Turkish Journal of Forecasting vol. 8 no. 1 (2024) pp. 23-39

Content list available at JournalPark

Turkish Journal of Forecasting

Metaheuristic Algorithms to Forecast Future Carbon Dioxide Emissions of Turkey

O.A. Arık[1,*], G. Canbulut[2], E. Köse[2]

[1]*Erciyes University, Faculty of Engineering, Department of Industrial Engineering, 380180 Kayseri, Turkey*
[2]*Nuh Naci Yazgan University, Faculty of Engineering, Department of Industrial Engineering, 38170 Kayseri, Turkey*

A R T I C L E   I N F O

Article history:

Received	10	November	2023
Accepted	16	February	2024

Keywords:
Genetic algorithm
Artificial bee colony
Simulated annealing
Hybrid metaheuristics
Carbon dioxide emission,
Forecasting,
Multiple linear regression.

R E S E A R C H   A R T I C L E

A B S T R A C T

This paper proposes the use of five different metaheuristic algorithms for forecasting carbon dioxide emissions (MtCO2) in Turkey for the years between 2019 and 2030. Historical economic indicators and construction permits in square meters of Turkey between 2002 and 2018 are used as independent variables in the forecasting equations, which take the form of two multiple linear regression models: a linear and a quadratic model. The proposed metaheuristic algorithms, including Artificial Bee Colony (ABC), Genetic Algorithm (GA), Simulated Annealing (SA), as well as hybrid versions of ABC with SA and GA with SA, are used to determine the coefficients of these regression models with reduced statistical error. The forecasting performance of the proposed methods is compared using multiple statistical methods, and the results indicate that the hybrid version of ABC with SA outperforms other methods in terms of statistical error for the linear equation model, while the hybrid version of GA with SA performs better for the quadratic equation model. Finally, four different scenarios are generated to forecast the future carbon dioxide emissions of Turkey. These scenarios reveal that if construction permits and the population is strictly managed while the economical wealth of Turkey keeps on improving, the CO2 emissions of Turkey may be less than in other possible cases.

1. Introduction

Greenhouse gases trapped in the atmosphere cause the earth's temperature to rise, resulting in climate change or global warming. Rapid industrialization, fuelled by increased use of fossil fuels, is a major contributor to this phenomenon. Growing awareness of the long-term consequences of global warming has led to international efforts to reduce greenhouse gas emissions, including carbon dioxide (CO2), methane, nitrous oxide, and fluorinated gases, of which CO2 is the most prevalent in the atmosphere. Fossil fuel combustion is the primary source of CO2 emissions, largely due to electricity generation, transportation, industrial processes, residential and commercial fuel use, and agriculture. Energy demand in the industrial and residential sectors, as well as population growth, drives the use of fossil fuels. The building sector, which requires high energy inputs for materials such as steel and cement, has a significant impact on CO2 emissions. This study uses the gross domestic product (GDP), export, import, population, and construction permits of Turkey as independent variables in estimating CO2 emissions. Halicioglu [1] found that income is the most significant variable in explaining carbon emissions in Turkey, followed by energy consumption and foreign trade. The number of construction permits, measured in square meters, reflects the energy-related CO2 emissions, while other economic indicators such as GDP, export/import, and population also impact energy-related CO2 emissions. The Turkish economy and its energy demand have been and are expected to continue having an

* Corresponding author.
E-mail addresses: oguzhanahmet@erciyes.edu.tr (Oğuzhan Ahmet Arık)

upward trend. The significant energy demand of the Turkish economy is mainly fulfilled by fossil fuels such as coal, fuel oil, and natural gas. Although Turkey has alternative energy sources like hydro, geothermal, wind, and solar power plants, the total electricity generated from these sources is still less than that generated from thermal power plants. While Turkey currently lacks nuclear power plants, it plans to build new ones in the near future. Decreasing CO2 emissions is critical in the fight against climate change and developed countries are committed to reducing CO2 emissions by decreasing their reliance on fossil fuels in energy production. As a developing country, Turkey's growth is dependent on energy and its energy needs have rapidly increased as its production has grown since the 1980s. However, this has also led to an uncontrolled balance of payments and a current account deficit, which is a major macroeconomic problem for Turkey [2]. Therefore, this study aims to examine the relationship between CO2 emissions in Turkey and its economic indicators, population, and its major business area, the building sector. In the future, if Turkey increases its use of alternative energy sources, a new forecasting model will be needed to predict its CO2 emissions.

This study is to analyse the intricate relationship between economic indicators, population dynamics, and the influential building sector on CO2 emissions in Turkey. The primary goal of this study is to establish a robust forecasting model that elucidates the impact of these factors on carbon emissions, facilitating informed policy-making and guiding Turkey towards sustainable energy pathways. In the global pursuit to combat climate change, understanding the nuanced interplay between economic growth, population dynamics, and key sectors like construction in driving CO2 emissions is pivotal. Turkey stands at a crucial juncture, grappling with rapid industrialization and surging energy demands, predominantly met by fossil fuels. Forecasting CO2 emissions isn't just about foreseeing environmental impact; it's a compass guiding Turkey's sustainable development trajectory. By unravelling these relationships, we can pave the way for strategic interventions, prioritize alternative energy sources, and forge a blueprint for a greener, more resilient future. Ultimately, this study aims not only to forecast CO2 emissions but to empower Turkey in making informed, sustainable choices that safeguard the environment while nurturing continued economic growth. In this study, we employ two multiple linear regression models to estimate Turkey's CO2 emissions. The independent variables in the linear and quadratic forecasting equations are GDP, exports, imports, population, and construction permits, while the dependent variable is Turkey's CO2 emission in MtCO2. To calculate the forecasting coefficients, we use three metaheuristics named Artificial Bee Colony (ABC), Genetic Algorithm (GA), and Simulated Annealing (SA) and two hybrid metaheuristics (ABC-SA and GA-SA). To compare the results, we use four statistical measures: mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R-squared (R^2). The results show that the ABC-SA algorithm outperforms others in terms of statistical error for the first linear model, while GA-SA is the best choice for the second quadratic model. The remainder of the paper is as follows: Sect. 2 presents a literature review about forecasting CO2 emissions. Section 3 introduces our forecasting equations and data. Section 4 presents solution approaches. Section 5 includes the experimental results and comparison of solution approaches. Section 6 presents our forecasting scenarios for CO2 emissions of Turkey. Section 7 draws a conclusion and future projection for the reader.

2. Literature Review

In this section, we present a brief literature review about forecasting CO2 emissions. Since each country has its own strategy to decrease CO2 emissions, forecasting CO2 emissions have been investigated by researchers to help policymakers with their strategies about CO2 emissions. Behrang et al. [3] proposed a two-step approach for forecasting world CO2 emissions. In the first step of their research, they used the bee colony algorithm to forecast the world's oil, natural gas, coal, and primary energy consumption by using historical data for the world's population, GDP, oil trade movement, and natural gas trade movement. In the second step, they used artificial neural networks (ANN) to forecast world CO2 emissions by using the forecasted world's oil, natural gas, coal, and primary energy consumption. Chang et al. [4] proposed a novel quantum harmony search algorithm-based discounted mean square forecast error combination model to forecast the world's CO2 emissions. Sun et al. [5] present a grey prediction model to forecast CO2 emissions in China. Samsami [6] applied GA, particle swarm optimization (PSO), and ant colony optimization techniques to forecast NOx emission in Iran based on the values of oil, natural gas, coal, and primary energy consumption. Abdullah & Mohd Pauzi [7] proposed a multi-layer perceptron type of ANNs by using different learning algorithms to forecast CO2 emissions in Malaysia. Wang et al. [8] proposed a hybrid nonlinear grey-prediction and quota allocation model for supporting optimal planning of China's carbon intensity reduction at both departmental and provincial levels in 2020. Sun and Xu [9] used an improved PSO algorithm to train backpropagation ANN to forecast CO2 emissions of Hebei Province in China. Zhao et al. [10] proposed a hybrid method including a whale optimization algorithm and improved least squares support vector machine to forecast China's CO2 emissions by using GDP, energy consumption, and population as independent variables. Wen and Liu

[11] used the PSO algorithm to train backpropagation ANN to forecast CO2 emissions of Beijing Province in China. Sun and Sun [12] proposed a novel hybrid model that combined principal component analysis with regularized extreme learning machine to make CO2 emissions in China. Sun et al. [13] proposed an extreme learning machine to predict CO2 emissions and they also used a PSO algorithm to train their extreme learning machine. Baareh [14] proposed genetic programming to predict CO2 emissions by using global oil, natural gas, coal, and primary energy consumption. Zhao et al. [15] suggested a hybrid of the mixed data sampling regression model and backpropagation neural network to forecast carbon dioxide emissions of the USA. Assareh and Nedaei [16] proposed a two-step approach for forecasting world CO2 emissions. In the first step of their research, they used Ray Optimization (RO) algorithm to forecast the world's oil, natural gas, coal, and primary energy consumption by using historical data for the world's population, GDP, oil trade movement, and natural gas trade movement. In the second step, they used ANN to forecast world CO2 emissions by using forecasted data by the RO algorithm. Dai et al. [17] proposed a grey model and a least-squares support vector machine model for forecasting CO2 emissions in China accurately. Zhao et al. [18] proposed a hybrid approach including a swarm algorithm and least squares support vector machine model to forecast CO2 emissions in China. They used GDP, population, energy consumption, economic structure, energy structure, urbanization rate, and energy intensity as the input variables in their forecasting equation. Ameyaw and Yao [19] implemented a bidirectional long short-term memory sequential algorithm formulation for CO2 emissions in five west African countries. Guo et al. [20] forecasted CO2 emissions and CO2 intensities in China during 2030 by using three scenarios, seven indicators, and a back-propagation neural network. Lin et al. [21] proposed a two-stage forecasting approach consisting of multivariable grey forecasting model and genetic programming to forecast CO2 emissions in Taiwan.

