Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Modeling electricity generation and consumption in cameroon

Blaise Fombuwing* [1], **Neyre Tekbıyık Ersoy** [1]

[1] *Cyprus International University, Energy Systems Engineering, 99258, Nicosia, Mersin 10, Turkey, fombuwingb@gmail.com; nersoy@ciu.edu.tr*

Cite this study: Fombuwing, B., & Tekbıyık-Ersoy, N. (2024). Modeling Electricity Generation and Consumption in Cameroon. Turkish Journal of Engineering, 8 (4), 593-602.

https://doi.org/10.31127/tuje.1434866

Keywords
Cameroon
Correlation Analysis
Electricity Consumption
Electricity Generation
Regression Analysis

Research/Review Article

Received: 20.02.2024
Revised: 29.03.2024
Revised: 03.07.2024
Accepted: 17.07.2024
Published: 31.10.2024

Abstract

Currently, there is a significant gap between electricity generation and consumption in Cameroon. Research has shown that electricity consumption in the country is estimated to increase by 965.7 GWh in five years, from 2020 to 2024 due to demographic and economic growth. Hence, this study aims to find methods that can be useful in developing strategies to balance the energy supply and demand in the country. This is done by developing models that can predict future electrical power consumption and generation. Correlation analysis and regression analysis were performed by using data obtained from various databases, and related models were developed accordingly. The model parameters were carbon dioxide emissions, electricity consumption per capita, final consumption expenditures, electricity installed capacity, fossil fuel installed capacity, labor force, and GDP. The models' results demonstrated excellent performance coefficients with RMSE of 0.17041, 0.23893, 0.27571, and 0.2465 for hydroelectricity generation, fossil fuel electricity generation, net electricity generation, and net electricity consumption respectively. Also, hydroelectricity generation, net electricity generation, and net electricity consumption models showed very good RRMSE performance indicating that the models can make predictions with only 4.26%, 5.26%, and 5.77% deviation from the mean values of hydroelectricity generation, net electricity generation, and net electricity consumption, respectively.

1. Introduction

Electricity has a significant contribution to the Cameroon's economy. It is used for lighting, heating, cooling, device operations, machinery, and refrigeration purposes. Due to increasing electricity demand, climate change, and the poor state of power transmission and distribution networks, the country has had more power outages over the years, causing major damage to households and industries. Each year, the absence of electricity in industry slows economic growth by 2% and results in a loss of more than USD 12 million in foreign exchange revenues [1, 22]. This is a result of the imbalance between power generation and consumption, and mainly the losses on the system. Consequently, generators and kerosene are increasingly being used to provide electricity and lighting [23].

Understanding the driving forces behind electricity generation and consumption in Cameroon may help in providing long-term strategies necessary to manage electricity demand and ensure the continued supply of energy to the grid. While there has been much research on electricity consumption modeling in Cameroon [1 – 3], to the best of the knowledge of the author of this study, none have focused specifically on modeling electricity generation from different sources in Cameroon. It is critical to gain a thorough understanding of the generation sector in order to resolve the issue of power outages/shortages in Cameroon. Therefore, this study investigates the existing relationships between different parameters and electricity generation/consumption in Cameroon to formulate multiple linear regression models that can be used to predict electricity generation and consumption. Utilization of such models is expected to control power demand, control the amount of fuel required for thermal generation, and thus improve electrical energy efficiency in the country.

The republic of Cameroon endows 15 oil-fired power plants and 2 gas fired power plants (Kribi and Logbaba) [1]. There are also 3 hydropower plants in operation; Song Loulou (384MW), Edea (276.2MW), and Lagdo (72MW) [1]. Most of the electricity supply (9.149 GWh in 2019) came from hydro (57%), followed by gas (27%) and oil (15%), with other renewable energies accounting around 1% [4]. The country's electricity installed capacity in 2021 was around 1402 MW [5]. Furthermore, the country's electrification rate in 2020 was around 64.72%, with 94.03% and 24.98% of urban and rural populations having access to electricity, respectively [6]. Only 20% of the population has access to electricity networks [24], with a large majority of communities in rural areas not being connected [25]. Considering these, the government's new goal is to boost domestic generation capacity by 3500MW by 2030 and achieve 100% access by 2035 [4].

Section 2 presents a review of previous work on electricity consumption modeling in Cameroon and electricity generation modeling in other countries. It also contains the authors' viewpoints related to the studies. Section 3 explains the method used in this study and the data used in the analysis. Section 4 addresses the results obtained and their role in achieving the study objectives. Section 5 evaluates the performance of the developed models and compares them with those in the literature. Section 6 concludes the study.

2. Literature review

Many studies on electricity generation and consumption have been conducted around the World. Recent studies in Cameroon have focused on electricity consumption forecasts as well as the existing relationship between electricity consumption and economic growth. Dieudonne, et al. [3] for instance aimed at determining the prediction of electricity consumption in Cameroon for the period 1975- 2019. The article examined three models which are the Vector error correction model (VECM), the Holt-Winters exponential smoothening (HES), and the Hybrid model (VECM-HES). The study parameters were population, GDP per habitant, electricity consumption per habitant, and the expenses on final consumption in households. The analysis resulted in the MAPE of 12.65%, 7.32%, and 1.59% with the RMSE of 395.4, 229, and 6.74 for the VECM, HES, and VECM-HES model respectively. Based on their respective MAPEs, the author concluded that the VECM-HES model projected electricity consumption to rise from 7169.031GWh in 2020 to 8134.772GWh in 2024. Also, Guefano, et al. [1] examined electricity consumption in the residential sector by using the grey (GM(1,1)) model, vector autoregressive (VAR(p)) econometric model, and the hybrid model (GM(1,1)-VAR(p)) over the period 1994-2017. The parameters included the GDP, GDP per capita, population, number of subscribers, and number of households. The study indicated a MAPE of 3.96%, 7.73%, and 1.629% for the GM (1,1), VAR (p), and GM(1,1)-VAR(p) models respectively. Based on their respective MAPEs, the author concluded that the GM (1,1)-VAR(1) hybrid model was more accurate than the other models. Next, Guefano, et al. [2] also focused on forecasting electricity demand in Cameroon's residential sector using multiple regression over the period 1994-2014. The authors used the same parameters as [1]. They observed a strong correlation of more than 96% between the variables. According to the results, residential electricity consumption was expected to increase from 1721GWh in 2014 to 2481GWh in 2020.

