Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-021-13431-6

RESEARCH ARTICLE

Carbon footprint forecasting using time series data mining methods: the case of Turkey

Müge Akyol[1] · Emine Uçar[2]

Received: 11 June 2020 / Accepted: 9 March 2021

Abstract
In the globalizing world, many factors such as rapidly increasing population, production and consumption habits, and economic growth cause climate changes. The carbon footprint is a measure of CO_2 emissions released into the atmosphere, which increases day by day, causing glaciers to melt and increase sea level, reduce water resources, and global warming. For Turkey, as a country trying to complete its economic development, signed international agreements such as the Paris Climate Agreement and Kyoto Protocol to reduce the carbon footprint give great importance to the studies estimating carbon footprint and making policies to reduce it. For this reason, in this study it is aimed to estimate the greenhouse gas emissions of Turkey in the year 2030 and to determine its damages to the economy. Time series forecasting algorithm in the WEKA data mining software was used for analysis, and population, gross domestic product, energy production, and energy consumption were used as independent variables. As a result of analysis using data from the years 1990–2017, as long as Turkey continues its course of gradually increasing the amount of current greenhouse gas emissions in the year 2030, 728.3016 metric tons of CO_2 equivalent will be reached. It appears that these estimates remain below the rate of Turkey's commitments at the Paris Climate Agreement that is considered to be promising for Turkey. However, the estimations in other studies should not be ignored; policy makers should determine policies accordingly.

Keywords Carbon footprint · Renewable energy · Time series data mining · SMOreg

Introduction

Mankind has lived with nature since its creation and has been a part of nature. The protection of the world in which it lives is necessary for the sustainability of human existence. Realizing that environmental problems started to threaten the living conditions with economic development, humanity started to work towards increasing the quality of the environment. The course of these initiatives over the years has been as follows;

- *Stockholm Conference (1972):* By organizing the conference, the United Nations argued that environmental problems and protection of the environment are the duty of all the countries of the world, and therefore the measures to be taken should be guided. Environmental problems became universal via this conference.
- *Brundtland Commission Report (1987):* It was prepared by the World Environment and Development Commission and it is important to use the term sustainable development in this report for the first time (Sezer 2015).
- *Kyoto Protocol (1997):* This is a protocol that Turkey has also signed and with this protocol EU countries pledged to cut greenhouse gas emissions by 8% between 2008 and 2012 (Özdemir 2011).
- *Paris Climate Agreement (2015):* With this agreement, 196 countries came together and an agreement was reached to reduce the greenhouse gas emission rates,

Responsible Editor: Philippe Garrigues

✉ Müge Akyol
mugeakyol.sbe19@iste.edu.tr

Emine Uçar
emine.ucar@iste.edu.tr

[1] Department of Logistics Management, Faculty of Business and Management Sciences, İskenderun Technical University, Hatay, Turkey

[2] Department of Management Information Systems, Faculty of Business and Management Sciences, İskenderun Technical University, Hatay, Turkey

Published online: 18 March 2021

which increase the global temperature, by the rate committed by 2030, so that the increase in temperature should be limited to 1.5–2° for a sustainable world (Pabuçcu and Bayramoğlu 2017).

As mentioned in the Brundtland report for the first time, there are different indicators of sustainable development. One of these indicators is the ecological footprint resulting from people's production and consumption activities. Ecological footprint is defined as the fertile water and soil area required to eliminate waste generated by the activities of an individual or community and to reproduce the resources consumed with current technologies. The carbon footprint constitutes 46% of Turkey's ecological footprint (WWF 2012). Carbon footprint is the measure of the carbon dioxide emission that occurs at all stages of production, transportation, use, and degradation of a product. This measure is expressed by the term global hectare (kha), which indicates the production capacity of one hectare of land with a size of approximately 1.5 football fields. The carbon footprint is divided into primary (direct) and secondary (indirect). While the primary footprint refers to the energy consumption resulting from household appliances or the CO_2 emissions released from fossil fuels burned as a result of transportation during the day; the secondary footprint represents damage to nature throughout the entire life cycle. It is possible to learn how much carbon footprint is left to nature with carbon footprint calculation methods. In these calculations, un-harvested forest land that is required to remove every ton of carbon emitted into the atmosphere through photosynthesis is used. If the required forest land is not sufficient to eliminate the carbon emitted in terms of size and yield, it appears in the carbon capture category (Erden Özsoy 2015). Ultimately, the fact that the world exceeds its carrying capacity causes the ozone layer to perforate, causing the average temperature to rise, melting glaciers, and increasing sea level. All these troubles threaten sustainable world existence.

When the greenhouse gas emissions, including CO_2 emissions, are examined, it is understood that the most important source comes from the energy sector. Unconscious use of sectors such as electricity production and consumption, industrial production, transportation, construction, and agricultural and forestry lands also increase the amount of toxic gas in the atmosphere (Pabuçcu and Bayramoğlu 2017).

Many methods are used for future carbon footprint estimates, but studies using data mining methods have become important in recent years. Data mining methods generally focus on planning and decision making processes such as marketing strategies, and cost estimations. Data mining models can be divided into two categories: one of them is a descriptive model that provides explanatory information about the data set and the other one is a predictive model that tries to estimate the value of the dependent variable in accordance with a target (Ecemiş and Irmak 2018).

Many countries are working on the green economy within the scope of sustainable development models planned with the goal of leaving a habitable world for future generations. Turkey still supplies a large part of its energy needs from environmentally enemy fossil fuels. For this reason, it is important for our country to carry out a measurement and forecasting study for the transition to the green economy and to offer solutions in this direction. The aim of this study is threefold: to estimate the measurement of the carbon footprint in the 2018–2030 periods by using variables like population, primary electricity consumption, gross domestic product (GDP), and electricity generation based on the 1990–2017 period by using time series data mining methods and determine its damages to the economy and offer solutions.

The article consists of five sections. In "Brief review of literature" section a literature review is presented. In "Methodology" section the research methodology and dataset are all explained in detail. In "Results and discussions" section the results of time series analysis are presented and discussed. The fifth and final section offers the solutions and concludes the paper.

Brief review of literature

With the technological transformation developing in the economic literature, many studies on carbon dioxide emissions have been carried out. Most of these studies are towards reducing carbon emission rates with the transition to renewable energy, and policy recommendations have been developed in this context. Some of these studies are briefly discussed below.

