



Research article

Modeling and forecasting carbon dioxide emission in Pakistan using a hybrid combination of regression and time series models

Hasnain Iftikhar^{a,c,*}, Murad Khan^b, Justyna Żywiołek^c, Mehak Khan^d,
Javier Linkolk López-Gonzales^e

^a Department of Statistics, Quaid-i-Azam University, Islamabad, 45320, Pakistan

^b Department of Statistics, Abdul Wali Khan University Mardan, Mardan, 23200, Pakistan

^c Faculty of Management, Czestochowa University of Technology, Czestochowa, 42-200, Poland

^d Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, 5063, Norway

^e Escuela de Posgrado, Universidad Peruana Unión, Lima, 15468, Peru

ARTICLE INFO

Dataset link: <https://databank.worldbank.org>

Keywords:

Carbon emission forecasting

Regression methods

Time series methods

Hybrid combination models

Energy policy

ABSTRACT

Carbon dioxide (CO₂) emissions continue to rise globally despite efforts to combat climate change. Energy industry emissions are a pressing global issue, causing devastating impacts. Hence, it is vital to accurately and efficiently forecast CO₂ emissions. Thus, this study comprehensively analyzes forecasting CO₂ emissions by comparing various hybrid combinations of regression and time series methods to explore the CO₂ emissions in Pakistan. First, divide the yearly time series of CO₂ emissions into the long-run curve trend series and the residual subseries. The long-run curve trend subseries is modeled using parametric and nonparametric regression methods, while various standard time series models are used to forecast the residual subseries. However, the forecasts of each subseries will be combined to obtain the final forecast of CO₂ emissions. This work used four different accuracy mean errors, a statistical test, and a graphical analysis as performance measures to evaluate the proposed hybrid forecasting technique. The findings confirmed that the proposed hybrid combination forecasting technique is highly accurate and efficient in forecasting CO₂ emissions. Likewise, according to the proposed final optimal hybrid combination forecasting model, Pakistan's per capita CO₂ emissions will be 1.130215 metric tons in 2030. Pakistan's escalating emission trend signals that creative solutions must be implemented to curb it. Thus, the government must price carbon footprints, regulate electricity from zero-carbon sources, reduce population, encourage afforestation in densely populated areas, adopt clean technology, and fund research.

1. Introduction

Human activities have significantly increased greenhouse gas (GHG) concentration and emissions since 1750 [1]. GHGs trap heat near the Earth's surface and are crucial for maintaining Earth's habitable temperature [2]. However, excessive consumption of fossil fuels and other resources contributes to GHG overburden, potentially disrupting the Earth's carbon cycle and contributing to

* Corresponding author at: Department of Statistics, Quaid-i-Azam University, Islamabad, 45320, Pakistan.

E-mail address: hasnainchill3@gmail.com (H. Iftikhar).

<https://doi.org/10.1016/j.heliyon.2024.e33148>

Received 20 March 2024; Received in revised form 14 June 2024; Accepted 14 June 2024

Available online 19 June 2024

2405-8440/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

global warming. The primary human-emitter activities include electricity generation, heat, and transport, with energy consumption contributing around 73% to emissions. Deforestation, the fossil-fuel industry, livestock, and fertilizer also contribute to 18% of emissions of CO₂, methane, and nitrous oxide [3–5]. The global CO₂ emissions in 2022 were 36.1 ± 0.3 GtCO₂, reflecting 2%, 7.9%, and 1.5% growth compared to 2019, 2020, and 2021, respectively [6,7]. The Carbon Monitor data suggested 1.5% (0.9–2.6%) growth for the year 2021 to 2022, along with other energy consumption-based projected growth of $1.0 \pm 0.9\%$ by Global Carbon Project (CGP) and the International Energy Agency (IEA) [8–10].

In 2020, COVID-19 restrictions decreased global CO₂ emissions, but they have since increased beyond pre-pandemic levels [11, 12]. This suggests that the peak CO₂ emissions may not have reached yet, and the temperature may hit 5 °C before 2030 pre-industrial levels, the limit set by G7 countries and the Paris Agreement [13–15]. Developed countries contribute more CO₂ emissions due to their energy industries and transportation, causing higher carbon emissions. China is the world's largest emitter of CO₂, emitting around 10 billion tonnes annually, accounting for 28.8% of global emissions [16–18]. This rapid increase in emissions has led to catastrophes like floods, droughts, rising healthcare issues, and other global warming and climate change calamities [19–22]. Pakistan, the 19th country with the highest GHG emissions, has been severely affected by devastating floods, affecting 33 million people and causing 1,700 deaths [23–25]. With a growing population of 220 million, Pakistan is expected to emit 2.9 tonnes per capita of CO₂ by 2025 and 5.4 tonnes per capita by 2050. To address these issues, stakeholders need to adopt productive and efficient policies. Pakistan's agriculture sector is also affected by population growth, leading to deforestation and increased CO₂ emissions [26–28].

The growing population also causes an increase in automobiles, with carbon emissions used in the country. The government of Pakistan is trying to mitigate and reduce CO₂ emissions by adopting specific programs that include technology-based innovations and nature-based solutions and launching a mass-scale program of planting ten billion trees Tsunami. These were found very beneficial in providing livelihood opportunities to the youth and reducing the emission to some extent from the high-emission industry and energy sector [29,30]. However, these steps can further be strengthened by proper planning, policy-making, and time decision-making to meet the conditionally targeted goals set in the Paris Agreement to reduce an overall emission of 50% by 2030, with a reduction of 15% from own resources and 35% from the expected grants from international finance. In addition, Pakistan aims to shift 30% on electric vehicles, 60% on renewable energy, and to impose a ban on coal imports along with expanding nature-based solutions by 2030 [31,32]. Therefore, these challenges motivate us to choose Pakistan for this research, and it aims to assist governments in making policies and achieving their targeted goals in the light of efficient modeling and forecasting of CO₂ emission in Pakistan.

Statistical, machine learning (ML), time series, and hybrid models are widely used in the literature for modeling and forecasting CO₂ emissions [33–37]. Pakistan is one of the top victims of climate change, and it is very crucial to forecast CO₂ emissions here; therefore, the researchers in [38] used a univariate model, ARIMA (autoregressive integrated moving average) model for CO₂ emission forecast of 2030 based on the developed scenario of CPEC (China Pakistan Economic Corridor) in which most of the projects are energy oriented. They attempted to estimate emission reduction percentages and assist the government in adopting suitable policies accordingly. Their findings last revealed the forecast with less than 10% mean absolute percentage error. The authors in [39] used the time series models, including the ARIMA model, the Holt-Winters models, and the seasonal ARIMAX model with exogenous factors and ML and deep learning models such as Linear Regression, Random forest, long short-term memory (LSTM) models to predict ten years CO₂ emission in India. Their results revealed that LSTM, SARIMAX, and Holt-Winters are the best performers of the six used models regarding the highest accuracy and lowest mean errors in predicting CO₂ emissions. The energy sector, i.e., transport and manufacturing sectors, is a leading factor in CO₂ emissions in Pakistan. Therefore, the researchers in [40] evaluated the performance of statistical and ML models for forecasting sector-wise CO₂ emission in Pakistan till 2030 and provided valuable suggestions to the government for making better decisions based on the produced results.

Unlike statistical, ML, and time series models, many researchers have created hybrid models to increase forecasting accuracy and efficiency by combining features from many methods [41–45]. For example, the researchers in [46] developed a hybrid approach of ETS and ANN based on combining linear and nonlinear exponential smoothing models to obtain various compositions of linear and nonlinear patterns in time series forecasting. It achieved the best results compared to ARIMA, ETS, MLP, and the hybrid ARIMA-ANN model [47]. Further, the researchers in [48] used the Bi-LSTM model to predict CO₂ emissions in South Asian countries and China from 2022 to 2030. The effects of technological, industrial, and scientific sectors on CO₂ emissions are studied. The results of Bi-LSTM yield the comparatively lowest mean errors compared to LSTM and Gated Recurrent Unit (GRU). Their finding declared carbon emissions a big issue, and it will become worse if India and China fail to control high emissions in the next decade. In further studies [49] to find the relationship of gross domestic product (GDP) in Bangladesh with CO₂ emission due to the consumption of electrical energy by using the fully modified ordinary least squares (FMOLS) technique, convolutional neural network (CNN), dense neural network (DNN), LSTM, and CNN-LSTM. These models were comparatively evaluated with performance measures including RMSE, MAE, and MAPE, and the results suggest a significant effect of CO₂ emissions over GDP. The findings revealed in [50] that the main contributor to CO₂ emissions is energy demand in the case of Iran, achieved by applying the Logarithmic Mean Divisia Index decomposition technique.