Other studies in Cameroon analyzed the relationship between energy consumption and economic growth. Tamba, et al., [7] researched electricity consumption and its relationship with economic growth from 1971 to 2013. They used the VAR model and the Granger causality test. Results demonstrated that the use of electricity and economic growth were not causally related. Also, Joel and Cyrille [8] examined the relationship between energy consumption, economic growth, and CO_2 emissions in Cameroon by using the autoregressive distributed lag model. According to the findings, there was no long-term relationship between CO_2 emissions and energy consumption. In the short term, however, fossil and electricity consumption positively affected emissions.

Although most Cameroonian researchers focused solely on the electricity consumption sector, other researchers from around the world also examined the power generation sector. For instance [9 – 11] aimed at comparing the regression model and the artificial neural network (ANN) model. Based on their finding they concluded that the ANN model estimated and predicted with more accuracy and reliability than the regression model. Although machine learning models are more accurate than regression models, their complexity makes them difficult for people with little background in the field to understand and use them. In order to use machine learning related modelling or forecasting, a graphical user interface that can be used by the user (who does not have modelling background) should be implemented and should be provided to the user. Thus, regression is preferred over machine learning in this study because it provides equations that can easily be used by people in the utility company or government due to their simplicity. These equations can also help the utility company or government during transition to adopting more sophisticated models and forecasting approaches.

As discussed above, the generation sector was disregarded in earlier studies on Cameroon's power modeling, which mostly focused on the consumption sector. However, power production modeling is the key to ensure future energy supply and maintain a balance between power demand and supply. Therefore, in addition to consumption, this study also considers generation modelling. Finally, this study proposes a novel method (step-by-step approach by Fombuwing (SSAF)) for selecting the variables that are to be used in the models.

3. Material and method

This study aims to formulate multiple linear regression models that can be used to predict electricity generation and consumption by using Microsoft Excel. In regression analysis, one variable is predicted from another based on the existing relationship between these variables. The predicted variable is generally referred to as the dependent (outcome) variable, whereas the other variables are called independent variables or predictors. A regression analysis with one independent variable is called a univariate regression, whereas a regression analysis with more than one independent variable is called a multivariate regression [12]. The following is a formula for multivariate regression analysis.

$$Y = B_0 + B_1 X_1 + \cdots . + B_k X_k + \varepsilon \quad (1)$$

Where Y = $\tilde{Y}$ + ε, with Y and $\tilde{Y}$ representing the real and predicted dependent variables respectively. ε stands for the error and B_0 is the constant. B_k and X_k represent the regression coefficient, and the value of the independent variable respectively.

3.1. Data collection

This study covers a 23-year period (1997-2019) due to data availability. The study used data from the U.S. Energy Information Administration [13], the World Bank development indicators database [6], and electricity consumption per capita data from previous studies [3]. The values of electricity consumption per capita data were shifted a unit backward during modeling due to its direct link with electricity consumption and generation. Overall, 22 variables were considered, of which 4 were dependent variables. The details of the data can be seen in Table 1. The renewable electricity generation (REG) variable was not considered for modeling because 99% of renewable electricity in Cameroon comes from hydropower (HG). Therefore, the hydroelectricity generation (HG), fossil fuels electricity generation (FFEG), net electricity generation (EG), and net electricity consumption (EC) dependent variables were the outcome variables in this study. A total of 17 independent variables that could impact electricity generation and consumption in Cameroon were selected. The aim was to examine the existing relationship between these variables and the outcome variables in order to select the best-fit predictors for the models.

Table 1. Specific details of the data used in this study.

Variables	Minimum	Maximum	Standard deviation
Pop (Population)	14344444	25876387	3575468.721
Labor force	6451990	11333454	1407864.624
Fcon expend (final consumption expenditure) {current US$}	8360145043	33618648566	8697283826
GDP (gross domestic product) {current US$}	10789457915	39973839065	10166222482
CO_2 emissions {kt}	4860	9590.00	1669.459
FFEI cap (fossil fuels electricity installed capacity) {million kW}	0.07	1.26	0.418
EC per capita (electricity consumption per capita) {kWh}	172	286	47.472
EI cap (electricity installed capacity) {million kW}	0.817	2.03	0.421
GDP/capita (GDP per capita) {current US$}	752.16	1604.21	324.336
DNG con (dry natural gas consumption) {BCF}	0	28	9.608
LPG con (liquid petroleum gas consumption) {Mb/d}	22	42	7.262
RE con (renewable energy consumption) {%}	76.79	86.31	3.361
Precipitation {mm}	1467.07	1718.41	67.579
HEI cap (hydroelectricity installed capacity) {million kW}	0.719	0.80	0.035
EP (electricity price) {USD/MWh}	113.45	160.68	18.409
Mean Temp (mean temperature) {Celsius}	24.57	25.21	0.177
REI cap (renewable electricity installed capacity) {million kW}	0.71	0.80	0.036
HG (hydroelectricity generation) {billion kWh}	3.06	5.34	0.621
REG (renewable electricity generation) {billion kWh}	3.06	5.35	0.626
FFEG (fossil fuels electricity generation) {billion kWh}	0.03	3.36	1.167
EC (net electricity consumption) {billion kWh}	2.48	6.50	1.400
EG (net electricity generation) {billion kWh}	3.09	8.35	1.747

3.2. Analysis of the data

In this section, combo charts are drawn to reveal the certain trends in data used in this study. In Figure 1 and Figure 2, the bar graph shows the pattern of data on the primary axis while the line graph illustrates the data on the secondary axis.

