ten datasets evaluated with the CARiDRO modeling tool, the *aexsk* dataset consistently outperformed all others across both stations (Georgetown and Timehri) and timescales (SPI3 and SPI12), achieving optimal values in both ROC metrics and traditional indices. It was therefore selected for generating the final forecast for the period 2024–2034.

These findings confirm that conventional statistical performance metrics are not always reliable for dataset selection, as they may produce varying rankings across timescales. In contrast, the ROC–Youden approach proved methodologically consistent, ensuring dataset selection is based on robust classification performance rather than variable statistical fit. This operational consistency enhances the reliability of drought predictions and supports more informed resource allocation in agricultural systems.

Author's certificate: The manuscript or its part is not under consideration for publication elsewhere and the same has been approved by all co-authors.

Conflict of Interests: The authors declare that there is no conflict of interest regarding the publication of this article.

Funding: The authors received no financial support for the research.

Data Availability: The rainfall data used in the study were accessed through the Ministry of Agriculture Hydrometeorological Service of Guyana, solely for the purpose of this research.

Author's contribution: **Esequiel Rolando Jiménez Espinosa:** Conceptualization, Methodology, Data Analysis, Writing and Editing; **Donessa Lucinda David:** Data Processing, Data Analysis, Writing and Editing.

REFERENCES

Beharry, S.L., Gabriels, D., Lobo, D., & Clarke, R.M., (2019). A 35-year meteorological drought analysis in the Caribbean Region: case study of the small island state of Trinidad and Tobago. *SN Applied Sciences,* 1. https://doi.org/10.1007/s42452-019-1296-4.

Bhardwaj, J., Kuleshov, Y., Chua, Z.W., Watkins, A.B., Choy, S., & Sun, Q. (2022). Evaluating satellite soil moisture datasets for drought monitoring in Australia and the South-West Pacific. *Remote Sensing*, 14, 3971. https://doi.org/10.3390/rs14163971

Caribbean Climate Information Education and Tools (CLIE'nT). (n.d.). CARiDRO: Caribbean Assessment of Regional Drought. Available from: https://theclient.app/caridro

Ceglar, A., Moran-Tejeda, E., & Vicente-Serrano, S.M. (2012).

Assessment of multi-scale drought datasets to quantify drought severity and impacts in agriculture: A case study for Slovenia. *International Journal of Spatial Data Infrastructures Research*, 7, 464–487. https://doi.org/10.2902/1725-0463.2012.07.art21

Centella, A., Bezanilla, A., Vichot, A., & Silva, M. (2014). *User's guide CARiDRO: An online tool for drought assessment*. INSMET, Cuba.

Centella, A., Bezanilla, A., Vichot, A., & Silva, M. (2017). CARiDRO: The Caribbean Assessment Regional Drought Tool. Case Study Report 1 of the CARIWIG Project. Institute of Meteorology, Cuba. Available from: https://cdkn.org/resource/case-study-caridro-caribbean-assessment-regional-drought-tool?

Farrell, D., Trotman, A. & Cox, C. (2010): Drought Impacts and Early Warning in The Caribbean, The Drought of 2009-2010. *Global Assessment Risk Reduction,* 22pp.

Gupta, H.V., Kling, H., Yilmaz, K.K., & Martinez, G.F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Guttman, N.B. (1994). On the sensitivity of sample L moments to sample size. *Journal of Climate*, 7(6), 1026–1029.

Guttman, N.B. (1999). Accepting the standardized precipitation index: A calculation algorithm. *Journal of the American Water Resources Association*, 35(2), 311–322.

Hao, Z. & AghaKouchak, A. (2013). Multivariate Standardized Drought Index: A parametric multi-index model. *Advances in Water Resources,* 57: 12–18. https://doi.org/10.1016/j.advwatres.2013.03.009

Hayes, M., Svoboda, M., Wall, N., & Widhalm, M. (2011). The Lincoln Declaration on Drought Indices: Universal Meteorological Drought Index Recommended. *Bulletin of the American Meteorological Society*, *92*(4), 485-488. https://doi.org/10.1175/2010BAMS3103.1

IPCC, (2013). *Climate change 2013: The physical science basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, USA.

IPCC, (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. P´ean, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2391 pp, https://doi.org/10.1017/9781009157896.

Junge, M.R., & Dettori, J.R. (2024). ROC solid: Receiver operator characteristic (ROC) curves as a foundation for better diagnostic tests. *Global Spine Journal*, 8(4), 424–429. https://doi.org/10.1177/2192568218778294

Kheyruri, Y., Sharafati, A., & Neshat, A. (2023). Predicting agricultural drought using meteorological and ENSO parameters in different regions of Iran based on the LSTM model. *Stochastic Environmental Research and Risk Assessment*, https://doi.org/10.1007/s00477-023-02465-6

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Madrigal, J., Solera, A., Suárez-Almiñana, S., Paredes-Arquiola, J., Andreu, J., & Sánchez-Quispe, S.T. (2018). Skill assessment of a seasonal forecast model to predict drought events for water resource systems. *Journal of Hydrology,* 564, 574–587. https://doi.org/10.1016/j.jhydrol.2018.07.046

McKee, T.B., Doesken, N.J., & Kleist, J. (1993). Relationship of drought frequency and duration to time scales. *Eighth Conference on Applied Climatology*, 17–22 January 1993, Anaheim, CA, USA. Available from: www.droughtmanagement.info/literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf

