Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.30
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

THRESHOLD SPECTRAL-BASED FOURIER REGRESSION MODEL FOR SEASONAL-CYCLICAL TIME SERIES

*[1]Taylor Jimoh I., [2]Taiwo Abass I. and [3]Olatayo Timothy O.

[1]Department of Mathematical Sciences, Lagos State University of Science and Technology, Ikorodu, Lagos, Nigeria
[2,3]Department of Statistics, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria

*Corresponding Author Email Address: jimohtaylor119@gmail.com

ABSTRACT

The Threshold Spectral Fourier Regression (TSFR) model is a hybrid model that integrates Fourier methods with the nonlinear dynamics of threshold models. It was developed to simultaneously model a series that exhibits nonlinear, nonstationary, cyclical, and seasonal behaviours, as well as structural breaks. This study presents the theory and development of the TSFR model. The model leverages spectral estimation and converts the dataset from the time domain to the frequency domain using methods such as the Fourier Transform. A threshold value of $\tau = 0.85$ is determined using the quantile method; to split the series into two regimes categorised as Low and High, frequencies exceeding the threshold are retained for inclusion in the model. The TSFR model is developed using the retained frequencies and Fourier terms as regressors, while model parameters are estimated using the Ordinary Least Squares (OLS) method. The fitted model is assessed for goodness-of-fit and model accuracy; the residuals are examined to ensure they are randomly distributed, and are used to predict future values. Nigerian rainfall data, covering the period 1981- May, 2025, obtained from climate hazards, UC Santa Barbara WEP; made up of 533 data points $(N = 533)$ are used for empirical analysis. The estimated parameters for the rainfall dataset include five values per category. Low rainfall parameters are 0.2169, 0.0045, 0.0055, 0.0013 and -0.0015 with Coefficient of Determination R^2 = 0.8818, Adjusted Coefficient of Determination $\bar{R}^2$ = 0.7486, Mean Absolute Percentage Error (MAPE) = 0.3, Akaike Information Criterion (AIC) = -6427.32 and Bayesian Information Criterion (BIC) $= -0.0002$; while the parameters for High rainfall are 0.9998, -0.00005, -0.0002, 0.00003 and -0.00006 with R^2 = 0.9484, Adjusted $\bar{R}^2$= 0.90388, MAPE = 0.05, AIC = -8832.82 and BIC = -8808. The findings in this study demonstrate that the TSFR model is suitable for modeling time series with pronounced nonlinear and periodic features. The model is flexible and robust, provides a good fit, exhibits minimal overfitting, and achieves better forecasting accuracy than traditional and non-threshold models.

Keywords: Fourier Regression, Periodogram, Threshold, Ordinary Least Squares, Spectral density

INTRODUCTION

Time series is a crucial area of statistical and econometric modeling that focuses on understanding data collected sequentially over time. A time series is a set of data points recorded at successive, equally spaced intervals (Box et al., 2016) or a sequence of measurements of the same variable recorded at equal time intervals; it is vital for understanding the temporal behaviors of variables measured at equal time intervals. Time series exhibit dependencies between past and present values (Kumar, 2013); the existing patterns in time series datasets over time make them essential for forecasting and decision-making. policy evaluation, and system control across various fields, such as economics, meteorology, and environmental science. climatology, hydrology, and finance (Hyndman & Athanasopoulos, 2021). The objectives of time series modeling are to discover the patterns or structures in the series, develop appropriate models, and generate reliable forecasts of future behaviours.

Traditional or classical models such as the Autoregressive (AR), Moving Average (MA), and Autoregressive Integrated Moving Average (ARIMA) models, which assume a linear, continuous seasonal structure and constant parameters over time, have been widely applied to model and forecast stationary and linear time series datasets (Wei, 2019; Jones, 2018).

Several real-world datasets, such as economic and environmental datasets including rainfall patterns and inflation rates, exhibit nonlinear, nonstationary, cyclical, and seasonal dynamics that violate the assumptions of linearity and stationarity (Shumway & Stoffer, 2017), the complex nature of time series datasets, led to the development or formation of a more advanced techniques, such as threshold models, spectral analysis, and Fourier-based regression, to capture structural breaks, cyclic and periodic fluctuations within the series (Tong, 2011; Zhang et al., 2025). The introduction of the threshold principle allows the model to accommodate regime shifts-situations where the behavior of a process changes once a certain threshold is crossed (Tsay, 1989). Fourier regression represents time series in the form of an infinite sum of sine and cosine functions to adequately model periodic and seasonal variations (Taiwo & Olatayo, 2018; Lima et al., 2023). A Threshold Spectral Fourier Regression (TSFR) model was developed to simultaneously model nonlinear, cyclical, seasonal behaviors and structural breaks in a series. It is a hybrid model that integrates Fourier methods or analysis with the nonlinear dynamics of threshold models; it leverages spectral analysis estimation to enhance forecasting accuracy and encourages transferability and adaptability across disciplines. The TSFR model contributes to both theoretical and empirical advancements in time series modeling, giving better performance in capturing regime-dependent periodicities observed in variables, is very simple to interpret, and is robust as a tool for the analysis of nonlinear periodic processes across complex datasets.

Thresholds in time series are key values that influence how data is interpreted, modelled, or acted upon, especially when identifying regime shifts, anomalies, or decision points. It is also referred to as a specific value or level that serves as a boundary or limit for certain behaviours or decisions within a dataset.

