IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

**ISSN: 3007-1658 (Online)**

# Climatic Time Series Forecasting by using the Hybrid TF-SMO Model

Najlaa Khalid Ahmed[1] , Sabah Manfi Ridha[2]

[1,2] Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

**ARTICLE INFO**

***Article history:***

Received 10/11/2024
Revised 10/11/2024
Accepted 13/1/2025
Available online 15/5/2025

***Keywords:***

Time series forecasting
TF-SMO model
Hybrid TF-SMO
Maximum , minimum and evaporation temperatures
Spider Monkey Optimization (SMO)

**ABSTRACT**

Forecasting is considered an estimation of future values based on past data, and its importance lies in future planning. Necessity often calls for finding a model that logically describes the dynamic relationships connecting a single output series to one or more input series. In this research, daily time series data for several years of maximum and minimum temperatures were used as input variables, and evaporation as a response variable for the Baghdad governorate during the period. (2012-2022). Environmental and climatic data often suffer from problems such as heterogeneity and unreliability, which are results of the non-linearity of that type of data. It is possible to use the Transfer function model, denoted by the symbol TF, to model the causal relationship between the output variable and one or more input variables after synchronizing the data temporally to achieve its homogeneity. The aim of this research is to improve the forecasting of time series data, and one of the proposed methods for this is the hybrid model combining the TF model and the Spider Monkey Optimization (SMO) Algorithm, referred to as the TF-SMO hybrid model. The outputs of the TF model are utilized as inputs in the spider monkey algorithm to integrate linear and nonlinear hybrid effects for data processing, modeling, and improving forecasting accuracy. To improve modeling and forecasting results, the (SMO) Algorithm was used as a method for optimizing the modeling of non-linear data patterns and enhancing forecasting. The hybrid model TF-SMO was proposed, which is a hybrid model of the SMO algorithm as an optimization method with the TF model. As for the issue of data heterogeneity, the Time Stratified (TS) method was used to temporally stratify the data into hot and cold seasons. Where the data for the hot and cold seasons for maximum and minimum temperatures and evaporation were divided into two groups: the first group for the training period and the second group for the testing period, approximately 65% for training and approximately 35% for testing .In this research ,data similar to real data was generated with the same sample size for the hot and cold seasons for the transfer function. The results indicate that the hybrid model achieves better accuracy compared to the TF model. Therefore, it is possible to use the TF-SMO hybrid model structure for similar accuracy in forecasting climatic time series data.

## 1.Introduction

Forecasting climatic and environmental data is essential to prevent the world from unexpected natural hazards such as extreme high and low temperatures, evaporation, rainfall, winds, and others. The repeated rise and fall in maximum and minimum temperatures, evaporation, and the non-linear nature of the data, along with the issue of data heterogeneity, make forecasting a complex process for accurately forecasting climatic data. Therefore, some researchers have suggested using the Transfer Function (TF) model, which is a traditional statistical model for forecasting this data. To address the issue of data heterogeneity, the data was divided into hot and cold seasons for analysis and

* Corresponding author E-**mail** address: **njlaalzydy962@gmail.com**

https://doi.org/10.62933/mk1bpd71

forecasting. Daily data for the study was obtained from the General Authority for Meteorology and Seismic Monitoring for the province of Baghdad over eleven years from 2012 to 2022. In this research, the data for the hot and cold seasons were divided into two groups. The first group is called the Training period dataset, which is used to build the models, and the second group, called the Testing period dataset, is used to test the models. The ratio of training data to testing data used was approximately 65% for training and approximately 35% for testing. In this research, the TF model was combined with the Spider Monkey Optimization algorithm, which is one of the swarm intelligence algorithms inspired by the behavior of monkey troops in search of food. The hybrid TF-SMO model was used to improve forecasting accuracy, and some forecasting accuracy metrics, including the MSE criterion, were employed to evaluate the models' forecasting accuracy. The results obtained indicate that the SMO algorithm can enhance the forecasting accuracy of the TF model, as it yielded the lowest MSE for forecasting errors.

The researchers Sentas and Psilovikos[1] used the TF transfer function for modeling and forecasting, utilizing daily measurements over three years (2005-2007) for water temperature data (Tw) at four different depths (1 m, 20 m, 40 m, 70 m) from the Thesaurus Dam Lake in the Nestos River to obtain the best models, and MSE was calculated to evaluate the forecasting accuracy.

The researcher Cho, Hwang et al. [2] used a method for forecasting using the ARIMA model and the TF transfer function model for short-term load forecasts by considering the weather-load relationship for four types of customers in Taiwan's power system: residential loads, commercial loads, office loads, and industrial load customers. They proposed a ARIMA model transfer function model for short-term load forecasting over one week and verified the forecasting accuracy of the transfer function for forecasting load on weekdays and weekends. To improve forecasting accuracy, the impact of temperature was considered in the transfer function. To demonstrate the effectiveness of the proposed method, the results of the TF model were compared with the ARIMA model and traditional regression. It was concluded that the proposed method achieves better load forecasting accuracy than the ARIMA model by considering the causal relationship between energy consumption and temperature.