To motivate the above hybrid models, this work proposes various hybrid forecasting models based on parametric and nonparametric regression models with the combination (hybridization) of linear and nonlinear time series models to enhance the forecasting accuracy and efficiency of CO₂ emissions in Pakistan. The polynomial regression and spline regression models are considered in regression models. However, in the time series models, consider four standard models: two linear models, the autoregressive model, the autoregressive moving average, and two nonlinear models, the nonparametric autoregressive model and the neural network autoregressive model. In contrast, in the literature on hybrid models, the direct hybridization of two or more models was not primarily the per-define components; they only applied the model. To this end, the main difference between the current work and the above

literature is that this work discusses the time series database capture by the parametric and nonparametric regression models. The reaming series is used to model different time series models. Obtained the forecast for these regression and time series models and considered them the final estimates. In addition, the proposals can also be used for hourly, daily, weekly, and monthly time series datasets because they are worth handling other seasonality in the database, which the mentioned literature hybrid models did not offer.

Thus, this study thoroughly examines predicting CO₂ emissions by comparing sixteen distinct hybrid combinations of regression and time series methods to investigate CO₂ emissions in Pakistan. The polynomial regression and spline regression models are considered in regression models. However, in the time series models, consider four standard models: two linear models, the autoregressive model, the autoregressive moving average, and two nonlinear models, the nonparametric autoregressive model, and the neural network autoregressive model. In this context, the performance of all possible hybrid combinations of the parametric and nonparametric regression models and the considered time series will be compared using the accuracy metrics, a statistical test, and visual analysis within the proposed hybrid combination methodology. In addition, the performance of the proposed hybrid combination forecasting approach compared to the established benchmark models will be evaluated. However, within the proposed hybrid combination technique, the final best hybrid combination model is compared to the unique time series models (the autoregressive, the nonparametric autoregressive, the autoregressive moving average, and the neural network autoregressive), and the outcomes are noted. Thus, the reported results demonstrated that the proposed final best hybrid combination model is very accurate and efficient for projecting CO₂ emissions compared to all the considered benchmark models. Finally, while this study only employs the CO₂ emission information from Pakistan, it may be extended and adapted to other countries to ensure the efficacy and accuracy of the proposed hybrid combination modeling and forecasting technique.

The rest of the paper is organized in the following order: Section 2 describes the models used in developing the proposed hybrid combination modeling and forecasting strategy. Section 3 presents a practical implementation of the proposed hybrid combination forecasting technique utilizing Pakistan's annual CO₂ emissions. Section 4 discusses the best combination model with a comparison to some of the best time series. Also, it discusses the final observations policy advice for stakeholders. Finally, Section 5 addresses the conclusion, limits, and future research endeavors.

2. Method and materials

This section thoroughly reviews the models and methods for building the proposed hybrid combination forecasting technique. As a result, turn to the following subsections for further information on each model and approach.

2.1. The proposed hybrid modeling and forecasting procedure

This section will discuss the whole procedure of the proposed hybrid combination forecasting technique. Let the C_y represent the time series of CO₂ for the year y^{th} . The graphical presentation of the C_y time series is shown in 3. It can be seen from this figure the C_y time series have an increasing upward trend. To achieve this, divide the yearly time series of CO₂ emissions into two new series: the long-run curve trend series and the residual series. However, the mathematical equation can be written as;

$$C_y = L_y + R_y \quad (1)$$

In the equation (1), C_y represents the time series of CO₂ emissions; L_y , a long-run trend curve series, and R_y , a residual series, to model the long-run curve trend series (L_y) using parametric and nonparametric regression methods. Parametric regression includes polynomial regression, whereas nonparametric regression includes the smoothing regression splines method. However, on the other hand, for residual series (R_y) modeling, we consider four different univariate time series models: two linear time series models, the autoregressive model, and the autoregressive moving average; and two nonlinear models: the nonparametric autoregressive model and the neural network autoregressive model. Hence, the trend and residual series modeling details are in the coming section.

2.1.1. Modeling the long-run trend series

In this section, we will discuss how to forecast the long-run curve trend series using parametric and nonparametric regression methods. To achieve this, we will model the response variable C_y linearly by estimating the long-run trend component L_y using a cubic polynomial regression technique for time y . The mathematical form is given by

$$L_y = \phi_0 L_y + \phi_1 L_y^2 + \phi_2 L_y^3 \quad (2)$$

The regression coefficients of the above equation are determined using the ordinary least square method; after obtaining $\hat{\phi}_0$, $\hat{\phi}_1$, and $\hat{\phi}_2$, the polynomial regression model coefficients, the final estimated equation is presented.

$$\hat{L}_y = \hat{\phi}_0 L_y + \hat{\phi}_1 L_y^2 + \hat{\phi}_2 L_y^3 \quad (3)$$

The long-run trend series can be modeled nonparametrically in the nonparametric case.

$$L_y = f_1(L_y) + f_2(L_y) + f_3(L_y) \quad (4)$$

In the equation (4), each f_i is a smoothing function of L_y . As a function of time y , the long-term trend component L_y is estimated. Cubic regression splines assess the long-run trend part for smoothing functions. The number of knots and their position are the most

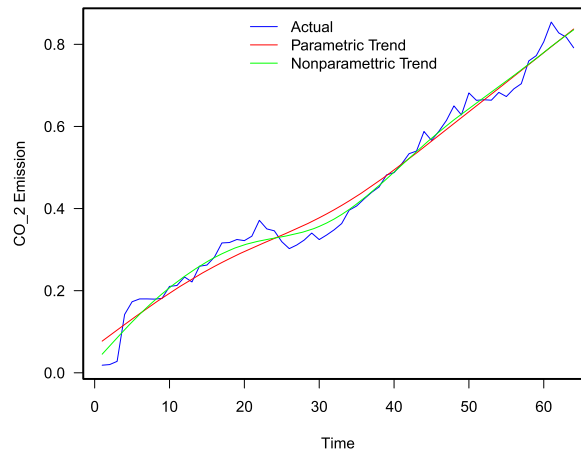


Fig. 1. Pakistan CO₂ emission time series with superimposed fitted long run trend series (L_y) for polynomial regression model (red curve) and spline regression model (green line).

critical choices in the regression spline technique since they dictate the smoothness of the approximation. For this problem, we employ the cross-validation approach. However, the ordinary least squares method is used to calculate regression coefficients.

The observed CO₂ emission series and the estimated long-run trend curve series are depicted in Fig. 1, with a parametric model (red line) and a nonparametric model (green curve) to visualize the performance of the approaches discussed above used for the estimation of long-run trend curve series L_y (both parametric and nonparametric). As the growing (upward) trend can be observed in the figure, both models employed to estimate L_y adequately capture the long-run trend curve series of CO₂ emission series (C_y).

Once the estimated long-run trend curve is obtained, the residual series can be derived as follows:

$$R_y = C_y - (\hat{L}_y) \quad (5)$$

2.1.2. Modeling to residual series

After modeling the long-run trend series (L_y) from the original time series of CO₂ (C_y) using the polynomial regression (parametric case) and regression spline (nonparametric case), the next step is to forecast the residual series (R_y) using four commonly time series models including, the parametric autoregressive model, the autoregressive moving average model, the nonparametric autoregressive model, and the neural network autoregressive model. These models are detailed in the following subsections.

2.1.3. Parametric autoregressive model

A linear and parametric autoregressive (AR) model, which uses a linear combination of the previous observations of the R_y time series and represents the short-term dynamics of this series, is referred to as:

$$R_y = \alpha + \vartheta_1 R_{y-1} + \vartheta_2 R_{y-2} + \dots + \vartheta_n R_{y-n} + \varepsilon_n \quad (6)$$

In the above formula, α is an intercept term, and ϑ_j ($j = 1, 2, \dots, n$) is the slope parameter of the underlying AR process, and ε_d is the disturbance term [51]. However, the present study uses the maximum likelihood method to estimate the parameters of the AR model. The model includes lags 1, 2, and 3 based on their significant results in the series' correlogram (autocorrelation and partial autocorrelation functions).

2.1.4. Autoregressive moving average model

The autoregressive moving average (ARMA) model integrates the target variable's previous values and vital information as moving averages (the error lags). In our scenario, the study variable R_y is explained by the prior terms, as are the delayed residual values [52]. Mathematically,

$$R_y = \alpha + \vartheta_1 R_{y-1} + \vartheta_2 R_{y-2} + \dots + \vartheta_n R_{y-n} + \varepsilon_n + \zeta_1 \varepsilon_{y-1} + \zeta_2 \varepsilon_{y-2} + \dots + \zeta_m \varepsilon_{y-m} \quad (7)$$

In the last equation, α denotes intercept, ϑ_j ($j = 1, 2, \dots, n$) and ζ_k ($k = 1, 2, \dots, m$) are the parameters of AR and MA process respectively, and ε_n is a Gaussian white noise series with mean zero and variance σ_ε^2 . Hence, the maximum likelihood method estimates the unknown population parameters. The ARMA model order selection, which is the number of past values and the past error term value, is established by examining the correlograms. In the MA part, the first two lags are significant, while in the AR part, only lags 1, 2, and 3 are significant for the residual time series (R_y).

