CO₂ emissions in the USA: new insights based on ANN approach

Mihai Mutascu¹,²,³

Received: 22 February 2022 / Accepted: 30 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
The paper’s main aim is to forecast the carbon dioxide (CO₂) emissions in the USA and its related components, analysing the contributions of each of those components to CO₂ total volume. The empirical ground is a mix of non-linear tools, combining the artificial neural network (ANN) parametric method with a vector autoregressive (VAR) estimator. ANN includes 1 layer and 20 neurons, forecasting being based on the economic growth and net trade effects doubled by different types of renewable energy consumption. The accuracy of estimations for 14 targeted categories of CO₂ emissions is ensured by 4360 observations, with 10 types of inputs over 1984M01–2020M04. ANN seems to offer superior forecasting accuracy compared with the widely used autoregressive methods, such as VAR model, but seems to be weak in capturing the output ‘spike’ forms. The main findings show that, although economic growth and net trade have an important contribution to the targeted outputs, the more prominent ones are wind, solar and total biomass energy consumption. Therefore, the CO₂ emissions can be better controlled through non-polluting capacities, in parallel with the use of wind, solar and total biomass energies. The tool excellently predicts the CO₂ emissions during pandemic crises being a good instrument in policy decisions. Modest contributions to CO₂ prediction seem to have energy consumption generated by waste, hydroelectric power and renewable geothermal systems. This underlines an unclear current status given their collateral effects in environmental damages and high investment costs. The paper contributes to the literature in several ways. It is one of the first works focused on CO₂ emissions forecasting in the USA based on a mixed approach by ANN and VAR types, considering an extended pallet of inputs to predict the volume of total CO₂ emissions but also its components. As a novelty, the inputs combine both economic and environmental determinants. Not at least, the estimations are performed based on a large span, with monthly frequency.

Keywords CO₂ emissions · Determinants · ANN · Learning machine · USA

JEL classification Q30 · Q53 · C15

Introduction
During the last decades, gas emissions started to occupy high positions in the international policy agenda. Their destructive effects preoccupy not only the policymakers and researchers from all over the world but also many non-governmental organisations, companies or individuals. Therefore, the research interest in the field rapidly grew up, being focused on both theoretical and empirical explorations.

In this context, the USA is of particular interest facing high gas emissions levels over the last decades. Currently, this country is second-ranked globally after China in terms of gas emissions, followed by India, Russia and Japan. According to the Emissions Database for Global Atmospheric Research (Edgar 2019), the total CO₂ gas emissions registered in the USA reached 5275.48 mega tonnes in 2018,
compared with 11255.88 mega tonnes in China. In the same year, India registered 2621.92 mega tonnes, Russia 1748.35 mega tonnes and Japan 1198.55 mega tonnes. The total energy CO₂ emissions in the USA also exhibited an interesting evolution, as Figure 1 plots.

Figure 1 clearly shows that the total energy CO₂ emissions continuously increased until May 2004, reducing during the crisis from 2008 to 2009. A flat sinusoidal trend is registered until 2019, with a strong rebound in 2020, as the effect of the coronavirus disease (COVID-19). Not at least, it is noteworthy that the USA does not ratify the Kyoto Protocol, signed in 1997. With several amendments, the main target of the agreement is the reduction of greenhouse gas emissions in signatory countries (i.e. 37 industrialised countries and the European Union (EU)).

Many viable solutions were proposed to reduce gas emissions in the last years, using renewable energy being the most important. Hence, a wide range of renewable energy is considered, from hydroelectric, geothermal and solar types to biofuels and biomass ones. This substantially improved the modelling techniques of CO₂ emissions and their related forecasting methods.

In this light, the paper aims to predict the CO₂ emissions in the USA, taking into account the economic growth, renewable energy consumption and net trade. In order to obtain accurate forecasting, the artificial neural network (ANN) methodology is used as the principal tool. A classical econometric vector autoregressive (VAR) model is employed in parallel for comparison. This allows reinforcing the results, highlighting the valence of ANN approach. Both proposed methodologies perform nonlinear estimations, and the ANN generally offering better quality results (Gallo et al. 2014). The sample covers the period 1984M01–2020M04.

Grossman and Krueger’s (1991) effects are given the theoretical ground, controlled through different types of renewable energy consumption and net international trade. The core is the link between gas emissions and economic growth evidenced by Grossman and Krueger (1991). There are three theoretical transmission channels. First channel is called the ‘scale effect’, showing that growth generates pollution due to increased output. This output requires additional inputs and, therefore, more natural resources, stimulating pollution. Second channel is the ‘technological effect’, revealing that using new technologies is not the result of environmental worries but the race for profit maximisation. The new technologies improve productivity which means fewer inputs. This lowers the prices, reducing pollution. Third channel is represented by ‘composition effects’. It supposes that the government and private follow the services as the economy grows-up as a vector of ‘technological effect’. Further, this allows reaching a low level of pollution (i.e. growth does not generate pollution if the output structure changes). An important hypothesis precedes Grossman and Krueger’s (1991) theory, namely the environmental Kuznets curve (EKC). EKC has its origin in well-known Kuznets’s (1955), claiming an inverted U-shape link between pollution and economic growth.

The valence of renewable energy consumption in the CO₂ emissions is clearly evidenced by IREA (2013). Different types of energies, such as wind, solar and hydro, promote friendly environments. In parallel, this creates, in the same time, new jobs and improves skills, reducing income inequality (Zoundi 2017). Several papers also prove a negative influence of renewable energy consumption on CO₂ emissions (Bilgili et al. 2016; Wesseh and Lin 2016) or mixed results (Jebli et al. 2015).

Finally, the trade also is an essential ingredient for CO₂ emissions, mainly because of its EKC implications via the growth component. Two reasons are underlined by Cole (2004). On the one hand, the intensification of trade reduces pollution as the resources are used more efficient under the pressure of greater competition. On the other hand, the liberalisation of trade often required the restriction of imports of damaging environmental goods.
The contribution of the paper is fivefold. First, to the best of our knowledge, the paper is one of the first works devoted to predicting CO₂ emissions in the USA, based on a mix of methodologies. ANN method is employed in parallel with a classical VAR model, helping to reinforce the results by highlighting the valence of ANN against the classical econometric methodologies. Compared with classical tools, the ANN is superior in prediction accuracy. As Appiah et al. (2018) note, the ANN can be easily used for any arbitrary function, works well with any type of model, allows for performing nonlinear estimations and, not least, considers all categories of inputs. However, they are erroneous, incomplete or incoherent. Second, it also analyses the contributions of each considered determinant to CO₂ total volume, running from the gross domestic product (GDP) to different types of consumption. Third, unlike the existing studies, the paper predicts not only the volume of total CO₂ emissions but also its components derived from coal, natural gas, aviation gasoline, distillate fuel oil, hydrocarbon gas liquids, jet fuel, kerosene, lubricants, motor gasoline, petroleum coke, residual fuel oil, other petroleum products and petroleum, excluding biofuels. Fourth, also as a novelty, the study considers a generous set of renewable types of consumption, as follows: hydroelectric power, geothermal energy, solar energy, wind energy, wood energy, waste energy, biofuels and biomass energy, respectively. Finally, the proposal uses an extended dataset with monthly frequency, improving the accuracy of estimations (i.e. 436 observations and 4796 inputs).

The remainder of the paper is organised as follows: the ‘Literature’ section reviews the literature, the ‘Data and methodology’ section describes the data and methodology, while the ‘Results’ section presents the empirical results. Finally, the ‘Conclusions’ section concludes.

Literature

The literature regarding CO₂ emissions forecasting is generous, with a wide range of methodologies, types of predicated gas emissions, targeted countries, covered periods of exploration or dataset frequencies.