In spectral analysis, data is transformed from the time domain to the frequency domain, and the frequency properties of the time series are analyzed (Wilfredo, 2016). Frequency domain analysis,

https://dx.doi.org/10.4314/swj.v21i2.30

or spectral analysis, is useful for identifying seasonal patterns and hidden cycles.
The threshold regression model is a nonlinear regression model that allows the relationship between the predictor (independent variable) and the outcome (dependent variable) to change at a certain point known as the threshold. The threshold parameter (sample split value) can be treated as unknown (Kourtellos *et al.*, 2011) or a linear combination of variables (Liziong, 2021).

The purpose of time series analysis is to predict future values based on previously observed time series values (Sherrod, 2015). Forecasting using a time series model involves collecting historical data, identifying patterns (trend, seasonal patterns, cycles) in the data, developing an appropriate model, assessing the model, and interpreting the results to make informed decisions.

MATERIALS AND METHODS

The materials for the formation of the Threshold Spectral Fourier Regression model are listed as follows.

Fourier Series

Fourier series can be in two forms, the trigonometric polynomial form or the exponential form. Fourier series is well-suited for data that can be modeled as a distribution of sine and cosine waves or for modeling periodic patterns (Suparti, 2018). The trigonometric Fourier series form, relevant for this study, is defined as:

$$f(x) = a_o + \sum_{1}^{n}\left(a_n cos\frac{n\tau x}{T} + b_n sin\frac{n\pi x}{T}\right) \qquad (1)$$

$$a_0 = \frac{1}{T}\int_{t_0}^{t_0+T} f(x)\,dx \qquad (2)$$

$$a_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(x)cos(n\omega x)\,dx \qquad (3)$$

$$b_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(x)sin(n\omega x)\,dx \qquad (4)$$

The function is said to be periodic if the value of the function repeats itself at regular intervals
According to Damanik (2010), a function has a period of *T* or is periodic with a period *T > 0*, if for all *x*, equation (5) holds

$$f(x+T) = f(x) \qquad (5)$$

Based on formula 5 above, it can be concluded that

$$f(x+nT) = f(x) \qquad (6)$$

A function can be an odd function or an even function or a combination of both. A function is called an odd function if it satisfies the property:

$$f(-x) = -f(x) \qquad (7)$$

And called an even function if it satisfies the property

$$f(-x) = f(x) \qquad (8)$$

To determine the Fourier coefficients $a_0, a_{n,}$ and b_n of the periodic function, even and odd functions, the following formula 6 is used:
If $f(x)$ is even, then:

$$a_0 = \frac{1}{T}\int_{t_0}^{t_0+T} f(x)\,dx$$

$$a_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(x)cos(n\omega x)\,dx \qquad (9)$$

For this case, it is said the function $f(x)$ is even and decomposes in the cosine series ($b_n = 0$).
If $f(x)$ is odd, then $a_0 = 0,\ a_n = 0$

$$b_n = \frac{2}{T}\int_{t_0}^{t_0+T} f(x)sin(n\omega x)\,dx \qquad (10)$$

For this case, it is said the function $f(x)$ is odd and decomposes in the sine series ($a_n = 0$).

Spectral Density

$$f(\omega) = \frac{\sigma^2}{2\pi}\left|\varphi\left(e^{i\omega}\right)\right|^2 \qquad (11)$$

or

$$f(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\hat{\gamma}(h)\,e^{i\omega h} \qquad (12)$$

Periodogram

$$I(\omega) = \frac{1}{2n}\left|\sum_{t=1}^{n}(y_t-\bar{y})\,e^{i\omega t}\right|^2 \qquad (13)$$

or

$$E[I(\omega)] = \frac{1}{2\pi}E\left[\sum\hat{\gamma}(h)\,e^{i\omega t}\right] \qquad (14)$$

Fourier Transform

$$Y(f) = \int_{-\infty}^{\infty} y(t)\,e^{-2\pi ift}dt \qquad (15)$$

$$H(w) = \sum_{-\infty}^{\infty} h(t)exp^{(-i\omega t)}, -\pi \le \omega \le \pi \qquad (16)$$

Thresholding

$$\hat{Y}(f) = \begin{cases} Y(f), & [X(t)] > d \\ 0, & otherwise \end{cases} \qquad (17)$$

$$T_t = \begin{cases} f_1(t) & if\ T_{t-d} \le k \\ f_2(t) & if\ T_{t-d} > k \end{cases} \qquad (18)$$

Fourier Regression model

$$Y_t = \beta_0 + \sum_{n=1}^{N}\left(a_n\cos\left(\frac{2\pi nt}{T}\right) + b_n\sin\left(\frac{2\tau nt}{T}\right)\right) + \epsilon_t \qquad (20)$$

(i) Time-Varying Threshold Model

$$y_t = \begin{cases} \beta_1' X_t + e_t, & if\ \ q_t \le \gamma_t \\ \beta_2' X_t + e_t, & if\ \ q_t > \gamma_t \end{cases} \qquad (21)$$

(ii) Threshold Autoregressive Regression

Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.30
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Model

array script phi sub 1 plus sum from i equals 1 to p of , script phi sub t $y_t =$

$$\begin{cases} \varphi_1 + \sum_{i=1}^{p} \varphi_{1i} y_{t-1} + \varepsilon_t, if q_{t-d} \le \tau \\ \varphi_2 + \sum_{i=1}^{p} \varphi_{2i} y_{t-1} + \varepsilon_t, if q_{t-d} \le \tau \end{cases} \quad (22)$$

