Statistical Modeling of Emission Factors of Fossil Fuels Contributing to Atmospheric Carbon Dioxide in Africa

Mohamed Ali Abu Sheha[1,2], Christ P. Tsokos[1]

[1]Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA
[2]Department of Health Informatics, Faculty of Public Health, University of Benghazi, Benghazi, Libya
Email: mabusheha@mail.usf.edu, ctsokos@usf.edu

Abstract

Global warming is majorly caused by an increase in atmospheric temperature and carbon dioxide (CO_2) emissions due to the rise in the temperature. The continued accumulation of CO_2 into the atmosphere is a massive part of the climate change problem. This study aims to develop a data-driven statistical model using Africa's fossil-fuel CO_2 emissions real data to identify the significant attributable variables and their interaction that produce the carbon dioxide emissions. However, we have considered five attributable variables in our statistical modeling and they are **Liquid fuels** (Li), **Solid fuels** (So), **Gas fuels** (Ga), **Gas flares** (Gf) and **Cement production**. The development of the statistical model that contains the different emissions of fossil fuels and their interactions have been specified and ranked based on a percentage of their annual contributions to carbon dioxide in the atmosphere. Our proposed statistical model is compared with a different penalization method since multicollinearity among the risk factors exists and which provided excellent results according to the root mean square errors (RMSE) statistic. The results of the proposed model are compared to previous results of different countries of the world.

Keywords

Global Warming, Greenhouse Effect, Statistical Modeling, Africa Continent

1. Introduction

Global warming is not a phenomenon that could happen; it is a phenomenon that is happening. We are witnessing the effects of climate changes in the Arctic ice levels that have been the lowest since scientists have ever recorded. The cir-

culation of radiation that warms the earth is referred to as the greenhouse effect and the gases involved are called the greenhouse gases, which mainly include Carbon Dioxide, Methane, Water Vapor, Chlorofluorocarbons, etc.

Recently, the UN's Intergovernmental Panel on Climate Change (IPCC, 2018) [1], reports that human activities are expected to add approximately 1.0˚C of global warming above the pre-industrial age levels, with a probable range of 0.8˚C to 1.2˚C. The IPCC also reports that, as early as 2030 the planet will reach 1.5˚C (2.7˚F) above pre-industrial levels, leading to wildfires, extreme drought, floods and poverty for hundreds of millions of people. Furthermore, global temperatures are already measuring about 1.0˚C which means the planet is two-thirds of the way there.

The warmest year on record since 1850 is 2016 with a central estimate of 1.15˚C above the same baseline [2]. Scientists around the globe have gathered tons of evidence telling us that the earth is rapidly warming up. They believe that as the concentration of carbon dioxide in earth's atmosphere CO_2 increases, so is the temperature, and both are directly connected. However, the latest report that has been prepared by the UK's Met Office Hadley Centre (office's Richard Betts 2019) [2], pointed out that in 2019 the average CO_2 concentration in the earth's atmosphere is expected to increase by 2.8 ppm to reach 411 ppm, and that it will be the most significant rise in the concentration of atmospheric carbon dioxide in 62 years of records.

Indeed, carbon dioxide is released into the atmosphere from both natural and human's emission such as fossil fuels that people are burning for energy. Besides, in 2017 Li *et al.* [3] have found that economic growth, resident population growth, and energy intensity enhancement were the major significant growth factors of carbon emissions in Beijing.

This is consistent with the IPCC report (April 2007) [4] that "Africa was not acting quickly enough to stem the dire economic and environmental consequences of greenhouse gas emissions". Continuing with the same report South Africa has been ranked as the 13th largest Carbon Dioxide emitter among all the countries in the world in 2008 based on the record of fossil-fuel CO_2 consumptions and cement productions with 119 million metric tons of carbon CO_2 emissions. Thus, South Africa is considered the largest CO_2 emitting country on the continent of Africa.

According to the United Nations Fact Sheet on Climate Change [5], Africa is the continent's most vulnerable into the impacts of climate change. Most vulnerable are the Seychelles islands, Cape Verde, and Mauritius, as well as large African deltas such as the Niger Delta, Nile delta in Egypt, the Kalahari and Okavango deltas in Botswana. Most of the continent already is experiencing temperature increases of approximately 0.7˚C, and with predictions that the temperatures will rise further, in addition Africa is facing a wide range of impacts, including increased drought and floods. The impact of climate change has already aggravated parts of Africa. For example, in the large basins the total available water in Senegal, Lake Chad, and Niger has decreased by 40 to 60 per-

cent and many climate models project declining precipitation in the already-dry regions of Southern Africa [5].

Regions that are facing inadequate supply of water, especially in North Africa, would have climate change further threatening sustainable development due to demands of water. On the other hand, African countries that are affected by AIDS, HIV, fighting poverty, political instability, internal/civil wars, drawbacks in policy making and economic reforms may lack the funds/resources to tackle these expected significant climate change problems.

Usually, atmospheric CO_2 concentrations that are emitted from fossil fuel combustion and industrial operation are divide into seven sources [6] (based on the chemical form of fossil-fuels) namely: **Solid fuels** (So) include wood, charcoal, coal, and others; **Liquid fuels** (Li) is the gasoline that we regularly use to create mechanical energy, **Gas fuels** (Ga) carry gas consisting essentially of methane, and **Gas flares** (Gf) are the vertical stack on oil wells or natural gas well completion activities. **Cement production** (Ce), **oxidation of non-fuel hydrocarbons** (Hy), and fuel from **bunkers** (Bu) used for shipping and air transportation. Thus, these seven emissions are considered as the attributable variables to the atmospheric CO_2 concentration in our statistical modeling with their interactions. **Bunkers** (Bu) and **oxidation of non-fuel hydrocarbons** (Hy), information are not available in the Africa data base, so our model is utilizing five attributable variables in this study.

