Study on the Impact of Population and Economic Activities on Environmental Indicators through Pca

Junchao Zhang [1], Junmeng Yan [1], Xin Li [2]

[1] Guilin university of technology, Guangxi, 541004, China

[2] Guangxi Financial Vocational and Technical College, Guangxi, 530007, China

Abstract. Principal Component Analysis (PCA) is applied in environmental policy planning to effectively identify and assess key factors affecting carbon peaking and emissions reduction. By analysing multidimensional data covering GDP, population, average temperature, oil consumption and price, and CO2 emissions, PCA reveals complex correlations of environmental challenges. Preliminary regression analyses expose complex relationships between variables, with a particular focus on anomalous oil consumption data for 2020, emphasising the importance of precise analytical methods.The PCA highlights the significant impacts of economic activity, population and oil-related factors on environmental indicators by simplifying the high-dimensional data, retaining the key information, and extracting the principal components with the highest variance.The KMO and Bartlett test confirms the applicability of PCA, with variance interpretation and dendrograms guiding principal component selection to ensure adequate retention of data information. This approach enabled precise identification of factors affecting environmental indicators, construction of a weighted composite model to quantitatively analyse environmental impacts on a yearly basis, and strengthened the understanding of the overall role of population and economic activities on the environment. Overall, the application of PCA highlights its key role in refining complex environmental data, clarifying the drivers of carbon emissions and energy consumption, and in the development of targeted environmental policies, enriching the environmental planning toolkit and demonstrating the contribution of statistical techniques to environmental governance.

Keywords: Principal Component Analysis (PCA); Environmental Policy Planning; Peak Carbon; Data Analysis; Energy Consumption.

1. Introduction

The use of Principal Component Analysis (PCA) in the field of environmental policy planning is a compelling way to dissect and understand the multifaceted nature of environmental challenges. The methodological shift to PCA is based on a comprehensive data collection and pre-processing exercise that gathered a decade's worth of data on China's GDP by sector, demographic indicators, average temperature, oil consumption and prices, and carbon dioxide emissions. The first step in the analytical journey is to conduct regression analysis, which aims to explore the relationship between the independent and dependent variables, laying the foundation for predictive modelling and the identification of anomalies [1]. basis for predictive modelling and identification of anomalies. This process highlights the need for powerful statistical tools such as PCA when dealing with the complexity of factors that influence environmental outcomes. By standardising the data collected and using PCA, the analysis goes beyond traditional data reduction techniques and is able to extract the principal components that contain the core differences in the dataset. The strength of PCA lies not only in its ability to streamline complex datasets, but also in its ability to shed light on the dynamics of economic activity, population growth, and their collective impacts on environmental indicators such as carbon dioxide emissions and energy consumption. Through a detailed study of variance interpretation, factor loading and weighted composite model construction, this approach not only reveals the main drivers of environmental change, but also provides a quantifiable framework for policy analysis and strategic planning [2]. The validation of the applicability of PCA through KMO and Bartlett's test further reinforces its importance in environmental policy planning, providing policy makers with a refined perspective to identify, prioritise and address key levers for achieving sustainable environmental outcomes. The integration of PCA into environmental policy planning demonstrates the transformative potential of statistical techniques to enhance our understanding and management of environmental challenges, paving the way for informed and targeted policy interventions.

2. Related Work

The application of principal component analysis (PCA) in environmental policy planning has proved to be very effective in identifying and assessing key factors in peak carbon achievement and mitigation efforts. The methodology is based on a decade of detailed data collection and pre-processing, including variables such as gross domestic product (GDP) by sector, population size, average temperature, oil consumption and prices, and carbon dioxide (CO_2) emissions, highlighting the complexity of the environmental challenge and the multifaceted nature of the factors involved [3]. The initial steps of regression analysis helped to identify patterns and anomalies in the data, in particular flagging significant discrepancies in the 2020 oil consumption data, thus highlighting the need for powerful analytical tools like PCA for accurate environmental assessments.

The strength of PCA lies in its ability to reduce the dimensionality of complex datasets while retaining the most important information. This is particularly important in the field of environmental policy planning, where a myriad of factors interact to influence outcomes such as carbon emissions and energy consumption. By normalising the original dataset and constructing a correlation matrix, the PCA approach helps to extract the principal components that represent the greatest variance in the dataset [4]. The results of the study, particularly through the analysis of explained variance and factor loadings, revealed the significant impact of GDP, industrial sector, population and oil-related variables on key environmental indicators such as carbon dioxide emissions and energy consumption.