Durbin Watson Statistic (DWS)

$$d = \frac{\sum_{t=2}^{T} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T} (\varepsilon_t)^2} \quad (23)$$

Adjusted Coefficient of Determination

$$\bar{R}^2 = \frac{1}{n-k} (nR^2 - k) \quad (24)$$

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

(i) Akaike Information Criterion (AIC)

$$AIC = -2 \log L(\hat{\theta}) + 2r \quad (25)$$

(ii) Bayesian Information Criterion (BIC) is defined as:

$$BIC = -2 \log L(\hat{\theta}) + r logn \quad (26)$$

where: $x \in [0, \frac{n\pi}{T}]$, $a_n: a_1, a_{2,\dots}$, $b_n: b_1, b_2, \dots$, a_n and $b_n$are sequences of real numbers, T: period of the seasonal cycle. $\omega: \frac{n\pi}{T}$, $Y_t: y_n = \{y_1, y_2, \dots, y_n\}$ is the value of the time series at time t, $h(t)$ is a set of harmonic regression coefficients of the real-valued, β_0: intercept term. a_0 *and* b_0: Fourier coefficients for the Fourier term, ϵ_t : error term, q_t: threshold variable (used to split the sample into two subgroups), e_t: regression disturbance. γ_t: time-varying threshold, φ_{ji}: time-varying threshold, *k* = No. of fitted parameters, d = delay parameter, R^2 = Coefficient of Determination, $L(\hat{\theta})$ = likelihood of the data, $\hat{\theta}$ = maximum likelihood estimate. $k =$ Number of estimated parameters for the models and $\bar{R}^2 =$ Adjusted coefficient of determination.

METHODS

A Threshold Spectral Fourier Regression (TSFR) model combines the ideas from Fourier terms, Threshold mechanisms and Regression structure is developed through the following steps; (i) Identifying the Time series structure (ii) Extracting dominant frequencies (iii) Building or specifying Fourier Regression model (iv) Introducing a threshold variable and defining threshold value (v) Splitting the model into regimes based on threshold (v) Estimating the threshold value and the model parameters (vi) Validating and Diagnosing the model.

The Periodogram or spectral density estimation is applied to the time series $y(t)$ to detect the dominant frequencies; these frequencies are used in the Fourier terms. The Fourier Regression Model is specified as in equation (20) to capture the cyclical or seasonal behaviour in the series. The Threshold variable q_t is chosen, using the quantile method, which is insensitive to outliers and missing values, and the threshold value τ is defined to split the model into regimes. The Threshold Spectral based-Fourier Regression (TSFR) model is specified in (27).

$$y_t = \begin{cases} \beta_0 + \sum_{k=1}^{N} \left[\beta_{1k} cos\left(\frac{2\pi kt}{T}\right) + \gamma_{1k} sin\left(\frac{2\pi kt}{T}\right)\right] + \varepsilon_t, & if\ d \le \tau \\ \alpha_0 + \sum_{k=1}^{N} \left[\alpha_{2k} cos\left(\frac{2\pi kt}{T}\right) + \delta_{2k} sin\left(\frac{2\pi kt}{T}\right)\right] + \varepsilon_t, & if\ d > \tau, \end{cases} \quad 27)$$

where y_t: dependent variable (time series value at time t), β_0 *and* α_0: baseline (mean) component, β_{1k} and γ_{1k}: Fourier coefficients for the first regime (where $d \le \tau$), α_{2k} and δ_{2k}: Fourier coefficients for the second regime (where $d > \tau$), T = period of the periodic dataset

$d \le \tau$ *and* $d > \tau$ are indicator functions for when *d* is below or above the threshold value τ

ε_t = residual/ error term (assumed to be white noise), $k =$ harmonic number/frequency of the sinusoidal terms.

RESULTS

Research Data

The data used by the authors in this study is Nigerian monthly rainfall data, covering the period January 1981 to May, 2025. This dataset contains rainfall indicators, computed from Climate Hazards Group Infrared Precipitation satellite imagery short-term rainfall forecasts, aggregated by subnational administrative units.

Descriptive Statistics

The time series data for this study is monthly rainfall, it covers forty-four years and five months (approximately five hundred and thirty-three months), and the dependent variable is the amount of rainfall.

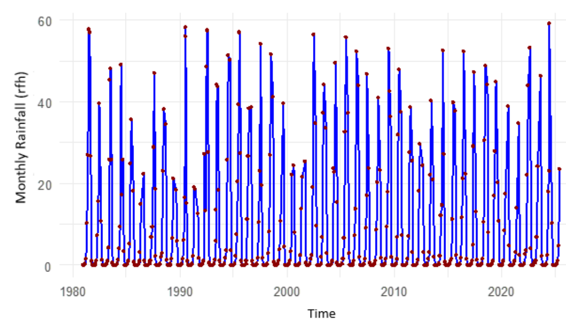

Figure 1. Time Plot of Monthly Rainfall

The time series plot of the monthly rainfall in figure 1.0 provides a temporal view of the dataset and suggests a nonstationary, nonlinear series with periodic structure, including seasonality, outliers, and possible regime or structural changes.