Salisu et al. (2018) state that the methodologies of CO₂ emissions forecasting are various, running from univariate models (i.e. autoregressive integrated moving average (ARIMA) and autoregressive fractionally integrated moving average (ARFIMA)) to univariate volatility models (i.e. generalised autoregressive conditional heteroscedasticity (GARCH)) and multivariate models (i.e. vector autoregressive (VAR) and vector error correction (VEC)). For example, Magazzino (2014) combines growth, CO₂ emissions and energy use by following a panel VAR analysis. Targeting the ASEAN group (i.e. Indonesia, Malaysia, the Philippines, Singapore, Thailand and Brunei Darussalam), over 1971–2007, he finds that the error variances in the CO₂ emissions are sensitive to disturbances in both growth and in CO₂ equations. Additionally, the author proves that the energy use Granger causes economic growth, validating the ‘growth hypothesis’. In the same vein, Magazzino (2016a) explores South Caucasus and Turkey over 1992–2013. His panel VAR estimations show that, in all three equations, the error variances in the CO₂ emissions are sensitive to disturbances. Herein, there is no causality between energy use and growth, suggesting that the ‘neutrality hypothesis’ holds. Unlike his previous studies, Magazzino (2016b) prefers the cointegration technic with breaks for analysing the same South Caucasus and Turkey area. The findings evidence a long-run relationship between growth, CO₂ emissions and energy use over 1992–2013. Finally, the author validates the ‘conservation hypothesis’ for Armenia, ‘neutrality hypothesis’ for Turkey and mixed results for Turkey (i.e. both ‘feedback hypothesis’ and ‘growth hypothesis’).

A handy and extended literature review regarding the methods of CO₂ emissions forecasting is offered by Abdul-lah and Pauzi (2015). They identify more heterogeneous tools, such as the ANN, grey models, computer-based simulation models, optimal growth models, linear regressions, pinch analyses, ‘top-down’ methods or adaptive neuro-fuzzy intelligent system approaches.

In the last years, the literature regarding CO₂ emissions forecasting based on ANN quickly extended with the development of computational models, having origins in the structure and functions of biological neural networks.

For example, Radojević et al. (2013) follow the ANN to forecast the CO₂ emissions in Serbia, also performing parallel estimations for Bulgaria. The used software tool NeuroShell 2 offers excellent predictions covering 1999–2007. The main inputs are as follows: the share of renewable sources of energy, GDP, gross energy consumption and energy intensity, respectively. Liu et al. (2012) choose China to forecast CO₂ emissions with good results. Their sample covers the period 1990–2010, while the input parameters are the GDP, exports, consumer price index (CPI), investment in fixed assets and population. The CO₂ emissions have origins in fossil fuels and cement production. China also is investigated by Li et al. (2010) proposing a unique ANN tool, namely the radial basis function (RBF). Herein, the total energy consumption is converted to CO₂ emissions from 1990 to 2010. The findings show that the RBF approach offers a superior approximation of CO₂ emissions compared to the classical ANN techniques. A similar study is proposed by Sun and Sun (2017) regarding China.

Baareh (2013) constructs an ANN for CO₂ emissions forecasting by selecting oil, natural gas, coal and primary energy consumption as inputs. He claims that the ANN is a powerful and efficient tool for predicting CO₂ emissions. Sözen et al. (2007) focus their study on Turkey, covering 2007–2020. Based on an ANN, the authors obtain accurate
estimations for the greenhouse gases (GHGs) by considering sectoral energy consumption as an indicator. The GHG is captured via CO$_2$, carbon monoxide (CO), sulphur dioxide (SO$_2$), nitrous dioxide (NO$_2$) and emissions of non-methane volatile organic compound (E), respectively. Alternatively, the scaled conjugate gradient and Levenberg-Marquardt (LM) tools are also considered.

Unlike them, Behrang et al. (2011) propose an integrated multi-layer perception neural network and Bees Algorithm (BA) approach. The CO$_2$ emissions are successfully forecasted based on a world-wide aggregate sample with 27 years, having as inputs the world’s population, GDP, oil trade movement and natural gas trade movement, respectively. As a novelty, Han et al. (2022a) recently propose a dendrite net method that relied on the adaptive mean square gradient. Their target is to predict the energy consumption of buildings in order to attenuate carbon dioxide emissions. Additionally, the authors show that the method also improves energy saving. In the same vein, Han et al. (2022b) link building energy consumption and carbon dioxide by using the input-output-based data envelopment analysis. Their saving proposed method supports carbon dioxide reduction and offers the ground for a competitive socio-economic sustainable development. In the same idea of carbon reduction, the complex industrial processes are investigated by Wu et al. (2022), developing a radial basis function neural network. They integrate multi-dimensional scaling to ensure energy optimisation and carbon emission analysing. As the authors claim, the findings are reinforced by University of California Irvine (UCI) datasets. A more advanced approach suggests Geng et al. (2021) in the industrial process. The authors develop a new gated convolutional neural network (CNN), showing that the tool excellently models the polypropylene and purified terephthalic acid industrial processes.

Curiously, according to Pérez-Suárez and López-Menéndez (2015), the theoretical core of those contributions is not centred on the EKC, despite its popularity. Salisu et al. (2018) show that many studies validate the existence of the EKC (e.g. Shafic and Bandyopadhyay 1992; Panayotou 1993; Grossman and Krueger 1995; Ang 2007; Fodha and Zaghdoud 2010; Kohler 2013; Aydoğan and Vardar 2020; Malik et al. 2020), while a significant group of works does not empirically support the effect (e.g. Paudel et al. 2005; Pao and Tsai 2010; Robalino-López et al. 2014). Dinda (2004) offers a valuable review of the literature regarding the EKC. Moreover, the World Commission on Environment and Development (WCED 1987) advances a contrary EKC vision, stressing that there is a U-shape link between environmental degradation and growth. Called the Brundtland Curve hypothesis, this supposes that in poor countries, environmental degradation is high initially, decreases as the economy expands, but rises after a turning point of growth.

In many analyses, the EKC is sustained by additional control determinants. Out of them, renewable energy consumption is often considered (Bilgili et al. 2016; Zoundi 2017), agriculture size (Liu et al. 2017), foreign trade (Chen et al. 2019; Dogan and Turkekul 2016) urbanisation or financial development (Dogan and Turkekul 2016).

The literature devoted to CO$_2$ emissions forecasting in the USA is relatively scarce. Officially, the prediction of CO$_2$ emissions in the USA is made by Energy Information Administration (EIA 2020b). The primary gas emission sources are petroleum, natural gas, coal and total fossil fuels. There are three central modelling systems, as follows: World Energy Projection System (WEPS), offering annual global energy projections; Short-Term Integrated Forecasting System (STIFS), outputting mootherly forecasting; and National Energy Modeling System (NEMS), generating annual US energy estimations. Various tools are used, ranging from linear regressions, method of solving for equilibrium, electricity load shapes, cointegration analysis, time-series benchmarking and mathematical programming.

Silva (2013) proposes alternative estimates to the official ones by entering a battery of methods, such as the ARIMA, Holt-Winters and exponential smoothing, before introducing the singular spectrum analysis (SSA). Auffhammer and Steinhauser (2012) propose a different approach using the US state-level panel dataset of CO$_2$ emissions. They claim that the well-known models in the literature and models considering the emissions per capita or different in-sample selection criteria allow obtaining less accurate estimations than the constructed model based on the out-of-sample loss measure defined over aggregate emissions. Bennedsen et al. (2021) estimate the CO$_2$ emissions in the USA based on a structural augmented dynamic factor model, with 226 observations. Their main finding shows that the Residential Utilities Index (RUI) is crucial for the CO$_2$ emissions predictions. In a very recent work, Eberle and Heath (2020) estimate the carbon dioxide emissions from electricity generation in the USA. The authors offer an innovative approach with two scenarios of future emissions: one with low emissions and another with high emissions. The core of their estimations is represented by a mix of ‘ingredients’, combining the capacity forecasts of US electric generation with technology- and fuel-specific capacity factors and emission factors. They found that if no shifts in electricity sources are made, four of the six data products could exclude 1.1 to 3.3% of CO$_2$ emissions from US electricity generation in 2040.