The applicability of PCA to the dataset was verified through KMO and Bartlett's test, which showed significant relationships between the variables, emphasising the effectiveness of PCA in environmental policy planning. In addition, interpreted ANOVA and dendrograms guided the selection of principal components, ensuring that more than 85 per cent of the information was retained [5]. This strategic reduction of data complexity allowed for the clear identification of the factors that have the greatest impact on emissions reductions and energy savings, highlighting the inverse relationship between oil consumption and winter temperatures, as well as the direct correlation between economic activity and CO_2 emissions.

A composite model was developed to quantify the composite scores for each year of the study by weighting them according to the variance explained by the principal components. The model not only provides insight into the time trends of environmental factors, but also serves as a powerful framework for policy analysis and planning. The weighted analysis further highlights the dominant role of the first principal component, reflecting the overall impact of demographic and economic activities on environmental outcomes.

In summary, the application of PCA in this context demonstrates its far-reaching role in distilling complex environmental data into actionable insights. It highlights the interconnections between economic activity, population dynamics and environmental impacts, providing a nuanced perspective for policy development [6]. The methodology identifies the main drivers of carbon emissions and energy consumption, providing a solid foundation for targeted policy interventions aimed at achieving sustainable environmental outcomes. This analysis not only enriches the methodological toolkit for environmental policy planning, but also exemplifies the key role of statistical techniques in addressing contemporary environmental challenges, demonstrating the critical role of PCA in promoting informed and effective environmental policy planning.

3. Data Analysis

Before solving the problem, we need to perform data collection and pre-processing. The data are collected from authoritative websites such as the National Statistical Yearbook and the Prospective Database, while the pre-processing is related in this paragraph. We collect data on China's GDP (including primary, secondary, and tertiary industries), population size, average temperature, oil consumption and prices, and CO2 emissions in the past ten years to solve this problem.

First, we perform regression analysis of the data. Regression analysis is a statistical method used to explore the relationship between independent and dependent variables. It predicts or explains changes in the dependent variable by building a mathematical model. In regression analysis, the independent variable is usually continuous, while the dependent variable can be continuous, discrete or binary [7]. The central objective is to determine the functional relationship between the independent and dependent variables and use this relationship to make predictions and explanations. Common regression methods include linear regression, polynomial regression, logistic regression, and ridge regression. Regression analysis requires the collection of data and the fitting of a mathematical model. Model selection and evaluation are important steps that can be used to test the significance of the effect of the independent variable on the dependent variable and to perform parameter estimation and hypothesis testing. In this paper, regressivity analysis is used to screen some anomalous data and to explain the anomalies and make preliminary predictions.

In regressivity analysis, the following formula is often used:

$$y = \sum_i \alpha_i x_i \tag{1}$$

For primary function regression, on the other hand, only calculations are required:

$$y = a + bx, b = \frac{\sum(x_i - \bar{x})(y_i - \overline{y)}}{\sum(x_i - \bar{x})^2} \tag{2}$$

According to the primary function regression, there are four groups of data showing good regressivity, which include GDP, primary industry, secondary industry and tertiary industry. According to the regressivity analysis, we can find that most of the data show good regressivity, indicating that most of the data do not have particularly serious flaws; while the crude oil consumption data appear a more prominent problem - in the data of 2020, there is a more serious implausibility.

4. Modelling

In solving the problem of what methods can be used to achieve carbon peaking and energy saving, we can use Principal Component Analysis to analyse all the variables. Principal Component Analysis (PCA) is a commonly used data dimensionality reduction technique used to transform a high dimensional data set into a low dimensional space. Its main purpose is to compress and downscale data by finding new features that preserve the maximum variance in the original data through linear transformation [8].

Firstly, we normalise the original indicator data by collecting p-dimensional random vectors and n samples:

$$= (x_1, x_2, \dots, x_p)^T,\ x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T, i = 1,2, \dots, n, n > p \tag{3}$$

Construct the sample array and perform the following normalisation transformation on the sample array:

$$Z_{ij} = \frac{x_{ij} - \overline{x_J}}{s_j}, i = 1,2, \dots, n; j = 1,2, \dots, p \tag{4}$$

Among them.

$$x_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}, s_j = \frac{\sum_{i=1}^{n}(x_{ij} - x_j)^2}{n-1} \tag{5}$$