Ameway et al. [22] investigated the economic growth and fossil fuel combustion of the USA, China, Canada, and Nigeria to forecast CO2 emissions. They proposed a long short-term memory algorithm to forecast selected countries' CO2 emissions. Hosseini et al. [23] used multiple linear regression and multiple polynomial regression models to forecast Iran's CO2 emissions. Huang et al. [24] applied the Elman neural network optimized by the Firefly Algorithm to forecast the CO2 emissions in China considering urbanization level, GDP of secondary industry, thermal power generation, real GDP per capita, and energy consumption per unit of GDP. Qiao et al. [25] proposed a novel hybrid algorithm, which combines lion swarm optimizer and GA to optimize the traditional least squares support vector machine model and they forecasted CO2 emissions of developed countries. Wu et al. [26] used a conformable fractional non-homogeneous grey model to forecast CO2 emissions in BRICS (Brazil, Russia, India, China, and South Africa) countries. Wu and Meng [27] proposed a new prediction model, which combines t-distribution, Gaussian perturbations bat algorithm, and a least-squares support vector machine to forecast CO2 emissions in China. Malik et al. [28] proposed an autoregressive integrated moving average to forecast CO2 emissions of Pakistan for different scenarios.

Ozturk and Acaravci [29] examined the long-run and causal relationship issues between economic growth, carbon emissions, energy consumption, and employment ratio in Turkey by using an autoregressive distributed lag-bounds testing approach of cointegration. Hamzacebi and Karakurt [30] proposed a grey model for forecasting energy-related CO2 emissions in Turkey in the years between 2013 and 2025. Bildirici and Bakirtas [31] analyze the relationship between carbon dioxide (CO2) emissions, economic growth, and coal and oil consumption in Brazil, Russia, India, China, Turkey, and South Africa by using the bounds test approach autoregressive distributed lag over the period from 1969 to 2011. Şahin [32] forecasted Turkey's electricity generation and CO2 emissions between 2017 and 2021 by using past data including capacity factors of thermal, hydro, geothermal, wind, and solar power plants from 2006 to 2016. Şahin [33] combined linear and nonlinear metabolic grey models with the optimization technique to obtain more accurate forecasting results for CO2 emissions in Turkey. Bakay and Ağbulut [34] forecast the GHG emissions of greenhouse gases using deep learning, support vector machine, and ANN algorithms from the electricity production sector in Turkey.

Ullah et al. [35] used a Quantile-on-Quantile regression approach to determine the dynamics between environmental taxes and ecological sustainability for top-seven green economies. Adebayo and Samour [36] investigated fiscal policy's impact on the load capacity factor in BRICS nations from 1990 to 2018. They find that economic growth and non-renewable energy sources contribute to environmental deterioration, while increased renewable energy promotes sustainability. Moreover, positive taxation revenue shocks improve environmental quality, whereas positive or negative government expenditure shocks decrease it. They advocated for leveraging fiscal policy in BRICS nations to prioritize renewable energy investments for enhanced ecological sustainability. Radmehr et al. [37] examined the impact of green technological innovation (GTI) and renewable energy (REC) on ecological sustainability across 20 EU nations from 1995 to 2018 using spatial panel econometrics. They found that both GTI and REC significantly improve domestic ecological sustainability and that neighboring nations' high levels of GTI, REC, and human capital benefit environmental quality. However, economic growth and financial globalization negatively impact environmental quality, with financial globalization indirectly contributing to increased ecological footprint. Kartal et al. [38] delved into the impact of nuclear and renewable energy consumption, economic growth, and financial development on ecological quality in the US. Using innovative causal analysis methods, they found that nuclear and renewable energy, alongside financial development, mitigate ecological deterioration in middle and higher levels, while economic growth negatively affects ecological quality in higher quantiles. Ultimately, their study emphasizes the significance of policies favoring nuclear energy transition for improved ecological sustainability and environmental quality.

This literature review shows that researchers have utilized ANNs, grey forecasting models, support vector machine models, and autoregressive integrated moving average models to forecast CO2 emissions globally or for specific countries. Common independent variables include economic indicators such as GDP, population, exports, imports, and building sector data. These indicators have proven to be effective in forecasting CO2 emissions with lower statistical error, as seen in their frequent use by researchers worldwide. Although metaheuristic algorithms are less commonly used for forecasting CO2 emissions, they have often been combined with other methods such as ANNs, support vector machines, and grey forecasting models. ANNs, support vector machines, and other machine learning approaches are capable of modelling nonlinear relationships between independent and dependent variables, while the autoregressive integrated moving average model is specifically designed for forecasting time series. The grey forecasting model is ideal for short-term forecasts with limited data. Regression analysis is a statistical process that seeks to identify the relationship between the dependent variable and independent variables. The most basic form of regression analysis is linear regression, which tries to fit the data to a line by minimizing the difference between the data and the line. However, linear regression assumes independence between variables and cannot account for multi-dependence. When selecting a forecasting method, factors such as data availability, projection period, and desired accuracy must be considered. Each method has its own strengths and weaknesses and the choice will depend on the specific situation.

In this study, we propose two multiple linear forecasting models counting on construction permits of Turkey in years since the buildings and construction sector has a large amount of the final energy use and energy-used related CO2 emissions of which have been results of producing building materials, cement, glass, and steel. According to the reported literature, no published paper considers the buildings and construction sector while building a forecast equation for CO2 emissions. To determine these models' coefficients, we apply metaheuristic algorithms such as ABC, SA, GA, and hybrids of them. Our forecasting equations are multiple linear regression models and the problem is polynomial time. Therefore, the complexity of the problem seems to be not required using metaheuristics which are mainly used for solving complex NP-hard problems. The linear regression models use the least squares method to fit the model by minimizing the squares of forecasting errors. However, no guarantee minimizing squares of errors presents the best forecasting equations so other forecasting performance indicators such as absolute values of errors and percentage deviations of errors should be considered simultaneously while generating the forecasting equation's coefficients. In this study, we apply strong metaheuristics to determine these coefficients in our models considering squares of errors, absolute values of errors, and percentage deviation of errors to forecast the CO2 emissions of Turkey. Even though other well-known methods such as ANN, support vector machine, deep learning, and grey forecasting algorithms have been applied to forecasting CO2 emissions, metaheuristics approaches that use data from the buildings and construction sector are first applied to the problem. The future projections with different scenarios in this study about the CO2 emissions of Turkey obtained with the proposed algorithms can be used in comparison for the performance of forecasting approaches.

3. Forecasting Equations and Data

In this section, we propose two multiple linear regression models as linear and quadratic forecasting equations to forecast CO2 emissions in Turkey. We used Turkey's historical data between 2002 and 2018 for GDP, population, export, import, and construction permits as independent variables in forecasting equations. These data are given in Table 1. Historical data given in Table 1 are collected from different sources (EnerData [35], WorldBank [36], and TurkStat [37]).

Table 1: Historical data of Turkey for forecasting equations

Year	CO_2 (MtCO2) [39]	GDP (10^9 USD) [40]	Population (10^3) [41]	Export (10^9 USD) [41]	Import (10^9 USD) [41]	Construction Permit (10^6 m^2) [41]
2002	192.919	238.428	66003	36.059	51.554	36.187
2003	203.722	311.823	66795	47.253	69.340	45.516
2004	210.297	404.787	67599	63.167	97.540	69.720
2005	217.949	501.416	68435	73.476	116.774	106.425
2006	242.649	552.487	69295	85.535	139.576	122.910
2007	267.292	675.770	70158	107.272	170.063	125.067
2008	267.616	764.336	71052	132.027	201.964	103.846
2009	265.316	644.640	72039	102.143	140.928	100.727
2010	269.829	771.902	73142	113.883	185.544	176.429
2011	290.997	832.524	74224	134.907	240.842	123.622
2012	303.627	873.982	75176	152.462	236.545	158.750
2013	292.015	950.579	76148	151.803	251.661	175.808
2014	314.110	934.186	77182	157.610	242.177	220.654
2015	326.542	859.797	78218	143.839	207.234	189.675
2016	347.102	863.722	79278	142.530	198.618	206.972
2017	377.077	852.677	80313	156.993	233.800	287.334
2018	385.264	771.350	81407	167.921	223.047	148.155

Let us assume $F_t(\vec{\beta})$ is the dependent variable that illustrates the forecasted CO2 emission of Turkey in year t by using $\vec{\beta}$ coefficient vector for independent variables and $\vec{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$ where k is the number of coefficients. In our proposed forecasting equations, five independent variables are used as follows:

$$F_t(\vec{\beta}) = \beta_0 + \sum_{i=1}^{5} \beta_i X_{i,t} \quad (1)$$

$$F_t(\vec{\beta}) = \beta_0 + + \sum_{i=1}^{5} \beta_i X_{i,t} + \beta_6 X_{1,t} X_{2,t} + \beta_7 X_{1,t} X_{3,t} + \beta_8 X_{1,t} X_{4,t} + \beta_9 X_{1,t} X_{5,t} + \beta_{10} X_{2,t} X_{3,t} + \beta_{11} X_{2,t} X_{4,t} + \beta_{12} X_{2,t} X_{5,t} + \beta_{13} X_{3,t} X_{4,t} + \beta_{14} X_{3,t} X_{5,t} + \beta_{15} X_{4,t} X_{5,t} \quad (2)$$

where $X_{1,t}$, $X_{2,t}$, $X_{3,t}$, $X_{4,t}$ and $X_{5,t}$ are GDP, population, export, import, and construction permit of Turkey in year t. Eq. (1) and Eq. (2) express linear and quadratic forecasting equations, respectively. We refer to Ozdemir et al. [42] for the linear equation and Toksarı [43] for the quadratic equation in this study. As seen from the equations, there are 6 coefficients in the linear equation and 16 coefficients in the quadratic equation. The dependent variables in forecasting equations are strongly correlated with CO2 emissions in Turkey. Table 2 shows Pearson Correlation coefficients between CO2 and dependent variables considering the data in Table 1. All independent variables are positively correlated to the CO2 emissions of Turkey. If any of these independent variables is increased, we expect that the CO2 emissions of Turkey will be increased.