Researchers have developed forecasting methods, and among these methods are hybrid models built from mixed linear and nonlinear models to address the problem of nonlinear data. These hybrid models help achieve greater forecasting accuracy. The researchers Yang, Zhai, Xu and Han [3] used the SMO algorithm to forecast two models of time series data: the number of sunspots and the Wolfer sunspot number, as well as the Box and Jenkins gas furnace data. They observed that the forecasting accuracy using the SMO algorithm is better than the forecasting accuracy of the neural network and Support Vector Regression using the QP optimization algorithm. The researchers Khair, Awang, Zakaraia and Mazlan [4] presented three algorithms to improve the accuracy of time series forecasting for water flow level data for the period (2001-2012), and the SMO algorithm yielded more accurate results compared to the other algorithms. The researcher Ragab [5] proposed the hybrid model CSMO-OKRR, which is a hybrid model of CSMO (chaotic spider monkey optimization) with the optimal kernel ridge regression model for predicting time series of rainfall data and improving SMO. The results showed the efficiency of the hybrid model. In this research, the TF model was hybridized with the Spider Monkey Optimization algorithm, which is one of the optimization algorithms inspired by the behaviour of a group of spider monkeys in searching for food. The hybrid TF-SMO model was used to improve forecasting accuracy. To evaluate the accuracy of the models, some forecasting accuracy metrics were used, including the MSE metric. The results showed that the SMO algorithm was able to improve the forecasting accuracy of the TF models, providing the lowest MSE for prediction errors.

In this research, the time stratification method was used, which involves stratifying the data according to seasonal effects by dividing the data into hot and cold seasons to achieve homogeneous data and reach more accurate forecasting results. Malig, Pearson, Chang, Broadwin, Basu, Green and Ostro[6] tested the relationship between ozone and asthma and respiratory emergency department visits (EDVs), particularly acute respiratory infections (ARI), asthma, pneumonia, chronic obstructive pulmonary disease (COPD), and upper respiratory tract infections (URTI). The correlations were stronger and more consistent in the hot season, which should be taken into account when evaluating the potential health benefits of reducing ozone concentrations. Both ALbazzaz and Shukur [7] proposed using some machine learning and deep learning models to classify time series of solar radiation. They used temporal data stratify, and their results were distinguished by more accurate classification performance. This research included a theoretical review of the methods used (TF, SMO and TS) and the practical application of these methods on maximum and minimum temperature and evaporation data for Baghdad Governorate to obtain better forecasting. The data was divided into two seasons, hot and cold, for maximum and minimum temperature and evaporation data into two groups: the first called the Training period dataset, which is used to build the models, and the second called the Testing period dataset, which is used to test the models. The ratio of training data to test data was approximately 65% for the training period and approximately 35% for the test period.

## 2. Methodology

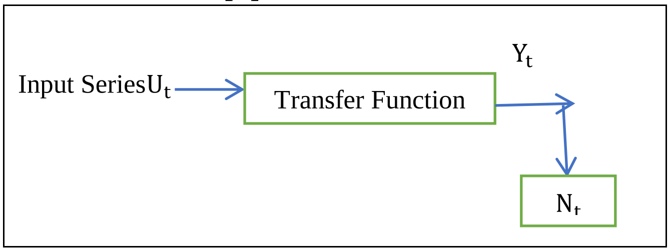
### 2.1.Transfer Function Models

Univariate time series models study a single time series, which leads to the neglect of some correlation relationships of this series with other time series. Therefore, it is necessary to study models that contain a set of time series that are interrelated with real reciprocal relationships. This means that the series will not only be affected by its past values but also by the present and past values of other time series. These models are called multivariate time series models or transfer function models. It is also called the Dynamic Regression model due to the presence of a dynamic relationship between the response variable and the explanatory variables.

Assuming that $U_t$ and $Y_t$ are stationary series, where the $U_t$ series represents the Input Series and the $Y_t$ series represents the Output Series, and assuming that :

1-The input series affects the output series through a relationship or function called the binary transformation function.

2- In addition to the effect of $U_t$ on the output series, there are other variables that affect the output series, referred to as noise or disturbance variables, denoted by the symbol $N_t$. The following figure illustrates the transfer function model[8]:




**Fig**ure (1).The structural model of the dynamic system for the transfer function.

The process of building transfer function models is the same as the stages of building ARIMA models established by researchers Box and Jenkins, which include model diagnosis, estimation, conducting diagnostic tests, and forecasting. The difference lies in the construction phase, where there is an intensive purification process for the two time series from known influences. The first stage of building a transfer function model is the diagnosis of the transfer function model, through which the stationary of the input and output series is determined in terms of the absence of a general trend in the series and variance through the absence of data dispersion. Any seasonal effect, if present, is also removed.