Figure 1 shows an increase in liquid petroleum gas (LPG) consumption over the years, whereas dry natural gas (DNG) consumption began in 2005 and has shown a rising trend over time. This is because dry natural gas production in Cameroon also began in 2005 [13]. As of 2017, Cameroon had proven gas reserves of 4.77 trillion cubic feet, equivalent to 148.4 times its annual consumption [14]. This means that the country has about 148 years of natural gas left excluding unproven reserves. The secondary axis shows insignificant variation for the mean temperature and a slight drop in RE consumption over the years from 84.5% to 79.4% of the total final energy consumption in the country.

Possibly this reduction is due to the rise in fossil fuel energy consumption for electricity production and the use of vehicles.

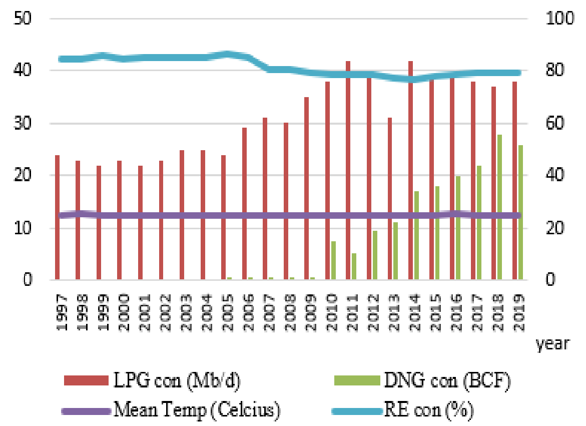

Figure1. Comparison of the LPG con, DNG cons, Mean temp, and RE con variables.

Figure 2 reveals that the HG and the REG rose over the years at almost the same rate, from 3 billion kWh each in 1997 to 5.88 billion kWh and 5.89 billion kWh respectively in 2020. This shows that more than 99% of the country's renewable electricity comes from hydro and the remaining 1% comes from other renewable energy sources. Also, the EC and the EG increased respectively from 2.48 billion kWh and 3.09 billion kWh in 1997 to 6.33 billion kWh and 8.29 billion kWh in 2020. This was due to an increase in power demand and fossil fuel electricity generation. The huge rise in the difference between total electricity generation and total electricity consumption over the years can also be observed from the figure. According to [4], ENEO's network loses approximately 30% of power in its grid due to both technical and commercial reasons. [3] estimated electricity consumption to rise by 965.7 GWh in five years, which is approximately 14% of Cameroon's annual electricity consumption. This means that by reducing the power losses from 30% to 15%, the government will save over 965.7 GWh of electrical energy annually. This will solve the problems of power cuts and the loss of more than 12 million USD in foreign exchange earnings. Moreover, since 2002, FFEG rose from 0.11 billion kWh to 2.84 billion kWh in 2020 due to a rise in fossil fuel electricity installation and dry natural gas production.

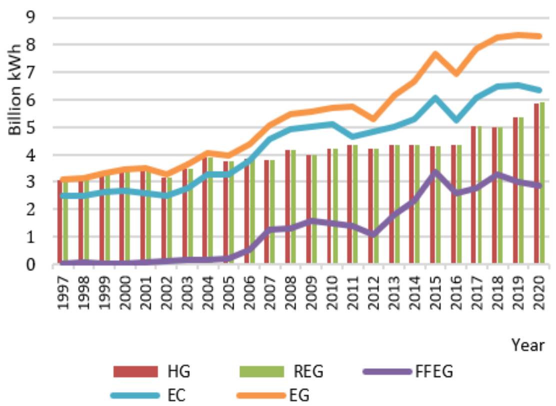

Figure 2. Comparison of the HG, REG, FFEG, EC, and EG variables.

3.3. Correlation analysis

In this study, correlation analysis was used to verify the existing relationship between all the variables in the dataset in Microsoft Excel. The strength of the linear relationship between two variables is indicated by the correlation coefficient (r). Its value ranges between -1 and 1, with -1 indicating a perfect negative relationship, 0 revealing no relationship, and 1 representing a perfect positive relationship between the variables [15]. Thus, the closer the correlation coefficient comes to +1 or -1, the stronger the correlation. A negative relationship signifies that as one variable increases, the other decreases while a positive relationship means that as one variable increases the other also increases and vice versa. The correlation coefficient is briefly interpreted in Table 2.

Table 2. Correlation coefficient and interpretation.

Correlation size (r)	Interpretation
0.9 to 1.0 / -0.9 to -1.0	Very high positive / negative correlation
0.7 to 0.9 / -0.7 to -0.9	High positive / negative correlation
0.5 to 0.7 / -0.5 to -0.7	Moderate positive / negative correlation
0.3 to 0.5 / -0.3 to -0.5	Low positive / negative correlation
0 to 0.3 / 0 to -0.3	Negligible correlation

Source: Makuka [15].

To develop models with high precision and predictability, variables with r > 0.8 or r < -0.8 were considered significant for the study's purposes. Those with r < 0.8 and r > -0.8 were considered insignificant and eliminated from the modeling datasets. A total of 12 variables were found to be strongly related to the outcome variables EG and FFEG. 11 variables were strongly related to the outcome variable EC, and 9 variables also had a strong relationship with the outcome variable HG. These variables were therefore significant and they were considered during the models' formulation. The significant variables obtained as a result of the correlation analysis are shown in Table 3.

Table 3. Significant data obtained from correlation analysis.