Table 2: Pearson Correlation coefficients between CO_2 and dependent variables

Pairs	Pearson correlation
CO_2 - GDP	0.815
CO_2 – Population	0.981
CO_2 – Export	0.917
CO_2 – Import	0.824
CO_2 – Construction permit	0.835

4. Solution Approaches

In this section, we propose three metaheuristic approaches and two-hybrid algorithms to determine coefficients of independent parameters in forecasting equations. We use the solution encoding schema that is proposed by Arık [2] and Toksarı [43] for our metaheuristic approaches. The performance criterion of the proposed algorithms is the mean square error (MSE) of forecasted CO2 emissions of Turkey as follows:

$$f(\vec{\beta}) = \frac{1}{n}\sum_{t=1}^{n}(F_t(\vec{\beta}) - A_t)^2 \quad (3)$$

where $f(\vec{\beta})$ is the MSE value of the method's results obtained with $\vec{\beta}$ coefficient vector, A_t is actual CO2 and n is the number of years in the experiment. Besides MSE values, other statistics are also used in comparisons. These are MAPE, MAE, and R2 statistics. These statistics are calculated as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2 \quad (4)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \quad (5)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|/A_t \quad (6)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(A_t - F_t)^2}{\sum_{t=1}^{n}(A_t - \bar{A})^2} \quad (7)$$

where n is the total number of years in the investigated period, A_t is the actual CO2 emission of Turkey in year t, F_t is the forecasted CO2 emission of Turkey in year t and $\bar{A}$ is the average value of actual CO2 emissions of Turkey in the investigated period.

4.1. Artificial Bee Colony Algorithm

ABC algorithm is an optimization algorithm based on a particular intelligent behavior of honeybee swarms [44]. ABC is a population-based and swarm-intelligence metaheuristic. Each induvial in the population indicates a food position. These individuals are evaluated with artificial bees to discover the best food resource or the area of good food resources. Each induvial is indexed with $j \in \{1,2,3,\ldots,m\}$ where m is the population size and each individual j is a coefficient vector $\overrightarrow{\beta_j} = (\beta_{j0}, \beta_{j1}, \beta_{j2}, \ldots, \beta_{jk})$ where β_{ji} is the ith coefficient in the coefficient vector $\overrightarrow{\beta_j}$. This solution representation is common for other metaheuristic algorithms in this study. The flowchart of the ABC algorithm is given in Fig. 1. Eq. (4) is for the initialization of β_{ji} $\forall$ i,j for all metaheuristics in this study as follows:

$$\beta_{ji} = l_i + rand(0,1) * (u_i - l_i) \; \forall i,j \quad (8)$$

where l_i and u_i are the lower and upper bounds of the parameter β_{ji}, respectively.

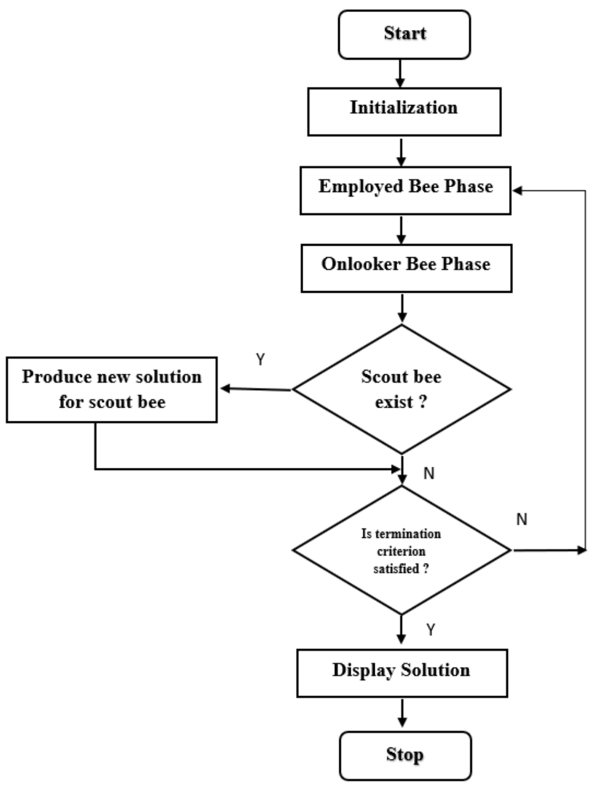

Fig. 1 The flowchart of the ABC algorithm (Arık [47])

Employed bees search for a new set of coefficients ($\vec{\upsilon}_j$) as a neighbor of the existing set of coefficients ($\vec{\beta}_j$). Employed bees find a new $\vec{\upsilon}_j$ and they evaluate their fitness. Eq. (9) is to determine a $\vec{\upsilon}_j$ by using existing $\vec{\beta}_j$.

$$\nu_{ji} = \beta_{ji} + \phi_{ji}(\beta_{ji} - \beta_{li}) \; \forall i, j \tag{9}$$

where $\vec{\beta}_l$ is a randomly selected set of coefficients, i is the index of coefficient that is selected randomly in $\vec{\beta}_j$ and ϕ_{ji} is a random real number where $\phi_{ji} \in [-1,1]$. After producing a new set of coefficients $\vec{\upsilon}_j$, its fitness is calculated and a greedy selection is applied between $\vec{\upsilon}_j$ and $\vec{\beta}_j$. The fitness value of $\vec{\beta}_j$ is illustrated with $fitness_j(\vec{\beta}_j)$ notation and it is calculated as follows:

$$fitness_j(\vec{\beta}_j) = 1/\left(1 + f(\vec{\beta}_j)\right) \tag{10}$$

In the ABC algorithm, employed bees give information about their own sets of coefficients to onlooker bees. Then, onlooker bees start selecting probabilistically their sets by using the feedback from employed bees. This selection phase is done with a fitness-based selecting technique and the probability p_j of the set $\vec{\beta}_j$ can be determined as follows:

$$p_j = \frac{fitness_j(\vec{\beta}_j)}{\sum_{j=1}^{m} fitness_j(\vec{\beta}_j)} \; \forall j \tag{11}$$

After a set $\vec{\beta}_j$ for an onlooker bee is probabilistically chosen, a neighborhood set $\vec{\upsilon}_j$ is determined by using Eq. (9), and its fitness value is computed by using Eq. (10). As in the employed bee's phase, a greedy selection is applied between $\vec{\upsilon}_j$ and $\vec{\beta}_j$. Thus, the number of onlooker bees recruiting better solution spaces is increased. In this phase, to disable to inefficient set, a counter $failure_j(\vec{\beta}_j)$ for each $\vec{\beta}_j$ takes places. If $fitness_j(\vec{\beta}_j)$ is better than $fitness_j(\vec{\upsilon}_j)$, then $failure_j(\vec{\beta}_j)$ increases one. If $fitness_j(\vec{\beta}_j)$ is not better than $fitness_j(\vec{\upsilon}_j)$, then $failure_j(\vec{\beta}_j)$ is set as zero. If $failure_j(\vec{\beta}_j)$ reaches a pre-determined $limit$, then the employee bee dealing with $\vec{\beta}_j$ becomes a scout bee that abandons $\vec{\beta}_j$ and finds a random set of coefficients by using Eq. (11). The number of scout bees is generally limited for preventing ABC from becoming a random search algorithm. The algorithm runs until a pre-determined stopping condition occurs.

4.2. Genetic Algorithm

GA is a population-based and evolutionary algorithm. It is inspired by evolutionary processes such as selection, crossover, and mutation that try to carry the best offspring to the next generation. The general flowchart of GA is given in Fig. 2. In the initialization phase of GA, population individuals are generated as introduced in Eq. (12). In the evaluation phase of GA, population members' MSE values are calculated as introduced in Eq. (3) and fitness values of population members are calculated as shown in Eq. (10). The population members are ordered in decreasing order of their fitness values. After this phase, the best population members are selected for the matching pool in the selection phase of GA. The proposed GA uses the roulette wheel selection mechanism. The crossover phase of the proposed GA uses the simulated binary crossover operator as shown in Eqs. (13-14). After selecting a pair of population members, if a randomly generated number $r \in [0,1]$ is less than the predetermined crossover probability p_c, the new individuals are generated as introduced in Eqs. (13-14).

$$\gamma_{ji} = \begin{cases} (2u_i)^{\frac{1}{\eta+1}} & if\ u_i \leq 0.5 \\ \left(\frac{1}{2(1-u_i)}\right)^{\frac{1}{\eta+1}} & otherwise \end{cases} \quad \forall\ i\ j \tag{12}$$

Where γ_{ji} is the spread factor for ith coefficient of jth population member, η is the determinator of the crowding degree of the above probability distribution of the spread factor and u_i is a random real number between 0 and 1.

$$\beta_{ji} = 0.5\left[(1 + \gamma_{ji})\beta_{ji} + (1 - \gamma_{ji})\beta_{li}\right] \; \forall\ i\ j \tag{13}$$

$$\beta_{li} = 0.5\left[(1 - \gamma_{ji})\beta_{ji} + (1 + \gamma_{ji})\beta_{li}\right] \; \forall\ i\ j \tag{14}$$

After the crossover phase, another random phase named mutation takes place to protect the diversity of the population. if a randomly generated number $r \in [0,1]$ is less than the predetermined mutation probability p_m, the existing individual j is mutated as follows:

$$\beta_{ji}{}^{*} = \beta_{ji} + \phi_{ji}\beta_{ji} \ \ \forall i,j \tag{15}$$

where $\beta_{ji}{}^{*}$ is the new coefficient after mutation and ϕ_{ji} is a randomly generated real number, it may take between –1 and 1.