This stage is divided into three parts, and the first part of the model diagnosis stage is the estimation of the impulse response weights. These weights are estimated using two methods, the first of which relies on the cross-correlation function. The weights of the transfer function $V_0, V_1, \ldots V_m$ are estimated through a statistical measure called the Cross Correlation Function, denoted by the symbol CCF, and can be expressed in the following form:

$$\hat{\rho}uy(m) = \frac{\gamma uy(m)}{\hat{\sigma}u\hat{\sigma}y} \quad (1)$$

$\gamma uy$ represents the cross-covariance between $U_t$ and $Y_t$, and $\sigma u, \sigma y$ are the standard deviations of the series $U_t$ and $Y_t$.
To identify the transfer function model using the cross-correlation function, the input and output series are prewhitened, meaning the data of the two series $U_t$ and $Y_t$ are prewhitened. This process helps eliminate the impact of changes occurring within the input and output series. For prewhitening, we assume that the input series $U_t$ follows an ARMA model as follows : [8,9]

$$\phi(B)U(t) = \theta(B)at \quad (2)$$

$$a_t = \frac{\phi u(B)}{\theta u(B)} U_t \quad (3)$$

The series of input whitening is called at. And then the output series $Y_t$is whitened in the same way as follows:

$$\phi(B)Y(t) = \theta(B)bt \quad (4)$$

$$b_t = \frac{\phi Y(B)}{\theta Y(B)} Y_t \quad (5)$$

$b_t$ represents the whitened output series.The aim of this whitening is to purify the two series $U_t$, $Y_t$ from any pattern resulting from an autoregressive or moving average process, leaving only the white noise, which are $b_t$, $a_t$.After obtaining the whitened residual series at for the inputs and $b_t$ for the outputs, the cross-correlation coefficient between the two whitened series was calculated as follows:

$$\rho_{ab}(m) = \frac{cov(a,b)}{\sigma_a \sigma_b} \quad (6)$$

And then the impulse response weights $(V_0, V_1, \ldots \ldots, V_m)$are calculated based on the cross-correlation results between the two filtered series as follows:

$$V(m) = \frac{\rho_{ab}\sigma_b}{\sigma_a} \quad (7)$$

The second method for estimating the impulse response weights of the transfer function is based on the Linear Transfer Function Method, denoted by the symbol LTF. Since the series $U_t$ and $N_t$ follow some properties of ARIMA models. [9,10]

$$Y_t = V_0 U_t + V_1 U_{t-1} + V_k U_{t-m} + N_t \quad (8)$$

$$= (V_0 + V_1 B + \ldots\ldots + V_m B^m)\, U_t + N_t \quad (9)$$

$$= V(B)\, U_t + N_t \quad (10)$$

Where $(V_0, V_1, \ldots \ldots, V_m)$The impulse response weights for the input series Ut represent the effect that occurs in the output series Yt as a result of a one-unit change in $U_t$. $Y_t$ represents the output series, $U_t$ represents the input series, and $N_t$ is the noise or disturbance term. m is the order of the transfer function and B is the back shift operator.
The second part of the model diagnosis phase is determining the Disturbance Model, which is done after estimating the impulse response weights $(V_0, V_1, \ldots \ldots, V_m)$The appropriate ARMA model for the disturbance is denoted as Nt using the following formula. [9]

$$Y_t = V(B)U_t + N_t \quad (11)$$

$$N_t = Y_t - V(B)U_t \quad (12)$$

And the third part of the diagnostic phase is determining the order of the transfer function model, which means calculating (r,s,b) .It is calculated using the cross-correlation function by using the following formula: [11]

$$V(B) = \frac{W(B)}{\delta(B)} = \frac{(W_0 - W_1 B - \cdots - W_s B^s)}{(1 - \delta_1 B - \cdots - \delta_r B^r)} \quad (13)$$

The information present in the input series $U_t$ and the output series $Y_t$ is finite, while the transfer function V(B) contains an unlimited number of parameters. Therefore, a general formula for the polynomial transfer function model has been established, requiring the least number of parameters through the following formula, which is considered the general formula for the transfer function:

$$Y_t = \frac{W_{s(B)} B^b}{\delta_{r(B)}} U_t + \frac{\theta(B)}{\phi(B)} at \quad (14)$$

Since:

$$W_{s(B)} = W_0 - W_1 B - \cdots \ldots . . -W_s B^s \quad (15)$$

$$\delta_r(B) = 1 - \delta_1 B - \cdots \ldots . \delta_r B^r \quad (16)$$

$$\theta(B) = 1 - \theta_1 B \ldots \ldots . -\theta_q B^q \quad (17)$$

$$\phi(B) = 1 - \phi_1 B - \cdots \ldots . . -\phi_p B^p \quad (18)$$

We note that: $W_s(B)$represents the Initial Effect, $\delta_r(B)$represents the Damping Factor, θ(B) is the Moving Average Operator, ϕ(B) is the Autoregressive Operator, and b represents the Delay parameter, also known as Dead Time. It is a positive integer that can be determined by plotting the cross-correlation function between the input series $U_t$ and the output series $Y_t$, where it represents the first significant correlation that falls outside the confidence limits [12]. q is the order of the Moving Average, and p is the order of the Autoregressive.