Table 1. Summary statistics of Rainfall data from 1981 to May,2025

Statistic	Value		
Count	533		
Mean	15.9		
Std Dev	24.96		
Variance	623		

Table 1 shows that the monthly average rainfall is 15.9 mm for the period, implying that the rainfall is moderate on average. The variance of 623 mm and the standard deviation of 24.96 mm indicate pronounced or extreme rainfall events (heavy storms) and

Science World Journal Vol. 21(No 2) 2026
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v21i2.30

very low rainfall periods (dry spells). The coefficient of variation of 157% (greater than 100%) indicates extremely high variability, suggesting that rainfall is unstable and inconsistent over time.

The statistical summary and the time plot of the rainfall data suggest that the Threshold Spectral Fourier Regression (TSFR) model is suitable for the data because the variance (623) is high, which needs a threshold to take care of the different rainfall regimes, seasonality that requires Fourier components to capture the wet (peak) and dry(trough) seasons, nonlinearity, and regime switching. When these two features (threshold and Fourier terms) are combined, the dry period will follow one pattern while the wet period will follow another.

Estimation of the Model Parameters Using Ordinary Least Squares (OLS)

The threshold effect in the TSFR model made the relationship between the dependent variable y sub t and the independent variable x sub t in the model nonlinear, such that the parameters of the TSFR model will be estimated using the Ordinary Least Squares method.

The TSFR model in equation (26) can be expressed as:

$$y_t = \begin{cases} \beta_0 + \beta_1 f(x_t) + \varepsilon_t, & if \ d \le \tau \\ \alpha_0 + \alpha_1 f(x_t) + \varepsilon_t, & if \ d > \tau \end{cases} \qquad (27)$$

where $y_t =$ Dependent variable at time t. $x_t =$ Independent variable at timet, $f(x_t) =$ Fourier transformation of x_t, τ = Threshold value, $\varepsilon_t =$ error term (assumed to be white noise), $\alpha_o, \alpha_1, \beta_o$ and β_1 are parameters to be estimated.

For the regime, $d \le \tau$.

$$y_t = \beta_0 + \beta_1 f(x_t) + \varepsilon_t \qquad (28)$$

And for the regime, $d > \tau$:

$$y_t = \alpha_0 + \alpha_1 f(x_t) + \varepsilon_t \qquad (29)$$

OLS procedure minimizes the sum of squared residuals (errors): Minimize

$$\sum\nolimits_{t \,\epsilon\, regime} (y_t - \hat{y})^2 \qquad (30)$$

where $\hat{y}$ is the predicted value from the Fourier-transformed regression model.

For $x_t \le \tau$, the OLS will estimate β_0 and β_1 such that

$$\widehat{\beta_o}, \widehat{\beta_1} = min \sum_{t\,;\,x_t \le \tau} \left(y_t - \beta_o - \beta_1 f(x_t)\right)^2 \qquad (31)$$

For $x_t > \tau$, the OLS will estimate $\alpha_0 and$ $\alpha_1$such that

$$\widehat{\alpha_o}, \widehat{\alpha_1} = min \sum_{t\,;\,d>\tau} \left(y_t - \alpha_o - \alpha_1 f(x_t)\right)^2 \qquad (32)$$

From equation (26), the error term is specified as:

$$\varepsilon_t = \beta_0 - \alpha_0 + \sum_{k=1}^{N}\left[\beta_k cos\left(\frac{2\pi kt}{T}\right) + \gamma_k sin\left(\frac{2\pi kt}{T}\right)\right] - \sum_{k=1}^{N}\left[\alpha_k cos\left(\frac{2\pi kt}{T}\right) + \delta_k sin\left(\frac{2\pi kt}{T}\right)\right] \qquad (33)$$

In matrix representation, the model in equation (26) can be written compactly as

$$y = X\beta + \varepsilon \qquad (34)$$

$$\varepsilon = y - X\beta \qquad (35)$$

where:

Let $= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{pmatrix}$ and X is the design matrix defined as:

$$X = \begin{pmatrix} 1 & P_{11}^{(1)} & D_{11}^{(1)} & \cdots & P_{k1}^{(1)} & D_{k1}^{(1)} & P_{11}^{(2)} & D_{11}^{(2)} & \cdots & P_{k1}^{(2)} & D_{k1}^{(2)} \\ 1 & P_{12}^{(1)} & D_{12}^{(1)} & \cdots & P_{k2}^{(1)} & D_{k2}^{(1)} & P_{12}^{(2)} & D_{12}^{(2)} & \cdots & P_{k2}^{(2)} & D_{k2}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & P_{1n}^{(1)} & D_{1n}^{(1)} & \cdots & P_{kn}^{(1)} & D_{kn}^{(1)} & P_{1n}^{(2)} & D_{1n}^{(2)} & \cdots & P_{k2}^{(2)} & D_{kn}^{(2)} \end{pmatrix}$$

The elements of X are the Fourier based functions in the TSFR model in (26), denoted as P_{kt}, D_{kt} and the indicator functions as I_{1t} and I_{2t}.

such that:

$P_{kt} = \sin\left(\frac{2\pi nt}{T}\right)$, $D_{kt} = \cos\left(\frac{2\pi nt}{T}\right)$, $I_{1t} = I(d_t \le \tau)$ and $I_{2t} = I(d_t > \tau)$