2.1.5. Nonparametric autoregressive model

The nonlinear or nonparametric alternative of the AR process leads to the additive model (NPAR), in which the relationship between R_y and its prior terms does not have any specified parametric form, allowing for any nonlinearity, which is stated as:

$$R_y = f_1(R_{y-1}) + f_2(R_{y-2}) + \dots + f_y(R_{y-n}) + \epsilon_n \quad (8)$$

where $f_j (j = 1, 2, \dots, n)$ depict smoothing functions and define the relationship between R_y and its past values [53]. As a result, the functions f_j are represented by cubic regression splines in this study and lags 1, lag 2, and lag 3 are employed for NPAR modeling.

2.1.6. Neural network autoregressive model

A neural network autoregressive (NNA) machine learning model forecasts future time series values based on past observations [54]. The model employs a mathematical function that takes into account the past values of the time series, denoted by $R_{y-1}, R_{y-2}, \dots, R_{y-n}$, where n is the time delay parameter. The NNA model is trained using backpropagation and the steepest descent strategy to reduce the discrepancy between anticipated and actual values.

An NNA model forecasting procedure consists of two parts. First, the order of autoregression is calculated, which refers to the number of prior values required to forecast the time series' present value. Second, the NN is trained with a training set that considers the order of autoregression. In univariate time series forecasting, the number of input nodes is determined by the order of autoregression, and the inputs are the prior, lagged data. The anticipated values represent the NN model's output. As there is no theoretical foundation for setting them, trial and error or experimentation frequently establish the number of hidden nodes. It is essential to select the number of iterations carefully to avoid overfitting. In this work, an NNA (3,2) design is used, which can be represented as $R_y = f(R_{y-1})$. Here, $R_{y-1} = (R_{y-1}, R_{y-2}, R_{y-3})$ is a vector containing past values of the residual time series of (R_y), and f is a neural network with 3 hidden nodes in a signal layer.

2.2. Accuracy measures

The proposed hybrid combination approach is evaluated using three criteria: a) accuracy mean errors, b) statistical tests, and c) graphical presentation. The considered accuracy mean errors are the following: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Root Mean Square Percentage Error (RMSPE). The functional form of each error is given by;

$$MAE = \frac{1}{y} \sum_{y=1}^y (|C_y - \hat{C}_y|), \quad (9)$$

$$MAPE = \frac{1}{y} \sum_{y=1}^y \left(\frac{|C_y - \hat{C}_y|}{|C_y|} \right) \times 100, \quad (10)$$

$$RMSPE = \sqrt{\frac{1}{y} \sum_{y=1}^y \left(\frac{|C_y - \hat{C}_y|}{|C_y|} \right)^2} \times 100, \quad (11)$$

$$RMSE = \sqrt{\frac{1}{y} \sum_{y=1}^y (C_y - \hat{C}_y)^2}, \quad (12)$$

where C_y is the actual value and \hat{C}_y is the projected value for the y^{th} value ($y = 1, 2, \dots, Y = 20$). However, The smaller the MAE, MAPE, RMSPE, and RMSE values, the better the model's forecasting accuracy.

In contrast to the performance indicators described above, the Diebold-Mariano (DM) [55] test is a frequently employed statistical test in the literature for assessing forecasts from different forecasting models [56–59]. For example, consider the two forecasts produced by two distinct time series models that include as \hat{C}_{1y} (forecasted values of model 1) and \hat{C}_{2y} (forecasted values of model 2). However, $Q_{1d} = C_y - \hat{C}_{1y}$ and $Q_{2y} = C_y - \hat{C}_{2y}$ are the respectively forecast errors. The loss associated with forecast error $\{Q_{iy}\}_{i=1}^2$ by $L(Q_{iy})$. For instance, time y absolute loss would be $L(Q_{iy}) = |Q_{iy}|$. For time y , the loss resulting from forecasts 1 and 2 is so $z_y = L(Q_{1y}) - L(Q_{2y})$. The null hypothesis with the same accuracy in forecasting for two forecasts is $E[z_y] = 0$. The DM test requires that the loss difference have a constant covariance, that is,

$$E[z_y] = u, \quad \forall y \quad (13)$$

$$\text{cov}(z_y - z_{y-n}) = \rho(n), \quad \forall y \quad (14)$$

$$\text{var}(z_y) = \sigma_z, \quad 0 < \sigma_z < \infty \quad (15)$$

Given these assumptions, a DM test with equivalent accuracy of forecasts is proposed.

$$DM = \frac{\bar{z}}{\hat{\sigma}_z} \xrightarrow{n} \text{Norm}(0, 1) \quad (16)$$

In the given formula (16), $\bar{z} = \frac{1}{y} \sum_{y=1}^y z_y$ is the mean sample loss difference, and $\hat{\sigma}_z$ is a consistent standard error estimate of z_y .

Ultimately, this study demonstrates the superiority of the suggested hybrid combination forecasting technique using various figures such as the line plot, correlogram plot (autocorrelation and partial autocorrelation functions), bar plot, and dot plot.

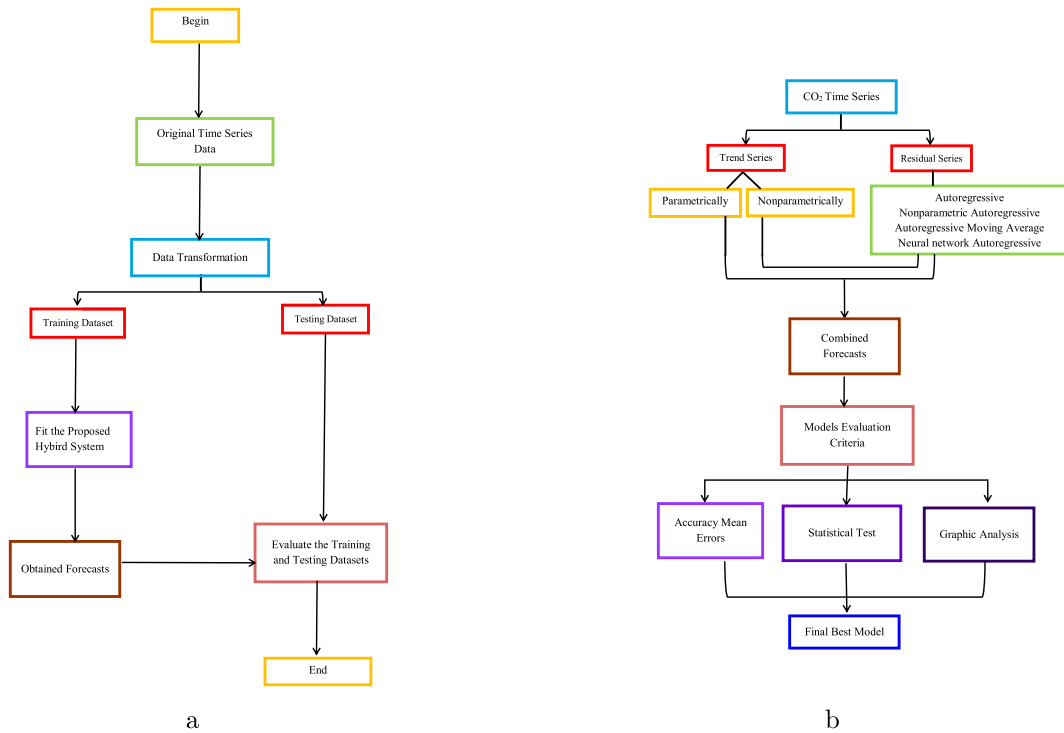


Fig. 2. Pakistan Carbon Dioxide Emission (metric tons per capita): (a) the training and testing phases of the proposal.; (b) the proposed hybrid combination modeling and forecasting approach is depicted in a flowchart.

