Ingeniería y Competitividad
Vol 28 (3) 2026 doi: 10.25100/iyc.v28i3.16014

Vol 28 (3)

ISSN
0123-3033
e- 2027-8284

Original Article

Prediction of CO_2 Emissions in OECD Countries Using Neural Networks and Demographic, Economic, and Circular Economy Indicators

Predicción de emisiones de CO_2 en países de la OCDE mediante redes neuronales e indicadores demográficos, económicos y de economía circular.

Elcira Solano-Benavides[1] Nelson Alandete-Brochero[2] Edie Pertuz-Díaz[3]

[1] Ph.D. in Economics, Research Professor, Universidad del Atlántico, Barranquilla, Colombia.
[2] M.Sc. in Economics, Research Professor, Universidad del Atlántico, Barranquilla, Colombia.
[3] Economist, Research Assistant, Universidad del Atlántico, Barranquilla, Colombia.

Abstract

Introduction: Mitigating climate change requires accurate monitoring and prediction of CO2 emissions, particularly in OECD countries, where industrialization and economic growth have historically driven high emission levels.
Objective: To predict CO2 emissions in OECD countries using a neural network approach that incorporates demographic and economic indicators, circular economy legislation, and plastics trade data.
Methods: A quantitative, longitudinal and predictive design was applied to a panel of 32 OECD countries (1960–2023; 2,048 country-year observations). GRU networks with one-, three- and five-year windows were optimized with Optuna, trained with a temporal split (training up to 2018, validation 2019–2021, test 2022–2023), and complemented by post-hoc calibration and Kernel SHAP. Results were compared with persistence, fixed-effects panel, Random Forest and LSTM benchmarks and checked across five random seeds.
Results: Optimization and calibration raised the test R^2 from 0.053 for the non-optimized network to 0.968 for the calibrated three-step model (MAE = 83.7 MtCO2). This model was the most stable across seeds (MAE = 86.7 ± 4.3 MtCO2), and window length did not yield consistent differences. Simple benchmarks, however, achieved lower errors: persistence (MAE = 11.0) and Random Forest (MAE = 45.8). Plastics trade, likely acting as a proxy for economic size, and labor force were the most influential variables, whereas circular economy laws had the weakest contribution.
Conclusions: Calibrated GRU networks provide an interpretable framework that links CO2 emissions to structural demographic, economic, trade and regulatory indicators without using past emissions as inputs. However, they do not outperform simple benchmarks for short-term forecasting. Their value lies in explanation and scenario analysis rather than in point prediction.

Keywords: CO2 emissions; Circular economy; GRU neural networks; OECD; Predictive models.

Resumen

Introducción: Mitigar el cambio climático requiere del monitoreo y la predicción precisa de las emisiones de CO2, en particular en los países de la OCDE, donde la industrialización y el crecimiento económico han impulsado históricamente altos niveles de emisiones.
Objetivo: Predecir las emisiones de CO2 en países de la OCDE mediante un enfoque de redes neuronales que incorpora indicadores demográficos y económicos, legislación de economía circular y datos de comercio de plásticos.
Métodos: Se aplicó un diseño cuantitativo, longitudinal y predictivo a un panel de 32 países de la OCDE (1960–2023; 2.048 observaciones país-año). Las redes GRU con ventanas de uno, tres y cinco años se optimizaron con Optuna y se entrenaron con una partición temporal (entrenamiento hasta 2018, validación 2019–2021 y prueba 2022–2023). Se complementaron con calibración post-procesamiento y análisis Kernel SHAP. Los resultados se compararon con modelos de persistencia, panel de efectos fijos, Random Forest y LSTM, y se verificaron con cinco semillas aleatorias.
Resultados: La optimización y la calibración elevaron el R^2 de prueba de 0,053 en la red sin optimizar a 0,968 en el modelo calibrado de tres pasos (MAE = 83,7 MtCO2). Este modelo fue el más estable entre semillas (MAE = 86,7 ± 4,3 MtCO2), y la longitud de la ventana no produjo diferencias consistentes. Sin embargo, modelos de referencia simples lograron errores menores: persistencia (MAE = 11,0) y Random Forest (MAE = 45,8). El comercio de plásticos, que probablemente actúa como indicador del tamaño de la economía, y la fuerza laboral fueron las variables más influyentes, mientras que las leyes de economía circular tuvieron la contribución más débil.
Conclusiones: Las redes GRU optimizadas mediante hiperparámetros, combinadas con calibración post-procesamiento e interpretabilidad basada en SHAP, constituyen un marco eficaz y transparente para predecir las emisiones de CO2 en países de la OCDE, apoyando una planificación de políticas ambientales más realista y basada en evidencia.

Palabras clave: OEmisiones de CO2; Economía circular; Modelos predictivos; OCDE; Redes neuronales GRU

How to cite?

Solano-Benavidez E, Alandete-Brochero N, Petruz-Díaz E. Prediction of CO□ Emissions in OECD Countries Using Neural Networks and Demographic, Economic, and Circular Economy Indicators. Ingeniería y Competitividad, 2026, 28(3)e-20516014

https://doi.org/10.25100/iyc.v28i3.16014

Received: 14/07/26
Reviewed: 8/10/26
Accepted: 9/11/26
Online: 110/11/26

Correspondence
nalandete@mail.uniatlantico.edu.co

Spanish version

CrossMark

CC BY NC SA

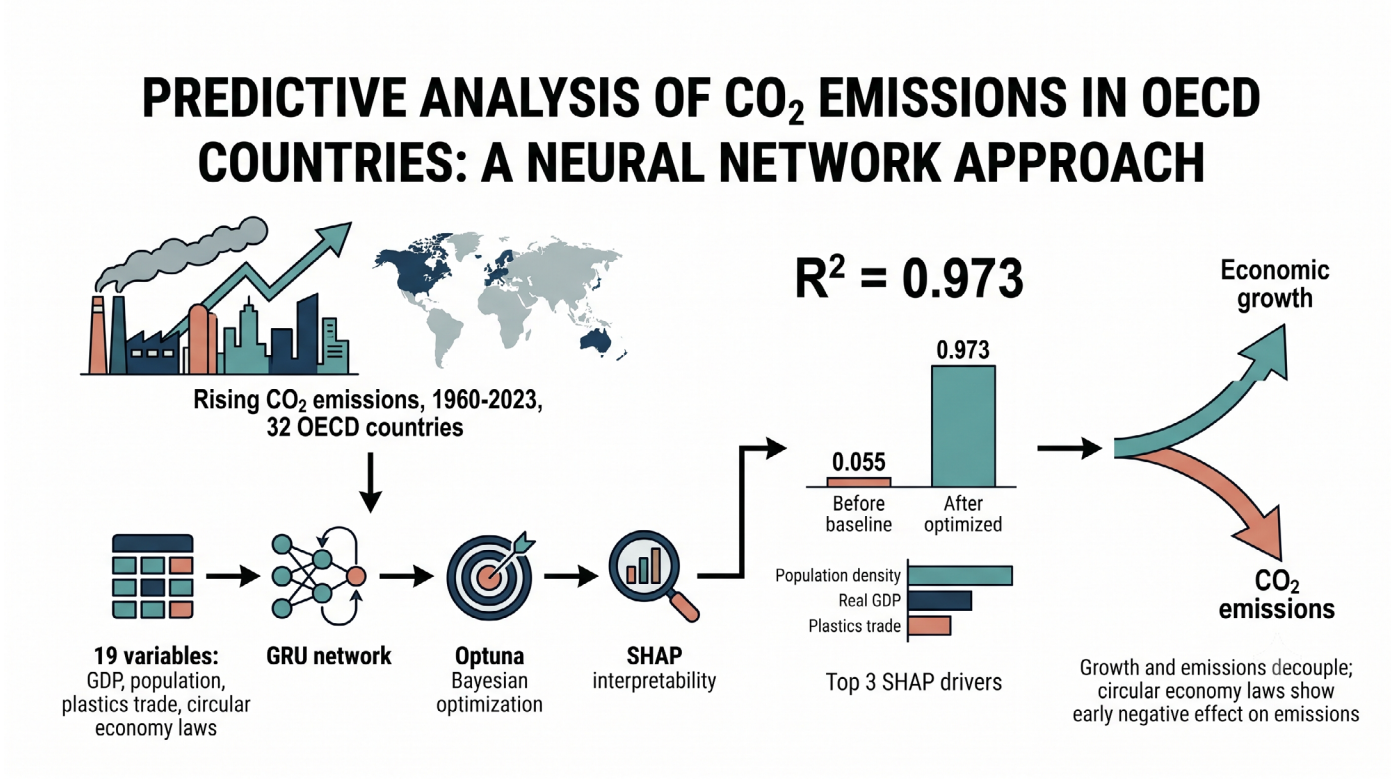
OPEN ACCESS

Why was the study conducted?
OECD countries account for a large share of global CO2 emissions, and most predictive models rely only on energy variables or on past emissions. This study assessed whether Gated Recurrent Unit (GRU) neural networks can predict emissions from demographic, economic, plastics trade and circular economy legislation indicators, and identified which of these factors most influence the predictions.

What were the key findings?
Hyperparameter optimization and calibration raised the R^2 in the test period (2022–2023) from 0.053 to 0.968. Longer time windows did not systematically improve the results. Simple benchmarks, such as persistence and Random Forest, achieved lower errors in short-term prediction. Plastics trade and the labor force were the most influential variables, while circular economy laws do not yet show a consistent effect on emissions.

What do these findings contribute?
The study provides an interpretable framework that links emissions to structural and regulatory factors without relying on past emissions. It also shows why complex models should be compared with simple benchmarks before claiming superiority. For environmental policy, the findings suggest that mitigation strategies should consider each country's structural context, and that the effect of circular economy laws needs a longer implementation period before it can be evaluated.

Graphical Abstract



doi: 10.25100/iyc.v28i3.16014

Introduction

Mitigating climate change is one of the most urgent challenges of the modern era (1), and accurate monitoring of carbon dioxide (CO2) emissions is critical to designing effective public policies. The historical evolution of emissions shows a complex structural dynamic, with a sustained upward trend for decades, followed by an inflection and recent volatility exacerbated by global events such as the COVID-19 pandemic (2).

In OECD (Organisation for Economic Co-operation and Development) countries, where industrialization and economic growth have historically driven high emission levels, understanding and predicting these trends is essential (3). Such analysis not only contributes to mitigating environmental impact but also supports the formulation of sustainable economic and trade policies, particularly in sectors such as the plastics trade, which plays an increasing role in the global economy.

In the literature, emission prediction has been based on statistical methods such as integrated moving average autoregressive models (ARIMA), but these approaches are often limited by assumptions of linearity (4). In this context, neural networks and deep learning manage to capture complex non-linear relationships directly from the data (5,6,7).

Despite efforts to reduce CO2 emissions, conventional statistical models often fail to capture the non-linear dynamics and complex interactions between demographic, economic, and trade variables in OECD countries (8). This limitation hinders effective environmental and economic policy planning, particularly where the plastics trade and industrialization remain major drivers of emissions (9).

In addition, recorded data suffer from spatial biases, delays in reporting, and inconsistent data, with official publications often delayed by one or two years, making them unsuitable for real-time decisions (10,11). Many official emission data sets are annual, resulting in small samples that can lead to overfitting in deep learning models.

Globally, population size, age structure and affluence are recognized drivers of aggregate CO2 emissions, particularly in developed countries (12). Plastics are also relevant, since their life cycle accounts for a substantial and growing share of global greenhouse gas emissions (13). Machine learning models that combine economic, energy and green-economy indicators have been used to forecast national emissions and to identify their main drivers (14); however, plastics trade has not yet been incorporated as a predictor in these models.