Three literature gaps can be identified. First, no paper is devoted to the CO$_2$ emissions forecasting in the USA based on the mix of tools by ANN and VAR types. Second, no work related to the USA uses the GDP and net trade for CO$_2$ emission prediction combined with additional control factors. Third, just a few papers consider the sources of CO$_2$ emissions in parallel with various types of renewable...
energy consumption as the main inputs for the CO$_2$ emission estimations in the USA. Not at least, no paper forecasts the components of CO$_2$ emissions estimations in the USA, in parallel with an analysis of the contribution of each type of renewable energy consumption as inputs.

Therefore, the paper addresses all gaps mentioned above using the ANN approach with an extended set of inputs. The Grossman and Krueger’s (1991) growth effects, net trade and different types of renewable energy consumption are the core determinants of CO$_2$ emissions forecasting in the USA, including CO$_2$-related components.

**Data and methodology**

**Data**

The CO$_2$ emissions forecasting in the USA based on the ANN follows an extended sample with 4360 observations, covering the period 1984M01–2020M04.

The targeted output is the total volume of CO$_2$ emissions (i.e. dependent variable), with its related components derived from coal, natural gas, aviation gasoline, distillate fuel oil, hydrocarbon gas liquids, jet fuel, kerosene, lubricants, motor gasoline, petroleum coke, residual fuel oil, other petroleum products and petroleum, excluding biofuels, respectively.

The predictions follow an extended set of inputs (i.e. independent variables), including the GDP in order to capture the Grossman and Krueger’s (1991) effects, net trade (Cole 2004; Chen et al. 2019; Dogan and Turkekul, 2019) and eighth types of renewable energy consumption (Jebli et al. 2015; Wesseh and Lin 2016; Bilgili et al. 2016; Zoundi and eighth types of renewable energy consumption (Jebli et al. 2015; Wesseh and Lin 2016; Bilgili et al. 2016; Zoundi 2017). Those renewable energy consumption inputs come from hydroelectric power, geothermal energy, solar energy, wind energy, wood energy, energy, energy, biofuels and total biomass energy.

Tables 3 and 4 (Appendix) describe the inputs and outputs, units of measure and data sources, while their descriptive statistics are shown in Table 5 (Appendix). All variables have been seasonally adjusted, with GDP and net trade exceptions already deseasonalised.

According to Sola and Sevilla (1997), good network performance and reduced errors in the training process can be reached by using the variables in their normalised range forms into [0,1].

Generally, the normalisation conversion has this form:

\[
x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]

where \(x'\) is the normalised variable \(x\), with \(x_{\text{max}}\) and \(x_{\text{min}}\) its maximum and minimum levels, respectively.

**Methodology**

The estimation of total CO$_2$ emissions and its related components (i.e. 14 total outputs) is performed based on a multilayer ANN perception approach, with 11 inputs (i.e. GDP, net trade and eight types of renewable energy consumption), covering 1984M01–2020M04. An alternative method is developed to reinforce the valence of the ANN method. In this way, an alternative VAR model is constructed,\(^1\) in order to follow the nonlinearity approach of ANN.

**ANN tool**

The use of ANN methodology in CO$_2$ emissions forecasting has gradually gained in popularity over the last decades. Unfortunately, the classical tools have a low propensity for modelling complex phenomena, such as those in meteorology or air pollution. Moreover, as Gallo et al. (2014, p. 33) note, the ANN methods ‘are widely used when you have no evident mathematical relationship between the variables. So, in the area of air quality forecasting the ANN model can be applied being not known a priori’.

The multilayer ANN perception model represents a nonlinear approach. Its architecture includes an extended network composed of neurons, each neuron being endowed with propagation, activation and output functions (Figure 2).

The key idea of the ANN tool is that the outputs (i.e. output layers) are predicted based on \(m\) inputs (i.e. input layers) through the ‘hidden layers’, while each layer performs a net output related to neuron \(j\) (\(\text{net}_j\)) that receives the outputs \(o_{i1}, o_{i2}, \ldots, o_{in}\) from neurons \(i1, i2, \ldots, in\) (Kreiner and Duca 2019). The slope of variables belonging to the activation function is calculated based on the weights \(w_{ij}\). The conversion of \(o_i\) into a weighted sum of the outputs is allowed based on the propagation function of neuron \(j\) supported by the aforementioned weights. Supposing \(b_j\) the intercept of activation function, the transformation process has this form:

\[
\text{net}_j = b_j + \sum_{i \in I} o_i w_{ij}
\]

The propagation function reduces the dimensionality of the neuron’s \(j\) inputs from \(n\) to \(l\), while an optimisation procedure can minimise the sum of in-sample squared errors.

In order to obtain accurate CO$_2$ emission predictions, the ANN configuration is designed by following the approach of Tümer and Akkuş (2018). Their proposal has excellent results, especially in the macroeconomic forecasting field.

\(^1\) Because of lack of space, only the simulation of the total volume of CO$_2$ is offered, the rest of the components being available upon request.
Consequently, the ANN is constructed by retaining the feed-forward back-propagation network type. The chosen training function is ‘Levenberg-Marquardt’ (trainlm), with ‘gradient descent’ (learngdm) as adaptive learning function. The mean squared error (MSE) represents the performance function, while ‘hyperbolic tangent’ (tansig) the transfer one. For all learning algorithms in the training process, the number of total iterations is set at 2000.

The optimal ANN architecture is chosen by alternatively testing different combinations of layers-neurons, more precisely: (1, 10), (1, 20), (2, 10), (2, 20), (3, 10) and (3, 20), respectively. Unfortunately, no very clear rule of selection exists. That is why optimal ANN architecture selection often follows the correlation coefficient ($R$) as the main ‘key control’, the coefficient testing the performance criteria of ANN training. Further, the root mean square error (RMSE) and coefficient of determination ($R^2$) allow for testing the accuracy of estimations. Both tests have high valence in assessing ANN’s performances. Finally, the model’s performance is validated in terms of accuracy by comparing the simulated data with the historical output values.

Additionally, the relative importance of input variables to the targeted outputs is calculated by following Garson’s Algorithm (Garson 1991). In other words, the algorithm allows identifying the contribution of each input to the target variables. The flowchart of the experimental process is plotted in Fig. 3.

The flowchart reveals an ANN routine, running from the dataset treatment to final outputs, supposing four different stages. In the first stage, the raw dataset, including both inputs and outputs, is normalised to ensure accurate estimations. In the second stage, the optimal ANN structure is created; training and testing further the sets of variables. The dataset with 4360 observations, covering 1984M01–2020M04, is used in the training sequence. The sample includes 11 inputs (i.e. GDP, net trade and eight types of renewable energy consumption) and 14 outputs (i.e. total CO$_2$ emissions and its components). The weight matrices are calculated in the third stage, allowing generating the main outputs in the last stage. Herein, the contribution of each input to the estimated output can also be identified.

**VAR method**

VAR tools are widely used to estimate CO$_2$ emissions, representing the classical methodologies in the field (Alam and AlArjani 2021).