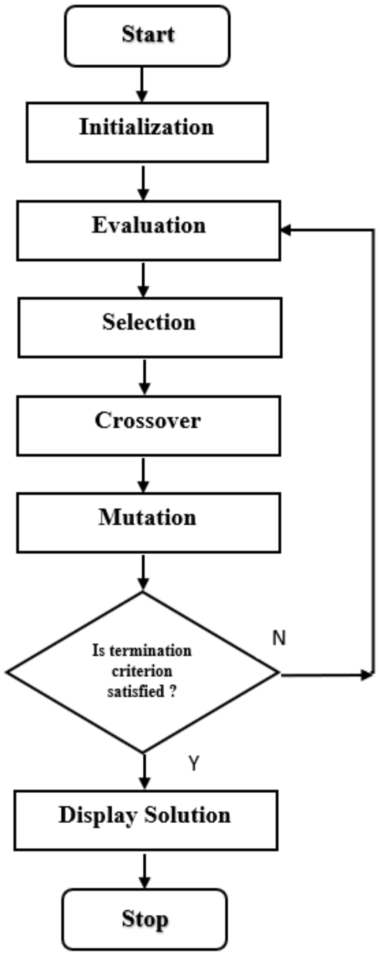

Fig. 2 The flowchart of the GA algorithm

4.3. Simulated Annealing

SA is inspired by the annealing process in metalwork. SA's best advantage by comparing it with other search methods is avoiding a local optimum by using a stochastic search [45]. SA starts with an initial system temperature T_{init} and this temperature is decreased at each iteration. At each system temperature T, SA looks for neighboring solutions in order to improve its current best solution. SA is a single solution-based metaheuristic method. In the initialization phase of SA, a solution $\vec{\beta}_{init}$ is generated as introduced in Eq. (8). SA assigns $\vec{\beta}_{init}$ as the best solution $\vec{\beta}_{best}$ at the beginning and it tries to improve $\vec{\beta}_{init}$ by generating new neighboring solutions. It uses Eq. (16) for generating new coefficient as follows:

$$\beta_{ji}{}^{*} = \begin{cases} \beta_{ji} + \phi_{ji}\beta_{ji} & if\ \beta_{ji} \neq 0 \\ \phi_{ji}\left(l_i + rand(0,1) * (u_i - l_i)\right) & otherwise \end{cases} \ \forall i,j \tag{16}$$

If the cost of the new solution $\vec{\beta}_{new}$ is better than $\vec{\beta}_{init}$ then $\vec{\beta}_{init}$ is replaced with $\vec{\beta}_{new}$. Otherwise, a random real number r between 0 and 1 is less than an acceptance probability, $\vec{\beta}_{init}$ is replaced with $\vec{\beta}_{new}$. Acceptance probability p_a of SA is calculated as follows:

$$p_a = e^{-\left(f\left(\vec{\beta}_{new}\right)-f\left(\vec{\beta}_{init}\right)\right)/T} \tag{17}$$

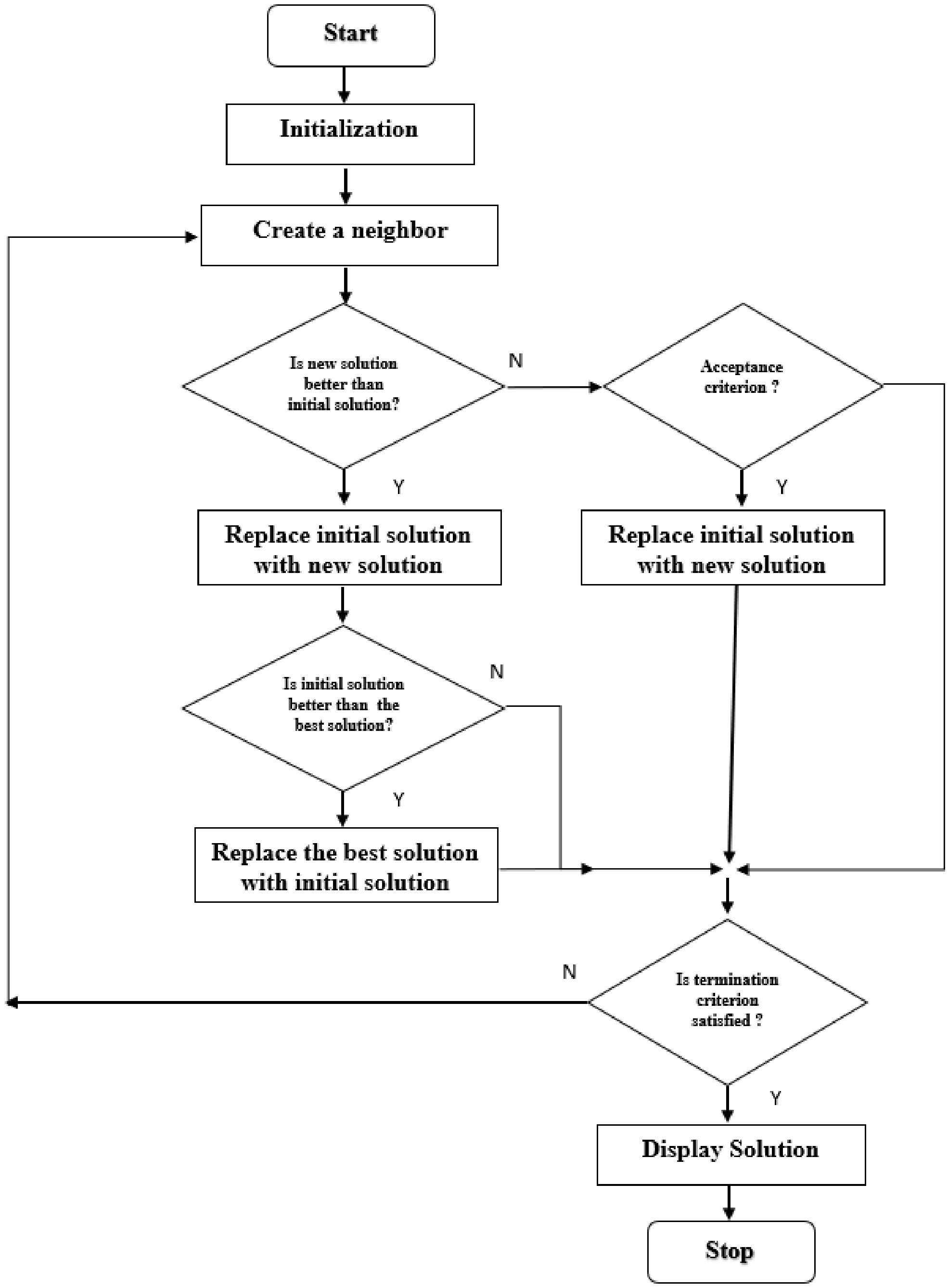

Fig. 3 The flowchart of the SA algorithm

This probabilistic acceptance helps SA to escape the local optimum. Also, if the cost of a new solution $\vec{\beta}_{new}$ is less than $\vec{\beta}_{best}$, then $\vec{\beta}_{best}$ is replaced with $\vec{\beta}_{new}$. While temperature is decreasing at each iteration, the number of movements for neighboring solutions at each iteration is being increased by SA. For each temperature T, the number of movements nm for neighboring solutions is calculated as follows:

$$nm = \left\lceil k^2 * \sqrt{\frac{T}{x}} \right\rceil + 1 \quad (18)$$

where k is the number of coefficients in the forecasting equations and x is the counter for iteration number? After searching nm neighbors, SA increases x value by one and decreases the system temperature as follows:

$$T = T * a \quad (19)$$

where a is the decrease factor between 0 and 1. This process ends when the system temperature reaches the stopping temperature or another stopping criterion is satisfied. Fig 3. shows the flowchart for SA in this study.

4.4. Hybrid Algorithms

In this section, we suggest using SA as a tool for escaping local optima and solution improvement in other metaheuristics GA and ABC, respectively. We used the same method proposed by Özmen et al. [46] to make a hybrid version of SA and GA to forecast Turkey's natural gas demand. Özmen et al.[46] used SA as a solution improvement module in their GA. In our GA-SA algorithms, the SA module is applied to the best solution after the evaluation phase of the proposed GA algorithm. In our proposed ABC-SA, the SA module is applied if the best new solution is found after the employed bee and onlooker bee phases. We use the same parameter setting suggested by Özmen et al. [46] for the SA module in hybrid algorithms. The SA module starts at 1000°C and continues, so as to make 10 iterations for each alteration of 1° and it ends when the temperature reaches 0°C. Also, the SA module in this study allows perturbation. Parameters and hybridization schemas of all algorithms are given in Table 3 for the reader.