In the present study, the real yearly CO_2 emissions data for each of the fossil-fuels for the African continent obtained from Carbon Dioxide Information Analysis Center (CDIAC), and this actual annual data has been collected from 1963 to 2014. All emission estimates are shown in metric tons of carbon (MT). In developing the statistical model, the response variable is the CO_2 in the atmosphere; hence, we develop an analytical model that contains the significant contributable variables and important interactions along with higher order of contributions if applicable.

The proposed model relies on several assumptions such as the linearity, multicollinearity, and the normality assumption that related to errors. Carbon dioxide dataset shows that the attributable variables are highly correlated; thus, the parameters are challenging to interpret. The parameters become very unstable when independent variables are highly correlated and leading to experiencing over-fitting the model. Moreover, we apply different penalization regression methods: Ridge Regression (L2) [7], Lasso Regression (L1) [8], and Elastic net (EN) [9]. These methods are widely used to address over-fitting of the model.

The proposed statistical model is useful in predicting the CO_2 in the atmosphere given the values of the significant attributable variables. Also, we rank the attributable variables according to the percent of contribution to CO_2 emissions in the atmosphere. The validation and quality of the proposed analytical model have been statistically evaluated using R square (R^2), R square adjusted (R_{adj}^2), root mean square error (RMSE) statistic and residual analysis. Eventually, its usefulness has been illustrated by utilizing different combinations of various at-

tributable variables.

To our knowledge, no such statistical model has been developed under the proposed logical structure in Africa. Also, we wanted to rank the explanatory variables according to their CO_2 contributions in the atmosphere and likely comparing them with those of the United States [10] [11], European Union [12], South Korea [13], and the Middle East [14]. Therefore, looking for an appropriate statistical model in predicting of carbon emissions is imperative.

2. Methodology

2.1. The Data

The CO_2 emission data was obtained from Carbon Dioxide Information Analysis Center (CDIAC), located at Oak Ridge National Lab (Division of US Department of Energy). The plot of the yearly CO_2 emissions in the atmosphere is shown in Figure 1, below.

The African CO_2 emissions show an increasing pattern over the years 1964 to 1988. However, the years from 1990 to 2005 show nonstationary phenomena behavior in CO_2 emissions as a function of time. The period 2002 to 2008 show a noticeable increase in CO_2 emissions before a slight decrease in the years 2010 to 2013. This was probably due to the socio-economic and political crises that Africa was experiencing during these periods.

In developing the statistical model for CO_2 emissions as a function of the attributable variables, one of the underlying assumptions is that the response variable should follow the Gaussian probability distribution. The mid-values of CO_2 in the atmosphere seem to be reasonably straight, but the ends are somewhat skew which can be seen from the QQ plot in Figure 2, below. The goodness-of-fit testing (Shapiro-Wilk normality test, A p-value = 8.952e−05) that the subject data does not follow the normal probability distribution as well. Therefore, the QQ plot supports the fact that natural phenomena such as atmospheric CO_2 are not following the Gaussian probability distribution.
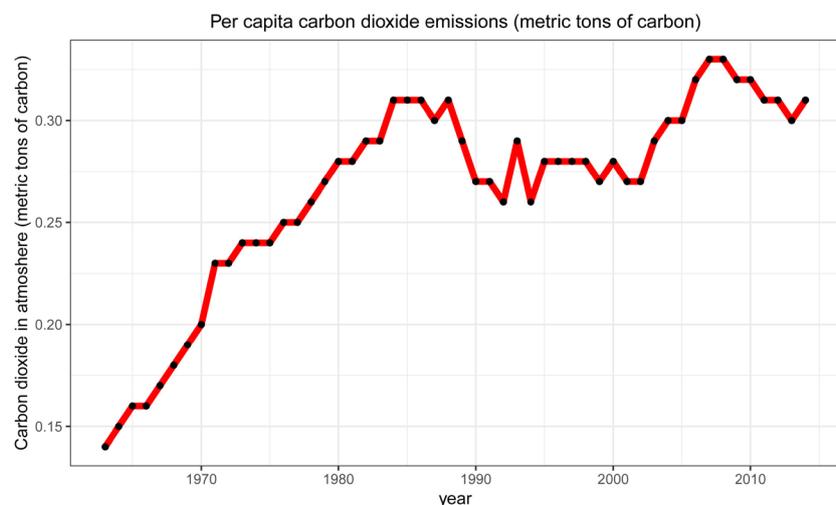
Figure 1. Annual CO_2 emission in Africa in metric tons from 1964 to 2014.
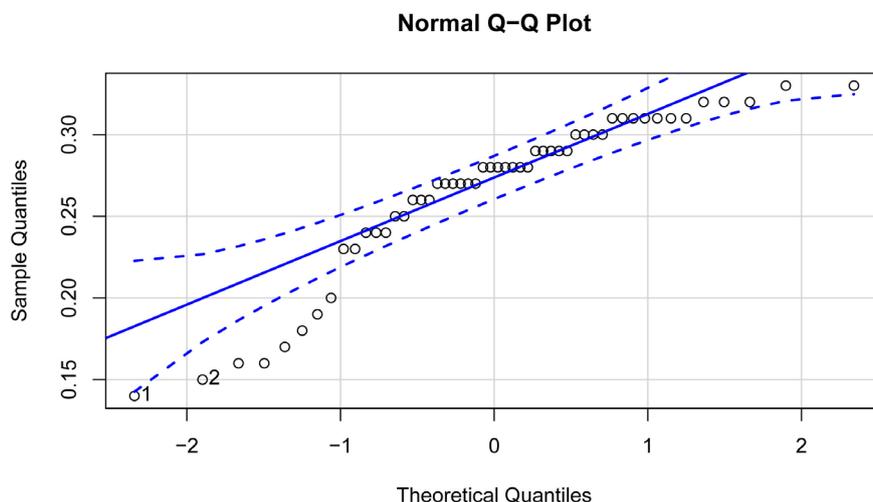
Normal Q−Q Plot

Figure 2. QQ plot for testing normality.