This type of stochastic estimator considers the interest variables in their lagged forms (i.e. endogenous variables), offering the possibility to isolate their interaction through different control determinants (i.e. exogenous variables). In this case, the GDP and CO$_2$ emissions are the endogenous variables, while the net trade and eight types of renewable energy consumption are used as control exogenous determinants. GDP and CO$_2$ emissions are interest determinants to follow the EKC link, the central key of study. Noteworthy, the GDP squared is not considered as GDP’s nonlinear effect is directly captured by VAR based on lags. In this case, the VAR has this shape:

$$CO_2_t = \alpha_1 + \sum_{i=1}^{k} \beta_i CO_2_{t-i} + \sum_{j=1}^{k} \gamma_j GDP_{t-j} + \omega_1 Z + \sigma_1 t$$

$$GDP_t = \alpha_2 + \sum_{i=1}^{k} \phi_i CO_2_{t-i} + \sum_{j=1}^{k} \mu_j GDP_{t-j} + \omega_2 Z + \sigma_2 t$$

where $\alpha_1$ and $\alpha_2$ are the constants; $\beta$, $\gamma$, $\phi$ and $\mu$ denote the coefficient of endogenous variables that vary over time $t$ with lags $i$ and $j$ ($i,j = 1$ to $k$). $Z$ is the vector of control variables, with coefficients $\omega_1$ and $\omega_2$, assuming that the errors
Fig. 3 Flowchart of the ANN experimental process

\[ \sigma_1 \text{ and } \sigma_2 \text{ are not correlated. The variables are treated in their logarithm forms, except GDP expressed as an index, while the variables with negative signs are upwardly rescaled to obtain strict positive values. VAR model considers variables as stationary processes to avoid any bias in estimations. Optimal common selected lag is suggested by the minimal values of Akaike and Schwarz criteria.} \]

Results

ANN estimations

In order to construct an optimal ANN configuration, different scenarios are tested through the correlation coefficients (\( R \)) by alternatively using different ‘layers-neurons’ configurations. Table 1 shows followed scenarios based on the repetitive training routine process and their related \( R \):

The repetitive ANN training routines for considered ‘layers-neurons’ configurations generate many closed levels of \( R \), as Table 1 reveals. Although the \( R \) values are centred on 0.95, the structures with 1 layer and 20 neurons, and 2 layer and 20 neurons, seem more appropriate to construct an optimal ANN to forecast the total volume of CO\(_2\) emissions and its components in the USA. More precisely, the highest \( R \) levels (i.e. \( \approx 0.95485 \) and 0.95387) are registered by the (1, 20) and (2, 20) scenarios, recommending those ANN configurations for accurate predictions.

Table 2, the RMSE and \( R^2 \) are calculated for each scenario to test the accuracy of estimations.

By following the highest \( R^2 \) and/or lower RMSE, both tests in Table 2 suggest that good predictions in terms of accuracy can be obtained by calibrating the ANN with 1 layer and 20 neurons (1, 20) or 2 layers and 20 neurons (2, 20). Corroborating with \( R \) level, it is clear that the more appropriate ANN architecture to accurately forecast the total volume of CO\(_2\) emissions and its components in the USA is

| Number of layers and neurons | Correlation coefficient (\( R \)) |
|-----------------------------|---------------------------------|
| 1 layer 10 neurons (1, 10)  | 0.94542                         |
| 1 layer 20 neurons (1, 20)  | **0.95387**                     |
| 2 layers 10 neurons (2, 10) | 0.94634                         |
| 2 layers 20 neurons (2, 20) | **0.95485**                     |
| 3 layers 10 neurons (3, 10) | 0.95163                         |
| 3 layers 20 neurons (3, 20) | 0.95251                         |
For this ANN configuration, the RMSE ≈ 0.070275, while $R^2 ≈ 0.9248$.

The simulated output regarding the total volume of CO$_2$ emissions in the USA is presented in parallel with afferent historical data in Fig. 4. The retained (1, 20) ANN has a forecasting power of almost 92.42%, with a related error of 7.58%.

Figure 4 evidences that the ANN simulation for the total volume of CO$_2$ emissions in the USA offers excellent results. The estimations follow the historical values with pretty high accuracy. This strongly supports the idea that the total volume of CO$_2$ emissions can be successfully predicted based on the ANN, having as main inputs the GDP, net trade and, very important, renewable energy use and its components.

Moreover, this approach excellently covers the economic crisis from 2008 to 2009, and the pandemic coronavirus disease (COVID-19) started at the end of 2019. Unfortunately, the proposed ANN does not capture all registered historical ‘spikes’, being more prone to ‘smooth’ dynamics. In this light, the ‘dynamic neural networks could have a better performance when it comes to estimating high and peak inflows’ (Hadiyan et al. 2020, p. 8). An extensive palette of models developed during the last years to better capture the peak inflows, the spiking neural network (SNN) type being the prominent one$^2$. Although such models offer good quality results, they seem to be more efficient in the machine learning algorithm in real life (Lee et al. 2020), including human or animal behaviours (Gerstner and Kistler 2002).

The forecasting of the components of CO$_2$ emissions is shown in Figs. 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17 (Appendix), maintaining the accuracy of the aggregated volume of CO$_2$ emissions. Interesting, the ANN excellently captures the fall of CO$_2$ components related to the most affected economic sectors during the pandemic coronavirus crisis. Concretely, those components are distillate

$^2$ A Matlab code is available and dated April 5th, 2022 (Versace 2022), at https://ch.mathworks.com/ matlabcentral/fileexchange/25931-spiking-neurons-simulator.
fuel, hydrocarbon gas liquids, jet fuel, lubricants and motor gasoline (without ethanol). Between March and April 2020, the generalised world-wide lockdown negatively altered the transportation industry, which uses the types mentioned above of energy. The total number of commercial flights collapsed by 73% in the USA, while truck tonnage transportation by 9.18% (Mack et al. 2021). Private transportation was also significantly reduced, with a decongestion of highway traffic by around 30% in 2020 compared with 2019. For example, the collapse has been by 36% in Los Angeles, 30% in New York and 25% in Miami (Kelly and Sharafedin 2021). Therefore, a strong rebound has also registered in the related energy use, the consumption redaction being more than significant.

Based on Garson’s Algorithm, the contribution of each input to targeted outputs is presented in Table 6 (Appendix). Wind energy consumption explains 13.61% of predicted outputs, while solar and total biomass energy consumptions contribute 13.24% and 12.61%, respectively. The GDP and biofuels consumption covers 10.48% and 9.97% of forecasting, while net trade and wood energy consumption are around 9.06% and 8.65%, respectively. The lowest contribution levels to the predicted outputs have the waste, hydroelectric power and geothermal renewable energy consumption, each with cca. 7–8% of the total.

The results highlight the importance of the GDP and net trade in the CO2 emissions forecasting, but with the strong support for renewable energy consumption. The outputs are in line with the contribution of Grossman and Krueger (1991), reinforcing the link between environment and growth. The valence of net trade is also underlined, validating the findings of Cole (2004), Chen et al. (2019) or Dogan and Turkekul (2019). But more important is the crucial contribution of renewable energy consumption to CO2 emissions forecasting, fully confirming Jebli et al. (2015), Wesseh and Lin (2016), Bilgili et al. (2016) and Zoundi (2017).

The essential explanatory renewable energy variables are wind energy consumption, followed by solar and biomass energy components. All those offer strong support for the future of green energy. The analysis highlights the low importance in CO2 emissions forecasting of waste, hydroelectric power and geothermal renewable energy consumption. Hydroelectric power production is ‘eco-friendly’ on the one side but can generate substantial natural damage on the other hand.