The regressors are:

$$P_{kt}^{(1)} = P_{kt} I_{1t} = \sin\left(\frac{2\pi nt}{T}\right) I(d_t \le \tau)$$

$$D_{kt}^{(1)} = D_{kt} I_{1t} = \cos\left(\frac{2\pi nt}{T}\right) I(d_t \le \tau)$$

$$P_{kt}^{(2)} = P_{kt} I_{2t} = \sin\left(\frac{2\pi nt}{T}\right) I(d_t > \tau)$$

$$D_{kt}^{(2)} = D_{kt} I_{2t} = \cos\left(\frac{2\pi nt}{T}\right) I(d_t > \tau)$$

The parameter vector is defined:

$$S = (\beta_0 \quad \alpha_{11} \quad b_{11} \quad \cdots \quad \alpha_{1k} \quad b_{1k} \quad \alpha_{2k} \quad b_{2k} \quad \cdots \quad \alpha_{2k} \quad b_{2k})^T$$

The OLS estimator minimizes the sum of squared error in (34)

$$\begin{aligned} \varphi(\beta) &= \varepsilon^T \varepsilon \\ &= (y - X\beta)^T (y - X\beta) \qquad (36) \\ &= (y^T - \beta^T X^T)(y - X\beta) \\ &= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta \\ &= y^T y - 2\beta^T X^T y + \beta^T X^T X\beta \end{aligned}$$

The result of the value $\varepsilon^T \varepsilon$ in equation (35) is reduced, such that

$$\frac{\partial \varphi(\beta)}{\partial \beta} = 0 \qquad (37)$$

$$\frac{\partial \varphi(\beta)}{\partial \beta} = -2X^T y + 2X^T X\beta = 0$$

$$-2X^T y + 2X^T X\beta = 0$$

$$X^T X\beta = X^T y$$

Since, $X^T X$ is a non-singular matrix, the OLS estimator $\hat{\beta}$, is

https://dx.doi.org/10.4314/swj.v21i2.30

specified as

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (38)$$

Equation (37) simultaneously estimates the:

(i) Intercept β_0
(ii) Fourier coefficient in regime 1,
(iii) Fourier coefficient in regime 2

The fitted values and residuals are respectively $\hat{y} = X\hat{\beta}$ and $y - \hat{y}$

The variance estimates and inference as given equations

(i) The unbiased estimator of the error variance is specified as

$$\hat{\sigma}^2 = \frac{1}{n-p} \hat{\varepsilon}^T \hat{\varepsilon}^2 \quad (38)$$

$$= \frac{1}{n_j - p} \sum_{t:d_{j=j-1}} (y_t - X^T \beta)^2 \quad , \mathrm{j} \in (1,2)$$

(ii) OLS covariance (conditional on regimes) or variance-covariance matrix of the estimator is given as:

$$Var(\hat{\sigma}^2) = \hat{\sigma}^2 (X^T X)^{-1} \quad (39)$$

Equation (39) is useful for conducting t-tests, confidence interval estimates, and regime-specific inference on the Fourier coefficients.

Selection of Threshold value (τ)

Threshold value or constant (τ), which filters the insignificant frequencies or splits the data for the TSFR model into regimes, to allow a separate Fourier regression model to be developed within each regime, is determined by the quantile method. This method is insensitive to outliers or missing values. The threshold value or constant $\tau = 0.85$ for rainfall data, which indicates a structural change point in the dynamics of the series, removes noise or irrelevant information, and improves the accuracy of the model.

Threshold Spectral-Based Fourier Regression Modeling

The Threshold Spectral-based Fourier Regression (TSFR) model in equation (26) is fitted to the dataset to assess its predictive accuracy and ability to model seasonal and cyclical time series. Threshold splits the rainfall dataset into two regimes, characterised as, Low (L) and High (H). The TSFR model can then be extended to these two regimes, denoted as $y_t^{(L)}$ and $y_t^{(H)}$ representing Low Rainfall and High Rainfall respectively.

$$\hat{y}_t^{(L)} = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{T}\right) + \gamma_1 \sin\left(\frac{2\pi t}{T}\right) + \beta_2 \cos\left(\frac{2\pi 2t}{T}\right) + \gamma_2 \sin\left(\frac{2\pi 2t}{T}\right) \qquad d \le 0.85 \quad (33)$$

$$\hat{y}_t^{(H)} = \alpha_0 + \alpha_1 \cos\left(\frac{2\pi t}{T}\right) + \delta_1 \sin\left(\frac{2\pi t}{T}\right) + \alpha_2 \cos\left(\frac{2\pi 2t}{T}\right) + \delta_2 \sin\left(\frac{2\pi 2t}{T}\right) \qquad d > 0.85 \quad (34)$$

T = period (e.g. 12 for monthly data for yearly seasonality) and n is the number of harmonics (2, 3, 4, …)

The Fourier coefficients of the TSFR models in equations (33) and (34) were estimated using the OLS method outlined in section 4.2, which fits a model to minimize the difference between observed rainfall values and those predicted by the models. The Fourier coefficients for the TSFR models based on the rainfall datasets are given in Table 2. The fitted TSFR models based on the dataset and the threshold values $(i)\, r \le 0.85$ and $(ii)\, r > 0.85$ are specified in equations (35) and (36).

Table 2. Values of Fourier coefficients for Threshold Spectral Fourier Regression Model (Low and High) Rainfall.