### 2.2. Spider Monkey Optimization (SMO) Algorithm

The Spider Monkey Algorithm (SMO) is an optimization algorithm derived from the foraging behaviour of a population of spider monkeys in search of food. These monkeys live in the tropical rainforests of Central and South America, in the northernmost region of Mexico. Spider monkeys are considered some of the most intelligent monkeys today, named for their spider-like appearance when suspended by their tails. These monkeys stay in a single group called the "mother group," where they unite or divide themselves based on the availability or scarcity of food. They communicate with each other through their gestures, postures, and calls, and this algorithm consists of several phases: Local Leader Phase, Global Leader Phase, Global Leader Learning Phase, Local Leader Learning Phase, Local Leader Decision Phase and Global Leader Decision Phase[13] . .

Initially, the SMO algorithm in the initialization phase generates a primary swarm of spider monkeys( SM)with a sample size N of dimension D, which is the number of variables. The initial swarm is initialized using the following equation [13,14].

$$SM_{ij} = SMmin_j + U(0,1) \times \left(SMmax_j - SMmin_j\right) \quad (19)$$

$SM_i$ - The spider monkey in the $i^{th}$swarm.

$SM_{ij}$ is the upper and lower limit for $SMmin_j$and $SMmax_j$, respectively, for the search space within the $j^{th}$ dimension. As for U(0,1), it is the uniform distribution Uniform(0,1) within the interval (0,1).

The first phase of the algorithm is the local leader phase, which is an important stage for the algorithm and is denoted by the symbol LLP. In this phase, all spider monkeys have the opportunity to update their position. The update in the position of the spider monkey depends on the experiences of the local leader and the members of the local group. After that, we calculate the fitness value for each spider monkey in its new position. If the new fitness is higher than the old fitness of the monkey, it is updated. However, if the new fitness is lower, it will not be updated and will remain at the initialization equation. The equation for updating the location can be expressed as follows:

$$\mathrm{SMnew_{ij}} = \mathrm{SM_{ij}} + \mathrm{U}(0,1) \times \left(\mathrm{LL_{kj}} - \mathrm{SM_{ij}}\right) + \mathrm{U}(-1,1) \times \left(\mathrm{SM_{rj}} - \mathrm{SM_{ij}}\right) \quad (20)$$

The term $LL_{kj}$ represents the local dimension of the $k^{th}$ group and $SM_{rj}$is the $j^{th}$ dimension of the spider monkey, which is randomly selected from the $k^{th}$ group. On the other hand, U(-1,1) follows a uniform distribution within the interval. (-1,1).

The Global Leader Phase (GLP) is the second phase, and in it, the location is updated based on the fitness function, which can be calculated using the following formula:

$$\text{fitness function} = \mathrm{fit_i} = \begin{cases} \frac{1}{1+\mathrm{f_i}}; & \text{if } \mathrm{f_i} \geq 0 \\ 1 + \mathrm{abs}(\mathrm{f_i}); & \text{if } \mathrm{f_i} < 0 \end{cases} \quad (21)$$

$f_i$ represents the objective function and $fit_i$ is the fitness value for the $i^{th}$ SM.If the fitness of the SM is suitable, the probability of selecting it is calculated using the following formula:

$$prob_i = 0.9 \times \frac{fit}{max\, fit} + 0.1 \quad (22)$$

$prob_i$ represents the selection probability and $\max fit$ is the maximum fitness in the group.

All spider monkeys update their location using the expertise of the global leader as well as the experience of the neighbouring spider monkeys and the update equation is:

$$\mathrm{SMnew_{ij}} = \mathrm{SM_{ij}} + \mathrm{U}(0,1) \times \left(\mathrm{GL_j} - \mathrm{SM_{ij}}\right) + \mathrm{U}(-1,1) \times \left(\mathrm{SM_{rj}} - \mathrm{SM_{ij}}\right) \quad (23)$$

$GL_j$ represents the global leader's position in the $j^{th}$dimension.

The local leader decision phase is the fifth phase, followed by the global leader decision phase, which is the sixth phase. If any local leader's position is not updated to a specific number of iterations, the spider monkeys update their positions either through random initialization or through the following equation:

$$SMnew_{ij} = SM_{ij} + U(0,1) \times (GL_j - SM_{ij}) + U(0,1) \times (SM_{ij} - LL_{kj}) \quad (24)$$

The process of updating the spider monkey locations continues until the stopping condition is met and we obtain the optimal solution. The figure below represents the stages of the SMO algorithm:

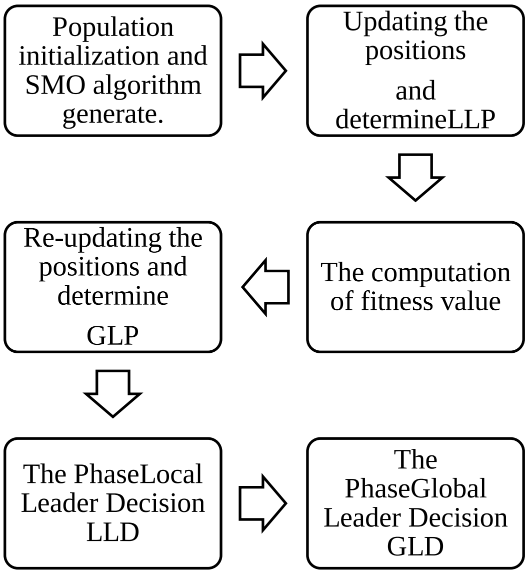



Fig 2.The stages of Spider Monkey Optimization(SMO)

**2.3. Hybrid TF-SMO Model**

In this research, a hybrid model was proposed, namely the Hybrid SMO-TF Model. In this hybrid model, TF was relied upon after estimating its parameters, and then the structure of the TF model was used to determine the variables entering the SMO algorithm. Where the structure of the TF model was taken and used as inputs in the SMO monkey algorithm, this hybridization between TF and SMO helped obtain a suitable method for dealing with nonlinear data, thereby improving forecasting. The framework for the hybrid model can be illustrated with the following steps using the SCA program and the Matlab program:

1-The data was divided into hot and cold seasons using time series stratification (TS).