The study of economic growth and income injustice by Simon Kuznets (1955) became a benchmark for studies that examine the relationship between environmental degradation and income, defined by the Kuznets curve in later years. Kuznets found an inverted *u*-shaped relationship between the income levels of the countries and growth and he upheld that the income injustice will decrease as the income level increases (Kuznets 1955). Environmental Kuznets curve has been used in environmental studies based on Kuznets' study, and findings have been reached that environmental degradation will increase as income level increases, but degradation will decrease if an increase continues after a level. The hypothesis that environmental pollution rates are higher compared to high-income countries with the increase in energy demands of developing countries, and the environmental degradation will decrease with the increase in income levels of countries, has been confirmed (Carson et al. 1997; Halıcıoglu 2009). Alam and others have addressed the causal relationship between energy consumption, CO_2 emissions, and economic growth of India using time series analysis with a dynamic modeling approach. In the study in which GDP, total energy consumption, total workforce, capital stock, and CO_2 emission rate were considered as variable, the period of 1971–2006 was evaluated by time series analysis and causality relationships were revealed. According to the results of the analysis, unlike the environmental Kuznets curve hypothesis, there is no long-term relationship between income and CO_2 emissions. A bidirectional causality relationship has been determined between energy consumption and CO_2, and no causal relationship has been found between energy consumption and income. For India, as a country consuming energy based on fossil fuels, it is recommended to produce energy-saving policies that can prevent global warming without ignoring the increase in income (Alam et al. 2011). Arı and Zeren examined the relationship between CO_2 emission and economic growth with the panel data analysis method. In the study, in which the data of Mediterranean countries for the period 2000–2005 was used to test the environmental Kuznets curve hypothesis, an *N*-shaped curve was obtained between per capita income and CO_2 emissions. The results of the study show that as the per capita income increases, CO_2 emission increases to a certain level, and the increase in energy use and population density per person increases CO_2 emission. In this context, it is recommended to produce incentive policies for companies to use clean technologies (Arı and Zeren 2011).

Apart from studies examining the income-environment relationship, there are studies conducted with different methods for estimating emission rates. Pabuçcu and Bayramoglu tried to estimate future greenhouse gas emission values of Turkey and selected 28 EU countries (2020–2030) using artificial neural networks. In the study consisting of many variables such as GDP, production and consumption of energy, population, and the energy used for transportation and greenhouse gas emission rates, it is predicted that greenhouse gas emission 1244.13 mt rates in 2030 will be above the promised rate (929 mt CO_2 equivalent) at Paris Climate Convention in case of continuation of the existing conditions in Turkey. The EU average is estimated to be 151.75 mt in 2030. As a result of the study, it is recommended for Turkey to use renewable energy to increase energy efficiency (Pabuçcu and Bayramoğlu 2017). Çam et al. conducted their studies on energy efficiency. In the study, gross GDP of the years 1990–2013, capital stock per workforce, energy consumption, and CO_2 emission rates were used as variables and energy efficiency has been estimated with artificial neural network model by using the energy efficiencies obtained by the TOPSIS method. According to the results the most important determinant of Turkey's energy efficiency has been the total energy consumption through the capital stock per labor. The long-term relationship between the ARDL boundary test and the variables was appointed, and it was found that capital stock per workforce, GDP, and energy consumption positively affect energy efficiency, and CO_2 energy efficiency negatively. It is recommended for Turkey to increase energy efficiency by using renewable energy sources (Çam et al. 2018). Hamzaçebi and Karakurt have analyzed Turkey's CO_2 emission data of the years 1965–2012, by using gray forecasting method based on energy and tried to estimate Turkey's CO_2 emission values for 2025. The results of the analysis show that the amount of CO_2 released into the atmosphere in 2025 will increase by 64%, reaching 496,404 mt compared to 2010. As a result, the gray estimation method can be used in CO_2 emission estimation both to establish energy policies and to ensure the sustainability of international agreements on climate changes (Hamzaçebi and Karakurt 2015). Shaikh et al. has tried to create a framework considering Turkey's water and carbon footprint consumption of existing and future. Based on the years of 1990–2013, energy production estimation was made for the year 2030 with the ARIMA model, and three different scenarios were established: the continuation of the current situation, the government plan, and the renewable energy development plan. In the continuation of the current situation, the water footprint for 2030 is 6.67 × 1011 gallons, and the carbon footprint is 2.05 × 108 tons, 7.5% less water footprint and 28% less carbon footprint than the first scenario in the government plan scenario. In the renewable energy development plan, there is 31.7% less water footprint and lowest carbon footprint estimation compared to the other two scenarios. It was determined that the strong correlation between water consumption and CO_2 emission decreases as we move towards the renewable energy plan. As a result of the study, it is recommended to consider the renewable energy plan scenario for future energy conversion policies in terms of both carbon and water footprints compared to other scenarios (Shaikh et al. 2017). Kaya et al. tried to determine the wind energy potential of Kastamonu using artificial neural networks. They have found that some of the wind turbines are better at efficiency and Kastamonu is potentially suitable for building wind turbines. The study is important for feasibility, considering cost (Kaya et al. 2016).

Studies to predict carbon emission trends are a subject of interest to researchers around the world. Rezaei et al. modeled the carbon emission of four Scandinavian countries including Finland, Sweden, Denmark, and Norway using the group method of data handling algorithm (GMDH) of the artificial neural network method. In this study, total primary energy consumption, type of utilized fuels, and GDP from 1990 to 2016 were considered as indicators. As a result of the research, it was found that the group method of data handling algorithm was a suitable method for carbon emission estimation (R^2 = 0.998) and the obtained output was very close to the actual data. In addition, variables considered as inputs were shown to be the most effective factor in estimating these Scandinavian countries (Rezaei et al. 2018). Jaforullah and King examined the effect of renewable energy sources on CO_2 emissions. In the research in which the data of the United States between 1965 and 2012 were used, the relationship between CO_2