Secondly, we find the matrix of correlation coefficients for the normalised matrix Z:

$$R = [r_{ij}]_p xp = \frac{Z^T Z}{n-1} \tag{6}$$

Among them.

$$r_{ij} = \frac{\sum z_{kj} \cdot z_{kj}}{n-1}, i, j = 1,2, \dots, p \tag{7}$$

Then, go ahead and solve the characteristic equation. This characteristic equation is the characteristic equation associated with the sample matrix R with the following equation:

$$|R - \lambda I_p| = 0 \tag{8}$$

Solving the equation gives p-eigenroots and determining the principal components determines the value of m according to the following equation in order to achieve more than 85 per cent utilization of the information:

$$\frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \geq 0.85 \tag{9}$$

And for each $\lambda_j, j = 1,2, \cdots, m$, we solve the system of equations $R_b = \lambda_j b$ to obtain the unit eigenvector b_j^0. Finally, we convert the standardised indicator variables into principal components:

$$U_{ij} = z_i^T b_j^0, j = 1,2, \dots, m \tag{10}$$

Where U_1 is the first principal component, U_2 is the second principal component and so on [9]. Finally, the principal components are weighted and summed to get the final evaluation value, and the weights are the variance contribution ratio of each principal component.

Through principal component analysis, we can get the most important features in the dataset and transform them into a new set of irrelevant variables called principal components. These principal components are linear combinations of the original features and they are arranged in decreasing order of importance in explaining the variance of the data. Principal component analysis helps us to reduce the dimensionality of the data while retaining as much information as possible [10]. With PCA, we can better understand the relationships between datasets, identify underlying patterns and structures, and provide valuable insights in tasks such as data visualisation, feature selection and predictive modelling.

5. Model Solution

Firstly, we perform the KMO and Bartlett's test, as shown in Table 1, to determine whether principal component analysis can be performed.

Table 1. KMO and Bartlett's test

KMO value		0.299
Bartlett's test of sphericity	approximate chi-square (math.)	221.94
	df	36
	P	0.000***
Note: ***, **, * represent 1 per cent, 5 per cent and 10 per cent significance levels, respectively.		

The result of KMO test shows that the value of KMO is 0.299, meanwhile, the result of Bartlett's spherical test shows that the significance p-value is 0.000, which presents significance at the level, rejecting the original hypothesis that there is a correlation between the variables and that the principal component analysis is valid [11].

Secondly, we can draw and analyse the variance explained table whose results are shown in Table 2.

Table 2. Explanation of variance table

Ingredient	Characteristic root	Explanation of variance (%)	Cumulative variance explained (%)
1	5.926	65.839	65.839
2	1.548	17.199	83.038
3	1.017	11.299	94.337
4	0.443	4.917	99.254
5	0.053	0.587	99.841
6	0.013	0.147	99.988
7	0.001	0.01	99.998
8		0.002	100
9			100

In the variance explained table, the eigenroot of total variance explained is lower than 1 at principal component 4, and the contribution rate of variable explanation reaches 99.254. Then the importance of hidden variables in each principal component can be analysed by analysing the principal component loading coefficients [12]. The specific results are shown in Table 3.

Table 3. Table of factor loading coefficients

Name	Factor loading factor				Commonality (common factor variance)
	Main component 1	Main Component 2	Main Component 3	Main Component 4	
Total Population	0.935	-0.174	-0.204	-0.174	0.977
GDP	0.996	0.079	-0.026	-0.009	0.999
Primary Industry	0.967	0.16	-0.09	0.156	0.993
Secondary sector	0.983	0.158	0.064	0.022	0.996
Tertiary sector	0.995	0.024	-0.074	-0.043	0.998
Average Winter Temperature	0.314	-0.792	0.228	0.47	0.999
Crude oil consumption	0.397	-0.438	0.725	-0.351	0.998
Crude oil price	-0.13	0.748	0.615	0.193	0.992
Carbon dioxide emissions	0.945	0.286	0.037	0.06	0.98

We can find the commonality of the variables between the two principal components, which facilitates us to analyse and solve the problem. We find that carbon dioxide emission, primary, secondary and tertiary industries, GDP, and total population are the variables with greater correlation, while crude oil consumption, average winter temperature, and crude oil price among the opposite variables are also the variables with greater correlation. And then we get the component matrix as shown in Table 4:

Table 4. Component matrix

Name	Ingredient			
	1	2	3	4
Total Population	0.158	-0.113	-0.2	-0.393
GDP	0.168	0.051	-0.026	-0.019
Primary Industry	0.163	0.103	-0.088	0.353
Secondary sector	0.166	0.102	0.063	0.051
Tertiary sector	0.168	0.015	-0.073	-0.096
Average Winter Temperature	0.053	-0.512	0.224	1.063
Crude oil consumption	0.067	-0.283	0.713	-0.793
Crude oil price	-0.022	0.483	0.605	0.437
Carbon dioxide emissions	0.159	0.185	0.036	0.137

Based on the component matrix table, we get the formula for the model:

$$F1 = 0.158 \times P + 0.168 \times G + 0.163 \times F + 0.166 \times S + 0.168 \times T$$
$$+0.053 \times D + 0.067 \times H - 0.022 \times V + 0.159 \times C \tag{11}$$
$$F2 = -0.113 \times P + 0.051G + 0.103 \times F + 0.102 \times S + 0.015 \times T$$
$$-0.512 \times D - 0.283 \times H + 0.483 \times V + 0.185 \times C \tag{12}$$
$$F3 = -0.2 \times P - 0.026 \times G - 0.088 \times F + 0.063 \times S - 0.073 \times T$$
$$+0.224 \times D + 0.713 \times H + 0.605 \times V + 0.036 \times C \tag{13}$$
$$F4 = -0.393 \times P - 0.019 \times G + 0.353 \times F + 0.051 \times S - 0.096 \times T$$
$$+1.063 \times D - 0.793 \times H + 0.437 \times V + 0.137 \times C \tag{14}$$

It can be obtained from the above:

$$F = \left(\frac{0.658}{0.993}\right) \times F1 + \left(\frac{0.172}{0.993}\right) \times F2 + \left(\frac{0.113}{0.993}\right) \times F3 + \left(\frac{0.049}{0.993}\right) \times F4 \tag{15}$$

The formula expresses the meaning of the energy saving and emission reduction issues, each year in a certain weight under the integrated quantitative score. And then we get the factor weight analysis as shown in Table 5:

Table 5. Factor weight analysis

Name	Explanation of variance (%)	Cumulative variance explained (%)	Weights (%)
Main Component 1	0.658	65.839	66.334
Main Component 2	0.172	83.038	17.328
Main Component 3	0.113	94.337	11.384
Main Component 4	0.049	99.254	4.954

The above table shows the principal component weight analysis of the principal component analysis based on the information such as loading coefficients, which is calculated by the formula of variance explained/rotated cumulative variance explained. The calculation results show that the weight of principal component 1 is 66.334%, the weight of principal component 2 is 17.328%, the weight of principal component 3 is 11.384%, and the weight of principal component 4 is 4.954%, where the maximum value of the indicator weight is principal component 1 (66.334%), and the minimum value of the weight is principal component 4 (4.954%). Finally, we get a composite score as shown in Table 6:

Table 6. Composite Score Table

Ranking	Year	Composite Score	Main Component 1	Main Component 2	Main Component 3	Main Component 4
1	2022	1.279	1.458	1.062	0.547	1.32
2	2021	0.968	1.378	-0.224	0.908	-0.204
3	2020	0.306	0.581	-0.647	0.508	-0.51
4	2019	0.178	0.541	0.632	-2.62	0.173
5	2018	0.066	0.132	0.344	0.159	-2.01
6	2017	-0.311	-0.197	-0.815	-0.076	-0.616
7	2014	-0.512	-1.085	0.51	0.632	0.964
8	2016	-0.593	-0.492	-1.785	-0.164	1.229
9	2013	-0.66	-1.466	1.545	0.435	-0.097
10	2015	-0.722	-0.85	-0.622	-0.329	-0.249

The comprehensive score is sorted in descending order according to the comprehensive score obtained from the F-value calculation, and the comprehensive score and ranking of each year can be obtained, and the results after factor condensation are output at the same time. The relevant results are shown in Figure 1:

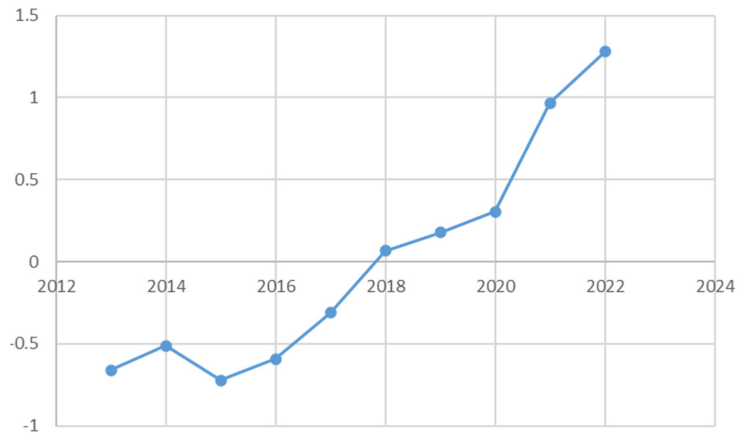

Fig 1. Results of composite scores for each year

We can clearly find that the index growth of each year is roughly in line with the yearly trend, so this modelling is successful and the principal component analysis is valid. Therefore, according to similarity of the two values of each principal component 1 and principal component 2, we can determine the factors related to energy saving and emission reduction as shown in Table 7:

Table 7. Relevant factors

Class	Relevant elements
Energy saving	Independent Variables: Crude Oil Price, Crude Oil Consumption (Negative); Dependent Variable: Winter Temperature
Emission reduction	Independent Variables: GDP, Primary Sector, Secondary Sector, Tertiary Sector, Population Size Dependent variable: CO_2 emissions

6. Conclusion

The application of Principal Component Analysis (PCA) in environmental policy planning has proved to be very effective in identifying and assessing key factors for carbon peaking and emission reduction. By meticulously collecting and pre-processing a decade's worth of data, including GDP (primary, secondary and tertiary sectors), population size, average temperature, oil consumption and prices, and carbon dioxide emissions, the study highlights the complexity of the environmental challenge and the multifaceted nature of the factors involved. Regression analysis, a preliminary step in this study, helped to identify patterns and anomalies in the data, highlighting in particular the problems present in the 2020 oil consumption data, thus emphasising the need for an accurate environmental assessment using powerful analytical tools such as PCA.

The strength of PCA lies in its ability to reduce the dimensionality of complex datasets while retaining the most important information. This is crucial in environmental policy planning, where numerous factors interact to influence outcomes such as carbon emissions and energy consumption. By normalising the raw data and constructing a correlation matrix, the PCA method helps to extract the principal components that represent the maximum variance in the dataset. The results of the study, particularly through variance interpretation and factor loading, revealed the significant impact of GDP, industrial sector, population and oil-related variables on environmental indicators such as carbon dioxide emissions and energy consumption.

The KMO and Bartlett's test verified the applicability of PCA to the dataset, indicating the importance of the relationship between the variables. In addition, the explained ANOVA and dendrograms guided the selection of principal components, ensuring that more than 85% of the information was retained. This strategic reduction in data complexity allowed for the clear identification of the factors that have the greatest impact on emissions reductions and energy savings, including the negative correlation between oil consumption and winter temperatures, and the positive correlation between economic activity and carbon dioxide emissions.

By weighting the principal components according to the explained variance, a composite model was developed to quantify the composite scores for each year of the study. The model not only facilitates an in-depth understanding of time trends in environmental factors, but also serves as a powerful framework for policy analysis and planning. The weighted analysis further emphasises the dominant role of the first principal component, reflecting the overall impact of demographic and economic activities on environmental outcomes.

In summary, the application of PCA in this context demonstrates its far-reaching role in distilling complex environmental data into actionable insights. It highlights the interconnections between economic activity, population dynamics and environmental impacts, providing a nuanced perspective for policy development. The methodology enables the identification of key drivers of carbon emissions and energy consumption, providing a solid basis for targeted policy interventions aimed at achieving sustainable environmental outcomes. The analysis not only enriches the methodological toolkit for environmental policy planning, but also demonstrates the critical role of statistical techniques in addressing contemporary environmental challenges.

7. Discussion

The use of Principal Component Analysis (PCA) in environmental policy planning provides a sophisticated methodology for dissecting the intricate web of factors that influence carbon peaking and emissions reductions, as evidenced by a decade of meticulous data collection and pre-processing. This includes a variety of variables such as gross domestic product (GDP) across different sectors, population figures, average temperatures, oil consumption and prices, and carbon dioxide emissions, highlighting the complex interplay between economic, demographic and environmental factors. The initial step of regression analysis is not only a tool for understanding the relationships between variables, but also a means of detecting anomalies in the dataset, such as the significant discrepancies in the 2020 oil consumption data, which underlines the urgent need for advanced analytical methods such as PCA in ensuring accurate environmental assessments.