In the context of OECD countries, regional disparities in emissions reflect differences in industrialization, environmental policies, and demographic structures (15,16). These variations require an adapted approach that considers local characteristics, such as the implementation of circular economy laws and trends in the plastics trade, to develop relevant and applicable predictive models, taking advantage of the abundance of data to advance climate modeling (6).

Therefore, the main objective of this study is to predict CO2 emissions in OECD countries with GRU neural networks that incorporate demographic, economic, circular economy legislation and plastics

doi: 10.25100/iyc.v28i3.16014

trade indicators, and to identify the variables that drive the predictions. The design is quantitative, longitudinal and predictive. Data were obtained from the World Bank, the OECD, the International Labour Organization and the United Nations. The sample comprises 32 of the 38 OECD member countries over 1960–2023.

Although statistical and machine learning models have been widely used to analyze CO2 emissions, the integration of GRU recurrent neural networks into long multidimensional time series (1960–2023) remains uncommon (5,7). This study compares GRU architectures with one, three and five-year input windows and, for the first time, incorporates circular economy legislation together with demographic, economic and plastics trade indicators. In addition, the models are compared against persistence, panel and machine learning benchmarks, and their predictions are interpreted with SHAP to estimate the contribution of each structural and regulatory variable (6,17).

State of the Art

Various studies on the prediction of CO2 emissions using machine learning models and neural networks, as observed in Table 1, reveal consistent patterns, including the prevalence of techniques such as Random Forest, Support Vector Machines and Long Short-Term Memory networks to handle non-linearities and achieve high accuracy, with R^2 or accuracy values above 0.93 (18,19,20,21).

Common input variables in these models include GDP, population, fossil fuel and renewable energy consumption, and industrial activity, which are frequently linked to economic growth and the impacts of transport on emissions (22,23,24,25,26,27,28).

Key forecast patterns indicate potential reductions in CO2 emissions through the adoption of renewable energies and policy interventions. However, recurring limitations involve data constraints, such as small sample sizes, inconsistencies in reports and dependence on univariate or limited factors, along with model challenges such as black-box interpretability, computational demands and overfitting risks, which require broader datasets and comparative validations for greater accuracy (18,19,23,24,25,29).

doi: 10.25100/iyc.v28i3.16014

Table 1. Research on the prediction of CO2 emissions

Author(s) and Year	Country/Region	Model type	Key Input Variables	Main Finding / Forecast	Key Limitation Mentioned
Li et al. (18)	Top 30 CO2 emitting countries	Support Vector Machine, Gradient Boosting Regressor	GDP, population, unemployment, sectoral added value, air passengers, vehicles, etc.	A GBR model combining transport and economic features predicts transport-related emissions with R^2=0.9943, underscoring the role of transport in climate change mitigation.	Lower performance in Tier 2 countries, requiring additional factors for environmental accuracy.
Adebowale & Uzel (22)	Sweden	Bayesian Neural Network	GDP, renewable consumption, oil consumption	CO2 emissions drop to 30M tons in 2025, emphasizing the role of renewables in environmental reduction.	Uncertainty due to input forecasts; data limited to fossil fuels.
Anonna et al. (19)	United States	Random Forest	Energy consumption, industrial activity, transport	Random Forest with 0.939 accuracy supports sustainable policies aligned with SDGs to mitigate climate.	Inconsistencies in reports; complexities in data integration.
Ma et al. (23)	China	Gaussian Process Regression	GDP, energy consumption, population, industrialization, income	More accurate Gaussian Process Regression, supporting emission reduction strategies in China.	Limited to five factors; needs comparison with other models.
Ene Yalçın (24)	Turkey	Multivariate regression, random forest, XGBoost	Year, population, GDP per capita, energy consumption, renewables	Maximum likelihood machine learning forecasts greenhouse gases with high precision; scenarios up to 2030 show reductions with renewables.	Small dataset limits machine learning methods.

doi: 10.25100/iyc.v28i3.16014

Tian et al. (20)	United States	Decision tree, Random Forest, Multiple linear regression, K-nearest neighbors, Support Vector Regression, Gradient Boosting	Energy consumption by source (coal, natural gas, petroleum, electricity)	Multiple linear regression achieved the highest accuracy (R^2 = 0.99 on test data); Random Forest and Gradient Boosting were most sensitive to coal use.	Decision tree overfitting; lower accuracy of SVR and KNN.
Nie et al. (25)	Beijing, China	Feedforward network with gray model	Carbon emissions, coal/oil consumption, GDP, population	Simulation and prediction errors below 10%; forecasts Beijing's emissions for 2019–2023.	Short series (2009–2018), with results sensitive to the length of the modeling data.
Feng Li et al. (29)	China	CNN-LSTM	GDP, population, energy structure	Decrease in 2023-2033 scenarios, supporting national reduction goals.	Small sample size.
Han et al. (30)	30 Chinese provinces	CNN-LSTM	Population, GDP per capita, energy intensity	Peak majority before 2030, facilitating carbon neutrality.	Difficulty of interpretation, subjective rates.
Jena et al. (26)	17 key CO2-emitting countries	Multilayer Artificial Neural Network (MLANN)	GDP, urban population, trade liberalization	96% accuracy, varied trends 2017-2019, supporting global reductions.	The MLANN model predicts CO2 emissions fairly accurately; however, there are cases where the prediction error was high.
Mustafa et al. (31)	Asia	Generalized Regression Neural Network (GRNN)	CO2 anomalies, O2 satellite monitoring system, primary productivity network (PPN)	Environmental satellite monitoring helps predict the concentration of anthropogenic carbon dioxide.	Greater prediction error in the face of uncertainty and satellite anomalies.

doi: 10.25100/iyc.v28i3.16014

Moayedi et al. (21)	Central Europe	Multilayer Perceptron (MLP) with optimizers	GDP, energy sources	Complex evolution shuffled-MLP best R^2=0.9999, forecasting emissions for green policies.	Data restrictions, model complexity, variable selection, external factors, algorithm sensitivity, validation and generalization.
Ajala et al. (32)	OECD	ARMA, ARIMA, SARMA, SARIMA, SVM, RF, GB, ANN, RNN variations (GRU, LSTM, BILSTM, CNN-RNN)	Daily CO2 data	Machine learning and deep learning models outperform statistics, increasing winter emissions, guiding policies.	Univariate data only.
Keerthana et al. (33)	United States	Multiple linear regression	Emissions, population, GDP	Decreases in most states through 2050.	Focused only on the states with the highest emissions.
Ahmadi et al. (27)	Latin America	Least squares support vector machine (LSSVM) and evolutionary methods	Oil, gas, coal, renewables, GDP	The best LSSVM model highlights renewable investments as a source of emission reductions.	The population was seven countries.
Chauhan (28)	Global	Multilayer Perceptron (MLP), Random Forest Regressor (RFR), Multiple Linear Regression (MLR)	Country, crop, year, population	The best model was RFR for CO2 per crop, key emitter profiles.	Scarce related literature and no baseline; FAOSTAT emission data of unknown accuracy (IPCC Tier 1).
Fatima et al. (34)	OECD	GMM quantile dynamic panel model	Globalization, institutional quality, economic growth, electricity consumption, renewable energy	Simultaneous quantile regression and panel quantile regression reveal that renewable energy consumption, institutional quality, and globalization help reduce CO2 emissions, while economic growth and energy consumption are responsible for climate and environmental loss.	Explanatory panel model (1991–2018), not designed for forecasting.

doi: 10.25100/iyc.v28i3.16014

Methodology

To achieve this objective, a quantitative, longitudinal and predictive research design was used, combining a descriptive analysis of the CO2 emissions variable with the estimation of a neural network model. The study analyzes panel data (time series for multiple countries) covering a period of 64 years, from 1960 to 2023, for 32 countries (Table 2), generating 2,048 observations. The study population comprised 32 of the 38 OECD member countries; Costa Rica, Estonia, Korea, Slovakia, Slovenia and Türkiye were excluded because of the lack of complete data series for the study period.

Table 2. Countries in the study population

Country	Region
Australia	Oceania
Austria	Europe
Belgium	Europe
Canada	North America
Chile	South America
Colombia	South America
Denmark	Europe
Finland	Europe
France	Europe
Germany	Europe
Greece	Europe
Hungary	Europe
Iceland	Europe
Ireland	Europe
Israel	Asia
Italy	Europe
Japan	Asia
Latvia	Europe
Lithuania	Europe
Luxembourg	Europe
Mexico	North America
Netherlands	Europe
New Zealand	Oceania
Norway	Europe
Poland	Europe
Portugal	Europe
Spain	Europe
Sweden	Europe
Switzerland	Europe
United Kingdom	Europe
United States	North America
Czech Republic	Europe

Data were collected from governmental and non-governmental sources, including the World Bank, the OECD Data Explorer, the International Labor Organization (ILO), and the United Nations (UN), ensuring comprehensive coverage of economic, geographic, demographic, environmental, and legal variables. A total of 19 variables (Table 3) were used, with CO2 emissions as output and the remaining 18 as inputs, including numerical variables such as historical real GDP, manufacturing

doi: 10.25100/iyc.v28i3.16014

GDP, historical exports and imports of plastics, unemployment rate, exchange rate, productivity, labor force, historical mining GDP, population metrics (total, growth, age groups 0-14, 15-64, +65), Gini index, GDP per capita and population density, plus a dichotomous variable for circular economy laws, coded as 1 in the year in which a country enacted a law supporting the circular economy and 0 otherwise (33,34).

The dataset covered annual observations from 1960 to 2023 for 32 countries (Table 2), generating 2,048 observations. Primary sampling was not required, as the data were secondary and comprehensive for the study population; however, imputation was planned for missing values (NA) identified in variables such as unemployment rate (589 NA; 28.8%), productivity (1,440 NA; 70.3%, available only from 2005), Gini index (1,262 NA; 61.6%) and exports and imports of plastics (1,408 and 1,428 NA; 68.8% and 69.7%, available only from 2004), among others, addressing common limitations in environmental datasets such as inconsistencies and delays (10,11). This method was chosen for its reliability in providing high-quality historical panel data suitable for learning applications.

Table 3. Statistical summary of the variables used

Variable	Unit		NA (%)	Coverage		Std. Dev.			Max
Emissions	MtCO2	WB	0.0	1960-2023	339	876	1	66	6,132
Real GDP	Billion constant 2010 US$	WB	6.7	1960-2023	757.5	2,173.5	0.3	157.1	
Manufacturing GDP	Billion US$	WB	40.5	1960-2023	103.3	173.6	1.0	38.8	2,123.2
Plastics Exports	Million US$	WITS	640	68.8	2004-2023		15,172.4		
Plastics Imports	Million US$	WITS	620	69.7	2004-2023		9,346.7		
Unemployment Rate	% of labor force	WB	28.8	1960-2023	7.2	4.2	0.2	6.4	27.7
Exchange Rate	LCU per US$	WB	5.1	1960-2023	92.86	355.32		3.21	
Labor Productivity	Thous. 2010 US$ per worker	WB	608	70.3	2005-2023	64.6	29.6	63.9	196.6
Econ. Active Pop.	Million persons	WB (ILO)	46.9	1990-2023	17.20	28.84		5.12	171.33
Mining GDP	% of GDP	WB	25.9	1970-2021	0.41	1.52		0.01	16.87
Population	Million persons	WB	0.0	1960-2023	31.50	50.86		10.30	334.91
Population Growth	Annual %	WB	1.6	1961-2023	0.75	0.82		0.66	6.02
Pop. 0-14 years	Million persons	WB	0.0	1960-2023	6.94	11.37		2.18	61.60
Pop. 15-64 years	Million persons	WB	0.0	1960-2023	20.45	33.39		6.78	217.61

doi: 10.25100/iyc.v28i3.16014

Variable	Unit			NA (%)	Coverage		Std. Dev.			Max
Pop. +65 years	Million persons	WB		0.0	1960-2023	4.11	7.16		1.38	58.38
Circular Econ. Laws	Binary (0/1)			0.0	1960-2023	0.018	0.133		0.000	1.000
GINI	Index (0-100)	WB	786	61.6	1963-2022	33.7	6.7		32.8	58.5
GDP per capita	Constant US$	WB		6.7	1960-2023		22,140			
Population Density	People per km²	WB		11.6	1961-2022	111.2	112.2	1.4	85.7	525.7

Note: Monetary and population variables are rescaled for readability. Sources: WB = World Bank, World Development Indicators; OECD = OECD Data Explorer; WITS = World Integrated Trade Solution (UN Comtrade, HS chapter 39); OWID = Our World in Data; Authors = authors' compilation from national legislation. LCU = local currency units.