emission, nuclear energy consumption, renewable energy consumption, and real GDP variables were analyzed by using the vector error correction model (VECM). As a result of the analysis, they found that nuclear energy consumption has an insignificant but positive effect on CO_2; there was a negative relationship between energy prices and emissions, and CO_2 emissions can decrease significantly by increasing the use of renewable energy sources. In the study conducted in the form of reassessment of previous studies, suggestions are given such as incentive usage of renewable energy instead of nuclear energy and carbon tax enforcement in order to reduce the emission rates (Jaforullah and King 2015). Wang and Ye (2017) tried to estimate the carbon emission rate using fossil-weighted energy consumption between 2014 and 2020. They used ARIMA and non-linear gray prediction model for estimating the impact of low, medium, and rapid economic growth on CO_2 emissions and offered solutions (Wang and Ye 2017). In a research conducted in China, one of the most greenhouse gas-emitting countries in the world, it was determined that energy intensity has a decreasing effect on CO_2 emission. Zhang et al. analyzed the effect of China's energy use intensity on CO_2 emissions based on 1991–2006 data and found that economic activities had an impact on greenhouse gas emissions and economic structure contributed little to changes in CO_2 emission intensity (Zhang et al. 2009). Li et al. conducted the carbon emission rate of the Beijing-Tianjin-Hebei region, which uses fossil fuel to represent the entire country. They developed a new forecasting method called SVM-ELM which optimized the extreme learning machine algorithm by using kernel of support vector machines. They used gray prediction method for the region's carbon use estimation and found that energy consumption seriously affects the amount of carbon emissions. Firstly, they estimated CO_2 emissions between 2017 and 2030 with gray prediction method. Then they used forecasting results as input for support vector machines. And thus, they proved that SVM-ELM methods had high accuracy of carbon emission prediction compared to support vector machines (SVM) or extreme learning machines (ELM). They stated that the amount of CO_2 emission would be about 97 million tonnes after 2027, so they recommended that the government must be encouraged for clean energy (Li et al. 2018). The research of Hertwich and Peters was about measuring the carbon footprint of 73 nations by dividing the world into 14 regions. It was divided into categories such as construction, shelter, clothing, service industry, mobility, and trade. CO_2 and other greenhouse gas emissions were estimated within the scope of the global trade analysis project with the global multiregional input-output model (MRIO). Countries with similar geographical characteristics were taken together and their estimations were stated as weighted averages. Operating result for the per capita carbon footprint that was the amount of about 28 t/yp for African countries, 33 t/yp for American countries, and 4.6 t/yp for Turkey had been detected. While there was a strong correlation between consumption expenditures and CO_2 emissions, it was found that the food industry did not have a significant effect on the amount of emissions. Analysis results show that rich countries have low greenhouse gas emission rates compared to poor countries (Hertwich and Peters 2009). The estimation study of the carbon footprint presented by Pandey et al. as a qualitative research with the current methods includes a wide literature review and different calculation methods. The study, which also shows the rates of CO_2 emissions made for enterprises, activities, and organizations of many countries, states that different calculations are caused by regional or calculation methods. It has been suggested that supervision is necessary due to the commercialization of carbon footprint calculations, and that the institutions performing these calculations should be encouraged by targeting transparent and sustainable development (Pandey et al. 2011). Vieilledent and his colleagues modeled Madagascar by separating 5 different regions where they examined the effect on deforestation and carbon emissions. For the deforestation and emission estimation of 2030, land cover change of 2000–2010 was divided into seven classes and used as a training dataset. For the future deforestation prediction, the demographic variable obtained from the using population census and the location factors determining the location of deforestation were obtained by logistic regression. With these models, estimation of deforestation was made until 2030. The relationship between this rate and CO_2 emissions was examined and it was found that the annual emission amount depends on the rate of deforestation (Vieilledent et al. 2013). Pao et al. analyzed China's CO_2 emission data of the years 1980–2009, by using non-linear gray Bernoulli model (NGBM) to predict China's CO_2 emission values for 2009–2020 period. They used energy consumption, real GDP, and CO_2 emissions as indicators. They used three comparative methods to estimate CO_2 emissions like ARIMA, (GM (1, 1)), and NGBM-OP. As a result of the obtained MAPE values in the study, it was found that the NGBM-OP was a stronger predictor than GM and ARIMA. Beside the prediction of CO_2 emissions, their study showed that there was a long-run equilibrium relationship between emissions, energy consumption, and real GDP (Pao et al. 2012). In the research conducted by Behrang et al. (2011), carbon emission prediction was tried to forecast with multilayer perceptron neural network and bees' algorithm. In the research, many socioeconomic indicators such as world population, primary energy demands, oil and natural gas trade movements, and GDP were used; an estimated value between 2007 and 2040 was created using the data of 1980–2006 period. For the study consisting of two stages, primarily GDP, population, oil, and natural gas trade movements were used as input variables depending on the bees' algorithm; oil, natural gas, coal consumption, and primary energy consumption were output. On the second step, CO_2 emission value was tried to be estimated by using the output variables of the first step. Two separate scenarios were used for the first step: in the first scenario, a linear model determined by the fitted polynomial values and in the second scenario, the artificial neural network model that was trained and tested; the predictive results of each input variable were obtained. The estimated CO_2 emission value for the first scenario is 77,027 mt, and for the second scenario it is 75,611 mt. The results show that the two models have been successful for emission rates, fossil fuels, and energy consumption, and that policy makers can engage in energy planning practices using this study (Behrang et al. 2011). Jones and Kammen calculated all the variables that cause carbon footprint by categorizing them into energy, transportation, waste, water, food, goods, and services, and calculated by carbon footprint calculation methods, and removing the marginal reduction cost curves, it identified areas where carbon footprint savings can be made in their study about carbon footprint reduction for the American household and society. According to the study, it is stated that carbon savings can be achieved with many small changes such as purchasing fuel-saving vehicles, using eco-driving technique in vehicles, using bicycles or buses instead of individual vehicles, reducing flights by 20%, turning to energy-saving when choosing a refrigerator, and using the thermostat effectively and efficiently (Jones and Kammen 2011).

As seen from the previous studies mentioned above, there are a limited number of studies on the greenhouse gas estimation for Turkey. In addition, in most of the studies, it has been observed that the carbon emission rate is tried to be estimated based on only one parameter. Therefore, in this comprehensive study, it is aimed to eliminate this gap by using various parameters that are thought to have an effect on the greenhouse gas emission rate.

Methodology

In this section detailed information about the dataset, methods, and evaluation criteria used in this the study is given.