PCA can reduce the complexity of high-dimensional datasets while retaining the most relevant information, which is critical in environmental policy where multiple factors combine to influence outcomes such as carbon emissions and energy consumption. By normalising the data and constructing correlation matrices, PCA can help extract the principal components with the highest variance in the dataset, providing insight into the impact of GDP, industrial sector, population and oil-related variables on key environmental indicators. Validation of the applicability of PCA through KMO and Bartlett's test further emphasises its validity, with interpreted ANOVA and dendrograms guiding the selection of principal components to ensure that more than 85% of the data information is retained.

This strategic reduction in data complexity not only facilitates the identification of important factors influencing emissions reductions and energy savings, but also highlights the inverse relationship between oil consumption and winter temperatures, as well as the direct correlation between economic activity and CO2 emissions. A composite model, weighted according to the variance explained by each principal component, quantified the composite score for each year studied. This not only deepens the understanding of time trends in environmental factors, but also establishes a robust framework for policy analysis and planning, highlighting the primary role of the first principal component in capturing the overall impact of demographic and economic activities on environmental outcomes.

In summary, the application of PCA to environmental policy planning emphasises its powerful ability to distil complex environmental data into actionable insights, highlighting the interconnections between economic activity, population dynamics and environmental impacts. The methodology helps to identify the key drivers behind carbon emissions and energy consumption, providing a solid foundation for targeted policy interventions to achieve sustainable environmental outcomes. The analysis enriches the toolkit available for environmental policy planning, demonstrates the critical role of statistical techniques in addressing contemporary environmental challenges, and proves the valuable contribution of PCA in enhancing informed decision-making and effective policy formulation in the field of environmental governance.

References

[1] Wang, T., Zhang, F., Gu, H., Hu, H., & Kaur, M. (2023). A research study on new energy brand users based on principal component analysis (PCA) and fusion target planning model for sustainable environment of smart cities. Sustainable Energy Technologies and Assessments, 57, 103262.

[2] Mohammadnazari, Z., Aghsami, A., & Rabbani, M. (2023). A hybrid novel approach for evaluation of resiliency and sustainability in construction environment using data envelopment analysis, principal component analysis, and mathematical formulation. Environment, Development and Sustainability, 25(5), 4453-4490.

[3] Zhang, X., & Dong, F. (2023). What affects residents' behavioral intentions to ban gasoline vehicles? Evidence from an emerging economy. Energy, 263, 125716.

[4] Wang, Z., Liang, F., Li, C., Xiong, W., Chen, Y., & Xie, F. (2023). Does China's low-carbon city pilot policy promote green development? Evidence from the digital industry. Journal of Innovation & Knowledge, 8(2), 100339.

[5] Ma, S., Huang, Y., Liu, Y., Kong, X., Yin, L., & Chen, G. (2023). Edge-cloud cooperation-driven smart and sustainable production for energy-intensive manufacturing industries. Applied Energy, 337, 120843.

[6] Xiao, J., Gao, D., Zhang, H., Shi, H., Chen, Q., Li, H., ... & Chen, Q. (2023). Water quality assessment and pollution source apportionment using multivariate statistical techniques: A case study of the Laixi River Basin, China. Environmental Monitoring and Assessment, 195(2), 287.

[7] He, X., Khan, S., Ozturk, I., & Murshed, M. (2023). The role of renewable energy investment in tackling climate change concerns: Environmental policies for achieving SDG-13. Sustainable Development.

[8] Ismaeel, W. S., & Lotfy, R. A. E. R. (2023). An integrated building information modelling-based environmental impact assessment framework. Clean Technologies and Environmental Policy, 25(4), 1291-1307.

[9] Fonseca, A., Fraga, H., & Santos, J. A. (2023). Exposure of Portuguese viticulture to weather extremes under climate change. Climate Services, 30, 100357.

[10] Okeke, D. C., Obasi, O., & Nwachukwu, M. U. (2023). Analysis of road transport response to COVID-19 pandemic in nigeria and its policy implications. Transportation Research Record, 2677(4), 851-864.

[11] Peng, J., Wen, L., Mu, X., & Xiao, J. (2023). The evolving centres of gravity in China's oil and gas industry: Evidence from infrared radiation imaging gas flaring data. Energy for Sustainable Development, 73, 263-279.

[12] Bouaakkaz, B., El Morjani, Z. E. A., & Bouchaou, L. (2023). Social vulnerability assessment to flood hazard in Souss basin, Morocco. Journal of African Earth Sciences, 198, 104774.