Energy Research Journal

Review

The use of Principal Component Regression and Time Series Analysis to Predict Nitrous Oxide Emissions in Ghana

Benjamin Odoi, Lewis Brew and Christopher Attafuah

Department of Mathematical Science, University of Mines and Technology, Tarkwa, Ghana

Article history
Received: 10-08-2020
Revised: 04-01-2021
Accepted: 06-01-2021

Corresponding Author:
Benjamin Odoi
Department of Mathematical Science, University of Mines and Technology, Tarkwa, Ghana
E-mail: bodoi@umat.edu.gh

Abstract: The disturbing pace of emanation of Nitrous Oxide (N_2O) into the atmosphere and its calamitous impact on the environment, as monitored by many governmental agencies and researchers has become a wellspring of worry for many nations and therefore needs due attention. The study deployed nitrous oxide emissions from the three sectors to the total nitrous oxide emissions in Ghana over the period 1990 to 2016. The sectors, energy sector, Agriculture Forestry and Other Land Use (AFOLU) and Waste sector were considered against the total N_2O emissions. Principal Component Regression (PCR) was applied to the input variables for the reduction of its large size to a few principal components to explain the variations in the original dataset since there was the presence of multicollinearity. Autoregressive Integrated Moving Average (ARIMA) was used to develop models to predict the total N_2O emissions and emissions from the sectors in Ghana. The appropriate models that fitted the data well were ARIMA (1,2,1) and ARIMA (1,1,2) based on information criteria (AIC, AICc and BIC). The ARIMA (1,2,1) model was found to be the most suitable model for predicting N_2O emission from Energy sector and Waste sector. 70% Of the dataset was used for the analysis and the results from the forecasted values mimic the original dataset. It was revealed that the AFOLU sector is the predominant sector that significantly contribute the overall N_2O emission in the atmosphere based on standardized coefficient. The model was adequate since its MAPE for AFOLU sector and the total N_2O emissions were 2.95 and 2.68% respectively, meaning the model explained 97.05 and 97.32% respectively. The predicted values mimic the trend of the current situation at hand.

Keywords: Principal Component Regression, Multicollinearity, Nitrous Oxide, Times Series

Introduction

The planet's continuing rise in temperature is really upsetting. This is root caused by global warming. Global warming is the long-term heating of Earth's climate system observed due to human activities, primarily fossil fuel burning, which increases heat-trapping greenhouse gas levels in Earth's atmosphere. The clouds, water particles, reflective ground surfaces and ocean surface then send back into space about 30% of sunlight, while the rest is absorbed by seas, air and land (Jacobson, 2014). Consequently, this heats up the planet's surface and atmosphere and makes life possible. As the earth warms up, the thermal radiation and infrared rays radiate this solar energy, propagating it directly out into space and thus cooling the earth down (Eppelbaum *et al*., 2014). However, some of the outgoing radiation is reabsorbed into the atmosphere by nitrous oxide, carbon oxide, water vapour, ozone, methane and other gases and is radiated back to the surface of the earth. Because of their heat-trapping capacity, these gases are commonly known as greenhouse gases (Shahzad, 2015).

Global warming has remained a topic of discussion and a debatable issue among politicians and the scientific community ever since it emerged in the early nineteenth century (Berlie, 2018). The special Eurobarometer 2009 report adds that the world's most serious problems at the moment include global warming, poverty and international terrorism. But most Europeans respond that global warming is by far the most serious challenge compared with any other threat.

N_2O is recognised as the most important ozone depleting substance (Ravishankara *et al*., 2009). Nitrous Oxide (N_2O) is the third most prevalent GHG, behind Carbon Dioxide (CO_2) and methane (CH_4). Since the early 1990 s, the concentration of this gas in the atmosphere has steadily increased and has an atmospheric life of 121 years. Nitrous Oxide (N_2O) also has a potential for global warming 300 times that of carbon dioxide over a 100-year timeframe (Griffis *et al*., 2017).

In the year 2016, the total national greenhouse gas emission in Ghana was 42.2 $MtCO_2e$ (based on nitrous oxide, carbon dioxide, methane, hydrofluorocarbon and perfluorocarbon). Nitrous oxide was the second largest greenhouse gas for that year, constituted about 18.3%. Nitrous Oxide (N_2O) is a very stable substance in the atmosphere and for several decades the emission can influence global atmospheric concentrations (Ogeya *et al*., 2018; Tiemeyer *et al*., 2016).

In fact, the findings of a recent scientific analysis show that nitrous oxide is the leading ozone depleting agent currently released. Legislation to restrict nitrous oxide emissions could therefore contribute to both protecting climate change and recovering ozone.

This study aims to use historical empirical data to examine various economic sectors that contribute to N_2O emissions in Ghana and make future predictions using ARIMA model to help Ghana government implement different policies and strategies to limit nitrous oxide emissions within its borders.

Some Related Literature

The following are some related works considered under this study.

Nyoni and Bonga (2019) predicted Carbon Dioxide (CO_2) emissions in India using Box-Jenkins ARIMA approach over the period 1960 to 2017 and established that the ARIMA (2,2,0) is the best fit model for predicting CO_2 emissions in India. They also found out that CO_2 emissions in India are likely to increase and thereby exposing India to climate related challenges.

Nyoni and Bonga (2019) used ARIMA in modeling and forecasting carbon dioxide emissions in China. They found out that ARIMA (1,2,1) is the optimal model for forecasting carbon dioxide emissions in china and also CO_2 emissions in china are likely to increase and thereby exposing china to plethora of climate change related challenges.