2- After splitting the data, the structure of the TF model is determined from the variables and parameters.

3-Identifying the response variable, which is the evaporation variable Y.

4- The forecasting of the TF models were compared with the forecasting of the hybrid TF-SMO model.

5-The Mean Squared Error (MSE) for the TF model and the hybrid TF-SMO model was calculated.

6- Measuring the performance and efficiency of the SMO algorithm.

The figure below represents the steps that were used to predict the TF-SMO hybrid model:

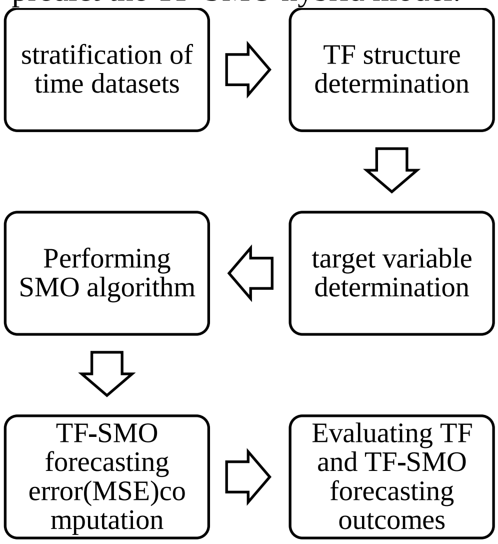



Fig 3.Framework plot of TF-SMO

**2.4. Forecasting error measurement**

Many measures exist to measure the accuracy of forecasting, including the Mean Square Error, which can be expressed in the form that follows [15]:

$$\text{MSE} = \frac{\sum_{t=1}^{n}(Y_t - \hat{Y})^2}{n} \quad (25)$$

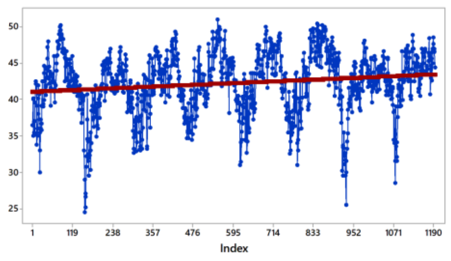
**3. Result and discussion**

**3.1.** ***Dataset description***

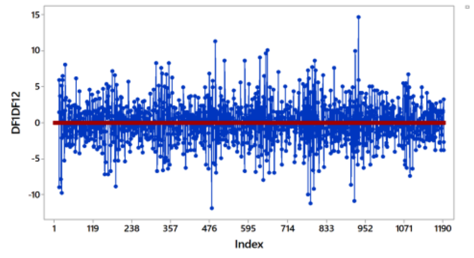
In this research, daily climatic time series data for maximum and minimum temperatures and evaporation for the province of Baghdad from 2012 to 2022 were used, taken from the General Authority for Meteorology and Seismic Monitoring. The temperature and evaporation data for the hot and cold seasons were divided into two groups. The first group includes most of the data at the beginning of the series and is called the Training period dataset. This data is used to build the models. The second group includes less data and is usually used to test the models built using the training data. This group is called the Testing period dataset. The ratio of training data to testing data is usually 80% to 20% or 70% to 30%, and it may increase or decrease depending on the nature of the data or the study. In this research, the data was divided into 65% for the training period and approximately 35% for the testing period. The number of observations for the hot season was 1683, with 1224 observations used for the training period and 459 observations for the testing period. The number of observations for the cold season was 1661 with 1208 observations used for the training period and 453 observations for the testing period.

### 3.2 *TF model*

In this research, maximum and minimum temperature and evaporation data for the Baghdad Governorate were used, obtained from the General Authority for Meteorology and Seismic Monitoring for the period. (2012-2022). Based on the data for maximum and minimum temperatures as input variables for the transfer function TF, which affects the output variable, evaporation. Where Y represents the output variable, which is evaporation, and$U_1$ represents the input variable for maximum temperature, and $U_2$ represents the input variable for minimum temperature. Two models of transfer functions were used: one for the hot season and one for the cold season.A ready-made software was used in the practical aspect of the transfer function models, which is the Statistical System SCA from a company (Scientific Computing Associates Crop.) From the United States of America. The data for maximum and minimum temperatures and evaporation for the hot and cold seasons were plotted and instability in the series was observed. Therefore, the first difference and seasonal difference were taken for the input and output variables, as shown in the following figures:

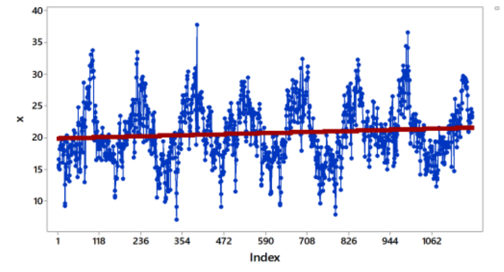



a

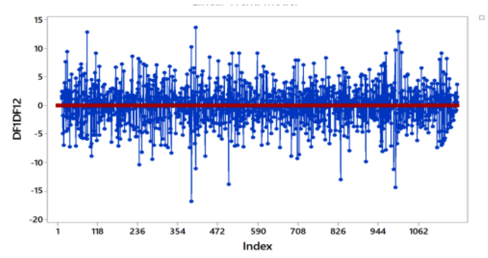



b

Fig (4). (a) represents the time series of maximum temperature for the hot season, and (b) represents the same series after taking the first difference and seasonal difference12.

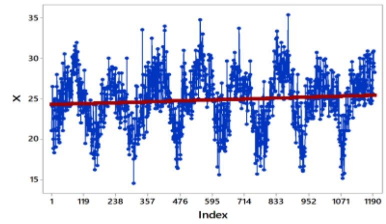



a

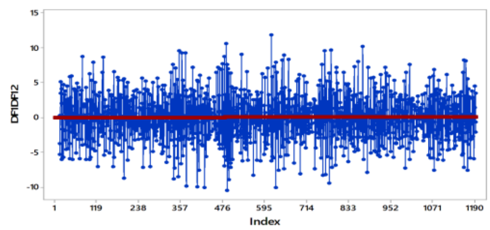



b

Fig (5). (a) represents the time series of maximum temperature for the cold season, and (b) represents the same series after taking the first difference and seasonal difference12.

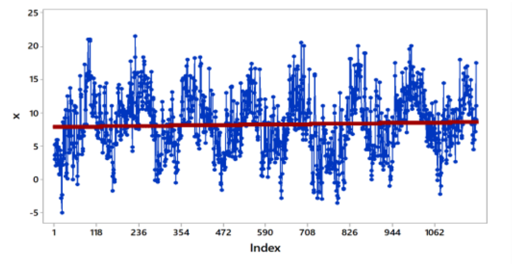



a

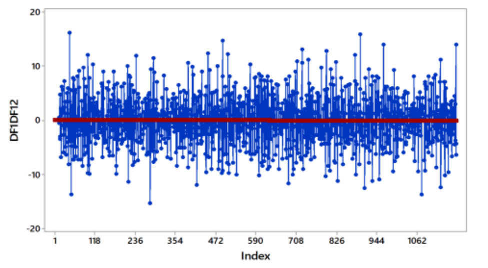



b

Fig (6).(a) represents the time series of minimum temperature for the hot season and (b) represents the same series after taking the first difference and seasonal difference12.

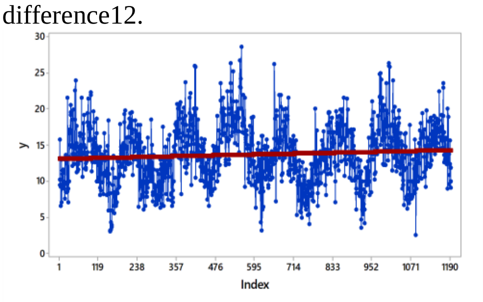



a

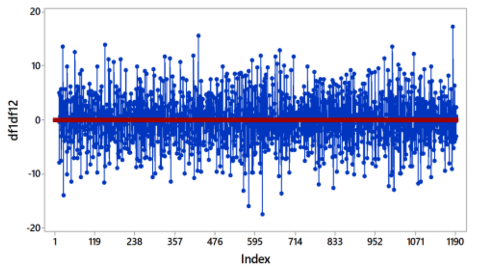



b

Fig (7).(a) represents the time series of minimum temperature for the cold season, and (b) represents the same series after taking the first difference and seasonal difference12.

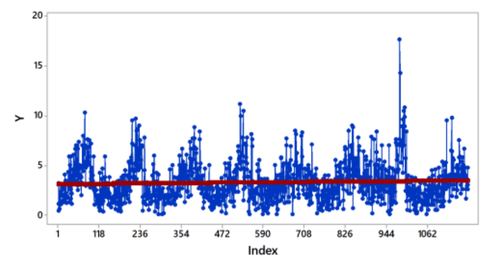



a

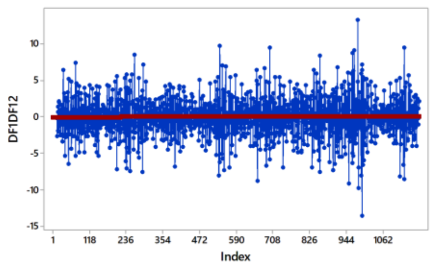



b

Fig (8). (a) represents the time series of evaporation for the hot season, and (b) represents the same series after taking the first difference and the seasonal difference12.




a




b

Fig (9). (a) represents the time series of evaporation for the cold season, and (b) represents the same series after taking the first difference and the seasonal difference 12.