The collinearity assumption of the model is shown in **Figure 3**, where negative correlations displayed in red and positive correlations in blue color. Color intensity and the degree of the relationship between each pair are proportional to the correlation coefficients. Thus, the variables *Gas-Fuels* (Ga), *Solid-Fuels* (So), *Liquid-Fuels* (Li), *Bunker-Fuels* (Bu), and *Cement* (Ce) have a positive high correlation, so at this point we would consider the regularization techniques such as Ridge Regression (L2), Lasso Regression (L1) and Elastic net penalties to address over-fitting. Hence, there are enough statistically significant relationships (Linearity) between CO_2 and Africa's fossil-fuel CO_2 emissions to build a high-quality multiple regression model.

However, a schematic diagram [15] that shows the relationship between the attributable variables and carbon dioxide in the atmosphere is shown in **Figure 4**.

2.2. Statistical Modeling

A statistical model describes the relationship of the response variable, (*i.e.* whose content we are trying to model) with the attributable variables. We proceed to develop the statistical model which is given by CO_2 in the atmosphere as a function of the five attributable variables and all possible interactions as previously presented. One of the pure forms of a model with all possible interactions and additive error structure, in the given particular case, could be expressed as follows:

$$CO_2 = \beta_0 + \sum_i \alpha_i x_i + \sum_j \gamma_j k_j + \varepsilon_i, \tag{1}$$

here β_0 is the intercept of the model, α_i is the coefficient of ith individual attributable variable x_i, γ_j is the coefficient of jth interaction term k_j, and ε_i denotes the random disturbance or residual error of the model.

One of the underlying assumptions to construct the above model is that the response variable should follow the Gaussian probability distribution. As we
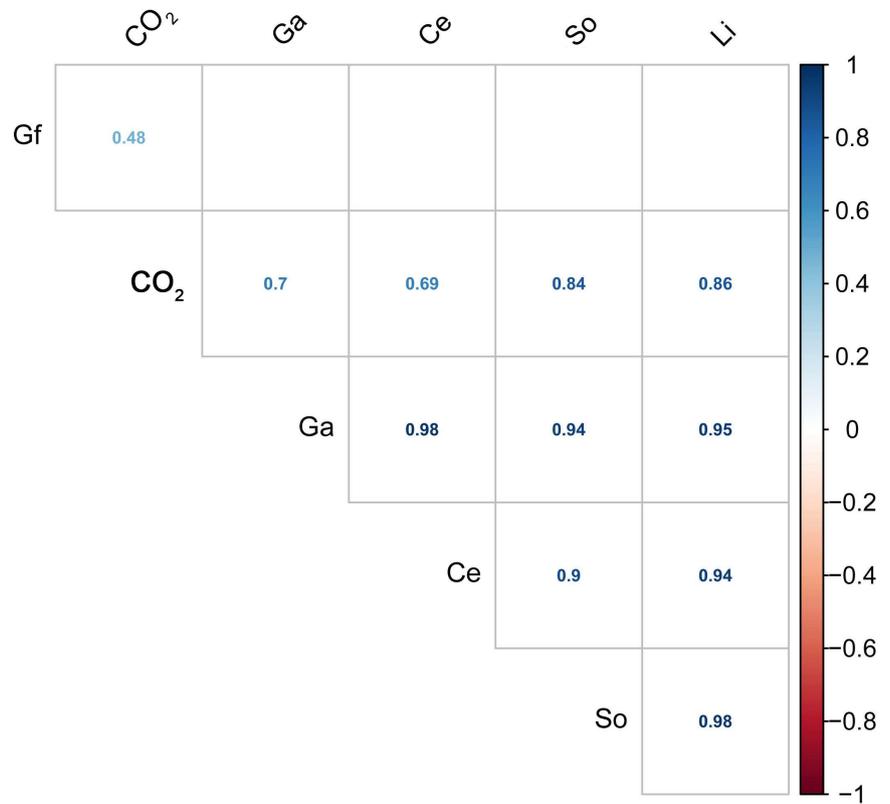
Figure 3. Correlation matrix of carbon dioxide and fossil fuel sources.
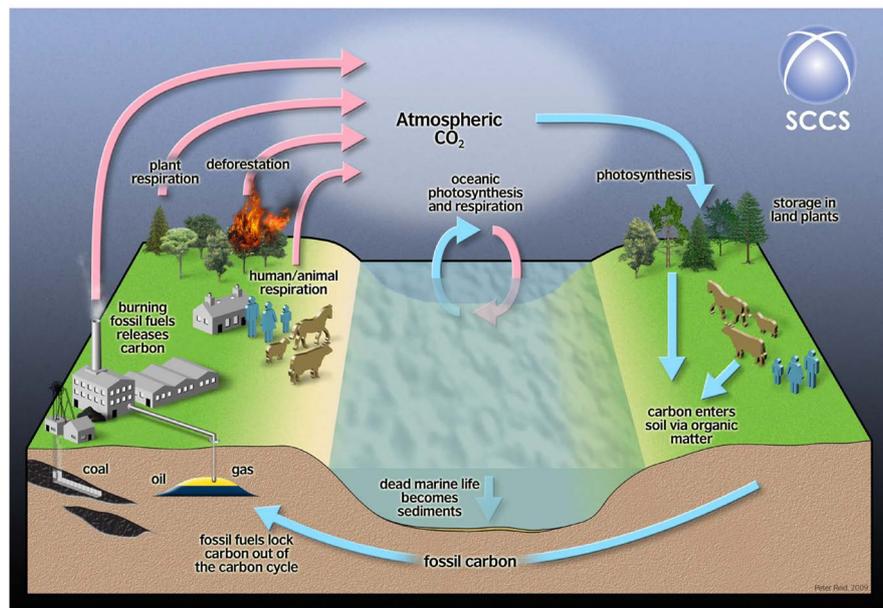
**Figure 4.** A schematic view of carbon dioxide in the atmosphere.

illustrated above, the dependent variable CO_2 emission does not follow the Gaussian probability distribution. Therefore, we must utilize the Johnson Transformation [16] to the carbon dioxide data to filter the data to follow normal probability distribution, which results in Equation (2), below:

$$z = \gamma + \delta \ln\left(\frac{x-\epsilon}{\lambda+\epsilon-x}\right),$$

and

$$\text{TCO}_2 = -4.4174 + 1.4857 \ln\left(\frac{\text{CO}_2 - (-0.9845)}{0.3444 - \text{CO}_2}\right). \tag{2}$$

Hence, TCO_2 represents the new response variable after the Johnson Transformation has been applied. Again, we check the normality condition on the TCO_2 data, which then follows the normal probability distribution as is clearly seen by **Figure 5**, thus, we proceed to estimate the approximate coefficients (weights) of the actual contributable variables for the transformed CO_2 atmosphere data in the Equation (2).