Table 3: Parameters of metaheuristics to forecasting CO_2 emissions of Turkey

Metaheuristic	Is it hybrid?	Parameters
GA	No	$p_c = 0.85, p_m = 0.15$ and the population size is 120
ABC	No	the number of employed bees and the number of onlooker bees are equal to 120, the number of scout bees equals 1, and the $limit$ for failures of each bee is 300.
SA	No	the initial temperature is 10^6 °C
ABC-SA	Yes	the initial temperature is 1000°C, other parameters are the same with ABC
GA-SA	Yes	the initial temperature is 1000°C other parameters are the same with GA

5. Experimental Results

In this section, we tested all proposed metaheuristics and their hybrid variants. Since all proposed methods are stochastic, we executed independently these algorithms more than once until the predetermined stopping criterion which is 600 seconds. If the elapsed time of an algorithm reaches 600 seconds, the algorithm stops and displays the best solution found so far. The number of independent runs for an algorithm is 30. That means an algorithm is executed 30 times until 600 seconds. Each algorithm was executed for both multiple linear regression models to minimize the MSE value of its best coefficient vectors. Besides MSE values, other statistics are also used in comparisons. These are MAPE, MAE, and R2 statistics. All proposed algorithms are coded with the C++ programming language in Visual Studio 2019. All experiments and executions of algorithms are done in a standard workstation that has Intel Xeon E-2136 CPU (3.30 GHz) with 16 GB RAM. For ABC and ABC-SA algorithms; the number of employed bees and the number of onlooker bees are equal to 120, the number of scout bees equals 1, and the $limit$ for failures of each bee is 300. These parameters are suggested by Arık [2]. For GA and GA-SA algorithms; $p_c = 0.85, p_m = 0.15$ and population size is 120. For SA algorithms, the initial temperature is set as 106 °C. We use the same parameter setting suggested by Özmen et al. [46] for the SA module in hybrid algorithms. The SA module starts at 1000°C and continues, so as to make 10 iterations for each alteration of 1° and it ends when the temperature reaches 0°C. Also, the SA module in this study allows perturbation.

Firstly, we tested our algorithms for the first model introduced in Eq. (1). Table 4 shows the results of all proposed algorithms for each execution in 30 runs. As seen in Table 4, the SA algorithm spreads less from its mean value compared with others. Despite this, the best MSE value for the linear forecasting equation is found by GA algorithm; if we look at Fig. 4.a and Fig. 4.b, the best average solution with less spread belongs to SA algorithm. The closest rival of SA is ABC-SA for linear forecasting equations. As seen from Fig. 4.a, the SA module within ABC increases the efficiency of ABC algorithm. However, the same conclusion cannot be made considering the solution quality difference between GA and GA-SA for the first model. To test the statistical significance of the used methods, we made an ANOVA test for MSE values of methods. In our ANOVA, the dependent variable is MSE and there is one independent variable (factor) that is the method used to determine coefficients of forecasting equations. The p-value of the method factor is zero and we can deduce from this variance test, the method factor is statistically significant for MSE values. In conclusion, we can say that SA method is the best method for determining coefficients in the first multiple linear regression model if we only consider the stability and reproductivity of results in the same algorithm. Table 5 shows other statistics of the best solutions of all algorithms for the linear forecasting equation. As seen in Table 5, ABC-SA outperforms its rivals in three of four statistics. Furthermore, the best MSE values of ABC-SA and GA algorithms are close to each other and the average MSE of ABC-SA algorithm is the second after the SA algorithm. Even SA uses random initial solutions and randomized local search while determining coefficients of the forecasting equation, the SA algorithm's MSE values spread less than others for the first regression model. This might occur because SA is trapped in a local optimal point and the temperature value decreases to a certain point where the new candidate solution cannot be accepted toward to the end of the algorithm execution time of 600 seconds. Even SA algorithm's MSE values spread less from a certain value, the other statistics in Table 5 suggest using ABC-SA algorithm. Thus, we may deduce the best algorithm for the first model is ABC-SA algorithm considering multiple criteria and statistics.

Table 4: MSE values of algorithms for the first model

#of Run	ABC	ABC-SA	GA	GA-SA	SA
1	106.1763	127.2678	178.0394	356.5419	116.2289
2	125.6026	156.0149	253.0427	233.3632	117.0659
3	177.0989	99.2987	186.9268	127.6559	117.5993
4	154.1555	128.5866	222.8125	665.9451	117.6067
5	164.5035	132.9636	75.6457	355.1885	117.6033
6	180.1259	117.8639	153.4897	169.8130	117.5987
7	160.8980	143.2739	183.0255	263.6544	117.6325
8	136.7153	129.3953	460.9474	112.7885	115.1901
9	191.4855	131.9556	171.2984	140.4948	117.6002
10	171.6920	79.6208	129.9576	196.7563	117.6113
11	114.6933	123.1059	129.1651	299.1079	117.6027
12	141.0669	201.8547	346.7039	119.3759	117.6245
13	127.3351	129.2685	1039.1824	220.1192	117.3401
14	139.8233	144.9369	135.5554	273.6354	116.1953
15	111.4894	130.1481	225.5580	138.0023	117.6145
16	200.1010	102.4391	314.3379	364.1640	117.6225
17	140.0858	118.0351	433.6952	172.6289	117.6175
18	150.5069	110.4908	284.0357	201.3590	117.6000
19	220.2732	170.0936	366.0341	140.1378	117.6015
20	110.2225	148.3965	122.7713	251.2801	117.6163
21	171.0220	104.3259	153.6240	96.6401	117.6071
22	112.4140	107.6973	114.6926	216.7187	116.1103
23	168.1861	95.3638	152.9097	167.0188	117.7709
24	146.0137	127.4391	92.1024	382.1067	117.6369
25	125.2055	118.3696	144.5533	596.6030	117.6151
26	181.4356	142.9075	301.5989	188.7173	117.6317
27	160.4035	162.2073	310.7408	173.2142	117.6098
28	121.1206	125.6397	330.7923	198.8662	117.6236
29	207.5171	118.5107	121.1803	677.6258	117.3483
30	121.2286	116.2088	268.7551	656.2185	117.6284
The best	106.1763	79.6208	**75.6457**	96.6401	115.1901
Average	151.2866	128.1227	246.7725	271.8581	**117.3585**

Table 5: All statistics of best solutions of all algorithms for the first model

Method	MSE	MAE	MAPE	R^2
ABC	106.1763	7.6375	0.0258	96.94%
ABC_SA	79.6208	**6.8102**	**0.0239**	**97.75%**
GA	**75.6457**	7.1599	0.0252	97.63%
GA_SA	96.6401	7.9623	0.0279	97.34%
SA	115.1901	7.9881	0.0274	96.33 %

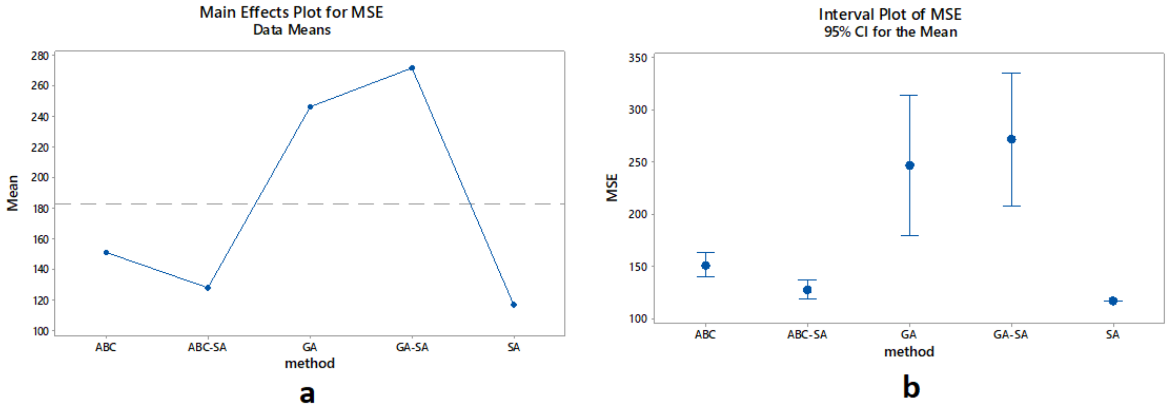

Fig. 4 (a) Main effect plot for MSE values, (b) Interval plot of MSE values for the first model

Secondly, we tested our algorithms for the second model introduced in Eq. (2). Table 6 shows the results of all proposed algorithms for each execution in 30 runs. As seen in Table 6, ABC-SA algorithm spreads less from its mean value compared with other algorithms and it has the best MSE value for the second model. Despite the spread of MSE results obtained ABC algorithm, the SA module in ABC-SA algorithm increases the efficiency of ABC considering the spread of MSE results and MSE values. If we look at Fig. 5.a and Fig. 5.b, the best average solution with less spread belongs to ABC-SA algorithm. The closest rival of ABC-SA is SA for the second model. As seen in Fig. 5.a, the SA module within ABC and GA increases the efficiency of ABC and GA algorithms. In order to test the statistical significance of the used methods, we made an ANOVA test for MSE values of the methods. In our ANOVA, the dependent variable is MSE and there is one independent variable (factor) that is the method used to determine coefficients of forecasting equations. The p-value of the method factor is zero and we can deduce from this variance test, the method factor is statistically significant for MSE values. Although ABC-SA seems like a better solution approach for the second regression model in view of the average MSE value, there are other statistics (MAE, MAPE, and R2) to be considered to determine the forecasting coefficients. The least MSE value (105.4845) is found by GA-SA algorithm. Table 7 shows all statistics of the best solutions of solution approaches. GA_SA has two of the four best statistics and SA has other two of the four best statistics. Even though GA_SA's MAE and MAPE value is more than SA, these values are so close to the best values. Thus, we may deduce the most preferable solution approach for the second regression model is GA-SA algorithm considering multiple criteria and statistics.