Data

The variables used in the study were selected among the strong determinants of CO_2 emission obtained as a result of the literature review. The selected variables were population, primary electricity consumption, electricity generation, and gross domestic product (GDP). These variables were used for the prediction of 2018–2030 based on the 1990–2017 period. The variables used in the study were obtained from different sources and a new dataset was created. This dataset was presented in Appendix Table 5. In addition, these variables and their sources were given in Table 1 in detail.

Time series data mining

Time series are the magnitudes in which the values of the variables are observed sequentially over time. They are used as a method that makes future predictions by obtaining a forward extension of the series (Emeç 2020). Time series are divided into two categories as continuous and discrete in the context of time. Continuously recorded series of engineering fields such as electrical signals, voltage, sound vibrations, and measurement of seismic movements are the continuous time series; interest series, sales volume, and production amount are the discrete time series (Can 2009). Depending on the number of variables, if the series contains the previous observation values of a single variable, they are called the univariate time series; if it contains the previous observation values of more than one variable, they are called the multivariable time series.

Many economic variables are collected at regular intervals. These variables are very closely related to each other, for example, unemployment rate, inflation, growth, exports, imports, and investments. It is possible to reveal their structures and predict future values by analyzing each of these variables one by one. However, each of these variables is explored together due to its close relationship with others and the state of the relationship between them is revealed. Examining this relationship is very useful for future forecasting and policy making.

In the analysis of the relationship between the variables in the multivariate time series, the series should be stationary, that is, constant over time in order to avoid the problem of spurious regression. In some cases, although the series are not stationary, any linear combination is stationary (Can 2009). In time series data mining, known as data mining methods used to analyze a time series, restrictions such as stationarity and linearity required in classical time series have been removed (Aydın et al. 2005). Here, a model based on time is developed in accordance with the purpose of the series and the most appropriate estimation of future data is provided (Erguvan Etgin 2017).

In this study, three most used algorithms in the time series data mining prediction models were selected and it was tried to be determined which of these algorithms produce better results. The algorithms used in time series analysis are briefly described below.

Linear regression (LR) LR algorithm uses standard least square multiple linear regression for prediction. Optionally, the qualification can be selected for this process, either using greedy or backward elimination method. Thus, it builds a model that includes all or only selected qualities. LR algorithm uses Akaike criteria for model selection and makes weighted sampling (Witten et al. 2005; WEKA 2020b; Akaike 1974).

Table 1 Details of the used variables and sources

Variables	Definitions	Sources
Population	Population of Turkey at 1990–2007 period	United Nations National Accounts https://unstats.un.org/unsd/snaama/Basic
Gross domestic product	2010 at fixed prices, in USD	United Nations National Accounts https://unstats.un.org/unsd/snaama/Basic
Primary electricity consumption	Million tons of oil equivalent	Balance Sheets of General Directorate of Energy Affairs https://www.eigm.gov.tr/tr-TR/Denge-Tablolari/Denge-Tablolari?page=1 Business Data Platform https://www.statista.com/statistics/892848/primary-energy-consumption-turkey/
Electricity generation	GWh	Energy Statistics of Turkish Statistical Institute. http://www.tuik.gov.tr/PreTablo.do?alt_id=1029
Greenhouse gas emission rate	Million tonnes of CO_2 equivalent	Environmental Statistics of Turkish Statistical Institute http://www.tuik.gov.tr/PreTablo.do?alt_id=1019

Multilayer perceptron The MLP algorithm implements supervised learning in classification and uses back propagation. An artificial neural network using feed forward back propagation method consists of the input layer, the hidden layer, and the output layer. The number of neurons in the layers varies according to the problem under consideration. Data from the input layer is processed in the hidden layers and the output layer. A supervised learning algorithm is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. With supervised learning, output values corresponding to each input and input are given, and the algorithm generates outputs by making generalizations and modeling from examples. The model with the least square error is determined as the last model and forecasting is realized through this model (Öztemel 2006).

Sequential minimal optimization The SMOreg algorithm is included in support vector machines (SVM), one of the machine learning methods, and is mainly based on statistical learning. Support vector machines, whose basic logic is the determination of the best separating hyperplane for linearly separable data, solve the non-linearly separable data by applying transformation technique. As it has the ability to generalize and classify unobserved data during training, it gives more successful results in many real problems compared to artificial neural networks and decision trees. The SMO algorithm proposed by Platt classifies two-dimensional working sets by repeatedly selecting and optimizing them. SMOreg does this by maximizing the margin between the two classes. In the data that cannot be separated linearly, the optimum linear hyperplane finds the original data by moving it to a higher dimensional space (Shevade et al. 2000; Güven and Bilgin 2014).

Evaluation metrics

In this study, mean absolute percent error (MAPE) and mean squared error (MSE), which are the most used performance criteria in the literature, have been used to compare the success of the models and determine the best model. The equations for these metrics are given in Eqs. 1 and 2, respectively.

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i-\widehat{y}_i}{y_i}\right| \quad (1)$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i-\widehat{y}_i\right)^2 \quad (2)$$

MAPE is a metric that measures the error in a prediction model as a percentage, and the closer the MAPE value to zero is, the higher the performance of the model is. MSE is a metric that measures the difference between actual data and predicted results, and a MSE value close to zero indicates that the model is successful.

Results and discussions

In the study, WEKA program was used to estimate the future values of greenhouse gas emissions. WEKA is a powerful tool in data mining and is an open source program. WEKA got its name from the initials of each word in the Waikato Environment for Knowledge Analysis (WEKA). In the study, firstly, WEKA, Time Series Forecasting plugin, was used to analyze time series. The three most used algorithms in the time series data mining prediction models were selected. These are linear regression, multilayer perceptron, and sequential minimal optimization (WEKA 2020a).

The dataset used in this study was divided into 2 parts. The data between 1990 and 2008 were used for training, and the data for 2009–2011, 2012–2014, and 2015–2017 were used for estimation in groups. In the study, models were first trained with the data between the years 1990–2008 and then, starting from 2009, it was tried to estimate greenhouse gas emission rates for 3-year periods. During these predictions, MSE and MAPE values were used to test the estimation success of each model. The MSE and MAPE values obtained by the models according to the determined periods were given in Table 2.

After this stage, the most successful model was determined by calculating the average of the MSE and MAPE values obtained by each model in all periods. The average MSE and MAPE values calculated for each model were presented in Table 3.