Rahman and Hagan (2017), using forty-four-year time series data from 1972-2015 based on ARIMA models, revealed that the ARIMA model (0,2,1) is the best model for carbon dioxide modelling and prediction in Bangladesh.

Hossain *et al*. (2017) forecasted carbon dioxide emissions in Bangladesh using Box-Jenkins ARIMA technique over the period 1972-2013 and identified that the ARIMA (12,2,12), ARIMA (8,1,13) and the ARIMA (5,1,5) are the best fits models for forecasting CO_2 emissions from Gaseous Fuel Consumption (GFC), Liquid Fuel Consumption (LFC) and Solid Fuel Consumption (SFC) rather the other methods of forecasting Holt-Winters Non Seasonal (HWNS) and Artificial Neural Networks (ANN) models.

Ismail and Abdullah (2016) combined Principal Component Regression (PCR) and Back-Propagation Neural Networks (BPNN) techniques in order to improve the accuracy of the electricity demand prediction rates.

Mendeș (2009) used multiple linear regression models based on principal components scores to predict slaughter weight of broiler.

Thupeng *et al*. (2018) used Principal Component Regression (PCR) technique to predict a day in advance the daily maximum 1 h average ambient ground level ozone concentration for Maun town.

Mishra and Vanli (2016) used principal component regression for extracting damage sensitive features of a lamb wave sensor signal and establish a relation between the features and measured areas.

Rahayu *et al*. (2017) used principal component analysis to reduce multicollinearity of the currency exchange rate of some countries in Asia, period 2004-2014.

Ganiyu and Zubairu (2010) developed a predictive cost model using principal component regression for public building projects in Nigeria.

Haque *et al*. (2013) developed principal component regression by combining multiple linear regression and principal component analysis to forecast future water demand in the Blue Mountains, water supply systems in New South Wales, Australia.

Lall *et al.* (2016) used principal component regression model for predicting acceleration factors for copper-aluminum wire bond, subjected to harsh environments.

Sousa *et al*. (2007) used multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. The aim of their study was to predict next day hourly ozone concentrations through a new methodology based on feedforward artificial neural networks using principal components as inputs. They found that the use of principal components as inputs improved model prediction by reducing complexity and eliminating data collinearity.

Asare *et al.* (2018) used principal component regression method to predict the water level of the Akosombo Dam.

Methods Used

The general multiple linear regression model with response Y and predictors $X_1,\dots,X_n$ will have the form of Eq. (1):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \qquad (1)$$

where, β_0 is the intercept point of the regression line and y-axis, β_1, $\beta_2 \ldots \beta_n$ are the regression coefficients associated with X_1, $X_2 \ldots X_n$ respectively. Each coefficient measures the effect of the corresponding predictor after taking account of the effect of all other predictors in the model and ε is the error.

Assumptions in Multiple Linear Regressions

Some assumptions are needed in the model $Y = X\beta + \varepsilon$ for drawing the statistical inferences. The following assumptions are made:

i. $E(\varepsilon) = 0$
ii. $E(\varepsilon\varepsilon^T) = \sigma^2 I_n$
iii. X is a non-stochastic matrix
iv. $\varepsilon \sim N(0, \sigma^2 I_n)$

These assumptions are used to study the statistical properties of estimators of regression coefficients.

Multicollinearity

Multicollinearity occurs when two or more predictors in a regression model are moderately or highly correlated with one another. Predictors multicollinearity occurs when your model includes multiple factors that are correlated not just to your response variable, but also to each other. This problem is more troublesome at smaller sample sizes, where the standard errors are usually larger due to sampling error. Multicollinearlity can be tackled by applying some multivariate techniques like principal component regression, factor analysis and so on. However, this study makes use of the principal component regression.

Principal Component Regression

Suppose that we have a random vector X:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \quad (2)$$

with population variance-covariance matrix:

$$\operatorname{var}(X) = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp}^2 \end{pmatrix} \quad (3)$$

Consider the linear combinations:

$$\begin{aligned} Y_1 &= e_{11}X_1 + e_{12}X_2 + \cdots + e_{1p}X_p \\ Y_2 &= e_{21}X_1 + e_{22}X_2 + \cdots + e_{2p}X_p \\ &\vdots \\ Y_p &= e_{p1}X_1 + e_{p2}X_2 + \cdots + e_{pp}X_p \end{aligned} \quad (4)$$

Each of these can be thought of as a linear regression, predicting Y_i from $X_i, X_2, \ldots X_p$. There is no intercept, but e_{ip}, $e_{i2}, \ldots, e_{ip}$ can be viewed as regression coefficients. Note that Y_i is a function of our random data and so is also random. Collect the coefficients e_{ij} into the vector:

$$e_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{ip} \end{pmatrix} \quad (5)$$

Therefore, it has a population variance:

$$\operatorname{var}(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} = e_i^T \sum e_i \quad (6)$$

It should be noted that Eq. (6) was obtained from the matrix var (X) in Eq. (3) and e_i^T is the transpose of e_i. Moreover, Y_i and Y_j have population covariance:

$$\operatorname{cov}(Y_i, Y_j) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{jk}\sigma_{kl} = e_i^T \sum e_j \quad (7)$$

First Principal Component (PCA1)

The first principal component Y_1 is the linear combination of X-variables (among all linear combinations) with maximum variance. It represents as much variation as possible in the data. Specifically, we define coefficients e_{11}, $e_{12}, \ldots, e_{1p}$ for the first component in such a way that its variance is maximized, subject to the constraint that the sum of the squared coefficients is equal to one. This constraint is required so that a unique answer may be obtained.