FFEG	HG	EC	EG
Population	Population	Population	Population
LPG con	LPG con	LPG con	LPG con
DNG con	DNG con	DNG con	DNG con
GDP	GDP	GDP	GDP
GDP per capita	GDP per capita	GDP per capita	GDP per capita
EC per capita	EC per capita	EC per capita	EC per capita
EI cap	Labor force	FFEI cap	EI cap
FFEI cap	CO_2 emissions	Labor force	FFEI cap
Labor force	Fcon expend	CO_2 emissions	Labor force
CO_2 emissions		RE con	CO_2 emissions
RE con		Fcon expend	RE con
Fcon expend			Fcon expend

3.4. Performance measurement criteria

Models were created based on performance criteria, such as; adjusted R square, the RMSE, and the RRMSE (CV). The p-value was also calculated during the analysis. In statistical analysis, the p-value indicates the significance of the parameters used in a model. The performance measurement criteria used in this study are as follows;

- R square: It is an indicator often used in statistics to estimate a model's performance, with values ranging from 0 to 1 [16]. The closer the value approaches 1, the better the model.
- Adjusted R Square: It is an indicator used to determine the goodness of fit of a linear model [17]. It ranges from 0 to 1. The closer the value is to 1, the better the model.
- RMSE (root mean square error): This is a measure commonly used to compare forecasting errors between models; the lower the RMSE value, the better the model's accuracy [16].
- RRMSE (relative root mean square error): This is an indicator calculated by dividing the RMSE by the average value of measured data [16]. The RRMSE is also often called the coefficient of variation (CV).

Using statistical significance ($p < 0.05$) in Excel, the test was conducted on all independent variables X and dependent variable Y to establish the regression equation ($\tilde{Y}$). Afterward, the R Square, the adjusted R Square, the RMSE, and the RRMSE were calculated in Excel to verify the model's goodness by using the formulas,

$$R^2 = \frac{\sum_{i=1}^{n}(\tilde{Y}_i - \bar{y})^2}{\sum_{i=1}^{n}(Y_i - \bar{y})^2} \tag{2}$$

$$\text{Adjusted } R^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k}\right) \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{\sum(Y_i - \tilde{Y}_i)^2}{n-k}} \tag{4}$$

$$\text{RRMSE} = \text{CV} = \frac{\text{RMSE}}{\bar{y}} \tag{5}$$

Where Y_i = the real value of the outcome, $\tilde{Y}_i$= the predicted outcome, $\bar{Y}$ = the mean of Y_i, n = the number of observations, and k = the number of coefficients in the model.

Table 4 interprets the performance of a model based on the RRMSE. This will be used in this study to verify the forecasting ability of the formulated models.

Table 4. Model accuracy and RRMSE.

RRMSE	Accuracy
RRMSE < 10%	Excellent
10% < RRMSE < 20%	Good
20% < RRMSE < 30%	Fair
RRMSE > 30%	Poor

Source: Despotovic, et al. [16].

3.5. Development of the models

In this section, the model predictors were selected from the significant independent variables following the step-by-step approach by Fombuwing (SSAF) developed for this study. The adjusted R Square was used in this section as the performance determination coefficient. The best fit models were developed following the SSAF described as follows;

Step1: As the first predictor, use the variable that has the highest correlation coefficient (correlation with the outcome variable). Then, using the first predictor and each significant variable in the outcome dataset, run a regression analysis on the outcome variable to find the second predictor. As the second predictor, choose the variable that causes the highest performance coefficient when used with the first predictor.

Step 2: To find the third predictor, run a regression analysis on the outcome variable using the two predictors from Step1, along with each significant variable in the outcome dataset. As the third predictor, choose the variable that causes the highest performance coefficient when used with the first two predictors.

Step 3: To identify the fourth predictor, perform a regression analysis on the outcome variable using the three predictors from Step 2 as well as each significant variable in the outcome dataset. Use the variable resulting the highest performance coefficient in relation to the outcome as the fourth predictor. More details about SSAF can be found in [18].

3.5.1. Developing the net electricity consumption model

The EC dataset consisted of 11 significant independent variables. The Fcon expend variable with the highest correlation coefficient of 0.9713 was considered as the first predictor for the net electricity consumption model. Other significant independent variables in the EC dataset were tested with the Fcon expend variable in different combinations using the regression analysis to obtain the best fit model following the SSAF.

Employing the step 1 approach, the EC-(Fcon expend-CO_2 emissions) showed the greatest performance coefficient and the CO_2 emissions variable was selected as the second independent variable. Applying the step 2 approach, the EC-(Fcon expend-CO_2 emissions-FFEI cap) with coefficient of 0.9603 showed the best performance and the FFEI cap variable was picked as the third independent variable. As determined by the performance coefficient of 0.969 by using step 3 approach, the EC-(Fcon expend-CO_2 emissions-FFEI cap-EC per capita) variables were the most suitable variables for the creation of electricity consumption model. As a result, the electricity consumption model was created as follows:

$$\tilde{Y} = -1.77848 + 0.0005X_1 - (0.8917)X_2 + 0.008124X_3 + (6.23 \times 10^{-11})X_4 \tag{6}$$

Where X_1 is the CO_2 emissions (kt) and X_2 represents the FFEI cap (million kW). X_3 is the EC per capita (kWh) and X_4 represent the Fcon expend (current US$).

3.5.2. Developing the hydroelectricity generation model

During the HG model formulation, the first predictor was selected as the labor force because it had the strongest correlation coefficient with the HG dependent variable. The other predictors were chosen by using the SSAF. Following step 1, the Fcon expend variable was selected as the second variable since HG-(labor force-Fcon expend) variables showed the greatest performance coefficient of 0.9191. Applying the step 2 approach, the EC per capita variable was chosen as the third variable due to having the greatest performance determination coefficient of 0.9209 with the labor force-Fcon expend variables. Following step 3, the CO_2 emission variable showed the greatest performance determination coefficient of 0.923468 with the labor force-Fcon expend-EC per capita variables and was selected as the fourth variable. Therefore, the hydroelectricity model was created as follows:

$$\tilde{Y}=1.3840 - (0.0001)X_1 + (3.75 \text{ x } 10^{-7})X_2 - (0.00422)X_3 + (4.19 \text{ x } 10^{-11})X_4 \quad (7)$$

Where X_1 is the CO_2 emissions (kt) and X_2 represents the labor force. X_3 is the EC per capita (kWh) and X_4 represent the GDP (current US$).