In order to develop our statistical model, we begin with the full statistical model, which included all five attributable variables as previously defined and ten possible interactions between each pair. Thus, initially, we start structuring our model with fifteen total terms that include the primary contribution of attributable variables and all possible interactions.

Since we started with the full statistical model (fifteen terms), as we mentioned above, we shall apply the backward elimination process to determine the significant contributions of both the individual attributable variables and interactions. Moreover, backward elimination is considered one of the best traditional methods in the case of having a small set of features to tackle overfitting and perform feature selection [17].

However, the estimation process of our statistical analysis has shown that four out of five risk factors significantly contribute and seven interaction terms. Thus, the best proposed statistical model with all significant attributable variables and interactions that estimates accurate CO_2 emissions in the atmosphere in Africa is given by Equation (3), below.
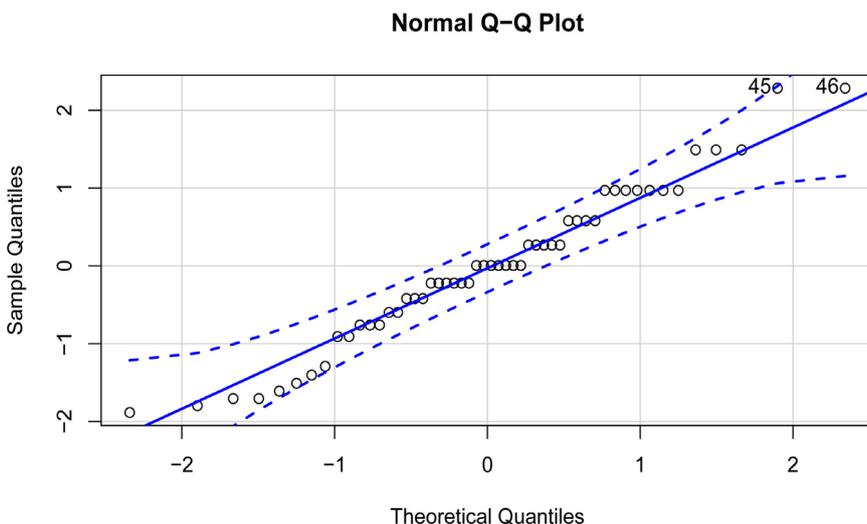
Figure 5. QQ plot for testing the normality of TCO_2.

$$\hat{T}CO_2 = -6.658 \times 10^{00} + 1.800 \times 10^{-4}\, So + 1.145 \times 10^{-4}\, Li$$
$$-8.821 \times 10^{-6}\, Ga - 1.949 \times 10^{-4}\, Ce$$
$$+1.259 \times 10^{-4}\, Gf - 1.276 \times 10^{-9}\, So.Li$$
$$-4.896 \times 10^{-9}\, So.Ga + 1.849 \times 10^{-8}\, So.Ce \qquad (3)$$
$$-6.158 \times 10^{-9}\, So.Gf + 5.616 \times 10^{-9}\, Li.Ga$$
$$-1.137 \times 10^{-8}\, Li.Ce + 5.723 \times 10^{-8}\, Ce.Gf$$

The TCO_2 estimate is obtained from Equation (3) is based on the Johnson transformation of the data, thus we will utilize the anti-transformation on Equation (3) to estimate the desired, actual CO_2 emissions in the atmosphere as follows:

$$\hat{C}O_2 = \frac{-0.05 + e^{0.673 \times \hat{T}CO_2}}{0.051 + e^{0.673 \times \hat{T}CO_2}}. \qquad (4)$$

The proposed model will help scientists understand how the typical value of the carbon dioxide emissions in the atmosphere in Africa changes when any one of the five attributable variables is varied, while the other attributable variables are held fixed. Similarly, with the significant interaction. Most commonly, it will estimate the conditional expectation of the carbon dioxide emissions given the attributable variables. Furthermore, we illustrate the percentage that the attributable variables and the interactions contributing to CO_2 in the atmosphere by Figure 6, below.
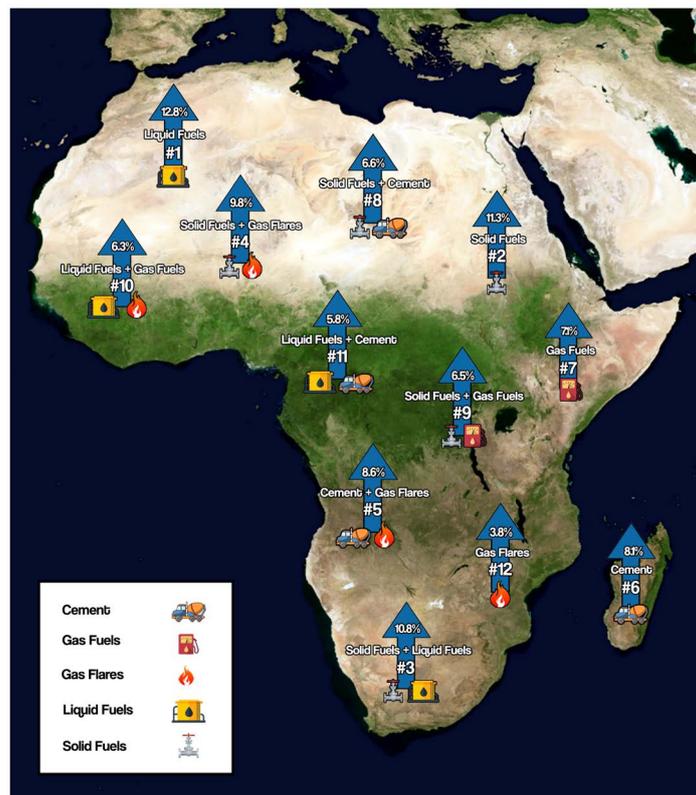
Figure 6. CO_2 in the atmosphere variable contribution diagram.