More formally, select e_{11}, $e_{12}, \ldots, e_{1p}$ that maximizes:

$$\operatorname{var}(Y_1) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{1l}\sigma_{kl} = e_1^T \sum e_1 \quad (8)$$

Subject to the constraint that:

$$e_1^T e_1 = \sum_{j=1}^{p} e_{1j}^2 = 1 \quad (9)$$

DOI: 10.3844/erjsp.2021.1.12

Second Principal Component (PCA2)

The second principal component Y_2 is the linear combination of x-variables, which represents as much of the remaining variation as possible, with the limitation that the correlation between the first and second components is 0. Select e_{21}, $e_{22},\ldots,e_{2p}$ select that maximise the variance of this new component:

$$\text{var}(Y_2) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{2k}e_{2l}\sigma_{kl} = e_2^T \sum e_2 \tag{10}$$

subject to the constraint that the sums of squared coefficients add up to one:

$$e_2^T e_2 = \sum_{j=1}^{p} e_{2j}^2 = 1 \tag{11}$$

along with the additional constraint that these two components are uncorrelated:

$$\text{cov}(Y_1, Y_2) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{2l}\sigma_{kl} = e_1^T \sum e_2 = 0 \tag{12}$$

All subsequent principal components have the same property-they are linear combinations accounting for as much of the remaining variation as possible and are not correlated with the other principal components.

Same procedure is carried out for each additional component for instance.

i^{th} Principal Component (PCAi)

For the i^{th} principal component Y_i we select e_{i1}, $e_{i2},\ldots,e_{ip}$ maximize:

$$\text{var}(Y_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} e_{ik}e_{il}\sigma_{kl} = e_i^T \sum e_i \tag{13}$$

subject to the constraint that the sums of squared coefficients add up to one, along with the additional constraint that this new component is uncorrelated with all the previously defined components:

$$e_i^T e_i = \sum_{j=1}^{p} e_{ij}^2 = 1 \tag{14}$$

$$\begin{aligned} \text{cov}(Y_1, Y_i) &= \sum_{k=1}^{p}\sum_{l=1}^{p} e_{1k}e_{il}\sigma_{kl} = e_1^T \sum e_i = 0 \\ \text{cov}(Y_2, Y_i) &= \sum_{k=1}^{p}\sum_{l=1}^{p} e_{2k}e_{il}\sigma_{kl} = e_2^T \sum e_i = 0 \\ &\vdots \\ \text{cov}(Y_{i-1}, Y_i) &= \sum_{k=1}^{p}\sum_{l=1}^{p} e_{i-1,k}e_{il}\sigma_{kl} = e_{i-1}^T \sum e_i = 0 \end{aligned} \tag{15}$$

Therefore, all principal components are uncorrelated with one another. The variance for the i^{th} principal component is equal to the i^{th} eigenvalue of matrix var (Y):

$$\text{var}(Y_i) = \text{var}(e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X) = \lambda_i \tag{16}$$

Box-Jenkins Model Approach

This time series forecasting is a step by step approach which apply ARMA or ARIMA to find the best fit of time series model to past values of a time series data (Box *et al*., 1994). The basic steps in Box-Jenkins methodology are:

i. Differencing the series to achieve stationarity
ii. Identification of tentative model
iii. Estimation of the model
iv. Diagnostic checking of the model; and
v. Using the model for forecasting

Preliminary Analysis of Data

The descriptive statistics from the data and correlations existing among the variables considered in the study are displayed in Table 1. In Table 1 the average values (means), the deviations from the mean (standard deviations), the minimum and maximum value for each of the variables considered in the study have been presented.

From Fig.1, there exists a structure in the data with one general class of relationship, thus positive (blue). It is observed that Waste sector and Agriculture, Forestry and Other Land Use (AFOLU) sector are positively related to each other.

To ascertain the dangers or consequences associated with Multicollinearity and as well validate the need to employ dimensional reduction (Principal Components). All the three sectors are regressed on the total nitrous oxide emissions. The result of the regression analysis is provided in Table 2. As observed, the p-value from the F-test (0.000), shows that the model is statistically significant (adequate). The adjusted R-squared indicates that about 100% of the total variability of total nitrous oxide is accounted for by the model. Also, two of the variables have Variance Inflation Factor (VIF) value greater than 5, hence an indication of multicollinearity in the model.

Therefore, an application of a direct Multiple Regression Analysis produced inaccurate results for interpretation and thus is normally called spurious regression. In order to solve the multicollinearity problem and perform a reliable regression analysis, Principal Component Analysis (PCA) is employed to help eliminate the level of multicollinearity in the dataset. Principal Component Analysis will also help identify appropriate variables (Principal Components) to be used as independent variables.

Formulation of Principal Components

The first step in formulating the principal components is an estimation of the correlation matrix of the independent variables considered in the study. You can also use Bartlett's test of Sphericity and Kaiser-Meyer-Olkin (KMO) test to determine whether performing principal component is necessary.