As can be seen from Table 3, the LR model, which has the highest MSE and MAPE values, was last, while the MLP model was second and the SMOreg model, which has the lowest MSE and MAPE values, was the first. In addition, the greenhouse gas emission rates estimated by each model for the years 2009 and 2017 were compared with the actual data in order to evaluate the success of the models. Detailed information about the actual data and the greenhouse gas emission rates estimated obtained by the models was presented in Table 4.

The values stated as bold in Table 4 were the closest estimated values to the actual greenhouse gas emission results. As can be seen from Table 4, the SMOreg model obtained very close values to actual results in most of the tested years. Therefore, in the study, the greenhouse gas emission rates to be released into the atmosphere between 2018 and 2030 were estimated using the SMOreg model that obtained the best results in both Table 3 and Table 4. The estimation results of the model for the next years were given in Fig. 1.

As seen in Fig. 1 the greenhouse gas emission rates of Turkey have been increasing each year. Energy production and primary energy consumption are the two most important parameters in estimating greenhouse gas emissions. It is known that most of the carbon emission that causes the greenhouse effect consists of fossil fuels used in energy production and consumption. As a matter of fact, in this study, it has been seen that there is a strong relationship between energy production and primary energy consumption and carbon dioxide emission, and it is thought that serious differences in carbon emission rates can be achieved by controlling energy consumption. And also the rapidly growing population every year is another cause of the increase in Turkey's greenhouse gas emissions. It is seen that the effect of another independent variable GDP on carbon emission is negative. This case proves the accuracy of the environmental Kuznets curve hypothesis. As a matter of fact, as mentioned in the Carson et al. (1997) study, the hypotheses that with the increasing energy demand of developing countries, environmental pollution rates are higher than high-income countries and that with the increase in the income levels of the countries, environmental degradation will decrease has been confirmed. These results are supported by the study of Okutan and Yamak (2018) and show that Turkey has reached the income level defined as scale effects then passed to the technological impact level.

Table 3 Average MSE and MAPE values of each model

	LR	MLP	SMOreg
Average MSE for all periods	5398.3520	104.6519	79.0241
Average MAPE for all periods	13.3842	2.3406	1.8888

Today, carbon emission is an important problem that scientists are trying to solve, and almost all countries are working to reduce greenhouse gas emissions in the world atmosphere. Carson et al. (1997), Halıcıoglu (2009), and Alam et al. (2011) examined the relationship between carbon emissions and income under the environmental Kuznets curve hypothesis, while Hamzaçebi and Karakurt (2015), Behrang et al. (2011), Hertwich and Peters (2009), and Shaikh et al. (2017) tried to estimate emissions by various methods. Zhang et al. (2009), Jaforullah and King (2015), and Çam et al. (2018) conducted studies examining the relationship between CO_2 emissions and other variables and determined policies in line with these studies.

In this study the greenhouse gas emission estimates of Turkey were obtained by time series mining methods using parameters such as population, GDP, energy production, and consumption variables. When examining the studies for Turkey, it was found that generally artificial neural network algorithms were used. Pabuçcu and Bayramoğlu (2017) estimated the CO_2 emissions for selected EU countries and Turkey by using artificial neural networks. Although almost

Table 2 MSE and MAPE values obtained by the models

Algorithms	Year	MSE			MAPE			Average of MSE	Average of MAPE
Linear regression (LR)	2009–2011	3,346.4374	417.6099	928.2114	13.3911	3.2400	7.3022	1,564.0862	7.9778
	2012–2014	1,457.4139	8,004.5650	1,669.8935	7.3828	18.1210	6.7962	3,710.6241	10.7667
	2015–2017	3,283.0074	15,438.0018	14,040.0282	11.7403	26.1076	26.3764	10,920.3458	21.4081
Multi-layer perceptron (MLP)	2009–2011	33.1515	38.1435	39.8875	1.5563	1.6733	1.7433	39.0608	1.6576
	2012–2014	46.7250	67.5869	65.5335	1.7306	1.9431	1.9089	59.9485	1.8609
	2015–2017	250.7600	193.3230	200.7560	4.0329	3.1904	3.2864	214.9463	3.5032
Sequential minimal optimization (SMOreg)	2009–2011	40.7380	49.8186	58.8298	1.3151	1.5863	1.7542	49.7955	1.5519
	2012–2014	56.2392	83.3988	93.4105	1.4645	2.0513	2.2511	77.6828	1.9223
	2015–2017	72.2825	120.1381	136.3618	1.6464	2.3079	2.6220	109.5941	2.1921

Table 4 Comparison of actual values with forecasting values of each model

Years	Actual values	Forecasting values		
		LR	MLP	SMOreg
2009	395.515	321.908	393.451	**395.390**
2010	398.661	340.958	386.199	**399.283**
2011	427.572	353.595	352.237	**394.080**
2012	446.935	297.805	**429.298**	421.472
2013	438.969	337.171	**435.890**	428.162
2014	457.962	389.265	433.282	**433.943**
2015	472.191	409.853	490.697	**475.696**
2016	498.469	431.825	**492.369**	489.132
2017	526.253	403.513	493.494	**507.349**

the same variables were used in our study, different results were obtained. It is thought that the reason for these results was the use of five years of data in their study. The results of this study were compared to Hamzaçebi and Karakurt (2015) the differences in the emission rates were noteworthy. In this study, emission results are estimated based on the energy variable, while considering other related variables in our study makes it possible to accept the difference of the results.

National income per capita in a country gives some clues about the carbon footprint of that country. For this reason, GDP has been considered as a socioeconomic indicator in many studies on carbon emissions in the world. On the other hand, electricity consumption has been included in the studies as the most important independent variable affecting the carbon footprint. In this context, this study was similar to studies of Alam et al. (2011), Behrang et al. (2011), Pao et al. (2012), Jaforullah and King (2015), and Rezaei et al. (2018), who used energy consumption and GDP variables. Since Alam et al. (2011) Jaforullah and King (2015) conducted a time series analysis; the range of dataset was higher than other forecasting studies.

Our study had similar aspects with the study of Li et al. (2018). The dataset used for training and testing, time interval, years of estimation, and energy consumption variables are very similar. Although the examined regions were different, there was strong evidence in both studies that energy consumption increased the CO_2 emissions, and solutions were presented with this evidence.