Procedure

The study adopted a structured process in six sequential phases to develop a robust, calibrated and interpretable predictive model for CO2 emissions. The methodological approach integrated advanced deep learning techniques, Bayesian optimization, post-hoc calibration, and interpretability analysis, informed by a review of the CO2 emissions literature (18,20,21,32).

Data were obtained from international sources such as the World Bank, the OECD, the International Labor Organization, and the United Nations, and compiled into a unified set covering the period 1960-2023. The unit of analysis corresponded to country-year observations, configuring an unbalanced panel, since the availability of the input variables differed between countries and years (Table 3), that allowed for the capture of temporal dynamics and heterogeneity between countries.

An exploratory analysis was carried out to characterize the statistical properties of the data set through graphical analysis of histograms, scatter plots and time series in order to visualize growth patterns, non-linear trends and possible points of structural change in CO2 emissions.

A total of 18 predictor variables were evaluated that captured economic, demographic, commercial, and environmental policy dimensions, along with the target variable of emissions in millions of tons of CO2. Economic variables included historical real GDP, historical manufacturing GDP, mining GDP, GDP per capita, productivity, and exchange rate. Trade variables included exports and imports of plastics as a proxy for carbon-intensive industrial activity. Demographic variables covered total population, population growth, age distribution in three groups (0-14 years, 15-64 years and over 65 years) and population density. Labor variables included the unemployment rate and economically active population. Policy and inequality variables included binary-coded circular economy laws and the Gini index as a measure of economic inequality. In addition, a numerical country identifier was included as an input to the neural networks so that the model could distinguish between countries, resulting in 19 input features.

doi: 10.25100/iyc.v28i3.16014

To ensure the validity of subsequent analyses and reduce systematic biases, the procedure followed the recommendations of Ajala et al. (32), who implemented a preprocessing step based on the recommendations for heterogeneous environmental data. A strict time division strategy was adopted to preserve the integrity of the time series and avoid information leakage, defining the training and calibration set for the period 1960-2021 and reserving the years 2022-2023 as a test set. This partition ensured that the model did not access future information during training, simulating realistic forward-looking prediction conditions.

The imputation of missing values was performed using the SimpleImputer algorithm (35), using the median as the imputation statistic. This choice was based on the robustness of the median in the face of extreme values when compared to the mean, its ability to preserve the central properties of asymmetric variables, and the numerical stability it provides by avoiding biases in variables with heavy tails (36). The imputer was fitted exclusively on the 1960-2021 data, using the global median of each variable across all countries, and then applied to the complete dataset, so that the test years did not inform the imputation. Since the exports and imports of plastics and productivity are only available from 2004-2005 onwards, their imputed values remain constant in earlier years; this limitation is addressed in the Discussion.

Robust Scaler was applied to both the predictor variables and the target variable, a method that uses the median and the interquartile range according to the formula $X_{scaled} = (X - \text{median}(X)) / IQR(X)$. This scaler offers significant advantages, including resistance to multivariate outliers, preservation of non-Gaussian distributions, gradient stabilization during neural network training, and suitability for economic variables with heteroscedasticity (18,36). The scaling was performed independently for each variable, adjusting parameters on the 1960-2021 data and applying them consistently to the test set.

In this paper, the Gated Recurrent Units (GRU) network is used as the base architecture due to its demonstrated ability to capture long-term temporal dependencies with lower computational complexity than Long Short-Term Memory networks (37). GRUs use gate mechanisms to regulate the flow of information, specifically an update gate that controls how much information from the previous state is retained and a reset gate that determines which past information is relevant to the calculation of the new state. This architecture has demonstrated efficacy in environmental time series with multiple time scales (30). To capture temporal patterns at different scales, three sliding window configurations were implemented.

The one-step model used the variables for the year t to predict the emissions for the year t, capturing immediate contemporary relationships. The three-step model used the sequence of variables for the years t-2, t-1 and t as input to predict the emissions for the year t, capturing short-term trends and recent momentum. The five-step model used the sequence of variables for the years t-4, t-3, t-2, t-1 and t as inputs to predict the emissions for the year t, capturing medium-term patterns and economic cycles.

Chronologically ordered time sequences were generated for each country, ensuring that data from different countries were not mixed in the same sequence. During hyperparameter optimization, 20% of the sequences of the period 1960-2021 were randomly extracted for validation, using

doi: 10.25100/iyc.v28i3.16014

a fixed seed for reproducibility. For the training of the final models, a temporal partition was adopted: the sequences whose target year fell between 1960 and 2018 formed the training set, those whose target year fell between 2019 and 2021 formed the validation set, used for early stopping and for estimating the calibration factors, and the sequences of 2022-2023 formed the test set, which was not used in the optimization, training or calibration. The final training used a fixed seed (42) with deterministic operations, and the three configurations were retrained with four additional seeds (7, 123, 2024 and 99) to assess the sensitivity of the results to initialization.

Hyperparameter optimization was performed using Optuna (38), a Bayesian optimization framework that uses the Tree-structured Parzen Estimator (TPE) algorithm. Unlike exhaustive or random searches, this algorithm builds a probabilistic model of the objective function that is iteratively updated, allowing for efficient exploration of the hyperparameter space (39).

Results and Discussion

The data were collected on an annual basis, with a historical record from 1960 to 2023, covering 64 years of data for the 19 variables for the 32 countries in the study population, thus processing a total of 2,048 observations in the research. The target or dependent variable was the CO2 emissions for each country measured in millions of tons. The remaining variables were used as inputs for the design of an artificial neural network whose components are explained below.

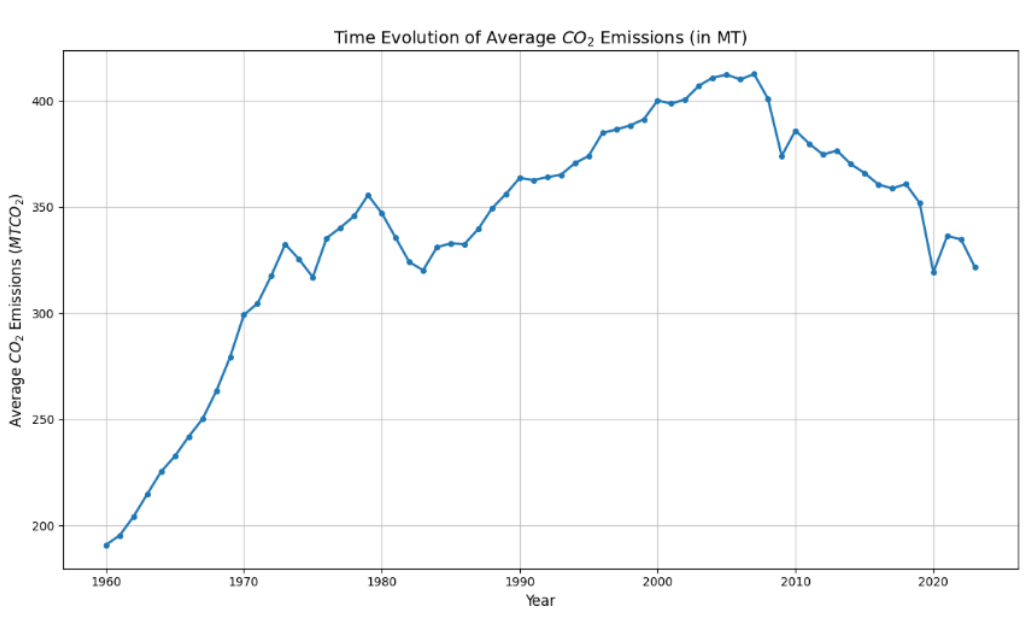
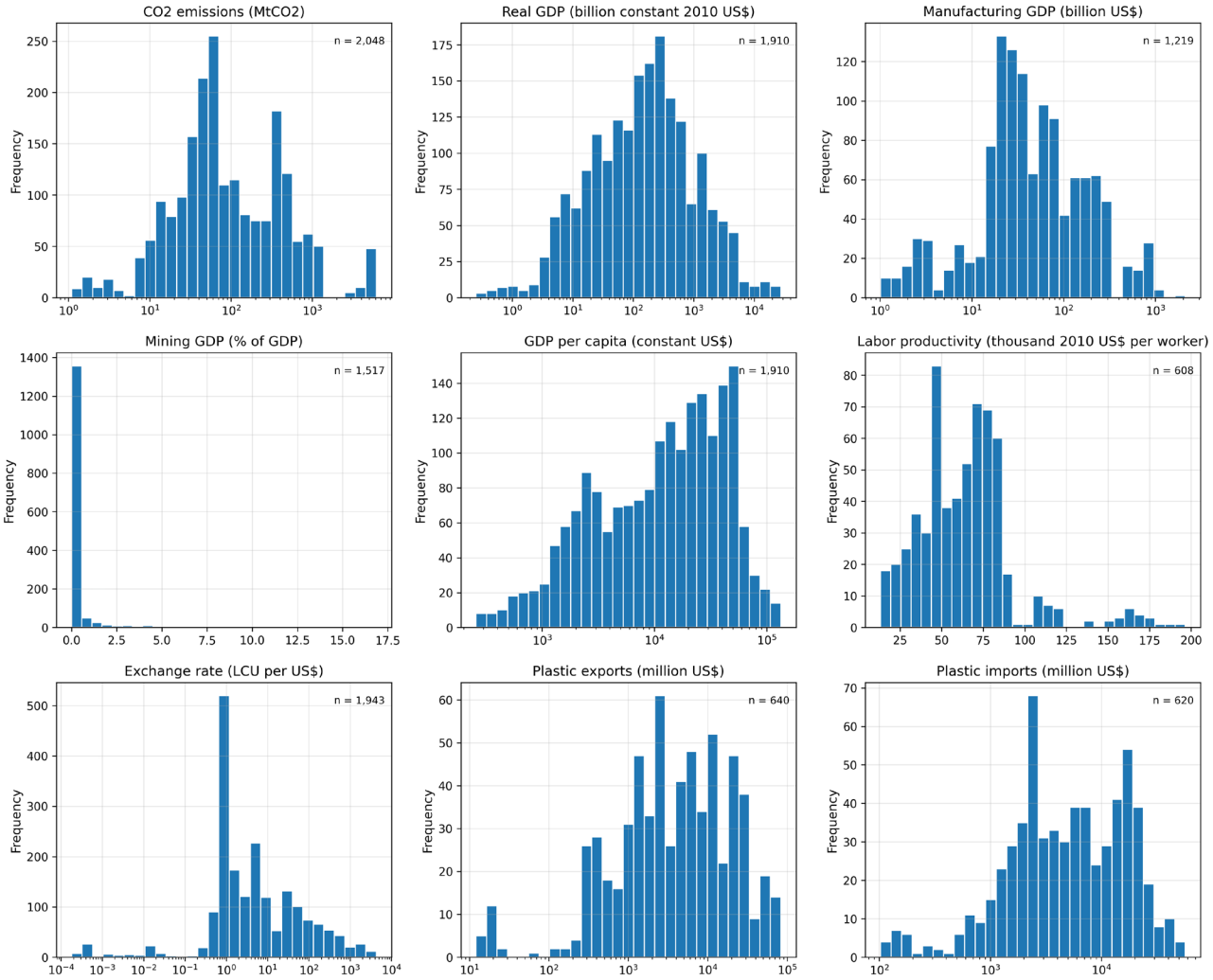


Figure 1. Historical evolution of CO2 emissions in millions of tons.