Miller, P.W., & Ramseyer, C.A. (2020). Did the Climate Forecast System Anticipate the 2015 Caribbean Drought? *Journal of Hydrometeorology,* 21(6), 1245-1258. https://doi.org/10.1175/JHM-D-19-0284.1

Mishra, A.K. & Singh, V.P. (2010) A Review of Drought Concepts. Journal of Hydrology, 391, 202-216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Murphy, A.H. (1996). The Finley affair: A signal event in the history of forecast verification. *Weather and Forecasting*, 11(1), 3–20. https://doi.org/10.1175/1520-0434(1996)011<0003:tfaase>2.0.co;2

Nandgude, N., Singh, T. P., Nandgude, S., & Tiwari, M. (2023). Drought Prediction: A Comprehensive Review of Different Drought Prediction Models and Adopted Technologies. *Sustainability,* 15(15), 11684. https://doi.org/10.3390/su151511684

Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models, part I - A discussion of principles. *Journal of Hydrology,* 10(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

National Drought Mitigation Center. (2018). SPI Generator [software]. University of Nebraska-Lincoln. Available from:https://drought.unl.edu/Monitoring/SPI/SPIProgram.aspx

Naumann, G., Barbosa, P., Garrote, L., Iglesias, A., & Vogt, J. (2014). Exploring drought vulnerability in Africa: An indicator-based analysis to be used in early warning systems. *Hydrology and Earth System Sciences*, 18, 1591–1604. https://doi.org/10.5194/hess-18-1591-2014

Orimoloye, I.R., Belle, J.A., Orimoloye, Y.M., Olusola, A.O., & Ololade, O.O. (2022). Drought: A common environmental disaster. *Atmosphere,* 13, 111. https://doi.org/10.3390/atmos13010111

Ortega-Gaucin, D., De la Cruz Bartolón, J., & Castellano Bahena, H. V. (2018). Drought Vulnerability Indices in Mexico. *Water,* 10(11), 1671. https://doi.org/10.3390/w10111671

Oyounalsoud, M.S., Abdallah, M., Yilmaz, A.G., Siddique, M., & Atabay, S. (2023). A new meteorological drought index based on fuzzy logic: Development and comparative assessment with conventional drought indices. *Journal of Hydrology*, 619, 129306. https://doi.org/10.1016/j.jhydrol.2023.129306

Paredes, F., & Guevara, E. (2010). Desarrollo y evaluación de un modelo para predecir sequías meteorológicas en los Llanos de Venezuela. *Bioagro*, 22(1), 3–10.

Paredes-Trejo, F., & Barbosa, H. (2017). Evaluation of the SMOS-derived soil water deficit index as agricultural drought index in Northeast of Brazil. *Water*, 9(6), 377. https://doi.org/10.3390/w9060377

Peres, D.J., & Cancelliere, A. (2014). Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. *Hydrology and Earth System Sciences*, 18(12), 4913–4931. https://doi.org/10.5194/hess-18-4913-2014

Peres, D.J., Iuppa, C., Cavallaro, L., Cancelliere, A., & Foti, E. (2015). Significant wave height record extension by neural networks and reanalysis wind data. *Ocean Modelling,* 94, 128–140. https://doi.org/10.1016/j.ocemod.2015.08.002

Trotman A., Joyette A., Van Meerbeeck C., Mahon R., Cox S., Cave N., and Farrell D. (2018). Drought Risk Management in the Caribbean Community: Early Warning Information and Other Risk Reduction Considerations. In: Wilhite, D. and Pulwarty, R. (Eds.) (2018). Drought and Water Crises. *Boca Raton: CRC Press,* https://doi.org/10.1201/b22009

Venturi, S., Dunea, D., Mateescu, E., Virsta, A., Petrescu, N., & Casadei, S. (2024). Drought monitoring in southeastern Romania based on the comparison and correlation of SPEI and SPI indices. *Scientific Papers. Series E. Land Reclamation, Earth Observation & Surveying*,Environmental Engineering. Vol. XIII, 2024 Print ISSN 2285-6064, CD-ROM ISSN 2285-6072, Online ISSN 2393-5138, ISSN-L 2285-6064.]

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696-1718. https://doi.org/10.1175/2009JCLI2909.1

Vicente-Serrano, S.M., Domínguez-Castro, F., Reig, F., Tomás-Burguera, M., Peña-Angulo, D., & Latorre, B. (2022). A global drought monitoring system and dataset based on ERA5 reanalysis: A focus on crop-growing regions. *Geoscience Data Journal*, 00, 1–14. https://doi.org/10.1002/gdj3.178

Wilhite, D.A. & Pulwarty, R. (Eds) (2018) Drought and Water Crises: Integrating Science, Management and Policy, 2nd edn. CRC Press, Boca Raton, FL, USA.

Wilhite, D.A., & Glantz, M. (1985). Understanding the drought phenomenon: The role of definitions. *Water International*, 10, 111–120.

World Meteorological Organization. (2012). *Standardized precipitation index: User guide*. WMO-No. 1090. ISBN 978-92-63-11091-6. Available from: https://library.wmo.int/doc_num.php?explnum_id=7768

World Meteorological Organization & Global Water Partnership. (2016). Handbook of drought indicators and indices (M. Svoboda & B. A. Fuchs). Integrated Drought Management Programme, Integrated Drought Management Tools and Guidelines Series 2, WMO-No. 1173. Geneva, Switzerland, and Stockholm, Sweden. ISBN 978-92-63-11173-9

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer*, 3, 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3