**VAR alternative estimations**

The variables’ stationarity tests are presented in Table 7 (Appendix). Augmented Dickey-Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are employed, considering both constants, and constant and trend. Corroborating the results, all variables are I(1), except for hydroelectric power consumption, geothermal energy consumption and wood energy consumption. Therefore, in the VAR model, all I(1) variables are considered in their first differences.

The complete VAR estimations are presented in Table 8 (Appendix). Optimal common selected lag is 4 being indicated by the minimal values of Akaike and Schwarz criteria. According to the LM test in Table 9 (Appendix), no residual serial correlations are observed at lag 4, while the residuals are homoskedastic at the limit (i.e. Chi2 = 134.9, with a probability of 0.0162).

The simulated output of CO2 emissions in the USA in parallel with its historical values is shown in Fig. 18 (Appendix). In order to ensure comparability with ANN simulations, VAR estimation is normalised to 1. Compared with ANN, the findings reveal that the VAR method offers a good prediction of trend dynamics but falls under high gaps between simulated and historical output values. The conclusion is reinforced by Fig. 19 (Appendix). This indicates that the simulated CO2 emissions with ANN have lower residual values than the VAR tool, despite not capturing better the ‘spike’ forms. The superior accuracy of ANN against the autoregressive models as VAR is in line with Lahane (2008) or Adebiyi et al. (2014).

The study has two significant limits. First, no extended set of inputs is considered because of the lack of data. Many of the main CO2 emissions factors evidenced by literature are not officially available with infra-annual frequency, such as the agriculture size, urbanisation or financial development. Second, forecasting quality is exclusively checked based on a classical econometric VAR method, not considering alternative ANN types. This allows us to compare the valence of ANN against the classical methodology and analyse the forecasting quality of similar ANN tools.

**Conclusions**

The paper offers a mixed tool approach for predicting the total volume of CO2 emissions and its related components in the USA by following a non-linear ANN parametric and VAR methodologies. The forecasting of 14 targeted categories of CO2 emissions is ensured by 4360 observations, including 10 types of inputs, over the period 1984M01–2020M04.

An optimal ANN architecture is proposed by selecting the more appropriate ‘layers-neurons’ pair from different computed combinations, such as (1, 10), (1, 20), (2, 10), (2, 20), (3, 10) and (3, 20). Following the criteria of highest R2 and lower RMSE tests (i.e. $R^2 \approx 0.9248$, and RMSE $\approx 0.070275$), the ANN is finally calibrated as (1, 20) configuration, with 1 layer and 20 neurons.
Overall, the forecasting power has 92.42%, with a related error of 7.58%. The tool is excellently adapted in capturing general crises, especially those pandemic ones, but partially falls under ‘spike’ forms. In parallel, a VAR stochastic method is also proposed to test the quality of ANN. The comparison between ANN and VAR simulations reveals that the ANN tool offers superior results from an accuracy point of view. Although the VAR generates good forecasting as trend dynamic, it completely falls in terms of residuals. Despite not capturing the ‘spike’ forms of CO2 emissions better, ANN experiences much lower residual values than VAR.

The results strongly support the idea that the CO2 emissions and their related components in the USA can be successfully predicted based on the growth effect by controlling for renewable energy consumption and net trade. Except for economic growth, the most significant contributions to targeted outputs are wind, solar and total biomass energy consumption. This supports the idea of CO2 control based on non-polluting capacities, corroborated by the intensively use of green energy by wind, solar and total biomass origins. Net trade and wood energy consumption are also not negligible, while waste, hydroelectric power and geothermal energy consumption have modest contributions. Herein, dam and reservoir construction can irreversibly alter water temperatures, water chemistry, river flow characteristics, and silt loads, negatively impacting both plants and animals (EIA 2021). As waste and geothermal renewable energy require high investment costs, their current use is modest.

In the light of policy implications, the proposed ANN approach based on growth effect, renewable energy consumption and net trade can be successfully used to predict the CO2 emissions and their related components in the USA. Several policy implications can be identified. First, such estimations are very useful for US policymakers in the environmental area and any other actor involved in the field (i.e. national statistical offices, international organisations and non-governmental organisations). The tool should be considered with caution given its lack of accuracy in the prediction of ‘spike’ values. Therefore, autoregressive methods seem to capture the dynamic trend better but are limited in their accuracy due to the high gap between simulated and historical output data. In this context, smoothed datasets seem more prone to be analysed via the ANN approach.

Second, in their CO2 predictions, the interested actors should focus on the growth and net trade but more pregnant on the wind, solar and total biomass energy consumptions. The energy consumption components coming from waste, hydroelectric power and renewable geothermal systems should be used in moderation given their unclear status (i.e. collateral environmental damages or high investment costs).

Thirds, the results offer important information for designing environmental policies. The input components are useful in CO2 predictions and in identifying the leading renewable energy sources that allow controlling the CO2 emissions.

As for further research, two directions should be identified. First research direction regards the insertion of more input variables with infra-annual frequency as soon as they are officially available. In this way, the agriculture size (Liu et al. 2017), urbanisation or financial development (Dogan and Turkekul, 2019) can be considered. Second research direction is related to the methodology. Herein, the dynamic ANN can be resorted as a gradual further step in using ANN tools in the CO2 emissions forecasting process. According to Hadiyan et al. (2020), such a technique generally adds more quality in estimations because all layers have feedback links with different time delays. However, the field of dynamic ANN research is still in progress, especially regarding the ‘spike’ behaviour (Versace 2022).
Table 3 Description of inputs

| Variables                              | Explanation                                           | U.M.                  | Source                  |
|----------------------------------------|-------------------------------------------------------|-----------------------|-------------------------|
| Input 1. GDP                           | Normalised index, seasonally adjusted                 | Index                 | OECD (2020)             |
| Input 2. Hydroelectric power consumption| Volume of hydroelectric power consumption            | Million metric tons of CO₂ | EIA (2020a, 2020b)     |
| Input 3. Geothermal energy consumption  | Volume of geothermal energy consumption              | Million metric tons of CO₂ | EIA (2020b)            |
| Input 4. Solar energy consumption       | Volume of solar energy consumption                   | Million metric tons of CO₂ | EIA (2020a)            |
| Input 5. Wind energy consumption        | Volume of wind energy consumption                    | Million metric tons of CO₂ | EIA (2020b)            |
| Input 6. Wood energy consumption        | Volume of wood energy consumption                    | Million metric tons of CO₂ | EIA (2020a)            |
| Input 7. Waste energy consumption       | Volume of waste energy consumption                   | Million metric tons of CO₂ | EIA (2020b)            |
| Input 8. Biofuel consumption            | Volume of biofuel consumption                        | Million metric tons of CO₂ | EIA (2020a)            |
| Input 9. Total biomass energy consumption| Volume of total biomass energy consumption           | Million metric tons of CO₂ | EIA (2020b)            |
| Input 10. Net trade                     | Net trade expressed in US dollar converted (billions), seasonally adjusted | US dollar (billions) | OECD (2020)             |