The average CO2 emissions (1960-2023) show an upward trend until the mid-2000s, attributed to industrialization and greater use of fossil energy. Between 2005 and 2010, a peak is reached, followed by a moderate decline due to stricter environmental policies, technological advances and programs such as EU Directive 2009/29/EC and US energy efficiency laws. In 2020, emissions fell sharply due to the pandemic, without fully recovering afterwards, suggesting a structural shift towards lower carbon intensity.

doi: 10.25100/iyc.v28i3.16014

(a) Emissions, economic and trade variables



doi: 10.25100/iyc.v28i3.16014

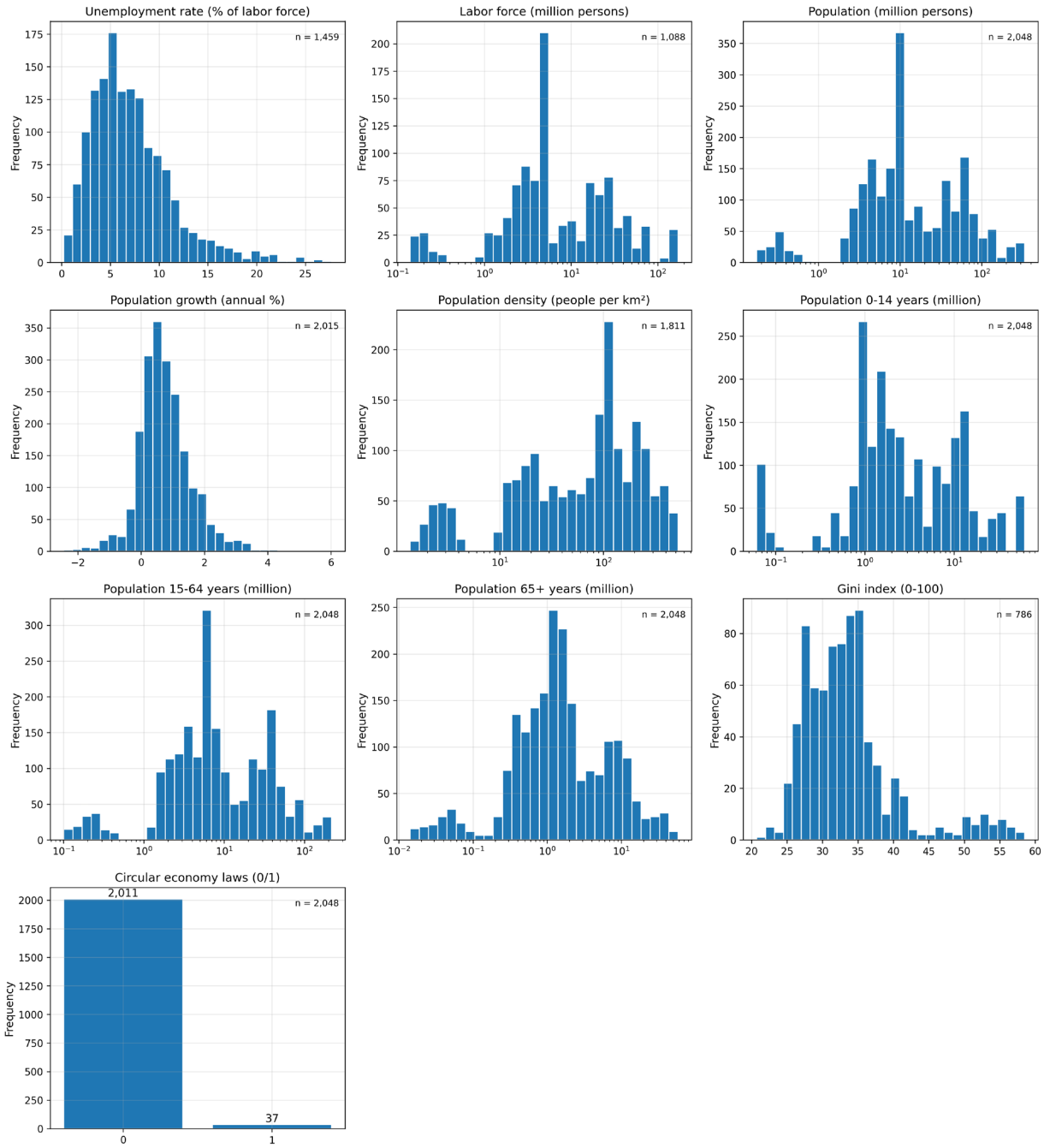
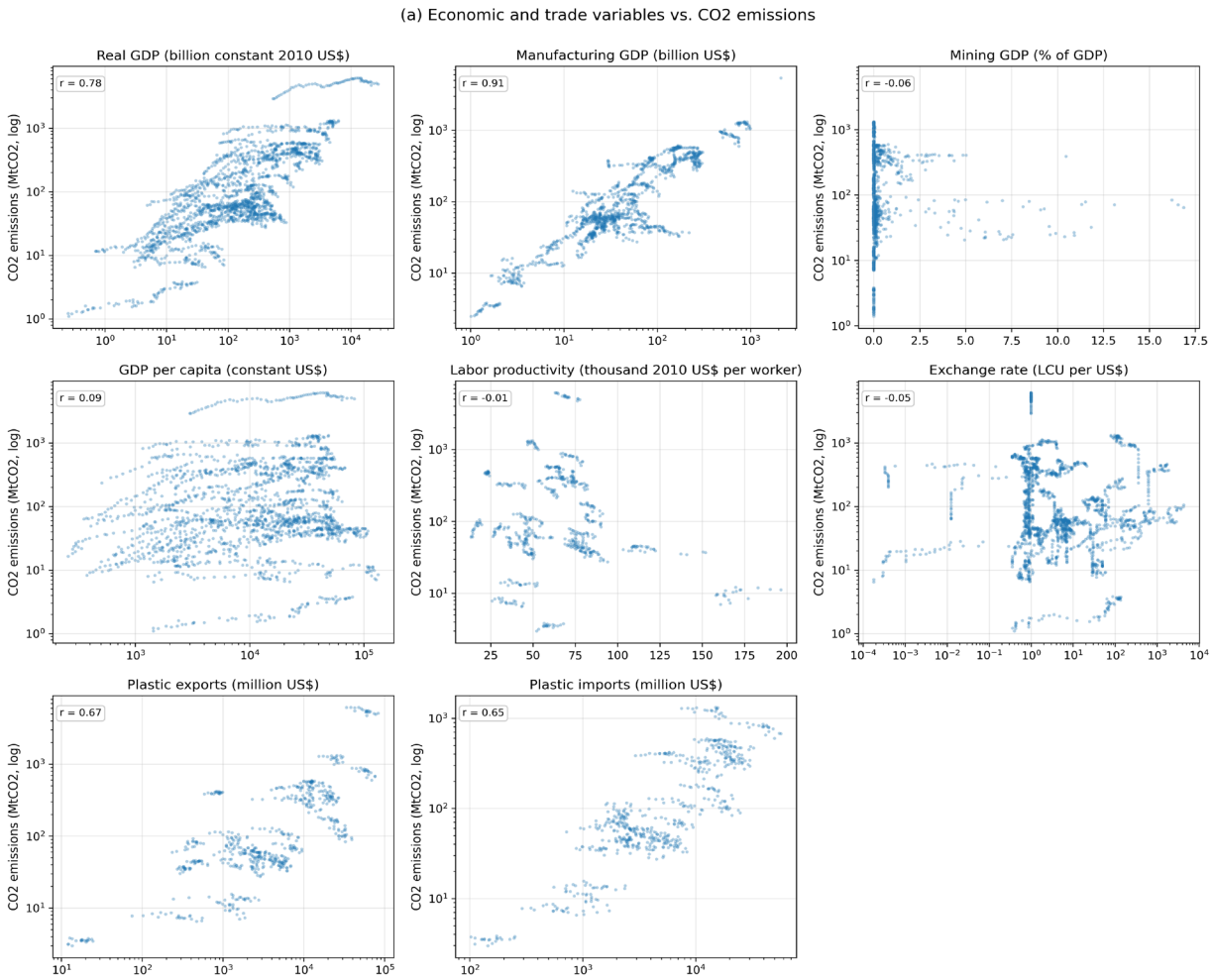


Figure 2a. Histograms of emissions, economic and trade variables. 2b. Histograms of labor, demographic, inequality and policy variables

The histograms of the variables (Figures 2a and 2b) reveal a dominance of positive asymmetry or right bias in the vast majority of the variables, including the target variable CO2 emissions and the main economic (GDP, exports, GDP per capita) and demographic (population) predictors. This distribution, characterized by a high concentration of data at low values and a long tail of high outliers, is a consistent pattern in the time series for OECD countries. For this reason, the most skewed variables are displayed on a logarithmic scale. In contrast, the Gini index and population growth show more centered and symmetrical distributions. Finally, the circular economy laws variable, which takes the value 1 in the year in which a country enacted a law supporting the

doi: 10.25100/iyc.v28i3.16014

circular economy and 0 otherwise, is heavily skewed toward zero: of the 32 countries in the study population, 25 enacted at least one law, with a combined total of 37 enactments over the entire study period.