Table 3 shows the Bartlett's test of Sphericity and KMO test. The *p*-value for the Bartlett's test of Sphericity (P-value <0.05) meaning the variables are not orthogonal (correlated) and overall Measure of Sampling Adequacy (MSA) of KMO (0.51>0.5). Collectively, these tests suggested that dataset is appropriate for Principal component regression. Table 4 is the correlation matrix of the independent variables considered in the study. The correlation matrix reveals that there is a strong correlation between the waste sector and AFOLU sector.

Table 5 contains information regarding the three possible principal components and their relative explanatory power as expressed by their eigenvalues. As expected, the component solution extracts the components in the order of their importance. Principal Components 1, 2 and 3 completely explains 72.01, 97.93 and 100% respectively of the dataset. Two criteria are evaluated in order to decide on the number of factors to retain. These are latent roots and the proportion of variance explained. Using the latent root and proportion of variance explained criteria, one component is retained, which explains about 72.01% of the dataset. Figure 2, which is the Scree plot also reveals that one component must be retained since it is the first component whose eigenvalue is greater than one.

Table 6 presents the principal component eigenvector. Column is the loading for the one principal component extracted with respect to each variable.

The result in Table 7 shows the varimax rotation for the components model. The results for the rotation are easy to interpret. One main component was extracted to represent the three components it shows that, one variable correlate well with component one and that variable is the waste sector.

Table 8 shows the principal component regression, after the extraction of the one main component, the eigenvectors for the one component was used as repressor for the regression analysis. The *F*-statistic was statistically significant at 5% significance level (F = 502.7, p-value = 0.0000). Also, the Adjusted R-Squared was approximately 95% to show how much the component can be explained on the dependent variable. The estimated Principal Component Regression (PCR) that fits the data gathered is given as:

$$Total\ N_2O = 5.59333 + 0.69798PC1 \tag{18}$$

Time Series Analysis

Time Series Analysis was analysed based on the three sectors and total nitrous oxide. Test for stationarity was performed to apply the method used on the three sectors and the total nitrous oxide. The energy sector and waste sector appear not to be stationary whilst AFOLU sector and Total N_2O appear to be stationary for the first differencing based on the KSPS, ADF and PP tests as shown in Table 9. However, after the second differencing the energy and waste sectors were stationary as shown in Table 10.

Table 1: Descriptive statistics of data

Variable	Mean	Standard Deviation	Minimum	Maximum
Total N_2O Emission (Total)	5.593	1.051	4.090	7.710
Sector				
Energy	0.282	0.050	0.180	0.380
Agriculture, Forestry and Other Land Use (AFOLU)	4.841	0.958	3.490	6.720
Waste	0.470	0.074	0.360	0.600

Table 2: Multiple regression on the raw data

Variable	Parameter estimate	Std. error	Std. coef.	t -value	P-value	VIF
Intercept	-0.0170	0.0101		-1.705	0.102	0.000
Energy	1.0120	0.0270	0.048	37.503	0.000	1.301
AFOLU	0.9990	0.0040	0.911	277.094	0.000	8.664
Waste	1.0350	0.0440	0.073	23.321	0.000	7.857

F-value = 267100, *P*-value = 0.0000, adjusted R-squared = 1

Table 3: Inspection of correlation matrix

Test	*P*-value	Chi-square
Bartlett's test of sphericity	0.0000	54.57756
Overall MSA		
Kaiser-Meyer-Olkin (KMO)	0.5100	

Table 4: Correlation matrix

	Energy	AFOLU	Waste
Energy	1.0000	0.4228	0.3080
AFOLU	0.4228	1.0000	0.9295
Waste	0.3080	0.9295	1.0000

Table 5: Eigenvalues of correlation matrix

Component	Eigenvalue	Proportion	Cumulative
1	2.1603	0.72010	0.72010
2	0.7775	0.25920	0.97930
3	0.0622	0.02074	1.00000

Table 6: Eigenvectors of the correlation matrix

Sectors	PC1
Energy	0.4077224
AFOLU	0.6567680
Waste	0.6343645

Table 7: Varimax rotation

	PC1
Energy	
AFOLU	
Waste	1.000

Table 8: Principal component regression

Variable	Parameter estimate	Standard error	*t*-value	*P*-value
Intercept	5.59333	0.04490	124.57	0.0000
PC1	0.69798	0.03113	22.42	0.0000

F-value = 502.7, *P*-value = 0.0000, adjusted R-squared = 0.9507

Table 9: Stationary test for the first differenced

	P-value		
Variable	ADF	KPSS	PP
Energy	0.30540	0.1	0.01
AFOLU	0.02330	0.1	0.01
Waste	0.21560	0.1	0.01
Total N_2O	0.03648	0.1	0.01

Table 10: Stationary test for the second differencing

	P-Value		
Variable	ADF	KPSS	PP
Energy	0.03313	0.1	0.01
Waste	0.01000	0.1	0.01

Model Selection

The formulation of the ARIMA models was based on the information triggered by the ACF and the PACF.

Based on the first difference achieving stationarity for AFOLU and Total N_2O sectors, the ACF plot shows an autocorrelation at lag 1 which exceeds the significance bound, but all other autocorrelation is below the significance bound whilst the PACF shows that the partial autocorrelation at lag 1 exceeds the significance bounds. Clearly, from the plots, AR and MA terms can be identified. Since the ACF plot of the first difference cut off after lag1, MA (1) can be assumed.

The PACF plot of the first difference tails off after lag 1, so AR (1) can be assumed. Hence, mixed model ARIMA (1,1,1) is formed by combining the AR and MA terms.