To assess the quality of the proposed statistical model we use both the coefficient of determination, R^2 and adjusted R^2 which are the key criteria to evaluate the model fitting.

The regression sum of squares (SSR), is a measure of the variation that is explained by the proposed model. The sum of squared errors (SSE), also called the residual sum of squares, is the variation that is left unexplained. The total sum of squares (SST) is proportional to the sample variance and equals the sum of SSR and SSE. The coefficient of determination R^2 is defined as the proportion of the total response variation that is explained by the proposed model and it measures how well the regression line approximates the real data points. Thus, R^2 is given by

$$R^2 = 1 - \frac{SSE}{SST}.$$

However, R^2 itself does not consider the number of variables in the model, plus there is that sticky problem of the ever increasing R^2. The R^2 adjusted will adjust for degree of freedom of the model and considers the number of parameters. The R^2 adjusted is

$$R^2_{adj} = 1 - \frac{SSE/df_{error}}{SST/df_{total}}.$$

For our final statistical model, the R squared is 0.9728 and R squared adjusted is 0.9644. Both R squared and R squared adjusted are very high (more than 90%) and very close to each other. That is, the developed statistical model explains 97.28% of the variation in the response variable, a very high-quality model. Similarly, the risk factor that we included in the model along with the relevant interactions estimate 97% of the Africa CO_2 emissions (metric tons per capita) in the atmosphere. These results show that the increase of the value of R squared is not due to the increase in the number of the predictors but to the good quality of the proposed statistical model.

In Table 1, we rank the individual attributable variables and interactions with

Table 1. Rank of variable according to their contributions.

Rank	Variables
1	Liquid Fuels
2	Solid Fuels
3	Liquid Fuels ∩ Solid Fuels
4	Solid Fuels ∩ Gas Flares
5	Cement ∩ Gas Flares
6	Cement
7	Gas Fuels
8	Solid Fuels ∩ Cement
9	Solid Fuels ∩ Gas Fuels
10	Liquid Fuels ∩ Gas Fuels
11	Liquid Fuels ∩ Cement
12	Gas Flares

respect to their contribution of CO_2 in the atmosphere. That is, (we listed those terms based on their percentage of contribution to CO_2 in the atmosphere) as we expected, Li ranks number one which is one of the risk factors from the emissions from fossil fuels.

Again the percentage of their contributions is shown in Figure 6.

Penalized Regression Models

The presence of collinearity which leads to overinflating the standard errors of the estimated coefficients; as well as it makes some attributable variables statistically insignificant when they should be significant and stable. Basically, in developing the proposed statistical model for CO_2 emissions, the ordinary least squares method (OLS) has been used to obtain an approximate estimate of the coefficients of the contributable variables.

To address the multicollinearity problem, the Regularization methods are used and whereas these methods are based on adding the regularization parameter (two small penalty equal λ and α) to the regression coefficients of the individual attributable variables, so that the model generalizes the data and prevent over-fitting. This can be explained with a cost function of the form

$$CO_2 = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2.$$

Hence, we can characterize these proposed developed models into three categories as following: Ridge regression regularization method that adds squared magnitude of coefficient as penalty term to the loss function that can be explained by

$$CO_2 = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2, \tag{5}$$

where Lasso regression method, adds absolute value of magnitude of coefficient as penalty term to the loss function that can be expressed by

$$CO_2 = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right|, \tag{6}$$

and the Elastic Net regression method which is the mix of Ridge and Lasso technique can be defined by

$$CO_2 = \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \left[(1-\alpha)\sum_{j=1}^{p} \beta_j^2 + \alpha \sum_{j=1}^{p} \left| \beta_j \right| \right]. \tag{7}$$

However, in the above Equations (5)-(7) the constructions of the three models will be the same structure as our proposed model in Equation (1) with only the coefficient estimation will be different because of the randomness of choosing the training data set. Also, they will include optimal two hyper-parameters, which are lambda = 0.0001 and alpha = 1 (penalty term) that give the smallest RMSE, as shown in Table 2, below.

Table 2. Different techniques with respect to RMSE.

Technique	RMSE
Proposed Model	0.261
Ridge Model	0.484
Lasso Model	0.307
Elastic Net Model	0.307

2.3. Validation of the Proposed Models

We utilize two methods to perform the model validation. The first method is to use the proposed model to calculate the predicted value for each individual data, CO_2, and then calculate the residuals.

Thus, the residual analysis of the complete model used to attest the quality of the developed statistical model, that is, the observed annual CO_2 emission in the atmosphere (response) minus the model estimate of CO_2 emission.

The residual analysis also justifies the model assumptions of normality and constant error variance. For the developed statistical model, where the mean residual is equal **zero** indicates that the predictions from our statistical model are very good, variance of the residual is **0.03**, standard deviation is **0.16** and standard error of the residuals is **0.19**, that are very good statistics that support the high quality of the model. The results are shown in Q-Q plot in **Figure 7** and scatter plot in **Figure 8**, below.

From the Q-Q plot, we can clearly see an approximate normality distribution of the residual within 95% confidence interval and the scatter plot illustrates an approximate zero mean and no clear pattern or trend in the residuals.