Conclusions and policy implications

Ecological footprint reduction studies, which are regarded as one of the most important indicators of sustainable development, have reached a different dimension over the years with technological innovations and research and development planning. In this context, reducing the carbon footprint is a great sensitivity for all countries of the world and agreements are made in the international arena today. Under the Paris Climate Agreement which Turkey is a signatory, countries are committed to reducing greenhouse gas emission rates in accordance with adjusted rates. In accordance with this commitment, developed and developing countries determine energy policies by considering economic growth.

Countries decide according to some criteria in determining energy policies. In terms of world countries, it is possible to divide use of energy into two categories as developed and developing countries. Every country, which is trying to realize its economic growth, tries to increase its production while the technology is stable in the first stage, and therefore uses more resources and energy in the production process. Due to the excessive energy and resource usage, the amount of CO_2 released into

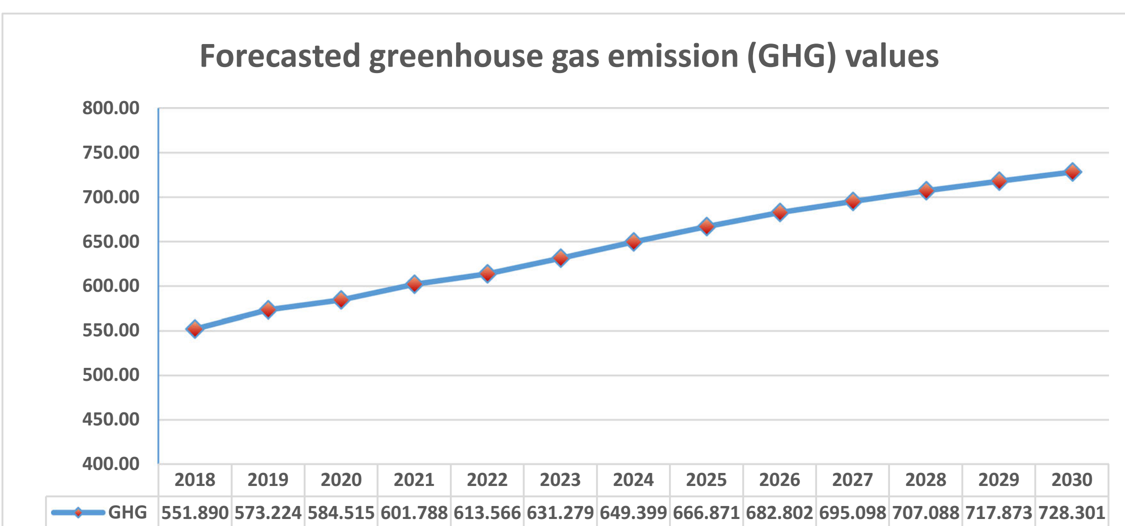

Fig. 1 Forecasted greenhouse gas emission (GHG) values for 2018–2030 with the SMOreg model

the atmosphere increases, causing deterioration of environmental quality. This condition, which is also called the effect of scale, takes place in the early stages of growth. As the growth continues, there is a structural transformation in the economy, the transition from labor-intensive production to technology-intensive production is made with the effect of growth, and environmental destruction starts to decrease with the use of technology in the energy sector, which is known as the structural effect. Finally, with the share allocated to R&D investments in developed countries, a technological transformation is experienced and the environmental quality is improved by reducing the carbon footprint by using clean technology. This transformation, which is also called technological effect, continues in the last period in developed countries where per capita income is high (Koçak 2014).

G20 countries, which constitute 85% of the world economy, continue to reduce their carbon footprint by increasing their energy efficiency in many sectors with the effect of technological transformation. For example, in the transportation sector, standards for reducing the emission rate arising from fuel consumption are implemented, and thus, up to 20% efficiency is achieved (Civelekoğlu and Bıyık 2018). Apart from the transportation sector, there were also major transformations in renewable energy. For example, Tesla founder, Elon Musk, founded Solarcity Company, arguing that the world's 1-year energy needs could be met by absorbing and storing 1 min of solar energy falling on the earth's surface. The aim of this company is to meet the energy needs of the houses very cheaply by equipping the houses with solar cells. Solarcity, which was gone to the public in 2012, reduced the solar energy installation cost by 50% to 2.50 dollars per watt (Yılmaz 2016).

After the 24th January 1980 decisions, a rapid industrialization process began in Turkey, which established an outward development model and struggled to integrate with the global economy. This process, with the increasing population of our country, has also increased the energy use required to increase production. Energy demand has started to increase with the increasing use of energy, and accordingly, the rate of destruction given to the environment has increased. This result is common, considering that overall economic growth increases environmental degradation in general. It should be remembered that many other variables such as population growth, open trade, and personal income are among the determinants of CO_2 emission besides energy consumption. Therefore, individuals and firms also have duties in the development of environmental quality as well as country policies.

Industry sector which has a high proportion of energy use constitutes 70% of the rates of greenhouse gas emissions. Turkey, which is in the category of developing countries, wants to complete the economic development accompanied by sustainable development policies and has many renewable energy potentials such as hydraulic, solar, wind, and geothermal to achieve this. But it can only use a very small part of this potential. Turkey should first go to reduce energy use in the sector and it should be directed to renewable energy in order to complete the growth by making efficient use of energy with less carbon emissions.

According to the carbon footprint calculations made in recent years, as stated in the 2010 issue of the Living Planet Report, it is estimated that until 2050, the world will be far above its carbon holding capacity and at least two more planets will be needed to survive unless we change our production and consumption habits. Since these planets have not been found yet, protecting the planet lives seems to be the only cure for human beings. For this reason, the studies carried out to develop appropriate strategies by making predictions about the future are important.

In this study, the greenhouse gas emission of Turkey was examined over time changes with the effect of independent variables such as population, GDP, energy production, and energy consumption, and future forecasts were obtained by using time series mining. By looking at the comparison of the actual values and test results, it is seen that the SMOreg model predicts greenhouse gas emission at acceptable error levels. As a result of analysis using the data as long as 1990–2017 years, if Turkey continues the trend today greenhouse gas emissions gradually increasing the amount of 728.3016 was determined as CO_2 equivalent metric in the year 2030. It appears that these estimates remain below the rate of Turkey's commitments at the Paris climate treaty that is considered to be promising for Turkey. However, the estimates in other studies should not be ignored; policy makers should determine policies accordingly.