Again, the energy and waste sectors achieved stationarity at second differencing hence, AR and MA terms was identified. Since the ACF plot of the second difference cut off after lag 1, MA (1) was assumed. Likewise, the PACF tails off after lag 1, thus AR (1) was also assumed. Hence, mixed model ARIMA (1,2,1) is formed by combining the AR and MA terms. Figures 3 and 4 show the ACF and PACF for the energy sector at second differencing.

After model identification, the need arises to select a model based on the reliability of prediction. Three Information criteria (AIC, AICc and BIC) were considered for the model selection. The thumb rule is that the best model is the one with the minimum information criteria. It was revealed from the analysis that, ARIMA (1,2,1) was the model that best fits Energy sector and Waste sector while ARIMA (1,1,2) was the model that best fits AFOLU sector and the Total N_2O. Table 11 shows the model selection criteria used to select a good predictive ARIMA model.

The estimated parameters and the best fitted models based on the selection criteria for the Energy, AFOLU, Waste and Total N_2O sector is shown in Table 12 and Eqs. 19 to 22.

The fitted model for the Energy sector will be expressed as:

$$X_t = 2.06X_{t-1} - 1.06X_{t-2} + 1.0\varepsilon_{t-1} + \varepsilon_t \tag{19}$$

the fitted model for the AFOLU sector will be expressed as:

$$X_t = 1.98X_{t-1} - 0.98X_{t-2} + 1.41\varepsilon_{t-1} - 0.54\varepsilon_{t-2} + \varepsilon_t \tag{20}$$

the fitted model for the Waste sector will be expressed as:

$$X_t = 2.00X_{t-1} - 1.00X_{t-2} + \varepsilon_{t-1} + \varepsilon_t \tag{21}$$

the fitted model for the Total N_2O will be expressed as:

$$X_t = 1.98X_{t-1} - 0.98X_{t-2} + 1.34\varepsilon_{t-1} - 0.49\varepsilon_{t-2} + \varepsilon_t \tag{22}$$

Diagnostic Checking

For correlation on the standardized tests, the Ljung-Box Test was used. The hypothesis states that residuals are not correlated (null) and residuals are correlated (alternative). It was deduced that the p-values for all the sectors were greater than 0.05, hence the null hypothesis is not rejected and conclude that residuals are not correlated, as shown Table 13. This implies that the models are adequate.

Table 11: Model selection criteria

VAR	Criteria	ARIMA (1,2,1)	ARIMA (2,2,1)	ARIMA (2,2,2)	ARIMA (3,2,1)	ARIMA (2,2,3)
Energy	AIC	-96.16	-94.33	-92.33	-92.33	-94.67
	AICc	-95.02	-92.33	-89.17	-89.17	-90.00
	BIC	-92.51	-89.45	-86.23	86.23	-87.36
Waste	AIC	-130.87	-128.88	-126.97	-125.88	-126.43
	AICc	-129.73	126.88	-123.81	-122.72	-121.77
	BIC	-127.22	124.01	-120.87	-119.79	-119.72
VAR	Criteria	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (1,1,3)	ARIMA (1,1,4)
AFOLU	AIC	10.80	4.19	5.41	5.92	7.72
	AICc	11.89	6.09	7.32	8.92	12.15
	BIC	14.57	9.22	10.44	12.21	15.27
Total N_2O	AIC	11.85	6.10	6.81	7.93	9.84
	AICc	12.94	8.00	8.72	10.93	14.26
	BIC	15.63	11.13	11.85	14.22	17.38

Table 12: Parameter estimate for the best model for various sectors

Sectors	Parameter estimate
Energy	0.0559*(0.201)
	0.999**(0.549)
AFOLU	0.9806*(0.043)
	1.4136**(0.1670)
	0.5350***(0.1529)
Waste	0.0013*(0.2035)
	-1.000**(0.1257)
Total N_2O	0.9769*(0.0524)
	-1.3393**(0.1705)
	0.4895***(0.1705)

*AR (1), **MA (1), ***MA (2), **(_)** standard errors

Table 13: Ljung-box test

Variable	Chi-squared (χ^2)	P-value
Energy	5.2285	0.6321
AFOLU	1.1335	0.9924
Waste	1.3556	0.9869
Total N_2O	1.5680	0.9799

Table 14: Forecasted values

Year	Energy	AFOLU	Waste	Total N_2O
2017	0.385	6.843	0.609	7.871
2018	0.389	7.002	0.618	8.060
2019	0.395	7.157	0.628	8.244
2020	0.400	7.309	0.637	8.424
2021	0.405	7.459	0.646	8.599
2022	0.410	7.605	0.655	8.771
2023	0.415	7.749	0.665	8.938
2024	0.420	7.890	0.674	9.103
2025	0.425	8.028	0.683	9.263
MAPE (%)	7.548	2.952	1.624	2.681