3.5.3. Developing the fossil fuel related generation model

The CO_2 emissions variable was used as the first independent variable because it had the greatest correlation coefficient of 0.975 with the dependent variable FFEG. Following the step 1 approach, the EC per capita variable was chosen as the second variable since FFEG-(CO_2 emissions-EC per capita) had the greatest performance determination coefficient of 0.9583. As per the step 2, the FFEG-(CO_2emissions-EC per capita-GDP) variables showed the highest performance determination coefficient of 0.9598 and so the GDP variable was selected as the third independent variable. According to step 3, FFEG-(CO_2emissions-EC per capita-GDP-FFEI cap) showed the greatest performance coefficient. Hence, FFEI cap was selected as the fourth independent variable. Thus, the fossil fuel electricity generation model was developed as follows:

$$\tilde{Y} = -4.20995 + 0.000643X_1 + 0.144145X_2 + 0.008155X_3 - (3.2 \text{ x } 10^{-11})X_4 \quad (8)$$

Where, X_1 is the CO_2 emissions (kt) and X_2 represents the FFEI cap (million kW). X_3 is the EC per capita (kWh) and X_4 represent the GDP (current US$).

3.5.4. Developing the net electricity generation model

This model was also created using regression analysis by applying the SSAF. The Pop variable was picked as the first independent variable due to having the highest performance coefficient of 0.980577 with the dependent variable EG. The remaining 11 significant independent variables were then tested in different combinations as per the SSAF. As in step 1, EG-(pop-CO_2 emissions) showed the strongest performance with a coefficient of 0.9709. Therefore, the CO_2 emissions variable was selected as the second predictor for the EG model. Following the step 2, EG-(pop-CO_2 emissions-EC per capita) showed the best performance coefficient of 0.9734. Therefore, EC per capita variable was chosen as the third independent variable. As per step 3, EG-(pop-CO2 emissions-EC per capita-labor force) showed the highest performance coefficient of 0.974, so the labor force variable was picked as the fourth independent variable. To obtain the fifth predictor, the remaining 8 significant variables to EG were tested with the (pop-CO_2 emissions-EC per capita-labor force) variables. As EG-(pop-CO_2 emissions-EC per capita-labor force-EI cap) showed the highest coefficient, the EI cap variable was picked as the fifth predictor. However, it was observed that the performance coefficient, in step 4, was lower than that of the previous steps despite having more predictors. It was also noticed that as the analysis progressed, the p-value for the pop variable was continuously increasing, reaching the value of 0.9391 in the EG-(pop-CO_2 emissions-EC per capita-labor force-EI cap) analysis, which is far greater than the p-significance value of 0.05. So, when the pop variable was removed, the model provided a better result with a performance coefficient of 0.97511 with four predictors which are CO_2 emissions-EC per capita-labor force-EI cap. Therefore, the net electricity generation model was created as follows:

$$\tilde{Y} = -3.48558 + (0.000635)X_1 + (4.64 \text{ x } 10^{-7})X_2 + (0.003719)X_3 - (0.30931)X_4 \quad (9)$$

Where X_1 is the CO_2 emissions (kt) and X_2 represents the labor force. X_3 is the EC per capita (kWh) and X_4 represent the EI cap (million kW).

3.6. Significance of the variables used in the formulated models and their impact on electricity generation and consumption

According to the created models, electricity production/consumption in Cameroon is affected by CO_2 emissions, electricity consumption per capita, labor force, electricity installed capacity, final consumption expenditure and GDP variables. Each variable plays an important role in electricity production/consumption as described below:

- CO_2 emissions: It is the stemming released from the burning of fossil fuels and the production of cement [6]. More emissions mean more fossil fuel combustion and the more fossil fuels are burned for electricity production, the more electricity is produced and consumed.
- Electricity consumption per capita: It is the amount of electricity consumed per person in a country each year. In general, the greater the electricity consumption per capita, the greater the net electricity production/consumption.
- Labor force: It consists of people aged 15 and up who provide labor to produce goods and services over a

specified period [6]. As the labor force grows, more people can afford to purchase electricity. Therefore, electricity consumption/ production also increases.

- Electricity installed capacity: It refers to the maximum amount of electricity that can be produced from all electrical energy sources combined [13]. As installed capacity increases, electricity generation/consumption also increases.
- Final consumption expenditure: It is the sum of household final consumption expenditure and general government final consumption expenditure [6]. It is also directly related to electricity production/consumption.
- GDP: It is the total value of goods and services produced in a country over a particular period of time [6]. GDP positively affects electricity generation/consumption, as the ability to produce or consume electricity increases with GDP. In the modern world, electricity consumption has an influential role in the growth of economies [19].

4. Results

This section presents the results of the power generation and consumption models and their performance values such as the RMSE, the RRMSE, the R square, the adjusted R square, and the significance F value. It also presents the comparison of this study with previous studies on power modeling.

4.1. Results of the net electricity consumption model

Table 5 shows the results obtained from the formulation of the net electricity consumption model. The analysis resulted in an excellent RMSE of 0.246 and a RRMSE of 0.0577. This indicated that this model could predict net electricity consumption with a 5.77% error in its mean value. Furthermore, the adjusted R square of 0.9690 with a value very close to 1 also confirmed the goodness of fit of the model.

Table 5. Regression statistics for the net electricity consumption model.