Parameters	Low	High
β_0	0.216904278	0.9998647
β_1	0.004977601	-0.00005097524
γ_1	0.005457208	-0.0001918983
β_2	0.001329586	0.00003468025
γ_2	-0.001515835	-0.0000607448

For $r \le 0.85$

$$\hat{y}_t^{(L)} = 0.216904278 + 0.004977601 \cos\left(\frac{\pi t}{6}\right) + 0.005457208 \sin\left(\frac{\pi t}{6}\right) + 0.001329586 \cos\left(\frac{\pi t}{3}\right) - 0.001515835 \sin\left(\frac{\pi t}{3}\right)$$

(35)

$$with\ R^2 = 0.8878, Adjusted\ \bar{R}^2 = 0.7486\ and\ Durbin\ Watson\ statistics = 1.957$$

For threshold value, $r > 0.85$

$$\hat{y}_t^{(H)} = 0.99986 - 0.00005 \cos\left(\frac{\pi t}{6}\right) - 0.00019 \sin\left(\frac{\pi t}{6}\right) + 0.00003 \cos\left(\frac{\pi t}{3}\right) - 0.00006 \sin\left(\frac{\pi t}{5}\right)$$

(36)

$$with\ R^2 = 0.9484, Adjusted\ \bar{R}^2 = 0.9038\ and\ Durbin\ Watson\ statistics = 1.973$$

Model Evaluation

The values of AIC and BIC for Threshold Spectral Fourier Regression model (Low and High) are given in Table 3, and the values of the Coefficient of Determination (R^2) and Adjusted Coefficient of Determination $\bar{R}^2$ in Table 4.

Table 3. Values of Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) for Threshold Spectral Fourier Regression Model (Low and High) Rainfall

Models	Log Lik	AIC	BIC
TSFR (Low)	3417.66	-6827.32	-6806.23
STFR (High)	4421.41	-8832.82	-8808.38
FR	176.11	-342.21	-321.53
AR	-480.21	972.43	1004.08

Science World Journal Vol. 21(No 2) 2026
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University
https://dx.doi.org/10.4314/swj.v21i2.30

Table 4. Values of Co-efficient of Determination and Adjusted Co-efficient of Determination for Threshold Spectral Fourier Regression Model (Low and High) for Rainfall

Model	Co-efficient of Determination (R^2)	Adjusted Co-efficient of Determination ($\bar{R}^2$)
TSFR (Low)	0.8818	0.7486
TSFR (High)	0.9484	0.9038
FR	0.7806	0.6738
AR	0.7193	0.5142

Forecast Evaluation Metrics
The forecast evaluation or error metrics, Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Average Percentage Error (MAPE) for the TSFR models were determined and given in Table 5.

Table 5. Forecast Evaluation Metrics for Threshold Spectral Fourier Regression Model (Low and High) for Rainfall dataset.

Model	MAE	RMSE	MAPE (%)
TSFR (Low)	0.0844	0.1301	0.03
TSFR (High)	0.0005	0.0027	0.05
FR	0.2968	0.3375	21.97
AR	0.0020	0.0226	0.17

Residual Analysis
The residual analysis of the Threshold Spectral Fourier Regression (TSFR) model for both low rainfall and high rainfall regimes was carried out using the Autocorrelation Test (Durbin-Watson) and the Ljung-Box p-values test; the results are given in Table 6.

Table 6. Autocorrelation Test (Durbin-Watson) for Threshold Spectral Fourier Regression Model (Low and high) for Rainfall data

Regime	DW Statistic	p-value	Interpretation
Low Rainfall	1.957	6.22e-2	No autocorrelation
High Rainfall	1.973	6.22e-2	No autocorrelation

FORECAS

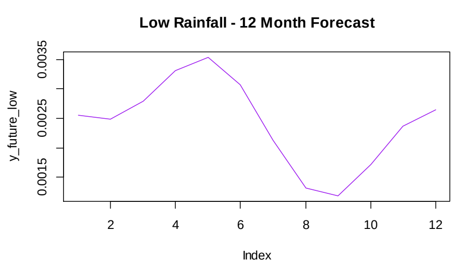

Figure 2. 12-Month Forecast of Low Rainfall

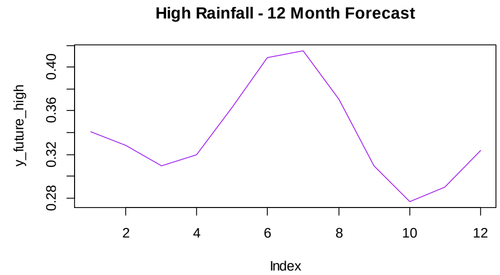

Figure 3. 12-Month Forecast of High Rainfall

Figures 2 and 3. The 12-month forecast for the two regimes shows a smooth cyclical pattern, reflecting the continuation of the series 'underlying seasonal behaviour. The forecast predicts the highest rainfall values between months 4 and 6, corresponding to a mild wet period, and the lowest values between months 9 and 10, representing the peak of the dry season. The amplitude of variation remains small, suggesting that low-rainfall months exhibit limited fluctuation around their generally low mean levels. This pattern implies that although dry conditions dominate, there is a predictable annual oscillation in which slightly wet months recur at consistent intervals. The high rainfall forecast shows a more pronounced and regular seasonal pattern than the low rainfall case. The predicted peak occurs around months 6–7, marking the wet season, while the lowest values are projected between months 9–11, indicating the onset of the dry season. The amplitude of the forecasted oscillation is much greater, capturing the intensity and variability typical of high rainfall months.
In the context of TSFR, the two results align with the spectral and time-series analyses, confirming a strong annual rhythm in rainfall behaviour. Overall, the model provides a credible forecast of seasonal rainfall trends, highlighting the predictable timing of wet and dry cycles in the region.