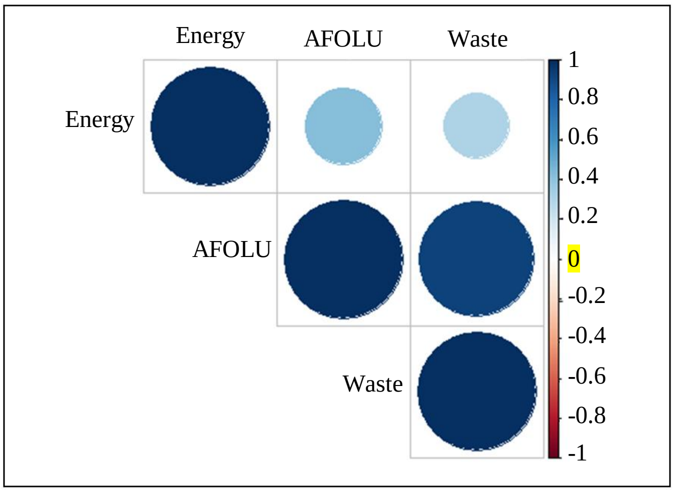

Fig. 1: Plot of correlation existing among variables

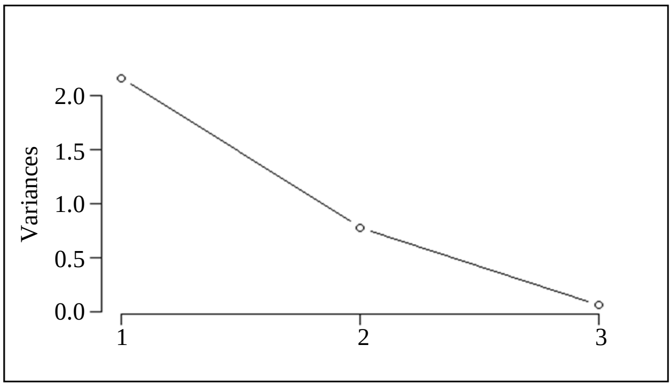

Fig. 2: Scree plot

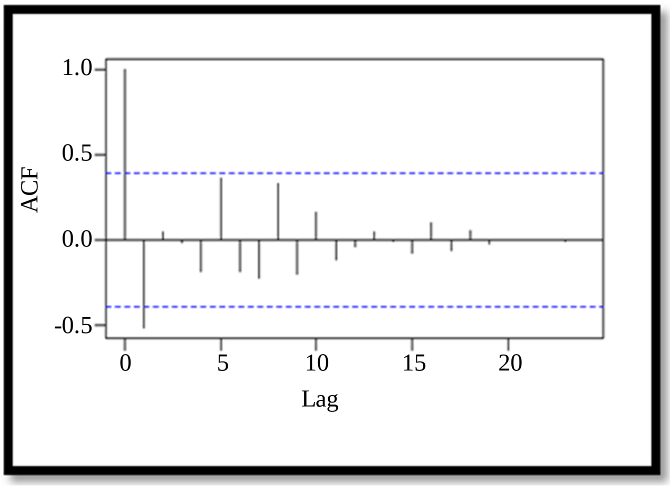

Fig. 3: ACF plot of the second differenced energy sector

The forecasted values (in million metric tons carbon dioxide equivalent) for 2017 to 2025 and the Mean Absolute Percentage Error (MAPE). The forecasted values show a significance increase from 2017 to 2025 as shown in Table 14. The forecasted values mimic the trend of the current situation at hand. Figures 5 to 8 are the forecasted plot for the three sectors and Total N_2O.