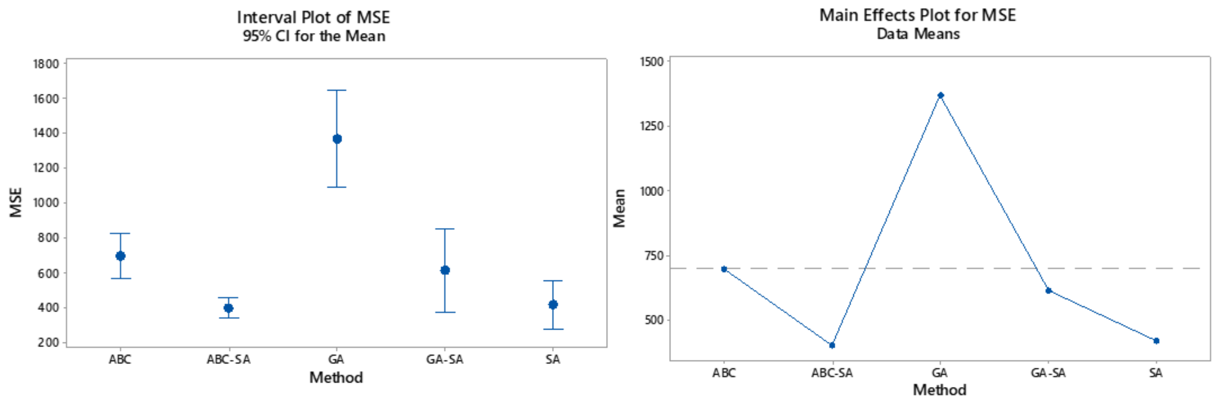

Fig. 5 (a) Main effect plot for MSE values, (b) Interval plot of MSE values for the second model

We test the performance of hybridization of SA within GA and ABC against regular GA, ABC, and SA algorithms using the Wilcoxon signed-rank test for each forecasting model. The results of the test are given in Table 8. According to these comparisons, ABC with SA is different from regular ABC for both regression models because the p-value of pairwise comparison via the Wilcoxon signed-rank test is less than 0.05. Thus, we may say that hybridization of SA within ABC decreases the forecasting error in the MSE metric. While hybridization of SA within GA presents different solutions than the regular GA algorithm for the first model, the same deduction cannot be done for the second regression model as seen P-values of GA vs GA-SA comparison in Table 8. For the first regression model, there is a difference between MSE values obtained SA and ABC-SA/GA-SA because the P-values of these comparisons are less than 0.5. On the contrary, the same deduction cannot be done for the second regression model because the P-values of the pairwise comparisons are more than 0.05. As a result of this analysis, the hybridization of SA within ABC is better than ABC algorithm for both regression models.

Table 6: MSE values of algorithms for the second model

#of Run	ABC	ABC-SA	GA	GA-SA	SA
1	1137.9125	486.3050	723.0015	319.2283	179.0639
2	258.6887	337.4784	851.1617	315.1948	276.7880
3	1079.7197	473.1336	2108.9993	777.0607	387.7509
4	441.6496	288.1705	2886.9478	460.4249	1056.6896
5	504.9791	533.3440	1162.1664	2079.6716	243.5872
6	659.9951	350.5450	2991.0275	705.1254	222.8892
7	647.5035	372.4091	1699.2213	295.0072	558.1693
8	761.1007	251.5996	2740.4217	307.4757	112.9112
9	732.3610	716.7250	754.3703	105.4845	331.9269
10	324.6624	527.6271	1644.0370	503.0628	518.6929
11	692.5116	500.0446	1150.8222	228.8617	332.5319
12	237.6280	292.4084	2437.3865	341.5527	179.0639
13	309.0866	296.8112	1191.0471	117.6504	839.6081
14	723.2274	290.4987	1160.2913	613.6297	1513.4387
15	935.0708	594.5475	584.9005	187.1277	166.1859
16	1072.5401	185.5897	1004.2143	491.2965	510.1504
17	479.9619	252.1061	1352.1674	346.4666	242.6722
18	438.3009	346.6410	317.8724	461.3946	248.4537
19	1787.2330	529.3630	608.8776	422.7305	164.8701
20	690.9215	541.8214	1430.9877	506.1792	117.6393
21	516.9811	858.6601	2161.7278	1131.0027	291.4638
22	1099.6612	398.1761	853.4295	459.6797	332.7212
23	606.7738	306.1484	1156.9702	533.9398	1567.0071
24	658.6952	473.1435	457.4306	3406.7695	145.5180
25	267.6760	500.0447	1499.8940	851.2355	386.2456
26	970.1562	166.3319	925.1500	545.8308	375.5327
27	489.8375	333.0847	867.7975	354.6408	356.2661
28	420.1794	321.1805	1157.6822	255.8513	186.7382
29	963.4858	159.7582	692.5668	734.0122	663.8368
30	1088.3012	365.8338	2478.1558	558.9605	135.2423
The best	237.628	159.7582	317.8724	**105.4845**	112.9112
Average	699.8934	**401.651**	1368.358	613.8849	421.4552

Table 7: All statistics of best solutions of all algorithms for the second model

Method	MSE	MAE	MAPE	R^2
ABC	237.6280	0.0381	10.6201	92.42%
ABC_SA	159.7582	0.0372	10.6668	94.90%
GA	317.8724	0.0450	13.4267	89.86%
GA_SA	**105.4850**	0.0288	8.1868	**96.63%**
SA	112.9112	**0.0285**	**7.9442**	96.40%

Table 8: Results of Wilcoxon signed-rank tests: pairwise comparisons of solution approaches for each model

Model	Comparison	P-value
The first model	ABC vs ABC-SA	0.0080
	GA vs GA-SA	0.5510
	SA vs ABC-SA	0.0220
	SA vs GA-SA	0.0000
The second model	ABC vs ABC-SA	0.0000
	GA vs GA-SA	0.0000
	SA vs ABC-SA	0.3760
	SA vs GA-SA	0.1000

6. Experimental Results

In this section, we propose four different scenarios for the economic indicators of Turkey and the number of construction permits in square meters. Since Turkey is a developing country, its economic growth needs to continue. However, independent variables such as population and construction permits are not strongly required to protect the development of the economy, especially construction permits. The first scenario includes a strong growth expectation in economic indicators, population, and construction permits. The second scenario expects strong growth in economic indicators, and a bit less slow growth in population and construction permits than the first scenario. The third scenario uses the same expectations except for construction permits. On the contrary, the fourth scenario includes a strong decrease in expectations in economic indicators. The fourth scenario also includes a growth expectation for population and construction permits. All scenarios for independent variables are given in Table 9.

Table 9: Data for scenarios

Scenario	GDP	Population	Export	Import	Construction Permit
Scenario#1	4%	1%	4%	4%	15%
Scenario#2	4%	0.75%	4%	4%	8%
Scenario#3	4%	0.75%	4%	4%	1%
Scenario#4	-2%	1%	-2%	-2%	15%

Since we have two forecasting equations and the best algorithms with fewer statistical errors for these equations, we forecast the CO2 emissions of Turkey in the years between 2019 and 2030 with these equations and algorithms. For the first model, the coefficients determined by ABC-SA algorithm are used for scenarios. GA-SA algorithm is the best algorithm among all five algorithms and coefficients that are found by GA-SA are used in the second model. Table 10 shows the forecasted CO2 emissions of Turkey in the years 2019 and 2030 with two different approaches and three different scenarios. At first look, we may understand the second model with GA-SA presents fewer CO2 emissions than the linear model with ABC-SA for all scenarios. The most aggressive scenario among all scenarios is scenario#1 and both approaches present the worst CO2 emissions in Turkey. While expected increases for population and construction permits decrease in scenarios such as Scenario#2 and Scenario#3, forecasted CO2 emissions of Turkey decrease. This conduction can be also observed in Fig. 6.

As seen in Table 10 and Fig. 6; the CO2 emission in 2030 will be 410 MtCO2 at least rising by 6.5% from the CO2 emission in 2018 even for the best scenario of CO2 emissions of Turkey. For the worst scenario, it may be 550 MtCO2 at least rising by 43% from the CO2 emission in 2018. Since Tukey is a developing country, other positively correlated parameters such as GDP and export values are vital to its economic wealth. However, increases in population and construction permits are not so vital for the wealth of Turkey. Hence CO2 is so energy-related and Turkey's economic wealth is so dependent on energy; if the energy demand of Turkey continues to be met from today's sources which are mainly coal, fuel-oil, and natural gas, Turkey needs to find new ways to decrease its CO2 emissions and the remaining option for the policymaker can be to decrease the activities of the building sector. Otherwise, Turkey has to diversify and increase its clean energy production plants such as solar and wind energy plants to accomplish both goals of keeping its economic wealth and decreasing CO2 emissions. In this study, we investigate four different scenarios where the energy source diversification of Turkey will be the same in the future. The experimental study showed that if construction permits and the population is strictly managed while the economical wealth of Turkey keeps on improving, the CO2 emissions of Turkey may be less than in other possible cases.