After the data was plotted, the differences were taken, and a stationary time series was obtained, Liu's method, which relies on the SCA statistical system, was used. The transfer function model for the hot and cold seasons was obtained, its parameters were estimated, and the predictive values for the evaporation phenomenon (the dependent variable) were obtained. The model formula for the hot season can be expressed by the following equation:

$$Y = 0.0900U_{1t-1} + 0.1859U_{2(t-1)} + 0.0.1235U_{2(t-2)} + 0.0787U_{2(t-3)} + a_t - 0.3292 * a_{t-1} \quad (26)$$

Where Y represents the output variable, $U_1$is the input variable for maximum temperature, $U_2$ represents the input variable for minimum temperature and $(a_t - 0.3292 * a_{t-1})$ represents the disturbance term $N_t$.

Similarly, the transfer function model was found and its parameters were estimated for the cold season and the model formula for the cold season can be expressed through the following equation:

$$Y = 0.1320U_{1t-1} - 0.9546 * 0.1320 * U_{1(t-2)} - 0.9546 * 0.0445U_{1(t-3)} + (-0.0469) * U_{2(t-1)} - 0.9546 * (-0.0469) * U_{2(t-2)} + a_t - 0.8126 * a_{t-1} \quad (27)$$

Where Y represents the output variable, $U_1$is the input variable for maximum temperature, $U_2$ represents the input variable for minimum temperature and $(a_t - 0.8126 * a_{t-1})$represents the disturbance term $N_t$.

Table (1) below includes the MSE values for forecasts errors using the TF model for the training and testing periods for both seasons for maximum and minimum temperature and evaporation data.

**Table 1.MSE values of TF model for hot and cold season**

| | Training | Testing |
|---|---|---|
| **Hot** | 13.4895 | 13.5695 |
| **Cold** | 3.1879 | 3.5432 |

It can be observed from Table (1) that the MSE values for the hybrid TF-SMO model for the input data of maximum and minimum temperatures and the output data of evaporation for the hot and cold seasons during the training period were better than during the testing period. This indicates that the hybrid model was able to improve the forecast accuracy of the TF models, as it provided forecast results with lower MSE.

### 3.3. *TF-SMO*

After obtaining the TF transfer function models for the hot and cold seasons, the variables present in the TF models are used as inputs in the SMO algorithm to obtain the hybrid TF model. The forecasting method using the hybrid model can be explained in the following steps:

1.The data for the hot and cold seasons were stratified with the hot season including the months of May to September, while the cold season included the months of November to March. The months of April and October, which fluctuate between hot and cold from year to year, were neglected.

2. The data was divided into two groups for each season: the first group is the training data and the second group is the test data.

3. Introducing the response variable, which is the evaporation variable Y. Introducing the response variable, which is the evaporation variable Y.

4.Input the input variables and parameters based on TF models in equations (26) and (27) in MATLAB separately as a matrix for the input variables, a vector for the dependent variable, and a vector for the forecasting obtained from the SCA program.

5. Obtaining a specific forecasting vector for the hybrid model. Obtaining a special vector for predictions for the hybrid model.

6. Measuring the performance and efficiency of the hybrid TF-SMO model.

7. Using the Mean Squared Error (MSE) forecasting metric for the TF-SMO model.

8.The forecasting results of the TF models are compared with the forecasting results of the hybrid TF-SMO model.

The framework for implementing the TF-SMO model on the data can be illustrated with the following diagram:

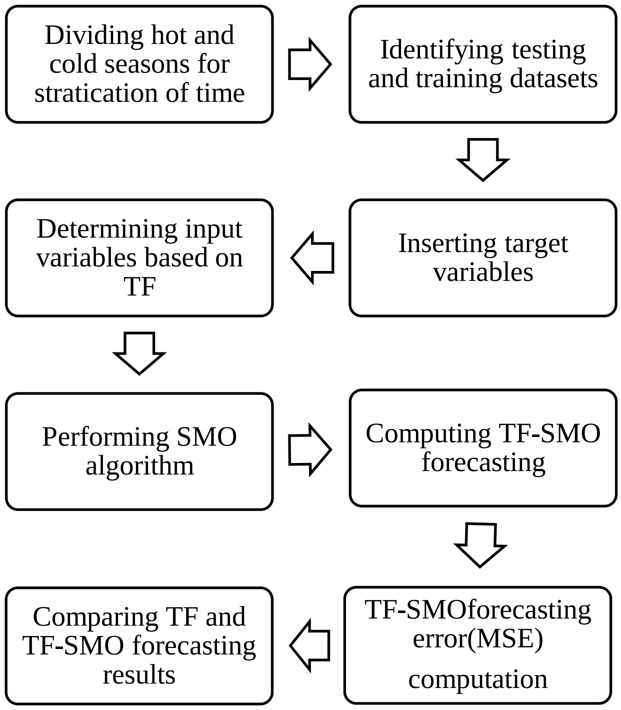



Fig 10.The framework for applying datasets by using TF-SMO

Table (2) includes the MSE values for forecast errors using the TF-SMO model for the training and testing periods for both seasons for maximum and minimum temperature and evaporation data.