Finally, combining the long-run curve trend series forecast and the residual series forecast leads to sixteen possible combinations for comparison purposes ($4^2 = 16$). Each combination model with parametric and nonparametric regression models is denoted in this study by ${}^{L_y}M^{R_y}$, where the L_y at the top-left is related with a long run curve trend series, and the R_y represents the residual series at the top-right. This work assigns a code to each model in the forecasting models: “a” for the AR, “b” for the ARMA, “c” for the NPAR, and “d” for the NNA. For example, ${}^0M^{a1}$ represents the estimate of the long-term trend (L_y) with the parametric regression (polynomial regression model), and the residual series (R_y) estimated using the AR model and ${}^1M^{a1}$ represents the estimate of the long-term trend (L_y) with the nonparametric regression (smoothing spline regression model), and the residual series (L_y) estimated using the AR model similarly ${}^0M^{b1}$ shows parametric regression using ARMA model and ${}^1M^{b1}$ presents nonparametric model with ARMA and so on ${}^0M^{d1}$ denotes parametric with neural network auto-regressive and ${}^1M^{d1}$ represents nonparametric with NNA respectively. Therefore, the different forecasting models are integrated to provide a final year-ahead CO₂ emission forecast. The final equation is as follows:

$$\hat{C}_{y+1} = (\hat{L}_{y+1} + \hat{R}_{y+1}) \quad (17)$$

At the end of this section, Fig. 2 depicts the proposed hybrid combination modeling and forecasting strategy to finish this part.

3. Case study outcomes

This research study compiles CO₂ emissions in Pakistan from 1949 to 2021. According to international standards, CO₂ was measured in metric tons per capita, and the dataset was taken from the World Bank website (<https://databank.worldbank.org>). The dataset's features (descriptive statistics) are shown numerically in Table 1, and the CO₂ emissions time series is illustrated visually in Fig. 3. According to the descriptive data in Table 1, the mean CO₂ emissions time series from 1949 to 2021 are 0.4974 metric tons per capita, with a low of 0.0185 metric tons per capita and a maximum of 0.9990 metric tons per capita. On the other hand, the skewness has a 0.1518, and kurtosis has a 0.1997. These values show that the dataset is not normally distributed. In addition, from Fig. 3, it can be observed that there is an increasing upward trend, which indicates a seasonal direction in the time series of CO₂ emission. In addition, with a gradual increase in CO₂ emission from 1949 to 1971 and attempted the first peak in 1972. After this, the trend slightly decreased during 1975, and again it took an increasing trend and the 2nd highest peak attempt in 2007. After the second peak, again a downward fall trend for a short period in 2011. Similarly, there was an increasing CO₂ emission trend and another peak attempt in 2019, while a slightly downward trend was observed in 2020. The region's separation mainly caused these fluctuations, the unhealthy political environment, and unsuccessful plan implementations. Hence, in the proposed hybrid combination forecasting approach, we remove the long-run trend series using the parametric and nonparametric models to achieve this. It is worth noting that the stationarity of time series data is generally checked using the Augmented Dickey-Fuller, Unit root, and Philips-Perron tests.

Table 1
Pakistan Carbon Dioxide Emission (metric tons per capita): CO₂ time series descriptive statistics in Pakistan from 1949 to 2021.

S. No	Statistic	Value
1	Minimum	0.0185
2	Q1	0.3168
3	Median	0.4527
4	Mean	0.4974
5	Q3	0.6978
6	Max	0.999
7	ADF	0.4774 (-2.2405)
8	Skewness	0.1518223
9	Kurtosis	1.977317

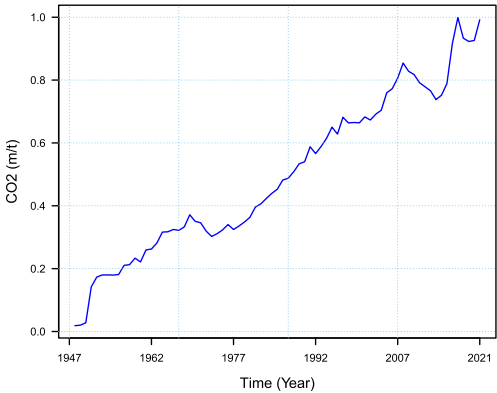


Fig. 3. Pakistan Carbon Dioxide Emission (metric tons per capita): The CO₂ emission time series plot from 1949 to 2021.

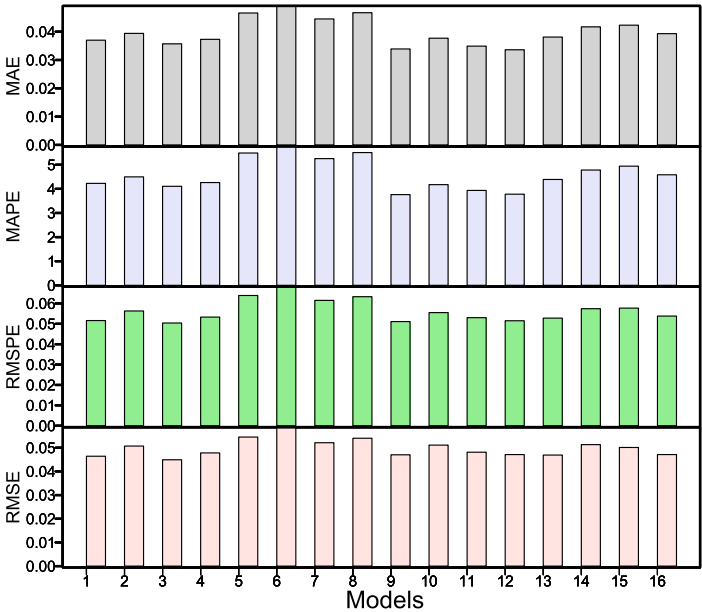


Fig. 4. Pakistan Carbon Dioxide Emission (metric tons per capita): The bar plot of the out-of-sample accuracy mean errors of CO₂ emission forecast for all sixteen hybrid combination models.

However, some studies have demonstrated that these tests may provide skewed and deceptive findings due to the possibility of structural discontinuities in the time series data [60]. We did not employ these tests since the residual series is practically stationary after filtering the CO₂ emission series for long run curve trend series.

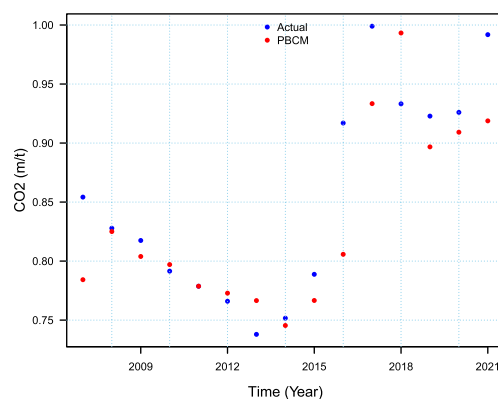


Fig. 5. Pakistan Carbon Dioxide Emission (metric tons per capita): Actual and forecasted CO₂ emission for only the final best model ¹M^a₀ for January 2007–December 2021.

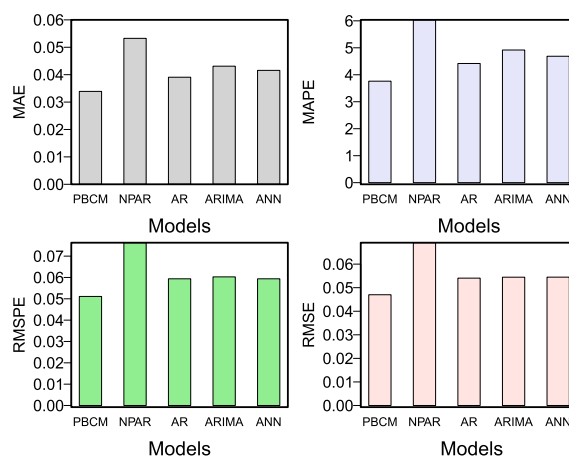


Fig. 6. Pakistan CO₂ emission: The bar plot comparing the proposed best model versus considered benchmark models: Out-of-sample CO₂ emission accuracy mean forecast errors.

Table 2

Pakistan Carbon Dioxide Emission (metric tons per capita): One-year-ahead out-of-sample accuracy mean errors of CO₂ emission forecast for all sixteen hybrid combination models.