Table 10: Future projection of CO_2 emissions of Turkey in the years 2019 and 2030 considering multiple scenarios

	ABC-SA for the first model				GA-SA for the second model			
Years	Scenario1	Scenario2	Scenario3	Scenario4	Scenario1	Scenario2	Scenario3	Scenario4
2019	400.4738	398.1450	397.6216	396.7157	391.1350	389.8516	389.8671	385.1470
2020	411.6626	406.8532	405.7592	404.0713	400.1788	397.5249	397.5586	387.9699
2021	423.2221	415.7581	414.0426	411.7178	409.3104	405.1937	405.2487	390.6723
2022	435.1861	424.8685	422.4765	419.6846	418.5081	412.8313	412.9109	393.2601
2023	447.5932	434.1935	431.0653	428.0053	427.7470	420.4071	420.5154	395.7389
2024	460.4869	443.7426	439.8141	436.7185	436.9984	427.8872	428.0287	398.1140
2025	473.9168	453.5262	448.7277	445.8686	446.2296	435.2334	435.4131	400.3903
2026	487.9394	463.5550	457.8114	455.5068	455.4031	442.4027	442.6264	402.5723
2027	502.6189	473.8405	467.0706	465.6917	464.4764	449.3473	449.6215	404.6646
2028	518.0283	484.3948	476.5109	476.4908	473.4011	456.0140	456.3461	406.6709
2029	534.2513	495.2309	486.1381	487.9816	482.1226	462.3432	462.7415	408.5951
2030	551.3831	506.3625	495.9582	500.2534	490.5790	468.2691	468.7431	410.4405

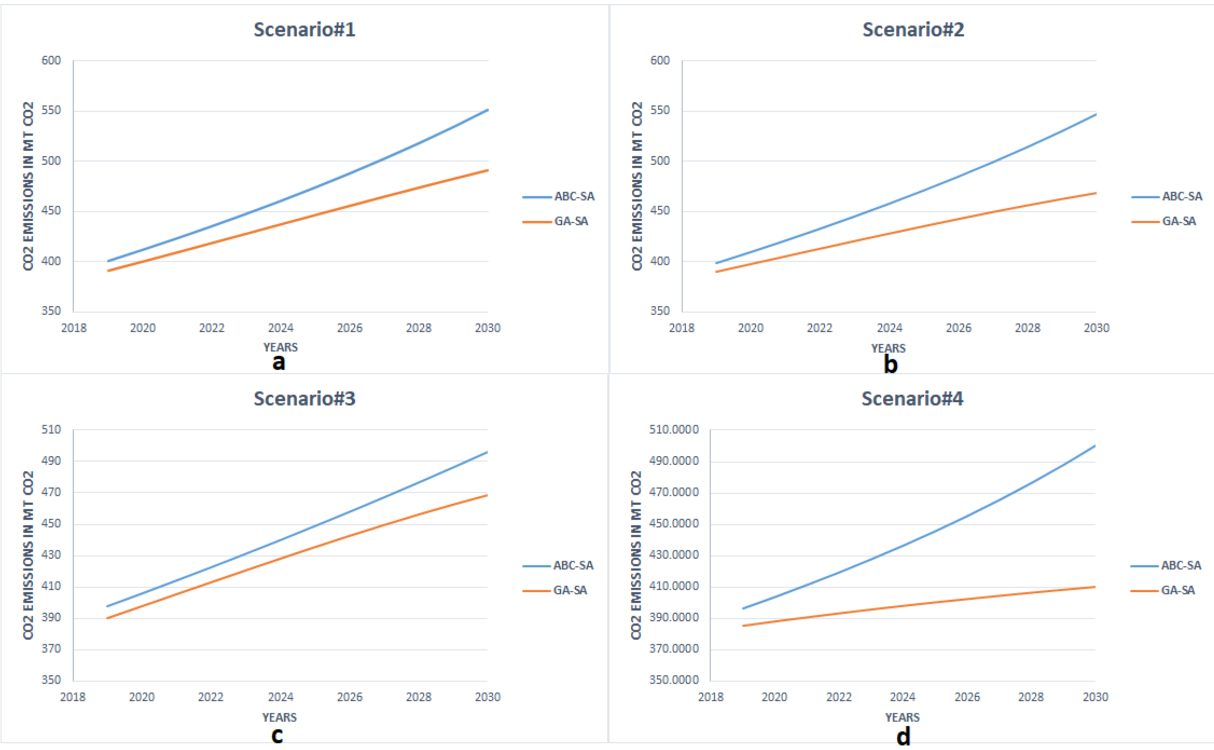

Fig. 6 Forecasted CO_2 emissions of Turkey with proposed algorithms and equations for (a) Scenario#1, (b) Scenario#2, (b) Scenario#3, and (c) Scenario#4

7. Conclusion

In this study, we investigate two different forecasting equations for CO2 emissions in Turkey. We use GDP, population, export, import, and construction permits as independent variables to forecast CO2 emissions of Turkey. We propose three different metaheuristics and two hybrid variants of those metaheuristics to determine coefficients of linear and quadratic forecasting equations. The experimental study reveals that ABC-SA algorithm outperforms ABC, GA, GA-SA, and SA algorithms considering four statistics and other criteria for the first multiple linear regression model. Furthermore, we reveal that GA-SA algorithm may be the most preferable solution approach considering four statistics and other criteria for the second multiple regression model. By using two different forecasting equations whose coefficients are determined by ABC-SA and GA-SA algorithms, we make a future projection of CO_2 emissions of Turkey in $MtCO_2$ in the years between 2019 and 2030 by using three different scenarios. In these different scenarios, we assume the energy production sources of Turkey will be the same and dependent on mostly fuel oil combustion in the future. The CO_2 emission in 2030 will be 410 $MtCO_2$ at least rising by 6.5% from the CO_2 emission in 2018 for the best scenario and it may be 550 $MtCO_2$ at least rising by 43% from the CO_2 emission in 2018. These scenarios reveal that if construction permits and the population is strictly managed while the economical wealth of Turkey keeps on improving, the CO_2 emissions of Turkey may be less than in other possible cases. Construction permit in squared meters is an indicator of how much steel and cement are required for the building sector and the required energy amount is extremely high for these materials. As a developing country, Turkey needs to keep increasing in GDP and export for its economic wealth. Nevertheless, increases in construction permits and the population are not so vital for the wealth of Turkey. In a conclusion, our study reveals that if the energy demand of Turkey continues to be met from today's sources which are mainly coal, fuel-oil, and natural gas, Turkey needs to find new ways to decrease its CO_2 emissions, the remaining option for the policymaker can be to decrease the activities of the building sector. Otherwise, Turkey has to diversify and increase its clean energy production plants such as solar and wind energy plants to accomplish both goals of keeping its economic wealth and decreasing CO_2 emissions. In our scenarios, we assume the energy production sources of Turkey will be the same and dependent on mostly fuel oil combustion in the future. Considering alternative development paths with clean and renewable energy sources is a good potential immediate extension for this study. Alternative forecasting methods such as ANN can be used to forecast future CO_2 emissions of Turkey for alternative scenarios including development paths with clean and renewable energy sources. Furthermore, different artificial intelligence methods and world data can be used for future studies.

ABC-SA presents the best average MSE values for both multiple linear regression models. Since we use multiple statistics to compare solution approaches, it may be difficult to state which model and its extension are the best for forecasting. An immediate future study can be to clarify which inputs lead to the least statistical error in different experiments on a larger scale. For future research, proposed algorithms can be implemented to forecast other countries' CO_2 with different forecasting equations like nonlinear or logistic. Our contribution to the literature can be shortlisted as follows:

- Inclusion of Buildings and Construction Sector: This study addresses a critical gap in existing research by considering the buildings and construction sector in the development of a CO_2 emissions forecasting equation. Unlike previous studies, we specifically focus on incorporating this significant sector, recognizing its impact on emissions.

- Utilization of Metaheuristic Algorithms: We employ advanced metaheuristic algorithms, including ABC, SA, GA and hybrid variants of these algorithms. These powerful optimization techniques are applied to determine the coefficients of our forecasting equations.

- Comprehensive Evaluation Metrics: In contrast to relying solely on minimizing the squares of forecasting errors using the least squares method, we introduce a more comprehensive approach. We consider multiple performance indicators, including squared errors, absolute errors, and percentage deviations of errors simultaneously when determining the forecasting equation's coefficients.

- Application to Turkey's CO_2 Emissions: Our study applies these strong metaheuristics and comprehensive evaluation metrics to forecast CO_2 emissions in Turkey. By doing so, we extend the applicability of metaheuristic approaches to the domain of CO_2 emissions prediction, particularly with a focus on the buildings and construction sector.

- Novelty in Methodology: While previous research has explored various techniques like Artificial Neural Networks (ANN), Support Vector Machines (SVM), Deep Learning, and Grey Forecasting for CO_2 emissions forecasting, our work pioneers the use of metaheuristics that incorporate data from the buildings and construction sector. This novel methodology offers a unique perspective and potentially improved forecasting accuracy.

Acknowledgments

Dr. Oğuzhan Ahmet Arık designed the investigated problem and he contributed to the design of metaheuristic algorithms and experimental study in the paper. Dr. Oğuzhan Ahmet Arık also edited and wrote the paper. Dr. Oğuzhan Ahmet Arık read and approved the final manuscript.

References

[1] F. Halicioglu, An econometric study of CO2 emissions, energy consumption, income and foreign trade in Turkey, Energy Policy. 37 (2009) 1156–1164. doi:10.1016/j.enpol.2008.11.012.

[2] O.A. Arık, Artificial bee colony algorithm to forecast natural gas consumption of Turkey, SN Applied Sciences. 1 (2019) 1–10. doi:10.1007/s42452-019-1195-8.

[3] M.A. Behrang, E. Assareh, M.R. Assari, A. Ghanbarzadeh, Using bees algorithm and artificial neural network to forecast world carbon dioxide emission, Energy Sources, Part A: Recovery, Utilization and Environmental Effects. 33 (2011) 1747–1759. doi:10.1080/15567036.2010.493920.

[4] H. Chang, W. Sun, X. Gu, Forecasting energy CO2 emissions using a quantum harmony search algorithm-based dmsfe combination model, Energies. 6 (2013) 1456–1477. doi:10.3390/en6031456.

[5] W. Sun, J. Wang, H. Chang, Forecasting carbon dioxide emissions in China using optimization grey model, Journal of Computers (Finland). 8 (2013) 97–101. doi:10.4304/jcp.8.1.97-101.

[6] R. Samsami, Comparison between genetic algorithm (GA), particle swarm optimization (PSO) and ant colony optimization (ACO) techniques for NOx emission forecasting in Iran, World Applied Sciences Journal. 28 (2013) 1996–2002. doi:10.5829/idosi.wasj.2013.28.12.1155.

[7] L. Abdullah, H. Mohd Pauzi, An effective model for carbon dioxide emissions prediction: Comparison of artificial neural networks learning algorithms, International Journal of Computational Intelligence and Applications. 13 (2014). doi:10.1142/S146902681450014X.