doi: 10.25100/iyc.v28i3.16014

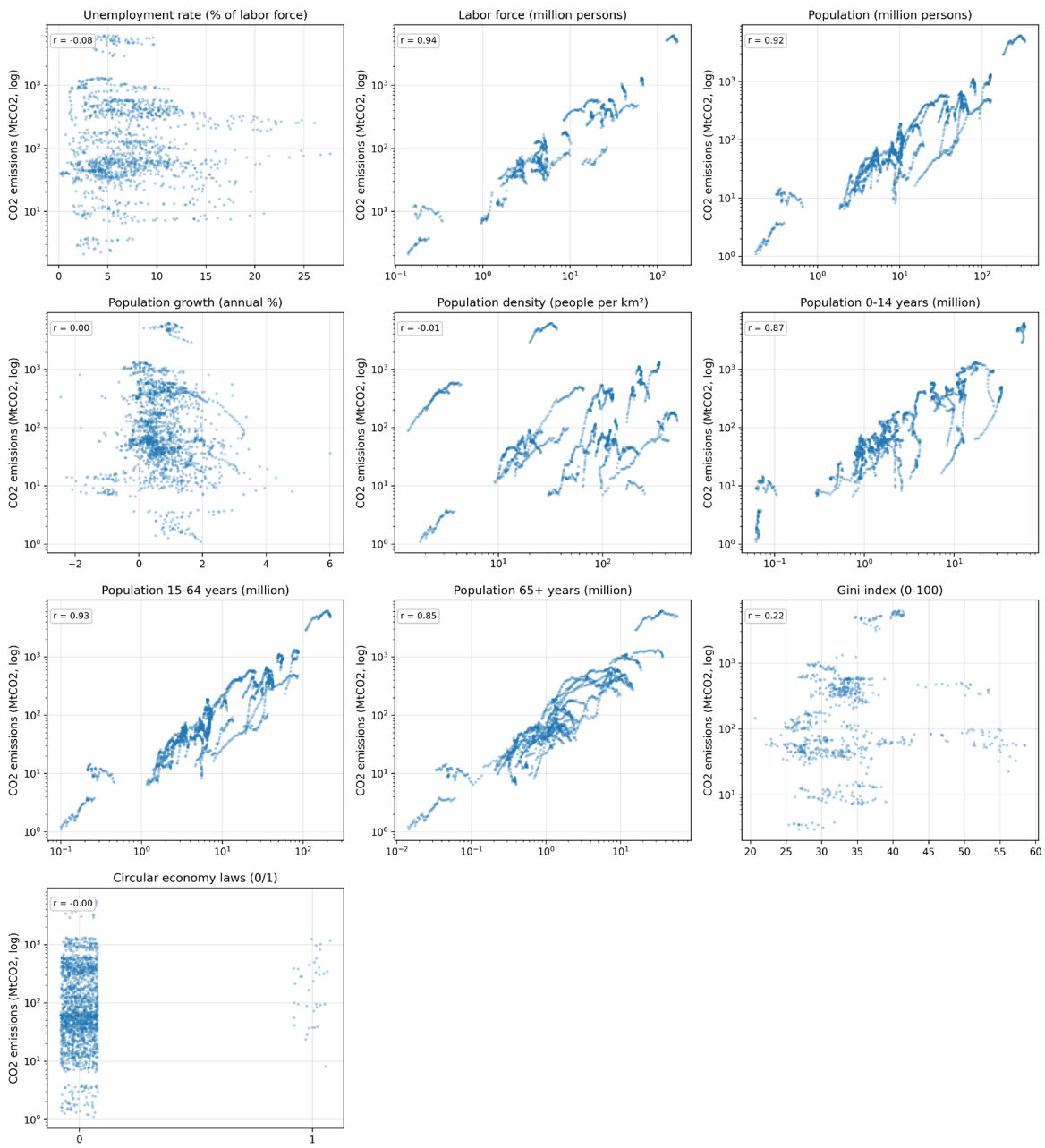


Figure 3a. Scatter plots between CO2 emissions and economic and trade variables. 3b. Scatter plots between CO2 emissions and labor, demographic, inequality and policy variables.

Figures 3a and 3b show the relationship between CO2 emissions and each predictor, with emissions on a logarithmic scale and the Pearson correlation coefficient (r) in each panel. Strong positive relationships are observed with the variables associated with the size of the economy and the population, such as labor force (r = 0.94), working-age population (r = 0.93), total population (r = 0.92), manufacturing GDP (r = 0.91) and Real GDP (r = 0.78), as well as with plastic exports and imports (r = 0.67 and 0.65). In contrast, GDP per capita, productivity, unemployment rate, exchange rate, mining GDP, population density, population growth and circular economy laws show correlations close to zero (|r| < 0.10), and the Gini index shows a weak positive relationship (r = 0.22). Considerable heterogeneity across countries produces substantial dispersion around these

doi: 10.25100/iyc.v28i3.16014

trends, which suggests that country-specific structural and unobserved factors moderate these relationships in ways that are not captured by linear associations alone.

GRU Model Architecture

An architecture based on Gated Recurrent Units (GRU) was used to model the time sequence present in the data. GRUs are a type of recurrent neural network (RNN) that, unlike traditional RNNs, incorporate gates that allow the flow of information to be controlled, facilitating the learning of long-term dependencies and mitigating the problem of gradient fading.

In particular, GRUs have a simpler and more efficient structure than LSTMs (Long Short-Term Memory), since they have only two gates (reset and update), which reduces the number of parameters and the computational cost, maintaining a comparable or superior performance in many tasks. For these reasons, the GRU architecture is considered particularly suitable for sequential problems with noisy data or moderately long time series.

To optimize the model's performance, Optuna was used, a hyperparameter optimization framework based on Bayesian search and efficient sampling algorithms. Optuna uses Bayesian optimization to build probabilistic models of the error function based on the hyperparameters, iteratively selecting the most promising configurations that make it possible to find the minimum of the objective function, which in this case corresponds to the minimum error. In this study, this approach allowed for an automatic and directed search over the space of key hyperparameters, including the number of GRU units (hidden state dimension), the number of recurrent layers, the learning rate, the dropout rate, and the batch size.

The use of Optuna made it possible to find combinations of hyperparameters that minimize the loss in the validation set, without using the test set, avoiding manual selection and reducing the risk of overfitting. In addition, its integration with techniques such as early stopping (which stops training when the validation error stops improving to prevent overfitting) and automatic metric logging ensured efficient and reproducible exploration.

For each time window configuration (1, 3 and 5 steps), an optimization object was initialized with the Tree-structured Parzen Estimator, running 25 independent trials. In each trial, a GRU architecture was built with the proposed hyperparameters, trained with Early Stopping (patience of 10 epochs monitoring validation loss), and evaluated on the validation set. The search space comprised 2 to 4 layers, 64 to 256 units, dropout rates between 0.2 and 0.5, learning rates between 10^{-4} and 10^{-2} (logarithmic scale), batch sizes of 8, 16 or 32, and L2 regularization between 10^{-3} and 2×10^{-2} (logarithmic scale). The Tree-structured Parzen Estimator algorithm iteratively adjusted its objective function model until it completed the 25 trials. All three optimization studies consistently converged toward two-layer GRU architectures, suggesting that this depth is sufficient for the complexity of the data without incurring overfitting. The final models were then trained for up to 200 epochs with early stopping (patience of 20 epochs) on the temporal validation set.

As can be seen in Table 4, the five-step model requires significantly fewer units (66 versus 198) than the three-step model, indicating that the additional temporal information compensates for the need for greater capacity in the hidden layer. The L2 regularization is highest for the five-step

doi: 10.25100/iyc.v28i3.16014

model (1.67×10^{-3}), suggesting greater effective complexity of the model. All models converge at learning rates on the order of 10^{-4}, favoring stable convergence over speed.

Table 4. Optimal hyperparameters by model

Hyperparameter	**M1**	**M3**	**M5**
No. of layers	2	2	2
No. units	169	198	66
Dropout rate	0.495	0.300	0.374
Learning rate	5.29×10^{-4}	3.59×10^{-4}	4.54×10^{-4}
Batch size	32	16	32
L2 reg	1.33×10^{-3}	1.02×10^{-3}	1.67×10^{-3}

Table 5 presents the actual CO2 emissions (in millions of tons) for 2022 and 2023 in the 32 countries of the study population, along with the predictions and differences generated by the original model (without optimization), by the three versions optimized using Optuna with time windows of 1, 3 and 5 steps (called M1, M3 and M5, respectively), and by the calibrated five-step model. Differences are calculated as the actual value minus the predicted value, so that negative values indicate overestimation and positive values indicate underestimation. The original model collapses towards an almost constant prediction (close to 360 Mt for most countries), which produces strong overestimation in low- and medium-emission countries (such as Switzerland, Sweden or Hungary, with differences close to -320 Mt) and severe underestimation in high-emission nations (such as the United States, with +4,718.97 Mt in 2022 and +4,551.49 Mt in 2023). Optimized predictions substantially reduce this bias; for example, M5 reduces the difference in Spain from -125.24 to -1.18 Mt and in Mexico from +105.44 to -16.22 Mt in 2022, and predicts the United States with differences between -73.14 and -240.62 Mt. However, high errors persist in countries with atypical dynamics: the optimized models overestimate several small and medium European emitters, such as Sweden (-258.70 Mt with M5 in 2022), Hungary or Belgium, and underestimate Germany and Japan (+429.59 and +745.04 Mt with M5 in 2022). Calibration reduces the overestimation in small countries (Sweden: -73.27 Mt; Switzerland: -22.62 Mt in 2022), but does not correct the underestimation of Germany and Japan.

doi: 10.25100/iyc.v28i3.16014

Table 5. Actual vs. predicted emissions: original model, Optuna models and calibrated model (MtCO2)

Year 2022

Country	Real	Orig.	Diff.	M1	Diff.	M3	Diff.	M5	Diff.	M5 cal.	Diff.
Iceland	3.61	31.28	-27.66	13.37	-9.76	-4.58	8.20	20.58		4.57	-0.96
Latvia	6.62	33.41	-26.79	13.37	-6.75		19.65	20.58		30.66	
	7.28	31.74	-24.47	0.80	6.48	1.19	6.09	6.85	0.43	16.94	-9.66
Lithuania	12.86	31.77	-18.91	3.35	9.51	-7.91	20.76	5.68	7.18	15.77	-2.91
Denmark	28.43	311.57								89.86	
New Zealand	31.61	48.79	-17.18	86.39		26.10	5.51	39.01	-7.40	45.01	
	32.82	359.90								55.44	
Finland	36.37	30.90	5.47	3.35	33.02		59.57	-2.73	39.09	10.18	26.19
Ireland	36.71	113.64	-76.93			82.05		66.42		57.78	
Sweden	37.47	359.90								110.74	
Portugal	40.69	359.90						84.29		72.47	
Norway	40.81	54.94	-14.13	98.33		89.69		72.78		54.75	
Hungary	44.84	359.90								108.78	
Greece	58.08	49.88	8.20	77.06		50.01	8.07	62.61	-4.53	72.63	
Israel	60.95	31.28	29.67	17.03	43.91	4.31	56.64	3.99	56.95	13.00	47.94
Austria	61.49	359.90								94.91	
Chile	82.54	54.51	28.03			83.33	-0.79	81.32	1.21	91.34	-8.81
Belgium	89.00	359.90								167.74	
Czech Republic	95.11	359.90								219.06	
Colombia	98.29	55.47	42.82	98.33	-0.04	32.47	65.82	52.49	45.80	62.51	35.78
		359.90								224.41	
Spain		359.90							-1.18	241.61	-6.95
France		359.90	-66.40						2.39	337.50	
United Kingdom		359.90	-46.06		4.28		9.52			218.98	94.85
Poland		359.90	-44.86				2.20		58.68	302.75	12.29
Italy		359.90	-19.23		7.99		25.77			207.97	
Australia		359.89	24.48							300.05	84.31
Mexico		359.90	105.44							527.95	
Canada		359.90	190.71							297.86	
Germany		359.90	311.57							324.18	
Japan		359.90	672.79							334.04	
United States		359.90								5,198.40	

doi: 10.25100/iyc.v28i3.16014

Year 2023

Country	Real	Orig.	Diff.	M1	Diff.	M3	Diff.	M5	Diff.	M5 cal.	Diff.
Iceland	3.76	31.28		13.37	-9.61	-4.59	8.35	20.58		4.57	-0.81
Latvia	6.51	33.21		13.37	-6.86		24.46	20.58		30.66	
	6.99	31.74		0.80	6.19	1.19	5.81	-8.24	15.23	12.70	-5.71
Lithuania	12.46	32.23		0.80	11.66		28.32	2.75	9.71	12.83	-0.37
Denmark	27.26	84.24								89.86	
New Zealand	30.12	41.90		70.22		43.57		39.01	-8.89	45.01	
Finland	31.62	31.74	-0.12	9.60	22.02		73.91		48.22	7.63	23.99
	32.74									55.44	
Ireland	34.43	54.51						66.79		57.78	
Sweden	36.56									110.74	
Portugal	37.23			0.77	36.46					83.57	
Norway	38.90	61.92		91.47				62.21		54.75	
Hungary	40.19									108.78	
Greece	55.20	43.11	12.08	72.06		28.82	26.37	61.76	-6.56	71.78	
Austria	58.58									94.91	
Israel	59.49	32.35	27.14	0.80	58.69	1.71	57.78	15.36	44.13	16.26	43.23
Chile	77.29	58.21	19.08	88.58				58.06	19.22	68.51	8.78
Belgium	83.37									167.74	
Czech Republic	85.62									219.42	
		41.90	63.34	70.22	35.02	65.36	39.88	54.22	51.02	64.24	41.00
										224.41	
Spain									-3.47	230.86	-9.24
France										337.48	
Poland									55.46	239.60	49.68
United Kingdom					-4.41		-6.53			209.88	95.27
Italy					11.11		25.82		99.83	219.40	94.06
Australia			28.37		39.79					308.51	74.46
Mexico									2.84	526.18	
Canada										253.92	
Germany										324.18	
Japan										342.57	
United States										5,198.40	