The second method we will utilize repeated cross-validation. The basic idea is; we will use 10-fold cross-validation, then just repeating cross-validation five times where in each of the repetition folds are split differently. In 10-fold cross-validation, the training set is divided into ten equal subsets. One of the subsets is taken as a testing set in turn and (10-1) subsets are taken as a training set in the proposed model.

Besides, after each repetition of the cross-validation, the model assessment metric is computed, whereas root mean square errors (RMSE) selected as the cost function, which is given by:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}. \tag{8}$$

We construct our model using only the training set, and the constructed model will have the same structure as our proposed model with only the weights of the attributable variables will be different. To enhance the reliability of the training results; we use this model to predict the CO_2 value using the testing sets of the attributable variables. However, we repeated this procedure to verify which regularization technique can be considered to improve the prediction and
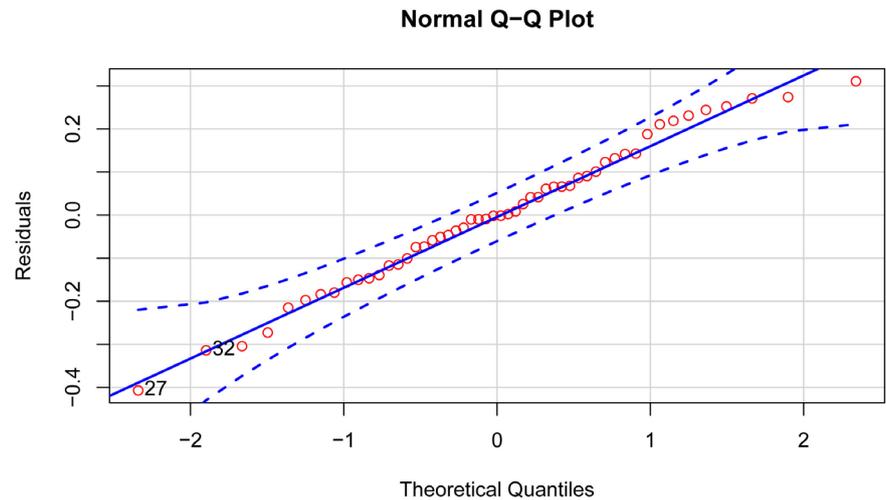
Normal Q−Q Plot

Figure 7. QQ plot for testing the normality of residuals with 95% confidence limits.

Residuales Plot
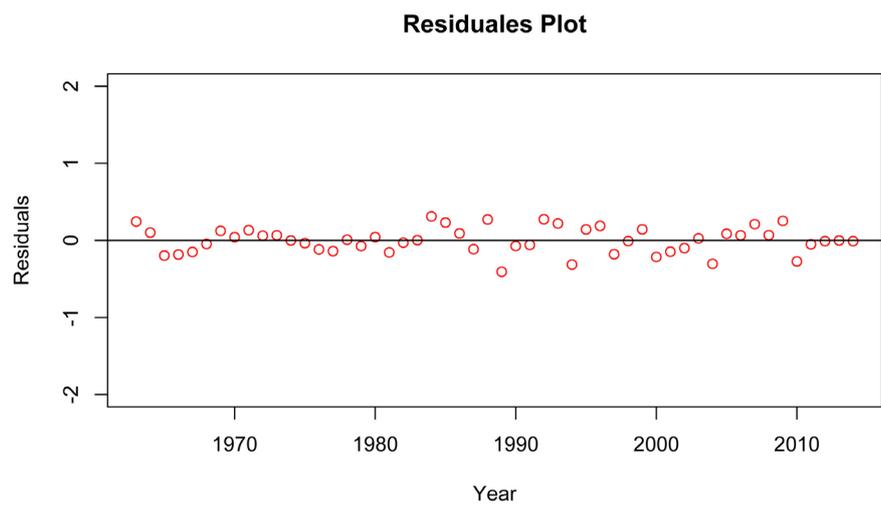
Figure 8. Scatter plot for testing the pattern of residuals.

then compare it with our proposed model we had on the RMSE. The results are shown in **Table 2**, above.

We compare the statistical models in terms of the root mean square errors; RMSE, of the prediction of the CO_2. The proposed nonlinear statistical model performed better than the other models with the smallest RMSE 0.261. Also, since the hyper-parameter α tuning using cross-validation in Equation (7) equal one, the RMSE was the same in both methods Lasso and Elastic net. Thus, the proposed underlying statistical model is very high in quality to predict CO_2 in the atmosphere.

2.4. Results and Discussion

- **Ranking of the Contributing Variables—Africa**

We use the R^2 criteria to rank the attributable variables along with the significant interactions with respect to the percent of contribution of CO_2 emissions in

the atmosphere. Table 3 below shows the rankings of these risk factors along with their percent of the overall contribution.

The risk variable that has the biggest contribution to the CO_2 emission in Africa is **Liquid-Fuels**, which contributes 13% of the CO_2 emission. The next largest contribution is **Solid-Fuels** with 11% contribution. Note that numbers (rankings) 3, 4, and 5 are interactions of Li ∩ So, So ∩ Gf, and Ce ∩ Gf, respectively. Hence, summing these risk factors up we identify that they contribute 97.5% of CO_2 emissions in Africa.

• **Ranking of the Contributing Variables—United States**

Xu and Tsokos [10] [11] structured a nonlinear statistical model that identified the significant risk factors along with the significant interactions that contribute to the CO_2 in the atmosphere in the continental United States. The ranks of the contributing variables with the rate of CO_2 contribution in the atmosphere are listed in Table 4. Thus, these variables and interactions contribute 98.98% of emissions in United States.

Table 3. Rank of attributing variables (Africa).

Rank	Variables	Contribution (%)
1	Liquid Fuels (Li)	12.8
2	Solid Fuels (So)	11.3
3	Li ∩ So	10.8
4	So ∩ Gf	9.8
5	Ce ∩ Gf	8.6
6	Cement (Ce)	8.1
7	Gas Fuels (Ga)	7.1
8	So ∩ Ce	6.6
9	So ∩ Ga	6.5
10	Li ∩ Ga	6.3
11	Li ∩ Ce	5.8
12	Gas Flares (Gf)	3.8

Table 4. Rank of attributing variables (USA).