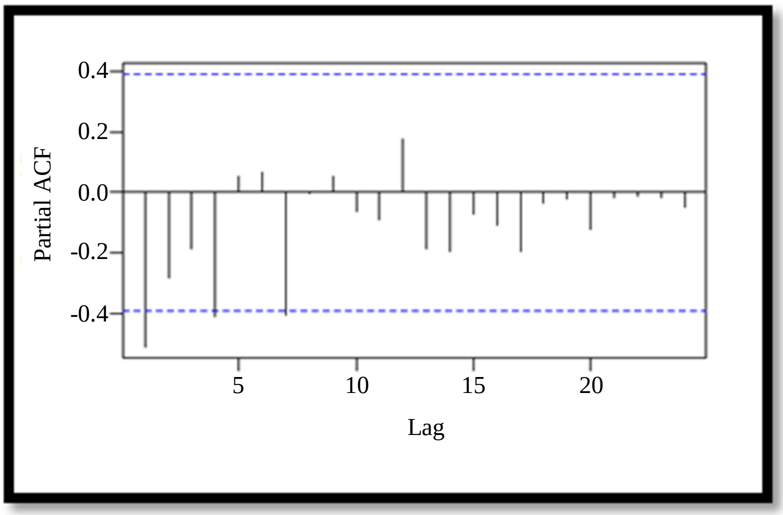

Fig. 4: PACF plot of the second differenced energy sector

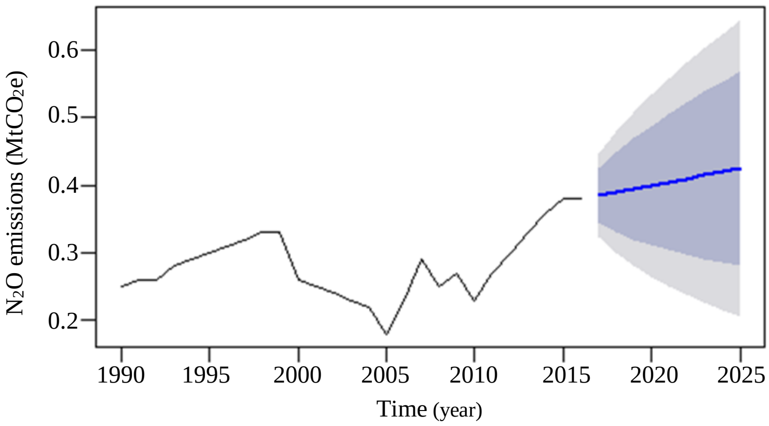

Fig. 5: Forecast plot for energy sector

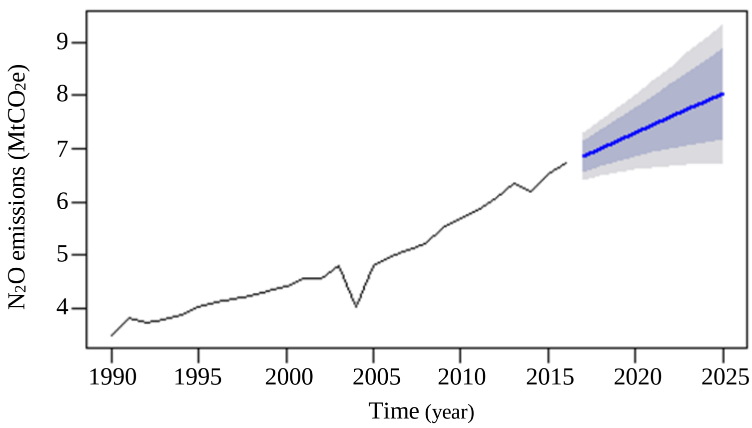

Fig. 6: Forecast plot for AFOLU sector

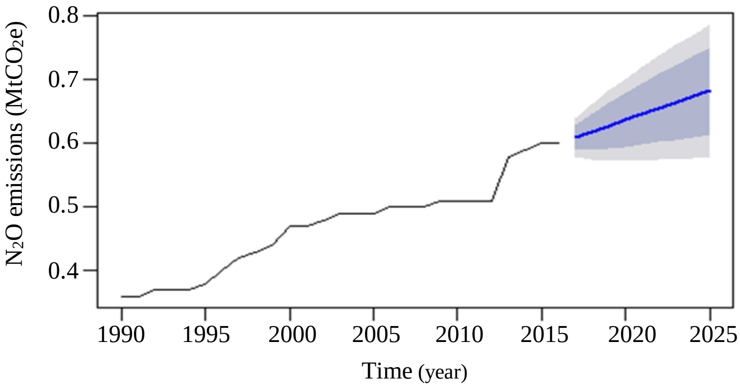

Fig. 7: Forecast plot for waste sector

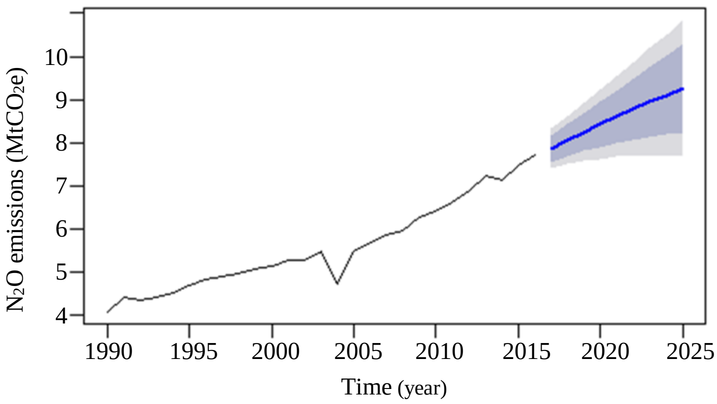

Fig. 8: Forecast plot for total N_2O emissions

Conclusion and Recommendation

The variance that was explained by the one main component was 72.01% as indicated Table 5. The approach presented here is efficient and appropriate for classification of nitrous oxide emissions that make up the total nitrous oxide emissions in Ghana. After the classification, the eigenvectors were regressed on the total nitrous oxide emissions and the result show that PC1 has a significant impact on the total nitrous oxide emissions. This means that, when there are more nitrous oxide emissions in the Waste Sector, it will have significant impact on the total nitrous oxide emissions and from the standardized coefficient, it was observed that Agriculture Forestry and Other Land Use (AFOLU) sector is the major contributor of overall nitrous oxide emission, followed by Waste sector and Energy sector. The study also provided an appropriate model for predicting N_2O emissions from the three sectors and the annual total nitrous oxide emissions in Ghana. Findings of the study have established that ARIMA (1,2,1) is the best fitted model for predicting N_2O emissions from energy and waste sector while ARIMA (1,1,2) is the best fitted model for predicting N_2O emissions from AFOLU sector and the annual total N_2O emissions in Ghana. The models were deemed accurate for prediction based on their small Mean Absolute Percentage Error (MAPE) values. It is expected that N_2O emissions from the three sectors and the total N_2O emissions will continue to increase.

Recommendation

In order to curb high nitrous oxide emissions from Agriculture, Forestry and Other Land Use (AFOLU) sector, it is recommended that nitrogen-based fertilizer application should be reduced, minimum tillage for cropping and reducing emissions from livestock as well as modifying a farm's manure management practices. It is also appropriate for policy makers to put in plays some mechanism to control the emissions of nitrous oxide from the other two sectors thus the Energy sector and Waste sector since as indicated in this study as a key factor of the increase in nitrous oxide.

Acknowledgement

We acknowledge the technical help of Dr Sampson Twumasi-Ankrah of Kwame Nkrumah University of Science and Technology (KNUST), Ghana. Ghana Statistical Service and the Energy sector of Ghana.