**Table 2. MSE values of TF-SMO model for hot and cold season**

| | Training | Testing |
|---|---|---|
| **Hot** | 0.0794 | 0.0824 |
| **Cold** | 0.1831 | 0.2114 |

It can be observed from Table (2) that the MSE values for the hybrid TF-SMO model for the input data of maximum and minimum temperatures and the output data of evaporation for the hot and cold seasons during the training period were better than during the testing period. This indicates that the hybrid model was able to improve the forecast accuracy of the TF models, as it provided forecast results with lower MSE.

Table (3) includes the MSE values for forecast errors using the TF and TF-SMO models for both seasons for the generated data on maximum and minimum temperatures and evaporation.

**Table 3.** MSE values of TF-SMO model for hot and cold season for generated **data**

| | TF | TF-SMO |
|---|---|---|
| **Hot** | 1.7702 | 0.0528 |
| **Cold** | 1.5280 | 0.1754 |

It can be observed from Table (3) that the MSE values for the hybrid TF-SMO model for the input data of maximum and minimum temperatures and the output data of evaporation for the hot and cold seasons were better than the TF model. This indicates that the hybrid model was able to improve the forecast accuracy of the TF models, as it provided forecast results with lower MSE.

## 4. Conclusions

In this research, the TF model was used as a traditional linear method to forecast maximum and minimum temperatures and evaporation. The combination of TF and the SMO algorithm was proposed as a nonlinear intelligent method to improve the forecasting of the transfer function. Time stratification was used to divide the data into hot and cold seasons, obtained from meteorological observations, to achieve homogeneity in the studied data. The data for the hot and cold seasons were divided into two groups: the first for the training period and the second for the testing period. The forecasting results of the hybrid TF-SMO model were better than the traditional method. It can be concluded that the proposed hybrid method for forecasting maximum and minimum temperatures and evaporation was more accurate in prediction results and the TF-SMO hybrid method can be used with any type of data.

## References

[1] Sentas, A., & Psilovikos, A. (2022). Comparison of ARIMA and transfer function (TF) models in water temperature simulation in dam—lake Thesaurus, eastern Macedonia, Greece. In Environmental Hydraulics. Volume 2 (pp. 929-934). CRC Press.

[2] Cho, M. Y., Hwang, J. C., & Chen, C. S. (1995, November). Customer short term load forecasting by using ARIMA transfer function model. In Proceedings 1995 International Conference on Energy Management and Power Delivery EMPD'95 (Vol. 1, pp. 317-322). IEEE.

[3] Yang, J. F., Zhai, Y. J., Xu, D. P., & Han, P. (2007, August). SMO algorithm applied in time series model building and forecast. In 2007 International Conference on Machine Learning and Cybernetics (Vol. 4, pp. 2395-2400). IEEE.

[4] Khair, A. F., Awang, M. K., Zakaraia, Z. A., & Mazlan, M. (2017). Daily streamflow prediction on time series forecasting. Journal of Theoretical and Applied Information Technology, 95(4), 804.

[5] Ragab, M. (2022). Spider Monkey Optimization with Statistical Analysis for Robust Rainfall Prediction. *Computers, Materials & Continua*, *72*(2).

[6] Malig, B. J., Pearson, D. L., Chang, Y. B., Broadwin, R., Basu, R., Green, R. S., & Ostro, B. (2016). A time-stratified case-crossover study of ambient ozone exposure and emergency department visits for specific respiratory diagnoses in California (2005–2008). Environmental health perspectives, 124(6), 745-753.

[7]. ALbazzaz, Z. M., & Shukur, O. B. (2024, June). Using LSTM Network Based on Logistic Regression Model for Classifying Solar Radiation Time Series. In International Conference on Explainable Artificial Intelligence in the Digital Sustainability (pp. 375-388). Cham: Springer Nature Switzerland.

[8] Box, G., Jenkins, G.., Reinsel, G., & Ljung G. (2016). Time series analysis: forecasting and control. John Wiley & Sons,Inc.Hoboken,New Jersey.

[9] liu,L.M.,(2006),"Time series analysis and forecasting ",2nd ed.,Scientific Computing Associates Crop.,Illinois,USA.

[10] Jenkins, G. M., & Box, G. E. (1976). Time series analysis: forecasting and control.

[11] Wei, W. W. (2006). Time series analysis: univariate and multivariate. Methods. Boston, MA: Pearson Addison Wesley

[12]Yaffee, R. A., & McGee, M. (2000). An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®. Elsevier.

[13]Bousnane, S. (2019). Parallelization of Spider Monkey Optimization (SMO) algorithm (Doctoral dissertation, Ministry of Higher Education).

[14] Sharma, H., Hazrati, G., & Bansal, J. C. (2019). Spider monkey optimization algorithm. Evolutionary and swarm intelligence algorithms, 43-59.

[15] Shukura, O. B. (2020). Using the MLR and neuro-fuzzy methods to forecast air pollution datasets. Int J Adv Sci Eng Inform Technol, 10(4), 1457-1464.