Considering the results obtained from the study, it is seen that the variables that affect the amount of emissions most are electricity generation and primary energy consumption. Although Turkey has high energy production potential due to the geographical features, it is an energy-dependent country due to the inability to use this potential efficiently. As Turkey realizes its growth, it must strive to use this potential to the full end. It should develop environmental protection policies in this direction by identifying possible carbon emission targets related to climate change. This target should be to reduce the carbon emission rate to 2010 levels by taking the necessary measures until 2030 according to the Base Path Scenario. Turkey needs to determine by scientific methods its share of global emission reduction, how to implement which low carbon development policies, and the costs and effects of these mitigate policies on the country's economy. For this reason, a fair share of the global carbon budget should be determined with the Climate Equity Reference Calculator (WWF 2015). Under the target scenario, a climate policy package consisting of carbon tax, renewable energy investment fund, and energy efficiency tools should be implemented. Accordingly, it is envisaged that the share of wind and sun will be 44% by using the collected carbon taxes as a source for renewable energy applications (WWF 2015). Although Turkey ranks 12th in the world as the installed power of the wind farm, it still uses little of its potential. In this context, the wind stands can be established by determining potential locations with feasibility analyses. The country's presence in mild climate and dominance of the sun in all 4 seasons ensures maximum use of the sun. Work on the designs of new solar panels should be encouraged to increase the share of the sun in electricity production. Surrounded by seas on all 3 sides, Turkey can turn this advantage into electrical energy with wave energy systems. This should help reduce Turkey's energy dependence on the outside. The work of environmental protection organizations in Turkey should be taken into account and people need to be conscious to reduce the secondary footprint.

The most important limitation of this research is that it is unpredictable how much the emission rates will change as a result of the unforeseen global Covid-19 outbreak. The failure to know the normal process of return to life and the end of the economic downturn in the return process may lead to a rapid increase in emission rates.

Authors' contribution Conceptualization, E. UÇAR. Data curation, M. AKYOL. Investigation, M. AKYOL, E. UÇAR. Methodology, M. AKYOL, E. UÇAR. Writing—original draft, M. AKYOL, E. UÇAR. Writing—review and editing, M. AKYOL, E. UÇAR.

Data availability All data generated or analyzed during this study are included in this manuscript.

Declarations

Ethical approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare that they have no competing interests.

Appendix

Table 5 Dataset used in this study

Years	Population	GDP (ABD dollar)	Primary energy consumption (Million tonnes of fuel oil equivalent)	Production of energy (GWh)	The rate of greenhouse gas emission (million tonnes of CO_2 equivalent)
1990	53,921,699	363,953,141,840	52,465	57,543	219,202
1991	54,840,531	367,325,226,581	53,262	60,246	226,578
1992	55,748,875	389,307,006,602	55,698	67,342	232,802
1993	56,653,729	420,615,154,835	59,471	73,808	240,149
1994	57,564,132	397,667,810,036	58,238	78,322	234,127
1995	58,486,381	426,264,971,410	62,968	86,247	247,585
1996	59,423,208	456,126,194,264	68,717	94,862	267,232
1997	60,372,499	490,467,027,515	72,614	103,296	278,607
1998	61,329,590	505,631,597,875	73,306	111,022	280,288
1999	62,287,326	488,494,200,606	72,451	116,440	277,759
2000	63,240,121	520,930,514,794	73,500	124,922	298,890
2001	64,191,474	489,871,019,020	66,900	122,725	280,411
2002	65,143,054	521,371,090,335	73,100	129,400	286,073
2003	66,085,803	550,610,910,697	77,500	140,581	305,596
2004	67,007,855	603,713,603,545	82,800	150,698	314,951
2005	67,903,406	658,107,313,056	84,900	161,956	337,213
2006	68,763,405	704,896,790,945	94,300	176,300	358,155
2007	69,597,281	740,356,326,377	100,400	191,558	391,423
2008	70,440,032	746,614,198,898	100,800	198,418	387,593
2009	71,339,185	711,489,987,901	102,200	194,813	395,515
2010	72,326,914	771,876,791,246	107,700	211,208	398,661
2011	73,409,455	857,659,284,076	115,100	229,395	427,572
2012	74,569,867	898,740,650,963	122,300	239,497	446,935
2013	75,787,333	975,055,500,275	121,600	240,154	438,969
2014	77,030,628	1,025,433,602,157	125,600	251,963	457,962
2015	78,271,472	1,087,840,328,670	137,500	261,783	472,191
2016	79,512,426	1,122,475,332,193	144,400	274,408	498,469
2017	80,745,020	1,206,000,890,969	157,700	297,278	526,253

References

Akaike H (1974) A new look at the statistical model identification. IEEE Explore 19(6):716–723

Alam MJ, Begum IA, Buysse J, Rahman S, Van Huylenbroeck G (2011) Dynamic modelling of causal relationship between energy consumption, CO_2 emissions and economic growth in India. Renew Sust Energ Rev 15(6):3243–3251

Arı A, Zeren F (2011) CO_2 Emisyonu ve Ekonomik Büyüme: Panel Veri Analizi. J Faculty Econ Admin Scie Celal Bayar Univ 18(2):37–47

Aydın İ, Karaköse M, Akın E (2005) Zaman Serileri Veri Madenciliği Tekniği Kullanılarak Asenkron Motorlarda Sarım Ve Sürtünme Hatalarının Teşhisi. Electr Electron Comput Eng Nati Congr 11: 124–127

Behrang MA, Assareh E, Assari MR, Ghanbarzadeh A (2011) Using bees algorithm and artificial neural network to forecast world carbon dioxide emission. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 33(19):1747–1759

Çam S, Sigeze Ç, Ballı E (2018) Türkiye'nin Enerji Verimliliğinin Yapay Sinir Ağı ve ARDL Yaklaşımı İle Analizi. J Ege Acad Rev 18(4): 661–670

Can M (2009) İşletmelerde Zaman Serisi Analizi ile Tahmin. Unpublished doctoral dissertation Istanbul University. Institute of Social Sciences, Istanbul

Carson R, Jeon Y, McCubbin D (1997) The relationship between air pollution emissions and income: USA data. Environ Dev Econ 2: 433–450