Rank	Variables	Contribution (%)
1	Liquid Fuels (Li)	17.59
2	Li ∩ Ce	16.36
3	Ce ∩ Bu	15.73
4	Bunker Fuels (Bu)	15.06
5	Cement	10.77
6	Gas Flares (Gf)	8.95
7	Gas Fuels (Ga)	6.82
8	Ga ∩ Gf	5.43
9	Li ∩ Ga	2.25
10	Li ∩ Bu	0.02

- **Ranking of the Contributing Variables—European Union**

In 2013, Teodorescu and Tsokos [12] developed a data driven nonlinear statistical model using CO_2 emissions data for the European Union Countries (EU). They have found that Gas-Fuels contribute 48.72% of the overall CO_2 emissions. Table 5 below contains the other individual contributions of CO_2 emission along with the significant contributing interactions for EU.

- **Ranking of the Contributing Variables—South Korea**

Similarly, in 2015, Kim and Tsokos [13] have structured a data driven statistical model that identified the individual attributable variables along with significant interactions terms that contribute to atmospheric in South Korea. Their proposed statistical model explained 99.41% of the CO_2 in the atmosphere. The ranking of the explanatory variables and significant interactions with their percentages of overall contribution are shown in Table 6, below.

- **Ranking of the Contributing Variables—Middle East**

Recently, Habadi and Tsokos [14] have built a nonlinear statistical model using CO_2 emissions data for the Middle East Countries (ME). They identified that Gas-Fuels contributes 95% of the overall CO_2 emissions. Table 7, below illustrates the other individual contributions of CO_2 emission along with the significant contributing interactions for ME.

Table 5. Rank of attributing variables (EU).

Rank	Variables	Contribution (%)
1	Gas Fuels (Ga)	48.72
2	Li ∩ Bu	12.41
3	Li^2	11.79
4	Bu^2	7.78
5	Gas Flares (Gf)	6.66
6	Li ∩ Gf	5.06
7	Li ∩ Bu	4.71
8	Liquid Fuels (Li)	2.86

Table 6. Rank of attributing variables (South Korea).

Rank	Variables	Contribution (%)
1	Liquid Fuels (Li)	75.37
2	Solid Fuels (So)	18.61
3	So ∩ Bu	2.008
4	Ga ∩ Bu	1.534
5	Li ∩ Bu	0.912
6	Bunker Fuels (Bu)	0.47
7	Gas Fuels (Ga)	0.224
8	Li ∩ So	0.207
9	Li ∩ Ga	0.062
10	Li ∩ So ∩ Bu	0.004

Table 7. Rank of attributing variables (Middle East).

Rank	Variables	Contribution (%)
1	Cement (Ce)	15.28
2	Gas Fuels (Ga)	14.7
3	Li ∩ So	13.66
4	Ga ∩ So	13.47
5	Ga ∩ Ce	12.56
6	Liquid Fuels (Li)	10.63
7	So ∩ Gf	9.65
8	Gf ∩ Ce	7.9

- **Global Comparison: USA, EU, S. Korea, ME and Africa**

Table 8, below gives an interesting comparison of what contributes to the CO_2 emissions in the atmosphere in the United States, European Union, South Korea, Middle East, and Africa. An important fact from this comparison is that 12.8% of the CO_2 emissions in Africa, 17.59 in the US and 75.37% in South Korea are caused by Liquid fuels, whereas in the EU and ME **Liquid fuels** contribute to only 2.86% and 10.63% of emissions, respectively.

Furthermore, Liquid-Fuels is the number one attributable variable of the emission of CO_2 in the atmosphere in Africa, the US, and South Korea, whereas it is the last in the EU and the 6th in the Middle East.

Moreover, Gas-fuels ranked as the number one attributable variable in the EU; however, it is the 7th in Africa, the US, and South Korea with a contribution 7.1%, 6.82%, and 0.224% respectively while in the Middle East is ranked as number Two with only 14.7% contribution.

Similarly, Cement is ranked as the number one attributable variable in the Middle East; however, it is the 6th in Africa with a contribution 8.1%, whereas it is the 5th in the US with a contribution 10.77%.

As well, it is interestingly to identify that Africa has seven significant contributing interactions of the risk factors while the US and South Korea identified five, whereas the Middle East has Four significant interactions and EU has only three contributing interactions to CO_2 emissions.

3. Conclusions

In the present study we investigated fossil fuels risk factors that contribute to the widespread of the most common air pollutant namely carbon dioxide in the atmosphere in Africa. Previous data obtained from Carbon Dioxide Information Analysis Center (CDIAC) shows that there are five attributable variables that are contributing to the emission of carbon dioxide into the atmosphere in Africa. These attributable variables are **Liquid fuels** (Li), **Solid fuels** (So), **Gas fuels** (Ga), **Gas flares** (Gf) and **Cement production**, in addition to seven interaction among them.

Table 8. Global comparison of ranks in five continents.

Rank	USA	EU	S. Korea	ME	Africa
1	Li	Ga	Li	Ce	Li
2	Li ∩ Ce	Li ∩ Bu	So	Ga	So
3	Ce ∩ Bu	Li²	So ∩ Bu	Li ∩ So	Li ∩ So
4	Bu	Bu²	Ga ∩ Bu	Ga ∩ So	So ∩ Gf
5	Ce	Gf	Li ∩ Bu	Ga ∩ Ce	Ce ∩ Gf
6	Gf	Li ∩ Gf	Bu	Li	Ce
7	Ga	Li ∩ Bu	Ga	So ∩ Gf	Ga
8	Ga ∩ Gf	Li	Li ∩ So	Gf ∩ Ce	So ∩ Ce
9	Li ∩ Ga	-	Li ∩ Ga	-	So ∩ Ga
10	Li ∩ Bu	-	Li ∩ So ∩ Bu	-	Li ∩ Ga
11	Li ∩ Bu	-	Li ∩ So ∩ Bu	-	Li ∩ Ce
12	-	-	-	-	Gf

In our study, we build a data-driven statistical model in which we discovered that all five attributable variables significantly contribute to the emission of carbon dioxide in the atmosphere along with seven significant interactions which were unknown to be part of factors that significantly cause the emission of the carbon dioxide in the atmosphere of the Africa continent.