S.No	Models	MAE	MAPE	RMSPE	RMSE
1	⁰ M ^a ₀	0.0370	4.2284	0.0516	0.0464
2	⁰ M ^b ₀	0.0394	4.4982	0.0563	0.0507
3	⁰ M ^c ₀	0.0357	4.1071	0.0504	0.0449
4	⁰ M ^d ₀	0.0373	4.2621	0.0533	0.0478
5	⁰ M ^a ₁	0.0466	5.4875	0.0639	0.0545
6	⁰ M ^b ₁	0.0491	5.7648	0.0682	0.0587
7	⁰ M ^c ₁	0.0445	5.2540	0.0615	0.0521
8	⁰ M ^d ₁	0.0467	5.5058	0.0633	0.0540
9	¹ M ^a ₀	0.0333	3.7621	0.0511	0.0470
10	¹ M ^b ₀	0.0377	4.1753	0.0555	0.0511
11	¹ M ^c ₀	0.0349	3.9356	0.0530	0.0481
12	¹ M ^d ₀	0.0336	3.7798	0.0515	0.0471
13	¹ M ^a ₁	0.0381	4.3899	0.0528	0.0469
14	¹ M ^b ₁	0.0417	4.7810	0.0574	0.0513
15	¹ M ^c ₁	0.0423	4.9423	0.0577	0.0501
16	¹ M ^d ₁	0.0393	4.5840	0.0538	0.0471

To produce a year-ahead forecast for CO₂ emissions using the proposed hybrid combination forecasting approach outlined in Section 2, the following actions had to be taken: First, the parametric and nonparametric regression models were employed to generate a long-run curve trend series (L_y) and a residual series (R_y). Second, four conventional time series models were applied to each subseries: the parametric autoregressive, the autoregressive moving average model, the nonparametric autoregressive model, and

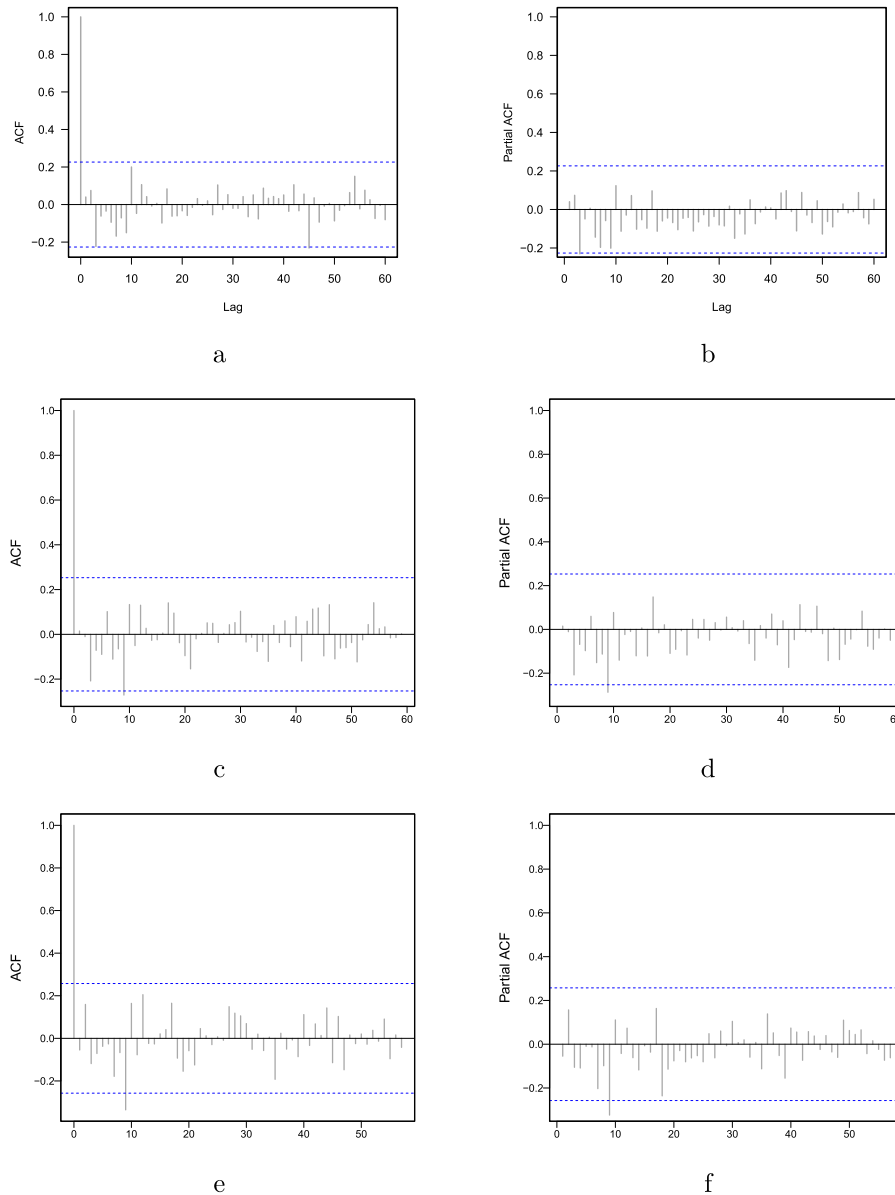


Fig. 7. Pakistan Carbon Dioxide Emission (metric tons per capita): The Autocorrelation plot (left), and the partial autocorrelation plot (right) for the three final best hybrid combination models ($^1M^{a_0}$, $^1M^{d_0}$, and $^1M^{c_0}$), respectively.

the neural network autoregressive model. The model parameters were then determined, and the one-year prediction was derived using the expanding window approach. Thus, Equation (17) produced the final one-year-ahead CO₂ emission forecasts. Finally, combining the long-run curve trend series forecast and the residual series forecast leads to sixteen possible combinations for comparison purposes ($4^2 = 16$). To evaluate and compare the performance of these sixteen hybrid combination models using three criteria: a) accuracy mean errors, b) statistical tests, and c) graphical presentation. The one-year ahead out-of-sample accuracy mean errors (MAE, MAPE, RMSE, RMSPE) for all sixteen hybrid combination models are listed in Table 2. It is confirmed from this table that the best hybrid combination forecasting model within all the sixteen hybrid combination models is $^1M^{a_0}$, which obtained 0.0333, 3.7621, 0.0511, and 0.0470 for MAE, MAPE, RMSPE, and RMSE, respectively. It can be observed that the MAE, MAPE, RMSPE, and RMSE are the lowest values compared to all other combinations. In the same way, the model $^1M^{d_0}$ produced the second-best result compared to all other used combination models with 0.0336, 3.7798, 0.0515, and 0.0471 for MAE, MAPE, RMSPE, and RMSE, accordingly. However, the $^1M^{c_0}$ model and $^0M^{c_0}$ model showed the third and fourth best results (MAE = 0.0349, 0.0357; MAPE = 3.9356, 4.1071; RMSPE = 0.0504, 0.0530; RMSE = 0.0481, .0449). Once the best hybrid combination model was obtained using the accuracy mean errors, check their consistency using a statistical test. To do this, the DM test was carried out for each pair of models to confirm the superiority of the findings stated in Table 2. The DM test findings (emphasis p-values) are shown in Table 3. Each

Table 3

Pakistan Carbon Dioxide Emission (metric tons per capita): Results (p-values) of the DM test for all combination models within the proposed filtering and ensemble technique.