Table 4 Description of outputs

| Variables                              | Explanation                                           | U.M.                  | Source                  |
|----------------------------------------|-------------------------------------------------------|-----------------------|-------------------------|
| Output 1. Total CO₂ emissions          | Volume of total energy CO₂ emissions                  | Million metric tons of CO₂ | EIA (2020a)            |
| Output 2. Coal CO₂ emissions           | Volume of coal, including coal coke net imports, CO₂ emissions | Million metric tons of CO₂ | EIA (2020b)           |
| Output 3. Natural gas CO₂ emissions    | Volume of natural gas, excluding supplemental gaseous fuels, CO₂ emissions | Million metric tons of CO₂ | EIA (2020a)            |
| Output 4. Aviation gasoline CO₂ emissions | Volume of aviation gasoline CO₂ emissions             | Million metric tons of CO₂ | EIA (2020a)            |
| Output 5. Distillate fuel oil CO₂ emissions | Volume of distillate fuel oil, excluding biodiesel, CO₂ emissions | Million metric tons of CO₂ | EIA (2020a)            |
| Output 6. Hydrocarbon gas liquids, CO₂ emissions | Volume of hydrocarbon gas liquids, CO₂ emissions | Million metric tons of CO₂ | EIA (2020b)            |
| Output 7. Jet fuel CO₂ emissions       | Volume of jet fuel CO₂ emissions                      | Million metric tons of CO₂ | EIA (2020a)            |
| Output 8. Kerosene CO₂ emissions       | Volume of kerosene CO₂ emissions                      | Million metric tons of CO₂ | EIA (2020b)            |
| Output 9. Lubricants, CO₂ emissions    | Volume of lubricants, CO₂ emissions                   | Million metric tons of CO₂ | EIA (2020a)            |
| Output 10. Motor gasoline, excluding ethanol CO₂ emissions | Volume of motor gasoline, excluding ethanol, CO₂ emissions | Million metric tons of CO₂ | EIA (2020b)            |
| Output 11. Petroleum coke CO₂ emissions | Volume of petroleum coke CO₂ emissions                | Million metric tons of CO₂ | EIA (2020a)            |
| Output 12. Residual fuel oil CO₂ emissions | Volume of residual fuel oil CO₂ emissions            | Million metric tons of CO₂ | EIA (2020b)            |
| Output 13. Other petroleum products, CO₂ emissions | Volume of other petroleum products, CO₂ emissions   | Million metric tons of CO₂ | EIA (2020a)            |
| Output 14. Petroleum, excluding biofuels, CO₂ emissions | Volume of petroleum, excluding biofuels, CO₂ emissions | Million metric tons of CO₂ | EIA (2020a, 2020b)     |
### Table 5: Descriptive statistics of parameters

| Variables                                      | Mean   | Median  | Maximum  | Minimum  | Std. Dev. | Obs. |
|------------------------------------------------|--------|---------|----------|----------|-----------|------|
| Input 1. GDP                                   | 100.0577 | 100.0027 | 101.8458 | 91.90456| 0.953175  | 436  |
| Input 2. Hydroelectric power consumption       | 14.54856 | 15.0695  | 19.182   | 5.094    | 3.087591  | 436  |
| Input 3. Geothermal energy consumption         | 14.58346 | 5.97     | 115.836  | -0.001   | 23.23624  | 436  |
| Input 4. Solar energy consumption              | 50.78641 | 7.367    | 270.486  | 0        | 73.25908  | 436  |
| Input 5. Wind energy consumption               | 189.3772 | 186.672  | 252.903  | 128.7    | 21.27616  | 436  |
| Input 6. Wood energy consumption               | 36.43025 | 37.2965  | 54.461   | 16.371   | 7.792112  | 436  |
| Input 7. Waste energy consumption              | 70.88745 | 23.5685  | 205.896  | 6.125    | 71.58952  | 436  |
| Input 8. Biofuel consumption                   | 296.6949 | 263.8605 | 456.707  | 178.544  | 72.95196  | 436  |
| Input 9. Total biomass energy consumption      | 610.0198 | 547.92   | 1059.944 | 395.833  | 157.6737  | 436  |
| Input 10. Net trade                            | 233.4066 | 229.3015 | 357.387  | 145.715  | 44.95899  | 436  |
| Output 1. Total energy CO2 emissions           | 446.9781 | 447.8565 | 559.581  | 307.125  | 44.65145  | 436  |
| Output 2. Coal CO2 emissions                   | 151.9561 | 154.2535 | 203.435  | 50.809   | 28.01467  | 436  |
| Output 3. Natural gas CO2 emissions            | 101.9618 | 97.754   | 185.55   | 50.21    | 26.05317  | 436  |
| Output 4. Aviation gasoline CO2 emissions      | 0.198321 | 0.187    | 0.652    | 0.065    | 0.074976  | 436  |
| Output 5. Distillate fuel oil CO2 emissions    | 45.82784 | 46.9605  | 57.763   | 32.128   | 6.025373  | 436  |
| Output 6. Hydrocarbon gas liquids, CO2 emissions | 7.375319 | 7.0545   | 12.639   | 4.301    | 1.706332  | 436  |
| Output 7. Jet fuel CO2 emissions               | 18.6427  | 18.7675  | 23.145   | 8.335    | 1.978329  | 436  |
| Output 8. Kerosene CO2 emissions               | 0.598489 | 0.461    | 3.172    | -0.202   | 0.563768  | 436  |
| Output 9. Lubricants, CO2 emissions            | 0.976445 | 0.9845   | 1.365    | 0.434    | 0.160283  | 436  |
| Output 10. Motor gasoline, excluding ethanol CO2 emissions | 90.34002 | 90.909   | 107.257  | 59.289   | 8.1998    | 436  |
| Output 11. Petroleum coke CO2 emissions        | 6.670239 | 6.609    | 11.723   | 2.228    | 1.552649  | 436  |
| Output 12. Residual fuel oil CO2 emissions     | 11.96178 | 11.809   | 30.381   | 1.676    | 5.748825  | 436  |
| Output 13. Other Petroleum products, CO2 emissions | 9.703564 | 9.6645   | 14.344   | 5.441    | 1.604561  | 436  |
| Output 14. Petroleum, excluding biofuels, CO2 emissions | 192.2947 | 191.11   | 233.509  | 133.104  | 14.86774  | 436  |
Table 6  Relative contribution of each input neuron to each hidden neuron, relative importance, and rank of inputs