Civelekoğlu G, Bıyık Y (2018) Ulaşım Sektöründen Kaynaklı Karbon Ayak İzi Değişiminin İncelenmesi. Bilge Int J Sci Technol Res 2(2): 157–166

Ecemiş O, Irmak S (2018) Paslanmaz Çelik Sektörü Satış Tahmininde Veri Madenciliği Yöntemlerinin Karşılaştırılması. J Soc Sci 8(1): 148–169

Emeç H (2020) Ekonometri Semineri: Durağan ve Durağan Olmayan Zaman Serileri s:3 http://kisi.deu.edu.tr//hamdi.emec/Ekonometri%20Semineri/Duragan_ve_Duragan_Olmayan_Zaman_Serileri.pdf. Date of Access: 07.01.2020

Erden Özsoy C (2015) Düşük Karbon Ekonomisi ve Türkiye'nin Karbon Ayak İzi. Hak İş International. J Labor Soc 4(9):198–215

Erguvan Etgin, E (2017) Zaman Serilerinde Veri Madenciliği Öngörü Algoritmalarının Etkinlik Ve Verimliliğinin Bıst100 Hisse Senedi Üzerinde Gerçeklenmesi. Master's thesis, Maltepe University, Institute of Science

Güven Z B, Bilgin T (2014) Zaman Serileri Madenciliği Kullanılarak Nüfus Artışı Tahmin Uygulaması. Acad Inform , 5-7

Halıcıoglu F (2009) An econometric study of CO_2 emissions, energy consumption, income and foreign trade in Turkey. Energy Policy 37(3):1156–1164

Hamzaçebi C, Karakurt I (2015) Forecasting the energy-related CO_2 emissions of Turkey using a grey prediction model. Energy Source 37(9):1023–1031

Hertwich EG, Peters GP (2009) Carbon footprint of nations: a global, trade-linked analysis. Environ Sci Technol 43(16):6414–6420

Jaforullah M, King A (2015) Does the use of renewable energy sources mitigate CO_2 emissions? A reassessment of the US evidence. Energy Econ 49:711–717

Jones CM, Kammen DM (2011) Quantifying carbon footprint reduction opportunities for US households and communities. Environ Sci Technol 45(9):4088–4095

Kaya Ü, Caner M, Oğuz Y (2016) Rüzgar Türbin Modelleri Kullanarak Kastamonu İli Rüzgar İle Elektrik Üretim Potansiyeli Tahmini. Technological. Appl Sci 11(3):65–74

Koçak E (2014) Türkiye'de Çevresel Kuznets Eğrisi Hipotezinin Geçerliliği: ARDL Sınır Testi Yaklaşımı. J Bus Econ Stat 2(3):62–73

Kuznets S (1955) Economic growth and income inequality. Am Econ Rev 45(1):1–28

Li M, Wang W, De G, Zi X, Tan Z (2018) Forecasting carbon emissions related to energy consumption in Beijing-Tianjin-Hebei region based on grey prediction theory and extreme learning machine optimized by support vector machine algorithm. Energies 11(9):2475

Okutan Ş, Yamak R (2018) Türkiye'nin Çevresel Kuznets Eğrisi: ARDL Sınır Testi Yaklaşımı. İktisat Seçme Yazılar:157–181

Özdemir V (2011) Türkiye'nin Karbonizasyon İndeksinin Temel Enerji Göstergelerine Bağlı Olarak Yapay Sinir Ağları İle Tanımlanması. J Fac Eng Archit Gazi Univ 26(1)

Öztemel E. (2006) Yapay Sinir Ağları. Papatya Publishing, IBAN 975-67-97-39-8, İstanbul, October

Pabuçcu H, Bayramoğlu T (2017) Yapay Sinir Ağları ile CO_2 Emisyonu Tahmini: Türkiye Örneği. J Faculty Econ Admin Sci Gazi Univ 18(3):762–778

Pandey D, Agrawal M, Pandey JS (2011) Carbon footprint: current methods of estimation. Environ Monit Assess 178(1-4):135–160

Pao HT, Fu HC, Tseng CL (2012) Forecasting of CO_2 emissions, energy consumption and economic growth in China using an improved grey model. Energy 40:400–409

Rezaei MH, Sadeghzadeh M, Nazari MA, Ahmadi MH, Astaraei FR (2018) Applying GMDH artificial neural network in modeling CO_2 emissions in four Nordic countries. Int J Low-Carbon Technol 13(3):266–271

Sezer Ö (2015) Küresel Konferanslar ve Çevre Sorunları: Çevre Kalkınma ve Etik Açısından Eleştirel Bir Değerlendirme. Environment, Urbanization problems and solutions Declaration 761-780

Shaikh MA, Kucukvar M, Onat NC, Kirkil G (2017) A framework for water and carbon footprint analysis of national electricity production scenarios. Energy 139:406–421

Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK (2000) Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Netw 11(5):1188–1193

Vieilledent G, Grinand C, Voudry R (2013) Forecasting deforestation and carbon emissions in tropical developing countries facing demographic expansion: a case study in Madagascar. Ecol Evol 3(6): 1702–1716

Wang ZX, Ye DJ (2017) Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models. J Clean Prod 142:600–612

WEKA, https://www.cs.waikato.ac.nz/ml/index.html, Date of Access: 07.01.2020a

WEKA Sourceforge, http://weka.sourceforge.net/doc.dev/overview-summary.html Date of access: 07.01.2020b

Witten I H, Frank E, Hall M A (2005) Practical machine learning tools and techniques. Morgan Kaufmann 578

WWF, Turkey's footprint report (2012) http://awsassets.wwftr.panda.org/downloads/turkiyenin_ekolojik_ayak_izi_raporu.pdf Date of access: 07.02.2021

WWF, Türkiye İçin Düşük Karbonlu Kalkınma Yolları ve Öncelikleri Raporu (2015) http://awsassets.wwftr.panda.org/downloads/20151007_turkiye_icin_duuk_karbonlu_kalknma_yollar_ve_oncelikleri_rapor.pdf Date of access: 07.02.2021

Yılmaz İ (2016) Elon Musk. Science and Technical Journal 49(584):48–50

Zhang M, Mu H, Ning Y (2009) Accounting for energy-related CO_2 emission in China, 1991–2006. Energy Policy 37(3):767–773