The identification of the significance of the five attributable variables and the seven interactions were based on a well-structured statistical data analysis. The data we obtained did not follow the Gaussian probability distribution. We then used the Johnson transformation to transform the response variable (*i.e.* carbon dioxide) to make it Gaussian, so that we could proceed with statistical modeling.

There was the presence of multicollinearity among the risk factors. However, our model was compared with a different penalization technique which provided very good results according to the RMSE statistic. In statistical modeling, specifically in regression modeling, the parameter coefficients and p-values are affected by multicollinearity. However, this does not affect our predictions and how precisely the predictions are, as well as the goodness of fit of our model. We do not have to be concerned about the severity of multicollinearity in our model if our main aim is to make predictions [18].

The proposed model has high predictive accuracy which is supported by the high values of R^2 and adjusted R^2. Furthermore, we ranked the attributable variables in descending order by their percentages of contribution to the emission of CO_2 in the atmosphere. **Liquid fuel** was ranked the highest contributor of the emission of CO_2 in Africa representing 12.8%, whereas **Gas flares** is the least contributor with 3.8%. Interestingly, countries like the United States and South Korea also have Liquid fuel as the leading cause of CO_2 in air [10] [11] [13].

We can address the usefulness of the proposed model in the subject area. First,

we can obtain excellent predictions of CO_2 emissions in the atmosphere given the values of the attributable variables. Second, we identify the individual attributable variables. Third, we have identified the significant interactions that exist in the model. Fourth, we rank the individual attributable variables and interactions as a percentage of contribution in the response namely CO_2 emissions in the atmosphere.

Furthermore, having this proposed model one can proceed to perform surface response analysis that is with a high degree of accuracy what are the values of attributable variables that would be at the acceptable level which will not lead the CO_2 in the atmosphere to go above the critical value.

Thus, we want to obtain the values of those attributable variables, so that we will not exceed the specified value of CO_2 in the atmosphere. Thus, we want to be at least 95% certain what are the values of the attributable variables to be within the minimum appropriate, acceptable CO_2 in the atmosphere.

In addition, we cannot have a world policy for Global warming because we have studied five different regions of the world and seem to be responding differently with respect to CO_2. Our findings show that it would be a waste of time and resources to manage the world increasing global warming base through Global uniform policies. It is clear from our study that Global environmental policies are not applicable, but rather regional well-structured policies will address the world problem of Global warming.

Finally, our proposed statistical model is highly useful for decision making and strategic planning on controlling the air pollutant CO_2 in the atmosphere in Africa.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., *et al.* (2018) Global Warming of 1.5 ˚C: An IPCC Special Report on the Impacts.
https://report.ipcc.ch/sr15/pdf/sr15_spm_final.pdf

[2] Press Office (2019) Faster CO_2 Rise Expected in 2019.

[3] Li, J., Zhang, B. and Shi, J. (2017) Combining a Genetic Algorithm and Support

Vector Machine to Study the Factors Influencing CO_2 Emissions in Beijing with Scenario Analysis. *Energies*, **10**, 1520. https://doi.org/10.3390/en10101520

[4] Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. and Intergovernmental Panel on Climate Change (2007) Climate Change 2007: Synthesis Report. Contribution of Working Group I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers. Intergovernmental Panel on Climate Change (IPCC).

[5] United Nations (2006) United Nations Fact Sheet on Climate Change.

[6] Goreau, T.J. (1990) Balancing Atmospheric Carbon Dioxide. *Ambio*, **19**, 230-236. https://www.jstor.org/stable/4313702

[7] Hoerl, A.E. and Kennard, R.W. (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, **12**, 55-67. https://doi.org/10.1080/00401706.1970.10488634

[8] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B* (*Methodological*), **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[9] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B* (*Statistical Methodology*), **67**, 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

[10] Xu, Y. and Tsokos, C.P. (2013) Attributable Variables with Interactions That Contribute to Carbon Dioxide in the Atmosphere. *Frontiers in Science*, **3**, 6-13.

[11] Xu, Y. and Tsokos, C.P. (2011) Statistical Models and Analysis of Carbon Dioxide in the Atmosphere. *Problems of Nonlinear Analysis in Engineering Systems*, **2**, e1.

[12] Teodorescu, I. and Tsokos, C. (2013) Contributors of Carbon Dioxide in the Atmosphere in Europe.

[13] Kim, D. and Tsokos, C.P. (2013) Statistical Significance of Fossil Fuels Contributing to Atmospheric Carbon Dioxide in South Korea and Comparisons with USA and EU. *Journal of Applied Statistical Science*, **21**, 337-347.

[14] Habadi, M.I. and Tsokos, C.P. (2016) Statistical Analysis and Modeling of the Atmospheric Carbon Dioxide in the Middle East and Comparisons with USA, EU and South Korea. *SCIREA Journal of Environment*, **1**, 32-47.

[15] Reid, P. (2009) Scottish Carbon Capture and Storage.

[16] Farnum, N.R. (1996) Using Johnson Curves to Describe Non-Normal Rocess Data. *Quality Engineering*, **9**, 329-336. https://doi.org/10.1080/08982119608919049

[17] Hocking, R.R. (1976) A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics*, **32**, 1-49. https://doi.org/10.2307/2529336 https://www.jstor.org/stable/2529336

[18] Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., *et al.* (2005) Applied Linear Statistical Models. 5th Edition, McGraw-Hill, Boston.