| Explanation | Input 1 | Input 2 | Input 3 | Input 4 | Input 5 | Input 6 | Input 7 | Input 8 | Input 9 | Input 10 | Sum |
|-------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------|-----|
| Hidden 1    | 0.170   | 0.027   | 0.053   | 0.117   | 0.051   | 0.027   | 0.125   | 0.115   | 0.058   | 0.258     | 1   |
| Hidden 2    | 0.105   | 0.009   | 0.007   | 0.093   | 0.082   | 0.175   | 0.177   | 0.001   | 0.259   | 0.091     | 1   |
| Hidden 3    | 0.067   | 0.058   | 0.068   | 0.156   | 0.282   | 0.101   | 0.115   | 0.013   | 0.129   | 0.011     | 1   |
| Hidden 4    | 0.022   | 0.076   | 0.026   | 0.320   | 0.095   | 0.045   | 0.142   | 0.148   | 0.060   | 0.067     | 1   |
| Hidden 5    | 0.075   | 0.240   | 0.048   | 0.204   | 0.128   | 0.174   | 0.052   | 0.003   | 0.13    | 0.062     | 1   |
| Hidden 6    | 0.042   | 0.075   | 0.198   | 0.137   | 0.161   | 0.043   | 0.069   | 0.090   | 0.071   | 0.115     | 1   |
| Hidden 7    | 0.215   | 0.022   | 0.104   | 0.147   | 0.144   | 0.133   | 0.036   | 0.126   | 0.071   | 0.002     | 1   |
| Hidden 8    | 0.100   | 0.072   | 0.121   | 0.081   | 0.090   | 0.075   | 0.005   | 0.128   | 0.118   | 0.209     | 1   |
| Hidden 9    | 0.047   | 0.041   | 0.015   | 0.148   | 0.331   | 0.038   | 0.108   | 0.094   | 0.105   | 0.074     | 1   |
| Hidden 10   | 0.102   | 0.077   | 0.170   | 0.032   | 0.105   | 0.099   | 0.072   | 0.033   | 0.184   | 0.126     | 1   |
| Hidden 11   | 0.126   | 0.081   | 0.049   | 0.116   | 0.154   | 0.099   | 0.118   | 0.133   | 0.006   | 0.117     | 1   |
| Hidden 12   | 0.150   | 0.360   | 0.138   | 0.120   | 0.106   | 0.008   | 0.003   | 0.079   | 0.003   | 0.033     | 1   |
| Hidden 13   | 0.033   | 0.019   | 0.089   | 0.270   | 0.058   | 0.036   | 0.062   | 0.232   | 0.129   | 0.073     | 1   |
| Hidden 14   | 0.062   | 0.025   | 0.004   | 0.006   | 0.221   | 0.149   | 0.006   | 0.397   | 0.125   | 0.125     | 1   |
| Hidden 15   | 0.090   | 0.127   | 0.126   | 0.103   | 0.152   | 0.105   | 0.049   | 0.034   | 0.078   | 0.137     | 1   |
| Hidden 16   | 0.097   | 0.000   | 0.005   | 0.170   | 0.264   | 0.059   | 0.057   | 0.104   | 0.183   | 0.061     | 1   |
| Hidden 17   | 0.180   | 0.041   | 0.088   | 0.072   | 0.032   | 0.022   | 0.009   | 0.384   | 0.094   | 0.078     | 1   |
| Hidden 18   | 0.109   | 0.009   | 0.033   | 0.266   | 0.192   | 0.131   | 0.052   | 0.055   | 0.143   | 0.009     | 1   |
| Hidden 19   | 0.132   | 0.105   | 0.028   | 0.013   | 0.179   | 0.112   | 0.024   | 0.181   | 0.216   | 0.010     | 1   |
| Hidden 20   | 0.172   | 0.028   | 0.047   | 0.076   | 0.110   | 0.028   | 0.145   | 0.036   | 0.205   | 0.153     | 1   |
| Sum        | 2.095   | 1.492   | 1.417   | 2.648   | 2.722   | 1.730   | 1.568   | 1.994   | 2.521   | 1.811     | 20  |
| Relative importance | 10.48% | 7.46% | 7.09% | 13.24% | 13.61% | 8.65% | 7.84% | 9.97% | 12.61% | 9.06% | 100% |

$R_{i} = \frac{\sum_{j=1}^{n} |w_{ij}w_{jz}|}{\sum_{j=1}^{m} |w_{ij}w_{jz}|}$ denotes the relative importance ($R_{i}$) of neuron $x$ by $n$ number, with total number $m$ of hidden neurons $y$ and output neurons $z$ (Ibrahim 2013, p. 5697).
### Table 7 Stationarity tests

| Variables                                | ADF          |          | PP           |          | KPSS        |          |
|------------------------------------------|--------------|----------|--------------|----------|-------------|----------|
|                                          | Constant     | Constant+Trend | Constant     | Constant+Trend | Constant     | Constant+Trend |
| CO₂ (level)                              | −1.438***    | −0.663***| −1.913***    | −1.406***| 0.998       | 0.635    |
| CO₂ (1st difference)                     | −22.97       | −23.12   | −34.91       | −37.07   | 0.501*      | 0.092*** |
| GDP (level)                              | −1.032***    | −1.255***| −1.726***    | −1.941***| 0.141***    | 0.048*** |
| GDP (1st difference)                     | −7.106       | −7.217   | −7.032       | −7.151   | 0.245***    | 0.103*** |
| Net trade (level)                        | −1.571***    | −3.417*  | −2.159***    | −6.926   | 2.162       | 0.149*  |
| Net trade (1st difference)               | −22.13       | −22.11   | −38.22       | −38.19   | 0.033***    | 0.022*** |
| Hydroelectric power consumption (level)  | −3.803       | −4.229   | −4.464       | −5.269   | 0.794       | 0.104*** |
| Hydroelectric power consumption (1st difference) | −19.15       | −19.12   | −26.06       | −26.02   | 0.044***    | 0.038*** |
| Geothermal energy consumption (level)     | −3.643       | −3.581*  | −4.623       | −6.341   | 1.877       | 0.262   |
| Geothermal energy consumption (1st difference) | −12.37       | −12.56   | −40.91       | −45.32   | 0.387**     | 0.501   |
| Solar energy consumption (level)          | −0.987***    | −1.962***| −1.469***    | −2.898***| 1.922       | 0.261   |
| Solar energy consumption (1st difference) | −11.88       | −11.87   | −20.67       | −20.65   | 0.037***    | 0.036*** |
| Wind energy consumption (level)           | −0.651***    | −5.341   | −0.878***    | −6.102   | 2.590       | 0.258   |
| Wind energy consumption (1st difference)  | −13.83       | −13.81   | −88.01       | −88.12   | 0.135***    | 0.135** |
| Wood energy consumption (level)           | −3.389**     | −3.672*  | −7.292       | −8.784   | 1.117       | 0.425   |
| Wood energy consumption (1st difference)  | −15.61       | −15.59   | −35.19       | −34.38   | 0.038***    | 0.018*** |
| Waste energy consumption (level)          | −3.225*      | −2.601***| −3.627       | −2.942***| 0.912       | 0.346   |
| Waste energy consumption (1st difference) | −20.56       | −20.71   | −35.61       | −36.83   | 0.421**     | 0.111*  |
| Biofuel consumption (level)               | −1.081***    | −0.573** | −1.232***    | −0.635***| 2.587       | 0.346   |
| Biofuel consumption (1st difference)      | −28.64       | −28.67   | −28.74       | −28.76   | 0.279***    | 0.226   |
| Total biomass energy consumption (level)  | −0.882***    | −2.050***| −1.304***    | −3.039***| 2.061       | 0.519   |
| Total biomass energy consumption (1st difference) | −10.13       | −10.11   | −33.75       | −33.73   | 0.105***    | 0.104*** |

(1) ADF augmented Dickey-Fuller test (H₀ = the series has a unit root); PP Phillips-Perron test (H₀ = the series has a unit root); and KPSS Kwiatkowski–Phillips–Schmidt–Shin test (H₀ = the series is stationary); (2) *, ** and *** denotes 1, 5 and 10% level of significance.
### Table 8: VAR estimations

| Variables | CO₂   | GDP   |
|-----------|-------|-------|
| CO₂(−1)  | −0.463593 (0.04953) | −0.032220 (0.50092) |
| CO₂(−2)  | −0.370388 (0.05463) | 1.625707 (0.55240) |
| CO₂(−3)  | −0.085699 (0.05540) | 1.274405 (0.56020) |
| CO₂(−4)  | −0.079641 (0.04992) | 0.708397 (0.50482) |
| GDP(−1)  | 0.014334 (0.00515)  | 0.578838 (0.05209)  |
| GDP(−2)  | −0.284512 (0.14224) | 2.001605 (1.43842) |
| GDP(−3)  | 0.468976 (0.27199)  | −2.377649 (2.75051) |
| GDP(−4)  | −0.186003 (0.14286) | 0.661612 (1.44464) |
| Constant | −0.065103 (0.10376) | −0.509724 (1.04925) |
| Net trade| 0.003677 (0.00460)  | 0.049036 (0.04649)  |
| Hydroelectric power consumption | 0.006580 (0.00892) | 0.153880 (0.09021) |
| Geothermal energy consumption | −0.004224 (0.00595) | −0.026246 (0.06017) |
| Solar energy consumption | −0.001638 (0.00592) | 0.011332 (0.05985) |
| Wind energy consumption | 0.002545 (0.00437)  | 0.044559 (0.04422)  |
| Wood energy consumption | 0.007669 (0.01612)  | −0.052347 (0.16300) |
| Waste energy consumption | −0.045728 (0.01973) | −0.047953 (0.19957) |
| Biofuel consumption | 0.055119 (0.01985)  | 0.448048 (0.20076)  |
| Total biomass energy consumption | 0.069590 (0.02734) | 0.461488 (0.27652) |
| R-squared | 0.289823 | 0.398269 |
| Adj. R-squared | 0.260591 | 0.373501 |
| Akaike AIC | −4.636199 | −0.008644 |
| Schwarz SC | −4.466385 | 0.161171 |