Table 6 shows the performance metrics in the test set (2022-2023) for the original model, the optimized models before and after calibration, and the benchmarks. Optimization raised the R^2 from 0.053 for the original model to between 0.946 and 0.953 for the optimized models, and calibration increased it to between 0.968 and 0.969, reducing the Mean Absolute Error by 28% to 40%. With the reference seed, the five-step model obtained the lowest error before calibration

doi: 10.25100/iyc.v28i3.16014

(MAE = 118.02 MtCO2); however, this result was not stable: across five random seeds, the Mean Absolute Error of M5 was 151.98 ± 56.32 MtCO2, compared to 140.22 ± 7.95 MtCO2 for M1 and 136.06 ± 3.75 MtCO2 for M3. After calibration, the three models performed similarly, with M3 being the most accurate and stable (MAE = 86.65 ± 4.33 MtCO2). Therefore, the results do not support a systematic advantage of longer time windows. The benchmarks also show that the GRU models do not outperform simpler alternatives in short-term prediction: persistence, which predicts each country's emissions with the value of the previous year, achieved MAE = 10.97 MtCO2 and R^2 = 0.999, and Random Forest, which uses the same predictors as the GRU models, achieved MAE = 45.83 MtCO2 and R^2 = 0.989. The LSTM network obtained errors similar to those of the GRU models (MAE = 122.25 MtCO2). Overfitting occurs when the model captures idiosyncratic noise from the training set, deteriorating its ability to generalize to unseen data (5), manifesting as increasing divergence between training and validation metrics. For each optimized model, the learning curves of the final training (up to 200 epochs with early stopping) were analyzed, monitoring training loss and validation loss, and the training and validation R^2 were calculated at the best epoch. Two quantitative indicators were calculated: the first is the loss gap, defined as validation loss minus training loss, calculated on the scaled data without the regularization terms; since this indicator depends on the scale of the variable, it is reported descriptively. The second indicator is the R^2 difference, defined as training R^2 minus validation R^2, where values below 0.02 indicate no overfitting, values between 0.02 and 0.05 indicate slight overfitting, and values above 0.05 indicate severe overfitting.

Table 6. Performance metrics in the test set (2022-2023)

Model	MAE	RMSE	R^2	MAPE (%)
Not Optimized	283.64	844.74	0.053	243.43
M1	140.64	202.33	0.946	197.83
M3	138.98	201.20	0.946	196.18
M5	118.02	188.31	0.953	156.21
M1 calibrated	87.05	153.43	0.969	72.37
M3 calibrated	83.74	156.16	0.968	68.40
M5 calibrated	85.11	154.79	0.968	72.21
LSTM (5 steps)	122.25	192.65	0.951	183.55
Persistence (t-1)	10.97	26.11	0.999	5.11
Persistence (t-5)	33.00	74.49	0.993	15.53
Fixed-effects panel	74.08	128.23	0.978	125.13
Random Forest	45.83	91.47	0.989	27.48

Table 7 shows that the one-step model exhibits no overfitting (ΔR^2 = 0.002), while the three- and five-step models show slight overfitting, with R^2 differences of 0.031 and loss gaps of 0.262, consistent with their greater capacity to fit the training sequences. Even so, the validation R^2 remains high in the three models (0.938 to 0.946), and early stopping selected the best epochs (83, 27 and 65, respectively) before the validation loss diverged. Similarly, in Figure 4, the loss curves show a rapid decrease during the first 20 epochs, oscillations in the validation loss of the one- and five-step models, and a gradual separation between training and validation loss in the three-step model after epoch 20, which justifies the use of early stopping and of the dropout and L2 regularization strategies.

doi: 10.25100/iyc.v28i3.16014

Table 7. Overfit analysis results

Model	Best Epoch	Gap Loss	R^2 Train	R^2 Val	ΔR^2	Diagnosis
M1	83	0.007	0.940	0.938	0.002	No overfitting
M3	27	0.262	0.977	0.946	0.031	Slight overfitting
M5	65	0.262	0.977	0.946	0.031	Slight overfitting

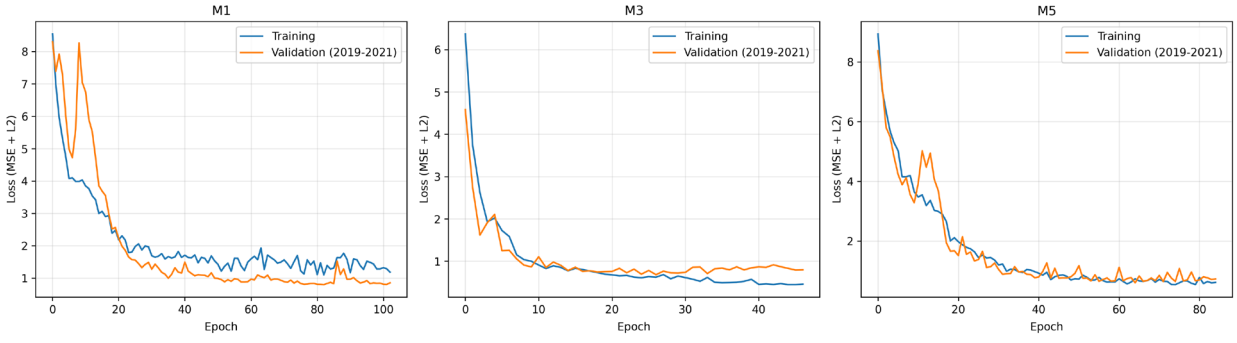


Figure 4 Overfit analysis results: Loss curves for the three models

Post-processing calibration corrects these biases without retraining, improving operational reliability. Following the five-time-step model, a three-stage sequential calibration pipeline was implemented (40). The first stage of calibration by quartiles aims to correct heteroscedasticity in magnitude-dependent prediction errors, using predictions in the calibration set corresponding to the years 2019 to 2021 as a proxy for future real values. Residuals were calculated as the difference between the actual value and the predicted value, stratifying these residuals by quartiles of the predicted value, since the actual value is unknown at the time of prediction. The correction factor was calculated as the mean of residuals per quartile, and additive correction was applied where the calibrated prediction equals the base prediction plus the corresponding quartile factor.

Table 8 reveals that the model underestimates emissions in the four quartiles of the predicted value during the calibration period, with a moderate bias in the first three quartiles (+5.77 to +10.09 MtCO2) and a much larger bias in the highest quartile (+46.39 MtCO2). This behavior is consistent with the compression of the predictions towards the center of the distribution observed in Table 5, which is common in Mean Squared Error-trained neural networks. The second stage of post-processing by historical ranges aims to avoid physically implausible predictions, or those residing outside the country's historical ranges. For each country, historical emissions from the period 1960 to 2021 were extracted, and limits were calculated with a tolerance margin, where the lower limit is 0.5 times the historical minimum, and the upper limit is 1.2 times the historical maximum. The lower limit (50% of the historical minimum) allows for severe economic contractions while avoiding implausible near-zero predictions, and the upper limit (120% of the historical maximum) allows for moderate growth while preventing extreme extrapolations.

doi: 10.25100/iyc.v28i3.16014

Table 8. Quartile correction factors

Quartile	Range of predicted value (MtCO2)of predicted value (MtCO2)	Factor (MtCO2)
Q1	≤ 50.4	+10.09
Q2	50.4-204.5	+10.02
Q3	204.5-249.9	+5.77
Q4	> 249.9	+46.39

Note. Positive factors indicate underestimation and are added to the base prediction.

The third stage of temporal smoothing aims to ensure year-on-year consistency by penalizing unrealistic abrupt changes, defining a maximum annual change of 25%. For each country and year, the percentage change is calculated as the absolute value of the current year's prediction minus the previous year's prediction, divided by the previous year's prediction.

Table 9 shows that the calibration reduces the Mean Absolute Error of the five-step model by 27.9% in the test set (from 118.02 to 85.11 MtCO2) and the MAPE by 53.8%, without requiring retraining of the model. To evaluate the differential effect of the calibration, countries were stratified into three categories according to 2023 emissions, finding that low emitters (under 100 million tons of CO2; 19 countries) present a Mean Absolute Error improvement of 57.2% (from 84.85 to 36.30 MtCO2), mainly because the historical ranges limit the overestimation of small countries. Medium emitters (100 to 500 million tons of CO2; 9 countries) show an improvement of 3.8%, and high emitters (over 500 million tons of CO2; 4 countries) an improvement of 6.9%, since the calibration does not correct the underestimation of Germany and Japan. Calibration is therefore particularly effective in low-emission countries, where the bias of the base model is more pronounced.

Table 9. Calibration impact on model performance

Year 2022

Metric	Pre-Cal.	Post-Cal.	Improv.	%
MAE	116.47	83.17	33.29	28.6
RMSE	191.42	156.08	35.34	18.5
R^2	0.953	0.969	0.016	1.7
MAPE (%)	146.94	70.21	76.73	52.2

Year 2023

Metric	Pre-Cal.	Post-Cal.	Improv.	%
MAE	119.58	87.04	32.54	27.2
RMSE	185.16	153.48	31.68	17.1
R^2	0.953	0.968	0.015	1.5
MAPE (%)	165.48	74.22	91.26	55.2

doi: 10.25100/iyc.v28i3.16014

SHAP Analysis

The black-box nature of deep neural networks makes it difficult to interpret underlying causal mechanisms, limiting confidence and adoptability in public policy contexts (19). To mitigate this limitation, SHAP (SHapley Additive exPlanations) was implemented, an interpretability method based on cooperative game theory that assigns contribution values to each predictor variable (41,42). The mathematical properties of SHAP include additivity, where the sum of contributions equals the prediction minus the expected value of the function; local consistency, where if one function favors one variable more than another function, then the SHAP value of the first is greater than or equal to that of the second; and missingness, where missing variables have a SHAP value equal to zero. Given the prohibitive computational cost of calculating exact Shapley values in networks with 19 input features over five time steps, Kernel SHAP (41,42), an efficient approximation based on weighted local regression, was used. The Kernel SHAP explainer was applied to the five-step model before calibration, using as background 20 k-means centroids of the training observations for 2004-2018, the period with observed plastics trade data, and 4,000 perturbation samples for each of the 64 test predictions (32 countries × 2 years). The contribution of each variable was obtained as the sum of its SHAP values over the five time steps, expressed in MtCO2.