Multiple R	R Square	Adjusted R Square	RMSE	RRMSE	Significance F
0.987	0.974	0.969	0.246	0.057	4.22 x 10^{-14}

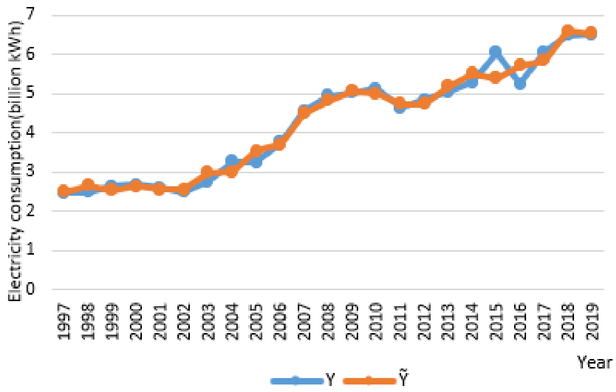

Figure 3. Real (Y) vs predicted (Ỹ) curve for the net electricity consumption.

Figure 3 shows the curve of the predicted electricity consumption model in orange (Ỹ) and the real electricity consumption in blue (Y). As electricity consumption increases over time, both curves show a similar trend and are almost in line with each other. Therefore, based on this graph and the model performance displayed in Table 5 it can be concluded that this model has an excellent forecasting ability of the net electricity consumption in Cameroon.

4.2. Results of the hydroelectricity generation model

The regression statistics for the hydroelectricity model are shown in Table 6. The results showed a very good performance for the RMSE and the RRMSE (CV) of 0.1704 and 0.04268 respectively. This means that the error obtained from the use of this model represents 4.268% of the mean value of hydroelectricity generation. Thus, also representing an excellent model. The model also provided an excellent fit to the data used in its creation indicated by the significance F value of 1.20 x 10^{-10}. It also showed an adjusted R square of 0.92478, which is an additional tool to confirm the stability of the model.

Table 6. Regression statistics for hydroelectricity generation.

Multiple R	R Square	Adjusted R Square	RMSE	RRMSE	Significance F
0.968	0.938	0.924	0.170	0.042	1.20 x 10^{-10}

Figure 4 shows the overall results of the hydroelectricity model in graphical form.

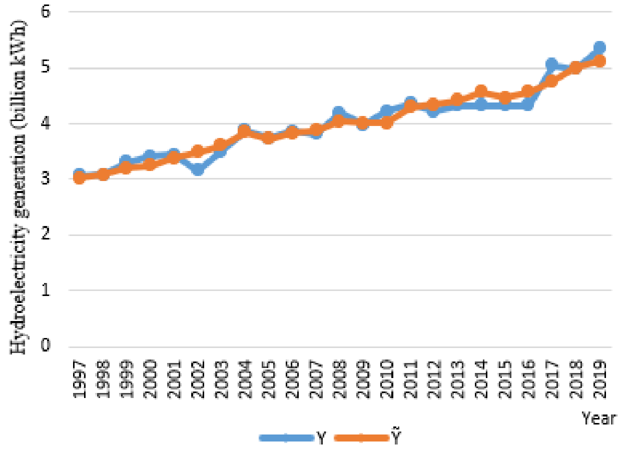

Figure 4. Real (Y) vs predicted (Ỹ) curve of hydroelectricity generation

The actual curve is depicted in blue (Y), while the predicted curve is shown in orange (Ỹ). Both curves show the evolution of hydroelectricity generation almost at the same rate. In accordance with the excellent performance coefficient displayed in Table 6, it can be concluded that the model was very successful and can be used for hydroelectricity forecasts.

4.3. Results of the fossil fuel electricity generation model

Table 7 presents the regression statistics for the fossil fuels electricity generation model. The results revealed that the model presented excellent goodness of fit based on the performance coefficient R square and adjusted R square of 0.965 and 0.958 respectively. Furthermore, the RMSE of 0.238 was very low. RRMSE (CV) of 0.1918 suggested that the model could make predictions with 19% error of the mean value of fossil fuel electricity generation. This error is less than 20%, indicating a good forecasting ability according to [16].

Table 7. Regression statistics for fossil fuel electricity generation.

Multiple R	R Square	Adjusted R Square	RMSE	RRMSE	Significance F
0.982	0.965	0.958	0.238	0.191	6.31 x 10^{-13}

Figure 5 shows the graphical representation of the fossil fuel electricity generation model. Based on its performance coefficient shown in Table 7, it can be concluded that the formulated model is statistically valid and can be used in forecasting fossil fuel electricity generation.

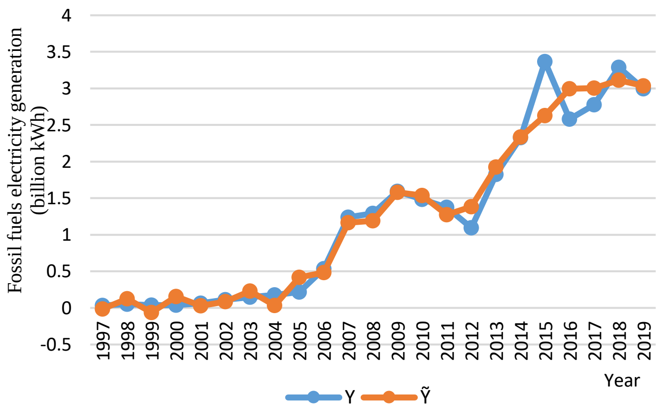

Figure 5. Real (Y) vs predicted (Ỹ) curve of fossil fuels electricity generation.

4.4. Result of the net electricity generation model

The results of the regression analysis for the net electricity generation are presented in this section. Table 8 indicates the performance of the model. It was observed that the RRMSE was equal to 0.0526 or 5.26% which is less than 10%, thus indicating an excellent model as stated by [16]. Also, the adjusted R square of 0.97511 was very close to 1. Consequently, a significance F of 5.91 x 10^{-15} indicated that the model provided a better fit to the data used in its formulation, and thus was ideal for electrical energy predictions.