[8] X. Wang, Y. Cai, Y. Xu, H. Zhao, J. Chen, Optimal strategies for carbon reduction at dual levels in China based on a hybrid nonlinear grey-prediction and quota-allocation model, Journal of Cleaner Production. 83 (2014) 185–193. doi:10.1016/j.jclepro.2014.07.015.

[9] W. Sun, Y. Xu, Using a back propagation neural network based on improved particle swarm optimization to study the influential factors of carbon dioxide emissions in Hebei Province, China, Journal of Cleaner Production. 112 (2016) 1282–1291. doi:10.1016/j.jclepro.2015.04.097.

[10] H. Zhao, S. Guo, H. Zhao, Energy-related CO2 emissions forecasting using an improved lssvm model optimized by whale optimization algorithm, Energies. 10 (2017). doi:10.3390/en10070874.

[11] L. Wen, Y. Liu, A research about Beijing's carbon emissions based on the IPSO-BP model, Environmental Progress and Sustainable Energy. 36 (2017) 428–434. doi:10.1002/ep.12475.

[12] W. Sun, J. Sun, Prediction of carbon dioxide emissions based on principal component analysis with regularized extreme learning machine: The case of China, Environmental Engineering Research. 22 (2017) 302–311. doi:10.4491/eer.2016.153.

[13] W. Sun, C. Wang, C. Zhang, Factor analysis and forecasting of CO2 emissions in Hebei, using extreme learning machine based on particle swarm optimization, Journal of Cleaner Production. 162 (2017) 1095–1101. doi:10.1016/j.jclepro.2017.06.016.

[14] A.K. Baareh, Evolutionary design of a carbon dioxide emission prediction model using Genetic Programming, International Journal of Advanced Computer Science and Applications. 9 (2018) 298–303. doi:10.14569/IJACSA.2018.090341.

[15] X. Zhao, M. Han, L. Ding, A.C. Calin, Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA, Environmental Science and Pollution Research. 25 (2018) 2899–2910. doi:10.1007/s11356-017-0642-6.

[16] E. Assareh, M. Nedaei, A metaheuristic approach to forecast the global carbon dioxide emissions, International Journal of Environmental Studies. 75 (2018) 99–120. doi:10.1080/00207233.2017.1374075.

[17] S. Dai, D. Niu, Y. Han, Forecasting of energy-related CO2 emissions in China based on GM(1,1) and least squares support vector machine optimized by modified shuffled frog leaping algorithm for sustainability, Sustainability (Switzerland). 10 (2018). doi:10.3390/su10040958.

[18] H. Zhao, G. Huang, N. Yan, Forecasting energy-related CO2 emissions employing a novel ssa-lssvm model: Considering structural factors in China, Energies. 11 (2018). doi:10.3390/en11040781.

[19] B. Ameyaw, L. Yao, Analyzing the impact of GDP on CO2 emissions and forecasting Africa's total CO2 emissions with non-assumption driven bidirectional long short-term memory, Sustainability (Switzerland). 10 (2018). doi:10.3390/su10093110.

[20] D. Guo, H. Chen, R. Long, Can China fulfill its commitment to reducing carbon dioxide emissions in the Paris Agreement? Analysis based on a back-propagation neural network, Environmental Science and Pollution Research. 25 (2018) 27451–27462. doi:10.1007/s11356-018-2762-z.

[21] C.-C. Lin, R.-X. He, W.-Y. Liu, Considering multiple factors to forecast CO2 emissions: A hybrid multivariable grey forecasting and genetic programming approach, Energies. 11 (2018). doi:10.3390/en11123432.

[22] B. Ameyaw, L. Yao, A. Oppong, J.K. Agyeman, Investigating, forecasting and proposing emission mitigation pathways for CO2 emissions from fossil fuel combustion only: A case study of selected countries, Energy Policy. 130 (2019) 7–21. doi:10.1016/j.enpol.2019.03.056.

[23] S.M. Hosseini, A. Saifoddin, R. Shirmohammadi, A. Aslani, Forecasting of CO2 emissions in Iran based on time series and regression analysis, Energy Reports. 5 (2019) 619–631. doi:10.1016/j.egyr.2019.05.004.

[24] Y. Huang, H. Wang, H. Liu, S. Liu, Elman neural network optimized by firefly algorithm for forecasting China's carbon dioxide emissions, Systems Science and Control Engineering. 7 (2019) 8–15. doi:10.1080/21642583.2019.1620655.

[25] W. Qiao, H. Lu, G. Zhou, M. Azimi, Q. Yang, W. Tian, A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer, Journal of Cleaner Production. 244 (2020). doi:10.1016/j.jclepro.2019.118612.

[26] W. Wu, X. Ma, Y. Zhang, W. Li, Y. Wang, A novel conformable fractional non-homogeneous grey model for forecasting carbon dioxide emissions of BRICS countries, Science of the Total Environment. 707 (2020). doi:10.1016/j.scitotenv.2019.135447.

[27] Q. Wu, F. Meng, Prediction of energy-related CO2 emissions in multiple scenarios using a least square support vector machine optimized by improved bat algorithm: a case study of China, Greenhouse Gases: Science and Technology. 10 (2020) 160–175. doi:10.1002/ghg.1939.

[28] A. Malik, E. Hussain, S. Baig, M.F. Khokhar, Forecasting CO 2 emissions from energy consumption in Pakistan under different scenarios: The China–Pakistan Economic Corridor, Greenhouse Gases: Science and Technology. (2020) ghg.1968. doi:10.1002/ghg.1968.

[29] I. Ozturk, A. Acaravci, CO2 emissions, energy consumption and economic growth in Turkey, Renewable and Sustainable Energy Reviews. 14 (2010). doi:10.1016/j.rser.2010.07.005.

[30] C. Hamzacebi, I. Karakurt, Forecasting the energy-related CO2 emissions of Turkey using a grey prediction model, Energy Sources, Part A: Recovery, Utilization and Environmental Effects. 37 (2015) 1023–1031. doi:10.1080/15567036.2014.978086.

[31] E.M. Bildirici, T. Bakirtas, The relationship among oil and coal consumption, carbon dioxide emissions, and economic growth in BRICTS countries, Journal of Renewable and Sustainable Energy. 8 (2016). doi:10.1063/1.4955090.

[32] U. Şahin, Forecasting of Turkey's electricity generation and CO2 emissions in estimating capacity factor, Environmental Progress and Sustainable Energy. 38 (2019) 56–65. doi:10.1002/ep.13076.

[33] U. Şahin, Forecasting of Turkey's greenhouse gas emissions using linear and nonlinear rolling metabolic grey model based on optimization, Journal of Cleaner Production. 239 (2019). doi:10.1016/j.jclepro.2019.118079.

[34] M.S. Bakay, Ü. Ağbulut, Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms, Journal of Cleaner Production. 285 (2021). doi:10.1016/j.jclepro.2020.125324.

[35] S. Ullah, R. Luo, T. S. Adebayo, and M. T. Kartal, "Dynamics between environmental taxes and ecological sustainability: Evidence from top-seven green economies by novel quantile approaches," Sustain. Dev., vol. 31, no. 2, pp. 825–839, Apr. 2023, doi: 10.1002/SD.2423.

[36] T. S. Adebayo and A. Samour, "Renewable energy, fiscal policy and load capacity factor in BRICS countries: novel findings from panel nonlinear ARDL model," Environ. Dev. Sustain., pp. 1–25, Jan. 2023, doi: 10.1007/S10668-022-02888-1/FIGURES/6.

[37] R. Radmehr, S. Shayanmehr, E. A. Baba, A. Samour, and T. S. Adebayo, "Spatial spillover effects of green technology innovation and renewable energy on ecological sustainability: New evidence and analysis," Sustain. Dev., 2023, doi: 10.1002/SD.2738.

[38] M. T. Kartal, A. Samour, T. S. Adebayo, and S. Kılıç Depren, "Do nuclear energy and renewable energy surge environmental quality in the United States? New insights from novel bootstrap Fourier Granger causality in quantiles approach," Prog. Nucl. Energy, vol. 155, p. 104509, Jan. 2023, doi: 10.1016/J.PNUCENE.2022.104509.

[39] EnerData, EnerData, (2020). https://www.enerdata.net/ (accessed March 28, 2020).

[40] WorldBank, GDP (current US$) - Turkey | Data, (2020). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=TR (accessed March 28, 2020).

[41] TurkStat, Turkish Statistical Institute, (2020). http://www.turkstat.gov.tr (accessed March 28, 2020).

[42] G. Ozdemir, E. Aydemir, M.O. Olgun, Z. Mulbay, Forecasting of Turkey natural gas demand using a hybrid algorithm, Energy Sources, Part B: Economics, Planning and Policy. 11 (2016) 295–302. doi:10.1080/15567249.2011.611580.

[43] M.D. Toksarı, Ant colony optimization approach to estimate energy demand of Turkey, Energy Policy. 35 (2007) 3984–3990. doi:10.1016/j.enpol.2007.01.028.

[44] D. Karaboga, B. Basturk, On the performance of artificial bee colony (ABC) algorithm, Applied Soft Computing. 8 (2008) 687–697. doi:10.1016/J.ASOC.2007.05.007.

[45] O.A. Arık, Comparisons of metaheuristic algorithms for unrelated parallel machine weighted earliness/tardiness scheduling problems, Evolutionary Intelligence. 13 (2020) 415–425. doi:10.1007/s12065-019-00305-7.

[46] A. Özmen, Y. Yılmaz, G.-W. Weber, Natural gas consumption forecast with MARS and CMARS models for residential users, Energy Economics. 70 (2018) 357–381. doi:10.1016/j.eneco.2018.01.022.

[47] O.A. Arık, Artificial bee colony algorithm including some components of iterated greedy algorithm for permutation flow shop scheduling problems, Neural Computing and Applications. 33 (2021) 3469–3486. doi:10.1007/s00521-020-05174-1.