Models	${}^0M^{a_0}$	${}^0M^{b_0}$	${}^0M^{c_0}$	${}^0M^{d_0}$	${}^0M^{a_1}$	${}^0M^{b_1}$	${}^0M^{c_1}$	${}^0M^{d_1}$	${}^1M^{a_0}$	${}^1M^{b_0}$	${}^1M^{c_0}$	${}^1M^{d_0}$	${}^1M^{a_1}$	${}^1M^{b_1}$	${}^1M^{c_1}$	${}^1M^{d_1}$
${}^0M^{a_0}$	0	0.8799	0.2161	0.6342	0.7093	0.8895	0.8425	0.5554	0.9552	0.9639	0.8749	0.9363	0.5684	0.9111	0.6384	0.5448
${}^0M^{b_0}$	0.1201	0	0.1375	0.3406	0.1494	0.7486	0.4599	0.3255	0.7823	0.9585	0.5921	0.6845	0.2588	0.5443	0.3647	0.3411
${}^0M^{c_0}$	0.7839	0.8625	0	0.8207	0.8465	0.8757	0.9696	0.7207	0.9637	0.9582	0.9361	0.9841	0.7284	0.9036	0.7784	0.6619
${}^0M^{d_0}$	0.3658	0.6594	0.1793	0	0.4127	0.6825	0.7534	0.3978	0.8216	0.8712	0.7419	0.9	0.4013	0.7224	0.5472	0.387
${}^0M^{a_1}$	0.2907	0.8506	0.1535	0.5873	0	0.874	0.8214	0.5196	0.9695	0.9685	0.893	0.9459	0.5107	0.8506	0.5932	0.5141
${}^0M^{b_1}$	0.1105	0.2514	0.1243	0.3175	0.126	0	0.4241	0.3017	0.7685	0.9714	0.5577	0.6576	0.2475	0.4777	0.3447	0.3239
${}^0M^{c_1}$	0.1575	0.5401	0.0304	0.2466	0.1786	0.5759	0	0.1313	0.7927	0.8564	0.6703	0.8809	0.2569	0.5663	0.3155	0.2783
${}^0M^{d_1}$	0.4446	0.6745	0.2793	0.6022	0.4804	0.6983	0.8687	0	0.8593	0.8831	0.8011	0.9631	0.49	0.708	0.595	0.4987
${}^1M^{a_0}$	0.0448	0.2177	0.0363	0.1784	0.0305	0.2315	0.2073	0.1407	0	0.8763	0.1848	0.4459	0.1564	0.3065	0.2226	0.2104
${}^1M^{b_0}$	0.0361	0.0415	0.0418	0.1288	0.0315	0.0286	0.1436	0.1169	0.1237	0	0.1183	0.2477	0.1043	0.1495	0.1574	0.1569
${}^1M^{c_0}$	0.1251	0.4079	0.0639	0.2581	0.107	0.4423	0.3297	0.1989	0.8152	0.8817	0	0.7272	0.2439	0.444	0.3074	0.2801
${}^1M^{d_0}$	0.0637	0.3155	0.0159	0.1	0.0541	0.3424	0.1191	0.0369	0.5541	0.7523	0.2728	0	0.1467	0.3461	0.1797	0.1556
${}^1M^{a_1}$	0.4316	0.7412	0.2716	0.5987	0.4893	0.7525	0.7431	0.51	0.8436	0.8957	0.7561	0.8533	0	0.8955	0.6571	0.5113
${}^1M^{b_1}$	0.0889	0.4557	0.0964	0.2776	0.1494	0.5223	0.4337	0.292	0.6935	0.8505	0.556	0.6539	0.1045	0	0.2888	0.2769
${}^1M^{c_1}$	0.3616	0.6353	0.2216	0.4528	0.4068	0.6553	0.6845	0.405	0.7774	0.8426	0.6926	0.8203	0.3429	0.7112	0	0.3272
${}^1M^{d_1}$	0.4552	0.6589	0.3381	0.613	0.4859	0.6761	0.7217	0.5013	0.7896	0.8431	0.7199	0.8444	0.4887	0.7231	0.6728	0

Table 4

Pakistan CO₂ emission Data: Comparison of the proposed best model versus the considered benchmark models: Out-of-sample CO₂ emission accuracy mean forecast errors.

S.No	Models	MAE	MAPE	RMSPE	RMSE
1	${}^1M^{a_0}$	0.0333	3.7621	0.0511	0.0470
2	NPAR	0.0533	6.0354	0.0763	0.0691
3	AR	0.0391	4.4181	0.0594	0.0540
4	ARMA	0.0431	4.9178	0.0603	0.0544
5	NNA	0.0415	4.6803	0.0593	0.0544

item in the table represents the p-value of a hypothesis system. The null hypothesis assumes no difference in the model's accuracy in the column/row vs. the alternative hypothesis that the model in the column is more accurate than the model in the row. This table shows that, among all sixteen hybrid combination models, the ${}^1M^{a_0}$, ${}^1M^{d_0}$, and ${}^1M^{c_0}$ models at a 10% significance level are statically better than the rest, except when comparing them to ${}^0M^{c_0}$. Finally, to check the graphical performance of the best models, as mentioned previously, this work presents the bar plots for the accuracy of mean errors. For instance, in Fig. 4, the MAE, the MAPE, the RMSPE, and the RMSE are plotted. For this figure, can seen that the ${}^1M^{a_0}$, ${}^1M^{d_0}$, ${}^1M^{c_0}$, and ${}^0M^{c_0}$ show minimum accuracy mean errors as compare the rest of all models. In addition, forecasted values from the best hybrid combination model (${}^1M^{a_0}$) among all sixteen combination hybrid models are plotted in Fig. 5. This plot shows that the predicted values follow the observed values of CO₂ emission very well. On the other hand, Fig. 7 depicts the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the residuals for the final best hybrid combination model (${}^1M^{a_0}$). The series no longer shows a notable autocorrelation structure in both figures, indicating that it has been whitened and is satisfactory. Thus, it is concluded from these findings (mean errors, statistical test, and graphical analysis) that the proposed hybrid combination forecasting technique is highly accurate and efficient in forecasting CO₂ emissions.

4. Discussion

In this research work, the ${}^1M^{a_0}$ model was shown to be the most accurate and efficient for CO₂ emission forecasting in Pakistan based on the findings of the descriptive statistics, statistical test, and graphical analysis. It is worth mentioning here that the mean accuracy errors (RMSPE, MAPE, MAE, and RMSE) reported in this study were lower than those of the benchmark models, which included the nonparametric autoregressive (NPAR), autoregressive (AR), autoregressive moving integrated average (ARIMA), and neural network autoregression models. Table 4 presents a numerical comparison of the proposed model with the four benchmark models, while Fig. 6 gives a visual comparison. As both presentations prove, this work's proposed optimal combination model (${}^1M^{a_0}$) produced much lower accuracy mean errors than the Benchmark models. For example, the final best hybrid combination model obtained mean accuracy errors MAE = 0.0333, MAPE = 3.7621, RMSPE = 0.0511, and RMSE = 0.0470, all lower than the benchmark models. To summarize, based on all of these findings and the preceding section, the accuracy of the best hybrid combination (${}^1M^{a_0}$) model is relatively high and efficient compared to all sixteen hybrid combinations and all baseline models.

Once the best model ${}^1M^{a_0}$ was assessed by descriptive statistics, a statistical test, graphic evaluation, and baseline model comparison, this work proceeded to future forecasting with the superior model. The authors used the ${}^1M^{a_0}$ for the CO₂ emission in Pakistan and forecast from 2022 to 2030 (using an expanding window forecast). The forecasted CO₂ emission values are the following: 2022 (1.009671), 2023 (1.024739), 2024 (1.039807), 2025 (1.054875), 2026 (1.069943), 2027 (1.085011), 2028 (1.100079), 2029 (1.115147), 2030 (1.130215), respectively. However, As per UNDP, Pakistan wants to establish a cumulative aggressive conditional

objective of a 50% reduction in predicted emissions by 2030, with a 15% reduction coming from domestic resources and a 35% reduction contingent on foreign grant funding. As a result, our findings provide policymakers with valuable information to direct future resource allocation and mitigation efforts. Furthermore, the forecasting exercise will aid in understanding the spread and, as a result, the risk.

In addition, based on the proposed hybrid combination forecasting technique, Pakistan's CO₂ emissions will be 1.130215 metric tonnes per capita by 2030. It implies that the growing emission trend is a worrisome and unmistakable signal that novel strategies to limit this tendency are required. The ever-increasing emission trend is an alarming warning that action must be taken to halt this tendency. To do this, the authors will suggest the Pakistani government adopt the following measures:

- To charge the carbon footprint of enterprises and facilities per tonne and limit the amount of electricity generated from hydro-electricity, wind power, and many other zero-carbon sources.
- To work on population reduction and encourage afforestation in heavily populated areas instead of wooded regions.
- Encourage enterprises, organizations, groups, and people to employ clean technology, create clean and creative zero- or low-carbon technologies, and support research and other research.
- The government of Pakistan lacks sufficient engagement and support on a global platform to address the issue of climate change, and the level of commitment is unclear. Pakistan needs proactive engagement with international climate forums to secure funds and strengthen partnerships with the global community to tackle this global problem.
- Pakistan must engage dedicated and robust institutions to collaborate and initiate coordination efforts among various stakeholders, scientific organizations, civil society, and the private sector to implement, monitor policies, and streamline decision-making in impactful and coherent climate change strategies.
- Pakistan's current policy on climate change focuses only on adaptation measures. It should be expanded to other contributing factors in climate change, including urbanization, unsustainable practices in agriculture sectors, and emissions due to industries. It needs a comprehensive policy on promoting sustainable land use, responsible agricultural practices, urban planning, prioritization of green infrastructure, and strict regulations on industrial emissions.

5. Conclusion and future directions

Currently, the CO₂ emissions continue to rise globally despite efforts to combat climate change. Energy industry emissions are a pressing global issue, causing devastating impacts. Hence, it is vital to accurately and efficiently forecast CO₂ emissions. Thus, this study has comprehensively analyzed predicting CO₂ emissions by comparing various hybrid combinations of regression and time series methods to explore the CO₂ emissions in Pakistan and, first, divided the yearly time series of CO₂ emissions into the long-run curve trend series and the residual subseries. The long-run curve trend subseries were modeled using parametric and nonparametric regression methods, while various standard time series models were used to forecast the residual subseries. However, the forecasts of each subseries have been combined to obtain the final estimates of CO₂ emissions. This work used four different accuracy mean errors, a statistical test, and a graphical analysis as performance measures to evaluate the proposed hybrid forecasting technique. The findings confirmed that the proposed hybrid combination forecasting technique is highly accurate and efficient in forecasting CO₂ emissions. Likewise, according to the proposed final optimal hybrid combination forecasting model, Pakistan's per capita CO₂ emissions will be 1.130215 metric tons in 2030. Pakistan's escalating emission trend signals that creative solutions must be implemented to curb it. Thus, Pakistan's growing emission pattern signals that creative actions are required to mitigate its impact. The government must tax enterprises' carbon footprints per ton, regulate zero-carbon power generation, and reduce population. Advocacy efforts will center on afforestation in densely populated regions, clean technology adoption, breakthrough low- or zero-carbon technology development, and research financing.