DISCUSSION
The Fourier coefficients in Table 2 correspond to the estimated parameters of the Threshold Spectral Fourier Regression (TSFR) model applied to the dataset's Low and High Rainfall categories. These parameters contain five values for each category. The values, β_0 =0.216904278, and alpha sub 0 equals 0.9998647, are the respective baseline rainfall in the low and high regimes,o r dry and wet/rainy seasons. If these values are the same ($\beta_0 = \alpha_0$), it implies that there is a shift in mean across regimes. The Fourier cosine coefficient 0.004977601 and -0.00005097524 measure the magnitude of cosine-based periodic components in the two regimes; these parameters capture long-run smooth cycles; their large values indicate strong seasonality or smooth cyclical behaviours, and small values indicate weak seasonality or periodic influence. The Fourier sine coefficients capture phase-shifted cyclical motion and peak timing. The Fourier sine coefficients, along with Fourier cosine terms, allow the model to represent seasonal effects, oscillatory patterns, and complex nonlinear cycles.

The AIC and BIC values for the TSFR model (Low and High) for the dataset are given in Table 3. AIC and BIC both measure goodness of fit with a penalty for complexity; their lower values indicate a better model compared with the values for the Autoregressive (AR) and Fourier regression (FR) models. These

Science World Journal Vol. 21(No 2) 2026 https://dx.doi.org/10.4314/swj.v21i2.30
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

lower values imply that the TSFR model performs better and balances the fit better than the two models. It simply means that the model adequately explains the observed data by capturing seasonal (Fourier) patterns, that the threshold correctly divides rainfall into regimes, and that the model is appropriate.

The Coefficient of Determination (R) and the Adjusted Coefficient of Determination ($\bar{R}^2$) for the three models as shown Table 4, indicates that both values for TSFR (Low and High) are higher. This suggests that the TSFR model captures the structure more effectively, provides a better fit, and offers better predictive performance than the R and AR models. The high values of R and R^2 can be interpreted as indicating that the TSFR model has high explanatory power, as over 80% of the variation in rainfall data was explained by the model, and that the variables are used efficiently. The error metrics MAE, RMSE, and MAPE in Table 5 measure how well the model fits and predicts the data; lower values indicate a better fit and prediction. It is observed that the TSFR (Low and High) models have lower values than the FR and AR models. This suggests that the TSFR model handles seasonality better, avoids overfitting, provides a better understanding of the dynamics of structural breaks, and offers better forecasts.

The Durbin-Watson statistics for the Threshold Spectral Fourier Regression model for the low rainfall and high rainfall regimes are shown in Table 6. These statistics (1.957 and 1.973) are approximately equal to 2, which suggests the absence of autocorrelation. This outcome strongly indicates the absence of first-order correlation in the models' residuals. The associated Ljunp-Box p-values (6.22e-2) further confirm that the null hypothesis (Ho: No autocorrelation) could not be rejected at an alpha equals 0.05level of 0.05.

In the context of the Threshold Spectral Fourier Regression (TSFR) model, designed such that the Fourier terms capture seasonal and cyclical components and thresholds capture regime shifts, the presence of residual autocorrelation would imply mis-specified frequencies, missing regime dynamics, and an incomplete threshold structure. The Durbin-Watson Statistic (DW = 1.957 and 1.973) for rainfall and the corresponding Ljung-Box p-values in the two regimes exceeding the 0.05 cutoff, indicates that the null hypothesis of no residual autocorrelation cannot be rejected or there is no strong evidence of autocorrelation in the residuals of TSFR model for the dataset, this suggests that the models adequately capture the dynamics of the data, well specified, meaningful and possesses good balance between complexity and accuracy.

Model Comparison
Table 7 highlights the performance comparison metrics for the three models (TSFR, FR and AR). The high values of R^2 = 0.8818 and Adjusted $\bar{R}^2$ = 0.7486 in the low regime and R^2 = 0.9484 and Adjusted $\bar{R}^2$ = 0.9038 in the high regime, shows that the model explained over 80% of the variation in the rainfall, it means that TSFR model produced high explanatory powers, and thus reflects a robust fit to the rainfall data even after accounting for model simplicity. The lower MAPE, RMSE, and MAE values in both regimes of the TSFR model also indicate that predictions were much closer to actual rainfall outcomes, or that there is close alignment between the fitted and actual values. The lower AIC and BIC values (in both regimes of the TSFR model) further confirmed that the TSFR model substantially outperformed the FR and AR models. It effectively captured rainfall patterns, making it statistically robust and a practically reliable framework for forecasting.