For each prediction i and variable j, a SHAP value greater than zero indicates that the variable increases the predicted emissions for the observation, exemplified by a SHAP value of Real GDP of +39.78 MtCO2 for Italy in 2022, meaning that Real GDP increased the predicted emissions for that observation. A SHAP value less than zero indicates that the variable reduces the predicted emissions for the observation, exemplified by a SHAP value of Circular Economy Laws of -6.86 MtCO2 for Spain in 2022, the year in which this country enacted a circular economy law, meaning that this variable reduced the predicted emissions for that observation. The absolute value of the SHAP indicates the magnitude of the impact or importance, where values close to zero indicate negligible contribution. Absolute SHAP values were added across the 64 test predictions (2022-2023) to obtain a global ranking of importance, calculated as the mean of the absolute SHAP values across observations.

As shown in Table 10, the aggregate SHAP analysis of the 64 test predictions (32 countries × 2 years) of the five-step model reveals that exports and imports of plastics are by far the most influential variables (mean absolute SHAP values of 269.14 and 246.64 MtCO2), followed by the labor force (34.97 MtCO2), while the remaining variables, except circular economy laws, show mean absolute contributions between 7.80 and 10.74 MtCO2. However, the two trade variables should be interpreted with caution: they are highly correlated (r = 0.90 for 2004-2023), they act in opposite directions in most countries (for example, +382.74 and -427.30 MtCO2 for Germany in 2022), and their largest contributions are concentrated in the United States (+2,523.72 and +1,947.26 MtCO2 in 2022), which suggests that the model uses them mainly as proxies for the size of the economy rather than as independent drivers of emissions. In addition, both variables were imputed for the years prior to 2004, which limits the interpretation of their effect. The labor force shows a consistent reducing effect in 83% of the cases, while most economic and demographic variables, such as Real GDP, GDP per capita and the population age groups, show mixed effects

doi: 10.25100/iyc.v28i3.16014

whose sign varies between countries. In contrast, circular economy laws have the weakest impact of all the variables (mean absolute SHAP value of 2.15 MtCO2), with a zero contribution in 31 of the 64 predictions and an almost equal number of positive (16) and negative (17) cases, so the model does not identify a consistent effect of this legislation on emissions in the test period.

Table 10. SHAP values of the five-step model in the test set (2022-2023)

Variable	Mean |SHAP|	Mean SHAP				% Pos.*	Direction
Plastic exports	269.14	109.05	30	34	0	46.9	Mixed (varies by country)
Plastic imports	246.64	-101.66	12	52	0	18.8	Reduces emissions (consistent)
Labor force	34.97	-1.94	11	53	0	17.2	Reduces emissions (consistent)
Mining GDP	10.73	-2.50	31	33	0	48.4	Mixed (varies by country)
Real GDP	10.30	-1.59	32	32	0	50.0	Mixed (varies by country)
Pop. +65	10.27	-3.03	24	40	0	37.5	Mixed (varies by country)
Productivity	9.98	-1.10	33	31	0	51.6	Mixed (varies by country)
Population	9.95	0.19	33	31	0	51.6	Mixed (varies by country)
GINI	9.73	-0.72	29	35	0	45.3	Mixed (varies by country)
Country ID	8.95	2.54	35	29	0	54.7	Mixed (varies by country)
Exchange rate	8.88	-0.08	31	33	0	48.4	Mixed (varies by country)
Pop. density	8.84	0.28	40	24	0	62.5	Mixed (varies by country)
Manuf. GDP	8.80	-1.34	36	28	0	56.2	Mixed (varies by country)
Unemployment	8.79	-2.07	31	33	0	48.4	Mixed (varies by country)
GDP per capita	8.54	-1.98	26	38	0	40.6	Mixed (varies by country)
Pop. 15-64	8.45	-1.01	32	32	0	50.0	Mixed (varies by country)
Pop. growth	8.07	2.53	37	27	0	57.8	Mixed (varies by country)
Pop. 0-14	7.79	1.23	42	22	0	65.6	Mixed (varies by country)
Circular econ. laws	2.15	0.50	16	17	31	48.5	Mixed (varies by country)

Note. SHAP values in MtCO2. *Percentage of positive values among non-zero values; the direction is considered consistent when at least 70% of the non-zero values have the same sign.

Discussion

The findings of this study, which uses an optimized GRU model with one-, three-, and five-step time windows to predict CO2 emissions in 32 OECD countries during the period 1960-2023, present both significant convergences and divergences with the state of the art in predicting emissions using deep learning techniques.

The calibrated GRU models achieved coefficients of determination (R^2) between 0.968 and 0.969 in the test set (2022-2023), with the calibrated three-step model as reference (MAE = 83.74 MtCO2), values comparable to those reported in recent literature using machine learning techniques (18,20,21). For example, Moayedi et al. (21) obtained an R^2 of 0.9999 with an optimized multilayer perceptron (MLP) in Central Europe, while Tian et al. (20) reported an R^2 of 0.99 with multiple linear regression in the United States. However, this comparison must be interpreted with caution: annual CO2 emissions are highly persistent, so that a simple model that repeats the value of the previous year achieved an R^2 of 0.999 in this study. High R^2 values therefore do not by themselves

doi: 10.25100/iyc.v28i3.16014

demonstrate predictive superiority, and comparisons with simple benchmarks, rarely reported in the reviewed literature, are necessary to assess the real contribution of complex models.

The choice of the GRU architecture was based on its demonstrated ability to model long-term temporal dependencies with lower computational complexity than LSTM networks, representing a significant practical advantage. This decision is aligned with the trend identified in the state of the art of employing models capable of handling complex non-linearities and temporal dependencies (29,32). Like the LSTM and CNN-LSTM models cited in the literature, the GRU proved to be effective in capturing the dynamics of emissions over 64 years of historical data, with the additional advantage of requiring less computational capacity and fewer parameters, which reduces the risk of overfitting. In this study, the LSTM network with the same configuration as the five-step model obtained errors similar to those of the GRU models (MAE = 122.25 MtCO2), which supports the choice of the GRU as an equally accurate and more parsimonious alternative.

Contrary to the initial hypothesis, the comparison between time windows did not show a systematic advantage of longer histories. Although the five-step model obtained the lowest error before calibration with the reference seed (MAE = 118.02 MtCO2), its performance was the most sensitive to initialization (MAE = 151.98 ± 56.32 MtCO2 across five seeds), while the one- and three-step models were more stable, and after calibration the three configurations reached similar errors, with the three-step model being the most accurate and stable (MAE = 86.65 ± 4.33 MtCO2). This result suggests that, with annual data, the additional information provided by longer windows is limited and is offset by greater variability in training. The systematic comparison of window lengths and the evaluation of robustness across random seeds, rarely reported in previous studies, therefore constitute a methodological contribution of this work.

The SHAP analysis partially converges with the state of the art regarding the importance of economic and demographic variables. The labor force showed a consistent reducing effect in 83% of the predictions, while Real GDP, GDP per capita and the population age groups showed mixed effects whose sign varied between countries. These variables are recognized as key determinants of emissions in previous studies (12,24,26); however, their heterogeneous effects in this study suggest that their influence depends on the structural context of each country, such as its energy mix and productive specialization, which the model does not observe directly.

However, the SHAP analysis also revealed results that require caution. Exports and imports of plastics obtained the highest mean absolute contributions (269.14 and 246.64 MtCO2), but with opposite signs in most countries and with their largest values concentrated in the United States. Since both variables are highly correlated (r = 0.90) and were imputed for the years prior to 2004, these contributions more likely reflect the use of plastics trade as a proxy for the size of the economy than an independent effect of this trade on emissions. Future studies should evaluate this variable with complete series and with techniques that account for the correlation between predictors, such as grouped SHAP values or conditional permutation importance.

In terms of input variables, this study simultaneously incorporated 18 predictor variables covering demographic, economic, commercial, and environmental policy dimensions, representing a more comprehensive multidimensional approach than most studies reviewed. While Li et al. (18)

doi: 10.25100/iyc.v28i3.16014

mainly used transport and economic variables, and Adebowale & Uzel (22) focused on energy consumption, this study specifically integrated indicators from the plastics trade and circular economy legislation. This inclusion directly addresses the recurring criticism in the literature about the dependence on univariate or limited factors (32).

A limitation frequently noted in studies based on annual data is the small sample size. Although this study used a panel of 2,048 country-year observations, the test set includes only 64 observations (32 countries in 2022 and 2023), so the results should be confirmed with longer evaluation periods as new data become available.

Lastly, circular economy laws had the weakest contribution of all the variables (mean absolute SHAP value of 2.15 MtCO2), with no consistent direction, so the model does not identify an effect of this legislation on emissions in the test period. This result can be attributed to the fact that only 37 laws were enacted in the countries studied, most of them recently (half of them since 2018), and to their coding as a binary indicator of the year of enactment, which does not capture the accumulation or enforcement of the regulations. As these policies mature, their effects could become observable, and future studies could represent them through cumulative indicators of the laws in force.

This study has several limitations. First, simple benchmarks such as persistence and Random Forest achieved lower errors than the GRU models, so the contribution of the proposed approach lies in its ability to relate emissions to structural, trade and regulatory indicators without using past emissions as inputs, rather than in short-term forecasting accuracy. Second, the imputation by global median of variables with high proportions of missing data, such as plastics trade (68.8% and 69.7%) and productivity (70.3%), may introduce biases, particularly in the years prior to 2004. Third, the results of the neural networks were sensitive to initialization, which was addressed by reporting their variability across five random seeds. Fourth, energy consumption variables were intentionally excluded, since energy-related CO2 emissions are almost an accounting identity of fossil fuel consumption, and their inclusion would have obscured the contribution of the non-energy structural determinants that the study aimed to analyze; nevertheless, this decision limits the predictive accuracy of the model. Finally, SHAP values describe associations learned by the model and not causal effects, and the numerical country identifier used as input may introduce an artificial ordering between countries.

Conclusions

This study developed and validated a predictive model of CO2 emissions for 32 OECD countries, integrating economic, demographic, commercial and environmental policy indicators through a neural network architecture based on Gated Recurrent Units (GRU) optimized with the Optuna algorithm. The results show that Bayesian optimization and post-processing calibration substantially improve the accuracy of the networks, although the models must be compared with simple benchmarks to assess their real contribution.

doi: 10.25100/iyc.v28i3.16014

In terms of performance, optimization and calibration raised the R^2 in the test set (2022-2023) from 0.053 for the base model to 0.968 for the calibrated three-step model, with a mean absolute error (MAE) of 83.74 MtCO2. The three time windows reached similar errors after calibration, and the five-step model was the most sensitive to initialization, so the results do not support a systematic advantage of longer time windows. However, persistence (MAE = 10.97 MtCO2) and Random Forest (MAE = 45.83 MtCO2) achieved lower errors, which indicates that the value of the proposed models lies in relating emissions to structural indicators without using past emissions, rather than in short-term forecasting.

The SHAP analysis showed that plastics trade and the labor force were the variables with the greatest contribution to the predictions, although the effect of plastics trade more likely reflects its role as a proxy for the size of the economy than an independent driver of emissions. Most economic and demographic variables showed mixed effects between countries, which suggests that mitigation policies should consider the structural context of each country rather than uniform approaches.