Table 8. Regression statistics for the net electricity generation model.

Multiple R	R Square	Adjusted R Square	RMSE	RRMSE	Significance F
0.989	0.979	0.975	0.275	0.052	5.91 x 10^{-15}

Figure 6 shows the results of the model. It can be observed that predicted curves represented in orange (Ỹ) and the actual curve represented in blue (Y) for electricity generation are very close, with both showing an increasing trend. Taking this and its excellent performance coefficient indicated by the RRMSE into consideration, it can be concluded that the 95% confidence level for forecasts remained constant over time. Thus, confirming the stability of the model

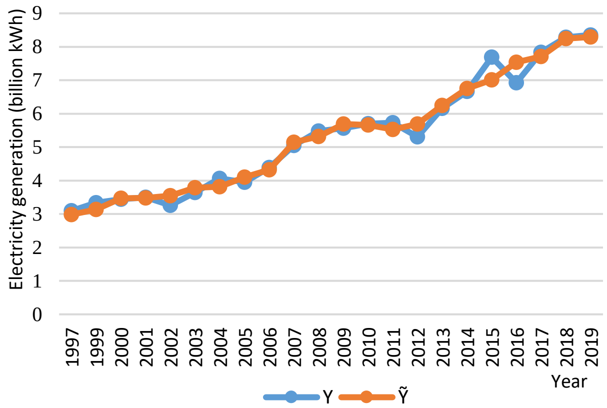

Figure 6. Real (Y) vs predicted (Ỹ) curve of the net electricity generation.

5. Discussion

5.1. Results comparison with previous studies

The MAPE is a straightforward comparative metric that makes error interpretation easier [20]. The MAPE for the net electricity generation and consumption was also calculated as 3.46% and 3.75% respectively. This was done in order to compare with previous studies in the literature. According to [21], a model is considered of high precision if MAPE < 10%, good if 10% < MAPE < 20%, reasonable if 20% < MAPE < 50% and inaccurate for MAPE > 50%. This reveals that developed models were of high precision based on their respective MAPE values. The results of this study were compared to those found in the literature on Cameroon power consumption modeling, as shown in Table 9. The results of this study were found to be superior to those in the literature, with a much lower RMSE value. The VECM-HES hybrid model, on the other hand, reported a lower RMSE than that in this study, indicating that this study reported the most reliable model after VECM-HES hybrid model. This confirms the accuracy of the model. Additionally previous studies in the literature were compared based on their respective MAPE values. The MAPE obtained from this study was found to be also superior to those from previous studies except for their respective hybrid models. However, the hybrid models are mathematically complex and challenging to understand by someone with little background in the field, especially those not present during model creation. It is therefore recommended to use the model developed in this study for electric power modeling in Cameroon since it outperforms other non-hybrid models in the literature.

Table 9. Comparison with similar studies on power consumption modeling in Cameroon.

Model	RMSE	MAPE	RRMSE	Author
VECM	395.4	12.65	-	Dieudonne et al., (2022)
HES	229	7.32	-	Dieudonne et al., (2022)
VECM-HES	0.0674	1.59	-	Dieudonne et al., (2022)
GM (1,1)	-	3.96	-	Guefano et al., (2021)
VAR (1)	-	7.73	-	Guefano et al., (2021)
GM (1,1)-VAR (1)	15	1.63	-	Guefano et al. (2021)
MLR	0.2465	3.75	5.77%	Developed in this study

These results were also compared to those on power generation in the literature (other countries) as shown in Table 10. Both the R square and adjusted R square obtained in this study were better than those reported in the literature. However, only Ibeh & Agwu [9] reported better RMSE than in this study.

Table 10. Comparison with previous studies on power generation modeling in other countries.

Model	RMSE	MAE	MAPE	R square	Authors
MLR	0.076	0.003	-	0.93	Ibeh and Agwu (2011)
LR	-	-	-	0.69	EI, et al., (2017)
MLR	0.275	0.0346	3.46	0.9796	Developed in this study

5.2. Models' accuracy and predictions

In comparison with actual values, the models developed in this study provide excellent predictions of electricity consumption and generation. As an example, the model predicted electricity generation in 2018 and 2019 at 8.239 billion kWh and 8.286 billion kWh, respectively. These values are close to the actual values of 8.28083 billion kWh and 8.35384 billion kWh in 2018 and 2019 respectively as reported by the EIA in 2021. The model also estimated electricity consumption in 2018 and 2019 at 6.588 billion kWh and 6.536 billion kWh respectively. These values are also very close to the actual values of 6.49483 billion kWh and 6.50884 billion kWh in 2018 and 2019 respectively, as reported by the EIA in 2021. This confirms the ability of the model to give estimations with excellent accuracy. Additionally, the model predicts electricity generation and consumption to be 9.6859 billion kWh and 8.0567 billion kWh (8056.77 GWh) respectively in 2024. Using Excel, predictor values were obtained by plotting each predictor variable against the year variable. The values were then used in the models to make predictions. Dieudonne, et al. [3] estimated electricity consumption in 2024 to be 8134.772 GWh based on their hybrid model. This shows an overestimation of 78 GWh when compared to this model. Considering that each predictor variable's value was estimated based on the scatter plot of each predictor variable and the year variable before using in the model, it can be concluded that this model actually provides very good forecasts.