Table 7. Model Performance Comparison of Threshold Spectral Fourier Regression (TSFR), Fourier Regression (FR) and Autoregressive (AR) Models for Rainfall

Metric	**TSFR (Low)**	**TSFR FOURIER (High) (FR)**	**REG.**	**AUTO REG. (AR)**
RMSE	**0.1301**	**0.0027**	0.3375	0.0226
MAE	**0.0844**	**0.0005**	0.2968	0.0020
MAPE	**0.0300**	**0.0500** 21.970		18.492
R^2	**0.8818**	**0.9484**	0.7806	0.7193
Adj. $\bar{R}^2$	**0.7486**	**0.9038** 0.6738		0.5142
AIC	**-6427.32**	**-8832.82** 342.21	-	972.43
BIC	**-6806.23**	**8808.38** 321.53	-	1004.03

Conclusion
Based on the study, the best threshold value for the Threshold Spectral Fourier Regression (TSFR) model for monthly rainfall data in Nigeria between 1981 and 2025 is 0.85.
The level of model accuracy is evidenced by the average MAPE values in the two regimes of the TSFR model, with the effect of monthly rainfall in Nigeria between the periods being 0.04, and the Adjusted coefficient of Determination is 0.8262, indicating that the model can be used to predict rainfall for Nigeria accurately.
The findings in this study reveal that the Threshold Spectral Fourier Regression model is a useful tool in time series analysis; it is a flexible model suitable for datasets that exhibit nonlinearity, regime switching behaviour, and seasonal or periodic patterns, such as climate data, economic variables, financial series, or series where cyclical patterns depend on threshold variables.
The Threshold Spectral Fourier Regression model has a better fit than non-threshold or standard models, as evidenced by the lower AIC/BIC and the higher Adjusted Coefficient of Determination. The model identifies the dynamics of regime specification, providing separate models for different threshold intervals; different regimes have different frequency components, thereby enhancing in-sample and out-of-sample forecasting accuracy, which is key to decision-making.
The threshold effect in the model is easily detected, since a specific threshold value (or values) significantly alters the model's structure.

REFERENCES
Anatansyah, A. M., Epha, D. S., & Muhammad, V. M. (2022). Fourier series nonparametric Regression modelling for Rainfall in West Java Province. *International Journal of Informatics for Development (IJID). 11(1)*, 142-151. DOI:10.14421/ijid.3300. e-ISSN: 2549-7448
Box, G.E.P., Jenkins, G.M., Reinsel, G. C., & Ljung, G.M (2016).

 https://dx.doi.org/10.4314/swj.v21i2.30

Time Series Analysis: Forecasting and Control. (5th ed.). Hoboken, N. J: John Wiley & Sons. ISBN 978-1-118-67502-1.712

Gyasi-Agyei, K. (2012). Analysis and Modeling of the Prevalence of Measles in the Ashanti Region of Ghana. West Africa: University of Science and Technology.

Hansen, B. E. (2011). Threshold Effects in Non-Dynamic Models: Estimation, Testing, and Inference. *Journal of Econometrics*, *93(2)*, 345-368.

Hyndman, R. J., & Athanasopoutos, G. (2021). Patterns of Seasonality and Cyclical Behaviours in Forecasting: Principles and Practice. (3rd.ed.). OText https://otext,Com//fpp.3/

Kumar, A. (2023, March 31). Different types of time series forecasting models. VitalFlux. hptts;//vitalflux.com/different-types-of-time-series-forecasting-models

Kourtellos, A., Stengos, T., & Tan, C. M. (2014). Structural Threshold Regression. *SSRN Electronic Journal.* DOI:10.2139/ssrn.1105931. 1-66.

Lima, S., Ferreira, P., & Santos, R. (2023). Predictive accuracy of time series models applied to retail sales: A comparison using RMSE. *Journal of Statistical Forecasting*, 12(3), 45–59.

Lixiong Yang, Chngnun Lee & I-PoChen (2020). Threshold Model with a Time-varying Threshold Based on Fourier Approximation. *Journal of Time Series Analysis 42 (4)* DOI: 10.1111/jtsa.12574

Lixiong Yang, Chngnun Lee and I-PoChen (2021). Time-varying threshold cointegration with an application to the Fisher hypothesis. *Studies in Nonlinear Dynamics and Econometrics. 26(2)*. DOI: 10.1515/snde-2018-0101

Qudratullah, M. F. (2007). Bayesian Information Criterion (BIC) dalam pemilihan model terbaik Feed Forward Neural Network (FFNN). https://etd.repository.ugm.ac.id/penelilian/detail/32974

Sherrod, P. H. (2015). DTREG time series analysis. https://www.dtreg.com

Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With R examples (4th ed.). Springer

Suparti, S. (2012). Perbandingan estimator regressi non-parametric menggunakan metode Fourier dan metode wavelet. *Matematika,* 8(3), 45. https://ejournal.undip.ac.id/index.php/matematika/article/view/1373

Szulczyk, K. R. and Sadique, S. (2018). Using Cyclical Components to Improve the Forecasts of the Stock Market and Macroeconomic Variables, *Journal of Modern Applied Statistical Methods*, 17(1), 1 – 23.

Taiwo A. I., & Olatayo T. O. (2018). An Improved Procedure for Fourier Regression Analysis. *Anale. Seria Informatică. 16(2)*. 108-112.

Tong, H. (2011). Threshold models in nonlinear time series analysis. Springer.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. Journal of the American Statistical Association, 84(405), 231–240. https://doi.org/10.1080/01621459.1989.10478760

Wei, W. W. S. (2019). Time series analysis: Univariate and multivariate methods (2nd ed.). Pearson. Wilfredo P. (2016). Time series analysis. Pontificia Universidad Católica de Chile. Willey Series and Son's. 121-277.

Zhang, X., Li, Q., & Chen, J. (2025). Threshold Fourier regression models for nonlinear seasonal data. International Journal of Forecasting, 41(2), 101-108.