However, the study's main weakness is that it only incorporates CO₂ emission data and does not include additional exogenous components such as power pricing, temperature, wind speed, natural gas prices, and so on, which might improve CO₂ forecasting accuracy. In contrast, the current study employed solely data from Pakistan. Consequently, it may be used in other countries to assess the utility of the suggested hybrid combination modeling and forecasting method. Furthermore, while this study only employed parametric and nonparametric univariate models, machine learning techniques like deep learning and artificial neural networks might be explored within the proposed forecasting framework. On the other hand, it may also be expanded and used for other datasets, such as electrical power systems [61,62], time-tracking control systems [63,64], and air pollution [65].

Funding

This research received no external funding.

CRediT authorship contribution statement

Hasnain Iftikhar: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Conceptualization. **Murad Khan:** Writing – review & editing, Visualization, Investigation, Formal analysis, Data curation. **Justyna Żywiłłek:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Mehak Khan:** Writing – review & editing, Validation, Investigation. **Javier Linkolk López-Gonzales:** Writing – review & editing, Supervision, Project administration, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this study is openly available on the World Bank website (<https://databank.worldbank.org>).

References

- [1] Y. Cui, W. Su, Y. Xing, L. Hao, Y. Sun, Y. Cai, Experimental and simulation evaluation of CO₂/CO separation under different component ratios in blast furnace gas on zeolites, *Chem. Eng. J.* 472 (2023) 144579.
- [2] Z. Li, B. Gan, Z. Li, H. Zhang, D. Wang, Y. Zhang, Y. Wang, Kinetic mechanisms of methane hydrate replacement and carbon dioxide hydrate reorganization, *Chem. Eng. J.* 477 (2023) 146973.
- [3] H. Ritchie, P. Rosado, M. Roser, Greenhouse gas emissions. Our world is in data, <https://ourworldindata.org/CO2-and-greenhouse-gas-emissions>. (Accessed 22 June 2023), 2023.
- [4] M. Shang, J. Luo, The tapio decoupling principle and key strategies for changing factors of Chinese urban carbon footprint based on cloud computing, *Int. J. Environ. Res. Public Health* 18 (4) (2021) 2101.
- [5] A. Xu, M. Song, Y. Wu, Y. Luo, Y. Zhu, K. Qiu, Effects of new urbanization on China's carbon emissions: a quasi-natural experiment based on the improved PSM-DID model, *Technol. Forecast. Soc. Change* 200 (2024) 123164.
- [6] Z. Liu, Z. Deng, S. Davis, P. Ciaia, Monitoring global carbon emissions in 2022, *Nat. Rev. Earth Environ.* 4 (4) (2023) 205–206.
- [7] J.J. Chen, H.C. Wang, W.X. Yin, Y.Q. Wang, J.Q. Lv, A.J. Wang, Deciphering Carbon Emissions in Urban Sewer Networks: Bridging Urban Sewer Networks with City-Wide Environmental Dynamics, *Water Research*, 2024, p. 121576.
- [8] P. Friedlingstein, M.W. Jones, M. O'Sullivan, R.M. Andrew, D.C. Bakker, J. Hauck, J. Zeng, Global carbon budget 2021, *Earth Syst. Sci. Data* 14 (4) (2022) 1917–2005.
- [9] R. Zhang, L. Yin, J. Jia, Y. Yin, Application of ATS-GWIFBM operator based on improved time entropy in green building projects, *Adv. Civ. Eng.* 2019 (2019).
- [10] Defying Expectations, CO₂ emissions from global fossil fuel combustion are set to grow in 2022 by only a fraction of last year's big increase (IEA, 2022), <https://www.iea.org/news/defying-expectations-CO2-emissions-from-global-fossil-fuel-combustion-are-set-to-grow-in-2022-by-only-a-fraction-of-last-year-s-big-increase>.
- [11] S.J. Davis, Z. Liu, Z. Deng, B. Zhu, P. Ke, T. Sun, P. Ciaia, Emissions rebound from the COVID-19 pandemic, *Nat. Clim. Change* 12 (5) (2022) 412–414.
- [12] R.B. Jackson, P. Friedlingstein, C. Le Quéré, S. Abernethy, R.M. Andrew, J.G. Canadell, G.P. Peters, Global fossil carbon emissions rebound near pre-COVID-19 levels, *Environ. Res. Lett.* 17 (3) (2022) 031001.
- [13] L.A. Yarzabal, L.M.B. Salazar, R.A. Batista-García, Climate change, melting cryosphere and frozen pathogens: should we worry...?, *Environ. Sustain.* 4 (3) (2021) 489–501.
- [14] C. Zheng, H. Chen, Revisiting the linkage between financial inclusion and energy productivity: technology implications for climate change, *Sustain. Energy Technol. Assess.* 57 (2023) 103275.
- [15] K. Shang, L. Xu, X. Liu, Z. Yin, Z. Liu, X. Li, W. Zheng, Study of urban heat island effect in Hangzhou Metropolitan area based on SW-TES algorithm and image dichotomous model, *SAGE Open* 13 (4) (2023), <https://doi.org/10.1177/21582440231208851>.
- [16] L. Zhu, Q. Xiong, Key influencing factor and future scenario simulation of China's CO₂ emissions from road freight transportation, *Sustain. Prod. Consump.* 37 (2023) 11–25.
- [17] C. Sun, J. Chen, B. He, J. Liu, Digitalization and carbon emission reduction technology R&D in a Stackelberg model, *Appl. Econ. Lett.* (2024) 1–6.
- [18] J. Fan, X. Zhou, Optimization of a hybrid solar/wind/storage system with bio-generator for a household by emerging metaheuristic optimization algorithm, *J. Energy Storage* 73 (2023) 108967.
- [19] I. Yousuf, A.R. Ghumman, H.N. Hashmi, M.A. Kamal, Carbon emissions from power sector in Pakistan and opportunities to mitigate those, *Renew. Sustain. Energy Rev.* 34 (2014) 71–77.
- [20] Z. Yin, Z. Liu, X. Liu, W. Zheng, L. Yin, Urban heat islands and their effects on thermal comfort in the US: New York and New Jersey, *Ecol. Indic.* 154 (2023) 110765.
- [21] B.K. Sovacool, The importance of comprehensiveness in renewable electricity and energy-efficiency policy, *Energy Policy* 37 (4) (2009) 1529–1541.
- [22] Z. Liu, P. Jiang, J. Wang, Z. Du, X. Niu, L. Zhang, Hospitality order cancellation prediction from a profit-driven perspective, *Int. J. Contemp. Hosp. Manag.* 35 (6) (2023) 2084–2112.
- [23] United Nations Pakistan Available online, <https://pakistan.un.org/en/217455-supporting-millions-impacted-pakistan-flooding>. (Accessed 23 June 2023).
- [24] United Nations Framework Convention on climate change available online, <https://unfccc.int/sites/default/files/NDC/2022-06/Pakistan>. (Accessed 23 June 2023).
- [25] H. Iftikhar, M. Khan, J.E. Turpo-Chaparro, P.C. Rodrigues, J.L. López-Gonzales, Forecasting stock prices using a novel filtering-combination technique: application to the Pakistan stock exchange, *AIMS Math.* 9 (2) (2024) 3264–3288, <https://doi.org/10.3934/math.2024159>.
- [26] M.O. Faruque, M.A.J. Rabby, M.A. Hossain, M.R. Islam, M.M.U. Rashid, S.M. Mueen, A comparative analysis to forecast carbon dioxide emissions, *Energy Rep.* 8 (2022) 8046–8060.
- [27] L. Kong, L. Wang, F. Li, J. Li, Y. Wang, Z. Cai, G. Wang, Life cycle-oriented low-carbon product design based on the constraint satisfaction problem, *Energy Convers. Manag.* 286 (2023) 117069.
- [28] Z. Liu, Z. Xu, X. Zhu, L. Yin, Z. Yin, X. Li, W. Zheng, Calculation of carbon emissions in wastewater treatment and its neutralization measures: a review, *Sci. Total Environ.* 169356 (2023).
- [29] K. Tawiah, M. Daniyal, M. Qureshi, Pakistan CO₂ emission modelling and forecasting: a linear and nonlinear time series approach, *J. Environ. Public Health* (2023).
- [30] P. Jiang, Z. Liu, M.Z. Abedin, J. Wang, W. Yang, Q. Dong, Profit-driven weighted classifier with interpretable ability for customer churn prediction, *Omega* 125 (2024) 103034.
- [31] United Nation Development Program Available online, <https://climatepromise.undp.org/what-we-do/where-we-work/pakistan>. (Accessed 24 June 2023).
- [32] Z. Liu, P. Jiang, K.W. De Bock, J. Wang, L. Zhang, X. Niu, Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction, *Technol. Forecast. Soc. Change* 198 (2024) 122945.
- [33] M. Ahmed, C. Shuai, M. Ahmed, Influencing factors of carbon emissions and their trends in China and India: a machine learning method, *Environ. Sci. Pollut. Res.* 29 (32) (2022) 48424–48437.