6. Conclusion and future work

Using multiple linear regression techniques, this study analyzed electricity generation and consumption in Cameroon by formulating linear models that can be used to predict future power demand. Based on quantitative analysis of the dependent and independent variables used in this study, it can be concluded that CO_2 emissions, electricity consumption per capita, labor force, electricity installed capacity, final consumption expenditure and GDP are important factors to consider when designing models for electrical power generation and consumption in Cameroon. According to the performance criteria, the results indicated excellent model formulation with (RMSE = 0.275, RRMSE = 5.26%, and Adjusted R square = 0.975) for the net electricity generation model and (RMSE = 0.246, RRMSE = 5.77%, and Adjusted R square = 0.969) for the net electricity consumption model. Additionally, the results revealed an excellent model formulation for the hydroelectricity generation model with (RMSE = 0.170, RRMSE = 4.269%, and Adjusted R square =0.924) and a good fossil fuel electricity generation model with (RMSE =0.238, RRMSE = 19.183%, and Adjusted R square = 0.958). This confirms its forecasting abilities. This study has added to the existing knowledge of electrical energy modeling in Cameroon by developing a model with high forecasting ability for electricity consumption. It has also considered the gap between electrical power supply and demand in Cameroon, and established electricity generation models with excellent accuracy. These models can be used to estimate and understand the requirements of the future power supply to the electric grid. The findings can assist government officials to implement efficient electrical energy management strategies in order to reduce power losses in the country and improve energy access.

Several studies have indicated that artificial intelligence is an advanced technique that is better suited to forecasting exercises and has demonstrated better results than regression in the literature. Also, it has been used for electric power modeling around the world, but has not yet been used for power production modeling in Cameroon. Further research is therefore needed to identify alternative methods for electrical power consumption/generation modeling, particularly those that use artificial intelligence to generate outputs that could improve the effectiveness of various modeling techniques.

Author contributions

Blaise Fombuwing: Conceptualization, Writing, Methodology, Results, Discussion and Conclusion. **Neyre Tekbıyık-Ersoy:** Visualization, Supervision, Guiding, Assisting, Reviewing and Editing

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Guefano, S., Tamba, J. G., Azong, T. E. W., & Monkam, L. (2021). Forecast of electricity consumption in the Cameroonian residential sector by Grey and vector autoregressive models. *Energy, 214*.
2. Guefano, S., Tamba, J. G., Monkam, L., & Bonoma, B. (2020). Forecast for the Cameroon's residential electricity demand based on the multilinear regression model. *Energy and Power Engineering, 12*(05).
3. Dieudonné, N. T., Armel, T. K. F., Vidal, A. K. C., & René, T. (2022). Prediction of electrical energy consumption in Cameroon through econometric models. *Electric Power Systems Research, 210*.
4. African Development Bank Group. (2021). Country priority plan and diagnostic of the electricity sector Cameroon 3.
5. Kidmo, D. K., Deli, K., & Bogno, B. (2021). Status of renewable energy in Cameroon. *Renewable Energy and Environmental Sustainability, 6*, 2.
6. World Bank Database. (2022).
7. Tamba, J. G., Nsouandélé, J. L., Fopah Lélé, A., & Sapnken, F. E. (2017). Electricity consumption and economic growth: Evidence from Cameroon. *Energy Sources, Part B: Economics, Planning and Policy, 12*(11), 1007–1014.
8. Joel, S., & Nguepdjio Cyrille, N. (2019). Consommation d'énergie, croissance économique et émissions de Co2 au Cameroun: une analyse de causalité. *African Integration and Development Review*.
9. Ibeh, L., & Agwu, C. (2011). Comparison of regression model and artificial neural network model for the prediction of electrical power generated in Nigeria. *Spatial Dynamics of Urban Poverty and Environment Nexus*.
10. El, I., Bannari, R., Abouabdellah, A., & Guerrero, J. M. (2017). Energy production: A comparison of forecasting methods using the polynomial curve fitting and linear regression. In *2017 International Renewable and Sustainable Energy Conference (IRSEC)*.
11. Verma, T., Tiwana, A. P. S., Reddy, C. C., Arora, V., & Devanand, P. (2016). Data analysis to generate models based on neural network and regression for solar power generation forecasting. *Proceedings - International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, 97–100.
12. Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences, 106*, 234–240.
13. U.S. Energy Information Administration Database. (2022).
14. Worldometers. (2022). Cameroon natural gas.
15. Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal, 24*(3).
16. Despotovic, M., Nedic, V., Despotovic, D., & Cvetanovic, S. (2015). Review and statistical analysis of different global solar radiation sunshine models. *Renewable and Sustainable Energy Reviews, 52*, 1869–1880.
17. Liu, R. X., Kuang, J., Gong, Q., & Hou, X. L. (2003). Principal component regression analysis with SPSS. *Computer Methods and Programs in Biomedicine, 71*(2), 141–147.
18. Fombuwing, B. (2023). Modeling electricity generation and consumption in Cameroon. *MSc Thesis, Cyprus International University*.
19. Comert, M., & Yildiz, A. (2022). A novel artificial neural network model for forecasting electricity demand enhanced with population-weighted temperature mean and the unemployment rate. *Turkish Journal of Engineering, 6*(2), 178-189.
20. Ayyildiz, E., & Murat, M. (2024). A lasso regression-based forecasting model for daily gasoline consumption: Türkiye case. *Turkish Journal of Engineering, 8*(1), 162-174.
21. Yadav, A. K., & Chandel, S. S. (2014). Solar radiation prediction using artificial neural network techniques: A review. *Renewable and Sustainable Energy Reviews, 33*, 772–781.
22. Tatietse, T. T., Kemajou, A., & Diboma, S. (2010). Electricity self-generation costs for industrial companies in Cameroon. *Energies, 3*(7).
23. USAID. (2016). Cameroon Power Africa fact sheet.
24. MINEPDED. (n.d.). Cameroon climate investment fund, Ministry of Environment and protection of nature. *Sustainable Development*, 1-10.
25. Wirba, A. V., Mas'ud, A., Abdullahi, A., & Firdaus, M. (2015). Renewable energy potentials in Cameroon: Prospects and challenges. *Renew Energy*.