Standard errors are presented in [...].
Fig. 5  Simulated data and historical data for coal CO₂ emissions in the USA

Fig. 6  Simulated data and historical data for natural gas CO₂ emissions in the USA
Fig. 7  Simulated data and historical data for aviation gasoline CO₂ emissions in the USA

Fig. 8  Simulated data and historical data for distillate fuel oil CO₂ emissions in the USA
Fig. 9  Simulated data and historical data for hydrocarbon gas liquids CO₂ emissions in the USA

Fig. 10  Simulated data and historical data for jet fuel CO₂ emissions in the USA
**Fig. 11** Simulated data and historical data for kerosene CO₂ emissions in the USA

**Fig. 12** Simulated data and historical data for lubricants, CO₂ emissions in the USA
Fig. 13  Simulated data and historical data for motor gasoline, excluding ethanol, CO₂ emissions in the USA

Fig. 14  Simulated data and historical data for petroleum coke CO₂ emissions in the USA
**Fig. 15** Simulated data and historical data for residual fuel oil CO₂ emissions in the USA

**Fig. 16** Simulated data and historical data for other petroleum products, CO₂ emissions in the USA
Fig. 17  Simulated data and historical data for petroleum, excluding biofuels, CO₂ emissions in the US

Fig. 18  VAR simulated and historical data for total volume of CO₂ emissions in the USA
Fig. 19  ANN and VAR simulated data and historical values of total volume of CO₂ emissions in the USA
Author contribution The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Data availability The data that support the findings of this study are openly available in OECD dataset, at https://doi.org/10.1787/data-00285-en, and in US Energy Information Administration, at https://www.eia.gov/totalenergy/data/browser/.

Declarations

Ethics approval Not applicable, because this article does not contain any studies with human or animal subjects.

Consent to participate Not applicable

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

References

Abdullah L, Pauzi HM (2015) Methods in forecasting carbon dioxide emissions: a decade review. J Teknol (Sciences & Engineering) 75(1):67–82

Adebiyi AA, Adewumi AO, Ayo CK (2014) Comparison of ARIMA and artificial neural networks models for stock price prediction. J Appl Math 2014:614342

Alam T, AlArjani A (2021) A comparative study of CO₂ emission forecasting in the gulf countries using autoregressive integrated moving average, artificial neural network, and holt-winters exponential smoothing models. Adv Meteorol 2021:8322590

Ang JB (2007) CO₂ emissions, energy consumption, and output in France. Energy Policy 35(10):4772–4778

Appiah K, Du I, Appah R, Quaicoe D (2018) Prediction of potential carbon dioxide emissions of selected emerging economies using artificial neural network. J Environ Sci Eng A7:321–335

Auffhammer M, Steinhauser R (2012) Forecasting the path of U.S. CO₂ emissions using state-level information. Rev Econ Stat 94(1):172–185

Aydoğan B, Vardar G (2020) Evaluating the role of renewable energy, economic growth and agriculture on CO₂ emission in E7 countries. Int J Sustain Energy 39(4):335–348

Baareh AK (2013) Solving the carbon dioxide emission estimation problem: an artificial neural network model. J Softw Eng Appl 6:338–342

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

Bennedesen M, Hillebrand E, Koopman SJ (2021) Modeling, forecasting, and nowcasting U.S. CO₂ emissions using many macroeconomic predictors. Energy Econ 96:1–17. [105118]. https://doi.org/10.1016/j.eneco.2021.105118

Bilgili F, Koçak E, Bulut Ü (2016) The dynamic impact of renewable energy consumption on CO₂ emissions: a revised Environmental Kuznets Curve approach. Renew Sust Energ Rev 54:838–845

Chen Y, Zheng W, Zhangji Z (2019) CO₂ emissions, economic growth, renewable and non-renewable energy production and foreign trade in China. Renew Energy 131(C):208–216

Cole MA (2004) Trade, the pollution haven hypothesis and the environmental Kuznets curve: examining the linkages. Ecol Econ 48:71–81

Dinda S (2004) Environmental Kuznets curve hypothesis: a survey. Ecol Econ 49(4):431–455

Dogan E, Turkekul B (2016) CO₂ emissions, real output, energy consumption, trade, urbanisation and financial development: testing the EKC hypothesis for the USA. Environ Sci Pollut Res Int 23:1203–1213

Eberle AL, Heath GA (2020) Estimating carbon dioxide emissions from electricity generation in the United States: how sectoral allocation may shift as the grid modernises. Energy Policy 140:111324

Edgar (2019) Emissions database for global atmospheric research. In: Crippa M, Orenggioni G, Guizzardi D, Muntean M, Scafe A, Lo Vullo E, Solazzo E, Monforti-Ferrario F, Olivier JGI, Vignati E (eds) Fossil CO₂ and GHG emissions of all world countries - 2019 Report. Office of the European Union, Luxembourg

EIA (2020a) US energy information administration online database. https://www.eia.gov/international/data/world. Accessed October 2020

EIA (2020b) Analysis & projections, Handbook of Energy Modeling Methods, October, 14th. https://www.eia.gov/analysis/handbook/

EIA (2021) Hydropower explained, hydropower and the environment. https://www.eia.gov/energyexplained/hydropower/hydropower-and-the-environment.php

Fodha M, Zaghdoud O (2010) Economic growth and pollutant emissions in Tunisia: an empirical analysis of the environmental Kuznets curve. Energy Policy 38(2):1150–1156

Gallo C, Contò F, Fiore M (2014) A neural network model for forecasting CO₂ emission. Agris On-line Pap Econ Inform 6(2):31–36

Garson GD (1991) Interpreting neural network connection weights. AI Expert 6:47–51

Geng Z, Chen Z, Meng Q, Han Y (2021) Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes. IEEE Trans Ind Inform. https://doi.org/10.1109/TII.2021.3086798

Gerstner W, Kistler WM (2002) Spiking neuron models: single neurons, populations, plasticity. Cambridge University Press, Cambridge

Grossman GM, Krueger AB (1991) Environmental Impacts of a North American Free Trade Agreement. WP No. 3914, National Bureau of Economic Research (NBER). https://www.nber.org/papers/w3914

Grossman GM, Krueger AB (1995) Economic growth and the environment. Q J Econ 110(2):353–377

Hadiyan PP, Moeini P, Ehsanzadeh E (2020) Application of static and dynamic artificial neural networks for forecasting inflow discharges, case study: Setifdroud Dam reservoir. Sustain Comp: Inform Syst 27:100401

Han Y, Li J, Lou X, Fan C, Geng Z (2022a) Energy saving of buildings for reducing carbon dioxide emissions using novel dendrite net integrated adaptive mean square gradient. Appl Energy 309:118409

Han Y, Li J, Lou X, Feng M, Geng Z, Chen L, Ping W, Lu G (2022b) Energy consumption analysis and saving of buildings based on static and dynamic input-output models. Energy 239(Part C):122240

Ibrahim OM (2013) A comparison of methods for assessing the relative importance of input variables in artificial neural networks. J Appl Sci Res 9(11):5692–5700

IREA, (2013). Africa’s renewable future: the path to sustainable development, International Renewable Energy Agency - IRENA