Three main limitations persist: 1) the imputation of variables with high proportions of missing values, such as productivity (70.3%) and plastics trade (68.8% and 69.7%), may introduce systematic biases; 2) large errors persist in countries with atypical dynamics, such as Japan and Germany, which remain underestimated by more than 270 MtCO2 after calibration, indicating structural factors not captured, such as the energy mix, whose variables were intentionally excluded; 3) SHAP describes associations learned by the model and does not identify causal effects or complex interactions between variables.

Methodologically, this research provides evidence that the combination of recurrent neural networks, Bayesian optimization, sequential calibration and robustness analysis across random seeds constitutes a transparent framework for modeling emissions in extensive time panels. It also reinforces the relevance of interpretable approaches (such as SHAP) to transform “black box” models into useful tools for the formulation of evidence-based environmental policies. Circular economy laws showed the weakest contribution and no consistent effect on emissions in the test period, which is consistent with their recent enactment; future studies should evaluate their effect with cumulative indicators of the laws in force and longer post-implementation periods.

CRediT authorship contribution statement

Conceptualization - Ideas: Elcira Solano Benavides; Nelson Alandete. Data curation: Nelson Alandete, Edie Pertuz Diaz. Formal Analysis: Elcira Solano Benavides; Nelson Alandete. Investigation: Elcira Solano Benavides; Nelson Alandete. Methodology: DElcira Solano Benavides; Nelson Alandete, Edie Pertuz Díaz. Project Management: Elcira Solano Benavides. Resources: Nelson Alandete, Edie Pertuz Díaz. Software: Nelson Alandete, Edie Pertuz Diaz Supervision: Elcira Solano Benavides; Nelson Alandete. Validaction: Elcira Solano Benavides; Nelson Alandete Writing - original draft - Preparation: Nelson

doi: 10.25100/iyc.v28i3.16014

Alandete, Edie Pertuz Diaz. Writing - revision and editing - Preparation: Elcira Solano Benavides; Nelson Alandete, Edie Pertuz Díaz.

Funding: This research received no external funding.

Conflict of Interest Statement: The authors declare no conflicts of interest.

References

(1) IPCC (Intergovernmental Panel on Climate Change). Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2021.

https://doi.org/10.1017/9781009157896

(2) Friedlingstein P, O'Sullivan M, Jones MW, Andrew RM, Bakker DCE, Hauck J, et al. Global carbon budget 2023. Earth Syst Sci Data. 2023;15(12):5301-5369.

https://doi.org/10.5194/essd-15-5301-2023

(3) Organisation for Economic Co-operation and Development (OECD). OECD environmental outlook to 2050: the consequences of inaction. Paris: OECD Publishing; 2012.

https://doi.org/10.1787/9789264122246-en

(4) Tian Y, Ren X, Li K, Li X. Carbon Dioxide Emission Forecast: A Review of Existing Models and Future Challenges. Sustainability. 2025;17(4):1471.

https://doi.org/10.3390/su17041471

(5) Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge (MA): MIT Press; 2016.

https://mitpress.mit.edu/9780262035613/deep-learning/

(6) Geissdoerfer M, Savaget P, Bocken NMP, Hultink EJ. The circular economy—a new sustainability paradigm? J Clean Prod. 2017;143:757-768.

https://doi.org/10.1016/j.jclepro.2016.12.048

(7) Houdou A, El Badisy I, Khomsi K, Abdala SA, Abdulla F, Najmi H, et al. Interpretable machine learning approaches for forecasting and predicting air pollution: a systematic review. Aerosol Air Qual Res. 2024;24(1):230151.

https://doi.org/10.4209/aaqr.230151

(8) Pachauri RK, Allen MR, Barros VR, Broome J, Cramer W, Christ R, et al. Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva: IPCC; 2014.

https://www.ipcc.ch/site/assets/uploads/2018/05/SYR_AR5_FINAL_full_wcover.pdf

(9) Ellen MacArthur Foundation. Completing the picture: how the circular economy tackles climate change (Internet). 2019 (cited 2026 Jul 11). Available from: https://www.ellenmacarthurfoundation.org/publications/completing-the-picture

(10) Zheng X, Streimikiene D, Balezentis T, Mardani A, Cavallaro F, Liao H. A review of greenhouse gas emission profiles, dynamics, and climate change mitigation efforts across the key climate change players. J Clean Prod. 2019;234:1113-33.

https://doi.org/10.1016/j.jclepro.2019.06.140

(11) Bajic A, Kiesel R, Hellmich M. Handle with care: challenges in company-level emissions data for assessing financial risks from climate change. J Clim Finance. 2023;5:100017.

https://doi.org/10.1016/j.jclimf.2023.100017

(12) Liddle B, Lung S. Age-structure, urbanization, and climate change in developed countries: revisiting STIRPAT for disaggregated population and consumption-related environmental impacts. Popul Environ. 2010;31(5):317-343.

https://doi.org/10.1007/s11111-010-0101-5

(13) Zheng J, Suh S. Strategies to reduce the global carbon footprint of plastics. Nat Clim Change. 2019;9:374-378.

https://doi.org/10.1038/s41558-019-0459-z

(14) Costantini L, Laio F, Mariani MS, Ridolfi L, Sciarra C. Forecasting national CO2 emissions worldwide. Sci Rep. 2024;14:22438.

https://doi.org/10.1038/s41598-024-73060-0

(15) UNEP (United Nations Environment Programme). Emissions gap report 2020 (Internet). 2020 (cited 2026 Jul 11). Available from:

https://www.unep.org/resources/emissions-gap-report-2020

(16) Wood R, Grubb M, Anger-Kraavi A, Pollitt H, Rizzo B, Alexandri E, et al. Beyond peak emission transfers: historical impacts of globalization and future impacts of climate policies on international emission transfers. Clim Policy. 2020;20(sup1):S14-S27.

https://doi.org/10.1080/14693062.2019.1619507

(17) Fan R, Zhang X, Bizimana A, Zhou T, Liu JS, Meng XZ. Achieving China's carbon neutrality: predicting driving factors of CO2 emission by artificial neural network. J Clean Prod. 2022;362:132331.

https://doi.org/10.1016/j.jclepro.2022.132331

(18) Li X, Ren A, Li Q. Exploring patterns of transportation-related CO2 emissions using machine learning methods. Sustainability. 2022;14(8):4588.

https://doi.org/10.3390/su14084588

(19) Anonna FR, Mohaimin MR, Ahmed A, Nayeem MB, Akter R, Alam S, et al. Machine learning-based prediction of U.S. CO2 emissions: developing models for forecasting and sustainable policy formulation. J Environ Agric Stud. 2023;4(3):85-99.

https://doi.org/10.32996/jeas.2023.4.3.12

(20) Tian L, Zhang Z, He Z, Yuan C, Xie Y, Zhang K, et al. Predicting energy-based CO2 emissions in the United States using machine learning: a path toward mitigating climate change. Sustainability. 2025;17(7):2843.

https://doi.org/10.3390/su17072843

(21) Moayedi H, Mukhtar A, Alshammari S, Boujelbene M, Elbadawi I, Thi QT, et al. CO2 prediction for central European countries through five metaheuristic optimization techniques helping the multilayer perceptron. Eng Appl Comput Fluid Mech. 2024;18(1):2327437.

https://doi.org/10.1080/19942060.2024.2327437

(22) Adebowale K, Uzel R. Forecasting CO2 emissions in Sweden with a Bayesian neural network (bachelor's thesis). Stockholm: KTH Royal Institute of Technology, DiVA portal; 2023.

https://www.essays.se/essay/bfbf619c31/

(23) Ma N, Shum WY, Han T, Lai F. Can machine learning be applied to carbon emissions analysis: an application to the CO2 emissions analysis using Gaussian process regression. Front Energy Res. 2021;9:756311.

https://doi.org/10.3389/fenrg.2021.756311

(24) Ene Yalçın S. Development of a forecasting framework based on advanced machine learning algorithms for greenhouse gas emissions. Systems. 2024;12(12):528.

https://doi.org/10.3390/systems12120528

(25) Nie W, Ao O, Duan H. A novel grey prediction model with a feedforward neural network based on a carbon emission dynamic evolution system and its application. Environ Sci Pollut Res. 2023;30(8):20704-20.

https://doi.org/10.1007/s11356-022-23541-4

(26) Jena PR, Managi S, Majhi B. Forecasting the CO2 emissions at the global level: a multilayer artificial neural network modelling. Energies. 2021;14(19):6336.

https://doi.org/10.3390/en14196336

(27) Ahmadi MH, Dehghani Madvar M, Sadeghzadeh M, Rezaei MH, Herrera M, Shamshirband S. Current status investigation and predicting carbon dioxide emission in Latin American countries by connectionist models. Energies. 2019;12(10):1916.

https://doi.org/10.3390/en12101916

(28) Chauhan N. Predicting CO2 emission per crop-type using machine learning and neural network algorithms (master's thesis). Tilburg: Tilburg University; 2020.

https://arno.uvt.nl/show.cgi?fid=156400

(29) Li F, Hu Y, Wang L. Carbon emission prediction using CNN-LSTM. Int J Multidiscip Res Publ. 2022;5(6):30-5.

https://ijmrap.com/wp-content/uploads/2022/11/IJMRAP-V5N5P146Y22.pdf

(30) Han Z, Cui B, Xu L, Wang J, Guo Z. Coupling LSTM and CNN neural networks for accurate carbon emission prediction in 30 Chinese provinces. Sustainability. 2023;15(18):13934.

https://doi.org/10.3390/su151813934

(31) Mustafa F, Bu L, Wang Q, Yao N, Shahzaman M, Bilal M, et al. Neural-network-based estimation of regional-scale anthropogenic CO2 emissions using an Orbiting Carbon Observatory-2 (OCO-2) dataset over East and West Asia. Atmos Meas Tech. 2021;14(11):7277-7290.

https://doi.org/10.5194/amt-14-7277-2021

(32) Ajala AA, Adeoye OL, Salami OM, Jimoh AY. An examination of daily CO2 emissions prediction through a comparative analysis of machine learning, deep learning, and statistical models. Environ Sci Pollut Res. 2025;32(5):2510-35.

https://doi.org/10.1007/s11356-024-35764-8

(33) Keerthana KB, Wu SW, Wu ME, Kokulnathan T. The United States energy consumption and carbon dioxide emissions: a comprehensive forecast using a regression model. Sustainability. 2023;15(10):7932.

doi: 10.25100/iyc.v28i3.16014

https://doi.org/10.3390/su15107932

(34) Fatima N, Zheng Y, Guohua N. Globalization, institutional quality, economic growth and CO2 emission in OECD countries: an analysis with GMM and quantile regression. Front Environ Sci. 2022;10:967050.

https://doi.org/10.3389/fenvs.2022.967050

(35) Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30.

https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

(36) Rousseeuw PJ, Hubert M. Robust statistics for outlier detection. Wiley Interdiscip Rev Data Min Knowl Discov. 2011;1(1):73-9.

https://doi.org/10.1002/widm.2

(37) Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); 2014. p. 1724-34.

https://doi.org/10.3115/v1/D14-1179

(38) Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2019. p. 2623-31.

https://doi.org/10.1145/3292500.3330701

(39) Bergstra J, Yamins D, Cox D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In: International Conference on Machine Learning. PMLR; 2013. p. 115-23.

https://proceedings.mlr.press/v28/bergstra13.pdf

(40) Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. In: International Conference on Machine Learning. PMLR; 2017. p. 1321-30.

https://proceedings.mlr.press/v70/guo17a/guo17a.pdf

(41) Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-74.

https://doi.org/10.48550/arXiv.1705.07874

(42) Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56-67.

https://doi.org/10.1038/s42256-019-0138-9