- [34] J. Luo, W. Zhuo, B. Xu, A deep neural network-based assistive decision method for financial risk prediction in carbon trading market, *J. Circuits Syst. Comput.* 2450153 (2023).
- [35] J. Luo, W. Zhuo, S. Liu, B. Xu, The optimization of carbon emission prediction in low carbon energy economy under big data, *IEEE Access* (2024).
- [36] C. Ziyuan, Y. Yibo, Z. Simayi, Y. Shengtian, M. Abulimiti, W. Yuqing, Carbon emissions index decomposition and carbon emissions prediction in Xinjiang from the perspective of population-related factors, based on the combination of STIRPAT model and neural network, *Environ. Sci. Pollut. Res.* (2022) 1–16.
- [37] Z. Liu, Y. Zhao, Q. Wang, H. Xing, J. Sun, Modeling and assessment of carbon emissions in additive-subtractive integrated hybrid manufacturing based on energy and material analysis, *Int. J. Precis. Eng. Manuf.-Green Technol.* (2024) 1–15.
- [38] A. Malik, E. Hussain, S. Baig, M.F. Khokhar, Forecasting CO₂ emissions from energy consumption in Pakistan under different scenarios: the China–Pakistan economic corridor. Greenhouse gases, *Sci. Technol.* 10 (2) (2020) 380–389.
- [39] S. Kumari, S.K. Singh, Machine learning-based time series models for effective CO₂ emission prediction in India, *Environ. Sci. Pollut. Res.* (2022) 1–16.
- [40] H.U. Rehman, R. Rafique, M. Nasir, M.A.Z. Chudhery, Forecasting CO₂ emissions from energy, manufacturing and transport sectors in Pakistan: statistical vs. machine learning methods, *Mach. Learn. Methods* (2017) (November 28, 2017).
- [41] Z. Hajirahimi, M. Khashei, Hybridization of hybrid structures for time series forecasting: a review, *Artif. Intell. Rev.* 56 (2) (2023) 1201–1261.
- [42] J.F. de Oliveira, E.G. Silva, P.S. de Mattos Neto, A hybrid system based on dynamic selection for time series forecasting, *IEEE Trans. Neural Netw. Learn. Syst.* 33 (8) (2021) 3251–3263.
- [43] N. Carbo-Bustanza, H. Iftikhar, M. Belmonte, R.J. Cabello-Torres, A.R.H. De La Cruz, J.L. López-Gonzales, Short-term forecasting of Ozone concentration in metropolitan Lima using hybrid combinations of time series models, *Appl. Sci.* 13 (18) (2023) 10514.
- [44] H. Iftikhar, A. Zafar, J.E. Turpo-Chaparro, P. Canas Rodrigues, J.L. López-Gonzales, Forecasting day-ahead Brent crude oil prices using hybrid combinations of time series models, *Mathematics* 11 (16) (2023) 3548.
- [45] J.F. de Oliveira, E.G. Silva, P.S. de Mattos Neto, A hybrid system based on dynamic selection for time series forecasting, *IEEE Trans. Neural Netw. Learn. Syst.* 33 (8) (2021) 3251–3263.
- [46] S. Panigrahi, H.S. Behera, A hybrid ETS–ANN model for time series forecasting, *Eng. Appl. Artif. Intell.* 66 (2017) 49–59.
- [47] G.P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing* 50 (2003) 159–175.
- [48] M. Aamir, M.A. Bhatti, S.U. Bazai, S. Marjan, A.M. Mirza, A. Wahid, U.A. Bhatti, Predicting the environmental change of carbon emission patterns in South Asia: a deep learning approach using BiLSTM, *Atmosphere* 13 (12) (2022) 2011.
- [49] M.O. Faruque, M.A.J. Rabby, M.A. Hossain, M.R. Islam, M.M.U. Rashid, S.M. Mueyen, A comparative analysis to forecast carbon dioxide emissions, *Energy Rep.* 8 (2022) 8046–8060.
- [50] B. Mousavi, N.S.A. Lopez, J.B.M. Biona, A.S.F. Chiu, M. Blesl, Driving forces of Iran's CO₂ emissions from energy consumption: an LMDI decomposition approach, *Appl. Energy* 206 (2017) 804–814.
- [51] I. Shah, H. Iftikhar, S. Ali, Modeling and forecasting medium-term electricity consumption using component estimation technique, *Forecasting* 2 (2020) 163–179.
- [52] H. Iftikhar, J.E. Turpo-Chaparro, P. Canas Rodrigues, J.L. López-Gonzales, Day-ahead electricity demand forecasting using a novel decomposition combination method, *Energies* 16 (18) (2023) 6675.
- [53] I. Shah, H. Iftikhar, S. Ali, Modeling and forecasting electricity demand and prices: a comparison of alternative approaches, *J. Math.* 2022 (2022) 3581037.
- [54] H. Iftikhar, M. Khan, M.S. Khan, M. Khan, Short-term forecasting of monkeypox cases using a novel filtering and combining technique, *Diagnostics* 13 (11) (2023).
- [55] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, *J. Bus. Econ. Stat.* 20 (2002) 134–144.
- [56] H. Iftikhar, M. Khan, Z. Khan, F. Khan, H.M. Alshanbari, Z. Ahmad, A comparative analysis of machine learning models: a case study in predicting chronic kidney disease, *Sustainability* 15 (3) (2023) 2754.
- [57] H.M. Alshanbari, H. Iftikhar, F. Khan, M. Rind, Z. Ahmad, A.A.A.H. El-Bagoury, On the implementation of the artificial neural network approach for forecasting different healthcare events, *Diagnostics* 13 (7) (2023) 1310.
- [58] H. Iftikhar, N. Khan, M.A. Raza, G. Abbas, M. Khan, M. Aoudia, A. Emara, Electricity theft detection in smart grid using machine learning, *Front. Energy Res.* 12 (2024) 1383090.
- [59] H. Iftikhar, N. Bibi, P. Canas Rodrigues, J.L. López-Gonzales, Multiple novel decomposition techniques for time series forecasting: application to monthly forecasting of electricity consumption in Pakistan, *Energies* 16 (6) (2023) 2579.
- [60] M. Shahbaz, H.H. Lean, The dynamics of electricity consumption and economic growth: a revisit study of their causality in Pakistan, *Energy* 39 (2012) 146–153.
- [61] C. Guo, J. Hu, Time base generator based practical predefined-time stabilization of high-order systems with unknown disturbance, *IEEE Trans. Circuits Syst. II, Express Briefs* (2023).
- [62] H. Iftikhar, J.E. Turpo-Chaparro, P. Canas Rodrigues, J.L. López-Gonzales, Forecasting day-ahead electricity prices for the Italian electricity market using a new decomposition–combination technique, *Energies* 16 (18) (2023) 6669.
- [63] B. Chen, J. Hu, B.K. Ghosh, Finite-time tracking control of heterogeneous multi-AUV systems with partial measurements and intermittent communication, *Sci. China Inf. Sci.* 67 (5) (2024) 152202.
- [64] K. Liu, G. Nie, S. Jiao, B. Gao, H. Ma, J. Fu, G. Wu, Research on fault diagnosis method of vehicle cable terminal based on time series segmentation for graph neural network model, *Measurement* (2024) 114999.
- [65] F. Quispe, E. Salcedo, H. Iftikhar, A. Zafar, K.J.E. Khan, J.E. Turpo-Chaparro, P. Canas Rodrigues, J.L. López-Gonzales, Multi-step ahead ozone level forecasting using a component-based technique: a case study in Lima, Peru, *AIMS Environ. Sci.* 11 (3) (2024) 401–425.