DOI: 10.3844/erjsp.2021.1.12

Author's Contributions

Benjamin Odoi: Participated in all experiments, coordinated the data-analysis and contributed to the writing of the manuscript. Coordinated the mouse work. Disigned the research plan and organized the study.

Lewis Brew: Coordined the mouse work. Designed the research plan and organized the study.

Christopher Attafuah: Designed the research plan and organized the study.

Ethics

We hereby state that this study is the authors' own original work, which has not been previously published elsewhere. The results are appropriately placed in the context of prior and existing research. All authors' have been personally and actively involved in substantial work leading to the paper and will take public responsibility for its content.

References

Asare, I. O., Frempong, D. A., & Larbi, P. (2018). Use of Principal Components Regression and Time-Series Analysis to Predict the Water Level of the Akosombo Dam Level.

Berlie, A. B. (2018). Global Warming: A Review of the Debates on the Causes, Consequences and Politics of Global Response. Ghana Journal of Geography, 10(1), 144-164.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994). Time series analysis: forecasting and control, prentice hall, Englewood cliffs, USA, 282-285.

Eppelbaum, L., Kutasov, I., & Pilchin, A. (2014). Applied geothermics (pp. 99-149). Springer Berlin Heidelberg.

Ganiyu, B. O., & Zubairu, I. K. (2010). Project cost prediction model using principal component regression for public building projects in Nigeria. Journal of Building Performance ISSN, 1(1), 2010.

Griffis, T. J., Chen, Z., Baker, J. M., Wood, J. D., Millet, D. B., Lee, X., ... & Turner, P. A. (2017). Nitrous oxide emissions are enhanced in a warmer and wetter world. Proceedings of the National Academy of Sciences, 114(45), 12081-12085.

Haque, M. M., Rahman, A., Hagare, D., & Kibria, G. (2013). Principal component regression analysis in water demand forecasting: An application to the Blue Mountains, NSW, Australia. J. Hydrol. Environ. Res, 1(1), 49-59.

Hossain, A., Islam, M. A., Kamruzzaman, M., Khalek, M. A., & Ali, M. A. (2017). Forecasting carbon dioxide emissions in Bangladesh using Box-Jenkins ARIMA models. Department of Statistics, University of Rajshahi.

Ismail, N., & Abdullah, S. (2016). Principal component regression with artificial neural network to improve prediction of electricity demand. Int. Arab J. inf. Technol., 13(1), 196-202.

Jacobson, M. Z. (2014). Effects of biomass burning on climate, accounting for heat and moisture fluxes, black and brown carbon and cloud absorption effects. Journal of Geophysical Research: Atmospheres, 119(14), 8980-9002.

Lall, P., Deshpande, S., & Nguyen, L. (2016, May). Principal Components Regression Model for Prediction of Acceleration Factors for Copper-Aluminum Wirebonds Subjected to Harsh Environments. In 2016 IEEE 66th Electronic Components and Technology Conference (ECTC) (pp. 637-647). IEEE.

Mendeş, M. (2009). Multiple linear regression models based on principal component scores to predict slaughter weight of broiler. Archiv für Geflügelkunde, 73(2), 139-144.

Mishra, S., & Vanli, O. A. (2016). Remaining useful life estimation with lamb-wave sensors based on wiener process and principal components regression. Journal of Nondestructive Evaluation, 35(1), 11.

Nyoni, T., & Bonga, W. G. (2019). Prediction of co2 emissions in india using arima models. DRJ-Journal of Economics & Finance, 4(2), 01-10.

Ogeya, M., Nyambane, A., & Wanjiru, H. (2018). Technical report: Nordic Green to Scale for countries: Unlocking the potential of climate solutions in Kenya and Ethiopia. Nordic Council of Ministers.

Rahayu, S., Sugiarto, T., Madu, L., & Subagyo, A. (2017). Application of Principal Component Analysis (PCA) to Reduce Multicollinearity Exchange Rate Currency of Some Countries in Asia Period 2004-2014. International Journal of Educational Methodology, 3(2), 75-83.

Rahman, A., & Hasan, M. M. (2017). Modeling and forecasting of carbon dioxide emissions in Bangladesh using Autoregressive Integrated Moving Average (ARIMA) models. Open Journal of Statistics, 7(4), 560-566.

Ravishankara, A. R., Daniel, J. S., & Portmann, R. W. (2009). Nitrous oxide (N2O): the dominant ozone-depleting substance emitted in the 21st century. science, 326(5949), 123-125.

Shahzad, U. (2015). Global warming: Causes, effects and solutions. Durreesamin Journal, 1(4).

Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environmental Modelling & Software, 22(1), 97-103.

DOI: 10.3844/erjsp.2021.1.12

Thupeng, W. M., Mothupi, T., Mokgweetsi, B., Mashabe, B., & Sediadie, T. (2018). A Principal Component Regression Model, for Forecasting Daily Peak Ambient Ground Level Ozone Concentrations, in the Presence of Multicollinearity Amongst Precursor Air Pollutants and Local Meteorological Conditions: A Case Study of Maun. International Journal of Applied Mathematics & Statistical Sciences (IJAMSS) Vol, 7, 1-12.

Tiemeyer, B., Albiac, B., E., Auustin, J., Bechtold, M., Beetz, S., Beyer, C., Drosler, M., Ebil, M., Eickenscheidt, T., Fiedler, S. and Forster, C. (2016). High Emissions of Greenhouse Gases from Grasslands on Peat and Other Oranic Soils. Global Change Biology, 22 (12), 4134-4149.